Squalene-Hopene Cyclase Catalyzed Isomerization of Monoterpenes

Squalene-Hopene Zyklase katalysierte Isomerisierung von Monoterpenen

Von der Fakultät 4: Energie-, Verfahrens- und Biotechnik der Universität Stuttgart zur Erlangung der Würde eines Doktors der Naturwissenschaften (Dr. rer. nat.) genehmigte Abhandlung.

Vorgelegt von

Dipl. biochem. Svenja Diether, geb. Kaspari

aus Schwabach

Hauptberichter:	Prof. Dr. Bernhard Hauer
Mitberichter:	Prof. Dr. Karl-H. Engesser
Vorsitzender:	Prof. Dr. Stephan Nussberger
Tag der mündlichen Prüfung:	29.10.2020

Institut für Biochemie und Technische Biochemie der Universität Stuttgart

Abteilung für Technische Biochemie

2020

Opening statement

The presented work was developed at the suggestion and under supervision of Prof. Dr. Bernhard Hauer from January 2016 to December 2018 at the Institute of Biochemistry and Technical Biochemistry, Department of Technical Biochemistry at the University of Stuttgart.

Declaration of authorship/ Eigenständigkeitserklärung

I hereby certify that the dissertation entitled "Squalene-Hopene Cyclase Catalyzed Isomerization of Monoterpenes" is entirely my own work except where otherwise indicated. Passages and ideas from other sources have been clearly indicated.

I hereby declare that this submission is entirely of my own and to the best of my knowledge has not been submitted, either in part or whole, for a degree at this or any other educational institution except where due acknowledgement is made in this work.

Hiermit erkläre ich, dass ich die Dissertation mit dem Titel „Squalene-Hopene Zyklase katalysierte Isomerisierung von Monoterpenen" selbstständig verfasst und keine anderen als die angegebenen Hilfsmittel genutzt habe. Alle wörtlich oder inhaltlich übernommenen Stellen aus fremden Quellen sind als solche kenntlich gemacht.

Ich versichere außerdem, dass ich die beigefügte Dissertation nur in diesem und keinem anderen Promotionsverfahren eingereicht habe und, dass diesem Promotionsverfahren keine endgültig gescheiterten Promotionsverfahren vorausgegangen sind.

Name: Svenja Diether

Date/Datum: _____

Signature/ Unterschrift: _____

Danksagung

Mein größter Dank gilt Herrn Prof. Hauer dafür, dass er von Anfang an an mich geglaubt und in mich vertraut hat. Dies hat sich nicht nur in der Aufnahme in seinen Arbeitskreis ausgedrückt, sondern auch in der Überlassung eines sehr spannenden, aber auch herausfordernden Themas. Dabei hat er mir stets den Rücken gestärkt, aber mir auch große kreative Freiheiten im Bezug auf mein Forschungsthema gelassen und wie ich letztlich zur Lösung der gegebenen Aufgaben gelange. Die gestellten Herausforderungen haben mich fachlich und persönlich wachsen lassen und dafür bin ich sehr dankbar.

Des Weiteren möchte ich Herrn Prof. Engesser für die freundliche Übernahme des Nebengutachtens und spannenden Diskussionen und Herrn Prof. Nussberger für die Übernahme des Prüfungsvorsitzes danken.

Dr. Bettina Nestl und Dr. Bernd Nebel möchte ich dafür danken, dass die beiden weit über ihre eigentlichen Aufgaben hinaus immer für die Mitglieder des Instituts da sind und so für eine wertvolle Atmosphäre und den guten Zusammenhalt in der Arbeitsgruppe sorgen und sie fördern. Einen besonderen Dank möchte ich dabei Bettina für die anregenden Diskussionen und für ihr kritisches und ehrliches Feedback aussprechen. Trotz eines übervollen Tages findet sie immer Zeit, ein offenes Ohr und hilfreiche und aufbauende Worte. Vielen Dank dafür!

Bernd möchte ich besonders danken für die fortwährende Begleitung in der Zeit als „meine“ GC/MS und ich zusammengewachsen sind. Ich habe von Bernd sehr viel über die Funktionsweise und die Wartung eines Gaschromatographen und Massenspektrometers gelernt. Dabei hat auch er bereits zu Anfang großes Vertrauen in mich gesetzt und mir oft mehr zugetraut als ich mir selbst und mir so die Chance gegeben an der Aufgabe zu wachsen. Dafür und für die fortwährende -oft sehr spontane- Unterstützung bin ich ihm sehr dankbar.

Dr. Stephan Hammer danke ich für die Mitbegründung dieses spanenden Themas, die mir die Erkundung eines sehr breiten und spannenden Feldes ermöglicht hat. Nach seiner Rückkehr an das Institut hatten wir einen regen Austausch und ich danke ihm für die spannenden Diskussionen und wertvollen Impulse in dieser Zeit.

Ich möchte diese Gelegenheit nutzen und allen aktuellen und ehemaligen Mitgliedern des Instituts, die ich kennen lernen durfte, für die tolle Atmosphäre danken. Für jegliches Thema, sei es fachlicher, gerätetechnischer oder organisatorischer Art hat sich stets ein hilfsbereiter und kompetenter Kollege gefunden. Trotz all der Arbeit kam das Sozialleben nie zu kurz, sei es bei den üppigen und häufigen Kuchentreffen in der Kaffeeecke, gemeinsamem Grillen oder Basteln für und Feiern nach Verteidigungen. Es war immer ein schönes Miteinander und dafür danke ich euch allen von Herzen!

Im Sinne von Nobelpreisträger John Bardeen, der einmal sagte „ The combined results of several people working together is often much more effective than could be that of an individual scientist working alone” möchte ich allen danken die direkt zum Erfolg dieser Arbeit beigetragen haben. Angefangen mit Dr. Anja Manthey, geb. Eichler die mir in der Anfangsphase des Projektes hilfreich zur Seite stand und für die gute und konstruktive

Zusammenarbeit in dieser Zeit. Des Weiteren möchte ich meinen Studenten Jan Müller, Melanie Buntrock, Josip Tulumovic und Julius Knerr für die gute Zusammenarbeit, die spannenden Impulse und die tollen Ergebnisse, die wir gemeinsam erzielen konnten, danken. Es hat sehr viel Spaß gemacht und nicht nur sie, sondern auch ich habe wertvolles in unserer Zusammenarbeit gelernt.

Mein besonderer Dank gilt Prof. Sílvia Osuna und Christian Curado von der Universitat de Girona in Spanien. Ich durfte beide als sehr kompetente und hilfsbereite Projektpartner kennen lernen. Ich habe unsere Zusammenarbeit und unseren tollen Austausch sehr genossen, der uns gegenseitig neue Impulse für unsere Forschungsarbeiten gegeben hat.

In diesem Zusammenhang sind auch meine guten Freunde und Kollegen Dr. Sabrina Henche und Dr. Nico Kress besonders hervorzuheben. Sabrina Henche hat parallel mit mir an unserem interessanten Enzym gearbeitet. So konnten wir nicht nur wunderbar Ideen, sondern auch die eine oder andere Variante austauschen. Wir hatten eine wundervolle Zeit mit- und nebeneinander, in der wir eine gute Balance zwischen privatem Austausch, gegenseitiger Unterstützung und produktivem Arbeiten gefunden haben.

Nico ist während meines verletzungsbedingten Ausfalls nicht nur eingesprungen, mir beim Projekt zu helfen, sondern hat auch wertvolle Impulse für das Projekt gegeben. Die Zusammenarbeit mit ihm war geprägt von einem spannenden und begeisterten Austausch von Ideen und der Diskussion von Ergebnissen. Auch nach seinem Weggang vom Institut hat er keine Mühen gescheut, mich in dieser Arbeit weiter durch die kritische Durchsicht dieses Manuskripts zu unterstützen.

Gemeinsam mit Jens Schmid waren die beiden Teil der „Fantastic Four“. Aus diesem Gespann sind tiefe Freundschaften geworden und ohne sie wäre die Zeit am ITB nur halb so schön gewesen!

Ein großer Dank gilt auch meiner Familie für ihre fortwährende Unterstützung!

„Es lohnt sich, die Entdeckungen anderer zu studieren,
da für uns selbst eine neue Quelle für Ideen entspringt.“
(Gottfried W. Leibniz)

Summary

In nature the squalene-hopene cyclase from *Alicyclobacillus acidocaldarius* (*Aac*SHC), a class II triterpene cyclase, catalyzes the cyclization of squalene to hopene and hopanol. In addition, it accepts a broad range of non-natural substrates with different chain lengths and functional groups. This makes it an ideal starting point for the investigation of new biocatalysts for Brønsted-acid-catalyzed reactions.

Preliminary studies revealed that *Aac*SHC is able to convert different pinenes with (+)-ß-pinene showing the highest conversion. A screen of an already existing library of *Aac*SHC variants in the first shell of the active site showed the possibility of shifting product distributions. The goal of this thesis was therefore to investigate the reaction in detail, establish a robust analytical system to confirm initial hits and then create and identify improved variants of *Aac*SHC with higher activities and product selectivities towards the main products (+)-α-pinene, ß-camphene and (*R*)-limonene. For this, different approaches were chosen.

The first approach comprised a semi-rational library design in the second shell of the enzymes active site. Another approach was to transfer key residues of identified intrinsically more selective and active SHCs enzymes from other organisms into the *Aac*SHC. Moreover, positions of the so far most selective variants for each of the products were investigated by saturation mutagenesis. Promising variants concerning selectivity and activity of these approaches were further combined.

As a result, it was confirmed by using a robust analytical setup that *Aac*SHC converts (+)-ß-pinene (10% conversion in 40 h) into the main products (+)-α-pinene (39% product selectivity), ß-camphene (25%) and (*R)*-limonene (20%). The highest selectivity towards (+)-α-pinene was achieved with the variant Y609W_Q366E which shows a product selectivity of 84% and a 1.8-fold higher activity compared to the wild type. The variant L36A was already present in the in-house library and was confirmed to show the highest product selectivity in the formation of (*R*)-limonene with 67% and a 1.3-fold higher activity than the wild type. The double variant L36A_Y420F was created and identified to show a 1.6-fold higher activity than the wild type, however, at the expense of the product selectivity (54% towards (*R*)-limonene). The most selective variant in the formation of ß-camphene (53%) was created and identified as D422P with a 1.4-fold higher activity than the wild type. The activity was improved with the creation of variant D442N_Y420F (1.7-fold). However, this was accompanied by a minor drop in product selectivity towards ß-camphene (47%) compared to the variant D422P (53%).

In additional studies it was shown that the acidic isomerization of monoterpenes by *Aac*SHC is applicable to other substrates with strained rings, as well. Here, sabinene was accepted as substrate by the *Aac*SHC resulting in the formation of terpinen-4-ol, α-thujene, γ-terpinene, sabinene hydrate or *cis*-ß-terpineol and others. Variants with higher selectivities towards some of the products were identified.

Taken together, an increase of selectivity and activity in the isomerization of (+)-ß-pinene with *Aac*SHC towards different products was achieved and a robust analytical system

established. It was further shown that this principle can be applied to sabinene as substrate.

Zusammenfassung

Die Squalen-Hopen Zyklase aus *Alicyclobacillus acidocaldarius* (*Aac*SHC) ist eine Klasse II Triterpenzyklase, die in der Natur die Zyklisierung von Squalen zu Hopen und Hopanol katalysiert. Zusätzlich akzeptiert sie eine große Bandbreite an nicht-natürlichen Substraten mit unterschiedlichen Kettenlängen oder funktionellen Gruppen. Dadurch ist sie ein idealer Startpunkt für die Untersuchung neuer Biokatalysatoren für Brønstedsäure-katalysierte Reaktionen.

Vorarbeiten konnten zeigen, dass *Aac*SHC in der Lage ist verschiedene Pinene umzusetzen, wobei (+)-ß-Pinen davon den höchsten Umsatz zeigte. Eine Sichtung einer bereits existierenden Bibliothek verschiedener *Aac*SHC Varianten in der ersten Schale des aktiven Zentrums des Enzyms zeigte, dass es prinzipiell möglich ist Produktverteilungen zu verschieben. Das Ziel dieser Arbeit war es daher die Reaktion im Detail zu untersuchen, ein robustes analytisches System zu etablieren, um die initialen Funde zu bestätigen und verbesserte *Aac*SHC Varianten mit höheren Aktivitäten und Produktselektivitäten gegenüber den Hauptprodukten (+)-α-Pinen, ß-Camphen und (*R*)-Limonen herzustellen und zu identifizieren. Hierzu wurden verschiedene Herangehensweisen gewählt.

Der erste Ansatz umfasste das semi-rationale Design einer Mutantenbibliothek in der zweiten Schale des aktiven Zentrums des Enzyms. Ein anderer Ansatz war der Transfer von Schlüsselaminosäuren, die in SHC Enzymen aus anderen Organismen identifiziert wurden, da diese bereits intrinsisch höhere Aktivitäten oder Produktselektivitäten als der Wildtyp der *Aac*SHC zeigten. Des Weiteren wurden die Positionen der Varianten gesättigt, die zu diesem Zeitpunkt die höchsten Produktselektivitäten aufwiesen. Vielversprechende Varianten bezüglich Selektivität und Aktivität wurden in einem weiteren Ansatz kombiniert.

Als Ergebnis konnte, unter Benutzung eines robusten analytischen Aufbaus, bestätigt werden, dass *Aac*SHC (+)-ß-Pinen (10% Umsatz in 40 h) in die Hauptprodukte (+)-α-Pinen (39% Produktselektivität), ß-Camphen (25%) und (*R)*-Limonen (20%) umsetzt. Die höchste Selektivität bezüglich (+)-α-Pinen konnte mit der Variante Y609W_Q366E erreicht werden, welche eine Produktselektivität von 84% und eine 1,8-fach höhere Aktivität als der Wildtyp aufwies. Die Variante L36A, welche bereits in der institutseigenen Mutantenbibliothek vorlag, konnte darin bestätigt werden, dass sie die höchste Selektivität gegenüber der Bildung von (*R)*-Limonen aufweist mit 67% Produktselektivität und einer 1,3-fach höheren Aktivität als der Wildtyp. Die Doppelvariante L36A_Y420F wurde neu geschaffen und zeigte eine 1,6-fach höhere Aktivität als der Wildtyp, allerdings auf Kosten der Produktselektivität bezüglich (*R)*-Limonen, die auf 54% sank. Die Varianten mit der höchsten Produktselektivität gegenüber ß-Camphen (54%) wurde erschaffen und identifiziert als D442P, die eine 1,4-fach höhere Aktivität gegenüber dem Wildtyp aufweist. Die Aktivität konnte durch die Schaffung der Doppelvariante D442N_Y420F noch erhöht werden (1,7-fach). Dies resultierte jedoch in einem geringen Absinken der Produktselektivität für ß-Camphen (47%) im Vergleich zur Variante D442P (53%).

In einer zusätzlichen Studie konnte gezeigt werden, dass sich die sauer induzierte Isomerisierung von Monoterpenen durch die *Aac*SHC auch auf weitere Substrate mit gespannten Ringen anwenden lässt. Hierbei wurde gezeigt, dass Sabinen ebenfalls als Substrat von *Aac*SHC akzeptiert wird, was unter Anderem in der Bildung von Terpinen-4-ol, α-Thujen, γ-Terpinen, Sabinenhydrat oder *cis*-ß-Terpineol resultiert. Hier konnten Varianten mit höheren Selektivitäten bezüglich einiger der Produkte identifiziert werden.

Zusammengefasst wurde eine Erhöhung der Produktselektivität und der Aktivität in der Isomerisierung von (+)-ß-Pinen durch die *Aac*SHC gegenüber verschiedenen Produkten erreicht und ein robuster analytischer Aufbau etabliert. Es konnte weiter gezeigt werden, dass dieses Prinzip auch auf andere Substrate, wie Sabinen, zutrifft.

Table of Contents

1 Introduction

1.1 Terpenoids

Terpenes and terpenoids represent the most diverse family of natural products, concerning structure and stereochemistry. In 2017 over 80 000 different terpenes and terpenoids were known[1–3]. Terpenes consist of a different number of isoprene units ranging from 5 to over 45 carbon atoms[4]. The great diversity is accomplished by different combinations of elongation reactions, cyclization reactions, branching and addition of functional groups, e.g. alcohols, ketones, or classes of biomolecules such as fatty acids or sugars. This makes terpenes and terpenoids a good example for combinatorial chemistry in nature[1–5].

1.1.1 Terpene biosynthesis

All terpenes originate from isopentenyl diphosphate (IPP, **1**) and dimethylallyl diphosphate (DMAPP, **2**, Figure 1). IPP can be generated by the mevalonate pathway (some prokaryotes, eukaryotes: in the cytosol and peroxisomes) or the desoxyxylulose-5-phosphat (DXP) pathway (most prokaryotes, plastides of plants). The mevalonate pathway starts from three acetyl-CoA units and results solely in IPP, which can be isomerized to DMAPP by isopentenyl diphosphate isomerase. The DXP pathway starts from pyruvate and glyceraldehyde-3-phosphate and results in IPP and DMAPP (5:1 ratio)[6,7].

A prenyltransferase adds IPP in a head-to-tail fashion to an allylic diphosphate (mostly DMAPP) to generate terpene precursors of different lengths, namely geranyl diphosphate (C_{10}, **3**, Figure 1), farnesyl diphosphate (C_{15}, **4**) and geranylgeranyl diphosphate (C_{20}, **5**). Two farnesyl diphosphate units can in turn be condensed in a tail-to-tail fashion by a squalene synthase to form squalene (C_{30}, **6**, Figure 1). These linear terpene precursors can be further cyclized for example by a squalene-hopene cyclase (SHC) to hopene and hopanol or after squalene oxidation by oxidosqualene cyclase (OSC) to lanosterol[6,8]. In eukaryotes lanosterol is further converted to cholesterol, which is used to influence membrane fluidity. In prokaryots hopanoids fulfill this function[9,10].

Different intermediates in the terpene biosynthesis are used as precursors for branching terpene classes like monoterpenes or triterpenes. For example geranyl diphosphate (**3**) is the precursor for pinenes (**11**, **12**), limonene (**14**) or (-)-menthol (**10**, Figure 3)[2,6]. The biosynthetic pathway of terpenes and some selected branches are shown in Figure 1.

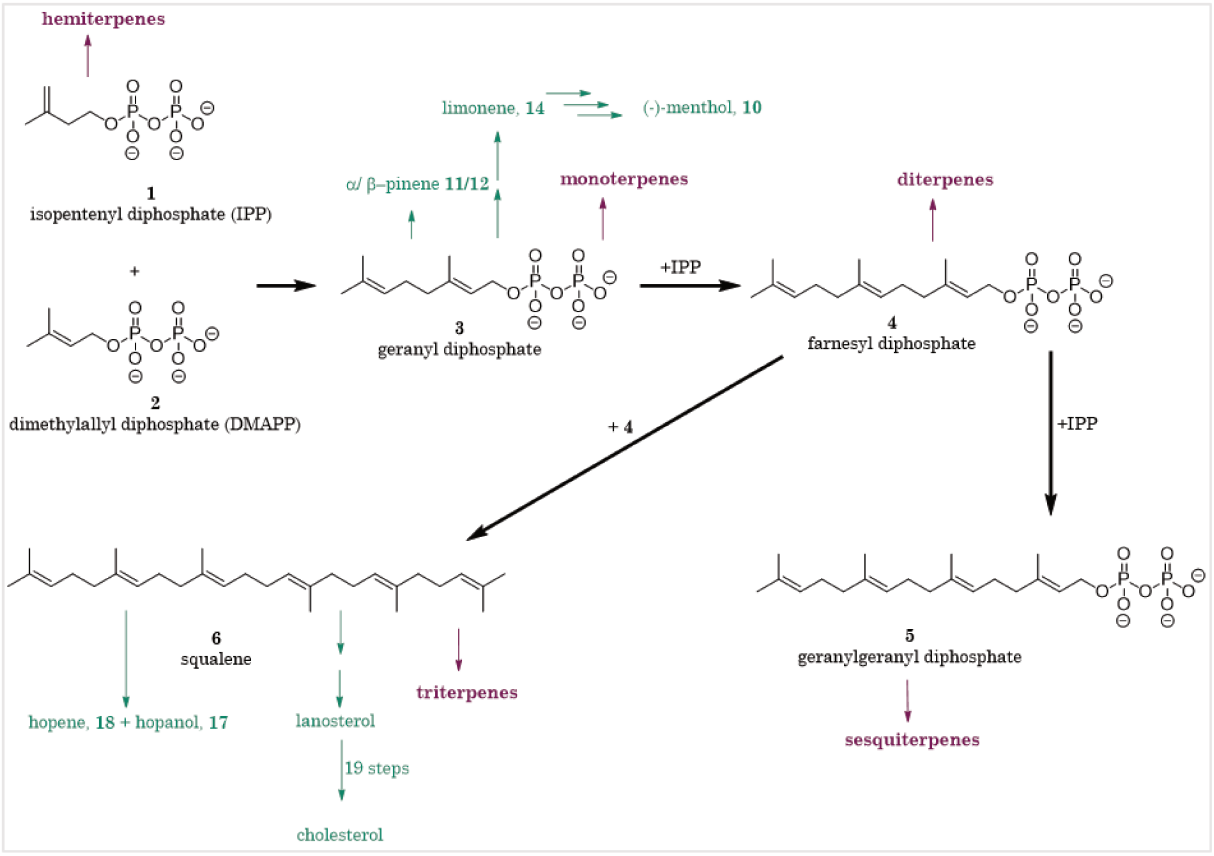


Figure 1: General scheme of terpene biosynthesis and chemical structures of some key players. The main pathway is depicted in bold arrows. The branching terpene groups are shown in purple. Interesting products are colored green. Compiled and modified based on literature[1,2,6,11].

1.1.2 Natural occurrence, utilization and applications

Terpenes are widespread in nature. They have been found in plants, fungi, insects, nematodes, mollusks, some marine organisms, gram-positive and gram-negative bacteria and others[12]. Terpenes and terpenoids serve diverse functions in physiology as hormones, antioxidants, maintaining the membrane structure or in ecology as means of communication to defeat predators or attract allies[13]. Terpenes are prone to act as messengers, since especially the shorter terpenes are of low-molecular weight with a high vapor pressure and can thus deliver information over great distances[4,12,14–16].

Given the different properties of various terpenoids, there are multiple possibilities for applications. Throughout human history they have constantly been used as scents in the form of myrrh and frankincense or for practical purposes such as rubber and pitch to water-proof boats. Most importantly they were used for medicinal reasons and people's diets. They serve as provitamins (ß-carotene), antioxidants (lycopene, astaxanthin), colors and flavors and can be found in every natural food. Especially spices were most valuable for their antimicrobial and medical activities, which are again a result of their rich terpene contents[13,15].

Given the abundance of terpenes and their broad properties and applications, only a few prominent examples will be highlighted in detail in the following. Artemisinin (**7,** Figure

2), a plant sesquiterpene, is now used as an anti-malaria agent. It kills all asexual stages of *Plasmodium falciparum*, which causes the malaria disease by inhibiting the Ca^{2+} ATPase of the sarco-endoplasmatic reticulum[12]. Artemisinin was discovered by Youyou Tu[17], for which she received the Nobel Prize in medicine in 2015. The therapy could reduce mortality caused by malaria by over 20% and is estimated to save 100, 000 lives every year in Africa alone[18]. Another remedy is the anti-tumor drug Taxol containing the diterpene paclitaxel (**8**, Figure 2), which kills tumor cells by binding to tubulin and thereby interfering with the microtubule dynamics and inhibiting proliferation[12]. The triterpene ambroxan (**9**, Figure 2) is a potent fragrance. Naturally it is produced in whales and can be found on coasts or in the sea and whales were hunted for it. Since 1950 it can be synthesized chemically. However, a more elegant way is to synthesize it enzymatically from homofarnesol (**22**e) using the squalene-hopene cyclase[3,19,20].

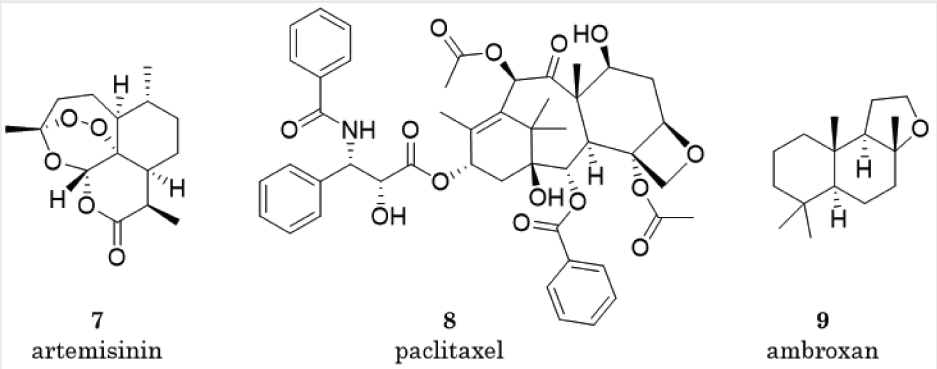


Figure 2: Examples for terpenes with valuable applications. Adapted from literature[3,12,18–20].

1.2 Monoterpenes

Monoterpenes are the smallest members of the terpene family regarding their size. They are built up of only ten carbon atoms which leads to a high lipophilicity, a low molecular-weight and contributes to their volatile nature and high vapor pressure at ambient temperatures[12,16,21].

1.2.1 Natural occurrence, properties and applications

Monoterpenes are solely produced by plants[13] and are mostly extracted in the form of essential oils[22]. They are generally reported to have repellant and toxic features against bacteria, fungi, nematodes, insects, tumors and other threats to human health[23–27]. However, it is difficult to asses certain properties to one single monoterpene since they are mostly applied as plant oils or extracts and therefore contain various mixtures of monoterpenes. Moreover, the contents of individual monoterpenes can vary within these mixtures and some effects are believed to be synergistic and can therefore not always be attributed to a single component[28].

Apart from their medicinal properties, monoterpenes are mainly used as fragrances and flavors. A good example is (-)-menthol (**10,** Figure 3), which is frequently used e.g. in tooth paste, chewing gums, sweets and drugs[29]. As a rare example, a synthetic approach for menthol was found by Ryoji Noyori for which he received the nobel prize in chemistry in 2001[30]. Recently, progress was reported for a chemoenzymatic production of (-)-menthol using a cascade of an engineered ene-reducatase (NADH-dependent 2-cyclohexen-1-one

reductase, NCR) and squalene-hopene cyclase (SHC). It should be highlighted that the starting material is an isomeric mixture of citral and the ene-reductase was modified to selectively produce *R*-citronellal as cascade intermediate. Whereas, the SHC was engineered to solely produce (-)-isopulegol out of four possible diastereomers. The last step is the hydration to (-)-menthol[31–33].

As marginalia it should be noted that monoterpenes are largely emitted into the atmosphere. One report finds up to 2.5 mg/g dry leaf weight limonene emission per hour for plants in California[34]. Another report states α-pinene and carene flux of up to 500 $\frac{ng}{m^2 s}$ in a Finnish forest during the summer[35]. In the atmosphere monoterpenes and sesquiterpenes, representing the most abundant volatile organic compounds, are oxidized to form secondary organic aerosols. In turn, these serve as cloud condensation nuclei thereby influencing the formation of clouds and climate[36,37].

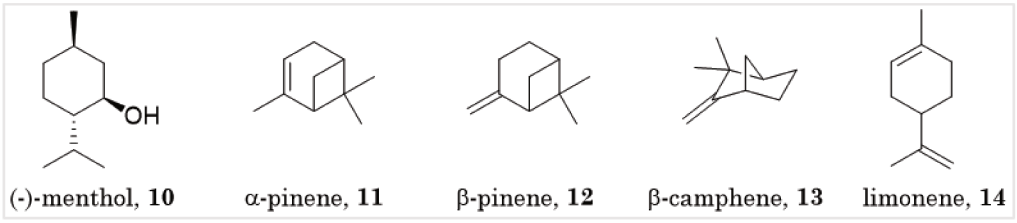


Figure 3: Structural depiction of relevant monoterpenes.

The following monoterpene representatives are of great importance for this work and are therefore reviewed in more detail.

1.2.1.1 α- and β-Pinene

α- and ß-pinene are both colorless liquids that are insoluble in water but soluble in organic solvents and ethanol. The odor and taste of α-pinene's (**11**, Figure 3) is described as citric, spicy, woody and turpentine-like. ß-Pinene (**12,** Figure 3) is reported to additionally have a minty and eucalyptus like aroma[38]. Together they are the dominant odorous compounds emitted by grasses, flowers, shrubs and trees[28]. Especially conifer trees are large producers of pinenes and their essential oils can contain up to 75-90% of pinenes[38]. Another important source is turpentine, a byproduct of the wood pulp industry[39,40]. There are many reports attributing pinenes anti-bacterial[41], anti-fungal[42], anti-tumor[43], anti-inflammatory[44] and anti-oxidative properties[28], just to name a few. Additionally, they are used as paints and varnishes and as starting materials for fine chemicals and other valuable monoterpenes e.g. menthol[38,45]. Moreover, investigations are under conduction to asses pinene dimers as biofuels since they show adequate volumetric energies[39].

1.2.1.2 ß-Camphene

ß-Camphene (**13**, Figure 3) forms colorless crystals at room temperature which emit a pungent smell reminiscent of fir needles. It is mainly found in laurel tree but also in spices like nutmeg[46]. It is used as a fragrance and intermediate for the synthesis of camphor via isobornyl acetate and other pharmaceuticals[40,47,48]. Camphor is an important aroma compound with various medical applications in the treatment of diseases like rheumatism, asthma, bronchitis, kidney weakness, fever, measle and others, making use of its anti-microbial, anti-viral, anti-tussive and cooling properties[46,49,50]. ß-Camphene itself was

shown to lower lipid and cholesterol contents in the blood and inhibit the cholesterol biosynthesis in a different, yet unknown manner than statines[51]. It also exhibited pain relieving and antioxidative effects[52,53].

1.2.1.3 Limonene

Limonene (**14**, Figure 3) is a colorless liquid, which is insoluble in water but miscible with organic solvents. The *R*-enantiomer smells like lemon or orange, whereas the *S*-enantiomer smells pine-turpentine like[54]. It can be found in caraway, dill but largely in oranges, where it makes up to 90% of orange peel oil. *R*-limonene is widely used as a flavor in drinks, sweets and bakery products or as fragrance in cosmetics, soaps and cleaning products. Additionally it is applied as solvent for paints, as monomer from renewable sources in polymerization processes or as cheap starting material for organic synthesis[22,38,55,56]. Medically it is administered for the treatment of some cancer types[13,22].

1.2.2 Chemical isomerization of monoterpenes

Despite some reported successes many of the industrially important monoterpenes are not synthetically accessible and are still isolated from natural sources[29] or have to be interconverted starting from other monoterpenes. However, these processes usually result in mixtures and therefore require a cost-intensive work-up[57]. Thus, there is a high demand for (bio-)synthetic access.

The industrially applied isomerization of α-pinene to ß-camphene uses TiO_2 hydrate and sulfuric acid under 156 °C-162 °C reflux. However, ß-camphene is only produced with a typical selectivity of 35-45% and the catalyst must be activated *in situ*. One of the major side products in this process is tricyclene with 13%-17%, which can react in a similar way as ß-camphene and it is therefore crucial to be removed afterwards[58,59].

Several attempts to increase the selectivity towards ß-camphene were made employing various catalysts, such as zeolithes[60,61], general Brønsted acdis[62], sulphated zirkonia[63], clays and other acidic porous structures[64–67]. Mostly, the selectivities for ß-camphene stayed under ~55%[61,63,67] or where otherwise accompanied by great reduction in conversion[62,64]. A thorough summary of different catalysts can also be found by Mäki-Arvela et al[40].

The main products for the isomerization of α-pinene are for most cases ß-camphene and limonene[61,64,67,68]. Common side products are tricyclene and different monocyclic monoterpenes[62,65]. Even though there are several attempts to rationalize the reactions and to find ways to steer the selectivities e.g. by variation of the catalyst, reaction conditions, acidic strength, activation temperature and time[45,69] selective catalysts are still missing. This highlights the difficulties to control the reaction selectively. One reason for this lies within the intermediate carbocations.

The intermediate carbocations in these reactions pose several challenges to a catalyst. First, they are very similar regarding reactivity and energy levels[58,70]. Second, the cations are not distinguishable by any functional groups and differ only slightly in their conformation.

Third, they are highly reactive and therefore short lived[71]. For example, Pemberton et al[72] calculated an average lifetime of 40 fs for the bornyl cation coming from the pinanyl cation and resulting in the camphyl cation.

Overall, these characteristics make these carbocations very hard to control which is why usually mixtures of different products result[45]. Despite all these challenges, the general reaction is still of interest due to the low cost and high accessibility of starting material and the demand for the resulting higher value products[45].

In nature, specifically tailored enzymes evolved to control these intermediate carbocations in the (bio-)synthesis of monoterpenes. They are presented in the following chapter.

1.2.3 Monoterpene synthases

Monoterpene synthases are Mg-dependent enzymes that catalyze the formation of various acyclic, mono-, bi- and some tri-cylic monoterpene structures out of geranyl diphosphate (**3**, Figure 4)[1]. They use precisely coordinated Mg^{2+} ions to abstract the pyrophosphate group from geranyl diphosphate and generate a highly reactive intermediate carbocation. The reaction is initiated by the abstraction and re-addition of pyrophosphate to form a linalyl diphosphate intermediate. This is necessary for bond-rotation. After re-dissociation of the pyrophosphate and subsequent cation formation it can cyclize to form α-terpinyl cation, the central cation to the monoterpene variety[73,74] (Figure 4).

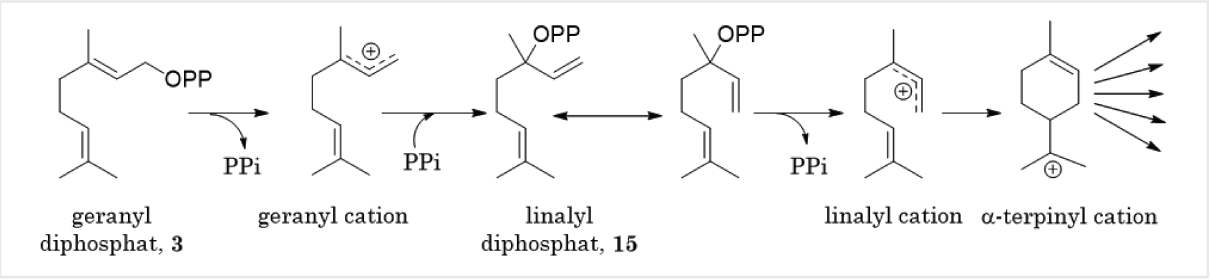


Figure 4: Formation of α-terpinyl cation within monoterpene synthases. Adapted from literature[74].

The α-terpinyl cation is the starting point for various other possible isomeric cations that access a variety of monoterpenes after deprotonation or quenching by water. In Figure 5 only some intermediate cations and their transformational paths are shown. Noteworthy is the re-addition of pyrophosphate to form bornyl phosphate for the production of borneol and camphor[75–77].

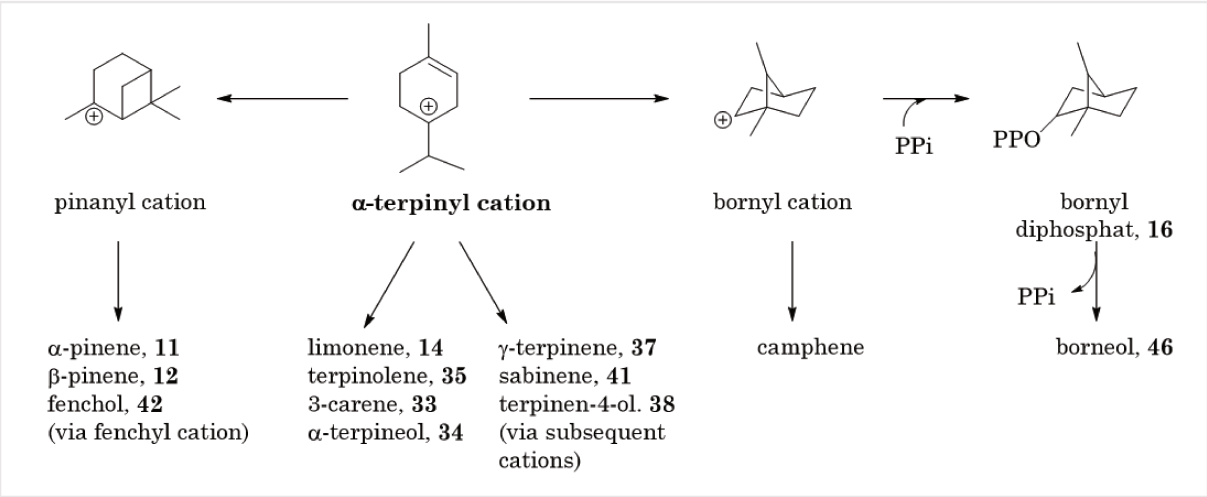


Figure 5: Follow-up reactions for the biosynthesis of monoterpenes and exemplary products out of α-terpinyl cation. Compiled and modified based on literature[74,78].

The vast number of rearrangements and the structural diversity of the resulting monoterpenes is not only a result of the large number of monoterpenes cyclases that have evolved but is mainly due to their ability to form multiple products besides their main product. In fact the production of mixtures is even more common than the formation of a single product[74,78]. However, nature also managed to evolve some very specific monoterpene cyclases[73]. For biotechnological applications it would be desirable to have more single-product enzymes, however engineering monoterpenes synthases towards higher selectivities proves to be challenging due to the little knowledge in correlation between the active site sequence and the product outcome[77].

It seems to be advantageous for plants to produce mixtures of different monoterpenes instead of specific single products. First, when used for communication, a more complex signal could be thought of as more specific. Second, it might be more difficult for predators to develop resistance to mixtures that address several modes of actions for defense. Third, the maintenance of a single enzyme that produces a multitude of products might be more cost-effective for the organism. Fourth, the composition of these mixtures and thus their effects can easily be changed by a single point mutation and therefore give an advantage to individuals within a population[12,79]. This plasticity was also demonstrated *in vitro* when single or only a few amino acid changes in monoterpene synthases shifted the product distribution dramatically[76,80].

1.3 Squalene-hopene cyclase

There are several terpene and terpenoid cyclases for different cyclization reactions of terpenes and terpenoids. Among them are monoterpene (C_{10}) cyclases (see chapter 1.2.3), sesquiterpene (C_{15}) cyclases, diterpene (C_{20}) cyclases, triterpene (C_{30}) cyclases, and many more. The cyclases differ not only in the length of their substrates and consequently in the number of rings formed but also in the activation of their substrates and thus in their catalytic mechanism[1,5].

1.3.1 Classification

The entire terpene cyclase enzyme family is divided into class I and class II, based on the catalytic mechanism and the phosphorylation status of the substrate. Class I is ionization-dependent. It initiates the carbocation formation using Mg^{2+} as Lewis acid and to dissociate diphosphate from the substrate. Additionally, the reaction proceeds in a tail-to head fashion. Although the amino acid sequences differ, a common fold was found for class I terpene cyclases, the common aspartic acid-rich motif DDxxD and sometimes an additional NSE/DTE motif. Both motifs form a cluster with catalytic active Mg^{2+} ions. In total the active site resides in an α-helix bundle that all class I cyclases share[2,81–83]. Examples of class I enzymes are monoterpene synthases (see chapter 1.2.3).

Class II is protonation-dependent and initiates the carbocation formation using an aspartic acid residue as Brønsted acid, which protonates the terminal isoprene unit leading to a head-to-tail cyclization. The catalytic aspartic acid lies within a signature sequence DxDD. Class II terpene cyclases possess a double α-barrel fold, which is topologically not related to the class I fold[2,81–83].

This work focuses on a class II triterpene cyclase, more precisely on the squalene-hopene cyclase (SHC) from *Alicyclobacillus acidocaldarius* (*Aac*SHC) belonging to the triterpene cyclase family[1].

1.3.2 Structure

*Aac*SHC is a dimeric, 71.6 kDa monotopic membrane protein. This means it is inserted into the inner side of the cytoplasmic membrane with helix 8 (yellow, Figure 6), facing towards the cytoplasm[8]. The first crystal structure of *Aac*SHC and the first structure of an SHC in general was solved in 1997 (PDB code: 1SQC)[84] and was published again in 2004 with a higher resolution (PDB code: 1UMP)[85]. Since then *Aac*SHC has become a model for SHC investigations. Domain 1 (green, Figure 6) is composed of an α_6-α_6 barrel, stabilized by QW-motifs, which can be found in all triterpene cyclases and are believed to increase the thermostability of the enzyme during the exergonic (48 kcal/mol) cyclization reaction[8,84,86,87]. Domain 2 (purple, Figure 6) is

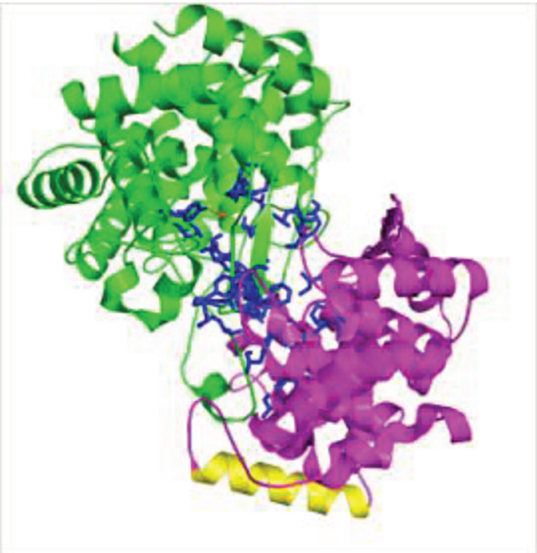


Figure 6: Cartoon structure of *Aac*SHC, domains are labeled in green and purple, the helix inserted into the membrane is shown in yellow. Residues in the active center are in blue. Adapted from 1UMP (PDB)[85] using the software package PyMOL.

composed of an α-α-barrel. The hydrophobic active site has a volume of about 1200 Å^3 and is located in between the two domains[84] (blue, Figure 6). The substrate squalene enters and the product hopene exits the active site via a flexible hydrophobic tunnel towards the membrane, which is believed to widen for hopene release due to the high energy release during the reaction[5,85]. The so-called protonating machinery can be found in a barrel fold together with the catalytic active D376. The catalytic active aspartic acid 376 is surrounded by a specific hydrogen bonding network (green, Figure 7) which facilitates an *anti*-orientation of the hydrogen and therefore makes it a strong Brønsted acid. The network consists of D374, D377, H451, Y495 and a well-defined water which is thought to be involved in the reprotonation via a salt bridge formed by Q454 and R623 connected to the solvent water[84,85,88,89] (not depicted in Figure 7).

Mutations at the protonation machinery inactivate the enzymes catalytic ability or reduce it dramatically[88]. The hydrophobic amino acids lining the inner surface of the active site are anchored in different loops and can be mutated without affecting the protonation ability but eventually altering the substrate specificity and product outcome[33,81,90–92]. Moreover, these loops are believed to be flexible, which would explain why mutations further (14 Å) away from D376 can have an influence on substrate specificity[90].

In Figure 7 some of the known important residues in the active site and their respective functions are shown. The aspartic acid residue at position D376, which protonates the substrates is shown in green. The residues D447 and H451, thought to enhance the acidity, and D374 and D377, which also stabilize the first carbocation intermediate, are highlighted in green. Residues that further stabilize the intermediate cations via cation-π interactions are highlighted in red. For example, F365 and Y420 stabilize the cation at position 10 of the substrate and F601 stabilizes the cation at position C15. The residues in deep blue have been found to help bind the substrate, for example L607. Residues that might either help bind the substrate or stabilize intermediate carbocations are depicted in orange. The residues shown in light blue are responsible for the terminal deprotonation step and I261, which helps with stereocontrol is shown in purple[8,86].

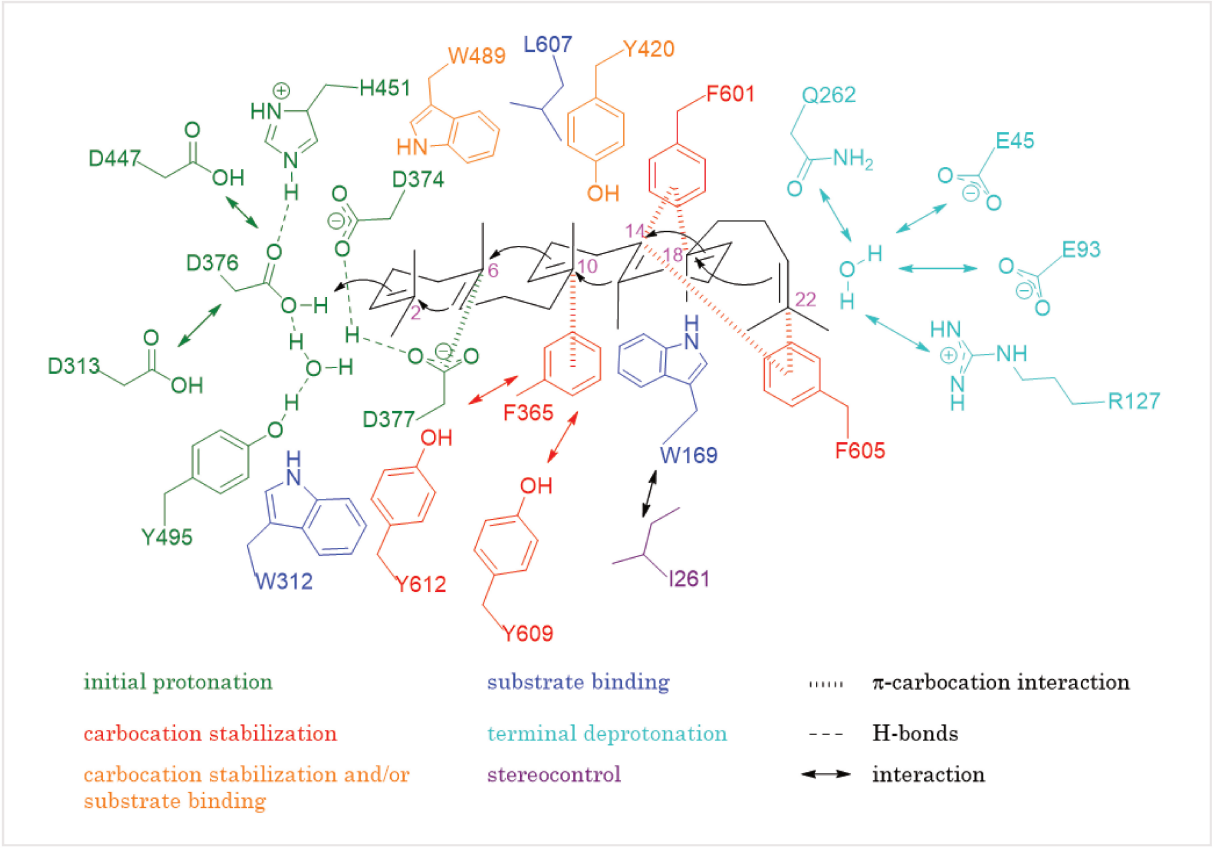


Figure 7: Important residues in the active site of *Aac*SHC, adapted from [86] and updated[8].

1.3.3 Catalytic mechanism

SHCs naturally catalyze the conversion of squalene (C_{30}, **6**) to hopene (**18**) and hopanol (**17**), which is one of the most complex and most demanding single-step reactions in biochemistry. The enzyme establishes five new C-C bonds and nine chiral centers with hopene being one out of 512 possible stereoisomers[1,8]. Although the turnover number of 0.3 s^{-1} is low, the catalytic rate enhancement over the uncatalyzed reaction is enormous[1,84]. The natural reaction is shown schematically in Figure 8.

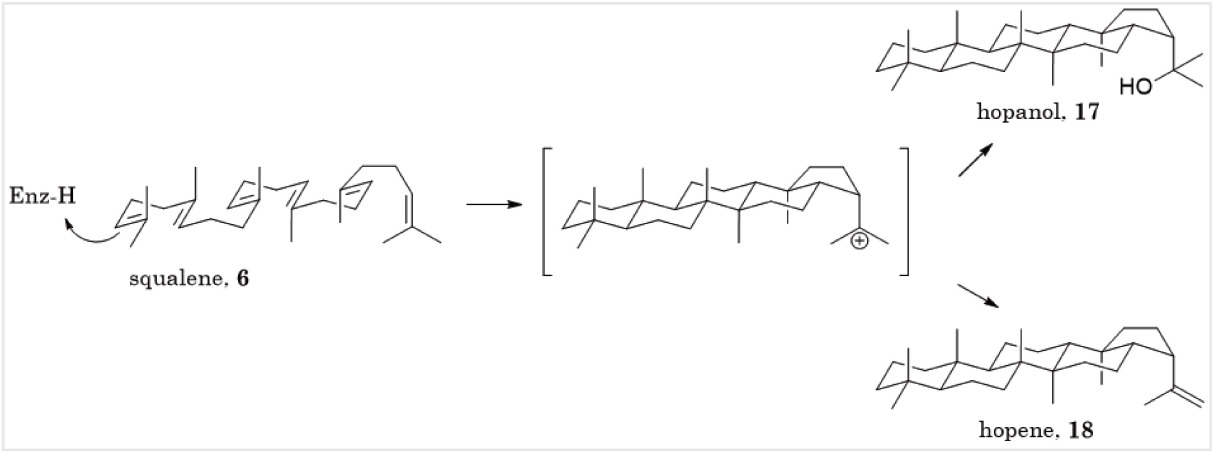


Figure 8: Schematic conversion of squalene to hopene and hopanol, adapted from literature[86].

The cyclization of squalene to either hopene (**18**) or hopanol (**17**) starts with the protonation of the terminal carbon-carbon double bond by an aspartic acid at position D376 (for *Aac*SHC) in the active site. This results in a carbocation, which is electrophilically attacked by an adjacent double bond, resulting in a new carbocation. These steps are repeated until the last carbocation at the opposite side of the molecule is either deprotonated to yield hopene (**18**) or nucleophilic attacked by water to yield hopanol (**17**). The distance between a cation and the attacking double bond is 4 Å, each[85,86].

There are several special features to this mechanism, described in the following by using the example of *Aac*SHC. First, the substrate already adopts an all pre-chair conformation prior to the protonation step, so only small rearrangements are necessary during the cyclization cascade[93].

Second, the proton from the catalytic active aspartic acid residue 376 is faced towards the substrate in *anti*-orientation, which makes this aspartic acid a specific Brønsted acid that is 10^4 times more acidic than the in nature more common *syn*-oriented carboxylic proton[90,94]. This *anti*-orientation is facilitated by a specific hydrogen bonding network consisting of H451, D376, Y495 and a well-oriented water molecule[8]. *Anti*-oriented protons are a very rare concept in enzyme catalysis. However general acid-base catalysis is very common[94].

Third, the active site is hydrophobic and excludes water through water channels upon substrate binding. Thereby it mimics an inert solvent in synthetic organic chemistry and prevents a premature determination through nucleophilic attack to an intermediate carbocation[1,81,91]. In a recent study it appeared as if *Aac*SHC also uses water exiting through these water channels to pay the enthalpic and high entropic cost of prefolding the substrate. This entropy gain seems to be a key aspect of the SHC-catalytic mechanism. This is supported by the fact that variants of different water channels showed a shift in their product spectrum[95].

Fourth, the enzyme functions as a chaperone by stabilizing the intermediate carbocations with π-bonds of aromatic amino acid side-chains[96]. The number of the carbocationic intermediates can be estimated by the number of different side products, which appear only in small amounts[97]. Cation-π interactions are by now acknowledged as a major force for intermolecular recognition and protein stabilization. Thereby, tryptophane residues form more stable interactions than tyrosines and phenylalanines[98–100]. This principle is not restricted to enzyme catalysis but also applied in related fields[101,102].

Fifth, whether the terminating water acts as a nucleophile is depending on the near-by-water channel network and the exact position of the water molecule. Usually it is connected to a well-defined water molecule, which is connected to the cytosol via other water molecules[85].

1.3.4 Promiscuity and evolvability

Many terpene cyclases are considered to be promiscuous or multiproduct enzymes[1,103,104]. Sometimes a promiscuous or multiproduct activity is first revealed after minor alterations in the active site[76,105,106]. In general, this is thought to pose an evolutionary advantage since new cyclic products can easily be made without establishing a whole new biosynthetic pathway or inventing a new catalyst from scratch[81,107]. An extreme example of multiproducts among terpene cyclases is the γ-humulene cyclase, which can synthesize up to 52 different products[107]. However, while there are many examples for multiproduct-terpene cyclases there are only a few for cyclases converting non-natural substrates[108]. One prominent example is the *Aac*SHC.

To date, there are over 30 known substrates for *Aac*SHC other than the natural substrate squalene. An exemplary list of substrates that were accepted by *Aac*SHC and its variants is depicted in Figure 9. Among them are substrates that are significantly shorter (C_{10}, **19**) than squalene (C_{30}, **6**), longer (C_{35}, **30**), or posses other functional groups to attack an intermediate carbocation such as hydroxyl groups, ketones, amines or (hetero-)aromatic rings[3,8,112–115,86,91–94,109–111]. Squalene analoga and long alkyl chains were especially used to investigate the mechanism[86,93,113,114]. While, farnesyldimethylallyl-derivatives containing pyrrol (**22g**) und indolrings (**22h**) mainly proved the size and plasticity of the active site[8,111,112]. Moreover, polyprenyl phenyl ethers (**25a**) were observed to be converted with the phenyl moiety attacking an intermediate carbocation (Friedel-Crafts alkylation) first in the related enzyme *Zmo*SHC[91] and later also for the *Aac*SHC[94,109,116]. The conversion of homofarnesol (**22e**) is particularly useful for the synthesis of ambroxan, a high value fragrance and is one example for a patent on using the *Aac*SHC for industrial applications[8,117]. With the smaller farnesol (**22**) the first bi-molecular ether formation within the cyclase was shown. This was also observed for geraniol (**19**) and further expanded to alcohols of varying chain length as nucleophiles. Also variants with high selectivities for each of the resulting products were found[92].

Smaller substrates proved to be more demanding. The conversion of geraniol (**19**) and epoxygeraniol (**20**) was initially reported not to be possible[93] but was enabled by introducing point mutations, which facilitate a productive substrate prefolding[90,118]. Citronellal (**21**) was seen as substrate in a Prins reaction type with a variant of *Zmo*SHC first[119]. In the search for *Aac*SHC variants that convert either of the citronellal enantiomers, several variants with high product specificities were identified[33]. It should be highlighted that a triple variant of *Aac*SHC was engineered that produces (-)-isopulegol from *R*-citronellal, which is a precursor for the industrially relevant (-)-menthol (**10**) production[33] (see also chapter 1.2.1).

In general, the formation of new C-C bonds is of special interest for applications in synthetic organic chemistry. So far, only a few C-C bond forming enzymes with promiscuous activities are known, mostly lipases, proteases aldolases, and a few tautomerases and nucleases[120]. This makes *Aac*SHC's observed catalytic promiscuity combined with its evolvability even more impressive and highlights its unique features as a valuable tool in biosynthesis.

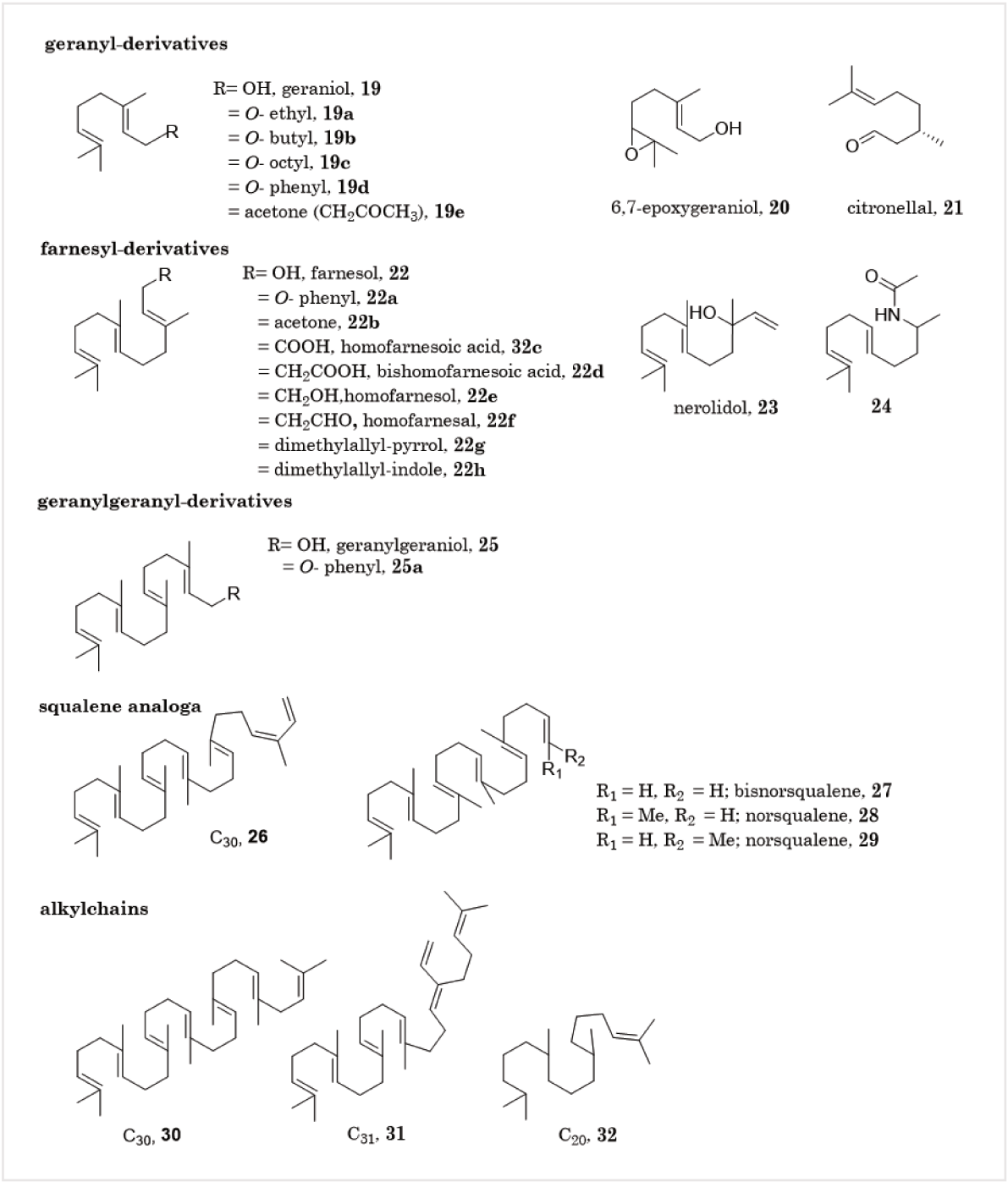


Figure 9: Accepted substrates by *Aac*SHC wild type and variants. Compiled and modified based on literature[3,8,111–115,81,86,91–94,109,110].

1.4 Engineering towards novel enzyme functions

Many reactions in classical organic chemistry make use of toxic metals and excess amounts of organic solvents. This can lead to a high production of unwanted waste, which is critical in an industry that is becoming more and more aware of its ecological consequences. Additionally, less waste production can also have economic benefits and is therefore highly desirable. A solution on the way to more environmentally-friendly processes is the use of enzymes. Their merits are for example a high stereoselectivity, the possibility to omit protection and deprotection steps, as well as their origin from renewable sources, biodegradability and non-toxic properties. In the past, enzymes have been associated with several limitations like not being available in large quantities, low stability or having a narrow substrate scope. However, these limitations have in part been overcome by advances in biotechnology and biocatalysis. Still, there is a great demand to find tailored enzymes for transformations that have so far not been addressed by enzymes[121–123].

1.4.1 Finding new biocatalysts

In the early days of biocatalysis, the main focus was on adjusting the reaction conditions to the enzyme needs to exploit their full potential. With the advent of new techniques in molecular biology it was possible to adjust the enzyme via enzyme engineering to meet the requirements of the desired reaction[124,125]. This can include optimizing an enzyme's stability, activity, stereoselectivity or alter the substrate specificity to accept non-natural substrates[126–128]. The most obvious way at hand was to mimic natural evolution in the lab which was developed into the field of directed evolution. Directed evolution is now a common and successful tool, which has been successfully applied in numerous cases to engineer proteins. It can also be extended to whole pathways or cells, for example to achieve higher tolerances towards a certain product[126,129]. To acknowledge the importance of directed evolution, Frances Arnold was awarded with the Nobel Prize in 2018 for her great contribution to the field[130].

Directed evolution generally starts with the introduction of random mutations into the protein of interest, followed by the evaluation of the improvement in the phenotype and the selection of the best variant. This process is repeated iteratively for five to ten rounds until the desired function is achieved[126,131,132]. Traditionally, UV-irradiation, error prone polymerase chain reaction (epPCR) and gene shuffling were the methods of choice for the creation of vast libraries (1000-2000 clones per round), which were screened by high-throughput methods in microtiter plates. They usually made use of spectroscopic properties like fluorescence absorption or coupling with an NADH-dependent reaction. Recently, new screening methods like FACS and microfluidics further increased the number of clones that can be screened in a certain amount of time[124,129].

However, despite its countless successes this method has several limitations. First, the introduced mutations in epPCR are biased, so not all the possible sequence space will be exploited for library creation. Second, since the best variant is selected after each evolutionary round for the next one, possible beneficial mutations will not be discovered or evolution can get trapped in a local minimum in the fitness landscape, without reaching the best possible variant. Moreover, screening systems with a very high throughput can get less accurate and one always needs a distinguishable feature to screen for. To obtain

such a suitable assay it can be tempting to use a method leading to an engineered enzyme, which does not exactly match the criteria for the desired application[126,129,131,133].

The strength of directed evolution lies in the fact that no prior knowledge on the function, the three-dimensional structure, nor the structure-function relationship of introduced mutations is necessary. However, due to detailed biochemical analyses on structure-function relationships over the past decades, more and more knowledge has been generated, which can now be used to create smarter libraries that cause less screening effort[134]. This strategy to use prior knowledge on the enzyme to select promising target sites for mutations to create smaller, more focused and functionally-rich libraries is termed semi-rational design[126,127,131]. When a beneficial mutation is found, the position is usually explored by site-saturation mutagenesis. Possibilities also emerged to reduce time effort and increase efficiency for this step. For example degenerated codons like NNK or NDT can be used or the 22c-trick method, which reduces the redundancy of the genetic code[135,136].

The advantage of semi-rational protein design lies in the fact that fewer iterative rounds are necessary and mutant libraries can be smaller, which reduces the screening effort dramatically. Additionally, semi-rational protein design can tailor the biocatalyst for the desired requirements more specifically without inadvertently affecting the catalytic efficiency[127,129,131]. On the other hand, extensive knowledge on the structure and biochemistry of a protein is necessary and some activating residues were found in the periphery in the past, where they never would have been addressed by semi-rational design[126,137].

Lately, the field of computational design has seen major achievements, which now contribute more and more to the success of protein design[127]. Within the field several new techniques were developed, focusing on protein flexibility, transition states (OptZyme), exploring the sequence space (ORBIT) or identifying tunnels (CAVER) and many more. The rationalization can be achieved on different levels of resolution, ranging from quantum mechanics (QM) calculations over hybrid quantum mechanics/molecular mechanics (QM/MM) calculations to molecular dynamics (MD) simulations. So far, the greatest strength of computational methods is to rationalize the impact of newly introduced mutations. Computationally-designed enzymes are still being outcompeted by laboratory-engineered enzymes, while the underlying reasons are still not completely understood. Nevertheless, it can be assumed that the field is at the advent of being able to reliably predict beneficial mutations and thus becoming an essential and inseparable part of protein engineering[138,139].

Despite recent developments and accomplishments, we are still beginning to understand structure-function relationships and the impact of newly introduced mutations, nevertheless complex interactions within the enzyme. Therefore, it must be carefully evaluated which design strategy is suitable for each individual case of protein engineering to benefit most of it.

However, independent from the design strategy, selecting a suitable starting point is still a critical bottleneck. Therefore, it can be beneficial to identify a promiscuous enzyme for

the desired reaction. Other than having a promiscuous activity, the enzyme also needs to be evolvable, e.g. to have both plasticity and robustness towards mutations[126,133,140].

1.4.2 Catalytic promiscuity

Catalytic promiscuity is an enzyme's ability to convert other substrates than its physiologically relevant ones[124,133,141,142]. Until several decades ago this was recognized only as a secondary function and enzymes were in general regarded highly specific for one substrate. By now, enzyme promiscuity is considered rather the rule than the exception and to be an inherent function of proteins[140,143,144]. In fact, it is estimated that every enzyme has on average ten promiscuous substrates, which they never encounter under physiological conditions[133]. Usually the conversion of these non-natural substrates show dramatically reduced activities due to the lack of evolutionary pressure and a sub-optimal binding in the active site[124,133,142,145,146].

