Investigations Into the Function of Two Plant Sesquiterpene Synthases: δ-Cadinene Synthase and (*E*)-β-Farnesene Synthase

by

Amang Li

A thesis submitted to
Cardiff University
for the degree of
DOCTOR OF PHILOSOPHY

School of Chemistry
Cardiff University
November, 2011

UMI Number: U585520

DECLARATION

This work has not previously been accepted in substance for any degree and is not concurrently submitted in candidature for any degree.

Signed .... Date 26 / 03 / 2012

STATEMENT 1
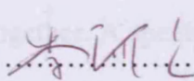
This thesis is being submitted in partial fulfillment of the requirements for the degree of PhD.

Signed .... Date 26 / 03 / 2012

STATEMENT 2
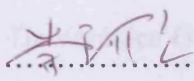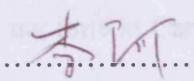
This thesis is the result of my own independent work/investigation, except where otherwise stated. Other sources are acknowledged by explicit references.

Signed .... Date 26 / 03 / 2012

STATEMENT 3

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed .... Date 26 / 03 / 2012

Acknowledgements

Abstract

Terpenoids represent the most structurally and stereochemically diverse family of natural products with more than 55,000 terpenoid structures discovered to date from all life forms. Sesquiterpenes are a class of the terpenoid family, and their formation from farnesyl diphosphate is catalyzed by sesquiterpene synthases. This project focuses on trying to decipher the reaction mechanisms of two sesquiterpene synthases, δ-cadinene synthase from *Gossypium arboreum* and (*E*)-β-farnesene synthase from *Mentha x piperita* and to provide a method for the generation of unnatural terpenes with potential commercial applications in both the pharmaceutical and agrochemical industries.

Modifications by substitution of residues around the active site of δ-cadinene synthase did not lead to any functional divergence, indicating an unusual structural component that determines the product specificity of this enzyme. Domain-swapping experiments based on phylogenetic information suggested that the subdomain encoded by exon 4 is most likely the key structure element controlling the product specificity of this enzyme. Manipulation of the active site volume of (*E*)-β-farnesene synthase by site-directed mutagenesis revealed a rigid active site cavity that is precisely defined for generating mainly acyclic products. The active site hybrid constructed by replacing the active surface of (*E*)-β-farnesene synthase with the corresponding part from δ-cadinene synthase lost activity, suggesting the catalytic specificity of this enzyme is modulated at a distance by residues surrounding the active site, which may have a huge influence on the active site volume. Mechanistic studies utilizing a substrate analogue revealed a new reaction mechanism for (*E*)-β-farnesene synthase. Functional approaches to explore the N-termini region of δ-cadinene synthase and (*E*)-β-farnesene synthase provided direct evidence that suggested dual roles for this region.

TABLE OF CONTENTS

LIST OF SCHEMES

LIST OF TABLES

LIST OF ABBREVIATIONS

A	Adenosine
AA	Amino acid
Abs	Absorbance
ADP	Adenosine diphosphate
AMP	Ampicillin
APS	Ammonium persulfate
AS	Aristolochene synthase
AT-AS	*Aspergillus terreus* Aristolochene synthase
ATP	Adenosine triphosphate
β-ME	β-Mercaptoethanol
bp	Base pair
BSA	Bovine serum albumin
C	Cytosine
c	Concentration
CD	Circular dichroism
cDNA	Complementary DNA
CoA	Coenzyme A
CV	Column volume
DCS	δ-Cadinene synthase
DEAE	Diethylaminoethyl
dH$_2$O	Deionised water
DMAPP	Dimethyallyl diphosphate

DMSPP	Dimethylallyl-*S*-thiolodiphosphate
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphate
dsDNA	Double stranded DNA
DTT	Dithiothreitol
EBFS	(*E*)-β-Farnesene synthase
EDTA	Ethylenediaminetetraacetic acid
EI	Electron ionisation
ES	Electron spray
FHP	Farnesyl hydroxyphosphonate
FPLC	Fast protein liquid chromatography
FDP	(*2E,6E*)-Farnesyl diphosphate
2F-FDP	(*2Z,6E*)-2-Fluoro-farnesyl diphosphate
2F-GDP	*2Z*-2-Fluoro-geranyl diphosphate
G	Guanine
GC	Gas chromatography
GC-MS	Gas chromatography-mass spectrometry
GCS	Germacrene C synthase
GGDP	Geranylgeranyl diphosphate
GDP	(*E*)-Geranyl diphosphate
hr	Hours
HPLC	High-performance liquid chromatography
HPS	*Hyoscyamus muticus* premnaspirodiene synthase

IPP	Isopentenyl diphosphate
IPTG	Isopropyl-β-D-1-thiogalactopyranoside
ISPS	Isoprene synthase
k_{cat}	Turnover number
K_M	Michaelis constant
l	Pathlength
LB	Luria-Bertani growth media
LDP	Linalyl diphosphate
MEP	Methylerythritol phosphate
MOPS	3-(N-morpholino)propanesulfonic acid
MS	Mass spectrometry
MVA	Mevalonic acid
MW	Molecular weight
NADPH	Reduced nicotinamide adenine dinucleotide phosphate
NDP	(*E*)-Nerolidyl diphosphate
OD	Optical density
ORF	Open reading frame
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PR-AS	*P. roqueforti* Aristolochene synthase
RNA	Ribose nucleic acid
rpm	Revolutions per minute
SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis

Chapter 1: Introduction

1.1 Terpenoids

Terpenoids represent the structurally and stereochemically most diverse family of natural products. More than 55,000 terpenoids have been discovered to date from nature (*1*). The extraordinarily diverse carbon skeletons of terpenoids are formally derived from the branched C5 skeleton of isoprene (**1**) (*2*). Dimethylallyl diphosphate (DMAPP, **2**) and isopentenyl diphosphate (IPP, **3**) are biological equivalents of isoprene and are the universal precursors of the isoprenoids in all living organisms (Scheme 1.1) (*3*).

The most significant building reaction in the terpenoids pathway is a chain elongation reaction catalysed by isoprenyl diphosphate synthases to yield increasingly longer polyisoprenoid diphosphates, such as geranyl diphosphate (GDP C10, **4**), farnesyl diphosphate (FDP C15, **5**) and geranylgeranyl diphosphate (GGDP C20, **6**). In addition, longer acyclic intermediates, such as squalene (C30, **7**), phytoene (C40, **8**), all-trans polyprenyl diphosphates (C5n, **9**) and *cis*-polyprenyl diphosphates (C5n, **10**) can be further derived from either FDP or GGDP. All these acyclic precursors can then be modified to generate acyclic, monocyclic or multicyclic products (*4-6*).

Thus, terpenoids can be grouped based on the number of C5 units: hemiterpenes (C5), monoterpenes (C10), sesquiterpenes (C15), diterpenes (C20), triterpenes (C30) and so on. Different attachment patterns such as non-head-to-tail chain elongation or irregular conformations have also been discovered (Scheme 1.2). The biosynthesis of sterols and carotenoids goes through a cyclopropanation reaction as the first step in the specific pathway among eukarya, archaea and some bacteria (*7, 8*). A branching reaction has only been identified from a few plants (*9*) and a cyclobutanation reaction has only been reported in the study of mealybug mating pheromones (*10*).
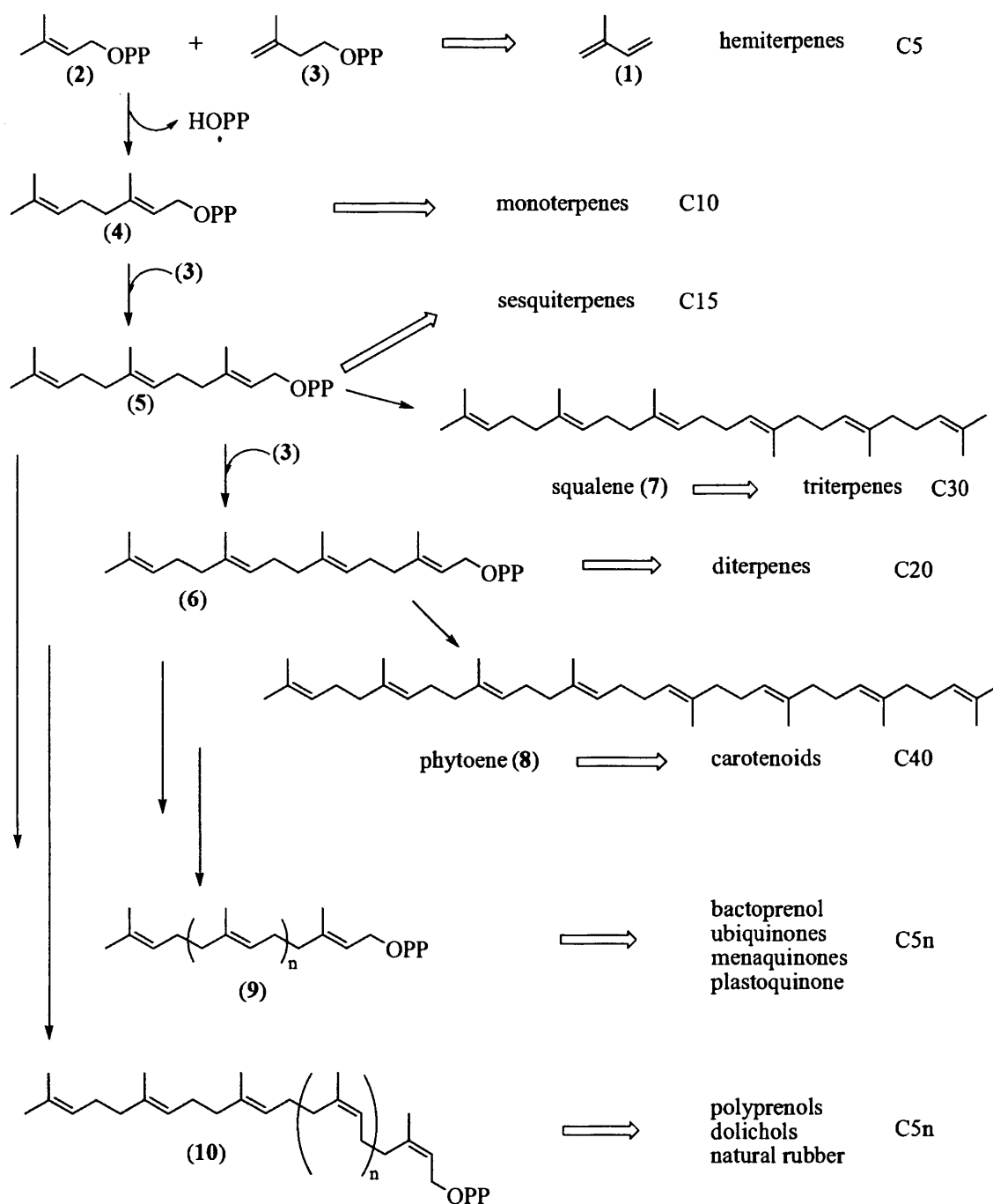
Scheme 1.1: *Biogenetic scheme for the formation of the main terpenoid series through the regular head to tail fusion.*

The fundamental chain elongation of terpene biosynthesis involves condensations of DMAPP **(2)** with IPP **(3)**, corresponding to the ionization of DMAPP **(2)** triggered by the enzyme's metal cluster (Scheme 1.3) (*11*). The following electrophilic attack between the resulting carbocation and C3-C4 π bond of IPP generates a tertiary carbocation at C3. GDP is formed after stereospecific elimination of the C2-H_R proton

3

(*11*, *12*)). Further chain elongations forming FDP and GGDP were suggested to share similar coupling mechanisms based on the crystal structures of both farnesyl diphosphate synthase and geranylgeranyl diphosphate synthase (*13*, *14*). On the other hand, to form cyclobutanation, branching and cyclopropanation products, one molecule of IPP is replaced by DMAPP. These reactions also employ the allylic carbocation derived from DMAPP by allowing alternative trajectories of carbon-carbon bond formation (Scheme 1.3) (*15*).
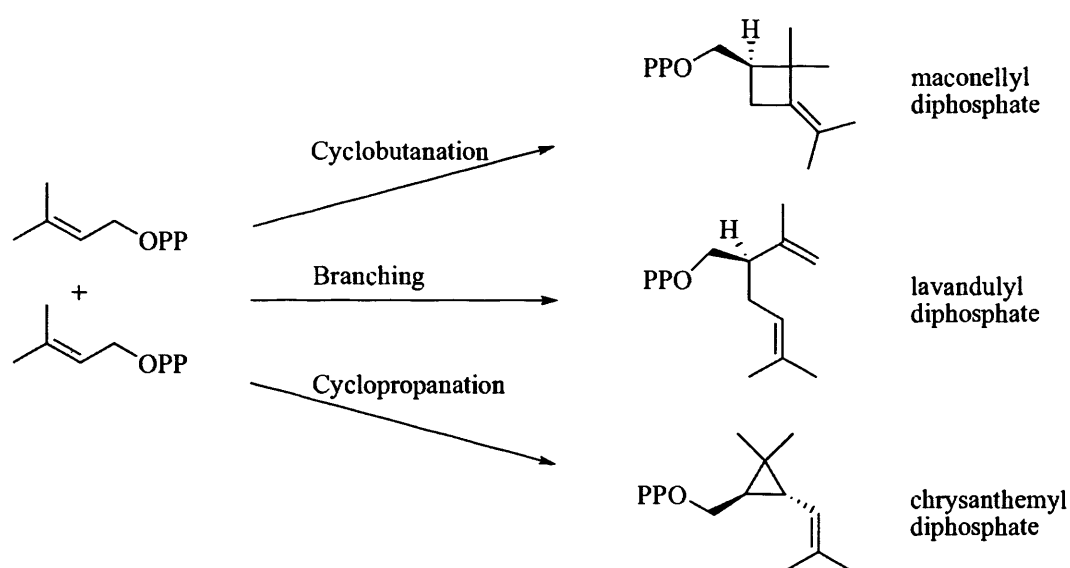
Scheme 1.2: *Scheme for the formation of non-head-to-tail conformation structures.*

The enzymes that catalyse chain elongation can be divided into two subfamilies according to the stereochemistry of the newly formed double bond. Those that generate *E* double bond conformations usually synthesise shorter chain products (*e.g.* GDP, FDP and GGDP) found early in the pathway. Others that generate *Z* double bond conformations synthesize longer chain diphosphates (*e.g.* natural rubber).
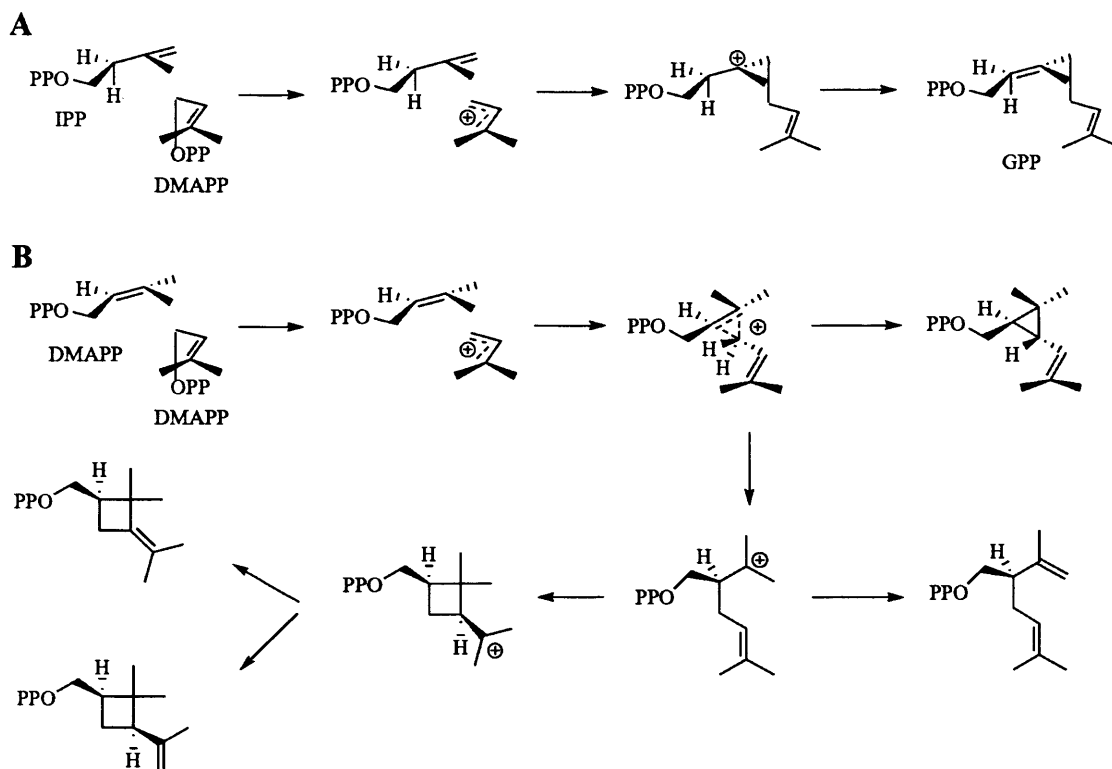
Scheme 1.3: *A dissociative electrophilic alkylation mechanism for chain elongation (A), cyclopropanation, branching and cyclobutanation (B).*

1.2 Biological functions of terpenoids

Low molecular weight organic compounds produced by organisms are generally divided into two groups. The ones that are required for reproduction or growth are defined as primary metabolites, while those that have no apparent function in the basic processes of growth and development are defined as secondary metabolites (*16*). Very little was known about the functionality of terpenoids before the middle of the last century and the compounds were even defined as products of detoxification or overflow metabolism (*16*). However, since the 1970s, quite a few terpenoids were identified to be toxins, repellents or attractants (*17*).

1.2.1 Primary metabolites

Sterols, carotenoids, growth regulators and polyprenol substituents of dolichols, quinones and proteins are classified as primary metabolites (*16*). Generally, these terpenoids are important for membrane integrity, photoprotection and orchestration of

developmental programs and biochemical functions of specific membrane systems.

Sterols are an essential component for the formation of liquid-ordered membrane states (membrane fluidity), which has been suggested to be related to many important biological functions, such as signal transduction, cellular sorting, cytoskeleton reorganization, asymmetric growth and infectious disease (18). Sterols are widely dispersed among organisms and all of them contain the common sterol ring system, which is derived from C30 terpenene squalene (7) or its analogue squalene oxide. Animals employ cholesterol (11) as the major sterol; ergosterol (12) is found in fungi and more than 40 plant sterols have been discovered to contain more complex sterol compositions, in which the major constituents are stigmasterol (13) and sitosterol (14) (Scheme 1.4).

Cholesterol (11) was first isolated from gallstones and is an essential membrane component in higher eukaryotes. With the help of cholesterol (11), the cell membrane becomes semipermeable between cellular compartments and it also alters the functions of membrane proteins. In addition, it participates in several membrane trafficking and transmembrane signalling processes (19). Recent research indicates an important role for cholesterol (11) in several diseases, such as the pathogenesis of cardiac and brain vascular diseases, dementias, diabetes and cancer (20, 21).

Stigmasterol (13) and sitosterol (14) from plants are involved in the polarized growth of pollen tube and root hair. The asymmetric distribution of these components in the membrane causes the asymmetric growth of plant cells (18). These sterols are also suggested to have a temperature dependent ordering effect on membranes; in other words they have a significant disordering effect when temperatures are below the phase transition. In addition, plant sterols can lower the concentrations of cholesterol in serum (22). The appearance of plant sterols in the cell could reduce the absorption of cholesterol from the gut by a competition effect. This significant effect makes plant sterols useful drug candidates for cholesterol-lowering purposes.
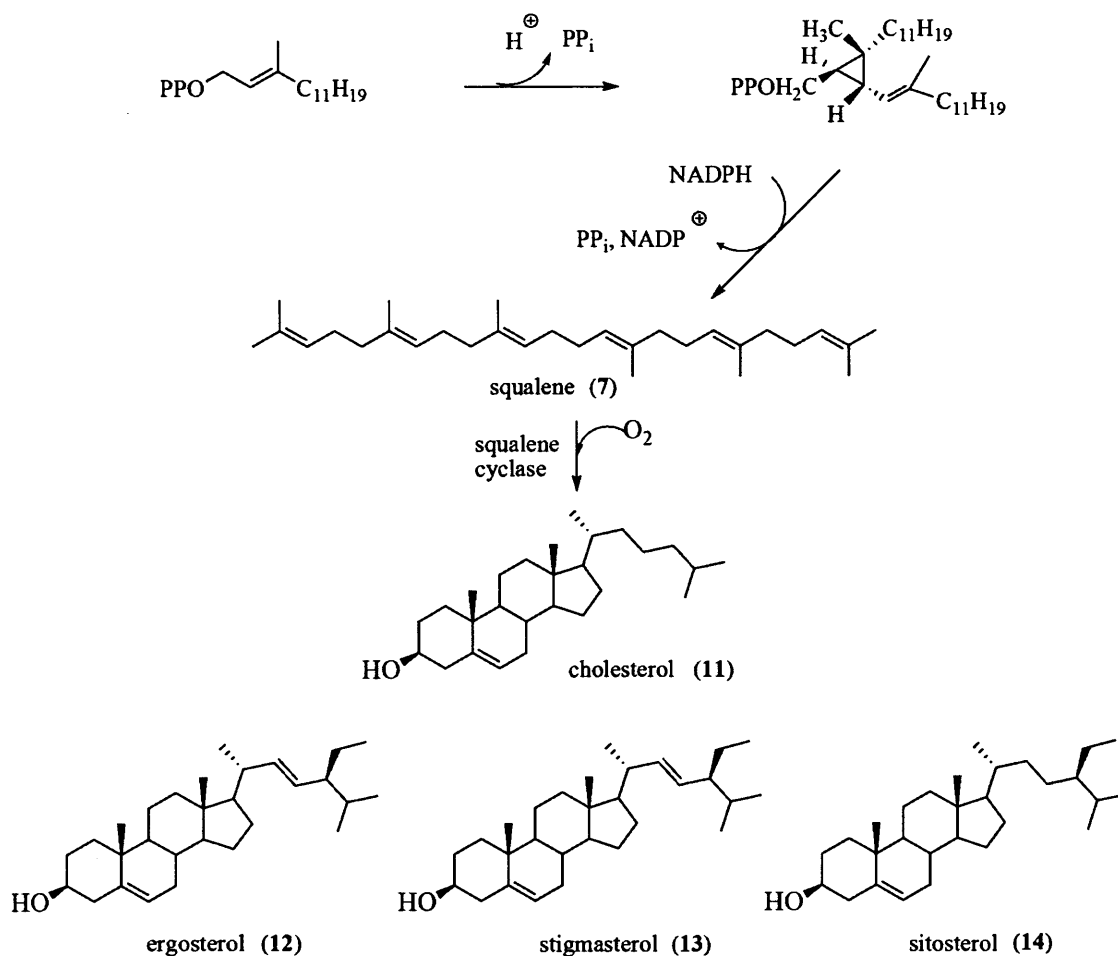
Scheme 1.4: *Molecular structures of animal, fungus and plant sterols.*

The carotenoids are traditionally thought to be a group of natural plant pigments, although they have been identified also in animals and microorganisms. More than 600 different carotenoids have been isolated and characterized from natural sources and they are all biosynthesized from two GGDP units (*23*). The basic acyclic and cyclic structures are illustrated by lycopene (**15**) and β,β-carotene (**16**) (Figure 1.1). A long chain of alternating double and single bonds forms the central part of the molecule and the variable structures can be further modified by cyclization at one or both ends. Also, modifications can be carried out by hydrogenation and addition of oxygen-containing functional groups. The conjugated polyene structure can determine many functions such as light absorption properties, photochemical properties, consequent light-harvesting and photoprotective action (*23*).

A

lycopene   (15)

β,β-carotene   (16)

B

CH₂R
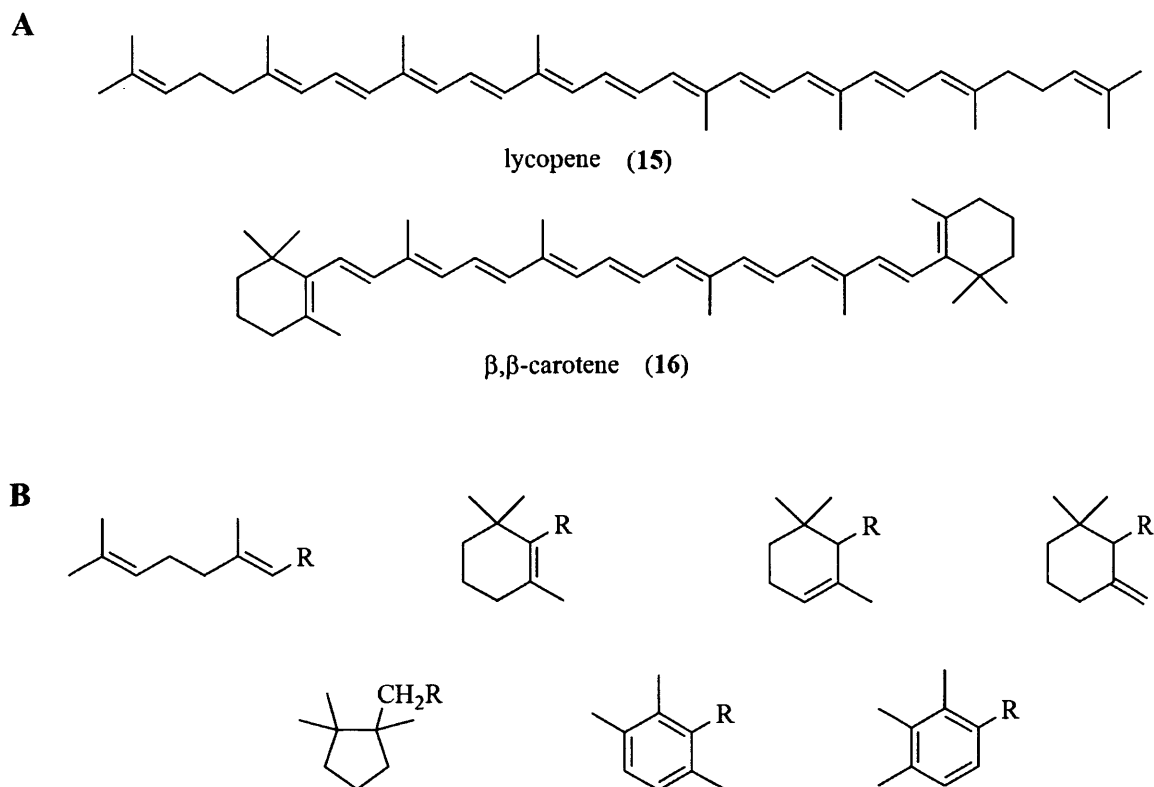
Figure 1.1: *(A) Structures of an acyclic lycopene and a dicyclic β,β-carotene. (B) The seven different end groups found in natural carotenoids.*

Carotenoids can also serve as antioxidants *via* the corresponding carotenoid radicals that react with oxidizing agents (*24*). The conjugated structure of carotenoids allows the unpaired electron to be highly delocalized over the whole structure, providing a stabilizing effect to facilitate subsequent reactions. The radicals can be removed by reacting with each other to generate harmless products or by disrupting the free radical chain reaction.

Quinones are another type of natural terpenoid compounds that play a role in nature as the electron transfer mediator in many essential biochemical phenomena, such as iron respiration by microorganisms and nutrient acquisition by plant and microorganisms (*25, 26*). More than a thousand different quinones are produced by bacteria, fungi, plants and insects, most of which contain a species-specific polyisoprene chain (*27*). Quinones can undergo facile and reversible electron transfer reactions, which profits from the unique structural characteristics (Figure 1.2) allowing resonance stabilization

of semiquinone radical intermediates during redox chemistry (*28*). The electron transfer reactions can also be irreversible due to side-reactions. Nowadays, quinone containing compounds are widely used for their antitumour and anticancer activity (*27*).
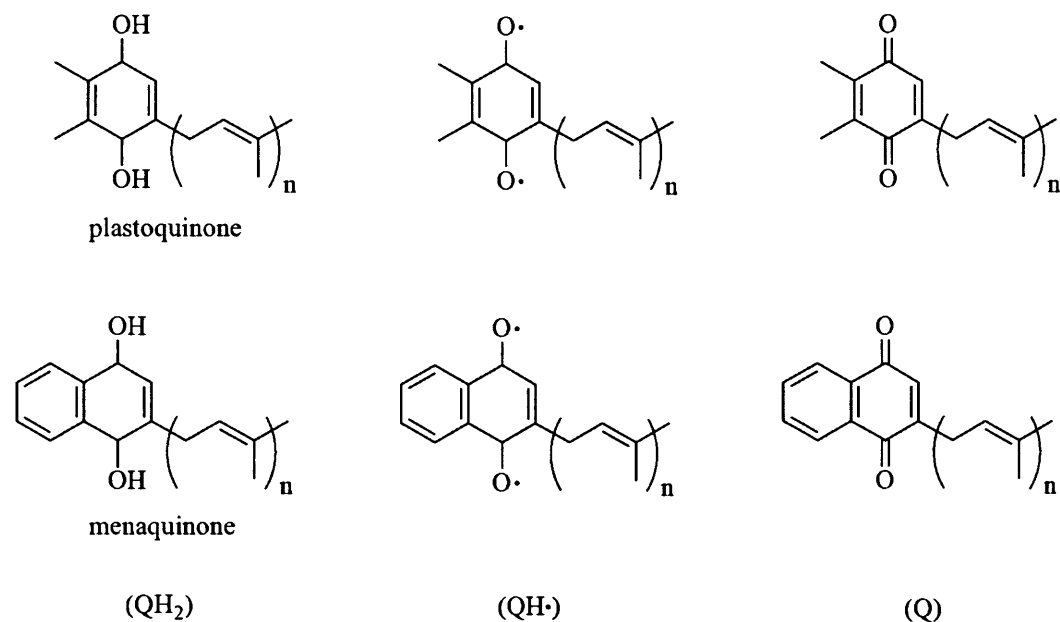
Figure 1.2: *Structures of quinones in different redox conditions.* (QH₂): Hydroquinone. (QH•): semiquinone radical. (Q): Quinone.

1.2.2 Secondary metabolites

Many monoterpenes, sesquiterpenes and diterpenes are classified as secondary metabolites, which are important for ecological viability among organism. However, these substances have been demonstrated to play important roles among organisms and have valuable Applications in medicine, agriculture and industry (*17*).

Using genetically transformed organisms, in which certain terpenoids are overproduced without an effect on other traits, can test functions of terpenoids. In a recent example, an engineered *Arabidopsis thaliana* plant was used to emit large amounts of the monoterpene alcohol linalool (**17**) (Figure 1.3) (*29*). Compared with wild-type *A. thaliana*, the transgenic plants showed significantly increased repellent activity towards aphids, suggesting a defense function for this monoterpenoid.

(S)-linalool (17)
a monoterpenoid

avenacoside B (18)
a saponin

chrysomelidial (19)
an iridoid monoterpenoid

caulerpenyne (20)
a sesquiterpenoid

artemisinin (21)
a sesquiterpenoid

paclitaxel (22)
a diterpenoid

geraniol (23)
a monoterpenoid

nerolidol (24)
a sesquiterpenoid

(E,E)-farnesol (25)
a sesquiterpenoid

(E)-β-caryophyllene (26)
a sesquiterpenoid

(+)-strigol (27)
an apocarotenoid

Figure 1.3: *Structures of terpenoids with established functions.* Rha = rhamnose, Glu = glucose.

Plant terpenoids also have important roles in resistance to diseases caused by fungi and bacteria. A study of the function of triterpenoid saponins 18 was carried out by employing an oat species (*Avena strigosa*). Saponins are terpene glycosides that serve

as detergents (Figure 1.3). They are toxic to fungi due to their ability to form complexes with sterols in fungal membranes, leading to the loss of membrane integrity (*30*). Compared with wild-type lines, mutants that are deficient in generating saponins are much more easily attacked by fungal pathogens (*31*).

Animals, especially insects, can employ terpenoids as protective substances, an example of which is the iridoid monoterpene **19** of leaf beetles (Figure 1.3) (*32*). The terpenoids can be directly sprayed at enemies as effective repellents for small predators such as fire ants.

Terpenoids may also function as defense agents in the marine world. The study of caulerpenyne (**20**) and certain diterpenes of marine algae revealed unusual defense effects by converting polyacetates into the potent feeding deterrent halimedatrial when algal tissue is wounded (Figure 1.3) (*33*). Other research showed some marine organisms employ terpenoids not only to combat predators and pathogens, but also to prevent their surfaces from being colonized by bacteria and fungi (*34*).

Based on the terpenoid defences present in nature, many have been further developed for medical use. Artemisinin (**21**) (Figure 1.3), a sesquiterpene lactone from *Artemisia annua* and its derivatives have been widely used as antimalarial drugs. The compounds can inhibit the growth of the sarco-endoplasmic reticulum in all asexual stages of the malarial parasite, leading to the death of the parasite (*35*). Another widely used medicine with a diterpene structure is the anticancer compound paclitaxel **22** (Taxol) (Figure 1.3). This compound can kill tumour cells by binding to tubulin, which interferes with microtubule dynamics and arrests cell division during mitosis (*36*).

As well as defense functions, terpenoids also play essential roles in interactions among organisms by acting as messengers. The most important reason for organisms to choose terpenoids (especially monoterpenes and sesquiterpenes) to exchange

information is due to their low molecular weight and lipophilic character with high vapour pressures at ordinary temperatures, which allow them to serve as good conveyors of information over distances. In addition, the significant structural variety of terpenes allows messages to be very specific.

Terpenoids are one of the major components of fruit and flower volatiles (*37*). Studies of *Manduca sexta* (the tobacco hornworm) revealed that this pollinator contains a group of receptor cells that can respond to certain oxygenated monoterpenes and sesquiterpenes, such as geraniol (**23**), nerolidol (**24**) and farnesol (**25**) (Figure 1.3) (*38*). In the other plants, emission of blends of terpenoid components has also been discovered when the foliage is under attack (*39, 40*). These compounds act as a 'call for help', attracting predators and parasitoids that feed on herbivores.

Terpenoids can also be employed by the underground parts of plants. The sesquiterpene (*E*)-β-caryophyllene (**26**) was shown to be released by maize roots when the plant is under attack, which attracts nematodes that prey on insect larvae (Figure 1.3) (*41*). Nonvolatile terpenoids, such as strigol (**27**) and other strigolactones can be involved in underground communications (Figure 1.3). These compounds stimulate the growth of *Arbuscular mycorrhizal* fungi from mutualistic, which facilitates nutrient acquisition of the plant (*41*). In some cases, terpenoids released by plants may also help their enemies, such as herbivorous insects or parasitic plants, to locate their host. Guided by a blend of monoterpenes, seedlings of parasitic plant dodder (*Cuscuta pentagona*) grow toward nearby host tomato plants, while other parasitic plants may use strigolactones for a similar function (*42, 43*).

1.3 Biosynthesis of terpenoids

In the 19th century, the hemiterpene isoprene was believed to be the fundamental component for the generation of longer terpenoids, such as monoterpenes and sesquiterpenes. Isoprene itself was first isolated from the low boiling fraction of the

dry distillation of rubber in 1826 (*44*), while the term 'isoprene' was first used by Williams in 1860 (*45*). Based mainly on studies of monoterpenoids, Otto Wallach proposed the 'isoprene rule', which suggests that monoterpenoids are generated by attachment of two isoprene units either in a regular head to tail fashion or in an irregular sequence (*46*). The first formulation that supports the isoprene rule was provided in 1897, which was considered as an early example of a Diels-Alder condensation (Scheme 1.5). The corresponding organic synthesis of this theory was carried out in the same year (*45*). Until now, the isoprene rule has already been identified among more complex terpenoids structures, such as sesquiterpenoids and triterpenoids (*45*).
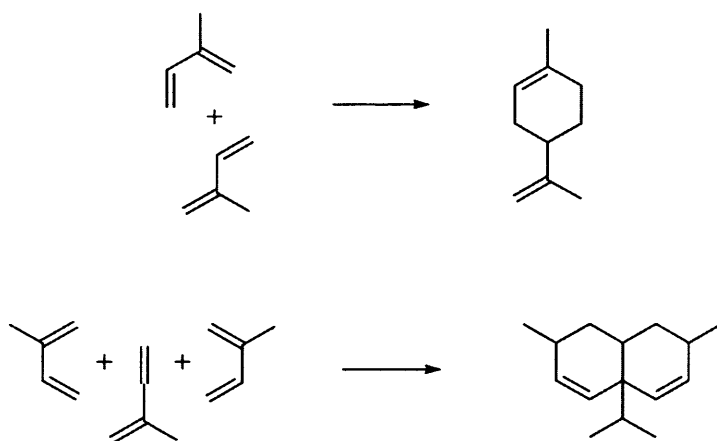
Scheme 1.5: *Proposed mechanism for the formation of monoterpene and sesquiterpene by isoprene units by Otto Wallach.*

The isoprene unit can be identified in many terpenoid structures as an essential component and some simple terpenoids can be also chemically synthesized from it. However, it is not the biological precursor for terpenoids. DMAPP (2) and IPP (3) were found to correspond to the biological equivalents of isoprene and then defined as the universal precursors of terpenoids in all living organisms(*47*). The formation of the isoprene unit was first studied in liver tissues and in yeast in the 1950s, leading to the discovery of the well-known mevalonic acid (MVA) pathway (Scheme 1.6) (*48*). This pathway is widespread for production of polyisoprenoids and sterols in fungi, plants, animals, most other eukaryotes, archaea and some eubacteria.
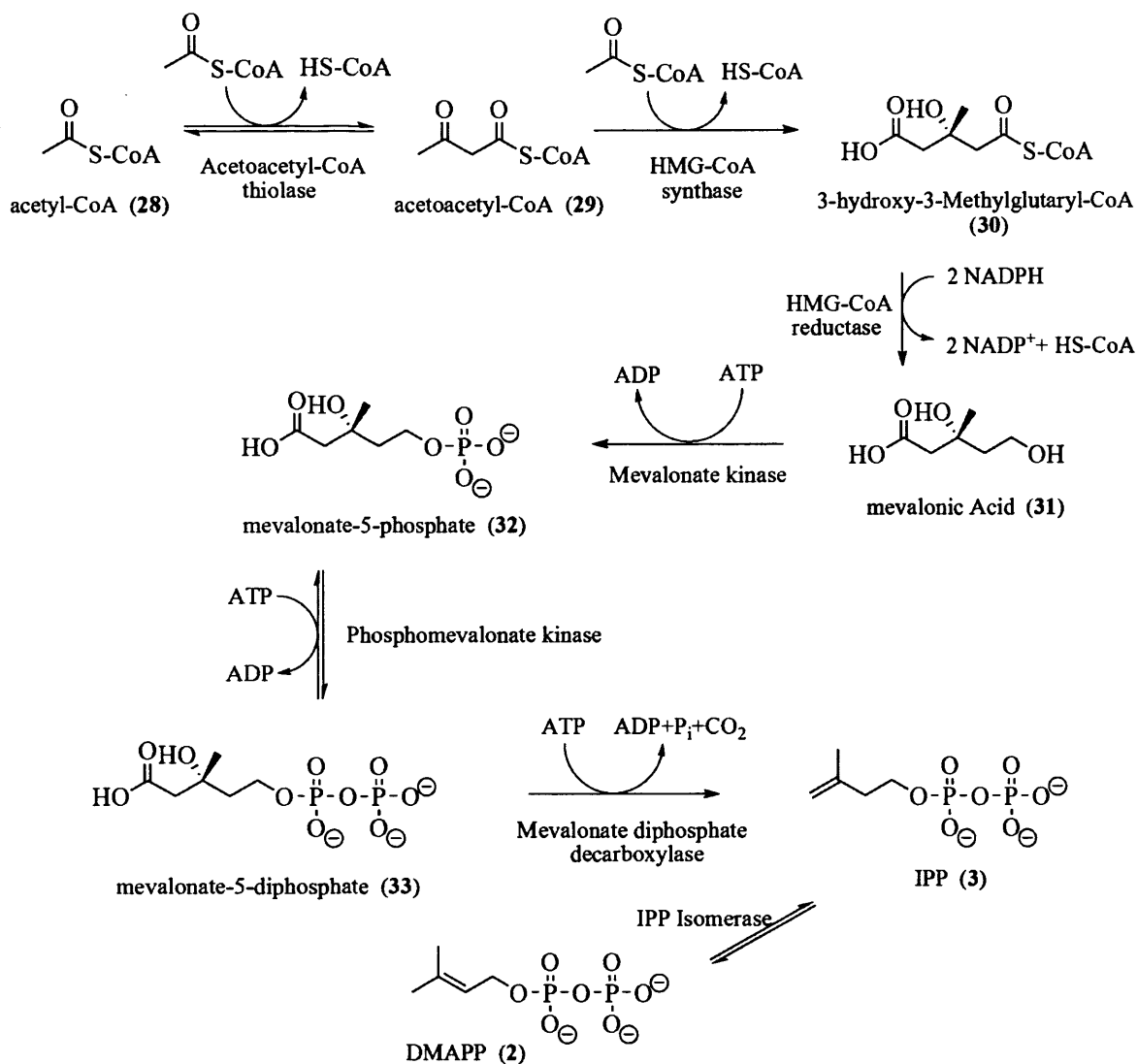
Scheme 1.6: *The mevalonate (MVA) pathway.*

Acetyl-CoA (**28**) is first extended by a two carbon acetyl modules *via* a Claisen condensation reaction catalyzed by acetoacetyl-CoA thiolase (Scheme 1.6). Another similar chain extension is catalyzed by HMG-CoA synthase to form 3-hydroxy-3-methylglutaryl-CoA (**30**) *via* an aldol-type reaction. The most important precursor, mevalonic acid (**31**), is then formed by a reduction using 2 equivalents of NADPH. A diphosphate group is added onto the resulting acid from two ATP molecules and mevalonate diphosphate decarboxylase removes CO_2 to yield IPP (**3**) after a second P_i is added. Isomerisation catalysed by IPP isomerase then leads to DMAPP (**2**).

However, the MVA pathway is not the only pathway employed by organisms to produce isoprenoids. An alternative route toward biosynthesis of IPP and DMAPP was discovered during a study of the unique features of bacterial hopane triterpenoids using ^{13}C-labeled acetate as the only carbon source (49). Incorporation of [1-^{13}C]- and [2-^{13}C]-acetate into the hopane series from three bacteria (*Rhodopseudomonas palustris, Rhodopseudomonas acidophila and Methylobacterium organophilum*) revealed an unusual labelling pattern observed in the isoprene unit that did not fit well with a direct incorporation of acetate into the MVA pathway. A further investigation by incorporation of a whole series of ^{13}C-labeled glucose isotopomers into hopanoids of the bacterium *Zymomonas mobilis* demonstrated that the carbon skeleton is derived from a C2 subunit formed from pyruvate (34) by decarboxylation and a C3 subunit derived from a triose phosphate derivative (Scheme 1.7) (50, 51). In addition, incorporation of doubly labelled [4,5-^{13}C$_2$] glucose and uniformly labelled [U-^{13}C$_6$] glucose showed that an intramolecular rearrangement was involved in the reaction (52). This rearrangement is incompatible with the MVA pathway. Incorporation of ^{13}C-labeled pyruvate and ^{13}C-labeled glycerol using *E. coli* mutants, each lacking a single enzyme of the triose phosphate metabolism, showed that D-glyceraldehyde phosphate (GAP, 35) together with pyruvate (34) are the precursors of the alternative pathway (51).

The first step of this route is the formation of 1-deoxy-D-xylulose 5-phosphate (DXP, 36) by the condensation of GAP and pyruvate catalyzed by the thiamine diphosphate-dependent DXP synthase (dxs). The second step of this biosynthetic pathway is catalysed by the DXP reductoisomerase (dxr). This enzyme converts DXP into 2-C-methyl-D-erythritol 4-phosphate (MEP, 37) through an acid catalyzed rearrangement coupled with an NADPH-dependent reduction. In the next step, a cytidyl moiety is introduced into MEP *via* a diphosphate bond to produce 4-diphosphocytidyl-2-*C*-methyl-D-erythritol (38) and in turn this reaction was catalysed by 2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase (ygbP). The resulting intermediate **38** is phosphorylated by 4-(cytidine

5'-diphospho)-2-C-methyl-D-erythritol kinase (ychB) and then cyclised to form 2-C-methyl-D-erythritol 2,4-cyclodiphosphate (40) after loss of the cytidyl moity in a reaction catalysed by 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (ygbB). The two final dehydration steps were catalyzed by the 2-C-methyl-D-erythritol 2,4-cyclodiphosphate reductase (gcpE) and 4-hydroxy-3-methylbut-2-enyl diphosphate reductase (lytB), resulting DMAPP (2) and IPP (3) as final products.
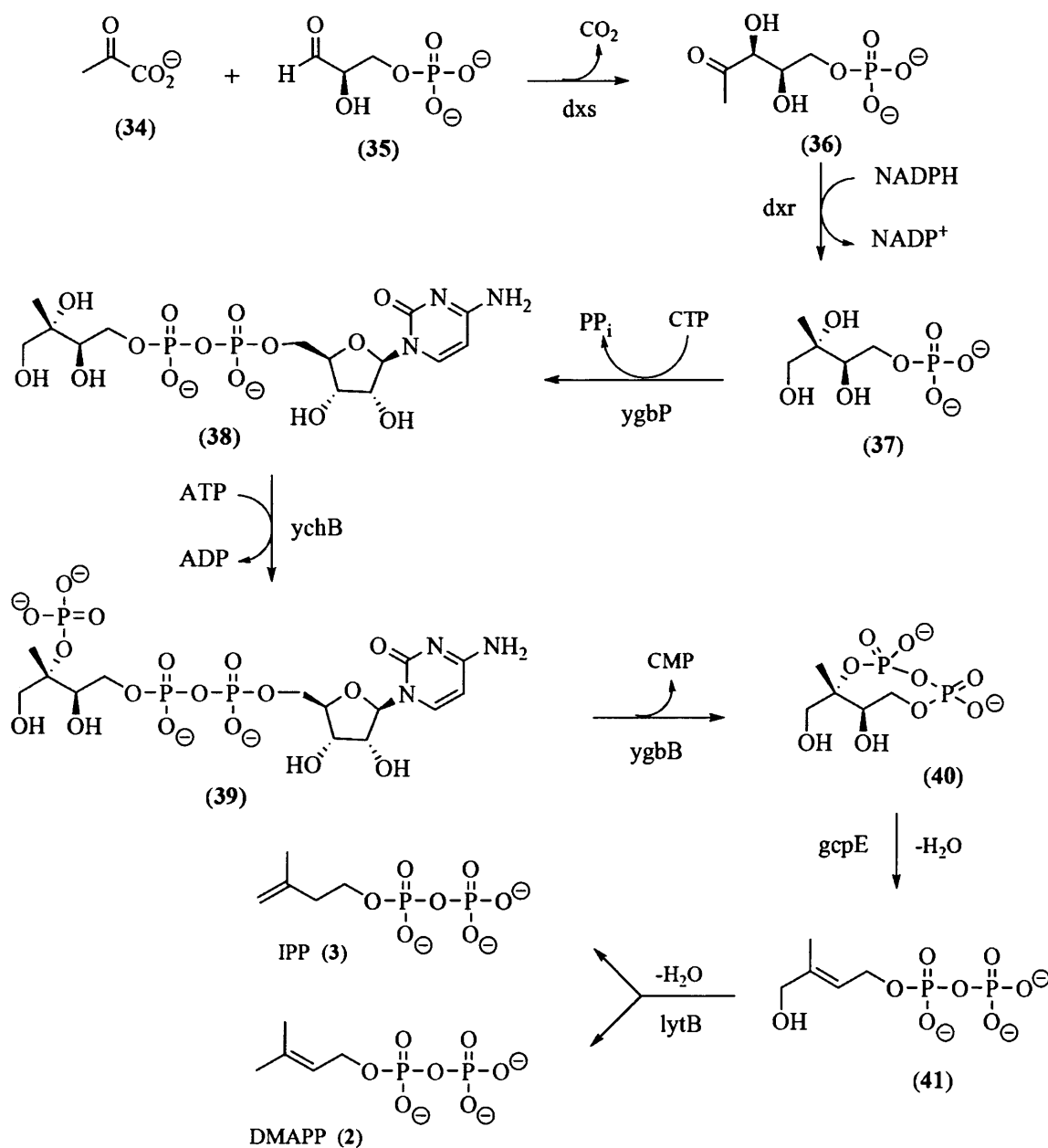
Scheme 1.7: Methylerythritol phosphate (MEP) pathway.

The MVA pathway has been well-established to account for production of polyisoprenoids and sterols in fungi, plant cytoplasm, animals, most other eukaryotes, archaea and some bacteria. Much work towards the MVA pathway related enzymes has been carried out (*53*). These enzymes are essential for animals and human, therefore they have been identified as potentially useful drug targets for modulation of polyisoprenoid and sterol biosynthesis. Genetic work has also illustrated that disruption of genes encoding various enzymes in the MVA pathway showed positive effects on treatment of human pathogenic diseases (*54*).

The alternative MEP pathway of isoprenoid biosynthesis is widely distributed among pathogenic bacteria and is the predominant pathway (*55*). As exceptions, some bacteria possess both pathways (*54, 56*). Such bacteria utilise the MVA pathway for the production of isoprenoids classified as secondary metabolic products and the MEP pathway for the biosynthesis of essential isoprenoids that are classified as primary metabolic products. Genetic studies showed that the yeasts (*Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*), the microsporidian intracellular parasite (*Encephalitozoon cuniculi*), the cryptomonad (*Guillardia theta*) and all completely sequenced animal species (*Homo sapiens, Mus musculus, Rattus norvegicus, Danio rerio, Drosophila melanogaster, Anopheles gambiae, Caenorhabditis elegans*) have complete sets of mevalonate genes, while genes of the MEP pathway are absent (*55*). The fact that animals exclusively use the MVA pathway makes the alternative pathway a valuable target for the development of new specific antibacterial and antiparasitic drugs.

1.4 Terpene synthases

The essential function of a terpene synthase is to catalyse the formation of terpenoid products from linear polyisoprenoid substrate(s) in a reaction that usually shows remarkable structural and stereochemical precision. Enzymatic terpenoid formation generally proceeds in three steps: generation of a highly reactive carbocation,

transformation of the resulting carbocation and quenching of the terminal carbocation by a base (addition of a diphosphate group/hydroxyl group or deprotonation). In terms of the first reaction step, the terpenoid synthases can be divided into two classes (*57*). Most class I terpenoid synthases generate an allylic carbocation by the departure of the diphosphate group. The isoprenyl diphosphate synthases, monoterpene synthases and sesquiterpene synthases belong to this class. Class II terpenoid synthases generate the carbocation by the protonation of the carbon-carbon double bond or the corresponding epoxide ring. Enzymes from this class include the diterpene, triterpene and tetraterpene synthases.

Terpenoid synthases have similar physical and chemical properties, such as requiring a divalent metal ion for catalysis and the universal electrophilic reaction mechanism. As an exception, the presence of metal ions are not absolutely required for the class II terpenoid synthases (*58*). The specificity of the terpenoid synthase reaction can vary dramatically. High fidelity enzymes generate mostly one product with extraordinary structural and stereochemical precision. For example, δ-cadinene synthase from *G. arboreum* generates exclusively (+)-δ-cadinene as the product (*59*). In contrast, promiscuous enzymes could generate different structures such as γ-humulene synthase, which generates 52 different cyclization products (*60*).

1.4.1 Isoprenyl diphosphate synthases

Isoprenyl diphosphate synthases (also known as prenyltransferases) catalyse isoprenoid chain elongations to generate prenyl diphosphates with various chain lengths using the C5 precursors DMAPP (**2**) and IPP (**3**) (*11*). The carbon chain length of linear terpenoids in nature varies from geranyl diphosphate (GDP, C10) to natural rubber (C>100,000) and many of these linear compounds can then be cyclized to generate cyclic or multicyclic products by terpene synthases.

The isoprenyl diphosphate synthases can be divided into two subgroups based on their protein structure and product stereochemistry (*61*). (*E*)-isoprenyl diphosphate synthase catalyse the synthesis of all-*trans*-prenyl diphosphates with various chain lengths from C10 to C50. On the other hand, (*Z*)-isoprenyl diphosphate synthases catalyze the successive *cis*-condensation of IPP with the allylic diphosphate to generate more diverse long chain prenyl diphosphates, from *cis*-FDP (C15) to natural rubber (C>100,000).

The crystal structure of avian farnesyl diphosphate synthase reveals the first view of a tertiary structure of sesquiterpenoid synthase (*13*). This enzyme is a homodimer with a hydrophobic active site surrounded by 10 α-helices. A more recent crystal structure of farnesyl diphosphate synthase from *E. coli* complexed with substrate IPP and the nonreactive substrate analogue dimethylallyl-*S*-thiolodiphosphate (DMASPP) revealed two aspartate-rich DDXX(XX)D motifs on helices D and H liganded with a trinuclear magnesium cluster (Figure 1.4) (*62*). Mg^{2+}_A and Mg^{2+}_B each form six-membered ring chelate structures with two unesterified diphosphate oxygens. Asp105 and Asp111 from *E. coli* farnesyl diphosphate synthase make bidentate coordination interactions with Mg^{2+}_A and Mg^{2+}_C. Although the first aspartate-rich motif ($D^{105}DXXXXD^{111}$) in the bacterial enzyme is different form that of the avian enzyme (DDXXD), the first and last aspartate residues of this motif in each enzyme make identical interactions with Mg^{2+}_A and Mg^{2+}_C. The third Mg^{2+}_B ion coordinates with the first aspartate in the second aspartate-rich motif (Asp244 of *E. coli* farnesyl diphosphate synthase). Thus, both motifs together bind a trinuclear magnesium cluster that triggers the departure of the diphosphate group and the initial carbocation formation.
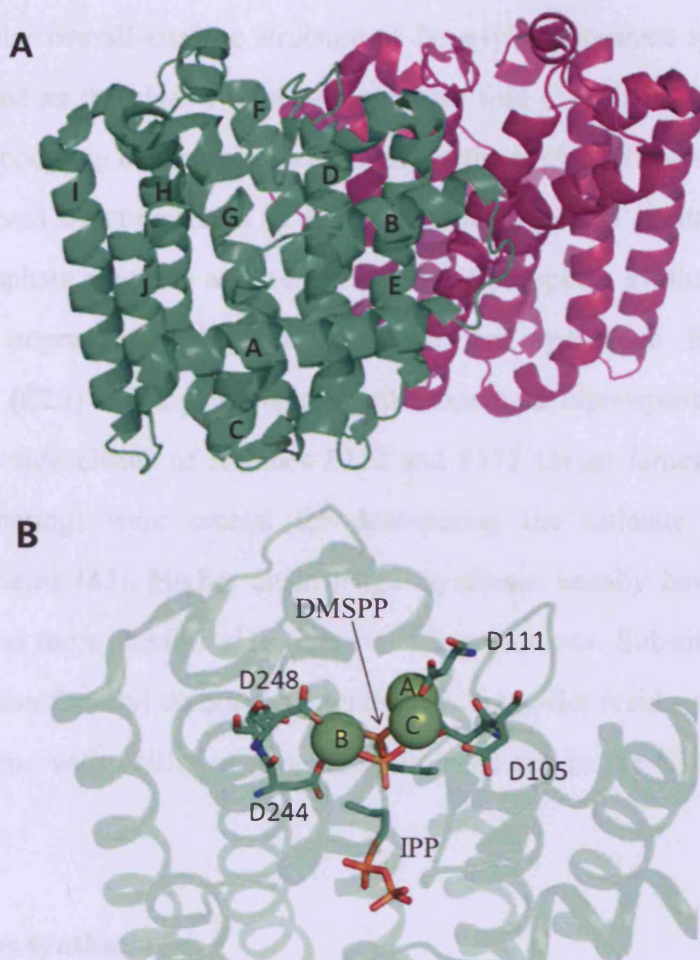
Figure 1.4: *Cartoon representations of* E. coli *farnesyl diphosphate synthase (PDB 1RQI) (A) and (B) the active site of the enzyme complexed with three magnesium ions (shown as green spheres), dimethylallyl-S-thiodiphosphate (DMASPP) and IPP*. Chain A of the dimer is coloured cyan and chain B is green. Helices of chain B are labeled with capital letters. Aspartate residues from the two magnesium binding motifs are shown as sticks.

The tertiary enzyme-Mg^{2+}-substrate complex observed from the crystal structure of farnesyl diphosphate synthase suggests that the first step of this chain elongation is the metal-triggered ionization of DMAPP, forming an allylic carbocation intermediate. Alkylation from the distal double bond of IPP to C1 of DMAPP generates a 10-carbon intermediate with a positive charge at C3 position. The first chain elongation that yields GDP is terminated by a stereospecific elimination of the pro-R hydrogen (*62*). The second chain elongation reaction takes place by addition of another molecule of IPP to GDP, forming farnesyl diphosphate at the end of the catalytic cycle (Scheme 1.1). The recently solved crystal structure of geranylgeranyl diphosphate synthase

displays a similar overall enzyme structure as farnesyl diphosphate synthase, which has been defined as the class I terpenoid synthase fold (*14*). In addition, they also share a similar coupling mechanism and a conformational change upon ligand binding was also observed by comparison of unliganded and liganded crystal structures of farnesyl diphosphate synthase and geranylgeranyl diphosphate synthase (*14, 62*). A study of 35 isoprenyl diphosphate synthases that synthesize farnesyl (C15), geranylgeranyl (C20) and higher chain length isoprenoid diphosphate revealed that the size of the side chains of residues F112 and F113 (avian farnesyl diphosphate synthase numbering) were crucial for determining the ultimate length of the hydrocarbon chains (*63*). Higher chain length synthases usually have amino acids with smaller and more flexible side chains at these positions. Substitutions in F112 and F113 of *avian* farnesyl diphosphate synthase with smaller residues also conferred the avian enzyme with ability of generating diphosphates in the C20 to C70 range (*63*).

1.4.2 Isoprene synthases

Isoprene represents the simplest structural unit of the terpene family. Roughly 100 Tg of isoprene is generated from natural sources each year (*64*) and it is suggested that isoprene emission protects against environmental stresses such as transiently high temperatures and oxidative damage (*65, 66*). According to the computational studies, isoprene helps stabilise lipid membranes against thermally induced phase transitions (*67*). It also influences the defense mechanisms of herbivore infested plants (*68, 69*). In addition, isoprene and its derivatives are currently being investigated for their industrial values in fuel science and materials science (*70, 71*).

Isoprene synthase (ISPS) catalyses the metal-dependent conversion from DMAPP (3) to isoprene (1). This simple elimination reaction catalysed by ISPS may go through an allylic carbocation intermediate as the initiation step proposed for most class I terpenoid synthases, or it could also proceed *via* concerted fashion. In either case, a

proper active site base is required, while no such candidate residue was spotted in the active site that could function as the general base from analysis of the crystal structure of ISPS from gray poplar hybrid *Populus* × *canescens* (PcISPS). Surprisingly, analysis of the structure of PcISPS complex with DMASPP that serves as the inhibitor in the crystal structure studies, which suggests the diphosphate group itself could serve as the general base (Scheme 1.8) (*72*). Specifically, the PcISPS-DMASPP complex indicates that the substrate analogue binds to the enzyme with a favourable seven-membered-ring chair-like conformation, which brings one of the charged oxygen atom close to the C3 atom of DMASPP.
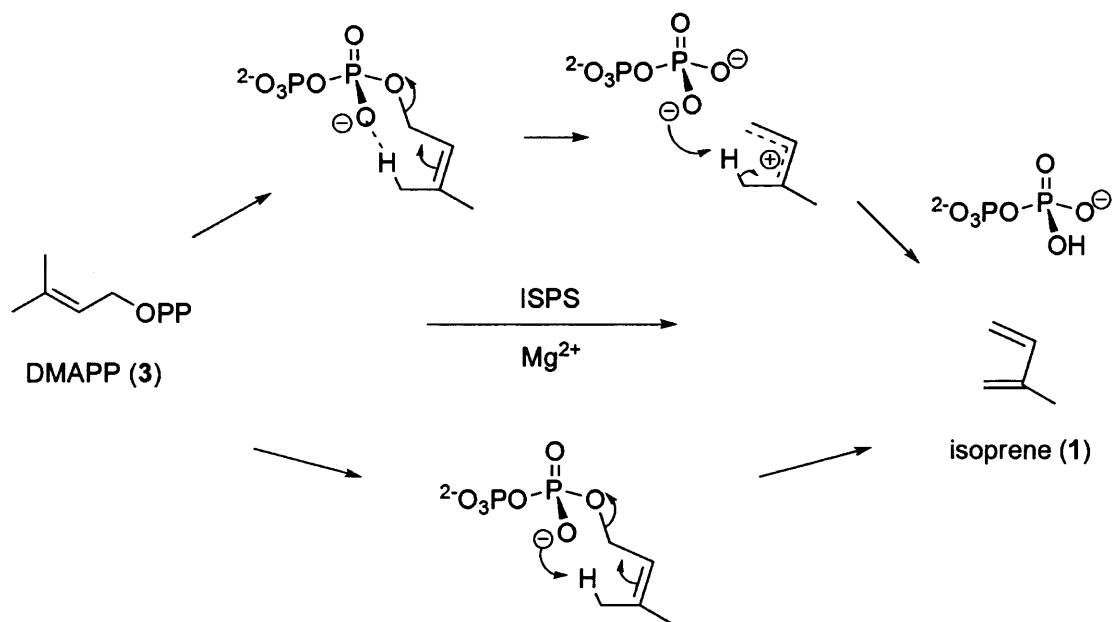
Scheme 1.8: *Proposed catalytic mechanism for ISPS (72).*

The crystal structure of PcISPS and the structure of its complex with the unreactive substrate analogue DMASPP and three Mg^{2+} ions were solved recently as homodimers (*72*). The PcISPS structure is constructed with two α-helical domains. Similar to the structure of farnesyl diphosphate synthase, the C-terminal domain of PcISPS is functionally active and adopts the α-helical class I terpenoid synthase fold. The N-terminal domain adopts an α-barrel class II terpenoid synthase fold with no known catalytic activity (Figure 1.5). The aspartate-rich metal binding motif ($D^{345}DXXD$) that is conserved in most class I terpene synthases was found on the

helix D of PcISPS and another characteristic NSE/DTE motif (N[489]DXXSXXXE) was observed on helix H (Figure 1.5). Both motifs chelate the trinuclear magnesium cluster, in which Mg^{2+}_A and Mg^{2+}_C are mainly coordinated by D345 and Mg^{2+}_B by E497 (Figure 1.5). The magnesium ions also coordinate with oxygen atoms of the DMASPP diphosphate group and this interaction could initiate the potential ionisation step.
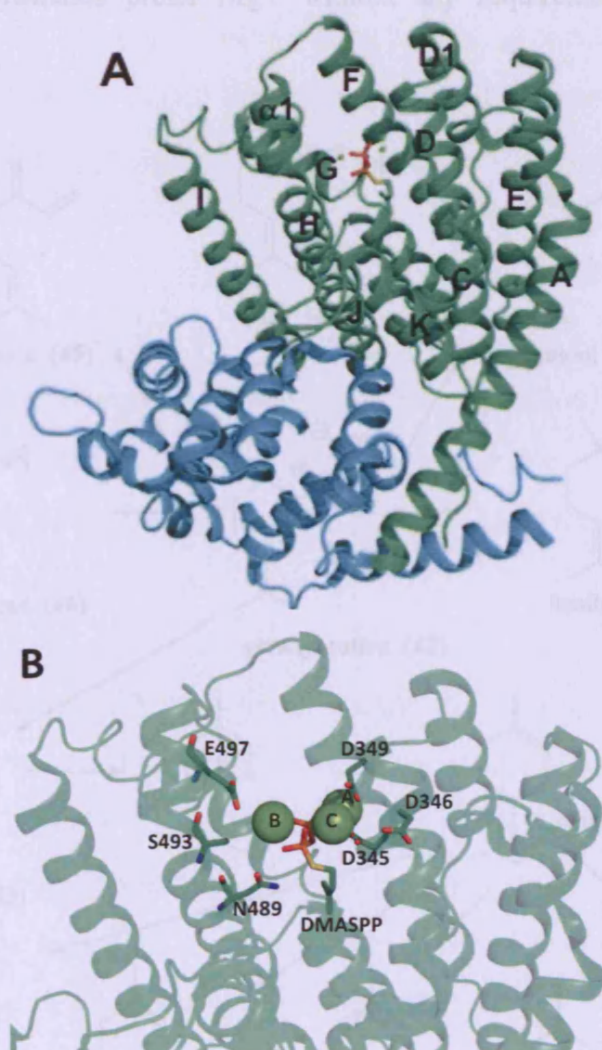
Figure 1.5: *Cartoon representations of chain B of PcISPS complexed with DMASPP (shown as sticks) (PDB 3N0G) (A) and the active site of the enzyme complexed with three magnesium ions (shown as green spheres) and DMASPP (B).* The helices of monomer are labeled with primed capital letters. Residues from the two magnesium binding motifs are shown as sticks.

1.4.3 Monoterpene synthases

Monoterpene synthases catalyze the metal-dependent formation of acyclic, monocyclic and bicyclic C10 olefins, alcohols, ethers and diphosphate esters (Scheme 1.9). Monoterpenes have been isolated almost exclusively from plants, including angiosperms and gymnosperms. The enzymes from gymnosperms require monovalent cations such as K^+ together with divalent cations Mn^{2+} or Fe^{2+} as cofactor, whereas the angiosperm synthases prefer Mg^{2+} without any requirement for monovalent cations (*4*).
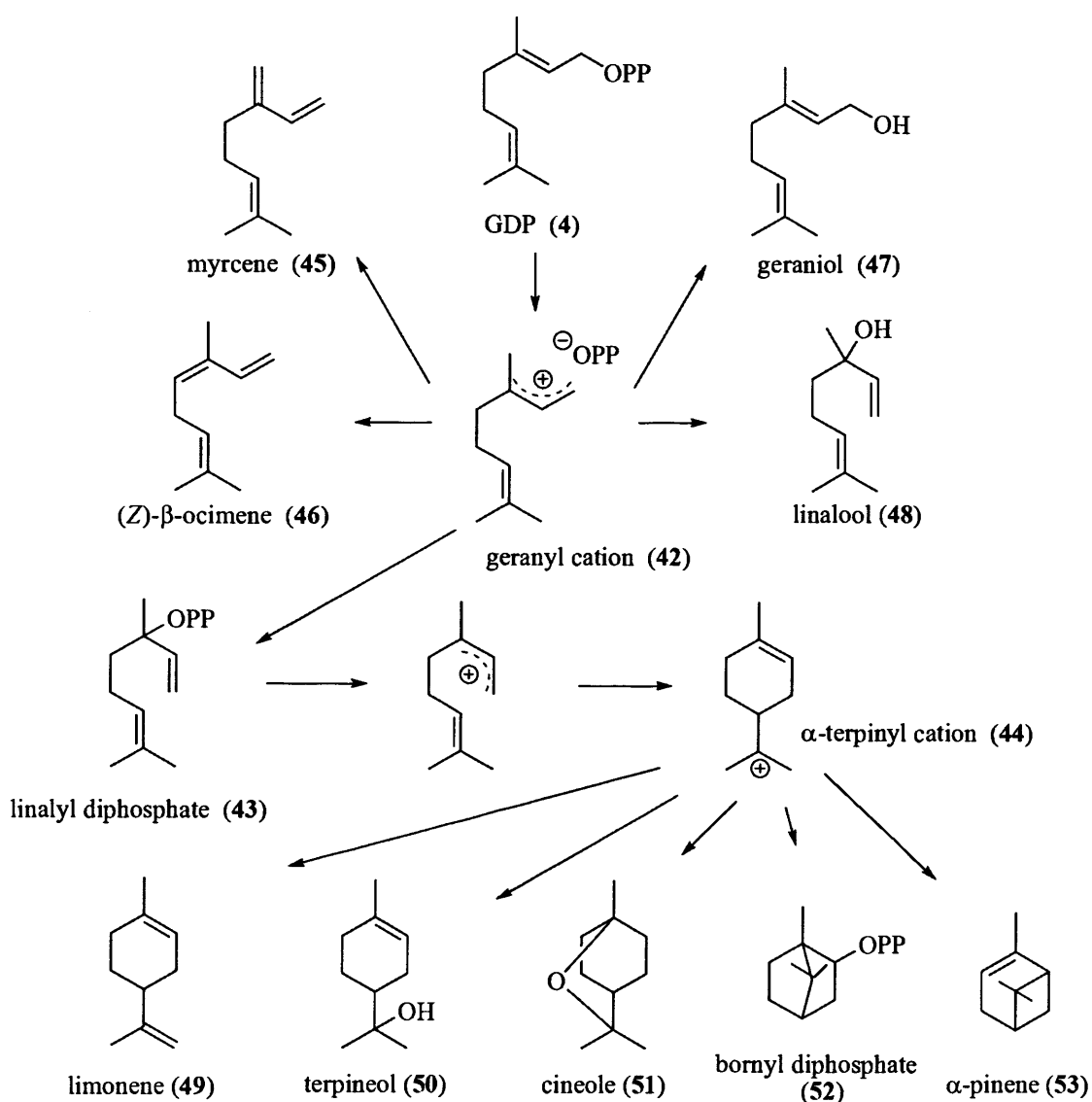
Scheme 1.9: *Biosynthesis of representative monoterpenes from GDP.*

All monoterpene synthases follow a common carbocationic reaction mechanism initiated by the departure of diphosphate group to form a geranyl cation (**42**). From here, acyclic products are formed by proton loss or addition of water (Scheme 1.9). The formation of cyclic products, on the other hand, needs preliminary conversion of the geranyl cation to linalyl diphosphate (LDP, **43**), facilitating the cyclization to a six-membered ring. Kinetic studies with LDP as an alternative substrate in combination with GDP support the intermediacy of LDP and also indicate the isomerization reaction as the slow step during catalysis (*73*). Subsequent electrophilic attack of C1 by the C6-C7 double bond of the cisoid conformer of LDP yields the universal monocyclic intermediate, α-terpinyl cation (**44**). From this intermediate, monoterpene synthases catalyse loss of protons, hydride shifts, water capture and internal cyclizations to form different structures (Scheme 1.9). The fact that the individual intermediate (**44**) may undergo multiple modifications suggests a rationale for the ability of monoterpene synthases to produce multiple products. Studies showed that the minor products of the monoterpene synthase have fewer cyclisations than the major product, suggesting that the minor products are formed from premature termination of the reaction cascade (*74*).

The first crystal structure of a monoterpene synthase was solved for the homodimeric (+)-bornyl diphosphate synthase (*75*). The monomer contains two α-helical domains (Figure 1.6). The C-terminal domain exhibits the characteristic class I terpenoid synthase fold, catalyzing the metal-dependent cyclization of GDP (Scheme 1.9) (*75*). The N-terminal domain is homologous to the glycosyl hydrolase domain and has no well-defined function; nevertheless, the N-terminal polypeptide segment caps the active site in the C-terminal domain upon ligand binding (*57*).

Two arginines (R55, R56 of (+)-bornyl diphosphate synthase) (Figure 1.6) have been found in many monoterpene synthases (*75*). Polypeptide deletion studies on the limonene synthase of *Mentha spicata* revealed that the enzyme activity was not affected by deletion of the N-terminal region before this tandem arginine motif.

Deletion of this motif rendered the limonene synthase unable to accept GDP as a substrate. However, LDP is still usable by the deletion mutant, suggesting that the arginine motif might participate in the isomerization of GDP to linalyl cation (76).
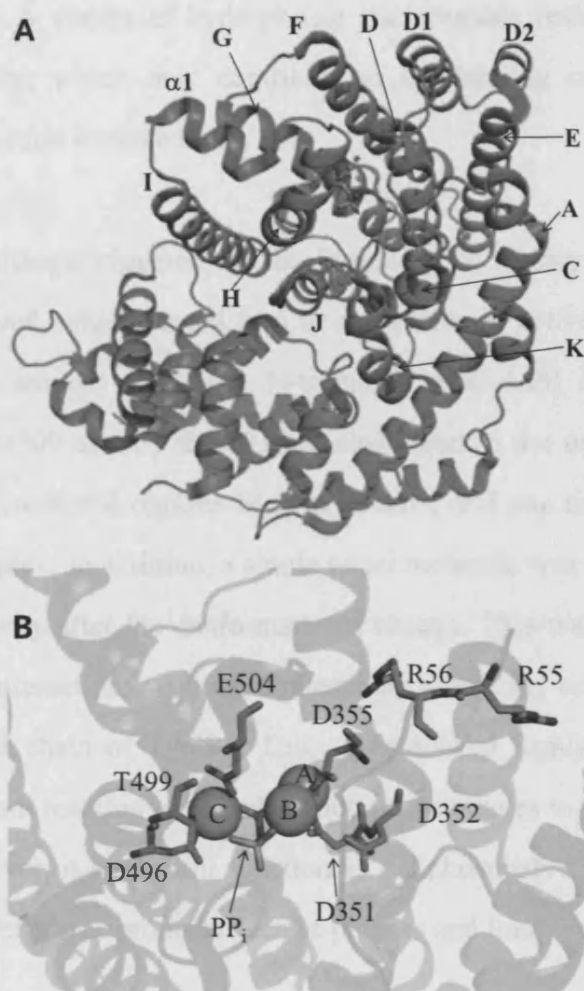
Figure 1.6: *Cartoon representations of chain A of bornyl diphosphate synthase complexed with diphosphate group (shown as orange sticks) (PDB 1N1Z) (A) and the active site of the enzyme complexed with three magnesium ions (shown as green spheres) and diphosphate group (PP$_i$) (B).* The helices of monomer are labeled with primed capital letters. Aspartate residues from the two magnesium binding motifs and the tandem arginines are shown as cyan sticks.

The structure of (+)-bornyl diphosphate synthase complexed with PP$_i$ reveals the important molecular recognition of the diphosphate moiety involving a magnesium cluster and several hydrogen bond interactions (Figure 1.6) (75). The first and last aspartate residues in the DDXXD motif of helix D, Asp351 and Asp355 contact Mg^{2+}$_A$ and Mg^{2+}$_C$. The second magnesium binding motif (NSE/DTE motif), Asp496,

26

Thr500 and Glu504, chelates Mg^{2+}_B on the opposite site of the active site cleft. This pattern is very similar to that observed with *E. coli* farnesyl diphosphate synthase and suggests a universal molecular recognition of isoprenoid diphosphates by class I terpenoid synthases. A variety of hydrophobic and aromatic residues were found in the active site cavity, which may contribute to the binding conformation of the substrate and carbocation intermediates.

A conformational change triggered by the formation of tertiary enzyme-PP_i-Mg^{2+}_3 complex was observed, which would lead to a sequestered active site pocket that is isolated from bulk solvent (*75*). The N-terminus (E50-A63) and loop segments D227-D234, T500-D509 and F578-S583 are disordered in the unliganded structure, while all of these disordered regions become ordered and cap the active site in the enzyme-ligand complex. In addition, a single water molecule was found to be trapped in the active site cavity after the conformational change. This water molecule makes hydrogen bonding interactions with the diphosphate group, the backbone carbonyl of Ser451 and the side chain of Tyr426. Thus, it is docked firmly in the active site, which prevents it from reacting with carbocation intermediates to prematurely quench the reactions. However, it could still function as a diphosphate-assisted general base to generate cyclic olefin side products such as pinenes and limonene (*77*).

1.4.4 Sesquiterpene synthases

The formation of sesquiterpenes from FDP follows similar carbocation-based reaction mechanisms as those of monoterpene synthases. However, the longer carbon chain skeleton of FDP and the additional double bond greatly increase the structural diversity of products, many of which contain bicyclic structures. Often, the first step of sesquiterpene formation is believed to be the ionization of FDP to the corresponding transoid farnesyl cation (**54**) (*5*). The reaction can be terminated here by proton abstraction or water capture to form acyclic products (Scheme 1.10).
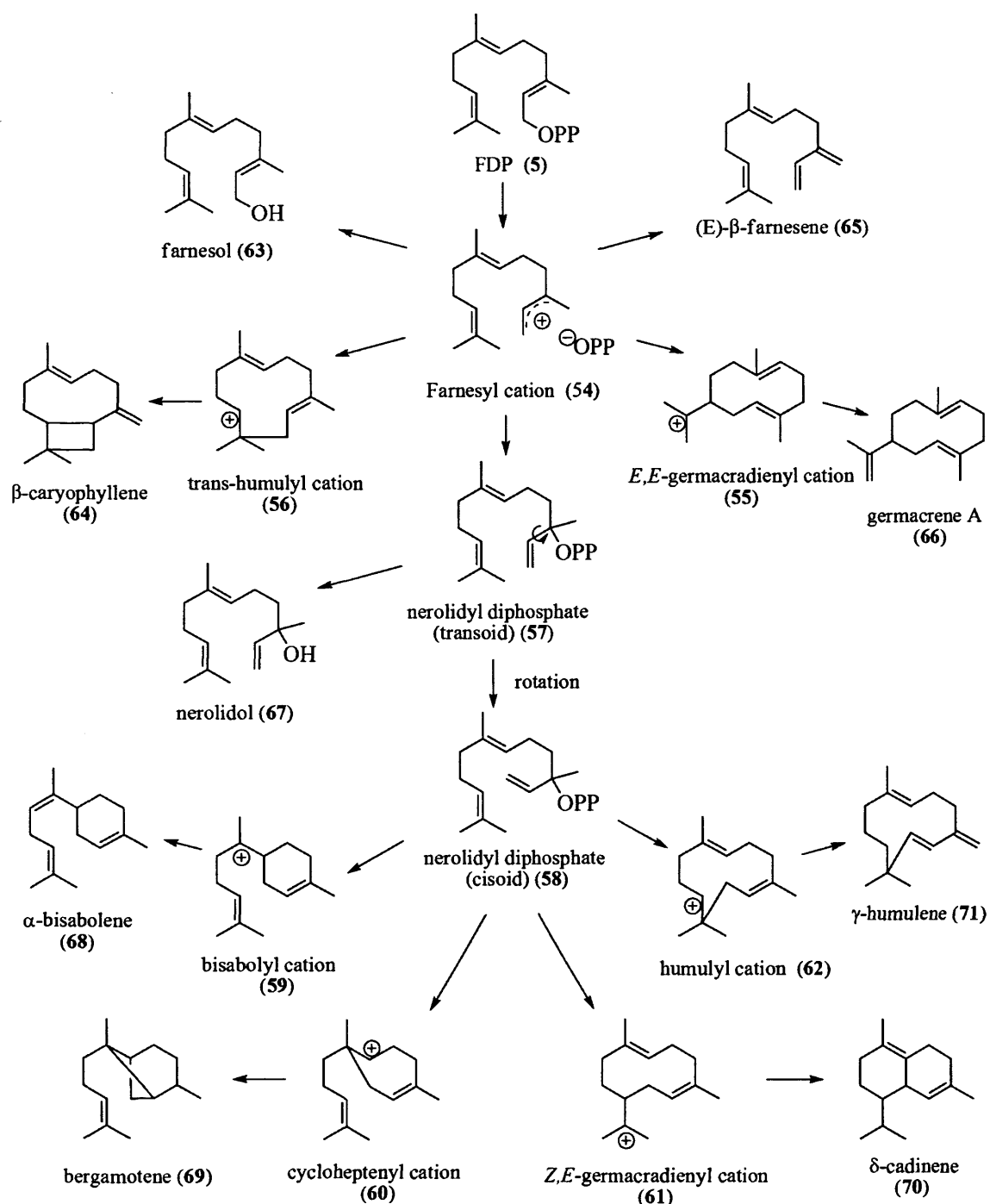
Scheme 1.10: *Biosynthesis of representative sesquiterpenes from FDP.*

From the resulting farnesyl cation, the cyclization reactions can go through three different pathways (Scheme 1.10). Reaction of the C1 cation with the distal double bond could occur by two different pathways forming either 10-membered ((*E,E*)-germacradienyl cation, **55**) or 11-membered ((*E*)-humulyl cation, **56**) rings. However, due to the (E)-geometry of the C2-C3 double bond of FDP, a

1,6-cyclization including the central double bond is precluded by steric constraints (*78*). To accomplish 1,6-cyclizations, sesquiterpene synthases generate the farnesyl cation (**54**) and recapture diphosphate group at C3 to form transoid nerolidyl diphosphate (**57**). A tightly bound cisoid nerolidyl diphosphate (**58**) is then formed *via* preliminary isomerization of the C2-C3 double bond. The cisoid conformer of the nerolidyl cation can then be cyclized from either the central or distal double bond to form 1,6-bisabolyl cation (**59**), 1,7-cycloheptanyl cation (**60**), 1,10-(Z,E)-germacradienyl cation (**61**) or 1,11-(Z)-humulyl cation (**62**) products. Subsequent cyclisations can involve electrophilic attack on either of two remaining double bonds. In addition to cyclization, the diverse modifications of the sesquiterpene skeletons also include hydride shifts, methyl migrations or rearrangements prior to termination of the reaction by deprotonation or nucleophile capture. Generally, the sesquiterpene synthases require divalent metal ions Mg^{2+} or Mn^{2+} as cofactor (*79, 80*).

