Investigating the Biosynthesis of Bioactive Compounds in Fungi

Von der Naturwissenschaftlichen Fakultät der

Gottfried Wilhelm Leibniz Universität Hannover

zur Erlangung des Grades

Doktor der Naturwissenschaften (Dr. rer. nat.)

genehmigte Dissertation

von

Vjaceslavs Hrupins, M. Sc.

2021

Referent:	Prof. Dr. Russell J. Cox
Korreferent:	PD Dr. Carsten Zeilinger
Tag der Promotion:	02.07.2021

Abstract

Fungal Type I polyketide synthases (PKS) are large multifunctional enzymes with a complex structure consisting of multiple functional domains. They produce complex bioactive products such as squalestatin S1 **27** and byssochlamic acid **106**. The main focus of the projects lies on highly reducing iterative polyketide synthases, especially on their structure and function. Previous work has shown that generating structural information from cross-linked polyketide domains can be the first step towards the elucidation of function and programming.

In this thesis we have cloned and heterologously expressed the squalestatin tetraketide synthase (SQTKS) KS/AT didomain, standalone DH domain and the ACP. The SQTKS KS/AT didomain and the ACP were heterologously expressed in *E. coli* but were obtained as insoluble (KS/AT) or inactive (ACP) proteins. The ACP and the standalone DH domain from the *Byssochlamys fulva* nonadride PKS (bfPKS) were expressed for the first time. Both proteins were successfully expressed in *E. coli* and showed catalytic activity as measured by mass spectrometry. The bfPKS ACP was used for crosslinking attempts. Biosynthetic enzymes were expressed and used for the enzymatic conversion of chemically synthesized pantetheine linker analogues into CoA linker analogues. A phosphopantetheine transferase (PPTase, also expressed) was used to load the linkers onto the bfPKS ACP. Cross-linking between bfPKS ACP and DH domains of SQTKS, bfPKS and the *Strobilurin tenacellus* strobilurin PKS (stPKS) were tried but no linking was observed.

In vitro investigations of the *B. fulva* citrate synthase (bfCS) revealed its substrate preference. The bfCS showed conversion of CoA bound substrate but no conversion for the bfPKS ACP bound one. The SQTKS standalone methyltransferase (*C*-MeT) domain did not convert the natural substrate which was bound to the bfPKS ACP and was inactive as in previous investigations.

Keywords: Polyketides, Squalestatin S1, Acyl Carrier Protein, Byssochlamic Acid

Zusammenfassung

Typ I Polyketid Synthasen (PKS) aus Pilzzellen sind große multifunktionale Enzyme, komplex in der Struktur und bestehen aus mehreren funktionalen Domänen. Sie produzieren komplexe, bioaktive Verbindungen wie Squalestatin S1 **27** und Byssochlaminsäure **106**. Das Hauptaugenmerk der Projekte liegt auf der Struktur- und Funktionsaufklärung der reduzierenden, iterativ agierenden Polyketid Synthasen. Vorherige Arbeiten haben gezeigt, dass die Strukturaufklärung von miteinander verbundenen Polyketid Domänen der erste Schritt zum Verständnis der Funktion und Programmierung dieser Proteine sein kann.

In dieser Arbeit haben wir die Squalestatin Tetraketid Synthase (SQTKS) KS/AT Didomäne, die alleinstehende DH Domäne und das ACP kloniert und heterolog exprimiert. Die SQTKS KS/AT Didomäne und das ACP wurden in *E. coli* heterolog exprimiert, resultierend in unlöslichem Protein (KS/AT) oder inaktivem (ACP) Protein. Das ACP und die alleinstehende DH Domäne von der *Byssochlamys fulva* Nonadrid PKS (bfPKS) wurden zum ersten Mal exprimiert. Beide Proteine konnten in *E. coli* erfolgreich exprimiert und waren katalytisch aktiv, was durch massenspektrometrische Messungen gezeigt werden konnte. Das bfPKS ACP wurde zur Verbindung mit anderen Domänen verwendet. Zusätzliche Enzyme wurden exprimiert, um die chemisch synthetisierten Pantethein Linker in CoA Analoga zu überführen. Eine Phosphopantethein Transferase (PPTase, ebenfalls exprimiert) wurde verwendet, um die Linker auf das bfPKS ACP zu beladen. Die Verbindung mittels Linker zwischen bfPKS ACP und den DH Domänen von SQTKS, bfPKS und der *Strobilurin tenacellus* strobilurin PKS (stPKS) wurden versucht, scheiterten jedoch.

In vitro Untersuchungen der *B. fulva* Zitrat Synthase (bfCS) gaben Aufschluss über deren Substratpräferenz. Die bfCS akzeptierte CoA gebundene Substrate, jedoch keine bfPKS ACP gebundenen. Die alleinstehende SQTKS Methyltransferase (*C*-MeT) konnte das natürliche Substrat, welches an dem bfPKS ACP gebunden war, nicht in das Produkt überführen und war, wie bereits in vorherigen Untersuchungen, inaktiv.

Schlagwörter: Polyketide, Squalestatin S1, Acyl Carrier Protein, Byssochlamsäure

Acknowledgement

First, I would like to thank Prof. Dr. Russell J. Cox for offering me this interesting and challenging project. I appreciate the good supervision, patience and the professional environment he was offering me.

Thank you to PD Dr. Carsten Zeilinger for being the second examiner, sharing his experience in protein expression and purification and for having interesting conversations.

A special thanks goes to Prof. Dr. Sascha Beutel for taking part as the third examiner.

I would like to thank the former and current Cox group members (Liz, Haili, Dongsong, Chongqing, Steffen, Lei, Yunlong, Mary, Katharina, Hao, Carsten, Karen, Lukas, Eman, Raissa, Verena, Juliane and Henrike) making the time really enjoyable. Moreover, I want to thank especially Oliver Piech for many funny moments inside and outside of the lab. Sen Yin and Jin Feng gave me during my PhD time a small excursus into the chinese culture and language, thanks for this.

The time in Frankfurt with Dr. Eric Kuhnert will be kept in good memory, thank you Eric.

A special thanks goes to the media kitchen and the store, especially Katja Körner. A fast progress would not be possible without this team.

Finally, I want to thank my fiancee, Nina Stehle, and my family for giving me the support during challenging times.

Abbreviations

2TY	2x tryptone yeast medium
ACP	acyl carrier protein
AT	acyl transferase
ATP	adenosine triphosphate
C-MeT	*C*-methyltransferase
CoA	Coenzyme A
Da	Dalton
DEBS	6-deoxyerythronolide B synthase
DH	dehydratase
DNA	desoxyribonucleic acid
EDTA	ethylenediaminetetraacetic acid
EIC	extracted ion chromatogram
ER	enoyl reductase
ESI	electron spray ionisation
ELSD	evaporative light scattering detector
FAS	fatty acid synthase
FPLC	fast protein liquid chromatography
GST	glutathione-*S*-transferase
HPLC	high performance liquid chromatography
hr	highly reducing
IPTG	isopropyl-β-D-1-thiogalactopyranoside
KS	ketosynthase
kb	kilo base pairs
kDa	kilo Daltons
KR	ketoreductase
KS	ketosynthase
LC	liquid chromatography
MALDI	matrix-assisted laser desorption ionization
m/z	mass to charge ratio
MS	mass spectrometry
NADPH	nicotinamide adenine dinucleotide phosphate
Ni-NTA	nickel-charged affinity resin

NMR	nuclear magnetic resonance
nr	non-reducing
NSAS	norsolorinic acid synthase
OD_{600}	optical density at 600 nm
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PDB	protein data bank
PKS	polyketide synthase
pr	partially-reducing
PT	product template
Q-TOF	quadrupole time-of-flight
SAM	*S*-adenosylmethionine
SAT	starter unit:ACP transacylase
SAXS	small angle light scattering
SDS	sodium dodecyl sulphate
SEC	size exclusion chromatography
Sfp	4'-phosphopantetheinyl trnasferase
SNAC	S-*N*-acetylcysteamine
SOC	super optimal broth with catabolite repression
SQTKS	squalestatin tetraketide synthase
SUMO	small ubiquitin-like modifier
TAE	tris-acetate-EDTA buffer
TE	thioesterase
TEMED	*N,N,N,N*-tetramethylethylenediamine
TIC	total ion current
Tris	tris(hydroxymethyl)aminomethane
UV	ultraviolet

Table of Contents

1. Introduction

1.1 Natural Products

Broadly, natural products can be defined as a set of small molecules derived from living organisms.[1] Metabolites which are required for the growth and maintenance of cellular function are called primary metabolites, while secondary metabolites increase the survival fitness of a living organism, but are not essential.[2] For instance, plants use secondary metabolites for defence purposes against herbivores, pests, pathogens and also as signal compounds to attract pollen and seed dispersing animals.[3,4] The bioactivity of natural products in the human body has inspired the pharmaceutical development.[5] For example, taxol **1**, also called paclitaxel, is one of the most prominent anticancer drugs and was isolated from the western yew, *taxis brevifolia*.[6] The effectiveness of the diterpenoid is based on a chromosome missegregation on multipolar spindles, which causes cell death to the tumor cell.[7] A prominent group of natural products with strong bioactivity are the alkaloids. Cocaine **2**, which works as a central nerve system stimulant, was isolated from dried leaves of *Erythroxylon coca* and is a prominent example for drug abuse (Figure 1.1).[8,9]

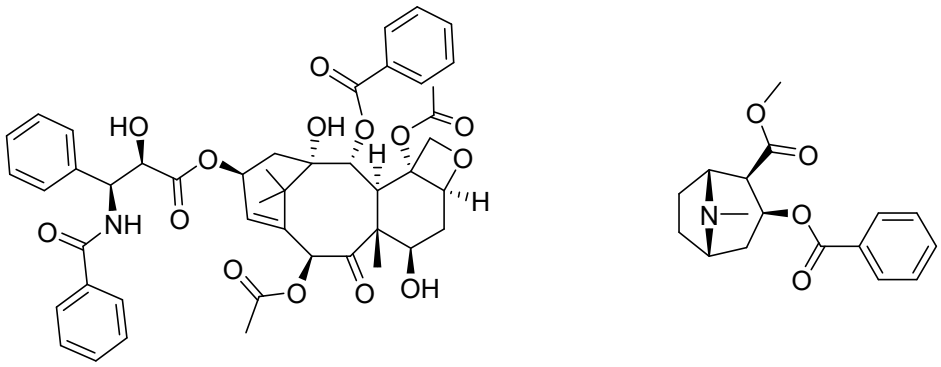
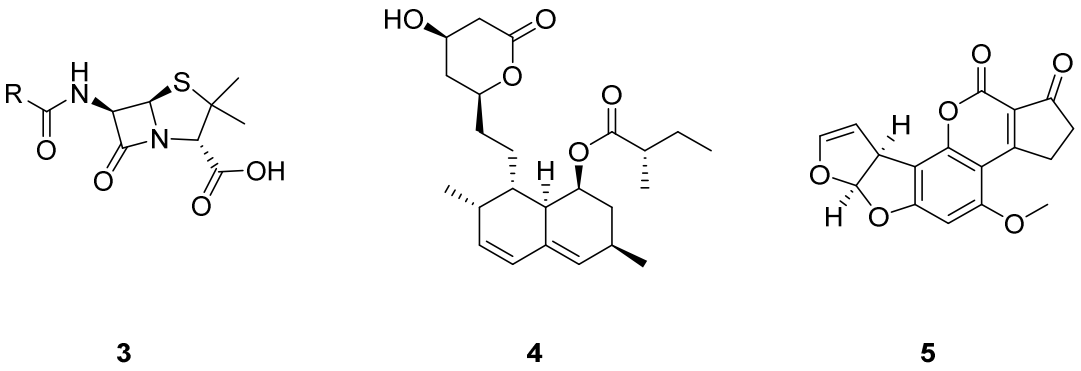


Figure 1.1: Structures of taxol and cocaine.

1.2 Natural Products from Fungi

Fungi produce a high diversity of natural products.[10] Many of them show a bioactivity which can be used for the treatment of diseases. Penicillins **3**, for instance, are β-lactam antibiotics that are the most widely used antimicrobials in the world (Figure 1.2).[11] These compounds are characterized by a thiazolidine ring, an attached β-lactam ring and a side chain. Penicillins inhibit cell wall synthesis of bacteria by binding to the enzymes that produce the protein crosslinks.[12] Another important class of bioactive compounds from fungi are the cholesterol-lowering agents, the statins.[13] The most popular representative is lovastatin **4**. The compound acts through selective inhibition of HMG-CoA reductase, which catalyses an early step in cholesterol biosynthesis.[14] A further class of natural products derived from fungi are the mycotoxins.[15,16] Aflatoxin B1 **5** is one of the most dangerous mycotoxins that is produced by certain fungi of the *Aspergillus* species and can be found in soil, hay and grains. The toxin leads to the development of hepatocellular carcinoma in humans and animals.[17]



3 **4** **5**

Figure 1.2: Structures of a penicillin, lovastatin and aflatoxin B1.

1.2.1 Main Classes of Fungal Secondary Metabolites

Secondary metabolites derived from fungi can be classified in four distinct groups based on the starting material and the general biosynthetic pathway. The starting material can be for example dimethylallyl diphosphate **6**, resulting in an terpenoid (*e.g.* ergosterol **10**);[18] an amino acid (**7**, **8**) leading to peptides (*e.g.* cyclosporine A **11**)[19] and alkaloids (*e.g.* communesin F **12**);[20] or acetyl-CoA **9** from which fatty acids (*e.g.* oleic acid **14**)[21] and polyketides (*e.g.* tetrahydroxynaphtalene **13**)[22] are produced (Figure 1.3). Combinations of different biosynthetic pathways lead to even more complex secondary metabolites.

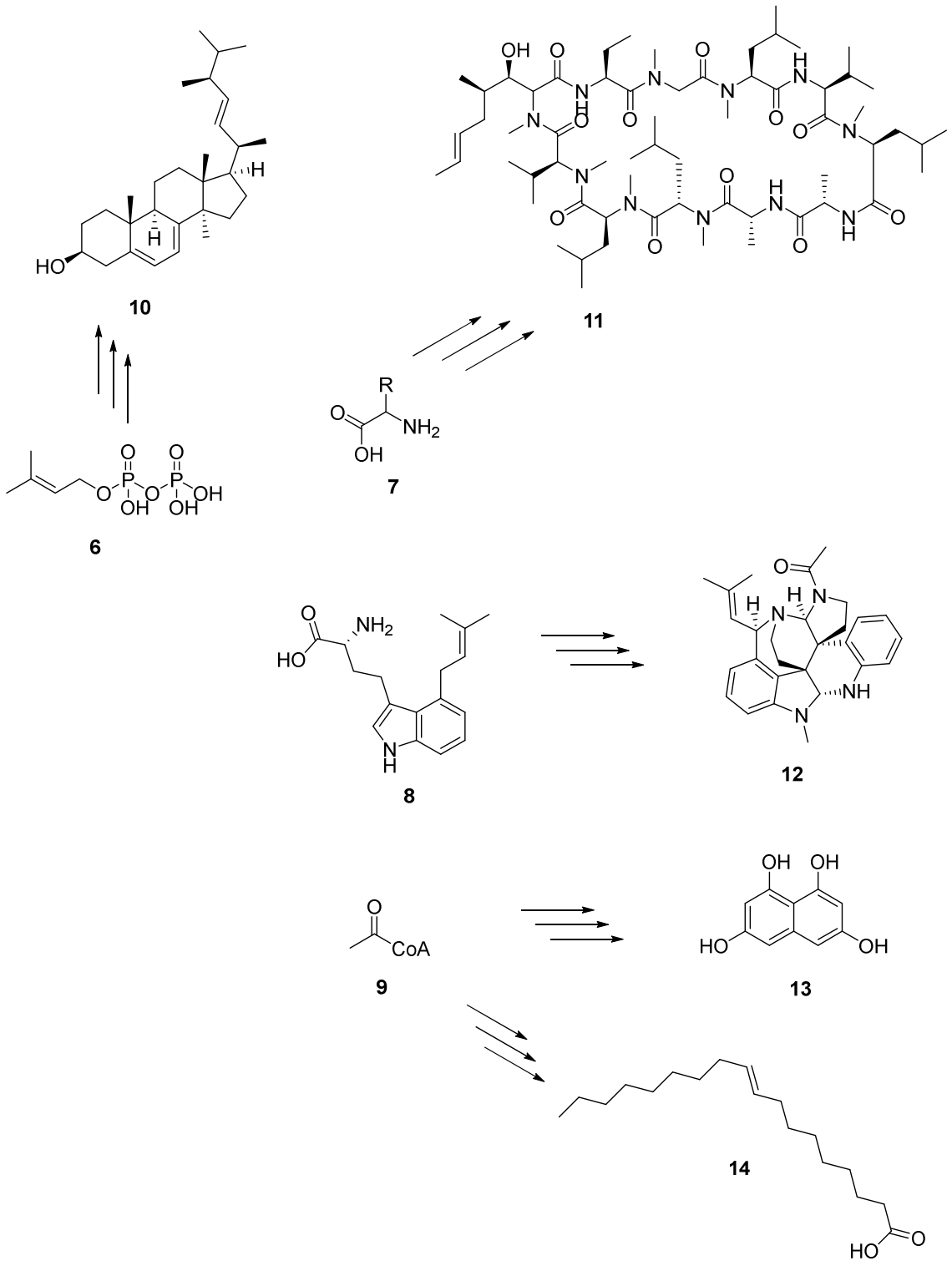


Figure 1.3: Fungal secondary metabolites from different biosynthetic pathways.