The catalytic amino acids for a promiscuous reaction are usually the same as for the natural reaction. However, the results differ. For example, many proteases can break an ester bond in contrast to their usual C-N cleavage. Or metalloproteases that usually break P-O bonds can also cleave P-F bonds. This can be further expanded by the use of site-directed mutagenesis or the exchange of a metal cofactor to introduce new catalytic functions[146,147]. Additionally, a promiscuous activity of one enzyme can be the natural reaction of a related enzyme or be newly introduced by a single mutation. The same promiscuous activity can also appear quite often within one superfamily[140,146]. In general, catalytic promiscuity is considered an advantage since it can be the starting point for divergent evolution in nature or directed evolution in the laboratory, which can both lead to completely new enzymes or metabolic pathways. This can result in a more rapid and flexible adaption of an organism to changing environmental conditions or the development of new biocatalysts[124,128,133,142,147].

2 Motivation

2.1 Project idea

Monoterpenes are valuable compounds (chapter 1.2.1). Their selective synthesis is highly important and remains a mainly unsolved challenge, despite its evident demand. Therefore, it is desirable to synthesize them selectively. There are some reports on single-product monoterpene cyclases and attempts to metabolically engineer strains for a higher precursor flux since these enzymes strongly depend on a phosphorylated precursor[148,149]. However, in a large scale such processes could constitute a competition to human nourishment[16].

In general, many monoterpenes are interconvertible. Hence, an ideal process would make use of already present cheap monoterpenes (e.g. pinenes) which occur in waste products like turpentine and convert them into higher value ones. This was already attempted by utilizing engineered microorganisms growing on waste products. However, they were limited by the toxicity of the products or precursors[16,38,150,151]. Chemical isomerization on the other hand is unaffected by this and several catalysts with high acidic properties were reported for successful isomerizations. Despite intensive research these reactions still result in difficult to separate product mixtures, which can largely increase the downstream processing costs[152]. This is mainly attributed to intermediate carbocations that are energetically very similar and therefore hard to control by a catalyst (chapter 1.2.2)[45,58,70].

Mutational studies on monoterpene synthases already showed that they are capable of chaperoning reactions involving these carbocations and are able to address the slight differences in energy[73]. By fine tuning the steric and electrostatic surrounding of the intermediate cations, these enzymes are principally able to steer the reaction towards a desired product. Moreover, this fine-tuned cavity is subject to manipulation by mutagenesis[76,80] (chapter 1.2.3).

The motivation for this thesis is therefore to combine the potential of a protein scaffold to fine tune the catalyst cavity and therefore address the subtle differences in intermediate monoterpene cations with the power of a strong acid that was so far exploited for chemical isomerizations and can therefore make use of already available cheap monoterpenes.

A suitable enzyme for this strategy is the squalene-hopene cyclase (SHC) which possesses an in nature unusual high acidic strength combined with the ability to stabilize intermediate carbocations[88]. The size of the hydrophobic active site can also be altered without affecting the protonation machinery. This was already proven by exchanging hydrophobic amino acids with different sizes and thus changing the substrate specificity[90] (chapter 1.3). All in all, this makes SHC the perfect tool for the task.

2.2 Preliminary studies

In preliminary studies[153] it was shown that the wildtype *Aac*SHC is able to convert different isomers of pinene, with (+)-ß-pinene showing the highest conversion of the four pinene isomers. Further, an already present mutant library[90] in the first shell of the enzyme was investigated for the conversion of (+)-ß-pinene. The products obtained were α-pinene, ß-camphene, terpinolene, α-terpineol and further unidentified side products with

varying product distribution ratios, depending on the variant[153]. This showed that the *Aac*SHC is an ideal starting point for the project, since minor changes in the protein scaffold already had a great influence on the resulting product distribution. However, it was observed that the biotransformation setup seems to be crucial for product distributions[153].

Further, the influence of temperature and the nature of the detergent on the reaction was investigated. Biotransformations performed at 30 °C and in the presence of Triton-X-100 were found to be most suitable for the project since it produced the least background reactions[153].

2.3 Aim of the project

Based on these initial screening data, the first part of this thesis focused on improving the setup for biotransformations and the analytical methods. The second part of this thesis was in consequence dedicated to the creation of further SHC variants by protein engineering methods to improve selectivities. During this thesis further goals arose, as to create variants with higher activities, test other substrates and cyclases from other organisms. Finally, a hypothesis for the observed results was targeted in collaboration with the group of Prof. Sílvia Osuna (University of Girona, Spain).

3 Methods and materials

3.1 Materials

3.1.1 Media and buffers

All media and buffers used for the work with microorganisms were sterilized by heating to 121 °C and 1.3 bar for 20 minutes. For all works, either dH_2O or ddH_2O obtained by an Milli-Q Integral Water Purification System (Merck Millipore) were used. All buffers and media are listed in the following or are referenced in the corresponding chapter.

LB media (lysogeny broth)[154]: 10 g tryptone, 5 g yeast extract, 5 g NaCl in 1 L water.

TB media (adapted)[155]: consists of TB solution and Kpi buffer in a 9:1 ratio which have to be sterilized separately.

TB solution (terrific broth, adapted)[155]: 12 g tryptone, 24 g yeast extract, 4 g glycerol, 8 g lactose (see chapter 4.1.1.1) in 900 mL in water.

Kpi buffer (adapted)[155]: 0.17 mM KH_2PO_4, 0.72 mM K_2HPO_4 in water, pH 7.5.

SOC media (super optimal broth with catabolite repression)[156]: part 1: 20 g tryptone, 5 g yeast extract, 580 mg NaCl, 180 mg potassium chloride in 500 mL water, part 2: 950 mg $MgCl_2$, 3.6 g glucose in 500 mL water. Part 1 and 2 were autoclaved separately and then mixed 1:1.

2% *w/v* Agar media: 5g Agar in 250 mL LB media.

Lysis buffer[157]: 200 mM citrate, 0.1 mM EDTA in water, pH 6.

Solubilization buffer[157]: 60 mM citrate, 1% Triton X-100 in water, pH 6.

IEX-regeneration buffer[157]: 12 mM citrate, 500 mM NaCl, 0.2% Triton X-100 in water, pH 6.

IEX-wash buffer[157]: 12 mM citrate, 0.2% Triton X-100 in water, pH 6.

IEX elution buffer[157]: 12 mM citrate, 200 mM NaCl, 0.2% Triton X-100 in water, pH 6.

Whole cell reaction buffer[153]: 60 mM citrate, 0.2% Triton X-100 in water, pH 6.

SDS sample buffer (adapted)[31]: 0.1 M Tris-HCl, 100 mM $MgCl_2$, 0.05% glycerol, 140 mM sodium dodecyl sulfate (SDS), 150 µM bromophenol blue, 100 mM dithiothreitol (DTT) in water.

Coomassie staining solution (adapted)[31]: 2.25% (w/v) Coomassie brilliant blue in water.

Destaining solution[158]: 30% ethanol, 10% acetic acid, 60% water.

DNA loading dye[31]: 2 g sucrose, 10 mg Orange G in 5 mL water.

TAE buffer[31]: 2 M Tris acetate, 100 mM ethylenediaminetetraacetic acid (EDTA) in water, pH 8.

3.1.2 Strains

In this work different *E. coli* strains were employed. The strain *E. coli* BL21(DE3) is a commonly used strain due to its convenient properties for the expression of heterologous proteins. The genotype is *B F– ompT gal dcm lon hsdSB(rB–mB–) λ(DE3 [lacI lacUV5-T7p07 ind1 sam7 nin5]) [malB+]K-12(λS)*[159]. For the generation of this strain several proteases such as Lon or OmpT were deleted, to prevent the degradation of heterologous-expressed proteins. Furthermore, the *hsdSB* mutation inhibits plasmid loss and the gene for the T7 RNA polymerase was inserted into the chromosome under the control of the *lacUV5* promoter[160,161]. In general, the *lac* promoter is induced by lactose but inhibited in the presence of glucose due to low levels of cAMP. Therefore, all glucose has to be metabolized for the promoter to be activated. Yet, the variant *lacUV5* tolerates low concentrations of glucose which makes it an ideal tool for autoinduction[160–162]. The T7 RNA polymerase is derived from the bacteriophage T7 and frequently used for the expression of heterologous proteins for several reasons. First, it is very specific for its own promoter which is not occurring naturally in *E. coli*. Second, it can transcribe five times faster than the endogenous *E. coli* polymerase and thus saturates the ribosome after a few hours. This ensures an overexpression of the protein of interest. Third, it has a very low basal level and therefore the unintended expression of potentially toxic proteins is prevented[161].

The strain *E. coli* DH5α is a commonly-used strain for introducing new mutations in a plasmid, the genotype is *F– endA1 supE44 thi-1 recA1 relA1 gyrA96 deoR φ80dlacZΔM15 deoR Δ(lacZYA-argF)U169, hsdR17(rK–mK+), λ–*[163]. It is especially suited since the e*ndA1* mutation prevents the degradation of an inserted plasmid by intracellular nucleases and the *hsdR17* mutation inhibits the degradation of non-methylated DNA which is present after a PCR reaction, just to name a few beneficial mutations[164].

3.1.3 Plasmids and sequences

The *Aac*SHC gene is encoded on a pET22b(+) vector (Merck), containing an ampicillin resistance (*AmpR*) and is shown in Figure 70 (chapter 8.6.1). Variants that were generated in the course of this thesis are listed in Table 23 (chapter 8.6.2) . The corresponding primers are shown in Table 22 (chapter 8.5). A more detailed view on the *Aac*SHC sequence on the plasmid can be found in Figure 71 (chapter 8.6.1). The plasmid containing the sequence was obtained from the ITB in-house plasmid library and previously used in several studies[31,109,110,153,157,158].

In the course of this thesis SHCs from other organism were tested, as well. The plasmids originated from the ITB in-house library and were obtained in BL-21 cells. The protein sequences were received by unpublished ITB in-house documents from previous studies. To draw a comparison between those SHCs and the *Aac*SHC a pairwise sequence alignement[165] was conducted to learn the sequence identity. Therefore, the EMBOSS Needle alignment, using the Needleman-Wunsch alignment algorithm was employed on the website www.ebi.ac.uk from the European Molecular Biology Laboratory in Hinxton.

3.1.4 Enzymes

For molecular biology work, the polymerase Pfu Ultra II Fusion (Agilent technologies) was used in the corresponding buffer. To degrade parental DNA after a PCR reaction, Dpn1 (Thermo Fisher Scientific) was added to the reaction.

3.1.5 Designed primers

All primers were obtained from Metabion International, diluted to a working concentration of 10 µM and numbered using the prefix PR. Primers for sequencing were adopted from previous works[109] and are listed in Table 1. Primers that were used to introduce site-directed mutations can be found in Table 22 in chapter 8.5.

Table 1: Primers used for sequencing, adopted from previous works[109].

primer	sequence
PR01_T7 _Seq	TAATACGACTCACTATAGGG
PR02_In_Seq	CGCTGAGCATTGTGATGAGCCGC
PR03_In2_Seq	GCGTGAACTATCTGTATGGCACCGGCG
PR04_pET22_RP	GTTAGCAGCCGGATCTC
PR06_InR_Seq	CCAGCCACATGCGGGTAAACACGC

3.1.6 Chemicals

Chemicals were obtained from Alfa-Aesar, Carl-Roth, Fluka, Sigma-Aldrich and Thermo Fisher Scientific if not otherwise stated.

3.1.7 Consumables

Plastic reaction tubes in sizes 1.5 mL, 2 mL, 15 mL and 50 mL were obtained from Sarstedt, 5 mL reaction tubes from Eppendorf. Glass reaction tubes in 2 mL size with plug were purchased from Roth. Glas vials for GC analysis with a volume of 2 mL were obtained from Wicom.

3.1.8 Kits

For plasmid isolation the Zyppy™ Plasmid Miniprep Kit from ZymoResearch was used. The protocol was modified to the following. Three mL bacterial culture were centrifuged, the supernatant discarded, and the pellet resuspended in 600 µL ddH_2O and mixed with 100 µL 7x Zyppy™ Lysis Buffer. The tube was inverted and 350 µL Zyppy™ Neutralization Buffer added. The resulting mixture was inverted several times and centrifuged at 13 000 xg for 4 minutes. The supernatant was transferred into a Zymo-Spin™ IIN column and shortly centrifuged. The column was washed using 300 µL Zyppy™ Endo-Wash Buffer and centrifuged for 30 seconds and washed two times using 400 µL Zyppy™ Wash Buffer and centrifuged for 1 minute. The plasmid was eluted by adding 30 µL ddH_2O, incubating for five minutes and centrifuging for 1 minute.

For purification after QuikChange™ reaction the DNA Clean & Concentrator Kit from ZymoResearch was used according to protocol.

3.1.9 Instruments

Different centrifuges were used for various purposes. To centrifuge glass or plastic vials with volumes up to 2 ml an Eppendorf Centrifuge 5417C was used. For plastic tubes with volumes up to 50 mL an Eppendorf Centrifuge 5810R was chosen. To harvest cells in up to 500 mL a Beckman Coulter Avanti J-265 XP was used.

Reactions that needed shaking and temperature control were conducted using an Eppendorf Thermomixer comfort or an Infors HT Multitrin Pro, which was also used to incubate bacterial cells.

For electroporation a Biorad Pulse Controler was used in combination with a Biorad Gene Pulser. Bacterial cells were disrupted with a Branson Sonfier 250. Gels were visualized with a CanoScan LIDE 200.

For analysis an Agilent Technologies system was used consisting of a 7890A gas chromatograph (GC) system, a 7693 Autosampler and a 5975C inert MSD with Triple-Axis Detector. The combined system is referred to as GC/MS. The default column installed was an Agilent DB-5 or DB-5ms column. To determine enantioselectivities an Agilent Cyclosil B column was utilized. For headspace measurements a Shimadzu GC-2010 with a GCMS.QP2010 mass detector was employed with an installed Shimadzu HP-5msi column.

To measure wavelengths either a POLARstar Omega from BMG LABTECH, an Eppendorf BioPhotometer or a Thermo Scientific NanoDrop 2000 were used.

3.1.10 Software

The Microsoft Office Professional Plus 2016 package including Word, Power Point and Excel was used to generate most office related documents. The personal literature collection and citations were organized by Mendeley Desktop version 1.19.1. Chemical drawings were created using PerkinElmers ChemDraw Professional in the version 16.0.1.4(77). To depict the enzyme structure The PyMOL Molecular Graphics System, Version 2.0.6 Schrödinger LLC was employed and the software package YASARA by Elmar Krieger, Version 14.5.6 to create dockings.

For GC/MS analysis the software MSD ChemStation from Agilent Technologies was used in the version E.02.02.1431 and the NIST Mass Spectral Search Program for the NIST/EPA/NIH Mass Spectral Library in the version 2.0.

To operate the Thermo Scientific NanoDrop 2000 the corresponding software was utilized. The software MP Navigator Ex Cano Scan LIDE 200 coupled with a scanner created pictures of gels.

DNA sequences were depicted in SnapGene® 3.1.4 to verify mutations and to design primers.

3.2 Methods

Most methods, based on previous works[31,109,153,157,158], were adjusted and improved over the course of this thesis. The detailed implementations can be found in the corresponding experiments for the improvements and are listed as numbers in this chapter. It was shown that the results of the biotransformations are not affected but improvements were made concerning the time, analytical sensitivity or ease of handling.

3.2.1 Molecular biology methods

3.2.1.1 Site-directed mutagenesis

The QuikChange™ II Site-Directed Mutagenesis Kit (Agilent Technologies) was applied to introduce new mutations[166,167]. For single mutations the pET22b(+) vector containing the gene for *Aac*SHC served as a template. For each reaction the protocol shown in Table 2 with the temperature program depicted in Table 3 was used. The appropriate primers are listed in Table 22 in the appendix.

Table 2: Protocol for site-directed mutagenesis using QuikChange™.

	amount for one reaction	**final concentration**
template DNA (30 ng/µL)	1 µL	0.6 ng/µL
forward primer (10µM)	1.25 µL	0.25 µM
complement reverse primer (10 µM)	1.25 µL	0.25 µM
dNTP mix (10 mM for each NTP)	1.25 µL	250 µM each
PfuUltra Fusion Buffer (10x)	5 µL	1x
PfuUltra II Fusion Polymerase 0.05 U/µL	1 µL	0.001 U/µL
water	39.25 µL	

Table 3: Temperature program for QuikChange™ reaction.

reaction	**temperature**	**time**
initial denaturation	95 °C	2 min
number of cycles	16	-
denaturation	95 °C	30 sec
annealing	primer annealing temperature minus 5 °C	1 min
elongation	72 °C	4min
storage	8 °C	-

To digest parental DNA a mixture of 1 µL Dpn1, 6 µL 1x Tango buffer and 2 µL water were added to each reaction and incubated over night at 37 °C.

To control and evaluate the success of the Dpn1 digestion, one QuikChange reaction was conducted without polymerase, since this is the most expensive ingredient and should give no colonies.

3.2.1.2 Purification and transformation

The products of the QuikChange reaction and the Dpn1 digestion were purified using the ZymoResearch Clean&Concentrator™ kit according to the protocol. The purified DNA was eluted using 20 µL dH_2O.

For the transformation pre-cooled cuvettes for electroporation were used. Fifty µL electrocompetent *E. coli* BL-21 or DH5α cells were thawed on ice, mixed with 10 µL purified DNA and transferred into the cuvette. The electroporation was performed using 2.5 V, 200 Ω and 25 µFD. For a good result the electroporation should result in a time constant of about 4.5. The cells were then incubated in 1 mL warm SOC media for one hour and then plated on LB-agar plates containing ampicillin as selection marker.

As a control water was used for electroporation instead of DNA after previous transformations to check for aerosol and other contaminations.

3.2.1.3 Plasmid preparation, sequencing and analysis

For plasmid preparation colonies were picked and incubated in 5 mL LB media containing 100 µg/mL ampicillin over night. The plasmid was prepared using the ZymoResearch MiniPrep kit according to protocol or by dialysis using a nitrocellulose membrane filter with a pore size of 0.025 µm. DNA was applied on top of a filter floating on dH_2O and dialyzed for 15 minutes. The resulting concentration was determined using a Nanodrop spectrometer, diluted to 30 ng/µL and sent to GATC Biotech for sequencing with appropriate primers from Table 1. Exemplary sequencing of the whole plasmid showed no spontaneous mutations, therefore only mutated regions were checked by sequencing. Desired mutations were further used while unwanted mutations were discarded. During the preparation of this thesis GATC Biotech was taken over by Eurofins Scientific.

For qualitative analysis 1% agarose gels were used. Agarose was heated in TAE buffer and poured into the chamber together with 5 µL PeqGreen™ (VWR) staining solution. After solidification the samples together with DNA loading dye and GeneRuler™ 1 kb Plus DNA Ladder (Fermentas) were inserted into the pockets and run with TAE buffer at 120V for 60 minutes. UV light was used for visualization.

3.2.1.4 Preparation of electrocompetent cells

E. coli cells of choice were incubated in 5 mL LB media at 37 °C over night with shaking. Afterwards, they were subcultured in 500 mL LB media to an OD_{600} of 0.5, placed on ice for 15-30 min and centrifuged at 4000 xg at 4 °C for 10 minutes. The pellet was resuspended in 5 mL ice-cold water and filled up to 500 mL with cold water. The solution was centrifuged as before, the pellet resuspended in 5 mL ice-cold water and filled up to 250 mL with ice-cold water. The solution was centrifuged as before, the pellet resuspended in 5 mL ice-cold water and filled up to 25 mL with cold 10% glycerol. The solution was centrifuged as before, the pellet resuspended in 2 mL 10% glycerol and aliquoted in 50 µL fractions. They were directly frozen in liquid nitrogen and stored at −80 °C.

3.2.2 Biochemical methods

3.2.2.1 Cultivation

Different cultivation conditions (CC1-CC3, Table 4) were used to obtain the enzyme of choice.

Table 4: Cultivation conditions and identifier. All solutions contain 100 µg/mL ampicillin.

time scale/identifier	CC1	CC2	CC3
over night	5 mL LB media	5 mL LB media	-
during the day	-	50 mL TB media	5 mL LB media
over night	500 mL TB media	500 mL TB media	50 mL TB media

Cells containing the respective plasmid were inoculated in growth media containing ampicillin to maintain the plasmid. All cells were grown at 37 °C with continuous shaking at 180 rpm. Afterwards they were harvested by centrifugation at 10 543 xg (500 mL culture) or 3 220 xg (50 mL culture). For long term storage 500 µL of a grown cell culture was mixed with 500 µL 10% glycerol, frozen and stored at −80 °C.

3.2.2.2 Cell disruption and IEX purification

After harvesting the cell pellets, they were resuspended in 4 mL lysis buffer, 5 µL 0.1 mM Dnase solution and 10 µL 0.1M PMSF (phenylmethylsulfonyl fluoride) solution per gram cell pellet and sonified using a Branson Sonifier 250. The solution was sonified on ice-water for 4.5 minutes at 30% duty cycle and an output of 4 at the micro tip. The cell debris was collected by centrifugation at 6036 xg for 30 minutes at 4 °C. The supernatant was discarded since the *Aac*SHC is membrane anchored and can therefore be found within the cell debris. The resulting pellet was resuspended using 1 mL solubilization buffer per gram pellet and incubated at 4 °C over night while rotating slowly. The next step was to denature most of the *E. coli* proteins using a heat shock in a water bath at 50 °C for 30 minutes and centrifuging for 60 minutes at 43667 xg. Afterwards the supernatant was diluted 1:5 with water to reduce the detergent concentration and filtered using a 0.2 µm pore size filter.

To prepare the ion exchange column (IEX) purification diethylaminoethyl sephacel was regenerated in an appropriate column using 5 column volumes of IEX regeneration buffer and subsequently washed using 5 column volumes of IEX wash buffer. The protein solution was applied onto the column and afterwards washed using 5 column volumes of IEX wash buffer. The protein was eluted using IEX elution buffer. The protein solution was either directly used for biotransformations or stored at −80 °C. Preliminary tests (not shown) showed that the frozen enzyme is stable over weeks without losing activity which is not applicable to frozen pellets.

The IEX purification was improved during this thesis and the bachelor thesis of Josip Tulumovic [168] under supervision of the author. Different variations (IEX1-IEX3) can be found in Table 5.

Table 5: Variations on IEX chromatography and identifier.

options/identifier	IEX1	IEX2	IEX3
column	1 mL syringes	6 mL syringes	adaptable columns
column volume	1 mL	2 mL	7-9 mL
elution	5 mL	6 mL	collecting 1 mL fractions

3.2.2.3 Protein analysis

Protein solutions were analyzed qualitatively using standard SDS-PAGE[169] (sodium dodecylsulfate polyacrylamide gel electrophoresis) and quantitatively using the BradfordUltra assay (Expedeon Inc), an adaption of the original bradford assay[170].

For SDS-PAGE the samples were mixed with an appropriate amount of SDS sample buffer, which reduces disulfide bonds and applies an overall negative charge and heated to 95 °C for 5 minutes to denature the proteins. Afterwards, the samples were applied to a precast 12% ExpressPlus PAGE gel from GenScript together with a PageRuler Unstained Protein Ladder. The gel was run with the provided buffer at 140 V for 45 minutes which separates the linearized proteins by size[169]. The bands were stained by shaking in a Coomassie staining solution for several hours and subsequent shaking in destaining solution for several hours.

To determine the protein concentration the samples were diluted 1:10 with buffer and 10 µL of it were mixed in triplicates with 150 µL BradfordUltra from Expedeon Inc. A calibration curve with BSA (bovine serum albumin) concentrations ranging from 0 to 1 mg/mL was treated the same. All solutions were measured at 595 nm and the protein concentrations calculated accordingly.

3.2.3 Biotransformations

For biotransformations, the catalyst was either resuspended in whole cell reaction buffer for whole cell biotransformations to a concentration of 2 mg/mL wet weight or used in IEX elution buffer for IEX-purified enzymes with a maximum obtained concentration. To this a 200 mM stock solution containing the substrate solved in the co-solvent was added in a 1:100 ratio, resulting in a 2 mM final concentration. The reaction was incubated at 30 °C at different speeds for 40 hours. The biotransformation was then extracted by adding the same volume of extraction agent containing a standard, vortexing for 40 seconds and centrifuging at 7000 rpm for two minutes. The organic phase was collected and analyzed by GC/MS. For the two-phase system, the biotransformation with the extraction agent overlaid was vortexed for 40 seconds and centrifuged at 4000 rpm for ten minutes. The details on different variations of this protocol can be found in Table 6. All experiments were performed at least in biological duplicates (different cell cultures on different days) and technical triplicates with (+)-ß-pinene as substrate, if not otherwise stated. The IEX-purified enzyme can be stored at −80 °C before usage in biotransformations. For the work with whole cells the cells have to be freshly harvested.

Table 6: Variations on the biotransformation protocol and corresponding identifiers. * The speed of 600 rpm refers to an Eppendorf Thermomixer comfort. ‡ The speed of 180 rpm refers to an Infors HT Multitron shaker.

options/ identifier	BT1	BT2	BT3	BT4	BT5	BT6	BT7	BT8
catalyst	IEX purified	IEX purified	IEX purified	whole cells	whole cells	whole cells	whole cells	IEX purified
reaction vessel	2 mL with plug	2 mL with plug	2 mL with plug	2 mL with plug	2 mL with plug	2 mL gastight GC vial	1.1 mL vials	1.1 mL vials
reaction volume	500 µL	500 µL	500 µL	500 µL	500 µL	1 mL	1.1 mL	1.1 mL
incubation	600 rpm*	600 rpm*	600 rpm*	600 rpm*	600 rpm*	180 rpm‡	600 rpm*	600 rpm*
co-solvent	DMSO	DMSO	ethanol	DMSO	ethanol	ethanol	ethanol	ethanol
extraction	2 times	1 time	1 time	1 time	1 time	two-phase system	1 time after transfer into 2 mL vials	1 time after transfer into 2 mL

3.2.4 Data analysis

Different temperature programs for GC/MS were tested during this thesis. In the following (Table 7-13) the programs are listed, which were used for most of the experiments. Except for headspace measurements or to determine enantioselectivities an Agilent DB-5 column ((5%-phenyl)-methylpolysiloxane, 30 m length, 0.25mm diameter, 0.25 µm film) or Agilent DB-5ms column was installed in the GC oven of the 7890A GC system with helium as carrier gas. To determine enantioselectivities an Agilent Cyclosil B column (30% heptakis (2,3-di-*O*-methyl-6-*O*-*t*-butyl dimethylsilyl)-β-cyclodextrin in DB-1701, 30 m length, 0.25mm diameter, 0.25 µm film) was installed. The injection temperature was 250 °C and the system was operated at 15.99 psi pressure.

Early in this thesis it was discovered that a splitless injection with 1.5 µL injection volume yields more reliable results, therefore the listed temperature programs were applied almost exclusively with a splitless injection, except for GC-P3 and GC-P4 temperature programs. The detection occurred in two ways, via FID and mass detection. The FID (flame ionization detection) was operated at 320 °C. The mass detection was operated at 230 °C and 70 eV in an EI (electron ionization) mode. Most measurements were recorded in the SCAN mode, usually ranging from *m/z* 50-200. Few measurements were conducted in the SIM mode, usually containing the masses *m/z* 91 for *p*-xylene and *m/z* 121 and *m/z* 136 for monoterpenes.

For headspace measurements a Shimadzu GC-2010 with a GC/MS.QP2010 mass detector was employed with an installed Shimadzu HP-5msi column, equivalent to the DB-5 column.

For data analysis the resulting chromatograms were always compared with the buffer control and D376C negative control (used in preliminary studies[153]) or AxAA negative control and only peaks that could be specifically assigned to *Aac*SHC or variant activity were considered for further processing. The data were obtained as peak area percentages.

Table 7: Temperature program GC-P1 for pinene.

rate in °C/min	temperature in °C	hold time in min
	70	1
6	85	1
3	90	1
3	110	2
3	130	1
25	140	0
50	310	1

retention times on an Agilent DB-5 column: 4 min *p*-xylene, 5.1 min α-pinene, 5.3 min camphene, 5.8 min ß-pinene, 5.2 min limonene, 7.4 min terpinolene, 8.9 min terpineol.

Table 8: Temperature program GC-P2 for pinene.

rate in °C/min	temperature in °C	hold time in min
	70	1
6	85	1
3	90	1
3	100	1
75	190	0
120	310	1

retention times on an Agilent DB-5 column: 3.7 min *p*-xylene, 4.6 min α-pinene, 4.8 min camphene, 5.3 min ß-pinene, 6.4 min limonene, 7.2 min γ-terpinene, 8.1 min terpinolene.

Exemplary chromatograms for temperature program GC-P1 can be found in Figure 15 and Figure 16 and for GC-P2 in Figure 42 and Figure 41.

Table 10: Temperature program GC-P3 for pinene.

rate in °C/min	temperature in °C	hold time in min
	50	0
2.5	60	0
7	105	0
60	300	1

retention times on an Agilent DB-5 column: 5.9 min *p*-xylene, 8.3 min α-pinene, 8.7 min camphene, 9.4 min β-pinene, 10.3 min α-terpinene, 10.5 min limonene, 11 min γ-terpinene, 11.2 min unidentified 1, 11.4 min terpinolene, 11.7 min fenchol, 11.8 min unidentified 2, 11.9 min unidentified 3, 12.0 min unidentified 4, 12.05 min unidentified 5, 12.2 min borneol, 12.2 min terpinene-4-ol, 12.4 min α-terpineol.

Table 9: Temperature program GC-P4 for pinene.

rate in °C/min	temperature in °C	hold time in min
	50	0
2.5	60	0
5	105	1
10	160	1
60	300	0

retention times on an Agilent DB-5 column: 7.0 min *p*-xylene, 8.6 min α-pinene, 9.1 min camphene, 9.8 min β-pinene, 11.2 min α-terpinene, 11.4 min limonene, 12.1 min γ-terpinene, 12.9 min unidentified, 13.2 min terpinolene, 13.8 min unidentified, 14.3 min unidentified, 14.5 min fenchol, 15.3 min unidentified, 15.4 min unidentified, 15.5 min unidentified, 16.0 min borneol, 16.1 min terpinene-4-ol, 16.6 min α-terpineol.

For temperature program GC-P3 exemplary chromatograms can be found in Figure 18 and Figure 17 and from Figure 72 to Figure 76 in the appendix in chapter 8.8.1.

For temperature program GC-P4 exemplary chromatograms are shown in Figure 31 and Figure 32 and from Figure 45 to Figure 49 in the appendix in chapter 8.8.2.

Table 12: Temperature program GC-Q1 for squalene.

rate in °C/min	temperature in °C	hold time in min
	120	3
30	310	15

retention times on an Agilent DB-5 column: 4.9 min *p*-xylene 10.9 min squalene, 14.4 min hopene, 16.4 min hopanol.

Table 11: Temperature program GC-P5 for enantioseparation.

rate in °C/min	temperature in °C	hold time in min
	70	5
2	115	3
50	200	2

retention times on an Agilent Cyclosil B column: 8.5 min *p*-xylene, 13 min (-)-α-pinene, 13.2 min (+)-α-pinene, 15 min camphene, 15.5 min (+)-β-pinene, 18.2 min *S*-limonene, 18.8 min *R*-limonene.

An exemplary chromatogram for temperature program GC-Q1 can be found in Figure 11 and for temperature program GC-P5 in Figure 20 and Figure 21.

Table 13: Temperature program GC-S1 for sabinene.

rate in °C/min	temperature in °C	hold time in min
	70	1
15	88	0
8	130	0
120	310	1

retention times on an Agilent DB-5 column: 3.3 min *p*-xylene, 3.8 min unidentified P1, 4.3 sabinene, 4.4 min ß-pinene, 4.7 min unidentified P8, 4.9 min α-terpinene, 5.1 min unidentified P9, 5.4 min γ-terpinene, 5.6 min unidentified P4, 6.1 min unidentified P5, 6.9 min unidentified P10, 7.3 min terpinen-4-ol, 7.5 min α-terpineol.

For temperature program GC-S1 an exemplary chromatogram is shown in Figure 59.

Most measurements were conducted in the SCAN mode. The advantage of the SCAN mode is that small peaks are, additionally to their retention time, verified by their mass fragments and structure prediction with the NIST data base. Thus, monoterpenes with similar retention times could be distinguished by their mass fragmentation.

Moreover, the GC-MS occasionally showed minor deviations in retention times in-between runs. After intensive troubleshooting this was subjected to minor variations in gas pressure. When minor variations in retention times occurred, the peak was still unambiguously identified by mass fragmentation.

Detection of many solvent ions can pollute the mass detector and thus lower its sensitivity. Therefore, in most chromatograms recording started after several minutes due to technical requirements. Towards the end of the chromatograms, many exhibit peaks that are not assigned to specific molecules. These are usually fatty acids and other molecules with a high boiling point. At the end of a run technical requirements demanded heating to very high temperatures. Otherwise the extracted fatty acids would stay on the column and thus lower the performance over time.

Due to software reasons, depicted overlaid chromatograms show the TIC (total ion count) signals. The TIC is sufficient for a representation regarding quality and relative quantity in signals of the monoterpenes. However, integrations were always performed in the FID signal.

Furthermore, an indol peak can be spotted in some chromatograms (e.g. Figure 31, Figure 41 or Figure 45). Indol was identified as a stress response in *E. coli* due to the presence of an antibiotic such as ampicillin[171] and in some chromatograms solvent peaks are visible (e.g. Figure 20, retention time 4-7 min) due to an impure solvent.

3.2.5 Data processing

The obtained data from the GC/MS were processed by the following arithmetic operations. All peaks areas were integrated in the FID signal. The equation

$$product\ distribution\ ratio(x) = \left(\frac{peak\ area(x)}{\sum all\ product\ peak\ areas}\right) \cdot 100$$

was used to determine product distribution ratios. Whereas the equation

$$analyte\ distribution\ ratio(x) = \left(\frac{peak\ area(x)}{\sum all\ peak\ areas}\right) \cdot 100$$

was used to determine analyte distribution ratios. The conversion was calculated by

$$conversion = 100\% - analyte\ distribution\ ratio\ (of\ the\ substrate)$$

The term conversion is also sometimes referred to as activity.

A calculation via the molecular concentration was not considered productive due to the high volatility of the substrate and products.

In general, conversions should be handled with care. Although various efforts were made (see chapter 4.1.1), conversion rates cannot be compared absolutely from batch to batch despite identical handling and same expression levels. The reason for this is attributed to evaporation. Whenever a reproducible trend for a higher conversion rate was observed, the changes were put in relative numbers to describe them. The numbers given are calculated in relation to one another within one batch. Additionally, they were calculated in relation to the wildtype for comparisons between batches. Absolute numbers for conversions are omitted purposely.

Product distribution ratios are very stable and are maintained between the protocols BT1-BT8. Minor variations can be attributed to the number of total products detectable with the applied protocols.

4 Results

4.1 SHC-catalyzed isomerization of (+)-β-pinene

In preliminary studies[153] it was shown that *Aac*SHC can convert pinenes, whereof (+)-ß-pinene showed the highest conversion. This isomerization based on a protonation of (+)-ß-pinene lead to the formation of α-pinene, ß-camphene, limonene and α-terpineol as the main products. An initial screening of an already present mutant library[90] led to the identification of several variants that shifted the product distribution. These results were very promising and proved that the *Aac*SHC is a suitable tool for an enzymatic isomerization of pinene with a selective product outcome. However, in preliminary studies differences in product selectivities depending on the biotransformation setup were noticed[153]. Therefore, the first task in this thesis was to improve biotransformations and the analytical setup.

4.1.1 Reaction engineering

At first, the main focus was on increasing analytical sensitivity since *Aac*SHC catalyzed-conversion of (+)-ß-pinene showed generally low conversion rates. Therefore, various measures were taken to improve the detection of monoterpene. During the search for improved variants the setup for biotransformations and analysis was further improved to reduce time effort, increase the throughput or when new trace products appeared. However, with every improvement and alteration in the setup it was investigated if and how this change affects the result of the biotransformations. A variation of up to ~5% in product distribution was considered acceptable.

In the beginning of this thesis α-pinene, ß-camphene and limonene were the main products. The variant D376C was considered to have no activity based on previous studies[109,110,153,158] and was therefore treated as negative control. This led to the identification of α-terpineol as a side product not produced by the cyclase, which later turned out not to be true (see chapter 4.1.2.3). Moreover, trace products were not detectable with the early methods.

4.1.1.1 Cultivation of E. coli cells

In order to obtain as much *Aac*SHC protein as possible in a reasonable amount of time, the cultivation was investigated and improved.

The gene for *Aac*SHC is encoded on a pET22b(+) plasmid. The plasmid contains an ampicillin resistance and the expression of the *Aac*SHC gene is linked to T7 RNA polymerase, which is under the control of a *lac*UV5 promoter. The expression can be induced by isopropyl-ß-D-galactosid (IPTG) or by autoinduction using lactose[160]. The advantage of autoinduction is that the expression starts once the cells reach a certain metabolic state and the need for OD_{600}-measurements and the presence of an experimenter is omitted, while the metabolic stress to the cells is also lowered[172]. However, it is strongly dependent on the presence of lactose. This is usually guaranteed by the presence of yeast extract in the media. However, Nair et al showed that yeast extracts from different suppliers show various lactose contents, which can affect the autoinduction[173]. Therefore, an additional lactose content in TB media and its effect on the expressed protein

concentration was tested. Investigated were 8 g/L, 12 g/L lactose *versus* no additional lactose. The results were analyzed by Bradford concentration determination and visualized on an SDS-gel (not shown). It turned out that there was no significant change in protein concentration when using no lactose (~3,4 mg/mL), 8 g/L lactose (~3,5 mg/mL) or 12 g/L lactose (~3,8mg/mL). However, all further experiments were conducted with additional 8 g/L lactose in TB media since the addition has no negative influence and makes the expression of high protein contents independent of the yeast extract batch.

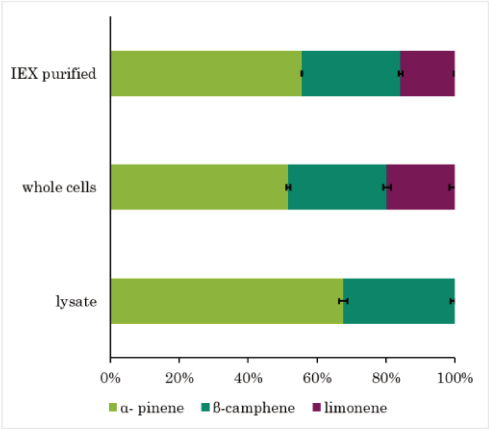
Furthermore, Josip Tulumovic investigated in his bachelor thesis, under authors supervision, a 2-step cultivation with dilutions each 1:10 instead of a single 1:100 dilution from CC1 to CC2 protocol (see Table 4). This improved the protein yield of about 23%[168] and inspired a single 1:10 dilution (CC3 protocol) for whole cell biotransformations, which reduces time effort by 60% and yields enough biomass for biotransformations (~1000 mg (baffles) to ~600 mg (without baffles) in 50 mL cultures.

The next step after obtaining enough protein was to improve the biotransformation setup.

4.1.1.2 Biotransformations

A very time consuming step is the purification of *Aac*SHC protein by ion-exchange chromatography (IEX) (see chapter 3.2.2.2). Therefore, it was investigated whether whole cells or lysate yield the same biotransformation results as IEX purified protein. Working with whole cells is well established in biocatalysis[121,124] and was shown by co-workers for the *Aac*SHC before (unpublished). However, it should be noted that the natural substrate squalene can not be converted when using whole cells since it gets degraded by the metabolism of the cell[8] (data not shown). For the experiment 3 mg IEX purified protein (BT2 protocol) and 50 mg/mL cell pellet in whole cell buffer (BT4 protocol) were used. The lysate after heat shock was diluted 1:5 with water to reach a Triton X-100 concentration of 0.2%. The data in Table 14 shows that the results of the biotransformations using IEX protein or whole cells are quite comparable concerning product distribution ratios. The reaction using whole cells shows a slightly higher content of limonene which is, negligible for screening purpose. The biotransformation with lysate shows no limonene and completely different product distribution ratios. Therefore, diluted lysate was no longer considered for the use in biotransformations.

Table 14: Comparison of different purification methods for biotransformations of β-pinene. Product distributions are given in % and are visualized: ■ α-pinene, ■ β-camphene, ■ limonene

	α-pinene	β-camphene	limonene	visualized
IEX purified	55.5 ± 0.2	28.7 ± 0.5	15.8 ± 0.3	
whole cells	51.9 ± 0.6	28.7 ± 1	19.9 ± 1	
lysate	67.6 ± 1	32.4 ± 1	0	

Further, it should be mentioned that the GC peak areas were ~6 times higher for IEX purified enzyme compared to whole cells, which might be due to higher catalyst concentrations (see chapter 4.1.1.4). However, the higher conversion using IEX purified protein (~6%) compared to whole cells (1%) can not be explained by the amount of catalyst alone and the trend was also observed in follow up experiments.

In order to omit protein purification processes the use of whole cells as catalyst was investigated. Therefore, different cell concentrations were investigated using BT4 protocol. A cell pellet concentration of 100 mg/mL produced 1.3 times higher yields than 50 mg/mL but was not further increased using 150 mg/mL cell pellet. Therefore, 100 mg/mL cell pellet concentration were further used. In this context it should also be noted that it was observed that frozen cell pellets used as catalyst loose activity over time, which is why only fresh cells were used for biotransformations.

After a successful biotransformation it is necessary to find a suitable way to extract the products. Therefore, solvents and extraction procedures were investigated.

4.1.1.3 Solvent and extraction evaluation

For a sufficient extraction of biotransformations, different organic solvents were investigated. Therefore, biotransformations without a catalyst or cells were prepared using BT2 protocol and directly extracted 1:1 with either ethyl acetate (EtOAc), cyclohexane or methyl-*tert*-butylether (MTBE). After mixing and centrifuging, milky phases were observed for the solutions containing cyclohexane or MTBE (see Figure 10). Therefore, they were dismissed as possible extracting agents. Ethyl acetate formed two clear phases that enabled a good extraction. Further ratios of biotransformation to EtOAc were tested ranging from 1:0.5 to 4:1. A ratio of 1:1 proved to have the highest retrieval of the substrate. During these experiments it was observed that a single extraction is not sufficient for a full retrieval. Biotransformations without catalyst or cells containing either the substrate ß-pinene or one of the products α-pinene, limonene or ß-camphene were extracted ten times. It was shown that 4 extractions are necessary to retrieve 90%

α-pinene, 62% ß-camphene and 6 times extraction to retrieve 86% ß-pinene and 90% limonene. A full retrieval was not possible for any of the substances. However, it was shown that the product distribution ratios are not influenced by the number of extraction steps and that the first extraction is representative for all further extraction steps. Therefore, a single extraction was considered sufficient for screening.

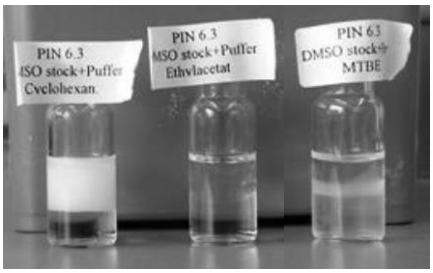


Figure 10: Different solvents for extraction after mixture with buffer, DMSO and β-pinene. From left to right: cyclohexane, ethyl acetate, and MTBE.

During these experiments it was observed that ß-camphene is not very soluble in dimethyl sulfoxide (DMSO) in which stock solutions were prepared. This might affect the outcome of the biotransformation. Therefore, alternatives were investigated. Most monoterpenes are reported to be soluble in ethanol, which was verified experimentally for the monoterpenes of interest in this thesis. Moreover, it was already shown that *Aac*SHC tolerates 10% methanol, however, with some loss of activity[109].

To verify this assumption, squalene was converted using IEX purified enzyme and either BT2 (DMSO) or BT3 (ethanol) protocol. The results were analyzed with GC-Q2 program. It was shown that the activity of the cyclase is not affected by using 1% ethanol.

As quality control an external standard is added to the extracting agent. A well established standard for *Aac*SHC was the use of 1-decanol as external or internal standard[109,110,153,157]. However, the peaks in the GC/MS were broad due to the relatively low temperatures necessary for monoterpene separation. A peak that is hard to integrate is not suitable as a standard. Therefore, *p*-xylene was tested as possible standard. It was reported as standard before in other works[174] and shows a similar boiling point to the monoterpenes of interest. In the chromatograms *p*-xylene showed a sharp peak. It was further tested to use *p*-xylene as internal standard. Therefore, biotransformations were prepared using the natural substrate squalene and BT2 protocol either with or as a control without equimolar amounts of *p*-xylene. The results were analyzed with GC-Q1 temperature program. It was shown that the activity of the cyclase is not inhibited in the presence of *p*-xylene (Figure 11). Therefore, it was also used as an internal standard.

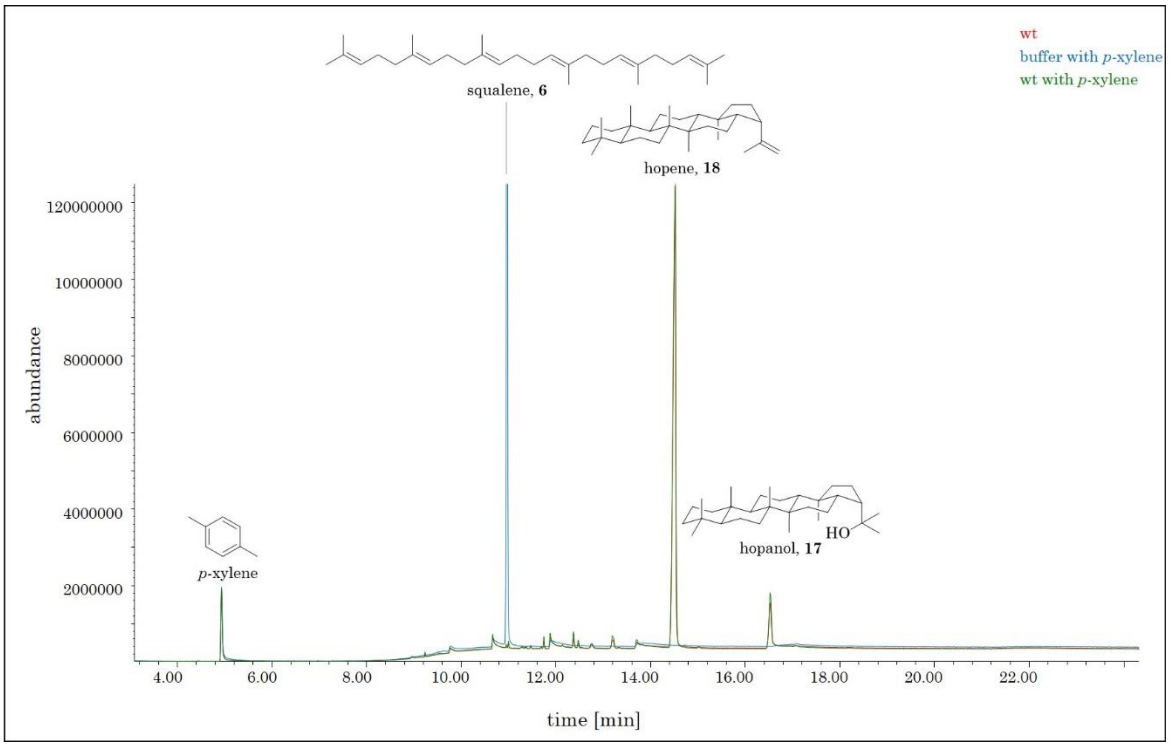


Figure 11: Full length overlaid chromatogram of wild type with or without *p*-xylene and buffer using temperature program GC-Q1. The standard *p*-xylene elutes at 4.9 min and squalene at 10.9 min, hopene at 14.4 min and hopanol at 16.4 min.

To follow a successful extraction, samples have to be measured with a suitable setup.

4.1.1.4 Detection evaluation

A major issue when converting ß-pinene using *Aac*SHC is the low conversion and the resulting small peaks in the GC/MS chromatograms. Especially, since this can cause unjustified variability upon integration of these small peaks. Therefore, extensive efforts were dedicated to increase analytical sensitivity. One attempt was to vary the substrate concentration using BT2 protocol and GC-P1 program. Higher substrate concentrations (up to 10 mM) resulted in higher peaks by a factor of 1.5 although the substrate concentration was increased by a factor of 5 and the conversion dropped by 64% compared to 2 mM substrate conversion. Lower concentrations (up to 0.25 mM) decreased the signal intensities by a factor of 3.2, while the substrate concentration was lowered by a factor of 4. The conversion increased by 38% compared to 2 mM substrate conversion. However, the product distribution ratios were not affected by changes in substrate concentration.

Considering signal intensities, mass detector saturation with analytes and conversion rates, 2 mM substrate concentration were ideal. Aiming at an increase in signal intensities it is suitable to lower the substrate concentration even though the conversion is increased. However, increasing the substrate concentration, the resulting product peaks are not increased by the same factor. Moreover, high substrate concentrations can potentially harm the MS detector of the GC/MS over time.

The concentration of the purified enzyme has a direct influence on the peak intensities, as well. Experiments were conducted using the same purified enzyme solution either undiluted or diluted (1:1 or 1:3) with the corresponding buffer (IEX elution buffer), BT2

protocol and temperature program GC-P2. The resulting signals decreased almost linear to the dilution of the enzyme but without affecting the product distribution ratios. Therefore, maximum protein concentrations after purification can be used, without the need to adjust to the same and therefore the lowest obtained concentration. Further, these findings support the possibility to use whole cell catalysts, since slight variations in enzyme expression do not affect the relative product distributions.

The first major improvement was the establishment of a two-phase system (BT6 protocol), which uses GC vials for biotransformations. The biotransformation solution containing whole cells as catalyst is directly overlaid with the extraction agent containing the standard *p*-xylene. Thus, the evaporation of substrate and products is greatly reduced. The following shaking of completely filled vessels during the incubation is less vigorous to ensure minimal contact between the two phases and thus minimizes enzyme denaturation. As a result to the slow shaking the utilized whole cells gradually sediment. This procedure reduces the conversion by ~3% but increases the signal intensities of the volatile analytes by a factor of 100.

Towards the end of this thesis, signal intensities were further increased by a method (BT7 protocol) adapted from Nico Kress[31]. This method uses completely filled 1.1 mL glass reaction vessels for biotransformation. Afterwards the reaction is transferred into larger vials for extraction. The advantage of this method is to leave no gas volume during the incubation, which greatly reduces evaporation. Moreover, since the organic layer is not overlaid, the vials can be agitated more rigidly, which increased the cyclase activity by 16-fold. With this method even trace products were visible, which required a new analysis method (GC-P3 and GC-P4), as well.

4.1.1.5 Quantification of monoterpene depletion

As described in previous chapters, it was noted that not all of the substrate was retrieved or found as product after biotransformation. It was anticipated that the loss was due to evaporation[35]. To quantify the depletion, calibration curves were recorded under different conditions. First, different concentrations were instantly measured in ethyl acetate. For the second curve a sample incubated under BT2 protocol conditions without cells or enzyme was directly extracted and measured. The third samples were treated the same as a biotransformation but without catalyst. Hence, the samples without enzyme or cells were incubated in 2 mL glass vials for 40 hours and afterwards extracted.

Although, it was already known that significant substrate and product deficiency is caused by extracting only once (see chapter 4.1.1.3), an even greater deprivation was found applying incubation conditions (40 h, 30 °C). This seems to be caused by evaporation. Potential solutions to this issue are the usage of the two-phase system (BT7 protocol) or completely filled vials (BT8 protocol). The different curves in Figure 12 show that due to the insufficient extraction 61% of the initially added substrate cannot be redetected and 89% are lost due to evaporation during the incubation. In total this amounts to 96% over the course of the reaction when using BT2 protocol. This is in good accordance with observations in preliminary studies and experiments on improving the reaction conditions. This highlights the need for the here performed conditional improvements as a basis for further studies.

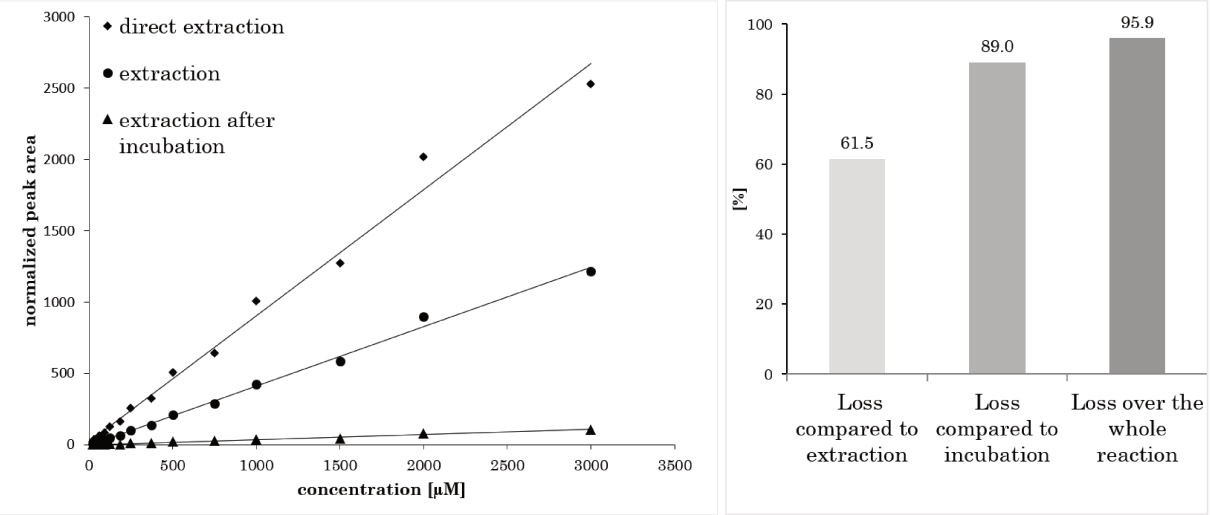


Figure 12: Quantification of substrate depletion after incubation and extraction under biotransformation conditions (BT2 protocol) for a quantitative estimation on analyte deprivation. Left: Concentration curves measured directly in ethyl acetate (♦ direct concentration), after extraction from buffer (● extraction) and after incubating the buffer for 40 h at 30 °C, hence reaction conditions (▲ extraction after incubation).
Right: Visualization of the total depletion during the different procedures.

Another approach to address the evaporation problem was to use headspace measurements. Therefore, biotransformations were performed in 21 mL headspace vials. The vials were heated to 145 °C for 8 minutes and 1 mL gas volume was injected using a pre-heated syringe (150 °C) on a Shimadzu GC-2010 with a GCMS-QP2010 mass detector and the temperature program GC-P1 and a split of 1:8. The resulting product distributions

differed dramatically from extraction experiments. The resulting product distribution for α-pinene was 46.4% and thus close to extractions, whereas the product distribution for ß-camphene was greatly increased (44.6%). Limonene on the other hand was heavily reduced (8.8%), which is due to the higher boiling point of 177 °C, whereas α-pinene and ß-camphene show boiling points of 155 °C and 160 °C, respectively. Since the device could not heat higher to reach the boiling point of limonene, this method had to be discarded.

4.1.1.6 Time dependency of the reaction

In order to improve the reaction regarding yield and time efficiency, the conversion of the reaction was determined every hour when possible for 50 hours. BT3 protocol and GC-P2 temperature program were applied. The product distribution ratios remained constant and the standard deviation for the product distribution was <4% over the entire reaction. This indicates that no products serve significantly as new substrates, which might influence the product distribution ratios over time. The conversion steadily increased over the reaction time and is shown in Figure 13. These results show that the reaction can in theory be stopped after a few hours with respect to the product distribution ratios. However, the resulting peak areas can be very low depending on the applied method and therefore lead to inconclusive results. Therefore, it was advised (using BT3 protocol) to keep a reaction time of 40 hours as a compromise of time efficiency and high enough peak areas for significant peak integration results. Moreover, it should be noted that although the conversions were determined with adequate accuracy in technical duplicates, the given numbers are not to be considered absolute. Due to the high evaporation and monoterpene depletion during the incubation these results might not be completely reproducible concerning the conversion numbers.

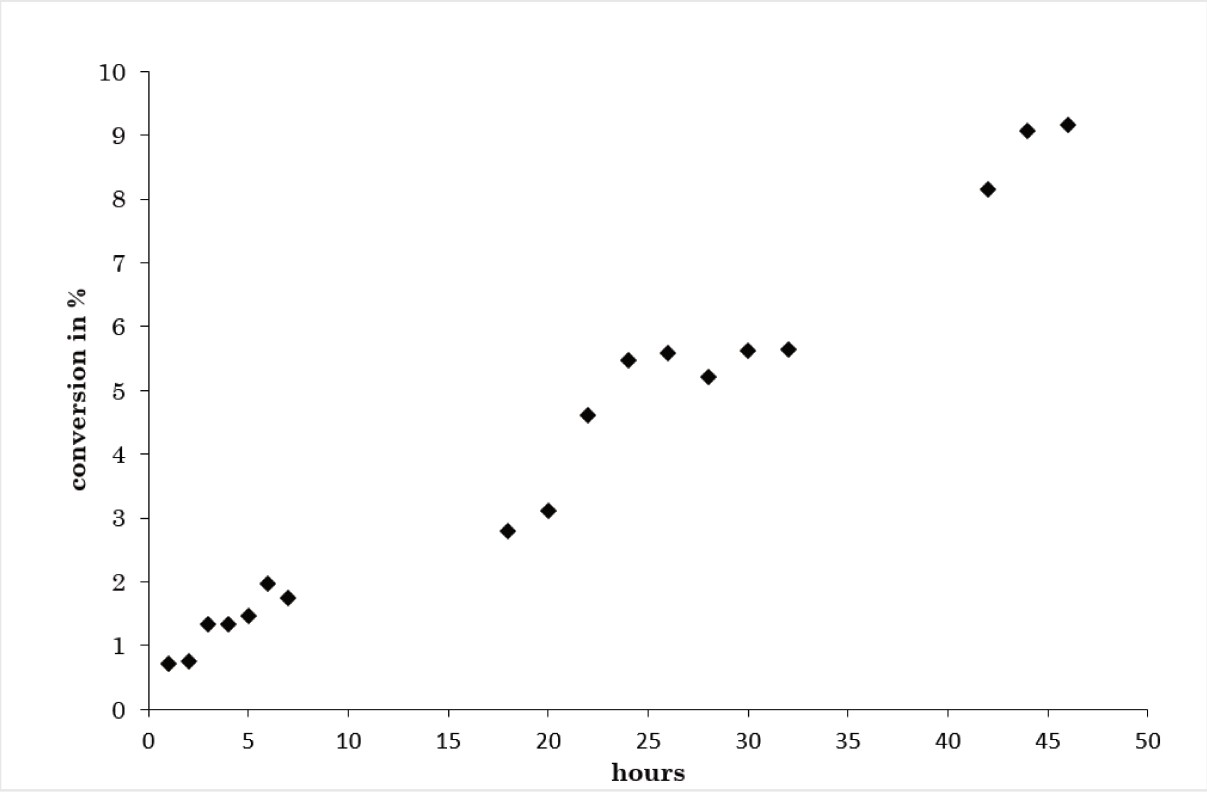


Figure 13: Increasing conversion over time during biotransformation using BT3 protocol.

4.1.2 Reaction characterizations

A good reaction setup is vital but it is beneficial to determine relevant reaction parameters before finding new variants by mutagenesis. Therefore, various aspects were investigated e.g. which products can hypothetically be expected, trace products after improvement of the setup, the enantioselectivity of the reaction, as well as revisiting negative controls and enzyme variants from previous studies.

4.1.2.1 Concept of possible products

Based on preliminary studies (see chapter 2.2), literature research[45,71,74,175] and the principals of standard organic chemistry, an overview of possible products out of ß-pinene isomerization was elaborated, as well as their hypothetical-intermediate carbocations and transformation routes (Figure 14). After protonation of the double bond, ß-pinene (**12**) forms an intermediate pinanyl cation. This can be directly quenched by deprotonation to α-pinene (**11**), by nucleophilic attack of water to pinene hydrate (**48**) or cyclization to tricyclene (**47**).

Alternatively, the pinanyl cation can react to *p*-menthenyl cation (alias: α-terpinyl cation, pink arrow) by cleavage of the bridging bond. The *p*-menthenyl cation can be quenched to limonene (**14**), 3-carene (**33**), α-terpineol (**34**) and α-terpinolene (**35**) or undergo a 6,7-hydride shift to terpinen-4-yl cation. The terpinen-4-yl cation can be quenched to form α-terpinolene (**35**), α-terpinene (**36**), γ-terpinene (**37**), terpinen-4-ol (**38**) or rearrange by a second ring-closure to thujyl cation (orange arrow). The thujyl cation can be quenched to thujene (**39**), sabinene hydrate (**40**) or sabinene (**41**).

Another possibility for the pinanyl cation is to undergo a Wagner-Meerwein shift to form fenchyl cation (red arrow), which can in turn be quenched by water to fenchol (**42**).

A further rearrangement of the pinanyl cation is a widening of the four-membered ring to a five-membered ring to yield isobornyl cation (green arrow). The isobornyl cation can be quenched to form bornylene (**44**), isoborneol (**45**), borneol (**46**) or tricyclene (**47**). Another possibility for the isobornyl cation is to undergo a rearrangement to form camphyl cation (blue arrow) which can be deprotonated to ß-camphene (**13**). Alternatively, the isobornyl cation can undergo a 1,2-methyl shift (pink arrow) to form 2,7,7-trimethylbicyclo[2.2.1]heptan-1-yl cation which can rearrange to 2,7,7-trimethylbicyclo[2.2.1]heptan-2-yl cation which can be deprotonated to yield fenchene (**43**).

These products can generally be formed during the biotransformation of ß-pinene with the *Aac*SHC and were assessed analytically as potential products.