Aristolochene synthase (AS) is the only known sesquiterpene synthase that does not follow the farnesyl cation (**54**) mechanism (Scheme 1.11). This was shown using fluorinated FDP analogues, mechanism based inhibitors and site-directed mutagenesis (*81, 82*).
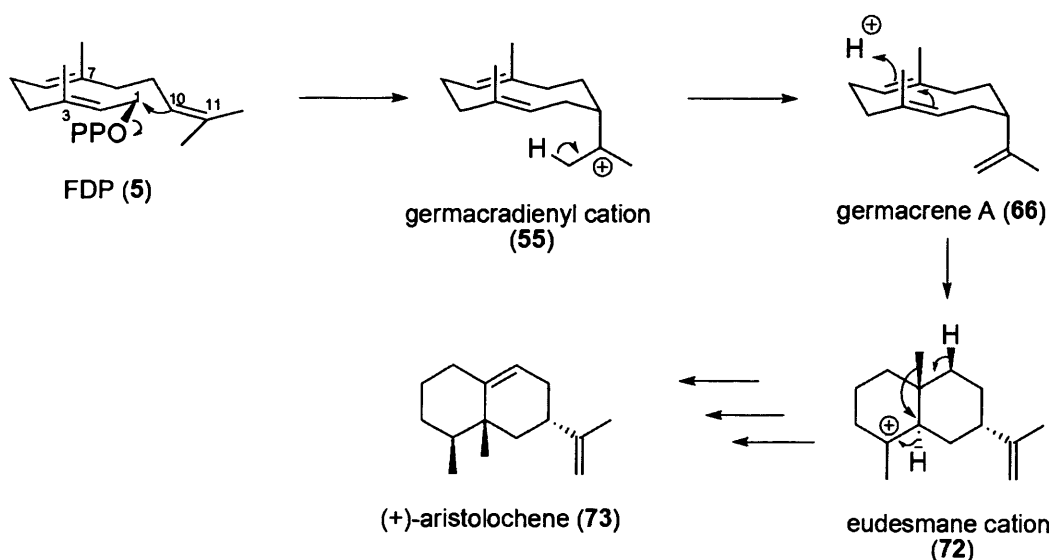
Scheme 1.11: *AS-mediated conversion of FDP to (+)-aristolochene.*

The mechanistic studies by employing substrate analogue 12,13-difluoro-FDP (**74**) rule out the possible existence of the farnesyl cation in the enzymatic reaction catalysed by AS from both *Penicillium roqueforti* and *Aspergillus terreus* (*81, 82*). Due to the electronic effects of fluoro substituents, the electrophilic attack of C1 by the C10, C11 π-bond would be inhibited if farnesyl cation was generated in the first place. In this case, the accumulation of the 12,13-difluorofarnesyl cation would go through an elimination, forming 12,13-difluoro-(3E,6E)-α-farnesene (**75**) and 12,13-difluoro-(E)-β-farnesene (**76**) instead (Figure 1.7). However, the observation that **74** acts as an competitive inhibitor of AS indicates that the formation of the germacradienyl cation (**55**) takes place in a concerted reaction, in which farnesyl diphosphate ionisation is accompanied by electrophilic attack of C1 by the C10, C11 π-bond. In addition, another fluorinated substrate analogue, 2-fluoro-FDP (**77**) has been shown to be a weak substrate (*81*), also suggesting the absence of farnesyl cation during the cyclisation pathway.

A

12,13-difluoro-(3E,6E)-α-farnesene
(**75**)

12,13-difluoro-FDP (**74**)

12,13-difluoro-(E)-β-farnesene
(**76**)

B

2-fluoro-FDP (**77**)
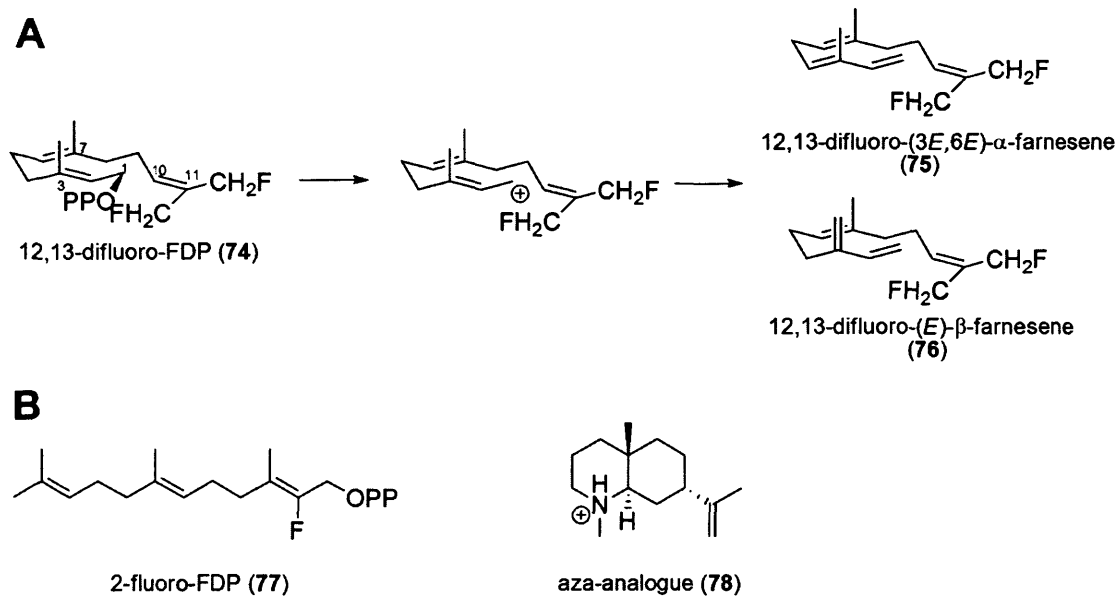
aza-analogue (**78**)

Figure 1.7: *(A) Outcome of the hypothetical mechanism for the conversion of 12,13-difluoro-FDP (74) through an initial ionization step. (B) Structure of 2-fluoro-FDP (77) and the aza-analogue (78) that mimics the eudesmane cation (72).*

Further investigations toward understanding the reaction mechanism of AS were

performed to identify the intermediates during the catalysis. Mutagenesis studies of aromatic residues Y92, F112 and F178 (*83-85*) suggest these residues play an important role in anchoring the distal double bond in the precatalytic conformation of FDP and hence influence formation of the neutral intermediate, germacrene A (**66**). Replacement of Y92 of *P. roqueforti* aristolochene synthase (PR-AS) with residues of decreasing steric bulk (*e.g.* Phe, Val, Cys and Ala) led to a increasing production of acyclic sesquiterpene products. The site-specific mutant AS-F112A produces 36% (*E*)-β-farnesene and 53.5% (*3E,6E*)-α-farnesene as the major products. Another mutant AS-F178V generates 50% germacrene A as the major product, which strongly suggests the appearance of germacrene A as an intermediate in the catalytic pathway. A recent study (*86*) further evaluated the contributions of aromatic residues by amino acid replacements in positions 88, 89 and 108 of PR-AS. These residues may have the potential to interact with the three aromatic residues described above (Y92, F112 and F178). The substitution of either V88 or T89 by the bulkier phenylalanine led to an increase in germacrene A production. Replacement of L108 with serine or alanine led to a suppression of the native aristolochene synthase activity. Instead, the enzymatic reaction gave a mixture of (*E*)-β-, (*3E,6E*)-α- and (*3Z,6E*)-α-farnesene as the predominant reaction products. These results demonstrated the existance of the germacrene A intermediate and the involvement of the aromatic and aliphatic residues around the active site of PR-AS.

The existence of eudesmane cation (**72**) had been probed for a long time with little evidence to support its existence. Our group recently provided an approach to identify **72** by employing the aza analogue (**78**) that effectively mimics the topological and electrostatic properties of the eudesmane cation (**72**) (*87*). **78** acted as a competitive inhibitor with a K_i value (0.35 ± 0.12 μM) similar to the Michaelis constant (K_M = 0.53 ± 0.21 μM)) determined for PR-AS. This observation greatly supports the intermediacy of eudesmane cation as an intermediate in AS catalysis. Mutagenetic approaches further demonstrated that the indole ring of active site residue W334 of PR-AS is responsible for stabilisation of the eudesmane cation (**72**) through cation-π

interactions (*88*). Substitutions of residues without aromatic side chains (*e.g.* Leu) greatly altered the product specificity towards the neutral intermediate germacrene A. Replacement of tryptophan by *para*-substituted phenylalanines with strong electron-withrawing groups (e.g. Cl, CF_3 and NO_2) had only minor effects on the K_M values but led to 30-250 folds reduction on the k_{cat} values and the main product distribution of these mutants shifted from aristolochene to germacrene A.

Bacterial and fungal sesquiterpene synthases such as aristolochene synthase from *Aspergillus terreus* and *Penicillium roqueforti*, pentalenene synthase from *Streptomyces* UC5319 and trichodiene synthase from *Fusarium sporotrichioides* are usually active as single domain enzymes (*89, 90*). Plant sesquiterpene synthases such as *epi*-aristolochene synthase from *Nicotiana tabacum* and δ-cadinene synthase from *gossypium arboretum* contain an additional N-terminal domain (*91*). The C-terminal domain of sesquiterpene synthases from bacteria, fungi and plants adopt the characteristic class I terpenoid synthase fold. This type of enzyme usually contains the aspartate-rich DDXXD(E) and another NSE/DTE metal ion binding motifs, suggesting the binding of the trinuclear magnesium cluster is required for catalysis. The additional N-terminal domain presents a fold most related to that of the glycosyl hydrolases and weakly homologous to the class II terpenoid synthase fold (*92*).

The crystal structure of aristolochene synthase from *Aspergillus terreus* is employed here to present the important tertiary structural features from microbial sesquiterpene synthases (*93*). Aristolochene synthase adopts the α-helical class I terpene synthase fold and is reported to be a monomer in solution (*94*). The crystal structure reveals the assembly of two dimers in the asymmetric unit to form a tetramer (Figure 1.8). The observed attenuation of enzyme activity at high concentrations (greater than 27 nM) could be due to the oligomerization (*95*).
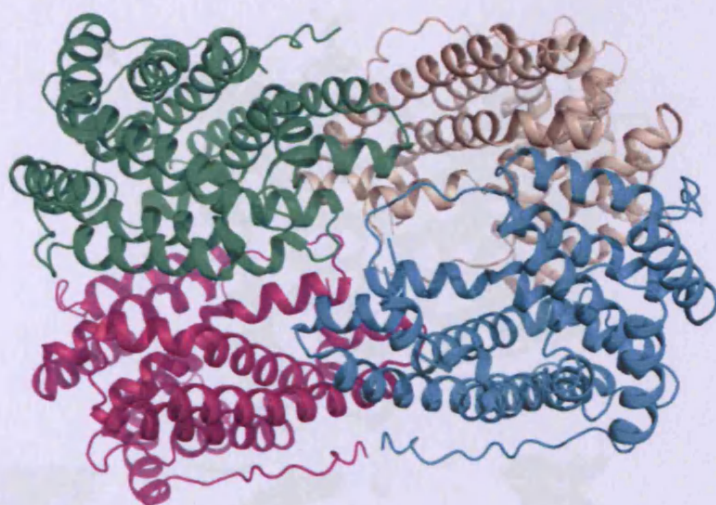
Figure 1.8: *Cartoon representation of tetramer of aristolochene synthase from A. terreus (PDB 2OA6).* Chain A is coloured tint, chain B is magenta, chain C is cyan and chain D is green.

The tertiary structure of *A. terreus* aristolochene synthase contains 13 α-helices, connected by short loops. The aspartate-rich $D^{90}DXXE$ motif is located on helix D on the upper wall of the active site cleft; the other metal binding motif, $N^{219}S^{223}E^{227}$ motif, is located on helix H on the opposite wall of the active site cleft. The enzyme-Mg^{2+}_3-PP_i complex was only observed in subunit D (Figure 1.9). In this complex, Asp90 was observed to coordinate directly to Mg^{2+}_A and Mg^{2+}_C. Asp91 makes a salt bridge interaction with Arg314, which donates a hydrogen bond to PP_i. Mg^{2+}_A and Mg^{2+}_C are also coordinated by PP_i and water molecules. The NSE motif coordinates to Mg^{2+}_B in a fashion identical to that observed in bornyl diphosphate synthase (see section 1.4.3). In addition to the metal coordination interactions mentioned above, the PPi anion also makes hydrogen bonds with some basic residues, including Arg175, Lys226, Arg314 and also aromatic residue Tyr315.
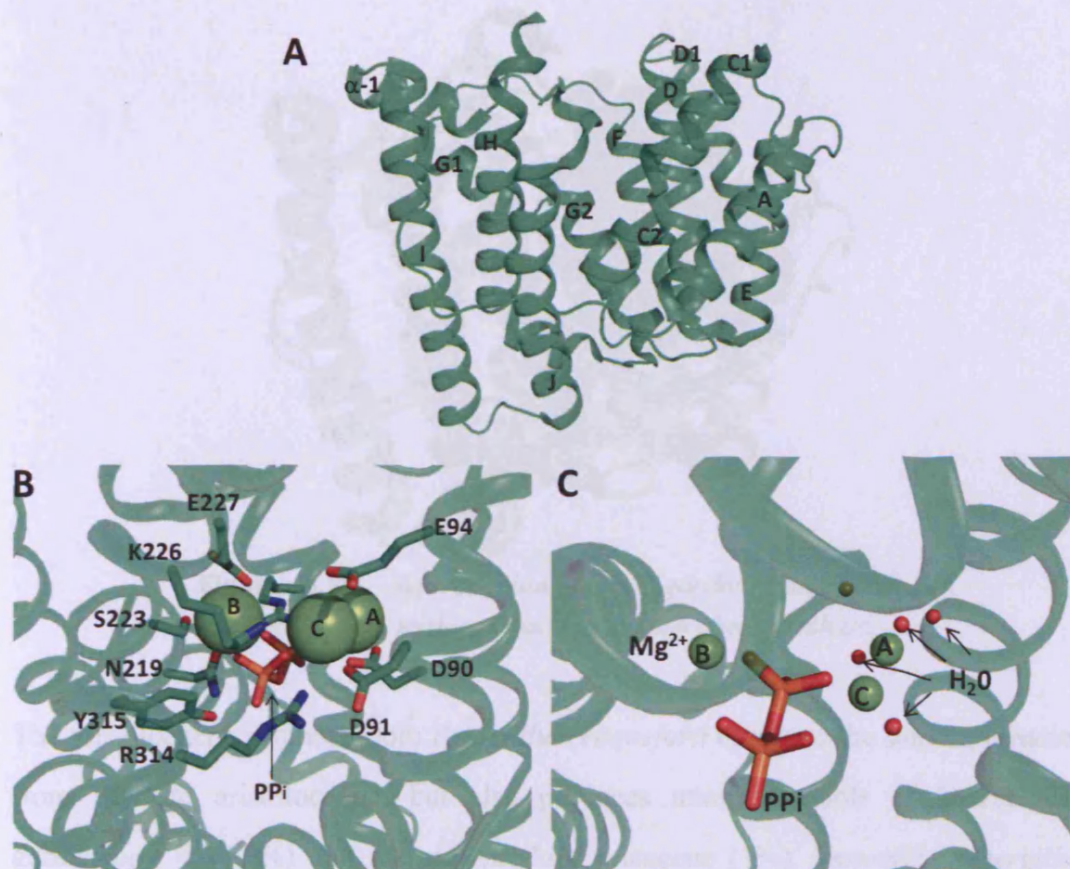
Figure 1.9: *Cartoon representations of chain D of aristolochene synthase (PDB 2OA6) (A), the active site of the enzyme complexed with three magnesium ions (shown as green spheres) and PP$_i$ (shown as sticks) (B) and magnesium ions (shown as green nb_spheres) coordinated with PP$_i$ and water molecules (shown as red nb_spheres) (C).* The helices of the monomer are labeled with capital letters. Residues from the two magnesium binding motifs and the ones ordinated with PP$_i$ are shown as sticks.

The binding of Mg$^{2+}_3$-PP$_i$ to monomer D triggers a significant conformational change, which causes helices C1, D1, F, G1 and H to shift inward and the previously disordered H-α-1 loop becomes ordered (Figure 1.10) (*93*). This conformational change leads to a closed active site, bringing ligand-binding residues closer together and sequestering the active site from bulk solvent.

Figure 1.10: *Superposition of unliganded (tint) and liganded (green) tertiary structures aristolochene synthase.*

The aristolochene synthase from *Penicillium roqueforti* catalyses the same cyclisation from FDP to aristolochene, but also produces minor amounts of intermediate germacrene A (7.5%) and the side-product valencene (1%). However, *Aspergillus terreus* aristolochene synthase generate exclusive aristolochene as the only product. The overall structure of both enzymes is very similar and they share 61% amino acid sequence identity. The majority of amino acid substitutions between the two enzymes are found on the protein surfaces, while most active site residues are highly conserved (*93*). This fact suggests that formation of the minor products may arise either from an adjustments of residues surrounded the active site (to change the active site shape and volume), or by rearrangements of hydrogen bond networks.

Unlike microbial sesquiterpene synthases, all plant sesquiterpene synthases contain two domains. The first plant sesquiterpene synthase for which a crystal structure was solved was *epi*-aristolochene synthase from *Nicotiana tabacum*, which displayed two α-helical domains (Figure 1.11) (*91*).
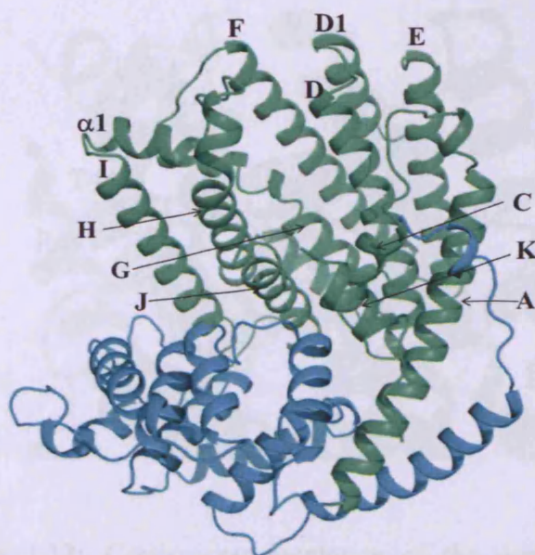
Figure 1.11: *Cartoon representation of 5-epi-aristolochene synthase (PDB 5EAS).* The C-terminal domain is coloured green and all the helices from C-terminal domain are labeled with capital letters. The N-terminal domain is coloured cyan.

The crystal structure of *epi*-aristolochene synthase complexed with farnesyl hydroxyphosphonate (Figure 1.12) showed that Asp301 and Asp305 from helix D (the first and the last residues from DDXXD motif) are liganded with $Mg^{2+}{}_A$ and $Mg^{2+}{}_C$, which is different from coordination interactions observed from trichodiene synthase but identical to that of (+)-bornyl diphosphate synthase. $Mg^{2+}{}_B$ is liganded by Asp444, Thr448 and Glu452 from the NSE/DTE motif (on the helix H) following the same pattern as had been observed for (+)-bornyl diphosphate synthase and trichodiene synthase. The variety of metal ions coordination with diphosphate group among these terpenoid synthases indicates a strategy for the evolution of divergent cyclization products in each enzyme active site. In addition, two arginines, R441 and R264, seem to facilitate binding of the diphosphate group by providing positive charge localized at the opening of the active site. The active site closure upon ligand binding was also observed, in which the movement of the A-C loop could promote ionization of FDP by interaction with R264. The movement of J-K loop and N-terminal region may also help to close the active site.

Figure 1.12: *Cartoon representation of the active site of 5-epi-aristolochene synthase complexed with three magnesium ions (shown as green spheres) and substrate analogue farnesyl hydroxyphosphonate (FHP) (PDB 5EAT).* Residues from the two magnesium binding motifs and two arginines are shown as sticks.

1.4.4.1 δ-Cadinene synthase

δ-Cadinene synthase (DCS) is a sesquiterpene cyclase that catalyses the cyclization of FDP to δ-cadinene (**79**). Most genes coding for DCS have been isolated from the cotton species (*59, 96-100*). In addition, another DCS gene has been isolated from melon (*Cucumis melo L.*) (*101*). In cotton plants, δ-cadinene is the biosynthetic precursor of gossypol (**80**), which consists of two highly oxygenated forms of δ-cadinene joined together by a binaphthyl linkage (Scheme 1.12). Gossypol is an important secondary metabolite among cotton plants, which defends the plant from bacterial and fungal pathogens (*16*). It accumulates in lysigenous glands found in stems, leaves and the cotyledons of seeds, resulting in cotton plants being much more resistant to microbe and insect predation. In addition, gossypol reduces the growth of cancer cells (*102*) and inhibits male fertility in humans (*99*). Desoxyhemigossypol (**81**) and hemigossypol (**82**) also accumulate in wild type cotton plants, serving as phytoalexins or defense compounds against microbial infection.
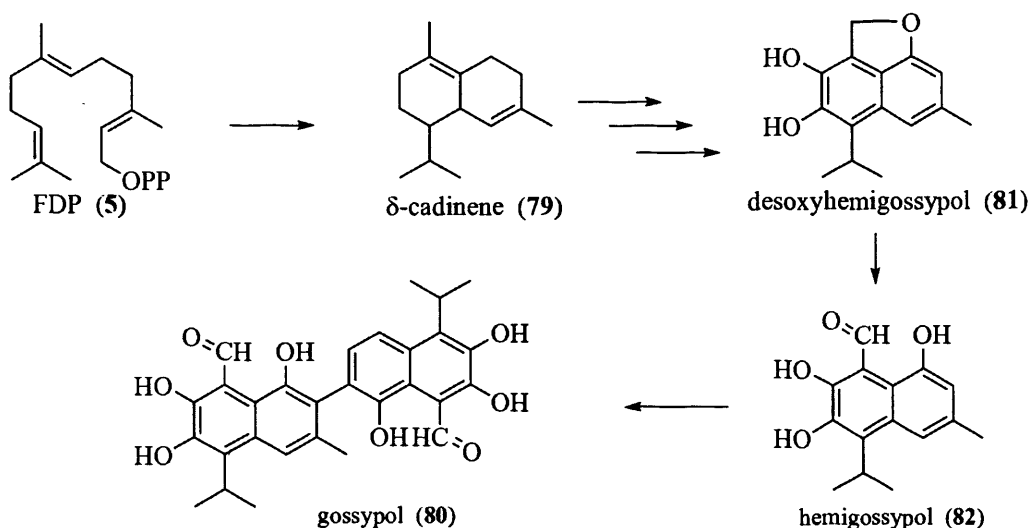
Scheme 1.12: *Biosynthesis of gossypol from δ-cadinene.*

The biosynthetic mechanism of generating δ-cadinene from FDP has been investigated by several groups (*99, 104-109*). It has been proposed that the mechanism begins with formation of NPP (**57**), followed by rotation of the C2, C3-bond to bring C1 within bonding distance of C10 (*107*) (Scheme 1.13). NPP then undergoes a C1-C10 ring closure to generate *cis*-germacradienyl cation (**61**) with the stereochemistry found in δ-cadinene. Prior to the second ring closure, the carbocation has to be repositioned within the macrocycle by a 1,3-hydride shift. The reaction is terminated by deprotonation from C6 to release δ-cadinene (**79**).
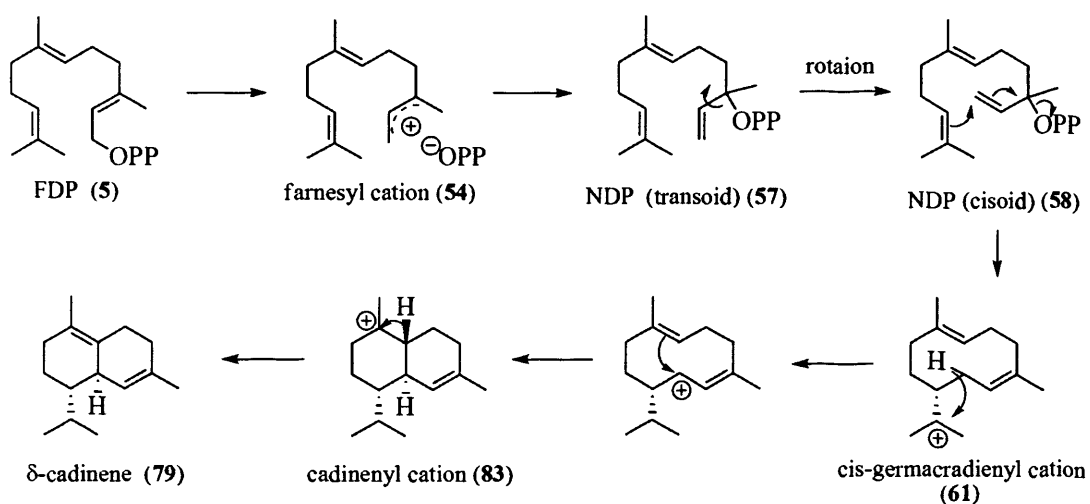
Scheme 1.13: *Proposed catalytic mechanism for δ-cadinene synthase via NDP as intermediate.*

In agreement with the proposed mechanism, NPP (57) has been shown to be a substrate of DCS (*104*, *105*) and its formation within the active site of the enzyme has been demonstrated by [1,2-^{13}C$_2$]acetate and [2-^{14}C] and [4-^{14}C]mevalonate labelled feeding experiments (*108*, *110*).

The biosynthesis of cadinene-type sesquiterpenes could also be achieved *via* a 1,10-cyclisation but without going through NPP (57) as intermediate (*107*) (Scheme 1.14). In this alternative pathway, the first ring closure goes through a direct displacement of the diphosphate group by the distal double bond of FDP, forming a *trans*-germacradienyl cation (55). The neutral germacrene D (84) is formed by a 1,3-hydride shift and deprotonation from C15 (*105*). The second ring closure upon reprotonation of germacrene D leads to the formation of cadinanyl cation (83), as in the NDP pathway, followed by deprotonation to form the final product.
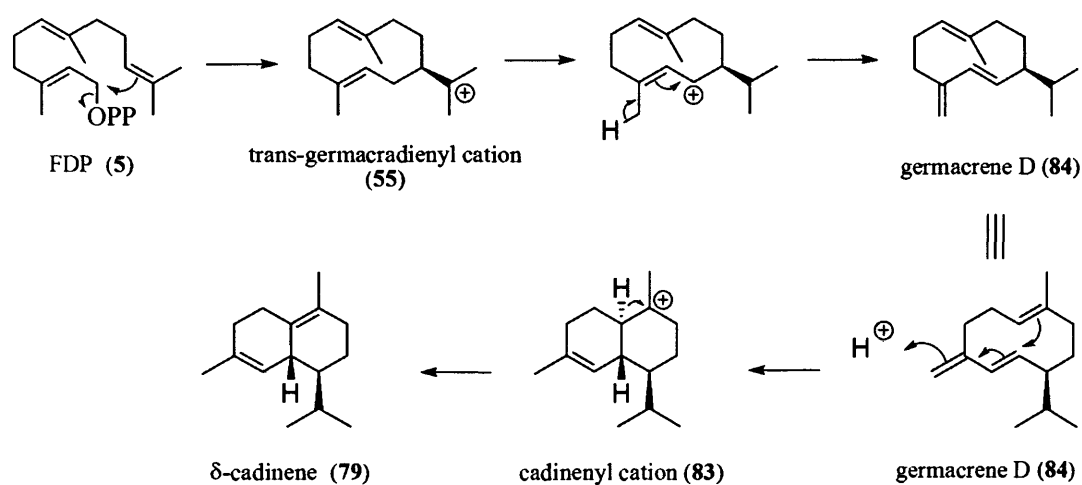
Scheme 1.14: *Proposed catalytic mechanism for δ-cadinene synthase via germacrene D as intermediate.*

Formation of δ-cadinene (79) *via* this proposed mechanism was first suggested by Arigoni (*107*) based on mechanistic studies of avocettin formation *via* γ-cadinene (*107*). Tritium-labelling experiments showed that all-*E*-FDP was not isomerized to the 2,3-*Z* isomer. Instead, a 1,3-hydride shift was observed, which led to the postulation of the intermediacy of germacrene D. In this mechanism, there is no interconversion

between (2*E*,6*E*)-FDP and its *Z*-isomer NDP, while the conversion of the *E*-double bond to a *Z*-double bond is achieved by a ring flip of germacrene D.

However, a recent investigation of the catalytic mechanism of DCS using a variety of substrate analogues strongly challenged the mechanisms described above (*109*). The catalytic mechanism of DCS involving germacrene D as an intermediate was ruled out by investigating the hypothetical protonation step. Since formation of δ-cadinene along this pathway would require protonation of the exocyclic double bond of germacrene D, the reaction cascade is expected to be pH-dependent. An isotopic exchange with the medium is also expected. However, the production of δ-cadinene was shown not to be pH-dependent and the isotopic exchange with the media was not observed either by incubations of DCS with FDP and [15-^{2}H$_3$]FDP (**85**) (Figure 1.13) in deuterated and aqueous buffer respectively.

[15-^{2}H$_3$]FDP (**85**)

(2*Z*,6*E*)-2F-FDP (**86**)

(2*E*,6*E*)-2F-FDP (**87**)

6F-FDP (**88**)
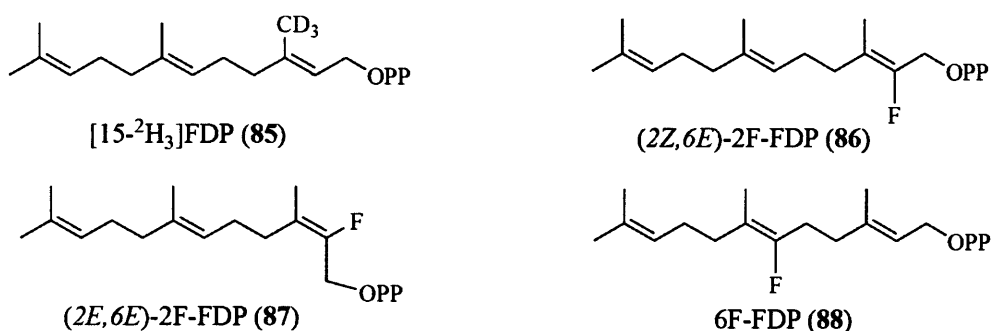
Figure 1.13: *Structures of FDP analogues.*

Investigation of the NDP pathway was first carried out by employing fluorinated substrate analogues. (2*Z*,6*E*)-2F-FDP (**86**) and (2*E*,6*E*)-2F-FDP (**87**) that contain a fluorine atom on C2 would greatly prevent the initial ionization step *via* depletion of electron density in the allylic moiety (*111*). Indeed, **86** and **87** acted as efficient inhibitors, suggesting the involvement of NDP (**57**) during the catalysis. Another fluorinated compound 6F-FDP (**88**) was designed to inhibit the second ring closure and this compound has been demonstrated to disrupt the catalytic cycle of AS (*112*)

and TEAS (*113*). Thus, if the catalytic mechanism of DCS goes through intermediates **61** and then **83**, the corresponding deprotonated forms of **61** (**89, 90** and **91**) would be expected (Scheme 1.15). While, incubation of 6F-FDP (**88**) with DCS generates no sesquiterpene product, suggesting an early involvement of the central C6, C7 double bond in the catalysis.

6F-FDP (**88**)

6F-helminthogermacrene A
(**89**)

6F-germacrene D
(**90**)
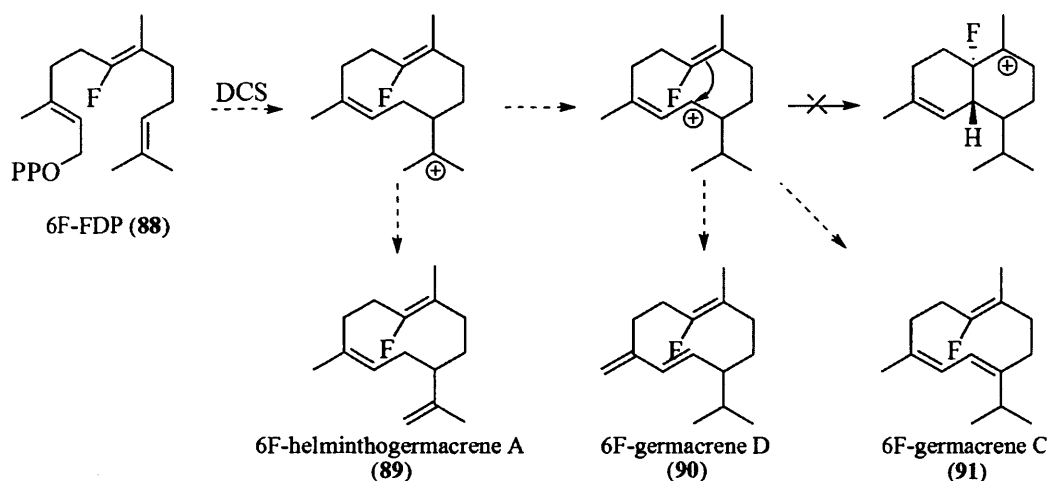
6F-germacrene C
(**91**)

Scheme 1.15: *Outcome of the hypothetical derailment of the DCS catalytic cycle with 6F-FDP at the germacrene cation stage.*

Additional evidence supporting the possible 1,6-ring closure as the first ring closure step in δ-cadinene biosynthesis was obtained from incubations with (2*Z*,6*E*)-FDP (**92**). Substrate analogue **92** can be recognised as the isomer of all-tran FDP and cisoid NDP, thus it should present as cisoid farnesyl cation after ionisation by DCS, which will greatly facilitate either 1,6 or 1,10 ring closure. DCS convert **92** into a mixture of the bisabolenes (**93, 94** and **95**) and δ-cadinene (40%) (Scheme 1.16), which is comparable to that observed with feeding NDP directly to DCS.

Z-γ-bisabolene (93)   β-bisabolene (94)   Z-α-bisabolene (95)

(2Z,6E)-FDP (92)   +
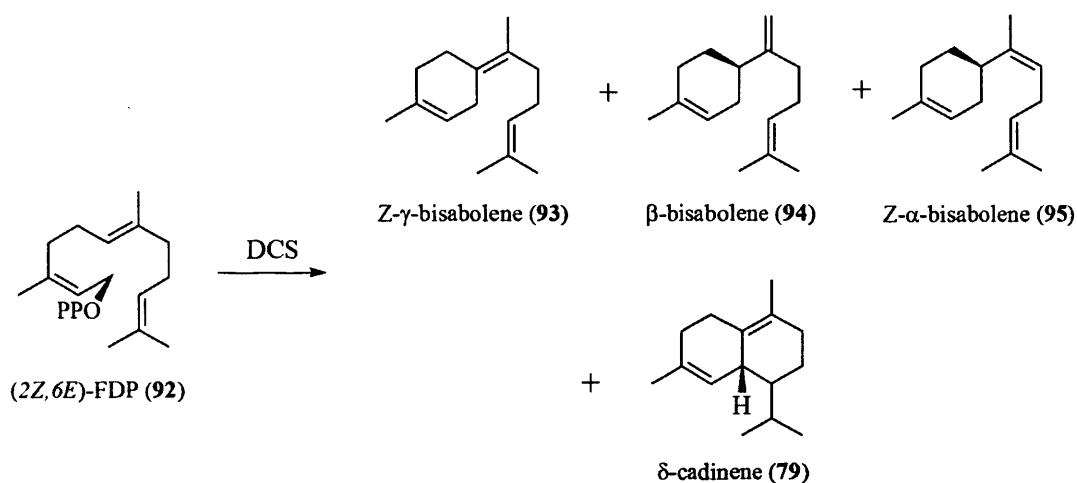
δ-cadinene (79)

Scheme 1.16: *Hydrocarbons generated by incubation of DCS with (2Z,6E)-FDP.*

These results, especially the inhibition of the enzyme by 6F-FDP, indicate a 1,6-ring closure mechanism *via* NPP (**58**) and α-bisabolyl cation (**96**) as a plausible means in δ-cadinene biosynthesis (Scheme 1.17).

(2E,6E)-FDP (5)   cisoid-NDP (58)   α-bisabolyl cation (96)

δ-cadinene (79)   cadinenyl cation (83)
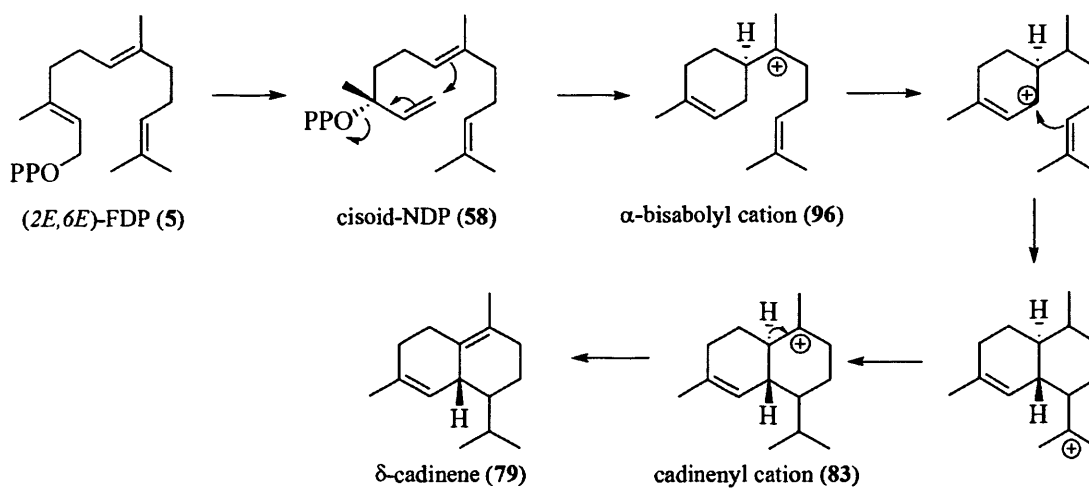
Scheme 1.17: *Possible biosynthesis of δ-cadinene via NDP and α-bisabolyl cation.*

1.4.4.1.1   Characterisation of DCS

The cDNA encoding DCS for our study was obtained from *Gossypium arboretum* (tree cotton) (*59*). The gene encodes 554 amino acids corresponding to a protein with a molecular weight of 64096. The crystal structure of this protein has recently been solved, revealing a two domain overall structure (Figure 1.14) (*114*).
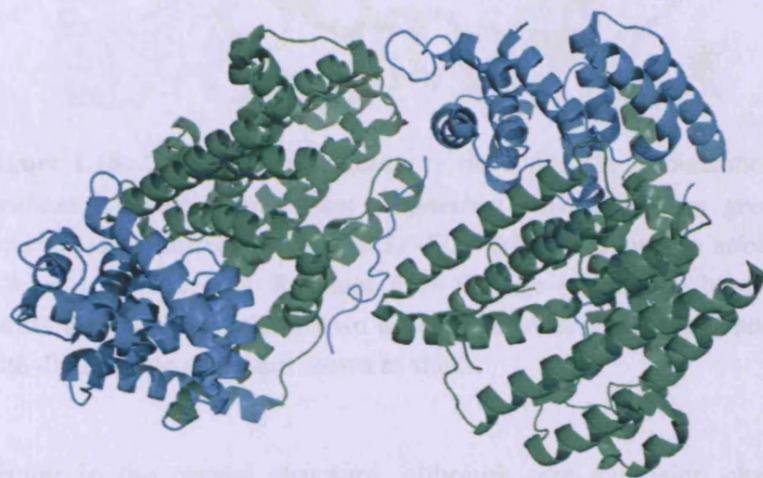
Figure 1.14: *Cartoon representation of δ-cadinene synthase (PDB 3G4D).* The C-terminal domain is coloured green and the N-terminal domain is coloured cyan.

The C-terminal domain is defined as the catalytically active domain, which adopts the α-helical class I terpenoid synthase fold. The active site consists of six α-helices, which form an 18 Å deep active site cleft (Figure 1.15). The crystal structure confirms that the characteristic aspartate-rich motif ($D^{307}DTYD^{331}$) is located on helix D, which interacts with Mg^{2+}_A and Mg^{2+}_C. However, the typical second magnesium binding motif (NSE/DTE) is replaced by a unique second aspartate-rich motif, $D^{451}DVAE^{455}$ that chelates Mg^{2+}_B. In terms of this feature, DCS is more similar to the farnesyl diphosphate synthase, which also contains two aspartate-rich motifs, rather than the majority of class I terpenoid synthases. The additional N-terminal domain of DCS adopts an α-helical fold, which shows some similarity to a class II terpenoid synthase fold (*92*). The function of this domain is still unknown, although it has been proposed to cap the active site during catalysis (*57*).

43
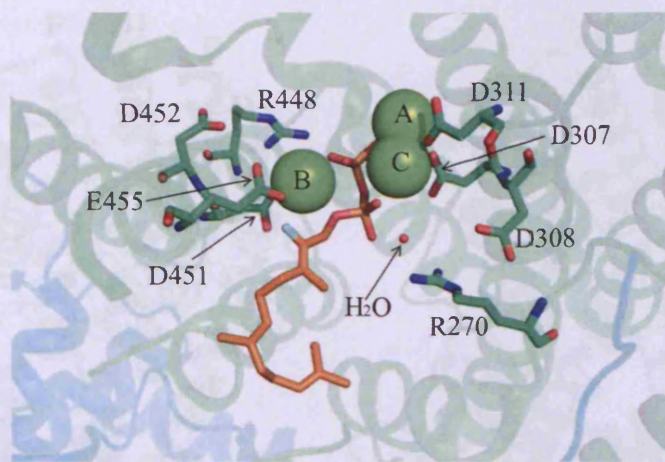
Figure 1.15: *Cartoon representation of the active site of δ-cadinene synthase complexed with three magnesium ions (shown as green spheres) and substrate analogue 2F-FDP (shown as orange sticks) (Chain A, PDB 3G4F).* Residues from the two magnesium binding motifs are shown as sticks. Two arginine residues that have contact with diphosphate group are shown as sticks.

DCS is a dimer in the crystal structure, although size exclusion chromatography experiments show that DCS is a monomer in solution, suggesting that the crystallographic dimer is not biologically relevant (*106*). In monomers A and B, several polypeptide regions are disordered and excluded from the final model, including the N-termini (M1-K24 and M1-P29, respectively) and the loop segments K42-I44 and K42-D45, F460-D464 and F460-D464, and G530-T534 and Y533-V536, respectively (Figure 1.16). Most of these segments are adjacent to the entrance of the active site in the C-terminal domain except the segments K42-I44 or K42-D45. According to previous studies of monoterpene and sesquiterpene synthases, one or more such disordered polypeptide segments become ordered upon ligand binding and the conformational change leads to the closure of active site that shields the active site pocket from bulk solvent (*1, 57*).
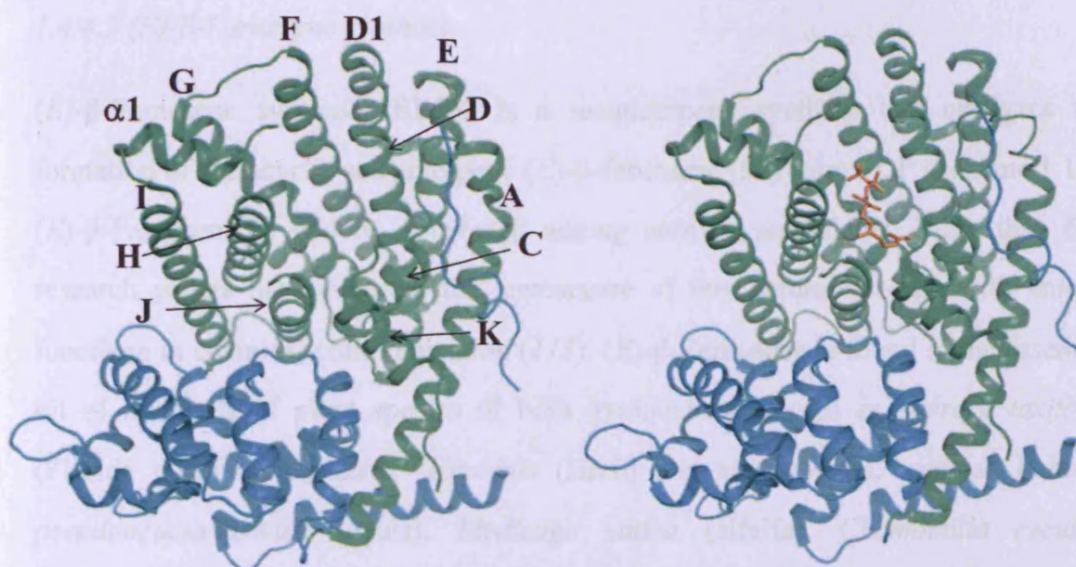
Figure 1.16: *Cartoon representations of δ-cadinene synthase (left) (PDB 3G4D) and the enzyme complexed with substrate analogue 2F-FDP (right) (PDB 3G4F).* The C-terminal domain is coloured green. The helices of C-terminal domain of unliganded enzyme are labeled with capital letters. The N-terminal domain is coloured cyan. Substrate analogue 2F-FDP is shown as orange sticks.

In the case of DCS, binding of the magnesium cluster and the substrate analogue 2F-FDP do not lead to an obvious conformational change; the N-terminal region M1-K24, loop segments K42-I44, F460-D464 and G530-T534 remain disordered (the structures of these polypeptides did not show up in the electron-density map) in each monomer (Figure 1.16) (*114*). This suggests that the conformational changes between liganded and unliganded catalytic domains of plant enzymes are relatively smaller than those observed from fungal enzymes, which may relate to the additional N-terminal domain.

The binding of the substrate analogue 2F-FDP to each monomer is similar. Some differences are related to the formation of hydrogen bonds. In monomer A, an oxygen atom of the diphosphate group appears to form a hydrogen bond with R448 and another charged oxygen atom donates a hydrogen bond to a water molecule, which accepts a hydrogen bond from R270 (Figure 1.15). In monomer B, the corresponding interactions are too long to be defined as hydrogen bonds.

1.4.4.2 (E)-β-Farnesene synthase

(*E*)-β-Farnesene synthase (EBFS) is a sesquiterpene synthase that catalyzes the formation of the acyclic sesquiterpene (*E*)-β-farnesene (**57**) from FDP (Scheme 1.18). (*E*)-β-Farnesene is widely distributed among animals and plants. More than 600 research papers have reported the appearance of this natural product with unique functions in chemical communication (*115*). (*E*)-β-Farnesene is found in the essential oil of hundreds of plant species of both gymnosperms, such as *Torreya taxifolia* (Florida torreya) and *Larix leptolepis* (larch) and angiosperms, such as *Robinia pseudoacacia* (black locust), *Medicago sativa* (alfalfa), *Chamomilla recutita* (chamomile), *Vitis vinifera* (grapes), *Cannabis sativa* (hemp), *Zea mays* (corn), *Piper nigrum* (black pepper), *Daucus carota* (carrot) and *Mentha x piperita* (peppermint), as reviewed (*115*).

In insects and plants, (*E*)-β-farnesene serves as a semiochemical in most cases. It was found as a pheromone (chemical substances could trigger a social response in members of the same species) in the urine of male mice (*116*) and is also emitted by andrenid bees (*117*) and by several genera of ants (*118*) to serve both as a defensive allomone (chemical substance that produced and released by an individual of one species of organism that affect the behaviour of a member of another species to benefit of the emitter) and as a trail pheromone. As an allomone, it functions as a feeding stimulant to the sand fly *Lutzomyia longipalpis*, which is an important vector of the blood disease leishmaniasis (*119*). The compound can serve as a prey-finding kairomone (chemical substances that are emitted by an organism in a way that benefits an individual of another species which receives it, without benefiting the emitter) among several species of predatory carabid beetles (*120*).

(*E*)-β-Farnesene also plays an important role in chemical communication of plants and animals. The compound stimulates kairomonal oviposition of the European corn borer when emitted by corn (*121*). Increasing amounts of (*E*)-β-farnesene were

produced and released by corn when it was attacked by lepidopterans. (E)-β-farnesene then serves as a synomone (chemical substances that are producted by one organism that affects the behaviour of a member of another species) and can be detected by the parasitic wasp to locate the hosts (122).

The most significant role of (E)-β-farnesene in plant-insect interactions is as an alarm pheromone in aphids (123). Aphids are one of the most destructive pests in agriculture, often causing plant virus infections in their host. Because of the development of insecticide resistance and the ability of some species to reproduce asexually, it is hard to control their population. Aphids become agitated and disperse from their host plant when exposed to (E)-β-farnesene (124), preventing colonization. High concentrations of (E)-β-farnesene can also be toxic to aphids (125). However, little progress has been made in controlling aphid population by application of (E)-β-farnesene to plants, mainly due to its volatility and instability. The unsaturated carbon chain structure can be easily oxidized in air.

In common with most sesquiterpenes, the formation of (E)-β-farnesene catalysed by EBFS was proposed to be initiated by loss of the diphosphate group, leaving an allylic carbocation with positive charge delocalized between C1 and C3 (115). The reaction is then terminated by proton abstraction from the C15 methyl group to yield (E)-β-farnesene (Scheme 1.18).

FDP (5) (E)-β-farnesene (57)
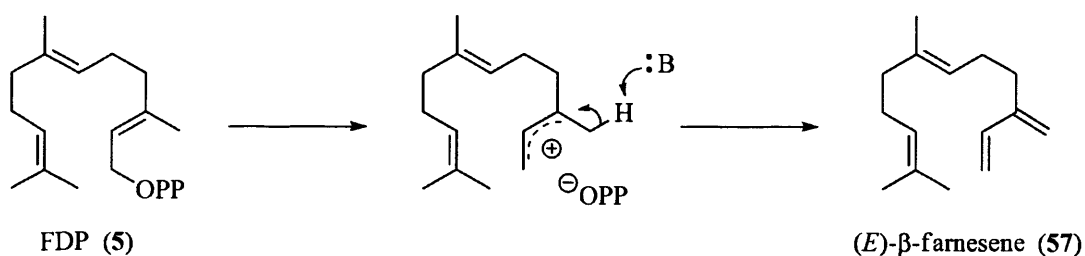
Scheme 1.18: *Proposed catalytic mechanism for (E)-β-farnesene synthase.*

During the last 15 years, cDNAs encoding EBFS have been isolated from *Mentha x piperita* (*115*), *Citrus junos* (*126*), *Artemisia annua* (*127*), *Pseudotuga menziesii* (*128*), *Zea mays* (*129*). The enzymes from *Citrus junos*, *Artemisia annua* and *Pseudotuga menziesii* generate essentially one product from FDP, while the enzyme from *Mentha x piperita* generates (*E*)-β-farnesene (**65**) (~85%) along with (*Z*)-β-farnesene (**97**) (8%), δ-cadinene (**79**) (5%) and another three unidentified minor products and the maize (*Zea mays*) enzyme produces (*E*)-β-farnesene (~50%), (*E*)-α-bergamotene (**98**) (~30%) and the other four minor products (**94, 99, 100** and **101**) (Figure 1.17). Formation of multiple products by a single enzyme is a common feature of terpenoid synthases and this fact may be caused by the typical electrophilic reaction mechanism. A number of highly reactive carbocationic intermediates were generated during enzymatic reaction, which would increase the chance of side reactions to produce minor products (*78*).

A

(*E*)-β-farnesene (**65**)   (*Z*)-β-farnesene (**97**)   δ-cadinene (**79**)

B

(*E*)-α-bergamotene   sesquisabinene (**99**)   zingiberene (**100**)   β-bisabolene (**94**)   sesquiphellandrene (**101**)
(**98**)
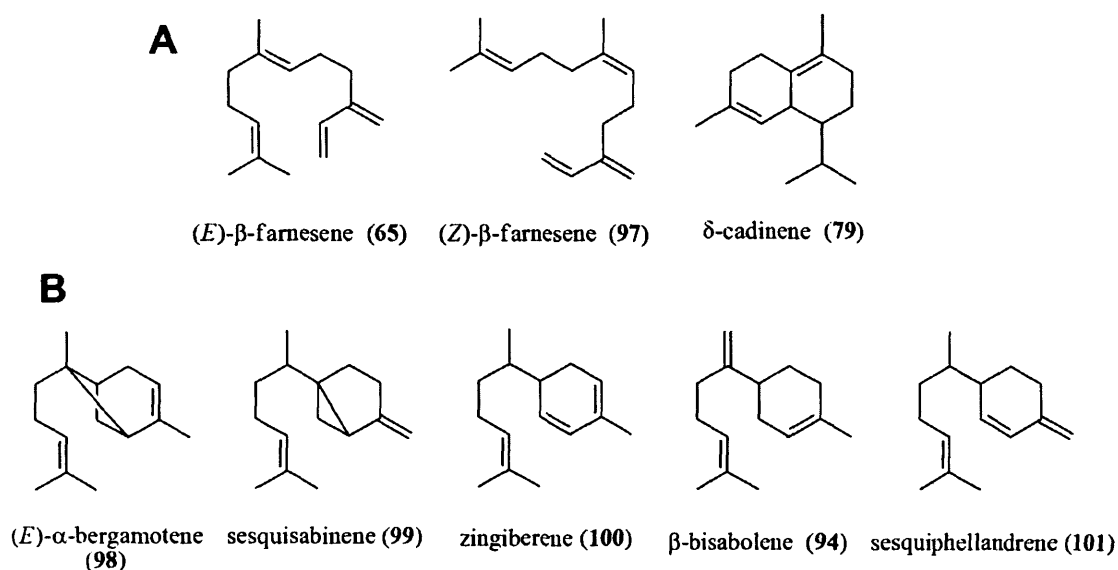
Figure 1.17: *Structure of hydrocarbons generated by EBFS from* Mentha x piperita *(A) and* Zea mays *(B)*.

Comparison of the amino acid sequence of EBFS from *A. annua* with EBFSs from other sources reveals an identity/similarity between 21/39% and 42/61%. However, the amino acid sequence of EBFS from *A. annua* shows considerably high similarity to other sesquiterpene synthases from the same plant (the four *Artemisia*

48

sesquiterpene synthases show an identity/similarity between 47/65% and 59/78% (*127*)). This fact suggests that the level of amino acid sequence similarity between the sesquiterpene synthases containing identical catalytic specificities from different species can be lower than the similarity between sesquiterpene synthases catalysing completely different reactions but from the same plant.

The cDNA encoding EBFS from peppermint (*Mentha x piperita*) was utilized in our studies. The 1653 bp open reading frame encodes a protein of 550 amino acids with deduced molecular mass of 63829. Like most other known class I terpenoid synthases, the magnesium-binding aspartate-rich ($D^{301}DXXD^{305}$) and $N^{444}S^{447}E^{452}$ motifs are present (*57*).

The deduced amino acid sequence of EBFS is quite similar to some other plant sesquiterpene synthases (see Appendix). The primary structure of EBFS is most similar to that of *epi*-aristolochene synthase from tobacco (exhibiting identity/similarity of 49/62%). It also closely resembles three other angiosperm sesquiterpene synthases: vetispiradiene synthase from *Hyoscyamus muticus*, δ-cadinene synthase from cotton and germacrene C synthase from tomato (showing an identity/similarity between 34/57% and 40/63%) (*115*).

Interestingly, peppermint EBFS exhibits a comparable degree of similarity to both angiosperm and gymnosperm terpenoid synthases. For example, peppermint EBFS shows an identity/similarity of 30/51% with limonene synthase from spearmint and 29/50% with sabinene synthase from culinary sage. It also resembles terpenoid synthases from the gymnosperm *Abies grandis* with identity/similarity of 28/49% for monoterpene synthases, 29/53% for sesquiterpene synthases and 28/51% for diterpene synthases (*115*). These amino acid sequence-based relationships indicate a bifurcation in the evolution of the monoterpene synthases from the higher terpenoid synthases that is as ancient as the separation between the angiosperms and gymnosperms.

1.4.5   Diterpene synthases

Diterpene synthases catalyze the cyclization of geranylgeranyl diphosphate (C20) to form a variety of cyclic and polycyclic products. Plant diterpene synthases are triple-domain enzymes, which makes them the largest terpenoid synthases (800-900 residues) (*130*). The three domains have been suggested to be formed by fusion of single-domain and double-domain bacterial diterpene synthases (*131*).

The first crystal structure of diterpene synthase was solved for taxadiene synthase from *Taxus brevifolia* in 2011 (*130*). Taxadiene synthase contains three α-helical domains including both class I and class II terpenoid synthase folds (Figure 1.18). The conserved motifs (DDXXD and (N,D)XX(S,T)XXXE) were identified in the C-terminal domain, suggesting that it functions as a class I terpenoid synthase. The N-terminal domain (M107-I135 and S349-Q552) together with the insertion domain (S136-Y348) adopt the class II terpenoid synthase fold with the unique double α-barrel fold tertiary structures, but the conserved DXDD motif among class II terpenoid synthases was absent in taxadiene synthase. The unique insertion domain was found in most plant diterpene synthases with conserved amino acid sequence and position (*132*). However, the absence of the characteristic DXDD motif and an active site cavity makes it a non-functional domain. On the other hand, the C-terminal domain was determined to be the functional active domain and the analysis of N-terminal truncation variants showed an essential role for the N-terminal domain in catalysis too (*133*).

Active site

C-terminus
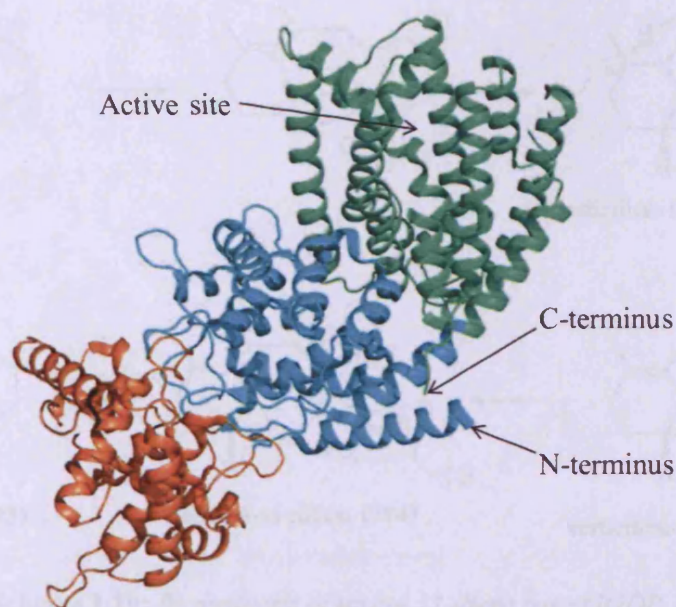
N-terminus

Figure 1.18: *Cartoon representation of taxadiene synthase (PDB 3P5R).* The C-terminal domain is coloured green, the N-terminal domain is coloured cyan and the insertion domain is coloured orange.

Investigations with different substrate analogues suggest a cyclization mechanism (*133-135*), in which the C14-C15 double bond, the C10-C11 double bond and the diphosphate leaving group are optimally aligned for departure of the diphosphate group before the formation of a verticillen-12-yl carbocation intermediate (**102**). Subsequent modification steps involve a proton transfer and a ring closure, followed by a deprotonation terminating step to yield taxa-4(5),11(12)-diene (**105**) (Scheme 1.19).

GGDP (6) → verticillen-12-yl cation (102)

verticillen-8-yl cation (103)

taxa-4,11-diene (105) ← taxen-4-yl cation (104) ← verticillen-8-yl cation (103)
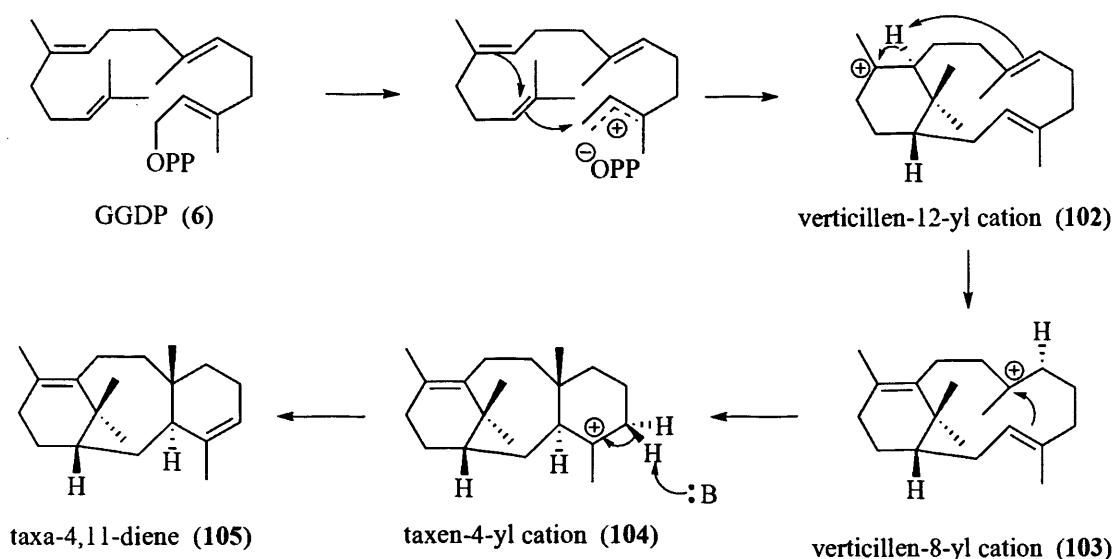
Scheme 1.19: *Biosynthesis of taxa-4,11-diene from GGDP.*

The crystal structure of taxadiene synthase complexed with the substrate analogue 2-fluoro-geranylgeranyl diphosphate reveals a similar magnesium cluster-diphosphate moiety binding fashion as other class I terpenoid synthases (Figure 1.19) (*130*). Asp613 and Asp617 (the first and the last aspartate residues from DDXXD motif) coordinate with Mg^{2+}_A and Mg^{2+}_C; Mg^{2+}_B ion is chelated by N757, T761 and E765 from the other magnesium binding motif on the opposite side of the active site. The diphosphate moiety can also form a hydrogen bond with R754 and makes water-mediated hydrogen bonds with Y688, E691, Y835, S713, R768 and Q770, which would possibly further stabilize the diphosphate group upon binding in the active site. In addition, the conformational change initiated by ligand binding that has been observed for the other class I terpenoid synthases is also proposed for the active site conformation of taxadiene synthase (*130*). The active site volume of taxadiene synthase is determined to be larger than the volume of the product taxadiene and this could explain the promiscuity of the enzyme since the shape of the active site contour should be more product-like for high fidelity terpenoid synthases.
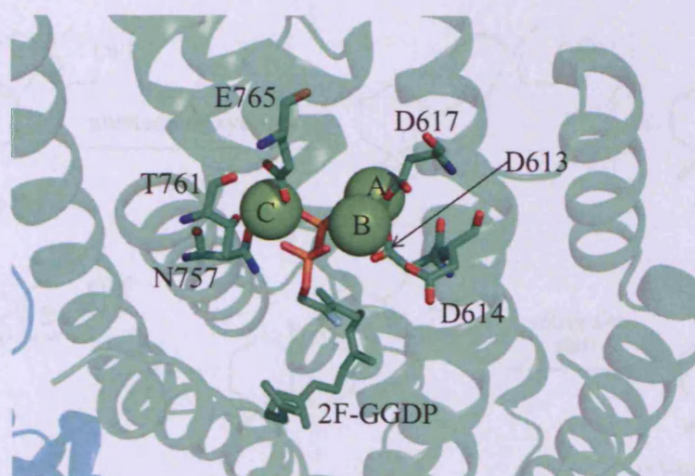
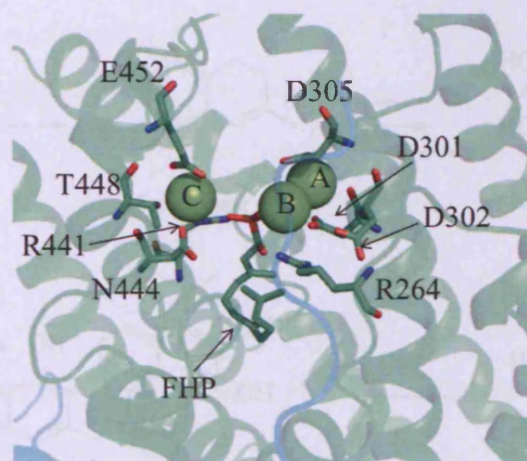**Figure 1.19:** *Cartoon representation of the active site of taxadiene synthase complexed with three magnesium ions (shown as green spheres) and substrate analogue 2-fluoro-geranylgeranyl diphosphate (2F-GGDP) (PDB 3P5R).* Residues from the two magnesium binding motifs are shown as sticks.

The structure of taxadiene synthase could generally illustrate the structure-function relationships in other diterpene synthases that contain catalytically active class II terpenoid synthase domains. For example, the bifunctional diterpene synthase abietadiene synthase from the grand fir tree is constructed from both class I and class II terpenoid synthase functional domains (*136*). The cyclization of GGDP is first carried out in the class II terpenoid synthase domain by protonation to form (+)-copalyl diphosphate as an intermediate and the class I terpenoid synthase can then catalyze the ionization dependent cyclization of (+)-copalyl diphosphate to generate abietadiene (*137*) (Scheme 1.20). In another diterpene synthase, copalyl diphosphate synthase, only the class II terpenoid synthase is functional; the class I terpenoid synthase domain has even lost the signature metal ion binding motifs (*138*).
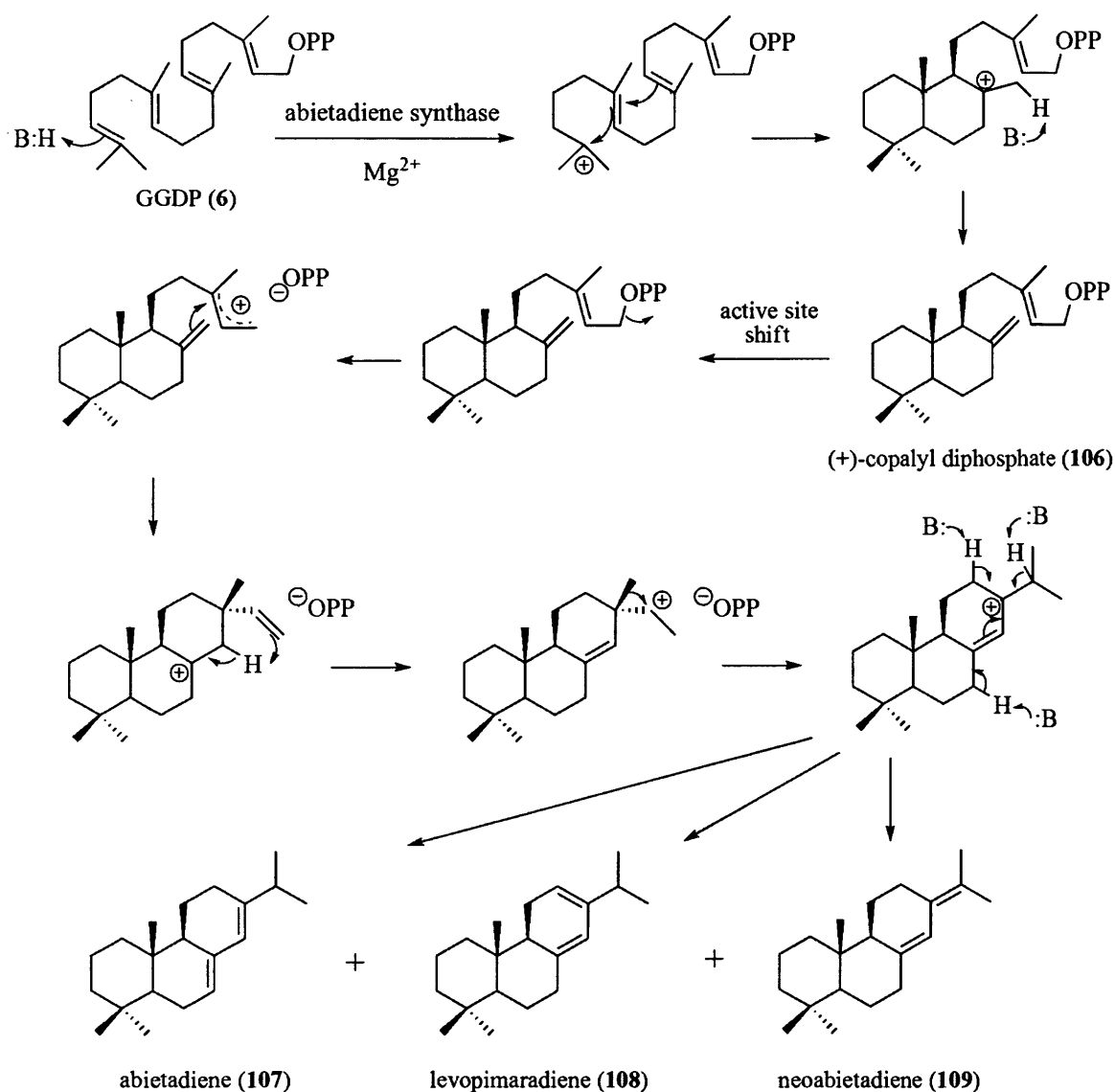
Scheme 1.20: *Biosynthesis of (+)-copalyl diphosphate and abietadiene from GGDP.*

1.4.6 Triterpene synthases

Triterpene synthases catalyze the cyclization of the linear C30 substrates squalene or 2,3-oxidosqualene to generate various fused-ring compounds (*139*, *140*). The triterpene synthases are defined as class II terpenoid synthases due to their unique double α-barrel fold tertiary structures, which are distinct from the characteristic α-helical fold of class I terpenoid synthases (*92*). The reaction cascade was initiated by protonation of a carbon-carbon double bond in squalene or by epoxide protonation and ring opening in squalene oxide (*57*). Although triterpene synthases catalyse

cyclisations by employing a longer substrate than monoterpene synthases and sesquiterpene synthases, they can also generate products with high structural and stereochemical precision. A key structure among triterpene synthases is the aspartate-rich motif (DXDD), which does not chelate metal ions as for class I terpenoid synthase, but functions as the proton donor to initiate the cyclization. For example, squalene-hopene synthase catalyses the cyclization of squalene (7) to form the pentacyclic hydrocarbon hopene (110) (Scheme 1.21).
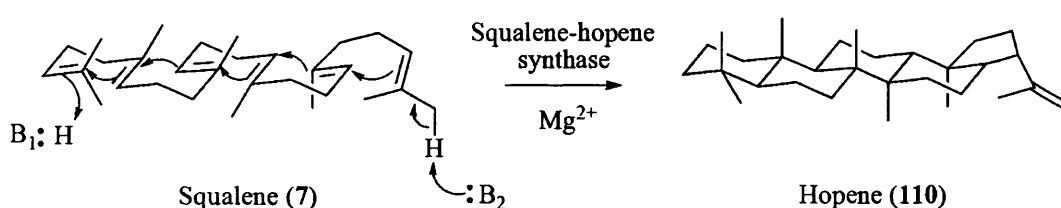
Scheme 1.21: *Biosynthesis of hopene from squalene.*

The crystal structure of squalene-hopene synthase from *Alicyclobacillus acidocaldarius* reveals a dimeric enzyme and each subunit contains two domains (Figure 1.20) (*141*). Domain 1 adopts a double α_6-barrel fold with helices within each barrel parallel to each other. Domain 2 contains a similar α-α barrel fold but each layer arranges in an irregular pattern. Some long loops from the inner barrels of both domains form a small β structure and enclose a central hydrophobic active site cavity (*141*). Eight QW-sequence motifs were discovered from the amino acid sequence of squalene-hopene synthase, in which seven of the eight motifs assume virtually identical polypeptide conformations. The side chains of the Q and W sequences can form hydrogen bonds with the amino end group of the adjacent outer barrel helix and with the carbonyl end moiety of the preceding outer barrel helix respectively, thus stabilizing the whole protein (*142*).
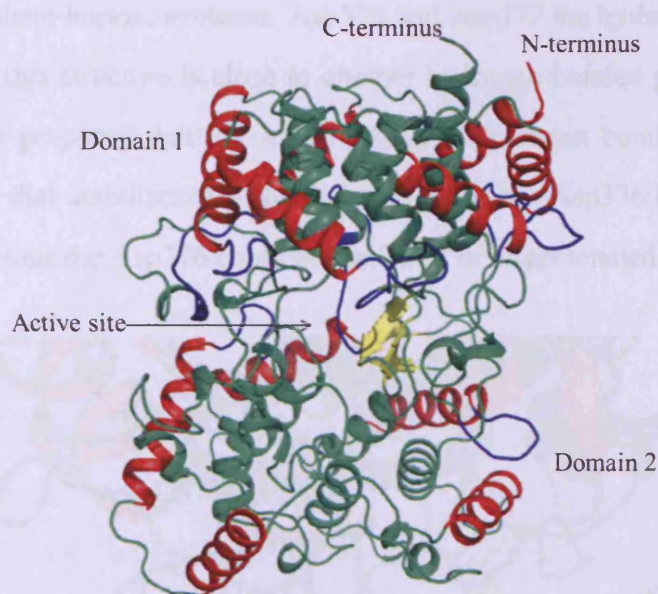
Figure 1.20: *Cartoon representation of squalene-hopene synthase (PDB 2SQC).* The external barrel helices are coloured red, the internal barrel helices are coloured green, the β-sheets are coloured yellow and the QW-motifs are coloured blue.

The active site of squalene-hopene synthase contains a variety of hydrophobic and aromatic residues, defining a hydrophobic active site template. The crystal structure of the enzyme complex with the nonreactive substrate analogue 3,6,9,12-tetraoxaicosan-1-ol showed a product-like binding, suggesting the active site template is able to enforce the pre-catalytically productive conformation of the large, flexible substrate squalene (Figure 1.21) (*143*). The aspartate motif (DXDD) has been identified at the polar top of the active site cavity. The catalytic function of this motif in squalene-hopene synthase has been established by site-directed mutagenesis studies, in which Asp376 (the second aspartate residue of the motif) seems to help protonate the C3 atom of substrate squalene. The mutation D376E reduces the catalytic activity by 10-fold and the D376G and D376R mutants are inactive (*144*).

Since the pK$_a$ of the aspartate side chain is around 4 and the enzymatic reaction was performed in the sodium citrate buffer system (pH 6), the Asp374, Asp376 and Asp377 were expected to be in their deprotonated form. However, in the crystal

structure of squalene-hopene synthase, Asp374 and Asp377 are hydrogen-bonded (2.6 Å distance) and this structure is close to another hydrogen-bonded pair (Asp376 and His 451). It was proposed that the Asp374:Asp377 hydrogen bonded pair carries a negative charge that stabilises the positive charge on the Asp376:His451 hydrogen bonded pair and thus the Asp376 could be stabilised in its protonated form (*141*).
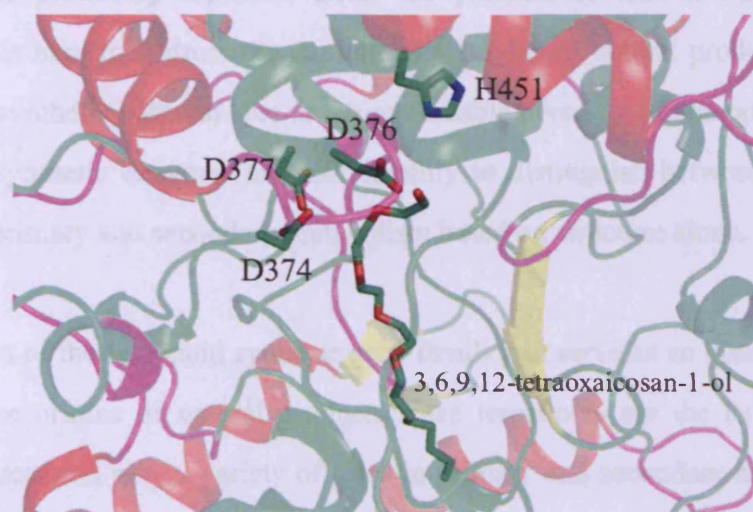
Figure 1.21: *Cartoon representation of the active site of squalene-hopene synthase complexed with substrate analogue 3,6,9,12-tetraoxaicosan-1-ol (shown as sticks) (PDB 2SQC).* Residues from the aspartate-rich motif are shown as sticks.

Notably, the amino acid residues in the upper region of the active site (*i.e.*, the polar Asp376 region) are more conserved than the ones in the lower region compared to other triterpene synthases. This suggests that the protonation step common to squalene and oxidosqualene synthases occurs in the conserved region of the cavity. The divergence of amino acid composition in the lower region of the active site cavity among triterpene cyclases will thus determine the divergent biosynthetic reactions (*141*).