1.3 Fatty Acid Synthases

The enzymatic reactions for fatty acid biosynthesis are very similar to those of polyketide biosynthesis. Both systems utilize common precursors and similar chemistry.[23] Even today the fatty acid biosynthesis guides the polyketide research. It is helpful to introduce the functionality of the domains on a simpler system like the FAS and proceed with a more complex system - the PKS.[24]

1.3.1 Architectures of Fatty Acid Synthases

Fatty acid synthases (FAS) are divided in two types and have at least three distinct architectures (known to date). The Type I FASs are large multifunctional proteins which are found in animals and fungi.[24] They consist of covalently linked functional catalytic domains. Gram-positive, mycolic acid producing bacteria, also contain Type I FASs.[25] Type II FAS systems consist of single proteins which form non-covalent complexes. Type II FAS are found in plants and bacteria.[25,26]

There are two main architectures known for Type I FAS. The crystal structure of the yeast FAS illuminates a large macromolecular assembly which functions as a six chambered reactor for fatty acid synthesis.[27] The synthase consists of two separate multifunctional polypeptides, encoded by two genes, and is assembled to an α_6-β_6 heterododecamer of 2.6 MDa, composed of 210 kDa α- and 230 kDa β-chains.[28,29] The FAS has a barrel-shaped structure formed by a central wheel, in which six α-chains are capped by two domes of three β-chains each (Figure 1.4 **B**).

The second type of FAS is found in vertebrates such as mammals (mFAS, Figure 1.4 **A**). mFAS is a large multienzyme complex which catalyses all reactions of fatty acid biosynthesis and is encoded by a single gene.[28,30] The resolved crystal structure reveals all domains except the acyl carrier protein (ACP) and the thiolesterase (TE). The homodimeric structure of the mFAS is important for the activity of the protein. The enzyme is divided into two fragments that are connected by a short linker region leading to an X-shaped enzyme. The lower part consists of the domains for chain elongation and the upper part is built of the chain modification domains. The ACP domain is not observed in this crystal structure but it has to have a position where every domain could be reached by the carrier.

The Type II bacterial FAS is encoded by the *fab*-genes. Crystal structures are solved for the single enzymes FabH[31] and FabG[32] and in a ACP bound state for the enzymes FabD,[33] FabB,[34] FabF,[35] FabA,[36] FabZ (some of them are described in sections 1.6 and 7.1).[37,38]

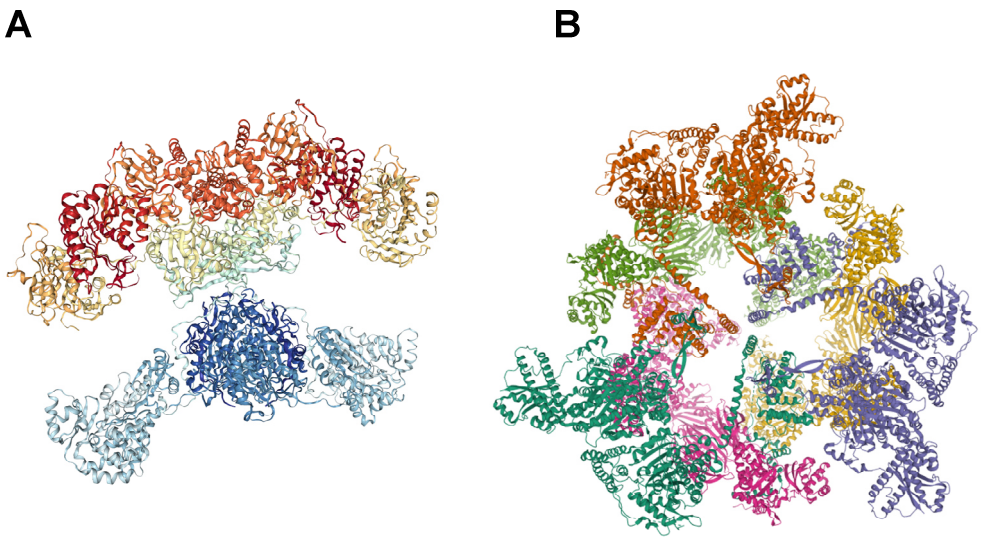


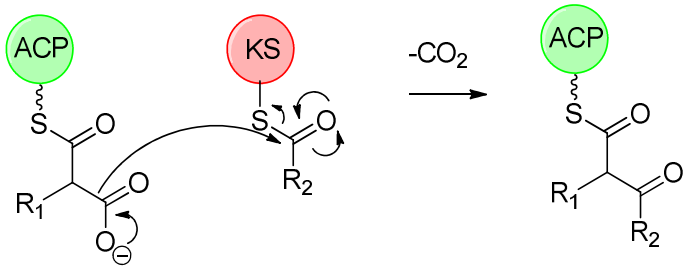
Figure 1.4: Crystal structures: **A**, structure of vertebrate FAS (PDB: 2VZ8); **B**, Structure of yeast FAS (PDB: 3HMJ).

1.3.2 Function of the Fatty Acid Synthase

Regardless of its quarternary structure, FAS from all organisms use the same set of enzymatic reactions. The reactions are directed by the acetyltransferase (AT), β-ketoacylsynthase (KS), ketoreductase (KR), dehydratase (DH) and the enoylreductase (ER). The acyl carrier protein (ACP) has an important role in transferring intermediates from one catalyst to another, and release is often catalysed by a thiolesterase (TE) or other hydrolytic enzyme.

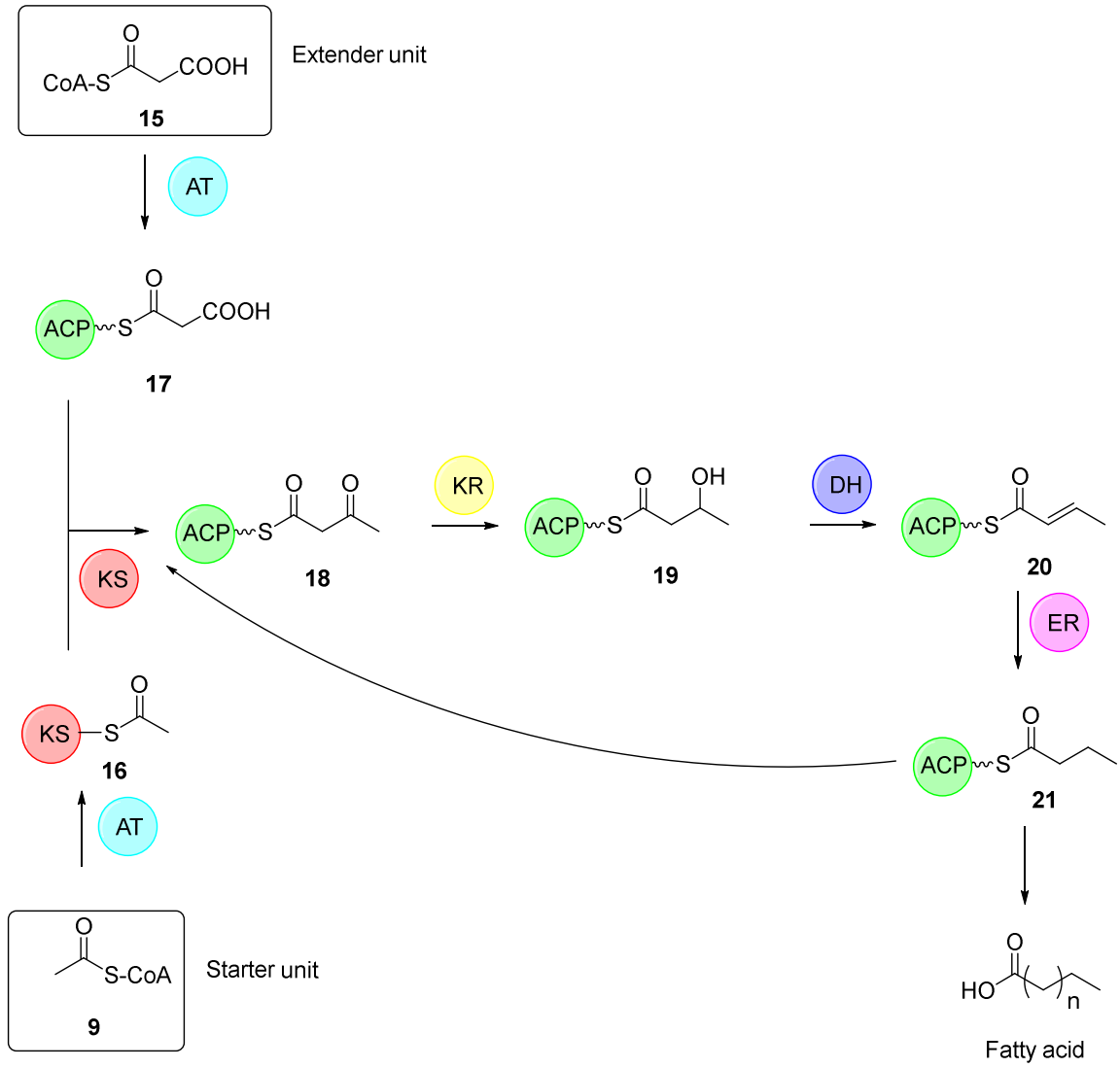
Acetyl CoA usually serves as the *starter unit*. Malonyl-CoA, required as the *extender unit*, is biosynthesized by the biotin containing enzyme acetyl-CoA carboxylase.[39] Both acetyl and malonyl CoAs are available from primary metabolism. The first domain working in the fatty acid biosynthesis is the AT. The AT domain loads acetyl-CoA **9** and malonyl-CoA **15** moieties in a competitive reaction.[40] AT enzymes have a conserved active site serine which is involved in transferring acyl groups as intermediate esters. Only if the appropriate substrate is loaded, the reaction can take place. The starter unit **9** is transferred from the AT to a conserved thiol in the active site of the

KS **16** which holds it as a thiolester. The extender unit, malonyl-CoA, is first loaded onto AT and then transferred to the ACP **17**. ACP possesses a phosphopantetheine (PP) prosthetic group attached to a conserved serine of the ACP. The terminal thiol of the PP group holds the malonyl unit as a thiolester. The chain elongation step is catalysed by a decarboxylative Claisen condensation. Decarboxylation of malonate creates a nucleophile which adds to the KS-bound starter unit, and results in transfer of the KS-bound group to the ACP (Scheme 1.1). During the first cycle of FAS this results in the creation of acetoacetyl ACP **18**.



Scheme 1.1: Mechanism of the Claisen condensation catalyzed reaction by a KS domain.

In the next step the reduction of the β-carbonyl **18** is performed by the ketoreductase (KR) to create a β-hydroxyl substituent **19**. The KR uses NADPH as the hydride donor. The resulting alcohol is then dehydrated by the dehydratase (DH) domain giving an α,β-unsaturated *E*-alkene **20**. Finally, the enoyl reductase (ER) domain reduces the alkene, creating a saturated product **21** which can be either released or transferred back to the KS and extended again (Scheme 1.2).[41–44] These reactions are described as the β-processing cycle.



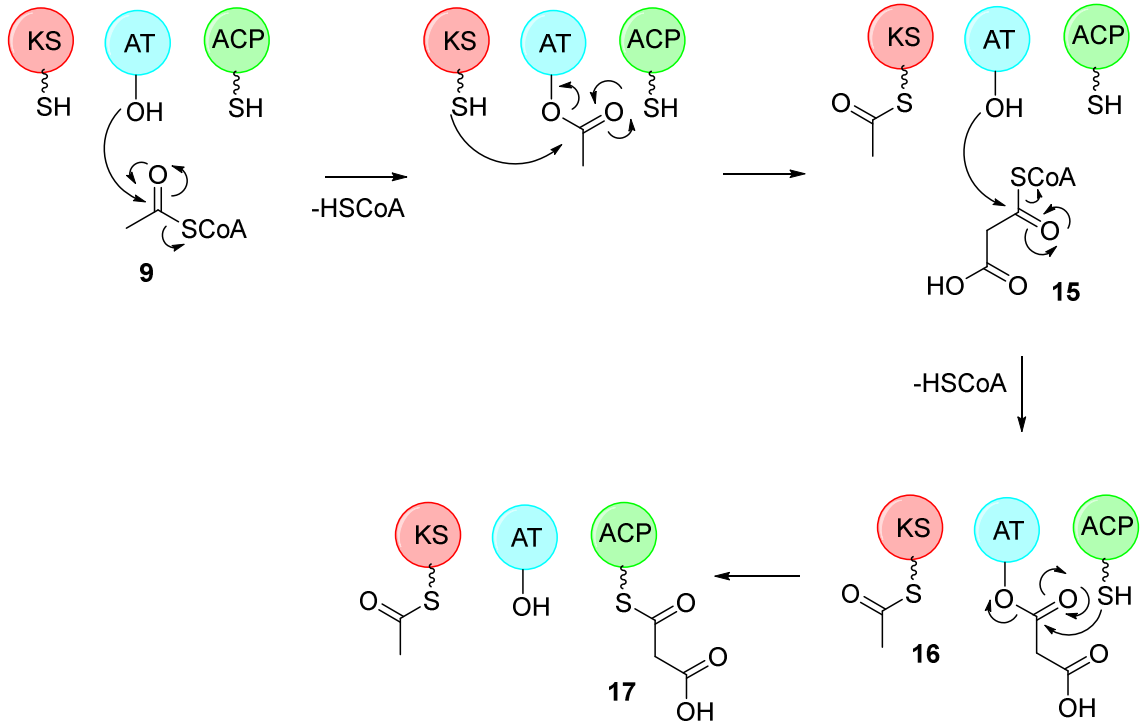
Scheme 1.2: Schematic overview of the fatty acid biosynthesis steps and the functions of each domain of the FAS.

1.4 Polyketides and their Biosynthesis

Polyketides are an important class of molecules produced by bacteria, fungi and plants. Many of them have a complex chemical structure and play an important role as pharmaceuticals. The natural compounds have a variety of biological activities like antitumor-, antiparasitic-, antibiotic-, antifungal- and anticholesterolemic-properties. Some of these compounds are in clinical use and have an outstanding importance as lead-structures for the design of new drugs.[45] The large diversity of these kinds of secondary metabolites is accomplished by directed chemical processes executed by multidomain proteins called polyketide synthases (PKS). The PKS are chemically highly similar to

FAS. Both systems utilize an acyl carrier protein to transport the acyl chain and have a very similar β-processing with the same basic chemistry.[46]

PKS proteins have at least the minimal PKS, consisting of the ketosynthase (KS), acyltransferase (AT) and the acyl carrier protein (ACP), which are essential. Polyketide synthesis starts with the loading procedure of the starter and extender unit (Scheme 1.3).

