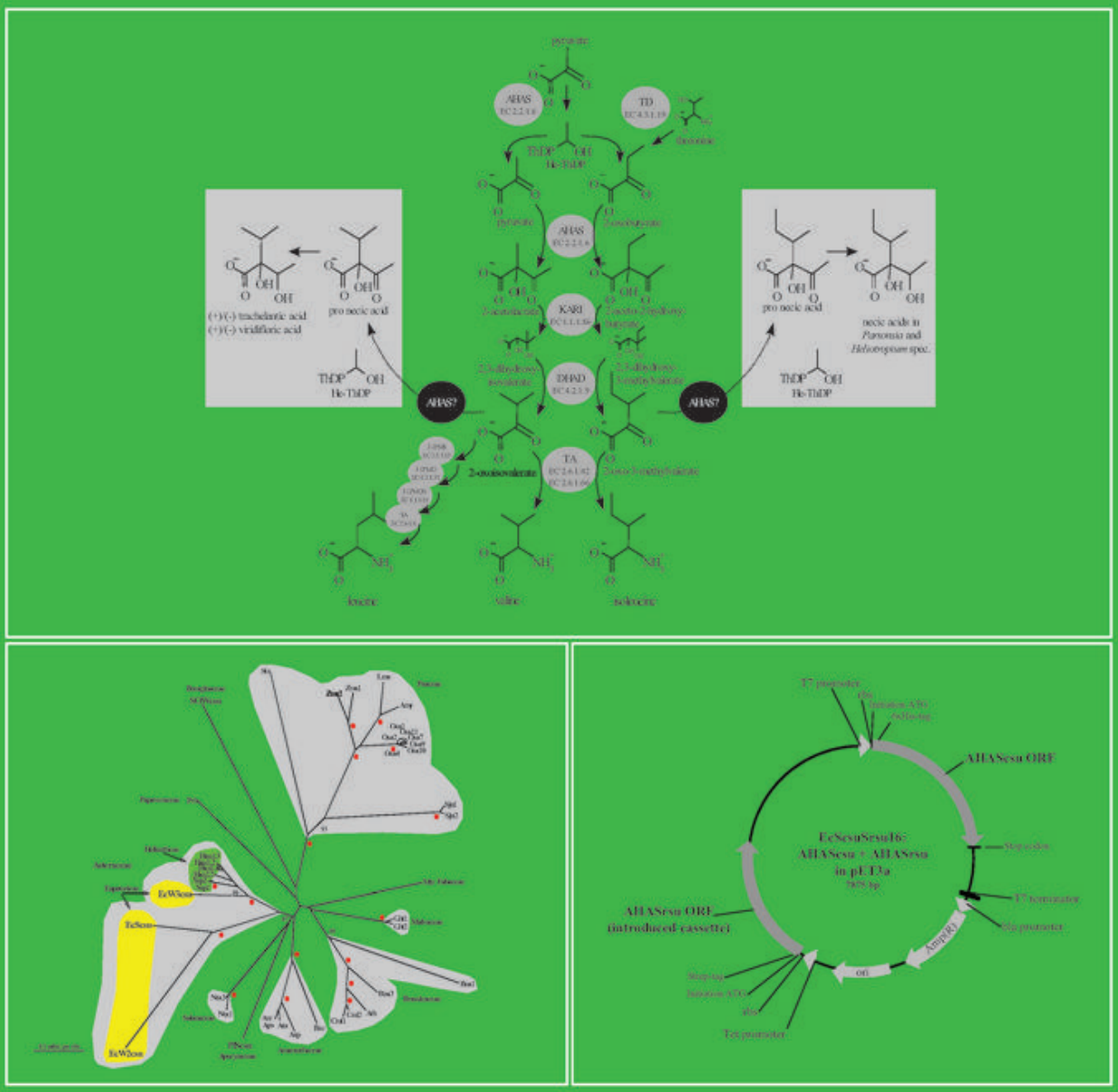
Dorothee Langel

Biosynthesis of the Unique Necic Acid Moiety in Lycopsamine Type Pyrrolizidine Alkaloids – a Molecular Approach –



Cuvillier Verlag Göttingen

Biosynthesis of the Unique Necic Acid Moiety in Lycopsamine Type Pyrrolizidine Alkaloids - a Molecular Approach -

Von der Fakultät für Lebenswissenschaften
der Technischen Universität Carolo-Wilhelmina
zu Braunschweig
zur Erlangung des Grades einer
Doktorin der Naturwissenschaften
(Dr. rer. nat.)
genehmigte

D i s s e r t a t i o n

von Dorothee Langel
aus Fulda

Bibliografische Information der Deutschen Nationalbibliothek
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.ddb.de abrufbar.
1. Aufl. - Göttingen : Cuvillier, 2008
Zugl.: (TU) Braunschweig, Univ., Diss., 2008

978-3-86727-646-7

1. Referent:	Professor Dr. Dietrich Ober
2. Referent:	Professor Dr. Thomas Hartmann
eingereicht am:	28.04.2008
mündliche Prüfung (Disputation) am:	20.06.2008

Druckjahr 2008

Nonnenstieg 8, 37075 Göttingen
Telefon: 0551-54724-0
Telefax: 0551-54724-21
www.cuvillier.de

1. Auflage, 2008
Gedruckt auf säurefreiem Papier

978-3-86727-646-7

Contents

Danksagung

An dieser Stelle danke ich allen sehr herzlich, die zur Entstehung der vorliegenden Arbeit beigetragen haben:

Prof. Dr. Dietrich Ober, für die Begeisterung weckende Einführung in Thema und Methodik, die ständige Bereitschaft zu Diskussion und Rat, die Übertragung von Verantwortlichkeit für "unsere" Diplomanden und das stete Vertrauen in meine Person;

Prof. Dr. Thomas Hartmann, für das in mich gesetzte Vertrauen, eine Lösung der Problemstellung zu erzielen, die stete Bereitschaft, über wirklich alle möglichen Lösungen nachzudenken, auch wenn sie der ursprünglichen Theorie diametral entgegen liefen und nicht zu letzt für die Übernahme des Co-Referats;

Dr. Till Beuerle, für die "Entwicklungshilfe" beim spezifischen Assay und die ständige Bereitschaft zur ultra schnellen GC-MS Vermessung der Proben;

Prof. Dr. Ronald G. Duggleby, University of Brisbane, Australia, for providing valuable information on recombinant AHAS and for supplying us with the pET GM construct harbouring the *E. coli* AHASII that was essentially helpful as a positive control for the enzyme assay;

Dr. Barbar J. Mazur and Dr. Robert A. LaRossa, Dupont, USA, for kindly providing us with the AHAS-deficient *E. coli* strain MF2000;

Prof. Dr. Jens Christoffers und Dipl.-Chemiker Michael Rössle, Carl von Ossietzky Universität Oldenburg (Niedersachsen) vormals Universität Stuttgart, für die Einladung zum Erlernen der Synthese der benötigten Referenzsubstanzen;

Anita Backenköhler, für die einleuchtende Einarbeitung, die ständige Bereitschaft, den Teufel im Detail zu finden und für die darüber hinaus entstandene außergewöhnliche Freundschaft;

allen Mitarbeitern des Instituts für Pharmazeutische Biologe der TU Braunschweig für das angenehme Arbeitsklima und die gute Zusammenarbeit; insbesondere Frau Dr. Hoda Mohagheghi, für die guten, erkenntnisreichen Stunden im Büro, bei Exkursionen und im Labor; Frau Dipl.-Apothekerin Dagmar Enß, für ihre unbeirrbare Zähigkeit und Ausdauer während ihrer Diplomarbeit und ihre Stützkraft bei noch so unerfreulichen Ergebnisquoten; Herrn Dipl.-Apotheker Uwe Genzel, für die tatkräftige Unterstützung bei den Expressionsversuchen im Rahmen seiner Diplomarbeit sowie Frau Apothekerin Jessica Steinert, für die Hilfe bei den katalytischen Untereinheiten.

allen Mitarbeitern der Arbeitsgruppe Ober am Botanischen Institut der Christian-Albrechts Universität zu Kiel für das herzliche Willkommen, der tatkräftigen Mithilfe bei Aufbau und Einrichtung von Labor und Arbeitsplätzen und für das außerordentlich gute Arbeitsklima.

1 Introduction

Plants produce a vast and diverse assortment of chemical compounds. These compounds can be classified as products or intermediates from primary and secondary metabolism. Primary metabolism is necessary for growth and development, is universal, uniform, and hardly changed during evolution. Examples of plant primary metabolites are simple carbonic acids and sugars, amino and nucleic acids (Croteau et al. 2000).

On the other hand, secondary metabolism is not necessary for growth and development but is essential for continued existence of a species within its environment (Hartmann 1985). Plant secondary metabolites serve as chemical signals that enable the sessile plant to respond to environmental cues. Secondary compounds can function in the defense against herbivores, pathogens, and competitors while others provide protection from sun radiation, aid in pollen and seed dispersal or can submit informations acting as pheromone like signals and phytohormones (Hartmann 1991, Harborne 1993).

Plant secondary compounds are derived from central primary metabolites and, based on their biosynthetic origins, they can be divided into three major groups: the terpenoids, the phenylpropanoids including phenolic compounds, and the alkaloids. Terpenoids originate from carbohydrate metabolism (carbonic acids and sugars) and functionally include toxins and feeding deterrents against herbivores like cardenolides or antibacterial, fungicide and viricidal mono- and sesquiterpenes in essential oils. In lower dosages, essential oil components are important to attract pollinators and seed dispersers. Those are visually attracted by many flavonoid pigments that belong to the phenolic compound group, e.g., anthocyanins can be responsible for red and blue, while chalcones count for yellow flower pigments. Besides this visual function, phenolic compounds like tannins, lignans, and flavonoids serve as defense against herbivores and pathogens. Moreover, lignins strengthen cell walls mechanically, and naphthaquinones like juglone have allelopathic activity and may adversely influence the growth of neighbouring plants.

The fascinating, chemical diverse group of alkaloids is synthesized from compounds of the amino and nucleic acid metabolism. These bitter-tasting, nitrogenous alkaloids protect plants from a variety of herbivorous animals, and many possess dramatic physiological effects on vertebrates including humans.

Most alkaloids are easily resorbed and many of them interfere with essential parts of the nervous system, acting on pre- and/or post-synaptic receptors, inactivating neurotransmitters, inhibiting the post-synaptic signal transduction pathways, or hampering a proper function of ion channels. Depending on the targeted nerve, this can result in paralysis of selected muscle tissues like skeleton muscles and heart muscles, or it can result in the paralysis of pain-conductive nerves providing an analgesic effect, or it can result in the paralysis of the central nervous system causing psychological effects. Moreover, some alkaloids interfere with the assembly or disassembly of essential structures within the cytoskeleton, hindering cell division and awarding some cytostatic effects while other

alkaloids are able to alkylate DNA probably causing cancer (Hänsel and Sticher 2004).

The first identified alkaloid was morphine from the latex of opium poppy (*Papaver somniferum*). Until today, morphine is used in medicine as an analgesic and cough suppressant. However, excessive use of this drug can lead to strong addiction and, if overdosed, it provokes respiratory paralysis and death.

More than 12,000 alkaloids have been isolated since the discovery of morphine by the German pharmacist Fiedrich Sertürner in 1806 and the questions, how and why alkaloids are made by plants, have fascinated generations of researchers within the fields of biology, chemistry, and pharmacy.

1.1 Pyrrolizidine Alkaloids - Typical Compounds of Plant Secondary Metabolism

Pyrrolizidine alkaloids (PAs) are characterized as typical compounds of plant secondary metabolism (Croteau et al. 2000). PAs do fulfill the four essential demands on secondary compounds as proposed by Hänsel and Sticher (2004).

1. Differential distribution among limited taxonomic groups within the plant kingdom

PAs were discovered in more than 6,000 plant species (Chou and Fu 2006) distributed among certain unrelated Angiosperm taxa. More than 95% of the PA-producing species investigated so far belong to just five families (marked with a red disc in Figure 1.1) namely the Asteraceae (1), the Boraginaceae (2), the Apocynaceae (3), the Orchidaceae (4), and the Fabaceae (5). Further isolated occurrences were reported from single species of other families (marked with a pink disc in Figure 1.1) like the Convolvulaceae (6), the Santalaceae (7), the Sapotaceae (8), the Ranunculaceae (9), the Celastraceae (10) (reviewed by Hartmann and Witte, 1995), and recently from the Lamiaceae (11) discovered by Nawaz et al. (2000).

The scattered occurrence of PAs among the Angiosperms has been particularly puzzling and provoked the question whether the PA biosynthetic pathway was invented just once and lost several times during evolution or whether it evolved several times independently in distinct, separate lineages. To date, this question was answered in favour of an independent evolution. Homospermidine synthase (HSS, EC 2.5.1.44), catalyzing the first step in PA biosynthesis, was found to be of polyphyletic origin within the Angiosperms (Ober and Hartmann 1999b). Up to now, the HSS has been shown to be recruited at least four times independently, once early in the evolution of the Boraginaceae, once within the monocots, and twice within the Asteraceae separating the tribes of Senecioneae and Eupatorieae (Reimann et al. 2004).

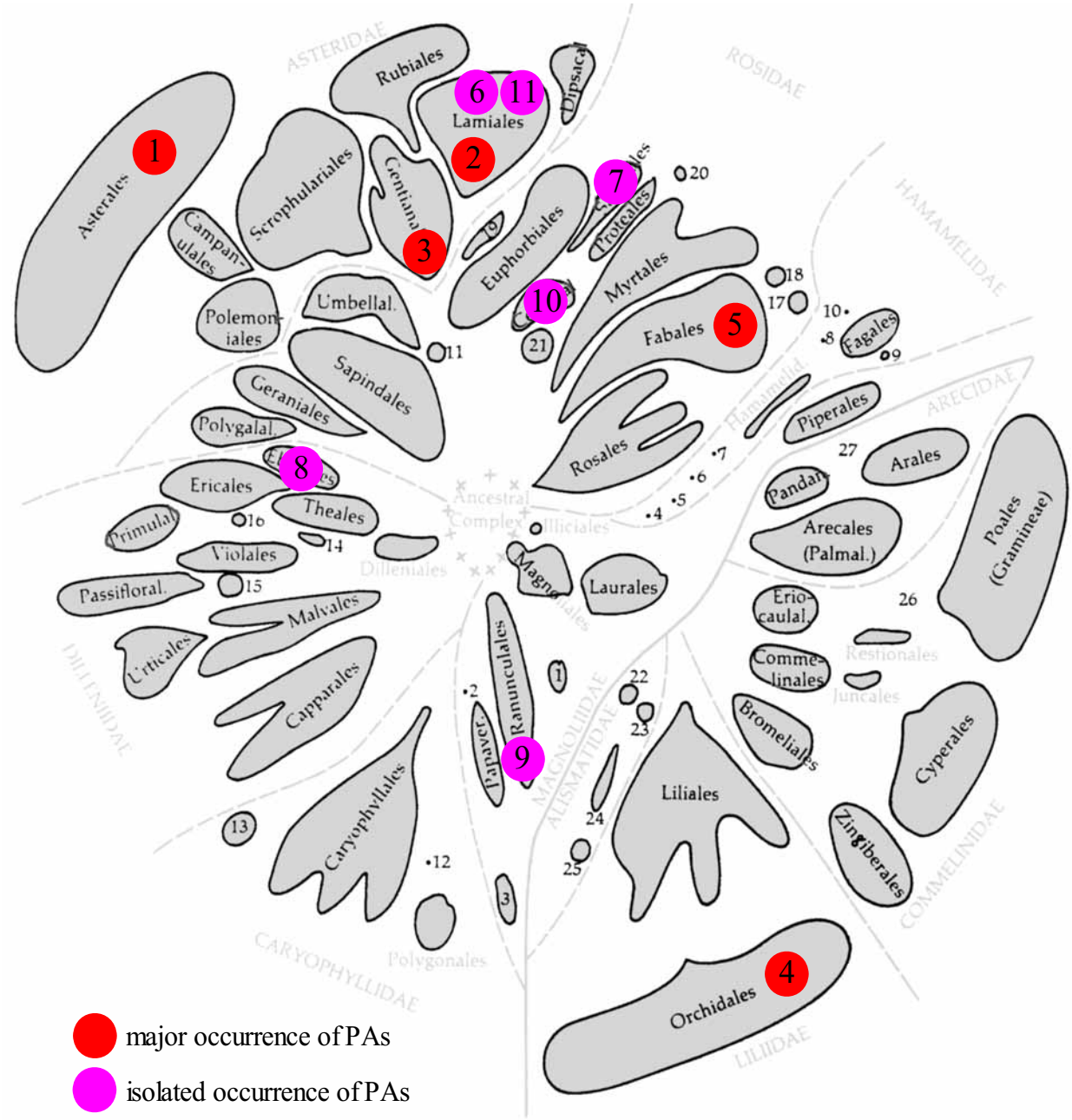


Figure 1.1: Major and isolated occurrences of pyrrolizidine alkaloids within the angiosperms, modified from Ober and Hartmann 1999b and Stebbins 1974: Asteraceae (1), Boraginaceae (2), Apocynaceae (3), Orchidaceae (4), Fabaceae (5), Convolvulaceae (6), Santalaceae (7), Sapotaceae (8), Ranunculaceae (9), Celastraceae (10), and Lamiaceae (11).

2. Great structural diversity and complexity

More than 660 different chemical structures of PAs are known so far (Jiang et al. 2006). Basically, plant PAs are composed of a "necine base" moiety that is esterified with one or more "necic acids" (Panel B of Figure 1.2, Crout 1966).

In 1995, the main PA-types (see Figure 1.2) comprising senecionine, triangularine, monocrotaline, lycopsamine, and phalaenopsine type PAs were described by Hartmann and Witte (1995) based on Culvenor's chemosystematic suggestions (Culvenor 1978). The growing number of discovered PA-structures as well as the increased understanding of the mechanisms involved in biosynthesis, biological activities and functions of PAs led to a new modified classification introduced by Hartmann (2006) which emphasizes the biosynthetic relationship within the necic acids:

I. Macrocyclic or open-chain diester types with eleven or more members including their related monoesters comprising senecionine (**S**), triangularine (**T**), and monocrotaline (**M**) types esterified with isoleucine derived necic acids (Panel I in Figure 1.2).

II. Lycopsamine (**L**) type PAs occurring as open-chain mono- and diesters, and macrocyclic triesters esterified with a unique C7 necic acid (Panel II in Figure 1.2).

III. Special type PAs characteristic to selected genera enclosing ipanguline (**I**) and phalaenopsine (**P**) types, occurring as mono- and open-chain di- and triesters with aryl, aralkyl, and rarely alkyl necic acids (Panel III in Figure 1.2).

The distribution of PA types within the Angiosperm families is shown in Table 1.1. Remarkably, the alkaloid pattern within the Asteraceae differs significantly between its two PA-producing tribes, the Senecioneae and the Eupatorieae. While the vast majority of PA types within the Eupatorieae belongs to the lycopsamine (L) type, PAs within the Senecioneae are exclusively found in class I with an accumulation of senecionine (S) types. These distinct alkaloid patterns are reflected in the finding that the homospermidine synthase (HSS) catalyzing the first step in PA biosynthesis was recruited independently within the Eupatorieae and the Senecioneae. Moreover, the inability of Senecioneae species to produce lycopsamine type PAs leads to the hypothesis that the HSS might not be the only enzyme specialized in PA biosynthesis. The formation of lycopsamine type PAs requires a second specialized enzyme that is important for the formation of the unique C7 necic acids characterizing the lycopsamine type PAs. Thus, this second specialized enzyme might not be present in the Senecioneae, the Orchidaceae, the Convolvulaceae (*Ipomoea*), Ranunculaceae, Celastraceae, and Lamiaceae but within the Eupatorieae as well as in the Boraginaceae, the Apocynaceae, the Santalaceae, the Sapotaceae, the Convolvulaceae (*Merremia*), and in the Fabaceae (*Laburnum anagyroides*).

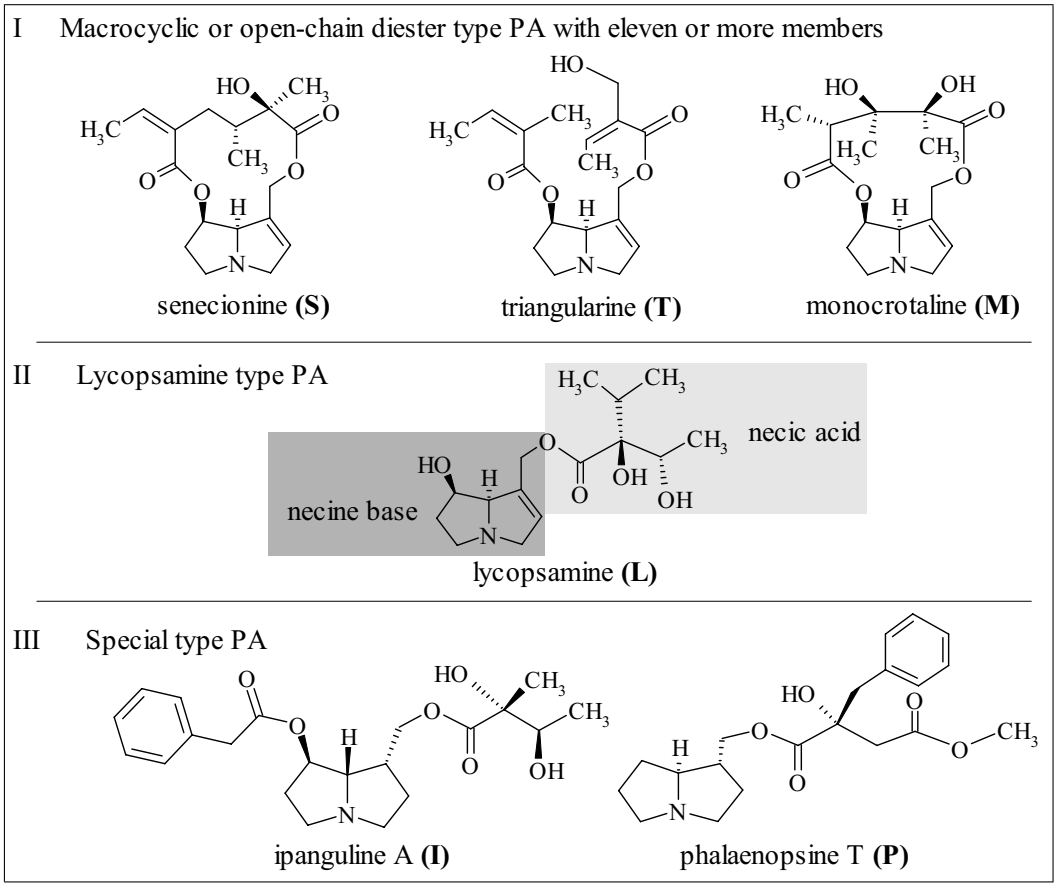


Figure 1.2: The three classes of pyrrolizidine alkaloids (PAs): **I.** Macrocyclic or open-chain diesters comprising the senecionine (S), triangularine (T), and monocrotaline (M) type PAs; **II.** Lycopsamine (L) type PAs; and **III.** Special type PAs enclosing the ipanguline (I) and phalaenopsine (P) type PAs.

family of PA occurrence	prevalent genera (Hartmann and Witte 1995)	distribution of PA types [%]						
		I			II	III		others
		S	**T**	**M**	**L**	**P**	**I**	
Asteraceae, tribe Senecioneae	*Senecio*	82	16	1	-	-	-	1
Asteraceae, tribe Eupatorieae	*Eupatorium*, *Ageratum*	8	5	-	87	-	-	
Boraginaceae	virtually all genera[1]	< 1	7	1	90	1	-	
Apocynaceae	*Parsonsia*	4	-	-	77	4	-	15
Orchidaceae	subfamily of Epidendroideae[2]	-	-	-	-	95	-	5
Fabaceae	*Crotalaria*	39	2	45	1	-	-	13
Convolvulaceae	*Ipomoea*, *Merremia*[3]	-	-	-	- / 100	-	100 / -	
Santalaceae	*Thesium*, *Amphorogyne*[4]	-	-	-	14	86	-	
Sapotaceae	*Planchonella*, *Minusops*	-	-	-	13	75	-	12
Ranunculaceae	*Caltha*, *Trollius*[5]	100	-	-	-	-	-	
Celastraceae	*Bhesa*	-	100	-	-	-	-	
Lamiaceae	*Ajuga*[6]	100	-	-	-	-	-	

[1]Röder (1995), [2]Frölich et al. (2006), [3]Mann et al. (1996), [4]Thu Huong et al. (1998), [5]Liddell and Stermitz (1994), [6]Nawaz et al. (2000)

Table 1.1: The distribution of PA types within the Angiosperm families.

3. Characteristic supply with PAs depending on the developmental stage

PA biosynthesis in the Asteraceae is strictly coordinated to root growth (Hartmann et al. 1988, Sander and Hartmann 1989) and is terminated when flowers open (Anke 2004). PA backbone structures like senecionine *N*-oxide are formed constitutively and under high constraint. Their biosynthesis is not inducible by wounding nor by microbial attack (van Dam and Vrieling 1994, Tinney et al. 1998). Additionally, PA backbones are diversified via specific one- or two-step reactions such as hydroxylation, epoxidation, O-acetylation, or otonecine formation which proceed in a molecule position-specific and stereoselective manner (Hartmann and Dierich 1998). Any genetic variability affecting the PA transforming enzymes modifies the PA bouquet without affecting the overall quantity. Such a highly plastic system may indicate a powerful strategy in constitutive plant defense against differential herbivory (Hartmann 1999, Hagen 2003).

4. Biosynthesis in specialized tissues, accumulation in unique, defined patterns

PAs are synthesized in specific tissues (see Table 1.2) and are accumulated in cell vacuoles of various tissues (Ehmke et al. 1987) with up to 80% of total plant PA content. Such accumulation tissues are important for species or individual survival, e.g., buds, seeds or roots (Hartmann and Zimmer 1986). Translocation from tissues of biosynthesis to those of accumulation is enabled by PAs existing as hydrophilic *N*-oxides (Figure 1.3) which allows transport via phloems (Hartmann et al. 1989, Witte et al. 1990). A specific PA *N*-oxide carrier, responsible for selective uptake into the vacuoles, has been characterized (Ehmke et al. 1988). Phloem loading and unloading is still under investigation and predicted to be carrier-mediated since species, which do not produce PAs, are unable to translocate PAs via the phloem (Hartmann et al. 1989). Stored PAs are stably retained and show no metabolic turnover except for PA-protected seedlings from *Crotalaria scassellatii* using a PA-specific *N*-oxygenase to start the mobilization of the bound nitrogen utilized for growth and development (Chang 1997).

1.1.1 Toxicity of Pyrrolizidine Alkaloids and Role in Plant Protection

In mammals, PAs are hepatotoxic and carcinogenic (Mattocks 1972, Culvenor et al. 1976, Wiedenfeld and Röder 1984, Steenkamp et al. 2001). The toxic compound is formed after ingestion via bioactivation by microsomal liver cytochrome P450 monooxygenases producing unstable pyrrolic intermediates (Figure 1.3). These highly reactive intermediates are only formed when the alkaloid substrates display the following essential features (Winter and Segall 1989): C1-C2 double bond (boxed area A in the pro-toxic PA structure in Figure 1.3), esterification of the allylic hydroxyl group at C9 (boxed area B), and free or esterified hydroxyl group at C7 (boxed area C). The necic acids are cleaved off and the remaining nucleophilic necine base derived compound is able to alkylate DNA (Fu et al. 2004). The attacked liver develops necrosis, fibrosis, and eventually neoplastic growth. The loss of liver function means an overflow of brain toxic substances like ammonia and

families	organs of PA synthesis	tissues of PA synthesis[1]	organs of highest[2] PA accumulation
Asteraceae, tribe Senecioneae	roots	specific groups of endodermis + adjacent cortex cells opposite the phloem[3]	infl.[4] (90%)
Asteraceae, tribe Eupatorieae	roots	cortex of annually sprouted roots[5]	infl.[6] (90%)
Boraginaceae	Sof roots[7]	endodermis[8]	?
	Cof shoots[9], roots[8]	endodermis + pericycle[8]	?
	Hin shoots[7], leaves[8] + buds[8]	epidermis of shoots + leaves[8]	infl. (71%)[7]
Orchidaceae	aerial root tips	basic meristem[10]	buds[11] (52%)
	buds	epidermis[10]	
Fabaceae	roots and/or leaves and/or shoot tips[12]	nd	seeds[13]
Convolvulaceae	shoots[14], roots[15]	nd	young leaves + shoot tips (60%)[14]

[1]deduced from the localization of the first pathway specific enzyme, the homospermidine synthase (HSS), [2]PAs were found in all parts of the plants
[3]Moll et al. (2002), [4]Hartmann and Zimmer (1986), [5]Anke et al. (2004), [6]Biller et al. (1994), [7]Frölich et al. (2007), [8]Niemüller (2007), [9]Van Dam et al. 1995, [10]Anke (2004), [11]Frölich et al. (2006), [12]Nurhayati and Ober (2005), [13]Toppel et al.(1988), [14]Jenett-Siems et al. (1998), [15]Jenett-Siems et al. (2005)

Table 1.2: Plant organs and tissue of PA biosynthesis vary significantly among the PA producing families or even genera while plant organs of PA accumulation are preferentiably reproductive tissues. Abbreviations: infl. = inflorescence (includes buds, flowers + fruits), Sof = *Symphytum officinale*, Cof = *Cynoglossum officinale*, Hin = *Heliotropium indicum*, nd = not determined, ? = unfortunately, most phytochemical reports do not contain any quantitive information on PA levels or even specifications of the analyzed plant organs.

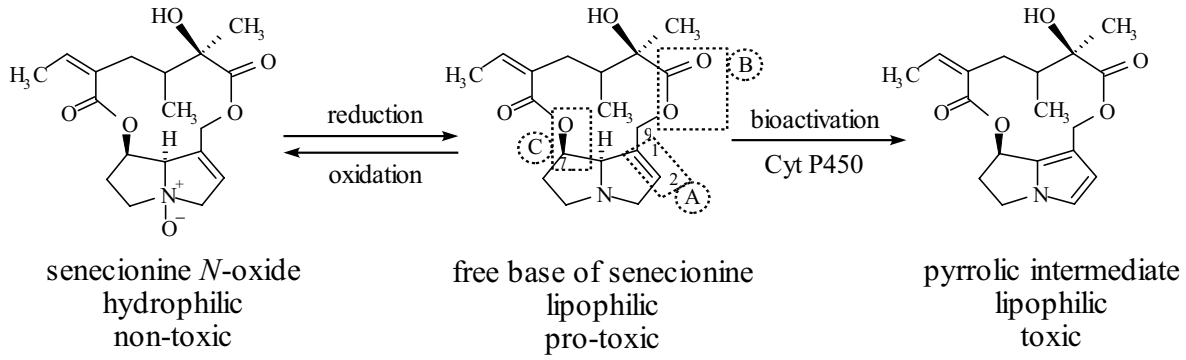


Figure 1.3: In plants, pyrrolizidine alkaloids exist in two molecular forms: the hydrophilic, non-toxic N-oxide and the lipophilic, pro-toxic free base. During bioactivation in vertebrates, the toxic pyrrolic intermediate is generated by liver microsomal monooxygenases. This toxic form is only produced when the pro-toxic form displays the following essential features: C1-C2 double bond (boxed area A), esterification of the allylic hydroxyl group at C9 (boxed area B), and free or esterified hydroxyl group at C7 (boxed area C).

results in a destroyed central nervous system, in an inability to breathe and finally in death. The potential health hazard for humans caused by unsaturated PAs was the reason for the Federal Health Department of Germany (Kommission E 1992) to recommend a limited use of leaves and buds from coltsfoot (*Tussilago farfara*) in bronchial tea (maximum use is 6 weeks per year with a daily uptake of 1 μg), in the use of roots from comfrey (*Symphytum officinale*) as extracts in ointments treating distortions and contusions (with external use on intact skin, daily dose is 100 μg of PAs), and of butterbur extracts (*Petasitis hybridus*) used for the treatment of migraine (petasitenine daily allowance is 0.1 μg).

The described effects on mammals work very slow and do not prevent the individual plant from being eaten. Hence, reported livestock intoxications with PA-producing plants (Cheeke 1989) might have been interpreted rather as a man-made accident with locked up cattle forced to feed on run-down pastures. Grazing animals, however, usually avoid PA plants, apparently because of their deterrent taste (Boppré 1986).

Considering herbivorous insects, PAs are hypothesized to play a protective role in plant defense as well. Most plant secondary compounds are thought to be co-evolved in the cretaceous period (130 million years before present) during the appearance and rapid diversification of both Angiosperms and insects which appeared to Darwin (Darwin and Seward 1903) as an "abominable mystery". While the Angiosperm-specific flower attracted the facultative herbivorous insect, essential for pollination, it was necessary for the plant to ensure intact tissues important for species survival and reproduction (Fraenkel 1959). Hence, protective PAs are cytotoxic for adult insects and mutagenic for larvae and pupae (Frei et al. 1992). Within one generation, the herbivore attack is completely stopped which is an optimal result for the plant.

The most convincing evidence favoring an anti-feeding role of plant PAs can be observed from specialized herbivorous insects. A number of these insects from diverse taxa have evolved adaptations not only to overcome the defensive barrier of PA-protected plants, but to sequester and utilize PAs for their own defense against predators. Plant-derived PA-sequestering species are found among butterflies and moths (Lepidoptera), leaf beetles (Coleoptera), aphids (Homoptera), and the African grasshopper *Zonocerus* (Orthoptera). These plant PA sequestering insect species advertise their unpalatability to potential insect predators by conspicuous warning coloration (Hartmann 1999) with the effect that predators with previous contact to the taste of PAs avoid preying on aposematic coloured insects (Hare and Eisner 1993).

The giant tropical orb-weaving spider *Nephila clavipes* (Araneae, Tetragnathinae), for instance, liberates PA sequestering adults of lepidopteran Ithomiinae, Danainae, and Arctiidae unharmed from its web (Belt 1888, Vasconcellos-Neto and Lewinsohn 1984, Eisner 1982). When PA-free, dead and wing-less workers of honey bee *Apis mellifera* (Apidae) were coated topically with methanolic solutions of various PAs and tossed onto the web of a female spider, the potential prey was released from the web depending on the type and amount of encoated PAs (Silva and Trigo 2002).

Besides of their function as antifeeding compounds, PAs play a role in mating behavior. In the arctiid species *Utetheisa ornatrix,* the content of PAs inside a male evolved as a criterion for attractiveness (Boppré 1986, Boppré 1990). PAs are reconstructed by males to the pheromone hydroxydanaidal and advertised to females that now can estimate the total systemic PA-load of males (Dussourd et al. 1991). During copulation PAs are transmitted from male to female. The eggs receive both the male and the female load and are protected in the best way their parents were able to do (Eisner et al. 2002).

1.1.2 Pyrrolizidine Alkaloid Biosynthesis

Pyrrolizidine alkaloid (PAs) biosynthesis was shown to root in the amino acid biosynthesis. Proven and proposed parts of the principle metabolic pathway of the lycopsamine type PAs are shown in Figure 1.4.

1.1.2.1 Necine Base Biosynthesis

At first sight, necine bases reveal the identity of PAs due to their characteristic bicyclic ring system containing a nitrogen at position 4 within the 1-hydroxymethyl pyrrolizidine (see Figure 1.3).

Tracer studies with ^{14}C-labelled putative precursors (Hartmann and Toppel 1987, Sander and Hartmann 1989) revealed a pathway that starts from arginine which is decarboxylated to agmatine and then transformed to *N*-carbamoyl putrescine (CAP, Figure 1.4). The amido group is removed and putrescine is released. It can be converted to spermidine and reverted to putrescine again. With the help of homospermidine synthase (HSS), homospermidine is produced by transferring the aminobutyl moiety from spermidine to putrescine, releasing diaminopropane (Böttcher et al. 1993, Ober and Hartmann

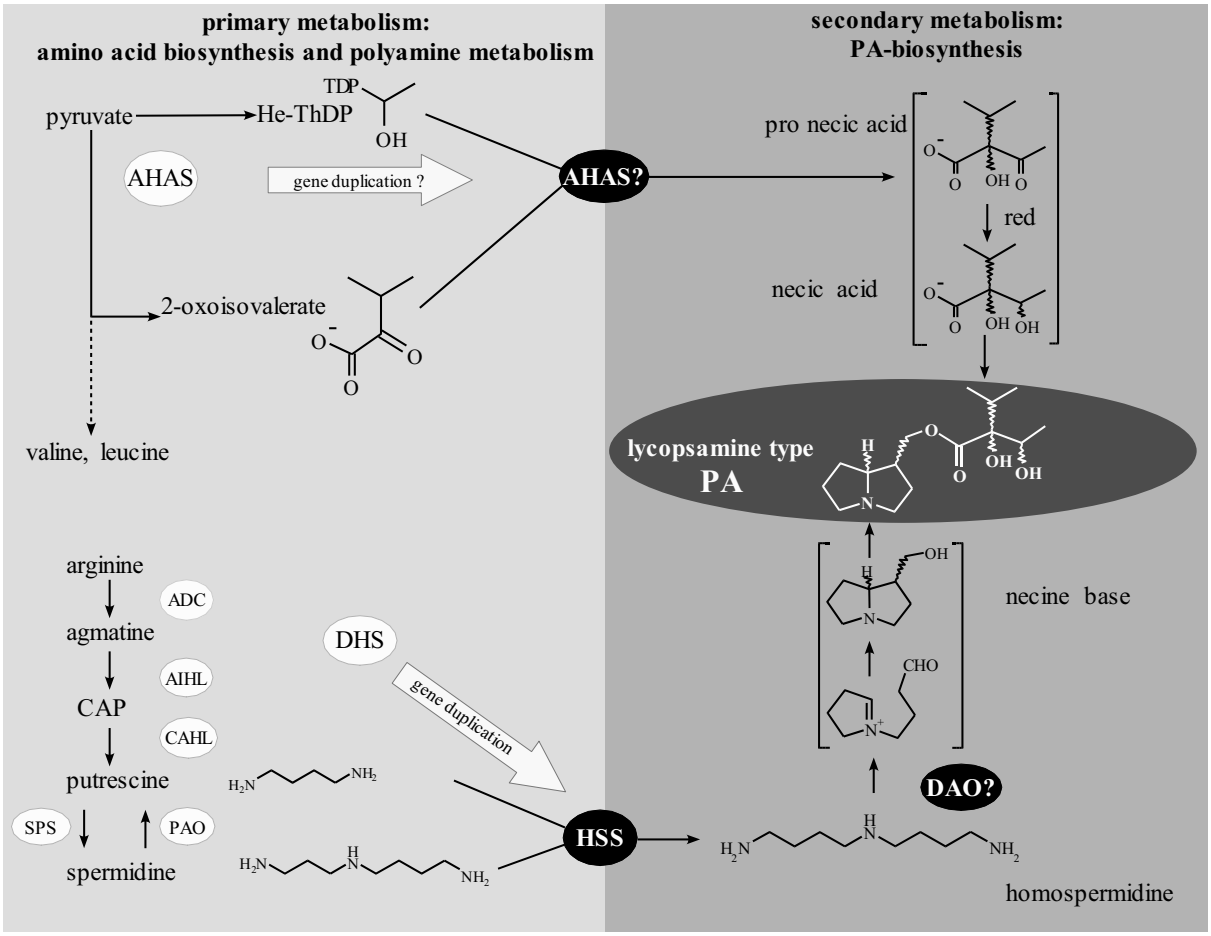


Figure 1.4: PA-biosynthesis is deduced from enzymes and substrates from primary metabolism, i.e., the amino acid biosynthesis and polyamine metabolism according to Hartmann and Ober 2000. Postulated parts of the pathway are tagged by a question mark, postulated intermediates are enclosed in square brackets (ADC = arginine decarboxylase, AHAS = acetohydroxyacid synthase, AIHL = agmatine iminohydrolase, CAHL = N-carbamoylputrescine amidohydrolase, CAP = N-carbamoylputrescine, DAO = diamine oxidase, He-ThDP = hydroxyethyl thiamine diphosphate, HSS = homosperimidine synthase, PA = pyrrolizidine alkaloid, PAO = a putrescine producing polyamine oxidase, SPS = spermidine synthase).

1999b, Hartmann and Ober 2000).

The next step in the biosynthesis of the necine base has not yet been clarified at the enzyme level. In the presence of β-hydroxyethylhydrazine (HEH), which is a diamine oxidase inhibitor, *Senecio* root cultures were shown to accumulate homospermidine (Böttcher et al. 1993). More recently, the same result were confirmed with various organs from *Heliotropium indicum*. Inhibiting trials combined with ^{14}C-putrescine tracer technique revealed that HEH completely interrupted PA biosynthesis directly after the homospermidine synthesis, homospermidine accumulated and appeared to function exclusively as unique necine base precursor (Frölich et al. 2007). Hence, the postulated intermediate dialdehyde might be formed by a diamine oxidase like enzyme (DAO? in Figure 1.4). The postulated dialdehyde might undergo cyclization resulting in the 1-hydroxymethylpyrrolizidine main body of necine bases via an intermediate iminium ion.

HEH inhibitor studies had provided evidence that the HSS is the first pathway specific enzyme in the formation of the necine base and, moreover, in the PA biosynthesis. Thus, the HSS is active at the interface between primary (amino acid) and secondary (PA) metabolism (Figure 1.4).

The HSS of *Senecio vernalis* was cloned, sequenced, and overexpressed in *E. coli* (Ober and Hartmann 1999b). High sequence similarities were shown between plant HSS and deoxyhypusine synthase (DHS) which catalyzes the first step in the activation of the eukaryotic initiation factor 5A (eIF5A, Chen and Liu 1997). DHS transfers an aminobutyl moiety from spermidine to the ϵ-amino group of a specific lysine residue in eIF5A. HSS catalyses exactly the same reaction but uses putrescine instead of eIF5A as an acceptor for the aminobutyl moiety (Ober et al. 2003b). While DHS is capable of accepting both substrates, eIF5A and putrescine, HSS only utilizes putrescine as an acceptor molecule for the transfer of an aminobutyl group. Hence, the HSS can be considered to be a modified DHS that lost its ability to bind one of the two original substrates with a still very well conserved catalytic activity. Noteworthy, the DHS main function is seen in the post-translational activation of eIF5A that was shown to be essential for cell growth and proliferation in *Saccharomyces cerevisiae* (Park et al. 1998) which is clearly part of the primary metabolism. The synthesis of homospermidine is considered a side reaction of the DHS which might explain the occurrence of homospermidine at least in traces even in non-PA-producing plants (Ober et al. 2003a). Hence, the detection of the PA precursor homospermidine in a plant is no indicator that this plant species is at the border of evolving the PA biosynthesis.

1.1.2.2 Necic Acid Biosynthesis

In contrast to the biosynthesis of the necine base, the knowledge concerning the biosynthesis of the necic acids is rather fragmentary and has not yet been clarified at the enzyme level. Tracer studies revealed that all aliphatic necic acids are derived from branched-chain amino acids (Hartmann and Witte 1995). Best examined are class I necic acids (Figure 1.2 on page 13) from senecionine, monocrotaline, and triangularine type PAs. Crout and co-workers showed that isoleucine is incorporated once into monocrotalic acid

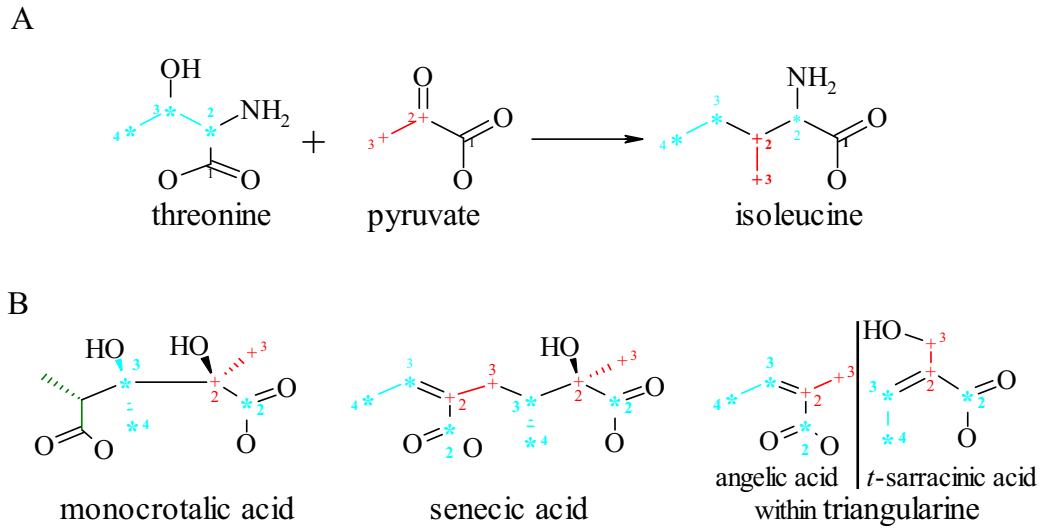


Figure 1.5: Isoleucine is incorporated into necic acids of group I PAs: carbon atoms highlighted in cyan (*) originate from threonine, while carbon atoms highlighted in red (+) originate from pyruvate.

(Robins et al. 1974) as well as into triangularine type necic acids (angelic acid: Crout 1967), where in macrocyclic senecionine type necic acids (seneciphyllic acid: Crout et al. 1966, 1970, senecic and isatinecic acid: Crout et al. 1972, Bale et al. 1978, Cahill et al. 1980) isoleucine is incorporated twice (Figure 1.5).

Moreover, leucine was shown to be introduced to senecioic acid (O'Donovan and Long 1975) while threonine/isoleucine labeled one half and valine/leucine labeled the other half of the trichodesmic acid from senecionine type PAs (Devlin and Robins 1984).

Studies by Crout (1966) on the biosynthesis of lycopsamine type necic acids performed on hound's tongue (*Cynoglossum officinale*, Boraginaceae) indicated that valine is incorporated, but as valine only provides 5 carbon atoms, Crout suggested that the missing two-carbon-atom unit for a complete C7 necic acid is inserted at the α-carbon atom of either valine or the corresponding product of transamination, 2-oxoisovalerate, via an "active acetaldehyde" in an acyloin condensation. Tracer studies with ^{13}C-labelled glucose (Weber et al. 1999) provided further indications that the "active acetaldehyde" introduced into lycopsamine type necic acids is the same used in branched-chain amino acid biosynthesis (Figure 1.6): the hydroxyethyl thiamine diphosphate (He-ThDP).

The enzyme transferring the He-ThDP in primary metabolism is an acetohydroxyacid synthase (AHAS, EC 2.2.1.6). In catalyzing the first step of the biosynthesis of valine, leucine, and isoleucine (Umbarger and Brown 1958), the AHAS is able to recognize two different substrates: pyruvate and 2-oxobutyrate (Umbarger and Brown 1958, Radhakrishanan and Snell 1960). Supposing an enlarged substrate specificity, the enzyme may accept 2-oxoisovalerate to form an intermediate in necic acid biosynthesis (Figure 1.6).

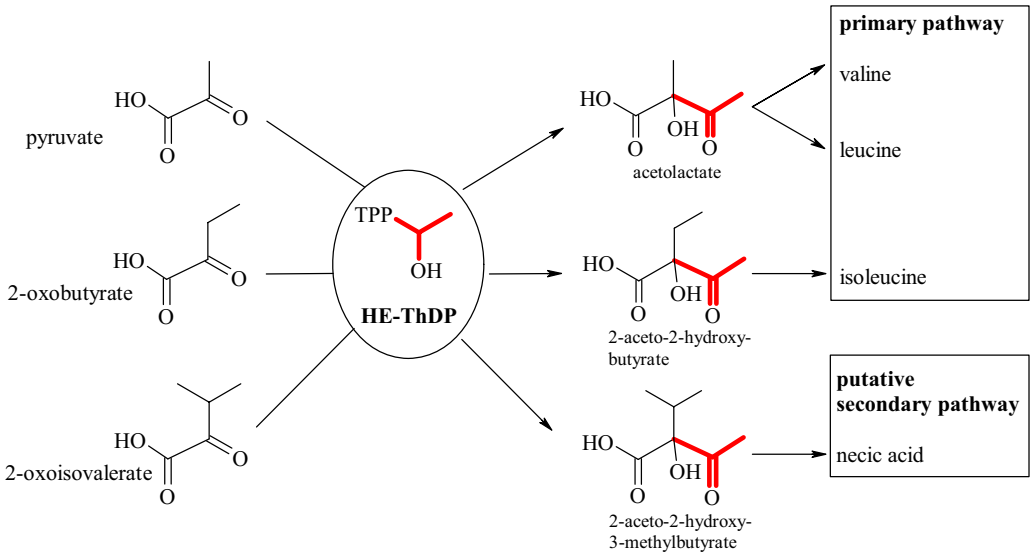


Figure 1.6: Postulated broadening of the substrate specificity of a postulated AHAS-like enzyme catalyzing the first step in necic acid biosynthesis within PA-producing plants. As shown by Crout 1966, a two-carbon-atom unit (highlighted in red) is inserted at the α-carbon atom of 2-oxoisovalerate via the "active acetaldehyde" He-ThDP (hydroxyethyl thiamine diphosphate). Tracer studies with ^{13}C-labbelled glucose, performed by Weber et al. (1999), confirmed that valine (precursor: acetolactate) shows the same labbelling patterns as the lycopsamine type necic acids (precursor: 2-aceto-2-hydroxy-3-metylbutyrate).

AHAS inhibitor herbicide	described by	minimal amounts and concentrations to apply
sulfonylureas	Levitt 1978	5 - 100 g/ha, nM-range
imidazolinones	Los 1984	5 - 100 g/ha, μM-range
triazolopyrimidine	Gerwick et al. 1990, Kleschick et al. 1992	9 g/ha (flumetsulam)
pyrimidinylsalicylic acids	Shimizu et al. 1994	70 g/ha (pyrithiobac)

Table 1.3: Classes of herbizides targeting AHAS

1.1.2.3 Acetohydroxyacid Synthase (AHAS)

In 1958, Umbarger and Brown discovered the enzyme catalyzing the first step in branched chain amino acid biosynthesis. The capability of this *de novo* synthesis is characteristic for all autotrophic species, namely, archaea, algae, and plants and, for some heterotrophs as well, bacteria and fungi. Animals including humans are not capable of synthesizing valine, leucine, or isoleucine. They have to supply these essential and proteinogenic amino acids with their diet.

In plants, the AHAS was determined to be the target enzyme of a novel class of herbicides (see Table 1.3), the AHAS inhibitors. Due to their polyploidy, AHAS from crops are resistant to those herbicides while, in the beginning, weeds had been sensitive. Today, more plants are resistant to AHAS inhibitors than to any other herbicide class (Tranel and Wright 2002). This is mainly due to a single amino acid mutation. Since the herbicide targeted enzyme, the AHAS, is absent in humans, these novel herbicides were believed to pose only a very little risk to health of consumers.

The AHAS either catalyzes the synthesis of (2S)-acetolactate from two molecules of pyruvate and (2S)-2-aceto-2-hydroxybutyrate from pyruvate and 2-oxobutyrate (Radhakrishnan and Snell 1960). The reactions are mediated by the cofactor thiamine diphosphate (ThDP). In Figure 1.7 (page 24), the catalytic cycle of the AHAS mediated reaction with ThDP is shown: Starting with mesomerically stabilized ylide molecule formed from ThDP in the active center of the AHAS, a non-oxidative decarboxylation of pyruvate yields the "active acetaldehyde", the 2-(1-hydroxyethyl)thiamine diphosphate (HE-ThDP). Next, a second 2-ketoacid is condensed to He-ThDP while ThDP is restored for a new reaction (Holzer et al. 1960, Tittmann et al. 2003). A divalent metal ion like Mg^{2+} is required anchoring ThDP to the AHAS. Additionally, flavin adenine dinucleotide (FAD) is necessary which does not participate in the reaction but functions as an enzyme structure preserving element (Stormer and Umbarger 1964).

Two different types of subunits form the holoenzyme. Two catalytic subunits (csu) form an intimate dimer which harbours the active site, while the regulatory subunits (rsu) mediate feedback inhibition from the endproducts valine, leucine or isoleucine. However, it is still unknown how many regulatory subunits constitute a native plant holoenzyme (Table 1.4 on page 26). Isoenzymes of AHAS catalytic subunits are found in plants

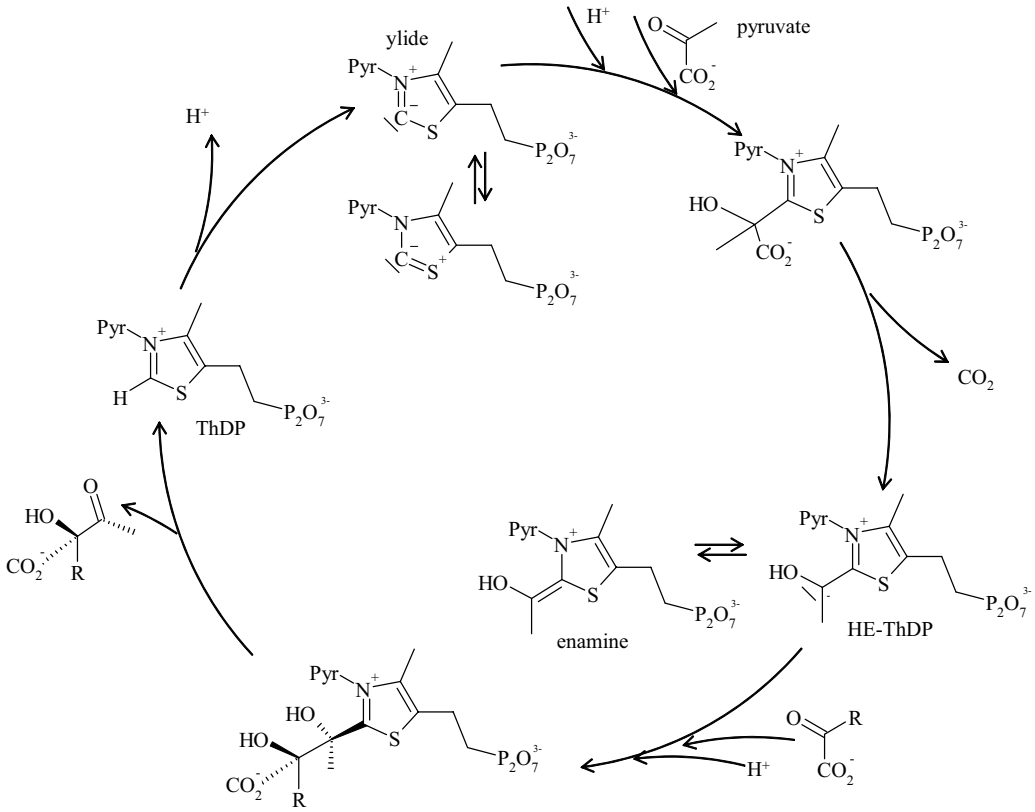


Figure 1.7: Catalytic cycle of AHAS mediated reaction with thiamine diphosphate (ThDP). A non-oxidative decarboxylation of pyruvate yields the He-ThDP (2-(1-hydroxyethyl)thiamine diphosphate) that is stabilised as enamine. Apart from pyruvate, a second 2-ketoacid is condensed to He-ThDP while ThDP is restored for the next transfer of a hydroxyethyl group to an α-ketoacid.

and bacteria, but not in fungi. In *E. coli*, only homodimers of catalytic subunits were found, while in plants they can occur as heteromers, e.g., *Brassica napus* (Bekkaoui et al. 1993), or one isomere type of catalytic subunit forms holoenzymes with various regulatory subunits (*Arabidopsis thaliana*). The number of isoenzymes is correlated with a plants ploidy: Allotetraploid *Nicotiana tabacum* has two isoenzymes both expressed in a constitutive manner (Chang et al. 1997). Allotetraploid *Brassica napus* was reported by Grula et al. (1995) to own 5 AHAS encoding genes (2 housekeeping, 1 exclusively in reproductive tissues, 1 pseudogene, and one of unknown function) while allotetraploid *Gossypium hirsutum* has 6 AHAS encoding genes (4 housekeeping, 1 or 2 in anthers).

As reviewed by McCourt and Duggleby (2006), the main body of the *Arabidopsis* catalytic subunit is composed of 3 domains: α (85 – 269), β (281 – 458) and γ (463 - 639). N-terminally (1-84), a chloroplast-targeting transit peptide (cTP) is located. The nuclear encoded AHAS from plants, cyanobacteria, green algae, and fungi are guided to chloroplasts or mitochondria (in case of fungi) by a signal peptide which is cleaved when entering the organelles to yield the mature protein (Duggleby and Pang 2000). Despite of this, red algae contain a plastid encoded AHAS that does not include a signal peptide. C-terminally, a structured tail (568-583) is found which covers the active site during the enzyme reaction and uncovers it to enable release of the product (Pang et al. 2003).

In enterobacteria like *E. coli*, three isoenzymes enable the organism to compensate various metabolic states imposed through the environment. According to Barak et al. (1987), substrate preference is quantifyable using the specificity ratio R (Figure 1.8, page 27). Under normal growing conditions (full medium), cellular concentrations of pyruvate are much greater than that of 2-oxobutyrate ([pyruvate] $\gg$ [2-oxobutyrate]) requiring enzymes with high R-ratios (R = formed 2-acetohydroxybutyrate / [2-oxobutyrate] proportional to formed 2-acetolactate / [pyruvate]) to secure the synthesis of all three branched chain amino acids (BCAAs). AHAS II displayed a strong preference for 2-ketobutyrate (R = 185; Barak et al. 1987) while AHAS III revealed a moderate preference for 2-ketobutyrate (R = 53; Vyazmensky et al. 1996). In contrast to AHAS I and AHAS III, AHAS II is insensitive to feedback inhibition by any BCAA (Lawther et al. 1987) which prevents the cell from death due to isoleucine starvation. Hence, AHAS II and III are regarded as "house-keeping"-enzymes. In case of FAD deficiency, AHAS II keeps up the alimentation with BCAAs, because it does not require FAD. AHAS I showed little preferences for either substrate (R = 1; Barak et al. 1987) and, thus, is the most useful enzyme at intracellularly low pyruvate concentrations (acetate minimal medium).

1.2 Objectives of This Work

The differential distribution of PA types between the Senecioneae and the Eupatorieae implies (Table 1.1, page 14) that the formation of lycopsamine type PAs requires a specialized enzyme that is not present in the Senecioneae. Comparing the main structures of the two PA types, the necic acids make the difference: Necic acids like tiglic acid within the Senecioneae (group I in Figure 1.2, page 13) had been found esterified in many other

organisms	native enzyme [kDa]	iso-enzyme	csu [kDa]	rsu [kDa]	pH-optimum	spec. activ. [nkat/mg]
Escherichia coli[1]	145	I II III	60 59 63	11.0 9.5 17.9		667 879 500
Arabidopsis thaliana[2]	548	I	65	53	6.5 - 8.5	131
Brassica napus[3]	130	I III	66 65		7.0 - 7.4 7.2 -7.6	
Eupatorium clematideum[4]					7.0	0.0205*
Glycine max[5]					6.0 - 8.0	60*
Hordeum vulgare[6]	440	I	58 65		6.3 - 8.3	27 - 92*
Nicotiana tabacum[7]			65			96
Nicotiana plumbaginifolia[8]			65	50		108
Oryza sativa[9]			69	52		
Pisum sativum[10]						19*
Sinapis alba ssp. *alba*[10]						9*
Triticum aestivum[11]	520	I II	58 57			1*
Xanthium spec.[12]						70
Zea mays[13]	900 440	I II	55 55		6 - 7	12

[1] Grimminger and Umbarger 1979, Eoyang and Silverman 1988, Weinstock et al. 1992, Vyazmensky et al. 1996, Hill et al. 1997, Hill and Duggleby 1998;
[2] Chang and Duggleby 1997, 1998, Lee and Duggleby 2001; [3] Bekkaoui et al. 1993;
[4] Weber 1997; [5] Durner 1991; [6] Durner and Böger 1988, 1990, Durner 1991, Chong et al. 1997b, Yoon et al. 2003; [7] Oh et al. 2001; [8] Hershey et al. 1999; [9] McCourt and Duggleby 2006;
[10] Durner 1991; [11] Southan and Copeland 1996; [12] Bernasconi et al. 1995
[13] Muhitch et al. 1987, Singh et al. 1988a, 1988b, Durner 1991; * native enzyme

Table 1.4: Currently known features of *E. coli* and plant AHAS catalytic (csu) and regulatory subunits (rsu)

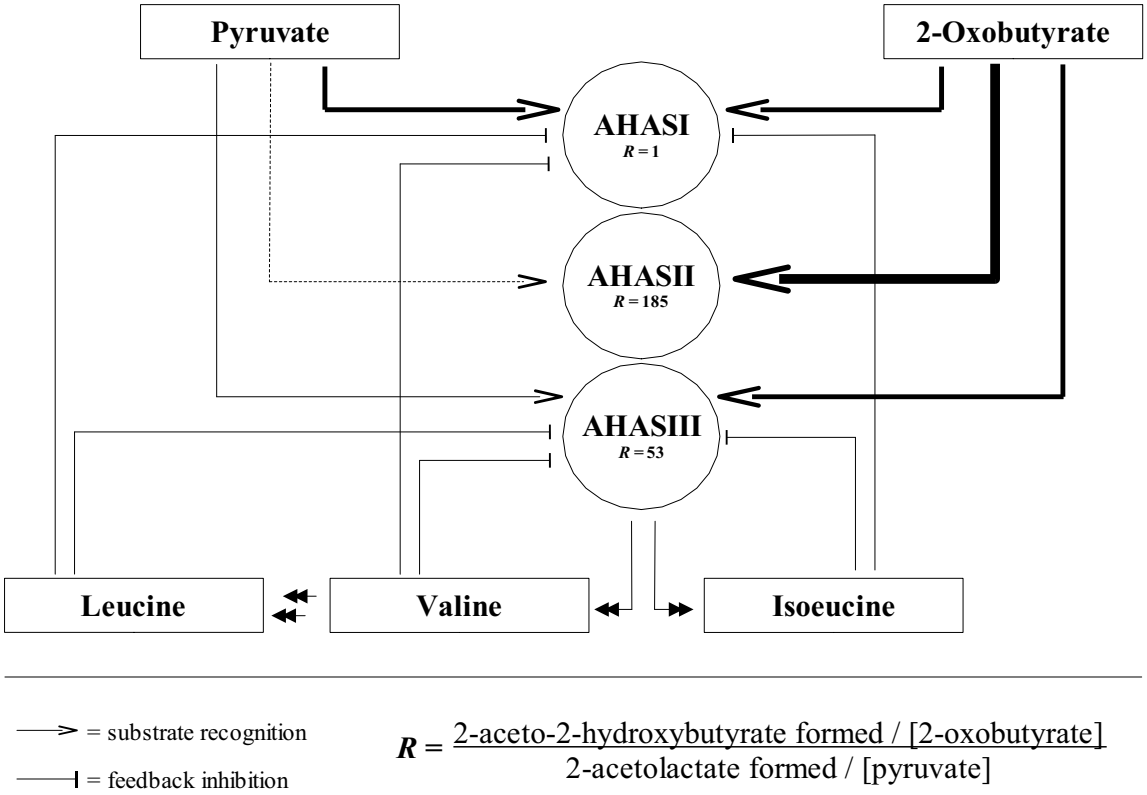


Figure 1.8: Regulation of the three AHAS isoenzyme activities in *E. coli*. Substrate preferences are indicated by the thickness of arrows: AHAS II shows the highest preference for 2-oxobutyrate followed by AHAS III with a moderate preferrence for 2-oxobutyrate. AHAS I has no preferences. Unlike AHAS II, AHAS I and AHAS III require FAD and are sensitive to feedback inhibition.

alkaloids, e.g., as ditigloyloxytropane within the tropane alkaloids of *Datura meteloides* (Evans and Woolley 1965). On the other hand, necic acids of lycopsamine type PAs within the Eupatorieae (group II in Figure 1.2) are very unique. They possess a remarkable C7 carbon skeleton of 2-isopropylbutyric acid and they occur almost exclusively within PA-producing plants with one exception: in asarumins of *Asarum forbesii*, Araceae (Kobayashi et al. 1989).

Tracer studies by Crout (1966) and Weber et al. (1999) suggested that the 2-isopropyl butyric acid skeleton is formed with intermediates and enzymes from the branched-chain amino acid biosynthesis (BCAAB, see Figure 1.9). The intermediate 2-oxoisovalerate that is educt of the transamination reaction resulting in valine was postulated to be fused to the "active acetaldehyde" He-ThDP. In primary metabolism (BCAAB), He-ThDP is formed and used as a substrate by the AHAS, the first enzyme in the biosynthesis of valine and leucine. The general conserved catalytic activity of the AHAS is characterized by transferring a hydroxyethyl group to α-ketoacids. Not surprisingly, the AHAS is involved in the biosynthesis of isoleucine as well, accepting 2-oxobutyrate as an additional substrate to pyruvate (section 1.1.2.3, Figure 1.7). Hence, an AHAS-like enzyme was postulated to show an enlarged or altered substrate specificity to catalyze the first step in the lycopsamine type necic acid biosynthesis. In order to test this hypothesis, the following experimental steps shall be carried out:

1. AHAS homologous cDNAs from lycopsamine type PA-producing plants shall be discovered in full-length. Similarity of the novel sequences to already known AHAS has to be determined with computer assisted sequence analyses. Proof of functionality shall be given by heterologous expression followed by biochemical characterization. Investigations by Weber (1997) on the native AHAS from *Eupatorium clematideum* root-tissue cultures could not reach the necessary purity due to the enzymes' instability and low abundance. For sufficient quantities of stabile enzyme, heterologous expression of the novel determined plant AHAS sequences has to be conducted.

2. Recombinant AHAS from PA-producing plants have to be checked for PA-specific enzymes utilizing an especially designed assay that verifies the acceptance of substrates for necic acid biosynthesis and, moreover, that differentiates between products from branched chain amino acid biosynthesis and necic acid biosynthesis.

3. RT-PCR analyses shall determine the tissues of mRNA occurrence which will contribute to select AHAS-like enzymes with a differing substrate specificity.

In the case that the AHAS-like enzyme shows an altered substrate specificity, a clear distinction between enzymes from primary and secondary metabolism should be possible. As a long range perspective, the evolutionary origin of the PA-specific AHAS-like enzyme should be traced back and will reveal a mono- or polyphyletic origin when further

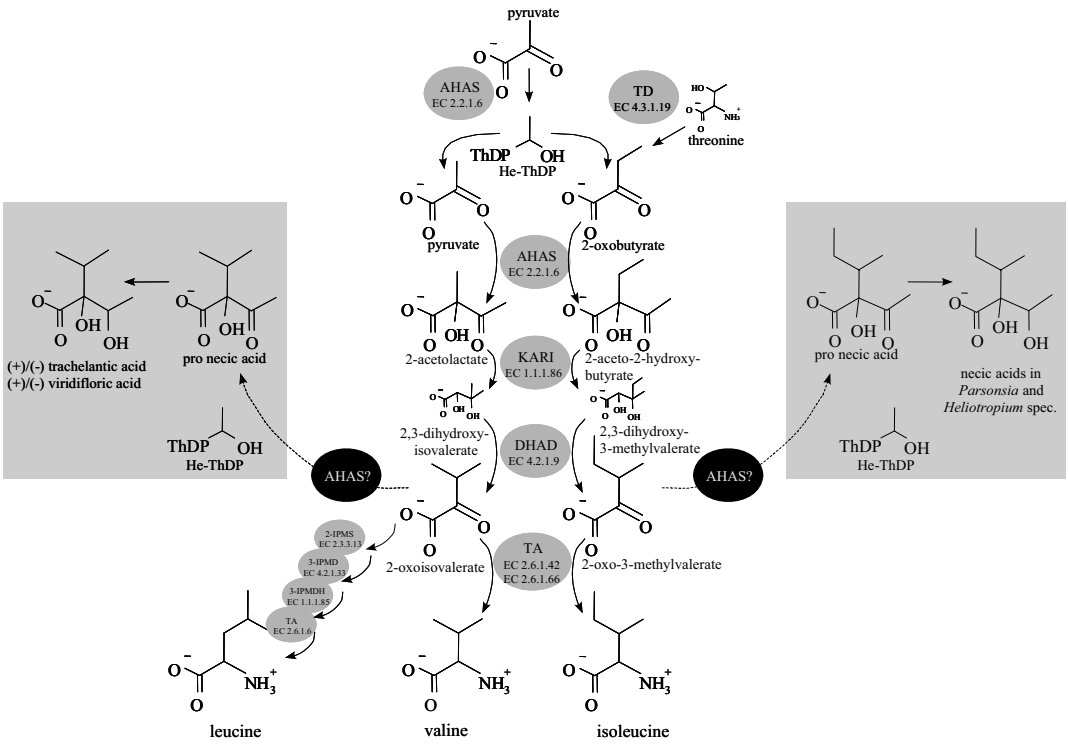


Figure 1.9: The branched-chain amino acid biosynthesis (BCAAB) and proposed junction to necic acids. The educts of the transamination reactions leading to valine and isoleucine, respectively, are postulated to receive a hydroxyethyl group transferred by an AHAS-like enzyme. All necic acid structures resulting from this hypothesized pathway are esterified moieties of actually existing pyrrolizidine alkaloids. Abbreviations: AHAS = acetohydroxy acid synthase, DHAD = dihydroxyacid dehydratase, He-ThDP = hydroxyethyl thiamine diphosphate, 3-IPMD = 3-isopropylmalate dehydratase, 3-IPMDH = 3-isopropylmalate dehydrogenase, 2-IPMS = 2-isopropylmalate synthase, KARI = ketol-acid reductoisomerase, TA = transaminase, TD = threonine deaminase.

AHAS-like enzymes from other lycopsamine type PA-producing plants have been identified. These results can be compared to the polyphyletic origin of the homospermidine synthase (HSS), the first characterized PA-specific enzyme (section 1.1, Reimann et al. 2004, Ober 2007).

2 Material and Methods

In this chapter, used materials are listed and comprehensive descriptions of all applied methods are placed at the readers disposal.

2.1 Chemicals

Commercially available chemicals and gases were used in grades *pro analysi* or *pure*. In general, they were purchased from companies named below if not stated otherwise.

Aldrich	Fermentas	New England Biolabs	Roth
Amersham-Bioscience	Fluka	Novagen	Serva
AppliChem	Gerbu	Promega	Sigma
Biomol	Invitrogen	Qiagen	Westfalen
Bio-Rad	Machery-Nagel	Roche Diagnostics	
Calbiochem	Merck	Riedel de Haën	

Individually designed primers and rare primers for sequencing were produced by MWG Biotech in HPSF quality. Disposable plastic and glass ware was ordered from Biozym, Nalgene, Neolab, Renner, or Sarstedt.

All aqueous solutions and dilutions were prepared with purified water (Milli-Q System, Millipore) and sterilized by autoclaving for 20 minutes at 121°C and 2 bar or by filtration through sterile cellulose acetate filters of 0.22 μm mesh.