1.4.7 Genomic organization of plant terpenoid synthase

It has been suggested that the precursors and pathways for the generation of natural products most likely arose from mutations of enzymes involved in the synthesis of primary metabolites (*145*). The natural products evolved to increase the survival fitness of the producing organism under the pressure of natural selection (*146*). However, it is hard to address the evolutionary origins of natural products and their complex biosynthetic pathway due to the remarkable diversity of compounds, lack of relevant biosynthetic enzymes and the inability to distinguish between compounds involved in primary and secondary metabolism based on structure alone.

The evolution of the terpenoid synthase gene family can serve as an instructive model to present the origins of natural products since terpenoids are the largest class of natural products and play a variety of roles in primary and secondary metabolism. A study based on 21 terpenoid synthases gene proposed that the ancestral terpenoid synthase contains 12-14 introns and 13-15 exons (Figure 1.22) (*145*). It was also pointed out that the plant terpene synthases involved in secondary metabolism have a common ancestral origin from a terpenoid synthase involved in primary metabolism and terpenoid synthases went through gene duplication and divergence several times to create enzymes of both primary and secondary metabolism. The structural variation among the terpenoid synthase genes is consistent with an experimentally demonstrated process of concerted intron loss, thus providing a possible explanation for the derivation of terpenoid synthases of natural products biosynthesis from those of primary metabolism (*147*).
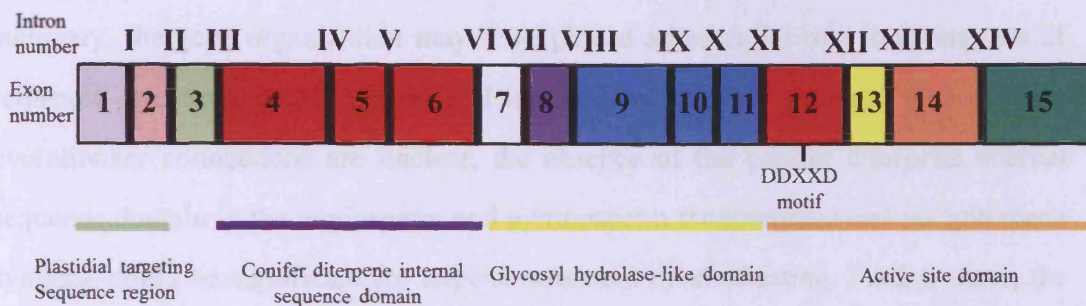
Figure 1.22: *General structure of plant terpene synthase genes (145)*. Introns 1-14 (Roman numerals in figure) are represented by black vertical bars and exons 1-15 are depicted by coloured boxes.

Repeated duplication of the ancestral gene and divergence by functional and structural specialization have been proposed as the common evolutionary process for the derivation of large gene families (*148, 149*). In addition, it seems most reasonable to assume that the precursors and pathways for the generation of secondary metablic products arose from mutations of enzymes involved in the synthesis of primary metabolites (*148*). Under the pressure of natural selection, secondary metabolic products evolved as compounds with important ecological functions, which increased the chance of survival of the organism. This assumption provides an explanation for the mechanism of the origin and evolutionary relationship of plant terpenoid synthase genes and their encoded enzymes.

In addition, the analysis suggests that the genes encoding terpenoid synthases from both gymnosperms and angiosperms derive from a common ancestor (*145*). The subsequent ancestral gene products diverged in function under the pressure of natural selection to generate the large family of terpenoid synthases involved in both primary and secondary metabolic pathways. Although speculative, it is plausible that the early terpenoid synthase ancestors were designed to be less specific enzymes, able to utilize different sized prenyl diphosphate substrates for the production of multiple products. The subfamilies of terpenoid synthases such as monoterpene synthases, sesquiterpene synthases and diterpene synthases were developed much later during evolution. In

59

summary, the gene organization may have played an essential role in divergence of terpenoid structures and the ecological interactions that they mediate. Although the evolutionary connections are unclear, the absence of the conifer diterpene internal sequence domain in the angiosperm and gymnosperm monoterpene and sesquiterpene synthase could be significant for terpene structural diversification. Further more, the genetic changes are closely related with an important structural aspect of the terpene synthases. Thus, considerable genetic variation occurs in the N-terminal region (with no known function) of the terpene synthase genes, whereas the C-terminal region (formation of the active site) remains highly conserved in organization and catalytic function.

Phylogenetic analysis of the deduced amino acid sequences of 33 terpenoid synthases from both angiosperms and gymnosperms allows subdivision of the terpenoid synthase (Tps) gene family into six subfamilies, designated Tpsa through Tpsf, each distinguished by sharing a minimum of 40% identity among members (*132*). The majority of terpene synthases analysed produce secondary metabolites, including members of subfamilies Tpsa (sesquiterpene and diterpene synthases from angiosperms), Tpsb (monoterpene synthases from angiosperms of the Lamiaceae), Tpsd (11 gymnosperm monoterpene, sesquiterepene and diterpene synthases) and the distant and possibly ancient Tpsf branch containing linalool synthase. The other three subfamilies, Tpsc, Tpse and Tpsf, are represented by the single angiosperm terpene synthase types copalyl diphosphate synthase, kaurene synthase and linalool synthase, respectively. The first two subfamilies are diterpene synthases involved in early steps of gibberellin biosynthesis. These two are grouped into a single clade and are involved in primary metabolism.

1.5 Aims

Comparison of the deduced amino acid sequences of the isolated plant sesquiterpene synthases reveals significant conservation among them, which can be explained by the

conserved exon organization observed between their genes. The derivation of this large gene family was proposed to be related to functional mutations that had been retained during gene duplication by natural selection. This proposal provides a reasonable explanation for the mechanism of origin and evolutionary relationship of plant sesquiterpene synthase genes and their encoded enzymes.

This project consisted of three parts. The first part of the study focused on DCS, which was aimed to ascribe functional roles for particular amino acid positions and subdomains encoded by conserved exons. A homology model of DCS based on the crystal structure of epi-aristolochene synthase (PDB 5EAT) was created using the SWISS-MODEL automated homology modelling server. The mutagenesis studies mainly relied on this model because the crystal structure of DCS was solved two years after this work had been started (*114*). Mutational replacement of variable residues within and surrounding the active site between DCS and germacrene C synthase (GCS) was intended to provide an insight into the relationship between the chemical environment of the active site and the product outcome. The role of the DDXXD motif was also investigated by site-directed mutagenesis studies. In addition, the critical structural elements in defining product outcome were to be determined by domain swap experiments.

The second part of the project was aimed at examining the influence of the active site volume of EBFS on the product outcome. A homology model of EBFS based on the crystal structure of DCS (PDB 3G4F) was created using the SWISS-MODEL automated homology modelling server and the mutagenesis studies mainly relied on this model. Mutants with enlarged or refined active site were designed to try to facilitate cyclization reactions in the modified active site pocket. The active site cavity was further modified by swapping residues from the active site of EBFS with corresponding ones from DCS to demonstrate the importance of active site volume.

The final part of the project was aimed to investigate the function of N-terminal

domains of EBFS and DCS, which were previously defined as non-catalytic domains. Truncated mutants, chimeras and site-directed mutants were designed to determine the role of N-terminal domain in catalysis.

Chapter 2: Materials and Methods

2.1 Materials

All chemicals were purchased from Sigma-Aldrich UK, Fisher Scientific Ltd or Melford, unless otherwise stated. FDP (*150*), 2F-FDP (*112*), 2F-GDP (*151*) and (*E*)-β-2-fluoro-farnesene (*112, 135*) were synthesised by Dr. Juan A. Faraldos and Dr. David J. Miller (Cardiff University). [1-^{3}H]-FDP (20 Ci/mmol) was purchased from American Radiolabeled Chemicals, Inc. DEAE anion exchange resins were purchased from GE Healthcase. Amicon YM30 membranes were purchased from Millipore. EcoScint scintillation fluid was purchased from National Diagnostics. All restriction enzymes were purchased from New England Biolabs or Promega. All chemicals were used in accordance with their material safety data sheets, COSHH and risk assessments. All oligonucleotides were purchased from Operon Technologies (Cologne, Germany). The wild type gene encoding δ-cadinene synthase was obtained from Dr. Xiao-Ya Chen (*100*) and subcloned into pET21d by Dr. Susan E. Taylor (University of Birmingham). The wild type cDNA of (*E*)-β-farnesene synthase was obtained from Dr Linda M. Field (Rothamsted Research, UK) and subcloned into pET32b by Dr. Athina Deligeorgopoulou (University of Birmingham).

2.2 Media

2.2.1 Luria-Bertani medium

LB medium was prepared by dissolving 10 g tryptone, 10 g NaCl, 5 g bacto-yeast extract in 1L deionised water and adjusted the *p*H to 7. The solution was sterilised in an autoclave at 121 °C, 15 lb.(sq. in)$^{-1}$ for 20 min.

2.2.2 LB agar

LB agar plates were made up by mixing 7.5 g of agar with 500 mL of LB media. The solution was autoclaved and cooled to under 50 °C and 100 mg of ampicillin was added. The mixture was then divided into Petri dishes (around 20 mL per dish) and left to solidify. The plates were stored at 4 °C.

2.3 Sterile solutions

2.3.1 Antibiotics

Ampicillin was dissolved in ethanol (50% v/v) to a concentration of 100 mg/mL. The solution was sterilised using a 0.2 μm syringe filter, aliquoted and stored at -20 °C. A working concentration of 0.1 mg/mL was used.

2.3.2 Isopropyl-β-D-1-thiogalactopyranoside (IPTG)

IPTG (360 mg) was dissolved in 1 mL of deionised water for use at a working concentration of 0.12 mg/mL. The solution was sterilised using a 0.2 μm syringe filter, aliquoted and stored at -20 °C.

2.3.3 Ethylenediaminetetraacetic acid (EDTA)

EDTA (1.46 g) was added to 450 mL deionised water. The solid was dissolved by adjusting the pH to 8.0 with 5 M sodium hydroxide and the total volume taken to 500 mL with deionised water. The solution was sterilised using a 0.2 μm syringe filter and stored at room temperature. Stock and working concentrations were 10 mM and 0.1 mM respectively for enzyme storage.

2.3.4 Dithiothreitol (DTT)

DTT (154 mg) was dissolved in 100 mL deionised water to give a stock concentration of 10 mM. The solution was sterilised using a 0.2 μm syringe filter and stored at 4 °C. The working concentration was 1 mM for enzyme storage.

2.3.5 Competent cell solutions

2.3.5.1 Rubidium chloride preparation of competent cells

Rubidium chloride solution 1 (Rb1) was prepared by dissolving potassium acetate (294 mg, 30 mM), rubidium chloride (1.21 g, 100 mM), calcium chloride (110.98 mg, 10 mM), manganese chloride (692 mg, 55 mM) and glycerol (15% v/v) in 80 mL of deionised water. The pH was adjusted to 5.8 with dilute acetic acid and the total volume adjusted to 100 mL with deionised water. The solution was sterilised using a 0.2 μm syringe filter and stored at 4 °C.

Rubidium chloride solution 2 (Rb2) was prepared by dissolving 3-(N-morpholino) propanesulfonic acid (MOPS) (209 mg, 10 mM), rubidium chloride (121 mg, 1 mM), calcium chloride (832 mg, 7.5 mM) and glycerol (15% v/v) in 80 mL of deionised water. The pH was adjusted to 6.5 with dilute sodium hydroxide and the total volume adjusted to 100 mL with deionised water. The solution was sterilised using a 0.2 μm syringe filter and stored at 4 °C.

2.4 Non-sterile solutions

2.4.1 dNTPs

Stock solutions of each dNTP were purchased at a concentration of 100 mM. These were diluted to a working concentration of 10 mM using deionised water, aliquoted and stored at -20 °C.

2.4.2 Ethidium bromide

8.1 g of ethidium bromide was dissolved in 100 mL of deionised water to make a 25 μM stock solution and stored in the dark at 4 °C. Immediately prior to gel staining, a working concentration of 6 μM was prepared by diluting 48 μL of stock solution into 200 mL of deionised water.

2.4.3 10x DNA loading dye

Bromophenol blue (2.5 mg) and sucrose (400 mg) were dissolved in 1 mL of deionised water and store at room temperature. This gave stock concentrations of 3.7 mM and 1.2 mM respectively. The dye was diluted 1:9 with each DNA sample immediately prior to use.

2.4.4 TAE buffer stock (50×) for agarose gels

0.5 M EDTA solution was prepared by dissolving 14.61 g EDTA in 80 mL of deionised water. The pH was adjusted to 8.0 with 5 M NaOH and the total volume was adjusted to 100 mL.

Tris base (242 g, 2 M), glacial acetic acid (57.1 mL, 7.5 M) and 0.5 M EDTA solution (100 mL) were mixed with 800 mL of deionised water. The total volume was adjusted to 1 L with deionised water after the entire Tris base was completely dissolved. The solution was stored at room temperature and was diluted 50 fold with deionised water prior to use.

2.4.5 SDS stacking buffer

Tris base (6 g, 0.5 M) was dissolved in 80 mL of deionised water and the pH adjusted to 6.8 with 6 M HCl. The total volume was adjusted to 100 mL with deionised water and stored at 4 °C.

2.4.6 SDS resolving buffer

Tris base (27.23 g, 1.5 M) was dissolved in 100 mL of deionised water and the pH adjusted to 8.8 with 6 M HCl. The total volume was adjusted to 150 mL with deionised water and stored at 4 °C.

2.4.7 10% (w/v) Sodium dodecyl sulfate (SDS)

Sodium dodecyl sulfate (10 g) was dissolved in 90 mL of deionised water. After the solid was fully dissolved, the total volume was adjusted to 100 mL and the solution was stored at room temperature.

2.4.8 10% (w/v) Ammonium persulfate

Ammonium persulfate (100 mg) was dissolved in 1 mL of deionised water. The solution was stored at 4 °C.

2.4.9 SDS electrode running buffer (10×)

Tris base (30.3 g, 250 mM), glycine (150 g, 2 M) and sodium dodecyl sulfate (10.0 g) were fully dissolved in 900 mL of deionised water. The total volume was adjusted to 1 L and stored at 4 °C. The concentrated solution was diluted 10 fold prior to use with deionised water.

2.4.10 SDS gel stain solution

Coomassie brilliant blue (0.25 mL) and glacial acetic acid (10 mL) were fully dissolved with 40 mL deionised water and 45 mL ethanol. The total volume was adjusted to 100 mL with deionised water and the solution was stored at room temperature.

2.4.11 SDS gel destain solution

Glacial acetic acid (100 ml) and isopropanol (120 ml) were mixed with 750 mL deionised water. The total volume was adjusted to 1 L and stored at room temperature.

2.4.12 Protein purification lysis buffers

The lysis buffer for purification of wild-type DCS was prepared by adding Tris base (2.42 g, 20 mM), EDTA (1.75 g, 6 mM) and β-mercaptoethanol (350 μL) to 900 mL deionised water. The pH was adjusted to 8 with 5 M NaOH and the total volume adjusted to 1 L with deionised water.

The lysis buffer for purification of wild-type EBFS was prepared by adding MOPS (4.2 g, 20 mmol), EDTA (1.75 g, 6 mM) and β-mercaptoethanol (350 μL) to 900 mL deionised water. The pH was adjusted to 7.2 with 5 M NaOH and the total volume adjusted to 1 L with deionised water.

Both lysis buffers were degassed via vacuum pumping (Vacuubrand GmbH +CO KG, MD4C, Wertheim, Germany) and stored at room temperature.

2.4.13 Protein dialysis buffer

For wild-type DCS, Tris-base (4.8 g, 10 mM) and β-mercaptoethanol (1.4 mL) were added to 3.8 L deionised water. The pH was adjusted to 7.5 with 5 M NaOH and the total volume adjusted to 4 L with deionised water.

For wild-type EBFS, MOPS (8.4 g, 10 mM) and β-mercaptoethanol (1.4 mL) were added to 3.8 L deionised water. The pH was adjusted to 7.2 with 5 M NaOH and the total volume adjusted to 4 L with deionised water.

Both dialysis buffers were prepared fresh each time and dialysis was performed at 4 °C.

2.4.14 DNA miniprep buffers

2.4.14.1 Buffer P1 (suspension buffer)

Tris-HCl (157.6 mg 50 mM), EDTA (58.448 mg, 10 mM) and RNase A (50 µg/mL final concentration) were dissolved in 15 mL of deionised water. The pH was adjusted to 8.0 and the total volume adjusted to 20 mL with deionsed water. The solution was stored at 4 °C.

2.4.14.2 Buffer P2 (lysis buffer)

NaOH (4 g, 0.4 M) and SDS (5 g, 2% w/v) were dissolved separately in 200 mL of deionised water. Both solutions were combined and the total volume was adjusted to 500 mL. The solution was stored at room temperature.

2.4.14.3 Buffer N3 (neutralization and binding buffer)

Guanidine hydrochloride (7.6 g, 4 M) and potassium acetate (981.4 mg, 0.5 M) were dissolved in 15 mL of deionised water. The pH was adjusted to 4.2 and the total volume was adjusted to 20 mL with deionised water and the solution was stored at room temperature.

2.4.14.4 Buffer PB (wash buffer)

Guanidine hydrochloride (9.5 g, 5M), Tris-HCl (63.0 mg, 20 mM) and ethanol (7.6 mL, 38% v/v) were added to 8 mL of deionised water. The pH was adjusted to 6.6 and the total volume was adjusted to 20 mL with deionised water. The solution was stored at room temperature.

2.4.14.5 Buffer PE (wash buffer)

NaCl (23.4 mg, 20 mM), Tris-HCl (6.3 mg, 2 mM) and ethanol (16 mL, 80% v/v) were added to 2 mL of deionised water. The pH was adjusted to 7.5 and the total

volume was adjusted to 20 mL with deionised water. The solution was stored at room temperature.

2.4.14.6 Buffer EB (Elution buffer)

Tris-HCl (32 mg, 10 mM) was dissolved in 15 mL of deionised water. The pH was adjusted to 8.5 and the total volume was adjusted to 20 mL with deionised water. The solution was stored at room temperature.

2.4.15 Agarose gel DNA isolation buffers

2.4.15.1 Buffer QG (gel solubilisation buffer)

Guanidine thiocyanate (13 g, 5.5 M) and Tris-HCl (6.3 mg, 20 mM) were dissolved in 15 mL of deionised water. The pH was adjusted to 6.6 and the total volume was adjusted to 20 mL with deionised water. The solution was stored at room temperature.

2.4.15.2 Buffer PE (wash buffer)

NaCl (23.4 mg, 20 mM), Tris-HCl (6.3 mg, 2 mM) and ethanol (16 mL, 80% v/v) were added to 2 mL of deionised water. The pH was adjusted to 7.5 and the total volume was adjusted to 20 mL with deionised water. The solution was stored at room temperature.

2.4.15.3 Buffer EB (Elution buffer)

Tris-HCl (32 mg, 10 mM) was dissolved in 15 mL of deionised water. The pH was adjusted to 8.5 and the total volume was adjusted to 20 mL with deionised water. The solution was stored at room temperature.

2.5 *E. coli* strains and their preparation

2.5.1 Cloning strain

XL1-Blue competent cells (Stratagene, CA, USA) with genotype (recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac [F′ proAB lacI^{q}ZΔM15 Tn10 (Tet^r)]) were used for routine cloning purposes, such as following site-directed mutagenesis. High quality miniprep DNA could also be produced from this strain for sequencing purposes.

2.5.2 Expression strains

BL21(DE3) competent cells (Stratagene, CA, USA) with genotype (E. coli B F$^-$ dcm ompT hsdS(r$_B^-$ m$_B^-$) gal λ(DE3)) were used for expression of wild-type DCS, whilst BL21-CodonPlus(DE3)-RP competent cells (Stratagene, CA, USA) with genotype (E. coli B F$^-$ ompT hsdS(r$_B^-$ m$_B^-$) dcm+ Tet^r gal λ(DE3) endA Hte [argU proL Cam^r]) were used for expression of wild-type EBFS. BL21-CodonPlus(DE3)-RIL competent cells (Stratagene, CA, USA) with genotype (E. coli B F$^-$ ompT hsdS(r$_B^-$ m$_B^-$) dcm+ Tet^r E. coli gal λ (DE3) endA Hte [argU ileY leuW Cam^r]) were used for expression of N-terminal region chimera D24AAE. All strains contain the λDE3 lysogen that carries the gene encoding T7 RNA polymerase under the control of the *lac*UV5 promoter.

2.5.3 Preparation of competent cells

Competent cells were streaked on a non-selective plate by a wire loop. The same competent cells were also streaked on an ampicillin plate as a negative control. Both plates were incubated at 37 °C overnight. 5 mL of LB medium without antibiotics was inoculated with a single colony from the non-selective plate and incubated at 37 °C until an optical density at 600 nm of 0.6 was reached. The cultures were then cooled down on ice for 15 min. Cells were harvested in a Thermo IEC 243 Centra CL3R centrifuge (Thermo Fisher Scientific Inc, USA) at 4000 RPM and 4 °C for 10 min and

the supernatant discarded. The cell pellet was resuspended in 5 mL Rb1 solution (Section 2.3.2), incubated on ice for 20 min and the centrifugation step repeated. The resulting cell pellet was resuspended with 6 mL Rb2 solution, aliquoted (50 µL) into sterile Eppendorf tubes and flash frozen in liquid nitrogen for storage at -80 °C.

2.5.4 Transformation protocol and controls

The frozen competent cells were allowed to thaw on ice for about 5 min. 2 µL of ice-cold DNA solution was mixed with the competent cells using a pipette tip under sterile conditions. The cell/DNA mixture was incubated on ice for 20 min, heat shocked at 42 °C in a water bath for 40 s and returned to ice for a minimum of 2 min. 500 µL of LB medium without antibiotics was added to the cell/DNA mixture and incubated at 37 °C for 1 hr. Cells were harvested by centrifugation for 1 min at 15,700 RPM (Eppendorf centrifuge 5415) and the supernatant discarded. The cell pellet was resuspended in 100 µL of LB medium, plated on agar plates containing the appropriate antibiotic and incubated at 37 °C overnight. A negative control was always included by replacing the DNA solution with sterile water. A positive control was also included by transforming cells with wild-type DNA.

2.6 DNA manipulation

2.6.1 Polymerase chain reaction (PCR)

The gene amplification and site-directed mutagenesis work was carried out by employing the Polymerase Chain Reaction (PCR) using short oligonucleotide primers. For all cloning work described here, *Pfu* polymerase from the bacterium *Pyrococcus furiosus* was used at 72 °C. This enzyme has a 3' to 5' proofreading activity and produces blunt-ended PCR products. The low error rate of this polymerase in PCR is roughly 10^6 per base pair duplicated (*152*).

The double stranded template DNA (dsDNA) was first denatured at 95 °C to form

two single stranded DNA (ssDNA) molecules. The temperature was then lowered to between 45-65 °C to allow primers to anneal to the template ssDNA. DNA polymerases synthesised new DNA strands complementary to the template DNA by adding dNTPs that were complementary to the template in a 5' to 3' fashion. The primers would be extended along the whole length of the DNA template and a 'daughter' DNA duplex was formed and the cycle could start again. A final elongation step was applied after the last cycle to finish synthesis of any partially completed fragments.

The primers were designed according to the instruction manual of the QuikChangeTM Site-Directed Mutagenesis Kit. Both primers (forward and reverse) must contain the desired mutation and anneal to the same sequence on complementary strands of the DNA. The primers should contain between 25 and 45 bases and the desired mutation should be in the middle of the primer. The percentage of GC should be at least 40%.

To amplify genes of interest and to make site mutations, a general PCR recipe was used (Tables 2.1 and 2.2).

Table 2.1: *Components of the PCR reaction.*

Component	Volume (µL)
dNTPs (40 mM total, 10 mM each NTP)	1
Forwards primer (100 µM)	4
Backwards primer (100 µM)	4
Template dsDNA (~0.01 pmol)	1
10× *Pfu* buffer	5
MgSO$_4$ (50 mM stock, 1.0 mM working concentration)	1
Pfu polymerase (2-3U/µL)	1
Deionised water	33

Table 2.2: *Temperature and duration of the PCR cycles.*

PCR cycles	Temperature (°C)	Time (min)
Initial denaturation	95	2
Denaturation	95	1
Annealing	~55	0.5
Primer extension	72	2 min per kilobase
Final extension	72	5

*Dpn*I restriction enzyme (target sequence: 5'-GATC-3' with methylated alanine) was used to digest template DNA following site directed mutagenesis, ensuring only mutated DNA was left intact. The digestion was performed immediately after site directed mutagenesis at 37 °C for 1 hr.

2.6.2 Restriction digestion of DNA and controls

For each restriction digest, the manufacturer's guidelines were followed with regard to operating buffers, temperatures and denaturation steps. *Eco*RI and *Sac*I were used sequentially for excision of the aristolochene synthase gene from pET21 vector when required. Both restriction endonucleases were incubated at the same temperature (37 °C) for 6 hours and then were heated inactivated at 65 °C for 20 minutes. *Nhe*I and *Bam*HI were used sequentially for excision of the C-terminal EBFS gene from pET32 vector when required. Both restriction endonucleases were incubated at the same temperature (37 °C) for 6 hours and then were heated inactivated at 65 °C for 20 minutes.

2.6.3 DNA visualisation, isolation and purification

Agarose gels were prepared by w/v ratio (usually 1%) with 1×TAE buffer (Section 2.4.3). Agarose gel electrophoresis was run at a constant electric current of 100 mA for 60 min and stained in diluted ethidium bromide solution for 10 min. DNA bands

were visualised using a Syngene GeneFlash UV light box (Syngene, UK). The size of sample bands was estimated by comparison to a 1kb DNA ladder. For agarose gel purification, the appropriate bands were excised from the gel using a clean scalpel blade and stored in a sterile eppendorf tube. DNA extraction was performed using the QIAquick gel extraction kit (QIAGEN, UK) according to the manufacturer's instructions and stored at -20 °C.

For the determination of the concentration of DNA, a UV/VIS spectrophotometer was used. Nucleic acid, has a strong ultraviolet absorption at 280 nm, however absorbs much more strongly at 260 nm. Most proteins exhibit a characteristic ultraviolet light absorption at 280 nm due to the presence of tyrosine and tryptophan. An A260nm value of 1.0 indicates an approximate concentration of 50 μg ml-1 double stranded DNA, 40μg ml-1 single stranded DNA or 20μg ml-1 single stranded oligonucleotides. The ratio OD260nm / OD280nm is indicative of the purity of the DNA sample. Thus, when this ratio takes values between 1.8 and 2.0, the sample is pure, whilst lower values indicate contamination with proteins and/or lipids.

2.6.4 Constructions of DNA hybrids and truncated mutants

2.6.4.1 DCS-hybrid1

The construct of DCS-hybrid1 was prepared by replacing the native GGT codon of G276 with TTC by site-directed mutagenesis. Together with the codon GAA that encodes amino acid E275, this mutation creates an *Eco*RI restriction site. Another *Sac*I restriction site located after the stop codon was also employed to form the hybrid construct. The gene encoding aristolochene synthase was amplified from the genomic DNA *via* PCR using short complementary oligonucleotide primers designed to flank the gene of interest and provide overhanging *Eco*RI and *Sac*I restriction sites at the 5' and 3' end respectively. The restriction sites were introduced immediately prior to the start codon and immediately after the stop codon. The gene fragment that encodes the C-terminal domain of DCS was removed from the mutated plasmid by a double

digestion with *Eco*RI and *Sac*I restriction enzymes. The opened plasmid was then ligated with the newly formed DNA fragment encoding aristolochene synthase, creating a plasmid containing the gene encoding DCS-hybrid1.

2.6.4.2 DCS-hybrid2

The gene of DCS-hybrid2 was constructed using two native *Sac*I restriction sites identified in the plasmid containing the gene encoding DCS; one *Sac*I restriction site sequence encodes amino acids E318 and L319 and the other *Sac*I restriction site is located after the stop codon. A synthetic DNA fragment was designed to encode the amino acid sequence from E311 to the stop codon of GCS, with *Sac*I restriction sites at both the 5' and 3' end. The plasmid containing the gene encoding DCS was digested with *Sac*I restriction enzyme, dephosphorylated with Calf Intestinal Alkaline Phosphatase (CIP) and the longer DNA fragment was purified by agarose gel electrophoresis. The longer opened plasmid was then ligated with the synthetic DNA fragment encoding the 3'-terminal portion of the GCS gene.

2.6.4.3 EBFS active site hybrid

The construct of the hybrid was created by mutating the native codon TAT that encodes amino acid Y271 to AGC; together with the codon GCT that encodes amino acid A270, this mutation creates a *Nhe*I restriction site. A synthetic DNA fragment was designed to encode the amino acid sequence after V272 of EBFS with all the chosen mutations and constructed with *Nhe*I and *Bam*HI restriction sites at the 5' and 3' end respectively. Double digestion of the mutated plasmid with *Nhe*I and *Bam*HI restriction enzymes and subsequent ligation with the synthetic DNA fragment formed a plasmid containing the gene encoding the hybrid.

2.6.4.4 EBFSΔ14

The full-length cDNA for EBFS (1.6 kb) was initially cloned into the vector pET32b (5.9 kb) with an *Nco*I restriction site at the 5' end. To construct the gene of EBFSΔ14,

another *Nco*I restriction site was introduced into the DNA sequence of the wild-type enzyme. The foreign *Nco*I restriction site was formed by replacing the DNA fragment 5'-GTAAGGCCA-3' that encodes Val14, Arg15 and Pro16, with 5'-GCCATGGCA-3' using site-directed mutagenesis. The mutated plasmid was then treated with *Nco*I and the resulting large DNA fragment was intramolecularly ligated to give the desired plasmid.

2.6.4.5 EBFSΔ24

This mutant was prepared following a similar method to that used for constructing EBFSΔ14, by introducing an *Nco*I restriction site into the DNA sequence of the wild-type EBFS. The mutated plasmid was then treated with *Nco*I and the resulting large DNA fragment was intramolecularly ligated to give the desired plasmid.

2.6.4.6 D24AAE

The cDNA of D24AAE was created by ligation of the DNA encoding EBFSΔ24 following *Nco*I digestion with a DNA fragment encoding the chosen amino acids from DCS. The smaller DNA fragment was designed to anneal with *Nco*I restriction sites at both ends.

2.6.4.7 E24AAD

The cDNA of E24AAD was prepared by ligation of the DNA encoding DCSΔ30 (constructed by Dr. Veronica Gonzalez in our group) following *Nco*I digestion with a DNA fragment encoding the first 24 amino acids of EBFS. The insert DNA fragment was designed to anneal with the *Nco*I restriction sites at both ends.

2.6.5 Plasmid purification

5 mL of ampicillin selective LB medium was inoculated by a single colony that was picked from an ampicillin selective plate containing the appropriate transformed cells. The mixture was incubated at 37 °C overnight, whilst shaking at 150 rpm in an Innova® 43 shaker (New Brunswick Scientific, UK). Cells were harvested by centrifugation for 10 min at 5000 RPM (Eppendorf centrifuge 5415R). Isolation of pure plasmid was performed using the QIAprep spin miniprep kit (QIAGEN, UK), following the manufacturer's instructions.

2.6.6 DNA sequencing

Usually, the standard plasmid purification procedure described in Section 2.6.3 will give pure DNA at a concentration of ~50 μg/μL. DNA sequencing of plasmid constructs and site-directed mutagenesis products was performed by Lark Technologies (Cogenics, UK) and in the School of Biosciences, Cardiff University.

2.7 Protein production and purification

2.7.1 Small scale test expression

Prior to large scale production of any protein (Section 2.7.2), a small scale test expression was performed to ensure cells could be brought to an optical density of 0.6 AU at 600 nm and the expression could be induced by IPTG. 100 mL of ampicillin selective LB medium was inoculated with 1 mL of overnight culture and allowed to grow at 37 °C to an optical density of 0.6 AU at 600 nm. IPTG (0.5 mM) was used to induce gene expression, which was allowed to grow at optimised temperature. Protein production was monitored by taking a 200 μL sample every hour for SDS-PAGE analysis (Section 2.7.8).

2.7.2 Large scale expression

For protein production, the selected strain of competent cells (Section 2.5.2) were transformed with the appropriate DNA solution and incubated overnight on ampicillin selective agar plates at 37 °C. A single colony was picked and used to inoculate 100 mL of ampicillin selective LB medium by incubating at 37 °C overnight whilst shaking at 150 rpm in an Innova® 43 shaker (New Brunswick Scientific, UK). The overnight pre-culture was used to inoculate ampicillin selective LB medium (5 mL per 500 mL final culture) and incubated at 37 °C whilst shaking as previously described until an optical density of 0.6 AU at 600 nm was reached. IPTG (0.5 mM) was then added into cells culture. For the expression of wild-type DCS, the cell culture was allowed to grow for another 4 hr at 37 °C. For the expression of wild-type EBFS, the cell culture was allowed to grow for another 7 hr at 16 °C.

Large scale protein production used 1.5 L of culture. Cells were harvested by centrifugation in a Sorvall RC5C Plus centrifuge (Thermo Fisher Scientific, USA) using an SLA-3000 rotor at 9000 RPM for 10 min and stored at -20 °C.

2.7.3 Glycerol stocks

0.8 ml of an overnight culture containing the plasmid of interest was aseptically transferred to a sterile 1.5 ml eppendorf. 0.15 ml of sterile glycerol was added and the resultant mixture was vortexed briefly and stored at − 80 ° C. Prior to use, the glycerol stock should be defrosted on ice and by the use of a sterile inoculating loop the sample was streaked on an agar plate containing proper antibiotics.

2.7.4 Base extraction

Frozen cell pellets (~10 g per 1.5 L culture) (See section 2.7.2) were allowed to thaw on ice. Cells were resuspended in ~100 ml protein purification lysis buffer (Section 2.4.9) and lysed by sonication on ice (3 sec sonication period followed by a 3 sec rest period, repeated for a total of 5 min). The lysate was centrifuged at 12,000 RPM

(Sorvall RC5C Plus centrifuge, Thermo Fisher Scientific Inc, USA, using an SS-34 rotor) for 30 min and the supernatant solution discarded. The pellets were resuspended in 150 mL of fresh lysis buffer and titrated on ice with 5 M NaOH until the solution became clear (about pH 12). After stirring on ice for 30 min, the pH was lowered to 7.2 with 1 M HCl and β-mercaptoethanol was added to a final concentration of 5 mM. The resulting solution was stirred for 30 min on ice and centrifuged at 12,000 RPM (Sorvall RC5C Plus centrifuge, Thermo Fisher Scientific, Inc, USA, using an SS-34 rotor) for 30 min at 4 °C, after which the protein was found in the supernatant solution.

2.7.5 Anion exchange chromatography

Ion-exchange chromatography, which separates compounds based on their net surface charge, is often used in protein purification. The stationary phase contains charged functional groups, of which there are four main types: strong acidic (*e.g.* sulfonic acid groups), strong basic (*e.g.* quaternary amino groups), weak acid (*e.g.* carboxylic acid groups) and weak basic (*e.g.* primary, secondary and/or tertiary amino groups).

The crude cell extract is prepared in a buffer of pH at least one unit above (for anion exchange) or below (for cation exchange) its isoelectric point (pI) and passed over the stationary phase resin. The desired protein and other molecules of appropriate charge are retained on the resin, while other molecules are eluted. After all the non-binding molecules have been washed through, the buffer conditions are changed in order to elute the target protein. The buffer conditions can be modified by adjusting the pH or the ionic strength. For this work, diethylaminoethyl (DEAE), a weakly basic anion-exchange resin that remains charged in a range of pH 3-9, was used.

The protein solution was loaded onto a diethylaminoethyl (DEAE) anion exchange column (column volume (CV) ~75 ml). The column was washed for at least 2 CV

until baseline absorbance was reached and the flow-through was collected. A two-stage NaCl gradient was applied to the column, from 0-0.5 M NaCl over 4 CV, followed by 0.5-1 M NaCl over 1 CV. The column was then washed by another 1 CV lysis buffer with 1 M NaCl solution. The absorbance was monitored at 280 nm. To remove the NaCl used to elute the protein, the solution was dialyzed using MediCell membranes Size 3500/2 (cut-off of 3500) against buffer that contained 10 mM Tris and 5 mM β-mercaptoethanol at pH 7.5, without NaCl.

2.7.6 Concentration of the dialysed protein using Amicon ultrafiltration

The dialysate was transferred to a 50 ml Amicon^{TM} ultracentrifugation apparatus containing a 44.5 mm Millipore 30 kDa cutoff ultrafiltration membrane and concentrated at a pressure of 1.5 bar at 4 °C.

2.7.7 Size exclusion chromatography

The apparent molecular weight of the wild-type EBFS was determined by size exclusion chromatography. Typical commercially available resins for sizing column include crosslinked compounds such as dextran and epichlorohydrin (*e.g.* sepharose), dextran covalently bound to agarose (*e.g.* superdex), highly crosslinked agarose (*e.g.* superose) and *bis*-acrylamide/detran copolymers (*e.g.* sephacryl). The packed resin in the sizing column would retard different proteins in the mixture to different degrees whilst passing through the column based on their molecular size. Larger molecules will be retarded much less than the smaller ones due to their inability to enter the small pores of the particles in the gel matrix. Smaller molecules will be trapped in the pores while diffusion happens between the stationary and mobile phase, which leads to a longer retention time. Therefore, the separation is dependent on the length and density of the column, the type of resin used, the sample volume and the flow rate applied.

1 mL of concentrated protein solution was applied to a Superdex™ 200 size exclusion column (CV ~24 mL), which had been previously equilibrated with 2 CV of protein purification lysis buffer with 150 mM NaCl. The column was then washed with another 2 CV of the equilibration buffer. The absorbance was monitored at 280 nm and the fractions were analysed by SDS-PAGE.

2.7.8 Determination of protein concentration

Sodium phosphate buffer (5 mM, pH 7) in the presence of 50 mM sodium sulfate was filtered and degassed using a vacuum pump (MD4C, Vacuubrand GMBH + CO KG, Germany) prior to use in protein concentration determination. Purified protein was dialysed against dialysis buffer (4L) overnight before the measurement. The spectrometer was blanked with 1 mL of buffer. Then, 10 µL of the dialysed protein sample was mixed with 1 mL of phosphate buffer and a blank sample was prepared by mixing 1 - 10 µL of the dialysis buffer with 1 mL of phosphate buffer. Both samples were prepared immediately before use by gently pipetting several times. The increase in absorption at 205 nm was noted once the readings were stable. This process was then repeated at 280 nm using a 10 times more concentrated sample. To ensure the accuracy of measurements, the abosorbance was kept down to no more than 0.5. The extinction coefficient at 205 nm was calculated by the formula:

$$\varepsilon_{205}^{1\ mg\ mL^{-1}} = 27 + 120 \times (A_{280}/A_{205})$$

where: ε = extinction coefficient, A = absorbance at given wavelength

The Beer-Lambert law was used to calculate the concentration (mg/mL) of the protein solution.

$$A = \varepsilon \times c \times l \text{ and therefore } c = A \times df / \varepsilon \times l$$

where: ε = extinction coefficient, A = absorbance at given wavelength, c = sample concentration in mg/mL and l = pathlength of cuvette in cm and df is the dilution factor. The calculated protein concentration can be used to recalculate the ε_{280} and the value of this ε_{280} can be used to determine the concentration of the same protein sample next time.

2.7.9 SDS-PAGE protocol

Monomer solutions were prepared by mixing reagents as described in Table 2.3. The gels were cast between two glass plates. 100 μL of 10% APS and 10 μL of N,N,N',N'-tetramethylenediamine (TEMED) were added to both resolving and stacking gels immediately prior to pouring the gel. Solutions were mixed gently by pipetting to initiate polymerization. The resolving gel solution was covered with a layer of isopropanol when it was poured and allowed to polymerize. Isopropanol was removed with blotting paper once the resolving gel was set. The stacking gel solution was poured and a comb was inserted to create wells for sample loading. Typically, a 9% stacking and 13% resolving gel were used in my experiments.

Table 2.3: *SDS-PAGE monomer solutions.*

Gel (%)	dH₂O (mL)	30% Degassed Acrylamide-Bis (mL)	Gel Buffer (mL)	10% (w/v) SDS (mL)
6	5.4	2.0	2.5	0.1
7	5.1	2.3	2.5	0.1
8	4.7	2.7	2.5	0.1
9	4.4	3.0	2.5	0.1
10	4.1	3.3	2.5	0.1
11	3.7	3.7	2.5	0.1
12	3.4	4.0	2.5	0.1
13	3.1	4.3	2.5	0.1
14	2.7	4.7	2.5	0.1
15	2.4	5.0	2.5	0.1
16	2.1	5.3	2.5	0.1

2.8 Enzyme characterization

2.8.1 Enzyme kinetics

Enzyme kinetics is the study of the chemical reactions that are catalysed by enzymes. In enzyme kinetics, the reaction rate is measured and the effects of varying the conditions of the reaction are investigated as well. Studying an enzymes's kinetics can reveal the catalytic mechanism of this enzyme.

In biochemistry, Michaelis-Menten kinetics is one of the most frequently used model of enzyme kinetics. Most of the biochemical reactions are bimolecular reactions involving two reactants, substrate 'S' and enzyme 'E'. In all enzymatic reactions, the enzyme is recycled and not consumed in the reactions. Before proceeding the *in vitro* experiments, some conditions must be met to ensure the resulting data is valid. As long as initial velocity is considered, the concentration of product can be neglected (compared to that of the substrate) and the concentration of substrate is in large excess over that of the enzyme. Here is a simple model of enzyme reaction:

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES \xrightarrow{k_{cat}} E + P$$

In this model, the substrate S reversibly associates with the enzyme E in a first step and some of the resulting complex ES is allowed to break down and yield the product P and the free enzyme back. Michaelis-Menten made three assumptions to derive and explain the kinetics of an enzyme reaction.

Assumption 1: reaction is in equilibrium.

In the above reaction once the product is formed, it does not go back to substrate, the reaction is in equilibrium.

Assumption 2: Steady-state assumption.

In a steady state, the concentration of the intermediates [ES] is constant even if the concentration [S] and [P] are changing. This steady state occurs when the rate of formation and breakdown of ES complex are equal.

Assumption 3:

Maximal rate is obtained when the catalytic sites on the enzyme are saturated with the substrate.

2.8.1.1 Buffers for kinetics

Kinetic parameters of DCS based proteins were determined in 20 mM Tris buffer containing 1 mM DTT and 5 mM $MgCl_2$ at pH 7.5. Kinetic parameters of EBFS based proteins were determined in 20 mM MOPS buffer containing 1 mM DTT and 5 mM $MgCl_2$ at pH 7.2. Buffers were made freshly for each experiment.

2.8.1.2 Activity assays of column fractions

The aim of this assay was to quantify the product formed by the enzyme from different fractions of the chromatography column, when incubated with $[1-^3H]$ FDP. The substrate remained in the aqueous phase while the cyclic products went into the organic solvent (hexane) used for the extraction. Constant amounts of substrate were incubated at 22 °C with constant amounts of enzyme eluted at different times from the chromatography column in activity buffer. The final volume for the reaction was 250 μl. The reaction was quenched by the addition of 100ml EDTA (100 mM). Products were extracted with 3×700 μL hexane. The organic solution was then passed through a small silica column (typically containing 500 mg of silica) and mixed with 15 mL

EcoScint™ (National Diagnostics) scintillation fluid before analysis on a scintillation counter (Packard 2500 TR™) in ³H mode for 4 min per sample. The percentage conversion of [1-³H] FDP to radioactive hexane extractable products was determined by comparing the observed radioactivity with a known concentration of [1-³H]FDP and with control samples with no enzyme.

2.8.1.3 Steady state kinetics

Steady state kinetics were performed by radioactive assays using tritiated FDP ([1-³H] FDP) as substrate. Reactions (250 µL) were initiated by addition increasing concentrations of labeled FDP to a constant concentration of enzyme solution. The highest concentration of FDP should reach the saturation state for each enzyme. The reaction time and enzyme concentration were optimized to ensure the reaction was in the initial linear phase and was not reaching saturation during the experiment. The incubation temperature was 22 °C for all measurements. The reactions were quenched by addition of 100 µL of 100 mM EDTA. Products were extracted with 3×700 µL hexane. The organic solution was then passed through a small silica column (typically containing 500 mg of silica) and mixed with 15 mL EcoScint™ (National Diagnostics) scintillation fluid before analysis on a scintillation counter (Packard 2500 TR™) in ³H mode for 4 min per sample. The percentage conversion of tritiated FDP to radioactive hexane extractable products was determined by comparing the observed radioactivity level with a known concentration of tritiated FDP without enzyme in scintillation fluid. Data were fitted to the Michaelis-Menten equation,

$$v = (v_{max} [S])/(K_M + [S])$$

using Systat SigmaPlot 10. k_{cat} was calculated by the formula:

$$k_{cat} = v_{max}/(A \times t \times [E])$$

where: A is the specifici activity of substrate (counts per min corresponding to 1 μM FDP). t is the reaction time (s). [E] is the enzyme concentration (μM).

2.8.2 GC-MS

Product profiles were determined by incubating 10 μM of purified enzyme with 1 mM (E,E)-FDP in the present of 5 mM Mg^{2+}. Reactions were overlaid with 700 μL HPLC grade pentane and typically left over two nights. Sesquiterpene products were extracted using 3×700 μL pentane (HPLC grade) and pentane extracts were passed through a small silica column (containing approximately 500 mg silica). According to the previously reported methods (114), the resulting solutions were then analysed by GC-MS using a Hewlett-packard 6890 gas chromatography fitted with a J&W Scientific DB-5MS column and a Micromass GCT Premiere detecting in the range of m/z 50-800 in EI^+ mode.

'Relative abundance' of GC or MS readings was defined as the ion current. The presented data has been normalized to that the maximum value displayed equals 100.

2.8.3 Circular dichroism spectroscopy

All circular dichroism spectroscopy was performed on a Chirascan^{TM} spectrophotometer (Applied Photophysics Ltd, UK). Potassium phosphate buffer (10 mM, pH 7) was used for all the measurements. A scan of the buffer alone was taken before every actual run as a blank. In the experiment, the appropriate amount of enzyme was mixed with the potassium phosphate buffer to give a final enzyme concentration of 10 μM.

Converting the signal obtained from CD experiments into mean residue ellipticity (MRE) was performed by the formula:

$$\Theta_{MRE} = \Theta/10 \times n \times c \times l$$

where: Θ = CD signal in millidegree, n = number of backbone peptide bond, c = molar concentration of sample and l = path length of cuvette used in cm.

Typically, a 0.1 cm path length cuvette was used and the CD spectrum was recorded usually from 190-400 nm.

2.8.4 Errors and their propagation

2.8.4.1 Standard deviation and standard error of the mean

The errors in this work are described as the standard error of the mean (σ_{SEM}) that is defined as the standard deviation (σ) of the values in the sample divided by the square root of the sample size.

$$\sigma_{SEM} = \frac{\sigma}{\sqrt{n}}$$

$$\sigma = \sqrt{\frac{\sum(X-M)^2}{(n-1)^2}}$$

where: X = each value measured in the sample, M = mean of the sample, n = sample size.

2.8.4.2 Propagation of errors

The propagation of the errors for k_{cat}/K_M was calculated as follow,

$$Z = X/Y$$

$$\Delta Z = Z \sqrt{\left(\frac{\Delta X}{X}\right)^2 + \left(\frac{\Delta Y}{Y}\right)^2}$$

where X and Y represent independent values directly measured experimentally, ΔX and ΔY are their errors, Z is the calculated value and ΔZ is its propagated error.

2.9 Amino acid sequence(s) alignment

Program ClustalX was used for two or multiple amino acid sequence(s) alignment. Sequences (and profiles) are input using the FILE menu. Invalid options will be disabled. All sequences must be in 1 file, one after another. 7 formats are automatically recognised: NBRF/PIR, EMBL/SWISSPROT, Pearson (Fasta), Clustal (*.aln), GCG/MSF (Pileup), GCG9 RSF and GDE flat file. All non-alphabetic characters (spaces, digits, punctuation marks) are ignored except "-" which is used to indicate a GAP ("." in MSF/RSF).

Table 2.4: Default parameters for protein sequence(s) alignment.

Fast pairwise alignment parameters	
K-tuple (word) size	1
Window size	5
Scoring method	percentage
Number of top diagonals	5
Multiple alignment parameters	
Wight matrix	blosum
Gap opening penalty	10.0
Gap extension penalty	0.05
Hydrophilic gaps	on
Hydrophilic residues	GPSNDQERK
Residue-specific gap penalties	on

90

Chapter 3: δ-Cadinene Synthase

3.1 Characterisation of the wild-type δ-cadinene synthase (WT-DCS)

As described in Section 1.5, a considerable amount of work has previously been devoted to investigating both the mechanistic and the structural characteristic of DCS (*104-110, 114*). To provide a benchmark for our mutagenesis and cloning studies, some basic experiments were repeated to characterize DCS from *Gossypium arboreum* (tree cotton).

3.1.1 Purification of WT-DCS

A cDNA of WT-DCS (*Gossypium arboreum*) was previously subcloned into the expression vector pET21d by Dr. S. Taylor. *E. coli* BL21(DE3) cells transformed with the resulting plasmid were used for protein production. After cell lysis, the protein was isolated from the insoluble fraction and refolded using the base extraction protocol (Section 2.7.3) (Figure 3.1).
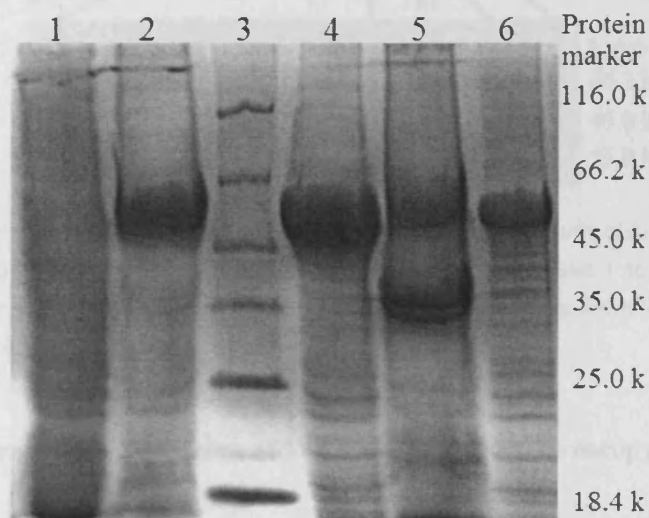
Figure 3.1: *13% SDS-PAGE analysis of DCS purification from BL21(DE3) cells.* Lane 1: supernatant solution after sonication and centrifugation. Lane 2: pellet after sonication and centrifugation. Lane 3: protein markers. Lane 4: supernatant solution after base extraction and centrifugation. Lane 5: pellet after base extraction and centrifugation. Lane 6: DCS fractions after concentration by Amicon.

The refolded protein mixture was loaded onto a DEAE column and a 0-1 M NaCl gradient was applied to elute the bound proteins. Two peaks could be identified from

the FPL chromatogram (Figure 3.2A). Based on the SDS-PAGE analysis of the purified protein, the first peak contained the majority of the protein with the correct molecular weight representing DCS (Figure 3.2B). The fractions that contained pure protein were pooled and dialyzed against 4 L dialysis buffer overnight. The dialysate was concentrated to approximately 5 ml using an Amicon^{TM} ultrafiltration apparatus (Figure 3.1).
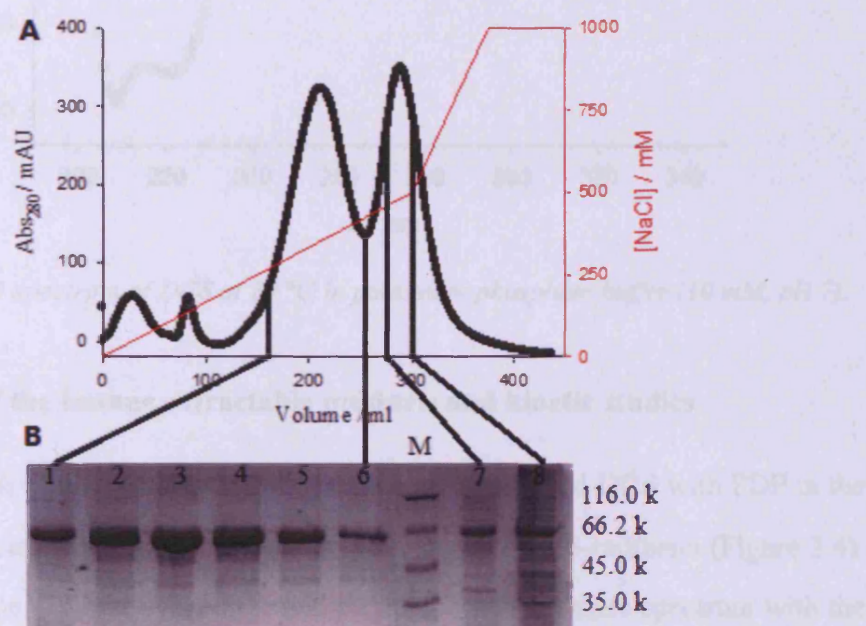
Figure 3.2: *(A) Chromatogram for DEAE purification of wild-type DCS. (B) 13% SDS-PAGE analysis of the purified protein.* Lane 1 to 6: eluent fractions from the first peak. Lane 7 and 8: eluent fractions from the second peak. Lane M: protein marker.

3.1.2 Structural characterisation of WT-DCS by CD spectroscopy

The studies of the crystal structure of WT-DCS revealed an overall α-helical fold (*114*) and this was confirmed by the circular dichroism spectrum of the protein with two minima at 208 and 222 nm (Figure 3.3). The α-helicity was calculated to be 44% at 20 °C using K2d software (*153*).
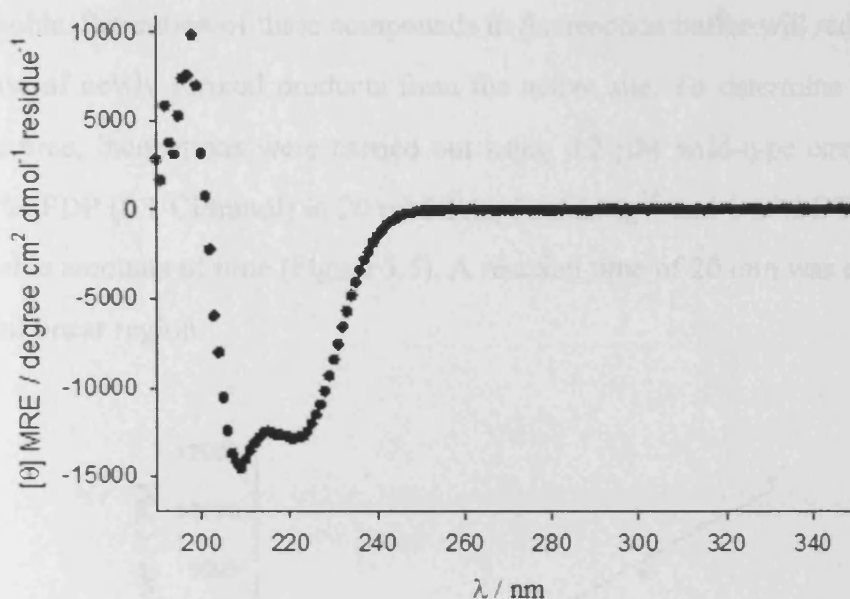
Figure 3.3: *CD spectrum of DCS at 20 °C in potassium phosphate buffer (10 mM, pH 7).*

3.1.3 Analysis of the hexane extractable products and kinetic studies

In agreement with a previous study (*59*), incubation of purified DCS with FDP in the presence of cofactor Mg^{2+} gave only the bicyclic compound δ-cadinene (Figure 3.4). The identity of the product was confirmed by comparing the mass spectrum with the Wiley database available on the mass spectrometer software (*154*).
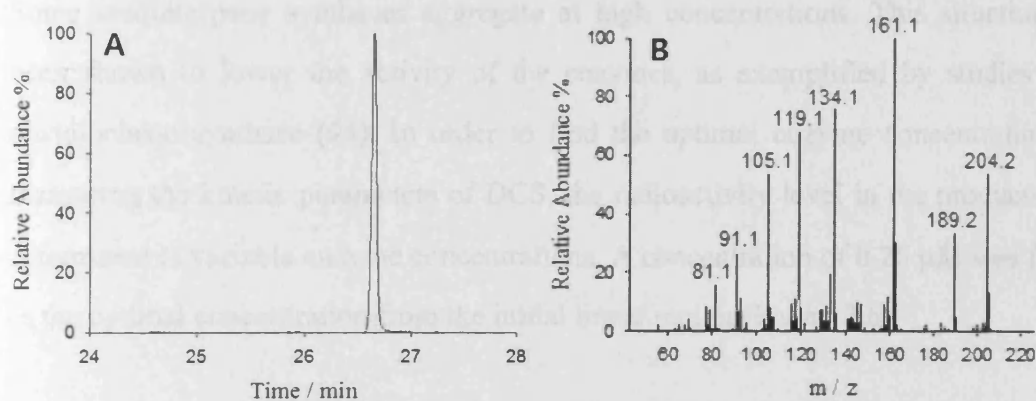
Figure 3.4: *(A) GC trace for hexane-extractable products following incubation of 10 µM wild-type DCS with 1 mM FDP in the presence of 5 mM Mg^{2+}. (B) EI-mass spectrum of δ-cadinene peak.*

Most of the compounds formed by sesquiterpene synthases from FDP are hydrophobic. Saturation of these compounds in the reaction buffer will reduce the rate of release of newly formed products from the active site. To determine the optimal reaction time, incubations were carried out using 0.2 μM wild-type enzyme with 5 μM [1-^{3}H]FDP (0.1 Ci/mmol) in 20 mM Tris, 5 mM Mg^{2+} and 1 mM DTT at pH 7.5 for variable amounts of time (Figure 3.5). A reaction time of 20 min was chosen from the initial linear region.
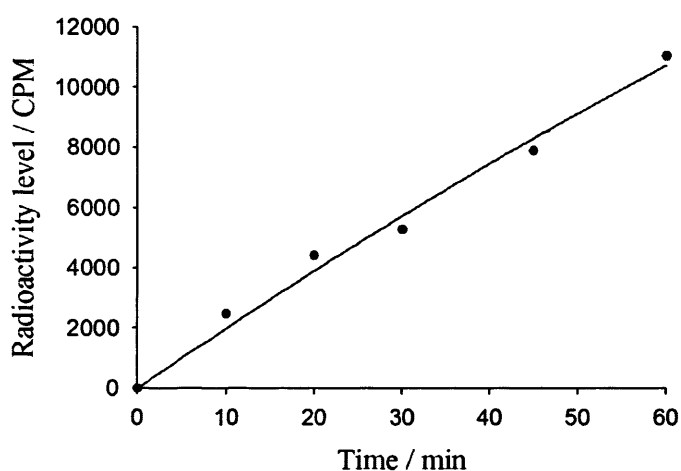
Figure 3.5: *Plot of radioactivity level in hexane-extracted products formed by 0.2 μM wild-type DCS versus reaction time.*

Some sesquiterpene synthases aggregate at high concentrations. This situation has been shown to lower the activity of the enzymes, as exemplified by studies with aristolochene synthase (*95*). In order to find the optimal enzyme concentration for measuring the kinetic parameters of DCS, the radioactivity level in the products was determined at variable enzyme concentrations. A concentration of 0.26 μM was found as the optimal concentration from the initial linear region (Figure 3.6).
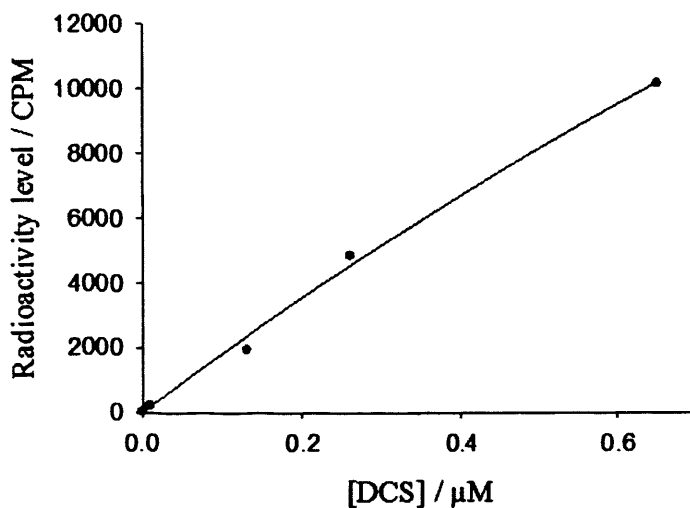
Figure 3.6: *Plot of radioactivity level in products formed in 20 min by wild-type DCS versus enzyme concentration.*

Kinetic assays were subsequently carried out by incubating 0.26 µM WT-DCS with variable concentrations of [1-^{3}H]FDP (from 1 to 40 µM) (Figure 3.7). K_M and v_{max} values for each individual run were calculated using the program SigmaPlot 10 by fitting to the Michaelis-Menten equation, $v = (v_{max} [S])/(K_M + [S])$.
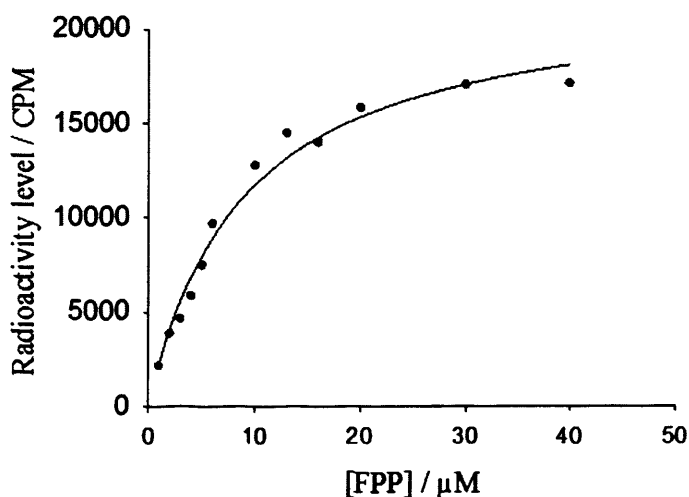
Figure 3.7: *Michaelis-Menten plot for of wild-type DCS, incubation of 0.26 µM enzyme with different concentrations of [1-^{3}H]FDP in the presence of 5 mM Mg^{2+}.*

Based on five individual kinetic runs, the K_M value was calculated to be 9.82 ± 0.33 μM and the k_{cat} value 0.0155 ± 0.0005 s^{-1} (based on the equation, $k_{cat} = V_{max}/(A \times T \times [E])$, (section 2.8.1.2)). Both k_{cat} and K_M value are comparable to the published data (7 μM for K_M and 0.039 s^{-1} for k_{cat}) (100).

3.2 Mutagenesis studies of DCS

3.2.1 Introduction

Promiscuous enzyme activities are believed to have divergently evolved to acquire higher specificity and activity (155). This process is usually carried out by a small number of amino acid substitutions (156). Some studies have reported the importance of contributions from residues within and surrounding the active site (84, 157-159). For example, Deligeorgopoulou et al. observed that a single amino acid substitution within the active site of aristolochene synthase was sufficient to change the product spcificity more towards the acyclic sesquiterpene (E)-β-farnesene (84). Active site saturation mutagenesis of γ-humulene synthase, a promiscuous sesquiterpene synthase generating more than 50 products, created mutants that generated alternative predominant products by substitution of only three to five active site residues at a time (155).

Recently, Greenhagen et al. accomplished a systematic interconversion of Nicotiana tabacum 5-epi-aristolochene synthase (TEAS) and Hyoscyamus muticus premnaspirodiene synthase (HPS) activities, which established a minimal set of residues responsible for the divergent biosynthetic properties in these two enzymes (160). Molecular comparisons of TEAS and HPS, two plant sesquiterpene synthases that share 72% amino acid identity and similar enzymatic mechanism (Scheme 3.1), have shown to be useful in identifying structural elements for determining product specificity (161).

FDP (5)

eudesmyl cation (72)

epi-aristolochene (70)

TEAS

HPS

premnaspirodiene (111)
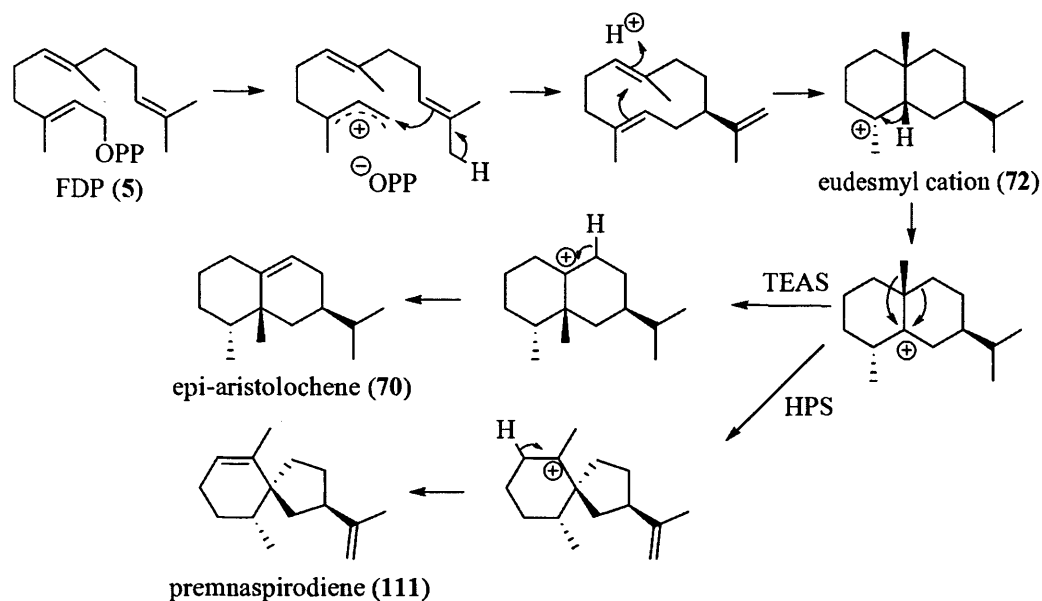
Scheme 3.1: *Proposed catalytic mechanisms for TEAS and HPS via the common eudesmyl cation as intermediate.*

To have a more systematic investigation of identifying functional residues, a contact mapping strategy was developed by using three-dimensional coordinates of TEAS complexed with an unreactive FDP analogue, farnesylhydroxy-phosphonate (FHP). This strategy created concentric tiers of amino acid residues that have contact with each carbon atom of FHP (Figure 3.8). The first tier of the active site of TEAS was defined as the residues within van der Waals radii (3.5 Å) of FHP and ten residues were identified in this tier. The second tier was constructed by the next 18-residues shell of contact. A three-dimensional model of HPS was created based on the TEAS template structure. By comparing the amino acid composition of both enzymes, the first tier residues of TEAS and HPS are identical, while only 4 different residues were spotted in the second tier. Experiments by mutating these residues of TEAS to the corresponding ones of HPS resulted in a growth of premnaspirodiene production, but not a full transmutation of activity. In this case, another five residues differing between TEAS and HPS within 12.5 Å of the active site center that may have potential to contribute to the reaction were targeted for further analysis. Sequential mutagenesis at these five positions on the previously obtained quadruple mutant

showed a gradual growth of premnaspirodiene production. The mutant with all nine amino acids mutated catalyses the formation of 75% premnaspirodiene and the catalytic efficiency of this mutant is comparable to the wild type HPS. Since all of the sesquiterpene synthases share the same class I terpenoid fold, this successful interconversion could serve as a model for sesquiterpene synthases to identify amino acids within and surrounding the active site that make direct and indirect contributions to catalysis.
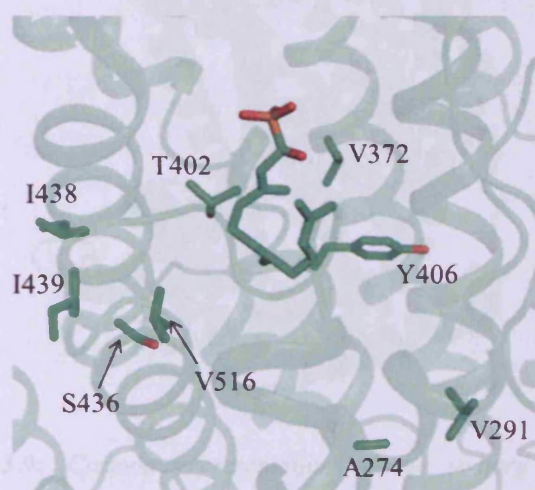
Figure 3.8: *Identity and spatial relationships of residues that distinguish TEAS and HPS activitys.* FHP and selected residues are shown as sticks.

The aim of our mutagenesis work was to convert DCS to germacrene C synthase (GCS) by analogy. The amino acid sequence of DCS (from *Gossypium arboretum*) shows 68% similarity and 50% identity with that of GCS (from *Lycopersicon esculentum*), which is the closest homologue of DCS (*59*). Thus, GCS would be the most suitable target to perform the transmutation of activity from DCS and hence to identify structural elements underlying the evolution of product selectivity.

A homology model of DCS based on the crystal structure of epi-aristolochene synthase (PDB 5EAT) was created using the SWISS-MODEL automated homology modelling server. The mutagenesis studies mainly relied on this model because the

crystal structure of DCS was solved two years after this work had been started (*114*). The tertiary structure alignment between the homology model and the crystal structure reveals very little difference in the overall conformation (Figure 3.9), demonstrating that the structural rationale behind all of the mutagenesis studies performed here was robust.
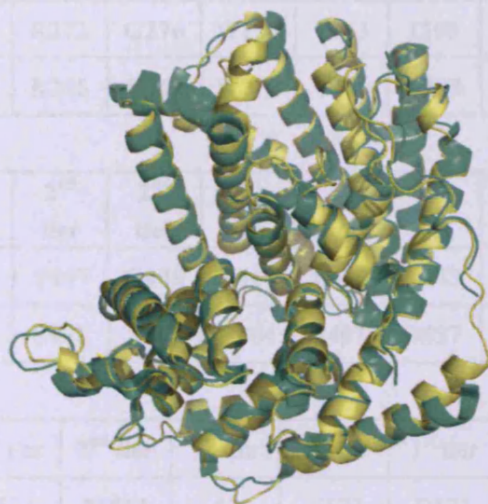
Figure 3.9: *Cartoon representation of the tertiary structure alignment between the homology model generated for this work (coloured yellow) and the x-ray crystal structure of DCS (coloured green) (PDB 3G4F (114)).*

The substrate FDP was docked into the active site by the program FlexX and the residues within and surrounding the active site of DCS were defined here as the ones that in the first tier (*i.e.* within van der Waals radii (about 3.5 Å) of the substrate carbon atoms) and second tier (*i.e.* within van der Waals radii of the first tier residues) of the active site sequence alignment to identify the corresponding residues was carried out using ClustalX with the default parameters (*162*). The proposed active site residues of DCS were then compared with corresponding ones of GCS and 7 different residues between DCS and GCS were identified, which potentially have the ability to affect the enzymatic reaction (Table 3.1). The contribution of these residues to the proposed active site was later confirmed by the crystal structure of DCS (*114*); the

tertiary structure alignment revealed that these residues were positioned in a similar fashion in both homology model and crystal structure (Figure 3.10).

Table 3.1: *Comparison of wild-type DCS active site contact residues with corresponding residues from GCS.* Residue differences between DCS and GCS are shown as bold characters.

	2nd tier	1st tier	2nd tier	**2nd tier**	1st tier	1st tier	**1st tier**	2nd tier	2nd tier	2nd tier
DCS	F27	R270	R272	**G276**	W279	V283	**I300**	S304	D308	L378
GCS	F15	R263	R265	**C269**	W272	V276	**L293**	S297	D301	L371

	2nd tier	1st tier	**2nd tier**	**2nd tier**	1st tier	**1st tier**	2nd tier	2nd tier	1st tier	**2nd tier**
DCS	Y382	E385	**T407**	**C408**	Y410	**L413**	S443	I446	R448	**E455**
GCS	Y275	E378	**S401**	**A402**	Y404	**I407**	S437	I440	R442	**G449**

	1st tier	2nd tier	**2nd tier**	1st tier	1st tier	1st tier	1st tier
DCS	L519	V522	**M523**	L526	Y527	D531	Y533
GCS	L513	V516	**A517**	L520	Y521	D525	Y527
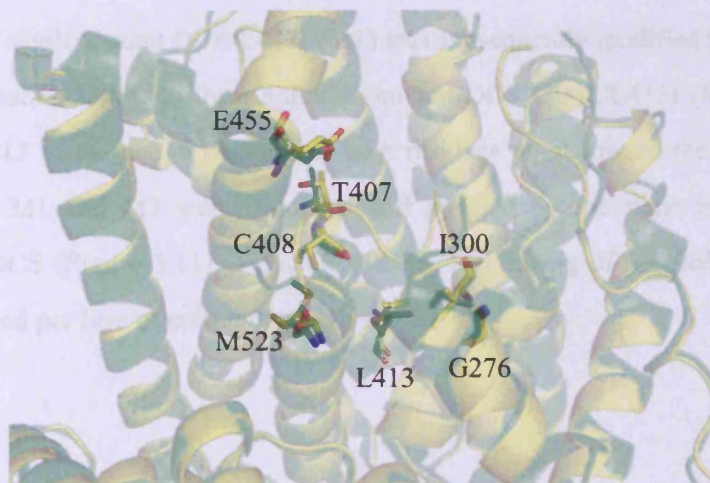
Figure 3.10: *Cartoon representation for the comparison of active site contact residues (shown as sticks) between the homology model (coloured yellow) and the x-ray crystal structure (coloured green) (PDB 3G4F (114)).*

Substitution of DCS active site residues to the corresponding residues in GCS was performed to identify the essential set of residues responsible for divergent biosynthetic properties in DCS.

3.2.2 Expression and purification of active site mutants

Active site mutations were constructed by site-directed mutagenesis. The native and the substituted codons for each mutation are listed in Table 3.2. The presence of the mutations was verified by DNA sequencing.

Table 3.2: *Comparison of native codons from wild-type DCS with substituted codons for the active site contact residues.*

	G276C	I300L	T407S	C408A	L413I	E455G	M523A
Native codon	GGT	ATA	ACT	TGT	CTT	GAA	ATG
Substituted codon	TGT	AAT	AGT	GCT	ATT	GGA	GCG

Stepwise mutagenesis of the active site of DCS started from the first tier residues. L413 of DCS, located deep in the middle of the active site, was mutated to Ile. The constructed single mutant DCS-L413I (M1) was subsequently modified by addition of another mutation I300L to form a double mutant DCS-I300L/L413I (M2). I300 sits close to L413 in the active pocket and both residues point toward the centre of the active site. M1 and M2 were produced and purified by the same method as the wild-type DCS (Figure 3.11). Typically, 12 mg and 35 mg of purified M1 and M2 were obtained per litre of culture respectively.
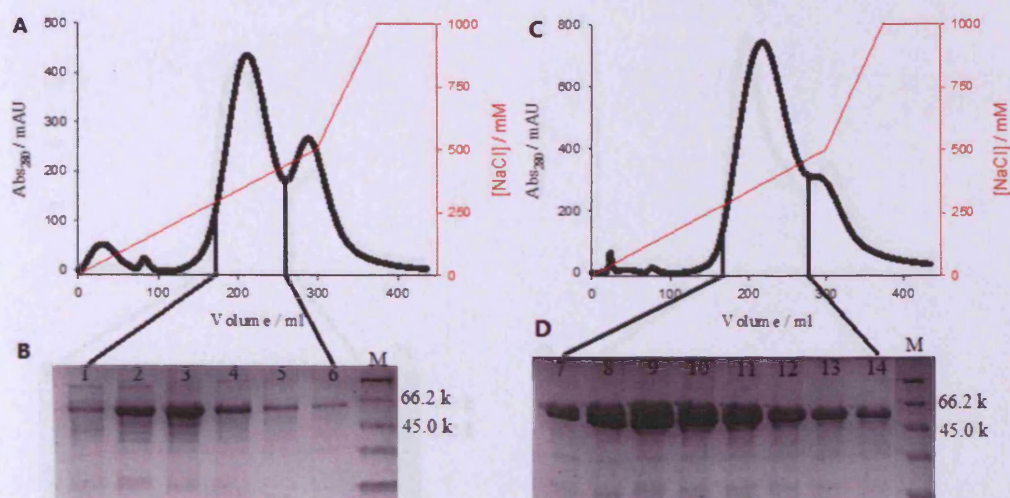
Figure 3.11: *Chromatograms for DEAE purification of M1 (A) and M2 (C) and 13% SDS-PAGE analyses of each purification (B) and (D).* Lane 1-6: eluent fractions for M1. Lane 7-14: eluent fractions for M2. Lane M: protein marker.

Mutational replacement of the second tier residues of DCS to the corresponding residues of GCS was carried out by sequential addition of the other five mutations onto the double mutant M2. First, E455 of DCS was mutated to the corresponding Gly residue of GCS to form a triple mutant DCS-I300L/L413I/E455G (M3). Next, the mutant DCS-I300L/T407S/C408A/L413I/E455G (M5) was constructed by combining another two mutations T407S and C408A with M3. As these two residues are adjacent to one another, this mutagenesis was performed in a single step and no mutant M4 was generated. E455, T407 and C408 are located near the entrance of the active site. In addition to the contribution of forming active site contour, these residues may also involve in stabilizing the magnesium ion and diphosphate moiety. The mutants M3 and M5 were produced and purified following the same method as wild-type DCS (Figure 3.12).
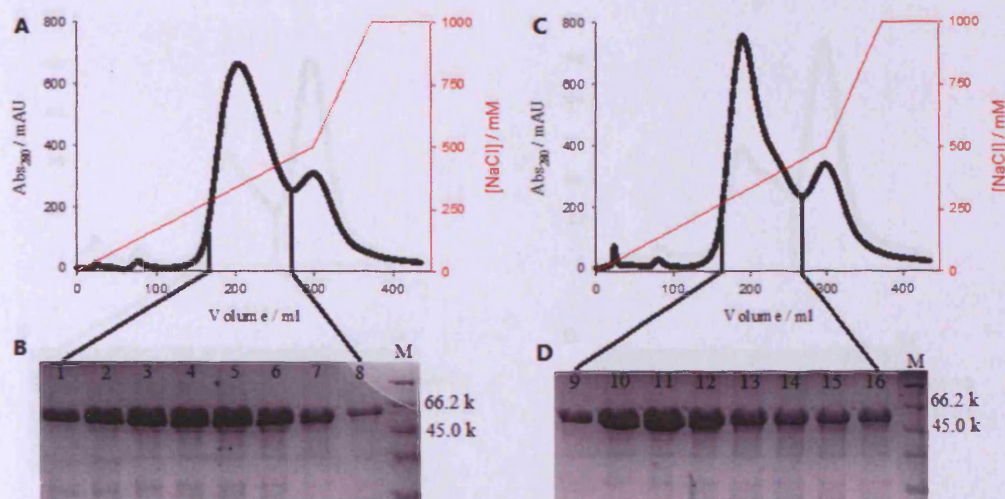
Figure 3.12: *Chromatograms for DEAE purification of M3 (A) and M5 (C) and 13% SDS-PAGE analyses of each purification (B) and (D).* Lane 1-8: eluent fractions for M3. Lane 9-16: eluent fractions for M5. Lane M: protein marker.