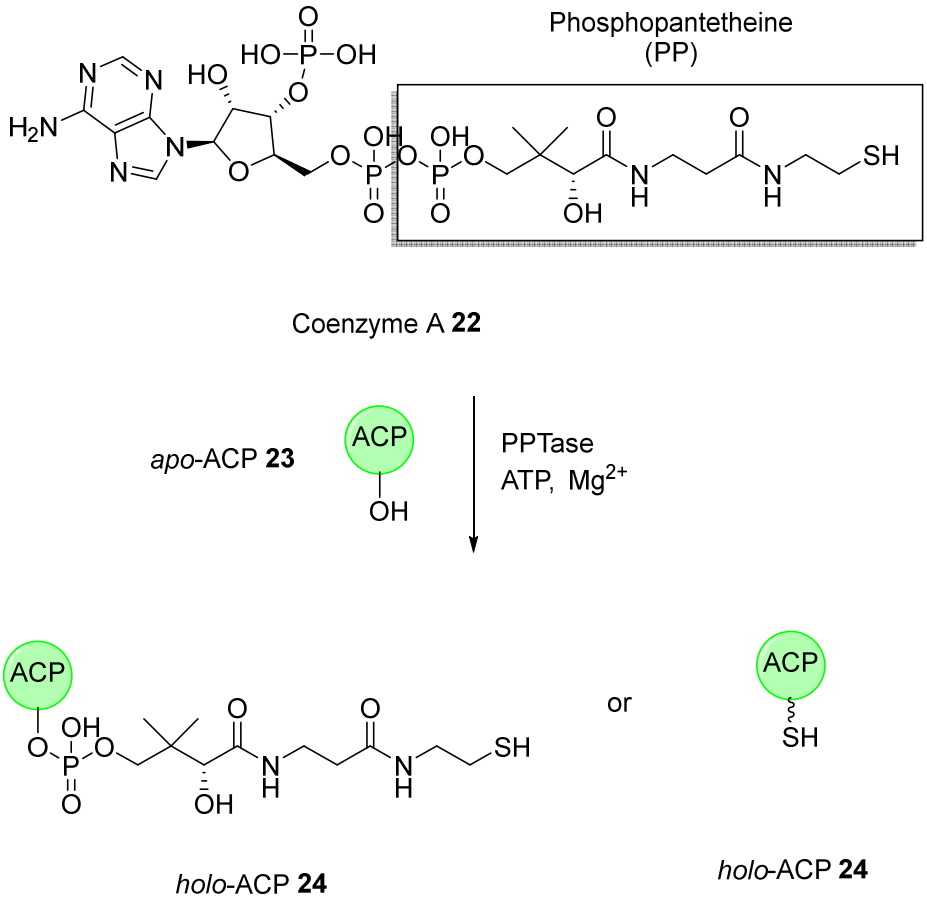


Scheme 1.3: Loading and transfer reactions of the starter and extender units.

The most common starter and extender units are acetyl-CoA **9** and malonyl-CoA **15**. However, in contrast to FAS these can vary. The loading of the starter unit acyl-CoA is performed by the AT and is subsequently transferred to the KS **16**. The extender malonyl unit is also selected by the AT and transferred to the ACP **17** which carries the growing polyketide chain. The carbon-carbon bond forming reaction works in the same way as described for FAS. Each KS catalysed reaction lengthens the growing polyketide chain with one more ketide, which is named, depending on the length of the polyketide chain, as diketide, triketide, tetraketide *etc*.

For the functionality of the ACP a posttranslational modification is necessary. The prosthetic group phosphopantetheine (PP) is transferred from the Coenzyme A (CoA) **22** to a highly conserved serine of the ACPs active site in an ATP dependent reaction (Scheme 1.4). An ACP domain without the PP is called *apo*-ACP **23** and with an attached PP it is known as *holo*-ACP **24**. The PP group works like an arm which reaches inside the

active site pockets of the corresponding domains and offers substrates which are bound to the thiol as a thiolester. The transfer of the PP moiety is catalysed by a phosphopantetheine transferase (PPTase).

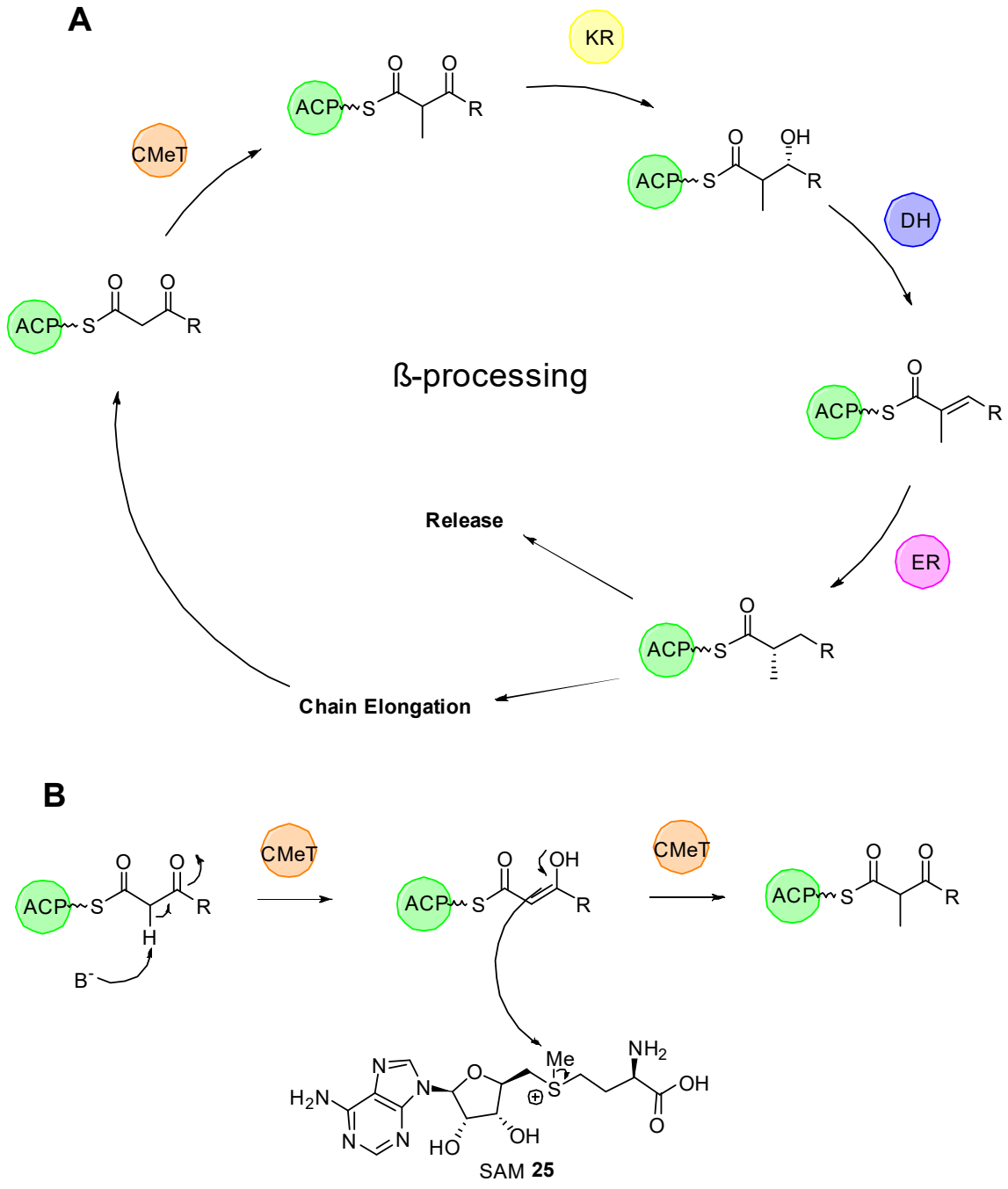


Scheme 1.4: Activation of *apo*-ACP by the transfer of phosphopantetheine, catalyzed by a PPTase.

The most significant difference between FAS proteins and PKS proteins are the β-modification reactions. While in FAS all these domains (KR, DH, ER) are active in each extension cycle leading to a fully saturated carbon chain, in PKS they may be inactive, depending on the programming, in one or several extension cycles.[24] This leads to a variability in reduction and dehydration pattern and in consequence to more complex molecules with a variety of functional groups. There are also PKS proteins existing which have no or only a limited set of β-modification domains which makes the resulting molecule more complex compared to a saturated carbon chain of a fatty acid.

Some PKS can perform a methylation reaction during the β-processing cycle. The methyltransferase (*C*-MeT) catalyses the transfer of a methyl group to the polyketide. The methyl group comes from the cofactor *S*-adenosyl methionine **25** (SAM) and is attached to the α-carbon of the growing polyketide chain (Scheme 1.5 **B**). The following β-

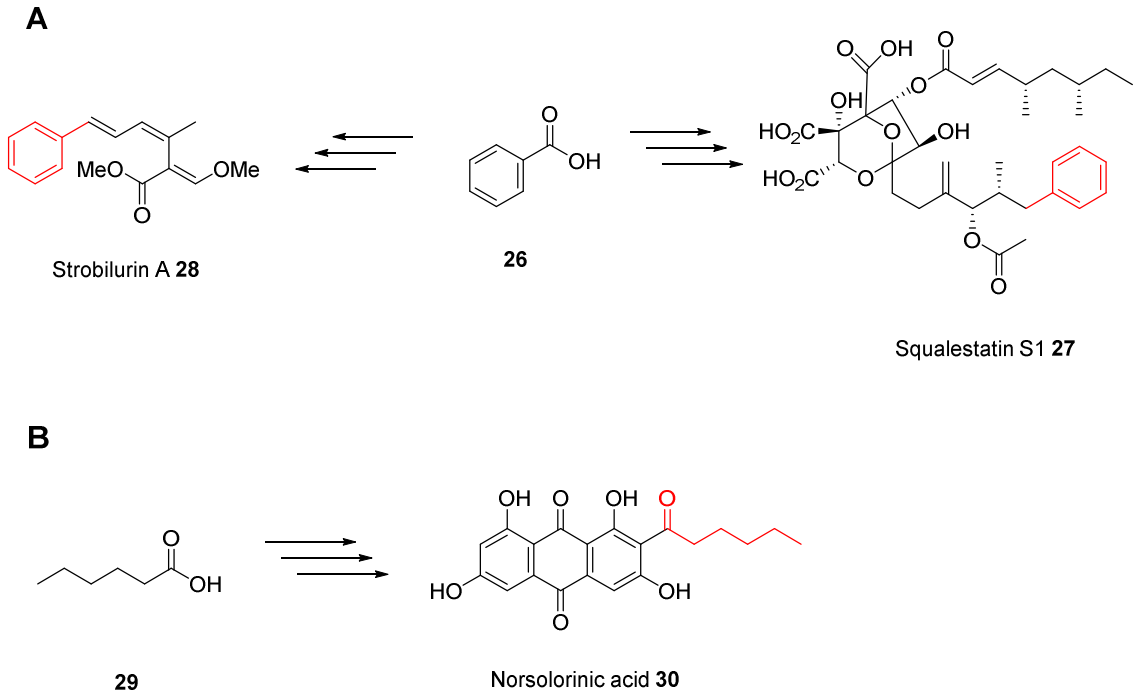
modification domains (KR, DH, and ER) catalyse the same reactions as already described for the fatty acid biosynthesis (Scheme 1.5 **A**). After the biosynthesis is complete the polyketide is usually released by hydrolysis, reduction or lactonisation.[47] Additionally to the regular β-modification domains, some PKS contain domains such as cyclases and aromatases which define the folding pattern of the polyketide.



Scheme 1.5: β-processing during polyketide synthesis: **A**, most common processing steps; **B**, mechanism of the methyl transfer catalysed by the methyltransferase.

1.4.1 Unusual Starter Units in Fungal Polyketides

In general, polyketides are built from simple precursors, such as acetyl-CoA and malonyl-CoA as the extender unit. The carbon-carbon bond forming is based, like in FAS, on a decarboxylative Claisen thiolester condensation.[48] The structural diversity of polyketides consists of chain length (*i.e.* number of extension cycles), variations in the oxidation state at β-positions, the stereochemistry of the resulting functional groups, the mechanism of off-loading and modifications of the carbon skeleton introduced by post PKS tailoring enzymes. In addition to the usage of simple precursors there are unusual starter units which increase structural diversity and chemical complexity within the synthesis of polyketides, expanding the variety of resulting polyketides. Benzoate **26** is one unusual starter unit which is used by the hexaketide synthase of the squalestatin S1 **27** pathway (Scheme 1.6 **A**).[49–52] The same first building block **26**, which derived by the degradation of phenylalanine *via* cinnamate, is also incorporated into strobilurin A **28**, an antifungal antibiotic.[53–55]



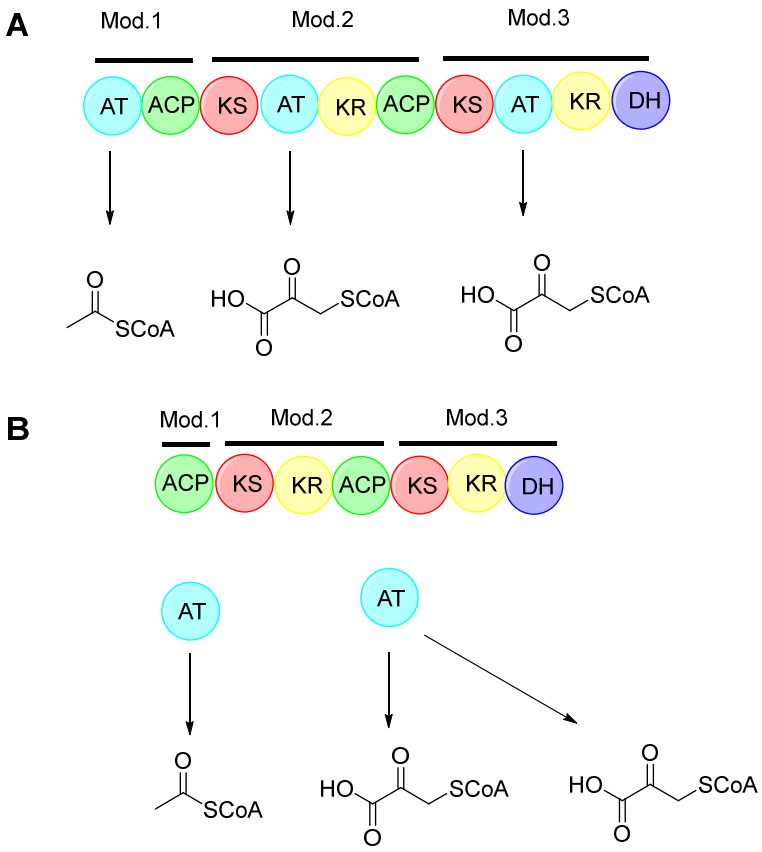
Scheme 1.6: Unusual starter units: **A**, hexanoate is incorporated into norsolorinic acid; **B**, Benzoate is incorporated into strobilurin A and squalestatin S1.

Another unusual starter unit used during fungal polyketide biosynthesis is hexanoate **29**. The groups of Townsend and Simpson showed that this starter unit is incorporated into norsolorinic acid **30**, which is a key intermediate in the aflatoxin B_1 biosynthesis (Scheme 1.6 **B**).[56–58] The hexanoate comes from a dedicated fatty acid synthase which is part of a FAS-PKS hybrid complex.[59]

1.4.2 Classification of Polyketide Synthases

Polyketide synthase proteins (PKS) are characterized by several classes and subclasses based on their architectures, similar to the FAS systems. Mainly they are divided in Type I, II and III PKS proteins. Type I PKS proteins are characterized by covalently linked catalytic domains within large multifunctional enzymes. Type I PKS are found in bacteria and fungi.[24,48] Type II PKS proteins consist of a dissociable complex of discrete, mostly monofunctional enzymes and are found in bacteria.[60] Type III PKS is a special type of PKS proteins, has no corresponding FAS system, and is a very simple type of synthases. It consists of a ketosynthase (KS) only and does not require other catalytic domains. Type III PKS use CoA-bound substrates rather than ACP-bound substrates used by Type I and II PKS.[24,61] Type III PKS proteins are found in bacteria, fungi and plants.

