TECHNISCHE UNIVERSITÄT MÜNCHEN
DEPARTMENT CHEMIE
LEHRSTUHL FÜR BIOCHEMIE

Structural and Functional Characterization of Class I Terpene Cyclases

Philipp Baer

Vollständiger Abdruck der von der Fakultät für Chemie der Technischen Universität München zur Erlangung des akademischen Grades eines Doktors der Naturwissenschaften (Dr. rer. nat.) genehmigte Dissertation.

Vorsitzender: Univ.-Prof. Dr. J. Buchner

Prüfer der Dissertation

1. Univ.-Prof. Dr. M. Groll
2. Univ.-Prof. Dr. T.A.M. Gulder

Die Dissertation wurde am 25.06.2015 bei der Technischen Universität München eingereicht und durch die Fakultät für Chemie am 03.08.2015 angenommen.

Table of Contents

Table of Contents

Summary

The work at hand comprises the biochemical and structural characterization of the bacterial class I sesquiterpene cyclases selina-4(15),7(11)-diene synthase[1] (SdS, *Streptomyces pristinaespiralis*) and (2*Z*,6*E*)-hedycaryol synthase[2] (HcS, *Kitasatospora setae*). Class I terpene cyclases are the key players for introducing the structural diversity into terpenoids, which represent the largest class of natural products on earth. In course of this, a limited number of linear polyprenyl diphosphates (~ 4) are converted into a large number (> 30,000) of structural distinct terpenes, hereby forming the scaffold molecules for the class of terpenoids. The underlying chemistry is based on highly reactive carbocations whose regulation is a great challenge, especially in aqueous solution. The cyclization reaction catalysed by class I terpene cyclases represents nature's way of utilizing the advantages of combinatorial chemistry. Moreover, the class of terpenoids includes prominent members like Artemisinin (anti-malaria) and Taxol (anti-cancer) and therefore, they play an important role for medicine and industry, as well.

The biochemical characterization (*in vitro*, GC-MS based) of purified SdS revealed an exclusive consumption of farnesyl diphosphate (FPP) as a substrate which was specifically converted into selina-4(15),7(11)-diene. Thermal shift assays (TSA) were conducted to characterise the protein's binding preferences. Hereby, the FPP analogue dihydrofarnesyl diphosphate (DHFPP) displayed the strongest binding towards SdS. It was possible to crystallize SdS in its apo state, in complex with PP_i-$(Mg^{2+})_3$ and in complex with DHFPP-$(Mg^{2+})_3$. The phase information for these structures was obtained experimentally by selenomethionine substitution and single-wavelength anomalous dispersion methods (SAD). It was possible to identify an **induced-fit mechanism** which for the first time explained substrate activation and carbocation formation in class I terpene cyclases. By this molecular rearrangement, these enzymes can control carbocation chemistry in aqueous solution. Underlying this mechanism is a novel effector triad, comprising a pyrophosphate sensor (R178), a linker (D181) and an effector residue (Gly182). Notably, this sophisticated architecture is present in all crystal structures of class I terpene cyclases. Thus, the induced-fit mechanism presumably applies for all class I terpene cyclases. The design of 28 SdS mutants and their analysis revealed in most cases an alteration of the corresponding product spectra (*in vitro*, GC-MS based) which turned out to be most valuable for proposing and proofing advanced mechanistic models.

In addition, HcS could be cloned, expressed, purified, crystalized and its structure determined. This sesquiterpene cyclase could be analysed regarding its substrate preference and its product spectrum, as well. HcS turned out to be a highly specific class I terpene cyclase. Circular dichroism based thermal shift assays (TSA) revealed that its strongest binder was 2-fluoro-farnesyl diphosphate. It was possible to capture the reaction intermediate analogue **(*R*)-nerolidol** in course of the protein purification from *Escherichia coli*. Crystals of HcS were obtained in its apo form and in complex with the metabolic by-product. Hereby the phase problem was solved experimentally by soaking native HcS crystals with $HgCl_2$ and applying SAD-methods. The achieved high resolution data of HcS at 1.5 Å data represented for the first time a crystal structure of a class I terpene cyclases in complex with a reaction surrogate lacking the diphosphate moiety. It is shown that the ligand (*R*)-nerolidol is deeply inserted into the active site, where the compound adopts a conformational rearrangement that closely resembles the product (2*Z*,6*E*)-hedycaryol. Interestingly, the pre-folding of the molecule takes place prior to the first intramolecular ring closure. The complex structure of (*R*)-nerolidol bound to HcS allows a structure based interpretation of a class I terpene cyclase's active site. Importantly, the crystallographic data revealed that the helix-dipole of helix G1 and the carbonyl oxygen of Val179 (effector residue) both contribute to the initial cyclisation of the substrate which exclusively takes place at the C1 atom. In addition, the orientation of the bound (*R*)-nerolidol ligand within the active site renders the existence of a nerolidyl diphosphate reaction intermediate, as proposed in the literature, rather to be unlikely. The mechanistic models could be proven by 12 distinct point mutants, which were analysed regarding their individual product spectra at two different pH values.

Taking together the biochemical and crystallographic data of SdS and HcS, an entire structure based catalytic cycle for class I terpene cyclases can be provided.

Zusammenfassung

Die vorliegende Arbeit umfasst die biochemische und strukturelle Charakterisierung der bakteriellen Klasse I Terpenzyklasen Selina-4(15),7(11)-diene Synthase[1] (SdS, *Streptomyces pristinaespiralis*) und (2*Z*,6*E*)-Hedycaryol Synthase[2] (HcS, *Kitasatospora setae*). Klasse I Terpenzyklasen sind die Schlüsselenzyme für die Generierung der strukturellen Vielfalt in der Naturstoffklasse der Terpenoide, welche die größte Naturstoff Familie der Erde darstellt. Hierbei wird eine kleine Zahl an linearen Polyprenyl Diphosphat Substraten (~4) in eine sehr große Anzahl an strukturell einzigartigen Terpenen umgewandelt (> 30.000), welche die Gerüstmoleküle für die Naturstoffklasse der Terpeniode darstellen. Die zugrunde liegende Chemie basiert auf hochreaktive Carbokationen. Die Kontrolle von diesen ist aus chemischer Sicht äußerst schwierig, zumal die Reaktion in wässriger Lösung stattfindet. Diese durch Klasse I Terpenzyklasen katalysierte Zyklisierungsreaktion stellt die in der Natur vorkommende Variante der kombinatorischen Chemie dar, welche auch für biologische Systeme große Vorteile bietet. Die Naturstoffklasse der Terpenoide umfasst bekannte Vertreter wie Artemisinin (anti-Malaria) und Taxol (anti-Krebs) und ist deshalb von größter Bedeutung für Medizin und Industrie.

Die biochemische Charakterisierung (*in vitro*, GC-MS basiert) von aufgereinigter SdS hat eine exklusive Umsetzung des Substrates Farnesyldiphosphat (FPP) gezeigt, welches spezifisch in Selina-4(15),7(11)-diene umgewandelt wurde. Um die bevorzugte Ligandenbindung zu analysieren wurden *Thermal Shift Assays* (TSA) durchgeführt. Hierbei zeigte sich, dass das Substratanalogon Dihydrofarnesyldiphosphat (DHFPP) am stärksten an SdS bindet. Es war möglich SdS in seiner Apoform, in Komplex mit PP_i-$(Mg^{2+})_3$ und in Komplex mit DHFPP-$(Mg^{2+})_3$ zu kristallisieren. Die Phaseninformation wurde experimentell durch Einbau von Selenomethionin erhalten, die Auswertung des anormalen Datensatzes wurde mittels *single-wavelength anomalous dispersion methods* (SAD-Methoden) durchgeführt. Die strukturellen Daten haben einen neuartigen ***Induced-Fit*** **Mechanismus** gezeigt, welcher es zum ersten Mal ermöglicht die Substrataktivierung und die Initiierung der Carbokationenchemie in Klasse I Terpenzyklasen zu verstehen. Dieser molekulare Mechanismus ist Grundvoraussetzung, dass Klasse I Terpenzyklasen die Carbokationenchemie in wässriger Lösung kontrollieren können. Die zugrunde liegende strukturelle Architektur umfasst eine zuvor unbeschriebene Effektor-Triade, welche aus einem Pyrophosphatsensor (R178), einem Linker (D181) und einem Effektor (G182) besteht.

Diese strukturelle Anordnung findet sich in allen verfügbaren Klasse I Terpenzyklasen Strukturen wieder. Aus diesem Grund ist es sehr wahrscheinlich, dass der beschriebene *Induced-Fit* Mechanismus für alle Klasse I Terpenzyklasen zutrifft. Das Designen und die Analyse von 28 verschiedenen Punktmutanten von SdS hat in den meisten Fällen ein abgewandeltes Produktspektrum gezeigt (*in vitro*, GC-MS basiert), was von großer Bedeutung für die Formulierung und Überprüfung von fortgeschrittenen mechanistischen Modellen war.

Zusätzlich wurde HcS kloniert, exprimiert, aufgereinigt und kristallisiert. Hierbei wurden die Ligandenbindungspreferenz und das dazugehörige Produktspektrum von HcS analysiert. Diese Untersuchungen haben gezeigt, dass HcS ebenfalls eine hochspezifische Sesquiterpenzyklase ist. Circulardichroismus basierte *thermal shift assays* (TSA) haben ergeben, dass das Substratanalogon 2-fluoro-Farnesyldiphosphat am stärksten an HcS bindet. Des Weiteren war es im Zuge der Proteinaufreinigung möglich ein Reaktionsintermediat Analogon (**(*R*)-Nerolidol**) abzufangen. Es wurden Apo-Kristalle und Kristalle in Komplex mit diesem Metabolit von HcS erhalten. Die Phasen für diese Strukturen wurden experimentell ermittelt. Hierfür wurden native HcS Kristalle mit $HgCl_2$ inkubiert und die anormalen Datensätze mit SAD-Methoden ausgewertet. Diese Strukturdaten zeigen zum ersten Mal eine Klasse I Terpenzyklase, welche in Komplex mit einem Reaktionsintermediat Analogon ist, das keine Diphosphatgruppe aufweist. Hierdurch kann der Ligand tief in das aktive Zentrum binden und vollständig vorgefaltet werden. Die Vorfaltung ist derart ausgeprägt, dass der Ligand stark dem Produkt (2*Z*,6*E*)-Hedycaryol ähnelt. Dies findet noch vor dem ersten Ringschluss statt. Der (*R*)-Nerolidol Ligand erlaubt zum ersten Mal eine strukturbasierte Interpretation des aktiven Zentrums einer Klasse I Terpenzyklase. Hierbei konnten wir zeigen, dass der Helixdipol von Helix G1 und der Carbonylsauerstoff von V179 (Effektor) den initialen Ringschluss, welcher exklusiv an der C1 Position stattfindet, katalysiert. Zusätzlich lässt die Orientierung des gebundenen (*R*)-Nerolidol Liganden vermuten, dass das in der Literatur hypothetisierte Reaktionsintermediat Nerolidyldiphosphat nicht existiert. Um unsere mechanistischen Modelle zu überprüfen, haben wir 12 Punktmutanten erzeugt und deren individuellen Produktspektren bei zwei unterschiedlichen pH Werten analysiert (*in vitro*, GC-MS basiert).

Durch die Kombination der biochemischen und strukturellen Daten von SdS und HcS ist es möglich, einen vollständigen, strukturbasierten Katalysezyklus für Klasse I Terpenzyklasen zu formulieren.

1. Introduction

"What is life?" - this fundamental question, which is also the title of Erwin Schrödinger's famous book first published in 1944, is as prevailing as it was in the past. Although people have always found elaborate answers to this topic in their respective eras, it is obvious that there will never be an all-encompassing answer. Despite the technical advances taking place in the last 20 years and the associated increase of information it is still not possible to decipher the complex molecular networks even in a single, bacterial cell. It is important to keep in mind that those complex constructs, from a single molecule to a living organism, are the temporary result of continuous optimization, achieved by evolution over millions of years. These mechanics of variation and selection (as described by Charles Darwin in his famous work "On the Origin of Species") are principally based on altering existing systems. It is therefore highly improbable that genes, proteins, natural products, molecular machines and cell types are introduced from scratch in a single step. For this reason biological systems are, on the one hand, always comprising just a minor selection of all theoretically possible configurations and conformations (DNA, amino acids, protein folds, natural products and cellular arrangement); on the other hand, there is a very high selection pressure towards biologically relevant alterations (biological activity). The outdated opinion that something like junk DNA (DNA introns) or metabolic waste (secondary metabolites) exists, clearly contradicts this fundamental principle. Therefore, natural products represent highly optimized compounds which show an intrinsic affinity towards biological systems. They target proteins, including enzymes, and display bioactivity. As a result, many promising drugs under investigation are based on natural product scaffolds (20% of all small molecule drug launches between 2005- 2007)[3][4].

When taking a closer look at the intracellular organisation of life, a general differentiation can be made on a molecular level: (**1**) DNA/RNA constitutes the information storage, -flow and -access, forming the genome. (**2**) Proteins represent the link between the informational- (Genome, 1-dimensional) and the functional space (Proteome, 3-dimensional), mainly acting on an intracellular level and constituting the proteome. (**3**) Metabolites are formed by enzymes, comprising natural products of low to high molecular weight. Their function not only covers the intracellular context but especially the intercellular one, hereby controlling a cell's communication with the outside world. Metabolites comprise the major part of biologically active compounds in chemical space, which covers all theoretically possible configurations and conformations of small molecules[5]. It is noteworthy that the above

mentioned classification represents both, a chronological order of development (DNA first, metabolites last) as well as an increasing degree of complexity taking place during evolution, which simply reflects the continuous specialisation of organisms (**Figure 1**).

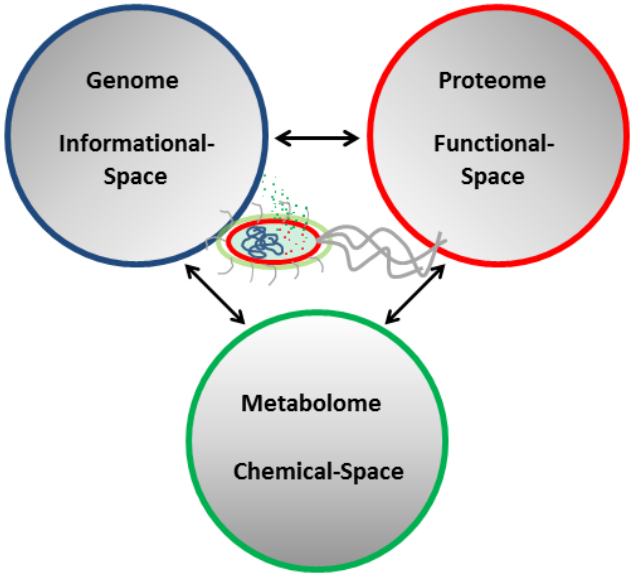


Figure 1. A scheme of the general organisation of life. The **Informational Space** is based on DNA/ RNA. The **Functional Space** comprises the proteome which represents the linker between information and function. The metabolome is part of the **Chemical Space**. It is generated by enzymes and features chemical compounds with various biological functions.

Many metabolites act as signalling substances, attractants, repellents or chemical weapons against hostile organisms[6]. Since the main survival strategy of all species is to continuously adapt to ever changing surrounding conditions, it is beneficial to develop modular, molecular systems which feature a high degree of flexibility and diversity. Listed below are a number of advantageous aspects of modular molecular systems: (**1**) common key- precursor molecules among different species reduce the amount of genetic information necessary to synthesize a certain compound. As a result, (**2**) not all genes have to be horizontally transferred to spread the essential genetic information among organisms. (**3**) In order to generate new products, just a few key- enzymes have to be genetically varied; the "common- substrate" biosynthesis can remain unchanged. **Figure 2** gives an overview of the terpenoid biosynthesis, which covers all aspects of a perfect, modular molecular system.

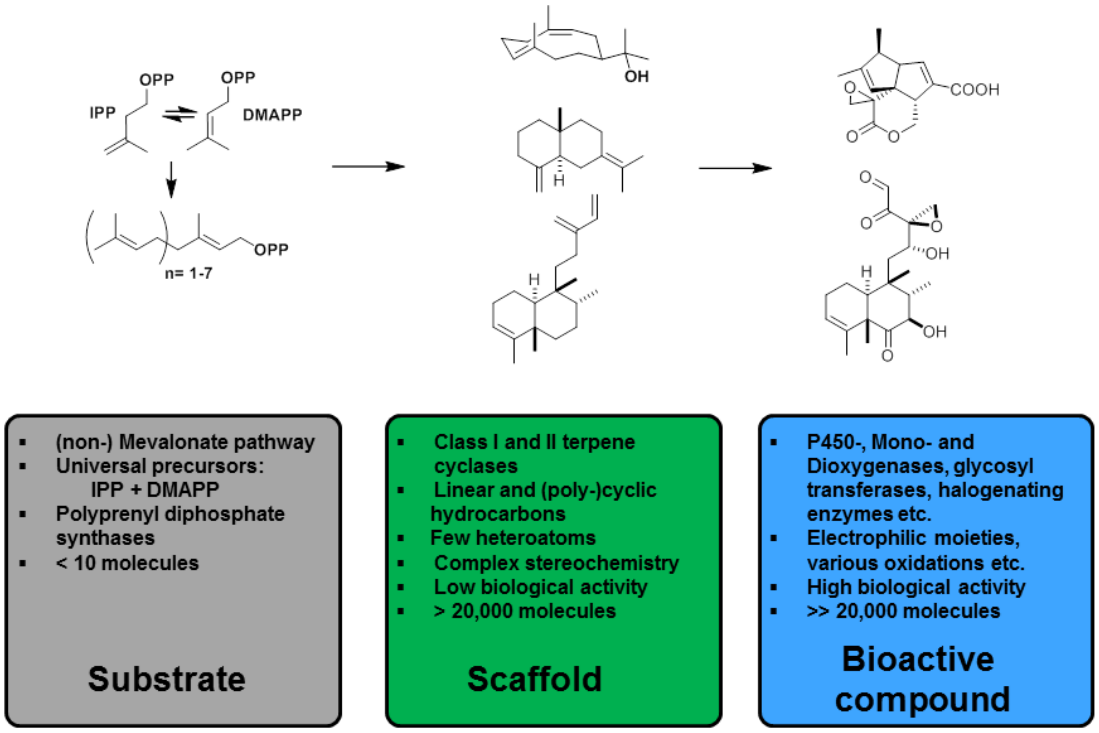


Figure 2. Overview of the terpenoid biosynthesis. The substrate stage (grey) comprises the generation of the universal precursor molecules isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP) *via* the mevalonate- or the 1-Deoxy-D-xylulose 5-phosphate (DXP/ non- mevalonate) pathway[7][8]. These molecules are condensed to form the different polyprenyl diphosphates[9]. In the scaffold stage (green), these linear substrates are converted to linear/(poly-) cyclic terpenes. This chemical conversion is achieved by class I and class II terpene cyclases[10][11]. This way, the compounds' structural diversity is significantly increased. In the last stage (blue), the chemical decoration[12][13] of the scaffold molecules yields bioactive compounds[14]. At this point of the pathway, oxidoreductases are the key players[15].

Terpenoids represent the world's largest group of natural products and can be found among all three domains of life including bacteria, fungi, plants, insects and mammals[16][17]. The biosynthesis of terpenoids can be sub-divided into three stages (modules).

The first module covers the biosynthesis of the universal precursor molecules isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP), which are subsequently converted into polyprenyl diphosphates *via* the corresponding synthases. **Figure 3** shows the two main routes which accomplish the biosynthesis of these two key substrates, the mevalonate pathway (MAD) and the 1-deoxylulose-5-phosphate (DXP) pathway. Since the latter does not existent in mammals it represents a promising target for drug development

(Malaria, herbicides etc.)[18][19]. Moreover, both routes have been genetically engineered in *Escherichia coli* and *Saccharomyces cerevisiae* for increasing the overall yield of various, biotechnologically relevant terpenoids[15][20][21][22][23][24][25][26][27].

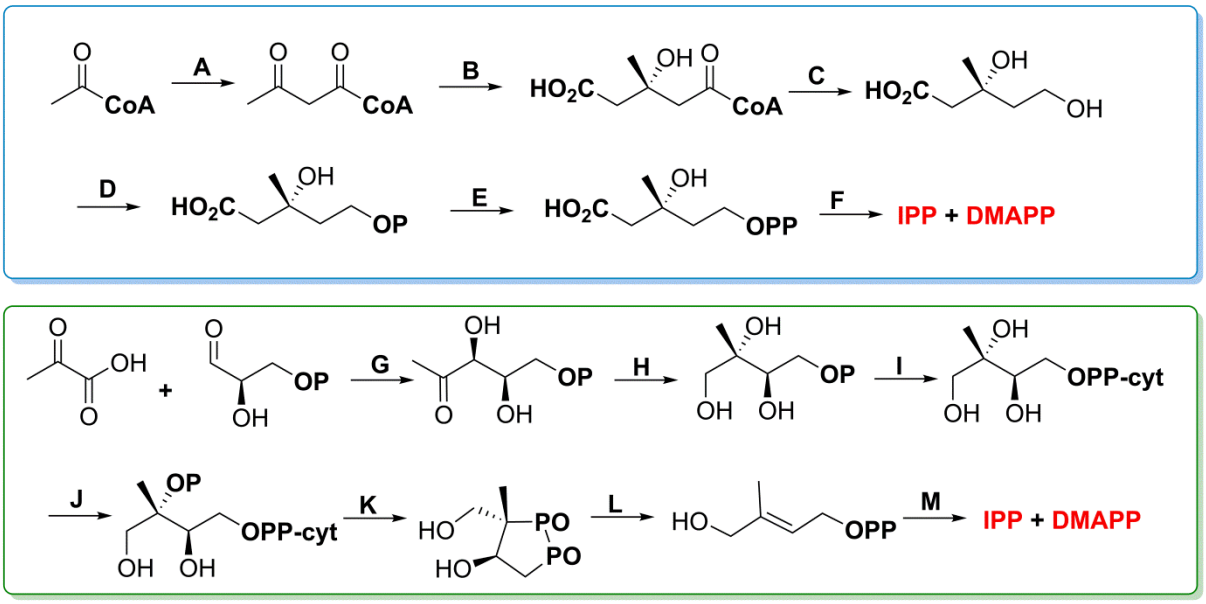


Figure 3. Scheme of the mevalonate- pathway (blue) and the 1-deoxylulose-5-phosphate-pathway (green), both of which generate isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP). The former starts from an acetyl- CoA precursor and comprises six different enzymes (A-F): acetoacteyl-CoA synthase (**A**), hydroxymethylglutaryl-CoA synthase (**B**), 3-hydroxy-3-methyl-glutaryl-CoA reductase (**C**), mevalonate kinase (**D**), phosphomevalonate kinase (**E**) and mevalonate diphosphate decarboxylase (**F**). The latter one utilizes pyruvate and glyeraldehyde-3-phosphate as starting compounds. The conversion into DMAPP and IPP is conducted by seven different enzymes (G-M): 1-deoxy-D-xylulose 5-phosphate synthase (**G**), 1-deoxy-D-xylulose-5-phosphate reductase (**H**), 2-C-methyl-D-erythritol-4-phosphate cytidylyltransferase (**I**), 4-(cytidine-5'-diphospho)-2C-methyl-D-erythritol kinase (**J**), 2C-methyl-D-erythritol-2,4-cyclodiphosphate synthase (**K**), (E)-4-Hydroxy-3-methyl-but-2-enyl pyrophosphate synthase (**L**) and 4-hydroxy-3-methylbut-2-en-1-yl diphosphate reductase (**M**)[20].

In a next step, these universal precursor molecules IPP and DMAPP are converted into polyprenyl diphosphates which is accomplished by the corresponding polyprenyl diphosphate synthases[28]. This class of enzymes comprises two different families, the *E*- branching family (e.g. (2*E*)-geranyl-, (2*E*,6*E*)-farnesyl- and (2*E*,6*E*,10*E*)-geranylgeranyl diphosphate) and the *Z*- branching family (e.g. (2*Z*,6*E*)-farnesyl diphosphate)[29][30]. A structural superposition

(Combinatorial Extension Alignment[31]) of a farnesyl diphosphate synthase (PDB code: 1FPS[9]) and hedycaryol synthase[32] (PDB code: 4MC3[2]) results in a root mean square deviation (RMSD) value of 5.4 (over 216 residues), indicating a close structural relationship. Moreover, the $(Mg^{2+})_3$ coordinating primary sequence motif DDxxD[33] is conserved in both structures. The overall architecture of the two enzymes is entirely based on helices which are connected *via* short loops. This again highlights how structural motives may be rearranged during evolution to generate enzymes which completely differ in their product spectra (in this case the conversion of a farnesyl diphosphate synthase into a farnesyl diphosphate cyclase). A reaction mechanism of geranyl diphosphate biosynthesis (C_{10}) is exemplary shown in **Figure 4**.

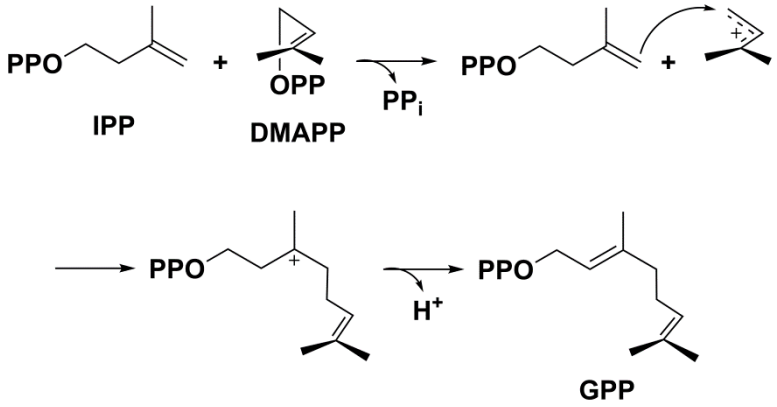


Figure 4. The proposed mechanism of geranyl diphosphate (GPP) biosynthesis, starting from IPP and DMAPP. Similar to class I terpene cyclases, the reaction relies on carbocation chemistry[29].

The third module of terpenoid biosynthesis covers the chemical decoration of the terpene scaffold molecules. By this, individual polarity patterns are introduced to the different molecules, determining the compounds chemical reactivity, binding preferences and overall bioactivity. The key reaction in this process is the introduction of the first heteroatoms, namely oxygens. This is accomplished in the first place by cytochrome P450 monooxygenases, which show the high oxidation potential necessary for activating unreactive hydrocarbons[34]. Next, electrophilic groups such as Michael systems, aldehydes, ketones, epoxides and peroxides are introduced[35][36]. Hydroxyl groups act as proton donors and contribute to the overall steric properties of the compounds. In addition, they are also targeted in downstream modifications such as glycosylations [12]. In summary, the terpenoid class covers almost all possible chemical reactions present in biological systems. Therefore, a profound understanding of the biological oxidation machineries is of great importance for

accessing terpenoids by biotechnological approaches and to implement them as working horses in synthetic chemistry *via* semisynthetic strategies. A great challenge in utilizing these enzymes is the need for their specific redox partners[37]. These are often derived from the cells primary metabolism and their coding sequences are therefore not located close to the different oxidases on the genome[38]. **Figure 5** shows the different modification steps of Artemisinin biosynthesis, starting from α-Amorphene, which is the prime example for the implementation of a terpenoid biosynthesis into a biotechnological process.

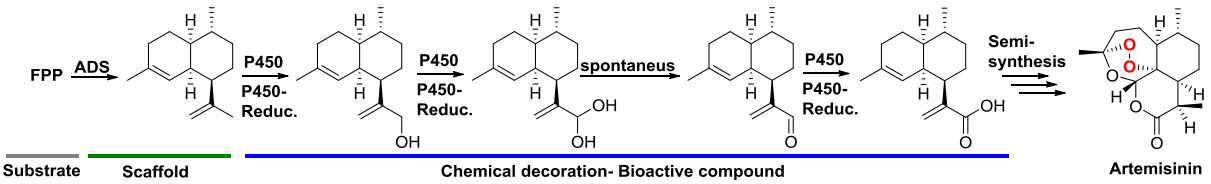


Figure 5. Synthesizing Artemisinin is one of the most successful examples for the implementation of terpenoid enzymes into a biotechnological process. The semisynthetic production of this natural product proves the principal concept of generating a high value (terpenoid-) compound in a biotechnological approach[23][25].

The thesis at hand deals with the second stage of terpenoid biosynthesis, the scaffold stage. Here, the polyprenyl diphosphate substrates are converted into linear-/(poly-) cyclic hydrocarbons which exhibit a complex stereochemistry with little or no heteroatoms. This reaction is catalysed by class I and II terpene cyclases. The most remarkable feature of this reaction step is the conversion of a limited number of educts (~ 10 polyprenyl diphosphates) into a large number of scaffolds molecules (more than 30,000)[16][19]. This remarkable increase in number of distinct molecules is achieved by the highly reactive carbocation chemistry present in class I/II terpene cyclases which is strictly controlled and guided by the enzymes' active sites[10][39][40][41]. In a first step, abstraction of the diphosphate is catalysed enzymatically and the primary carbocation is formed (mesomeric stabilization). Subsequently, an intramolecular nucleophilic attack of a double bond at the C1 position takes place (C1,x cyclisation), resulting in the first intramolecular cyclisation. In the following reaction cascade, which is controlled and guided by the enzymes' active sites, various carbocation intermediates are formed by Wagner/Meerwein[42] and Cope rearrangements[43], hereby giving rise to the thousands of different configurations and conformers. These reactions are ultimately stopped by the elimination of a proton or by addition of a water molecule. **Figure 6** is giving a partial overview of the general mechanism of class I terpene cyclases.

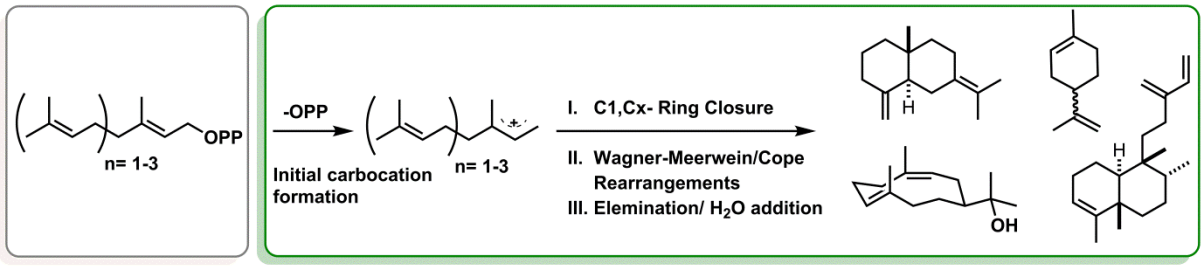


Figure 6. A scheme of the class I terpene cyclases' general mode of action. The substrate stage is highlighted in grey and the scaffold stage is coloured in green.

The polyprenyl diphosphate (grey, substrate stage) is activated by abstraction of the diphosphate, catalysed through the class I terpene cyclase (green, scaffold stage). The primary carbocation is mesomerically stabilized. At this point, the substrate can switch from *trans*- to *cis* conformation. This may be achieved by way of the reaction intermediate nerolidyl diphosphate[44] or spontaneously, as proposed recently[45]. Supported by the enzyme, the first ring closure takes place exclusively at the substrate's C1 atom. Wagner- Meerwein and Cope rearrangements are guided by the enzyme, producing distinct compounds with complex stereochemistry[42][46][43]. The basic mechanistic concept of Wagner- Meerwein and Cope rearrangements is shown in **Figure 7**.

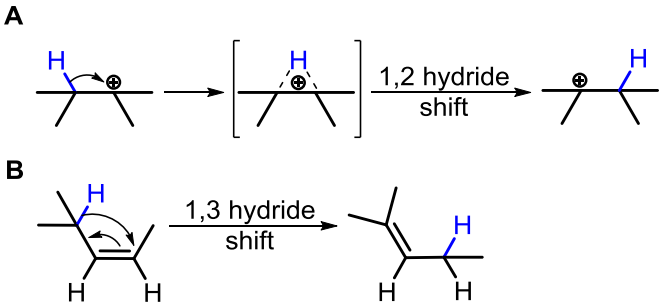


Figure 7. Mechanistic concepts of Wagner-Meerwein- (**A**) and Cope rearrangements (**B**).

The carbocation chemistry is ultimately quenched by an elimination or water addition. The first published structure of a terpene cyclase was the 5-Epi-Aristolochene synthase from *Nicotiana tabacum* (PDB codes: 5EAS, 5EAT and 5EAU)[10]. The overall structure of this $(Mg^{2+})_3$ dependent enzyme displays several α- helices connected *via* short loops. The second terpene cyclase structure solved, this time from a bacterium, was the pentalenene synthase from *Streptomyces exfoliates* (PDB codes: 1PS1, 1HM4 and 1HM7)[47]. A structural

superposition of these two synthases results in a RSMD of 3.8 (over 240 residues) and highlights a common core structure motif, present in all kind of class I terpene cyclases[48]. The central motif comprises 11 α–helices which are connected by short loop regions. However, the overall primary sequence identity between class I terpene cyclases is quite low (< 25%), even though the general topology of this class of enzymes is well conserved among species. Strictly conserved sites of the primary sequence are the $(Mg^{2+})_3$ coordinating residues DDxxD and ND(L,I,V)xSxxxE[49]. A structural key feature present in all class I terpene cyclases is the characteristic helix-break motif located between helix G1 and G2. Upon substrate binding and diphosphate-$(Mg^{2+})_3$ coordination, a closure of the active site takes place, shielding the latter from solvent molecules[50]. It has been demonstrated that within the active site transient carbocations are stabilized by aromatic amino acids. Thereby, the carbocation reaction cascade along a beneficial energy landscape is carried out[49]. Even though many different crystal structures from various class I terpene cyclases (partly in complex with ligands) have been reported since the first structure was published, the understanding of the structure/function relationship is still controversial[51][52][53][54][55].

Class I terpene cyclases are the key players for introducing the structural diversity into the natural product class of terpenoids. This catalytic step represents the first nodal point and the first committed step in terpenoid biosynthesis. The most striking feature of these enzymes is their capacity to generate thousands of different compounds utilizing a limited number of linear educts. Moreover, most class I terpene cyclases demonstrate a unique product specificity, generating just a single compound with a distinct stereo chemistry. The driving forces underlying this powerful chemistry are carbocations, which are formed after diphosphate abstraction. The application and the regulation of this kind of chemistry in aqueous solution is a great challenge which has to be overcome by class I terpene cylcases. Therefore, it is fascinating to understand the structure/function relationship of these enzymes at the molecular level. High resolution crystal structures of these enzymes in complex with ligands (substrate analogues, reaction intermediates and products) might allow insights how class I terpene cyclases can guide and orchestrate highly reactive carbocation chemistry in aqueous solution. Certainly, a thoroughly understanding of class I terpene cyclases will greatly contribute to their application in biotechnological processes, which is the primary focus of the present Ph.D.-thesis.

2. Aim of this work

Class I terpene cyclases are the key players for introducing the exceptional structural diversity into the natural product class of terpenoids. In addition, they represent the first committed step in the biosynthesis of a distinct terpenoid and constitute the first nodal point in biotechnological terpenoid production. Therefore, a fundamental understanding of the enzymes' mode of action is certainly of great interest and would significantly contribute to the field of enzymology and to bio-industrial applications of terpene cyclases. The overall aim of this work is to establish a structure based enzymatic model which explains the carbocation chemistry catalysed in class I terpene cyclases. Thus, understanding the enzymatic mechanisms of these sophisticated enzymes and to contribute to their application in biotechnological processes was the primary focus once started with the project.

The aim of the work at hand was the **Structural and Functional Characterization of Class I Terpene Cyclases**. Hereby, new insights into the structure/function relationship of this class of enzymes should be gained. Therefore, I was instructed to clone, purify, crystalize and characterise two class I terpene cyclases as my Ph.D.-thesis's main project. These class I sesquiterpene cyclases were selina-4(15),7(11)-diene synthase (SdS) and (2*Z*,6*E*)-hedycaryol synthase (HcS). It was expected to investigate these enzymes regarding their ligand binding preferences. For this, thermal shift assays should be conducted. In order to monitor the temperature dependent unfolding process, fluorescence- (for SdS) and circular dichroism spectroscopy (HcS) had to be applied. Furthermore, the enzymes' substrate specificity and their respective product spectra were expected to be analysed conducting *in vitro* assays and GC-MS measurements. Soon, the major focus of my studies turned out to be achieving high resolution mechanistic insights into the class I terpene cyclases' structure/function relationship by producing crystal structures from SdS and HcS in their apo- and their ligand bound forms. In case of HcS, it was attempted to disrupt ongoing substrate conversion during protein purification to capture reaction intermediates within the active site. Since molecular replacement in general does not work for class I terpene cyclases, the implementation of experimental phases were planned from the beginning. Once, first mechanistic models would have been suggested by the crystal structures, experiments were envisioned to create and to analyse point mutants in respect to the enzymes' activity and product spectra. This should proof the proposed mechanistic models in the end. Moreover, it was planned to set up an expression system in *Escherichia coli* to produce terpenoids in vivo.