Typically, 25 mg and 30 mg of purified M3 and M5 were obtained from per litre of culture respectively. The mutant M5 was subsequently modified by addition with another mutation M523A to form a new variant DCS-I300L/T407S/C408A/L413I/E455G/M523A (M6). The last mutant DCS-G276C/I300L/T407S/C408A/L413I/E455G/M523A (M7) was prepared by introducing the mutation G276C into the mutant M6. The mutants M6 and M7 were produced and purified based on the same method as wild-type DCS (Figure 3.13). Typically, 15 mg and 10 mg of purified M6 and M7 were obtained from per litre of culture respectively.
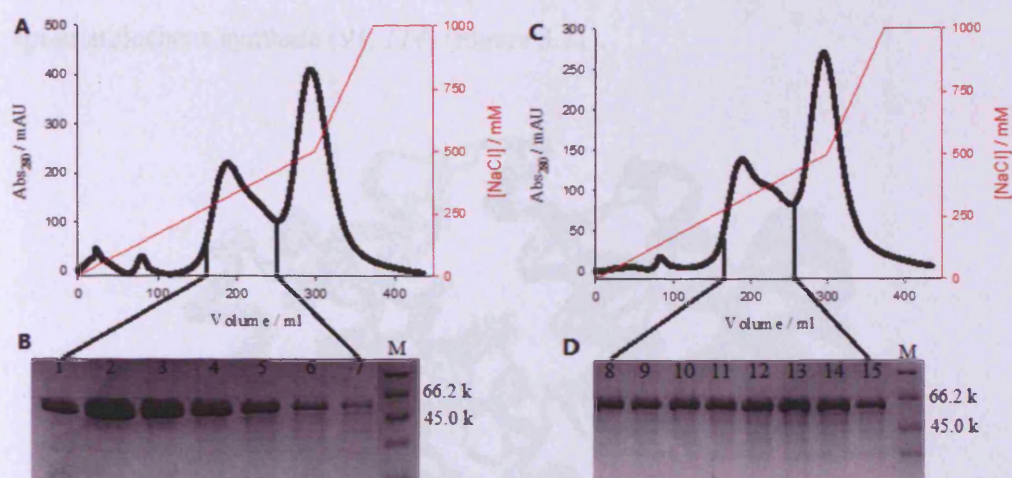
Figure 3.13: *Chromatograms for DEAE purification of M6 (A) and M7 (C) and 13% SDS-PAGE analyses of each purification (B) and (D).* Lane 1-7: eluent fractions for M6. Lane 8-15: eluent fractions for M7. Lane M: protein marker.

3.2.3 Analysis of the sesquiterpene products and kinetic studies

All the *in vitro* reactions were performed by incubating 10 μM enzyme and 1 mM FDP in 20 mM Tris buffer containing 5 mM Mg^{2+} and 1 mM DTT at pH 7.5. All the kinetic assays were carried out using radiolabelled [$1-^3$H]FDP and the same Tris buffer system as for *in vitro* reactions. The incubation time and the enzyme concentration for the kinetic assays were optimized before the actual runs (See section 2.8.1.2 for details).

The mutagenesis studies related to the active site of DCS revealed a totally different pattern of the function of active site residues from those previously reported mutagenesis studies of other sesquiterpene synthases (*e.g.* AS, TEAS and γ-humulene synthase). It has been shown that active site modification by substitutions of active site residues of terpene synthases had a significant influence on the product profile and/or catalytic efficiency (*84, 155, 157-159*). However, all of the mutants within and surrounding the active site of DCS gave δ-cadinene as the exclusive sesquiterpene product (Figure 3.14).

105
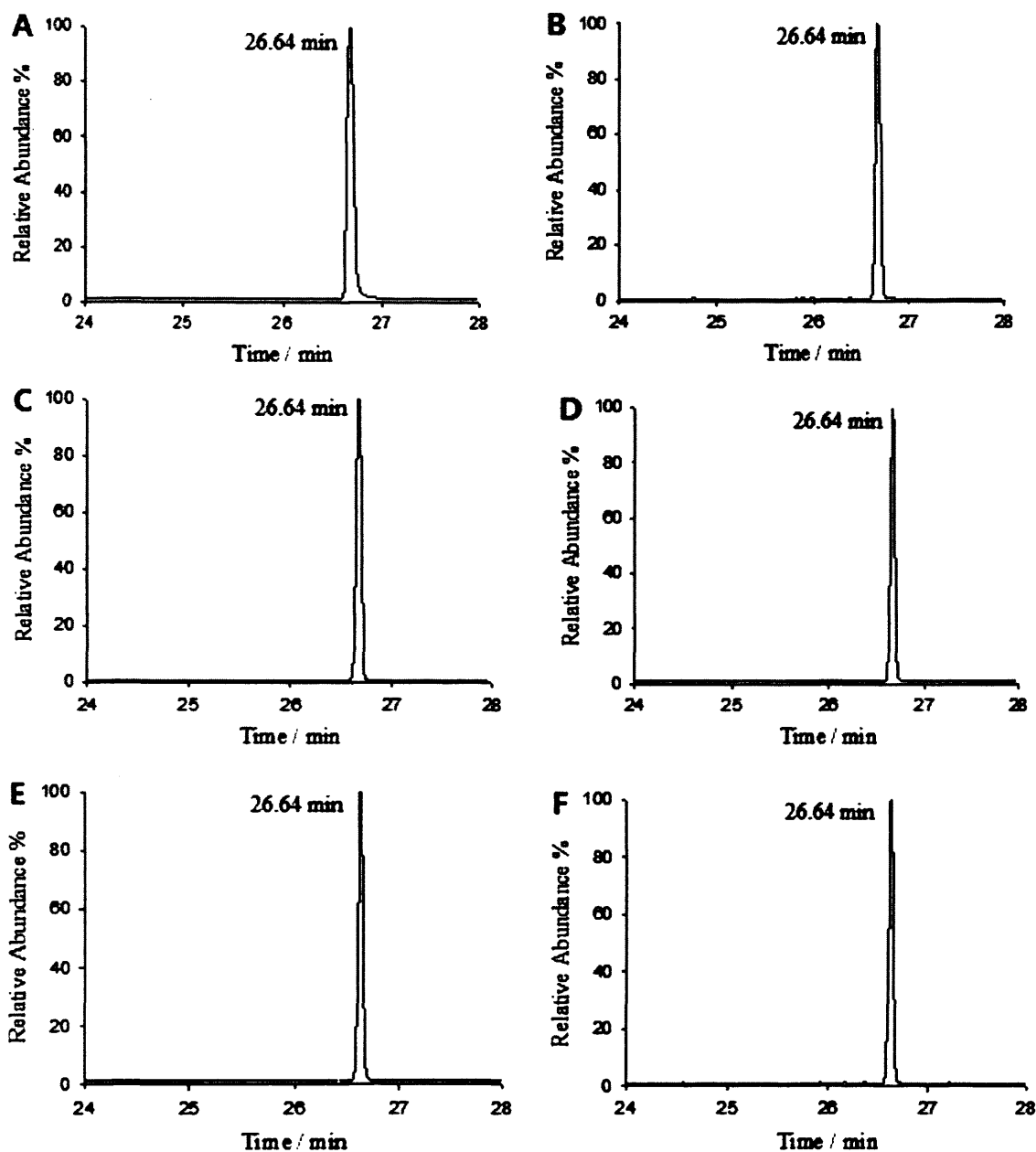
Figure 3.14: *Product profiles for incubation of FDP with M1 (A), M2 (B), M3 (C), M5 (D), M6 (E) and M7 (F).*

In addition to the unchanged product outcome, the mutants that contain modification(s) in the first tier of the active site even showed similar kinetic parameters to the wild-type enzyme. Kinetic assays revealed a k_{cat} value of 0.0203 ± 0.0009 s^{-1} and a K_M value of 13.34 ± 2.56 μM for the mutant M1 (Figure 3.15). The catalytic efficiency (k_{cat}/K_M) of this mutant (1.52 ± 0.31 s^{-1} mM^{-1}) was very similar to that of the wild-type DCS (1.58 ± 0.07 s^{-1} mM^{-1}).
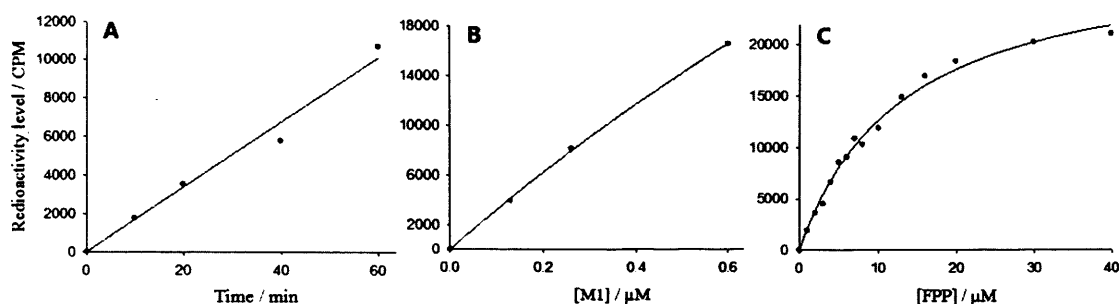
Figure 3.15: *Kinetic analysis of M1.* (A) Time course (0.26 μM enzyme, 5 μM FDP); (B) Plot of radioactivity level against enzyme concentration (5 μM FDP, 20 min incubation time); (C) Michaelis-Menten plot (incubation of 0.26 μM enzyme with radiolabelled substrate for 20 min at 22 °C).

The double mutant M2 was found to be an even more efficient enzyme. A k_{cat} value of 0.0223 ± 0.0037 s^{-1} and a K_M value of 2.89 ± 0.94 μM were obtained from the assay (Figure 3.16). The decrease of the K_M value leads to a roughly 5-fold improvement of catalytic efficiency comparing to the wild-type DCS.
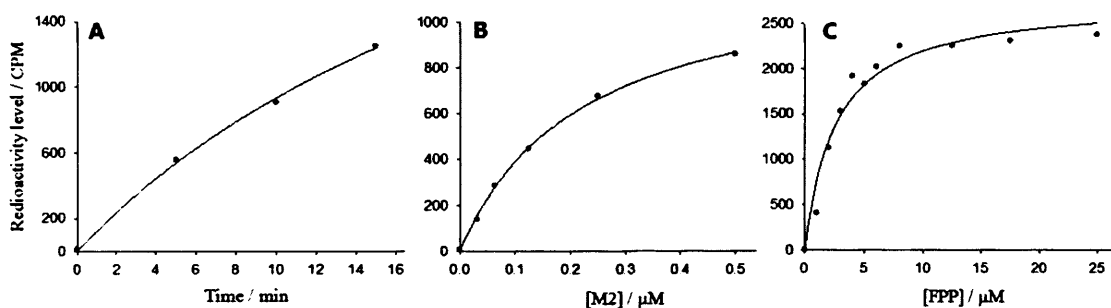
Figure 3.16: *Kinetic analysis of M2.* (A) Time course (0.062 μM enzyme, 5 μM FDP); (B) Plot of radioactivity level against enzyme concentration (5 μM FDP, 10 min incubation time); (C) Michaelis-Menten plot (incubation of 0.062 μM enzyme with radiolabelled substrate for 10 min at 22 °C).

Generally, the mutants containing mutations in both the first and the second tier of the active site experienced a reduction in catalytic activity. The k_{cat} value of M3 was determined to be $(0.04 ± 0.012) × 10^{-3}$ s^{-1}, which was reduced by about 400 fold compared to the wild-type enzyme. The K_M value was increased to 4.65 ± 1.63 μM (Figure 3.17). The catalytic efficiency was therefore reduced by about 200 fold.
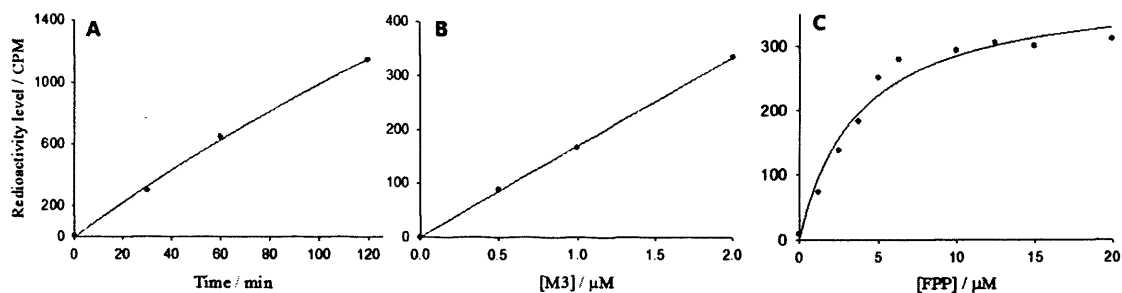
Figure 3.17: *Kinetic analysis of M3.* (A) Time course (1 μM enzyme, 5 μM FDP); (B) Plot of radioactivity level against enzyme concentration (5 μM FDP, 20 min incubation time); (C) Michaelis-Menten plot (incubation of 1 μM enzyme with radiolabelled substrate for 20 min at 22 °C).

The k_{cat} value of M5 was found to be (0.009 ± 0.0008) × 10^{-3} s^{-1} which represented the lowest catalytic activity among all the mutants. However, this mutant showed a much stronger substrate binding affinity compared to wild-type DCS. The K_M value was determined to be 0.84 ± 0.04 μM, which was roughly 12 times smaller than that of wild-type DCS (Figure 3.18). Overall, the catalytic efficiency was reduced by about 140 fold compared to the wild-type enzyme.
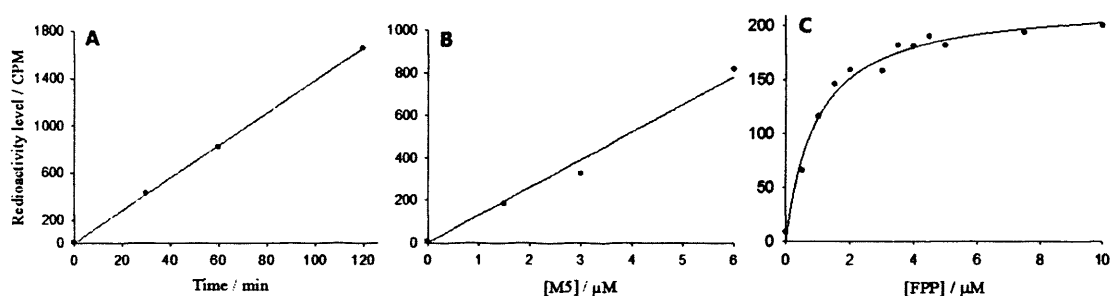
Figure 3.18: *Kinetic analysis of M5.* (A) Time course (2 μM enzyme, 5 μM FDP); (B) Plot of radioactivity level against enzyme concentration (5 μM FDP, 60 min incubation time); (C) Michaelis-Menten plot (incubation of 2 μM enzyme with radiolabelled substrate for 60 min at 22 °C).

The k_{cat} value of M6 was found to be (0.035 ± 0.0003) × 10^{-3} s^{-1} and the K_M value was 1.86 ± 0.31 μM (Figure 3.19). The catalytic efficiency of this mutant (0.019 ± 0.003 s^{-1} mM^{-1}) is comparable to that of the mutants M3 (0.009 ± 0.004 s^{-1} mM^{-1}) and M5 (0.011 ± 0.004) s^{-1} mM^{-1}).
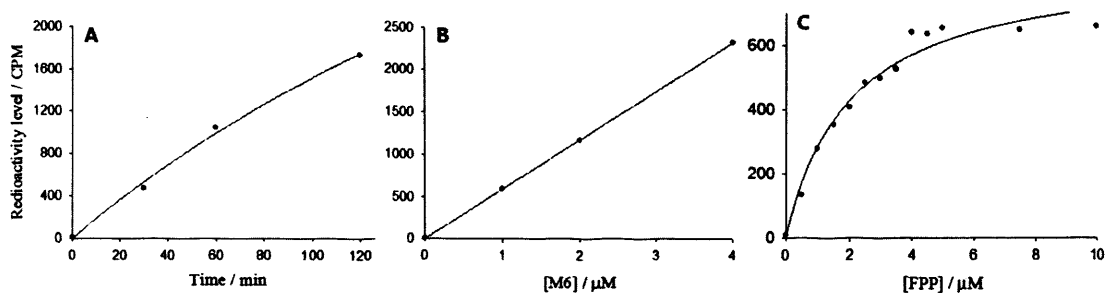
Figure 3.19: *Kinetic analysis of M6.* (A) Time course (2 µM enzyme, 5 µM FDP); (B) Plot of radioactivity level against enzyme concentration (5 µM FDP, 60 min incubation time); (C) Michaelis-Menten plot (incubation of 2 µM enzyme with radiolabelled substrate for 60 min at 22 °C).

The mutant M7, containing all seven chosen mutations, showed a k_{cat} value of $(0.030 \pm 0.006) \times 10^{-3}$ s^{-1} and a K_M value of 0.52 ± 0.07 µM (Figure 3.20). Compared to the wild-type enzyme, the catalytic efficiency of this mutant is only reduced 27 fold.
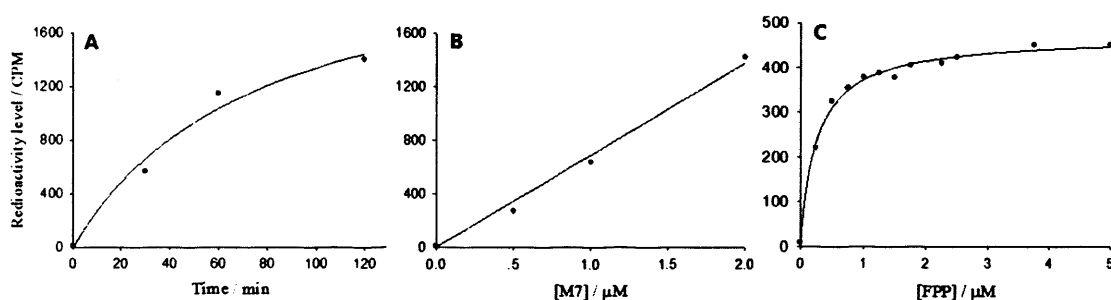
Figure 3.20: *Kinetic analysis of M7.* (A) Time course (1 µM enzyme, 5 µM FDP); (B) Plot of radioactivity level against enzyme concentration (5 µM FDP, 60 min incubation time); (C) Michaelis-Menten plot (incubation of 1 µM enzyme with radiolabelled substrate for 60 min at 22 °C).

The residues identified from the proposed active site of DCS were based on the crystal structure of epi-aristolochene synthase. However, modification of the active site surface showed no influence on the product specificity, leading to a speculation that the enzymatic reaction catalysed by DCS may not take place in the proposed active site. The recently solved crystal structure of DCS (*114*) provides some indications to support this assumption. An obviously different orientation of the farnesyl chain moiety can be identified by comparing the binding conformations of the substrate analogue 2F-FDP in DCS and farnesyl hydroxyl phosphonate (FHP) in

epi-aristolochene synthase (*91, 114*) (Figure 3.21).
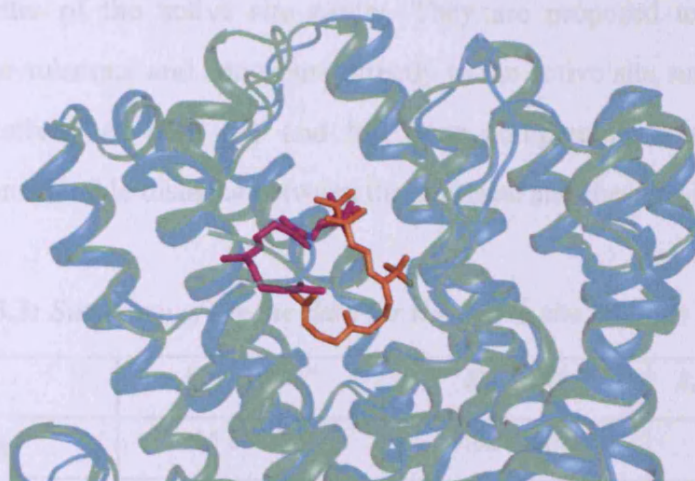
Figure 3.21: *Cartoon representation for the comparison of 2F-FDP (shown as magenta sticks) bound in the crystal structure of DCS (coloured green) (PDB 3G4F, Chain A; (114)) and farnesyl hydroxyl phosphonate (shown as orange sticks) bound in the crystal structure of 5-epi aristolochene synthase (coloured cyan) (PDB 5EAT; (91)). Tertiary structures of DCS and 5-epi aristolochene synthase were aligned using the Pymol program.*

For epi-aristolochene synthase, the farnesyl chain of FHP is docked deeply into the active site pocket, adopting a conformation conductive to the cyclization between C1 and C10. This kind of productive substrate binding fashion can be found in most of the crystal structures that have been solved including monoterpene, sesquiterpene and diterpene synthases (*57, 130*). However, the farnesyl chain of the substrate analogue 2F-FDP was observed binding more towards the entrance of the active site in the crystal structure of DCS. The actual initial binding conformation of substrate FDP may be different from the observed conformation for 2F-FDP because the distal double bond of 2F-FDP is not in close proximity to C1, therefore precluding the proposed C1-C10 ring closure. Nevertheless, it suggests that the native substrate FDP may not be docked deeply into the active site as observed for 5-epi aristolochene synthase.

The kinetic data (summarized in Table 3.3) also provide some support for this

argument. According to the crystal structure of DCS, residues I300 and L413 are found in the middle of the active site and the hydrophobic side chains are positioned toward the centre of the active site cavity. They are proposed to be in a close proximity to the substrate and contribute directly to the active site surface. However, the catalytic activity of both M1 and M2 were comparable to wild-type DCS, suggesting a considerable distance between the substrate and these residues.

Table 3.3: *Summary of kinetic data for the active site mutants of DCS.*

Protein	k_{cat} / 10^{-3} s^{-1}	K_M / µM	k_{cat}/K_M / s^{-1} mM^{-1}
WT-DCS	15.5 ± 0.5	9.82 ± 0.33	1.58 ± 0.07
M1 (DCS-L413I)	20.3 ± 0.9	13.34 ± 2.56	1.52 ± 0.31
M2 (M1-I300L)	22.3 ± 3.7	2.89 ± 0.94	7.7 ± 2.8
M3 (M2-E455G)	0.04 ± 0.012	4.65 ± 1.63	0.009±0.004
M5 (M3-T407S/C408A)	0.009 ± 0.0008	0.84 ± 0.04	0.011±0.004
M6 (M5-M523A)	0.035 ± 0.0003	1.86 ± 0.31	0.019±0.003
M7 (M6-G276C)	0.03 ± 0.006	0.52 ± 0.07	0.058±0.014

E455 of DCS was defined to be outside of active site. The triple mutant M3, constructed by introducing mutation E455G into the double mutant M2, leads to a huge catalytic activity reduction. This is because the E455 is part of the second magnesium binding motif. This mutation could directly compromise the magnesium ion binding affinity, which may also lead to the mispositioning of the farnesyl chain and thus affect catalytic activity. The role of E455 related to the magnesium ion binding has been explained in the crystal study of DCS and the E455A single mutant also shows a significant catalytic activity reduction (*114*).

T407 and C408 of DCS are two residues that also form unique structures in the crystal structure of DCS. T407 makes a water-mediated hydrogen bond with the diphosphate moiety, which would further stabilize the diphosphate group upon binding in the

active site (*114*). C408 and C477 were observed to form a disulfide linkage in the crystal structure of DCS between helices G and H (*114*). The function of this disulfide linkage has not been identified, but this bond could possibly stabilize the two helices and so help to form a well-defined active site contour. These two structures were both compromised in the mutant M5. The changes did not allow any new sesquiterpene hydrocarbons to be formed. In addition, it led to a further four-fold reduction in catalytic activity compared to M3.

G276 and M523 of DCS are found in the middle of the active site pocket. Similar to the function of residues I300 and L413, these two residues may also contribute to the overall active site shape and volume. The catalytic activities of mutants M6 and M7 were better than that of M5 and this was most likely because the modified active site could serve as a better template to direct substrate binding into a suitable conformation for catalysis. On the other hand, the catalytic activities of mutants M6 and M7 were similar to that of M3, suggesting the activity loss caused by disturbance of magnesium binding area cannot be re-established by manipulating the structure of the active site contour.

Generally, the major catalytic activity loss among the mutants was caused by the mutation E455G. Other modifications to the proposed active site did not make an obvious difference in terms of catalytic activity. In addition, the K_M values were found to be smaller in most of the mutants compared to the wild-type DCS. This suggests the overall structure change allowed the mutants to have better binding affinity towards substrate FDP.

Similar results have been reported by manipulating active site residues of amorphadiene synthase (*163*). In this study, the author compared the active site amino acid composition between amorphadiene synthase and TEAS. The only different residue in the active site pocket of the two enzymes was F525 in amorphadiene synthase that corresponds to Y527 in TEAS. In addition, the modelling study revealed

a different positioning of the W271 of amorphadiene synthase in comparison to the corresponding W273 of TEAS and this residue may be essential for positioning of the substrate FDP in the active site of amorphadiene synthase. The different conformation of this tryptophane residue can be caused by the surrounding structures, including residues F514 and Q518 from amorphadiene synthase. Mutagenesis studies toward F514, Q518 and F525 of amorphadiene synthase were carried out, but the substitutions replacing the original amino acid for the corresponding amino acid from TEAS did not lead to any new sesquiterpene product. Only reductions or complete losses of activities were observed among those mutants. These results are in agreement with our findings, suggesting the differences in product specificity of amorphadiene synthase and DCS are not only determined by the core of active sites. To find out the essential component that controls the product specificity of DCS, four chimeras were designed and analysed in the following experiments.

3.3 Domain-swapping studies

Generally, the functional domain of a terpenoid synthase could provide a template that binds flexible isoprenoid substrate(s) with proper orientation and conformation, especially for high-fidelity enzymes. Thus, the closed active site cavity could enforce the generation of reactive carbocations and specific intermolecular or intramolecular carbon-carbon bond forming reactions.

The proposed active site for DCS is in the C-terminal domain. However, mutagenesis studies by sequential substitution of the active site surface showed an unchanged product profile for all the mutants (Section 3.2.3). These results may lead to two hypotheses. One possibility is that the enzymatic reaction catalysed by DCS does not take place in the proposed active site. In this case, the mutations that were introduced into the enzyme may only bring some minor influence to the conformation of the real reaction cavity rather than affecting the substrate or intermediates directly. This could explain the observed activity loss from mutants. The other possibility is that the

113

product specificity of DCS is determined by an unidentified structure rather than the proposed active site contour if the C-terminal domain is the catalytically active domain. Domain-swapping studies were therefore performed to clarify the functional domain of DCS and the residues or subdomain responsible for determining the product specificity.

3.3.1 N-Terminal domain hybrid

The first speculation that arose from the unexpected mutagenesis results was that the N-terminal domain of DCS might catalyse the enzymatic reaction. It has been shown that class I and class II terpenoid synthase folds can be combined in multidomain terpenoid synthases to form larger active site cavities or even bifunctional enzymes (*1*). For example, squalene-hopene synthase is formed by joining two class II domains together in 'face-to-face' fashion, which form a large active site pocket to accommodate the C30 substrate (*141*). The plant sesquiterpene synthase 5-epi aristolochene synthase contains a catalytically active class I terpenoid synthase domain and an inactive class II terpenoid synthase domain (*91*), while the plant diterpene synthase abietadiene synthase, which is structurally homologous to 5-epi aristolochene synthase and taxadiene synthase, is constructed with both domains showing catalytic activity and catalysing tandem cyclization reactions (*57*).

To characterize the function of the N-terminal domain of DCS, a hybrid (hybrid 1) was prepared by substitution of the C-terminal domain of DCS with that of the fungal sesquiterpene synthase aristolochene synthase (from *Penicillium roqueforti*) (Figure 3.22). Aristolochene synthase is active as a single domain, which is homologous to the C-terminal domain of DCS (*164*). Therefore, if the N-terminal domain of DCS is catalytically active, the product profile of this hybrid should give both δ-cadinene and aristolochene as enzymatic products.
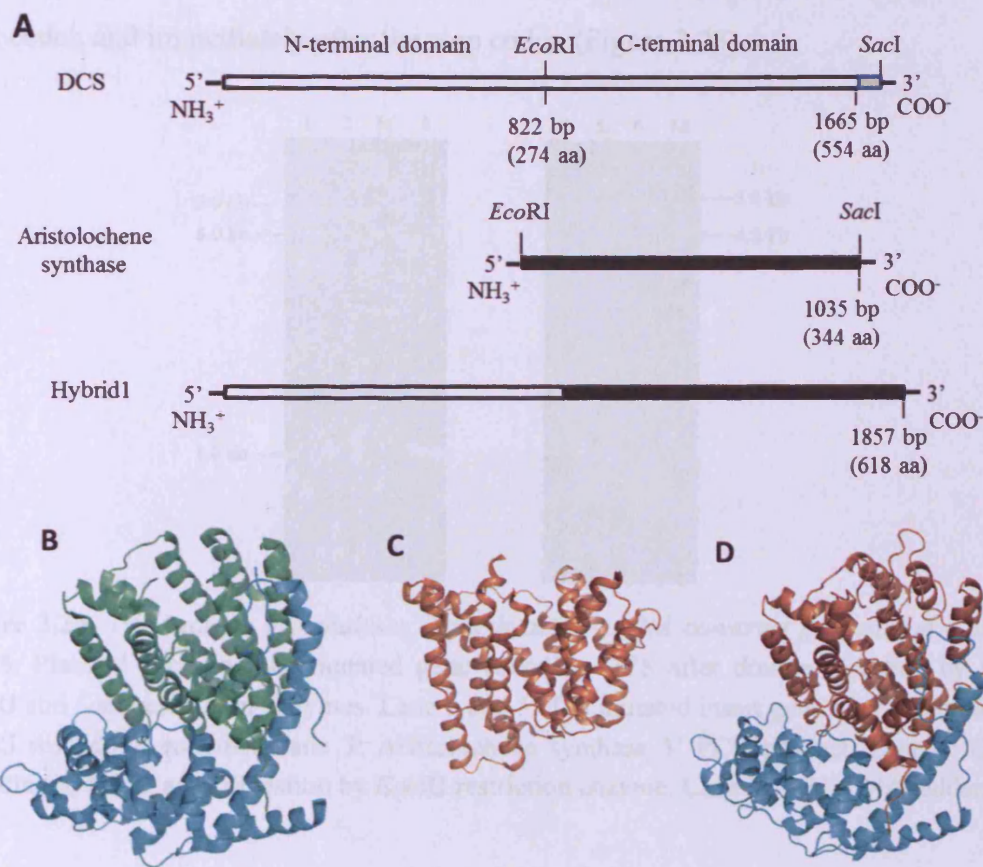
Figure 3.22: *Schematic diagram of the N-terminal domain-swapping test (A) and cartoon representation of homology model of d-cadinene synthase (B), crystal structure of aristolochene synthase (C) and homology model of hybrid1 (D).* Line drawings depict composite diagrams for wild-type DCS (the insert of DCS is coloured white and the vector part is coloured gray), wild-type aristolochene synthase (coloured black) and hybrid 1 genes. The N-terminal domain of DCS is coloured cyan and the C-terminal is green. The structure of AS is coloured orange.

The construct of hybrid 1 was prepared by replacing the native GGT codon of G276 with TTC by site-directed mutagenesis. Together with the codon GAA that encodes amino acid E275, this mutation creates an *Eco*RI restriction site. Another *Sac*I restriction site located after the stop codon was also employed to form the hybrid construct. The gene encoding aristolochene synthase was amplified from the genomic DNA *via* PCR using short complementary oligonucleotide primers designed to flank the gene of interest and provide overhanging *Eco*RI and *Sac*I restriction sites at the 5'

and 3' end respectively. The restriction sites were introduced immediately prior to the start codon and immediately after the stop codon (Figure 3.23).
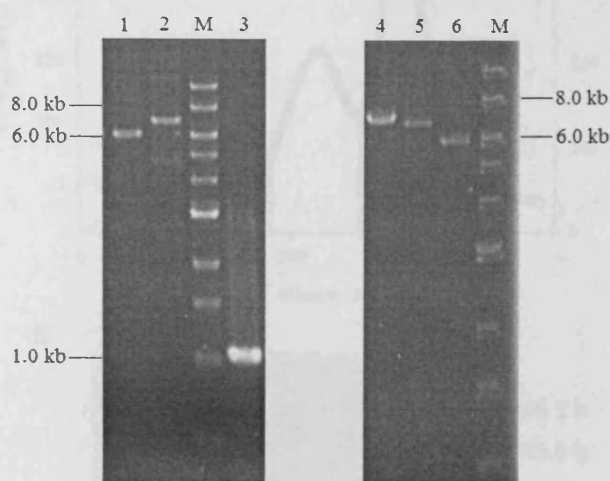
Figure 3.23: 1% *Agarose gel analyses of the hybrid 1 cDNA construct preparation.* Lane 1 and 6: Plasmid encoding the mutated gene encoding DCS after double digestion by both *Eco*RI and *Sac*I restriction enzymes. Lane 2 and 5: The mutated insert gene after digestion by *Eco*RI restriction enzyme. Lane 3: Aristolochene synthase 3' PCR product. Lane 4: Gene encoding hybrid 1 after digestion by *Eco*RI restriction enzyme. Lane M: 1 kb DNA ladder.

The gene fragment that encodes the C-terminal domain of DCS was removed from the mutated plasmid by a double digestion with *Eco*RI and *Sac*I restriction enzymes. The opened plasmid was then ligated with the newly formed DNA fragment encoding aristolochene synthase, creating a plasmid containing the gene that encodes hybrid 1. Formation of the constructed plasmid was checked by a single digestion with *Eco*RI (Figure 3.23) and the correct construct was confirmed by DNA sequencing.

Hybrid 1 was produced and purified in the same manner as wild-type DCS (Figure 3.24), typically yielding 20 mg of purified enzyme per litre of culture. The *in vitro* assay was carried out by incubating 10 μM of purified enzyme with 1 mM of FDP in the same buffer system used for wild-type DCS. GC-MS analysis of the hexane extractable products showed a nearly identical product profile to wild-type aristolochene synthase and no δ-cadinene was observed (Figure 3.25).
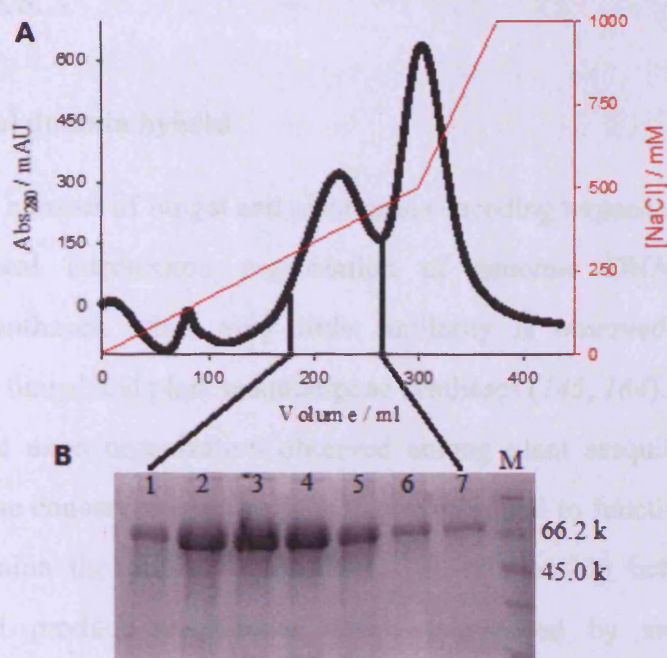
Figure 3.24: *(A) Chromatogram for DEAE purification of hybrid 1. (B) 13% SDS-PAGE analysis*. Lane 1-7: eluent fractions. Lane M: protein marker.
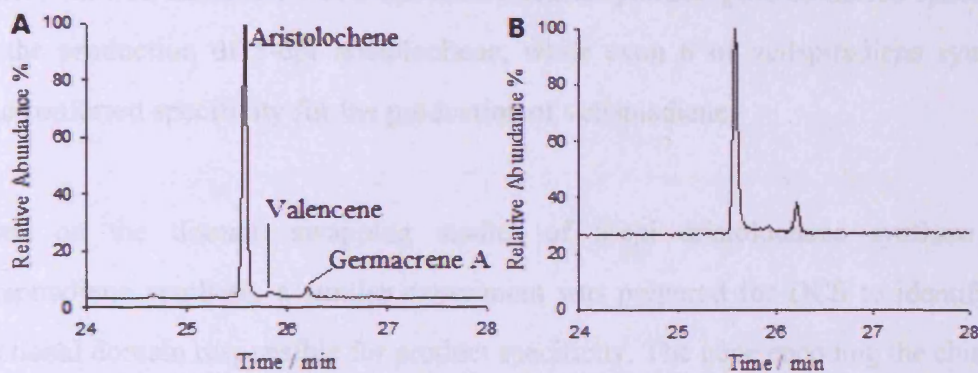
Figure 3.25: *Product profiles for incubation of FDP with wild-type aristolochene synthase (A) and hybrid 1 (B)*.

The study of hybrid 1 showed that the N-terminal domain of DCS is unlikely to be responsible for the conversion of FDP to δ-cadinene. The C-terminal domain of DCS must therefore be the catalytically active domain. In this case, it is very likely that the previously proposed active site contour is not the key component for determining the product specificity of DCS. Further experiments were therefore performed to aid identification of the essential component that controls the specific formation of

117

δ-cadinene by DCS.

3.3.2 C-Terminal domain hybrid

Comparison of a number of fungal and plant genes encoding terpene synthases reveals a nearly identical intron/exon organization of genomic DNA for all plant sesquiterpene synthases, while very little similarity is observed between DNA sequences of the fungal and plant sesquiterpene synthases (*145, 164*). One implication of the conserved exon organization observed among plant sesquiterpene synthase genes is that these conserved exon regions may correspond to functional subdomains that may determine the product specificity. The relationship between functional subdomains and product specificities was demonstrated by swapping regions approximating exons between a 5-epi aristolochene synthase gene and a vetispiradiene synthase gene (*161*). The results from characterization of the resulting hybrids showed that exon 4 of 5-epi aristolochene synthase gene conferred specificity for the production of 5-epi aristolochene, while exon 6 of vetispiradiene synthase gene conferred specificity for the production of vetispiadiene.

Based on the domain swapping studies of 5-epi aristolochene synthase and vetispiradiene synthase, a similar experiment was prepared for DCS to identify the functional domain responsible for product specificity. The gene encoding the chimeric protein (hybrid 2) was designed by substituting the gene regions of DCS that approximate exons 5, 6 and 7 with the corresponding gene regions for GCS (Figure 3.26).
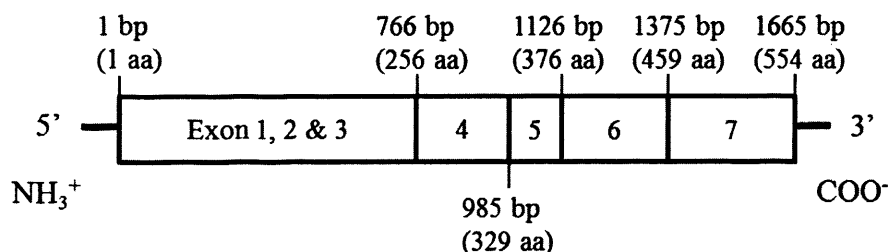
Figure 3.26: *Genomic organization of DCS gene (145).*

The gene of hybrid 2 was constructed using two native SacI restriction sites (Figure 3.26) identified in the plasmid containing the gene encoding DCS; one SacI restriction site sequence encodes amino acids E318 and L319 and the other SacI restriction site is located after the stop codon. A synthetic DNA fragment was designed to encode the amino acid sequence from E311 to the stop codon of GCS, with SacI restriction sites at both the 5' and 3' end (Figure 3.27).
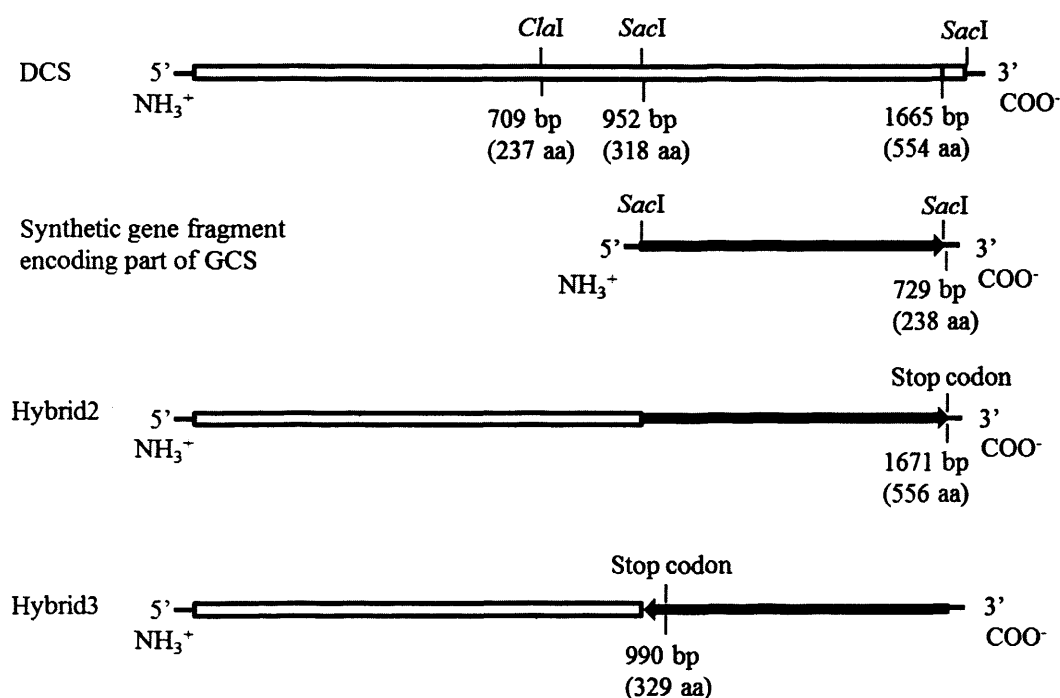
Figure 3.27: *Schematic diagram of the C-terminal domain-swapping tests.* Line drawings depict composite diagrams for wild-type DCS (coloured white), the synthetic gene fragment encoding part of GCS (shown as black arrow), hybrid 2 and hybrid 3 genes.

The plasmid containing the gene encoding DCS was digested with SacI restriction enzyme, dephosphorylated with Calf Intestinal Alkaline Phosphatase (CIP) and the longer DNA fragment was purified by agarose gel electrophoresis (Figure 3.28). The longer opened plasmid was then ligated with the synthetic DNA fragment encoding the 3'-terminal portion of the GCS gene. Formation of the constructed plasmids were checked by a single digestion with ClaI restriction enzyme, which cuts the resulting plasmid only once (Figure 3.28).
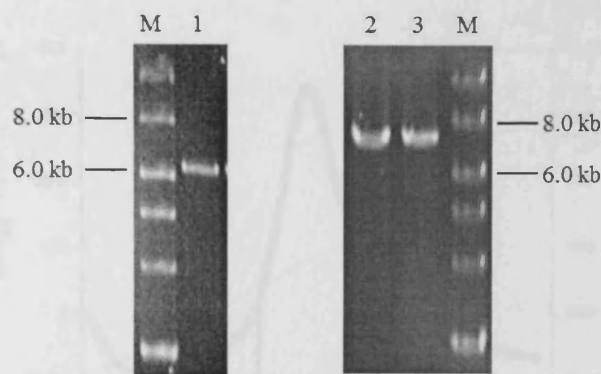
Figure 3.28: 1% *Agarose gel analyses of the hybrid2 and hybrid3 cDNA construct preparation.* Lane 1: The plasmid containing the gene encoding DCS after digestion by *Sac*I restriction enzyme. Lane 2 and 3: The ligation products after digestion by *Cla*I restriction enzyme. Lane M: 1 kb DNA ladder.

Since both 5' and 3' restriction sites employed the same *Sac*I site, the synthetic DNA fragment can be ligated into the open insert gene in two different manners (Figure 3.27). The desired gene construct for hybrid 2 was confirmed by DNA sequencing. The other construct was also identified, in which the synthetic DNA fragment was ligated into the insert vector in a reverse pattern. The gene sequence of this chimeric protein (hybrid 3) contains a stop codon after that encoding N329 amino acids, truncating the protein at that position.

Hybrid 2 was produced and purified by the same method as wild-type DCS (Figure 3.29). Typically, 15 mg of purified protein can be obtained per litre of culture. The *in vitro* assay was carried out by incubating 10 µM purified enzyme with 1 mM FDP in the same buffer system used for wild-type DCS. GC-MS analysis of the hexane extractable products revealed δ-cadinene as the sole sesquiterpene product (Figure 3.30).
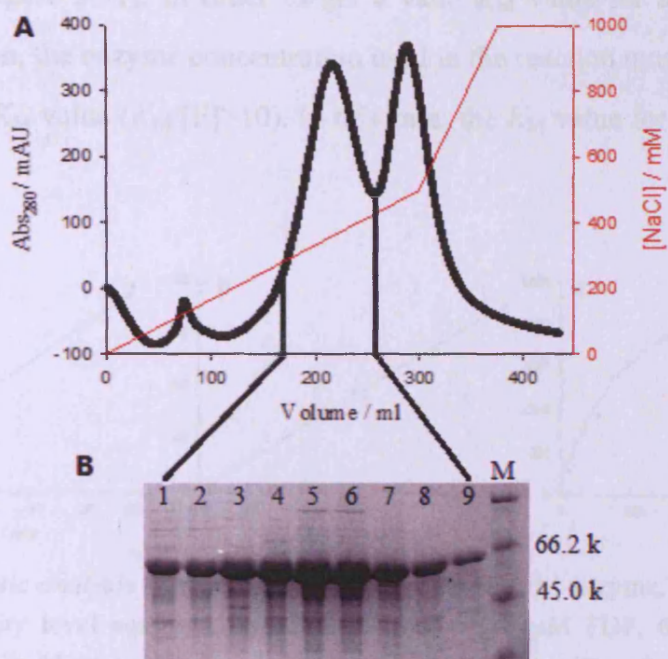
Figure 3.29: *(A) Chromatogram for DEAE purification of hybrid 2.*
(B) 13% SDS-PAGE analysis. Lane 1-9: eluent fractions. Lane M:
protein marker.
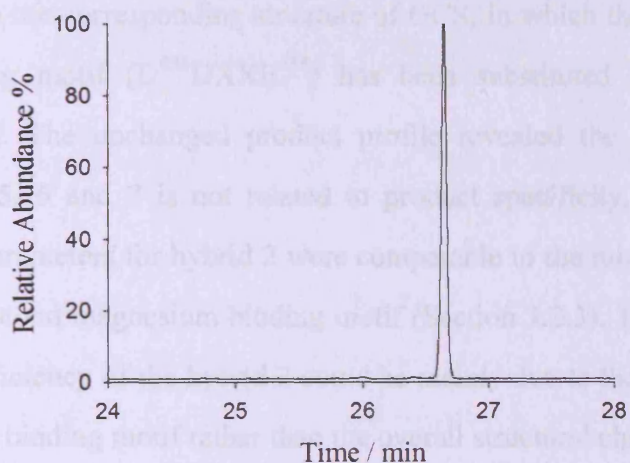
Figure 3.30: *GC trace representing product profile for incubation of FDP with hybrid 2.*

For the kinetic study of hybrid 2, the reaction time and the enzyme concentration were
optimized to be 1 hr and 1 μM respectively before the actual run. A k_{cat} value of
$(0.025 \pm 0.002) \times 10^{-3}$ s^{-1} and a K_M value of 0.44 ± 0.04 μM were obtained from the

kinetic assay (Figure 3.31). In order to get a valid K_M value for an enzyme using Michaelis-Menten, the enzyme concentration used in the reaction must be much lower than the actural K_M value (K_M/[E]>10). In this case, the K_M value for the hybrid2 was approximated.
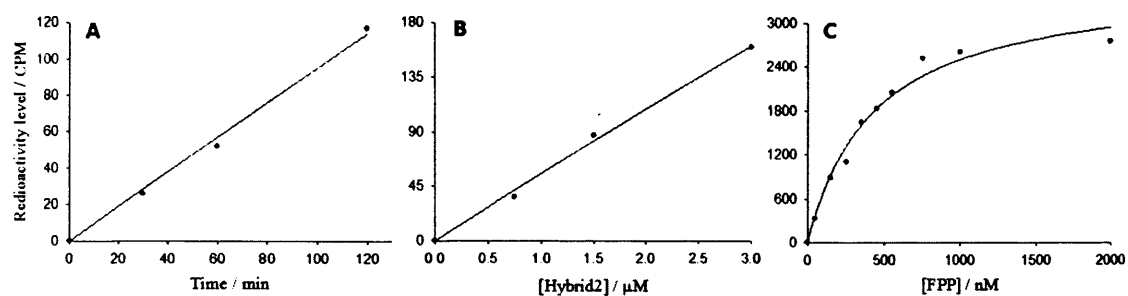
Figure 3.31: *Kinetic analysis of hybrid 2.* (A) Time course (1 μM enzyme, 10 μM FDP); (B) Plot of radioactivity level against enzyme concentration (10 μM FDP, 60 min incubation time); (C) Michaelis-Menten plot (Incubation of 1 μM enzyme with radiolabelled substrate for 60 min at 22 °C).

In hybrid 2, roughly 2/3 of the functional C-terminal domain structure of DCS has been swapped with the corresponding structure of GCS, in which the original second magnesium binding motif ($D^{451}DXXE^{455}$) has been substituted with an unusual $N^{450}G^{454}E^{458}$ motif. The unchanged product profile revealed the protein structure encoded by exon 5, 6 and 7 is not related to product specificity. In addition, the observed kinetic parameters for hybrid 2 were comparable to the mutants that contain mutations in the second magnesium binding motif (Section 3.2.3). Thus, it seems the loss of catalytic efficiency of the hybrid 2 could be mainly due to the alteration of the second magnesium binding motif rather than the overall structural change.

It has been shown in hybrid 1 that the N-terminal domain of DCS, encoded by exon 1, 2 and most of exon 3, does not contribute to the product specificity. The result from hybrid 2 indicates that amino acids from E318 to L554 of DCS, corresponding to exons 5, 6 and 7, do not determine the product specificity either. These results strongly suggest that exon 4 of the DCS gene confers specificity for the production of δ-cadinene.

Hybrid 3 was produced and purified by the same method as wild-type DCS (Figure 3.32A). The protein with a molecular weight around 38,000 Da was identified by the SDS-PAGE analysis (Figure 3.32B). Typically, 15 mg of purified protein can be obtained per litre of culture.
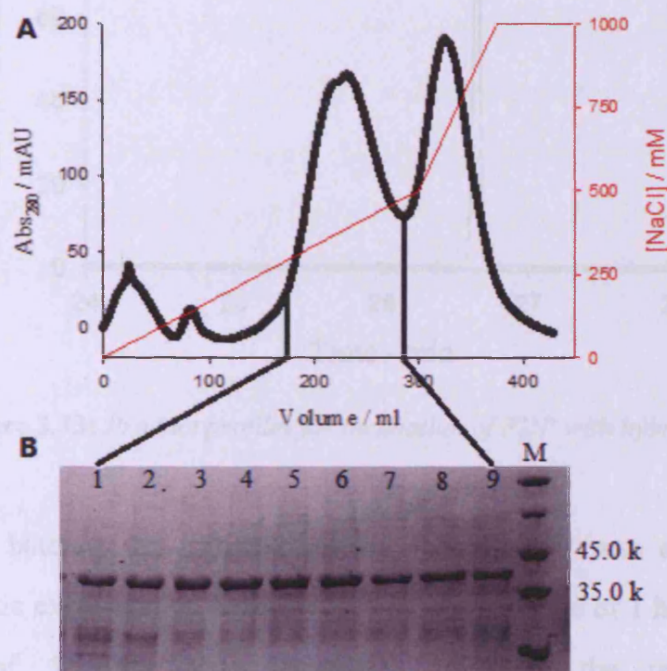
Figure 3.32: *(A) Chromatogram for DEAE purification of hybrid 3. (B) 13% SDS-PAGE analysis.* Lane 1-9: eluent fractions. Lane M: protein marker.

Hybrid 3 can be defined as a truncated mutant of DCS; roughly 2/3 of the C-terminal domain has been removed from the original protein structure, resulting in a protein that encoded by exons 1 to 4. This mutant was expected to be inactive because the truncated active site is very likely exposed to the solvent and unable to protect the cationic intermediates. Surprisingly, the *in vitro* assays of this mutant with FDP suggest a different story. In total, eight individual protein productions and purifications were performed for hybrid 3. Half of these purifications yielded enzyme that showed catalytic activity towards FDP, with δ-cadinene as the only product according to GC-MS analysis (Figure 3.33); while half of the time, the hybrid 3 was

inactive.
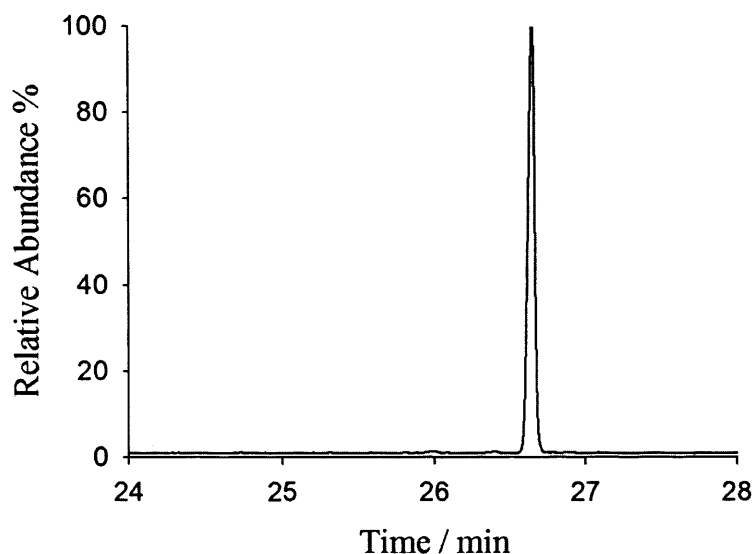
Figure 3.33: *Product profiles for incubation of FDP with hybrid 3.*

Hybrid 3 from batches that showed activity toward FDP was employed in the subsequent kinetic experiments. An optimised incubation time of 1 hr and an enzyme concentration of 1 µM were determined to meet the requirements for Michaelis-Menten kinetic assay before the actual run (Figure 3.34). The k_{cat} and K_M were determined to be $(0.073 \pm 0.013) \times 10^{-3}$ s^{-1} and 0.43 ± 0.13 µM respectively. Again, since the enzyme concentration value used in this reaction is bigger than the calculated K_M value, the calculated K_M value is only approximated here.
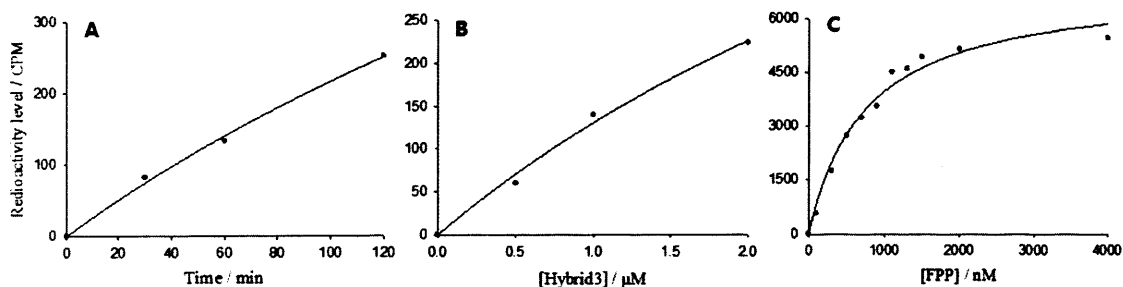
Figure 3.34: *Kinetic analysis of hybrid 3.* (A) Time course (1 µM enzyme, 10 µM FDP); (B) Plot of radioactivity level against enzyme concentration (10 µM FDP, 60 min incubation time); (C) Michaelis-Menten plot (Incubation of 1 µM enzyme with radiolabelled substrate for 60 min at 22 °C).

The approximate K_M value was reduced by roughly 20 fold than that of wild-type DCS and the k_{cat} value was 3 times greater than that of the hybrid 2, indicating the activity observed from hybrid 3 was not due to the contamination either from wild-type enzyme or the hybrid 2. The unpredictable catalytic activity of hybrid 3 may be because of the unstable tertiary structure. The remaining C-terminal domain and the N-terminal domain may adjust their orientations to form an enclosed space that could catalyze the cyclization from FDP to δ-cadinene. On the other hand, the formation of the active conformation of hybrid 3 may not be easily achieved, thus the mutant sometimes behaves as an inactive enzyme probably due to the inappropriate tertiary structure. In any case, the active form of hybrid 3 provides further evidence to support the theory that exon 4 of the DCS gene confers specificity for the production of δ-cadinene. However, more truncated mutants could be designed to verify this result conclusively.

3.4 Mutagenesis of the first aspartate rich motif

The gene sequence encoding the aspartate rich $D^{307}DTYD^{311}$ motif was identified in the exon 4 and this motif was proposed to interact with the putative Mg^{2+}_A and Mg^{2+}_C ions (57, 145). To further probe the role of this magnesium binding motif, three aspartate residues were substituted with alanine respectively.

All the three mutants D307A, D308A and D311A were prepared by substitution of the native codon that encodes the relevant Asp residue with a GCG codon of Ala by site-directed mutagenesis. These mutants were produced and purified in the same way as wild-type DCS (Figure 3.35). Typically, 20 mg of purified protein was obtained per litre of culture for each mutant. The *in vitro* assays followed the standard method in that 10 μM of the purified mutant was incubated with 1 mM FDP overnight. δ-Cadinene was identified as the sole hexane extractable product for all three mutants by GC-MS analysis (Figure 3.36).
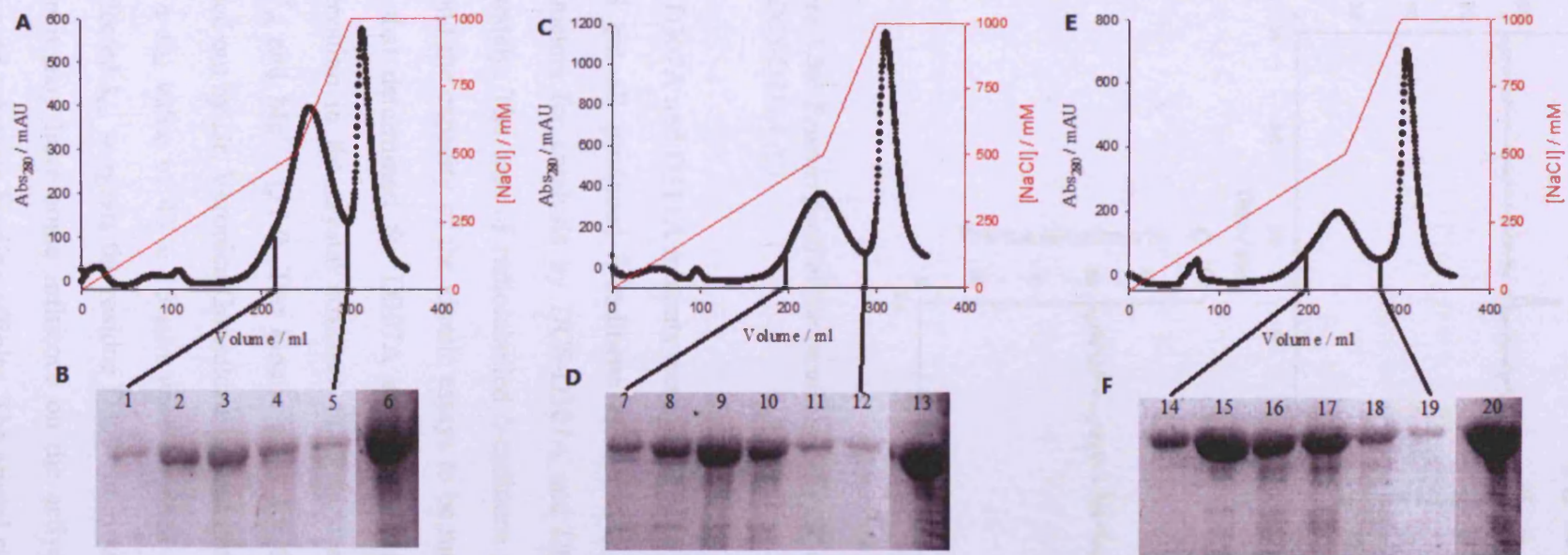
Figure 3.35: *Chromatograms for DEAE purification of DCS-D307A (A), DCS-D308A (C) and DCS-D311A (E) and 13% SDS-PAGE analyses of each purification (B), (D) and (E).* Lane 1-5: eluent fractions for DCS-D307A. Lane 7-12: eluent fractions for DCS-D308A. Lane 14-19: eluent fractions for DCS-D311A. Lane 6, 13 and 20: wild-type DCS sample.
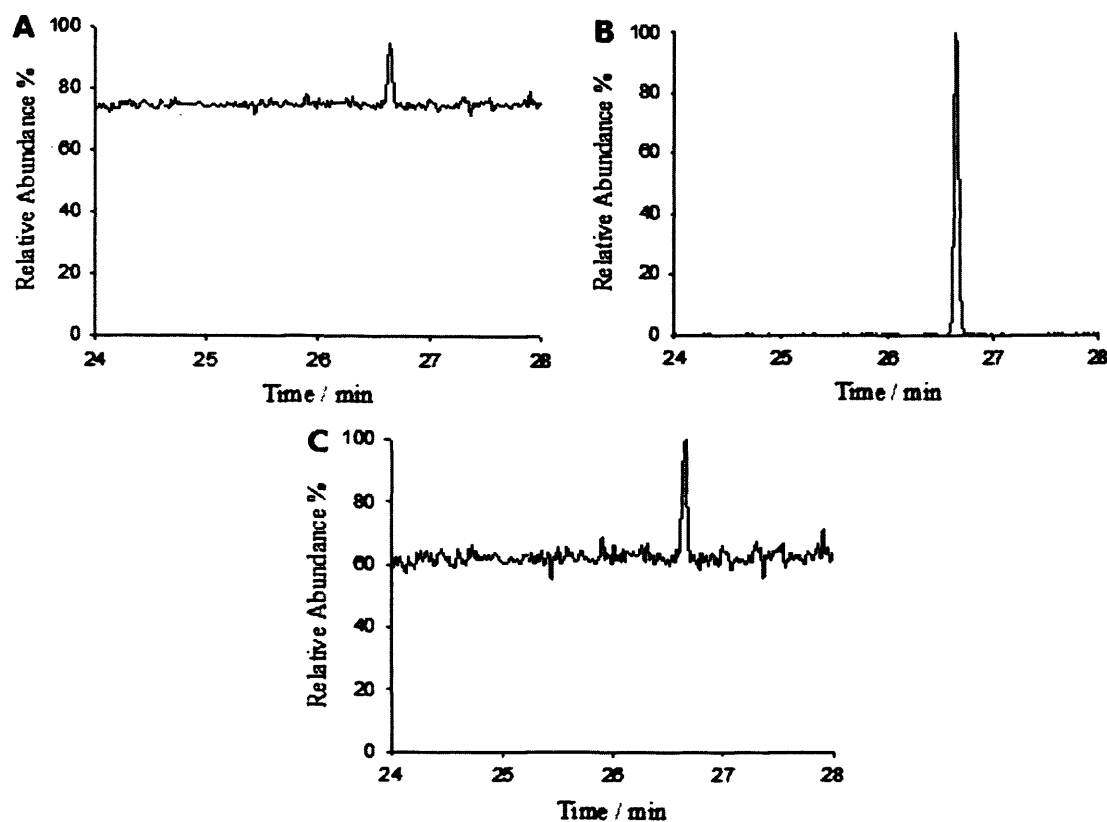
Figure 3.36: *Product profiles for incubation of FDP with DCS-D307A (A), DCS-D308A (B) and DCS-D311A (C).*

The D307A and D311A mutants were found to be much less active than the wild-type DCS but all produced δ-cadinene as the only product. The steady-state kinetic parameters for catalysis by DCS-D307A and DCS-D311A could not be measured accurately. The rate of radiolabelled δ-cadinene production was so low that it was beyond the capacity of the kinetic assays to be measured. The huge catalytic activity loss that determined for D307A and D311A substitutions is consistent with the observation in the crystal structure that residues D307 and D311 coordinate with Mg^{2+}_A and Mg^{2+}_C (*114*). The kinetic assay for the other mutant DCS-D308A was carried out by Dr. Veronica Gonzalez from our group. A k_{cat} value of 0.012 ± 0.002 s^{-1} and a K_M value of 43 ± 16 μM were obtained from the kinetic assay (*114*). The unaffected k_{cat} suggests the residue D308 contributes very little to metal ion binding, while it may have some influence on the active site contour, as reflected by the decreased substrate binding affinity. The crystal structure of DCS also confirmed that D308 is not involved in the interactions that chelate the magnesium ions (*114*).

127

3.5 Conclusions

Wild-type DCS from *Gossypium arboretum* was shown to be constructed with an overall α-helical fold by CD spectroscopy and confirmed to be a high-fidelity enzyme by producing δ-cadinene as the sole product. The relationship between the product specificity and the protein structure was investigated first by mutagenesis studies based on the sequence alignment and homology modelling. Seven residues from the proposed active site were chosen to be substituted by the corresponding residues of GCS and to study a possible enzyme activity interconversion between DCS and GCS for this study. However, the product profiles displayed by all the variants were the same as WT-DCS, suggesting that the active site contour was not the key component for determining the product specificity of DCS.

In addition, the determined kinetic constants for the mutants suggest most of these mutations did not make obvious contributions to catalysis. The catalytic activities of mutants M1 and M2 were comparable to wild-type DCS. Mutant M3 showed a significant catalytic activity reduction of three orders of magnitude compared to wild-type DCS. The observed activity loss of M3 was caused by the newly introduced mutation E455G. This mutation would greatly compromise the interaction between the second magnesium binding motif and the Mg^{2+}_B. A further catalytic activity reduction was observed in the mutant M5, probably due to the additional mutation T407S. This mutation may compromise the potential water-mediated hydrogen bond between T407 and the diphosphate moiety. The mutants M6 and M7 contain additional modifications in the proposed active site, while the catalytic activities of both mutants are comparable to that of M3.

Generally, the mutations were sequentially introduced into the wild-type DCS, but the mutations are non-additive in terms of k_{cat}. The major activity loss was caused by alteration of the magnesium binding motif rather than amino acid substitution of the

proposed active site surface. These results suggest that the product specificity of DCS is possibly determined by an unidentified structure rather than the immediate active site contour.

The possibility that the N-terminal domain of DCS is the major catalytically active domain was ruled out by characterising the function of hybrid 1. Hybrid 1 contains the C-terminal domain from aristolochene synthase and the N-terminal domain from DCS. The product profile of hybrid 1 is identical to that of aristolochene synthase, indicating the enzymatic reaction of DCS should take place in the C-terminal domain. Nevertheless, it is still possible that the N-terminal domain of DCS could catalyse the conversion from FDP to δ-cadinene, but in a very slow rate.

As the above studies showed that either the active site contour or the N-terminal domain could not affect the product specificity of DCS; thus, the product specificity was expected to be related to a specific region from the C-terminal domain. The study of hybrid 2 suggested that the structure of amino acid sequence from E318 to L554 of DCS did not contribute in regard to product specificity, since hybrid 2 was constructed by substitution of this region with the corresponding region of GCS and the mutant converted FDP into δ-cadinene exclusively. The C-terminal domain is mainly encoded by exons 4 to 7 and the examined region from E318 to L554 of DCS is encoded by exons 5, 6 and 7. Therefore, the exon 4 of DCS gene most likely confers the specificity for the production of δ-cadinene. This assumption was supported by the result from hybrid 3 which was a truncated mutant encoded only by exons 1 to 4. This mutant was able to convert FDP into δ-cadinene, indicating that either the protein structure encoded by exon 4 was the key component for controlling specificity, or this partial C-terminal domain combining with the whole N-terminal domain could form a new active site, which also possesses the catalytic activity to convert FDP into δ-cadinene.

The role of the first aspartate rich motif was confirmed by alanine substitutions. The

significantly catalytic activity loss of D307A and D311A mutants indicated the reactions between these two residues with magnesium ions. This is consistent with the observation in the crystal structure that residues D307 and D311 coordinate with Mg^{2+}_A and Mg^{2+}_C. The unaffected catalytic activity of D308A suggests that Mg^{2+}_A and Mg^{2+}_C are only chelated by D307 and D311, which is also confirmed in the crystal structure.

Chapter 4: (*E*)-β-Farnesene Synthase

4.1 Characterisation of the wild-type (*E*)-β-farnesene synthase (WT-EBFS)

This chapter aimed at examining the influence of the active site volume of EBFS on the product outcome. Mutants with enlarged or refined active site were designed to try to facilitate cyclization reactions in the modified active site pocket. The active site cavity was further modified by swapping residues from the active site of EBFS with corresponding ones from DCS to demonstrate the importance of active site volume.

4.1.1 Expression and purification of WT-EBFS

The cDNA clone of WT-EBFS was previously subcloned into the expression vector pET32b by Dr Athina Deligeorgopoulou with *Nco*I and *Bam*HI restriction sites at the 5' and 3' ends. The gene was expressed in BL21-CodonPlus(DE3)-RP cells. To optimise the expression conditions, different expression temperatures were investigated. Cells were grown at 37 °C to an OD_{600} of 0.6 before expression was induced by addition of IPTG (0.1 mg/ml) at 16 °C and 37 °C, respectively (Figure 4.1). As indicated from the SDS gel, very little WT-EBFS was observed at 37 °C until 7 hr, but a large amount of the protein was produced at 16 °C after 7 hr.
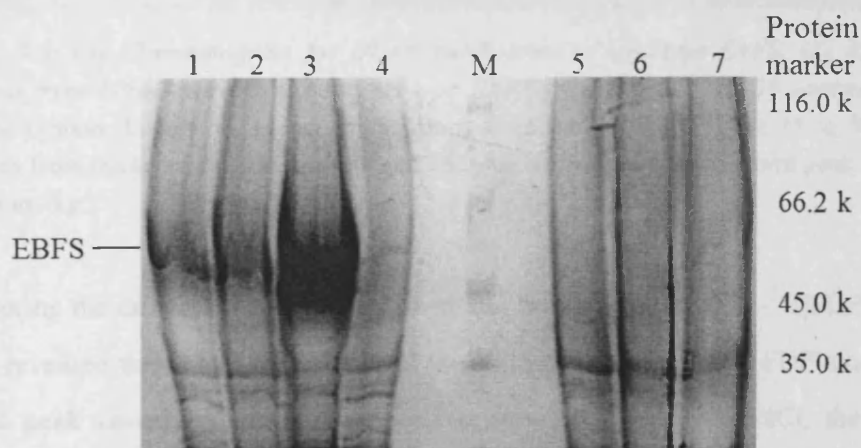
Figure 4.1: *13% SDS-PAGE analysis of wild-type EBFS production in comparison to total protein production in* E. coli *BL21-CodonPlus(DE3)-RP at 16 and 37 °C respectively.* Lanes 1 to 3: 2, 4 and 7 hr at 16 °C after induction of expression. Lanes 4 and 5: 37 °C immediately prior to induction of expression. Lane 6 and 7: 2 and 7 hr at 37 °C after induction of expression.

The inclusion bodies that formed during protein production were pelleted by centrifigation. Refolding of inclusion body proteins followed the same base extraction protocol as described for wild-type DCS (Section 2.7.3). The refolded protein was purified on a DEAE column. Three peaks can be identified from the FPL chromatogram (Figure 4.2A). According to the SDS gel, the first two peaks contained the majority of the enzyme with the correct molecular weight (Figure 4.2C).

6 7 8 9 10 11 12 13 14 15 16 17 18 19   20 21 22 23 24 25 26 27 28 29 30 M 37 38

66.2 k
45.0 k
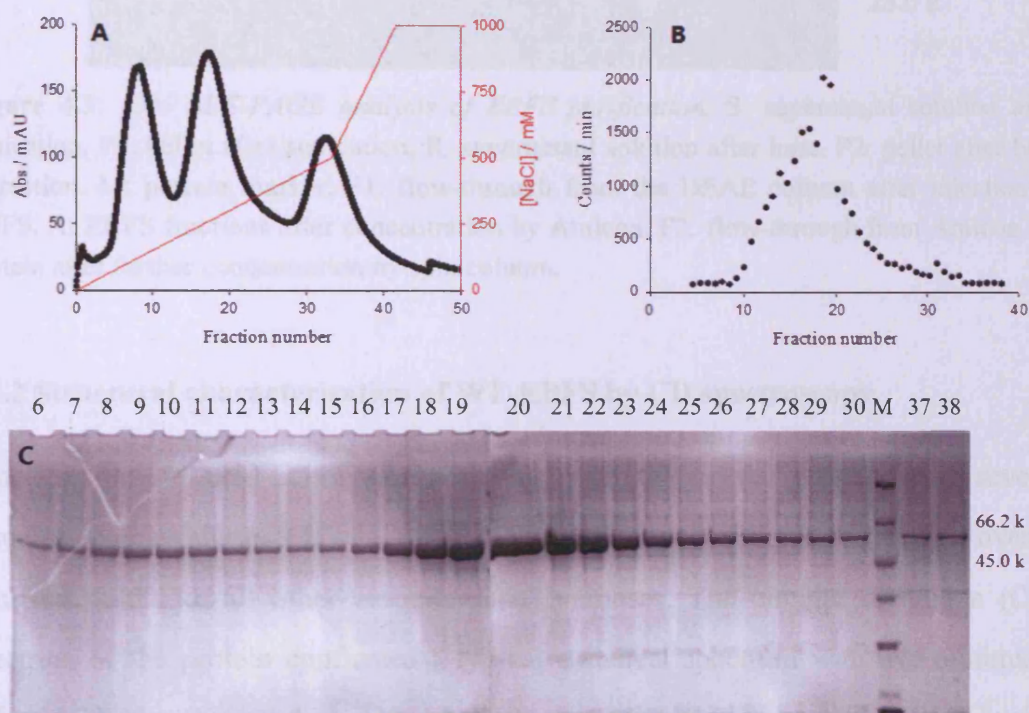
Figure 4.2: *(A) Chromatogram for DEAE purification of wild-type EBFS. (B) Activity of fractions from DEAE purification of wild-type EBFS. (C) 13% SDS-PAGE analyses of the purified protein.* Lane 6 to 14: eluent fractions from the first peak. Lane 15 to 30: eluent fractions from the second peak. Lane 37 and 38: eluent fractions from the third peak. Lane M: protein marker.

Monitoring the enzymatic activity of the eluted fractions utilizing $[1-^3H]FDP$ (Figure 4.2B) revealed that, while the first and the third peak turned over FDP slowly, the second peak contained most of the active protein. To remove NaCl, the protein solution from the second peak was combined and dialyzed using MediCell membranes Size 3500/2 (cut-off of 0-3500) against buffer that contained 10 mM MOPS and 5 mM β-mercaptoethanol at pH 7.2. The dialysate was concentrated to

133

approximately 5 ml using an Amicon™ ultrafiltration apparatus containing a Millipore 30 k cutoff membrane, at a pressure of 1.5 bar at 4 ℃ (Figure 4.3).
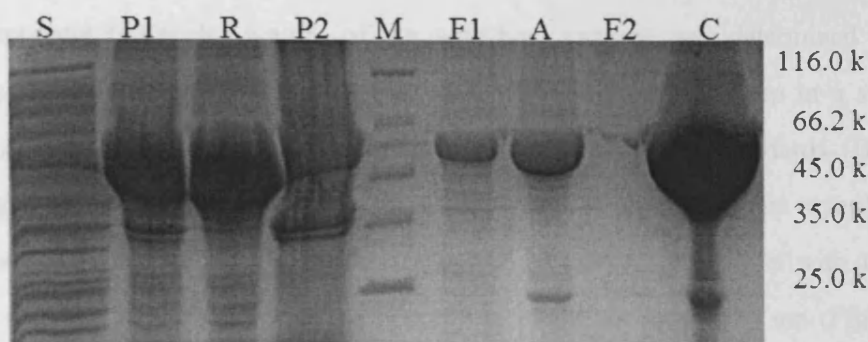
Figure 4.3: *13% SDS-PAGE analysis of EBFS purification.* S: supernatant solution after sonication. P1: pellet after sonication. R: supernatant solution after base. P2: pellet after base extraction. M: protein marker. F1: flow-through from the DEAE column after injection of EBFS. A: EBFS fractions after concentration by Amicon. F2: flow-through from Amicon. C: protein after further concentration by spin column.