Type I PKS proteins can be further subdivided into multimodular proteins, which work non-iteratively and monomodular, iterative synthases (Table 1.1). Multimodular synthases are found in bacteria and can be differentiated into *cis*- and *trans*-AT PKS proteins. The difference between *cis*- and *trans*-AT PKS proteins is that *cis*-AT PKS have an AT in each and every module (Scheme 1.7 **A**) while *trans*-AT PKS lack integral ATs in each module and have separate *trans-acting* ATs (Scheme 1.7 **B**) which supply each module with simple malonyl extender units.[46] *Trans* ATs are not incorporated into large multienzyme complexes. They usually transfer malonyl-CoA to *holo*-ACP with high specificity and function as individual catalysts. Their transferase activity is reserved to a single multienzyme module. In comparison to *trans*-ATs, *cis*-ATs have an increased tolerance for different substrates and the acyl transferase activity is reserved to a single multienzyme module.



Scheme 1.7: Comparison between *cis*- and *trans*-AT PKS: **A**, hypothetical *cis*-AT PKS; **B**, hypothetical *trans*-AT PKS.

The projects in this work are based on fungal Type I systems. Fungal Type I PKS proteins work iteratively, which means they use a single set of active sites through multiple catalytic cycles, and are subdivided according to the level of reductive processing in non-reducing (nr-PKS), partially reducing (pr-PKS) and highly reducing (hr-PKS) polyketide synthases.

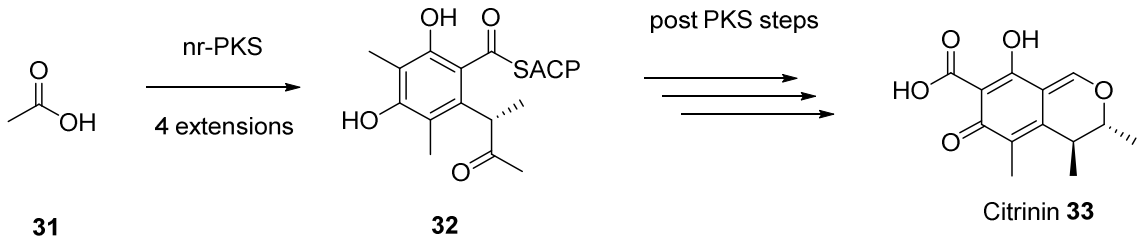
Table 1.1: Subclasses of Type I PKS.

Type I	Fungal Monomodular (iterative)	Non-Reducing (nr-PKS)
		Partially-Reducing (pr-PKS)
		Highly-Reducing (hr-PKS)
	Multimodular (non-iterative)	*cis*-AT
		trans-AT

1.4.3 Non-Reducing Iterative Polyketide Synthases

The non-reducing polyketide synthases (nr-PKS) lack all β-processing domains and use the primary polyketide chain for aromatic cyclization. Non-reducing PKS proteins usually contain, in addition to the minimal PKS requirements of AT, ACP and KS, a starter unit ACP transacylase (SAT) and product template (PT) domains. They may also contain a methyltransferase (*C*-MeT) domain as well as domains for the product release, for instance a thiolester reductase (R) or a thiolesterase (TE).[24,62]

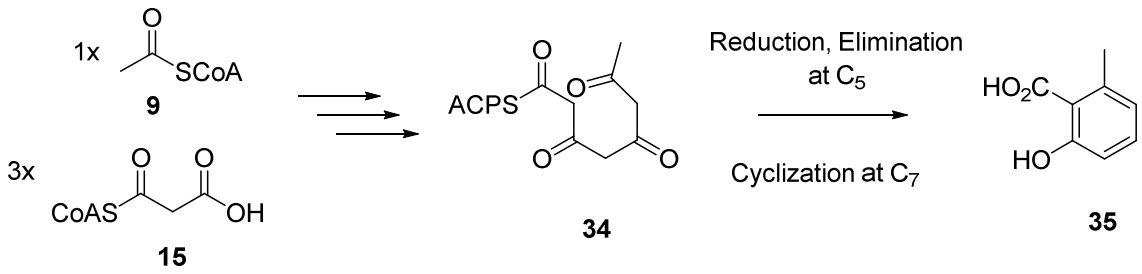
Citrinin is an example of a secondary metabolite which is made by a nr-PKS. The Cox group elucidated the biosynthesis of citrinin **33** and showed that the starter unit is acetate **31**.[63] The biosynthesis starts with an acetate that is extended four times, cyclised and methylated to reach ACP bound intermediate **32** (Scheme 1.8). Further modifying steps (reductions, dehydrations) are not performed by the PKS which lacks all β-processing catalysts.



Scheme 1.8: Biosynthetic steps of citrinin.

1.4.4 Partially Reducing Iterative Polyketide Synthases

Compared to the other iterative PKSs less research has been done for partially reducing synthases (pr-PKS). 6-Methylsalicylate synthase (6-MSAS) is one of the best studied pr-PKS and was the first fungal PKS gene to have been sequenced.[64] The synthase has only one β-processing domain, the KR. The 6-MSAS builts a tetraketide **34** in which the keto group at C-5 is reduced to an alcohol, followed by an elimination and cyclizes in an aldol like reaction to 6-methylsalicyclic acid **35** (Scheme 1.9).[65]



Scheme 1.9: Biosynthetic steps of 6-methylsalicylic acid.

1.4.5 Highly Reducing Iterative Polyketide Synthases

Highly reducing PKS proteins contain all necessary β-processing domains to form a fully saturated carbon-carbon bond. The typical domain order in hr-PKS on the gene level is KS-AT-DH-*C*-MET-ER-KR-ACP. A hr-PKS which is relevant for this work is the squalestatin tetraketide synthase (SQTKS).

Squalestatin S1 (SQS1) **27**, also known as zaragozic acid, shows a broad antifungal activity and is a potent inhibitor of squalene synthase (Figure 1.5).[66] Squalene synthase is a key part of the cholesterol biosynthesic pathway. The chemical structure of SQS1 is characterized by an 2,8-dioxabicyclo[3.2.1]octane-4,6,7-trihydroxy-3,4,5-tricarboxylic acid ring system.

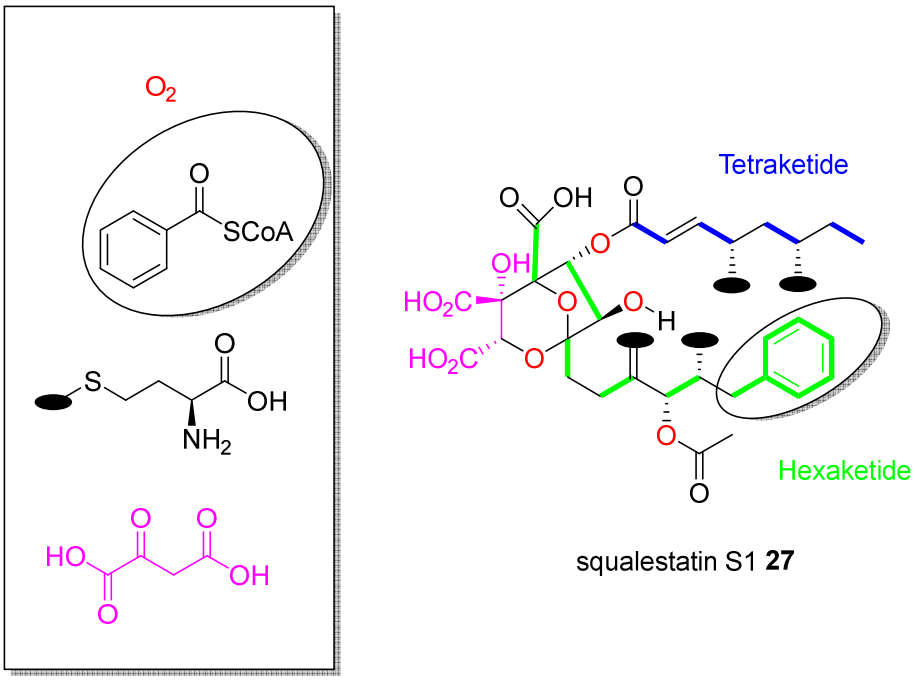


Figure 1.5: Squalestatin S1 consisting of the hexaketide part (green) and the tetraketide part (blue) and the biosynthetic origin of carbon and oxygen atoms.

The origin of the carbon and oxygen atoms of SQS1 **27** was examined by feeding experiments with isotope labelled precursors.[67,68] The group of Kaplan showed that acetate and methionine were incorporated into SQS1 **27** using ^{13}C-labeled substrates. The direction of incorporation and positions of the acetate units were identified by feeding of [1-^{13}C]- and [1,2-$^{13}C_2$] labelled acetic acids. Supplementation with [*methyl*-^{13}C]-methionine resulted in high levels of enrichment at all four methyl/methylene groups indicating that the origin of these groups derives from *S*-adenosylmethionine (SAM). The research for the incorporation of aromatic precursors revealed that *L*-[U-^{14}C]phenylalanine was incorporated as benzoyl-CoA **36** into SQS1 **27** in the starting position of the hexaketide.[68] The origin of oxygen atoms was identified by feeding experiments with $^{18}O_2$ in which five oxygen atoms were incorporated in **27** (Figure 1.5, red marked oxygen atoms).[67]

The incorporation of the remaining carbon atoms was analysed by feeding of different ^{13}C labelled compounds of the citric acid cycle such as aspartic acid, succinic acid, citric acid and oxaloacetate. Labelled succinic acid showed a two-fold higher incorporation into **27** compared to labelled citric or aspartic acid. It was proposed that succinic acid is incorporated into the bicyclic ring of **27**. Later the group of Cox identified the encoding genes for the SQS1 **27** biosynthesis, in which one of the genes encodes a citrate synthase-like protein. This enzyme uses oxaloacetate as a substrate which is more likely incorporated into **27** than succinic acid.[52]

The biosynthesis of SQS1 **27** is encoded by a biosynthetic gene cluster (BGC) which encodes two highly reducing iterative PKS (Figure 1.6 **A**).[49,52] The two PKS produce both chains of SQS1 **27**, the tetraketide and the hexaketide. The tetraketide (Figure 1.5, marked in blue) consists of acetyl-CoA **9** and malonyl-CoA **15** units. The longer chain hexaketide (Figure 1.5, marked in green) is built from benzoyl-CoA **36**, SAM and acetyl/malonyl-CoA units.

The full biosynthetic pathway of SQS1 **27** was elucidated by the groups of Cox and Tang using a combination of directed gene knockout and heterologous expression experiments (Figure 1.6 **B**).[51,69] The hexaketide synthase produces the compound **37** together with the hydrolase (Mfm8, Hyd) and a citrate synthase-like protein (Mfr3, CS). **37** is then oxidized stepwise by the oxygenase Mfr1 presumably *via* the alcohol **38**, *via* the ketone **39**, to the unsaturated ketone **40**. The next oxidation might lead to the epoxide **41**. Next, the second oxygenase Mfr2 hydroxylates two times (**42** and **43**) which then presumably induces a Payne rearrangement forming compound **44** that then reacts

spontaneously to the bicyclic compound **45**. In the next step the oxygenase Mfm1 probably performs an epoxidation resulting in the alcohol **47** that is then acetylated by Mfr4 to give **48**.[52] The last biosynthetic step is the attachement of the tetraketide-CoA **49** to the bicyclic compound **48**, which is catalysed by the acyltransferase Mfm4 and leading to the product SQS1 **27**.[70]

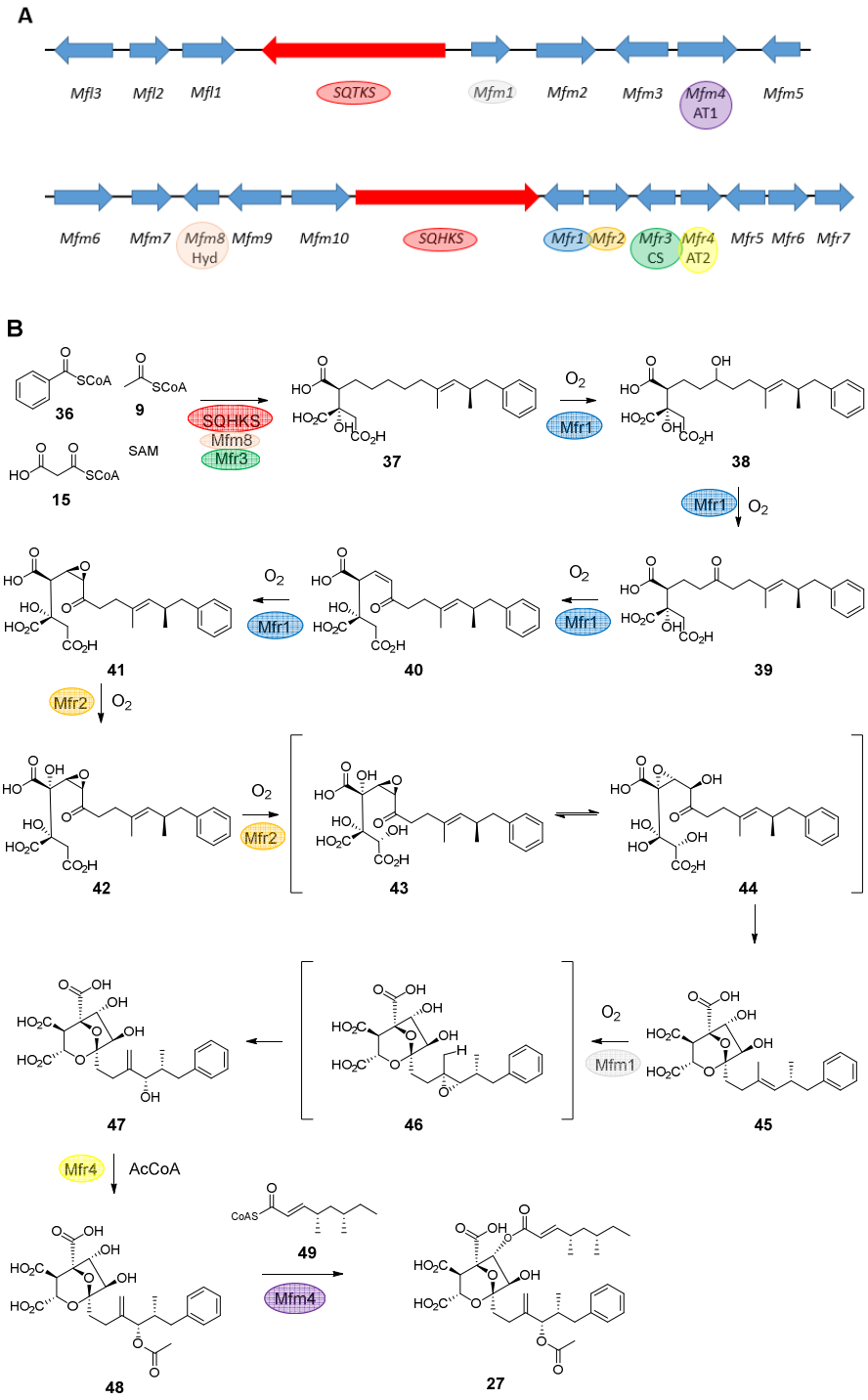
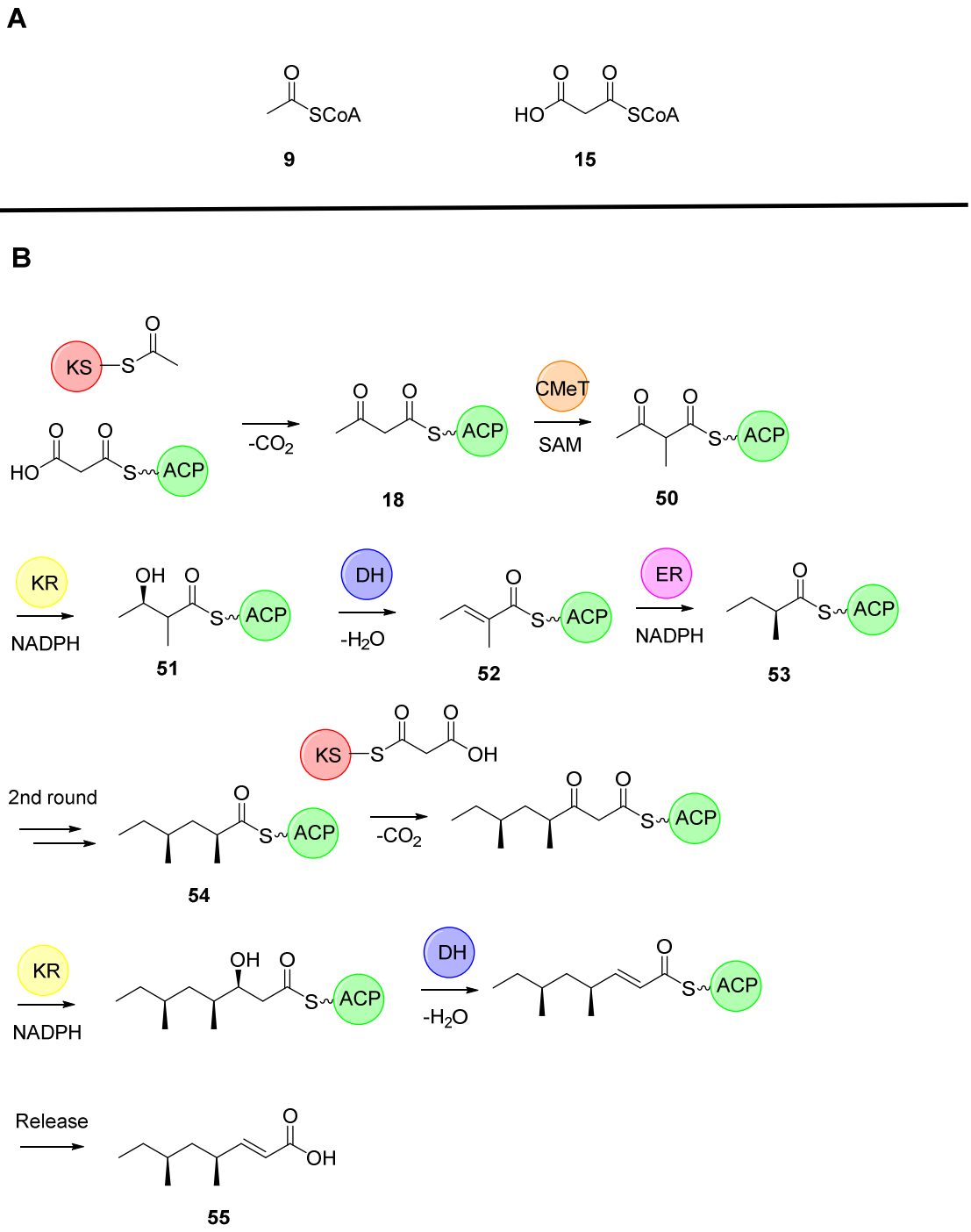