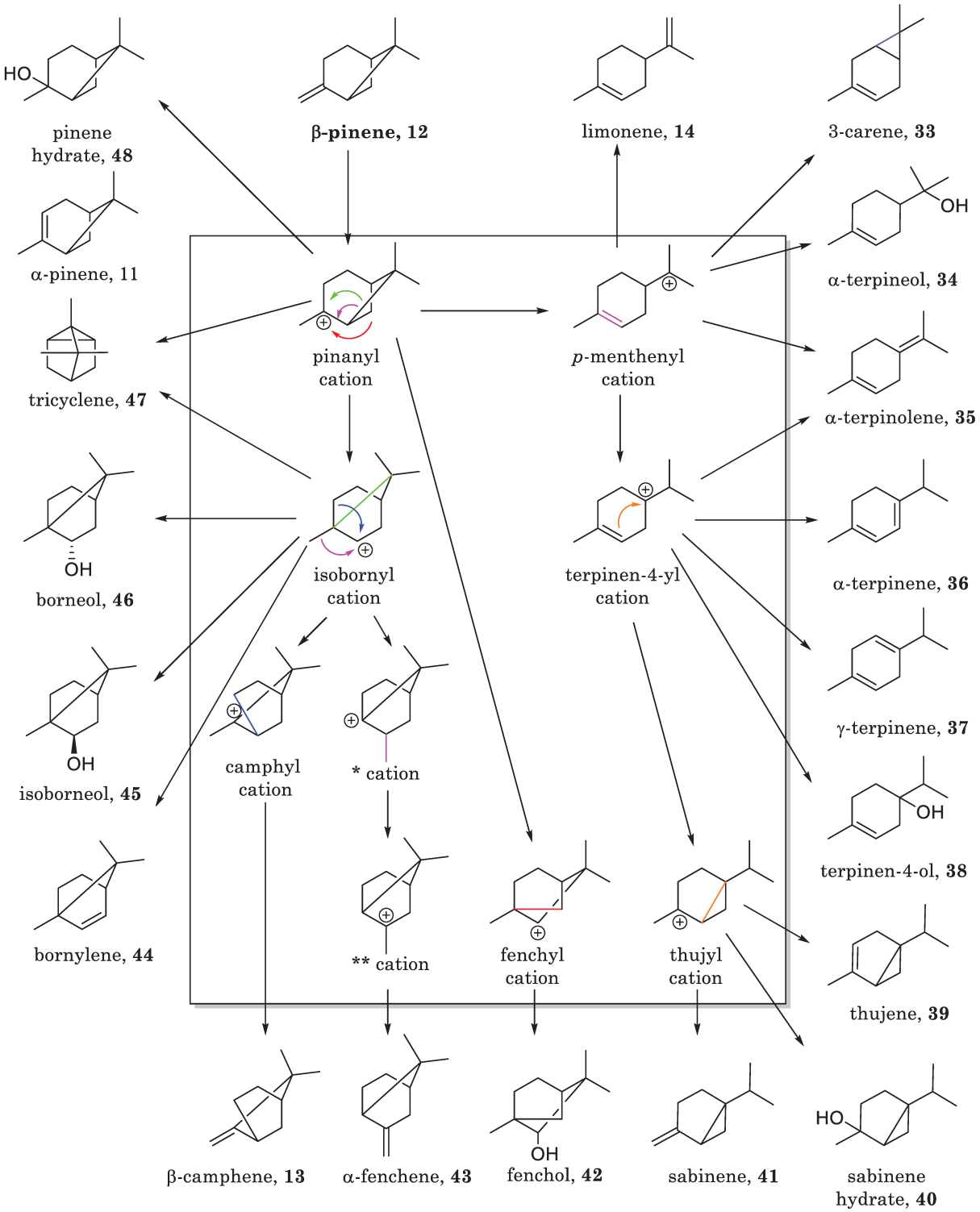


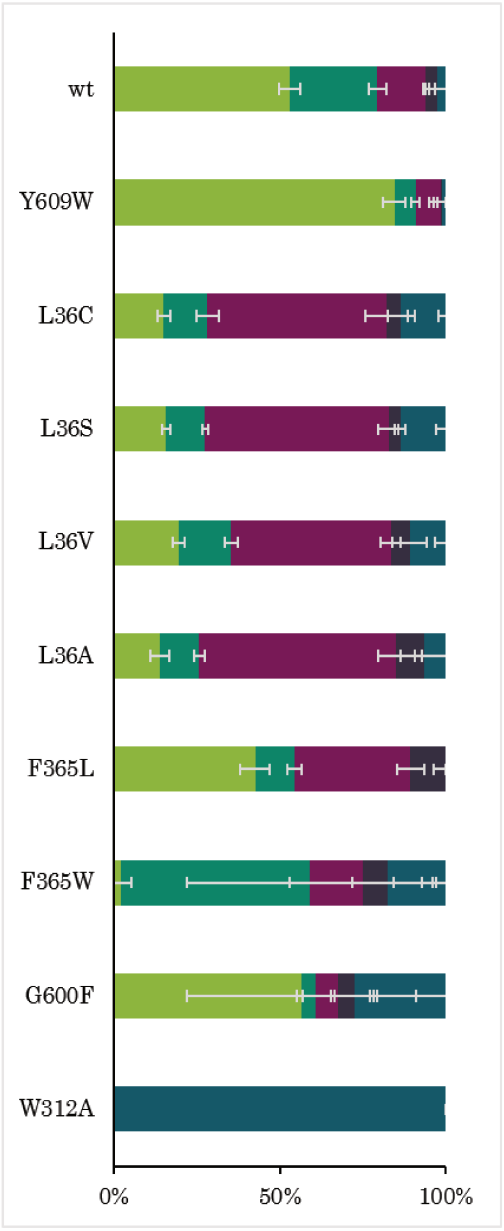
Figure 14: Possible products out of the acidic isomerization of β-pinene (12) and their hypothetical intermediate carbocations. Cations are shown in the box. Newly formed bonds and their corresponding arrows are color-coded for reasons of clarity.
*** 2,7,7-trimethylbicyclo[2.2.1]heptan-1-yl**
**** 2,7,7-trimethylbicyclo[2.2.1]heptan-2-yl**

4.1.2.2 Re-screening of preliminary studies

One of the first steps in the timeline of this thesis was to re-screen initial hits that had been identified in previous studies (see chapter 2.2). Therefore, wild type (wt), variants Y609W, L36C/S/V/A, F365L/W, G600F and W312A that were found in the initial screening were investigated using BT1 protocol and GC-P1 temperature program at least in biological duplicates and technical triplicates. The results are shown in Table 15.

Table 15: Re-screening of previously identified variants using BT1 protocol and GC-P1 temperature program. Product distribution ratios are given in % and are visualized: ■ α-pinene, ■ β-camphene, ■ limonene, ■ terpinolene, ■ terpineol.

	α-pinene	β-camphene	limonene	terpinolene	α-terpineol	visualized
wt	52.9 ± 3.2	26.5 ± 2.6	14.6 ± 0.9	3.4 ± 3.4	2.6 ± 3.2	
Y609W	84.4 ± 3.4	6.4 ± 1.3	7.5 ± 3.3	0.4 ± 1.3	1.3 ± 3.8	
L36C	14.8 ± 1.9	13.3 ± 3.5	53.7 ± 6.4	4.5 ± 4.1	13.6 ± 2.1	
L36S	15.7 ± 1.2	11.8 ± 0.9	55.2 ± 3.0	3.6 ± 1.6	13.7 ± 3.0	
L36V	19.3 ± 1.8	15.9 ± 2.1	48.1 ± 3.0	5.7 ± 5.1	11.0 ± 3.4	
L36A	13.7 ± 2.7	11.9 ± 1.6	59.4 ± 5.6	8.4 ± 7.0	6.6 ± 7.3	
F365L	42.4 ± 4.3	11.9 ± 2.2	35.0 ± 4.0	10.7 ± 3.7	-	
F365W	2.0 ± 3.3	57.0 ± 37.0	15.9 ± 22.1	7.3 ± 10.4	17.8 ± 15.8	
G600F	58.0 ± 35.4	4.2 ± 5.8	7.0 ± 10.8	5.0 ± 7.2	28.5 ± 23.7	
W312A	-	-	-	-	100	



Variant Y609W was confirmed in having a high selectivity (84.4%) towards α-pinene and is the most selective variant for this product. Product distributions for variants at position L36 were also in agreement with the observed trends in preliminary studies. All variants at position L36 showed a shift in product distribution towards limonene, with L36A being the most selective one (59.4%). Chromatographic spectra of these two variants are shown in Figure 15 and Figure 16 in comparison to the wild type and buffer.

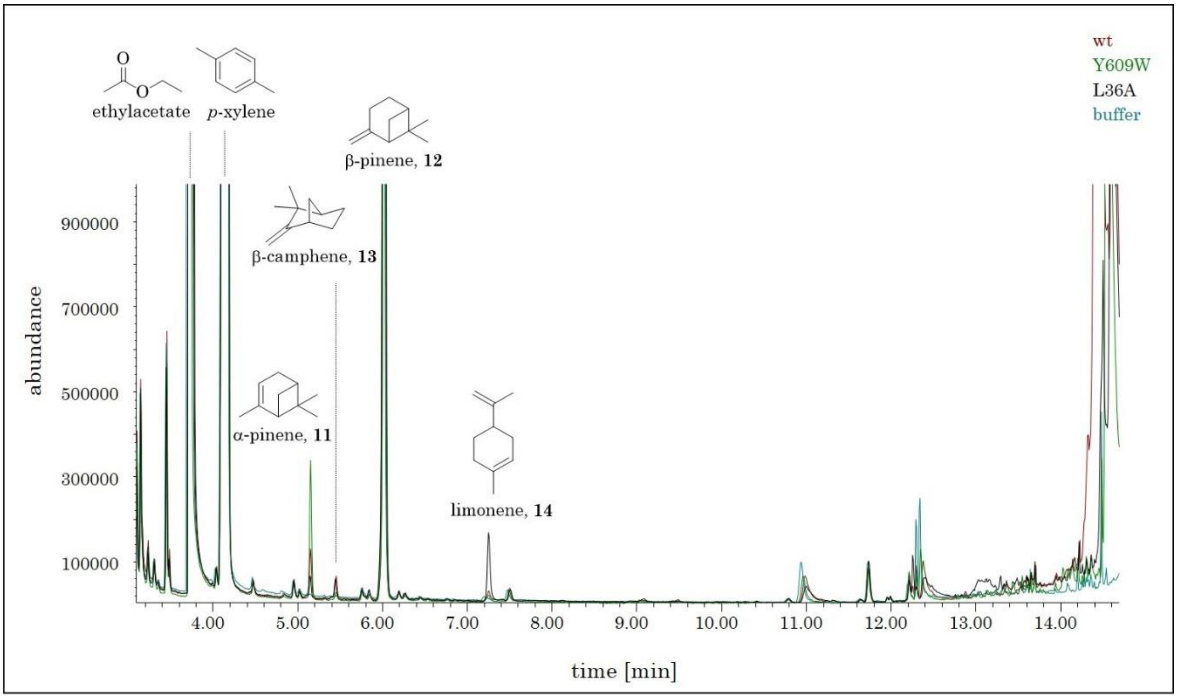


Figure 15: Full length overlaid chromatograms of variants L36A, Y609W and wild type enzyme, compared to the buffer using temperature program GC-P1. The standard *p*-xylene elutes at 4 min and the monoterpenes α-pinene at 5.1 min, β-camphene at 5.4 min, β-pinene at 5.9 min and limonene at 7.2 min.

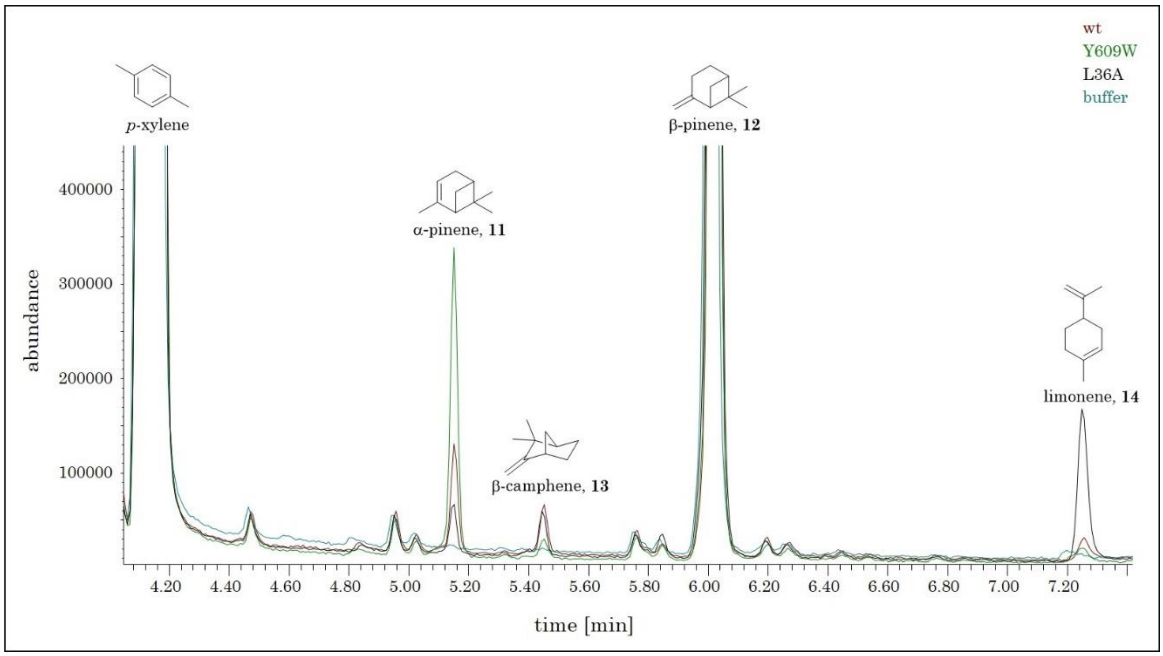


Figure 16: Section of overlaid chromatograms of variants L36A, Y609W and wild type enzyme, compared to the buffer using temperature program GC-P1. The standard *p*-xylene elutes at 4 min and the monoterpenes α-pinene at 5.1 min, β-camphene at 5.4 min, β-pinene at 5.9 min and limonene at 7.2 min.

On the other hand, variants F365W and G600F showed great standard deviations due to their low activity. This low activities in combination with method BT1 and CG-P1 caused the observed large variability. This was the stimulus for ever improving methods (see chapter 4.1.1). With better detection methods and lower evaporation, the peaks in the chromatograms can be much higher, which leads to more reproducible product distributions after integration and thus more accurate data.

Due to their low activity which resulted in high error bars, the variants F365W and G600F were not further considered. Variant W312A appears to be selective in the formation of α-terpineol. However, the activity of the variant was also very low and with method BT1 α-terpineol was the only visible product, which was identified not to be a product of the enzyme catalyzed reaction, based on preliminary studies (see chapter 4.1.2.3). Therefore, it was not further considered until the emergence of highly improved methods.

4.1.2.3 Investigation of negative controls

As mentioned in chapter 4.1, variant D376C was considered to have no activity based on previous studies[109,110,153,158] and was therefore treated as negative control. The variant D376C was initially used for crystallographic studies on *Aac*SHC[84,176] and was tested with different substrates with a similar size to monoterpenes, where it showed no conversion[90]. The advantage of this negative control lies in its high similarity to the native *Aac*SHC. In contrast to an empty vector or a denatured enzyme, reactions on the surface of the enzyme can also be excluded when using D376C as negative control.

However, it was observed that variant D376C forms considerable amounts of α-terpineol when using ß-pinene as substrate. Therefore, Dr. Nico Kress investigated the activity further. He created different variants substituting the DxDD motif with alanines, namely D374A, D376A, D377A and the triple variant D374A_D376A_D377A named "AxAA", thereby completely destructing the protonation machinery. The single variants, as well as the triple mutant show only slight formation of α-terpineol when using BT8 protocol and GC-P3 temperature program, whereas the variant D376C forms significant amounts of α-terpineol. A section of the chromatogram showing the retention time of α-terpineol is shown in Figure 18 and the full chromatogram in Figure 17.

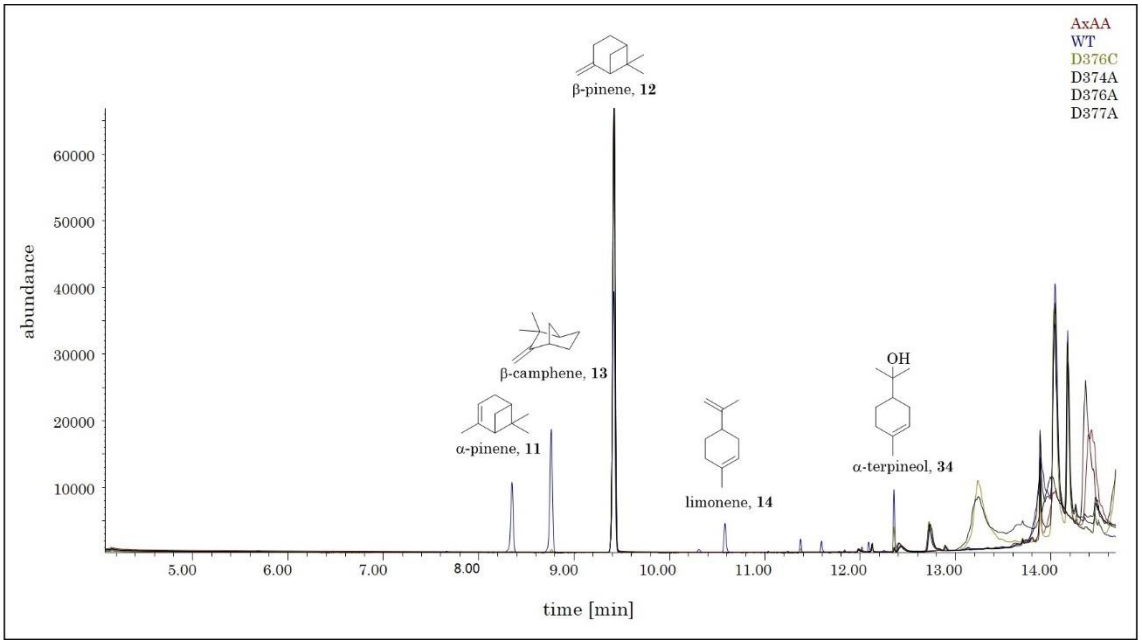


Figure 17: Comparison of the chromatograms of alanine substitutions at the DxDD motif for the formation of α-terpineol. The full chromatogram is depicted showing α-pinene (8.3 min), camphene (8.7 min), β-pinene (9.4 min), limonene (10.5 min) and α-terpineol (12.3 min). The data were obtained using BT8 protocol and GC-P3 temperature program.

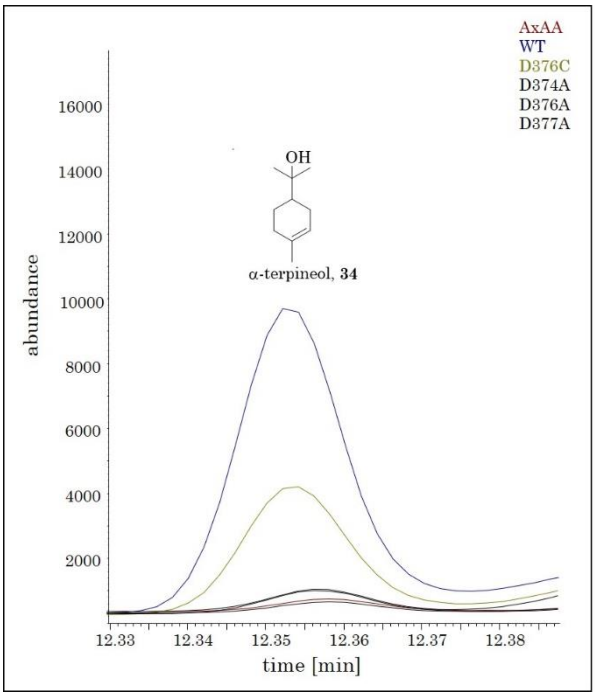


Figure 18: Comparison of the chromatograms of alanine substitutions at the DxDD motif for the formation of α-terpineol. Only the section for the retention time of α-terpineol (12.3 min) is depicted. The data were obtained using BT8 protocol and GC-P3 temperature program.

Consequently, for further experiments AxAA was treated as negative control and for selected previous results a re-evaluation was performed including α-terpineol as *Aac*SHC-derived product and the data are presented as such in the corresponding chapters. This also led to the reconsideration of the variant W312A and it was re-tested using the improved method BT7.

4.1.2.4 Detection of trace products

With the improved protocols BT7 and BT8 it was possible to detect further products that are only formed in trace amounts. Therefore, they were investigated in more detail using BT7 protocol and GC-P4 temperature program. For the wild type, besides the main products α-pinene, ß-camphene, limonene and α-terpineol, the side products α-terpinene, γ-terpinene, α-terpinolene, fenchol, borneol and terpinen-4-ol, as well as five unidentified monoterpenes were found in trace amounts. The unknown products were predicted to be different monoterpenes by mass fragmentation but could not be verified by retention time of a reference sample.

Promising variants were investigated for their side products (for details see Table 38 in chapter 8.7.3). Thereby, attention was drawn to variant W312A which produced seemingly only α-terpineol in the first evaluation using BT1 protocol (see chapter 4.1.2.2) and was reinvestigated with the improved methods after it was shown that α-terpineol is in fact a product of the enzymatic reaction of *Aac*SHC.

In Figure 19 it can be seen that it forms considerable amounts of α-terpineol but also many side products in greater amounts than the wild type. Borneol, fenchol, γ-terpinene, terpinen-4-ol, α-terpinolene and three unidentified monoterpenes are produced by W312A in considerable amounts which were only produced in traces in the wild type. On the other hand, it produces only small amounts of α-pinene, limonene and ß-camphene which are the main products of the wild type.

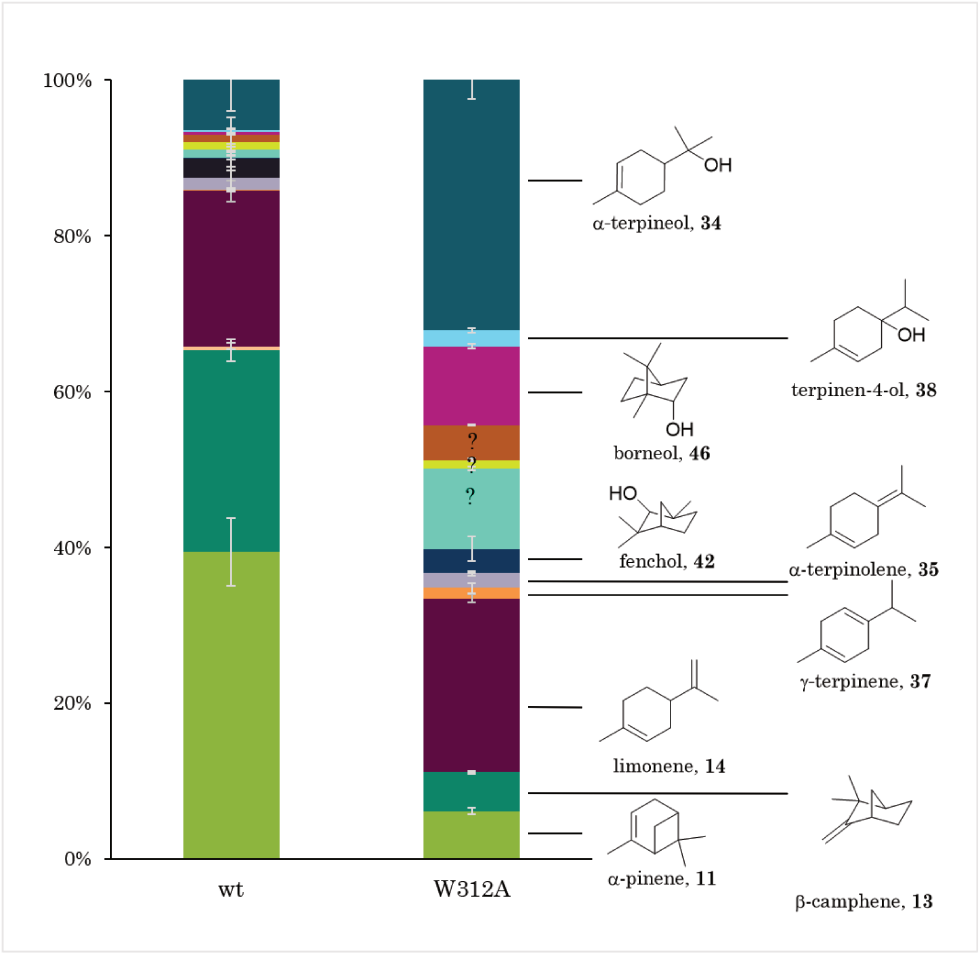


Figure 19: Product distribution of W312A in comparison to the wild type. Identified products are given, unidentified ones are marked with a question mark. All data were obtained using BT7 protocol and temperature program GC-P3.

The main focus of the enzyme engineering part (chapter 4.1.3) is to find selective variants for the main products. For reasons of clarity and since the side products are only detectable with the elaborate methods BT7 and BT8, they are not always shown or otherwise summarized as 'side products'.

4.1.2.5 Determination of enantioselectivity

There are two possible isomers for α-pinene and limonene. Therefore, the enantioselectivity of the products was determined using a wild type reaction (BT6 protocol) and applying it on a Cyclosil-B column with GC-P5 temperature program. The column and temperature program were adapted from literature[177]. The standard (-)-α-pinene (**11B**, Figure 20 and Figure 21) elutes at 13.0 minutes and (+)-α-pinene (**11A**) elutes at 13.25 minutes. The standard (*S*)-limonene (**14B**) elutes at 18.2 minutes and *(R*)-limonene (**14A**) elutes at 18.9 minutes. Mixtures of the enantiomers showed good separation. The products of the conversion using *Aac*SHC, elute solely at 13.3 minutes ((+)-α-pinene, **11A**), 19.95 minutes (ß-camphene, **13**) and 18.95 minutes ((*R*)-limonene, **14A**). Thus, it can be concluded that *Aac*SHC forms (+)-α-pinene (**11A**) and (*R*)-limonene (**14A**) out of (+)-ß-pinene (**12A**) with >99% enantiomeric excess. The corresponding chromatograms are shown in full length in Figure 20 and a section in Figure 21.

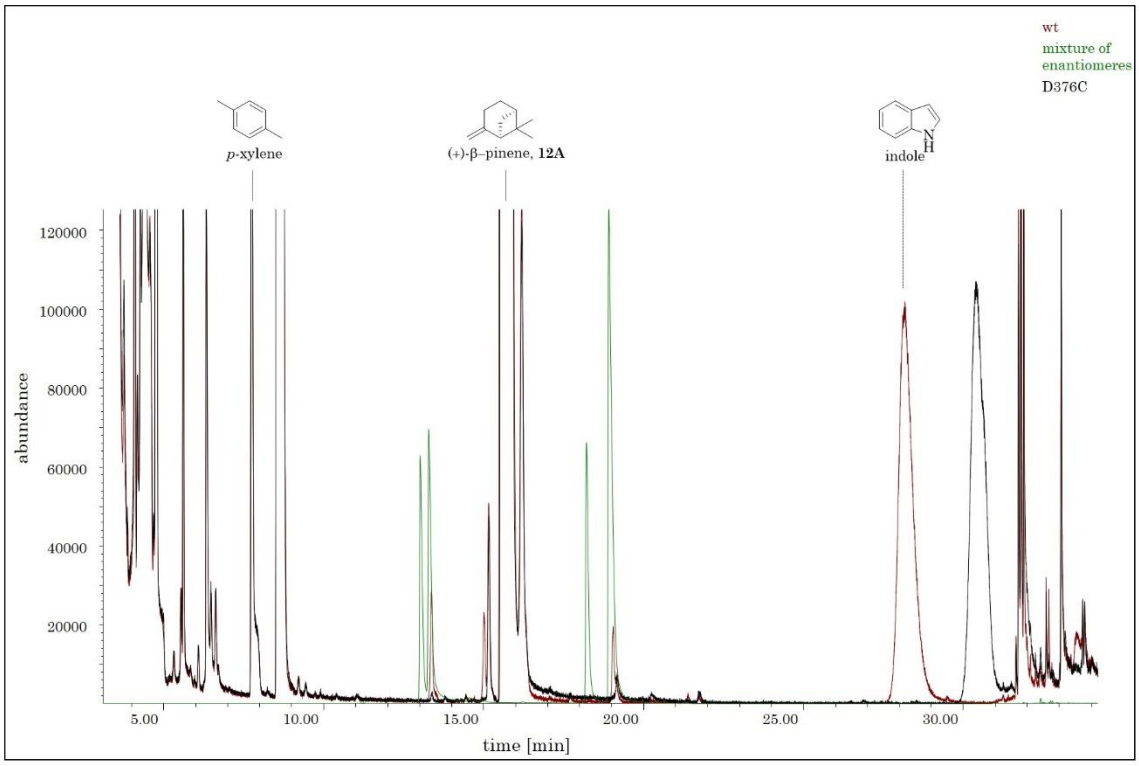


Figure 20: Overlay of full length chromatograms for mixtures of enantiomers of limonene and α-pinene (green) and a biotransformation of (+)-β-pinene (12A, 15.5 min) with the wild type (wt, red) and variant D376C (black) showing a section of the relevant products. The standard (-)-α-pinene (11B) elutes at 13.0 minutes and (+)-α-pinene (11A) elutes at 13.25 minutes. The standard (*S*)-limonene (14B) elutes at 18.2 minutes and *(R*)-limonene (14A) elutes at 18.9 minutes. The temperature program GC-P5 was used.

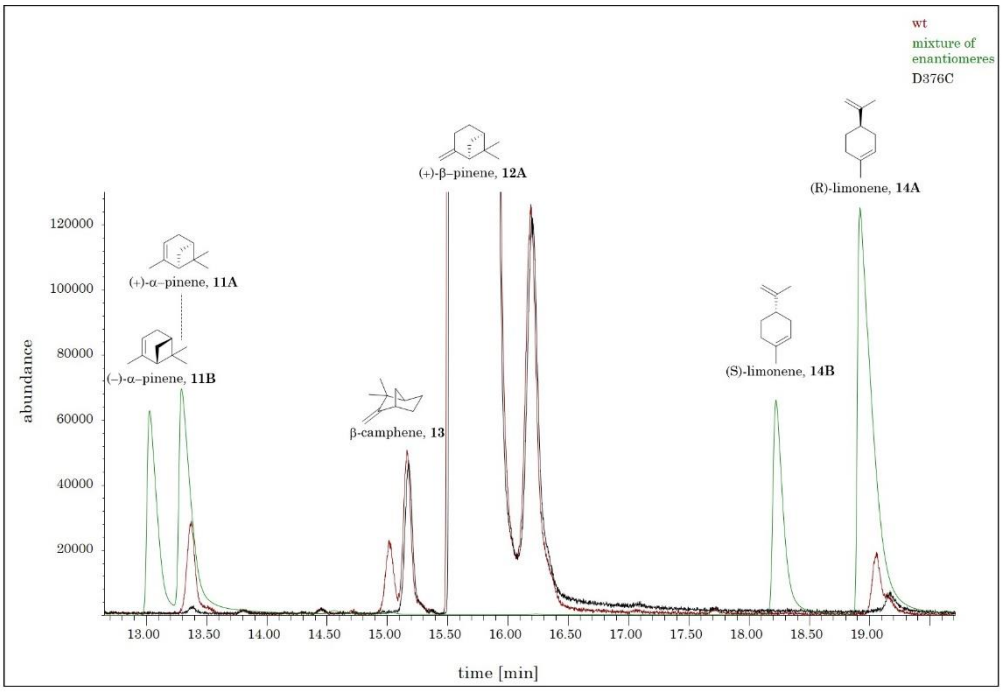


Figure 21: Section of overlaid chromatograms for mixtures of enantiomers of limonene and α-pinene (green) and a biotransformation of (+)-β-pinene (12A, 15.5 min) with the wild type (wt, red) and variant D376C (black) showing a section of the relevant products. The standard (-)-α-pinene (11B) elutes at 13.0 minutes and (+)-α-pinene (11A) elutes at 13.25 minutes. The standard (*S*)-limonene (14B) elutes at 18.2 minutes and *(R)*-limonene (14A) elutes at 18.9 minutes. The temperature program GC-P5 was used.

4.1.3 Enzyme engineering

After thorough investigation and optimization of the reaction, the next goal was to find more selective variants by semi-rational enzyme-design. As stated in chapter 2, there is a high demand for a selective isomerization of pinene into various valuable monoterpenes. In previous chapters it was shown that *Aac*SHC can isomerize (+)-ß-pinene by means of protonation. Further, the reaction was thoroughly investigated and conditions optimized to decrease the evaporation of substrate and products for an accurate and reproducible analytical setup.

Some variants with shifted product distributions towards a main product were already identified in preliminary studies and confirmed (see chapter 4.1.2.2). In a quest to identify variants with further improved selectivities or towards other products, different approaches were chosen. A combinatorial approach focused on a strategy to rationally transfer principals from a so-called 'polar pocket' in a monoterpene synthase[106] into *Aac*SHC. Further, the selected variants were combined to mimic a 'polar pocket'. Another approach was based on the success of the in-house first shell library, which was used to identify variants with increased selectivities in preliminary studies. The idea was to further fine tune the shape of the active site by modifying residues in the second shell. Also, a classical path was followed in saturating the positions of variants with shifted product distributions to explore the variability at these positions and its effect on product selectivities. Forth, a collaboration with Prof. Sílvia Osuna and her group to investigate variants with improved selectivities by computational methods inspired further variants. In an attempt to further vary the sequence space, SHCs from other organisms were investigated for their ability to isomerize (+)-ß-pinene and prompted the transfer of point mutations into *Aac*SHC. Fifth, the most selective and most active variants were combined.

The main focus of this enzyme engineering part was to find selective variants for the main products. For reasons of clarity and since side products are only detectable with the elaborate methods BT7 and BT8, they are not always shown or otherwise summarized as 'side products'.

4.1.3.1 Combinatorial approach

The limonene synthase from *Mentha spicata* is highly selective for the production of S-limonene (94%). In a mutational study Xu et al[106] identified 8 polar residues (C321, W324, N345, T349, S454, M458, H579 and Y573) that are responsible for the stabilization of the intermediate *p*-menthenyl cation. They suggested the term 'polar pocket' for it. A similar structural arrangement can be found in related species but mostly conserved among the angiosperms clade.[106]

The idea was to transfer this 'polar pocket' to *Aac*SHC in a rational approach, to increase the selectivity in the formation of limonene. Therefore, the crystal structure (PDB:2ONG)[73] of the limonene synthase from *M. spicata* co-crystallized with 2-fluorogeranyl pyrophosphate was analyzed using PyMOL. It appeared that residues forming the polar pocket are all in a distance of 3.9-6.3 Å around the terminal side of the 2-fluorogeranyl pyrophosphate, which is supposed to be equivalent to the location of the cation in the p-menthenyl cation (Figure 22A, left).

To transfer these results to *Aac*SHC, ß-pinene was docked into the active site of *Aac*SHC using the YASARA software package. The enzyme was energy minimized and a simulation cell was chosen with a size of 12 Å around the acid of D376. Further, the *p*-menthenyl cation was modelled using the protein structure software package PyMOL, taking into account its formation by ring opening of the docked ß-pinene (see chapter 4.1.2.1). Next, all residues in a distance of ~6 Å around the cation in *p*-menthenyl cation or the corresponding position of the cation in pinanyl cation were selected and rationally analyzed on their ability to stabilize or destabilize the cation (Figure 22B) by either electronic effects or cation-π interactions.

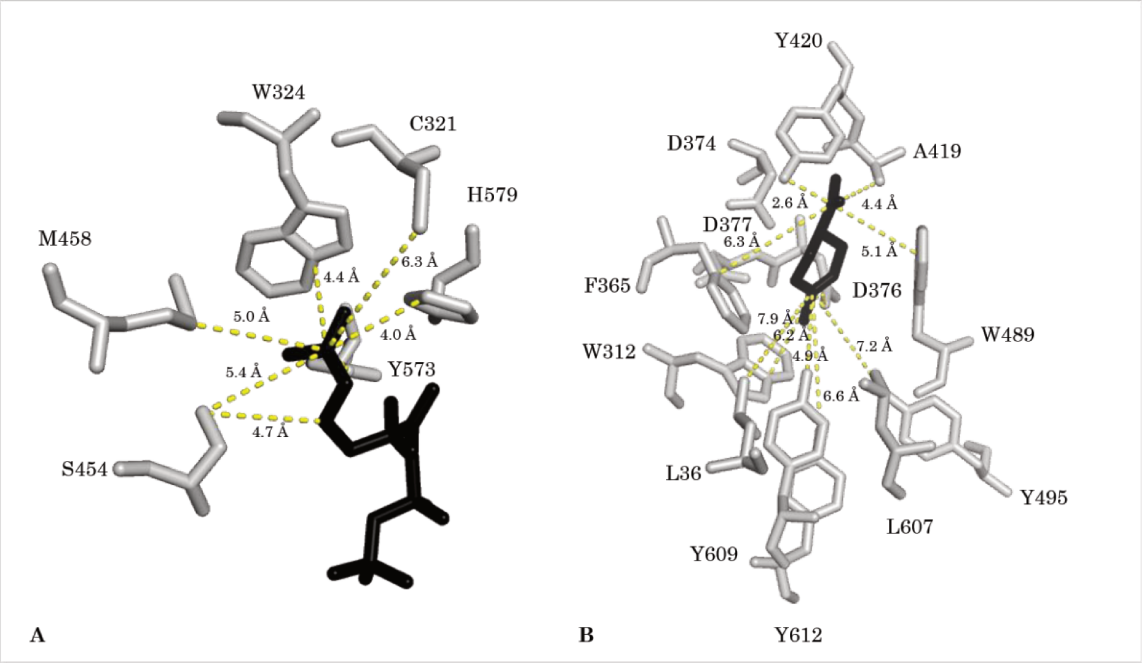


Figure 22: Comparison of the stabilizing residues in *S*-limonene synthase from *M. spicata* identified by Xu et al[106] (A, left) and residues in *Aac*SHC in a similar distance around a docked *p*-menthenyl cation (B, right). For residues in *Aac*SHC (right, B) distances from either a heteroatom or the plane of an aromat are depicted to the side of the cation in *p*-menthenyl cation or the corresponding position in pinanyl cation.

The identified residues were W312, F365, D374, D377, A419, Y420, W489, Y495, L607, Y609 and Y612. The residues Y420, W489 and Y612 were believed to have a stabilizing effect by cation-π interactions and were therefore not considered for mutagenesis. The residues W312, F365 and Y495 seem to favor the pinanyl cation by cation-π interactions and were therefore mutated to destabilizing amino acids such as W312L, W312I, F365L, Y495I and Y609I. The residues A419 and L607 were in a good position to stabilize the *p*-menthenyl cation but had no directing effect. Therefore, they were changed to polar residues A419S, A419D, L607S and L607Y due to sterical reasons. Polar residues were chosen for these positions since aromatic side chains were considered to be too bulky. The positions D374 and D377 are part of the protonation machinery and were therefore left unchanged. A comparison of the believed stabilizing or destabilizing effects of different amino acids on either *p*-menthenyl or pinanly cation is depicted in Figure 23.

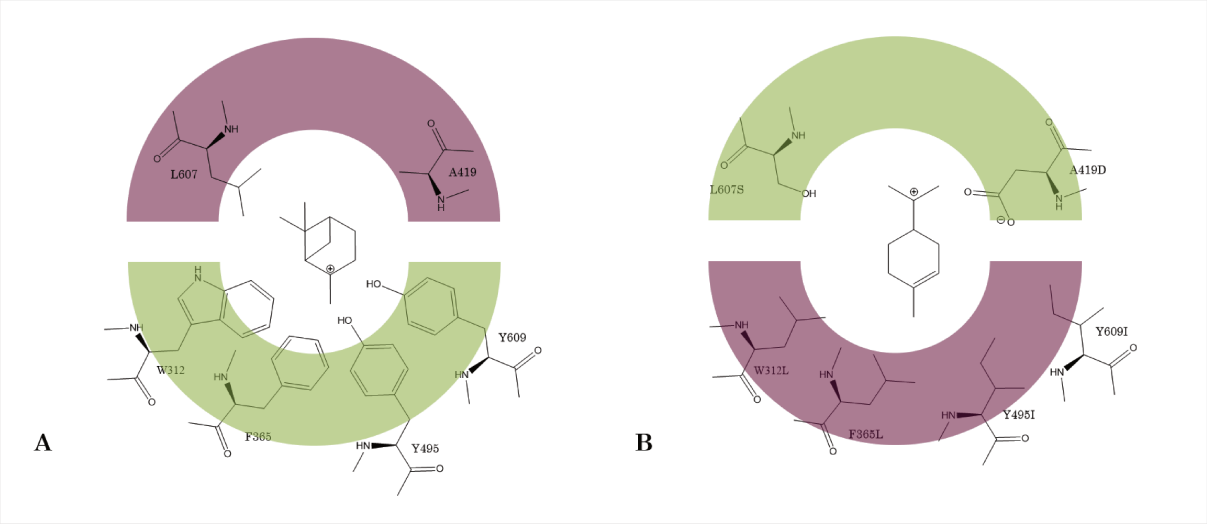


Figure 23: Schematic drawing of hypothesized stabilizing (green) or destabilizing (purple) effects of different amino acids in the active site of *Aac*SHC on either *p*-menthenyl or pinanly cation are shown.
A: The pinanly cation is thought to be stabilized by aromatic side chains facing the cationic charge (green) and non-stabilizing residues on the opposite site of the cation.
B: The *p*-menthenyl cation is thought to be stabilized by introduced polar residues facing the cationic charge (green). Residues at the other side of the cation, when changed to aliphatic amino acids (purple), are thought to have no stabilizing effect.

Figure 23 **A** shows a situation that is thought to stabilize the pinanyl cation. This situation can be found in the wild type which produces α-pinene as main product. However, the hypothesis is to introduce mutations that flip the cation stabilization to the opposite site (Figure 23, **B**) which leads to a stabilization of the *p*-menthenyl cation, which results in a higher formation of limonene.

A summary of the wild type amino acids (grey) and the introduced variants (black) are depicted in Figure 24.

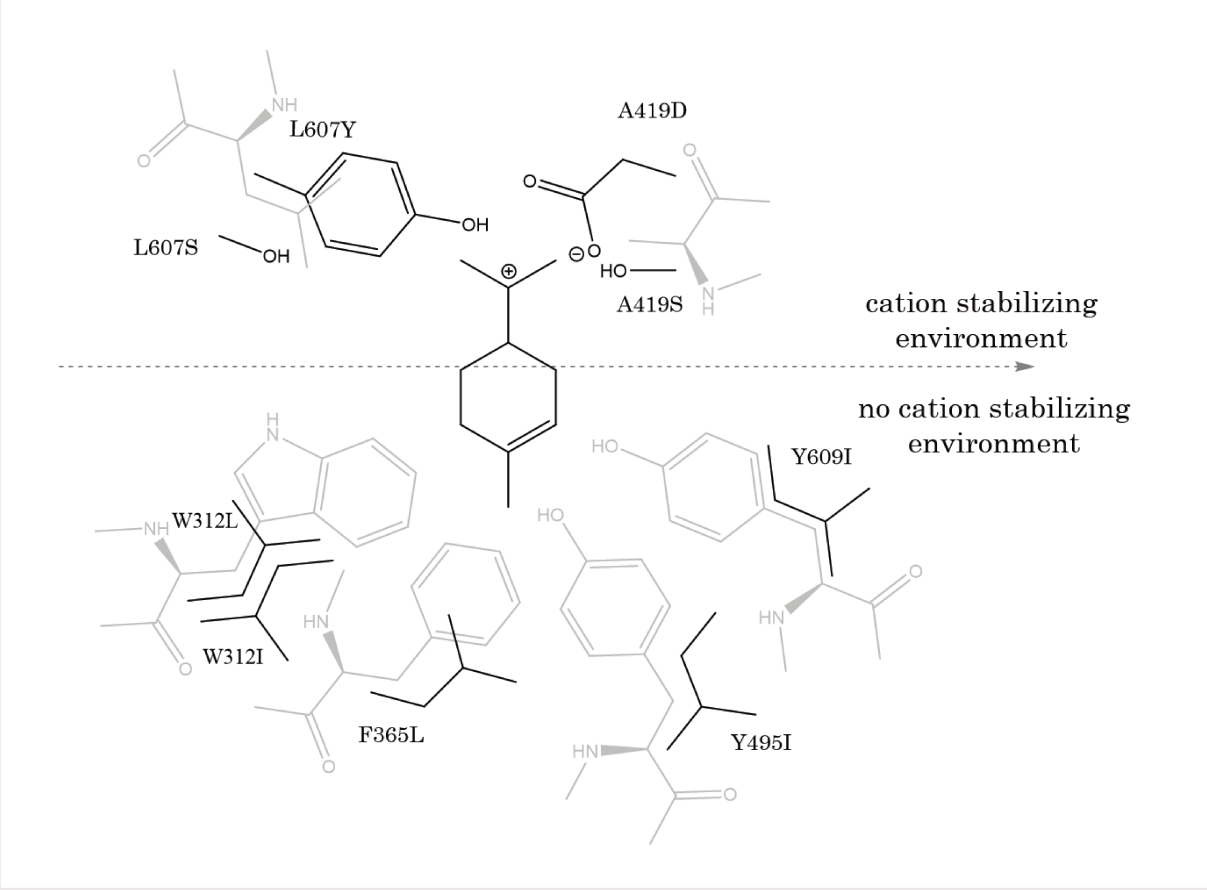


Figure 24: Graphical depiction of all introduced single and multiple mutations in the active site of *Aac*SHC in one picture. Original amino acids are depicted in grey and changes in black.

The variants were created as single variants and in various combinations, including substitutions of all six residues, to mimic a complete polar pocket. They were also combined with L36A, the best variant for limonene formation.

The creation of variants and their testing was gratefully executed by Josip Tulumovic as part of his bachelor thesis[168] under the authors supervision. The data were obtained using BT5 protocol and measured using GC-P2 temperature program. The selectivities of the combinations were lower or almost as good as L36A. The single variants L607Y, A419D and W312I were inactive. Surprisingly, most combinations showed no activity, or showed lower selectivities than the single mutants alone[168]. However, all active single variants showed a higher selectivity towards limonene than the wild type (Figure 25).

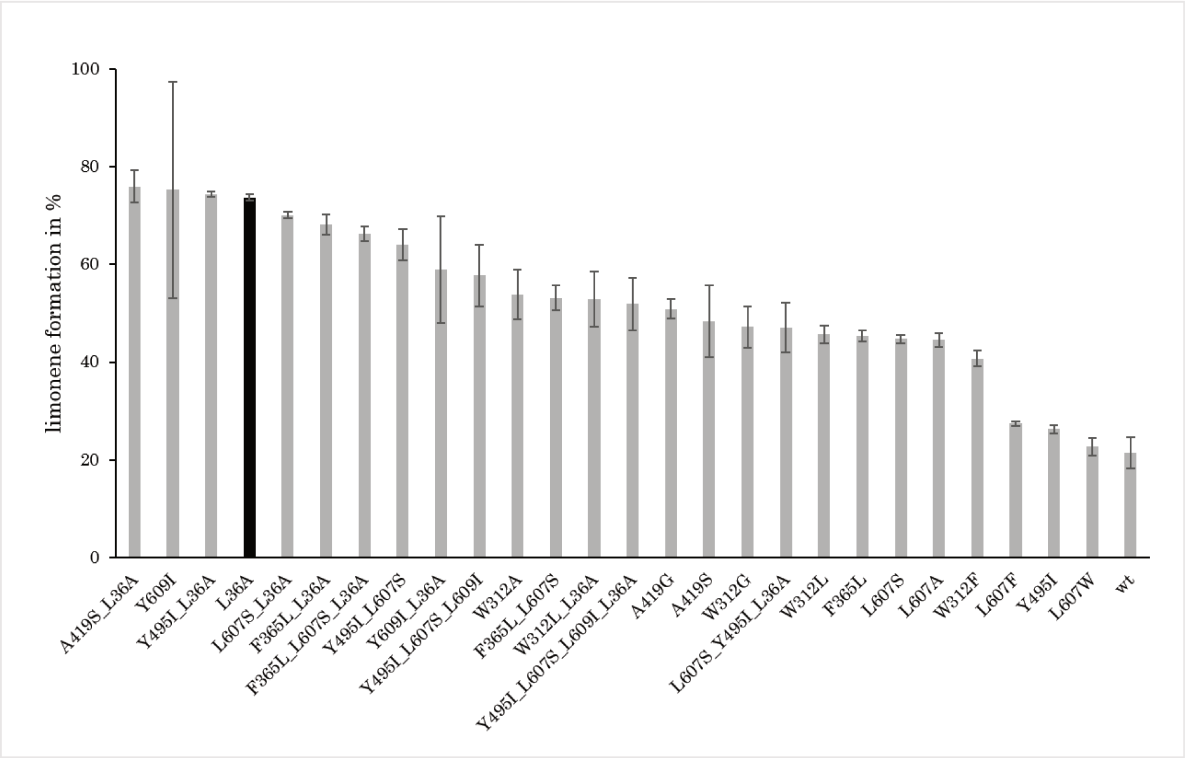


Figure 25: Selectivities for the formation of limonene from (+)-β-pinene for different *Aac*SHC variants obtained by the combinatorial approach. The so far best variant L36A is highlighted in black. In cooperation with Josip Tulumovic[168].

The double variants A419S_L36A and Y495I_L36A showed slightly higher selectivities than L36A alone. However, since the difference is within the normal variation, the effect might mostly be due to L36A. The high selectivity of Y609I towards limonene was surprising sinceY609W shows a high selectivity towards α-pinene (in cooperation with Josip Tulumovic[168]).

4.1.3.2 Second shell library

A library in the second shell of the enzyme's active site was created to identify new variants with increased selectivities after the success of the first shell library[90] in previous studies[153,157,158]. The intention was to introduce more subtle changes in the active site by changing residues in the second shell. This would alter the first shell only slightly, addressing the fine differences between intermediate carbocations. The affected positions are highlighted in Figure 26. Amino acid mutations in the first shell are marked in yellow and areas representing changes in the second shell in blue.

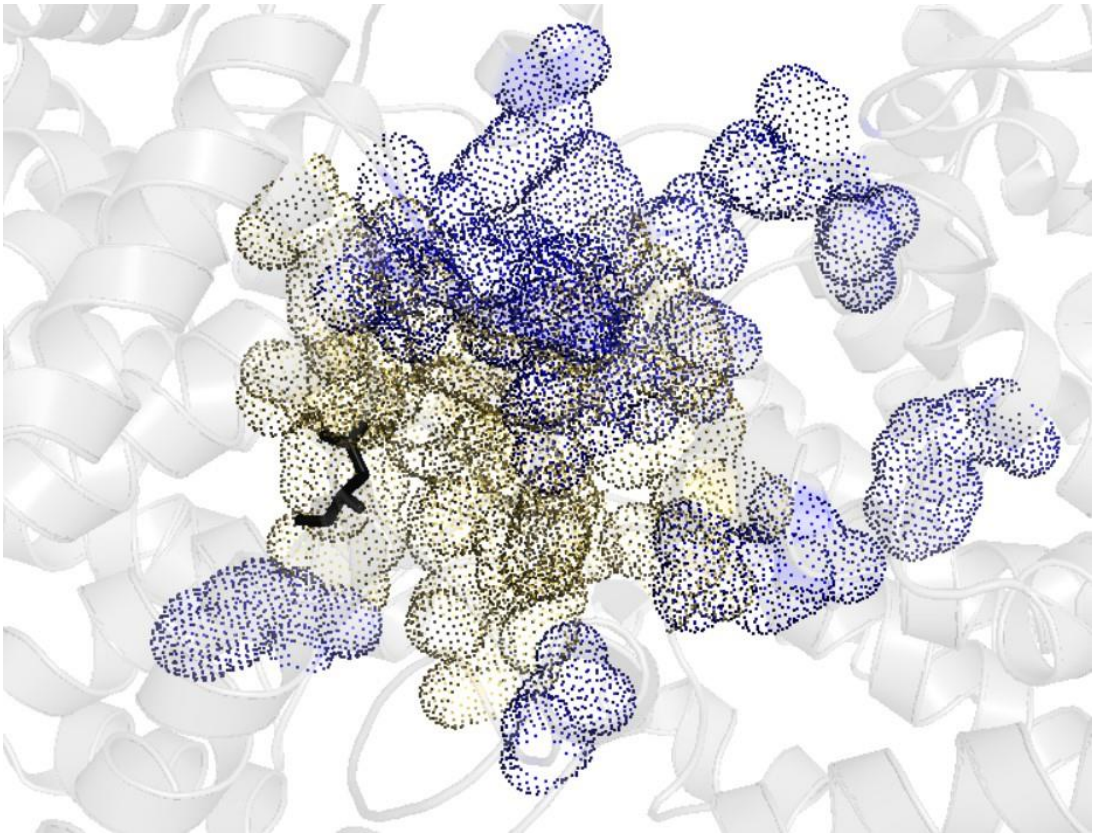


Figure 26: Structural depiction (PDB:1UMP[85]) of *Aac*SHC in part to compare first and second shell libraries. Residues mutated in the first shell library are highlighted in yellow, modified second shell residues are depicted in blue. The catalytic active aspartic acid residue 376 is colored in black. The greatest distances in the second shell from D376 are 27 Å to D89, whereas the greatest distances in the first shell library from D376 are 18 Å to I261.

Residues in the second shell of *Aac*SHC were already identified by Dr. Silvia Fademrecht in previous studies[178,179]. She classified all residues that are in contact with the substrate as first shell and all interacting residues within a 5 Å radius of the substrate as second shell. Moreover, she performed a thorough analysis of conserved residues in both shells[178,179].

The focus of the first shell library was to keep the hydrophobicity in the active site constant. Therefore, only hydrophobic substitutions were introduced. In contrast the aim of the second shell library was to maintain the functional groups while introducing changes in the size of the residues. This should limit the perturbations within the enzyme and therefore the number of inactive variants. Additionally, to increase the chance of a

beneficial outcome, conserved amino acids within the triterpene cyclase family and variants that were previously identified to be interesting were included, as well.

The second shell was determined to comprise 71 residues. If each of these 71 residues were to be mutated into one smaller, one larger and one conserved amino acid, this library would be 213 mutants large. Therefore, the number of variants was reduced in a rational approach using the protein structure software packages PyMOL and YASARA. The bulky limonene was chosen as a model monoterpene and docked into an energy-minimized structural model of *Aac*SHC. Further, the structure software package PyMOL was used to identify all residues within a 5 Å radius of the monoterpene. Next, all residues within a 5 Å radius of these contacting amino acids were aligned with residues classified as second shell. The identified residues, as well as chosen variants can be found in Table 16.

Table 16: Rationally-reduced library by identifying 1st shell residues within a 5 Å radius of limonene, residues contacting them in a 5 Å radius (2nd shell) and suggested variants. The variants were either smaller or larger while maintaining the functional group or contained conserved or interesting amino acid functional groups (see comments). Positions that were identified multiple times are shown in brackets. Variants with an asterisk could not be created. In cooperation with Julius Knerr[180].

1st shell residues contacting limonene within 5 Å	2nd shell residues within 5 Å of 1st shell residues	chosen variants	comments
W169	R171	R171K R171G	
	D442	D442E D442N	D442N reduces the temperature optimum of the *Aac*SHC[88]
	P443	P443G P443Y P443H	P443H and Y conserved in OSCs[179]
	R488	R488K R488S	R488S conserved in OSCs[179]
I261	N369	N369D N369Q	N369D conserved in OSCs[179]
P263	W133	W133L W133F W133G	W133L conserved in SHCs[179]
F437	L80	L80A	
	D89	D89E	
	K211	K211Y* K211R* K211A*	K211Y conserved in OSCs[179]
	(D442)		
W489	(D442)		
	S445	S445T S445Y	
	G487	G487A G487L	
	(R488)		
	G490	G490A G490L	
	Y540	Y450S Y450T*	
	T597	T597I T597S T597Y	T597I conserved in OSCs[179]
	T599	T599G T599S T599Y	T599G conserved in OSCs[179]
G600	(R488)		
	(T599)		

	Y606	Y606M Y606S Y606T	Y606M conserved in OSCs[179]
F601	(T599)		
	(Y606)		
F605	S38	S38T S38Y	
	N39	N39P N39Q	N39P conserved in OSC[179]
	(L80)		
	(T599)		

The majority (34) of the chosen variants were gratefully created and tested by Julius Knerr in his Bachelor thesis[180] under the authors supervision. Of the tested variants 94% showed activity[180], which highlights the high plasticity of *Aac*SHC. The most interesting ones were reproduced for this work. The data were obtained using BT5 protocol and measured using GC-P2 temperature program.

The most interesting results are variants S38Y, T599Y and D442N. The variant S38Y exhibited the highest selectivity within the library in the formation of α-pinene with 66.3%, T599Y produced limonene with a selectivity of 48.3%. However, for α-pinene and limonene variants with higher selectivities were already found in the first shell library before (see chapter 4.1.2.2). D442N showed the highest selectivity for ß-camphene with 58.5% and is therefore superior to any first shell variant. A direct comparison of each of the best variants from the libraries can be found in Figure 27.

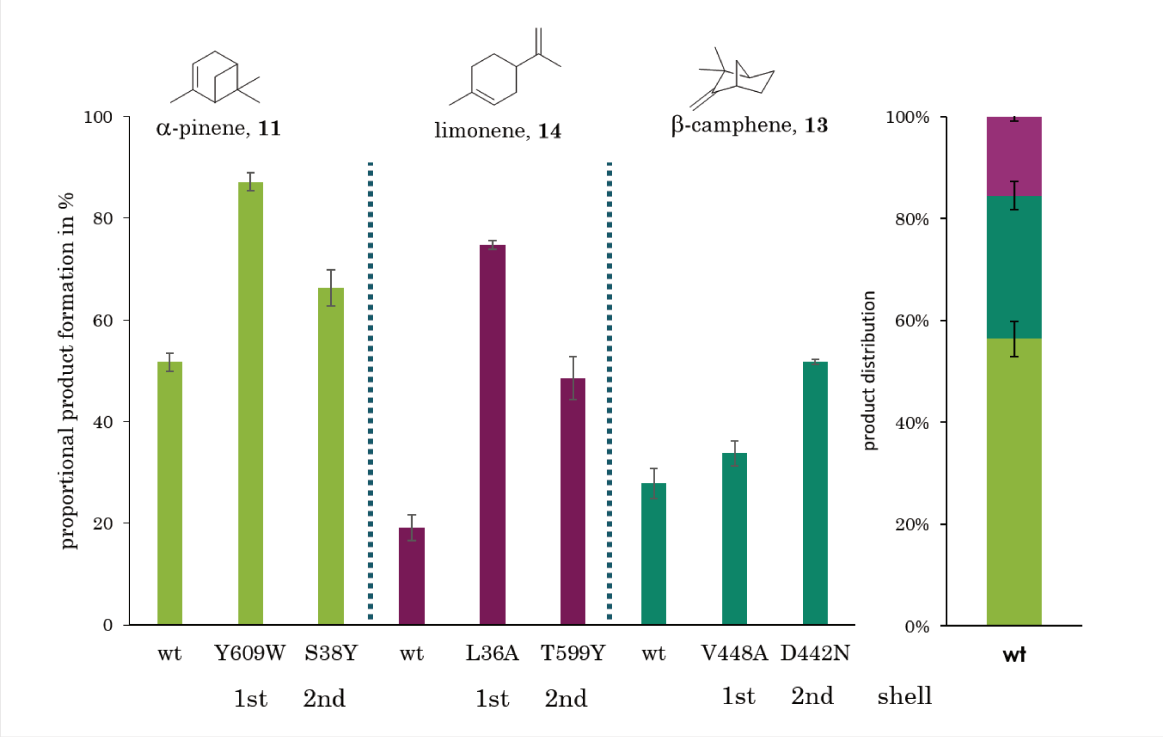


Figure 27: Comparison of the most selective variants of first and second shell libraries for ■ α-pinene, ■ β-camphene and ■ limonene, using β-pinene as substrate. The wild type (wt) product distribution is shown for comparison (in cooperation with Julius Knerr[180]). The data were obtained using BT5 protocol and measured using GC-P2 temperature program.

4.1.3.3 Saturation mutagenesis

Since at positions L36, D442 and Y609 the most selective variants for the products limonene (L36A), α-pinene (Y609W) and ß-camphene (D442N) were found, the positions were further investigated by saturation mutagenesis to exploit their full potential. Especially, the already high selectivities for limonene of the variants L36A/C/S/V (48-59%, see chapter 4.1.2.2) prompted the saturation of position L36. All data were obtained using BT7 protocol and GC-P4 temperature program. The results are shown in Figure 28-Figure 30 and listed in Table 39-41 in the appendix.

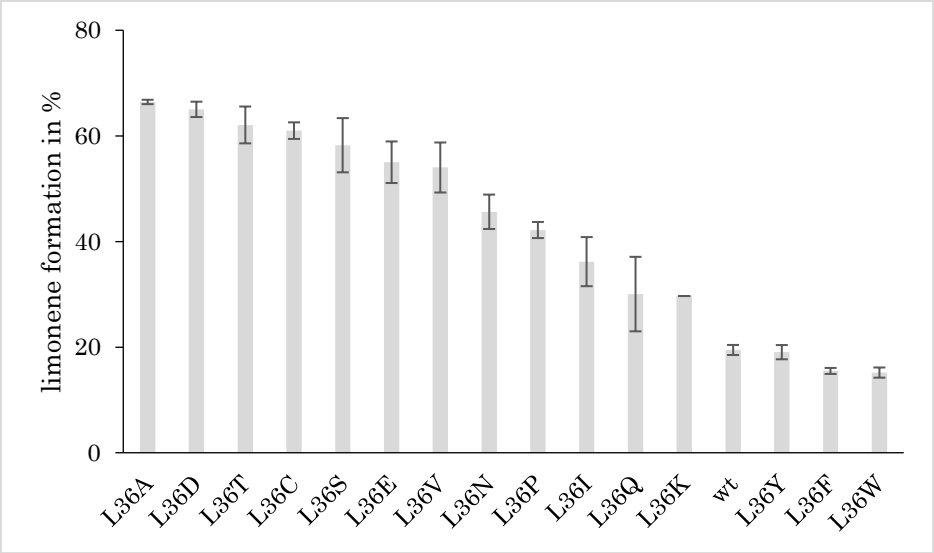


Figure 28: Results of the saturation mutagenesis for position L36, sorted by product selectivity for limonene using β-pinene as substrate. The variants L36G/H/M/R showed no activity and are therefore not depicted. All data were obtained using BT7 protocol and GC-P4 temperature program.

Variant L36A still shows the highest selectivity for limonene formation. Although the variants L36D/T/C/S/E display similar high selectivities. Only variants L36Y/F/W show lower selectivities than the wild type. The variants L36G/H/M/R showed no activity.

For the formation of α-pinene from ß-pinene using saturation mutagenesis at position Y609 a different picture emerges. Here, only Y609W presents a high selectivity for α-pinene, whereas all other active variants at this position show formations similar to the wild type.

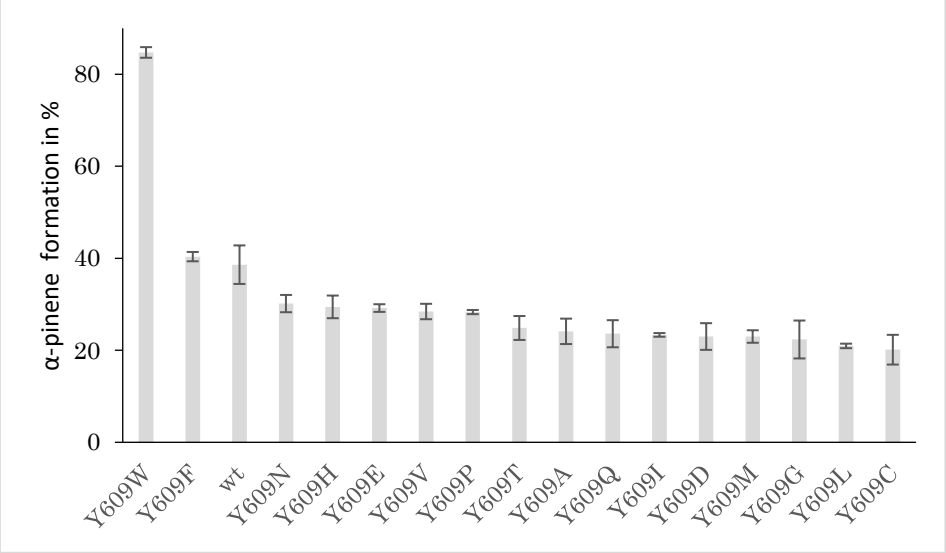


Figure 29: Results of the saturation for position Y609, sorted by product selectivity for α-pinene using β-pinene as substrate. The variants Y609K/R/S showed no activity and are therefore not depicted. All data were obtained using BT7 protocol and GC-P4 temperature program.

For the saturation mutagenesis at position D442 the variants D442P/C were found that are slightly more selective for the formation of ß-camphene than the so far best variant D442N. Additionally, all other active variants at this position are more selective than the wild type.

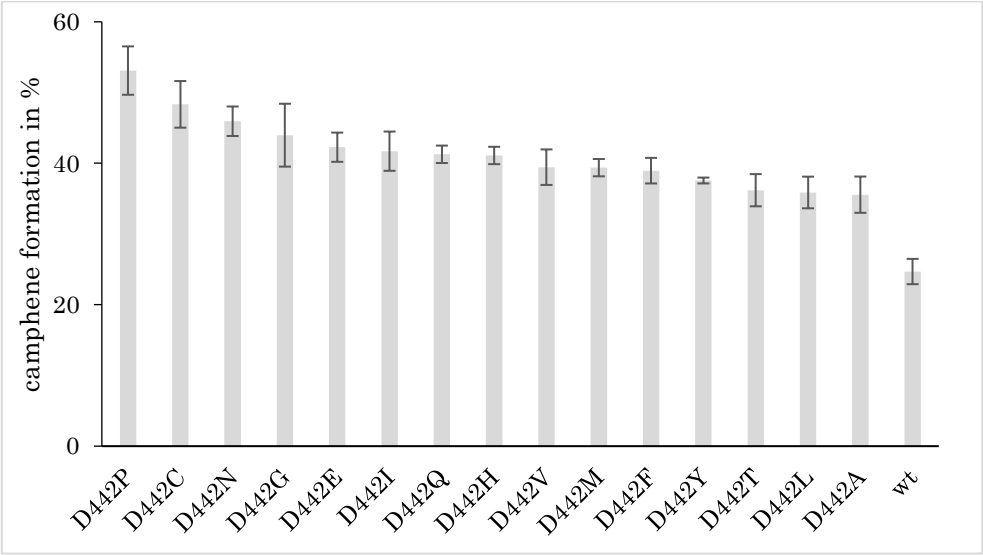


Figure 30: Results of the saturation for position D442, sorted by product selectivity for β-camphene using β-pinene as substrate. The variants D442K/R/S/W showed no activity and are therefore not depicted. All data were obtained using BT7 protocol and GC-P4 temperature program.