4.1.2 Structural characterisation of WT-EBFS by CD spectroscopy

Based on amino acid sequences homology and the crystal structures of several sesquiterpene synthases (*57, 93, 114, 165*), EBFS was predicted to have an overall α-helical fold like all other sesquiterpene synthases. The circular dichroism (CD) spectrum of the protein confirmed a typical α-helical spectrum with two minima at 208 and 222 nm (Figure 4.4). The a-helicity was calculated to be 34% at 20 °C using K2d software (*153*).
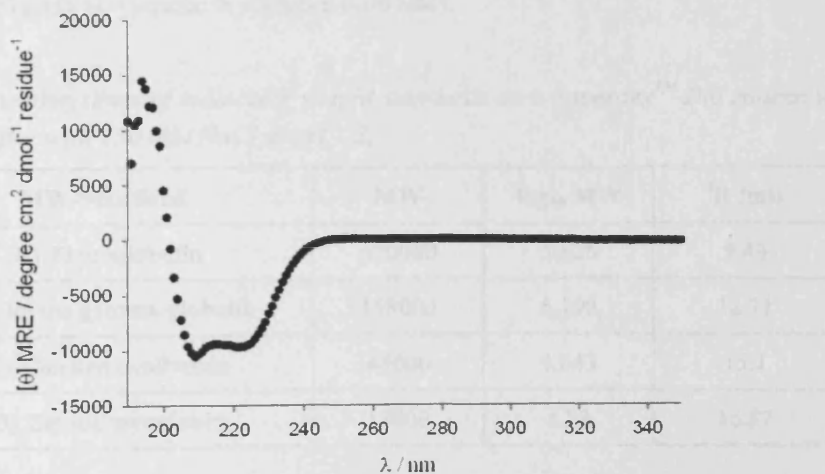
Figure 4.4: *CD spectrum of EBFS at 20 °C in potassium phosphate buffer (10 mM, pH 7).*

4.1.3 Determination of the apparent molecular mass of WT-EBFS

The apparent molecular weight of the wild-type enzyme was determined by size exclusion chromatography, which is commonly used as a final step in a series of chromatographic procedures. The retention times of protein standards (Bio-Rad) thyroglobulin, bovine gamma-globulin, chicken ovalbumin and equine myoglobin on a Superdex^{TM} 200 column were determined in 20 mM MOPS buffer with 150 mM NaCl at pH 7.2 by monitoring the absorption of the eluent at 280 nm (Figure 4.5, Table 4.1). The dependence of the log of the molecular mass on a retention time was determined by linear regression (Figure 4.6).
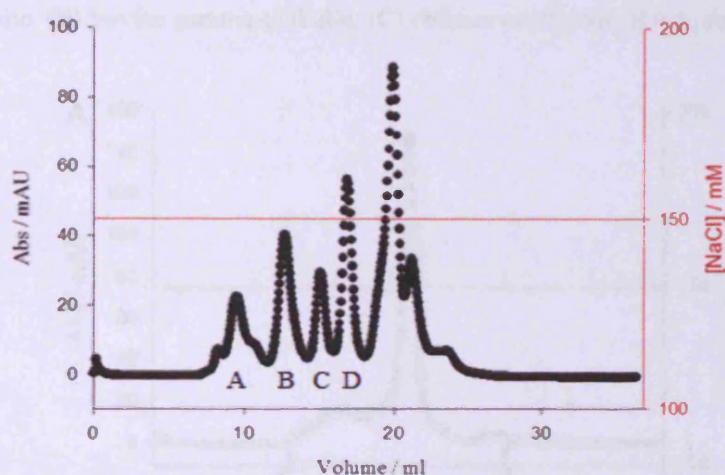
Figure 4.5: *Chromatogram for elution of molecular mass standards (50 µl) from a Superdex^{TM} 200 size exclusion column at 20 °C in MOPS buffer (20 mM MOPS, 150 mM NaCl, pH 7.2).* (A) thyroglobulin (20 µM). (B) bovine gamma-globulin (80 µM). (C) chicken ovalbumin (300 µM). (D) equine myoglobin (400 µM).

Table 4.1: *Retention times of molecular weight standards on a Superdex^{TM} 200 column in 20 mM MOPS buffer with 150 mM NaCl at pH 7.2.*

MW-Standard	MW	log$_{10}$ MW	^{t}R (ml)
(A) Thyroglobulin	670000	5.826	9.49
(B) Bovine gamma-globulin	158000	5.199	12.71
(C) Chicken ovalbumin	44000	4.643	15.1
(D) Equine myoglobin	17000	4.23	16.87

$$y = -0.2166x + 7.9079$$
$$R^2 = 0.9977$$
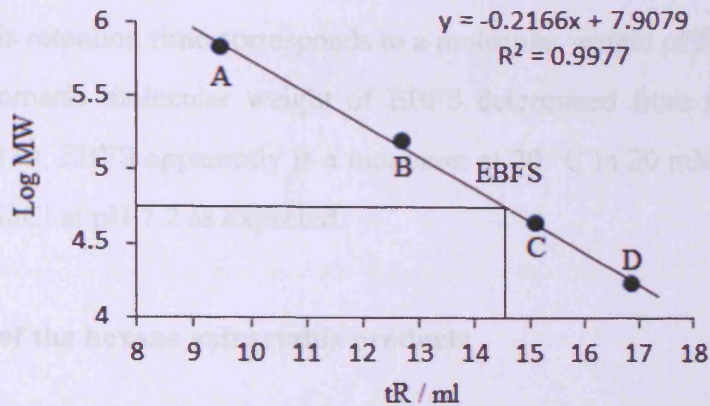
Figure 4.6: *Retention times (in ml) of molecular weight standards and EBFS on a Superdex^{TM} 200 column in 20 mM MOPS buffer with 150 mM NaCl at pH 7.2.* (A) thyroglobulin. (B) bovine gamma-globulin. (C) chicken ovalbumin. (D) equine myoglobin.
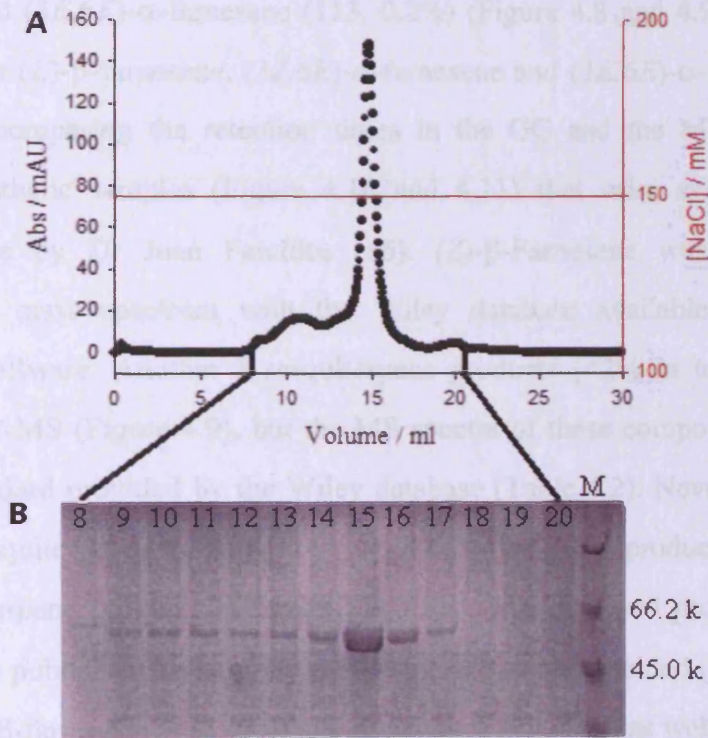
Figure 4.7: *(A) Chromatogram for Superdex^{TM} 200 purification of wild-type EBFS. (B) 13% SDS-PAGE analysis of the purified protein.* Lane 8 to 20: fractions from 8 to 20 (1 ml per fraction). Lane M: protein marker.

Figure 4.7 showed the purification of EBFS when 100 μl of 250 μM enzyme was loaded onto the size exclusion column. Different concentrations (20, 40, 250 and 600

μM) of EBFS all showed the same retention time of 14.56 ml and eluted as a single major peak. This retention time corresponds to a molecular weight of 56781 for EBFS. Since the monomeric molecular weight of EBFS determined from the amino acid sequence is 63830, EBFS apparently is a monomer at 20 °C in 20 mM MOPS buffer with 150 mM NaCl at pH 7.2 as expected.

4.1.4 Analysis of the hexane extractable products

Incubation of purified EBFS with FDP in the presence of 5 mM Mg^{2+} gave (E)-β-farnesene (65, 95%) as the dominant product, along with a small amount of three other acyclic sesquiterpenes, (Z)-β-farnesene (97, 1.5%), (3Z,6E)-α-farnesene (112, 1.3%) and (3E,6E)-α-farnesene (113, 0.2%) (Figure 4.8 and 4.9). The identity of the peaks for (E)-β-farnesene, (3Z,6E)-α-farnesene and (3E,6E)-α-farnesene were confirmed by comparing the retention times in the GC and the MS spectra with chemically synthetic samples (Figure 4.10 and 4.11) that were synthesized from farnesyl acetate by Dr Juan Faraldos (86). (Z)-β-Farnesene was identified by comparing the mass spectrum with the Wiley database available on the mass spectrometer software. Another 5 sesquiterpene products (<2% in total) were also detected by GC-MS (Figure 4.9), but the MS spectra of these compounds could not match any standard provided by the Wiley database (Table 4.2). Nevertheless, these unidentified sesquiterpene structures are very likely to be cyclic products, since all the acyclic sesquiterpene products had been identified. The observed product profile is different to that published for partially purified EBFS, which yielded (E)-β-farnesene (97, 85%), (Z)-β-farnesene (112, 8%) and δ-cadinene (79, 5%), as well as three other minor cadinene-type products (less than 1% each) based on mass spectrometry (MS) (115).

(E)-β-farnesene (65)   (Z)-β-farnesene (97)   (3Z,6E)-α-farnesene (112)   (3E,6E)-α-farnesene (113)
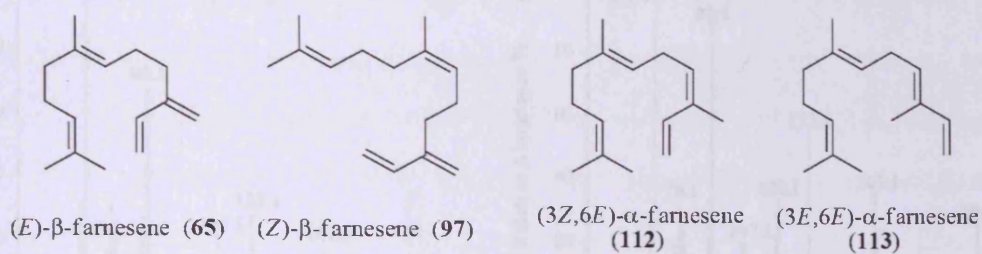
Figure 4.8: *Acyclic sesquiterpene products generated by WT-EBFS from substrate FDP.*
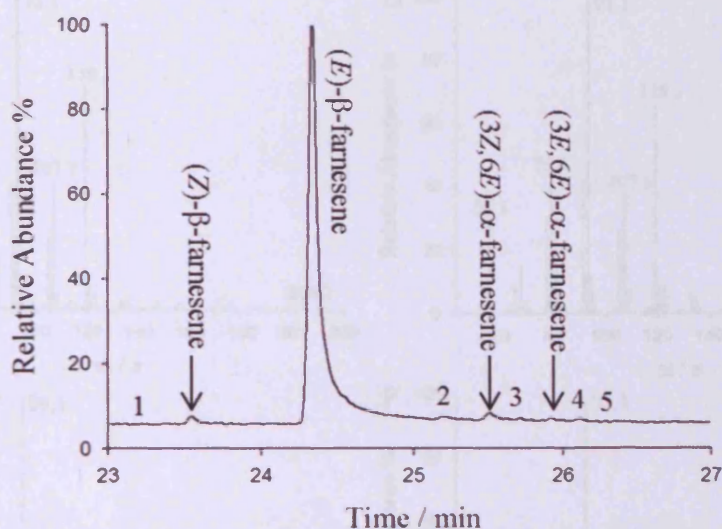
Figure 4.9: *GC trace for hexane-extractable products following incubation of 10 μM WT-EBFS with 1 mM FDP in the presence of 5 mM Mg^{2+}. Peaks 1 to 5: 5 unknown sesquiterpene products.*
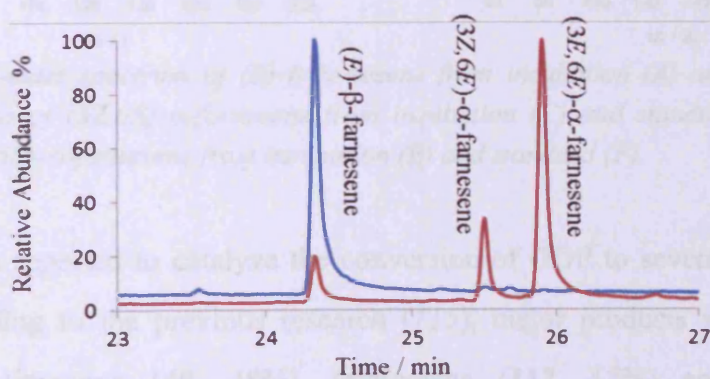
Figure 4.10: *Overlay of the chromatograms for wild-type EBFS incubation products (blue) with chemically synthetic (E)-β-farnesene, (3Z,6E)-α-farnesene and (3E,6E)-α-farnesene samples (red).*

138
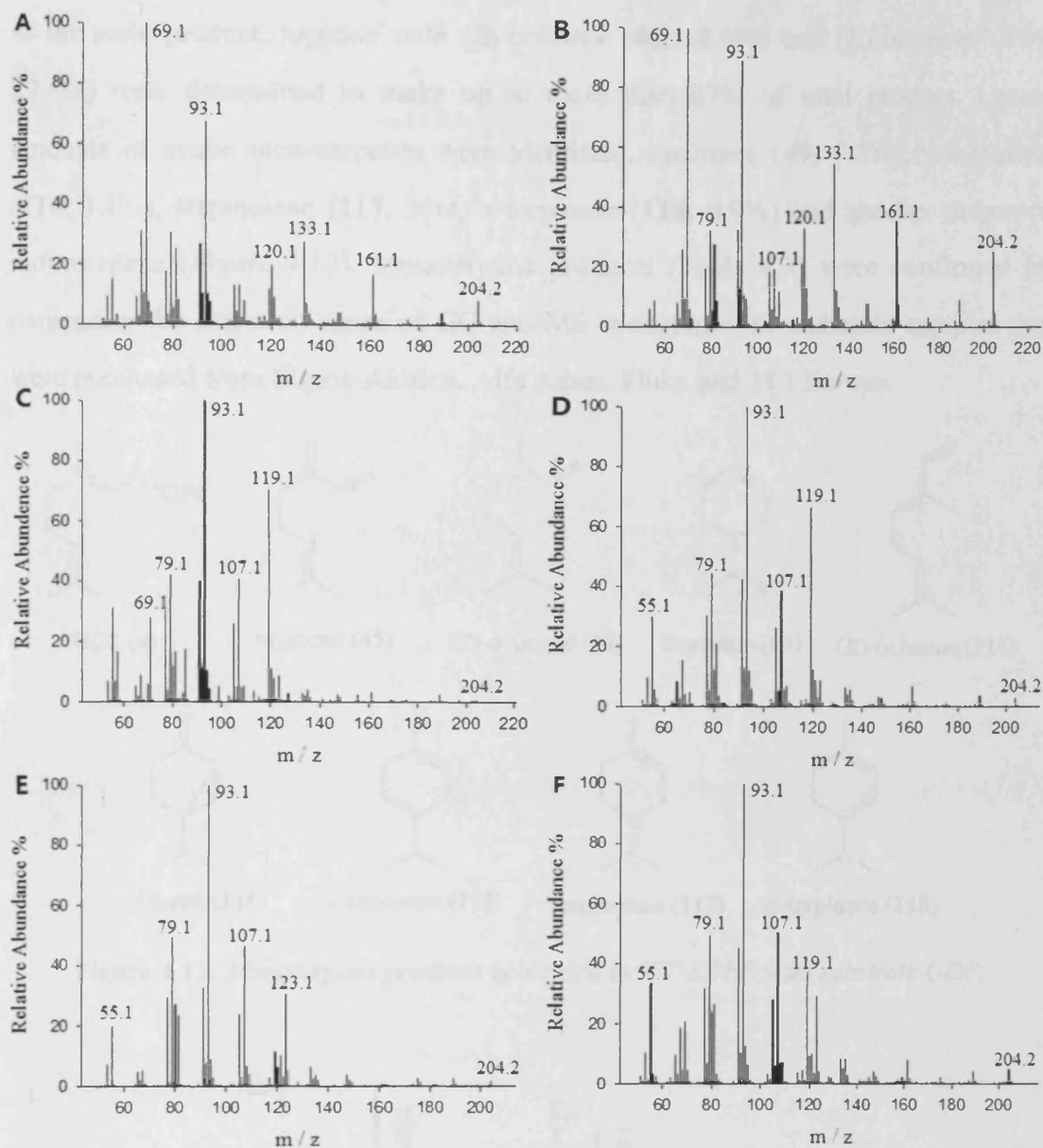
Figure 4.11: *EI-mass spectrum of (E)-β-farnesene from incubation (A) and standard (B). EI-mass spectrum of (3Z,6E)-α-farnesene from incubation (C) and standard (D). EI-mass spectrum of (3E,6E)-α-farnesene from incubation (E) and standard (F).*

EBFS has been reported to catalyze the conversion of GDP to several monoterpene olefins. According to the previous research (*115*), major products were the cyclic monoterpenes limonene (**49**, 48%), terpinolene (**117**, 15%) and the acyclic monoterpene myrcene (**45**, 15%). The other minor products were γ-terpinene (**116**, 7%), (Z)-ocimene (**46**, 6%), (E)-ocimene (**114**, 7%) and sabinene (**115**, 3%) (Figure 4.12). This observation is dramatically different from our result. As expected, mainly acyclic monoterpenes were confirmed as products (Figure 4.13). Myrcene (**45**, 38.3%),

139

as the main product, together with (Z)-ocimene (**46**, 18.9%) and (E)-ocimene (**114**, 29.4%) were determined to make up to more than 87% of total product. Lesser amounts of cyclic monoterpenes were identified, limonene (**49**, 7.2%), γ-terpinene (**116**, 1.1%), terpinolene (**117**, 3%), α-terpinene (**118**, 0.9%) and another unknown monoterpene (Figure 4.12). Monoterpene products (Table 4.3) were confirmed by comparing the retention times of GC and MS spectrum with authentic samples that were purchased from Sigma-Aldrich, Alfa Aesar, Fluka and TCI Europe.

GDP (4) myrcene (**45**) (Z)-ocimene (**46**) limonene (**49**) (E)-ocimene (**114**)

sabinene (**115**) γ-terpinene (**116**) terpinolene (**117**) α-terpinene (**118**)
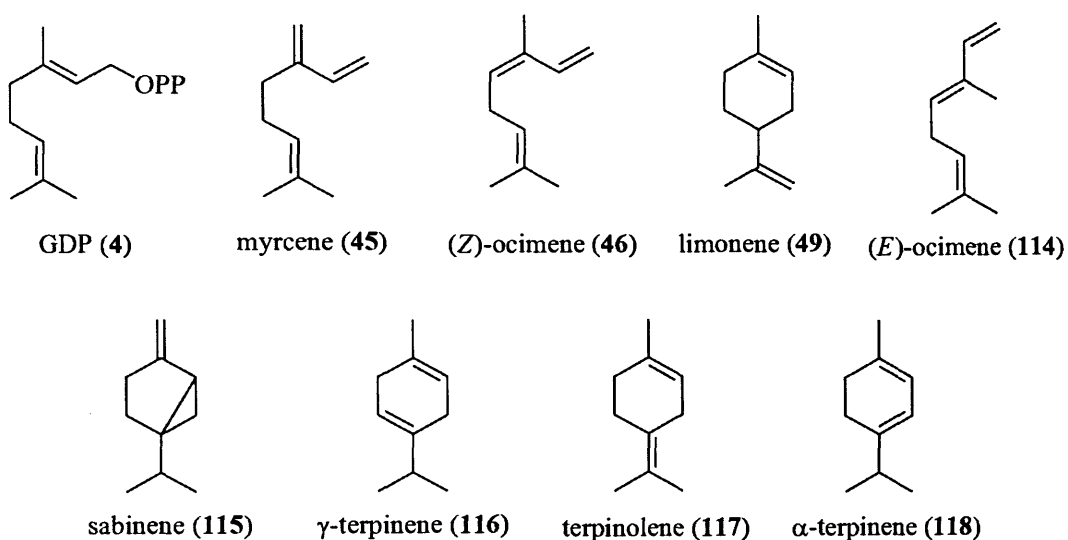
Figure 4.12: *Monoterpene products generated by WT-EBFS from substrate GDP.*
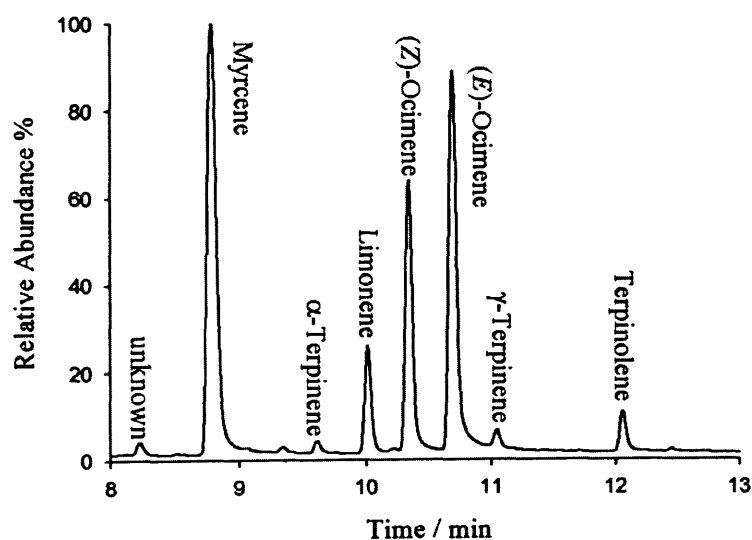
Figure 4.13: *GC trace for the hexane extractable products following incubation of 10 mM EBFS with 1 mM GDP.* All monoterpene products have been labeled.

Table 4.2: *GC-MS analysis of the sesquiterpene products from the incubation of wild-type EBFS with FDP.*

Retention time / min	Product	Percentage %
23.24	unknown-1	trace
23.54	(Z)-β-farnesene (97)	1.5
24.34	(E)-β-farnesene (65)	95
25.22	unknown-2	trace
25.50	(3Z,6E)-α-farnesene (112)	1.3
25.79	unknown-3	trace
25.91	(3E,6E)-α-farnesene (113)	0.2
26.10	unknown-4	trace
26.37	unknown-5	trace

Table 4.3: *GC-MS analysis of the sesquiterpene products from the incubation of wild-type EBFS with GDP.*

Retention time / min	Product	Percentage %
8.32	unknown	1.2
8.86	Myrcene (45)	38.3
9.69	α-terpinene (118)	0.9
10.08	limonene (49)	7.2
10.41	(Z)-ocimene (46)	18.9
10.76	(E)-ocimene (114)	29.4
11.12	γ-terpinene (116)	1.1
12.14	terpinolene (117)	3

4.1.5 Kinetic studies of WT-EBFS

All the kinetic assays were carried out using radiolabelled [1-^{3}H]FDP and MOPS buffer system (20 mM MOPS, 5 mM MgCl$_2$ and 5 mM β-mercaptoethanol at pH 7.2).

To determine the optimal incubation time, the reaction was carried out by incubating 0.06 µM wild-type enzyme with 10 µM [1-^{3}H]FDP for variable amounts of time (Figure 4.14A). A reaction time of 15 min was chosen for the actual kinetic runs since it is in the initial region of the graph and a relatively good radioactivity level was obtained in this time. The optimal enzyme concentration was determined by plotting variable enzyme concentrations against radioactivity level of product formation. As a result, the concentration of 0.06 µM was defined as the optimal concentration from the initial linear region (Figure 4.14B) and this result also confirmed that the time course was carried out with the optimised enzyme concentration conditions.
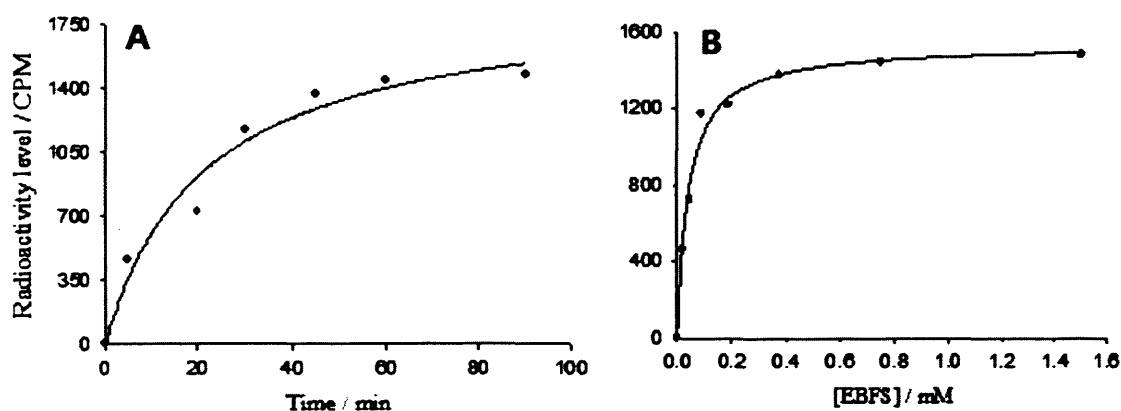
Figure 4.14: *Plots of radioactivity level in products formed by 0.06 µM WT-EBFS versus incubation time (A) and in 15 min versus concentration of WT-EBFS (B).*

Previous studies showed a slight substrate inhibition when the substrate concentration went over 10 µM (*115*). To make sure the inhibition was not caused by the cofactor (Mg^{2+}) or the inorganic diphosphate released during reaction (as the aqueous solution was covered by hexane during reaction, most of the organic products would go into the organic phase and diphosphate is therefore the only product that could accumulate in the aqueous phase), a set of kinetic assays were carried out in the presence of variable MgCl$_2$ and diphosphate concentrations (Figure 4.15). No inhibition was observed up to Mg^{2+} concentration of 200 mM or for diphosphate concentrations of up to 400 µM.
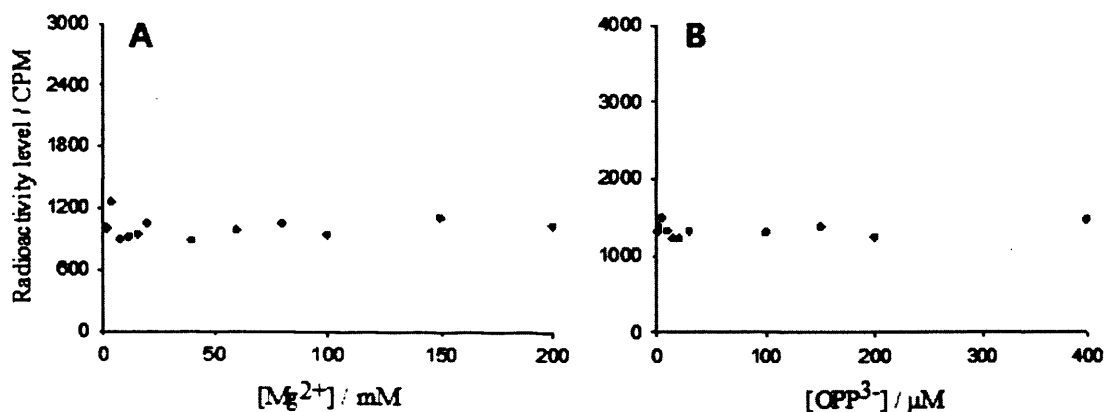
Figure 4.15: *Plots of radioactivity level in products formed in 15 min formed by 0.06 mM WT-EBFS and 10 mM [1-³H]FDP against Mg²⁺ (A) and inorganic diphosphate (B) concentration.*

Kinetic assays were subsequently carried out by incubating 0.06 μM wild-type EBFS with variable concentrations of radiolabelled FDP (from 0.1 to 120 μM) (Figure 4.16). K_M and k_{cat} values for each individual run were calculated using the programme SigmaPlot 10 by fitting the data to the Michaelis-Menten equation, $v = (v_{max} [S])/(K_M + [S])$.
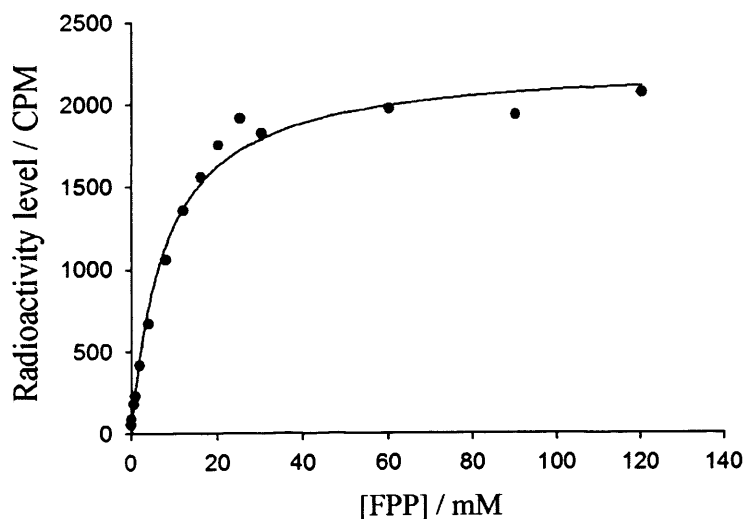
Figure 4.16: *Michaelis-Menten profile of WT-EBFS, incubation of 0.06 μM enzyme with different concentrations of radiolabelled FDP in the presence of 5 mM Mg²⁺.*

143

Based on five individual kinetic runs, the K_M value was calculated to be 6.4 ± 1.7 μM and the k_{cat} value 0.015 ± 0.004 s^{-1}. The K_M value was higher than the previously reported value of 0.6 μM (*115*) but very close to the value of 5 μM that was determined for the EBFS from maritime pine (*166*). The k_{cat} value (calculated based on the equation, $k_{cat} = v_{max}/(A \times T \times [E])$, (section 2.8.1.2)) obtained here is in the normal range for plant sesquiterpene synthases (*167*).

4.1.6 Investigation of the reaction mechanism using the substrate analogues 2F-FDP and 2F-GDP

Previous research into EBFS has proposed an enzymatic reaction mechanism *via* the ionization of *trans-trans*-farnesyl diphosphate and formation of (*E*)-β-farnesene (**65**) (Scheme 1.18) (*115*). 2F-FDP (**86**) should be a good inhibitor for producing both products because the fluorine at C$_2$ position can greatly inhibit the formation of allylic cation (Scheme 4.1).
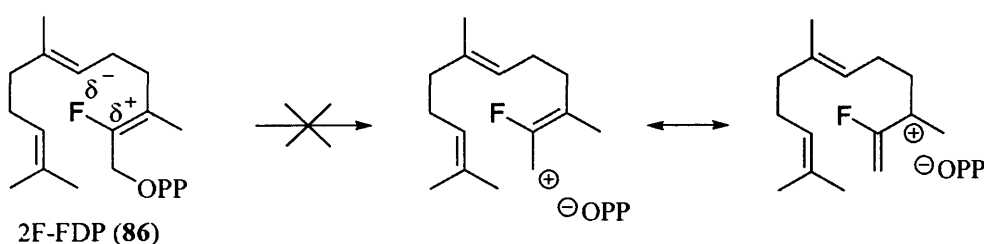
2F-FDP (**86**)

Scheme 4.1: *Inhibition step of forming 2F-transoid allylic cation.*

It has recently been shown that 2F-FDP (**86**) is an inhibitor of DCS and the crystal structure of the complex of DCS with 2F-FDP has been solved (*114*). In contrast, **86** is a very good substrate for EBFS. The product, (*E*)-β-2F-farnesene (**119**) (Figure 4.17), was confirmed by comparing retention time and MS spectrum with an authentic sample that was synthesized by Dr Juan Faraldos (*112, 135*). Under identical conditions (10 μM enzyme, 1 mM substrate, 5 mM Mg^{2+}), EBFS produces approximately twice as much (*E*)-β-farnesene as (*E*)-β-2F-farnesene based on the peak intensity in the GC-MS spectrum (Figure 4.17).

The proposed mechanism of forming both δ-cadinene and (E)-β-farnesene goes through the same intermediate, transoid allylic cation (54) (115). To form δ-cadinene, this cation is converted to NDP and then to the bicyclic product (114). The C2 fluorine can strongly inhibit the formation of transoid allylic cation through an inductive electron withdrawing effect. In this case, 1,6 ring closure can hardly happen and therefore, δ-cadinene is unlikely to be formed. The fact that EBFS converts 2F-FDP into (E)-β-2F-farnesene strongly suggests that the transoid allylic cation is not an intermediate in (E)-β-farnesene formation and therefore implies a concerted elimination mechanism rather than the two steps E1 elimination.
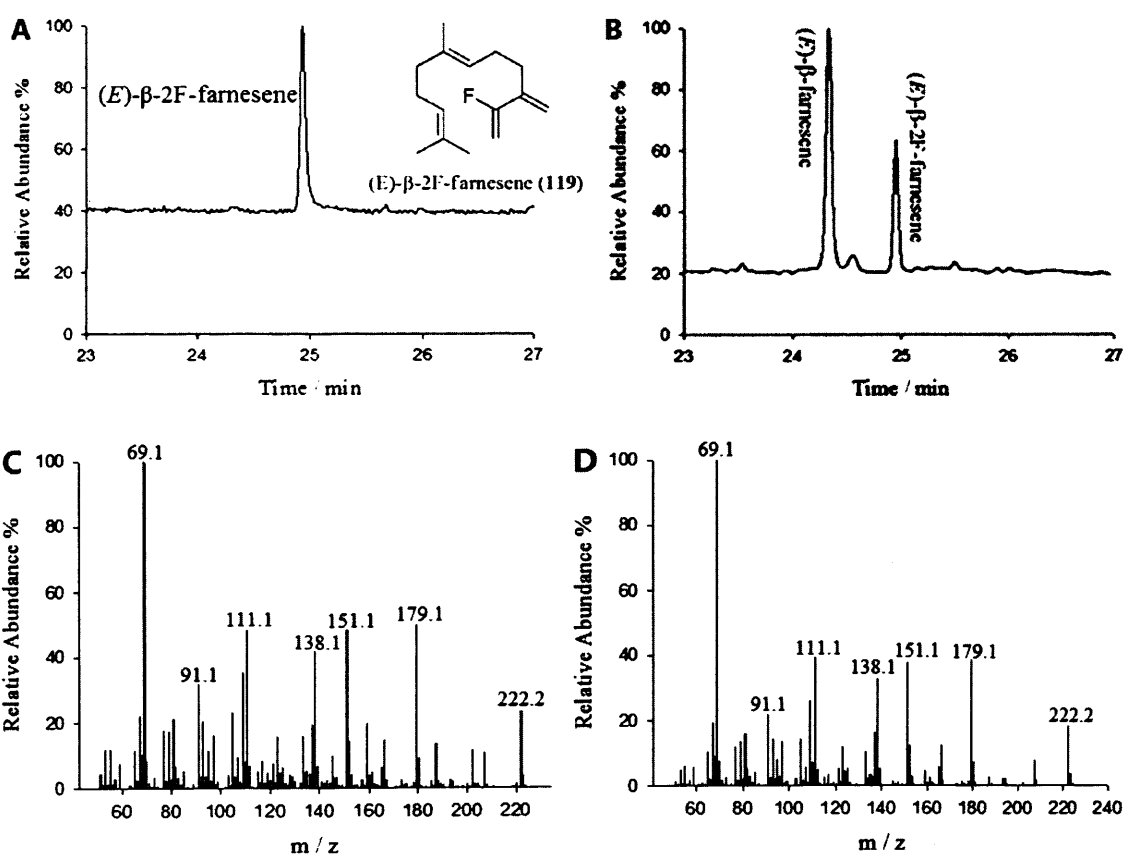
Figure 4.17: (A) GC trace for the hexane extractable products from incubation of 10 mM EBFS 1 mM 2F-FDP. (B) GC trace for the hexane extractable products from incubation of 10 mM EBFS with 1 mM FDP and 1 mM 2F-FDP. (C) EI-mass spectrum of (E)-β-2F-farnesene extracted from incubation. (D) EI-mass spectrum of (E)-β-2F-farnesene from authentic standard.

The yield of (E)-β-2F-farnesene is slightly lower than that of (E)-β-farnesene itself, probably because the fluorine at C2 position increases the transition state energy of diphosphate group release. The concerted elimination mechanism requires involvement of an active site base. These are usually histidine residues and the diphosphate group (57, 167). In the case of EBFS, there are no histidine residues around the active site pocket, suggesting that the diphosphate group would most likely be the active site base. Here, we proposed a syn-elimination mechanism for the enzymatic reaction catalysed by EBFS (Scheme 4.2). In agreement, the diphosphate group was also proposed as the catalytic base in the enzymatic reactions catalysed by isoprene synthase, farnesyl diphosphate synthase and aristolochene synthase (62, 72, 93), which provides another hint for the described enzymatic strategy of EBFS.
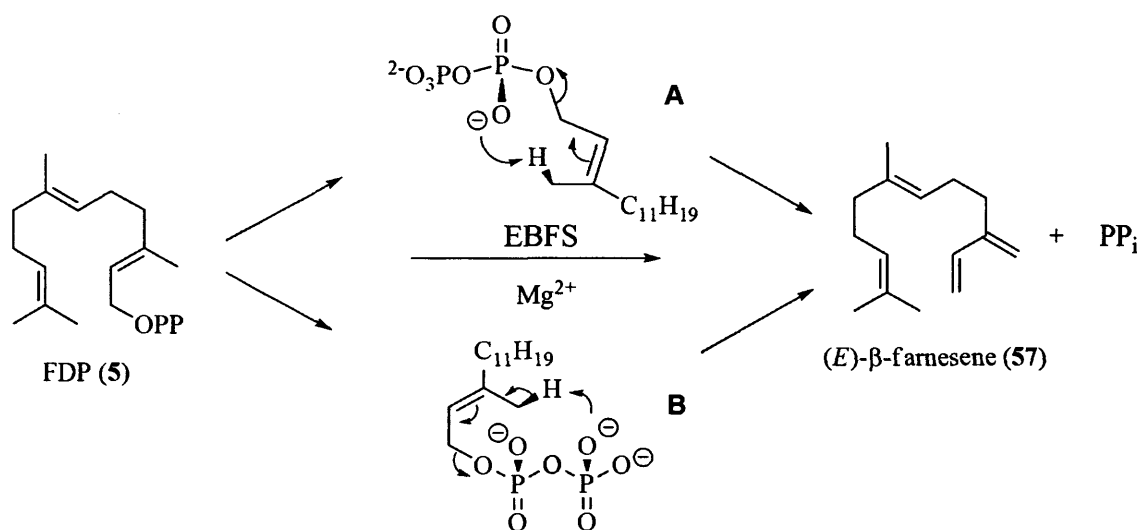
Scheme 4.2: *Proposed mechanism for the formation of (E)-β-farnesene from an intramolecular reaction of FDP via either pathway A or pathway B.*

Since the structures of those cyclic products have not been identified, it is hard to explain the details about the mechanism of enzymatic conversion of FDP to cyclic product by EBFS. Nevertheless, it suggests that the first ring closure of this catalytic mechanism goes through a concerted fashion as decribed for aristolochene synthase (81, 82)

.

4.2 Mutagenesis studies of (E)-β-farnesene synthase

4.2.1 Introduction

The deduced amino acid sequence of wild-type (E)-β-farnesene synthase has a significant level of similarity with other plant sesquiterpene synthases. It is most similar to epi-aristolochene synthase from tobacco (62% similarity and 49% identity) (168) and also closely resembles δ-cadinene synthase from tree cotton with 60% similarity and 37% identity (59) and the germacrene C synthase from tomato (57% similarity and 34% identity) (169).

Several liganded and unliganded crystal structures of sesquiterpene synthases from both fungi and plants have been solved, revealing a conformational change triggered by Mg^{2+}_3-FDP complex binding to the enzymes (81, 90, 93, 170). Several loop regions participate in this movement upon ligand binding, forming a sequestered active site cavity with the proper conformation for terpenoid production. It has also been suggested that the active site cavity could enforce the substrate to dock into the enzyme with a product-like conformation for high fidelity sesquiterpene synthases, whereas the promiscuous synthases contain a larger active site pocket that allows different substrate binding conformations, leading to mutiple reaction products (1). Based on the above discussion, it is reasonable to assume that altering the active site volume would lead to a different product outcome.

A homology model for EBFS was created based on the crystal structure of δ-cadinene synthase using the SWISS-MODEL automated homology modelling server (Figure 4.18). Sequence alignment with other plant terpene synthases was carried out with ClustalX using default parameters (162). Residues in the first tier of the active site (i.e. within van der Waals radii (about 3.5 Å) of the substrate (160)) have been listed in Table 4.4.

147
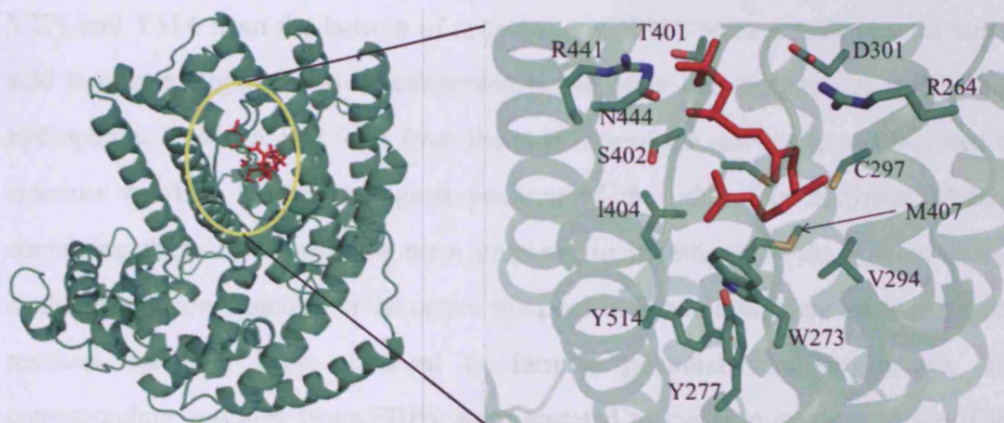
Figure 4.18: *Cartoon representations of the EBFS homology model and the expansion of the active site.* FDP (red sticks) was docked into the active site by the program FlexX.

Table 4.4: *Comparison of wild-type EBFS active site contact residues with other plant sesquiterpene synthases.* GAS: germacrene A synthase from *Crepidiastrum sonchifolium* (ABB00361.1); GCS: germacrene C synthase from *Solanum lycopersicum* (AAC39432.1); TEAS: 5-*epi*-aristolochene synthase from *Nicotiana tabacum* (AAA19216.1); DCS: δ-cadinene synthase from *Gossypium arboretum* (AAA93064.1); EBFS: (*E*)-β-farnesene synthase from *Mentha* x *piperita*, L. (AAB95209.1).

Enzyme	Active site contact residues							
GAS	R275	W284	V288	T305	L308	V309	D312	T412
GCS	R263	W272	V276	L293	T296	S297	D300	S401
TEAS	R264	W273	V277	I294	I297	S298	D301	T401
DCS	R270	W279	V283	I300	A303	S304	D307	T407
EBFS	R264	W273	Y277	V294	C297	G298	D301	T401

Enzyme	Active site contact residues						
GAS	S413	A414	Y415	I418	R453	D456	L524
GCS	A402	G403	Y404	I407	R442	N445	L513
TEAS	T402	T403	Y404	L407	R441	D444	L512
DCS	C408	G409	Y410	L413	R448	D451	L519
EBFS	S402	C403	I404	M407	R441	N444	Y514

Y277 and Y514 from the bottom of active site of EBFS were revealed by the amino acid sequence alignment to be conserved in the other four enzymes as smaller size hydrophobic residues; and I404 from the side face of the active site was replaced by aromatic residues in the analogous position of the other four enzymes. Mutants containing these three residues were prepared to investigate if the volume change could bring a new function to the active site pocket. In addition, there are two cysteine residues that had been observed to form a potential disulfide linkage. The corresponding residues from EBFS were mutated to cysteine residues to see if the changes could affect the product outcome. Finally, a hybrid was constructed by swapping all the active site residues of EBFS with the corresponding ones of DCS.

4.2.2 EBFS-Y514L

Y514 of EBFS that sits at the bottom of the active site pocket has been found to be a conserved Leu in the four other sesquiterpene synthases (Figure 4.19). The reason EBFS forms mainly acyclic products may be that the distal isoprene unit is mispositioned in the active site due to steric hindrance, attenuating the efficiency of cyclisation. The mutant EBFS-Y514L could release some free space around the bottom of the active site, which might allow the substrate to fold into a different conformation.
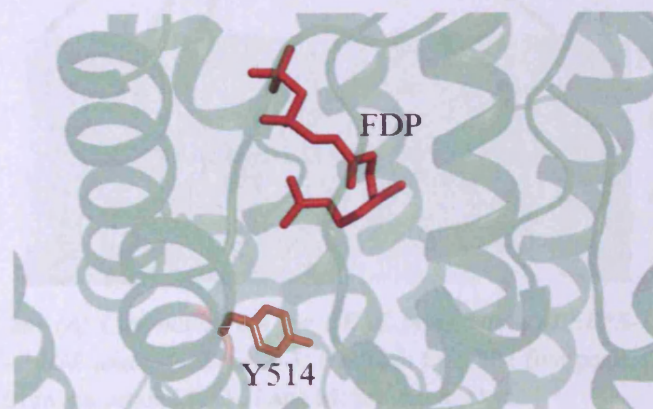
Figure 4.19: *Cartoon representation of the active site of the EBFS homology model.* FDP (shown as red sticks) was docked into the active site by the program Flex X. Residue Y514 was shown as orange sticks.

The mutant was made by substitution of the native TAC codon of Tyr with CTG of Leu by site-directed mutagenesis. EBFS-Y514L was purified in the same way as WT-EBFS (Figure 4.20). Typically, 15 mg of EBFS-Y514L was obtained per litre of culture. 10 μM of the purified mutant was incubated with 1 mM FDP overnight and the hexane extractable products were analysed by GC-MS. The product outcome was not significantly affected (Figure 4.21). The mutant generates (E)-β-farnesene as the main product (95%) and another four minor products, (Z)-β-farnesene (0.7%), (3Z,6E)-α-farnesene (1%), (3E,6E)-α-farnesene (1.6%) (structures are confirmed by comparing GC-MS results of wild-type EBFS) and an unknown sesquiterpene with the retention time of 25.22 min (1.7%).
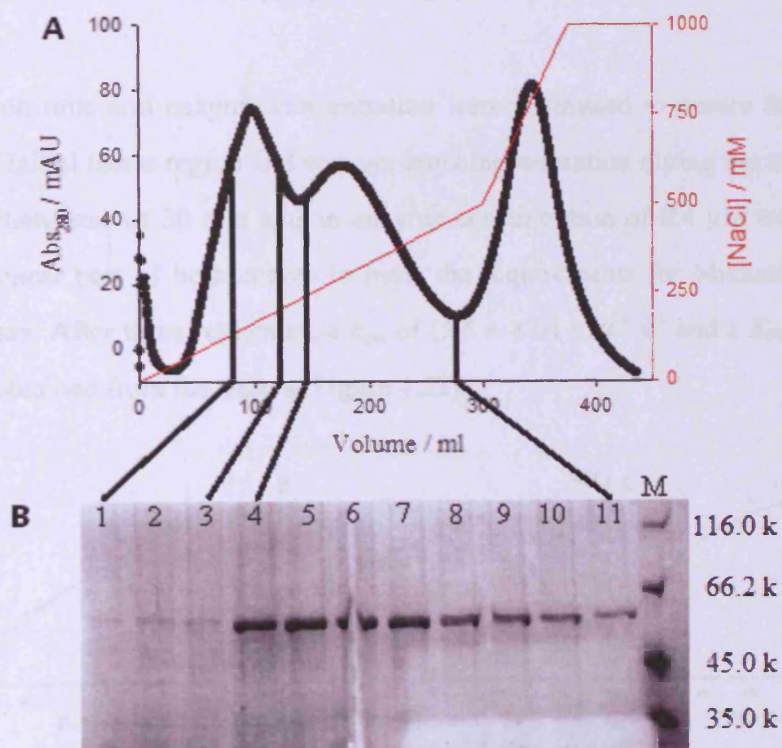
Figure 4.20: *(A) Chromatogram for DEAE purification of EBFS-Y514L. (B) 13% SDS-PAGE analysis.* Lane 1-3: fractions from the first peak. Lane 4-12: fractions from the second peak. Lane M: protein marker.

150
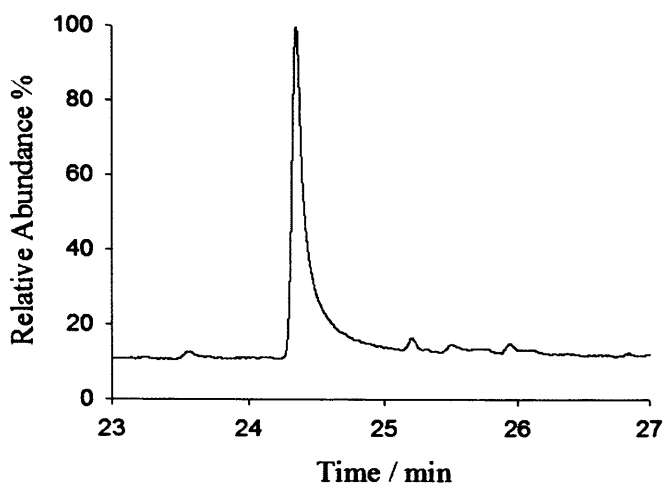
Figure 4.21: *GC trace for incubation of 10 μM EBFS-Y514L with 1 mM FDP in the presence of 5 mM Mg²⁺.*

The reaction time and enzyme concentration were optimised to ensure the reaction was in the initial linear region and was not reaching saturation during the experiment. An incubation time of 30 min and an enzyme concentration of 0.4 μM were chosen from the linear part of both curves to meet the requirements for Michaelis-Menten kinetic assay. After three replicates, a k_{cat} of $(5.5 \pm 1.9) \times 10^{-4}$ s^{-1} and a K_M of 23 ± 8 μM were obtained from the assays (Figure 4.22).
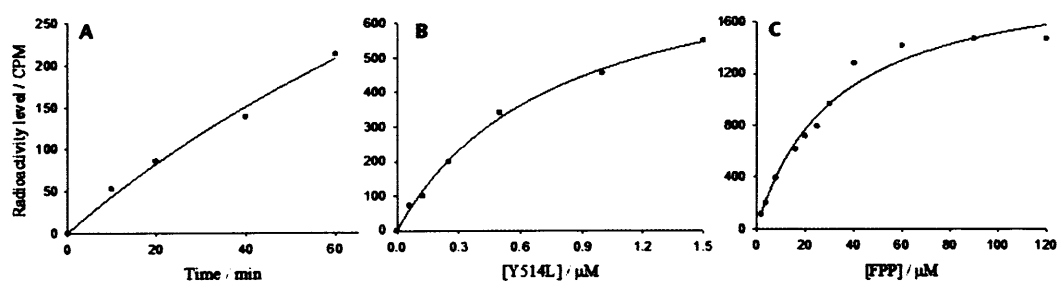
Figure 4.22: *Kinetic analysis of EBFS-Y514L.* (A) Time course (0.125 μM enzyme, 10 μM FDP); (B) Plot of velocity against enzyme concentration (10 μM FDP, 30 min incubation time); (C) Michaelis-Menten plot (incubation of 0.4 μM enzyme with radiolabelled substrate for 30 min at 22 °C).

Although the product outcome was not affected, the k_{cat} was reduced 27-fold and the

K_M was 4 times greater than that of the wild-type enzyme, suggesting the volume increase does affect the substrate binding, but not in the way of generating any other new structures. In contrast, this change may compromise the orbital alignment between the electron lone pair of nucleophile and empty s-obital of the proton by increasing the mobility of the farnesyl chain in the active site.

4.2.3 EBFS-I404Y

I404 of EBFS is in helix G and the side chain of I404 points to helice H in close proximity (Figure 4.23). The corresponding residue was found to be a conserved Tyr in the other four sesquiterpene synthases. Different residues at this position may either affect the conformation of helices G and H, or directly influence the plasticity of active site surface.
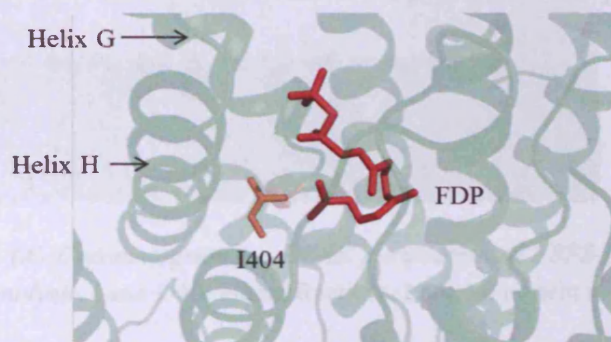
Figure 4.23: *Cartoon representation for the active site of EBFS homology model.* FDP (shown as red stick) was docked into the active site by the program Flex X. Residue I404 was shown as orange sticks.

The gene of mutant EBFS-I404Y was prepared by replacing the native Ile codon ATT with a Tyr codon of TAT using site-directed mutagenesis. The mutant was purified in the same manner as the wild-type EBFS (Figure 4.24). Typically, 20 mg of EBFS-I404Y was obtained per litre of culture. The *in vitro* assay was carried out by incubating 10 μM of purified mutant protein with 1 mM of FDP in a total volume of 500 μl reaction solution at room temperature overnight and the hexane extractable products were analysed by GC-MS. Rather than producing another 8 minor products (as seen for the wild type), in addition to the predominant (*E*)-β-farnesene only a

152

small amount of the unknown sesquiterpene at the retention time of 25.22 min was detected (Figure 4.25).
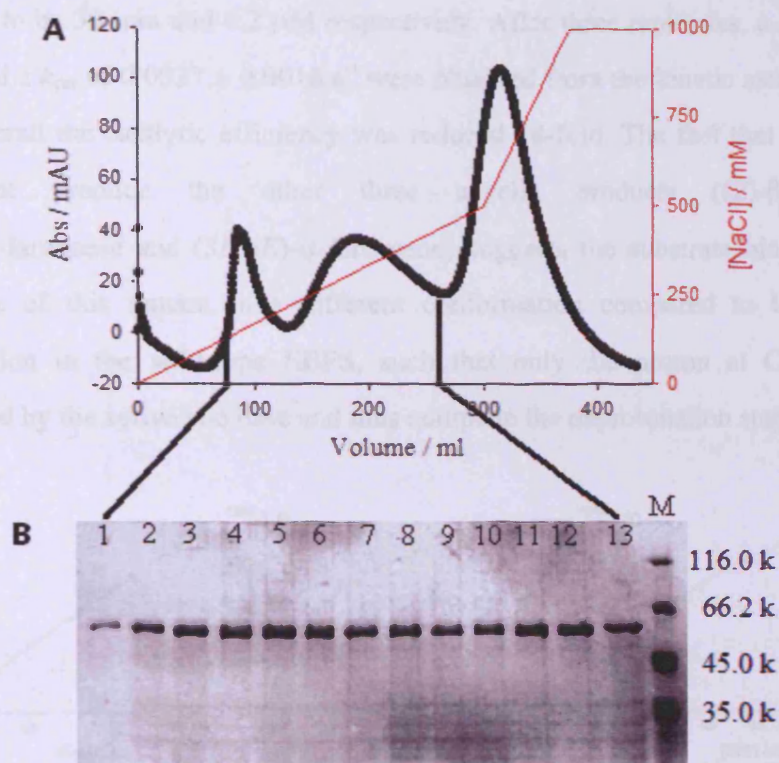
Figure 4.24: *(A) Chromatogram for DEAE purification of EBFS-1404Y. (B) 13% SDS-PAGE analysis.* Lane 1-13: eluent fractions. Lane M: protein marker.
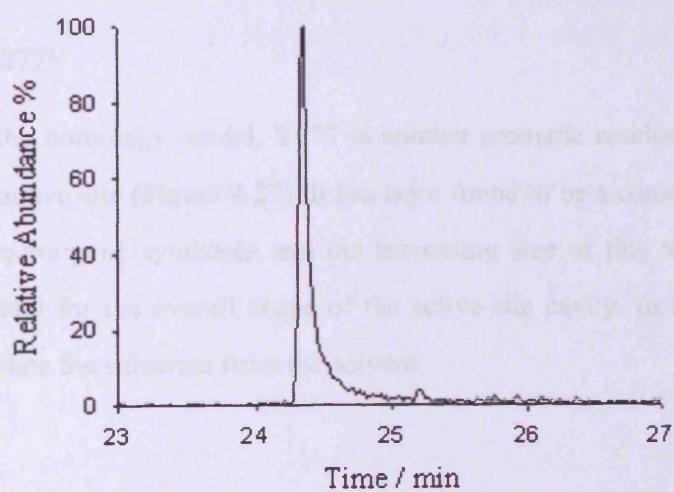
Figure 4.25: *GC trace for incubation of 10 μM EBFS-1404Y with 1 mM FDP in the presence of 5 mM Mg^{2+}.*

153

EBFS-I404Y was also found to be a less active enzyme compared to the wild-type EBFS. The incubation time and the enzyme concentration for the kinetic assay were optimised to be 30 min and 0.2 μM respectively. After three replicates, a K_M of 29 ± 12 μM and a k_{cat} of 0.0037 ± 0.0016 s^{-1} were obtained from the kinetic assays (Figure 4.26). Overall the catalytic efficiency was reduced 18-fold. The fact that the mutant could not produce the other three acyclic products ((Z)-β-farnesene, (3Z,6E)-α-farnesene and (3E,6E)-α-farnesene) suggests the substrate binds into the active site of this mutant in a different conformation compared to the binding conformation in the wild-type EBFS, such that only the proton at C15 can be approached by the active site base and thus complete the deprotonation step.
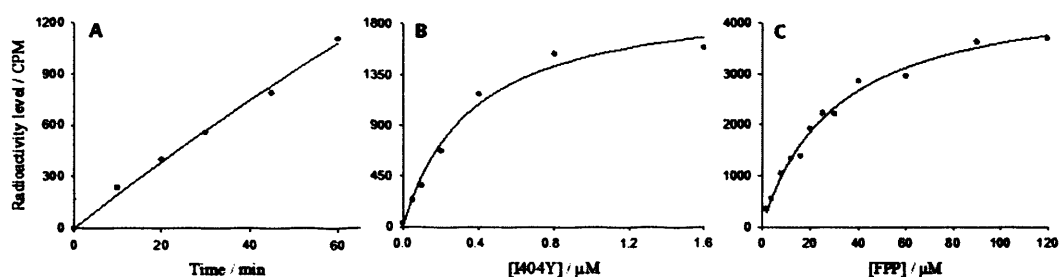
Figure 4.26: *Kinetic analysis of EBFS-I404Y.* (A) Time course (0.2 μM enzyme, 10 μM FDP); (B) Plot of velocity against enzyme concentration (10 μM FDP, 30 min incubation time); (C) Michaelis-Menten plot (incubation of 0.2 μM enzyme with radiolabelled substrate for 30 min at 22 °C).

4.2.4 EBFS-Y277V

According to the homology model, Y277 is another aromatic residue that sits in the bottom of the active site (Figure 4.27). It has been found to be a conserved Val in the four other sesquiterpene synthases and the increasing size of this residue in EBFS may be important for the overall shape of the active site cavity. In addition, it may also help to isolate the substrate from the solvent.
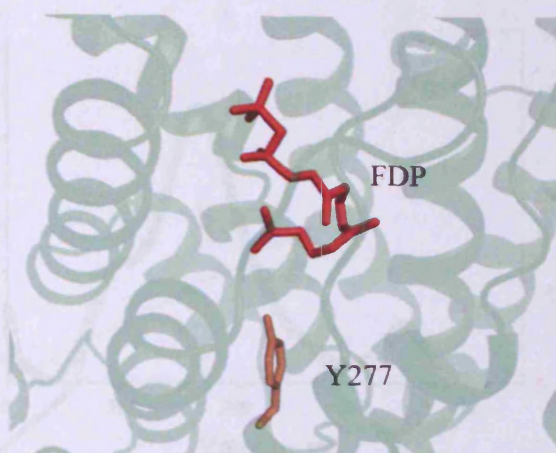
Figure 4.27: *Cartoon representation for the active site of EBFS homology model.* FDP (shown as red stick) was docked into the active site by the program Flex X. Residue Y277 was shown as orange stick.

The corresponding residues from the other four sesquiterpene synthases are found to be a conserved Val. Thus, the mutant EBFS-Y277V was designed to check if the modified active site cavity would affect the product selectivity and catalytic activity. The native Tyr codon TAC was substituted by a Val codon GTG using site-directed mutagenesis. The mutant was produced and purified by the same method for the wild-type enzyme (Figure 4.28). Typically, 13 mg of EBFS-Y277V was obtained per litre of culture. GC-MS analysis following overnight incubation of EBFS-Y277V (10 μM) with FDP (1 mM) showed an unchanged product profile from the wild-type enzyme (Figure 4.29).
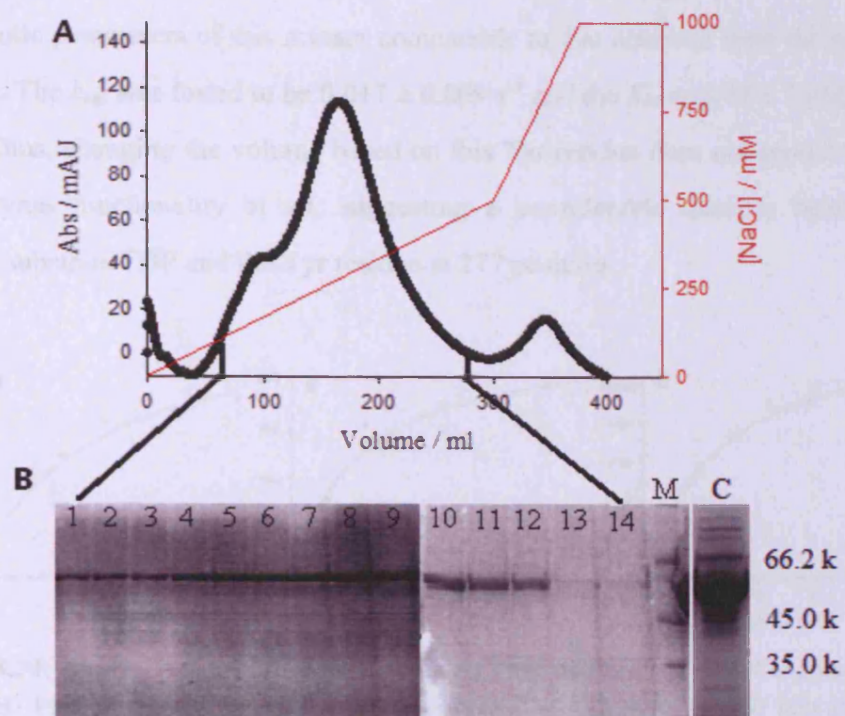
Figure 4.28: *(A) Chromatogram for DEAE purification of EBFS-Y277V. (B) 13% SDS-PAGE analysis.* Lane 1-14: eluent fractions. Lane M: protein marker. Lane C: concentrated protein.
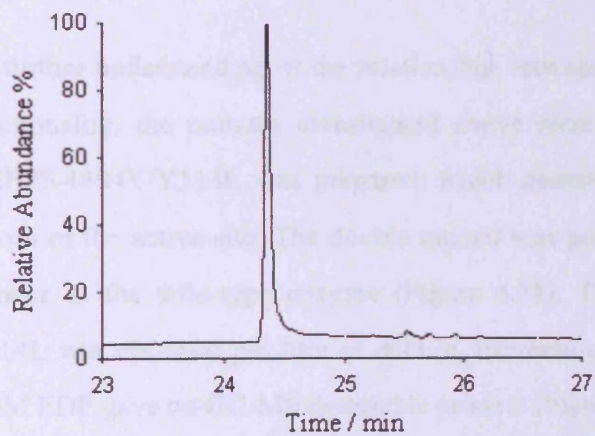
Figure 4.29: *GC trace for incubation of 10 µM EBFS-Y277V with 1 mM FDP in the presence of 5 mM Mg^{2+}.*

The kinetic parameters of this mutant comparable to that obtained from the wild-type enzyme. The k_{cat} was found to be 0.017 ± 0.009 s^{-1} and the K_M was 18 ± 7 μM (Figure 4.30). Thus, changing the volume based on this Tyr residue does not appear to affect the enzyme functionality at all, suggesting a considerable distance between the binding substrate FDP and the Tyr residue at 277 position.
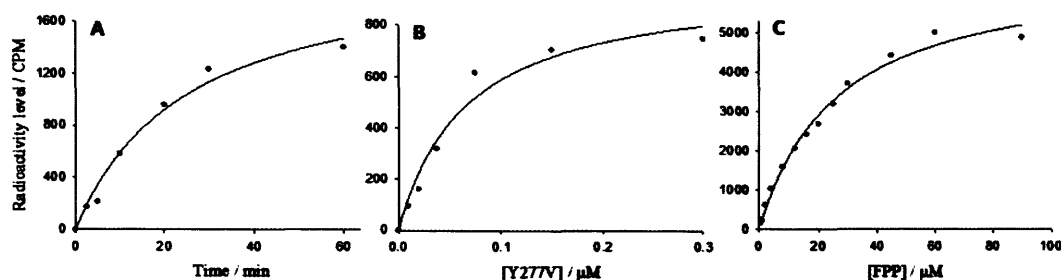
Figure 4.30: *Kinetic analysis of EBFS-Y277V.* (A) Time course (0.0375 μM enzyme, 10 μM FDP); (B) Plot of velocity against enzyme concentration (10 μM FDP, 20 min incubation time); (C) Michaelis-Menten plot (incubation of 0.375 μM enzyme with radiolabelled substrate for 20 min at 22 °C).

4.2.5 EBFS-I404Y/Y514L

In order to get a further understanding of the relationship between active site volume and enzyme functionality, the mutants investigated above were combined. First, a double mutant EBFS-I404Y/Y514L was prepared, which contains modifications to the side and bottom of the active site. The double mutant was produced and purified in the same manner as the wild-type enzyme (Figure 4.31). Typically, 20 mg of EBFS-I404Y/Y514L was obtained per litre of culture. Incubation of 10 μM purified enzyme with 1 mM FDP gave no GC-MS detectable product (Figure 4.32).

The observed inactivation of the double mutant indicates that the closure of the active site is obstructive. The modified active site may or may not be able to complex with FDP substrate, as FDP is docked into the active site with an conformation that is highly unfavorable, if any, reaction to take place. Given that the active site was designed based on EBFS-I404Y/Y514L, it may not be the active conformation. An enzyme capable of oxidized bound to FDP with canonical active site can expectedly be constructed with site-specific mutations.

4.2.6 EBFS-F277I/I404Y/Y514L

The substrate promiscuity of EBFS was further examined with the triple mutant EBFS-I404Y/Y514L, which showed the highest activity among all the mutant (Figure 4.27). The purification gave a single protein band on SDS-PAGE, as wild-type enzyme. Incubation of 10 μM EBFS-F277I/I404Y/Y514L with 1 mM was...

Figure 4.31: *(A) Chromatogram for DEAE purification of EBFS-I404Y/Y514L. (B) 13% SDS-PAGE analysis.* Lane 1-11: eluent fractions. Lane M: protein marker.

Figure 4.32: *GC trace for incubation of 10 μM EBFS-I404Y/Y514L with 1 mM FDP in the presence of 5 mM Mg²⁺.*

158

The observed activity loss of the double mutant indicates a significant change of the active site conformation. The modified active site cavity may not be able to complex with FDP anymore, or FDP is docked into the active site with a conformation that is highly unfavored for any reaction to take place. Another three triple mutants were designed based on EBFS-I404Y/Y514L, trying to restore the catalytic activity of this enzyme against FDP. These mutants with additional alteration were also expected to be constructed with different functions.

4.2.6 EBFS-Y277V/I404Y/Y514L

The substitution Y277V was first introduced into the double mutant EBFS-I404Y/Y514L to confer the active site with more free space at the bottom (Figure 4.27). This triple mutant was overexpressed and purified by the same method as wild-type enzyme (Figure 4.33). Typically, 17 mg of EBFS-Y277V/I404Y/Y514L was obtained per litre of culture.

Figure 4.33: *(A) Chromatogram for DEAE purification of EBFS-Y277V/I404Y/Y514L. (B) 13% SDS-PAGE analysis.* Lane 1-12: eluent fractions. Lane M: protein marker. Lane C: concentrated protein.

159

Although the mutation Y277V did not change either product selectivity or catalytic activity compared to wild-type EBFS, it restored the activity of the double mutant EBFS-I404Y/Y514L. In addition, the triple mutant showed an obviously different product distribution. (E)-β-Farnesene is still the main product but it only makes up 83% of the total sesquiterpenes and an increasing amount of $(3Z,6E)$-α-farnesene (2.5%) and $(3E,6E)$-α-farnesene (11.5%) was observed by GC-MS (Figure 4.34). Besides, a newly formed peak at the retention time of 23.95 min was identified. Comparing the mass fragmentation pattern of this compound with data from the Wiley database available on the mass spectrometer software suggests a cyclic sesquiterpene structure. However, the actual structure of this compound cannot be assigned solely by it mass spectrum because the mass spectra of cyclic sesquiterpenes are very similar. An activity test was carried out by incubation of 0.5 μM purified mutant with 60 μM radiolabelled FDP for 30 min and the activity of enzyme reaction was roughly at the same level as the enzyme-free control. Thus, the catalytic activity was defined as too low to be measured reliably by the kinetic assay using radiolabelled substrate.
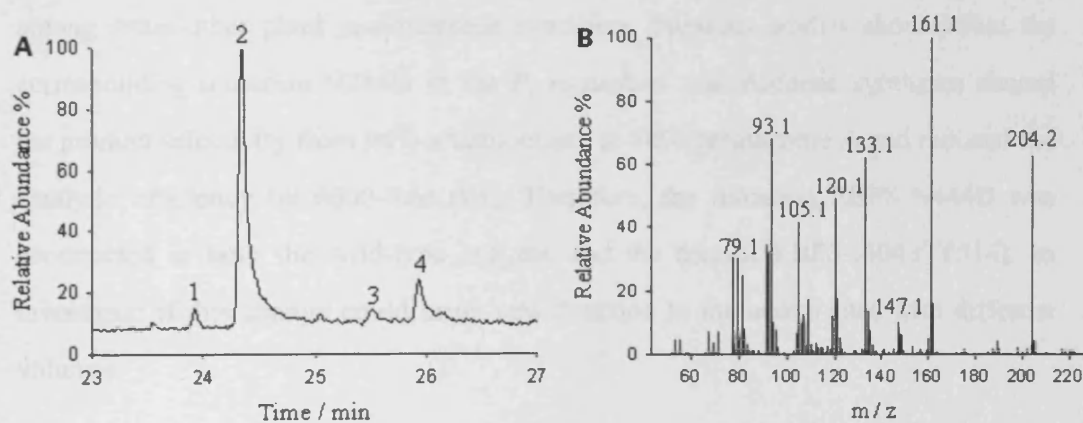
Figure 4.34: *(A) Sesquiterpene product profile of EBFS-Y277V/I404Y/Y514L.* Peak 1: unknown sesquiterpene product. Peak 2: (E)-β-farnesene. Peak 3: $(3Z,6E)$-α-farnesene. Peak 4: $(3E,6E)$-α-farnesene. *(B) EI-MS of the peak 1.*

In summary, all the mutants studied with the aim of manipulating the active site volume suggest that the bound farnesyl chain is close to the side formed by helices H and G since the mutations (I404Y and Y514L) on this side significantly reduce the

enzyme activity and the double mutant even lost catalytic activity, possibly due to impaired substrate binding. After releasing some more space on the other side of active site by the mutation Y277V in the triple mutant, the substrate may have been forced to bind more towards the opposite side of helices H and G with a different conformation, allowing a new cyclic product to form. However, the original active site contour is precisely defined for the optimised reaction conditions, thus the modification leads to a huge reduction of enzyme activity.

4.2.7 N444D

The work so far was focussed on manipulating the space inside the active site pocket. On the other hand, the conformation of the substrate binding area is also crucial for the product selectivity and activity and mutations in the Mg^{2+} binding motif could change both the enzyme activity and product profile (*95, 170-172*). According to the homology model and the sequences alignment, N444 was identified as part of the second Mg^{2+} binding motif N^{444}DTSSQLRE (equivalent position to other plant sesquiterpene synthases) and it was found to be an Asp in the corresponding position among some other plant sesquiterpene synthases. Previous studies showed that the corresponding mutation N244D in the *P. roqueforti* aristolochene synthases altered the product selectivity from 94% aristolochene to 80% germacrene A and reduced the catalytic efficiency by 6600-fold (*95*). Therefore, the mutation EBFS-N444D was constructed in both the wild-type enzyme and the mutant EBFS-I404Y/Y514L to investigate if this change could bring new function to the active sites with different volumes.

Both mutants were purified by the same method as wild-type EBFS (Figure 4.35). Typically, 20 mg of EBFS-N444D and 10 mg of EBFS-I404Y/N444D/Y514L was obtained per litre of culture. Incubation of EBFS-N444D with FDP gave no GC-MS detectable sesquiterpene product (Figure 4.36). The observed loss of wild-type enzyme activity resulting from this mutation in the Mg^{2+} binding motif is probably due to disturbance of the precise binding conformation of the substrate by

161

compromising the enzyme:Mg^{2+}:FDP tertiary complex. It also suggests a precise active site template, which is only defined for certain substrate binding conformation. EBFS-I404Y/N444D/Y514L showed very weak enzyme activity; only three small peaks (barely higher than the noise level) representing (E)-β-farnesene, (3Z,6E)-α-farnesene and (3E,6E)-α-farnesene were identified by GC-MS analysis after a typical overnight incubation (Figure 4.36). Although the misfolded substrate in the modified active site of the triple mutant did not lead to a total loss of activity, the catalytic efficiency was lowered to a negligible level.
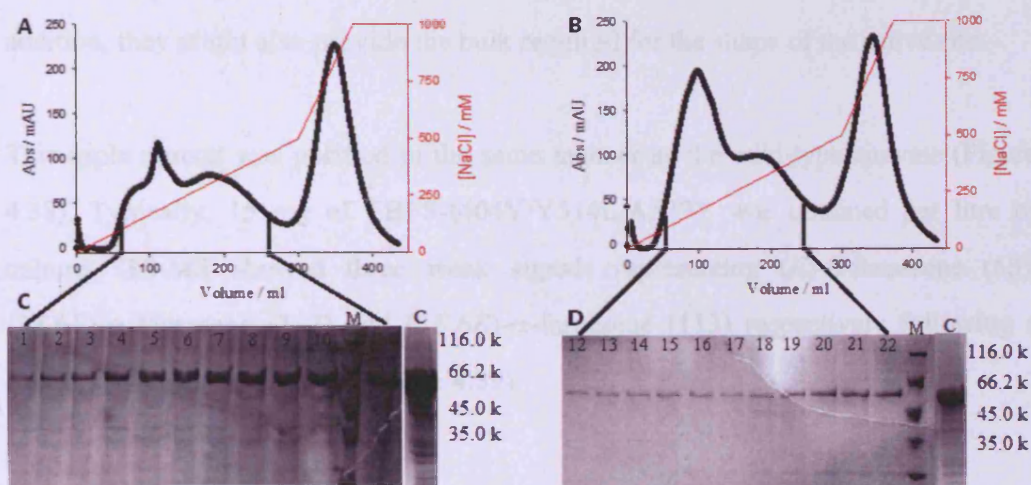
Figure 4.35: *Chromatograms for DEAE purification of EBFS-N444D (A) and EBFS-I404Y/N444D/Y514L (B) and 13% SDS-PAGE analyses of each purification (C) and (D). Lane 1-11: eluent fractions for EBFS-N444D. Lane 12-22: eluent fractions for EBFS-I404Y/N444D/Y514L. Lane M: protein marker. Lane C: concentrated protein.*
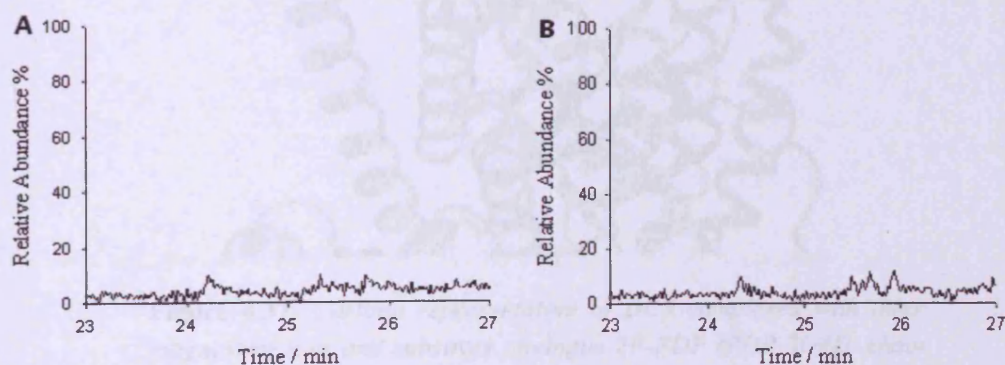
Figure 4.36: *Product profiles for incubation of FDP with EBFS- N444D (A) and EBFS-I404Y/N444D/Y514L (B).*

4.2.8 EBFS-I404Y/Y514L/A529Y

Another substitution that was introduced into the double mutant EBFS-I404Y/Y514L was A529Y. The residue A529 belongs to the J-K loop, which is located near the entrance of the active site and has been shown to be involved in the movement upon ligand binding, capping the active site pocket during catalysis (*1*). This residue was proposed to be close to the Mg^{2+}-PPi complex. The corresponding residues to A529 are either Tyr or Phe among the other four sesquiterpene synthases (Figure 4.37). The aromatic group could provide negative or positive electrostatic potential to the local area and this might be related to adjusting the orientation of the Mg^{2+} cluster. In addition, they might also provide the bulk required for the shape of the active site.