Figure 1.6: Squalestatin S1 biosynthesis: **A**, biosynthetic gene cluster; **B**, proposed sequence of biosynthetic steps.



Scheme 1.10: Biosynthesis of the squalestatin tetraketide: **A**, Starter and extender units; **B**, the biosynthesis of the squalestatin tetraketide.

The special interest in this work is focussed on the squalestatin tetraketide synthase (SQTKS). This PKS consists of seven catalytic domains (KS, AT, DH, *C*-MeT, ER, KR, ACP) which work iteratively. The produced tetraketide is built up from four units by three iterative extension cycles (Scheme 1.10 **B**). The biosynthesis starts with acetyl-CoA **9** (Scheme 1.10 **A**) which is selected by the AT-domain and transferred to the KS domain.

The extender unit malonyl-CoA **15** is loaded to the AT domain and transferred further to the ACP. A carbon-carbon bond is formed between the acetyl and malonyl unit in a decarboxylation reaction leading to the diketide **18**. In the next step *C*-MeT adds a methyl group to the *α*-carbon atom of the diketide **50**. The KR domain reduces the keto group at the *β*-carbon atom to a hydroxyl group **21**. The hydroxyl group is eliminated in the next step by the DH domain, leaving an alkene **52**.

In the last step of the first extension cycle the ER domain reduces the double bond to a fully saturated chain **53**. In the first and second round of the synthesis, all of the β-processing domains (*C*-MeT, KR, DH and ER) are active, generating the product **54**. In the final extension step the *C*-MeT and the ER domains are inactive leading to the unsaturated tetraketide **55**.

1.5 Protein Structure and Mechanisms of Selected Domains

1.5.1 The Ketosynthase Domain

The most crucial step during polyketide synthesis is done by the ketosynthase which catalyzes the carbon-carbon bond forming reaction. The bond forming takes place in the active site of the ketosynthase and is achieved by a decarboxylative Claisen condensation. The starter unit or the growing chain is attached *via* a thiolester linkage. The extender unit is bound to the ACP and acts as a nucleophile.[23]

The mammalian FAS KS has a homodimeric structure where each subunit consists of two subdomains. These subdomains have a βαβαβαββ topology. The dimer interface is stabilized by a pair of hydrogen-bonded antiparallel β-strands.[23] Although many crystal structures of KS domains for FAS and PKS enzymes are available and this domain is one of the best studied domains overall, there is still a disagreement concerning the role of the conserved residues and the proposed mechanism.

The latest KS mechanism is proposed by Khosla and coworkers and is based on structural information of the module 1 KS domain of the 6-deoxyerythronolide B (6-DEBS). The structural information of various site-directed mutants was supported by biochemical analysis which led to the proposed mechanism (Figure 1.7).[71] A propionyl and a methylmalonyl unit are connected during the first cycle.

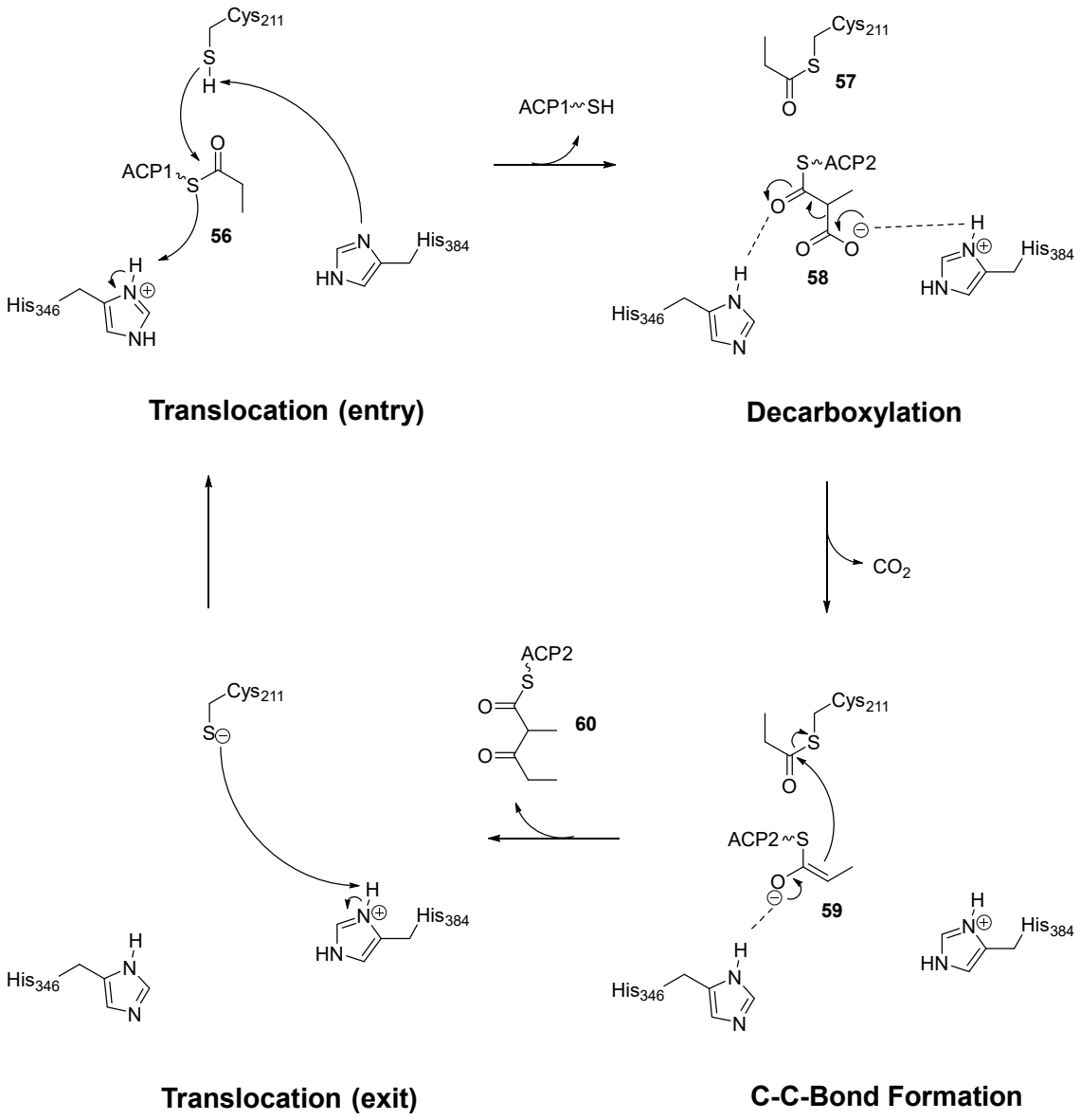


Figure 1.7: Proposed mechanism of the module 1 KS domain in 6-DEBS.[71]

In the first step of the mechanism the propionyl unit **56** is translocated from the upstream ACP (ACP1) of the loading didomain onto the active site cysteine (C211) of the KS **57**. In the second step the methylmalonyl unit **58**, which is already bound to ACP2 undergoes the decarboxylation reaction, forming the enolate **59**. Histidine 346 and Histidine 384 may stabilize the transition state for this reaction. In the third step the carbon-carbon bond forming reaction takes place, which forms the diketide **60**. A proton transfer between the H384 and C211 takes place in the last step and the KS domain is able to perform the next catalytic cycle. The role of the two histidines of the active site has not been extensively studied and their role as acid base catalysts are not verified. The group showed that the mutants H384A and H211A had no or only decreased substrate turnover. Their exact

function or the reason for the decreased activity was not investigated in detail. The group did not consider in their discussion the concerted mechanism of the yeast FAS KS which was proposed in 1975 by the group of Lynen based on investigations with deuterated substrates.[72]

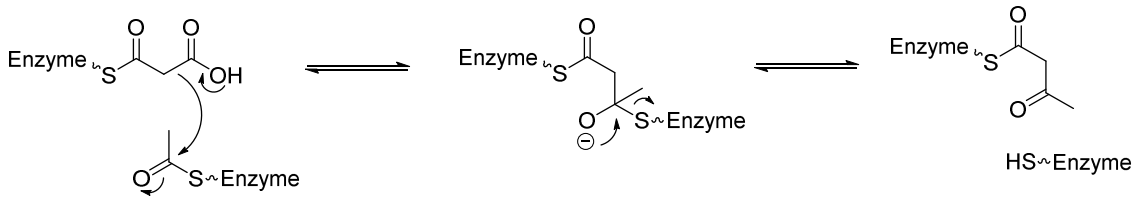


Figure 1.8: Proposed mechanism of the yeast FAS KS domain.[72]

The investigations of Lynen and coworkers showed that deuteromalonyl-ACP had no kinetic isotope effect on the reaction velocity, which favors the proposed concerted mechanism of the reaction (Figure 1.8). This work does not exclude the possibility of a proton transfer from residues of the active site leading to the formation of the KS product but the missing kinetic isotope effect favors the concerted mechanism. The missing kinetic isotope effect loses its significance in the case that the proton transfer reaction is not the rate limiting step in the reaction cascade. The work of Lynen also showed that no protons from the solvent (water) were incorpoarated but the possibility of a back and forth proton transfer from the active site residues was not excluded. The exact mechanism of the KS domain is still not examined in full detail. It is also possible that the mechanism of the KS DEBS module 1 is catalyzed slightly different compared to the yeast FAS KS domain. Anyway, more detailed investigations are necessary in future to fully understand the mechanism of the KS domain.

1.5.2 The Acyltransferase Domain

The acyltransferase (AT) is an essential part of the PKS and FAS as it selects the precursors and loads the units onto the ACP.[73] In FAS and iterative fungal PKS the AT always specifies malonyl CoA. However, bacterial modular PKS ATs very often use methyl (or other alkyl) malonyl CoAs as extender units. The AT domain transfers acyl moieties between CoA and ACP by the *ping-pong bi-bi* mechanism. A highly conserved region, the GHSXG motif is responsible for the catalytic activity. It lies between a β sheet and an α helix and is called the 'nucleophilic elbow'. The nucleophilicity of the active

site serine is enhanced by the helix dipole moment. The serine attacks the CoA substrate **61** in the first step. Two backbone amides function as an oxyanion hole which stabilizes the negative charge of the substrate thiolester carbonyl group **62**. The next important residue is the highly conserved histidine working as an acid/base catalyst deprotonating the nucleophilic serine. The active side histidine protonates the CoA moiety and the serine bound substrate **63** is replaced by ACP (Figure 1.9) leading to the ACP bound product **64**.[23]

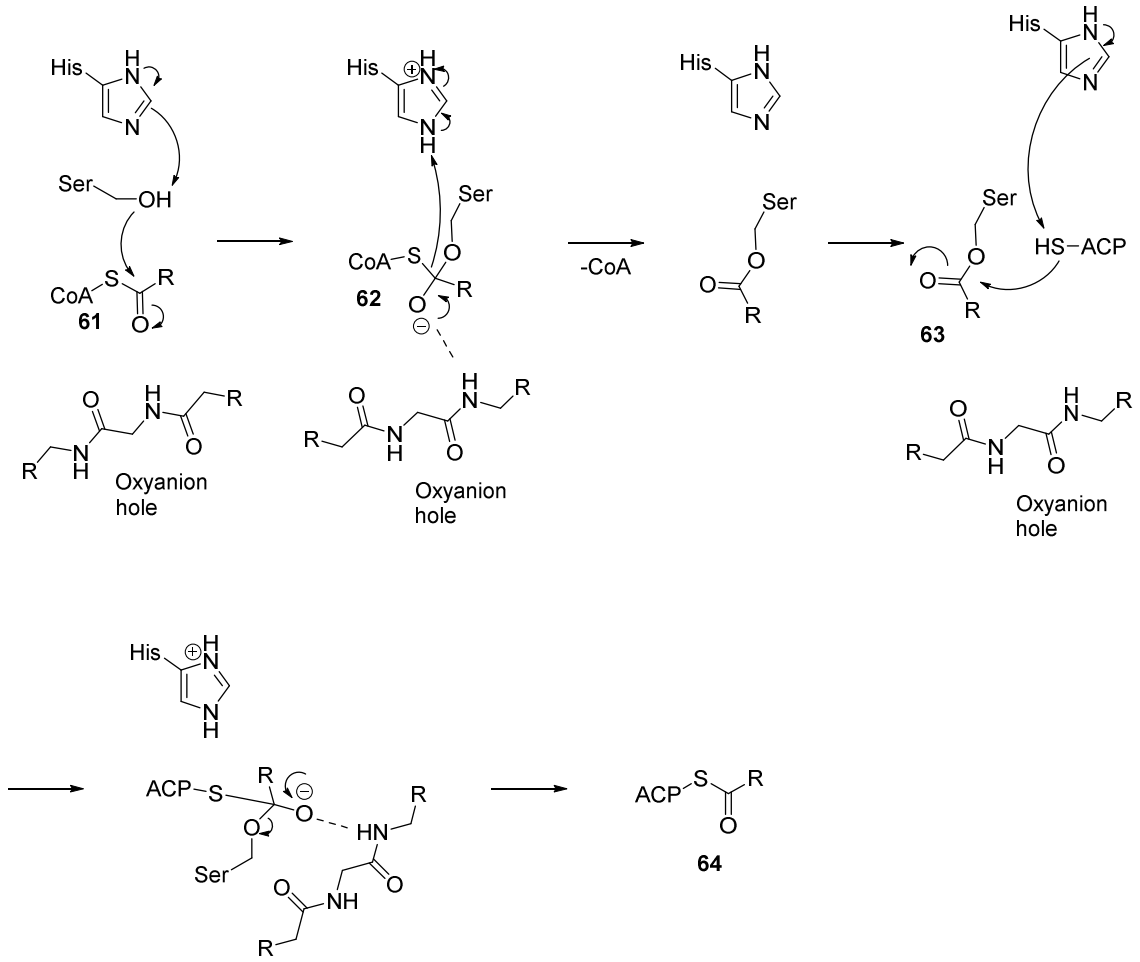


Figure 1.9: Mechanism for the Acyltransferase domain.[23]

In modular PKS the substrate selectivity of AT domains (*i.e.* towards malonyl or alkyl CoAs) is highly influenced by a conserved arginine in the active site which interacts with the carboxylate and stabilizes the charge. Some ATs like the AT from 6-DEBS 1 or avermectin module 2 accept different CoA substrates and do not share the active site arginine. Instead they have a tryptophan residue which allows more flexibility.[46]

There are crystal structures of a number of ATs from polyketide synthases and the related fatty acid synthases which have been determined, for example the AT domains of

6-DEBS modules 3 and 5.[74,75] Concerning structures of ATs of iterative PKSs, there is a crystal structure of a bacterial Type I iterative PKS (Figure 1.10).[76] The enediyne PKS transfers only malonyl units to the ACP but not acetyl starter units, unlike the AT domains of other iterative PKSs. The structure of the enediyne AT domain is composed of three globular α/β domains. The large subdomain is composed of a four stranded parallel β-sheet and eight α-helices. The small subdomain adopts a ferredoxin like fold comprising a four-stranded anti-parallel β-sheet and two α-helices. The enediyene PKS possess a domain organization that is different to other known PKSs. Due to the low sequence similarity and different domain organization this PKS cannot be compared to fungal iterative PKSs. A crystal structure of a fungal Type I iterative PKS is still awaited.