3. Materials & Methods

In the following section, the materials used and the methods applied will be described.

3.1. Materials

3.1.1. Chemicals

All chemicals used are of microbiological grade, reaction grade or of HPLC grade (fine chemicals) and were not further purified.

Table 1. Chemicals

Chemical	Source	**Chemical**	Source
Acetic acid, 100 %	Roth, Karlsruhe, DE	**Imidazole**	Merck ,Darmstadt, DE
Acrylamide/Bis-solution, 40 %, 29:1	Roth, Karlsruhe, DE	**Isopropyl alcohol**	Merck ,Darmstadt, DE
Agar	Merck,Darmstadt, DE	**Isopropyl β-D-1-thiogalactopyranoside, IPTG**	Sigma-Aldrich ,St. Louis, US
Agarose	Roth ,Karlsruhe, DE	**Kanamycin**	AppliChem ,Darmstadt, DE
Ammoniumperoxodisulfate, APS	Merck ,Darmstadt, DE	**Magnesium chloride hexahydrate**	Merck ,Darmstadt, DE
Ampicillin	AppliChem ,Darmstadt, DE	**2-Mercaptoethanol**	Merck ,Darmstadt, DE
Bromophenol Blue S	Serva ,Heidelberg, DE	**Methanol**	Merck ,Darmstadt, DE
Carbenicillin	Applichem, St. Louis, US	**Pefabloc SC**	Roche ,Risch, CH
Chloramphenicol	Applichem, St. Louis, US	**Peptone**	Merck ,Darmstadt, DE
Coomassie Brilliant Blue R-250	Serva ,Heidelberg, DE	**Sodium chloride**	Merck ,Darmstadt, DE
Ethanol, 96 %	Merck ,Darmstadt, DE	**Sodium dodecyl sulfate, SDS**	Roth ,Karlsruhe, DE
Ethidium bromide	Sigma-Aldrich ,St. Louis, US	**Sodium hydroxide**	Merck ,Darmstadt, DE
Ethylenediaminetetraacetic acid, EDTA	Merck ,Darmstadt, DE	**Tetramethylethylenediamine, TEMED**	Roth ,Karlsruhe, DE
Glycerol, anhydrous	Sigma-Aldrich ,St. Louis, US	**Tween 20**	Merck ,Darmstadt, DE
Glycine, 99 %	Sigma-Aldrich ,St. Louis, US	**Tris,hydroxymethyl-aminomethane, Tris**	Merck ,Darmstadt, DE
(4-)2-hydroxyethyl-1-piperazineethanesulfonic acid , HEPES	Amresco, Ohio, US	**Yeast extract**	Merck ,Darmstadt, DE
Hydrochloric acid	Merck ,Darmstadt, DE		

3.1.2. Molecular biology kits and standards

For molecular biology, the following products were used.

Table 2. Molecular biology kits and standards

Kit	Source	Standard	Source
peqGOLD Plasmid Miniprep I & II	Peqlab, Erlangen, DE	**DNA-Ladder Mix**	Peqlab, Erlangen, DE
peqGOLD Gel Extraction	Peqlab, Erlangen, DE	**Roti-Mark Standard**	Roth, Karlsruhe, DE
peqGOLD Cycle-Pure	Peqlab, Erlangen, DE	**Roti-Mark Prestained**	Roth, Karlsruhe, DE

3.1.3. Protein chromatography

For protein purification, a variety of chromatographic columns were used. These are listed below.

Table 3. Chromatography columns

Device	Source	Device	Source
HisTrap FF crude 5ml	GE Healthcare, Chalfont St. Giles, GB	**BioPro Q30**	YMC, München, DE
Superdex 75 10/300	GE Healthcare, Chalfont St. Giles, GB	**Superdex 200 10/300**	GE Healthcare, Chalfont St. Giles, GB
Superdex 75 16/600	GE Healthcare, Chalfont St. Giles, GB	**Superdex 200 16/600**	GE Healthcare, Chalfont St. Giles, GB
Superose 6 10/300	GE Healthcare, Chalfont St. Giles, GB		

3.1.4. Crystallography

For protein crystallography, different devices were used. These are shown next.

Table 4. Crystallography devices

Device	Source	Device	Source
X8 Proteum in-house beamline	Bruker AXS, Karlsruhe, DE	**Zoom stereo microscope SZX10/KL 1500 LCD**	Olympus, Tokio, JP
Crystallization Screening Suites	Quiagen, Hilden, DE	**SuperClear Pregreased 24 Well Plate**	Crystalgen, New York, US
Glass Cover Slides	Hampton, Aliso Viejo, US	**CrystalCap HT for CryoLoop**	Hampton, Aliso Viejo, US
Mounted Cryo Loop	Hampton, Aliso Viejo, US	**CrystalWand Magnetic**	Hampton, Aliso Viejo, US
Magentic caps and vials	Molecular Dimensions, Newmarket, UK	**Vial Tongs**	Molecular Dimensions, Newmarket, UK
Micro Tool Box	Molecular Dimensions, Newmarket, UK	**Foam Dewers**	Spearlab, San Francisco, US

3.1.5. Technical devices

Table 5 gives an overview of the technical devices used.

Name	Device	Manufacturer
3-30K	Centrifuge	Sigma
1-14K	Centrifuge	Sigma
4-15K	Centrifuge	Sigma
8K	Centrifuge	Sigma
6-16K	Centrifuge	Sigma
DynaPro NanoStar	DLS	Wyatt
NanoDrop2000c	UV-Vis Spectrometer	Thermo Scientific
WTW series	pH-Meter	inoLab
ÄKTApurifier 900	Pump- system	GE Healthcare
ÄKTAprime plus	Pump- system	GE Healthcare
MR Hei-Standard	Stirrer	Heidolph
Thermomixer comfort	1.5 ml tube shaker	Eppendorf
TE124S	Scale	Sartorius
MyCycler	Thermal cycler	Biorad
EPS 600	Electrophoresis power supply	Pharmacia Biotech
G:BOX	UV detection chamber	Syngene
DB 2A	Heating block	Techne
PowerPac Basic	Electrophoresis power supply	Biorad
Phoenix	Sitting drop pipetting crystallization robot	Art Robbins Instruments
Quick-Combi Sealer Plus	Sealing device	HJ-Bioanalytik GmbH
Microlab Star	Buffer pipetting robot	Hamilton
RU MED	Tempered cabinet	Rubarth Apparate GmbH
KL 1500 LCD	Light Microscope	Olympus
Oryx 8	Sitting-, Hanging drop pipetting robot + micro seeding	Douglas Instruments
Digital Sonifier	Ultrasonic cell disruption	Branson
Infors HT	Heating-, shaking cabinet	Multitron
Cell disruption	French press	Constant Cell Disruption Systems

3.1.6. Software

Different software was used for the experiments and for writing the thesis.

Table 6. Software used for research and illustration of results

Software	Source	Software	Source
Adope Photoshop CS4	Adobe, San Jose, US	**PyMol**	www.pymol.org
Adobe Acrobat XI Pro	Adobe, San Jose, US	**Unicorn control software**	GE Healthcare, Chalfont St. Giles, GB
Microsoft Office 2010	Microsoft, Redmond, US	**CCP4 Software Suite**	www.ccp4.ac.uk
Zotero	https://www.zotero.org/	**ApE**	http://biologylabs.utah.edu/jorgensen/wayned/ape/
Tm Calculator	New England Biolabs, Frankfurt a.M., DE	**Double Digest Finder**	New England Biolabs, Frankfurt a.M., DE
Bioinformatics Toolkit	http://toolkit.tuebingen.mpg.de/user/welcome	**Protparam**	http://web.expasy.org/protparam/

3.1.7. Enzymes

TEV- and SUMO protease have been produced and purified in house according standard protocols.

Table 7. Enzymes used in experiments

Enzyme	Source	Enzyme	Source
Restriction Enzymes	New England Biolabs, Frankfurt a.M., DE	**Phusion Polymerase**	New England Biolabs, Frankfurt a.M., DE
T4 DNA Ligase	New England Biolabs, Frankfurt a.M., DE	**Q5 Polymerase**	New England Biolabs, Frankfurt a.M., DE
T4 Polymerase	New England Biolabs, Frankfurt a.M., DE		

3.1.8. Oligonucleotides

All oligonucleotides for PCRs have been purchased from Eurofins MWG (Ebersberg, DE) and Biomers (Ulm, DE).

3.1.9. Plasmids

Different plasmid based expression systems were used for protein characterization.

Table 8. Plasmids used for protein expression

Plasmid	Source	Plasmid	Source
pACYC-Duet	Novagen, Darmstadt, DE	**pET28a**	Agilent, Santa Clara, US
pET-Duet	Novagen, Darmstadt, DE	**pET28c**	Agilent, Santa Clara, US

3.1.10. Bacterial strains

A variety of bacterial strains were used to address different problems of protein production.

Table 9. Bacterial strains used for protein expression

Strain	Source	Strain	Source
Xl1blue	Agilent, Santa Clara, US	**BL21 Star DE3**	Merck, Darmstadt, DE
Bl21 DE3	Agilent, Santa Clara, US	**Rosetta2**	Novagen, Darmstadt, DE

3.1.11. Media

Different microbial culture media were used. The composition of these is listed below.

Table 10. List of microbial culture media

Name	Ingredients	Quantity
LB	Peptone	10 g/l
	Yeast extract	5 g/l
	NaCl	5 g/l
	(Agar)	20 g/l
2TY	Peptone	16 g/l
	Yeast extract	10 g/l
	NaCl	5 g/l
TB	Peptone	12 g/l
	Yeast extract	24 g/l
	Glycerol	4 ml
	KH_2PO_4	0.17 M
	K_2HPO_4	0.72 M
SOC	Peptone	20 g/l
	Yeast extract	5 g/l
	Glucose	20 mM
	NaCl	10 mM
	KCl	0.25 mM
	$MgCl_2$	10 mM
	$MgSO_4$	10 mM

3.1.12. Antibiotics

For plasmid selection different antibiotics were used.

Table 11. Antibiotics used in microbial experiments

Name	Stock (1000x)	Name	Stock (1000x)
Ampicillin	100 mg/ml	**Carbenicillin**	100 mg/ml
Chloramphenicol	25 mg/ml (EtOH)	**Kanamycin**	50 mg/ml

3.2. Methods

3.2.1. Chemically competent cells

For producing chemically competent *E.coli* XL1blue and Bl21 (Star) DE3 cells, a sterile-filtrated **TSS** (transformation and storage solution) solution is used: 85% LB- medium (vol/vol), 10 % PEG 8000 (wt/vol), 5% DMSO (vol/vol) and 50 mM $MgCl_2$ (pH 6.5). First, 5 ml of overnight LB- culture are inoculated. The next day, 200 ml of LB media is inoculated with 2 ml of overnight culture and cells are grown to an OD_{600} = 0.6 in a baffled flask at 37°C. Afterwards, the bacterial culture is immediately cooled down in ice water and harvested at 4°C. The cell pellet is resuspended with 20 ml of ice-cold TSS solution on ice. Subsequently, 100 µl aliquots are made with pre-cooled 1.5 ml tubes and frozen in liquid nitrogen.

3.2.2. Plasmid transformation using chemically competent cells

For plasmid transformation into chemically competent cells, an aliquot of competent cells (100 µl) is thawed on ice. Next, 30 ng of plasmid DNA (for a re- transformation) or 5 µl of T4- reaction mixture are added and incubated for 15 min on ice. This is followed by a heat shock at 42 °C for 45 s. The cells are again put on ice for two minutes and 600 µl of SOC-medium is added. After incubating the cells for 1 h at 37 °C, they are streaked out on agar plates with the appropriate antibiotic.

3.2.3. Polymerase chain reaction (PCR)

For amplification of genes from genomic- or plasmidic DNA a standard PCR protocol and thermocycler program are used.

Table 12. PCR protocol and thermocycler settings

PCR protocol		**Thermocycler settings**		
Substance	**Quantity (µl)**	Step, (°C)	**Time (sec)**	Cycles
H_2O	x (total 100 µl)	**1. Melting, 98**	30	**1**
HF/GC buffer (5x)	20	**2. Melting, 98**	10	**Steps 2-4 x 35**
dNTPS (10 mM)	2	**3. Annealing, 50-72**	10	**Steps 2-4 x 35**
(+ 5% DMSO)	(5)	**4. Elongation, 72**	60/kb	**Steps 2-4 x 35**
Fw Primer (0.1 mM)	0.5	**5. Elongation, 72**	600	**1**
Rv Primer (0.1 mM)	0.5	**6. Storage, 4**	∞	**1**
Template DNA	1 (plasmids, 1 ng), 2-3 (genomic)			
Polymerase	1			

The primers used for the PCR reactions are designed with help of the Ape software, Neb primer designing tool and the Neb builder software. PCR products are purified using the peqGOLD Cycle-Pure kit. In case of a colony PCR, the reaction volumes of 20 µl are prepared. The template DNA is replaced by directly adding a small amount of a single colony to the reaction mixture. Each colony is numbered on the LB agar plate. After running the PCR and checking positive clones by agarose gel electrophoresis, the corresponding colonies can be picked from the LB agar plate and used for inoculation of 5 ml LB overnight cultures.

3.2.4. Plasmid preparation

For plasmid production, an XL1blue *E.coli* strain featuring the plasmid of interest is inoculated overnight at 37 °C in 5 ml LB medium at 160 rpm shaking. The next day, the plasmid is isolated using the peqGOLD Plasmid Miniprep I or II. DNA concentrations are measured with a NanoDrop2000c at 260 nm (E_{260} = 1 ≙ 50 ng/µl).

3.2.5. Agarose gel electrophoresis

PCR products and digested plasmids are analysed by running an agarose gel (1%) electrophoresis. Hereby, DNA samples are prepared by adding 6x DNA loading buffer (+ 1 µ: 30 % glycerol, 0.25 % bromophenol blue). The electrophoresis is performed with 1x TAE buffer (40 mM Tris, 20 mM acetic acid and 1 mM EDTA), at 120 V (constant) for 30 min. Subsequently, the gel is stained in an ethidium bromide solution for 10 min. DNA bands are visualized in a G:Box detection system (365 nm).

3.2.6. DNA digestion

For linearization of plasmids or to generate sticky ends (PCR products), DNA is digested with an appropriate restriction enzyme. The following protocol describes the quantities used for an analytical- and preparative digestion.

Table 13. Protocol for analytical and preparative DNA digestion

Analytical		**Preparative**	
DNA solution	5 µl	**DNA solution**	20-59 µl
H_2O	3.5 µl	**H_2O**	x µl
Enzyme buffer (10x)	1 µl	**Enzyme buffer (10x)**	7 µl
Enzyme	0.5 µl	**Enzyme**	2-4 µl
Temperature	37°C	**Temperature**	37°C
Time	2-3 h	**Time**	24 h

After digestion, the DNA is purified with a Cycle-Pure kit.

3.2.7. DNA ligation

For ligation, a total amount of about 50 ng DNA is used. Hereby, a plasmid/insert ratio of 1:3 or 1:5 yields the best results. The ligation protocol is described next.

Table 14. Ligation protocol

Substance	Quantity (µl)
H_2O	x (total 7.5 µl)
Plasmid DNA	x (~ 10 ng)
Insert DNA	x (~ 40 ng)
10 min, 55 °C ➔ 5 min, 4 °C	
T4 ligase buffer	2
T4 ligase	0.5
2-24 h, 24 °C	

After ligation, 1-2 µl of the mixture are transformed into chemical competent cells.

3.2.8. SLIC cloning

Sequence and **L**igand **I**ndependent **C**loning was first been described by Elledge and co-workers[56]. Based on this work, an individual protocol for integration of genes into DNA vectors has been established. **Table 15** shows two example primer pairs for sequentially integrating two different genes into a pACYC- Duet vector (Novagen).

Table 15. SLIC- Primer

	Gene 1 (NcoI, C'CATG'G)	Gene 2 (NdeI, CA'TA'TG)
Fw-Primer	GTTTAACTTTAATAAGGAGATATACC **ATGTCTGCTGTTAACGTTGCACC**	GTTAAGTATAAGAAGGAGATATACAT **ATGCCGTTTGGAATAGACAACAC**
Rv-Primer	GATGATGGTGATGGCTGCTGCCCATG**TTA ACCAATCAACTCACCAAACAAAAATG**	GCCGATATCCAATTGAGATCTGCCATA**TT ACCAGACATCTTCTTGGTATCTACCTG**

Each primer can be divided into two parts: The first part (highlighted in red) comprises the sequence homolog to the linearized vector. The second part (bold) is the sequence homolog to the gene of interest (this part is used for the PCR). The underlined nucleotides are the restriction sites used. In case of NcoI, the restriction site is destroyed if the base after the ATG- start codon is not "G". Therefore, the use of NdeI as a restriction enzyme always conserves the restriction site since there is no possibility to change the ATG- start codon. These intrinsic features can be used to modify the SLIC cloning procedure in the desired way. If NcoI is used and the restriction site destroyed, it may be re-introduced downstream of the gene of interest to enable utilization of this restriction enzyme again. This way, another gene

can be integrated into the vector with the same restriction enzyme. The applied cloning procedure is described below:

- **a.** PCR reaction with the appropriate primer
- **b.** Purification of the PCR product with the Cycle-Pure kit
- **c.** Linearization of the DNA vector with the appropriate restriction enzyme
- **d.** Purification of the plasmid with the Cycle-Pure kit
- **e.** 30- 50 ng of linearized plasmid is mixed with 150- 200 ng of PCR product, 1 µl NEB2 buffer and 1 µl of BSA (10x). Water is added to reach a final volume of 9.5 µl and the solution is mixed. Finally, 0.5µl of T4-DNA- polymerase is added, mixed with a pipette and incubated for 2 minutes at room temperature. The 3'- DNA digestion is stopped by putting the reaction tube on ice. The reaction volume can be divided in half (5 µl).
- **f.** 5 µl of the reaction mixture are used for the transformation into chemical competent cells

The next day, 5 colonies are used for inoculation of 5 ml LB- media, grown overnight at 37 °C and plasmid DNA is isolated the next day. An analytic digestion with an appropriate restriction enzyme is done and the digestion pattern is analysed on a 1% agarose gel. The positively identified clones are sent for sequencing.

The main advantage of SLIC cloning over classic cloning strategies, which rely on digestion enzymes and ligases, is its robustness. Especially, if one aims to integrate more than one gene into a vector, SLIC cloning routinely yields a much higher percentage of positive clones (90-100%). Moreover, there is no need to purify material *via* an agarose gel at any cloning step, which eliminates the risk of introducing mutations caused by UV radiation. The required vectors may be prepared in advance (digestion for 24 h), which reduces working time significantly. The steps following the PCR reaction can be performed in less than half an hour, followed by the transformation into cells. There is no need for digesting the PCR product. The treatment with T4-DNA-polymerase takes less than 5 min. In summary, SLIC cloning represents a fast, cheap and robust cloning strategy, which is especially useful for cloning several multi- gene constructs in parallel.

3.2.9. DNA sequencing

All DNA constructs were sequenced by GATC (Konstanz, DE) prior usage. Hereby, 20 µl of plasmid DNA was sent to GATC and the appropriate sequencing primers were used. For

aligning the experimental DNA sequence with the theoretical one, the ApE software was used again.

3.2.10. Protein expression

For protein expression, freshly transformed cells or cells from an LB agar plate are used for inoculation of an overnight LB medium culture ($Vol_{preculture}$ = $Vol_{expression\ culture}$/50). The next day, the expression culture is supplemented with the appropriate antibiotics and the preculture is added. Cells are grown at 37 °C and 150 rpm of shaking. When the OD_{600} reaches half of the induction $OD_{600\text{-}induction}$ ($OD_{600\text{-}induction}$= 0.5 for LB medium, $OD_{600\text{-}induction}$= 0.8 for 2TY medium and $OD_{600\text{-}induction}$= 1.5 for TB medium), the temperature is lowered to 20 °C. When finally the $OD_{600\text{-}induction}$ is reached, 0.5-1 mM IPTG are added. The cells are grown overnight. The next day, cells are harvested for 15 min at 5000 xg and 4°C. Ice cold saline (0.9% NaCl (w/v)) is used for the resuspension of the cell pellet. This suspension is centrifuged in 50 ml falcons for 10 min at 5000 xg at 4°C. The supernatant is removed and the falcons are stored at -20 °C.

3.2.11. Protein purification

In a first step, the frozen cell pellets are resuspended in 5x Vol_{pellet} of lysis buffer. Next, 0.5 mM of protease inhibitors and DNase are added. The cell suspension is either disrupted using an ultra-sonic device (largest tip, 30% amplitude, 0.1 s on/off, 3x 1 min, 4 °C) or a cell disruption device (1.6 kBar, 10 °C, 1 passage). The latter one is more suitable for larger cell suspension volumes and less stable proteins. Subsequently, the disrupted cells are centrifuged for 30 min at 40,000 xg and 4 °C. The supernatant is applied with 3 ml/min flow rate *via* an Äkta prime plus on a 5 ml Ni^{2+} HisTrap column which is pre-equilibrated with lysis buffer. Next, the column is washed with low salt buffer to remove salt. Finally, the flow rate is reduced to 1 ml/min and a gradient over 25 ml and a final concentration of elution buffer of 100 % is run. Next, the pooled protein is either dialysed overnight in a low salt buffer, supplemented with sumo protease to remove the sumo tag, or it is directly applied on an anion exchange column. The protein solution is applied with a flow rate of 1 ml/min on an anion exchange column (5 ml Q- sepharose) again running an Äkta prime plus. The column is pre-equilibrated with buffer A. After binding the protein to the column, a gradient over 25 ml and a final concentration of buffer B = 100 % is run. Subsequently, the pooled fractions are concentrated using a centrifugal filter (30 kDa cut off). The final volume should be maximum 500 µl for a 10/300- and 5 ml for 16/600 size exclusion column. The concentrated sample is

applied on the size exclusion column by a loop (500 µl for a 10/300 and 5 ml for a 16/600 size exclusion). The flow rate is used as recommended by the size exclusion's manufacturer. In the following the buffers used are summarized.

Table 16. Protein purification buffer solutions

Buffer	Ingredients
Lysis buffer	50 mM Tris/HCl, pH 7.4, 500 mM NaCl, 25 mM imidazole, 10 % glycerol, 0.02 % Na-azide
Wash buffer	20 mM Tris/Hcl, pH 8.0, 50 mM NaCl, 10 % glycerol, 0.02 % Na-azide
Elution buffer	20 mM Tris/Hcl, pH 8.0, 50 mM NaCl, 400 mM imidazole, 10 % glycerol, 0.02 % Na-azide
Buffer A	20 mM Tris/Hcl, pH 8.0, 50 mM NaCl, 10 % glycerol, 0.02 % Na-azide
Buffer B	20 mM Tris/Hcl, pH 8.0, 500 mM NaCl, 10 % glycerol, 0.02 % Na-azide
Gel filtration buffer	20 mM HEPES/NaOH, pH 7.5, 100 mM NaCl, 10 % glycerol, 0.02 % Na-azide

3.2.12. Polyacrylamide gel electrophoresis (PAGE)

For analysing protein samples and protein expression levels from whole cells, PAGE was applied. Therefore, the protein sample is mixed with 5-fold Laemmli buffer[57] and boiled for 5 min at 95. A 12 % polyacrylamide gel is prepared and put inside the gel electrophoresis chamber and running buffer is added. The PAGE is run at constant ampere (35 mA/gel) for 40 min. Afterwards, staining solution is added to the gel and the liquid is shortly boiled. After 10 min of staining, the staining solution is removed, the gel is washed with water and the de-staining solution is added for at least 10 min.

Table 17. SDS-PAGE buffer solutions

Buffer	Ingredients
Laemmli buffer	200 mM Tris/HCl, pH 6.8, 10 % (w/v) SDS, 10 mM DTT, 20 % (v/v) glycerol and 0.05 % (w/v) bromphenolblue
Running buffer	25 mM Tris/HCl, pH 6.8, 200 mM glycine, 0.1 % (w/v) SDS

3.2.13. Protein concentration

The protein concentration is measured with a NanoDrop2000c at 280 nm. The extinction coefficient ε and the molecular weight are calculated with the software Protparam[58].

3.2.14. Thermofluor based thermal shift assays

Protein crystallography strongly relies on medium- or high throughput crystallization buffer screens, aided by commercial sparse-matrix screens and crystallization robots. Despite these

technological advantages, it is important to keep in mind that approximately 50 % of the initial crystallization setup is based on the protein sample itself. Hereby, parameters like concentration, monodispersity, ligands bound and the protein buffer formulation are a source of variation. A perfect buffer for protein crystallization stabilizes the protein (in crystallization setups, the protein is in destabilizing precipitant conditions), weakens unspecific intermolecular interactions and enhances the protein's overall compactness/rigidity. Thermofluor based thermal shift assays (Thermofluor, Sigma- Aldrich) allow a medium throughput (96- well plate) screening process which requires little protein (4 µg/well), cheap devices (a simple real- time PCT cycler is sufficient) and for which well-designed commercial screens (EMBL-Hamburg) are available. When the optimal protein buffer condition is identified, a second round of ligand/additive screening can be carried out. In principle, this assay is based on the binding of a fluorophore (Sypro Organge) to hydrophobic patches of a protein, generating a fluorescence signal. Since hydrophobic parts are mainly buried inside the folded protein, the starting fluorescence signal is low. The unfolded protein, however, exposes more hydrophobic patches, generating a stronger fluorescence signal. Such unfolding in dependence of a temperature gradient (20- 95°C°) can be followed by the setup shown below (**Figure 8**).

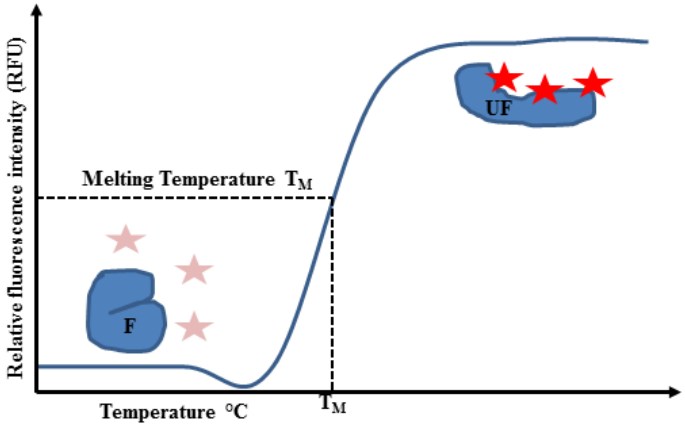


Figure 8. The principal of a thermofluor based thermal shift assay (TSA) is explained. The folded protein (F) covers its hydrophobic patches on the inside; the fluorophore (stars, Sypro Orange) displays only weak fluorescence intensity. A temperature gradient (0.5 °C steps, 20-95 °C) unfolds the protein and a transition temperature (T_M), where 50 % of the protein sample is unfolded (UF), can be determined by applying a sigmoid Boltzmann fit[59].

Every parameter of the buffer (pH, ionic strength, and ligands) will (de-)stabilize the protein and, as a result, a thermal shift to lower/higher melting temperatures can be observed. This rational approach of optimizing the protein buffer and screening for ligands, greatly enhances the probability of successfully crystalizing a certain protein[59].

3.2.15. Dynamic light scattering (DLS)

The preparation of a perfect protein sample is the most challenging task in protein crystallization. After weeks of cloning the gene of interest into a DNA vector, optimizing the expression condition, purifying the protein in a multi-chromatography procedure and identifying the final storage buffer, the protein is concentrated to match the needs for crystallization trials (8- 50 mg/ml). This last step of concentrating the protein sample might be trivial, but it can be a source of significant problems. The intrinsic properties of the protein, the buffer parameters and the process of concentrating the sample itself can lead to an unordered aggregation of the protein. Using such a sample will greatly enhance the amorphous precipitation, thereby lowering the concentration of soluble protein. By this, the probability of forming ordered crystals will be reduced. Therefore, it is worthwhile to analyse the final protein sample shortly before setting up a crystallization screen. Since the protein sample is of great value at this point in time, an analytical method is needed which does not waste protein material. The sample analysis *via* DLS is the perfect solution. Based on the Einstein-Stokes equation, the hydrodynamic radius of proteins can be calculated and a statement about the sample quality can be made (number of protein species, oligomerization, mono- or polydispersity, formation of aggregates). The underlying concept is the time resolved observation of fluctuation in the scattering intensity (positive and negative interference of the coherent and monochromatic laser light). This varies due to diffusion of the particles in solution, the so-called Brownian motion. If the liquid's viscosity is known, the diffusion coefficient can be used to calculate the spherical object's hydrodynamic radius (**Equation 1**)[60][61].

$$D = \frac{K_B \cdot T}{6\pi\eta r}$$

Equation 1. Einstein- Stokes equation: D (diffusion coefficient), K_B (Boltzmann constant), T (temperature), η (viscosity) and r (radius).

The device used for this work is the Wyatt NanoStar. For measuring the dynamic light scattering, sample volumes of 5-20 µl are needed. The minimal concentration for lysozyme is

0.1 mg/ml. In case of larger proteins, the scattering intensity is higher; therefore the protein concentration can be even lower. Prior to DLS measurement and crystallization trials, the protein sample should always be filtrated (pore size = 0.2 µm) using a centrifugal device. Particles which differ by a factor of three or more with respect to their radius can be separately measured by the NanoStar DLS device, yielding two independent hydrodynamic radii. Particles more similar in their size will show a mixed radius of both species. In summary, DLS- measurements allow a rapid and non-invasive method to ultimately check the quality of a protein sample prior to crystallization screens. Hereby, the degree of mono- or polydispersity and the amount of aggregates can provide a first hint for an unsuccessful crystallization attempt. If the first round of crystallization fails it is a rational approach to optimize the protein sample based on the DLS analytics.

3.2.16. Crystallization

For obtaining well diffracting protein crystals, a general procedure described in the following was applied:

1. A monodisperse protein sample with a thermofluor optimized buffer formulation is prepared. The final concentration is between 8-25 mg/ml.

2. For co-crystallization, a ligand solution or the ligand in its solid form is added to the protein sample. The latter approach was used for SdS. The protein:ligand ratio should be around 1:5. After 1-2 h of incubation at 4 °C, the protein sample is filtrated with a centrifugal device (0.2 µm) and once more analysed by DLS.

3. The apo- or liganded protein sample is applied to an initial crystallization screen in a 96-well plate sitting drop approach. Hereby, the initial crystallization screens from Quiagen are used: Classic Suite I, Classic Suite II and JCSG suite. The protein: crystallization buffer ratio used is 0.1µl:0.1µl, 0.2µl:0.2µl and 0.2µl:0.1µl. For pipetting a Phoenix robot is applied.

4. The 96-well initial screen plates are stored at 20 °C and 4 °C. After one hour, the plates are checked for the first time to estimate the percentage of heavy precipitate formation. If more than 80 % of the conditions are precipitated the protein concentration has to be lowered or the protein is not stable enough in general. If the precipitation is below 40 %, the protein concentration has to be increased. In the first

case, a protein buffer optimization with the thermofluor thermal shift assay should be considered.

5. Initial screening hits are further optimized, either in a 96-well sitting drop format or in a 24-well hanging drop vapour diffusion format. In the latter case, two parameters of the initial screening hit condition are varied. For example, in column 1-6 the precipitant concentration is varied and in rows A-D, the pH is altered. Moreover, the volume of the drops is increased in the hanging drop method. The protein:buffer ratio used is: 1µl:1µl, 2µl:1µl and 3µl:1µl. In general, the precipitant concentration should be lowered by approximately 30% compared to the initial hit conditions.

6. After crystal optimization, the crystals are transferred into a 2-4 µl drop containing the crystallization buffer + 30 % glycerol (in general, a suitable cryoprotectant). After 1-3 min of incubation, the crystals are frozen in liquid nitrogen and stored until measurement.

3.2.17. Selenomethionine substituted crystals for SAD methods

If phases cannot be calculated using an existing homolog protein model, they have to be obtained experimentally. The most successful approach to incorporate heavy metal atoms into a protein is to use selenomethionine in place of methionine. Therefore, we produced, purified and crystallized selenomethionine substituted SdS (based on the protocol of Fusinita van den Ent and Jan Löwe).

First, freshly transformed BL21 *E.coli* are grown in 15 ml of 2TY medium (supplemented with kanamycin and incubated at 37°C, overnight). The next day, a 3 l baffled flask containing 1.5 l of M9 medium (160 ml of sterilized 10x M9-medium stock, 16 ml of sterile-filtrated 100x trace elements, 1.6 ml of sterile-filtrated 1000x vitamin mixture, kanamycin (50 µg/ml), 30 ml of sterilized 20% glucose and 3 ml of $MgSO_4$ (1 M stock)) are inoculated with the overnight culture and are grown at 37°C and 140 rpm. At OD_{600} = 0.4, the temperature is lowered to 20 °C and at OD_{600} = 0.6 an amino acid mixture (75 mg L-selenomethionine, 75 mg Leu, Ile, Val and 150 mg Lys, Thr and Phe) to supress endogenous methionine production is added. After 20 minutes, the protein expression is induced by addition of IPTG to a final concentration of 1 mM. The cells are harvested the next day. The selenomethionine protein is purified as described before (cf. 3.2.11.). The only exception is the addition of 5 mM DTT directly after the protein is eluted from the HisTrap-Ni^{2+} column.

3.2.18. $HgCl_2$ substituted crystals for SAD methods

For obtaining experimental phases, a heavy metal soak with native crystals can be done. Depending on the number of certain amino acids, different heavy metal derivatives can be used, e.g. platinum for histidines and mercury for cysteines. In case of HcS, native HcS:**2** crystals are transferred into a 4 µl drop of crystallization buffer and small amounts of solid $HgCl_2$ are added. The crystals are incubated for 2 h. Subsequently, the soaked crystals are transferred into a fresh drop of crystallization buffer to remove excess $HgCl_2$. This step is repeated 2-times. Finally, the crystal is treated with cryoprotectant and frozen in liquid nitrogen[62].

3.2.19. Microseeding

If crystalline precipitate or crystals are obtained in an initial screen, just the first of many steps is achieved. Subsequently, crystals have to be optimized with respect to their size, the packing, their number, their conformation-/oligomerization state and apo- and liganded protein crystals have to be generated. To improve and accelerate this process, it is highly beneficial to produce a seeding stock. For this purpose, 40 µl of the crystallization buffer are transferred into a 1.5 ml reaction tube together with a seeding bead (Hampton research). From this solution, 5 µl are removed and added to the crystallisation drop containing the crystals. With a modified glass Pasteur pipette, the crystals are crashed rigorously. Subsequently, the suspension is completely transferred into the 1.5 ml tube containing the seeding bead. The crystals are further sheared by vortexing the tube for 6x 30 sec, with cooling down the sample on ice between each round. Finally, a dilution series is made up to a factor of 1: 10^6 [63][64]. The seeding stock can be stored at -20 °C over long periods. An appropriate setup using the seeding stock is a mixture of 0.5 µl of protein, 0.3 µl of buffer screen and 0.2 µl of seeding stock, applying the Oryx 8 crystallization robot (Douglas Instruments).

4. Results

4.1. Selinadiene Synthase

Selinadiene synthase (SdS) is a $(Mg^{2+})_3$ -dependent class I sesquiterpene cyclase from *Streptomyces pristinaespiralis* (Uniprot code: B5HDJ6, EC 4.2.3., lyase) which selectively converts farnesyl diphosphate (FPP) into Selina-4(15), 7(11)-diene (Figure 4)[32][1]. Though structural information of bacterial class I terpene cyclases are available for many years (e.g. pentalenene synthase)[47], the structure/function relationship of this fascinating type of enzymes is still unclear. Particularly, the underlying carbocation chemistry conducted in aqueous solution and the enzymes' active site, controlling this process is rarely understood from a structural point of view. SdS has been chosen from a set of bacterial class I terpene cyclases for its high expression rate in *E.coli*, a beneficial prerequisite for protein crystallography. The gene counts 1098 base pairs (68 % GC content) and the protein product comprises 356 amino acids (M_W = 41 kDa, ε = 64,400.00). The corresponding DNA- and amino acid sequence are shown in the appendix. **Table 18** display a primary sequence HHPred alignment (profile hidden Markov model based alignment type, *http://toolkit.tuebingen.mpg.de/hhpred*)[65] of SdS, identifying structurally homolog terpene cyclases. The PDB entry codes are given on the left, whereas the E- value indicates the degree of similarity. This short collection of different structures from various enzyme types already indicates a common topology of these proteins. Key features of the so-called α-fold are its anti-parallel arranged alpha helices (11 alpha helices, annotated A- K) which are connected by short loops. Thereby, a central reaction chamber is formed which can be closed upon coordination of $(Mg^{2+})_3$-PP_i towards the conserved **DD**xx**D**- and the **ND**(L,I,V)x**S**xxx**E** motif[10]. **Figure 9** shows the corresponding primary sequence alignment of these different structures, highlighting the sequence key features (presented with Jalview[66]). Despite this obvious structural relationship of these different enzymes, the primary sequences displays no significant conservation, expect from the two above mentioned motifs. In case of the SdS, a variation for the **DD**xx**D** (^{82}**DD**GH**C**^{86}) exist, whereas the **ND**(L,I,V)x**S**xxx**E** motif (for SdS, ^{224}**ND**IF**S**YHK**E**^{232}) is according to the literature. The primary sequence of SdS features an unique C- terminal elongation (amino acids 310- 365) which cannot be found in the compared terpene cyclases.

