Isolation of cDNAs encoding (+)-and (–)-Germacrene D-Synthase, α-Gurjunene- and Cascarilladiene Synthase from *Solidago canadensis* L. and their functional Expression in *E. coli*

Dissertation

zur Erlangung des Doktorgrades
des Fachbereichs Chemie der
UNIVERSITÄT HAMBURG

vorgelegt von

Iris Altug

Hamburg 2003

1. Gutachter Prof. Dr. W.A. König
Institut für Organische Chemie, Universität Hamburg

2. Gutachter Prof. Dr. B. Bisping
Abteilung für Lebensmittelmikrobiologie und Lebensmittelhygiene, Universität Hamburg

1. Prüfer Prof. Dr. W.A. König
Institut für Organische Chemie, Universität Hamburg

2. Prüfer Prof. Dr. U. Hahn
Abteilung für Biochemie und Molekularbiologie, Universität Hamburg

3. Prüfer Dr. W. Maison
Institut für Organische Chemie, Universität Hamburg

Tag der Disputation: 16. Januar 2004

Die vorliegende Arbeit wurde in der Zeit von November 1999 bis Dezember 2003 unter der Leitung von Herrn Prof. Dr. W.A. König erstellt. Die Durchführung erfolgte am Institut der Organsichen Chemie der Universität Hamburg unter der Leitung von Prof. Dr. W.A. König und in der Gruppe Cellcybernatics am Forschungsintitut Plant Research International (vormals AB-DLO), Wageningen in den Niederlanden unter der Leitung von Dr. H.J. Bouwmeester.

Für meine Eltern:

Filiz & Özer Altug

Contents

Abbreviations

(+)GDS	(+)-germacrene D-synthase
(–)GDS	(–)-germacrene D-synthase
αGS	α-gurjunene-synthase
aa	amino acid/s
ADP	adenosine diphosphate
AMV	avian myeloblastosis virus
APS	ammonium peroxodisulfate
aq	aqueous
ATP	adenosine triphosphate
bp	base pairs
BLAST	Basic Local Alignment Search Tool
BSA	bovine serum albumin
CAPS	3-(cyclohexylamino)-1-propanesulfonic acid
cDNA	complementary DNA
CDIS	conifer diterpene internal sequence domain
CS	cascarilladiene-synthase
CoA	coenzyme A
dATP	2′-deoxyadenosine 5′-triphosphate
dCTP	2′-deoxycytidine 5′-triphosphate
DMF	dimethyl formamide
DMSO	dimethyl sulfoxide

dGTP	2′-deoxyguanosine 5′-triphosphate
DMAPP	dimethylallyl diphosphate
DNA	deoxy ribonucleic acid
dNTP	nucleotide mixture of dATP, dCTP, dGTP, dTTP
dpm	disintegrations per minute
dsDNA	double strand DNA
DTT	dithiothreitol
DXP	1-deoxy-D-xylulose 5-phosphate
E. coli	*Escherichia coli*
EDTA	ethylenediamine tetraacetate
EtBr	ethidium bromide
ER	endoplasmatic reticulum
dTTP	2′-deoxythymidine 5′-triphosphate
FPP	*E,E*-farnesyl diphosphate
g	gram
GC	gas chromatograph
GC-EAG	gas chromatography linked to electroantenogramm recordings
GC-MS	gas chromatography with mass selective detector
GGPP	geranylgeranyl diphosphate
GPP	geranyl diphosphate
h	hour/s
HMG	*(S)*-3-hydroxy-3-methylglutaryl
IPP	isopentenyl diphosphate
IPTG	isopropyl-β-D-thiogalactopyranoside
IUB	International Union of Biochemistry
kDa	kilo Dalton
L	liter
LiDS	lithium dodecylsulfoxide

MEP	2*C*-methyl-D-erythritol 4-phosphate
min	minute
MMLV	Moloney murine leukemia virus
MOPSO	3-(*N*-morpholino)-2-hydroxypropane sulfonic acid
mRNA	messenger RNA
MS	mass spectrometry
NAD	nicotinamide adenine dinucleotide (oxidized)
NADP	nicotinamide adenine dinucleotide phosphate (oxidized)
NADPH	nicotinamide adenine dinucleotide phosphate (reduced)
NCBI	National Center for Biotechnology Information
NPP	nerolidyl diphosphate
nt	nucleotide/s
ORF	open reading frame
PCR	polymerase chain reaction
pfu	plaque forming unit
Pipes	piperazine-1,4-bis(2-ethane sulfonic acid)
PMSF	phenylmethylsulfonyl fluoride
PVPP	polyvinyl polypyrrolidone
r	standard deviation
RNA	ribonucleic acid
RT-PCR	reverse transcriptase polymerase chain reaction
SDS	sodium dodecylsulfate
SDS-PAGE	sodium dodecylsulfate polyacrylamide gel electrophoresis
sec	second
SSC	sodium chloride–sodium citrate buffer
TAE	Tris(hydroxymethyl)aminomethane–acetate–EDTA buffer
TBE	Tris(hydroxymethyl)aminomethane–borate–EDTA buffer
TEMED	N, N, N',N'-tetramethyl-ethylenediamine

Tricine	N-[Tris(hydroxymethyl)methyl]glycine
Tris	Tris(hydroxymethyl)aminomethane
UTR	untranslated region
x-gal	5-bromo-4-chloro-3-indolyl-β-D-galactoside

Amino acid notation

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

Nucleotide base notation

A	adenine
C	cytosine
G	guanine
T	thymine

Chapter 1

Introduction

Plants produce a vast and diverse assortment of organic compounds like terpenoids, phenols or alkaloids, the great majority of which do not appear to participate directly in growth and development. These substances are referred to as secondary metabolites or natural products. Although noted for the complexity of their chemical structures and biosynthetic pathways, secondary metabolites have been widely perceived as biologically insignificant and have historically received little attention from biologists.

Organic chemists, however, have long been interested in these novel phytochemicals and have investigated their chemical properties extensively since the 1850s. Studies of natural products stimulated the development of separation techniques, spectroscopic approaches to structure elucidation, and synthetic methodologies that now constitute the fundamentals of organic chemistry.

The interest in natural products was not purely academic but was rather prompted by their great utility as dyes, fibers, waxes, perfumes and drugs. Recognition of the biological properties of natural products has fueled the search for new drugs, antibiotics, insecticides, herbicides, and pheromones. Importantly, the diverse biological effects produced by natural products has prompted a reevaluation of the roles these compounds play in plants, especially in the context of ecological interactions. Many

of these compounds have now been shown to be important in protection against herbivores [Gibson and Pickett, 1983] and microbial infection [Stoessl et al., 1976], as attractants for pollinators [Dudareva and Pichersky, 2000] and seed-dispersing animals, and as allelopathic agents [Jacyno et al., 1991], that influence competition among plant species. These ecological functions affect plant survival profoundly. [Buchanan et al., 2002]

1.1 Terpenoids

The terpenoids, also called isoprenoids are one of the largest groups of natural products found in nature with more than 30 000 known examples and their number is growing steadily [Davis and Croteau, 2000]. They are perhaps the most structurally diverse class of secondary metabolites.

The name terpenoid, or terpene, derives from the fact that the first members of the class were isolated from turpentine ("Terpentin" in German). At the turn of the 20^{th} century, structural investigations of many terpenoids led Otto Wallach in 1887 to formulate the "Isoprene Rule" ([Wallach, 1887], [Wallach, 1909]), which postulated that most terpenoids could be constructed hypothetically by repetitively joining C_5 isoprene units (2-methylbuta-1,3-diene). This principle provided the first conceptual framework for a common structural relationship among terpenoid natural products.

Wallach's idea was refined in the 1950s, when Leopold Ruzicka formulated the "Biogenetic Isoprene Rule" [Ruzicka et al., 1953], emphasizing mechanistic considerations of terpenoid synthesis in terms of electrophilic elongations, cyclizations and rearrangements. This hypothesis ignores the precise character of the biological precursors and assumes only that they are "isoprenoid" in structure. As a working model for terpenoid biosynthesis, the biogenetic isoprene rule has proved essentially correct. Despite great

diversity in form and function, the terpenoids are unified in their common biosynthetic origin.

The biosynthesis of all terpenoids involves the sequential assembly of branched five carbon isopentenoid units (isopentenyl diphosphate and dimethylallyl diphosphate). Classification of the different families of terpenoids is based on the number of five carbon units present in the skeleton of the compound reviewed in [Buchanan et al., 2002] and [McCaskill and Croteau, 1997].

- Hemiterpenes are the smallest terpenes, containing a single isoprene unit. The best known hemiterpene is isoprene itself, a volatile product released from photosynthetically active tissues.

- Monoterpenes consist of two isoprene units. As the first terpenoids isolated from turpentine in the 1850s, they were considered to be the base unit from which the subsequent nomenclature is derived. The monoterpenes are best known as components of the essential oil of flowers, herbs and spices.

- Sesquiterpenes (one and one-half terpene) originate from three isoprene units containing 15 carbon atoms. Like monoterpenes, many sesquiterpenes are found in essential oils. In addition, numerous sesquiterpenoids act as phytoalexins inhibiting the growth of pathogenic fungi, antibiotic compounds in response to microbial attack, and as antifeedants.

- Diterpenes, which contain 20 carbon atoms (four isoprene units), include phytol (the hydrophobic side chain of chlorophyll) and phytohormones.

- Triterpenes, contain 30 carbon atoms and include the large group of steroids, various toxins, feeding deterrents and components of surface waxes.

- Tetraterpenes (40 carbons) are among others the carotenoid pigments which perform essential functions in photosynthesis.

- The polyterpenes (containing more than eight isoprene units) include the prenylated quinone electron carriers (plastoquinone and ubiquinone) and enormously long polymers such as latex (average molecular mass greater than 106 kDa).

An example for each class of the terpenes is shown in figure 1.1.

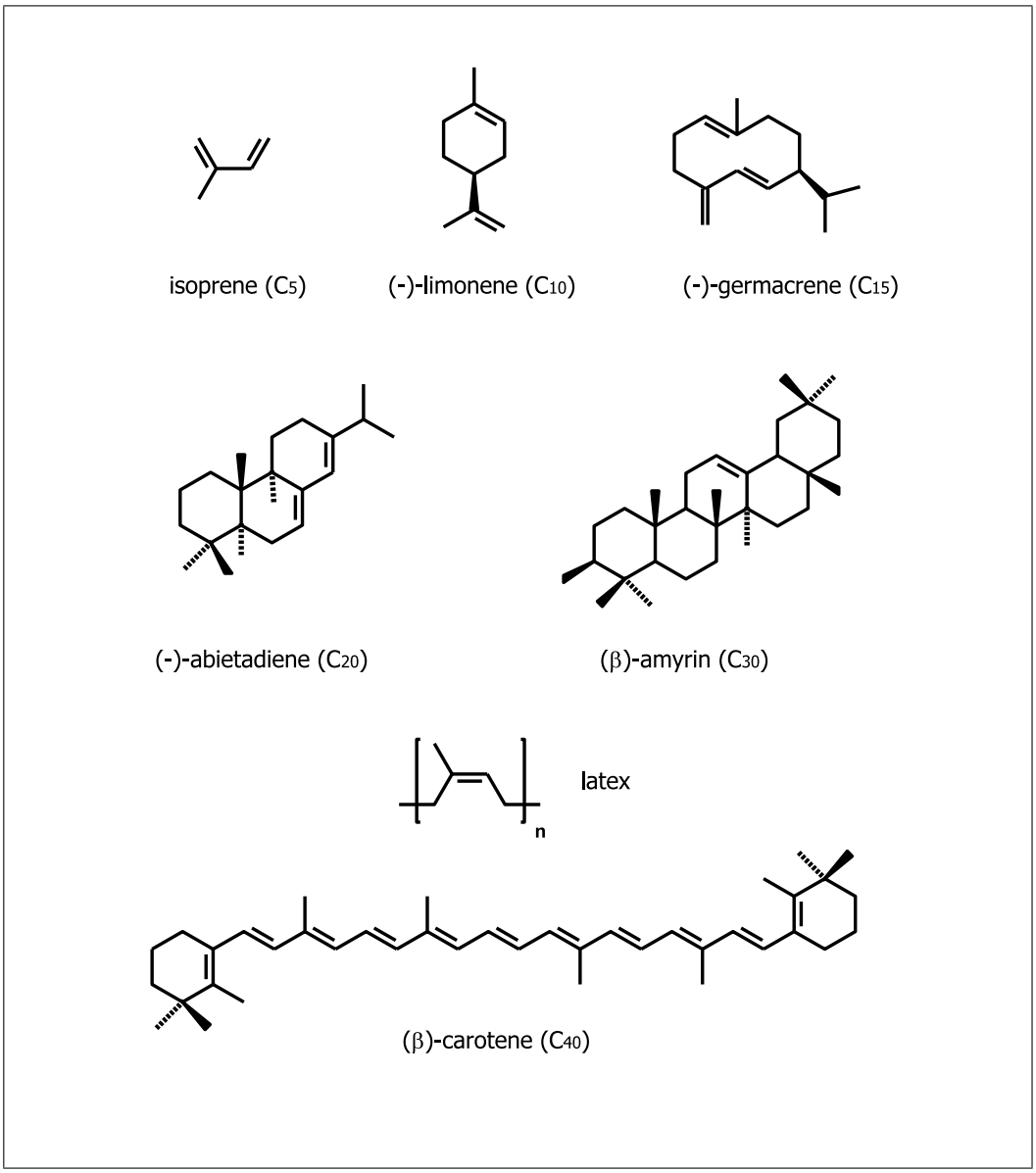


Figure 1.1: structure of selected terpenes

1.2 Chirality

A definition of chirality was given by Lord Kelvin in 1904 [Kelvin, 1904]: *"I call any geometrical figure, or group of points, chiral, and say it has chirality, if its image in a plane mirror, ideally realized, cannot be brought to coincide with itself."*

One of the reasons for the occurrence of chirality is the tetrahedral geometry of the carbon atom. Normally when four different substituents are bound to a carbon atom the compound becomes chiral, its mirror image can not be superimposed on top of it. The two possible compounds are called enantiomers.

An example is shown in figure 1.2

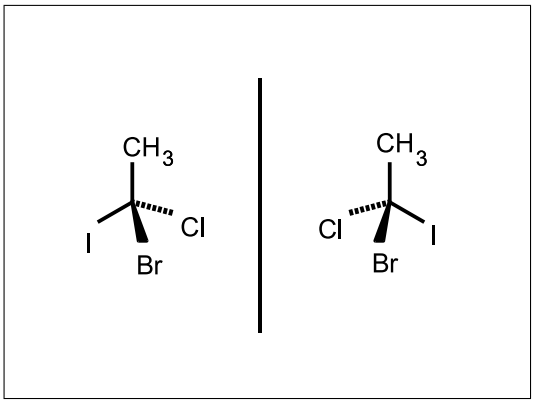


Figure 1.2: example for central chirality

Enzymatically catalyzed reactions commonly generate only one of the possible enantiomers. This is due to the catalytic site of enzymes, which is chiral too due to the chirality of the amino acids. This asymmetry leads to enzyme enantiospecificity, their ability to distinguish between substrate enantiomers and catalyze reactions with only one of the enantiomers. Some common examples are the occurrence of mainly D-sugars and L-amino acids in nature.

Enantiomers do not differ in their physical properties as long as they exist in a non-chiral environment, besides in the phenomenon of the optical activity. Solutions containing each enantiomer rotate the plane of polarized light in the opposite direction.

Many terpenes possess chiral centers. The common analysis of terpenes (mono-, sesqui-, and diterpenes) by gas chromatography using non chiral stationary phases does not differentiate enantiomers as their physical properties are identical. For the separation of enantiomers stationary phases with chiral selectors are needed. Good resolution of chiral terpenes was found on stationary phases composed of modified cyclodextrins [König, 1998]. Cyclodextrins are cyclic oligosaccharides consisting of 6, 7 or 8 D-glucopyranosyl units connected by α-1,4-glycosidic linkage. These are named α-, β-, and γ-cyclodextrins. An example is shown in figure 1.3. As cyclodextrins are chiral itself, they form different diastereomeric complexes with the enantiomers, which may result in different retention times of the enantiomers in a chromatographic system. For decreasing the melting point of cyclodextrins modified cyclodextrins are used as chiral stationary phases e.g. heptakis(6-O-*tert*-butyldimethylsilyl-2,3-di-O-methyl)-β-cyclodextrin (6-T-2,3-Meβ-CD).

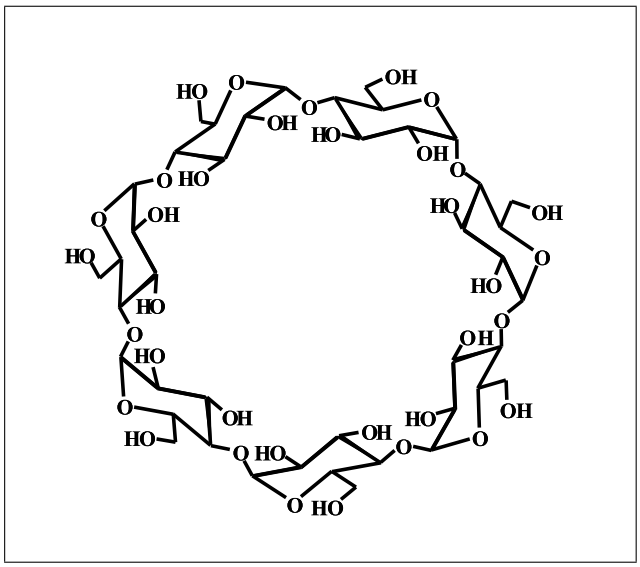


Figure 1.3: structure of β-cyclodextrin

Chapter 2

Scope of the thesis

The biosynthesis of sesquiterpenes is initiated by cyclization of farnesyl diphosphate. The enzymes catalyzing the formation of sesquiterpenes are called sesquiterpene synthases. These enzymes usually work enantiospecifically, catalyzing the formation of one of the possible enantiomers of a sesquiterpene.

Interestingly *Solidago canadensis* and other *Solidago* species are exceptional in producing both enantiomers of the sesquiterpene germacrene D. This led to the question, if one germacrene D-synthase is producing both enantiomers non-specifically or if two individual germacrene D-synthases are existing in *Solidago canadensis*, each of them catalyzing enantiospecifically one of the enantiomers of germacrene D. In earlier investigations by Schmidt [Schmidt, 1998], it was shown, that *Solidago canadensis* possesses two individual synthases, each of them catalyzing enantiospecifically the formation of either (+)-germacrene D or (–)-germacrene D. Furthermore, Schmidt also demonstrated the mechanism of the formation of the two enantiomers by using the substrate farnesyl diphosphate, deuterium labeled on different positions, combined with GC-MS and linked-scan mass spectrometric measurements. He showed that the (–)-germacrene D formation proceeds by a 1,3-hydride shift after the ring closure, while (+)-germacrene D is formed by two 1,2-hydride shifts .

The aim of this work was to isolate the encoding cDNAs of (+)- and (–)-germacrene D-synthase from *Solidago canadensis*. Isolation of the encoding genes should start by generating first fragments of sesquiterpene cDNAs from *Solidago canadensis* by PCR using degenerate primers, designed on conserved regions of known terpene synthase cDNAs. Isolation of the full-length cDNAs should then be performed either by RACE-PCR or by screening a cDNA library. For the latter a λ-phage cDNA library from mRNA of *Solidago canadensis* leaf material has to be prepared.

As the specific function of a gene or cDNA can not be assigned only on knowledge of the nucleotide sequence, the function of the cDNA has to be demonstrated by functional expression of the protein in a suitable system followed by an enzyme assay with the substrate farnesyl diphosphate. The enzymatically catalyzed product has then to be identified by gas chromatography with mass selective detector (GC-MS) and GC measurements with an enantioselective stationary phase to assign the function of the expressed enzyme. A suitable system to express the enzymes is the heterologous expression in *Escherichia coli*, which was demonstrated for plant terpene synthases before [Back et al., 1994].

By having a convenient system to produce the enzymes in sufficient amounts in *E. coli*, X-ray crystallographic experiments can be performed to explore their 3-D-structure and to identify the enantiospecifically working active site of both germacrene D-synthases. Based on the known crystal structure of *epi*-aristolochene synthase from *Nicotiana tabacum* [Starks et al., 1997], it is feasible to perform comparative homology protein modeling of the isolated sesquiterpene synthases. The protein homology modeling in combination with substrate docking experiments can lead to first impressions of the enantioselectively working active sites of both germacrene D-synthases. Amino acid residues within the active site, involved in the enantiospecificity of the catalysis, can be identified.

The following aims were set within the present work:

- isolation of sesquiterpene synthases encoding cDNAs from *Solidago canadensis* (specifically (+)-germacrene D-synthase and (–)-germacrene D-synthase)
  - generating first sesquiterpene cDNA fragments by degenerate primer PCR
  - retrieving full-length cDNAs:
    - either by RACE-PCR
    - or by screening a cDNA library. For this a λ-phage cDNA library of mRNA from *Solidago canadensis* has to be prepared
- functional expression of isolated full-length cDNAs in *E. coli*
- identification of heterologously expressed enzyme catalyzed sesquiterpenes by GC-MS and enantioselective GC measurements
- protein homology modeling and substrate docking to investigate the enantiospecifically working active sites of both germacrene D-synthases

Chapter 3

General Part

3.1 Biosynthesis of Terpenoids

The biosynthesis of terpenoids from simple, primary metabolites can be conceptually divided into three overall steps:

- the synthesis of the fundamental precursor isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP)
- assembly of DMAPP and IPP to the terpene skeleton precursors by prenyl transferases
- formation (often cyclization) of the terpene skeleton precursor to the terpene skeletons by terpene synthases
- secondary enzymatic modifications to the skeletons (largely redox reactions)

Although terpenoid biosynthesis in plants, animals, and microorganisms involves similar classes of enzymes, important differences exist among these processes. In particular, plants produce a much wider variety of terpenoids than do either animals or

microorganisms, a difference reflected in the complex organization of plant terpenoid biosynthesis at the tissue, cellular, subcellular, and genetic levels. The production of large quantities of terpenoid natural products as well as their subsequent accumulation, emission, or secretion is almost always associated with the presence of anatomically highly specialized structures.

The glandular trichomes and secretory cavities of leaves and the glandular epiderms of flower petals generate and store or emit fragrant terpenoid essential oils to encourage pollination by insects. The resin ducts and blisters of conifer species produce and accumulate a defensive resin consisting of turpentine (monoterpene olefins) and resin (diterpenoid resin acids). These specialized structures sequester natural products away from sensitive metabolic processes and thereby prevent autotoxicity. Most structures of this type are non photosynthetic and must therefore rely on adjacent cells to supply the carbon and energy needed to drive terpenoid biosynthesis reviewed in [Buchanan et al., 2002] and [McCaskill and Croteau, 1997].

3.2 Formation of IPP and DMAPP

All terpenoids are assembled biosynthetically from only two precursors: isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP) [Eisenreich et al., 1998]. Presently two individual biosynthesis pathways have been elucidated in plants to generate IPP and DMAPP: the mevalonate pathway and the deoxyxylulose pathway.

3.2.1 The mevalonate pathway

Initial observations by Folkers, Tavormina and coworkers indicated that the isoprenoid monomers are biosynthetically derived from mevalonate [Folkers and Wolf, 1956], [Tavormina et al., 1956]. Subsequent work by several scientists elucidated in detail the steps of the mevalonate pathway (reviewed in [Bach, 1995], [Bloch, 1992]).

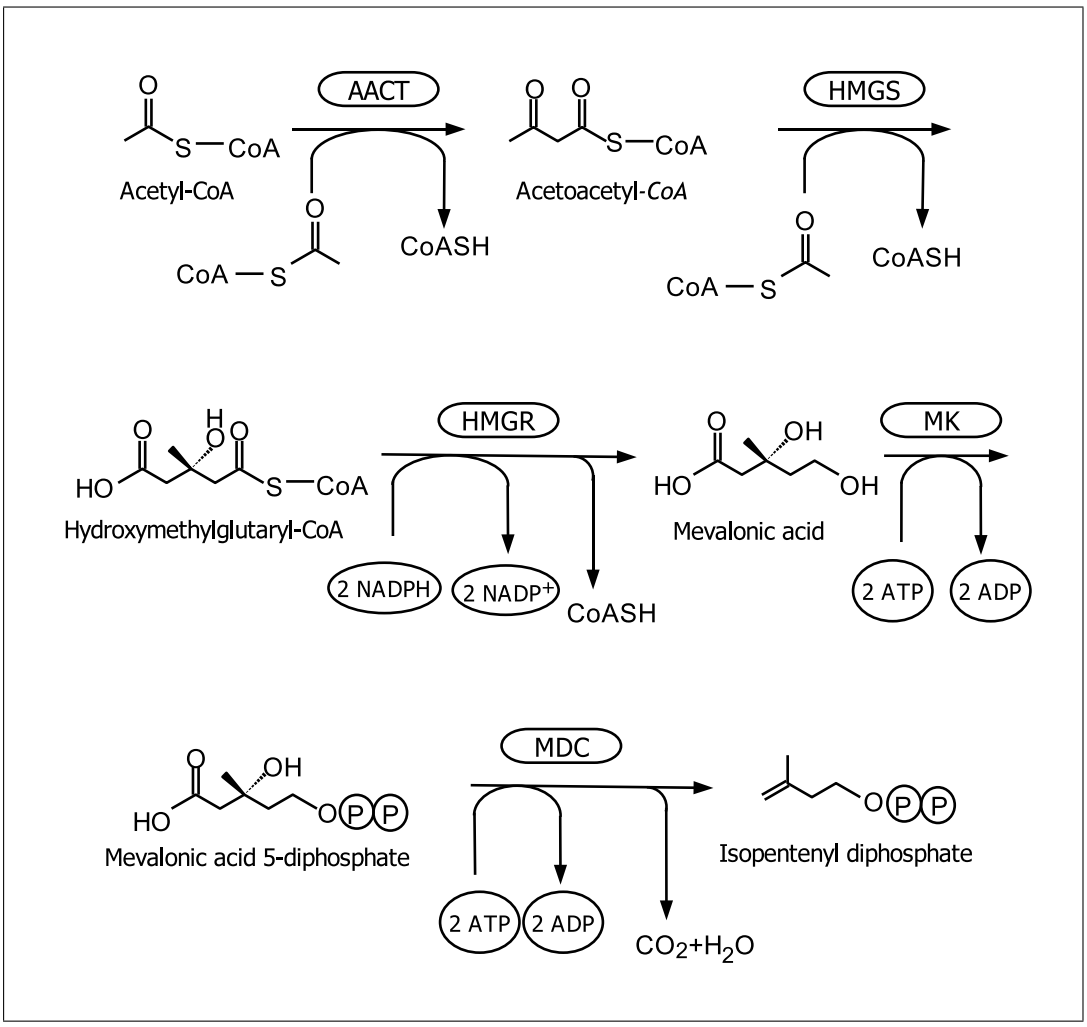


Figure 3.1: The mevalonate pathway.
AACT = acetyl-CoA acetyltransferase, HMGS = hydroxymethyl glutaryl-synthase, HMGR = hydroxymethyl glutaryl-reductase, MK = mevalonic acid kinase, MDC = mevalonic acid decarboxylase

As shown in figure 3.1, the reaction consist of the following steps:

- a Claisen-type condensation of two acetyl CoA molecules yields acetoacetyl-CoA formed by an acetyl-CoA acetyltransferase.

- An aldol addition of another acetyl CoA catalyzed by 3-hydroxymethylglutaryl-synthase yields *(S)*-3-hydroxy-3-methylglutaryl-CoA (HMG-CoA),
- which is subsequently reduced by HMG-CoA reductase under use of NADPH to *(R)*-mevalonic acid.
- *(R)*-mevalonic acid is then phosphorylated to mevalonic acid 5-diphosphate,
- followed by a phosphorylation-assisted decarboxylation yielding isopentenyl diphosphate (IPP).

3.2.2 Isopentenyl isomerase

Isopentenyl diphosphate is utilized in the formation of the allylic isomer dimethylallyl diphosphate, which serves as the immediate precursor for isoprenoid biosynthesis. Isomerization of isopentenyl diphosphate to dimethylallyl diphosphate is catalyzed by isopentenyl-diphosphate-Δ-isomerase. The reaction proceeds via an acid-base catalyzed carbocationic mechanism as shown in figure 3.2. [Eisenreich et al., 1998], [Bach, 1995].

3.2.3 The deoxyxylulose pathway

Since the discovery of the mevalonate pathway, many studies on the biosynthesis of terpenoids in a wide variety of species have been published.

In many cases, however, the experimental data on the biosynthesis of specific terpenoids in plants and in microorganisms could not be explained easily by the mevalonate pathway. It was often found that mevalonate was not incorporated efficiently into the terpenoids under study. The reassessment of the isoprenoid biosynthesis in plants and microorganisms was initiated in the late 1980s by ^{13}C-labeling studies of Rohmer and others [Rohmer et al., 1996], [Schwarz, 1994]. The unexpected labelling

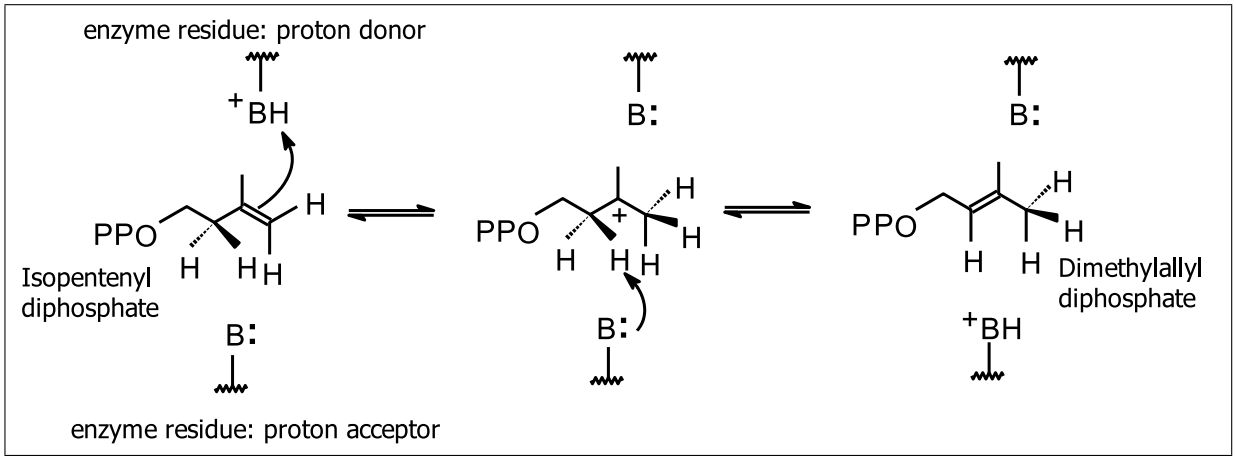


Figure 3.2: Isomerization of isopentenyl diphosphate to dimethylallyl diphosphate catalyzed by isopentenyl-diphosphate-Δ-isomerase.

patterns in the terpenoid moiety of the molecules led eventually to a novel pathway, the deoxyxylulose pathway [Eisenreich et al., 1998], shown in figure 3.3.

The starting substrates of the deoxyxylulose pathway are pyruvate and glyceraldehyde 3-phosphate.

- In a thiamine-dependent transketolase-type reaction, a C_2-unit derived from pyruvate is transferred to glyceraldehyde 3-phosphate, whereby 1-deoxy-D-xylulose 5-phosphate (DXP) is formed. This step is catalyzed by the enzyme DXP-synthase.
- In a second step, 1-deoxy-D-xylulose 5-phosphate is converted into the branched-chain polyol 2-*C*-methyl-D-erythritol 4-phosphate (MEP) by DXP-reductoisomerase.
- In a cytidine 5'-triphosphate-dependent reaction MEP is converted into 4-(cytidine 5'diphospho)-2-*C*-methyl-D-erythritol.
- The ATP-dependent phosphorylation at the C-2 hydroxy group yields 2-phospho-4-(cytidine 5'-diphospho)-2-*C*-methyl-D-erythritol,

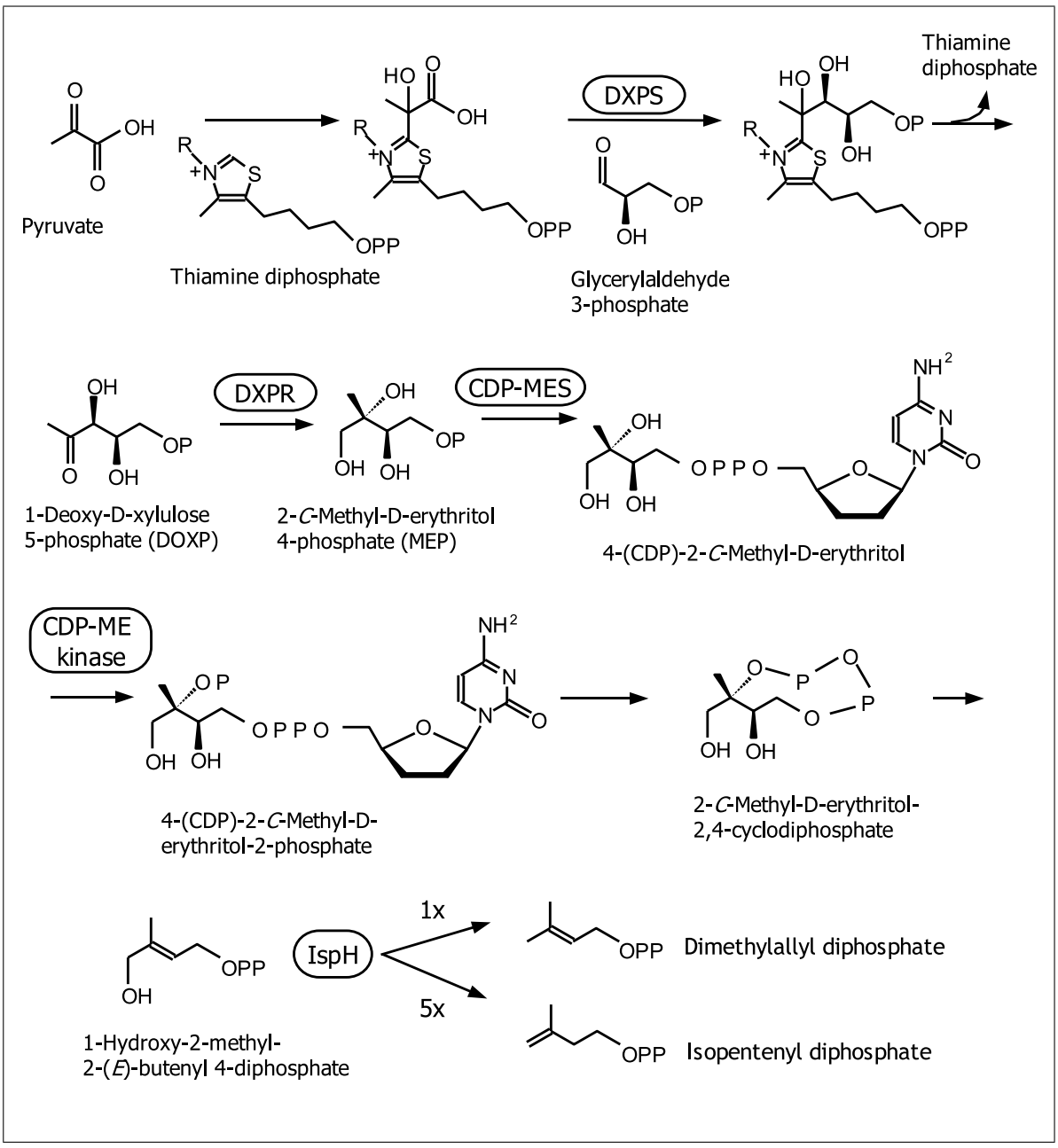


Figure 3.3: Deoxyxylulose pathway.

DXPS = deoxyxylulose-5-phosphate-synthase, DXPR = deoxyxylulose-5-phosphate-reducto-isomerase, CDP-MES = cytidine-diphospho methyl-D-erythritol-synthase, CDP = cytidine 5'-diphospho, CDP-ME kinase = cytidine-diphospho methyl-D-erythritol kinase

- which is further converted into 2-*C*-methyl-D-erythritol 2,4-cyclodiphosphate.
- A reductive ring-opening reaction yields 1-hydroxy-2-methyl-2-(*E*)-butenyl 4-diphosphate.
- In a last branching step 1-hydroxy-2-methyl-2-(*E*)-butenyl 4-diphosphate is converted into IPP and DMAPP as mixture in a ratio of 5:1 [Rohdich et al., 2002].

3.3 Prenyltransferases

Dimethylallyl diphosphate serves as the immediate precursor of the different families of terpenoids. As shown in figure 3.4 it undergoes elongation by the sequential addition of one, two or three additional IPP molecules to form either geranyl diphosphate (GPP), farnesyl diphosphate (FPP) or geranylgeranyl diphosphate (GGPP). Geranyl diphosphate is the 10-carbon precursor for monoterpenes, farnesyl diphosphate the 15-carbon precursor for sesquiterpenes and geranylgeranyl diphosphate the 20-carbon precursor for diterpenes.

A family of enzymes known collectively as prenyltransferases catalyzes this electrophilic elongation sequence. Specific prenyltransferases exist for the formation of GPP, FPP and GGPP.

- The reaction catalyzed by prenyltransferases involves the initial ionization of the allylic diphosphate DMAPP to generate a delocalized allylic carbocation.
- This enzyme bound cation then attacks the double bond of IPP
- followed by deprotonation to generate the next allylic diphosphate homologue.

As shown in figure 3.4 in the case of GPP synthase, the first condensation product is released from the enzyme. In the case of FPP and GGPP synthases, the resulting

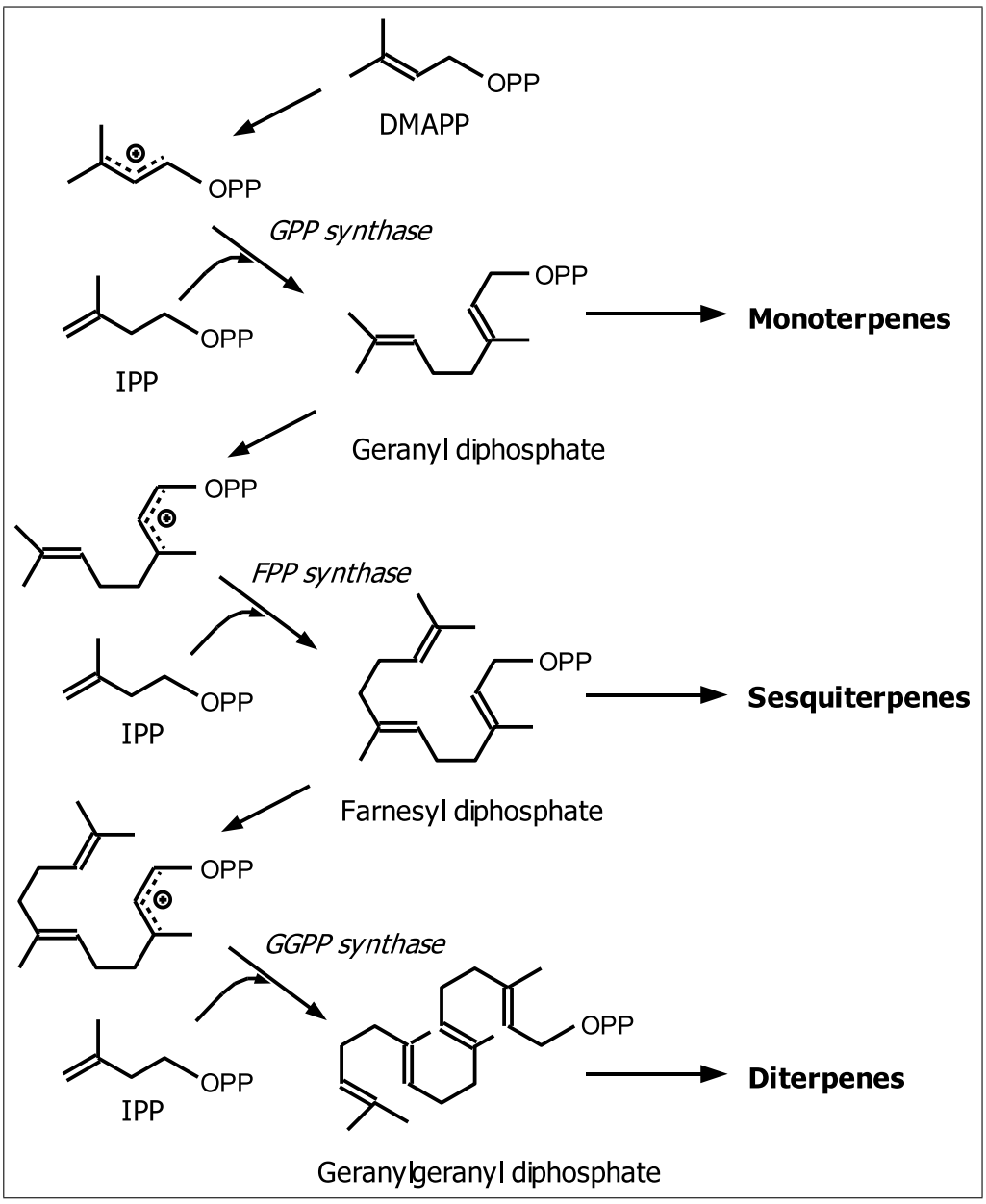


Figure 3.4: The reaction of prenyltransferases.
DMAPP = dimethylallyl diphosphate, IPP = isopentenyl diphosphate

enzyme-bound GPP undergoes further reaction with the addition of another IPP to generate FPP, which is either released in the case of FPP synthase or which undergoes

reaction with a third IPP to generate GGPP in the case of GGPP synthase, reviewed in [McCaskill and Croteau, 1997].

3.4 Compartmentation of isoprenoid biosynthesis

A more fundamental, and perhaps universal, feature of the organization of terpenoid metabolism exists at the subcellular level. The sesquiterpenes (C_{15}), triterpenes (C_{30}), and polyterpenes appear to be produced in the cytosolic and endoplasmatic reticulum (ER) compartments, whereas isoprene, the monoterpenes (C_{10}), diterpenes (C_{20}), tetraterpenes (C_{40}) originate largely from the plastids. A schematic overview is shown in figure 3.5. The biosynthetic pathways for the formation of the fundamental precursor IPP and DMAPP differs in these compartments, with the classical mevalonate pathway being active in the cytosol and ER and the deoxyxylulose pathway operating in the plastids [Buchanan et al., 2002]. The compartmental separation of the two different IPP and DMAPP biosynthetic pathways is not absolute as metabolites can be exchanged between the compartments [Schwarz, 1994], [Laule et al., 2003]. The extent of this crosstalk depends on the species as well as on the presence and concentration of exogenous precursors. Generally crosstalk contributions appear to be small in intact plants under physiological conditions (<1%). Higher values have been found in plant cell cultures in the presence of exogenous deoxyxylulose or mevalonate [Eisenreich et al., 1999]. The crosstalk of precursors does not only appear on IPP and DMAPP level, but also on e.g. GPP level as shown for the biosynthesis of sesquiterpenes in chamomile [Adam et al., 1999].

Recent biosynthetic labeling studies in *Solidago canadensis* have been done by Steliopoulos et al. [Steliopoulos et al., 2002]. They found that the bulk biosynthesis of the sesquiterpene germacrene D in *Solidago canadensis* involves the deoxyxylulose pathway, showing that this disagreement seems to be more common than originally postulated.

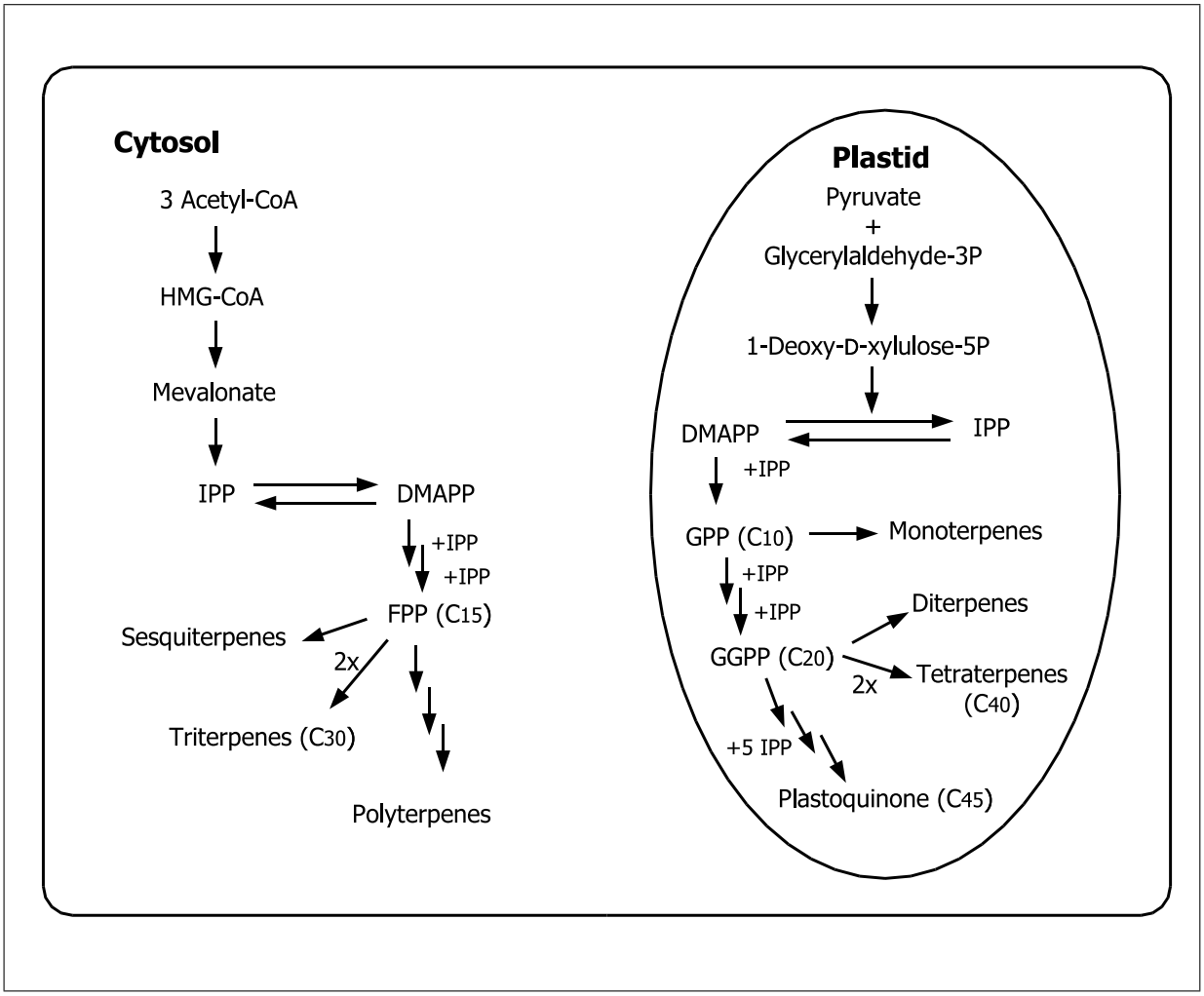


Figure 3.5: Schematic overview of the compartmentation of terpene biosynthesis [Lichtenthaler, 1999]

3.5 Terpene synthases

The family of enzymes responsible for the formation of terpenoids from GPP, FPP and GGPP are known as monoterpene, sesquiterpene, and diterpene synthases. These synthases use the acyclic, achiral allylic diphosphates as substrates to form the enormous chemical diversity of carbon skeletons characteristic of terpenoids. Most terpenoids are cyclic, and often contain multiple ring systems. The regio- and stereochemistry of

the basic ring structure of the terpenoids are in most cases determined by these highly specific synthases. The type of reaction catalyzed by the terpene synthases is the intermolecular equivalent of the reaction of the prenyltransferases.

The reaction generally starts with an ionization of the allylic diphosphate, proceeded by an intramolecular cyclization. The reactive intermediate then undergoes often different rearrangements like internal electrophilic addition or hydride shift before termination of the reaction by proton loss or capture of a nucleophile, e.g. water, reviewed in [McCaskill and Croteau, 1997] and [MacGarvey and Croteau, 1995].

The exploitation of electrophilic additions for carbon-carbon bond formation is a biochemical rarity. Conservation of this unusual carbocationic mechanism suggests a common evolutionary origin of the prenyltransferases and the terpenoid synthases.

3.5.1 Sesquiterpene synthases

More than 300 distinct sesquiterpene carbon skeletons have been identified to date, and thousands of naturally occurring oxidized or otherwise modified derivatives have been isolated. These metabolites display a broad range of physiological properties, including antibiotic, antitumor, antiviral, cytotoxic, phytotoxic, antifungal, insect antifeedant, and hormonal activities. Remarkably all of these substances are derived from a single acyclic precursor, farnesyl diphosphate (FPP). Enzymes catalyzing the formation of sesquiterpenes from FPP are known as sesquiterpene synthases or sesquiterpene cyclases.

Farnesyl diphosphate undergoes ionization followed by electrophilic attack of the resultant allylic cation on either the central double bond to form six- or seven-membered ring intermediates or on the distal double bond to form 10- or 11-membered ring intermediates as shown in figure 3.6.

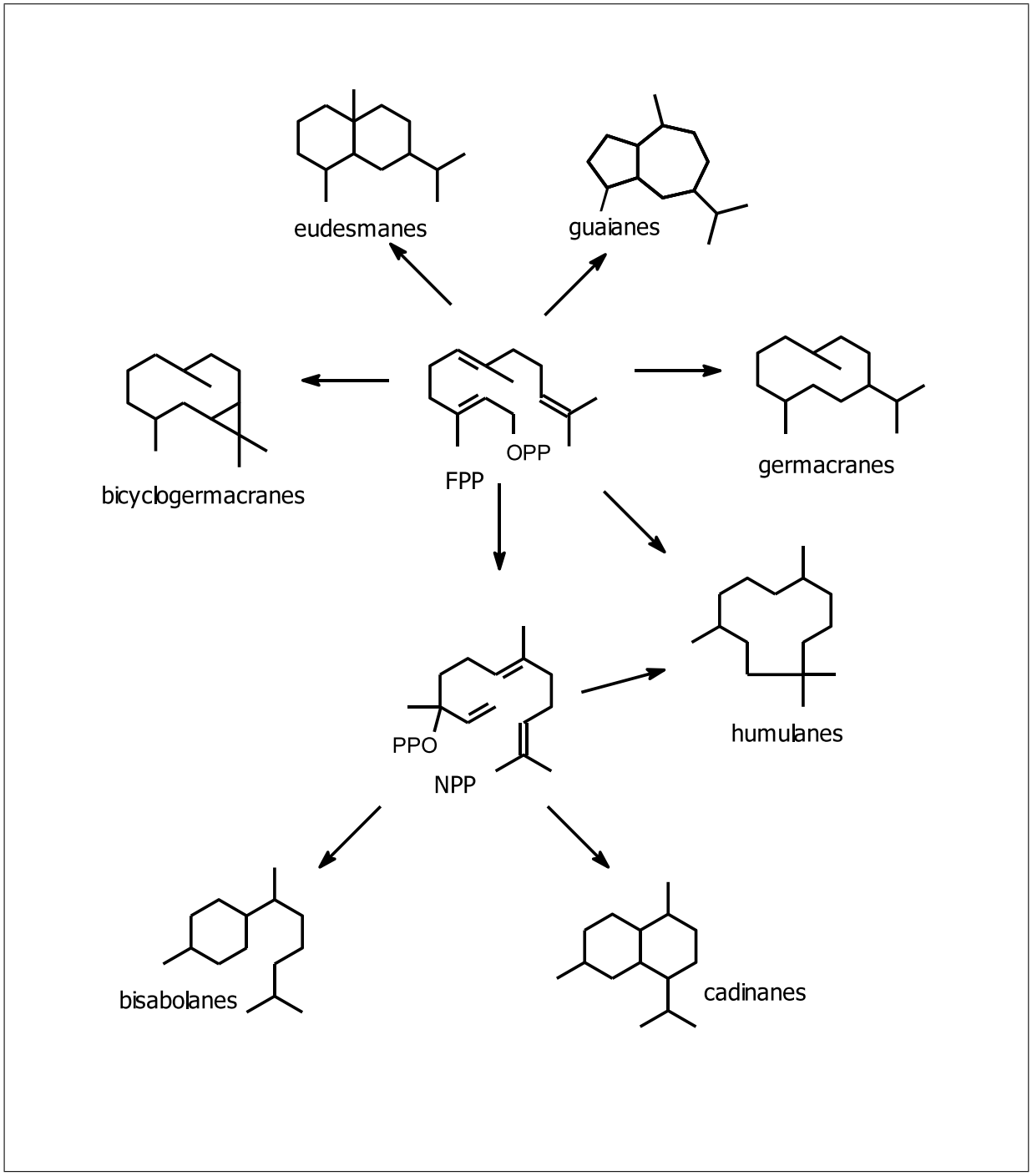


Figure 3.6: Cyclization scheme of farnesyl diphosphate (FPP) and Nerolidyl diphosphate (NPP) to some sesquiterpene carbon skeletons.

Recognizing that there is a geometric barrier to direct cyclization of *E,E*- farnesyl diphosphate to six-membered rings, as well as to 10- and 11-membered rings containing a cis double bond, it was subsequently proposed that FPP would in such cases undergo an initial isomerization to the corresponding tertiary allylic ester, nerolidyl diphosphate (NPP), which has the conformational flexibility and appropriate reactivity to allow formation of the derived cyclic products as shown in figure 3.6.

The derived cationic intermediates can then undergo further cyclizations and rearrangements, including methyl migrations and hydride shifts with the reaction being terminated by quenching of the positive charge, either by removal of a proton or capture of an external nucleophile, such as water or the original pyrophosphate anion.

Subsequent modifications of the basic parent skeletons produced by the terpenoid synthases are responsible for generating the myriad of different terpenoids produced by plants. These secondary transformations commonly involve oxidation, reduction, isomerization, and conjugation reactions. Many of the reactions introducing oxygen atoms into the terpenoid skeletons are performed by cytochrome P450 mixed-function oxidases.

All of the enzymes examined to date, from both plant and microbial sources, are operationally soluble proteins of molecular weights in the range of 40 kDa-100 kDa (550-850 aa). Several are monomers of 40 kDa-60 kDa while some are homodimers of subunits around 40 kDa, e.g. trichodiene synthase and patchoulol synthase. The enzymes themselves, which are all moderately lipophilic, require as cofactor a divalent metal ion, Mg^{2+} usually being preferred. The apparent Km values for the acyclic substrate, farnesyl diphosphate, are generally in the 0.5-15 μM range. In all these respects the sesquiterpene synthases strongly resemble prenyl transferases, as well as the biogenetically closely related monoterpene and diterpene synthases, which display similar patterns of molecular weight, lipophilicity, cofactor requirements, substrate binding affinity, and turnover rates [Barton et al., 1999], [Cane, 1990].

Based on the crystal structure of *epi*-aristolochene synthase [Starks et al., 1997], it appears that terpene synthases are composed of two distinct structural domains, a C-terminal active site domain, and an N-terminal domain that structurally resembles the catalytic cores of glycosyl hydrolysases. The protein is composed entirely of α-helices with short, connecting loops and turns, forming a two-layer α-barrel active site. The hydrophobic, aromatic residue-rich active-site pocket of *epi*-aristolochene synthase is formed by α-helices of the C-terminal domain to accommodate the olefin chain of the substrate, whereas the diphosphate moiety is complexed by two Mg^{2+} ions at the entrance of the active site. The two Mg^{2+} ions are coordinated by conserved aspartate residues, which are part of the conserved aspartate-rich motif DDxxD of terpene synthases [Bohlmann et al., 1998].

3.6 The plant *Solidago canadensis* L.

Solidago canadensis L., also called canadian golden rod (in German: kanadische Goldrute) is a perennial herb of the genus *Solidago* (family: *Asteraceae*). About 120 species belong to the genus. The stems erect 30-150 cm, glabrous at the base and pubescent or scabrid at least in the upper half. The leaves are alternate, lanceolate, long-attenuate with two prominent lateral veins. Capitula on erecto-patent to patent curved branches forming terminal pyramidal yellow inflorescence. A drawing of *Solidago canadensis* is shown in figure 3.7. The plant originates from North America, but is now widespread in Europe as a naturalized escape [Stace, 1991], [Tutin et al., 1976].

Solidago canadensis including *Solidago gigantea* is used medicinally as Solidaginis herba. The tea made of dried flowers, leaves and stems, collected during flowering time, has diuretic, spasmolytic and anti-inflammatory activities [Hänsel et al., 1994], [Leuschner, 1995]. The containing flavonoids and saponins are claimed to be responsible for these effects [Reznicek et al., 1991].

Figure 3.7: Solidago canadensis [Rothmaler et al., 1991]

3.7 Sesquiterpenes in *Solidago canadensis*

3.7.1 (+)- and (–)-Germacrene D

Germacrene D is a sesquiterpene widespread in nature. While (–)-germacrene D is common in higher plants like *Asteraceae* [Yoshihara et al., 1969], (+)-germacrene occurs mainly in lower plants e.g. liverwort [König et al., 1996]. Some exceptions were found in the gymnosperm *Podocarpus spicatus* [Lorimer and Weavers, 1987] and the soft coral *Sinularia mayi* [Beechan and Djerassi, 1978]. Though the latter is controversial, because the sesquiterpenes in *Sinularia* are discussed to be generated by symbionts or by a symbiotic relationship with microorganisms [Rudi et al., 1998].

Solidago species as a higher plant is exceptional in containing both enantiomers of germacrene D (figure 3.8), often in equal amounts and furthermore germacrene D being the main sesquiterpene [Niwa et al., 1980]. The proportion of both enantiomers present in the plant is varying widely depending on the collection site [Bülow and König, 2000]. Formation of all sesquiterpenes is catalyzed by sesquiterpene synthases from the general precursor farnesyl diphosphate (FPP). The existence of both enantiomers led to the question, if two separate germacrene D-synthases enantiospecificly catalyze the formation of each enantiomer or one synthase catalyzes the formation of both enantiomers. The answer was given by the work of C.O. Schmidt by isolating and separating two enantiospecifically working germacrene D-synthases from *Solidago canadensis* [Schmidt, 1998], [Schmidt et al., 1998]. Moreover he elucidated the biosynthetic pathway by incubation of the isolated enzymes with isotopically labeled FPP and mass spectrometric analysis of the products.

Already 1977 Nishino et al. [Nishino et al., 1980], [Nishino et al., 1977] showed that germacrene D is a mimic of the sex pheromone of the American cockroach, that gave antennal and behavioral responses. Antennal responses of insects to germacrene D have been shown by the use of GC-EAG (gas chromatography linked to electroan-

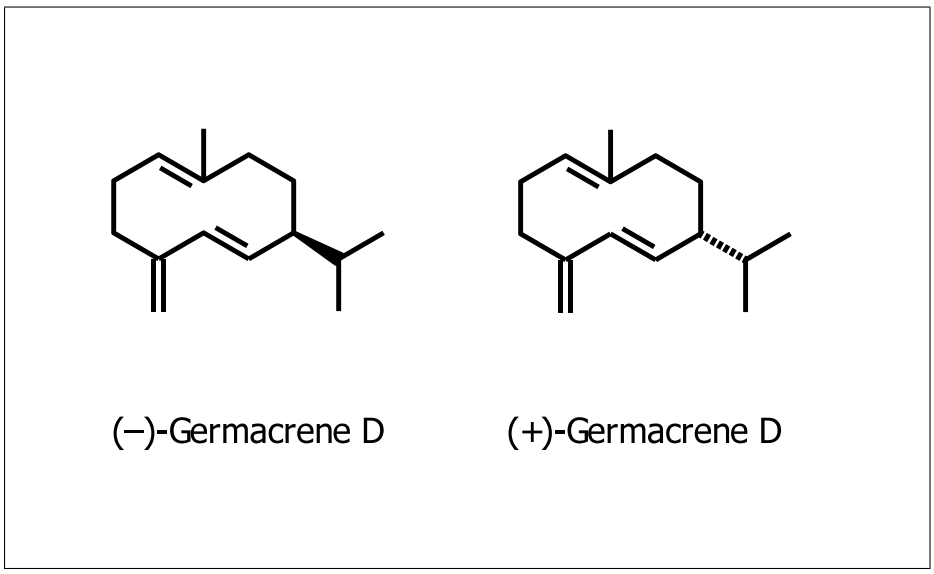


Figure 3.8: Both enantiomers of germacrene D

tenogram recordings) for many other insects e.g. for the female codling moth *Cydia pomonella* L. [Bäckmann et al., 2001], the two-spotted stinkbug *Perillus bioculatus* F. [Weissbecker et al., 2000] and the moth *Helicoverpa armigera*. For the latter it was found, that the receptor neurons responded to both enantiomers of germacrene D, but (–)-germacrene D had a 10 times stronger effect than (+)-germacrene D [Stranden et al., 2002]. Recently host plant recognition effects have been also reported. (–)-Germacrene D increases attraction of and oviposition by the mated tobacco budworm moth *Heliothis virescens* F. [Mozuraitis et al., 2002]. Increased attraction was also found to damaged, germacrene D enriched host plants for the leaf beetle *Oreina cacaliae* Schrank [Kalberer et al., 2001]. In contrast, a masking effect of (–)-germacrene D on host attraction is found for the cerambycid beetle *Monochamus alternatus* Hope [Yamasaki et al., 1997].