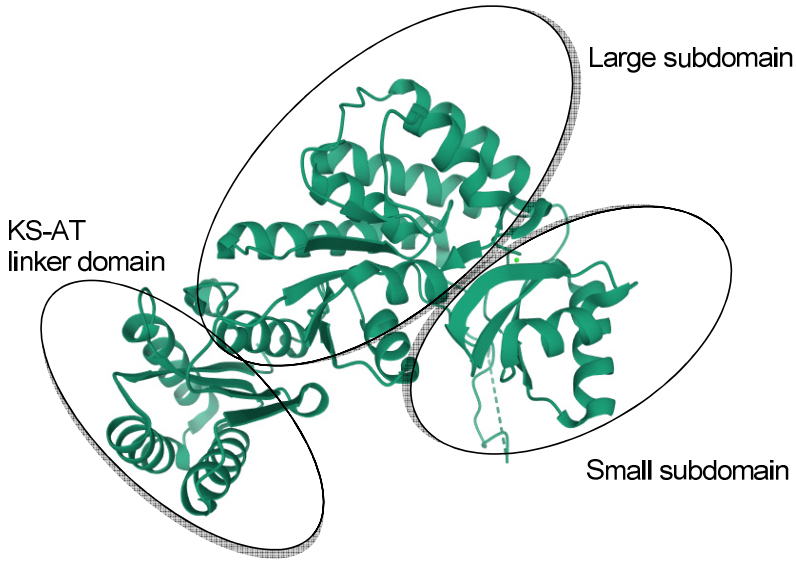


Figure 1.10: Structure of the AT domain of the DynE8 PKS (PDB: 4AMM).[76]

1.5.3 The Acyl Carrier Protein

The ACP is a central and essential component of all FAS and Type I and II PKS pathways. ACPs are small, globular proteins that consist of less than 100 amino acids.[77,78] The ACP serves as an intermediate transporter and interacts with other modifying domains. For its biological function a CoA derived 4'-phosphopantetheine (PP) prosthetic group, to which the intermediates are bound, is necessary (Chapter 3.3.5).

There are several structures for acyl carrier proteins available which have been solved by NMR[79–81] or X-ray crystallography.[82,83] Protein structures were analyzed in *apo*- (lacking PP) or *holo*-(with covalent PP) form and with or without substrate**.** The typical structure of an ACP consists of three or four α-helices and contains a conserved

serine residue, to which PP is attached, located at the N-terminus of the second helix. ACPs have been shown to be very dynamic and flexible which may be a reason for the property of the recognition by a series of proteins.[78] A structure for an iterative Type I ACP is available. The structure of a bacterial Type I iPKS CalE8 was solved (Figure 1.11).[84] Using NMR techniques the group of Yang showed that the CalE8 ACP is a highly dynamic protein which consists of the α-helices 1-3. A mutant with a disulphide linkage between helix 1 and helix 3 was more rigid, with a lower flexibility compared to the wild type. Despite the reduction of flexibility of the ACP, the mutant showed activity and was able to synthesize the PKS product at the same rate as the wild type. In this particular case the dynamic properties of the ACP seem not to be critical for the function of the PKS.

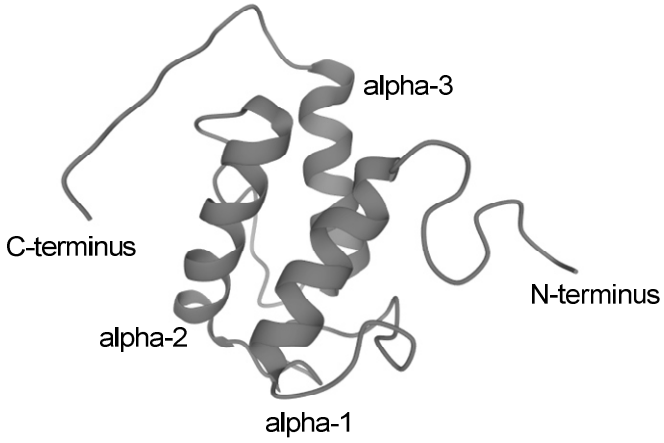


Figure 1.11: Protein structure of an iterative Type I PKS CalE8, solved by NMR (PDB: 2L9F).

1.6 Structural Analysis of Crosslinked FAS/PKS Domains

Protein-protein interactions are essential processes during fatty acid and polyketide biosynthesis. Differences in size, shapes, and electrostatics influence the protein-protein interactions. It is known that these interactions are almost always fast and weak, involving strong recognition rather than strong (and lasting) binding. Such dynamic and transient interactions are hard to study by crystallography or even by NMR. Connecting catalytic domains with the help of chemical tools is a first step in understanding domain interactions, because it forces the proteins to remain in proximity. Crosslinkers have been used to connect the interacting domains in their active sites by an irreversible covalent bond. This causes a static position of the domains to each other. Nevertheless, the

crosslink shows the most important step, a snapshot of the dynamic process, of an interaction between two active sites and can be used for structural analysis.

1.6.1 Crosslinking between an Acyl Carrier Protein and a Ketosynthase Domain

The first crosslinking probes were based on known inhibitors and substrates of a particular domain. For example, epoxypantetheine **68** is a crosslinker designed to link the ACP and the KS domain and was inspired by cerulenin **69**, an inhibitor of KS domains (Figure 1.13).[85–87] The active site cysteine of the KS domain attacks the epoxide, forming an irreversible covalent adduct with the enzyme.[88] Hydrolysis of the epoxide moiety was a problem in aqueous solution. Improved crosslinkers were designed which react with thiols. *Cis* **70** as well as *trans* β-chloroacrylate linker **71**, crosslinked a carrier protein with a KS domain *via* an irreversible 1,4-conjugate addition. All mentioned crosslinkers were converted from pantetheine analogues **65** into CoA analogues **66** using PanK, PPAT and DPCK (enzymes and their detailed function are described in chapter 5) and loaded by a PPTase to ACP **67**.

The Townsend group linked an ACP with a KS domain of a non-reducing iterative Type I PKS, CTB1. They elucidated the structure of the crosslinked protein by cryo-electron microscopy with a resolution of 7.1 Å. (Figure 1.13 **A**).[89] The thiol of the active site cysteine of the KS reacts with the bromohexyl moiety **72** forming an irreversible covalent bond.[90]

Another successful ACP-KS crosslink was performed with a KS domain (FabB) of a Type II fatty acid synthase and the *E. coli* acyl carrier protein (AcpP, Figure 1.13 **B**). Structural information was obtained by X-ray crystallography, NMR and molecular simulations.[34] For the crosslinking again the *trans* β-chloroacrylate linker **36** was used. The structural results revealed that helix II of AcpP is relevant for anchoring to facilitate correct orientation of AcpP for productivity. Helix III of AcpP may be important for the chain translocation of the fatty acid.

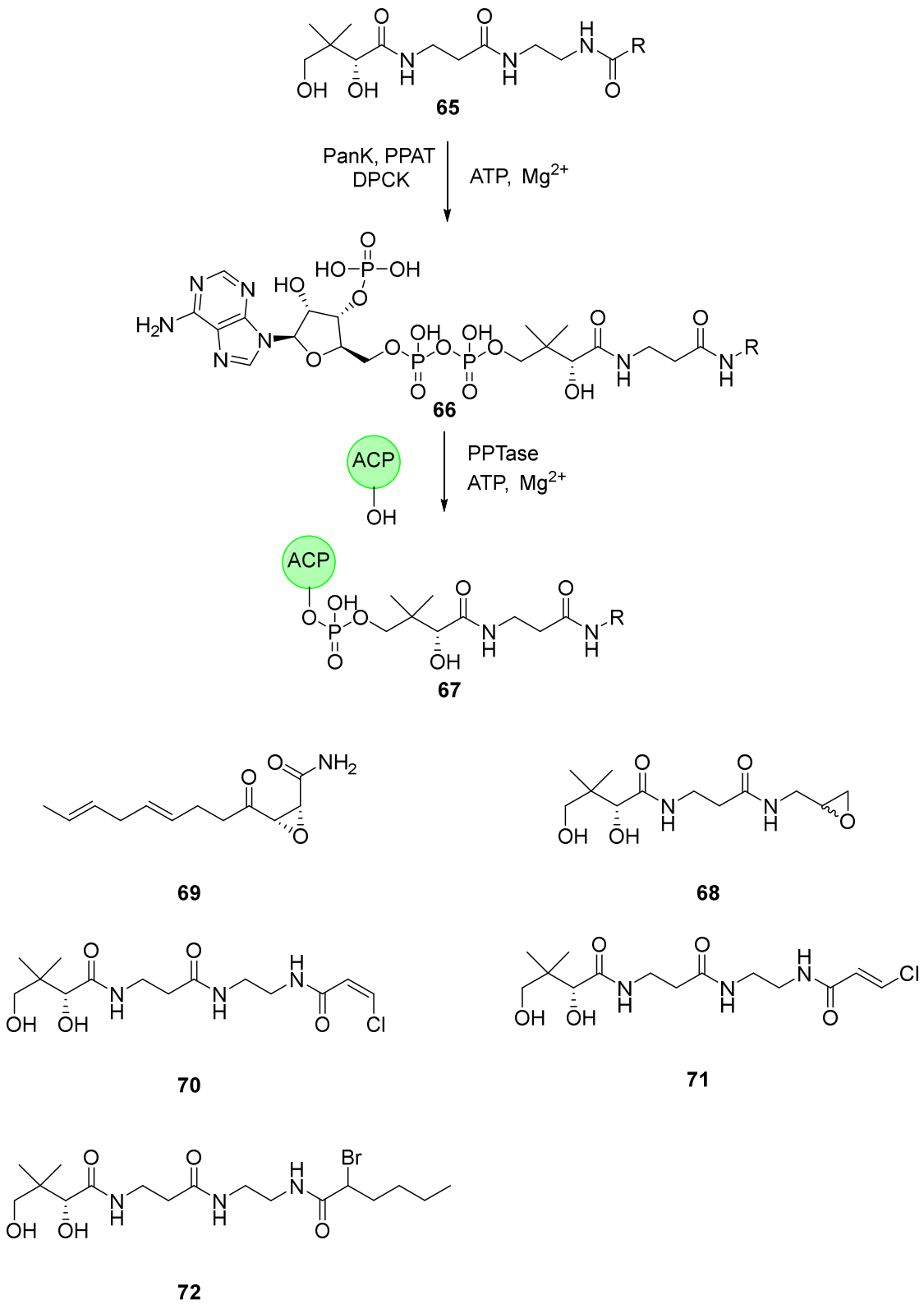


Figure 1.12: Structures of ACP-KS crosslinkers and of the inhibitor cerulenin.

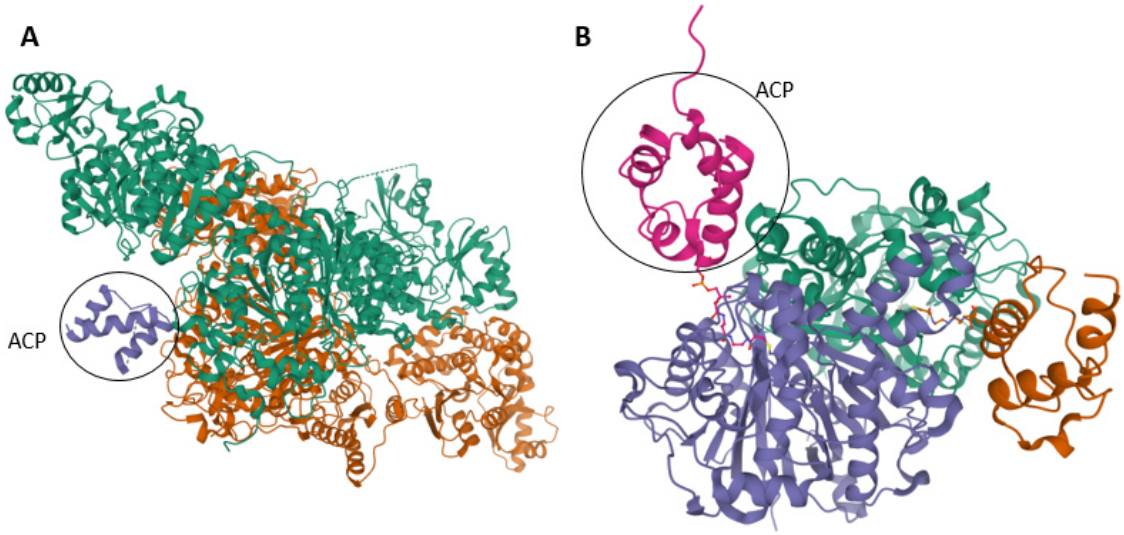
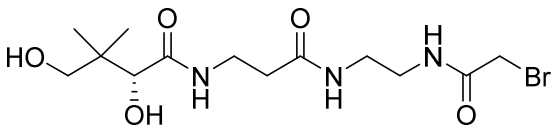


Figure 1.13: Structures of crosslinked ACP-KS proteins: **A**, structure of ACP2 crosslinked with CTB1 KS domain (PDB: 6FIK); **B**, structure of linked KS and ACP of a FAS II system (PDB:5KOF).

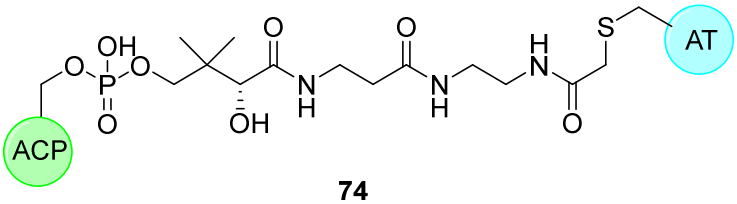
1.6.2 Crosslinking between an Acyl Carrier Protein and an Acyltransferase Domain

Crosslinking between an ACP and an AT domain was performed for a FAS Type II system by the group of Burkhart. The ACP (AcpP) and the AT (FabD) were connected by the cysteine specific linker **73** (Figure 1.14).[33] The linker was transformed into a CoA analogue and loaded onto the ACP as already described for the KS crosslinking (section 1.6.1).To use this similar approach for the KS crosslinking a mutation was introduced into the active site of the AT domain (S92C) to generate a thiol and use it for the linkage **74**.

The X-ray crystal structure with a resolution of 1.9 Å is supported by a combination of mutations, kinetic analyses and long-timescale molecular dynamics simulations. The combined data suggests that AcpP adopts multiple, productive conformations at the AT binding interface, with rigid body subdomain motions within the FabD structure that may play a key role in AT activity and substrate selectivity (Figure 1.15). The MD simulations demonstrate that the ACP can undergo rotational and translational motion at the AT interface, while still positioning its PPant bound substrate for the catalytic reaction.