Table 18. HHPred alignment of SdS with various terpene cyclases

PDB-Code	Name	Prob	E-vlaue	P-value	Score	Query HMM	Template	HMM
4okz	**Selinadiene syn. (SdS)**	**100**	**1.10E-84**	**3.00E-89**	**623.2**	**1-365**	**1-365**	**365**
4mc3	Putative terpene cyclase	100	2.90E-59	8.30E-64	444.3	1-321	1-327	346
3kb9	EPI-isozizaene syn.	100	1.70E-59	4.70E-64	451.4	1-330	38-381	382
1ps1	Pentalenene syn.	100	1.50E-56	4.40E-61	424.1	1-316	1-320	337
3v1v	2-MIB syn.	100	1.70E-53	4.90E-58	413.5	2-315	107-431	433
4kwd	Aristolochene syn.	100	2.10E-52	6.10E-57	392	6-314	7-312	314
1di1	Aristolochene syn.	100	1.30E-50	3.60E-55	377.5	8-309	2-300	300
1yyq	Trichodiene syn.	100	1.00E-41	2.90E-46	326.2	2-313	19-307	374
4gax	Amorpha-4,11-diene syn.	100	2.30E-37	6.70E-42	307.7	21-315	255-546	563
3g4d	(+)-delta-cadinene syn.	100	8.60E-37	2.50E-41	303.7	21-315	246-537	554
2ong	4S-limonene syn.	100	6.80E-36	1.90E-40	297.2	21-316	235-528	543
3n0f	Isoprene syn.	100	4.50E-36	1.30E-40	298.6	21-297	244-524	555
4xly	Putative terpene cyclase	100	8.20E-35	2.40E-39	270.6	22-275	4-255	300
3m00	Aristolochene syn.	100	9.90E-36	2.80E-40	295.6	21-297	242-521	550
2j5c	1,8-cineole syn.	100	6.90E-36	2.00E-40	297.5	21-298	262-538	569
1n1b	(+)-bornyl syn.	100	2.70E-35	7.90E-40	292.9	21-316	241-534	549
3s9v	Abietadiene syn.	100	6.60E-35	1.90E-39	300.6	20-297	477-757	785
3p5p	Taxadiene syn.	100	4.10E-34	1.20E-38	293.9	20-298	448-725	764
3sdr	Alpha-bisabolene syn.	100	2.10E-34	6.10E-39	298.2	21-298	506-786	817
4lix	ENT-copalyl syn.	99.2	2.50E-11	7.20E-16	123.7	21-250	453-672	727
4omg	Geranylgeranyl syn.	98.9	1.50E-07	4.40E-12	78.5	64-313	92-297	318
4hd1	Squalene syn.	96.9	0.26	7.60E-06	44.9	55-297	31-256	294
3rmg	Octaprenyl-diphosphate syn.	96.8	0.39	1.10E-05	44.7	52-296	54-323	334

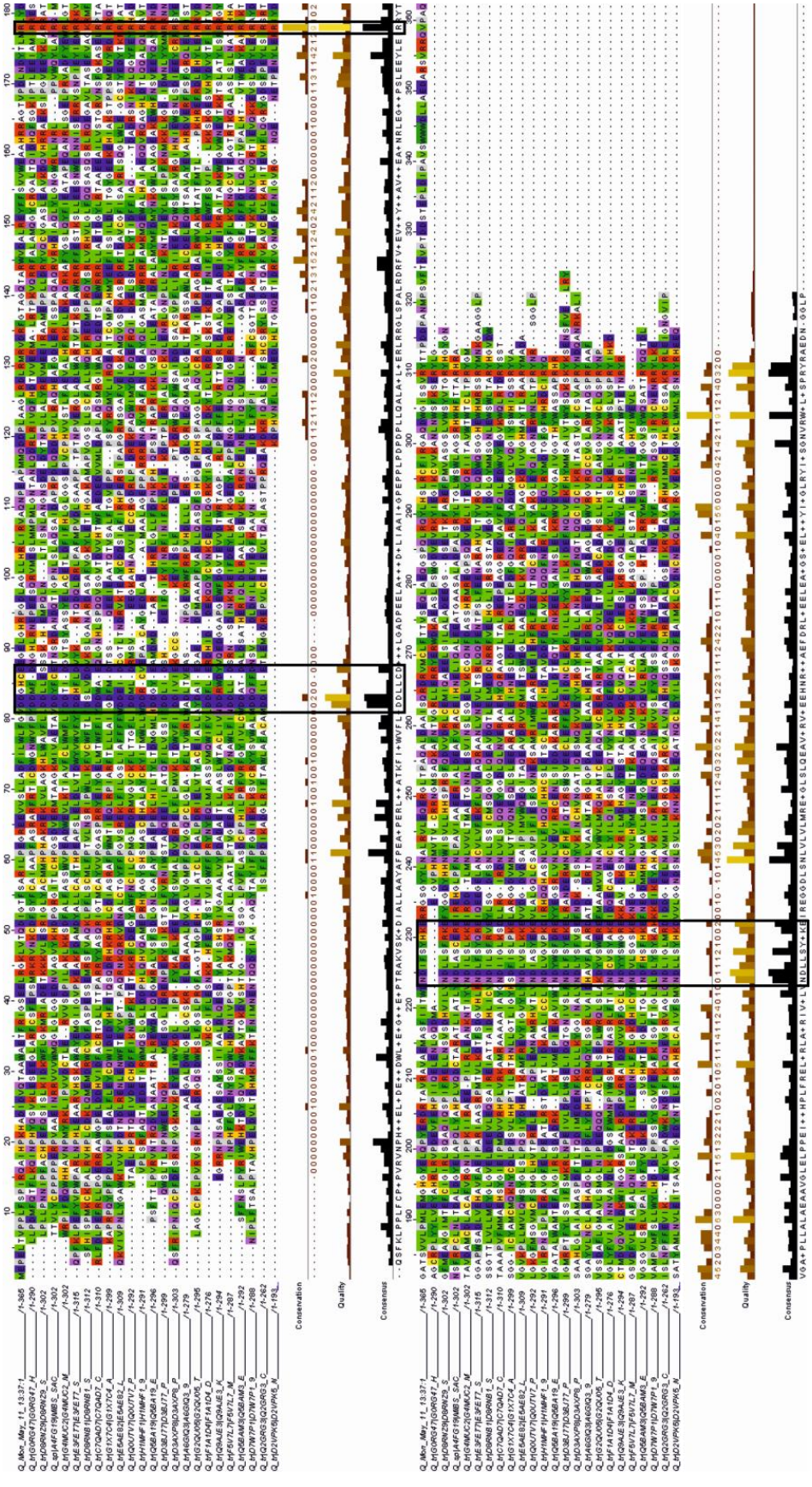


Figure 9. HHPred (protein homology detection by HMM-HMM comparison) based Jalview alignment of the primary sequence of SdS (first line) with various class I terpene cyclases[66].

The conservation, the consensus and the alignments' overall quality are displayed as well. Black boxes indicate highly conserved amino acids. The first box (amino acids ~80- 90) displays the **DD**xx**D** motif, which is central for the $(Mg^{2+})_3$ coordination. The second box highlights an arginine residue (amino acids ~175- 180), one of the few strictly conserved amino acids among class I terpene cyclases with hitherto unknown function. The last box (amino acids ~220- 235) points out the **ND**(L,I,V)x**S**xxx**E** motif, which also participates in PP_i-$(Mg^{2+})_3$ coordination[49].

The alignment of the primary and tertiary sequences of class I terpene cyclases highlights on the one hand the strong conservation of the class I terpene cyclases' 3-dimensional architecture. On the other hand, it points out the weak conservation of the primary sequence, which just comprises two conserved motifs. This peculiar aspect of class I terpene cyclases complicates the bioinformatic analysis by primary sequence based search algorithms.

4.1.1. Cloning and Purification

The SdS gene was cloned into a high copy pET-His_6- sumo vector *via* the restriction enzymes SpeI and PstI, leaving out the starting methionine. The N- terminal sumo tag was used to improve the overall expression rate, to enhance the protein folding and to remove the affinity tag prior to the crystallization trials[67]. After expression of the His_6- sumo- SdS protein, a Ni^{2+} based affinity chromatography yielded pure sumo- SdS (according to SDS- PAGE[68]) with a major band at around 43 kDa and some minor bands at lower molecular weight. Subsequently, the protein has been dialyzed overnight to exchange the affinity chromatography buffer for anion exchange chromatography. Additionally, sumo- protease was added to remove the affinity tag. Subsequently, an anion exchange chromatography was applied to remove DNA, to further purify the protein and to concentrate the sample. The third purification step was a size exclusion chromatography (Superdex 75, HiLoad 16/600), displaying a sharp single peak at an elution volume of 56 ml, thus indicating a monomeric protein. **Figure 10** shows a SDS PAGE of the nickel affinity and anion exchanger purification steps.

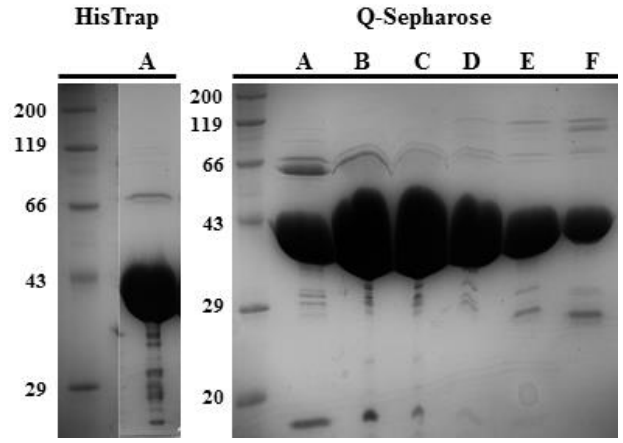


Figure 10. 12% SDS PAGE of the SdS purification by nickel affinity chromatography (left panel) and Q- Sepharose (anion exchange, right panel) chromatography. The molecular weight is indicated by a protein marker (in kDa).

Lane A of the left panel displays a protein sample from the Ni^{2+}imidazole elution fraction (43 kDa). This purification step yielded pure protein. Lanes A- F (right panel) correspond to the SdS anion exchanger elution peak. Though the sample's purity wasn't significantly improved, this second chromatography proofed to be beneficial for decreasing the sample's volume and to remove residual DNA contaminations.

4.1.2. Thermal Shift Assays

The purified protein was investigated with respect to its binding preference towards different oligoprenyl diphosphate ligands (farnesyl diphosphate = **FPP**, *cis*/*trans*-farnesyl diphosphate = **Z,E-FPP**, 2-fluoro-farnesyl diphosphate = **2F-FPP**, diphosphate = **PP_i** and 2,3-dihydrofarnesyl diphosphate = **DHFPP**). For this purpose, thermal shift assays (TSA, thermofluor based) were conducted. The data were analysed using a Boltzmann sigmoidal fit (least squares) with Graph Pad software for determining the melting point T_M. Note, data points before the minimum and after the maximum fluorescence intensity were removed prior to fitting to improve data analysis (**Figure 11**)[59].

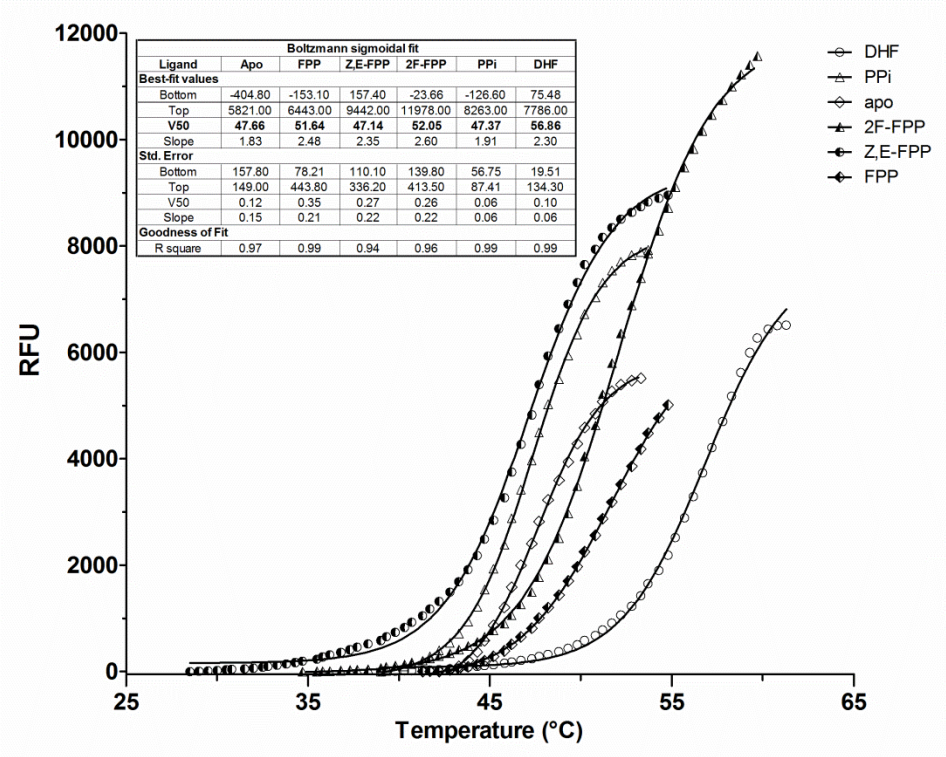


Figure 11. The results of the thermofluor based thermal shift assay. The binding of (*E*,*E*)-farnesyl diphosphate, (*Z*,*E*)- FPP, 2-fluoro- FPP (2F-FPP), PPi and dihydrofarnesyl diphosphate (DHFPP) to SdS were investigated. With a thermal shift of 9.2 °C relative to the apo enzyme, the DHFPP ligand shows the strongest binding of all compounds. For data fitting, a sigmoid Boltzmann fit was applied (GraphPad Prism 5 software). The high R square values indicate the overall exactness of the applied fit[59]. Adapted from Baer et al[1].

According to the observed melting points, PP_i (T_M = 47.4 °C) and *Z*, *E*- FPP (T_M = 47.1 °C) might display low binding preference for SdS (compared to the apo form, T_M = 47.7 °C). In contrast, FPP (T_M = 51.6 °C), 2F- FPP (T_M = 52.0 °C) and DHFPP (T_M = 56.9 °C) look like to possess high affinity to the enzyme. It is noteworthy, that the PP_i apparently does not bind to SdS, though in general the PP_i-$(Mg^{2+})_3$ cluster is heavily coordinated to the enzyme, as shown in previous crystal structures[10]. Presumably, the interaction of the hydrophobic part of the natural substrate (FPP) with the active site is prerequisite for proper binding. The non-binding of the *Z*,*E*-FPP substrate analogue can easily be explained by sterical clashes caused by the alternative conformation of the 2,3-double bond. This highlights the need for exact coordination of the substrate.

4.1.3. Enzymatic Activity

For an enzymatic characterization of SdS purified protein has been incubated with various oligoprenyl diphosphate substrates (dimethylallyl-PP (DMAPP), geranyl-PP (GPP), farnesyl-PP (FPP), geranylgeranyl-PP (GGPP), *Z,E*-farnesyl-PP (*Z,E*-FPP), 2-fluoro-farnesyl-PP (2F-FPP), dihydrofarnesyl-PP (DHFPP)) as described in the *Materials & Methods* section. SdS turns out to be a highly specific class I terpene cyclase, exclusively converting FPP into Selina-4(15),7(11)-diene. **Figure 12** depicts a GC-MS chromatogram of an incubation experiment of SdS with FPP and the underlying reaction mechanism.

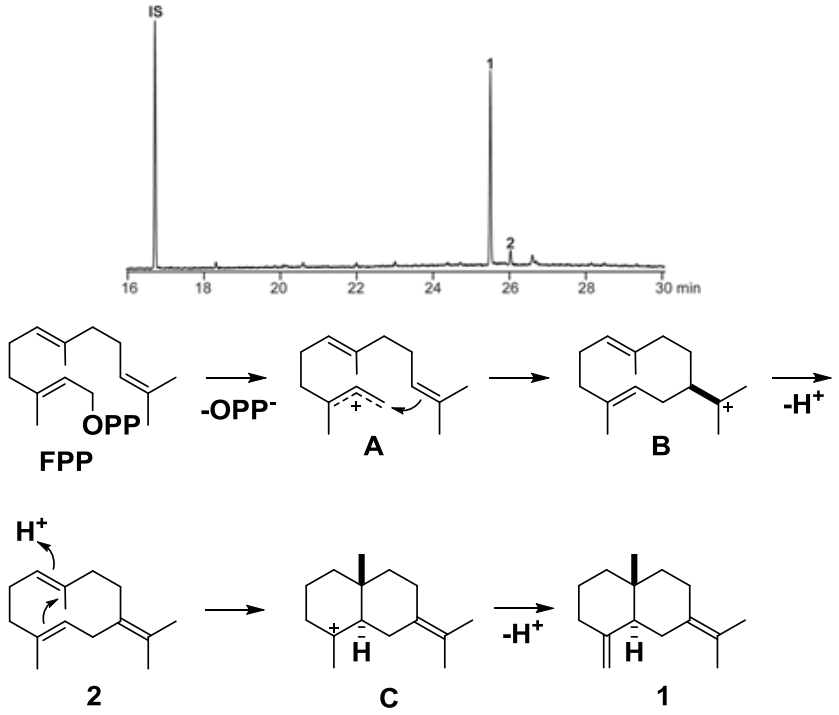


Figure 12. GC-MS chromatogram of purified SdS incubated with FPP. **1** correspondents to selina-4(15),7(11)-diene and **2** to germacrene B. The internal standard (IS) used is dodecane. **A** is the primary carbocation after diphosphate abstraction, **B** is a germacrene B cation, and **C** represents the carbocation intermediate which is ultimately converted into selina-4(15),7(11)-diene. Adapted from Baer et al[1].

The *in vitro* activity assay from SdS incubated with FPP in presence of Mg^{2+} indicates the enzyme's high product specificity. 90 % of the total product is the sesquiterpene selina-4(15),7(11)-diene, whereas 10 % are germacrene B.

4.1.4. Crystal Structure Determination of SdS

Initial crystallization trials were done with wild type SdS as described in the *Materials & Methods* section. In this first round, 96 well plate sitting drop screens (Quiagen) using a Phoenix robot (Art Robbins Instruments) were carried out. Based on initial hits, a fine screen was performed applying the hanging drop vapour diffusion method (SdS:buffer ratio = 2µl:2µl, 15 mg/ml), to optimize the crystallization conditions (**Table 19**).

Table 19. Crystallization conditions for SdS

SdS:PP_i crystals	200 mM $MgCl_2$, 100 mM Tris/ HCl, pH 7.8, 24 % PEG 3350 **+ FPP**
SdS:DHFPP crystals	200 mM $MgCl_2$, 100 mM Tris/ HCl, pH 7.8, 24 % PEG 3350 + **DHFPP**
SdS:Selenomethionine crystals	200 mM $MgCl_2$, 100 mM Tris/ HCl, pH 7.8, 24 % PEG 3350, 200 mM NaCl + **FPP**

Crystals appeared after two weeks at 20 °C. **Figure 13** shows a picture of a representative SdS crystal. The macroscopic crystal appearance was similar for all crystallization conditions.

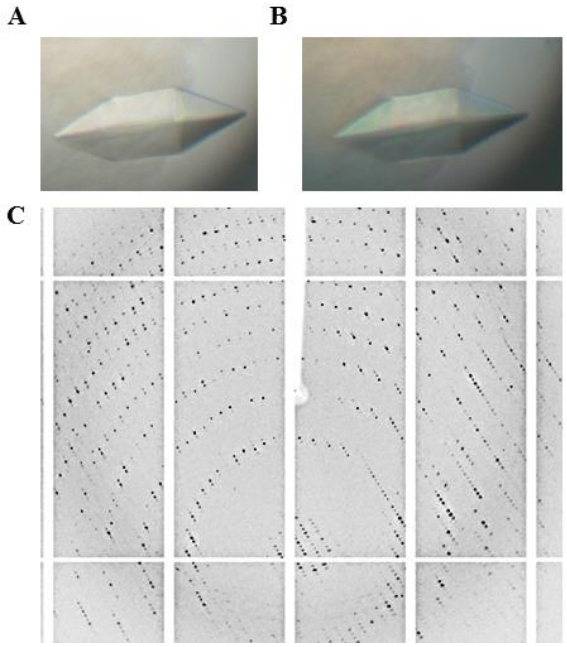


Figure 13. Picture of a SdS:DHFPP crystal in a hanging drop (4µl total volume). A) Normal picture, B) Polarization filter picture and C) A representative diffraction pattern picture.

Crystals were measured at Swiss Light Source synchrotron (Villigen, Switzerland) as described in the *Materials & Methods* section. Native data sets were recorded at 1 Å wavelength. To obtain experimental phases, an anomalous data set was collected using

selenomethionine substituted crystals. The fluorescence scan revealed the Se-absorption peak at 0.9793, thus allowing single-wavelength anomalous dispersion methods (SAD)[69]. The collected data set was processed with the XDS program suite [70], yielding a monoclinic space group $P2_1$ with the cell constants (Å) a = 76.6, b = 121.4, c = 189.6 and β = 90.6°. Solvents predictions and self-rotation functions suggested 8 subunits in the asymmetric unit cell (MOLREP[71]). Subsequently, using SHELXD[72] 72 selenium sites could be positioned at a resolution of 3 Å. These results succeeded in phasing with initial starting phases with SHARP-SAD[73] and solvent flattening with SOLOMON[74]. Next, phases were improved by cyclic 8-fold non-crystallographic symmetry averaging methods which allowed unambiguous fitting of most secondary structure elements by a polyalanine model. Subsequently, the model phases were combined with the experimental phases, hereby visualizing the last missing secondary structures and loop connections. Using the defined Se-sites, side chains could be traced into the $2F_O$-F_C electron density map. The model has been improved in successive rounds using MAIN[75] and was finalized in successive rounds using *Translation Libration Screw-motion* (TLS) parameters and REFMAC5[76]. Having determined the SeMet-SdS crystal structure, the coordinates were used as molecular replacement for calculating the phases of the SdS:**DHFPP** and SdS:$\mathbf{PP_i}$ datasets. Water molecules have been placed automatically running ARP/wARP[77]. The evaluation tool PROCHECK[78] revealed an excellent stereochemistry of the built model. The structural data of SdS:$\mathbf{PP_i}$ revealed three subunits (A, B and C) in the closed conformation (SdS:$\mathbf{PP_i}$:$(\mathbf{Mg^{2+}})_3$). Hereby, the $2F_O$-F_C electron density clearly depicts the $(Mg^{2+})_3$-PP_i cluster after refinement. The packing of the crystal structure probably supports the SdS:$\mathbf{PP_i}$ closed conformation state which is not visible in solution. Surprisingly, the last subunit (D) of SdS:$\mathbf{PP_i}$ displays the enzyme's open apo-conformation. From a crystallographic point of view, this conformation is much more difficult to crystallize since large parts of the molecule are more flexible as compared to the closed conformation. In combination with the three other tightly folded molecules (chains A. B and C), the fourth subunit D could be incorporated into the crystalline lattice as well. The SdS:**DHFPP** dataset was treated according to the SdS:$\mathbf{PP_i}$ structure, resulting in four subunits in the closed conformation, each in complex with DHFPP-$(Mg^{2+})_3$. Again, this ligand could be unambiguously fitted into the $2F_O$-F_C electron density map. This was supported by the asymmetric arrangement of the ligand's methyl groups which exhibit a characteristic appearance in the electron density map. All values of structure determination are given in **Table 20**.

Table 20. Data collection and refinement statistics for SdS. Adapted from Baer et al.[1]

		SdS:PP_i (peak; Se)	SdS:PP_i	SdS:DHFPP
Space group		$P2_1$	$P2_12_12_1$	$P2_12_12_1$
Cell constants (Å)		a=76.6, b=121.4, c=189.6 β=90.6	a=74.8,b=119.1,c=186.1	a=75.1,b=117.8,c=185.6
Anomalous scatterers		72 Selenium	-	-
Molecules in asym. unit		8	4	4
Disordered regions	Chain A		1-3/351-365	1-3/351-365
	Chain B		1-3/350-365	1-3/350-365
	Chain C		1-3/350-365	1-3/350-365
	Chain D		1-6/84-94/230-238/ 308-322/349-365	1-3/351-365
X-ray source		SLS, X06SA	SLS, X06SA	SLS, X06SA
Wavelength (Å)		0.9793	1.0	1.0
Resolution range (Å)[a]		40-2.3(2.4-2.3)	40-2.1 (2.2-2.1)	40-1.9 (2.0-1.9)
No. observations		1053339	314699	576654
No. unique reflections[b]		302273[b]	94771	129570
Completeness (%)[c]		99.6 (99.1)	97 (95.7)	99.6 (99.6)
R_{merge} (%)[a,c]		8.8 (50.3)	7 (48.7)	4.7 (41.8)
I/ δ (I)[a]		12.0 (2.5)	13.9 (2.5)	23.8 (4.3)
Resolution range (Å)			15-2.1	15-1.9
No. reflections working set			89987	123091
No. reflections test set			4737	6479
No. non hydrogen atoms (protein)			10707	11006
No. of heteroatoms: Mg^{2+}			9	12
	Ligand		27	96
	Water		784	1197
R_{work}/R_{free} (%)[d]			15.7/ 20.3	14.3/ 17.7
rmsd bond lengths (Å) / (°)[e]			0.005/ 1.01	0.007/ 1.13
Average B-factor ($Å^2$)				
	Protein		31.4	28.2
	Ligand		24.4	29.9
Ramachandran Plot (%)[f]			99.7 / 0.3 / 0.0	99.9 / 0.1 / 0.0
PDB accession code			**4OKM**	**4OKZ**

[a] The values in parentheses of resolution range, completeness, Rmerge and I/σ (I) correspond to the last resolution shell. [b] Friedel pairs were treated as different reflections. [c] $R_{merge}(I) = [\sum_{hkl}\sum_j |[I(hkl)_j - I(hkl)]|]/ [\sum_{hkl}\sum_j I_{hkl,j}]$ where $I(hkl)_j$ is the measurement of the intensity of reflection *hkl* and $<I(hkl)>$ is the average intensity. [d] $R = \sum_{hkl} | |F_{obs}| - |F_{calc}| |/\sum_{hkl} |F_j|$, where R_{free} is calculated without a sigma cut off for a randomly chosen 5% of reflections, which were not used for structure refinement, and R_{work} is calculated for the remaining reflections. [e] Deviations from ideal bond lengths/angles. [f] Number of residues in favoured region / allowed region / outlier region.

4.1.5. SdS:apo and SdS:PP_i- Complex Structures

The open- and closed conformations of SdS display a typical class I terpene cyclase fold, comprising 11 anti-parallel α-helices (A-K) connected via short loops. Helix A (^{22}HADIDVQTAAWAETF^{36}) is preceded by an 18 amino acids long segment (^{3}ELTVPPLFSPIRQAIHPK^{21}) which shows no secondary structure. Next, a short loop sequence (^{37}RIGS^{40}) follows. Helix B (^{41}EELRGKLVTQDIGTFSARI^{59}) is connected to helix C (^{65}EEVVSLLADFILWLFGVDDGHCEE^{88}) by the loop region ^{60}LPEGR^{64}. In the apo structure, the C-terminus of helix C and the adjacent loop sequence (^{85}HCEEGELGHR^{94}) are structurally not defined. Upon coordination of PP_i-$(Mg^{2+})_3$ to the **DD**xx**D**-motif (^{82}**DD**GHC^{86}) and active site closure, that segment of the primary sequence is getting defined in the electron density map as well. This rearrangement contributes to the structural switch between the open conformation, which is accessible for substrate binding and the closed conformation, which is inaccessible for solvent molecules. Helix D comprises the amino acids ^{95}PGDLAGL LHRLIRVAQ^{110} and is connected to helix E (^{121}PLAAGLRDLRMRVDRF^{136}) by two amino acids ^{137}GT^{138}. Helix F (^{139}AGQTARWVDALREYFFSVVWEAAHRRA^{165}) is connected to helix G1/2 (^{171}LNDYTLMRLYDGATSVVLPMLEMGH^{195}) through the loop ^{166}GTVPD^{170}. Helix G1/2 displays a characteristic helix-break motif (^{182}GAT^{184}) which is structurally strictly conserved in all class I terpene cyclases. Interestingly, this helix-break exhibits major structural rearrangements between the open- and the closed conformation upon substrate binding. Still, its biological function is so far not understood. Helix G1/2 is adjacent to helix H (^{201}PYERDRTAVRAVAEMASFIITWDNDIFSYHKERR^{234}) and connected to it *via* the sequence ^{196}GYELQ^{200}. Helix H features the second conserved primary sequence motif "**ND**(L,I,V)x**S**xxx**E**" (^{224}**ND**IF**S**YHK**E**^{232}). As for helix C, parts of the C-terminus of helix H and the succeeding loop (^{235}GSGYYLN^{241}) are just structurally defined upon substrate coordination (^{230}HKERRGSG^{237}). Next, helix I (^{242}ALRVLEQER^{250}) and helix J (^{254}PAQALDAAISQRDRVMCLFTTVSEQLAEQ^{282}), which are connected to each other by the short loop sequence ^{251}GLT^{253}, follow. Helix K (^{285}PQLRQYLHSLRCFIRGAQ DWGISSVRYT^{312}) is the last secondary structure element. As helices C and H, parts of it are structured just upon ligand coordination (^{307}SSVRYTTPDDPANMPS^{322}), thereby contributing to the active site closure as well. The C-terminal sequence ^{313}TPDDPANMPS VFTDVPTDDSTEPLDIPAVSWWWDLLA^{349} is well defined in the electron density map but exhibits no secondary structure. The first three amino acids ^{1}MEP^{3} and the last sixteen ^{350}EDARSVRRQVPAQRSA^{365} are not defined in the crystal structure. In summary, the SdS:apo conformation features flexible C-termini of helices C, H and K which grant access to

the active site. These elements are structurally defined upon substrate binding, therewith closing the active site (SdS:**PP_i**, **Figure 14** and **15**).

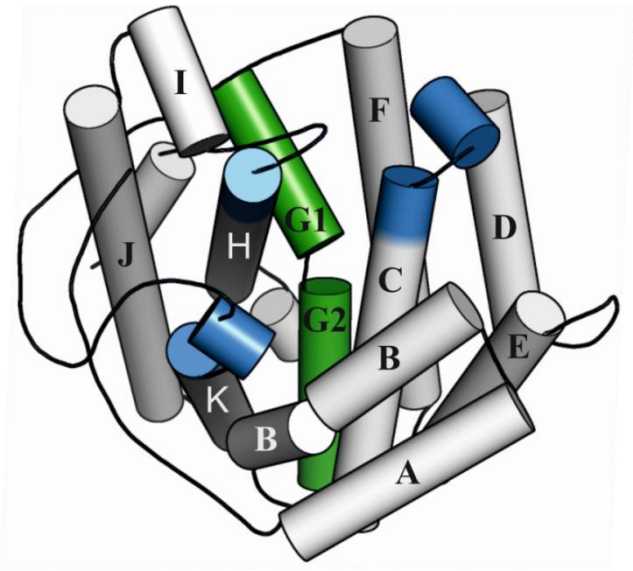


Figure 14. A scheme of helices A- K of SdS:**PP_i**. This α-fold is typical for all class I terpene cyclases. Helix G1/2 and the corresponding helix-break are depicted in green. The blue parts highlight amino acid sequences for which distinct electron density can only be observed upon substrate binding and active site closure.

The active site of SdS is formed by residues located on the different α-helices, which is described in the following: helix B (F55), C (I75, L78 and F79), F (Y152), G2 (V186 and V 187), H (I220), and K (F297, W304 and Y311). Residues involved in the $(Mg^{2+})_3$-PP_i coordination consist of: Helix C (D82, D83 and E87), helix F (E159), helix G1 (R178), helix H (N224, D225, S228, K231 and E232) and helix K (R310) (**Figure 15**).

Figure 15. An alignment of the primary- and secondary structure of selina-4(15),7(11)-diene synthase (SdS)[1]. The dark bars indicate α- helices. Amino acids (one letter code) forming the active site are illustrated in grey, residues coordinating the $(Mg^{2+})_3$ cluster are displayed in blue and amino acids coloured in green are involved in substrate activation. Red arrows mark residues which were targeted by point mutations. Adapted from Baer et al[1].

The active site architecture of class I terpene cyclases as summarized above is highly complex. Therefore, it is quite difficult to assign amino acids to the active site solely based on the primary sequence. This is exemplary illustrated by the short sequence ^{74}FILWLF^{79}. The underlined amino acids are orientated towards the active site, whereas the remaining amino acids are directed outside the catalytic chamber. Every single amino acid of these six residues is hydrophobic and could potentially line the active site. Thus, a prediction of the active site amino acids, solely based on the primary sequence, is hardly possible. Since the function (e.g. product outcome) of an enzyme is always depending on its structure (structure-function relationship), this kind of alignment would greatly enhance the identification of distinct class I terpene cyclases within an organism's genome. For class I terpene cyclases, the only way to link the primary- with the tertiary sequence are to determine the terpene cyclases' structure.

A central structural and functional feature of class I terpene cyclases is the finely tuned coordination of the PP_i-$(Mg^{2+})_3$ cluster, which locks up the active site upon substrate binding. The electron density of the bound PP_i ligand (SdS:PP_i) is depicted in **Figure 16**.

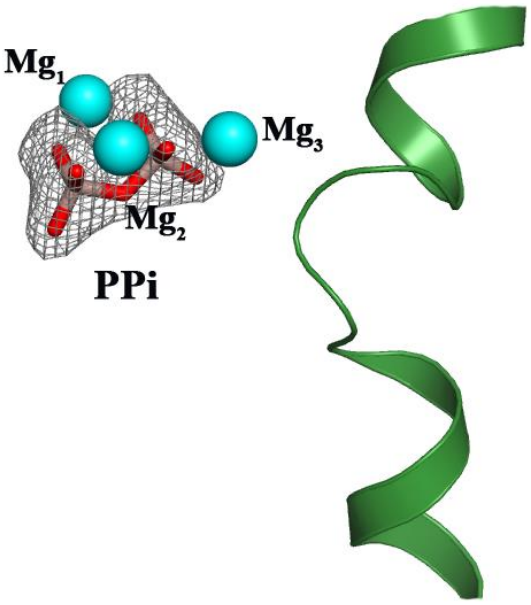


Figure 16. Calculated electron $2F_O$-F_C density map of the ligand in the SdS:**PP_i** complex (contoured at 1σ). For clarity, the electron density map is not shown for the Mg^{2+} ions.

The H-bond network formed upon PP_i-$(Mg^{2+})_3$ coordination is very complex. Hereby, we can distinguish between amino acids which are fixed and structurally defined in the open conformation as well as in the closed one (D82, D83, R178, N224, S228 and E232). The other group of amino acids comprises residues located within the primary sequence segments which are structurally defined only upon substrate binding and thus absent in the SdS:**PP_i** structure (helix C = **E87**, helix H = **K231** and helix K = **R310**, **Y311**). The latter four amino acids can presumably be considered as the driving force of active site closure (**Figure 17**). Hereby, **E87** is directly connected to Mg-1 and Mg-2 by forming two H-bonds (2.1 Å each). Also located on helix C are amino acids D82 (one H-bond to Mg-3, 2.1 Å) and D83 (two H-bonds linked to the hydrate shell of Mg-2, 2.7 Å and 2.8 Å; the other two H-bonds bind to R310, 2.9 Å and 3.0 Å). Located on helix G1, R178 binds to the diphosphate group with two H-bonds (2.9 Å and 3.1 Å) and to D181 with one H-bond (3.0 Å). **K231** connects to the diphosphate moiety *via* one H-bond of 2.8 Å length. Also located on helix H are amino acids N224 (linked to R178 via one H-bond, 3.0 Å and to Y311, one H-bond, 3.2 Å) and S228 (one H-bond to Mg-1, 2.3 Å). **R310** is heavily complexed with H-bonds, two of them being connected with the carboxyl group of D83 (2.9 Å and 3.0 Å) and one with the diphosphate (2.9 Å). By this arrangement, R310 connects helix C with helix K, supporting an orchestrated closure of the active site. **Y311** forms two H-bonds, one with the diphosphate (2.6Å) and another with N224 (3.2 Å). Hereby, a linkage of helix K and H takes place. The third amino acid involved in substrate binding and located on helix K is E232, which binds with one H-bond to Mg-1 (2.1

Å). In summary, the coordination of the PP_i-$(Mg^{2+})_3$ cluster leads to an intramolecular network between helices C, G, H and K. This ultimately results in a concerted rearrangement of the C-termini of helices C, H and K, which tightly close and shield the active site.