This triple mutant was purified in the same manner as the wild-type enzyme (Figure 4.38). Typically, 15 mg of EBFS-I404Y/Y514L/A529Y was obtained per litre of culture. GC-MS showed three weak signals representing (*E*)-β-farnesene (**65**), (3*Z*,6*E*)-α-farnesene (**112**) and (3*E*,6*E*)-α-farnesene (**113**) respectively following a typical overnight incubation (Figure 4.39).
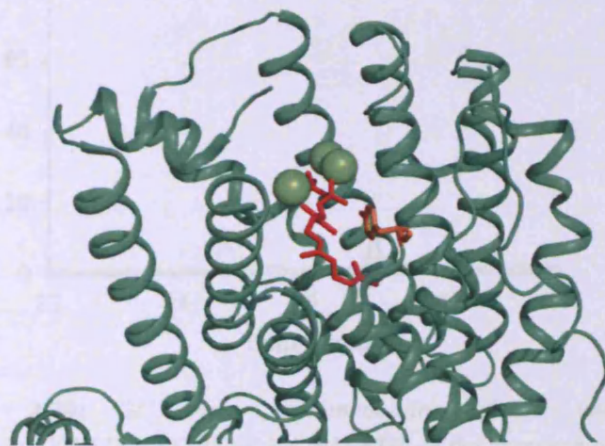
Figure 4.37: *Cartoon representation of DCS complexed with three magnesium ions and substrate analogue 2F-FDP (PDB 3G4D, chain A).* Magnesium ions are shown as green spheres and bound 2F-FDP is shown as red sticks. Residue Y535 is shown as orange sticks.
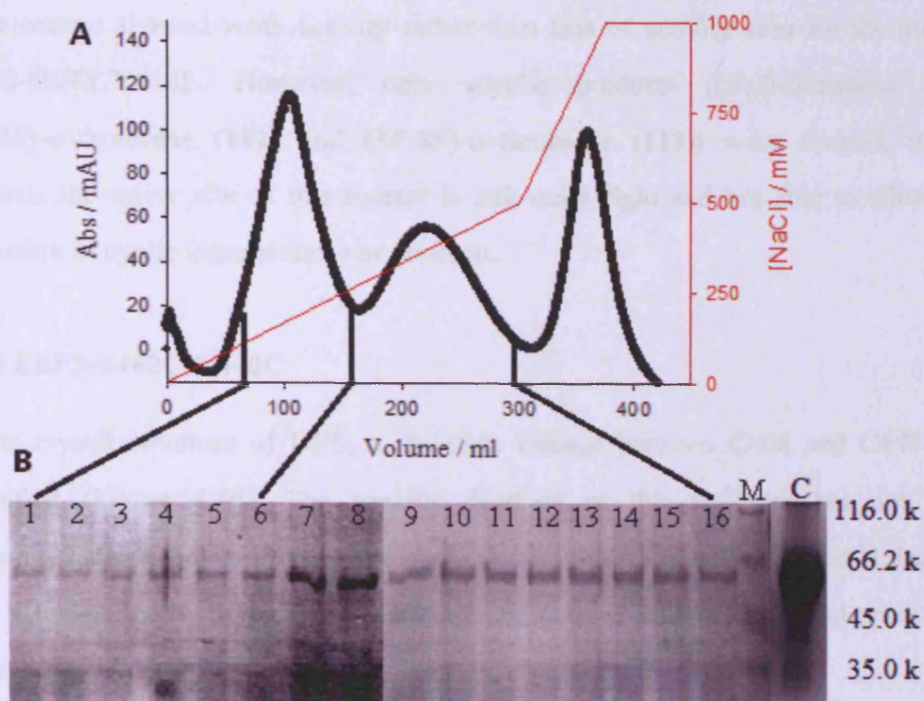
Figure 4.38: *(A) Chromatogram for DEAE purification of EBFS-I404Y/Y514L/A529Y. (B) 13% SDS-PAGE analysis.* Lane 1-6: fractions from the first peak. Lane 7-16: fractions from the second peak. Lane M: protein marker. Lane C: concentrated protein.
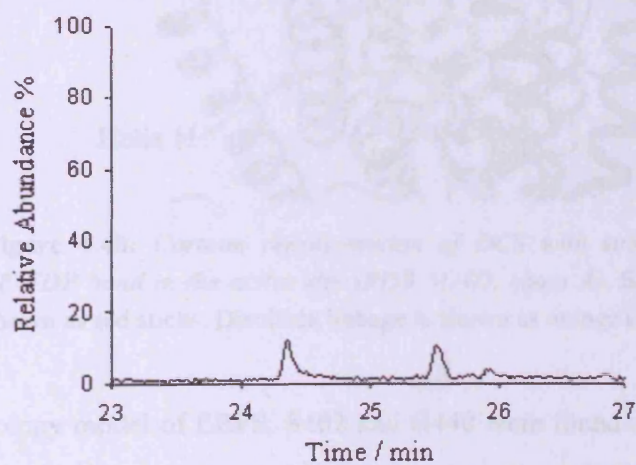
Figure 4.39: *GC trace for incubation of 10 μM EBFS-I404Y/Y514L/A529Y with 1 mM FDP in the presence of 5 mM Mg²⁺.*

Similar to the mutant EBFS-I404Y/N444D/Y514L, EBFS-I404Y/Y514L/A529Y seems to affect the precise binding of the substrate, supported by the fact that the

164

triple mutant showed weak activity rather than loss of activity seen for the mutant EBFS-I404Y/Y514L. However, only acyclic products ((E)-β-farnesene (**65**), (3Z,6E)-α-farnesene (**112**) and (3E,6E)-α-farnesene (**113**)) were formed, which suggests the active site of this mutant is still quite rigid and not able to allow the formation of cyclic intermediates or products.

4.2.9 EBFS-S402C/G440C

In the crystal structure of DCS, a disulfide linkage between C408 and C447 was identified (Figure 4.40). The specific function of this structure has not been discovered; nevertheless this linkage could possibly stabilize helices G and H, leading to a better-defined active site template for the precise intermediates and product to form.
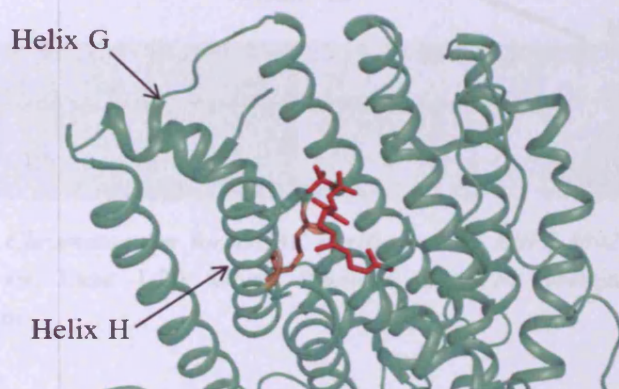
Figure 4.40: *Cartoon representation of DCS with substrate analogue 2F-FDP bond in the active site (PDB 3G4D, chain A).* Bound 2F-FDP is shown as red sticks. Disulfide linkage is shown as orange sticks.

In the homology model of EBFS, S402 and G440 were found at analogous positions to the two cysteine residues from DCS. A double mutant (EBFS-S402C/G440C) was designed to introduce the disulfide bridge into the structure of EBFS. To construct the cDNA of the double mutant, the native codons AGC for S402 and GGT for G440 were replaced by Cys codons TGC and TGT, respectively, by site-directed mutagenesis. The double mutant was purified in the same way as wild-type enzyme (Figure 4.41). Typically, 10 mg of EBFS-S402C/G440C was obtained per litre of

culture. Sesquiterpene products profile was obtained by an incubation of 10 μM EBFS-S402C/G440C with 1 mM FDP at room temperature overnight. GC-MS analysis showed that mainly (*E*)-β-farnesene (**97**) and (3*Z*,6*E*)-α-farnesene (**112**) were identified as sesquiterpene products with a roughly 2 to 1 ratio (Figure 4.42). Kinetic assay of this mutant was not performed.
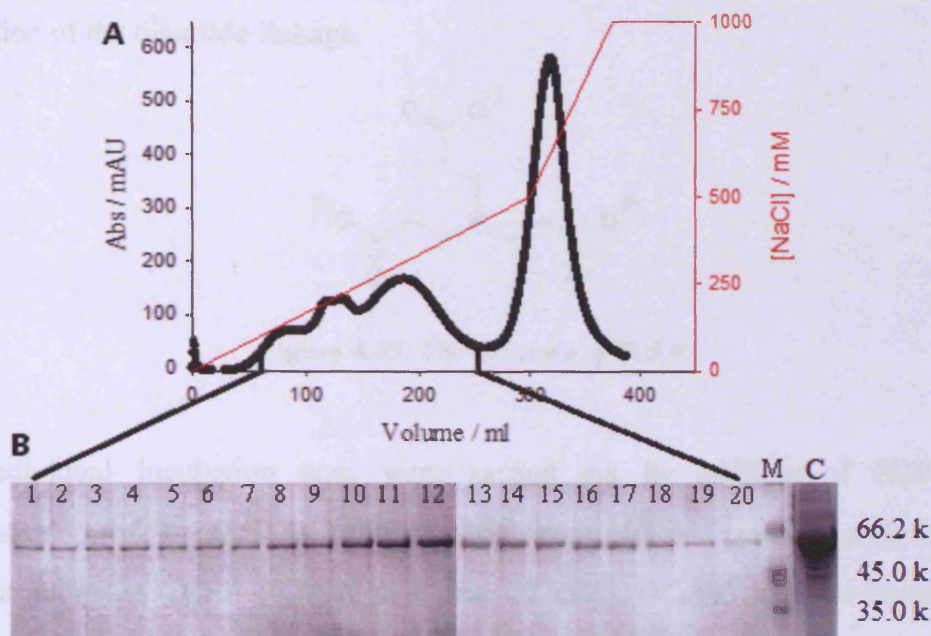
Figure 4.41: *(A) Chromatogram for DEAE purification of EBFS-S402C/G440C. (B) 13% SDS-PAGE analysis*. Lane 1-20: eluent fractions. Lane M: protein marker. Lane C: concentrated protein.
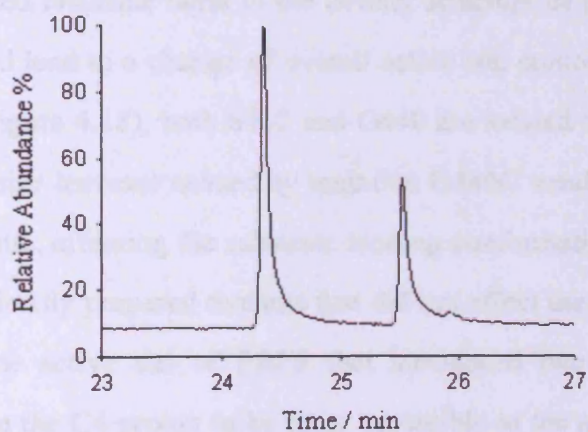
Figure 4.42: *GC trace for incubation of 10 μM EBFS-S402C/G440C with 1 mM FDP in the presence of 5 mM Mg²⁺*.

166

Further investigation of the role of the proposed disulfide linkage was performed by employing tris(2-carboxyethyl)phosphine (TCEP) (Figure 4.43). TCEP is a commonly used reducing agent for the purpose of breaking disulfide bonds between and within proteins (*173*). Thus, a different product profile was expected after addition of TCEP into the reaction solution, if the disturbed product distribution was caused by the formation of the disulfide linkage.
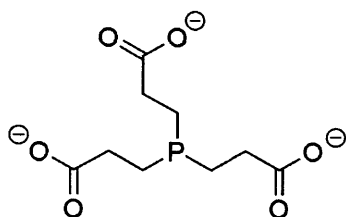
Figure 4.43: *The structure of TCEP.*

Six individual incubation tests were carried out by addition of FDP (final concentration of 1 mM) to solutions containing 10 μM protein with variable concentrations of TCEP (0.1, 1, 3, 8 and 20 mM). GC-MS analysis revealed no obvious change of product profile up to 8 mM TCEP (Figure 4.44), while protein precipitation was observed when the concentration of TCEP was increased to 20 mM. The unchanged product profiles strongly indicate that the two cysteine residues did not form the proposed disulfide bond in the tertiary structure of EBFS. Nevertheless, both mutations could lead to a change of overall active site contour. According to the homology model (Figure 4.18), both S402 and G440 are located in the middle part of the active site. The size increase caused by mutation G440C would greatly reduce the width of the active site, affecting the substrate binding conformation in the active site. Differing from previously prepared mutants that did not affect the product distribution, the alterations in the active site of EBFS that introduced two cysteine mutations would possibly force the C4 proton to be more accessible to the active site base. This was supported by the increased production of (*3Z,6E*)-α-farnesene. In addition, it also revealed a critical role for the structure of helices G and H, which could be the essential component to determine the product specificity of EBFS.
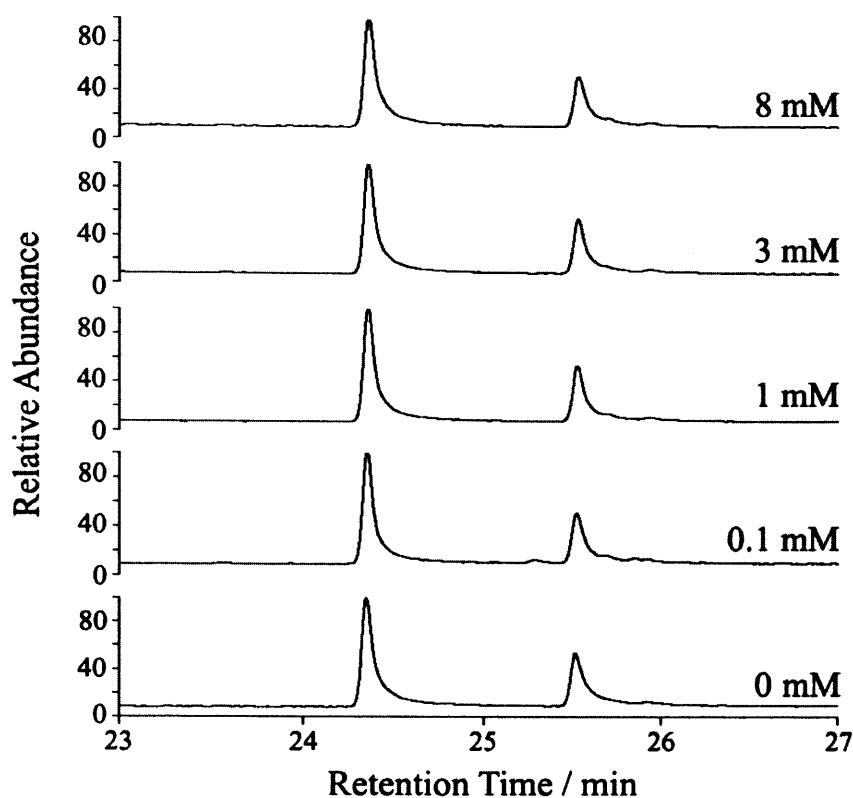
Figure 4.44: *The TCEP concentration dependence of product profiles.* GC traces for incubation of 10 μM EBFS-S402C/G440C with 1 mM FDP at the presence of variable TCEP concentrations.

The enzymatic reaction catalysed by the double cysteine mutant was also performed in non-reducing conditions. The mutant protein was purified without the presence of any reducing agent. 200 μl of the purified protein (20 μM) was first left in the air at room temperature for 12 h and then mixed with 250 μl buffer (20 mM MOPS, 5 mM MgCl$_2$, pH7.2) and 50 μl FDP (10 mM). The incubation was left at room temperature overnight. GC-MS analysis showed an unchanged product profile compared to that obtained under reducing conditions (Figure 4.45).
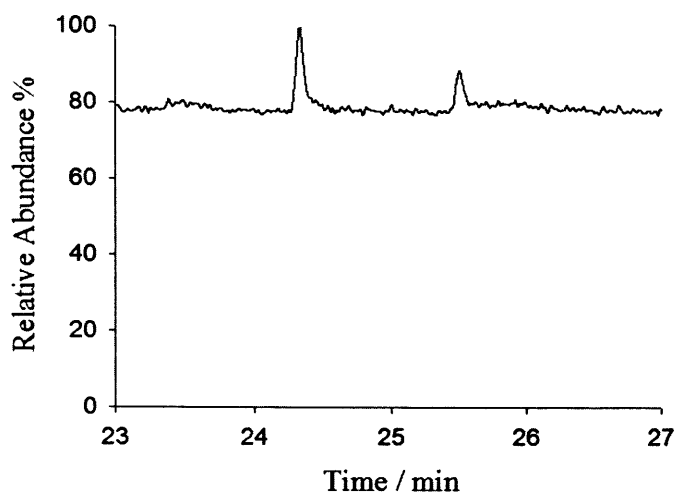
Figure 4.45: *GC trace for incubation of 10 µM EBFS-S402C/G440C with 1 mM FDP in the buffer system without any reducing agent.*

4.2.10 EBFS-S402C/G440C/Y514L

The substitution Y514L was first introduced into the double cysteine mutant to test if the increasing volume at the bottom of active site pocket would have any particular effect on the selectivity of the double cysteine mutant. The triple mutant was produced and purified by the same procedure as the wild-type enzyme (Figure 4.46) typically yielding 23 mg of EBFS-S402C/G440C/Y514L per litre of culture. Incubation of the purified enzyme with FDP showed an unchanged product profile from that of the double cysteine mutant (Figure 4.47). Together with the result from mutant EBFS-Y514L, the unchanged product profile of this triple mutant suggests the residue Y514 has nearly no contribution to the product specificity of EBFS. Kinetic assay of this mutant was not performed.
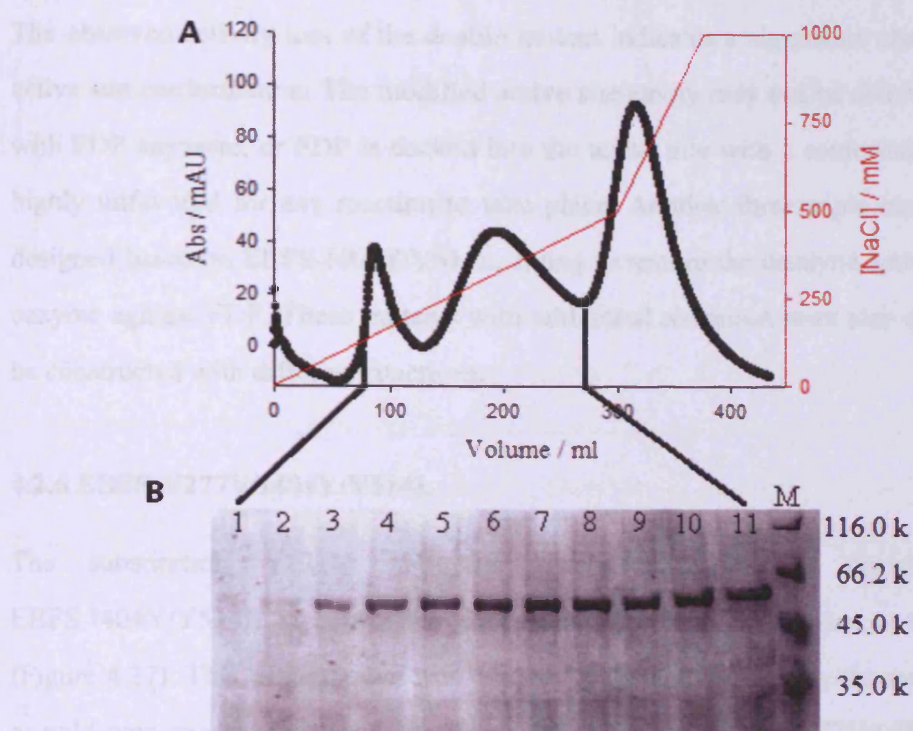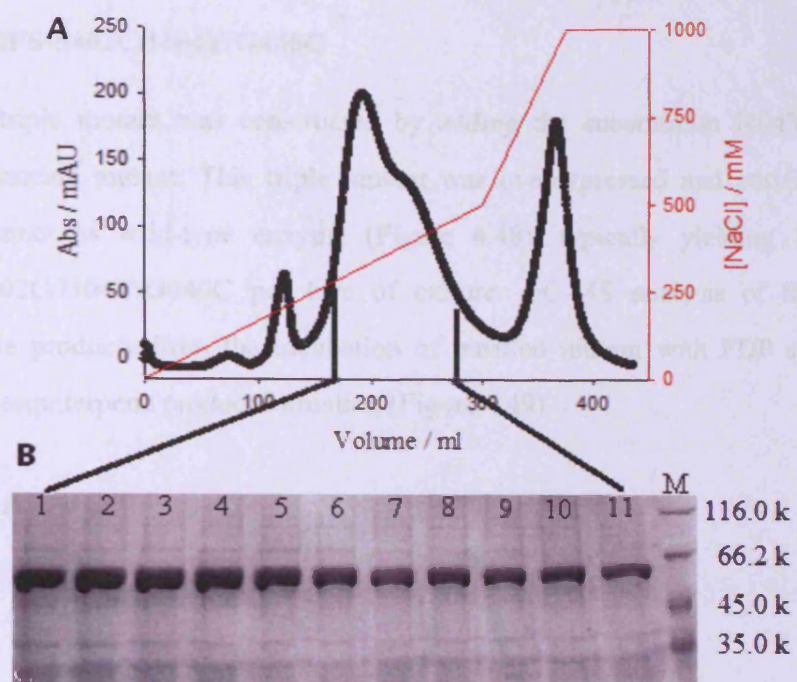
Figure 4.46: *(A) Chromatogram for DEAE purification of EBFS-S402C/G440C/Y514L. (B) 13% SDS-PAGE analysis.* Lane 1-11: eluent fractions. Lane M: protein marker.
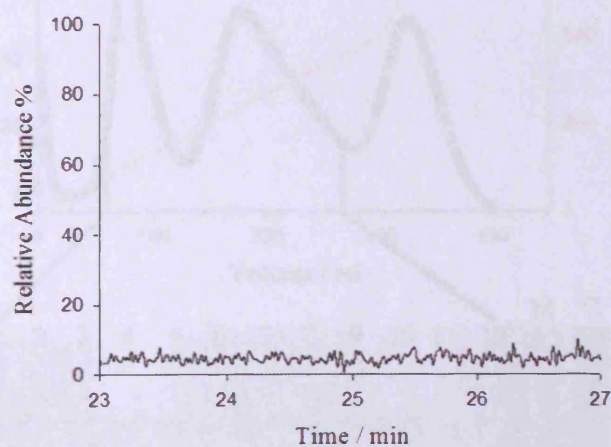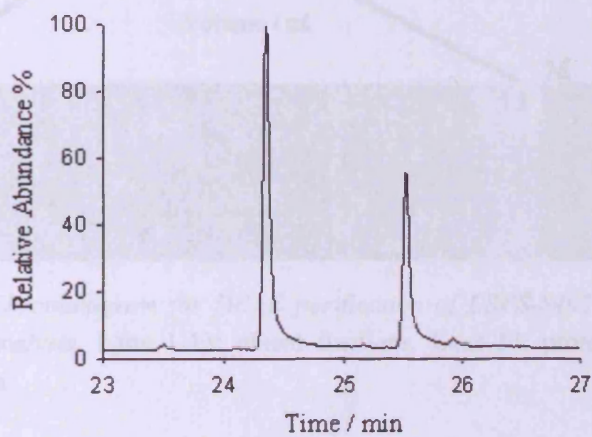

Figure 4.47: *GC trace for incubation of 10 μM EBFS-S402C/G440C/Y514L with 1 mM FDP in the presence of 5 mM Mg²⁺.*

4.2.11 EBFS-S402C/I404Y/G440C

Another triple mutant was constructed by adding the substitution I404Y into the double cysteine mutant. This triple mutant was overexpressed and purified by the same manner as wild-type enzyme (Figure 4.48), typically yielding 10 mg of EBFS-S402C/I404Y/G440C per litre of culture. GC-MS analysis of the hexane extractable products from the incubation of purified mutant with FDP showed no obvious sesquiterpene product formation (Figure 4.49).
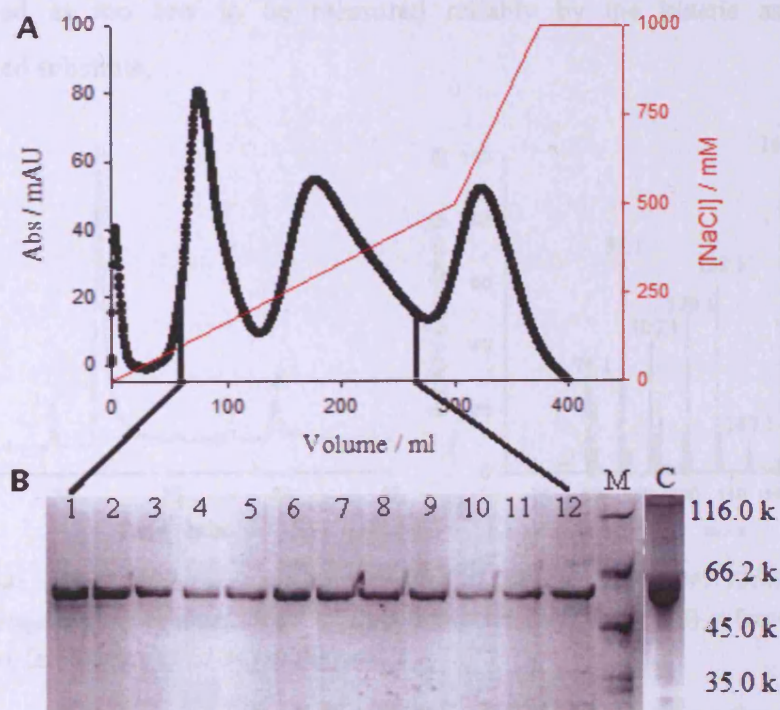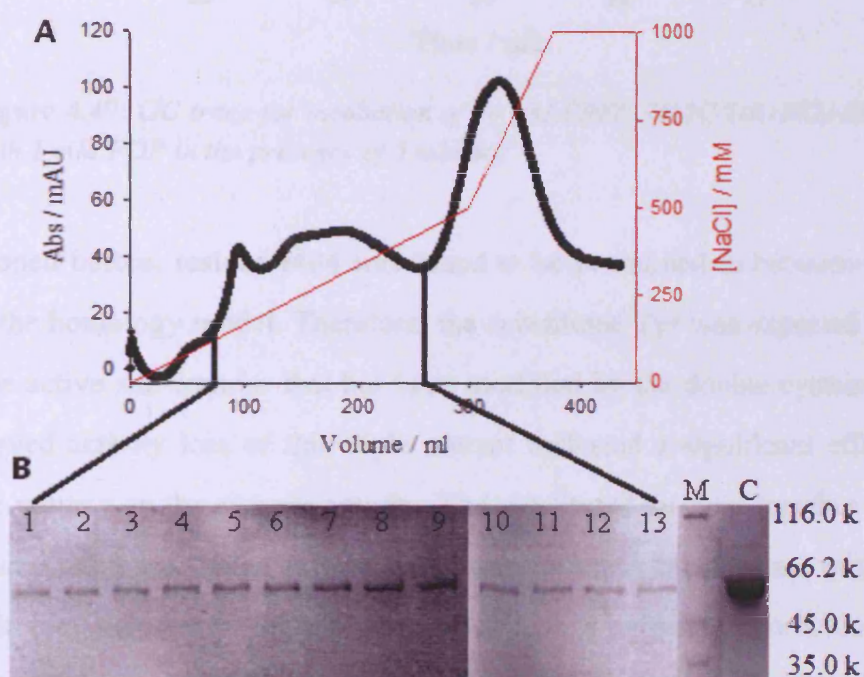

Figure 4.48: *(A) Chromatogram for DEAE purification of EBFS-S402C/I404Y/G440C. (B) 13% SDS-PAGE analysis.* Lane 1-13: eluent fractions. Lane M: protein marker. Lane C: concentrated protein.
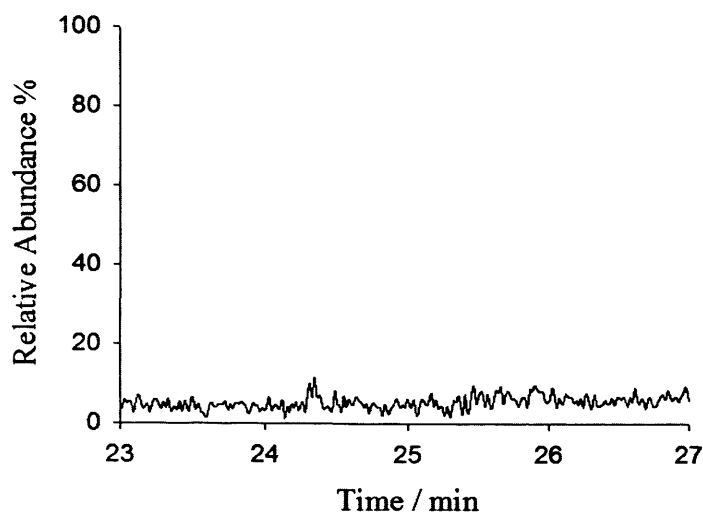
Figure 4.49: *GC trace for incubation of 10 μM EBFS-S402C/I404Y/G440C with 1 mM FDP in the presence of 5 mM Mg^{2+}.*

As mentioned before, residue I404 was found to be positioned in between helices G and H in the homology model. Therefore, the substituted Tyr was expected to further disturb the active site contour that has been modified by the double cysteine mutant. The observed activity loss of this triple mutant indicated a significant effect of the active site volume on the enzyme activity. The substituted residues together brought a local volume reduction to the side of active site cleft and the substrate might not be able to fit into the newly formed active site with a proper conformation for the reaction. In addition, this fact supports the assumption that substrate FDP binds more closely to the side of the active site formed by helices G and H (see Section 4.2.6) and modifications in this area will directly lead to the product promiscuity and the reduction of enzyme activity.

4.2.12 EBFS-S402C/I404Y/G440C/Y514L

The mutant EBFS-I404Y/Y514L has been proved to be a non-catalytic enzyme. Here, a quadruple mutant was prepared by introducing the double cysteine mutations into the mutant EBFS-I404Y/Y514L to check if the further modification to the active site would confer the enzyme with ability to make new sesquiterpenoid products. The

mutant was purified by the same method as wild-type enzyme (Figure 4.50), typically yielding 10 mg of EBFS-S402C/I404Y/G440C/Y514L per litre of culture.

Remarkably, the quadruple mutant showed activity towards FDP rather than the observed activity loss of double mutant EBFS-I404Y/Y514L and triple mutant EBFS-S402C/I404Y/G440C. The product profile of this mutant was very close to the one from the double cysteine mutant, consisting of mainly (*E*)-β-farnesene and (3*Z*,6*E*)-α-farnesene with a tiny amount of (3*E*,6*E*)-α-farnesene (Figure 4.51). A slight increase in production of (3*Z*,6*E*)-α-farnesene was observed from 30% of the double cysteine mutant to 40% of the quadruple mutant. These results indicated that the modifications further altered the binding orientation of farnesyl chain in the active site, which made the proton elimination from C15 became less favoured.
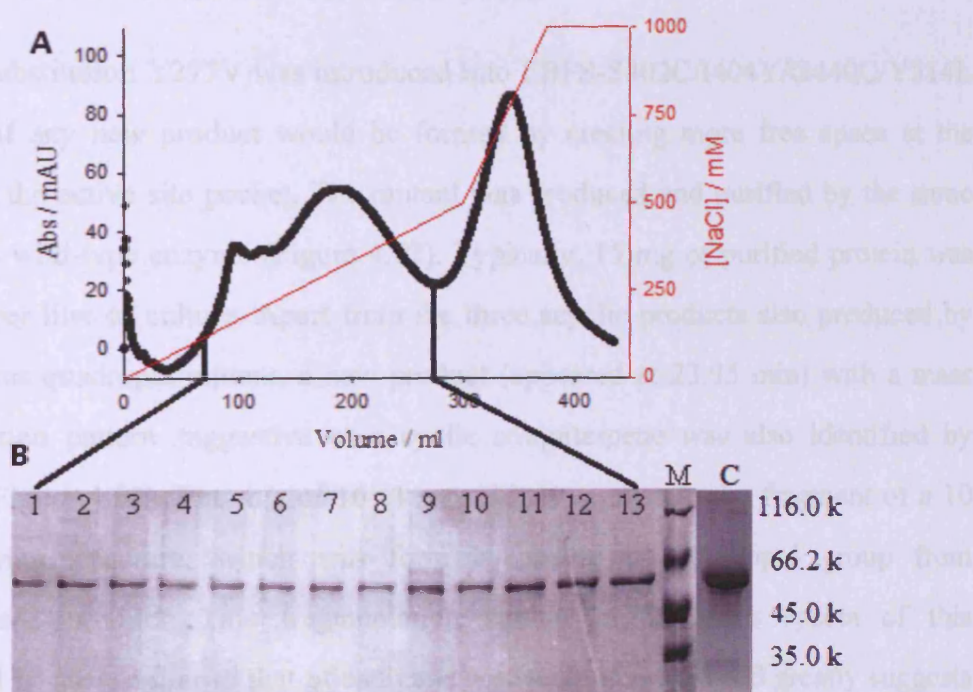
Figure 4.50: *(A) Chromatogram for DEAE purification of EBFS-S402C/I404Y/G440C/Y514. (B) 13% SDS-PAGE analysis*. Lane 1-13: eluent fractions. Lane M: protein marker. Lane C: concentrated protein.

173
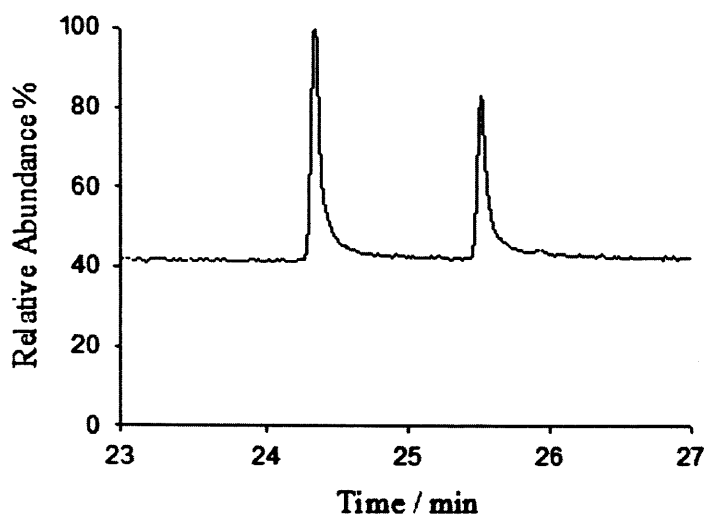
Figure 4.51: *GC trace for incubation of 10 μM EBFS-S402C/I404Y/G440C/Y514L with 1 mM FDP in the presence of 5 mM Mg²⁺.*

4.2.13 EBFS-Y277V/S402C/I404Y/G440C/Y514L

Another substitution Y277V was introduced into EBFS-S402C/I404Y/G440C/Y514L to check if any new product would be formed by creating more free space at the bottom of the active site pocket. The mutant was produced and purified by the same method as wild-type enzyme (Figure 4.52). Typically, 15 mg of purified protein was obtained per litre of culture. Apart from the three acyclic products also produced by the previous quadruple mutant, a new product (appeared at 23.95 min) with a mass fragmentation pattern suggestive of a cyclic sesquiterpene was also identified by GC-MS (Figure 4.53). The m/z of 161.1 most likely represents the fragment of a 10 member ring structure, which was form by losing the isopropyl group from sesquiterpene product. This fragmentation pattern in the mass specta of this compound is very similar to that of cadinane type sesquiterpenes and greatly suggests that a 1,10-cyclisation is involved in the formation of this compound.
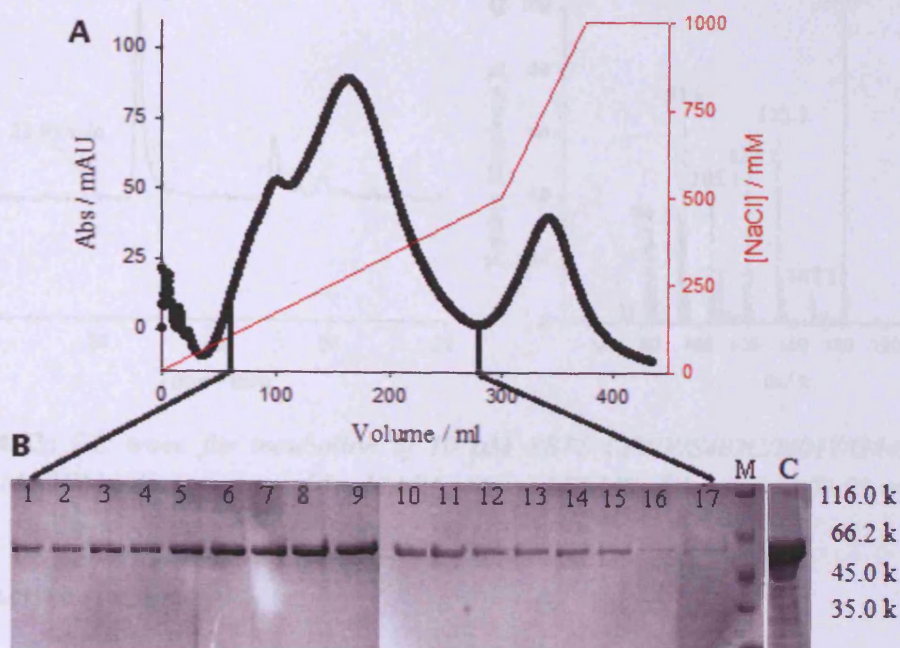
Figure 4.52: *(A) Chromatogram for DEAE purification of EBFS-Y277V/S402C/I404Y/G440C/Y514. (B) 13% SDS-PAGE analysis.* Lane 1-17: eluent fractions. Lane M: protein marker. Lane C: concentrated protein.

This product profile is close to that of mutant EBFS-Y277V/I404Y/Y514L, whereas the increased amount of (3Z,6E)-α-farnesene (16%) represents the impact of the double cysteine mutant. In addition, the increased amount of the unknown sesquiterpene product (7%) supports the previously proposed explanation that modifications on the helices G and H forces the substrate to move towards the other side of the active site and the enlarged space created by mutation Y277V could possibly accommodate substrate in a cyclic form to initiate cyclisation. We speculated that the mutant produces mainly acyclic compounds probably because the π-orbital of the distal double bond and the empty p-orbital of C1 could not reach a perfect alignment required for the alkylation. Thus, only a small amount of substrate FDP can be converted into cyclic product.
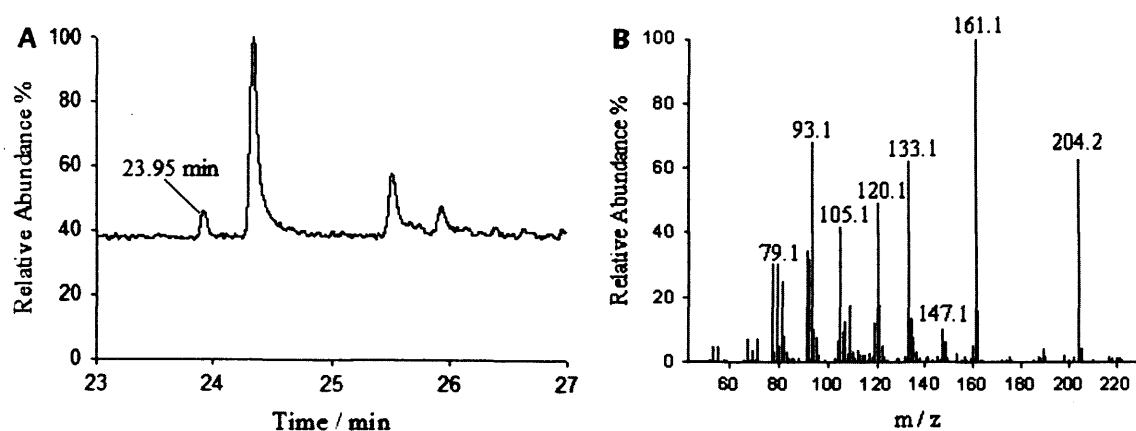
Figure 4.53: *GC trace for incubation of 10 μM EBFS-Y277V/S402C/I404Y/G440C/Y514L with 1 mM FDP in the presence of 5 mM Mg²⁺ (A) and EI-MS of the peak at 23.95 min.*

4.2.14 Active site hybrid

Plant sesquiterpene synthases exhibit a significant degree of similarity at the amino acid level and this feature leads to some successful examples of manipulating enzyme function through a small amount of amino acid substitutions in the active site (*155, 174*). However, in our studies all the attempts to modify the active site volume to mimic other plant sesquiterpene synthases failed to alter the function of EBFS from generating cyclic sesquiterpene product as major product. To get a better understanding of the relationship between active site contour and product selectivity, a hybrid was prepared by substitution residues of EBFS that have contact with the substrate FDP within 8 Å with corresponding ones from DCS (Figure 4.54). The whole design was based on the deduced amino acid sequence alignment and the homology model of EBFS. In total, 50 residues of EBFS were identified have contact with FDP within 8 Å, in which 34 residues were different to the corresponding residues from DCS (Figure 4.55). Thus, these 34 residues were chosen to be substituted.
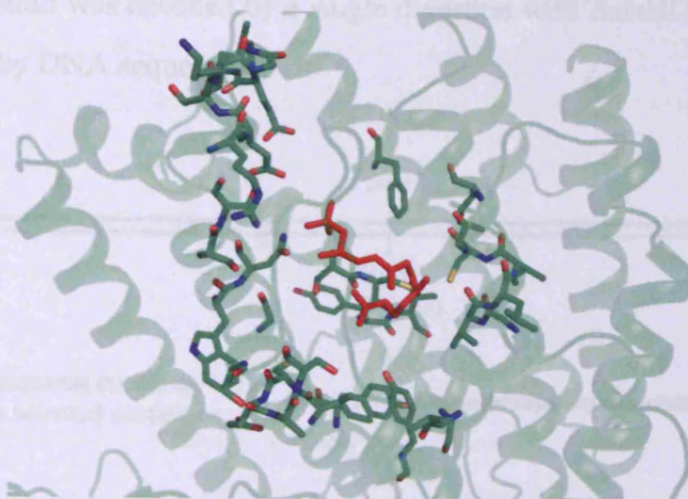
Figure 4.54: *Cartoon representation of the EBFS homology model.* FDP was docked into the active site by the program MOE (red). Selected mutations are shown as sticks.

```
  1   MATNGVVISCLREVRPPMTKHAPSMWTDTFSNFSLDDKEQQKCSETIEALKQEARGMLMA
 61   ATTPLQQMTLIDTLERLGLSFHFETEIEYKIELINAAEDDGFDLFATALRFRLLRQHQRH
121   VSCDVFDKFIDKDGKFEESLSNNVEGLLSLYEAAHVGFREERILQEAVNFTRHHLEGAEL
181   DQSPLLIREKVKRALEHPLHRDFPIVYARLFISIYEKDDSRDELLLKLSKVNFKFMQNLY
241   KEELSQLSRWWNTWNLKSKLPYARDRVVEAYVWGVGYHYEPQYSYVRMGLAKGVLICGIM
301   DDTYDNYATLNEAQLFTQVLDKWDRDEAERLPEYMKIVYRFILSIYENYERDAAKLGKSF
361   AAPYFKETVKQLARAFNEEQKWVMERQLPSFQDYVKNSEKTSCIYTMFASIIPGLKSVTQ
421   ETIDWIKSEPTLATSTAMIGRYWNDTSSQLRESKGGEMLTALDFHMKEYGITKEEAASKF
481   EGLVEETWKDINKEFIATTNYNVGREIAITFLNYARICEASYSKTDGDAYLDPNVAKANV
541   VALFVDAIVF
```

Figure 4.55: *The deduced amino acid sequence of WT-EBFS.* The mutations on the active site hybrid are coloured orange.

The construct of the hybrid was created by mutating the native codon TAT which encodes amino acid Y271 to AGC; together with the codon GCT that encodes amino acid A270, this mutation creates a *Nhe*I restriction site. A synthetic DNA fragment was designed to encode the amino acid sequence after V272 of EBFS with all the chosen mutations and constructed with *Nhe*I and *Bam*HI restriction sites at the 5' and 3' end respectively. Double digestion of the mutated plasmid with *Nhe*I and *Bam*HI restriction enzymes and subsequent ligation with the synthetic DNA fragment formed a plasmid containing the gene encoding the hybrid (Figure 4.56). Formation of the

constructed plasmid was checked by a single digestion with *Bam*HI (Figure 4.57) and also confirmed by DNA sequencing.
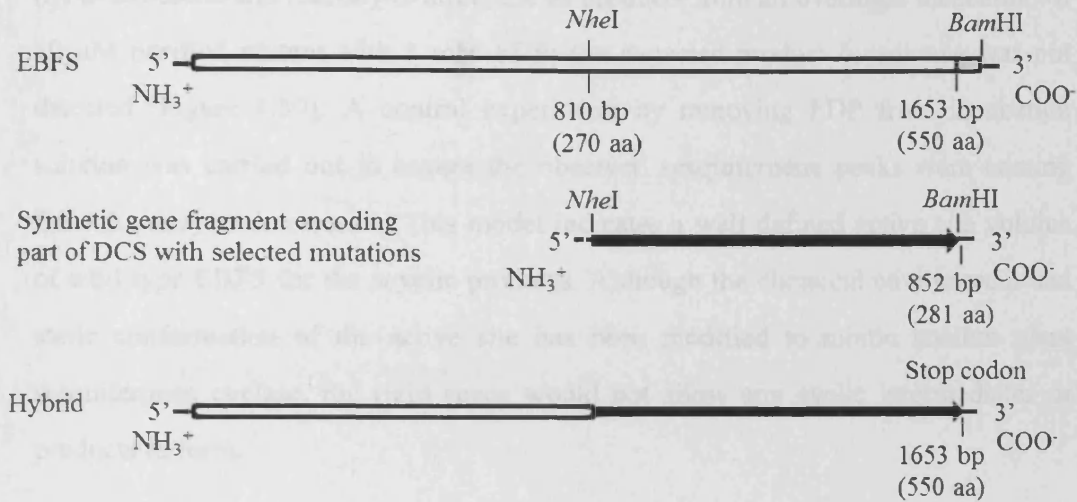
Figure 4.56: *Schematic diagram of construction of the active site hybrid.* Line drawings depict composite diagrams for wild-type EBFS (coloured white), synthetic gene fragment encoding part of DCS with selected mutations (coloured black) and hybrid gene.
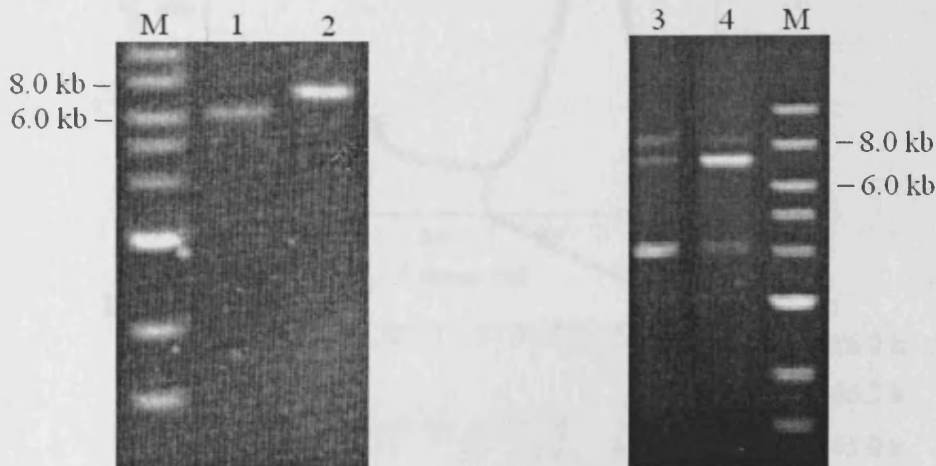
Figure 4.57: 1% *Agarose gel analyses of the active site hybrid cDNA construct preparation.* Lane 1: The mutated insert gene after double digestion by both *Nhe*I and *Bam*HI restriction enzymes. Lane 2: The mutated insert gene after digestion by *Bam*HI restriction enzyme. Lane 3: The gene plasmid of the hybrid. Lane 4: The gene of the hybrid after digestion by *Bam*HI restriction enzyme. M: 1 kb DNA ladder.

The hybrid was purified by the same method as wild-type EBFS (Figure 4.58), typically yielding 6 mg of hybrid per litre of culture. Surprisingly, the overall active

site modification leads to a huge reduction in enzyme activity. GC-MS analysis showed only two weak signals (barely higher than the noise level) representing (*E*)-β-farnesene and (3*Z*,6*E*)-α-farnesene as products from an overnight incubation of 10 μM purified mutant with 1 mM FDP; the expected product δ-cadinene was not detected (Figure 4.59). A control experiment by removing FDP from incubation solution was carried out to ensure the observed sesquiterpene peaks were coming from the enzymatic reaction. This model indicates a well-defined active site volume of wild-type EBFS for the acyclic products. Although the chemical environment and steric conformation of the active site has been modified to mimic another plant sesquiterpene cyclase, the rigid space would not allow any cyclic intermediates or products to form.
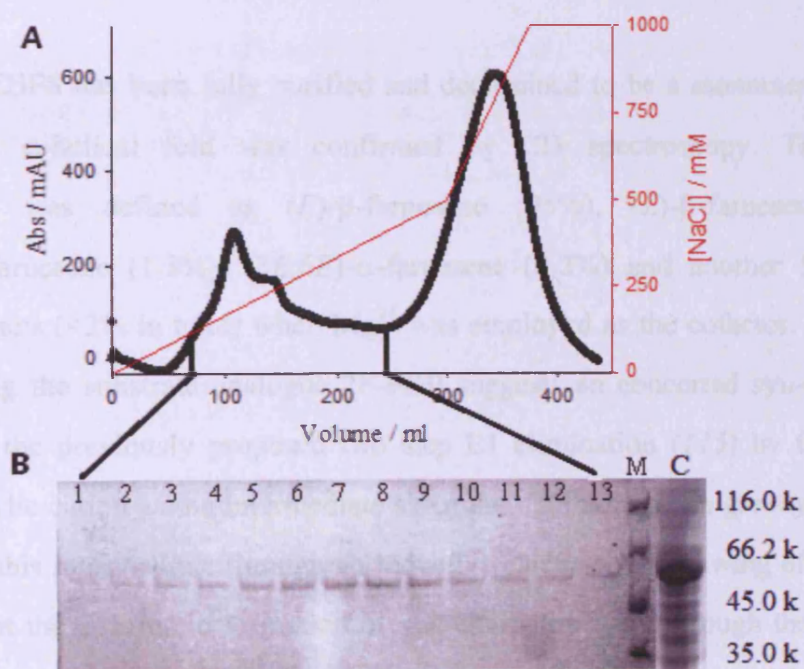
Figure 4.58: *(A) Chromatogram for DEAE purification of the active site hybrid. (B) 13% SDS-PAGE analysis.* Lane 1-13: eluent fractions. Lane M: protein marker.
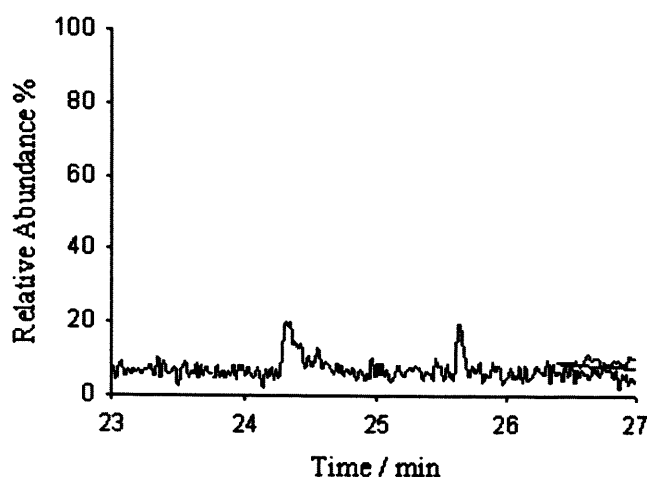
Figure 4.59: *GC trace for incubation of 10 μM the active site hybrid with 1 mM FDP in the presence of 5 mM Mg^{2+}.*

3.3 Conclusions

Wild-type EBFS has been fully purified and determined to be a monomer at pH 7.2. An overall α-helical fold was confirmed by CD spectroscopy. The product distribution was defined as (E)-β-farnesene (95%), (Z)-β-farnesene (1.5%), ($3Z,6E$)-α-farnesene (1.3%), ($3E,6E$)-α-farnesene (0.2%) and another 5 unknown cyclic products (<2% in total) when Mg^{2+} was employed as the cofactor. Mechanitic studies using the substrate analogue 2F-FDP suggests an concerted syn-elimination rather than the previously proposed two step E1 elimination (*115*) by forming the transoid allylic cation as the intermediate since the C2 fluorine can greatly inhibit the forming of this intermediate through an inductive electron withdrawing effect. It was believed that the enzymatic formation of sesquiterpenes went through the ionization of *trans-trans*-farnesyl diphosphate to the corresponding transoid allylic cation. However, the E2 elimination suggests an alternative route to (E)-β-farnesene. It seems that the enzymatic reaction catalysed by EBFS is under kinetic control since the predominant product (E)-β-farnesene is not the most stable product. In addition, the formation of transoid allylic cation intermediate in sesquiterpene cyclization reactions was also reported to be under kinetic rather than thermodynamic control (*175*).

Alternatively, purified wild-type EBFS catalyses the conversion of C_{10} analogue GDP to mainly acyclic products (myrcene (38.3%), (E)-ocimene (29.4%) and (Z)-ocimene (18.9%)), rather than the previously reported major product limonene (48%) that was generated by the partially purified enzyme. The broad distribution towards three acyclic products is probably due to the large active site volume which leads to different substrate GDP binding conformation, allowing the active site base to approach protons from both C4 and C10.

In terms of the proposed active site volume theory (Section 4.2.1), three single mutants (Y277V, I404Y and Y514L) were prepared to mimic the conformation of this region in the other four plant sesquiterpene synthases, but none of them had a significant influence on the product outcome, only a 20-100 fold catalytic efficiency reduction was observed for mutants I404Y and Y514L. The observed activity loss in the double mutant EBFS-I404Y/Y514L and restored activity of the triple mutant EBFS-Y277V/I404Y/Y514L indicate that the position of the farnesyl chain in the active site of wild-type EBFS should be close to the helices G and H. After introducing mutations I404Y and Y514L, the active site contour could not accomodate the farnesyl chain in a reactive conformation, leading to activity loss. Mutation Y277V brought more space into the active site of the double mutant and the substrate can be accommodated into the enlarged active site again but more towards the opposite site of helices G and H with a slightly different conformation. The modified active site contour plus the possibly different substrate binding conformation led to the formation of an unknown cyclic sesquiterpene by the intramolecular attack of the 10,11- or 6,7- double bond on C1.

Mutagenesis studies at the two conserved Mg^{2+} binding motifs showed the possibility of altering both the enzyme activity and product profile (95, 170-172). When the mutation N444D (from the second Mg^{2+} binding motif N^{444}DTSSQLRE) was introduced into both wild-type EBFS and the double mutant I404Y/Y514L, only a huge enzyme activity reduction was observed. The observed activity loss by

manipulating the Mg^{2+}-binding site suggests a significant disturbance of the precise substrate binding conformation. This change did not generate any new product, possibly due to the rigid active site pocket, which would barely accommodate any kind of cyclic intermediate or product.

The active site contour was also modified by introducing two cysteine substitutions, which have been observed to form a disulfide linkage in the crystal structure of DCS. However, treatment with TCEP suggested that the newly introduced two cysteine residues did not form the proposed disulfide linkage in the tertiary structure of EBFS. Nevertheless, the double cysteine mutant alters the product distribution from (*E*)-β-farnesene (95%) to 65% (*E*)-β-farnesene, 33% (3*Z*,6*E*)-α-farnesene and 1% (3*E*,6*E*)-α-farnesene. The increased production of (3*Z*,6*E*)-α-farnesene suggests a possible active site volume reduction, which would squeeze the farnesyl chain more towards the entrance of the pocket, leading the C4 proton to be more accessible to the diphosphate group (proposed active site base). This is consistent with the apparent function of the two cysteine mutations that could reduce the width of the active site pocket. In addition, the alteration of the product distribution by modifications on the helices G and H also indicates the structure of helices G and H may be the key component that determines the product specificity for the enzyme.

Additional modifications toward manipulating the active site volume did not confer the enzyme with obvious new function. The activity loss of the triple mutant EBFS-S402C/I404Y/G440C suggests that the modified active site was not able to accommodate the substrate with the proper conformation due to the volume reduction caused by the mutations. This can be supported by the fact that releasing some space in the active site of the non-catalytic enzyme by mutating the aromatic residue Y514 into a smaller hydrophobic residue Leu restored catalytic ability of the enzyme. The quintuple mutant EBFS-Y277V/S402C/I404Y/G440C/Y514L even reveals a combined effect of the mutants EBFS-Y277V/I404Y/Y514L and EBFS-S402C/G440C by showing an increased production of (3*Z*,6*E*)-α-farnesene and

the newly formed unknown sesquiterpene (retention time of 23.95 min) compared to the wild-type enzyme.

The observed activity loss of the active site hybrid simply illustrates the importance of the active site volume. In other words, the active site of the hybrid does not contain a proper cavity to dock cyclic intermediates and products, although it contains a nearly identical chemical environment to DCS. In addition, the modification also compromises the ability to form (E)-β-farnesene, suggesting either the conformation of the Mg^{2+} binding area has been severely disturbed or the modified active site contour cannot accommodate the native substrate FDP.

Table 4.5: *Summary of kinetic data for the mutants of EBFS.*

Protein	k_{cat} / s^{-1}	K_M / μM
WT-EBFS	0.015 ± 0.004	6.4 ± 1.7
EBFS-Y514L	$(5.5 ± 1.9) \times 10^{-4}$	23 ± 8
EBFS-I404Y	0.0037 ± 0.0016	29 ± 12
EBFS-Y277V	0.017 ± 0.009	18 ± 7
EBFS-I404Y/Y514Y	Inactive	
EBFS-Y277V/ I404Y/Y514Y	Poor actvity	
EBFS-N444D	Inactive	
EBFS-I404Y/N444D/Y514Y	Inactive	
EBFS-I404Y/Y514Y/A529Y	Poor actvity	
EBFS-S402C/G440C	Active	
EBFS-S402C/G440C/Y514L	Active	
EBFS-S402C/I404Y/G440C	Inactive	
EBFS-S402C/I404Y/G440C/Y514L	Poor activity	
EBFS-Y277V/S402C/I404Y/G440C/Y514L	Poor activity	
Active site hybrid	Inactive	

Chapter 5: Dual roles of the N-Terminal Region in Catalysis

5.1 Introduction

Sesquiterpene synthases from bacteria and fungi are single-domain enzymes. These enzymes contain a class I terpenoid synthase fold domain constructed with several α-helices connected by short loops. Plant sesquiterpene synthases usually contain an additional N-terminal domain (*91*). This domain adopts an α-helical fold closely related to glycosyl hydrolases and has a structure similar to the N-terminal domain of the class II terpenoid synthase fold (*92*). Recently, the crystal structure of *A. grandis* a-bisabolene synthase (AgBIS) was solved (*176*), which is the first three-domain plant sesquiterpene structure. Unlike the previous structurally characterised plant sesquiterpene synthases (TEAS and DCS) that have only two domains, AgBIS composes an additional domain (insertion domain), which can be found in diterpene synthases such as taxadiene synthase and copalyl diphosphate synthase (*130, 138*). The C-terminal domain of AgBIS contains the class I terpenoid synthase fold and is determined to be the functionally active domain. While, the N-terminal domain together with the insertion domain of AgBIS exhibits a non-functional class II terpenoid synthase fold, since AgBIS lacks the conserved DXDD motif that initiates the cyclization of linear isoprenoids. Thus, AgBIS is proposed to be potentially an intermediate in the evolutionary transition from diterpene to sesquiterpene synthase (*176*). In this sense, the N-terminal domain of plant sesquiterpene synthase may function as part of the class II terpenoid synthase active site for the ancestral origin of this enzyme family. However, the function of the N-terminal domain among two-domain plant sesquiterpene synthases is still not clear.

It has been proposed (*57*) that the N-terminal domain of plant sesquiterpene synthase plays a role in capping the active site of the C-terminal class I terpenoid synthase domain. In the crystal structures of two plant sesquiterpene synthases (TEAS and DCS), the N-termini are above the entrance of the active site, which were proposed to protect the active site from exposure to the solvent during catalysis (*75, 114*). Based on the crystal structure of TEAS, it has also been suggested that the N-terminal region

may have contact with A-C, D-D1, J-K and H-H-α1 loops for the purpose of protecting the active site (*91*). Thus it is tempting to suggest that the N-terminal region could have influence on catalysis. Based on the sequence alignment of EBFS with other plant sesquiterpene synthases, the initial part (first 14 amino acids) of the N-terminal region (M1–K38) shares nearly no similarity, yet the rest part of sequence shares a considerable amount of similarity and identity (Figure 5.1).

```
GAS     1  MAAVEAKGTLQANTKTTTEPVKPLANFPPSVWGDRFLSFSLDMKELEGYAKAMEEPKSELRRLIVDSTMDSNKKLSLIYSVHRLGLTYLF
GCS     1     MAASS.....ADKCRPLANFHPSVWGYHFLSYT.HEITNQ.EKVEVDEYKETIRKPLVETCDNSTQKLVLIDAMQRLGVAYHF
TEAS    1     MASAAVANYEEEIVRPVADFSPSLWGDQFLSFSIKMNQVAEKYAKEIEALKEQTRNPLLATGMKLADTLNLIDTIERLGISYHF
DCS     1  MASQVSQMPSSSPLSSNKDEMRPKADFQPSIWGDLFLNCPDKNIDAE.TEKRHQQLKSEVRKPIVAPMANSTQKLAFIDSVQRLGVSYHF
EBFS    1     MATNGVVISCLREVRPPMTKHAPSMWTDTFSNFSLDDKEQQKCSETIEALKQRARGPLMAATTP.LQQMTLIDTLERLGLSFHF

GAS    91  LQEIEAQLEKLFKEFWLQDYDEVDLYTTSINFQVFRHLGHKLPPDVFNKFKDSTSGKFKEYITNDVKGMLGLYESAQLKVRGESILDEAS
GCS    77  DKEIETSIQNIFDASSKQNDNDMNLYVVSLRPRLVRQQGHYMSSDVFKQFTN.QDGKFKETLTNDVQGLLSLYEASHLRVRKEEILEEAL
TEAS   84  EKEIDDILDQIYNQN.....SNCNDLCTSALQFRILRQHGENISPKIFSKFQD.ENGKFKESIASRVLGLLNLYEASHVRTHADDILEDAL
DCS    90  TKEISDELENIYHNK...KDAEMDLYTTSIRFRLLREHGYNVSPDVFNKFKD.EQGNFKSSVTSDVRGLLELYQASYLRVHGEDILDEAI
EBFS   84  ETEIEYKIELINAAE.....DRGFDLFATALRFRLLRQHQRHVSPRVFDKFID.KDGKFKESLSNNVEGLLSLXEAAKVGFRERRILQEAV

GAS   181  AFTETKLKSVVNTLEGK...LESQVMQSLRRPFHQGMPMVEARLYFS.NYTEECATHASLLKLAKLHFSYLELQQKEELRIVSQWWKDME
GCS   166  TFTTTHLEBIVSNLSNWNNWSLKVEVGEALTQPIRMTLPRMGARKYIS.IYENKDAHHHLLLKFAKLDFNMLQKFKQRSLSDLTRWWKDLD
TEAS  169  AFSTIHLESAAPHLKSP...LREQVTHALEQCLHKGVPRVETRFFISSIYDKEQSKNNVLLRFAKLDFNLLQMLHKQELAQVSRWWKDLD
DCS   176  SFTTHKLSLAVASLDHP...LSEEVSHALKQSIRRGLPRVEARHYLS.VYQDIESHNKALLEFAKIDFNMLQFLRKELSEICRWWKDLD
EBFS  169  NFTRHHLRG..AELDQSPLLIREKVKRALEHPLHRDFPIVYARLFIS.IYEKDDSRDELLLKLSKVNFKFMQNLYKEELSQLSRWWNTWN

GAS   267  FQTSVPYIRDRVPEIYLWILGLYFEPRYSLARIIATKITLFLVVLDDTYDAYATIEEIRLLTDAIKRWDISAMEQIPEYIRPFYKILLEE
GCS   255  FANKYPYARDRLVECYFWILGVYFEPKYSRARKPMTKVLKLTSIIDDTFDAYATFDELVTFKDAIQRWDANAIDSIQPYRPAIQALLDI
TEAS  256  FVTTLPYARDRVVECYFWALGVYFEPQYSQARVPLVKTISMISIVDDTFDAYGTVKELEAYTDAIQRWDINEIDRLPDYMKISYKAILDL
DCS   262  FQRKLPYARDRVVEGYFWISGVYFEPQYSLGRKPLTKVIAMASIVDDTYDSYATYEELIPYTWAIERWDIKCIDEIPEYMKPSYKALLDV
EBFS  256  LKSKLPYARDRVVRAYVMGVGYHYEPQYSYVRMGLAKGVLICGIMDDTYDNYATLNEAQLFTQVLDKWDRDEAERLPEYMKIVYRFILSI

GAS   357  YAELEKQLAKEGREKSVIASKQAFQDIANGYLEEAEWTNSG.YVATFPEYMKNGLITSAYNVISKSALVGGEMVSEHALAWYESHPKTL
GCS   345  YSEMEQVLSKEGKLDRVYYAKNEPKKLVRAIFKETQWLNDCDHIPKYEEQVEWAIVSAGYMMISTTCLVGIEEFISHETFEWLMNESVIV
TEAS  346  YKDYEKELSSAGRSHIVCHAIERPKEVVRNYNVESTWFIEG.YTPPVSEYLSWALATTTYYYLATTSYLGPKS.ATEQDFEWLSKNPKIL
DCS   352  YEEMVQLVAEHGRQYRVEYAKNAPIRLAQSYLVEAKWTLQN.YKPSFEEFKANALPTCGYAMLAITSFVGPGDIVTPETFKWAASDPKII
EBFS  346  YENYERDAAKLGKSFAAPIYFKETVKQLARAFNEEQKWVMER.QLPSFQDYVKNSEKTSCIYTMFASIIPGLKS.VTQETIDWIKSEPTLA

GAS   446  QASELISRLQDDVMTYQFERERGQSATGVDSYIKTYGVSEKEAIDELNKMIEKAWKDINEGCLKPR..EVSMDVLAPILNLARMIDVVYR
GCS   435  RASALIARAMNDIVGHEDEQERGHVASLIECYPKDYGABKQETYIKFLKEVTKAWKDINKQFSRPT..EVPMFVLERVLELTRVADTLYK
TEAS  434  EASVIICRVIDDTATYEVEKSRGQIATGIEPCPRDYGISTKEAMAKFQNMAETAWKDINEGLLRPT..PVSTEFLTPILNLARIVEVTYI
DCS   441  QASTIICRFMDDVAEHKFKHRREDDCSAIECYIEEYGVTAQEAYDVFNKHVESAMKDLMQEFLKPT..SMPTEVLKRSLNLARVMDVLYR
EBFS  434  TSTAMIGRYWNDTSSQLRESKGGEMLTALDFHMIKEYGITREEAASKFEGLVEETWKDINKEFIATTNYNVGREIAITFLNYARICEASYS

GAS   534  Y.DDGFTFPGKTMKEYTTLLFVGSSPM
GCS   523  E.KDTYSTAKGKLKMINPILIESVKI
TEAS  522  HNLDGYTHPEKVLKPHIINLLVDSIKI
DCS   529  E.GDGYTYVGKAAKGGITSLLIEPIAL
EBFS  524  KTDGDAYLDPKVAKANVVALFVDAIVF
```

Figure 5.1: *Amino acid sequence alignment of five plant sesquiterpene synthases.* GAS: germacrene A synthase from *Crepidiastrum sonchifolium*; GCS: germacrene C synthase from *Solanum lycopersicum*; TEAS: 5-*epi*-aristolochene synthase from *Nicotiana tabacum*; DCS: δ-cadinene synthase from *Gossypium arboreum*; EBFS: (*E*)-β-farnesene synthase from *Mentha x piperita*. Similarity and identity between residues are represented by different colours.

The N-terminal regions (M1-S23 of TEAS, M1-P29 of DCS) are missing in the crystal structures of these plant sesquiterpene synthases and the mobility of this region may relate to the function of covering the active site. Based on the homology model and amino acid sequence alignment (Figure 5.1), the analogous region of EBFS is defined from M1 to S24 and this part of the N-terminal region could share the same function and therefore protecting the active site from bulk solvent.

Previous examination of the role of the N-terminal domain regions revealed that swapping the first 152 amino acids of a *Hyoscyamus muticus* vetispiradiene synthase with the analogous part of *Nicotiana tabacum* 5-*epi*-aristolochene synthase (TEAS) gave no novel product but only a slight decrease in specific activity (*161*). Another study of the N-terminal region of *Ixeris dentate* germacrene A synthase indicates an important role of this region towards the enzyme catalytic activity (*177*).

Our investigations based on the plant sesquiterpene synthase EBFS indicate dual roles of the N-terminal region in the catalytic reaction. Kinetic studies with mutated, truncated and chimeric enzymes demonstrate the importance of this loop for catalysis and suggest that this region regulates product release.

5.2 N-Terminal region truncated mutants

5.2.1 EBFSΔ14

According to the sequence alignment (Figure 5.1), the initial part of the N-terminal region (first 14 amino acids) of EBFS shares less than 20% similarity with other plant sesquiterpene synthases. In order to find out if this part of the N-terminal region could make contribution to catalysis, a truncated mutant (EBFSΔ14) was designed with the first 14 amino acids of EBFS removed. The full-length cDNA for EBFS (1.6 kb) was initially cloned into the vector pET32b (5.9 kb) with an *Nco*I restriction site at the 5' end (Figure 5.2).

Figure 5.2: *pET-32b vector including the cDNA sequence for EBFS.* The N-terminal and the C-terminal domains are separated by residue E223. The N-terminal region is coloured blue, including sequence from M1 to K38.

To construct the gene of EBFSΔ14, another *Nco*I restriction site was introduced into the DNA sequence of the wild-type enzyme. The foreign *Nco*I restriction site was formed by replacing the DNA fragment 5'-GTAAGGCCA-3' that encodes Val14, Arg15 and Pro16, with 5'-GCCATGGCA-3' using site-directed mutagenesis (Figure 5.3).

```
                      3'-CGAATTCCCTTCGGTACCGTGGATACTGCTTCG-5'
 1 ATGGCTACAAACGGCGTCGTAATTAGTTGCTTAAGGGAAGTAAGGCCACCTATGACGAAGC
62 ATGCGCCAAGCATGTGGACTGATACCTTTTCTAACTTTTCTCTTGACGATAAG - 3'
```

Figure 5.3: *The 5' to 3' N-terminal region of EBFS gene sequence (red) and the forward primer (purple, primer sequence is reversed and complementary) that generates NcoI restriction site (underlined) for construction of EBFSΔ14.*

The mutated plasmid was then treated with *Nco*I and the resulting large DNA fragment was intramolecularly ligated to give the desired plasmid (Figure 5.4). As a result of introducing the *Nco*I restriction sites, Arg15 of wild-type EBFS was changed to Ala2 in EBFSΔ14. The truncated mutant was produced and purified by the same protocol for the wild-type enzyme (Figure 5.5). Typically, 16 mg of EBFSΔ14 was obtained per litre of culture.

Figure 5.4: 1% *Agarose gel analyses of the EBFSΔ14 construct preparation.*
Lane 1: linearized plasmid that encodes EBFSΔ14 after *Nco*I digestion. Lane
2: plasmid that encodes EBFSΔ14 after ligation. M: 1 kb DNA ladder.
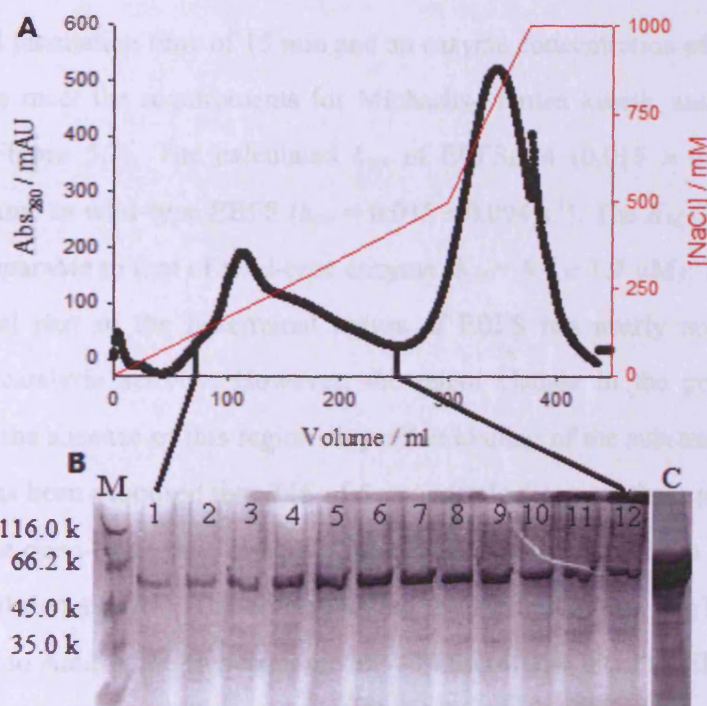
Figure 5.5: *(A) Chromatogram for DEAE purification of EBFSΔ14. (B) 13% SDS-PAGE
analysis.* Lane 1-12: eluent fractions. Lane M: protein marker. Lane C: concentrated protein.

The product profile was analysed based on an overnight incubation of 10 μM enzyme
with 1 mM FDP. GC-MS revealed that the main product (*E*)-β-farnesene still took up
94% of the total production (Figure 5.6). In addition, an increased amount of the
unknown cyclic product at retention time of 25.22 min was observed.
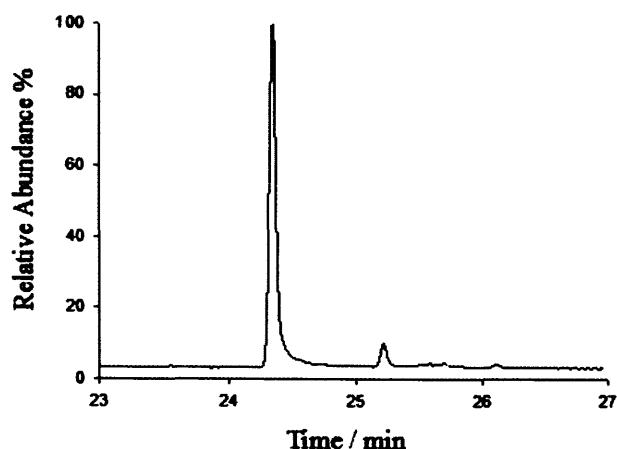
Figure 5.6: *GC trace for incubation of 10 μM EBFSΔ14 with 1 mM FDP in the presence of 5 mM Mg^{2+}.*

An optimised incubation time of 15 min and an enzyme concentration of 0.1 μM were determined to meet the requirements for Michaelis-Menten kinetic assay before the actual run (Figure 5.7). The calculated k_{cat} of EBFSΔ14 (0.015 ± 0.002 s^{-1}) was exactly the same as wild-type EBFS (k_{cat} = 0.015 ± 0.004 s^{-1}). The K_M (3.7 ± 0.6 μM) was also comparable to that of wild-type enzyme (K_M = 6.4 ± 1.7 μM). This indicates that the initial part of the N-terminal region of EBFS has nearly no contribution towards the catalytic activity. However, the minor change in the product profile suggests that the absence of this region may affect binding of the substrate slightly. In addition, it has been proposed that R15 of 5-*epi*-aristolochene synthase plays a role in controlling the open-close state of active site conformation by hydrogen bonding with the C-terminal domain (*91*). The corresponding residue in EBFS is Arg15, which has been mutated to Ala2 in the truncated mutant EBFSΔ14. The fact that EBFSΔ14 does not lose catalytic efficiency rules out the proposed function for R15 in the catalysis of EBFS.
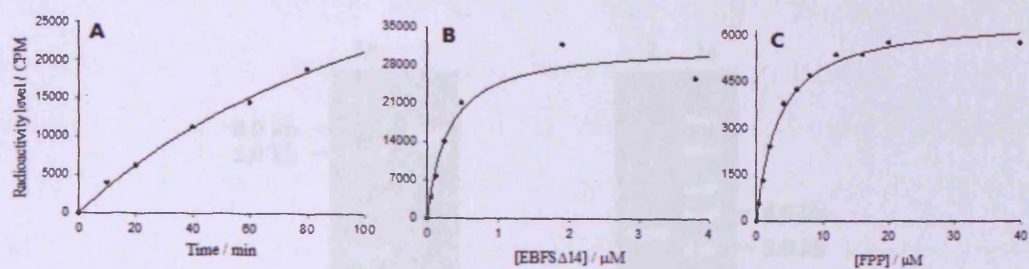
Figure 5.7: *Kinetic analysis of EBFSΔ14.* (A) Time course (0.1 μM enzyme, 10 μM FDP); (B) Plot of radioactivity level against enzyme concentration (10 μM FDP, 15 min incubation time); (C) Michaelis-Menten plot (incubation of 0.1 μM enzyme with radiolabelled substrate for 15 min at 22 °C).

5.2.2 EBFSΔ24

To further define the function of the N-terminal region of EBFS, another truncated mutant (EBFSΔ24) was generated by deletion of the first 24 amino acids from the wild-type sequence. The analogous polypeptide segments of TEAS and DCS were found missing in the crystal structures probably due to the mobility that relates to the function of covering active sites.

This mutant was prepared following a similar method to that used for constructing EBFSΔ14, by introducing an *Nco*I restriction site into the DNA sequence of the wild-type EBFS (Figure 5.8). The mutated plasmid was then treated with *Nco*I and the resulting large DNA fragment was intramolecularly ligated to give the desired plasmid (Figure 5.9).

```
 1  ATGGCTACAAACGGCGTCGTAATTAGTTGCTTAAGGGAAGTAAGGCCACCTATGACGAA
60  GCATGCGCCAAGCATGTGGACTGATACCTTTTCTAACTTTTCTCTTGACGATAAG - 3'
 3'-CGTACGCGGTTGGTACCCCTGACTATGG- 5'
```

Figure 5.8: *The 5' to 3' N-terminal region of EBFS gene sequence (red) and the forward primer (purple, primer sequence is reversed and complementary) that generates NcoI restriction site (underlined) for construction of EBFSΔ24.*
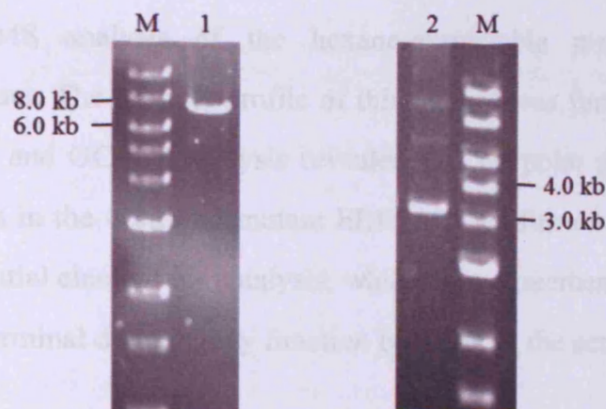
Figure 5.9: 1% *Agarose gel analyses of the EBFSΔ24 construct preparation.*
Lane 1: linearized plasmid that encodes EBFSΔ24 after *Nco*I digestion. Lane
2: plasmid that encodes EBFSΔ24 after ligation. M: 1 kb DNA ladder.