2.2 Plant Material

Sterile cultures of whole plants were already available at the institute. Seeds sterilized in hydrogenperoxide or sodium hypochlorite solutions had been allowed to germinate on solid MS0H medium (see below for plant media composition, page 31).

2.2.1 Media for Plants

All sterile cultures were kept in MS-media (Murashige and Skoog 1962). Slight differences in the amount of components are given in the Tables 2.1 and 2.2. MS20 contained 20% of the amount of NH_4NO_3 in MS0H, MS10 only 10%. None of these media contained any phytohormones (in MS0H, "0H" means: "zero hormones").

The ingredients were solved in deionized water and the pH was adjusted to 5.8 for MS0H and MS20 or to 6.0 for MS10. For preparation of solid media, 0.72 g agar (Bernd Euler) per 80 ml solution were added and transferred into wide necked Erlenmeyer flasks

ingredients	MS0H [g/l]	MS20 [g/l]	MS10 [g/l]
NH_4NO_3 (Merck)	1.65	0.33	0.165
$CaCl_2$ x $2H_2O$ (Merck)		0.44	
$MgSO_4$ x $7H_2O$ (Merck)		0.37	
KH_2PO_4 (Roth)		0.17	
sucrose (Serva)		30.0	
NaFe EDTA (Merck)		0.04	
myoinositol (Sigma)		0.1	
vitamin stock solution [1]		1 ml	
KI-stock solution [2]		1 ml	
micronutrients stock solution [3]		1 ml	

[1 - 3] For composition of stock solutions see Table 2.2

Table 2.1: Ingredients for MS-media

stock solutions for	ingredients	[g/100 ml]
vitamins	nicotinic acid (Merck)	0.10
	thiamine HCl (Serva)	1.00
	pyridoxine HCl (Merck)	0.10
	inositol (Sigma)	10.0
KI	KI (Sigma)	0.075
micronutrients	H_3BO_3 (Roth)	0.62
	$MnSO_4 \times H_2O$ (Fluka)	1.00
	$ZnSO_4 \times 7H_2O$ (Merck)	0.86
	$NaMoO_4 \times 2H_20$ (Merck)	0.025
	$CuSO_4 \times 5H_20$ (Merck)	0.0025
	$CoCl_2 \times 6H_2O$ (Merck)	0.0025

Table 2.2: Composition of stock solutions for plant media

with a capacity of 250 - 300 ml. Liquid media used for root-tissue cultures (RTC) were poured out directly after pH adjustment into 25 - 200 ml wide necked Erlenmeyer flasks. After sealing the flasks with an aluminium foil, the whole batch was sterilized by autoclaving. Media preserved like this and stored at room temperature were stable for at least half a year while stock solutions for vitamins, KI, and micronutrients were stored unpreserved at -20°C for more than one year but only thawed once.

2.2.2 Growing Conditions of Plants

Sterile plants from *Eupatorium cannabinum* and *Parsonsia laevigata* were grown in solid MS0H or from *Symphytum officinale* in MS20 at about 23°C (± 2°C) with 16 hours of light per day in a plant chamber. While sterile plants from *Eupatorium cannabinum* and *Parsonsia laevigata* (panel A in Figure 2.1) consisted of roots submerged in the solid medium, stalks, leaves, and shoot tips, sterile plants from *Symphytum officinale* grew as

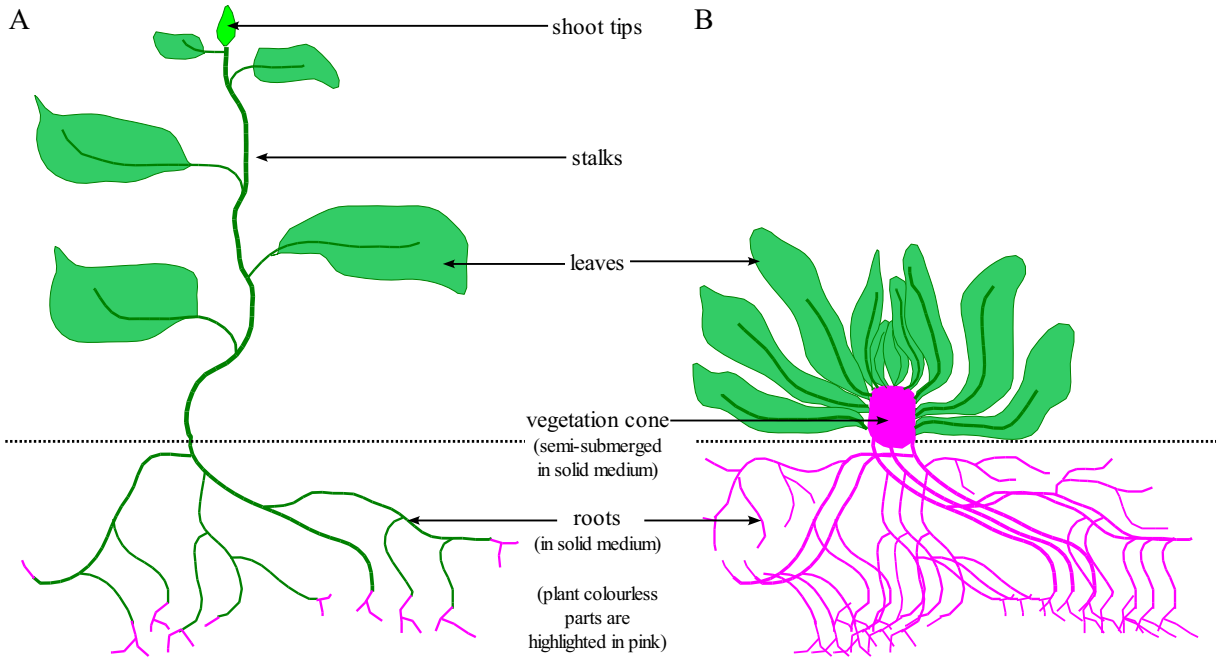


Figure 2.1: Two types of sterile plants were used as plant material. While *Eupatorium cannabinum* and *Parsonsia laevigata* grew according to panel A with only the youngest root tips still chloroplast-free (highlighted in pink), sterile total plants from *Symphytum officinale* grew according to panel B as rosette plants consisting of leaves, a medium semi-submerged vegetation cone (colourless), and naturally chloroplast-free roots. Type B roots were directly harvested while type A roots were used for growing root-tissue cultures in the dark obtaining as well chloroplast-free roots.

rosette plants without any internodes, stalks, or shoot tips, just with roots submerged in the solid medium, the semi-submerged vegetation cone, and the leaves (panel B in Figure 2.1).

Sterile root-tissue cultures (RTC) were generated by cutting off the roots from whole plants and transferring them to liquid MS0H for *Parsonsia laevigata* or to MS10 for *Eupatorium cannabinum*. The RTCs were grown in the dark at 25°C and shaken with 80 rpm. The liquid medium was exchanged every 3 to 4 weeks. Freshly sprouted root tips were almost colourless or white and were apparently free from chloroplasts ("etiolated"). They were used for isolation of total RNA (see page 58). Roots from *Symphytum officinale* appeared already chloroplast-free when grown in solid medium. They were harvested directly from the plant and were not grown in liquid medium.

2.3 Cloning Material

In this section, host cells for cloning and heterologous expression are introduced. Their basic but special genetic properties are lined out. Usually performed combinations with vectors are described. The application of media is described for different culturing purposes.

2.3.1 Host Cells

Representatives of two different domains were used as host cells, namely bacteria and eukaryotic fungi.

2.3.1.1 Properties of Bacterial Strains (*Escherichia coli*)

The following list gives the used strains. At the end of this section, an overview of host strain properties is given in Table 2.3 on page 37.

TOP 10 (Grant et al. 1990) (Invitrogen): *E. coli* K12 strain with F^- *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) Φ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *deo*R *ara*D139 Δ(*ara-leu*)7697 *gal*U *gal*K *rps*L(Str^r) *end*A1 *nup*G

In combination with an appropriate vector, this strain can be used for blue/white screening without IPTG (Φ80*lac*ZΔM15). The introduced vector should code for the complementing α-part of the β-galactosidase while the host genome confers the ω-part. These two parts of the enzyme are constitutively expressed so that there is no need for induction. The reconstituted holoenzyme of β-galactosidase cleaves the artificial substrate X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside). The resulting two compounds are galactose and 5-bromo-4-chloro-indoxyl, the latter appears blue after oxidative dimerization to 5,5'-dibromo-4,4'-dichloroindigo. If a PCR product is ligated successfully into the vector, the open reading frame of the α-part is interrupted and no functional α-part can be expressed. The colourless X-gal remains uncleaved and positive transformants appear white or colourless as well. If no PCR product is ligated successfully, the α-part is expressed, β-galactosidase is reconstituted and active, and produces the above mentioned blue colour when using X-gal as a substrate.

Further properties of TOP 10 cells are a safe propagation of plasmids due to the lack of endonuclease I and reduced general recombination abilities (*end*A1, *rec*A1). The transformation efficiency is increased by inhibited expression of restriction enzymes that digest unmethylated DNA (*hsd*RMS) as generated from PCR or strangely methylated DNA (*mcr*A, *mcr*BC, *mrr*) like genomic DNA from plants.

The strain was recommended for cloning PCR products with the pCR 2.1-TOPO TA cloning system (Invitrogen) but was also used with the pGEM-T Easy cloning vector (Promega), described on page 45. A detailed description of the cloning procedure is given on page 75. The TOP 10 strain is also a suitable host for propagating large plasmids (*deo*R). It is resistant to streptomycin (*rps*L) which can be used to maintain a pure host strain.

JM109 (Yanisch-Perron et al. 1985) (Promega): *E. coli* K12 strain with *rec*A1 *end*A1 *gyr*A96(Nal^r) *thi* *hsd*R17(r_k^-, m_k^+) *rel*A1 *sup*E44 Δ(*lac*$^-$*pro*AB) [F' *tra*D36 *pro*AB *lac*I^{q}ZΔM15]

Like TOP 10, this strain was used for cloning PCR products. This time it was recommended for the pGEM-T Easy cloning vector (Promega, for more information on the vector, see page 45 "TOP10") that enabled blue/white screening (*lac*I^{q}ZΔM15). Here, IPTG has to be added to induce expression of the complementing α-part of the β-galactosidase (for detailed information on blue /white screening, see page 34). The strain was used for cloning PCR products with the pCR 2.1-TOPO vector as well. JM109 has an increased transformation efficiency of unmethylated DNA (*hsd*R17) as it is obtained from PCR because this type of DNA is let uncleaved. It safely propagates plasmids due to the *end*A1- and *rec*A1-mutations. According to Sambrook et al. (1989), this strain is slightly defect in bacterial cell wall synthesis. This does not effect the function or the properties of the strain but produces mucoid formed colonies on plates. It was observed that cells which were made competent as described below (see page 42) had a low transformation efficiency; so competent JM109 were purchased from Promega.

As recommended by IBA Göttingen, the strain was used for expression with the pASK IBA32 expression vector (see page 47). JM109 is resistant to nalidixic acid (*gyr*A96).

DH5α (Hanahan 1983) (Invitrogen): *E. coli* K12 strain with F$^-$ Φ80*lac*ZΔM15 Δ(*lac*ZYA-*arg*F) U169 *end*A1 *rec*A1 *hsd*R17(r_k^-, m_k^+) *sup*E44 *thi-1* *gyr*A96(Nal^r) *rel*A1 *pho*A

This strain was used during subcloning for maintenance of expression constructs (see page 76) and for safe vector propagation (see page 61) due to mutations in the *end*A1 and *rec*A1 genes. Additionally, it was used for expression from T7 *lac* promoter independent vectors, e.g., pASK IBA vectors. It can be used for expression with the thermoinducible promoter p_L (Coligan et al. 2003). Expression is improved by a partial defect in heat shock response. The strain can be applied to blue/white screening without IPTG (Φ80*lac*ZΔM15) together with the appropriate vector (see page 34). The transformation with DNA from PCR is simplified (*hsd*R17), and it is resistant to nalidixic acid (*gyr*A96) that can be used to maintain a pure strain.

XL1-Blue (Bullock et al. 1987) (Stratagene): *E. coli* K12 strain with F'::Tn*10*(Tet^r) *pro*A$^+$B$^+$ *lac*I^qΔ(*lac*Z)M15/ *rec*A1 *end*A1 *gyr*A96(Nal^r) *thi* *hsd*R17(r_k^-, m_k^+) *gln*V*44* *rel*A1 *lac*

Like DH5α, this strain was used for subcloning, for safe vector propagation (*end*A1, *rec*A1), and for expression from T7 *lac* promoter independent vectors. The strain can also be used for blue/white screening (see page 34) together with the appropriate vector, but, in this host, induction with IPTG is necessary due to a *lac*I^{q}ZΔM15 genotype. The transformation of unmethylated DNA from PCR is simplified (*hsd*R17), and it is resistant to both tetracycline (Tn*10*) and nalidixic acid (*gyr*A96).

BL21(DE3) (Studier and Moffatt 1986) (Novagen): *E. coli* B strain with F$^-$ *omp*T *gal* [*dcm*] [*lon*] *hsd*S$_B$($r_B^-m_B^-$) galλ(DE3)

BL21(DE3) is the standard strain for a strong T7 *lac* promoter driven expression. The DE3 gene includes the *lac*I gene coding for the repressor of the *lac*UV5 promoter that controls T7 RNA polymerase expression which results in inhibited heterologous expression. When isopropyl-β-D-thiogalactopyranoside (IPTG, a lactose structural analog) is added, the repressor molecule (encoded by *lac*I) is inactivated. The *lac*UV5 promoter enables the production of T7 RNA polymerase that starts expression from the gene of interest cloned in the vector. This tight regulation, that separates cell growth from heterologous protein expression via IPTG induction, should be advantageous for production of toxic proteins.

Degradation of produced recombinant protein is reduced due to defect proteases (*omp*T, *lon*). Plasmids are not methylated at cytosines (*dcm*), nor is unmethylated DNA cleaved ($hsd\mathrm{S_B}$).

The strain is not suitable for subcloning or vector propagation due to endonuclease and recombination activity: plasmids and inserts can be mutated.

ER2566 (NEB): *E. coli* B strain with $\mathrm{F}^-\lambda^-$ *fhu*A2 [*lon*] *omp*T *lac*Z::T7 gene*1 gal sul*A11 Δ(*mcr*C-*mrr*)114::*IS10* R(*mcr*-73::*mini*Tn*10*-TetS)2 R(*zgb*-210::Tn*10*)(TetS) *end*A1 [*dcm*]

Like BL21(DE3), this strain was used also for T7 *lac* promoter driven expression (*lac*Z::T7 *gene1*). It was supplied with pTYB1 expression kit (IMPACT-CN from NEB, see page 46). It has a reduced degradation of produced recombinant protein (*omp*T, *lon*) and plasmids are not methylated at cytosines (*dcm*). In contrast to BL21(DE3), plasmids are a little bit safer propagated (*end*A1) but recombination is still possible.

MF2000 was kindly provided by Dr. Barbara J. Mazur and Dr. Robert A. LaRossa (both Dupont). It was created by Martin Freundlich and co-workers (Friden et al. 1982): *ilv*B800::mu-1 Bgl32 *ilv*115 *thi-1 arg*E3 *rps*L31 Δ(*ara-leu ilv*HI)863 *mtl*-1;*xyl*-5 *gal*K2 *lac*Y1 *rec*A1. This strain is completely devoid of endogenous AHAS activity (*ilv*B800::mu-1 Bgl32 *ilv*115 *thi-1 arg*E3 *rps*L31 Δ(*ara-leu ilv*HI)863 *mtl*-1). MF2000 cells were used for *in vivo* expression of functional AHAS.

2.3.1.2 Properties of Yeast Strains

Saccharomyces cerevisiae **INV*Sc*1** (Invitrogen): Genotype:*MAT*α/*MAT*α *his3*Δ*1/his3*Δ*1 leu2/leu2 trp1-289/trp1-289 ura3-52/ura3-52,* phenotype: His^-, Leu^-, Trp^-, Ura^-.

This diploid strain is auxotrophic for histidine, leucine, tryptophan, and uracil. After complementation with pYES2, auxotrophy for uracil is abolished. Hence, only histidine, leucine, and tryptophan have to be supplemented with the minimal medium when selecting on positive transformants. The strain was used for heterologous expression of ORFs cloned in vector pYES2 (see page 52).

Kluyveromyces lactis **GG799** (NEB): This strain has no auxotrophies, grows rapidly to high cell density, and is able to secrete heterologously expressed proteins. It was

property	TOP10	JM109	DH5α	XL1-Blue	BL21 (DE3)	ER 2566	MF 2000
blue/white screening	+ without IPTG	+ with IPTG	+ without IPTG	+ with IPTG	-	-	-
improved transform. efficiency	++ unmeth. + meth.	+ unmeth.	+ unmeth.	+ unmeth.	+ unmeth.	+ meth.	i.n.a.
transform. of large plasmids is simplified	+	-	-	-	-	-	-
increased yield of recomb. protein	-	-	-	-	+	+	-
safe plasmid propa-gation	++ both	++ both	++ both	++ both	-	+ endA1	+ recA1
methylates DNA	yes	yes	yes	yes	no (dcm)	no (dcm)	yes
resist. to streptomycin (Str^r) nalidixic acid (Nal^r) tetracycline (Tet^r)	Str^r	Nal^r	Nal^r	Nal^r Tet^r	-	i.n.a.	Str^r

Table 2.3: Overview of host strains used in this work and their properties. meth. = methylated, resist. = resistance, transform. = transformation, unmeth. = unmethylated, i.n.a. = information was not available

supplied with the *K. lactis* Protein Expression Kit (NEB) and was used for expression of pKLAC1 (see page 53) cloned ORFs.

2.3.2 Host Media

Pro- and eukaryotic hosts have different demands on their nutrition. Below, find a list showing which kind of medium was used for which purpose:

Media for Growth of Host Cells

medium	**host**	**type of medium and manner of use**
LB	*E. coli*	full medium for propagation and expression
M9 and 0.5 M9		minimal medium to influence expression
M9 BALH		minimal medium for selection
SOC		used during transformation
YPG	*S. cerevisiae*	full medium for propagation
SC-U		minimal medium for selection or induction
YPGlu = YPG	*K. lactis*	full medium for propagation
YPGal		full medium for expression
YCBA		minimal medium for selection

All host media were prepared according to Sambrook et al. (1989). Media were sterilized by autoclaving or filtration as indicated on page 31. The pH was adjusted with NaOH and HCl before adjusting the final volume. Solid media were poured out sterilely on Petri dishes with a volume of 25 ml for each dish. This volume was taken into account when calculating the correct amount of antibiotic solutions (see Table 2.4 on page 42). Plates were allowed to harden at room temperature and stored at 4°C, media with glycerin for permanent stocks were kept at 4°C as well, while autoclaved liquid media were stored at room temperature until used.

2.3.2.1 Media for *Escherichia coli*

Luria Bertani full medium (LB)	**liquid [g/l]**	**solid [g/l]**	**permanent stocks [g/l]**
tryptone (Roth)	10	10	10
yeast extract (Roth)	5	5	5
NaCl	10	10	10
dissolved in water and adjusted to pH 7.5			
agar (Bernd Euler)	-	15	-
glycerin (Roth)	-	-	600 ml
sterilized by autoclaving			

M9 minimal media	**M9 liquid**	**0.5 M9 liquid**	**M9-BALH solid**
5 x M9 salt solution[1]	200 ml	100 ml	200 ml
sterilized by autoclaving			
1 M $MgSO_4$	2 ml	1 ml	2 ml
1 M $CaCl_2$	0.1 ml	0.05 ml	0.1 ml
20% (w/v) Glucose	20 ml	10 ml	20 ml
thiamine (vit **B**1)	-	-	0.0002%
arginine	-	-	0.016%
leucine	-	-	0.01%
histidine	-	-	0.012%
sterilized by filtration and added to sterilized and cooled (< 60°C) M9 solt solution			

1

5 x M9 salt solution	**[g/l]**
NaH_2PO_4 x 7 H_2O	64
K_2HPO_4	15
NH_4Cl	5
NaCl	2.5

SOC part 1:	**[g/l]**	part 2:	**[mM]**
tryptone (Roth)	20	glucose	20
yeast extract (Roth)	5	$MgCl_2$	10
NaCl	10 mM	$MgSO_4$	10
KCl	2.5 mM		
dissolved in water, adjusted to pH 7.5, and autoclaved		sterilized by filtration and added to part 1 cooled down under 60°C	

2.3.2.2 Media for *Saccharomyces cerevisiae*

YPG full medium	**liquid [g/l]**	**permanent stocks [g/l]**
ryptone (Roth)	20	20
yeast extract (Roth)	10	10
glycerin (Roth)	-	600 ml
sterilized by autoclaving		
20% (w/v) Glucose	100 ml	100 ml
sterilized by filtration and added after cooling down under 60°C		

SC-U (without uracil) **minimal medium**	**selection medium solid [g/l]**	**selection medium liquid [g/l]**		**induction medium liquid [g/l]**
yeast nitrogen base without amino acids (Sigma)	6.7	6.7		6.7
dropout powder[1]	1.15	1.15		1.15
agar (Bernd Euler)	20	-		-
sterilized by autoclaving				
	[ml/l]	[ml/l]		[ml/l]
10% (w/v) raffinose	200	200[2]	-	-
20% (w/v) glucose	-	-	100[3]	-
20% (w/v) galactose	-	-	-	100
sterilized by filtration and added after cooling down under 60°C				

[1]dropout powder according to Ausubel et al. (1994):
1 g of each: adenine, arginine, cysteine, leucine, lysine, threonine, and tryptophan
0.5 g of each: aspartic acid, histidine, isoleucine, methionine, phenylalanine, proline, serine, tyrosine, and valine
were mixed well in a mortar and stored at room temperature.
[2]indifferent to transcription, [3]transcription is repressed

2.3.2.3 Media for *Kluyveromyces lactis*

YPG full medium	**YPGlu liquid [g/l]**	**YPGal liquid [g/l]**
tryptone (Roth)	20	20
yeast extract (Roth)	10	10
sterilized by autoclaving		
20% (w/v) glucose	100 ml	-
20% (w/v) galactose	-	100 ml
sterilized by filtration and added after cooling down under 60°C		

YCBA minimal medium	**selection medium solid [g/l]**	**selection medium liquid [g/l]**
yeast carbon base (Sigma)	11.7	11.7
1M $Na_xH_xPO_4$-stock solution [1]	30 ml	
agar (Bernd Euler)	20	
sterilized by autoclaving		
0.5 M acetamid	10 ml	
sterilized by filtration and added after cooling down under 60°C		

[1] 423 ml of solution A (138 g/l Na_2HPO4 x H_2O) + 577 ml of solution B(142 g/l Na_2HPO_4) were mixed, autoclaved, and resulted in pH7

2.3.3 Growth and Manipulation of Host Cell Cultures

Bacterial and fungal hosts had to be treated differently in respect to growing conditions, the preparation and transformation of competent cells, and the preparation of permanent stocks. All cultures were manipulated under sterile conditions.

antibiotic working conc. / to be dissolved in	antibiotic action	mechanism of resistance, coferred by the underlined enzyme	size of enzyme conferring resistance [kDa]
ampicillin (Amp) 50-125 μg/ml / in water	disturbance of bacterial cell wall synthesis; bactericidal	*bla* gene encodes for β-lactamase, cleaves Amp	31.5
kanamycin (Kan) 50 μg/ml / in water	binds to 70S ribosomes, misreading of mRNA, no protein biosynthesis; bactericidal	*kan* gene encodes for aminoglycoside phosphotransferase, Kan is modified, interaction with ribosomes is prevented	29.1
chloramphenicol (Cam) 20-170 μg/ml / in ethanol	binds to 50S subunits of ribosomes, prevents peptide bond formation; bacteriostatic	*caf* gene encodes for acetyltransferase, acetylated Cam is inactive	25.7

Table 2.4: Utilized antibiotics, their antibiotic principle, the corresponding bacterial mechanisms of resistance, and the size of the produced "resistance" enzyme for identification on SDS-PAGE.

2.3.3.1 Manipulation of Bacterial Cultures

This section deals with the specific needs for propagation, preparation of and subsequent transformation of chemical competent *E. coli* hosts.

Growing conditions for bacteria.

Strains were cultivated in LB as full or in M9 as minimal medium (for ingredients, see section 2.3.2.1 on page 39). While the initial growth was performed at 37°C, expression was carried out at lower temperatures as determined in this work (page 156). Resistances conferred by the hosts themselves or by incorporated vectors were selected with sterilely filtrated solutions added to already autoclaved media in the appropriate working concentrations as given in Table 2.4.

Preparation of competent bacterial hosts.

The modified method from Cohen et al. (1972) produced chemical competent *E. coli* cells. Cell walls are partially and reversibly digested which enables an up-take of foreign DNA. A culture, grown to log phase (with OD_{600} of 0.4 to 0.5) at 37°C, was harvested cautiously: cooled down for 10 minutes on ice, the culture was pelleted at 4°C for 10 minutes at 2,000 x g (Beckmann GS-6R with swinging-bucket rotor GH 3.8). The medium was removed and the cells were resuspended in precooled 0.1 M calcium chloride half the

amount of the original growing medium. Resuspension was achieved by adding the calcium chloride solution in small portions. Each time after adding a small portion, the tube was tapped softly to the ball of the thumb and cooled down again. Once resuspended, the cells were pelleted again and resuspended in 0.1 part of the starting volume of the culture with 0.1 M calcium chloride/15% (v/v) glycerol (for composition, see LB medium - permanent stocks on page 39). The cells were kept at 4°C over night to reach high competence. The next day, cells were aliquoted, snap frozen in liquid nitrogen, and stored at -80°C until used for transformation (see next section, page 43). Since they were competent for at least 12 h, it was possible to refreeze the cells once again. This procedure guaranteed a great number of intact cells and contributed to a good transformation efficiency.

Transformation of competent bacterial hosts.

Plasmids carrying the DNA fragment of interest were introduced into chemical competent cells (see above, page 42) with the help of the following protocol:

Cells were taken from -80°C and thawed on ice. An amount of 5 μl of ligation product (see page 75) or 1 - 2 μl of isolated plasmid (see page 59) were added and incubated for 20 to 30 minutes on ice. The tube was transferred to 42°C for 60 to 90 seconds. This so-called "heat shock" was carried out in a waterbath to make sure that the desired temperature spread uniformly over the mixture. Recovery of the cells was reached by adding 400 μl of prewarmed SOC medium (see section 2.3.2.1, page 39) and by shaking for 1 hour at 37°C. Now, positively transformed cells were able to express vector-encoded resistance genes for antibiotic selection. This was tested in transferring the cells to selective LB plates containing the appropriate amount of antibiotic (Table 2.4) and growing them over night at 37°C.

2.3.3.2 Manipulation of Yeast Cultures

Propagation, preparation of competent cells, and transformation of yeast cultures are described.

Growing conditions for yeast.

The two yeast strains, *Saccharomyces cerevisiae* INV*Sc*1 and *Kluyveromyces lactis* GG799 were grown in the same full medium (see pages 40 and 41: YPG liquid is the same as YPGlu liquid with 10 g/l yeast extract, 20 g/l peptone, 20 g/l glucose) while selective media differed. Positive transformants of *S. cerevisiae* INV*Sc*1 containing the pYES2 vector could grow on SC-U, a minimal medium supplemented with the amino acids histidine, leucine, and tryptophan and the purine base adenine. Uracil was not added because the ability to produce uracil was transmitted with pYES2 to the uracil-auxotrophic strain INV*Sc*1. *K. lactis* GG799 could grow on YCBA when the linearized vector pKLAC1 was successfully integrated into its genome. YCBA is a nitrogen-free minimal medium except for acetamide that was metabolized to ammonia by acetamidase encoded in the *amd*S-gene on the pKLAC1 vector.

Preparation of competent yeast hosts.

According to Dohmen et al. (1991) and modified by Dr. Martens, University Marburg, cells were grown to an OD_{600} of 0.6 to 1.0. Cells were harvested and washed in 0.5 volume

Dohmen solution I	Dohmen solution II	Dohmen solution III
1 M sorbitol 10 mM bicine-NaOH/ pH 8.35 3% ethylene glycol 5% DMSO	40% PEG 1000 0.2 M bicine-NaOH/ pH 8.35	0.15 M NaCl 10 mM bicine-NaOH/ pH 8.35

Table 2.5: Solutions for preperation of competent fungal hosts according to Dohmen et al. (1991)

of Dohmen solution I (see Table 2.5) and resuspended in 0.02 volume of Dohmen solution I. The crucial step was freezing for long-term storage: Aliquots of 0.2 ml were first cooled down to 4°C, after half an hour transferred to -20°C, tightly wrapped up with thermally insulating material to enable a very slow freezing. When freezing was complete, cells were transferred to -80°C. The two types of yeast strains were made competent with the same Dohmen-method.

Transformation of competent yeast hosts.

Up to 5 μg of plasmid DNA in a volume of 15 μl (minimum was 0.1 μg) were mixed with 50 μg of single stranded carrier DNA (herring sperm DNA [10 μg/μl], denatured to single strand at 95°C for 5-10 minutes and let cool down) and the mixture was added to still frozen competent cells. It was thawed at 37°C in a water bath. The critical parameter was to efficiently mix during and after thawing: Every 10 to 15 seconds the mixture had to be vortexed. After adding 1.4 ml of Dohmen solution II (see Table 2.5) cells were shaken for 1 minute. During incubation at 30°C for 60 minutes, sedimented cells were resuspended every 15 minutes. Cells were harvested at 3.000 x g for 5 seconds, resuspended in 1.0 ml Dohmen solution III (see Table 2.5), and plated onto appropriate selective medium (sections 2.3.2.2 and 2.3.2.3).

2.3.3.3 Permanent Stocks from Bacterial and Yeast Cultures

A single bacterial colony was inoculated in 7 ml of liquid LB medium mixed with the appropriate antibiotic, while a single yeast colony was transferred to 7 ml of liquid YPG. The two types of colonies were grown over night with 250 rpm on a horizontally rotating plate within a culturing chamber of Innova 4230 shaker from New Brunswick, for bacterial cultures at 37°C and for fungal cultures at 30°C. The next day, 750 μl of these over night cultures were transferred to 250 μl LB/ glycerin or YPG/glycerin, respectively, into a 2 ml screw capped microcentrifuge tube. While the bacterial culture was immediately stored away at -80°C, the yeast culture was frozen very slowly as outlined on page 43 in section "Preparation of competent yeast hosts".

vector	prom	rep	ind	sel	tags	leaders
pET3a	T7	lacI^q	IPTG	Amp	-	-
pET22b-23a (both Novagen)					C-His	N-pelB
pTYB1 (NEB)[1]	T7	lacI	IPTG	Amp	C-Intein	-
pGEX-5x-3 (Amersham Bioscience)[2]	*tac*	lacI^q	IPTG	Amp	N-GST	-
pASK IBA3C	*tetA*	tetR	AHTC	Cam	C-Strep	-
pASK IBA7+				Amp	N-Strep	-
pASK IBA32 (IBA Göttingen)				Amp	C-His	N-ompA

[1] Chong et al.(1997a), [2] Smith and Johnson (1988)

Table 2.6: Features of vectors used for heterologous protein expression in *E. coli* (prom = promoter, rep = repressor, ind = induction with, sel = antibiotic selection). Nomenclature of *E. coli* genotypes and abbriviations are given in the appendix.

vector	suitable host cells / blue/white screening enabled	selection	size [bp]	copy numbers
pCR 2.1-TOPO (Invitrogen) Shuman 1994	Top10, DH5α / without IPTG	Amp	3,908	500 - 700 $\triangleq$ high copy vector
pGEM-T Easy (Promega) Summerton et al. 1983	JM109, XL1 blue / with IPTG	Amp	3,000	300 - 500 $\triangleq$ high copy vector

Table 2.7: Features of *E. coli* cloning vectors.

2.3.4 Vectors

As autonomic genetic information, plasmids provide to their hosts useful additional capabilities like antibiotic resistance or expression of cloned genes. For a quick overview of cloning vectors, see Table 2.7 on page 45. Expression vectors for *E. coli* are listed in Table 2.6 on page 45, while expression vectors for fungal hosts are shown in Table 2.10 on page 52.

2.3.4.1 Vectors for Cloning

The following vectors (see Table 2.7 on page 45) were used during determination of unknown open reading frames (ORF) for maintenance and sequencing of PCR products (see page 59).

The two vectors belong to the class of high copy plasmids although the number of copies differs a lot: pCR 2.1-TOPO normally contains 500 - 700 plasmids per single cell while pGEM-T Easy harbours 300 - 500 plasmids. For preparation of plasmids (see section 2.4.3.2 on page 59) from cells containing pGEM-T Easy vectors, the amount of culture was doubled compared to that harbouring the pCR 2.1-TOPO vector. The two vectors provide ampicillin resistance to their hosts. They differ in the way, blue/white screening is enabled: pCR 2.1-TOPO does not need to be induced with IPTG for expressing the α-part of the β-galactosidase that produces a blue colour in colonies not harbouring the insert of interest (for more details on the mechanism of blue/white screening, see section 2.3.1 on page 34). In contrast to this, vector pGEM-T Easy needs to be induced with IPTG.

2.3.4.2 Vectors for Heterologous Protein Expression in *E. coli*

Vectors listed below are shortly described in respect to their most prominent features. An overview is given in Table 2.6 on page 45.

pET3a (4,640 bp) and **pET22b-23a** (3,666 bp)

The pET vectors from Novagen were already available at the institute. Vector pET22b-23a is a chimera of the vectors pET22b and pET23a. The multiple cloning site (MCS) of vector pET23a has been substituted with the one from vector pET22b. Vector pET22b-23a offers the sequence for a N-terminal leader peptide called pelB and a C-terminal 6xHis-tag for purification on Ni-NTA columns (see section 2.5.2.1 on page 83). Vector pET3a contains no sequence encoding a tag for purification which means that it had to be introduced with the primers. In both vectors, the phage T7 promoter is recognized by T7 RNA polymerase but not by *E. coli* RNA polymerase. The T7 gene*1* is not located on the pET vector and has to be supplied chromosomally by the host, e.g., BL21(DE3)-encoded in λ(DE3) or ER2566 equipped with T7 gene*1* (see section 2.3.1.1, page 35).

Sequencing was performed with either T7 universal primer that annealed to the T7 promoter region and decoded in 3'-direction or with the T7 terminator primer that was complementary to the T7 terminator sequence and decoded in 5'-direction. The two primers were already available as "standard primers" at MWG, where constructs were sequenced.

pTYB1 (7,477 bp) vector is part of the IMPACT-CN system (NEB). The vector pTYB1 encodes a C-terminal Intein-tag with an enclosed chitin binding domain (ChBD) of together 55 kDa (454 amino acids, 403 + 51). The tag was used for purification with Chitin beads (NEB). The Intein-tag contains an inducible self-cleavage activity and allows to cleave on-column with DTT (Gerbu). With the tag still sticking to the column, the recombinant protein is released without any artificial amino acids with the help of elution buffer (see section 2.5.2.2 on page 84). For inducible self-cleavage activity, some C-terminal residues of the target protein which promote *in vivo* cleavage (aspartate (100%), arginine (75%), histidine, glutamate (both 50%), and threonine (25%)) had to be exchanged by indifferent but additional and artificial amino acids. This is inevitable as premature *in*

vivo cleavage takes place within the *E. coli* host after translation into recombinant protein and results in a loss of expressed protein during purification due to a loss of the purification tag. Suitable hosts for expression are ER2566, as recommended by the manufacturer, or BL21(DE3). Clones harbouring pTYB1 can be selected with ampicillin. Sequencing was performed with T7 universal primer of which the binding site was upstream the MCS.

pGEX-5x-3 (4,974 bp) vector from Amersham Biosciences encodes a N-terminal glutathione S-transferase tag (N-GST tag) which is 26 kDa in size and removable by cleavage with factor Xa. Purification was done with the help of glutathione sepharose as stationary phase and reduced glutathione as eluting agent on FPLC (see section 2.5.2.3 on page 85). Vector pGEX-5x-3 provides ampicillin resistance to its host. As described by the manufacturer, expression was performed in BL21(DE3) or in XL1-Blue. The advantage of BL21(DE3) host strain is its protease deficiency (*omp*T, *lon*) which spares recombinant proteins from degradation, while XL1blue is not protease deficient but propagates the plasmid safely due to deficient endonuclease and recombinase (*end*A1, *rec*A1). While pTYB1 and pET vectors can not be used for heterologous protein expression without an appropriate host that supports the T7 promoter genetically (BL21(DE3), ER2566), the pGEX vectors contain a *tac* promoter which is independent from genetical features of its host. Hence, pGEX vectors can be used for expression together with hosts like XL1blue or DH5α. Qualitatively positive expression results can be seen right after transformation into and induction of hosts for cloning and maintenance like XL1blue or DH5α. This saves a second transformation into and induction of specialized expression hosts during the quick check on expression. When preparing protein expression quantitatively, specialized expression hosts like BL21(DE3) were preferred.

Sequencing was performed with primers recommended by the manufacturer.

pASK IBA3C (3,001 bp), **pASK IBA7+** (3,267 bp), and **pASK IBA32** (3,289 bp) were purchased from IBA GmbH, Göttingen. Expression is recommended in JM109 or BL21(DE3) host cells. The *tetA* promoter system is induced with a subantibiotic amount of anhydrotetracycline (AHTC, 200 ng/ml). Vectors pASK IBA3C and 7+ encode a Strep-tag II sequence for purification of the expressed protein. For pASK IBA3C and 7+, this tag was located C-terminally and N-terminally, respectively. The N-terminal tag was removable by cleavage with factor Xa. Vector pASK IBA32 encodes a C-terminal 6xHis-tag allowing purification of the recombinant protein via Ni-NTA agarose (see section 2.5.2.1, page 83). Moreover, it might enhance correct protein folding in guiding the translated recombinant protein to the periplasmic space with the help of a N-terminally introduced ompA-leader sequence. In the periplasmic space, redox conditions are different from those in the cytoplasm which can enhance correct folding of the protein. The leader sequence, necessary for reaching and entering the periplasmic space, is then cleaved when the protein enters the periplasmic space. This cleavage reduces the number of artificially introduced amino acids. The cleavage is conducted by a membrane bound protease which is encoded in *omp* genes.

While pASK IBA3C confers chloramphenicol resistance to its host, pASK IBA7+ and 32 encode ampicillin resistance for selection of positive recombinants.

vectors encoding chaperonins	promoter	induction	selection	coding for... (size)
pREP4-groESL (5,884 bp) Caspers et al. 1994	*lac*	IPTG	Kan	GroEL (57 kDa) GroES (12 kDa)
pRDKJG (7,937bp) Caspers et al. 1994	*lac*	IPTG	Kan	DnaK (71 kDa) DnaJ (47 kDa) GrpE (25 kDa)

Table 2.8: Features of used chaperonin encoding vectors.

Sequencing was performed with primers recommended by the manufacturer. These primers were suitable for all pASK IBA vectors used in this study.

2.3.4.3 Further Vectors for Use in *E. coli*

Vectors **pREP4-groESL** and **pRDKJG** described by Caspers et al. (1994) code for chaperonins and were used to promote correct folding of recombinant proteins by co-expression with AHAScsu-containing constructs (section 3.4.1.5 on page 164). The chaperonins GroEL and GroES, encoded by pREP4-groESL, are members of the heat shock protein family with around 60 kDa in size (Hsp60) although only GroEL nearly reaches the size of 60 kDa (see Table 2.8 on page 48). Chaperonins DnaK, DnaJ, and GrpE, encoded by vector pRDKJG, are members of the Hsp70 family and only DnaK among the three encoded chaperonins on pRDKJG is fitting to the size of 70 kDa. Chaperonins are subunits of chaperones which only operate as holoenzymes. Chaperones detect their potential substrate protein in binding to a short, hydrophobic peptide sequence. Numerous hydrophobic protein sequences are exposed on the surface of un- or misfolded proteins that lends insolubility to the misfolded protein.

The composition and the catalytic cycle of chaperone GroESL is described in the following: 7 chaperonins of GroEL form a ring. Two ring layers form together a cylinder (see Figure 2.2 on page 49). Two of these cylinders stick together bottom to bottom and their tops are sealed each with one GroES protein (see Figure 2.3 on page 50). Each of the resulting two cavities can incorporate proteins with a maximum size of 60 kDa. The lid of GroES is mobile and is closed during processing. Every 15 seconds one peptide is refolded in each cavity. One cavity is processing a misfolded protein while the other cavity is releasing a correctly folded protein. There is always only one substrate in process. Each refolding process requires 7 ATP in each cavity. A short video of this catalytic cycle can be seen at http://www.cryst.bbk.ac.uk/~ubcg16z/chaperone.html (Roseman et al. 1996).

The Hsp70 chaperones require ATP as well. DnaK consists of an ATPase domain and a substrate binding domain (see Figure 2.4 on page 51). Recognition and binding is carried out by DnaJ while release is controlled by GrpE. One binding-and-release-cycle takes approximately one second. It is repeated as often as necessary to obtain correctly

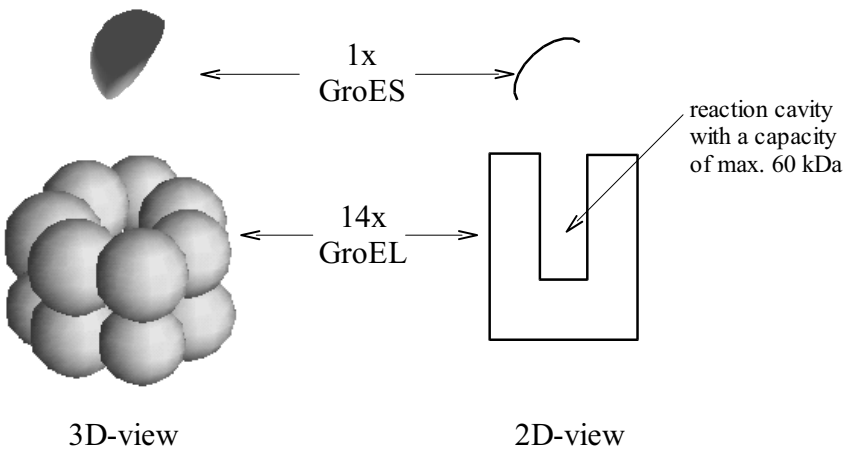


Figure 2.2: Structure of GroESL: 7 chaperonins of GroEL form a ring, two layers of rings create a cylinder which is covered with a mobile lid from GroES. Both chaperonins are encoded on vector pREP4-groESL.

folded protein (Mogk et al. 2001). In contrast to GroESL, the size is no limiting factor and substrate proteins larger than 60 kDa can be refolded with DnaK- and co-chaperonins.

The two vectors pREP4-groESL and pRDKJG provide kanamycin resistance to their hosts. They only can be co-expressed with vectors providing different selection markers like ampicillin or chloramphenicol resistance. Expression of chaperonins needs to be started with IPTG.

2.3.4.4 Plasmid Used in Assay Control

pET-GM (Hill et al. 1997) was kindly provided by Prof. R. G. Duggleby, University of Queensland, Brisbane, Australia (see Table 2.9). This plasmid contains the *ilv*GM gene encoding the AHAS II from *E. coli*. Translation of *ilv*GM results in two homologously expressed proteins: the catalytic (*ilv*G) and the regulatory subunit (*ilv*M) of AHAS II from *E. coli*. The catalytic subunit is fused N-terminally to a 6xHis-tag followed by 32-bp-spacer coding for a thrombin cleavage site and other artificial amino acids introduced when using the *Bam*HI/*Eco*RI restriction sites of pET30a(+) expression vector from Novagen. The expected mass of this fusion protein is 64.7 kDa with 5.4 kDa from artificially introduced amino acids. The native catalytic subunit is expected to show 59.3 kDa. The regulatory subunit (*ilv*M) is expressed in its native form (9.6 kDa). The vector pET30a(+) confers kanamycin resistance to the host.

2.3.4.5 Vectors for Heterologous Protein Expression in Yeast

Two different expression systems for yeast strains were tested. Vector pYES2 belongs as 2μm-DNA derivative to an extrachromosomally replicating system, while vector pKLAC1

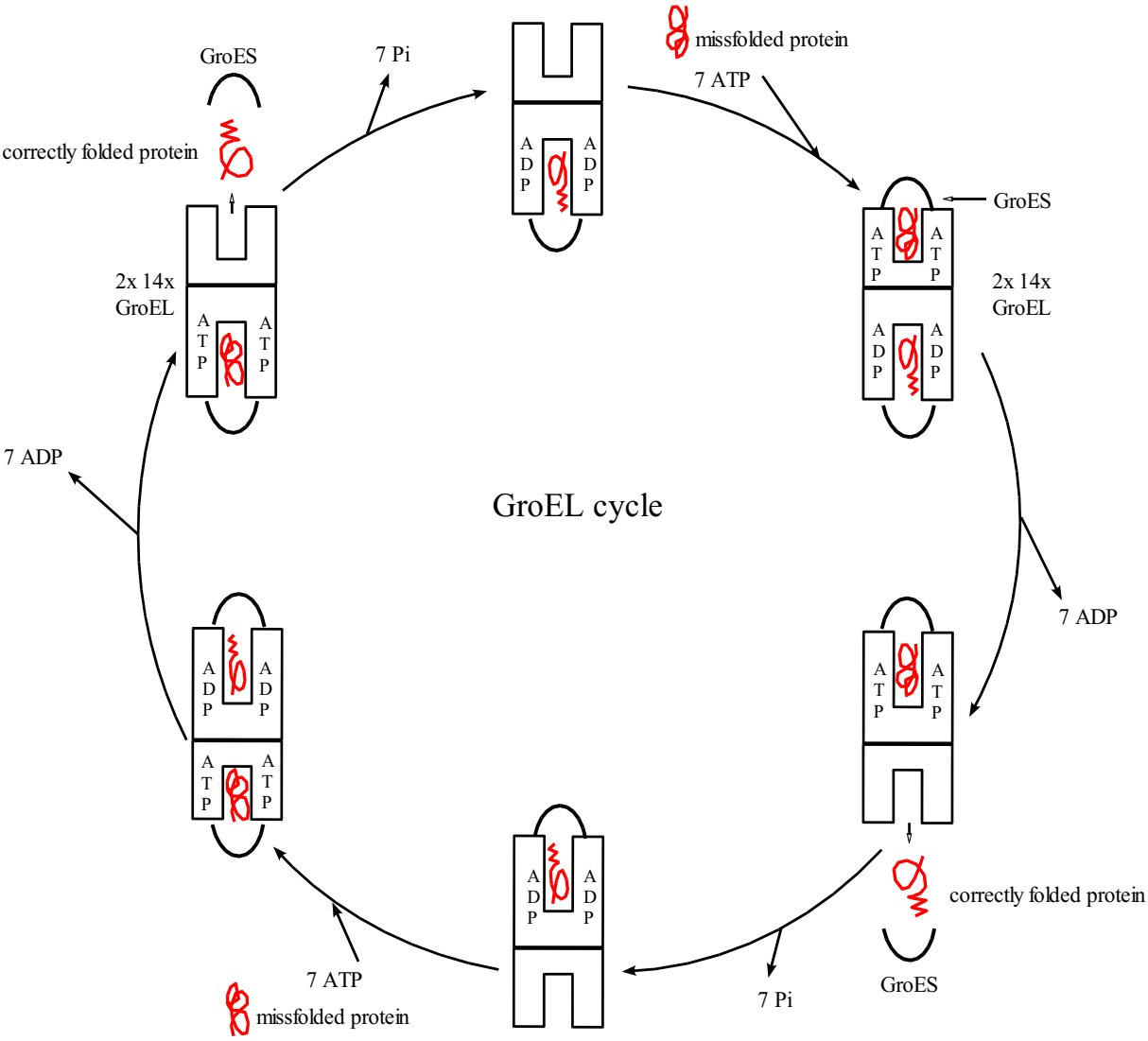


Figure 2.3: The GroEL cycle (modified from Mogk et al. 2001) uses two cavities to refold proteins up to 60 kDa. Cycling consumes 7 ATP per cavity.

vector	gene	cloned in	induction	selection	coding for (size)
pET-GM (5,884 bp) Hill et al. 1997	*ilv*GM	pET30a(+)	IPTG	Kan	AHASIIcsu (59.3 kDa) AHASIIrsu (9.6 kDa)

Table 2.9: Features of vector pET-GM encoding *E. coli* acetohydroxyacid synthase II (AHASII, csu = catalytic subunit, rsu = regulatory subunit) used as positive control when establishing a specific assay for necic acid synthase activity.

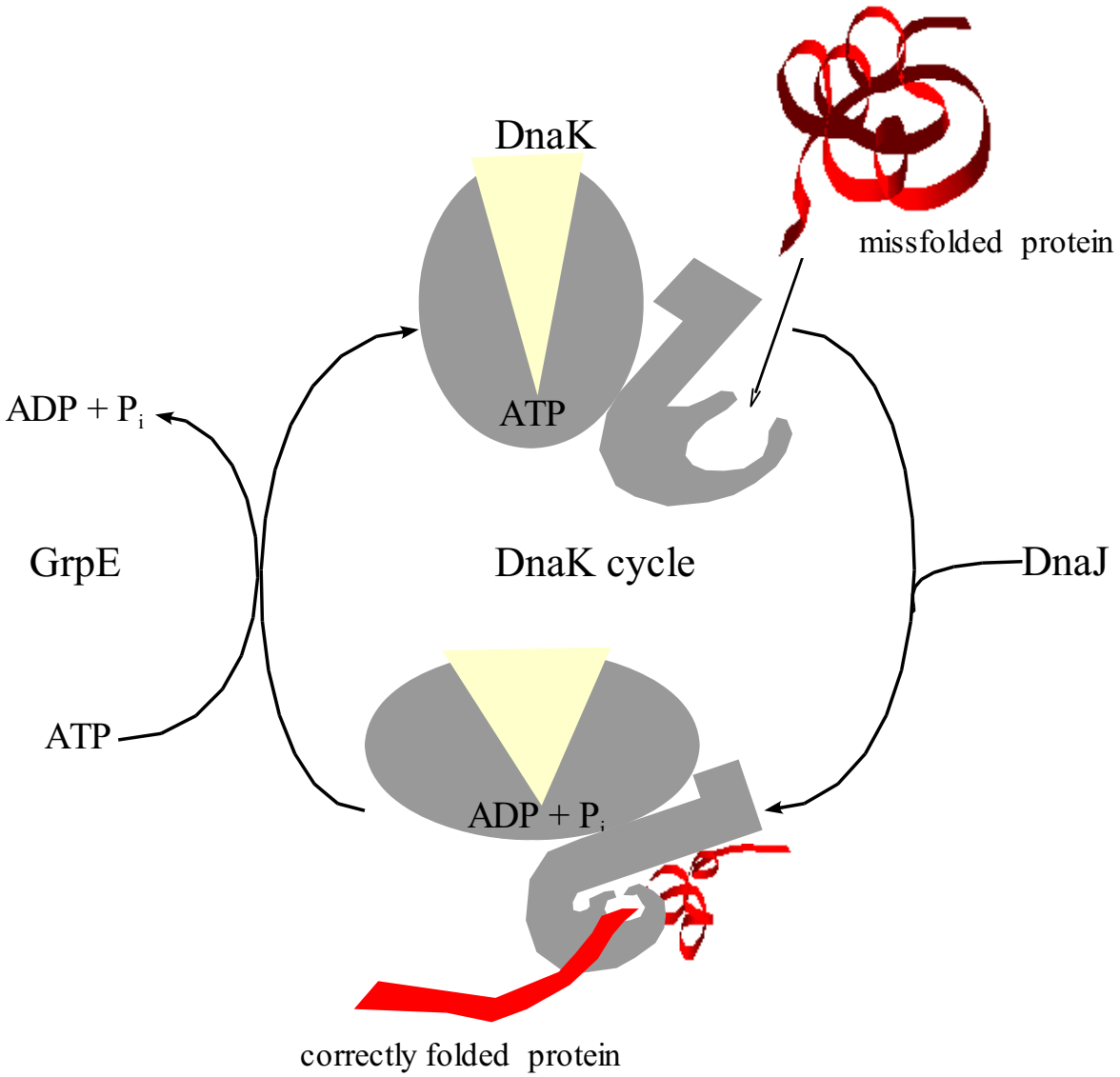


Figure 2.4: The DnaK cycle is supported by the co-chaperonins DnaJ and GrpE, modified from Mogk et al. (2001). These three proteins are encoded on vector pRDKJG.

vector	host strain	promoter	repressor	induction	selection (yeast)	selection (*E. coli*)
pYES2 (5,856 bp) Invitrogen	INV*Sc*1	*GAL1*	glucose	galactose	URA3	Amp
pKLAC1 (9,091 bp) NEB	GG799	*LAC4*	glucose	galactose	*amdS*	Amp

Table 2.10: Features of vectors used for heterologous protein expression in yeast. Nomenclature of phenotypes and abbriviations are given in the appendix. Gene *amdS* encodes for acetamidase

belongs to the group of integrative vectors that needs to be integrated in the host's genome. 2μm-DNA vectors show high copy-numbers (30 - 50 copies per cell) enabling high rates of heterologously expressed proteins together with other factors like strength of promoters. The disadvantage of 2μm-DNA vectors is that it needs to be kept under high selection pressure to reduce the number of hosts to a minimum which had not received any vectors during mitosis. With vector pKLAC1, this disadvantage is excluded: once it is integrated homologously into the host genome, it is mitotically stabile and under control of Mendel's inheriting rules. The number of genome integrated vectors is much lower than the copy numbers of 2μm-DNA derivatives in their hosts which contributes to a lowered rate of expressed protein from vector pKLAC1.

In contrast to bacteria, selection of positive transformants is achieved via minimal media: Auxotrophic hosts like INV*Sc*1 depend on the vectors' information for complementing their genetical defect; in case of non-auxotrophic yeasts like GG799, a nitrogen-free medium is offered together with a source of nitrogen, which is exploitable with the help of the vector encoded enzyme. While vector pYES2 is used for expression in *S. cerevisiae* strain INV*Sc*1 (auxotrophic), vector pKLAC1 is supplied together with the *K. lactis* strain GG799 (non-auxotrophic). For more information on the yeast host strains, see section 2.3.1.2 on page 36.

The two vectors pYES2 and pKLAC1 are shuttle vectors containing a bacterial origin of replication enabling the cloning in *E. coli*. Features of the yeast expression vectors are listed in Table 2.10.

pYES2 (5,856 bp) vector belongs to the pYES expression system (Invitrogen). For proper initiation of translation, the insert produced by PCR has to contain a Kozak translation initiation sequence and an ATG initiation codon. As advised by the manufacturer, the forward primer needs to contain the sequence ANN <u>ATG</u> G where the A at position -3 and the G at position +4 is essential for the Kozak sequence (ATG initiation codon is underlined and not variable either). The reverse primer needs to contain a stop codon. Since pYES2 does not offer any encoded tags, they have to be introduced with the primers as well. Propagation, selection, and maintenance of positive transformants need to be carried out in recombination (*rec*A) and endonuclease deficient (*end*A) *E. coli* strains like

name of primers	sequence	T_m [°C]
pYES2seqFOR	5'-dTAC CTC TAT ACT TTA ACG TC-3'	51.2
pYES2seqREV	5'-dATA GGG ACC TAG ACT TCA-3'	51.4

Table 2.11: Sequencing primers for pYES2: pYES2seqFOR was used for sequencing the 5'-end in 3'-direction, pYES2seqREV for sequencing the 3'-end in 5'-direction

TOP10, DH5α, XL1blue or JM109. pYES2 confers ampicillin resistance to *E. coli* hosts which is exploited for selection of positive transformants. Although recommended by the manufacturer, proper integration of the insert into the vector could not be checked with the standard T7 universal primer. Hence, two new sequencing primers were designed (see Table 2.11 on page 53) allowing to sequence in both directions, each starting the sequencing from vaster boundaries as the recommended T7 primer. Constructs harbouring the PCR product correctly ligated were transformed into yeast INV*Sc*1 cells (as described in section 2.3.3.2 on page 44) and were grown on SC-U minimal medium for selection (section 2.3.2.2).

pKLAC1 (9,091 bp) was obtained together with the *K. lactis* protein expression kit from NEB. It offers the opportunity to obtain heterologously expressed proteins secreted to the medium. Expression primers enable integration of the sequence directly downstream of the Kex-cleavage site that is located again downstream of the α-mating factor (α-MF) sequence. During protein biosynthesis, this factor (α-MF) directs the protein into the endoplasmic reticulum where a part of this factor is cut off with a signal peptidase. With the help of secretory vesicles, the protein is guided to the Golgi where the remaining α-MF pro-domain is cleaved by Kex endoprotease (N-Lys-Arg↓-C) resulting in a mature protein without any artificial amino acids. In the Golgi, the mature protein is packed in vesicles, transported to the plasma membrane of the host cell, and secreted to the medium. In order to achieve a mature protein without any artificial amino acid, the *Xho*I restriction site has to be used to integrate the 5'-end of the insert. The encoded Kex-cleavage site is downstream of the *Xho*I recognition site and is cut off from the vector. Hence, the sequence AAA AGA encoding the Kex-cleavage site needs to be re-integrated with the help of the forward primer. If any tags for purification are wanted to be fused to the mature protein, they have to be integrated with the primers as well because the vector offers no tag-possibilities. The constructs were cloned in *E. coli* for selection of positive transformants and for maintenance. Plasmids isolated from positive transformants were linearized by *Sac*II and transformed into *K. lactis* cells (as described in section 2.3.3.2 on page 44). Successful integration into the *K. lactis* genome allowed the host cells to grow on YCBA minimal medium (section 2.3.2.3).

2.4 Methods of Molecular Biology

In the following section, methods for identification and for handling of nucleic acids are described right to the moment, the genetic informations are heterologously expressed

into proteins. The section 2.5 proceeds with the methods used for manipulation of the heterologously expressed proteins.

2.4.1 Overview of General Procedures

Determination of unknown open reading frames (ORFs).

There are three common methods that help to discover unknown coding sequences:

1. isolate the protein of interest, sequence it partially (Edmann degradation), and design degenerate primers or

2. construct a cDNA library and screen it with primers or sequencing,

3. browse the database, create an alignment with known homologous sequences extracted from the database, and design degenerate primers.

The latter is the most convenient one if the underlying sequences from the database display a high degree of similarity. This is the case with most enzymes from primary metabolism due to the necessary maintenance of function. Figure 2.5 gives an overview of the standard approach during this project how to identify an ORF with degenerate primers.

Total RNA was isolated (section 2.4.3.1) and then transcribed to cDNA with an oligo(dT)-primer (section 2.4.8). This reduces all possible RNAs to those that contain a poly-A tail, one of the characteristic features of mRNA. Only transcribed parts of the genome, on their way from the active gene to become a functional protein, were considered. In addition, all introns were already spliced off. Once, PCR with degenerate primers was successful (section 2.4.9.1), DNA with a single 3'A overhang was cloned for maintenance and further multiplication into a vector system with sticky single 3'T overhangs at the insertion site (TA-cloning, section 2.4.10.1). Positive ligation was checked twice with blue/white selection (section 2.4.11.1) and an additional restriction analysis. Sequenced clones were checked for similarities to the database. Based on these detected internal sequences, new primers were designed for 5' and 3' rapid amplification of cDNA ends (RACE, section 2.4.9.2). Successful PCRs with the gene specific primers were again cloned for maintenance, screened for positive ligation and sequenced. Finally, the three (or more) resulting sequences were stick together virtually to complete the ORF (2.4.13).

Heterologous Expression.

The generation of protein in a foreign organism is called heterologous expression, while the production of protein from the same species in itself is named homologous expression. Once the ORF was identified, the translation start- and stop-codon regions determine the design of expression primers (EXP, section 2.4.9.4). The following "full-length PCR" (section 2.4.9.4) was conducted by a DNA polymerase with 3'→5'exonuclease activity (proof-reading) to avoid introduction of any artificial point mutations. PCR product and expression vector were digested and purified separately, and ligated with

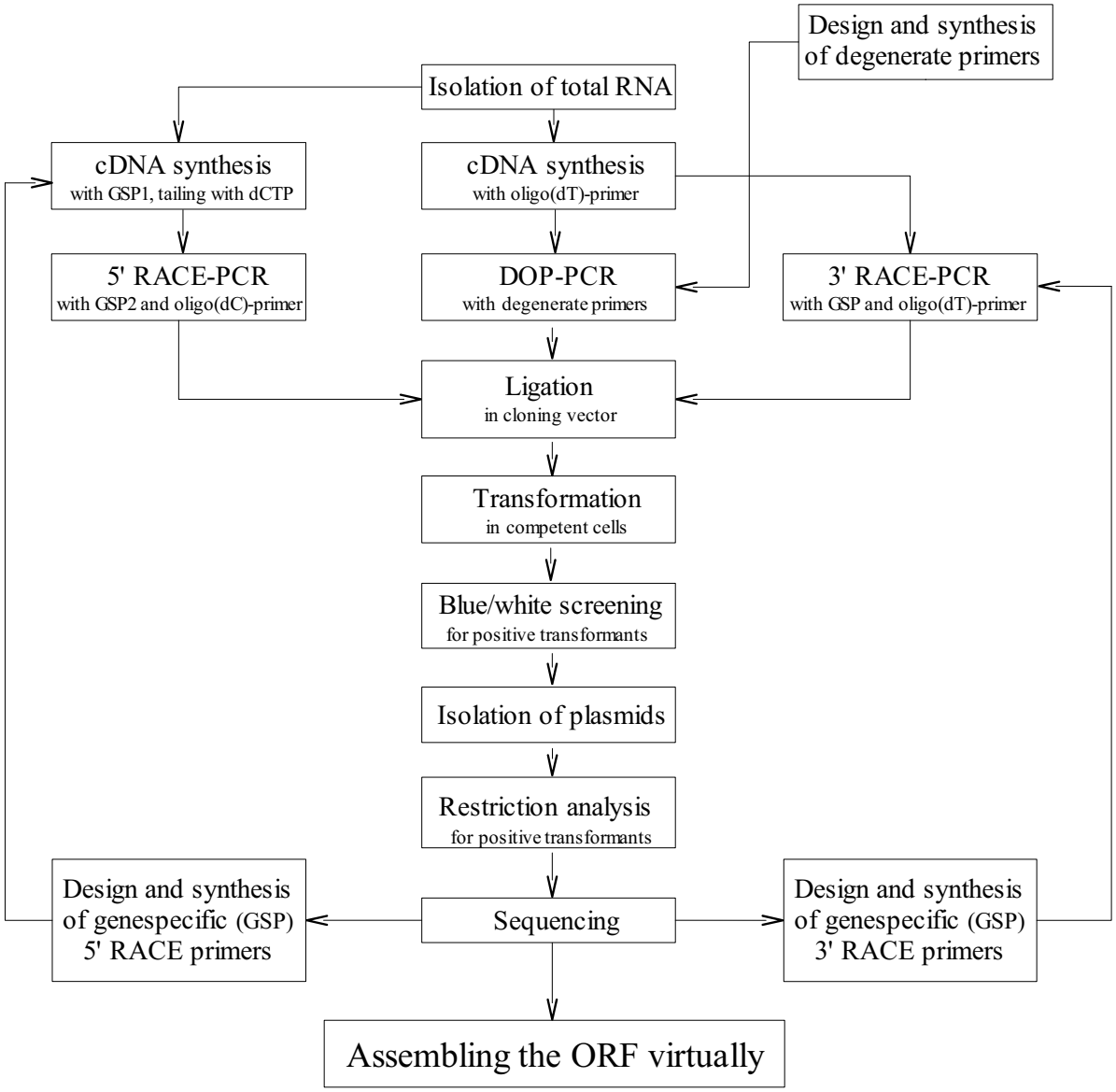


Figure 2.5: Procedure for detecting unknown ORFs. (RACE = rapid amplification of cDNA ends, GSP = gene specific primer, DOP = degenerate oligo nucleotide primed)

each other (section 2.4.10.2). The ligation product was transformed into competent host cells (section 2.3.3.1) and plated onto selective medium (section 2.3.2). Screening for successful transformants was done either by restriction analysis or PCR clone screening (2.4.11). Positive constructs were introduced to expression hosts and expression was induced (section 2.4.14). After extracting and purifying the recombinant protein (sections 2.5.1 and 2.5.2), its activity was characterized (section 2.5.10). For an easy overview of the steps involved in heterologous protein expression, see Figure 2.6.

2.4.2 Computer Assisted Sequence Analysis

Apart from **GCG Wisconsin Software Package** (Devereux et al. 1984,
http://www.gwdg.de), amino and nucleic acid sequences were analyzed with the following programmes available at the institute or freely accessible in the World Wide Web:

Genedoc, a multiple sequence alignment editor, analyzer, and shading utility for Windows used for editing and visualizing multiple sequence alignments. Cave: Genedoc treats some amino acids as equal due to their similar chemical behavior (valine = isoleucine = leucine = methionine; lysin = arginine; tyrosine = phenylalanine; threonine = serin).
http://www.psc.edu/biomed/genedoc/

BLAST (Altschul et al. 1990, Gish and States 1993), a set of programmes searching similarity in all available sequence databases used for both DNA and protein queries.
http://www.ncbi.nlm.nih.gov/BLAST

Restriction Mapper, a programme to display a map of restriction sites of nucleotide sequences.
http://www.arabidopsis.org/cgi-bin/patchmatch/RestrictionMapper.pl

FastA program searches for homologies in Uniprot database for peptides obtained by protein sequencing (see page 92).
http://www.ebi.ac.uk/fasta33/

ClustalX (Thompson et al. 1997) is a multiple sequence alignment programme for nucleic or amino acid sequences and was used for calculating alignments that were transfomed to phylogenetic trees by PHYLIP V 3.6 and presented by TreeView (Win32).
http://inn-prot.weizmann.ac.il/software/ClustalX.html and
http://www.igbmc.u-strasbg.fr/BioInfo/ClustalX/Top.html

TreeView (Win32) uses alignments to draw phylogenetic trees and was applied together with ClustalX and PHYLIP V 3.6.
http://taxonomy.zoology.gla.ac.uk/rod/rod.html

PHYLIP programme package V 3.6 (Felsenstein 2001) was used to estimate phylogenies from alignments resulting from ClustalX applying the following programmes:

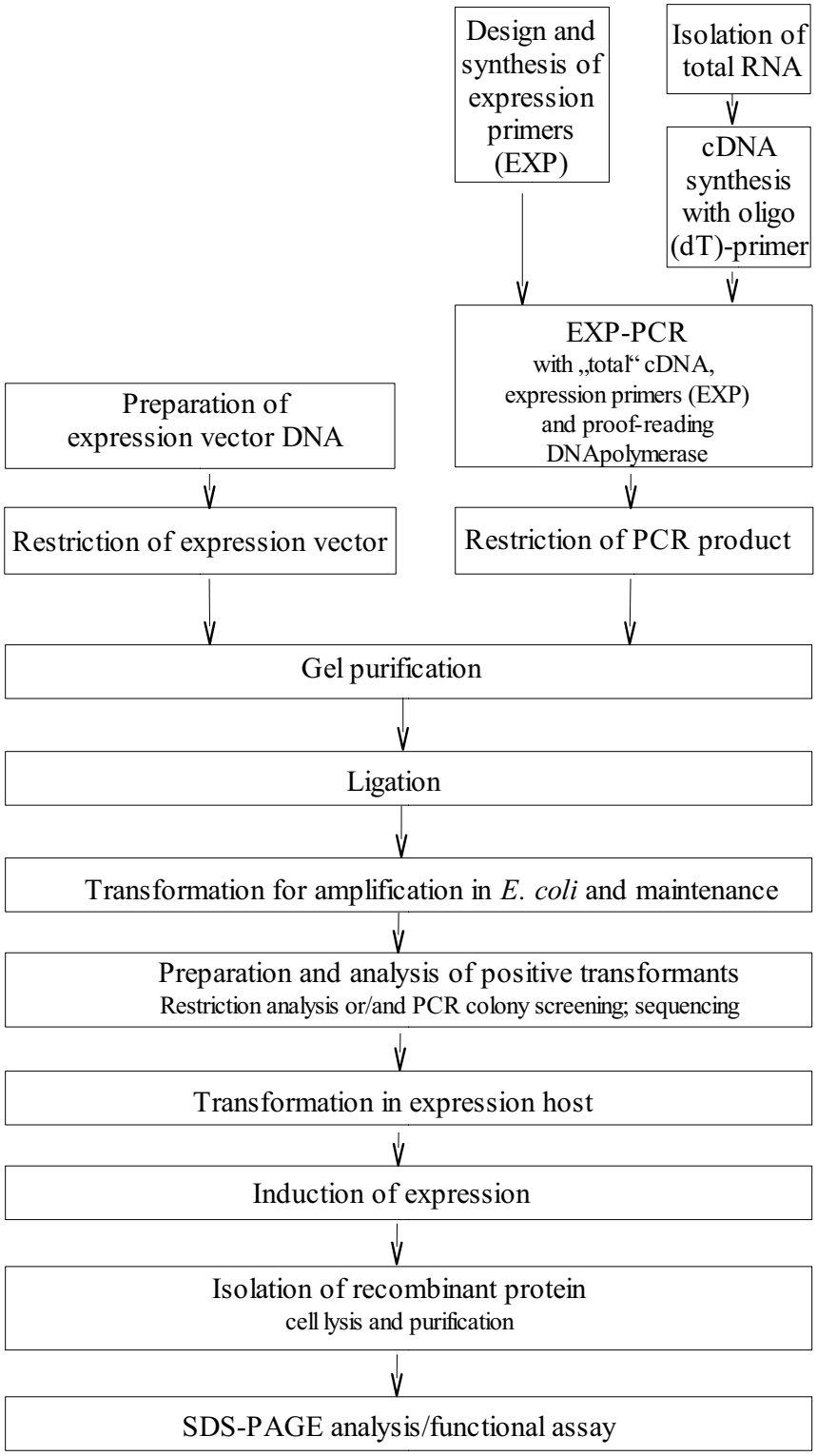


Figure 2.6: Overview of the procedure of heterologous expression (EXP-PCR = polymerase chain reaction with expression primers).

DNADIST (Kimura 1980), PROTDIST (Jones et al. 1992), both combined with NEIGHBOR as a neighbour-joining method established by Saitou and Nei (1987). Bootstrap values were estimated with SEQBOOT and CONSENSE. All bootstrap estimates result from 1,000 replicates.

TargetPV 1.1 (Emanuelsson et al. 2000, Nielsen et al. 1997), a neural network-based tool for large-scale subcellular location prediction of eukaryotic protein sequences. Considering the N-terminal sequence information, it discriminates between chloroplast (cTP) or mitochondrial (mTP) transit peptides or signal peptides for the secretory pathway (SP).
http://www.cbs.dtu.dk/services/TargetP/

SignalP V1.1 (Nielsen et al. 1997) predicts signal peptides in sequence information of pro- and eukaryotes.
http://www.cbs.dtu.dk/services/SignalP/

PSORT Predicition, a programme predicting protein localization in cell compartments or organelles.
http://www.psort.nibb.ac.jp/form/html

gcua. The graphical codon usage analyzer easily reveals parts of sequences with cumulating appearance of rarely used codons. It constructs a codon usage table from a given sequence and compares this table to those available from other species.
http://www.gcua.schoedl.de

2.4.3 Isolation of Nucleic Acids

Three major types of nucleic acids were isolated: RNA from plant material, plasmid DNA from bacteria, and total DNA from yeast.