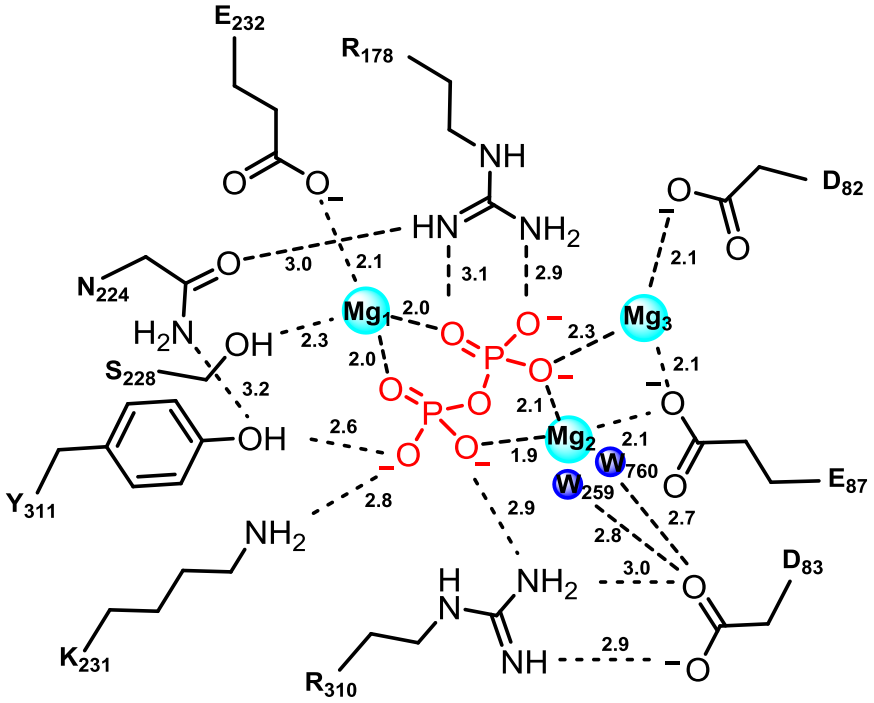


Figure 17. The coordination sphere of SdS:PP_i:$(Mg^{2+})_3$ is displayed. Distances are given in Å. Amino acids are labelled according the one letter code, the Mg^{2+} are coloured in cyan and numbered 1- 3. The two water molecules W_{259} and W_{760}, which are part of the hydrate shell of Mg-2, are coloured in blue.

From a biological point of view, the most pronounced difference between the SdS:**apo** and the SdS:**PP_i** complex structure is the switch between open and the closed conformation. Whereas the former one is accessible for substrate binding, the latter one is a prerequisite for substrate turn over. The formation of the underlying H-bond network and the structural rearrangement described above is the driving force of this process. Next, the formed reaction cavity of SdS:**PP_i** is described in detail. First, a cartoon representation and a surface calculation of the different conformational states are shown in **Figure 18**.

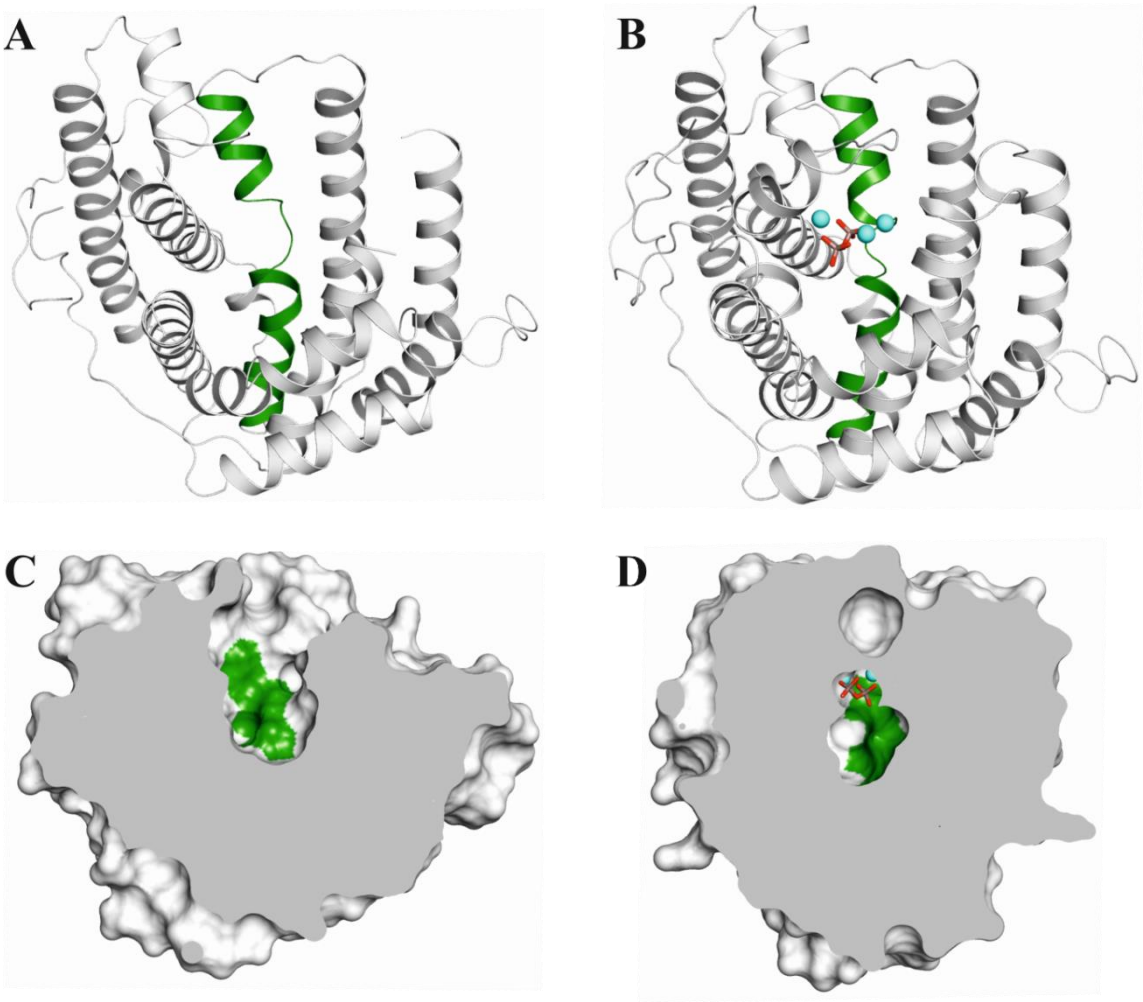


Figure 18. A cartoon (**A**, **B**) and surface (**C**, **D**) representation of SdS. **A**) Shows the open apo- conformation of SdS. Helices G1/2 and the connecting helix break motif are highlighted in green. **B**) Displays the closed SdS:**PP$_i$** complex structure. The three Mg^{2+} ions are coloured in cyan. **C**) The cross section of SdS:**apo** reveals a deep cavity, accessible for substrate binding. **D**) The cross section of SdS:**PP$_i$** reveals a tightly closed active site upon coordination of the PP_i-$(Mg^{2+})_3$ cluster.

The carton representations illustrate the binding of the PP_i-$(Mg^{2+})_3$ cluster (**Figure 18**, **B**) upon the enzyme's central cavity (**Figure 18**, **A**). In addition to this, the corresponding surface calculations **C**, **D** indicate the form and size of the two conformational states' cavities (**C** = open, **D** = closed). It is noteworthy, that this conformational shift is just partly achieved by rearranged amino acids. The crucial structural feature of this active site closure is the bound PPi-$(Mg^{2+})_3$-cluster which in this case acts like a cap. The active site is formed by the following residues: F55, I75, L78, F79, Y152, V186, V187, I220, F297, W304 and Y311.

This perfectly hydrophobic pocket is on the one hand water-repellent; on the other hand it strongly interacts with the hydrophobic hydrocarbon tail of farnesyl diphosphate (FPP) *via* van der Waals forces. By this, the ligand is aligned to the active site after diphosphate abstraction and prefolding takes place. Another remarkable feature of this hydrophobic folding chamber is the distinct arrangement of aromatic amino acids. The active site of class I terpene cyclases is contoured by aromatic residues, which can take part in the carbocation chemistry by stabilizing these reaction intermediates. In case of SdS those aromatic amino acids are F55, F79, Y152, F297, W304 and Y311, as **Figure 19** illustrates.

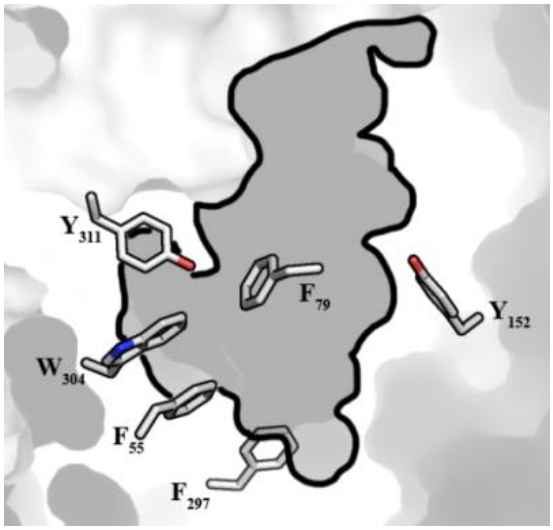


Figure 19. Aromatic amino acids lining the active site of SdS:**PP$_i$** are displayed. The framed section represents the hollow reaction chamber where substrate turnover takes place. The aromatic residues are highly hydrophobic and potentially interact with the substrate double bonds *via* π- stacking. Thus, these residues follow all necessary demands to stabilize transient carbocations during the reaction trajectory.

In summary, the active sites of class I terpene cyclases are mainly formed by hydrophobic amino acids. These residues are distributed among the whole range of the enzymes' primary sequence. The exact form of this folding chamber is thought to control the product outcome. Generally, it is assumed that more promiscuous class I terpene cyclases (e.g. taxadiene synthase and epi-isozizaene synthase) harbour a larger active site, compared to more specific ones[52]. There is a noteworthy number of aromatic amino acids located within the active site of class I terpene cyclases. These amino acids certainly influence the ligand binding by interacting with the substrate's double bonds *via* π-stacking, but they also might stabilize carbocation reaction intermediates with their negative polarity located above and below their

ring plane. By this, substrate conversion can be guided along an energetically favoured reaction pathway.

There are two major kind of structural rearrangements taking place in course of the transition from the open- to the closed conformation. One takes place at the enzyme's active site entrance, as described above. The other one relates to a molecular rearrangement of helix-break G1. This is illustrated in **Figure 20** and described in the following.

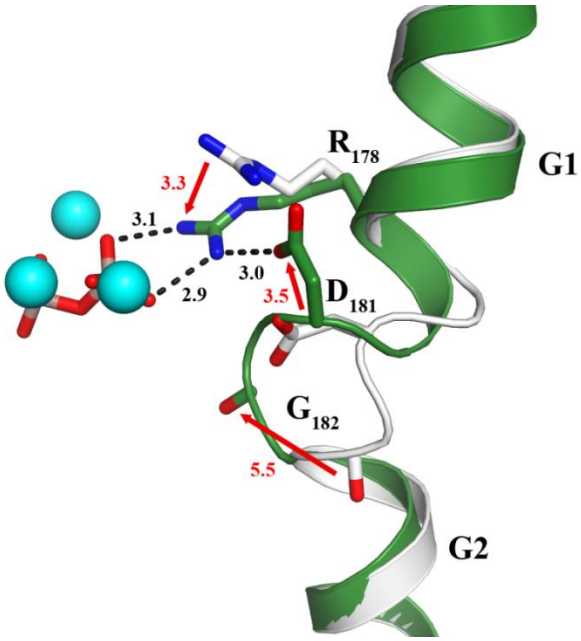


Figure 20. A close- up view of an alignment of helix G1 and its corresponding helix-break motif from SdS:**apo** and SdS:**PP$_i$**. Hereby, SdS:**apo** represents the open conformation, whereas SdS:**PP$_i$** corresponds to the closed conformation. The apo- form is coloured in grey, the closed SdS:**PP$_i$** structure is shown in green. Upon substrate coordination and active site closure, R178 shifts by 3.3 Å, D181 by 3.5 Å and G182 by 5.5 Å. In the course of these structural reorientations, the overall helix-break motif rearranges into an alternative conformation.

The coordination of the PP_i-$(Mg^{2+})_3$-cluster triggers a rearrangement of R178 by 3.3 Å, forming two H- bonds with the diphosphate (3.1 Å and 2.9 Å). In concert with this, the D181 shifts by 3.5 Å, forming an H- bond with R178 (3.0 Å). These structural movements ultimately lead to a reorganization of the overall G1/2 helix-break motive, shifting the carbonyl group of G182 by 5.5 Å. Interestingly, the backbones of helices G1 and G2 are not altered by this shift, as disclosed in the structural alignment. This substrate triggered molecular rearrangement does not only contribute to the coordination of the diphosphate, but

places the carbonyl group of G182 within the active site. This kind of conformational switch has never been observed in any class I terpene cyclase structure before. In order to evaluate the conservation of these key amino acids and to estimate its biological importance, an sequence alignment of G1 helix and its corresponding helix-break motif from 200 different bacterial terpene cyclases was conducted[79]. The results are displayed in **Figure 21**.

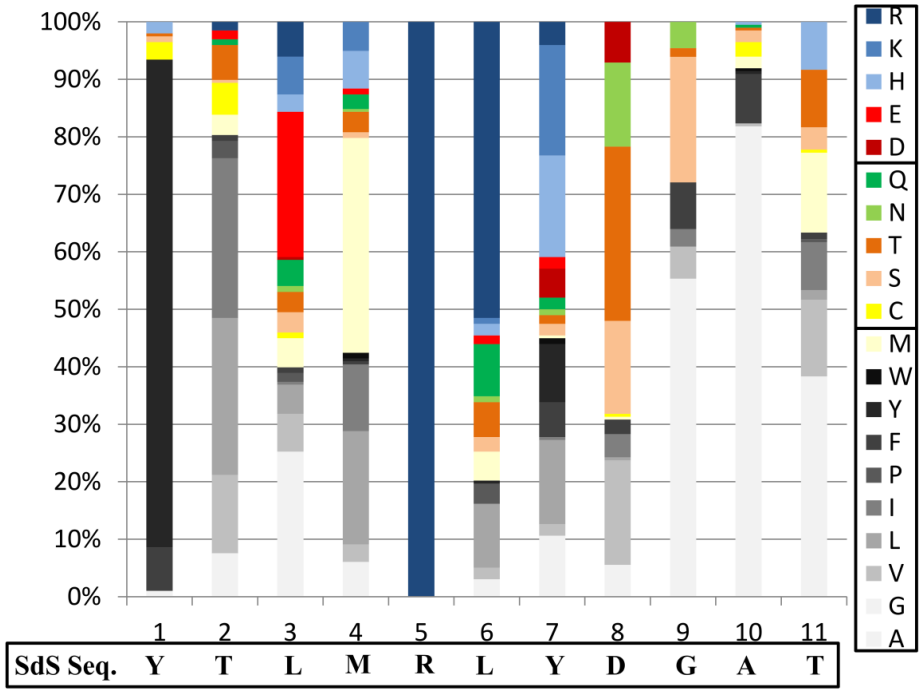


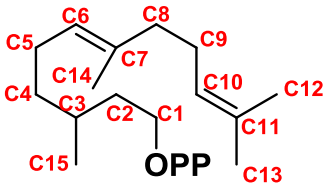
Figure 21. An alignment of the primary sequence of SdS helix G1 and its corresponding helix-break (^{174}YTLMRLYDGAT^{184}) with 200 bacterial sequences (SdS BLAST[80] search, OMEGA alignment, Uniprot[81][79]). Each amino acid is coloured differently and every column amounts to 100 % sequence conservation in total. R178 (dark blue, position 5) is strictly conserved (100 %), whereas the site of the linker residue D181 (position 8) displays an amino acid capable of H- bond formation in up to 70 % of the sequences. Positions 9 and 10 (G182 and A183) mainly display small and hydrophobic amino acids[1]. Adapted from Baer et al[1].

In order to investigate the biological importance of the three amino acids (R178, D181 and G182) involved in the structural rearrangement of the helix-break motif in G1, an alignment (OMEGA alignment[79]) of the SdS sequence ^{174}YTLMRLYDGAT^{184} with 200 bacterial class I terpene cyclases (BLAST search, Uniprot[81]) was performed. Hereby, Y174 is in 94% of all investigated primary sequences conserved which is the second highest value, after R178 (100% conservation). Interestingly, the hydroxyl group of this residue forms a H-bond (2.5 Å) to D225 from the **ND**(L,I,V)x**S**xxx**E** motif upon substrate binding and active site closure.

Thereby, helix G1 is linked to helix H which contributes to the active site closure, too. The next position (T175) is in 80% of all cases occupied by a hydrophobic amino acid of variable size. L176 is weakly conserved. At the position of M177 80% of all amino acids are hydrophobic, with 40% of them being a methionine. R178 is the only strictly conserved amino acid in this primary sequence segment. Its triple coordination upon substrate binding (two H-bonds with PP_i and one with D181) highlights the structural importance of the guanidine moiety, which can't be replaced by any other amino acid. This indicates that also the H-bond formation with D181 is of great functional importance for class I terpene cyclases. The position of L179 and Y180 features in 40% of all cases a positively charged amino acid like arginine or lysine. This positive charge probably stabilizes the negative polarity of helix G1's C-terminus, which originates from its helix dipole. At the site of D181, most class I terpene cylcases (70%) possess an amino acid which is capable of H-bond formation. This is important for linkage with R178 upon substrate binding. The site of G182 marks the start of the helix-break motif. In 55% this position is occupied by a glycine or alanine. The adjacent amino acid position (A183) exhibits an even higher degree of conservation (80%) of these two residues. The last amino acid from the helix-break motif is T184. Again, in 40% of all cases a glycine or alanine residue is located here. The latter three amino acids are forming the helix-break. As we demonstrated, this structural feature is highly conserved among class I terpene cyclases. In course of substrate binding, a molecular rearrangement of this helix-break takes place, which requires a certain structural flexibility. This explains the high frequency of glycine and alanine at this position. As shown in the alignment, the structural importance of Arg178 is underlined by its strict conservation. In contrast, amino acids Asp181 and Gly182 alter, which is caused by the requested structural/chemical function required from these residues, e.g. H- bond formation (Asp 181) and the presence of carbonyl oxygen (Gly182), properties provided by many amino acids.

4.1.6. DHFPP- Complex Structure

In order to understand the observed molecular rearrangement of helix-break G1 upon substrate binding, we aimed to co-crystallize or soak SdS:**PP_i** crystals with a FPP substrate analogue. Through this, we were curious to see if some kind of interaction between the helix-break in the closed conformation and the substrate's hydrocarbon backbone takes place. Since soaking of SdS:**PP_i** crystals failed, we performed co-crystallization with the FPP analogue dihydrofarnesyl diphosphate (DHFPP, **Figure 22**). After several attempts, we succeeded in obtaining well diffracting SdS:**DHFPP** crystals. In contrast to the SdS:**PP_i** crystals, all four subunits of SdS:**DHFPP** displayed the closed conformation. Each subunit was in complex with DHFPP and $(Mg^{2+})_3$.



Dihydrofarnesyl diphosphate

Figure 22. The chemical structure of dihydrofarnesyl diphosphate (DHFPP). Red numbers label the ligand's carbon atoms.

DHFPP acts as an inhibitor for SdS, because it is missing the first double bond located between C2 and C3 compared to the natural substrate FPP. The carbocation which is formed upon diphosphate abstraction cannot be mesomerically stabilized and therefore, the enzymatic conversion of this compound by the class I terpene cyclase is prohibited. Another comparable inhibitor, 2-fluoro-farnesyl diphosphate (2F-FPP), which arrests the enzymatic activity of SdS possibly in the Michaelis complex, was tested with the thermal shift assays regarding its binding properties, as well. Here, the 2,3-double bond is kept but the electron withdrawing fluorine group prevents a diphosphate abstraction. The main difference between these two substrate mimics is the first double bond which is conserved in the latter. Therefore, the characteristic planarity of the original substrate is kept in 2F-FPP. Notably, DHFPP inherits a higher degree of rotational freedom between the diphosphate moiety and the first three hydrocarbons (C1, C2 and C3). These rotational less restricted structural features of DHFPP resulted in a significantly stronger binding towards SdS, as demonstrated in the thermal shift assays (52°C for 2F-FPP and 57°C for DHFPP). This might be one explanation for the successful co-crystallization of SdS with DHFPP and its failure with 2F-FPP.

The SdS:**DHFPP** complex structure is almost identical compared to the SdS:**PP_i** structure, except their bound ligands which electron density is shown for DHFPP in **Figure 23**.

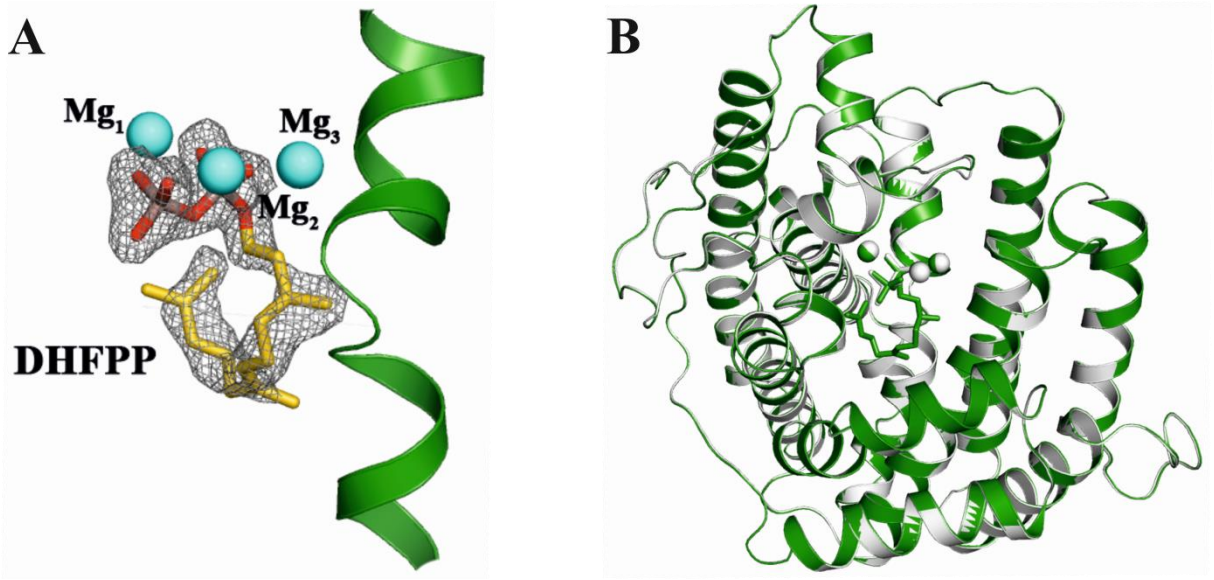


Figure 23. A) The $2F_O$-F_C electron density map of SdS:**DHFPP** is shown for the ligand (contoured at 1σ). For clarity, the electron density map is not shown for the protein and the Mg^{2+} ions. **B**) A structural superposition of SdS:**DHFPP** (green) and SdS:**PP_i** (grey) is depicted. The closed conformations are almost identical with an r.m.s.d of < 0.2 Å of C_α-atoms.

The SdS:**DHFPP** complex structure could be determined at a high resolution of 1.9 Å. The occupancy of the ligand is close to 100%, thus allowing an unambiguously assigning of the asymmetric distributed methyl groups of the ligand.

The coordination sphere of the $(Mg^{2+})_3$-PPi moiety is identical to the one observed for the SdS:**PP_i** complex structure (**Figure 17**). The additional interactions between the hydrocarbon backbone of the DHFPP inhibitor and SdS are depicted in **Figure 24**.

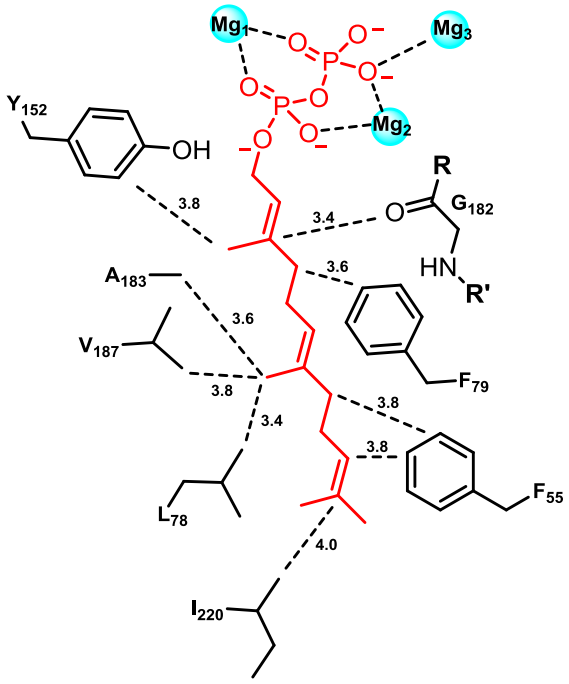


Figure 24. The interactions between the active site residues from SdS and DHFPP are illustrated. Distances (dashed lines) are given in Å. For clarity reasons, the coordination of the $(Mg^{2+})_3$-PP_i cluster towards SdS is not shown.

The hydrogen carbon backbone of DHFPP interacts with the following active site residues *via* van der Waals forces: F55, L78, F79, Y152, A183, V187 and I220. Hereby, F55 interacts with the ligand's C8 and C10 positions (3.8 Å each). Amino acid F79 points towards the C4 position (3.6 Å). The substrate's methyl group C14 is stabilized by amino acids L78 (3.4 Å), A183 (3.6 Å) and V187 (3.8 Å). Y152 is in close contact to C15, the first methyl group (3.8 Å). I220 is 4.0 Å away from the substrates C11 carbon atom. Interestingly, the only non-hydrophobic interaction between the DHFPP ligand and SdS is the carbonyl group of G182, which is 3.4 Å distant to the ligand's C3 position. It is noteworthy to mention that the SdS:**DHFPP** complex represents the Michaelis substrate-enzyme complex. In this binding state, the substrate is still anchored to the entrance of the active site due to the diphosphate moiety. After abstraction of this chemical group, the entire hydrophobic hydrogen carbon backbone of the substrate can bind deeper into the active site. Therefore, not all amino acids important for catalysis within the active site are in close contact or in the correct orientation towards the substrate in this binding state.

4.1.7. SDS- Mutants and Product Spectra

In order to prove our mechanistic models and to rationally change the product spectra, we designed 28 different mutants. The corresponding protein variants were purified and the enzymatic conversion of FPP was analysed with respect to their individual product spectra. The purification of the SdS mutants and the corresponding product spectra analysis was conducted in joint collaboration with Patrick Rabe from the group of Prof. Dr. Jerome Dickschat at the University of Bonn. The different products of the mutants' and their proposed reaction mechanism are shown in **Figure 25**.

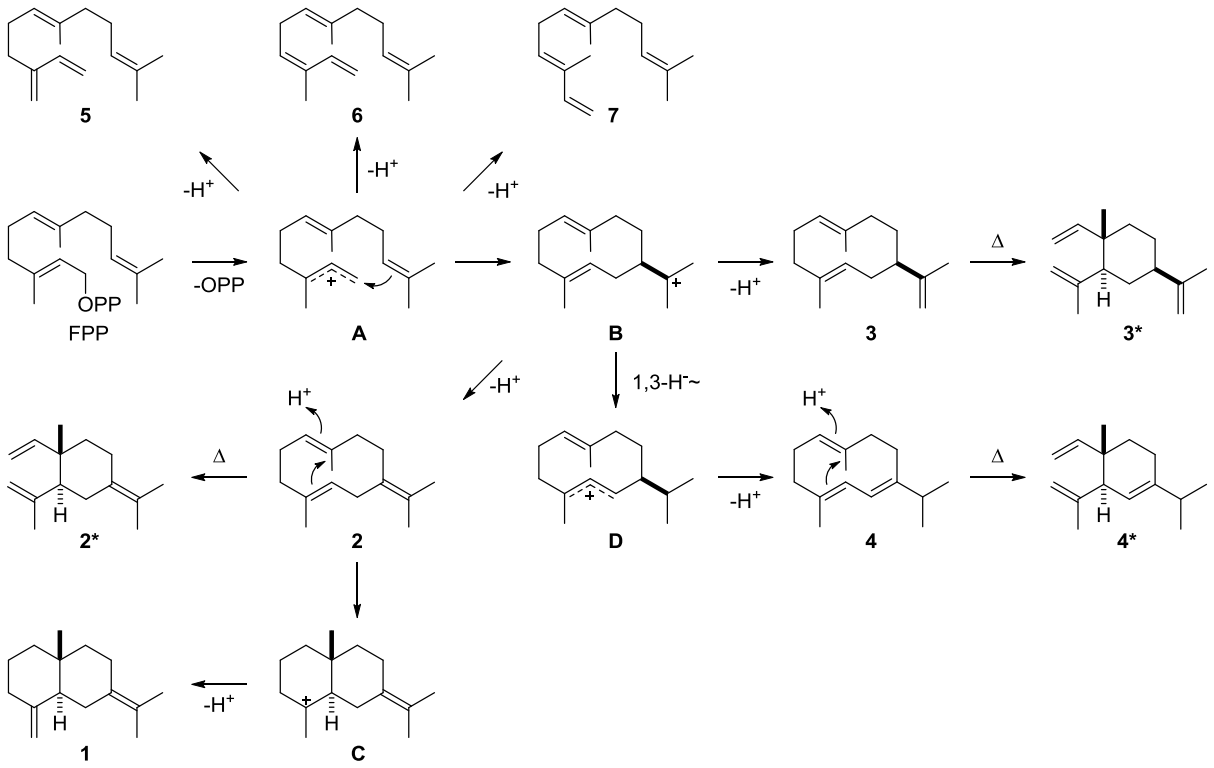


Figure 25. Biosynthesis of **1** and side products of SdS. **1**: Selina-4(15),7(11)-diene, **2**: germacrene B, **2***: γ-elemene, **3**: germacrene A, **3***: β-elemene, **4**: δ-elemene, **5**: (*E*)-β-farnesene, **6**: (2*Z*,6*E*)-α-farnesene, **7**: (2*E*,6*E*)-α-farnesene. Compounds **2***, **3***, and **4*** are artifacts (Cope rearrangement products of **2** – **4**) formed during GC/MS analysis. Adapted from Baer et al[1].

Wild type SdS and mutants abstract in the first step of catalysis the diphosphate from the FPP substrate, forming the primary carbocation **A**. By different elimination reactions, this intermediate can be converted into (*E*)-β-farnesene (**5**), (2*Z*,6*E*)-α-farnesene (**6**) and (2*E*,6*E*)-α-farnesene (**7**). These distinct products are formed depending on the proton eliminated. In case of wild type SdS, carbocation reaction intermediate **A** is transformed into the germacrene

B carbocation **B**, by a nucleophilic attack of double bond C10,11 towards substrate position C1. From a thermodynamic point of view, the mesomerically stabilized carbocation is preferentially localized at the higher substituted C3 position. Therefore, the double bond's attack at the C1 position (for terpenoids, this first attack always takes place at C1) has to be catalysed by SdS, exclusively forming the anti- Markovnikov product. Next, intermediate **B** can be converted into **3** by an elimination of a proton at the C12 or C13 position. An alternative reaction pathway of **B** is its conversion into **D**. This is achieved by a 1,3-hydride shift (Cope rearrangement) between C11 and C2, which is followed by a further elimination reaction, yielding δ-elemene (**4**). In course of **1** formation, **B** is transformed into germacrene B (**2**) upon deprotonation (at C10). Subsequently, the C6,7 double bond is reprotonated and carbocation **C** is produced. A final elimination reaction (C15) yields the wild type SdS's main product, selina-4(15),7(11)-diene (**1**). This reaction mechanism of FPP conversion into Selina-4(15),7(11)-diene by SdS is a prime example of terpene biosynthesis. All class I terpene cyclases follow this general reaction pathway, hereby generating thousands of different products from a view substrate molecules.

In order to investigate the role of key amino acids and to proof mechanistic models, we aimed to rationally modify the enzyme. Moreover, we attempted to alter the product spectra by this. From a biological chemistry point of view, this latter approach should be feasible since the underlying carbocation chemistry is highly reactive and the structural diversity is a key component of terpenoids. The mutants analysed can be divided into three different groups: **1.**) Amino acids involved in $(Mg^{2+})_3$-PP_i coordination (Asp83, Glu159, blue), **2.**) Residues acting in the induced-fit mechanism, which will be explained later on in the Discussion section (Arg178, Asp 181,Gly 182, Ala183 and Tyr152, green) and **3.**) Amino acids contouring the active site and guiding the carbocation chemistry (Phe55, Phe79, Trp304 and Tyr311, black). **Figure 26** gives an overview of the mutated amino acids and their relative positions within the enzyme.

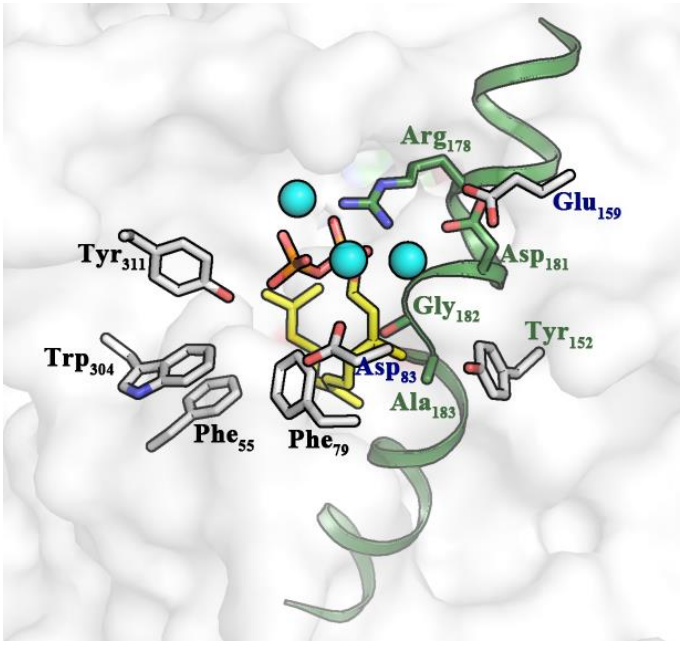


Figure 26. A scheme of the mutated amino acids in SdS, relative to the DHFPP-$(Mg^{2+})_3$ ligand and helix G1/2. Labels for residues coordinating the $(Mg^{2+})_3$-PP_i cluster are shown in blue, side chains involved in the induced-fit mechanism are indicated in green and aromatic amino acids forming the active site are highlighted in black.

Group 1 comprises amino acids which coordinate the $(Mg^{2+})_3$-PP_i cluster. All four mutants showed a significant decrease in the GC-MS signal for compound **1** (Selina-4(15),7(11)-diene, main product of wild type) and a strong increase of **2** (germacrene B). The mutants investigated were: D83N, D83E, E159Q and E159D (**Figure 27**).

Target	Mutant	Activity[a]	1[b]	2	3	4	5	6	7	Group
wt		100	89.7	10.3	-	-	-	-	-	
Asp83	D83N	34	0.9	96.4	0.7	1.9	-	-	-	1
	D83E	203	6.0	86.4	3.1	4.6	-	-	-	
Glu159	E159Q	21	2.2	80.0	10.8	6.9	-	-	-	1
	E159D	38	1.5	87.8	3.3	7.3	-	-	-	

[a]Activity relative to wild type SdS. [b] Selina-4(15), 7(11)-diene. Red fields highlight the main product of the distinct mutant.