A comparison of the chromatograms of D442P, D442N, AxAA, wild type and buffer are shown in Figure 31 and Figure 32.

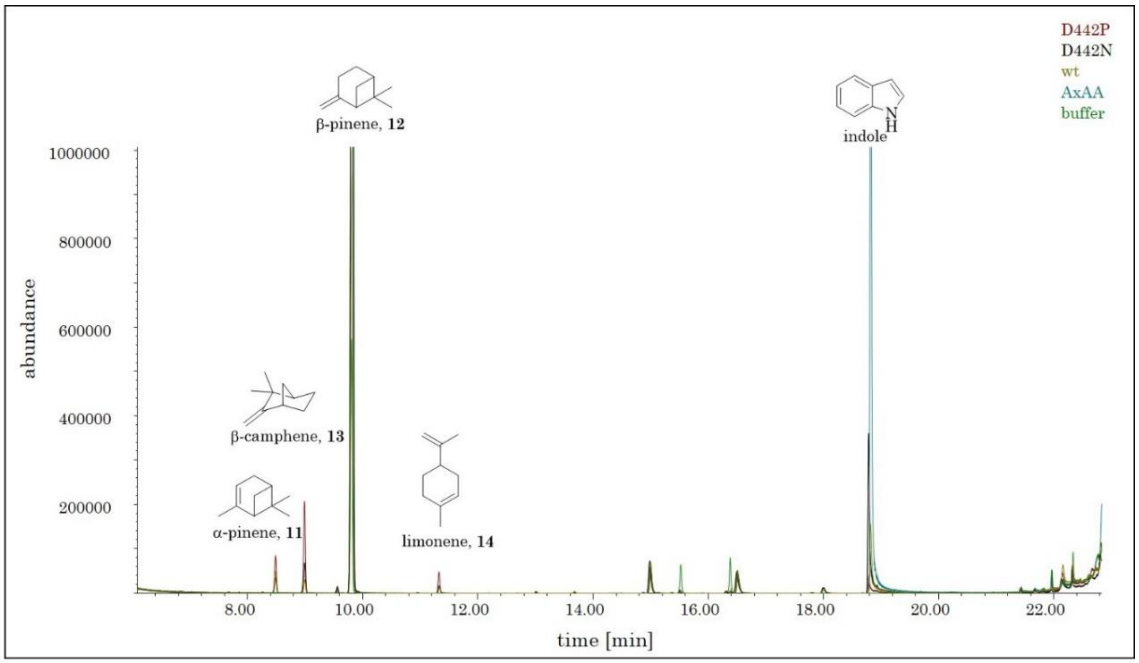


Figure 31: Full length of overlaid chromatograms of variants D442P, D442N, AxAA, wild type and buffer using temperature program GC-P4. The monoterpenes α-pinene (8.4 min), β-camphene (8.9 min), β-pinene (9.7 min) and limonene (11.3 min) are highlighted.

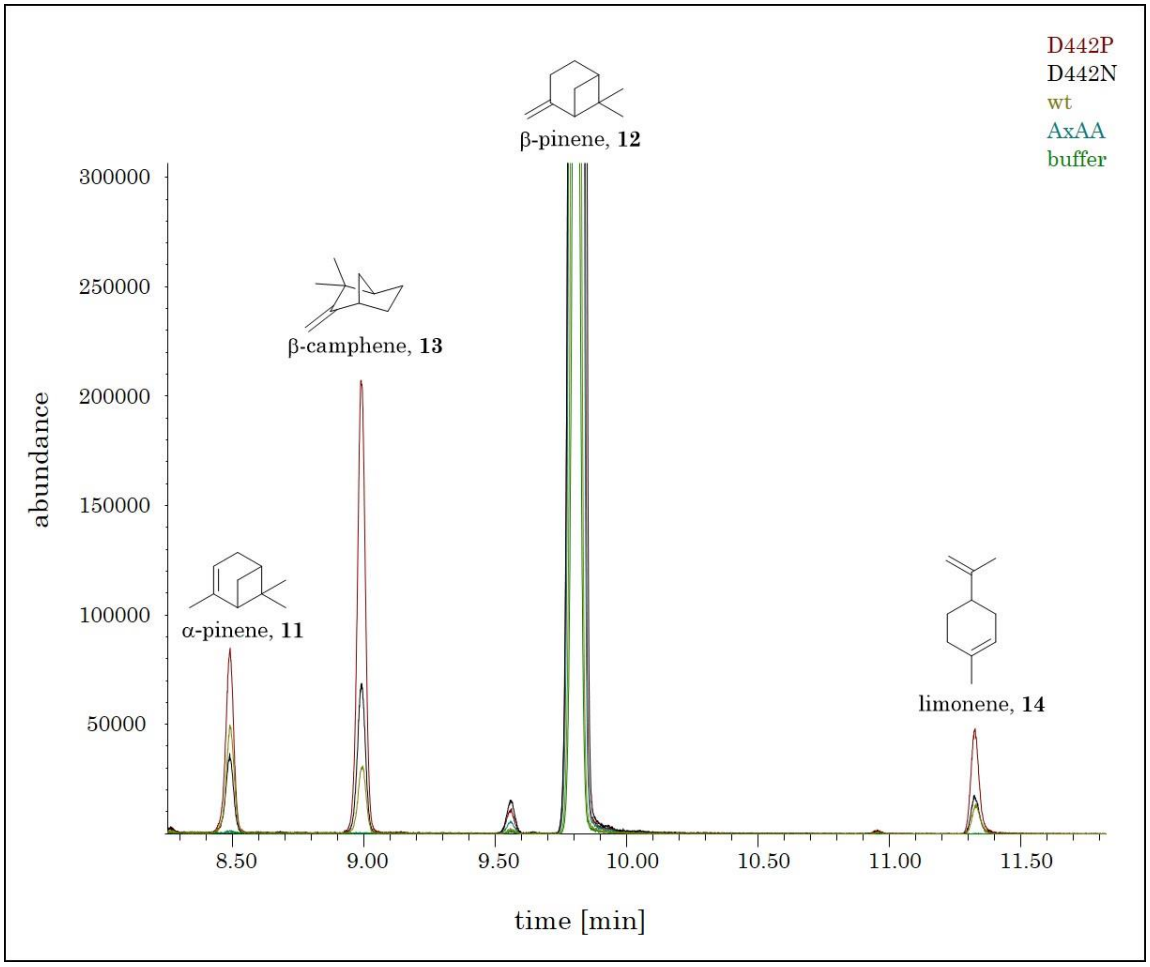


Figure 32: Section of overlaid chromatograms of variants D442P, D442N, AxAA, wild type and buffer using temperature program GC-P4. The monoterpenes α-pinene (8.4 min), β-camphene (8.9 min), β-pinene (9.7 min) and limonene (11.3 min) are highlighted.

4.1.3.4 In-silico studies

In a collaboration with Prof. Sílvia Osuna and Christian Curado from the University of Girona in Spain, the influence of variants L36A, Y609W and D442N on the product distribution was investigated. A computational model of *Aac*SHC was set up by Christian Curado based on the crystal structure PDB:1UMP[85] embedded in a DMPC membrane using the web-server Charmm-GUI. The positioning of the different intermediate carbocations (pinanyl, *p*-menthenyl, and camphene carbocation) was considered to be crucial for studying the reaction. Therefore, they were investigated by Christian Curado as ligands in the active site of the enzyme and their structures parametrized by means of the ANTECHAMBER module of AMBER using the General Amber Force Field (GAFF). The charges of each intermediate were obtained using Gaussian 09 at HF/6-31G(d) level by the RESP model. MD simulations were performed by Christian Curado using the AMBER software package using the ff14SB Amber force field for the protein, and lipid14 force field for the DMPC membrane.

Since the residues D374, D376 and D377 are in close contact with one another and in a protonation equilibrium, it was further investigated by Christian Curado which protonation state for each of the aspartic acids is in best agreement with experimental results. For this purpose, 400 ns MD simulations on wild type, L36A and Y609W variants were performed by Christian Curado, embedded in the DMPC membrane, locating the proton either on the aspartic acid at position D374 or D376. With this 400 ns MD-simulations with D376 in a deprotonated state and a 6 Å threshold for the distance between the intermediate cation and D376 were performed. Thus, catalytically active from inactive frames were discriminated and the number of counts of the docked intermediate carbocations close to D376 gave a similar distribution of cations as experimental results for product distributions of the resulting products for the variants L36A and Y609W. However, for wild type a setting with D374 in a deprotonated state showed similar numbers of count of docked intermediate carbocations within 6 Å that accurately fit with experimental data.

In a detailed analysis of variants L36A and Y609W by Christian Curado, F365 and W489 were identified as key players in orienting all investigated intermediate carbocations. The introduced mutations mainly seem to alter the conformation of these key aromatic amino acids (Figure 33).

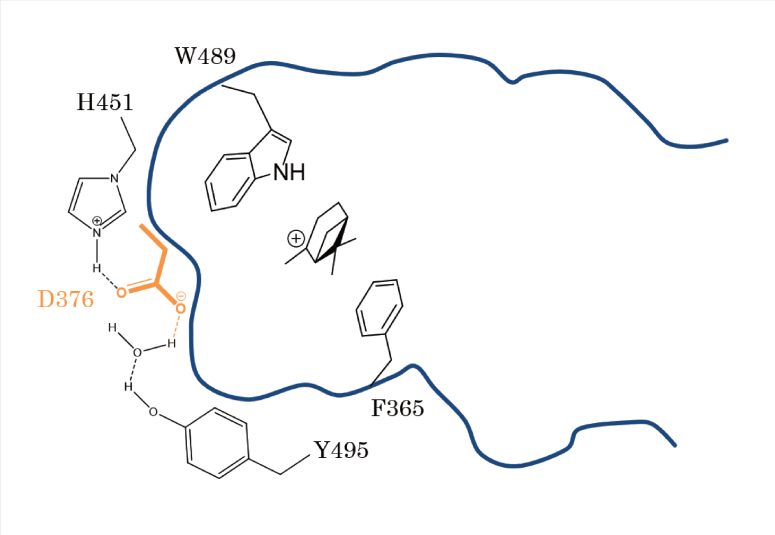


Figure 33: Schematic depiction of the active site. Shown are the catalytic active aspartic acid 376 in orange, as well a hydrogen-bonding network consisting of water, tyrosine 495 and histidine 451, responsible for the *anti*-orientation of the proton of the aspartic acid. The key players F365 and W489, identified by Christian Curado, are depicted within the active site, stabilizing an exemplary pinanyl cation.

The mutation to Y609W seems to introduce a hydrogen bond between Y609W and S309, moving S309 into a different conformation, resulting in additional room for F365 to adopt another conformation, favoring the stabilization of pinanyl cation.

In an attempt to prove the influence of the hydrogen bonds, variants Y609W_S309A and S309A were generated and measured using BT7 protocol and GC-P4 temperature program.

It was expected that variant Y609W_S309A would show a wild type like product distribution since the Hydrogen bond between Y609W and S309 is broken and the newly introduced alanine would allow F365 a more wild type like conformation and thus a wild type product distribution (Figure 34).

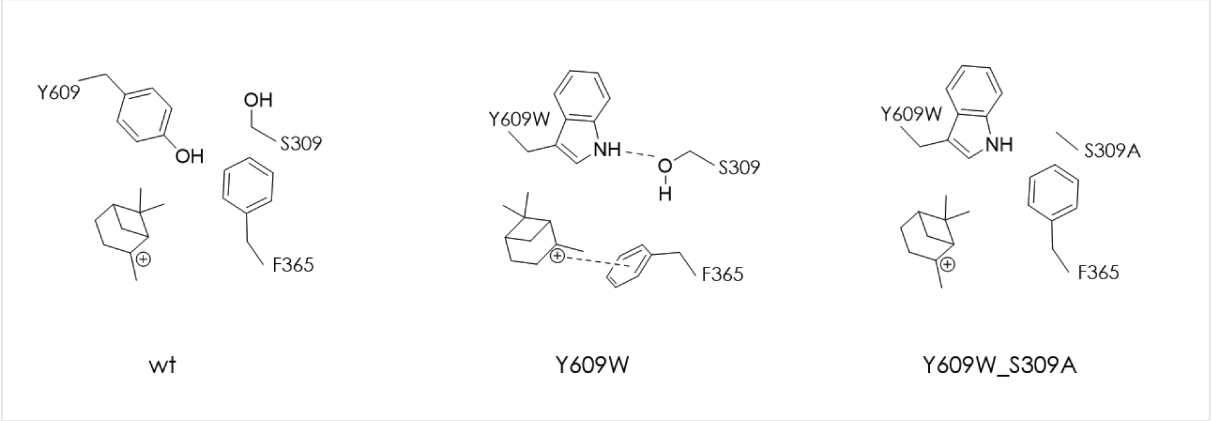


Figure 34: Schematic drawing of the hypothesized influencing hydrogen-bonding networks by introduction of Y609W and Y609W_S309A on pinanyl cation in the active site of *Aac*SHC.

The most obvious result was a dramatic reduction in relative conversion when introducing S309A both as single and as double variant Y609W_S309A (Figure 35 and Table 42). Y609W_S309A shows a decrease in the production of α-pinene but not as much as expected, whereas S309A shows a wild type like product distribution.

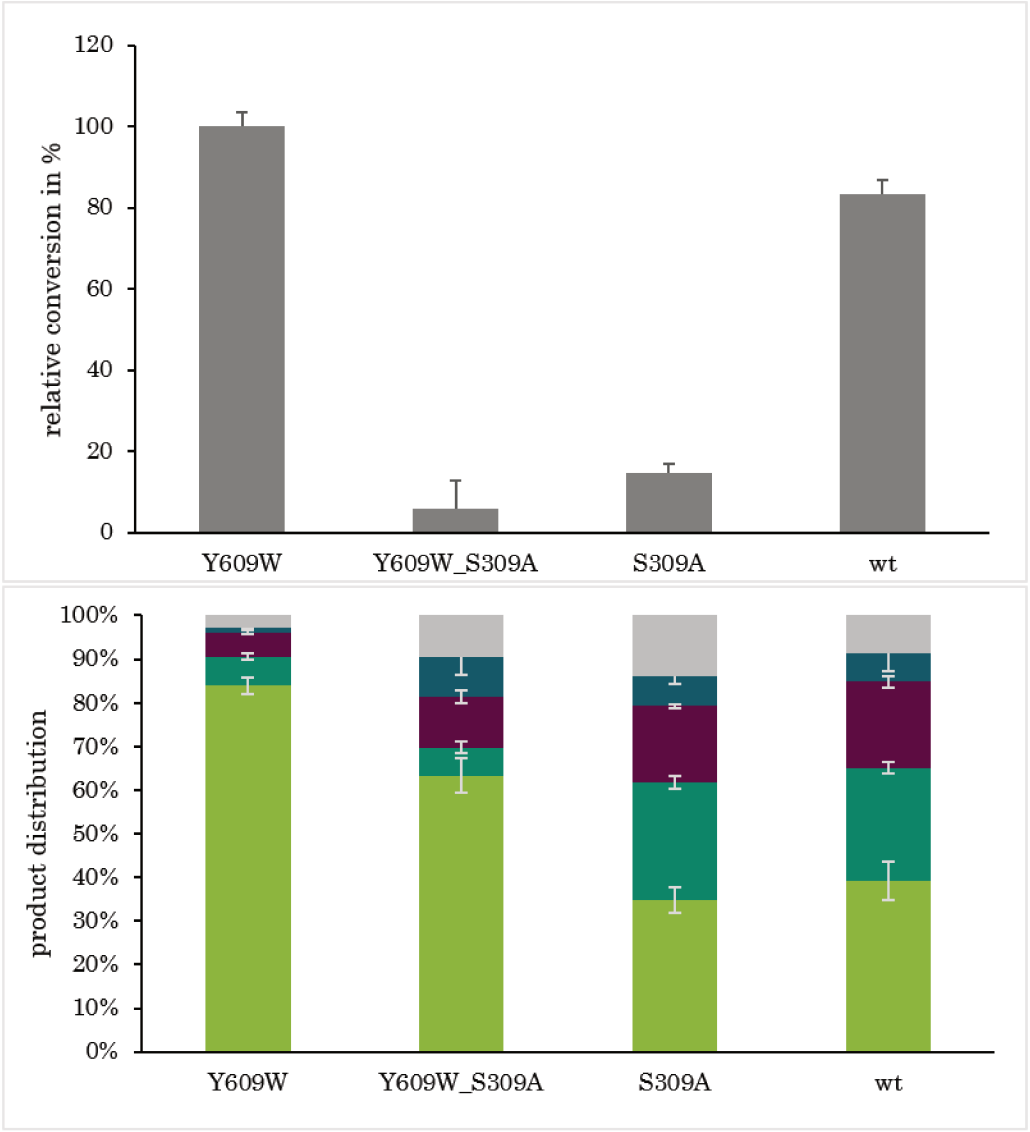


Figure 35: Result of Y609W_S309A in comparison to Y609W and S309A alone and wt for the conversion of β-pinene, α-pinene, β-camphene, limonene, α-terpineol, minor products. Relative conversions are depicted in dark grey on the top with Y609W set to 100% for reference. All data were obtained using BT7 protocol and GC-P4 temperature program.

While Y609W was found to produce α-pinene in larger amounts than the wild type, the variant L36A was found to produce more limonene. This finding can be explained similarly using MD-simulations by Christian Curado.

The smaller alanine at position L36 enables for its backbone to make a hydrogen bond with the side chain of S307. Consequently, the extra space generated by such an interaction allows F365 to adopt yet another conformation, which is hence in favor of stabilizing *p*-menthenyl cation, which in turn can result in the formation of limonene.

In an attempt to prove the influence of the hydrogen bonds, variants L36A_S307A, S307A were generated and measured using BT7 protocol and GC-P4 temperature program.

Similar to the expectation for Y609W_S309A, it was anticipated that variant L36A_S307A would show a wild-type like product distribution since the hydrogen bond between L36A and S307 is broken and the newly introduced alanine would allow F365 a more wild type like conformation and thus a wild type like product distribution (Figure 36).

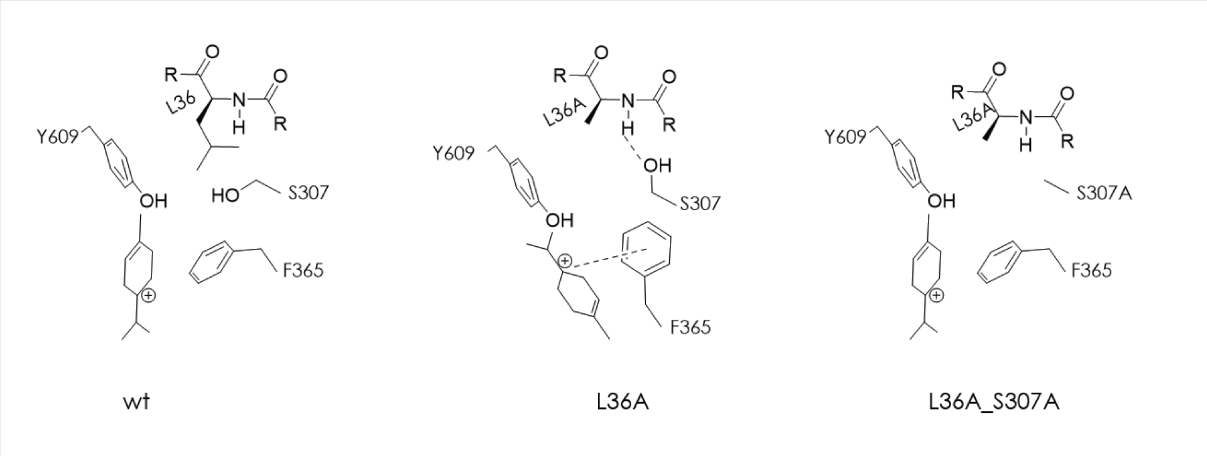


Figure 36: Schematic drawing of the hypothesized influencing hydrogen-bonding network by introduction of L36A and L36A_S307A on *p*-menthenyl cation in the active site of *Aac*SHC.

The variant S307A showed a highly decreased relative conversion, which leads to completely disordered product distribution ratios since the peaks could no longer be integrated reliably due to the low conversion. The variant L36A_S307A on the other hand seemed to be quite unaffected by the introduction of S307A concerning both product distribution and relative conversion (Figure 37 and Table 42). This result seems to contradict the hypothesis on the influence of L36A on F365.

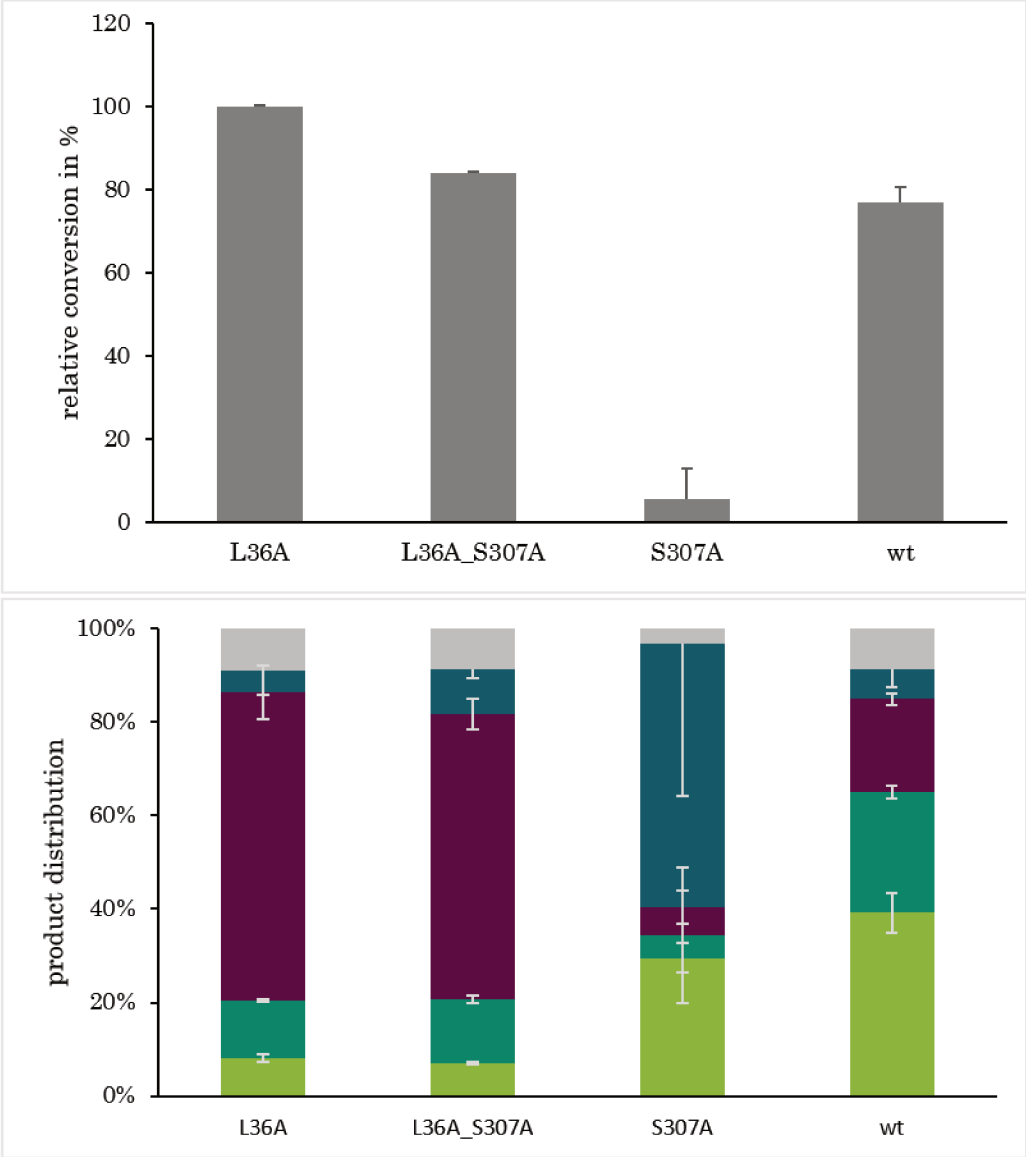


Figure 37: Results of L36A_S307A in comparison to L36A and S307A alone and wt for the conversion of β-pinene, α-pinene, β-camphene, limonene, α-terpineol, minor products. Relative conversions are depicted in dark grey on the top. with L36A set to 100% for reference. All data were obtained using BT7 protocol and GC-P4 temperature program.

The case for D442N is more complicated. The mutation of aspartic acid to an asparagine results in a network of hydrogen bond interactions which leads to a side-chain rotation of F365 that better accommodates a camphyl carbocation.

Based on his simulations, Christian Curado further suggested to create variants P444D and Q366S for an increased ß-camphene production after using Shortest Path Map tool[181] and running 300 ns MD simulations. The variants were created and measured using BT7 protocol and GC-P4 temperature program.

Unfortunately, the results of the single variants yielded wild type like product distributions (Figure 38 and Table 42). Therefore, variants D442N and Q366S were combined in an attempt to use the identified potential of this variant for the formation of ß-camphene. A combination with P444D was not tested since the positions are only two amino acids away from each other and the introduced mutations might disrupt fine-tuned networks. The double variant D442N_Q366S yielded ß-camphene product distributions similarly to D442N but did not exceed them.

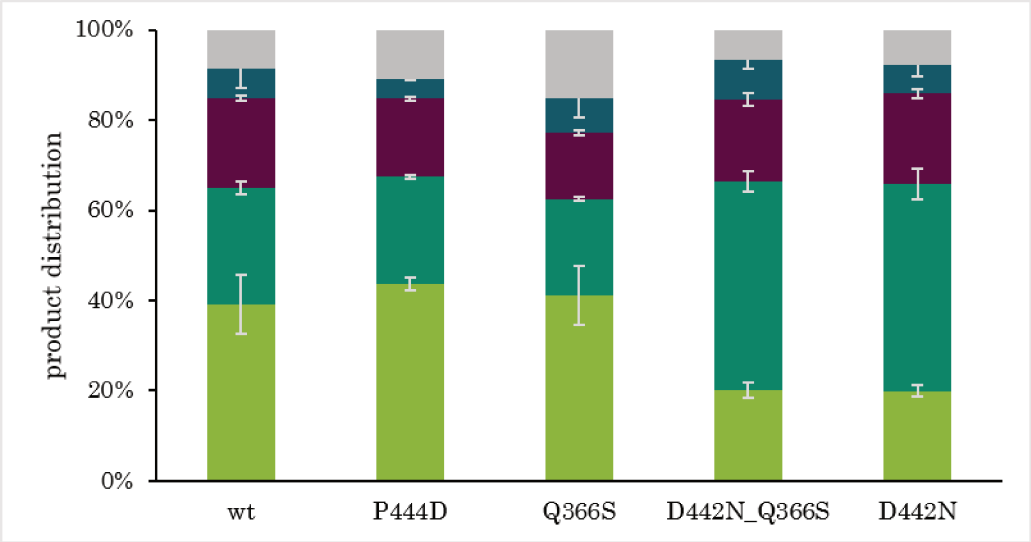


Figure 38: Results for variants P444D, Q366S, D442N_Q366S in comparison to D442N and wild type for the conversion of β-pinene, α-pinene, β-camphene, limonene, α-terpineol, minor products. All data were obtained using BT7 protocol and GC-P4 temperature program.

The expression levels of all variants were similar, thus supporting that the observed effects on the relative conversion are not due to different amounts of catalysts (Figure 39).

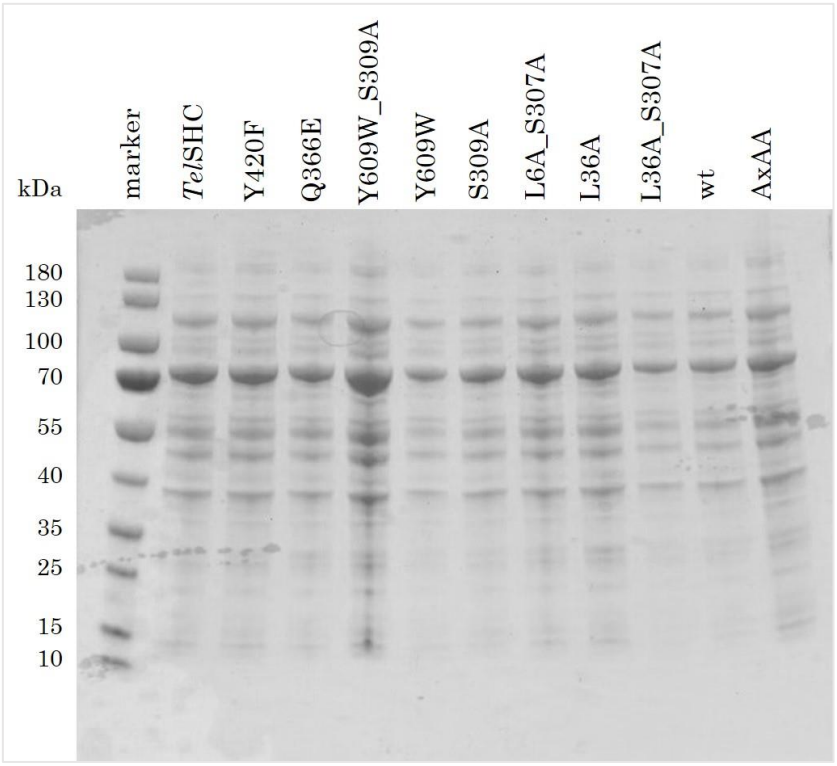


Figure 39: SDS-PAGE gel of *E. coli* extracts containing *Tel*SHC, and *Aac*SHC Y420F, Q366E and variants. A pre-stained Thermo Scientific™ PageRuler™ with a range from 10-180 kDa was used as reference. The *Aac*SHC has molecular weight of 77 kDa.

4.1.3.5 Exploration of the natural diversity of SHCs

In previous chapters it was shown that shifted-product selectivities can be achieved by introduction of mutations. Also, a natural diversity within an enzyme family can lead to altered product selectivities and activities. Therefore, members of the SHC family were investigated for their ability to convert ß-pinene.

Sabrina Henche[116] showed that some alternate cyclases can have superior activities for specific non-natural reactions like Friedel-Crafts alkylation. In the scope of this thesis it was hence tested how this can be applied to the conversion of ß-pinene. All twelve in-house cyclases (see chapter 8.6.3) were tested for higher activities towards ß-pinene using BT6 protocol and measured using the temperature program GC-P2. The buffer served as negative control. The enzymes *Sfu*SHC from *Syntrophobacter fumaroxidans*, *Bam*SHC1 from *Burkholderia ambifaria* and *Ttu*SHC from *Teredinibacter turnerae* showed no activity. It should be noted that GC-P2 protocol was not optimized for a detection of α-terpineol since it was not considered a product of the enzymatic reaction at the time the experiment was conducted (see chapter 4.1.2.3).

Interestingly, SHCs from different organisms already show great variations in their product distributions (Figure 40 and Table 43). The SHC from *Thermesynechococcus elongatus* (*Tel*SHC) showed a high selectivity for the formation of α-pinene (77.0%), which is close to variant Y609W in *Aac*SHC (84.3%). Also, the SHC from *Bradyrhizobium japonicum* (*Bja*SHC) demonstrates good selectivity towards limonene (67.0%), which is similarly to variant L36A in *Aac*SHC (65.9%). Additionally, a production of ɣ-terpinene (1.8%) was observed for *Bja*SHC, which was slightly higher than for W312A in *Aac*SHC (1.3%) and thus the highest observed.

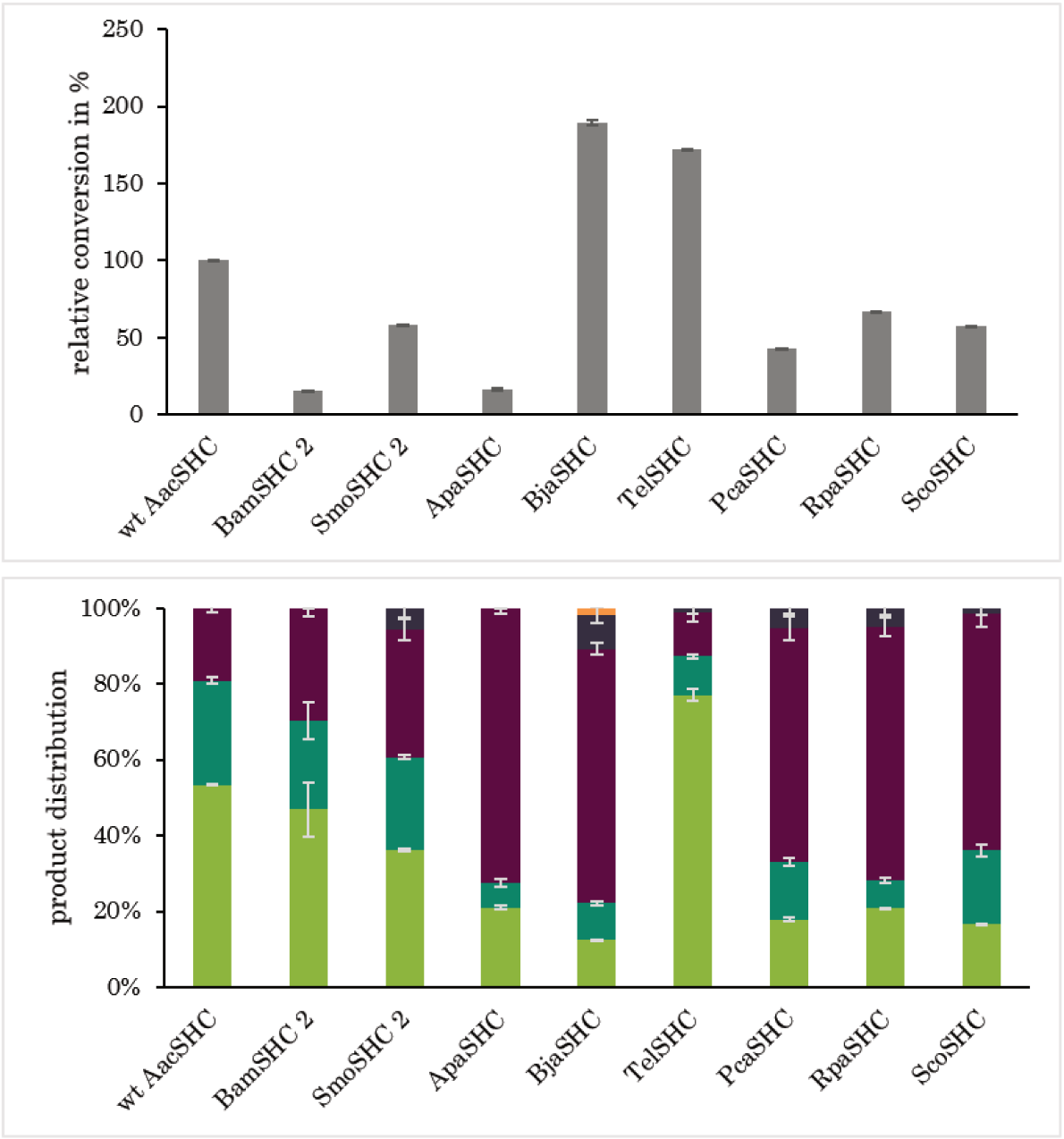


Figure 40: Product distributions of SHCs from different organisms for the conversion of (+)-β-pinene, α-pinene, β-camphene, limonene, terpinolene, γ-terpinene. Relative conversions are depicted in dark grey on the top. All data were obtained using BT6 protocol and temperature program GC-P2.

Only *Tel*SHC and *Bja*SHC showed higher relative conversions than wild type from *Aac*SHC with 172% and 189%, respectively (Figure 40). In combination with their inherent product selectivities they exceed the variants L36A and Y609W in their absolute production of limonene and α-pinene, respectively (Figure 41 and Figure 42).

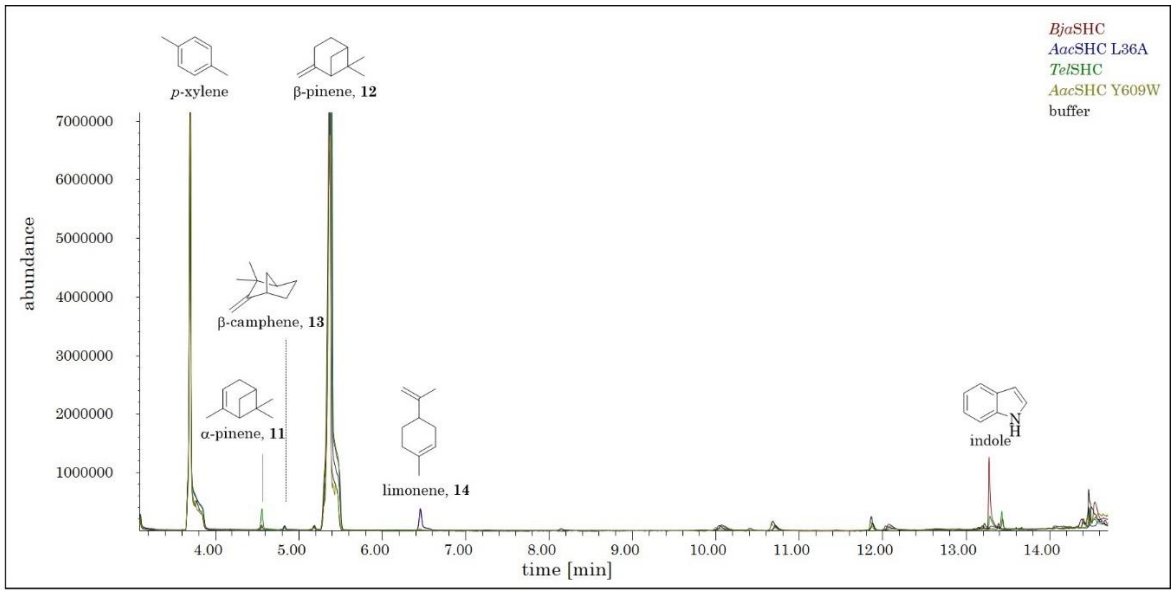


Figure 41: Full length overlaid chromatograms for the conversion of β-pinene (5.3 min) with *Tel*SHC, *Bja*SHC and *Aac*SHC variants Y609W and L36A. *Tel*SHC produces more α-pinene (4.5 min) than the *Aac*SHC variant Y609W, *Bja*SHC forms more limonene (6.4 min) than *Aac*SHC variant L36A, and β-camphene is shown at 4.8 min and *p*-xylene at 3.7 min. All data were obtained using BT6 protocol and temperature program GC-P2.

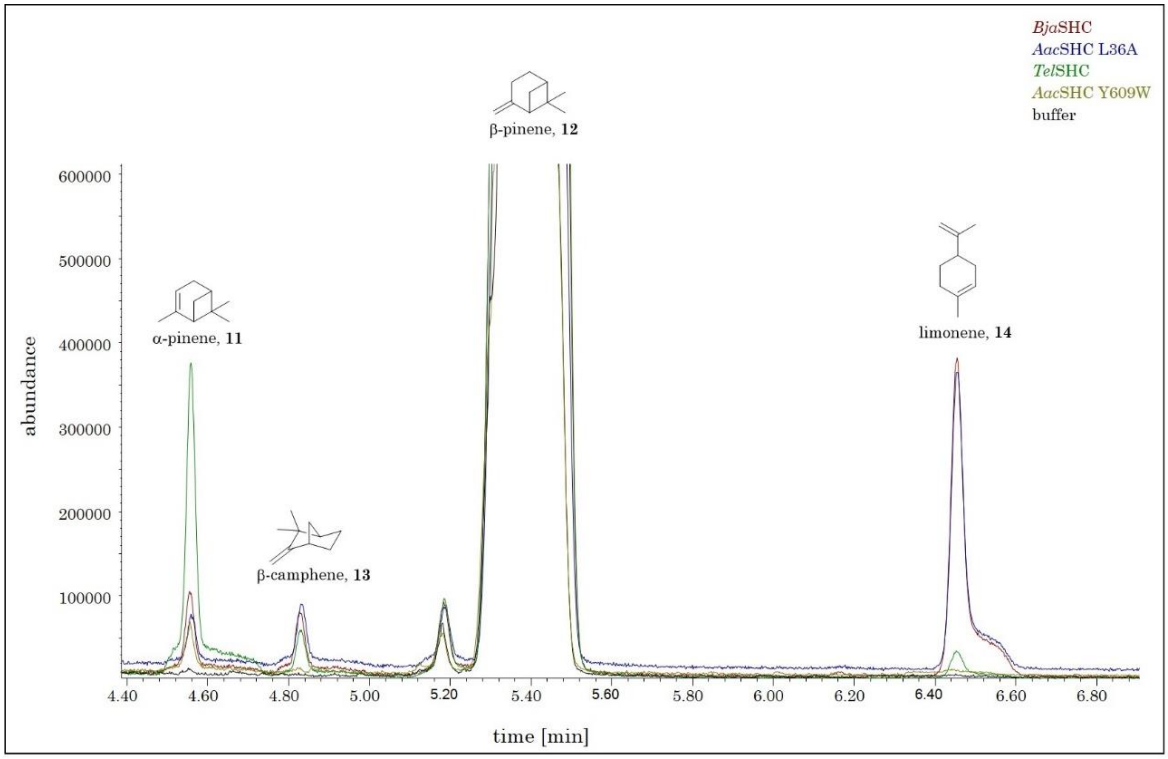


Figure 42: Section of overlaid chromatograms for the conversion of β-pinene (5.3 min) with *Tel*SHC, *Bja*SHC and *Aac*SHC variants Y609W and L36A. *Tel*SHC produces more α-pinene (4.5 min) than the *Aac*SHC variant Y609W, *Bja*SHC forms more limonene (6.4 min) than *Aac*SHC variant L36A and camphene elutes at 4.8 min. All data were obtained using BT6 protocol and temperature program GC-P2.

The next step was therefore to transfer key positions of *Tel*SHC to *Aac*SHC to increase the activity of *Aac*SHC. Therefore, ß-pinene was docked into *Aac*SHC using the protein structure software package YASARA and residues in a distance of 8 Å were compared using Emboss Needle Sequence Alignment*. Positions that showed differences in their amino acid sequence were changed to the ones found in *Tel*SHC. As a result the variants S307C, I308V, Q366E and Y420F were created in *Aac*SHC (in collaboration with Dr. Sabrina Henche[116]) and tested using BT7 protocol and GC-P4 temperature program.

The variants Y420F, S307C and I308V in *Aac*SHC showed a very similar product distribution to the wild type, whereas Q366E produced more minor products (Figure 43 and Table 44). None of the single variants seemed to be solely responsible for the high selectivity of *Tel*SHC in the formation of α-pinene. However, variants Y420F and Q366E showed high relative conversions compared to wild type *Aac*SHC and even *Tel*SHC alone.

*The EMBOSS Needle alignement, using the Needleman-Wunsch alignement algorithm was employed on the website www.ebi.ac.uk from the European Molecular Biology Laboratory in Hinxton.

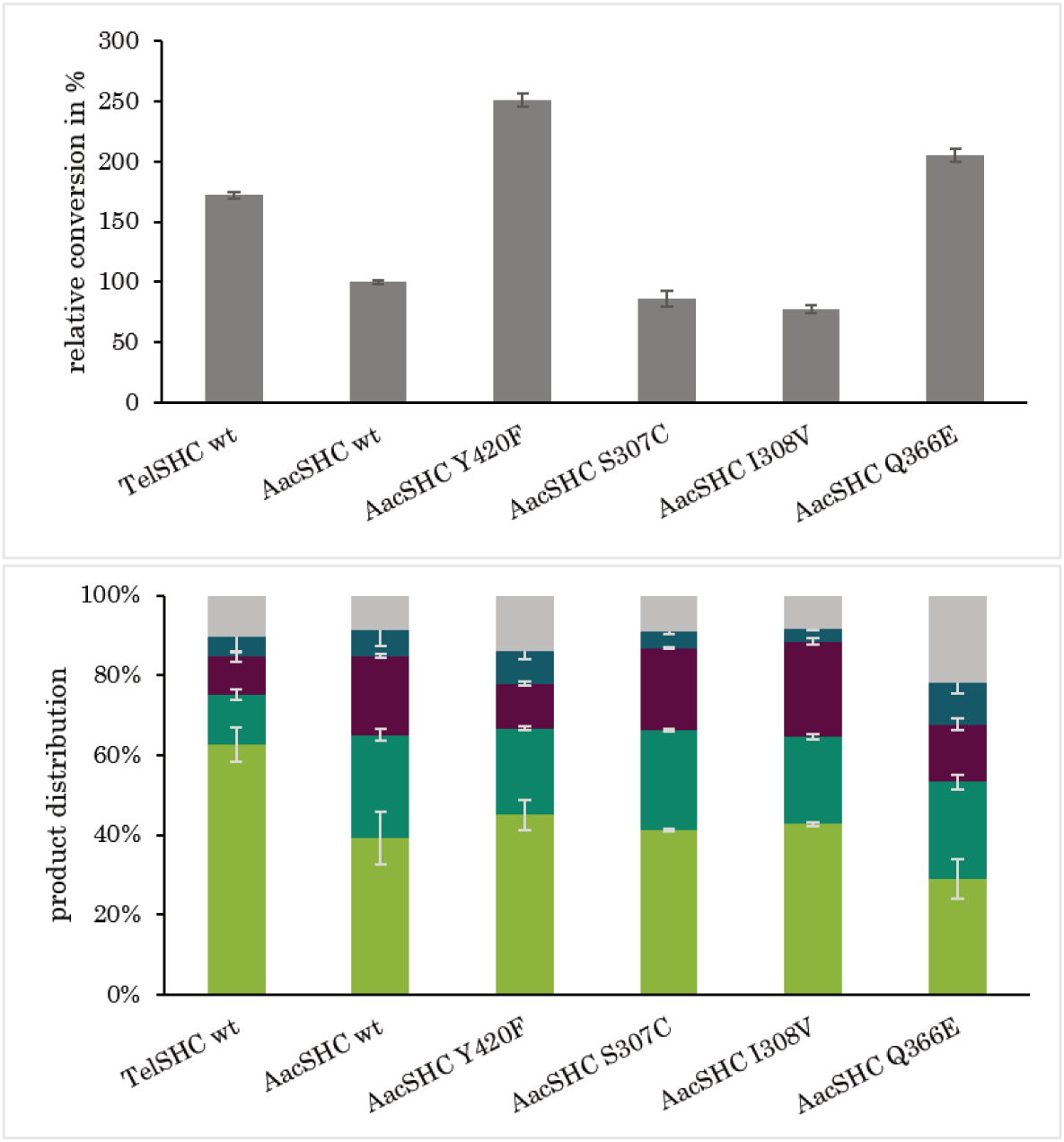


Figure 43: Product distributions for the conversion of β-pinene of wt and variants in *Aac*SHC (Y420F, S307C, I308V and Q366E) in comparison to *Tel*SHC, ■ α-pinene, ■ β-camphene, ■ limonene, ■ α-terpineol, ■ minor products. Relative conversions are depicted in dark grey on the top. All data were obtained using BT7 protocol and GC-P4 temperature program.

Thus, the most selective variants in *Aac*SHC L36A, Y609W and D442N were combined with the most active variants in *Aac*SHC Y420F and Q366E. The results are shown in the following.

Variant L36A_Q366E maintained the product selectivity of L36A but showed lower relative conversion than L36A. Variant L36A_Y420F showed slightly decreased selectivities but increased relative conversion (Figure 44). The respective chromatograms are shown in Figure 45 and Figure 46.

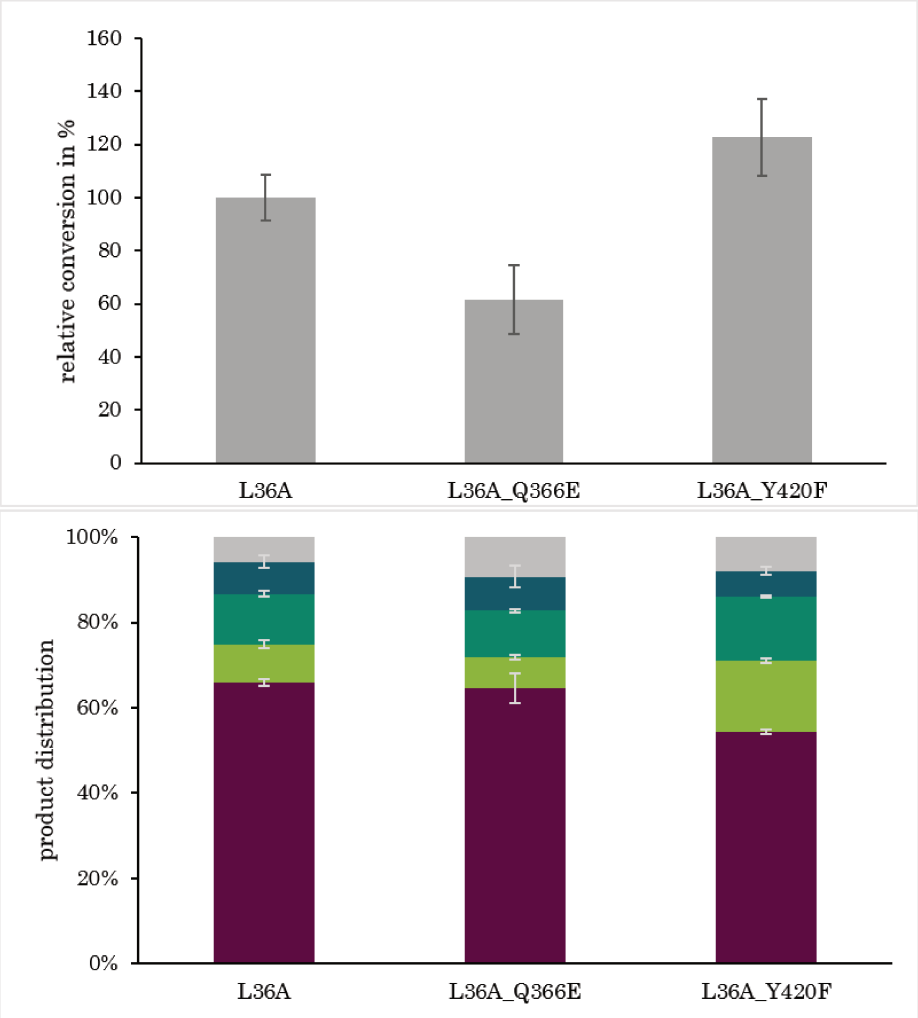


Figure 44: Product distributions and relative conversions for combinatorial variants in *Aac*SHC L36A_Q366E and L36A_Y420F in comparison to L36A for the conversion of β-pinene, ■ α-pinene, ■ β-camphene, ■ limonene, ■ α-terpineol, ■ minor products. Relative conversions are each depicted in dark grey on the top. All data were obtained using BT7 protocol and GC-P4 temperature program.

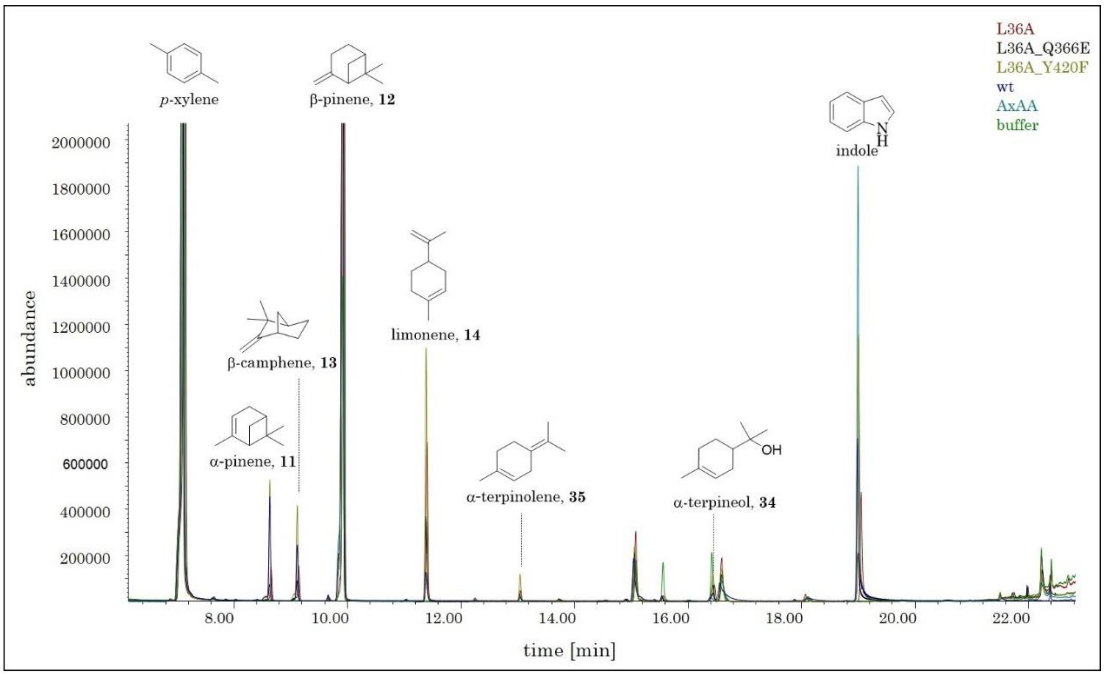


Figure 45: Full length of overlaid chromatograms of variants L36A, L36A_Q366E, L36_Y420F, AxAA, wild type and buffer using temperature program GC-P4. The monoterpenes α-pinene (8.6 min), β-camphene (9.1 min), β-pinene (9.8 min), limonene (11.4 min) α-terpinolene (13.2 min) and α-terpineol (16.6 min) and the standard *p*-xylene (7 min) are highlighted.

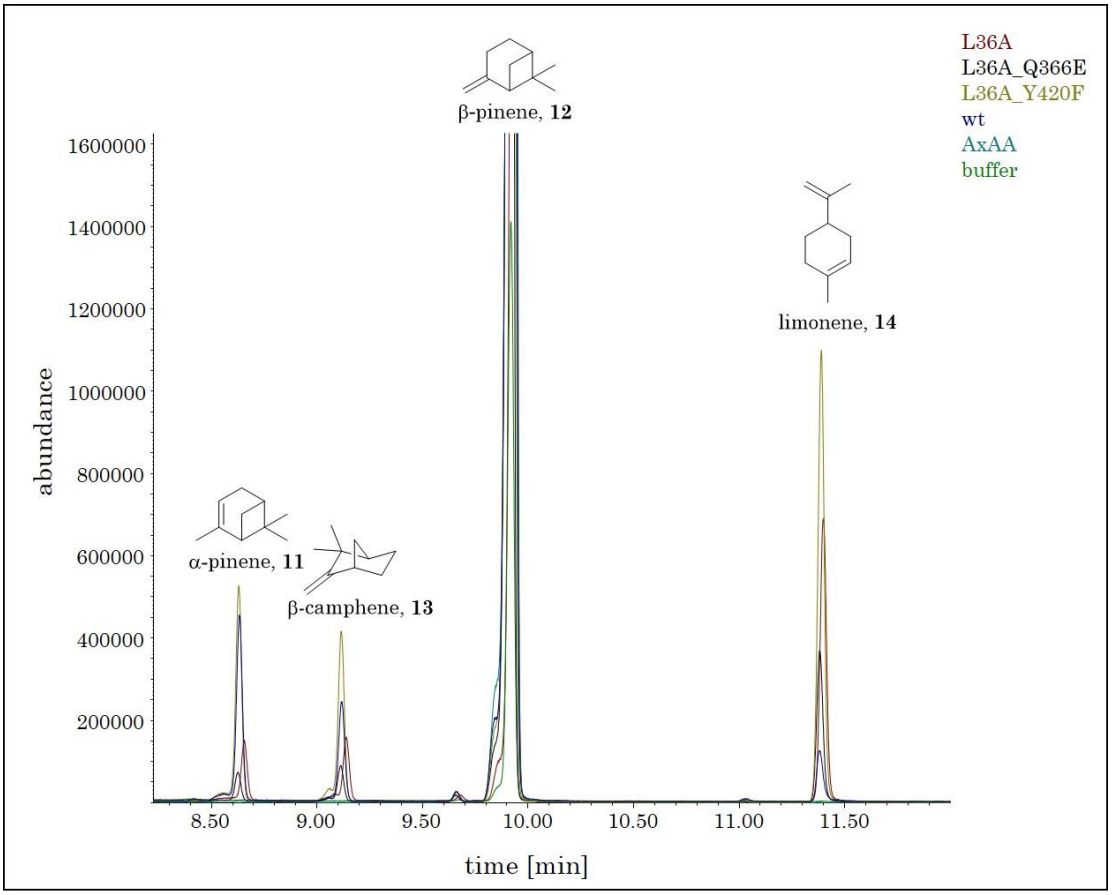


Figure 46: Section of overlaid chromatograms of variants L36A, L36A_Q366E, L36_Y420F, AxAA, wild type and buffer using temperature program GC-P4. The monoterpenes α-pinene (8.6 min), β-camphene (9.1 min), β-pinene (9.8 min), limonene (11.4 min) are highlighted.

The combined variants for D442N with Q366E, Y420F or Q366S maintained a good selectivity for ß-camphene and showed both increased relative conversion rates (Figure 47). The respective chromatograms are shown in Figure 48 and Figure 49.

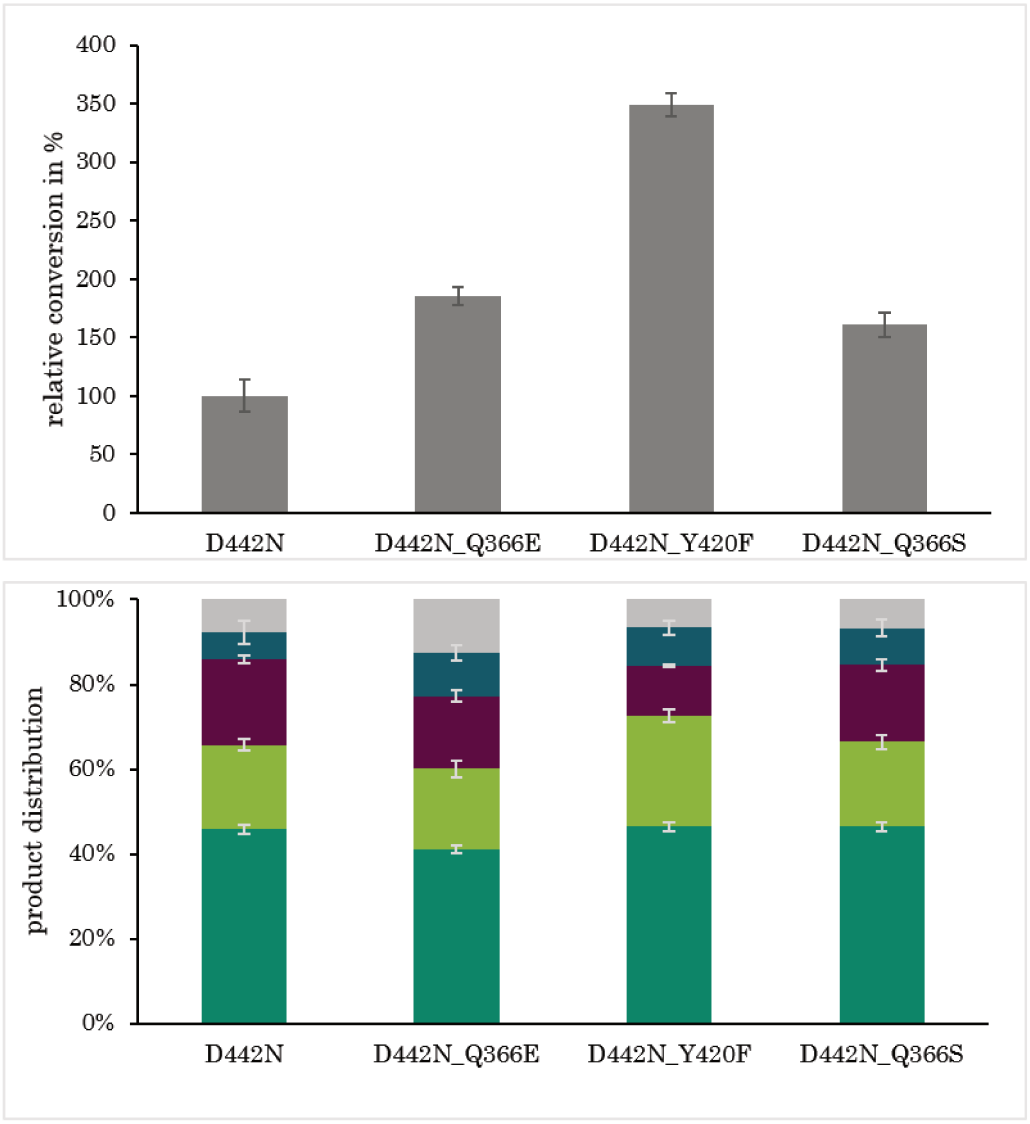


Figure 47: Product distributions and relative conversions for combinatorial variants in *Aac*SHC D442N_Q366E and D442N_Y420F in comparison to D442N for the conversion of β-pinene, ■ α-pinene, ■ β-camphene, ■ limonene, ■ α-terpineol, ■ minor products. Relative conversions are each depicted in dark grey on the top. All data were obtained using BT7 protocol and GC-P4 temperature program.

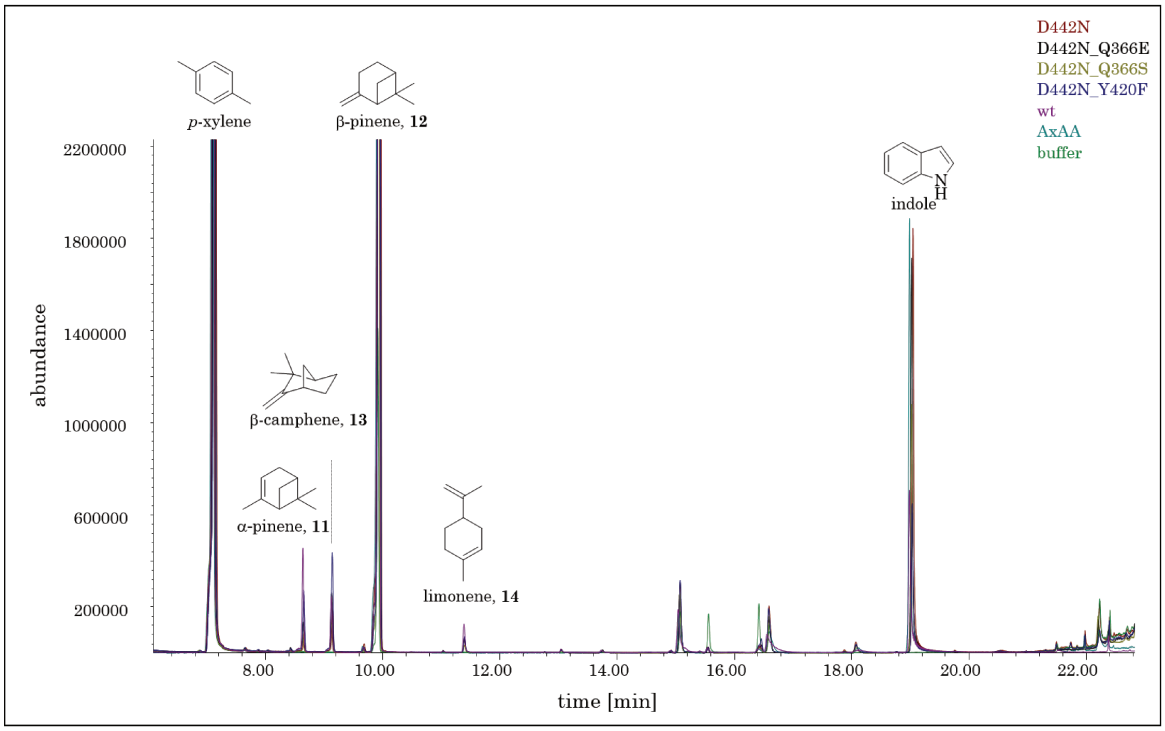


Figure 48: Full length of overlaid chromatograms of variants D442N, D442N_Q366E, D442N_Q366S, D442N_Y420F, AxAA, wild type and buffer using temperature program GC-P4. The monoterpenes α-pinene (8.7 min), β-camphene (9.2 min), β-pinene (9.8 min) and limonene (11.4 min) and the standard *p*-xylene (7 min) are highlighted.