2.4.3.1 Isolation of Total RNA from Plants

In the search for AHAS, total RNA was isolated from young parts of the plant, as the highest level of AHAS transcription was found in the metabolically active meristematic tissues (Keeler et al. 1993). High levels of transcription provide higher proportion of mRNA in total RNA that was isolated with the help of RNeasy Plant Mini Kit from Qiagen. In order to exclude as many ubiquitous RNases as possible, all instruments involved in the isolation procedure were autoclaved if possible or carefully cleaned. The isolation procedure was carried out as fast as possible to reduce the time during which RNases would be able to destroy or shorten RNA.

About 100 mg plant material were harvested and immediately frozen in liquid nitrogen. With the help of a mortar, the frozen plant material was ground to powder. The amount of total RNA being isolated corresponded directly to the crushing degree of the plant material: The smaller the particles, the more total RNA could be isolated. It was

decided not to exceed the grounding in favour of a shortened procedure which enabled the isolation of high quality RNA instead of high quantities. The still frozen and powdered plant material was solubilized in lysis buffer containing freshly added ß-mercaptoethanol. This mixture was centrifuged through a QIAshredder spin column to get rid of insoluble material and to reduce viscosity of the lysate. Ethanol, added to the cleared lysate, created conditions to promote selective binding of RNA to the silica gel based membrane of RNeasy mini spin columns. Contaminants were washed away and RNA was eluted in 2 x 30 μl RNase-free water. Immediately after eluting the RNA, two aliquots for checking quality and quantity (see section 2.4.6 on page 63) were taken. The remaining stock was stored at -80°C until used for RT-reactions (see section 2.4.8 on page 65).

The quality was checked with a 1.5% agarose gel (see section 2.4.7 on page 63): Two sharp prominent bands formed by several rRNAs (the larger band consisted out of 26S (mitochondrial), 25S (cytosolic), 23S (chloroplastidal) rRNA and the slightly smaller band out of 18S (mitochondrial), 17S (cytosolic), 16S (chloroplastidal) rRNA). A third broad but weak band contained 5.8S (cytosolic), 5S, and 4.5S (chloroplastidal) rRNA was typical for plant total RNA (Heldt 1996). High quality RNA is devoid of genomic DNA which can be seen on the agarose gel as bands of approximately 10 kDa. When the preparation did not show such bands, the amount of co-isolated genomic DNA was stated as negligible. When the two major prominent bands were not sharp but appeared as a smear of smaller sized RNAs, the preparation was discarded for the RNA sample suffered of major degradation during preparation. The mRNA itself is not visible in ethidium bromide stained gels as mRNAs have no stretches of double strands. Ethidium bromide needs double strands to intercalate in between and mark the location of nucleic acids in an agarose gel when detected under UV-light.

2.4.3.2 Isolation of Plasmid DNA from Bacteria

Three main kinds of methods were used differing in grades of purity and in the total amount of isolated plasmid DNA: manual minipreparations for high throughput of clones, kit minipreparations for isolation of utmost clean plasmid DNA applied in sequencing, or midipreparations for large amounts of pure plasmid used in cloning and expression.

Minipreparations: hand methods

Hand methods were used during cloning of PCR fragments (section 2.4.10.1) and full-length clones (section 2.4.10.2). The obtained plasmids were checked by restriction analysis or PCR-screening. The yield was sufficient (estimated by agarose gel), the elapsed time was 1.5 to 3.5 hours and the purity was tolerable, judged from detectable bands of genomic DNA on the agarose gel (section 2.4.3.1).

Hand method I (Birnboim and Doly 1979, Ish-Horowicz and Burke 1981): The pellet of an *E. coli* over-night culture (see section 2.3.3.3 on page 44) was collected in a single microcentrifuge tube (1.5 ml size). The culture volume of 1.5 ml was enough to isolate sufficient amounts of high copy plasmids while for low copy plasmids a volume of 4.5 ml was needed. LB-medium was removed completely before the pellet was resuspended in 200 μl solution I (see Table 2.12 on page 60), shortly placed on ice and lysed with 400 μl

Hand method I	Hand method II
Solution I:	STET solution:
50 mM Glucose	8% (w/v) Sucrose
25 mM Tris	50mM Tris
10 mM Sodium EDTA	50 mM Sodium EDTA
adjusted to pH 8 with HCl	0.5% (v/v) Triton X-100
and autoclaved	adjusted to pH 8 with HCl
	and autoclaved
Solution II:	Lysozyme solution:
40 μl 2 N Sodium hydroxide	10 mg/ml Lysozyme
40 μl 10% (w/v) SDS	resolved in
ad 400 μl H_2O	10 mM Tris/HCl pH8

Table 2.12: Solutions required for Minipreparations - hand methods I and II

of freshly mixed solution II (see Table 2.12), gently mixed by converting the tube three times and placed on ice for exactly 5 minutes. Genomic DNA was precipitated by adding 300 μl of autoclaved 7.5 M ammonium acetate. Ten minutes on ice were followed by 15 minutes of speed centrifugation (18,000 - 21,000 x g). Precipitated genomic DNA and cell debris sedimented while the plasmid DNA remained in the supernatant. Transferred to a fresh tube containing already isopropanol in an amount of 0.6 to 1.0 volume parts of the supernatant, the plasmid DNA was precipitated during 10 to 35 minutes at room temperature. Plasmid DNA was collected by centrifugation (15 minutes at 18,000 - 21,000 x g) and was washed with 500 μl of 70% ethanol. The resulting pellet was dried at room temperature until translucency and was resuspended in 10 mM Tris/HCl pH8 containing 10 μg/ml of RNase A to degrade RNA at 37°C for 30 minutes. The plasmid DNA was stored at -20°C until used. This method took about 3 hours for 24 preparations.

Hand method II: The rapid boiling method (Holmes and Quiqley 1981) was faster (1.5 hours/ 24 preparations), reached almost double yield and less but sufficient purity than hand method I. It was preferred when more than 24 preparations were carried out simultaneously. An amount of 1.5 ml of *E. coli* over-night culture was sufficient for all kinds of vectors (high or low copy). Harvested cells were resuspended in 500 μl STET solution (see Table 2.12 on page 60). 50 μl of freshly prepared lysozyme solution (see Table 2.12) were added and after 2 to 3 minutes at room temperature transferred to 95°C for 90 seconds for cell lysis. After centrifugation for 10 minutes (18,000 - 21,000 x g), the cell debris and genomic DNA were removed with an autoclaved wooden toothpick. The remaining plasmid DNA was precipitated with 50 μl of 7.5 M ammonium acetate and 500 μl isopropanol. Thoroughly mixed, the precipitate was sedimented by centrifugation (18,000 - 21,000 x g) for 15 minutes, air dried and resuspended in 10 mM Tris/HCl pH8 containing 10 μg/ml of RNase A for RNA degradation, as mentioned in hand method I (see page 59).

Minispin preparations using a kit

Plasmid preparation with a kit provided the necessary high purity for sequencing plasmids (2.4.12). On host cells which were not free from endonucleases like *E. coli* BL21(DE3), no hand methods could be performed. In order to gain stable and storable plasmid DNA, isolation was carried out with a kit to get rid of all endonucleases. The amount of isolated DNA tended to be rather low (100 - 1,000 ng/μl), but the elapsed time was also very short (0.5 - 1 hour).

NucleoSpin Plasmid (Macherey-Nagel): An amount of 40 μg of plasmid DNA can be cleaned with one spin column. Steps were carried out according to the manufacturer's protocol (version 03/2002/rev 01) except for the following modifications: Up to 10 ml of the *E. coli* over-night culture were treated with one spin column without clogging the membrane. The additional washing step with prewarmed buffer AW was performed routinely to decrease the amount of endonucleases to a minimum. The plasmid was eluted twice with 50 μl 10 mM Tris/HCl pH8 and quantified by UV as described on page 63. The average amount of purified plasmid DNA was between 100 to 200 ng/μl.

Wizard Plus SV Minipreps DNA Purification System (Promega): This kit turned out to yield higher amounts of purified plasmid DNA (about 200 to1,000 ng/μl). Most steps were carried out according to the manufacturer's protocol except for the following: The plasmid DNA was eluted with 50 μl of 10mM Tris/HCL pH8 instead of nuclease-free water. The concentration of eluted DNA could be increased by applying prewarmed (70°C) elution solution and by a prolonged stay (10 minutes) before eluting by centrifugation. Additionally, the elution volume could be decreased down to 25 μl.

PureYield Plasmid Midiprep System (Promega): This kit managed up to 250 ml of culture. It was observed that volumes larger than 150 ml resulted in no further increase of the total yield. According to the manufacturer's protocol, a lysed culture was cleared within two steps. First, larger particles were sedimented by centrifugation at 15,000 x g and second, particles remaining unsedimentable swimming on the surface were left back after filtration under vacuum. Elution was carried out with prewarmed (70°C) 10 mM Tris/HCl pH 8 instead of nuclease-free water. The reason for this was that DNA is stable much longer at an alkaline buffered pH of about 7.5 to 8.5 that can neutralize absorbed carbon dioxide, which otherwise would lower the pH value and destabilize the DNA.

2.4.3.3 Isolation of Total DNA from Yeast

The DNeasy Tissue Kit (Qiagen) is based on a special silica gel membrane that binds DNA selectively, while contaminants such as proteins and polysaccharides are washed away. In order to avoid clogging the membrane, the recommended maximum number of 5 x 10^7cells were determined spectrophotometrically by measuring the optical density at 600 nm. OD values vary from species to species measuring light scattering rather than absorption. Measured light scattering is dependent on the distance between the sample and the detector and, therefore, readings vary between different types of photometers. The photometer was calibrated for *Kluyveromyces lactis* by counting the cells directly in a hemacytometer chamber as proposed by the manufacturer and Ausubel et al. (1994). The appropriate amount of cell culture was harvested and resuspended in sorbitol buffer

Ammonium acetate-method	final conc.	resulting ratio
2 volumes plasmid 1 volume NH_4OAc [7.5 M]	2.5 M	1 volume of buffered plasmid
6 volumes EtOH [96%]		2 volumes of EtOH

Sodium acetate-method	final conc.	resulting ratio
9 volumes plasmid 1 volume NaOAc pH 5.2 [3 M]	0.3 M	1 volume of buffered plasmid
20 volumes EtOH [96%]		2 volumes of EtOH

Table 2.13: Precipitation of plasmid DNA: use of ammonium or sodium as monovalent cations. The resulting ratio of buffered plasmid to ethanol is always 1:2

(1 M sorbitol, 100 mM sodium EDTA, and 14 mM ß-mercaptoethanol). Lyticase was added to lyse the cell wall of yeast cells at 30°C for 30 minutes. Resulting spheroplasts were pelleted cautiously by centrifuging for 10 minutes at 300 x g. The samples were lysed over night at 55°C with proteinase K. The next day, the solution was almost translucent. RNA was digested with RNase A for 2 minutes at room temperature. The solution was prepared for transferring the DNA onto the membrane by adding buffer with high concentrated caotropic salt incubated for 10 minutes at 70°C. Absolute ethanol was added and the whole mixture including already formed precipitate was transferred to the spin column. DNA was transferred to the membrane by centrifugation, washed twice, and spun dry. Elution was done according to the manufacturer's instructions. The DNA was ready for screening on positive clones by PCR (see section 2.4.11.2 on page 78).

2.4.4 Precipitation of Plasmid DNA

For concentrating plasmid DNA, precipitation with ethanol was used as described by Sambrook et al. (1989). Two monovalent cations were used in different concentrations resulting in the same ratios of volumes (see Table 2.13).

The mixture was kept at 4°C for 30 minutes to allow precipitation. Subsequently, the solution was centrifuged in a precooled (4°C) centrifuge at 21,000 x g for 30 minutes. The supernatant was removed and the remaining DNA-pellet was rinsed with 70% ethanol. The plasmid was collected by centrifugation for 15 minutes at 21,000 x g. After removal of ethanol, the pellet was dried at room temperature and redissolved in an appropriate buffer diluted to the desired concentration.

The ammonium acetate-method was applied to sequencing analysis, while the sodium acetate-method was needed for cell-free translation and expression of AHAS (see section 2.4.14.3 on page 81). Plasmids used with the cell-free system were strongly recommended to contain no ammonium cations since this would have inhibited translation reactions.

2.4.5 Purification of DNA

PCR products needed to be purified to enable subsequent cloning in pGEM T easy vector (see section 2.4.10.1 on page 76) or to enable digestion and cloning in expression vectors (see section 2.4.10.2 on page 76). cDNAs obtained in a reverse transcription reaction (see section 2.4.8 on page 65) were purified before tailing (see section 2.4.9.2 on page 72).

Direct DNA clean up was carried out with the help of NucleoSpin Extract II (Macherey-Nagel) kit. An amount of 15 μg DNA with 50 - 10,000 bp in size was cleaned with one spin column. The PCR product was transferred to the membrane according to the manufacturer's protocol. The DNA was washed three times and dried for 3 minutes at 70°C. Elution was carried out with 15 to 25 μl of prewarmed (70°C) 10 mM Tris/HCl pH 8 and, after incubation for 1 minute at room temperature, the purified DNA was spun off the column.

Agarose gel extraction was done with QIAEX II (Qiagen) kit. About 100 mg excised gel obtained from agarose gel electrophoresis (see section 2.4.7 on page 63), containing PCR products in a range of 100 to 4,000 bp, were extractable with 10 μl matrix supplied with the kit. As described by the manufacturer, the PCR product excised from the agarose gel was transferred to the matrix. The matrix was washed three times and the pellet was dried until it turned completely white. For elution, the dry pellet was resolved in 20 μl of 10mM Tris/HCl pH 8, incubated at room temperature for at least 5 minutes, and spun off for 30 seconds. Although this method was more time consuming than the spin column method previously described, there was one main reason to use it: In case there was more than one distinct band on the agarose gel, it was possible to separate those bands from each other.

2.4.6 Quantification of Nucleic Acids

RNA- or DNA- concentrations were measured with an UV/Visible spectrophotometer (Ultraspec 3100 pro, Amersham Pharmacia Biotech) at 260 and 280 nm. At 260 nm, nucleic acids have a local maximum, while at 280 nm, contaminating proteins have their maximum (Lottspeich and Zorbas 1998). With the ratio A_{260}/A_{280} the purity of nucleic acid solutions is determined: For high purity, the ratio should lay between 1.8 or 2.0 depending on the kind of nucleic acid (see Table 2.14). Samples spoiled with proteins were not used for sequencing. The nucleic acids were diluted 1:100 in a volume of 500 μl and measured in a 10 mm light path quartz cuvette.

2.4.7 Separation of Nucleic Acids: Agarose Gel Electrophoresis

As a standard method for separating nucleic acids in respect to their molecular size, agarose gel electrophoresis was applied. When applied to isolated total RNA, this method informed about the quality of samples (see section 2.4.3.1 on page 58), when applied to PCR products, the method helped to distinguish between expected PCR products and artefacts (see section 2.4.9 on page 67), and when conducted during restriction analysis (see section 2.4.11.3 on page 79) positive recombinants were identified. Despite of this

kind of nucleic acid	conc. [μg/ml] when $A_{260} = 1$	demanded purity determined as ratio $\frac{A_{260}}{A_{280}}$
ssRNA	40	≥ 2.0
dsDNA	50	≥ 1.8
ssDNA, oligos	30	≥ 1.8

Table 2.14: Concentration and purity of nucleic acid solutions determined with UV/Vis spectrometry. ss = single stranded, ds = double stranded

qualitative use, agarose gel electrophoresis was used as a preparative method as well: Separated bands from PCR products or from digested vectors were cut out and submitted to purification with the gel extraction method (see section 2.4.5 on page 63).

Gelatinous agarose forms a matrix with defined size of meshes adjustable by altering the concentration of agarose. Phosphate groups within the nucleic acids are charged negatively and draw the complete molecules through the matrix towards the anode when the power is switched on. The longer a molecule is, the slower it moves through the meshes of the matrix. The size of separated bands was estimated with the help of standard DNA molecules of known size which run on the same gel as an internal standard.

The concentration of agarose was chosen according to the size of nucleic acids which had to be separated. Fragments larger than 2 kb were chromatographed in gels of 0.8% - 1.0% agarose. Smaller fragments from 2 to 1 kb separated in 1.0%- to 1.5%-agarose gels, and for fragments smaller than 1kb, 1.5%- up 3%-agarose gels were used. Two types of agarose were in use: agarose from Roth and a low melting type (Helena Bioscience) applied to TOPO TA cloning (see page 75).

Agarose was dissolved in 1x TBE buffer (see Table 2.15), the fill level of the solution was marked and the mixture was heated in a microwave oven with 600 watt until solubilization was complete (1 - 3 minutes, depending on the total amount of solution). The mixture was allowed to cool down and an evaporated buffer was refilled up to the mark. Under a hood, 1.5 μg ethidium bromide per 1 ml gel were added before the gel was cast onto a gel tray equipped with a comb forming slots. The gel was allowed to solidify at room temperature or at 4°C which helped to pull out the comb much easier.

A sample of nucleic acids was mixed 1:1 (v/v) with the loading buffer (see Table 2.15) before it was transferred to the gel which was submersed in 1x TBE buffer. The loading buffer had two functions: First, it facilitated the application of the samples to the slots because the density of the samples was increased (responsible: Ficoll 400, copolymer of sucrose and epichlorohydrin), i.e. the samples sank to the bottom of the slots instead of floating away. Second, the loading buffer visualized the migration of samples through the matrix with the help of its coloured compounds xylencyanol and bromphenol blue.

Electrophoresis was performed in a Sub-Cell GT agarose gel electrophoresis chamber from Biorad at a voltage of 80 mV supplied by Consort E815 (LTF). The separating process took 25 minutes to 1 hour depending on the density of the gel. It had to be taken into account that the gel was prestained which meant that the staining agent ethidium bromide migrated in opposite direction compared to the nucleic acids. A prolonged

10x TBE Buffer		
Trisaminomethane (Roth)		108 g
Boric Acid (Roth)		55 g
EDTA (Roth)		7.3 g
H_2O	up to	100 ml

Loading Buffer		
Xylencyanol		0.25 g
Bromphenol Blue		0.25 g
Ficoll 400		25.0 g
EDTA (Roth)		1.46 g
H_2O	up to	100 ml

Table 2.15: Buffers for agarose gel electrophoresis. Loading buffer solution was filtered and sterilized by autoclaving.

Reverse Transcriptase	optimal reaction temperature [°C]
M-MuLV RT [200 U/μl] (Fermentas)	42
SSII [200 U/μl] (Invitrogen)	42 - 50
SSIII [200 U/μl] (Invitrogen)	42 - 55
eAMV RT [20 U/μl] (Sigma)	42 - 65

Table 2.16: Reverse transcriptases, used in this work, differ in their optimal reaction temperature

separating process could cause a loss of ethidium bromide in lower bp areas of the gel. Separated and stained nucleic acids were detected under UV-light (300 nm) and documented with a multi-image light cabinet (Alpha-Imager, Biozym).

2.4.8 Synthesis of First Strand cDNA with Reverse Transcription

Reverse transcription is carried out by reverse transcriptases which need an oligonucleotide (primer) as a starting molecule. Reverse transcriptases are RNA depending DNA polymerases which use single stranded RNA as a template for synthesizing the complementary DNA strand. The DNA-primer anneals to a complementary RNA strand and the resulting double strand is recognized by the reverse transcriptase to start the reaction that results in RNA-DNA hybrids. Depending on the enzyme in use, the original RNA strand was destroyed either directly after formation of RNA-DNA hybrids (RevertAidM-MuLV Reverse Transcriptase, shortly named M-MuLV RT and enhanced Avian Reverse Transcriptase, eAMV RT) because these enzymes include RNase activity, or with the help of extra added RNase H (SuperScriptII, shortly named SSII and SuperScriptIII, SSIII). The main reason for the use of four different reverse transcriptases laid in the enzymes' ability to stand higher reaction temperatures (see Table 2.16) which helped to minimize secondary structures in the RNA. Secondary structures are known to stop the amplification because RNA double strands are not accessible to the enzyme. The resulting cDNA is shorter than its RNA matrix.

Total RNA obtained from sterile plant material (see section 2.4.3.1 on page 58) was reversely transcribed into copy DNA (cDNA). The resulting single stranded cDNAs were used as a template in polymerase chain reactions (see page 67). For degenerate oligonu-

cleotide primed (DOP) PCR, 3'RACE-PCR, or for full-length PCR (see pp 69, 71, 73), the reverse transcription was started with an oligo(dT)-primer (see Table 2.19 on page 71) which annealed to the poly-A tail of all mRNAs and created a pool of cDNAs. For 5'RACE-PCR, a gene specific primer (GSP1) was used (see section 2.4.9.2, page 71) which minimized the number of different cDNAs (see Figure 2.7).

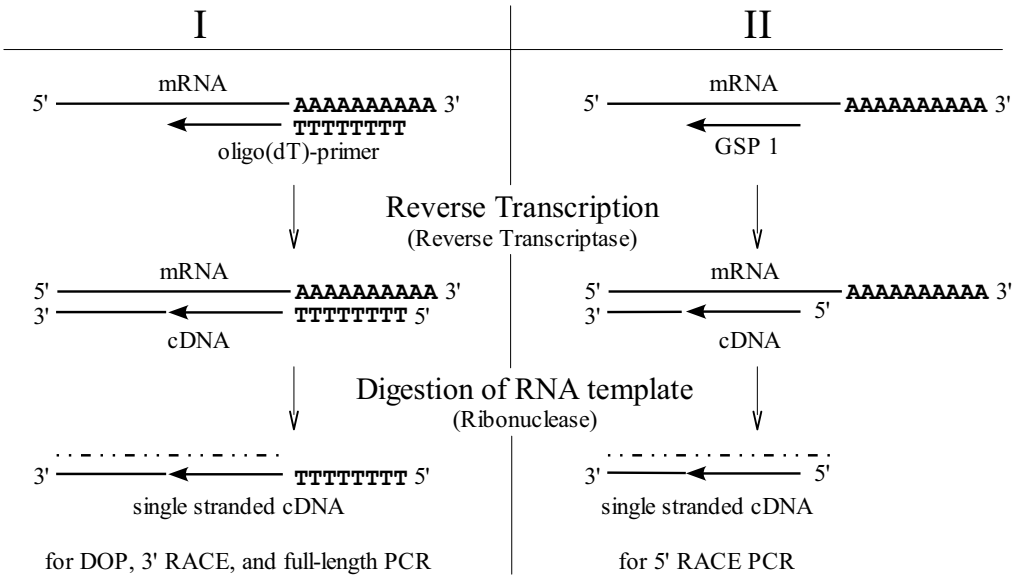


Figure 2.7: Synthesis of first strand cDNA with reverse transcriptase: I = oligo(dT)-primed for DOP, 3' RACE and full-length PCR, II = GSP 1-primed for 5' RACE PCR

The following components were mixed in a microcentrifuge tube and heated to 70°C for 10 minutes to denature secondary structures in the RNA:

content of tube 1	M-MuLV RT	SSII, SSIII or eAMV RT
total RNA	1 μg	1-5 μg
oligo(dT)-primer [1 pM]	2.5 μl	-
GSP1 [1 pM]	-	2 μl
dNTPs [10 mM each]	2 μl	1 μl
H_2O	ad 13 μl	ad 14 μl

Meanwhile, a second tube was prewarmed for 2 minutes to 42°C (M-MuLV RT or SSII) or to 50°C (SSIII or eAMV RT). This tube contained:

content of tube 2	M-MuLV RT	SSII, SSIII or eAMV RT
5x 1st strand buffer	4 μl	4 μl
DTT [0.1 M]	-	1 μl
H_2O	2 μl	-

The contents of the two tubes were combined. Finally, 1 μl of enzyme was added and

incubated at 42°C (M-MuLV RT), 50°C (SSII, SSIII), 55°C (SSIII, eAMV RT) or 60°C (eAMV RT) for 1 hour.

The reaction was finished by heat inactivation of the enzyme at 70°C for 15 minutes. In case of SSII or SSIII, a digestion of the RNA strand was performed in adding 1 μl of RNase H [2 U/μl] and incubating at 37°C for 20 minutes. The cDNA was stored at -20°C until used for PCR or tailing reactions (see pages 67 and 72).

2.4.9 Polymerase Chain Reaction (PCR)

In contrast to reverse transcription (see above, page 65) which is catalyzed by RNA depending DNA polymerases, the polymerase chain reaction is catalyzed by a thermostable DNA depending DNA polymerase. With this method, DNA is amplified exponentially (Saiki et al. 1985, Mullis and Faloona 1987, Saiki et al. 1988). The amplification requires a single stranded DNA matrix, desoxynucleotides as a substrate, and oligonucleotide primers that form double stranded regions within the single stranded matrix as a recognition and starting site for the enzyme reaction.

Typical PCR reaction mixtures depended on the DNA polymerase in use and were composed as given below:

components of the PCR reaction mixture	*Taq* [5 U/μl] (Promega, Invitrogen) Accu*Taq* (Sigma) [μl]	Platinum *Pfx* [2.5 U/μl] (Invitrogen) [μl]
forward primer [10 pmol/μl]	2	2
reverse primer [10 pmol/μl]	2	2
template	1	2
10x PCR buffer	2.5	2.5
Mg^{2+} [50 mM]	0.75[1]	0.50
dNTPs [10 mM each]	0.50	0.50
DNA polymerase	0.13	0.25
H_2O	ad 25	ad 25

Composition of standard PCR reaction mixtures depending on the DNA polymerase in use; [1] For reactrions with Accu*Taq*, Mg^{2+} was already supplied with 10x PCR buffer.

Double stranded DNA template is denatured to single stranded DNA by heating to 94°C (denaturing step). Double stranded regions within the single stranded matrix, which initiate the reaction, are formed when the temperature is lowered and the primers anneal to their complement site (annealing step). Each of the primers anneals to one of the opposite strands. Temperature is increased to the optimum of the enzyme and the primers

	constant PCR programme		touch down PCR programme	
I. initial denaturation	94°C	5 min	94°C	5 min
II. cycling			cycling no.1 repeated 21x:	
1. denaturing step			94°C	45 sec
2. annealing step			60 - 50°C[1]	1 min
3. extension step			72°C	2 min
			cycling no.2 repeated 14x:	
1. denaturing step	94°C	45 sec	95°C	45 sec
2. annealing step	T_m - 5°C	90 sec	50°C[2]	1 min
3. extension step	68°C	10 min	72°C	2 min
	repeated 35x			
III. final extension	68°C	10 min	72°C	10 min
	15°C	∞	15°C	∞

[1]Annealing temperature was gradually decreased 0.5°C/cycle
[2]Annealing temperature was kept constant at the level of the last cycle in cycling II

Table 2.17: PCR cycling programms. Tm = primer melting temperature

are prolonged in 5'direction (extension step). In general, this sequence of altering the temperature (cycling = II., see Table 2.17) was repeated 35 times.

Before the first PCR cycle was started, an initial denaturation (I.) was carried out to ensure complete denaturation of all double strands. Although in case of cDNA (see section 2.4.8 on page 65), this might have not been necessary because RT-reactions only produce single stranded cDNAs, it was performed to resolve all other secondary structures that might inhibit correct annealing of primers. In case of double stranded plasmid DNA, denaturing was essential. The PCR programme was finished with a final extension (III.) at 68°C for 10 minutes ensuring to complete all generated DNA strands. The PCR product was kept at 15°C in the cycler until it was stored at 4°C. PCR products were analyzed using agarose gel electrophoresis (see page 63).

In case of significantly different primer melting temperatures (T_m), a touch down PCR programme was applied (see Table 2.17). In cycling no.1, the annealing temperature was reduced 0.5°C in each cycle in order to promote synthesis of PCR products resulting from specific primer annealing. The last annealing temperature of cycling no.1 was chosen for a constant annealing temperature in cycling no.2. which was repeated as many times as necessary to yield 35 cycles.

PCR cycling was carried out in Eppendorf Mastercycler personal or Mastercycler gradient and was used to synthesize sufficient amounts of DNA for cloning (section 2.4.10.1) or to screen for successfully cloned plasmid DNA (section 2.4.11.2).

2.4.9.1 Degenerate Oligonucleotide Primed (DOP) PCR

Degenerate oligonucleotide primed PCR was used to discover internal parts of unknown open reading frames (DOP-PCR DNA) with the help of degenerate primers designed as described below. It was possible to discover several similar internal parts of cDNAs with just one PCR. As a template, a pool of cDNAs was used which was obtained from reverse transcription (section 2.4.8) with an oligo(dT)-primer and a pool of mRNAs. Suitable PCR conditions had to be figured out individually for each pair of primers and every plant species. DOP-PCR was conducted with *Taq* DNA polymerase (section 2.4.9).

Design of degenerate oligonucleotide primers

Each amino acid is encoded in a triplet of nucleic acids (see Table 2.18) where most amino acids have several possible encoding triplets since the genetic code is degenerate.

Degenerate primers are a mixture of several slightly varying primers. The number of different primers is reflected in the degree of degeneracy. This degree should be kept as low as possible, since the concentration of a single type of primer decreases with the increase of degeneracy. If this concentration is too low, PCR will not work. The degree of degeneracy is calculated by counting the number of different nucleic acids (max. 4) at each position in the putative primer sequence and by multiplying these degeneracy values for the whole primer sequence. The amino acid serin, for example, has a degeneracy of 2 (T, A) x 2 (G, C) x 4 (G, A, T, C) = 16.

Using the method described for discovery of desoxyhypusine synthase (DHS) from *Nicotiana tabacum* by Ober and Hartmann (1999a), the database NCBI was searched for amino acid sequences from plant AHAS. With the help of the software GCG, a multiple alignment of amino acids was created. Sections with high conservation were sought in the amino acid alignment to create primers of low degeneracy. The search was supported by the software Genedoc that shades black all significantly similar sections of the amino acid alignment. Selected sections in the amino acid alignment were translated into nucleic acids and were synthesized.

Degeneracy of degenerate primers was reduced by using an inosine in the primer sequence. Inosine is a purine that forms hydrogen bonds to cytidine, thymidine, and adenosine but does not bind to guanosine. Inosine was put in the middle of the primer where a tight linkage to the complementary strand was not as necessary as at the end positions. If inosine is used in positions that demands all the four nucleic acids, the degeneracy of the primer pool can be reduced 4-fold. However, the use of an inosine promotes the occurrence of I:G mismatches. Moreover, it has to be taken into account that the DNA polymerase, which is working together with the inosine-containing primer in the PCR reaction, must be capable of synthesizing DNA over the inosine position. *Taq* DNA polymerase is able to do so, proof-reading DNA polymerase is not and PCR will end up in a reaction stop. The degenerate primers were used for degenerate oligonucleotide primed PCR (see page 69).

	Ala (A)			Arg (R)			Asn (N)			Asp (D)			Cys (C)		
	G	C	G	C	G	G	A	A	T	G	A	T	T	G	T
			A	A		A			C			C			C
			T			T									
			C			C									
	1x	1x	4	2x	1x	4	1x	1x	2	1x	1x	2	1x	1x	2
degen	4x			8x			2x			2x			2x		

	Gln (Q)			Glu (E)			Gly (G)			His (H)			Ile (I)		
	C	A	G	G	A	G	G	G	G	C	A	T	A	T	A
			A			A			A			C			T
									T						C
									C						
	1x	1x	2	1x	1x	2	1x	1x	4	1x	1x	2	1x	1x	3
degen.	2x			2x			4x			2x			3x		

	Leu (L)			Lys (K)			Met (M)			Phe (F)			Pro (P)		
	C	T	G	A	A	G	A	T	G	T	T	T	C	C	G
	T		A			A						C			A
			T												T
			C												C
	2x	1x	4	1x	1x	2	1x	1x	1	1x	1x	2	1x	1x	4
degen.	8x			2x			1x			2x			4x		

	Ser (S)			Thr (T)			Trp (W)			Tyr (Y)			Val (V)		
	T	C	G	A	C	G	T	G	G	T	A	T	G	T	G
	A	G	A			A						C			A
			T			T									T
			C			C									C
	2x	2x	4	1x	1x	4	1x	1x	1	1x	1x	2	1x	1x	4
degen.	16x			4x			1x			2x			4x		

Ala = alanine, Arg = arginine, Asn = asparagine, Asp = aspartic acid, Cys = cysteine, Gln = glutamine, Glu = glutamic acid, Gly = glycine, His = histidine, Ile = isoleucine, Leu = leucine, Lys = lysine, Met = methionine, Phe = phenylalanine, Pro = proline, Ser = serine, Thr = threonine, Trp = tryptophan, Tyr = tyrosine, Val = valine
A = adenosine, C = cytidine, G = guanosine, T = thymidine
degen. = degeneracy

Table 2.18: The degenerate genetic code

name / direction	function	sequence	length [bp]	T_m [°C]
oligo(dT) / reverse	internal part, 3'RACE	5'-dGTC GAC TCG AGA ATT CTT TTT TTT TTT TTT TTT-3'	33	59.5
AAP / forward	5'RACE	5'-dGGC CAC GCG TCG ACT AGT ACG GGI IGG GII GGG IIG-3'	36	77.4
AUAP / forward	5'RACE	5'-dGGC CAC GCG TCG ACT AGT AC-3'	20	65.7

A = adenosine, C = cytidine, G = guanosine, I = inosine, T = thymidine

Table 2.19: Standard primers for RACE-PCR

2.4.9.2 Rapid Amplification of cDNA Ends (RACE) PCR

RACE-PCR was performed with gene specific primers (GSP) to complete the internal fragments from DOP-PCR in 5' and 3'direction (Frohmann et al. 1988). While GSPs were designed individually as described below (see section "Design of gene specific primers (GSP)" on page 71), the complementing primers for PCR are shown in Table 2.19. Templates for RACE-PCR were synthesized as shown in section "Synthesis of first strand cDNA with reverse transcription" on page 65.

Design of gene specific primers (GSP)

Unlike degenerate primers which represent a pool of different oligonucleotides, gene specific primers (GSPs) are primers with one defined sequence. They were designed according to a stretch of known sequence. The primers were positioned in the known sequences to produce an overlap of about 100 to 150 bp. In order to increase annealing specificity, GSPs were created the way that within the last 5 nucleic acids a minimum of 3 nucleic acids belonged to the A/T-group. A/T nucleotides form only two hydrogen bonds with the complementary strand while C/G create three bonds. If in the last 5 nucleic acids, there are only Cs or Gs, they stick strongly to the complementary strand with 15 hydrogen bonds; if 3 of these C/G are replaced by A/T, the number of hydrogen bonds is reduced to 12 and the binding is less sticky. This reduced stickiness avoids mismatches. The GSP length of about 25 to 30 nucleic acids was adjusted to reach a T_m of around 55° up to 65°C. The T_m was calculated with A/T counting for 2°C and C/G counting for 4°C (2°/4° rule). They were used in RACE-PCRs (see page 71). In case of the 3'RACE-PCR, the GSP was constructed as forward primer while in case of 5'RACE-PCR, GSPs were designed as reverse primer.

3'RACE-PCR

3'RACE-PCR completes the ORF in 3'direction. The 3'end of an ORF is recognizable by a stop codon. In cDNAs, this stop codon is followed downstream by a poly-A tail. The poly-A tail originates from a reversely transcribed mRNA. The reverse primer for 3'RACE-PCR (oligo(dT)-primer) anneals to this poly-A tail region, while the gene specific forward primer is positioned upstream within an region of known sequence. 3'RACE-PCRs were

carried out with *Taq* DNA polymerase (see page 67).

5'RACE-PCR

5'RACE-PCR completes the ORF in 5'direction. The 5'end of an ORF is characterized by a putative ATG initiation codon. Unlike 3'RACE-PCR that can exploit a natural occurring poly-A tail for annealing, templates used in 5'RACE-PCR need to obtain a primer annealing region. This region is introduced via tailing (reaction is described below in section "Tailing of cDNA" on page 72). After tailing, the forward primer (AAP = abridged anchor primer, oligo(dG)-primer) can anneal to an artificially added poly-C tail at the 3'end of a cDNA. 20 nucleotides of AAP function as an adapter for AUAP (= abridged universal anchor primer) which allows to perform a nested PCR (see section 2.4.9.3 on page 72). 5'RACE-PCRs were carried out with Accu*Taq* DNA polymerase (see page 67). For a quick overview of 5'RACE procedures, see Figure 2.8 on page 74.

Tailing of cDNA

Tailing introduces a homopolymeric sequence at the 3'end of a purified, single stranded cDNA. This sequence is complementary to the forward primer (AAP) used in 5'RACE-PCR. cDNA was obtained from reverse transcription (as described on page 65) and was purified with the help of a kit (see section 2.4.5 on page 63). The reaction is catalyzed by a terminal deoxynucleotidyl transferase (TdT, Invitrogen) and was composed of:

5x tailing buffer (Invitrogen)	5 μl
dCTPs [2 mM]	2.5 μl
cDNA, purified	10 μl
H_2O	6.5 μl
TdT [15 U/μl] (Invitrogen)	1 μl

All reaction compounds were premixed except for TdT. The mixture was heated to 94°C for 3 minutes to denature cDNA and chilled on ice for 1 minute. The enzyme was added and the reaction was performed at 37°C for 10 minutes. The enzyme was inactivated at 65°C for 10 minutes and the tailed cDNA was stored at -20°C until used for 5'RACE-PCR.

2.4.9.3 Nested PCR

In order to increase specificity, nested PCRs are performed. Nested PCRs re-use the PCR-DNA of a first PCR reaction but, this time, the second PCR is carried out with different primers. These primers anneal to regions within the PCR-DNA of the first reaction. The resulting PCR product is shorter than the starting PCR-DNA. In case of 5'RACE technique, the first PCR was performed with the primers AAP and either GSP1 or GSP2 (GSP1 was already used during reverse transcription, see above on page 65) and the nested PCR was done with AUAP and GSP2 or GSP3, respectively. Strictly speaking, nested PCR performed with the pair of primers AUAP/GSP3 or AUAP/GSP2 are no real nested PCRs because AUAP does not anneal within, but directly on the sequence yielded from AAP. Nested PCRs were performed routinely on 5'RACE-PCR-DNA but they are

applicable to other types of PCR-DNA as well. PCR-DNA taken from the first reaction was diluted 1:100 if not indicated otherwise. Nested PCRs were carried out with Accu*Taq* DNA polymerase (see page 67).

2.4.9.4 Full-Length PCR

Full-length PCRs were carried out with expression primers (EXPs) and resulted in full-length PCR-DNA that was cloned in expression vectors. The design of EXPs is described below on page 73. Full-length PCRs were catalyzed by Platinum *Pfx* DNA polymerase (Invitrogen) that offers a 3'→5' exonuclease activity (proof-reading) to avoid the introduction of artificial point mutations. Standard cycling parameters of a constant PCR programme (see Table 2.17 on page 68) were slightly modified to gain sufficient amounts of PCR products for cloning: Annealing temperature was 10°C lower than T_m, annealing time was extended to 4 minutes, and elongation time was 20 minutes in the case of an elongation temperature of 68°C or 10 minutes in the case of 72°C.

Design of expression primers (EXP)

Expression primers are GSPs designed to amplify a whole ORF for cloning into an expression vector. In addition to the specific sequence of the proven cDNA, a unique restriction site at the 5'end of each forward and reverse primer allowed directed ligation into the expression vector. The forward and reverse primers contained the initiation and the stop codon of the ORF, respectively, if not indicated otherwise. To facilitate the restriction and ligation procedure (see section 2.4.10.2 on page 76), two restriction enzymes were chosen that cut in the same buffer system.

A typical EXP was constructed like this: 5'—ATAT—restriction site—gene specific sequence ($T_m \approx 70$°C, according to the 2°/4° rule mentioned above on page 71)—3'. The 5' "ATAT"-element ensured efficient digestion close to the cDNA end. A list with the number of necessary base pairs between 5'end and restriction site is given in the appendix (see Table B.5 on page 222). If necessary, the sequence coding for a tag could be introduced by the primer. In most cases, tags were already supplied with the expression vector.

2.4.9.5 PCR Colony Screening

PCR colony screening was used to screen for positive recombinants and could be performed with any kind of GSP fitting to the insert. For more details, see section "Selection of positive transformants" on page 78.

2.4.9.6 RT-PCR for Tissue-Specific Transcript Localization

For tissue-specific transcript localization, total RNA was harvested from different tissues of a single plant (except for the roots from root-tissue culture) at once. It was possible to assign the occurrence of a specific mRNA to a specific tissue. RT-PCR for tissue-specific transcript localization was performed with primers of sufficient specificity. The reaction was catalyzed by Accu*Taq* DNA polymerase.

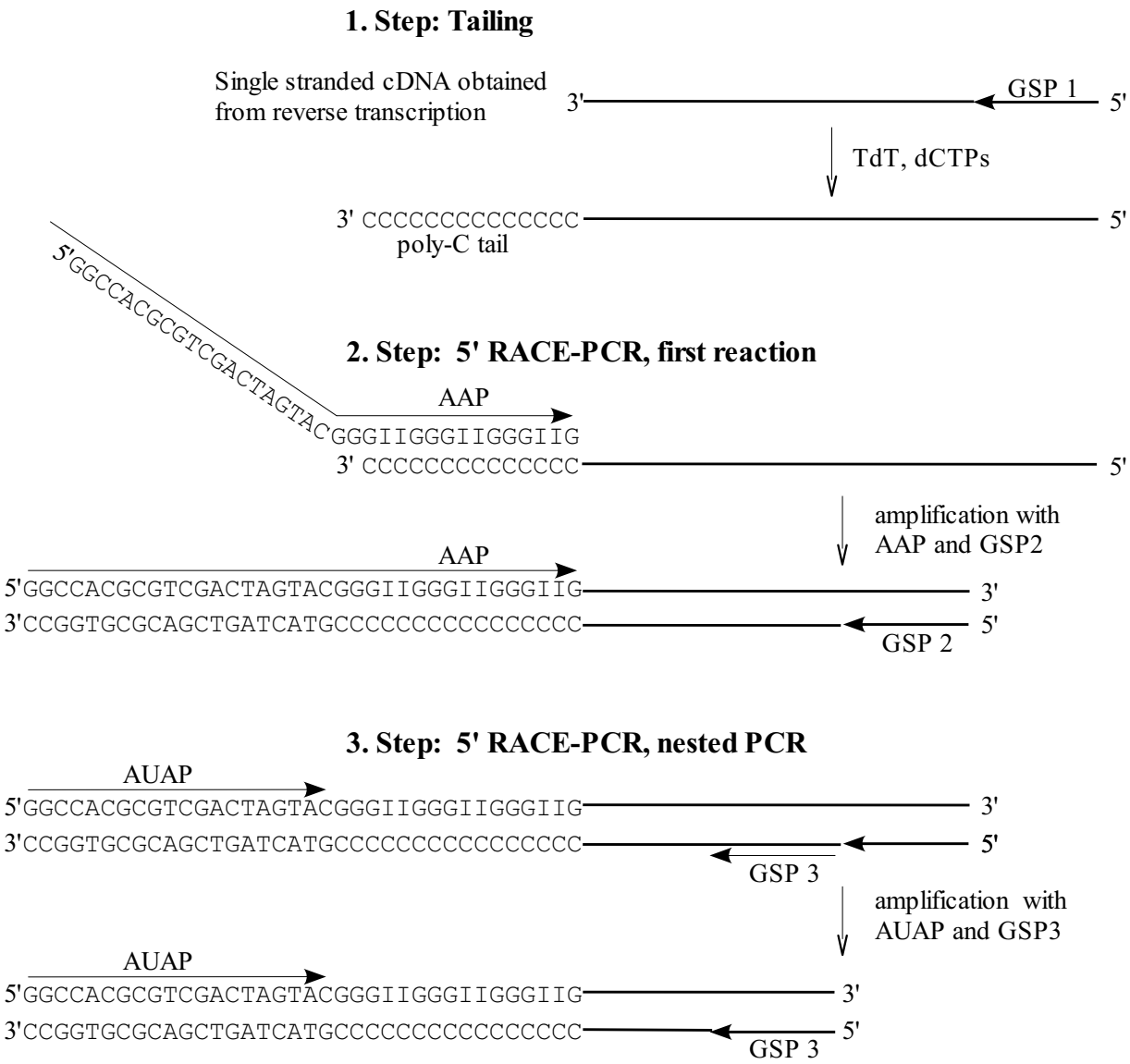


Figure 2.8: Procedures of 5'RACE technique. 1. step: tailing of the cDNA obtained from reverse transcription reaction with GSP 1 introduces a poly-C tail at the 3'end, 2. step: first PCR reaction with AAP and GSP 2, 3. step: second PCR reaction = nested PCR with AUAP and GSP 3 (AAP = abridged anchor primer, AUAP = abridged universal anchor primer, dCTPs = deoxycytosine triphosphates, GSP = gene specific primer, RACE = rapid amplification of cDNA ends, TdT = terminal deoxynucleotidyl transferase).

2.4.10 Cloning of DNA fragments

PCR products represent a pool of different DNA fragments (see section 2.4.9 on page 67). Cloning is used to separate and amplify these fragments. Separation is carried out by ligation into a vector. The resulting chimera of PCR- and vector-DNA (plasmid) is submitted to a host cell via transformation (see section 2.3.3.1 on page 43 or section 2.3.3.2 on page 44). Only one plasmid should reach one host cell. The host cell amplifies the plasmid and multiplies itself in creating a colony. In this section, cloning reactions are differentiated in respect to the used PCR products. In general, DOP-PCR DNA, 3'RACE-PCR DNA, and 5'RACE-PCR DNA were cloned either with the help of the TOPO TA cloning or the pGEM-T Easy cloning kit, full-length PCR-DNA was cloned in expression vectors. As an exception from this rule, full-length PCR-DNA was subcloned for maintenance in vector pCR 2.1-TOPO (TOPO TA cloning kit) or pGEM-T Easy as well.

2.4.10.1 Cloning of DOP-, 3'RACE-, and 5'RACE-PCR DNA

Two cloning kits were in use to clone DOP-, 3'RACE-, or 5'RACE-PCR DNA. TOPO TA Cloning kit and pGEM-T Easy cloning kit supplied already linearized vectors. Their 3'ends contained single deoxythymidine (dT) overhangs that allowed single deoxyadenosines at the 3'ends of PCR-DNA to ligate sticky ended with the vector (TA Cloning). Single 3'deoxyadenosine overhangs are only created if the *Taq* DNA polymerase has a nontemplate-dependent terminal transferase activity. Full-length PCR-DNA, that were produced by proof-reading DNA polymerases, had to be incubated for 10 minutes with 0.25 units *Taq* DNA polymerase at 72°C to add the single deoxyadenosine (dA) overhang if TA cloning was performed subsequently. Due to the fact that a single overhang is less stabile than a double strand and easily gets lost, all PCR products, that were stored longer than 24 h, were supplied with 3'single deoxyadenosine by incubation with *Taq* DNA polymerase. While with the TOPO TA Cloning kit, ligation was conducted with the help of a topoisomerase I activated vector (Shuman 1994), the pGEM-T Easy cloning kit needed the help of a T4 DNA ligase. The two kits supplied blue-white screening for selection of positive transformants (see section 2.4.11.1 on page 78).

TOPO TA cloning kit (Invitrogen)

DOP-, 3'RACE-, or 5'RACE-PCR DNA were ligated into vector pCR 2.1-TOPO. A typical ligation reaction was carried out as described below: PCR-DNA was separated in a low-melting agarose (Helena Bioscience) gel (see section 2.4.7 on page 63). The appropriate single discrete DNA-band was excised under low UV-emission arranged by using only one third of UV-lamps in the transilluminator. Since UV-radiation is known to cause DNA strand breaks, the gel was put on a glass plate that adsorbed UV-radiation partially during excission of the band. The gel was molten at 65°C for 3 - 5 minutes, transferred to 37°C to avoid re-solidification, and was immediately diluted 1:1 with water. An amount of 4 μl molten and diluted agarose gel containing the PCR-band was mixed

with 0.5 μl salt solution and 0.5 μl vector. The mixture was allowed to sit at 37°C while competent TOP 10 cells were taken from -80°C and thawed on ice. 25 μl of competent cells were transferred to the ligation reaction, stirred gently, and kept on ice for 30 minutes. The transformation procedure was continued by the heat shock reaction (as described in section 2.3.3.1 on page 43). For selection, 150 μl ampicillin [10 mg/ml] were spread on plates with 25 ml LB and 40 μl of X-gal [40 mg/1 ml DMF] were added to enable blue-white screening. In order to avoid overgrowing, only half of this mixture was spread onto one plate.

pGEM-T Easy cloning kit (Promega)

The ligation reaction was performed in accordance with the manufacturer's protocol: Briefly, PCR products were purified as described in section 2.4.5 (see page 63). 1.5 μl of purified PCR product was mixed with 2.5 μl 2x ligation buffer, 0.5 μl vector, and 0.5 μl T4 DNA ligase. The mixture was incubated for 1 hour at room temperature or over night at 4°C. 25 μl JM109 or TOP10 competent cells were added after thawing on ice. Transformation was carried out as described in section 2.3.3.1 on page 43. Selective LB plates were prepared with 150 μl ampicillin [10 mg/ml], 25 μl of X-gal [40 mg/1 ml DMF] and 20 μl 0.5 M IPTG to enable and to start blue-white screening.

2.4.10.2 Cloning in Expression Vectors

In contrast to the convenient cloning of PCR-DNA in already pre-cut vectors (see above 2.4.10.1 on page 75), expression vectors had to be cut and ligated manually. For more details on expression vectors which had been used in this work, see section 2.3.4.2 on page 46.

Restriction.

Linearization of expression vectors was carried out individually depending on the sequence and on the multiple cloning site (MCS). Unique restrictions sites were introduced into the full-length PCR-DNA with the help of expression primers (see section 2.4.9.4 on page 73). Full-length PCR-DNA and expression vector were digested in separate tubes for 3 hours and a "double digestion" was performed when the two restriction enzymes (one for 5' and one for 3'end) worked equally well in one buffering system and when each of them had the same temperature optimum. If one of these demands was not fulfilled, a sequential digestion was necessary. The first restriction reaction was carried out with the kind of buffer that contained lower salt concentrations followed by the second digestion with the higher salted buffer in just adding the appropriate amount of salt solution together with the second enzyme. If the buffering systems were not compatible at all, the PCR-DNA had to be purified after the first reaction with one of the kit-methods for purifying PCR products (see page 63). If the buffering system was the same but the optimal temperature was different, the restriction reaction was conducted first at lower temperatures with one enzyme, continued by adding the second enzyme, and transferring the mixture to the appropriate higher temperature. The reaction mixture was set up

according to the enzymes' description with purified PCR-DNA and expression vectors, respectively.

Ligation.

Digested vector and PCR-DNA (= insert) were purified with QIAEX II gel purification kit (see page 63) to remove buffer components, DNA, and vector fragments. The purification gel was loaded with an uncut vector as negative control that helped to make sure to cut out the correct bands for ligation. Uncut vectors migrate in several different bands in an agarose gel while completely linearized vectors form one distinct band. The different bands from uncut vectors are due to various topoisomeres, differing in their number of helical turns. The more turns a vector contains, the easier and faster it migrates through the net structure of the gel. The relaxed shape is the slowest while the supercoiled shape is the fastest topoisomere. The behavior of the vector is not predictable from its size.

Purified vector and insert were ligated with a molar vector-to-insert ratio of 1:10. Molarities of the two components were estimated with the help of an agarose gel. A typical ligation reaction did not exceed a total amount of 100 ng DNA. Compounds were mixed with the following amounts:

insert	x μl
vector	y μl
5 x ligation buffer (Invitrogen)	2.0 μl
T4 DNA-ligase (Invitrogen) [2 U/μl]	0.5 μl
H_2O	ad 10.0 μl

The reaction was mixed except for the ligase. DNA was denatured at 55°C for 2 minutes followed by 2 minutes on ice. Then, T4 DNA-ligase was added. T4 DNA-ligase from Fermentas was supplied with 10 x ligation buffer and a 50% PEG 4000 solution. For 10 μl reaction mixture, 1 μl of each ligation buffer and PEG solution was necessary.

Ligation conditions were tested individually. Three different temperatures and durations were applied: at 25°C for 1 hour, at room temperature over night, or at 16°C over night. The reaction was stopped by heating to 65°C for 15 minutes. This step breaks off the plasmid-ligase complex and increases transformation efficiency because the plasmid is much smaller without ligase on its back and easily enters a host cell. Transformation was carried out as described in section 2.3.3.1 on page 43.

In order to reduce the number of negative transformants, an additional restriction reaction was carried out after ligation previous to the enclosed transformation reaction. This additional restriction reaction called "re-cut" was aimed for destroying all incompletely digested vectors. While properly digested vectors had lost a part of their MCS, incompletely digested vectors still contained an intact MCS. The enzyme that conducted this additional restriction reaction was chosen to cut within the lost part of the MCS but not within the insert. Hence, correctly restricted vectors that resulted in correctly ligated plasmids were left untouched while incompletely digested vectors were linearized and subsequently unable to be transformed into host cells.

2.4.11 Selection of Positive Transformants

Successful ligation was checked with three different methods: DOP-, 3'RACE-, or 5'RACE-PCR plasmids were checked with blue/white screening and restriction analysis, full-length PCR plasmids were checked with restriction analysis and optionally, a PCR colony screening could be performed.

2.4.11.1 Blue-White Screening

DNA fragments ligated into vector pCR 2.1-TOPOor pGEM-T Easy (section 2.4.10) were submitted to blue-white screening. Together with an appropriate host strain like TOP10, JM109, DH5α, or XL1blue, the substrate X-Gal, and the inducer IPTG in case of JM109 or XL1blue hosts, positive colonies appeared white (or colourless) while the negative ones turned blue. For more information on formation of the blue colour, see section 2.3.1 on page 34.

2.4.11.2 PCR Colony Screening

with bacteria. Since blue/white screening was not possible with the expression vectors used in this work, PCR colony screening was applied instead to select positive clones. Colonies were picked using 10 μl pipette tips. Each of them was resuspended with 20 μl of sterile water. Half of this suspension (10 μl) was applied as template in the PCR, while the remaining suspension was used to grow a liquid, over-night culture for subsequent plasmid isolation and restriction analysis. When the insert of interest was integrated into the vector, amplification with primers, that annealed only to the insert, was successful. As a negative control, 10 μl of water were dropped on the plate where no colony had formed. These drops solved all untransformed DNA from the ligation reaction that may cause background in PCR reactions. The cells were lysed during the initial 5 minutes of denaturation of the PCR temperature programme. PCR conditions were the same as with successful full-length PCR (see section 2.4.9.4 on page 73).

with *Kluyveromyces lactis*. Successful integration of expression cassettes into the genome of *K. lactis* could be determined by either submitting the whole cell or the isolated chromosomal DNA (section 2.4.3.3 on page 61) to PCR screening. It was possible to either identify one copy or multi tandem copies. Primers were enclosed in the *K. lactis* protein expression kit (NEB). Integration primer 1, directed forwardly, annealed to the promoter of the hosts' *LAC4* locus, while integration primer 2 hybridized as reverse primer to the integrated expression cassette. Together, primers 1 and 2 could detect a one-fold integration resulting in a 1,900 bp diagnostic PCR product. With primers 2 and 3 the two- or more-fold integration was indicated by a 2,300 bp fragment. This fragment was copied from a *LAC4* promoter that was rebuild of two expression cassettes. Integration primer 3 annealed downstream of primer 2 within the first expression cassette, while primer 2 annealed in the second expression cassette upstream of primer 3.

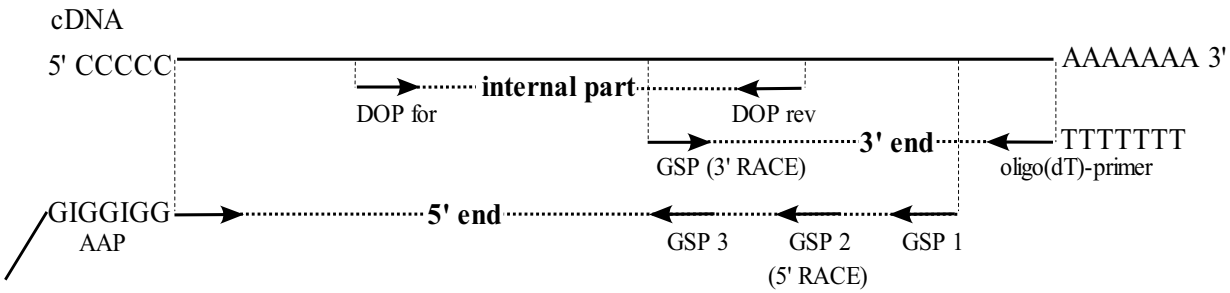


Figure 2.9: Localization of primers used during discovery of partial sequences. (AAP = abridged anchored primer, DOP for/DOP rev = degenerate oligonucleotide primer in forward/reverse direction, GSP = gene specific primer, RACE = rapid amplification of cDNA ends)

2.4.11.3 Restriction Analysis

The most reliable method, to select clones containing the insert of choice, was restriction analysis. 3 μl of isolated plasmid DNA (see section 2.4.3.2 on page 59) were mixed with 0.25 units of restriction enzyme (for restriction sites and reaction temperatures, see appendix Table B.5 on page 222). Buffer and additional components like bovine serum albumin (BSA) were added to a 1x final concentration as recommended by the manufacturer. Digestion was carried out for 1 hour at the advised temperature. Restriction patterns were visualized by agarose gel electrophoresis (see page 63).

2.4.12 Sequencing

Sequencing was carried out by MWG Biotech. Up to 1.5 μg of pure plasmids were isolated with a kit (see section 2.4.3.2 on page 60) and dried over night at room temperature. Drying at increased temperatures was avoided as overdried DNA was badly redissolved and caused failed sequencing.

2.4.13 Assembling the ORF

Partial sequences (internal fragments, 3' and 5'ends) with high similarities to AHAS from the database were joined together virtually. During discovery of the sequences, primers were placed to create overlaps of 100 to 150 bp between 5'ends and internal fragments on one hand and between internal fragments and 3'ends on the other hand (see Figure 2.9). In theory, these overlaps should display an identity of 100% but in practice overlap identities ranged from 96% to 100%. Final sequencing of the full-length clones was performed to prove which parts really form together an ORF.

2.4.14 Induction of Heterologous Protein Expression

Host cells harbouring the gene of interest can be forced to produce protein from this gene via transcription and translation. Starting this process is called induction. Induction and expression conditions differed in accordance with the used expression systems like bacteria (section 2.4.14.1), yeast (section 2.4.14.2), or without any host organism but with *E. coli* lysate (cell-free, section 2.4.14.3). Induction and expression conditions influence correct folding of the resulting proteins.

2.4.14.1 Inducing Bacterial Hosts

A single colony was picked up from a plate to inoculate liquid LB medium containing the appropriate antibiotic selection (see "Growing conditions for bacteria", page 42). This pre-culture was grown over night at 37°C shaking with 250 rpm. The amount of overnight culture that needed to be pelleted to obtain a starting OD_{600} of 0.2 was calculated with the following equation modified after Ausubel et al. (1994):

$$\frac{\text{starting OD}_{600} \cdot \text{expression culture volume [ml]}}{\text{pre-culture OD}_{600}} = \text{pre-culture to be pelleted [ml]}$$

Harvested and washed cells from the pre-culture were transferred to the appropriate medium with the chosen volume. The culture was grown at 37°C and 250 rpm until OD_{600} reached 0.4 up to 1.2. Protein expression was induced with 0.1 mM - 2.0 mM IPTG or with 200 ng/ml AHTC depending on the applied vector (see Table 2.6 on page 45). Expression culture was grown over night at a temperature between 25°C and 37°C and an aeration of 250 rpm. The culture was harvested by centrifugation at 4°C and 1,500 x g for 10 minutes. Protein extraction and purification was carried out immediately after harvesting to avoid any disturbing influence of storage (sections 2.5.1 and 2.5.2).

2.4.14.2 Inducing Fungal Hosts

S. cerevisiae: According to the pYES2 manual, cells were induced in transferring them from SC-U minimal selection medium to SC-U minimal induction medium (see section 2.3.2.2 on page 40). Induction medium contained galactose instead of glucose or raffinose compared to the SC-U minimal selection medium. Glucose represses every transcription while raffinose behaves indifferently. In galactose containing medium repression of the *GAL1* promoter is removed and transcription is allowed to start. Harvested cells were submitted to protein extraction (as described in section 2.5.1.2 on page 83) and extracted proteins were purified with the same methods as described for recombinant proteins from bacteria only depending on the kind of fused purification tag (section 2.5.2).

K. lactis: PCR determined positive transformants (section 2.4.11.2 on page 78) were induced by a change of medium. An area of approximately 2 mm^2 from the selecting minimal medium plate YCBA (section 2.3.2.3 on page 41) was scraped with a sterile pipette tip and cells were resuspended in 2 ml of YPGal medium as inducing agent. Cells

were grown at 30°C and 250 rpm. Within 2 to 4 days the culture reached a density of > 30 OD_{600} units/ml. As recombinant protein was secreted into the medium, cells were pelleted and discarded while the supernatant was concentrated by Ultra-filtration (as described in section 2.5.1.2 on page 83).

2.4.14.3 Starting a Cell-Free Expression

Protein synthesis reaction was started with the help of ExpresswayPlus Expression System (Invitrogen) according to the manufacturer's protocol. The reaction mixture contained 20 μl *E. coli* extract (IVPS Plus), 20 μl 2.5x reaction buffer, 1 μl T7 RNA polymerase, 1 μl 75 mM methionine, and 8 μl plasmid DNA containing a minimum amount of 1 μg. Plasmid DNA was isolated with the help of a minispin preparation kit (see section 2.4.3.2 on page 60) because the DNA had to be free of RNases. Preparations of plasmids isolated with a hand method contained RNase A which destroyed the ribosomes in the *E. coli* extract. The plasmid purification kit had to be a spin preparation type that did not use a gel-purification method. DNA solutions obtained from agarose gels significantly inhibited the protein synthesis reaction. In order to reach a concentration of 1 μg/8 μl for the isolated plasmid DNA, only sodium acetate was recommended for precipitation (see section 2.4.4). The expression reaction was conducted at 37°C in a thermomixer at 1,400 rpm for 4 hours and was stopped by adding 5 μl RNase A and incubating at 37°C for 5 minutes. The resulting product was submitted to Westerfeld enzyme assay with UV detection (see section 2.5.10.1) to check the AHAS-activity.

2.5 Biochemical Methods

In the following sections, all methods are described that were used to handle heterologously expressed proteins.

2.5.1 Preparation of Crude Protein Extracts

Target proteins were expressed *in vivo* in bacterial (*E. coli*) or eukaryotic (yeast: *S. cerevisiae* and *K. lactis*) hosts or *in vitro* in a cell-free system. The *in vivo* systems demanded disruption of the producing cells except for the medium-secreting *K. lactis* system while proteins from the *in vitro* cell-free system were already accessible for further investigation.

2.5.1.1 Crude Protein Extracts from Bacteria

Depending on the used vector, the expressed protein accumulated either in the cytoplasm or in the periplasm of its host. In each case, cells were harvested by moderate centrifugation with 8,000 x g at 4°C for 10 minutes in order to avoid an early release and decay of heterologously expressed proteins. The resulting cell pellet could be stored over night at 4°C when extraction and purification procedures were carried out the next day. Longer

termed storage at -80°C was not practised to avoid a putative loss of activity (Hill et al. 1997). Soluble proteins were separated by centrifugation at 4°C and 21,000 x g for 30 minutes. The supernatant contained the crude extract of soluble proteins and was submitted to purification (see section 2.5.2, page 83).

Cytoplasmic proteins from bacteria

For extraction of cytoplasmic proteins from bacteria, three different methods were in use: sonication only, lysozyme treatment followed by sonication, or treatment with CelLytic reagent. Cell lysis was randomly monitored by measuring the optical density at 590 nm applying the following equation (IBA BioTAGnology manual 2005, page 21):

$$\% \text{ lysis} = \left(\frac{1 \text{ - } \mathrm{OD}_{590} \text{ (treated culture)}}{\mathrm{OD}_{590} \text{ (untreated culture)}}\right) \cdot 100$$

Sonication. Cell pellets were resuspended in 15% volume of induced culture in resuspension buffer (RB) that corresponded to the intended purification method (for composition of the appropriate resuspension buffer, see Table 2.20 on page 86). The resuspended cell paste was transferred in a 50 ml, stand-alone Falcon tube with a conical bottom placed on ice water. Sonication was performed with a Branson Sonifier II W-250 using a standard-microtip with 3 mm in diameter. The instrument was adjusted to an output level of 20% - 50%, to an intensity level of 2 - 4, and to a duration of 5 minutes.

Lysozyme. Lysozyme (EC 3.2.1.17) hydrolyses the cross-linking muramic acid of bacterial cell walls. The resulting tiny little holes in the cell wall created tangential stress on the cell membrane that finally tears apart releasing the whole cytoplasm. Depending on the animal source, lysozyme has a size of approximately 14.5 kDa. Egg white lysozyme (Roth) was added in a concentration of 75-100 μg per ml of resuspended cell pellet. The mixture was gently shaken with 60 rpm at 25°C for 15 minutes. Afterwards, sonication was performed as described above (see page 82).

CelLytic. With the help of 10x CelLytic reagent (Sigma), lysis of bacterial hosts could be achieved without sonication. Pellet from 50 ml culture was mixed with 300 μl 10x CelLytic, 23 μl lysozyme solution [75 mg/ml], and 1 μl (= 350 U) benzonase (EC 3.1.30.2, Sigma) in a total volume of 1.5 ml of resuspension buffer (RB), (see Table 2.20 on page 86). The mixture was shaken with 100 rpm at room temperature for 30 minutes. Within this time, the mixture turned from milky to translucent. Without benzonase, which is a type of nuclease that selectively removes nucleic acids from proteins, the yield of soluble protein was only half of that compared to cell lysis with benzonase.

Periplasmic proteins from bacteria

Expressed from either pET22bmod or pASK IBA32 vectors, proteins could be fused to a N-terminal signal peptide (pelB or ompA) guiding the expressed protein to the periplasmic space. Located between the outer and inner cell membrane of *E. coli,* the periplasmic surrounding should enhance protein folding due to a simplified disulfide formation (Raina and Missiakas 1997, Sone et al. 1997). In order to simplify harvesting of the recombinant protein, only the periplasmic space was broken off while the inner membrane containing

the majority of the host derived proteins and nucleic acids remained intact. Induced cells were harvested as described earlier (see 2.4.14.1, page 80). The following preparation was carried out according to the osmotic shock protocol by Ausubel et al (1994). The pellet of 50 ml cultures was resuspended in 30 ml of 30 mM Tris/HCl pH 8, 20% (w/v) sucrose. Then, 60 μl 0.5 M EDTA pH 8 (final concentration of 1 mM) was added and stirred slowly for 10 minutes at room temperature. Cells were recollected by centrifugation at 10,000 x g at 4°C for 10 minutes and the supernatant was removed. Next, cells were suspended in 30 ml of ice-cold 5 mM $MgSO_4$ and stirred slowly for 10 minutes on ice. At this stage, the recombinant protein was released osmotically from the periplasm into the buffer. Cell debris and protoplasts were removed by centrifugation at 4°C for 10 minutes at 10,000 x g. The supernatant containing the periplasmic fraction was transferred to a clean tube and immediately submitted to the following purification procedures (section 2.5.2) and activity assay (section 2.5.10). Storage was avoided at this stage to avoid a decrease in activity.

2.5.1.2 Crude Protein Extracts from Yeast

S. cerevisiae accumulated expressed protein in the cytoplasm while *K. lactis* was capable of excretion.

Cell lysis for cytoplasmic proteins from S. cerevisiae

With the help of CelLyticY reagent (Sigma), *S. cerevisiae* cells INVSc1 were extracted. For 1 g of wet pellet, 2.5 ml reagent and 25 μl 1M DTT were added. Cells were resuspended and gently shaken for 30 minutes at room temperature.

Medium secreted proteins from *K. lactis*

The *K. lactis* yeast expression system using the vector pKLAC1 allowed secretion of the target protein into the medium. The protein was harvested from the cell-free medium, purified, and concentrated in one step with Ultra-filtration ProVario-3 (Filtron) as described below (see page 93)

2.5.2 Purification of Recombinant Proteins with Affinity Chromatography

Soluble recombinant protein fused to purification tags was purified under native conditions. Four different tags were in use and purification was carried out according to each purification kit manual. As recommended by Hill et al. 1997, all steps were performed at 4°C and under exclusion of daylight covering the probes with an aluminium foil as soon as the protein was released from its host in order to prevent a loss of activity. All buffers used for purification are listed in Table 2.20 at the end of this section on page 86.

2.5.2.1 His-Tag Purification with Ni-NTA Agarose

This type of immobilized-metal affinity chromatography (IMAC) applies a nickel-nitrilo triacetic acid (Ni-NTA) matrix (Qiagen) for purification of his-tagged recombinant pro-

teins obtainable from cloning into vectors pET3a, pET22b-23a, and pASK IBA32 (for information on the vectors, see section 2.3.4.2 on page 46). The His-tag is composed of six artificially introduced consecutive histidine residues (the so-called "6xHis-tag") on either N- or C-terminal side of the amino acid chain. Two histidine residues of the 6xHis-tag interact with one nickel ion that is anchored to the nitrilotriacetic acid matrix with its four remaining ligand binding sites. As a structural analog of histidine, imidazole has a higher affinity to nickel ions and displaces histidine during elution.

200 μl of Ni-NTA agarose were used for crude protein extracts from 50 ml cultures. Ni-NTA was equilibrated with 13.5 ml column buffer (CB, buffer compositions are listed at the end of this section in Table 2.20 on page 86). Spun down cells were resuspended in RB and soluble crude protein extracts were prepared as outlined in section 2.5.1 on page 81. Harvested protein was transferred to the matrix and was allowed to bind for 1 hour at 4°C under constant inversion with 0.1 rpm. In small scale purification, the charged matrix was collected by centrifugation for 2 minutes at 10,000 x g. Larger amounts of matrix-protein-slurry were transferred to an empty PD10 column or a syringe equipped with a frit at the outlet. Washing was performed 4 times with 1.5 ml washing buffer (WB). In order to get rid of highly concentrated NaCl, an additional washing step with 1.5 ml CB was performed. Recombinant protein was eluted in three to six fractions of 750 μl elution buffer (EB). All fractions were analyzed by SDS-PAGE (see page 88). Elution fractions harbouring the protein of interest were pooled, desalted, and transferred to the Westerfeld assay buffer (section 2.5.10) via PD10 (section 2.5.8.1) or spin filtration (section 2.5.8.2).

2.5.2.2 Intein-Tag Purification with Chitin Beads

As outlined in section 2.3.4.2 on page 46, expression of sequences cloned in vector pTYB1 results in a C-terminally fused Intein-tag. This Intein-tag contains an inducible self-cleavage activity that allows to remove the tag in the presence of thiols such as DTT resulting in a recombinant protein without any artificial amino acids. The Intein-tag itself is fused to a chitin binding domain (ChBD) which allows affinity purification on a chitin column.