Moreover, germacrene D is considered to be an important intermediate in the biosynthesis of sesquiterpenoid compounds [Yoshihara et al., 1969]. It is capable of a large variety of rearrangements. Bülow and König succeeded in identifying a large num-

ber of acid catalyzed, photochemically and thermally induced rearrangement products [Bülow and König, 2000].

Biosynthesis of sesquiterpene enantiomers (+)- and (–)-germacrene D

The biosynthesis of the enantiomers (+)- and (–)-germacrene D was first proposed by Niwa et al. [Niwa et al., 1980]. Schmidt et al. [Schmidt et al., 1999b] investigated the biosynthesis of (+)- and (–)-germacrene D in *Solidago canadensis*. He incubated partially purified enzyme preparations from *Solidago canadensis* with farnesyl diphosphate (FPP), deuterium labeled on different positions, and analyzed the product formation using gas chromatography with mass spectrometry (GC-MS) and linked-scan mass spectromic methods. The investigation of Schmidt et al. confirmed the biosynthesis proposed by Niwa et al. for (+)- and (–)-germacrene D. A schematic overview is given in figure 3.9.

(–)-Germacrene D is formed by an initial cyclization of FPP (**1**) under elimination of the diphosphate group to germacrenyl cation (**2**) carrying a positive charge at position 11. During this step the configuration at C-7 is already fixed. The second step is a 1,3-hydride shift of the pro-*R*-H-6 into the isopropyl group. Furthermore the positive charge is shifted to position 6, generating the germacrenyl cation (**3**). Finally, deprotonation at C-15 takes place and two electron pair shifts quench the positive charge to generate (–)-germacrene D (**4**).

The initial step of the formation of (+)-germacrene D is the same as in the formation of the (–)-enantiomer (**4**) by abstraction of the diphosphate under ring closure to yield the germacrenyl cation (**2**). The next step is a 1,2-hydride shift of H-7 into the isopropyl group, generating germacrenyl cation (**5**) with the positive charge at position 7. A second 1,2-hydride shift of H-6 yields the cation (**6**) with the positive charge at position 6. During this step the configuration at C-7 is fixed. Finally (**6**) is deprotonated at

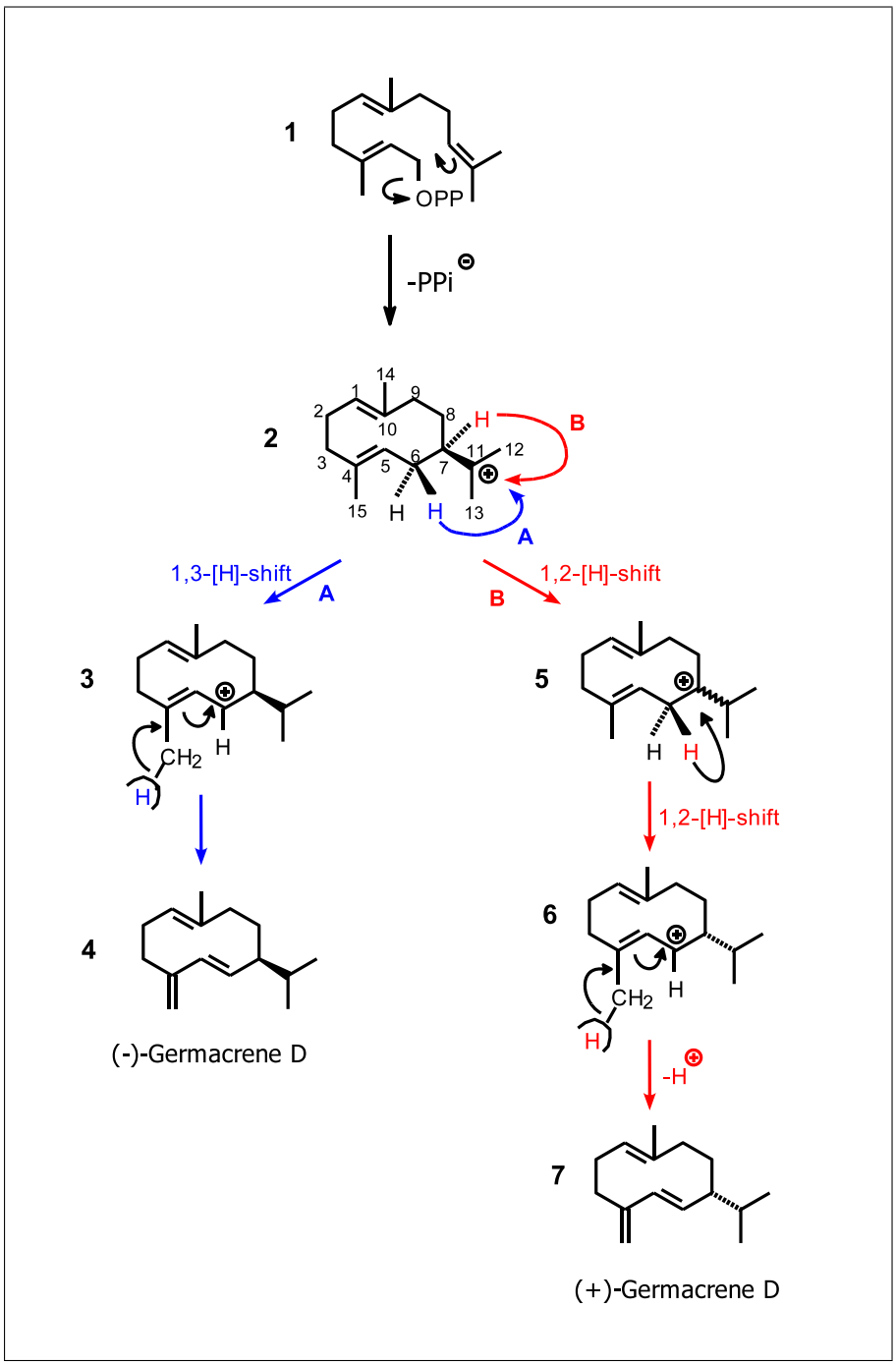


Figure 3.9: Biosynthesis of (–)-germacrene D (**4**) and (+)-germacrene D (**7**).
1 is the general precursor farnesyl diphosphate and **2**, **3**, **5**, **6** are germacrenyl cation intermediates.

C-15 and the positive charge is quenched by two electron pair shifts to generate (+)-germacrene D (**7**).

3.7.2 Cyclocolorenone and (–)-α-gurjunene

Another major component of the essential oil of *Solidago canadensis* is cyclocolorenone. Schmidt et al. [Schmidt et al., 1999a] showed, that (–)-α-gurjunene is the sesquiterpene hydrocarbon intermediate in the biosynthesis of cyclocolorenone. He proposed, that *in planta* (–)-α-gurjunene is oxidized to cyclocolorenol by a cytochrome P450-dependent hydroxylase and subsequently further oxidized by an NAD^+- or $NADP^+$-utilizing dehydrogenase to cyclocolorenone as shown in figure 3.10.

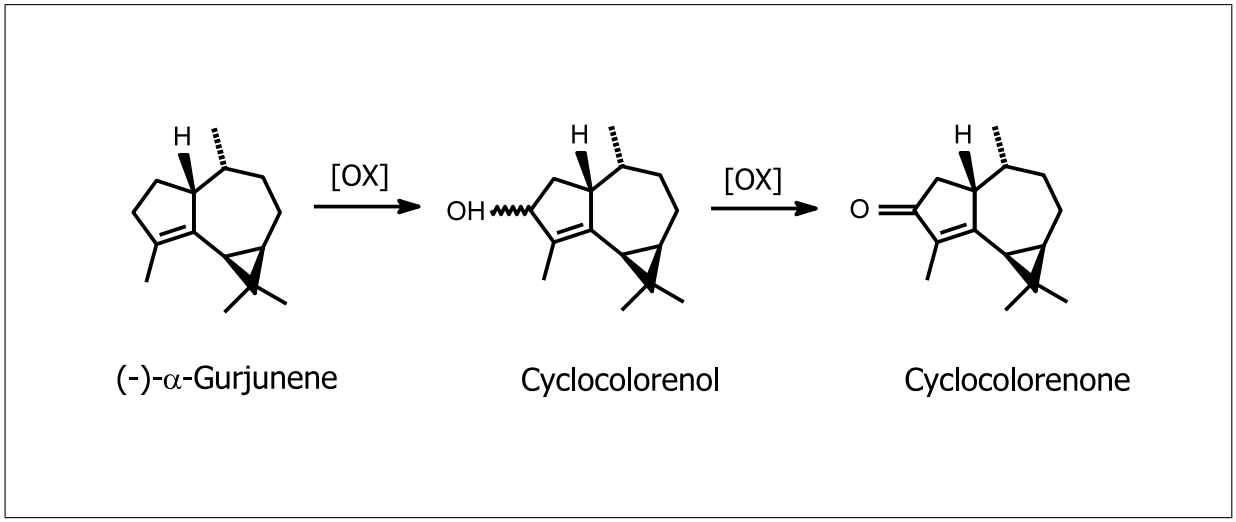


Figure 3.10: Biosynthesis of cyclocolorenone

α-Gurjunene itself is the major component of the gurjun balsam oil [Ruzicka et al., 1923], which is isolated from various *Dipterocarpus* species, like *Dipterocarpus dyeri* [Palmade et al., 1963], *Dipterocarpus alatus* and *Dipterocarpus turbinatus* [Streith et al., 1962]. The gurjun balsam oil is largely used in perfume industry as the light woody flavor suits well to other essential oils. In addition the oil has good fixative properties, needed

especially in soap perfumes [Bauer et al., 1990]. Further α-gurjunene is used as starting material for the guaiazulene synthesis [Treibs, 1952]. It is also a constituent of the essential oils of various plants like *Pogostemon cablin* Benth. (patchouli oil) [Tsubaki et al., 1967] and *Myroxylon balsamum* L. (Tolu balsam or Balsamum tolatum) [Friedel and Matusch, 1987].

Cyclocolorenone has been isolated and characterized from *Pseudowintera colorata* [Corbett and Speden, 1958], *Solidago canadensis* [Krepinsky and Herout, 1962], *Porella vernicosa* [Asakawa and Arantani, 1976] and *Magnolia grandiflora* [Jacyno et al., 1991]. Jacyno et al. [Jacyno et al., 1991] demonstrated also, that cyclocolorenone has phytotoxic, antibacterial and antifungal properties.

Biosynthesis of (–)-α-gurjunene

The formation of (–)-α-gurjunene in *Solidago canadensis* is induced by an (–)-α-gurjunene synthase using the general precursor FPP. The (–)-α-gurjunene synthase catalyzes mainly the formation of (–)-α-gurjunene with (+)-γ-gurjunene as a side product. The biosynthetic mechanism was investigated by Schmidt et al. [Schmidt et al., 1999a] using partially purified enzymes and FPP, deuterium labeled in different positions. The product formation was analyzed by GC-MS. Schmidt et al. proposed the following mechanism, shown in figure 3.11.

The initial step is the formation of the germacrenyl cation (**2**) from FPP (**1**) after abstraction of the diphosphate group. Internal shifts of electron pairs lead to the patchoulenyl cation (**3**), which then undergoes a double 1,2-hydride shift to yield cation (**5**). During the second shift only H_A migrates stereospecifically, while H_B remains in its initial position. Rearrangement of (**5**) leads to the aromadendrenyl cation (**6**), which then undergoes a 1,2-hydride shift to generate cation (**7**). Abstraction of proton H_A by the enzyme leads isoledene (**8**). Isoledene is then reprotonated at the active site of the enzyme to yield cation (**9**). At this junction two proton abstraction

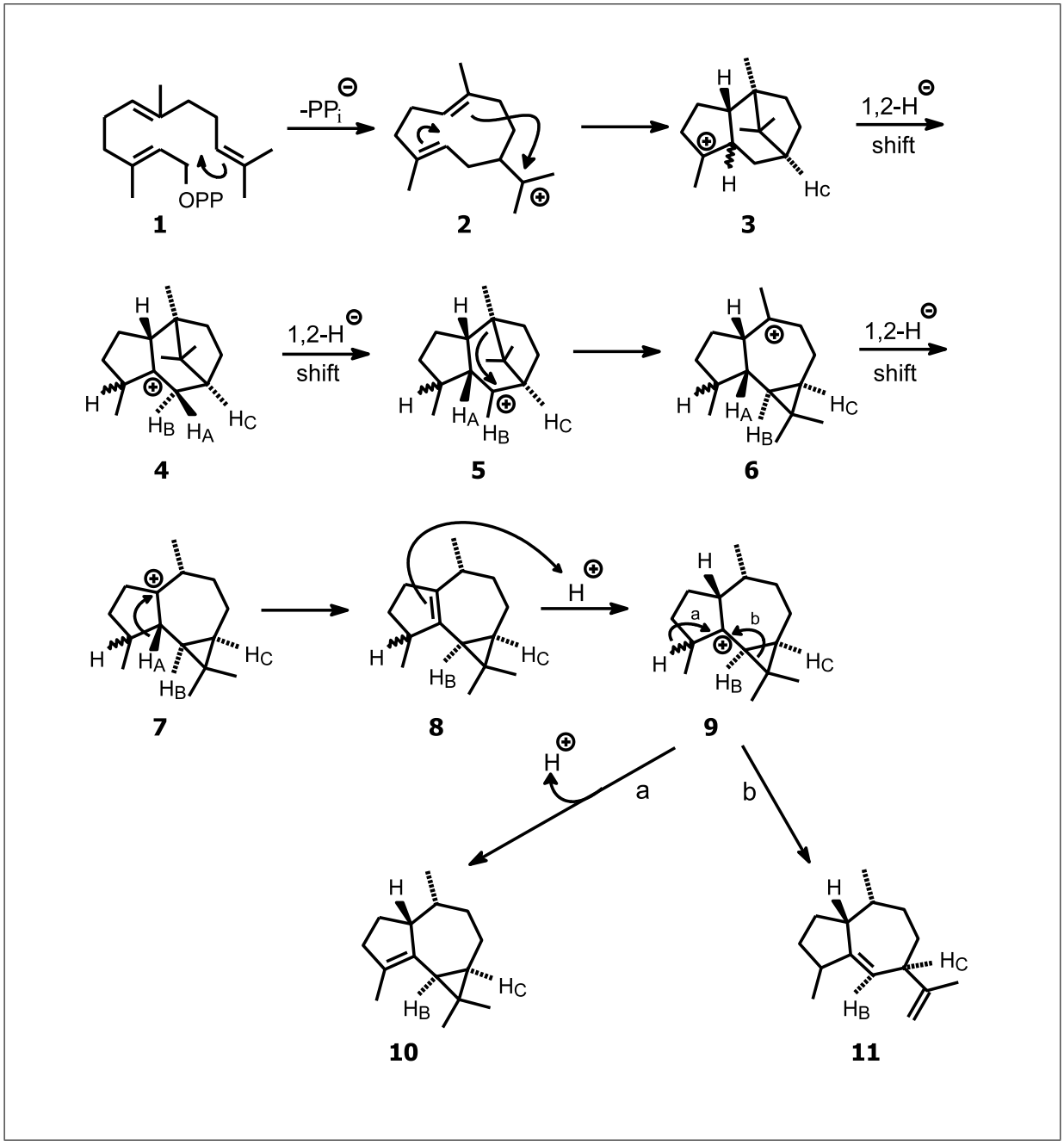


Figure 3.11: Biosynthesis of (–)-α-gurjunene (**10**) and (+)-γ-gurjunene (**11**).

1 is the general precursor farnesyl diphosphate, **2** germacrenyl cation, **3** patchoulenyl cation, **6** aromadendryl cation, **8** isoledene and **4**, **5**, **7**, **9** are cation intermediates.

are possible, one leading to (–)-α-gurjunene (**10**) as the main product and the other to (+)-γ-gurjunene (**11**) as a side product.

3.7.3 Cascarilladiene

A minor component, which was recently found in extracts from the roots of *Solidago canadensis* is cascarilladiene. Cascarilladiene is a component of cascarilla oil, which is prepared by steam distillation of the bark of *Croton eluteria* Benett [Claude-Lafontaine et al., 1976], [Hagedorn and Brown, 1991]. The structure of cascarilladiene was revised in 1989 by Weyerstahl et al. [Weyerstahl et al., 1989] and is shown in figure 3.12. Besides that, cascarilladiene was also detected in Haitian vetiver oil (*Vetiveria zizanioides*) [Weyerstahl et al., 2000] and in the liverwort *Preissia quadrata* [König et al., 1996].

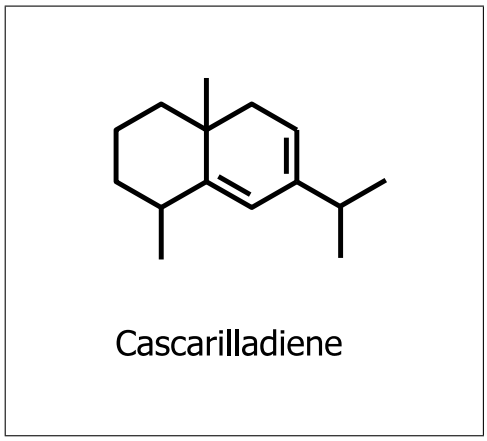


Figure 3.12: Structure of cascarilladiene

Biosynthesis of cascarilladiene

The biosynthesis of cascarilladiene was not investigated yet. A hypothesized pathway is shown in figure 3.13, which includes germacrene C (**4**) as a possible intermediate.

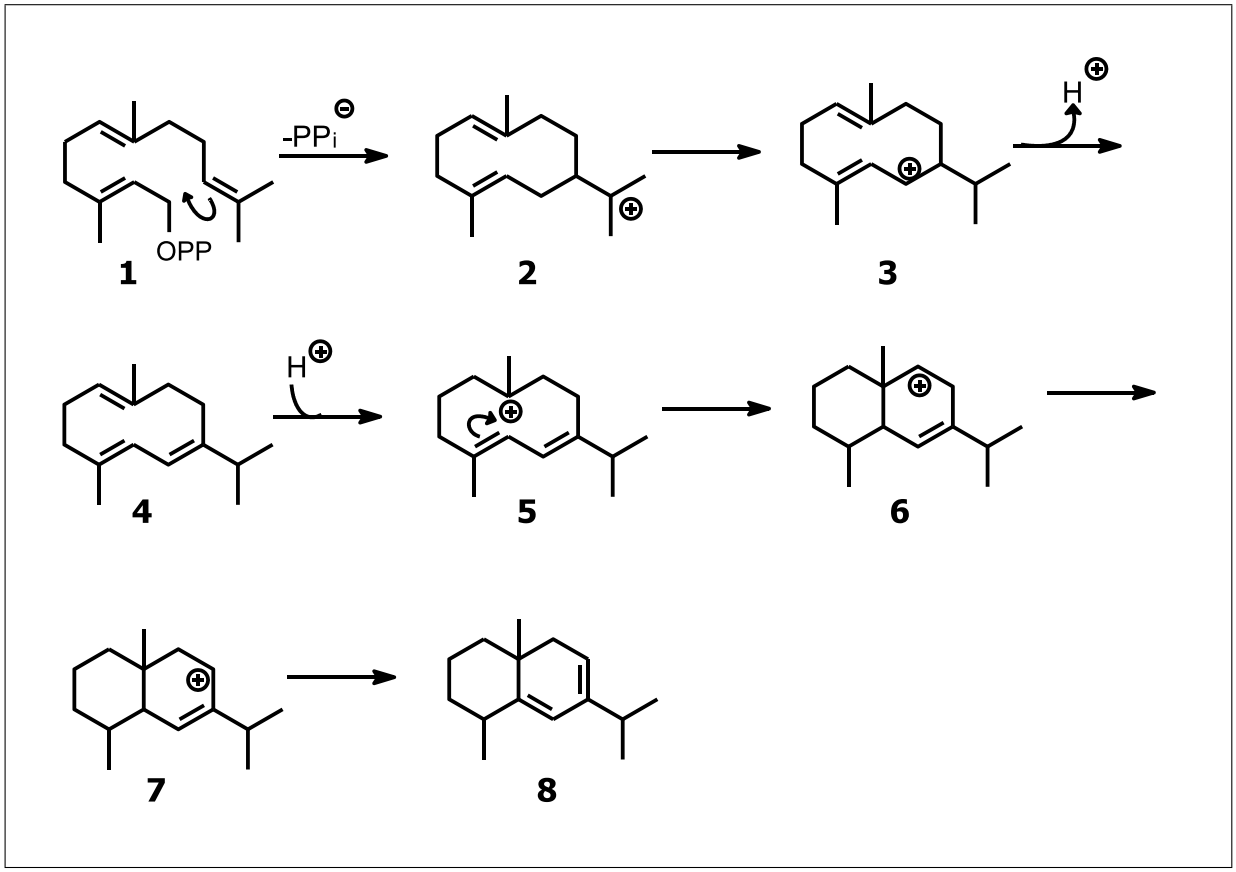


Figure 3.13: Biosynthesis of cascarilladiene (**8**).
1 is the general precursor farnesyl diphosphate, **4** is the intermediate germacrene C and **2**, **3**, **5**, **6**, **7** are cation intermediates.

3.8 Genomic organization of plant terpene synthases

One of the most important discoveries in terpene synthase research is the observation that all of the known plant terpene synthases appear to be closely related. Plant terpene synthase similarities include regions of sequence conservation and the positioning of intron sequences [Mau and West, 1994]. Recognition of plant terpene synthase sequence similarities has had important implications for studies of these enzymes and efforts to isolate plant terpene synthase genes. Common to most terpene synthase amino acid sequences is the aspartate-rich motif DDxxD, which is important for the

binding of the covalent metal ion (generally Mg^{2+}) involved in the initial ionization step of the reaction [Starks et al., 1997].

Comparison of the genomic structure of the plant terpene synthase family by Trapp and Croteau [Trapp and Croteau, 2001] indicates that the terpene synthase gene family consists of three classes based on intron/exon patterns. A schematic overview of the terpene synthase gene family is shown in figure 3.14.

Class I terpene synthase genes comprise 12-14 introns and 13-15 exons (including the conifer diterpene internal sequence domain (CDIS)) and consist primarily of diterpene synthases found in gymnosperms (secondary metabolism) and angiosperms (primary metabolism). Class II terpene synthase genes comprise 9 introns and 10 exons and consist of only gymnosperm monoterpene and sesquiterpene synthases involved in secondary metabolism. Class III terpene synthase genes comprise 6 introns and 7 exons and consist of angiosperm monoterpene, sesquiterpene, and diterpene synthases involved in secondary metabolism.

Trapp and Croteau concluded that the terpene gene family derives from an ancestral class I type terpene synthase gene of primary metabolism common to both gymnosperms and angiosperms. This evolved into the class II terpene synthases in gymnosperms and into the class III terpene synthases in angiosperms.

3.9 Cloning of terpene synthase genes

Purification of plant terpene synthases is often complicated by factors such as localized expression within specific tissues and subcellular organelles and low levels of enzyme expression. Despite these obstacles, several plant terpene synthases have been purified from plant tissues e.g. [Schmidt et al., 1998], [de Kraker et al., 1998], [Pichersky et al., 1995] and [Alonso et al., 1992]. Initial efforts to isolate terpene synthase genes from plants employed probes based on either amino acid sequence data from

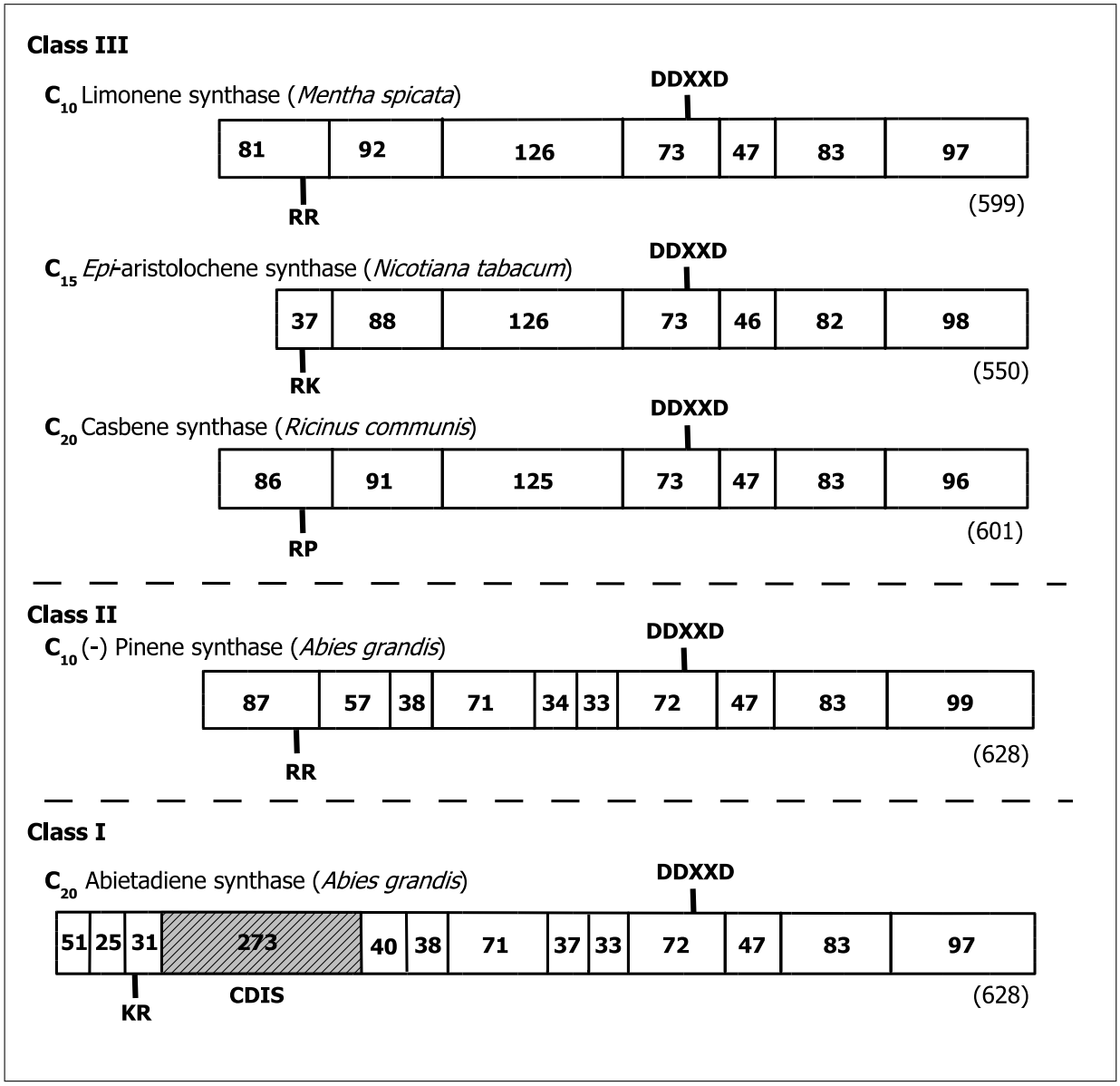


Figure 3.14: Schematic amino acid sequence alignment of the three terpene synthase gene family classes . Vertical bars represent introns, blocks represent exons with specified length. Shown is also the conserved motif DDXXD and the conserved sequence motif (K/R)(R/P/K), downstream the plastidial targeting sequence. CDIS= conifer diterpene internal sequence domain.

purified enzymes [Colby et al., 1993] or specific antibody preparations [Lois and West, 1990]. Comparisons between the first three available terpene synthase sequences es-

tablished that all three enzymes are related [Mau and West, 1994]. These enzymes represented a monoterpene synthase (limonene synthase from spearmint) [Colby et al., 1993], a sesquiterpene synthase (5-*epi*-aristolochene synthase from tobacco) [Facchini and Chappell, 1992], and a diterpene synthase (casbene synthase from castor bean) [Mau and West, 1994]. The alignment is shown in figure 5.2 on page 63, a schematic overview is given in figure 3.14. Identities in global alignments between the three enzymes ranged from 31% to 42%. Nowadays higher levels of identity (around 90%) have been observed between enzymes from closely related plant species e.g. germacrene A-synthase from chicory [Bouwmeester et al., 2002] and lettuce [Bennett et al., 2002]. Based on available terpene synthase sequence information, mono-, sesqui-, and diterpene synthases constitute a gene superfamily within plants and, in many instances, the similarities between enzymes are sufficient to permit the application of sequence homology-based gene cloning techniques.

3.9.1 Heterologous expression of terpene synthase genes

Difficulties associated with the isolation of terpene synthases were seen as limiting factors for studies of many enzymes. Expression of terpene synthase genes in *Escherichia coli* has greatly reduced this barrier and promises to facilitate all aspects of terpene synthase research.

Many plant terpene synthases have been expressed in *E. coli* up to now. Sufficient enzyme was often obtained to confirm gene identity by GC-MS analysis of the resulting enzyme reaction products [Bohlmann et al., 1998]. Expression of monoterpene and diterpene synthase genes is often complicated by the presence of plastidial target sequences at the N-terminus, which target the nuclear-encoded preproteins to the plastids for proteolytic processing to the mature forms. High-yield expression of soluble terpene synthases can be achieved by truncation of the cDNAs to remove the targeting sequences. Another commonly encountered problem with expression of plant

terpenoid synthases in *E. coli* relates to the frequency of arginine residues that use rare tRNAs in the prokaryotic host. Coexpression of the terpenoid synthase cDNA with the required tRNA can eliminate translational difficulty and yield high-level expression of active recombinant enzymes. Still terpene synthases not showing any function upon heterologous expression are found e.g. [Aubourg et al., 2002] and [van Gelder et al., 2000].

Identification of terpene synthase genes isolated from plants only on knowledge of the nucleotide or amino acid sequence is not possible up to date, because plants frequently possess numerous terpene synthase genes and many terpene synthases are closely related. For this reason, heterologous expression of putative terpene synthase genes is essential and often critical for gene identification. [Barton et al., 1999]

Recently Martin et al. [Martin et al., 2003] have managed to engineer *Echerichia coli* to enable high level production of terpenoids *in vivo*. This opens the possibility for the cheap production of rare terpenoids with high impact for pharmaceutical or flavor and fragrance companies in large quantities, when the coding sequence is known.

Chapter 4

Methods

In the following chapter a brief introduction is given to techniques in molecular biology, which were used within the present work. The content of this chapter refers to the standard study books [Griffiths et al., 2000] and [Schmid, 2002].

4.1 The genetic code

All genetic information of an organism is written in its DNA (deoxyribonucleic acid). The DNA is composed of four nucleotide building blocks. Each building block consists of a purine base adenine (A), guanine (G), or a pyrimidine base thymine (T), cytosine (C), which is N-glycosidically linked to a deoxyribose-5'-phosphate in position 1. The nucleotide building blocks are linked over phosphate bridges between the 5'-C-atom of one nucleotide to the 3'-C-atom of the other resulting in a sugar-phosphodiester polymer (see figure 4.1). Because of the 5'-3'-binding each strand has a direction. The polymer has a double helix structure, with the purine and pyrimidine bases directed inwards. Adenine binds to thymine over 2 hydrogen bonds and guanine binds to cytosine over 3 hydrogen bonds. This results in two hybridizing DNA strands with com-

plementary bases. The complementary sequence of the bases allows the conservation and amplification of the genetic information.

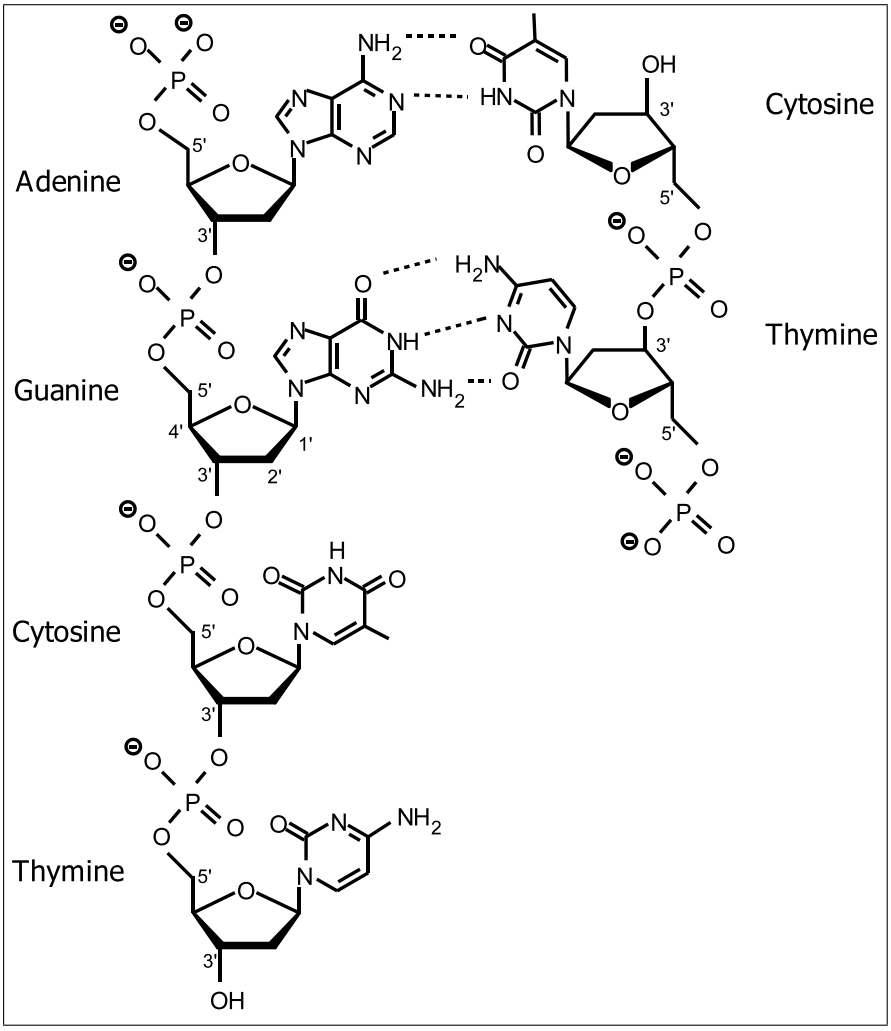


Figure 4.1: Structure of DNA, dashed lines are hydrogen bonds

The genetic information included in the DNA is coding for the biosynthesis of proteins. This process comes off in two steps. The transcription of the coding part of the DNA into mRNA (messenger ribonucleic acid) followed by translation of the information to proteins. The differences between prokaryotes and eukaryotes is, that in eukaryotes the mRNA has to undergo a maturing process (splicing) before translation,

because the eukaryotic DNA contains coding (exon) and non-coding (intron) parts. The intron parts are discarded during the splicing process, see figure 4.2.

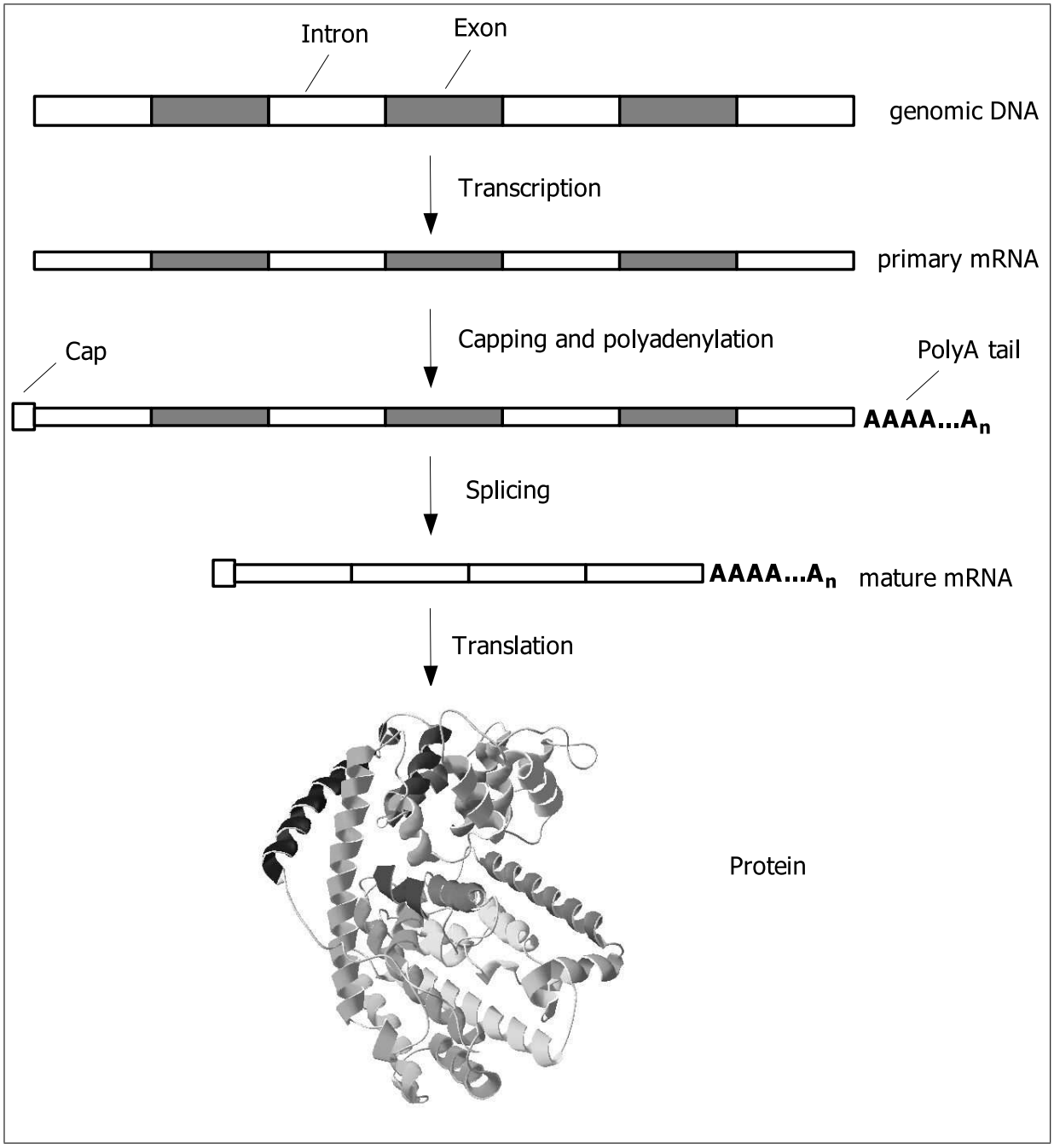


Figure 4.2: Schematic overview of transcription, splicing and translation.

The genetic code is, besides of some exceptions, universal. Three nucleotides (triplets) code for one amino acid. Because the genetic code is unified for all species, genetic information can be transferred from one organism to another organism beyond the border of species. This applies among others for viral infections and gene technology.

Theoretical triplets of three nucleotides with the possibility of four variations for the four bases results in $4^3 = 64$ possibilities. But only 20 amino acid are used for proteins, therefore several triplets code for the same amino acid. Additionally one triplet codes for the start (start codon ATG = methionine (Met)) of an open reading frame (coding part) and three triplets (TAA, TAG, TGA) are stop codons, which mark the end of the open reading frame. The genetic code is shown in figure 4.3.

4.2 Polymerase chain reaction (PCR)

With the PCR technique a specific part of the DNA can be amplified *in vitro*. For the reaction two oligonucleotides (primers) are needed for the ends of the desired DNA part (one for each DNA end). For this either the DNA sequence or the protein sequence has to be known. The DNA itself is needed as template, a mixture of the four deoxynucleotides as building blocks and a temperature resistant polymerase to synthesize the complementary strands. A schematic overview is given in figure 4.4.

The PCR undergoes three steps:

- **Denaturation:** heating of the DNA doublestrand to 94°C to separate the two strands
- **Annealing:** cooling down the temperature to 40-60°C to allow the primers to anneal to the single strand DNA

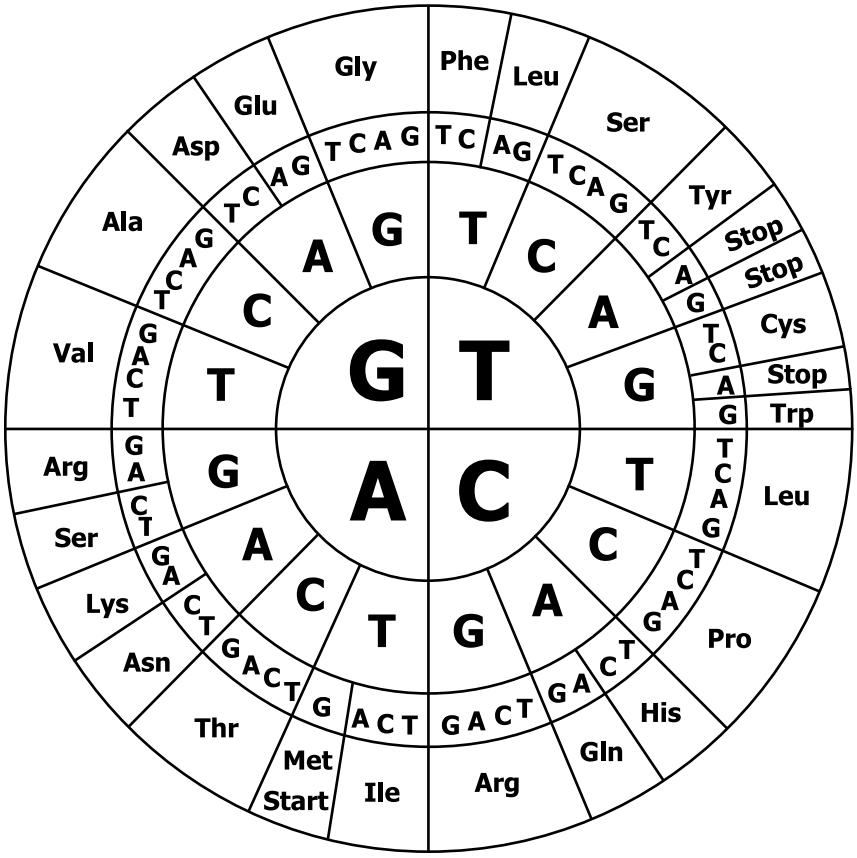


Figure 4.3: The genetic code.
Start stands for the start codon (ATG) of a coding sequence (open reading frame (ORF)),
Stop stands for the stop codon (TAA, TAG, TGA), which signals the end of an open reading frame

- **Extension:** increasing the temperature to 72°C to allow the polymerase to synthesize the two new complementary strands in 5′- to 3′-direction

A new heating up to 94°C allows a denaturation of the newly formed DNA strands and the reaction is repeated. As the reaction follows exponential growth, within 1-2 h time $2^{24} - 2^{40}$ copies are produced.

The basis for PCR is a polymerase, which doesn't get heat inactivated during the denaturation step. Therefore polymerases from thermophilic bacteria as *Thermus aquaticus* (e.g. *Taq* polymerase) are used. The size and the amount of the PCR products are analyzed by gel electrophoresis.

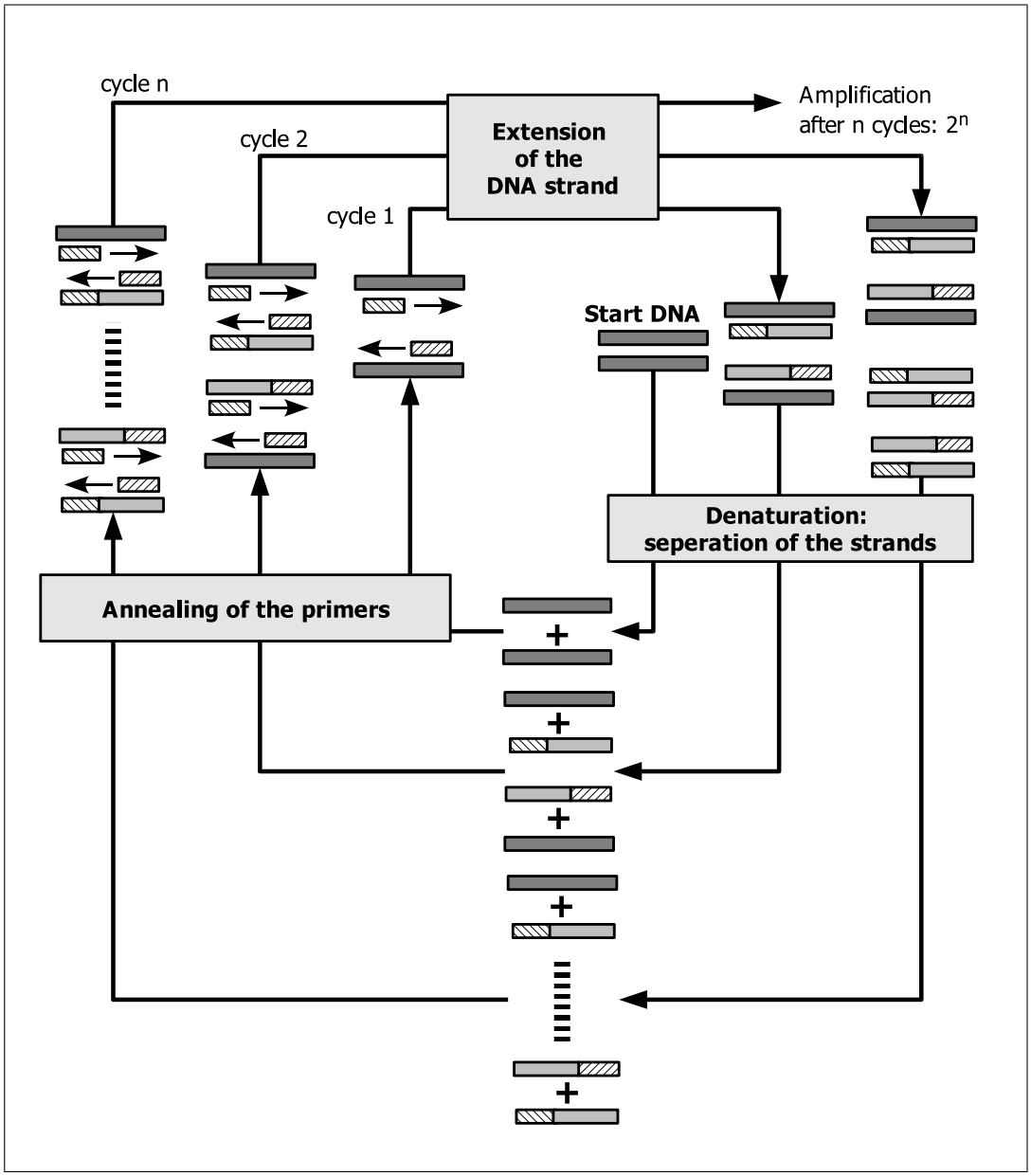


Figure 4.4: Polymerase chain reaction (PCR)

4.2.1 RT-PCR

The reverse transcriptase PCR (RT-PCR) is a specific PCR, in which mRNA is used as starting material. The first step in this PCR is the reverse transcription of the mRNA to a complementary DNA (cDNA) by a reverse transcriptase and a poly-dT primer, which anneals to the polyA tail of the mRNA. Then a normal PCR reaction follows amplifying the cDNA. The reverse transcriptase is isolated from retroviruses e.g. avian myeloblastosis virus (AMV) or moloney murine leukemia virus (MMLV).

4.2.2 Proofreading-PCR

The error rate (mutation per nucleotide per duplication) of *Taq* polymerase is around $8 \cdot 10^{-6}$, in other words, about 56% of a 200 bp amplification product will contain at least a single error after 1 million fold amplification. To reduce the error rate proof-reading polymerases are used e.g. *pfu* from *Pyrococcus furiosus*. The proof-reading polymerases exhibit next to the 5′- to 3′-DNA polymerase activity also 3′- to 5′-exonuclease activity. The exonuclease activity splits off mismatched nucleotides at the 3′-end. By this the error rate of the *pfu* polymerase is six-fold lower than the *Taq* polymerase. The PCR reaction itself is the same as the *Taq* polymerase, but often with lower fidelity resulting in less amplified product. Therefore *Taq* polymerase is used for general PCR reactions, while proofreading polymerase like *pfu* are used for specific applications like amplification of cDNA for heterologous expression.

4.2.3 RACE-PCR

Another specialized PCR is the RACE-PCR (rapid amplification of cDNA ends). This PCR technique is used to generate full-length coding cDNAs. To employ RACE-PCR a part of the cDNA sequence has to be known. The PCR is then performed from the

known part to the 3'-end and to the 5'-end. To perform RACE-PCR it is necessary to assemble adaptor sequences at the ends of the cDNA or mRNA.

For the 3'-end, reverse transcription of the mRNA is carried out with a poly-dT_{25} primer possessing an adapter sequence as shown in figure 4.5.

For the 5'-end a certain reverse transcriptase (here: MMLV) is used, that exhibits a terminal transferase activity by adding 3-5 residues, predominantly dC to the 3'-end. An overview is given in figure 4.6. The terminal stretch of dG residues of the 5'-adaptor oligonucleotide can anneal to the dC-rich cDNA 3'-end and serves as an extended template for the reverse transcriptase. A complete cDNA copy of the original mRNA is synthesized with the additional adaptor sequence at the 5'-end.

4.2.4 Cycle Sequencing

Cycle sequencing is done according to the method of Sanger [Sanger et al., 1970]. The sequencing of the DNA starts with a special type of PCR. For the sequencing PCR only one downstream primer (not a primer pair) is needed. Next to the mixture of deoxynucleotides, also a mixture of dideoxynucleotide-derivatives (e.g. dRhodamine dye terminator, see figure 4.8) with each of the base labeled with different fluorescent markers are used. Whenever during the PCR reaction a dideoxynucleotide-derivative is introduced, the extension reaction is terminated. This reaction statistically results in a mixture of fragments of all possible strand lengths. The mixture is then separated by capillary electrophoresis according to their fragment length (= mass) and the fragments are detected by their terminal fluorescence marker. The detection signal is directly readable as the DNA sequence. An overview is given in figure 4.7.

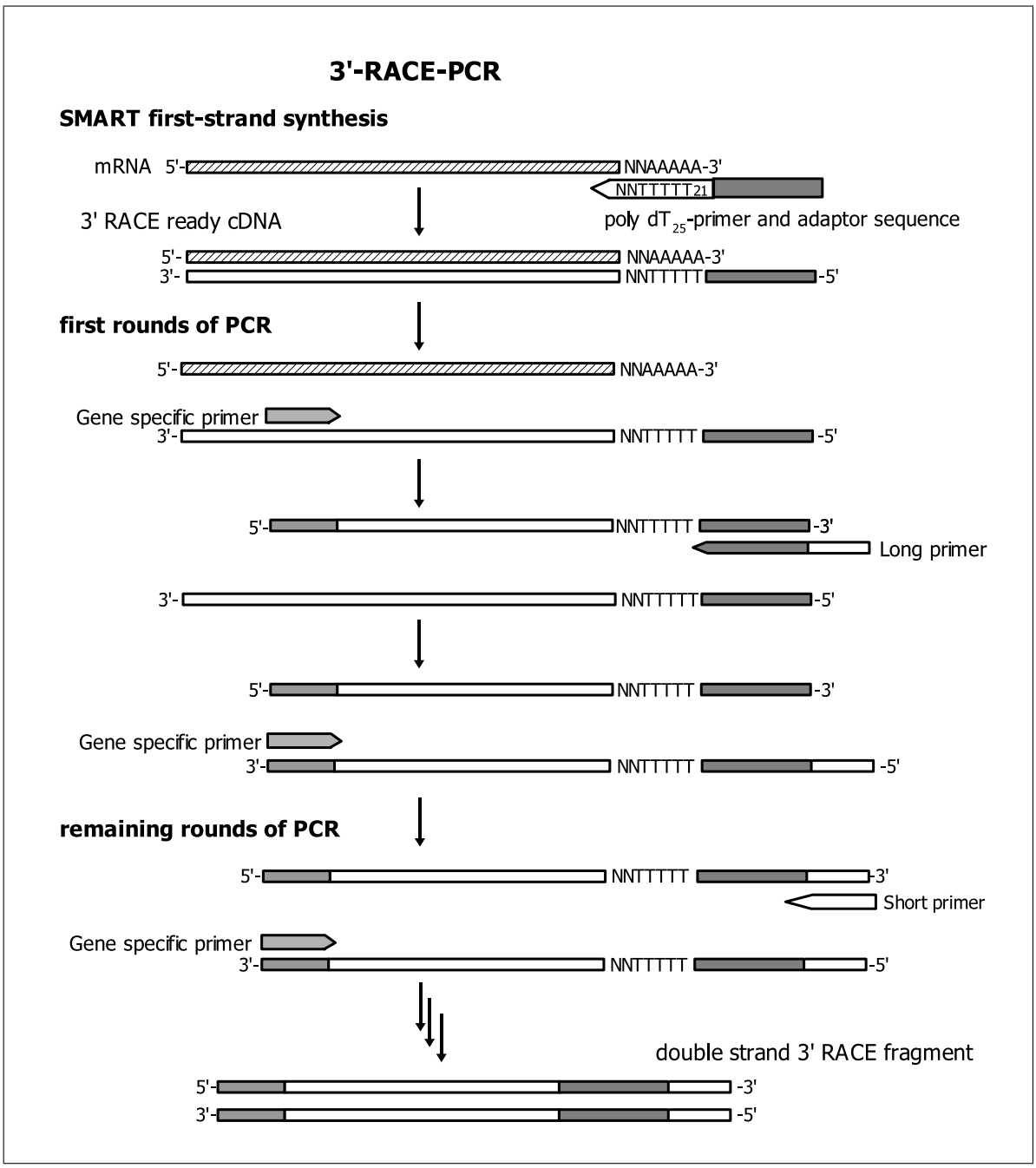


Figure 4.5: Schematic overview of the mechanism of the 3′-RACE reaction. To achieve higher product specificity the adaptor sequence is extended by the long primer during the PCR reaction.

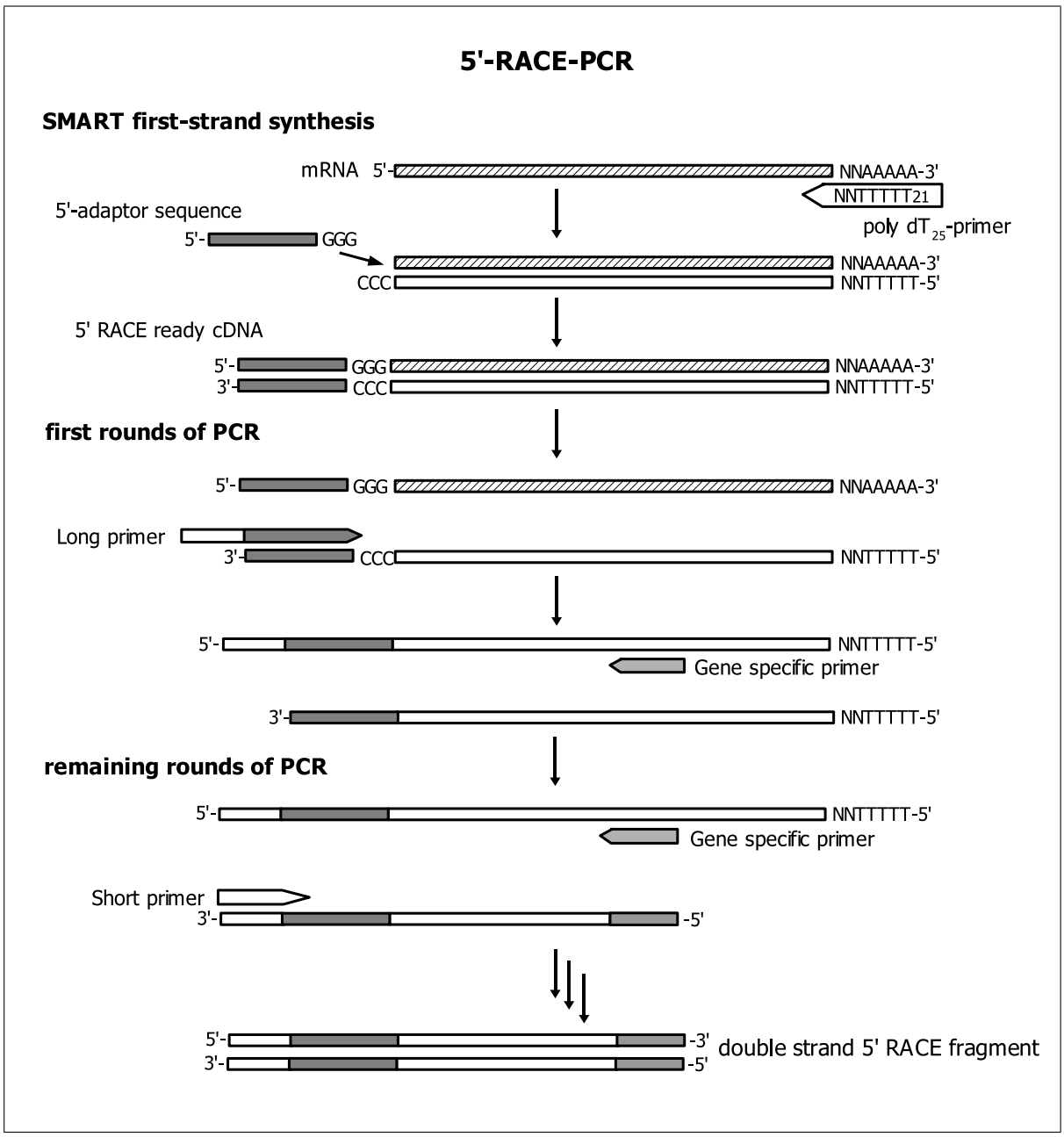


Figure 4.6: Schematic overview of the mechanism of the 5′-RACE reaction. To achieve higher product specificity the adaptor sequence is extended by the long primer during the PCR reaction.

4.3 Cloning and heterologous Expression in *E. coli*

Cloning is the isolation of a coding DNA section. As eukaryotic genomic DNA contains coding (exon) and non-coding (intron) parts, the mature mRNA containing only coding regions (exons) is taken as starting material.

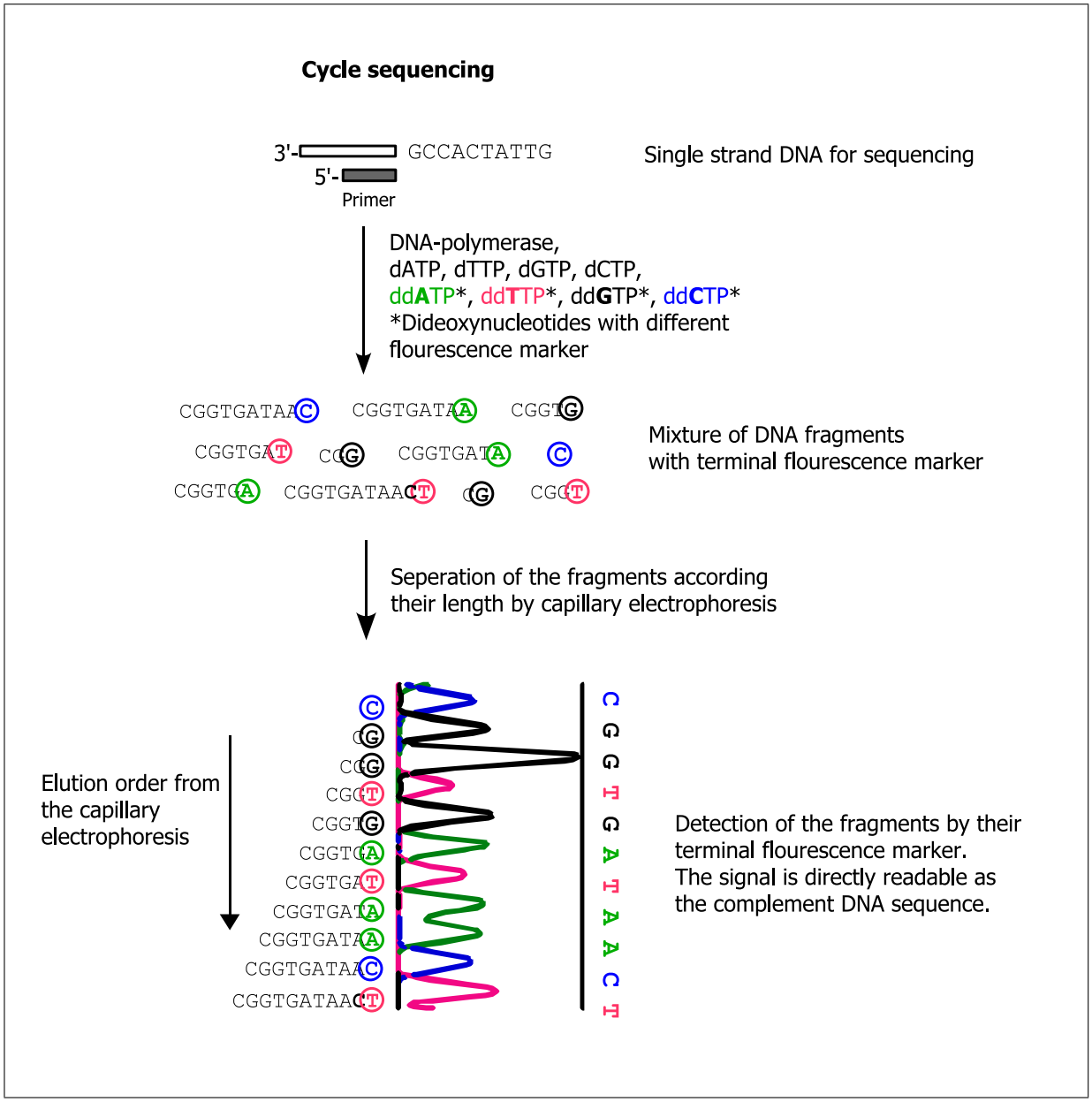


Figure 4.7: Cycle sequencing

DNA or cDNA can be transferred into a host cell (e.g. *E. coli*) in that way, that it will be replicated during cell division, transcribed into mRNA and translated to proteins ("gene expression"). Often bacteria are used for hosts as the transformation of non-host DNA in plasmids (vectors containing the introduced DNA) is a convenient method.