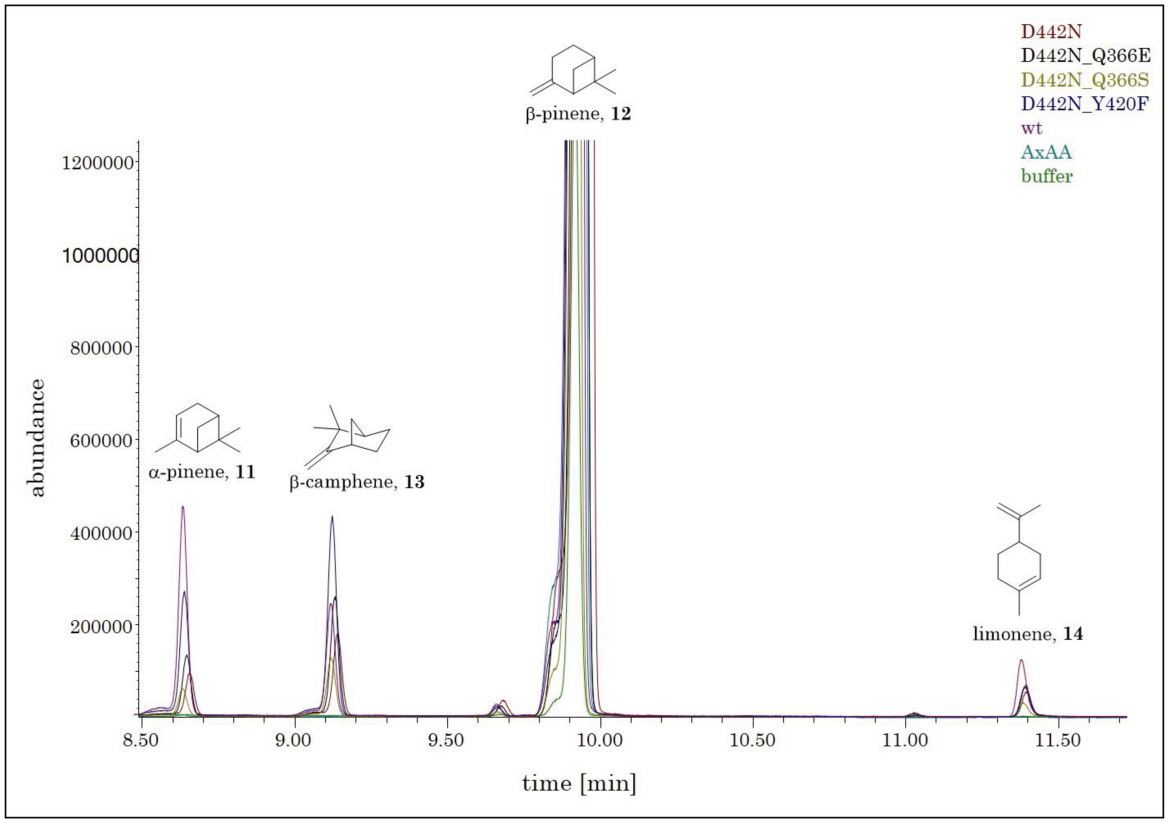


Figure 49: Section of overlaid chromatograms of variants D442N, D442N_Q366E, D442N_Q366S, D442N_Y420F, AxAA, wild type and buffer using temperature program GC-P4. The monoterpenes α-pinene (8.6 min), β-camphene (9.1 min), β-pinene (9.8 min) and limonene (11.4 min) are highlighted.

Combinations of Y609W with Q366E or Y420F showed high product selectivities towards α-pinene and Y609W_Q366E showed higher relative conversions than Y609W (Figure 50). The respective chromatograms are shown in Figure 51 and Figure 52.

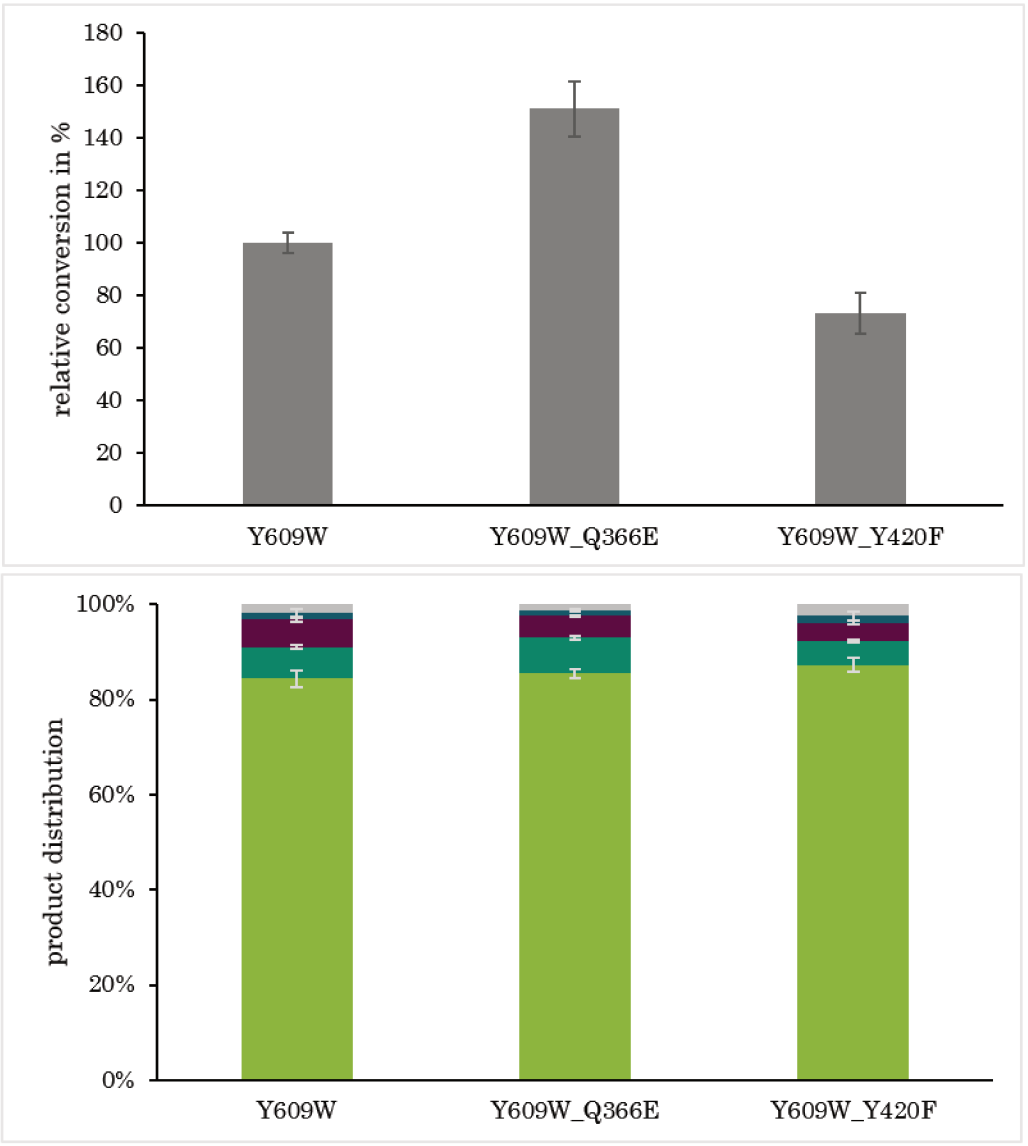


Figure 50: Product distributions and relative conversions for combinatorial variants in *Aac*SHC Y609W_Q366E and Y609W_Y420F in comparison to Y609W for the conversion of β-pinene, ■ α-pinene, ■ β-camphene, ■ limonene, ■ α-terpineol, ■ minor products. Relative conversions are each depicted in dark grey on the top. All data were obtained using BT7 protocol and GC-P4 temperature program.

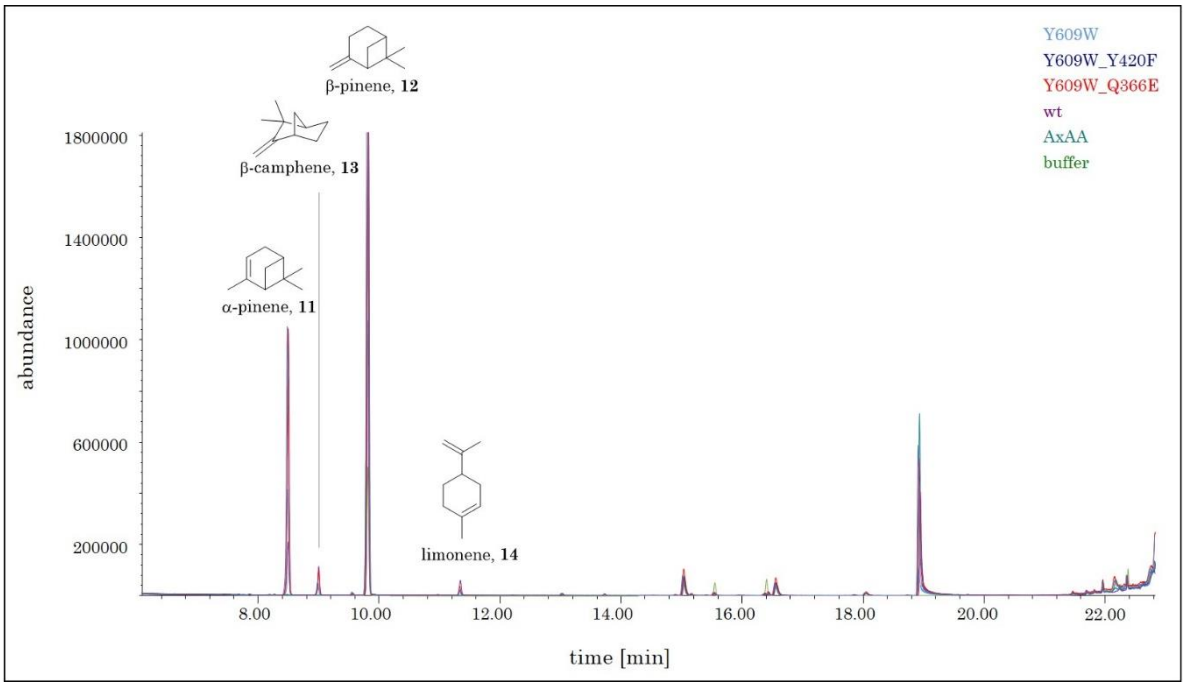


Figure 51: Full length of overlaid chromatograms of variants Y609W, Y609W_Q366E, Y609W_Y420F, AxAA, wild type and buffer using temperature program GC-P4. The monoterpenes α-pinene (8.4 min), β-camphene (8.9 min), β-pinene (9.8 min) and limonene (11.4 min) and the standard *p*-xylene (7 min) are highlighted.

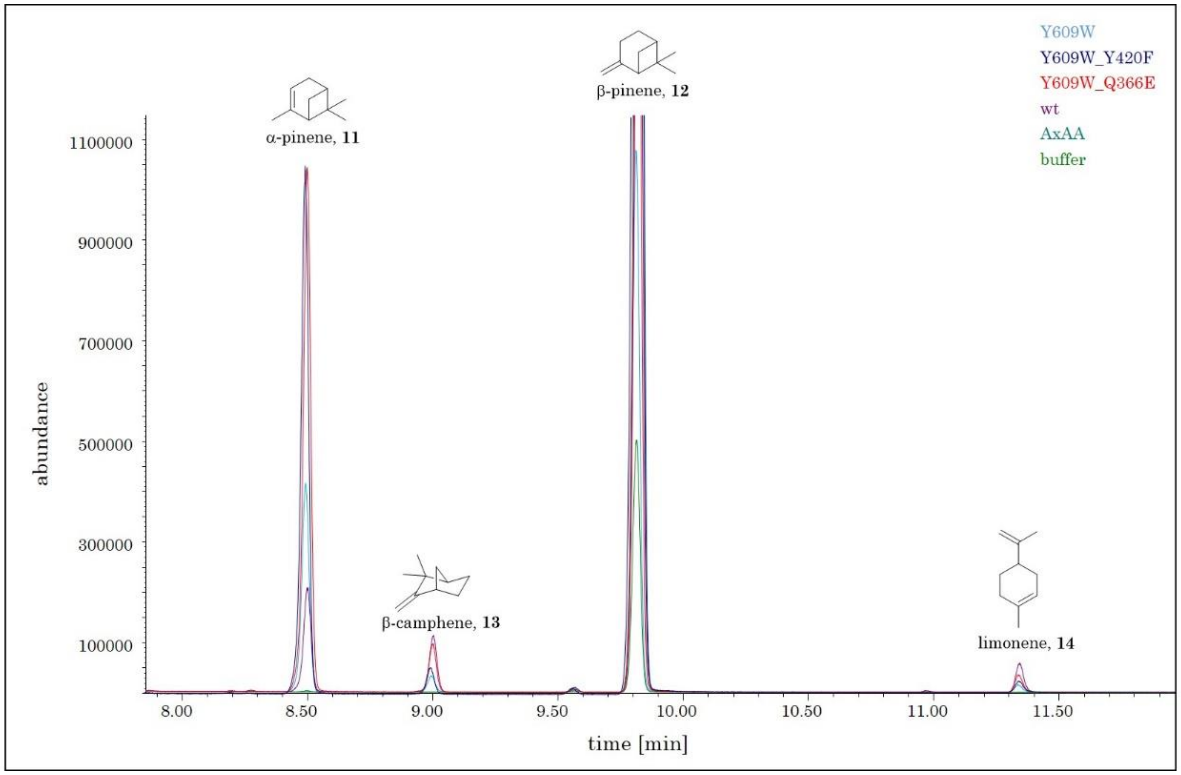


Figure 52: Section of overlaid chromatograms of variants Y609W, Y609W_Q366E, Y609W_Y420F, AxAA, wild type and buffer using temperature program GC-P4. The monoterpenes α-pinene (8.4 min), β-camphene (8.9 min), β-pinene (9.8 min) and limonene (11.4 min) are highlighted.

The expression levels of all variants were similar, thus supporting that the observed effects on relative conversions are not due to different amounts of catalysts (Figure 53).

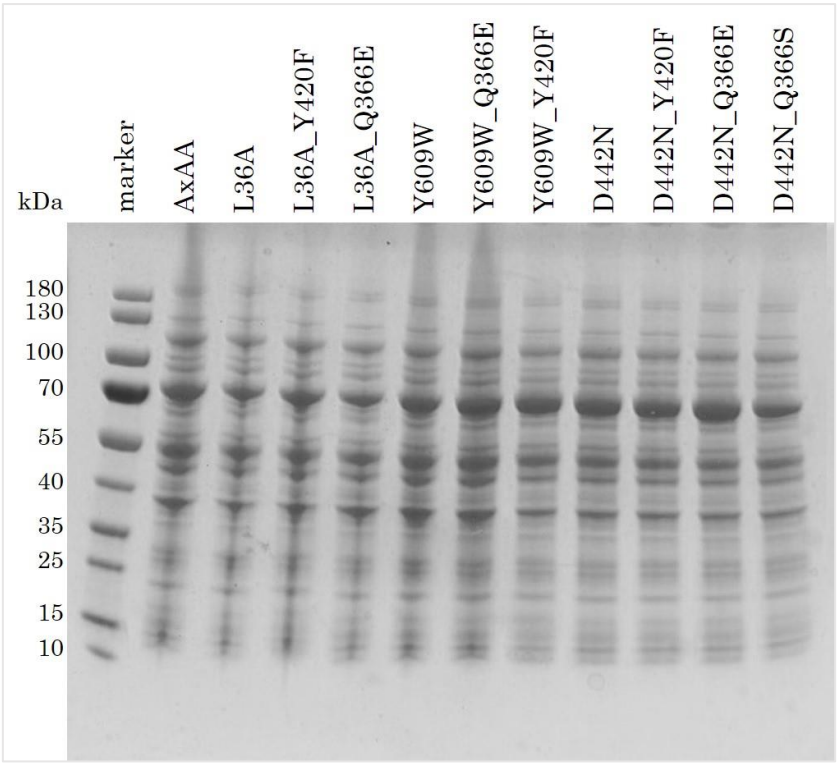


Figure 53: SDS-PAGE gel of *E. coli* extracts from different variants of *Aac*SHC. A pre-stained Thermo Scientific™ PageRuler™ with a range from 10-180 kDa was used as reference. The *Aac*SHC has a molecular weight of 77 kDa.

In general, it was shown that high selectivities and higher relative conversions can be combined.

4.1.4 Summary of the most selective and active variants

The best variants regarding selectivity and activity are summarized in Figure 54 and Table 45. All data were obtained using BT7 protocol and temperature program GC-P4.

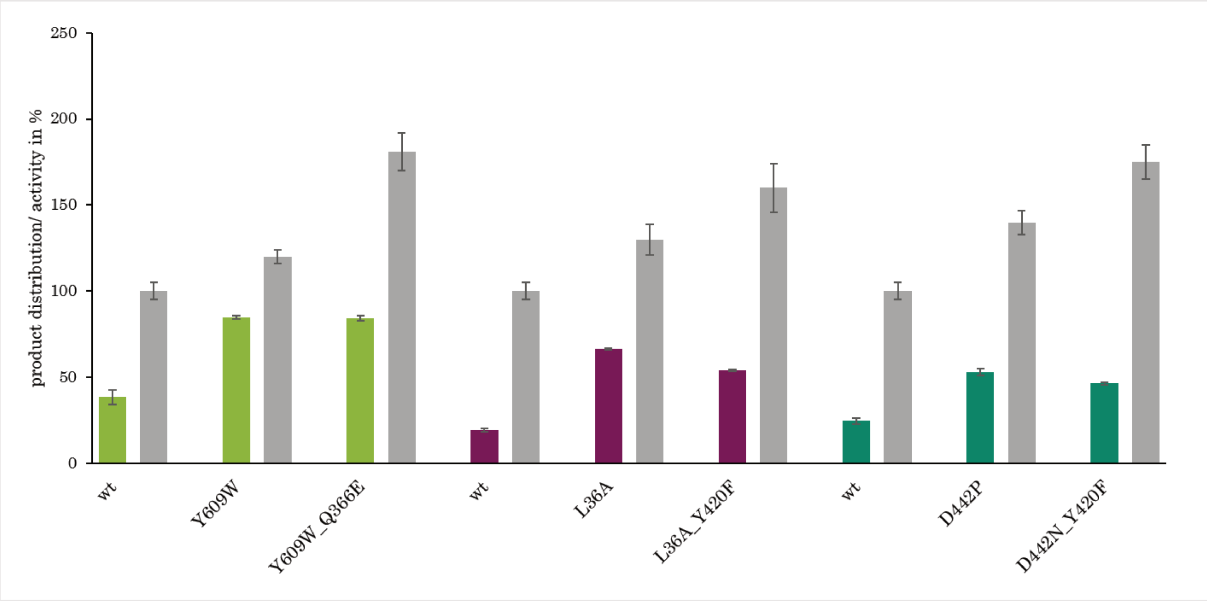


Figure 54: Product distribution and relative activities for the most selective and active variants converting β-pinene are shown in %. Product distributions are depicted for different products ■ α-pinene, ■ β-camphene, ■ limonene. Relative activities are shown in grey and are in reference to wild type, which is set to 100%. All data were obtained using BT7 protocol and temperature program GC-P4.

The wild-type *Aac*SHC showed ~10% overall conversion, depending on the biotransformation protocol used, which is set here as 100% for reference. The highest relative conversion towards α-pinene was achieved with Y609W (84.8%) but with only slightly higher relative conversion to wild type with 120%. This was surmounted by variant Y609W_Q366E, which achieved a relative conversion of 181% similar selectivity (84.4%).

L36A, which was discovered in previous works, was confirmed as best variant (66.4%) for the production of limonene with a relative conversion of 130% to the wild type. A higher relative conversion was achieved in a combination with Y420F to variant L36A_Y420F (160%) but at the expense of selectivity (54.2%).

The most selective variant in the formation of ß-camphene is D422P (53.1%) with a relative conversion of 140%. The relative conversion was improved with variant D442N_Y420F (175%) but again at the cost of selectivity (46.5%).

4.2 *Aac*SHC-catalyzed isomerization of sabinene

4.2.1 Target molecule: Terpinen-4-ol

Terpinen-4-ol is another valuable monoterpene and the main ingredient in tee tree oil. It was shown to have *anti*-tumor[182], *anti*-bacterial[183], *anti*-inflammatory[184] and *anti*-mite[185] activities and was of interest for an industrial partner. Therefore, it was attempted to produce terpinen-4-ol in larger quantities using *Aac*SHC and different substrates. Several candidates, such as α-terpinolene, limonene and γ-terpinene were considered as potential substrates. After protonation these substrates are expected to lead to the formation of terpinen-4-yl cation, which results in terpinen-4-ol after quenching by water. Conceivable pathways and possible products are shown in Figure 55 and Figure 56.

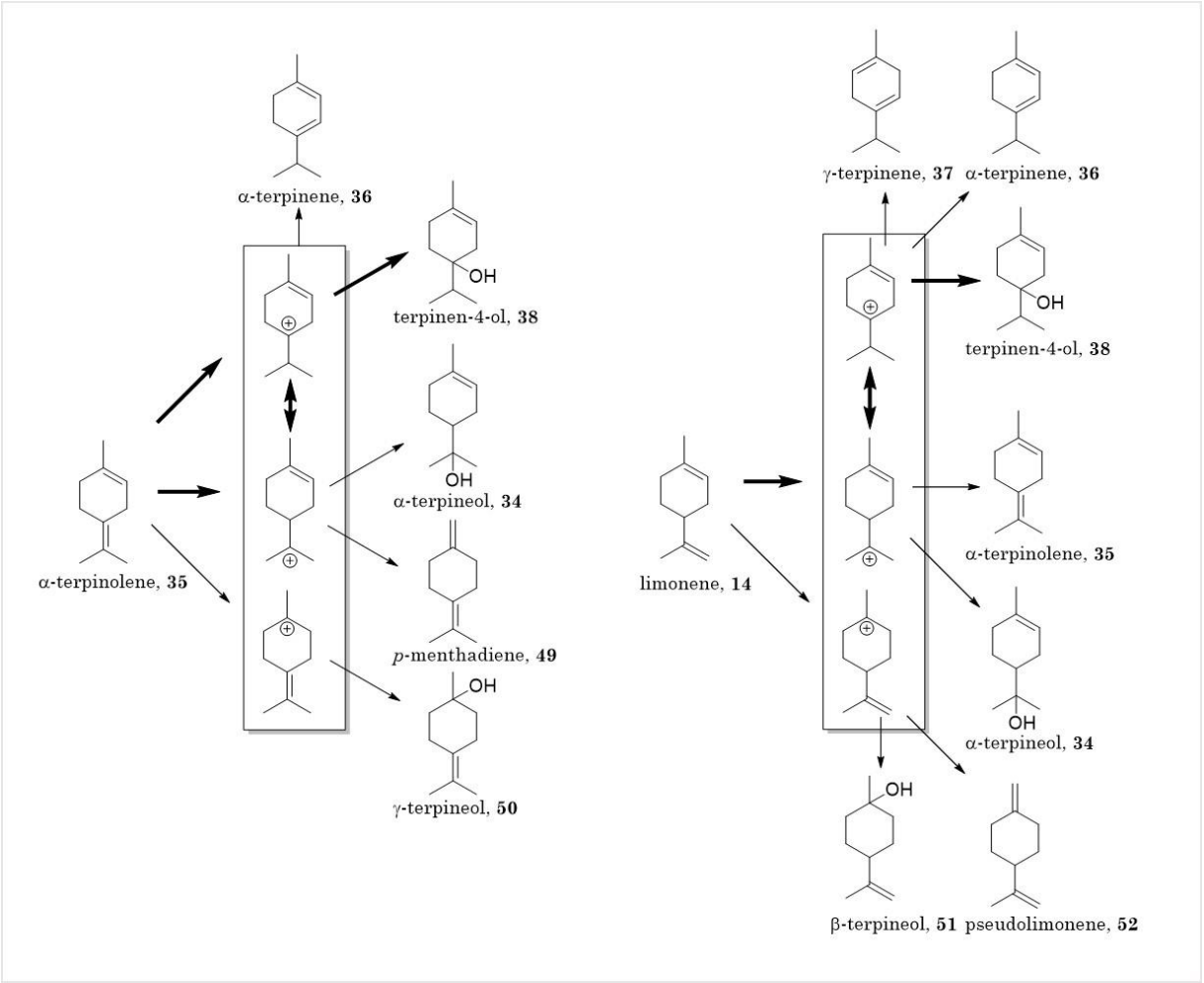


Figure 55: Possible products for the protonation and rearrangement of α-terpinolene and limonene. Pathways leading to terpinen-4-ol are highlighted.

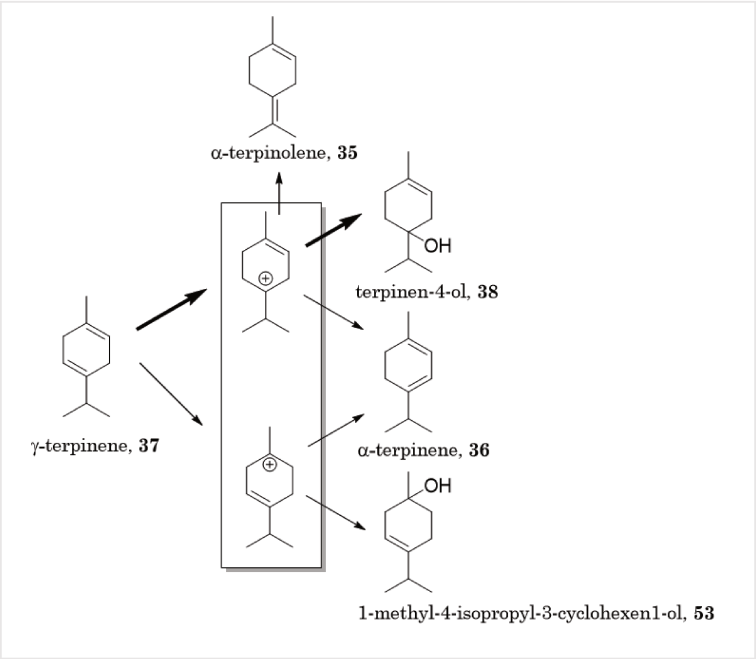


Figure 56: Possible products for the protonation and rearrangement of γ-terpinolene. Pathways leading to terpinen-4-ol are highlighted.

An initial test using protocol BT5 yielded no conversion of any of the substrates by *Aac*SHC.

Literature describes that essential oil from Australian tea tree (*Melaleuca alternifolia*) contains at least 30% terpinen-4-ol, which is produced by a rearrangement of sabinene hydrate. Several attempts were made to isolate the responsible synthase but so far with no success[186,187]. The substrate sabinene hydrate of this synthase makes an inadequate substrate for SHC, however, sabinene is suitable. It consists of a double bond and a strained ring system. Thus, showing some similarities to ß-pinene, which is already accepted as substrate. Moreover, it is expected that after protonation a ring opening leads to the desired terpinen-4-yl cation, which can be quenched to terpinen-4-ol. This route and other possible products are depicted in Figure 57. A docking analysis using the protein structure software package YASARA shows that sabinene adapts a productive binding mode, in which the double bond is oriented towards the catalytic active D376 with a distance of 3.8 Å (Figure 57).

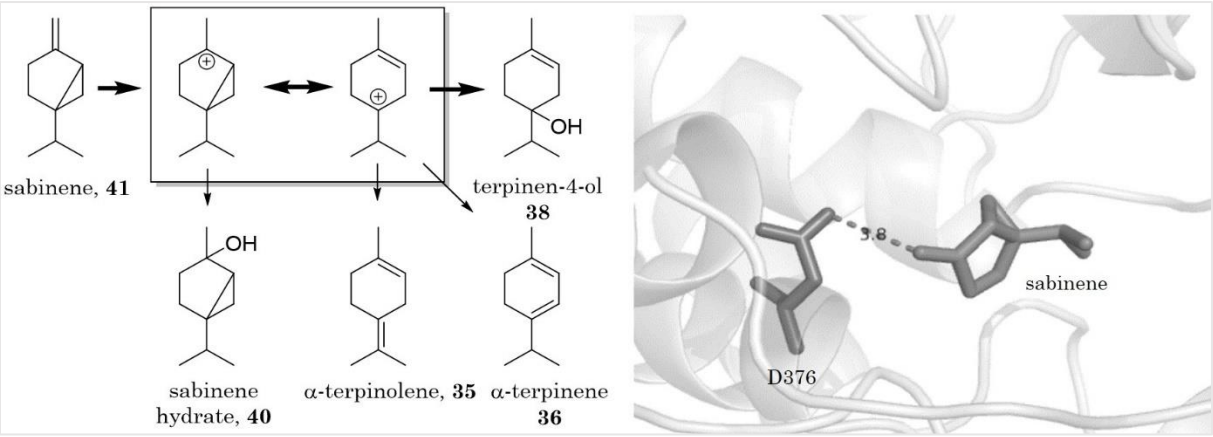


Figure 57: Left: Possible products for the protonation and rearrangement of sabinene. Pathways leading to terpinen-4-ol are highlighted.

Right: Docking of sabinene in *Aac*SHC using the software package YASARA. The double bond and the aspartic acid from D376 have a distance of 3.8 Å.

The conversion of sabinene with *Aac*SHC (BT5 protocol) was successful and yielded among other products terpinen-4-ol. Additionally, the conversion of sabinene was much higher than the conversion of (+)-ß-pinene (see chapter 4.1).

4.2.2 Library screening and product characterization

The in-house libraries for the 1st shell[90] and 2nd shell libraries (see chapter 4.1.3.2) were screened in collaboration with Julius Knerr[180] using BT5 protocol and GC-S1 temperature program. A detailed compilation of the results can be found in his bachelor thesis[180] conducted under the authors supervision. It was observed that almost all variants produce terpinen-4-ol with product selectivities ranging from 16% to 37% with variant T597I being the most selective one (Figure 59). Interesting variants are shown in Figure 58. The variant T597I also exhibited the highest relative conversion with 208% relative to the wild type.

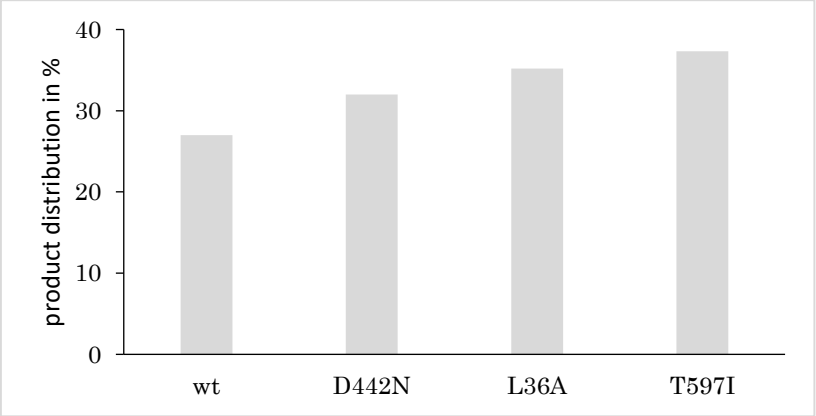


Figure 58: Interesting *Aac*SHC variants for the production of terpinen-4-ol using sabinene as substrate. All data were obtained using BT5 protocol and GC-S1 temperature program in cooperation with Julius Knerr[180].

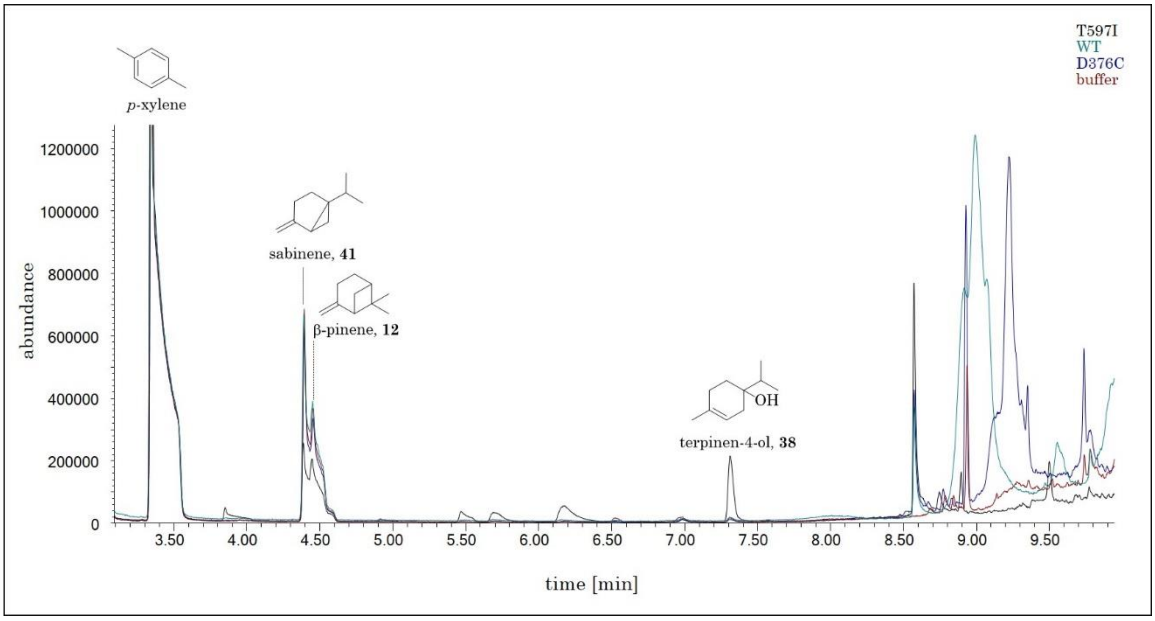


Figure 59: Full length overlaid chromatograms of variant T597I, D376C, wild type and buffer using temperature program GC-S1. The standard p-xylene (3.3 min) and the monoterpenes sabinene (4.3 min), β-pinene (4.4 min) and terpinen-4-ol (7.3 min) are highlighted.

To date, only some of the products could be identified. Table 17 gives a summary of the products that were found, how they were identified and which variant produced most of

the corresponding product. Thereby, not only the selectivity but also the relative conversion are considered (in collaboration with[180]).

Table 17: Identified products for the conversion of sabinene and their most selective *Aac*SHC variants. All data were obtained using BT5 protocol and GC-S1 temperature program, in cooperation with Julius Knerr[180].

product number	product name	identified by standard	suggested by NIST database (GC/MS)	best variant	selectivity [%]	wt selectivity [%]	relative conversion (variant) [%]
P1	α-thujene		X	V448F	54.5 ± 1.7	20 ± 4.8	63.5 ± 1.1
P2	α-terpinene	X	X	Y609A	12.8 ± 0.8	1.3 ± 0.7	12.5 ± 2.0
P3	γ-terpinene	X		L36A	22.2 ± 3.5	4.9 ± 22	16.7 ± 2.6
P4	sabinene hydrate		X	F365W	49.3 ± 4.1	19.1 ± 2.8	23.6 ± 9.7
P5	*cis*-ß-terpineol		X	G600F	44.3 ± 3.5	23.3 ± 2.9	62.4 ± 2.7
P6	terpinen-4-ol	X		T597I	37.3 ± 3.1	27.3 ± 3.7	14.5 ± 8.3
P7	α-terpineol	X		L36W	3.4 ± 0.2	-	38.5 ± 3.6
P8	α-phellandrene		X	V448F	1.0 ± 0.8	-	63.5 ± 1.1
P9	ß-thujene		X	V448F	2.3 ± 0.4	-	63.5 ± 1.1
P10	*cis*-2-menthenol		X	G600F	3.5 ± 0.3	-	62.4 ± 2.7

The improvements of selectivities towards several products compared to wild type are remarkable. Some of the products were not even produced by the wild type, thus showing potential to address these products.

Additionally, it was observed that in contrast to ß-pinene as substrate, the product distributions vary over time. Some variants that were extracted after 20 hours and 40 hours showed a differing product distribution. This is shown exemplary for variant Q366F in Figure 60.

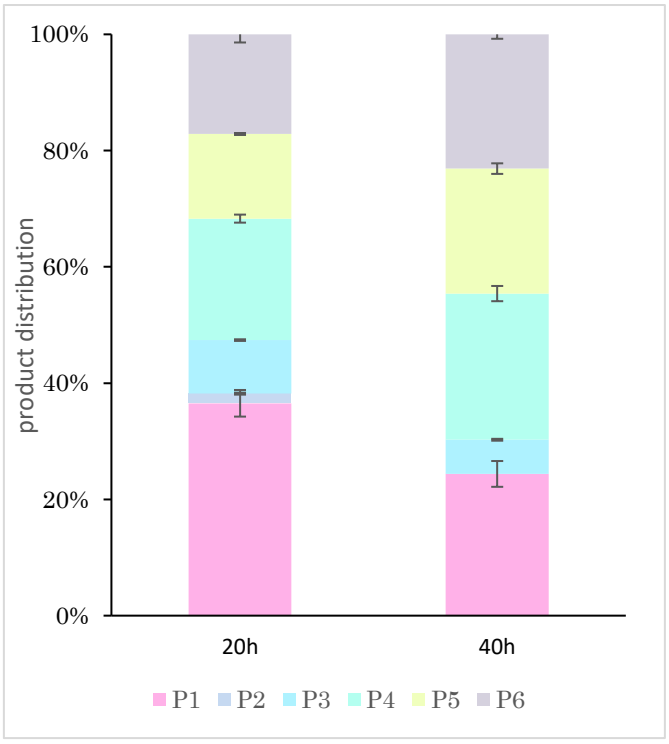


Figure 60: Product distribution for variant Q366F converting sabinene after 20h and 40h. All data were obtained using BT5 protocol and GC-S1 temperature program, in cooperation with Julius Knerr[180] under authors supervision.

It is noticeable that the products containing a double bond (P1: α-thujene, P2: α-terpinene, P3: γ-terpinene) decrease over time, whereas the products quenched by water (P4: sabinene hydrate, P5: *cis*-ß-terpineol, P6: terpinene-4-ol) increase over time. The exact numbers can be found in Table 46 in chapter 8.7.8.

5 Discussion

5.1 SHC-catalyzed isomerization of (+)-β-pinene

5.1.1 General reaction

The conversion of (+)-ß-pinene with *Aac*SHC yielded α-pinene, ß-camphene, limonene and α-terpineol as main products with several minor side products including borneol, fenchol, terpinolene, γ-terpinene and terpinen-4-ol.

This is no surprise. Several reports can be found that describe a similar product spectrum after biotransformation of pinenes with microorganisms. However, none of these studies mention an enzyme that might be responsible for the emergence of these products. Taking a closer look at these reported transformations it is noticeable that all of these microorganisms contain squalene-hopene cyclases: Yoo et al[188] report the conversion of α- and ß-pinene into α-fenchene, ß-camphene, limonene, α-terpinolene, fenchyl alcohol, terpinen-4-ol, α-terpineol, borneol and some oxidation products using a *pseudomonas* strain isolated from pine bark, sap, and surrounding terpene soaked soil. There are 55 entries in the uniprot database[189†] for SHCs in pseudomonas, giving a hint that SHCs are responsible for the mentioned transformation. Rottava et al[190] tested 400 microorganisms for their ability to grow on (-)-ß-pinene as sole carbon source. They identified only fungal strains that were able to convert (-)-ß-pinene and focused on *Aspergillus niger* strains for further investigations. The strains produced mainly α-terpineol and traces of fenchol and *trans*-pinocarveol. However, no monocyclic products were observed. This was consistent with previous findings of the same group of researchers[191]. The uniprot database[189†] lists 9 SHCs for *A.niger* strains with sequence similarities of 29-52% to *Aac*SHC[‡]. Priya et al[192] observed the conversion of (+)-α-pinene to ß-camphene, limonene, α-terpineol, verbenone, isoborneol and *trans*-sobrerol using the bacterial strain *Gluconobacter japonicus*. SHCs from *G. japonicus* showed 6 entries in the uniport database[189†] with sequence similarities to *Aac*SHC of 40-57%[‡].

As shown in chapter 4.1.3.5, SHCs from other organisms can generally convert (+)-ß-pinene to similar product spectra as *Aac*SHC. Therefore, it can be assumed that previous reports on conversions of pinenes with microorganisms can be traced back to the promiscuous activity of an inherent SHC. This further supports the results of this thesis.

However, it should be noted that Narushima et al[193] already noted that the conversion of α-pinene into ß-camphene and borneol by the strain *Pseudomonas maltophilia* could be the result of an "acid catalysis-like reaction" [193]. Thereby, unconsciously already predicting the mode of action of the probably inherent SHC.

† The uniport database was accessed on 7th January 2019.
‡The EMBOSS Needle alignement, using the Needleman-Wunsch alignement algorithm was employed on the website www.ebi.ac.uk from the European Molecular Biology Laboratory in Hinxton.

5.1.2 Reaction engineering

It was shown that large amounts of products and substrate evaporate over the course of the reaction (chapter 4.1.1.5). Therefore, every further improvement should aim at reducing evaporation during the reaction or improving conversion rates to reduce the reaction time.

In total, improvements for the reaction setup were achieved in the course of this thesis, which improved signal intensities and thereby revealed so far undetected products. A great reduction in preparation time was accomplished by using whole cells instead of purified enzyme.

However, the question arises where the SHC resides in *E. coli* during whole cell reactions and how the substrate gets into contact with the enzyme. The original host *A. acidocaldarius* is a Gram-positive bacterium[194], whereas *E. coli* is Gram-negative[195], meaning *E. coli* has an additional membrane[196]. Since the SHC is inserted into the membrane with one helix in *A. acidocaldarius*[89] it can be assumed that this is also true for *E. coli*. In general, proteins find their destination via signal sequences[197]. Therefore, *Aac*SHC was investigated for an accidental signal sequence for the outer membrane in *E. coli*[198]. Since no such signal sequence was found and predicting online tools like PFAM[199] and SMART[200] could not find a signal sequence, it can be assumed that *Aac*SHC can not cross the inner membrane and is therefore anchored to the inner membrane of *E. coli*.

Thus, the monoterpenic substrate and products have to cross the outer membrane and intermembrane space to insert into the inner membrane to be accessible for the cyclase. After the reaction they have to migrate back through the outer membrane into the extracellular space. Pichersky et al. suggested that plant volatiles can cross membranes freely[201]. Since this refers to plant cells, it is questionable whether this applies to bacterial membranes, as well. However, due to similar evaporation rates of the products (see chapter 4.1.1) the investigated monoterpenes seem to be able to cross *E. coli* outer membranes prior to extraction.

This additional membrane-crossing in *E. coli* might be a reason why shaking velocity affects conversion rates between BT6 and BT7 protocol. A more vigorously shaking might therefore enhance diffusion and crossing of the outer membranes on the way to and from the cyclase.

Nevertheless, the local distribution of the substrate in *E. coli*, products and the catalyst *in vivo* still needs to be investigated. New insights on diffusion enhancement or bringing cyclase and substrate closer together might lead to improved conversion rates.

5.1.3 Reaction characterization

5.1.3.1 Concept of possible products

It should be noted that several of the predicted products (see chapter 4.1.2.1) were found throughout the experimental course of this thesis. The concept is again shown in Figure 61 and previously known products are highlighted in a bright green color, whereas newly identified trace products are colored dark green. A special case is α-terpineol, which was always observed but not considered an enzymatic product of *Aac*SHC since it was already produced by variant D376C, which was treated as negative control due to preliminary studies. With variant D374A_D376A_D377A (AxAA) it was shown that α-terpineol is indeed a product of the enzymatic reaction and that D376C is not a suitable negative control in this reaction (see chapter 4.1.2.3).

Comparing this concept of possible products (Figure 61) with published reaction paths in biosynthesis by monoterpene synthases the following points are noticeable.

First, the order of some rearrangements is different. For example, the central carbocation in monoterpene synthases is *p*-menthenyl (see chapter 1.2.3). It is rearranged into pinanyl or bornyl cation, whereas the scheme in chapter 4.1.2.1 and Figure 61 predicts that both carbocations emerge from pinanyl cation after protonation by *Aac*SHC. This is supported by theoretical investigations on the chemical isomerization of pinene to ß-camphene, which is also based on a protonation initiation[58,70]. This interchangeability in the order of cations that can be passed through highlights the general low energy barriers of these carbocations (see chapter 1.2.2) and how the catalyst influences the reaction path.

Second, monoterpene synthases use the re-addition and -abstraction of pyrophosphate to stabilize bornyl cation to form borneol (see chapter 1.2.3), as demonstrated with variant W312A. *Aac*SHC can form borneol without pyrophosphate and is believed to use aromatic residues to stabilize and direct the cation.

Third, the scheme can also be accessed by protonation of other monoterpenes with a double bond other than ß-pinene. This was demonstrated for sabinene as substrate. The results suggested that deprotonated products serve as new substrates over time, as well (see chapter 0). Thus, the challenge remains to control the product outcome and not to initially create cations.

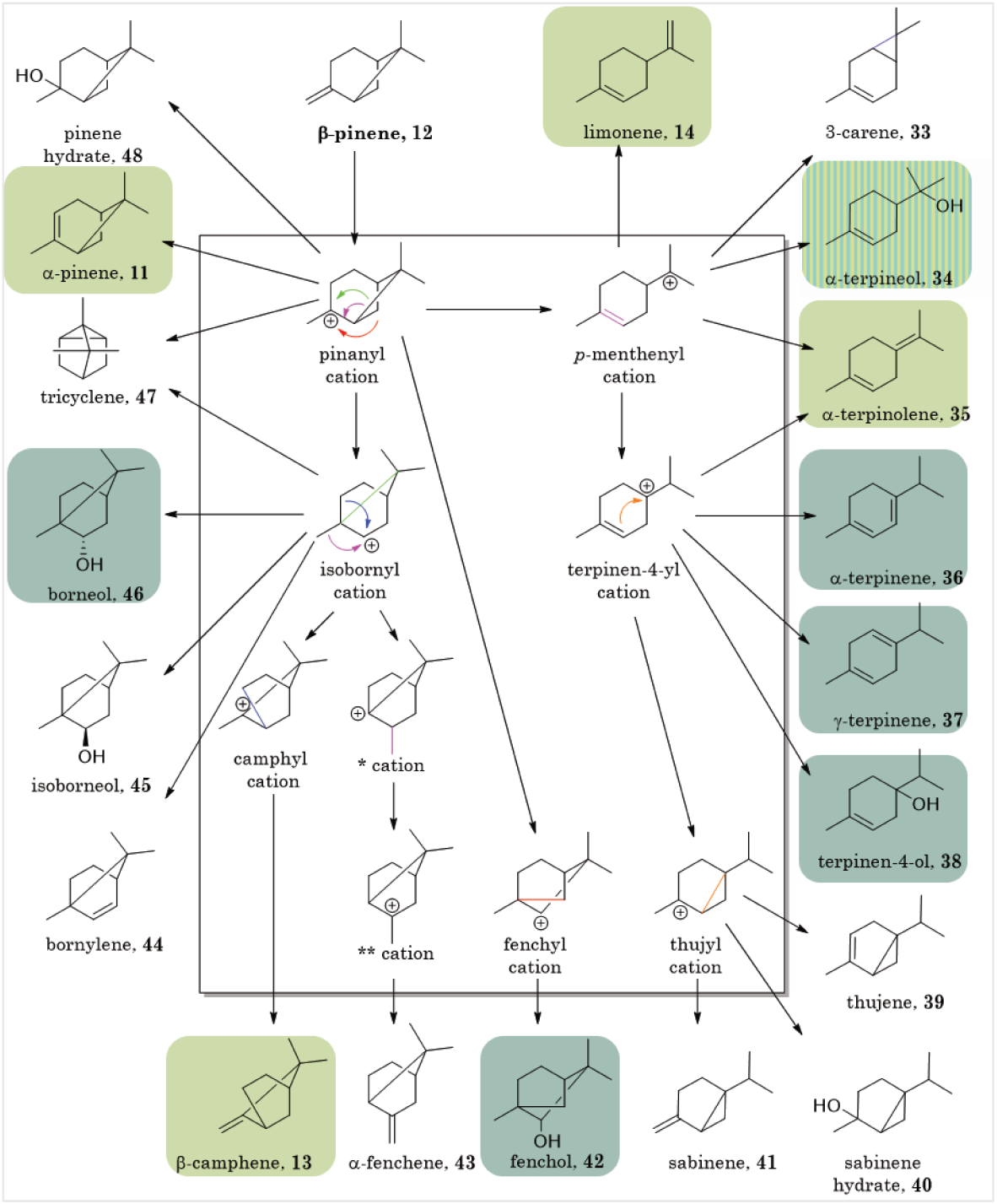


Figure 61: Concept of possible products. Products that were already found in previous studies are highlighted in light green. Products that were found in the course of this thesis are colored dark green. α-Terpineol was not considered an enzymatic product of *Aac*SHC at the beginning of these studies but confirmed to be a product of the enzymatic reaction. It is therefore colored in stripes.
*** 2,7,7-trimethylbicyclo[2.2.1]heptan-1-yl**
**** 2,7,7-trimethylbicyclo[2.2.1]heptan-2-yl**

Comparing this concept and the results of *Aac*SHC variants with reported chemical isomerizations, other aspects come into focus.

The industrially-applied and well-studied isomerization of α-pinene to ß-camphene using TiO_2 hydrate shows a product selectivity of 35-45% towards ß-camphene[58,59]. This was surpassed by *Aac*SHC variant D442P achieving a product selectivity of 53%. The activity was enhanced with the variant D442N_Y420F while maintaining a selectivity of 46.5%. However, reports on ongoing research can be found that list 73% product selectivity with activated clay but at the cost of activity (21% conversion)[64]. Higher conversions were only achieved by reducing the product selectivity. For example, Comelli et al[63] state almost full conversion of α-pinene to ß-camphene with sulfated zirkonia while achieving a product selectivity of 56%.

For the formation of limonene out of ß-pinene Ma et al[202] proclaimed a 76% product selectivity with almost full conversion using an MCM-22 catalyst. Slightly lower selectivities (72.6%) were achieved with wild type *Apa*SHC, however with lower activities than *Aac*SHC. The wild type enzyme *Bja*SHC and variant L36A in *Aac*SHC showed higher activities but with lower product selectivities of 67% and 66.4%, respectively.

Some of the reported results in literature are very promising and show especially higher conversion rates. Nevertheless, single mutations in *Aac*SHC can already shift the product spectrum and relative conversion dramatically. With additional mutations the activity could also be enhanced while mostly maintaining product selectivities (see chapter 4.1.4). Moreover, in contrast to classical chemical catalysts, *Aac*SHC is more accessible for manipulations due to it's proteinous nature. This demonstrates the high potential of SHCs for a high product selectivity and good conversion rates.

Chemical isomerizations and isomerizations within cyclases also use different tools to direct the reaction path. It is reported for classical chemical isomerizations that non-aqueous conditions favor ring expansion reactions to bornyl and fenchyl cations, whereas aqueous solvents favor ring opening reactions to *p*-menthenyl cation.[175] Within *Aac*SHC it seems that this is addressed more subtle by conformational changes of directing aromats, which is more manipulable for a more precise direction.

Another approach found in literature already combines the acidic moiety with a defined scaffold. The Tiefenbacher group produced a self-assembled resorcinarene capsule wherein the reaction takes place. The capsule possesses Brønsted acid functionalities and is able to stabilize cations via cation-π interactions[203]. Further, they showed an acid-based cyclization of linear monoterpenes within the capsule[204] and control of the products by different leaving groups of the substrate[205]. These results are promising and highlight the need for a sterically demanding environment to steer intermediate carbocations. However, in contrast to these reactions, which require reagents with different leaving groups to steer the reaction, the results of this thesis are solely achieved by catalyst design while using a singular substrate.

5.1.3.2 Re-screening of preliminary studies

It was important to identify that integration of small peaks in the chromatograms is responsible for some of the variations in product selectivities in preliminary studies and in chapter 4.1.2.2. There are always minor variations when manually integrating peaks. However, if the resulting peak areas are very small the signal to noise ratio has a huge impact on the overall result, which leads to major differences in product distributions. Therefore, it is important to increase peak areas. With higher peak areas minor variations in the integration have not such a significant impact.

One possibility is to increase the activity of variants, which was accomplished in chapter 4.1.3.5. Further, it was crucial to improve sample preparation procedures to reduce evaporation. Interestingly, the increase in signal intensities also led to the identification of ever new trace products (chapter 4.1.2.4). It seems *Aac*SHC is able to produce various products out of (+)-ß-pinene and there might be even more than so far detected or anticipated (chapter 4.1.2.1). Thus, it became a game of cat-and-mouse between identifying new trace products and developing new and improved protocols to be able to integrate them reliably. On the other hand, improved methods lead to the identification of new trace products, which would start the cycle again.

Variants that were identified to be selective towards one product remain specific throughout all these changes in protocols since newly emerged side products normally do not account for many percentage points of the total products in relation to the main product. Therefore, the trend always remains, whereas the exact numbers vary slightly. A summary of this is schematically depicted in Figure 62.

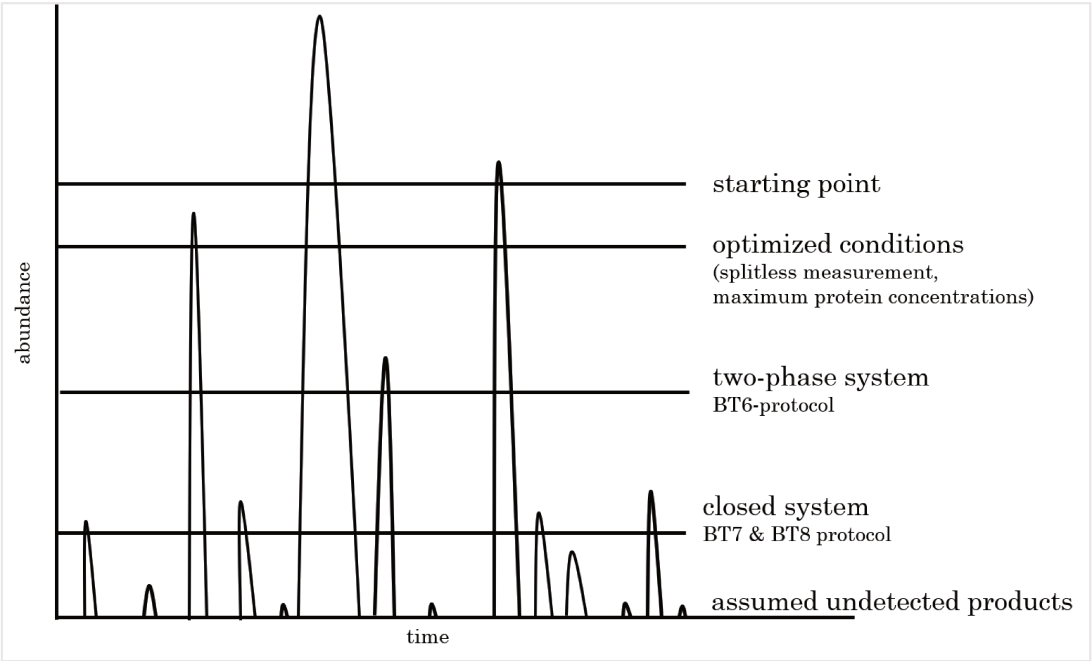


Figure 62: Schematic depiction of a fictional chromatogram to highlight the detection limits of various protocols. The horizontal lines represent what was detectable with the respective protocol.

5.1.3.3 Investigation of negative controls

The reason for a slight conversion of ß-pinene to α-terpineol with variant D376C is unclear. The double bond of ß-pinene is expected to require a strong acid for protonation for which a cysteine does not qualify. One explanation is that surrounding amino acids that are responsible for the *anti*-orientation of D376's carboxylic acid increase also the proton-donating ability of cysteine. A subsequent question is why only α-terpineol is formed and none of the other products. Especially, α-pinene should kinetically be favored. Moreover, in other previous studies using different substrates and D376C variant as negative control, a formation of corresponding products was not observed[109,110,153,158]. On the other hand, a complete destruction of the protonation machinery, termed 'AxAA' (D374A_D376A_D377A) abolishes the formation of α-terpineol compared to variant D376C. This points to an enzymatically catalyzed reaction. Moreover, it was observed that variant L36A, which shows a shifted product distribution towards limonene also produces greater amounts of α-terpineol than the wild type (Figure 63). Both products are formed by the same intermediate carbocation (*p*-menthenyl cation, Figure 64). Additionally, variant Y609W mainly produces α-pinene while the formation of *p*-menthenyl cation is suppressed, which results in lower amounts of limonene and α-terpineol. Taken together, this suggests a controlled reaction.

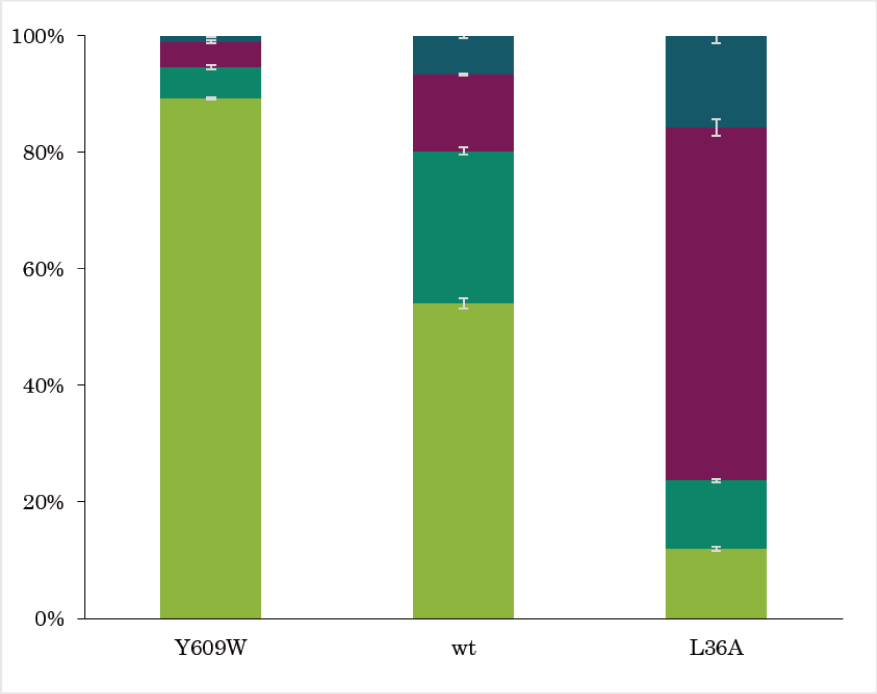


Figure 63: Comparison of variants Y609W, L36A and wt for the formation of α-terpineol for the conversion of β-pinene. Product distributions are shown for ■ α-pinene, ■ β-camphene, ■ limonene, ■ α-terpineol. All data were obtained using BT2 protocol and GC-P1 temperature program.

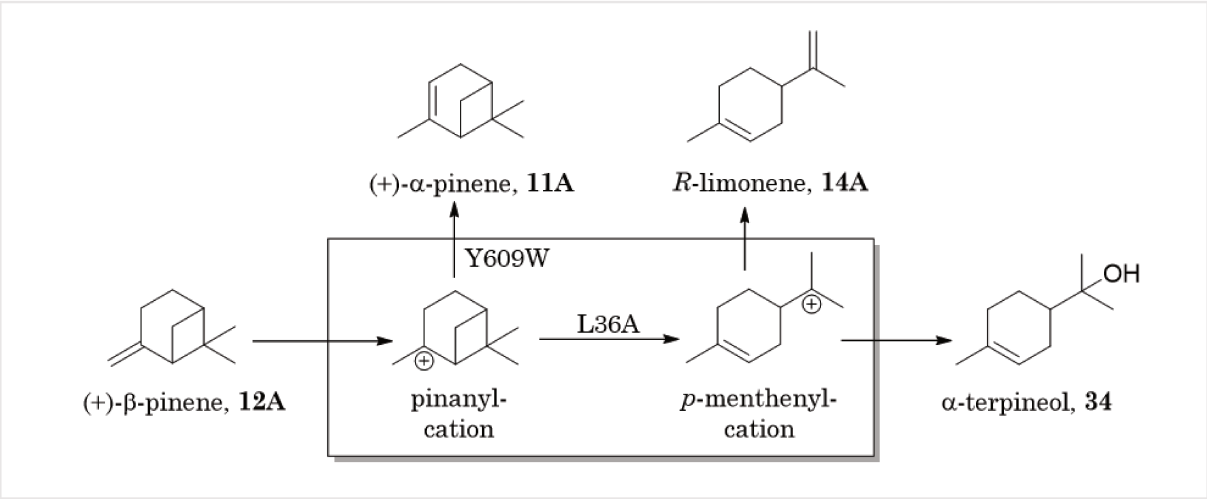


Figure 64: Formation of α-terpineol with respect to intermediate carbocations shown in the box. The pinanyl cation seems to be a crossroad here.

Since α-terpineol can be found in almost all variants except for D374A, D376A, D377A and D374A_D376A_D377A this strongly suggests that these positions are crucial for the formation of α-terpineol and it should be further investigated why D376C is able to form α-terpineol and no other products from ß-pinene.

5.1.3.4 Detection of trace products

The most interesting result of biotransformations using W312A is the suppression of the otherwise main products α-pinene, ß-camphene and limonene. Limonene is still the main product with 22%. However, more astonishing is the product diversity and presence of monoterpenes that require complex rearrangements, like fenchol and borneol. It is also remarkable how many different products from *p*-menthenyl cation like α-terpinolene and α-terpineol, are formed. Further, products are formed from terpinen-4-yl cation which originate from *p*-menthenyl cation as well, like γ-terpinene, α-terpinene and terpinen-4-ol (Figure 14).

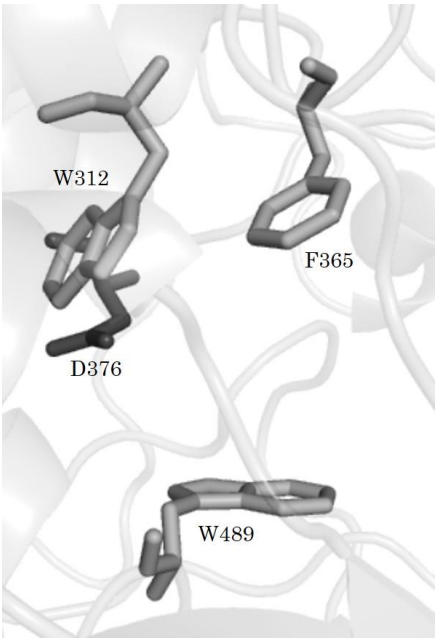


Figure 65: Residues F365, W489, W312 and D376 in the crystal structure PDB:1UMP. The aromatic residues can be seen as an "open cage" around an intermediate carbocation (not shown).

One possible explanation for this plethora of products might be the removal of a directing aromatic residue. The aromatic residues F365 and W489 were identified by Christian Curado in chapter 4.1.3.4 to stabilize and direct intermediate carbocations. Looking at the crystal structure PDB:1UMP[85], W312 can be seen as an additional barrier, forming an "open cage" of aromats with F365 and W489 (Figure 65). By changing W312 to a non-aromatic residue (W312A) allows intermediate carbocations to react along other pathways, thus enabling new rearrangements. This might also explain the lower activity of W312A. The combination with variants such as Y420F or Q366E that increase activity, could access new reaction paths and variants that are selective for e.g. borneol.

5.1.3.5 Determination of enantioselectivity

*Aac*SHC forms (+)-α-pinene (**11A**) and (*R*)-limonene (**14A**) out of (+)-ß-pinene (**12A**) with >99% enantiomeric excess. This supports the concept of possible reaction paths and cation rearrangements, elaborated in chapter 4.1.2.1. If the stereocenter in this reaction never becomes positively charged and thus never shows a sp^2 orbital hybridization, the original stereo information in maintained throughout the reaction (Figure 66).