Purification of recombinant proteins harvested from 500 ml cultures was carried out with 10 ml of chitin beads. The beads were equilibrated with 100 ml of column buffer (CB, buffer compositions are listed at the end of this section in Table 2.20 on page 86). Cells were spun down and resuspended in 7.5 ml CB. Crude protein extracts were prepared according to section 2.5.1 on page 81 with the help of lysozyme. As an exception from the general procedure with lysozyme, an amount of 0.01 g lysozyme per 1 g pellet was not exceeded since egg white lysozyme is known to bind and digest chitin. The mixture was shaken at 240 rpm for 15 minutes at 24°C and sonicated as outlined in section 2.5.1. The supernatant was obtained by centrifugation at 27,000 x g for 30 minutes. It was transferred to the equilibrated chitin beads and gently shaken at 4°C for 30 minutes to allow coupling to the matrix. The flow through was removed and the beads were washed with 100 ml CB. 30 ml of cleavage buffer (CLB) were quickly flushed through the beads to evenly distribute DTT. This first flow through was discarded. On-column cleavage was

then performed with 10 ml CLB at 4°C over night. The recombinant protein was eluted in 6 fractions each of 1 ml CB. Purification was monitored by SDS-PAGE from every fraction. To check cleavage efficiency, 200 μl of used chitin beads were boiled with 100 μl 2 x SDS sample buffer for 10 minutes at 95°C. The resulting supernatant was analyzed on SDS-PAGE for any proteins.

2.5.2.3 GST-Tag Purification with Glutathione Sepharose

With the help of vector pGEX-5x-3, recombinant proteins tagged with an N-terminal glutathione S-transferase (N-GST tag) were produced as described earlier in section 2.3.4.2 on page 47. This GST-tag was used for purification with glutathione sepharose as stationary phase and reduced glutathione as eluting agent. All buffers for FPLC-purification were prepared with bidestilled water. Solutions were filtered through 0.45 μm mesh and stirred under reduced pressure to remove all gaseous parts. Purification with FPLC was performed at 4°C with precooled buffers. Column buffer (CB, buffer compositions are listed at the end of this section in Table 2.20 on page 86) was used to resuspend the harvested cell paste as well as to equilibrate the FPLC column. Cells were extracted according to section 2.5.1 on page 81. The crude protein extract was transferred to the column to allow coupling to the resin. Washing was performed with CB followed by washing buffer (WB), both steps with a flow rate of 1.00 ml/min for 10 minutes. The protein of interest was eluted with elution buffer (EB) at a flow rate of 1.00 ml/min for 5 minutes. Eluted protein was detected with UV.

2.5.2.4 Strep-Tag Purification with Strep-Tactin Sepharose

As described previously in section 2.3.4.2 on page 47, vectors pASK IBA3C and pASK IBA7+ allowed to produce strepII-tagged recombinant proteins. While strep-tactin sepharose is binding strepII-tagged proteins selectively, desthiobiotin is the eluting agent showing higher affinity to the sepharose than strep-tag II. Cell pellet from 100 ml cell culture was resuspended in 10 ml washing buffer (WB, buffer compositions are listed at the end of this section in Table 2.20 on page 86) and extracted according to section 2.5.1 on page 81. For use with lysozyme or CelLytic reagent during extraction (see page 82), WB could be supplemented appropriately as long as the resulting pH was not lower than 7.5. 1 ml of strep-tactin sepharose (IBA) was equilibrated with 2 ml WB. The supernatant of protein extraction was submitted to equilibrated sepharose and was allowed to bind for 1 hour at room temperature. The matrix was washed 5 times with 1 ml WB and elution was performed in 6 fractions each of 500 μl elution buffer (EB). All fractions were checked with SDS-PAGE for pure recombinant protein.

applied buffers	purification method (p.)			
	His-tag p.	Intein-tag p.	GST-tag p.	Strep-tag p.
RB resus-pension buffer	50 mM NaH_2PO_4, 300 mM NaCl, 5 mM β-mercapto-ethanol (optional), 1% Triton X 100, pH 8	RB = CB see below	RB = CB see below	RB = WB see below
CB column buffer	50 mM NaH_2PO_4, 300 mM NaCl, 5 mM imidazole, pH 8	20 mM Tris/HCl pH8.5, 500 mM NaCl, 10 μM FAD	0.1 M K_2HPO_4 pH 7.5	CB = WB see below
WB washing buffer	50 mM NaH_2PO_4, 1 M NaCl, 25 mM imidazole, pH 8	-	1 x PBS: NaCl 8.0 g KCl 0.2 g Na_2HPO_4 1.4 g KH_2PO_4 0.2 g dissolved in H_2O adjusted to pH 7.2 filled up 1,000 ml	100 mM Tris/HCl pH 8.0, 150 mM NaCl, 1 mM EDTA
EB elution buffer	50 mM NaH_2PO_4, 300 mM NaCl, 200 mM imidazole, pH 8.5	-	50 mM Tris/HCl pH 8.0, 10 mM red glutathione	100 mM Tris/HCl pH8.0, 150 mM NaCl, 1 mM EDTA, 2.5 mM desthiobiotin
CLB cleavage buffer	-	20 mM Tris/HCl pH8.5, 500 mM NaCl, 10 μM FAD, 1 M DTT	-	-

Table 2.20: Overview of buffers for extraction and purification of crude protein extracts

2.5.3 Isolation, Solubilization and Renaturation of Inclusion Bodies

One of the major problems with heterologous protein expression in the bacterial host *Escherichia coli* is the occurrence of insoluble expressed protein that accumulates to small aggregates named "inclusion bodies". These inclusion bodies lack a correctly established secondary, tertiary, and quarternary structure. Refolding of inclusion bodies is based on the assumption that the 3-dimensional structure of a protein is a direct product of its primary structure leading towards a soluble and active enzyme because this soluble form contains the lowest amount of conformation energy. The procedure comprises two major steps:

1. Secondary, tertiary, and quarternary structures are destroyed under strong denaturing conditions (solubilizing buffer) while the primary structure is maintained.
2. Solubilized protein in solubilizing buffer is diluted and the denaturing buffer is replaced step by step with a buffer assisting renaturation (dialysis buffer).

The composition of these buffers has to be figured out individually for each type of protein to promote spontaneous correct folding. Buffer components are given in Table 2.21 on page 88.

The following procedure was taken from the manual of the inclusion body isolation kit from Novagen and modified according to the special requirements of the expressed catalytic subunit of AHAS. Harvested cells were resuspended in 0.1-fold culture volume of washing buffer (see Table 2.21) and sonicated as described above (see page 82). The pellet was washed three times with washing buffer. Solubilization was conducted with 1 ml 1 x solubilizing buffer (Table 2.21) per 10 mg wet pellet. The mixture was kept at room temperature for 15 to 45 minutes and inverted occasionally until the suspension turned from milky to opalescent. The remaining insoluble particles were removed by centrifugation at 21,000 x g and 4°C for 30 minutes. The supernatant containing the solubilized protein was submitted to renaturation. It was transferred to a dialysis tubing (Visking from Roth) and allowed to stay for 4 days in 25-fold culture volume of 1 x dialysis buffer (Table 2.21) at 4°C under permanent slow stirring.

2.5.4 Quantification of Proteins (Bradford)

The photometric assay (Bradford 1976) was used to determine the concentration of proteins. Protein concentrations were used in calculating the enzyme specific activity. To prepare the Bradford reagent, 100 mg Coomassie Brilliant Blue G-250 (Serva) were dissolved in 50 ml of 95% (v/v) ethanol before 100 ml of 85% (w/v) phosphoric acid were added. The resulting solution was made up to 1,000 ml with H_2O and filtered subsequently. 20 μl of sample or standard protein solution were thoroughly mixed with 1 ml Bradford reagent, incubated at room temperature for 10 minutes, and measured at 595 nm with UV/Visible spectrophotometer (Ultraspec 3100 pro, Amersham Pharmacia Biotech). On the dye binding to the protein, the solution turns blue. For setting the reference, 20 μl of H_2O were used instead of protein sample. The standard protein solution was prepared

washing buffer 20 mM Tris/HCl pH 7.5 10 mM EDTA-sodium salt 1% (v/v) Triton X-100	
10 x solubilizing buffer 1 M glycine pH 11 1 x solubilizing buffer freshly prepared before use diluting solution: 3% (w/v) n-laurosylsarcosine sodium salt and 1 mM DTT. 9 volumes of diluting solution were mixed with 1 volume of 10 x solubilizing buffer	50 x dialysis buffer 1 M Tris/HCl pH 8.5 1 mM red-glutathione, 1 mM ox-glutathione, 0.05 mM FAD, 0.5 mM $MgCl_2$ x 6 H_2O, and 0.5 mM ThDP. 1 x dialysis buffer 49 volumes of water were mixed with 1 volume of 50 x dialysis buffer

Table 2.21: Buffers for isolation, solubilization, and renaturation of inclusion bodies

from bovine serum albumin in concentrations of 0.1 to 1.0 mg/ml.

2.5.5 Sodium Dodecyl Sulfate - Polyacrylamide Gel Electrophoresis (SDS-PAGE)

Using SDS-PAGE, protein mixtures are denatured and separated solely in respect to their molecular weight. Disulfide bonds are cleaved with a thiol reagent (ß-mercaptoethanol) that dissociates the protein into its subunits. SDS-anions form micelles around the proteins, presenting the anions' negative charge on the surface of the resulting SDS-protein complexes. The negatively charged proteins separate by moving towards the anode. The smaller these proteins, the closer they move towards the anode. This method was established by Laemmli (1970). It is called "discontinuous buffered " because the sample has to cross two different types of gels: A non-restrictive large-pore gel, called stacking gel (pH 6.8) overlays a small-pore resolving gel (pH 8.8). The major advantage of a discontinuous buffer system is that relatively large volumes of dilute protein can be applied to the gel and resolution is increased compared to continuous systems. Diluted proteins concentrate to narrow starting zones during migration through the large-pore stacking gel which gives an optimal starting point for migration through the small-pore resolving gel.

Stacking and resolving gel were composed of :

components of SDS-PAGE gels	resolving gel 10% [ml]	resolving gel 12% [ml]	stacking gel 5% [ml]
H_2O	4.00	3.30	3.40
1.5 M Tris/HCl pH 8.8	2.50	2.50	-
0.5 M Tris/HCl pH 6.8	-	-	0.63
10% (w/v) SDS	0.10	0.10	0.05
30% acrylamide (Roth)[1]	3.30	4.00	0.83
TEMED	4 μl	4 μl	5 μl
10% (w/v) ammonium persulfate	0.10	0.10	0.05

[1]30% acrylamide: 30% (w/v) acrylamide plus 0.8% (w/v) bisacrylamide (37.5:1)

Samples were diluted 1:2 with 2x sample buffer and boiled for at least 5 minutes at 95°C. Depending on the size of wells, 15 to 20 μl of the boiled mixture were loaded onto 0.75 mm gels in a mini-protean II electrophoresis system (Bio-Rad). The 10x running buffer was diluted 1:10 and proteins were separated for 45 to 60 minutes at constant 200 V, starting 75 mA, and 15 W with a powersupply Multidrive XL (Pharmacia). 2x sample buffer and 10x SDS running buffer were composed of:

2x sample buffer	[ml]
0.5 M Tris/HCl pH 6.8	1.0
glycerol	2.0
10% (w/v) SDS	3.3
ß-mercaptoethanol	0.5
0.5% (w/v) bromophenol blue	0.5
H_2O	2.7

10x SDS running buffer	
Tris base	30 g
glycine	144 g
SDS	10 g
H_2O	ad 1,000 ml
pH 8.3 should be measured but not adjusted	

This method was used to check whether or not recombinant proteins were expressed and, if yes, whether they had approximately the right size. Moreover, it was applied to control successful protein purification.

2.5.5.1 Staining of SDS Gels

After electrophoresis, the stacking gel was removed and the resolving gel was transferred to the staining procedure. Two different staining methods were used.

Staining with Coomassie

The gel was kept for 60 minutes in a Coomassie blue staining solution in which proteins were fixed with 10% acetic acid and dyed blue. Destaining of protein-free zones within the gel was carried out in two steps: 30 minutes in destaining solution I was followed by a minimum of 60 minutes in destaining solution II which was used as storage solution as well.

Components of staining and destaining solutions were the following:

	Coomassie blue staining solution	destaining solution I	destaining solution II
Coomassie blue R-250 (Serva) 1% (w/v) in water	25 ml	-	-
methanol	100 ml	100 ml	10 ml
acetic acid, glacial	20 ml	20 ml	14 ml
H_2O	ad 200 ml	ad 200 ml	ad 200 ml

Alternatively, destaining could be performed in a much quicker procedure with heated water: After staining for 60 minutes, the gel was transferred to water and heated for 1 minute at 600 W in a microwave oven. Two further cycles of heating were carried out with fresh water. The results were of lower contrast but saved almost half an hour. This method allows to detect minimal amounts of 0.1 μg protein/band.

Staining with silver nitrate

Less than 0.001 μg protein/band can be detected with this method as detailed by Heukeshoven and Dernick (1988a and 1988b) and modified by Ober (1997). Proteins were precipitated in the gel with trichloroacetic acid fixing solution for 45 minutes on a shaking table. The fixing solution was rinsed off with 50% ethanol. The gel was soaked with sodium thiosulfate impregnation solution for 1 minute. Transferred to a silver nitrate staining solution for 40 minutes, proteins in the gel bound silver ions which were reduced to atomic silver germs. The process was enhanced by the previous impregnation step. The staining solution was washed away with water.

The gel was transferred to the developing solution that reduced all silver ions to silver metal. This process is much faster in the presence of the silver germs sitting on the proteins than everywhere else in the gel. Protein bands turned dark brown to black. In order to obtain sufficient contrast, the developing reaction was stopped in time by rinsing with water twice and by shifting to acidic pH (7% acetic acid). Colour developing was carried out on ice in order to slow down the developing process to give time to choose the right moment when to stop it. When the right time for stopping was missed, a destaining solution could be used and the procedure described above could be repeated. The readily stained gel was stored in 30% methanol.

For components of fixing, impregnation, staining, developing, and destaining solution,

please see the Tables below:

fixing solution	
trichloroacetic acid	40 g
ethanol	100 ml
H_2O	ad 200 ml

staining solution	
silver nitrate	100 mg
37% formaldehyde	33 μl
H_2O	ad 50 ml

destaining solution	
potassium hexacyanoferrate (III)	1 g
sodium thiosulfate	1.6 g
H_2O	ad 100 ml

impregnation solution	
37% formaldehyde	100 μl
43% (w/v) sodium thiosulfate	75 μl
H_2O	ad 150 ml

developing solution	
sodium carbonate (waterfree)	15 g
37% formaldehyde	250 μl
43% (w/v) sodium thiosulfate	5 μl
H_2O	ad 250 ml

2.5.6 Protein Precipitation

In order to obtain detectable amounts of protein for SDS-PAGE (page 88), diluted samples were precipitated.

2.5.6.1 TCA Precipitation

The TCA precipitation method was modified after Bennet (1967). The sample was mixed with 1/10 sample volume of 0.1% (w/v) sodium deoxycholate (DOC) in 0.02% (w/v) NaN_3 and 1/10 sample volume of 55% (w/v) trichloroacetic acid (TCA). After vigorously vortexing, the sample was allowed to precipitate over night at 4°C and was harvested at 15,000 x g for 30 minutes. The pellet was resolved in equal volumes of 1.5 M Tris/HCl pH 8.8 and 2x SDS sample buffer to ensure an alkaline pH. If buffering was omitted, the SDS sample buffer turned orange (bromophenol blue changes colours from orange to blue at pH range from 3 to 4.7) and separation on SDS-PAGE was impossible due to uncharged proteins that are not moving in direct current. Negatively charged micelles from SDS-anions around proteins are neutralized with positively charged hydrogen cations from acids like TCA and result in uncharged proteins.

2.5.6.2 Acetone Precipitation

The sample was mixed with 4 sample volumes of acetone and transferred to -20°C. After 1 hour the supernatant was removed after centrifugation at 15,000 x g for 30 minutes, and the pellet was analyzed by SDS-PAGE.

2.5.7 Protein Sequencing

Protein sequencing was kindly performed by Dr. M. Nimtz, Departement of Structural Biology, Helmholtz-Zentrum für Infektionsforschung, former GBF, Braunschweig. Sequencing was conducted with heterologously expressed protein. Before sequencing, the protein was purified by His-tag affinity chromatography, separated on SDS-PAGE, and finally cut from the SDS-PAGE gel. The gel fragment was washed 5 times with sterile water and dried at room temperature for 4 days. After tryptic digestion of the dried protein, two types of sequencing were carried out: MALDI-MS/MS and ESI-MS/MS. Homologies were searched with FastA programme, Uniprot database (for www-address, see page 56).

2.5.8 Buffer Exchange, Desalting, and Concentration of Protein Samples

Enzyme activity had to be determined in an assay buffer differing from the elution buffers used for purification of recombinant proteins (see section 2.5.2 on page 83). The assay buffer was composed of:

Assay buffer
0.1 M K_2HPO_4
0.5 mM $MgCl_2 \cdot 6\ H_2O$
0.5 mM ThDP
10 μM FAD
solubilized in water and pH adjusted to pH 7.0

2.5.8.1 PD10 Desalting and Buffer Exchange

Prepacked PD10 columns (Pharmacia) contain Sephadex G-25 covered with a frit. The column was equilibrated with 25 ml of assay buffer. A sample volume of 2.5 ml was allowed to enter the gel bed and was eluted with 3.5 ml assay buffer. Smaller samples, e.g., 0.5 ml, again first had to enter the gel bed completely, then 2 ml assay buffer were allowed to enter the gel bed, and elution was carried out with 1 ml of assay buffer. The column was regenerated with 15 ml 0.2 N NaOH followed by 30 ml 0.5 N NaCl. For storage the columns were flushed with 0.05% NaN_3 to prevent microbial growth. Columns treated like this could be reused when stored at 4°C.

2.5.8.2 Spin-Filtration

Samples of 6 ml in volume or smaller could be concentrated to 60 μl in a centrifuge at 8,000 x g within 15 minutes at 4°C. The size exclusion limit with concentrators like Vivaspin 6 (Satorius) was 30 kDa.

2.5.8.3 Ultra-Filtration with ProVario-3

Large samples ($\geq$ 50 ml) were concentrated and re-buffered using a tangential flow filter (Filtron) excluding all proteins smaller than 30 kDa. At a pumping rate of 250 ml/min and with a counterpressure of 0.5 bar, proteins were washed with 4 sample volumes of assay buffer and concentrated down to 30 ml. The filter was rinsed intensely with H_2O and 0.5 N NaCl and stored in 0.05% (w/v) NaN_3.

2.5.9 Enzymatic Synthesis of ^{14}C-2-oxoisovaleric Acid

As described by Weber (1997), ^{14}C-labelled 2-oxoisovaleric acid was synthesized with the help of L-leucine dehydrogenase from *Bacillus cereus* (Calbiochem). 20 μl of 0.01 μCi/μl ^{14}C-labelled valine were mixed with 10 μl of 200 mM NAD^+, 10 μl of 55.3 U/ml L-leucine dehydrogenase, and 60 μl of 0.2 M glycine/KCl/KOH-buffer pH 10.7 and incubated at 37°C for 30 minutes. Reaction was stopped with 25 μl 5N HCl. The mixture was diluted to 600 μl with 1M HCl and 2-oxoisovaleric acid was extracted 4 times with ethyl acetate. The organic solvent was evaporated at room temperature under a constant flow of synthetic air. The yield of the synthesis was in a range of 20% to 38%.

2.5.9.1 Rapid Intelligence TLC Analyzer (RITA)

Thin layer chromatography (TLC) with ^{14}C-labelled substrates were analyzed with the Rapid Intelligence TLC Analyzer (RITA) from Raytest. The analyzer contained a Geiger-Müller counter that was run with argon-methane (90:10) from Westfalen. The 100%-method was applied for quantifying the peaks.

2.5.10 Enzyme Assays

In order to characterize heterologously expressed catalytic subunits of acetohydroxyacid synthase (AHAScsu) by enzyme activity, the following detection methods were applied routinely. An appropriate amount of purified (see section 2.5.2) and assay-buffered enzyme was diluted 1:10 with reaction cocktail 1 or with reaction cocktail 2 (see Table 2.22). In contrast to the previously mentioned assay buffer (see section 2.5.8 on page 92), the reaction cocktails contained pyruvate which is a substrate for AHAScsu (for details about the AHAScsu enzyme reaction see Figure 2.10 on page 94). Consequently, the reaction cocktails started the enzyme reactions which were carried out at 37°C for 30 minutes. Reactions performed with reaction cocktail 1 were stopped with 0.1 volumes of 6N H_2SO_4 when the reaction product actolactate was detected with UV (see section 2.5.10.1). When reaction products and educts were analyzed with GC-MS (section 2.5.10.2), reactions were carried out with reaction cocktail 1 or 2, but had to be stopped with 0.1 volumes of concentrated HCl.

reaction cocktail 1	reaction cocktail 2
50 mM K_2HPO_4 pH 7.9	50 mM K_2HPO_4 pH 7.9
1 mM ThDP	1 mM ThDP
10 mM $MgCl_2$	10 mM $MgCl_2$
10 μM FAD	10 μM FAD
40 mM pyruvate	10 mM pyruvate
	30 mM 2-oxoisovaleric acid

Table 2.22: Components of reaction cocktails 1 and 2 for AHAScsu activity assays

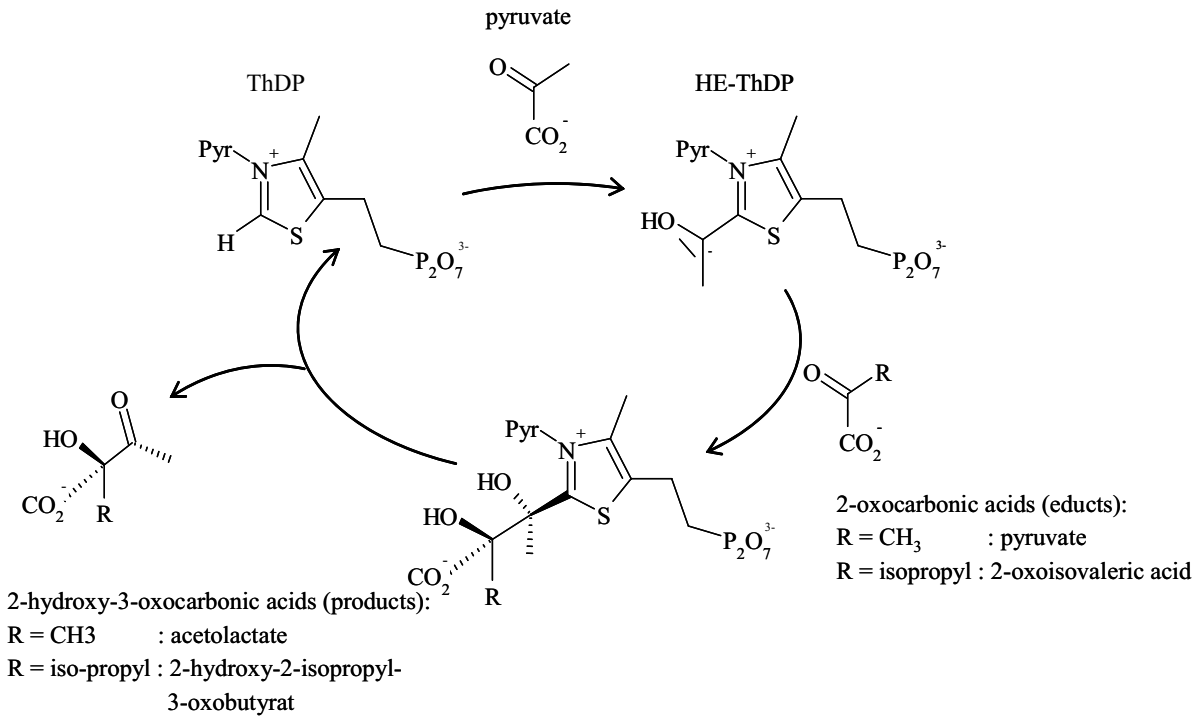


Figure 2.10: Carboligation reaction of a hydroxyethyl group fused to 2-oxocarbonic acids: the catalytic cycle of the acetohydroxyacid synthase (AHAS) reaction (modified from Tittmann et al. 2003 and McCourt and Duggleby 2006). 1. Pyruvate forms the active acetaldehyde (He-ThDP) together with thiamin diphosphate (ThDP). 2. He-ThDP attacks a 2-oxocarbonic acid creating a new C–C-bond. 3. A 2-hydroxy-3-oxocarbonic acid is released and ThDP is recycled.

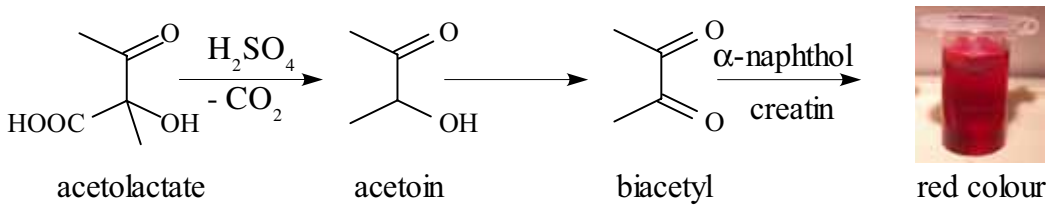


Figure 2.11: Colour formation for detection with UV. Acetolactate is decarboxylated to acetoin and oxidized to biacetyl. Together with α-naphthol and creatine, a red colour of unknown structure develops in a Sakagucchi-like reaction.

2.5.10.1 Detection with UV (Westerfeld assay)

This method was used to detect acetolactate which is the product of the first step in valine and leucine biosynthesis in primary metabolism (Figure 1.9). It was developed by Westerfeld (1945) and modified by Duggleby and Pang (2000). The stopped enzyme reaction was incubated at 60°C for 15 minutes to decarboxylate the acetolactic acid (= acetolactate in the presence of 6N H_2SO_4) to acetoin and subsequently to oxidize acetoin to biacetyl (see Figure 2.11 on page 95). For colour formation, 500 ml of this stopped and heated reaction solution were mixed with 500 ml of 0.5% creatine solution and 500 ml of freshly prepared α-naphthol solution (5% in 4M NaOH). This mixture was incubated for 15 minutes at 60°C under occasional inverting. When the test was positive, the red colour formed within the first 5 minutes. The colour was allowed to develop for further 15 minutes at room temperature and was read at 525 nm (ε_M = 16.164 M^{-1} cm^{-1}).

2.5.10.2 Detection with GC-MS

GC-MS measurements were kindly carried out at the institute by Dr. T. Beuerle, TU Braunschweig. Settings were designed to detect substrates and products from primary (valine and leucine biosynthesis) and secondary metabolism (postulated necic acid biosynthesis, see section "State of knowledge on necic acids biosynthesis" on page 18). The enzyme reaction was stopped with HCl because it was inert to methylation with diazomethane. The acidification not only inactivated the enzyme but also transferred all organic salts to their corresponding acids. Now, these acids, representing either substrates or products of AHAS, were extracted four times with ethyl acetate. The pooled organic phases were dried over Na_2SO_4, concentrated under a constant flow of synthetic air, and methylated with freshly synthesized diazomethane-diethyl ether-solution (diazomethane generator is shown in Figure 2.12) until the yellow colour persisted in the resulting solution. The mixture was concentrated under nitrogen flow to 1 mg/ml, separated via GC (injection: 1/25 splitless, column: DB5-30m, temperature programme: 40°C for 3 min; heating rate 5°C/min up to 100°C) and detected with MS. In addition to the samples, two negative controls were treated in the same way. Negative controls were composed of the same

solution A	solution B
6.25 g *N*-methyl-*N*-nitroso-p-toluenesulfonamide in 18.75 ml ethylene glycol were mixed with diethyl ether 1:1	5 g KOH in 50 ml methanol were mixed with H_2O 1:1

Table 2.23: Components of solutions A and B for preparation of diazomethane

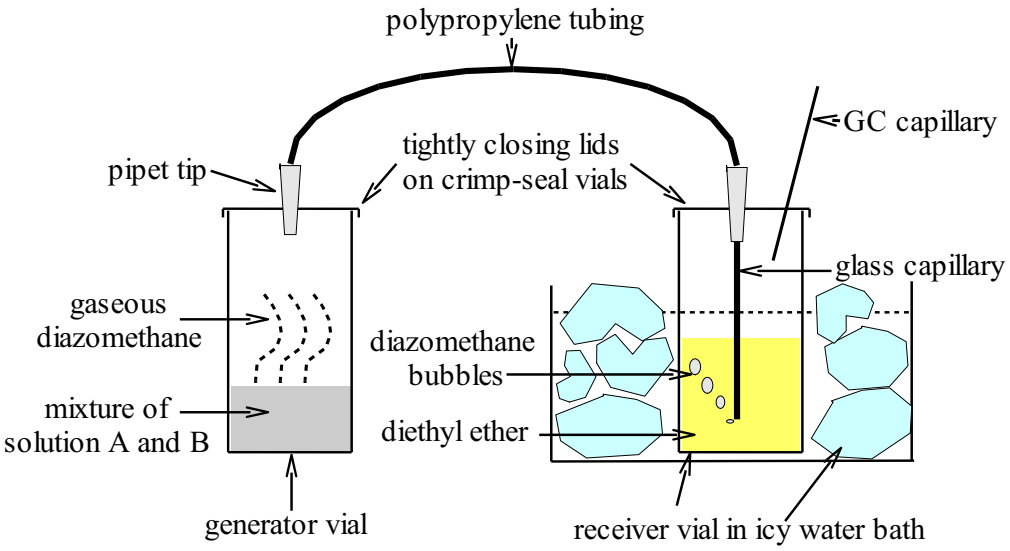


Figure 2.12: Diazomethane generator: Gaseous diazomethane is produced in the generator vial by the reaction of solution A with solution B, it is transferred and solubilized into diethyl ether in the receiver vial.

components as the samples except for the enzyme: One control contained water instead of enzyme while the second control contained heat inactivated enzyme.

The diazomethane generator

Two crimp-seal vials with tightly closing lids were connected with a polypropylene tubing (see Figure 2.12 on page 96). The vial that should receive the gaseous diazomethane was equipped with a GC capillary for equalizing pressure. The receiver vial was filled up to 75% of its volume with diethyl ether without submerging the GC capillary. When the diethyl ether was sufficiently pre-cooled in an icy water bath, the reaction was started in the generator vial. 750 μl of solution A and 750 μl of solution B (see Table 2.23) were placed in this order in the generator vial. Developing gaseous diazomethane was detected in the receiver vial by the formation of bubbles in diethyl ether which turned yellow by solubilized diazomethane. The reaction was finished when the bubbling stopped. Remaining reactants in the generator vial were destroyed with acetic acid.

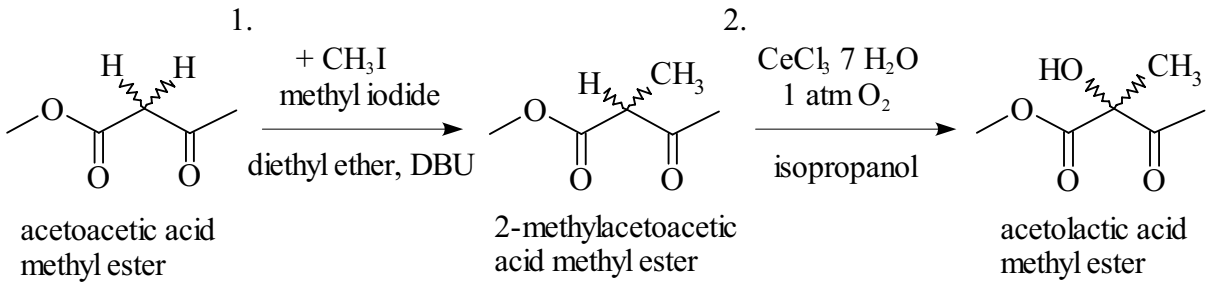


Figure 2.13: Synthesis of methylated acetolactic acid in two steps: 1. methyl iodide is added to acetoacetic acid methyl ester resulting in 2-methylacetoacetic acid methyl ester. 2. oxygen is inserted via Cer(III)-catalysis resulting in acetolactic acid methyl ester

2.6 Methods of Organic Chemistry

For unequivocal identification of predicted products from recombinant AHAS reaction, references were synthesized and purified for both primary and secondary metabolic pathways. This was inevitable due to commercial unavailability of these standards. All standards were stabilized as methyl esters.

2.6.1 Synthesis of Acetolactic Acid Methyl Ester As a Reference for Products from Primary Metabolism

Acetolactic acid methyl ester was used as a stabilized reference for AHAS reaction in primary metabolism. AHAS catalyzes the first step in branched chain amino acid biosynthesis where two molecules of pyruvate are combined to acetolactic acid (Figure 1.7). The synthesis of acetolactic acid methyl ester was conducted in a two step reaction (see Figure 2.13) according to Christoffers et al. (2003).

First step.

24.9 mmol acetoacetic acid methyl ester and 24.9 mmol DBU (1,8-diazabicyclo[5.4.0] undec-7-ene) were dissolved separately in 17.5 ml diethyl ether. The solvated components were combined under constant stirring. 30.4 mmol methyl iodide was added drop by drop and the mixture was evacuated and floated with gaseous nitrogen. While stirring over night at room temperature, a white and voluminous precipitate developed. It was resolved in 20 ml H_2O, the organic phase was separated, and kept for further manipulations. The aqueous phase was extracted three times with 15 ml dichloromethane. The pooled organic phases (diethyl ether and dichloromethane and solvents) were dried over $MgSO_4$, filtered and concentrated under reduced pressure (300 mbar/40°C) to approximately 15 ml. The remaining clear and oily liquid was purified by column chromatography (column of 30 cm length and 5 cm effective diameter) with $SiO_2$60 as stationary phase and petroleum ether/diethyl ether (2:1) as mobile phase. Fractions of 8 ml were collected and checked for the product 2-methylacetoacetic methyl ester.

The product 2-methylacetoacetic acid methyl ester first left the column in fractions number 14 to 18 and showed a R*f* of 0.30 in the enclosed TLC-checks (silica TLC F_{254}, developed with petroleum ether/diethyl ether (2:1), and detected with UV light at 254 nm). The educt acetoacetic acid methyl ester left second in fractions 21 to 33 with a R*f* of 0.23. Fractions containing the product were pooled, concentrated under reduced pressure, and stored at 4°C.

Second step.

1.62 mmol 2-methylacetoacetic methyl ester, product of the first step, were mixed with 0.081 mmol cerium(III)chloride heptahydrate and solved in 0.5 ml isopropanol. The yellow salt crystals of cerium(III)chloride heptahydrate dissolved poorly. The flask was evacuated up to 400 mbar and floated with oxygen. A constant oxygen stream of 50 cm^3/h was set up. After approximately one hour, the preparation turned milky-purple and the salt crystals were dissolved. The mixture was stirred over night at room temperature. The solvent was evaporated under reduced pressure and the remaining mixture of products and educts was separated by column chromatography with SiO_2 60 as stationary phase. Column and fraction size were the same as described in the first step (page 97). As mobile phases, two different compositions of petroleum ether/ethyl acetate were used sequentially. The first mobile phase was petroleum ether/ethyl acetate (5:1) which was used for fractions 1 to 25. The educt 2-methylacetoacetic methyl ester left the column in fractions 15 to 25. The second mobile phase was used for fractions 26 to 41 and was composed of petroleum ether/ethyl acetate (2:1). The reaction product acetolactic acid methyl ester left the column in fractions 26 to 41. For identification of reaction educts and products, each fraction was spotted on TLC, developed with petroleum ether/ethyl acetate (2:1), and detected with both UV light at 254 nm and phosphomolybdic acid reagent.

phosphomolybdic acid reagent	
phosphomolybdic acid	4 g
concentrated H_2SO_4	8 ml
glacial acetic acid	ad 100 ml
The mixture was stirred over night before use. Mode of employ: The reagent was sprayed on the developed and dried TLC plate and shortly heated to approximately 200°C until blue spots developed.	

The educt 2-methylacetoacetic acid methyl ester showed a R*f* of 0.37 detected with UV light while the product acetolactic acid methylester was detected with phosphomolybdic acid reagent and showed a R*f* of 0.25. Fractions containing the product acetolactic acid methyl ester were pooled, concentrated under reduced pressure, and stored at 4°C until used as a reference for GC-MS detected enzyme assay (see section 2.5.10.2 on page 95).

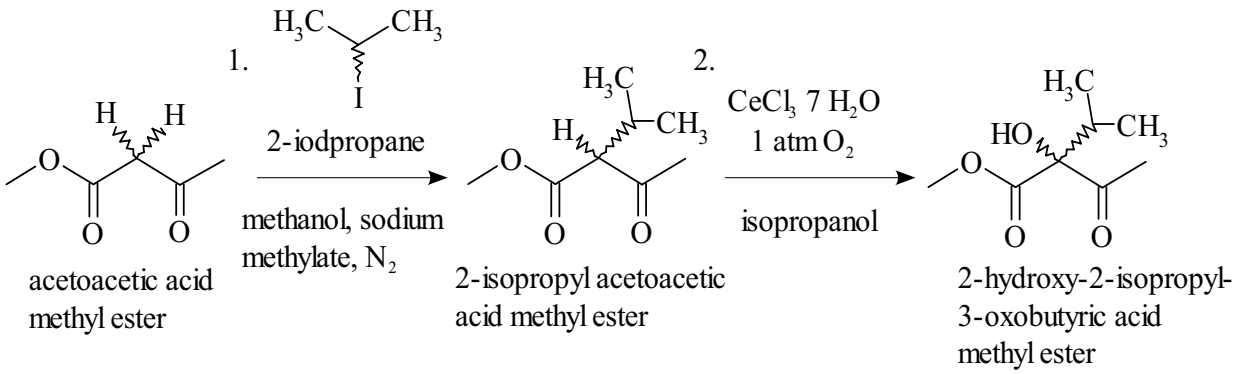


Figure 2.14: Synthesis of methylated pronecic acid in two steps: 1. 2-iodpropane is added to acetoacetic acid methyl ester resulting in 2-isopropylacetoacetic acid methyl ester. 2. oxygen is inserted via Cer(III)- catalysis resulting in 2-hydroxy- 2-isopropyl-3-oxobutyric acid methyl ester, the methylated pronecic acid

2.6.2 Synthesis of 2-hydroxy-2-isopropyl-3-oxobutyric Acid Methyl Ester As a Reference for Postulated Products from Secondary Metabolism

The synthesis of 2-hydroxy-2-isopropyl-3-oxobutyric acid methyl ester was carried out in the same manner as described above for acetolactic acid methyl ester with slight modifications in the first step (see Figure 2.14 on page 99).

First step.

8.6 mmol acetoacetic acid methyl ester were dissolved in 6 ml waterfree methanol. 8.6 mmol sodium methylate were added cautiously under constant stirring and N_2-flow. 12.9 mmol 2-iodopropane were added in the same way. The reaction chamber was composed of a 50 ml round bottomed flask and a reflux condenser. The system was closed, evacuated, and floated with N_2. The reaction mixture was stirred over night at 80°C. The mixture was concentrated under reduced pressure (300 mbar/40°C). The remaining liquid contained precipitated yellow sodium iodide. Water was added until the salt was resolved. The aqueous phase was extracted three times with diethyl ether. The pooled organic phases were dried over $MgSO_4$, filtered and concentrated under reduced pressure (150 mbar/40°C). Purification was achieved with the help of column chromatography with SiO_2 60 as stationary phase. This time petroleum ether/ethyl acetate (5:1) was used as mobile phase. The product 2-isopropyl acetoacetic acid methyl ester left the column first in fractions 10 to 13 while the educt acetoacetic acid methyl ester left the column in fractions 14 to 18. For identification of reaction educts and products, each fraction was spotted on TLC, developed with petroleum ether/ethyl acetate (2:1), and detected with

UV light at 254 nm. The product 2-isopropyl acetoacetic acid methyl ester showed a R*f* of 0.44 and the educt acetoacetic acid methyl ester showed a R*f* of 0.3. Product fractions were pooled, concentrated, and stored at 4°C.

Second step.

The second step of the synthesis of 2-hydroxy-2-isopropyl-3-oxobutyric acid methyl ester was carried out according to the second step of the synthesis for acetolactic acid methyl ester. 1.59 mmol 2-isopropyl acetoacetic acid methyl ester (product from the first step) were mixed with 0.081 mmol cerium(III) chloride heptahydrate and solved in 0.5 ml isopropanol. The flask was evacuated up to 400 mbar and floated with oxygen. A constant oxygen stream of 50 cm^3/h was set up. The preparation turned milky-orange instead of milky-purple. The working up was performed by column chromatography with SiO_2 60 as stationary phase. This time, three sequentially used mobile phases assured a proper separation: Petroleum ether/ethyl acetate (10:1) was used for fractions 1 to 23, petroleum ether/ethyl acetate (5:1) was used for fractions 24 to 38, and petroleum ether/ethyl acetate (2:1) was used for fractions 39 to 48. The product 2-hydroxy-2-isopropyl-3-oxobutyric acid methyl ester left the column in fractions 22 to 36 and showed a R*f* of 0.41 in the TLC-screening with petroleum ether/ethyl acetate (2:1) and detection with phosphomolybdic acid reagent. Fractions containing the product were pooled, concentrated under reduced pressure, and stored at 4°C until used as a reference for GC-MS detected enzyme assay (see section 2.5.10.2 on page 95).

3 Results

In the following, the identification of several cDNAs, coding for putative AHASs, is described (section 3.1) and these cDNA sequences are further characterized in section 3.2. The results of various efforts in heterologous expression are shown in section 3.4 that, finally, resulted in a first biochemical characterization of the recombinant enzyme.

3.1 cDNA Identification Coding for Putative AHAS

The search for cDNAs coding for putative AHAS was performed on three PA-containing plants from different families: *Eupatorium cannabinum* (Asteraceae, tribe Eupatorieae), *Symphytum officinale* (Boraginaceae), and *Parsonsia laevigata* (Apocynaceae). All three species belong to not closely related families but nevertheless, they all produce the same type of PAs that contains a very unique necic acid. This scattered occurrence of such a unique compound in three non-related taxa can be due to biochemical convergence (polyphyletic origin) or to a monophyletic origin of the specific enzyme with a partial loss in some taxa during evolution. As the basis for phylogenetic analysis is formed by sequence data, cDNAs were identified.

Former studies feeding ^{13}C-labelled glucose to PA-producing *Eupatorium clematideum* had shown that the postulated PA-specific enzyme performs an AHAS-like reaction utilizing pyruvate as an AHAS-typical substrate (Weber et al. 1999). Consequently, studies to further characterize the putative PA-specific enzyme had been conducted with native AHAS from *Eupatorium clematideum*. As described for many other isolation trials of plant AHAS (Durner and Böger 1988, Muhitch 1988, Southan and Copeland 1996), purification had been hampered by the labile nature of the enzyme: Within 24 hours the ammonium sulfate precipitated, resolved, and desalted enzyme had lost 80% activity. The most disturbing fact had been that a co-purified pyruvate decarboxylase (PDC) activity had not been removable from the AHAS activity. Unfortunately, PDC activity is able to mimic AHAS activity if the assay is based on acetoin formation as described for the Westerfeld assay (section 2.5.10.1 on page 95). Hence, the strategy in this work was to start a new approach of molecular biology with sufficient amounts of heterologously expressed plant AHAS that should exclude the PDC activity and provide an increased stability for extended biochemical characterization.

The AHAS is known to perform the first step in branched chain amino acid biosynthesis in plants (Umbarger and Brown 1958) and is classified as an enzyme from primary metabolism. In this work, an AHAS with either enlarged or altered substrate specificity was expected to function in PA biosynthesis. In case of an altered substrate specificity, the enzyme would be designated necic acid synthase (NAS). Due to its exclusive activity in secondary metabolism (PA biosynthesis), the NAS was believed to be very rare. The methodical answer to this problem was to pre-select PA-producing plant tissues (root-tissue

primer	direction	sequence	degen-eracy	T_m[°C]
P01	forward	5'-dTTY CAR GAR ACN CCN RTH GT-3'	768 x	55.9
P02	forward	5'-dATG YTN GGN ATG CAY GGN AC-3'	256 x	58.3
P03	reverse	5'-dYTG NGC NGC CCA CAT YTG RTG YTG-3'	256 x	66.1
P04	reverse	5'-dACC ATN CCN ARR TGY TGR TTR TT-3'	512 x	58.0

For IUPAC Ambiguity Code, see Table B.3 on page 221

Table 3.1: Degenerate primers for identification of AHAScsu: sequences, degree of degeneracy, and predicted melting temperature. Sequences are written in IUPAC Ambiguity Code

cultures) and to find as many different AHAS as possible.

Once the specific necic acid synthase will be identified and discovered from the three lycopsamine type PA containing families, the phylogenetic analysis should reveal a mono- or polyphyletic origin of the NAS. In combination with the findings for the first necine base specific enzyme, the homospermidine synthase (HSS), that up to now is likely to be invented in evolution several times independently (Reimann et al. 2004), the information about the first necic acid specific enzyme will contribute to the knowledge about the evolution of pyrrolizidine alkaloid biosynthesis.

3.1.1 AHAS Catalytic and Regulatory Subunits of *Eupatorium cannabinum*

The AHAS is known to consist out of two different types of subunits (see section 1.1.2.3): the catalytic subunit (csu) which conducts the thiamine diphosphate catalyzed reaction and the regulatory subunit (rsu) which is responsible for regulation and substrate specificity. In this work, three different catalytic subunits (csu) and two different regulatory subunits (rsu) were discovered from *Eupatorium cannabinum* (Ec). The name of the sequences indicates the type of tissues in which the sequence was discovered (shoot = S, root-tissue culture = W).

3.1.1.1 AHAS Catalytic Subunit of *Eupatorium cannabinum* from Shoot-Derived cDNA (EcScsu-AHAShyp)

As outlined in the "Overview of general procedures" (Figure 2.5, page 55), the internal fragment of the unknown sequence was discovered with the DOP-PCR technique. Subsequently, sequences were completed in 3'- and 5'-direction with the RACE technique.

Degenerate primers for identification of internal fragments

Due to their essential function as first enzyme in the biosynthesis of branched-chain amino acids, AHASs show a significant similarity in large parts of their amino acid sequence even when comparing plant-, fungi-, algae-, and bacterial-AHAS (Duggleby and Pang 2000). Postulating that this would be the same with AHAS of PA-producing plants,

four degenerate primers were designed according to an amino acid alignment of 17 plant AHAScsu (for NCBI accession numbers of amino acid sequences, see appendix Table B.6 on page 223) as described in section "Design of degenerate oligonucleotide primers (DOP)" on page 69.

Primers were positioned in the amino acid alignment as shown in Figure 3.1. With these four primers, four different combinations were possible. The primer combinations and the expected size of the PCR products according to the amino acid alignment of known catalytic subunits of AHAS are shown in Table 3.2 on page 104.

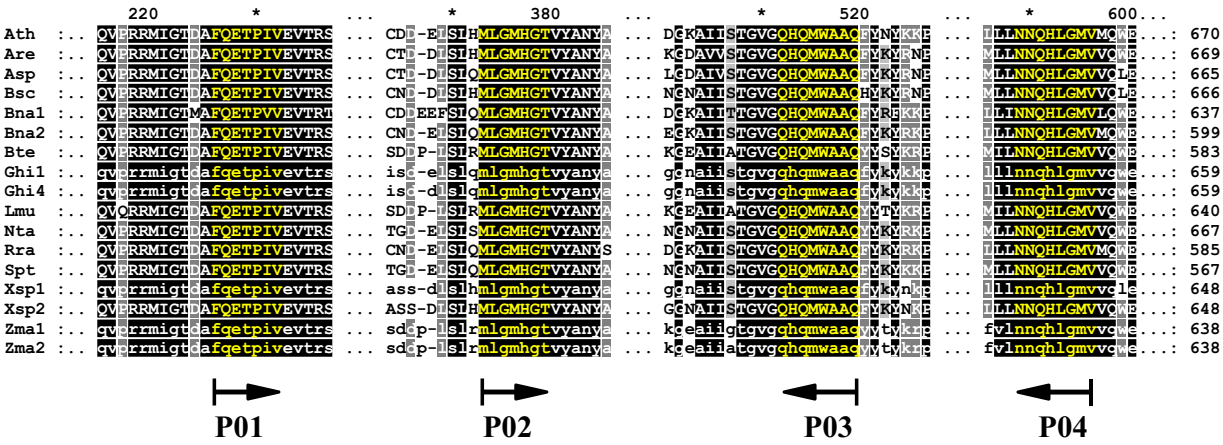


Figure 3.1: Extract from the amino acid alignment of 17 plant AHAScsu used to design degenerate primers. Yellow shaded amino acids show the sequences from which primers were deduced. Positions of primers P01 to P04 are indicated with an arrow below the alignment. Ath = *Arabidopsis thaliana*, Are = *Amaranthus retroflexus*, Asp = *Amaranthus* spec., Bsc = *Bassia scoparia*, Bna1 = *Brassica napus* 1, Bna2 = *Brassica napus* 2, Bte = *Bromus tectorum*, Ghi1 = *Gossypium hirsutum* 1, Ghi4 = *Gossypium hirsutum* 4, Lmu = *Lolium multiflorum*, Nta = *Nicotiana tabacum*, Rra = *Raphanus raphanistrum*, Spt = *Solanum ptychanthum*, Xsp1 = *Xanthium* spec. 1, Xsp2 = *Xanthium* spec. 2, Zma1 = *Zea mays* 1, Zma2= *Zea mays* 2 (for NCBI accession numbers of amino acid sequences, see appendix Table B.6 on page 223, for one letter code of amino acids, see Table B.4).

Internal fragment of clone EcScsu-AHAShyp

Total RNA was isolated from young leaves of sterile cultures (section 2.2 on page 31) of *Eupatorium cannabinum* as described in section 2.4.3.1 on page 58. RNA quality was checked with agarose gel electrophoresis (section 2.4.7 on page 63) and quantified by UV as described in section 2.4.6 on page 63. From 1 μg of total RNA only naturally occurring mRNA (poly-A tailed) was reversely transcribed with the oligo(dT)-primer (for primer sequence, see Table 2.19 on page 71) with the reverse transcriptase M-MuLV RT under

degenerate primer combination		calculated size of PCR products according to the amino acid alignment positions of primers, number of amino acids	predicted size of PCR products: corresponding number of bp
1.	P01 + P03	519 - 226 = 293	879
2.	P01 + P04	596 - 226 = 370	1,110
3.	P02 + P03	519 - 373 = 146	438
4.	P02 + P04	596 - 373 = 223	669

Table 3.2: Expected size of PCR products according to the amino acid alignment of known catalytic subunits of AHAS

PCR parameter	modification
PCR reaction mixture:	
cDNA synthesis	1 μg RNA from shoot tissues, reversely transcribed with M-MuLV RT
Mg^{2+} concentration	1.6 mM
DNA polymerase	Promega *Taq*
PCR programme:	constant
annealing	55°C 1 min
elongation	72°C 2 min
resulting fragment:	EcSA650alt6

Table 3.3: Discovery of the internal fragment of clone EcScsu-AHAShyp: PCR modifications differing from the standard methods described in section 2.4.9

its appropriate conditions as described in section 2.4.8 on page 65. The resulting (dT)-tailed cDNA was submitted to DOP-PCR with primers P01 to P04 in all four possible combinations (Table 3.2 on page 104). The PCR reaction and cycling programme were carried out as described in section "Polymerase chain reaction" on page 67 except for the modifications summarized in Table 3.3 on page 104. Only primer combination no. 4 (P02 + P04, see Table 3.2) resulted in a PCR product with the expected size of 669 bp. It was cloned in vector pCR 2.1-TOPO and propagated in TOP10 cells (for details on vector and host cells, see section "Cloning Material" on pages 34 pp).

Plasmids were isolated with hand method I of minipreparations as described in section 2.4.3.2 on page 59. Plasmids harbouring the PCR product were identified with *Eco*RI restriction analysis (see section 2.4.11.3 on page 79) and submitted to sequencing (see section 2.4.12 on page 79). On the amino acid level, the fragment EcSA650alt6 revealed

primer	applied to	design based on sequence	sequence	T_m[°C]
P05	3'RACE	EcSA650alt6	5'-dGGT TGA CAT TGA CGG TGA TGG AAG TTT TA-3'	63.9
P08	5'RACE as GSP1	EcSA650alt6	5'-dCGA ATG TTT TAA ATG TCA AA-3'	47.1
P09	5'RACE as GSP2	EcSA650alt6	5'-dATC AGT TCT TGC TTC CAT GTC GAG AAA T-3'	62.2
P10	5'RACE as GSP3	EcSA650alt6	5'-dCTA TCT CAG CAG CGT CGA TGT CAA TAT-3'	63.4
P11	5'RACE as GSP1 and in RT	EcS3race2, EcW3race3SP, EcW3race7SP	5'-dTGT ATC CAA CAT YTT CTG AAT-3'	51.0
P14	5'RACE as GSP3	EcS3race2	5'-dTTG TCG GGT TCC CCA AGT AAG TGT GT-3'	64.8

Table 3.4: 3'RACE and 5'RACE PCR gene specific primer for the identification of cDNA ends of clone EcScsu -AHAShyp: application of the primer, sequence according to which the primer is based, primer sequence, and predicted melting temperature. Y=T,C

high similarity to other catalytic subunits of AHAS from plants. Plasmids were screened once again, this time with *Sca*I restriction analysis. Sequencing had provided the information that fragment EcSA650alt6 had contained a *Sca*I restriction site, thus, clones without this restriction site were believed to reveal further AHAS-like clones but different to EcSA650alt6. Unfortunately, sequencing of the selected clones did not detect any additional AHAS-like fragments.

3'end of clone EcScsu-AHAShyp

The P05 gene specific primer for 3'RACE-PCR (see section 2.4.9.2 on page 71) was designed according to the sequence of fragment EcSA650alt6 (see Table 3.4 on page 105). 3'RACE-PCR with P05 as forward primer and with oligo(dT)-primer as reverse primer was performed with the same cDNA and the same composition of PCR reaction mixture as described for the internal fragment EcSA650alt6 (see page 103). The constant PCR programme was adjusted to primer P05 melting temperature: Annealing was carried out for 1 minute at 60°C and elongation was done for 2 minutes at 72°C. The resulting PCR product was cloned, propagated, submitted to restriction analysis, and sequenced as described for the internal fragment EcSA650alt6.

Considering the underlying AHAScsu alignment (Figure 3.1 on page 103) the 3'RACE-PCR product was expected to show a size of 342 bp. Actually, the 3'RACE fragment, designated EcS3race2, had a length of 586 bp. The first 416 bp completed the open reading frame (ORF) of the internal fragment which left a 3' untranslated region (UTR) of 170 bp including the poly-A tail.

5'end of clone EcScsu-AHAShyp

Three gene specific primers for 5'RACE-PCR were designed based on the internal fragment EcSA650alt6 (see Table 3.4 on page 105). P08 was used as gene specific primer 1 (GSP1) in reverse transcription (as described in section 2.4.9.2 on page 71), while P09 was used for the first PCR reaction and P10 was used for the nested PCR reaction (see section 2.4.9.3 on page 72). According to the underlying AHAScsu alignment PCR products from the nested PCR with P10 were expected to result in 1,282 bp large fragments.

In the following, the generation of two 5'RACE clones (EcS5race43 and EcS5race80) is described reflecting the development of a successful 5'RACE method for AHAScsu sequences.

EcS5race43

5'RACE-PCR was performed as described in section 2.4.9 on page 67 with the following modifications: For the first PCR reaction, 3.06 μg of total RNA were reversely transcribed with the reverse transcriptase Superscript II at 50°C and with the help of primer P08. 2 μl of the resulting cDNA were used as template for the first PCR together with 1 μl of primer AAP as forward primer and with 1 μl of primer P09 as reverse primer. The first PCR reaction was supplemented with 0.6 mM Mg^{2+}-ions and was carried out with Invitrogen *Taq* DNA polymerase. Cycling was performed with a constant PCR programme performing annealing at 62°C for 2 minutes and elongation at 72°C for 4 minutes. The resulting PCR reaction mixture was diluted 1:10 and served as a template for the nested PCR reaction. 2 μl of this template were mixed with 1.5 μl AUAP as forward primer and with 1.5 μl P10 as reverse primer in a final concentration of 1.2 mM Mg^{2+}-ions. The nested PCR reaction was carried out as the first PCR step with an modified annealing step at 63.4°C for 2 minutes. The resulting PCR product was cloned, propagated, checked via restriction analysis, and sequenced as described for the internal fragment EcSA650alt6 on page 103.

The 5'RACE fragment EcS5race43 had a length of 957 bp which was 325 bp shorter than predicted and did not include any putative ATG initiation codon. The fragment EcS5race43 showed high similarity to plant AHAScsu and the overlap with the internal fragment EcSA650alt6 consisted of 163 bp as predicted from positioning of primer P10. Thus, the fragment EcS5racce43 was considered to be the correct elongation in 5'direction of the internal fragment EcSA650alt6, but was incomplete and did not provide a putative ATG initiation codon.

In the course of several 5'RACE-PCR trials, the resulting sequences that did not contain a putative ATG initiation codon of the ORF. Sequence shortage could have occurred during the process of 5'RACE technique at four levels:

1. at the level of total RNA's quality: A slow extracting process and/or the use of an overstored RNA in reverse transcription may have resulted in shortened fragments due to RNA fragmentation by RNase activity.

2. at the level of mRNA's features and reverse transcription: Putative secondary structures especially at the 5'end may have caused a premature stop of the reverse tran-

scription since the employed "standard" reverse transcriptases were not recommended for the use with higher temperatures which were necessary to solve secondary structures.

3. at the level of PCR: *Taq* DNA polymerase made mismatches which may have resulted in amplification products being too short in 5'direction.

These possible reasons for truncated PCR products were considered and resulted in the following changes in the 5'RACE procedure:

- RNA quality was improved in harvesting very young and still growing material with high cell division rates and high numbers of unfragmented mRNAs as shown by agarose gel electrophoresis (section 2.4.7 on page 63). Harvesting was carried out as quick as possible and total RNA was aliquoted and submitted to a maximum of 3 freeze-thaw cycles.
- The type of reverse transcriptase was changed to enzymes tolerating temperatures exceeding 42°C.
- Ordinary *Taq* DNA polymerase was replaced by a mixture of *Taq* DNA polymerase and proof-reading polymerase (Accu*Taq* DNA polymerase, Sigma). With Accu*Taq* DNA polymerase mistakes of the *Taq* DNA polymerase were corrected by included proof-reading features which should result in a continued and elongated amplification.
- The elongation time was prolonged in each PCR cycle to allow completion of all strands while the elongation temperature was reduced to 68°C to extend the life time of the *Taq* DNA polymerase.
- Decisive improvements in obtaining elongated 5'RACE fragments were made when using the GSP1 for reverse transcription as well as for the first PCR reaction of a nested PCR. In general, GSP1 was 10 bp shorter than GSP2 or GSP3 which allowed a simplified annealing to a partially inaccessible template probably covered in secondary structures.

<u>EcS5race80</u>

A new primer (P11, see Table 3.4 on page 105) was designed on the basis of the 3'end sequence EcS3race2. P11 was positioned 190 bp downstream of P04 to have a change of the starting sequence for reverse transcription. This should allow an improved access to mRNA that was thought to be covered in secondary structures at the 5'end.

5'RACE-PCR was performed as described above for sequence EcS5race43 on page 106 except for the modifications summarized in Table 3.5 on page 109. One extraordinary modification shall be mentioned here: Primer P11, that was used previously in reverse transcription, now served as reverse primer in the first PCR.

An improved quality of the new cDNA was visualized by agarose gel electrophoresis (see Figure 3.2 on page 108): a test-PCR was conducted with degenerate primers P02 and P04 (described in section "Internal fragment of clone EcScsu-AHAShyp" on page 103) and showed diffuse and weak bands (lane 1 and 3) obtained from the previously used ("old") cDNA that had been generated with P08 and resulted in the shortened fragment EcS5race43, while test-PCR products obtained from new cDNA which was generated with P11 and resulted in fragment EcS5race80 were bright and clear at the expected size of approximately 670 bp (lane 2 and 4).

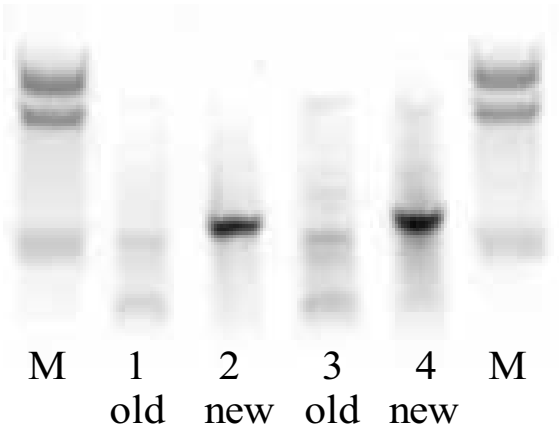


Figure 3.2: Comparing the quality of cDNA for 5'RACE PCR produced by reverse transcription with either gene specific primer P08, named "old", or with primer P11, named "new": 1.5% agarose gel stained with ethidium bromide showing separated PCR products from cDNA of shoot tissues (lane 1 and 2) and of root tissue cultures (lane 3 and 4). M = marker: 100 bp ladder (Invitrogen).

The size of the 5'RACE nested PCR product was expected to be 1,282 bp, which was the same as described for sequence EcS5race43 due to the use of primer P10. The resulting 5'RACE fragment EcS5race80 had a length of 1149 bp, showed the expected overlap with the internal fragment EcSA650alt6, and contained a putative ATG initiation codon including a stretch of 2 bp from 5'UTR. As shown in the alignment of various Angiosperm plant species (Figure 3.15 on page 140) the 5'end within the ORF displays a rather variant length differing in 108 bp when considering *Arabidopsis thaliana* and *Zea mays* AHAScsu. Hence, further PCRs were conducted, but none of them was able to reveal an upstream located ATG initiation codon.

The resulting hypothetical AHAS catalytic subunit, isolated from shoot material of *Eupatorium cannabinum* (EcScsu-AHAShyp), was composed of the three following partial sequences: EcSA650alt6 as internal part, EcS3race2 as 3'end, and EcS5race80 as 5'end (see Figure 3.3). The ORF of EcScsu-AHAShyp included 1,9530 bp coding for 650 aa.

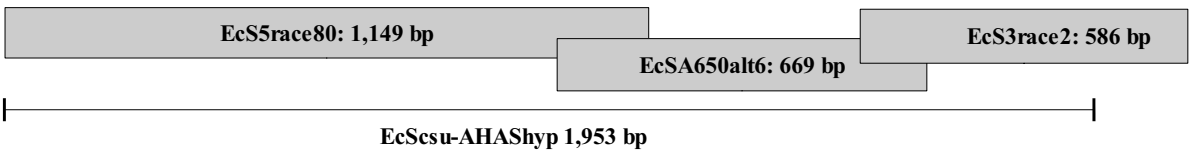


Figure 3.3: Sequence EcScsu-AHAShyp is composed of three separately detected parts: EcSA650alt6 as internal fragment, EcS3race2 as 3'end, and EcS5race80 as 5'end. The line resembles the length of the ORF starting with the putative ATG initiation codon and ending with the stop codon.

PCR parameter	modifications	
PCR reaction mixture:	first PCR	nested PCR
cDNA synthesis	1.62 μg RNA from shoot tissues, reversely transcribed with P11 and SS III at 55°C	
forward primer	AAP: 2 μl	AUAP: 1.5 μl
reverse primer	P11: 2 μl	P10: 1.5 μl
template	2 μl	2 μl of first PCR diluted 1:100
Mg^{2+} concentration	1.6 mM	3.5 mM
DNA polymerase	Invitrogen *Taq*	Accu *Taq*
PCR programme:	constant	constant
annealing	48.4°C 1 min	67.5°C 1 min
elongation	68°C 10 min	68°C 10 min
resulting fragment:		EcS5race80

Table 3.5: 5'RACE PCR modifications for clone EcScsu-AHAShyp differing from the standard methods described in section 2.4.9

3.1.1.2 AHAS Catalytic Subunits of *Eupatorium cannabinum* from Root-Derived cDNA (EcW2csu-AHAShyp and EcW3csu-AHAShyp)

The biosynthesis of necic acids is postulated to take place close to the biosynthesis of the necine base as one moiety of the necine base is esterified with one or more moieties of necic acids to form a complete pyrrolizidine alkaloid (PA). The homospermidine synthase (HSS), responsible for the formation of the first intermediate in the necine base biosynthesis, had been immuno-localized by Anke et al. (2004) for *Eupatorium cannabinum* in the cortical parenchyma of young roots from plants grown in the institute garden. Hence, the place of protein biosynthesis for the necic acid-specific AHAS-like enzyme was postulated to be in the roots, too. According to Jones et al. (1985), the AHAS is known to be active in the chloroplasts. Its mRNA is nuclear-encoded and contains a chloroplast-targeting transit peptide. In order to reduce the number of mRNA coding for AHAS from primary metabolism and to increase the percentage of mRNA coding for AHAS from secondary metabolism, may be localized in other plastids than chloroplasts, chloroplast-free root tissues were used for isolation of RNA.

While clone EcScsu-AHAShyp from shoot tissues was composed of three parts, the two root-derived clones EcW2csu-AHAShyp and EcW3csu-AHAShyp (described in the following sections) were composed of only two fragments: the 3'end and the 5'end. An internal fragment from root tissues could be discovered by DOP-PCR (not described), but could not be prolonged in 3'direction although two different gene specific primers for

3'RACE-PCR were designed and tested. The resulting 3'RACE PCR products did not show an overlap with the underlying internal fragment but confirmed the 3'end of clone EcScsu-AHAShyp. The internal fragment turned out to contain multiple frame shifts at its 3'end but confirmed clone EcW2csu-AHAShyp with its 5'end.

Meanwhile, 3'RACE-PCR with primer P05 produced two different but AHAScsu-like fragments with root-tissue culture derived template, although P05 was designed as gene specific primer on the basis of the internal fragment EcSA650alt6 from shoot tissues.

3'ends of clones EcW2csu-AHAShyp and EcW3csu-AHAShyp

3'RACE-PCR with primer P05 was carried out under the same conditions as described above for the 3'RACE fragment EcS3race2 (page 105) except for the origin of the template which was harvested from root-tissue cultures (for generation and maintenance of root-tissue cultures, see "Growing conditions of plants" on page 33). The single resulting PCR product was cloned in pCR 2.1-TOPO and propagated in TOP10 cells as described earlier. Isolated plasmids were submitted to extended restriction analysis with *Eco*RI, *Bsa*AI, *Eco*RV, and *Afl*III that had been chosen according to the restriction map of the firstly discovered 3'end EcS3race2. Based on these informations, two plasmids could be chosen for sequencing and actually showed to be different from each other as well as from the shoot-derived 3'end EcS3race2. The fragment EcW3race3SP had a length of 548 bp providing a translation termination codon at position 432 bp which left a 116 bp for a 3'UTR ending with the poly-A tail. The fragment EcW3race7SP measured 582 bp in length, the translation termination codon finished with bp number 417, and the 3'UTR had a length of 165 bp including the poly-A tail.

5'end for EcW2csu-AHAShyp

The gene specific primer P16 (see Table 3.6) was designed for 5'RACE-PCR in order to prolong the 3'end EcW3race3SP (identified as described on page 110) in 5'direction. Primer P11 was used for reverse transcription and for the first PCR in the 5'RACE procedure (as described above on page 107 for EcScsu-AHAShyp) since P11 fitted to either sequence EcW3race3SP and EcS3race2, respectively (see Table 3.4). Primer P16 was positioned 88 bp upstream of P11 which should have resulted in a 1,722 bp PCR fragment for sequence complementation judging from sequence EcScsu-AHAShyp.

The first PCR in the 5'RACE procedure was carried out as described for EcS5race80 on page 107 except for the following modifications listed in Table 3.7 on page 112. Previous PCR trials (data not shown) had revealed that AUAP had annealed to the our template in both forward and reverse directions and had produced artefacts. Thus, primer AUAP was not used any more.

Nested PCR was performed with a touch down programme as described in 2.4.9 on page 67 with the following modifications: Cycling no.1 comprised 16 cycles decreasing the annealing temperature from 71°C to 63.5°C, and cycling no.2 was carried out for 19 cycles at 63°C. Sequencing of the resulting PCR-product showed a fragment of 1,725 bp and was designated EcW5r3.2W2.15. The fragment included a putative ATG initiation codon and 8 bp of 5' UTR.

primer	applied to	design based on sequence	sequence	T_m[°C]
P15	5'RACE as GSP3	EcW3race7SP	5'-dCGT TTG ACG GGT TTC CTA AGT AGG TAT-3'	63.4
P16	5'RACE as GSP3	EcW3race3SP	5'-dTTT CTT TTG TTG GGA TTC CTA AGA ATG AT-3'	59.6
P23	confirmation of EcW3csu	EcW5r3.2W3.19	5'-dCCC ACC GCC ATC CTC CGC CGC ATA A-3'	71.2
P24	confirmation of EcW3csu	EcW5r3.2W3.19	5'-dAAA CCG CAC CCC AAA TGC AAG TAA AAC ATC-3'	65.4

Table 3.6: 5'RACE PCR gene specific primer for *E. cannabinum*-csu from root-tissue cultures: application of the primer, sequence according to which the primer is based, primer sequence, and predicted melting temperature.

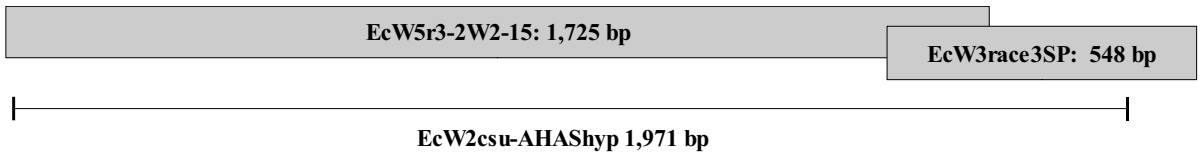


Figure 3.4: Clone EcW2csu-AHAShyp is composed of two separately detected parts: EcW3race3SP as 3'end and EcS5r3-2W2-15 as 5'end. The line resembles the length of the ORF starting with the putative ATG initiation codon and ending with the stop codon.

Unfortunately, a frame shift of 2 bp might have occurred at bp no. 1,013. In order to find out whether this was due to mistakes from the polymerase or from the reverse transcriptase, the nested PCR was carried out once again with the following modifications: As a template, the first PCR-product was diluted 1:500 and cycling was carried out as constant PCR programme, annealing was performed at 66°C for 1 minute, and elongation at 68°C for 10 minutes. Sequencing of the resulting fragment, named EcW5r3.2W2.4, revealed that at position 1,013 two bp had to be inserted in sequence EcW5r3.2W2.15. It was concluded that the two previously missing bp where due to a polymerase mistake since the correcting fragment EcW5r3.2W2.4 was obtained from the same cDNA under altered PCR-cycling conditions. Accu*Taq* DNA polymerase is known to produce a rate of 0.3% mistakes, i.e., 43.85×10^{-6} mistakes/bp/cycle: Hence, findings of 2 bp out of 1,725 bp (0.12%) are within the limits.