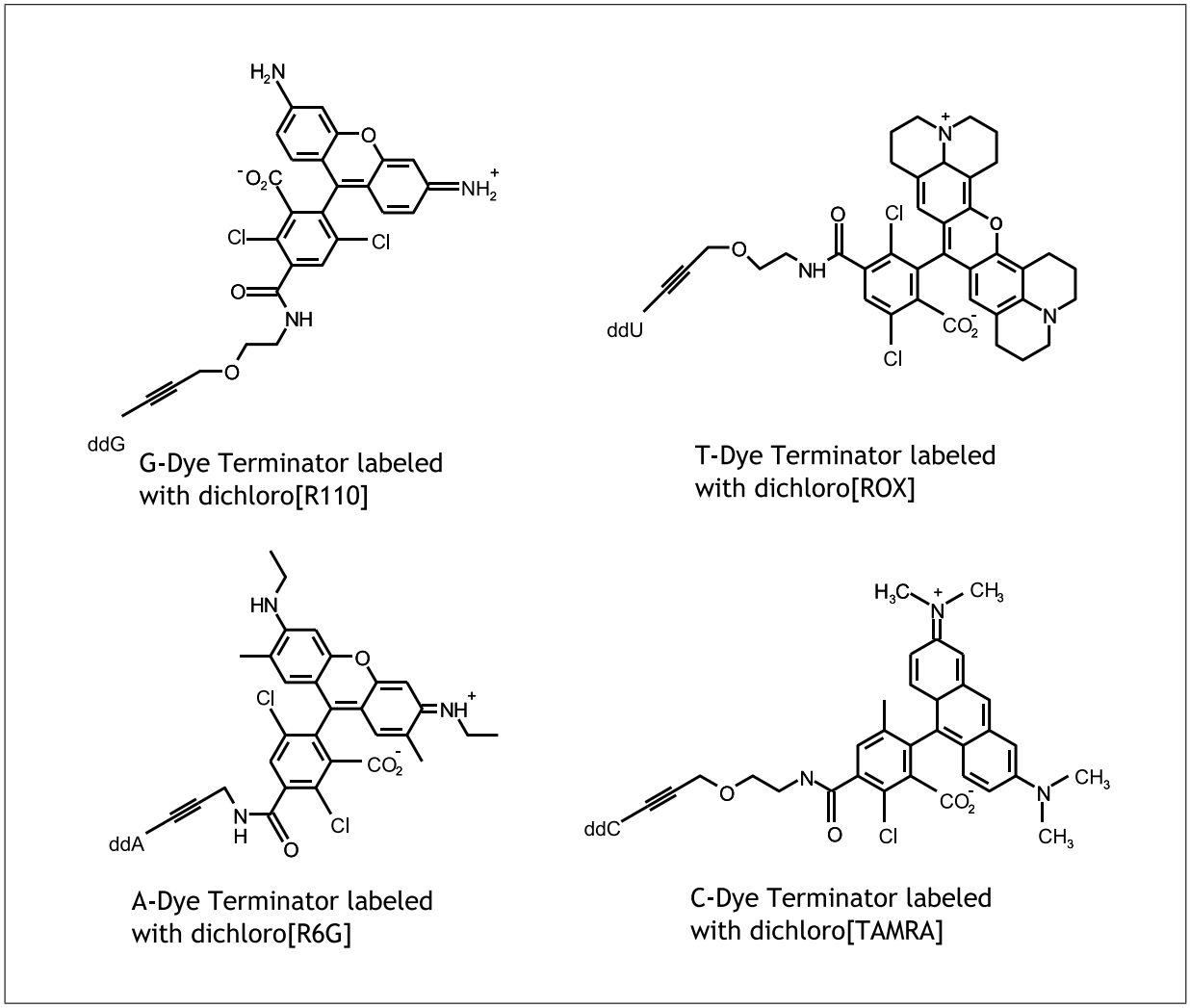


Figure 4.8: The structures of the dRhodamine dye terminators.

For heterologous expression of a coding cDNA in a host (e.g. *E. coli*) the DNA has to be transferred into the host. To do this in bacteria cloning- or expression-vectors are used. They contain an origin of replication (ori) for the replication of the plasmid in the host and a multiple cloning site (MCS) to ligate the desired non-host DNA. The plasmid also contains a marker for selection (e.g. antibiotic resistance) to select successfully cloned cells. For heterologous expression a promoter is needed, which regulates the transcription and translation of the introduced non-host DNA.

An often used promoter is the *lac* promoter, which is inducible upon addition of the inductor IPTG (isopropyl-β-D-thiogalactopyranoside) to the media. The open reading frame (ORF) of the cDNA including a start codon (ATG) and a stop codon is ligated into the multiple cloning site.

For this purpose the cDNA is prepared with recognition sites for restriction endonucleases (restriction enzymes). The restriction enzyme recognition site can be introduced to the DNA by PCR using primers, which contain the recognition site. Restriction enzymes recognize a specific DNA sequence (in a palindrome form) and cut the double strand DNA either blunt or with an overhang (also called sticky end). Doublestrand DNA cut with an overhang can be ligated with a ligase into a plasmid with the same overhang. A schematic overview is shown in figure 4.9.

Transformation of the plasmid into *E. coli* is often done by the heat-shock method. The cells have been made competent (cell wall gets porous for DNA) by treating with e.g. calcium ions. After adding the plasmid to the prepared cells a short heating (e.g. 42°C for 30 sec.) transfers the plasmid into the cells. The plasmid contains next to the introduced DNA also an antibiotic resistance marker. Only cells transformed with an plasmid can survive on media containing the antibiotic. A marker for the identification of successfully cloned DNA into the plasmid is the blue-white-screening method. A plasmid containing the *lacZ'*-gene is used in combination with host cells, containing the *lacZ* gene, but missing the *lacZ'* part (deletion *lacZ*ΔM15) (e.g. *E. coli* DH5α). Only transformed cells with an intact *lacZ'*-gene can produce β-galactosidase, which splits the leuko dye 5-bromo-4-chloro-3-indolyl-β-D-galactoside (x-gal) to the dark blue 5,5'-dibromo-4,4'-dichloro-indigo. The cloning site is in the middle of the *lacZ'*-gene on the plasmid. If DNA is successfully ligated into the plasmid, the *lacZ'*-gene will be interrupted and looses its function. Only white colonies have successfully ligated the DNA into the plasmid. The other colonies get blue, shown in figure 4.10. This technique is used e.g. in the $pGEMT_{easy}$ vector system.

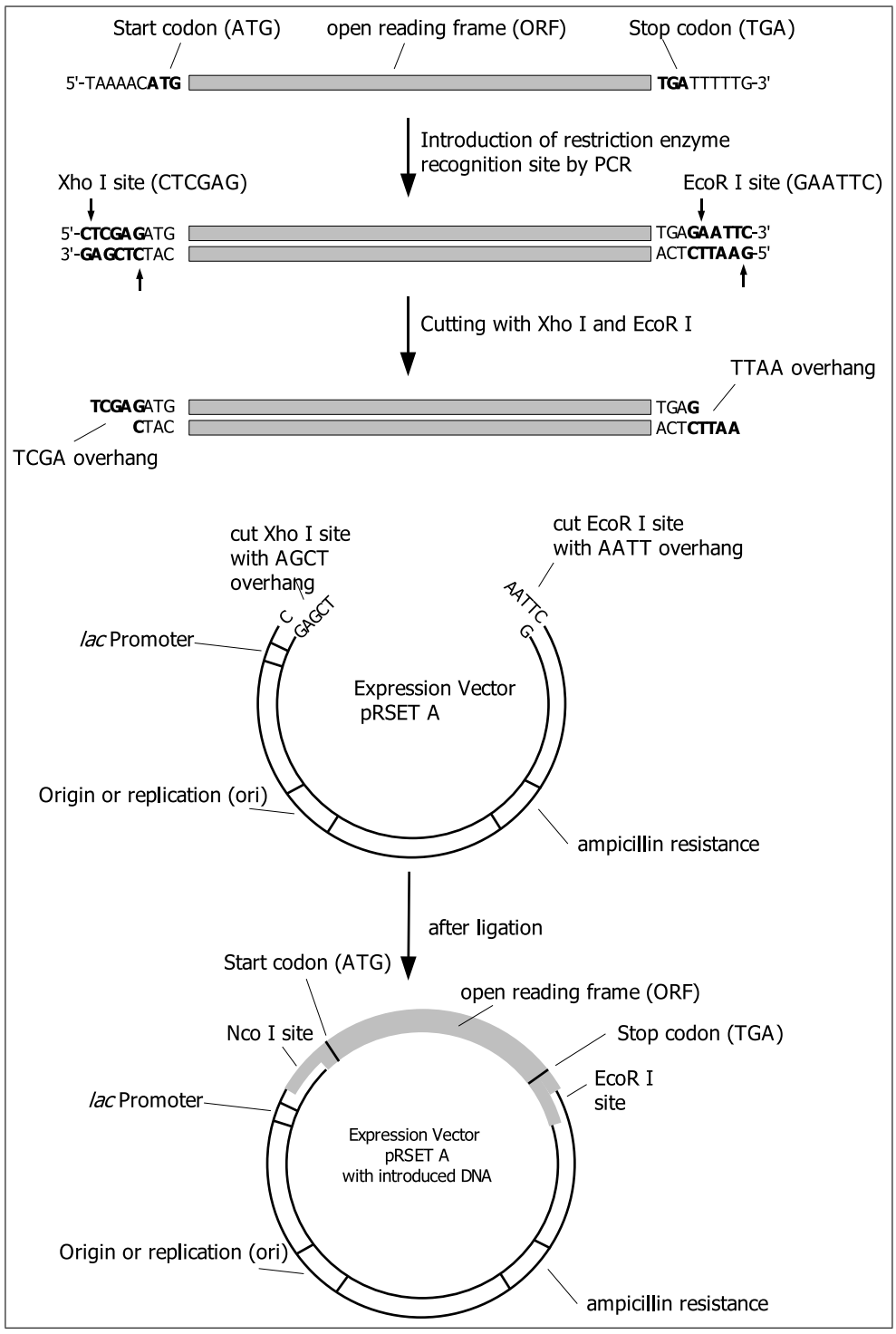


Figure 4.9: Schematic overview of cloning

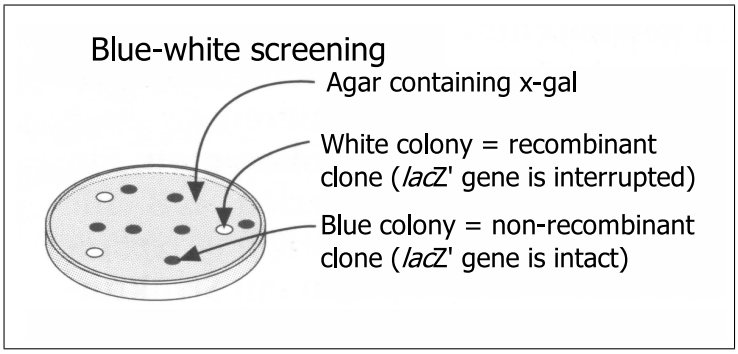


Figure 4.10: Blue-white screening

Expression vectors often add a tag to the expressed protein (e.g. His-tag). The His-tag is a tag of 6 histidines either at the N-terminal or C-terminal end of the protein without disturbing the function of the heterologously expressed protein. Proteins tagged with a His-tag can be purified by affinity chromatography on Nickel-columns. This is useful to purify the heterologously expressed protein and separate it from the endogenous proteins of the host.

4.4 cDNA library

A cDNA library is a collection of generally full-length cDNAs, which should represent the whole coding genome of an organism. For easy handling of the cDNA collection, they are transferred into bacteriophages. Bacteriophages are viruses of bacteria, in this case of *E. coli*.

The cDNA library is generated from mRNA. After reverse transcription of the mRNA to cDNA, restriction enzyme recognition sites are introduced to both ends of the cDNA by a short PCR with modifying primers. Then the cDNA is cut with the restriction enzymes and ligated into λ-phage vector arms. The phage DNA is not circular like DNA of *E. coli*, but straight. Therefore the vector in which the cDNAs are incorporated into the genome of the virus are called "arms". The λ-phage is a specific class

of *E. coli* bacteriophages. After ligation of the cDNA collection into the λ-phage vector arms they are transferred into λ-bacteriophages. The DNA of the phage (including the cloned cDNA) can either be introduced into the *E. coli* genome, amplified and excised as a plasmid (lysogen cycle) or the phage can be replicated in the *E. coli* cell until cell lysis (lytic cycle) producing so called plaques as shown in figure 4.11.

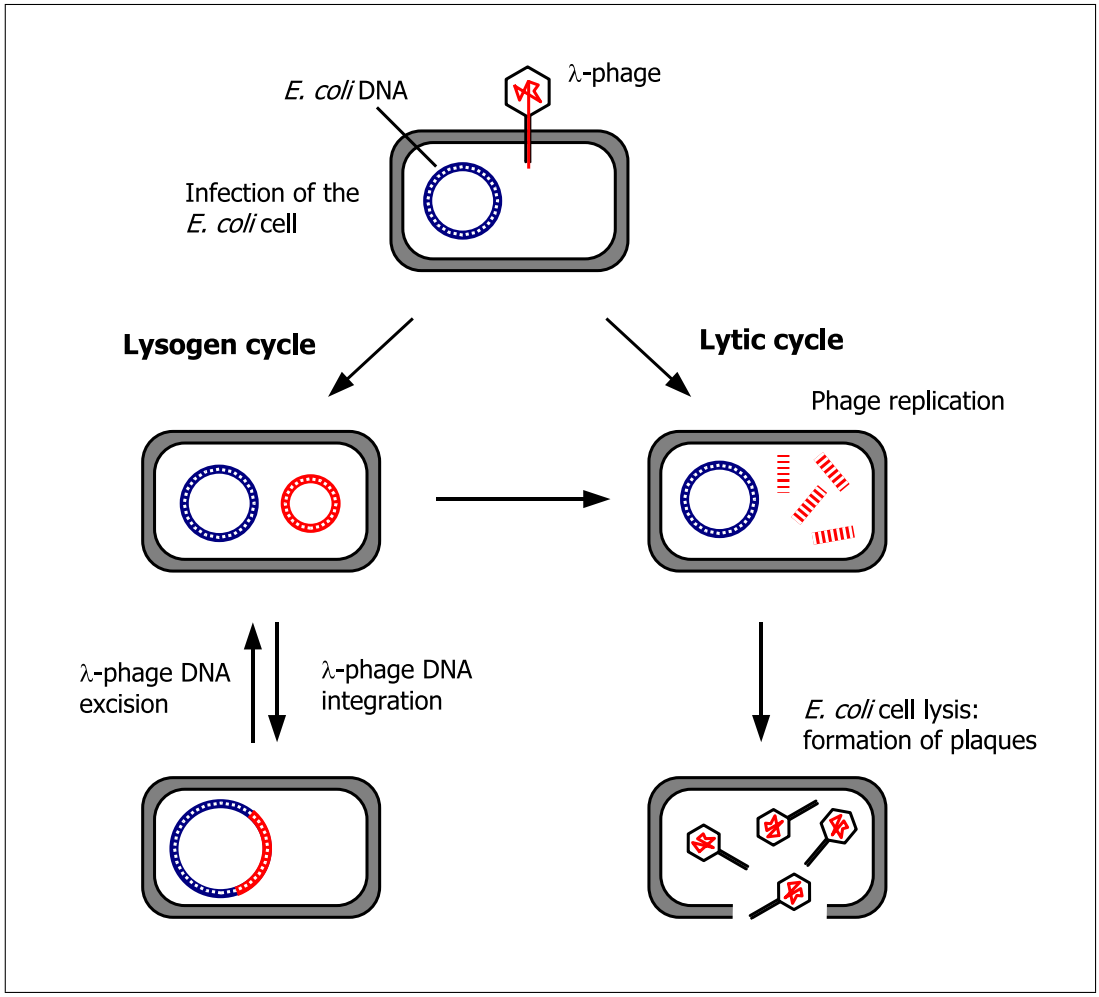


Figure 4.11: Phage cycle

For screening the cDNA library *E. coli* is infected by the phages and spread on agar plates to form plaques. Each *E. coli* cell is infected by one phage, containing one of the cloned cDNA. This results in several thousand different cloned full-length cDNAs

present as plaques on the agar plate. A print of the plaques is made on a membrane filter. The cDNA of the plaques is fixed on the filter and can be screened by hybridization with a labeled cDNA fragment (probe), generally with radioactively labeled nucleotides (e.g. [α^{32}P]dATP).

Labeling is performed by random priming. The double strand cDNA is denatured and a DNA polymerase (in this case the Klenow fragment from *E. coli* DNA polymerase I) synthesizes the labeled DNA by extension of random hexamer primers. The template strand remains unlabeled, but the enzyme can synthesize several new strands from every template, resulting in a large amount of labeled product.

Hybridization reaction is the formation of partial or complete double strand nucleic acid molecules by sequence specific interaction of two complementary single strand nucleic acids. Hybridization using labeled probes is the way to detect the presence of a specific nucleic acid sequence in a complex nucleic acid mixture. Screening the cDNA library with a cDNA fragment by hybridization under less stringent conditions leads to full-length sequences identical or homologous to the fragment. After hybridization of the probes, the membranes are washed to remove non-hybridized or mismatched hybridized probes, before detection. The stringency of the hybridization is determined by the temperature of the hybridization and the salt concentration of the washing buffer. Plaques hybridizing to the probe can be isolated from the agar plate and the introduced cDNA can be isolated (excised as plasmid) by infecting *E. coli* cells according to the lysogen cycle.

4.5 Electrophoresis and Western Blotting

Electrophoresis is a method of separation of charged substances by migration in an electrical field. As biomolecules (proteins and nucleic acids) are often charged this

principal is the simplest way to separate them by mass in e.g. polyacrylamide or agarose gel.

Polyacrylamide gels are obtained by radical polymerization of acrylamide with the crosslinking N, N'-methylenbisacrylamide. The SDS-polyacrylamide gel electrophoresis is used for the separation of proteins in the presence of sodium dodecylsulfate (SDS) according to Laemmli [Laemmli, 1970]. The proteins are heat denatured in the presence of the anionic detergent SDS, which forms around the protein anionic micelles with a constant net charge per mass. Because of the linear relationship between the logarithm of the protein mass and the migration speed of the micelle in the polyacrylamide gel, the proteins can be separated by their mass.

Detection of the protein in the gel is done by coloring with Coomassie brilliant blue R-250. This dye binds unspecifically to all proteins, but not to the gel.

Another way of detection of specific proteins is the immunodetection with antibodies. For the immunodetection, the SDS-PAGE separated proteins have to be transferred onto a nitrocellulose membrane (blotting). The protein blotting is done by electrophoretic transfer in a transfer buffer solution under an electrical field. The transfer has to be done in vertical direction to the protein separation. After the proteins are transferred onto the membrane, the membrane is blocked with milk protein to prevent unspecific binding of the antibody to the membrane. Then the membrane is incubated with the antibody (e.g. anti-His-tag-antibody). The antibody itself is conjugated to peroxidase. After the antibody has bound specifically (e.g. to protein containing a His-tag), the visualization is achieved by chemiluminescence. The conjugated peroxidase reacts with the substrate (peroxide) by producing a light-emitting product (chemiluminescence), which can be detected.

DNA is separated on agarose gels. Agarose gel can separate DNA of a broader length range than polyacrylamide gels. Agarose gels are prepared by heating the polysaccharide agarose in buffer solution until it dissolves. After pouring the hot solution into

a tray, it forms a gel upon cooling. A 1% agarose gel separates e.g. DNA of a strand length between 500 until 10,000 bp. The detection of DNA is performed by the use of ethidium bromide. It binds to the DNA double helix and is visible as ethidium-DNA-complex after excitation under UV light at 254 nm.

Chapter 5

Special Part

5.1 Isolation of the encoding genes of Sesquiterpene synthases from *Solidago canadensis*

An overview of the strategy to isolate full-length cDNAs encoding sesquiterpene synthases is shown in the flowchart 5.1. For the isolation of the encoding genes mRNA is used as starting material, isolated from *Solidago canadensis* leafs. The mRNA can be used for different working steps. It starts with a reverse transcription of the mRNA to cDNA. The cDNA is used for a PCR with degenerate primers to isolate first cDNA fragments of sesquiterpene synthase. The degenerate primers are designed on conserved regions of known plant terpene synthases.

Having first cDNA fragments of sesquiterpene synthases, the full-length cDNAs can either be obtained by RACE-PCR or by screening a cDNA library.

The RACE-PCR generates the missing 5′- and 3′-ends starting from the sequence information of the isolated fragments. The RACE-PCR is carried out on specific RACE cDNA retrieved by transcription of mRNA under assembly of adaptor sequences at the 5′- and 3′-end. After obtaining the full-length sequence information, full-length PCR

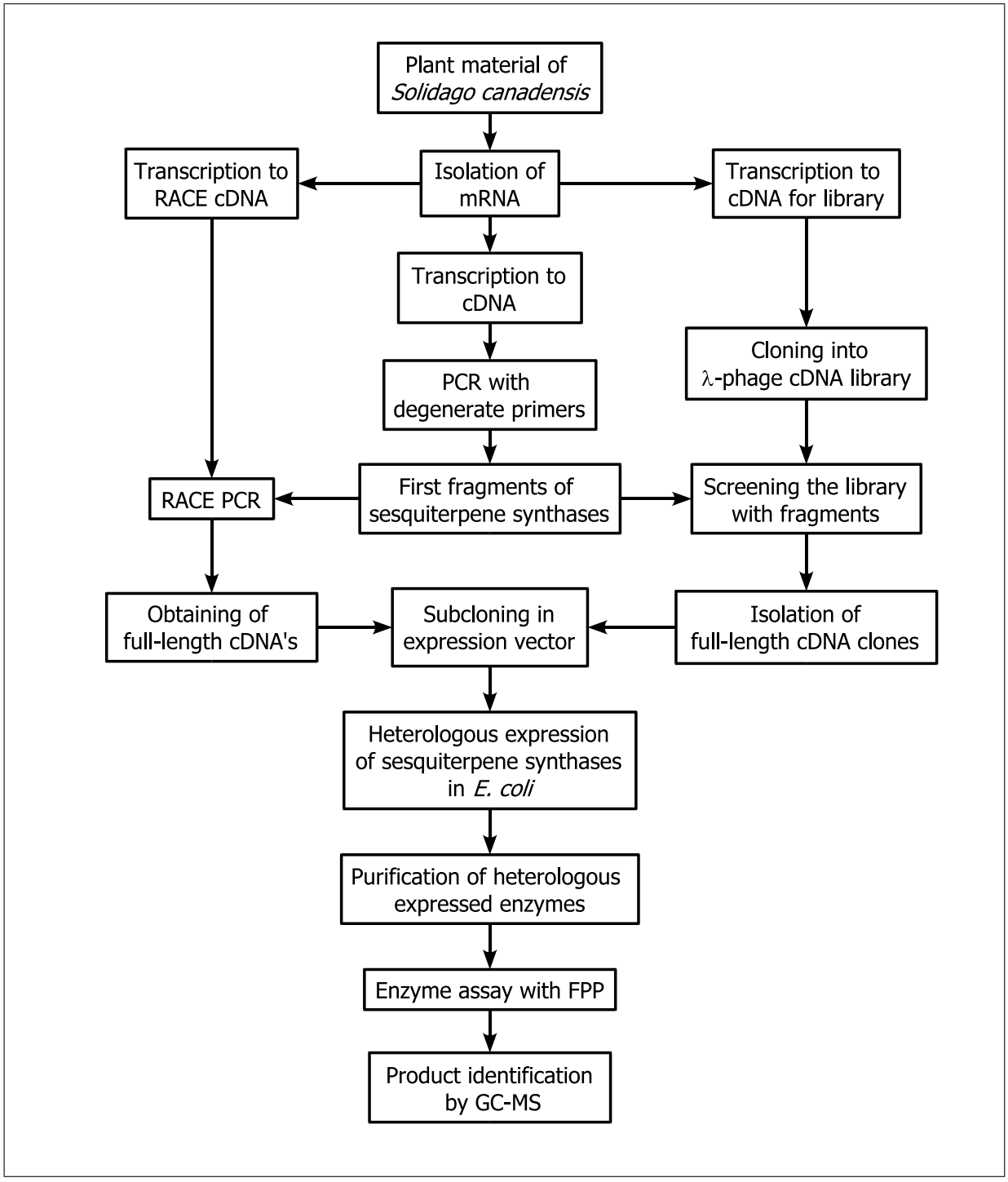


Figure 5.1: Flowchart overview of the strategy

with proof-reading enzyme has to be performed to generate full-length sequences for subcloning into expression vectors for heterologous expression.

To isolate full-length cDNAs by screening a cDNA library, the cDNA library has first to be prepared from mRNA. The mRNA is introduced into λ-phage vector arms after transcription and transferred into bacteriophages. The library is then screened for full-length sesquiterpene sequences with the fragments obtained from the degenerate primer PCR.

Isolated full-length clones either obtained from the RACE-PCR or from the cDNA library are then subcloned in expression vectors and transferred into *E. coli*. Heterologous expression of the cDNA is performed under induction with IPTG. The heterologously expressed enzymes are expressed with a Histidine-tag. Purification of the enzymes is done by affinity chromatography on columns containing Nickel, which binds to the Histidine-tag. The function of the heterologously expressed enzyme is tested in enzyme assays under addition of the sesquiterpene substrate FPP. Possibly formed products are identified by gas chromatography with mass selective detector. The absolute configuration of products is determined by gas chromatography with enantioselective stationary phase.

In this chapter the experiments are described only in a brief form with an emphasis on the results. The detailed description of the experimental procedure can be found in chapter 8.

5.2 Plant material

Different chemotypes of the plant *Solidago canadensis* L. were available. The difference of the chemotypes was determined due to the different ratio of the (+)-germacrene D and (–)-germacrene D in the essential oils of the plants.

Plant 1 had an excess of (+)-germacrene D
with a ratio of (+)-germacrene D to (–)-germacrene D of 5.5 to 1

Plant 2 had an excess of (–)-germacrene D
with a ratio of (+)-germacrene D to (–)-germacrene D of 1 to 6.1

Plant 3 had a ratio of (+)-germacrene D to (–)-germacrene D of 1 to 2

All three plant types were used for the experiments. The plants were freshly frozen in liquid nitrogen and stored at -80°C until use.

5.3 Isolation of first fragments

5.3.1 Degenerate primer design

Degenerate Primers were designed to generate first cDNA fragments from *Solidago canadensis* with high homology to sesquiterpene synthase genes. To identify conserved regions a multiple sequence alignment of deduced amino acid sequences of *S*-limonene synthase *(Mentha spicata)* [Colby et al., 1993], 5-*epi*-aristolochene synthase (*Nicotiana tabacum*) [Starks et al., 1997] and casbene synthase (*Ricinus communis*) [Mau and West, 1994] was performed. As shown in figure 5.2 two highly conserved regions framed in boxes were selected for designing the degenerate primers [Wallaart et al., 2001].

The amino acid sequence (D/N)(S/E)(D/E/N)G(K/E)FKE yielded the downstream degenerate primer `5'-GAY-GAR-AAY-GGN-AAR-TTY-AAR-GA-3'`.
The amino acid sequence DD(T/I)(I/Y/F)D(A/V)Y(A/G), containing the characteristic DDxxD motif of terpene synthases, yielded the upstream degenerate primer `5'-CC-RTA-NGC-RTC-RAA-NGT-RTC-RTC-3'`.
The letters `R` (`A` or `G`) and `Y` (`C` or `T`) designate IUB codes for the variable nucleotide sites, and `N` denotes inosine.

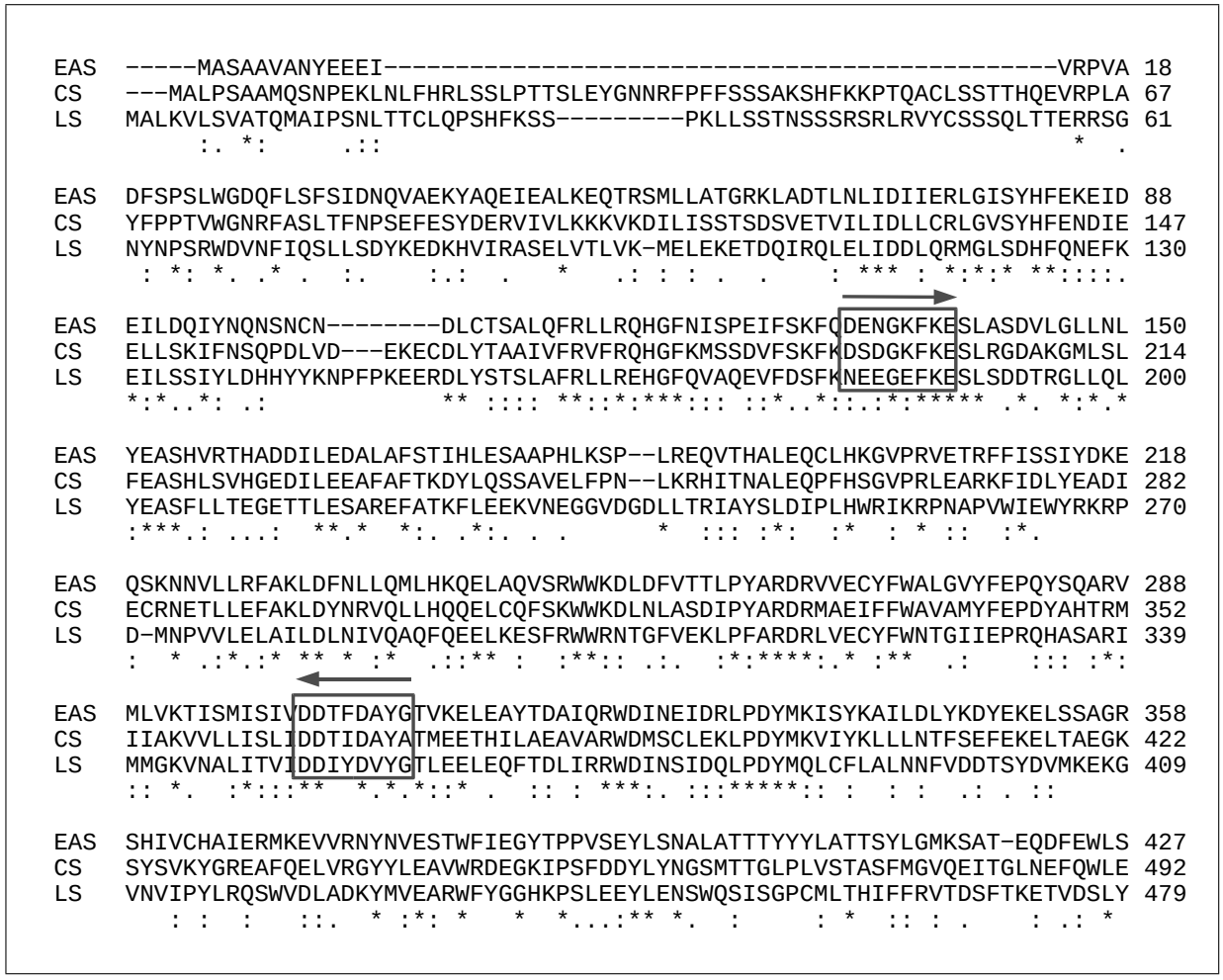


Figure 5.2: Alignment of terpene synthases for selecting highly conserved regions.

EAS = 5-*epi*-aristolochene synthase *(Nicotiana tabacum)* [Starks et al., 1997]

CS = casbene synthase *(Ricinus communis)* [Mau and West, 1994]

LS = *S*-limonene synthase *(Mentha spicata)* [Colby et al., 1993]

5.3.2 Generating and cloning of first fragments of sesquiterpene genes from *Solidago canadensis*

The first strand cDNA was synthesized from mRNA isolated from total RNA from leaf material of the different chemotypes of *Solidago canadensis*. To generate a first fragment

RT-PCR was performed with the degenerate primer pair from section 5.3.1 at a low annealing temperature of 42°C.

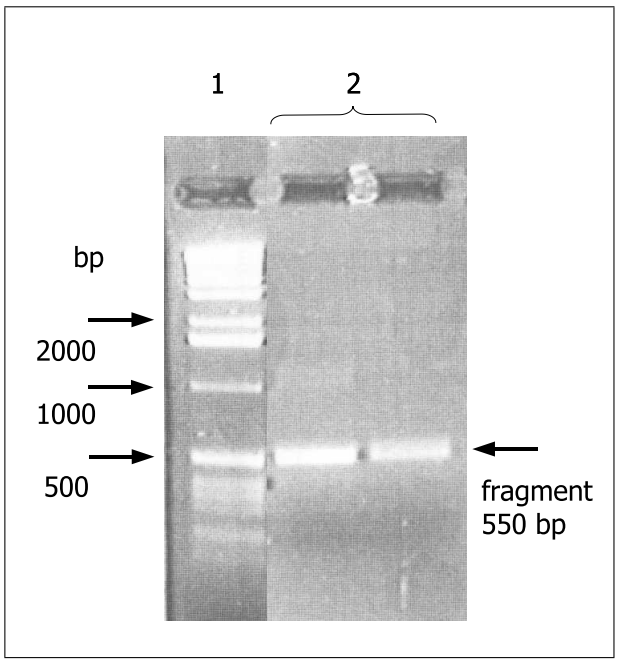


Figure 5.3: Agarose gel separation of the RT-PCR product with the degenerate primer pair. **1** = 1 kB DNA ladder, **2** = degenerate primer RT-PCR product

The resulting PCR product was separated on a 1% agarose gel and a band around the size of 550 bp (see figure 5.3) was isolated. The isolated fragment was ligated into the pGEM-T_{easy} vector and transformed into *E. coli* DH5α. After blue-white screening around 30 positive colonies were picked and the cloned fragments sequenced.

The sequencing revealed three different types of fragments shown in an alignment in figure 5.4. Fragment A is included in double, as it was present in two different splicing versions.

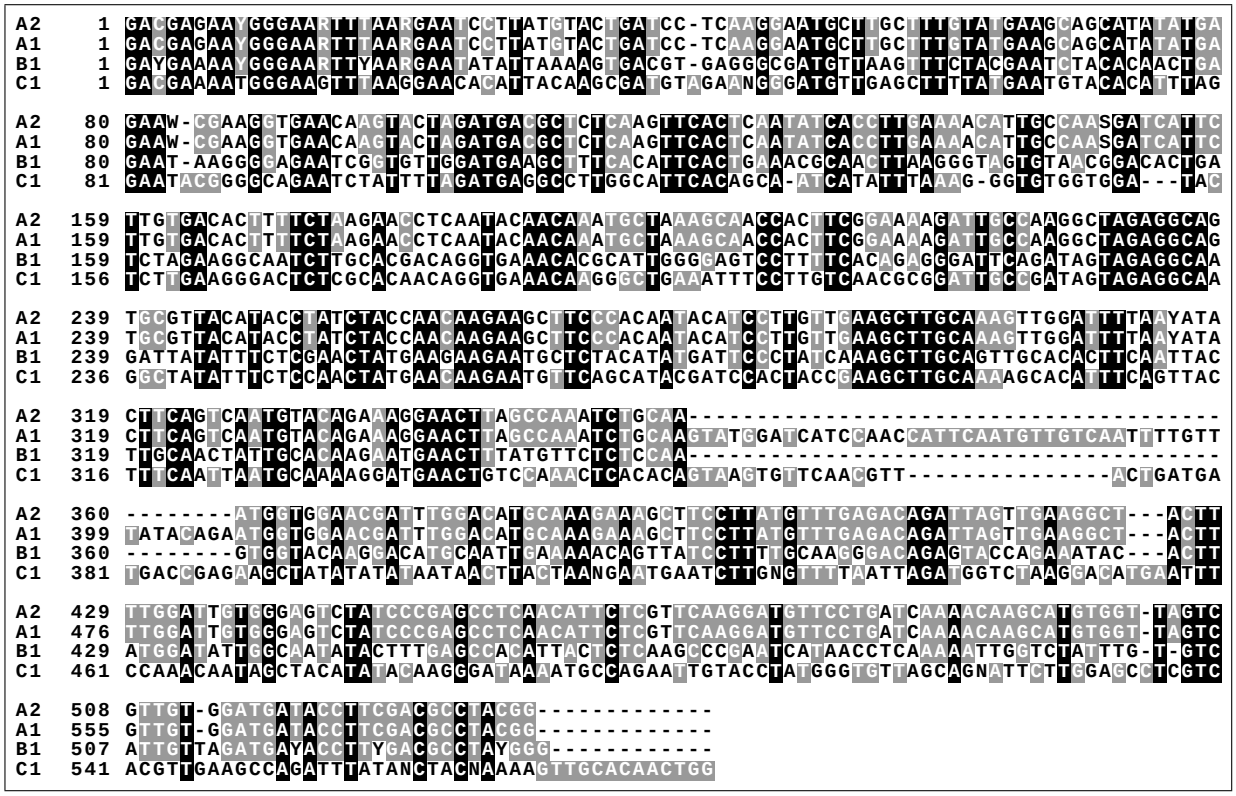


Figure 5.4: Alignment of the three fragments A, B, C including both splicing versions of fragment A.

5.4 Retrieving full-length cDNA by RACE-PCR

To retrieve full-length cDNAs in a fast way the technique of RACE-PCR was applied. The SMART RACE cDNA Amplification Kit from CLONTECH was used. This is a mechanism to generate full-length cDNAs by using the SMART technology (Switching Mechanism At 5′-end of RNA Transcript) and RACE-PCR (rapid amplification of cDNA ends).

As explained in section 4.2.3 the 5′-RACE reaction starts from mRNA. The MMLV reverse transcriptase exhibits a terminal transferase activity by adding 3-5 residues (predominantly dC) to the 3′-end of the first strand cDNA. The terminal stretch of dG residues of the SMART adaptor oligonucleotide can anneal to the dC-rich cDNA tail

and serves as an extended template for the reverse transcriptase. After the reverse transcriptase switches templates from the mRNA to the SMART adaptor oligonucleotide, a complete cDNA copy of the original mRNA is synthesized with the additional adaptor sequence at the 5′-end.

For the 3′-RACE-PCR the reverse transcription of the mRNA is carried out with a poly-dT_{25} primer possessing an adapter sequence at the 5′-end.

To try out the RACE-PCR fragments A and B from section 5.3.2 were chosen and the following gene specific primers for 5′- and 3′-RACE-PCR were designed:
For Fragment A:
downstream primer: `5′-GTG-AAC-AAG-TAC-TAG-ATG-ACG-CTC-TC-3′`
upstream primer: `5′-GGA-TGT-TGA-GGC-TCG-GGA-TAG-ACG-C-3′`
For Fragment B:
downstream primer: `5′-CTG-AGA-ATA-AGG-GGA-GAA-TCG-GTG-TTG-3′`
upstream primer: `5′-TTG-AGG-TTA-TGA-TTC-GGG-CTT-GAG-AG-3′`

Following reverse transcription, the first-strand cDNA was used for 5′- and 3′-RACE-PCR to generate the missing cDNA ends. After performing the RACE-PCR, the generated fragments were separated by gel electrophoreses and the bands of interest (marked with an arrow) were isolated. The isolated fragments were cloned into pGEM-T_{easy} vector and transformed into *E. coli* DH5α. Several positive clones were picked and sequenced.

RACE-PCR products of fragment A

Figure 5.5 shows the separated fragments after 5′-RACE-PCR and 3′-RACE-PCR of fragment A. The sequencing result revealed, that a problem occurred with the RACE reaction for fragment A. One type of RACE fragment was obtained for the 5′-end, but three different types of fragments were obtained for the 3′-end, all having an overlapping sequence part to the 5′-end RACE fragment. Furthermore the alternating splicing,

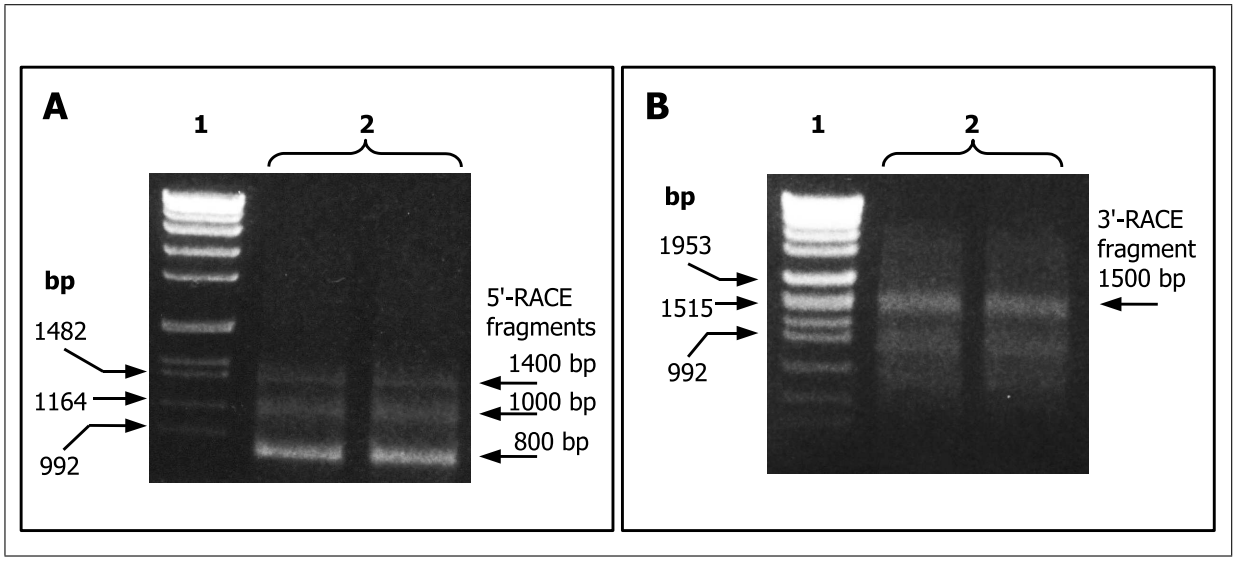


Figure 5.5: Agarose gel separated fragments after RACE-PCR of fragment A.
A = 5′-RACE-PCR product, **B** = 3′-RACE-PCR product.
1 = DNA molecular weight marker VII, **2** = RACE-PCR product of fragment A, with arrow marked bands were isolated, subcloned and sequenced.

already observed in the fragments generated with degenerate primer PCR, appeared again in the 5′-end RACE fragment as well as in the 3′-end RACE fragment. The sequencing results are shown in a schematic overview in figure 5.6.

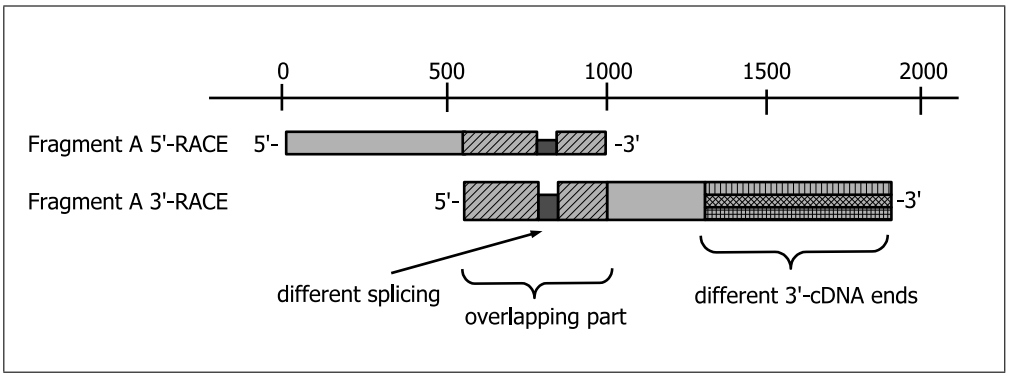


Figure 5.6: Schematic RACE-PCR results for fragment A.

A part of the 3′-RACE product of fragment A including a part of the 5′-RACE product is shown in an alignment in figure 5.7. The figure shows one sequence of each different type of the 3′-RACE product and one of the 5′-RACE product. A part of the overlapping region of the 3′-RACE sequence with the 5′-RACE sequence, including the different splicing version, is also shown in the alignment.

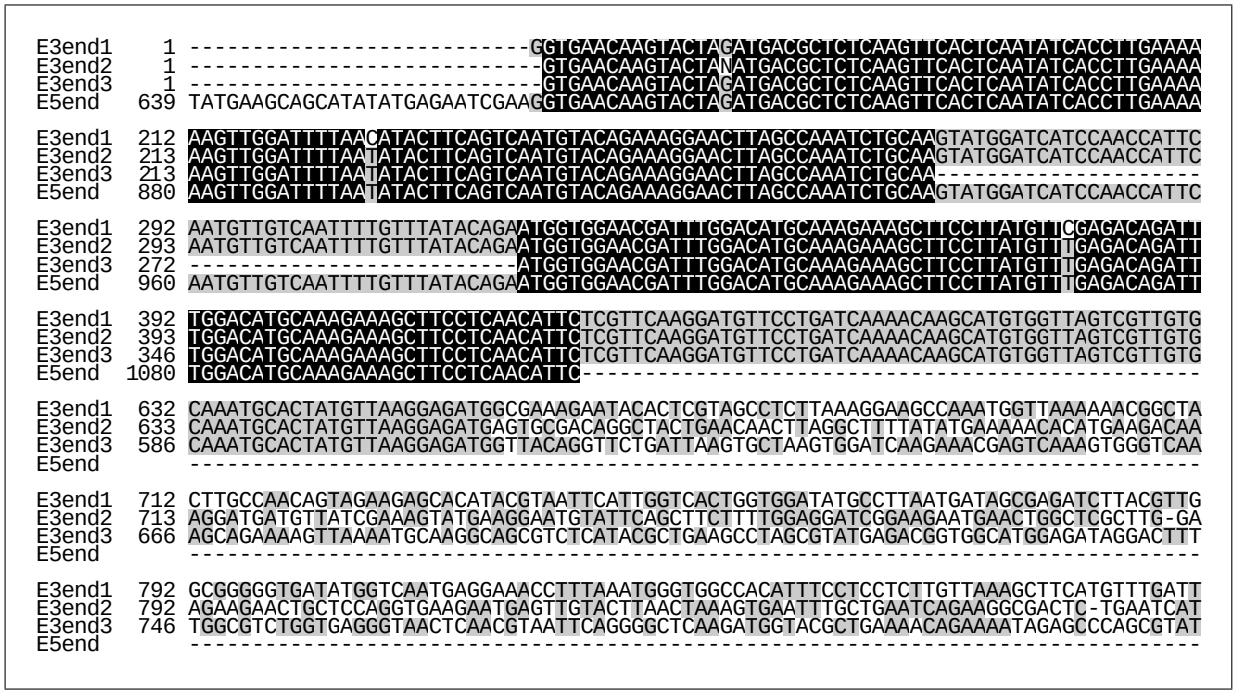


Figure 5.7: Part of the alignment of fragment A 3′-RACE product including the end of the 5′-RACE product. Shown is only one sequence of each type of the 3′-RACE sequence. One 5′-RACE sequences is included to show the overlapping region and the alternating splicing. **top block** = overlapping part of the 5′-RACE sequence (nt 639-719) with the 3′-RACE sequence (nt 1-51), **2nd block to 3rd block** = alternating splicing part: 5′-RACE sequence (nt 880-1160) and 3′-RACE sequence (nt 212-472/213-426), **4th block** = overlapping part of the 5′- RACE sequence (nt 1080-1111) and 3′-RACE sequence (nt 392-472/346-426), **5th to bottom block** = diversification of the 3′-RACE sequences into three different types (nt 632-872/586-826).

RACE-PCR product of fragment B

The RACE-PCR went well for fragment B. One type of RACE fragment was obtained for the 5′-end and the 3′-end of fragment B and an open reading frame could be detected. Figure 5.8 shows the gel separated fragments after 5′-RACE-PCR and after 3′-RACE-PCR.

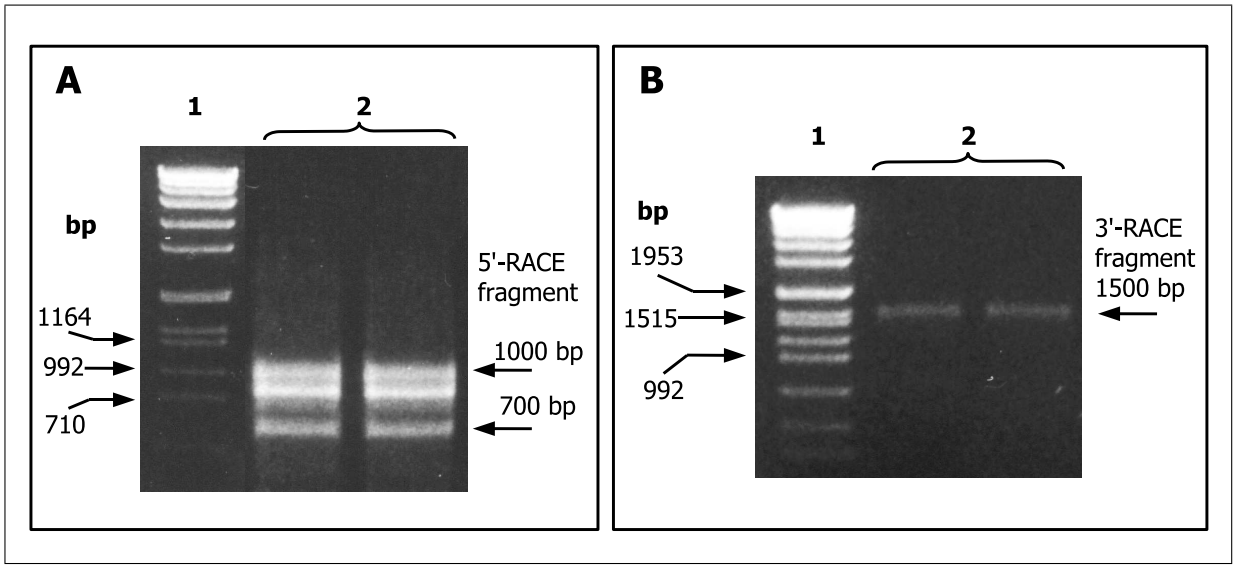


Figure 5.8: Agarose gel separated fragments after RACE-PCR of fragment B.
A = 5′-RACE-PCR product, **B** = 3′-RACE-PCR product,
1 = DNA molecular weight marker VII, **2** = RACE-PCR product of fragment B, with arrow marked bands were isolated, subcloned and sequenced.

A part of the alignment of the products for the 5′-RACE and 3′-RACE reaction is shown in figure 5.9. The alignment shows an example sequence for the 5′- and 3′-RACE reaction with the overlapping part. The complete full-length sequence including the open reading frame is shown in the appendix A.1.

```
B-5end  420 ACAACATGGTWATAAATTGTCTTGTGATGTGTTCAATAAGTTCAAGGATTGTCACTCTGGTAAATTTAAGGAATATATTA
B-3end    1 ------------------------------------------------------------------------------

B-5end  500 AAAGTGACGTGAGGGCGATGTTAAGTTTCTACGAATCTACACGACTGAGAATAAGGGGAGAATCGGTGTTGGATGAAGCT
B-3end    1 ----------------------------------------CTGAGAATAAGGGGAGAATCGGTGTTGGATGAAGCT

B-5end  580 TTCACATTCACTGAAACGCAACTTAAGGGTAGTGTAACGGACACTGATCTAGAAGGCAATCTTGCACGACAGGTGAAACA
B-3end   37 TTCACATTCACTGAAACGCAACTTAAGGGTAGTGTAGCGGACACTGATCTAGAAGGCAATCTTGCACGACAGGTGAAACA

B-5end  660 CGCATTGGGGAGTCCTTTTCACAGAGGGATTCAGATAGTAGAGGCAAGATTATATTTCTCGAACTATGAAGAAGAATGCT
B-3end  117 CGCATTGGGGAGTCCTTTTCACAGAGGGATTCAGATAGTAGAGGCAAGATTATATTTCTCGAACTATGAAGAAGAATGCT

B-5end  740 CTACATATGATTCCCTATCAAAGCTTGCAGTTGCACACTTCAATTACTTGCAACTATTGCACAAGAATGAACTTTATGTT
B-3end  197 CTACATATGATTCCCTATCAAAGCTTGCAGTTGCACACTTCAATTACTTGCAACTATTGCACAAGAATGAACTTTATGTT

B-5end  820 CTCTCCAAGTGGTACAAGGACATGCAATTGAAAAACAGTTATCCTTTTGCAAGGGACAGAGTACCAGAAATACACTTATG
B-3end  277 CTCTCCAAGTGGTACAAGGACATGCAATTGAAAAACAGTTATCCTTTTGCAAGGGACAGAGTACCAGAAATACACTTATG

B-5end  900 GATATTGGCAATATACTTTGAGCCACATTACTCTCAAGCCCGAATCATAACCTCAA--------------------------
B-3end  357 GATATTGGCAATATACTTTGAGCCACATTACTCTCAAGCCCGAATCATAACCTCAAAAATTGGTCTAT-TTGTGTCATTG

B-5end      ------------------------------------------------------------------------------
B-3end  436 TTAGATGACATTTTTGATGCATATGGTACTATTGACGAGCTTCGACTTCTAACTGATGCATTAAACAGGTGGGAAATTAG

B-5end      ------------------------------------------------------------------------------
B-3end  516 CGCTATGGAGCAACTTCCAGAATATATTAAACCATTTTACAAAATTGTCTTGTC-TATGTATACTGAACTTGAAGAACAA
```

Figure 5.9: Part of the alignment of the 5′- and 3′-RACE product of the fragment B.

5.5 cDNA library

5.5.1 Generating a cDNA library

Because of the problems in retrieving 5′-end and 3′-end sequences for fragment A by RACE-PCR, the method of screening a cDNA library to obtain full-length cDNAs was applied. Therefore a cDNA library was prepared and screened with the fragments of the degenerate primer PCR from section 5.3.1.

A λ-phage cDNA library was constructed from mRNA isolated from leaf material of *Solidago canadensis*. For the library construction the three different chemotypes of *Solidago canadensis* were used in equal amounts. mRNA was isolated directly from plant lysate with magnetic GenoPrep mRNA beads (GenoVision). The SMART cDNA Library Construction Kit (Clontech) was used to construct a λ-phage cDNA library with 1 µg of mRNA. As shown in figure 5.10 the protocol utilizes an adaptor oligonucleotide in the first strand synthesis to generate full-length cDNAs similar to the RACE-PCR in section 5.4. After double strand cDNA synthesis and introduction of restriction enzyme recognition sites of *Sfi* I by primer extension, the doublestrand cDNA is digested

with the restriction enzyme *Sfi* I, ligated into the λ-phage vector arms (provided from the SMART cDNA Library Construction Kit) and packed in λ phages.

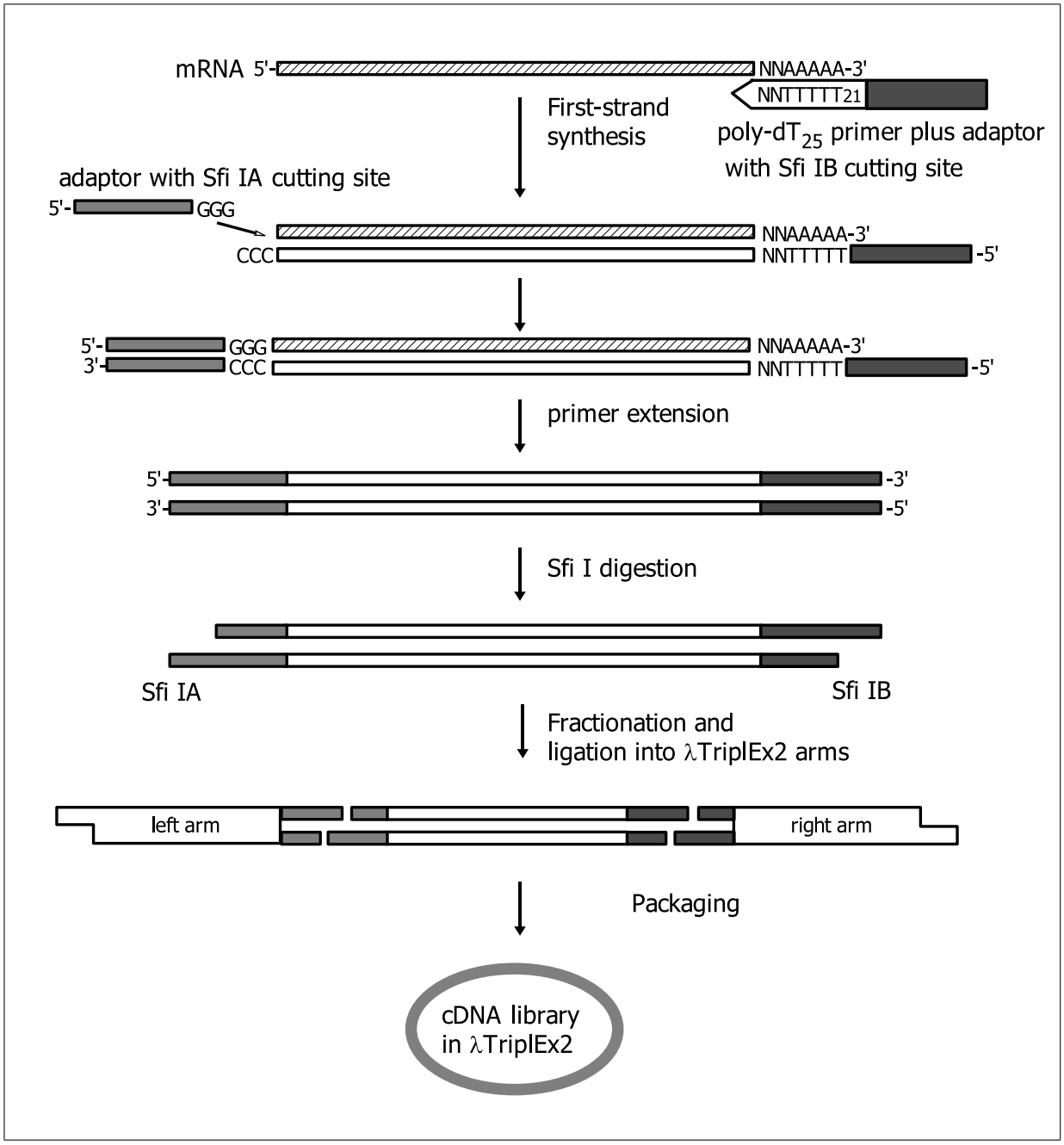


Figure 5.10: Flow chart of the SMART cDNA Library Construction Kit protocol

The library was packed into the bacteriophages with Packagene Lambda DNA Packaging System (Promega) and the packaging efficiency was checked. The constructed library had approximately $1.5 \cdot 10^6$ plaque forming units per ml (pfu/mL). The plaque forming unit means phages, which have included the vector arm and are capable of infecting *E. coli* to form a plaque. As the starting amount of mRNA to construct the cDNA library was known, it can be calculated how many phages per µg cDNA are obtained. The obtained cDNA library has $3 \cdot 10^6$ recombinant cells/µg DNA.