73



74

Figure 1.14: Crosslinking between AcpP and FabD.

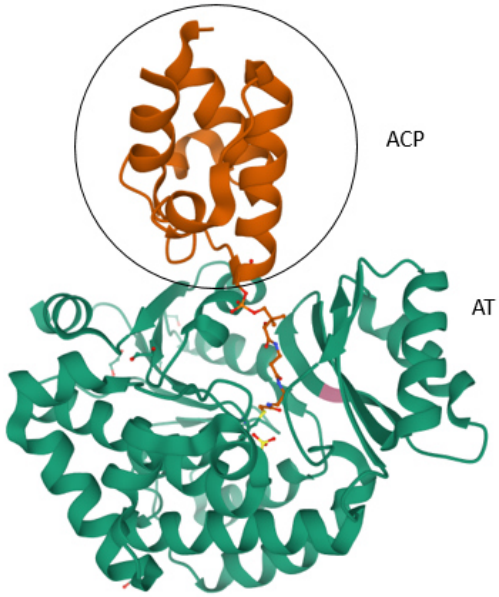


Figure 1.15: Protein structure of crosslinked AcpP and FabD (PDB: 6U0J).

Two other ACP-AT structures are known and both belong to Type I *trans*-AT PKS systems, VinK-VinL and disorazole synthase (DSZS).[91,92] A comparison to these ACP-AT structures revealed that the orientation differs between the three complexes. Hydrophobic interactions are more prominent at the *trans*-AT complexes than at the AcpP-FabD interface.

1.6.3 Crosslinking between an Acyl Carrier Protein and a Dehydratase Domain

The group of Zhang successfully obtained the crystal structure of the *H. pylori* FAS ACP with the DH of a Type II FAS (FabZ).[38] No crosslinker was used to obtain this crystal structure. Instead of using an alkyne linker (section 7.1) for the DH FabA, *holo*-ACP was heterologously expressed, mixed with FabZ in a molar ration of 7:1, purified by gel filtration and crystallized within 30 days under 277 K in hanging drop. The crystal structure (Figure 1.16) revealed that the protein-protein interaction between these proteins are dominated by electrostatic interactions. An unexpected 2:1 (DH:ACP) stoichiometric binding was observed. One ACP molecule binds to a FabZ dimer.

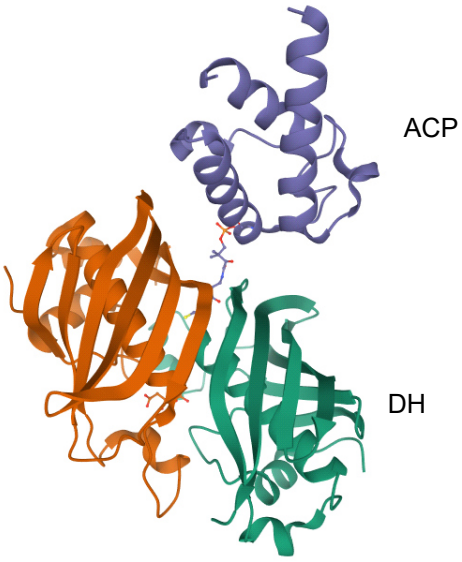


Figure 1.16: Structure of *H. pylori holo*-ACP in interaction with the DH (FabZ, PDB: 4ZJB).

Structural analysis together with biophysical and computational results showed a novel dynamic seesaw-like ACP binding and catalysis mechanism for the DH in the FAS system. The DH domain has a pseudodimeric structure and can only catalyse one ACP at a time. The binding of the first ACP (ACP1) with the alpha2 helix to the β-sheet layer of the active site of monomer A induces a conformational change to the β-sheet layer of monomer B, causing an inability to bind a second ACP at the same time (Figure 1.17). Past the dissociation of ACP1, ACP2 is now able to bind to monomer B and causing the same effect to monomer A.

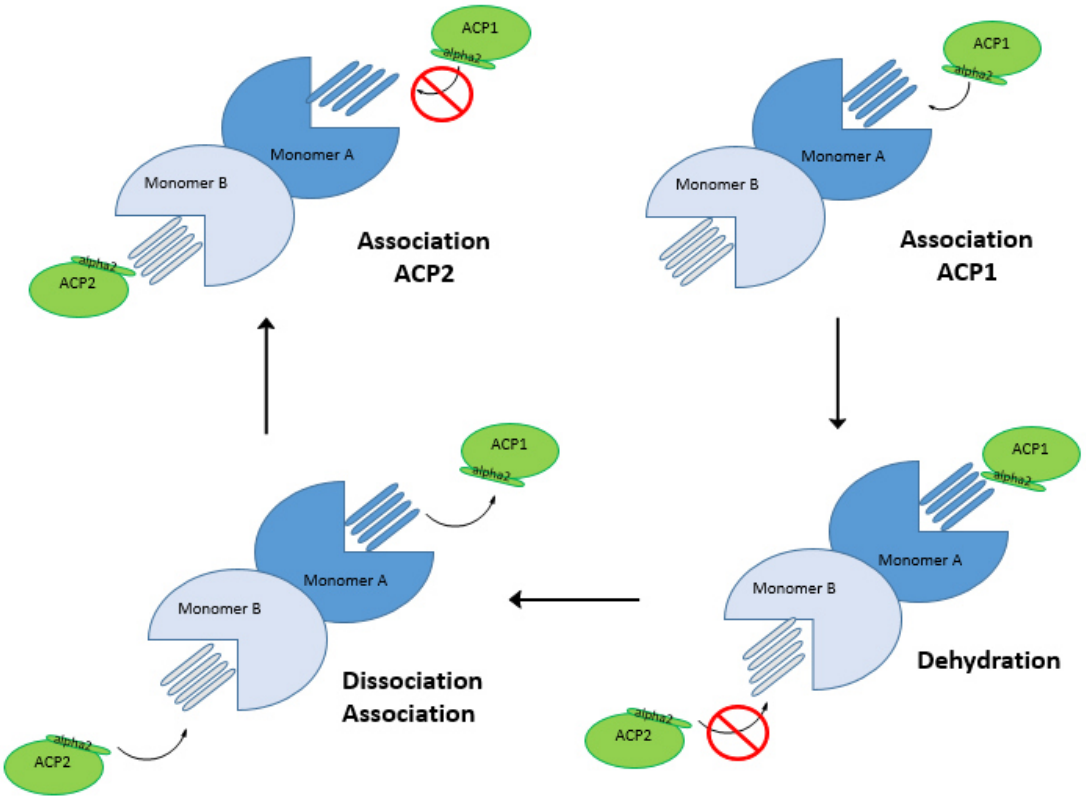


Figure 1.17: Schematic diagram of the dynamic seesaw-like catalytic mechanism of the FabZ pseudo dimer (monomers in light blue and blue) with two ACP molecules (green).[38]

1.7 Main Aims for the Projects

Although a lot of structural information is available for fatty acid synthases and components of modular PKS, the progress in this field was limited for polyketide synthases, especially for fungal iterative PKS, and hr-PKS in particular. By crosslinking fungal iterative Type I PKS domains with the acyl carrier protein of the SQTKS pathway and the byssochlamic acid pathway it was hoped to get first structural information similar to fatty acid related research. For this approach different types of linkers will be synthesized, in cooperation with the Kirschning group, as pantetheine derivatives and two fungal hr iPKS ACPs will be expressed. The ACPs belong to the squalestatin and byssochlamic acid pathways. A successful loading of these linkers need four steps. The first three steps are catalysed by biosynthetic enzymes (PanK, PPAT, DPCK) which create the CoA analogue of the linker. In the fourth step a phosphopantetheine transferase (Sfp) is needed to load the CoA linker to the ACP. These four additional enzymes need also to be expressed first to reach this goal. As the linkage partners of the ACPs the KS/AT didomain of the squalestatin pathway and DH domains of the squalestatin and byssochlamic acid pathway will be expressed heterologously. A third DH domain of the

strobilurin pathway is also available for the linkage experiments. The ACPs will be linked to these domains and it is planned to gain structural information by X-ray scattering techniques, performed by to cooperation of the Weissman group.

In addition to structural investigations, *in vitro* testings using substrate loaded ACP will be performed. The substrate loading will be performed in a similar way as the linker loading. With the substrate loaded ACP in hand we want to determine the substrate selectivity of the citrate synthase of the byssochlamic acid pathway. In the squalestatin pathway the methyltransferase will be checked for activity using an ACP loaded substrate due to inactivity to previously offered pantetheine substrate.

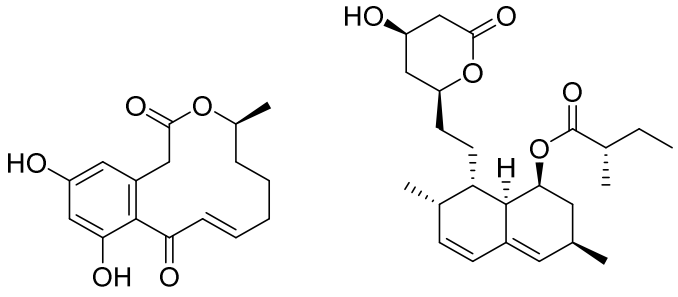
2. SQTKS KS/AT Di-Domain

2.1 Introduction

Understanding substrate specificity and regulatory effects of highly reducing iterative PKSs (HR iPKS) remains a challenge. *In vitro* experiments of single and multi catalytic domains of this type of PKS, in combination with structural information, can help mastering these challenges.

In comparison to modular PKS, fewer catalytic domains of HR iPKSs have been heterologously expressed. Even fewer of them were analysed in detail with *in vitro* studies. The expression of the KS/AT didomain from the lovastatin **4** (Figure 2.1) nonaketide synthase in a soluble and active form is one example of successful work with this type of PKS.[93] In that work the activity and substrate selectivity of the isolated KS and also of the KS/AT didomain was analysed using labelled substrates. A KS/AT mutant with an inactive AT domain was used to assay the KS activity. The expression of the didomain was performed in *E. coli* BL21 DE3 cells and soluble protein was obtained in high concentration (20 mg/ml) with the His-tag at the *C*-terminus using the pSMa30 plasmid. The purification was performed with a Ni-NTA and an anion exchange chromatography. The didomain borders were at the positions M1 and S940.

Other successfully expressed HR iPKSs are Dhc3 and CazF. Dhc3 a PKS involved in the biosynthesis of the phytotoxic macrolide 10,11 dehydrocurvularin **75** (DHC, Figure 2.1). The whole PKS was expressed in yeast and was used for phylogenetic analysis and homology modelling.[94]



75 **4**

Figure 2.1: Structures of DHC and Lovastatin.

CazM, a PKS involved in the biosynthesis of an azaphilone natural product was expressed and analysed by *in vitro* assays revealing the critical role of the SAT domain in a substrate transfer process to a nonreducing PKS.[95]

The *in vitro* investigation of the SQTKS catalytic domains has been in progress for many years. The first expression trials by David Ivison led to the observation of several soluble and functional domains,[96] however, the KS/AT didomain was not one of these. There is no crystal structure available for SQTKS and finding the correct domain boundaries is complex and only an approximation to the real ones. For this reason, Ivison used two different constructs for the KS/AT didomain, both leading to mainly insoluble proteins.

2.2 Aims

The dehydratase (DH) and enoyl reductase (ER) catalytic domains of SQTKS have been studied in detail as isolated domains. This has led to a better understanding of SQTKS.[96–101] To have a deeper knowledge of the complex multi-domain protein it is necessary to analyse the base of it – the KS/AT didomain. Previous work of Hao Yao showed that an expression and isolation of multidomains of SQTKS is possible.[100] The first aim of this project is to have a border structural prediction for the KS/AT didomain which should be as precise as possible. With this in hand an *E. coli* codon optimized template sequence will be synthesized. Diverse starting and end positions of the KS/AT can be generated by PCR. The next aim is the heterologous expression of the KS/AT didomain in *E. coli*. In case of soluble protein it can be purified and characterized with activity assays. *In vitro* experiments with natural and unnatural substrates such as acyl SNACS, pantetheines, CoAs and generating structural information with SAXS would be the last step. If successful these assays could also be the basis for future crosslinking and structural experiments.

2.3 Results

2.3.1 Bioinformatic Solubility Prediction

Protein-Sol is a web server for predicting protein solubility.[102,103] The server calculates 35 sequence-based properties based on the amino acid sequence in the expression system *E. coli*. The algorithm considers the composition of 20 amino acid; 7 composite amino acid measures (K-R, D-E, K+R, D+E, K+R-D-E, K+R+D+E, F+W+Y); overall length; overall pI; overall hydropathy; absolute charge at pH 7; fold propensity; disorder; sequence entropy; and β-strand propensity. The calculated values are then correlated to

values for the population of experimental solubilities for 2395 known proteins and used to assess combinations of features. The final prediction scheme consists of 10 features (H, L, V, K-R, D+E, F+W+Y, length, absolute charge, fold propensity and sequence entropy with a correlation coefficient of 0.621 between calculated and experimental values). Each feature has a positive or negative influence on the solubility. Display of the extent to which each feature deviates from population average allows to select features that could be targeted to improve solubility.

Protein-Sol was performed to analyze the solubility properties of the SQTKS KS/AT protein in the expression system *E. coli*. Solubility prediction on the server is given in the 0-1 range with 0 = predicted insoluble and 1 = predicted soluble. The protein sequence of SQTKS KS/AT was analyzed and gave a predicted scaled solubility value of 0.162 (Figure 2.2, column 2) which is far below the average soluble *E. coli* protein with the value of 0.45 (Figure 2.2, column 1). This means that the theoretically calculated solubility of KS/AT expressed in *E. coli* is more than two times lower compared to proteins of the database. An example of a protein with a high solubility value is thioredoxin, predicted at 0.76, consistent with its wide use in co-expression or as a fusion partner.[102]

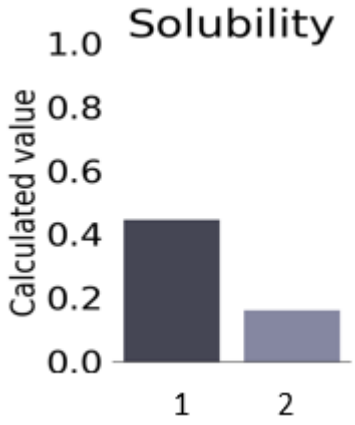


Figure 2.2: Solubility prediction for *E. coli* expressed SQTKS KS/AT didomain using Protein-Sol web server with column **1:** population average solubility of proteins of the database and column **2:** the solubility of SQTKS KS/AT.

Features such as sequence length, disorder propensity and sequence entropy, calculated by Protein-Sol showed a strong deviation compared to the population average. All of the mentioned values are higher than the average values of the database. These properties lead to an increase in the probability of getting insoluble protein expression in *E. coli*. A higher value in sequence entropy and disorder propensity corresponds with a higher probability of an incorrectly folded protein which has an influence on the solubility.

E. coli cells very often have problems to express large protein complexes and that's why a high sequence length is influencing the solubility value in a negative way (Figure 2.3).

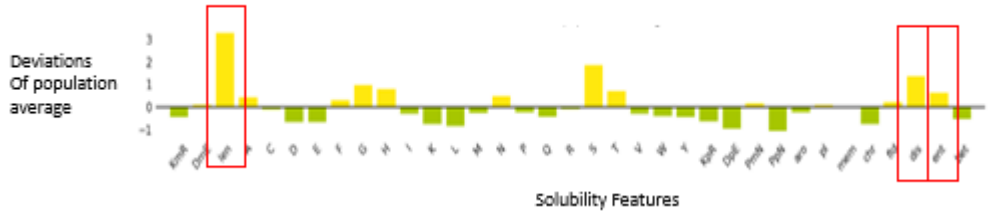


Figure 2.3: Solubility prediction for in *E. coli* expressed KS/AT di-domain with Protein-Sol software. Noticeable values: len: length; dis: disorder propensity; ent: sequence entropy.

Another powerful feature for understanding protein folding is the *windowed fold propensity* per amino acid.[102] This property shows how strongly a defined amino acid region is folded in comparison to average fold propensity of the database. Regions of negative values (Figure 2.4, blue) in *windowed fold propensity* per amino acid suggest the presence of unstructured regions. The region between residues 400 and 450 (Figure 2.4, red box) corresponds to a likely 'linker' region with low structure between the KS and AT domains and fits with the information from the alignment in section 2.3.5. Significant are also the first 30 residues (Figure 2.4, green box), showing a likely disordered region at the N-terminus of the protein. Smaller disordered regions at the C-terminus may also represent a potential linker sequence.