Several additional PCR approaches confirmed this insertion and none of them was able to prove a more upstream located ATG initiation codon.

5'RACE sequence EcW5r3.2W2.15 was combined with 3'RACE sequence EcW3race3SP to the hypothetical sequence EcW2csu-AHAShyp (see Figure 3.4 on page 111). This sequence, isolated from root-tissue cultures, contained an ORF of 1,971 bp that encodes 656 amino acids.

PCR parameter	modifications	
PCR reaction mixture:	first PCR	nested PCR
cDNA synthesis	2 μg RNA from root tissues, reversely transcribed with P11 and SS III at 55°C	
forward primer	AAP 2 μl	AAP 1.5 μl
reverse primer	P11: 2 μl	P16: 1.5 μl
template	1 μl	2 μl of first PCR diluted 1:1,000
Mg^{2+} concentration	2.5 mM	2.5 mM
DNA polymerase	Accu*Taq*	Accu*Taq*
PCR programme:	constant	touch down
		cycling no.1 repeated 16x:
annealing	51°C 20 sec	71°C → 63.5°C 1 min
elongation	68°C 20 min	68°C 10 min
		cycling no.2 repeated 19x:
annealing		63°C 1 min
elongation		68°C 10 min
resulting fragment:		EcW5r3.2W2.15

Table 3.7: 5'RACE PCR modifications for clone EcW2csu-AHAShyp differing from the standard methods described in section 2.4.9

PCR parameter	modifications	
PCR reaction mixture:	nested PCR	standard PCR
cDNA synthesis		5 μg RNA from root tissues, reversely transcribed with oligo(dT)-primer and eAMV RT at 60°C
forward primer	AAP: 1.5 μl	P23: 2 μl
reverse primer	P15: 1.5 μl	P24: 2 μl
template	2 μl of first PCR in Table 3.7 diluted 1:1,000	1 μl cDNA
Mg^{2+} concentration	2.5 mM	2.5 mM
DNA polymerase	Accu*Taq*	Accu*Taq*
PCR programme:	constant	constant
annealing	65°C 1 min	68°C 90 sec
elongation	68°C 10 min	72°C 8 min
resulting fragment:	EcW5r3.2W3.19	EcW3-5r6check

Table 3.8: 5'RACE and standard PCR modifications for clone EcW3csu-AHAShyp differing from the standard methods described in section 2.4.9

5'end for EcW3csu-AHAShyp

Primer P15 (see Table 3.6 on page 111) was designed for 5'prolongation of the second out of two 3'ends identified from *Eupatorium cannabinum* root-tissue cultures, i.e. fragment EcW3race7SP (page 110). The first PCR in the 5'RACE procedure was carried out with primer P11 in the same way as described for EcW2csu-AHAShyp (see Table 3.7 on page 112). As indicated in Table 3.4, primer P11 was designed to fit to all three 3'ends from *Eupatorium cannabinum*, i.e. EcScsu-AHAShyp (EcS3race2), EcW2csu-AHAShyp (EcW3race3SP), and EcW3csu-AHAShyp (EcW3race7SP) and, thus P11 was designed as a two-fold degenerate primer.

The PCR-product from the first PCR reaction with P11 was diluted 1:1,000 and 2 μl were used as template in the nested PCR reaction. A constant PCR cycling programme was performed with AAP as forward primer and P15 as reverse primer as detailed in Table 3.8.

The resulting fragment EcW5r3.2W3.19 consisted of 1,770 bp with 52 bp 5'UTR and formed the expected overlap with the 3'end EcW3race7SP. Unfortunately, an early stop at amino acid 131 occurred in the translated amino acid sequence. Hence, a second PCR was conducted with a new pair of primers (P23, P24, see Table 3.6) carried out as standard PCR (modifications detailed in Table 3.8) to re-read the critical part and to find out

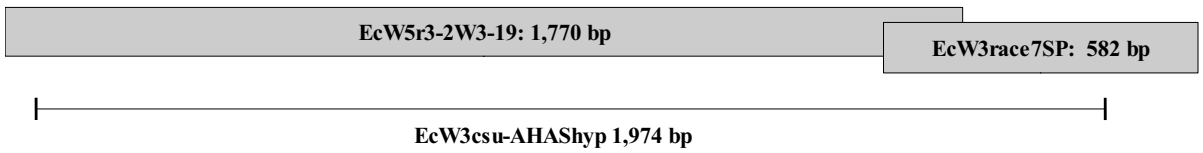


Figure 3.5: Clone EcW3csu-AHAShyp is composed of two separately detected parts: EcW3race7SP as 3'end and EcS5r3-2W3-19 as 5'end. The line resembles the length of the ORF starting with the putative ATG initiation codon and ending with the stop codon.

whether there was really a stop within the genetic information or just a polymerase mistake that had occurred during PCR. A similar situation was reported earlier for sequence EcW5r3.2W2.15 (see page 110). The resulting fragment EcW3-5r6check from PCR with primers P23 and P24 revealed a frame shift in sequence EcW5r3.2W3.19 at bp 264 due to an omitted G caused by the polymerase. The corrected 5'RACE sequence EcW5r3.2W3.19 was combined with the 3'RACE sequence EcW3race7SP to the hypothetical sequence EcW3csu-AHAShyp (see Figure 3.5 on page 114). This sequence, isolated from root-tissue cultures, contained an ORF of 1,974 bp that coded for 657 amino acids.

3.1.1.3 AHAS Regulatory Subunit of *Eupatorium cannabinum* from Shoot-Derived cDNA (EcSrsu-AHAShyp)

The search for regulatory subunits of acetohydroxyacid synthase (AHASrsu) from *Eupatorium cannabinum* was carried out as outlined in the "Overview of general procedures" on page 54: The internal fragments were discovered with DOP-PCR and the completion in 3' and in 5'direction was managed with RACE technique. All PCR products obtained in this search for AHASrsu were cloned in vector pGEM-T Easy instead of pCR 2.1-TOPO which was used for the catalytic subunits.

Degenerate primers for identification of internal fragments

Two degenerate primers were designed according to an amino acid alignment of 6 plant AHASrsu (for NCBI accession numbers of amino acid sequences, see appendix Table B.7 on page 224) as described in section 2.4.9.1 on page 69. As observed in the alignment of catalytic subunits from AHAS (see page 103), amino acid sequences from regulatory subunits of AHAS show large areas of high similarities as well. According to these conserved regions, two DOP-PCR primers were positioned in the alignment as shown in Figure 3.6 on page 115. In order to decide whether a DOP-PCR product with these two primers was likely to be a part of the regulatory subunit of AHAS, the expected size of the PCR product was calculated. P26 ended at amino acid 472, while P25 started with amino acid 131 of the alignment. The numbers of amino acids lying in between were 341 that corresponded to 1,023 bp. The exact sequences of the two degenerate primers are given in Table 3.9.

```
          120         *       140          ..........  440         *       460         *       ......
Ath  :    GYNIESLAVGLNEDKALFTIVVLGTDKVLQQV .......... KAIDVSDHTITLEVTGDLRKMSALQTQLEAYGICEVART ......: 477
Ath5 :    GYNIESLAVGLNRDKALFTIVVCGTERVLQQV .......... KAVDVSDHTITLQLTGDLDKMVALQRLLEFYGICEVART ......: 491
Npl  :    GYNIESLAVGLNKDKALFTIVVSGTERVLQQV .......... KAVDVSDHTITLELTGDLHKMVRLQRLLEFYGICEVART ......: 455
Osa1 :    GYNIESLAVGLNKDKALFTIVVSGTEKILKQV .......... KTVDVSDHTVTLQLTGDLDKMVALQRMLEFYGICEVART ......: 476
Osa2 : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ .......... KSVDVSDHTVTLQLTGDLDKMVALQRLLEFYGICEVART ......: incomplete
Zma  : ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ .......... KPIDVSDHTVTLQLTGDLDKMVALQRLLEFYGICEVART ......: incomplete
```

P25 **P26**

Figure 3.6: Extract from the amino acid alignment of 6 plant AHASrsu used to design degenerate primers. Amino acid sequences are shaded yellow from which primers were deduced. Positions of primers 25 and 26 (P25, P26) are indicated with an arrow below the alignment. Ath = *Arabidopsis thaliana*, Npl = *Nicotiana plumbaginifolia*, Osa = *Oryza sativa*, Zma = *Zea mays*. Amino acids are shown in the one letter code (Table B.4).

primer	direction	sequence	degen-eracy	T_m[°C]
P25	forward	5'-dGAY AAR GCI YTN TTY ACN AT-3'	512 x	51.2
P26	reverse	5'-dGCI ACY TCR CAD ATN CCR TA-3'	96 x	55.9

For IUPAC ambiguity code, see Table B.3 on page 221

Table 3.9: Degenerate primers for identification of AHASrsu: sequences, degree of degeneracy, and predicted melting temperature. Sequences are written in IUPAC Ambiguity Code.

Internal fragment of clone EcSrsu-AHAShyp

For identification of internal parts from regulatory subunits of AHAS, degenerate oligonucleotide primed PCR (DOP-PCR) was carried out as described for the catalytic subunit on page 103 with modifications summarized in Table 3.10. The resulting PCR-product was cloned and plasmids were isolated with hand method II of minipreparations (described in section 2.4.3.2 on page 60), submitted to restriction analysis, and sequenced. The fragment EcS03_M_17 revealed high similarity to other regulatory subunits of AHAS from plants and had a length of 1,018 bp.

3'end and 5'end of clone EcSrsu-AHAShyp

On the internal sequence EcS03_M_17, three gene specific primers were designed (sequences are shown in Table 3.11 on page 116). 3'RACE-PCR primer P27 was positioned 414 bp downstream of the forward degenerate primer P25, while 5'RACE-PCR primer P30 was located 502 bp downstream of P25. Consequently, the resulting RACE-PCR fragments should give an overlap with each other. 5'RACE primer P29 was used in reverse transcription as well as in the first PCR of the 5'RACE procedure.

3'RACE-PCR was carried out with 2 μl of the same cDNA as used for the internal sequence EcS03_M_17. A constant PCR cycling programme was performed as described in 2.4.9 on page 67 with modifications detailed in Table 3.12 on page 117. The resulting PCR-product was designated EcS03_3R-07: It comprised 838 bp that contained 657 bp

PCR parameter	modification
PCR reaction mixture:	
cDNA synthesis	1 μg RNA from shoot tissues, reversely transcribed with SS III at 55°C
template	3 μl cDNA
DNA polymerase	Accu*Taq*
PCR programme:	constant
annealing	50°C 3 min
elongation	68°C 20 min
resulting fragment:	EcS03_M_17

Table 3.10: Discovery of the internal fragment of clone EcSrsu-AHAShyp: PCR modifications differing from the standard methods described in section 2.4.9 and from the settings chosen for the identification of the internal fragment of the catalytic subunit EcSA650alt6 described in section 3.1.1.1

primer	applied to	design based on sequence	sequence	T_m[°C]
P27	3'RACE	EcS03_M_17	5'-dAAG AGA ACA GAG TAT CAG GAA TCT G-3'	59.7
P29	5'RACE as GSP1 and for RT	EcS03_M_17	5'-dAAC CCC CCA ATG TGC ATC AA-3'	57.3
P30	5'RACE as GSP2	EcS03_M_17	5'-dCTT GAT TGA ATG AAT AGT CAT CTC TAG A-3'	59.3

Table 3.11: 3'RACE and 5'RACE PCR gene specific primer for identification of the cDNA ends of clone EcSrsu -AHAShyp (P27 = 3'RACE primer, P29 and P30 = 5'RACE primers): application of the primer, sequence according to which the primer is based, primer sequence, and predicted melting temperature.

PCR parameter	modifications for EcSrsu-AHAShyp		
	3'RACE	5'RACE	
PCR reaction mixture:		first PCR	nested PCR
cDNA synthesis		1 μg RNA from shoot tissues, reversely transcribed with P29 and SS III at 55°C	
forward primer	P27	AAP	AAP 1.5 μl
reverse primer	oligo(dT)	P29	P30: 1.5 μl
template	2 μl	4 μl	2 μl of first PCR diluted 1:10
DNA polymerase	Accu*Taq*	Accu*Taq*	Accu*Taq*
PCR programme:	constant	constant	constant
annealing	54°C 90 sec	50°C 4 min	54.6°C 4 min
elongation	72°C 2 min	68°C 10 min	68°C 10 min
resulting fragment:	EcS03_3R_07		EcS03_5R_05

Table 3.12: 3'RACE and 5'RACE PCR modifications for clone EcSrsu-AHAShyp differing from the standard methods described in section 2.4.9

within the ORF and 181 bp as 3'UTR including the poly-A tail.

Details for the 5'RACE procedure are summarized in Table 3.12 on page 117. The resulting 5'RACE-PCR product, designated EcS03_5R_05, had a length of 870 bp that included the putative ATG initiation codon and 11 bp of 5'UTR.

3'RACE fragment EcS03_3R_07 and 5'RACE fragment EcS03_5R_05 showed the expected overlap and formed the hypothetical full-length sequence EcSrsu-AHAShyp (see Figure 3.7 on page 118). The ORF of EcSrsu-AHAShyp included 1,431 bp coding for 476 amino acids.

3.1.1.4 AHAS Regulatory Subunit of *Eupatorium cannabinum* from Root-Derived cDNA (EcWrsu-AHAShyp)

Internal fragment of clone EcWrsu-AHAShyp

The internal fragment of clone EcWrsu-AHAShyp was discovered from sterile root tissues of *Eupatorium cannabinum* with DOP-PCR as described earlier for internal sequences. The DOP-PCR set-up was the same as described for the identification of the internal fragment of the shoot originating clone EcSrsu-AHAShyp (see above, page 115) except for the template which was obtained by reverse transcription from 2 μg of total RNA from root-tissue cultures (see section 2.2 on page 31) and except for the annealing

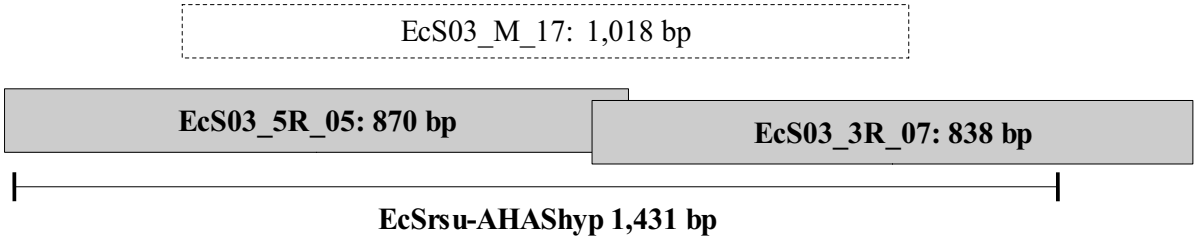


Figure 3.7: Clone EcSrsu-AHAShyp is composed of two separately detected parts: EcS03_3R_07 as 3'end and EcS03_5R_05 as 5'end. The internal fragment EcS03_M_17 was used for RACE primer design, showed similarity with each 3'end and 5'end but was not considered for construction of the full-length sequence. The line resembles the length of the ORF starting with the putative ATG initiation codon and ending with the stop codon.

primer	applied to	design based on sequence	sequence	T_m[°C]
P28	3'RACE	EcW01_M_12	5'-dAGC TAC ATA AAG GGT CAA GTT CTT GAT G-3'	62.2
P31	5'RACE as GSP1 and for RT	EcW01_M_12	5'-dATC TTC CCT GGA TCC CCT GTA-3'	59.8
P32	5'RACE as GSP2	EcW01_M_12	5'-dTAT TCT GAT GCA TCC ACC ACG TTG CCT-3'	65.0

Table 3.13: 3'RACE and 5'RACE PCR gene specific primer for the identification of cDNA ends of clone EcWrsu -AHAShyp (P28 = 3'RACE primer, P31 and P32 = 5'RACE primers): application of the primer, sequence according to which the primer is based, primer sequence, and predicted melting temperature.

temperature at 52°C. The resulting fragment EcW01_M_12 had a length of 1,018 bp, it showed similarity to other regulatory subunits of AHAS from plants, and, moreover, it differed from the internal fragment EcS03_M_17 which led to the conclusion that both sequences encode different regulatory subunits of *Eupatorium cannabinum*.

3'end and 5'end of clone EcWrsu-AHAShyp

The internal sequence EcW01_M_12 served as a basis to design three gene specific primers which were positioned at sequence stretches differing significantly from sequence EcS03_M_17: P28 was used in 3'RACE-PCR while P31 and P32 were employed in 5'RACE procedure. Primer sequences are given in Table 3.13 on page 118.

3'RACE-PCR with P28 was carried out with the same template as used for identification of the internal fragment EcW01_M_12. PCR modifications were the same as applied for 3'RACE of the shoot-derived sequence EcS03_3R_07 (see Table 3.12 on page 117) except for the annealing temperature at 55°C. The resulting sequence EcW01_3R_04

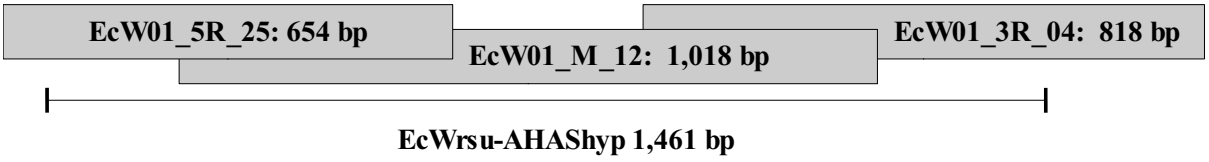


Figure 3.8: Clone EcWrsu-AHAShyp is composed of three separately detected parts: EcW01_M_12 as internal fragment, EcW01_3R_04 as 3'end, and EcW01_5R_05 as 5'end. The line resembles the length of the ORF starting with the putative ATG initiation codon and ending with the stop codon.

comprised 818 bp with 589 bp coding region and 229 bp of 3'UTR including the poly-A tail and formed the expected overlap with the internal fragment EcW01_M_12.

The 5'RACE procedure for the regulatory subunit of the root-derived sequence only differed from the procedure for the shoot-derived sequence in the template and the primers (see Table 3.12 on page 117). For the first PCR, 2 μg total RNA from root-tissue cultures were reversely transcribed with P31 which subsequently was used in the first PCR as reverse primer. The resulting PCR-product was diluted 1:10 and served as template in the nested PCR that was primed with P32 as reverse primer. The annealing temperature was adjusted to P32 at 60°C. The resulting fragment EcW01_5R_25 showed the expected overlap with the internal fragment EcW01_M_12 and had a length of 654 bp including the putative ATG initiation codon and a 61 bp 5'UTR.

The hypothetical full-length sequence EcWrsu-AHAShyp was composed of three fragments (see Figure 3.8 on page 119): the internal fragment EcW01_M_12, the 3'end EcW01_3R_04, and the 5'end EcW01_5R_25. The whole sequence encodes an ORF of 1,461 bp (486 aa, respectively).

3.1.2 AHAS Catalytic Subunit of *Parsonsia laevigata*

Sterile cultures from plants (see section 2.2 on page 31) formed the starting material for the discovery of AHAS catalytic subunits from *Parsonsia laevigata* (Pl). Total RNA was freshly harvested (as described in section 2.4.3.1) from very young and still soft leaves since RNA from full-grown and hard leaves showed poor quality and quantity. For discovery of the internal fragments with DOP-PCR and for their prolongation in 3'direction (3'RACE), 5 μg RNA were reversely transcribed with Superscript III (SSIII) at 55°C (as described in section 2.4.8 on page 65). The resulting cDNA was submitted to DOP-PCR with primers P02 and P04 designed for the discovery of AHAS catalytic subunits (see section 3.1.1 on page 102). DOP-PCR was carried out with modifications as summarized in Table 3.15 on page 121. The resulting PCR-product was cloned for propagation in vector pCR 2.1-TOPO. The selected clone revealed a 671 bp large internal fragment named PlSmitte23 with high similarities to AHAScsu.

According to the sequence of PlSmitte23, gene specific primers for 3'RACE and

primer	applied to	design based on sequence	sequence	T_m[°C]
P33	3'RACE	PlSmitte23	5'-dGTG CTG CTG TTG CAA GAC CTG ATT CA-3'	64.8
P34	5'RACE as GSP1 and for RT	PlSmitte23	5'-dCAT TTT CCT TGC TTC CAA CAT-3'	54.0
P35	5'RACE as GSP2	PlSmitte23	5'-dGCC ACC TTG ATA TCA GCA CAA ATA GA-3'	61.6

Table 3.14: 3'RACE and 5'RACE PCR gene specific primer for the identification of cDNA ends of clone PlScsu -AHAShyp (P33 = 3'RACE primer, P34 and P35 = 5'RACE primers): application of the primer, sequence according to which the primer is based, primer sequence, and predicted melting temperature.

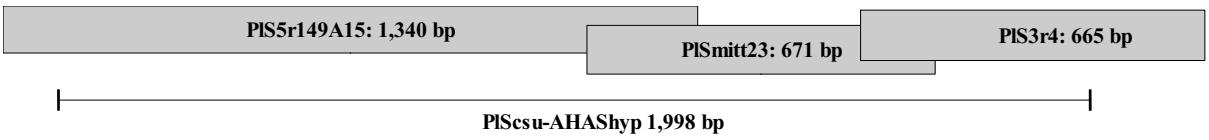


Figure 3.9: Clone PlScsu-AHAShyp is composed of three separately detected parts: PlSmitte23 as internal fragment, PlS3r4 as 3'end, and PlS149A15 as 5'end. The line resembles the length of the ORF starting with the putative ATG initiation codon and ending with the stop codon.

5'RACE procedures were designed. Other gene specific primers from previous RACE procedures with cDNA from *Eupatorium cannabinum* (see section 3.1.1.1 on page 105 pp) did not fit to the internal fragment from *Parsonsia laevigata*, although the amino acid sequences showed similarities. Primer P33 for 3'RACE-PCR was located 147 bp upstream of P04 while for 5'RACE-PCR primer P34 was positioned 251 bp downstream of P02 and primer P35 was placed 212 bp downstream of P02. For gene specific primer sequences, see Table 3.14.

The 3'RACE procedure was carried out as summarized in Table 3.15 on page 121. The resulting 3'end was designated PlS3r4. It comprised 665 bp and was composed of 448 bp ORF and 217 bp 3'UTR including the poly-A tail.

The 5'RACE procedures are summarized as well in Table 3.15 on page 121 and resulted in a sequence named PlS5r149A15 that had a length of 1,340 bp with the ORF including the putative ATG initiation codon and 105 bp of 5'UTR.

The 3'end and the 5'end both showed the expected overlap with the internal fragment PlSmitte23. Together, these three partial sequences formed the hypothetical full-length sequence of the AHAS catalytic subunit from *Parsonsia laevigata* designated PlScsu-AHAShyp with a resulting ORF of 1,998 bp (see Figure 3.9 on page 120) coding for 665 amino acids.

PCR parameter	modifications			
	DOP	3'RACE	5'RACE	
PCR reaction mixture:			first PCR	nested PCR
cDNA synthesis	5 μg RNA from shoot tissues, reversely transcribed with oligo(dT) and SS III at 55°C		3 μg RNA from shoot tissues, reversely transcribed with P34 and SS III at 55°C	
forward primer	P02	P33	AAP	AAP
reverse primer	P04	oligo(dT)	P34	P35
template	1 μl	1 μl	1 μl	2 μl of first PCR diluted 1:10
PCR programme:	constant	constant	constant	constant
annealing	50°C 1 min	61°C 1 min	50°C 2 min	65°C 1 min
elongation	72°C 4 min	72°C 2 min	68°C 6 min	68°C 10 min
resulting fragment:	PlSmitte23	PlS3r4		PlS5r149A15

Table 3.15: DOP-, 3'RACE and 5'RACE PCR modifications for clone PlScsu-AHAShyp differing from the standard methods described in section 2.4.9

In order to find more AHAS catalytic subunits of *Parsonsia laevigata*, RNA from root-tissue cultures (see section 2.2 on page 31) was investigated as described for the internal sequences from shoot tissues (PlSmitte23). The resulting sequences were 99.6% similar to PlSmitte23 (3 nucleotides difference) and were considered to belong to clone PlScsu-AHAShyp. There may be two reasons for this result: First, any further AHAScsu is of rather low abundance, or second, there is no further AHAScsu as it is the case with *Arabidopsis thaliana* where the genome contains only a single AHAScsu gen.

3.1.3 AHAS Catalytic Subunits of *Symphytum officinale*

Two different catalytic subunits (csu) were discovered from *Symphytum officinale* (So), one from shoot tissues and the other from root tissues as detailed in the following sections.

3.1.3.1 AHAS Catalytic Subunit of *Symphytum officinale* from Shoot-Derived cDNA (SoScsu)

Internal fragment of clone SoScsu

Sterilely grown rosette plants (see section 2.2.1 on page 31) were used for harvesting leaves. Total RNA was isolated from these leaves as described in section 2.4.3.1 on page 58. An amount of 4 μg total RNA was reversely transcribed with M-MuLV RT at 42°C as shown in section 2.4.8 on page 65. DOP-PCR (section 2.4.9.1) with primers P02 and P04 (see Table 3.1 on page 102) was successful with the constant PCR reaction as described in section 3.1.1.1 on page 102 with modifications summarized in Table 3.16 on page 123. The resulting PCR-product was cloned and propagated as described in section 3.1.1.1 on page 103. Selection of positive transformants with *Eco*RI-restriction analysis and subsequent sequencing yielded an AHAScsu-like internal fragment named SoS650-5 that comprised 671 bp in length. Plasmids were screened once again with *Dra*I restriction analysis in the search for sequences differing from SoS650-5, but no additional AHAScsu-like fragments were identified.

3'end of clone SoScsu

The gene specific 3'RACE primer P36 was designed according to the internal fragment SoS650-5 (sequence is given in Table 3.17) and was located 97 bp upstream of P04. The 3'RACE procedure is summarized in Table 3.16 on page 123. The resulting cloned and sequenced PCR-product named SoS3'r-9 had a length of 666 bp that contained the expected overlap with the internal fragment. A 397 bp large part coded for the ORF including the translation termination codon and a 269 bp large 3'UTR that harboured the poly-A tail.

5'end of clone SoScsu

Two gene specific primers P38 and P39 (see Table 3.17 on page 124) were designed for 5'RACE-PCR based on the internal fragment SoS650-5. Primer P38 was positioned 324 bp downstream of primer P02 and was used as gene specific primer 1 (GSP1) in reverse transcription as described in 2.4.9.2 on page 71 as well as for the first PCR reaction. P39

PCR parameter	modifications	
	DOP	3'RACE
PCR reaction mixture:		
cDNA synthesis	4 μg RNA from shoot tissues, reversely transcribed with oligo(dT) and M-MuLV RT at 42°C	
forward primer	P02	P33
reverse primer	P04	oligo(dT)
Mg^{2+} concentration	2.0 mM	1.6 mM
DNA polymerase	Invitrogen *Taq*	
PCR programme:	constant	constant
annealing	55°C 1 min	68.5°C 1 min
elongation	72°C 2 min	72°C 2 min
resulting fragment:	SoS650-5	SoS3'r-9

Table 3.16: DOP- and 3'RACE PCR modifications for clone SoScsu-AHAShyp differing from the standard methods described in section 2.4.9

primer	applied to	design based on sequence	sequence	T_m[°C]
P36	3'RACE	SoS650-5	5'-dGCA GCT TCC TCA TGA ACG TGC AGG AAT T-3'	66.6
P38	5'RACE as GSP1 and for RT	SoS650-5	5'-dAAA CGA CAA CGG GTA TTT CAC TTT-3'	57.6
P39	5'RACE as GSP2	SoS650-5	5'-dATT CAG CCC CGC CAA AGA TAA CTT AAT AT-3'	62.4
P41	5'RACE as GSP1 and for RT	SoS5r500-2	5'-dAAG ACC CCA AAC CCA TCA AAG TA-3'	58.9
P42	5'RACE as GSP2	SoS5r500-2	5'-dTCC CCG TAA GCT CAA CAA ACC TCT TCA A-3'	65.1
P43	5'RACE as GSP3	SoS5r500-2	5'-dAAA GCA CCG GCT TTT TTG CCT CAG AAA T-3'	63.7

Table 3.17: 3'RACE and 5'RACE PCR gene specific primer for the identification of cDNA ends of clone SoScsu-AHAShyp (P36 = 3'RACE primer, P38 to P43 = 5'RACE primers): application of the primer, sequence according to which the primer is based, primer sequence, and predicted melting temperature.

was used for the nested PCR reaction (see 2.4.9.3) and was positioned 228 bp downstream of P02.

Template was obtained from 4.3 μg freshly harvested total RNA from leaves, this time reversely transcribed by eAMV RT (section 2.4.8 on page 65) at 65°C primed with P38. Further details of this first 5'RACE procedure is detailed in Table 3.18 on page 125. Although fragments of 1,100 bp were expected, the PCR resulted in significantly shorter bands. A 471 bp large band was cloned in vector pCR 2.1-TOPO and sequencing of the resulting clone SoS5'r500-2 showed the expected overlap with the internal fragment SoS650-5 of 210 bp. The remaining new 290 bp did not provide a putative ATG initiation codon but were used to design a new set of gene specific 5'RACE primers P41, P42, and P43 (see Table 3.17 on page 124). Primer P41 was used for reverse transcription and was positioned 28 bp upstream of P02, while P42 and P43 were located 64 bp and 117 bp, respectively, upstream of P02. Hence, PCR-products, that were able to complete the ORF, should display a length of 900 bp. Conditions for reverse transcription were altered as well: 2.6 μg of total RNA were transcribed with primer P41 and reverse transcriptase SSIII at 55°C. Modifications of this second 5'RACE procedure are summarized in Table 3.18 on page 125. The resulting PCR-product was cloned in vector pGEM T Easy, the clone was named SoS5r6jes and sequencing revealed a 830 bp large fragment that formed the expected overlap with sequence SoS5'r500-2, but still did not comprise the putative ATG initiation codon.

Compared to the root-derived sequence (see below, section 3.1.3.2, page 125), there

PCR parameter	modification			
PCR reaction mixture:	first 5'RACE		second 5'RACE	
	first PCR	nested PCR	first PCR	nested PCR
cDNA synthesis	4.3 μg RNA from leaves, reversely transcribed with P38 and eAMV RT at 65°C		2.6 μg RNA from leaves, reversely transcribed with P41 and SSIII at 55°C	
forward primer	AAP	AAP	AAP	AAP
reverse primer	P38	P39	P42	P43
PCR programme:	constant	constant	constant	constant
annealing	55.5°C 1 min	61.8°C 1 min	60°C 3 min	57°C 2 min
elongation	72°C 2 min	72°C 2 min	68°C 20 min	72°C 5 min
resulting fragment:		SoS5'r500-2		SoS5r6jes

Table 3.18: 5'RACE PCR modifications for clone SoScsu-AHAShyp differing from the standard methods described in section 2.4.9

were still 70 bp missing to finish the ORF in 5'direction. At this point of the work, it was already clear that heterologous expression did not demand the N-terminally located signal sequence encoded in the missing fragment. Hence, it was decided to focus on trials for heterologous expression and to leave the identification of the 5'end aside.

As shown in Figure 3.10 on page 126, four fragments formed the 5'-unfinished clone SoScsu-AHAShyp: SoS650-5 served as internal fragment, SoS3'r-9 as 3'end, both SoS5'r500-2 and SoS5r6jes as 5'end. Clone SoScsu-AHAShyp comprised an ORF of 1,929 bp coding for 642 amino acids.

3.1.3.2 AHAS Catalytic Subunit of *Symphytum officinale* from Root-Derived cDNA (SoWcsu)

Internal fragment of clone SoWcsu

Plants from *Symphytum officinale* growing in solid medium developed white roots that apparently contained no chloroplasts. Hence, there was no necessity to cultivate chloroplast-free root-tissue cultures (RTC) from *Symphytum officinale* as described in section 2.2.1 on page 33 for *Eupatorium cannabinum*.

An amount of 1.2 μg total RNA harvested from roots was reversely transcribed with SSIII at 55°C as described in section 2.4.8. DOP-PCR was carried out as summarized

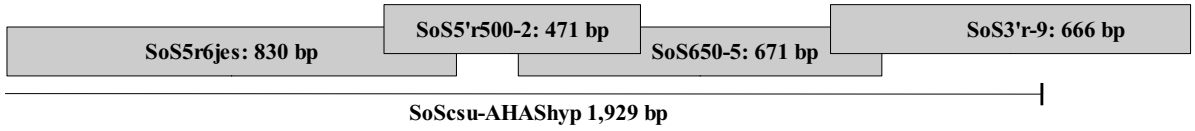


Figure 3.10: Clone SoScsu-AHAShyp is composed of four separately detected parts: SoS650-5 as internal fragment, SoS3'r-9 as 3'end, and both, SoS5'r500-2 and SoS5r6jes as 5'ends. The line resembles the length of the ORF starting without a putative ATG initiation codon and ending with the stop codon.

in Table 3.19. Interestingly, PCR products were obtained with the degenerate forward primer P01 used in a concentration of 200 pmol/μl that is a 100 times higher than usual. The resulting PCR product was cloned in vector pCR 2.1-TOPO and sequencing showed high similarity to catalytic subunits from AHAS. The fragment was designated SoW2-12 and had a length of 1,108 bp.

3'end and 5'end of clone SoWcsu

Three gene specific primers were designed according to the internal fragment SoW2-12 (sequences are shown in Table 3.20 on page 127). Primer P37 was positioned 166 bp upstream of the degenerate primer P04 and was used in 3'RACE procedure. Primer P44 was used for reverse transcription as well as in the first PCR of 5'RACE procedure. Primer P45 was located 226 bp downstream of degenerate primer P02 and was used in the nested PCR.

Details on the 3'RACE procedure are given in Table 3.19. Sequencing of the cloned PCR-product revealed a 641 bp large fragment and was named SoW3r-6, which comprized the expected overlap with the internal fragment SoW2-12, 466 bp within the ORF including the translation termination codon, and an 3'UTR of 175 bp comprizing the poly-A tail.

The 5'RACE procedure was carried out as described for the 5'end from *Parsonsia laevigata* in section 3.1.2 with fragment PlS5r149A15 on page 119 in Table 3.15 except for modifications that are summarized in Table 3.21. The resulting 5'end had a length of 888 bp large and was named SoW5r150-4. It showed the expected overlap with the internal sequence, contained an ORF of 810 bp that included the putative ATG initiation codon and a 5'UTR of 78 bp.

The hypothetical full-length sequence of root tissues from *Symphytum officinale*, SoWcsu-AHAShyp, was composed of three fragments: the internal fragment SoW2-12, the 3'end SoW3r-6, and the 5'end SoW5r150-4. They formed together an ORF of 2,013 bp coding for 670 amino acids (see Figure 3.11 on page 129).

PCR parameter	modifications	
	DOP	3'RACE
PCR reaction mixture:		
cDNA synthesis	1.2 μg RNA from root tissues, reversely transcribed with oligo(dT) and SSIII at 55°C	1 μg RNA from root tissues, reversely transcribed with oligo(dT) and SSII at 50°C
forward primer	P01 [200 pmol/μl]	P37
reverse primer	P04 [20 pmol/μl]	oligo(dT)
Mg^{2+} concentration	1.6 mM	2.5 mM
DNA polymerase	Invitrogen *Taq*	Accu *Taq*
PCR programme:	touch down	constant
	cycling no.1 repeated 15x:	
annealing	57°C → 50°C	63.6°C 1 min
elongation	72°C 3 min	72°C 5 min
	cycling no.2 repeated 20x:	
annealing	50°C 1 min	
elongation	72°C 10 min	
resulting fragment:	SoW2-12	SoW3r-6

Table 3.19: DOP- and 3'RACE PCR modifications for clone SoWcsu-AHAShyp differing from the standard methods described in section 2.4.9

primer	applied to	design based on sequence	sequence	T_m[°C]
P37	3'RACE	SoW2-12	5'-dATT GCC TGC AGC TAT GGG TGC TAT TAT A-3'	63.7
P44	5'RACE as GSP1 and for RT	SoW2-12	5'-dCAA CCT GGA GAT ATA ACC ATA TA-3'	55.3
P45	5'RACE as GSP2	SoW2-12	5'-dTGG TTC ATC CCA ATT AGG ACC CCA ATT A-3'	63.7

Table 3.20: 3'RACE and 5'RACE PCR gene specific primer for the identification of the cDNA ends of clone SoWcsu-AHAShyp (P37 = 3'RACE primer, P44 and P45 = 5'RACE primers): application of the primer, sequence according to which the primer is based, primer sequence, and predicted melting temperature.

PCR parameter	modifications	
	5'RACE	
PCR reaction mixture:	first PCR	nested PCR
cDNA synthesis	4 μg RNA from root tissues, reversely transcribed with P44 and SS III at 50°C	
forward primer	AAP	AAP
reverse primer	P44	P45
template	1 μl	2 μl of first PCR diluted 1:100
PCR programme:	constant	constant
annealing	50.1°C 2 min	63.4°C 2 min
elongation	68°C 6 min	68°C 10 min
resulting fragment:		SoW5r150-4

Table 3.21: 5'RACE PCR modifications for clone SoWcsu-AHAShyp differing from the standard methods described in section 2.4.9

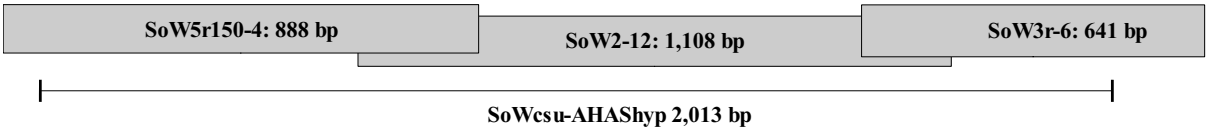


Figure 3.11: Clone SoWcsu-AHAShyp is composed of three separately detected parts: SoW2-12 as internal fragment, SoW3r-6 as 3'end, and SoW5r150-4 as 5'end. The line resembles the length of the ORF starting with the putative ATG initiation codon and ending with the stop codon.

3.1.4 Summary of Results from the Identification of cDNAs Coding for Putative AHAS

For a better overview, all full-length cDNAs coding for putative AHAS, described in this work, are summarized in Table 3.22 on page 130.

3.2 Characterization of cDNAs Coding for Putative AHAS

In this section, the novel cDNAs, coding for putative catalytic and regulatory AHAS subunits from pyrrolizidine alkaloid (PA) producing plants, were examined more closely in respect to their characteristics and their identity with all so far known plant AHASs. The patterns of transcription were determined from selected tissues applying the method of RT-PCR (see section 2.4.9.6 on page 73). This should help to predict whether one of the three putative catalytic subunits from *Eupatorium cannabinum* is likely to be involved in PA biosynthesis.

In preparation of heterologous expression, the length of signal peptides was determined with different methods to find the correct start of the mature protein (section 3.2.4), while functionality of the putative catalytic AHAS was proven *in vivo* by complementing *E. coli* mutant MF2000 strain devoid of AHAS activity (section 3.2.5), and *in vitro* by cell-free expression (section 3.2.6).

3.2.1 Sequence Characteristics

All six catalytic subunits from PA-producing plants, discovered as described in section 3.1, showed the same functional essential and, therefore, highly conserved residues as known for *Arabidopsis thaliana* (Duggleby and Pang 2000, McCourt and Duggleby 2006) and as already indicated in section 1.1.2.3 (residue numbering according to *Arabidopsis thaliana* AHAScsu as proposed by Tranel and Wright 2002): 1. In the catalytic center the expected glutamate was found (E144). 2. For substrate recognition, the same residues R377, F206, and M351 were discovered which are essential for interaction with the carboxylate group of the ketoacid and the maintenance of the correct orientation for ionic interaction. 3.

name of hypothetical full-length sequence	ORF bp (aa)	partial sequences: 5'end / internal part / 3'end	tissue of discovery	5'UTR bp	3'UTR bp from stop to poly-A
EcScsu-AHAShyp	1,953 (650)	EcS5race80 / EcSA650alt6 / EcS3race2	shoot, root[1]	2	117
EcW2csu-AHAShyp	1,971 (656)	EcW5r3.2W2.15 / - / EcW3race3SP	root	8	112
EcW3csu-AHAShyp	1,974 (657)	EcW5r3.2W3.19 / - / EcW3race7SP	root	52	129
PlScsu-AHAShyp	1,998 (665)	PlS5r149A15 / PlSmitte23 / PlS3r4	shoot, root	105	160
SoScsu-AHAShyp	1,929 (642)	SoS5'r500-2 + SoS5r6jes / SoS650-5 / SoS3'r-9	shoot	- [2]	230
SoWcsu-AHAShyp	2,013 (670)	SoW5r150-4 / SoW2-12 / SoW3r-6	root	78	138
EcSrsu-AHAShyp	1,431 (476)	EcS03_5R_05 / (EcS03_M_17) / EcS03_3R_07	shoot	11	157
EcWrsu-AHAShyp	1,461 (486)	EcW01_5R_25 / EcW01_M_12 / EcW01_3R_04	root, shoot[1]	61	198

[1] results not shown, [2] incomplete sequence

Table 3.22: Summary of all hypothetical full-length cDNAs determined in this work. ORF = open reading frame, UTR = untranslated region, bp = base pairs, aa amino acids

Residues W574 and M570 were found as well which are responsible for the recognition of the second substrate. 4. The thiamine diphosphate co-factor binding motif GDG X_{24} NN (Hawkins et al. 1989) showed that in the stretch X_{24} 22 out of 24 amino acids were identical. The thiamine diphosphate co-factor binding motif was detailed as GDG SFLMNIQELATIR A/V/S ENLPVK M/L NN (differences are underlinded). 5. The V-conformation was imposed on the thiamine diphosphate with the help of the expected M513. 6. Mg^{2+} which anchors the thiamine diphosphate was connected to the enzyme via residues D538, N565, and H567. 7. The FAD was anchored with G350 and W491. Interestingly in sequence EcW2csu, one of the three Mg^{2+}-anchoring residues was muted from histidine to proline (H567Y). 8. Moreover, all six sequences identified in this work provided N-terminally a putative chloroplast-targeting transit peptide (cTP) which will be discussed later in more detail in section 3.2.4 on page 137.

Both regulatory subunits showed the same major sequence features as already known for the regulatory subunit from *Arabidopsis thaliana* (Lee and Duggleby 2001): The overall structures started with a putative cTP (detailed in section 3.2.4 on page 142) and proceeded with the first domain providing the putative leucine binding site that is connected via a non-conserved linker of 25 amino acids to the second domain. This second domain is a repeat of the first domain but confers the valine/isoleucine binding site for feed back inhibition.

3.2.2 Comparison of Plant AHAS Primary Structure

For AHAS primary structure comparison, cDNA sequences were translated into amino acid (aa) sequences. Only sequences providing a complete ORF were considered. Identities were determined with the help of the bestfit programme within the GCG Wisconsin Software Package (see page 56).

AHAS is an enzyme of primary metabolism catalyzing the first step in the branched chain amino acid biosynthesis and shows highly conserved regions. The average level of identity was 74% among AHAScsu sequences of Angiosperm species which is in accordance with the average identities from other enzymes of the branched chain amino acid biosynthetic pathway, namely the threonine deaminase (EC 4.3.1.19) with 70%, the ketol-acid reductosiomerase (EC 1.1.1.86) with 85%, and the dihydroxy-acid dehydratase (EC 4.2.1.9) with 82% (for details on the branched chain amino acid biosynthesis, see Figure 1.9 on page 29).

Species with more than one known AHAS catalytic subunit (Table 3.23) displayed an even higher percentage of identity between the paralogous sequences. Here, the average identity was 83%.

In the first four rows in Table 3.24, the novel AHAScsu from PA-producing plants were compared among each other. Paralogous sequences from the species *Eupatorium cannabinum* (EcScsu/EcW2csu and EcScsu/EcW3csu: 81%) had a significant higher identity to each other than sequences from different species (69% - 75%). Surprisingly, the two

Species with two or more AHAScsu	% Identity (aa)
Bna 1 : Bna 3	80
Cmi 1 : Cmi 2	98
Ghi 1 : Ghi 2	97
Han 1.1 : Han 2.1	96
Nta 1 : Nta 2	98
Osa 1 : Osa 3	99
Sju 1 : Sju 2	99
Xsp 1 : Xsp 2	99
Zma 1 : Zma 2	95
EcScsu : EcW2csu	81
EcScsu : EcW3csu	81
EcW2csu : EcW3csu	74

For explanation of species abbreviations, see Table 3.24

Table 3.23: Identities on amino acid level of AHAS-csu from species with two or more AHAScsu.

root-derived sequences from *Eupatorium cannabinum*, EcW2csu and EcW3csu showed a remarkable low degree of identity (74%).

Additionally, sequences from the database of non-PA-producing plants were compared to the new sequences from PA-producing plants (see Table 3.24). Sequence EcW3csu (Asteraceae, tribe Eupatorieae) showed a high identity towards the *Xanthium* sequence (90%) that is a member of the family of Asteraceae, tribe Heliantheae. A modified amino acid sequence may enable an altered substrate specificity. Thus, EcW3csu might be a good candidate to be involved in PA biosynthesis. Alternatively, sequence EcW3csu might be probably the less modified sequence, whereas sequences EcScsu and EcW2csu are then likely to play the predicted role in PA biosynthesis.

3.2.3 Tissue-Specific Transcript Localization of AHAS with RT-PCR

Tissue-specific transcript localization was used to determine the type of tissue in which the mRNA of interest is located. The mRNA is regarded as a marker for protein biosynthesis and tissue-specific transcript localization was performed to get an additional hint which one of the three discovered AHAS cDNA sequences of *Eupatorium cannabinum* was a likely candidate to code for an enzyme involved in PA biosynthesis.

As mentioned in section 1.2, the first step in necine base biosynthesis is catalyzed by the homospermidine synthase (HSS) which was immuno-localized by Anke (2004) for *Eupatorium cannabinum* in the cortical parenchyma of freshly sprouted, earth grown roots. The biosynthesis of the necic acid was postulated to be co-localized with the biosynthesis

% Identity (aa)	PA-producing plants				
	EcScsu	EcW2csu	EcW3csu	SoWcsu	PlScsu
EcW2csu	81	100			
EcW3csu	81	**74**	100		
SoWcsu	75	69	75	100	
PlScsu	75	74	81	72	100
Amy	72	66	75	73	74
Apo	79	73	82	73	84
Are	78	73	82	73	84
Asp	81	73	83	73	82
Ath	76	70	78	71	75
Atu	79	72	82	79	84
Bna 1	67	66	72	68	72
Bna 3	79	74	84	79	82
Bsc	79	74	81	72	78
Cmi 1	75	69	82	72	75
Ghi 1	77	70	80	74	77
Han 1.1	81	74	89	79	83
Lmu	70	63	73	72	71
Mtr	76	70	78	74	77
Nta 1	77	71	81	76	82
Osa 1	73	67	74	73	74
Prh	75	72	82	72	75
Sju 1	70	65	72	73	70
Str	70	66	73	74	71
Xsp 1	81	75	**90**	76	80
Zma 1	70	67	72	73	71

For NCBI accession numbers of amino acid sequences
see appendix Table B.6 on page 223

Table 3.24: Identities on the amino acid level of the novel AHAScsu from PA producing plants compared to non-PA-producing plants (bold numbers are mentioned in the text). csu = catalytic subunit, EcS = *Eupatorium cannabinum* shoot-derived, EcW2 = *Eupatorium cannabinum* root-derived, seq. no. 2, EcW3 = *Eupatorium cannabinum* root-derived, seq. no. 3, all three of them are Asteraceae, PlS = *Parsonsia laevigata* shoot derived, Apocynaceae, SoW = *Symphytum officinale* root derived, Boraginaceae, Amy = *Alopecurus myosuroides*, Poaceae, Apo = *Amaranthus powellii*, Are = *Amaranthus retroflexus*, Asp = *Amaranthus* spec., all Amaranthaceae, Ath = *Arabidopsis thaliana*, Brassicaceae, Atu = *Amaranthus tuberculatus*, Amaranthaceae, Bsc = *Bassia scoparia*, Chenopodiaceae, Bna = *Brassica napus*, Cmi = *Camelina microcarpa*, both Brassicaceae, Ghi = *Gossypium hirsutum*, Malvaceae, Han = *Helianthus annuus*, Asteraceae, Lmu = *Lolium multiflorum*, Poaceae, Mtr = *Medicago trunculata*, Fabaceae, Nta = *Nicotiana tabacum*, Solanaceae, Osa = *Oryza sativa*, Poaceae, Prh = *Papaver rhoeas*, Papaveraceae, Str = *Sagittaria trifolia*, Sju = *Schoenoplectus juncoides*, both Poaceae, Xsp = *Xanthium* spec., Asteraceae, Zma = *Zea mays*, Poaceae

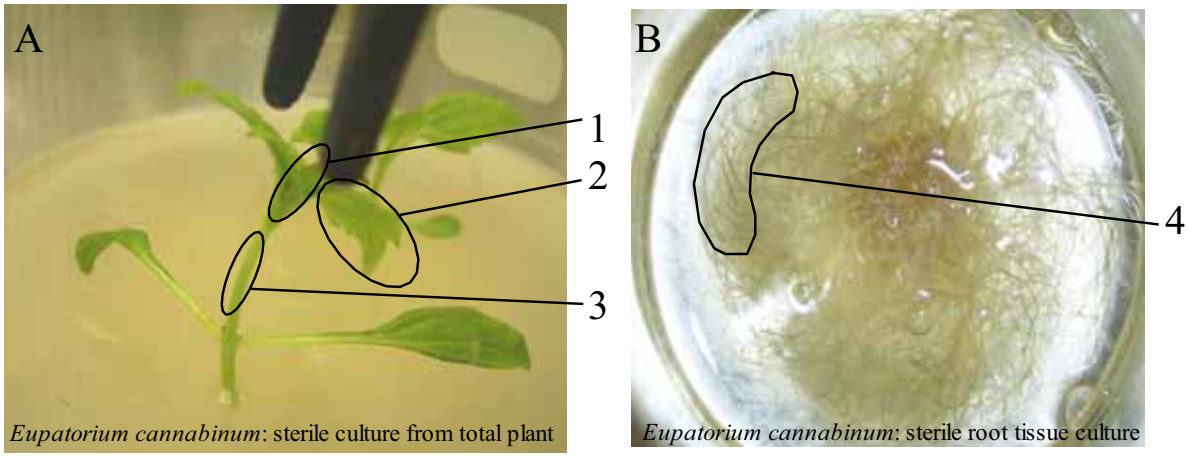


Figure 3.12: Tissues for tissue-specific transcript localization with sterile cultures from *Eupatorium cannabinum*. Panel A: plant in solid medium, panel B: root tissue culture in liquid medium. Total RNA was harvested from 1 = shoot tips, 2 = leaves, 3 = stipes, and 4 young roots.

of the necine base, as one moiety of necine base is esterified with one or more moieties of necic acids to form a complete PA. In case of lycopsamine type PA, the special lycopsamine type necic acid is believed to be produced by a modified AHAS (secondary metabolism). Hence, the place of protein biosynthesis for the modified AHAS was considered to be in the roots as well as it is for the HSS allowing an efficient biosynthesis of PAs.

Experimental set-up

Total RNA was isolated from four different tissues of *Eupatorium cannabinum* sterile cultures, as described in section 2.4.3.1 on page 58. While shoot tips (1), leaves (2), and stipes (3) were taken from one sterilely grown plant (Figure 3.12 panel A on page 134), the root tissue (4, panel B) was taken from young parts of a sterile root-tissue culture (see section 2.2 on page 33) to exclude artefacts produced from chloroplastidal nucleotides.

Reverse transcription (see section 2.4.8 on page 65) of each tissue was performed with SuperscriptIII at 55°C with 1 μg total RNA and the oligo(dT)-primer. 3 μl of the resulting cDNA were used as template in a standard PCR reaction mixture with Accu*Taq* DNA polymerase (see section 2.4.9). With the help of primers annealing to different stretches, the three sequences of catalytic subunits and the two sequences of regulatory subunits were amplified selectively. PCR cycling was performed with a constant PCR programme as described on page 67, modified in the annealing procedure that was carried out for 2 minutes and in the elongation procedure that took 10 minutes at 72°C. After 25, 30, 35, and 40 cycles, 5 μl were taken from the PCR reaction mixture and analyzed on an agarose gel as shown in Figure 3.13. In each of these gels, columns 1 to 4 show the RT-PCR product from tissues 1 to 4, while primer specificity is shown in columns 5 to 7,

controlled in a PCR reaction with a full-length cDNA as template. A negative control, shown in column 8, contained water instead of cDNA.

RT-PCR for EcScsu-AHAShyp

With P53 (see Table 3.35 in section 3.4.1.4) as forward primer, P14 (see Table 3.4 on page 105) as reverse primer, and an annealing temperature of 64°C, a fragment of 1,521 bp revealed the site of transcription for sequence EcScsu-AHAShyp. As shown in Figure 3.13, EcScsu-AHAShyp transcripts were identified in all analyzed tissues (columns 1, 2, 4) except for the stipes (column 3). Column 5 shows the PCR-product with plasmid EcScsupTYB1 (cloned in section 3.4.1.4) that functioned as a positive control for the primer combination P53 and P14. In columns 6 and 7, the utilized templates were plasmid EcW2csupET3a (cloning is described in section 3.4.1.1) and plasmid EcW3csupET22b (cloned in section 3.4.1.2), respectively. Here (in columns 6 and 7), the lack of any PCR-product proved the expected primer specificity.

RT-PCR for EcW2csu-AHAShyp

For the pair of primers P48 and P49 (see Table 3.29), the annealing temperature was determined at 55°C as shown in section 3.4.1.1. As expected, the PCR product had a size of 1,773 bp. The positive control in this trial was plasmid EcW2csupET3a (column 6 in the lower panel of Figure 3.13). In Figure 3.13, column 4, the occurrence of sequence EcW2csu-AHAShyp was shown to be restricted to the roots. This finding is very remarkable since AHAS genes are usually expressed constitutively and in all tissues as housekeeping genes. In *Brassica napus*, mRNA of functionally distinct AHAS genes was detected only in reproductive tissues (Ouellet et al. 1992). This might serve as an example that a tissue-specific abundance of AHAS mRNA occurs simultaneously with an altered functionality of AHAS genes as discussed in more detail in section 4.1.3 on page 185 making this sequence a candidate for PA-specific AHAS.

RT-PCR for EcW3csu-AHAShyp

The pair of primers P61 and P62 (see Table 3.25) required an annealing temperature of 55°C for detecting the fragment of 1,750 bp from clone EcW3csu-AHAShyp. The transcription patterns were the same as found for clone EcScsu-AHAShyp (thus not shown), but with a reduced intensity. Primer specificity was demonstrated in the absence of PCR-products with plasmids EcScsupTYB1 and EcW2csupET3a while the positive control resulted in the expected PCR product. It was estimated that EcW3csu-AHAShyp was of low abundance as EcW2csu-AHAShyp, but EcW3csu-AHAShyp was not restricted to any tissue.

RT-PCR for the regulatory subunits EcSrsu- and EcWrsu-AHAShyp

EcSrsu-AHAShyp was amplified with the pair of primers P75 and P74 (see Tables 3.25 and 3.41 on pages 137 and 173) and PCR resulted in fragments of 1,420 bp, while EcWrsu-AHAShyp was detected with primers P76 and P77 (see Table 3.41) and produced a PCR product with 1,460 bp in size. The two PCR reactions were carried out with an annealing temperature of 55°C. The two regulatory subunit sequences were transcribed in the same

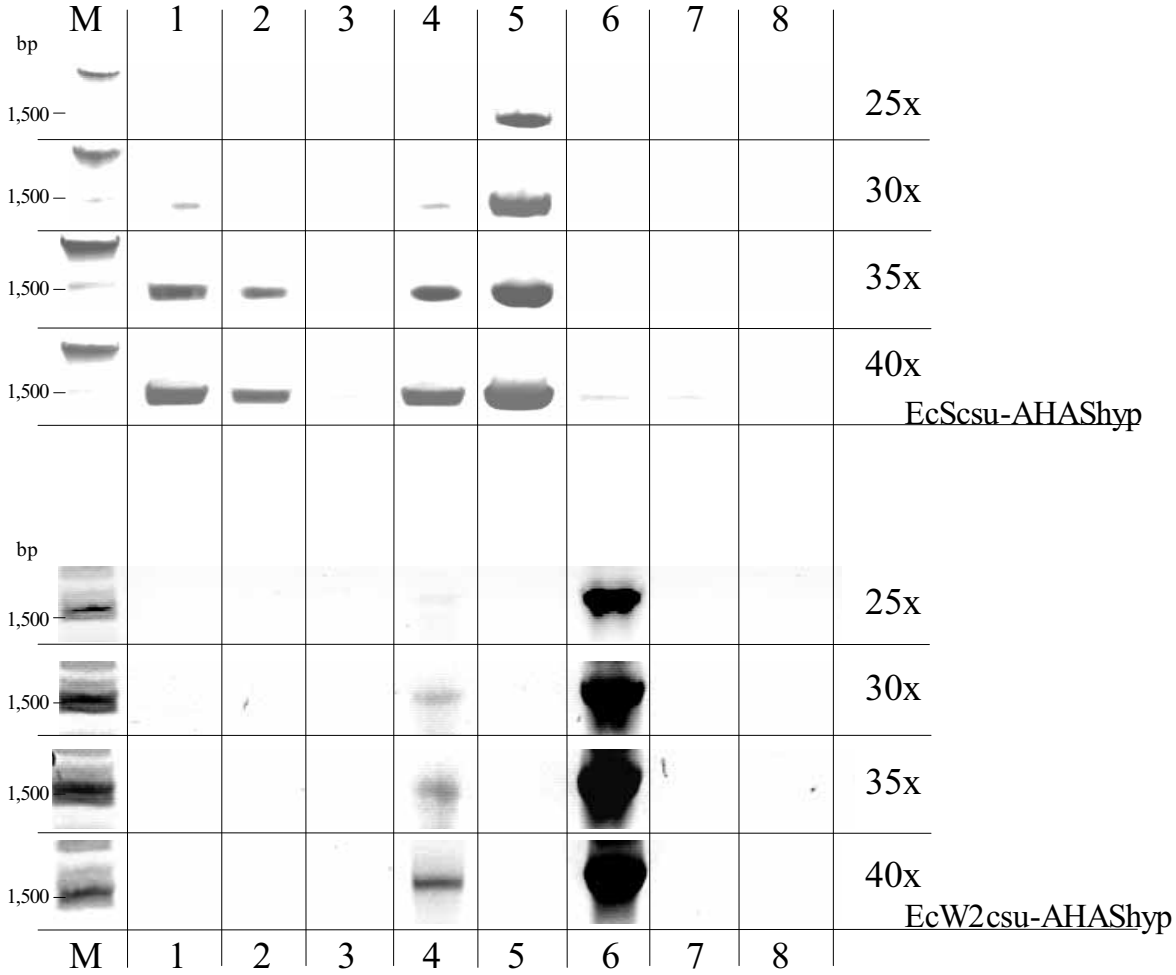


Figure 3.13: Agarose gels (1.5%, stained with ethidium bromide) for RT-PCR analysis after 25, 30, 35, and 40 cycles: RT-PCR for clone EcScsu-AHAShyp (above) and for clone EcW2csu-AHAShyp (below). M = 100 bp ladder, 1 = shoot tips, 2 = leaves, 3 = stipes, and 4 young roots, 5 = EcScsu-AHAShyp positive control, 6 = EcW2csu-AHAShyp positive control, 7 = EcW3csu-AHAShyp positive control, 8 = negative control (water).

primer	type	sequence	T_m [°C]
P61	forward	5'-dTAT AGA GCT CAC CAG AGA TAT TCG TTT CCC GAT TCG CCC AA-3'	72.4
P62	reverse	5'-dTAT CTC GAG ATA CTT CGT TCT ACC ATC ACC CTC GGT GAT CAC-3'	73.3
P75	forward	5'-dTAT ATA CGT CTC AAA TGG CGG CCG TCG TAT CAA-3'	68.2

Table 3.25: Gene specific primers P61 and P62 applied in RT-PCR for sequence EcW3csu-AHAShyp, genespecific primer P75 applied in RT-PCR for sequence EcSrsu-AHAShyp.

tissues as found for EcScsu-AHAShyp (therefore not shown) except for EcWrsu-AHAShyp being transcribed in the stipes as well. Primer specificity and positive controls worked as predicted with plasmids EcSrsupIBA7 and EcWrsupIBA3C (constructed as described in section 3.4.3.1 on page 173). It was concluded that the transcription of both putative regulatory subunits was not restricted to any tissue. Up to now, it is not known which combination of catalytic and regulatory subunits does exist *in vivo*. This result makes no indications in favor or against any possible combination.

Conclusions from primary structure comparison and tissue-specific transcript localization

In order to make a prediction whether one of the three putative catalytic subunits from *Eupatorium cannabinum* (EcScsu-AHAShyp, EcW2csu-AHAShyp, and EcW3csu-AHAShyp) was likely to be involved in PA biosynthesis, the two applied methods of primary structure comparison and tissue-specific transcript localization led to different results. In primary structure comparison, sequence EcW3csu-AHAShyp showed a lower degree of identity to AHAS amino acid sequences from the same species (i.e. EcScsu-AHAShyp 81%, EcW2csu-AHAShyp 74%) than to those of other species (e.g. *Xanthium* spec. 90%). Furthermore, the tissue-specific transcript localization of sequence EcW2csu-AHAShyp turned out to be restricted to the roots, as predicted from the findings for the PA-specific homospermidine synthase, making sequence EcW2csu-AHAShyp a promising candidate for PA-specific AHAS.

3.2.4 Determination of Signal Peptides

In the following section, it was necessary to insert a comprehensive overview about existing reports that presented experiences with heterologous expression. These reports were taken into account for the decision whether to include or to omit the chloroplast-targeting signal peptide to obtain an active recombinant enzyme.

Signal peptides had been first discovered in the early 1970s in secreted proteins that had been co-translational translocated across the endoplasmic reticulum (ER) membrane prior to their transport to the Golgi apparatus. The deduced signal hypothesis had

postulated that an N-terminal leader peptide serves as a targeting peptide that directs the growing polypeptide chain to the ER membrane and is then cleaved off by a special protease in the ER membrane even before the protein is completed. In contrast to this co-translational translocation of proteins across the ER membrane, proteins targeting the chloroplasts and mitochondria are translocated post-translationally after the protein folded completely. The protruding signal peptide (highlighted in red in Figure 3.14 on page 139) contacts the signal peptide recognition site (S, Figure 3.14) of an integral membrane protein (IMP, Figure 3.14) that is located at a special contact zone where inner and outer membranes of the chloroplast are joined (Pain et al 1988). The IMP allows the protein to cross the two membranes at once, but needs to be unfolded. Hence, a common mechanism of translocation across the chloroplast and mitochondria double membranes is postulated since all unfolded proteins show the same linear conformation. Unfolding is performed by an energy-driven protein pump (hsp70, Figure 3.14) that forces the entire protein through the membrane, transiently unfolding the polypeptide chain. In yeast, this protein-pump has been found to belong to the hsp70 stress-response proteins, which resembles the same subclass as the chaperones DnaJ, DnaK, and GrpE used in the expression experiments (section 3.4.1.5, page 164). After import to the stroma-side, a special signal peptide protease (P, Figure 3.14) cleaves off the signal peptide and the protein refolds. Hence, there are two recognition sites involved in the translocation process: a signal peptide and a signal peptide protease cleavage site.

Signal peptides of AHAS catalytic subunits

Plant and fungal sequences of AHAS catalytic subunits contain signal peptides for targeting the nuclear encoded enzymes to the appropriate organelle. Fungal AHAS activity had been localized in the mitochondria matrix by Cassady et al. (1972) in *Neurospora crassa* and by Ryan and Kohlhaw (1974) in *Saccharomyces cerevisiae*, while plant AHAS activity had been entirely found in the chloroplast stroma of *Pisum sativum* by Jones et al. (1985). Since transit peptides are cleaved off *in vivo* after translocation in eukaryotes and, moreover, bacterial hosts are not able to process eukaryotic proteins correctly, the length of an N-terminally located chloroplast-targeting transit peptide was determined in order to position the primers for heterologous expression in the way that a completely mature protein was produced in the bacterial host. Nevertheless, it was not clear whether omitting the signal peptide in heterologous expression was useful or not, as the literature provided evidence for both:

Wiersma et al. (1990) had been able to functionally complement AHAS devoid mutants from *Salmonella typhimurium* with plant AHAS from *Brassica napus* in omitting the transit peptide coding sequences (start of translation at S65 is highlighted in red in Figure 3.15 on page 140 in line 10), while further trials with truncated and full-length copies of the plant AHAS genes had resulted in inactive (M1, K16, R73, and L79 highlighted in turquoise) or less active enzymes (R52 with 90% enzyme activity, A45 with 50% enzyme activity, highlighted in red).

On the other hand, Bernasconi et al. (1995) had shown that heterologous expression of AHAS from *Xanthium* spec. in *E. coli* omitting the signal sequence (start of

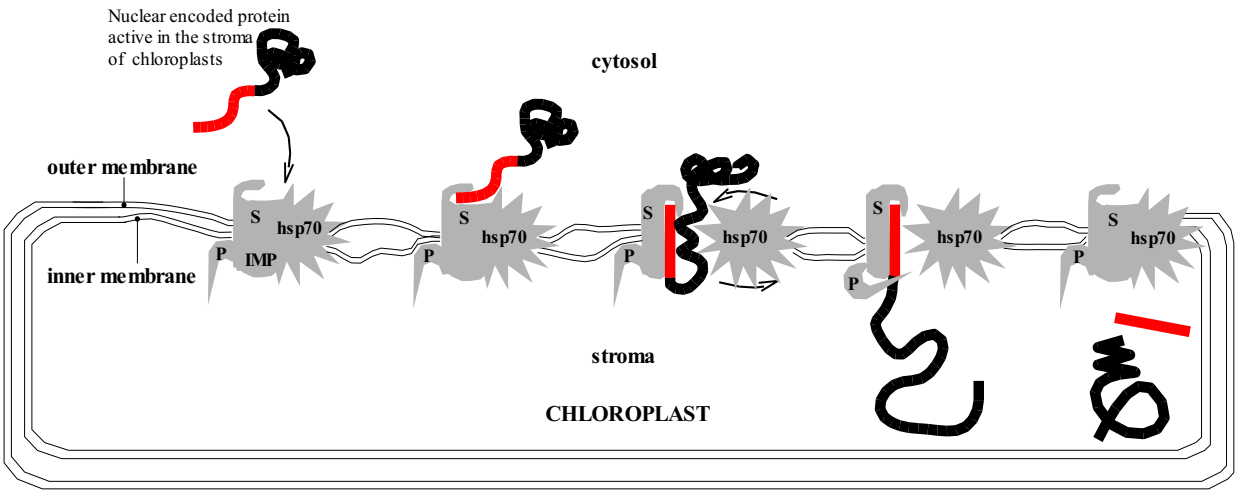


Figure 3.14: Translocation of nuclear encoded proteins across the chloroplast membranes into the stroma. An integral membrane protein (IMP) allows the simultaneous translocation across outer and inner membrane. The IMP is composed of a signal peptide recognition site (S) to let insert the signal peptide of the protein (highlighted in red) into the IMP. Furthermore, the IMP encloses an energy-driven protein pump (hsp70) that forces the entire protein through the membrane under transient unfolding. The integrated signal peptide protease (P) cleaves off the signal peptide (scheme adopted and modified from Alberts et al. 1989).

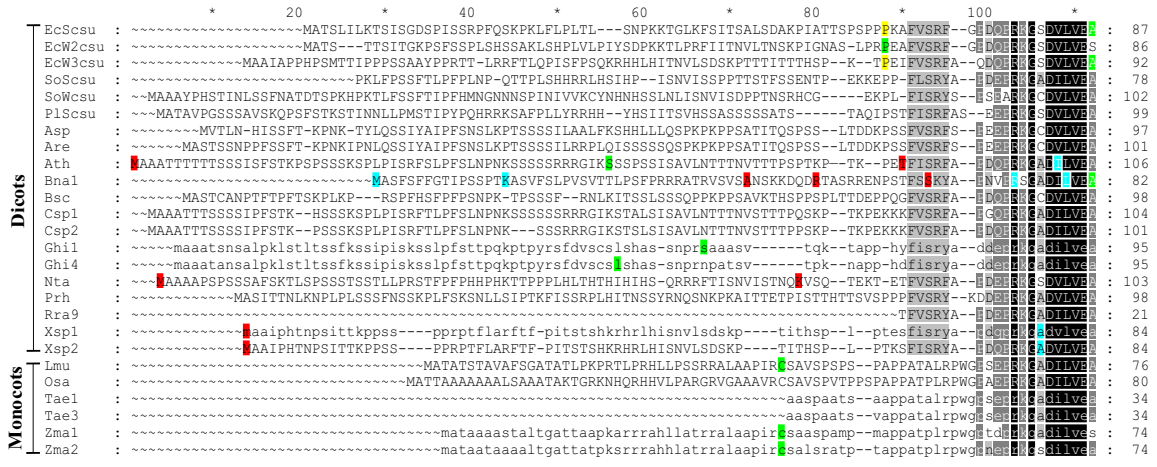


Figure 3.15: N-terminal stretch of an amino acid alignment of plant AHAS catalytic subunits. Sequences shown in lines 1 to 6 were determined in this work as described in section 3.1, sequences from line 7 to 26 were obtained from the NCBI database. Alignment parts highlighted in green show the starting amino acids of a mature protein predicted by TargetP1.1. Amino acids, highlighted in red, show the start of translation for a heterologously expressed active AHAS, while amino acids highlighted in turquoise, show the start of translation for inactive and/or insoluble proteins. The start of the heterologously expressed catalytic subunits from *Eupatorium* described in this work are highlighted in yellow. csu = catalytic subunit, dicots: EcS = EcScsu-AHAShyp, EcW2 = EcW2csu-AHAShyp, EcW3 = EcW3csu-AHAShyp, SoS = SoScsu-AHAShyp, SoW = SoWcsu-AHAShyp, PlS = PlScsu-AHAShyp, Asp = *Amaranthus* spec., Are = *Amaranthus retroflexus*, Ath = *Arabidopsis thaliana*, Bsc = *Bassia scoparia*, Bna = *Brassica napus*, Csp = *Camelina* species, Ghi = *Gossypium hirsutum*, Nta = *Nicotiana tabacum*, Prh = *Papaver rhoeas*, Rra = *Rapha raphanistrum*, Brassicaceae, Xsp = *Xanthium* species; Poaceae: Lmu = *Lolium multiflorum*, Osa = *Oryza sativa*, Tae = *Triticum aestivum*, Zma = *Zea mays*. For NCBI accession numbers of amino acid sequences see appendix Table B.6 on page 223; for one letter code of amino acids, see Table B.4.

translation at A78 is highlighted in turquoise in Figure 3.15 in lines 19 and 20) had ended up in inclusion bodies, while expression including the complete signal peptide had created a soluble and active enzyme (M1 highlighted in red).