5.5.2 Screening the cDNA library

The cDNA library was screened with a probe labeled with [α^{32}P]dATP (Amersham Biosciences). The probe consisted of the fragment-mixture generated by PCR on cDNA from *Solidago canadensis* with the degenerate primers according section 5.3.1.

In addition a PCR fragment of 833 nucleotides was generated with specific primers to the meanwhile known sequence of germacrene A-synthase from *Solidago canadensis* (GenBank accession AJ304452). For this the following primers were designed and used in a normal temperature PCR on cDNA of *Solidago canadensis*:
the downstream primer: `5'-ATA-TCT-ACA-AGC-TTG-ATA-ATG-GAT-CT-3'` on the ORF position 374-399
the upstream primer: `5'-CAC-CAG-TTA-TGT-TTG-TAA-CTG-AGT-TA-3'` on the ORF position 1182-1207.
The generated PCR fragment was isolated and subcloned into pGEM-T_{easy} vector and the resulting construct was then transformed into *E. coli* DH5α for propagation. One clone was taken and the plasmid isolated. The isolated plasmid was used as template for PCR to generate enough of the fragment for the screening. The sequence of the generated fragment is shown in an alignment with the germacrene A sequence from the GenBank in appendix A.2. It shows, that the PCR generated fragment has minor nu-

cleotide changes compared to the GenBank published sequence. Two of the nucleotide changes even result in different amino acids.

The probe was labeled with the RadPrime DNA labeling system (Gibco-Life Technologies / Invitrogen) using the method of random priming, see section 8.6.3. Hybond-N Nylon membrane (Amersham Biosciences) replicas of plated phage virons (plaques) on *Escherichia coli* XL1-Blue were pre-hybridized for 3-5 h at 58°C in hybridization buffer and hybridized after probe addition for 24 h at 58°C. The membranes were washed for 10 min at 58°C with 2x sodium chloride-sodium citrate buffer, 0.1% SDS followed by a double washing at room temperature for 15-30 min. Hybridized plaques were identified by auto-radiography at -70°C, isolated and purified by another round of hybridization after diluted plating.

In total around 10^6 plaque forming units were screened on 15 membranes giving about 24 clones with high homology to sesquiterpene synthases. The isolated clones pair in three groups. Three of the clones pair in group 1, 12 pair in group 2 and 9 in group 3.

5.5.3 Results of the cDNA library screening

Group 1

The isolated sequences from the cDNA library, which pair in group 1 are of the same type as the fragment C from section 5.3.2. An alignment of the three isolated clones of group 1 (L2.2, L3.1 and L4.2) is shown in the appendix A.3. The alignment shows minor differences between clone L2.2 and the other two clones. It is likely that the clones of group 1 exist in two different alleles. Nevertheless the differences on the cDNA level does not cause any differences in the amino acid sequence.

A BLAST search at NCBI (http://www.ncbi.nlm.nih.gov/BLAST/) with blastx (nucleotide query against protein database) shows, that the isolated sequences of group

1 seem to be 5′-truncated. The highest BLAST hit was given to the germacrene A-synthase sequence of *Cichorium intybus* (GenBank accession AF498000). The alignment of the group 1 sequence starts at the nucleotide position 3, while the germacrene A-synthase sequence starts at the amino acid position 114. The result of the BLAST search shown below for the germacrene A-synthase sequence of *Cichorium intybus* indicates, that about 300 to 350 nucleotides of the coding sequence at the 5′-end are missing.

```
BLAST search result:

Score = 599 bits (1544), Expect = e-170
Identities = 286/445 (64%), Positives = 362/445 (81%)
Frame = +3

Gr1:  group 1 amino acid sequence,
Cich:  germacrene A-synthase sequence of Cichorium intybus

Gr1:..3....LYTTSVNFLVFRQHGYKLSCDVFNKFKDVSTGKFKEHITSDVKGMLSFYECTHLGIRGES.182
...........LYTTS+NF.VFR..G+KL.CDVFNKFKD.S+G.FKE.IT+DVKGML..YE...L.+RGE.
Cich:.114..LYTTSINFQVFRHLGHKLPCDVFNKFKDSSSGTFKESITNDVKGMLGLYESAQLRLRGEP.173

Gr1:..183..ILDEALAFTESYLKGVVDTLEGTLAQQVKQGLKFPCQRGLPIVEARLYFSNYEQECSAYD.362
...........ILDEA.AFTE+.LK.VV+TLEG.LA+QV.Q.L+.P..+G+P+VEAR+YFSNY++ECS.++
Cich:.174..ILDEASAFTETQLKSVVNTLEGNLAKQVMQSLRRPFHQGMPMVEARMYFSNYDEECSTHE.233

Gr1:..363..PLPKLAKAHFSYFQLMQKDELSKLTQWSKDMNFQTIATYTRDKMPELYLWVLAVFLEPRH.542
............LPKLAK.HF+Y.QL.QK+EL..+++W.KDM.FQ....Y.RD++PE+YLW+L.++.EPR+
Cich:.234..SLPKLAKLHFNYLQLQQKEELRIVSKWWKDMRFQETTPYIRDRVPEIYLWILGLYFEPRY.293

Gr1:..543..VEARFITTKVAQLVLVLDDTFDAYATIEELRLLTDAISRWEISCMEQLPEYIKPFYQIIL.722
............AR.I.TK+...++VLDDT+DAYATIEE+RLLTDAI+RW+IS.MEQ+PEYI+PFY+I+L
Cich:.294..SLARIIATKITLFLVVLDDTYDAYATIEEIRLLTDAINRWDISAMEQIPEYIRPFYKILL.353

Gr1:..723..NEYAEWEKQLAKEGRENVVYASKKAFQELARAYLREAEWRHSGTVPSFQEYYENGLATST.902
...........+EYAE.EKQLAKEGR...V.ASK+AFQ++AR.YL.EAEW.+SG.V.SF.EY.+NGL.TS.
Cich:.354..DEYAELEKQLAKEGRAKSVIASKEAFQDIARGYLEEAEWTNSGYVASFPEYMKNGLITSA.413

Gr1:..903..YNLLGKSCLIGMGKIVDEEALAWYDSHPKILEASELIARLHNDVVSFEFERE REHRATGI.1082
...........YN++.KS.L+GMG++V.E+ALAWY+SHPK.L+ASELI+RL.+DV++++FERER...ATG+
Cich:.414..YNVISKSALVGMGEMVGEDALAWYESHPKTLQASELISRLQDDVMTYQFERERGQSATGV.473

Gr1:..1083.DAYMKTFGVTEDVAVKELKEMIENAWKDINEGCLKPTKVSMDLLYPIVNLSRVIYVAYRF.1262
...........D+Y+KT+GVTE..A+.EL.+MIENAWKDINEGCLKP.+VSMDLL.PI+NL+R+I.V.YR+
Cich:.474..DSYIKTYGVTEKEAIDELNKMIENAWKDINEGCLKPREVSMDLLAPILNLARMIDVVYRY.533

Gr1:..1263.NDGFTFSDLTLKDYISLLFEASVPV.1337
...........+DGFTF...T+K+YI+LLF..S.P+
Cich:.534..DDGFTFPGKTMKEYITLLFVGSSPM.558
```

Group 2

The 12 sequences isolated from the cDNA library, which cluster in group 2 are not completely identical. They share very high homology, with sequence identity above 89 % on nucleotide level. Four of the isolated clones were full-length clones. The remaining 8 clones were truncated at the 5'-end. The sequences of group 2 show very high homology to the known germacrene A-synthase sequence from *Solidago canadensis* (GenBank accession AJ304452).

The sequences of the full-length clones are shown in an alignment including the germacrene A-synthase sequence from *Solidago canadensis* in figure 6.2 in chapter 6 on page 119. The sequence of each individual full-length clone including the translated open reading frame is shown in the appendix A.5.

The table 5.1 shows the percentage of identities of the four full-length clones 2a, 2b, 2c and 2d and GA (germacrene A-synthase sequence from *Solidago canadensis*) for the ORF part of the nucleotide sequence and of the amino acid sequence in brackets.

sequence	**2a**	**2b**	**2c**	**2d**
2b	92 % (85 %)			
2c	89 % (80 %)	92 % (84 %)		
2d	92 % (84 %)	94 % (89 %)	92 % (84 %)	
GA	94 % (87 %)	95 % (90 %)	92 % (85 %)	94 %(88 %)

Table 5.1: Identities of group 2 sequences on nucleotide level of the open reading frame. In brackets identities on amino acid level.

Group 3

The remaining sequences clustering into group 3 are from the same type as the fragment A from the degenerate primer PCR from section 5.3.2. None of these sequences

were full-length. Most of them are 5′-truncated. Only one seems not to be 5′-truncated, but instead 3′ truncated. The alternating splicing as well as the diversification of the 3′-end sequence appeared again as already obtained from the RACE-PCR results. Furthermore no open reading frame was detectable. As no additional information was obtained, the sequences are not shown in detail.

5.6 Full-length PCR and subcloning

5.6.1 5′-RACE-PCR on the sequence of group 1

To obtain the missing 5′-end of the group 1 sequences, 5′-RACE-PCR was performed on cDNA from the library lysate. For the 5′-RACE-PCR gene specific primers in combination with primers designed on the λ-phage vector arm sequence functioning as adaptor or anchor sequence were used. To have more PCR specificity, after the first RACE-PCR (primary RACE-PCR) a second, so called nestered RACE-PCR was performed with a second set of primers. An overview of the strategy is shown in figure 5.11.

The following gene-specific primers were designed:
For primary RACE-PCR:
upstream primer (PRIMER1): `5′-GTA-GTG-GAT-CGT-ATG-CTG-AAC-ATT-G-3′`
and downstream primer (VECTOR PRIMER1): `5′-CTC-CGA-GAT-CTG-GAC-GAG-C-3′`
For nestered RACE-PCR:
upstream primer (PRIMER2): `5′-TGT-TTC-ACC-TGT-TGT-GCG-AGA-GT-3′` and downstream primer (VECTOR PRIMER2): `5′-TTT-CTC-GGG-AAA-GCG-CC-3′`

After isolating the cDNA from the library lysate, a primary RACE-PCR was performed with PRIMER1 and VECTOR PRIMER1. A small aliquot was separated on a 1% agarose gel. Several bands were already visible, especially a band around 750 bp. To

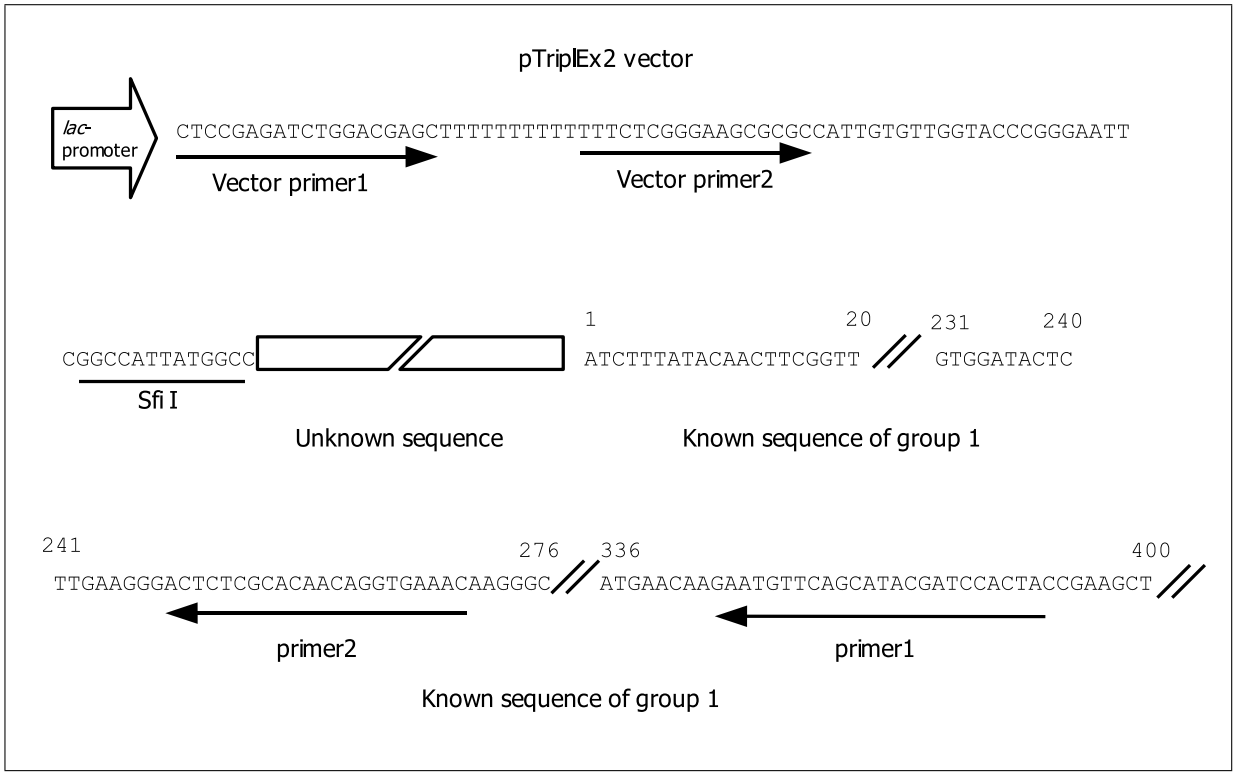


Figure 5.11: Strategy for obtaining the missing 5′-end of the sequence of group 1

get more specificity a nestered RACE-PCR was performed with the primer combination PRIMER2 and VECTOR PRIMER2. After separation on a 1% agarose gel again a band around 750 bp appeared. This band was isolated, subcloned into pGEM-T_{easy} vector, transformed into *E. coli* DH5α and sequenced. The sequencing revealed the missing 5′-end of the group 1 sequence. A picture of the agarose gel separated fragments of the primary and nestered RACE-PCR reaction is shown in figure 5.12

The RACE-PCR revealed the missing 5′-end. The result of the RACE-PCR is shown below in an alignment giving the RACE-PCR result and the overlapping part with the library sequence of group 1. The ORF is translated in the 3-letter code amino acid sequence. In total 408 additional nucleotides were acquired by RACE-PCR, of which 108 nucleotides belong to the UTR (untranslated region) and 330 nucleotides belong to

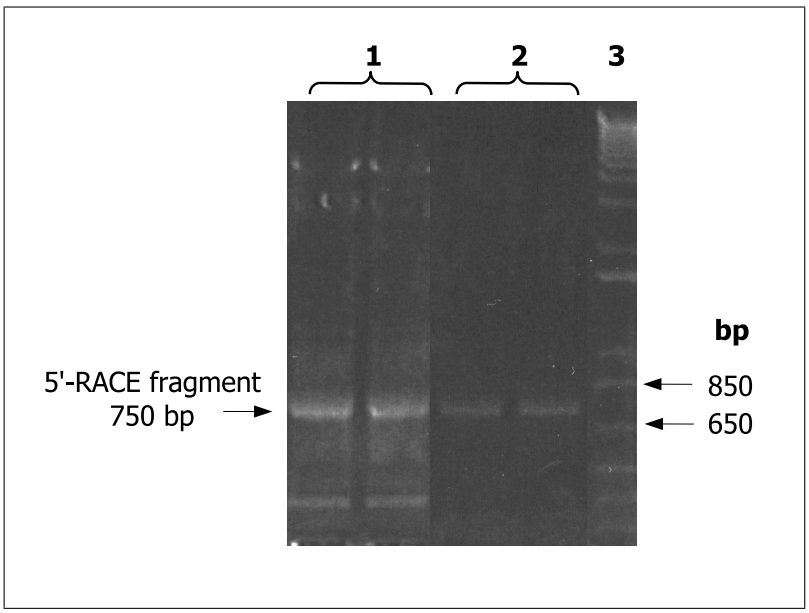


Figure 5.12: Agarose gel separated fragments of the primary and nestered 5′-RACE-PCR products of group 1.

1 = primary 5′-RACE-PCR products of group 1,

2 = nestered 5′-RACE-PCR products of group 1, **3** = 1 Kb plus DNA ladder,

bands marked with arrow were isolated

the ORF (open reading frame) of the 5′-end. The full sequence information is given in appendix A.4.

G1race = sequence found from 5′-RACE PCR of group 1,
G1lib = group 1 sequence from the library

```
G1race GTACCCGGGAAATCGGCCCATTACGGCCGGGGATCACACATTNCATTTTTAACAACTTTAATTATTAGTCT
.........................Untranslated region (UTR).....................

G1race CAAACAAACCTTTAAACACTTTAATTTGATTCAAGaATGGCCATGGTTGACGCCAATACTACTACTGTCCA
.............................MetAlaMetValAspAlaAsnThrThrThrValGlnGlnGlyGlu

G1race GCAAGGAGAGACCAAAACCACCATCGTGCGTCCTATTGCTAAGTTCCCTCCTTCGTTATGGGGTGATCGTT
.......ThrLysThrThrIleValArgProIleAlaLysPheProProSerLeuTrpGlyAspArgPheLeuSerPh

G1race TCCTATCATTCGACCTCGACAACTTGGAATTGGATGATTGCGTTGCAGCCATGAAAGAACCAAGTGATGAG
.......eAspLeuAspAsnLeuGluLeuAspAspCysValAlaAlaMetLysGluProSerAspGluIleGlnArgL

G1race ATACAAAGATTAATCGTTGATCCGGACATGGATTCAAATGAAAAACTACGTTTAATTAACTGCGTGTGTCG
.......euIleValAspProAspMetAspSerAsnGluLysLeuArgLeuIleAsnCysValCysArgLeuGlyLeu
```

```
G1race TCTTGGTTTAAGCTATCTTTTCATAAAAGATATTGAATGTCAACTCGAAAAACTTTTCAAGGAGCTTAATA
.......SerTyrLeuPheIleLysAspIleGluCysGlnLeuGluLysLeuPheLysGluLeuAsnMetGluGluTy

G1race TGGAAGAGTATAATGAATTTGATCTTTATACAACTTCGGTTAACTTTCTTGTTTTTAGACAACATGGATAT
G1lib.......................ATCTTTATACAACTTCGGTTAACTTTCTTGTTTTTAGACAACATGGATAT
.......rAsnGluPheAspLeuTyrThrThrSerValAsnPheLeuValPheArgGlnHisGlyTyrLysLeuSerC

G1race AAACTATCTTGTGATGTGTTTAACAAATTCAAGGATGTTAGCACCGGCAAATTCAAGGAACACATTACAAG
G1lib..AAACTATCTTGTGATGTGTTTAACAAATTCAAGGATGTTAGCACCGGCAAATTCAAGGAACACATTACAAG
.......ysAspValPheAsnLysPheLysAspValSerThrGlyLysPheLysGluHisIleThrSerAspValLys

G1race CGATGTGAAGGGGATGTTGAGCTTTTATGAATGTACACATTTAGGAATACGGGGCGAATCTATTTTGGATG
G1lib..CGATGTGAAGGGGATGTTGAGCTTTTATGAATGTACACATTTAGGAATACGGGGCGAATCTATTTTAGATG
.......GlyMetLeuSerPheTyrGluCysThrHisLeuGlyIleArgGlyGluSerIleLeuAspGluAlaLeuAl

G1race AGGCCTTGGCATTCACAGAATCATATTTAAAGGGTGTGGTGGATACACTTGAAGGGACTCTCGCACAAC
G1lib..AGGCCTTGGCATTCACAGAATCATATTTAAAGGGTGTGGTGGATACTCTTGAAGGGACTCTCGCACAAC....
.......aPheThrGluSerTyrLeuLysGlyValValAspThrLeuGluGlyThrLeuAlaGlnGlnValLys....
```

5.6.2 Full-length PCR of group 1

To obtain the full-length coding sequence of the group 1 for subcloning, proof-reading full-length PCR was performed. The cDNA template for the PCR was isolated from the library lysate. The PCR was performed with the proof-reading polymerase *PfuTurbo* (Stratagene). One downstream primer was designed at the beginning of the coding sequence and one upstream primer was designed at the end of the coding sequence behind the stop codon. The pRSET A vector was chosen for the heterologous expression of the sesquiterpene synthase. To be able to subclone the PCR generated fragment into the pRSET A expression vector, the downstream primer introduced the endonuclease restriction enzyme recognition site *BamH* I (5′-GGATCC-3′) in front of the starting codon ATG and the upstream primer the recognition site *Nco* I (5′-CCATGG-3′) at the position of the stop codon.

Primers for the group 1 sequence for full-length PCR:

downstream primer: 5′-TTT-GAT-GGA-TCC-ATG-TCG-ATG-GTT-GAC-3′

upstream primer: 5′-TGC-ATC-CCA-TGG-GGA-CAG-AAG-CCT-CG-3′

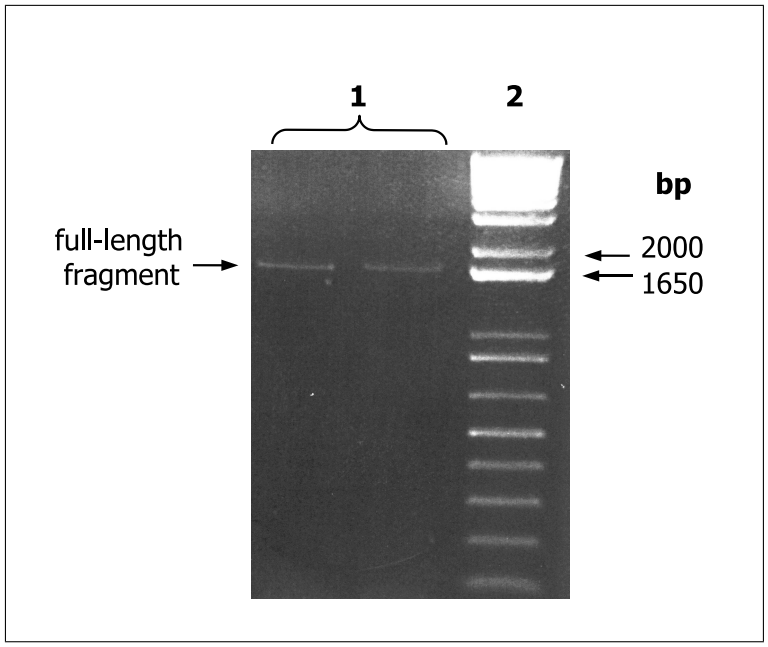


Figure 5.13: Agarose gel separated reaction after full-length PCR on group 1.
1 = full-length PCR on group 1, **2** = 1 Kb plus DNA ladder,
band marked with arrow was isolated

After the PCR reaction, the generated full-length band was separated on 1% agarose gel (see picture 5.13) and isolated. Restriction endonuclease reaction was performed with *BamH* I and *Nco* I in sequence. After the first and after the second incubation with the restriction endonuclease, the sample was each time purified with the PCR purification kit. The fragment was then ligated into the expression vector pRSET A in frame with the His-tag sequence of the vector.

The pRSET A vector is a vector suitable for high level heterologous expression of protein in *E. coli*. The vector has the feature of adding six histidines (His-tag) at the N-terminal end of the expressed protein. The His-tag is used to purify the expressed pro-

tein from endogenous proteins by using affinity chromatography at Nickel-columns: the histidine binds covalent to the nickel. A more detailed description of the vector is given in section 5.7 and 7.5.2. After the ligation, the construct was named **CSpRSET**.

5.6.3 Full-length PCR of fragment B

To obtain the full-length sequence of fragment B from section 5.4, proof-reading full-length PCR was performed. The mRNA for the RT-PCR was isolated from leaf material of *Solidago canadensis*. One downstream primer was designed at the beginning of the coding sequence and one upstream primer was designed at the end of the coding sequence behind the stop codon. To be able to subclone the PCR generated fragment into the pRSET A expression vector, the downstream primer contained the endonuclease restriction enzyme recognition site *BamH* I (5'-GGATCC-3') in front of the starting codon ATG in frame with the coding sequence and the upstream primer the site *BspLU11* I (5'-ACATGT-3'), which has the same overlapping sequence as the *Nco* I cutting site of the pRSET A vector.

Primers for full-length PCR for RACE fragment B:
downstream primer: `5'-AGG-AGA-GGA-TCC-ATG-GCC-ACT-GGT-3'`
upstream primer: `5'-GCA-GCC-ACA-TGT-GAC-TTA-TTG-AGT-AT-3'`

After the PCR reaction, the generated full-length fragment was separated on 1% agarose gel (see figure 5.14) and isolated. Restriction endonuclease reaction was performed with *BamH* I and *BspLU11* I in sequence. The fragment was then ligated in frame with the His-tag sequence into the pRSET A expression vector, which was previously cut with *BamH* I and *Nco* I. After the ligation, the construct was named **FragBpRSET**.

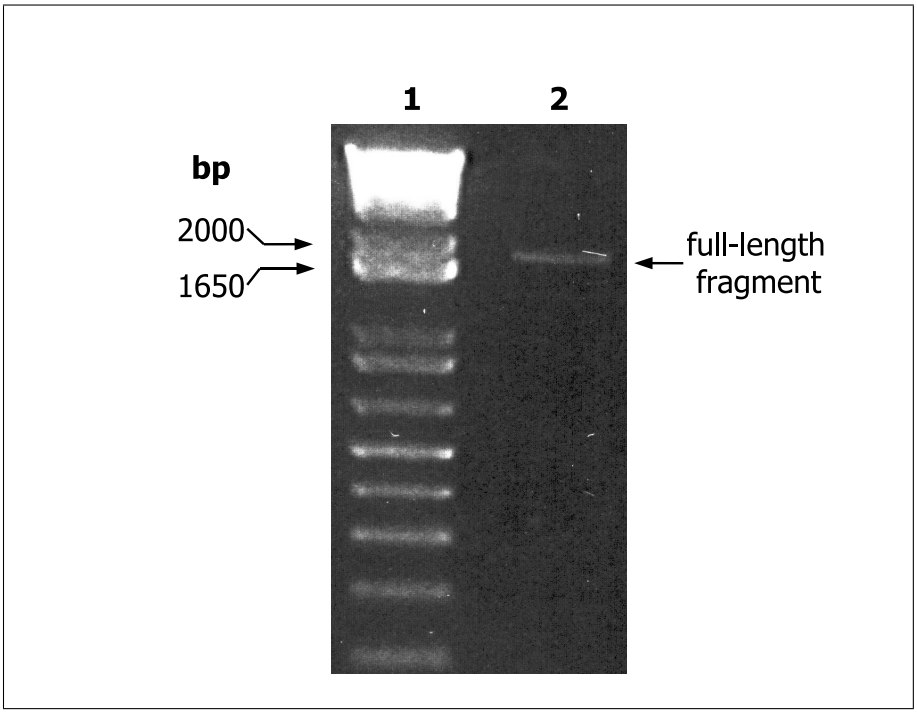


Figure 5.14: Agarose gel separated fragment after full-length PCR of fragment B.
1 = 1 Kb plus DNA ladder, **2** = full-length PCR of fragment B,
band marked with arrow was isolated

5.6.4 Subcloning of full-length library clones

The four full-length clones of group 2 (2a, 2b, 2c, and 2d), isolated by screening the cDNA library (see section 5.5.3) were subcloned into the pRSET A expression vector. For subcloning the coding sequence had to be amplified by proof-reading PCR. The isolated library clones were each used as template DNA for the PCR. To be able to subclone the PCR generated fragment into the pRSET A expression vector, the downstream primer introduced the endonuclease restriction enzyme recognition site *Xho* I (5'-CTCGAG-3') in front of the starting codon ATG and the upstream primer the site *EcoR* I (5'-GAATTC-3') downstream after the stop codon. The following primers were designed for the full-length PCR.

Primers for the group 2 sequence for full-length PCR:

downstream primer: 5′-GTC-CAT-ACT-CGA-GAT-GGC-TGC-TAA-AC-3′

upstream primer: 5′-GTC-ATG-AAG-GAC-GAA-TTC-TCA-AAC-AC-3′

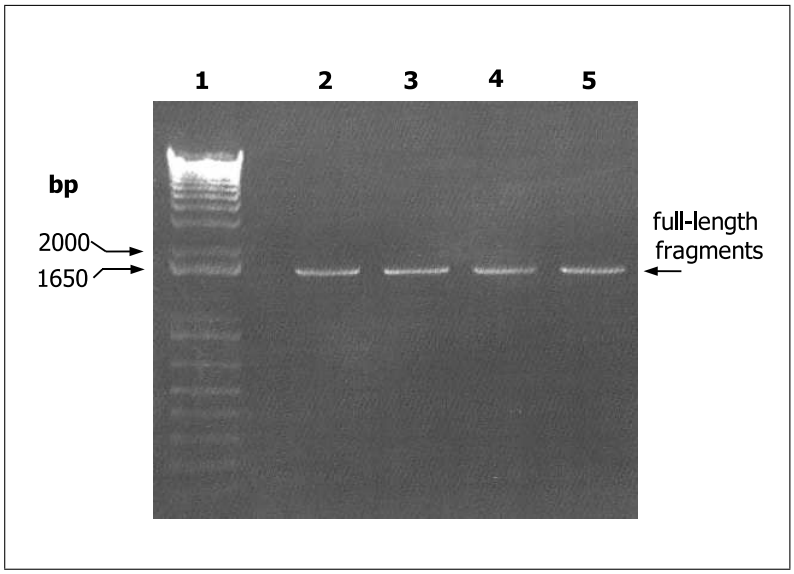


Figure 5.15: Agarose gel separated fragments after full-length PCR on group 2.
1 = 1 Kb plus DNA ladder, **2** = full-length PCR on clone 2a, **2** = full-length PCR on clone 2b,
3 = full-length PCR on clone 2c, **4** = full-length PCR on clone 2d,
bands marked with arrow were isolated.

After the PCR reaction the generated full-length fragments were separated on a 1% agarose gel (see figure 5.15) and isolated. Restriction endonuclease reaction was performed with *Xho* I and *EcoR* I in sequence. The fragments were then ligated in frame with the His-tag sequence into the pRSET A expression vector, which was also previously cut with *Xho* I and *EcoR* I. Only sequence 2c was ligated instead into the pET 32c vector again in frame with the His-tag sequence. The use of a different vector had no specific reason. After the ligation, the constructs were named as given in table 5.2.

2a	(+)GDSpRSET
2b	(–)GDSpRSET
2c	αGSpET
2d	Gr2dpRSET

Table 5.2: Construct name of group 2 sequences

5.7 Heterologous Expression

The isolated full-length cDNAs are expressed heterologously in *E. coli*. The heterologously expressed enzymes are purified and assayed with the sesquiterpene precursor farnesyl diphosphate. The formed sesquiterpenes are then identified by gas chromatography coupled with mass selective detector.

5.7.1 Micro enzyme assays with radioactively labeled substrate

For the heterologous expression, the prepared full-length constructs in the expression vector pRSET A (pET 32c vector for fragment 2c) were transformed into the *E. coli* strain BL21DE3, which is suitable for expression since it has less proteolytic enzymes. The pRSET A vector (and the pET 32c vector) adds a tag of 6 histidine units (Histidine-tag) to the N-terminal end of the expressed protein. After growing the cells until OD_{600}=0.5, the expression of the protein is induced by addition of the inductor IPTG. The cells are lysed by sonication to release the heterologously expressed protein. The non-soluble cell debris is centrifuged and discarded. The heterologously expressed protein can be purified by affinity chromatography on Nickel-columns as the Histidine-tag binds covalently to the immobilized nickel ions in the column matrix. The microassay with radioactively labeled substrate is useful to check, if the heterologously expressed enzyme is active and capable of processing the substrate to a terpene.

20 µL of the purified protein, corresponding to approx. 30-50 µg protein was used in a micro enzyme assay with tritium labeled FPP to check if the heterologously expressed enzymes show any activity. An empty vector (closed pRSET vector without any insert) was used during the experiment as negative control. Detection of any product formation was done by scintillation counting. The results are shown in the table 5.3.

	dpm (disintegration per min)
empty Vector	33
CSpRSET	439
(+)GDSpRSET	34292
(–)GDSpRSET	7616
αGSpET	2773
Gr2dpRSET	40
FragBpRSET	45

Table 5.3: Micro assay scintillation counting

The result indicates, that expressed protein from construct (+)GDSpRSET, (–)GDSpRSET and αGSpRSET show good activity. The expressed protein from construct CSpRSET shows some activity, but not very high. The expressed proteins from construct Gr2dpREST and FragBpRSET seem to be inactive at all.

5.7.2 Product identification by GC-MS measurement

To identify the formed products, a non-radioactive macro enzyme assay was performed with the expressed protein. For the macroassay 100 µL of the expressed protein (approx. 150-250 µg protein) are used. The terpenes formed in the assays were extracted and product identification was achieved by GC-MS.

Product identification of CSpRSET

The GC-MS analysis revealed, that CSpRSET construct is producing a sesquiterpene synthase, which catalyzes the formation of cascarilladiene (77.3%) as main product and δ-selinene (14.5%) as side product. It seems, that a third minor side product with 8.2% is produced, which remains unidentified. The heterologously expressed enzyme is therefore named cascarilladiene-synthase. The chromatogram is shown in figure 5.16. The mass spectrum of peak 12.98 min (cascarilladiene) is shown in figure 5.17 and the mass spectrum of peak 13.88 min (δ-selinene) is shown in figure 5.18.

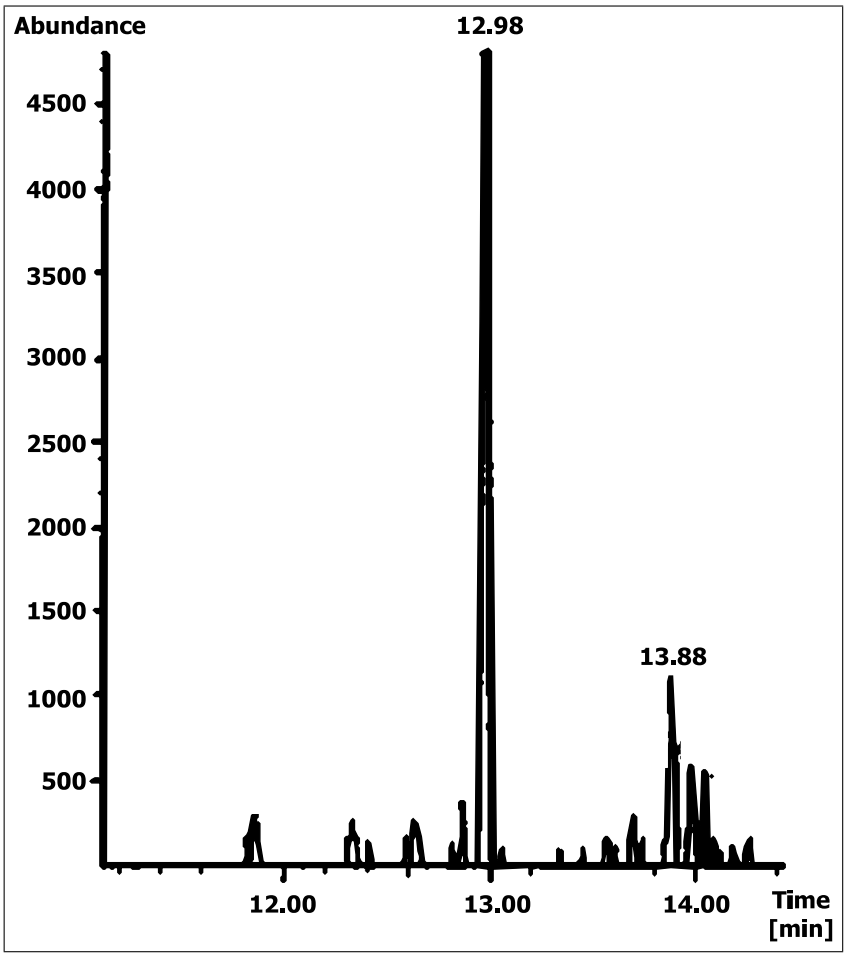


Figure 5.16: GC-MS chromatogram: single ion monitoring (ion 161) of the product formation of CSpRSET

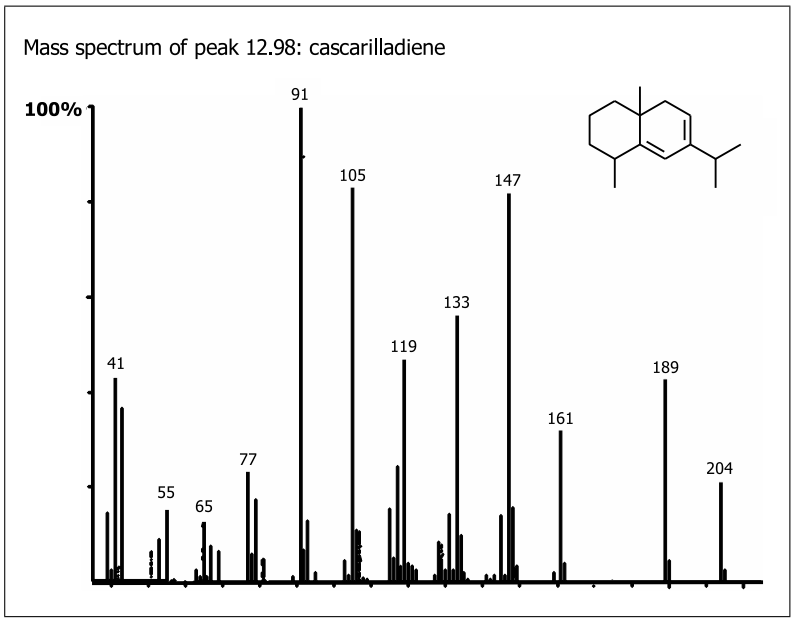


Figure 5.17: Mass spectrum of peak 12.98, identified as cascarilladiene.

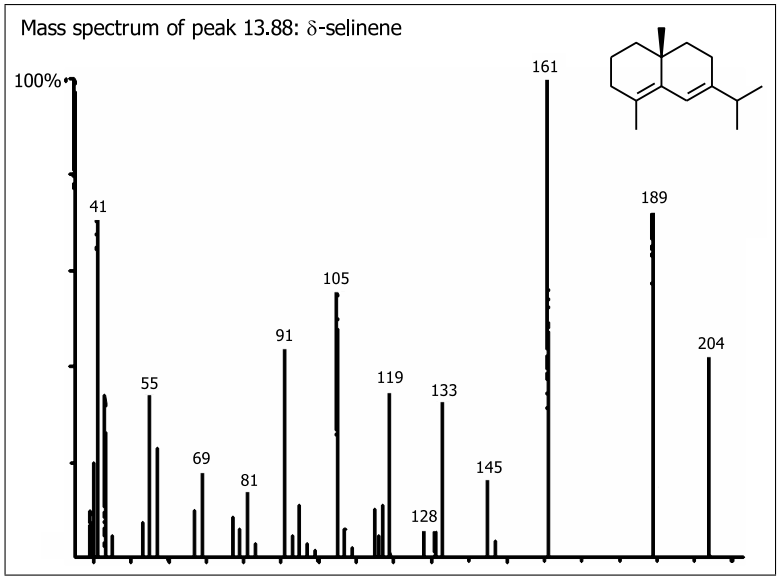


Figure 5.18: Mass spectrum of peak 13.88, identified as δ-selinene.

Product identification of (+)GDSpRSET and (–)GDSpRSET

The (+)GDSpRSET construct and the (–)GDSpRSET construct are both encoding a sesquiterpene synthase, which is catalyzing mainly the formation of germacrene D (90 %). Two minor side products of 5% remain unidentified. The chromatogram is shown in figure 5.19 and the mass spectrum of germancrene D (peak 13.80 min) is shown in figure 5.20.

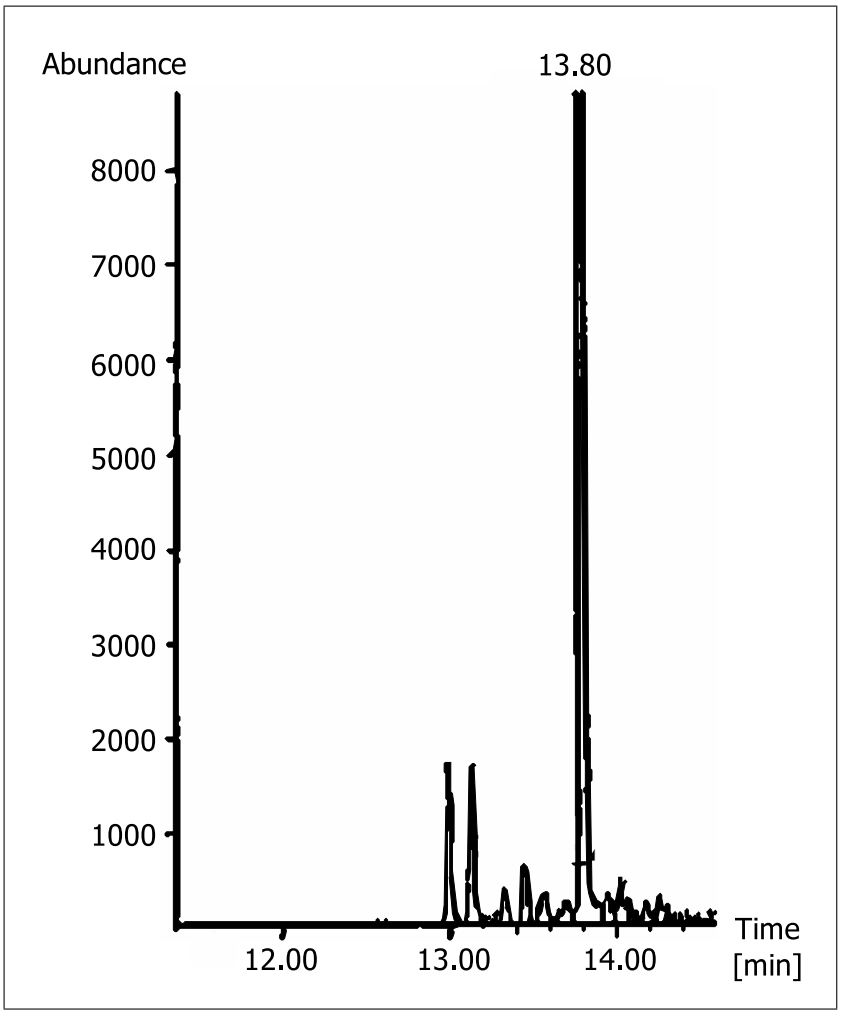


Figure 5.19: GC-MS chromatogram: single ion monitoring (ion 161) of the product formation of (+)GDSpRSET identical to (–)GDSpRSET.

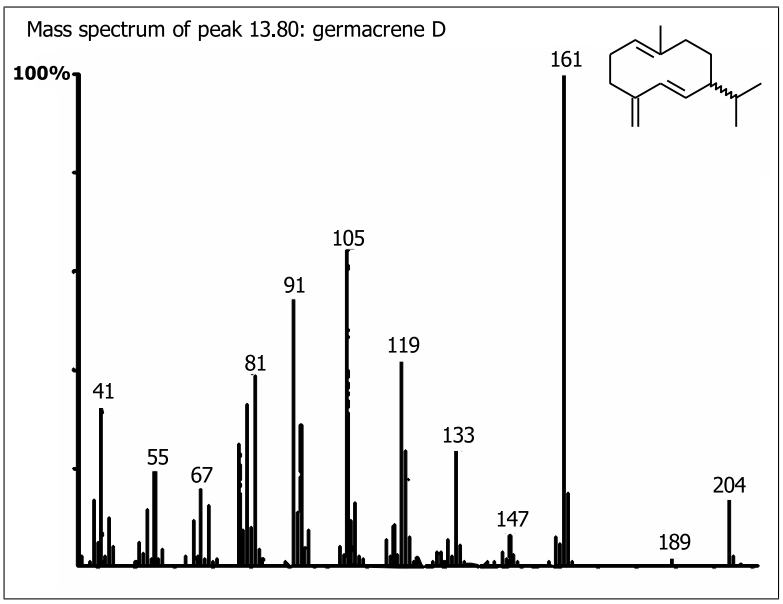


Figure 5.20: Mass spectrum of peak 13.80, identified as germacrene D.

The enantiospecificity of the two germacrene D-synthases was then tested by enantioselective gas chromatography. As enantioselective stationary phase 6T-2,3-Me-β-cyclodextrin was used. The analysis revealed that the (+)GDSpRSET construct expressed a germacrene D-synthase, which is producing exclusively (+)-germacrene D while the (–)GDSpRSET construct expressed a germacrene D-synthase, which is forming exclusively (–)-germacrene D. Therefore the (+)GDSpRSET expressed enzyme was named (+)-germacrene D-synthase and the (–)GDSpRSET expressed enzyme was named (–)-germacrene D-synthase. The chromatogram of the enantioselective gas chromatography is shown in figure 5.21.

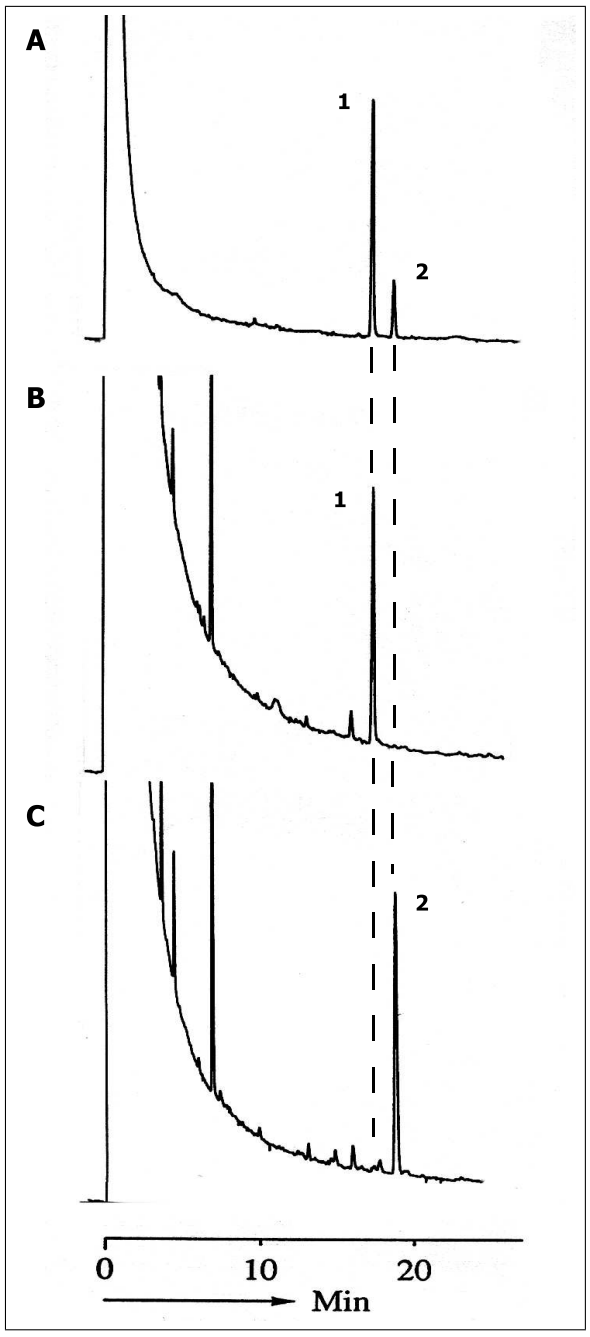


Figure 5.21: Chromatograms of the enantioselective gas chromatography

A: reference containing (+)-germacrene D (**1**) and (–)-germacrene D (**2**),

B: chromatogram of (+)GDSpRSET producing exclusively (+)-germacrene D (**1**) and

C: chromatogram of (–)GDSpRSET producing exclusively (–)-germacrene D (**2**).

Product identification of αGSpET

The αGSpET construct was expressing a sesquiterpene synthase, which was forming mainly α-gurjunene (81.3%), next to γ-gurjunene (8.8%) and bicyclogermacrene (7.1%). A fourth unidentified product was produced in traces (3%). The chromatogram is shown in figure 5.22. The mass spectrum of α-gurjunene (peak 12.89 min) is shown in figure 5.23. The mass spectrum of peak 13.68 min, identified as γ-gurjunene is shown in figure 5.24 and the mass spectrum of peak 13.99 min, identified as bicyclogermacrene is shown in figure 5.25.

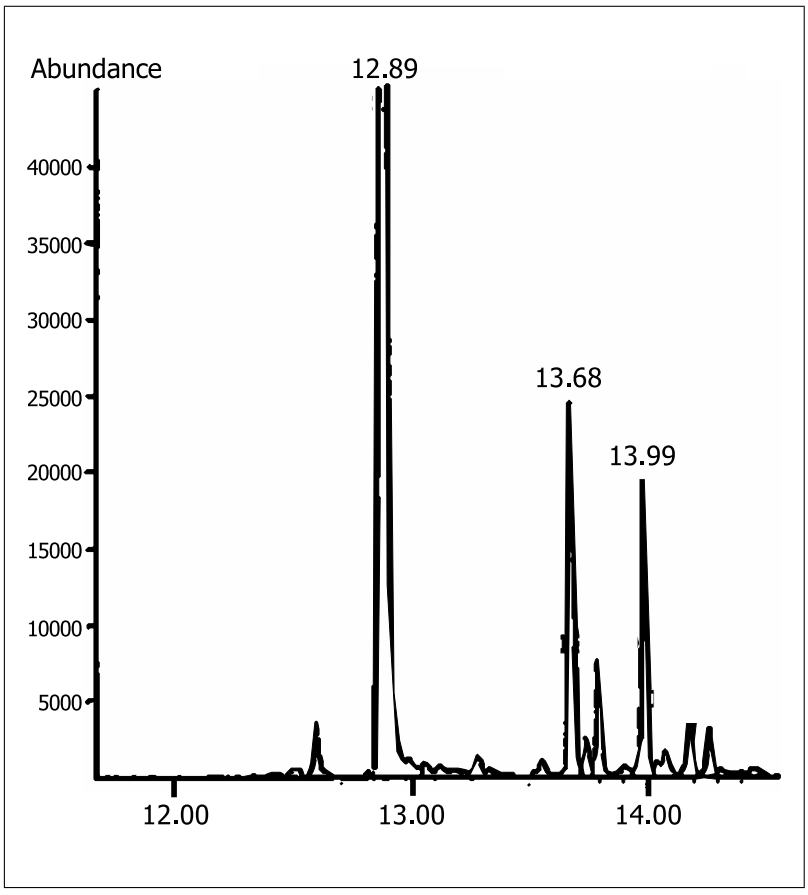


Figure 5.22: GC-MS chromatogram: single ion monitoring (ion 161) of the product formation of αGSpET.

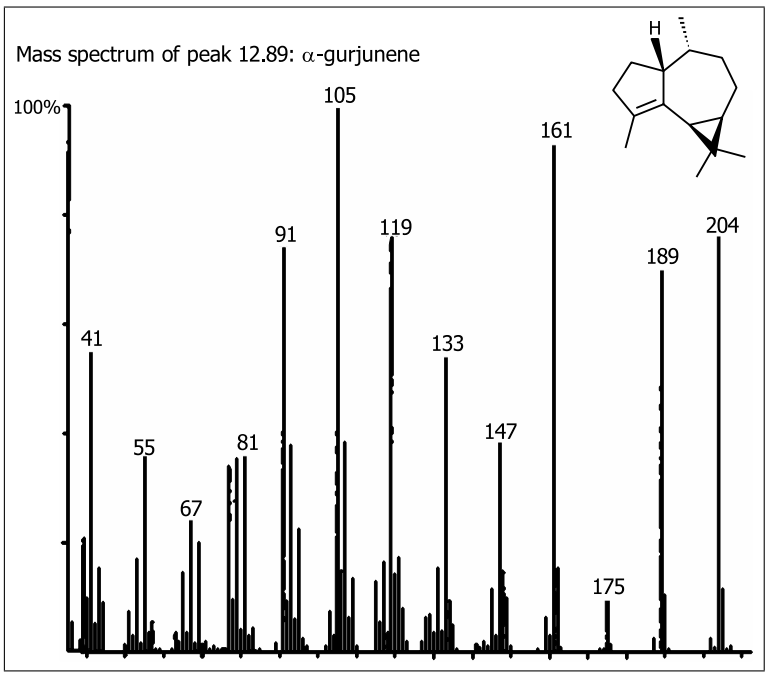


Figure 5.23: Mass spectrum of peak 12.89 from αGSpET, identified as α-gurjunene.

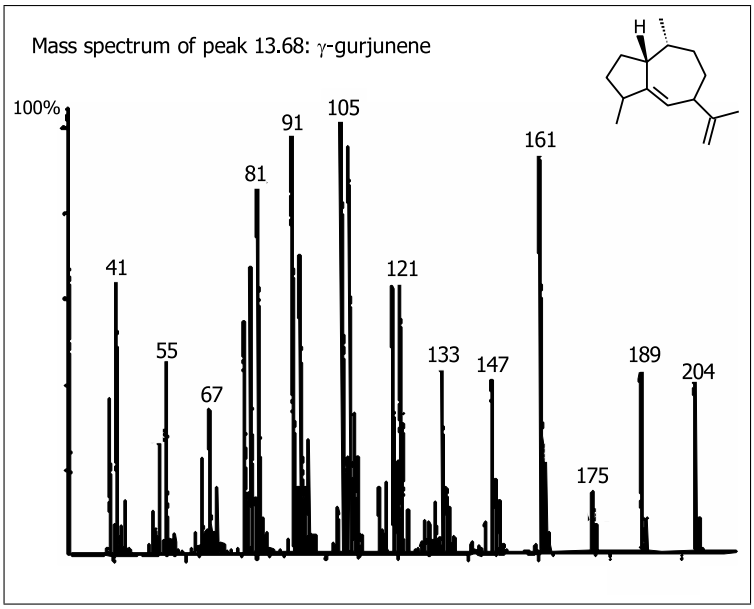


Figure 5.24: Mass spectrum of peak 13.68 from αGSpET, identified as γ-gurjunene.

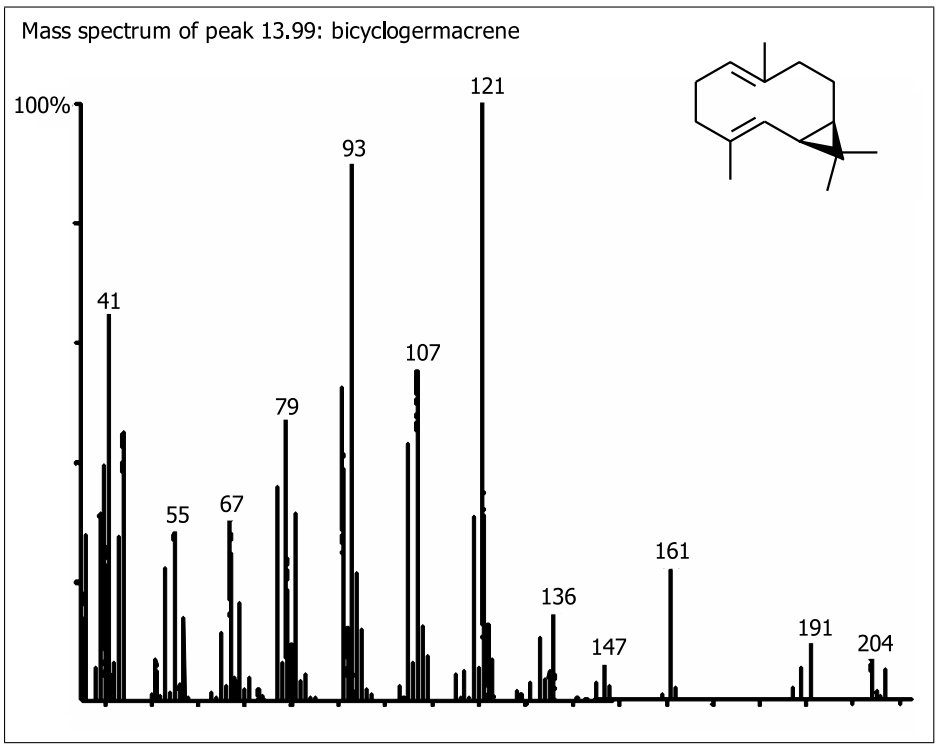


Figure 5.25: Mass spectrum of peak 13.99 from αGSpET, identified as bicyclogermacrene.

Remaining constructs

The remaining two constructs (FragBpRSET and Gr2dpRSET) didn't express any active sesquiterpene synthases. No formation of product was detectable by GC-MS.

5.8 Characterization of the expressed enzymes

5.8.1 SDS-PAGE and Western Blotting

The theoretical mass calculated on the isolated nucleotide sequence of the functional sesquiterpene enzymes is shown in the table 5.4. Calculation was carried out with the Lasergene Software package from DNAstar:

enzyme	theoretical size in kDa
(+)GDS	64.78
(–)GDS	65.04
αGS	64,97
CS	64.65

Table 5.4: Theoretical mass of the enzymes:
(+)GDS = (+)-germacrene D-synthase
(–)GDS = (–)-germacrene D-synthase
αGS = α-gurjunene-synthase
CS = cascarilladiene-synthase

To confirm the size of the heterologously expressed enzymes and to check the efficiency of the His-tag purification a SDS-PAGE was performed. On the gel the following fraction of the His-tag purification steps of each of the heterologously expressed enzymes were loaded:
– Non purified protein
– the centrifuged cell debris after cell lysis
– the His-tag purified enzyme eluate
– the flow through of the nickel column
– the first wash of the nickel column and
– the second wash of the nickel column

Besides the SDS-PAGE, a Westernblot was performed. The gel was developed in double, one was visualized by Coomassie-blue staining, the other one was used for Westernblotting. For the Westernblotting the separated proteins were transferred onto a nitrocellulose membrane and detection was performed by immunoblotting against the His-tag with the monoclonal anti-polyHistidine-peroxidase conjugate antibody. The figures 5.26 and 5.27 show the SDS gel and the Westernblot for the four sesquiterpene synthases.

Figure 5.26 shows the different purification steps of the His-tag purification of the α-gurjunene-synthase and the (+)-germacrene D-synthase. Only the His-tag purified enzymes, lane 5 and 12 on the SDS-gel and on the Westernblot, give a single protein band. During the His-tag purification some heterologously expressed enzyme gets lost, this is especially visible on the Westernblot lane 4 and lane 10 and 11. Apparently the binding conditions for the His-tag protein to the Nickel-column are not optimal. A lot of the heterologously expressed enzyme is also present in an insoluble form in the cell debris (lane 1 and 8 of the Westernblot) probably due to wrong folding. The size of the (+)-germacrene D-synthase is approximately around 65 kDa (lane 12), while the size of the α-gurjunene-synthase lies around 75 kDa (lane 5). The heterologously expressed α-gurjunene-synthase is larger than the other sesquiterpene synthases, because the enzyme was heterologously expressed in the pET32c expression vector, which adds in addition to the His-tag a thioredoxin tag. The thioredoxin tag adds about 10 kDa extra to the expected protein size. Therefore the protein size determined by SDS-gel electrophoresis corresponds to the theoretical size of 65 kDa + 10 kDa thioredoxin tag.

Figure 5.27 shows the different purification steps of the His-tag purification of the cascarilladiene-synthase and the (–)-germacrene D-synthase. The gel has a slope to the top right corner. Therefore the band of the His-tag purified (–)-germacrene D-synthase appears to be much higher than 65 kDa. This holds also for the Westernblot. By correcting for the slope the actual size of the (–)-germacrene D-synthase is determined around 65 kDa as well as for the cascarilladiene-synthase.