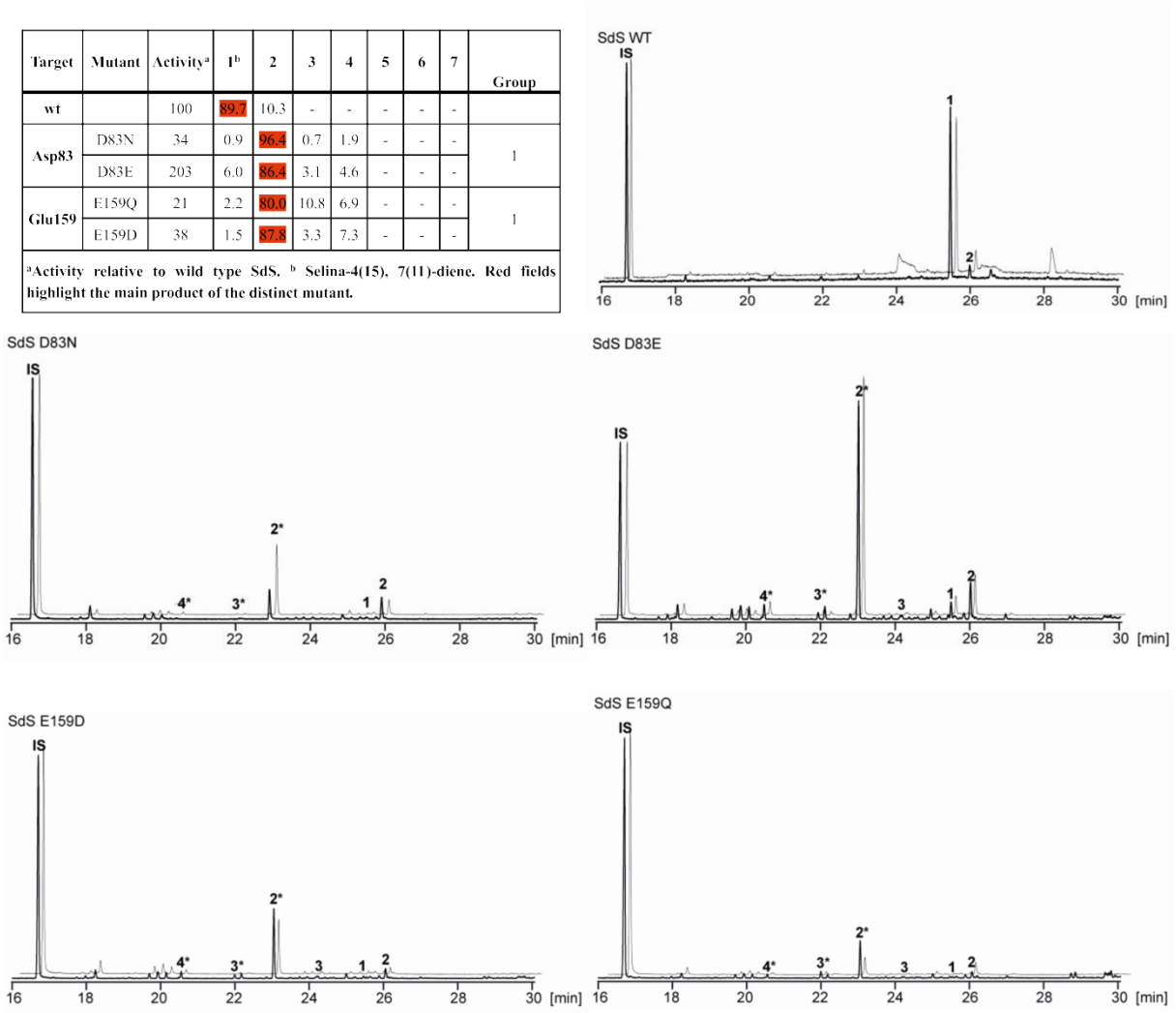


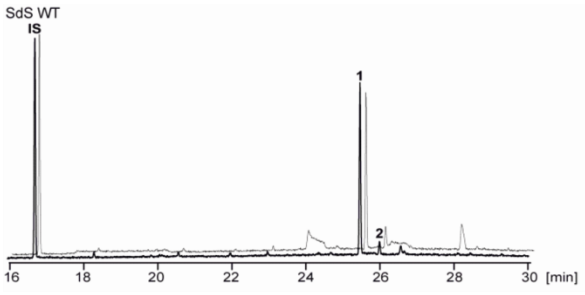
Figure 27. GC-MS product spectra analysis of the SdS mutants D83N, D83E, E159Q and E159D. Red fields indicate the main product of the distinct mutant. Bold numbers refer to compounds shown in **Figure 25**. Adapted from Baer et al[1].

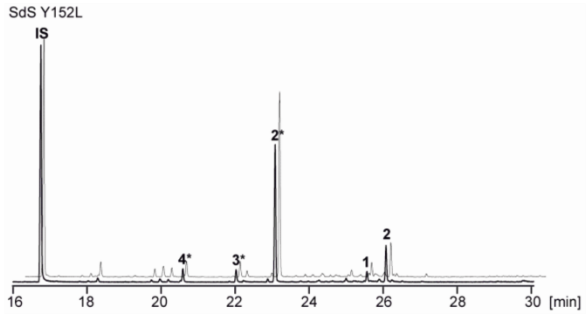
Mutant D83N exhibits a residual enzymatic activity of 34% (compared to the wild type) and forms almost exclusively **2**. Interestingly, D83E shows a 2-fold higher substrate turn over, producing **1** (minor product, 6%) and **2** (major product, 86%). D83 is part of the conserved **DD**xx**D** motif. Therefore, it is noteworthy that both mutants still show enzymatic activity and in case of D83E, this activity is even doubled. Since D83E represents a catalytically active class I terpene cyclase, it is most likely that this kind of variation of the **DD**xx**D** motif should also occur naturally. E159 is coordinated to the hydrate shell of Mg-3 (2.6 Å and 2.9 Å) and therefore contributes to the PP_i-$(Mg^{2+})_3$ coordination. Since both mutants (E159Q, E159D) display a low activity (21% and 38%, respectively) and product **2** is their main product (80% and 87%), E159 can be considered as being important for the enzyme's structure/function relationship.

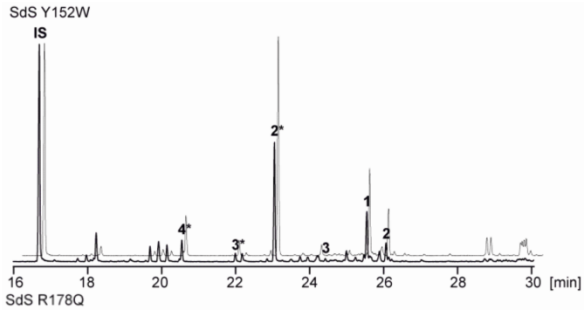
Group 2 covers amino acids which are essential for the Induced- Fit Mechanism and substrate activation (described in detail in the *Discussion* section). Hereby, the structural importance of Arg178 is highlighted which is invariant. The mutants investigated were: Y152W, Y152F, Y152L, R178K, R178Q, D181N, D181S, G182A, G182V, G182P, A183F and A183V (**Figure 28**).

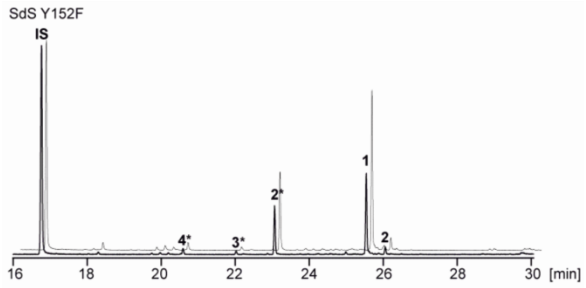
Target	Mutant	Activity[a]	1[b]	2	3	4	5	6	7	Group
wt		100	89.7	10.3	-	-	-	-	-	
Tyr152	Y152W	254	20.1	63.5	6.8	9.7	-	-	-	2
	Y152F	107	59.4	34.8	2.1	3.7	-	-	-	
	Y152L	111	5.5	81.8	6.4	6.4	-	-	-	
Arg178	R178K	8	-	100	-	-	-	-	-	2
	R178Q	0	-	-	-	-	-	-	-	
Asp181	D181N	98	4.4	88.9	3.5	3.2	-	-	-	2
	D181S	145	70.6	24.1	1.2	4.2	-	-	-	
Gly182	G182A	97	24.3	70.2	4.8	0.7	-	-	-	2
	G182V	0	-	-	-	-	-	-	-	
	G182P	0	-	-	-	-	-	-	-	
Ala183	A183F	0	-	-	-	-	-	-	-	2
	A183V	0	-	-	-	-	-	-	-	

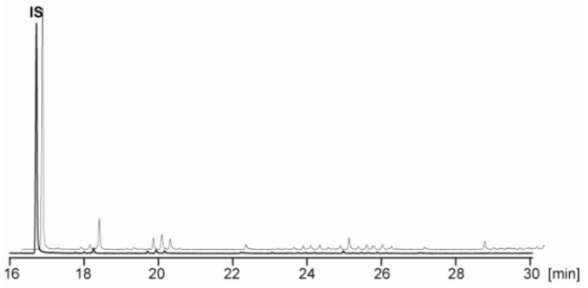
[a]Activity relative to wild type SdS. [b] Selina-4(15), 7(11)-diene. Red fields highlight the main product of the distinct mutant.

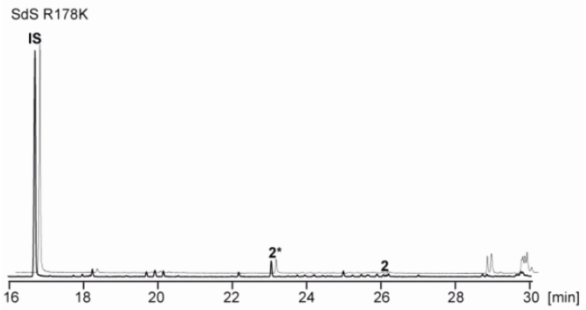












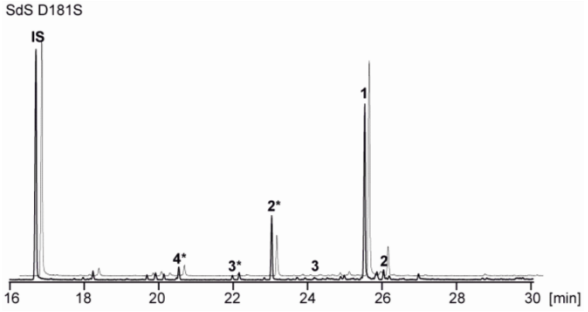


Figure continues on the next page

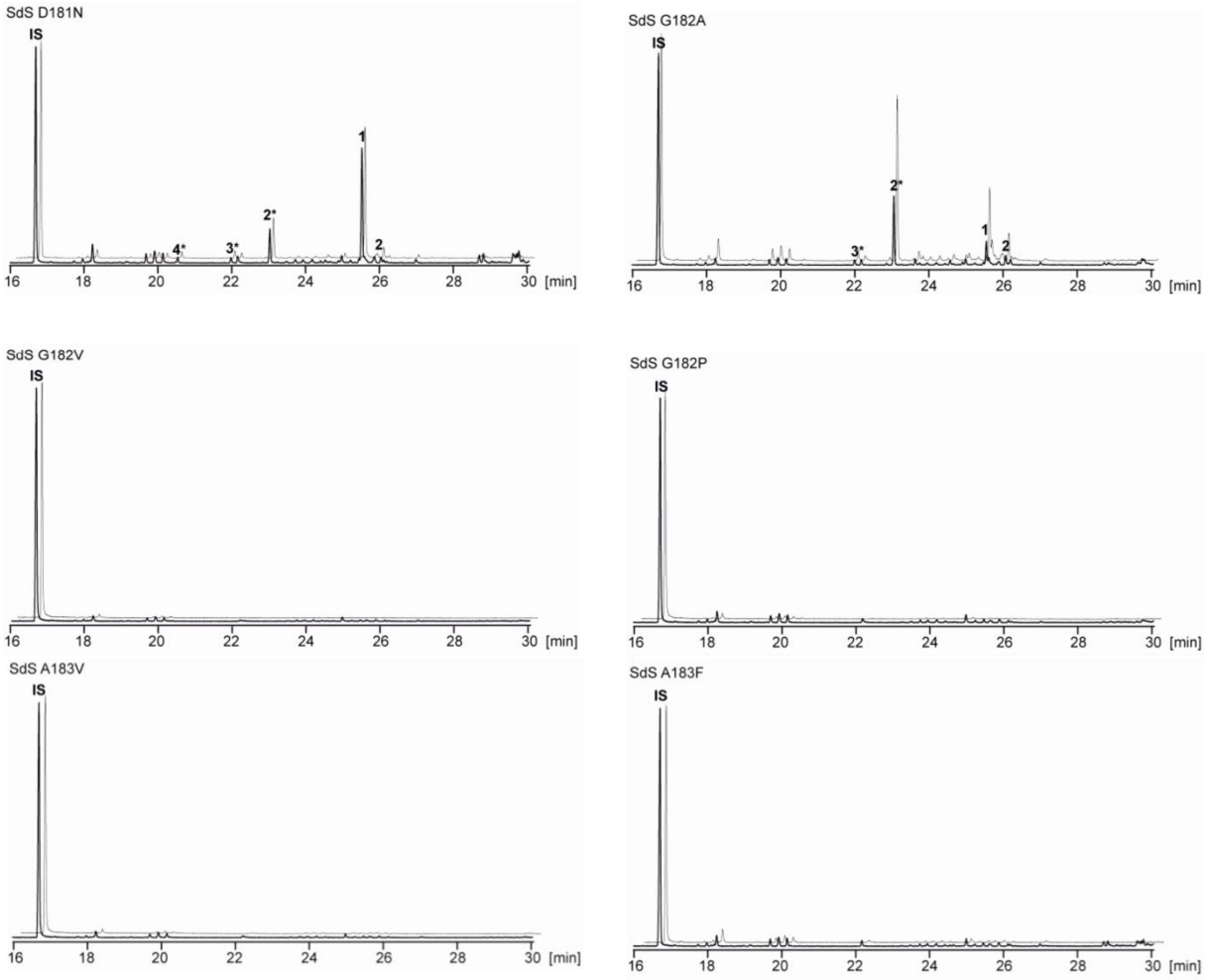


Figure 28. GC-MS product spectra analysis of the SdS mutants Y152W, Y152F, Y152L, R178K, R178Q, D181N, D181S, G182A, G182V, G182P, A183F and A183V. Red fields indicate the main product of the distinct mutant. Bold numbers refer to compounds shown in **Figure 25**. Adapted from Baer et al[1].

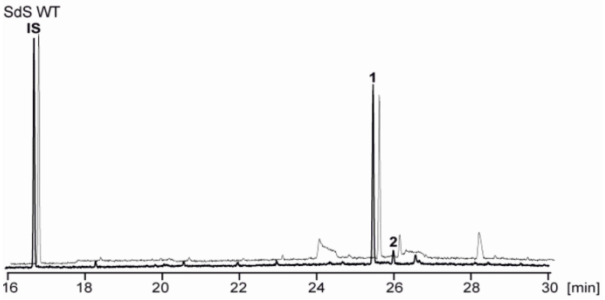
Y152 is adjacent to the substrate's C3 position (3.8 Å). Y152W displays a 2.5-fold increase in activity compared to wild type SdS, producing **1** (20%) and **2** (63%). Mutant Y152F does not alter the enzyme's catalysis speed but shifts the product spectra from **1** (60%) to **2** (35%). The last mutant, Y152L, shows again no decrease in activity, but a strong shift towards **2** (82%). The fact that a non-aromatic mutant does not result in a significant loss of activity rules out the importance of Y152 in carbocation stabilization. Both Arg178 mutants exhibit a complete loss in activity, therefore this amino acid can be considered as being essential for the enzyme's function. Its characteristic guanidine moiety which links different parts of the active site cannot be substituted by any other kind of amino acid. Like Arg178, D181 is part of the

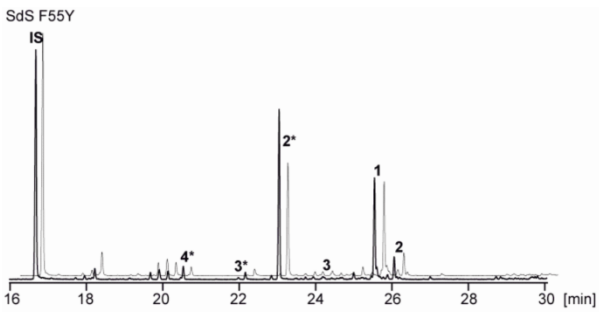
helix-break motif and is involved in the molecular rearrangement taking place upon substrate binding. Mutants D181N and D181S display wild type activity (the latter even exhibits a 1.45-fold increase). Whereas D181S shows no significant alteration of the product spectra, D181N considerably raises the fraction of **2** (89%). As the helix-break sequence alignment indicates a conversion of G182 into A182 does not have an impact on the enzyme's turnover rate. Still, the ratio of **1**:**2** is shifted again in favor of **2** (70%). The four mutants G182V, G182P, A183F and A183V display a complete loss of enzymatic activity. No substrate activation and carbocation formation takes places in these mutants.

The last group of amino acids investigated was group 3, including those amino acids contouring the active site and guiding the carbocation chemistry *via* aromatic cation stabilization. The mutants investigated were: F55W, F55Y, F55L, F79Y, F79W, F79L, W304Y, W304F, W304L, Y311W, Y311F and Y311L (**Figure 29**).

Target	Mutant	Activity [a]	1 [b]	2	3	4	5	6	7	Group
wt		100	89.7	10.3	-	-	-	-	-	
Phe55	F55W	13	5.2	89.0	1.4	4.4	-	-	-	3
	F55Y	131	36.6	58.1	1.2	4.1	-	-	-	
	F55L	138	17.4	52.0	14.2	1.5	1.1	11.7	2.2	
Phe79	F79Y	103	33.0	45.1	3.3	18.6	-	-	-	3
	F79W	29	5.4	80.5	4.3	9.7	-	-	-	
	F79L	153	42.1	3.3	5.3	4.0	9.9	32.0	3.3	
Trp304	W304Y	166	49.8	38.2	4.8	7.2	-	-	-	3
	W304F	75	57.3	36.1	2.8	3.9	-	-	-	
	W304L	130	3.5	81.9	8.7	5.8	-	-	-	
Tyr311	Y311W	17	33.3	58.9	6.6	1.1	-	-	-	3
	Y311F	137	63.9	27.9	3.4	4.8	-	-	-	
	Y311L	1	-	100	-	-	-	-	-	

[a]Activity relative to wild type SdS. [b] Selina-4(15), 7(11)-diene. Red fields highlight the main product of the distinct mutant.





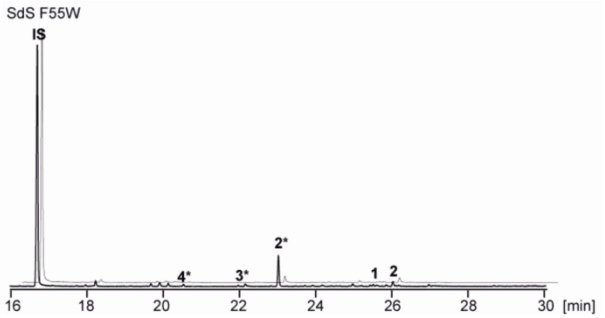


Figure continues on the next page

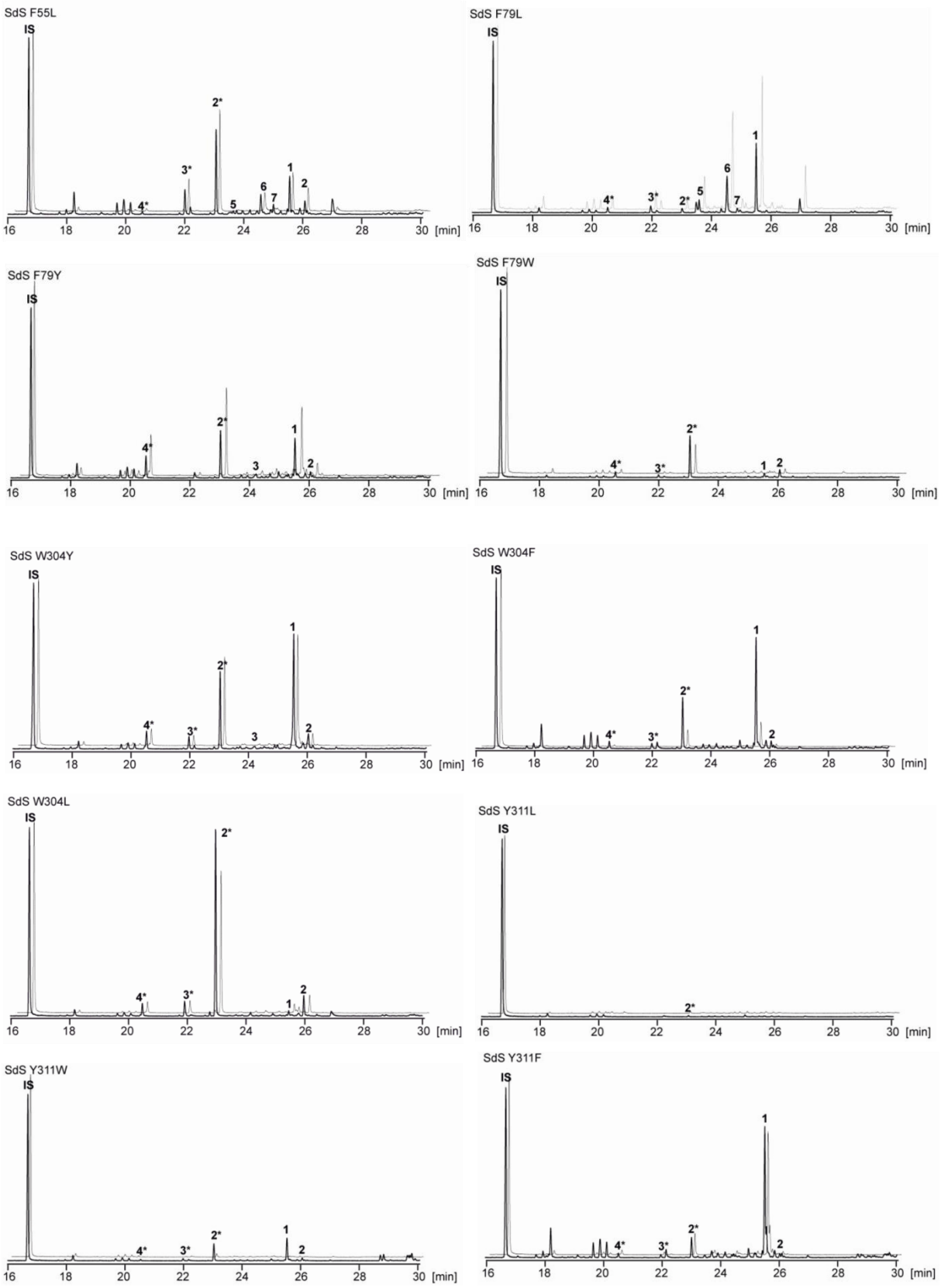


Figure 29. GC-MS product spectra analysis of the SdS mutants F55W, F55Y, F55L, F79Y, F79W, F79L, W304Y, W304F, W304L, Y311W, Y311F and Y311L. Red fields indicate the

main product of the distinct mutant. Bold numbers refer to compounds shown in **Figure 25**. Adapted from Baer et al[1].

The first mutant described is F55, which is located within the catalytic chamber. Mutants F55Y and F55L exhibit a slightly higher substrate turnover compared to wild type enzyme and a pronounced increase in production of **2** (58% and 52%, respectively). The non-aromatic leucine mutant rules out the participation of F55 in carbocation stabilization. It is noteworthy that this mutant produces compounds **1- 7**. Mutant F55W shows a significant loss in activity (13%), which is presumably caused by deteriorating substrate binding. F79 is another active site aromatic residue. Again, the tryptophan mutant displays a significant lower substrate turn over (29%) and a shift towards **2** (80%). F79Y remains kinetically unchanged; products **1** and **2** are produced in almost equal yields (33% and 45%, respectively). F79L results in a 1.5-fold increase of activity and its main product remains to be **1** (42%). Similar to F55L, the F79L mutants produces the whole range of products (**1- 7**). W304 is part of the active site as well and three different mutants have been designed and analyzed. W304Y displays a 1.6 fold increase in activity and it produces to almost equal parts **1** and **2** (50% and 38%). Mutant W304F exhibits a slight decrease in activity (75%), it main product remains **1** (57%). W304L exhibits an increased substrate turnover (1.3-fold). Interestingly, its main product is **2** (82%). The almost complete loss of **1** could be explained with the inability of the leucine mutant to interact with **2** *via* π-stacking. Therefore, **2** is less stabilized and its further conversion into **1** might be prohibited. Despite this, the W304L mutant produces terpenes as well. Therefore its participation in carbocation stabilization is unlikely. The last mutant investigated is Y311, which takes part in PPi-$(Mg^{2+})_3$ coordination. The Y311W mutant is less active compared to the wild type enzyme (17%), which is probably caused by steric clashes with the substrate. The Y311L mutant is inactive. Therefore, it is most likely that this amino acid is crucial for carbocation stabilization *via* an aromatic rest. This is confirmed by mutant Y311F, which exhibits almost unaltered enzymatic parameters (1.3-fold activity, main product **1** (64%)). Therefore it's not the H-bond between the tyrosine's hydroxyl group and the diphosphate which is of importance, but the aromatic rest.

In summary, 28 mutants of SdS were designed, purified and *in vitro* analysed regarding their corresponding product spectra (GC-MS). Hereby, we were able to identify amino acids important for substrate binding (group 1), substrate activation (group 2), product specificity and regulation of the carbocation chemistry (group 3). The most striking mutants proofed to be the ones targeting R178 (inactive), G182 (inactive), A183 (inactive), F55 and F79

(increase of product range), and Y311 (inactive). **Table 21** gives an overview of all mutants investigated. Hereby, red fields highlight the mutant's main product. Green fields indicate the mutant which produces the largest quantities of one of the seven different possible products (1- 7). Of all mutants investigated, Y152W exhibits the highest enzymatic activity.

Table 21. SdS mutants product spectra GC-MS analysis

Target	Mutant	Activity[a]	1[b]	2	3	4	5	6	7	Group
wt		100	89.7	10.3	-	-	-	-	-	
Asp83	D83N	34	0.9	96.4	0.7	1.9	-	-	-	**1**
	D83E	203	6.0	86.4	3.1	4.6	-	-	-	
Glu159	E159Q	21	2.2	80.0	10.8	6.9	-	-	-	**1**
	E159D	38	1.5	87.8	3.3	7.3	-	-	-	
Tyr152	Y152W	254	20.1	63.5	6.8	9.7	-	-	-	**2**
	Y152F	107	59.4	34.8	2.1	3.7	-	-	-	
	Y152L	111	5.5	81.8	6.4	6.4	-	-	-	
Arg178	R178K	8	-	100	-	-	-	-	-	**2**
	R178Q	0	-	-	-	-	-	-	-	
Asp181	D181N	98	4.4	88.9	3.5	3.2	-	-	-	**2**
	D181S	145	70.6	24.1	1.2	4.2	-	-	-	
Gly182	G182A	97	24.3	70.2	4.8	0.7	-	-	-	**2**
	G182V	0	-	-	-	-	-	-	-	
	G182P	0	-	-	-	-	-	-	-	
Ala183	A183F	0	-	-	-	-	-	-	-	**2**
	A183V	0	-	-	-	-	-	-	-	
Phe55	F55W	13	5.2	89.0	1.4	4.4	-	-	-	**3**
	F55Y	131	36.6	58.1	1.2	4.1	-	-	-	
	F55L	138	17.4	52.0	14.2	1.5	1.1	11.7	2.2	
Phe79	F79Y	103	33.0	45.1	3.3	18.6	-	-	-	**3**
	F79W	29	5.4	80.5	4.3	9.7	-	-	-	
	F79L	153	42.1	3.3	5.3	4.0	9.9	32.0	3.3	
Trp304	W304Y	166	49.8	38.2	4.8	7.2	-	-	-	**3**
	W304F	75	57.3	36.1	2.8	3.9	-	-	-	
	W304L	130	3.5	81.9	8.7	5.8	-	-	-	
Tyr311	Y311W	17	33.3	58.9	6.6	1.1	-	-	-	**3**
	Y311F	137	63.9	27.9	3.4	4.8	-	-	-	
	Y311L	1	-	100	-	-	-	-	-	

[a]Activity relative to wild type SdS. [b] Selina-4(15), 7(11)-diene. The chemical structure and biosynthesis of compounds **1- 7** are shown in Figure 25. Red fields highlight the main product of the distinct mutant; greens fields highlight the mutant which produces the largest quantities of one of the seven different products.

4.1.8. Discussion SdS

In summary, we determined high resolution crystal structures of Selina-4(15), -7(11)-diene Synthase (SdS), in complex with $(Mg^{2+})_3$-PP_i and $(Mg^{2+})_3$-DHFPP, in its closed conformation. Moreover, we obtained the open conformation structure of SdS:apo. Phases were determined experimentally, using selenomethionine substitution and SAD-methods. The biochemical characterization of SdS comprised the investigation of ligand binding properties (by thermofluor based thermal shift assay) and the enzymatic activity (product spectra analysis *via* GC-MS). Ultimately, mechanistic models have been probed by designing 28 different point mutants. In the following section, the different results will be discussed and linked to each other.

Experimental phases had to be obtained since class I terpene cyclases, despite having a general overall conserved architecture, display a low (< 25 %) conservation of the primary sequence. Therefore, the C_α carbon atoms of the different α-helices from distinct class I terpene cyclases are not well aligned. For this reason, phase information cannot be obtained using molecular replacement. The selenomethionine substituted SdS crystals grew fast and in large numbers, only displaying an overall small size. This problem could be resolved by the addition of 200 mM NaCl to the initial crystallization conditions. The observed positive effect can most likely be explained with the weakening of unspecific electrostatic interactions. By this, the nucleation tendency is slower and the total number of crystals is reduced.

It is well known for class I terpene cyclases that the highly reactive carbocation chemistry is taking place in a "dry" environment (closed active site). After abstraction of the diphosphate, the carbocation reaction cascade is guided by the chemical landscape of the active site, forming the specific product(s). Hereby, aromatic amino acids like Phe, Tyr and Trp[39][40] play an important role. Still, it is not understood how substrate activation takes place and by which means this is conducted in aqueous solution. If we compare the SdS:**apo** structure (open conformation) with the SdS:**DHFPP** complex structure (closed conformation), we can observe the same structural rearrangement of helix-break G1 as described for the SdS:**apo**/SdS:**PP_i** structures (**Figure 20**). Interestingly, the hydrocarbon backbone of the DHFPP ligand now allows a biologically reasonable explanation for this molecular rearrangement, as shown in **Figure 30**.

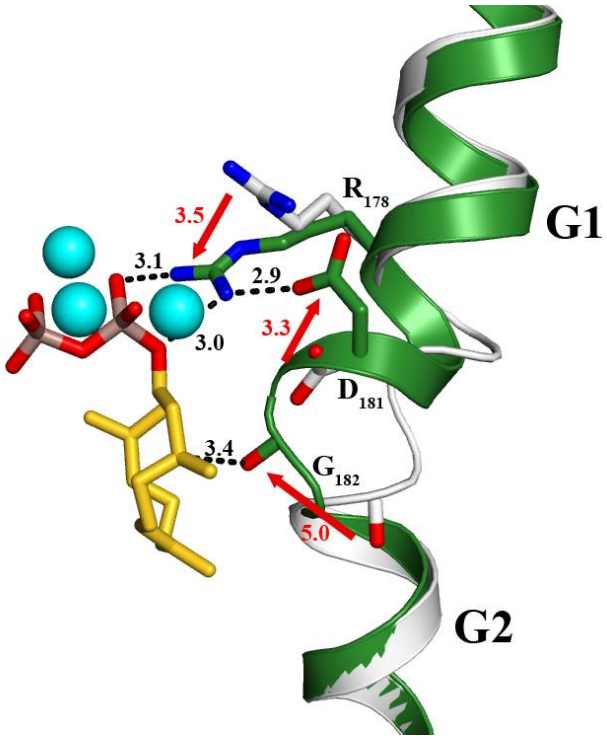


Figure 30. Structural superposition of helix G1/2 and the corresponding helix-break motif of SdS:DHFPP and SdS:apo. The apo- form (open conformation) is coloured in grey, the closed SdS:DHFPP structure in green. Upon substrate coordination and active site closure, Arg178 shifts by 3.5 Å, Asp181 by 3.3 Å and Gly182 by 5.0 Å. In the course of these rearrangements, the overall helix-break motif rearranges into an alternative conformation, bringing the carbonyl oxygen of Gly182 in close contact (3.4 Å) to the substrate's C3 position. Adapted from Baer et al[1].

Upon substrate binding, Arg178 (3.5 Å), Asp181 (3.3 Å), Gly182 (5.0 Å) and the overall G1/2 helix-break motive are rearranged relative to the apo structure. This distinct movement brings the carbonyl oxygen of Gly182 in close contact (3.4 Å) to the C3 atom of the DHFPP substrate. Hereby, the free electrons (negative polarity) of the carbonyl moiety are presumably interacting with the π* molecular orbital of the substrate's first double bond (C2/3). This way, the double bond is weakened, abstraction of the diphosphate is triggered and the primary carbocation is formed (**Figure 31**). It is noteworthy that the effector's carbonyl group shifts by 5.0 Å for the SdS:**DHFPP** closed conformation and 5.5 Å in case of the SdS:**PP_i** complex structure (upon substrate binding). This marginal variation of the helix-break motif, which can just be observed with high resolution crystal structures, indicates a structural tension between the bound ligand and SdS. This implicates a pushing of the carbonyl group of Gly182 towards the substrate's C3 position, therewith contributing to the diphosphate abstraction. This

sophisticated molecular mechanism perfectly matches the requirements of the carbocation chemistry in aqueous solution ([H_2O] = 55 M). In a first step, the open conformation is accessible for substrate binding and $(Mg^{2+})_3$ coordination, which is accompanied with a closure of the active site, thus forming a dry reaction chamber. Simultaneously, the molecular mechanism described above, comprising the pyrophosphate sensor Arg178, the linker Asp181 and the effector residue Gly182, initiates diphosphate abstraction and carbocation formation. These three amino acids were termed **effector triad**.

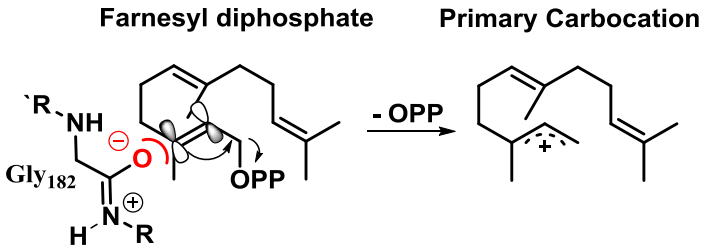


Figure 31. Interaction between the free electron pairs of Gly182 and the π*- molecular orbital of the substrate's first double bond. The anti-binding molecular orbitals are indicated. Pushing electrons into the these molecular orbitals weakens the π- bond[82] and, as a consequence, the diphosphate leaving group is abstracted.

The observed molecular mechanism and the identification of a novel effector triad (**pyrophosphate sensor** Arg178, **linker** Asp181 and **effector** Gly182) explains substrate activation and carbocation formation for class I terpene cyclases in aqueous solution. The underlying enzymatic reaction mechanism is a classic **induced-fit mechanism**. To prove the general applicability of this Induced-fit mechanism and the mode of substrate activation in class I terpene cyclases, the closed SdS:**DHFPP** complex structure is compared with various other structures of the same enzyme class (mono-, sesqui- and diterpene cyclases from bacteria, fungi and plants) (**Figure 32**).

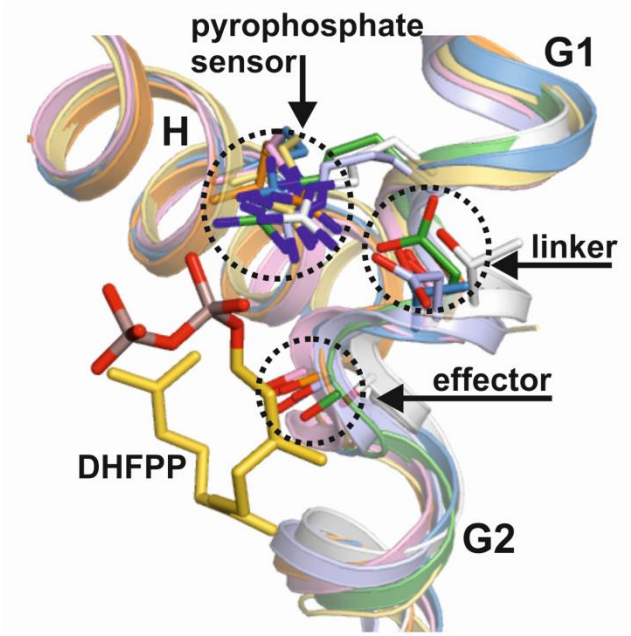


Figure 32. A structural alignment of the H helix and the G1/2- helix (and its corresponding helix-break) of various class I terpene cyclases is shown[1]. PDB codes: SdS (4OKZ, green)[1], epi-isozizaene synthase (3KB9, grey)[83], aristolochene synthase (4KUX, light blue)[41], bornyl diphosphate synthase (1N20, pink)[40], limonene synthase (2ONG, orange)[39], 5-epi aristolochene synthase (5EAT, blue)[10] and taxadiene synthase (3P5R, yellow)[52]. Adapted from Baer et al[1].

There is an apparent strict conservation of the structural arrangement of the effector triad in all class I terpene cyclases, occurring in bacteria, fungi and plants. Interestingly, only Arg178 is strictly conserved in the primary sequence of bacteria and fungi (helix G1). In plants this residue is located on helix H. This indicates that the class I terpene cyclases are a rather old enzyme class since the overall primary sequence conservation, including key residues like Arg178, is low. Nonetheless, the tertiary structure and the corresponding biological activity are highly conserved among class I terpene cyclases.

To proof these mechanistic models, we investigated 28 different mutants of SdS. As described in the *Results* section, we hereby distinguish 3 different groups, according to their role in the enzymatic mechanism.