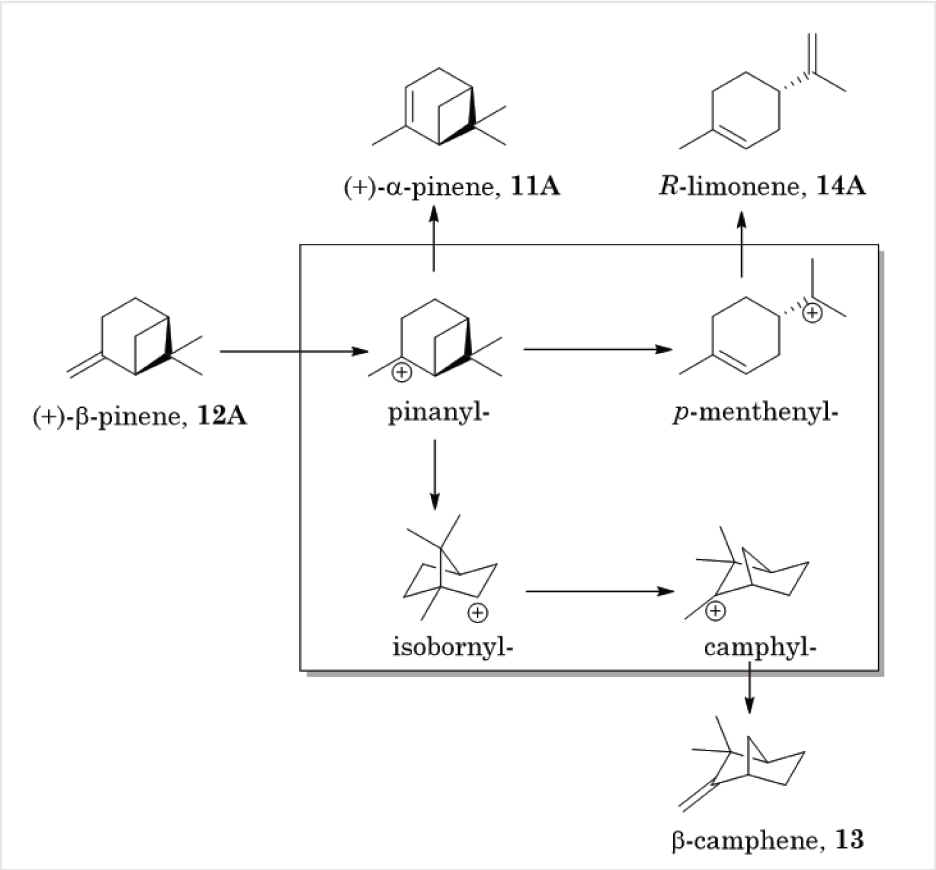


Figure 66: Reaction path from (+)-β-pinene to different products with respect to the stereo information.

5.1.4 Enzyme engineering

In contrast to the general availability of directed evolution methods, a semi-rational approach for enzyme engineering was conducted in this thesis. Usually, a distinct feature is necessary for a high-throughput screening that is vital to measure the vast number of mutants generated by an evolutionary approach[134]. Such features are missing for this project. The initial main products α-pinene, ß-camphene and limonene differ only slightly in their physio-chemical behavior and are thus difficult to distinguish. A fast GC method is also not feasible due to very similar elution temperatures. Yet, the more products were discovered, the longer the methods had to become to separate them. Additionally, initial issues with reproducibly integrating small peaks shifted the focus to developing a robust and reliable system rather than developing a fast screening system. Nevertheless, some time-saving methods were developed, as well (see chapter 4.1.1). This resulted in the development of semi-rational approaches with small and focused libraries and the attempt to rationalize the system by computational calculations with the group of Prof. Sílvia Osuna.

5.1.4.1 Combinatorial approach

The variant Y609I showed a high selectivity for limonene, however a high error bar was measured in the bachelor thesis of Josip Tulumovic.[168] Therefore, the results of the saturation at position Y609 were investigated for their selectivity towards limonene. The results are depicted in Figure 67.

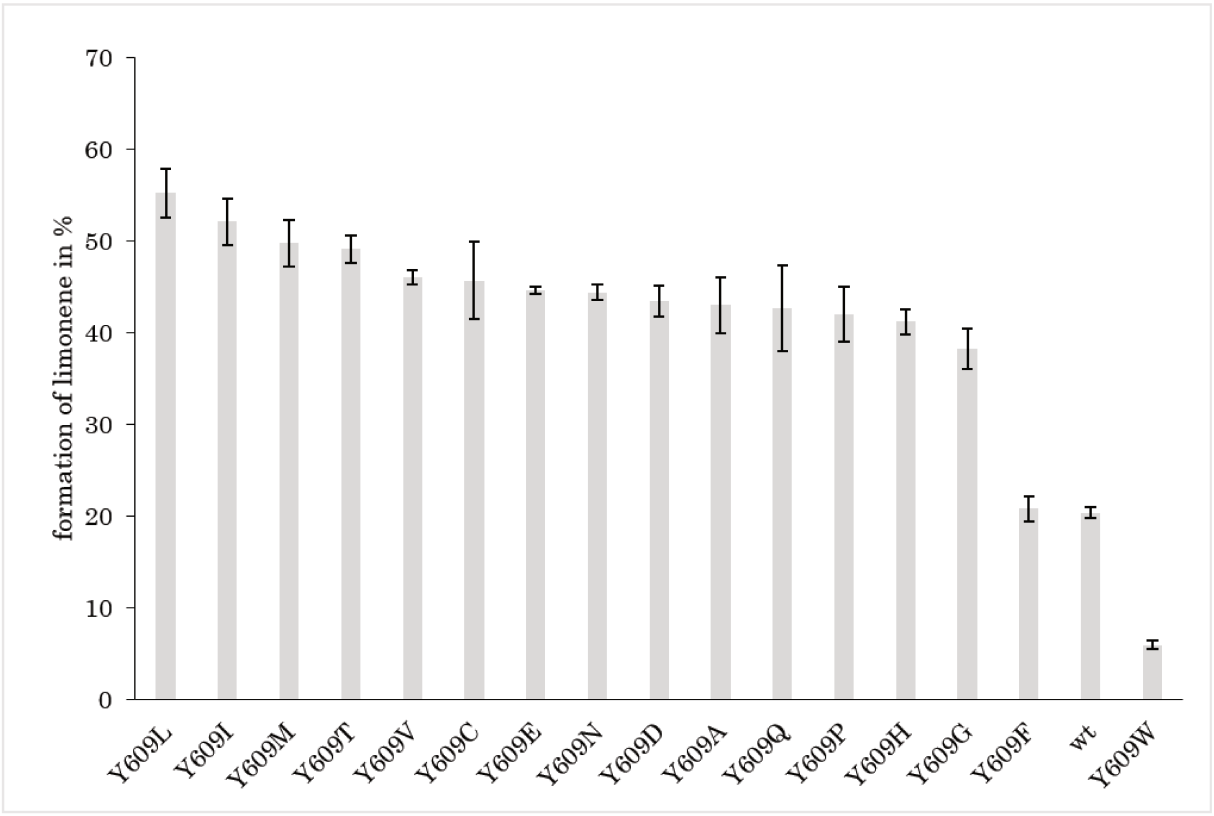


Figure 67: Selectivity of the saturation mutagenesis at position Y609 in the formation of limonene from β-pinene. All data were obtained using BT7 protocol and GC-P4 temperature program.

It is remarkable, that variants with no aromatic side chain show quite high selectivities in the formation of limonene (38-55%), whereas all aromatic residues at this position dramatically decrease the formation of limonene (5-21%). This confirms the initial idea of depleting a cation-π interaction at this position that was in favor of pinanyl cation. Since no influence of Y609I on the directing aromatic residues F365 and W489 was found by MD simulations (performed by Christian Curado) it can be concluded that the effects of introduced mutations in *Aac*SHC are more complex. They can originate from indirect interactions by directing aromats, as well as simple electrostatic interactions with the cations. Moreover, the active site of the enzyme is located on flexible loops, which makes MD simulations with short lived cations und hypothesis based on these simulations even more difficult. Additionally, it is interesting to see how position Y609 seems to act as a "switch" between the preference of either pinanyl or *p*-menthenyl cation depending on whether an aromatic moiety is present at the position, or not.

All combinatorial variants in combination with A419D and A419D itself were inactive. A deeper analysis using the protein structure software package PyMOL suggested sterical clashes of all rotamers with other residues. This might hamper a correct folding of the enzyme and thus lead to the observed loss of catalytic activity.

The residue F365 was addressed in the approach as well, which was later discovered to be important for directing intermediate carbocations (see chapter 4.1.3.4). A mutation to leucine did not affect the activity or the selectivity in combination with L36A significantly. MD simulations (chapter 4.1.3.4) predicted that the introduction of L36A brings F365 in a conformation which stabilizes the *p*-menthenyl cation by cation-π interactions of the aromat. Based on this hypothesis a drop in the selectivity for limonene was expected for L36A_F365L, which did not occur. This is in contrast to computational predictions and can so far not be explained.

In general, the approach to transfer a 'polar pocket' with an increased selectivity towards limonene from a monoterpene synthase (class I enzyme) to *Aac*SHC, a triterpene cyclase (class II enzyme), was not successful. Important residues and thus their selectivities were successfully exchanged in between monoterpene synthases[80,206]. However, the leap from a class I to a class II enzyme seemed to be too great.

5.1.4.2 Second shell library

Massive interferences were introduced into the enzyme by creating the 2nd shell library. Thus, it was surprising that 94% of variants were active. The SHC evolved as selective catalyst for the cyclization of squalene. Thus, the isomerization of pinene, which still needs to be optimized, leaves room for changes without affecting the general reaction.

5.1.4.3 Saturation mutagenesis

It was noticeable that almost all active variants at position L36 showed higher selectivities towards limonene than the wild type. Moreover, a clear correlation between the size of amino acids at position L36 and their effect on the formation of limonene was remarkable. Smaller amino acids such as alanine seem to favor limonene formation, whereas selectivity continuously reduced with the length of the side chain. However, the smallest amino acid glycine showed no conversion at all.

A similar behavior was observed for variants at position D442. D442P showed the highest product selectivities towards ß-camphene. Additionally, all active variants were more selective than the wild type. This comes as quite a surprise since the position is in the second shell of the active site and not at all in contact with the substrate. It is unknown if the same influencing network identified for D442N by Christian Curado (chapter 4.1.3.4) is also responsible for the selectivity of D442P. Moreover, the formation of camphyl cation requires several precise rearrangements than for the formation of the *p*-menthenyl cation. Thus, making variants with high ß-camphene content valuable.

Furthermore, the fact that for the saturation mutagenesis at position Y609 only Y609W produced significant amounts of α-pinene supports the hypothesis of a precise hydrogen bond network and the hydrogen bond network hypothesis by Christian Curado (chapter 4.1.3.4.)

5.1.4.4 In-silico studies

The identification of F365 and W489 by Christian Curado as positioning aromatic residues was a revelation. In the natural reaction, intermediate carbocations are also stabilized by aromatic residues in the cyclase. Therefore, it was interesting to identify that aromatic residues are again responsible for the stabilization and direction of intermediate carbocations. The attempt to cleave the hydrogen bond between Y609W and S309 in variant Y609W_S309A to verify the hypothesis was not successful. However, it revealed that position S309 is important for the activity of the enzyme in the conversion of ß-pinene. Unfortunately, it is not possible to break the hydrogen bond by mutagenesis to test the hypothesis without affecting the sterical environment of the active site. Moreover, knowledge on fine tuning the aromatic residues F365 and W489 by variants L36A, Y609W and D442N fits well with the model of the energetically very similar intermediate carbocations (see chapter 1.2.2). This also highlights the advantage of an enzyme to address subtle differences, since the complex sterical situation in an enzyme can provide a better scaffold for the necessary fine tuning.

The variants P444D and Q366S were suggested for a high selectivity towards ß-camphene. Unfortunately, this was not verified experimentally. However, position Q366 proved to be important for increasing the relative conversion, as shown for Q366E (chapter 4.1.3.5.) This shows that simulations are powerful tools. However, they still need to be refined for more accurate predictions to substitute in part traditional experimental approaches.

5.1.4.5 Exploration of the natural diversity of SHCs

Although the investigated SHCs originated from very different hosts they showed similar behavior in their product spectrum. In Figure 68, a phylogenetic tree for bacteria is shown that highlights how much the hosts differ from one another. The *Aac*SHC originates from A*licyclobacillus acidocaldarius,* which belongs to the class of bacilli[207]. Most SHCs of the other host strains in this thesis belong to the phylum of proteobacteria. *Zymomonas mobilis*, *Rhodopseudomonas palustris*, *Acetobacter pasteurianus* and *Bradyrhizobium japonicum* belong to α-proteobacteria[208], whereas *Pelobacter carbinolicus* is a representative from δ-proteobacteria[209]. Phylogenetically further away are *Thermosynechococcus elongatus,* which belongs to cyanobacteria[210] and *Streptomyces coelicolor,* which belongs to actinobacteria[211].

This leads to the hypothesis that the ability to isomerize ß-pinene among SHCs seems to be quite universal. Even similar isomerization products were obtained. However, distribution ratios differed. This highlights again how easily the equilibria between intermediate carbocations can be shifted by minor variations in the protein scaffold.

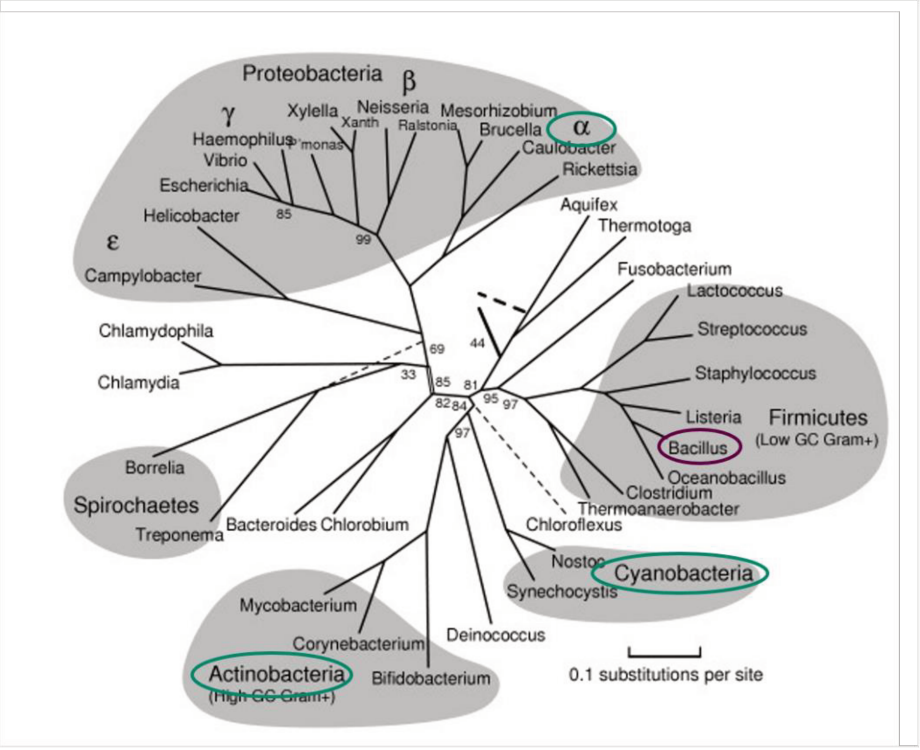


Figure 68: Phylogenetic tree of bacteria[212]. Highlighted are the phylogenetic classifications α-proteobacteria, actinobacteria, cyanobacteria (green) and bacillus (purple).

The higher activities of *Bja*SHC and *Tel*SHC were fortunate and the discovery that variations with high selectivities can be combined with variants that show higher relative conversion was a breakthrough. This opens up whole new possibilities for future mutations and might also lead to the identification of new terpene products. So far, it still needs to be tested whether beneficial positions of *Bja*SHC can also be transferred to *Aac*SHC, as was shown for *Tel*SHC. If these higher activities can also be transferred, L36A and W312A can be a good starting point. *Bja*SHC already showed a high selectivity towards limonene. Combination with L36A or W312A from *Aac*SHC, that both favor the *p*-menthenyl cation, could lead to a very selective and active variant for limonene or other *p*-menthenyl cation derived products.

The variants Q366E and Y420F in *Aac*SHC are both suitable for increasing the relative conversion. However, Y420F is more effective in combination with L36A or D442N, whereas Q366E is more active in combination with Y609W.

Interestingly, 96% of SHCs possess a phenylalanine at the corresponding position to Y420 in *Aac*SHC. In the SHC from *Zymomonas mobilis* this is even associated with a more broad product spectrum[178]. Sato et al[100] also investigated the effect of Y420F in *Aac*SHC on the natural reaction of SHCs. They reported that variant Y420F is slightly more thermostable than the wild type. Additionally, it showed a higher affinity to the substrate squalene but a 3.6-fold reduced reaction velocity. Overall, the position is responsible to stabilize a

carbocation at C10 for the formation of the C ring in the conversion of squalene (see chapter 1.3.2). Thus, it can be assumed that smaller monoterpenes will not come into direct contact with this residue during the reaction. This makes the effect even more puzzling. Since Y420F in *Aac*SHC slightly increases the thermostability of the enzyme[100] it can only be assumed that the mutation alters the whole structure slightly and thus also affects the conversion efficiency of monoterpenes. This makes such effects and mutations even harder to predict in the future.

5.1.5 Target molecule: Terpinen-4-ol

It should be noted that for initial experiments sabinene was used as a 75% pure liquid with ß-pinene being its main contaminant. However, no typical products of a ß-pinene conversion like α-pinene, ß-camphene or limonene were observed. Additionally, the conversion of sabinene was much higher than the general conversion of pure (+)-ß-pinene but when using a pure powdery sabinene no conversion was observed at all.

Even more puzzling is the presence of ß-pinene in the 75% pure sample and that none of the previous main products for it were found in the product spectrum. One reason might be that the cyclase prefers sabinene over ß-pinene. Another explanation might be that the products are again stemming from ß-pinene, which reacts differently in the presence of sabinene in the active site. In the case of two monoterpenes present in the active site a dimerization would be more likely. No dimerization products were observed for any of the substrates.

Furthermore, it should be noted that although limonene was not observed to be converted using BT5 protocol, it was later shown by Dr. Nico Kress (unpublished) that a slight conversion can be observed when using BT8 protocol and a great amount of purified catalyst. The resulting products were ß-camphene, α-terpinene, γ-terpinene, terpinolene, α-terpineol and terpinen-4-ol in trace amounts. This proves that the basic concept is valid and that *Aac*SHC can in general accept more monoterpenes as substrates than ß-pinene and sabinene, however, with dramatically reduced conversion rates. The goal should therefore be to enhance the activity towards other monoterpenes.

Moreover, it is interesting to see that besides products stemming from *p*-menthenyl and terpinen-4-ol cation also ß-camphene was formed. This requires major rearrangements and the formation of a strained ring, proving that the cyclase can enhance these rearrangements.

On the other hand, a strained ring system seems to pose an advantage as a substrate due to the inherent reactivity. This is supported by the fact, that both substrates (ß-pinene and sabinene), which showed good conversions, had strained ring systems. This tension was released for the formation of mainly single six membered ring products. Therefore, other substrates with strained rings like 3-carene or ß-camphene might be suitable substrates as well, although initial tests with BT5 protocol showed so far no conversions of these substrates.

5.1.6 Library screening and product characterization

Some of the observed products of a biotransformation with sabinene were not verified by a standard. They are either not commercially available or too expensive. However, this would make their biotechnological production using *Aac*SHC even more attractive. V448F is suggested for the identification of P1, P8 and P9, G600F for P5 and P10, L36W for P7 and F365W for P4 (see Table 17). In the future these variants could then be coupled with a microbial production of sabinene for which some success was reported[213]. Thus, valuable monoterpenes that are so far not commercially available could be synthesized from inexpensive media as feed stock.

The difference in product distribution over time can be explained that some products can act also in substrates in AacSHC-catalyzed reactions. This would fit the observed results nicely since deprotonation products were seen to decrease over time, whereas products with a hydroxyl group increase over time. Only products with a double bond can be protonated again. A hypothetical scheme is presented in Figure 69. Thereby, the potential products α-thujene, α-terpinene and γ-terpinene would serve as both products and substrates, whereas sabinene hydrate and terpinen-4-ol would only be products. The supposed product *cis*-ß-terpineol is excluded from this depiction since it is not clear which carbocation it passes through.

However, it should be noted that previous attempts to convert α-terpinene and γ-terpinene as sole substrates were not successful, which might be again due to the BT5 protocol used.

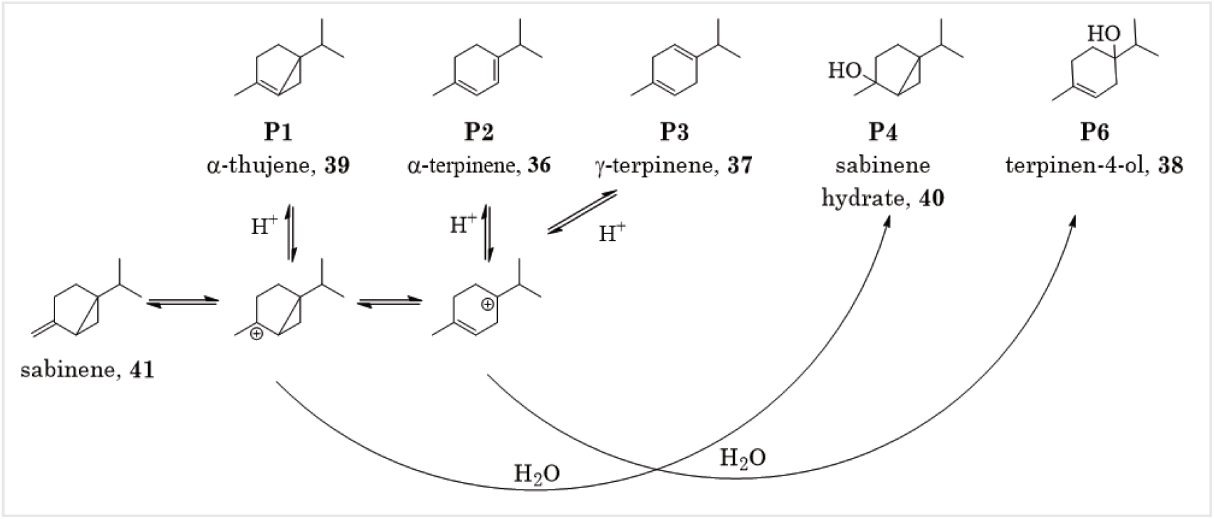


Figure 69: Hypothetical transformation over time for the conversion of sabinene.

6 Outlook

In chapter 4.1.2.4 it was shown that W312A produces a variety of interesting terpene products that appear only in trace amounts in other variants that were investigated in this study. Thus, the most obvious next step is to saturate position W312. With an improved reaction setup and negative control, it was shown that it produces a considerable amount of α-terpineol but also a bouquet of interesting side products. Therefore, a saturation at this position might reveal high product selectivities towards one of these products. Additionally, it would be worthwhile combining W312A with variants that proved to increase the activity such as Y420F or Q366E. Moreover, an *in-silico* analysis of this variant and it's influence on the directing aromatic amino acids F365 and W489 could be of interest.

The transfer of positions from *Tel*SHC into *Aac*SHC achieved success in respect to increase the activity of *Aac*SHC variants. However, in the initial analysis the tested variants were identified within a radius of 8 Å around the docked substrate. In the same analysis 42 potential variants were detected in a radius of 12 Å around the substrate. Out of these, 10 are of special interest. The variants G34A, P35E, Q262I, W265M and A306P are in close contact to S307, which was already identified to have a great impact on the reaction. In close proximity to D442, a position that was determined to influence the selectivity towards ß-camphene, are G418A, E439A and T441I. Further, A499G, F363W and D368E were selected, which are in close contact to the directing aromatic residues F365 and W489. Therefore, it would be very interesting to create and test these variants.

Moreover, a similar analysis for *Bja*SHC could provide interesting results. In contrast to *Tel*SHC, *Bja*SHC showed an inherent product selectivity towards limonene. Variants originating from a transfer analysis with *Bja*SHC would also be suitable parents for combinations with L36A or W312A. Both seem to favor a transition via *p*-menthenyl cation and such combinations could thus further promote the corresponding products.

Additionally, a continued collaboration with Prof. Sílvia Osuna and her group would be most valuable to increase the understanding of the complex interactions introduced by variants and to rationally predict suitable variants.

With ever improving methods to reduce the evaporation of both substrate and products during the biotransformation, it would be important to quantify the reaction. Several attempts were made on accurately and reproducibly quantifying the products. For a precise quantification a retrieval of >90% should be aimed for, which was so far hard to achieve (see chapter 4.1.1.5).

For the conversion of sabinene, it would be valuable to identify the products by preparative biotransformations, which could so far not be properly characterized. This harbors the great potential to biotechnologically produce products that can so far not be purchased commercially. Subsequently, similar approaches that were successful for ß-pinene are advised for sabinene as substrate. An *in-silico* analysis of the reaction might reveal interesting new insights, especially compared to findings from the conversion of ß-pinene. Experimentally, it would be valuable to further investigate the course of the reaction over time and test α-terpinene and γ-terpinene as substrates with the improved methods BT7

and BT8. Moreover, a screen with sabinene using SHCs from other organisms might be similarly successful regarding selectivity and activity as it was for the conversion of (+)-ß-pinene. Particularly, since sabinene already showed higher inherent relative conversion than ß-pinene.

7 Literature

1. Christianson, D. W. Structural and Chemical Biology of Terpenoid Cyclases. *Chem. Rev.* **117**, 11570–11648 (2017).
2. Christianson, D. W. Unearthing the roots of the terpenome. *Curr. Opin. Chem. Biol.* **12**, 141–150 (2008).
3. Seitz, M. *et al.* Synthesis of Heterocyclic Terpenoids by Promiscuous Squalene-Hopene Cyclases. *ChemBioChem* **14**, 436–439 (2013).
4. Jiang, Z., Kempinski, C. & Chappell, J. Extraction and Analysis of Terpenes/Terpenoids. *Curr. Protoc. Plant Biol.* **1**, 345–358 (2016).
5. Christianson, D. W. Structural Biology and Chemistry of the Terpenoid Cyclases. *Chem. Rev.* **106**, 3412–3442 (2006).
6. Withers, S. T. & Keasling, J. D. Biosynthesis and engineering of isoprenoid small molecules. *Appl. Microbiol. Biotechnol.* **73**, 980–990 (2007).
7. Rohdich, F., Kis, K., Bacher, A. & Eisenreich, W. The non-mevalonate pathway of isoprenoids: genes, enzymes and intermediates. *Curr. Opin. Chem. Biol.* **5**, 535–540 (2001).
8. Siedenburg, G. & Jendrossek, D. Squalene-Hopene Cyclases. *Appl. Environ. Microbiol.* **77**, 3905–3915 (2011).
9. Poralla, K., Kannenberg, E. & Blume, A. A glycolipid containing hopane isolated from the acidophilic, thermophilic bacillus acidocaldarius , has a cholesterol-like function in membranes. *FEBS Lett.* **113**, 107–110 (1980).
10. Schmerk, C. L., Bernards, M. A. & Valvano, M. A. Hopanoid Production Is Required for Low-pH Tolerance, Antimicrobial Resistance, and Motility in Burkholderia cenocepacia. *J. Bacteriol.* **193**, 6712–6723 (2011).
11. Bae, S.-H., Lee, J. N., Fitzky, B. U., Seong, J. & Paik, Y.-K. Cholesterol Biosynthesis from Lanosterol. *J. Biol. Chem.* **274**, 14624–14631 (1999).
12. Gershenzon, J. & Dudareva, N. The function of terpene natural products in the natural world. *Nat. Chem. Biol.* **3**, 408–414 (2007).
13. Wagner, K.-H. & Elmadfa, I. Biological Relevance of Terpenoids. *Ann. Nutr. Metab.* **47**, 95–106 (2003).
14. Rezanka, T., Siristova, L., Melzoch, K. & Sigler, K. Hopanoids in Bacteria and Cyanobacteria – Their Role in Cellular Biochemistry and Physiology, Analysis and Occurrence. *Mini. Rev. Org. Chem.* **7**, 300–313 (2010).
15. Pichersky, E. & Raguso, R. A. Why do plants produce so many terpenoid compounds? *New Phytol.* **220**, 692–702 (2018).
16. Ajikumar, P. K. *et al.* Terpenoids: Opportunities for Biosynthesis of Natural Product Drugs Using Engineered Microorganisms. *Mol. Pharm.* **5**, 167–190 (2008).
17. Tu, Y. The discovery of artemisinin (qinghaosu) and gifts from Chinese medicine. *Nat. Med.* **17**, 1217–1220 (2011).
18. Nobel-Prize. Nobelprize for her discoveries concerning a novel therapy against Malaria. *Tu Youyou* Nobel Media AB 2014 (2014). Available at: http://www.nobelprize.org/nobel_prizes/medicine/laureates/2015/press.html. (Accessed: 17th February 2016)
19. Schäfer, B. Ambrox®. *Chemie unserer Zeit* **45**, 374–388 (2011).
20. Breuer, M., Horster, A. & Hauer, B. BIOCATALYTIC PRODUCTION OF AMBROXAN. *US 8,759,043*

B2 United States Patent (2014).

21. Pichersky, E., Noel, J. P. & Dudareva, N. Biosynthesis of Plant Volatiles: Nature's Diversity and Ingenuity. *Biosynth. Plant Volatiles Nature's Divers. Ingen.* **311**, 808–811 (2006).
22. Crowell, P. L. Prevention and Therapy of Cancer by Dietary Monoterpenes. *J. Nutr.* **129**, 775S-778S (1999).
23. Garcia, R. *et al.* Antimicrobial activity and potential use of monoterpenes as tropical fruits preservatives. *Braz. J. Microbiol.* **39**, 163–8 (2008).
24. de Cássia da Silveira e Sá, R., Andrade, L. N. & de Sousa, D. P. A Review on Anti-Inflammatory Activity of Monoterpenes. *Molecules* **18**, 1227–1254 (2013).
25. Sobral, M. V., Xavier, A. L., Lima, T. C. & de Sousa, D. P. Antitumor Activity of Monoterpenes Found in Essential Oils. *Sci. World J.* **2014**, 1–35 (2014).
26. Echeverrigaray, S., Zacaria, J. & Beltrão, R. Nematicidal Activity of Monoterpenoids Against the Root-Knot Nematode Meloidogyne incognita. *Phytopathology* **100**, 199–203 (2010).
27. Zahran, H. E.-D. M. & Abdelgaleil, S. A. M. Insecticidal and developmental inhibitory properties of monoterpenes on Culex pipiens L. (Diptera: Culicidae). *J. Asia. Pac. Entomol.* **14**, 46–51 (2011).
28. Mercier, B., Prost, J. & Prost, M. The essential oil of turpentine and its major volatile fraction (α- and β-pinenes): a review. *Int. J. Occup. Med. Environ. Health* **22**, 331–342 (2009).
29. Corma, A., Iborra, S. & Velty, A. Chemical Routes for the Transformation of Biomass into Chemicals. *Chem. Rev.* **107**, 2411–2502 (2007).
30. Nobel-Prize. Nobelprize for their work on chirally catalysed hydrogenation reactions. *Ryoji Noyori* Nobel Media AB 2001 (2001). Available at: http://www.nobelprize.org/nobel_prizes/chemistry/laureates/2001/noyori-lecture.pdf. (Accessed: 1st March 2016)
31. Kreß, N. Development of a Chemoenzymatic (-)-Menthol Synthesis. (Universität Stuttgart, 2018).
32. Kress, N., Rapp, J. & Hauer, B. Enantioselective Reduction of Citral Isomers in NCR Ene Reductase: Analysis of an Active-Site Mutant Library. *ChemBioChem* **18**, 717–720 (2017).
33. Bastian, S. A., Hammer, S. C., Kreß, N., Nestl, B. M. & Hauer, B. Selectivity in the Cyclization of Citronellal Introduced by Squalene Hopene Cyclase Variants. *ChemCatChem* **9**, 4364–4368 (2017).
34. Filipsson, F. A., Bard, J. & Karlsson, S. *Concise International Chemical Assessment Document 5: Limonene*. *World Health Organization* (1998).
35. Räisänen, T., Ryyppö, A. & Kellomäki, S. Monoterpene emission of a boreal Scots pine (Pinus sylvestris L.) forest. *Agric. For. Meteorol.* **149**, 808–819 (2009).
36. Zhao, D. F. *et al.* Environmental conditions regulate the impact of plants on cloud formation. *Nat. Commun.* **8**, 14067 (2017).
37. Huff Hartz, K. E. *et al.* Cloud condensation nuclei activation of monoterpene and sesquiterpene secondary organic aerosol. *J. Geophys. Res. Atmos.* **110**, (2005).
38. Vespermann, K. A. C. *et al.* Biotransformation of α- and β-pinene into flavor compounds. *Appl. Microbiol. Biotechnol.* **101**, 1805–1817 (2017).
39. Sarria, S., Wong, B., Martín, H. G., Keasling, J. D. & Peralta-Yahya, P. Microbial Synthesis of Pinene. *ACS Synth. Biol.* **3**, 466–475 (2014).
40. Mäki-Arvela, P., Holmbom, B., Salmi, T. & Murzin, D. Y. Recent Progress in Synthesis of Fine and Specialty Chemicals from Wood and Other Biomass by Heterogeneous Catalytic Processes.

Catal. Rev. **49**, 197–340 (2007).

41. Martins, A. *et al.* Essential Oil Composition and Antimicrobial Activity of Santiria trimera Bark. *Planta Med.* **69**, 77–79 (2003).

42. Uribe, S., Ramirez, J. & Pena, A. Effects of β-pinene on yeast membrane functions. *J. Bacteriol.* **161**, 1195–1200 (1985).

43. Chen, W. *et al.* Anti-tumor effect of α-pinene on human hepatoma cell lines through inducing G2/M cell cycle arrest. *J. Pharmacol. Sci.* **127**, 332–338 (2015).

44. Perry, N. S. L. *et al.* In-vitro activity of S. lavandulaefolia (Spanish sage) relevant to treatment of Alzheimer's disease. *J. Pharm. Pharmacol.* **53**, 1347–1356 (2001).

45. Il'ina, I. V., Volcho, K. P. & Salakhutdinov, N. F. Acid-catalyzed transformations of pinane terpenoids. New prospects. *Russ. J. Org. Chem.* **44**, 1–23 (2008).

46. Hamidpour, R., Hamidpour, S., Hamidpour, M. & Shahlari, M. Camphor (Cinnamomum camphora), a traditional remedy with the history of treating several diseases. *Int. J. Case Reports Images* **4**, 86 (2013).

47. Krishnaswamy, N. R. Learning organic chemistry through natural products. *Resonance* **1**, 56–62 (1996).

48. Ponomarev, D. & Mettee, H. Camphor and its Industrial Synthesis. *Chem. Educ. J.* **18**, 2–5 (2016).

49. Chen, W., Vermaak, I. & Viljoen, A. Camphor—A Fumigant during the Black Death and a Coveted Fragrant Wood in Ancient Egypt and Babylon—A Review. *Molecules* **18**, 5434–5454 (2013).

50. Rahman, F. A., Priya, V., Gayathri, R. & Geetha, R. V. In vitro antibacterial activity of camphor oil against oral microbes. *Int. J. Pharm. Sci. Rev. Res.* **39**, 119–121 (2016).

51. Vallianou, I., Peroulis, N., Pantazis, P. & Hadzopoulou-Cladaras, M. Camphene, a Plant-Derived Monoterpene, Reduces Plasma Cholesterol and Triglycerides in Hyperlipidemic Rats Independently of HMG-CoA Reductase Activity. *PLoS One* **6**, e20516 (2011).

52. Tiwari, M. & Kakkar, P. Plant derived antioxidants – Geraniol and camphene protect rat alveolar macrophages against t-BHP induced oxidative stress. *Toxicol. Vitr.* **23**, 295–301 (2009).

53. Quintans-Júnior, L. *et al.* Antinociceptive Activity and Redox Profile of the Monoterpenes (+)-Camphene, p -Cymene, and Geranyl Acetate in Experimental Models. *ISRN Toxicol.* **2013**, 1–11 (2013).

54. Ibrahim, M. A. & Tiilikkala, K. Insecticidal , repellent , antimicrobial activity and phytotoxicity of essential oils : With special reference to limonene and its suitability for control of insect pests. *Agric. Food Sci. Finl.* **10**, 243–259 (2001).

55. Thomas, A. F. & Bessière, Y. Limonene. *Nat. Prod. Rep.* **6**, 291–309 (1989).

56. Zhu, Y., Romain, C. & Williams, C. K. Sustainable polymers from renewable resources. *Nature* **540**, 354–362 (2016).

57. Longo, M. A. & Sanromán, M. A. Production of food aroma compounds : microbial and enzymatic methodologies. *Food Technol. Biotechnol.* **44**, 335–353 (2006).

58. Flores-Holguín, N., Aguilar-Elguézabal, A., Rodríguez-Valdez, L. M. & Glossman-Mitnik, D. Theoretical study of chemical reactivity of the main species in the α-pinene isomerization reaction. *J. Mol. Struct. THEOCHEM* **854**, 81–88 (2008).

59. Gscheidmeier, M. & Häberlein, H. Process For The Preparation Of Camphene By The Rearrangement Of A-Pinene. *US 5,826,202* United States Patent (1998).

60. Findik, S. & Gündüz, G. Isomerization of α-Pinene to Camphene. *J. Am. Oil Chem. Soc.* **74**, 1145–1151 (1997).

61. Rachwalik, R. *et al.* Isomerization of α-pinene over dealuminated ferrierite-type zeolites. *J. Catal.* **252**, 161–170 (2007).

62. Alsalme, A., Kozhevnikova, E. F. & Kozhevnikov, I. V. α-Pinene isomerisation over heteropoly acid catalysts in the gas-phase. *Appl. Catal. A Gen.* **390**, 219–224 (2010).

63. Comelli, N. A., Ponzi, E. N. & Ponzi, M. I. Isomerization of α-pinene, limonene, α-terpinene, and terpinolene on sulfated zirconia. *JAOACS* **82**, (2005).

64. Beşün, N., Özkan, F. & Gündüz, G. Alpha-pinene isomerization on acid-treated clays. *Appl. Catal. A Gen.* **224**, 285–297 (2002).

65. Wang, J., Hua, W., Yue, Y. & Gao, Z. MSU-S mesoporous materials: An efficient catalyst for isomerization of α-pinene. *Bioresour. Technol.* **101**, 7224–7230 (2010).

66. Yadav, M. K., Chudasama, C. D. & Jasra, R. V. Isomerisation of α-pinene using modified montmorillonite clays. *J. Mol. Catal. A Chem.* **216**, 51–59 (2004).

67. Sidorenko, A. Y. *et al.* Catalytic isomerization of α-pinene and 3-carene in the presence of modified layered aluminosilicates. *Mol. Catal.* **443**, 193–202 (2017).

68. Allahverdiev, A. I., Irandoust, S. & Yu. Murzin, D. Isomerization of α-Pinene over Clinoptilolite. *J. Catal.* **185**, 352–362 (1999).

69. Swift, K. A. D. Catalytic Transformations of the Major Terpene Feedstocks. *Top. Catal.* **27**, 143–155 (2004).

70. Ebmeyer, F. Theoretical investigations towards an understanding of the α-pinene/camphene rearrangement. *J. Mol. Struct. THEOCHEM* **582**, 251–255 (2002).

71. Weitman, M. & Major, D. T. Challenges Posed to Bornyl Diphosphate Synthase: Diverging Reaction Mechanisms in Monoterpenes. *J. Am. Chem. Soc.* **132**, 6349–6360 (2010).

72. Pemberton, R. P. & Tantillo, D. J. Lifetimes of carbocations encountered along reaction coordinates for terpene formation. *Chem. Sci.* **5**, 3301 (2014).

73. Hyatt, D. C. *et al.* Structure of limonene synthase, a simple model for terpenoid cyclase catalysis. *Proc. Natl. Acad. Sci.* **104**, 5360–5365 (2007).

74. Degenhardt, J., Köllner, T. G. & Gershenzon, J. Monoterpene and sesquiterpene synthases and the origin of terpene skeletal diversity in plants. *Phytochemistry* **70**, 1621–1637 (2009).

75. Whittington, D. a *et al.* Nonlinear partial differential equations and applications: Bornyl diphosphate synthase: Structure and strategy for carbocation manipulation by a terpenoid cyclase. *Proc. Natl. Acad. Sci.* **99**, 15375–15380 (2002).

76. Kampranis, S. C. *et al.* Rational conversion of substrate and product specificity in a Salvia monoterpene synthase: structural insights into the evolution of terpene synthase function. *Plant Cell* **19**, 1994–2005 (2007).

77. Leferink, N. G. H. *et al.* Experiment and Simulation Reveal How Mutations in Functional Plasticity Regions Guide Plant Monoterpene Synthase Product Outcome. *ACS Catal.* **8**, 3780–3791 (2018).

78. Chen, F., Tholl, D., Bohlmann, J. & Pichersky, E. The family of terpene synthases in plants: a mid-size family of genes for specialized metabolism that is highly diversified throughout the kingdom. *Plant J.* **66**, 212–229 (2011).

79. Langenheim, J. H. Higher plant terpenoids: A phytocentric overview of their ecological roles. *J. Chem. Ecol.* **20**, 1223–1280 (1994).

80. Piechulla, B. *et al.* The α-Terpineol to 1,8-Cineole Cyclization Reaction of Tobacco Terpene Synthases. *Plant Physiol.* **172**, 2120–2131 (2016).

81. Hammer, S. C., Syrén, P.-O., Seitz, M., Nestl, B. M. & Hauer, B. Squalene hopene cyclases: highly promiscuous and evolvable catalysts for stereoselective CC and CX bond formation. *Curr. Opin.*

Chem. Biol. **17**, 293–300 (2013).

82. Gao, Y., Honzatko, R. B. & Peters, R. J. Terpenoid synthase structures: a so far incomplete view of complex catalysis. *Nat. Prod. Rep.* **29**, 1153 (2012).

83. Pronin, S. V. & Shenvi, R. A. Synthesis of highly strained terpenes by non-stop tail-to-head polycyclization. *Nat. Chem.* **4**, 915–920 (2012).

84. Wendt, K. U., Poralla, K. & Schulz, G. E. Structure and function of a squalene cyclase. *Science* **277**, 1811–1815 (1997).

85. Reinert, D. J., Balliano, G. & Schulz, G. E. Conversion of Squalene to the Pentacarbocyclic Hopene. *Chem. Biol.* **11**, 121–126 (2004).

86. Hoshino, T. & Sato, T. Squalene–hopene cyclase: catalytic mechanism and substrate recognition. *Chem. Commun.* 291–301 (2002).

87. Poralla, K. *et al.* A specific amino acid repeat in squalene and oxidosqualene cyclases. *Trends Biochem. Sci.* **19**, 157–158 (1994).

88. Sato, T. & Hoshino, T. Functional Analysis of the DXDDTA Motif in Squalene-Hopene Cyclase by Site-directed Mutagenesis Experiments: Initiation Site of the Polycyclization Reaction and Stabilization Site of the Carbocation Intermediate of the Initially Cyclized A-Ring. *Biosci. Biotechnol. Biochem.* **63**, 2189–2198 (1999).

89. Wendt, K. ., Lenhart, A. & Schulz, G. . The structure of the membrane protein squalene-hopene cyclase at 2.0 å resolution. *J. Mol. Biol.* **286**, 175–187 (1999).

90. Hammer, S. C., Marjanovic, A., Dominicus, J. M., Nestl, B. M. & Hauer, B. Squalene hopene cyclases are protonases for stereoselective Brønsted acid catalysis. *Nat. Chem. Biol.* **11**, 121–126 (2015).

91. Hammer, S. C., Dominicus, J. M., Syrén, P.-O., Nestl, B. M. & Hauer, B. Stereoselective Friedel–Crafts alkylation catalyzed by squalene hopene cyclases. *Tetrahedron* **68**, 7624–7629 (2012).

92. Kühnel, L. C., Nestl, B. M. & Hauer, B. Enzymatic Addition of Alcohols to Terpenes by Squalene Hopene Cyclase Variants. *ChemBioChem* **18**, 2222–2225 (2017).

93. Hoshino, T., Kumai, Y., Kudo, I., Nakano, S. & Ohashi, S. Enzymatic cyclization reactions of geraniol, farnesol and geranylgeraniol, and those of truncated squalene analogs having C 20 and C 25 by recombinant squalene cyclase. *Org. Biomol. Chem.* **2**, 2650–2657 (2004).

94. Syrén, P.-O., Henche, S., Eichler, A., Nestl, B. M. & Hauer, B. Squalene-hopene cyclases—evolution, dynamics and catalytic scope. *Curr. Opin. Struct. Biol.* **41**, 73–82 (2016).

95. Syrén, P., Hammer, S. C., Claasen, B. & Hauer, B. Entropy is Key to the Formation of Pentacyclic Terpenoids by Enzyme-Catalyzed Polycyclization. *Angew. Chemie Int. Ed.* **53**, 4845–4849 (2014).

96. Morikubo, N. *et al.* Cation−π Interaction in the Polyolefin Cyclization Cascade Uncovered by Incorporating Unnatural Amino Acids into the Catalytic Sites of Squalene Cyclase. *J. Am. Chem. Soc.* **128**, 13184–13194 (2006).

97. Pale-Grosdemange, C., Feil, C., Rohmer, M. & Poralla, K. Occurrence of Cationic Intermediates and Deficient Control during the Enzymatic Cyclization of Squalene to Hopanoids. *Angew. Chemie Int. Ed.* **37**, 2237–2240 (1998).

98. Dougherty, D. A. The Cation−π Interaction. *Acc. Chem. Res.* **46**, 885–893 (2013).

99. Mecozzi, S., West, A. P. & Dougherty, D. A. Cation-pi interactions in aromatics of biological and medicinal interest: electrostatic potential surfaces as a useful qualitative guide. *Proc. Natl. Acad. Sci.* **93**, 10566–10571 (1996).

100. Sato, T., Sasahara, S., Yamakami, T. & Hoshino, T. Functional Analyses of Tyr420 and Leu607 of Alicyclobacillus acidocaldarius Squalene-Hopene Cyclase. Neoachillapentaene, a Novel

Triterpene with the 1,5,6-Trimethylcyclohexene Moiety…. *Biosci. Biotechnol. Biochem.* **66**, 1660–1670 (2002).

101. Neel, A. J., Hilton, M. J., Sigman, M. S. & Toste, F. D. Exploiting non-covalent π interactions for catalyst design. *Nature* **543**, 637–646 (2017).

102. Kennedy, C. R., Lin, S. & Jacobsen, E. N. The Cation-π Interaction in Small-Molecule Catalysis. *Angew. Chemie Int. Ed.* **55**, 12596–12624 (2016).

103. Vattekkatte, A., Garms, S. & Boland, W. Alternate Cyclization Cascade Initiated by Substrate Isomer in Multiproduct Terpene Synthase from Medicago truncatula. *J. Org. Chem.* **82**, 2855–2861 (2017).

104. Xu, J. *et al.* Mutational analysis and dynamic simulation of S-limonene synthase reveal the importance of Y573: Insight into the cyclization mechanism in monoterpene synthases. *Arch. Biochem. Biophys.* **638**, 27–34 (2018).

105. Li, R. *et al.* Reprogramming the Chemodiversity of Terpenoid Cyclization by Remolding the Active Site Contour of epi -Isozizaene Synthase. *Biochemistry* **53**, 1155–1168 (2014).

106. Xu, J. *et al.* Converting S-limonene synthase to pinene or phellandrene synthases reveals the plasticity of the active site. *Phytochemistry* **137**, 34–41 (2017).

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

108. Vattekkatte, A. *et al.* Substrate geometry controls the cyclization cascade in multiproduct terpene synthases from Zea mays. *Org. Biomol. Chem.* **13**, 6021–6030 (2015).

109. Hammer, S. C. Zur Anwendbarkeit von Squalen-Hopen-Zyklasen als chirale Brønsted-Säuren in der asymmetrischen Katalyse. (Universität Stuttgart, 2014).

110. Seitz, M. Characterization of the substrate specificity of squalene-hopene cyclases (SHCs). (Universität Stuttgart, 2012).

111. Tanaka, H., Noguchi, H. & Abe, I. Enzymatic Formation of Indole-Containing Unnatural Cyclic Polyprenoids by Bacterial Squalene:Hopene Cyclase. *Org. Lett.* **7**, 5873–5876 (2005).

112. Tanaka, H., Noma, H., Noguchi, H. & Abe, I. Enzymatic formation of pyrrole-containing novel cyclic polyprenoids by bacterial squalene:hopene cyclase. *Tetrahedron Lett.* **47**, 3085–3089 (2006).

113. Cheng, J. & Hoshino, T. Cyclization cascade of the C33-bisnorheptaprenoid catalyzed by recombinant squalene cyclase. *Org. Biomol. Chem.* **7**, 1689 (2009).

114. Abe, I., Tanaka, H. & Noguchi, H. Enzymatic Formation of an Unnatural Hexacyclic C35 Polyprenoid by Bacterial Squalene Cyclase. *J. Am. Chem. Soc.* **124**, 14514–14515 (2002).

115. Hauer, B., Hammer, S. C. & Syrén, P.-O. Substrate pre-folding and water molecule organization matters for terpene cyclase catalyzed conversion of unnatural substrates. *Chem. Sel.* **1**, 3589–3593 (2016).

116. Henche, S. Squalen-Hopen Zyklasen vermittelte Friedel-Crafts Alkylierung. (Universität Stuttgart, 2019).

117. Eichhorn, E., Schilling, B., Wahler, D., Fourage, L. & Locher, E. Enzymes and applications thereof. *WO 2016170099 A1* International Patent (2016).

118. Hammer, S. C., Syrén, P.-O. & Hauer, B. Substrate Pre-Folding and Water Molecule Organization Matters for Terpene Cyclase Catalyzed Conversion of Unnatural Substrates. *ChemistrySelect* **1**, 3589–3593 (2016).

119. Siedenburg, G., Breuer, M. & Jendrossek, D. Prokaryotic squalene-hopene cyclases can be converted to citronellal cyclases by single amino acid exchange. *Appl. Microbiol. Biotechnol.* **97**,

1571–1580 (2013).

120. Miao, Y., Rahimi, M., Geertsema, E. M. & Poelarends, G. J. Recent developments in enzyme promiscuity for carbon–carbon bond-forming reactions. *Curr. Opin. Chem. Biol.* **25**, 115–123 (2015).

121. Reetz, M. T. Biocatalysis in Organic Chemistry and Biotechnology: Past, Present, and Future. *J. Am. Chem. Soc.* **135**, 12480–12496 (2013).

122. Sheldon, R. A. The E factor 25 years on: the rise of green chemistry and sustainability. *Green Chem.* **19**, 18–43 (2017).

123. Sheldon, R. A. & Woodley, J. M. Role of Biocatalysis in Sustainable Chemistry. *Chem. Rev.* **118**, 801–838 (2018).

124. Sheldon, R. A. & Brady, D. The limits to biocatalysis: pushing the envelope. *Chem. Commun.* **54**, 6088–6104 (2018).

125. Bornscheuer, U. T. *et al.* Engineering the third wave of biocatalysis. *Nature* **485**, 185–194 (2012).

126. Romero, P. A. & Arnold, F. H. Exploring protein fitness landscapes by directed evolution. *Nat. Rev. Mol. Cell Biol.* **10**, 866–876 (2009).

127. Lutz, S. Beyond directed evolution—semi-rational protein engineering and design. *Curr. Opin. Biotechnol.* **21**, 734–743 (2010).

128. Renata, H., Wang, Z. J. & Arnold, F. H. Expanding the enzyme universe: Accessing non-natural reactions by mechanism-guided directed evolution. *Angew. Chemie - Int. Ed.* **54**, 3351–3367 (2015).

129. Cheng, F., Zhu, L. & Schwaneberg, U. Directed evolution 2.0: improving and deciphering enzyme properties. *Chem. Commun.* **51**, 9760–9772 (2015).

130. Nobelprize.org. Nobel Prize for the directed evolution of enzymes. *Frances H. Arnold* Nobel Media AB 2019 (2018). Available at: https://www.nobelprize.org/prizes/chemistry/2018/press-release/. (Accessed: 12th January 2019)

131. Kaushik, M. *et al.* Protein engineering and de novo designing of a biocatalyst. *J. Mol. Recognit.* **29**, 499–503 (2016).

132. Reetz, M. T. & Carballeira, J. D. Iterative saturation mutagenesis (ISM) for rapid directed evolution of functional enzymes. *Nat. Protoc.* **2**, 891–903 (2007).

133. Copley, S. D. Shining a light on enzyme promiscuity. *Curr. Opin. Struct. Biol.* **47**, 167–175 (2017).

134. Chica, R. A., Doucet, N. & Pelletier, J. N. Semi-rational approaches to engineering enzyme activity: combining the benefits of directed evolution and rational design. *Curr. Opin. Biotechnol.* **16**, 378–384 (2005).

135. Reetz, M. T., Kahakeaw, D. & Lohmer, R. Addressing the Numbers Problem in Directed Evolution. *ChemBioChem* **9**, 1797–1804 (2008).

136. Kille, S. *et al.* Reducing Codon Redundancy and Screening Effort of Combinatorial Protein Libraries Created by Saturation Mutagenesis. *ACS Synth. Biol.* **2**, 83–92 (2013).

137. Davids, T., Schmidt, M., Böttcher, D. & Bornscheuer, U. T. Strategies for the discovery and engineering of enzymes for biocatalysis. *Curr. Opin. Chem. Biol.* **17**, 215–220 (2013).

138. Romero-Rivera, A., Garcia-Borràs, M. & Osuna, S. Computational tools for the evaluation of laboratory-engineered biocatalysts. *Chem. Commun.* **53**, 284–297 (2017).

139. Verma, R., Schwaneberg, U. & Roccatano, D. Computer-Aided Protein Directed Evolution: A Review Of Web Servers, Databases And Other Computational Tools For Protein Engineering. *Comput. Struct. Biotechnol. J.* **2**, (2012).

140. Khersonsky, O. & Tawfik, D. S. Enzyme Promiscuity: A Mechanistic and Evolutionary Perspective. *Annu. Rev. Biochem.* **79**, 471–505 (2010).

141. Khersonsky, O. & Tawfik, D. S. Enzyme promiscuity: a mechanistic and evolutionary perspective. *Annu. Rev. Biochem.* **79**, 471–505 (2010).

142. Bornscheuer, U. T. & Kazlauskas, R. J. Catalytic promiscuity in biocatalysis: Using old enzymes to form new bonds and follow new pathways. *Angew. Chemie - Int. Ed.* **43**, 6032–6040 (2004).

143. Gupta, R. D. Recent advances in enzyme promiscuity. *Sustain. Chem. Process.* **4**, 2 (2016).

144. Newton, M. S., Arcus, V. L., Gerth, M. L. & Patrick, W. M. Enzyme evolution: innovation is easy, optimization is complicated. *Curr. Opin. Struct. Biol.* **48**, 110–116 (2018).

145. Renata, H., Wang, Z. J. & Arnold, F. H. Expanding the Enzyme Universe: Accessing Non-Natural Reactions by Mechanism-Guided Directed Evolution. *Angew. Chemie Int. Ed.* **54**, 3351–3367 (2015).

146. Babtie, A., Tokuriki, N. & Hollfelder, F. What makes an enzyme promiscuous? *Curr. Opin. Chem. Biol.* **14**, 200–207 (2010).

147. Bornscheuer, U. T. & Kazlauskas, R. J. Catalytic Promiscuity in Biocatalysis: Using Old Enzymes to Form New Bonds and Follow New Pathways. *Angew. Chemie Int. Ed.* **43**, 6032–6040 (2004).

148. Leonard, E. *et al.* Combining metabolic and protein engineering of a terpenoid biosynthetic pathway for overproduction and selectivity control. *Proc. Natl. Acad. Sci.* **107**, 13654–13659 (2010).

149. Martin, V. J. J., Pitera, D. J., Withers, S. T., Newman, J. D. & Keasling, J. D. Engineering a mevalonate pathway in Escherichia coli for production of terpenoids. *Nat. Biotechnol.* **21**, 796–802 (2003).

150. Maróstica, M. R. & Pastore, G. M. Production of R-(+)-α-terpineol by the biotransformation of limonene from orange essential oil, using cassava waste water as medium. *Food Chem.* **101**, 345–350 (2007).

151. Korman, T. P., Opgenorth, P. H. & Bowie, J. U. A synthetic biochemistry platform for cell free production of monoterpenes from glucose. *Nat. Commun.* **8**, 15526 (2017).

152. Molina, G., Pimentel, M. R. & Pastore, G. M. Pseudomonas: a promising biocatalyst for the bioconversion of terpenes. *Appl. Microbiol. Biotechnol.* **97**, 1851–1864 (2013).

153. Bastian, S. Stereoselektive enzymatische Brønsted-Säure-Katalyse unter Verwendung einer Squalen-Hopen-Cyclase. (Universität Stuttgart, 2016).

154. Bertani, G. Studies on lysogenesis. I. The mode of phage liberation by lysogenic Escherichia coli. *J. Bacteriol.* **62**, 293–300 (1951).

155. Cold Spring Harbor Protocols. Terrific Broth (TB) Medium. *http://cshprotocols.cshlp.org/content/2015/9/pdb.rec085894.full* accessed 2020-02-24 (2015). Available at: http://cshprotocols.cshlp.org/content/2015/9/pdb.rec085894.full. (Accessed: 24th February 2020)

156. Hanahan, D. Studies on transformation of Escherichia coli with plasmids. *J. Mol. Biol.* **166**, 557–580 (1983).

157. Bischoff, J. Studien zum Einsatz von enzymatischen Lewis- und Brønsted-Säuren in Carbonyl-aktivierten organischen Reaktionen. (Universität Stuttgart, 2016).

158. Kühnel, L. Asparaginsäure-vermittelte enzymatische Reaktionen. (Universität Stuttgart, 2016).

159. Daegelen, P., Studier, F. W., Lenski, R. E., Cure, S. & Kim, J. F. Tracing Ancestors and Relatives of Escherichia coli B, and the Derivation of B Strains REL606 and BL21(DE3). *J. Mol. Biol.* **394**, 634–643 (2009).

160. Rosano, G. L. & Ceccarelli, E. A. Recombinant protein expression in Escherichia coli: advances and challenges. *Front. Microbiol.* **5**, 1–17 (2014).

161. Studier, F. W. & Moffatt, B. a. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol.* **189**, 113–130 (1986).

162. Pan, S. & Malcolm, B. A. Reduced Background Expression and Improved Plasmid Stability with pET Vectors in BL21 (DE3). *Biotechniques* **29**, 1234–1238 (2000).

163. Grant, S. G., Jessee, J., Bloom, F. R. & Hanahan, D. Differential plasmid rescue from transgenic mouse DNAs into Escherichia coli methylation-restriction mutants. *Proc. Natl. Acad. Sci.* **87**, 4645–4649 (1990).

164. Taylor, R. G., Walker, D. C. & Mclnnes, R. R. E. coli host strains significantly affect the quality of small scale plasmid DNA preparations used for sequencing. *Nucleic Acids Res.* **21**, 1677–1678 (1993).

165. Pearson, W. R. An Introduction to Sequence Similarity ("Homology") Searching. *Curr. Protoc. Bioinforma.* **42**, 3.1.1-3.1.8 (2013).

166. Hogrefe, H. H., Cline, J., Youngblood, G. L. & Allen, R. M. Creating randomized amino acid libraries with the QuikChange Multi Site-Directed Mutagenesis Kit. *Biotechniques* **33**, 1158–1165 (2002).

167. Xia, Y., Chu, W., Qi, Q. & Xun, L. New insights into the QuikChange process guide the use of Phusion DNA polymerase for site-directed mutagenesis. *Nucleic Acids Res.* **43**, 1–9 (2015).

168. Tulumovic, J. Optimierungen zur Umsetzung von Monoterpenen mit der Squalen-Hopen-Cyclase aus Alicyclobacillus acidocaldarius. (Universität Stuttgart, 2017).

169. Laemmli, U. K. Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. *Nature* **227**, 680–685 (1970).

170. Bradford, M. M. A Rapid and Sensitive Method for the Quantitation Microgram Quantities of Protein Utilizing the Principle of Protein-Dye Binding. *Anal. Biochem.* **254**, 248–254 (1976).

171. Han, T. H., Lee, J., Cho, M. H., Wood, T. K. & Lee, J. Environmental factors affecting indole production in Escherichia coli. *Res. Microbiol.* **162**, 108–116 (2011).

172. Kilikian, B. V., Suárez, I. D., Liria, C. W. & Gombert, A. K. Process strategies to improve heterologous protein production in Escherichia coli under lactose or IPTG induction. *Process Biochem.* **35**, 1019–1025 (2000).

173. Nair, R. *et al.* Yeast extract mediated autoinduction of lacUV5 promoter: an insight. *N. Biotechnol.* **26**, 282–288 (2009).

174. Gallart-Mateu, D., Rodriguez-Sojo, S. & de la Guardia, M. Determination of tea tree oil terpenes by headspace gas chromatography mass spectrometry. *Anal. Methods* **8**, 4576–4583 (2016).

175. Williams, C. M. & Whittaker, D. Rearrangements of pinane derivatives. Part I. Products of acid catalysed hydration of α-pinene and β-pinene. *J. Chem. Soc. B* 668–672 (1971).

176. Wendt, K. U., Feil, C., Lenhart, A., Poralla, K. & Schulz, G. E. Crystallization and preliminary X-ray crystallographic analysis of squalene-hopene cyclase from Alicyclobacillus acidocaldarius. *Protein Sci.* **6**, 722–724 (1997).

177. Shutava, H. G., Kavalenka, N. A., Supichenka, Nh. N., Leontiev, Nv. N. & Shutava, Nt. G. Essential Oils of Lamiaceae with High Content of α-, β-Pinene and Limonene Enantiomers. *J. Essent. Oil Bear. Plants* **17**, 18–25 (2014).

178. Racolta, S., Juhl, P. B., Sirim, D. & Pleiss, J. The triterpene cyclase protein family: A systematic analysis. *Proteins Struct. Funct. Bioinforma.* **80**, 2009–2019 (2012).

179. Fademrecht, S. Systematische Analyse von familienspezifischen Proteindatenbanken für die

biotechnologische Anwendung. (Universität Stuttgart, 2015).

180. Knerr, J. M. Mutationsstudien zur Umsetzung von (+) - β - Pinenund Sabinen mit der Squalen - Hopene - Cyclase aus Alicyclobacillus acidocaldarius. (Universität Stuttgart, 2017).

181. Romero-Rivera, A., Garcia-Borràs, M. & Osuna, S. Role of Conformational Dynamics in the Evolution of Retro-Aldolase Activity. *ACS Catal.* **7**, 8524–8532 (2017).

182. Shapira, S., Pleban, S., Kazanov, D., Tirosh, P. & Arber, N. Terpinen-4-ol: A Novel and Promising Therapeutic Agent for Human Gastrointestinal Cancers. *PLoS One* **11**, 1–13 (2016).

183. Lee, C.-J. *et al.* Correlations of the components of tea tree oil with its antibacterial effects and skin irritation. *J. Food Drug Anal.* **21**, 169–176 (2013).

184. Hart, P. H. *et al.* Terpinen-4-ol, the main component of the essential oil of Melaleuca alternifolia (tea tree oil), suppresses inflammatory mediator production by activated human monocytes. *Inflamm. Res.* **49**, 619–626 (2000).

185. Tighe, S., Gao, Y.-Y. & Tseng, S. C. G. Terpinen-4-ol is the Most Active Ingredient of Tea Tree Oil to Kill Demodex Mites. *Transl. Vis. Sci. Technol.* **2**, 1–8 (2013).

186. Shelton, D. *et al.* Isolation and partial characterisation of a putative monoterpene synthase from Melaleuca alternifolia. *Plant Physiol. Biochem.* **42**, 875–882 (2004).

187. Shelton, D., Leach, D., Baverstock, P. & Henry, R. Isolation of genes involved in secondary metabolism from Melaleuca alternifolia (Cheel) using expressed sequence tags (ESTs). *Plant Sci.* **162**, 9–15 (2002).

188. Yoo, S. K., Day, D. F. & Cadwallader, K. R. Bioconversion of α- and β-pinene by Pseudomonas sp. strain PIN. *Process Biochem.* **36**, 925–932 (2001).

189. Bateman, A. *et al.* UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* **45**, D158–D169 (2017).

190. Rottava, I. *et al.* Screening of microorganisms for bioconversion of (−)β-pinene and R-(+)-limonene to α-terpineol. *LWT - Food Sci. Technol.* **43**, 1128–1131 (2010).

191. Toniazzo, G., de Oliveira, D., Dariva, C., Oestreicher, E. G. & Antunes, O. a C. Biotransformation of (-)β-Pinene by Aspergillus niger ATCC 9642. *Appl. Biochem. Biotechnol.* **123**, 0837–0844 (2005).

192. Priya, D., Petkar, M. V. & Chowdary, G. V. Microbial biotransformation of α–(+)–pinene by newly identified strain of Gluconobacter japonicus MTCC 12284. *Int. J. Dev. Res.* **5**, 5270–5275 (2015).

193. Narushima, H., Omori, T. & Minoda, Y. Microbial Transformation of a-Pinene. *Eur. J. Appl. Microbiol. Biotechnol.* **16**, 174–178 (1982).

194. Groenewald, W. H., Gouws, P. A. & Witthuhn, R. C. Isolation, identification and typification of Alicyclobacillus acidoterrestris and Alicyclobacillus acidocaldarius strains from orchard soil and the fruit processing environment in South Africa. *Food Microbiol.* **26**, 71–76 (2009).