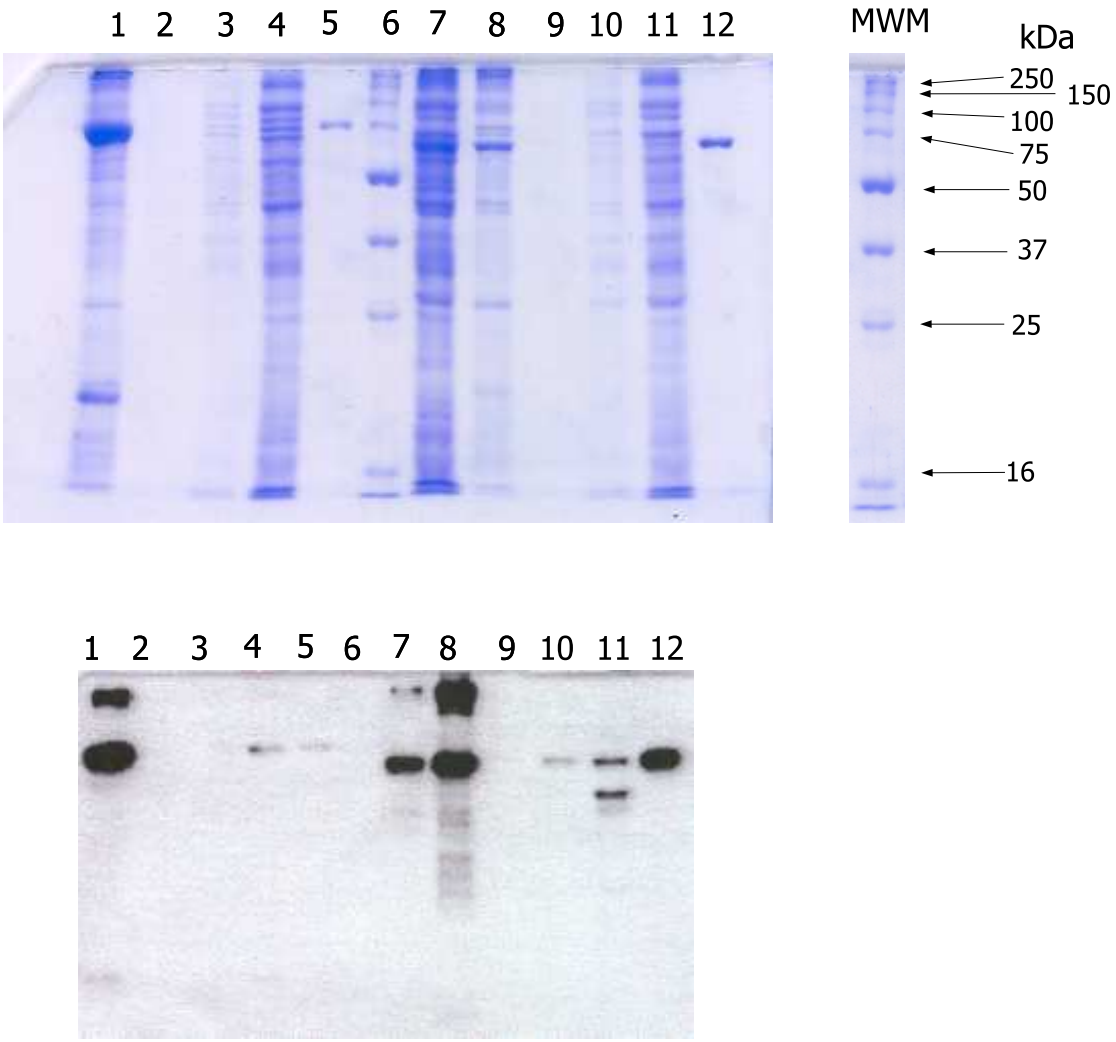


Figure 5.26: **Top** = SDS gel: coomassie stained, **Bottom** = Westernblot.

αGS = α-gurjunene-synthase, (+)GDS = (+)-germacrene D-synthase.

1 = cell debris αGS, **2** = second wash αGS, **3** = first wash αGS, **4** = flow through αGS, **5** = His-tag purified αGS.

6 = precision protein standard.

7 = non purified *E. coli* lysate (+)GDS, **8** = cell debris (+)GDS, **9** = second wash (+)GDS, **10** = first wash (+)GDS, **11** = flow through (+)GDS, **12** = His-tag purified (+)GDS.

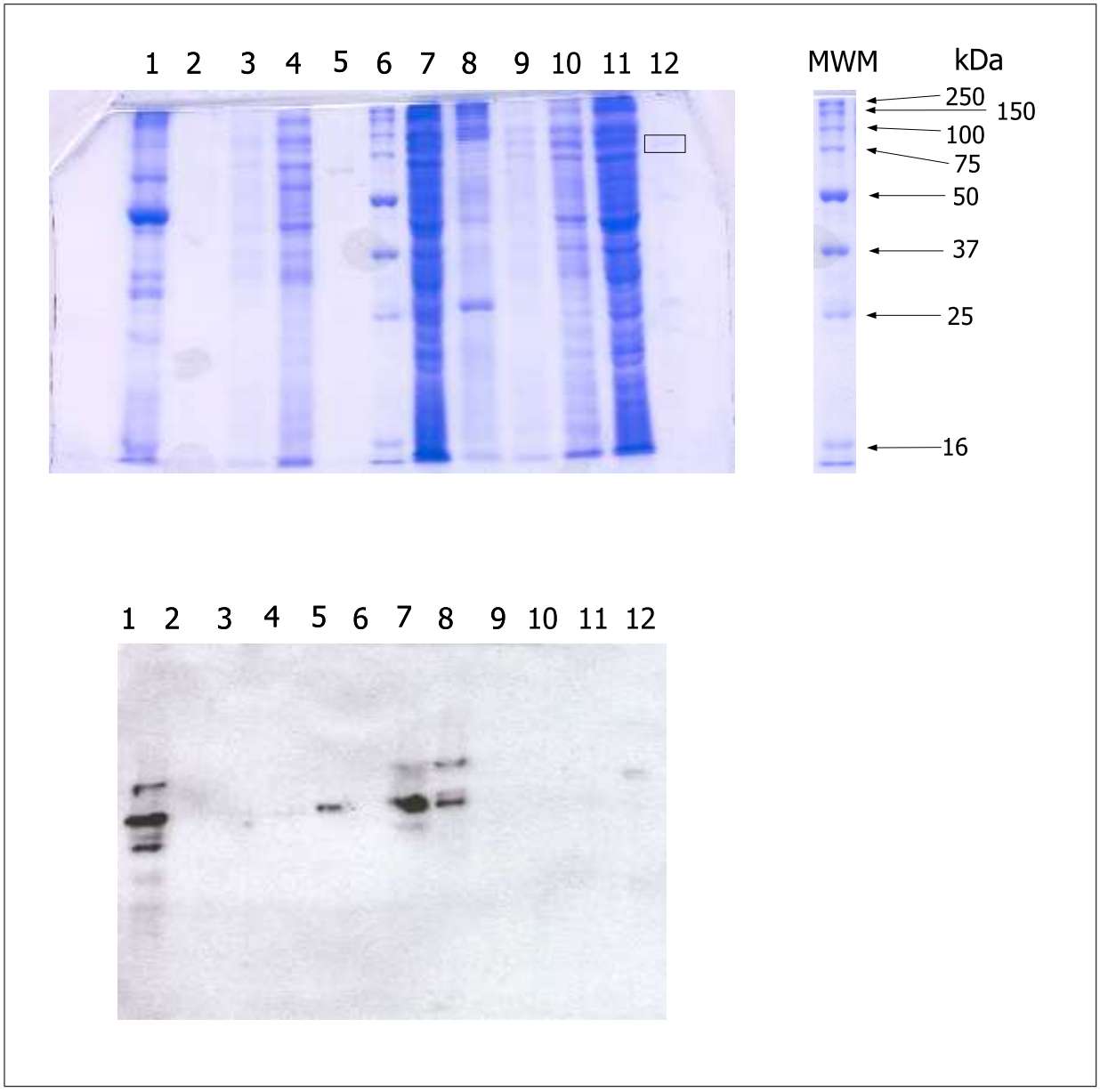


Figure 5.27: **Top** = SDS gel: coomassie stained, **Bottom** = Westernblot.

Faint band of lane 12 was boxed for better visibility.

CS = cascarilladiene-synthase, (–)GDS = (–)-germacrene D-synthase.

1 = cell debris CS, **2** = second wash CS, **3** = first wash CS, **4** = flow through CS, **5** = His-tag purified CS.

6 = precision protein standard.

7 = non purified *E. coli* lysate (–)GDS, **8** = cell debris (–)GDS, **9** = second wash (–)GDS, **10** = first wash (–)GDS, **11** = flow through (–)GDS, **12** = His-tag purified (–)GDS.

5.9 Enzyme kinetics

An enzyme reaction can be described as enzyme E and substrate S form an enzyme-substrate-complex X with the reaction velocity constant k_{+1}, which can either react to product P and enzyme E with the reaction velocity constant k_{+2} or dissociate without reaction to substrate S and enzyme E with the reaction velocity constant k_{-1}. The enzyme kinetics can therefore be described as the following [Kleber et al., 1997]:

$$S + E \underset{k_{-1}}{\overset{k_{+1}}{\rightleftharpoons}} X \overset{k_{+2}}{\longrightarrow} P + E$$

with:

$v =$ reaction velocity

$S =$ substrate

$E =$ enzyme

$X =$ enzyme-substrate-complex

$P =$ product

$k_{+1}, k_{-1}, k_{+2} =$ reaction velocity constants

As this reaction mechanism is quite complicated for further calculation it is simplified by some assumptions:
The first assumption is the steady-state-status. In the steady-state status the concentration of the enzyme-substrate complex [X] is assumed to be constant. This is fulfilled, when the substrate concentration is much higher than the enzyme concentration ($[X] = \text{const}$, when $[S] \gg [E]$). The reaction velocity can then be described as the following:

$$v = \frac{d[P]}{dt} = \frac{k_{+2} \cdot [E] \cdot [S]}{(\frac{k_{-1}+k_{+2}}{k_{+1}}) + [S]}$$

with:

$\frac{d[P]}{dt}$ = change of the concentration of product P in dependence of time

$[P] =$ concentration of product P

$[S] =$ concentration of substrate S

Every enzyme reaction will leave at a certain time point the steady-state status when the reaction proceeds long enough. Therefore to ensure the steady-state status some additional conditions have to be fulfilled. It has to be ensured that product formation is linear to the reaction time. Moreover the reaction velocity has to be independent from the substrate concentration but proportional to the enzyme concentration.

Another important assumption is, that the initial velocity at very high substrate concentration is only dependent on the enzyme concentration. Under this condition the maximum velocity (v_{max}) is reached:

$$v_{max} = k_{+2} \cdot [E]$$

This results in the Michaelis-Menten equation:

$$v = \frac{v_{max} \cdot [S]}{K_M + [S]}$$

with $K_M = \left(\frac{k_{-1}+k_{+2}}{k_{+1}}\right)$

The Michaelis-Menten constant K_M can be interpreted as the substrate concentration [S] at half maximum velocity ($\frac{v_{max}}{2}$):

with $v = \frac{v_{max}}{2}$ follows $K_M = [S]$:

$$\frac{v_{max}}{2} = \frac{v_{max} \cdot [S]}{K_M + [S]} \Leftrightarrow K_M + [S] = \frac{2 \cdot v_{max} \cdot [S]}{v_{max}} \Leftrightarrow K_M = [S]$$

5.9.1 pH-Optima

Before performing the kinetic studies, the pH optimum of each enzyme has to be determined to ensure, that the measurements are performed at the optimal pH. The measurements were done in a phosphate buffer with a pH-range between pH 6.5 and pH 8.0. Micro assays were performed and the amount of product formation was determined by scintillation counting after 30 min of incubation time with tritiated farnesyl diphosphate. In figure 5.28 the mean activities for the different pH-values are shown. Each measurement was done in duplicate. The single values of the measurements are given in appendix B.1.

The activity in dependence of the pH follows a gaussian curve. The gaussian curve is described by the following formula:

$$y = c + b \cdot e^{\left[-0.5 \cdot \left(\frac{x-a}{\sigma}\right)^2\right]}$$

with:
y = activity [dpm]
x = pH-value
c, b, a and σ = constants
$a = y_{max}$ = pH-optimum
turning point = $a \pm \sigma$
$c = \lim_{x\to\infty} f(x) = \lim_{x\to 0} f(x)$

The variable y corresponds to the enzyme activity measured in [dpm] and the variable x corresponds to the pH-value. The constant a is the pH optimum corresponding to pH value with highest enzyme activity ($a = y_{max}$). The pH values for half maximum enzyme activity are the turning points of the gaussian curve given by the formula $\frac{pH}{2} = a \pm \sigma$. And the constant c is the asymptote for $\lim_{x\to\infty} f(x)$ and $\lim_{x\to 0} f(x)$. The constants of the gaussian function were calculated by the slidewriter software for each of the enzymes. Table 5.5 is giving the values for each sesquiterpene synthase.

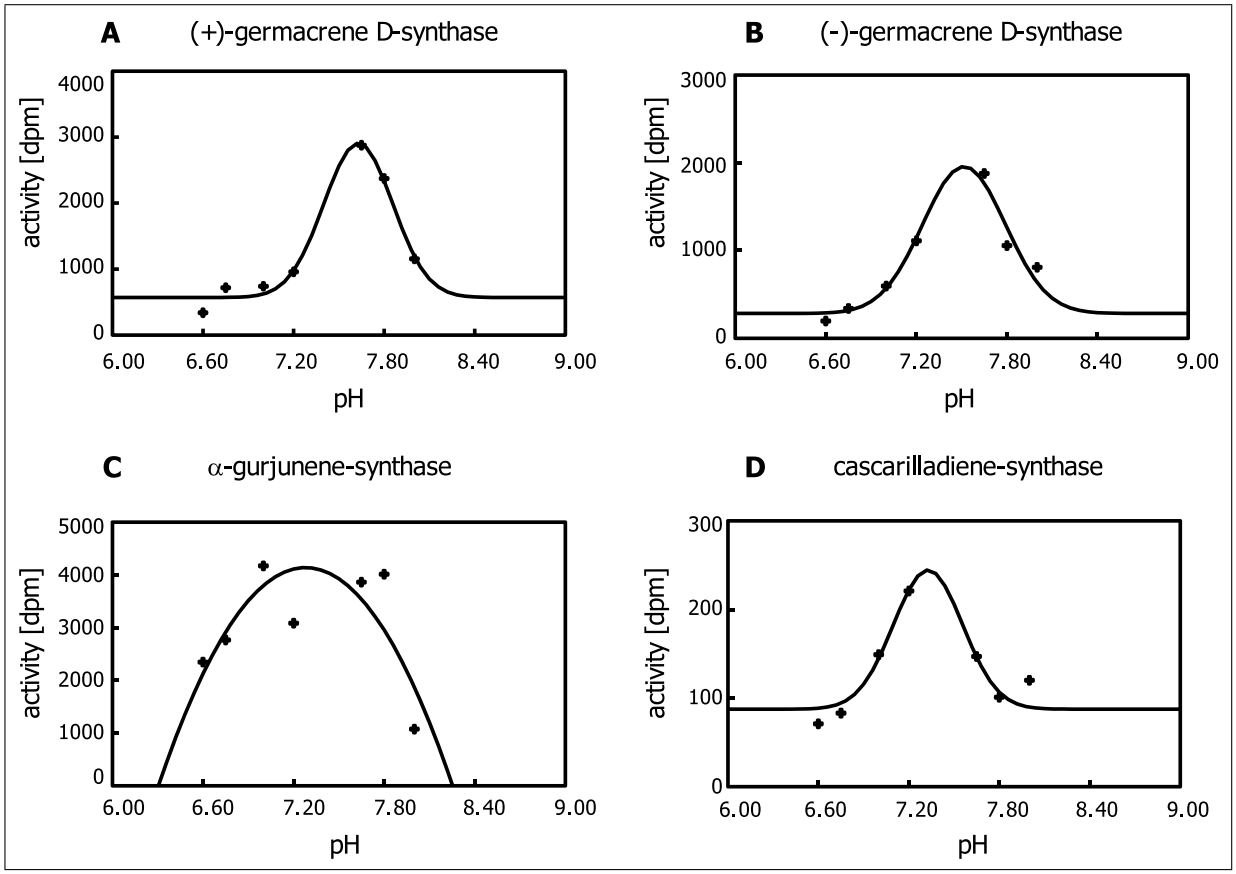


Figure 5.28: Determination of the pH optima of the four heterologously expressed enzymes. A gaussian curve was fitted by using the software "slidewriter" with the curve-fitter module "gaussian".

A = (+)-germacrene D-synthase, **B** = (–)-germacrene D-synthase, **C** = α-gurjunene-synthase, **D** = cascarilladiene-synthase

The gaussian curve is fitting well for all four enzymes besides for α-gurjunene-synthase. The measured data show a very broad pH range and a gaussian curve is not fitting very well. As a result the value for the pH optimum for α-gurjunene-synthase is very broad and with 7.3 not very precise. The other synthases have all a pH optimum around 7.5 with pH 7.6 for (+)-germacrene D-synthase, pH 7.5 for (–)-germacrene D-synthase and pH 7.3 for cascarilladiene-synthase.

enzyme	c	b	a = pH optimum	σ	r
(+)GDS	568.59	2336	7.63	0.23	0.9914
(–)GDS	283.31	1675	7.52	0.27	0.9750
αGS	-709485	713624	7.28	9.04	0.7568
CS	87.44	157	7.33	0.23	0.9536

Table 5.5: Measured pH-optimum.
(+)GDS = (+)-germacrene D-synthase, (–)GDS = (–)-germacrene D-synthase,
αGS = α-gurjunene-synthase, CS = cascarilladiene-synthase

5.9.2 Linearity

As described in section 5.9 product formation has to be linear to the reaction time and the reaction velocity has to be independent of the substrate concentration but proportional to the enzyme concentration, before kinetics measurements can be done. Therefore the linearity of the enzyme reaction has to be checked.

Time depending micro assays were carried out with the His-Tag purified heterologously expressed enzyme in phosphate buffer of pH 7.5. To ensure, that the time linearity of the enzyme reaction is concentration independent, the measurement is repeated at half concentration. The mean activity for each time point is shown in figure 5.29. Each measurement was done in duplicate. The values are given in the appendix B.2.

A linear curve was fitted for the values until 30 min. Longer incubation as shown in figure 5.29 doesn't follow linearity anymore. The micro assay was performed with 5 µL His-Tag purified enzyme for (+)-germacrene D-synthase and (–)-germacrene D-synthase in a 1:10 dilution (corresponding 1.5 µg enzyme) and 1:20 dilution (corresponding 0.75 µg enzyme). For α-gurjunene-synthase and cascarilladiene-synthase undiluted (corresponding to 15 µg enzyme) and 1/2 dilution (corresponding to 7 µg) enzyme was used. The cascarilladiene-synthase had already in undiluted form very little activity. The data for the 1/2 diluted were below the detection limit and are there-

fore not shown in figure 5.29. The data shows that linearity is given for an incubation time up to 30 min.

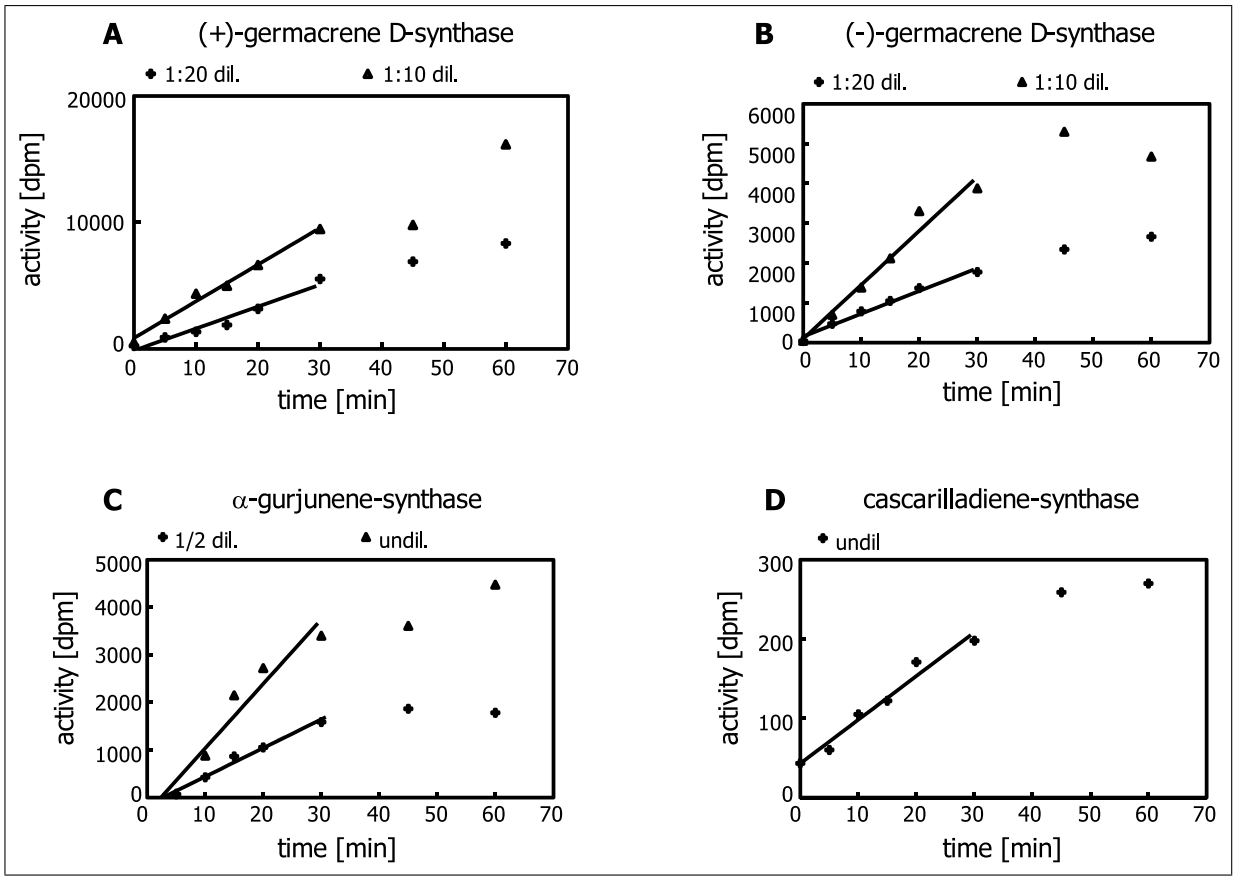


Figure 5.29: Determination of the activity of the four heterologously expressed enzymes in relation to time. A linear curve was fitted for the data until 30 min by using the software "slidewriter" with the curve-fitter module "linear".

A = (+)-germacrene D-synthase, **B** = (–)-germacrene D-synthase, **C** = α-gurjunene-synthase, **D** = cascarilladiene-synthase

.

The values shown in table 5.6 have been calculated according the linear function:

$$y = a + b \cdot x$$

with:

y = activity [dpm]

x = time [min]

a, b = constants

enzyme	dilution	a	b	r (standard deviation)
(+)GDS	1:10	853.4	290.4	0.9945
	1:20	-131.7	173.6	0.9777
(–)GDS	1:10	92.2	135.8	0.9844
	1:20	152.6	57.2	0.9914
αGS	undil.	-321.8	135.6	0.9642
	1/2	-163.7	60.2	0.9912
CS	undil.	42.9	5.5	0.9837

Table 5.6: Time linearity measurement

(+)GDS = (+)-germacrene D-synthase, (–)GDS = (–)-germacrene D-synthase,

αGS = α-gurjunene-synthase, CS = cascarilladiene-synthase

5.9.3 Kinetic studies

As linearity is given for product formation in relation to time until 30 min, the constants of the Michaelis-Menten function can be determined by measuring the amount

of product formation in relation to the substrate concentration. The function is defined as:

$$v = \frac{v_{max} \cdot [S]}{K_M + [S]} \text{ with } K_M = [S] \text{ when } v = \frac{v_{max}}{2}$$

with:

v= reaction velocity

v_{max} = maximum velocity at substrate saturation

[S] = substrate concentration

K_M = Michaelis-Menten-constant

The data shown in figure 5.30 were analyzed with the software "slidewriter" using the curve-fitter module "1-ligand", which corresponds to the Michaelis-Menten function. The measured data in detail are shown in appendix B.3

The values shown in table 5.7 for K_M and v_{max} were determined by the measurement. The determined K_M values are between 6 and 17 µM. These values are within the range for K_M values generally found for terpene synthases.

enzyme	Km [µM]	v_{max} [pmol/h]	r
(+)-GDS	9.28	183	0.9291
(–)-GDS	17.49	97	0.9809
α-GS	6.17	4961	0.9195
CS	16.07	379	0.9414

Table 5.7: Measured values for K_M and v_{max}

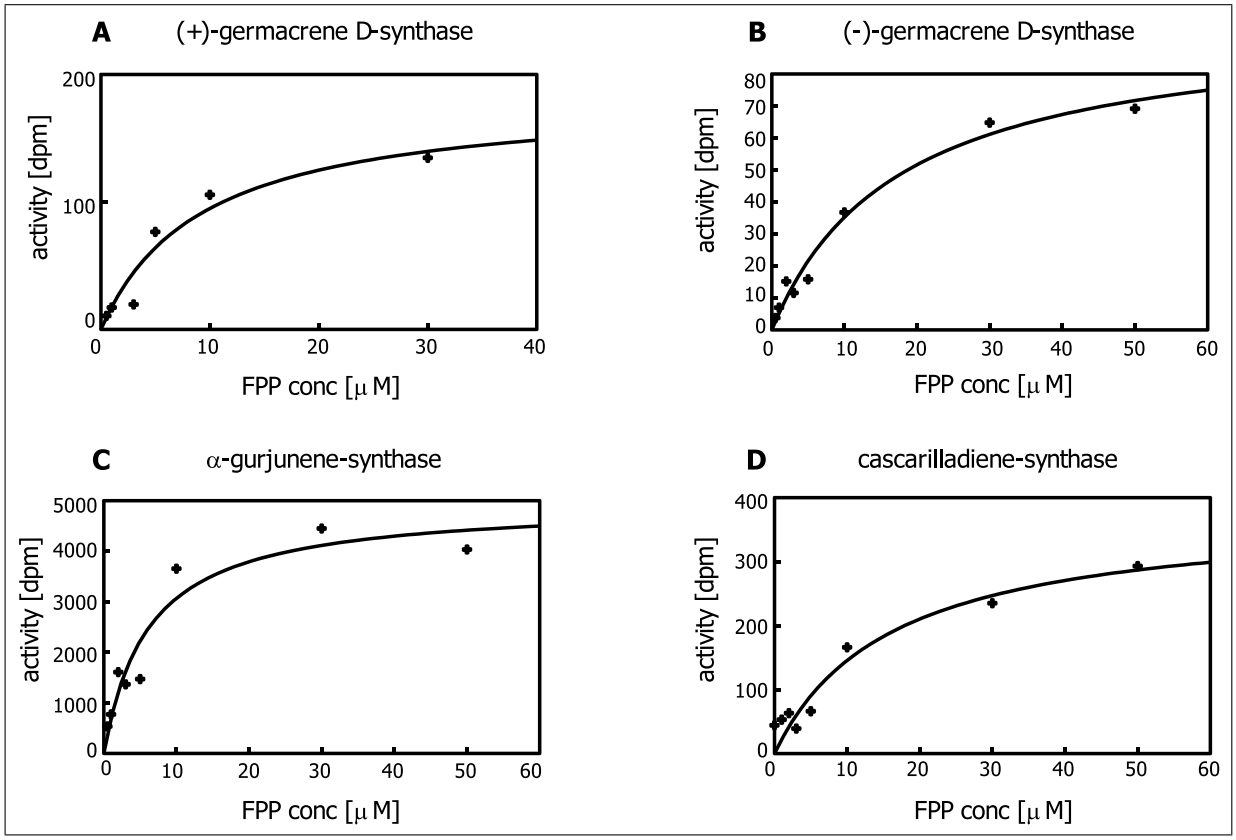


Figure 5.30: Determination of the Michaelis-Menten constant by determining the v_{max} of the four heterologously expressed enzymes. A curve was fitted for the data by using the software "slidewriter" with the curve-fitter module "1-ligand".

A = (+)-germacrene D-synthase, **B** = (–)-germacrene D-synthase, **C** = α-gurjunene-synthase, **D** = cascarilladiene-synthase

5.10 Homology Protein Modeling

A protein homology modeling approach was used to model the 3D-structure of (–)-germacrene D-synthase and (+)-germacrene D-synthase against the known structure of *epi*-aristolochene-synthase from *Nicotiana tabacum* [Starks et al., 1997]. The modeling was performed using the automated comparative protein modeling server SWISS-MODEL [Guex and Peitsch, 1997], [Peitsch, 1995] and [Peitsch, 1996]. In addi-

tion the germacrenyl cation intermediate (**2**), shown in figure 5.31, was docked into the active site pocket of (–)-germacrene D-synthase and (+)-germacrene D-synthase using DOCK 4.0 [Ewing and Kuntz, 1997]. The DOCK program is scoring for ligand orientations with the best energetic fit in the designated active site pocket. The orientations of the ligand proposed by DOCK for (–)-germacrene D-synthase and (+)-germacrene D-synthase were mostly similar to the one shown in picture 5.32.

The mechanisms of the formation of the two germacrene D enantiomers in *Solidago canadensis*, as elucidated by Schmidt et al. [Schmidt et al., 1999b], is identical until intermediate (**2**) shown in figure 5.31. The formation of (–)-germacrene D proceeds by a 1,3-hydride shift of pro-*R*-H-6 to C-11 of the isopropyl group without changing the configuration at C-7, while the formation of (+)-germacrene D proceeds via two 1,2-hydride shifts. The first 1,2-hydride shift of H-7 to the isopropyl group results in a loss of the chiral center and the second 1,2-hydride shift of pro-*R*-H-6 reverses the configuration at C-7. The difference in the formation of the two enantiomers should result from different amino acid residues involved in the catalytic action of both germacrene D-synthases. Therefore we analyzed the two models for amino acid residues that are positioned close to the docked intermediate (**2**), which are different in the two germacrene D-synthases. The multiple sequence alignment in figure 5.33 illustrates the very high homology between (–)-germacrene D-synthase ((–)GDS) and (+)-germacrene D-synthase ((+)GDS). The amino acid sequences share 85% identity. The identity to *epi*-aristolochene-synthase (EAS) is only 38% and 36%, respectively. In addition the amino acid sequence of (+)-germacrene A-synthase which was isolated earlier from *Solidago canadensis* [Prosser et al., 2002] is also included in the alignment. The (+)-germacrene A-synthase also shares very high homology to (–)-germacrene D and (+)-germacrene D-synthase with 90% and 87% identity on amino acid level, respectively. The (+)-germacrene A-synthase sequence was included, because the mechanism of the (+)-germacrene A formation is more similar to the formation of (–)-germacrene D then to the formation of (+)-germacrene D. The configuration

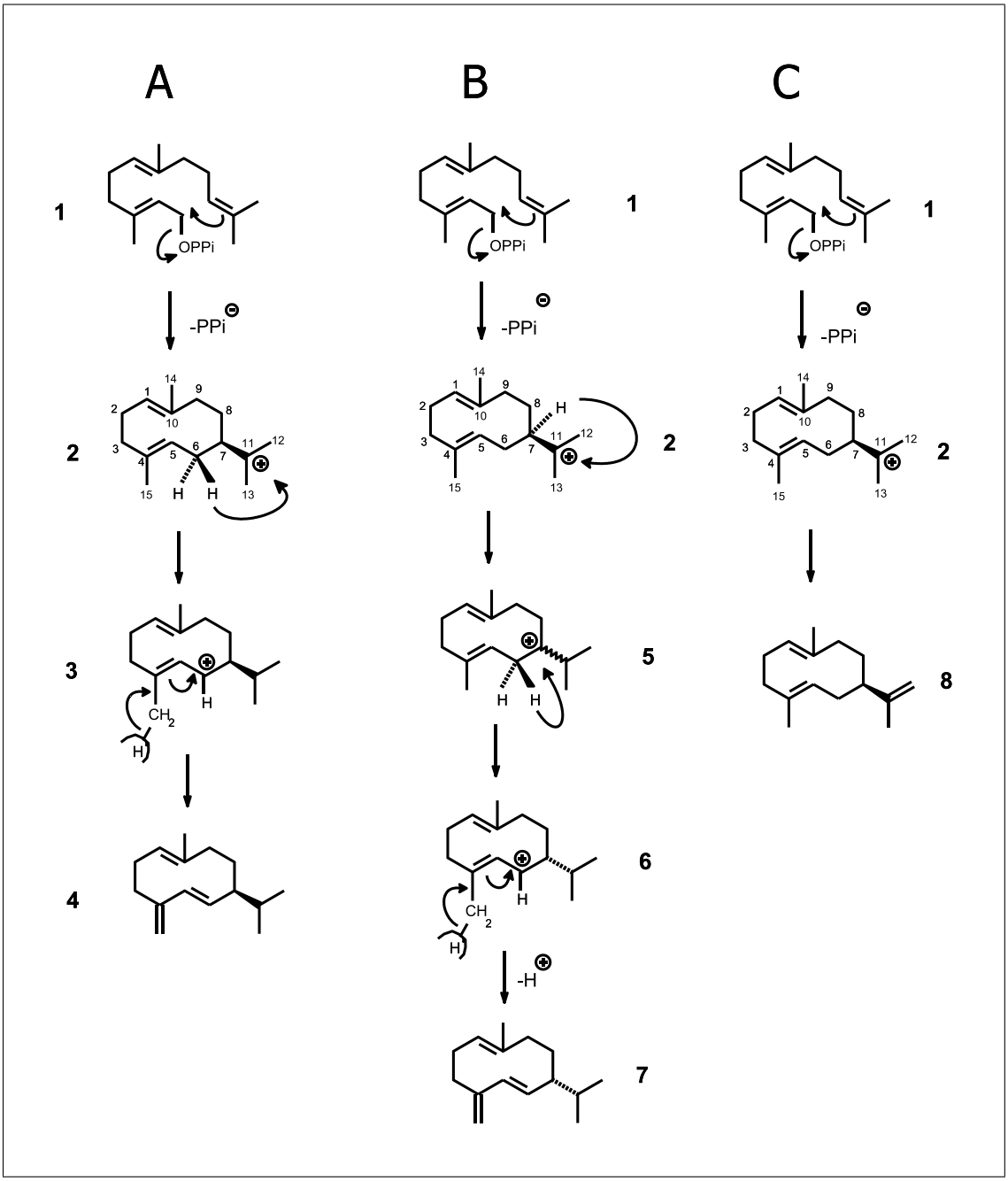


Figure 5.31: Catalytic mechanism of the formation of **A** = (–)-germacrene D (**4**),
B = (+)-germacrene D (**7**), and **C** = (+)-germacrene A (**8**).
(**1**) is the general precursor farnesyl diphosphate and (**2**), (**3**), (**5**), (**6**) are germacrenyl cation intermediates.

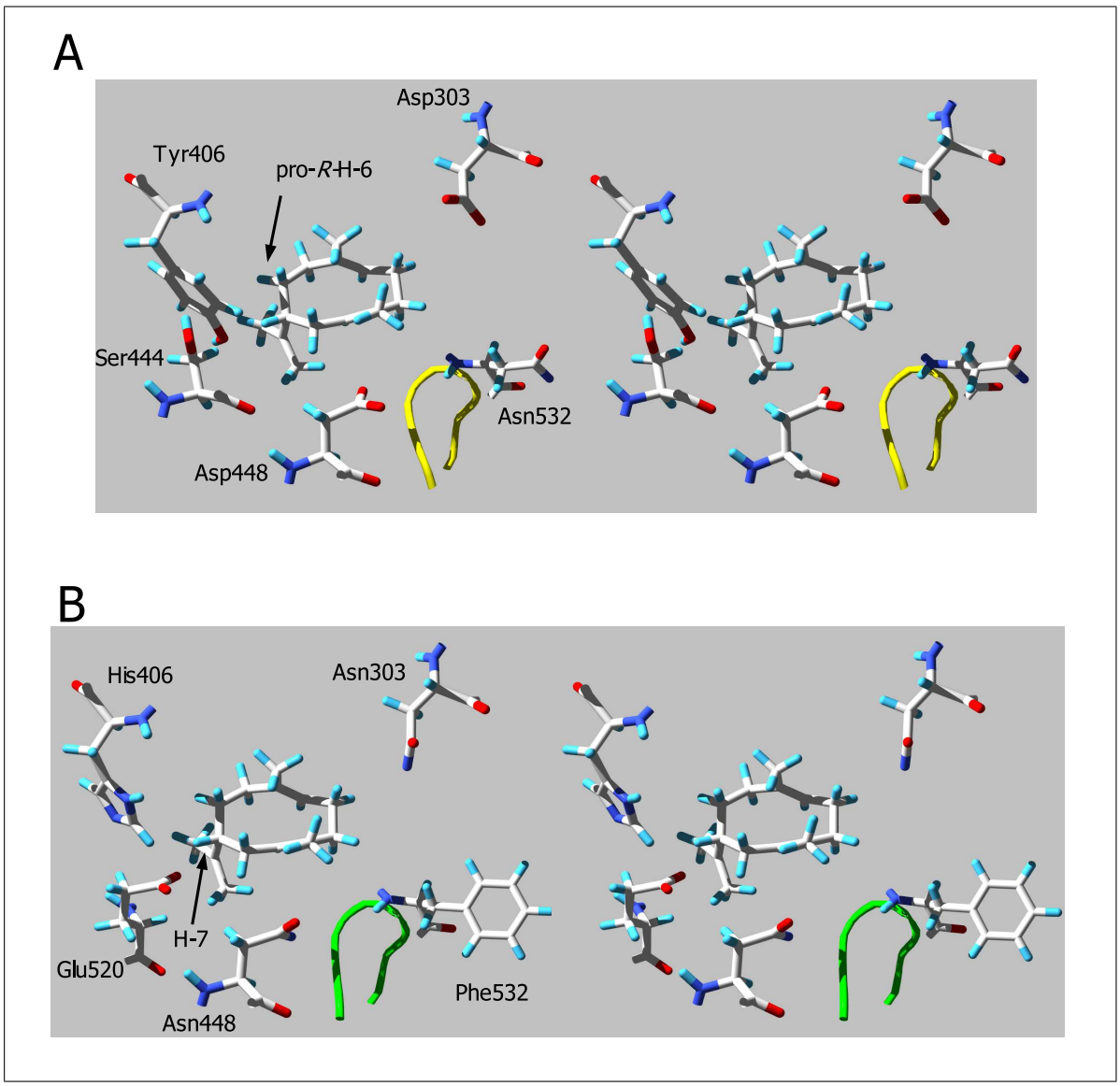


Figure 5.32: Stereoview of (–)-germacrene D-synthase (**A**) and (+)-germacrene D-synthase (**B**) with the docked germacrenyl cation (**2**) and displaying the residues proposed to be involved in the enantiospecificity.

Residue Ile 520 of (–)-germacrene D-synthase and residue Gly 444 of (+)-germacrene D-synthase are not displayed. Also displayed is the J/K loop with residue Asn 532 of (–)-germacrene D-synthase, respectively Phe 532 of (+)-germacrene D-synthase.

The models were analyzed with Sybyl and spdbViewer 3.7, rendering was done with PovRay.

of C-7 is already fixed in the same way as in the formation of (–)-germacrene D and thus might give some additional insight in the importance of certain residues.

The residues of EAS in figure 5.33 marked with an asterisk are residues proposed to be involved in the catalytic formation of *epi*-aristolochene according to Starks et al. [Starks et al., 1997]. Most of the residues of the (–)GDS and (+)GDS, which align to the proposed residues of EAS, seem not to be involved in the product- and stereospecificity as they were similar in both germacrene D-synthases. The remarkable change of the DDXXD motif in (–)GDS to NDXXD in (+)GDS is apparently not involved in the enantiospecificity either. The mutation Asp303 to Asn303 in (–)GDS didn't change the product specificity [Prosser et al., 2003]. As shown in figure 5.32 the residue Asp303 in (–)GDS, respectively Asn303 in (+)GDS is located at the backside of the germacrenyl cation (**2**) far away from the isopropyl group. Thus the orientation of the docked germacrenyl cation (**2**) seems to be confirmed by the experimental result of the mutation Asp303 to Asn300.

Five amino acid residues could be located near the active site to be different in the two germacrene D-synthases. Apparently four of these residues group together around the isopropyl group of the docked germacrenyl cation (**2**), which is another support for the assumed orientation of the docked ligand. These residues are Tyr 406, Ser 444, Asp 448, Ile 520 in (–)GDS and His 406, Gly 444, Asn 448, Glu 520 in (+)GDS. The Asp 448, respectively Asn 448, is the only residue, which aligns to one of the proposed residues for the catalysis of EAS. The hydrophobic residue Ile 520 in (–)GDS and the small Gly 444 in (+)GDS don't seem to be important for the catalysis. Eventually three residues (Tyr 406, Ser 444 and Asp 448 for (–)GDS and His 406, Asn 448 and Glu 520 for (+)GDS seem to be important for the enantiospecificity of the two germacrene D-synthases. The fifth residue Asn 532 (in (–)GDS), respectively Phe 532 (in (+)GDS) is located in the J/K-loop, which is proposed to function as a lid for the active site pocket [Starks et al., 1997]. Although these residues are not located close to the ligand

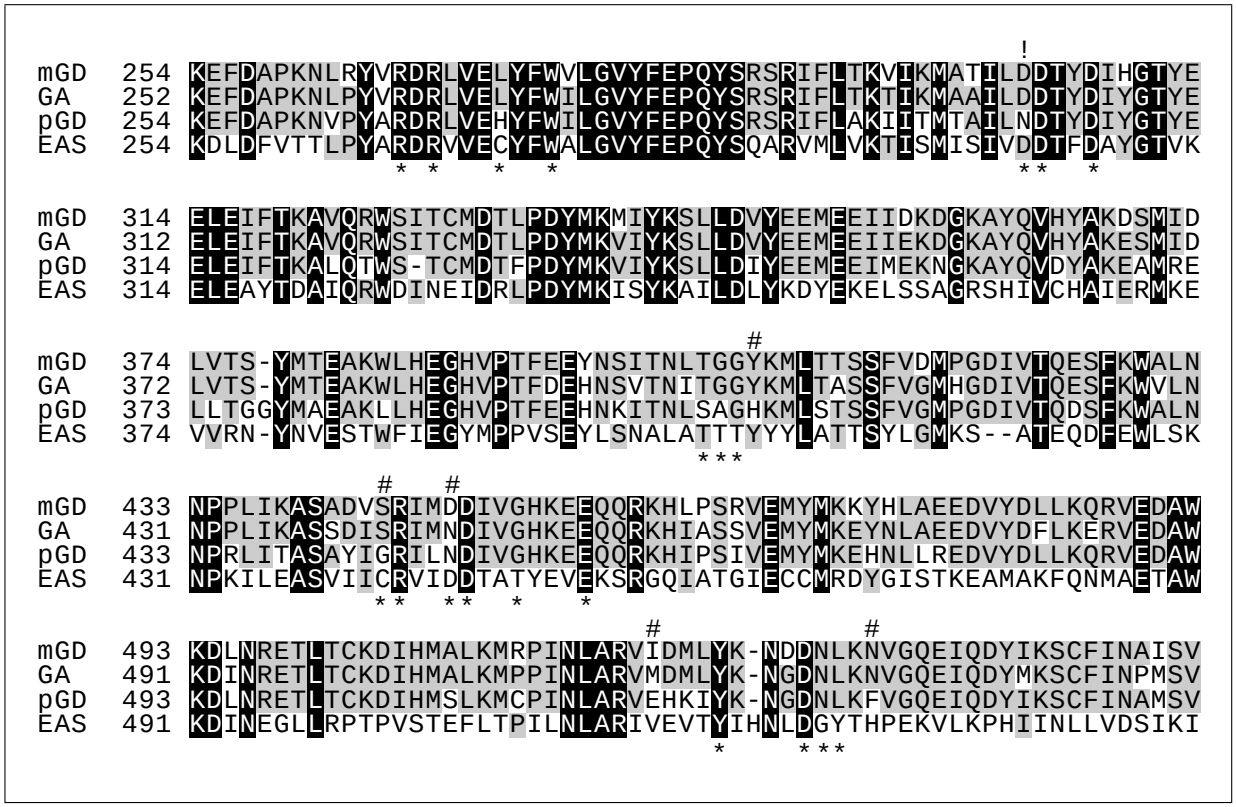


Figure 5.33: Alignment of the amino acid sequence of (–)-germacrene D-synthase ((–)GDS), (+)-germacrene D-synthase ((+)GDS), germacrene A synthase (GA) and *epi*-aristolochene synthase (EAS)

in the modeled structure, they might come in proximity to the ligand as the loop moves towards the active site pocket, and may then also be of importance for the catalysis.

Three of the four mentioned residues (Tyr 406, Ser 444 and Asn 532) of the (–)-germacrene D-synthase are identical to the corresponding residues in the (+)-germacrene A-synthase (see alignment in figure 5.33). As the mechanism of the formation of (+)-germacrene A is closer to the formation of (–)-germacrene D than to the formation of (+)-germacrene D, this can be interpreted as another proof for the importance of these residues for the enantiospecificity in the germacrene D-synthases. Especially the change of Tyr 406 in (–)GDS to His 406 in (+)GDS seems significant, as

histidine is a residue, which can be variably charged and is hence often involved in catalysis.

The different mechanisms for the formation of the germacrene D enantiomers might not only result from different polarization of a specific ligand in the active site, but also due to sterical hindering of certain hydride shifts. In both catalytic processes a hydride shift of pro-*R*-H-6 is observed: for (–)GDS in the 1,3-hydride shift and for (+)GDS in the second 1,2-hydride shift. Although the abstraction of pro-*R*-H-6 hydride is common to both synthases the type of shift is different. It is conceivable that His 406 in (+)GDS, in contrast to Tyr 406 in (–)GDS, supports the abstraction of the H-7 hydride and, moreover, prevents a 1,3-hydride shift of pro-R-H-6.

It has to be kept in mind, that the docking of the ligand is only a model and that the residues proposed to be involved in the enantiospecificity need to be verified e.g. by site-directed mutagenesis experiments.

Chapter 6

Discussion

Solidago canadensis contains several sesquiterpenes in its essential oil. Among others it is exceptional to contain both enantiomers (+)- and (–)-germacrene D. As shown in prior investigations *Solidago canadensis* possesses two individual germacrene D-synthases catalyzing the enantiospecific formation of (+)- or (–)-germacrene D from the common precursor farnesyl diphosphate. By this work, we succeded in retrieving the encoding cDNA of (+)-germacrene D-synthase and (–)-germacrene D-synthase from *Solidago canadensis* to get a better understanding of how protein structure determines the enantioselective catalysis of enzymes.

For the isolation of sesquiterpene terpene synthase encoding cDNAs a homology based cloning strategy was chosen to generate first cDNA fragments of sesquiterpene synthases from *Solidago canadensis* by degenerate primer PCR. The degenerate primers used for the PCR were already successfully used to clone sesquiterpene synthases from other plants [Bouwmeester et al., 2002], [Wallaart et al., 2001]. Several different fragments of cDNAs from *Solidago canadensis*, presumably encoding sesquiterpene synthases, were generated successfully.

The first strategy applied to obtain the full-length sequence information of the generated fragments was RACE-PCR. The method of RACE-PCR was used for fragments

A and B. By this approach the 5'- and 3'- cDNA ends starting from a known region are generated by PCR. For fragment B the RACE-PCR worked well. One type of fragment was generated for each end with a precisely fitting overlapping part. Problems occurred for fragment A. Although only one type of fragment was generated for the 3'-end, although with different splicing patterns, at least three types of fragments were generated for the 5'-end, all with a fitting overlap, but with sequence diversification towards the 3'- end. Besides, all fragments seemed to be pseudo-genes, as all of them had early stop codons in the open reading fame. A schematic overview is given in figure 5.6 in section 5.4 on page 67.

Because of the problems with RACE-PCR the strategy to obtain full-length sequence information was changed. Instead a λ-phage cDNA library was prepared from mRNA isolated from leaf material from *Solidago canadensis*. The cDNA library was then screened with the fragments generated by degenerate primer PCR. Additionally the library was screened with a PCR fragment generated with specific primers to the meanwhile known sequence of germacrene A-synthase from *Solidago canadensis* (GenBank accession AJ304452) [Prosser et al., 2002]. The sequence of the generated germacrene A fragment (834 nt) is shown in appendix A.2. This generated fragment had three changes in the nucleotide sequence compared to the GenBank published germacrene A-synthase sequence, two of them resulting in a different amino acid.

The screening of the library resulted in several sequences, some truncated at the 5'-end and some full-length. In total 24 cDNA sequences with high homology to sesquiterpene synthases were isolated. The isolated sequences pair in three different groups.

In group 1 no full-length sequence was isolated. Therefore a RACE-PCR strategy was approached to obtain the missing 5'-end with success. An alignment of the three isolated clones of group 1 (L2.2, L3.1 and L4.2) is shown in the appendix A.3. The alignment shows that one clone (clone L2.2) has some differences in the nucleotide sequence compared to the other two clones. Four nucleotides were different in the

open reading frame predominantly at the 3'-end and six nucleotides were different in the untranslated region of the 3'-end. It is likely that the sequence of group 1 exist in two different alleles. As the nucleotide differences do not cause any changes in the amino acid sequence, this was not further investigated. The heterologous expression of the full-length group 1 sequence produced a functional sesquiterpene synthase enzyme catalyzing the formation of cascarilladiene as main product. Thus the synthase was named cascarilladiene synthase (CS). The presence of cascarilladiene in the essential oil of *Solidago canadensis* was up to date unknown. In the meantime cascarilladiene has been determined in the essential oil of the root of *Solidago canadensis*.

The sequences of group 2 share very high sequence homology: identity above 89% on nucleotide level, shown in table 6.1. But they are not completely identical. Four different full-length clones were isolated. These sequences are also very highly homologous to the germacrene A sequence from *Solidago canadensis*, which was isolated by Prosser et al. [Prosser et al., 2002]. The sequences of the full-length clones are shown in an alignment including the germacrene A-synthase in figure 6.2. All isolated four full-length clones were subcloned in expression vectors and heterologously expressed in *E. coli*. Three of the four sequences expressed functional sesquiterpene synthases. The enzymes were identified upon their mainly produced product as (+)-germacrene D-synthase, catalyzing exclusively the (+)-enantiomer, (–)-germacrene D-synthase, generating exclusively the formation of the (–)-enantiomer and α-gurjunene-synthase. The fourth sequence didn't express any functioning terpene synthase. Full-length terpene synthase cDNAs showing no function on heterologous expression occur e.g. due to formation of inclusion bodies (misfolding of the protein) or other often unknown reasons [van Gelder et al., 2000] and [Aubourg et al., 2002]. The two germacrene D enantiomers represent the main sesquiterpenes in the essential oil of *Solidago canadensis*. α-Gurjunene is one of the sesquiterpene intermediates en route to cyclocolorenone, which is also one of the major components in the essential oil of *Solidago canadensis*.

This is the first time, that two sesquiterpene synthases, each of them catalyzing one enantiomer of the same sesquiterpene, in this case germacrene D, were isolated from one species. It was reported only recently, that the two enantiomers of the α-pinene synthase were isolated from loblolly pine [Phillips et al., 2003]. In addition the two α-pinene synthases from loblolly pine share only 66% identity on the amino acid level suggesting, that the synthases diverged long ago, while the two germacrene D-synthases from *Solidago canadensis* with 85% identity on the amino acid level are very closly related. Thus the two germacrene D-synthases give the possibility to get a better understanding of enantioselective catalysis of terpene enzymes.

The sequences of group 3 were from the same type as fragment A. None of these sequences were full-length. Most of them are 5'-truncated. Only one seemed to be 3'-truncated instead. The different splicing pattern, already recognized in the result of the RACE-PCR, was present again. This group of sequences seem to be pseudogenes. They are apparently translated in the cell to mRNA, but don't seem to have any function as terpene synthases, as no open reading frame was detectable.

The last type of sequence was the sequence of fragment B isolated by degenerate primer PCR. The full-length sequence information was obtained by RACE-PCR. Astonishingly no fragment or full-length sequence of this type was isolated by screening the library. It is likely, that fragment B was not present in sufficient copies in the cDNA library or in the probe, which was generated by degenerate primer PCR. As the full-length sequence was already obtained by RACE-PCR further screening of the cDNA library with only fragment B was not performed. Heterologous expression of the full-length sequence of fragment B didn't produce any active terpene synthase again for unknown reasons. The function of this sequence remains unidentified.

The phylogenetic analysis shown in figure 6.1 gives a graphical overview of the divergence of the sequences. The sequences of group 2 ((+)GDS, (–)GDS, αGS and sequence 2d), which share very high homology, group all together with the germacrene A-synthase, which was isolated by Prosser et al. [Prosser et al., 2002].

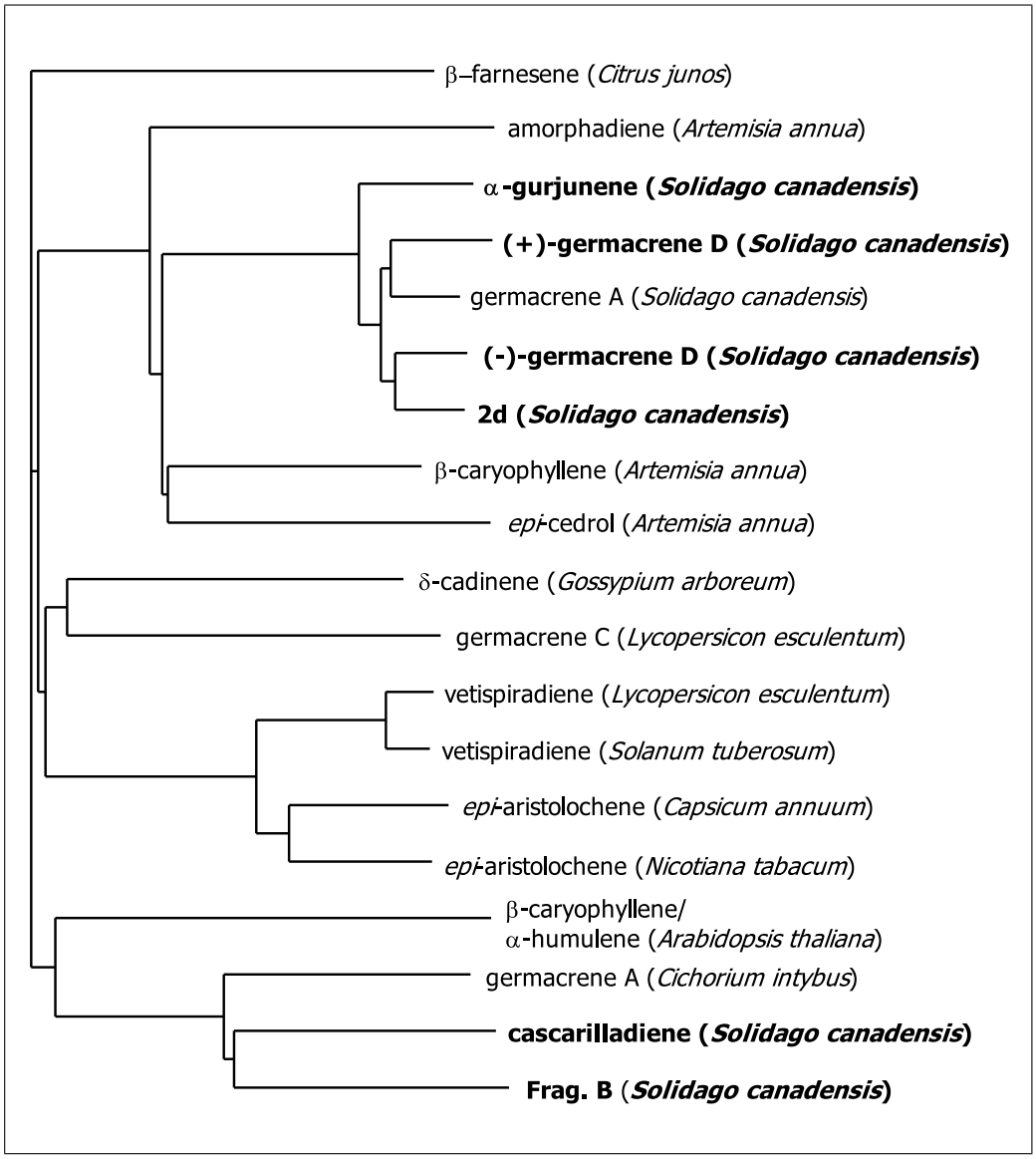


Figure 6.1: Phylogenetic tree of selected sesquiterpene synthase sequences including the isolated (+)-germacrene D-synthase, (–)-germacrene D synthase, α-gurjunene synthase and cascarilladiene synthase and also the non functional sequence of B and 2d (all in bold).

The phylogenetic tree was established by sequence alignment with ClustalX based on calculating the distances (percent divergence) and then neighbor joining according Saitou and Nei [Saitou and Nei, 1987].

In a broader context these sequences group together with sesquiterpene sequences isolated from *Artemisia annua*. *Artemisia annua* belongs, as *Solidago canadensis* does, to the plant family of *Asteraceae*. The phylogenetic tree in figure 6.1 shows, that sequence homology can be based on botanical relationship.

The other two isolated full-length sequences of cascarilladiene synthase and fragment B group also together, but further away from the group 2 sequences. These sequences group next to the germacrene A-synthase from *Cichorium intybus*, which belongs to the family of *Asteraceae*, too. Because this branch is grouped far away from the branch of the group 2 sequences, it is clear that they have diverged earlier in the evolution.

The group 2 sequences with their high homology but different catalytic function give an example for the hypothesis, that the huge variety of terpene synthases evolved by gene duplication followed by diversification [Aubourg et al., 2002]. The sequence identity within the group 2 sequences is higher on the nucleotide level than on the amino acid level (see table 6.1). As three nucleotides code for one amino acid, only a change of one nucleotide of three is sufficient to change the amino acid coding. Apparently the relationship between the sequences is closer on nucleotide level than expected from the amino acid sequence.

sequence	**(+)GDS**	**(–)GDS**	**αGS**	**2d**
(–)GDS	92 % (85 %)			
αGS	89 % (80 %)	92 % (84 %)		
2d	92 % (84 %)	94 % (89 %)	92 % (84 %)	
GA	94 % (87 %)	95 % (90 %)	92 % (85 %)	94 % (88 %)

Table 6.1: Identities of group 2 sequences on nucleotide level of the open reading frame. In brackets identities on amino acid level.

In total four synthases from *Solidago canadensis* with high homology are isolated up to date. These are the already mentioned and within this work isolated (–)GDS, (+)GDS and αGS and the germacrene A-synthase isolated by Prosser et al. [Prosser et al., 2002]. An alignment of the amino acid sequences is shown in figure 6.2. The alignment shows, that the residue variations take place especially in the second half of the sequence, where the catalytic domain of the enzymes are located. It also demonstrates, that minor amino acid residue changes (less than 15%) are sufficient to change the product specificity of the terpene synthases.

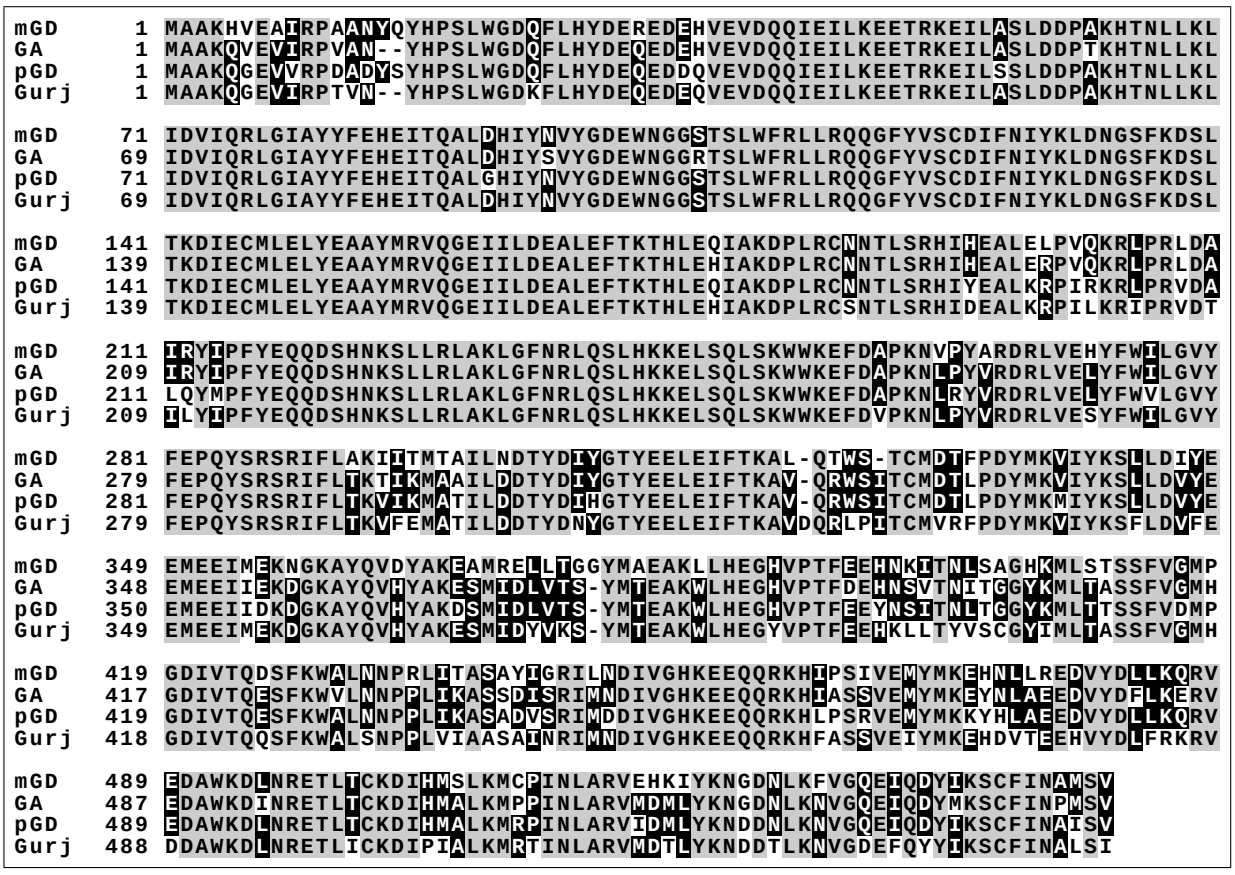


Figure 6.2: Alignment of the amino acid sequence of (–)-germacrene D-synthase (mGD), (+)-germacrene D-synthase (pGD), germacrene A-synthase (GA) and α-gurjunene synthase (Gurj).