Group 1 comprises amino acids which are coordinating the $(Mg^{2+})_3$-PP_i (D83N, D83E, E159Q and E159D). Interestingly, all mutants show an almost complete loss of product **1** and shift towards the production of **2**. This highlights the finely tuned coordination sphere of the $(Mg^{2+})_3$-PP_i cluster and its role in reprotonation of **2**, yielding **1**. Moreover, the D83E mutant

displays a 2- fold increase in enzymatic activity. This is surprising since D83 is part of the **DD**xx**D** motif, which is strictly conserved. Maybe, it could also be possible for other class I terpene cyclases to introduce this mutation to increase the enzymes' reaction rate, which would be a significant contribution to improving terpenoid biosynthesis form a biotechnological point of view.

Group 2 covers amino acids involved in the Induced-fit mechanism and substrate activation (Y152W, Y152F, Y152L, R178K, R178Q, D181N, D181S, G182A, G182V, G182P, A183F and A183V). The mutations of Y152W,F, L render the importance of carbocation stabilization at C3 for conversion of **2** to **1**. Accordingly, the Y152L variant shows the lowest activity in formation of 1. Interestingly, the Y152W mutant displays a 2.5 fold increase in activity compared to the wild type enzyme (Y152W generates **2** as its main product, whereas Y152F produces **1**), rendering this mutant even more effective in catalysing the first cyclization reaction. On the one hand, this could be explained with an improved stabilization of the primary carbocation after diphosphate abstraction and rearrangement of the substrate within the active site. On the other hand, a form of π- stacking between **2** (three double bonds) and Y152W could be imagined, thus favouring **2** over **1**. The R178K,Q mutants clearly highlight the importance of the guanidine moiety of the arginine residue and explain its strict structural conservation. The activity of both mutants drops significantly, underlining the importance of the distinct H- bond network between Arg178, the diphosphate moiety and the linker residue Asp181. Presumably, a disturbance of this well balanced molecular switch mechanism taking place upon substrate binding hinders substrate activation. This is in line with the observation that not even linear terpenes are formed by these mutants. For a long time it was assumed that substrate activation simply occurs by coordinating the ligand to different amino acids at the entrance of the active site, thereby initiating diphosphate abstraction[84]. There is no doubt that the three Mg^{2+} atoms are essential for catalysing diphosphate abstraction. Nonetheless, mutant R178K shows that the coordination of diphosphate is not sufficient for substrate activation. The described induced- fit mechanism ultimately controls activation of the substrate. Both mutants of linker Asp181N,S display activity similar to the wild type enzyme. D181S shows a 1.5- fold increase in activity suggesting an improved opening-closing of the active site. This might be explained by the weaker hydrogen bonding between Ser/Arg compared to Asp/Arg. Mutant D181N shows no change in activity, which is to be expected due to the two amino acids being structurally almost identical. The observed shift in production of **1** to **2** can again be explained with the last protonation step yielding **1**. Exchanging the acidic Asp with the

neutral Asn removes protons from the diphosphate, therefore probably preventing the conversion of **2** into **1**. The mutations of Gly182A,V,P point out the importance of the helix-break shift between open and closed conformation. Whereas Ala is similar to Gly regarding its size and often present as a substitute at this position in class I terpene cyclases (cf. Figure 12), Val and Pro disturb the well-balanced helix-break shift. Since the functionality of the effector residue relies on the carbonyl group of the backbone, which is present in all amino acids, the main catalytic property can be conducted by any amino acid. Still, there are amino acids favouring this kind of molecular mechanism over others. Therefore, many different amino acids can be found at this position in class I terpene cyclases. The individual architecture of helix G1 and the corresponding helix-break must be structured in a particular way, individually fitting the properties of the different amino acids. Therefore, an exchange of the Gly residue with a larger one like Val significantly disrupts the required conformational flexibility of the helix- break. The same observation and interpretation can be applied towards the A183F,V mutants. In both cases, the enzymatic activity is lost. It is noteworthy again to mention that not even linear terpenes are present as products. We therefore conclude that no substrate activation takes place, even though the mutations of Ala183 should not cause a sterically hindrance to substrate binding. Interestingly, a shift of Tyr152 caused by Ala183, can be observed upon active site closure. Larger amino acids like Phe and Val clash with Tyr152 at this point thus preventing substrate activation (**Figure 33**).

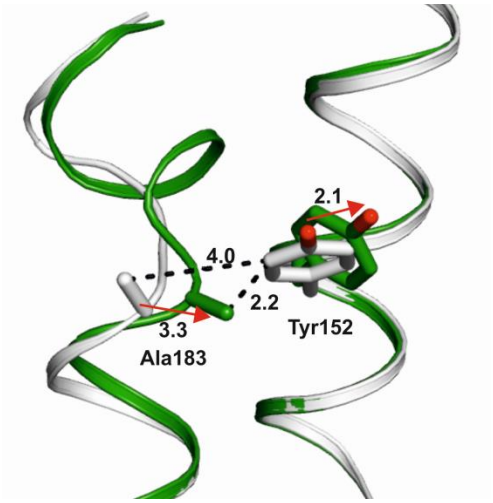


Figure 33. A close up view of the structural superposition of the helices G1 and C from SdS in its apo (grey) and closed (green) state. In the apo conformation, Ala183 is 4.0 Å distant from Tyr152. Compared to the closed conformation, Ala183 is shifting by 3.3 Å and Tyr152 by 2.1 Å. Therefore, larger amino acids sterically clash with Tyr152 upon active site closure,

hereby locking the enzyme in its open form. Accordingly, no substrate activation takes place and no (linear-) terpenes are formed. Adapted from Baer et al[1].

The last group investigated comprises amino acids forming the active site (F55W, F55Y, F55L, F79Y, F79W, F79L, W304Y, W304F, W304L, Y311W, Y311F and Y311). Upon comparison of the mutants F55L, F79L, W304L and Y311L with each other, it is striking that only the Y311L variant has completely lost its activity. Therefore, it is likely that only the stabilization of the secondary carbocation, occurring after the first ring closure, is stabilized by aromatic residues. All subsequent rearrangement reactions are presumably controlled by the pre-folded ligand and site- specific protonations. The wild type again displays by far the strongest formation of **1** compared to the mutants, showing once more the strict demands of the architecture of the active site residues. These prerequisites eventually enable the enzyme to specifically produce a single, distinct product. F55L and F79L both produce the whole range of products (**1**- **7**). A combination of these mutants could convert the specialist enzyme SdS into a more generalist like class I terpene cyclase which might serve as a starting point to generate novel terpenes by SdS.

The performed mutagenic study proved the proposed mechanistic models. In addition the achieved data clarified that only a limited number of aromatic residues within the active site control the carbocation chemistry. Moreover, the energetic and chemical landscape formed by the active site presumably functions in the sub-Angstrom range. An alteration based on a rational approach mainly leads to a loss of activity in the case of SdS, which clearly is a specialist. Still, there are a number of published examples of class I terpene cyclases where a shift in the product spectra was introduced by random mutagenesis[85]. The enzymes investigated in these works represented class I terpene cylcases with a broader product spectrum. In conclusion of this, we propose that product spectra modification of class I terpene cylcases are more likely to succeed when more generalistic enzymes like taxadiene synthase are applied. In summary, the combination of random mutagenesis and structure based rational design of generalistic class I terpene cyclase is the most promising approach to yield novel and interesting products.

4.2. Hedycaryol Synthase

Hedycaryol synthase (HcS) is a $(Mg^{2+})_3$ -dependent class I sesquiterpene cyclase from *Kitasatospora setae* (KM-6054, Uniprot code: E4MYY0, EC 4.2.3., lyase) which selectively converts farnesyl diphosphate (FPP) into (2*Z*,6*E*)-hedycaryol[2][32]. Despite the identification of an induced-fit mechanism in selinadiene synthase (SdS), which explains substrate activation and carbocation regulation in class I terpene cyclases in aqueous solution, it is still not understood by which means the active site contributes to the formation of a distinct product. So far, complex crystal structures of class I terpene cyclases were always liganded with substrate analogues (e.g. DHFPP and 2F-FPP) or with unnatural ligands. Therefore, an investigation of a pre-folded reaction intermediate, which would highlight the active site's interaction with the substrate, was not possible yet. In my Ph.D.-thesis, I attempted to target this issue by crystallizing HcS and applying techniques which trap reaction intermediates. I have chosen HcS among other sesquiterpene cyclases for its high protein expression level in *E.coli*. Its gene counts 1041 base pairs (68 % GC content) and the corresponding protein product comprises 338 amino acids (M_W = 37 kDa, ε = 61,420.00)[58]. The corresponding DNA- and amino acid sequence are shown in the appendix. An alignment of the primary sequences of HcS and SdS is shown in **Figure 34**. In summary, these two bacterial class I terpene cyclases feature a sequence identity of 18% and a sequence similarity of 29%, which represent rather low values. However, strictly conserved are the **DD**xx**D** (^{82}**DD**xx**D**^{86})- and the **ND**(L,I,V)x**S**xxx**E** (^{221}**ND**VF**S**VER**E**^{229}) motifs. Moreover, the effector triad comprising the pyrophosphate sensor (R175), the linker (S178) and the effector residue (V179) are well conserved in the HcS, too (red stars).

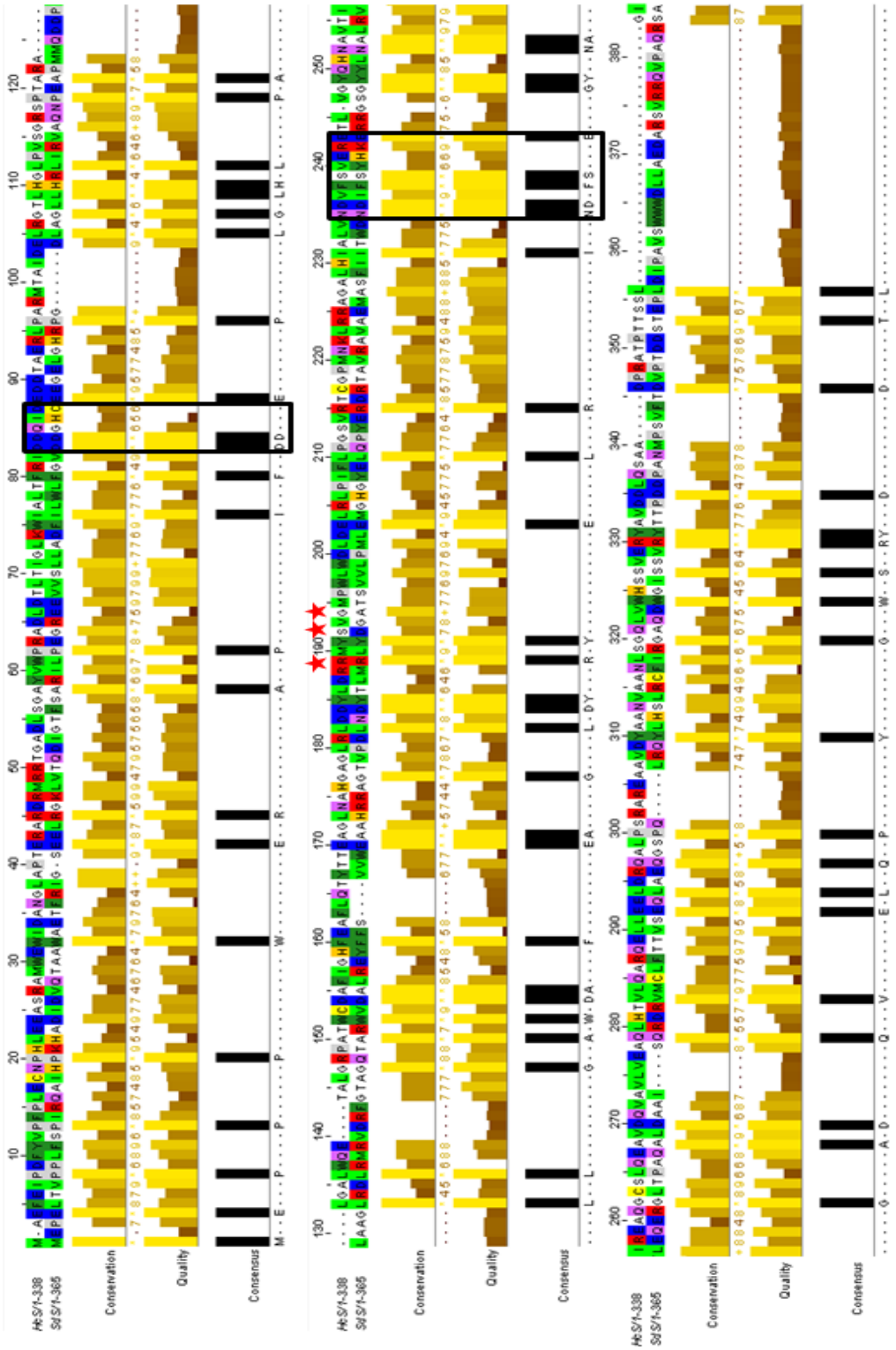


Figure 34. A primary sequence alignment of HcS and SdS (MSAProbs and Jalview). The alignment's conservation, quality and consensus are given as well. Black boxes indicate the **DD**xx**D-** and the **ND**(L,I,V)x**S**xxx**E** motifs. The red stars highlight the effector triad.

4.2.1. Cloning and Purification

The HcS gene was cloned into a high copy pET28b vector featuring a C-terminal LEHHHHHH sequence (His_6-tag). For capturing a reaction intermediate within the enzyme's active site, we decided to omit $MgCl_2$ during cell disruption to quench ongoing substrate turnover of endogenous farnesyl diphosphate. For purification of the apo enzyme, we added 10 mM $MgCl_2$ to the lysis buffer. After expression of the HcS-His_6 protein, a Ni^{2+} based affinity chromatography yielded pure HcS (according to SDS- PAGE[68]) with a major band at around 37 kDa. For protein elution a low salt buffer with pH 8.0 and supplemented with imidazole was used. Subsequently, the protein was loaded on an anion exchange chromatography column (BioPro Q30, YMC). A gradient between 50 mM- 400 mM NaCl over 25 ml was ran at 1 ml/min. HcS eluted at a NaCl concentration of around 300 mM. The third purification step comprised a size exclusion chromatography (Superdex 200, 10/300), displaying a sharp single peak at an elution volume of 15 ml, thus indicating a monomeric protein (**Figure 35**).

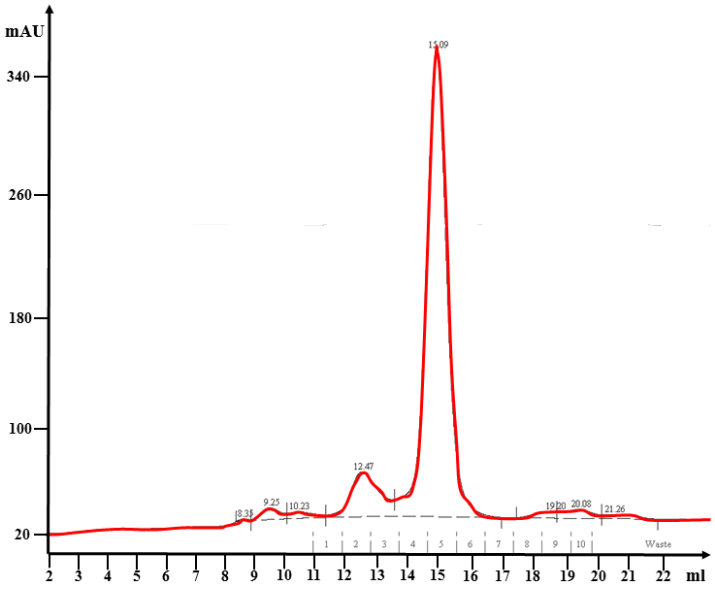


Figure 35. A Superdex 200 10/300 size exclusion chromatogram is displayed. The red graph shows the UV_{280} absorption. HcS elutes at 15 ml, thus displaying a monomeric protein in solution.

4.2.2. Circular Dichroism Thermal Shift Assay

In order to find suitable substrate analogues for co-crystallization with HcS we attempted to screen potential inhibitors utilizing thermal shift assays (Thermofluor based)[59]. Unfortunately, this approach failed due to the high fluorescence background generated by

HcS, 2-Fluoro-FPP and Sypro Orange (Sigma Aldrich). Therefore, we applied circular dichroism spectroscopy (CD) to follow the temperature dependent unfolding of the HcS α-helices. A higher melting temperature indicated a stronger binding and stabilization of the ligand. For data fitting a double Boltzmann fit was applied for HcS:apo and a normal Boltzmann fit for HcS:2F-FPP. HcS:apo displayed two points of inflection (40.7 °C and 49.1 °C) whereas HcS: 2F-FPP featured a single transition at 54.3 °C. Thus, 2F-FPP stabilized HcS by 5.2 °C, indicating a strong binding of the ligand (**Figure 36**).

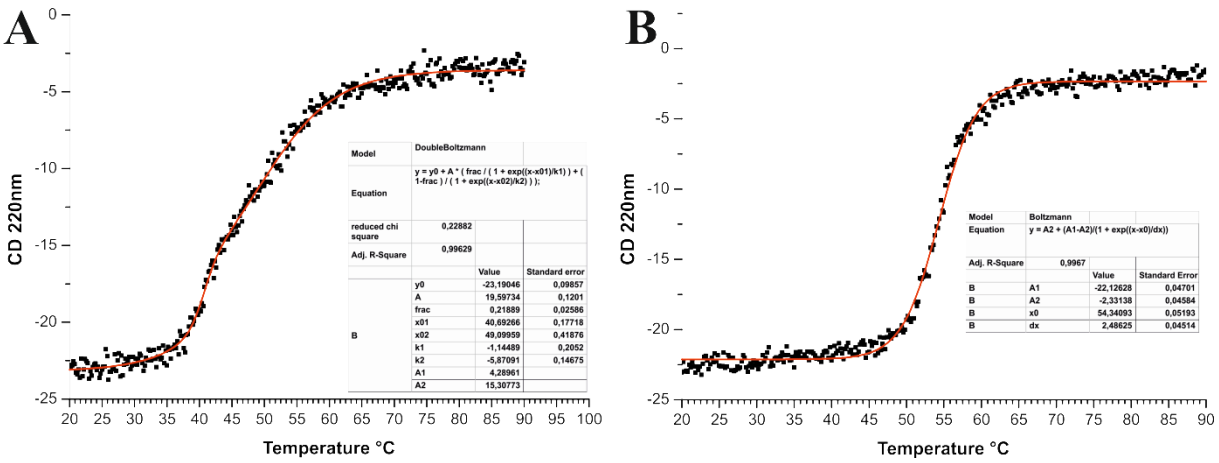


Figure 36. Thermal shift CD- spectra of apo- HcS (**A**) and HcS: 2F-FPP (**B**) are shown[2]. The apo form exhibits a two- step unfolding (T_{M1}= 40.7 °C and T_{M2}= 49,1 °C) which is analysed by applying a double Boltzmann sigmoid fit. Adapted from Baer et al[2].

The binding of the substrate analogue 2F-FPP towards HcS significantly stabilized the enzyme's overall structure (+ 5.2°C). Interestingly, the addition of PP_i in presence of Mg^{2+} again didn't show any effect on the protein's melting temperature, which is in line with the observations made for SdS. It is noteworthy to mention, that HcS:apo comprises two distinct inflection points. This is in contrast to HcS:2F-FPP, which features a single transition. An explanation might be the observations made for SdS, where PP_i-$(Mg^{2+})_3$ coordination linked large parts of the enzyme. Such a compact structure probably displays a concerted unfolding as we could observe for the HcS:2F-FPP complex (**Figure 36 B**).

4.2.3. Enzymatic Activity

For an enzymatic characterization of HcS purified protein was incubated with various oligoprenyl diphosphate substrates at different pH values (7.0 and 8.5): geranyl diphosphate (GPP), farnesyl diphosphate (FPP), (2*Z*,6*E*)-farnesyl diphosphate ((2*Z*,6*E*)-FPP), geranylgeranyl diphosphate (GGPP) and 2-fluoro-farnesyl diphosphate (2F-FPP) (**Figure 37**).

The incubation experiments were conducted in joint collaboration with Patrick Rabe from the group of Prof. Dr. Jerome Dickschat at the University of Bonn.

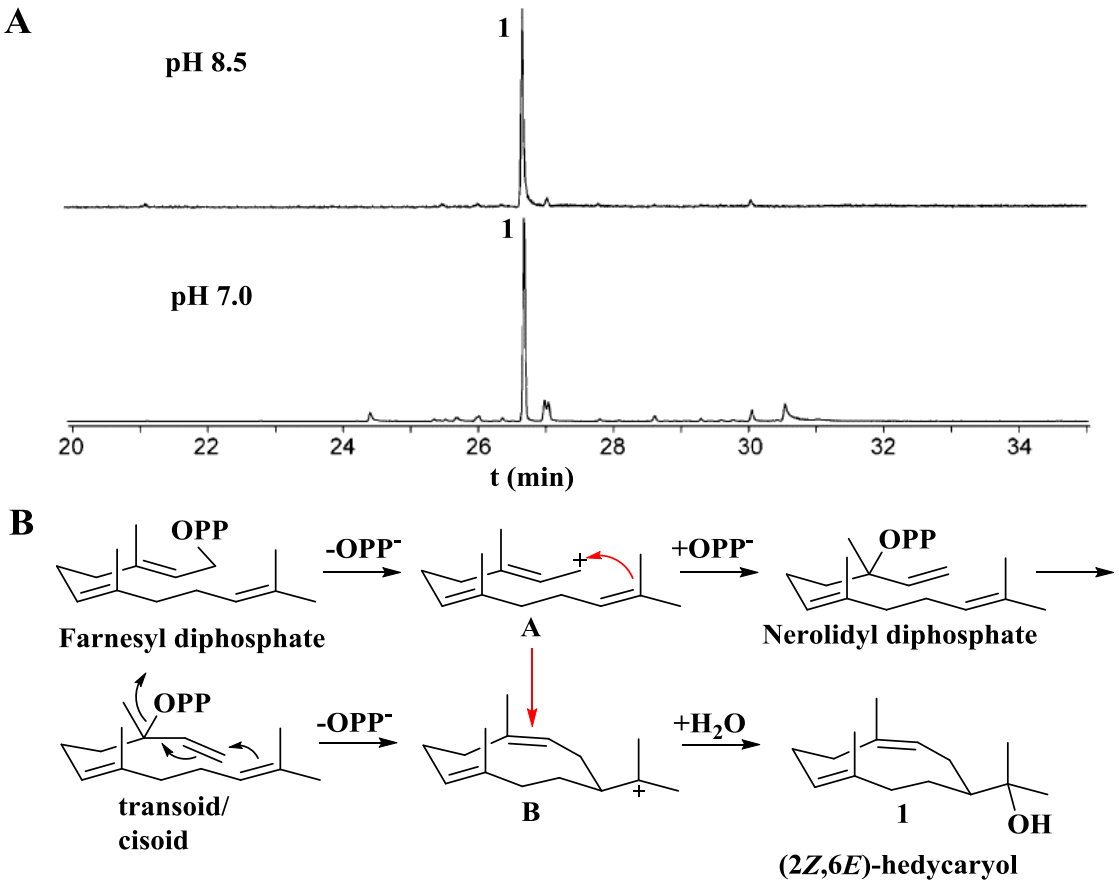


Figure 37. The enzymatic conversion of FPP by HcS at two different pH values (**A**) and the underlying reaction mechanism (**B**) are shown (GC-MS chromatogram). Both carbocation reaction intermediates A and B are displayed. The red arrow indicates the possibility of a direct conversion from the transoid A into the cisoid A reaction intermediate. Adapted from Baer et al[2].

The incubation experiments and the corresponding product spectra analysis *via* GC-MS classified HcS to be a highly specific class I terpene cyclase. The retention time of **1** was 26 min and 45 s. In a first step, substrate activation and primary carbocation formation takes place (induced-fit mechanism), yielding carbocation reaction intermediate **A**. In the literature it is proposed that in the succeeding reaction, the diphosphate re-attacks at the C3 position, forming the Markovnikov product nerolidyl diphosphate[44]. Hereby, the C2,3 double bond switches from *trans* to *cis*. Recently, based on a cyclization reaction conducted in organic solvent and an artificial folding chamber, it was suggested that this shift from *trans* to *cis* can also take place spontaneously in reaction intermediate **A** (red reaction arrow)[45]. Intermediate

B is yielded by an intramolecular attack of double bond C10,11 at the position C1. This carbocation is quenched by an addition of H_2O, yielding the final product **1**.

4.2.4. Crystal Structure Determination of HcS

Initial crystallization trials were conducted with wild type HcS-His_6 as described in the *Materials & Methods* section. In this first round, 96 well plate sitting drop screens (Quiagen) using a Phoenix robot (Art Robbins Instruments) were carried out. Based on initial hits, a fine screen was performed applying the hanging drop vapour diffusion method (20 mg/ml), to optimize the crystallization conditions (**Table 22**).

Table 22. Crystallization conditions for HcS

HcS:Hg crystals	4M Na-formiate, pH 7.0 + **$HgCl_2$**, 4µl:1µl (HcS:buffer)
HcS:apo crystals	100 mM Tris/HCl pH 8.0, 1 mM $MgCl_2$, 1.6 M ammonium sulfate, 1µl:1µl (HcS:buffer)
HcS:2 crystals	4M Na-formiate, pH 7.0, 4µl:1µl (HcS: buffer) ➔ **no $MgCl_2$** in the lysis buffer

The crystals appeared after one week at 20°C, displaying an overall large size. The HcS:**2** and HcS:apo crystals showed a distinct macroscopic appearance as **Figure 38** highlights.

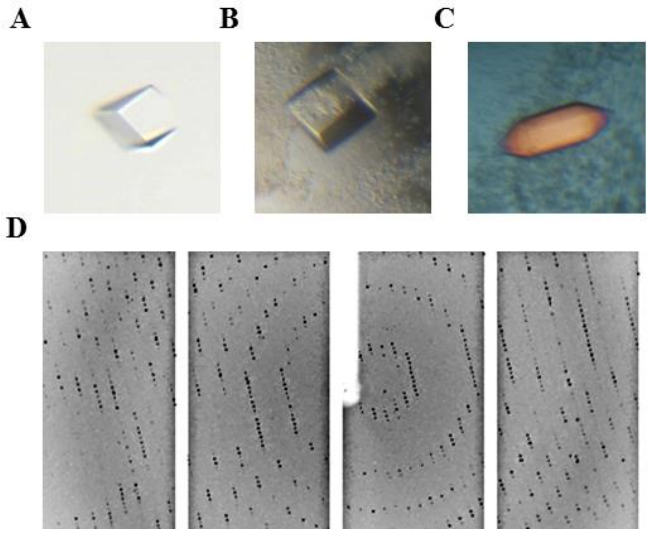


Figure 38. Crystals of HcS:**2** (A sitting drop, B hanging drop) and HcS:apo (C, polarization filter). A representative diffraction pattern is displayed (D).

Crystals were measured at the Swiss Light Source synchrotron (Villigen, Switzerland) as described in the *Materials & Methods* section. Native data sets were recorded at 1 Å wavelength. Since molecular replacement techniques failed, datasets of mercury soaked

HcS:**2**:**Hg**$^{2+}$ crystals were collected at the absorption wavelength of mercury (peak = 1.0061). The collected data sets of HcS:**2** and HcS:**2**:**Hg**$^{2+}$ were processed with the XDS program suite [70], yielding in both cases the $P3_121$ space group featuring the cell parameters a = b = 60 Å and c = 183 Å. Experimental phases were recorded of the anomalous measured HcS:**2**:**Hg**$^{2+}$ dataset applying single-wavelength anomalous dispersion methods (SAD)[69]. The solvent predictions (Matthews Coefficient) and the self-rotation function calculated with MOLREP[71] resulted in one subunit of HcS:**2** per asymmetric unit cell and a solvent content of 49%. SHELXD[72] was applied for identifying the positions of the heavy atom metals, thereby determining four Hg^{2+} binding sites. SHARP-SAD[73] phasing was performed and subsequent solvent flattening with SOLOMON[74] yielded proper phases at about 2.5 Å resolution, which were sufficient to build most secondary structure elements with a poly alanine model. The initial model was structurally superimposed with the coordinates of the pentalenene synthase (PDB code: 1PS1). Including the calculated PHI_{calc} values allowed to assign the correct sequence and to finish the entire HcS model by completing the missing secondary structure elements and the loops connecting these. Hereby, the interactive three-dimensional graphic program MAIN[75] was carried out in successive rounds. Water molecules were placed automatically running ARP/wARP[77]. Upon investigation of the active site, a distinct extra electron density could be observed. This electron density proofed to be nerolidol, a natural occurring side product of HcS. Once the ligand and all solvent molecules were added, the crystal structure was finalized running REFMAC5 with TLS parameters[76]. Model evaluation was done according a Ramachandran plot which was calculated with PROCHECK[78]. This revealed an excellent stereochemistry and 0% outliers (100% were in the favoured regions). In summary, the HcS:**2** dataset contains one subunit in the asymmetric unit which featured a nerolidol molecule within its active site. The HcS:apo crystal revealed an alternative space group namely C2. The corresponding unit cell parameters were a = 118 Å, b = 81 Å, c = 98 Å and β = 95°. Therefore, the model of HcS:**2** was used for as starting coordinates for the Patterson search calculations applying PHASER[86]. The positioned model was refined by REFMAC5, applying rigid body and TLS parameters. Again, the final structure showed excellent crystallographic values. There are two molecules in the asymmetric unit of the HcS-apo dataset, which active sites did not display any defined electron density. The data collection and refinement statistics for HcS:**2**:**Hg**$^{2+}$, HcS:**2** and HcS:apo are summarized in **Table 23**.

Table 23. Data collection and refinement statistics for HcS. Adapted from Baer et al[2].

	HcS:**2** (peak; Hg)	HcS:**2**	apo-HcS
Space group	$P3_121$	$P3_121$	C2
Cell constants (Å)	a=b=59.6, c=182.3	a=b=59.4, c=182.8	a=118.4, b=80.6.5, c=97.6, β=94.7
Anomalous scatterers	4 Hg	-	-
Molecules in asym. unit	1	1	2
Disordered regions		1-3 / 86-94 / 111-117 / 226-234 / 311-346	1-3 / 86-94 / 111-117 / 226-234 / 311-346
X-ray source	SLS, X06SA	SLS, X06SA	SLS, X06SA
Wavelength (Å)	1.0061	1.0	1.0
Resolution range (Å)[a]	30-2.5 (2.6-2.5)	30-1.5 (1.6-1.5)	30-2.7 (2.8-2.7)
No. observations	140981	386384	69457
No. unique reflections[b]	27524	61118	24119
Completeness (%)[c]	99.9 (100.0)	98.4 (96.8)	95.3 (97.3)
R_{merge} (%)[a,c]	5.7 (41.8)	4.3 (54.8)	7.5 (54.9)
I/σ (I)[a]	16.3 (2.8)	20.4 (2.6)	9.2 (2.5)
Resolution range (Å)	See HcS:**2**	10-1.5	15-2.7
No. reflections working set		57110	22913
No. reflections test set		3006	1206
No. non hydrogen (protein)		2267	4622
No. of heteroatoms		16	-
No. of solvent water		278	110
R_{work}/R_{free} (%)[d]		14.3 / 18.9	20.6 / 24.7
rmsd bond lengths (Å) / (°)[e]		0.019 / 1.61	0.005 / 0.97
Average B-factor (Å^2)		33.9	57.6
Ramachandran Plot (%)[f]		100 / 0.0 / 0.0	99.1 / 0.9 / 0.0
PDB accession code		4MC3	4MC0

[a] The values in parentheses of resolution range, completeness, Rmerge and I/ σ (I) correspond to the last resolution shell. [b] Friedel pairs were treated as different reflections. [c] $R_{merge}(I) = [\sum_{hkl}\sum_j |[I(hkl)_j - I(hkl)]|]/ [\sum_{hkl}\sum_j I_{hkl,j}]$, where $I(hkl)_j$ is the measurement of the intensity of reflection *hkl* and $\langle I(hkl)\rangle$ is the average intensity. [d] $R = \Sigma_{hkl} | |F_{obs}| - |F_{calc}| |/ \Sigma_{hkl} |F_{obs}|$, where R_{free} is calculated without a sigma cut off for a randomly chosen 5% of reflections, which were not used for structure refinement, and R_{work} is calculated for the remaining reflections. [e] Deviations from ideal bond lengths/angles. [f] Number of residues in favoured region / allowed region / outlier region.

4.2.5. HcS:apo Structure

The apo structure of HcS closely resembles the SdS:apo structure. An alignment (combinatorial extension alignment[31]) of these two structures results in a RMSD of 2.7 Å over 272 residues. The α-fold comprises 11 antiparallel helices (A-K), which will be described in the following. The first three amino acids ^{1}MAE^{3} are not defined in the crystal structure. Helix A (^{21}LEEASRAMWEWIDAN^{35}) is preceded by a sequence (^{4}FEIPDFYVPFPLECNPH^{20}) exhibiting no secondary structure. Helix B (^{41}ERARDRMRRTGADLSGAYV^{59}) is connected to helix A by the loop sequence ^{36}GLAPT^{40} and to helix C (^{65}LDTLTIGLKWIALTFRIDDQ^{84}) *via* the sequence ^{60}WPRAD^{64}. Parts of the C-terminus of helix C and the adjacent loop region (^{89}DTAERL^{94}) are structurally not defined (^{86}DEDDTAER^{93}). Therefore, the third aspartate from the ^{82}**DD**QI**D**^{86} motif is not visible in the crystal structure. Helix D (^{95}PARMTAIDELRGTLH^{109}) is connected to helix E (^{118}PTARALGALWQETA^{131}) by the loop sequence ^{110}GLPVSGRS^{117}. Hereby, residues ^{112}PVSGR^{116} are flexible and therefore in the electron density map not defined. The sequence ^{132}LGRP^{135} links helix E with helix F (^{136}ATWCDAFIGHFEAFLQTYTTEAGLN^{160}). Parts (^{161}AHGAG^{165}) of the subsequent loop region ^{161}AHGAGLR^{167} are not visible in the crystal structure. Helix G1/2 (^{168}LDDYLDRRMYSVGMPWLWDLDELR^{191}) comprises the effector triad, previously described for SdS (pyrophosphate sensor R175, linker S178 and effector residue V179). Next, Helix H (^{198}GSVRTCGPMNKLRRAGALHIALVNDVFS^{225}) is connected by loop ^{192}LPIFLP^{197} to helix G and with loop ^{226}VERETLVGYQHN^{237} to helix I (^{238}AVTIIREAQ^{246}). Parts (^{227}ERETLVG^{233}) of this loop are structurally not defined. Helix H features the **ND**(L,I,V)x**S**xxx**E** motif (^{221}**ND**VF**S**VER**E**^{229}). The last strictly conserved residue of this motif (E229) is not present in the structure. Helix I is connected to helix J (^{250}LQEAVDQVAVLVEAQLHTVLQARQELLEELDRQ^{282}) by the short sequence ^{247}GCS^{249}. Helix K (^{286}SRAREAAVDYAANVAANLSGQLVWH^{310}) is the last visible secondary structure whereas the C-terminus ^{311}SSVERYAVDDLQSAADPRATPTTSSLGI^{338} is again structurally distorted (**Figure 39** and **Figure 40**).

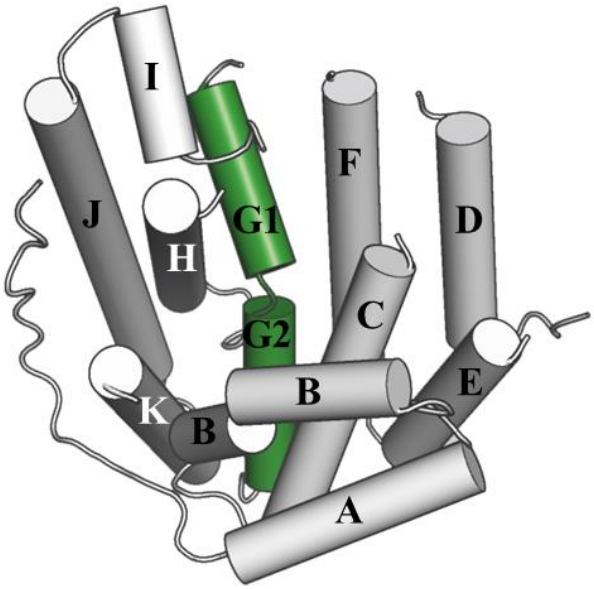


Figure 39. A scheme of helices A- K of HcS:apo. This α-fold is almost identical to the SdS:apo structure. Helix G1/2 and the corresponding helix-break are highlighted in green.

One obstacle of bioinformatic analysis of class I terpene cyclase is the low primary sequence conservation. Since the information of an enzyme's three dimensional structure is stored in its primary sequence, and the enzyme's product spectra is based upon it tertiary sequence, the information about the class I terpene cyclase's product(s) must ultimately be featured in the primary sequence as well. To investigate this, a primary sequence alignment (MSAProbs[87]) of HcS and SdS was done. Moreover, above these sequences the secondary structures (helices A- K) are shown (based on the crystal structures). In addition, residues which are part of the active site are highlighted in red. By this analysis, we can proof the correctness of the primary sequence alignment regarding its prediction of active site residues. Furthermore, the conservation of the class I terpene cylcases' secondary structures could be analysed (**Figure 40**).