Additionally, Hershey et al. (1999) had described two successful expressions of plant AHAS catalytic subunits. For *Arabidopsis thaliana* AHAScsu, the full-length sequence had been used (M1 highlighted in red in line 9), while for *Nicotiana plumbaginifolia* the sequence had been truncated and started with K74 (highlighted in red in line 16 within the *Nicotiana tabacum* sequence as the sequence from *Nicotiana plumbaginifolia* was not available). Despite of this, Kim and Chang (1995) had obtained an active AHAS from *Nicotiana tabacum* including the targeting peptide (M1 highlighted in red in line 16).

Hence, it was decided to obtain more information from the programme TargetP V1.1 (see section 2.4.2 on page 56), predicting types of signal sequences such as chloroplast (cTP) or mitochondria (mTP) transit peptides or signal sequences which target proteins to the secretory pathway (SP). Moreover, TargetP V1.1 supplies the user with information about the putative cleavage site of the signal sequence. The cleavage site of the sequences EcScsu-AHAShyp and EcW3csu-AHAShyp were positioned by TargetP V1.1 at amino acid E86 and E91, respectively, which put the start for the putatively processed, native protein at the same position (shown in the alignment in Figure 3.15 in lines 1 and 3, highlighted in green). A uniform start of the mature protein was expected since the signal peptide protease cleavage recognition site was hypothesized to be similar to all AHAScsu. For sequence EcW2csu-AHAShyp, the mature protein was predicted to start with amino acid P64 which did not fit into the unified picture. Thus, sequences from *Arabidopsis thaliana*, *Brassica napus*, *Gossypium hirsutum*, *Lolium multiflorum*, and *Zea mays* were submitted to the cleavage site prediction as well, but results did not provide a uniform start of the mature protein, especially not for dicotyledonous plants (see parts highlighted in green in Figure 3.15). Sequences were submitted to two more prediction programmes (SignalP V1.1 and PSORT Predicition) but results differed from those obtained by TargetP V1.1. However, the type of signal sequence was predicted uniformly by all three programmes with a cTP. Obviously, the signal peptide recognition site for chloroplasts is very well known and was integrated in the predicting programmes while the signal peptide protease cleavage site still needs to be discovered or is an unpredictable variable.

As the predicting programmes did not provide usable informations for the length of the cTPs, literature was studied once again. Chang and Duggleby (1997) had reported the heterologous expression of AHAS catalytic subunit from *Arabidopsis thaliana*. They had cloned the sequence in three different lengths: N-terminally unabridged (Ath-0), shortened by 85 amino acids (Ath-85E, start of the translation is highlighted in red in Figure 3.15 in line 9 at T86), and shortened by 101 amino acids (Ath-101D, translation start is highlighted in turquoise at I102). The specific activity of the shortened protein Ath-85E (5.9 nkat/mg) had been 3 times higher than that of the full-length protein Ath-0 (1.9 nkat/mg) while Ath-101D had been inactive. Hence, a difference of just 16 amino acids had been essential for activity.

An alignment by Duggleby and Pang (2000) with 6 plant, 2 fungal, 3 algal, and a num-

ber of bacterial AHAS catalytic subunits had shown that the cleavage site is probably close to the region where homology with the prokaryotic and algal sequences begins. For algae, it is known that the AHAS gene is located on the plastid genome (Douglas and Penny 1999) and the gene is expressed within the chloroplast and no intracellular trafficking is required. Duggleby and Pang (2000) had identified the sequence "TFXS[K/R][F/Y]AP" common to all the 6 plant AHAS. In the last years, the number of discovered AHAS sequences from plants had grown. For the dicots, the alignment in Figure 3.15 shows clearly that this common sequence (highlighted in light grey by Genedoc) nowadays shortened and modified to "F[V/I]S[K/R][F/Y][A/S]XP". For the monocots, the common sequence must be named with "APPAT[A/P]LRPWGP".

To sum up, 3 out of 4 reports had shown that heterologous expression including the signal peptide had resulted in an active enzyme (due to a very high similarity between the two *Xanthium* sequences with a difference of 1 amino acid, the *Xanthium* expression reports were counted as one single report). Furthermore, for the expression of truncated enzymes 5 out of 10 trials had resulted in an active enzyme. All inactive enzymes were due to a truncation downstream of the common dicot sequence "F[V/I]S[K/R][F/Y][A/S]XP". Moreover, Wiersma et al. (1990) and Chang and Duggleby (1997) had shown that enzyme activity decreases with larger transit peptides.

In order to obtain an optimal truncated enzyme with optimal activity, the following was decided: As shown for *Arabidopsis* and *Brassica*, the omission of 16 and 8 amino acids, respectively, has been the reason for the formation of insoluble and inactive enzymes. Thus, primers for heterologous expression (used in section 3.4) were positioned 3 amino acids upstream of the common dicot sequence, starting with P that is highlighted in yellow in the alignment: EcScsu-AHAShyp started at P65, for sequence EcW2csu-AHAShyp, P64 is already highlighted in green as advised by TargetP V1.1, and EcW3csu-AHAShyp started at P70.

With the "peptidesort" programme from the GCG Wisconsin Software Package (see page 56), features of the expected recombinant proteins were predicted. A comparison of these features is shown in Table 3.26: The isoelectric point (IEP) was shifted from basic to acidic values, when the cTP was omitted and the mass was lowered to 64 - 65 kDa which are typical dimensions for a native AHAScsu from plants (section 1.1.2.3).

Signal peptides of AHAS regulatory subunits

Signal peptides of AHAS regulatory subunits from *Eupatorium cannabinum* were determined with the help of information from the literature as well as with the help of the programme TargetP V1.1. In Figure 3.16, the amino acid sequences from the two regulatory subunits discovered in this work (lines 1 and 2) were aligned with the four to date known AHAS regulatory subunits from plants. The starting amino acids for the start of a mature protein, predicted from TargetP V1.1, were highlighted in green and showed a rather random distribution. Moreover, sequence EcSrsu-AHAShyp was strongly predicted to contain a chloroplast-targeting transit peptide, while for sequence EcWrsu-AHAShyp the existence of a cTP was predicted to be less reliable. Hence, it seemed more reasonable to relay on experimental evidence supplied with the literature. Up to now, only two

clone	length of cTP [aa]		total length [aa]	mass [kDa]	IEP [pH]
EcScsu-AHAShyp	64	full-length	650	71	7.9
		cTP omitted	586	64	6.8
EcW2csu-AHAShyp	63	full-length	656	72	7.6
		cTP omitted	593	65	6.2
EcW3csu-AHAShyp	69	full-length	657	73	7.6
		cTP omitted	588	64	6.6

Table 3.26: Comparison of AHAS catalytic subunits from *Eupatorium*: length of omitted cTP, length of enzyme with and without cTP, resulting mass and IEP.

papers are known to describe the heterologous expression and purification of plant AHAS regulatory subunits:

In 1999, Hershey and coworkers had been first to discover regulatory subunits from plants. A partial sequence from *Zea mays* served for screening the cDNA library from *Nicotiana plumbaginifolia*, where a positive clone had been completed in 5'direction with RACE technique. As shown in Figure 3.16 in line 5, the *Nicotiana* sequence did not provide an initiator ATG within the ORF. This had forced Hershey and coworkers to decide where to let start the translation during heterologous expression. Their choice of a starting site at amino acid I78 (highlighted in red) had been based on a consensus estimate with prokaryotic regulatory subunits and the decision to omit the plant transit peptide. Significant amounts of the obtained recombinant protein had been insoluble. Nevertheless, a small amount of soluble protein had been purified and successfully mixed with its corresponding catalytic subunit which had proved the functionality as discussed in more detail in section 4.1.4 on page 185.

In 2001, Lee and Duggleby had tested the expression of *Arabidopsis thaliana* regulatory subunit with four different lengths of the cTP. The level of expression for constructs comprising the cTP (M1 in line 4 highlighted in turquoise) or omitting 85 amino acids (K86) had been reported to be very low, while omission of 51 (M52, highlighted in red) or 70 amino acids (S71, highlighted in red) had resulted in increased amounts of soluble protein.

Primers for heterologous expression of the shoot-derived regulatory subunit from *Eupatorium cannabinum* EcSrsu-AHAShyp, determined in this work, were positioned at Q50 (highlighted in yellow in line 1 in Figure 3.16) in the middle of the already positively proved frame marked out from M52 to S71 in line 4 for Ath-rsu5. Expression primers for the root-derived sequence EcWrsu-AHAShyp were positioned at M1 (highlighted in yellow in line 2 in Figure 3.16) taking into account that the cTP was not reliably predicted. Moreover, it is commonly known for many multimeric proteins which are active in chloroplasts, e.g., the ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO), that one subunit is encoded in the nucleus (small subunit of the RuBisCO) and needs transfer across the chloroplast membrane, while the other is encoded in the organelle and needs no

```
                    *        20         *        40         *        60         *        80         *       100
EcSrsu   : ~~~~~~~~~~~~~~MAAVVSTYSAGLIHSRGKPSSFNLLPSSKTFDGLLSSNPKTFHFKRL--FVQASNGTVSTAVDSTSSMARP--KVRRHTISVFVGD :  82
EcWrsu   : MSSLSTAVETPIRFNNRCRVKRATGHLTNLKLLNRKCADRYNLKVVNAISDARD-GN---IVNGSPAVLPSILADSRAAAPIA---SKVKRHTISVFVGD :  93
Ath-rsu2 : MAATTTATSL---FSSRL-------HFQNQNQGYGFPAKTPNSLQVNQIIDGRKMRN---ATVLSAASTDKAITTAQSVAPTA--CDRVRRHTISVFVGD :  85
Ath-rsu5 : MAAISVSSSPSIRCLRSACSDSSPALVSS--TRVSF---PAKISYLSGISSHRGDEMGKRMEGFVRSVDGKISDASFSEASSATPKSKVRRHTISVFVGD :  95
Npl-rsu  : ~~~~~~~GFGTRAAVSTPFNSSLNKPFLSLYSSIPHLGFSQVVQFPTLKPKFLLSKIKTVVISAAKADGTYTNEDDSLTTFDSSIRQKTRRHTIQVFVGD :  93
Osa-rsu  : ~~~~~~~~~~~~~~~~~MATATAASRLAVAGAEPAR-ARRHRPTTVAVCGGARPRSRPAAVVAAAGAAAPSPATGGVAPVPPSPRGSIIKRHTISVFVGD :  82
```

Figure 3.16: N-terminal stretch from the amino acid alignment of plant AHAS regulatory subunits. Sequences shown in lines 1 and 2 were determined in this work as described in section 3.1, sequences from line 3 to 6 were obtained from the NCBI database. Alignment parts, highlighted in green, show the starting amino acids of a mature protein predicted by TargetP1.1. Amino acids, highlighted in red, show the start of translation for a soluble AHASrsu, while amino acids, highlighted in turquoise, show the start of translation for insoluble or no expressed proteins . The start of the heterologously expressed regulatory subunits from *Eupatorium*, described in this work, are highlighted in yellow. rsu = regulatory subunit, EcS = EcSrsu-AHAShyp, EcW = EcWrsu-AHAShyp, Ath = *Arabidopsis thaliana*, Npl = *Nicotiana plumbaginifolia,* Osa = *Oryza sativa.* For NCBI accession numbers of amino acid sequences, see appendix Table B.6 on page 224; for one letter code of amino acids, see Table B.4.

transfer and contains no transit peptide (Chua and Schmidt 1978). Hence, omitting a part of the sequence EcWrsu-AHAShyp would probably hinder the heterologous expression or interfere with the functionality of the protein.

3.2.5 *In vivo* Activity Test for Recombinant Plant AHAS Catalytic Subunit in *E. coli* MF2000

Functionality of the discovered cDNA encoding the putative catalytic subunit of *Eupatorium cannabinum* AHAS was tested *in vivo* by a functional complementation of the *E. coli* mutant MF2000 devoid of AHAS enzyme activity.

As reported in section 1.1.2.3 on page 23, *E. coli* owns three different AHAS, each of which is composed of catalytic and regulatory subunits. On gene *ilvBN*, AHASI is encoded with *ilvB* coding for the catalytic and *ilvN* coding for the regulatory subunit. Genes *ilvIH* and *ilvGM* code for AHASII and AHASIII, respectively. MF2000 cells are completely devoid of endogenous AHAS activity due to a three-fold mutation: AHASI is knocked out with an insertion in *ilvB* gene, AHASII is silenced by a frameshift in *ilvG*, and AHASIII has no activity due to a deletion in *ilvI*. The medium for MF2000 must supply leucine, isoleucine, and valine (for MF2000 strain description, see page 36).

Cells were made competent as described in section 2.3.3.1 on page 42. Competent cells were transformed, as described on page 43, with plasmid EcScsupTYB1 coding for EcScsu-AHAShyp (generated in section 3.4.1.4 on page 162), with vector pTYB1 as negative control, and with plasmid pET-GM as positive control (see section 2.3.4.4 on page 49) harbouring *E. coli* gene *ilvGM*. For recovery, transformed cells were grown over night at

37°C on full medium (LB) under antibiotic selection either prepared with ampicillin (100 μg/ml) for EcScsupTYB1 and pTYB1 or kanamycin (50 μg/ml) for pET-GM. Colonies from EcSpTYB1 and from pET-GM were transferred to M9-BALH plates (see section 2.3.2.1) containing glucose, vitamin B1, arginine, leucine, and histidine. Isoleucine and valine were omitted, but for a positive test on functional AHAS that was provided by the plasmids, growth on minimal medium should be enabled.

Additionally, minimal medium plates were supplemented with IPTG (24 μg/ml) to start expression from the vectors and prepared with the appropriate antibiotics as described for the LB plates. Colonies harbouring vector pET-GM did not grow, while colonies with EcScsupTYB1 were able to do so. A successfully growing colony harbouring plasmid EcScsupTYB1 was streaked out on a fresh M9-BALH plate supplied with ampicillin and IPTG (see Figure 3.17 on page 146, Panel A). On the other half of this plate (Panel B in Figure 3.17), a colony harbouring the negative control pTYB1 was transferred from full medium to minimal medium. Cells were grown for 5 days at 37°C. Colonies with EcSpTYB1 formed and proved functionality of the AHAS catalytic subunit of *Eupatorium cannabinum* in MF2000, while not complemented cells were not able to grow on minimal medium and thus showed to be devoid of any AHAS enzyme activity.

Plasmid EcScsupTYB1 was re-isolated from a liquid LB culture and submitted to PCR with the pair of primers P53/P14. PCR was carried out with *Taq* DNA polymerase as described in section 3.4.1.4. The resulting PCR product of 1,521 bp showed that plasmid EcScsupTYB1 really was transformed into MF2000 and really was responsible for functional activity.

The finding, that pET-GM coding for *E. coli* AHASII failed in enabling growth of MF2000 on minimal medium, might have been due to the supplied genetical information itself. It only codes for AHASII which prefers to contribute to isoleucine supplementation. In this experiment, firstly, the host was forced to transform glucose into pyruvate which is needed to either produce valine or leucine, and secondly, it was forced to transform glucose into 2-oxobutyrate which is used to obtain isoleucine. It might have been metabolically very "expensive" to produce 2-oxobutyrate as a substrate for AHASII the way that MF2000 died of valine starvation. Hence, it might have been worth a try only to omit isoleucine with the minimal medium instead of both valine and isoleucine.

3.2.6 Cell-Free Expression

With the cell-free expression system (see section 2.4.14.3), functionality of the putative AHAS sequence was tested very quickly. This system uses the translational machinery of *E. coli* including ribosomes and T7 RNA polymerase. As a major advantage of this expression system, cell-based protein production steps such as transformation, cell culture maintenance, and expression optimization steps need not to be performed. Moreover, the resulting recombinant protein neither needs to be harvested from host cells nor purified and is directly submitted to the characterizing Westerfeld assay (section 2.5.10). The main disadvantage of this cell-free expression system is the production of very low quantities of recombinant protein which are not sufficient for a complete biochemical characterization.

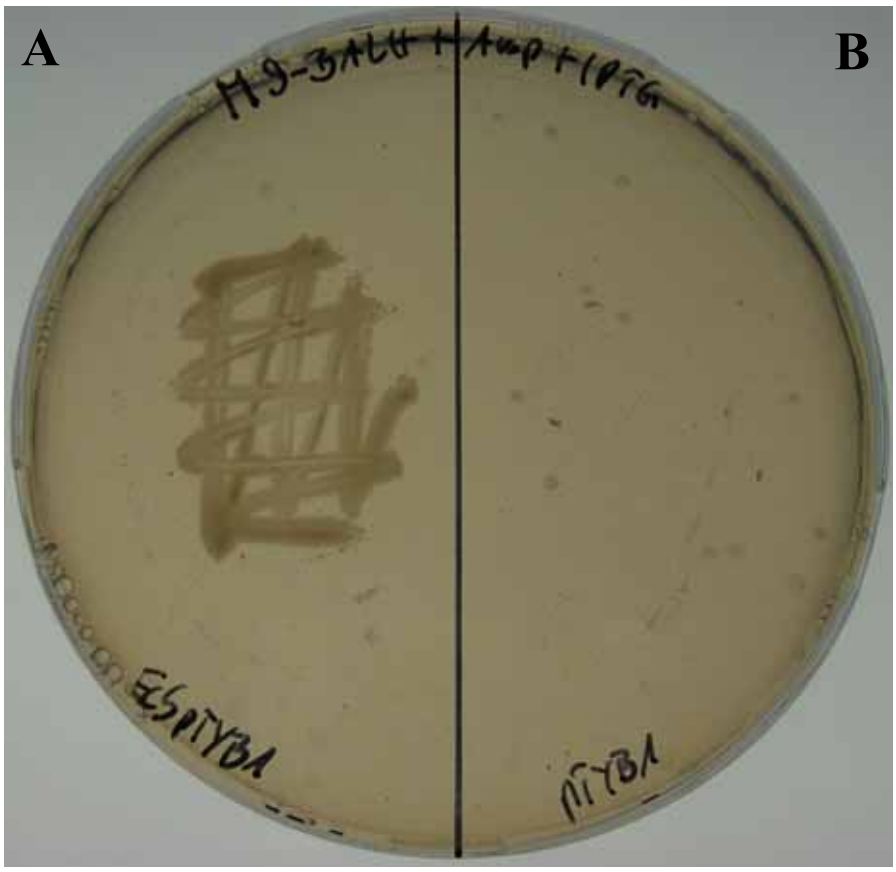


Figure 3.17: Growth on minimal medium M9-BALH supplemented with ampicillin and IPTG. **Panel A**: AHAS deficient *E. coli* strain MF2000 was transformed with plasmid EcSpTYB1-3 harbouring the AHAS catalytic subunit of *Eupatorium cannabinum*. **Panel B**: MF2000 transformed with vector pTYB1 as a negative control.

	concentration of plasmid [μg/ml]
EcScsupET3a	255
EcW2csupET3a	160
EcW3csu1750pET22b23a	200
EcSrsuIBA7	285
pET-GM	170

Table 3.27: Concentration of expression plasmids used in cell-free expression

According to the manufacturer, the Expressway Plus expression system (Invitrogen) should be used together with special vectors like pEXP1-DEST, pEXP2-DEST, pCR T7/CT-TOPO, pCR T7/NT-TOPO, or pRSET. The recommended vectors are characterized by a T7 promoter that was located 15 -20 bp upstream of the ribosome binding site, an ATG initiation codon that was positioned 9 -11 bp downstream of the ribosome binding site, and a stop codon that was found 4-100 bp upstream of the T7 transcription termination region. An alignment of the pET and pASK IBA vectors used in this work together with the recommended vectors showed that all features, the manufacturer had claimed, were realized within the pET and the pASK IBA vectors as well and, thus, were applied to the cell-free expression system.

Each of the three sequences from *Eupatorium cannabinum* coding for putative AHAS catalytic subunits were checked for functionality. Expression plasmids EcScsupET3a and EcW2csupET3a were obtained from cloning, as described in sections 3.4.1.1, harbouring the sequences EcScsu-AHAShyp and EcW2csu-AHAShyp, respectively, which were cloned in pET3a. Sequence EcW3csu-AHAShyp was cloned in pET22b-23a as shown in section 3.4.1.2 and yielded the expression plasmid EcW3csu1750pET22b23a. Expression plasmid EcSrsuIBA7 coded for the regulatory subunit EcSrsu-AHAShyp and was cloned in expression vector pASK IBA7+ as shown in section 3.4.3.1. Expression yielded proteins without any transit peptides. As a positive control, pET-GM was used, encoding for *E. coli* AHAS II regulatory and catalytic subunits. Two types of settings during expression were performed: Setting 1 solely used the catalytic subunit expression plasmid, while setting 2 contained a mixture of both catalytic and regulatory subunit expression plasmids in one reaction.

Translation reactions were set up and expression was carried out as described in section 2.4.14.3 on page 81. In Table 3.27, plasmid concentrations are listed. For setting 1 with only catalytic subunits, 8 μl of plasmids were used, while for the second setting, 4 μl of each subunit fit in the reaction mixture except for the positive control pET-GM that supplied the two types of subunits within a single expression plasmid.

The translation reaction was performed for 4 hours and was stopped with RNase A. The resulting mixture was submitted to Westerfeld enzyme assay with UV detection (see section 2.5.10.1) to check the AHAS-activity. In Table 3.28, results for both settings are

Setting 1: catalytic subunits only			Setting 2: catalytic and regulatory subunits mixed			resulting increase in acetoin formation
expression plasmid	A_{525}	produced μg acetoin /1 μg csu-plasmid	expression plasmids	A_{525}	produced μg acetoin /1 μg csu-plasmid	
EcScsu pET3a	0.28	0.75	EcScsupET3a + EcSrsuIBA7	0.34	1.81	2.4-fold
EcW2csu pET3a	0.32	1.36	EcW2csupET3a + EcSrsuIBA7	0.51	4.34	3.2-fold
EcW3csu1750 pET22b23a	0.32	1.09	EcW3csu1750 pET22b23a + EcSrsuIBA7	0.40	2.73	2.5-fold
			pET-GM	4.55	17.80	-

M_r(acetoin) = 88.11 g/mol, ε_M =16.164 $M^{-1}cm^{-1}$

Table 3.28: Comparison of acetoin formation from cell-free expressed catalytic AHAS without (setting 1) and with regulatory subunit (setting 2). csu = catalytic subunit, rsu = regulatory subunit

shown. For analysis of the results, a linearity between used amount of plasmid and amount of expressed enzyme was assumed. As a reference, the expression mixture contained H_2O instead of any plasmid. The Westerfeld detection (see page 95) was carried out in a volume of 1 ml.

AHAS catalytic and regulatory subunits were produced by *E. coli* translational machinery in an active form applying the novel sequences as matrix. The simultaneous translation of catalytic and regulatory subunits yielded a 2.4-fold increase in activity for the protein from the shoot-derived sequence EcScsu-AHAShyp and a 2.5-fold increase for the enzyme from the root-derived sequence EcW3csu-AHAShyp, while the only root-derived sequence EcW2csu-AHAShyp produced an AHAS with 3.2-fold increase in activity.

For biochemical characterization of the AHAS catalytic subunits, larger quantities of purified, recombinant protein were necessary, since the amount of produced protein during cell-free expression laid below the detection limit of the Bradford-method (see section 2.5.4 on page 87) and could not be determined. For calculating the specific activity of the recombinant enzyme (activity/mass protein [nkat/μg]), the mass of protein was needed. Expression trials of recombinant AHAS enzymes are described in the next section 3.4.

3.3 Development of a Specific Assay for Necic Acid Synthase Activity

The Westerfeld assay (section 2.5.10.1) was used for colourimetric detection of acetolactate which is the AHAS product of the first step in valine and leucine biosynthesis.

Acetolactate is decarboxylated at 60°C with the help 6N H_2SO_4 resulting in acetoin which spontaneously converts to biacetyl. In adding creatine and α-naphthol under strong alkaline conditions, an intense red colour develops. Until now, neither the mechanism nor the structure of this product has been fully elucidated. A guanidino group, probably formed from creatine in the presence of alkali, is postulated to react with biacetyl. Upon the addition of α-naphthol, the colour is intensified which leads to an increased sensitivity of the reaction.

Physiologically, the AHAS catalyzes a second competing reaction besides the condensation of two molecules of pyruvate to form acetolactate: the condensation of pyruvate and 2-oxobutyrate to form 2-aceto-2-hydroxybutyrate, which is an intermediate in the formation of isoleucine (Figure 3.18). Pyruvate plays a double role in this AHAS reaction: It is either converted to He-ThDP and it is the accepting molecule for the hydroxyethyl group to give acetolactate. In order to determine K_m values for 2-oxobutyrate, the AHAS reaction product 2-aceto-2-hydroxabutyrate needs to be detected selectively. Unfortunately, the red colour in the Westerfeld assay is formed with 2-aceto-2-hydroxybutyrate as well as from acetolactate but creates no additional, specific absorption maximum. Up to 40% of the colour yield was traced back to origin from 2-aceto-2-hydroxybutyrate when feeding 45 mM pyruvate and 55 mM 2-oxobutyrate to enterobacterial AHAS II (Epelbaum et al. 1990). Moreover, this specific maximum is very broad (Figure 3.18), which is reflected in a broad range of equally correct maxima applied: 540 nm were favoured by Westerfeld (1945) while Singh et al. (1988b) suggested 525 nm. Hence, a specific assay for the postulated product in necic acid biosynthesis was not obtainable from a modified Westerfeld assay.

3.3.1 ^{14}C-labelled Enzyme Assay with RITA (Rapid Intelligence TLC Analyzer)

In order to detect the postulated product 2-hydroxy-2-isopropyl-3-oxobutyrate (Figure 3.18) selectively, the enzyme reaction was performed with ^{14}C-labelled 2-oxoisovalerate (for synthesis see page 93) and ^{12}C-pyruvate in a ratio of 3 : 1 as previously determined by Weber (1997). Reaction cocktail 2 (see Table 2.22 on page 94) was modified using 5 mM ^{12}C-pyruvate and 15 mM ^{14}C-labelled 2-oxoisovalerate to avoid AHAS saturation. The reaction was stopped with 6N H_2SO_4, the volume of the reaction mixture was reduced under a constant flow of synthetic air and completely submitted to thin layer chromatography with 20 cm x 20 cm plates from Merck covered with 0.25 mm silica 60 F_{254}. Among several tested mobile phases, the mixture diethyl ether : n-hexane : formic acid (45:45:10) was chosen. High concentration of formic acid transformed all salts to acids and yielded clearly separated spots with following Rf-values:

	Rf-values
valine	0
pyruvic acid	0.25
2-oxoisovaleric acid	0.45

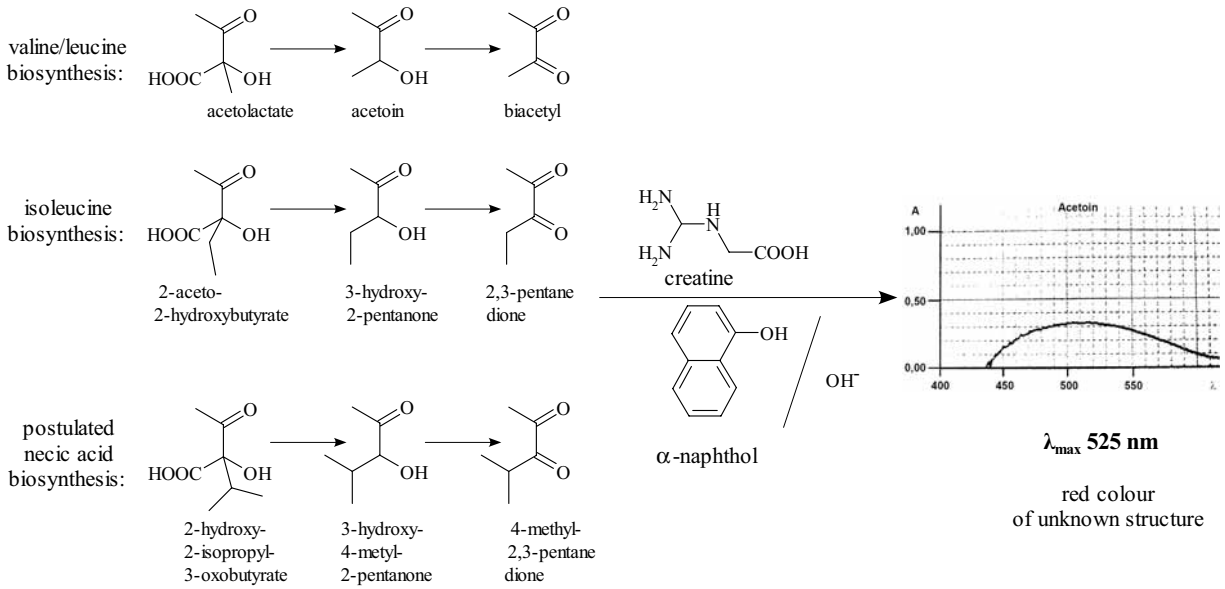


Figure 3.18: Westerfeld assay (Westerfeld 1945) used for quantification of reaction products with AHAS from physiological and postulated substrates: A broad maximum prevents the detection of differing substrates.

After 55 minutes of separation and a drying time of at least 2 hours, detection was performed with RITA (section 2.5.9.1 on page 93).

After two days, a second analysis of the same TLC plate revealed that the acids could not be detected anymore. Only valine was still detectable. It was assumed that the labile organic acids decarboxylated to volatile compounds and had vanished as gaseous CO_2. Since neither the radio-TLC trials were reproducible nor there were ^{14}C-labelled references (acetolactate and 2-hydroxy-2-isopropyl-3-oxobutyrate) available, the radio-TLC trials were abandoned.

3.3.2 GC-MS Based Enzyme Assay

In order to simplify the identification of educts and products from AHAS reactions with either pyruvate or pyruvate/2-oxoisovalerate mixtures (Figure 1.6) with MS-detection, both products (acetolactate and 2-hydroxy-2-isopropyl-3-oxobutyrate) were synthesized chemically in an cerium(III)-salt catalyzed, two step reaction (section 2.6) according to Christoffers et al. (2003). The resulting methylated products proved a higher stability than not derivatized substances. Hence, the AHAS reaction products were methylated immediately after incubation with diazald (section 2.5.10.2) to enable comparison as well as for stability reasons.

3.4 Expression of Recombinant AHAS

Due to a low abundance of plant AHAS, such as 0.005% of total protein (Durner 1991), and a very high lability, purification of AHAS from plant tissues is difficult. Moreover, previous attempts of AHAS purification from *Eupatorium clematideum* (Weber 1997) were hampered by the co-purification of pyruvate decarboxylase activity that mimics AHAS activity in the applied Westerfeld assay as detailed in section 3.1. Hence, the AHAS should be obtained from recombinant protein expression.

The necessary sequence information was discovered successfully as described in section 3.1. For biochemical characterization including the answer to the question whether one of the three novel AHAS catalytic subunits from *Eupatorium cannabinum* was exclusively responsible for catalyzing the first step in necic acid biosynthesis (section 1.2), the cDNAs were submitted to heterologous expression trials to obtain sufficient amounts of pure enzyme.

Heterologous expression was tested in the prokaryotic system of *Escherichia coli* and in the two eukaryotic systems of *Saccharomyces cerevisiae* and *Kluyveromyces lactis*. Finally, co-expression of catalytic and regulatory subunits in *E. coli* succeeded. For an overview of the methods carried out in heterologous expression, see Figure 2.6 on page 57.

3.4.1 Heterologous Expression in *Escherichia coli*

The following seven expression experiments with the prokaryotic system of *Escherichia coli* were performed to obtain active AHAS.

3.4.1.1 Standard Expression with the pET Vector System

The pET vector system supplies a very strong T7 promoter (see section 2.3.4.2) that was applied to produce sufficient amount of recombinant enzyme. Expression primers were designed according to section 2.4.9.4 with the forward primers, positioned as described in section 3.2.4, excluding the putative coding sequence for the chloroplast-targeting transit peptide (cTP). Moreover, with the help of the forward primer, a 6xHis-tag was fused to the recombinant protein that was exploited for purification (see section 2.5.2.1). Expression primer sequences are given in Table 3.29 on page 152.

Freshly harvested total RNA both from shoot- and root-tissue culture were transcribed (section 2.4.8) with eAMV reverse transcriptase at 55°C. With the help of primers P46 and P47, the sequence EcScsu-AHAShyp was amplified, while sequence EcW2csu-AHAShyp was obtained with the primers P48 and P49 in a full-length PCR conducted by the proof-reading DNA polymerase *Pfx* in a standard PCR reaction mixture (see section 2.4.9). The constant PCR programme was applied as described in section 2.4.9.4 and modified in the following: Annealing was carried out at 60°C for 90 seconds, elongation was performed at 72°C for 4 minutes, and cycling was repeated 40 times. Vector pET3a and PCR products were "double digested" (see section 2.4.10.2) with the restriction enzymes *Nde*I and *Bam*HI.

primer	direction	applied to expression of sequence	sequence	restriction enzyme	T_m [°C]
P46	forward	EcScsu-AHAShyp	5'-dATA C**A**T ATG CAT CAC CAT CAC CAT CAC CCA AAA GCT TTT GTC TCT CGG TTT GGC CCC-3'	*Nde*I	>75
P47	reverse	EcScsu-AHAShyp	5'-dTAT **G**GA TCC TCA TGT TTG TGT TCT GCC ATC ACC TTC AGT-3'	*Bam*HI	70.5
P48	forward	EcW2csu-AHAShyp	5'-dATA C**A**T ATG CAT CAC CAT CAC CAT CAC CCG GAG GCT TTC GTT TCT CGG TTC GGG CCG-3'	*Nde*I	>75
P49	reverse	EcW2csu-AHAShyp	5'-dTAT **G**GA TCC TTA CAT ATT TTT CTT GAT GAA TTG TGT TAG GCC-3'	*Bam*HI	67.5

restriction recognition sequence is underlined, restriction cleavage site is behind bold letters, temperature optimum for restriction enzymes is given in Table B.5

Table 3.29: Two pairs of primers for expression of the AHAS catalytic subunits from *Eupatorium cannabinum* EcScsu-AHAShyp (P46 + P47) and EcW2csu-AHAShyp (P48 + P49) in vector pET3a. Each pair of primers introduces an artificial ATG initiation codon and a N-terminal 6xHis-tag to the expressed protein replacing the native chloroplast-targeting transite peptide.

The two different inserts were ligated separately into the vector pET3a (section 2.4.10.2 on page 77) for 1 hour at 25°C. Ligation mixture was transformed into DH5α host cells (section 2.3.3.1 on page 43) and grown over night on LB plates pretreated with ampicillin (section 2.3.3.1). Colonies were screened by PCR (page 78) with the same PCR cycling programme and expression primers as used to produce the inserts. From all positive candidates, determined by colony-screening, plasmids were isolated with hand method I (section 2.4.3.2 on page 59) and checked for successful ligation once again. This time, restriction analysis with *Eco*RI was used.

This double check on successful ligation was applied since PCR colony screening is not as reliable as restriction analysis. Small quantities of the PCR product within the ligation mixture remaining on the plate are transferred into the PCR reaction during harvesting of the colony. These remnants are amplified during PCR which gives a false positive result. On the other hand, PCR colony screening is a very fast method to analyze large numbers of constructs. Hence, the two methods were combined in the chronological order: first, the less reliable but fast method to sort out most of the negative constructs, and second, the more reliable method to verify the positive ligation products.

Resulting positive plasmids were transformed into BL21(DE3) and induced, as described on page 80, in a 50 ml culture with 1 mM IPTG at an OD_{600} of 1.5. Initially, the cultures were grown for 3 hours at 37°C, followed by 14 hours at 25°C with an aeration of 250 rpm. Crude protein extracts were prepared by sonication (section 2.5.1 on page 81). Soluble protein was harvested by centrifugation, supernatant and pellet were analyzed by SDS-PAGE (see page 88).

As Figure 3.19 (page 154) shows, heterologously expressed protein from the sequence EcScsu-AHAShyp displayed the calculated mass of 64 kDa (calculated in section 3.2.4). The associated plasmid was designated EcScsupET3a. Proteins expressed from the sequence EcW2csu-AHAShyp showed the expected mass of 65 kDa and the corresponding plasmid was named EcW2csupET3a. Unfortunately, the expressed proteins were found as insoluble proteins in the pellet fractions (lanes P1 and P2 in Figure 3.19). His-tag purification of the supernatant fraction was carried out (as described in section 2.5.2.1) but no traces of soluble protein were detected with Coomassie staining (section 2.5.5.1). The recombinant protein was considered to be completely insoluble. The reason for the formation of insoluble recombinant protein as so-called "inclusion bodies" was not clear. There were three possible explanations and each was tested as described in the following sections (see Table 3.30). Additionally, solubilization and renaturing of inclusion bodies was tested as well (section 3.4.1.7).

3.4.1.2 Comparison of Expressions With and Without Plant-Derived Signal Sequence

Since three former heterologous expression trials had been successful when the plant signal sequence had not been omitted (Hershey et al. 1999, Bernasconi et al. 1995, and Kim and Chang 1995), it was not quite clear whether or not the signal peptide was necessary for an active enzyme (see section 3.2.4). Hence, a comparison between expression with and

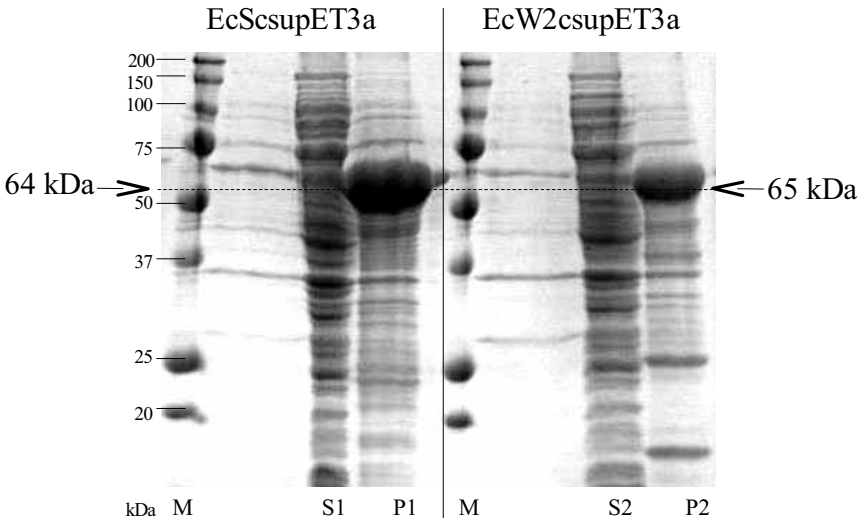


Figure 3.19: Fractions of *E. coli* crude extracts expressing AHAS catalytic subunits analyzed on SDS-PAGE gel (12%) stained with Coomassie blue. Recombinant proteins were obtained from expression of plasmid EcScsupET3a, coding for shoot-derived AHAScsu from *Eupatorium cannabinum*, and of plasmid EcW2csupET3a, coding for root-derived AHAScsu. Recombinant proteins only occurred in the pellet fractions (marked with P1 for EcScsupET3a and with P2 for EcW2csupET3a) while the supernatant fractions S1 and S2 did not contain any recombinant protein, as determined by His-tag purification. M = protein marker Dual color, Biorad, S1, S2 = supernatant fractions, P1, P2 = pellet fractions.

hypothesis	resulting experiment	see section
1. signal sequence is necessary for an active enzyme	1. expression is tested with and without the signal sequence	3.4.1.2
2. protein expression and extraction conditions are inadequate	2. modification of protein expression and extraction conditions	3.4.1.3
3. recombinant protein folds wrongly	3a) fusion to a solubilizing linker	3.4.1.4
	3b) co-expression with chaperones	3.4.1.5
	3c) co-expression with AHASrsu	3.4.3
	3d) expression in a eukaryotic host	3.4.2

Table 3.30: The three hypotheses probably explaning the formation of inclusion bodies and the resulting experiments necessary to reject or support the hypotheses

primer	direction	applied to expression of sequence	sequence	restriction enzyme	T_m [°C]
P50	forward	EcW3csu-AHAShyp	5'-dATA C**A**T ATG CCA GAG ATA TTC GTT TCC CGA TTC GCC CAA-3'	*Nde*I	70.5
P51	forward	EcW3csu-AHAShyp	5'-dATA C**A**T ATG ACT TCT GCA ACG ATG GCT GCC ATA GCT CCT CCA-3'	*Nde*I	73.3
P52	reverse	EcW3csu-AHAShyp	5'-dTAT **C**TC GAG ATA CTT CGT TCT ACC ATC ACC CTC GGT GAT CAC-3'	*Xho*I	73.3

restriction recognition sequence is underlined, restriction site is behind bold letters, temperature optimum for restriction enzymes is given in Table B.5

Table 3.31: Primers for expression of the AHAS catalytic subunit from *Eupatorium cannabinum* root-tissues with vector pET22b-23a

without plant chloroplast-targeting transit peptide (cTP) was carried out (hypothesis 1 in Table 3.30) with the cDNA EcW3csu-AHAShyp.

Freshly harvested RNA from root-tissue cultures was reversely transcribed with SSIII at 55°C (section 2.4.8), the resulting cDNA was used as a template for PCR in a constant PCR programme (section 2.4.9) that was altered in the annealing step conducted at 65.5°C for 90 seconds and in the elongation step performed at 72°C for 10 minutes. The reaction was carried out by *Pfx* proof-reading DNA polymerase and the reaction mixture was modified in containing 1 μl for each primer and 1 μl template instead of 2 μl. PCR with primers P50 (positioned as described in section 3.2.4) and P52 (Table 3.31) generated a sequence without the putative cTP coding region and comprised 1,750 bp, while PCR with primers P51 and P52 included the putative cTP coding stretch and resulted in a PCR product of 1,970 bp. A 6xHis-tag motif was introduced with the reverse primer P52 to each of the PCR products, providing a C-terminal purification tag fused to the expressed protein.

PCR products and the vector pET22b-23a (see section 2.3.4.2) were double digested with *Nde*I and *Xho*I within 3 hours at 37°C (section 2.4.10.2). The DNA was gel-purified (see section 2.4.5) and appropriate amounts of each insert and vector were ligated (page 77) at 25°C. Transformation in DH5α was followed by plasmid isolation with hand method I (section 2.4.3.2) and *Nde*I restriction analysis of the isolated plasmids. The plasmid, harbouring the sequence without the cTP coding region, was designated EcW3csu1750pET22b23a, while the plasmid including the cTP sequence was named EcW3csu1970pET22b23a. These plasmids were transformed into BL21(DE3) and the resulting colonies were induced in a 1 ml culture with 1 mM IPTG and grown as de-

scribed in section 3.4.1.1. SDS-PAGE analysis of the harvested supernatant and pellet fractions revealed that expression from plasmid EcW3csu1750pET22b23a produced a 64 kDa protein (as calculated in section 3.2.4, Figure 3.26) which was entirely found in the pellet fraction. Expression from plasmid EcW3csu1970pET22b23a was expected to produce a 73 kDa protein but showed no overexpression at all. Hence, hypothesis 1 (Table 3.30), postulating that the cTP is essential for an active enzyme, was rejected by the result that the cTP prevented overexpression. Moreover, the results from section 3.4.1.1 were confirmed for the third AHAS catalytic subunit of *Eupatorium cannabinum* as well that expression without the cTP was possible, but produced insoluble protein under the applied expression conditions.

3.4.1.3 Modification of *E. coli* Protein Expression and Extraction Conditions

In order to improve the quality of expressed protein, heterologous expression in BL21(DE3) by means of the three plasmids EcScsupET3a, EcW2csupET3a (created in section 3.4.1.1), and EcW3csu1750pET22b23a (section 3.4.1.2) was tested on different expression and extraction conditions.

Referring to the expression and extraction conditions in sections 3.4.1.1 and 3.4.1.2, these experiments were conducted with late log-phase cultures (OD_{600} 1.5), induced with 1 mM IPTG, incubated for 3 hours at 37°C followed by 14 hours at 25°C, and extracted by sonication (see line 1 in Table 3.33). Under these expression conditions, recombinant protein occurred as inclusion bodies. Induction at the late log-phase was chosen to obtain the maximum number of bacterial cells for a maximum amount of recombinant protein. A beginning shortage in alimentation and an increasing amount of metabolic end products may have put a strong stress on the host bacteria. Moreover, the amount of inducer was chosen very high compared to similar expression experiments (see Table 3.32) which had used a range from 0.3 mM to 1.0 mM IPTG. To sum up, all chosen parameters may have contributed to stressed bacteria that were deprived of time and energy to let the recombinant protein fold correctly, which is a commonly known phenomenon as outlined by Sambrook et al. (1989).

Hence, it was decided to reduce the stress by means of induction at an early log-phase (OD_{600} 0.3) with a reduced amount of IPTG (0.5 mM) under a constant growing temperature of 25°C. In order to impose a defined stress on the bacteria by supplementation, two minimal media (M9 and half of M9; for composition, see section 2.3.2.1) were tested besides the known full medium LB pH 7.5 (see line 2 in Table 3.33). Cells were harvested after 2, 4, 17, and 24 hours, sonicated, and the resulting proteins were separated by centrifugation. Fractions were checked by SDS-PAGE, but heterologously expressed protein was not detected in the soluble fraction. Moreover, protein mass did not increase up to 17 hours of induction compared to a culture harbouring a vector without insert. Thus, it was concluded, that the formation of inclusion bodies by a presumed stress was neither caused by the amount of inducer nor by the growth-phase in which the culture was induced.

Consequently, the next experimental set-up used similar conditions to the first expression trials (line 1 in Table 3.33). Cultures were grown to mid-log-phase (OD_{600} 0.6),

Expression conditions			method of	reported by
[°C][1]	OD_{600} [2]	IPTG[3] [mM]	cell lysis[4]	
37	0.8	1	lys. + son.	Ott et al. 1996
30	0.7	0.3	son.	Chang et al. 1997
28	0.5	1	son.	Dumas et al. 1997
30	0.8	0.5	lys. + son.	Chang and Duggleby 1997
30	0.6	0.5	son.	Hershey et al. 1999
30	0.8	0.3	son.	Oh et al. 2001

[1]growing temperature during expression
[2]optical density of the culture at the moment of induction
[3]IPTG concentration in the medium for induction
[4]lys. = lysozyme treatment, son. = sonication

Table 3.32: Overview of former successful conditions for heterologous expression of plant AHAS catalytic subunit

induced with 1 mM IPTG, and grown either at 37°C or at 25°C for 15.5 hours and 18 hours, respectively. The only parameter altered now, was the pH of the LB medium shifted from pH 7.5 to pH 8.5 (line 3 in Table 3.33). Since Kim and Chang (1995) and Chang et al. (1997) had reported to use a storage buffer of pH 8.1 and pH 9.6, respectively, for heterologously expressed and purified AHAScsu from *Nicotiana tabacum*, it was decided to test the formation of recombinant protein with an altered medium pH. This should influence the translation conditions for bacteria as the formation of inclusion bodies was shown to occur only during *in vivo* expression but not during expression in a cell-free system (section 3.2.6) of all three AHAScsu from *Eupatorium cannabinum*.

The LB medium was adjusted to pH 8.5 after autoclaving. Supernatant and pellet fractions were obtained, as described above, but did not show any additional band in SDS-PAGE analysis compared to the negative control of the host cells that were transformed with an "empty" vector, induced, and grown under the same conditions. Just in case the amount of soluble protein was only very small, a His-tag purification of the supernatant was performed as described in section 3.4.1.1. SDS-PAGE showed no purified protein even when stained with silver nitrate which is able to detect lower amounts of protein compared to staining with Coomassie (see section 2.5.5, page 89).

Nevertheless, these experiments (line 3, Table 3.33) differed from the preceding as they resulted in no detectable formation of inclusion bodies due to a change in the pH of the cultivating medium. Hence, the next experimental set-up (line 4 in Table 3.33) was chosen to optimize the method of cell lysis to find out whether the expressed protein, previously presumed insoluble, actually was insoluble or whether only the soluble portion was hidden behind an improperly working extraction method and, therefore, was not found in the supernatant.

An early log-phase culture (OD_{600} 0.4) was grown in LB pH 8.5, induced with 1.0 mM IPTG, and harvested after 18 hours at 37°C. Cells were forced open with a combined

	tested plasmid	Induction at OD_{600}	IPTG [mM]	temperature + duration of expression	medium	method of cell lysis	R
1	EcS EcW2 EcW3	1.5	1.0	37°C 3 h followed by 25°C 14 h	LB pH 7.5	sonication	A
2	EcW3	0.3	0.5	25°C 2 h, 4 h, 17 h, 24 h	LB pH 7.5 M9 0.5 M9	sonication	B
3	EcS	0.6	1.0	37°C 15.5 h 25°C 18 h	LB pH 8.5	sonication	B
4	EcS	0.4	1.0	37°C 18 h	LB pH 8.5	lysozyme treatment + sonication	C

R = results
A = expression in insoluble fraction, not cleanable from supernatant;
B = no detectable (SDS-PAGE) expression in supernatant;
C = expressed protein is cleanable from supernatant, but inactive

Table 3.33: Conducted modifications of *E. coli* protein expression and extraction conditions and their results. Each line shows a different experimental set-up that is described in more detail in the text. EcS = EcScsupET3a, EcW2 = EcW2csupET3a, EcW3 = EcW3csupET22b23a

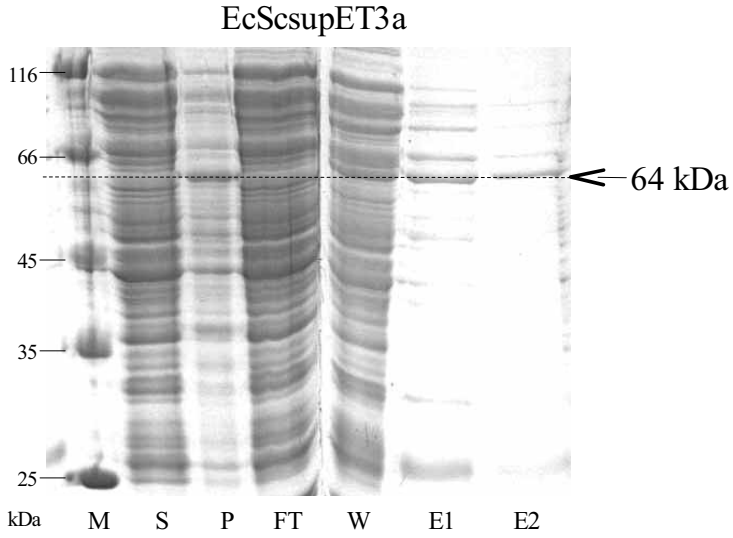


Figure 3.20: Overview of His-tag purification fractions from heterologously expressed AHAS catalytic subunit (AHAScsu) grown in LB pH 8.5 and analyzed on SDS-PAGE gel (12%) stained with Coomassie blue. Recombinant protein (64 kDa) was obtained from the expression of plasmid EcScsupET3a coding for shoot-derived AHAScsu from *Eupatorium cannabinum*. Soluble recombinant protein occurred in the elution fractions E1 and E2 while the pellet fraction displayed some insoluble protein. Enzyme activity was only found in the supernatant fraction. M = protein marker Fermentas, S = supernatant, P = pellet, FT = flow through, W = washing fraction, E1 and E2 = elution fractions 1 and 2.

treatment of lysozyme digestion (as described in section 2.5.1 on page 82) and sonication as performed in the previous trials. The enzyme was isolated by His-tag purification as described in section 3.4.1.1. The resulting fractions were submitted to SDS-PAGE analysis (Figure 3.20) and were examined on activity with the Westerfeld assay (described in section 2.5.10) that was zeronized with a negative control of host cell extract. The crude supernatant fraction (S) showed activity, while the washing (W) and elution (E1, E2) fractions were inactive.

Hence, there were two possible explanations: One, regarding the enzyme as soluble, cleanable and active unless it was purified which let to conversion into an inactive state, the other, supposing that the supernatant fraction contained two types of recombinant AHAScsu differing in their state of folding, one was active, but not cleanable due to an inaccessible His-tag, while the second type of enzyme was cleanable, but inactive, whereas both types of the AHAScsu were folded incorrectly.

In order to test the first possible explanation, active recombinant AHAScsu was prepared from heterologous expression of pET-GM (see section 2.3.4.4) in BL21(DE3) as described by Hill et al. (1997). The plasmid pET-GM encoded *E. coli* AHASII and purified fractions were incubated with 0.1 - 100 mM imidazole in the Westerfeld assay

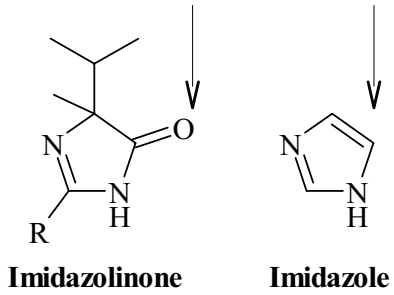


Figure 3.21: Structural relationship between imidazolinone, a class of AHAS targeting herbicides and imidazole, the eluting agent in His-tag purification. The exocyclic oxygen substitution in the imidazolinone (marked with an arrow) is absolutely necessary for herbicidal activity. The imidazole is missing the exocyclic oxygen.

reaction mixture. Imidazole is used in the elution process during His-tag purification (section 2.5.2.1) and is structurally related to the imidazolinones (see Figure 3.21) which form a class of AHAScsu inhibiting herbicides. An exocyclic oxygen substitution on the dihydroimidazolone ring of the imidazolinones is absolutely necessary for herbicidal activity (McCourt et al. 2006). Since the imidazole structure lacks this exocyclic oxygen, it should not be able to deactivate the AHAScsu, but this was tested never before. Actually, the activity of *E. coli* AHASII did not decrease in the presence of imidazole. Hence, the first explanation, that the recombinant enzyme was inactivated during purification, was refuted, assuming that no other factors of purification affected the enzyme activity.

The second idea that some molecules of the recombinant enzymes were active, but the His-tag was not accessible for purification, while another group of molecules was extractable via His-tag purification, but inactive *a priori*, is in accordance with the hypothesis of the folding funnel (see Figure 3.22, Schultz 2000). Protein folding is mainly driven by hydrophobic amino acids annealing to each other to exclude the polar solvent water. A protein, consisting out of 100 amino acids, is able to form 10^3 different 3D-structures, but only one is correct (Mogk et al. 2001). To find the way to the correct structure, the protein is guided by its native surrounding. In not native surroundings like it is the case for plant enzymes in *E. coli* host cells, proteins can misfold in several states. The way to the most energy-poor status can end in a local minimum that keeps the protein in a misfolded state. Some of those local minima are very close to the native status and may supply weak activity, while others do not, but may enable the access to a His-tag, for instance.

To sum up, the line of trials described in this section has shown that hypothesis 2 (section 3.4.1.1), postulating inadequate expression and/or extraction conditions, was true for the extraction conditions which demanded a combined treatment with lysozyme and sonication. Furthermore, expression conditions were modified and showed to influence the protein folding by using an alkaline (pH 8.5) medium, but suitable modifications were not

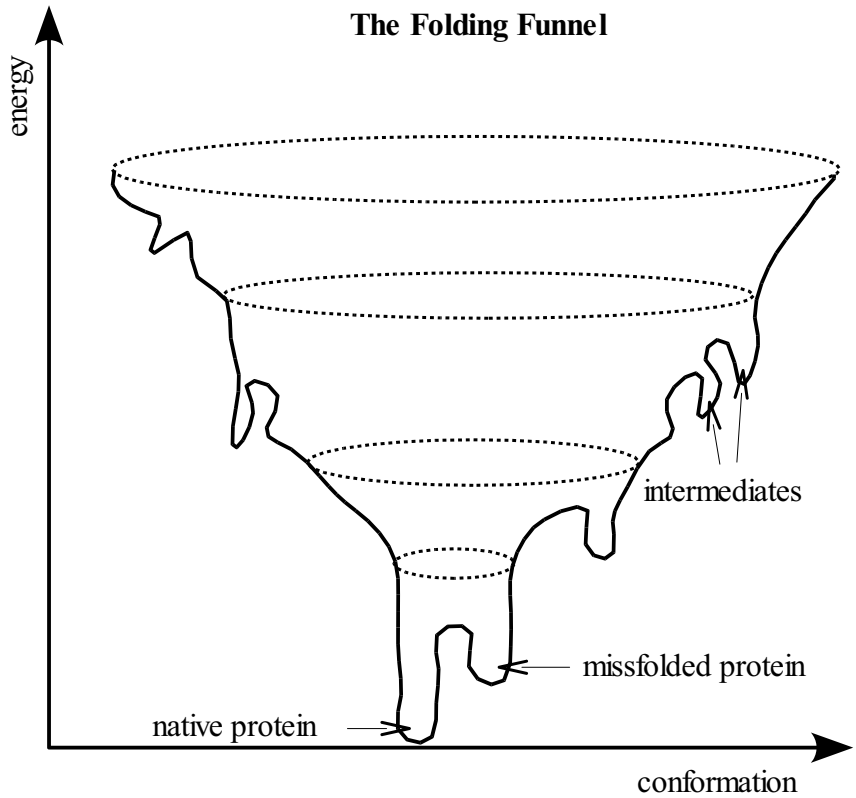


Figure 3.22: The folding funnel shows the free energy (vertical axis) of all possible protein structures as a function of the conformational degree of freedom (horizontial axis). The funnel supplies various local minima, some of them presenting intermediates while others function as a trapp keeping the protein in the missfolded status (modified from Schultz 2000).

solubilizing linker	plant AHAScsu from	reported by
ketol-acid reductoisomerase (50 kDa)	*Arabidopsis thaliana*	Dumas et al. 1997
glutathione S-transferase (26 kDa)	*Xanthium spec.*	Bernasconi et al. 1995
glutathione S-transferase (26 kDa)	*Arabidopsis thaliana*	Ott et al. 1996
glutathione S-transferase (26 kDa)	*Arabidopsis thaliana*	Hershey et al. 1999
glutathione S-transferase (26 kDa)	*Nicotiana tabacum*	Chang et al. 1997
glutathione S-transferase (26 kDa)	*Nicotiana tabacum*	Oh et al. 2001

Table 3.34: Overview of former successful heterologous expression experiments of plant AHAS catalytic subunits employing a linker

identified. For the expression of active enzymes, the hypothesis 3 (recombinant protein folds wrongly) was tested in the following sections.

3.4.1.4 Fusion to a "Solubilizing Linker"

"Solubilizing linkers" are large tags with more than 60 amino acids. They are known to increase solubility and stability of the target protein (Terpe 2004). Additionally, they can be used for purification with the appropriate matrix.

Six expression experiments with plant AHAS catalytic subunits, described in the literature (see Table 3.34), were successful when the AHAScsu was fused to such a tag.

Intein-tag

Fusing an insoluble protein to a highly soluble tag may result in a soluble protein if the tag was only large enough. In order to test this assumption, the IMPACT-CN expression system (NEB) with the vector pTYB1 was used. The vector pTYB1 fused a 55 kDa Intein-tag (see 2.3.4.2 page 46) C-terminally to the cTP-free target protein of 64 kDa. Hence, the fusion protein had a size of 119 kDa. Despite of promoting the solubility, the tag was used for purification with Chitin beads affinity chromatography (see section 2.5.2.2, page 84) as a chitin binding domain (ChBD) was included within the Intein-tag.

Primers P53 and P54 were designed according to section 2.3.4.2. The ORF of sequence EcScsu-AHAShyp (section 3.1.1.1, page 130) provided C-terminally a threonine that might have provoked an early *in vivo* cleavage of the tag (detailed in section 2.3.4.2). Hence, an additional glycine was introduced with P54 (indicated in italic letters in Table 3.35).

Freshly harvested shoot RNA was transcribed with SSIII at 60°C. Full-length PCR was carried out with the resulting cDNA applying the same modifications as described in section 3.4.1.2 except for an annealing temperature of 64°C. The resulting full-length DNA had the expected size of 1,750 bp.

The following restriction and ligation procedures were carried out as described in section 3.4.1.2 except for the following modifications: Vector pTYB1 and the full-length PCR product were restricted in a double digestion with *Sap*I and *Nde*I and ligation was incubated at 16°C over night. Plasmids were transformed directly into the expression host ER2566 (2.3.1) and positive transformants were determined by restriction analysis of the re-isolated constructs with *Eco*RV. The resulting plasmid was designated EcScsupTYB1

primer	direction	sequence	restriction enzyme	T_m [°C]
P53	forward	5'-dTAT AC**A** TAT GCC AAA AGC TTT TGT CTC TCG GTT TGG CCC C-3'	*Nde*I	71.5
P54	reverse	5'-ATA TAT GCT CTT C**C**G C**A***T CC*T GTT TGT GTT CTG CCA TCA CCT TCA GT-3'	*Sap*I (N_1/N_4)↓	73.8

restriction recognition sequence is underlined, restriction site is behind bold letters, artificially added nucleotides for avoiding a premature self-cleavage of the tag are in italic letters, temperature optimum for restriction enzymes is given in Table B.5

Table 3.35: Primers for expression of the AHAS catalytic subunit from *Eupatorium cannabinum* shoot tissues with vector pTYB1

and its clone was submitted to a small scale expression which confirmed that the clone was able to produce protein with the expected size of 119 kDa.

The expression culture was scaled up to 500 ml LB, induced with 0.3 mM IPTG final concentration at an OD_{600} of 0.4 for 5 hours at 30°C. Cell lysis was performed with the help of lysozyme and subsequent sonication as described in section 2.5.1. Soluble protein was harvested by centrifugation, purified as described in section 2.5.2.2, and the elution fractions were checked for activity. Unfortunately, no activity was detected. Moreover, the enclosed SDS-PAGE analysis showed that the complete recombinant protein was found in the pellet fraction and no protein had bound to and was eluted from the matrix. Hence, it was concluded that the "solubilizing" Intein-tag did not work for the AHAScsu. Probably, this was due to the fact that the recombinant protein had a total size of 119 kDa which slightly exceeded a discussed limit of solubility of 100 kDa for fusion proteins (Terpe 2004).

GST-tag

As shown in Table 3.34, five AHAScsu from plants had been already successfully expressed in the pGEX vector system. With vector pGEX-5x-3 (see section 2.3.4.2, page 47), a N-terminal glutathione S-transferase (GST) tag of 26 kDa was fused to the cTP-deprived protein from sequence EcScsu-AHAShyp. The heterologously expressed protein was expected to show a size of 90 kDa.

Full-length PCR was performed with the primers P55 and P56 as described above for expression with the Intein-tag (see page 162). The only difference was that the cDNA was transcribed at 55°C instead of 60°C with SSIII. The resulting fragment had the expected size of 1,750 bp. Vector and PCR product were double digested separately with *Bam*HI and *Not*I. The mixtures were gel-purified, appropriate amounts of vector and insert were ligated over night at 16°C, and transformed into XL1blue. Plasmids were re-isolated from successful transformants and screened with *Bam*HI restriction analysis. The successful plasmid was designated EcScsupGEX.

primer	direction	sequence	restriction enzyme	T_m [°C]
P55	forward	5'-dATA **G**GA TCC *CAC* CAA AAG CTT TTG TCT CTC GGT TTG GCC-3'	*Bam*HI	72.6
P56	reverse	5'-dTAT G**C**G GCC GCT CAT GTT TGT GTT CTG CCA TCA CCT TCA GT-3'	*Not*I	74.4

restriction recognition sequence is underlined, restriction site is behind bold letters, artificially added nucleotides for positioning the ORF are in italic letters, temperature optimum for restriction enzymes is given in Table B.5

Table 3.36: Primers for expression of the AHAS catalytic subunit from *Eupatorium cannabinum* shoot tissues with vector pGEX-5X-3

Expression was checked in a small scale as mentioned above for expression with Intein-tag. Cultures were extracted by lysozyme treatment followed by sonication and harvested soluble fractions were purified on FPLC as described in section 2.5.2.3 on page 85. Purification was not successful and the enclosed analysis on SDS-PAGE revealed that the major amount of expressed protein was found in the insoluble fraction while the soluble recombinant protein did not bind to the matrix. The recombinant protein from expression with vector pGEX-5x-3 was stated as soluble but still misfolded and, thus, not cleanable as the purification tag was misfolded as well.

3.4.1.5 Co-Expression with Chaperonins

Chaperonins are subunits of chaperones which are known to assist *in vivo* in protein folding. As mentioned in section 3.2.4, they are also involved in the translocation of proteins across the outer and inner membranes of chloroplasts. Moreover, chaperones had already shown to provide soluble and active recombinant enzyme from previously insoluble proteins when expressed simultaneously together with the enzyme of interest in a single host cell (Caspers et al. 1994). These features were tested with the intention to gain a soluble and active AHAS catalytic subunit.

As reported by Genzel (2006), the first trial to increase the number of *E. coli* chaperones endogenously in adding 2% ethanol to the medium had failed (section 3.4.1.6). Nevertheless, each of the two plasmids pREP4-groESL and pRDKJG, encoding a set of chaperonins, were co-expressed with either the plasmid EcScsupTYB1 or EcScsupGEX (constructed in section 3.4.1.4), both encoding the shoot-derived catalytic subunit of AHAS from *Eupatorium cannabinum*. For more information about the encoding chaperone vectors or the catalytic cycles of the resulting holoenzymes, please refer to section 2.3.4.3 on page 48.

BL21(DE3) host cells were transformed as described in section 2.3.3.1 on page 43 with either pRDKJG or pREP4-groESL. As both of the chaperone vectors conferred

kanamycin resistance to the host, only one type of vector was maintained securely in a host cell. Cultures harbouring one of the vectors were made competent (section 2.3.3.1 on page 42) and transformed with one of the AHAScsu encoding plasmids. Four different plasmid-vector combinations were created: (1) EcScsupTYB1 + pREP4-groESL, (2) EcScsupTYB1 + pRDKJG, (3) EcScsupGEX + pREP4-groESL, and (4) EcScsupGEX + pRDKJG. Positive transformants were submitted to protein expression (section 2.4.14.1) with the following experimental set ups:

As standard expression conditions, it was chosen to use a final concentration of 1 mM IPTG as inducer and to perform expression at 37°C. Cells were extracted with the combined lysozyme-sonication method (see page 82) and checked by SDS-PAGE (section 2.5.5). As none of the 4 plasmid-vector combinations was able to provide a soluble protein, expression conditions were modified to find out which of the 2 parameters (amount of inducer or growing temperature) was limiting: First, both, the amount of inducer (0.3 mM IPTG) and the growing temperature (30°C) were lowered moderately, and then, the standard amount of inducer (1 mM IPTG) was combined once with a significantly lowered temperature (25°C), and once with a moderately lowered temperature (30°C).

For all 4 plasmid-vector combinations, the largest amounts of soluble protein were obtained with a significantly lowered growing temperature (25°C) and the standard amount of inducer (1 mM IPTG). Conclusively, the growing temperature turned out to be a sensitive parameter in heterologous protein expression of AHAScsu.

In order to perform the Westerfeld assay (section 2.5.10), soluble proteins were submitted to purification. In case of EcScsupGEX, the protein did not bind to the glutathione sepharose during FPLC-supported purification (see section 2.5.2.3), although an additional sequencing of the plasmid had shown a perfectly inserted AHAS sequence.

In case of EcScsupTYB1, Intein-tag purification was carried out in the batch format as described in section 2.5.2.2. A soluble protein was extracted but the enclosed activity-test proved negative.

To sum up the various experimental lineages, both types of chaperones were able to provide a soluble protein when the growing temperature did not exceed 25°C. Surprisingly, the size of the heterologously expressed protein did not influence this result, since the chaperone system GroESL is known to be capable to process proteins with a maximum size of 60 kDa (section 2.3.4.3). Both recombinant proteins exceeded this limit with EcScsupTYB1 and EcScsupGEX showing a size of 119 kDa and 90 kDa, respectively. Hence, a new co-expression experiment with a smaller sized recombinant protein was performed taking the growing temperature limit of 25°C into account. Plasmid EcScsupIBA32 (constructed in section 3.4.1.6) harboured sequence EcScsu-AHAShyp including a C-terminally fused 6xHis-tag resulting in a size of 64 kDa expressed protein. Co-expression with chaperones GroESL provided soluble and purifyable but inactive protein as determined by Genzel (2006).

All results of this section considered, chaperones did not help to fold the enzyme correctly as the AHAScsu was either not cleanable, which was probably due to a misfolded purification tag, or it was cleanable but inactive, which then was due to a misfolded active center. In general, these findings are congruent to the results of expression in LB pH 8.5

described in section 3.4.1.3 on page 156.

3.4.1.6 Expression into the Periplasmic Space

The periplasmic space in gram negative bacteria is localized between the outer membrane and the cell's plasma membrane (inner membrane). The space harbours peptidoglycans that constitute the bacterial cell wall. Bacteria like *E. coli* are called "gram negative" because during the gram staining the blue dye is not retained in the cell wall as the latter is shielded by the outer membrane. Bacteria with single membranes but thicker cell walls are "gram positive" as their cell wall is not shielded by an outer membrane and the dye can enter the cell wall. Hence, gram positive bacteria have no periplasmic space.

The periplasmic space is known to offer redox conditions differing from those in the cytoplasm which can enhance correct folding of the protein (see section 2.3.4.2, page 47). According to Pickett and coworkers (Zhou et al. 2004), the heterologous expression in *E. coli* and translocation to the periplasmic space of the fruit fly *Drosophila melanogaster* odorant-binding protein had resulted in soluble protein without any modifications compared to the natural mature protein, while previous experiments had shown that expression in the host cytoplasm resulted in completely insoluble proteins.

In order to use the periplasmic space for production of soluble protein, the vector pASKIBA32 with ompA as a leading sequence was used (2.3.4.2). Cloning and expression of sequences EcScsu-AHAShyp (see Table 3.22 on page 130) were performed as detailed by Genzel (2006). The resulting plasmid was designated EcScsupIBA32 and coded for a recombinant protein that contained N-terminally the periplasmic ompA leader, and C-terminally, a 6xHis-tag was fused for purification facilities. The heterologous protein did not contain the plant putative cTP.

Extracting the periplasm of the induced BL21(DE3) culture (see page 82) showed that the protein did not enter the periplasmic space. Further investigation applying the combined sonication-lysozyme method (see page 82) checked the cytoplasm. Here, it was shown that the protein was found in the insoluble fraction. Altering the growing conditions in changing the pH of the LB medium from 8.5 to 7.5 did not change the previous result. The same happened when 2% ethanol was added during exponential growth half an hour prior to induction, although this should increase the amount of host derived chaperons (Thomas and Baneyx 1996) that are known to enhance a correct protein folding.

Reasons for the absence of the recombinant proteins from the periplasmic space may be due to the fact that translocation across the cell membrane of *E. coli* is still understood incompletely (Wickner et al. 1991) and, thus, was not performed properly. Fusion with a leader sequence is necessary but not sufficient for an export into the periplasm. The leader sequence must be recognized by a cell membrane standing protease. In case of the ompA-leader, it might have been much more promising not to use an *ompT*-deficient host like BL21(DE3) but, for example, JM83, as recommended by the manufacturer (personal communications with Dr. K. Terpe, IBA Göttingen, Germany). Moreover, it could not be excluded that the target protein is no good candidate for periplasmic localization, since the net charge of the N-terminal amino acids on the protein can inhibit translocation

(Kajava et al. 2000).