Because of the very high homology of the group 2 sequences, but different catalytic activity, it was possible to propose residues, which are essential for the specific reaction. Especially for (+)GDS and (–)GDS it was possible to identify the amino acid residues presumably involved in the enantiospecific formation. Four residues were identified to be responsible for the enantiospecificity of the catalysis of germacrene D based on the amino acid sequence alignment and on comparison of the homology modeled structure with the first cation intermediate docked. The identified residues are Tyr 406, Ser 444, Asp 448 and Asn 532 for (–)GDS and His 406, Asn 448, Glu 520 and Phe 532 for (+)GDS. These residues differ in the two germacrene D-synthases and are located at the active site pocket. Additionally these residues were compared to the corresponding residues in the (+)-germacrene A-synthase. Three of the four mentioned residues (Tyr 406, Ser 444 and Asn 532) of the (–)-germacrene D-synthase are identical to the corresponding residues in the (+)-germacrene A-synthase. As the mechanism of the formation of (+)-germacrene A is closer to the formation of (–)-germacrene D than to the formation of (+)-germacrene D, this can be interpreted as another indication for the importance of these residues for the enantiospecificity in the germacrene D-synthases. It has been shown before that the change of one single amino acid is sufficient to interrupt the biosynthetical pathway of sesquiterpene synthases as in 5-*epi*-aristolochene synthase to demonstrate the formation of germacrene A as intermediate [Rising et al., 2000] or interrupting at all the cyclization of farnesyl diphosphate in aristolochene synthase [Deligeorgopoulou and Allemann, 2003].

Though it has to be kept in mind, that the docking of the ligand is only a model and that the residues proposed to be involved in the enantiospecificity need to be verified by e.g. site-directed mutagenesis experiments.

Enzyme kinetic data were already previously determined with native enzymes isolated from *Solidago canadensis* plant material for (+)GDS, (–)GDS and αGS by C.O. Schmidt [Schmidt, 1998]. Within this work the enzyme kinetics were also measured, but with enzymes heterologously expressed in *E. coli*. The determined values for the

Michaelis-Menten constant K_M, measured by C.O. Schmidt for the native enzymes and within this work for the heterologously expressed enzymes is shown in table 6.2.

enzyme	Km [μM] native enzyme	Km [μM] recombinant enzyme
(+)-GDS	9.97	9.28
(–)-GDS	4.24	17.49
αGS	5.54	6.17

Table 6.2: Measured values for K_M with native enzyme by C.O. Schmidt [Schmidt, 1998] and with heterologously expressed (recombinant) enzymes isolated within this work.

The differences in the measured values for K_M are small, only the K_M value for (–)GDS has some greater variation. In general the differences of the values might be caused by different conditions. While the native enzyme measurements were done in MOPSO buffer with only Mg^{2+} as cofactors, the heterologously expressed enzymes were investigated in phosphate buffer with Mg^{2+} and Mn^{2+} as cofactor. It might also be that the different conditions under which the enzyme kinetic data were determined had somehow a bigger influence on the (–)GDS, than on the other sesquiterpene synthases, resulting in the greater variation of the K_M.

Chapter 7

Materials

7.1 Chemicals

All chemicals used were standard chemicals of analytical grade from the following companies: Acros, Fluka, Merck, Roche Diagnostics (Boehringer Mannheim), Sigma.

Water mentioned in the experiments is always nanupure water purified by MilliQ system from Millipore.

7.2 Apparatus

Electrophoresis

Agarose gel electrophoresis was done using vertical submerging electrophoresis systems from Bio-Rad laboratories.
SDS-PAGE and Western blotting was done with the Mini-PROTEAN electrophoresis system from Bio-Rad laboratories.

Gas chromatograph

GC-MS analysis was done using either a HP 5890 series II equipped with a mass selective detector model 5927A from Hewlett-Packard or coupled to a VG Analytical 70-250S mass spectrometer and a capillary HP-5MS column (30 m x 0.25 mm, film thickness of 0.25 μm) at helium flow rate of 0.969 mL min^{-1}. The splitless injection of 1 μL sample proceeded at an injection port temperature of 210°C. After an initial temperature of 55°C for 4 min, the column was programmed at 5°C min^{-1} to 210°C.

The mass spectra were recorded at 70 eV scanning from 30 to 250 atomic mass units.

Enantioselective gas chromatographic analysis was performed on a Carlo Erba Fractovap 2150 gas chromatograph with FID detector, equipped with a heptakis(6-*O*-*tert*-butyl-dimethylsilyl-2,3-di-*O*-methyl)-β-cyclodextrin phase, isothermal at 120°C. Injector port temperature was 200°C and FID detector temperature 250°C.

BioRobot

the pipetting Robot BIOROBOT 9600 from Qiagen was used for plasmid miniprepping.

Sequencer

ABI PRISM 310 Genetic Analyzer from PE Biosystems was used for cDNA sequencing.

Thermocycler

PCR was carried using thermocycler 96-Well GeneAmp PCR System 9700 from PE Applied Biosystems, or Robocycler 96 from Stratagene, or MJ research PTC Peltier thermal cycler from MJ Research.

Scintillation counter

TRI-CARB 2100 TR from Packard was used for measuring the radiolabeled microassay.

Lumi Imager

Lumi-Imager F1 Workstation from Roche Diagnostics was used for visualization of the Westernblots.

Photometer

Microtiter reader SmartSpec Spectrophotometer from Bio-Rad Laboratories was used for quantitative protein measurements.

Micro titer reader fluorescence

Fluoroskan II from MTX Lab Systems was used for quantitative measurements of mRNA.

7.3 Software

7.3.1 DNAstar

For sequence analysis, sequence assembling and primer design the software package Lasergene from DNAstar was used.

7.3.2 NCBI BLAST

BLAST (Basic Local Alignment Search Tool) analysis was done at NCBI at the site http://www.ncbi.nlm.nih.gov/BLAST/ to identify homologous sequences in the free accessible GenBank database at NCBI.

7.3.3 spdbViewer

For visualization and analysis of 3D homology models Deep View Swiss-PdbViewer version 3.7 was used. Docking experiment were performed with DOCK 4.0 [Ewing and Kuntz, 1997] in combination with SYBYL from Tripos.

7.3.4 Alignment with ClustalX

Sequence alignment was performed with ClustalX from http://www.ebi.ac.uk/clustalw. Shading of the alignment was performed by using boxshade version 3.1.1 available at http://www.ch.embnet.org/software/BOX_doc.html.

7.4 Nucleotides and Nucleic acids

7.4.1 DNA molecular weight marker

DNA molecular weight marker VII (0.37-8.0 kbp) from Roche Diagnostics or 1 kb or 1 kb-plus DNA ladder from Gibco-Life Technologies (Invitrogen) were used, see figure 7.1.

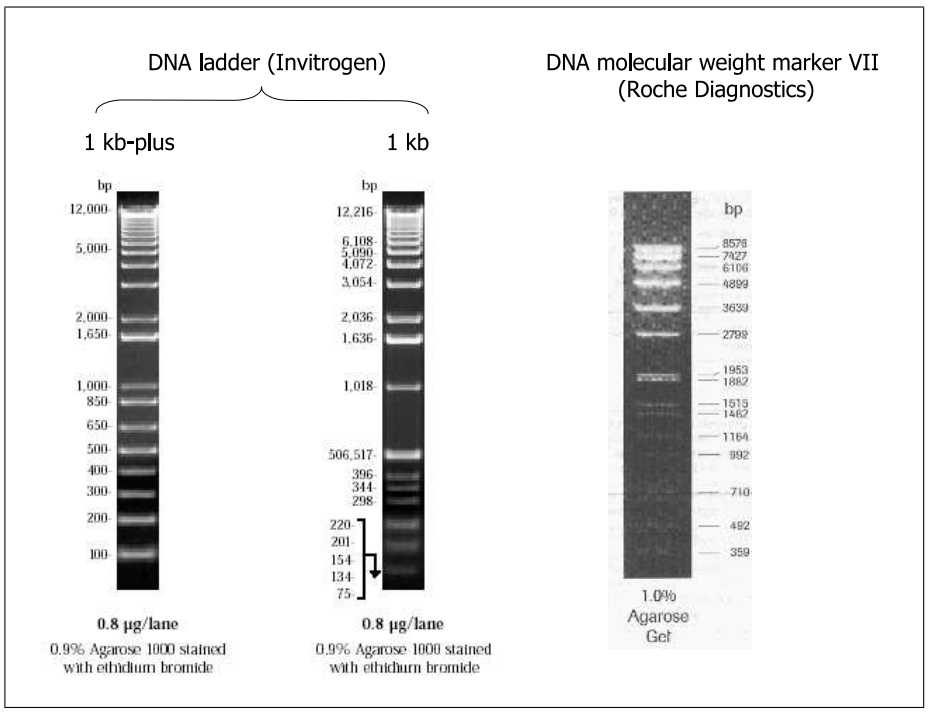


Figure 7.1: DNA ladder 1 kb and 1 kb-plus from Invitrogen and DNA molecular weight marker VII from Roche Diagnostics

7.4.2 Protein molecular weight marker

Unstained precision protein standard from Bio-Rad Laboratories was used for SDS-PAGE, see figure 7.2.

7.4.3 Deoxynucleotides (dNTP's)

100 mM dNTP sets from Gibco-Life Technologies (Invitrogen) were used for PCR.

7.4.4 Radioactive labeled nucleotides

As radioactive labeled nucleotide deoxyadenosine 5'-[α^{32}P] triphosphate (3,000 Ci/mmol) from Amersham Biosciences was used.

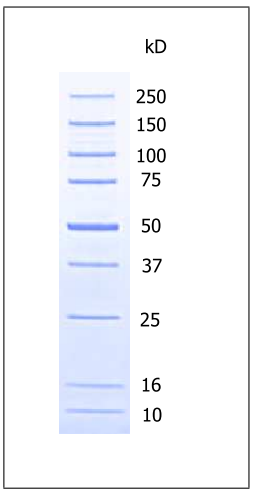


Figure 7.2: Precision protein standard, unstained from Bio-Rad Laboratories

7.4.5 Synthetic oligonucleotides

Synthetic oligonucleotides (primers) were purchased from Eurogentec, Belgium

7.5 Plasmids

7.5.1 pGEM-Teasy vector (3018 bp) from Promega

PCR generated fragments were ligated directly into the pGEM-T_{easy} vector from Promega. This vector has a 3′-terminal thymidine overhang at both ends preventing self-recircularization and providing a compatible overhang for PCR products generated by *Taq* polymerase. The *Taq* polymerase generally adds a single deoxyadenosine to the 3′-ends of amplified fragments. The vector contains T7 and SP6 RNA polymerase promoters flanking a multiple cloning region within the α-peptide coding region of the enzyme β-galactosidase. A successful cloning of an insert into the vector interrupts the coding sequence of β-galactosidase and recombinant clones can be identified by blue/white screening on indicator plates containing IPTG, x-gal and ampicillin.

7.5.2 pRSET A vector (2900 bp) from Invitrogen

The pRSET vectors are pUC-derived expression vectors designed for high-level protein expression and purification from cloned genes in *E. coli*. High levels of expression of DNA sequences cloned into the pRSET vectors are made possible by the presence of the T7 promoter. In addition, DNA inserts are positioned downstream and in frame with a sequence that encodes an N-terminal fusion peptide, a poly-histidine tag that functions as a metal binding domain in the translated protein. The metal binding domain of the fusion peptide allows simple purification of recombinant proteins by immobilized metal affinity chromatography (Nickel-column). A map of the vector is shown in figure 7.3

7.5.3 pET 32c vector (5900 bp) from Novagen

pET 32c vector is also an expression vector with similar features as pRSET vector, see above (section 7.5.2). In addition the pET 32c vector adds an additional thioredoxin tag (Trx), which helps producing soluble heterologously expressed protein. A vector map of pET 32c is shown in figure 7.4.

7.6 Enzymes

7.6.1 Polymerase

SuperTaq polymerase from HT Biotechnology LTD was used as *Taq* polymerase for PCR.
For proofreading PCR *pfuTurbo* polymerase from Stratagene was used.

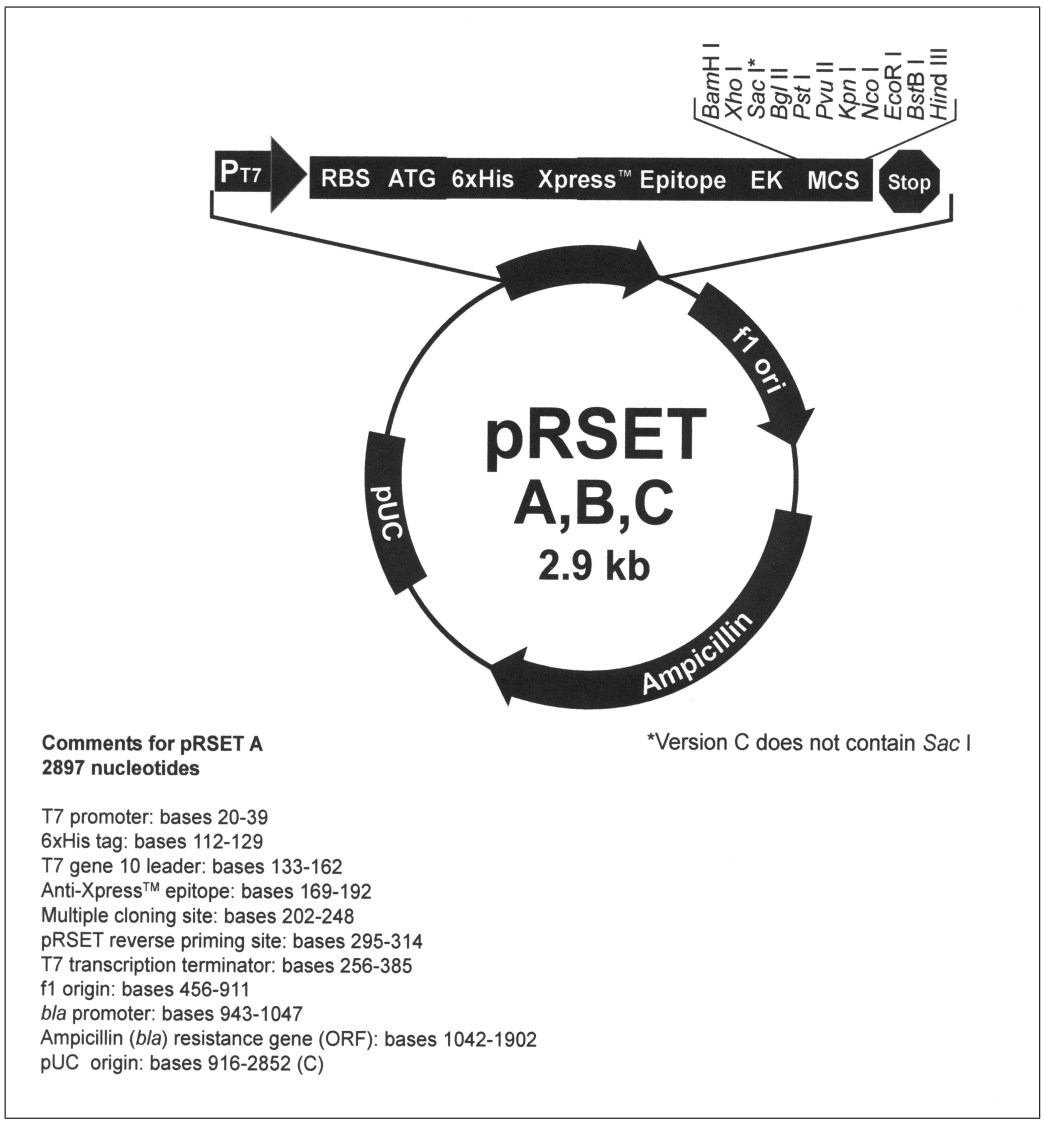
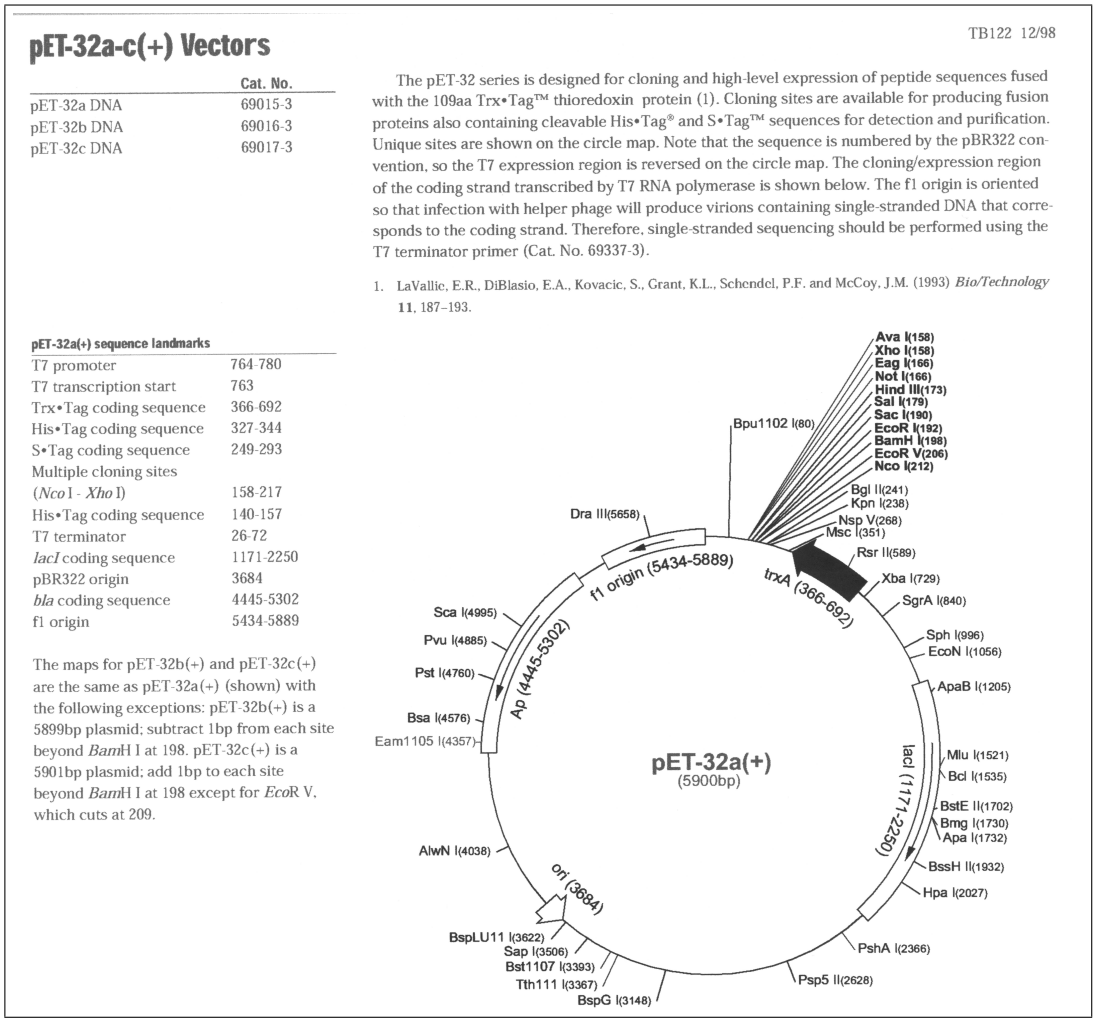


Figure 7.3: Vector map of pRSET A.

7.6.2 Reverse transcriptase

AMV reverse transcriptase and MMLV reverse transcriptase from Gibco-Lifetech (Invitrogen) were used for reverse transcription of mRNA to cDNA.



pET-32a-c(+) Vectors

TB122 12/98

	Cat. No.
pET-32a DNA	69015-3
pET-32b DNA	69016-3
pET-32c DNA	69017-3

The pET-32 series is designed for cloning and high-level expression of peptide sequences fused with the 109aa Trx•Tag™ thioredoxin protein (1). Cloning sites are available for producing fusion proteins also containing cleavable His•Tag® and S•Tag™ sequences for detection and purification. Unique sites are shown on the circle map. Note that the sequence is numbered by the pBR322 convention, so the T7 expression region is reversed on the circle map. The cloning/expression region of the coding strand transcribed by T7 RNA polymerase is shown below. The f1 origin is oriented so that infection with helper phage will produce virions containing single-stranded DNA that corresponds to the coding strand. Therefore, single-stranded sequencing should be performed using the T7 terminator primer (Cat. No. 69337-3).

1. LaVallie, E.R., DiBlasio, E.A., Kovacic, S., Grant, K.L., Schendel, P.F. and McCoy, J.M. (1993) *Bio/Technology* **11**, 187–193.

pET-32a(+) sequence landmarks	
T7 promoter	764-780
T7 transcription start	763
Trx•Tag coding sequence	366-692
His•Tag coding sequence	327-344
S•Tag coding sequence	249-293
Multiple cloning sites (*Nco* I - *Xho* I)	158-217
His•Tag coding sequence	140-157
T7 terminator	26-72
lacI coding sequence	1171-2250
pBR322 origin	3684
bla coding sequence	4445-5302
f1 origin	5434-5889

The maps for pET-32b(+) and pET-32c(+) are the same as pET-32a(+) (shown) with the following exceptions: pET-32b(+) is a 5899bp plasmid; subtract 1bp from each site beyond *Bam*H I at 198. pET-32c(+) is a 5901bp plasmid; add 1bp to each site beyond *Bam*H I at 198 except for *Eco*R V, which cuts at 209.

Figure 7.4: Vector map of pET 32c.

7.6.3 Restriction Endonuclease

Restriction endonuclease *BamH* I, *EcoR* I, *Nco* I and *Xho* I were used from Gibco-Lifetech (Invitrogen). *BspLU11* I was used from Roche Diagnostics. Incubation temperature and time was set according the manufacturers instruction and the recommended incubation buffers were used.

7.6.4 Ligase

T4 DNA ligase from Gibco-Lifetech (Invitrogen) was used to ligate DNA into vectors.

7.7 Solutions, Broths and Media

1% agarose gel

1 g agarose was dissolved under heating in 100 mL TBE or TAE buffer. Ethidium bromide was added to a final concentration of 0.01%, the gel was poured into a geltray and cooled down.

Ampicillin stock solution and working concentration

For the ampicillin stock solution 100 mg/mL ampicillin were dissolved in water and filter sterilized. The final concentration in the medium was 100 mg/L.

Carbenicillin stock solution and working concentration

For carbenicillin stock solution 50 mg/mL carbenicillin were dissolved in water and filter sterilized. The final concentration in the medium was 50 mg/L.

Enzyme Assay Buffer

15 mM MOPSO, 10 mM $MgCl_2$, 1 mM $MnCl_2$, 2 mM DTT, 1 mM sodium ascorbate, 6 mM sodium ortho-vanadate, 10% (w/v) glycerol, 0.1% (w/v) Tween 20, pH adjust to 7.0 was used as buffer for enzyme assays.

Ethidium bromide stock solution

1% ethidium bromide in water

His-tag purification: lysis buffer

50 mM NaH_2PO_4, 300 mM NaCl, 5 mM imidazole, 0.1 %(w/v) $TritonX_{100}$, 0.5 mM PMSF, pH adjust to 8.0.

His-tag purification: wash buffer

50 mM NaH_2PO_4, 300 mM NaCl, 20 mM imidazole, 0.1 %(w/v) $TritonX_{100}$, pH adjust to 8.0.

His-tag purification: elution buffer

50 mM NaH_2PO_4, 300 mM NaCl, 250 mM imidazole, 0.1 %(w/v) $TritonX_{100}$, pH adjust to 8.0.

IPTG (Isopropyl thiogalactopyranoside) stock solution

250 mg / mL water, filter sterile.

LB (Luria Bertani) medium

1% Bacto-Tryptone, 0.5% yeast extract, 0.5% NaCl were dissolved in water and autoclaved.

LB (Luria Bertani) plates

1% Bacto-Tryptone, 0.5% yeast extract, 0.5% NaCl and 1.5% agar-agar were dissolved in water and autoclaved. After cooling to 65°C the appropriate antibiotics were added and the medium was poured on Petri dishes.

LB (Luria Bertani) plates for blue/white screening

On each LB plate containing the appropriate antibiotics a mixture of 10 µL IPTG solution, 40 µL x-gal solution and 50 µL sterile water was spread and dried for 15 - 30 min under the laminar flow cabinet.

LB top agar

To 1L LB medium 10 mL filter sterile 1 M $MgSO_4$ and 0.7% agarose was added.

Loading buffer for agarose electrophoresis

15% Ficoll 400, 5 M Urea, 0.1 M Na_2EDTA pH 8.0, 0.01% Orange G were used as loading buffer for DNA agarose electrophoresis.

Minimum medium

For 100 mL minimum medium 1.5 g agar-agar were solved in 80 mL water, autoclaved and 20 mL 5x M9 salt, 2 mL 20% glucose solution, filter sterile, and 100 µL 1 M HCl-thiamine, filter sterile, were added. To prepare 5x M9 salt per 50 mL 3.2 g $Na_2HPO_4 \cdot 7H_2O$, 0.75 g KH_2PO_4, 1.125 g NaCl and 0.25 g NH_4Cl were dissolved and autoclaved.

10x Phage buffer

3.5 mM Tris-HCl (pH 7.5), 100 mM NaCl, 10 mM $MgSO_4$, 0.01% gelatine, autoclave.

SOB medium

2% Bacto-tryptone 0.5% yeast-extract, 10 mM NaCl, 10 mM $MgCl_2 \cdot 6H_2O$, adjust to pH 7.0. After autoclaving 10 mL of a filter sterile 1 M $MgSO_4 \cdot 7H_2O$ solution were added to 1 L.

SOC medium

The same solution as SOB medium but with added filter sterile glucose solution to a final concentration of 20 mM glucose.

20x SSC (sodium chloride-citrate buffer)

3 M NaCl, 0.3 M tri-sodium citrate, pH 7.0.

10x TAE (Tris-acetate EDTA) buffer

40 mM Tris, 10 mM acetic acid, 1 mM EDTA pH 8.5.

TB medium

10 mM Pipes, 15 mM $CaCl_2 \cdot H_2O$, 250 mM KCl, adjust to pH 6.7. After autoclaving filter sterile $MnCl_2$ solution to a final concentration of 55 mM was added.

10x TBE (Tris-borate EDTA) buffer

89 mM Tris-base, 89 mM boric acid, 2 mM EDTA pH 8.5.

10 xTBS (Tris-base sodium chloride) buffer

24,2 Tris base, 80 g NaCl, ad 1L, pH 7.5.

TE (Tris EDTA) buffer

10 mM Tris-HCl pH 8.0, 1 mM EDTA pH 8.5.

x-gal (5-bromo-4-chloro-3-indolyl-β-D-galactoside) solution

2% 5-bromo-4-chloro-3-indolyl-β-D-galactoside in dimethyl formamide .

Chapter 8

Experimental part

8.1 Isolation of nucleic acid

8.1.1 Total RNA isolation

Total RNA is isolated by using the SV Total RNA isolation System from Promega. To keep all equipment RNase free all tools are cleaned with 10% SDS solution. The liquid nitrogen frozen plant material is powdered in an ice-cold mortar. Around 50 mg powdered plant material is transferred to a cold 1.5 mL microcentrifuge tube containing 175 µL SV RNA Lysis Buffer (4 M guanidine thiocyanate, 0.01 M Tris pH 7.5, 0.97% β-mercaptoethanol) and thoroughly mixed by inversion. 350 µL SV RNA Dilution Buffer is added and the tube is placed for 3 min in a 70°C warm water bath. After centrifugation at 14,000 g for 10 min the cleared lysate is transferred into a new tube. 200 µL 95% ethanol is added and mixed. The mixture is transferred to the Spin Basket Assembly from Promega and centrifuged for 1 min at 14,000 g. The RNA and genomic DNA are retarded on the membrane of the spin basket. After discarding the eluate, 600 µL SV RNA Wash Solution (60 mM potassium acetate, 10 mM Tris-HCl pH 7.5 and 60% ethanol) is added and centrifuged for 1 min at 14,000 g. The eluate is again dis-

carded. The genomic DNA on the membrane is digested by adding 40 µL Yellow Core Buffer (22.5 mM Tris (pH 7.5), 1.123 M NaCl, yellow dye), 5 µL $MnCl_2$ (0,09 M) and 5 µL DNase I. After 15 min of incubation at room temperature 200 µL SV DNase Stop Solution (2 M guanidine isothiocyanate, 4 mM Tris-HCl (pH 7.5) and 57% ethanol) is added and centrifuged at 14,000 g for 1 min. The membrane is washed with 600 µL SV RNA Wash Solution. After centrifugation for 1 min at 14,000 g the eluate is discarded. The RNA is eluted with 100 µL nuclease-free water into a new tube by centrifugation for 1 min at 14,000 g.

To check the quality of the isolated total RNA a small aliquot solution (around 2 µL) is run on a 1% agarose gel in TBE buffer with 0.1% ethidium bromide. Two distinct bands of 18S and 28S r-RNA should be visible, otherwise the RNA isolation has to be repeated as the RNA has been degraded during isolation.

8.1.2 mRNA isolation

mRNA isolation from total RNA

For the isolation of mRNA from total RNA magnetic beads from Dynal (Dynabeads M-280 Streptavidin 10 mg/mL) and a magnetic rack for 1.5 mL microcentrifuge tubes are used. The beads are prepared by taking 50 µL of the vortexed bead solution into a new 1.5 mL microcentrifuge tube and placed in the magnetic rack. The storage buffer is discarded and the beads are washed twice with 50 µL of 2x STEX (2 M NACl, 20 mM Tris-HCl pH 9.0, 2 mM EDTA, 0.2% Triton X-100) and twice with 50 µL of 1x STEX. The beads are resuspended in 25 µL 1x STEX including 1 µL of poly-$dT_{25}V$ (1 µg/µL) and incubated for 30 min at room temperature. The beads are washed three times with 50 of µL 1x STEX and finally resuspended in 25 of µL 1x STEX.

Up to 100 µL of total RNA isolate are diluted 1:1 with 2x STEX and heated for 5 min at 65°C. 25 µL of the prepared bead solution are added, mixed and incubated for 20

min at room temperature. After washing three times with 100 µL 1x STEX the beads are taken up in 25 µL water. To elute the mRNA from the beads, they are incubated for 5 min at 65°C and the water and the eluted mRNA is taken in a new microcentrifuge tube.

Direct Isolation of mRNA from plant lysate

For the direct isolation of mRNA from plant lysate GenoPrep mRNA beads from GenoVision are used. The beads are prepared by taking 50 µL of the bead suspension in a new 1.5 mL microcentrifuge tube and placed into the magnetic rack. The storage buffer is discarded and the beads are resuspended in 100 µL lysis/binding solution A (100 mM Tris-HCl (pH 7.5), 500 mM LiCl, 10 mM EDTA, 5 mM DTT, 1% LiDS). A small piece of plant tissue (around 100 mg) is frozen and ground to powder in liquid nitrogen and 700 µL lysis/binding solution A is added. After lysis of the cells, the lysate is centrifuged for 5 min at 14,000 g. The cleared lysate is transferred into a new microcentrifuge tube and the prepared bead suspension is added. The mixture is incubated for 5 min at room temperature to allow hybridization. After the beads are collected in the magnetic rack, the solution is removed and the beads are washed twice in 500 µL washing solution B (10 mM Tris-HCl (pH 7.5), 150 mM LiCl, 1 mM EDTA, 0.1% LiDS) and twice in 500 µL washing solution C (10 mM Tris-HCl (pH 7.5), 150 mM LiCl, 1 mM EDTA). The mRNA is then eluted in 20 µL RNase-free water by incubation for 2 min at 65°C.

Isolation of DNA fragments from agarose gels

The DNA fragment band is excised from the agarose gel, chopped in small pieces and transferred into a 1.5 mL microcentrifuge tube. For the extraction the QIAEX II Agarose Gel Extraction kit from Qiagen is used. The agarose is dissolved in the chiatropic salt buffer QX1, while the DNA is bound to the silica beads QIAEX II during the incubation

at 50 °C for 10 min. The silica beads are collected at the bottom of the tube by centrifugation for 30 sec at 12,000 g. The supernatant is removed, the beads are washed with 500 µL buffer QX1 and then washed twice with the PE Buffer. The DNA is eluted from the silica beads by resuspension in 20 µL 10 mM Tris-HCl (pH 8.5) and incubation for 5 min at 50°C.

Isolation of plasmid DNA from *E. coli*

For the isolation of plasmid DNA QIAprep 8 Turbo Miniprep kit from Qiagen together with the BIOROBOT 9600 is used. An overnight culture is grown at 37°C in 5 mL LB medium containing the appropriate antibiotics. The cells are spun down at 2000 g and the supernatant is discarded. The bacterial pellet is resuspended in 250 µL buffer P1. 250 µL buffer P2 is added and gently mixed. After cell lysis 500 µL buffer N3 is added and the mixture is transferred into the wells of the TurboFilter strips. Vacuum is applied until the sample has passed through. The debris of the lysate is retarded on the top filter, while the plasmid DNA is retarded on the second filter. The flow-through solution and the top filter are discarded. The second filter is washed once with 1 mL PB buffer and once with 1 mL PE buffer, afterwards maximum vacuum is applied until the filter is dried. The filter is placed on a collection tube and the plasmid is eluted with 125 µL 10 mM Tris-HCl (pH 8.5). The final plasmid concentration is around 50 ng plasmid/µL.

8.1.3 PCR purification

For purification of PCR reaction products or of restriction endonuclease reaction products the QIAquick PCR purification kit from Qiagen is used. 5 volumes of PB buffer are added to 1 volume of sample and mixed. To bind the DNA the mixture is transferred to a QIAspin column and centrifuged at 12,000 g for 1 min, the flow-through is discarded. The bound DNA is washed with 750 µL PE buffer by centrifuging for 1 min

at 12,000 g. After discarding the flow-through, the column is centrifuged an additional 1 min at 12,000 g. The column is placed in a clean 1.5 mL microcentrifuge tube. The DNA is eluted with 30 µL elution buffer (10 mM Tris-HCl, pH 8.5) by centrifuging for 1 min at maximum speed.

8.2 Polymerase chain reaction (PCR)

8.2.1 Reverse transcriptase PCR (RT-PCR)

First strand cDNA is synthesized by taking up to 20 µL of the isolated mRNA solution and incubating for 1 min at 65°C with 3 µL of poly-$dT_{25}V$ (1 µg/µL). Then the temperature setting is shifted to 42°C. 28 µL stock solution containing 20.5 µL water, 5 µL 10x cDNA Buffer1 (500 mM Tris-HCl pH 9.0, 800 mM $MgCl_2$, 40 mM DTT), 2 µL dNTP's (25 mM each) and 0.5 µL AMV reverse transcriptase (Gibco-Lifetech /Invitrogen) is added and incubated for 3 h at 42°C.

8.2.2 Polymerase chain reaction with *Super Taq*

To perform a PCR with *Super Taq* enzyme (HT Biotechnology LTD), the following PCR mastermix is prepared for each reaction:
0.1 µL downstream primer (0.1 µg/µL)
0.1 µL upstream primer (0.1 µg/µL)
0.1 µL dNTP's (25 mM each)
0.1 µL *Super Taq* enzyme (5 units/µL)
1.0 µL 10x Super Taq PCR buffer (100 mM Tris-HCl pH 9.0, 15 mM $MgCl_2$, 500 mM KCl, 1.0% Triton X-100, 0.1% (w/v) stabilizer)
7.6 µL water

1.0 µL DNA sample solution

The low temperature program is used for PCR with degenerate primers. For all other PCRs the normal temperature program is used.

Low temperature program

94°C for 4 min	1 cycle
94°C for 30 sec 42°C for 30 sec 72°C for 1 min	 for 38 cycles
72°C for 8 min	1 cycle

Normal temperature program

94°C for 4 min	1 cycle
94°C for 30 sec 50°C for 30 sec 72°C for 1 min	 for 38 cycles
72°C for 8 min	1 cycle

8.2.3 Polymerase chain reaction with proof-reading enzyme *PfuTurbo*

To perform a proof-reading PCR with *PfuTurbo* enzyme from Stratagene, the following PCR mastermix is prepared for each reaction:
1.25 µL downstream primer (0.1 µg/µL)
1.25 µL upstream primer (0.1 µg/µL)
1.0 µL dNTP's (25 mM each)

0.5 µL *PfuTurbo* enzyme (2.5 units/µL)
5.0 µL 10x Pfu PCR buffer (200 mM Tris-HCl pH 8.8, 20 mM $MgSO_4$, 100 mM KCl, 100 mM $(NH_4)_2SO_4$,1.0% Triton X-100, 1 mg/mL nuclease-free BSA)
40 µL water
1.0 µL DNA sample

The following temperature program is used for proofreading PCR.

96°C for 1 min	1 cycle
96°C for 1 min 55°C for 1 min 72°C for 1 min	 for 25 to 30 cycles
72°C for 8 min	1 cycle

8.2.4 RACE-PCR

The SMART RACE cDNA Amplification Kit from CLONTECH is used for the RACE-PCR. This is a mechanism to generate full-length cDNAs by using SMART technology (Switching Mechanism At 5′ end of RNA Transcript) and RACE-PCR (rapid amplification of cDNA-ends). As shown in scheme 4.6 in section 4.2.3 the 5′-RACE reaction starts from mRNA transcribed to cDNA with a certain MMLV reverse transcriptase. When this MMLV reverse transcriptase variant reaches the end of the mRNA template, it exhibits a terminal transferase activity that adds 3-5 residues (predominantly dC) to the 3′-end of the first strand cDNA. The terminal stretch of dG residues of the SMART oligo adaptor nucleotide can anneal to the dC-rich cDNA tail and serve as an extended template for the reverse transcriptase. After the reverse transcriptase switches templates from the mRNA to the SMART oligo adaptor nucleotide, a complete cDNA copy of the original m-RNA is synthesized with the additional SMART adaptor sequence at the 5′ end.

For the RACE-PCR, gene specific upstream primers have to be designed. Following reverse transcription, the first-strand cDNA is used directly in 5'- and 3'-RACE-PCR to generate the doublestrand 5' RACE fragment, or the doublestrand 3'-RACE fragment.

First strand cDNA synthesis for RACE-PCR

Two separate reactions are needed, one for the preparation of 5'-RACE cDNA and one for the 3'-RACE cDNA:

For preparation of 5'-RACE cDNA
3 µL mRNA sample (containing around 1 µg mRNA), 1 µL 5'-CDS primer, 1 µL SMART II oligo nucleotide are mixed.

For preparation of 3'-RACE cDNA
3 µL mRNA sample (containing around 1 µg mRNA), 1 µL 3'-CDS primer, 1 µL water are mixed.

The reaction mixture is incubated for 2 min at 70°C. After cooling on ice, the following is added to each reaction and mixed:
2 µL 5x first-strand buffer (250 mM Tris-HCl (pH 8.3), 375 mM KCl, 30 mM $MgCl_2$), 1 µL DTT (20 mM), 1 µL dNTP mix (10 mM each), 1 µL MMLV reverse transcriptase Superscript II (200 units/µL) from Gibco-Life Technologies (Invitrogen).
The mixture is then incubated for 1.5 h at 42°C in an air incubator. Afterwards 20 µL Tricine-EDTA buffer (10 mM Tricine-KOH (pH 8.5), 1.0 mM EDTA) is added and the tubes are heated at 72°C for 7 min.

RACE-PCR

The following mastermix is prepared for each reaction:
34.5 µL water
5 µL 10x Advantage 2 PCR buffer (400 mM Tricine-KOH (pH 8.7), 150 mM KOAc, 35 mM MG $(OAc)_2$, 37.5 µg/mL BSA, 0.05% Tween-20, 0.05% Nonidet-P40)
1 µL dNTP mix (10 mM each)
1 µL 50x Advantage 2 Polymerase mix, containing TITANIUM Taq DNA Polymerase, a small amount of proofreading polymerase, TaqStart Antibody (1.1 µg/µL), 50% glycerol, 15 mM Tris-HCl (pH 8.0), 75 mM KCl, 0.05 mM EDTA

Add for the 5'-RACE reaction:
2.5 µL 5'-RACE cDNA, 5 µL UPM (universal primer mix), 1 µL gene specific primer.

Add for the 3'-RACE reaction:
2.5 µL 3'-RACE cDNA, 5 µL UPM (universal primer mix), 1 µL gene specific primer.

The following thermal cycling program is then used for the RACE-PCR:

94°C for 30 sec 72°C for 3 min	for 5 cycles
94°C for 30 sec 60°C for 30 sec 72°C for 3 min	for 5 cycles
94°C for 30 sec 55°C for 30 sec 72°C for 3 min	for 27 cycles
72°C for 8 min	1 cycle

The resulting PCR product is then separated on agarose gel. The biggest band of the desired size (generally the largest) is isolated and subcloned in a pGEM-T_{easy} vector. Several independent clones are picked and sequenced.

8.2.5 RACE-PCR on Library cDNA lysate

cDNA isolation from library lysate

To prepare cDNA from library lysate, about 500 µL library lysate are taken and incubated for 1 h at 37°C with 1 µL RNase I (DNase free, 10 ng/mL). The lysate is then extracted with 500 µL buffered phenol solution by shaking. After centrifugation for 2 min at 14.000 g the aqueous top layer is saved into a new tube. The phenol extraction is repeated with 500 µL buffered phenol and the aqueous top layer is again transferred into a new tube after centrifugation. Then the solution is extracted twice with 500 µL chloroform.

Primary RACE reaction

Add for the 5′-RACE reaction:
38.5 µL water
5 µL 10x Advantage 2 PCR buffer (see above)
1 µL dNTP mix (10 mM each)
1 µL 50x Advantage 2 Polymerase mix (see above)
2.5 µL cDNA from library lysate
1 µL gene specific primer 1 (0.1 µg/µL)
1 µL vector primer 1 (0.1 µg/µL)

The following thermal cycling program is then used for primary RACE-PCR:

94°C for 30 sec 72°C for 3 min	for 5 cycles
94°C for 30 sec 60°C for 30 sec 72°C for 3 min	for 5 cycles
94°C for 30 sec 55°C for 30 sec 72°C for 3 min	for 27 cycles
72°C for 8 min	1 cycle

Nestered RACE reaction

Add for the nestered RACE reaction:

38.5 µL water

5 µL 10x Advantage 2 PCR buffer (see above)

1 µL dNTP mix (10 mM each)

1 µL 50x Advantage 2 Polymerase mix (containing TITANIUM Taq DNA Polymerase, a small amount of proofreading polymerase, TaqStart Antibody (see above)

2.5 µL cDNA from the 1:20 diluted primary RACE reaction

1 µL gene specific primer 2 (0.1 µg/µL)

1 µL vector primer 2 (0.1 µg/µL)

The following thermal cycling program is then used for nestered RACE-PCR:

94°C for 30 sec 55°C for 30 sec 72°C for 3 min	for 25 cycles
72°C for 8 min	1 cycle

8.3 Subcloning

8.3.1 Restriction endonuclease reaction

A restriction endonuclease reaction is performed by using up to 50 µL DNA solution containing 50 ng DNA, 6 µL 10x reaction buffer and 1,5 µL enzyme (containing 1 U per µg DNA) and water to a final volume of 60 µL. In general the digestion is incubated for 1 h at 37°C. Only for the BspLU11 I restriction endonuclease incubation is performed at 50°C.

8.3.2 Ligation reaction

Ligation in pRSET vector

For ligation into the pRSET vector from Invitrogen, the insert and the vector have to be digested with the appropriate restriction endonucleases to obtain the fitting DNA overhang. Then 4 µL 5x ligase buffer from Invitrogen (250 mM Tris-HCl (pH 7.6) 50 mM MgCl2, 5 mM ATP, 5 mM DTT, 25% (w/v) polyethylene glycol-8000), 1 µL T4 DNA ligase from Invitrogen and in total 15 µL pRSET vector and insert DNA in a molar ratio of vector : insert between 1:3 to 1:1 are mixed together and incubated over night at 4°C.

Ligation in pGEM-T easy vector

If ligation of a PCR fragment with blunt ends, e.g. generated by proofreading enzyme is desired, an A-tailing procedure has to be performed first. 9 µL purified PCR fragment are mixed with 1 µL 10x PCR buffer (see section 8.2.2), 0.2 µL dATP (10 µM) and 1 µL Taq polymerase and incubated for 30 min at 70°C. The appropriate amount of this reaction is then used for the ligation.

For the ligation 5 µL 2x rapid ligation buffer, 0.5µL pGEMT$_{easy}$ vector (25 ng), 3.5 µL PCR fragment and 1 µL T4 DNA ligase (3 Weiss units/µL) are mixed. The molar ratio of vector : insert should be between 1:3 to 1:1. To obtain optimal ligation, the reaction is incubated over night at 4°C.

8.3.3 Preparation of competent *E. coli* cells

Preparation of competent *E. coli* DH5α cells

The preparation of competent *E. coli* DH5α cells is carried out according to the protocol of Inoue [Inoue et al., 1990]. Competent cells are capable of transferring vector DNA through their cell wall.

Streak a small portion of the frozen stock of DH5α cells on a minimal medium plate see section (7.7) and incubate it for 2 to 3 days at 37°C. A single colony is picked and incubated over night at 37°C in 10 mL SOB medium. The next day a 1:1000 dilution is prepared in 250 mL SOB medium and grown at 18°C until OD_{600} has reached 0.600. After cooling the cells for 10 min on ice, they are centrifuged at 2500 g for 10 min at 4°C and the supernatant is discarded. The cells are resuspended gently in 80 mL ice-cold TB medium and incubated for 10 min on ice. After another centrifugation for 10 min at 2500 g, discarding the supernatant, the cells are resuspended in 20 mL ice-cold TB medium. DMSO is added to a final concentration of 7% under vigorous shaking and placed again on ice for 10 min. The cells are aliquoted to 600 µL in microcentrifuge tubes, deep-frozen in liquid nitrogen and stored at -80°C

Preparation of competent *E. coli* BL21DE3 cells

To prepare competent *E. coli* BL21DE3 cells, a scratch of a frozen stock of these cells is made on a LB plate and grown at 37°C over night. The next day an isolated colony is

picked to inoculate 10 mL LB medium and grown under shaking over night at 37°C. 100 µL of the overnight culture are diluted in again 10 mL LB medium and grown for 1-2 h until OD_{600} has reached around 0.5. The cells are spun down for 5 min at 2000 g and resuspended in 5 mL (or half the starting volume) fresh, ice-cold 0.1 M $CaCl_2$ (in water, filter sterile) by vortexing. The cells are then incubated for 30 min on ice and then spun down again for 5 min at 2000 g. The cells are then resuspended in 1 mL (1/10 of the starting volume) ice-cold 0.1 M $CaCl_2$ solution by snipping with the finger-tip. The cells are now ready for transformation. Around 100 µL of the prepared competent cell solution is needed for one transformation.

8.3.4 Heat-shock transformation

About 100-200 µL suspension of competent cells is taken in a tube and incubated for 30 min on ice with up to 20 µL ligated plasmid solution. The tubes are heat-pulsed for 30 sec without agitation in a 42°C water bath. After cooling them on ice, 0.8 mL SOC medium is added and shaken for 1 h at 37°C. 125 µL of the grown cell solution is spread on LB plates containing the appropriate antibiotics or the needed indicators for e.g. blue/white screening. The left over solution can be kept for 1 day at 4°C in case a more concentrated or diluted spread is needed. The plates are incubated over night at 37°C.

8.3.5 Frozen Cell Stock

A frozen stock of cells is prepared by inoculating a single colony in the appropriate growth medium over night at the optimal growth temperature. 1.5 mL of the overnight culture are taken up in a screw-capped microcentrifuge tube and mixed with 0.5 mL glycerol. The tube is then deep frozen in liquid nitrogen and stored in a -80°C freezer up to 1 year.

8.4 Sequencing of cDNA

For sequencing a PCR is performed with ABI PRISM dRhodamine Terminator Ready Reaction mix. The mix contains:
A-Dye Terminator labeled with dichloro[R6G]
C-Dye Terminator labeled with dichloro[TAMRA]
G-Dye Terminator labeled with dichloro[R110]
T-Dye Terminator labeled with dichloro[ROX] deoxynucleoside triphosphates (dATP, dCTP, dGTP, dTTP)
AmpliTaq DNA Polymerase
$MgCl_2$
Tris-HCl buffer, pH 9.0

The structures of the dRhodamine dye terminator are shown in figure 4.8, section 4.2.4. 5 µL Rhodamine ready mix, 4 µL miniprepped plasmid (containing 50 ng - 500 ng plasmid) and 1 µL sequencing primer (0.1 µg/µL) is mixed and thermocycled according the following temperature program:

96°C for 3 min	1 cycle
96°C for 30 sec	
50°C for 15 sec	
60°C for 4 min	for 35 cycles

After PCR the product is purified by precipitation with 20 µL 2 mM $MgCl_2$ and 55 µL 95% ethanol. After incubation for 20 min at room temperature the pellet is centrifuged for 25 min at 16,000 g, rinsed with 250 µL 70% ethanol and dried on a heat block at 90°C for 1 min. The pellet is then resuspended in 15 µL TSR buffer from PE Biosystems,

vortexed, heated for 3 min at 95°C and after cooling on ice placed into the ABI PRISM 310 Genetic Analyzer from PE Biosystems.

8.5 Electrophoresis

8.5.1 Agarose gel electrophoresis

An agarose gel is prepared by dissolving 1% agarose in TBE or TAE buffer including 0.01% ethidium bromide solution. The same buffer type for preparing the gel is also used as running buffer in the electrophoresis box. Before loading the gel, 1% loading buffer (see section 7.7) is added to the samples. The samples are run at around 100 V until the dye has migrated until the end of the gel. A picture of the gel is taken under UV_{235} either with a Polaroid camera or a digital imaging system.

8.5.2 SDS-PAGE (polyacrylamide gel electrophoresis)

A 7% SDS-PAGE gel is prepared by mixing 5.35 mL water, 1.25 mL 1.5 M Tris buffer (pH 8.8), 50 µL 10% SDS, 1 mL acrylamide (30% acrylamide : 0.8 % bis-acrylamide), 50 µL 10% APS and 5 µL TEMED for the running gel. For stacking gel 1.65 mL water, 2.5 mL 0.25 M Tris buffer (pH 6.8), 50 µL 10% SDS, 0.8 mL acrylamide (30% acrylamide : 0.8 % bis-acrylamide), 50 µL 10% APS and 10 µL TEMED are mixed on ice. For pouring the gel the Mini-Protean cell assembly system from Bio-Rad is used. The samples are mixed with 4x loading buffer (200 mM Tris-HCl (pH 6.8), 8 % SDS, 0.4 % bromphenol blue, 40 % glycerol, 10 mM DTT). After the gel has polymerized, it is transferred into the electrophoresis box, containing the electrophoresis buffer (25 mM Tris (pH 8.3), 192 mM glycine, 0.1 %SDS). The gel is run for around 2 h at 50 mA. Then the gel is either stained with Coomassie or immunoblotting (Western blot) is performed with monoclonal anti-polyHistidine-peroxidase conjugate antibody (A7058) from Sigma.

Coomassie staining

Before staining, the proteins are fixed on the gel by incubation for 30 min in water/ethanol/acetic acid (50/40/10 (v/v/v)) solution. After washing the gel for 3 x 5 min in water, the gel is stained over night in 60 mL colloid Coomassie stain solution (75 g $(NH_4)SO_4$, 20 g H_3PO_4, 1 g Coomassie BB R (250) ad 1 L) mixed with 15 mL 95 % ethanol. The gel is then destained with 0.3 % Tween 20 solution for 30 - 90 min.

Western blotting

To transfer the protein on a nitrocellulose membrane, a sandwich is prepared, in which the SDS gel is overlayed by the nitrocellulose membrane assembled between 2 Whatman papers and 2 sponges. The sandwich is put into a electroblotting box from the Mini-Protean cell assembly system from Bio-Rad containing the electroblotting buffer (10 mM CAPS (pH 11), 10 % ethanol) and blotted for 60 min at 50 V. The gel is then discarded and the membrane washed in water. The membrane is blocked for 30 min with 1x TBS, 0.2 % Tween 20 and 2 % dry milk powder, fat-free, incubated then in the same solution as the block solution including a 1:2000 dilution of monoclonal anti-polyHistidine-peroxidase conjugate antibody (A7058) from Sigma. At the end the membrane is washed 3x for 5 min in water. The two component Lumi-Light western blotting substrate from Roche Diagnostic is used for the detection. The substrate consists of an enhancer solution and of a stable peroxide solution, which are mixed in a 1 : 1 ratio and poured over the blot. Detection of the chemiluminescence and imaging is carried out in the Lumi-Imager F1 Workstation from Roche Diagnostics.

8.6 Construction and screening of a cDNA library

8.6.1 Constructing the cDNA library

For constructing a cDNA library from mRNA from leaf material of *Solidago canadensis* the SMART cDNA Library Construction Kit from Clontech is used. As shown in figure 8.1 the protocol utilizes the SMART (Switching Mechanism At 5'-end of RNA Transcript) oligonucleotide in the first strand synthesis to generate full-length cDNA. After double strand (ds) cDNA synthesis by primer extension, the ds cDNA is digested with Sfi I, ligated into the λTriplEx phage vector arms and packed in λ-phages.

Preparation

For the first strand cDNA synthesis mRNA is isolated from leaf material of *Solidago canadensis*. The mRNA isolation is carried out directly from the plant cell lysate according section 8.1.2. The protocol was upscaled by the factor of 10 for the isolation, because at least 1 µg mRNA is needed as starting material for the cDNA library construction. The concentration of mRNA is measured by fluorescence measurement with the RiboGreen reagent from Molecular Probes. The determined concentration of the isolated mRNA was 3 µg/100µL.

35 µL of the mRNA solution are taken and precipitated with 4 µL 3 M sodium acetate, 1 µL glycogen (20µg/µL) and 105 µL 95% ethanol over night at -20°C. After centrifugation at 12,000 g for 20 min, the pellet is washed with 250 µL 80% ethanol, dried and dissolved in 3 µL water.

First strand synthesis

For the first strand cDNA synthesis 3 µL mRNA solution (containing 1 µg mRNA), 1 µL SMART IV Oligonucleotide (10 µM), 1 µL CDS III Primer (10 µM) are mixed in a

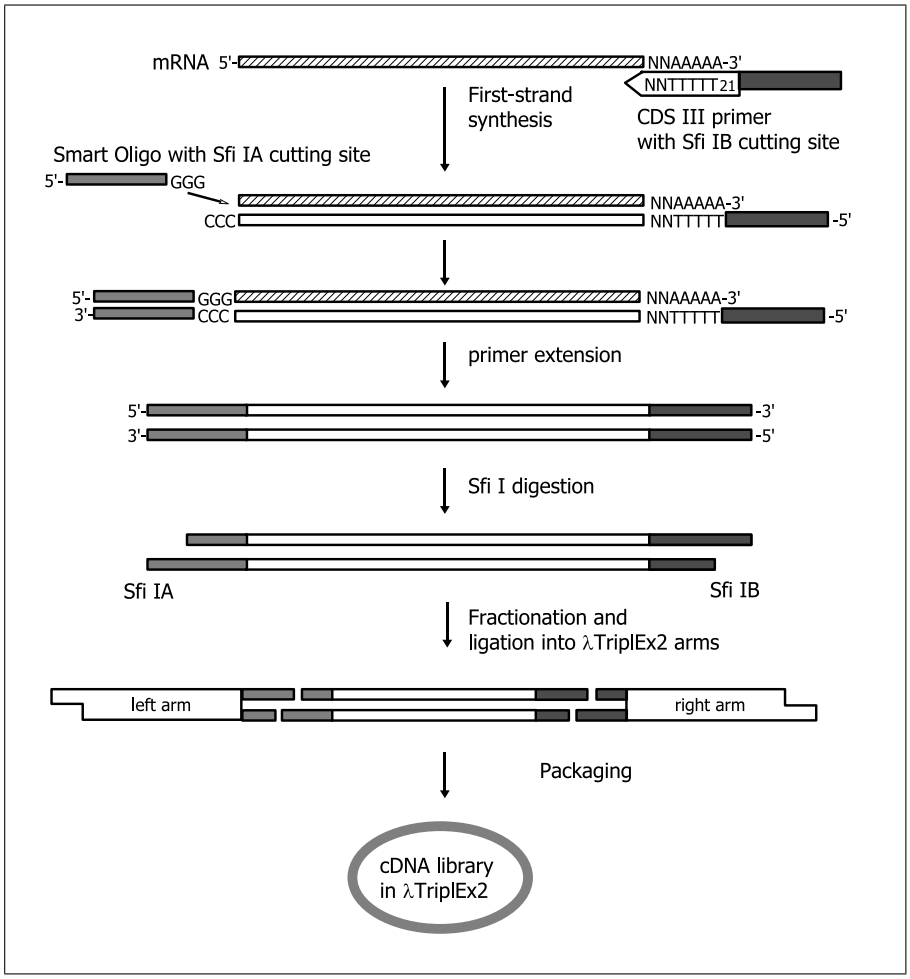


Figure 8.1: Flow chart of the SMART cDNA Library Construction Kit protocol

0.5 mL microcentrifuge tube, incubated for 2 min at 72°C and cooled on ice.
2 µL 5x First-Strand Buffer (250 mM Tris (pH 8.3), 30 mM $MgCl_2$, 375 mM KCl), 1 µL DTT (20 mM), 1 µL dNTP mix (each 10 mM), 1 µL of PowerScript Reverse Transcriptase are added, mixed and incubated at 42°C for 1 h in an air incubator. The first strand synthesis is terminated by placing the tube on ice, adding 1 µL 0.1 M sodium hydroxide and incubation for 30 min at 68°C.

Primer extension

For the double strand (ds) cDNA synthesis by primer extension 11 µL of the first strand cDNA are mixed with:
71 µL water
10 µL 10x advantage 2 PCR buffer
2 µL 50x dNTP mix
2 µL 5'PCR primer
2 µL CDS III primer
2 µL 50x advantage 2 polymerase mix

and the following temperature program is used:

72°C for 10 min	1 cycle
95°C for 20 sec	1 cycle
95°C for 5 sec 68°C for 8 min	3 cycles

50 µL of the PCR product is purified by using the QIAquick PCR purification kit from Qiagen (see section 8.1.3). After the purification 10 µL of 3 M sodium acetate, 1.3 µL glycogen (20µg/µL) and 260 µL of 95% ethanol are added. The sample is centrifuged for 20 min at 14,000 g at room temperature. The supernatant is carefully removed and the pellet is washed with 100 µL 80% ethanol. After the pellet is air dried the ds cDNA is resuspended in 79 µL water.

***Sfi* I digestion**

The ds cDNA *Sfi* I digestion is performed by combining the following in a 0.5 mL tube: 79 µL ds cDNA, 10 µL 10x *Sfi* I Buffer, 10 µL *Sfi* Enzyme, 1 µL 100x BSA. The mixture is then incubated for 2h in a 50°C water bath.

Size fractionation

2 µL of 1% xylene cyanol dye is added to the sample. A CHROMA SPIN column is taken and the matrix is resuspended by inverting several times. The storage buffer is drained through the column by gravity flow, 700 µL column buffer are added on the top and drained out. Then 100 µL mixture of the *Sfi* I-digested cDNA are applied to the top. After the sample is fully absorbed into the surface 100 µL column buffer are added on the top and also absorbed. Then 600 µL column buffer are added and the first 16 single drop-fractions are collected.

The profile of the fractions are checked by electrophoresis of 3 µL of each fraction on a 1.1% agarose/EtBr gel alongside a 1 kb DNA size marker. The first three fractions containing cDNA are taken. The DNA of each fraction is precipitated with 1/10 volume sodium acetate (3 M; pH 4.8), 1.3 µL glycogen (20 mg/mL) and 2.5 volume 95% ethanol. The tubes are placed over night at -20°C and then centrifuged for 20 min at room temperature. The supernatant is discarded and the pellet resuspended in 3 µL water after air drying.