The gene expression, protein overproduction and purification were carried out using
the same methods as wild-type EBFS (Figure 5.10). Typically, 11 mg of EBFSΔ24
was obtained per litre of culture.
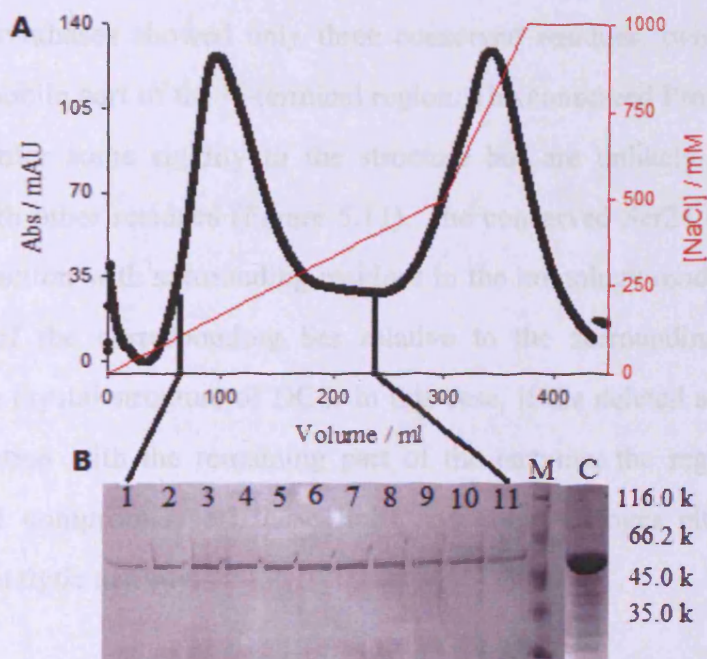
Figure 5.10: *(A) Chromatogram for DEAE purification of EBFSΔ24. (B) 13%
SDS-PAGE analysis.* Lane 1-11: eluent fractions. Lane M: protein marker. Lane C:
concentrated protein.

10 µM of purified EBFSΔ24 was incubated with 1 mM FDP overnight at room temperature. GC-MS analysis of the hexane-extractable products showed no sesquiterpene product. The product profile of this mutant was further investigated by an ether extraction and GC-MS analysis revealed that no polar product was formed either. Activity loss in the truncated mutant EBFSΔ24 confirmed that the N-terminal region was an essential element for catalysis, which is in agreement with the proposed theory that the N-terminal domain may function by capping the active site.

5.3 N-Terminal region chimeras

5.3.1 D24AAE

The truncated mutants established an important role for the N-terminal region in catalysis. However, the question still remains whether the N-terminal region has any contact with the surrounding loops as proposed, or if it only positions above the active site, capping the pocket. As mentioned before, sequence alignment among five plant sesquiterpene synthases showed only three conserved residues, two prolines and a serine, in the mobile part of the N-terminal region. The conserved Pro17 and Pro23 of EBFS may confer some rigidity to the structure but are unlikely to form strong interactions with other residues (Figure 5.11). The conserved Ser24 of EBFS shows no strong interaction with surrounding residues in the homology model and a similar conformation of the corresponding Ser relative to the surrounding residues was observed in the crystal structure of DCS. In this case, if the deleted segment has any specific interaction with the remaining part of the enzyme, the segment swapping mutants should compromise all these links and show changes either in product selectivity or catalytic activity.

```
GAS   1 MAAVEANGTLQANTKTTTEPVRPLANFPPSVWGDRFLSFSLDNK 44
GCS   1     MAASS.....ADKCRPLANFHPSVWGYHFLSYT.HEI 31
TEAS  1      MASAAVANYEEEIVRPVADFSPSLWGDQFLSFSIKNQ 37
DCS   1 MASQVSQMPSSSPLSSNKDEMRPKADFQPSIWGDLFLNCPDKNI 44
EBFS  1     MATNGVVISCLREVRPPMTKHAPSMWTDTFSNFSLDDK 38
                                ▲         ▲
                             EBFSΔ14    EBFSΔ24

D24AAE           1 MEPSSSPLSSNKDEMRPKADFQPSMWTDTFSNFSLDD 37
E24AAD           1 MATNGVVISCLREVRPPMTKHAPSIWGDLFLNCPDKN 37
D24AAE-S24T/W26G 1 MEPSSSPLSSNKDEMRPKADFQPTMGTDTFSNFSLDD 37
E24AAD-S24W      1 MATNGVVISCLREVRPPMTKHAPWIWGDLFLNCPDKN 37
```

Swapped N-terminal region

Figure 5.11: *Amino acid sequence alignment of N-terminal region of five plant sesquiterpene synthases and the N-terminal region for the chimeras.* Conserved residues are coloured. Solid triangles denote cut positions for N-terminal region truncated mutants.

Mutant D24AAE was constructed by replacing the first 24 amino acids of EBFS with residues 1 and 7-30 of DCS (as residues 1-9 of DCS were shown to have very little effect on the enzyme activity (*vide supra*)) (Figure 5.12). The cDNA of D24AAE was created by ligation of the DNA encoding EBFSΔ24 following *Nco*I digestion with a DNA fragment encoding the chosen amino acids from DCS. The smaller DNA fragment was designed to anneal with *Nco*I restriction sites at both ends. To check whether the small DNA fragment was ligated with the insert DNA, a colony PCR method was designed by measuring the DNA length between the T7 promoter and the 586th base pair position using T7 primer and primer EBFS-P586 (Figure 5.13). A blank experiment was carried out using the plasmid of EBFSΔ24 as template DNA.
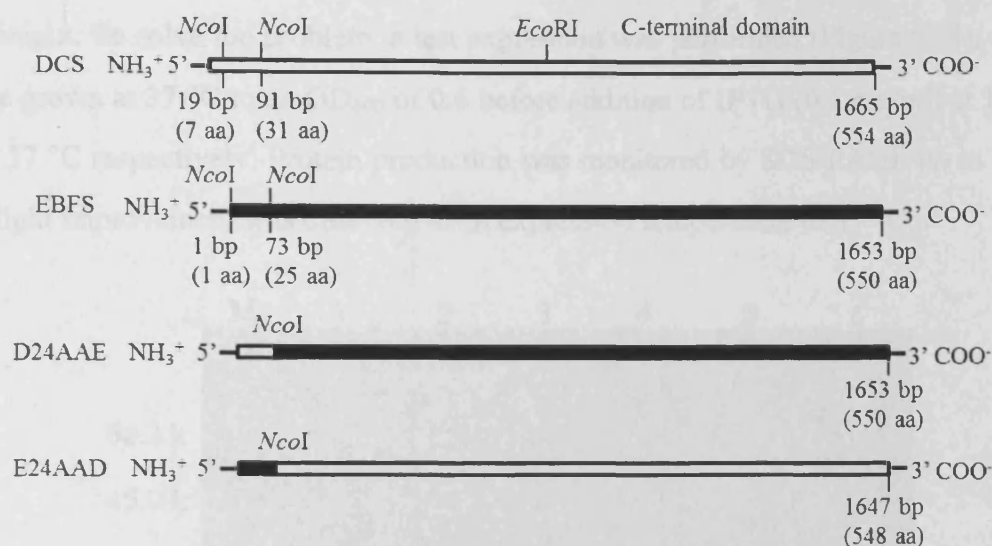
Figure 5.12: *Schematic diagram of the N-terminal region-swapping tests.* Line drawings depict composite diagrams for wild-type DCS (coloured white), wild-type EBFS (coloured black), D24AAE and E24AAD genes.

Figure 5.13: *1% Agarose gel analyses of colony PCR products.* Level 0: parent vector (i.e. encoding EBFSΔ24) with no oligonucleotide pair fragment ligated. Level 1: parent vector with one oligonucleotide pair fragment ligated. Level 2: parent vector with two oligonucleotide pair fragments ligated. Lane B: PCR product using only parent vector as DNA template.

Since both ends of the designed primer fragment had the same *Nco*I restriction site, the correct orientation of the construct was confirmed by DNA sequencing. As a result of the *Nco*I restriction sites introduced for construction of the chimera, residues 24 and 26 of D24AAE were initially Thr and Gly respectively rather than Ser and Trp in the wild type. Mutant D24AAE-S24T/W26G was overproduced the same way as WT-EBFS. However, very little protein was obtained after incubating cells at 16 °C

195

overnight. To solve the problem, a test expression was performed (Figure 5.14). Cells were grown at 37 °C to an OD_{600} of 0.6 before addition of IPTG (0.1 mg/ml) at 16 °C and 37 °C respectively. Protein production was monitored by SDS-PAGE up to 12 h. A slight improvement was observed at an expression temperature of 37 °C.
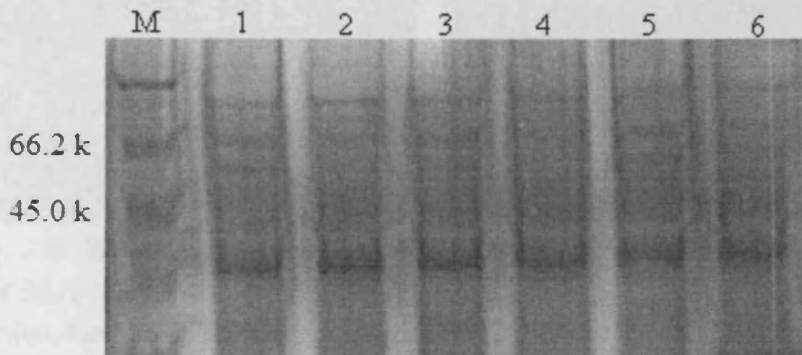
Figure 5.14: *13% SDS polyacrylamide gel of test expression under different temperature.* Lane 1: 37 °C for 1 hr. Lane 2: 16 °C for 2 hr. Lane 3: 37 °C for 2 hr. Lane 4: 16 °C for 6 hr. Lane 5: 37 °C for 4 hr. Lane 6: 16 °C for 12 hr. Lane M: protein maker.

In addition, an increased amount of IPTG (1 mg/ml) was applied to initiate expression. No obvious band around MW 63830 was observed (Figure 5.15).
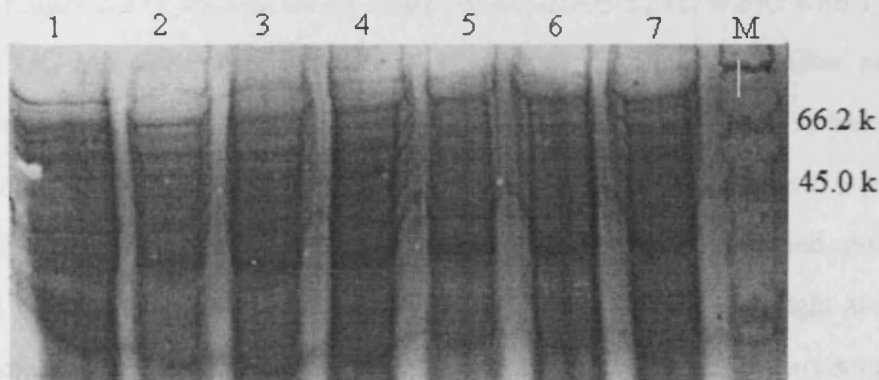
Figure 5.15: *13% SDS polyacrylamide gel of D24AAE-S24T/W26G test expression induced by IPTG (1 mg/ml) under different temperature.* Lane 1: before induction. Lane 2: 37 °C for 1 hr. Lane 3: 16 °C for 2 hr. Lane 4: 37 °C for 2 hr. Lane 5: 16 °C for 6 hr. Lane 6: 37 °C for 4 hr. Lane 7: 16 °C for 12 hr. Lane M: protein maker.

Further expression testing was performed by employing BL21-CodonPlus(DE3)-RIL cells and BL21-Star(DE3) cells. Expression was induced by addition of IPTG (0.1 mg/ml) at 16 or 37 °C respectively. BL21-CodonPlus(DE3)-RIL cells yielded the best

protein production (Figure 5.16). A good level of targeted protein was obtained after 4 hr at 37 °C
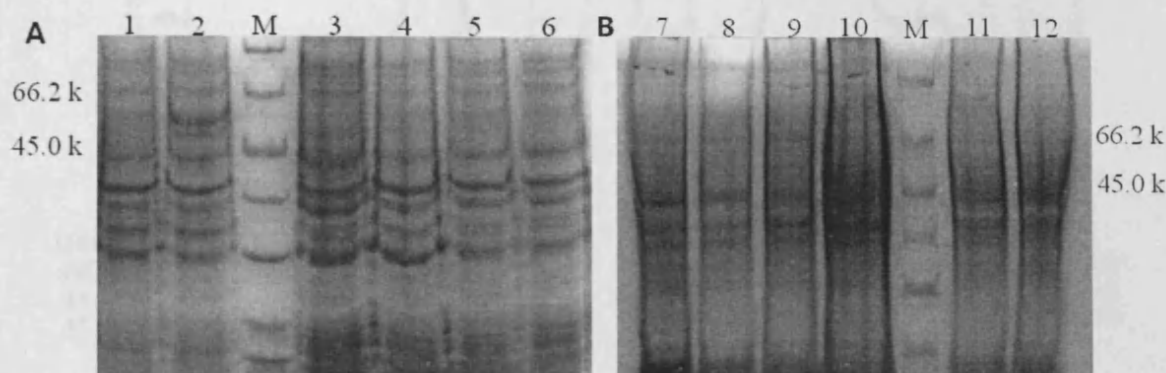
Figure 5.16: *13% SDS polyacrylamide gels of test expression with BL21-CodonPlus(DE3)-RIL cells (A) and BL21-Star(DE3) cells (B).* Lane 1: 37 °C for 2 hr. Lane 2: 37 °C for 4 hr. Lane 3: before induction. Lane 4: 16 °C for 2 hr. Lane 5: 16 °C for 6 hr. Lane 6: 16 °C for 12 hr. Lane 7: before induction. Lane 8: 16 °C for 2 hr. Lane 9: 16 °C for 6 hr. Lane 10: 16 °C for 12 hr. Lane 11: 37 °C for 2 hr. Lane 12: 37 °C for 4 hr. Lane M: protein maker.

The large-scale expression was carried out by growing cells to an OD_{600} of 0.6 at 37 °C. Expression was initiated by addition of IPTG (0.1 mg/ml) and cells were incubated for another 5 hr. The overexpressed protein was purified in the same way as wild-type EBFS and typically 10 mg of purified protein was obtained per litre of culture (Figure 5.17). Incubation of 10 μM of D24AAE-S24T/W26G with 1 mM FDP gave no GC-MS detectable sesquiterpene product (Figure 5.18). After performing site-directed mutagenesis to return position 24 and 26 the wild-type residues and the recombinant protein was produced and purified by the same manner as that of D24AAE-S24T/W26G (Figure 5.17). Typically, 20 mg of purified protein was obtained per litre of culture. Incubation of D24AAE with FDP overnight and analysis of the hexane extractable products by GC-MS showed that (*E*)-β-farnesene was the only sesquiterpene product (Figure 5.18).
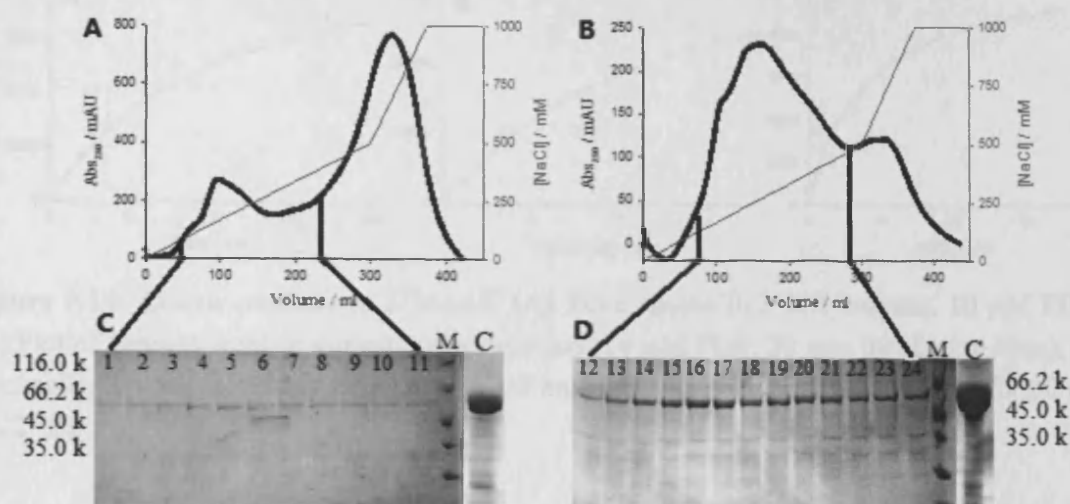
Figure 5.17: *Chromatograms for DEAE purification of D24AAE-S24T/W26G (A) and D24AAE (B) and 13% SDS-PAGE analyses of each purification (C) and (D).* Lane 1-11: eluent fractions for D24AAE-S24T/W26G. Lane 12-24: eluent fractions for D24AAE. Lane M: protein marker. Lane C: concentrated protein.
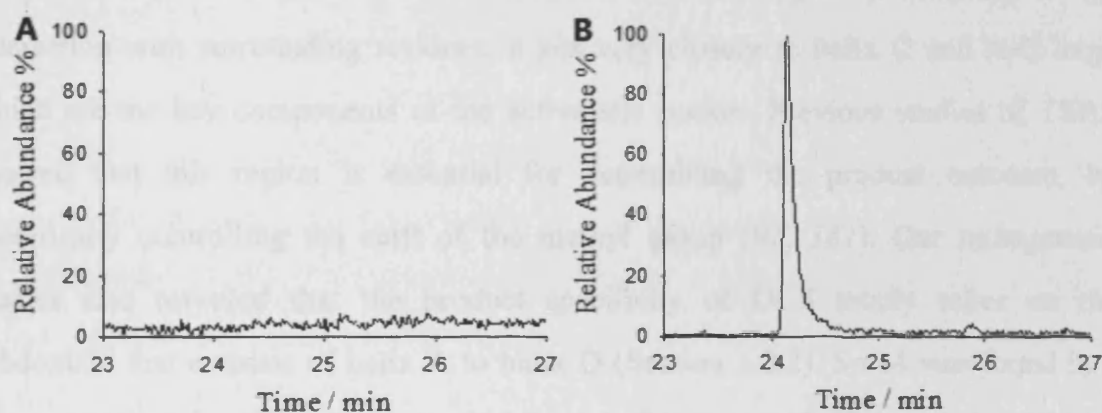
Figure 5.18: *GC traces for hexane extractable product profiles from an overnight incubation of 10 μM D24AAE-S24T/W26G (A) and 10 μM D24AAE (B) with 1 mM FDP.*

An optimised incubation time of 30 min and an enzyme concentration of 0.2 μM were determined to meet the requirements for Michaelis-Menten kinetic assay before the actual run (Figure 5.19). The k_{cat} and K_M were determined to be 0.005 ± 0.00005 s^{-1} and 7.5 ± 1.1 μM respectively. A slight reduction of turnover rate was observed, while generally the kinetic efficiency was comparable to the wild-type enzyme, suggesting that no strong interactions between the loop region (M1-S24) and C-terminal domain are existed or they are unimportant for catalysis.
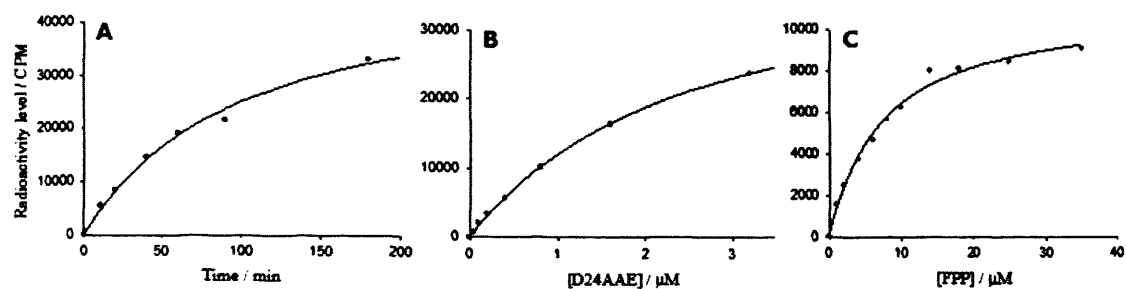
Figure 5.19: *Kinetic analysis of D24AAE.* (A) Time course (0.2 μM enzyme, 10 μM FDP); (B) Plot of velocity against enzyme concentration (10 μM FDP, 30 min incubation time); (C) Michaelis-Menten plot (incubation of 0.2 μM enzyme with radiolabelled substrate for 30 min at 22 °C).

Nevertheless, the conserved Ser and Trp residues are essential for the catalysis since the double mutant D24AAE-S24T/W26G lost catalytic ability. A closer look into the homology model of EBFS, reveals that the conserved Ser residue is facing into the C-terminal domain (Figure 5.20). Although it does not appear to have any strong interaction with surrounding residues, it sits very closely to helix C and A-C loop, which are the key components of the active site pocket. Previous studies of TEAS showed that this region is essential for determining the product outcome, by specifically controlling the shift of the methyl group (*91, 161*). Our mutagenesis studies also revealed that the product specificity of DCS totally relies on the subdomain that consists of helix A to helix D (Section 3.2.2). Ser24 was found in a distance of 3~4 Å with residues Asp265 from A-C loop and V268 from helix C in the homology model of EBFS, which suggested van der Waals interactions. Changes made to the Ser could introduce steric clash with A-C loop and helix C, leading to a change in volume of the active site cavity. Such change usually has a negative influence on catalytic activity, which was displayed before by the results of active site mutations of EBFS. The conserved Trp residue was also found to have a possible interaction with helix C, which was investigated by additional mutagenesis studies (Section 5.4.1).
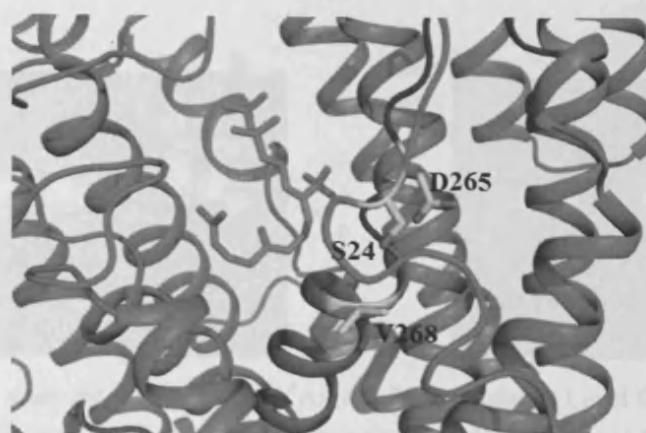
Figure 5.20: *Cartoon representation of EBFS showing the position of S24 relative to A-C loop and helix C in the homology model of EBFS.* Bound FDP is shown as cyan sticks. A-C loop is coloured blue and helix C is coloured magenta. S24, D265 and V268 are shown as yellow sticks.

5.3.2 E24AAD

The equivalent experiment of mutant D24AAE was also carried out based on DCS. The mutant E24AAD was prepared by replacing the first 30 amino acids of DCS with the first 24 amino acids of EBFS (Figure 5.12). The cDNA of E24AAD was prepared by ligation of the DNA encoding DCSΔ30 (constructed by Dr. Veronica Gonzalez in our group) following *Nco*I digestion with a DNA fragment encoding the first 24 amino acids of EBFS. The insert DNA fragment was designed to anneal with the *Nco*I restriction sites at both ends. DCSΔ30 was one of the N-terminal region truncated mutant based on DCS, of which the first 30 amino acids were eliminated. Formation of the targeted construct was monitored by colony PCR using forward T7 primer and reverse DCS-P557 primer based on the same method used for D24AAE (Figure 5.21). A blank experiment was carried out using the plasmid of DCSΔ30 as template DNA. Plasmids with the expected DNA length were sequenced and the confirmed construct was used for the following experiments.
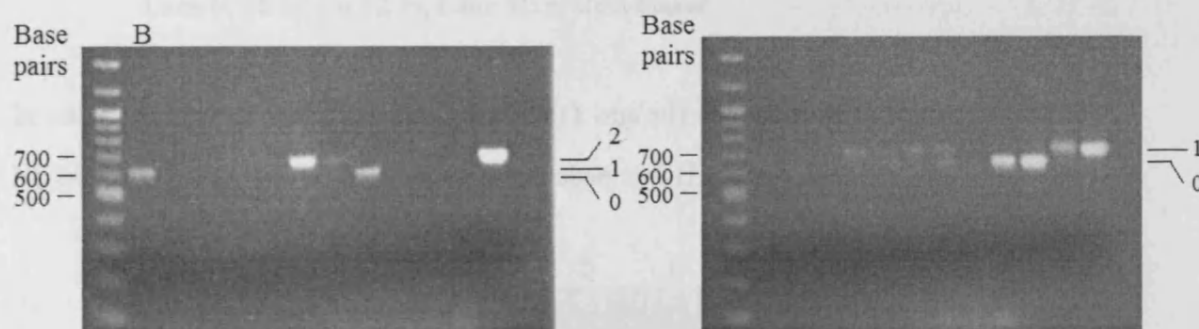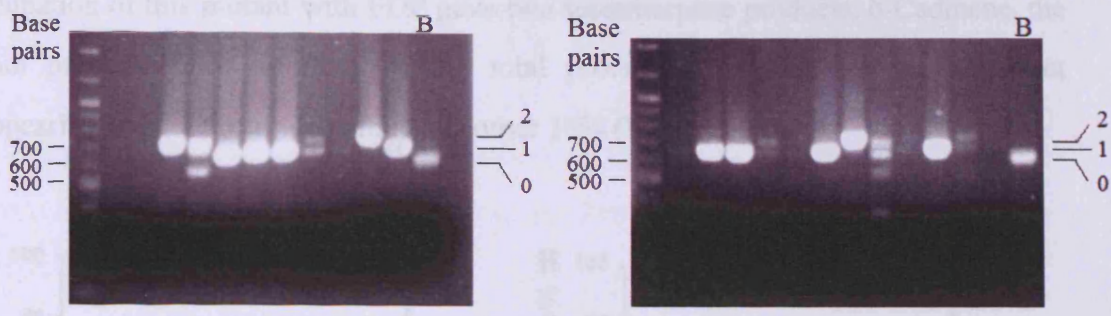
Figure 5.21: *1% Agarose gel analyses of colony PCR products.* Level 0: parent vector (*i.e.* encoding DCSΔ30) with no oligonucleotide pair fragment ligated. Level 1: parent vector with one oligonucleotide pair fragment ligated. Level 2: parent vector with two oligonucleotide pair fragments ligated. Lane B: PCR product using only parent vector as DNA template.

To introduce *Nco*I restriction sites for the swapping strategy, Ser30 of DCS was mutated to Trp; The S30 of DCS became S24 in the chimeric enzyme E24AAD since the first 6 amino acids of DCS were omitted for the swapping strategy. The tryptophan mutant, E24AAD-S24W, was produced and purified following the same protocol as the wild type DCS (Figure 5.22). Typically, 23 mg of purified protein was obtained per litre of culture.

Figure 5.22: *(A) Chromatogram for DEAE purification of E24AAD-S24W. (B) 13% SDS-PAGE analysis.* Lane 1-10: eluent fractions. Lane M: protein marker. Lane C: concentrated protein.

Incubation of this mutant with FDP gave two sesquiterpene products. δ-Cadinene, the main product takes up 90% of the total production and an unknown product (appearing at 24.87 min) contributes another 10% (Figure 5.23).
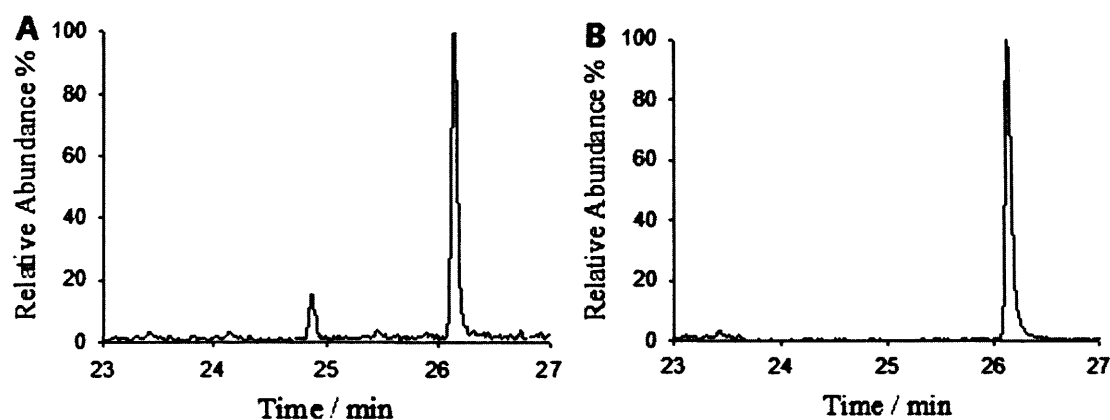
Figure 5.23: *GC traces for hexane extractable product profile from overnight incubations of 10 μM E24AAD-S24W (A) and 10 μM E24AAD (B) with 1 mM FDP.*

In addition to the change of product profile, the kinetic parameters were greatly affected. A K_M of 17.7 ± 3.9 μM and a k_{cat} of (1.4 ± 0.1) × 10^{-4} s^{-1} were obtained from kinetic analysis (Figure 5.24). Although the K_M was not affected much, the k_{cat} was reduced by about 100 fold.

Figure 5.24: *Kinetic analysis of E24AAD-S24W.* (A) Time course (0.34 μM enzyme, 10 μM FDP); (B) Plot of radioactivity level against enzyme concentration (10 μM FDP, 15 min incubation time); (C) Michaelis-Menten plot (incubation of 1.36 μM enzyme with radiolabelled substrate for 45 min at 22 °C).

The S24W mutation was corrected by site-directed mutagenesis and the gene of E24AAD was expressed following transformation into BL21-CodonPlus(DE3)-RIL

cells. Expression and purification of E24AAD were carried out by the same method as wild-type DCS (Figure 5.25). Typically, 30 mg of purified protein was obtained per litre of culture. In agreement with the previous study of chimera D24AAE, the mutant E24AAD maintained high-fidelity as parental enzyme DCS by converting FDP into δ-cadinene as the only product (Figure 5.23) and reserved similar catalytic activity at the same time. Kinetic analysis showed a K_M of 12.5 ± 1.3 μM and a k_{cat} of 0.031 ± 0.003 s^{-1} (Figure 5.26).
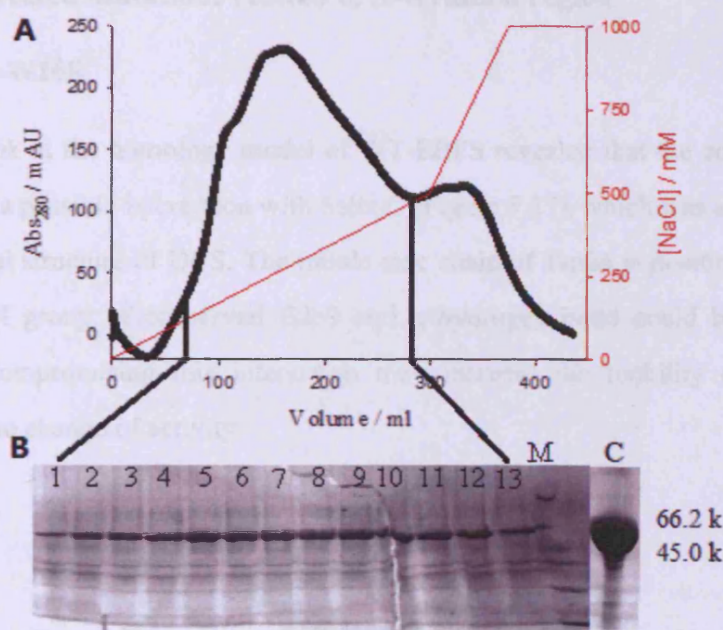
Figure 5.25: *(A) Chromatogram for DEAE purification of E24AAD. (B) 13% SDS-PAGE analysis.* Lane 1-13: eluent fractions. Lane M: protein marker. Lane C: concentrated protein.
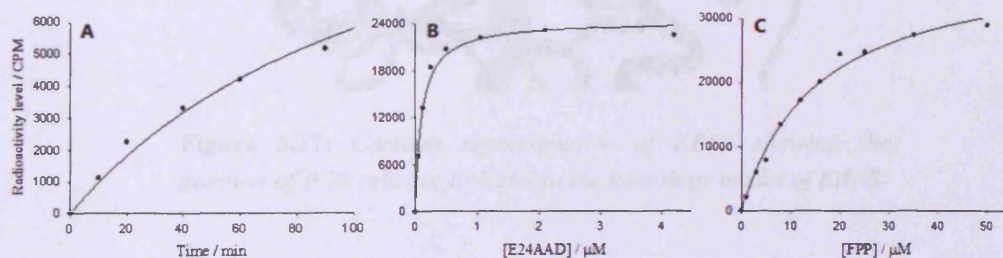
Figure 5.26: *Kinetic analysis of E24AAD.* (A) Time course (0.34 μM enzyme, 10 μM FDP); (B) Plot of velocity against enzyme concentration (10 μM FDP, 15 min incubation time); (C) Michaelis-Menten plot (incubation of 0.1 μM enzyme with radiolabelled substrate for 30 min at 22 °C).

203

A single mutation S24W in the chimera E24AAD changes both the product profile and the catalytic activity, supporting the predicted role of the conserved Ser residue in maintaining the precise conformation of helix C. The swapped N-terminal reigon does not affect the catalysis, supporting the assumption that the region before the conserved Ser has no strong interaction with the rest of the protein. The main function of this region appears to be sequestering the active site during catalysis.

5.4 Site-directed mutations related to N-terminal region

5.4.1 EBFS-W26E

A closer look at the homology model of WT-EBFS revealed that the conserved Trp residue had a possible interaction with helix C (Figure 5.27), which was also observed in the crystal structure of DCS. The indole side chain of Trp26 is positioned close to the carboxyl group of conserved E269 and a hydrogen bond could be formed in between. Compromising this interaction may increase the mobility of the loop, leading to the change of activity.
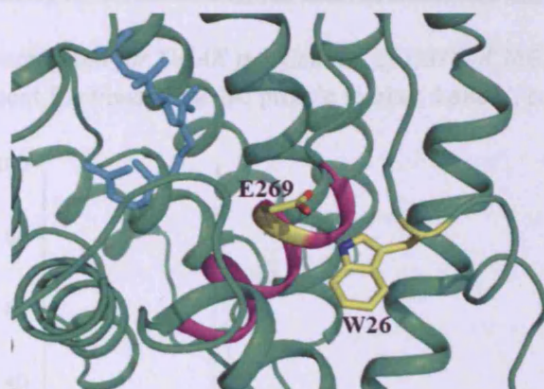
Figure 5.27: *Cartoon representation of EBFS showing the position of W26 relative to E269 in the homology model of EBFS.*

Mutation EBFS-W26E was designed to disrupt the proposed hydrogen bonding and instead introduce a repulsive interaction between the N-terminal region and helix C, to investigate the effect of this specific structure on enzyme selectivity or catalytic

efficiency. To make the mutant, the native codon TGG was replaced by GAG using site-directed mutagenesis. EBFS-W26E was produced and purified in the same way as the wild type EBFS (Figure 5.28). Typically, 30 mg of purified protein was obtained per litre of culture. GC-MS analysis showed a single peak for (E)-β-farnesene following an overnight incubation of EBFS-W26E with FDP (Figure 5.29).
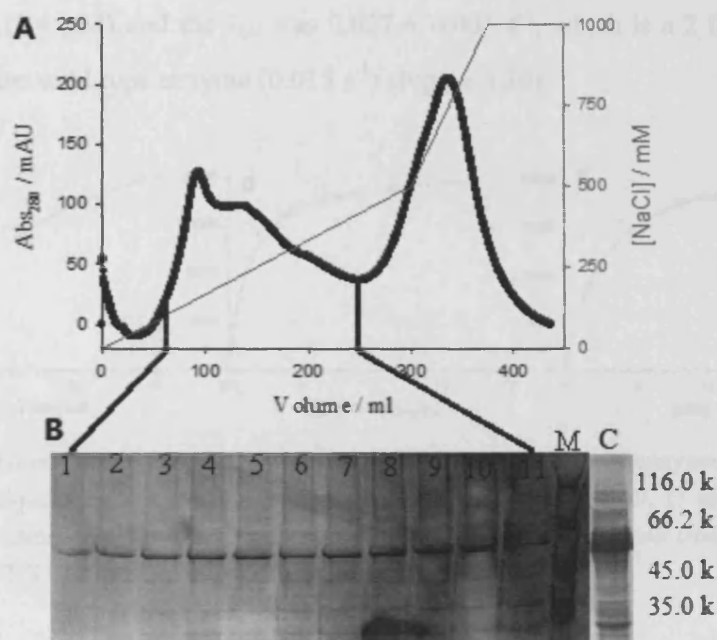
Figure 5.28: *(A) Chromatogram for DEAE purification of EBFS-W26E. (B) 13% SDS-PAGE analysis.* Lane 1-11: eluent fractions. Lane M: protein marker. Lane C: concentrated protein.
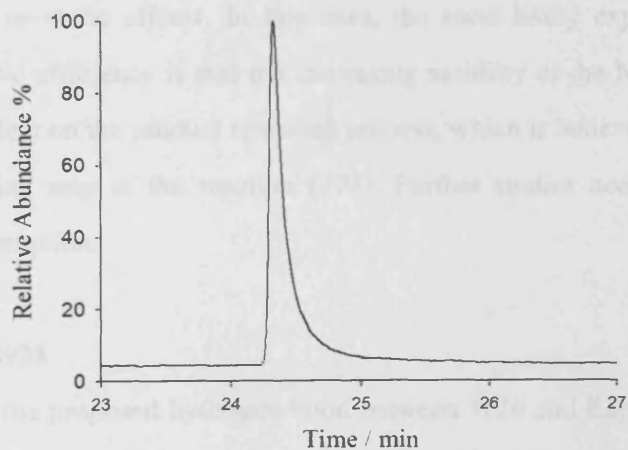
Figure 5.29: *GC trace for incubation of hexane extractable product profile from an overnight incubation of 10 μM EBFS-W26E with 1 mM FDP.*

This mutation was predicted to increase the mobility of the N-terminal region, which would possibly lead to a reduction of enzyme activity since all the previous studies defined that the only function of the N-terminal region is to cap the active site. However, EBFS-W26E was found to be more active than wild-type EBFS. The K_M of this mutant was determined to be 2.9 ± 0.6 μM, which is about 2 times lower than the wild enzyme (6.4 μM) and the k_{cat} was 0.027 ± 0.001 s^{-1}, which is a 2 fold increase compared to the wild type enzyme (0.015 s^{-1}) (Figure 5.30).
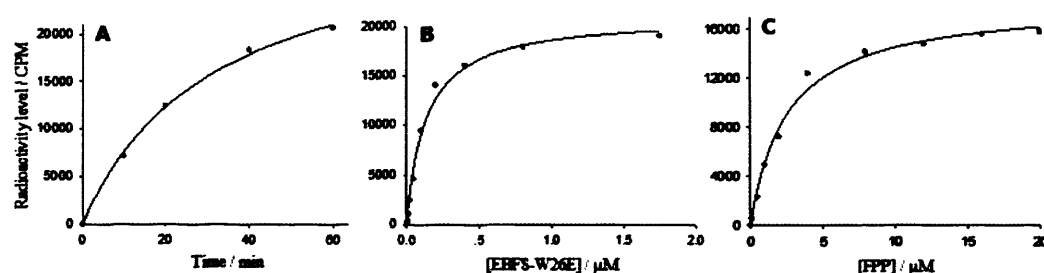
Figure 5.30: *Kinetic analysis of EBFS-W26E.* (A) Time course (0.1 μM enzyme, 4 μM FDP); (B) Plot of radioactivity level against enzyme concentration (4 μM FDP, 15 min incubation time); (C) Michaelis-Menten plot (incubation of 0.1 μM enzyme with radiolabelled substrate for 15 min at 22 °C).

According to the homology model of EBFS, this conserved Trp is located outside the C-terminal domain and is therefore unable to affect the actual reaction by contributing either chemical or steric effects. In this case, the most likely explanation for the increased catalytic efficiency is that the increasing mobility of the N-terminal region has a positive effect on the product releasing process, which is believed to be the most likely rate-limiting step of the reaction (*175*). Further studies need to be done to confirm this assumption.

5.4.2 EBFS-E269M

To compromise the proposed hydrogen bond between W26 and E269, the conserved Glu residue was mutated to Met, which has a similar side chain length. EBFS-E269M was prepared by replacing the native codon GAG with ATG. The mutant was produced and purified in the same manner as the wild-type EBFS (Figure 5.31).

Typically, 9 mg of purified protein was obtained per litre of culture. The EBFS-E269M mutant maintains the high fidelity since (*E*)-β-farnesene was the only hexane extractable sesquiterpene product after an overnight incubation with FDP (Figure 5.32).
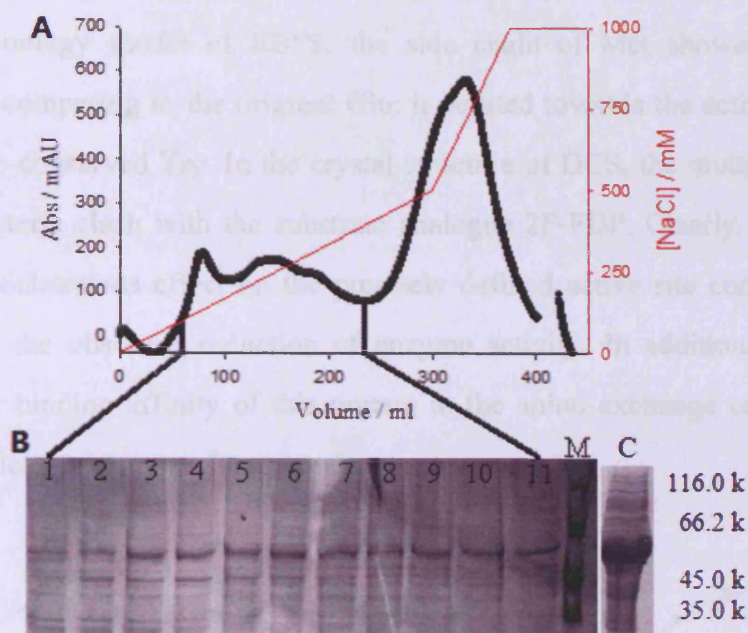
Figure 5.31: *(A) Chromatogram for DEAE purification of EBFS-E269M. (B) 13% SDS-PAGE analysis.* Lane 1-11: eluent fractions. Lane M: protein marker. Lane C: concentrated protein.
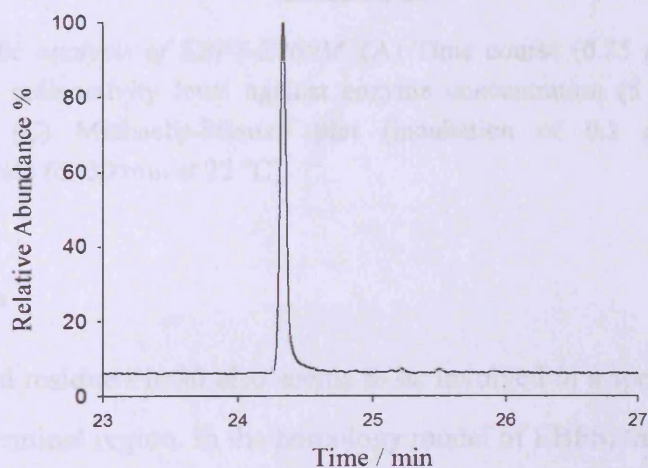
Figure 5.32: *GC trace for incubation of hexane extractable product profile from an overnight incubation of 10 µM EBFS-E269M with 1 mM FDP.*

Compared to the wild type enzyme, a 3 fold lower K_M (1.8 ± 0.4 μM) was seen from kinetic assay of EBFS-E269M, whereas the k_{cat} (0.00011 ± 0.00001 s^{-1}) was reduced by roughly 140 fold (Figure 5.33). The reduction in the catalytic activity of the mutant EBFS-E269M may be due to the change of the active site contour. According to the modelling studies by mutating conserved Asp into Met in both crystal structure of DCS and homology model of EBFS, the side chain of Met showed a different conformation comparing to the original Glu; it pointed towards the active site centre rather than the conserved Trp. In the crystal structure of DCS, the mutated Met even introduced a steric clash with the substrate analogue 2F-FDP. Clearly, the mutation has caused a deleterious effect on the precisely defined active site contour and this could explain the observed reduction of enzyme activity. In addition, SDS-PAGE indicates poor binding affinity of this mutant to the anion exchange column, which could also indicate a change of the protein structure.
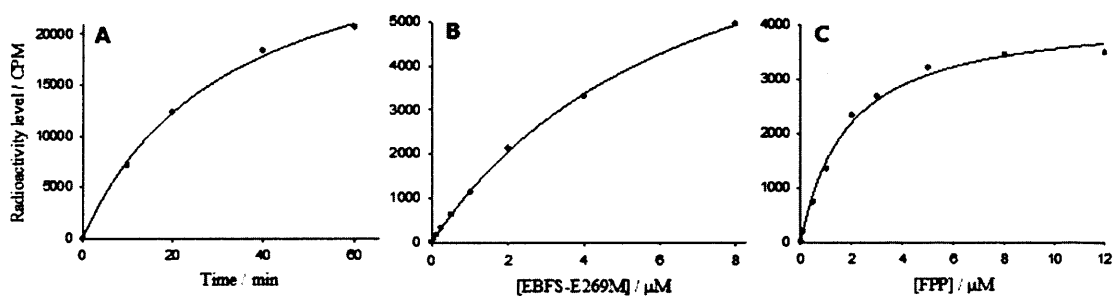
Figure 5.33: *Kinetic analysis of EBFS-E269M.* (A) Time course (0.25 μM enzyme, 5 μM FDP); (B) Plot of radioactivity level against enzyme concentration (5 μM FDP, 30 min incubation time); (C) Michaelis-Menten plot (incubation of 0.1 μM enzyme with radiolabelled substrate for 30 min at 22 °C).

5.4.3 EBFS-F30A

Another conserved residue Phe30 also seems to be involved in a special structure that stabilizes the N-terminal region. In the homology model of EBFS, this residue appears to be trapped in a pocket formed by residues from both the C-terminal domain and N-terminal loop and deletion of the phenyl group from the pocket could possibly increase the mobility of the N-terminal region (Figure 5.34). By introducing better mobility to the N-termini, an improvement of catalytic efficiency was expected for

this mutant as observed for the mutant EBFS-W26E.
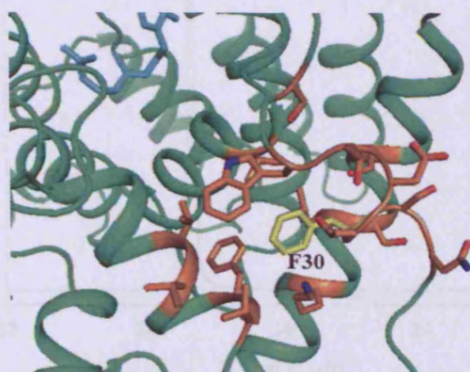
Figure 5.34: *Cartoon representation of EBFS showing the position of F30 and surrounding residues in the homology model of EBFS.*

EBFS-F30A was prepared by mutating the native codon TTT to GCT. The mutant was produced and purified by the same procedure used for the wild-type EBFS (Figure 5.35). Typically, 21 mg of purified protein was obtained per litre of culture. Incubation of purified EBFS-F30A with FDP gave (E)-β-farnesene as the only hexane extractable product (Figure 5.36).
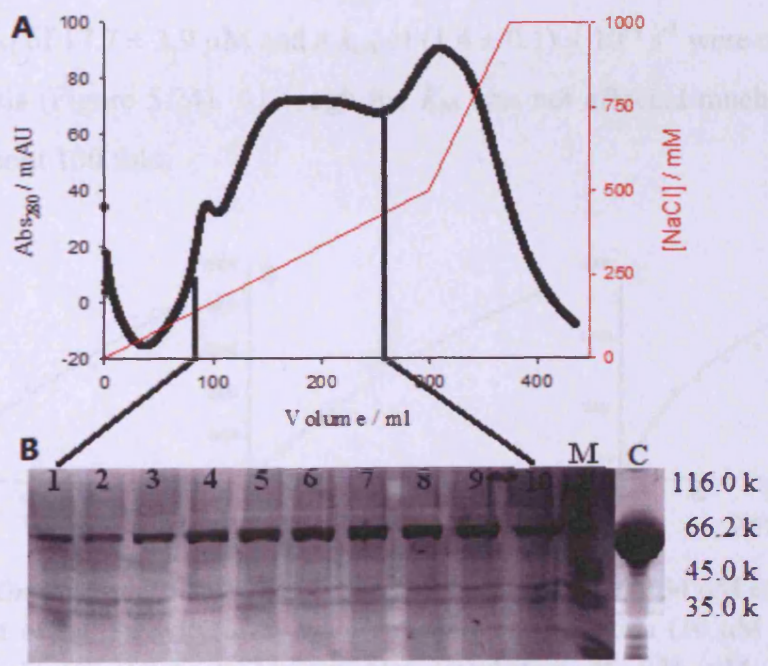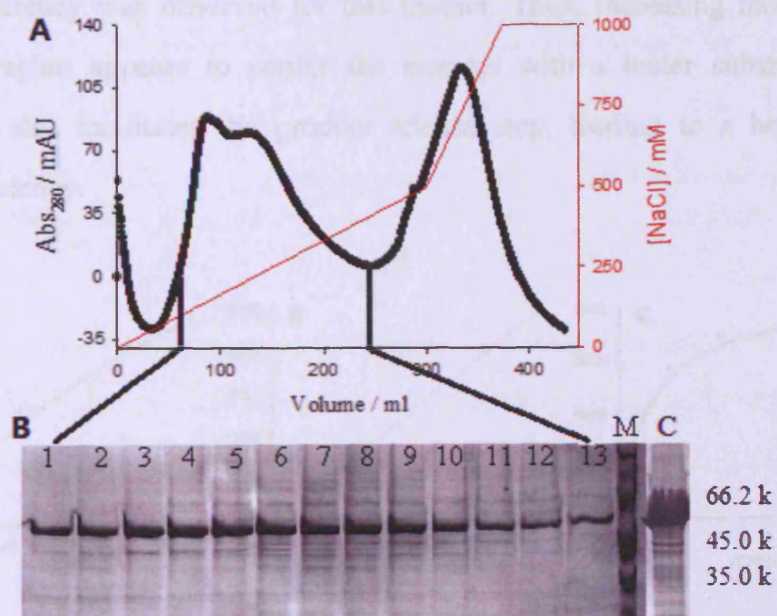

Figure 5.35: *(A) Chromatogram for DEAE purification of EBFS-F30A. (B) 13% SDS-PAGE analysis.* Lane 1-13: eluent fractions. Lane M: protein marker. Lane C: concentrated protein.
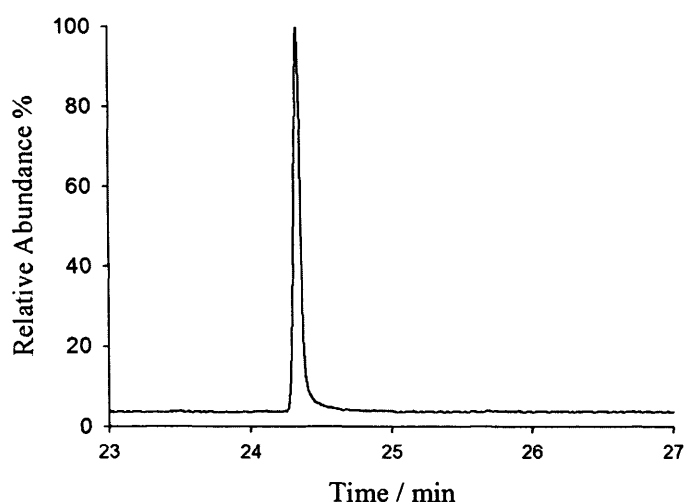
Figure 5.36: *GC trace for incubation of hexane extractable product profile from an overnight incubation of 10 μM EBFS-F30A with 1 mM FDP.*

Kinetic assay of EBFS-F30A showed an improvement in both substrate binding affinity and catalytic turnover. The K_M (1.5 ± 0.2 μM) was about 4 times lower than the wild type EBFS and the k_{cat} (0.044 ± 0.002 s^{-1}) increased by about 3-fold compared to the wild type enzyme (Figure 5.37); therefore a 12-fold increase in catalytic efficiency was observed for this mutant. Thus, increasing mobility of the N-terminal region appears to confer the enzyme with a better substrate binding affinity and also facilitates the product release step, leading to a better enzyme catalytic efficiency.
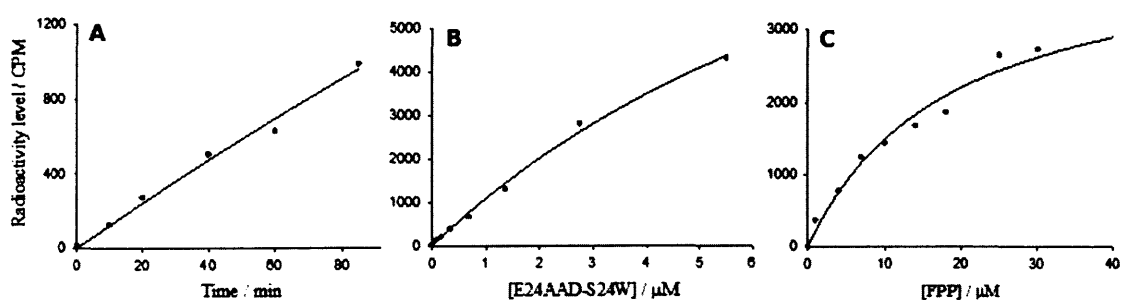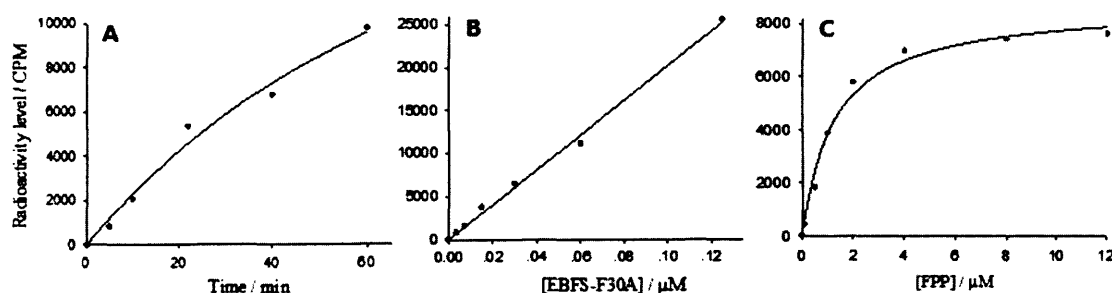

Figure 5.37: *Kinetic analysis of EBFS-F30A.* (A) Time course (0.015 μM enzyme, 5 μM FDP). (B) Plot of radioactivity level against enzyme concentration (5 μM FDP, 15 min incubation time). (C) Michaelis-Menten plot (incubation of 0.025 μM enzyme with radiolabelled substrate for 15 min at 22 °C).

210

5.4.4 DCS-F36A

The mutants EBFS-W26E, EBFS-269M and EBFS-F30A revealed improved enzyme catalytic efficiency, with EBFS-F30A showing the biggest improvement. The corresponding mutation DCS-F36A was carried out to investigate if this pattern can be also applied to DCS. The mutant was made by replacing the native codon TTC with GCC. DCS-F36A was produced and purified by Dr Veronica Gonzalez. When DCS-F36A was incubated with FDP overnight, the only hexane extractable product determined by GC-MS was δ-cadinene (Figure 5.38).
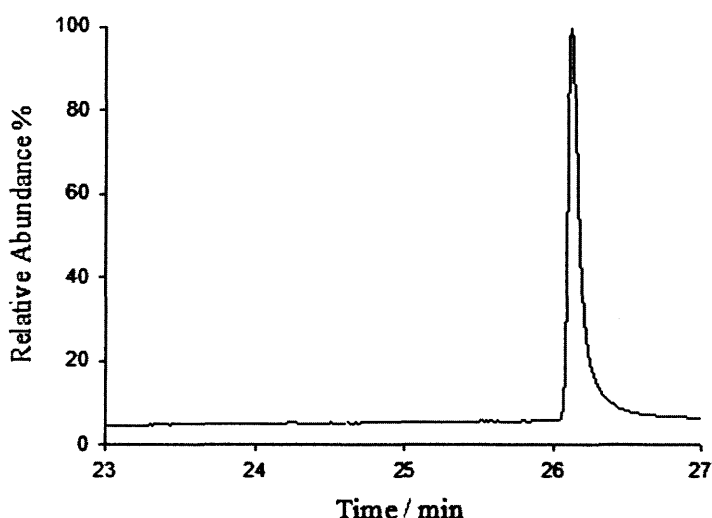
Figure 5.38: *GC trace for incubation of hexane extractable product profile from an overnight incubation of 10 μM DCS-F36A with 1 mM FDP.*

In agreement with previous results, improvement was found for both substrate binding affinity and catalytic turnover rate. The K_M value was found to be 1.9 ± 0.3 μM, which is about 5 times lower than the value of the wild type enzyme. The k_{cat} (0.043 ± 0.012 s^{-1}) increased 3-fold compared to the wild type enzyme (Figure 5.39). Both changes lead to an overall 15 times increase in catalytic efficiency. This mutant again supports the hypothesis that the N-terminal region could play an important role in regulating substrate binding and product release.
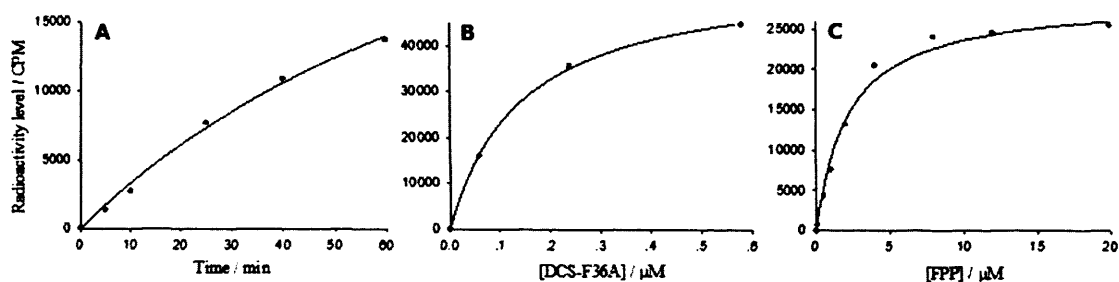
Figure 5.39: *Kinetic analysis of DCS-F36A.* (A) Time course (0.06 μM enzyme, 5 μM FDP); (B) Plot of radioactivity level against enzyme concentration (5 μM FDP, 15 min incubation time); (C) Michaelis-Menten plot (incubation of 0.06 μM enzyme with radiolabelled substrate for 15 min at 22 °C).

5.5 Conclusions

Plant sesquiterpene synthases have variable length N-terminal region and little amino acid sequence similarity before the first conserved Pro. Therefore, this part of the segment might contribute very little to the catalysis. This is consistent with the result for the truncated mutant EBFSΔ14, which showed very similar catalytic activity to its wild-type counterpart. Sequence similarity and identity appear after Pro17 (EBFS) and loss of residues 15-24 (EBFS) abolishes enzyme activity, suggesting that this part of the segment may have an important function causing it to be conserved among plant sesquiterpene synthases. In addition, the fact that EBFSΔ14 does not lose any catalytic efficiency rules out the proposed function for Arg15 of wild-type EBFS in closing the active site conformation.

Results from the chimeric enzymes indicate that the interactions between the N-terminal region and the C-terminal domain mainly start from the conserved Ser. The region before this residue is needed to cap the active site pocket but either no strong interactions exist between this region and the C-terminal domain or they are unimportant for catalysis.

The release of sesquiterpene product is believed to be the most likely rate-limiting step of the reaction (*175*). Site-directed mutagenesis studies toward the conserved Trp

and Phe residues suggest that better mobility of the N-terminal region could facilitate this step. In addition, these mutations may increase the hydrophilicity around the active site, which could also accelerate the departure of the hydrophobic product. Notably, while the regions from Met1 to Pro29 and Lys42 to Asp45 are missing from the crystal structure of DCS, residues 30-41 – *i.e.* around the proposed stabilising regions – are well defined. This observation further supports the role of the conserved Trp and Phe residues in stabilizing the N-terminal region.

Chapter 6: General Conclusions and Future Work

6.1 DCS

The catalytic specificity and activity of wild-type DCS (*Gossypium arboretum*) were in agreement to the previously published results (*100*). The secondary structure of DCS was determined by CD spectroscopy and agreed well with the proposed overall α-helical fold. The role of the aspartate-rich DDXXD motif was confirmed by alanine substitutions. The first and the last aspartate residues were shown to have potential interaction with magnesium ions, which is consistent with the observation from the crystal structure of DCS.

Despite reports of how individual or groups of amino acid residues contribute to reaction cascades of sesquiterpene synthases, several of these studies have actually reported on the premature termination of multistep reactions and the release of neutral reaction intermediates. In contrast, site-directed mutagenesis experiments revealed a different circumstance for DCS. Modifications by substitution of residues from both first and second tier of the active site did not bring any functional divergence. The enzymatic formation of δ-cadinene form FDP has been showed to involve a conversion of C2,C3-double bond geometry and a 1,6-ring closure. At the meantime, DCS can also catalyse 1,10 or 1,11 cyclisation of some FDP analogues with preserving the original C2,C3-double bond geometry. The mechanistic versatility of DCS may indicate a relatively flexible active site contour and the overall function of this active site may not be easily compromised by a few amino acid substitutions. Or, in another case, an unusual substructural component would determine the product specificity of DCS.

Domain-swapping experiments, based on the conserved exon organization among plant sesquiterpene synthase genes, have illustrated that the N-terminal domain and the subdomains that are encoded by exons 1, 2, 3, 5, 6 and 7 may have very little contribution towards the product specificity. However, the N-terminal combines with part of the exon 4 encoded C-terminal to form an alternative subdomain to the usual

sesquiterpene catalytic domain. This subdomain is also able to convert FDP to δ-cadinene but with lower activity. These findings suggest that the product specificity of DCS is most likely determined by the subdomain encoded by exon 4, which is similar to the observation for TEAS. In addition, it seems that the product specificity of DCS totally relies on this subdomain since no other sesquiterpene product was formed by hybrids. However, similar hybrids for TEAS could generate alternative products determined by the vetispiradiene synthase. Additional domain swapping experiments need to be done to verify the function of this assumption.

The importance of this subdomain was also illustrated by the mutant E24AAD-S24W. The mutation S24W could possibly alter the conformation of A-C loop and helix C that belong to the subdomain encoded by exon 4. This modification led to a changed product profile by showing an alternative sesquiterpene product compared with the parent counterpart (E24AAD). In addition, the catalytic activity of this mutant showed a significant reduction compared to the parent counterpart.

6.2 EBFS

Wild-type EBFS (*Mentha x piperita*) has been fully purified and determined to be a monomer in MOPS buffer (pH 7.2). An overall α-helical fold was confirmed by CD spectroscopy. In contrast to previously reports (*115*), the product distribution was defined as (*E*)-β-farnesene (95%), (*Z*)-β-farnesene (1.5%), (3*Z*,6*E*)-α-farnesene (1.3%), (3*E*,6*E*)-α-farnesene (0.2%) and another 5 unknown cyclic sesquiterpene products (<2% in total) when Mg^{2+} was employed as the cofactor.

A new mechanism was proposed for the enzyme by the study of substrate analogue 2F-FDP. It was believed that the enzymatic formation of (*E*)-β-farnesene went through the ionization of *trans-trans*-farnesyl diphosphate to the corresponding transoid allylic cation (*167*). In the case, an E1 elimination was proposed according to the ionization theory. However, the fact that this enzyme can catalyze the conversion

of 2F-FDP with a relatively good activity indicates that the proposed carbocation intermediate is unlikely to be formed in the process since the C2 fluorine can greatly inhibit formation of this intermediate through an inductive electron withdrawing effect. In other words, it seems the enzymatic reaction catalysed by EBFS simply goes through a concerted E2 elimination rather than the previously proposed E1 elimination.

The predominant production of (E)-β-farnesene suggests a rigid active site cavity for this enzyme, which is very likely because of the active site conformation keep the middle and distal double bonds in a distance that is not close enough for any electrophilic attack to take place. The precisely defined active site cavity was further demonstrated by mutations introduced to the magnesium binding motif. Previous studies showed that some promiscuous sesquiterpene synthases could be conferred with new function by manipulating the magnesium binding motif. These promiscuous enzymes usually contain large active site pockets and changes to the magnesium binding motif could lead to alternative substrate binding conformations which allow different structures to be formed. However, in the case of EBFS, mutations to the magnesium binding motif did not expand the product profile; a huge reduction in catalytic activity was the only effect, indicating that the active site contour of EBFS is constructed as product-like to ensure the high fidelity.

A variety of aliphatic and aromatic residues line the hydrophobic active site cavity of sesquiterpene synthases to ensure the proper binding conformation of the substrate and stabilize carbocation intermediates. For EBFS, the residues within the active site contour mainly contribute to constructing the active site geometry since no carbocation is likely to be formed during catalysis. Thus, product selectivity should mainly rely on the composition of these residues if the active site contour is the key component determining the product outcome. However, substitution of residues that directly contribute to the active site contour with obvious size differences did not confer the enzyme with improved ability to produce cyclic products. An increased

production of (3Z,6E)-α-farnesene was observed among certain variants, suggesting the modifications forced the substrate FDP to bind in the active site with a different orientation. Still, it seems modifications within the active site of EBFS are not sufficient to reconstruct the active site template to catalyse cyclisation reactions. Construction of an active site hybrid by replacing the active surface of EBFS with the corresponding part from DCS further supports this assumption. The hybrid contains a nearly identical active site surface to DCS, but the expected product δ-cadinene was not detected. In addition, the hybrid almost completely lost the ability to convert FDP into any sesquiterpene product. This result strongly suggests that the product specificity of EBFS is not entirely determined by residues within the active site; the catalytic specificity of this enzyme might be modulated at a distance by residues surrounding the active site, which may have a huge influence on the active site volume.

6.3 N-terminal region

The truncated mutants of EBFS suggest an important role for the N-terminal region in catalysis. The function of this segment mainly relies on the sequence after R15 since deletion of the first 14 amino acids has little effect on catalytic activity, whereas loss of residues 15-24 abolishes enzyme activity. Results from the chimeric enzymes further define the function for the initial part of the N-terminal region to be capping of the active site pocket only, since either no strong interactions exist between this region and the C-terminal domain, or they are unimportant for catalysis.

The 'communication' between the N-terminal region and the C-terminal domain was identified to begin with the first conserved Ser (DCS S30; EBFS S24). Mutagenesis and tertiary structure studies suggested a role for the Ser that it maintains the proper conformation of helix C that plays an important role in catalysis. The conserved Trp (DCS W32; EBFS W26) and Phe (DCS F36; EBFS F30) residues were discovered to be involved in specific structures that stabilize the N-terminal region. Mutagenesis of

these two residues could compromise the initial structures therefore potentially confer better mobility to the N-termini. Mutants of these residues in both EBFS and DCS showed improved catalytic efficiency compared to their parent enzymes. However, no alcohol products were produced by these mutants, suggesting either water molecules still could not enter the active site or the mechanism for producing (E)-β-farnesene goes through a concerted fashion (without generation of any carbocation). This improved catalytic efficiency is probably because the better mobility of the N-terminal region allows accelerated product release, which is believed to be the rate-limiting step in the parent enzymes. In addition, these mutations may increase the hydrophilicity around the active site, which could also accelerate the departure of the hydrophobic product.

6.4 Conclusions

The ultimate aim of this study is to explore the evolutionary pathway for creating a significant divergence of biological function among sesquiterpene synthases from a common ancestor. The most critical function of this enzyme family is to provide a template for orienting bound substrate and lowering the initial activation energy. The unique mechanistic character of this evolutionarily versatile enzyme family also depends on the repetitive electrophilic and nucleophilic functionalities in the substrate. Coupled with sufficient conformational flexibility, the intrinsic reactivity can be easily tuned by natural selection to favour formation of alternative sesquiterpene product. Therefore, in order to understand the evolutionary pathway of the sesquiterpene synthase family, we have to find out how the specific amino acid residues and unique tertiary structure components coordinate with each other. Based on this knowledge, it may be possible to predict the catalytic capability of a terpene synthase solely on the basis of sequence similarity, making it easier to trace back to the common ancestor(s) of terpene synthases along the evolutionary pathway.

The product specificity was believed to be determined mainly by the plasticity of

residues comprising the active site contour. Mutagenesis studies of 19 active site residues of γ-humulene synthase in various combinations revealed the biosynthetic array of sesquiterpenes generated by this enzyme can be altered to generate alternative predominant products by replacing only three to five active site residues at a time, which represented a perfect example supporting this theory. However, our studies toward EBFS and DCS illustrated that sesquiterpene synthases can evolve by alternative patterns to control the product outcome. In the case of EBFS, it seems that the active site cavity was constructed in the way that only acyclic structures can bind. Modification of the active site surface does not enable the enzyme to produce cyclic products. In contrast, (E)-β-farnesene is quite often found in the product profile of other sesquiterpene cyclases. These results indicate the overall conformation of the active site of EBFS is possibly modulated at a distance by residues surrounding the active site. In this regard, modifications toward residues within the active site may only slightly affect the orientation of the farnesyl chain rather than creating a totally different contour to accommodate the cyclic intermediate and product. Studies of the enzymatic chemistry of DCS revealed another potential method by which nature engineers the catalytic pathway for sesquiterpene synthases. Active site mutations showed that the residue composition of the active site surface could not determine the product specificity for DCS. In addition, the swapping of more than half of the functional domain with the phylogenetically related sesquiterpene synthase GCS did not bring any promiscuity to this high-fidelity enzyme. These findings are incompatible with all previous mechanistic studies for other sesquiterpene synthases. The unexpected results suggest the product specificity of DCS might be strictly defined by a subdomain encoded by a specific exon. Even though this assumption need to be verified by additional domain swapping tests, it indicates the possibility of engineering diversity in terpenoid biosynthesis by using phylogenetic information.

6.5 Future work

Plant sesquiterpenes have important uses in medicine, agriculture and industry,

however comparatively little is known about the biosynthesis of this large class of compounds. The studies described here suggest that the product specificity of sesquiterpene synthases could be determined by different patterns. The product specificity of DCS was closely related to the subdomain encoded by exon 4. The importance of this subdomain could be further investigated by translocating the DCS-specific subdomain to that region of GCS. In addition, mutations could be introduced into this subdomain to examine the effect of specific residues and mutagenesis work may also be used to identify the minimum number of residues or specific combinations of mutants that are necessary for interconversion of the enzyme activities. The mutagenesis work has been based on crystallographic studies of DCS, but the current crystal structure contains an unusual conformation of the bound substrate analogue (2F-FDP) which may not represent the realistic substrate binding conformation. Therefore, different substrate analogues could be co-crystalized with DCS to obtain alternative, potentially more realistic, substrate binding conformations.

Further studies of EBFS would be greatly assisted by a crystal structure of this enzyme. Work to obtain a crystal structure of EBFS is underway in collaboration with Prof. David Christianson (University of Pennsylvania, Philadelphia, USA). Visual inspection of the crystal structure with a substrate analogue could possibly confirm the function of residues investigated in this study. To further investigate the biosynthetic strategy of EBFS, the equivalent experiment to the formation of the active site hybrid described here could be performed by replacing the shell of the C-terminal domain of EBFS with the corresponding region of DCS while maintaining only the original active site surface of EBFS. If this hybrid can produce cyclic product(s), then site-directed mutagenesis could be employed to identify the minimum number of residues or specific combination of residues that are necessary for the interconversion of enzyme activities. And also, it may suggest that the product specificity of EBFS is determined by the structure outside active site which could have a great influence on the volume of the active site pocket. In that case, it would be clear that the limitation towards the product specificity of EBFS is due to the active

site volume.

The conclusion that the N-terminal region regulates the catalytic efficiency of certain sesquiterpene synthases is currently based only on studies of DCS and EBFS. The corresponding experiments may be performed on other plant sesquiterpene synthase to test the predictions made here. Such experiments would also test the universality of the conclusions made here among other plant terpene synthases in general. Further studies to optimize the function of N-terminal region can be done by making different combinations of the various mutations.

Most of the current studies focused on identifying residues within and surrounding the active site that contribute directly to the reaction specificity of terpenoid synthases. Modifications of these residues may be useful for designing novel functions for terpenoid synthases. However, such application is usually limited by the avalibility of an efficient screening method. In addition, it seems that the specificity for terpene synthase families is not always determined by the plasticity of residues. Therefore, in the long-term, exploring the structural features that govern reaction product specificity for various members of terpenoid synthase families should reveal a general perspective into the evolution of these enzymes. Such knowledge would also potentially assist the development in rational design and engineering of different catalytic activities into existing terpenoid synthases.

Chapter 7: References

1. Christianson, D. W. (2008) Unearthing the roots of the terpenome. *Curr. Opin. Plant Biol. 12*, 141-150.

2. Sacchettini, J. C. (1997) Creating Isoprenoid Diversity. *Science 277*, 1788-1789.

3. Christianson, D. W. (2007) Chemistry. Roots of biosynthetic diversity. *Science 316*, 60-61.

4. Davis, E. M., and Croteau, R. (2000) Cyclization enzymes in the biosynthesis of monoterpenes, sesquiterpenes, and diterpenes. *Top. Curr. Chem. 209*, 53–95.

5. Cane, D. (1999) Comprehensive natural products chemistry. *Compr. Nat. Prod. Chem.* (Amsterdam, E., Ed.).

6. Wendt, K., Schulz, G., Corey, E., and Liu, D. (2000) Enzyme mechanisms for polycyclic triterpene formation. *Angew. Chem. 39*, 2812-2833.

7. Altman, L. J., Ash, L., Kowerski, R. C., Epstein, W. W., Larsen, B. R., Rilling, H. C., Muscio, F., and Gregonis, D. E. (1972) Prephytoene pyrophosphate. New intermediate in the biosynthesis of carotenoids. *J. Am. Chem. Soc. 94*, 3257-3259.

8. Rilling, H. C., and Epstein, W. W. (1969) Mechanism of squalene biosynthesis. Presqualene, a pyrophosphorylated precursor to squalene. *J. Am. Chem. Soc. 91*, 1041-1042.

9. Gunawardena, K., Rivera, S. B., and Epstein, W. W. (2002) The monoterpenes of *Artemisia tridentata ssp. vaseyana*, *Artemisia cana ssp. viscidula* and *Artemisia tridentata ssp. spiciformis*. *Phytochemistry 59*, 197-203.

10. Bierl-Leonhardt, B. A., Moreno, D. S., Schwarz, M., Forster, H. S., Plimmer, J. R., and DeVilbiss, E. D. (1980) Identification of the pheromone of the Comstock mealybug. *Life Sci. 27*, 399-402.

11. Kellogg, B. A., and Poulter, C. D. (1997) Chain elongation in the isoprenoid biosynthetic pathway. *Curr. Opin. Chem. Biol. 1*, 570-578.

12. Poulter, C. D., and Rilling, H. C. (1978) The prenyl transfer reaction. Enzymic and mechanistic studies of the 1'-4 coupling reaction in the terpene biosynthetic pathway. *Acc. Chem. Res. 11*, 307-313.

13. Tarshis, L. C., Yan, M., Poulter, C. D., and Sacchettini, J. C. (1994) Crystal structure of recombinant farnesyl diphosphate synthase at 2.6-A resolution. *Biochemistry 33*, 10871-10877.

14. Kloer, D. P., Welsch, R., Beyer, P., and Schulz, G. E. (2006) Structure and reaction geometry of geranylgeranyl diphosphate synthase from *Sinapis alba.*, *Biochemistry 45*, 15197-15204.

15. Thulasiram, H. V., Erickson, H. K., and Poulter, C. D. (2007) Chimeras of two isoprenoid synthases catalyze all four coupling reactions in isoprenoid biosynthesis. *Science 316*, 73-76.

16. Gershenzon, J., and Dudareva, N. (2007) The function of terpene natural products in the natural world. *Nat. Chem. Biol. 3*, 408-414.

17. Langenheim, J. H. (1994) Higher-plant terpenoids-a phytocentric overview of their ecological rolesh. *J. Chem. Ecol.* Springer *20*, 1223-1280.

18. Dufourc, E. J. (2008) Sterols and membrane dynamics. *J. Chem. Biol. 1*, 63-77.

19. Ikonen, E. (2008) Cellular cholesterol trafficking and compartmentalization. *Nat. Rev. Mol. Cell Biol. 9*, 125-138.

20. Ikonen, E. (2006) Mechanisms for cellular cholesterol transport: defects and human disease. *Physiol. Rev. 86*, 1237-1261.

21. Maxfield, F. R., and Tabas, I. (2005) Role of cholesterol and lipid organization in disease. *Nature 438*, 612-621.

22. Law, M. R. (2000) Plant sterol and stanol margarines and health. *West. J. Med. 173*, 43-47.

23. Britton, G. (1995) Structure and properties of carotenoids in relation to function. *Fed. Am. Soc. Exp. Biol. 9*, 1551-1558.

24. Rao, A. V., and Rao, L. G. (2007) Carotenoids and human health. *Pharmacol. Res. 55*, 207-216.

25. Uchimiya, M., and Stone, A. T. (2009) Reversible redox chemistry of quinones: impact on biogeochemical cycles. *Chemosphere 77*, 451-458.

26. Raymond, K., Müller, G., and Matzanke, B. (1984) Complexation of iron by siderophores a review of their solution and structural chemistry and biological function. *Top. Curr. Chem. 123*, 49-102-102.

27. Thomson, R. (1996) Naturally occurring quinones IV, 4th Edition. Blackie, London

28. Patai, S. (1974) The Chemistry of the Quinonoid Compounds. *The Chemistry of the Functional Groups* (Patai, S., Ed.). Wiley, New York.

29. Aharoni, A., Giri, A. P., Deuerlein, S., Griepink, F., De Kogel, W.-J., Verstappen, F. W. A., Verhoeven, H. A., Jongsma, M. A., Schwab, W., and Bouwmeester, H. J. (2003) Terpenoid metabolism in wild-type and transgenic arabidopsis plants. *Plant Cell. 15*, 2866-2884.

30. Morrissey, J. P., and Osbourn, A. E. (1999) Fungal resistance to plant antibiotics as a mechanism of pathogenesis. *Microbiol. Mol. Biol. Rev. 63*, 708-724.

31. Papadopoulou, K., Melton, R. E., Leggett, M., Daniels, M. J., and Osbourn, A. E. (1999) Compromised disease resistance in saponin-deficient plants. *Proc. Natl. Acad. Sci. U. S. A. 96*, 12923-12928.

32. Laurent, P., Braekman, J.-C., Daloze, D., and Pasteels, J. (2003) Biosynthesis of defensive compounds from beetles and ants. *Eur. J. Org. Chem. 2003*, 2733-2743.

33. Jung, V., and Pohnert, G. (2001) Rapid wound-activated transformation of the green algal defensive metabolite caulerpenyne. *Tetrahedron 57*, 7169-7172.

34. Bhadury, P., and Wright, P. C. (2004) Exploitation of marine algae: biogenic compounds for potential antifouling applications. *Planta 219*, 561-578.