Figure 40. A primary sequence alignment of HcS and SdS. Green bars indicate α-helices from HcS, whereas the grey ones correspondent to SdS. Red residues highlight active site residues.

The inspection of both sequences reveals that helices A and B are conserved regarding their primary- and secondary structures. The active site residues located on helix B are S55, V59 for HcS and F55, I59 for SdS. These amino acids are predicted by the primary sequence alignment. Helix C seems to be conserved in both enzymes as well, still HcS:**apo** is partly missing the helix's C-terminus (open conformation). The first active site residue on this helix (for both enzymes I75) as well as the following amino acids T82 (HcS) and L82 (SdS) are aligned correctly, too. Interestingly, SdS exhibits the adjacent amino acid (F83) to be located within the active site, which is not the case for HcS. Interestingly, the primary structure illustrates that HcS actually features a phenylalanine at this position as seen in SdS. This observation indicates that neither the primary nor the secondary structure alignment can identify all active site residues located on this helix. The comparison of helices D reveals a gap in the primary sequence alignment for SdS, which upon investigation of the secondary

structures however proofs to be wrong. Here, the secondary structure alignment (in case of an unknown protein fold using a predicted model) is more accurate. These observations are true for helix D as well as helices E and F. In case of the latter, the misalignment of the primary sequence mismatches the active site residues of HcS (F149, Y153) and SdS (Y152, V156, V157). Notably, as observed for helix C, SdS exhibits an additional residue located on helix F which contributes to the active site compared to HcS. According to its functional importance, Helix G1/2 is in both enzymes highly conserved regarding its secondary structure. Remarkably all amino acids which are part of the catalytic center (HcS: V179, G180, M181, W183, L184; SdS: G182, A183, T184, V186, V187) are correctly positioned according to the primary sequence alignment. Since helix H features the important **ND**(L,I,V)x**S**xxx**E** motif, its secondary structure is conserved as well. In both cases the only amino acid contributing to the lower half of the active site located on this helix is I217 (HcS) and I220 (SdS). Helix I is well aligned (primary- and secondary structure) for both class I terpene cyclases. Helix J displays a small variation, being four amino acids longer in case of HcS compared to the situation in SdS. In case of SdS, Helix K comprises four amino acids contouring the active site: F297, A301, W304 and Y311. In contrast to this, HcS exhibits five amino acids forming the catalytic chamber located on helix K: N 302, Q306, W309, H310 and Y316. Except for H310 (HcS), all of these amino acids are well aligned according to the primary sequence. In summary, four of the seventeen (24%) active site amino acids from HcS are misaligned compared to SdS[87]. Partly, this mismatch can be overcome by aligning all secondary structure elements. But in some cases, the helices' characteristic twist (unique for every single class I terpene cyclase) includes or excludes amino acids from the active site, which is not predictable based on the primary- and secondary structure.

4.2.6. HcS:2 Structure

Similar as observed for the HcS:apo structure, the nerolidol complex also displayed the open conformation. By omitting the $MgCl_2$ from the lysis buffer, we aimed to disrupt the ongoing conversion of endogenous FPP, thereby possibly trapping reaction intermediates. Upon careful investigation of the electron density map, we could identify a non-proteinogenic ligand. The F_O-F_C electron density was perfectly shaped for (*R*)-nerolidol, a naturally occurring, non-physiological side product of HcS (**Figure 44**). **Figure 41** shows the underlying reaction scheme forming this compound.

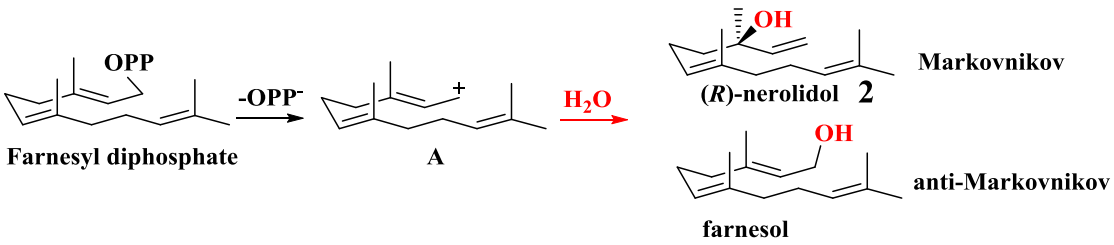


Figure 41. The reaction scheme of the water quenching reaction of carbocation **A** is shown. Hereby, the Markovnikov product (nerolidol, **2**) or the anti-Markovnikov product (farnesol) can be formed.

After diphosphate abstraction, the carbocation reaction intermediate **A** can be quenched with H_2O generating two different products: the Markovnikov product nerolidol (**2**) and the anti-Markovnikov product farnesol[88]. From a thermodynamic point of view, the production of nerolidol is favoured. The asymmetric distribution of nerolidol's methyl groups allowed its correct assignment due to the high resolution of the dataset (1.5 Å). Moreover, we were able to extract the ligand from the purified enzyme and to analyse it on a chiral GC-MS by comparing it to purchased (*R*)- and (*S*)-nerolidol (**Figure 42**). This experiment was conducted in joint collaboration with Patrick Rabe from the group of Prof. Dr. Jerome Dickschat at the University of Bonn.

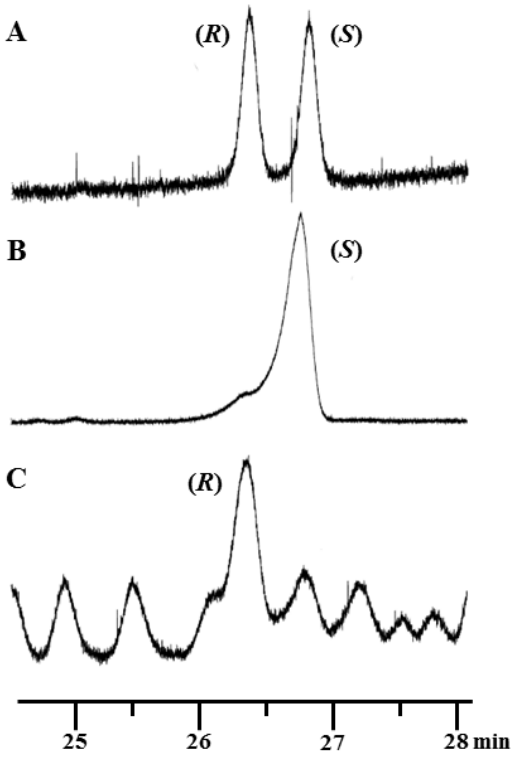


Figure 42. Chiral GC-MS chromatograms for the identification of (*R*)/(*S*)- nerolidol. **A)** Analysis of a racemic mixture of (*R*)- and (*S*)-nerolidol. **B)** Chiral GC-MS run, analyzing (*S*)-nerolidol (purchased). **C)** Product spectra of mutant R315K, which exhibits an increased production of (*R*)-nerolidol. Adapted from Baer et al[2].

In order to verify our conclusions drawn from the electron density map and the reaction mechanism catalysed by HcS we aimed to characterise the formation of nerolidol generated by HcS (R315K). Therefore, we first investigated the difference in retention time for (*R*)/(*S*)-nerolidol. This experiment clearly showed that (*R*)-nerolidol (around 26 min 20 sec) eluted prior (*S*)-nerolidol (upon comparison of pure (*S*)-nerolidol). Eventually, the nerolidol ligand extracted from the HcS:R315K mutant exhibits the retention time corresponding to (*R*)-nerolidol. Importantly, this ligand featured two important aspects, which so far were not described in a class I terpene cyclase structures. First, the diphosphate moiety was abstracted and therefore, the ligand was not restricted to the entrance of the active site. This allowed the ligand to deeply bind into the catalytic chamber, were the ligand's pre-folding could be visualized. Second, the ligand mimics a naturally occurring terpene product. By this, the significance could be considered to be higher compared to artificial, inhibitory substrate analogues. To investigate the importance of the active site's contour and its impact on the pre-folding of the ligand, the compound's coordination sphere will be described next. The coordinating amino acids in close contact to the ligand are: S55 (3.2 Å), V59 (4.6 Å), I75 (4.0 Å), T78 (3.8 Å), F149 (4.0 Å), Y153 (4.8 Å), V179 (3.7 Å and 2.9 Å), M181 (4.4 Å), W183 (4.1 Å), I217 (3.6 Å), N221 (3.6 Å), Q306 (3.6 Å) and H310 (3.8 Å, **Figure 43** and **Figure 44**).

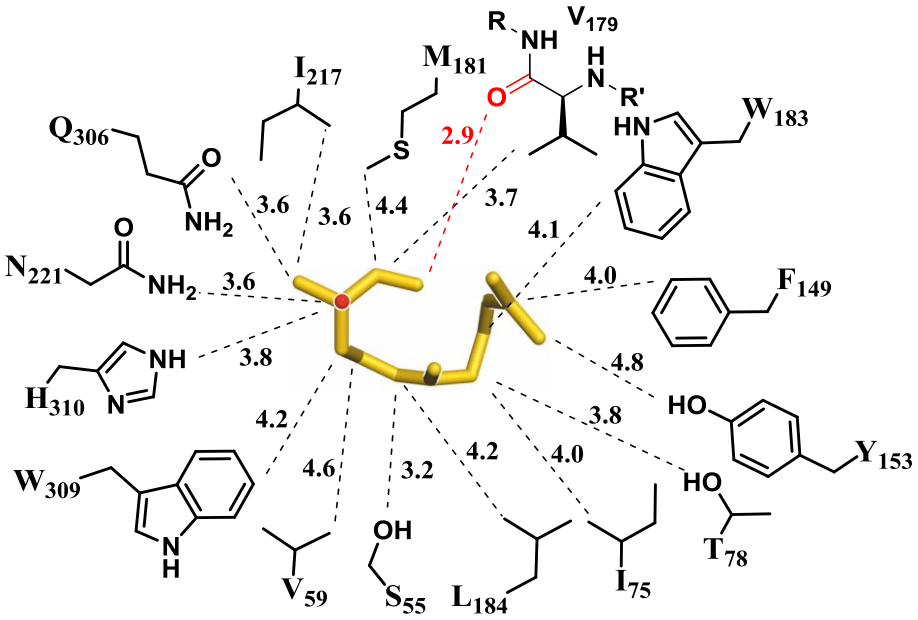


Figure 43. The coordination sphere of (*R*)-nerolidol (HcS:**2**) is displayed. Distances are given in Angstroms; dashed lines indicate the atoms closest to each other (amino acid/carbon atom). The ligand's oxygen is depicted in red, whereas the hydrogen carbon backbone is illustrated in gold.

Most of the active site residue described are highly hydrophobic and contribute to the enzyme's specificity by shaping the catalytic chamber. In contrast to SdS, HcS exhibits some polar amino acids within its central cavity, e.g. S55, T78 and Q306 as well. These residues could potentially interact with transient carbocations. All the other amino acids interact with the hydrophobic backbone *via* van der Waals forces. Similar to SdS, four aromatic amino acids are located within the active site and therefore can take part in carbocation stabilization (F149, Y153, W183 and W309). The most interesting chemical group interacting with **2** is the carbonyl oxygen of V179. This amino acid represents the effector of the effector triad, previously described in SdS. Surprisingly, in the HcS:**2** complex structure this oxygen is in close contact (2.9 Å) towards the ligand's C1 position (in contrast to SdS, where this carbonyl group is directed at the substrate's C3). Furthermore, the overall G1 helix points at C1 and carbocation stabilization supported by this architecture is very likely. The complex structure of HcS:**2** is illustrated in **Figure 44**.

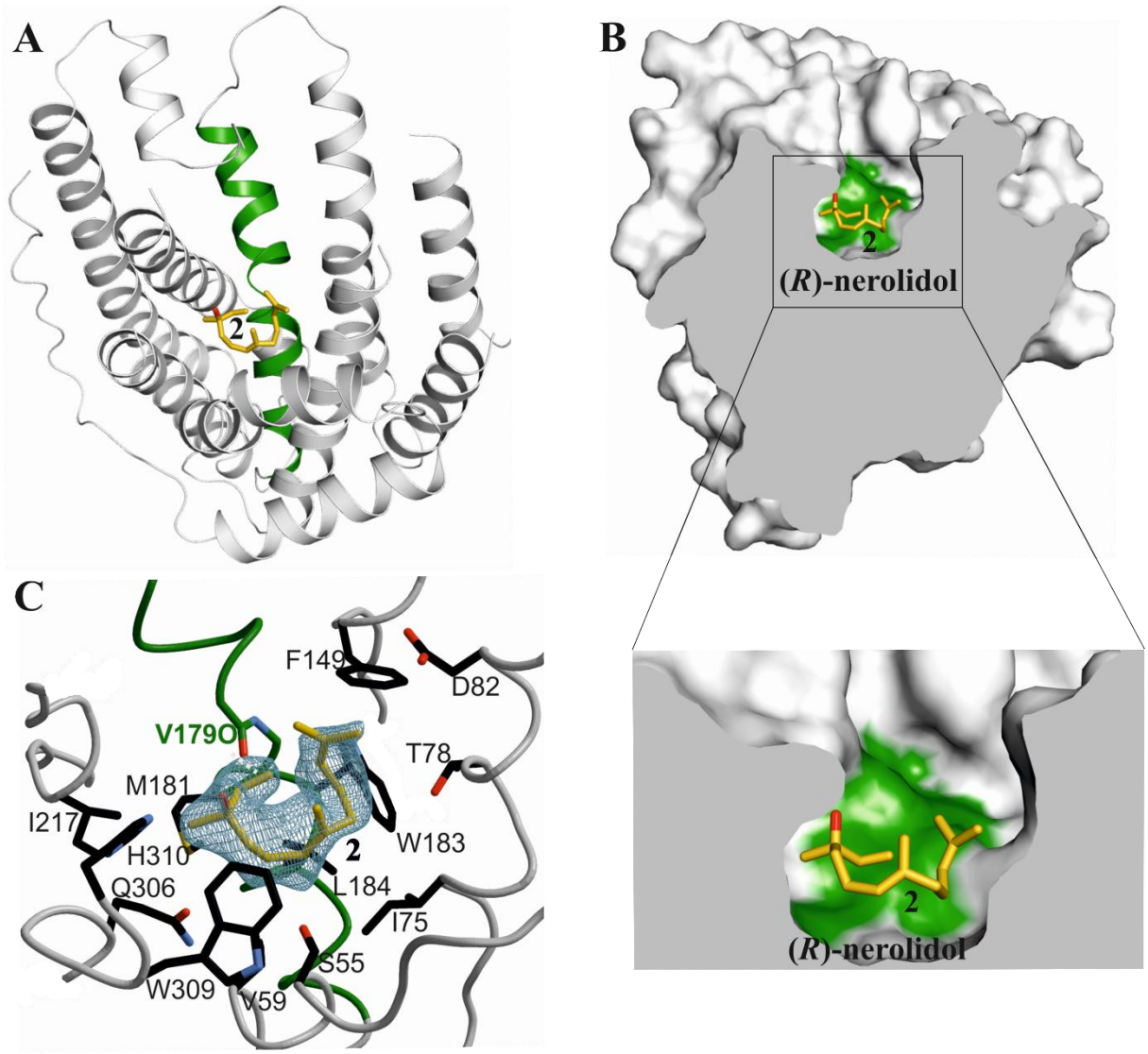


Figure 44. The HcS:**2** complex. **A**) Cartoon representation of the entire enzyme; helix G1/2 is highlighted in green, (*R*)-nerolidol (**2)** is coloured in gold. **B**) A cross section of HcS highlights the pre-folded ligand within the active site. **C**) The $2F_o$-F_c electron density omit map (contoured at 1σ) for (*R*)-nerolidol is shown. Adapted from Baer et al[2].

The inspection of the HcS:**2** complex shows the ligand to bind deep into the active site cavety. Compared to a closed conformation (for example SdS:**PP$_i$**), the PP$_i$-$(Mg^{2+})_3$-cluster is directly coordinated above the nerolidol ligand. The most striking feature of the metabolite is its high degree of pre-folding, which mimics the final product (2*Z*,6*E*)-hedycaryol. From a thermodynamic point of view the compound is still linear and not conformationally restricted by intramolecular bonds. There a three different driving forces which support this entropic restricted intermediate: 1) The active site is closed upon strong coordination of the PP$_i$-$(Mg^{2+})_3$-cluster (cf. with SdS:**PP$_i$**), thereby forming a spatially constrained reaction chamber.

2) The highly polar cap of the closed active site is repulsive towards the hydrophobic ligand. 3) The hydrophobic active site preferentially interacts with this kind of ligand *via* van der Waals forces. In combination of these three important parameters, it can be concluded that the conformation of the pre-folded ligand closely resembles the structure of the product even before the first intramolecular bond was formed. Hence, this kind of product control allows class I terpene cyclases to be specific even for a single compound though terpenoids are highly diverse in their overall structure. Therefore, the cyclisation reaction catalysed by specific class I terpene cyclases exhibits a product like reaction intermediate (slow kinetics) and not an educt like reaction intermediate (fast kinetics), which builds the basis for highly specific class I terpene cyclases. This is in line with the Hammond-Leffler postulate[89][90]. Furthermore, the HcS:**2** complex structures provides an explanation for the exclusive 1,x- ring closure reaction (x is one of the double bonds present in oligoprenyl diphosphates) taking place as a first step during terpenoid biosynthesis. This is remarkable since the nucleophilic attack of a double bond at the carbocation at the C1 position represents the thermodynamically less favoured anti- Markovnikov product. Hereby, the carbonyl oxygen of V179 and the helix dipole of helix G1 both contribute to stabilize the positive charge of the carbocation at the less preferred C1 position. Therefore, the enzymatic mechanism prevents formation of unwanted by-products (**Figure 45**).

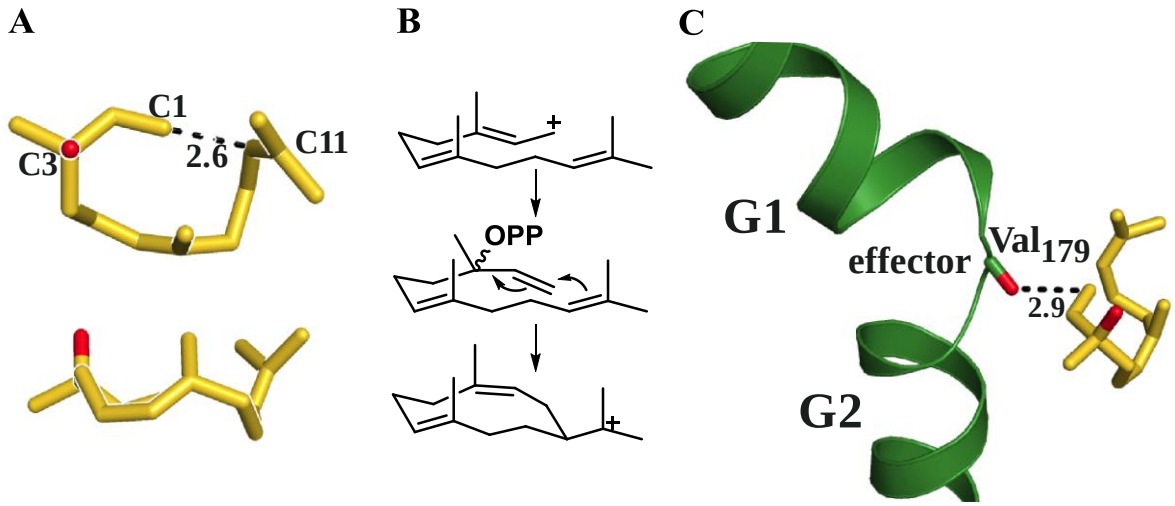
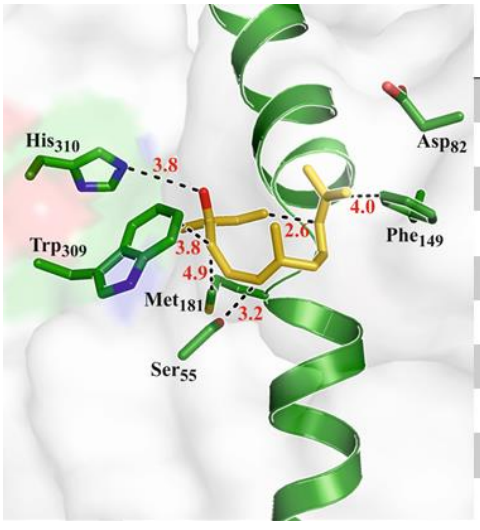


Figure 45. A) A close up view of the nerolidol ligand, **B**) Reaction scheme of the first ring closure and **C)** the carbonyl oxygen of the effector residue stabilizing the carbocation at the substrate's C1 position (Val179 for HcS). Distances are given in Å, and C1/C3/C11 refer to the ligand's carbon atoms. Adapted from Baer et al[2].

The pre-folded nerolidol ligand exhibits a distance between C1 and C10 of 2.6 Å. In principal, this ligand represents either the nerolidyl diphosphate (NPP) or the primary carbocation reaction intermediate. For a long time, it was thought that the NPP is a requirement for changing the substrate's conformation from *trans* to *cis*, but this transition is also feasible starting from the primary carbocation[44][45].

4.2.7. HcS Mutants and Product Spectra

The bound (*R*)- nerolidol ligand (**2**) allowed a mapping of the active site based on structural data for the first time. Therefore, we were aiming to identify residues stabilizing and guiding the carbocation cascade with a set of mutants and to rationally alter the product spectrum. This was done for each mutant at a neutral- and basic pH (7.0 and 8.5, respectively). The two different pH values were chosen to investigate whether the protonation state of the mutants is of importance or not. In summary, we conducted the following mutations: S55W, D82N, F149L, F149W, M181H, M181K, W309L, W309F, W309Y, H310S, R315K and Y316F. All mutants were purified, analysed regarding their FPP conversion activity and the corresponding product spectra. These experiments were conducted in joint collaboration with Patrick Rabe from the group of Prof. Dr. Jerome Dickschat at the University of Bonn. A summary of the GC-MS chromatograms and the amino acids' position relative to the ligand are shown in **Figure 46**.



Mutant	pH 7 (Activity)	pH 8.5 (Activity)
S55W	-	-
D82N	-	+
F149L	-	+
F149W	wt	wt
M181H	-	+
M181K	-	-
W309L	wt	wt
W309F	wt	wt
W309Y	wt	wt
H310S	-	-

Figure 46. Active site mutants involved in the (*R*)-nerolidol binding and their activity relative to the wild type enzyme. Distances are given in Å. All mutants were tested *in vitro* with two different pHs (7.0 and 8.5). "-" indicates inactive mutants, "+" highlight active mutants and "wt" marks the corresponding mutant to be as active as the wild type enzyme.

In the following section, the results of the incubation experiments are reported and discussed. Mutant S55W, which is located at the bottom of the active site, exhibits a complete loss in activity for both pHs. The next mutant investigated was D82N, which is part of the **DD**xx**D** motif. Again, there was no activity for pH 7.0 but residual activity for pH 8.5. F149 is located near C11 (4 Å distant) and therefore was thought to participate in stabilization of carbocation **B**. Since F149L exhibits no activity at pH 7.0 and just some minor substrate conversion activity at pH 8.5, the role of F149 in carbocation chemistry is confirmed. This observation is supported by mutant F149W, which displays for both pHs wild type activity. The larger size of W compared to F did not impaired with the active site. M181 is located 4.4 Å underneath the ligand's C2 atom. Mutant M181K is inactive at both pH values, whereas mutant M181H shows residual activity at pH 8.5 (at pH 7.0 it is inactive as well). This nicely demonstrates the inhibitory influence of a positive charge upon a carbocation reaction intermediate. In case of K (pH 7.0 and pH 8.5) and H (pH 7.0) the positive charge abolishes enzymatic activity, whereas for H (pH 8.5) residual activity could be observed. All mutants of W309 (W309L, F and Y) exhibit wild type enzymatic activity, therefore a participation of this residue in carbocation stabilization is ruled out. H310 is in close contact to the hydroxyl group of the nerolidol ligand (3.8 Å). In case of the closed conformation, this residue is probably directly involved in the diphosphate coordination. The H310S mutant exhibits for both pH values no activity, demonstrating that the H-bond network formed with the substrate's diphosphate moiety is of great importance. The last mutants investigated were R315K and Y316F. These residues are strongly conserved on the primary sequence of class I terpene cyclases and are incorporated in the H-bond forming network of PP_i-$(Mg^{2+})_3$ coordination, as described for SdS. Despite this conservation, both mutants exhibit wild type activity. The GC-MS chromatograms corresponding to these different mutants are shown in **Figure 47**. Compound numbers **1** and **2** refer to (2*Z*,6*E*)-hedycaryol and (*R*)-nerolidol, respectively. The negative control (no HcS) reveals a minor terpene formation which can be contributed to spontaneous hydrolysis of FPP in the *in vitro* experimental setup.

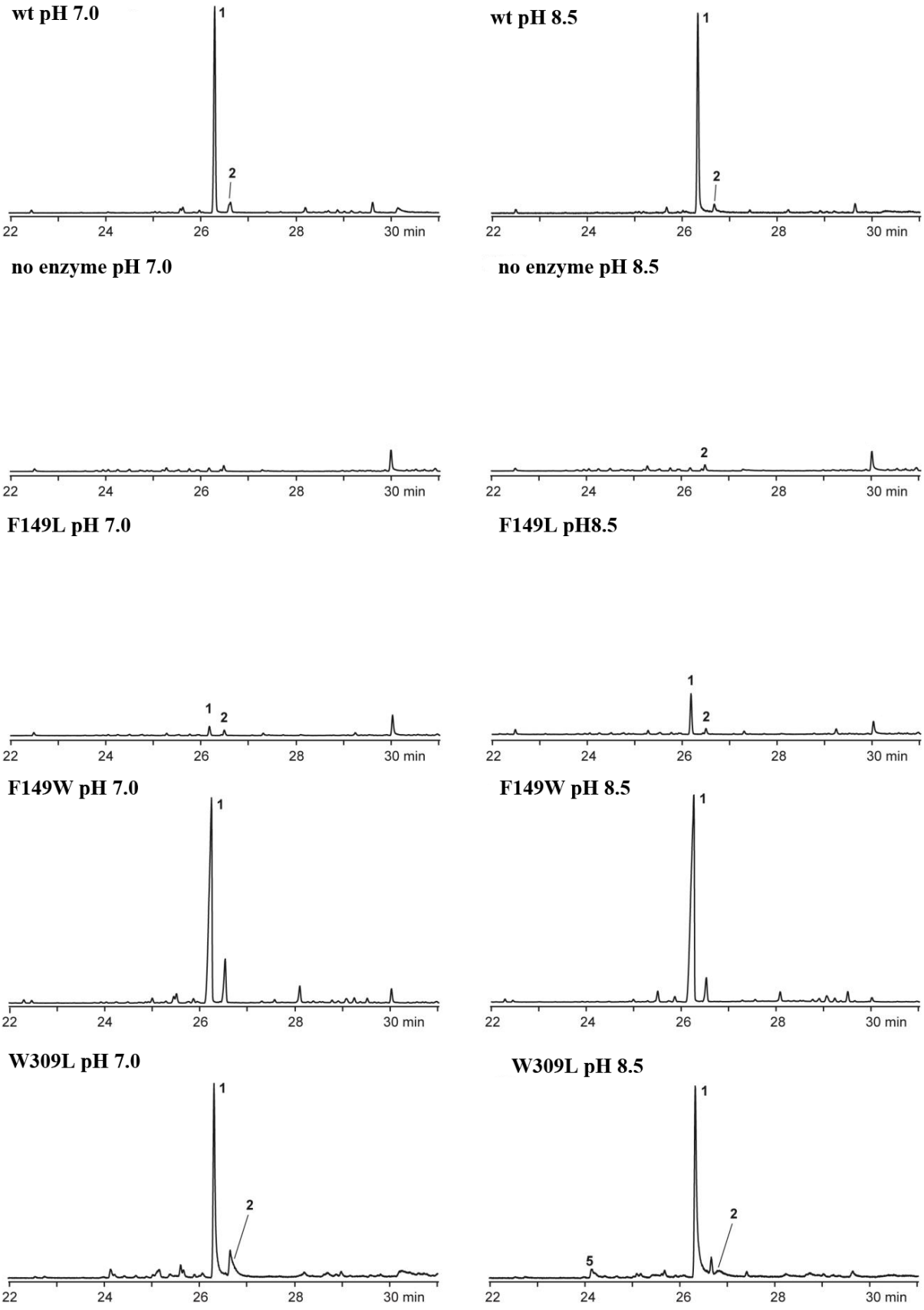


Figure continues on the next page

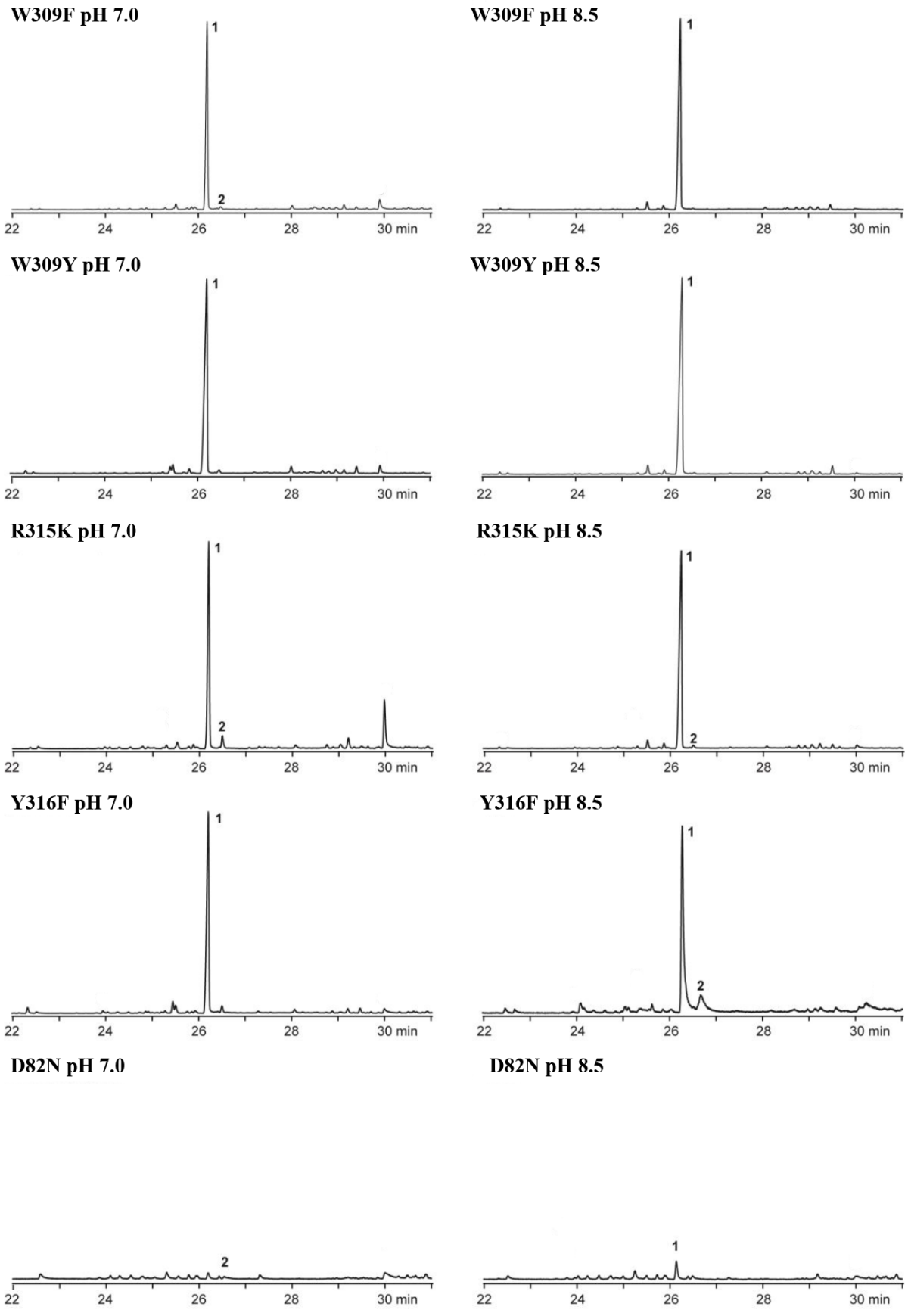


Figure continues on the next page

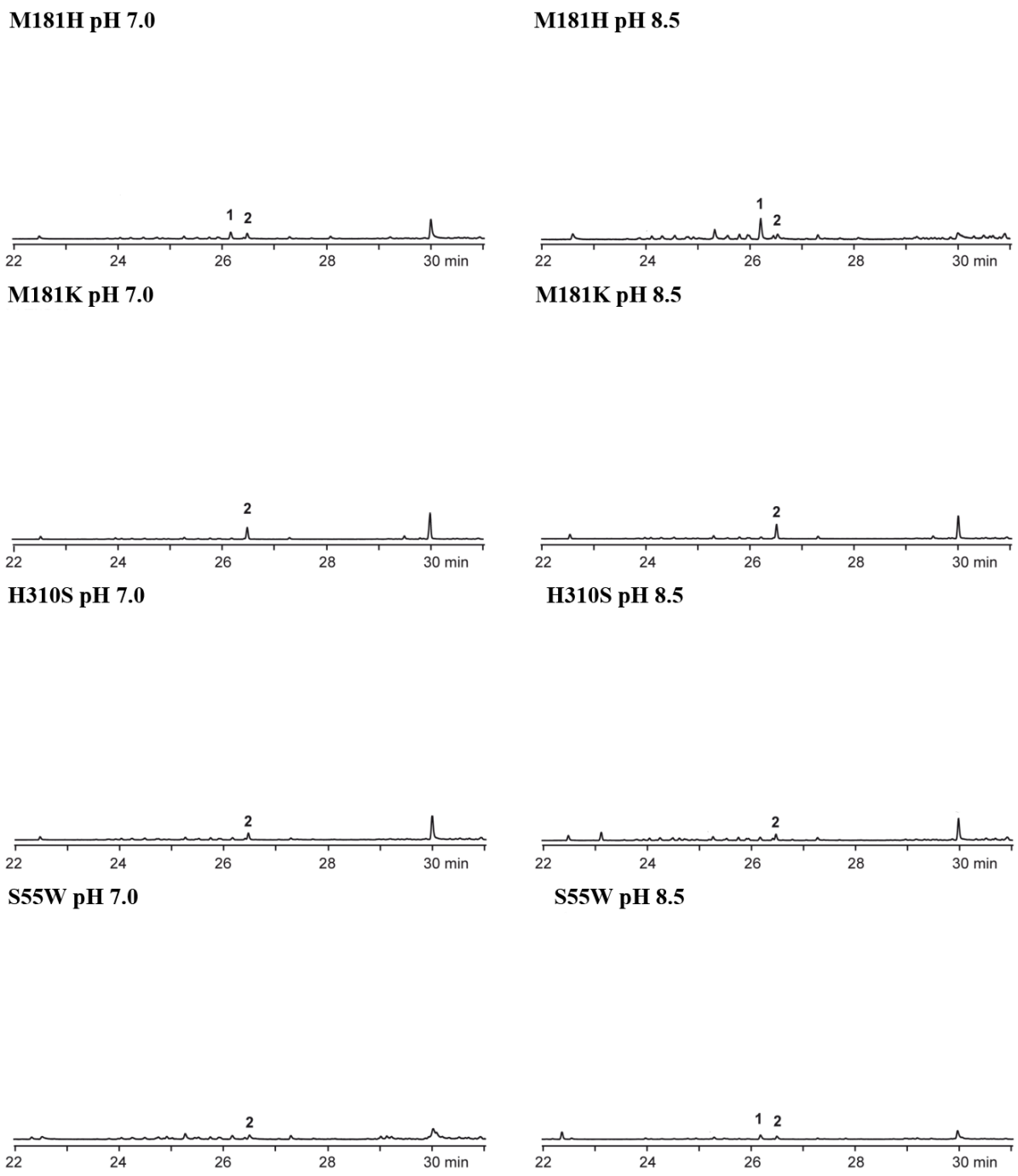


Figure 47. The GC-MS chromatograms of the following mutants are shown: S55W, D82N, F149L, F149W, M181H, M181K, W309L, W309F, W309Y, H310S, R315K and Y316F. Compound numbers **1** and **2** refer to (2*Z*,6*E*)-hedycaryol and (*R*)-nerolidol, respectively. Adapted from Baer et al[2].

4.2.8. Discussion HcS

The present study comprises the crystal structures of (2*Z*,6*E*)-hedycaryol synthase in its open conformation (apo) and in complex with (*R*)-nerolidol. Since class I terpene cyclases are not accessible for phase calculations based on molecular replacement, native HcS:**2** crystals were soaked with $HgCl_2$ and the anomalous datasets were analysed applying SAD-methods. We analysed the enzyme's substrate specificity and its product spectrum. Further, HcS's ligand binding preference was investigated applying circular dichroism based thermal shift assays. The mechanistic insights obtained and enzymatic models formulated were challenged and expanded by designing 12 different point mutants. These mutants were analysed regarding their activity and their individual product spectra at two different pH values (7.0 and 8.5).