3.4.1.7 Refolding

Up to this point, no recombinant enzyme was obtained in a soluble and active form regardless which modifications of heterologous expression trials had been performed (sections 3.4.1.1 to 3.4.1.6). On the other hand, large amounts of insoluble protein were available. Hence, solubilization and refolding of the enzyme was tested to obtain a correctly folded and active AHAScsu.

A single colony of BL21(DE3) harbouring plasmid EcW2csupET3a (see page 151) was inoculated in 7 ml LB/ampicillin and grown over night at 37°C. From this over-night culture, 2 ml were transferred to 100 ml LB/ampicillin and grown to an OD_{600} of 0.5 within 6 hours. The culture was induced with IPTG in a final concentration of 5 mM and grown for 17 hour at 37°C. Cells were harvested and stored at -20°C. Cell lysis, solubilization, and renaturing were performed as described in section 2.5.3 on page 87. 100 ml of induced culture gave a wet weight of 70 mg pellet. The solubilized and renatured protein was submitted to His-tag purification (described in section 2.5.2.1) but did not bind to the Ni-NTA matrix. It was stated that the enzyme refolded with the 6xHis-tag inaccessible to the purification matrix.

As known from refolding experiments with recombinant senecionine *N*-oxygenase (SNO) from the butterfly *Tyria jacobaeae* carried out by Naumann (2003), the refolded enzyme had been 28 times less active than the native SNO isolated from the insect (Lindigkeit et al. 1997). The native AHAScsu from *Eupatorium clematideum* had shown a very low specific activity of 0.0205 nkat/mg (Weber 1997) compared to 80.2 nkat/mg of native AHAScsu from *Hordeum vulgare* (Durner 1991), although Weber (1997) had tested 12 different parameters in various concentrations to improve the specific activity. A 28-fold decrease in activity of a refolded AHAScsu would have been below the detection limit. Moreover, Poulsen and Stougaard (1989) had reported about the first heterologous expression of a eukaryotic AHAS (*Saccharomyces cerevisiae*) that solubilization had been achieved by treatment with 8 M urea, but renaturing had yielded only a minor fraction. A refolding protocol had remained unpublished. Hence, it was concluded that the AHAScsu activity needs very special conditions that will not be obtained by refolding.

3.4.2 Heterologous Expression in Eukaryotes

The genetic code is degenerate. For one amino acid, there are up to six different triplets which can be used (arginine, leucine, serine). This usage is similar among closely related species. Not closely related species like bacteria and plants display a different codon usage. The previously described expression trials (sections 3.4.1.1 to 3.4.1.7) were all using *E. coli* as a host, but were not able to provide an active enzyme. The reason for this was hypothesized to be due to an incorrectly folded protein (section 3.4.1.3). Since expression in a cell-free system (section 3.2.6) using the *E. coli* translational machinery was able to produce active enzyme, it was concluded that wrong folding occurred *in vivo* in the

codon-usage tables from		mean difference [%] determined with gcua, see 2.4.2
E. coli, prokaryote		15.06
S. cerevisiae, eukaryote, unicellular		8.25
A. thaliana	eukaryote, dicotyl	8.61
N. tabacum		8.92
P. sativum		9.36
EcScsu-AHAShyp		0.00
O. sativa	eukaryote, monocot	16.39
Z. mays		17.33

Table 3.37: Mean differences in codon usage between *Eupatorium cannabinum* determined with sequence EcScsu-AHAShyp and other species.

E. coli expression host. Hence, it was postulated that the *E. coli* host might have been run short of rare tRNAs and used mismatching instead of correct but unavailable tRNAs. This would introduce wrong amino acids and result in an altered and probably inactive structure.

In order to decide whether expression in eukaryotic hosts might result in active enzymes, the codon usage was determined from the sequence EcScsu-AHAShyp (3.1.1.1) by the programme guac (see section 2.4.2). The resulting codon-usage table was compared to tables already available in the guac programme. The resulting mean difference is illustrated in Table 3.37.

The mean difference of the codon-usage tables from EcScsu-AHAShyp and *E. coli* was higher (15.06%) than the mean difference towards the codon-usage tables from *Saccharomyces cerevisiae* (8.25%) that was similar to those from *Arabidopsis thaliana* (8.61%), *Nicotiana tabacum* (8.92%), and *Pisum sativum* (9.36%). The codon usage of the monocots *Oryza sativa* (16.39%) and *Zea mays* (17.33%) seemed to be less similar. Hence, it was decided to test heterologous expression in eukaryotic hosts, although inclusion bodies are known to be produced in eukaryotes like insect or monkey cells as well (Mukhopadhyay 1997). Two different yeast strains were tested: *Saccharomyces cerevisiae* expressed the recombinant protein intracellularly, while *Kluyveromyces lactis* was able to secrete the expressed protein to the medium which allowed easy purification of the recombinant protein.

3.4.2.1 *Saccharomyces cerevisiae*

According to the manufacturer's protocol (pYES2 manual, Invitrogen), expression primers annealing at the 5'terminal sequence (forward primers) needed to contain a Kozak translation initiation sequence (ANN ATG G) and an ATG initiation codon for proper start of translation into the correct ORF (see section 2.3.4.5, Kozak 1987, Kozak 1990). Expression primers were designed by Enß (2006).

primer	direction	sequence	restriction enzyme	T_m [°C]
P65	forward	5'-dTAT AGA GC**T** C AT TAT GG G CCA TCA CCA TCA CCA TCA CGC TTT TGT CTC TCG GTT TGG C-3'	*Sac*I	77.9
P67	reverse	5'-dTAT AGC AT**G** CCT ATG TTT GTG TTC TGC CAT CAC CTT CAG T-3'	*Pae*I	70.5
P66	forward	5'-dTAT AGA GC**T** C AT TAT GG C TTT TGT CTC TCG GTT TGG CCC CGA C-3'	*Sac*I	74.2
P68	reverse	5'-dTAT AGC AT**G** CCT AGT GAT GGT GAT GGT GAT GTG TTT GTG TTC TGC CAT C-3'	*PaeI*	74.4

restriction recognition sequence is underlined, restriction site is behind bold letters, letters marked by boxes show the Kozak translation initiation sequence, temperature optimum for restriction enzymes is given in Table B.5

Table 3.38: Primers as designed by Enß (2006) used for expression of the AHAS catalytic subunit from *Eupatorium cannabinum* shoot tissues (EcScsu-AHAShyp) with vector pYES2 in *S. cerevisiae* INV*Sc*1

Full-length PCR was performed with both primer combinations: P65 + P66 resulting in a N-terminal and P67 + P68 coding for a C-terminal His-tag (see Table 3.38). Standard PCR cycling was carried out as described in section 3.4.1.2 except for the following modifications: Reversely transcribed RNA from shoot tissues was used as a template and the annealing procedure was carried out at 60°C for 4 minutes. The resulting full-length PCR product was incubated with *Taq* DNA polymerase to add a single deoxyadenosine overhang to the 3'ends (2.4.10.1) allowing subcloning in the pGEM T easy vector (page 76). Subcloning was checked with *Not*I restriction analysis. Successful constructs and vector pYES2 were submitted separately to a double digest with *Sac*I and *Pae*I (section 2.4.10.2). Purified restriction products were ligated over night at room temperature. An additional restriction cut into the MCS of unligated vectors was done with *Not*I (page 77) to reduce the number of false positive transformants. The resulting products were transformed into DH5α, their plasmids were isolated, and successful ligation was determined with restriction analysis. Plasmids were transformed into INVSc1 with the Dohmen protocol (see page 43), plated on selective minimal medium (SC-U, section 2.3.2.2) and grown for 4 days at 30°C.

Cells were transferred to liquid medium and pre-cultured at 30°C in 500 ml baffled-bottom-flasks with 0.1 volumes (50 ml) of SC-U including either raffinose or glucose. While the glucose supplemented cells reached the appropriate OD_{600} within 24 hours, the cells cultured in raffinose needed 48 hours. Cells were pelleted and transferred to 100 ml induction medium (2.3.2.2) in 1 l flasks where they grew 14 hours at 25°C and an aeration of 210 rpm. Cells were harvested and extracted with CelLytic Y reagent (described on page 83). The supernatant was harvested by centrifugation, desalted to the His-tag purification buffer via PD10 columns, and finally His-tag purified (section 2.5.2.1). All elution fractions showed neither activity in the Westerfeld assay (section 2.5.10) nor did they display a Coomassie detectable protein band on the SDS-PAGE gel. More concluding remarks are given at the end of the next section, summarizing this and the following yeast expression experiments.

3.4.2.2 *Kluyveromyces lactis*

Full-length primers (Table 3.39) for expression were designed as recommended by the manufacturer (NEB) for secretion of expressed protein into the medium (see section 2.3.4.5, page 53). The PCR reaction mixture was composed, as described in section 2.4.9, containing 3 μl cDNA of reversely transcribed RNA harvested from shoot tissues. Standard cycling parameters were applied as described in section 3.4.1.2 except for the annealing step that was carried out at 65°C for 5 minutes. The resulting PCR product showed the expected size of 1,750 bp and sticky ends for ligation were prepared with a *Bsa*I restriction reaction. As detailed by the manufacturer, the vector pKLAC1 did not offer a *Bsa*I restriction site. Thus, it was cut in a double digest with *Xho*I and *Not*I which enabled the *Bsa*I-restricted insert to fit into the vector. This way of cloning eliminated the whole MCS which meant that the Kex cleavage site encoding nucleotides had to be re-inserted via the forward primer. The resulting protein contained no artificial amino acids except

primer	direction	sequence	restriction enzyme	T_m [°C]
P69	forward	5'-dTAT AGG TCT C**C**T CGA G AAA AAG ACA CCA TCA CCA TCA CCA TCC AAA AGC TTT TGT CTC TCG GTT TGG CCC C-3'	*Bsa*I	> 75
P70	reverse	5'-dTAT AGG TCT C**A**G GCC TCA TGT TTG TGT TCT GCC ATC ACC TTC AGT-3'	*Bsa*I	74

restriction recognition sequence is underlined, restriction site is behind bold letters, letters marked by boxes show the Kex cleavage site sequence, temperature optimum for restriction enzymes is given in Table B.5

Table 3.39: Primers for expression of secreted AHAS catalytic subunit from *Eupatorium cannabinum* shoot tissues (EcScsu-AHAShyp) with vector pKLAC1 in *K. lactis*

for the N-terminally fused 6xHis-tag for purification facilities.

Restriction and ligation were carried out as described in section 2.4.10.2. Ligation was performed over night at room temperature and the resulting products were cloned in TOP 10 cells for maintenance. Plasmids were obtained from successful transformants by spin preparation (section 2.4.3.2) and were checked with *Eco*0109I restriction analysis. 26.4 μg of positive plasmids were cut with *Sac*II, spin purified (2.4.5) and transformed according to Dohmen et al. (1991) in *K. lactis* host cells (see page 43) with a concentration of 245 μg/ml. Cells were plated on selective medium YCBA (section 2.3.2.3) and grown for 4 days at 30°C. Developing colonies were grown over night in 15 ml liquid YPGlu, harvested, and total DNA was extracted with DNeasy Tissue kit (see page 61). DNA was submitted to PCR screening (section 2.4.11.2, page 78) for properly integrated fragments. Cycling was performed as recommended by the manufacturer except for the annealing temperature of 50°C with a duration of 20 seconds and the elongation temperature of 68°C carried out for 30 cycles. Induction of clones harbouring two- or more-fold integrated fragments was carried out as described in section 2.4.14.2 on page 80. The pre-culture was grown in 2 ml YPGlu over night at 30°C. The induction culture was then cultured in 75 ml YPGal inoculated with 750 μl pre-culture and cultivated in a 1 l baffled-bottom flask to increase aeration. After 4 days at 30°C and 250 rpm, the medium was harvested and the expressed protein was concentrated with Ultra-filtration ProVario-3 (see page 93). The protein was purified with the help of its N-terminal His-tag and submitted to SDS-PAGE. A single band of expected 65 kDa was displayed, but the fraction did not show any activity in the Westerfeld assay (section 2.5.10). Previously isolated total DNA used in PCR screening was submitted to PCR with expression primers as described in the beginning of this section. The resulting fragment was cloned into vector pGEM T easy and sequenced by MWG. Sequencing revealed the expected insert. In order to find out whether the protein was just not active after 4 days at 30°C in the medium and further expression trials might be successful when growing conditions were adapted, the protein was sequenced by Dr.

AHAS from	increase of catalytic activity	reported by
Ath	5.4-fold (recomb.)	Lee and Duggleby 2001
Npl	5-fold (recomb.)	Hershey et al. 1999
Sce *ilv*2, *ilv*6	7 - 12-fold (recomb.)	Pang and Duggleby 1999
Eco AHASIII	30 - 50-fold (native)	Weinstock et al. 1992
	20-fold (recomb.)	Vyazmensky et al. 1996

Ath = *Arabidopsis thaliana*, Npl = *Nicotiana plumbaginifolia*
Sce = *Saccharomyces cerevisiae*, Eco = *Escherichia coli*, recomb. = recombinant

Table 3.40: Increase of AHAS catalytic activity from various species by reconstitution of catalytic and regulatory subunits. Recombinant proteins had been expressed separately.

M. Nimtz, GBF Braunschweig (see page 92). Although the protein had shown the correct size, was soluble, and was cleanable by His-tag purification, it revealed to be no AHAS.

To sum up the results of both expression experiments with the yeast hosts, the recombinant protein production had failed completely. Neither inactive nor active enzyme was harvested. Cloning was checked and confirmed by sequencing. Probably the growing conditions were not suitable or the protein was toxic to the yeast cells. When consulting the literature, only one homologous expression had been performed with *Saccharomyces cerevisiae* host cells expressing the regulatory subunit (*ilv*6) of *Saccharomyces cerevisiae* itself (Pang and Duggleby 1999). Interestingly, the *in vivo* expressed protein had been used solely to determine the cleavage site of the transit peptide, but had not been applied to reconstitution trials with its catalytic subunit. This had been left to *E. coli* heterologously expressed proteins. Hence, a statement about the folding state of this homologously expressed regulatory subunit is not possible.

3.4.3 Co-expression of Regulatory and Catalytic Subunit

There are several examples that expression problems had been solved when, during refolding, the critical proteins had been diluted with its complex partner subunits (REFOLD database, Novagen). These findings are comparable to results from reconstitution trials with AHAS catalytic and regulatory subunits (see Table 3.40). In those reconstitution trials, each AHAS subunit was harvested separately either as native or recombinant enzyme and subsequently mixed with its corresponding partner subunit which increased significantly the AHAScsu catalytic activity.

A further hint had been given by Hill et al. (1997) who had expressed the *E. coli* AHASII from the complete gene *ilv*GM coding for both the catalytic (*ilv*G) and the regulatory subunit (*ilv*M). The two subunits are encoded on the single plasmid pET-GM ensuring simultaneous expression of both proteins in one host cell, probably influencing each other during translation and folding.

In the following section, first, the heterologous expression of both AHAS regulatory subunit sequences from *Eupatorium cannabinum* is described and second, the co-

primer	direction	sequence	restriction enzyme	T_m [°C]
P73	forward	5'-dTAT A**G**G ATC CCA AGC AAG CAA TGG GAC CGT TTC TAC T-3'	*Bam*HI	70.6
P74	reverse	5'-dTAT A**G**T CGA CAA CAG GAT AGG AAT ATC CAC GAA GAT A-3'	*Sal*I	67.2
P76	forward	5'-dTAT A**C**T CGA GAT GAG CTC ACT CTC CAC G-3'	*Xho*I	66.6
P77	reverse	5'-dTAT A**C**C ATG GTT ACA ACG GAA GAG AGT ACC CA-3'	*Nco*I	66.9

restriction recognition sequence is underlined, restriction site is behind bold letters, temperature optimum for restriction enzymes is given in Table B.5

Table 3.41: Primers for expression of the shoot-derived regulatory subunit EcSrsu-AHAShyp (P73 and P74) and of the root-derived regulatory subunit EcWrsu-AHAShyp (P76 and P77) from *Eupatorium cannabinum*.

expression of regulatory and catalytic subunit encoded on one plasmid is shown.

3.4.3.1 Expression of Regulatory Subunits

Heterologous expression of sequence EcSrsu-AHAShyp

The shoot-derived regulatory subunit sequence of *Eupatorium cannabinum* (EcSrsu-AHAShyp), determined in section 3.1.1.3 on page 114, was heterologously expressed in *E. coli*. The primers P73 and P74 (Table 3.41) were positioned as described in section 3.2.4 on page 142 to omit the chloroplast-targeting transit peptide (cTP) and were designed for the use with three different expression vectors: pET22b-23a, pASK IBA7+, and pASK IBA32. With pET22b-23a and pASK IBA 32, a C-terminal His-tag was introduced and ampicillin resistance was conferred. The vector pASK IBA7+ provided the same resistance gene, but introduced N-terminally a Strep-tag for purification of the recombinant protein. Additionally, the vectors pET22b-23a and pASK IBA32 provided the information coding for N-terminal leader peptides pelB and ompA, respectively, directing the recombinant protein into the periplasmic space to enhance protein folding (detailed in sections 2.3.4.2 and 3.4.1.6).

The cloning procedures were the same as described in section 3.4.1.2 except for the following: Total RNA was harvested from shoot tissues, the annealing of the full-length PCR reaction was carried out at 60°C for 5 minutes, the double digest was performed with *Bam*HI and *Sal*I for 4 hours, and the negative ligation products, still containing the unmodified MCS, were eliminated by a restriction into the MCS with *Xho*I. The resulting plasmids were transformed either into JM109 cells for constructs obtained from the vectors pET22b-23a and pASK IBA32 or into BL21(DE3) for constructs from vector

pASK IBA7+. Induction was carried out as described in section 2.4.14.1 with either IPTG (pET22b-23a) or AHTC (pASK IBA). A large scale expression check (100 ml) showed that with constructs from vectors pET22b-23a and pASK IBA32 no protein was detectable neither within the periplasmic (see page 82) nor within the cytoplasmic fraction. As shown for the catalytic subunit (section 3.4.1.6), expression into the periplasmic space was not successful.

On the other hand, expression with pASK IBA7+ resulted in a small band of protein with the expected size of 52 kDa that was detected in the crude cell extract on SDS-PAGE.

Heterologous expression of sequence EcWrsu-AHAShyp

The root-derived regulatory subunit sequence of *Eupatorium cannabinum* (EcWrsu-AHAShyp) was determined in section 3.1.1.4. In contrast to the shoot-derived regulatory subunit, EcWrsu-AHAShyp was expressed in full-length as the prediction for the existence of a cTP was not as reliable as for EcSrsu-AHAShyp. Moreover, it might have been just this subunit which is plastome encoded as it does not contain any transit peptide. Thus, the rsu might be already present in the organelle as reported for the large subunit of the RuBisCO as well. Since the catalytic subunit needs to unfold during transfer across the plastid membrane (detailed in section 3.2.4, Figure 3.14) and needs to fold up again at the plastid stroma side, it can be postulated that the presence of a correctly folded regulatory subunit is guiding the unfolded catalytic subunit to its correct folding.

Full-length PCR was carried out with primers P76 and P77 (see section 3.2.4) that enabled cloning into vector pASK IBA3C and supplied chloramphenicol resistance to its host (section 2.3.4.2). The PCR reaction and enclosed cloning procedures were carried out as described in section 3.4.1.2 except for the restriction reaction carried out by a *Xho*I/*Nco*I double digestion. Sequencing of the designated plasmid EcWrsupIBA3C confirmed the correct ligation. The plasmid was transformed into BL21(DE3) and induced with AHTC (section 2.4.14.1). SDS-PAGE analysis revealed that heterologous protein expression had failed: No additional bands were detected at a size of 53 kDa in the induced culture compared to the not induced culture. It was realized that either the new vector pASK IBA3C produced very low amounts of recombinant protein or the putative transit peptide prevented protein expression.

3.4.3.2 Expression of Regulatory and Catalytic Subunits Encoded on One Plasmid

The plasmid EcScsupET3a (constructed in section 3.4.1.1), encoding the shoot-derived catalytic subunit of *Eupatorium cannabinum,* was chosen for co-expression with the shoot-derived regulatory subunit encoded on plasmid EcSrsupIBA7 (constructed in section 3.4.3.1). As both expression plasmids conferred ampicillin resistance, a proper selection of clones harbouring both plasmids was impossible. Hence, the two expression plasmids were fused together resulting in a single construct and enabling co-expression of both enzyme subunits simultaneously by induction of their respective promoters.

Plasmid EcScsupET3a was linearized with *Bst*1107I, while plasmid EcSrsupIBA7 was cut twice with *Hin*cII which formed a 1,488 bp large fragment that contained the regu-

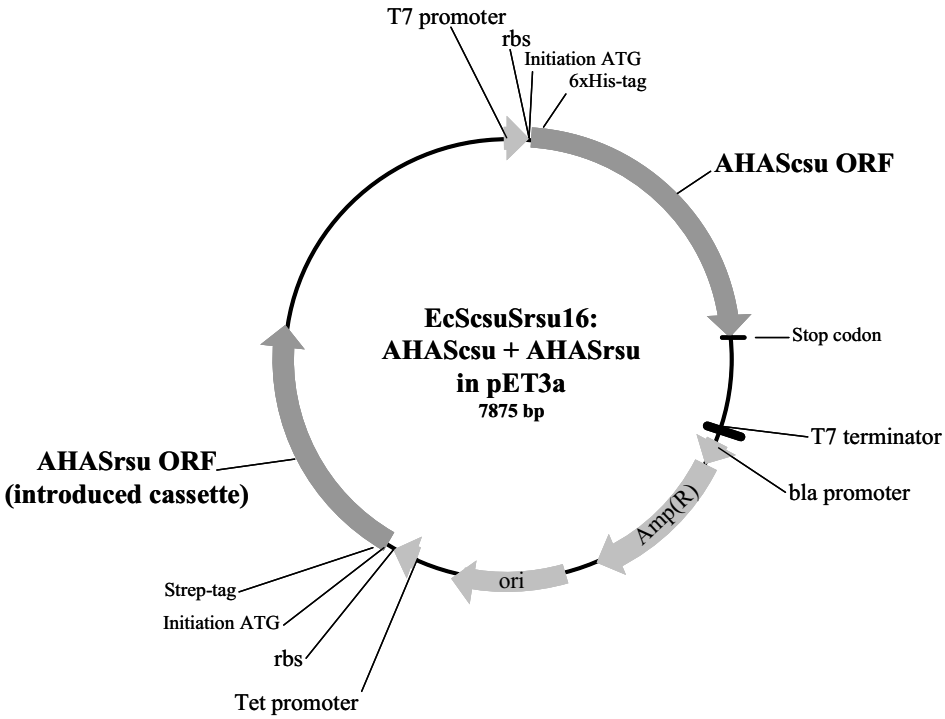


Figure 3.23: Plasmid EcScsuSrsu16 encoding AHAS catalytic and regulatory subunit constructed for co-expression in a single host cell. The plasmid was based on vector pET3a that harbours the catalytic subunit while the insert was taken from vector pASK IBA7+ encoding the regulatory subunit, each supplied with the corresponding promoter region.

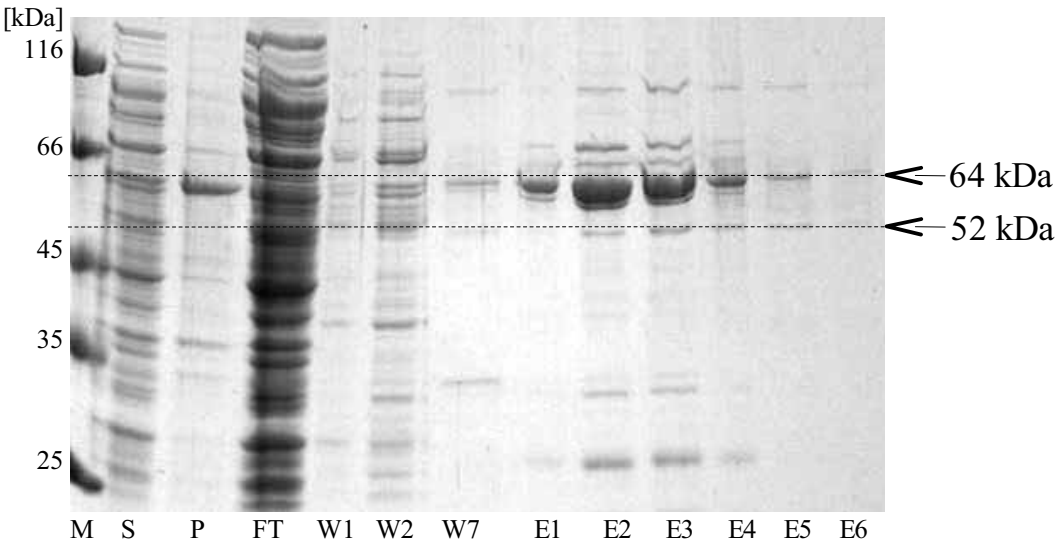


Figure 3.24: SDS-PAGE (12% gel, Coomassie blue stained) documentation of extraction and purification of *Eupatorium cannabinum* AHAS catalytic and regulatory subunits from sequence EcScsu-AHAShyp (64 kDa) co-expressed with the regulatory subunit from sequence EcSrsu-AHAShyp (52 kDa) encoded on plasmid EcScsuSrsu16. Heterologous expression of both sequences resulted in enzymes containing no putative cTP. In elution fractions 2 and 3, a co-purification of the 6xHis-tagged catalytic subunit (64 kDa) with the regulatory subunit (52 kDa, Strep-tagged) was observed, indicating that the interactions between the two subunits are relatively strong. M = marker, S = supernatant , P = pellet, FT = flow through, W = washing fraction, E = elution fraction.

latory subunit coding sequence but missed two nucleotides from the promoter sequence, excluded the transcription terminator and the repressor. The rsu-sequence containing fragment was ligated over night into the linearized plasmid EcScsupET3a harbouring the catalytic subunit. As both restriction enzymes had created blunt ends, the orientation of the introduced cassette was unpredictable. Hence, the orientation of the inserts was determined by two different restriction analyses, firstly with *Kpn*I and secondly with *Bam*HI. In the resulting plasmid EcScsuSrsu16, the rsu and the csu were orientated in the same direction as shown in Figure 3.23.

EcScsuSrsu16 was transformed into BL21(DE3), grown to mid-log phase at 37°C, induced with 1 mM IPTG and 200 ng/ml AHTC (section 2.4.14.1), and grown for 12 hours at 25°C. Recombinant protein was extracted by sonication after lysozyme treatment (section 2.5.1). The catalytic subunit was isolated with His-tag purification. SDS-PAGE analysis of the purification fractions showed that small amounts of the regulatory subunit co-purified with the catalytic subunit (fractions E2 and E3 in Figure 3.24). These findings were in good accordance with expression trials from Hill et al. (1997) who had reported that the recombinant, untagged *E. coli* AHASII regulatory subunit had been capable

of unspecific binding to the nickel matrix which had allowed to purify the native (thus untagged) catalytic subunit through the recombinant regulatory subunit. It was likely that in our case the recombinant AHAS catalytic subunit interacted with the nickel matrix via its 6xHis-tag as expected while the regulatory subunit bound to the catalytic subunit or directly to the nickel matrix as described by Hill et al. (1997) for the *E. coli* enzyme subunits.

Elution fraction E1 to E5 were pooled, desalted, and transferred to the assay buffer (section 2.5.10, page 93). The recombinant enzyme was diluted 1:10 with reaction cocktail 1 and submitted to UV detection as described in section 2.5.10.1. The mass of nearly purified protein was determined by the Bradford (1976) method (section 2.5.4) and the resulting specific activity was calculated with 3.4 nkat/mg for the turnover of pyruvate. Compared to specific activities from other plants, this was within the wide range from 0.7 nkat/mg (*Arabidopsis thaliana* csu GST-tag cleaved, Hershey et al. 1999) to 833 nkat/mg from the same plant species determined with an uncleaved GST-tag determined by Ott et al. (1996) and it differed from the result of 0.0205 nkat/mg determined by Weber (1997) for the native AHAS from *Eupatorium clematideum*. Probably, the native enzyme from *Eupatorium clematideum* had lost most of its activity due to a loss of its quarternary structure caused by enforced purification treatments.

For the analysis of the reaction product by GC-MS (section 2.5.10), the recombinant enzyme was diluted with reaction cocktail 2 that contained 1 part pyruvate mixed with 3 parts of 2-oxoisovaleric acid. The enzyme reaction was performed according to section 2.5.10 and was stopped with concentrated HCl. The resulting mixture was treated with diazald (produced as shown in Figure 2.12) to methylate all carbonic acids to esters. The esters were extracted four times with ethyl acetate, concentrated, and submitted to GC-MS analyses. As shown in Figure 3.25, the expected "pro-necic" acid 2-hydroxy-2-isopropyl-3-oxobutyric acid methyl ester had been produced by the shoot-derived AHAS catalytic subunit from *Eupatorium cannabinum*. For the first time, it was shown that the recombinant AHAS catalytic subunit is able to catalyze the first step in the necic acid biosynthesis of lycopsamine type pyrrolizidine alkaloids.

Moreover, the recombinant AHASII from *E. coli* (section 2.3.4.4) was submitted to the GC-MS detected assay as well. The same expression and purification procedures were applied as described for the plant AHAS and showed that the *E. coli* enzyme was capable of synthesizing the "pro-necic" acid as well. This extended substrate specificity was not unexpected since the *E. coli* AHASI had shown properties to accept quite a number of related substrates like aromatic or cyclic carboxaldehydes, namely benzaldehyd (Engel et al. 2003) and its monosubstitutes with ortho-, meta-, or para-positioned Cl-, Me-, Me-O-, and CN-groups, 1- or 2-naphthaldehydes, pyridine-2-, -3-, or -4-carboxaldehydes, furan-2- or -3-, thiophen-2- or -3-, and pyrrole-2-carboxaldehyds, phenylacetaldehyd, and cyclohexane carboxaldehyde (Engel et al. 2005) producing chiral hydroxyketones in almost enantiomeric purity (Engel et al. 2004a).

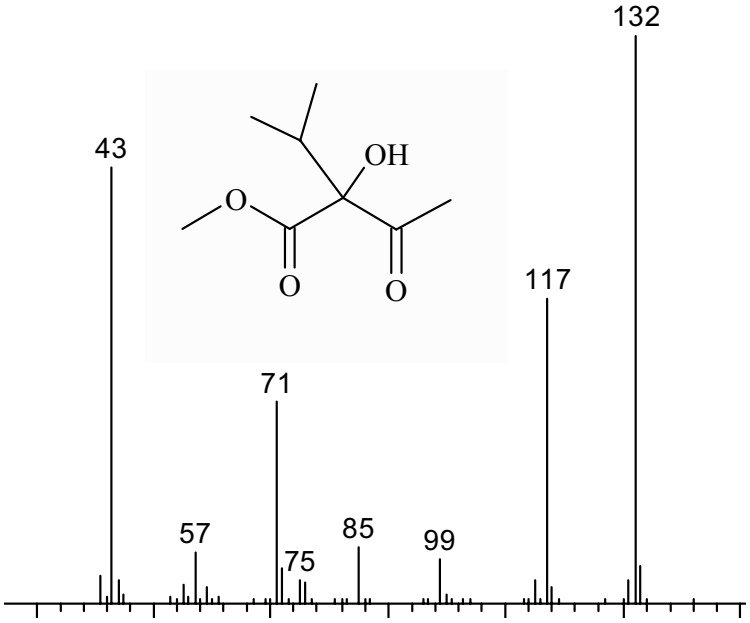


Figure 3.25: GC-MS-identified methylated "pro-necic" acid (2-hydroxy-2-isopropyl-3-oxobutyric acid methyl ester) at RI 17.12 obtained in a reaction of *Eupatorium cannabinum* recombinant AHAS catalytic subunit with pyruvate and 2-oxoisovalerate as substrates.

4 Discussion

Species of several unrelated families within the Angiosperms are able to constitutively produce pyrrolizidine alkaloids (PA). This scattered occurrence (Figure 1.1) provokes the question about the evolutionary origin of PA biosynthesis: Was the ability to synthesize PAs invented only once early in the evolution of the Angiosperms but lost in several unrelated lineages or was this ability invented several times independently? Though, this question cannot be answered for the whole biosynthetic pathway leading to PAs yet, a comparison of the first specific enzyme of PA biosynthesis, the homospermidine synthase (HSS), and of its ancestor, the deoxyhypusine synthase (DHS), allowed the conclusion that at least, the HSS was invented four times independently, once early in the evolution of the Boraginaceae, once within the monocots, and even twice within the Asteraceae splitting off the tribes from Senecioneae and Eupatorieae (Reimann et al. 2004). The extension of knowledge about the mono- or polyphyletic origin of further enzymes involved in the PA biosynthesis was the fundamental motivation to start this work.

PAs are known to consist of two moieties: the necine base and the necic acid(s). Since the HSS catalyzes the first step in the necine base biosynthesis, it was our motivation to elucidate the first specific step in the biosynthesis of the necic acid part of the PAs. Mechanistically, an acetohydroxyacid synthase (AHAS) was determined by Weber et al. (1999) as a candidate for catalyzing the formation of the unique necic acids which are characteristic for the lycopsamine type PAs. In general, AHAS from Angiosperms display a high degree of sequence identity coupled with a strongly preserved biochemical function (Duggleby and Pang 2000, Chipman et al. 1998) which parallels the features of the DHS system (Gordon et al. 1987, Bartig et al. 1992).

Plant AHAS amino acid sequences are identical to each other in a range from 69% to 75% (section 3.2.2 shown for catalytic subunits). The same degree of identity was found for the DHS/HSS protein family, e.g., the HSS of *Phalaenopsis* species (Orchidaceae, monocot) shares an identity of 74% with the DHS from *Crotalaria retusa* (Fabaceae, dicot) as found by Nurhayati (2004), the HSS of *Senecio vernalis* (Asteraceae, tribe Senecioneae, dicot) shows a sequence identity of 71% with the DHS from *Musa acuminata* (Musaceae, monocot) as detailed by Reimann et al. (2004).

In order to outline the strongly preserved biochemical function of DHS and HSS, the following details are noteworthy:

The DHS catalyzes the first step in the post-translational activation of the eukaryotic initiation factor 5A (eIF5A) by transferring the aminobutyl moiety of spermidine to a specific lysine residue of the eIF5A precursor protein (Park et al. 1997). The mode of action of the DHS is very well conserved which is shown by the fact that the eIF5A substrate protein of a given eukaryotic species is accepted as a substrate by the DHS enzymes from any other eukaryotic species. The same mode of action is found for the HSS which transfers the aminobutyl moiety of spermidine to putrescine, yielding homospermidine,

the first intermediate in the necine base biosynthesis (Ober and Hartmann 1999a, Ober et al. 2003b). The main difference between DHS and HSS is the fact that the DHS is capable to transfer the aminobutyl moiety of spermidine to eIF5A as well as to putrescine whereas the HSS only catalyzes the latter reaction. As the DHS was found to be the ancestor of HSS, the latter is believed to be recruited by gene duplication followed by neofunctionalization, i.e. the recruitment for PA-biosynthesis which in more detail means the loss of the binding capability for eIF5A (Moll et al. 2002) and a defined tissue-specific expression (Anke et al. 2004). This rare events are interpreted as "evolution by change of function" (Ober and Hartmann 2000). The vast majority of gene duplicates ends up as pseudogenes (Pichersky and Gang 2000, Nurhayati and Ober 2005).

The strongly preserved biochemical function of the AHAS lies in catalyzing the first step of branched chain amino acid biosynthesis. The AHAS transfers an active acetaldehyde to either pyruvate or 2-oxobutyrate which results at the end of a multistep biosynthesis in the formation of valine/leucine or isoleucine (section 1.1.2, Figure 1.9). Like the DHS, the AHAS is capable of accepting two different substrate molecules. When the DHS/HSS model for the necine base biosynthesis is transferred to the situation for the necic acid biosynthesis, the AHAS should resemble the DHS, while an AHAS-like NAS should resemble the HSS. The postulated necic acid synthase (NAS) should display a reduced substrate spectrum compared to the ancestor AHAS. Unlike the DHS, the two substrates recognized by the AHAS result in products of primary metabolism, whereas the DHS only forms one product in primary metabolism and the synthesis of homospermidine just resembles an uncontrollable side-reaction. Hence, the AHAS is postulated to catalyze the reaction with the proposed substrate in necic acid biosynthesis, namely the 2-oxoisovaleric acid, as an uncontrollable side-reaction as well, whereas the NAS is predicted to recognize the 2-oxoisovaleric acid exclusively. The AHAS is proposed to be a good candidate for an additional case of gene duplication followed by neofunctionalization. Thus, the molecular and biochemical proof that an AHAS-like NAS is capable to perform the first step in necic acid biosynthesis was to aim in this work.

Recently, the pathway for branched chain amino acid biosynthesis (primary metabolism) had been shown to provide further enzymes for secondary metabolism: The isopropylmalate synthase (IPMS) from the late steps of the leucine biosynthesis is the ancestor of the IPMS-like methylthioalkylmalate synthase (MAM) catalyzing the first step in aliphatic glucosinolate biosynthesis (De Kraker et al. 2007, Textor et al. 2007). Obviousely, enzymes from the branched chain amino acid biosynthesis are likely to have been recruited for secondary metabolism more than once.

4.1 Sequence Detection and Comparative Analysis of AHAS

In this work, cDNA sequences coding for AHAS catalytic and regulatory subunits were identified, heterologously expressed, and biochemically characterized from PA-producing plants. Angiosperm families chosen for these analyses are all characterized by lycopsamine

type PAs though they are not closely related to each other: Asteraceae, tribe Eupatorieae (*Eupatorium cannabinum*), Boraginaceae (*Symphytum officinale*), and Apocynaceae (*Parsonsia laevigata*). Sequences were identified with a combined method of degenerate oligonucleotide primed PCR and RACE-PCR.

4.1.1 Sequence Detection of AHAS Catalytic and Regulatory Subunits

During sequence detection, especially the method of the 5'RACE technique created several problems and the identification of the full-length cDNA sequences was hampered by shortened mRNA templates. This was probably due to an increased occurrence of secondary structures and was overcome by the application of increased temperatures during reverse transcription and the use of a *Taq* DNA polymerase with integrated proof-reading activity during PCR. Very rare templates could be multiplied in the 5'RACE-PCR when the primer used in reverse transcription was re-used in the first PCR of the nested PCR procedure. Hypothetical reasons why this worked just for this type of primer may be found in an increased primer concentration due to remnants from the reverse transcription.

In case of the regulatory subunits, we report the discovery of two cDNAs from the PA-producing plant *Eupatorium cannabinum*. These two cDNAs represent 50% of the worldwide heterologously expressed regulatory subunits of plant AHAS. The other two were identified and expressed from *Nicotiana plumbaginifolia* (Hershey at al. 1999) and *Arabidopsis thaliana* (Lee and Duggleby 2001). Due to the conserved sequences within the Angiosperms the degenerate primers designed in this work (Table 3.9, page 115) should allow the discovery of regulatory subunits from many more plants of interest.

Although designed independently, one pair of degenerate primers for the catalytic subunit was closely positioned to those from Bernasconi et al. (1995). This was due to a large number of highly conserved residues within the AHAS sequence that are encoded by a high number of codons (high degeneracy) and, consequently, were inappropriate for degenerate primer design. Nevertheless, a second pair of degenerate primers for detection of catalytic subunits was designed and successfully applied in this work. This pair of primers had the major advantage to deliver a large stretch of 1,110 bp which represents more than 55% of the nearly 2,000 bp comprising ORF and allows, generally speaking, an accelerated detection procedure.

4.1.2 Sequence Comparison of Plant AHAS Catalytic Subunits

For *Eupatorium cannabinum*, three putative catalytic subunit sequences were detected, whereas for *Symphytum officinale* two sequences were identified and for *Parsonsia laevigata* only one was discovered. The sequences were compared to each other and further plant AHAS (section 3.2). The deduced amino acid sequences showed that all new sequences had the necessary structural features for functionality (Table 4.1 on page 182). Two sequences displayed some interesting deviations from highly conserved residues:

role of residues according to McCourt and Duggleby (2006)	residue	mutation (in sequence)
catalytic center protonation of N1' in ThDP	E144	none
substrate recognition interaction with the carboxylate group of the ketoacid and	R377	none
maintenance of the correct orientation for ionic interaction by flanking R377	F206[2], M351[2]	none
proton transfer to the carbonyl oxygen of the second substrate	Q207	E208D (EcW2csu)
stabilization of the negative charge	G121[2]	none
resistance to imidazolinone herbicides	A122	A122N (EcW2csu)
recognition of the second substrate	W574[2], M570[2]	none
ThDP co-factor binding motif GDGX$_{24-27}$NN[1]	G539-N565	G541S (SoScsu)
imposing the V-conformation	M513	none
pyrimidine ring	Y118', P170', M513	none
4'-amino group	E144', Q207'	none
N1'	G120'	none
thiazolium ring and methylene groups	G120', M513, L568, M570, V571	V571A (SoWcsu)
diphosphate	D538, Q487, H488, G539, S540, H567, G569	H567Y (EcW2csu)
diphosphate anchored to Mg^{2+}	D538, N565, H567	H567Y (EcW2csu)
FAD binding site	G350, W491	none
mobile loop border	H567	H567Y (EcW2csu)
structured C-terminus, hypothesized	S653	S653P (EcW2csu)

[1] Hawkins et al. 1989, [2] Engel at al. 2004b,

Table 4.1: Important residues and their role in AHAS, residue numbering according to *Arabidopsis thaliana* as proposed by Tranel and Wright (2002)

1. SoWcsu-AHAShyp from root tissues of *Symphytum officinale* contained an insertion of the 4 amino acids "PNWG" behind V264 (*Arabidopsis thaliana* numbering favored by Tranel and Wright (2002), used in the following comparative analysis) which is located at the end of the α-domain as deduced from the crystal structure of *Arabidopsis thaliana* recently determined by McCourt et al. (2006). An 2D-model of the AHAS catalytic subunit is shown in Figure 4.1. This amino acid insertion may widen the substrate access funnel and influence activity and/or substrate specificity. The substrate access funnel is created in between all three domains of the catalytic subunit. It is hold open by an FAD molecule that does not interfere with the reaction but is crucial for enzyme activity (Duggleby and Pang 2000). Its role might be compared to a coronary stent. On day-light exposure, FAD decays. Hence, a loss of activity has been observed when the purified AHAS has not been protected from light (Chang and Duggleby 1997).

2. EcW2csu-AHAShyp from root tissues of *Eupatorium cannabinum* offered two probably critical mutations: H567Y and S653P (Figure 4.1). H567 is one of the three amino acids which anchor the essential magnesium ion that for its part fixes the thiamine diphosphate molecule (ThDP) which is the center of catalytic activity (McCourt and Duggleby 2006). Although the ThDP interferes with residues from the α-domain of a second catalytic subunit and, therefore, two catalytic subunits form an intimate dimer containing two active sites, each of which lying across a dimer interface and involving both monomers (Figure 4.3), this mutation may offer an altered position of the magnesium ion and may give space for an altered substrate or it probably may be fatal for activity. S653P may bee seen from the crystal structure of *Arabidopsis thaliana* to interact with the mobile loop that had been proposed to close the substrate access funnel during enzyme reaction in order to exclude the solvent which otherwise would interfere with the thiamine diphosphate and inhibit the original reaction (Pang et al. 2003). The mutation is critical because it inserts a proline at a position where the mobility of the loop is necessary to release the reaction product. Proline is the only amino acid that prevents free rotation around the peptide binding in amino acids as the α-amino group of proline is part of the pyrrolidine ring. Hence, this mutation will either keep the substrate access funnel permanently open or closed, in either case the enzyme reaction may be hampered spatially.

Sequence identities were calculated with the help of the bestfit programme in the GCG software package (Table 3.24 on page 133). Remarkably, sequence EcW3csu from *Eupatorium cannabinum* showed higher identities to *Xanthium* species (90%) and *Helianthus annuus* (89%), both Asteraceae, tribe Heliantheae, than to sequences EcW2csu (74%) or EcScsu (81%) both from *Eupatorium cannabinum*, Asteraceae, tribe Eupatorieae.

A neighbour joining tree (Figure 4.2 on page 186) showed that all 48 Angiosperm AHAS catalytic subunits cluster nicely in family branches. Sequence EcW3csu was found to cluster together with all other AHAS sequences of the Asteraceae species available from the database (highlighted in green in Figure 4.2). As these sequences are characterized by only short branch lengths, they are obviousely highly conserved in function and might represent the "housekeeping" AHAS involved in branched chain amino acid biosynthesis. Thus, it might be postulated that sequence EcW3csu shows "housekeeping" functions as

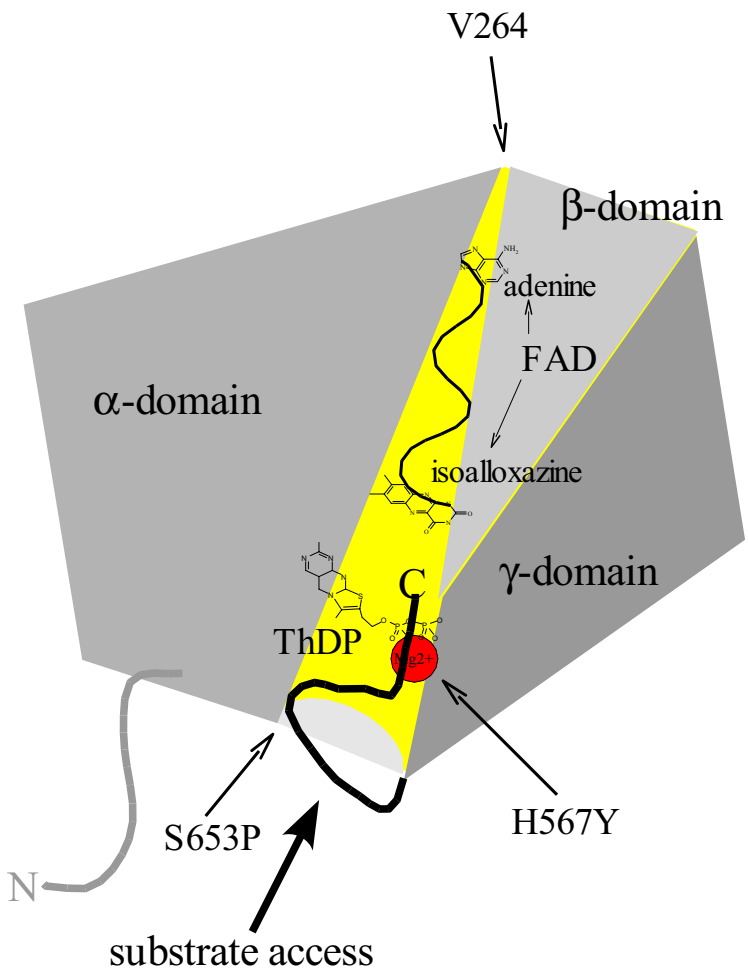


Figure 4.1: 2D-model of AHAS catalytic subunit modified from the original crystal structure (McCourt et al. 2006). α-, β-, and γ-domain are grey polygons, the substrate access tunnel is highlighted in yellow harbouring the active center (ThDP = thiamine diphosphate) that is anchored by a magnesium ion (red disc), whereas FAD is located at the end of the tunnel to keep the latter in shape. The C-terminal tail loops over the active site during catalysis. Marked mutations are described in the text in more detail.

well.

When compared to EcW3csu or to the other Asteraceae sequences, EcScsu and EcW2csu showed elongated branches indicating an increased rate of sequence substitutions, discussed as a feature of sequences that evolved new functions and are under modified selection pressure. Characteristically, such enlarged branch lengths were observed either for the HSS/DHS-system (Reimann et al. 2004) playing a key role in PA-biosynthesis and for ATPases with an altered substrate specificity (Axelsen and Palmgren 1988). Such gene duplication events had been shown to result in relaxed selective constraints on the gene copy supplying altered substrate specificity.

4.1.3 Tissue-Specific Transcript Localization

By tissue-specific transcript localization analyses (section 3.2.3, page 132), the abundance of sequence EcW2csu-AHAShyp was shown to be restricted to the roots. Despite the sequence mutations (discussed in section 4.1.2), this finding is a strong hint that EcW2csu-AHAShyp might be a good candidate to be the AHAS-like necic acid synthase (NAS) since the biosynthesis of the necine base producing homospermidine synthase of *Eupatorium cannabinum* was immuno-localized by Anke (2004) in the cortical parenchyma of roots. These findings are very remarkable as AHAS genes expressed in a non-tissue-specific and constitutive manner are known to function as housekeeping genes (Bna1, Grula et al. 1995). In *Brassica napus*, the mRNA (Bna3) of functionally distinct AHAS genes had been detected only in reproductive tissues (Ouellet et al. 1992) which might lead to the conclusion that a tissue-specific abundance of AHAS mRNAs occurs simultaneously with an altered function of AHAS genes.

4.1.4 Proof of Functionality with the *In vivo* Activity Test and of Activity with the Cell-Free Expression System

For functional characterization, it was decided to focus on the three identified cDNA sequences of *Eupatorium cannabinum*. It was postulated to detect at least two different AHAS enzymes, one involved in primary metabolism and a second involved in PA biosynthesis.

The proof of functionality was performed prior to intensive, heterologous expression studies ensuring to work with no pseudogenes, which are the more common products of gene duplication than neofunctionalized enzymes. On the other hand, pseudogene existence within the genome is, in general, of temporary nature only, due to a rapid decay by mutations. In case of DHS from *Crotalaria* species pseudogenes were found rather often (Nurhayati and Ober 2005) reproving for carefulness with PA-specific enzymes.

In the *in vivo* activity test, functionality of sequences was tested with the help of the *E. coli* mutant strain MF2000, kindly provided by Dr. B. Mazur, Dupont. It is devoid of AHAS activity due to mutations in all three native *E. coli* AHAS genes (section 3.2.5, page 144). The sequence EcScsu-AHAShyp was chosen as a model sequence for all putative AHAS catalytic subunits from *Eupatorium cannabinum*. Functional complementation of

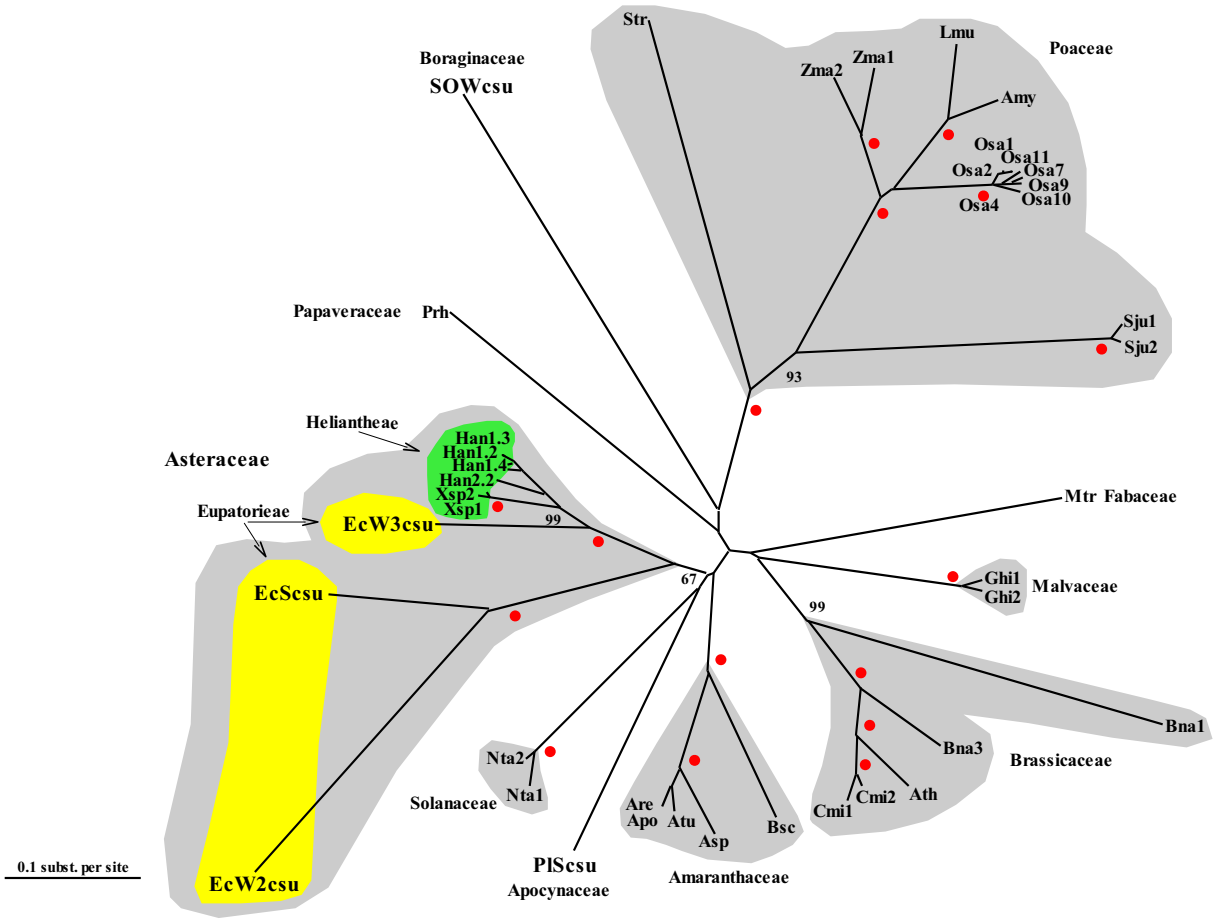


Figure 4.2: Unrooted neighbour joining tree based on 48 amino acid sequences of AHAS catalytic subunits from various Angiosperm species including the novel, full-length putative AHAScsu from PA-producing plants. The tree was calculated with PHYLIP V 3.6 as described in section 2.4.2 on page 56. Bootstrap supports resulting from 1,000 replicates are given below branches, red dots are shown for bootstrap values of 100%. csu = catalytic subunit, EcS = *Eupatorium cannabinum*, shoot-derived, EcW2 = *Eupatorium cannabinum*, root-derived, seq. no. 2, EcW3 = *Eupatorium cannabinum*, root-derived, seq. no. 3, all three of them are Asteraceae, tribe Eupatorieae, PlS = *Parsonsia laevigata* shoot derived, Apocynaceae, SoW = *Symphytum officinale* root derived, Boraginaceae, Amy = *Alopecurus myosuroides*, Poaceae, Apo = *Amaranthus powellii*, Are = *Amaranthus retroflexus*, Asp = *Amaranthus* spec., all Amaranthaceae, Ath = *Arabidopsis thaliana*, Brassicaceae, Atu = *Amaranthus tuberculatus*, Amaranthaceae, Bsc = Bassia scoparia, Chenopodiaceae, Bna = *Brassica napus*, Cmi = *Camelina microcarpa*, both Brassicaceae, Ghi = *Gossypium hirsutum*, Malvaceae, Han = *Helianthus annuus*, Asteraceae, tribe Heliantheae, Lmu = *Lolium multiflorum*, Poaceae, Mtr = *Medicago trunculata*, Fabaceae, Nta = *Nicotiana tabacum*, Solanaceae, Osa = *Oryza sativa*, Poaceae, Prh = *Papaver rhoeas*, Papaveraceae, Str = *Sagittaria trifolia*, Sju = *Schoenoplectus juncoides*, both Poaceae, Xsp = *Xanthium* spec., Asteraceae, Zma = *Zea mays*, Poaceae.

the *E. coli* mutant strain was achieved and enabled growth on minimal selective medium. Hence, the discovered putative cDNAs identified themselves as functional AHAS.

As discussed in section 4.1.2, sequence EcW2csu-AHAShyp contained two mutations that might be fatal for activity. Thus, a cell-free expression and enclosed activity assay for the substrate pyruvate were performed to exclude that EcW2csu-AHAShyp was a pseudogene (section 3.2.6, page 145).

All three catalytic subunits showed activity and undoubted proved functionality. Moreover, the simultaneous expression of catalytic and regulatory subunits yielded a 2.4-fold increase in activity for the protein from sequence EcScsu-AHAShyp, a 2.5-fold increase for the enzyme from sequence EcW3csu-AHAShyp, and a 3.2-fold increase in activity for the enzyme from the only root-localized and mutated sequence EcW2csu-AHAShyp. The finding that activity was increased when catalytic and regulatory subunits were united was in good accordance with the results from reconstitution experiments from Hershey et al. (1999) and Lee and Duggleby (2001). They had been able to identify, heterologously express, and purify the regulatory subunits from *Nicotiana plumbaginifolia* and *Arabidopsis thaliana*, and the reconstitution of the regulatory subunits with its corresponding catalytic subunits resulted in 5- and 5.4-fold increase in specific activity, respectively. Interestingly, the reconstitution of *Nicotiana plumbaginifolia* regulatory subunit with the catalytic subunit from *Arabidopsis thaliana* was possible as well and showed a 3-fold increase which confirmed the strongly preserved biochemical function of the AHAS-system.

4.2 Heterologous Expression of AHAS

AHAS from plants are the target enzymes of several advanced herbicides with no or little side effects on humans. This is due to the absence of AHAS-like enzymes in mammals. Intensive studies about the effects of herbicides on AHAS in plants resulted in a huge number of publications in this field.

As early as 1987, Mazur et al. (1987) had identified the first plant-derived cDNA putatively coding for a catalytic subunit of AHAS in *Arabidopsis thaliana*. While the first eukaryotic sequence from *Saccharomyces cerevisiae* had been heterologously expressed in *E. coli* in 1989 by Poulsen and Stougaard (1989), the first purified recombinant AHAS from plants was not described before 1995, when Bernasconi et al. (1995) expressed the full-length AHAS from *Xanthium* spec. including its cTP in pGEX-2T which had fused a glutathione *S*-transferase N-terminally to the enzyme. Meanwhile the sequences from *Arabidopsis thaliana* (Smith et al. 1989), *Brassica napus* (Wiersma et al. 1990), and *Nicotiana tabacum* (Kim and Chang 1995) had been functionally expressed and purified from AHAS deficient mutants of *E. coli* (strain MF2000 applied on *A. thaliana* and *N. tabacum*) or *Salmonella typhimurium* (strain DU2603 applied on *B. napus*).

For overexpression, the decision had to be taken which system, presented in the literature, was more reliable: the pGEX-2T-expression vector system which had been established by Bernasconi et al. (1995) and later was described in four more reports (Ott et al. 1996, Chang et al. 1997, Hershey et al. 1999, Oh et al. 2001), or the pET-expression

vector system which had been established by Chang and Duggleby (1997) for the catalytic subunit from *Arabidopsis thaliana* AHAS. Despite of this, it had to be decided whether the chloroplast-targeting transit peptide (cTP) should be omitted or not. In order to name the two most contradictory reports, Bernasconi et al. had strongly recommended to keep the cTP from *Xanthium* spec. while Chang and Duggleby (1997) had determined that the starting sequence of the mature protein for *Arabidopsis thaliana* is "TFISXFAPDQ" after N-terminally omitting 86 amino acids. For a detailed analysis of the literature concerning the transit peptide in AHAS catalytic and regulatory subunits, please refer to section 3.2.4 starting on page 137.

The pET-expression vector system was chosen for the first heterologous expression trials with only the catalytic subunits (csu) since the csu had proved activity when expressed with a cell-free method (section 3.2.6), since, in general, catalytic subunits are known to be the only place of enzyme activity.

On the other hand, regulatory subunits are known to provide sensitivity to feed-back inhibition to the catalytic subunit. Thus, the question was left open wether a regulatory subunit was responsible for substrate specificity of an AHAS-like enzyme from PA biosynthesis. Former HEH inhibiting trials (Böttcher et al. 1993, Frölich et al. 2007) blocking the necine base biosynthesis had shown that the content of necic acids within the PA-producing plant is not increased in comparison to non-inhibited cultures. The deduced conclusion was that HEH might probably inhibit the necic acid biosynthesis as well. Taking the feed-back inhibition facilities of the AHAS regulatory subunit into account, an alternative scenario might be considered: The AHAS-like enzyme from PA biosynthesis might be feed-back inhibited via its regulatory subunit by the smallest amounts of free necic acid. Moreover, this assumption fits to the observation that free necic acids are not accumulated in plants with the one exception of (-)-viridifloric acid in fruits of *Lappula echinata*, Boraginaceae (Wang et al. 1986).

Standard expression trials with the strong T7 promoter in the pET vector system ended up in insoluble and thus inactive enzymes (section 3.4.1.1). The reasons for this could have been the sequence itself (wrong length of the omitted signal sequence), could have been the result of improper handling of the enzyme (during expression: temperature, type and pH of the medium, extraction method, purification method with inadequate binding conditions, storage buffer, exposure to light, freezing), or could have been a problem of the expression system (protein expression promoter claims much of the host cell resources). As stated by Ausubel et al. (1994), "protein expression is an inexact science at present". Nevertheless, all possibilities were thoroughly tested (summarized in Table 4.2) with the result that the AHAS from *Eupatorium cannabinum* only folds correctly when catalytic and regulatory subunits are simultaneously induced and co-expressed in one *E. coli* host cell.

The deduced general hypothesis of these results is that the regulatory subunit is not only responsible for extended enzyme stability, increased specific activity, and regained feed-back inhibition facilities (when compared to the native enzyme, Lee and Duggleby 2001), but it is necessary for the correct folding of the catalytic subunit. The regulatory subunit is not necessary for the enzyme catalysis in itself as the purified catalytic subunit

variation	**effect**
expression in *E. coli*:	
expression hosts: BL21(DE3), ER2566	insoluble
media type: LB, M9, M9 half concentrated	insoluble
pH of LB media:	
7.5	insoluble
8.5	soluble, but inactive
amount of inducing IPTG [mM]: 0.1, 0.5, 1.0	insoluble
growing temperature [°C]:	
37, 30	insoluble
25, 16	soluble, but inactive
growing time [h]: 2, 4, 6, 9, 12, 18, 28	insoluble
way of cell lysis:	
sonication	insoluble
lysozyme + sonication	soluble, but inactive
CelLytic	soluble, but inactive
addition of ethanol during expression	insoluble
refolding of isolated inclusion bodies	soluble, but inactive
plant cTP was not omitted	no expression
plant cTP was omitted	insoluble
His-tag was placed N-terminally	insoluble
His-tag was placed C-terminally	insoluble or no expression
tags and leaders:	
Intein-, GST-, Strep-tag, ompA-leader	insoluble
promoters: T7, *tac*, *tet*	insoluble
co-expression with chaperonins	soluble, but inactive
expression in Yeast:	
pYES2 in *S. cerevisiae*	no expression
pKLAC1 in *K. lactis*	no expression
checking functionality (MF2000)	functional AHAS
cellfree	active
co-expression with rsu	soluble and active

Table 4.2: Listing of performed expression trials

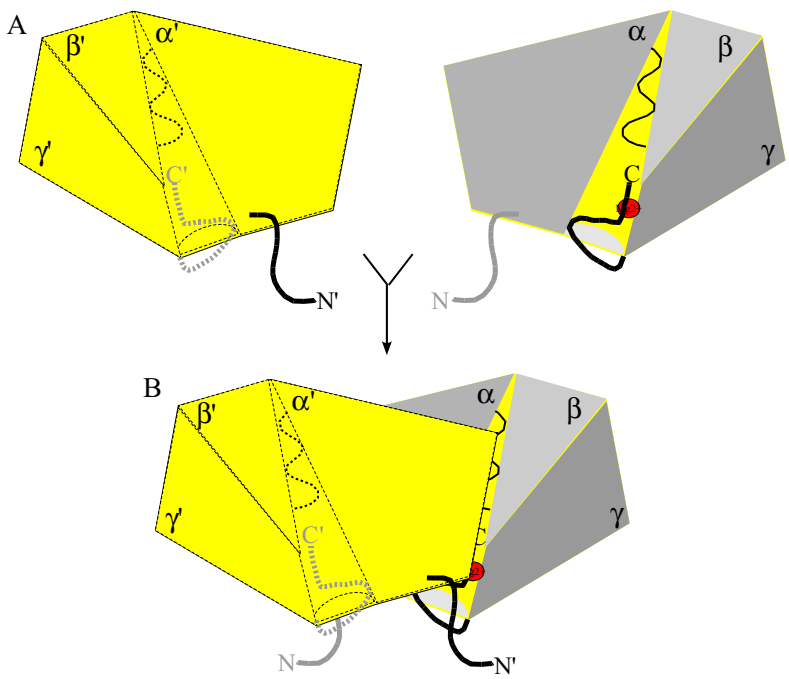


Figure 4.3: Two catalytic subunits form together the intimate dimer: The catalytic centers lie across the dimer interface. Catalytic domains highlighted in yellow are facing the backside, while catalytic domains highlighted in grey show the front side leaving open the substrate access funnel resulting in yellow and harbouring the active compound ThDP that is anchored by a magnesium ion (red disc), whereas FAD (shown as snakelike line) is located at the end of the funnel to keep it open.

was still active, but the regulatory subunit may help to maintain the optimal positions of the two catalytic subunits which form together the intimate dimer where the two catalytic centers lie across the dimer interface (Figure 4.3).

The necessity of the regulatory subunit for an optimal quarternary structure is supported by the following reports from the literature:

1. The native AHAS from *Arabidopsis thaliana* had shown a size of 548 kDa in gel-filtration (Lee and Duggleby 2001). The complex had been proposed to consist of four or five copies for each subunit. The size of the catalytic subunits had been determined with 65.1 kDa whereas the regulatory subunits had been 52.4 kDa. It occurred to our mind that a more precise composition of the quarternary structure was deducible from these results: As shown in Panel B of Figure 4.4, the proposed 6 catalytic subunits (3 x 2) may be tied together with 3 regulatory subunits which results in a calculated size of 547.8 kDa very close to the size of the native enzyme. Hence, the heteromeric complex AHAS from *Arabidopsis thaliana* is hypothesized to have an $\alpha_6\beta_3$ structure leading to a general $\alpha_{2n}\beta_n$ AHAS heteromeric pattern.

2. The regulatory subunit belongs to the ACT domain-containing proteins that

represent a growing family with different intracellular small molecule binding domains that function in the control of metabolism, solute transport, and signal transduction (Aravind and Koonin 1999). These ubiquitous proteins were detected in crystal structures of *E. coli* D-3-phosphoglycerate dehydrogenase (protein data bank code 1PSB), *Mycobacterium tuberculosis* ATP phosphoribosyltransferase (1NH8), and *E. coli* threonine deaminase (1TDJ). In these enzymes, the ACT domain is placed to stabilize the quarternary structure of the enzymes (Grant 2006).

Probably, these findings about the role of the regulatory subunit will inspire the work on the solvation of the crystal structure of plant AHAS holoenzymes.

4.3 Specific Assay for Necic Acid Synthase Activity

The AHAS-like necic acid synthase (NAS) in PA-producing plants is postulated to show a broadened or altered substrate specificity when compared with an AHAS from non-PA-producing plants just serving in primary metabolism (branched chain amino acid biosynthesis). Since PA-producing plants are equipped with AHAS from primary metabolism as well, a clear differentiation is necessary to identify AHAS-like NAS as the sequence characteristics might not be sufficient for identification. As detailed in section 3.3 on page 148, it was not possible to distinguish between an AHAS reaction transforming pyruvate (valine/leucine biosynthesis), 2-oxobutyrate (isoleucine biosynthesis), or 2-oxoisovalerate (necic acid biosynthesis) with the commonly applied Westerfeld assay (UV detection). Thus, a GC-MS based detection procedure was used with references synthesized specifically for this purpose.

In the future, the specific assay for NAS activity should be independent from MS detection for routine application. ^{14}C-labelled substrates might be fed to the purified, recombinant enzyme: For determination of primary metabolic activity, ^{14}C-labelled pyruvate is available, while for determination of the product from secondary metabolism, 2-hydroxy-2-isopropyl-3-oxobutyric acid, ^{14}C-labelled 2-oxoisovalerate is synthesized as reported in section 3.3.1 on page 93. The resulting products should be stabilized by methylation with diazald (section 2.5.10, Figure 2.12). Separation on TLC silica 60 F_{254} plates should be carried out with the mobile phase petroleum ether/ethyl acetate (2:1) that was determined during synthesis of the reference substances (section 2.6). Detection might be performed with Rapid Intelligence TLC Analyzer (RITA) as described in section 3.3.1 or with the help of phosphomolybdic acid reagent (see page 98). Rf-values of the products from primary (1.) and secondary (2.) metabolism are given in Table 4.3 as determined in section 2.6.

4.4 Perspectives

Future trials have to check activity with properly purified, co-expressed holoenzymes of AHAS-like NAS. The complete biochemical characterization should reveal all properties

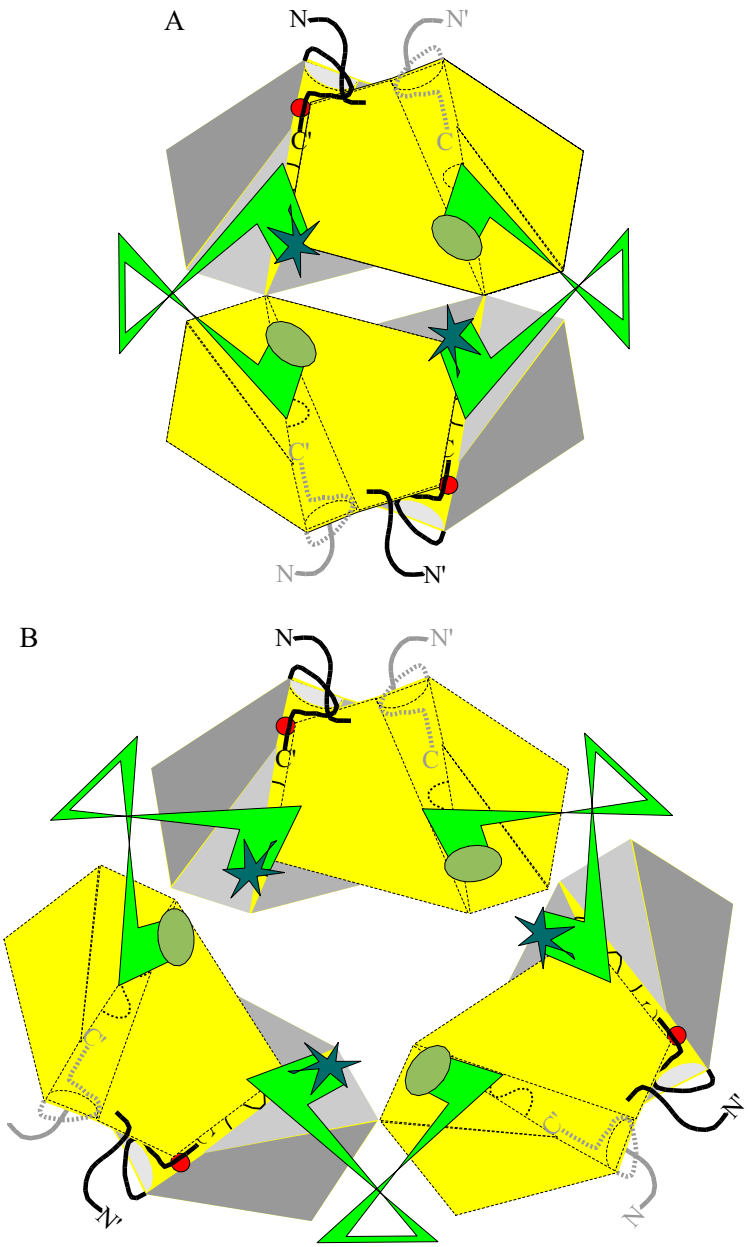


Figure 4.4: **Panel A** Proposal of the smallest possible quarternary structure of plant AHAS holoenzyme: Two catalytic subunits (csu), highlighted in grey or yellow according to Figure 4.3, form the intimate dimer responsible for catalytic activity. Two csu-dimers are grouped by two regulatory subunits (rsu, green) stabilizing and enabling feed-back inhibition. Leucine feed-back inhibition is mediated by the N-terminally located sequence stretch highlighted with a green disc while the valine/isoleucine feed-back inhibiting part is C-terminally located and marked with a dark green asterisk as suggested by Lee and Duggleby (2001). **Panel B** Proposal of the quarternary structure from *Arabidopsis thaliana* AHAS holoenzyme: Three dimers of catalytic subunits (yellow/grey) are grouped by three regulatory subunits (green) which may efficiently increase stability and moreover, reach the size of 548 kDa as determined for the native enzyme by Lee and Duggleby (2001).

substance	R*f*
1. acetolactic acid methyl ester	0.25
2. 2-hydroxy-2-isopropyl-3-oxobutyric acid	0.41

stationary phase: TLC silica 60 F_{254},
mobile phase: petroleum ether/ethyl acetate (2:1)

Table 4.3: Proposal for a future specific assay to determine the necic acid synthase activity with radio-labbelled substrates detected with Rapid Intelligence TLC Analyzer (RITA). Detection with phosphomolybdic acid reagent is unspecifically possible for both substances.

of the AHAS from PA-producing plants. The question should be answered which catalytic subunit belongs to which regulatory subunit. For this purpose, the specific NAS assay, established in this work, will clearly distinguish between enzymes recognizing the substrates from primary or from secondary metabolism.