Ligation into λTriplEx2 vector

Each fraction is ligated separately into the λTriplEx2 phage vector by mixing the following for each fraction:
3 µL cDNA of one fraction
0.5 µL vector (500 ng/µL)
0.5 µL 10x ligation buffer
0.5 µL ATP (10 mM)
0.5 µL T4 DNA ligase
The mixtures are incubated over night at 16°C. The fractions are pooled prior to λ-phage packaging reaction.

λ-phage packaging reaction

For the packaging the Packagene Lambda DNA Packaging System from Promega is used. A Packagene Extract of 50 µL is thawed gently on ice and 5 µL ligated DNA solution containing up to 0.5 µg ligated DNA is added. The extract is incubated for 3 h at 22°C. After incubation 445 µL phage buffer and 25 µL chloroform is added.

Bacterial culture plating

For the plating of the library the bacterial culture strain E. *coli* XL1-Blue is used. For the working stock plate a small portion of the frozen stock is streaked onto an LB agar plate containing 15 µg/mL tetracycline and incubated over night at 37°C. To prepare the bacterial culture solution a single colony is picked from the working stock plate and used to inoculate 15 mL LB broth containing 10 mM $MgSO_4$ and 0.2% maltose. Inoculation is done at 37°C until OD_{600} has reached 1.0. The cells are spun down and resuspended in 10 mM $MgSO_4$ solution.

Titering the library

The appropriate dilution of the packaging extract in Phage Buffer is prepared. 1 µL of the diluted phage solution is added to 200 µL bacterial culture solution (see section 8.6.1) and incubated for 20 min at 37°C. 2 mL melted LB top agar are added and poured on a pre-warmed LB plate containing 10 mM $MgSO_4$. The plates are incubated over night at 37°C. The plaques are counted and the titer is calculated according to the following formula:

$$\mathrm{pfu/mL} = \frac{\text{number of plaques} \cdot \text{dilution factor} \cdot 10^3 \mu\mathrm{L/mL}}{\mu\mathrm{L} \text{ of diluted phage plated}}$$

pfu/mL = plaque forming unit per mL

$$\text{recombinants}/\mu\text{gDNA} = \frac{\text{pfu/mL}}{\text{concentration of ligated DNA packaged}}$$

Library amplification

To 500 µL bacterial culture solution is the required amount of phage extract added to yield 6 to $7 \cdot 10^4$ plaques per 150 mm plate. The solution is incubated for 20 min at 37°C. 6 mL melted LB top agar are added and poured on a pre-warmed LB plate containing 10 mM $MgSO_4$. The plates are incubated over night at 37°C.

12 mL phage buffer are added to each plate and incubated over night at 4°C. After shaking the plates for 1 h at room temperature the λ-phage lysates are pooled.

To clear the phage lysate from the cell debris and for lysis of any remaining intact cells 10 mL chloroform is added to the lysate and vortexed for 2 min. After centrifuging for 10 min at 5000 g the clear supernatant is collected into a new tube.

8.6.2 Screening the cDNA library

Plaque transfer

For the primary screening around 70,000 plaques are plated on 150 mm agar LB plates containing 10 mM $MgSO_4$. For this up to 500 µL bacterial culture solution is taken and the required amount of phage extract is added. The solution is incubated for 20 min at 37°C. 6 mL melted LB top agar is added and poured on the pre-warmed LB plate. The plate is incubated over night at 37°C and then cooled for 2 h at 4°C. Hybond-N Nylon membrane from Amersham Biosciences is cut to the right size. The membrane is put on the agar plate and incubated for 2 min for plaque transfer. After lifting the membrane is submerged in the following solutions:

- For 5 min in the denature solution (1.5 M NaCl, 0.5 M NaOH).

- For 2x 5 min in the neutralization solution (1.5 M NaCL, 0.5 M Tris-HCl (pH 7.2), 1 mM EDTA).

- For 5 min in the wash solution 2x SSC.

The DNA is fixed on the membrane by crosslinking under UV light for 30 sec under 120,000 µJ. The agar plates are stored at 4°C.

For secondary screening the desired plaques are isolated by excision from the agar and incubated in 350 µL phage buffer at 4°C overnight to allow phage to elute. The appropriate dilution of the plaque solution is plated on 60 mm agar LB plates containing 10 mM $MgSO_4$. For this up to 100 µL bacterial culture solution (section 8.6.1) is taken and the required amount of phage extract is added. The solution is incubated for 20 min at 37°C. 3 mL melted LB top agar is added and poured on the pre-warmed LB plate and the protocol described above is followed.

Hybridization and washing

The membranes are pre-wetted with 2x SSC containing 0.1% SDS. The membranes are placed in 250 mL hybridization tubes and 15 mL of hybridization buffer (0.5% SDS, 6x SSC 5x Denhardts solution, 50x Denhardts solution: 10 g Ficoll, 10 g PVPP, 10 g BSA in 1 L) is added. The membranes are pre-hybridized for 3-5 h at 60°C. The radio labeled probe 8.6.3 is added and the membranes are hybridized for 24 h at 60°C.

The membrane is washed for 10 min at 60°C with 50 mL 2x SSC containing 0.1% SDS. Then the membrane is washed twice with 2x SSC containing 0.1% SDS at room temperature for 30 min. After wrapping the membrane in food barrier foil, it is exposed to an X-ray film (Kodak XAR) for 2-3 days at -80°C.

8.6.3 Nucleotide probe labeling

For labeling the DNA fragments the RadPrime DNA labeling system from Gibco-Life Technologies (Invitrogen) is used. 25 ng DNA are dissolved in 20 µL water and denatured by boiling for 5 min in a water bath. After cooling on ice, the following is added:
1 µL 500 µM dCTP
1 µL 500 µM dGTP
1 µL 500 µM dTTP
20 µL 2.5x RandPrime Buffer (125 mM Tris-HCl (pH 6.8), 12.5 mM $MgCl_2$, 25 mM 2-mercaptoethanol, 150 µg/ml oligodeoxyribonucleotide primers (random octamers))
2.5 µL (approx. 50 mCi) [α^{32}P]dATP
water to a total volume 49µL
1 µL Klenow fragment (40 U/µL Klenow fragment in 50 mM potassium phosphate buffer (pH 7), 100 mM KCl, 1 mM DTT, 50% glycerol)

The mixture is incubated for 10 min at 37°C. The labeled probe is purified with size exclusion ProbeQuant G-50 micro columns from Amersham Biosciences. The resin of the column is resuspended by vortexing and the column is drained by centrifugation for 1 min at 735 g. The column is transferred into a new 1.5 mL reaction tube and up to 50 µL of the labeled sample is applied on top of the column. The column is spun for 2 min at 735 g, the purified sample is collected in the reaction tube and the column is discarded. About 5 µL of the purified sample are used as probe per 250 mL hybridization tube.

In vivo excision

The conversion of a λTriplEx2 phage clone to a pTriplEx2 plasmid involves *in vivo* excision and circularization of a complete plasmid from the recombinant phage. The plasmid is released as a result of Cre recombinase-mediated site-specific recombina-

tion at the loxP sites flanking the embedded plasmid, see figure 8.2. The Cre/loxP recombination system derives from the P1 bacteriophage. The site-specific Cre (causes recombination) enzyme recognize a certain site (in this case the loxP site), cuts there the DNA and send it to the target DNAsequence [Lottspeich and Zorbas, 1998]. Release of the plasmid occurs automatically when the recombinant phage is transduced into a bacterial host in which Cre recombinase is being expressed e.g. *E. coli* BM25.8. Conversion is performed on individual positive plaques picked from the secondary screening plates. The excised plasmid can be propagated in *E. coli*.

For the *in vivo* excision a single, isolated colony from the working stock plate of BM25.8 host cells is picked and used to inoculate 10 ml of LB broth. Incubation is performed at 31°C overnight with shaking until the OD_{600} of the culture reaches 1. 100 ml of 1 M $MgCl_2$ solution are added to 10 mL culture. A well-isolated positive plaque from secondary screening plate is picked, and placed in 350 µL of phage buffer and incubated at 4°C overnight to allow phage to elute.

In a 20-ml test tube 200 ml of overnight cell culture are combined with 150 ml of the eluted positive plaque and incubated at 31°C for 30 min without shaking. 400 ml of LB broth are added and incubated at 31°C for 1 h with shaking. 100 µL of infected cell suspension are spread on an LB plates containing carbenicillin and grown at 31°C to obtain isolated colonies. A well-isolated colony is picked and the plasmid DNA is isolated by the standard method (see section 8.1.2). Information about the insert is obtained by sequencing.

8.7 Heterologous Expression

A culture from the transformed construct into *E. coli* BL21DE3 is taken to inoculate 50 mL LB medium with the appropriate antibiotics. The cells are grown until OD_{600} has reached around 0.5. The expression is then induced by adding 100 µL 2.5 M IPTG solu-

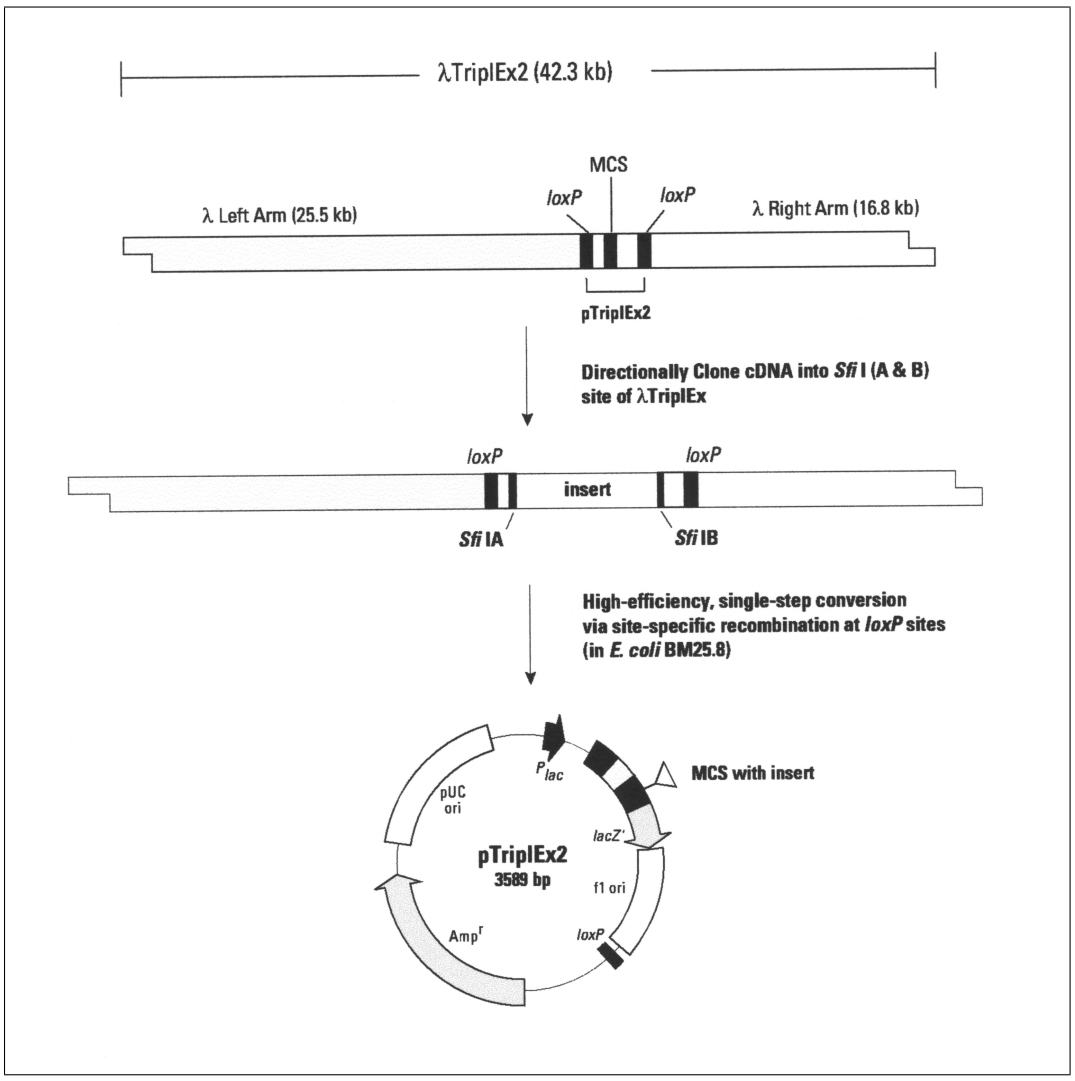


Figure 8.2: Flow chart of *in vivo* excision, **MCS**= multiple cloning site

tion and grown under shaking at 16°C over night. The next day the cells are harvested by centrifugation for 8 min at 2000 g and resuspended in 500 µL either in the enzyme assay buffer or in the His-Tag Lysis Buffer. The cells are then submitted to lysis by sonication for 6 cycles for 10 sec with 30 sec pause in between. The cell lysate is cleared by centrifugation for 5 min at 14,000 g. The expressed protein in the supernatant is then either purified by His-Tag affinity chromatography on Ni-NTA-spin columns from Qiagen or directly used for micro or macro enzyme assay.

8.7.1 His-Tag affinity chromatography

For the purification of the recombinantly expressed protein the Ni-NTA-spin columns from Qiagen are used. The expressed protein provided by the pRSET A vector or pET 32 vector possesses at the N-terminal end a His-Tag of 6 histidines, which binds to the immobilized nickel ions in the Ni-NTA resin. NTA (nitrilotriacetic acid) has a tetradentate chelating group that coordinates the nickel strongly to the resin.

The Ni-NTA spin columns are equilibrated with 600 µL His-Tag Lysis buffer by centrifugation for 2 min at 700 g. Up to 600 µL of the cleared lysate from section 8.7 are loaded onto the column and centrifuged for 2 min at 700 g preferably in a cooled centrifuge. The flow through is discarded or saved for loading an SDS gel. The column is washed twice with 600 µL His-Tag Wash Buffer by centrifugation for 2 min at 700 g. The wash solutions are discarded or saved for loading an SDS gel. The purified protein is then eluted twice with 200 µL His-Tag Elution Buffer by centrifugation for 2 min at 700 g and pooled. 20 µL of the eluate is used for a micro enzyme assay and 200 µL are used for a macro enzyme assay.

8.7.2 Macro enzyme assay

For the macro enzyme assay either 500 µL supernatant containing the expressed protein from section 8.7 diluted with 500 µL enzyme assay buffer, or 200 µL of the His-Tag purified protein from section 8.7.1 diluted with 800 µL enzyme assay buffer, are used. To the solution 2 µL 10 mM FPP (in 50% ethanol in 200 mM ammonium bicarbonate (v/v)) are added. The assay is overlayed with 1 mL pentane and incubated for 1 h at 30°C. Then the assay is vortexed, the phases are separated by a short spin and the pentane phase is passed over a short column of aluminum oxide overlayed with anhydrous magnesium sulfate. The aqueous assay phase is re-extracted with 1 mL pentane/ethyl ether (80:20, v/v), which is also passed over the column. The aqueous

assay phase is re-extracted with 1 mL ethyl ether to extract terpene alcohols, which is then also passed over the column. The column is then washed with 1.5 mL ethyl ether. If separation of terpene hydrocarbons and alcohols is desired, the column can be transferred into a new tube before the re-extraction and washing step with ethyl ether. For GC-MS analysis the organic phase is evaporated under nitrogen flow to a volume of 5 µL.

If a radioactive macro assay is desired the protocol described above is followed with the addition of 5 µL [1(n)^{3}H]farnesyl diphosphate, triammonium salt in 50 % ethanol in 100 mM ammonium bicarbonate solution (555-740 GBq/mmol, 15-20 Ci/mmol, 7.4 MBq/mL, 200 µCi/mL from Amersham Biosciences) together with 2 µL 10 mM FPP (in 50% ethanol in 200 mM ammonium bicarbonate (v/v)) before incubation. For radio-GC analysis the organic phase is evaporated under nitrogen flow until a volume of 5 µL.

8.7.3 Micro enzyme assay

100 µL supernatant containing the expressed protein from section 8.7 or 20 µL of the His-Tag purified protein diluted with 80 µL enzyme assay buffer from section 8.7.1 are used for the micro enzyme assay. To the solution 2 µL 1 mM FPP (in 50% ethanol in 200 mM ammonium bicarbonate (v/v)) and 5 µL of 1:10 diluted [1(n)^{3}H]farnesyl diphosphate, triammonium salt in 50 % ethanol in 100 mM ammonium bicarbonate solution (555-740 GBq/mmol, 15-20 Ci/mmol, 7.4 MBq/mL. 200 µCi/mL from Amersham Biosciences) is added. The assay is overlayed with 1 mL hexane and incubated for 1 h at 30°C. Then the assay is vortexed, the phases separated by a short spin and 750 µL of the hexane phase are transferred into a new tube containing 40 mg flush silica to bind the farnesol, produced by unspecific phosphotase activity present solution. After vortexing and centrifugation for 10 min at 12,000 g 500 µL of the hexane phase is transferred into a scintillation counting tube containing 4,5 mL scintillation cocktail Ultima

Gold from Packard and counted. The radioactivity is counted in dpm (disintegrations per minute, 1000 dpm = 0.45 nCi = 16.7 Bq)

8.8 Enzyme characterization

8.8.1 pH-optimum

To measure the pH optima of the enzymes, micro enzyme assays are performed with the His-Tag purified enzyme in phosphate buffer of different pH values ranging from 6.5 to 8.0. To achieve the pH range the potassium hydrogen-phosphate and di-sodium hydrogen-phosphate, each 133 mM, are mixed in different ratios shown in table 8.1.

pH	133 mM KH_2PO_4	133 mM Na_2HPO_4
6.6	96.4 mL	3.7 ml
6.75	88.9 ml	11.1 ml
7.0	73.35 ml	29.65 ml
7.2	41.3 ml	58.7ml
7.65	10.1 ml	89.9 ml
7.8	3.7 ml	96.3 ml
8.0	0	100 ml

Table 8.1: pH Buffer composition

To the phosphate buffer is then added 10 mM $MgCl_2$, 1 mM $MnCl_2$, 2 mM DTT, 1 mM sodium ascorbate, 10% (w/v) glycerol and 0.1% (w/v) Tween 20. The final volume is 200 mL

The micro enzyme assays are performed with 5 µL of His-Tag purified (+)-germacrene D-synthase, and 10 µL of each of the His-Tag purified (–)-germacrene D-synthase, α-gurjunene-synthase and cascarilladiene-synthase. The assay is incubated

for 30 min at 30°C. After the incubation a silica clean-up of the hexane phase is not needed prior to scintillation counting. Unspecific phosphatase activity is already eliminated by the His-Tag purification.

8.8.2 Time linearity measurements

Time depending micro assays are carried out with the His-Tag purified heterologously expressed enzyme in phosphate buffer of pH 7.5 (see section 8.8.1). The measurement is repeated with the half concentration to ensure, that the time linearity of the enzyme reaction is concentration independent. Samples are taken after 0 min, 5 min, 10 min, 15 min, 20 min, 30 min, 45 min and 60 min incubation time to measure the product formation by scintillation counting. The micro enzyme assays are performed with 5 µL of His-Tag purified (+)-germacrene D-synthase and (–)-germacrene D-synthase 1:10 and 1:20 diluted. For α-gurjunene-synthase and cascarilladiene-synthase 5 µL of the His-Tag purified enzyme were taken undiluted and half diluted. After the incubation a silica clean-up of the hexane phase is not needed prior to scintillation counting. Unspecific phosphotase activity is already eliminated by the His-Tag purification.

8.8.3 Kinetic study measurements

Substrate concentration depending micro assays are carried out with the His-Tag purified heterologously expressed enzyme in phosphate buffer of pH 7.5 (see section 8.8.1) and incubation time of 30 min. Incubation is done with FPP concentration of 0.5 µM, 1 µM, 2 µM, 3 µM, 5 µM, 10 µM, 20 µM, 30 µM, 50 µM, 100 µM, and 200 µM. The micro enzyme assays are performed with 5 µL of His-Tag purified (+)-germacrene D-synthase and (–)-germacrene D-synthase, 1:10 diluted. For α-gurjunene-synthase and cascarilladiene-synthase 10 µL of the His-Tag purified enzyme were taken. After the incubation a silica clean-up of the hexane phase is not needed prior to scintillation

counting. Unspecific phosphatase activity is already eliminated by the His-Tag purification.

8.8.4 Protein concentration measurements

The amount of enzyme (equivalent to protein) in the enzyme assays are measured spectrophotometrically based on the method of Bradford with the BioRad protein assay solution (Bio-Rad). It involves the addition of an acidic dye (Coomassie® Brilliant Blue G-250) to protein solution by shifting the absorbance maximum from 465 nm to 595 nm. The subsequent measurement at 595 nm with a spectrophotometer and comparison to a standard curve provides a relative measurement of protein concentration. The standard curve is made from a BSA solution diluted from 100 to 500 µg/mL. The samples are diluted in the appropriate way. After addition of 200 µL BioRad solution, the protein content is measured spectrophotometrically at 595 nm after 5 min of incubation time.

Chapter 9

Summary

In the present work several full-length cDNAs and cDNA fragments of sesquiterpene synthases from the plant *Solidago canadensis* were isolated. The cDNAs were isolated by screening a λ-phage cDNA library with a probe generated by degenerate primer PCR. Another strategy to isolated full-length cDNAs proceeded by RACE-PCR. Four of the isolated full-length cDNAs were functionally expressed in *E. coli*. The products catalytically formed by the heterologously expressed sesquiterpene synthases were generated by incubation with the general substrate farnesyl diphosphate followed by a product identification by GC-MS and GC with enantioselective stationary phase. The functionally expressed sesquiterpene synthases were identified as the following:

- (+)-germacrene D-synthase ((+)GDS) producing exclusively the (+)-germacrene D enantiomer (90%) and some unidentified minor side products.

- (–)-germacrene D-synthase ((–)GDS) producing exclusively the (–)-germacrene D enantiomer (90%) and also some unidentified minor side products.

- α-gurjunene-synthase (αGS) producing α-gurjunene (81.3%), next to γ-gurjunene (8.8%) and bicyclogermacrene (7.1%).

- cascarilladiene-synthase producing cascarilladiene (77.3%) and δ-selinene (14.5%) as side product.

The two germacrene D synthases represent the main sesquiterpenes in the essential oil of *Solidago canadensis*. α-Gurjunene is an essential sesquiterpene intermediate en route to cyclocolorenone, which is another major component of the essential oil of *Solidago canadensis*. Cascarilladiene is a minor sesquiterpene component, which was not known before to be existent in *Solidago canadensis*. The pH optima and the enzyme kinetic data were determined for each of the sesquiterpene synthases. The pH optima of the enzymes were around pH 7.5 and the K_M were between 6 and 16 μM, which are values generally found for sesquiterpene synthases.

This is the first time, that two sesquiterpene synthases, each producing one enantiomer of the same sesquiterpene, in this case germacrene D, were isolated from one plant species. Furthermore the full-length sequences of (+)GDS and (–)GDS share very high homology with amino acid sequence identity of 85%. This opened the possibility to get more insight into the enantiospecific catalytic action of the two enzymes. Comparative homology modeling was performed on the known structure of *epi*-aristolochene synthase, followed by substrate docking experiments by docking the germacrenyl cation intermediate common to the both synthases. In this way five amino acid residues were determined, which presumably are responsible in the enantiospecific catalysis of both germacrene D-synthases.

By isolating the two enantiospecific (+)GDS and (–)GDS in addition to successful heterologous expression in *E. coli*, the necessary amount of enzyme can be produced for X-ray crystallography experiments leading to the 3-D structure of the two synthases. This would give a more detailed insight into the enantiospecifically working active sites.

Furthermore a group of four sesquiterpene synthase sequences from *Solidago canadensis* with sequence identity above 84% were obtained, of which three were iso-

lated within this work ((+)GDS, (–)GDS and αGS) next to the known germacrene A-synthase [Prosser et al., 2002] showing that only 10% to 15% changes in the amino acid sequence are sufficient for product diversity of sesquiterpene synthases.

Chapter 10

Zusammenfassung

Im Rahmen der Doktorarbeit wurden mehrere Gesamtlängen-cDNAs und cDNA-Fragmente von Sesquiterpensynthasen aus der Pflanze *Solidago canadensis* isoliert. Die Isolierung erfolgte durch Screening einer λ-Phage cDNA-Bibliothek mit einer Sonde aus cDNA-Fragmente, die mittels PCR degenerierter Primer erzeugt wurden. Zusätzlich wurden Gesamtlängensequenzen mit Hilfe der RACE-PCR-Technik ermittelt. Vier der isolierten Gesamtlängen-cDNAs konnten funktional in *E. coli* heterolog exprimiert werden. Die Identifizierung der exprimierten Sesquiterpensynthasen erfolgte nach Inkubation mit dem Substrat Farnesyldiphosphat über die Identifizierung der katalytisch erzeugten Produkte. Die Produktidentifizierung wurde zum einem massenspektrometrisch und zum anderen durch gaschromatographische Untersuchungen an enantioselektiver stationärer Phase durchgeführt. Die Benennung der Sesquiterpensynthasen erfolgte nach dem katalytisch erzeugten Hauptprodukt. Es wurden folgende vier Sesquiterpensynthasen identifiziert:

- (+)-Germacren D-Synthase ((+)GDS) mit (+)-Germacren D als Hauptprodukt (90%) und einigen nicht identifizierten Nebenprodukten.

- (–)-Germacren D Synthase ((–)GDS) mit (–)-Germacren D als Hauptprodukt (90%) und einigen nicht identifizierten Nebenprodukten.
- α-Gurjunen-Synthase (αGS) mit dem Hauptprodukt α-Gurjunen (81.3%) und den zwei Nebenprodukten γ-Gurjunen (8.8%) sowie Bicyclogermacren (7.1%).
- Cascarilladien-Synthase mit dem Hauptprodukt Cascarilladien (77.3%) und dem Nebenprodukt δ-Selinen (14.5%).

Mit den zwei Germacren D-Synthasen wurden die Enzyme isoliert, welche für die Erzeugung der Sesquiterpen-Hauptkomponente (+)- und (–)-Germacren D von *Solidago canadensis* verantwortlich sind. α-Gurjunen ist die Sesquiterpenzwischenstufe für die Biosynthese von Cyclocolorenon, ebenfalls einer der Hauptkomponenten des ätherischen Öls von *Solidago canadensis*. Cascarilladien dagegen ist ein Nebenprodukt, dessen Existenz zuvor in *Solidago canadensis* unbekannt war und das erst kürzlich in der Terpenfraktion nachgewiesen werden konnte.

Das pH-Optimum und die kinetischen Daten der Enzyme wurden für alle vier heterolog exprimierten Sesquiterpensynthasen ermittelt. Das pH-Optimum der Enzyme lag bei pH = 7.5, für die Michaelis-Menten Konstante K_M wurden Werte zwischen 6 und 16 μM gefunden. Diese Werte liegen im Bereich der Werte, die auch für andere Sesquiterpensynthasen ermittelt worden sind.

In dieser Arbeit wurden zum erstenmal die codierenden Gesamtlängensequenzen zweier Sesquiterpensynthasen aus einem Organismus isoliert, die jeweils die Biosynthese der beiden Enantiomere desselben Sesquiterpens, in diesem Fall Germacren D, katalysieren.

Zudem ergab sich durch die hohe Sequenzhomologie mit einer Aminosäuresequenzidentität von 85% die Möglichkeit, die Aminosäuren der Aktivzentren der Enzyme zu ermitteln, die vermutlich für die Enantiospezifizität der Synthasen verantwortlich sind. Hierzu wurde die Methode der homologen Modellierung der räum-

lichen Struktur der Synthasen im Vergleich zu der bekannten räumlichen Struktur der Sesquiterpensynthase *epi*-Aristolochensynthase verwendet. Zusätzlich wurden Docking-Experimente mit dem ersten kationischen Zwischenprodukt, das für beide Germacren D-Enantiomere identisch ist, mit dem Aktivzentrum durchgeführt. Hierdurch konnten fünf Aminosäurereste ermittelt werden, die vermutlich für die enantiospezifische Katalyse der Germacren D-Synthasen verantwortlich sind.

Die Isolierung der enantiospezifischen (+)GDS und (–)GDS in Verbindung mit der erfolgreichen heterologen Expression in *E. coli* eröffnet weiterhin die Möglichkeit, die Enzyme in ausreichender Menge für Röntgenstrukturanalysen herzustellen. Durch die Aufklärung der 3-dimensionalen Struktur könnten die enantiospezifisch fungierenden Aktivzentren der Enzyme detailliert untersucht und der Mechanismus der enantiospezifisch ablaufenden Katalyse aufgeklärt werden.

Außerdem wurde durch die Identifizierung der (+)GDS, (–)GDS und αGS drei weitere Synthasen isoliert, die zusammen mit der schon bekannten Germacren A-Synthase eine Gruppe von vier verschiedenen Sesquiterpensynthasen in *Solidago canadensis* bilden, die alle untereinander eine Sequenzidentität über 84% besitzen. Dies zeigt, das schon ca. 10%-15% an Sequenzunterschieden zur Veränderung der Produktspezifizität der Sesquiterpensynthasen ausreichen.

Bibliography

[Adam et al., 1999] Adam, K.-P., Thiel, R., and Zapp, J. (1999). Incorporation of 1-[1-13c]Deoxy-d-xylulose in Chamomile Sesquiterpenes. *Archives of Biochemistry and Biophysics*, 369(1):127–132.

[Alonso et al., 1992] Alonso, W. R., Rajaonarivony, I. M., Gershenzon, J., and Croteau, R. (1992). Purification of 4*S*-limonene synthase, a monoterpene cyclase from glandular trichomes of peppermint (*Mentha* x *peperita*) and spearmint *Mentha spicata*. *The Journal of Biological Chemistry*, 267(11):7582–7587.

[Asakawa and Arantani, 1976] Asakawa, Y. and Arantani, T. (1976). Sesquiterpenes of *Porella vernicosa*. *Bull. Soc. Chim. France*, pages 1469–1470.

[Aubourg et al., 2002] Aubourg, S., Lecharny, A., and Bohlmann, J. (2002). Genomic analysis of the terpenoid synthase (AtTPS) gene family of *Arabidopsis thaliana*. *Molecular Genetic Genomics*, 267(6):730–745.

[Bach, 1995] Bach, T. J. (1995). Some new aspects of isoprenoid biosynthesis in plants - a review. *Lipids*, 30(3):191–202.

[Back et al., 1994] Back, K., Yin, S., and Chappell, J. (1994). Expression of a plant sesquiterpene cyclase gene in *Escherichia coli*. *Archives of Biochemistry and Biophysics*, 315(2):527–532.

[Bäckmann et al., 2001] Bäckmann, A.-C., Bengtsson, M., Borg-Karlsson, A.-K., Liblikas, I., and Witzgall, P. (2001). Volatiles from apple (*Malus domestica*) eliciting antennal responses in female codling moth *Cydia pomonella* L.: effect of plant injury and sampling techniques. *Zeitschrift für Naturforschung*, 56c:262–268.

[Barton et al., 1999] Barton, D., Nakanishi, K., and Meth-Cohn, O. (1999). *Comprehensive Natural Products Chemistry, Volume 2: Isoprenoids including carotenoids and steroids*. Elsevier Science Ldt. Oxford.

[Bauer et al., 1990] Bauer, K., Garbe, D., and Surburg, H. (1990). *Common Fragrance and Flavour Materials: Preparation, Properties and Uses*. VCH Verlaggesellschaft Weinheim.

[Beechan and Djerassi, 1978] Beechan, C. M. . and Djerassi, C. (1978). The sesquiterpenes from the soft coral *Sinularia mayi*. *Tetrahedron*, 34:2503–2508.

[Bennett et al., 2002] Bennett, M. H., Mansfield, J. W., Lewis, M. J., and Beale, M. H. (2002). Cloning and expression of sesquiterpene synthase genes from lettuce *(Lactuca sativa* L.). *Phytochemistry*, 60(3):255–261.

[Bloch, 1992] Bloch, K. (1992). Sterol molecule: structure, biosynthesis and function. *Steroids*, 57:378–382.

[Bohlmann et al., 1998] Bohlmann, J., Meyer-Grauen, G., and Croteau, R. (1998). Plant terpenoid synthases: Molecular biology and phylogenetic analysis. *Proceedings of the National Acadamy of Science*, 95:4126–4133.

[Bouwmeester et al., 2002] Bouwmeester, H. J., Kodde, J., Verstappen, F. W. A., Altug, I. G., de Kraker, J.-W., and Wallaart, T. E. (2002). Isolation and characterization of two germacrene A synthase cDNA clones from chicory. *Plant Physiology*, 129:134–144.

[Buchanan et al., 2002] Buchanan, B., Gruissem, W., and Jones, R. (2002). *Biochemistry & Molecular Biology of Plants, Chapter 24: Natural Products*. John Wiley & Sons Inc, New York.

[Bülow and König, 2000] Bülow, N. and König, W. A. (2000). The role of germacrene D as a precursor in sesquiterpene biosynthesis: investigations of acid catalyzed, photochemically and thermally induced rearrangements. *Phytochemistry*, 55:141–168.

[Cane, 1990] Cane, D. E. (1990). Enzymatic formation of sesquiterpenes. *Chemical Reviews*, 90:1089–1103.

[Claude-Lafontaine et al., 1976] Claude-Lafontaine, A., Rouillard, M., Cassan, J., and Azzaro, M. (1976). Composés terpéniques inédits, constituants de l'huile essentielle de Cascarille. *Bull. Soc. Chim. Fr.*, 1(8):88–90.

[Colby et al., 1993] Colby, S., Alonso, W. R., Katahira, E. J., McGarvey, D. J., and Croteau, R. (1993). 4*S*-Limonene Synthase from the Oil Glands of Spearmint *(mentha spicata)*. *Journal of Biological Chemistry*, 268(31):23016–23024.

[Corbett and Speden, 1958] Corbett, R. E. and Speden, R. N. (1958). The volatile oil of *Pseudowintera colorata*: The structure of cyclocolorenone. *Journal of the Chemical Society*, pages 3710–3715.

[Davis and Croteau, 2000] Davis, E. M. and Croteau, R. (2000). Cyclization enzymes in the biosynthesis of monoterpenes, sesquiterpenes, and diterpenes. *Topics in Current Chemistry*, 209:53–95.

[de Kraker et al., 1998] de Kraker, J.-W., Franssen, M. C. R., de Groot, A., König, W. A., and Bouwmeester, H. J. (1998). (+)-Germacrene A biosynthesis: the committed step in the biosynthesis of sesquiterpene lactones in chicory. *Plant Physiology*, 117:1381–1392.

[Deligeorgopoulou and Allemann, 2003] Deligeorgopoulou, A. and Allemann, R. K. (2003). Evidence for Differential Folding of Farnesyl Pyrophosphate in the Active Site of Aristolochene Synthase: A Single-Point Mutation Converts Aristolochene Synthase into an *E*-β-Farnesene Synthase. *Biochemistry*, 42:7741–7747.

[Dudareva and Pichersky, 2000] Dudareva, N. and Pichersky, E. (2000). Biochemical and molecular genetic aspects of floral scents. *Plant Physiology*, 122:627–633.

[Eisenreich et al., 1999] Eisenreich, W., Rohdich, F., and Bacher, A. (1999). Deoxyxylulose phosphate pathway to terpenoids. *Trends in Plant Science*, 6(2):78–84.

[Eisenreich et al., 1998] Eisenreich, W., Schwarz, M., Cartayrade, A., Arigoni, D., Zenk, M. H., and Bacher, A. (1998). The deoxyxylulose phosphate pathway of terpenoid biosynthesis. *Chemistry & Biology*, 5:R221–R233.

[Ewing and Kuntz, 1997] Ewing, T. J. A. and Kuntz, I. (1997). Critical evaluation of search algorithms for automated molecular docking and database screening. *Journal of computational chemistry*, 18(9):1175–1189.
http://www.cmpharm.ucsfedu/kuntz/dock4/dock4.html.

[Facchini and Chappell, 1992] Facchini, P. J. and Chappell, J. (1992). Gene family of an elicitor-induced sesquiterpene cyclase in tobacco. *Proceedings of the National Acadamy of Science*, 89:11088–11092.

[Folkers and Wolf, 1956] Folkers, K. and Wolf, D. E. (1956). Isolation of a new acetate replacing factor. *Journal of the American Chemical Society*, 78:5273–5275.

[Friedel and Matusch, 1987] Friedel, H. D. and Matusch, R. (1987). Neue Aromadendran-Derivate aus Tolu-Balsam. *Helvetica Chimica Acta*, 70:1753–1759.

[Gibson and Pickett, 1983] Gibson, R. W. and Pickett, J. A. (1983). Wild potato repels aphids by release of aphid alarm pheromone. *Nature*, 302:608–609.

[Griffiths et al., 2000] Griffiths, A. J. F., Miller, J. H., Suzuki, D. T., Lewontin, R. C., and Gelbart, W. M. (2000). *Genetic Analysis*. W. H. Freeman and Company, New York, seventh edition.

[Guex and Peitsch, 1997] Guex, N. and Peitsch, M. C. (1997). SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modelling. *Electrophoresis*, 18:2714–2723.

[Hagedorn and Brown, 1991] Hagedorn, M. L. and Brown, S. M. (1991). The constituents of Cascarilla oil (*Croton eluteria* Benett). *Flavour and Fragrance Journal*, 6:193–204.

[Hänsel et al., 1994] Hänsel, R., Kelter, K., Rimpler, H., and Schneider, G. (1994). *Hagers Handbuch der Pharmazeutischen Praxis - Bd. 6: Drogen P-Z*. Number 6. Springer-Verlag, Berlin, 5. edition.

[Inoue et al., 1990] Inoue, H., Nojimba, H., and Okayama, H. (1990). High efficient transformation of *E. coli* with plasmids. *Gene*, 96:23–28.

[Jacyno et al., 1991] Jacyno, J., Montemurrop, N., and Bated, A. D. Cutler, H. G. (1991). Phytotoxic and antimicrobial properties of cyclocolorenone from *Magnolia grandiflora* L. *Journal of Agriculture and Food Chemistry*, 39:1166–1168.

[Kalberer et al., 2001] Kalberer, N. M., Turlings, T. C. D., and Rahier, M. (2001). Attraction of a leaf beetle *Oreina cacaliae* to damaged host plants. *Journal of Chemical Ecology*, 27(4):647–661.

[Kelvin, 1904] Kelvin, W. T. L. (1904). *Baltimore Lectures on Molecular Dynamics and the Wave Theory of Light*. C. J. Clay and Sons, Cambridge University Press Warehouse, London.

[Kleber et al., 1997] Kleber, H.-P., Schlee, D., and Schöpp, W. (1997). *Biochemisches Praktikum*. Gustav Fischer Verlag, Jena.

[König, 1998] König, W. A. (1998). Enantioselective capillary gas chromatography in the investigation of stereochemical correlations of terpenoids. *Chirality*, 10:499–504.

[König et al., 1996] König, W. A., Bülow, N., Fricke, C., Melching, S., Rieck, A., and Muhle, H. (1996). The sesquiterpene constituents of the liverwort *Pressia quadrata*. *Phytochemistry*, 43(3):629–633.

[Krepinsky and Herout, 1962] Krepinsky, J. and Herout, V. (1962). Isolation of terpenic compounds from *Solidago canadensis* L. *Coll. Czech. Chem. Commun.*, 27:2459–2462.

[Laemmli, 1970] Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, 227:680–685.

[Laule et al., 2003] Laule, O., Fürholz, A., Chang, H.-S., Zhu, T., Wang, X. Heifetz, P. B., Gruissem, W., and Lange, B. M. (2003). Crosstalk between cytosolic and plastidial pathways of isoprenoid biosynthesis in *Arabidopsis thaliana*. *Proceedings of the National Acadamy of Science*, 100(11):6866–6871.

[Leuschner, 1995] Leuschner, J. (1995). Anti-inflammatory, spasmolytic and diuretic effects of a commercially available *Solidago gigantea herb.* extract. *Arzneimittel-Forsch./Drug Research*, 45(2):165–168.

[Lichtenthaler, 1999] Lichtenthaler, H. K. (1999). The 1-deoxy-D-xylulose-5-phosphate pathway of isoprenoid biosynthesis in plants. *Annual Reviews of Plant Physiology and Plant Molecular Biology*, 50:47–65.

[Lois and West, 1990] Lois, A. F. and West, C. A. (1990). Regulation of expression of the casbene synthetase gene during elicitation of castor bean seedlings with pectic fragments. *Archives of Biochemistry and Biophysics*, 276:270–277.

[Lorimer and Weavers, 1987] Lorimer, S. D. and Weavers, R. T. (1987). Foliage sesquiterpenes and diterpenes of *Podocarpus spicatus*. *Phytochemistry*, 26(12):3207–3215.

[Lottspeich and Zorbas, 1998] Lottspeich, F. and Zorbas, H. (1998). *Bioanalytik*. Spektrum Akademischer Verlag, Heidelberg.

[MacGarvey and Croteau, 1995] MacGarvey, D. J. and Croteau, R. (1995). Terpenoid metabolism. *The Plant Cell*, 7:1015–1026.

[Martin et al., 2003] Martin, V. J. J., Pitera, D. J., Withers, S. T., Newman, J. D., and Keasling, J. D. (2003). Engineering a mevalonate pathway in *Escherichia coli* for production of terpenoids. *Nature Biotechnology*, 21(7):796–802.

[Mau and West, 1994] Mau, C. J. D. and West, C. A. (1994). Cloning of a casbene synthase cDNA: Evidence for conserved structural features among terpenoid cyclases in plants. *Proceedings of the National Acadamy of Science*, 91:8497–8501.

[McCaskill and Croteau, 1997] McCaskill, D. and Croteau, R. (1997). Prospects for the bioengineering of isoprenoid biosynthesis. *Advances in Biochemical Engineering / Biotechnology*, 55:108–141.

[Mozuraitis et al., 2002] Mozuraitis, R., Stranden, M., Ramirez, M. I., Borg-Karlson, A.-K., and Mustaparta, H. (2002). (-)-Germacrene D increases attraction and oviposition by the tobacco budworm moth *Heliothis virescens*. *Chemical Senses*, 27:505–509.

[Nishino et al., 1980] Nishino, C., Takayanagi, H., and Kimura, R. (1980). Electronantennogram and behavioral responses of the *American cockroach* sex pheromone mimics. *Comparative Biochemical Physiolology*, 66A:393–398.

[Nishino et al., 1977] Nishino, C., Tobin, T. R., and Bowers, W. S. (1977). Electronantennogram responses of the American cockroach to germacrene D sex pheromone mimic. *Journal of Insect Physiology*, 23:415–419.

[Niwa et al., 1980] Niwa, M., Iguchi, M., and Yamamura, S. (1980). Co-occurrence of (-)- and (+)-germacrene D in *Solidago altissima* L.: determination of the optical rotation of optically pure germacrene D. *Chem. Pharm. Bull.*, 28(3):997–999.

[Palmade et al., 1963] Palmade, M., Pesnelle, P., Streith, J., and Ourisson, G. (1963). L'α-Gurjunéne - Structure et Stéréochimie. *Bull. Soc. chim. Fr.*, (1):1950–1960.

[Peitsch, 1995] Peitsch, M. C. (1995). Protein modelling by E-Mail. *Bio/Technology*, 13:658–660.

[Peitsch, 1996] Peitsch, M. C. (1996). ProMod and Swiss-Model: Internet-based tools for automated comparative protein modelling. *Biochemical Society Transaction*, (24):274–279.

[Phillips et al., 2003] Phillips, M. A., Wildung, M. R., Williams, D. C., Hyatt, D. C., and Croteau, R. (2003). cdna isolation, functional expression, and characterization of (+)-α-pinene synthase and (-)-α-pinene synthase from loblolly pine (*Pinus taeda*): Stereocontrol in pinene biosynthesis. *Archives of Biochemistry and Biophysics*, (411):267–276.

[Pichersky et al., 1995] Pichersky, E., Lewinsohn, E., and Croteau, R. (1995). Purification and characterization of *S*-linalool synthase, an enzyme involved in the production of floral scent in *Clarkia breweri*. *Archives of Biochemistry and Biophysics*, 2(316):803–807.

[Prosser et al., 2003] Prosser, I., Altug, I. G., Lewis, M. J., Phillips, A. L., König, W. A., Bouwmeester, H. J., and Beale, M. H. (2003). Enantiospecific (+)- and (-)-germacrene D synthases, cloned from goldenrod, reveal a functionally active variant of the universal isoprenoid-biosynthesis apartate-rich motif. *Proceedings of the National Acadamy of Science*. In submission.

[Prosser et al., 2002] Prosser, I., Phillips, A. L., Gittings, S., Lewis, M. J., Hooper, A. M., Pickett, J. A., and Beale, M. H. (2002). (+)-(10*R*)-germacrene A synthase from goldenrod, *Solidago canadensis*; cDNA isolation, bacterial expression and functional analysis. *Phytochemistry*, 60:691–702.

[Reznicek et al., 1991] Reznicek, G., Jurenitsch, J., Plasun, M., Korhammer, S., Haslinger, E., and Hiller, K. amd Kubelka, W. (1991). Four major saponins from *Solidago canadensis*. *Phytochemistry*, 30(5):1629–1633.

[Rising et al., 2000] Rising, K. A., Starks, C. m., Noel, J. P., and Chappell, J. (2000). Demonstration of Germacrene A as an Intermediate in 5-*Epi*-aristolochene Synthase Catalysis. *Journal of the american chemical society*, 122:1861–1866.

[Rohdich et al., 2002] Rohdich, F., Hecht, S., Gärtner, K., Adam, P., Krieger, C., Amslinger, S., Arigoni, D., Bacher, A., and Eisenreich, W. (2002). Studies on the non-mevalonate pathway of terpene biosynthetic pathway: metabolic role of IspH (LytB) protein. *Proceedings of the National Acadamy of Science*, 99(3):1158–1163.

[Rohmer et al., 1996] Rohmer, M., Seemann, M., Horbach, S., Bringer-Meyer, S., and Sahm, H. (1996). Glyceraldehyde 3-phosphate and pyruvate as precursors of isoprenic units in an alternative non-mevalonate pathway for terpenoid biosynthesis. *Journal of the American Chemical Society*, 118:2564–2566.

[Rothmaler et al., 1991] Rothmaler, W., Jäger, E., Schubert, R., and Werner, K. (1991). *Exkursionsflora, Bd.2. Atlas der Gefäßpflanzen*. Volk und Wissen Verlag. 8. Edition.

[Rudi et al., 1998] Rudi, A., Dayan, L.-A., Aknin, M., Gaydou, E. M., and Kashman, Y. (1998). Several new isoprenoids from the soft coral *Sinularia erecta*. *Journal of Natural Products*, 61:872–875.

[Ruzicka et al., 1953] Ruzicka, L., Eschenmoser, A., and Heuser, H. (1953). The isoprene rule and the biogenesis of terpenic compounds. *Experientia*, 357(9):357–367.

[Ruzicka et al., 1923] Ruzicka, L., Pontalti, S., and Balas, F. (1923). Höhere Terpenverbindungen: Über die Sesquiterpenfraktion des Ysop-Öls, des Eucalyptus globulus-Öls, Gurjunbalsams und über das Guajol. *Helvetica Chimica Acta*, 6:855–865.

[Saitou and Nei, 1987] Saitou, N. and Nei, M. (1987). The Neighbor-joining Method: A New Method for Reconstructing Phylogenetic Trees. *Molecular and Biological Evolution*, 4:406–425.

[Sanger et al., 1970] Sanger, F., S., N., and Coulson, A. (1970). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Acadamy of Science*, 74:4298–5467.

[Schmid, 2002] Schmid, R. D. (2002). *Taschenatlas der Biotechnologie und Gentechnik*. WILEY-VCH Verlag, Weinheim.

[Schmidt, 1998] Schmidt, C. O. (1998). *Biosynthese der Sesquiterpene (+)-Germacren D, (-)-Germacren D und (-)-α-Gurjunen in Solidago canadensis*. PhD thesis, University of Hamburg.

[Schmidt et al., 1999a] Schmidt, C. O., Bouwmeester, H. J., Bülow, N., and König, W. A. (1999a). Isolation, characterization and mechanistic studies of (-)-α-gurjunene synthase from *Solidago canadensis*. *Archives of Biochemistry and Biophysics*, 364(2):167–177.

[Schmidt et al., 1998] Schmidt, C. O., Bouwmeester, H. J., de Kraker, J.-W., and Koenig, W. A. (1998). Biosynthesis of (+)-and (-)-Germacrene D in *Solidago canadensis*: Isolation and Characterization of two Enantioselective Germacrene D Synthases. *Angewandte Chemie International Edition*, 37:1400–1401.

[Schmidt et al., 1999b] Schmidt, C. O., Bouwmeester, H. J., Franke, S., and König, W. A. (1999b). Mechanism of the biosynthesis of sesquiterpene enantiomers (+)- and (-)-germacrene D in *Solidago canadensis*. *Chirality*, 11:353–362.

[Schwarz, 1994] Schwarz, M. K. (1994). *Terpen-Biosynthese in Ginkgo biloba: Eine überraschende Geschichte (Terpene biosynthesis: a surprising story)*. PhD thesis, ETH Zürich, Switzerland. Thesis Nr. 10951.

[Stace, 1991] Stace, C. A. (1991). *New flora of the british isles*. Cambridge University Press.

[Starks et al., 1997] Starks, C. M., Back, K., Chappell, J., and Noel, J. P. (1997). Structural basis for cyclic terpene biosynthesis by tobacco 5-*epi*-aristolochene synthase. *Science*, 277:1815–1820.

[Steliopoulos et al., 2002] Steliopoulos, P., Wüst, M., Adam, K.-P., and Mosandl, A. (2002). Biosynthesis of the sesquiterpene germacrene D in *Solidago canadensis*: ^{13}C and ^{2}H labeling studies. *Phytochemistry*, 60:13–20.

[Stoessl et al., 1976] Stoessl, A., Stothers, J. B., and Ward, E. W. B. (1976). Sesquiterpenoid stress compounds of the *Solanaceae*. *Phytochemistry*, 15:855–873.

[Stranden et al., 2002] Stranden, M., Borg-Karlson, A.-K., and Mustaparta, H. (2002). Receptor neuron discrimination of the germacrene D enantiomers in the moth *Helicoverpa armigera*. *Chemical Senses*, 27:143–152.

[Streith et al., 1962] Streith, J., Pesnelle, P., and Ourisson, G. (1962). L'α-Gurjurnéné: Structure at configuration en c-6 et en c-7. *Tetrahedron Letters*, 15:677–682.

[Tavormina et al., 1956] Tavormina, P. A., Gibbs, M. H., and Huff, J. W. (1956). The utilization of β-hydroxy-β-methyl-valerolactone in cholesterol biosynthesis. *Journal of the American Chemical Society*, 78:4498–4499.

[Trapp and Croteau, 2001] Trapp, S. C. and Croteau, R. B. (2001). Genomic organization of plant terpene synthases and molecular evolutionary implications. *Genetics*, 158:811–832.

[Treibs, 1952] Treibs, W. (1952). Die Dehydrierung von Hydroazulenen zu Azulenen mit Halogen. *Justus Liebigs Annalen der Chemie*, 576:110–115.

[Tsubaki et al., 1967] Tsubaki, N., Nishimura, K., and Hirose, Y. (1967). Hydrocarbons in patchouli oil. *Bull. Chem. Soc. Jap.*, 15:597–600.

[Tutin et al., 1976] Tutin, T. G., Heywood, V. H., Burges, N. A., Moore, D. M., Valentine, D. H., Walters, S. M., and Webb, D. A. (1976). *Flora Europaea Vol 4.: Plantaginaceae to Compositae*. Cambridge University Press.

[van Gelder et al., 2000] van Gelder, E., de Pauw, I., van Montagu, M., and van den Eeckhout, E. (2000). Cloning and molecular analysis of two new sesquiterpene cyclases from *Artemisia annua* L. *Planta*, 158:163–171.

[Wallaart et al., 2001] Wallaart, T. E., Bouwmeester, H. J., Hille, J., Poppinga, L., and Maijers, N. C. A. (2001). Amorpha-4,11-diene synthase: cloning and functional expression of a key enzyme in the biosynthetic pathway of the novel antimalarial drug artemisinin. *Planta*, 212:460–465.

[Wallach, 1887] Wallach, O. (1887). Zur Kenntnis der Terpene und ätherischen Öle, vierte Abhandlung. *Justus Liebigs Annalen der Chemie*, 238:78–88.

[Wallach, 1909] Wallach, O. (1909). *Terpene und Campher: Zusammenfassung eigener Untersuchungen auf dem Gebiet der alicyclischen Verbindungen*. von Veit und Comp., Leipzig.

[Weissbecker et al., 2000] Weissbecker, B., Van Loon, J. J. A., Posthumus, M. A., Bouwmeester, H. J., and Dicke, M. (2000). Identification of volatile potato sesquiterpenoids and their olfactory detection by the two-spotted stinkbug *Perillus bioculatus*. *Journal of chemical ecology*, 26:1433–1445.

[Weyerstahl et al., 2000] Weyerstahl, P., Marschall, H., Splittgerber, U., Wolf, D., and Surburg, H. (2000). Constituents of Haitian vetiver oil. *Flavour and Fragrance Journal*, 15:395–412.

[Weyerstahl et al., 1989] Weyerstahl, P., Marschall-Weyerstahl, H., and Manteuffel, E. (1989). *Proceedings of the 11th International Congress of Essential Oils, Fragrances and Flavours*. Number 5. Oxford & IBH Publishing Co PVT. Ltd, New Delhi.

[Yamasaki et al., 1997] Yamasaki, T., Sato, M., and Sakaguchi, H. (1997). (-)-germacrene D: masking substance of attractants for the cerambycid beetle, *Monochamus alternatus* (Hope). *Applied Entomological Zoology*, 32(3):423–429.

[Yoshihara et al., 1969] Yoshihara, K., Ohta, Y., Sakai, T., and Hirose, Y. (1969). Germacrene D, a key intermediate of cadinene group compounds. *Tetrahedron Letters*, 27:2263–2264.