35. Eckstein-Ludwig, U., Webb, R. J., Van Goethem, I. D., East, J. M., Lee, A. G., Kimura, M., O'Neill, P. M., Bray, P. G., Ward, S. A., and Krishna, S. (2003) Artemisinins target the SERCA of Plasmodium falciparum. *Nature 424*, 957-961.

36. Jordan, M. A., and Wilson, L. (2004) Microtubules as a target for anticancer drugs. *Nat. Rev. Cancer 4*, 253-265.

37. Knudsen, J. T., Eriksson, R., Gershenzon, J., and Stahl, B. D. A. J.-M. (2006) Diversity and distribution of floral scent. *Bot. Rev. 72*, 1-120.

38. Shields, V. D. C., and Hildebrand, J. G. (2001) Responses of a population of antennal olfactory receptor cells in the female moth Manduca sexta to plant-associated volatile organic compounds. *J. Comp. Physiol. A. 186*, 1135-1151.

39. Dicke, M., Beek, T. A. V., Posthumus, M. A., Dom, N. B., Bokhoven, H. V., and Groot, A. D. (1990) Isolation and identification of volatile kairomone that affects acarine predator-prey interaction. *J. Chem. Ecol. 16*, 381-396.

40. Turlings, T. C., Tumlinson, J. H., and Lewis, W. J. (1990) Exploitation of herbivore-induced plant odors by host-seeking parasitic wasps. *Science 250*, 1251-1253.

41. Akiyama, K., Matsuzaki, K.-I., and Hayashi, H. (2005) Plant sesquiterpenes induce hyphal branching in arbuscular mycorrhizal fungi. *Nature. 435*, 824-827.

42. Runyon, J. B., Mescher, M. C., and De Moraes, C. M. (2006) Volatile chemical cues guide host location and host selection by parasitic plants. *Science. 313*, 1964-1967.

43. Bouwmeester, H. J., Matusova, R., Zhongkui, S., and Beale, M. H. (2003) Secondary metabolite signalling in host-parasitic plant interactions. *Curr. Opin. Plant Biol. 6*, 358-364.

44. Roberts, C. (1938) Roberts : The chemistry of natural rubber. Part I . *J. Rub. Res. 7*, 215-219.

45. Ruzicka, L. (1959) History of the isoprene rule. *Proc. Chem. Soc.* 341-376.

46. Wallach, O. (1887) Zur kenntniss der terpene und ätherischen oele. *Justus Liebig's Annalen der Chemie 238*, 78-89.

47. Ruzicka, L. (1953) The isoprene rule and the biogenesis of terpenic compounds. *Experientia. 9*, 357-367.

48. Bloch, K. (1965) The biological synthesis of cholesterol. *Science 150*, 19-28.

49. Flesch, G., and Rohmer, M. (1988) Prokaryotic hopanoids: the biosynthesis of the bacteriohopane skeleton. Formation of isoprenic units from two distinct acetate pools and a novel type of carbon/carbon linkage between a triterpene and D-ribose. *Eur. J. Biochem. 175*, 405-411.

50. Rohmer, M., Knani, M., Simonin, P., Sutter, B., and Sahm, H. (1993) Isoprenoid biosynthesis in bacteria: a novel pathway for the early steps leading to isopentenyl diphosphate. *Biochem. J. 295*, 517-524.

51. Schwender, J., Seemann, M., Lichtenthaler, H. K., and Rohmer, M. (1996) Biosynthesis of isoprenoids (carotenoids, sterols, prenyl side-chains of chlorophylls and plastoquinone) via a novel pyruvate/glyceraldehyde 3-phosphate non-mevalonate pathway in the green alga Scenedesmus obliquus. *Biochem. J. 316*, 73-80.

52.   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. *J. Am. Chem. Soc. 118*, 2564-2566.

53.   Miziorko, H. M. (2011) Enzymes of the mevalonate pathway of isoprenoid biosynthesis. *Arch. Biochem. Biophys. 505*, 131-143.

54.   Wilding, E. I., Brown, J. R., Bryant, A. P., Chalker, A. F., Holmes, D. J., Ingraham, K. A., Iordanescu, S., So, C. Y., Rosenberg, M., and Gwynn, M. N. (2000) Identification, evolution, and essentiality of the mevalonate pathway for isopentenyl diphosphate biosynthesis in gram-positive cocci. *J. Bacteriol. 182*, 4319-4327.

55.   Eisenreich, W., Bacher, A., Arigoni, D., and Rohdich, F. (2004) Biosynthesis of isoprenoids *via* the non-mevalonate pathway. *Cell. mol. life sci. 61*, 1401-1426.

56.   Lüttgen, H., Rohdich, F., Herz, S., Wungsintaweekul, J., Hecht, S., Schuhr, C. A., Fellermeier, M., Sagner, S., Zenk, M. H., Bacher, A., and Eisenreich, W. (2000) Biosynthesis of terpenoids: YchB protein of *E. coli* phosphorylates the 2-hydroxy group of 4-diphosphocytidyl-2C-methyl-D-erythritol. *Proc. Natl. Acad. Sci. 97*, 1062-1067.

57.   Christianson, D. W. (2006) Structural biology and chemistry of the terpenoid cyclases. *Chem. Rev. 106*, 3412-3442.

58.   Prisic, S., Xu, J., Coates, R. M., and Peters, R. J. (2007) Probing the role of the DXDD motif in Class II diterpene cyclases. *Chembiochem 8*, 869-874.

59.   Chen, X. Y., Chen, Y., Heinstein, P., and Davisson, V. J. (1995) Cloning, expression, and characterization of (+)-delta-cadinene synthase: a catalyst for cotton phytoalexin biosynthesis. *Arch. Biochem. Biophys. 324*, 255-266.

60.   Steele, C.L., Crock, J., Bohlmann, J., and Croteau, R. (1998) Sesquiterpene Synthases from *Grand Fir* (Abies). *Biochemistry 273*, 2078-2089.

61.   Ishii, K., Sagami, H., and Ogura, K. (1986) A novel prenyltransferase from *Paracoccus denitrificans. Biochem. J. 233*, 773-777.

62.   Hosfield, D. J., Zhang, Y., Dougan, D. R., Broun, A., Tari, L. W., Swanson, R. V., and Finn, J. (2004) Structural basis for bisphosphonate-mediated inhibition of isoprenoid biosynthesis. *J. Biol. Chem. 279*, 8526-8529.

63. Tarshis, L. C., Proteau, P. J., Kellogg, B. A., Sacchettini, J. C., and Poulter, C. D. (1996) Regulation of product chain length by isoprenyl diphosphate synthases. *Proc. Natl. Acad. Sci. U. S. A. 93*, 15018-15023.

64. Guenther, A., Hewitt, C. N., Erickson, D., Fall, R., Geron, C., Graedel, T., Harley, P., Klinger, L., Lerdau, M., Mckay, W. A., Pierce, T., Scholes, B., Steinbrecher, R., Tallamraju, R., Taylor, J., and Zimmerman, P. (1995) A global model of natural volatile organic compound emissions. *J. Geophys. Res. 100*, 8873-8892.

65. Behnke, K., Ehlting, B., Teuber, M., Bauerfeind, M., Louis, S., Hänsch, R., Polle, A., Bohlmann, J., and Schnitzler, J.-P. (2007) Transgenic, non-isoprene emitting poplars don't like it hot. *Plant J. 51*, 485-99.

66. Vickers, C. E., Possell, M., Cojocariu, C. I., Velikova, V. B., Laothawornkitkul, J., Ryan, A., Mullineaux, P. M., and Nicholas Hewitt, C. (2009) Isoprene synthesis protects transgenic tobacco plants from oxidative stress. *Plant cell Environ. 32*, 520-31.

67. Siwko, M. E., Marrink, S. J., de Vries, A. H., Kozubek, A., Schoot Uiterkamp, A. J. M., and Mark, A. E. (2007) Does isoprene protect plant membranes from thermal shock? A molecular dynamics study. *Biochim. Biophys. Acta. 1768*, 198-206.

68. Loivamäki, M., Mumm, R., Dicke, M., and Schnitzler, J.-P. (2008) Isoprene interferes with the attraction of bodyguards by herbaceous plants. *Proc. Natl. Acad. Sci. 105*, 17430-5.

69. Laothawornkitkul, J., Paul, N. D., Vickers, C. E., Possell, M., Taylor, J. E., Mullineaux, P. M., and Hewitt, C. N. (2008) Isoprene emissions influence herbivore feeding decisions. *Plant cell Environ. 31*, 1410-5.

70. Bohlmann, J., and Keeling, C. I. (2008) Terpenoid biomaterials. *Plant J. 54*, 656-669.

71. Rude, M. A., and Schirmer, A. (2009) New microbial fuels: a biotech perspective. *Curr. opin. Microbiol. 12*, 274-281.

72. Köksal, M., Zimmer, I., Schnitzler, J.-P., and Christianson, D. W. (2010) Structure of isoprene synthase illuminates the chemical mechanism of teragram atmospheric carbon emission. *J. Mol. Biol.* Elsevier Ltd *402*, 363-373.

73. Croteau, R., Felton, M., and Ronald, R. C. (1980) Biosynthesis of monoterpenes: conversion of the acyclic precursors geranyl pyrophosphate and neryl pyrophosphate to the rearranged monoterpenes fenchol and fenchone by a soluble enzyme preparation from fennel (*Foeniculum vulgare*). *Arch. Biochem. Biophys.* 200, 524-533.

74. Iijima, Y., Davidovich-Rikanati, R., Fridman, E., Gang, D. R., Bar, E., Lewinsohn, E., and Pichersky, E. (2004) The biochemical and molecular basis for the divergent patterns in the biosynthesis of terpenes and phenylpropenes in the peltate glands of three cultivars of basil1. *Plant Physiol.* 136, 3724-3736.

75. Whittington, D. A., Wise, M. L., Urbansky, M., Coates, R. M., Croteau, R. B., and Christianson, D. W. (2002) Bornyl diphosphate synthase: Structure and strategy for carbocation manipulation by a terpenoid cyclase. *Proc. Natl. Acad. Sci. U. S. A.* National Academy of Sciences 99, 15375-15380.

76. Williams, D. C., McGarvey, D. J., Katahira, E. J., and Croteau, R. (1998) Truncation of limonene synthase preprotein provides a fully active "pseudomature" form of this monoterpene cyclase and reveals the function of the amino-terminal arginine pair. *Biochemistry 37*, 12213-12220.

77. Wise, M. L., Savage, T. J., Katahira, E., and Croteau, R. (1998) Monoterpene synthases from common sage (*Salvia officinalis*). cDNA isolation, characterization, and functional expression of (+)-sabinene synthase, 1,8-cineole synthase, and (+)-bornyl diphosphate synthase. *J. Biol. Chem. 273*, 14891-14899.

78. Cane, D. E. (1990) Enzymatic formation of sesquiterpenes. *Chem. Rev. 90*, 1089-1103.

79. Munck, S. L., and Croteau, R. (1990) Purification and characterization of the sesquiterpene cyclase patchoulol synthase from *Pogostemon cablin*. *Arch. Biochem. Biophys. 282*, 58-64.

80. Vögeli, U., Freeman, J. W., and Chappell, J. (1990) Purification and characterization of an inducible sesquiterpene cyclase from elicitor-treated tobacco cell suspension cultures. *Plant physiology 93*, 182-187.

81. Shishova, E. Y., Yu, F., Miller, D. J., Faraldos, J. A., Zhao, Y., Coates, R. M., Allemann, R. K., Cane, D. E., and Christianson, D. W. (2008) X-ray crystallographic studies of substrate binding to aristolochene synthase suggest a metal ion binding sequence for catalysis. *J. Biol. Chem. 283*, 15431-15439.

82. Yu, F., Miller, D. J., and Allemann, R. K. (2007) Probing the reaction mechanism of aristolochene synthase with 12,13-difluorofarnesyl diphosphate. *Chem. commun.* 4155-4157.

83. Forcat, S., and Allemann, R. K. (2006) Stabilisation of transition states prior to and following eudesmane cation in aristolochene synthase. *Org. Biomol. Chem.* 4, 2563-2567.

84. 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)-beta-farnesene synthase. *Biochemistry 42*, 7741-7747.

85. Calvert, M. J., Taylor, S. E., and Allemann, R. K. (2002) Tyrosine 92 of aristolochene synthase directs cyclisation of farnesyl pyrophosphate. *Chem. commun.* 2384-2385.

86. Faraldos, J. A., González, V., Senske, M., and Allemann, R. K. (2011) Templating effects in aristolochene synthase catalysis: elimination versus cyclisation. *Org. Biomol. Chem.* The Royal Society of Chemistry 9, 6920-6923.

87. Faraldos, J. A., Kariuki, B., and Allemann, R. K. (2010) Intermediacy of eudesmane cation during catalysis by aristolochene synthase. *J. Org. Chem. 75*, 1119-25.

88. Faraldos, J. A., Antonczak, A. K., González, V., Fullerton, R., Tippmann, E. M., and Allemann, R. K. (2011) Probing eudesmane cation-π interactions in catalysis by aristolochene synthase with non-canonical amino acids., *Journal of the American Chemical Society. 133*, 13906-9.

89. Lesburg, C. A., Zhai, G., Cane, D. E., and Christianson, D. W. (1997) Crystal structure of pentalenene synthase: mechanistic insights on terpenoid cyclization reactions in biology. *Science 277*, 1820-1824.

90. Rynkiewicz, M. J., Cane, D. E., and Christianson, D. W. (2001) Structure of trichodiene synthase from Fusarium sporotrichioides provides mechanistic inferences on the terpene cyclization cascade. *Proc. Natl. Acad. Sci. U. S. A. 98*, 13543-13548.

91. 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.* AAAS 277, 1815-1820.

92. Wendt, K. U., and Schulz, G. E. (1998) Isoprenoid biosynthesis: manifold chemistry catalyzed by similar enzymes. *Structure 6*, 127-133.

93. Shishova, E. Y., Di Costanzo, L., Cane, D. E., and Christianson, D. W. (2007) X-ray crystal structure of aristolochene synthase from *Aspergillus terreus* and evolution of templates for the cyclization of farnesyl diphosphate. *Biochemistry* *46*, 1941-1951.

94. Cane, D. E., and Kang, I. (2000) Aristolochene synthase: purification, molecular cloning, high-level expression in *E. coli*, and characterization of the *Aspergillus terreus* cyclase. *Arch. Biochem. Biophys. 376*, 354-64.

95. Felicetti, B., and Cane, D. E. (2004) Aristolochene synthase: mechanistic analysis of active site residues by site-directed mutagenesis. *J. Am. Chem. Soc. 126*, 7212-21.

96. Townsend, B. J., Poole, A., Blake, C. J., and Llewellyn, D. J. (2005) Antisense suppression of a (+)-delta-cadinene synthase gene in cotton prevents the induction of this defense response gene during bacterial blight infection but not its constitutive expression. *Plant Physiology. 138*, 516-528.

97. Tan, X. P., Liang, W. Q., Liu, C. J., Luo, P., Heinstein, P., and Chen, X. Y. (2000) Expression pattern of (+)-delta-cadinene synthase genes and biosynthesis of sesquiterpene aldehydes in plants of *Gossypium arboreum L. Planta 210*, 644-651.

98. Meng, Y. L., Jia, J. W., Liu, C. J., Liang, W. Q., Heinstein, P., and Chen, X. Y. (1999) Coordinated accumulation of (+)-delta-cadinene synthase mRNAs and gossypol in developing seeds of *Gossypium hirsutum* and a new member of the cad1 family from *G. arboreum. J. Nat. Prod. 62*, 248-252.

99. Davis, E. M., Tsuji, J., Davis, G. D., Pierce, M. L., and Essenberg, M. (1996) Purification of (+)-delta-cadinene synthase, a sesquiterpene cyclase from bacteria-inoculated cotton foliar tissue. *Phytochemistry 41*, 1047-1055.

100. Chen, X. Y., Wang, M., Chen, Y., Davisson, V. J., and Heinstein, P. (1996) Cloning and heterologous expression of a second (+)-delta-cadinene synthase from *Gossypium arboreum. J. Nat. Prod. 59*, 944-951.

101. Portnoy, V., Benyamini, Y., Bar, E., Harel-Beja, R., Gepstein, S., Giovannoni, J. J., Schaffer, A. A., Burger, J., Tadmor, Y., Lewinsohn, E., and Katzir, N. (2008) The molecular and biochemical basis for varietal variation in sesquiterpene content in melon (*Cucumis melo L.*) rinds. *Plant Mol. Biol. 66*, 647-661.

102. González-Garza, M. T., Matlin, S. A., Mata-Cárdenas, B. D., and Said-Fernández, S. (1992) Further studies on the in vitro activity of gossypol as antiamebic agent. *Arch. Med. Res. 23*, 69-70.

103. Matlin, S. A., Zhou, R., Bialy, G., Blye, R. P., Naqvi, R. H., and Lindberg, M. C. (1985) (-)-Gossypol: an active male antifertility agent. *Contraception 31*, 141-149.

104. Benedict, C. R., Lu, J.-L., Pettigrew, D. W., Liu, J., Stipanovic, R. D., and Williams, H. J. (2001) The cyclization of farnesyl diphosphate and nerolidyl diphosphate by a purified recombinant delta-cadinene synthase. *Plant Physiol. 125*, 1754-1765.

105. Alchanati, I., Patel, J. A. A., Liu, J., Benedict, C. R., Stipanovic, R. D., Bell, A. A., Cui, Y., and Magill, C. W. (1998) The enzymatic cyclization of nerolidyl diphosphate by δ-cadinene synthase from cotton stele tissue infected with *verticillium dahliae*. *Phytochemistry 47*, 961-967.

106. Davis, G. D., and Essenberg, M. (1995) (+)-δ-Cadinene is a product of sesquiterpene cyclase activity in cotton. *Phytochemistry 39*, 553-567.

107. Arigoni, D. (1975) Stereochemical aspects of sesquiterpene biosynthesis. *Pure Appl. Chem. 41*, 219-245.

108. Masciadri, R., Angst, W., and Arigoni, D. (1985) A revised scheme for the biosynthesis of gossypol, *Chem. Commun.* 1573.

109. Faraldos, J.; Miller, D.; Gonzalez, V.; Yoosuf-Aly, Z.; Cascón, O.; Li, A.; Allemann, R. K. (in press). A 1,6-ring closure mechanism for (+)-δ-cadinene synthase? *J. Am. Chem. Soc.*

110. Stipanovic, R. D., Stoessl, A., Stothers, J. B., Altman, D. W., Bell, A. A., and Heinstein, P. (1986) The stereochemistry of the biosynthetic precursor of gossypol. *Chem. Commun.* 100.

111. Poulter, C. D., and King, C. H. R. (1982) Model studies of terpene biosynthesis. A stepwise mechanism for cyclization of nerol to alpha-terpineol. *J. Am Chem. Soc. 104*, 1422-1424.

112. Miller, D. J., Yu, F., and Allemann, R. K. (2007) Aristolochene synthase-catalyzed cyclization of 2-fluorofarnesyl-diphosphate to 2-fluorogermacrene A. *Chembiochem 8*, 1819-1825.

113. Faraldos, J. A., Zhao, Y., O'Maille, P. E., Noel, J. P., and Coates, R. M. (2007) Interception of the enzymatic conversion of farnesyl diphosphate to 5-epi-aristolochene by using a fluoro substrate analogue: 1-fluorogermacrene A from (2E,6Z)-6-fluorofarnesyl diphosphate. *Chembiochem 8*, 1826-1833.

114. Gennadios, H. A., Gonzalez Gonzalez, V., Di Costanzo, L., Li, A., Yu, F., Miller, D. J., Allemann, R. K., and Christianson, D. W. (2009) Crystal structure of (+)-delta-cadinene synthase from *Gossypium arboreum* and evolutionary divergence of metal binding motifs for catalysis. *Biochemistry 48*, 6175-6183.

115. Crock, J., Wildung, M., and Croteau, R. (1997) Isolation and bacterial expression of a sesquiterpene synthase cDNA clone from peppermint (*Mentha x piperita, L.*) that produces the aphid alarm pheromone (*E*)-beta-farnesene. *Proc. Natl. Acad. Sci. U. S. A. 94*, 12833-12838.

116. Novotny, M., Harvey, S., and Jemiolo, B. (1990) Chemistry of male dominance in the house mouse, *Mus domesticus*. *Experientia. 46*, 109-113.

117. Fernandes, A., Duffield, R. M., Wheeler, J. W., and LaBerge, W. E. (1981) Chemistry of the Dufour's gland secretions of North American andrenid bees (*Hymenoptera: Andrenidae*). *J. Chem. Ecol. 7*, 453-463.

118. Jackson, B. D., Morgan, E. D., and Billen, J. P. J. (1990) Contents of the pygidial gland of the ant *Myrmecia nigriceps*. *Naturwissenschaften. 77*, 187-188.

119. Tesh, R. B., Guzman, H., and Wilson, M. L. (1992) Trans-beta-farnesene as a feeding stimulant for the sand fly lutzomyia-longipalpis (*diptera, psychodidae*). *J. Med. Entomol. 29*, 226-231.

120. Kielty, J. P., Allen-Williams, L. J., Underwood, N., and Eastwood, E. A. (1996) Behavioral responses of three species of ground beetle (*Coleoptera: Carabidae*) to olfactory cues associated with prey and habitat. *J. Insect Behav. 9*, 237-250.

121. Binder, B. F., Robbins, J. C., and Wilson, R. L. (1995) Chemically mediated ovipositional behaviors of the european corn-borer, ostrinia-nubilalis (*lepidoptera, pyralidae*). *J. Chem. Ecol. 21*, 1315-1327.

122. Turlings, T. C. J., Tumlinson, J. H., Heath, R. R., Proveaux, A. T., and Doolittle, R. E. (1991) Isolation and identification of allelochemicals that attract the larval parasitoid, cotesia-marginiventris (*cresson*), to the microhabitat of one of its hosts. *J. Chem. Ecol. 17*, 2235-2251.

123. Bowers, W. S., Nault, L. R., Webb, R. E., and Dutky, S. R. (1972) Aphid alarm pheromone: isolation, identification, synthesis. *Science 177*, 1121-1122.

124. Wohlers, P. (1981) Effects of the alarm pheromone (E)-β-farnesene on dispersal behaviour of the pea aphid *acyrthosiphon pisum*. *Entomol. Exp. Appl.* *29*, 117-124.

125. Sun, Y., Qiao, H., Ling, Y., Yang, S., Rui, C., Pelosi, P., and Yang, X. (2011) New analogues of (E)-β-farnesene with insecticidal activity and binding affinity to aphid odorant-binding proteins. *J. Agric. Food Chem. 59*, 2456-2461.

126. Maruyama, T., Ito, M., and Honda, G. (2001) Molecular cloning, functional expression and characterization of (E)-beta farnesene synthase from *Citrus junos. Biol. Pharm. Bull. 24*, 1171-1175.

127. Picaud, S., Brodelius, M., and Brodelius, P. E. (2005) Expression, purification and characterization of recombinant (E)-beta-farnesene synthase from *Artemisia annua. Phytochemistry 66*, 961-967.

128. Huber, D. P. W., Philippe, R. N., Godard, K.-A., Sturrock, R. N., and Bohlmann, J. (2005) Characterization of four terpene synthase cDNAs from methyl jasmonate-induced *Douglas-fir, Pseudotsuga menziesii. Phytochemistry 66*, 1427-1439.

129. Schnee, C., Köllner, T. G., Held, M., Turlings, T. C. J., Gershenzon, J., and Degenhardt, J. (2006) The products of a single maize sesquiterpene synthase form a volatile defense signal that attracts natural enemies of maize herbivores. *Proc. Natl. Acad. Sci. U. S. A. 103*, 1129-1134.

130. Köksal, M., Jin, Y., Coates, R. M., Croteau, R., and Christianson, D. W. (2011) Taxadiene synthase structure and evolution of modular architecture in terpene biosynthesis. *Nature. 469*, 116-120.

131. Cao, R., Zhang, Y., Mann, F. M., Huang, C., Mukkamala, D., Hudock, M. P., Mead, M. E., Prisic, S., Wang, K., Lin, F.-Y., Chang, T.-K., Peters, R. J., and Oldfield, E. (2010) Diterpene cyclases and the nature of the isoprene fold. *Proteins 78*, 2417-2432.

132. Bohlmann, J., Meyer-Gauen, G., and Croteau, R. (1998) Plant terpenoid synthases: molecular biology and phylogenetic analysis. *Proc. Natl. Acad. Sci.* T *95*, 4126-4133.

133. Lin, X., Hezari, M., Koepp, A. E., Floss, H. G., and Croteau, R. (1996) Mechanism of taxadiene synthase, a diterpene cyclase that catalyzes the first step of taxol biosynthesis in *Pacific yew. Biochemistry 35*, 2968-2977.

134. Jin, Q., Williams, D. C., Hezari, M., Croteau, R., and Coates, R. M. (2005) Stereochemistry of the macrocyclization and elimination steps in taxadiene biosynthesis through deuterium labeling. *J. Org. Chem. 70*, 4667-4675.

135. Jin, Y., Williams, D. C., Croteau, R., and Coates, R. M. (2005) Taxadiene synthase-catalyzed cyclization of 6-fluorogeranylgeranyl diphosphate to 7-fluoroverticillenes. *J. Am. Chem. Soc. 127*, 7834-7842.

136. Phillips, M., and Croteau, R. (1999) Resin-based defenses in conifers. *Trends plant sci. 4*, 184-190.

137. Ravn, M. M., Coates, R. M., Jetter, R., and Croteau, R. B. (1998) Stereospecific intramolecular proton transfer in the cyclization of geranylgeranyl diphosphate to (−)-abietadiene catalyzed by recombinant cyclase from grand fir (*Abies grandis*). *Chem. Commun.* 21-22.

138. Köksal, M., Hu, H., Coates, R. M., Peters, R. J., and Christianson, D. W. (2011) Structure and mechanism of the diterpene cyclase ent-copalyl diphosphate synthase. *Nat. chem. biol. 7*, 431-433.

139. Eschenmoser, A., Ruzicka, L., Jeger, O., and Arigoni, D. (1955) Zur kenntnis der triterpene. 190. mitteilung. eine stereochemische interpretation der biogenetischen isoprenregel bei den Triterpenen. *Helv. Chim. Acta 38*, 1890-1904.

140. Ourisson, G., and Rohmer, M. (1992) Biohopanoids: a novel class of bacterial lipids. *Acc. Chem. Res. 25*, 403-408.

141. Wendt, K. U. (1997) Structure and function of a squalene cyclase. *Science 277*, 1811-1815.

142. Poralla, K., Hewelt, A., Prestwich, G. D., Abe, I., Reipen, I., and Sprenger, G. (1994) A specific amino acid repeat in squalene and oxidosqualene cyclases. *Trends Biochem. Sci. 19*, 157-158.

143. Reinert, D. J., Balliano, G., and Schulz, G. E. (2004) Conversion of squalene to the pentacarbocyclic hopene. *Chem. Biol.* Elsevier *11*, 121-126.

144. Feil, C., Sussmuth, R., Jung, G., and Poralla, K. (1996) Site-Directed Mutagenesis of Putative Active-Site Residues in Squalene-Hopene Cyclase. *Eur. J. Biochem. 242*, 51-55.

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

146. Maplestone, R. A., Stone, M. J., and Williams, D. H. (1992) The evolutionary role of secondary metabolites--a review. *Gene 115*, 151-157.

147. Frugoli, J. A., McPeek, M. A., Thomas, T. L., and McClung, C. R. (1998) Intron loss and gain during evolution of the catalase gene family in angiosperms. *Genetics 149*, 355-365.

148. Fryxell, K. J. (1996) The coevolution of gene family trees. *Trends Genet. 12*, 364-369.

149. Clegg, M. T., Cummings, M. P., and Durbin, M. L. (1997) The evolution of plant nuclear genes, *Proc. Natl. Acad. Sci. U. S. A. 94*, 7791-7798.

150. Davisson, V. J., Woodside, A. B., Neal, T. R., Stremler, K. E., Muehlbacher, M., and Poulter, C. D. (1986) Phosphorylation of isoprenoid alcohols. *J. Org. Chem. 51*, 4768-4779.

151. Poulter, C.D.; Argyle, J. C.; Mash, E. A. (1978) Farnesyl pyrophosphate synthase. *J. Org. Chem. 253*, 7227-7233.

152. Slater, M. Selman, S. Hartnett, J. (1998) Pfu DNA polymerase: A high fidelity enzyme for nucleic acid amplification, *promega notes 68*, 7-13.

153. Andrade, M.A., Chacón, P., Merelo, J.J., and Morán, F. (1993) Evaluation of secondary structure of proteins from UV circular dichroism using an unsupervised learning neural network. *Prot. Eng. 6*, 383-390.

154. McLafferty, F. W. (2011) Wiley registry 9th edition / NIST 2011 mass spectral library. Wiley.

155. Yoshikuni, Y., Ferrin, T. E., and Keasling, J. D. (2006) Designed divergent evolution of enzyme function. *Nature 440*, 1078-1082.

156. Aharoni, A., Gaidukov, L., Khersonsky, O., Gould, S. M., Roodveldt, C., and Tawfik, D. S. (2005) The "evolvability" of promiscuous protein functions. *Nat. Genet. 37*, 73-76.

157. Yoshikuni, Y., Martin, V. J. J., Ferrin, T. E., and Keasling, J. D. (2006) Engineering cotton (+)-delta-cadinene synthase to an altered function: germacrene D-4-ol synthase. *Chem. Biol.* Elsevier *13*, 91-98.

158. Lodeiro, S., Segura, M. J. R., Stahl, M., Schulz-Gasch, T., and Matsuda, S. P. T. (2004) Oxidosqualene cyclase second-sphere residues profoundly influence the product profile. *Chembiochem 5*, 1581-1585.

159. Hyatt, D. C., and Croteau, R. (2005) Mutational analysis of a monoterpene synthase reaction: altered catalysis through directed mutagenesis of (-)-pinene synthase from *Abies grandis*. *Arch. Biochem. Biophys. 439*, 222-233.

160. Greenhagen, B. T., O'Maille, P. E., Noel, J. P., and Chappell, J. (2006) Identifying and manipulating structural determinates linking catalytic specificities in terpene synthases. *Proc. Natl. Acad. Sci. U. S. A. 103*, 9826-9831.

161. Back, K., and Chappell, J. (1996) Identifying functional domains within terpene cyclases using a domain-swapping strategy. *Proc. Natl. Acad. Sci. U. S. A. 93*, 6841-6845.

162. Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., and Higgins, D. G. (1997) The Clustal_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res. 25*, 4876-4882.

163. Julsing, M. K. (2006). Bioconversion and combinatorial biosynthesis of selected terpenoids and lignans. *Ph.D. Thesis. University of Groningen: Netherlands*.

164. Proctor, R. H., and Hohn, T. M. (1993) Aristolochene synthase. Isolation, characterization, and bacterial expression of a sesquiterpenoid biosynthetic gene (Ari1) from *Penicillium roqueforti*. *J. Biol. Chem. 268*, 4543-4548.

165. Aaron, J. A., Lin, X., Cane, D. E., and Christianson, D. W. (2010) Structure of epi-isozizaene synthase from Streptomyces coelicolor A3(2), a platform for new terpenoid cyclization templates. *Biochemistry 49*, 1787-1797.

166. Salin, F., Pauly, G., Charon, J., and Gleizes, M. (1995) Purification and characterization of trans-beta-farnesene synthase from maritime pine (*Pinus pinaster Ait.*) needles. *J. Plant Physiol. 146*, 203-209.

167. Cane, D. E. (1990) Enzymic formation of sesquiterpenes. *Chem. Rev. 90*, 1089-1103.

168. Facchini, P. J. (1992) Gene family for an elicitor-induced sesquiterpene cyclase in tobacco. *Proc. Natl. Acad. Sci. U. S. A. 89*, 11088-11092.

169. Colby, S. M. (1998) Germacrene C synthase from *Lycopersicon esculentum cv. VFNT* Cherry tomato: cDNA isolation, characterization, and bacterial expression of the multiple product sesquiterpene cyclase. *Proc. Natl. Acad. Sci. U. S. A. 95*, 2216-2221.

170. Rynkiewicz, M. J., Cane, D. E., and Christianson, D. W. (2002) X-ray crystal structures of D100E trichodiene synthase and its pyrophosphate complex reveal the basis for terpene product diversity. *Biochemistry 41*, 1732-1741.

171. Cane, D. E., Xue, Q., and Fitzsimons, B. C. (1996) Trichodiene synthase. Probing the role of the highly conserved aspartate-rich region by site-directed mutagenesis. *Biochemistry. 35*, 12369-12376.

172. López-Gallego, F., Wawrzyn, G. T., and Schmidt-Dannert, C. (2010) Selectivity of fungal sesquiterpene synthases: role of the active site's H-1 alpha loop in catalysis. *Appl. Environ. Microbiol. 76*, 7723-7733.

173. Dmitrenko, O., Thorpe, C., and Bach, R. D. (2007) Mechanism of SN2 disulfide bond cleavage by phosphorus nucleophiles. Implications for biochemical disulfide reducing agents. *J. Org. Chem. 72*, 8298-8307.

174. O'Maille, P. E., Malone, A., Dellas, N., Andes Hess, B., Smentek, L., Sheehan, I., Greenhagen, B. T., Chappell, J., Manning, G., and Noel, J. P. (2008) Quantitative exploration of the catalytic landscape separating divergent plant sesquiterpene synthases. *Nat. Chem. Biol. 4*, 617-623.

175. Cane, D. E., Chiu, H. T., Liang, P. H., and Anderson, K. S. (1997) Pre-steady-state kinetic analysis of the trichodiene synthase reaction pathway. *Biochemistry. 36*, 8332-8339.

176. McAndrew, R. P., Peralta-Yahya, P. P., Degiovanni, A., Pereira, J. H., Hadi, M. Z., Keasling, J. D., and Adams, P. D. (2011) Structure of a three-domain sesquiterpene synthase: a prospective target for advanced biofuels production. *Structure 19*, 1876-1884.

177. Kim, M.-Y., Chang, Y.-J., Bang, M.-H., Baek, N.-I., Jin, J., Lee, C.-H., and Kim, S.-U. (2005) cDNA isolation and characterization of (+)-germacrene a synthase fromIxerls dentata form.*albiflora Hara. J. Plant Biol. 48*, 178-186.

Appendices

A I Sequence alignment of DCSs, EBFSs and some other plant terpene synthases

Black: aliphatic residues. Purple: hydroxyl residues. Salmon: acidic residues. Teal: basic residues. Brown: aromatic residues. Mustard: imino residue. Olive: sulfur residues. Green: amide residues. For key to enzymes see key for A I.

```
1    1                                                        MASQVSQMPSSSPLSSNKDEMRPKADFQPSIW
2    1                                                        MASQVSQMPSSSPLSSNKDEIRPKADFQPSIW
3    1                                                        MASQVSQMPSSSPLSSNKDEMRPKADFQPSIW
4    1                                                        MASQVSQMPSSSPLSSNKDEMRPKADFQPSIW
5    1                                                        MASQASQVLAS.PHPAISSENRPKADFHPGIW
6    1                                                        MATNGVVISCLREVRPPMTKHAPSMW
7    1                                          MSTLPISSVSFSSSTSPLVVDDKVSTKPDVIRHTMNFNASIW
8    1                                                        MKDMSIPLLAAVSSSTEETVRPIADFHPTLW
9    1                                                        MAMPVKLTPASLSLKAVCCRFSSGGHALRF
10   1           MEFRVHLHADHEQKILQNQMKP.....................EHEASYLINQ.....RRSANYKPNIW
11   1                                                        MALAIPFNNEEEIVRPVANFSPSLW
12   1                                                        MASAAVANYEEEIVRPVADFSPSLW
13   1                                                        MAAVQANVTGIKANTKTSAEPVRPLANFPPSVW
14   1                                                        MAASS..........ADKCRPLANFHPSVW
15   1                                                        MGSEVNRPLADFPANIW
16   1           MSIISMNVSILSKPLNCLHNLERRP..SKALLVPCTAPTAR......LRASCSSKLQEAHQIRR.SGNYQPALW
17   1             MALKLLTSLP..MYNFSRVPVSSKDPILLVTSRTRNGYLARPVQCMVANKVSTSPDILRRSANYQPSIW
18   1  MALLSIVSLQVPKSCGLKSLISSSNVQKALCISTAVPTLRMRRRQKALVINMKLTTVSHRDDNGGGVLQRRIADHHPNLW
19   1  MSPVSVISLPS.DLCLPTSFIDRS.GRELIPLHITIPNVAMRRQGKLMTR..ASMSMNLRTAVSDDAVIRRRGDFHSNLW
```

```
1    33  GDLFLNCPDKNIDAET..EKRHQQLKEEVRKMIVAPMANS..........TQKLAFIDSVQRLGVSYHFTKEIEDELENI
2    33  GDFFLNCPDKNIDAGT..EKRHQQLKEEVRKMIVAPMANS..........TQKLAFIDSVQRLGVSYHFTKEIEDELENI
3    33  GDFFLNCPDKNIDAET..QKRHQQLKEEVRKMIVAPMANS..........TLKLAFIDSVQGLGVSYHFTKEIEDELENI
4    33  GDLFLNCPDKNIDAET..EKRHQQLKEEVRKMIVAPMANS..........TQKLAFIDSVQRLGVSYHFTKEIEDELENI
5    32  GDMFIICPDTDIDAAT..ELQYEELKAQVRKMIMEPVDDS..........NQKLPFIDAVQRLGVSYHFEKEIEDELENI
6    27  TDTFSNFSLDDKEQQK.CSETIEALKQEARGMLMAATTPL..........QQMTLIDTLERLGLSFHFETEIEYKIELI
7    43  GDQFLTYDEPEDLVMK..KQLVEELKEEVKKELITIKGSNEPMQ.....HVKLIELIDAVQRLGIAYHFEEEIEEALQHI
8    32  GNHFLKSAADVETIDAATQEQHAALKQEVRRMITTTANKL..........AQKLHMIDAVQRLGVAYHFEKEIEDELGKV
9    31  GSSLPCWRRTPTQRSTSSSTTRPAAEVSSGKSKQHDQEASEAT......IRQQLQLVDVLENMGISRHFAAEIKCILDRT
10   44  KNDFLDQSLISKYDGDEYRKLSEKLIEEVKIYISAETKDL..........VAKLELIDSVRKLGLANHFEKEIKEALDGI
11   26  GDRFHSFSLDNQVAEK.YAQEIETLKEQTRSLLSAAACGIT.......LAEKLNLIDIVERLGLAYHFEKQIDDMLDQI
12   26  GDQFLSFSIKNQVAEK.YAKEIEALKEQTRNML..LATGMK.......LADTLNLIDTIERLGISYHFEKEIDDILDQI
13   34  GDRFLSFSLDRSELER.YAIAMEKPKEDLRKLIVDPTMDS..........NEKLGLIYSVHRLGLTYMFLQEIESQLDKL
14   21  GYHFLSYT.HEITNQE..KVEVDEYKETIRKMLVETCDNS..........TQKLVLIDAMQRLGVAYHFDNEIETSIQNI
15   18  EDP.LTSFSKSDLGTETFKEKHSTLKEAVKEAFMSSKANP.........IENIKFIDALCRLGVSYHFEKDIVEQLDKS
16   66  DSNYI.QSLNTPYTEERHLDRKAELIVQVRILL.KEKMEP.........VQQLELIHDLKYLGLSDFFQDEIKEILGVI
17   68  NHDYI.ESLRIEFVGETCTRQINVLKEQVRMML.HKVVNP.........LEQLELIEILQRLGLSYHFEEEIKRILDGV
18   81  EDDFIQSLSSPYGGSS.YSERAVTVVEEVKEMFNSIPN.NRELFGSQNDLLTRLWMVDSIERLGIDRHFQNEIRVALDYV
19   77  DDDLIQSLSSPYGEPS.YRERAERLIGEVKNSFNSMSNEDGESITPLDDLIQRLWMVDSVERLGIDRHFKKEIKSALDHV
```

```
1    101 Y...HNNND.......AENDLYTTSIRFRLLREHGYNVSCDVFNKFKD....EQGNFKSSVTSDVRGLLELYCASYLRV.
2    101 Y...HNNND.......AENDLYTTSLRFRLLREHGYNVSCDVFNKFKD....EQGNFKSSVTSDVQGLLELYCASYLRV.
3    101 Y...HNNND.......AENDLYTTSLRFRLLREHGFHVSCDVFNKFKD....EQGNFKSSVTSDVRGLLELYCASYLRV.
4    101 Y...HNNND.......AENDLYTTSLRFRLLREHGFNVSCDVFNKFKD....EQGNFKSSVTSDVRGLLELYCASYLRV.
5    100 YRD.TMNND.......ADTDLYTTALRFRLLREHGFDISCDAFNKFKD....EAGNFKASLTSDVQGLLELYEASYMRV.
6     95 NAAEDDGFD............LFATALRFRLLRQHQRHVSCDVFDKFID....KDGKFEESLSNNVEGLLSLYEAAHVGF.
7    116 HVTYGEQWV.......DKENLQSISLWFRLLRQQGFNVSSGVFKDFMD....EKGKFESLCNDAQGILALYEAAFMRV.
8    102 SHDLDS...........DDLYVVSLRFRLFRQQGVKISCDVFDKFKD....DEGKFESLINDIRGMLSLYEAAYLAI.
9    105 YRSWLQRHEE......IMLDTMTCAMAFRILRLNGYNVSDELYHVVEAS..GLHNSLGGYLNDTRTLLELHKASTVSIS
10   114 AAIES...DNLG....TRDDLYGTALHFKILRQHGYKVSQDIFGRFMDE...KDT.LENHHFAHLKGMLELFEASNLGF.
11    97 YKADPNFDG.........HDLNTLSLQFRILRQHGYNISQKFFSRFQD....ANG.FKECLSNDIRGLLNLYEASHVRT.
12    95 YNQNSWCN..........DLCTSALQFRLLRQHGFNISPEIFSKFCD....ENGKFKESLASDVLGLLNLYEASHVRT.
13   103 FNKFSLQDY.......EEVDLYTISINFQVFRHVGYKLPCDVFNKFKDV...SSGTFKASITSDVG.VVGLYESAQLRI.
14    88 FDASSKQND.......NDNNLYVVSLRFRLVRQQGHYMSSDVFKQFTN....QDGKFKETLTNDVQGLLSLYEASHLRV.
15    87 FDCLDFPQMVRQ....EGCDLYTVGIIFQVFRQFGFKLSADVFEKFKD....ENGKFKGHLVTDAYGMLSLYEAAQWGT.
16   134 YNEHKCFHNNEV....EKMDLYFTALGFRLLRQHGFNISQDVFNCFKNE...KGIDFKASLAQDTKGMLQLYEASFLLR.
17   136 YNNDHGGDTWKA....EN..LYATALKFRLLRQHGYSVSQEVFNSFKDE...RGS.FKACLCEDTKGMLSLYEASFFLI.
18   159 YSYWKEKEGIGCGRDSTFPDLNSTALALRTLRLHGYNVSSDVLEYFKDQKGHFACPAILTEGQITRSVLNLYRASLVAF.
19   156 YRYWSEK.GIGCGRESVVTDLNSTALGLRTLRLHGYDVSADVLNHFKNQSGQFACTLKQTEDQIR.TVLNLYRASLIAF.


1    166 HGEDILDEAISFTTHHLSLAVA..SLDHP...LSEEVSHALKQSIRRGLPRVEARHYLS.VYQDIESHNK..........
2    166 HGEDILDEAISFTTNHLSLAVS..SLDHP...LSEEVSHALKQSIRRGLPRVEARHYLS.VYQDIESHNK..........
3    166 HGEDILDEAISFTSNHLSLAVA..SLDHP...LSEEVSHALKQSIRRGLPRVEARHYLS.VYQDIESHNK..........
4    166 HGEDILDEAISFTTNHLSLAVA..SLDYP...LSEEVSHALKQSIRRGLPRVEARHYLS.VYQDIESHNK..........
5    167 HGEDILDEAISFTTAQLTLALP..TLHHP...LSEQVGHALKQSIRRGLPRVEARNFIS.IYQDLESHNK..........
6    159 REERILQEAVNFTRHHLEGAELD....QSPLLIREKVKRALEHPLHRDFPIVYARLFIS.IYEKDDSRDE..........
7    184 EDETILDNALEFTKVHLDIIAK..DPSCDS.SLRTQIHQALKQPLRRRLARIEALHYMP.IYQQETSHNE..........
8    165 RGEDILDEAIVFTTHHLKSVISISDHSHANSNLAEQIRHSLQIPLRKAAARLEARYFLD.IYSRDDLHDE..........
9    177 EDESILDSIGSRSRTLLREQLE.SGGALRKPSLFKEVEHALDGPFYTTLDRLHHRWNIENFNIIEQHMLETPYLSNQHTS
10   182 EGEDILDEAKASLTLALRDSGH.ICYP..DSNLSRDVVHSLELPSHRRVQWFDVKWQINAYEKDICRVNA..........
11   162 HGEDILEEALVFSTAHLESAAP..HLESP...LSKQVTHALEQSLHKSIPRVETRYFIS.IYEEEEFKND..........
12   159 HADDILEDALAFSTIHLESAAP..HLKSP...LREQVTHALEQCLHKGVPRVETRFFISSIYDKEQSKNN..........
13   171 RGEKILDEASVFTEAKLKSVVN..TLEGD...LAQQVTQSLRRPFHQGMPLGIRQGSISLTMKKNVPLMT..........
14   156 RNEEILEEALTFTTHHLESIVS..NLSNNNNSLKVEVGEALTQPIRMTLPRMGARKYIS.IYENNDAHHH..........
15   158 HGEDIIDEALAFSRSHLEEIS.....SRSSPHLAIRIKNALKHPYHKGISRIETRQYIS.YYEEEESCDP..........
16   206 KGEDTLELAREFATKCLQKKLD.EGGNEIDENLLLWIRHSLDLPLHWRIQSVEARWFIDAYARRPD.MNP..........
17   205 EGENILEEARDFSTKHLEEYVK.QNK...EKNLATLVNHSLELPLHWRMLRLEARWFINIYRHNQD.VNP..........
18   238 PGEKVMEEAEIFSASYLKEVLQ....KIPVSSFSREIEYVLEYGWHTNLPRLEARNYIDVYGQDSYESSNEMPYV...NT
19   233 PGEKVMDEAESFSAKYLKEALQ....KIPVSSFSREIGDVLEYGWHTYLPRLEARNYIDVFGQDTENSKS...YM...KT
```

```
1   230  .ALLEFAKIDFNMLQFLHRKELSEICRWWKDLDFQRKLPYARDRVVEGYFWI.SGVYFEPQYSLGRKMLTKVIAMASIVD
2   230  .ALLEFAKIDFNMLQFLHRKELSEICRWWKDLDFQRKLPYARDRVVEGYFWI.SGVYFEPQYSLGRKMLTKVIAMASIVD
3   230  .VLLEFAKIDFNMVQLLHRKELSEISRWWKDLDFQRKLPYARDRVVEGYFWI.SGVYFEPQYSLGRKMLTKVIAMASIVD
4   230  .VLLEFAKIDFNMVQLLHRKELSEISRNWKDLDFQRKLPYARDRVVEGYFWI.SGVYFEPQYSLGRKMLTKVIAMASIVD
5   231  .SLLQFAKIDFNLLQLLHRKELSEICRWWKDLDFTRKLPFARDRVVEGYFWI.MGVYFEPQYSLGRKMLTKVIAMASIVD
6   224  .LLLKLSKVNFKFMQNLYKEELSQLSRWWNTWNLKSKLPYARDRVVEAYVWG.VGYHYEPQYSYVRMGLAKGVLICGIMD
7   250  .DLLKLAKLDFSVLQSMHKKELSHICKWWKDLDLQNKLPYVRDRVVEGYFWI.LSIYYEPQHARTRMFLMKTCMWLVVLD
8   234  .TLLKFAKLDFNILQAAHQKEASIMTRWWNDLGFPKKVPYARDRIIETYIWMLLGVSYEPNLAFGRIFASKVVCMITTID
9   256  RDILALSIRDFSSSQFTYQQELQHLESWVKECRLD.QLQFARQKLAYFYLSA.AGTMFSPELSDARTLWAKNGVLTTIVD
10  249  .TLLGLAKLNFNMVQAQLQKDLREASKWWANLGIADNLKFARDRLVECFACA.VGVAFEPEYSSFRICLTKVINLVLIID
11  226  .VLLRFAKLDYNLLQMLHKHELSEVSRWWKDLDFVTTLPYARDRAVECYFWT.MGVYAEPQYSQARVILAKTIAMISIVD
12  224  .VLLRFAKLDFNLLQMLHKQELAQVSRWWKDLDFVTTLPYARDRVVECYFWA.LGVYFEPQYSQARVMLVKTISMSIVD
13  236  .HCLKLAKLHFKYLELQQKEELRIVSKWWKDMRFHETTPYIRDRVPEIYLWI.LGLYFEPRYSLARIIATKITLFLVVLD
14  223  .LLLKFAKLDFNMLQKFHQRELSDLTRWWKDLDFANKYPYARDRLVECYFWI.LGVYFEPKYSRARKMTKVLNLTSIID
15  222  .TLLEFAKIDFNLLQILHREELACVTRWHHEMEFKSKVTYTRHRITEAYLWS.LGTYFEPQYSQARVITTMALILFTALD
16  274  .LIFELAKLNFNIIQATHQQELKDLSRWWSRLCFPEKLPFVRDRLVESFFWA.VGMFEPHQHGYQRKMAATIIVLATVID
17  270  .ILLEFAELDFNIVQAAHQADLKQVSTWWKSTGLVENLSFARDRPVENFFWT.VGLIFQPQFGYCRRMFTKVFALITTID
18  311  QKLLKLAKLEFNIFHSLQQKELQYISRWWKDSCSS.HLTFTRHRHVEYYTMA.SCISMEPKHSAFRLGFVKTCHLLTVLD
19  303  EKLLELAKLEFNIFHALQKRELEYLVRWWKGSGSP.QMTFCRHRHVEYYTLA.SCIAFEPQHSGFRLGFAKACHIITVLD


1   308  DTYDSYATYEELIPYTNAIERWDIKCIDEI..PEYMKPSYKALLDVYE.EMVQLVAEHGRQYRVEYAKNAMIRLA..QSY
2   308  DTYDSYATYEELIPYTNAIERWDIKCIDEL..PEYMKPSYKALLDVYK.EMEQLVAEHGRQYRVEYAKNAMIRLA..QSY
3   308  DTYDSYATYEELIPYTNAIERWDIKCIDEL..PEYMKPSYKALLDVYE.EMEQLVAEHGRQYRVEYAKNAMIRLA..QSY
4   308  DTYDSYATYEELIPYTKAIERWDIKCIDEL..PEYMKPSYKALLDVYE.EMEQLVAKHGRQYRVEYAKNAMIRLA..QSY
5   309  DTYDSYATYDELIPYTNAIERWDIKCMNQL..PNYMKISYKALLNVYE.EMEQLLANQGRQYRVEYAKKAMIRLV..QAY
6   302  DTYDNYATLNEAQLFTQVLDKWDRDEAERL..PEYMKIVYRFILSIYE.NYERDAAKLGKSFAAPYFKETVKQLA..RAF
7   328  DTFDNYGTYEELEIFTQAVERWSISCLDML..PEYMKLIYQELVNLHV.EMEESLGKGGKNISNSLCQGRWQKELGSQIT
8   313  DTFDAYGTFEELTLFTEAVTRWDIGLIDTL..PEYMKFIVKALLDIYR.EAEEELAKEGRSYGIPYAKQMMQELI..ILY
9   334  DFFDVAGSKEELENLVMLVEMWDEHHKVEFY.SEQVEIIFSSIYDSVN.QLGEKASLVQDRSITKHLVEIWLDLL..KSM
10  327  DVYDIYGSEEELKHFTNAVDRWDSRETEQL..PECMKMCFQVLYNTTCEIAHEIEKDNGWNQVLPQLTKVWADFC..KAL
11  304  DTFDAYGIVKELEVYTDAIQRWDISQIDRL..PEYMKVSFKALLDLYE.DYEKELSKDGRSDVVHYAKERMKEIV..RNY
12  302  DTFDAYGTVKELEAYTDAIQRWDINEIDRL..PDYMKISYRAILDLYK.DYEKELSSAGRSHIVCHAIERMKEVV..RNY
13  314  DTYDAYATIEEIRLLTDAINKWDISAMEQI..PEYIRPFYKILLDEYAGNWRRKWLKKGEQILLLLQKKRSKTLA..RGY
14  301  DTFDAYATFDELVTFNDAIQRWDANAIDSI..QPYMRPAYQALLDIYS.EMEQVLSKEGKLDRVYYAKNEMKKLV..RAY
15  300  DMYDAYGTMEELELFTDAMDEWLPVVPDEIPIPDSMKFIYNVTVEFYD.KLDEELEKEGRSGCGFHLKKSLQKTA..NGY
16  352  DIYDVYGTLDELELFTDTFKRWDTESITRL..PYYMQLCYWGVHNYISDAAYDILKEHGF.FCLQYLRKSVVDLV..EAY
17  348  DVYDVYGTLDELELFTDVVERWDINAMDQL..PDYMKICFLTLHNSVNEMALDTMKEQRF.HIIKYLKKAWVDLC..RSY
18  389  DMYDTFGTLDELQLFTTAFKRWDLSETKCL..PEYMKAVYMDLYQCLN.ELAQEAEKTQGRDTLNYIRNAYESHF..DSF
19  381  DMYDTFGTLDELELFTSAIKRWDPSATECL..PEYMKGVYMIVYNTVN.EMSQEADKAQGRDTLNYCRQAWEEYI..DAY
```

```
1   383  LVEAKWTLQN.YKPSFEEFKANALPTCGYAMLAITSFVGMGD.IVTPETFKWAASDPKIIQASTIICRFMDDVAEHKFKH
2   383  LVEARWTLQN.YKPSFEEFKANALPTCGYAMLAITSFVGMGD.IVTPETFKWAANDPKIIQASTIICRFMDDVAEHKFKH
3   383  LVEARWTLQN.YKPSFEEFKANALPTCGYAMLAITSFVGMGD.IVTPETFKWAANDPKIIQASTIICRFMDDVTEHKFKH
4   383  LVEARWTLQN.YKPSFEEFKANALPTCGYAMLAITSFVGMGD.IVTPETFKWAANDPKIIQASTIICRFMDDVAEHKFKH
5   384  LLEAKWTHQN.YKPTFEEFRDNALPTSGYAMLAITAFVGMGE.VITPETFKWAASDPKIIKASTIICRFMDDIAEHKFNH
6   377  NEEQKWVMER.QLPSFQDYVKNSEKTSCIYTMFASIIPGLKS..VTQETIDWIKSEPTLATSTAMIGRYWNDTSSQLRES
7   405  LVETKMAKRGVHAQPLEEYMSVSMVTGTYGLMIARSYVGRGD.IVTEDTFKWVSSYPPIIKASCVIVRLMDDIVSHKEEQ
8   388  FTEAKWLYKG.YVPTFDEYKSVALRSIGLRTLAVASFVDLGDFIATKDNFECILKNAKSLKATETIGRLMDDIAGYKFEQ
9   410  MTEVEWRLSK.YVPTEKEYMINASLIFGLGPIVLPALYFVGP..KISESIVKDPEYDELFKLMSTCGRLLNDVQTFEREY
10  403  LVEAEWYNKS.HIPTLEEYLRNGCDSSSVSILLVHSFFSITHE.GTKEMADFLHKNEDLLYNLSIIVRLNNDLGTSAAEQ
11  379  FVEAKWFIEG.YMPPVSEYLSNALATSTYYLTTTSYLGVKS..ATKEDFEWLATNPKILEANVTLCRVVDDIATYEVEK
12  377  NVESTWFIEG.YTPPVSEYLSNALATTTYYYLATTSYLGMKS..ATEQDFEWLSKNPKILEASVIICRVIDDTATYEVEK
13  390  LEEAEWTNSG.YVASFPEYMKNGLITSAYNVISKSALVGMGE.IVSEDALAWYESHPKTLQASELISRLQDDVMTYQFER
14  376  FKETQWLNDCDHIPKYEEQVEMAIVSAGYMMISTTCLVGIEE.FISHETFEWLMNESVIVRASALIARAMNDIVGHEDEQ
15  377  MQEAKWLKKD.YIATFDEYKEMAILSSGYYALIAMTFVRMTD.VAKLDAFEWLSSHPKIRVASEIISRFTDDISSYEFEH
16  427  FHEAKWYHSG.YTPSLDEYLNIAKISVASPAIISPTYFTFANASHDTAVIDSLYQYHDILCLAGIILRLPDDLGTSYFEL
17  423  LLEAKWYYNK.YTPSLQEYIENAWISISAPTILVHVYFFVTNP.ITKEALDCLEEYPNIIRWSSTILRLADDLGTSTDEL
18  464  MHEAKWISSG.YLPTFEEYLKNGKVSSGSRTATLQPILTLDVP.LPNYILQEIDYPSRFNDLASSLLRLRGDTRCYKADR
19  456  MQEAKWIASG.EVPTFEEYYENGKVSSGHRVSALQPILTTDIP.FPEHVLKEVDIPSQLNDLASAILRLRGDTRCYQADR


1   461  RREDDCSAIECYMEEY.GVTAQEAYDVFNKHVESAWKDLNQEFLKPT..EMPTEVLNRSLNLARVMDVLYRE.GDGYTYV
2   461  RREDDCSAIECYMEEY.GVSAQEAYDVFNKHVESAWKDVNQEFQKPT..EMPTEVLNRSLNLARVMDVLYRE.GDGYTYV
3   461  RREDDCSAIECYMEEY.GVTAQEAYDVFNKHVESAWKDVNQGFLKPT..EMPTEVLNRSLNLARVMDVLYRE.GDGYTYV
4   461  RREDDCSAIECYMEEY.GVTAQEAYDVFNKHVESAWKDVNKEFLKPT..EMPTEVLNRSLNLARVMDVLYRE.GDGYTYV
5   462  RREDDCSAIECYMKQY.GVTAQFAYNEFNKHIESSWKDVNEEFLKPT..EMPTFVLCRSLNLARVMDVLYRE.GDGYTHV
6   454  KGGEMLTALDFHMKEY.GLTKEEAASKFEGLVEETWKDINKEFIATTNYNVGREIAITFLNYARICEASYSKTDGDAYSD
7   484  ERGHVASSIECYSKES.GASEEEACEYISRKVEDAWKVINRESLRPT..AVPFPLLMPAINLARMCEVLYSV.NDGFTHA
8   467  KRGHNPSAVECYKNQH.GVSEEEAVKELLLEVANSWKDINEELLNPT..TVPLPMLQRLLYFARSGHFIYDDGHDRYTHS
9   487  NEGKLNSVSLLVLHGG.PMSISDAKRKLQKPIDTCRRDLLSLVLREE.SVVPRPCKELFWKMCKVCYFFYST.TDGFSSQ
10  481  ERGDSPSSIVCYMREV.NASEEIARKNIKGMIDNAWKKVNGKCFTTNQVPFLSSFMNNATNMARVAHSLYKD.GDGFGDQ
11  456  GRGQIATGIECYMRDY.DVSTEVAMEKFQEMAEIAWKDVNEGILRPT..PVSTEILTRILNLARIIDVTYKHNQDGYTHP
12  454  SRGQIATGIECCMRDY.GISTKEAMAKFQNMAETAWKDINEGLLRPT..PVSTEFLTPILNLARIVEVTYIHNLDGYTHP
13  468  ERGQSATGVDAYIKTY.GVSEKEAIDALKIMIENAWKDINEGCLKPR..QVSMDLLAPILNLARMIDVVYRY.DDGFTFP
14  455  ERGHVASLIECYMKDY.GASKQETYIKFLKEVTNAWKDINKQFFRPT..EVPMFVLERVLNLTRVADTLYKE.KDTYTNA
15  455  KREHVATGIDCYMQQF.GVSKERAVEVMGNIVSDAWKDLNQELMRPH..VFPFPLLMRVLNLSRVIDVFYRY.QDAYTNP
16  506  ARGDVPKTIQCYMKET.NASEEEAVEHVKFLIREAWKDMNTAIAAG..YPFPDGMVAGAANIGRVAQFIYLH.GDGFGVQ
17  501  KRGDVPKAIQCYMNET.GASEEDAREYIKYLISATWKKMNEDRVAS..SPFSHIFIEIALNLARMAQCMYQH.GDGHGHG
18  542  ARGEEASAISCYMKDHPGSTEEDALNHINVMISDAIRELNWELLRPD.SKSPISSKKHAFDITRAFHHLYKY.RDGYTVA
19  534  ARGEEASCISCYMKDNPGTTEEDALNHLNAMISDVIKGLNWELLKPN.SSVPISAKKHAFDISRAFHCGYKY.RDGYSVA
```

```
1   537  GKAAKGGITSLLIEPIAL
2   537  GKAAKGGITSLLIEPIAL
3   537  GKAAKGGITSLLIEPIAL
4   537  GKAAKGGITSLLIEPVAL
5   538  GKAAKGGITSLLIDPIQI
6   533  PNVAKANVVALFVDAIVF
7   560  EGDMKSYMKSFFVHPMVV
8   544  .LMMKRQVALLLTEPLAI
9   564  VERAK.EVDAVINEPLKLQGSHTLVSDV
10  559  EKGPRTHILSLLFQPLVN
11  533  EKVLKPHIIALLVDSIEI
12  531  EKVLKPHIINLLVDSIKI
13  544  RKDSERVYQSFCLWVLYPV
14  531  KGKLKNMINSILIESVKI
15  531  .KLLKEHIVSLLIETIPI
16  582  HSKTYEHIAGLLFEPYA
17  577  NHETKDRILSLLIQPIPLNKD
18  620  SSETKNLVMKTVLEPVAL
19  612  NIETKSLVKRTVIDPVTL
```

No.	Enzyme	Plant	ID number
1		G. arboreum (Tree cotton)	Q39761
2		G. arboreum (Tree cotton)	Q39512
3		Gossypium (upland cotton)	P93665
		G. arboreum	
4		G. arboreum (Tree cotton)	Q39760
5	(+)-δ-cadinene synthase isozyme A	G. arboreum (Tree cotton)	Q43714
6	(E)-β-farnesene synthase	Mentha piperita (Peppermint)	O48935
7	(E)-β-farnesene synthase	Artemisia annua (Sweet wormwood)	Q9V6I2
8	(E)-β-farnesene synthase	Citrus junos	Q94JS8
9	(E)-β-farnesene synthase	Zea mays (corn)	Q84ZW8
10	α-farnesene synthase	Pyrus communis (Pear)	Q2PGC0
11	Vetispiradiene synthase	Solanum tuberosum (Potato)	Q9SBJ0
12	5-epi-aristolochene synthase	N. tabacum (Common tobacco)	Q40577
13	Germacrene A synthase	A. annua	Q1PDD2
14	Germacrene C synthase	Solanum lycopersicum (Tomato) (Lycopersicon esculentum)	O64961
15	β-caryophyllene/α-humulene synthase	Arabidopsis thaliana (Mouse-ear cress)	Q84UU4

245

Key to alignment

No.	Enzyme name	Plant	ID number
1	(+)-δ cadinene synthase isozyme XC1	*G. arboreum* (Tree cotton)	Q39761
2	(+)-δ cadinene synthase isozyme C2	*G. arboreum* (Tree cotton)	O49853
3	(+)-δ cadinene synthase	*G. hirsutum* (Upland cotton) (*G. mexicanum*)	P93665
4	(+)-δ cadinene synthase isozyme XC14	*G. arboreum* (Tree cotton)	Q39760
5	(+)-δ cadinene synthase isozyme A	*G. arboreum* (Tree cotton)	Q43714
6	(*E*)-β-farnesene synthase	*Mentha piperita* (Peppermint)	O48935
7	(*E*)-β-farnesene synthase	*Artemisia annua* (Sweet wormwood)	Q4VM12
8	(*E*)-β-farnesene synthase	*Citrus junos*	Q94JS8
9	(*E*)-β-farnesene synthase	*Zea mays* (corn)	Q84ZW8
10	α-farnesene synthase	*Pyrus communis* (Pear)	Q2PQC0
11	Vetispiradiene synthase	*Solanum. tuberosum* (Potato)	Q9SBJ0
12	5-epi-aristolochene synthase	*N. tabacum* (Common tobacco)	Q40577
13	Germacrene A synthase	*A. annua*	Q1PDD2
14	Germacrene C synthase	*Solanum lycopersicum* (Tomato) (*Lycopersicon esculentum*).	O64961
15	β-caryophyllene/α-humulene synthase	*Arabidopsis thaliana* (Mouse-ear cress)	Q84UU4

16	(+)-bornyl diphosphate synthase, chloroplast	*Salvia officinalis* (Sage)	O81192
17	Pinene synthase	*Quercus ilex* (Holly oak)	A1JH12
18	Limonene/ α-pinene synthase, chloroplast	*Abies grandis* (Grand fir)	Q9M7C9
19	(-)-α-pinene synthase	*Pinus taeda*	Q84KL6

A 2 Primers and DNA fragments for mutagenesis work

DCS-L413I-Fwd 5'-GGTTATGCCATG<u>ATT</u>GCTATTACATCTTTC-3'
DCS-L413I-Rev 5'-GAAAGATGTAATAGC<u>AAT</u>CATGGCATAACC-3'
For construction of mutant DCS-M1. Annealing temperature: 55 °C.

DCS-I300L-Fwd 5'-GATGTTGACAAAAGTGC<u>T</u>AGCAATGGCATC-3'
DCS-I300L- Rev 5'-GATGCCATTGCT<u>AG</u>CACTTTTGTCAACATC-3'
For construction of mutant DCS-M2. Annealing temperature: 55 °C.

DCS-E455G- Fwd 5'-GTTTATGGATGATGTTGCT<u>GGA</u>CACAAGTTCAAG CATAGGAG-3'
DCS-E455G- Rev 5'-CTCCTATGCTTGAACTTGTG<u>TCC</u>AGCAACATCAT CCATAAAC-3'
For construction of mutant DCS-M3. Annealing temperature: 60 °C.

DCS-T407S/C408A-Fwd 5'-GGCTAATGCATTGCCA<u>AGTGCT</u>GGTTATGC CATGCTTGC-3'
DCS-T407S/C408A-Rev 5'-GCAAGCATGGCATAACC<u>AGCACT</u>TGGCAAT GCATTAGCC-3'
For construction of mutant DCS-M5. Annealing temperature: 63 °C.

DCS-M523A-Fwd 5'-CTAAACCTTGCAAGGGTG<u>GC</u>GGATGTGC TTTACAGGG-3'
DCS-M523A-Rev 5'-CCCTGTAAAGCACATC<u>CGC</u>CACCCTTGC AAGGTTTAG-3'
For construction of mutant DCS-M6. Annealing temperature: 63 °C.

DCS-G276C-Fwd 5'-GCAAGAGATAGAGTTGTTGAA<u>TGT</u>TACTTTTGGA
TCTCTGGAG-3'

DCS-G276C-Rev 5'-CTCCAGAGATCCAAAAGTA<u>ACA</u>TTCAACAACTCT
ATCTCTTGC-3'

For construction of mutant DCS-M7. Annealing temperature: 62 °C.

DCS-D307A-Fwd 5'-CAATGGCATCCATTGTA<u>GC</u>GGATACATATG
ACTCATATG-3'

DCS-D307A-Rev 5'-CATATGAGTCATATGTATCC<u>GC</u>TACAATGG
ATGCCATTG-3'

Annealing temperature: 60 °C.

DCS-D308A-Fwd 5'-CAATGGCATCCATTGTAGAT<u>GC</u>GACATATGACTC
ATATGCAAC-3'

DCS-D308A-Rev 5'-GTTGCATATGAGTCATATGT<u>CGC</u>ATCTACAATGG
ATGCCATTG-3'

Annealing temperature: 60 °C.

DCS-D311A-Fwd 5'-CCATTGTAGATGATACATATGCCTCATATGCAACAT
ATGAAGAGC-3'

DCS-D311A-Rev 5'-GCTCTTCATATGTTGCATATGAGGCATATGTATCAT
CTACAATGG-3'

Annealing temperature: 60 °C.

DCS-G276F-Fwd 5'-GCAAGAGATAGAGTTGTTGAA<u>TTC</u>TACTTTTGGATC
TCTGGAGTG-3'

DCS-G276F-Rev 5'-CACTCCAGAGATCCAAAAGTA<u>GAA</u>TTCAACAACTCT
ATCTCTTGC-3'

For construction of *EcoR*I restriction site on DCS. Annealing temperature: 63 °C.

EBFS-Y514L-Fwd 5'-CATCACATTCCTCAAC<u>CT</u>GGCTCGGATATGTGAAG-3'

EBFS-Y514L-Rev 5'-CTTCACATATCCGAGCC<u>AG</u>GTTGAGGAATGTGATG-3'

Annealing temperature: 60 °C.

EBFS-I404Y-Fwd 5'-GAGAAAACCAGCTGC<u>TA</u>TTATACCATGTTTGCTTC-3'

EBFS-I404Y-Rev 5'-GAAGCAAACATGGTATA<u>AT</u>AGCAGCTGGTTTTCTC-3'

Annealing temperature: 57 °C.

EBFS-Y277V-Fwd 5'-CTTATGTTTGGGGAGTAGGT<u>GTG</u>CATTACGAACC
CCAATACTC-3'

EBFS-Y277V-Rev 5'-GAGTATTGGGGTTCGTAATG<u>CAC</u>ACCTACTCCCC
AAACATAAG-3'

Annealing temperature: 63 °C.

EBFS-N444D-Fwd 5'-GGTCGGTATTGG<u>GAT</u>GACACCAGCTC-3'
EBFS-N444D-Rev 5'-GAGCTGGTGTC<u>ATC</u>CCAATACCGACC-3'
Annealing temperature: 57 °C.

EBFS-A529Y-Fwd 5'-CTGACGGAGAC<u>TAT</u>TATTCAGATCC-3'
EBFS-A529Y-Rev 5'-GGATCTGAATA<u>ATA</u>GTCTCCGTCAG-3'
Annealing temperature: 55 °C.

EBFS-S402C-Fwd 5'-GAATTCAGAGAAAACC<u>TGC</u>TGCATTTATACCATG-3'
EBFS-S402C-Rev 5'-CATGGTATAAATGCA<u>GCA</u>GGTTTTCTCTGAATTC-3'
Annealing temperature: 55 °C.

EBFS-G440C-Fwd 5'-CGACCGCTATGATC<u>TGT</u>CGGTATTGGAATG-3'
EBFS-G440C-Rev 5'-CATTCCAATACCG<u>ACA</u>GATCATAGCGGTCG-3'
Annealing temperature: 58 °C.

EBFS-Y271S-Fwd 5'-GAGTCGTGGAGGCT<u>AGC</u>GTTTGGGGAGTAG-3'
EBFS-Y271S-Rev 5'-CTACTCCCCAAACG<u>CT</u>AGCCTCCACGACTC-3'
For construction of *Nhe*I restriction site on EBFS. Annealing temperature: 62 °C.

EBFS Δ14-1-Fwd 5'-GTTGCTTAAGGGAAGC<u>CA</u>GGCCACCTATGAC-3'
EBFS Δ14-1-Rev 5'-GTCATAGGTGGCC<u>TGG</u>CTTCCCTTAAGCAAC-3'
Annealing temperature: 60 °C.
EBFS Δ14-2-Fwd 5'-GCTTAAGGGAAGCCA<u>TGG</u>CACCTATGACGAAGC-3'
EBFS Δ14-2-Rev 5'-GCTTCGTCATAGGTG<u>CCA</u>TGGCTTCCCTTAAGC-3'
Annealing temperature: 60 °C.

EBFS Δ24-Fwd 5'-GCATGCGCCAA<u>CCATGG</u>GGACTGATACC-3'
EBFS Δ24-Rev 5'-GGTATCAGTCC<u>CCATGG</u>TTGGCGCATGC-3'
Annealing temperature: 61 °C.

EBFS Δ24-G2W-Fwd
5'-GAAGGAGATATACCATG<u>TGG</u>ACTGATACCTTTTC-3'
EBFS Δ24-G2W-Rev
5'-GAAAAGGTATCAGT<u>CCA</u>CATGGTATATCTCCTTC-3'
Annealing temperature: 57 °C.

D24AAE-T24S/G26W-Fwd
5'-CAAAGCCGATTTTCAGCCT<u>AGC</u>ATG<u>TGG</u>ACTGATACCTTTTCTAAC-3'
D24AAE-T24S/G26W-Rev
5'-GTTAGAAAAGGTATCAGT<u>CCA</u>CAT<u>GCT</u>AGGCTGAAAATCGGCTTTG-3'
Annealing temperature: 63 °C.

249

E24AAD-W24S-Fwd 5'-GAAGCATGCGCCA<u>AGC</u>ATTTGGGGAGATC-3'
E24AAD-W24S-Rev 5'-GATCTCCCCAAAT<u>GCT</u>TGGCGCATGCTTC-3'
Annealing temperature: 59 °C.

EBFS-W26E-Fwd 5'-CATGCGCCAAGCATG<u>GAG</u>ACTGATACCTTTTC-3'
EBFS-W26E-Rev 5'-GAAAAGGTATCAGT<u>CTC</u>CATGCTTGGCGCATG-3'
Annealing temperature: 59 °C.

EBFS-E269M-Fwd 5'-GATCGAGTCGTG<u>AT</u>GGCTTATGTTTG-3'
EBFS-E269M-Rev 5'-CAAACATAAGC<u>CAT</u>CACGACTCGATC-3'
Annealing temperature: 55 °C.

EBFS-F30A-Fwd 5'-CATGTGGACTGATACC<u>GCT</u>TCTAACTTTTCTCTTG-3'
EBFS-F30A-Rev 5'-CAAGAGAAAAGTTAGA<u>AGC</u>GGTATCAGTCCACATG-3'
Annealing temperature: 57 °C.

DCS-F36A-Fwd 5'-CATTTGGGGAGATCTC<u>GCC</u>CTCAATTGTCCCGAC-3'
DCS-F36A-Rev 5'-GTCGGGACAATTGAG<u>GGC</u>GAGATCTCCCCAAATG-3'
Annealing temperature: 63 °C.

Synthetic gene fragment encoding part of GCS (Section 3.3.2):

    *Sac*I site
5'-<u>GAGCTC</u>GTGACTTTCAATGATGCAATCCAGAGATGGGATGCTAATGCA
ATTGATTCAATACAACCATATATGAGACCTGCTTATCAAGCTCTTCTAGA
CATTTACAGTGAAATGGAACAAGTGTTGTCCAAAGAAGGTAAACTGGAC
CGTGTATACTATGCAAAAAATGAGATGAAAAAGTTGGTGAGAGCCTATT
TTAAGGAAACCCAATGGTTGAATGATTGTGACCATATTCCAAAATATGA
GGAACAAGTGGAGAATGCAATCGTAAGTGCTGGCTATATGATGATATCA
ACAACTTGCTTGGTCGGTATAGAAGAATTTATATCCCACGAGACTTTTGA
ATGGTTGATGAATGAGTCTGTGATTGTTCGAGCTTCCGCATTGATTGCCA
GAGCAATGAACGATATTGTTGGACATGAAGATGAACAAGAAAGAGGAC
ATGTAGCTTCACTTATTGAATGTTACATGAAAGATTATGGAGCTTCAAAG
CAAGAGACTTACATTAAGTTCCTGAAAGAGGTCACCAATGCATGGAAGG
ACATAAACAAACAATTCTCCCGTCCAACTGAAGTACCAATGTTTGTCCTT
GAACGAGTTCTAAATTTGACACGTGTGGCTGACACGTTATATAAGGAGA
AAGATACATATTCAACCGCCAAAGGAAAACTTAAAAACATGATTAATCC
AATACTAATTGAATCTGTCAAAATATAAGGATCCGAATTC<u>GAGCTC</u>-3'
                                stop codon   *Sac*I site

Synthetic fragment for construction of EBFS-active site hybrid (Section 4.2.14):

*Nhe*I site
5'-GCTAGCTTTTGGATCGTAGGTGTGCATTACGAACCCCAATACTCATAT
GTTCGAATGGGACTTGCCAAAGTGATAGCAATGGCATCCATCGTAGACG
ATACATATGATAATTATGCTACACTCAATGAAGCTCAACTTTTTACTCAA
GTCTTAGACAAGTGGGATAGAGATGAAGCTGAACGACTCCCAGAATACA
TGAAAATCGTTTATCGATTTATTTTGAGTATATATGAAAATTATGAACGT
GATGCAGCGAAACTTGGAAAAAGCTTTGCAGCTCCTTATTTTAAGGAAA
CCGTGAAACAACTGGCAAGGGCATATAATGAGGAGCAGAAGTGGGTTAT
GGAAAGGCAGCTACCGTCATTCCAAGACTACGTAAAGAATTCAGAGAAA
ACCTGTGGTATTGCCATGATGTTTGCTTCTATCATCCCAGGCTTGAAATCT
GTTACCCAAGAAACCATTGATTGGATCAAGAGTGAACCCACGCTCGCAA
CATCGACCGCTATGATCTGTCGGTATATGGATGACACCGCTGAACAGCTC
TTCAAGCATAGGAGAGAAGACGATCTGACTGCGTTGGATTTCCACATGA
AAGAATATGGTCTGACGAAGGAAGAGGCGGCATCTAAGTTTGAAGGATT
GGTTGAGGAAACATGGAAGGATATAAACAAGGAATTCATAGCCACAACT
AATTATAATGTGGGTAGAGAAATTGCCATCACATTCCTCAACCTTGCTCG
GGTGATGGATGTGCTTTACAGGAAAACTGACGGAGACGCTTATTCAGAT
CCTAATGTTGCCAAGGCAAATGTCGTTGCTCTCTTTGTTGATGCCATAGT
CTTTTGAGGATCC-3'
stop codon BamHI site

D24AAE-Fwd
5'-CATGGAACCTTCTTCATCACCCCTTTCTTCCAATAAGGATGAAATGCGT
CCCAAAGCCGATTTTCAGCCTAC-3'
D24AAE-Rev
5'-CATGGTAGGCTGAAAATCGGCTTTGGGACGCATTTCATCCTTATTGGA
AGAAAGGGGTGATGAAGAAGGTTC-3'
For construction of N-terminal segment hybrid D24AAE that encoding *Nco*I at both
5' and 3' ends.

E24AAD-Fwd
5'-CATGGCTACAAACGGCGTCGTAATTAGTTGCTTAAGGGAAGTAAGGCC
ACCTATGACGAAGCATGCGC-3'
E24AAD-Rev
5'-CATGGCGCATGCTTCGTCATAGGTGGCCTTACTTCCCTTAAGCAACTA
ATTACGACGCCGTTTGTAGC-3'
For construction of N-terminal segment hybrid E24AAD that encoding *Nco*I at both
5' and 3' ends.