Though, HcS was crystallographically analysed only in its open conformation, the binding of the inhibitor (2F- FPP) was demonstrated applying circular dichroism spectroscopy based thermal shift assays. Interestingly, PP_i and Mg^{2+} on its own were not sufficient to bind towards HcS and to close the enzyme's active site. This highlights the importance of the substrate's hydrophobic hydrocarbons backbone for binding to the enzyme's active site. Without these extensive interactions of the ligand's diphosphate group and its hydrophobic backbone, the active site closure cannot be achieved. This observation is also crucial for understanding the enzyme's catalytic cycle, because at one point, the synthesised product has to be released. Since the substrate's diphosphate moiety and the hydrocarbon backbone are detached after primary carbocation formation, the closed active site is rather unstable and presumably re-opens after a short time, hereby releasing the product.

The apo conformation of HcS closely resembles the structure of pentalenene synthase[47]. It displays a G1 helix-break arrangement identical to the one observed for the closed conformation of SdS:**DHFPP**. Upon superposition of SdS:**apo**, SdS:**PP_i** and HcS:**apo**, all potential conformational states of the helix-break G1 can be visualized, as shown in **Figure 48**.

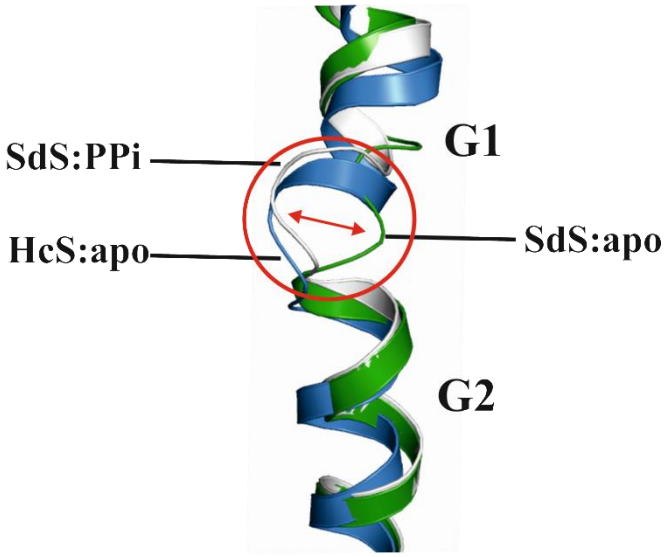


Figure 48. Structural superposition of the helix-break motif of helix G1/2. SdS:**PP$_i$** is coloured in grey, SdS:**apo** in green and HcS:**apo** in blue. The red arrow indicates the transition between the two different helix-break conformations.

Interestingly, the apo conformation is not necessarily displaying the helix-break arrangement observed for SdS:**apo**. This underlines the high structural flexibility of this peculiar architecture. Therefore, it would be very interesting to investigate, if the HcS:**apo** conformation is already capable of substrate binding. At least, the catalytic importance of the apo-conformation featured in SdS:**apo** was demonstrated in the presented work by SdS mutants like G182V and G182P as well as A183V and A183F.

The complex structure of HcS:**2** ((*R*)- nerolidol) represents to our knowledge the first class I terpene cyclase structure in complex with a naturally occurring linear sesquiterpene. Since the pyrophosphate group is not present in (*R*)-nerolidol, the ligand is deeply bound within the active site, thus exhibiting a complete pre-folding. This specific conformation is a prerequisite for a class I terpene cyclase to specifically produce a single product out of the thousands of possible products. In addition to these conclusions, the HcS:**2** complex reveals experimental, structural insights of the so far hypothesized nerolidyl diphosphate (NPP) reaction intermediate[44]. A superposition of HcS:**2** and SdS:**apo** displays that the nerolidol's C1 and C3 atoms are both more than 5 Å distant to the diphosphate group. Since PP_i is strongly coordinated to the active site's entrance, it cannot rearrange to come into close contact to the ligand's C1/C3 atoms for re-attacking the carbocation reaction intermediate **A** (for nomenclature cf. **Figure 41**). In addition, the HcS:**2** structure clearly exhibits the ligand's preferred mode of binding within the active site after diphosphate abstraction. These findings

exclude an alternative binding mode of **A**, which would be expected to be in contact with the attached diphosphate group as demonstrated in **Figure 49**.

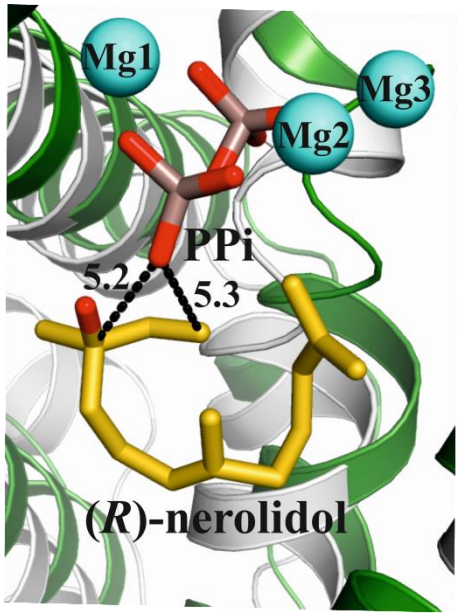


Figure 49. Close up view of the HcS:**2** (grey) and SdS:**PP$_i$** (green) active sites. The dashed lines indicate the distance between PP$_i$ and the (*R*)-nerolidol's C1 (5.3 Å) and C3 atoms (5.2 Å).

Upon investigation of the superimposed active sites of HcS:**2** and SdS:**PP$_i$**, it is obvious that both ligands (*R*)-nerolidol and PP$_i$ are far distant to each other. Therefore, the generation of nerolidyl diphosphate in course of the cyclization reaction to switch from *cis* to *trans* is expected to be unlikely according to the obtained crystal structures.

The bound (*R*)-nerolidol allows a mapping of the active site and thus significantly facilitated the design of suitable point mutants, which are most important to understand the underlying carbocation chemistry during catalysis. The designated carbonyl oxygen of Val179 of the effector triad, analogue to the Gly182 residue in SdS, is perfectly aligned to stabilize the positive charge at the substrate's C1 position, catalysing the 1,x- ring closure (hereby, a double bond of the substrate attacks nucleophilic at the C1 atom). Another structural feature enabling the reaction is the orientation of the negatively polarized helix dipole (C-terminus) of G1, explaining the formation of the anti-Markovnikov product. Both of these structural elements are able to stabilize the carbocation without being nucleophiles. Therefore, covalent bond-formation of the carbocation reaction intermediates to the enzyme does not take place.

To proof our proposed models and to investigate the structure-function relationship, we designed 12 different point mutants of HcS and investigated the *in vitro* FPP turn over for each. Ser55 was exchanged with a Trp side chain based on the primary sequence alignment between HcS and a (+)-caryolan-1-ol synthase (CpS[91]). Surprisingly, the S55W mutant didn't show detectable activity. However, upon inspection of different class I terpene cyclase structures, it turned out that a standard primary sequence alignment of these enzymes does not reliably reveal the active site forming residues. This is explained with the fact that the entire catalytic centre is formed by 5 different α helices. Since minor variations within each helix can turn a distinct amino acid into or out of the active site, it is difficult to predict the amino acids contouring the catalytic centre of class I terpene cyclases. The D82N mutant supports the role of this aspartate residue in water activation for the final quenching of the carbocation with H_2O in HcS. This assumption is based on the residual activity observed at pH 8.5, but which is absent at pH 7.0. Thus, at physiological pH a H_2O activating residue, to generate a nucleophilic water molecule, has to be close by for catalysing this reaction step. Phe149 is the key residue for stabilizing the carbocation at C11. Since in (2*Z*,6*E*)-hedycaryol formation the carbocation at C11 is the only one occurring during catalysis and it is subsequently quenched by water, a residual activity of F149L can be observed at pH 8.5. An immediate attack of OH^- instantaneously removes the unstable carbocation, this way resulting in a residual enzymatic activity at pH 8.5. F149W exhibits the same catalytic activity as the measured wild type enzyme. Mutants of Met181 allow the most fascinating insights of all tested variants. This amino acid is located below the substrate's C2 atom at a distance of 4.4 Å, a position on the substrate with a temporary positive polarity during catalysis. In accordance with this, the M181H mutant reveals substrate turnover at pH 8.5 (no protonation, no positive charge) while it is inactive at pH 7.0 (positive charge destabilizes carbocation chemistry). In contrast, the positively charged lysine in variant M181K prevents any catalysis independent of the pH value. This positive polarity destabilizes the carbocation **A**, hereby preventing substrate turnover. The Leu, Phe and Tyr mutants of Trp309 all exhibit wild type activity and the residue therefore can be considered to be not important for the carbocation chemistry. H310S is inactive as well. We propose that His310 stabilizes the diphosphate moiety which cannot be accomplished by a serine residue. Therefore, PP_i coordination is obstructed. The last two mutants tested are R315L and Y316F which both show wild type activity. These two amino acids are highly conserved among bacterial class I terpene cyclases. It is therefore rather surprising that no decrease in activity was observed for the respective mutants.

Based on these structural insights and the mutants analysed, calculations of the ligand cyclisation in the context of the specific active site architecture could be performed in future experiments. This will further contribute to the investigation of the structure-function relationship in class I terpene cyclases and the understanding of the link between primary- and tertiary structure. Therefore, it would be most desirable at some point to correlate the class I terpene cyclases' products with their respective primary sequence. The presented results in the Ph.D.-thesis might support this splendid purpose at least from a structural and mechanistic point of view.

5. Conclusion

Structural data of class I terpene cyclases have now been available for over 15 years[10]. Even though, the overall fold of this structurally well conserved enzyme class is known in detail, fundamental aspects could not be answered to date. It is known that the active site is closed upon substrate binding[50]; yet it is still obscure, *when* and *how* substrate activation takes place. The achieved structural data of SdS and the described **induced-fit mechanism** explain for the first time the interplay of substrate binding and activation by a sophisticated molecular mechanism. This perfectly trimmed conformational coordination allows this enzyme class to perform carbocation chemistry in aqueous solution. A novel **effector triad** was elucidated including the pyrophosphate sensor Arg178, the linker Asp181 and the effector Gly182, which enables this fundamental biological function in class I terpene cyclases. It is noteworthy, that this structural motif can be found in all class I terpene cyclases (mono-, sesqui- and diterpene cyclases) from bacteria, fungi and plants. Therefore, the induced-fit mechanism identified most likely applies to all members of this enzyme class. The HcS:**2** ((*R*)-nerolidol) complex structure illustrates that after abstraction of the diphosphate and primary carbocation formation, the linear hydrophobic ligand is re-orientating within the active site. Thereby, the pre-folded substrate mimics the final product (Hammond postulate), at least in case of specific class I terpene cyclases. This pre-folding of the ligand probably explains the high specificity of HcS. The effector residue (Val179 in HcS) and the G1 helix dipole stabilise the carbocation at atom C1, hereby catalysing the 1,x ring closure (anti-Markovnikov product). Subsequently, Wagner-Meerwein and Cope rearrangements are guided by the pre-folded ligand and the architecture of the active site. This has been investigated by assaying extensive mutant libraries. The HcS:**2** complex structure allows for the first time a mapping of the active site based on structural data. The complete enzymatic cycle of class I terpene cyclases, based on structural data of SdS and HcS, is shown in **Figure 50**.

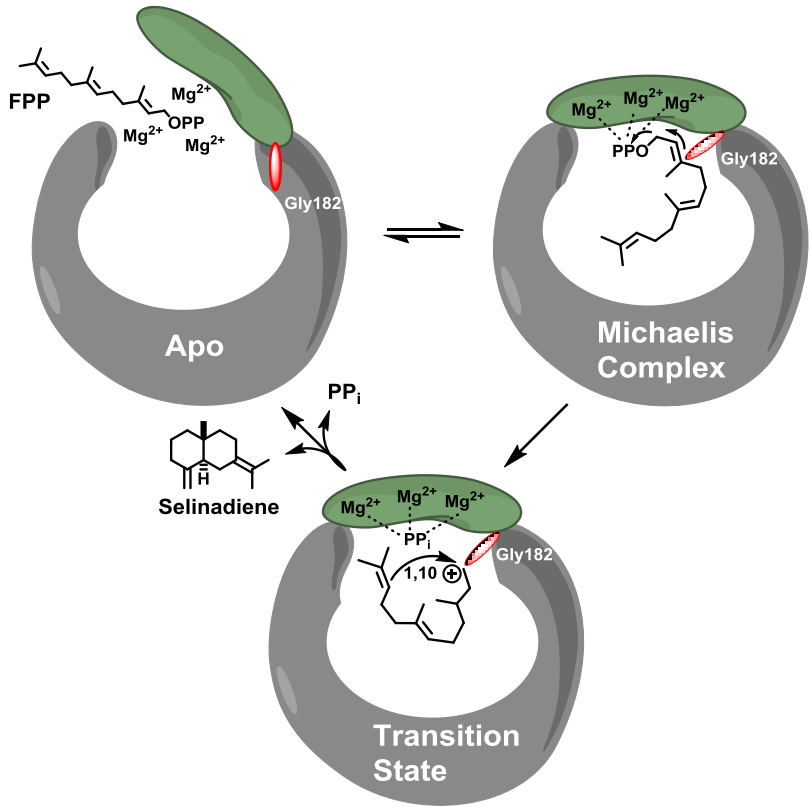


Figure 50. A scheme of the catalytic cycle of class I terpene cyclases, based on structural data of SdS and HcS. Adapted from Baer et al[1].

This catalytic cycle which accounts for the large family of class I terpene cyclases starts with the apo conformation (open). In this state, the catalytic centre is accessible for substrate binding. Hereby, the effector residue (Gly 182 in SdS, Val 179 in HcS) is turned away from the active site. Upon substrate binding and $(Mg^{2+})_3$ coordination, the active site closes and the induced-fit mechanism rearranges helix-break G1, bringing the effector residue in close contact to the substrate's C3 atom. This step represents the Michaelis complex. The molecular restructuring of helix-break G1 and the effector's carbonyl group leads to abstraction of the diphosphate, generating the primary carbocation (transition state). Subsequently, the ligand rearranges inside the active site, now bringing the substrate's C1 position close to the effector carbonyl group. This and the negative polarity of the helix G1 dipole stabilises the carbocation at C1, favouring the first ring closure to exclusively form 1,x rings (anti-Markovnikov product). The pre-folded ligand and the lining of the active site with aromatic residues guide the subsequent Wagner- Meerwein and Cope rearrangements, yielding the final terpene. The product is released and the class I terpene cyclase shifts back into the apo conformation[1].

In terpenoid biosynthesis, the scaffold stage is followed by the chemical decoration of these molecules to yield the bioactive compounds. It is of great interest to understand and utilize the underlying oxygenases, which are conducting these final chemical transformations[92]. The most central question to this process is the source of electrons needed to activate the inherently unreactive hydrocarbon scaffolds. In most cases, these redox- partners (for P450 oxidases or FAD- dependent monooxygenases) are derived from the central metabolism of the organisms[37]. Therefore, they are most of the time not located within or nearby the biosynthetic gene cluster of a certain terpenoid[38]. This complicates the identification of the correct partner protein. The regioselective oxidation of an inactivated hydrocarbon is a great challenge from a chemical point of view. Since oxidations are the prerequisite for subsequent modifications, e.g. glycosylations[12] or the introduction of electrophilic groups, these chemical reactions are of great biotechnological importance[37]. Their application in synthetic biology and semi-synthetic strategies will greatly contribute to the production of bioactive natural products. Therefore, it is important to address these enzymes in future experiments.

6. References

[1] P. Baer, P. Rabe, K. Fischer, C. A. Citron, T. A. Klapschinski, M. Groll, J. S. Dickschat, *Angew. Chem. Int. Ed.* **2014**, *53*, 7652–7656.

[2] P. Baer, P. Rabe, C. A. Citron, C. C. de Oliveira Mann, N. Kaufmann, M. Groll, J. S. Dickschat, *ChemBioChem* **2014**, *15*, 213–216.

[3] M. S. Butler, *Nat. Prod. Rep.* **2008**, *25*, 475–516.

[4] G. W. Huisman, S. J. Collier, *Curr. Opin. Chem. Biol.* **2013**, *17*, 284–292.

[5] C. M. Dobson, *Nature* **2004**, *432*, 824–828.

[6] A. Kessler, I. T. Baldwin, *Science* **2001**, *291*, 2141–2144.

[7] J. L. Goldstein, M. S. Brown, *Nature* **1990**, *343*, 425–430.

[8] W. Eisenreich, M. Schwarz, A. Cartayrade, D. Arigoni, M. H. Zenk, A. Bacher, *Chem. Biol.* **1998**, *5*, R221–R233.

[9] L. C. Tarshis, M. Yan, C. D. Poulter, J. C. Sacchettini, *Biochemistry (Mosc.)* **1994**, *33*, 10871–10877.

[10] C. M. Starks, K. Back, J. Chappell, J. P. Noel, *Science* **1997**, *277*, 1815–1820.

[11] K. U. Wendt, G. E. Schulz, *Structure* **1998**, *6*, 127–133.

[12] L. Caputi, E.-K. Lim, D. J. Bowles, *Chem. – Eur. J.* **2008**, *14*, 6656–6662.

[13] A. J. Farlow, P. V. Bernhardt, J. J. De Voss, *Tetrahedron Asymmetry* **2013**, *24*, 324–333.

[14] J. F. Andersen, J. K. Walding, P. H. Evans, W. S. Bowers, R. Feyereisen, *Chem. Res. Toxicol.* **1997**, *10*, 156–164.

[15] P. K. Ajikumar, W.-H. Xiao, K. E. J. Tyo, Y. Wang, F. Simeon, E. Leonard, O. Mucha, T. H. Phon, B. Pfeifer, G. Stephanopoulos, *Science* **2010**, *330*, 70–74.

[16] P. Dzubak, M. Hajduch, D. Vydra, A. Hustova, M. Kvasnica, D. Biedermann, L. Markova, M. Urban, J. Sarek, *Nat. Prod. Rep.* **2006**, *23*, 394.

[17] J. Gershenzon, N. Dudareva, *Nat. Chem. Biol.* **2007**, *3*, 408–414.

[18] S. A. Ralph, M. C. D'Ombrain, G. I. McFadden, *Drug Resist. Updat.* **2001**, *4*, 145–151.

[19] F. Rohdich, A. Bacher, W. Eisenreich, *Biochem. Soc. Trans.* **2005**, *33*, 785.

[20] M. C. Y. Chang, J. D. Keasling, *Nat. Chem. Biol.* **2006**, *2*, 674–681.

[21] S.-W. Kim, J. D. Keasling, *Biotechnol. Bioeng.* **2001**, *72*, 408–415.

[22] D. J. Pitera, C. J. Paddon, J. D. Newman, J. D. Keasling, *Metab. Eng.* **2007**, *9*, 193–207.

[23] D.-K. Ro, E. M. Paradise, M. Ouellet, K. J. Fisher, K. L. Newman, J. M. Ndungu, K. A. Ho, R. A. Eachus, T. S. Ham, J. Kirby, et al., *Nature* **2006**, *440*, 940–943.

[24] C. E. Vickers, J. B. Y. H. Behrendorff, M. Bongers, T. C. R. Brennan, M. Bruschi, L. K. Nielsen, in *Microorg. Biorefineries* (Ed.: B. Kamm), Springer Berlin Heidelberg, **2015**, pp. 303–334.

[25] H. Tsuruta, C. J. Paddon, D. Eng, J. R. Lenihan, T. Horning, L. C. Anthony, R. Regentin, J. D. Keasling, N. S. Renninger, J. D. Newman, *PLoS ONE* **2009**, *4*, e4489.

[26] D. Morrone, L. Lowry, M. K. Determan, D. M. Hershey, M. Xu, R. J. Peters, *Appl. Microbiol. Biotechnol.* **2009**, *85*, 1893–1906.

[27] J. A. Chemler, M. A. Koffas, *Curr. Opin. Biotechnol.* **2008**, *19*, 597–605.

[28] K. Wang, S. Ohnuma, *Trends Biochem. Sci.* **1999**, *24*, 445–451.

[29] C. D. Poulter, *Phytochem. Rev.* **2006**, *5*, 17–26.

[30] Y. Kharel, T. Koyama, *Nat. Prod. Rep.* **2003**, *20*, 111–118.

[31] I. N. Shindyalov, P. E. Bourne, *Protein Eng.* **1998**, *11*, 739–747.

[32] P. Rabe, J. S. Dickschat, *Angew. Chem.* **2013**, *125*, 1855–1857.

[33] D. W. Christianson, *Chem. Rev.* **2006**, *106*, 3412–3442.

[34] S. G. Burton, *Trends Biotechnol.* **2003**, *21*, 543–549.

[35] R. Bernhardt, *J. Biotechnol.* **2006**, *124*, 128–145.

[36] V. B. Urlacher, in *Handb. Green Chem.*, Wiley-VCH Verlag GmbH & Co. KGaA, **2010**.

[37] V. B. Urlacher, M. Girhard, *Trends Biotechnol.* **2012**, *30*, 26–36.

[38] S. G. Bell, N. Hoskins, F. Xu, D. Caprotti, Z. Rao, L.-L. Wong, *Biochem. Biophys. Res. Commun.* **2006**, *342*, 191–196.

[39] D. C. Hyatt, B. Youn, Y. Zhao, B. Santhamma, R. M. Coates, R. B. Croteau, C. Kang, *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 5360–5365.

[40] D. A. Whittington, M. L. Wise, M. Urbansky, R. M. Coates, R. B. Croteau, D. W. Christianson, *Proc. Natl. Acad. Sci.* **2002**, *99*, 15375–15380.
[41] M. Chen, Al-lami Naeemah, M. Janvier, E. L. D'Antonio, J. A. Faraldos, D. E. Cane, R. K. Allemann, D. W. Christianson, *Biochemistry (Mosc.)* **2013**, *52*, 5441–5453.
[42] H. Meerwein, *Justus Liebigs Ann. Chem.* **1914**, *405*, 129–175.
[43] A. C. Cope, E. M. Hardy, *J. Am. Chem. Soc.* **1940**, *62*, 441–444.
[44] D. E. Cane, M. Tandon, *J. Am. Chem. Soc.* **1995**, *117*, 5602–5603.
[45] Q. Zhang, K. Tiefenbacher, *Nat. Chem.* **2015**, *7*, 197–202.
[46] G. Wagner, K. Slawinski, *Berichte Dtsch. Chem. Ges.* **1899**, *32*, 2064–2083.
[47] C. A. Lesburg, G. Zhai, D. E. Cane, D. W. Christianson, *Science* **1997**, *277*, 1820–1824.
[48] E. Oldfield, F.-Y. Lin, *Angew. Chem. Int. Ed.* **2012**, *51*, 1124–1137.
[49] J. Chappell, *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **1995**, *46*, 521–547.
[50] M. J. Rynkiewicz, D. E. Cane, D. W. Christianson, *Proc. Natl. Acad. Sci.* **2001**, *98*, 13543–13548.
[51] R. Janke, C. Görner, M. Hirte, T. Brück, B. Loll, *Acta Crystallogr. Sect. D* **2014**, *70*, 1528–1537.
[52] M. Koksal, Y. Jin, R. M. Coates, R. Croteau, D. W. Christianson, *Nature* **2010**, *469*, 116–120.
[53] M. Koksal, K. Potter, R. J. Peters, D. W. Christianson, *Biochim.Biophys.Acta* **2013**, *1840*, 184–190.
[54] R. Li, W. K. Chou, J. A. Himmelberger, K. M. Litwin, G. G. Harris, D. E. Cane, D. W. Christianson, *Biochemistry (Mosc.)* **2013**, *53*, 1155–1168.
[55] M. Koksal, I. Zimmer, J. P. Schnitzler, D. W. Christianson, *J.Mol.Biol.* **2010**, *402*, 363–373.
[56] M. Z. Li, S. J. Elledge, *Nat. Methods* **2007**, *4*, 251–256.
[57] U. K. Laemmli, *Nature* **1970**, *227*, 680–685.
[58] E. Gasteiger, C. Hoogland, A. Gattiker, M. R. Wilkins, R. D. Appel, A. Bairoch, others, in *Proteomics Protoc. Handb.*, Springer, **2005**, pp. 571–607.
[59] S. Boivin, S. Kozak, R. Meijers, *Protein Expr. Purif.* **2013**, *91*, 192–206.
[60] K. Ahrer, A. Buchacher, G. Iberer, D. Josic, A. Jungbauer, *J. Chromatogr. A* **2003**, *1009*, 89–96.
[61] A. D. Hanlon, M. I. Larkin, R. M. Reddick, *Biophys. J.* **2010**, *98*, 297–304.
[62] E. Garman, J. W. Murray, *Acta Crystallogr. D Biol. Crystallogr.* **2003**, *59*, 1903–1913.
[63] E. A. Stura, I. A. Wilson, *Methods* **1990**, *1*, 38–49.
[64] P. D. Shaw Stewart, S. A. Kolek, R. A. Briggs, N. E. Chayen, P. F. Baldock, *Cryst. Growth Des.* **2011**, *11*, 3432–3441.
[65] J. Söding, *Bioinformatics* **2005**, *21*, 951–960.
[66] M. Clamp, J. Cuff, S. M. Searle, G. J. Barton, *Bioinformatics* **2004**, *20*, 426–427.
[67] T. R. Butt, S. C. Edavettal, J. P. Hall, M. R. Mattern, *Protein Expr. Purif.* **2005**, *43*, 1–9.
[68] H. Schägger, G. Von Jagow, *Anal. Biochem.* **1987**, *166*, 368–379.
[69] J. P. Abrahams, A. G. W. Leslie, *Acta Crystallogr. D Biol. Crystallogr.* **1996**, *52*, 30–42.
[70] W. Kabsch, *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66*, 125–132.
[71] A. Vagin, A. Teplyakov, *J. Appl. Crystallogr.* **1997**, *30*, 1022–1025.
[72] G. M. Sheldrick, *Acta Crystallogr. A* **2008**, *64*, 112–122.
[73] E. de La Fortelle, G. Bricogne, *Methods Enzymol.* **1997**, *276*, 472–494.
[74] C. Giacovazzo, D. Siliqi, *Acta Crystallogr. A* **1997**, *53*, 789–798.
[75] D. Turk, *Acta Crystallogr. D Biol. Crystallogr.* **2013**, *69*, 1342–1357.
[76] G. N. Murshudov, P. Skubák, A. A. Lebedev, N. S. Pannu, R. A. Steiner, R. A. Nicholls, M. D. Winn, F. Long, A. A. Vagin, *Acta Crystallogr. D Biol. Crystallogr.* **2011**, *67*, 355–367.
[77] G. Langer, S. X. Cohen, V. S. Lamzin, A. Perrakis, *Nat. Protoc.* **2008**, *3*, 1171–1179.
[78] R. A. Laskowski, M. W. MacArthur, D. S. Moss, J. M. Thornton, *J. Appl. Crystallogr.* **1993**, *26*, 283–291.
[79] F. Sievers, A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Söding, et al., *Mol. Syst. Biol.* **2011**, *7*, 539.
[80] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, *J. Mol. Biol.* **1990**, *215*, 403–410.
[81] U. Consortium, others, *Nucleic Acids Res.* **2008**, *36*, D190–D195.
[82] R. B. Woodward, R. Hoffmann, *Angew. Chem. Int. Ed. Engl.* **1969**, *8*, 781–853.
[83] J. A. Aaron, X. Lin, D. E. Cane, D. W. Christianson, *Biochemistry (Mosc.)* **2009**, *49*, 1787–1797.

[84] L. S. Vedula, D. E. Cane, D. W. Christianson, *Biochemistry (Mosc.)* **2005**, *44*, 12719–12727.
[85] Y. Yoshikuni, T. E. Ferrin, J. D. Keasling, *Nature* **2006**, *440*, 1078–1082.
[86] A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, R. J. Read, *J. Appl. Crystallogr.* **2007**, *40*, 658–674.
[87] Y. Liu, B. Schmidt, D. L. Maskell, *Bioinformatics* **2010**, *26*, 1958–1964.
[88] W. Markownikoff, *Justus Liebigs Ann. Chem.* **1870**, *153*, 228–259.
[89] J. E. Leffler, *Science* **1953**, *117*, 340–341.
[90] G. S. Hammond, *J. Am. Chem. Soc.* **1955**, *77*, 334–338.
[91] C. Nakano, S. Horinouchi, Y. Ohnishi, *J. Biol. Chem.* **2011**, *286*, 27980–27987.
[92] M. C. Y. Chang, R. A. Eachus, W. Trieu, D.-K. Ro, J. D. Keasling, *Nat. Chem. Biol.* **2007**, *3*, 274–277.

7. Appendix

7.1. Selinadiene Synthase

DNA- Sequence

ATGGAGCCCGAGCTGACCGTTCCGCCGCTCTTCTCTCCGATCCGGCAGGCGATCCATCCGAAACAT
GCCGACATCGACGTCCAGACAGCGGCCTGGGCGGAAACGTTCAGGATCGGATCCGAGGAACTGCG
CGGCAAACTCGTCACCCAGGACATCGGCACGTTCTCCGCACGGATCCTCCCGGAGGGCCGTGAAGA
GGTCGTGTCGCTGCTCGCGGACTTCATCCTCTGGCTGTTCGGCGTCGACGACGGCCACTGCGAAGA
GGGTGAGCTCGGCCACCGGCCGGGCGATCTGGCCGGGCTCCTGCACCGCCTGATACGCGTGGCGCA
GAACCCCGAGGCCCCGATGATGCAGGACGATCCCCTGGCGGCGGGCCTGCGGGACCTGCGTATGC
GGGTGGACCGCTTCGGCACGGCCGGCCAGACGGCCCGGTGGGTCGACGCCCTGCGTGAGTACTTCT
TCTCCGTCGTGTGGGAGGCCGCGCACCGGCGTGCGGGCACGGTCCCCGACCTCAACGACTACACCC
TGATGCGCCTCTACGACGGCGCGACCTCTGTGGTCCTCCCGATGCTGGAGATGGGCCACGGCTACG
AACTCCAGCCCTACGAGAGGGACCGGACCGCGGTACGGGCCGTGGCCGAGATGGCGTCGTTCATC
ATCACCTGGGACAACGACATCTTCTCGTACCACAAGGAGCGCAGGGGTTCCGGCTACTACCTCAAC
GCCCTGCGCGTGCTCGAGCAGGAACGCGGTCTGACCCCCGCTCAGGCGCTCGACGCGGCGATCTCG
CAGCGGGACCGGGTGATGTGCCTGTTCACGACCGTGAGCGAACAGCTCGCCGAACAGGGCAGCCC
CCAGCTGCGGCAGTACCTCCACAGCCTGCGGTGCTTCATCCGCGGCGCCCAGGACTGGGGCATCAG
CTCGGTCCGCTACACGACGCCGGACGACCCGGCGAACATGCCGTCGGTGTTCACCGACGTCCCGAC
CGACGACAGTACAGAGCCGCTGGACATCCCCGCGGTCTCCTGGTGGTGGGATCTCCTCGCCGAGGA
CGCGCGCTCCGTCCGCAGGCAGGTGCCGGCCCAGCGTTCCGCGTAA

Amino acid- Sequence

MEPELTVPPLFSPIRQAIHPKHADIDVQTAAWAETFRIGSEELRGKLVTQDIGTFSARILPEGREEVVSLL
ADFILWLFGVDDGHCEEGELGHRPGDLAGLLHRLIRVAQNPEAPMMQDDPLAAGLRDLRMRVDRFGT
AGQTARWVDALREYFFSVVWEAAHRRAGTVPDLNDYTLMRLYDGATSVVLPMLEMGHGYELQPYER
DRTAVRAVAEMASFIITWDNDIFSYHKERRGSGYYLNALRVLEQERGLTPAQALDAAISQRDRVMCLF
TTVSEQLAEQGSPQLRQYLHSLRCFIRGAQDWGISSVRYTTPDDPANMPSVFTDVPTDDSTEPLDIPAVS
WWWDLLAEDARSVRRQVPAQRSA

7.2. Hedycaryol Synthase

DNA- Sequence

ATGGCCGAGTTCGAGATACCGGACTTCTACGTCCCCTTCCCCCTGGAGTGCAATCCGCACCTGGAG
GAGGCGTCCCGGGCGATGTGGGAGTGGATCGACGCAAACGGCCTCGCGCCCACAGAACGGGCACG
CGACAGGATGCGGCGCACGGGAGCCGACCTCTCGGGGGCGTATGTGTGGCCCCGCGCCGACCTCG
ACACACTGACGATCGGTCTGAAATGGATCGCGCTGACCTTCCGGATCGACGACCAGATCGACGAG
GACGACACCGCGGAGCGGCTGCCGGCCCGGATGACAGCCATCGACGAGCTGCGCGGCACCCTGCA
CGGACTCCCGGTCTCCGGGCGGTCACCGACCGCCCGGGCCCTGGGCGCCCTGTGGCAGGAGACCGC
CCTCGGACGGCCCGCTACCTGGTGCGATGCCTTCATTGGGCACTTCGAGGCGTTCCTCCAGACCTAC
ACAACCGAGGCCGGCCTCAACGCCCACGGCGCCGGACTCCGCCTCGACGACTACCTCGACCGCAG
GATGTACTCGGTCGGCATGCCCTGGCTTTGGGACCTCGACGAACTGCGCCTTCCGATCTTCCTGCCC
GGCTCCGTACGAACCTGCGGCCCGATGAACAAACTGCGCCGGGCCGGCGCGCTGCACATCGCGTTG
GTGAACGACGTCTTCTCCGTCGAACGGGAGACCCTCGTCGGGTACCAGCACAACGCGGTCACCATC
ATCCGAGAAGCACAGGGCTGCTCGCTGCAGGAAGCGGTGGACCAAGTGGCGGTCCTCGTCGAAGC
CCAGCTCCACACGGTGCTGCAAGCCCGGCAGGAACTCCTCGAAGAACTCGACAGGCAAGCCCTGC
CGTCACGGGCTCGCGAGGCCGCAGTCGACTACGCGGCCAACGTCGCCGCCAACCTGAGCGGGCAG
CTCGTTTGGCACTCCTCGGTCGAACGGTATGCCGTCGACGACCTCCAGTCCGCGGCCGATCCACGG
GCTACCCCGACGACCTCCTCTCTGGGAATACTCGAGCACCACCACCACCACCACTGA

Amino acid- Sequence

MAEFEIPDFYVPFPLECNPHLEEASRAMWEWIDANGLAPTERARDRMRRTGADLSGAYVWPRADLDT
LTIGLKWIALTFRIDDQIDEDDTAERLPARMTAIDELRGTLHGLPVSGRSPTARALGALWQETALGRPAT
WCDAFIGHFEAFLQTYTTEAGLNAHGAGLRLDDYLDRRMYSVGMPWLWDLDELRLPIFLPGSVRTCG
PMNKLRRAGALHIALVNDVFSVERETLVGYQHNAVTIIREAQGCSLQEAVDQVAVLVEAQLHTVLQAR
QELLEELDRQALPSRAREAAVDYAANVAANLSGQLVWHSSVERYAVDDLQSAADPRATPTTSSLGI

8. Publications

Work of the thesis at hand has been conducted from November 2011 until Mai 2015 under the supervision of Prof. Dr. Michael Groll, Chair of Biochemistry, Technical University of Munich (TUM).

Parts of the thesis have been published:

I. **Induced-fit mechanism in Class I Terpene Cyclases**, Baer et al., **2014**, *Angewandte Chemie*, DOI: 10.1002/anie.201403648

II. **Hedycaryol Synthase in Complex with Nerolidol Reveals Terpene Cyclase Mechanism**, Baer et al., **2014**, *ChemBioChem*, DOI: 10.1002/cbic.201300708

Acknowledgement

First of all, I want to thank my supervisor Prof. Dr. Michael Groll for his constant support and his open mind for new ideas, which enabled me to realize my research projects. All the subsequent professional steps in my future career wouldn't be possible without that liberty in research I had at the Chair of Biochemistry.

Next, I want to thank Astrid König for her help and for keeping up the good spirits.

Moreover, I want to thank my students Katrin Fischer, Dirk Hoffmann, Dominik Renn, Nicola Tabertshofer and Iana Gadjalova for their help of realizing my projects.

Furthermore, I want to thank Patrick Rabe and Prof. Dr. Jerome Dickschat (Universität Bonn) for their contributions to our successful collaboration.

Many thanks go to Dr. Annika Frank for proofreading the thesis.

Special credit goes to my colleagues/friends Dr. Ferdinand Alte, Dr. Philipp Beck, Christian Dubiella, Florian Praetorius and Haissi Cui who were always at hand for solving problems and drinking beer.

Last of all, I want to thank all my colleagues from the Chair of Biochemistry for the good times I had during the last three and a half years.

A special thank goes to my family which always supported me, not only during my PhD thesis but all the former years, too.

Finally, I want to thank Carina for all her help and always having an open ear for me, even for my most ridiculous interpretation of data and project planning.

Declaration

I, Philipp Baer, hereby declare that I independently prepared the present thesis, using only the references and resources stated. This work has not been submitted to any examination board yet. Parts of this work have been published in scientific journals.

_____

Philipp Baer, Munich, April 2015