Probably, the AHAS-like NAS activity will be found in an unusual combination of catalytic and regulatory subunits. The ancestor of the regulatory subunit is seen in the threonine deaminase (Lee and Duggleby 2001) which is the enzyme converting threonine to 2-oxobutyrate, finally resulting in the formation of isoleucine. Probably, the recruitment of threonine deaminase took place twice: Once within the branched chain amino acid biosynthesis converting to a regulatory subunit for the AHAS, and once within the PA biosynthesis becoming the regulatory subunit for an AHAS-like enzyme being solely responsible for the exclusive necic acid formation.

The composition of the heteromeric AHAS and NAS can be tested in all possible combinations of catalytic and regulatory subunits including the combination of two different catalytic subunits with two different regulatory subunits. For *Brassica napus*, it was shown that two different catalytic subunits worked together (Bekkaoui et al. 1993), while *Arabidopsis thaliana* catalytic subunits were able to increase the specific activity even when combined with regulatory subunits from *Nicotiana plumbaginifolia* (Hershey et al. 1999).

The most promising combination of catalytic and regulatory subunits discovered in this work from *Eupatorium cannabinum* for PA-biosynthesis will be formed with the catalytic subunit EcW2csu-AHAShyp and the regulatory subunit EcWrsu-AHAShyp. The catalytic subunit EcW2csu-AHAShyp was determined to be restricted to root-tissues (section 3.2.3) and S653P-mutated (section 4.1.2), while the regulatory subunit EcWrsu-AHAShyp probably is encoded in the plastids, since the chloroplast-targeting transit peptide (cTP) was not reliably predicted (section 3.2.4). As a prominent example of heteromeric enzymes encoded with one subunit in the nucleus and the other subunit in the plastids, the ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) must be named and had been found to be active in the chloroplasts. The small subunit of the RuBisCO encoded in the nucleus needs transfer across the chloroplast membrane with the help of a cTP, while the large subunit is encoded in the organelle, needs no transfer and no transit peptide, and awaits

the nuclear encoded subunit to fold up correctly at the stroma side after unfolding for transfer across the chloroplast membrane (Chua and Schmidt 1978).

When considering the fact that the above mentioned most promising catalytic and regulatory AHAS subunits from *Eupatorium cannabinum* (EcW2csu-AHAShyp, EcWrsu-AHAShyp) were isolated from specifically grown root-tissue cultures which did not contain any chloroplasts, it can be proposed that the AHAS-like NAS might be localized in plastids in the roots. Otherwise, the cTP of the catalytic subunit cannot be cleaved and the AHAS might be inactive when remaining in the cytoplasm. Since the first enzyme in PA biosynthesis, the homospermidine synthase (HSS) was localized in the cytoplasm of the cortical parenchyma cells from young roots of *Eupatorium cannabinum* localized by Anke (2004), it may be assumed that either the product of HSS, the homospermidine, or the product of NAS, the pro-necic acid, need transfer from the cytoplasm into the special plastids or vice versa, respectively. According to Pichersky and Gang (2000), biosynthetic pathways of alkaloids are "sometimes split over more than one subcellular compartment or across two or more types of cells or even tissues".

Hence, biochemical analyses have to show whether the esterification of the necic acid with the necine base is performed before the formation of the necic acid is finished or afterwards. Three different scenarios are proposed as shown in Figure 4.5:

1) The necic acid precursor 2-oxoisovalerate is esterified directly with a free necine base. The resulting, postulated "pre pro pyrrolizidine alkaloid" is then accepted by the AHAS-like NAS and transformed into "pro pyrrolizidine alkaloid" and finally reduced at position C3' (necic acid moiety) to the pyrrolizidine alkaloid. Except for the free necine base, none of the postulated intermediates has ever been isolated from plants so far, but probably they are much too instable to be detected by random compound research.

2) The 2-oxoisolvalerate is converted by NAS to the "pro necic acid", reduced to the free necic acid, and esterified with a necine base to form the PA. This scenario is supported by a single detection of free necic acid, namely (-)-viridifloric acid, in fruits of *Lappula echinata*, Boraginaceae (Wang et al. 1986).

3) The "pro necic acid" is obtained as detailed in 2) but esterification takes place before reduction. Once again, it has to be realized that a "pro pyrrolizidine alkaloid" has never been detected.

Probably, the three detected catalytic and the two regulatory subunits from *Eupatorium cannabinum* will not display an exclusive activity for the PA biosynthesis as known from the homospermidine synthase (HSS). Further promising AHAS-like sequences might be detected from reproductive tissues since special AHAS had been detected by Grula et al. (1995) in the anthers of *Gossypium hirsutum* and by Ouellet et al. (1992) in mature ovules and extra embryonic tissues of developing seeds of *Brassica napus* (sequence Bna3 in Figure 4.2). These special AHAS were predicted to work outside the housekeeping function. Hence, reproductive tissues are likely to be a good source of extraordinary AHAS. Having in mind, that root tissue cultures had been proved to synthesize PAs (Hartmann and Zimmer 1986, Biller et al. 1994), a localization of the NAS in the inflorescence is rather unlikely and might demand the transport of the necic acids or the necine base to

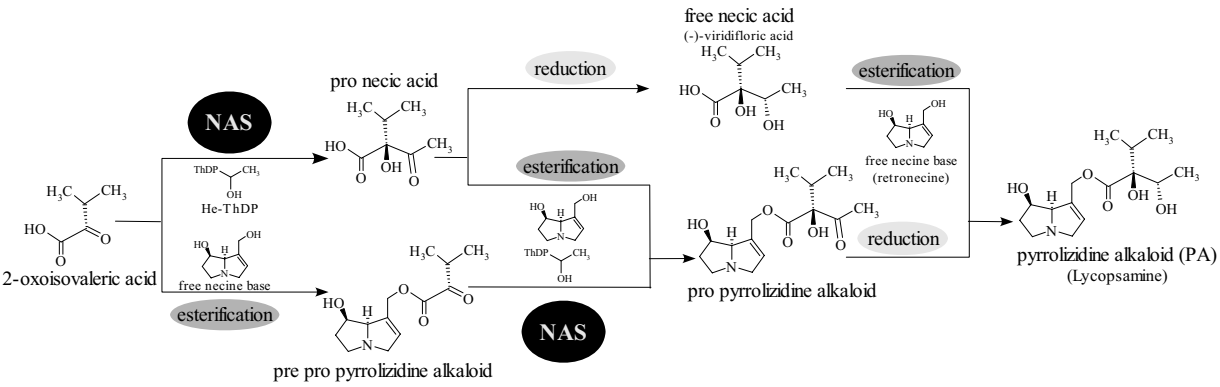


Figure 4.5: The three possible scenarios of PA biosynthesis proposing that the three essential steps "pro necic acid biosynthesis", esterification and reduction are interchanged. Determination of the substrate specificity of the putative AHAS-like necic acid synthase (NAS) should reveal the true scenario.

an unknown place of esterification.

For *Parsonsia laevigata* only one single catalytic subunit was identified up to now (section 3.1.2) probably indicating that the discovery of AHAS in this plant was incomplete. Assuming for one moment that there is just one catalytic subunit, as it is true for *Arabidopsis thaliana*, this might indicate that PA-producing plants are not dependent on a specialized NAS. Moreover, it is conceivable that one enzyme catalyzes both, the branched chain amino acid biosynthesis and the lycopsamine type necic acid biosynthesis. This gene sharing (Piatigorsky et al. 1988) may be an alternative mechanism of the recruitment of enzymes for secondary metabolism. Nevertheless, this can easily be tested by growing *Parsonsia laevigata* on a medium containing AHAS-inhibiting herbicides but supplemented with branched chain amino acids. In case *Parsonsia laevigata* only owns one AHAS that is responsible for branched chain amino acid and PA biosynthesis and in case this AHAS is sensitive to that class of herbicides, plants which are able to grow under these conditions are postulated to be unable to produce lycopsamine type PAs any longer.

One fact supporting to the gene sharing theory is that *E. coli* AHAS I showed properties to accept quite a number of related structures (Engel et al. 2004a). In this work it was shown that *E. coli* AHAS II is capable to accept the pro-necic acid precursor 2-oxoisovalerate besides pyruvate. In general, this shows a rather low substrate specificity for the bacterial enzyme but, on the other hand, this allows the formation of pro-necic acids mechanistically (section 3.4.3.2 on page 174).

Hence, there might be three possible scenarios: Firstly, PA-producing plants are not dependent on a specialized enzyme due to the very low substrate specificity of AHAS, or, secondly, PA-producing plants are dependent on a specialized enzyme which is more likely since, in general, the regulation of primary and secondary metabolism differs significantly, or thirdly, there might be a mixture of both scenarios. In some families, there might be just

one AHAS, catalyzing both the reactions from primary and from secondary metabolism, and in other families, a specialized, neofunctionalized AHAS with exclusive activity in secondary metabolism is precisely regulated according to the plant developmental stage.

Finally, the finding of families harbouring specialized NAS will probably contribute to elucidate the so far determined polyphyletic origin of the pyrrolizidine alkaloid biosynthesis in more detail whereas the finding of families with non-specialized, "all-round" AHAS might reveal a very young origin of lycopsamine type PA biosynthesis capabilities.

5 Summary

1. cDNAs with high similarity to already known AHAS were successfully identified with a combined method of degenerate oligonucleotide primed PCR and RACE technique. Three species from lycopsamine type PA-producing families were chosen and showed different numbers of AHAS-like cDNAs (Table 5.1).

2. Identities of 48 full-length Angiosperm putative or proven AHAScsu-like sequences were compared on the amino acid level (Table 3.24). In a neighbour joining tree (Figure 4.2) the sequences cluster nicely in Angiosperm family branches. Sequence EcW3csu-AHAShyp that displayed higher identities to sequences from other species than to its paralogous sequences branched close to sequences from Asteraceae species and was considered to be a good candidate for a putative housekeeping enzyme in branched chain amino acid biosynthesis. Sequences EcScsu-AHAShyp and EcW2csu-AHAShyp displayed an increased sequence substitution rate that might suggest a different function.

3. The expression patterns of cDNAs from *Eupatorium* were investigated with RT-PCR. Only one catalytic subunit was expressed specifically in root-tissues while all the other catalytic and regulatory subunits were expressed in all investigated plant organs (root-tissue culture, leaf, shoot, shoot tips, leaves, stipes, root-tissue culture), too. This exclusively expressed catalytic subunit (EcW2csu-AHAShyp) was considered to be a good candidate for a putative specific necic acid synthases. Moreover, this sequence displayed two critical mutations in the amino acid sequence (H567Y and S653P) that may result in an altered substrate specificity.

4. Reconstitution of AHAS deficient *E. coli* strain MF2000 proved a functional AHAS for the catalytic subunit sequence EcScsu-AHAShyp.

5. All three catalytic subunits from *Eupatorium cannabinum* showed activity after cell-free expression in the Westerfeld activity assay for the substrate pyruvate and proved functionality.

	catalytic subunits	regulatory subunits
Eupatorium cannabinum (Asteraceae)	3	2
Symphytum officinale (Boraginaceae)	2	n.e.
Parsonsia laevigata (Apocynaceae)	1	n.e.

n.e. = not evaluated

Table 5.1: Numbers of identified cDNAs with high similarties to AHAS catalytic and regulatory subunits

6. Heterologous expression in *E. coli* resulted in an active enzyme only when both, the catalytic (EcScsu-AHAShyp) and regulatory subunit (EcSrsu-AHAShyp), were co-expressed simultaneously in one host cell. The AHAS specific activity for pyruvate as substrate was determined for the purified catalytic subunit with 3.4 nkat/mg. This was within the surprising wide range from 0.7 nkat/mg (*Arabidopsis thaliana* csu GSTtag cleaved, Hershey et al. 1999) to 833 nkat/mg from the same plant determined with an uncleaved GST tag determined by Ott et al. (1996).

7. For secondary metabolism, the proposed ability to accept an altered substrate was proven qualitatively by GC-MS analysis with the help of references synthesized specifically for this purpose. First trials were hampered by a quick and rapid decay of substrates (2-ketoacids) and products (2-hydroxy-3-ketoacids). This was overcome by transferring the compounds to the corresponding methylesters prior to detection.

A Bibliography

Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K., and Watson, J.D. (1989) *Molecular Biology of The Cell.* 2nd ed., Garland Publishing, Inc. New York & London

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990) *Basic local alignment search tool.* J. Mol. Biol. 215, 403-410

Anke, S. (2004) *Molekularbiologische und immunologische Untersuchungen zur gewebe- und zellspezifischen Lokalisation der Homospermidinsynthase, des Eingangsenzyms der Pyrrolzidin-Alkaloid-Biosynthese in Eupatorium cannabinum (Asteraceae) und Phalaenopsis species (Orchidaceae).* PhD thesis TU Braunschweig, Germany

Anke, S., Niemüller, D., Moll, S., Hänsch, R., and Ober, D. (2004) *Polyphyletic origin of pyrrolizidine alkaloids within the Asteraceae. Evidence from differential tissue expression of homospermidine synthase.* Plant Physiol. 136, 4037-4047

Aravind, L. and Koonin, E.V. (1999) *Gleaning non-trivial structural, functional and evolutionary information about proteins by iterative database searches.* J. Mol. Biol. 287(5), 1023-1040

Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., and Struhl, K. (1994) *Current Protocols in Molecular Biology.* New York, Greene Publishing Associates and Wiley-Interscience

Axelsen, K.B. and Palmgren, M.G. (1998) *Evolution of substrate specificities in the P-type ATPase superfamily.* J. Mol. Evol. 46, 84-101

Bale, N.M., Cahill, R., Davies, N.M., Mitchell, M.B., Smith, E.H., and Crout, D.H.G. (1978) *Biosynthesis of the necic acids of pyrrolizidine alkaloids. Further investigations on formation of senecic and isatinecic acids in Senecio species.* J. Chem. Soc., Perkin Trans. I, 101-110

Barak, Z., Chipman, D.M., and Gollop, N. (1987) *Physiological implications of the specificity of acetohydroxy acid synthase isoenzymes of enteric bacteria.* J. Bacteriol. 169, 3750-3756

Bartig, D., Lemkemeier, K., Frank, J., Lottspeich, F., and Klink, F. (1992) *The archaebacterial hypusine-containing protein. Structural features suggest common ancestry with eukaryotic translation initiation factor 5A.* Eur. J. Biochem. 204, 751-758

Bekkaoui, F., Condie, J.A., Neustaedter, D.A., Moloney, M.M. and Crosby, W.L (1991) *Isolation, structure and expression of a cDNA for acetolactate synthase from Brassica napus*, Plant Mol. Biol. 16 (4), 741-744

Bekkaoui, F, Schorr, P., and Crosby, E.L. (1993) *Acetolactate synthase from Brassica napus: immunological characterization and quarternary structure of the native enzyme.* Physiol. Plant 88, 475-484

Belt, T. (1888) *The naturalist in Nicaragua*, 2nd ed. Edward Bumbus, London (reprinted by University of Chicago Press, 1985)

Bennet, T.P. (1967) *Membrane filtration for determining protein in the presence of interfering substances.* Nature 213 (5081), 1131-1132

Bernasconi, P., Woodworth, A.R., Rosen, B.A., Subramanian, M.V. and Siehl, D.L. (1995) *A naturally occurring point mutation confers broad range tolerance to herbicides that target acetolactate synthase.* J. Biol. Chem. 270 (29), 17381-17385

Biller, A., Boppré, M., Witte, L., and Hartmann, T. (1994) *Pyrrolizidine alkaloids in Chromolaena odorata: Chemical and chemoecological aspects.* Phytochemistry 35, 615-619

Birnboim, H.C. and Doly, J (1979) *A rapid alkaline extraction procedure for screening recombinant plasmid DNA.* Nucleic Acids Research 7, 1513-1523

Böttcher, F. Adolph, R., and Hartmann, T. (1993) *Homospermidine synthase, the first pathway-specific enzyme in pyrrolizidine alkaloid biosynthesis.* Phytochemistry 32, 679-689

Boppré, M. (1986) *Insects pharmacophagously utilizing defensive plant chemicals (pyr- rolizidine alkaloids).* Naturwissenschaften 73, 17-26

Boppré, M. (1990) *Lepidoptera and pyrrolizidine alkaloids: exemplification of complexity in chemical ecology.* J. Chem. Ecol. 16, 165-185

Bradford, M.M. (1976) *A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding.* Anal. Biochem. 72, 248-254

Bullock, W.O., Fernandez, J.M., and Short, J.M. (1987) *XL1-Blue, a high efficiency plasmid transforming recA Escherichia coli strain with beta galactosidase selection.* BioTechniques 5, 376-379

Cahill, R., Crout, D.H.G., Mitchell, M.B., and Müller, U. (1980) *Isoleucine biosynthesis and metabolism: stereochemistry of the formation of L-isoleucine and of its*

conversion into senecic and isatinecic acids in Senecio species. J. Chem. Soc., Chem. Comm., 419-421

Caspers, P., Stieger, M., and Burn, P. (1994) *Overproduction of bacterial chaperones improves the solubility of recombinant protein tyrosine kinase in Escherichia coli.* Cell. mol. Biol. 40, 635-644

Cassady, W.E., Leiter, E.H., Bergquist, A., and Wagner, R.P. (1972) *Separation of mitochondrial membranes of Neurospora crassa. II. Submitochondrial localization of the isoleucine-valine biosynthetic pathway.* J. Cell Biol. 53, 66-72

Chang, A. (1997) *Nachweis, Solubilisierung und Charakterisierung einer membrangebundenen, spezifischen Pyrrolizidin-Akaloid-N-Oxygenase aus Keimlingen von Crotalaria scassellatii.* PhD thesis, TU Braunschweig, Germany.

Chang, A.K. and Duggleby, R.G. (1997) *Expression, purification and characterization of Arabidopsis thaliana acetohydroxyacid synthase.* Biochem. J. 327, 161-169

Chang, A.K. and Duggleby, R.G. (1998) *Herbicide-resistant forms of Arabidopsis thaliana acetohydroxyacid synthase: characterization of the catalytic properties and sensitivity to inhibitors of four defined mutants.* Biochem. J. 333, 765-777

Chang, S.-I., Kang, M.-K., Choi, J.-D., and Namgoong, S.K. (1997) *Soluble overexpression in Escherichia coli, and purification and characterization of wild-type recombinant tobacco acetolactate synthase.* Biochem. Biophys. Res. Comm. 234, 549-553

Cheeke, P.R (1989) *Pyrrolizidine alkaloid toxicity and metabolism in laboratory animals and livestock.* In: Toxicants of plant origin (Cheeke, P.R., Ed.) vol 1: Alkaloids., CRS Press, Boca Raton, 1-22

Chen, K.Y. and Liu, A.Y. (1997) *Biochemistry and function of hypusine formation on eukaryotic initiation factor 5A.* Biol. Signals 6 (3), 105-109

Chipman, D., Barak, Z., and Schloss, J.V. (1998) *Biosynthesis of 2-aceto-2-hydroxy acids: acetolactate synthases and acetohydroxyacid synthases.* Biochim. Biophys. Acta 1385, 401-419

Chong, S., Mersha, F.B., Comb, D.G., Scott, M.E., Landry, D., Vence, L.M., Perler, F.B., Benner, J., Kucera, R.B., Hirvonen, C.A., Pelletier, J.J., Paulus, H., and Xu, M.-Q. (1997a) *Single-column purification of free recombinant proteins using a self-cleavable affinity tag derived from protein splicing element.* Gene 192, 277-281

Chong, C.K.; Chang, S.I.; Choi, J.D. (1997b) *Purification and characterization of acetolactate synthase from barley.* J. Biochem. Mol. Biol. 30, 274-279

Chou, M.W. and Fu, P.P. (2006) *Formation of DHP-derived DNA adducts in vivo from dietary supplements and Chinese herbal plant extracts containing carcinogenic pyrrolizidine alkaloids.* Toxicol. Industr. Health 22, 321-327

Christoffers, J., Werner, T., Unger, S., and Frey, W. (2003) *Preparation of acyloins by Cerium-catalyzed, direct hydroxylation of ß-dicarbonyl compounds with molecular oxygen.* Eur. J. Org. Chem, 425-431

Chua, N.-H. and Schmidt, G.W. (1978) *Post-translational transport into intact chloroplasts of a precursor to the small subunit of ribulose-1,5-bisphosphate carboxylase.* Proc. Natl. Acad. Sci. USA 75, 6110-6114

Cohen, S.N., Chang, A.C., and Hsu, L. (1972) *Nonchromosomal antibiotic resistance in bacteria: genetic transformation of Escherichia coli by R-factor DNA.* Proc. Natl. Acad. Sci. 69, 2110-2114

Coligan, J.E., Dunn, B.M., Speicher, D.W., and Wingfield, P.T. (2003) *Short protocols in protein science - a compendium of methods from current protocols in protein scienc*e. USA, Wiley

Croteau, R., Kutchan, T.M., and Lewis, N.G. (2000) *Natural products (secondary metabolites).* In: Biochemistry & molecular biology of plants (Buchanan, B., Gruissem, R., and Jones, R., Eds.), American Soc. Plant Phys., Rockville Maryland, 1250-1318

Crout, D.H.G. (1966) *Pyrrolizidine alkaloids: the biosynthesis of echimidinic acid.* J. Chem. Soc. (C), 1968-1972

Crout, D.H.G. (1967) *Pyrrolizidine alkaloids: the biosynthesis of the angelate component of heliosupine.* J. Chem. Soc. (C), 1233-1234

Crout, D.H.G., Benn, M.H, Imaseki, H., and Geissman, T.A. (1966) *Pyrrolizidine alkaloids: the biosynthesis of seneciphyllic acid.* Phytochemistry 5, 1-21

Crout, D.H.G., Davies, N.M., Smith, E.H., and Whitehouse, D. (1970) *Biosynthesis of the C_{10} necic acid of the pyrrolizidine alkaloids.* J. Chem. Soc., Chem. Comm., 635-636

Crout, D.H.G., Davies, N.M., Smith, E.H., and Whitehouse, D. (1972) *Pyrrolizidine alkaloids: the biosynthesis of senecic acid.* J. Chem. Soc., Perkin Trans. I, 671-680

Culvenor, C.C.J. (1978) *Pyrrolizidine alkaloids - occurence and systematic importance in Angiosperms.* Bot. Notiser 131, 473-486

Culvenor, C.C.J., Edgar, J.A., Jago, M.V., Outteridge, A., Peterson, J.E., and Smith, L.W. (1976) *Hepato- and pneumotoxicity of pyrrolizidine alkaloids and derivatives in relation to molecular structure.* Chem. Biol. Interactions 12, 299-324

Darwin, F and Seward, A.C. (1903) *More letters of Charles Darwin: a record of his work in a series of hitherto unpublished letters.* Vol. 2

De Kraker, J.-W., Luck, K., Textor, S., Tokuhisa, J.G., and Gershenzon, J. (2007) *Two Arabidopsis genes (IPMS1 and IPMS2) encode isopropylmalate synthase, the branchpoint step in the biosynthesis of leucine.* Plant Physiol. 143, 970-986

Demerec, M, Adelberg, E.A., Clark, A.J., and Hartman, P.E. (1966) *A proposal for a uniform nomenclature in bacterial genetics.* Genetics 54, 61-76

Devereux, J., Haeberli, P., and Smithies, O. (1984) *A comprehensive set of sequence analysis programs for the VAX.* Nucl. Acids Res. 12, 387-395

Devlin, J.A. and Robins, D.J. (1984) *Pyrrolizidine alkaloids. Biosynthesis of trichodesmic acid.* J. Chem. Soc. Perkin Trans. I, 1329-1332

Dohmen, R.J., Strasser, A.W.M., Höner, C.B., and Hollenberg, C.P. (1991) *An efficient transformation procedure enabling longterm storage of competent cells of various yeast genera.* Yeast 7, 691-692

Douglas, S.E. and Penny, S.L. (1999) *The plastid genome of the cryptophyte alga, Guillardia theta: complete sequence and conserved synteny groups confirm its common ancestry with red algae.* J. Molec. Evol. 48, 236-244

Duggleby, R.G., and Pang, S.S. (2000) *Acetohydroxyacid synthase.* J. Biochem. Mol. Biol. 33-1, 1-36

Dumas, R., Biou, V., and Douce, R. (1997) *Purification and characterization of a fusion protein of plant acetohydroxy acid synthase and acetohydroxy acid isomeroreductase.* FEBS Letters 408, 156-160

Durner, J. (1991) *Reinigung und Charakterisierung der Acetolactatsynthase aus Gerste (Hordeum vulgare L.).* PhD thesis, Univ. Konstanz, Germany

Durner, J. and Böger, P. (1988) *Acetolactate synthase from barley (Hordeum vulgare L.): purification and partial characterization.* Z. Naturforsch. C 43, 850-856

Durner, J. and Böger, P. (1990) *Oligomeric forms of plant acetolactate synthase depend on flavin adenine dinucleotide.* Plant Physiol. 93, 1027-1031

Dussourd, D.E., Harvis, C.A., Meinwald, J., and Eisner, T. (1991) *Pheromonal advertisement of a nuptial gift by a male moth Utetheisa ornatrix.* Proc. Natl. Acad. Sci. USA 88, 9224-9227

Ehmke, A., von Borstel, K., and Hartmann, T. (1987) *Specific uptake of the N-oxides of pyrrolizidine alkaloids by cells, protoplasts and vacuoles from Senecio cell cultures.* In: Plant vacuoles, their importance in solute compartmentation in cells and their application in plant biotechnology (Marin, B., Ed.), Plenum Publ. Corp., New York, 301-304

Ehmke, A., von Borstel, K., and Hartmann, T. (1988) *Alkaloid N-oxides as transport and vacuolar storage compounds of pyrrolizidine alkaloids in Senecio vulgaris L.* Planta 176, 83-90

Eisner, T. (1982) *For love of nature: exploration and discovery at biological field stations.* Bioscience 32, 321-326

Eisner, T., Rossini, C., Gonzales, A., Iyengar, V.K., Siegler, M.V.S., and Smedley, S.R. (2002) *Paternal investment in egg defence.* In: Chemoecology of insect eggs and egg deposition (Hilker, M., Meiners, T. Eds.), Blackwell, Oxford, 91-116

Emanuelsson, O., Nielsen, H., Brunak, S., and von Heijne G. (2000) *Predicting subcellular localization of proteins based on their N-terminal amino acid sequence.* J. Mol. Biol. 300, 1005-1016

Engel, S., Vyazmensky, M., Geresh, S., Barak, Z., and Chipman, D. (2003) *Acetohydroxyacid synthase: a new enzyme for chiral synthesis of R-phenylacetylcarbinol.* Biotechnol. Bioeng. 83, 833-840

Engel, S., Vyazmensky, M., Berkovich, D., Barak, Z., and Chipman, D (2004a) *Substrate range of acetohydroxyacid synthase I from Escherichia coli in the stereoselectvie synthesis of α-hydroxy ketones.* Biotechnol. Bioeng. 88, 825-831

Engel, S., Vyazmensky, M., Vinogradov, M., Berkovich, D., Bar-Ilan, A., Qimron, U., Rosiansky, Y., Barak, Z., and Chipman, D. (2004b) *Role of a conserved arginine in the mechanism of acetohydroxyacid synthase.* J. Biol. Chem. 279, 24803-24812

Engel, S., Vyazmensky, M., Berkovich, D., Barak, Z., Merchuk, J., and Chipman, D (2005) *Column flow reactor using acetohydroxyacid synthase I from Escherichia coli as catalyst in continuous synthesis of R-phenylacetylcarbinol.* Biotechnol. Bioeng. 89, 733-740

Enß, D. (2006) *Expression und biochemische Charakterisierung zweier pflanzlicher Enzyme mit möglicher Beteiligung an der Pyrrolizidin-Alkaloid-Biosynthese.* Diplomarbeit, University of Halle-Wittenberg, Germany

Eoyang, L. and Silverman, P.M. (1988) *Purification and assays of acetolactate synthase I from Escherichia coli K12.* Methods Enzymol. 166, 435-445

Epelbaum, S., Chipman, D.M., Barak, Z. (1990) *Determination of products of acetohydroxy acid synthase by the colorimetric method, revisted.* Anal. Biochem. 191, 96-99

Evans, W.C. and Woolley, J.G. (1965) *The alkaloids of Datura meteloides D.C.* J. Chem. Soc. 4936-4939

Fang, L.Y., Gross, P.R., Chen, C.H. and Lillis, M. (1992) *Sequence of two acetohydroxyacid synthase genes from Zea mays.* Plant Mol. Biol. 18 (6), 1185-1187

Felsenstein, J. (2001) *PHYLIP, Phylogeny Interference Package, Version 3.6 (alpha2).* Seatle University of Washington

Fraenkel, G.S. (1959) *The raison d'être of secondary plant substances.* Science 129, 1466-1470

Frei, H., Lüthy, J., Brauchli, J., Zweifel, U., Würgler, F.E., and 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.* Chem. Biol. Interactions 83, 1-22

Friden, P., Newman, T., and Freundlich, M. (1982) *Nucleotide sequence of the ilvB promotor-regulatory region: a biosynthetic operon controlled by attenuation and cyclic AMP.* Proc. Natl. Acad. Sci. USA 79, 6156-6160

Frölich, C., Hartmann, T., and Ober, D. (2006) *Tissue distribution and biosynthesis of 1,2-saturated pyrrolizidine alkaloids in Phalaenopsis hybrids (Orchidaceae).* Phytochemistry 67, 1493-1502

Frölich, C., Ober, D., and Hartmann, T. (2007) *Tissue distribution, core biosynthesis and diversification of pyrrolizidine alkaloids of the lycopsamine type in three Boraginaceae.* Phytochemistry, 68, 1026-1037

Frohmann, M.A., Dush, M.K., and Martin, G.R. (1988) *Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer.* Proc. Natl. Acad. Sci. USA 85, 8998-9002

Fu, P.P., Xia, Q., Lin, G., and Chou, M.W. (2004) *Pyrrolizidine alkaloids - Genotoxicity, metabolism enzymes, metabolic activation, and mechanisms.* Drug Metab. Rev. 36, 1-55

Genzel, U. (2006) *Optimierung eines E. coli Expressionssystems für pflanzliche Aceto- lactat-Synthasen mit möglicher Bedeutung für die Pyrrolizidin-Alkaloid-Biosynthese.* Diplomarbeit, University of Halle-Wittenberg, Germany

Gerwick, B.G., Subermanian, V.I., Loney-Gallant, V.I., and Chander, D.P. (1990) *Mechanism of action of the 1,2,4-triazolo(1,5-a)pyrimidines.* Pest. Sci. 29, 357-364

Gish, W. and States, D.J. (1993) *Identification of protein coding regions by database similarity search.* Nat. Genet. 3, 266-272

Gordon, E.D., Mora, R., Meredith, S.C., Lee, C., and Lindquist, S.L. (1987) *Eukaryotic initiation factor 4D, the hypusine-containing protein, is conserved among eukaryotes.* J. Biol. Chem. 262, 16585-16589

Grant, G.A. (2006) *The ACT domain: a small molecule binding domain and its role as a common regulatory element.* J. Biol. Chem. 281 (45), 33825-33829

Grant, S.G.N., Jessee, J., Bloom, F.R., and Hanahan, D.(1990) *Differential plasmid rescue from transgenic mouse DNAs into Escherichia coli methylation-restriction mutants.* Proc. Natl. Acad. Sci. USA 87, 4645-4649

Grimminger, H. and Umbarger, H.E. (1979) *Acetohydroxyacid synthase I of Escherichia coli: purification and properties.* J. Bacteriol. 137, 846-853

Grula, J.W., Hudspeth, R.L., Hobbs, S.L., and Anderson, D.M. (1995) *Organization, inheritance and expression of acetohydroxyacid synthase genes in the cotton allotetraploid Gossypium hirsutum.* Plant Mol. Biol. 28 (5), 837-846

Hänsel, R. and Sticher, O. (2004) *Pharmakognosie - Phytopharmazie.* 7st ed., Springer, Berlin Heidelberg New York

Hagen, J.(2003) *Genetisch und modifikativ bedingte Variabilität der Pyrrolizidinalkaloide in Senecio jacobaea L.*, PhD thesis, TU Braunschweig, Germany

Hanahan, D. (1983) *Studies on transformation of Escherichia coli with plasmids.* J. Mol. Biol. 166, 557-580

Harborne, J. (1993) *Introduction to ecological biochemistry.* San Diego, Academic Press

Hare, J. and Eisner, T. (1993) *Pyrrolizidine alkaloid deters ant predators of Utetheisa ornatrix eggs: Effects of alkaloid concentration, oxidation state, and prior exposure of ants to alkaloid-laden prey.* Oecologia 96, 9-18

Hartmann, T. (1985) *Prinzipien des pflanzlichen Sekundärstoffwechsels.* Plant. Syst. Evol. 150, 15-34

Hartmann, T. (1991) *Alkaloids.* In: Herbivores: Their interactions with secondary metabolites (Rosenthal, G.A. and Berenbaum, M.R., Eds.), 2nd ed., vol. 1: The chemical participants, Academic Press, 79-121

Hartmann, T. (1999) *Chemical ecology of pyrrolizidine alkaloids.* Planta 206, 443-451

Hartmann, T. (2006) *Plant Pyrrolizidine alkaloids (PAs) structures, toxicity, utilization by insects (Arctiids).* Personal communications 31.03.2006, Inst. f. pharm. Biologie, TU Braunschweig

Hartmann, T. and Zimmer, M. (1986) *Organ-specific distribution and accumulation of pyrrolizidine alkaloids during the life history of two annual Senecio species.* J. Plant Physiol. 122, 67-80

Hartmann, T. and Toppel, G. (1987) *Senecionine N-Oxide, the primary product of pyrrolizidine alkaloid biosynthesis in root culture of Senecio vulgaris.* Phytochemistry 26, 1639-1643

Hartmann, T. and Witte, L. (1995) *Chemistry, biology and chemoecology of the pyr- rolizidine alkaloids.* In: Alkaloids: Chemical and biological perspectives (Pelletier, S.W., Ed), vol. 9, chapt. 4, Pergamon Press, Oxford, 155-233

Hartmann, T. and Dierich, B. (1998) *Chemical diversity and variation of pyrrolizidine alkaloids of the senecionine type: biological need or coincidence?* Planta 206, 443-451

Hartmann, T. and Ober, D. (2000) *Biosynthesis and metabolism of pyrrolizidine alkaloids in plants and specialized insect herbivores.* Topics in Current Chemistry 209, 207-243

Hartmann, T., Sander, H., Adolph, R., and Toppel, G. (1988) *Metabolic links between the biosynthesis of pyrrolizidine alkaloids and polyamines in root cultures of Senecio vulgaris L.* Planta 175, 82-90

Hartmann, T., Ehmke, A., Eilert, U.,von Borstel, K., and Theuring, C. (1989) *Sites of synthesis, translocation and accumulation of pyrrolizidine alkaloid N-oxides in Senecio vulgaris.* Planta 177, 98-107

Hawkins, C.F., Borges, A., and Perham, R.N. (1989) *A common structural motif in thiamine pyrophosphate.binding enzymes.* FEBS Lett. 255, 77-82

Heldt, H. W.(1996) *Pflanzenbiochemie*, Spektrum, Akademischer Verlag, Heidelberg, Berlin, 505

Hershey, H.P., Schwartz, L.J., Gale, J.P. and Abell, L.M. (1999) *Cloning and functional expression of the small subunit of acetolactate synthase from Nicotiana plumbaginifolia.* Plant Mol. Biol. 40 (5), 795-806

Heukeshoven, J. and Dernick, R. (1988a) *Vereinfachte und universelle Methode zur Silberfärbung von Proteinen in Polyacrylamidgelen; Bemerkungen zum Mechanismus der Silberfärbung.* Pharmacia LKB Sonderdruck, RE-034, 92-97

Heukeshoven, J. and Dernick, R. (1988b) *Improved silver staining procedure for fast staining in PhastSystem development unit: staining of sodium dodecyl sulfate gels.* Electrophoresis 9, 28-32

Hill, C.M., Pang, S.S., and Duggleby, R.G. (1997) *Purification of Escherichia coli acetohydroxyacid synthase isoenzyme II and reconstitution of active enzyme from its individual pure subunits.* Biochem. J. 327, 891-898

Hill, C.M. and Duggleby, R.G. (1998) *Mutagenesis of Escherichia coli acetohydroxyacid synthase isoenzyme II and characterization of three herbicide-insensitive forms.* Biochem. J. 335, 653-661

Holzer, H., Goedde, H.W., Goeggel, K.H., and Ulrich, B. (1960) *Identification of alpha- hydroxyethyl thiamine pyrophosphate ("active acetaldehyde") as an intermediate in the oxidation of pyruvate by pyruvic oxidase from yeast mitochondria.* Biochem. Biophys. Res. Commun. 3, 599-602

Holmes, D.S. and Quiqley, M. (1981) *A rapid boiling method for preparation of bacterial plasmids.* Anal. Biochem. 114 (1), 193-197

IBA BioTAGnology (2005) *Expression and purification of proteins using Strep-tag and/or 6xHis-tag - A comprehensive manual*

Ish-Horowicz, D. and Burke, J.F. (1981) *Rapid and efficient cosmid cloning.* Nucleic Acids Research 9, 2989-2998

Jenett-Siems, K., Schimming, T., Kaloga, M., Eich, E., Siems, K., Gupta, M.P., Witte, L., and Hartmann, T. (1998) *Phytochemistry and chemotaxonomy of the Convolvulaceae: Pyrrolizidine alkaloids of Ipomoea hederifolia and related species.* Phytochemistry 47, 1551-1560

Jenett-Siems, K., Ott, S.C., Schimming, T., Siems, K., Müller, F., Hilker, M., Witte, L., Hartmann, T., Austin D.F., and Eich, E. (2005) *Ipangulines and minalobines, chemotaxonomic markers of the infrageneric Ipomoea taxon subgenus Quamoclit, section Mina.* Phytochemistry 66, 223-231

Jiang, Y., Fu, P.P., and Lin, G. (2006) *Hepatotoxicity of naturally occuring pyrrolizidine alkaloids.* As. J. Pharmacodyn. Pharmacokin. 6 (3), 187-192

Jones, A.V., Young, R.M., and Leto, K.J. (1985) *Subcellular localization and properties of acetolactate synthase, target site of the sulfonylurea herbicides.* Plant Physiol. 77, S-55

Jones, D.T., Taylor, W.R., and Thornton, J.M. (1992) *The rapid generation of mutation data matrices from protein sequences.* Comput. Appl. Biosci. 8, 275-282

Kajava, A.V., Zolov, S.N., Kalinin, A.E., and Nesmeyanova, M.A. (2000) *The net charge of the first 18 residues of the mature sequence affects protein translocation across the cytoplasmic membrane of gram-negative Bacteria.* J. Bacteriol. 182 (8), 2163-2169

Keeler, S.J., Sanders, P., Smith, J.K., and Mazur, B.J. (1993) *Regulation of tobacco acetolactate synthase gene expression.* Plant Physiol. 102, 1009-1018

Kim, H.-J. and Chang, S.-I. (1995) *Functional expression of Nicotiana tabacum acetolactate synthase gene in Escherichia coli.* J. Biochem. Mol. Biol. 28/3, 265-270

Kimura, M. (1980) *A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences.* J. Mol. Evol. 16, 111-120

Kleschick, W.A., Gerwick, B.C., Carson, C.M., Monte, W.T., and Snider, S.W. (1992) *DE-498, a new acetolactae synthase inhibiting herbicide with multicrop selectivity.* J. Agric. Food Chem. 40, 1083-1085

Kobayashi, M., Asai, M., Ishiguro, T., Kishi, T., Bian, R., and Xie, Q. (1989) *Studies on the antianaphylactic components of Asarum forbesii Maxim.* Shoyakugaku-zasshi 43, 230-234

Kommssion E (1992, 2001) Monographien der Kommission E des Bundesinstitutes für Arzneimittel und Medizinprodukte (BfArM)

Kozak, M. (1987) *An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs.* Nuc. Acids Res. 15, 8125-8148

Kozak, M. (1990) *Downstream secondary structure facilitates recognition of initiator codons by eucaryotic ribosomes.* Proc. Natl. Acad. Sci. USA 87, 8301-8305

Laemmli, U.K. (1970) *Cleavage of structural proteins during the assembly of the head of Bacteriophage T4.* Nature 227, 680-685

Lawther, R.P., Wek, R.C., Lopes, J.M., Pereira, R., Taillon, B.E., and Hattfield, G.W. (1987) *The complete nucleotide sequence of the ilvGMEDA operon of Escherichia coli K-12.* Nucl. Acids Res. 15, 2137-2155

Lee, Y-T. and Duggleby, R.G. (2001) *Identification of the regulatory subunit of Arabidopsis thaliana acetohydroxyacid synthase and reconstitution with its catalytic subunit.* Biochem. 40, 6836-6844

Levitt, G. (1978) *Herbicidal sulfonamids.* US Patent 4127405

Liddell, J.R. and Stermitz, F.R. (1994) Pyrrolizidine alkaloids from Trollius laxus. In: Plant-Assoc. Toxins: Agricultural, phytchemical, and ecological aspects (Colegate, S.M., and Dorling, P.R., Eds.) Proc. Int. Symp. Poisonous Plants (ISOPP4, 1993), CAB International, Wallingford UK, 217-220

Lindigkeit, R., Biller, A., Schiebel, H.M., Boppré, M., and Hartmann, T. (1997) *The two faces of pyrrolizidine alkaloids: the role of the tertiary amine and its N-oxide in chemical defense of insects with acquired plant alkaloids.* Eur. J. Biochem. 245, 262-636

Los, M. (1984) *o-(5-oxo-2-imidazolin-2-yl)arylcarboxylates: a new class of herbicides.* In: Pesticide synthesis through rational approaches (Magee, P.S., Kohn, G.K., and Menn, J.J., Eds.) American Chem. Soc., New York, 29-44

Lottspeich, F. and Zorbas, H. (1998) *Bioanalytik*, Spektrum, Akademischer Verlag, Heidelberg, Berlin

Mann, P., Eich, E., Witte, L., and Hartmann, T. (1996) *GC-MS study on the alkaloid pattern of Merremia quinquefolia (L.) H. Hall.: first occurrence of retronecine esters, simple phenylethylamine derivatives and pyrrolidinyl-heptadecanamide in the Convolvulaceae.* In : Abstract Book, 44th Annual Congress of the Society for Medicinal Plant Research and a Joint Meeting with the Czech Biotechnology Society, Prague, 128

Mattocks, A.R. (1972) *Toxicity and metabolism of Senecio alkaloids.* In: Phytochemical ecology, annual proceedings of the phytochemical society (Harborne, J.B., Ed.), no.8, Academic Press, London, New York, 179-200

Mazur, B., Chui, C.-F., and Smith, J.K. (1987) *Isolation and characterization of plant genes coding for acetolactate synthase, the target enzyme of two herbizides.* Plant Physiol. 85, 1110-1117

McCourt, J.A. and Duggleby, R.G. (2006) *Acetohydroxyacid synthase and its role in the biosynthetic pathway for branched-chain amino acids.* Amino Acids 31(2), 173-210

McCourt, J.A., Pang, S.S., King-Scott, J., Duggleby, R.G., and Guddat, L.W. (2006) *Herbicides binding sites revealed in the structure of plant acetohydroxyacid synthase.* Proc. Natl. Acad. Sci. USA 103, 569-573

Mogk, A., Mayer, M.P., and Deuerling, E. (2001) *Mechanismen der Proteinfaltung. Molekulare Chaperone und ihr biotechnologisches Potential.* Biologie in unserer Zeit 3, 182-192

Moll, S., Anke, S., Kahmann, U., Hänsch, R., Hartmann, T., and 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 Physiol. 130, 47-57

Moreira, R.F. and Noren, C.J. (1995) *Minimum duplex requirements for restriction enzyme cleavage near the termini of linear DNA fragments.* Biotechniques 19 (1), 56-59

Muhitch, M.J. (1988) *Acetolactate synthase activity in developing maize (Zea mays L.) kernels.* Plant. Physiol. 86, 23-27

Muhitch, M.J., Shaner, D.L., and Stidham, M.A. (1987) *Imidazolinones and acetohydroxyacid synthase from higher plants. Properties of the enzyme from maize suspension culture cells and evidence for the binding of imazapyr to acetohydroxyacid synthase in vivo.* Plant. Physiol. 83, 451-456

Mukhopadhyay, A. (1997) *Inclusion bodies and purification of proteins in biologically active forms.* Adv. Biochem. Eng. Biotechnol. 56, 61-109

Mullis, K.B. and Faloona, F.A. (1987) *Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction.* Methods Enzymol. 155, 335-350

Murashige, T. and Skoog, F. (1962) *Medium for growth and bioassays with tobacco tissue cultures.* Physiol. Plant 15, 473-497

Naumann, C. (2003) *Klonierung und Expression einer für die Detoxifizierung von Pyrrolizidinalkaloiden verantwortlichen Flavin-abhängigen Monooxygenase aus Tyria jacobaeae.* PhD thesis, TU Braunschweig, Germany.

Nawaz, H.R., Malik, A., Muhammad, P., Ahmed, S., and Riaz, M. (2000) *Chemical constituents of Ajuga parviflora.* Z. Naturforsch. 55b, 100-103

Nielsen, H., Engelbrecht, J., Brunak, S., and von Heinje, G. (1997) *Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites.* Prot. Eng. 10, 1-6

Niemüller, D.H. (2007) *Vergleichende Lokalisation der Homospermidin-Synthase, Eingangsenzym der Pyrrolizidin-Alkaloid-Biosynthese, in verschiedenen Vertretern der Boraginaceae.* PhD thesis, TU Braunschweig, Germany

Nurhayati, N. (2004) *Identification, molecular cloning and characterisation of homospermidine synthase and deoxyhypusine synthase from Phalaenopsis and Crotalaria species.* PhD thesis, TU Braunschweig, Germany, Cuvillier Verlag, Göttingen

Nurhayati, N. and Ober, D. (2005) *Recruitment of alkaloid-specific homospermidine synthase (HSS) from ubiquitous deoxyhypusine synthase:does Crotalaria possess a functional HSS that still has DHS activity?* Phytochemistry 66, 1346-1357

Ober, D. (1997) *Strategien zur immunologischen und molekular-biologischen Untersuchung der Homospermidin-Synthase, dem Eingangsenzym der Pyrrolizidinalkaloid-Biosynthese.* PhD thesis, TU Braunschweig, Germany

Ober, D. and Hartmann, T. (1999a) *Deoxyhypusine synthase from tobacco: cDNA isolation, characterization, and bacterial expression of an enzyme with extended substrate specificity.* J. Biol. Chem. 274, 32040-32047

Ober, D. and Hartmann, T. (1999b) *Homospermidine synthase, the first pathway-specific enzyme of pyrrolizidine alkaloid biosynthesis, evolved from deoxyhypusine synthase* Proc. Natl. Acad. Sci. USA 96, 14777-14782

Ober, D. and Hartmann, T. (2000) *Phylogenetic origin of a secondary pathway: The case of pyrrolizidine alkaloids.* Plant Mol. Biol. 44, 445-450

Ober, D., Gibas, L., Witte, L., and Hartmann, T. (2003a) *Evidence for general occurrence of homospermidine in plants and its supposed origin as by-product of deoxyhypusine synthase.* Phytochemistry 62, 339-344

Ober, D., Harms, R. Witte, L., and Hartmann, T. (2003b) *Molecular evolution by change of function: Alkaloid-specific homospermidine synthase retained all properties of deoxyhypusine synthase except binding the eIF5A precursor protein.* J. Biol. Chem. 278, 12805-12812

O'Donovan, D.G. and Long, D.J. (1975) *The biosynthesis of senecioic acid.* Proc. R. I. A., Sect. B, 75, 465-468

Oh, K.-J., Park, E.-J., Yoon, M.-Y., Han, T.-R., and Choi, J.-D. (2001) *Roles of histidine residues in tobacco acetolactate synthase.* Biochem. Biophys. Res. Comm. 282, 1237-1243

Ott, K-H., Kwagh, J-G., Stockton, G.W., Sidorov, V., and Kakefuda, G. (1996) *Rational molecular design and genetic engineering of herbicide resistant crops by structure modeling and site-directed mutagenesis of acetohydroxyacid synthase.* J. Mol. Biol. 263, 359-368

Ouellet, T., Rutledge, R.G., and Miki, B.L. (1992) *Members of the acetohydroxyacid synthase multigene family of Brassica napus have divergent patterns of expression.* Plant J. 2, 321-330

Pain, D., Kanwar, Y.S., and Blobel, G. (1988) *Identification of a receptor for protein import into chloroplasts and its localization to envelope contact zones.* Nature 331, 232-237

Pang, S.S. and Duggleby, R.G. (1999) *Expression, purification, characterization, and reconstitution of the large and small subunits of yeast acetohydroxyacid synthase.* Biochem. 38, 5222-5231

Pang, S.S., Guddat, L.W., and Duggleby, R.G. (2003) *Molecular basis of sulfonylurea herbicide inhibition of acetohydroxyacid synthase.* J. Biol. Chem. 278, 7639-7644

Park, M.H., Lee, Y.B., and Joe, Y.A. (1997) *Hypusine is essential for eukaryotic cell proliferation.* Biol. Signals 6, 115-123

Park, M.h., Joe, Y.A., and Kang, K.R. (1998) *Deoxyhypusine synthase activity is essential for cell viability in the yeast Saccharomyces cerevisiae.* J. Biol. Chem. 273, 1677-1683

Piatigorsky, J., O'Brien, W.E., Norman, B.L., Kalumuck, K., Wistow, G.J., Borras, T., Nickerson, J.M., and Wawrousek, E.F. (1988) *Gene sharing by δ-crystallin and argininosuccinate lyase.* Proc. Natl. Acad. Sci. USA 85, 3479-3483

Pichersky, E. and Gang, D.R. (2000) *Genetics and biochemistry of secondary metabolites in plants: An evolutionary perspective.* Trends Plant Sci. 5, 439-445

Poulsen, C. and Stougaard, P. (1989) *Purification and properties of Saccharomyces cerevisiae acetolactate synthase from recombinant Escherichia coli.* Eur. J. Biochem. 185, 433-439

Radhakrishnan, A.N. and Snell, E.E. (1960) *Biosynthesis of valine and isoleucine. II. Formation of α-acetolactate and α-aceto-α-hydroxybutyrate in Neurospora crassa and Escherichia coli.* J. Biol. Chem. 235, 2316-2321

Raina, S. and Missiakas, D. (1997) *Making and breaking disulfide bonds.* Annu. Rev. Microbiol. 51, 179-202

Reimann, A., Nurhayati, N., Backenköhler, A., and Ober D. (2004) *Repeated evolution of the pyrrolizidine alkaloid-mediated defense system in separate angiosperm lineages.* Plant Cell 16, 2772-2784

Robins, D.J., Bale, N.M., and Crout, D.H.G. (1974) *Pyrrolizidine alkaloids. The biosynthesis of monocrotalic acid, the necic acid component of monocrotaline.* J. Chem. Soc., Perkin Trans. I, 2082-2086

Röder, E. (1995) *Medicinal plants in Europe containing pyrrolizidine alkaloids.* Pharmazie 50, 83-98

Roseman, A.M., Chen, S., White, H., Braig, K., and Saibil, H.R. (1996) *The chaperonin ATPase cycle: mechanism of allosteric switching and movements of substrate-binding domains in GroEL.* Cell 87, 241-251

Ryan, E.D. and Kohlhaw, G.B. (1974) *Subcellular localization of isoleucine-valine biosynthetic enzymes in yeast.* J. Bacteriol. 120, 631-637

Saiki, R K., Scharf, S., Faloona, F., Mullis, K.B., Horn, G.T., Erlich, H.A., and Arnheim, N. (1985) *Enzymatic amplification of β-globin genomic sequences and restriction site analysis of diagnosis of sickle cell anemia.* Science 230, 1350-1354

Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B., and Erlich, H.A. (1988) *Primer directed enzymatic amplification of DNA with a thermostable DNA polymerase.* Science 239, 487-491

Saitou, N. and Nei, M. (1987) *The neighbor-joining method: A new method for reconstructing phylogenetic trees.* Mol. Biol. Evol. 4, 406-425

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) *Molecular cloning: a laboratory manual.* 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York

Sander, H. and Hartmann, T. (1989) *Site of synthesis, metabolism and translocation of senecionine N-oxide in cultured roots of Senecio erucifolius.* Plant Cell Tissue Organ. Cult. 18, 19-32

Schultz, C.P. (2000) *Illuminating folding intermediates.* Nature Struct. Biol. 7, 7-10

Shimizu, T., Nakayama, I., Nakao, T., Nezu, Y., and Abe, H. (1994) *Inhibition of plant acetolactate synthase by herbicides, pyrimidinylsalicylic acids.* J. Pest. Sci. 19, 59-67

Shuman, S. (1994) *Novel approach to molecular cloning and polynucleotide synthesis using Vaccinia DNA topoisomerase.* J. Biol. Chem. 269, 32678-32684

Silva, K.L. and Trigo, J.R. (2002) *Structure-activity relationship of pyrrolizidine alkaloids in insect chemical defense against the orb-weaving spider Nephila clavipes.* Chem. Ecol. 28 (4), 657-668

Singh, B.K., Stidham, M.A., and Shaner, D.L. (1988a) *Separation and characterization of two forms of acetohydroxy acid synthase from black mexican sweet corn cells.* J. Chromatogr. 444, 251-261

Singh, B.K., Stidham, M.A., and Shaner, D.L. (1988b) *Assay of acetohydroxyacid synthase.* Anal. Biochem. 171, 173-179

Smith, D.B. and Johnson, K.S. (1988) *Single-step purification of polypeptides expressed in Escherichia coli as fusions with glutathione S-transferase.* Gene 67(1), 31-40

Smith, J.K., Schloss, J.V., and Mazur, B.J. (1989) *Functional expression of plant acetolactate synthase genes in Escherichia coli.* Proc. Natl. Acad. Sci. USA 86, 4179-4183

Sone, M., Akiyama, Y., and Ito, K. (1997) *Differential in vivo roles played by DsbA and Dsbc in the formation of protein disulfide bonds.* J. Biol. Chem. 272, 10349-10352

Southan, M.D. and Copeland, L. (1996) *Physical and kinetic properties of acetohydroxy acid synthase from wheat leaves.* Physiol.Plant. 98, 824-832

Stebbins, G.L. (1974) *Flowering plants: evolution above the species level.* Cambridge, MA

Steenkamp, V., Stewart, M.J., Van der Merwe, S., Zuckerman, M., and Crowther, N.J. (2001) *The effect of Senecio latifolius, a plant used as a South African traditional medicine, on a human hepatoma cell line.* J. Ethnopharmacology 78, 51-58

Størmer, F.C. and Umbarger, H.E. (1964) *The requirement for flavine adenine dinucleotide in the formation of acetolactate by Salmonella typhimurium extracts.* Biochem. Biophys. Res. Comm. 17, 587-592

Studier, F.W. and Moffatt, B.A. (1986) *Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes.* J. Mol. Biol. 189, 113-130

Summerton, J., Atkins, T., and Bestwick, R. (1983) *A rapid method for preparation of bacterial plasmids.* Anal. Biochem. 133, 79-84

Terpe, K. (2004) *Proteinseparation mit Affinitäts-Tags, Teil 3.* Laborjournal 10, 66-67

Terpe, K. (2006) *Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems.* Appl. Microbiol. Biotechnol. 72 (2), 211-222

Textor, S., de Kraker, J.-W., Hause, B., Gershenzon, J., and Tokuhisa, J.G. (2007) *MAM3 catalyzes the formation of all aliphatic glucosinolate chain length in Arabidopsis.* Plant Physiol. 144, 60-67

Thomas, J.G. and Baneyx, F. (1996) *Protein misfolding and inclusion body formation in recombinant Escherichia coli cells overexpressing heat-shock proteins.* J. Biol. Chem. 271 (19), 11141-11147

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., and Higgins,D.G. (1997) *The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools.* Nucl. Acids Res. 24, 4876-4882

Thu Huong, D.T., Martin, M.-T., Litaudon, M., Sevenet, T., and Pais, M. (1998) *Pyrrolizidine alkaloids from Amphorogyne spicata.* J. Nat. Prod. 61, 1444-1446

Tinney, G., Theuring, C., Paul, N., and Hartmann, T. (1998) *Effects of rust infection with Puccinia lagenophorae on pyrrolizidine alkaloids in Senecio vulgaris.* Phytochemistry 49, 1589-1592

Tittmann, K., Golbik, R., Uhlemann, K., Khailova, L., Schneider, G., Patel, M., Jordan, F., Chipman, D.M., Duggleby, R.G., and Hübner, G. (2003) *NMR analysis of covalent intermediates in thiamine diphosphate enzymes.* Biochemistry 42, 7885-7891

Toppel, G., Witte, L., and Hartmann, T. (1988) *N-oxidation and degradation of pyrrolizidine alkaloids during germination of Crotalaria scassellatii.* Phytochemistry 27, 3757-3760

Tranel, P.J. and Wright, T.R. (2002) *Resistance to ALS-inhbiting herbicides: what have we learned?* Weed Sci. 50, 700-712

Umbarger, H.E. and Brown, B. (1958) *Isoleucine and valine metabolism in Escherichia coli. VIII. The formation of acetolactate.* J. Biol. Chem. 233, 1156-1160

Van Dam, N.M. and Vrieling, K. (1994) *Genetic variation in constitutive and inducible pyrrolizidine alkaloid levels in Cynoglossum officinale L.* Oecologia 99, 374-378

Van Dam, N.M., Witte, L., Theuring, C., and Hartmann, T. (1995) *Distribution, biosynthesis and turnover of pyrrolizidine alkaloids in Cynoglossum officinale.* Phytochemistry 39, 287-292

Vasconcellos-Neto, J. and Lewinsohn, T.M. (1984) *Discrimination and release of unpalatable butterflies by Nephila clavipes, a neotropical orb-weaving spider.* Ecol. Entomol. 9, 337-344

Vyazmensky, M., Sella, C., Barak, Z., and Chipman, D.M. (1996) *Isolation and characterization of subunits of acetohydroxyacid synthase isoenzyme III and reconstitution of the holoenzyme.* Biochemistry 35, 10339-10346

Wang, X.C., Pei, Y.H., Li, X., and Zhu, T.R. (1986) *Antibacterial constituents in the fruit of Lappula echinata Gilib.* Yaoxue Xuebao (=Act. Pharm. Sin.) 21 (3), 183-186

Weber, S. (1997) *Untersuchungen zur Biosynthese der Pyrrolizidin-Alkaloide vom Lycopsamin-Typ an Wurzelkulturen von Eupatorium clematideum.* PhD thesis, TU Braunschweig, Germany

Weber, S., Eisenreich, W., Bacher, A., and Hartmann, T. (1999) *Pyrrolizidine alkaloids of the lycopsamine type: biosynthesis of trachelanthic acid.* Phytochemistry 50, 1005-1014

Weinstock, O., Sella, C., Chipman, D.M., and Barak,Z. (1992) *Properties of subcloned subunits of bacterial acetohydroxyacid synthases.* J. Bacteriol. 174, 5560-5566

Westerfeld, W.W. (1945) *A colorimetric determination of blood acetoin.* J. Biol. Chem. 161, 495-502

Wickner, W., Driessen, A.J.M., and Hartl, F.-U. (1991) *The enzymology of protein translocation across the Escherichia coli plasma membrane.* Annu. Rev. Biochem. 60, 101-124

Wiedenfeld, H. and Röder, E. (1984) Pyrrolizidinalkaloide. Dtsch. Apoth. Ztg. 43, 2116-2122

Wiersma, P.A., Schmiemann, M.G., Condie, J.A., Crosby, W.L., and Moloney, M.M. (1989) *Isolation, expression and phylogenetic inheritance of an acetolactate synthase gene from Brassica napus.* Mol. Gen. Genet. 219 (3), 413-420

Wiersma, P.A., Hachey, J.E., Crosby, W.L., and Moloney, M.M. (1990) *Specific truncations of an acetolactate synthase gene from Brassica napus efficiently complement ilvB/ilvG mutants of Salmonella typhimurium.* Gen. Genet. 224, 155-159

Winter, C.K. and Segall, H.J. (1989) *Metabolism of pyrrolizidine alkaloids.* In: Toxicants of plant origin (Cheeke, P.R., Ed.) vol 1: Alkaloids., CRS Press, Boca Raton, 23-40

Witte, L., Ehmke, A., and Hartmann, T. (1990) *Interspecific flow of pyrrolizidine alkaloids.* Naturwissenschaften 77, 540 - 543

Yanisch-Perron, C., Vieira, J., and Messing, J. (1985) *Improved M13 phage cloning vectors and host strains: nucleotide sequence of the M13mp18 and pUC19 vectors.* Gene 33, 103-109

Yoon, J.M., Yoon, M.Y., Kim, Y.T., Choi, J.D. (2003) *Characterization of two forms of acetolactate synthase from barley.* J. Biochem. Mol. Biol. 36 (5), 456-461

Zhou, J.-J., Zhang, G.-A., Huang W., Birkett, M.A., Field, L.M., Pickett, J.A., and Pelosi, P. (2004) *Revisting the odorant-binding protein LUSH of Drosophila melanogaster: evidence for odour recognition and discrimination.* FEBS Letters 558, 23-26

B Appendix

B.1 Abbreviations

B.2 IUPAC Ambiguity Code for Nucleotides

B.3 One Letter Code for Amino Acids

B.4 Applied Restriction Enzymes

B.5 AHAS NCBI Accessions

AAP	abridged anchor primer
AHAS	acetohydroxyacid synthase
AHTC	anhydrotetracycline
Amp	ampicillin
AUAP	abridged universal anchor primer
BCAA	branched chain amino acids
bp	base pairs
Cam	chloramphenicol
CB	column buffer
cDNA	copy DNA
ChBD	chitin binding domain
csu	catalytic subunit
cTP	chloroplast transit peptide
DMF	dimethylformamide
DHS	deoxyhypusine synthase
DOC	deoxycholate
DOP-PCR	degenerate oligonucleotide primed PCR
DTT	dithiotreitol
EA	ethyl acetate
EDTA	ethylenediaminetetraacetic acid
EE	diethyl ether
EXP	expression primer
FAD	flavin adenine dinucleotide
FPLC	fast protein liquid chromatography
GSP	gene specific primer
GST	glutathione S-transferase
He-ThDP	hydroxythyl thiamine diphosphate
His	histidine
HSS	homospermidine synthase
IPTG	isopropyl-1-thio-β-D-galactoside
Kan	kanamycin
kDa	kilo Dalton
kb	kilo base pairs

Table B.1: Abbreviations A-K

LB	Luria-Bertani
MCS	multiple cloning site
mRNA	messenger RNA
mTP	mitochondrial transit peptide
NAS	necic acid synthase
nt	nucleotides
OD	optical density
ORF	open reading frame
PA	pyrrolizidine alkaloid
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PDC	pyruvate decarboxylase
PE	petroleum ether
PEG	polyethylene glycol
Pfu	*Pyrococcus furiosus*
Pfx	*Pyrococcus species*
POX	pyruvate oxidase
RITA	rapid intelligence TLC analyzer
RNA	ribonucleic acid
rsu	regulatory subunit
RTC	root-tissue culture
RT-PCR	reverse transcription PCR
SAP	shrimp alkaline phosphatase
SDS	sodium dodecyl sulfate
SP	secretory pathway
Strep	StrepII tag
Taq	*Thermus aquaticus*
TCA	trichloroacetic acid
T_m	melting temperature
UTR	untranslated region
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactoside

Table B.2: Abbreviations L - Z

symbol	**represents**
B	C, G, or T
D	A, G, or T
H	A, C, or T
K	G or T
M	A or C
N	A, C, G, or T
R	A or G
S	C or G
V	A, C, or G
W	A or T
Y	C or T

Table B.3: IUPAC Ambiguity Code for nucleotides. A = adenosine, C = cytidine, G = guanosine, T = thymidine

symbol	**represents**	**symbol**	**represents**
A	alanine	M	methionine
C	cysteine	N	asparagine
D	aspartate	P	proline
E	glutamate	Q	glutamine
F	phenylalanine	R	arginine
G	glycine	S	serine
H	histidine	T	threonine
I	isoleucine	V	valine
K	lysine	W	tryptophan
L	leucine	Y	tyrosine

Table B.4: One letter code for amino acids

restriction enzyme (required temp.)[1]	**restriction site**	**base pairs between 5'end and restriction site required for an optimal cut according to**	
		NEB[2]	Fermentas
*Afl*III	A↓CRYGT	6	n.i.
*Bam*HI	G↓GATCC	1	1
*Bgl*II	A↓GATCT	3	2
*Bsa*I (50°C)	GGTCTC(N_1/N_5)↓	6	3
*Bsa*AI	YAC↓GTR	6	1
*Bsm*BI (55°C)	CGTCTC(N_1/N_5)↓	6	1
*Bst*1107I	GTA↓TAC	6	2
*Dra*I	TTT↓AAA	6	3
*Eco*0109I	RG↓GNCCY	6	1
*Eco*RI	G↓AATTC	1	1
*Eco*RV	GAT↓ATC	1	2
*Hinc*II	GTY↓RAC	6	1
*Kpn*I	GGTAC↓C	2	1
*Nco*I	C↓CATGG	2	2
*Nde*I	CA↓TATG	6	3
*Not*I	GC↓GGCCGC	4-7	2
*Pae*I	GCATG↓C	2	5
*Sac*I	GAGCT↓C	1	1
*Sac*II	CCGC↓GG	6	1
*Sal*I	G↓TCGAC	3	3
*Sap*I	GCTCTTC(N1/N4)↓	6	1
*Sca*I	AGT↓ACT	6	2
*Xba*I	T↓CTAGA	1	2
*Xho*I	C↓TCGAG	1	2

[1] If no reaction temperature is given, 37°C are the optimum, Y = C or T; N = A, C, G, or T
[2] Moreira and Noren 1995
n.i. = no information

Table B.5: Restriction enzymes used in this project

AHAScsu plant source	**NCBI accession**	**AHAScsu plant source**	**NCBI accession**
Alopecurus myosuroides	AJ437300	*Nicotiana tabacum* 1	X07644[1]
Amaranthus powellii	AF363370	*Nicotiana tabacum* 2	X07645[1]
Amaranthus retroflexus	AF363369	*Oryza sativa* 1	AB049823
Amaranthus sp.	ASU55852	*Oryza sativa* 2	AB049822
Amaranthus tuberculatus	EF157821	*Oryza sativa* 3	AY885674
Arabidopsis thaliana	NM_114714[1]	*Oryza sativa* 4	AY885673
Bassia scoparia	AF094326	*Oryza sativa* 5	DQ516977
Brassica napus 1	X16708 M20498[1]	*Oryza sativa* 6	DQ516979
Brassica napus 2	M60068 M37786[1,2]	*Oryza sativa* 7	AY885675
Brassica napus 3	Z11526	*Oryza sativa* 8	DQ516980
Bromus tectorum	AAM03119[2]	*Oryza sativa* 9	DQ516981
Camelina microcarpa 1	AY428880	*Oryza sativa* 10	DQ516978
Camelina microcarpa 2	AY428947	*Oryza sativa* 11	AK242817
Gossypium hirsutum 1	Z46960	*Papaver rhoeas*	AJ577316
Gossypium hirsutum 2	Z46959	*Raphanus raphanistrum*	AJ344984[2]
Helianthus annuus 1.1	AY541451	*Sagittaria trifolia*	AB301496
Helianthus annuus 1.2	AY541452	*Schoenoplectus juncoides* 1	AB257441
Helianthus annuus 1.3	AY541453	*Schoenoplectus juncoides* 2	AB257443
Helianthus annuus 1.4	AY541454	*Solanum ptychanthum*	AF308650[2]
Helianthus annuus 1.5	AY541455	*Triticum aestivum*	AY210405[2]
Helianthus annuus 2.1	AY541456	*Xanthium* sp. 1	XSU16279[1]
Helianthus annuus 2.2	AY541457	*Xanthium* sp. 2	XSU16280[1]
Lolium multiflorum	AF310684	*Zea mays* 1	X63554
Medicago trunculata	EU292216	*Zea mays* 2	X63553 S38034

[1] functionality was proven either by deficient mutant complementation or by heterologous expression and characterization, [2] incomplete sequence

Table B.6: Acetohydroxy acid synthase catalytic subunit amino acid sequences from the database NCBI used to create an alignment and degenerate primers

AHASrsu plant source	**NCBI accession**
Arabidopsis thaliana	NM_180498
Arabidopsis thaliana 2	NM_121634
Arabidopsis thaliana 3	NM_179843
Arabidopsis thaliana 4	NM_179842
Arabidopsis thaliana 5	NM_179841[1]
Nicotiana plumbaginifolia	AJ234901[1,2]
Oryza sativa	AC145365
Oryza sativa 2	XM_466654[2]
Zea mays	BT019179[2]

[1] heterologously expressed and functionally characterized, [2] incomplete sequence

Table B.7: Acetohydroxy acid synthase regulatory subunit amino acid sequences from the database NCBI used to create an alignment and degenerate primers