Appendix A

Sequence data

A.1 Full-length sequence of fragment B

Full-length sequence with translated open reading frame of fragment B. The UTR (untranslated region) is shown in lower case, the ORF (open reading frame) in upper case and the translated protein sequence in 3-letter code.

```
atcaaaaccggatcgaggtttcactagaccaatccacttaaagaa 45

ATGGCCaCTGGTGAAGCTAGTACCAACCAGCTTGTTCGTCCTCTAGCCAATTTCCCTCCTTCCaTATGGGGTGATCACTTC 126
MetAlaThrGlyGluAlaSerThrAsnGlnLeuValArgProLeuAlaAsnPheProProSerIleTrpGlyAspHisPhe 27

CTATCTTTCTCTCATGACGATTCGAAATTGGAAGCATATGCTAAAGCCATGGAAGAACCAAAAGAAGATATAAGAAGATTG 207
LeuSerPheSerHisAspAspSerLysLeuGluAlaTyrAlaLysAlaMetGluGluProLysGluAspIleArgArgLeu 54

TTACTTAGCCCAACCATGGATTCAAATACAAAACTAAGATTGATTAATAGTGTGTATCGCCTTGGTTTGAAATATCTTTTC 288
LeuLeuSerProThrMetAspSerAsnThrLysLeuArgLeuIleAsnSerValTyrArgLeuGlyLeuLysTyrLeuPhe 81

CTTGAAGAGATCGAAGGTCAACTAGACAAACTTTTCAAGGATTTTCACTTGAAAGATTATGATGAAGCCGATCTTTACACA 369
LeuGluGluIleGluGlyGlnLeuAspLysLeuPheLysAspPheHisLeuLysAspTyrAspGluAlaAspLeuTyrThr 108

ATTTCGGTTAACTTCCAGGTTTTCAGACAACATGGTTATAAATTGTCTTGTGATGTGTTCAATAAGTTCAAGGATTGTCAC 450
IleSerValAsnPheGlnValPheArgGlnHisGlyTyrLysLeuSerCysAspValPheAsnLysPheLysAspCysHis 135

TCTGGTAAATTTAAGGAATATATTAAAAGTGACGTGAGGGCGATGTTAAGTTTCTACGAATCTACACAACTGAGAATAAGG 531
SerGlyLysPheLysGluTyrIleLysSerAspValArgAlaMetLeuSerPheTyrGluSerThrGlnLeuArgIleArg 162

GGAGAATCGGTGTTGGATGAAGCTTTCACATTCACTGAAACGCAACTTAAGGGTAGTGTAACGGACACTGATCTAGAAGGC 612
GlyGluSerValLeuAspGluAlaPheThrPheThrGluThrGlnLeuLysGlySerValThrAspThrAspLeuGluGly 189

AATCTTGCACGACAGGTGAAACACGCATTGGGGAGTCCTTTTCACAGAGGGATTCAGATAGTAGAGGCAAGATTATATTTC 693
AsnLeuAlaArgGlnValLysHisAlaLeuGlySerProPheHisArgGlyIleGlnIleValGluAlaArgLeuTyrPhe 216

TCGAACTATGAAGAAGAATGCTCTACATATGATTCCCTATCAAAGCTTGCAGTTGCACACTTCAATTACTTGCAACTATTG 774
```

```
SerAsnTyrGluGluGluCysSerThrTyrAspSerLeuSerLysLeuAlaValAlaHisPheAsnTyrLeuGlnLeuLeu 243

CACAAGAATGAACTTTATGTTCTCTCCAAGTGGTACAAGGACATGCAATTGAAAAACAGTTATCCTTTTGCAAGGGACAGA 855
HisLysAsnGluLeuTyrValLeuSerLysTrpTyrLysAspMetGlnLeuLysAsnSerTyrProPheAlaArgAspArg 270

GTACCAGAAATACACTTATGGATATTGGCAATATACTTTGAGCCACATTACTCTCAAGCCCGAATCATAACCTCAAAAATT 936
ValProGluIleHisLeuTrpIleLeuAlaIleTyrPheGluProHisTyrSerGlnAlaArgIleIleThrSerLysIle 297

GGTCTATTTGTGTCATTGTTAGATGACATTTTTGATGCATATGGTACTATTGACGAGCTTCGACTTCTAACTGATGCATTA 1017
GlyLeuPheValSerLeuLeuAspAspIlePheAspAlaTyrGlyThrIleAspGluLeuArgLeuLeuThrAspAlaLeu 324

AACAGGTGGGAAATTAGCGCTATGGAGCAACTTCCAGAATATATTAAACCATTTTACAAAATTGTCTTGTCTATGTATACT 1098
AsnArgTrpGluIleSerAlaMetGluGlnLeuProGluTyrIleLysProPheTyrLysIleValLeuSerMetTyrThr 351

GAACTTGAAGAACAACTATCTAAAGAAGGAAAAGCATATCTGGTTAATGCTTCAAAACAGGCGCTTCAAGAACTAACTAGA 1179
GluLeuGluGluGlnLeuSerLysGluGlyLysAlaTyrLeuValAsnAlaSerLysGlnAlaLeuGlnGluLeuThrArg 378

GCCGTACTATCaAAGAGGCTCAGTGGAGACATATTCGAATACATGCCATCATTTCAGGAGTATGTTAAGAATGGACTAATT 1260
AlaValLeuSerLysArgLeuSerGlyAspIlePheGluTyrMetProSerPheGlnGluTyrValLysAsnGlyLeuIle 405

ACATCTGCCTACGATGTCTTTGCGAGATCTGCTGTTATAGGTATGGGCAAGTTTGCCAGCGCAGATGTTTtCGATTGGtAT 1341
ThrSerAlaTyrAspValPheAlaArgSerAlaValIleGlyMetGlyLysPheAlaSerAlaAspValPheAspTrpTyr 432

GAAAGTCATCCCGAAATTCTTGAAGCTTCAAACTTAATTGCAAGACTTCACAATGATGTTGTGAGTTTCAAGTTTGAGGGT 1422
GluSerHisProGluIleLeuGluAlaSerAsnLeuIleAlaArgLeuHisAsnAspValValSerPheLysPheGluGly 459

GAAAGAGGATATCAAGTCACAGGCGTGGATGCATATATGAAGACTTTTGGGGTGCCAGAAAGTATAGCCAAGGAAGAAGTC 1503
GluArgGlyTyrGlnValThrGlyValAspAlaTyrMetLysThrPheGlyValProGluSerIleAlaLysGluGluVal 486

ATGAAAATTGTTGAAAAaTCATGGAAAGACATGAATGAGGGATATCTTAATAAGCCGACTGAAATTTCATTGGAGGTGCTT 1584
MetLysIleValGluLysSerTrpLysAspMetAsnGluGlyTyrLeuAsnLysProThrGluIleSerLeuGluValLeu 513

ACAACAATTGTTAATCTTGCACGAATAATAAACGTGGCATATAAGTATAATGATGCCTATACTTTTTCGGATGACACATTT 1665
ThrThrIleValAsnLeuAlaArgIleIleAsnValAlaTyrLysTyrAsnAspAlaTyrThrPheSerAspAspThrPhe 540

AAAGACTATATTACTCTCTTATTTATCGCTGACATCCCCATGGCATGTATGTAAtactcaataagtcggagtgattgtaat 1782
LysAspTyrIleThrLeuLeuPheIleAlaAspIleProMetAlaCysMetStop 558

gtatgttttgttcctaataatgaataactaactagtccttagactagtaggatccgtaaaaaaaaaaaaaaa 1854
```

A.2 Germacrene A fragment

Alignment of **frag17** = PCR generated fragment with primers designed on the GenBank published germacrene A sequence from *Solidago canadensis* and **Germ A** = germacrene A sequence from *Solidago canadensis* from position 375 of the ORF (GenBank accession AJ304452)

```
frag17 ATATCTACAAGCTTGATAATGGATCTTTCAAGGATTCCTTAACCAAGGATATTGAATGCATGCTTGAGTTATATGAGGCA 80
GermA ATATCTACAAGCTTGATAATGGATCTTTCAAGGATTCCTTAACCAAGGATATTGAATGCATGCTTGAGTTATATGAGGCA 455

frag17 GCCTATATGAGGGTGCAAGGCGAAATCATTCTAGATGAGGCTCTTGAGTTTACAAAAACTCATCTTGAACACATTGCAAA 160
GermA GCCTATATGAGGGTGCAAGGCGAAATCATTCTAGATGAGGCTCTTGAGTTTACAAAAACTCATCTTGAACACATTGCAAA 535
```

```
frag17 GGATCCACTTCGTTGCAACAACACGCTCTCTAGACACATACATGAAGCACTAGAGCGGCCTGTACAGAAAAGGTTGCCAA 240
GermA GGATCCACTTCGTTGCAACAACACGCTCTCTAGACACATACATGAAGCACTAGAGCGGCCTGTACAGAAAAGGTTGCCAA 615

frag17 GACTAGATGCAATACGATACATACCTTTCTATGAACAACAAGATTCTCACAACAAGTCCTTACTAAGACTTGCAAAGTTG 320
GermA GACTAGATGCAATACGATACATACCTTTCTATGAACAACAAGATTCTCACAACAAGTCCTTACTAAGACTTGCAAAGTTG 695

frag17 GGGTTCAACaGGCTTCAATCCTTGCATAAGAAGGAGCTTAGCCAACTTTCCAAATGGTGGAAAGAATTTGATGCTCCAAA 400
GermA GGGTTCAACcGGCTTCAATCCTTGCATAAGAAGGAGCTTAGCCAACTTTCCAAATGGTGGAAAGAATTTGATGCTCCAAA 775

frag17 GAATCTACCTTACGTAAGAGATAGATTGGTTGAACTCTACTTTTGGATACTAGGTGTCTACTTCGAACCTCAATATTCTC 480
GermA GAATCTACCTTACGTAAGAGATAGATTGGTTGAACTCTACTTTTGGATACTAGGTGTCTACTTCGAACCTCAATATTCTC 855

frag17 GTTCGAGAATATTCTTGACAAAAACAATTAAAATGGCAGCAATTCTAGACGACACGTATGATATCTATGGTACTTACGAA 560
GermA GTTCGAGAATATTCTTGACAAAAACAATTAAAATGGCAGCAATTCTAGACGACACGTATGATATCTATGGTACTTACGAA 935

frag17 GAACTTGAGATATTCACCAAAGCCGTTCAAAGGTGGTCAATTACCTGCATGGATACGCTTCCAGATTACATGAAAGTGAT 640
GermA GAACTTGAGATATTCACCAAAGCCGTTCAAAGGTGGTCAATTACCTGCATGGATACGCTTCCAGATTACATGAAAGTGAT 1015

frag17 TTATAAGAGCCTCTTGGATGTTTATGAAGAAATGGAGGAAATCATAGAAAAGGATGGAAAAGCATATCAAGTTCACTATG 720
GermA TTATAAGAGCCTCTTGGATGTTTATGAAGAAATGGAGGAAATCATAGAAAAGGATGGAAAAGCATATCAAGTTCACTATG 1095

frag17 CAAAAGAGTCGATGACAGATTTGGTTACAAGTTATATGACCGAAGCAAAATGGcTACATGAGGGTCATGTGCCAACATTT 800
GermA CAAAAGAGTCGATGATAGATTTGGTTACAAGTTATATGACCGAAGCAAAATGGtTACATGAGGGTCATGTGCCAACATTT 1175
........difference Thr/Ile
frag17 GACGAGCGTAACTCAGTTACAAACATAACTGGTG 834
GermA GACGAGCATAACTCAGTTACAAACATAACTGGTG 1209
difference Arg/His
```

A.3 Sequences of group 1 isolated from the cDNA library

Alignment of the three fragments L2.2, L3.1 and L4.2 of group 1 isolated from the cDNA library.

```
L2.2 ............................................................GATGTGTTTAACAAATTC
L3.1 ATCTTTATACAACTTCGGTTAACTTTCTTGTTTTTAGACAACATGGATATAAACTATCTTGTGATGTGTTTAACAAATTC 80
L4.2 ..........................................................................
.......LeuTyrThrThrSerValAsnPheLeuValPheArgGlnHisGlyTyrLysLeuSerCysAspValPheAsnLysPhe

L2.2 AAGGATGTTAGCACCGGCAAATTCAAGGAACACATTACAAGCGATGTGAAGGGGATGTTGAGCTTTTATGAATGTACACA
L3.1 AAGGATGTTAGCACCGGCAAATTCAAGGAACACATTACAAGCGATGTGAAGGGGATGTTGAGCTTTTATGAATGTACACA 160
L4.2 ..........................................................................
.....LysAspValSerThrGlyLysPheLysGluHisIleThrSerAspValLysGlyMetLeuSerPheTyrGluCysThrHi

L2.2 TTTAGGAATACGGGGCGAATCTATTTTAGATGAGGCCTTGGCATTCACAGAATCATATTTAAAGGGTGTGGTGGATACTC
L3.1 TTTAGGAATACGGGGCGAATCTATTTTAGATGAGGCCTTGGCATTCACAGAATCATATTTAAAGGGTGTGGTGGATACTC 240
L4.2 ..........................................................................
.....sLeuGlyIleArgGlyGluSerIleLeuAspGluAlaLeuAlaPheThrGluSerTyrLeuLysGlyValValAspThrL

L2.2 TTGAAGGGACTCTCGCACAACAGGTGAAACAAGGGCTGAAATTTCCTTGTCAACGCGGATTGCCGATAGTAGAGGCAAGG
L3.1 TTGAAGGGACTCTCGCACAACAGGTGAAACAAGGGCTGAAATTTCCTTGTCAACGCGGATTGCCGATAGTAGAGGCAAGG 320
L4.2 ..........................................................................
.....euGluGlyThrLeuAlaGlnGlnValLysGlnGlyLeuLysPheProCysGlnArgGlyLeuProIleValGluAlaArg

L2.2 CTATATTTCTCCAACTATGAACAAGAATGTTCAGCATACGATCCACTACCGAAGCTTGCAAAAGCACATTTCAGTTACTT
L3.1 CTATATTTCTCCAACTATGAACAAGAATGTTCAGCATACGATCCACTACCGAAGCTTGCAAAAGCACATTTCAGTTACTT 400
```

```
L4.2 ..........................................................................
.....LeuTyrPheSerAsnTyrGluGlnGluCysSerAlaTyrAspProLeuProLysLeuAlaLysAlaHisPheSerTyrPh

L2.2 TCAATTAATGCAAAAGGATGAACTGTCCAAACTCACACAATGGTCTAAGGACATGAATTTCCAAACAATAGCTACATATA
L3.1 TCAATTAATGCAAAAGGATGAACTGTCCAAACTCACACAATGGTCTAAGGACATGAATTTCCAAACAATAGCTACATATA 480
L4.2 ..........................................................................
.....eGlnLeuMetGlnLysAspGluLeuSerLysLeuThrGlnTrpSerLysAspMetAsnPheGlnThrIleAlaThrTyrT

L2.2 CAAGGGATAAAATGCCAGAATTGTACCTATGGGTGTTAGCAGTATTCTTGGAGCCTCGTCACGTTGAAGCCAGATTTATA
L3.1 CAAGGGATAAAATGCCAGAATTGTACCTATGGGTGTTAGCAGTATTCTTGGAGCCTC.....................  560
4.2 ...........................................................TGAAGCCAGATTTATA
.....hrArgAspLysMetProGluLeuTyrLeuTrpValLeuAlaValPheLeuGluProArgHisValGluAlaArgPheIle

L2.2 ..........................................................................
L3.1 ..........................................................................
L4.2 ACTACAAAAGTTGCACAACTGGTGTTGGTGTTAGATGACACATTTGATGCATATGCAACTATTGAGGAGCTTCGACTTCT 640
.....ThrThrLysValAlaGlnLeuValLeuValLeuAspAspThrPheAspAlaTyrAlaThrIleGluGluLeuArgLeuLe

L2.2 ..........................................................................
L3.1 ..........................................................................
L4.2 AACGGATGCCATAAGCAGATGGGAAATAAGTTGTATGGAGCAACTTCCAGAGTATATTAAACCATTTTACCAAATTATCC 720
.....uThrAspAlaIleSerArgTrpGluIleSerCysMetGluGlnLeuProGluTyrIleLysProPheTyrGlnIleIleL

L2.2 ..........................................................................
L3.1 ..........................................................................
L4.2 TCAATGAGTATGCTGAATGGGAGAAACAACTGGCTAAAGAAGGAAGGGAAAATGTGGTTTATGCTTCAAAAAAAGCATTT 800
.....euAsnGluTyrAlaGluTrpGluLysGlnLeuAlaLysGluGlyArgGluAsnValValTyrAlaSerLysLysAlaPhe

L2.2 ..........................................................................
L3.1 ..........................................................................
L4.2 CAAGAACTTGCCAGAGCCTATCTTAGGGAGGCTGAGTGGAGACATAGTGGAACTGTACCATCCTTTCAAGAGTACTATGA 880
GlnGluLeuAlaArgAlaTyrLeuArgGluAlaGluTrpArgHisSerGlyThrValProSerPheGlnGluTyrTyrGl

L2.2 .....................................CTTGTTTAATAGGCATGGGCAAAATTGTTGATGAAGAAG
L3.1 ..........................................................................
L4.2 GAATGGATTGGCAACTTCTACCTACAATCTTCTTGGAAAATCTTGTTTAATAGGCATGGGCAAAATTGTTGATGAAGAAG 960
.....uAsnGlyLeuAlaThrSerThrTyrAsnLeuLeuGlyLysSerCysLeuIleGlyMetGlyLysIleValAspGluGluA

L2.2 CTTTGGCTTGGTATGATAGTCATCCAAAAATACTTGAAGCCTCTGAACTAATTGCAAGACTCCAcAATGATGTTGTCAGC
L3.1 CTTTGGCTTGGTATGATAGTCATCCAAAAATACTTGAAGCCTCTGAACTAATTGCAAGACTCCAtAATGATGTTGTCAGC
L4.2 CTTTGGCTTGGTATGATAGTCATCCAAAAATACTTGAAGCCTCTGAACTAATTGC.....................  1040
.....laLeuAlaTrpTyrAspSerHisProLysIleLeuGluAlaSerGluLeuIleAlaArgLeuHisAsnAspValValSer

L2.2 TTTGAGTTCGAGCGTGAAAGAGAACATAGAGCTACAGGTATAGATGCATATATGAAGACTTTTGGAGTGACAGAAGATGT 1120
L3.1 TTTGAGTTCGAGCGTGAAAGAGAACATAGAGCTACAGGTATAGATGCATATATGAAGACTTTTGGAGTGACAGAAGATGT
L4.2 TTTGAGTTCGAGCGTGAAAGAGAACATAGAGCTACAGGTATAGATGCATATATGAAGACTTTTGGAGTGACAGAAGATGT
.....PheGluPheGluArgGluArgGluHisArgAlaThrGlyIleAspAlaTyrMetLysThrPheGlyValThrGluAspVa

L2.2 AGCTGTCAAAGAaCTCAAGGAAATGATTGAAAATGCATGGAAAGACATAAACGAGGGATGTCTAAAGCCAACCAAAGTGT 1200
L3.1 AGCTGTCAAAGAgCTCAAGGAAATGATTGAAAATGCATGGAAAGACATAAACGAGGGATGTCTAAAGCCAACCAAAGTGT
L4.2 AGCTGTCAAAGAgCTCAAGGAAATGATTGAAAATGCATGGAAAGACATAAACGAGGGATGTCTAAAGCCAACCAAAGTGT
.....lAlaValLysGluLeuLysGluMetIleGluAsnAlaTrpLysAspIleAsnGluGlyCysLeuLysProThrLysValS

L2.2 CGATGGATTTGCTTTATCCAATTGTTAATCTTTCACGGGTGATATATGTGGCCTACAGGTTCAACGATGGGTTCACCTTC 1280
L3.1 CGATGGATTTGCTTTATCCAATTGTTAATCTTTCACGGGTGATATATGTGGCCTACAGGTTCAACGATGGGTTCACCTTC
L4.2 CGATGGATTTGCTTTATCCAATTGTTAATCTTTCACGGGTGATATATGTGGCCTACAGGTTCAACGATGGGTTCACCTTC
.....erMetAspLeuLeuTyrProIleValAsnLeuSerArgValIleTyrValAlaTyrArgPheAsnAspGlyPheThrPhe

L2.2 TCAGACtTGACCTTGAAAGACTATATTAGTCTCTTATTCGAGGCTTCTGTCCCtGTgTaATtgATCATTtAACCaTcCGAT 1360
L3.1 TCAGACcTGACCTTGAAAGACTATATTAGTCTCTTATTCGAGGCTTCTGTCCCcGTaTgATccATCATaAACCgTtCGAT
L4.2 TCAGACcTGACCTTGAAAGACTATATTAGTCTCTTATTCGAGGCTTCTGTCCCcGTaTgATccATCATaAACCgTtCGAT
.....SerAspLeuThrLeuLysAspTyrIleSerLeuLeuPheGluAlaSerValProValSTOP

L2.2 CGATcCAATATTTTCTATTATTATCATGTCGTTTGTAAGAC....ATTTGCCAATAAAGGAATGTTGTAT.....AAAAAAAAA 1440
L3.1 CGATaCAATATTTTCTATAATTATCATGTCGTTTGTAAGACATTAATTTGCCAATAAAGGAATGTTGTATAAGATAAAAAAAAA
```

```
L4.2 CGATaCAATATTTTCTATAATTATCATGTCGTTTGTAAGACATTAATTTGCCAATAAAGGAATGTTGTATAAGATAAAAAAAAA
```

A.4 Full-length sequence of cascarilladiene-synthase (group 1)

The full-length sequence of group 1 was identified as cascarilladiene-synthase. The UTR (untranslated region) is shown in lower case, the ORF (open reading frame) in upper case and the translated protein sequence in 3-letter code.

```
gtacccgggaaatcggcccattacggccggggatcacacattncatttttaacaactttaattattagtctcaaacaaacc 81
tttaaacactttaatttgattcaagaATGGCCATGGTTGACGCCAATACTACTACTGTCCAGCAAGGAGAGACCAAAACC 80
.......................MetAlaMetValAspAlaAsnThrThrThrValGlnGlnGlyGluThrLysThr 18

ACCATCGTGCGTCCTATTGCTAAGTTCCCTCCTTCGTTATGGGGTGATCGTTTCCTATCATTCGACCTCGACAACTTGGAA 161
ThrIleValArgProIleAlaLysPheProProSerLeuTrpGlyAspArgPheLeuSerPheAspLeuAspAsnLeuGlu 45

TTGGATGATTGCGTTGCAGCCATGAAAGAACCAAGTGATGAGATACAAAGATTAATCGTTGATCCGGACATGGATTCAAAT 242
LeuAspAspCysValAlaAlaMetLysGluProSerAspGluIleGlnArgLeuIleValAspProAspMetAspSerAsn 72

GAAAAACTACGTTTAATTAACTGCGTGTGTCGTCTTGGTTTAAGCTATCTTTTCATAAAAGATATTGAATGTCAACTCGAA 323
GluLysLeuArgLeuIleAsnCysValCysArgLeuGlyLeuSerTyrLeuPheIleLysAspIleGluCysGlnLeuGlu 99

AAACTTTTCAAGGAGCTTAATATGGAAGAGTATAATGAATTTGATCTTTATACAACTTCGGTTAACTTTCTTGTTTTTAGA 404
LysLeuPheLysGluLeuAsnMetGluGluTyrAsnGluPheAspLeuTyrThrThrSerValAsnPheLeuValPheArg 126

CAACATGGATATAAACTATCTTGTGATGTGTTTAACAAATTCAAGGATGTTAGCACCGGCAAATTCAAGGAACACATTACA 485
GlnHisGlyTyrLysLeuSerCysAspValPheAsnLysPheLysAspValSerThrGlyLysPheLysGluHisIleThr 153

AGCGATGTGAAGGGGATGTTGAGCTTTTATGAATGTACACATTTAGGAATACGGGGCGAATCTATTTTGGATGAGGCCTTG 566
SerAspValLysGlyMetLeuSerPheTyrGluCysThrHisLeuGlyIleArgGlyGluSerIleLeuAspGluAlaLeu 180

GCATTCACAGAATCATATTTAAAGGGTGTGGTGGATACACTTGAAGGGACTCTCGCACAACAAGTGAAACAAGGGCTGAAA 647
AlaPheThrGluSerTyrLeuLysGlyValValAspThrLeuGluGlyThrLeuAlaGlnGlnValLysGlnGlyLeuLys 207

TTTCCTTGTCAACGCGGATTGCCGATAGTAGAGGCAAGGCTATATTTCTCCAACTATGAACAAGAATGTTCAGCATACGAT 728
PheProCysGlnArgGlyLeuProIleValGluAlaArgLeuTyrPheSerAsnTyrGluGlnGluCysSerAlaTyrAsp 234

CCACTACCGAAGCTTGCAAAAGCACATTTCAGTTACTTTCAATTAATGCAAAAGGATGAACTGTCCAAACTCACACAATGG 809
ProLeuProLysLeuAlaLysAlaHisPheSerTyrPheGlnLeuMetGlnLysAspGluLeuSerLysLeuThrGlnTrp 251

TCTAAGGACATGAATTTCCAAACAATAGCTACATATACAAGGGATAAAATGCCAGAATTGTACCTATGGGTGTTAGCAGTA 890
SerLysAspMetAsnPheGlnThrIleAlaThrTyrThrArgAspLysMetProGluLeuTyrLeuTrpValLeuAlaVal 288

TTCTTGGAGCCTCGTCACGTTGAAGCCAGATTTATAACTACAAAAGTTGCACAACTGGTGTTGGTGTTAGATGACACATTT 971
PheLeuGluProArgHisValGluAlaArgPheIleThrThrLysValAlaGlnLeuValLeuValLeuAspAspThrPhe 315

GATGCATATGCAACTATTGAGGAGCTTCGACTTCTAACGGATGCCATAAGCAGATGGGAAATAAGTTGTATGGAGCAACTT 1052
AspAlaTyrAlaThrIleGluGluLeuArgLeuLeuThrAspAlaIleSerArgTrpGluIleSerCysMetGluGlnLeu 342

CCAGAGTATATTAAACCATTTTACCAAATTATCCTCAATGAGTATGCTGAATGGGAGAAACAACTGGCTAAAGAAGGAAGG 1133
ProGluTyrIleLysProPheTyrGlnIleIleLeuAsnGluTyrAlaGluTrpGluLysGlnLeuAlaLysGluGlyArg 369
```

```
GAAAATGTGGTTTATGCTTCAAAAAAAGCATTTCAAGAACTTGCCAGAGCCTATCTTAGGGAGGCTGAGTGGAGACATAGT 1214
GluAsnValValTyrAlaSerLysLysAlaPheGlnGluLeuAlaArgAlaTyrLeuArgGluAlaGluTrpArgHisSer 396

GGAACTGTACCATCCTTTCAAGAGTACTATGAGAATGGATTGGCAACTTCTACCTACAATCTTCTTGGAAAATCTTGTTTA 1295
GlyThrValProSerPheGlnGluTyrTyrGluAsnGlyLeuAlaThrSerThrTyrAsnLeuLeuGlyLysSerCysLeu 423

ATAGGCATGGGCAAAATTGTTGATGAAGAAGCTTTGGCTTGGTATGATAGTCATCCAAAAATACTTGAAGCCTCTGAACTA 1376
IleGlyMetGlyLysIleValAspGluGluAlaLeuAlaTrpTyrAspSerHisProLysIleLeuGluAlaSerGluLeu 450

ATTGCAAGACTCCAcAATGATGTTGTCAGCTTTGAGTTCGAGCGTGAAAGAGAACATAGAGCTACAGGTATAGATGCATAT 1457
IleAlaArgLeuHisAsnAspValValSerPheGluPheGluArgGluArgGluHisArgAlaThrGlyIleAspAlaTyr 477

ATGAAGACTTTTGGAGTGACAGAAGATGTAGCTGTCAAAGAaCTCAAGGAAATGATTGAAAATGCATGGAAAGACATAAAC 1538
MetLysThrPheGlyValThrGluAspValAlaValLysGluLeuLysGluMetIleGluAsnAlaTrpLysAspIleAsn 504

GAGGGATGTCTAAAGCCAACCAAAGTGTCGATGGATTTGCTTTATCCAATTGTTAATCTTTCACGGGTGATATATGTGGCC 1619
GluGlyCysLeuLysProThrLysValSerMetAspLeuLeuTyrProIleValAsnLeuSerArgValIleTyrValAla 531

TACAGGTTCAACGATGGGTTCACCTTCTCAGACcTGACCTTGAAAGACTATATTAGTCTCTTATTCGAGGCTTCTGTCCCC 1700
TyrArgPheAsnAspGlyPheThrPheSerAspLeuThrLeuLysAspTyrIleSerLeuLeuPheGluAlaSerValPro 558

GTATGAtccatcataaaccgttcgatcgatacaatattttctataattatcatgtcgtttgtaagacattaatttgccaat 1781
ValSTOP 560

aaaggaatgttgtataagataaaaaaaaaa 1823
```

A.5 Sequences of group 2 isolated from cDNA library

A.5.1 cDNA sequence of (+)-germacrene D-synthase

The cDNA sequence 2a was identified as (+)-germacrene D-synthase. The UTR (untranslated region) is shown in lower case, the ORF (open reading frame) in upper case and the translated protein sequence in 3-letter code.

```
atcacaaaattttaatcaacataagaatctaagtccatataaaac 45

ATGGCTGCTAAACATGTAGAGGCTATTCGCCCAGCTGCAAACTATCAATATCATCCAAGCCTTTGGGGAGATCAGTTTCTA 126
MetAlaAlaLysHisValGluAlaIleArgProAlaAlaAsnTyrGlnTyrHisProSerLeuTrpGlyAspGlnPheLeu 27

CACTATGATGAGCGAGAGGATGAGCACGTTGAGGTAGATCAACAAATCGAAATTTTGAAGGAAGAAACACGAAAAGAAATA 207
HisTyrAspGluArgGluAspGluHisValGluValAspGlnGlnIleGluIleLeuLysGluGluThrArgLysGluIle 54

CTTGCAAGTTTGGATGATCCAGCAAAACATACAAATTTGCTGAAGCTGATTGATGTTATCCAACGTCTCGGTATAGCCTAC 288
LeuAlaSerLeuAspAspProAlaLysHisThrAsnLeuLeuLysLeuIleAspValIleGlnArgLeuGlyIleAlaTyr 81

TACTTTGAACATGAGATTACGCAAGCATTGGACCATATTTATAATGTATACGGTGACGAATGGAATGGTGGAAGTACTTCC 369
TyrPheGluHisGluIleThrGlnAlaLeuAspHisIleTyrAsnValTyrGlyAspGluTrpAsnGlyGlySerThrSer 108

CTTTGGTTTCGGCTCCTCCGACAACAAGGCTTTTACGTTTCATGTGATATTTTCAATATATACAAGCTTGATAATGGATCT 450
LeuTrpPheArgLeuLeuArgGlnGlnGlyPheTyrValSerCysAspIlePheAsnIleTyrLysLeuAspAsnGlySer 135

TTCAAGGATTCCTTAACCAAGGATATTGAATGCATGCTTGAGTTATATGAGGCAGCCTATATGAGGGTGCAAGGCGAAATC 531
```

```
PheLysAspSerLeuThrLysAspIleGluCysMetLeuGluLeuTyrGluAlaAlaTyrMetArgValGlnGlyGluIle 162

ATTCTAGATGAGGCTCTTGAGTTTACAAAAACTCATCTTGAACAAATTGCGAAGGATCCCCTTCGTTGCAACAACACGCTC 612
IleLeuAspGluAlaLeuGluPheThrLysThrHisLeuGluGlnIleAlaLysAspProLeuArgCysAsnAsnThrLeu 189

TCTAGACACATACATGAAGCACTAGAGCTGCCTGTACAGAAAAGGTTGCCAAGACTAGATGCAATACGATACATACCTTTC 693
SerArgHisIleHisGluAlaLeuGluLeuProValGlnLysArgLeuProArgLeuAspAlaIleArgTyrIleProPhe 216

TATGAACAACAAGATTCTCACAACAAGTCCTTACTGAGACTTGCAAAGTTGGGGTTCAACCGGCTTCAATCCTTGCATAAG 774
TyrGluGlnGlnAspSerHisAsnLysSerLeuLeuArgLeuAlaLysLeuGlyPheAsnArgLeuGlnSerLeuHisLys 243

AAGGAGCTTAGCCAACTTTCCAAATGGTGGAAAGAATTTGATGCTCCAAAGAATGTACCTTACGCAAGAGATAGATTGGTT 855
LysGluLeuSerGlnLeuSerLysTrpTrpLysGluPheAspAlaProLysAsnValProTyrAlaArgAspArgLeuVal 270

GAACACTACTTTTGGATACTAGGTGTCTACTTCGAACCCCAATaTTCTCGTTCTAGAATATTCTTGGCAAAAATAATTACA 936
GluHisTyrPheTrpIleLeuGlyValTyrPheGluProGlnTyrSerArgSerArgIlePheLeuAlaLysIleIleThr 297

ATGACAGCAATTCTAAACGACACGTATGATATCTATGGTACTTACGAAGAACTTGAGATATTCACCAAAGCCCTTCAAACG 1017
MetThrAlaIleLeuAsnAspThrTyrAspIleTyrGlyThrTyrGluGluLeuGluIlePheThrLysAlaLeuGlnThr 324

TGGTCAACCTGCATGGATACGTTTCCAGATTACATGAAAGTGATTTATAAGAGCCTCTTGGATATTTATGAAGAAATGGAG 1098
TrpSerThrCysMetAspThrPheProAspTyrMetLysValIleTyrLysSerLeuLeuAspIleTyrGluGluMetGlu 351

GAAATCATGGAAAAGAATGGAAAAGCATATCAAGTTGACTATGCAAAAGAGGCGATGAGAGAGTTGCTTACAGGCGGTTAT 1179
GluIleMetGluLysAsnGlyLysAlaTyrGlnValAspTyrAlaLysGluAlaMetArgGluLeuLeuThrGlyGlyTyr 378

ATGGCCGAAGCGAAATTGTTACATGAGGGTCATGTGCCAACATTTGAGGAGCATAACAAAATTACAAACTTAAGTGCtGGC 1260
MetAlaGluAlaLysLeuLeuHisGluGlyHisValProThrPheGluGluHisAsnLysIleThrAsnLeuSerAlaGly 405

CATAAAATGCTTTCAACATCAAGCTTTGTTGGCATGCCTGGTGATATAGTTACACAAGATTCTTTCAAATGGGCTCTCAAC 1341
HisLysMetLeuSerThrSerSerPheValGlyMetProGlyAspIleValThrGlnAspSerPheLysTrpAlaLeuAsn 432

AATCCTCGACTTATAACAGCTTCAGCTTACATTGGTAGGATTTTGAATGATATCGTCGGTCACAAGGAGGAGCAGCAAAGA 1422
AsnProArgLeuIleThrAlaSerAlaTyrIleGlyArgIleLeuAsnAspIleValGlyHisLysGluGluGlnGlnArg 459

AAGCATATTCCATCTATTGTGGAAATGTACATGAAAGAACATAATCTTTTGAGGGAGGACGTTTATGATTTGCTCAAACAA 1503
LysHisIleProSerIleValGluMetTyrMetLysGluHisAsnLeuLeuArgGluAspValTyrAspLeuLeuLysGln 486

AGAGTTGAAGATGCATGGAAAGATCTAAACCGAGAAACATTAACATGTAAAGACATTCATATGTCTCTTAAGATGTGTCCG 1584
ArgValGluAspAlaTrpLysAspLeuAsnArgGluThrLeuThrCysLysAspIleHisMetSerLeuLysMetCysPro 513

ATAAACCTCGCACGCGTAGAACATAAGATATACAAAAACGGTGATAACTTAAAATTTGTTGGACAAGAAATCCAAGATTAT 1665
IleAsnLeuAlaArgValGluHisLysIleTyrLysAsnGlyAspAsnLeuLysPheValGlyGlnGluIleGlnAspTyr 540

ATCAAATCTTGTTTCATTAATGCTATGAGTGTTTGA 1701
IleLysSerCysPheIleAsnAlaMetSerValstop 551

tttttggtccttcatgaccaattatcacgtctcatagatctgaaagatgttggttcgcactttgaaagatgcaaggttat 1782
ggtacccccctaaataagacctcaagtccgccggagctataaaactatatatacgtattgtatttgtggattttaatttatg 1863
tattgtatttcatgtatatgtggatttcaatttatgtattgtatttcataaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1943
```

A.5.2 cDNA sequence of (–)-germacrene D-synthase

The cDNA sequence 2b was identified as (–)-germacrene D-synthase. The UTR is shown in lower case, the ORF in upper case and the translated protein sequence in 3-letter code.

```
aaattttaatcagcatcataagaatctaagtccatataaaac 42

ATGGCTGCTAAACAAGGAGAGGTTGTTCGCCCAGATGCAGACTACTCGTATCATCCAAGCCTTTGGGGAGATCAGTTTCTA 123
MetAlaAlaLysGlnGlyGluValValArgProAspAlaAspTyrSerTyrHisProSerLeuTrpGlyAspGlnPheLeu 27

CACTATGATGAGCAAGAGGATGATCAAGTTGAGGTAGATCAACAAATCGAAATTTTGAAGGAAGAAACGAGAAAAGAAATA 204
HisTyrAspGluGlnGluAspAspGlnValGluValAspGlnGlnIleGluIleLeuLysGluGluThrArgLysGluIle 54

CTCTCAAGTTTGGATGATCCAGCAAAACATACAAATTTGCTGAAGCTGATTGATGTTATCCAACGTCTCGGTATAGCCTAC 285
LeuSerSerLeuAspAspProAlaLysHisThrAsnLeuLeuLysLeuIleAspValIleGlnArgLeuGlyIleAlaTyr 81

TATTTTGAACATGAGATTACGCAAGCATTGGGCCATATTTATAATGTATACGGTGATGAATGGAATGGTGGTAGTACTTCC 366
TyrPheGluHisGluIleThrGlnAlaLeuGlyHisIleTyrAsnValTyrGlyAspGluTrpAsnGlyGlySerThrSer 108

CTTTGGTTTCGGCTCCTCCGACAACAAGGCTTTTATGTTTCATGTGATATTTTCAATATCTACAAGCTTGATAATGGATCT 447
LeuTrpPheArgLeuLeuArgGlnGlnGlyPheTyrValSerCysAspIlePheAsnIleTyrLysLeuAspAsnGlySer 135

TTCAAGGATTCCTTAACCAAGGATATTGAATGCATGCTTGAGTTATATGAGGCAGCCTATATGAGGGTGCAAGGCGAAATC 528
PheLysAspSerLeuThrLysAspIleGluCysMetLeuGluLeuTyrGluAlaAlaTyrMetArgValGlnGlyGluIle 162

ATTCTAGATGAGGCTCTTGAGTTTACAAAAACTCATCTTGAACAAATTGCAAAGGATCCCCTTCGTTGCAACAACACGCTC 609
IleLeuAspGluAlaLeuGluPheThrLysThrHisLeuGluGlnIleAlaLysAspProLeuArgCysAsnAsnThrLeu 189

TCTAGACACATATATGAAGCGCTGAAGCGGCCTATACGGAAAAGGCTACCAAGAGTGGATGCATTACAATACATGCCTTTC 690
SerArgHisIleTyrGluAlaLeuLysArgProIleArgLysArgLeuProArgValAspAlaLeuGlnTyrMetProPhe 216

TATGAACAACAAGATTCACACAACAAGTCCTTACTAAGACTTGCAAAGTTGGGGTTCAACCGGCTTCAATCCTTGCATAAG 771
TyrGluGlnGlnAspSerHisAsnLysSerLeuLeuArgLeuAlaLysLeuGlyPheAsnArgLeuGlnSerLeuHisLys 243

AAGGAGCTTAGCCAACTTTCCAAATGGTGGAAAGAATTTGATGCTCCAAAGAATCTACGTTACGTAAGAGATAGATTGGTT 852
LysGluLeuSerGlnLeuSerLysTrpTrpLysGluPheAspAlaProLysAsnLeuArgTyrValArgAspArgLeuVal 270

GAACTCTACTTTTGGGTACTAGGTGTCTACTTCGAACCTCAATATTCTCGTTCAAGAATATTCTTGACGAAGGTAATTAAA 933
GluLeuTyrPheTrpValLeuGlyValTyrPheGluProGlnTyrSerArgSerArgIlePheLeuThrLysValIleLys 297

ATGGCCACAATTCTAGACGACACGTATGATATCCATGGCACTTACGAAGAACTTGAGATATTCACCAAAGCCGTTCAAAGG 1014
MetAlaThrIleLeuAspAspThrTyrAspIleHisGlyThrTyrGluGluLeuGluIlePheThrLysAlaValGlnArg 324

TGGTCAATTACCTGCATGGATACGCTTCCAGATTACATGAAAATGATTTATAAGAGCCTCTTGGATGTTTATGAAGAAATG 1095
TrpSerIleThrCysMetAspThrLeuProAspTyrMetLysMetIleTyrLysSerLeuLeuAspValTyrGluGluMet 351

GAGGAAATCATAGACAAGGATGGAAAAGCATATCAAGTTCACTATGCAAAAGACTCGATGATAGATTTGGTTACAAGTTAT 1176
GluGluIleIleAspLysAspGlyLysAlaTyrGlnValHisTyrAlaLysAspSerMetIleAspLeuValThrSerTyr 378

ATGACCGAAGCGAAATGGTTACATGAGGGTCATGTGCCAACATTTGAGGAGTATAACTCAATTACAAACTTAACTGGTGGC 1257
MetThrGluAlaLysTrpLeuHisGluGlyHisValProThrPheGluGluTyrAsnSerIleThrAsnLeuThrGlyGly 405

TATAAAATGCTTACAACATCAAGTTTTGTTGACATGCCTGGTGATATAGTTACACAAGAGTCTTTCAAATGGGCTCTTAAC 1338
TyrLysMetLeuThrThrSerSerPheValAspMetProGlyAspIleValThrGlnGluSerPheLysTrpAlaLeuAsn 432

AATCCTCCACTTATAAAAGCTTCAGCTGACGTTAGTAGGATTATGGATGATATCGTCGGGCACAAGGAGGAGCAACAAAGA 1419
AsnProProLeuIleLysAlaSerAlaAspValSerArgIleMetAspAspIleValGlyHisLysGluGluGlnGlnArg 459

AAGCATCTTcCATCTAGGGTGGAAAtGTACATGAAAAAATATCATCTTGCGGAGGAGGACGTCTATGATTTGCTCAAACAA 1500
LysHisLeuProSerArgValGluMetTyrMetLysLysTyrHisLeuAlaGluGluAspValTyrAspLeuLeuLysGln 486

AGAGTTGAAGATGCATGGAAAGATCTAAACCGAGAAACATTAACATGTAAAGACATTCATATGGCTCTTAAGATGCGTCCG 1581
ArgValGluAspAlaTrpLysAspLeuAsnArgGluThrLeuThrCysLysAspIleHisMetAlaLeuLysMetArgPro 513

ATCAACCTGGCACGCGTAATAGATATGCTATACAAAAACGATGATAACTTAAAAAATGTTGGACAAGAAATCCAAGATTAT 1662
IleAsnLeuAlaArgValIleAspMetLeuTyrLysAsnAspAspAsnLeuLysAsnValGlyGlnGluIleGlnAspTyr 540

ATCAAATCTTGTTTCATTAATGCTATTAGTGTTTGA 1698
IleLysSerCysPheIleAsnAlaIleSerValstop 551
```

```
tttttggtccttcatgaccaattatcacgtctcatagatctgaaagatgttggttcgcacttttgaaagatgcaaggttat 1779
ggtaccatactaaataagacctcaagtccataggagctataaaactatatatagctcatacgtattgtatttgtggatttt 1860
aatttatgtattgtatttcatgttaaaaaaaaaaaaaaaaaaaaawaaaaaaaa 1914
```

A.5.3 cDNA sequence of α-gurjunene-synthase

The cDNA sequence 2c was identified as α gurjunene-synthase. The UTR is shown in lower case, the ORF in upper case and the translated protein sequence in 3-letter code.

```
ttttaatcaacataagacctaagtccatataaaac 35

ATGGCTGCTAAACAAGGAGAGGTTATTCGTCCAACTGTAAACTATCATCCAAGCCTTTGGGGAGATAAGTTTCTACACTAT 116
MetAlaAlaLysGlnGlyGluValIleArgProThrValAsnTyrHisProSerLeuTrpGlyAspLysPheLeuHisTyr 27

GATGAGCAAGAGGATGAGCAAGTTGAGGTAGATCAACAAATCGAAATTTTGAAGGAaGAAACGAGAAAAGAAATACTCGCA 197
AspGluGlnGluAspGluGlnValGluValAspGlnGlnIleGluIleLeuLysGluGluThrArgLysGluIleLeuAla 54

AGTTTGGATGATCCAGCAAAACATACAAATTTGCTGAAGCTGATTGATGTTATCCAACGTCTCGGTATAGCCTACTATTTT 278
SerLeuAspAspProAlaLysHisThrAsnLeuLeuLysLeuIleAspValIleGlnArgLeuGlyIleAlaTyrTyrPhe 81

GAACATGAGATTACGCAAGCATTGGACCATATTTATAATGTATACGGTGATGAATGGAATGGTGGTAGTACTTCCCTTTGG 359
GluHisGluIleThrGlnAlaLeuAspHisIleTyrAsnValTyrGlyAspGluTrpAsnGlyGlySerThrSerLeuTrp 108

TTTCGACTGCTTCGACAACAAGGCTTTTACGTTTCATGTGATATTTTCAATATCTACAAGCTTGATAATGGATCTTTCAAG 440
PheArgLeuLeuArgGlnGlnGlyPheTyrValSerCysAspIlePheAsnIleTyrLysLeuAspAsnGlySerPheLys 135

GATTCCTTGACCAAGGATATTGAATGCATGCTTGAGTTATATGAGGCAGCCTATATGAGGGTGCAAGGCGAAATCATTCTA 521
AspSerLeuThrLysAspIleGluCysMetLeuGluLeuTyrGluAlaAlaTyrMetArgValGlnGlyGluIleIleLeu 162

GATGAGGCTCTTGAGTTTACAAAAACTCATCTTGAACACATTGCTAAGGATCCCCTTCGTTGCAGCAACACGCTCTCTAGA 602
AspGluAlaLeuGluPheThrLysThrHisLeuGluHisIleAlaLysAspProLeuArgCysSerAsnThrLeuSerArg 189

CACATAGATGAAGCGCTGAAGCGACCTATACTGAAAAGGATACCAAGAGTAGATACAATACTATACATACCTTTCTACGAA 683
HisIleAspGluAlaLeuLysArgProIleLeuLysArgIleProArgValAspThrIleLeuTyrIleProPheTyrGlu 216

CAACAAGATTCTCACAACAAGTCCTTACTAAGACTTGCAAAGTTGGGGTTCAACCGACTTCAATCCTTGCATAAGAAGGAG 764
GlnGlnAspSerHisAsnLysSerLeuLeuArgLeuAlaLysLeuGlyPheAsnArgLeuGlnSerLeuHisLysLysGlu 243

CTTAGCCAACTTTCCAAATGGTGGAAAGAATTTGATGTTCCAAAGAATCTACCTTACGTAAGAGATAGATTGGTTGAGAGC 845
LeuSerGlnLeuSerLysTrpTrpLysGluPheAspValProLysAsnLeuProTyrValArgAspArgLeuValGluSer 270

TACTTTTGGATACTAGGTGTCTACTTTGAACCTCAATATTCTCGTTCTAGAATATTCTTGACAAAAGTATTTGAAATGGCA 926
TyrPheTrpIleLeuGlyValTyrPheGluProGlnTyrSerArgSerArgIlePheLeuThrLysValPheGluMetAla 297

ACAATTCTAGACGACACTTATGATAACTATGGTACTTACGAAGAACTTGAGATATTCACCAAAGCTGTTGATCAAAGGTTG 1007
ThrIleLeuAspAspThrTyrAspAsnTyrGlyThrTyrGluGluLeuGluIlePheThrLysAlaValAspGlnArgLeu 324

CCAATTACCTGCATGGTACGCTTCCcAGATTACATGAAAGTGATTTATAAGAGCTTCTTGGATGTTTTTGAAGAAATGGAG 1088
ProIleThrCysMetValArgPheProAspTyrMetLysValIleTyrLysSerPheLeuAspValPheGluGluMetGlu 351

GAAATCATGGAAAAGGATGGAAAAGCATATCAAGTTCACTATGCAAAAGAGTCGATGATAGATTATGTTAAAAGTTATATG 1169
GluIleMetGluLysAspGlyLysAlaTyrGlnValHisTyrAlaLysGluSerMetIleAspTyrValLysSerTyrMet 378

ACCGAAGCGAAATGGTTACATGAGGGTTATGTGCCAACATTTGAGGAGCATAAATTACTTACATACGTAAGTTGCGGCTAT 1250
ThrGluAlaLysTrpLeuHisGluGlyTyrValProThrPheGluGluHisLysLeuLeuThrTyrValSerCysGlyTyr 405

ATAATGCTTACAGCATCAAGCTTTGTTGGCATGCATGGTGATATAGTTACACAACAGTCGTTCAAATGGGCTCTCAGCAAT 1331
```

```
IleMetLeuThrAlaSerSerPheValGlyMetHisGlyAspIleValThrGlnGlnSerPheLysTrpAlaLeuSerAsn 432

CCTCCACTTGTAATAGCTGCATCTGCCATTAATAGGATTATGAATGATATCGTCGGTCACAAGGAAGAGCAGCAAAGAAAG 1412
ProProLeuValIleAlaAlaSerAlaIleAsnArgIleMetAsnAspIleValGlyHisLysGluGluGlnGlnArgLys 459

CATTTTGCATCTAGTGTGGAAATTTACATGAAAGAACACGATGTCACGGAGGAGCACGTCTACGATTTGTTCAGGAAAAGA 1493
HisPheAlaSerSerValGluIleTyrMetLysGluHisAspValThrGluGluHisValTyrAspLeuPheArgLysArg 486

GTTGATGATGCATGGAAAGATCTAAACCGTGAAACGTTAATATGTAAAGACATTCCAATAGCTCTTAAGATGCGTACGATC 1574
ValAspAspAlaTrpLysAspLeuAsnArgGluThrLeuIleCysLysAspIleProIleAlaLeuLysMetArgThrIle 513

AACCTGGCACGCGTAATGGATACGCTATACAAAAATGATGATACTTTAAAAAATGTTGGAGACGAATTCCAATATTATATA 1655
AsnLeuAlaArgValMetAspThrLeuTyrLysAsnAspAspThrLeuLysAsnValGlyAspGluPheGlnTyrTyrIle 540

AAATCTTGTTTCATTAATGCTTTGAGCATTTGA 1688
LysSerCysPheIleAsnAlaLeuSerIlestop 550

tatgatttttggttcttcatcaccaattgtcatgtctcataaatttggaagatgttggttcgccctttaaaagatgcaag 1796
gttatggtactacgctaaataagacctcaagtcaattggagttggagctataaaactatatgtagttgatacgtgttgtat 1850
ttgtgggttttaatttatgtattgtattttatgtatttatggattttaatttatgtattgtatttcatgactaaaaaaaaa 1931
aaaaaaaaaaaaaaaaaaaa 1951
```

A.5.4 cDNA sequence of 2d

The functin of the cDNA sequence 2d couldn't be identified as the heterologous expresses enzyme didn't show any activity in the enzyme assay. The UTR is shown in lower case, the ORF in upper case and the translated protein sequence in 3-letter code.

```
gatcacaaaattttaatcaacataagaatctaagtccatataaaac 46

ATGGCTGCTAAACAAGGAGAGGTTATTCGTCCAACTGCAAACTATCATCCAAGCCTTTGGGGAGATCAGTTTCTACACTAT 127
MetAlaAlaLysGlnGlyGluValIleArgProThrAlaAsnTyrHisProSerLeuTrpGlyAspGlnPheLeuHisTyr 27

GATGAGCAAGAGGATGAGCAAGTTGAGGTAGATCAACAAATCGAAATTTTGAAGGAAGAAACGAGAAAAGAAATACTCGCA 208
AspGluGlnGluAspGluGlnValGluValAspGlnGlnIleGluIleLeuLysGluGluThrArgLysGluIleLeuAla 54

AGTTTGGATGATCCAGCAAAACATACAAATTTGCTGAAGCTGATTGATGTTATCCAACGTCTCGGTATAGCCTACTATTTT 289
SerLeuAspAspProAlaLysHisThrAsnLeuLeuLysLeuIleAspValIleGlnArgLeuGlyIleAlaTyrTyrPhe 81

GAACATGAGATTACGCAAGCATTGGACCATATTTATAATGTATACGGTGATGAATGGAATGGTGGTAGTACTTCCCTTTGG 370
GluHisGluIleThrGlnAlaLeuAspHisIleTyrAsnValTyrGlyAspGluTrpAsnGlyGlySerThrSerLeuTrp 108

TTTCGGCTCCTCCGTCAACAAGGCTTTTACGTTTCATGTGATACTTTCAATATCTACAAGCTTGATAATGGATCTTTCAAG 451
PheArgLeuLeuArgGlnGlnGlyPheTyrValSerCysAspThrPheAsnIleTyrLysLeuAspAsnGlySerPheLys 135

GATTCCTTAACCAAGGATATTGAATGCATGCTTGAGTTATATGAGGCAGCCTATATGCGGGTGCAAGGCGAAATCATTCTA 532
AspSerLeuThrLysAspIleGluCysMetLeuGluLeuTyrGluAlaAlaTyrMetArgValGlnGlyGluIleIleLeu 162

GATGAGGCTCTTGAGTTTAcTAAAACTCATCTTGAACACATTGCAGATCCTCTTCGTTGCAACAACACGCTCTCTAGA 611
AspGluAlaLeuGluPheThrLysThrHisLeuGluHisIleAlaAspProLeuArgCysAsnAsnThrLeuSerArg 188

CACATAAATGAAGCGCTGAAGCGGCCTATACGGAAAAGGCTACCAAGAGTGGATGCAGTACAATACATACCTTTCTATGAA 692
HisIleAsnGluAlaLeuLysArgProIleArgLysArgLeuProArgValAspAlaValGlnTyrIleProPheTyrGlu 216

CAACAAGATTCTCACAACAAGTCCTTACTAAGACTTGCAAAGTTGGGGTTCAACCGACTTCAATCCTTGCATAAGAAGGAG 775
GlnGlnAspSerHisAsnLysSerLeuLeuArgLeuAlaLysLeuGlyPheAsnArgLeuGlnSerLeuHisLysLysGlu 243
```

```
CTTAGCGAACTTTCCAAATGGTGGAAAGAACTTGATGCTCCAAAGAATCTACCTTACGTAAGAGATAGATTGGTTGAACTC 856
LeuSerGluLeuSerLysTrpTrpLysGluLeuAspAlaProLysAsnLeuProTyrValArgAspArgLeuValGluLeu 270

TACTTTTGGGTACTAGGTGTCTACTTCGAACCTCAATATTCTCGTTGTAGAATATTCaTGACAAAaGTaATTAAAACGATA 937
TyrPheTrpValLeuGlyValTyrPheGluProGlnTyrSerArgCysArgIlePheMetThrLysValIleLysThrIle 297

ACAATTCTAGACGACACGTATGATATCTATGGCACTTACGAAGAACTTGAGATATTCACCAAAGCCGTTCAAAGGTGGTCA 1018
ThrIleLeuAspAspThrTyrAspIleTyrGlyThrTyrGluGluLeuGluIlePheThrLysAlaValGlnArgTrpSer 324

ATTACCTGCATGGATACGCTTCCAGCTGATTACATGAAAGTGATTTATAAGACCCTCTTGGATGTTTACGAAGAAATGGGG 1099
IleThrCysMetAspThrLeuProAlaAspTyrMetLysValIleTyrLysThrLeuLeuAspValTyrGluGluMetGly 351

GAAATCATGGAAAAGGATGGAAAAGCATATCAAGTTGACTATGCAAAAGAGTCGATGATAGATTCGGTTACAAGTTATATG 1180
GluIleMetGluLysAspGlyLysAlaTyrGlnValAspTyrAlaLysGluSerMetIleAspSerValThrSerTyrMet 378

ACCGAAGCAAAATGGTTACATGAGGGATATGTGCCAACATTGGAGGAACATTACACAATTACAAACTTAACTGCTGGCTAT 1261
ThrGluAlaLysTrpLeuHisGluGlyTyrValProThrLeuGluGluHisTyrThrIleThrAsnLeuThrAlaGlyTyr 405

ATGCTCACAGCATCAAGCTTTGTTGGCATGCCTGGTGATATAGTTACACAAGAGTCTTTCAAATGGGCTCTCAACAATCCT 1342
MetLeuThrAlaSerSerPheValGlyMetProGlyAspIleValThrGlnGluSerPheLysTrpAlaLeuAsnAsnPro 432

AAACCACTCATAAAAGCTTCATCTGAAATTGGTAGGATTATGGATGATATCGTCGGTCACAAGGAGGAGCAAAAAAGAAAG 1423
LysProLeuIleLysAlaSerSerGluIleGlyArgIleMetAspAspIleValGlyHisLysGluGluGlnLysArgLys 459

CATATTGTATCTAGGGTGGAAATTTACATGAAAGAACGTCATCTTGCGGAGGAGGACGTCTATGATTTGCTCAAACAAAGT 1504
HisIleValSerArgValGluIleTyrMetLysGluArgHisLeuAlaGluGluAspValTyrAspLeuLeuLysGlnSer 486

GTTGAAGATGCATGGAAAGATCTAAACCGAGAAACATTAACATGCAAAGACATTCATATTGATCTTAAGATGCGTACGATC 1585
ValGluAspAlaTrpLysAspLeuAsnArgGluThrLeuThrCysLysAspIleHisIleAspLeuLysMetArgThrIle 513

AACCTGGCACGCGTAGTAGATATCCTATACAAAAGCGATGATAACTTAAAAAATGTTGGACAAGAAATCAAAGATTATATC 1666
AsnLeuAlaArgValValAspIleLeuTyrLysSerAspAspAsnLeuLysAsnValGlyGlnGluIleLysAspTyrIle 540

AAATCTTGTTTCATTAATGCTATGAGTGTTTGA 1697
LysSerCysPheIleAsnAlaMetSerValstop 549

tttttgggtccttcatgaccaattatcacgtctcatagatttgaaaaatgttggttccacccttttgaaagatgcaaggtc 1778
atgttaccacactaaataagacctcaagtgccgttctaaaaaaaataaaataagacctcaagtctactggagctataaaac 1859
tctatatagctgatacgtattgtatttgtggattttaatttatgtattgtatttncatgtatttgtggattttaatttata 1940
tattgtatttcatgtttttcggaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa 1992
```

Appendix B

Enzyme kinetics data

B.1 pH optimum

pH	(+)GDS [dpm]	(–)GDS [dpm]	αGS [dpm]	CS [dpm]
6.6	339.28	146.67	2543.03	63.33
6.6	339.01	246.34	2150.77	78.1
6.75	862.7	409.36	3106.07	76.01
6.75	573.32	273.3	2423.61	90.28
7.0	739.55	614.12	4355.69	110.69
7.0	3280.85	580.75	3986.54	188.12
7.2	717.91	1040.5	2773.73	254.65
7.2	1197.44	1180.5	3394.44	187.98
7.65	2830.16	1673.08	3493.23	129.64
7.65	2921.84	2084.23	4230.98	164.73
7.8	3109.98	889.71	3702.34	88.38
7.8	1634.59	1225.73	4323.66	113.06
8.0	1016.88	713.47	1109.92	117.95
8.0	1295.38	903.8	1040.06	121.95

B.2 Time linearity

[min]	(+)GDS undil. [dpm]	(+)GDS half dil. [dpm]	(–)GDS undil. [dpm]	(–)GDS half dil. [dpm]	α-GS undil. [dpm]	α-GS half dil. [dpm]	CS undil. [dpm]	CS half dil. [dpm]
0	444.02	201.23	38.06	43.43	–	–	34.53	45.87
0	560.31	271.95	46.09	38.46	–	–	51.4	53.88
5	2519.28	1100.75	707.12	443.71	78.48	63.91	67	58.36
5	2249.5	728.05	682.12	500.6	92.33	73.01	52.14	69.42
10	4455.7	1377.65	1302.64	956.84	943.28	303.22	93.03	74.64
10	4260.19	1329.05	1462.32	611.68	817.49	546.97	117.19	88.25
15	3865.21	1763.85	1909.68	946.8	2047.1	581.88	128.42	135.53
15	6136.42	2015.25	2320.1	1146.46	2249.35	1149.01	116.34	115.95
20	5651.26	3701.5	3304.75	1498.67	2620.4	828.68	161.34	138.9
20	7610.51	2636.65	4871.99	1245.61	2824.06	1279.07	181.13	139.03
30	16153.9	6739.35	3942.86	2063.03	2908.93	1366.36	178.88	161.76
30	9476.54	4337.35	3809.66	1482.51	3402.5	1810.3	217.81	150.57
45	8581.55	7310.2	5539.78	2197.22	3621.71	1985.51	196.82	172.96
45	11025.9	6495.03	5061.69	2486.97	3605.16	1748.1	322.08	178.78
60	15996.3	7852.95	4605.24	2143.67	3557.62	1769.99	288.3	188.11
60	16371	8817.73	4747.69	3177.83	5400.21	1794.01	251.69	187.1

B.3 Michaelis-Menten constant

FPP[μM]	(+)-GDS [dpm]	(–)-GDS [dpm]	α-GS [dpm]	CS [dpm]
0.5	609.34	189.22	260.33	10.93
0.5	633.92	248.37	368.55	40.98
1	863.32	299.51	346.94	32.82
1	1166.0	514.4	563.01	29.13
2	2697.6	784.81	999.04	39.22
2	1342.25	1002.12	896.12	34.62
3	1146.07	710.35	945.27	27.34
3	1164.05	649.11	663.55	18.23
5	5689.04	881.11	862.85	29.8
5	3324.18	979.94	868.59	47.66
10	6558.58	2315.12	2447.25	90.88
10	5894.52	2012.72	1859.01	104.63
20	–	8130.19	–	–
20	–	4603	–	–
30	6900.4	4361.47	2959.51	160.5
30	8970.64	3268.84	2283.79	116.61
50	12957.8	4174.37	2431.17	151.5
50	12699.9	3974.96	2320.84	193.47
100	23561.9	7688	4022.64	378.81
100	16899.2	6771.52	3569.84	363.98
200	27162.4	11313	7945.93	715.25
200	18148.5	11397.2	8248.12	969.37

Appendix C

Hazardous chemicals

Compound	Hazard symbol	R (risk) phrases	S (safety) phrases
Chloroform	Xn	22-38-40-48/20/22	36/37
Diethyl ether	F^+	12-19	9-16-29-33
N,N-Dimethyl-formamide	T	61-E20/21-36	53-45
1,4-Dithiothreitol	Xn	22-36/38	–
EDTA	Xi	22	–
Ethanol	F	11	7-16
Ethidium bromide	T^+	22-26-36/37/38-68	26-28.2-36/37-45
(1(n)-^{3}H)-FPP	F	11	7-16
Guanidine-thiocyanate	Xi	20/21/22-32	13
Glycerol	Xn	22	–
n-Hexane	Xn, F, N	11-38-48/20-51/53-62-65-67	9-16-29-33-36/37-61-62
Hydrogen	F	12	9-16-33
Imidazole	C	22-34	22-26-36/37/39-45

Compound	Hazard symbol	R (risk) phrases	S (safety) phrases
IPTG	–	–	22-24/25
2-Mercaptoethanol	T, N	22-24-34-51/53	26-36/37/39-45-61
Methane	F	12	9-16-33
Methanol	T, F	11-23/25	7-16-24-45
MOPSO	Xn	36/37/38	26-36
Tris(hydroxymethyl)-aminomethane	Xi	36/38	–
Triton X-100	Xn	22-41	24-26-39

Acknowledgement

Ich danke Herrn Prof. W.A. König für die Bereitstellung des überaus interessanten Themas, für sein fortwährenden Interesse sowie seine stete Unterstützung zur Ermöglichung der vorliegenden Arbeit.

I would like to thank Dr. H.J. Bouwmeester for his continuos advices and help in getting the project running and for the Terpene-Discussion-Meetings.

Herrn Prof. B. Bisping danke ich für die Anfertigung des Zweitgutachtens. Herrn Prof. U. Hahn und Herrn Dr. W. Maison danke ich für die Bereitschaft zum 2. und 3. Prüfer.

Ich möchte mich bei Minoo Khedmati-Moghaddam bedanken, die mir die Doktorandenstelle beim Arbeitskreis König vermittelt hat.
Ich danke allen Mitarbeitern des Arbeitskreis König, insbesondere Simla Başar, Julia Pruns, Fernando Campos und Frank Werner für Ihre allzeite Hilfsbereitschaft bei meinen Besuchen.
Frau A. Meiners und Herrn M. Preuße danke ich für die Aufnahme einiger GC-MS Messungen.
Prof. B. Meyer danke ich für die Nutzung der SGI-Maschinen und Jan-Christoph Westermann danke ich für die Einführung ins Substrat-Docking.

I would like to thank all colleagues of Plant Research International in the "de Haaff" and in the "Born Zuid" building for the nice work atmosphere, for their helpfulness and for the "gezellige borrels". I would like to thank all the foreign guest visitors of PRI for the unique multicultural atmosphere and for the nice parties.

In special I would like to thank Jan Kodde and Iris Kappers for the introduction in molecular biology techniques.
Francel Verstappen and Jan-Willem deKraker for introducing me in biochemistry work, radioactive work and for running the GC-MS.
Asaph Aharoni, Cinzia Bertea, Ashok Giri, Mazen El Tamer and Joost Lucker for their help in all kind of molecular biology questions.
Per Mercke for helping me with the cDNA library.

I would like to thank Prof. Gert Vriend (University of Nijmegen) for giving me the opportunity to participate the Bioinformatic and the Protein Homology Modelling Courses
and Mike Beale and Ian Prosser (both at BBRC, UK) for their collaboration on the publication.

Zum Schluß möchte ich Dir, Jens Fransson danken, daß Du so geduldig über die Distanz hinweg zu mir gehalten hast.

Curriculum Vitae

Persönliche Daten	
Name	Iris Altug
Geburtsdatum	28. September 1973
Schulausbildung	
1980–1984	Gorch-Fock Grundschule, Hamburg
1984–1993	Gymnasium Hochrad, Hamburg
Hochschulausbildung	
10.1993–04.1998	Studium der Lebensmittelchemie, Universität Hamburg
02.1997–10.1997	Diplomarbeit: Phenolische Carbonsäuren als Strukturelement in unlöslicher Nahrungsfaser: Gaschromatographische Untersuchungen an Hafer und Haferprodukten
04.1989	Erstes Staatsexamen in Lebensmittelchemie
08.1999	Zweites Staatsexamen in Lebensmittelchemie
Promotion	
11.1999–12.2003	Doktorarbeit an der Universität Hamburg bei Prof. W.A. König in Zusammenarbeit mit Plant Research International, Niederlande
Tätigkeiten	
05.1998–10.1998	Forschungsaufenthalt am Central Science Laboratory, England
11.1998–04.1999	Aufenthalt am Hygiene Institut Hamburg
12.1999–06.2002	Forschungsaufenthalt am Plant Research International, Niederlande