195. Mergulhão, F. J. M., Summers, D. K. & Monteiro, G. A. Recombinant protein secretion in Escherichia coli. *Biotechnol. Adv.* **23**, 177–202 (2005).

196. Silhavy, T. J. Classic spotlight: Gram-negative bacteria have two membranes. *J. Bacteriol.* **198**, 201 (2016).

197. Rapoport, T. A. Protein translocation across the eukaryotic endoplasmic reticulum and bacterial plasma membranes. *Nature* 663–669 (2007).

198. Movva, N. R., Nakamura, K. & Inouye, M. Amino acid sequence of the signal peptide of ompA protein, a major outer membrane protein of Escherichia coli. *J. Biol. Chem.* **225**, 27–29 (1980).

199. El-Gebali, S. *et al.* The Pfam protein families database in 2019. *Nucleic Acids Res.* **47**, 427–432 (2018).

200. Letunic, I. & Bork, P. 20 years of the SMART protein domain annotation resource. *Nucleic Acids Res.* **46**, D493–D496 (2018).

201. Pichersky, E. Biosynthesis of Plant Volatiles: Nature's Diversity and Ingenuity. *Science* **311**, 808–811 (2006).

202. Ma, X. *et al.* Highly selective isomerization of biomass β-pinene over hierarchically acidic MCM-22 catalyst. *Microporous Mesoporous Mater.* **237**, 180–188 (2017).

203. Zhang, Q. & Tiefenbacher, K. Hexameric Resorcinarene Capsule is a Brønsted Acid: Investigation and Application to Synthesis and Catalysis. *J. Am. Chem. Soc.* **135**, 16213–16219 (2013).

204. Zhang, Q. & Tiefenbacher, K. Terpene cyclization catalysed inside a self-assembled cavity. *Nat. Chem.* **7**, 197–202 (2015).

205. Zhang, Q., Catti, L., Pleiss, J. & Tiefenbacher, K. Terpene Cyclizations inside a Supramolecular Catalyst: Leaving-Group-Controlled Product Selectivity and Mechanistic Studies. *J. Am. Chem. Soc.* jacs.7b04480 (2017). doi:10.1021/jacs.7b04480

206. Kampranis, S. C. *et al.* Rational Conversion of Substrate and Product Specificity in a Salvia Monoterpene Synthase: Structural Insights into the Evolution of Terpene Synthase Function. *Plant Cell* **19**, 1994–2005 (2007).

207. Mavromatis, K. *et al.* Complete genome sequence of Alicyclobacillus acidocaldarius type strain (104-IAT). *Stand. Genomic Sci.* **2**, 9–18 (2010).

208. Gupta, R. S. & Mok, A. Phylogenomics and signature proteins for the alpha Proteobacteria and its main groups. *BMC Microbiol.* **7**, 106 (2007).

209. Aklujkar, M. *et al.* The genome of Pelobacter carbinolicus reveals surprising metabolic capabilities and physiological features. *BMC Genomics* **13**, 1–24 (2012).

210. Nakamura, Y. *et al.* Complete Genome Structure of the Thermophilic Cyanobacterium Thermosynechococcus elongatus BP-1. *DNA Res.* **9**, 135–148 (2002).

211. Chandra, G. & Chater, K. F. Developmental biology of Streptomyces from the perspective of 100 actinobacterial genome sequences. *FEMS Microbiol. Rev.* **38**, 345–379 (2014).

212. Bern, M. & Goldberg, D. Automatic selection of representative proteins for bacterial phylogeny. *BMC Evol. Biol.* **5**, 1–17 (2005).

213. Cao, Y. *et al.* Biosynthesis and production of sabinene: current state and perspectives. *Appl. Microbiol. Biotechnol.* **102**, 1535–1544 (2018).

214. Reipen, I. G., Poralla, K., Sahm, H. & Sprenger, G. A. Zymomonas mobilis squalene-hopene cyclase gene (shc): cloning, DNA sequence analysis and expression in Escherichia coli. *Microbiology* **141**, 155–161 (1995).

8 Appendix

8.1 List of abbreviations

Table 18: List of abbreviations used in this thesis.

abbreviation	full name
°C	degree Celsius
µ	micro
Å	Ångström
*Aac*SHC	SHC from *Alicyclobacillus acidocaldarius*
BSA	bovine serum albumin
dH_2O	deionized water
ddH_2O	double deionized water
DMAPP	dimethylally diphosphate
DMSO	dimethyl sulfoxide
dNTP	2´-desoxyribonucleosid-5´-triphosphate
DTT	dithiothreitol
DXP	desoxyxylulose-5-phosphat
e.g.	lat. *exempli gratia*: for example
EDTA	ethylenediaminetetraacetic acid
EI	electron ionization
epPCR	error prone polymerase chain reaction
et al	lat. *et alii*: and others
EtOAc	ethyl acetate
FID	flame-ionization detection
g	gram
GC	gas chromatograph
GC/MS	gas chromatograph with mass detector
h	hour
IEX	ion exchange
IPP	isopentenyl diphosphate
IPTG	isopropyl-ß-D-galactosid
ITB	Institute of Biochemistry and Technical Biochemistry, Department of Technical Biochemistry
kb	kilo base
kDa	kilo Dalton
L	liter
LB	lysogeny broth
m	milli or meter
MD simulations	molecular dynamics simulations
min	minute
MTBE	*methyl-tert*-butylether
n	nano
NCR	NADH-dependent 2-cyclohexen-1-one reductase
OD600	optical density at 600 nm
OSC	oxidosqualene cyclase
p	para

PCR	polymerase chain reaction
PMSF	phenylmethylsulfonyl fluoride
QM/MM	quantum mechanics/molecular mechanics
R	lat. *rectus*: right
rpm	rounds per minute
S	lat. *sinister*: left
SDS	sodium dodecylsulfate
SDS-PAGE	sodium dodecylsulfate polyacrylamide gel electrophoresis
SHC	squalene-hopene cyclase
SOC	super optimal broth with catabolite repression
TAE	tris(hydroxymethyl)aminomethane acetate
TB	terrific broth
TIC	total ion count
TM	trademark
Triton X-100	trade name for (p-tert-octylphenoxy)-polyethoxyethanol
V	Volt
wt	wild type

8.2 List of bacterial species

Table 19: List of abbreviations for bacterial species used in this thesis.

abbreviation	bacterial species
Aac	*Alicyclobacillus acidocaldarius*
Apa	*Acetobacter pasteurianus*
Bam	*Burkholderia ambifaria*
Bja	*Bradyrhizobium japonicum*
E. coli	*Escherichia coli*
Pca	*Pelobacter carbinolicus*
Rpa	*Rhodopseudomonas palustris*
Sco	*Streptomyces coelicolor*
Sfu	*Syntrophobacter fumaroxidans*
Tel	*Thermosynechococcus elongatus*
Ttu	*Teredinibacter turnerae*
Zmo	*Zymomonas mobilis*

8.3 List of nucleobases

Table 20: List of abbreviations for nucleobases used in this thesis.

single letter code	nucleobase
A	adenine
C	cytosine
G	guanine
K	T or G
N	A, C, G or T
T	thymine
U	uracil

8.4 List of proteinogenic amino acids

Table 21: List of abbreviations of amino acids used in this thesis.

single letter code	amino acids
A	alanine
C	cysteine
D	aspartic acid
E	glutamic acid
F	phenylalanine
G	glycine
H	histidine
I	isoleucine
K	lysine
L	leucine
M	methionine
N	asparagine
P	proline
Q	glutamine
R	arginine
S	serine
T	threonine
V	valine
W	tryptophane
Y	tyrosine

8.5 Primers for mutagenesis

Primers were numbered in the order of their creation and abbreviated PR + number.

Table 22: Primers used for site-directed mutagenesis, sorted by PR numbering.

primer name	sequence
PR19_F365W_forw	CCGGGCGGCTTTGCGTGGCAGTTTGATAACGTG
PR20_F365W_rev	CACGTTATCAAACTGCCACGCAAAGCCGCCCGG
PR21_W32A_forw	GATGAAGGCTATGCTTGGGGCCCGCTG
PR22_W32A_rev	CAGCGGGCCCCAAGCATAGCCTTCATC
PR25_N39Q_forw	CCGCTGCTGAGCCAGGTGACCATGG
PR26_N39Q_rev	CCATGGTCACCTGGCTCAGCAGCGG
PR29_N39P_forw	CCCGCTGCTGAGCCCTGTGACCATGGAAG
PR30_N39P_rev	CTTCCATGGTCACAGGGCTCAGCAGCGGG
PR31_L80A_forw	CACCTGGGCGGCCTATCCGGGCGGCCCGC
PR32_L80A_rev	GCGGGCCGCCCGGATAGGCCGCCCAGGTG
PR33_S38T_forw	GGGCCCGCTGCTGACTAACGTGACCATGGAAG
PR34_S38T_rev	CTTCCATGGTCACGTTAGTCAGCAGCGGGCCC
PR35_S38Y_forw	GGGCCCGCTGCTGTATAACGTGACCATGGAAGC
PR36_S38Y_rev	GCTTCCATGGTCACGTTATACAGCAGCGGGCCC
PR37_D89E_forw	GCCGGATCTGGAAACCACCATTGAAGCG
PR38_D89E_rev	CGCTTCAATGGTGGTTTCCAGATCCGGC
PR39_W133F_forw	GTTTACCCGCATGTTTCTGGCGCTGGTGGGC
PR40_W133F_rev	GCCCACCAGCGCCAGAAACATGCGGGTAAAC
PR41_W133L_forw	GTTTACCCGCATGTTACTGGCGCTGGTGGGC
PR42_W133L_rev	GCCCACCAGCGCCAGTAACATGCGGGTAAAC
PR43_W133G_forw	GTTTACCCGCATGGGTCTGGCGCTGGTGGGC
PR44_W133G_rev	GCCCACCAGCGCCAGACCCATGCGGGTAAAC
PR45_W312L_forw	GCATTAGCCCGGTGCTGGATACCGGCCTGGC
PR46_W312L_rev	GCCAGGCCGGTATCCAGCACCGGGCTAATGC
PR47_W312I_forw	GCATTAGCCCGGTGATTGATACCGGCCTGGC
PR48_W312I_rev	GCCAGGCCGGTATCAATCACCGGGCTAATGC
PR49_F365L_forw	GGCGGCTTTGCGCTGCAGTTTGATAACGTG
PR50_F365L_rev	CACGTTATCAAACTGCAGCGCAAAGCCGCC
PR53_A419D_forw	CAACGGCGGCTGGGGCGATTATGATGTGGATAAC
PR54_A419D_rev	GTTATCCACATCATAATCGCCCCAGCCGCCGTTG
PR55_Y495I_forw	GTGAACTATCTGATCGGCACCGGCGCGGTG
PR56_Y495I_rev	CACCGCGCCGGTGCCGATCAGATAGTTCAC
PR57_L607Y_forw	CCCGGGCGATTTTTATTATGGCTATACCATG
PR58_L607Y_rev	CATGGTATAGCCATAATAAAAATCGCCCGGG
PR59_L607S_forw	CCCGGGCGATTTTTATTCTGGCTATACCATG
PR60_L607S_rev	CATGGTATAGCCAGAATAAAAATCGCCCGGG
PR61_Y609I_forw	GCGATTTTTATCTGGGCATCACCATGTATCGCC
PR62_Y609I_rev	GGCGATACATGGTGATGCCCAGATAAAAATCGC
PR63_Y609I_ L607S_forw	GCGATTTTTATTCTGGCATCACCATGTATCGCC

PR64_Y609I_ L607S_rev	GGCGATACATGGTGATGCCAGAATAAAAATCGC
PR65_R171K_forw	TTGGCAGCTGGGCGAAAGCGACCGTGGTGG
PR66_R171K_rev	CCACCACGGTCGCTTTCGCCCAGCTGCCAA
PR67_R171G_forw	TTTGGCAGCTGGGCGGGTGCGACCGTGGTGG
PR68_R171G_rev	CCACCACGGTCGCACCCGCCCAGCTGCCAAA
PR69_N369Q_forw	GCGTTTCAGTTTGATCAGGTGTATTATCCGGAT
PR70_N369Q_rev	ATCCGGATAATACACCTGATCAAACTGAAACGC
PR71_N369D_forw	GCGTTTCAGTTTGATGATGTGTATTATCCGGATG
PR72_N369D_rev	CATCCGGATAATACACATCATCAAACTGAAACGC
PR77_P443G_forw	GGCGAAGTGACCGATGGCCCGAGCGAAGATG
PR78_P443G_rev	CATCTTCGCTCGGGCCATCGGTCACTTCGCC
PR79_P443Y_forw	GGCGAAGTGACCGATTATCCGAGCGAAGATG
PR80_P443Y_rev	CATCTTCGCTCGGATAATCGGTCACTTCGCC
PR81_P443H_forw	GGCGAAGTGACCGATCATCCGAGCGAAGATG
PR82_P443H_rev	CATCTTCGCTCGGATGATCGGTCACTTCGCC
PR83_S445T_forw	GACCGATCCGCCGACAGAAGATGTGACCGC
PR84_S445T_rev	GCGGTCACATCTTCTGTCGGCGGATCGGTC
PR85_S445Y_forw	GACCGATCCGCCGTATGAAGATGTGACCGC
PR86_S445Y_rev	GCGGTCACATCTTCATACGGCGGATCGGTC
PR87_G487A_forw	GGCAGCTGGTTTGCGCGCTGGGGCGTGAAC
PR88_G487A_rev	GTTCACGCCCCAGCGCGCAAACCAGCTGCC
PR89_G487L_forw	GGCAGCTGGTTTCTGCGCTGGGGCGTGAAC
PR90_G487L_rev	GTTCACGCCCCAGCGCAGAAACCAGCTGCC
PR91_R488K_forw	GCAGCTGGTTTGGCAAATGGGGCGTGAACTATC
PR92_R488K_rev	GATAGTTCACGCCCCATTTGCCAAACCAGCTGC
PR93_R488S_forw	GCAGCTGGTTTGGCTCTTGGGGCGTGAACTATC
PR94_R488S_rev	GATAGTTCACGCCCCAAGAGCCAAACCAGCTGC
PR95_G490A_forw	GGTTTGGCCGCTGGGCGGTGAACTATCTGT
PR96_G490A_rev	ACAGATAGTTCACCGCCCAGCGGCCAAACC
PR97_G490L_forw	TGGTTTGGCCGCTGGCTGGTGAACTATCTGT
PR98_G490L_rev	ACAGATAGTTCACCAGCCAGCGGCCAAACCA
PR101_Y609I_ L607Y_forw	GCGATTTTTATTATGGCATCACCATGTATCGCC
PR102_Y609I_ L607Y_rev	GGCGATACATGGTGATGCCTATATAAAAATCGC
PR103_A419S	CAACGGCGGCTGGGGCTCTTATGATGTGG
PR104_A419S	CCACATCATAAGAGCCCCAGCCGCCGTTG
PR105_T597I_forw	ATGAACCGTATTATATTGGCACCGGGTTCCCG
PR106_T597I_rev	CGGGAACCCGGTGCCAATATAATACGGTTCAT
PR107_T599S_forw	GTATTATACCGGCTCTGGGTTCCCGGGCGA
PR108_T599S_rev	TCGCCCGGGAACCCAGAGCCGGTATAATAC
PR109_T599Y_forw	CGTATTATACCGGCTATGGGTTCCCGGGCGA
PR110_T599Y_rev	TCGCCCGGGAACCCATAGCCGGTATAATACG
PR111_T599G_forw	CGTATTATACCGGCGGTGGGTTCCCGGGCG
PR112_T599G_rev	CGCCCGGGAACCCACCGCCGGTATAATACG
PR117_Y606M_forw	CCCGGGCGATTTTATGCTGGGCTATACCATG
PR118_Y606M_rev	CATGGTATAGCCCAGCATAAAATCGCCCGGG

PR119_T597S_forw	TGAACCGTATTATTCTGGCACCGGGTTCCCG
PR120_T597S_rev	CGGGAACCCGGTGCCAGAATAATACGGTTCA
PR121_T597Y_forw	ATGAACCGTATTATTATGGCACCGGGTTCCCGG
PR122_T597Y_rev	CCGGGAACCCGGTGCCATAATAATACGGTTCAT
PR150_L36I_forw	GGCTATTGGTGGGGCCCGATCCTGAGCAACGTG
PR151_L36I_rev	CACGTTGCTCAGGATCGGGCCCCACCAATAGCC
PR152_L36M_forw	GGCTATTGGTGGGGCCCGATGCTGAGCAACGTG
PR153_L36M_rev	CACGTTGCTCAGCATCGGGCCCCACCAATAGCC
PR154_L36F_forw	GGCTATTGGTGGGGCCCGTTTCTGAGCAACGTG
PR155_L36F_rev	CACGTTGCTCAGAAACGGGCCCCACCAATAGCC
PR158_L36T_forw	GGCTATTGGTGGGGCCCGACCCTGAGCAACGTG
PR159_L36T_rev	CACGTTGCTCAGGGTCGGGCCCCACCAATAGCC
PR160_L36Q_forw	GGCTATTGGTGGGGCCCGCAGCTGAGCAACGTG
PR161_L36Q_rev	CACGTTGCTCAGCTGCGGGCCCCACCAATAGCC
PR162_L36N_forw	GGCTATTGGTGGGGCCCGAACCTGAGCAACGTG
PR163_L36N_rev	CACGTTGCTCAGGTTCGGGCCCCACCAATAGCC
PR164_L36G_forw	GGCTATTGGTGGGGCCCGGGCCTGAGCAACGTG
PR165_L36G_rev	CACGTTGCTCAGGCCCGGGCCCCACCAATAGCC
PR166_L36P_forw	GGCTATTGGTGGGGCCCGCCGCTGAGCAACGTG
PR167_L36P_rev	CACGTTGCTCAGCGGCGGGCCCCACCAATAGCC
PR168_L36R_forw	GGCTATTGGTGGGGCCCGCGTCTGAGCAACGTG
PR169_L36R_rev	CACGTTGCTCAGACGCGGGCCCCACCAATAGCC
PR172_L36D_forw	GGCTATTGGTGGGGCCCGGATCTGAGCAACGTG
PR173_L36D_rev	CACGTTGCTCAGATCCGGGCCCCACCAATAGCC
PR174_L36E_forw	GGCTATTGGTGGGGCCCGGAACTGAGCAACGTG
PR175_L36E_rev	CACGTTGCTCAGTTCCGGGCCCCACCAATAGCC
PR180_Y609S_forw	CGGGCGATTTTTATCTGGGCTCGACCATGTATCGC
PR181_Y609S_rev	GCGATACATGGTCGAGCCCAGATAAAAATCGCCCG
PR184_Y609N_forw	CGGGCGATTTTTATCTGGGCAACACCATGTATCGC
PR185_Y609N_rev	GCGATACATGGTGTTGCCCAGATAAAAATCGCCCG
PR200_D442Q_forw	GGCGAAGTGACCCAGCCGCCGAGCGAAGATG
PR201_D442Q_rev	CATCTTCGCTCGGCGGCTGGGTCACTTCGCC
PR202_D442H_forw	GGCGAAGTGACCCACCCGCCGAGCGAAGATG
PR203_D442H_rev	CATCTTCGCTCGGCGGGTGGGTCACTTCGCC
PR214_D442P_forw	GGCGAAGTGACCCCGCCGCCGAGCGAAGATG
PR215_D442P_rev	CATCTTCGCTCGGCGGCGGGGTCACTTCGCC
PR216_D442M_forw	GGCGAAGTGACCATGCCGCCGAGCGAAGATG
PR217_D442M_rev	CATCTTCGCTCGGCGGCATGGTCACTTCGCC
PR218_D442K_forw	GGCGAAGTGACCAAACCGCCGAGCGAAGATG
PR219_D442K_rev	CATCTTCGCTCGGCGGTTTGGTCACTTCGCC
PR220_D442C_forw	GGCGAAGTGACCTGCCCGCCGAGCGAAGATG
PR221_D442C_rev	CATCTTCGCTCGGCGGGCAGGTCACTTCGCC
PR222_D442R_forw	GGCGAAGTGACCCGTCCGCCGAGCGAAGATG
PR223_D442R_rev	CATCTTCGCTCGGCGGACGGGTCACTTCGCC
PR224_D442I_forw	GGCGAAGTGACCATCCCGCCGAGCGAAGATG
PR225_D442I_rev	CATCTTCGCTCGGCGGACGGGTCACTTCGCC
PR226_D442W_forw	GGCGAAGTGACCTGGCCGCCGAGCGAAGATG

PR227_D442W_rev	CATCTTCGCTCGGCGGCCAGGTCACTTCGCC
PR228_D442F_forw	GGCGAAGTGACCTTTCCGCCGAGCGAAGATG
PR229_D442F_rev	CATCTTCGCTCGGCGGAAAGGTCACTTCGCC
PR230_D442A_forw	GGCGAAGTGACCGCGCCGCCGAGCGAAGATG
PR231_D442A_rev	CATCTTCGCTCGGCGGCGCGGTCACTTCGCC
PR232_D442E_forw	GGCGAAGTGACCGAACCGCCGAGCGAAGATG
PR233_D442E_rev	CATCTTCGCTCGGCGGTTCGGTCACTTCGCC
PR234_D442N_forw	GGCGAAGTGACCAACCCGCCGAGCGAAGATG
PR235_D442N_rev	CATCTTCGCTCGGCGGGTTGGTCACTTCGCC
PR236_Y609P_forw	GCGATTTTTATCTGGGCCCGACCATGTATCGCCATGTG
PR237_Y609P_rev	CACATGGCGATACATGGTCGGGCCCAGATAAAAATCGC
PR238_Y609C_forw	GCGATTTTTATCTGGGCTGCACCATGTATCGCCATGTGTTTCCG
PR239_Y609C_rev	CGGAAACACATGGCGATACATGGTGCAGCCCAGATAAAAATCGC
PR240_Y609R_forw	CGGGCGATTTTTATCTGGGCCGCACCATGTATCGCCATGTGTTTCCG
PR241_Y609R_rev	CGGAAACACATGGCGATACATGGTGCGGCCCAGATAAAAATCGCCCG
PR246_Y609Q_forw	CGGGCGATTTTTATCTGGGCCAGACCATGTATCGCCATGTGTTTCCG
PR247_Y609Q_rev	CGGAAACACATGGCGATACATGGTCTGGCCCAGATAAAAATCGCCCG
PR248_Y609T_forw	CGGGCGATTTTTATCTGGGCACCACCATGTATCGCCATGTGTTTCCG
PR249_Y609T_rev	CGGAAACACATGGCGATACATGGTGGTGCCCAGATAAAAATCGCCCG
PR250_Y609K_forw	CGGGCGATTTTTATCTGGGCAAAACCATGTATCGCCATGTGTTTCCG
PR251_Y609K_rev	CGGAAACACATGGCGATACATGGTTTTGCCCAGATAAAAATCGCCCG
PR252_Y609D_forw	CGGGCGATTTTTATCTGGGCGACACCATGTATCGCCATGTGTTTCCG
PR253_Y609D_rev	CGGAAACACATGGCGATACATGGTGTCGCCCAGATAAAAATCGCCCG
PR254_L36K_forw	GGCTATTGGTGGGGCCCGAAACTGAGCAACGTGACCATGG
PR255_L36K_rev	CCATGGTCACGTTGCTCAGTTTCGGGCCCCACCAATAGCC
PR256_L36H_forw	GGCTATTGGTGGGGCCCGCACCTGAGCAACGTGACCATGG
PR257_L36H_rev	CCATGGTCACGTTGCTCAGGTGCGGGCCCCACCAATAGCC
PR262_S307A_forw	GGATGTTTCAGGCGGCGATTAGCCCGGTGTGGG
PR263_S307A_rev	CCCACACCGGGCTAATCGCCGCCTGAAACATCC
PR264_S309A_forw	GTTTCAGGCGAGCATTGCGCCGGTGTGGGATACCG
PR265_S309A_rev	CGGTATCCCACACCGGCGCAATGCTCGCCTGAAAC
PR266_Y609H_forw	GATTTTTATCTGGGCCATACCATGTATCGC
PR267_Y609H_rev	GCGATACATGGTATGGCCCAGATAAAAATC
PR270_L36Y_forw	CTATTGGTGGGGCCCGTATCTGAGCAACGTGAC

PR271_L36Y_rev	GTCACGTTGCTCAGATACGGGCCCCACCAATAG
PR272_D442Y_forw	GATTTTGGCGAAGTGACCTATCCGCCGAGCGAAGATG
PR273_D442Y_rev	CATCTTCGCTCGGCGGATAGGTCACTTCGCCAAAATC
PR274_D442L_forw	GATTTTGGCGAAGTGACCCTGCCGCCGAGCGAAGATG
PR275_D442L_rev	CATCTTCGCTCGGCGGCAGGGTCACTTCGCCAAAATC
PR276_D442V_forw	GATTTTGGCGAAGTGACCGTGCCGCCGAGCGAAGATG
PR277_D442V_rev	CATCTTCGCTCGGCGGCACGGTCACTTCGCCAAAATC
PR278_D442G_forw	GATTTTGGCGAAGTGACCGGCCCGCCGAGCGAAGATG
PR279_D442G_rev	CATCTTCGCTCGGCGGGCCGGTCACTTCGCCAAAATC
PR280_AXAA_forw	GATAACGTGTATTATCCGGCGGCGGCGGCGACCGCGGTGGTGGTGTG
PR281_AXAA_rev	CACACCACCACCGCGGTCGCCGCCGCCGCCGGATAATACACGTTATC
PR282_D374A_forw	GTATTATCCGGCGGTGGATGATACCG
PR283_D374A_rev	CGGTATCATCCACCGCCGGATAATAC
PR284_D376A_forw	GTATTATCCGGATGTGGCGGATACCGCGGTG
PR285_D376A_rev	CACCGCGGTATCCGCCACATCCGGATAATAC
PR286_D377A_forw	GTATTATCCGGATGTGGATGCGACCGCGGTGGTGGTGTG
PR287_D377A_forw	CACACCACCACCGCGGTCGCATCCACATCCGGATAATAC
PR288_D442S_forw	GATTTTGGCGAAGTGACCAGCCCGCCGAGCGAAGATGTGACCG
PR289_D442S_rev	CGGTCACATCTTCGCTCGGCGGGCTGGTCACTTCGCCAAAATC
PR290_D442T_for	GATTTTGGCGAAGTGACCACCCCGCCGAGCGAAGATGTGACCG
PR291_D442T_re	CGGTCACATCTTCGCTCGGCGGGGTGGTCACTTCGCCAAAATC
PR292_Q366S_forw	CCGGGCGGCTTTGCGTTTTCTTTTGATAACGTG
PR293_Q366S_rev	CACGTTATCAAAAGAAAACGCAAAGCCGCCCGG
PR302_Y540T_forw	GAAGATTGCCGCAGCACGGAAGATCCGGCGTATGCGG
PR303_Y540T_rev	CCGCATACGCCGGATCTTCCGTGCTGCGGCAATCTTC
PR304_Y606S_forw	CCCGGGCGATTTTTCCCTGGGCTATACCATG
PR305_Y606S_rev	CATGGTATAGCCCAGGGAAAAATCGCCCGGG
PR306_Y606T_forw	CCCGGGCGATTTTACCCTGGGCTATACCATG
PR307_Y606T_rev	CATGGTATAGCCCAGGGTAAAATCGCCCGGG
PR308_I308V_forw	GTTTCAGGCGAGCGTGAGCCCGGTGTGGG
PR309_I308V_rev	CCCACACCGGGCTCACGCTCGCCTGAAAC
PR342_P444D_forw	GCGAAGTGACCGATCCGGATAGCGAAGATGTGACCGCG
PR343_P444D_rev	CGCGGTCACATCTTCGCTATCCGGATCGGTCACTTCGC

8.6 Plasmids and sequences

8.6.1 *Aac*SHC

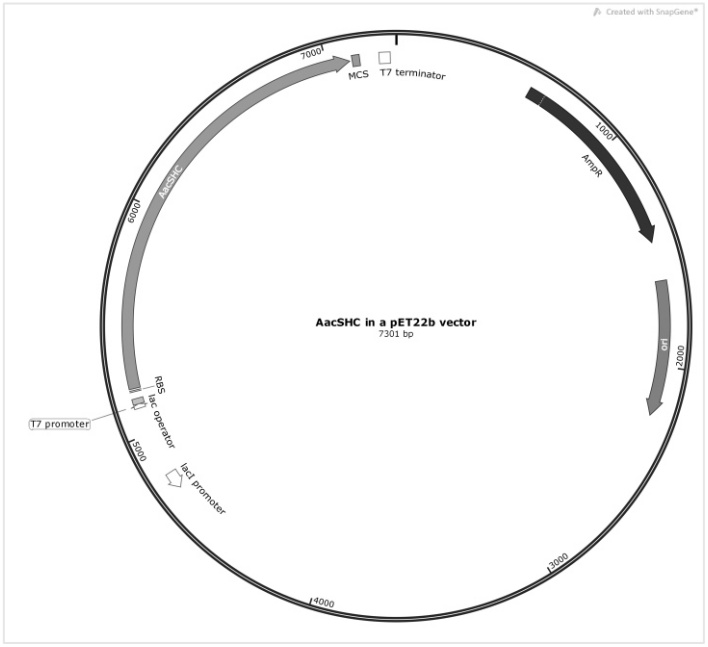


Figure 70: *Aac*SHC gene in a pET22b(+) vector.

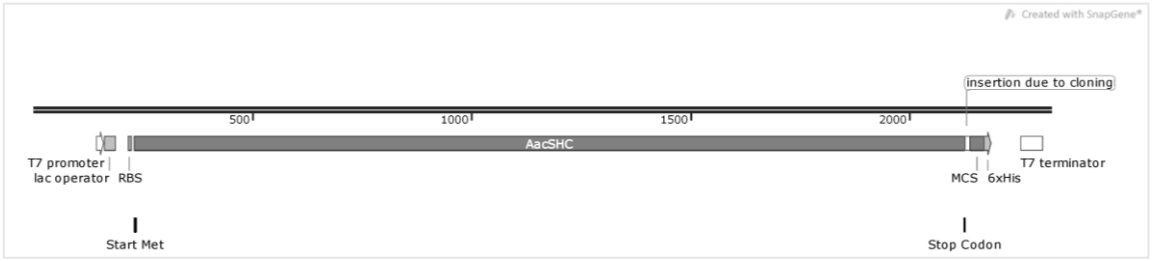


Figure 71: Details on the *Aac*SHC sequence.

The *Aac*SHC nucleotide sequence is the following:

ATGGCGGAACAGCTGGTGGAAGCGCCGGCGTATGCGCGCACCCTGGATCGCGCGG
TGGAATATCTGCTGAGCTGCCAGAAAGATGAAGGCTATTGGTGGGGCCCGCTGCTG
AGCAACGTGACCATGGAAGCGGAATATGTGCTGCTGTGCCATATTCTGGATCGCGT
GGATCGCGATCGCATGGAAAAAATTCGCCGCTATCTGCTGCATGAACAGCGCGAAG
ATGGCACCTGGGCGCTGTATCCGGGCGGCCCGCCGGATCTGGATACCACCATTGAA
GCGTATGTGGCGCTGAAATATATTGGCATGAGCCGCGATGAAGAACCGATGCAGAA
AGCGCTGCGCTTTATTCAGAGCCAGGGCGGCATTGAAAGCAGCCGCGTGTTTACCC
GCATGTGGCTGGCGCTGGTGGGCGAATATCCGTGGGAAAAAGTGCCGATGGTGCCG
CCGGAAATTATGTTTCTGGGCAAACGCATGCCGCTGAACATTTATGAATTTGGCAGC
TGGGCGCGCGCGACCGTGGTGGCGCTGAGCATTGTGATGAGCCGCCAGCCGGTGTT
TCCGCTGCCGGAACGCGCGCGCGTGCCGGAACTGTATGAAACCGATGTGCCGCCGC
GCCGCCGCGGCGCGAAAGGCGGCGGcGGCTGGATTTTTGATGCGCTGGATCGCGCG
CTGCATGGCTATCAGAAACTGAGCGTGCATCCGTTTCGCCGCGCGGCGGAAATTCG

CGCGCTGGATTGGCTGCTGGAACGCCAGGCGGGCGATGGCAGCTGGGGCGGCATT
CAGCCGCCGTGGTTTTATGCGCTGATTGCGCTGAAAATTCTGGATATGACCCAGCAT
CCGGCGTTTATTAAAGGCTGGGAAGGCCTGGAACTGTATGGCGTGGAACTGGATTA
TGGCGGCTGGATGTTTCAGGCGAGCATTAGCCCGGTGTGGGATACCGGCCTgGCGG
TGCTGGCGCTGCGCGCGGCGGGCCTGCCGGCGGATCATGATCGCCTGGTGAAAGC
GGGCGAATGGCTGCTGGATCGCCAGATTACCGTGCCGGGCGATTGGGCGGTGAAAC
GCCCGAACCTGAAACCGGGCGGCTTTGCGTTTCAGTTTGATAACGTGTATTATCCGG
ATGTGGATGATACCGCGGTGGTGGTGtGGGCGCTGAACACCCTGCGCCTGCCGGAT
GAACGCCGCCGCCGCGATGCGATGACCAAAGGCTTTCGCTGGATTGTGGGCATGCA
GAGCAGCAACGGCGGCTGGGGCGCGTATGATGTGGATAACACCAGCGATCTGCCGA
ACCATATTCCGTTTTGCGATTTTGGCGAAGTGACCGATCCGCCGAGCGAAGATGTGA
CCGCGCATGTGCTGGAATGCTTTGGCAGCTTTGGCTATGATGATGCGTGGAAAGTG
ATTCGCCGCGCGGTGGAATATCTGAAACGCGAACAGAAACCGGATGGCAGCTGGTT
TGGCCGCTGGGGCGTGAACTATCTGTATGGCACCGGCGCGGTGGTGAGCGCGCTGA
AAGCGGTGGGCATTGATACCCGCGAACCGTATATTCAGAAAGCGCTGGATTGGGTG
GAACAGCATCAGAACCCGGATGGCGGCTGGGGCGAAGATTGCCGCAGCTATGAAGA
TCCGGCGTATGCGGGCAAAGGCGCGAGCACCCCGAGCCAGACCGCGTGGGCGCTG
ATGGCGCTGATTGCGGGCGGCCGCGCGGAAAGCGAAGCGGCGCGCCGCGGCGTGC
AGTATCTGGTGGAAACCCAGCGCCCGGATGGCGGCTGGGATGAACCGTATTATACC
GGCACCGGGTTCCCGGGCGATTTTTATCTGGGCTATACCATGTATCGCCATGTGTTT
CCGACCCTGGCGCTGGGCCGCTATAAACAGGCGATTGAACGCCGCTAA

The protein sequence is the following:

MAEQLVEAPAYARTLDRAVEYLLSCQKDEGYWWGPLLSNVTMEAEYVLLCHILDRV
DRDRMEKIRRYLLHEQREDGTWALYPGGPPDLDTTIEAYVALKYIGMSRDEEPMQK
ALRFIQSQGGIESSRVFTRMWLALVGEYPWEKVPMVPPEIMFLGKRMPLNIYEFGSW
ARATVVAISIVMSRQPVFPLPERARVPELYDTDVPPRRRGAKGGGGRIFDALDRALHG
YQKLSVHPFRRAAEIRALDWLLERQAGDGSWGGIQPPWFYTLIALKILDMTQHPAFIK
GWEGLELYGVDLDYGGWMFQASISPVWDTGLAVLALRAAGLPADHDRLVKAGEWL
LDRQITVPGDWAVKRPNLKPGGFAFQFDNVYYPDVDDTAVVVWALNSLRLPDERRR
RDVMTKGFRWIVGMQSSNGGWGAYDVDNTSDLPNHIPFCDFGEVTDPPSEDVTAHV
LECFGSFGYDDAWKVIRRAVEYLKREQRPDGSWFGRWGVNYLYGTGAVVPALKAVG
IDVREPFIQKALDWVEQHQNPDGGWGEDCRSYEDPAYAGKGASTPSQTAWALMALI
AGGRAESDSVRRGVQYLVETQRPDGGWDEPYYTGTGFPGDFYLGYTMYRHVFPTLA
LGRYKQAIERR.

8.6.2 *Aac*SHC variants

All variants used in this thesis are listed in Table 23 containing the original source.

Table 23: *Aac*SHC variants used in this thesis with in-house numbering (ITB number) and source.

ITB number	plasmid	source (collaborator)
pITB5343	pET22b(+)_AacSHC_W32A	this thesis (J. Knerr)
pITB5344	pET22b(+)_AacSHC_N39Q	this thesis (J. Knerr)
pITB5345	pET22b(+)_AacSHC_S38T	this thesis (J. Knerr)
pITB5346	pET22b(+)_AacSHC_S38Y	this thesis (J. Knerr)
pITB5347	pET22b(+)_AacSHC_N39P	this thesis (J. Knerr)
pITB5348	pET22b(+)_AacSHC_N39Q	this thesis (J. Knerr)
pITB5349	pET22b(+)_AacSHC_L80A	this thesis (J. Knerr)
pITB5350	pET22b(+)_AacSHC_D89E	this thesis (J. Knerr)
pITB5351	pET22b(+)_AacSHC_W133F	this thesis (J. Knerr)
pITB5352	pET22b(+)_AacSHC_W133G	this thesis (J. Knerr)
pITB5353	pET22b(+)_AacSHC_W133L	this thesis (J. Knerr)
pITB5354	pET22b(+)_AacSHC_R171G	this thesis (J. Knerr)
pITB5355	pET22b(+)_AacSHC_R171K	this thesis (J. Knerr)
pITB5356	pET22b(+)_AacSHC_N369D	this thesis (J. Knerr)
pITB5357	pET22b(+)_AacSHC_N369Q	this thesis (J. Knerr)
pITB5358	pET22b(+)_AacSHC_D442E	this thesis (J. Knerr)
pITB5359	pET22b(+)_AacSHC_D442N	this thesis (J. Knerr)
pITB5360	pET22b(+)_AacSHC_P443G	this thesis (J. Knerr)
pITB5361	pET22b(+)_AacSHC_P443H	this thesis (J. Knerr)
pITB5362	pET22b(+)_AacSHC_P443Y	this thesis (J. Knerr)
pITB5363	pET22b(+)_AacSHC_S445T	this thesis (J. Knerr)
pITB5364	pET22b(+)_AacSHC_S445Y	this thesis (J. Knerr)
pITB5365	pET22b(+)_AacSHC_G487A	this thesis (J. Knerr)
pITB5366	pET22b(+)_AacSHC_G487L	this thesis (J. Knerr)
pITB5367	pET22b(+)_AacSHC_R488K	this thesis (J. Knerr)
pITB5368	pET22b(+)_AacSHC_R488S	this thesis (J. Knerr)
pITB5369	pET22b(+)_AacSHC_G490A	this thesis (J. Knerr)
pITB5370	pET22b(+)_AacSHC_G490L	this thesis (J. Knerr)
pITB5371	pET22b(+)_AacSHC_Y540S	this thesis (J. Knerr)
pITB5372	pET22b(+)_AacSHC_T597Y	this thesis (J. Knerr)
pITB5373	pET22b(+)_AacSHC_T597I	this thesis (J. Knerr)
pITB5374	pET22b(+)_AacSHC_T597S	this thesis (J. Knerr)
pITB5375	pET22b(+)_AacSHC_T599G	this thesis (J. Knerr)
pITB5376	pET22b(+)_AacSHC_T599S	this thesis (J. Knerr)
pITB5377	pET22b(+)_AacSHC_T599Y	this thesis (J. Knerr)
pITB5378	pET22b(+)_AacSHC_Y606M	this thesis (J. Knerr)
pITB5379	pET22b(+)_AacSHC_Y606T	this thesis
pITB5380	pET22b(+)_AacSHC_Y606S	this thesis
pITB5381	pET22b(+)_AacSHC_Y450T	this thesis
pITB5382	pET22b(+)_AacSHC_Y609W_F365W	this thesis (J. Knerr)
pITB5383	pET22b(+)_AacSHC_A419S	this thesis (J. Tulumovic)

pITB5384	pET22b(+)_AacSHC_A419D	this thesis (J. Tulumovic)
pITB5385	pET22b(+)_AacSHC_W312L	this thesis (J. Tulumovic)
pITB5386	pET22b(+)_AacSHC_W312I	this thesis (J. Tulumovic)
pITB5387	pET22b(+)_AacSHC_F365L	this thesis (J. Tulumovic)
pITB5388	pET22b(+)_AacSHC_Y495I	this thesis (J. Tulumovic)
pITB5389	pET22b(+)_AacSHC_L607Y	this thesis (J. Tulumovic)
pITB5390	pET22b(+)_AacSHC_L607S	this thesis
pITB5391	pET22b(+)_AacSHC_L36A_Y609I	this thesis (J. Tulumovic)
pITB5392	pET22b(+)_AacSHC_L36A_A419S	this thesis (J. Tulumovic)
pITB5393	pET22b(+)_AacSHC_L36A_W312L	this thesis (J. Tulumovic)
pITB5394	pET22b(+)_AacSHC_L36A_F365L	this thesis (J. Tulumovic)
pITB5395	pET22b(+)_AacSHC_L36A_Y495I	this thesis (J. Tulumovic)
pITB5396	pET22b(+)_AacSHC_L36A_L607S	this thesis (J. Tulumovic)
pITB5397	pET22b(+)_AacSHC_A419D_Y495I	this thesis (J. Tulumovic)
pITB5398	pET22b(+)_AacSHC_Y495I_L607S	this thesis (J. Tulumovic)
pITB5399	pET22b(+)_AacSHC_F365L_L607S	this thesis (J. Tulumovic)
pITB5400	pET22b(+)_AacSHC_Y495I_Y609W	this thesis (J. Tulumovic)
pITB5401	pET22b(+)_AacSHC_L36A_Y495I_L607S	this thesis (J. Tulumovic)
pITB5402	pET22b(+)_AacSHC_L36A_F365L_L607S	this thesis (J. Tulumovic)
pITB5403	pET22b(+)_AacSHC_L36A_Y495I_Y609W	this thesis (J. Tulumovic)
pITB5404	pET22b(+)_AacSHC_W312I_A419D_Y495I	this thesis (J. Tulumovic)
pITB5405	pET22b(+)_AacSHC_A419D_Y495I_L607S	this thesis (J. Tulumovic)
pITB5406	pET22b(+)_AacSHC_Y495I_L607S_Y609I	this thesis (J. Tulumovic)
pITB5407	pET22b(+)_AacSHC_A419S_L607S_Y609I	this thesis (J. Tulumovic)
pITB5408	pET22b(+)_AacSHC_W312L_L607S_Y609I	this thesis (J. Tulumovic)
pITB5409	pET22b(+)_AacSHC_L36A_Y495I_L607S_ Y609I	this thesis (J. Tulumovic)
pITB5410	pET22b(+)_AacSHC_L36A_A419S_L607S_ Y609I	this thesis (J. Tulumovic)
pITB5411	pET22b(+)_AacSHC_L36A_W312L_L607S_ Y609I	this thesis (J. Tulumovic)
pITB5412	pET22b(+)_AacSHC_W312I_F365L_A419D_ Y495I	this thesis (J. Tulumovic)
pITB5413	pET22b(+)_AacSHC_W312L_A419D_Y495I_ L607S	this thesis (J. Tulumovic)
pITB5414	pET22b(+)_AacSHC_A419D_W312I_F365L_ Y495I_L607Y	this thesis (J. Tulumovic)
pITB5415	pET22b(+)_AacSHC_W312L_F365L_A419_ Y495I_L607S	this thesis (J. Tulumovic)
pITB5416	pET22b(+)_AacSHC_W312L_A419S_Y495I_ L607S_Y609I	this thesis (J. Tulumovic)
pITB5417	pET22b(+)_AacSHC_W312L_F365L_A419D_ Y495I_L607S_Y609I	this thesis (J. Tulumovic)
pITB5418	pET22b(+)_AacSHC_W312L_F365L_A419D_ Y495I_L607Y_Y609I	this thesis (J. Tulumovic)
pITB5419	pET22b(+)_AacSHC_W312I_F365L_A419D_ Y495I_L607S_Y609I	this thesis (J. Tulumovic)
pITB5420	pET22b(+)_AacSHC_F365L__W312L_A419D _	this thesis (J. Tulumovic)

	Y495I_L607S_Y609I	
pITB1025	pET22b(+)_*Aac*SHC_L36A	S. C. Hammer
pITB1288	pET22b(+)_*Aac*SHC_L36C	S. Bastian
pITB5421	pET22b(+)_AacSHC_L36D	this thesis
pITB5422	pET22b(+)_AacSHC_L36E	this thesis
pITB5423	pET22b(+)_AacSHC_L36F	this thesis
pITB5424	pET22b(+)_AacSHC_L36G	this thesis
pITB5425	pET22b(+)_AacSHC_L36H	this thesis
pITB5426	pET22b(+)_AacSHC_L36K	this thesis
pITB5427	pET22b(+)_AacSHC_L36M	this thesis
pITB5428	pET22b(+)_AacSHC_L36N	this thesis
pITB5429	pET22b(+)_AacSHC_L36P	this thesis
pITB5430	pET22b(+)_AacSHC_L36Q	this thesis
pITB5431	pET22b(+)_AacSHC_L36R	this thesis
pITB1289	pET22b(+)_*Aac*SHC_L36S	S. C. Hammer
pITB5432	pET22b(+)_AacSHC_L36T	this thesis
pITB1290	pET22b(+)_*Aac*SHC_L36V	S. C. Hammer
pITB1026	pET22b(+)_*Aac*SHC_L36W	S. C. Hammer
pITB5433	pET22b(+)_AacSHC_L36Y	this thesis
pITB5434	pET22b(+)_AacSHC_D442A	this thesis
pITB5435	pET22b(+)_AacSHC_D442C	this thesis
pITB5436	pET22b(+)_AacSHC_W312L_A419D_Y495I	this thesis (J. Tulumovic)
pITB5438	pET22b(+)_AacSHC_D442F	this thesis
pITB5439	pET22b(+)_AacSHC_D442G	this thesis
pITB5440	pET22b(+)_AacSHC_D442H	this thesis
pITB5441	pET22b(+)_AacSHC_D442I	this thesis
pITB5442	pET22b(+)_AacSHC_D442K	this thesis
pITB5443	pET22b(+)_AacSHC_D442L	this thesis
pITB5444	pET22b(+)_AacSHC_D442M	this thesis
pITB5446	pET22b(+)_AacSHC_D442P	this thesis
pITB5447	pET22b(+)_AacSHC_D442Q	this thesis
pITB5448	pET22b(+)_AacSHC_D442R	this thesis
pITB5449	pET22b(+)_AacSHC_D442S	this thesis
pITB5450	pET22b(+)_AacSHC_D442T	this thesis
pITB5451	pET22b(+)_AacSHC_D442V	this thesis
pITB5452	pET22b(+)_AacSHC_D442W	this thesis
pITB5453	pET22b(+)_AacSHC_D442Y	this thesis
pITB1061	pET22b(+)_AacSHC_Y609A	S. C. Hammer
pITB5454	pET22b(+)_AacSHC_Y609C	this thesis
pITB5455	pET22b(+)_AacSHC_Y609D	this thesis
pITB5287	pET22b(+)_AacSHC_Y609E	S. Henche
pITB1317	pET22b(+)_AacSHC_Y609F	S. Bastian
pITB1318	pET22b(+)_AacSHC_Y609G	S. Bastian
pITB5456	pET22b(+)_AacSHC_Y609H	this thesis
pITB5457	pET22b(+)_AacSHC_Y609I	this thesis (J. Tulumovic)
pITB5458	pET22b(+)_AacSHC_Y609K	this thesis
pITB1319	pET22b(+)_AacSHC_Y609L	S. Bastian

pITB5288	pET22b(+)_AacSHC_Y609M	S. Bastian
pITB5459	pET22b(+)_AacSHC_Y609N	this thesis
pITB5460	pET22b(+)_AacSHC_Y609P	this thesis
pITB5461	pET22b(+)_AacSHC_Y609Q	this thesis
pITB5462	pET22b(+)_AacSHC_Y609R	this thesis
pITB5463	pET22b(+)_AacSHC_Y609S	this thesis
pITB5464	pET22b(+)_AacSHC_Y609T	A. Schneider
pITB5289	pET22b(+)_AacSHC_Y609V	S. Bastian
pITB1062	pET22b(+)_AacSHC_Y609W	S. Bastian
pITB5465	pET22b(+)_AacSHC_Y609W_Y420F	this thesis
pITB5466	pET22b(+)_AacSHC_L36A_S307A	this thesis
pITB5467	pET22b(+)_AacSHC_S309A	this thesis
pITB5468	pET22b(+)_AacSHC_S307A	this thesis
pITB5469	pET22b(+)_AacSHC_S307C	A. Schneider
pITB5470	pET22b(+)_AacSHC_Y609W_S309A	this thesis
pITB5471	pET22b(+)_AacSHC_Q366S	this thesis
pITB5472	pET22b(+)_AacSHC_P444D	this thesis
pITB5473	pET22b(+)_AacSHC_I308V	this thesis
pITB5474	pET22b(+)_AacSHC_D374A	this thesis (N. Kress)
pITB5475	pET22b(+)_AacSHC_D376A	this thesis (N. Kress)
pITB5476	pET22b(+)_AacSHC_D377A	this thesis (N. Kress)
pITB5477	pET22b(+)_AacSHC_D374A_D376_D377A	this thesis (N. Kress)
pITB5478	pET22b(+)_AacSHC_D442N_Q366E	this thesis
pITB5479	pET22b(+)_AacSHC_D442N_Q366S	this thesis
pITB5480	pET22b(+)_AacSHC_D442N_Y420F	this thesis
pITB5481	pET22b(+)_AacSHC_L36A_Q366E	this thesis
pITB5482	pET22b(+)_AacSHC_L36A_Y420F	this thesis
pITB5483	pET22b(+)_AacSHC_Y609W_Q366E	this thesis

8.6.3 SHCs from other organisms

SHCs from other organisms are listed in the following tables. Sequence similarities were calculated using the EMBOSS Needle alignment. The Needleman-Wunsch alignment algorithm was employed on the website www.ebi.ac.uk from the European Molecular Biology Laboratory in Hinxton.

Table 24: Compilation for *Tel*SHC.

***Tel*SHC**	
primary organism	*Thermosynechococcus elongatus*
vector	pET-22b(+)
ITB number	pITB0170
reference	literature[110,179]
protein sequence similarity with *Aac*SHC	29.4%
protein sequence	MPTSLATAIDPKQLQQAIRASQDFLFSQQYAEGYWWAELESNVTMTAEVILLHKIWG TEQRLPLAKAEQYLRNHQRDHGGWELFYGDGGDLSTSVEAYMGLRLLGVPETDPAL VKARQFILARGGISKTRIFTKLHLALIGCYDWRGIPSLPPWIMLLPEGSPFTIYEMSSW ARSSTVPLLIVMDRKPVYGMDPPITLDELYSEGRANVVWELPRQGDWRDVFIGLDRV FKLFETLNIHPLREQGLKAAEEWVLERQEASGDWGGIIPAMLNSLLALRALDYAVDD PIVQRGMAAVDRFAIETETEYRVQPCVSPVWDTALVMRAMVDSGVAPDHPALVKAG EWLLSKQILDYGDWHIKNKKGRPGGWAFEFENRFYPDVDDTAVVVMALHAVTLPN ENLKRRAIERAVAWIASMQCRPGGWAAFDVDNDQDWLNGIPYGDLKAMIDPNTADV TARVLEMVGRCQLAFDRVALDRALAYLRNEQEPEGCWFGRWGVNYLYGTSGVLTAL SLVAPRYDRWRIRRAAEWLMQCQNADGGWGETCWSYHDPSLKGKGDSTASQTAWA IIGLLAAGDATGDYATEAIERGIAYLLETQRPDGTWHEDYFTGTGFPCHFYLKYHYYQ QHFPLTALGRYARWRNLLAT

Table 25: Compilation for *Bam*SHC1.

***Bam*SHC1**	
primary organism	*Burkholderia ambifaria*
vector	pET-16b
ITB number	pITB0318
reference	Siedenburg et al[119]
protein sequence similarity with *Aac*SHC	27.9%
protein sequence	MNDLTEMATLSAGTVPAGLDAAVASATDALLAAQNADGHWVYELEADSTIPAEYVL LVHYLGETPNLELEQKIGRYLRRVQQADGGWPLFTDGAPNISASVKAYFALKVIGDD ENAEHMQRARRAIQAMGGAEMSNVFTRIQLALYGAIPWRAVPMMPVEIMLLPQWFP FHLSKVSYWARTVIVPLLVLNAKRPIAKNPRGVRIDELFVDPPVNAGLLPRQGHQSPG WFAFFRVVDHALRAADGLFPNYTRERAIRQAVSFVDERLNGEDGLGAIYPAMANAV MMYDVLGYAEDHPNRAIARKSIEKLLVVQEDEAYCQPCLSPVWDTSLAAHALLETGD ARAEEAVIRGLEWLRPLQILDVRGDWISRRPHVRPGGWAFQYANPHYPDVDDTAVV AVAMDRVQKLKHNDAFRDSIARAREWVVGMQSSDGGWGAFEPENTQYYLNNIPFS DHGALLDPPTADVSGRCLSMLAQLGETPLNSEPARRALDYMLKEQEPDGSWYGRWG MNYVYGTWTALCALNAAGLTPDDPRVKRGAQWLLSIQNKDGGWGEDGDSYKLNYR GFEQAPSTASQTAWALLGLMAAGEVNNPAVARGVEYLIAEQKEHGLWDETRFTATG FPRVFYLRYHGYRKFFPLWALARYRNLKRNNATRVTFGL

Table 26: Compilation for *Bam*SHC2.

***Bam*SHC2**	
primary organism	*Burkholderia ambifaria*
vector	pET-16b
ITB number	pITB0319
reference	Siedenburg et al[119]
protein sequence similarity with *Aac*SHC	27.4%
protein sequence	MIRRMNKSGPSPWSALDAAIARGRDALMRLQQPDGSWCFELESDATITAEYILMMHF MDKIDDARQEKMARYLRAIQRLDTHGGWDLYVDGDPDVSCSVKAYFALKAAGDSEH APHMVRARDAILELGGAARSNVFTRILLATFGQVPWRATPFMPIEFVLFPKWVPISMY KVAYWARTTMVPLLVLCSLKARARNPRNIAIPELFVTPPDQERQYFPPARGMRRAFL ALDRVVRHVEPLLPKRLRQRAIRHAQAWCAERMNGEDGLGGIFPPIVYSYQMMDVL GYPDDHPLRRDCENALEKLLVTRPDGSMYCQPCLSPVWDTAWSTMALEQARGVAVP EAGAPASALDELDARIARAYDWLAERQVNDLRGDWIENAPADTQPGGWAFQYANPY YPDIDDSAVVTAMLDRRGRTHRNADGSHPYAARVARALDWMRGLQSRNGGFAAFDA DCDRLYLNAIPFADHGALLDPPTEDVSGRVLLCFGVTKRADDRASLARAIDYVKRTQ QPDGSWWGRWGTNYLYGTWSVLAGLALAGEDPSQPYIARALAWLRARQHADGGWG ETNDSYIDPALAGTNAGESTSNCTAWALLAQMAFGDGESESVRRGIAYLQSVQQDDG FWWHRSHNAPGFPRIFYLKYHGYTAYFPLWALARYRRLAGGVSAAGAHAVPASTGA DAALA

Table 27: Compilation for *Zmo*SHC1.

***Zmo*SHC1**	
primary organism	*Zymomonas mobilis*
vector	pET-22b(+)
ITB number	pITB0280
reference	Dissertation Miriam Seitz[110] and previous identifications[214]
protein sequence similarity with *Aac*SHC	21.6%
protein sequence	MGIDRMNSLSRLLMKKIFGAEKTSYKPASDTIIGTDTLKRPNRRPEPTAKVDKTIFKT MGNSLNNTLVSACDWLIGQQKPDGHWVGAVESNASMEAEWCLALWFLGLEDHPLR PRLGNALLEMQREDGSWGVYFGAGNGDINATVEAYAALRSLGYSADNPVLKKAAAW IAEKGGLKNIRVFTRYWLALIGEWPWEKTPNLPPEIIWFPDNFVFSIYNFAQWARAT MVPIAILSARRPSRPLRPQDRLDELFPEGRARFDYELPKKEGIDLWSQFFRTTDRGLH WVQSNLLKRNSLREAAIRHVLEWIIRHQDADGGWGGIQPPWVYGLMALHGEGYQLY HPVMAKALSALDDPGWRHDRGESSWIQATNSPVWDTMLALMALKDAKAEDRFTPE MDKAADWLLARQVKVKGDWSIKLPDVEPGGWAFEYANDRYPDTDDTAVALIALSSY RDKEEWQKKGVEDAITRGVNWLIAMQSECGGWGAFDKDNNRSILSKIPFCDFGESID PPSVDVTAHVLEAFGTLGLSRDMPVIQKAIDYVRSEQEAEGAWFGRWGVNYIYGTGA VLPALAAIGEDMTQPYITKACDWLVAHQQEDGGWGESCSSYMEIDSIGKGPTTPSQT AWALMGLIAANRPEDYEAIAKGCHYLIDRQEQDGSWKEEEFTGTGFPGYGVGQTIKL DDPALSKRLLQGAELSRAFMLRYDFYRQFFPIMALSRAERLIDLNN

Table 28: Compilation for *Zmo*SHC2.

***Zmo*SHC2**	
primary organism	Zymomonas mobilis
vector	pET-16b
ITB number	pITB0282
reference	Dissertation Miriam Seitz[110]
protein sequence similarity with *Aac*SHC	27.3%
protein sequence	MTVSTSSAFHHSPLSDDVEPIIQKATRALLEKQQQDGHWVFELEADATIPAEYILLKH YLGEPEDLEIEAKIGRYLRRIQGEHGGWSLFYGGDLDLSATVKAYFALKMIGDSPDAP HMLRARNEILARGGAMRANVFTRIQLALFGAMSWEHVPQMPVELMLMPEWFPVHI NKMAYWARTVLVPLLVLQALKPVARNRRGILVDELFVPDVLPTLQESGDPIWRRFFS ALDKVLHKVEPYWPKNMRAKAIHSCVHFVTERLNGEDGLGAIYPAIANSVMMYDAL GYPENHPERAIARRAVEKLMVLDGTEDQGDKEVYCQPCLSPIWDTALVAHAMLEVG GDEAEKSAISALSWLKPQQILDVKGDWAWRRPDLRPGGWAFQYRNDYYPDVDDTAV VTMAMDRAAKLSDLHDDFEESKARAMEWTIGMQSDNGGWGAFDANNSYTYLNNIP FADHGALLDPPTVDVSARCVSMMAQAGISITDPKMKAAVDYLLKEQEEDGSWFGRW GVNYIYGTWSALCALNVAALPHDHLAVQKAVAWLKTIQNEDGGWGENCDSYALDYS GYEPMDSTASQTAWALLGLMAVGEANSEAVTKGINWLAQNQDEEGLWKEDYYSGG GFPRVFYLRYHGYSKYFPLWALARYRNLKKANQPIVHYGM

Table 29: Compilation for *Ttu*SHC2.

***Ttu*SHC2**	
primary organism	*Teredinibacter turnerae*
vector	pET-16b
ITB number	pITB0320
reference	Siedenburg et al[119]
protein sequence similarity with *Aac*SHC	29.9%
protein sequence	MEIQDEVDLLEPQESLTASADSAVDRALFWLLDAQYEDGYWAGILESNACMEAEWL LCFHVLGIANHPMSRGLVQGLLQRQRADGSWDVYYGARAGDINTTVEVYAALRCQG YAADHPDIKRARDWIQLQGGVKQVRVFTRFWLALIGEWPWEETPNLPPEILFFPRWF PFNIYHFAAWARATLVPLCILSARRMVVPLNKKSCLQELFPEDRSAVVALGKKAGAW STFFYHADRALKKYQRTFKRPPGRQQAIKMCLEWILRRQDADGAWGGIQPPWIYSLM ALKAEGYPVTHPVMAKGLAALDAHWSYERPGGARFVQACESPVWDTLLSSFALLDC GFSCTSSSELRKAVDWILDQQVLLPGDWQQKLPTVSPGGWAFERANVHYPDVDDTA VALIVLAKVRPDYPDTARVNLAIERGLNWLFAMQCRNGGWGAFDKDNDKDLLTKIP FSDFGETIDPASVDVTAHVLEALGLLGYRTTHPAVAKALEFIRSEQENDGCWFGRWG VNYIYGTAAVLPALASLNMNMNQEFIRRAANWILGKQNNDGGWGESCASYMDDTQ RGRGPSTASQTAWAMMSLLAVDGGTYAESLLRAEAYLKTTQTPEGTWDEPYYTGTG FPGYGIGRREIKRQRSLQQHAELSRGFMINYNLYRHYFPLMALGRLAALRGA

Table 30: Compilation for *Sfu*SHC.

***Sfu*SHC**	
primary organism	*Syntrophobacter fumaroxidans*
vector	pET-16b
ITB number	pITB0316
reference	Siedenburg et al[119]
protein sequence similarity with *Aac*SHC	33.4%
protein sequence	MNPIRGKRGSAADFLEEEYQWENLADHGESGRTPGGGHPAALKEYEAGSATEHTGH HCVHHLGVRNSWLRKIEKAIDNACGQLFKTQYEDGYWWSELESNVTITSEYIMLLYL LEVSRPEQQKSMVKYLLNQQRPDGSWGLYYGDGGNLSTTIEAYFALKLAGEHCESE PMRRAREFILSKGGIESARVFTKIWLALFSQYDWDKVPSMPVELVLLPSSLYFNIYEF SSWARGTVVPLSIVMSIRPRCPLPAKCSIKELYVPGSKHKNFASCTHKLFFLFDRIAKA FERRPVPSLRNKAVQAAETWVLDHQEDSGDWGGIQPPMVYSVLALYYLGYPLDHEV IVKGIKALDAFCMEDEEGTRMQSCVSPVWDTALTVLSMLDAGVAAEHPGLEKAGRW LLENQVLTGGDWQIKNDSLPGGWAFEFYNTRYPDVDDSAVVLSTLNRFNAERVEGL EFAKCRGMEWCLSMQSSNGGWAAFDKDNTLEILNRIPFADQEAMVDYPTADVTGRV LEAMGYLGYDGSHPRARKAIQFLKKRQERDGCWWGRWGVNYIYGTWSVLKGLISIG EDPRAAYIRAAVRWVKDHQNSDGGWGETCESYENPELRGQGPSTPSQTAWALMSLI ACGEMKSQEASRGIQYLLRTQKRDGTWEELHFTGTGFPKHFYIRYHNYRNCFPLMA LGQYLRALER

Table 31: Compilation for *Sco*SHC.

***Sco*SHC**	
primary organism	*Streptomyces coelicolor*
vector	pET16-b
ITB number	pITB0315
reference	Siedenburg et al[119]
protein sequence similarity with *Aac*SHC	32.7%
protein sequence	MTATTDGSTGASLRPLAASASDTDITIPAAAAGVPEAAARATRRATDFLLAKQDAEGW WKGDLETNVTMDAEDLLLRQFLGIQDEETTRAAALFIRGEQREDGTWATFYGGPGE LSTTIEAYVALRLAGDSPEAPHMARAAEWIRSRGGIASARVFTRIWLALFGWWKWDD LPELPPELIYFPTWVPLNIYDFGCWARQTIVPLTIVSAKRPVRPAPFPLDELHTDPARP NPPRPLAPVASWDGAFQRIDKALHAYRKVAPRRLRRAAMNSAARWIIERQENDGCW GGIQPPAVYSVIALYLLGYDLEHPVMRAGLESLDRFAVWREDGARMIEACQSPVWDT CLATIALADAGVPEDHPQLVKASDWMLGEQIVRPGDWSVKRPGLPPGGWAFEFHND NYPDIDDTAEVVLALRRVRHHDPERVEKAIGRGVRWNLGMQSKNGAWGAFDVDNT SAFPNRLPFCDFGEVIDPPSADVTAHVVEMLAVEGLAHDPRTRRGIQWLLDAQETDG SWFGRWGVNYVYGTGSVIPALTAAGLPTSHPAIRRAVRWLESVQNEDGGWGEDLRS YRYVREWSGRGASTASQTGWALMALLAAGERDSKAVERGVAWLAATQREDGSWDE PYFTGTGFPWDFSINYNLYRQVFPLTALGRYVHGEPFAKKPRAADAPAEAAPAEVKG S

Table 32: Compilation for *Rpa*SHC.

***Rpa*SHC**	
primary organism	*Rhodopseudomonas palustris*
vector	pET-16b
ITB number	pITB0314
reference	Siedenburg et al[119]
protein sequence similarity with *Aac*SHC	30.3%
protein sequence	MDSILAPRADAPRNIDGALRESVQQAADWLVANQKPDGHWVGRAETNATMEAQWC LALWFLGLEDHPLRVRLGRALLDTQRPDGAWHVFYGAPNGDINATVEAYAALRSLG HRDDEEPLRKARDWILSKGGLANIRVFTRYWLALIGEWPWEKTPNILPEVIWLPTWF PFSIYNFAQWARATLMPIAVLSAHRPSRPLAPQDRLDALFPQGRDSFNYDLPARLGAG VWDVIFRKIDTILHRLQDWGARRGPHGIMRRGAIDHVLQWIIRHQDYDGSWGGIQPP WIYGLMALHTEGYAMTHPVMAKALDALNEPGWRIDIGDATFIQATNSPVWDTMLSL LAFDDAGLGERYPEQVERAVRWVLKRQVLVPGDWSVKLPDVKPGGWAFEYANNFY PDTDDTSVALMALAPFRHDPKWQAEGIEDAIQRGIDWLVAMQCKEGGWGAFDKDN DKKILAKIPFCDFGEALDPPSADVTAHIIEAFAKVGLDRNHPSIVRALDYLKREQEPE GPWFGRWGVNYVYGTGAVLPALAAIGEDMRQPYIARACDWLIARQQANGGWGESCV SYMDAKQAGEGTATASQTAWALMALIAADRPQDRDAIERGCLYLTETQRDGTWQEV HYTGTGFPGYGVGQTIKLNDPLLSKRLMQGPELSRSFMLRYDLYRHYFPMMAIGRVL RQRGDRSGH

Table 33: Compilation for *Apa*SHC.

***Apa*SHC**	
primary organism	*Acetobacter pasteurianus*
vector	pET-16b
ITB number	pITB0312
reference	Siedenburg et al[119]
protein sequence similarity with *Aac*SHC	28.3%
protein sequence	MNMASRFSLKKILRSGSDTQGTNVNTLIQSGTSDIVRQKPAPQEPADLSALKAMGNS LTHTLSSACEWLMKQQKPDGHWVGSVGSNASMEAEWCLALWFLGLEDHPLRPRLG KALLEMQRPDGSWGTYYGAGSGDINATVESYAALRSLGYAEDDPAVSKAAAWIISKG GLKNVRVFTRYWLALIGEWPWEKTPNLPPEIIWFPDNFVFSIYNFAQWARATMMPLA ILSARRPSRPLRPQDRLDALFPGGRANFDYELPTKEGRDVIADFFRLADKGLHWLQSS FLKRAPSREAAIKYVLEWIIWHQDADGGWGGIQPPWVYGLMALHGEGYQFHHPVM AKALDALNDPGWRHDKGDASWIQATNSPVWDTMLSLMALHDANAEERFTPEMDKA LDWLLSRQVRVKGDWSVKLPNTEPGGWAFEYANDRYPDTDDTAVALIAIASCRNRP EWQAKGVEEAIGRGVRWLVAMQSSCGGWGAFDKDNNKSILAKIPFCDFGEALDPPS VDVTAHVLEAFGLLGLPRDLPCIQRGLAYIRKEQDPTGPWFGRWGVNYLYGTGAVLP ALAALGEDMTQPYISKACDWLINCQQENGGWGESCASYMEVSSIGHGATTPSQTAW ALMGLIAANRPQDYEAIAKGCRYLIDLQEEDGSWNEEEFTGTGFPGYGVGQTIKLDD PAISKRLMQGAELSRAFMLRYDLYRQLFPIIALSRASRLIKLGN

Table 34: Compilation for *Bja*SHC.

***Bja*SHC**	
primary organism	*Bradyrhizobium japonicum*
vector	pET-16b
ITB number	pITB0317
reference	Siedenburg et al[119]
protein sequence similarity with *Aac*SHC	30.9%
protein sequence	MTVTSSASARATRDPGNYQTALQSTVRAAADWLIANQKPDGHWVGRAESNACMEAQ WCLALWFMGLEDHPLRKRLGQSLLDSQRPDGAWQVYFGAPNGDINATVEAYAALRS LGFRDDEPAVRRAREWIEAKGGLRNIRVFTRYWLALIGEWPWEKTPNIPPEVIWFPL WFPFSIYNFAQWARATLMPIAVLSARRPSRPLPPENRLDALFPHGRKAFDYELPVKA GAGGWDRFFRGADKVLHKLQNLGNRLNLGLFRPAATSRVLEWMIRHQDFDGAWGG IQPPWIYGLMALYAEGYPLNHPVLAKGLDALNDPGWRVDVGDATYIQATNSPVWDTI LTLLAFDDAGVLGDYPEAVDKAVDWVLQRQVRVPGDWSMKLPHVKPGGWAFEYAN NYYPDTDDTAVALIALAPLRHDPKWKAKGIDEAIQLGVDWLIGMQSQGGGWGAFDK DNNQKILTKIPFCDYGEALDPPSVDVTAHIIEAFGKLGISRNHPSMVQALDYIRREQE PSGPWFGRWGVNYVYGTGAVLPALAAIGEDMTQPYIGRACDWLVAHQQADGGWGE SCASYMDVSAVGRGTTTASQTAWALMALLAANRPQDKDAIERGCMWLVERQSAGT WDEPEFTGTGFPGYGVGQTIKLNDPALSQRLMQGPELSRAFMLRYGMYRHYFPLMA LGRALRPQSHS

Table 35: Compilation for *Pca*SHC.

***Pca*SHC**	
primary organism	*Pelobacter carbinolicus*
vector	pET-16b
ITB number	pITB0313
reference	unpublished ITB in-house data
protein sequence similarity with *Aac*SHC	28.4%
protein sequence	MDKIKMKNINQPKFRVFRGGQKAATPCPGTTNERRGALDRGRLSASLKHSREWLLS LQADAGNWVFALEADTTIASEYVMLQRFLGRPLAPELQQRLANYLLSRQLPDGGWPL YAEDGFANISTTVKAYLALKLLGYPTHCDPLVRARQIVLALGGAEKCNVFTRIALALF GQIPWRTTPAMPVEIMLLPRWFYFHLSKISYWARTVVVPLLILYAKRPVCRLEPWEGI PELFVTPPDKLGYLDVCKPGQWRKNVFIWVDRLTRKMVRCVPRRLHNLALRAAETW TREHMQGAGGIGAIFPAMANAVMALRTLGCSPDDADYQRGLKALDDLLIDRCDVPPR EDTPVSPCWCTGTSAAPMLDPSPAGSHAQGGDQGICQPCASPIWDTGLALTALLEGG LDARHPAVDRAVRWLLDQQVDVKGDWAQRVPNLEAGGWAFQFENALYPDLDDTSK VLMSLIRAGAMDNPGYRQELSRAINWVIGMQNSDGGWGAFDVDNNYLYLNDIPFAD HGALLDPSTADVTGRCIEMLAMAGFGRDFLPIARGVDFLRREQEDFGGWYGRWGVN YIYGTWSALSGLIHAGEDLQAPYIRQAVGWLESVQNPDGGWGETCYSYDDPALAGRG VSTASQTAWALLGLMAAGEVDNLAVRRGIQYLVEEQNRAGGWDERHFTGTGFPRVF YLRYHGYSQYFPLWALGLYERLSSGNPSRQQMVRRAGPAGLHLPVLDRRKKLRRKR KA

8.7 Tables for product distributions and conversion

Product distributions and conversions that were previously shown in diagrams are listed in numbers in the following.

8.7.1 Time dependency of the reaction

A clear correlation between the reaction time and the conversion was observed using BT3 protocol and GC-P2 temperature program (see chapter 4.1.1.6) . The obtained conversions for several time points are listed in Table 36.

Table 36: Conversion in % at different time points in hours.

hours	conversion	standard deviation
1	0.7	0.1
2	0.8	0.5
3	1.3	0.1
4	1.3	0.1
5	1.5	0.1
6	1.9	0.3
7	1.8	0.3
18	2.8	0.1
20	3.1	1.8
22	4.6	0.9
24	5.5	1.1
26	5.6	0.4
28	5.2	0.1
30	5.6	0.1
32	5.6	0.1
42	8.2	0.4
44	9.1	0.2
46	9.2	0.2

8.7.2 Investigation of negative controls

Data shown for the discussion on formation of α-terpineol were obtained using BT2 protocol and temperature program GC-P1 (see chapter 4.1.2.3) . The resulting product distributions are shown in Table 37.

Table 37: Product distributions in % for the conversion of β-pinene using wt, L36A, Y609W with BT2 protocol and GC-P1 temperature program for a comparison of α-terpineol production.

product	wt	L36A	Y609W
α-pinene	54.0 ± 0.9	11.9 ± 0.3	89.3 ± 0.2
β-camphene	26.2 ± 0.6	11.7 ± 0.3	5.4 ± 0.4
limonene	13.1 ± 0.2	60.6 ± 1.4	4.3 ± 0.3
α-terpineol	6.7 ± 0.5	15.8 ± 1.3	1.07 ± 0.3

8.7.3 Detection of trace products

Using method BT8 several side products were detectable. Especially, variant W312A showed an increase in side product formation at expense of the main products typically observed for the wild type (see chapter 4.1.2.4). The products for wild type, W312A, L36A, D442N, Y609W and their distributions are given in Table 38.

Table 38: Product distributions for main and side products. Numbers are given in %. The data were obtained using BT8 protocol and GC-P3 temperature program.

product	wt	W312A	L36A	D442N	Y609W
α-pinene	39.2 ± 4.3	6.1 ± 0.4	8.4 ± 0.3	16.3 ± 0.9	75.6 ± 2.2
β-camphene	25.8 ± 1.3	5.0 ± 0.1	11.9 ± 0.6	39.0 ± 1.6	5.4 ± 0.3
α-terpinene	0.5 ± 0.4	-	-	-	-
limonene	19.8 ± 1.4	22.3 ± 0.5	37.3 ± 1.6	10.7 ± 0.6	2.8 ± 0.4
γ-terpinene	0.2 ±0.2	1.3 ± 0.6	0.3 ± 0.2	-	-
unidentified 1	0.5 ± 0.6	-	0.2 ± 0.2	1.1 ± 0.4	-
α-terpinolene	1.5 ± 1.4	1.9 ± 0.3	1.6 ± 0.3	1.0 ± 0.4	0.3 ± 0.1
fenchol	0.1 ± 0.2	3.1 ± 1.6	1.4 ± 0.1	5.5 ± 0.6	0.2 ± 0.1
unidentified 2	2.5 ± 1.5	-	2.5 ± 0.3	0.8 ± 0.8	-
unidentified 3	1.0 ± 0.6	10.3 ± 0.3	1.0 ± 0.8	0.9 ± 0.8	0.1 ± 0.2
unidentified 4	1.0 ± 1.1	1.1 ± 0.1	0.2 ± 0.2	1.3 ± 2.3	1.0 ± 0.7
unidentified 5	0.9 ± 2.1	4.5 ± 0.1	1.6 ± 0.4	2.0 ± 1.8	0.7 ± 1.1
borneol	0.4 ± 0.3	10.1 ± 0.2	1.0 ± 0.1	-	-
terpinen-4-ol	0.2 ± 0.3	2.0 ± 0.3	1.3 ± 0.2	1.0 ± 0.5	-
α-terpineol	6.5 ± 4.0	32.2 ± 2.5	20.3 ± 1.9	8.0 ± 1.2	5.0 ± 0.2

8.7.4 Saturation mutagenesis

The formation of α-pinene, ß-camphene and limonene are given for the saturation mutagenesis experiments at positions Y609, D442 and L36 (see chapter 4.1.3.3). The variants are sorted by their selectivity. The data were obtained using BT7 and protocol GC-P4 temperature program and are shown in 39-41.

Table 41: α-pinene formation for the saturation at position Y609. Numbers are given in %.

variant	α-pinene formation
Y609W	84.8 ± 1.2
Y609F	40.4 ± 1.0
wt*	38.6 ± 4.2
Y609N	30.2 ± 1.9
Y609H	29.4 ± 2.5
Y609E	29.2 ± 0.8
Y609V	28.4 ± 1.7
Y609P	28.3 ± 0.5
Y609T	24.9 ± 2.6
Y609A	24.1 ± 2.8
Y609Q	23.6 ± 2.9
Y609I	23.4 ± 0.4
Y609D	23.0 ± 2.9
Y609M	23.0 ± 1.4
Y609G	22.4 ± 4.1
Y609L	21.0 ± 0.5
Y609C	20.1 ± 3.2

Table 40: Limonene formation for the saturation at position L36. Numbers are given in %.

variant	limonene formation
L36A	66.4 ± 0.4
L36D	65.0 ± 1.5
L36T	62.1 ± 3.5
L36C	61.0 ± 1.6
L36S	58.2 ± 5.1
L36E	55.0 ± 3.9
L36V	54.0 ± 4.7
L36N	45.6 ± 3.3
L36P	42.2 ± 1.5
L36I	36.2 ± 4.7
L36Q	30.0 ± 7.1
L36K	29.7 ± 0.1
wt	19.5 ± 1.0
L36Y	19.0 ± 1.3
L36F	15.5 ± 0.6
L36W	15.2 ± 1.0

Table 39: β-camphene formation for the saturation at position D442. Numbers are given in %.

variant	camphene formation
D442P	53.1 ± 1.8
D442C	48.3 ± 3.3
D442N	45.9 ± 1.2
D442G	44.0 ± 2.1
D442E	42.3 ± 2.1
D442I	41.7 ± 1.2
D442Q	41.3 ± 0.4
D442H	41.1 ± 2.8
D442V	39.4 ± 2.2
D442M	39.4 ± 2.5
D442F	39.0 ± 4.4
D442Y	37.6 ± 2.6
D442T	36.2 ± 2.3
D442L	35.9 ± 1.2
D442A	35.6 ± 3.4
wt	24.7 ± 1.8

* The seemingly low α-pinene production compared to other results is due to the high number of side products that were detected with BT7 protocol. The overall product distribution trends are in agreement with results detected with previous BT protocols

8.7.5 In-silico insights

The product distributions for variants based on *in-silico* insights (see chapter 4.1.3.4) are given in Table 42 in %. Relative conversions are given in relation to the respective variants in %.

Table 42: Product distributions for variants of the *in silico* investigation. Numbers are given in %. Relative conversion are given in % in relation to the given variant. The data were obtained using BT7 protocol and GC-P4 temperature program.

	α-pinene	β-camphene	limonene	α-terpineol	minor products	relative conversion
Y609W	83.9 ± 1.9	6.7 ± 0.7	5.5 ± 0.4	1.2 ± 0.4	2.6	100 set to 100%
Y609W_S309A	63.6 ± 3.8	6.4 ± 1.4	11.7 ± 1.5	9.2 ± 4.2	9.5	6 to Y609W
S309A	34.8 ± 2.9	27.0 ± 1.4	17.4 ± 0.5	7.0 ± 1.8	13.8	15 to Y609W
wt	39.2 ± 4.3	25.8 ± 1.3	19.8 ± 1.3	6.5 ± 4.0	8.7	100 set to 100%
L36A	8.1 ± 0.8	12.4 ± 0.3	65.7 ± 5.7	4.6 ± 5.2	9.2	100 set to 100%
L36A_S307A	7.0 ± 2.7	13.7 ± 0.7	61.0 ± 3.2	9.4 ± 1.8	8.9	84 to L36A
S307A	29.6 ± 8.1	4.7 ± 12.4	6.0 ± 3.5	56.3 ± 32.0	3.5	5 to L36A
P444D	45.6 ± 1.5	24.7 ± 0.4	18.1 ± 0.5	4.5 ± 0.1	11.4	109 to wt
Q366S	44.6 ± 7.1	23.2 ± 0.4	15.9 ± 0.6	8.3 ± 4.6	16.6	220 to wt
D442N_Q366S	20.1 ± 1.7	16.4 ± 2.3	18.1 ± 1.4	8.7 ± 1.9	6.7	161 to D442N
D442N	20.0 ± 1.3	45.8 ± 3.3	20.1 ± 1.0	6.5 ± 2.7	7.7	100 set to 100%

8.7.6 Exploration of the natural diversity of SHCs

The results for a screening using SHCs from different host organisms (see chapter 4.1.3.5) are shown in Table 43. Product distributions are given in %. Relative conversions are given in % in relation to *Aac*SHC wild type. Variants that were created in *Aac*SHC based on a comparison with *Tel*SHC were tested and their product distributions are given in % in Table 44.

Table 43: Product distributions for the comparison of SHCs from different organisms given in %. Relative conversions are given in % in relation to *Aac*SHC wt. All data were obtained using BT6 protocol and GC-P2 where α-terpineol could not be detected.

	α-pinene	β-camphene	limonene	terpinolene	γ-terpinene	relative conversion
***Aac*SHC wt**	55.5 ± 0.2	28.7 ± 1.0	19.9 ±1.0	-	-	100
***Bam*SHC 2**	46.8 ± 7.1	23.6 ± 4.8	29.6 ± 2.3	-	-	15
***Smo*SHC**	36.1 ± 0.3	24.5 ± 0.6	33.6 ± 2.8	5.7 ± 2.6	-	59
***Apa*SHC**	21.0 ± 0.5	6.4 ± 1.1	72.6 ± 1.6	-	-	17
***Bja*SHC**	12.4 ± 0.1	9.8 ± 0.6	67.0 ± 1.6	9.0 ± 2.0	1.8 ± 0.1	189
***Tel*SHC**	77.0 ± 1.6	10.2 ± 0.7	11.7 ± 2.4	1.1 ± 1.5	-	172
***Pca*SHC**	17.8 ± 0.5	15.1 ± 1.1	61.7 ± 1.4	5.3 ± 1.4	-	43
***Rpa*SHC**	21.0 ± 0.1	7.2 ± 0.7	67.0 ± 2.5	4.9 ± 1.9	-	67
***Sco*SHC**	16.6 ± 0.2	19.5 ± 1.5	62.5 ± 3.7	1.4 ± 2.0	-	57

Table 44: Product distributions for the comparison of *Aac*SHC variants created on the basis of *Tel*SHC and combined with already selective variants, given in %. Relative conversion are given in % in relation to the given variant or set to 100%. The data were obtained using BT7 protocol and GC-P4 temperature program.

	α-pinene	β-camphene	limonene	α-terpineol	minor products	relative conversion
***Aac*SHC wt**	93.2 ± 4.3	25.8 ± 1.3	19.8 ± 1.3	6.5 ± 1.3	8.7	100 to *Aac*SHC
***Tel*SHC**	62.6 ± 6.5	12.5 ± 1.4	9.6 ± 0.5	4.9 ± 4.1	10.4	172 to *Aac*SHC
Y420F	48.8 ± 4.0	23.5 ± 0.5	12.2 ± 0.6	8.7 ± 2.2	15.2	251 to *Aac*SHC
S307C	42.1 ± 0.3	25.7 ± 0.3	20.9 ± 0.2	4.2 ± 0.4	9.4	86 to *Aac*SHC
I308V	44.2 ± 0.5	22.6 ± 0.7	24.7 ± 0.8	3.5 ± 0.2	8.5	78 to *Aac*SHC
Q366E	32.0 ± 5.4	26.8 ± 2.0	15.9 ± 1.6	11.5 ± 2.7	24.3	205 to *Aac*SHC
L36A	8.9 ± 1.0	11.7 ± 0.6	65.9 ± 0.8	7.6 ± 1.5	5.7	100 set to 100%
L36A_Q366E	7.2 ± 0.4	11.0 ± 0.4	64.5 ± 3.4	7.9 ± 2.6	9.2	62 to L36A
L36A_Y420F	16.6 ± 0.6	15.0 ± 0.2	54.2 ± 0.6	6.0 ± 0.9	7.8	123 to L36A
D442N	20.0 ± 1.3	45.8 ± 3.3	20.1 ± 1.0	6.5 ± 2.7	7.7	100 set to 100%
D442N_Q366E	18.9 ± 2.0	41.1 ± 0.9	17.1 ± 1.4	10.3 ± 1.8	12.5	186 to D442N
D442N_Y420F	26.3 ± 1.5	46.5 ± 0.7	11.6 ± 0.3	9.1 ± 1.7	6.6	349 to D442N
Y609W	84.3 ± 1.7	6.6 ± 0.4	5.8 ± 0.4	1.4 ± 0.8	1.8	100 set to 100%
Y609W_Q366E	85.4 ± 0.9	7.6 ± 0.4	4.6 ± 0.1	1.1 ± 0.1	1.3	151 to Y609W
Y609W_Y420F	87.2 ± 1.5	5.0 ± 0.3	3.8 ± 0.3	1.5 ± 0.9	2.4	73 to Y609W

8.7.7 Summary of the most selective and active variants

A summary of variants with a good combination of selectivity and relative conversion (see chapter 4.1.4) are listed in Table 45 with selectivities only given for the product of interest. Relative conversions are listed in relation to the wild type.

Table 45: Summary of the best variants concerning product selectivity and relative conversion given in %. The wild type is set to 100%. All data were obtained using BT7 protocol and temperature program GC-P4.

	product selectivity	relative conversion
	towards α-pinene	
wt	38.6 ± 4.2	100
Y609W	84.8 ± 1.2	120
Y609W_Q366E	84.3 ± 1.7	181
	towards limonene	
wt	19.5 ± 1.0	100
L36A	66.4 ± 0.4	130
L36A_Y420F	54.2 ± 0.6	160
	towards ß-camphene	
wt	24.7 ± 1.8	100
D442P	53.1 ± 1.8	140
D442N_Y420F	46.5 ± 0.7	175

8.7.8 Product distribution over time for the conversion of sabinene

The different product distributions for the conversion of sabinene with the variant Q366F at different time points (see chapter 4.2.2) are listed in Table 46.

Table 46: Product distribution for the conversion of Q366F with sabinene for 20h and 40h. Product distributions are given in %. All data were obtained using BT5 protocol and GC-S1 temperature program.

product	20 h	40h
P1	36.5 ± 2.3	24.4 ± 2.2
P2	1.7 ± 0.2	-
P3	9.2 ± 0.1	5.9 ± 0.2
P4	20.9 ± 0.7	25.2 ± 1.3
P5	14.6 ± 0.2	21.5 ± 0.9
P6	17.1 ± 1.4	23.1 ± 0.8

8.8 Chromatographic spectra

In this section some chromatographic spectra of different experiments are depicted.

8.8.1 Spectra for variant W312A and temperature program GC-P3

In the following are overlaid chromatograms of variant W312A, reference monoterpenes and buffer shown. They are either shown in full length (Figure 72 and Figure 74) or as sections (Figure 73, Figure 75 and Figure 76) for more details.

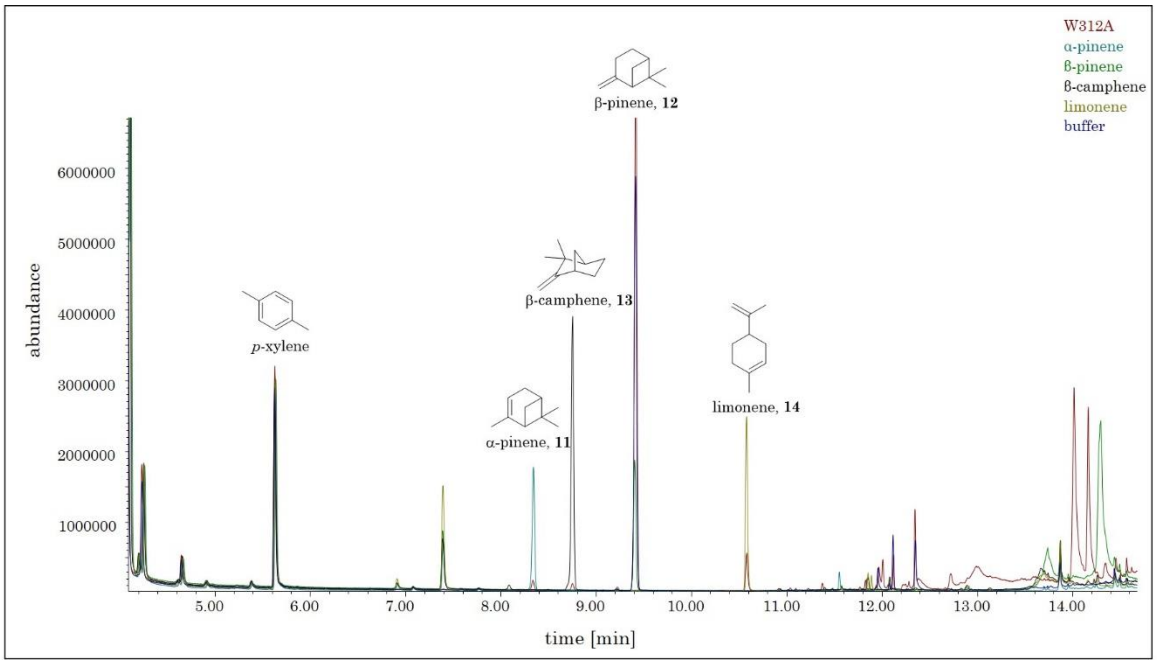


Figure 72: Full length of overlaid chromatograms of *Aac*SHC variant W312A and reference monoterpenes α-pinene (8.3 min), camphene (8.7 min), β-pinene (9.4 min), limonene (10.5 min) and buffer containing *p*-xylene (5.9 min) for comparison, using temperature program GC-P3.

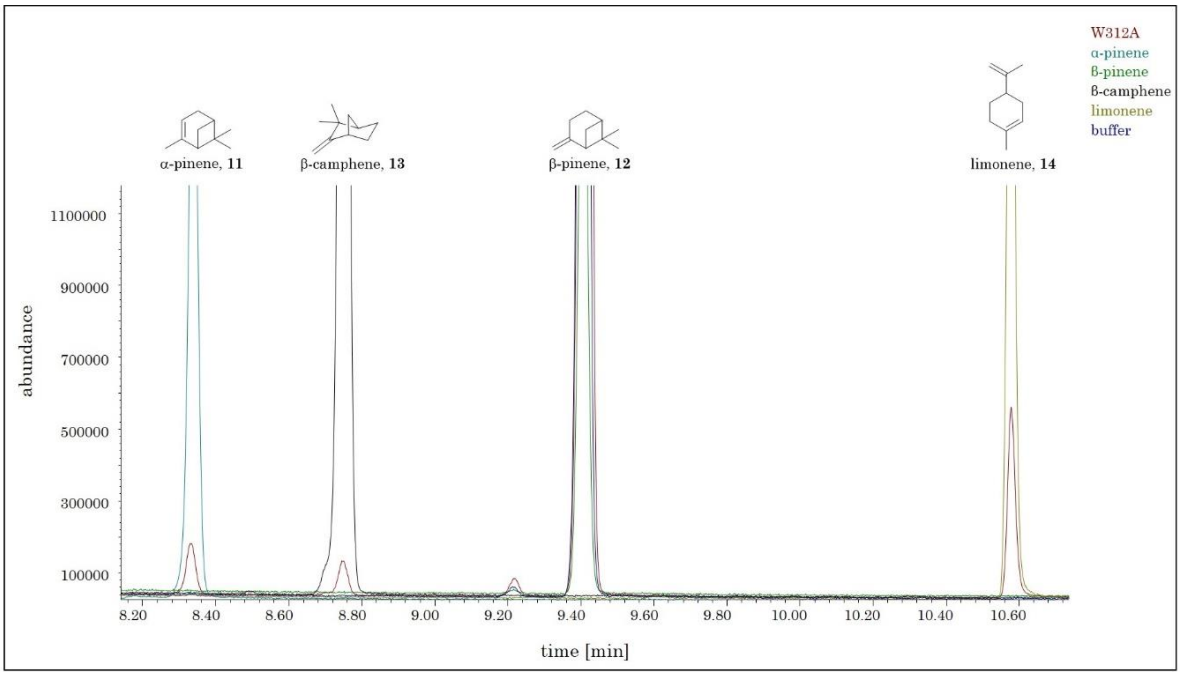


Figure 73: Section of overlaid chromatograms of *Aac*SHC variant W312A and reference monoterpenes α-pinene (8.3 min), camphene (8.7 min), β-pinene (9.4 min) and limonene (10.5 min) and buffer containing *p*-xylene (5.9 min) for comparison using temperature program GC-P3.

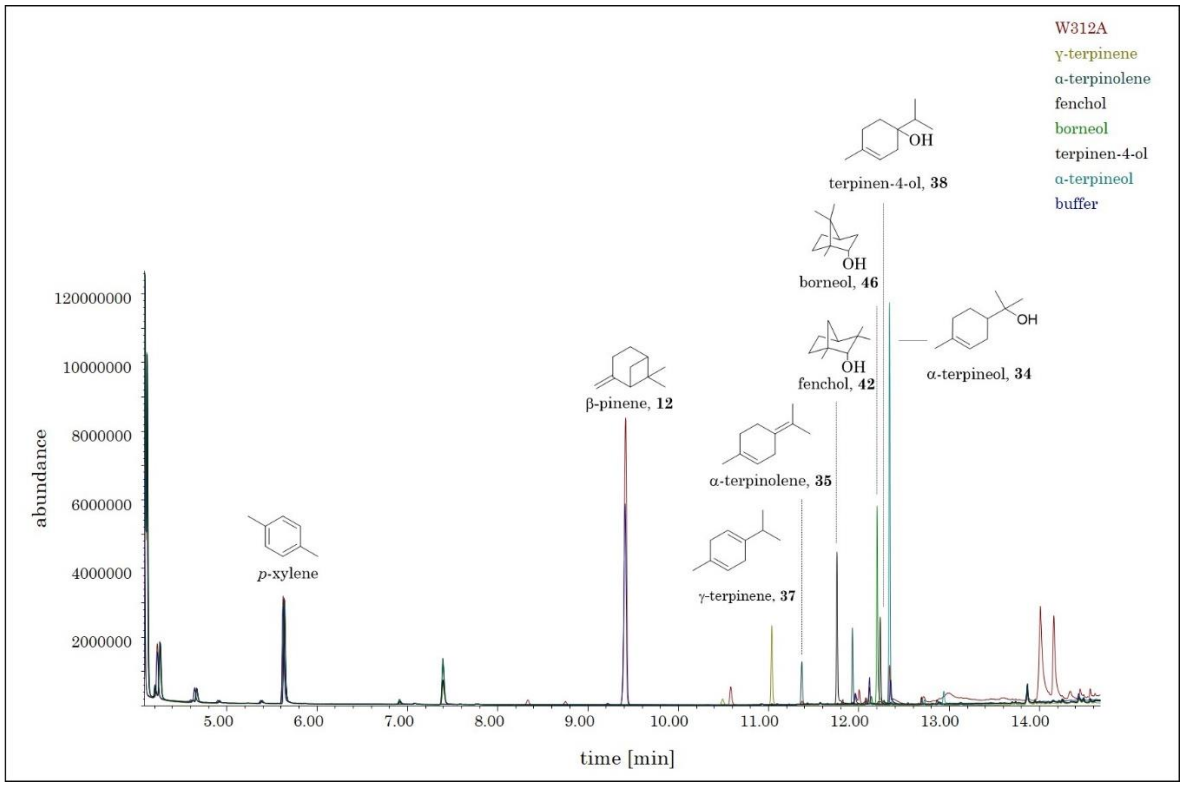


Figure 74: Full length of overlaid chromatograms of *Aac*SHC variant W312A and reference monoterpenes γ-terpinene (11.0 min), α-terpinolene (11.3 min), fenchol (11.7 min), borneol (12.1 min), terpinen-4-ol (12.2 min), α-terpineol (12.3 min) and buffer containing *p*-xylene (5.9 min) for comparison using the temperature program GC-P3.

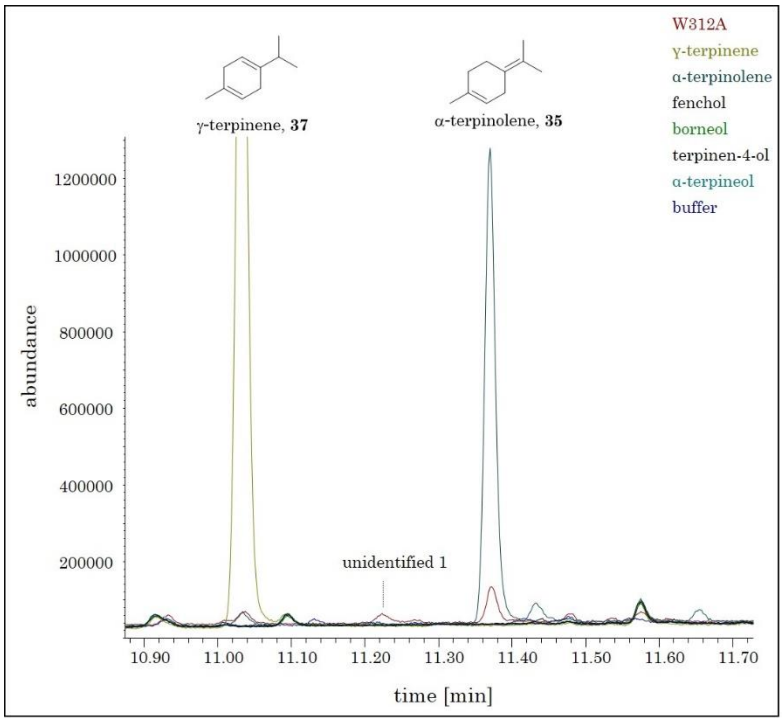


Figure 75: Section of overlaid chromatograms of *Aac*SHC variant W312A and reference monoterpenes γ-terpinene (11.0 min), α-terpinolene (11.3 min), fenchol (11.7 min), borneol (12.1 min), terpinen-4-ol (12.2 min), α-terpineol (12.3 min) and buffer containing *p*-xylene (5.9 min) for comparison using the temperature program GC-P3. Shown are γ-terpinene and α-terpinolene.

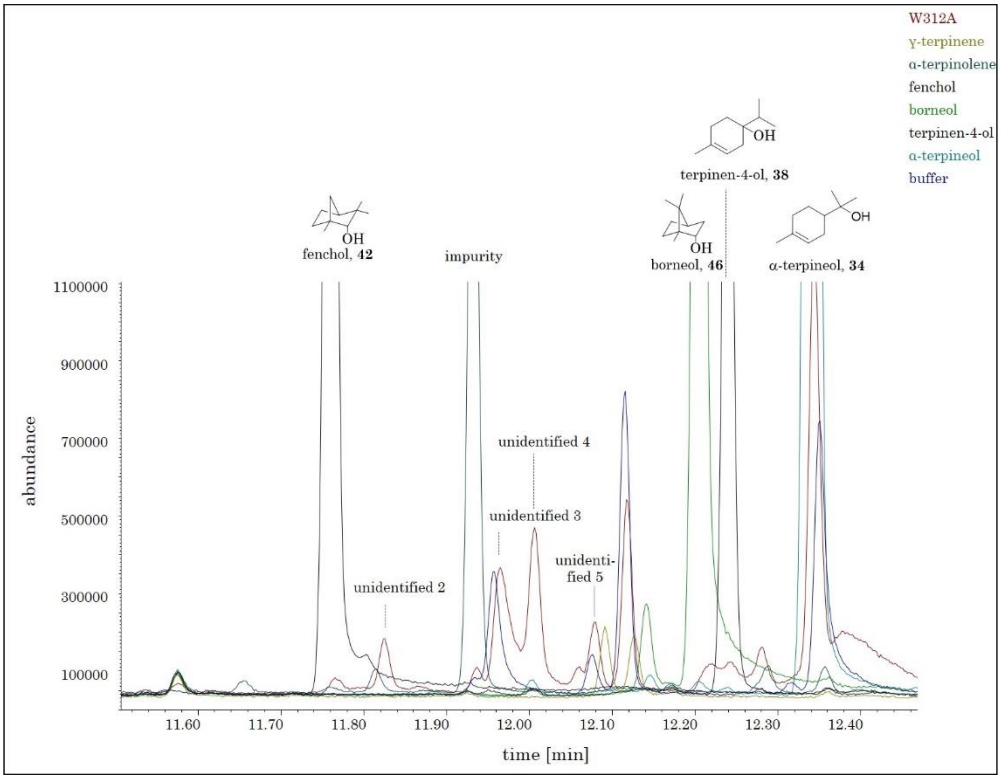


Figure 76: Section of overlaid chromatograms of *Aac*SHC variant W312A and reference monoterpenes γ-terpinene (11.0 min), α-terpinolene (11.3 min), fenchol (11.7 min), borneol (12.1 min), terpinen-4-ol (12.2 min), α-terpineol (12.3 min) and buffer containing *p*-xylene (5.9 min) for comparison using the temperature program GC-P3. Shown are fenchol, borneol, terpinen-4-ol and α-terpineol .

8.8.2 Spectra for temperature program GC-P4

In the following are overlaid chromatograms of reference monoterpenes shown for temperature program GC-P4. They are first shown without biotransformations for more clarity (Figure 77 - Figure 80) and then in comparison with the wildtype, buffer and variant AxAA for a better peak comparison (Figure 81 - Figure 84). The reference monoterpenes exhibit traces of other monoterpenes due to impurities. However, the main peak can be clearly assigned for each reference monoterpene and was cross-checked with NIST database.

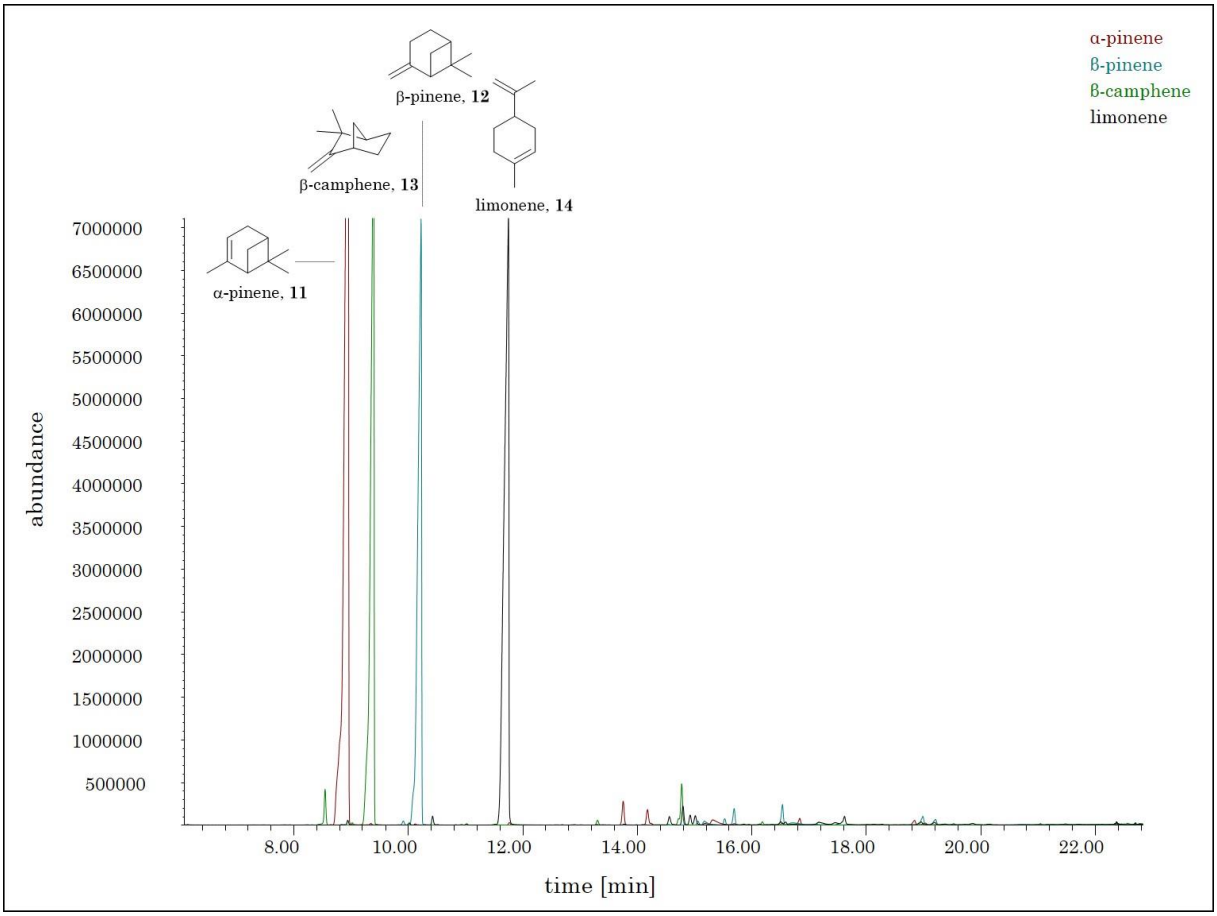


Figure 77: Full length overlaid chromatograms of reference monoterpenes α-pinene (8.7 min), β-camphene (9.2 min), β-pinene (10.0 min) and limonene (11.5 min), using temperature program GC-P4.

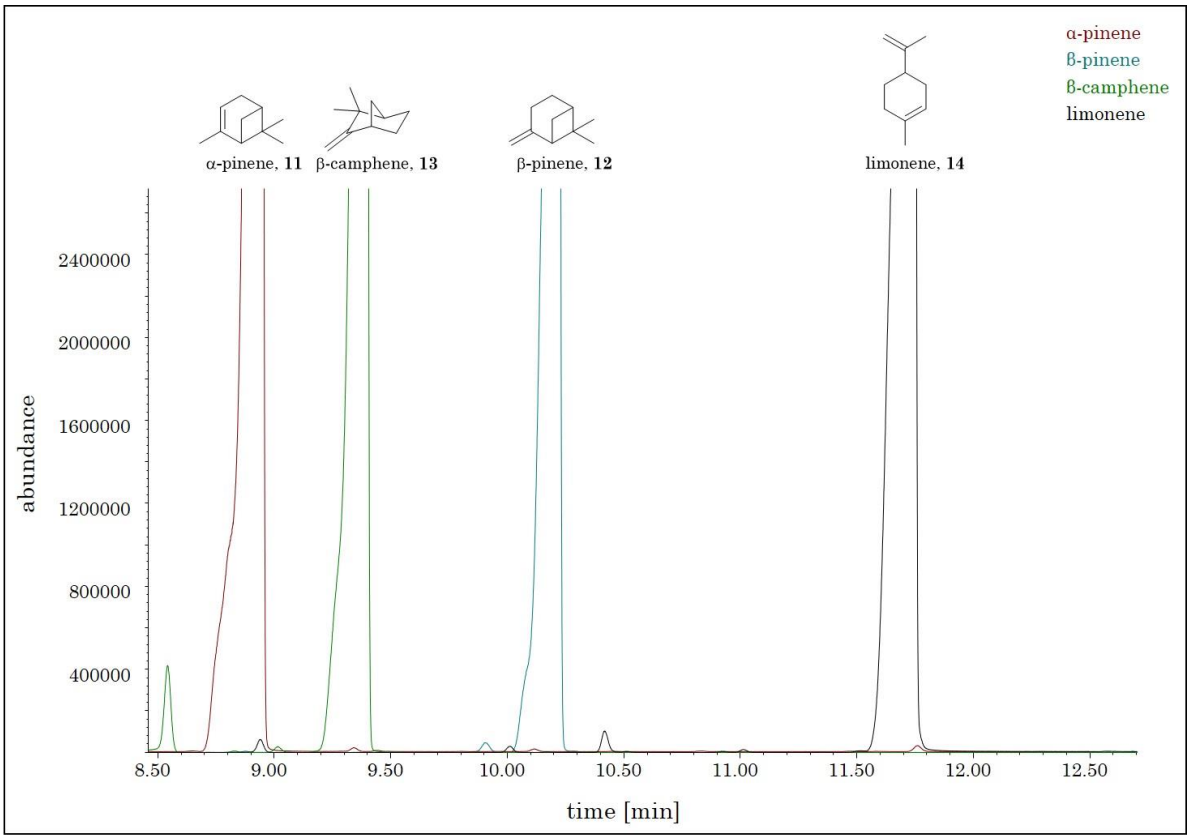


Figure 78: Section of overlaid chromatograms of reference monoterpenes α-pinene (8.7 min), β-camphene (9.2 min), β-pinene (10.0 min) and limonene (11.5 min), using temperature program GC-P4.

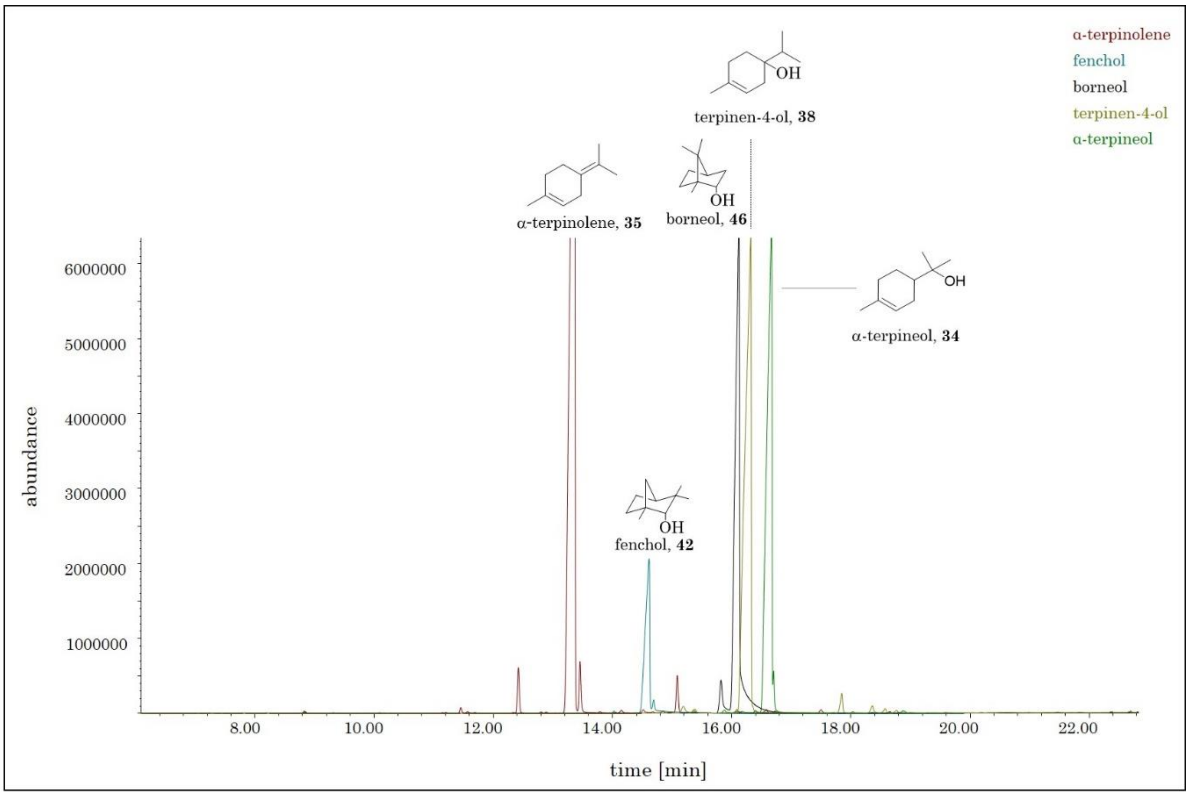


Figure 79: Full length of overlaid chromatograms of reference monoterpenes α-terpinolene (13.2 min), fenchol (14.5 min), borneol (16.0 min), terpinen-4-ol (16.1 min) and α-terpineol (16.5 min), using temperature program GC-P4.

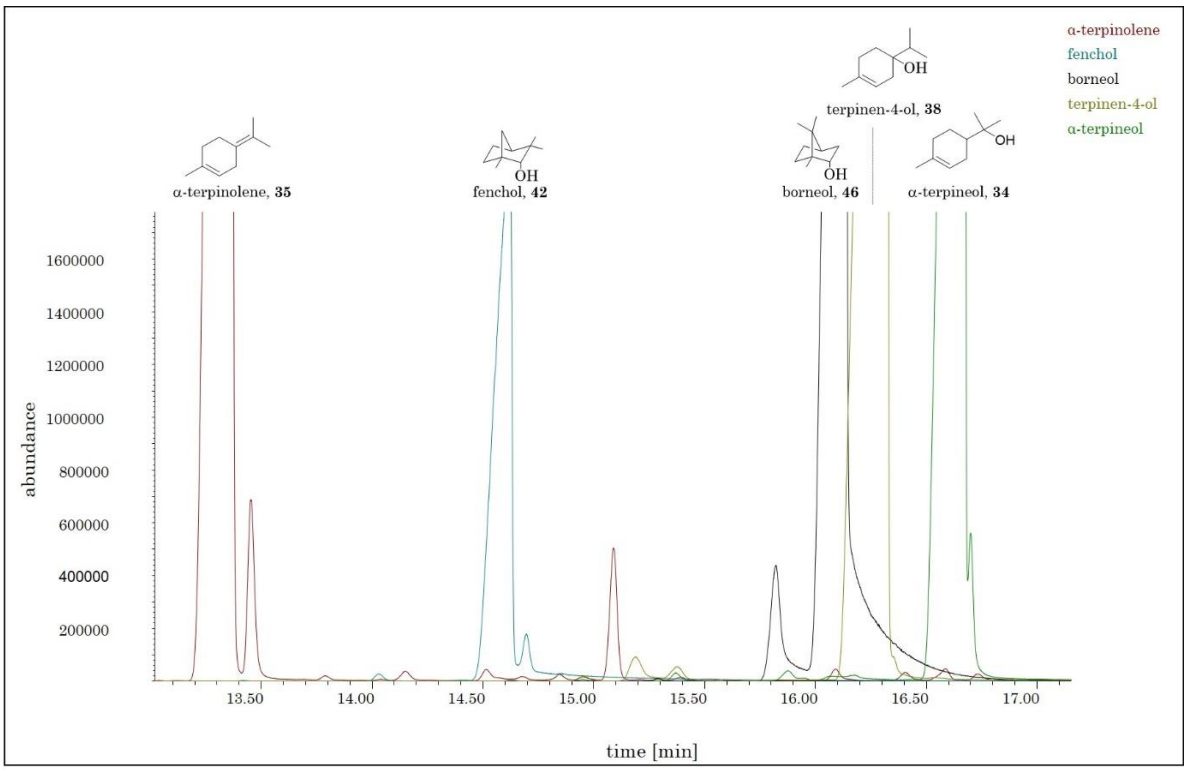


Figure 80: Section of overlaid chromatograms of reference monoterpenes α-terpinolene (13.2 min), fenchol (14.5 min), borneol (16.0 min), terpinen-4-ol (16.1 min) and α-terpineol (16.5 min), using temperature program GC-P4.

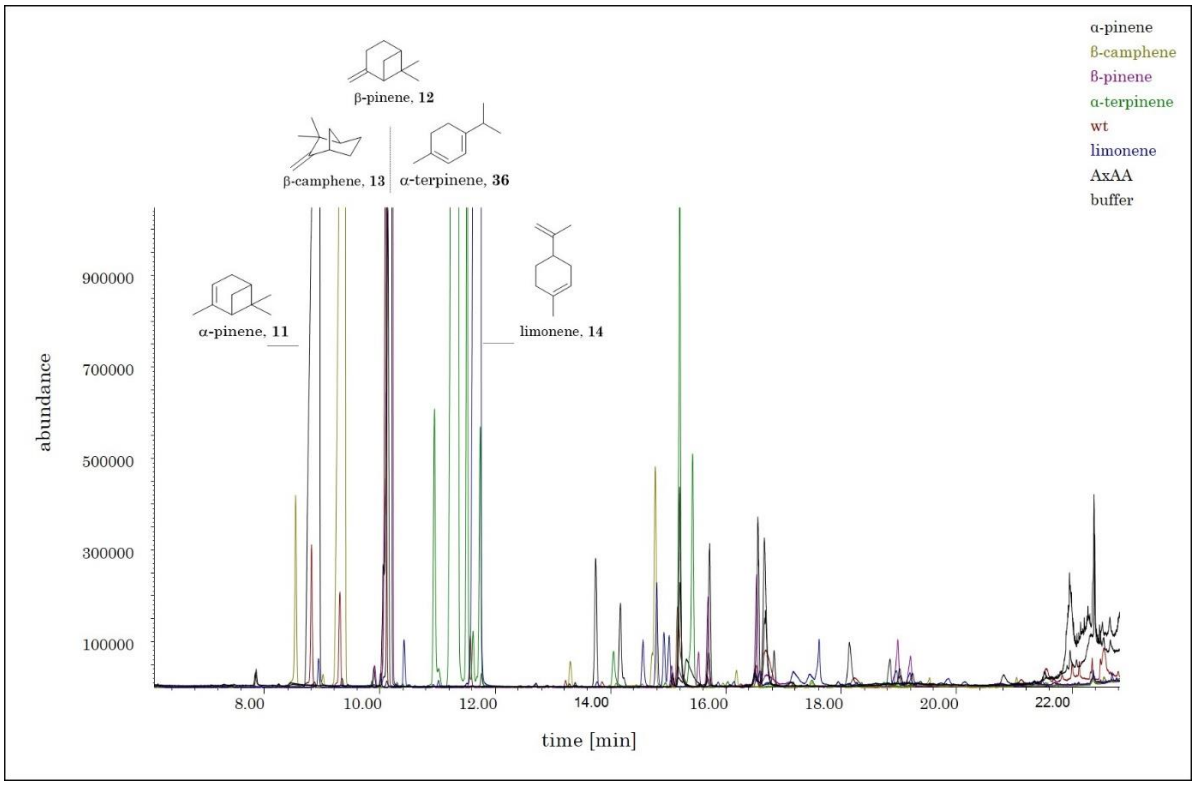


Figure 81: Full length overlaid chromatograms of reference monoterpenes α-pinene (8.7 min), β-camphene (9.2 min), β-pinene (10.0 min), α-terpinene (11.2 min), limonene (11.5 min), wildtype, variant AxAA and buffer, using temperature program GC-P4.

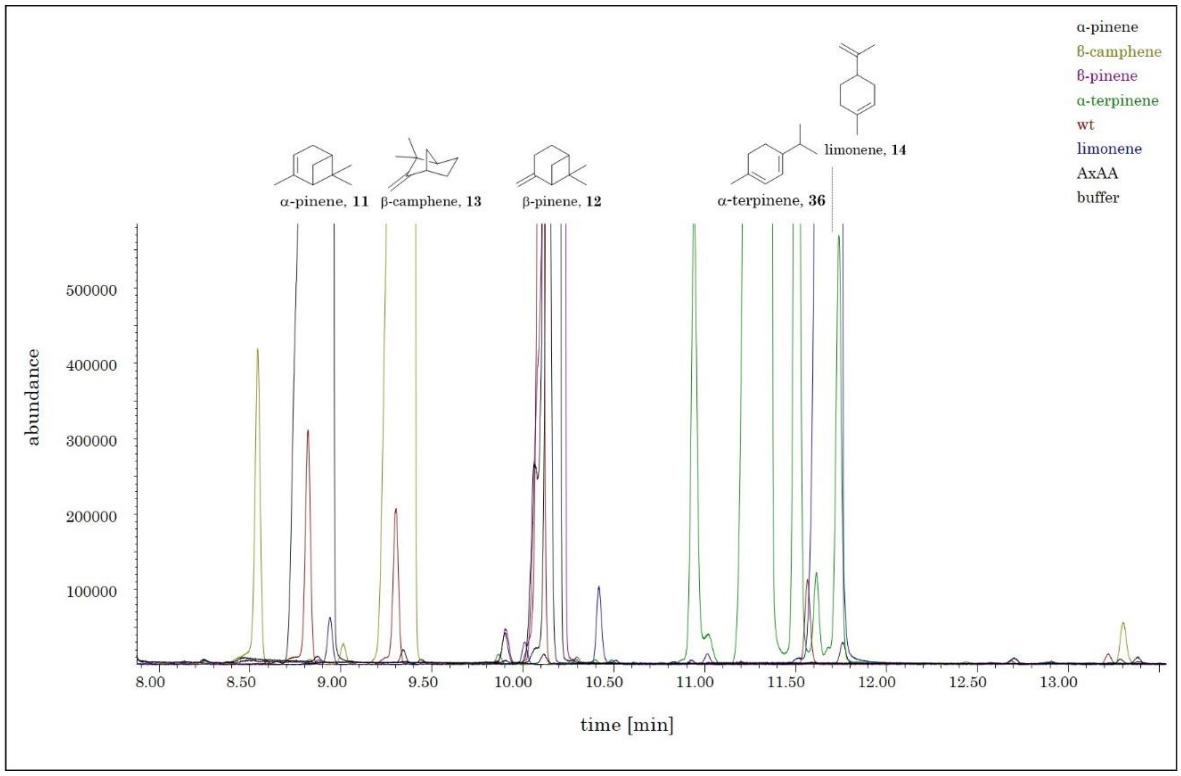


Figure 82: Section of overlaid chromatograms of reference monoterpenes α-pinene (8.7 min), β-camphene (9.2 min), β-pinene (10.0 min), α-terpinene (11.2 min), limonene (11.5 min), wildtype, variant AxAA and buffer, using temperature program GC-P4.

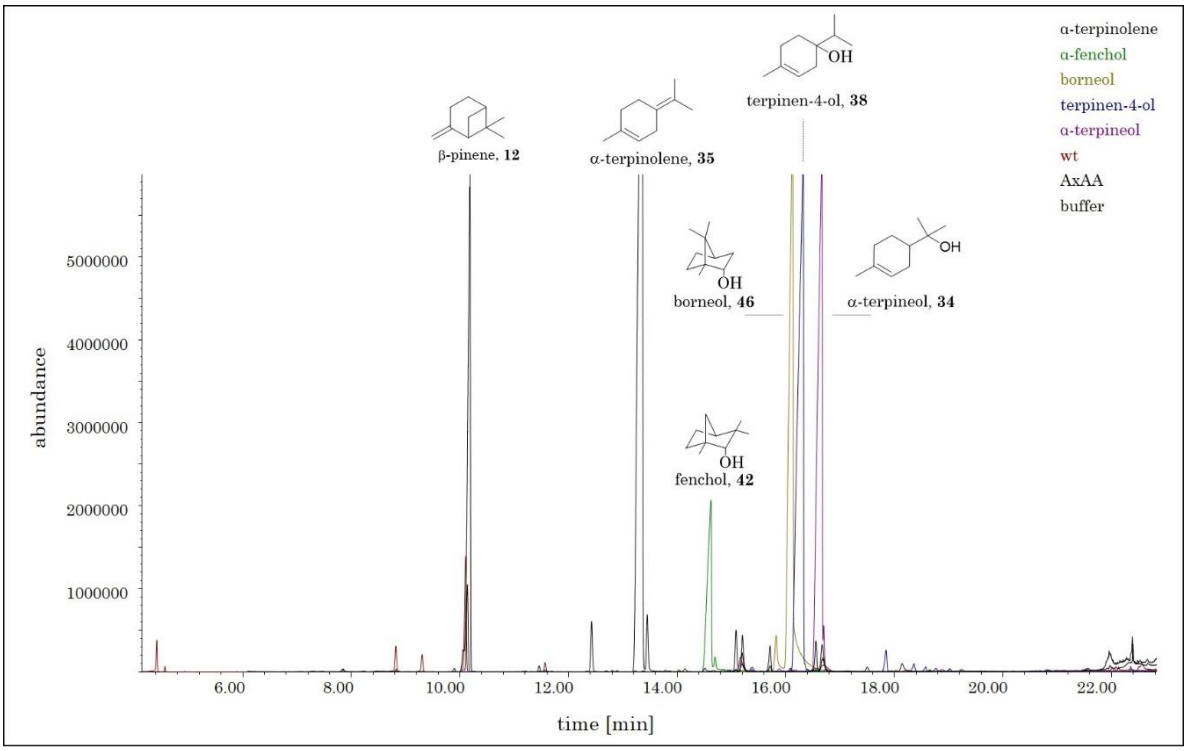


Figure 83: Full length of overlaid chromatograms of reference monoterpenes α-terpinolene (13.2 min), fenchol (14.5 min), borneol (16.0 min), terpinen-4-ol (16.1 min), α-terpineol (16.5 min), wildtype, variant AxAA and buffer, using temperature program GC-P4.

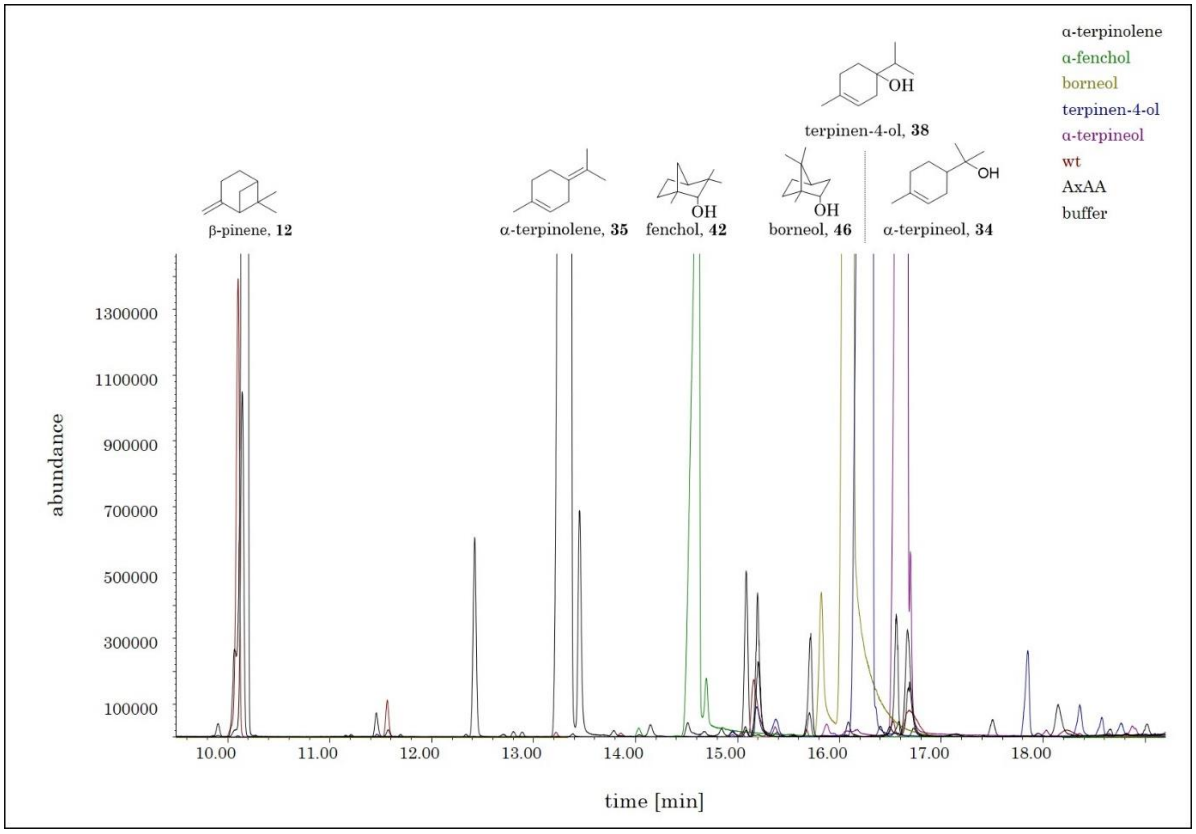


Figure 84: Section of overlaid chromatograms of reference monoterpenes α-terpinolene (13.2 min), fenchol (14.5 min), borneol (16.0 min), terpinen-4-ol (16.1 min), α-terpineol (16.5 min), wildtype, variant AxAA and buffer, using temperature program GC-P4.

8.9 Curriculum vitae

Persönliche Daten	
Name:	Svenja Diether, geb. Kaspari
Geburtsdatum/-ort:	28.10.1987 in Schwabach
Werdegang	
01/20016 – 12/2018	Promotion unter Prof. Hauer Institut für Biochemie und Technische Biochemie Abteilung Technische Biochemie Universität Stuttgart
05/2015 – 12/2015	Wissenschaftliche Mitarbeiterin unter Prof. Richert Institut für Organische Chemie Universität Stuttgart
04/2015	Diplom in Biochemie (1,4)
09/2014 – 04/2015	Diplomarbeit: “Investigation of the recognition of L-Lys(N(Me)Ac)-OH by deacetylases” unter Prof. Scharzer, Interfakultäres Institut für Biochemie, Eberhard-Karls Universität Tübingen
10/2011 - 02/2012	Auslandspraktikum: “Distribution of hnRNP A1 in premature senescent human diploid fibroblast cells after inhibition of p38 MAPK, MNK or ERK ” unter Prof. Hubbard, City College of New York, City University of New York, NY, USA
10/2008	Studiumsbeginn der Biochemie an der Eberhard-Karls Universität Tübingen
06/2008	Abitur (1,6) am Adam-Kraft Gymnasium in Schwabach