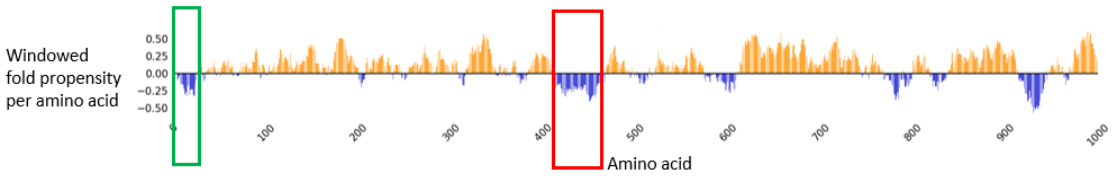


Figure 2.4: Windowed fold propensity per amino acid is shown for the whole sequence length.

The bioinformatic results lead to a conclusion that *E. coli* seems to be not a good organism for the production of the SQTKS KS/AT protein. The calculated solubility value is far below the average value and the probability to express soluble protein in this organism is very low. But it should be always considered that the database includes more bacterial related proteins and fewer fungal ones, and the prediction accuracy for fungal proteins is lower than for bacterial proteins.[102] Performing *in vitro* experiments requires a high amount of purified protein. Until today *E. coli* is the best expression system for producing large amounts of protein in a short time.

2.3.2 Expression of KS/AT Di-Domain in Bl21 DE3 Cells

The KS/AT gene sequence is positioned in the 5'-terminus of the SQTKS PKS gene (Figure 2.5). Finding the exact border between AT and the following DH domain is the most difficult part of the border prediction for the KS/AT didomain. The border prediction is based on an alignment with human and rat FAS and on a comparison with the crystal structure of mFAS (described in detail in 2.3.5).

Figure 2.5: Domain order in the highly reducing PKS of Squalestatin. KS: Ketosynthase; AT: Acyltransferase; DH: Dehydratase; CMeT: Methyltransferase; ER: Enoylreductase; KR: Ketoreductase; ACP: Acyl Carrier Protein.

A codon optimization for *E. coli* cell expression was performed using web tools to improve the expression rate and minimize expression aborts due to uncommon fungal codons.[104–106] The KS/AT gene includes 1000 amino acids beginning from the N-terminus (M1 – H1000), was synthesized by the company Baseclear and was subcloned into an expression vector pET28b including a stop codon at the C-terminus and a His-tag encoded at the N-terminus of the gene. The gene includes a small part of the beginning of the DH domain (probably 12 amino acids from the C-terminus) for having the possibility to create different constructs by cloning as the end of the AT domain cannot be predicted assuredly. The vector containing the KS/AT gene was successfully transformed into *E. coli* Top10 and BL21 cells.

A first KS/AT expression in *E. coli* BL21 was performed. For induction, 0.1 mM and 0.2 mM IPTG were used with an expression time of 22 h at 16 °C in 2TY medium. Low expression temperature and a moderate inducer concentration should lead to a slower expression and a higher probability to obtain soluble protein. The negative control contained the same KS/AT pET28b plasmid but was not induced by IPTG addition. In the sample of 0.1 mM as well as in 0.2 mM IPTG induction the protein of the expected size of 110 kDa was obtained as an intense band in the insoluble fraction (Figure 2.6).

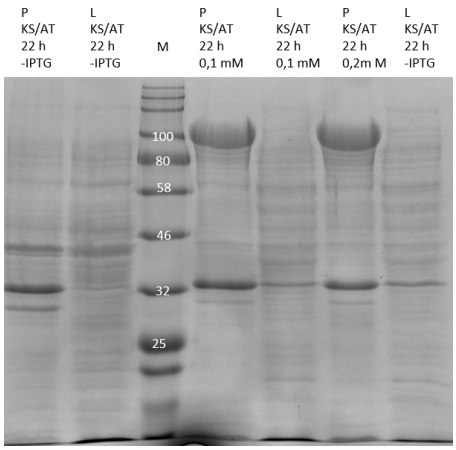


Figure 2.6: SDS-PAGE of full length KS/AT protein expressed in BL21 cells after induction with 0.1 and 0.2 mM IPTG and an induction time of 22 h at 16 °C with soluble (L) and insoluble (P) parts.

The IPTG concentration and the low expression temperature of 16 °C had no influence on the solubility. The variation of induction time (5 and 22 h) gave no improvement in terms of solubility either.

2.3.3 Variation of Expression Strains

The first approach for expressing KS/AT in *E. coli* BL21 gave poor results. An alternative to *E. coli* BL21 DE3 is the *E. coli* Arctic express (DE3). The strain is specialized for expression of proteins at low temperatures. Low temperatures lead to a slower expression in which the growing polypeptide chain has the time to fold in a different way than in a fast expression. In addition, these cells constitutively express the chaperonins Cpn10 and Cpn60 from *Oleispira Antarctica.* These proteins have high protein refolding activities in a temperature range between 4 and 12 °C.[107,108]

The pET28b plasmid containing the KS/AT construct was first transformed to *E. coli* Top10, amplified and transformed into *E. coli* Arctic Express. The heterologous expression was performed in 2TY medium. Cultures were grown at 30 °C to an OD_{600} of 0.4 and cooled to 10 °C for the induction of the cold shock chaperones. An induction followed with 0.1 mM and 1 mM IPTG for 21 h at 10 °C. After expression cells were lysed and checked by SDS-PAGE.

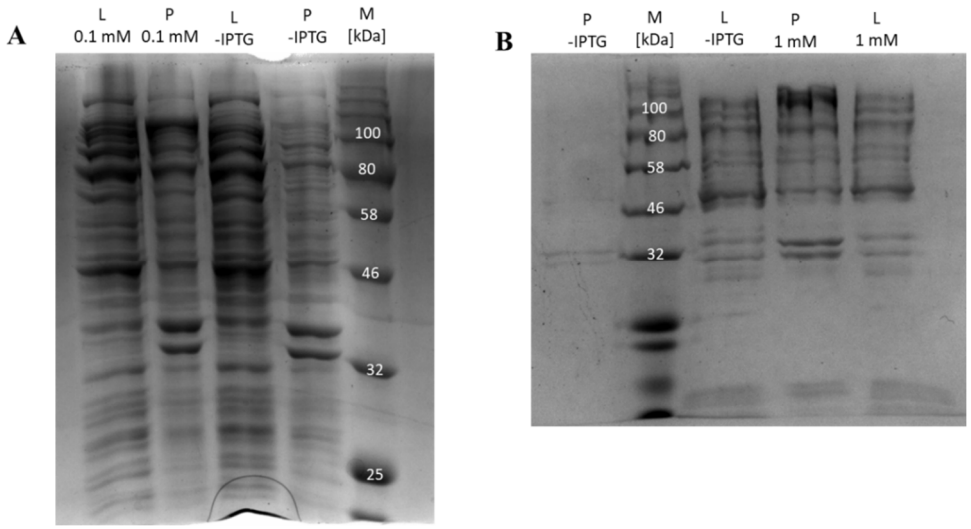


Figure 2.7: Expression of SQTKS KS/AT in *E. coli* arctic express (DE3) with soluble (L) and insoluble (P) parts: **A**, induction with 0.1 mM IPTG for 21 h at 10 °C; **B**, induction with 1 mM IPTG for 21 h at 10 °C.

The KS/AT protein at the expected size of 110 kDa is hard to distinguish from the background proteins produced by the *E. coli*. Some intense bands in the insoluble fractions and weak bands in soluble fractions in the expected range are visible (Figure 2.7). Unfortunately, *E. coli* arctic express cells did not give the desired amount of soluble KS/AT protein. For this reason, no large scale expression was performed.

The second alternative strain to BL21 DE3 is the *E. coli* TaKaRa strain. These cells contain additional plasmids with six possible chaperones which can be induced all at the same time or in different combinations. Chaperones assist in proper folding of other proteins or bring incorrectly folded proteins into the right conformation.[109] Inducible chaperones of the TaKaRa strain, their sizes and the corresponding inducers are listed in table 2.1.

For the induction of the chaperones L-arabinose (c = 0.5-4 mg/ml) and tetracycline (c = 1-10 ng/ml) were added. Optimal concentration and combination of inducers were screened in seven experiments (Table 2.2). After an OD_{600} of 0.6 was reached, the cultures were left for 30 min at 15 °C, before protein expression was induced by adding 1 mM IPTG. Cultures were kept for 24 hours at 15 °C, the cells were lyzed and the supernatant was separated from the cell pellet followed by SDS-PAGE analysis. According to table 3 the entries A-C and E-G showed a weak protein band in the soluble fraction at the expected size (Figure 2.8). In BL21 cells without any induced chaperones no soluble protein at 110 kDa could be observed. Further the soluble fraction of entry A showed a band at about 110 kDa which might be the desired protein.

Table 2.1: Chaperones coded on specific plasmids in *E. coli* TaKaRa. All of the plasmids have the Cm resistant marker.

No.	*Plasmid*	*Chaperone*	*Inducer (final conc.)*
1	pG-KJE8	dnaK(70kDa)-dnaJ(40kDa)-grpE(22kDa)-groES(10kDa)-groEL(60kDa)	L-Arabinose (0.5 mg/ml) Tetracycline (1-5 ng/ml)
2	pGro7	groES-groEL	L-Arabinose (0.5 mg/ml)
3	pKJE7	dnaK-dnaJ-grpE	L-Arabinose (0.5 mg/ml)
4	pG-Tf2	groES-groEL-tig	Tetracycline (1-5 ng/ml)
5	pTf16	tig	L-Arabinose (0.5 mg/ml)

Table 2.2: Screened L-arabinose and tetracycline concentration for the TaKaRa cells.

Entry	*L-Arabinose [mg/mL]*	*Tetracycline [ng/mL]*
A	0.5	1
B	1.5	3
C	2.5	6
D	4	10
E	2.5	-
F	-	6
G	0.5	5

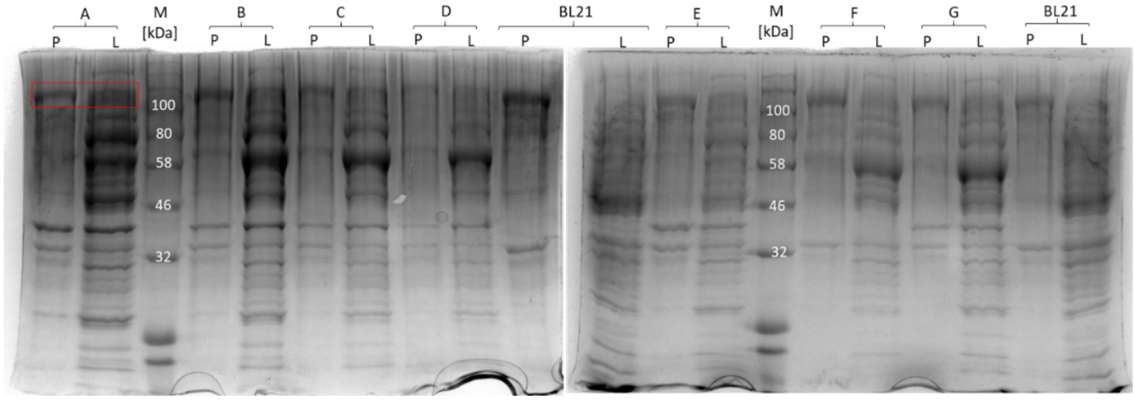


Figure 2.8: SQTKS KS/AT expression in *E. coli* BL21 DE3 and TaKaRa strain induced by 1 mM IPTG for 24 h at 15 °C with soluble (L) and insoluble parts (P): **A-G**, Experiments with chaperone inductions according to table 2.2.

Expression experiment A (0.5 mg/ml L-arabinose, 1 ng/ml tetracycline) was used for a large scale expression experiment. 1.2 l of culture were cultivated as in the small scale experiment and were purified with the FPLC. The first purification step was performed by a Ni-NTA column. The protein was then desalted *via* size exclusion chromatography and concentrated with a 30 kDa Amicon® Ultra centrifugal filter. A SDS-PAGE analysis was performed to analyze the purification success but no soluble protein could be observed at the expected size. The large scale expression combined with a purification showed that no soluble KS/AT was obtained. The observed band in the soluble fraction of experiment A was an *E. coli* correlated protein.

Protein expression in bacterial cells like *E. coli* results in expression of the desired protein at high translational rate, which can exhaust the bacterial protein quality control system. The partially folded and misfolded proteins aggregate to form inclusion bodies.[110] Inclusion bodies can be used for refolding approaches to solve the insolubility problem.[111] A successful example of one refolding approach is the refolding of procathepsin D from inclusion bodies.[112] The protein was expressed as a tag free protein and was also fused with the maltose binding protein (MBP). After the expression and cell lysis, the inclusion bodies were solubilized in 8 M urea. Refolding was achieved by removal of the urea *via* dialysis. The experiments showed that the tag free protein precipitated while the MBP fused protein was in soluble form after the refolding.

2.3.4 Refolding Approaches

As the change of expression strains did not give any improvement concerning solubility of the SQTKS KS/AT protein, refolding experiments were tried with the precipitated protein. For this approach different detergents were used. All detergents are characterized as containing a hydrophilic head region and a hydrophobic tail region. In solution protein detergent complexes are formed, where the detergent hydrophobic regions bind to the protein hydrophobic domains protecting them from aggregations.[113]

The first tried detergent was *N*-Lauroylsarcosine. The KS/AT protein was expressed for 22 h at 16 °C in *E. coli* BL21 DE3 cells, using 1 mM IPTG for induction. The cells were harvested by centrifugation and lysed by sonication. In the next step the pellet was separated from the supernatant by centrifugation. The pellet, containing the inclusion bodies, was resuspended in a Tris buffer containing 1% of *N*-Lauroylsarcosine and stirred for two hours. Analysis of the performed SDS-PAGE indicated no soluble protein of expected size in the soluble fraction. No improvement in comparison to the attempts without the detergent could be observed (Figure 2.9).

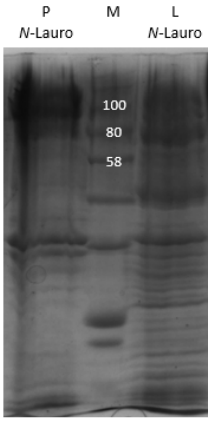


Figure 2.9: Solubilization attempt of in *E. coli* BL21 DE3 expressed SQTKS KS/AT with 1 % *N*-Lauroylsarcosine with soluble (L) and insoluble (P) part.

The second refolding approach was done with urea. This molecule has chaotropic properties and is able to disrupt the hydrogen bonding network between water molecules. During this process the protein gets reversibly denatured and the disulphide bonds are reduced with *β*-mercaptoethanol. The protein gets unfolded and has no activity. After slow removal of urea, the protein has a chance to fold in a different way.[114,115]

The SQTKS KS/AT protein was expressed in the same way as described for the *N*-lauroylsarcosin experiment. The cells were lysed in an EDTA containing buffer (buffer I, section 10.4.7.2) and the insoluble part was separated from the soluble by centrifugation. The pellet was divided in four aliquots. Each aliquot was suspended in buffer II (section 10.4.7.2) containing β-mercaptoethanol and different concentrations of urea. 2, 4, 6 and 8 M urea concentrations were used for the four aliquots. The aliquots were incubated for 1 h at room temperature and 40 °C. Incubation with urea at room temperature did not lead to solubilization of protein. The aliquots which were supplemented with 6 M or 8 M urea and incubated at 40 °C contain soluble protein of the predicted size of 110 kDa (Figure 2.10). Also the 4 M lysate contains small amounts of the expected protein. This results lead to the assumption that soluble KS/AT-protein was obtained at higher urea concentration.

The removal of urea was performed by dialysis. The concentration of urea was reduced stepwise by reducing the concentration from 6 M to 4 M within 24 h, then from 4 M to 2 M (24 h) and from 2 M to urea free buffer again in 24 h. Unfortunately, the protein precipitated during this process and no soluble protein was left.

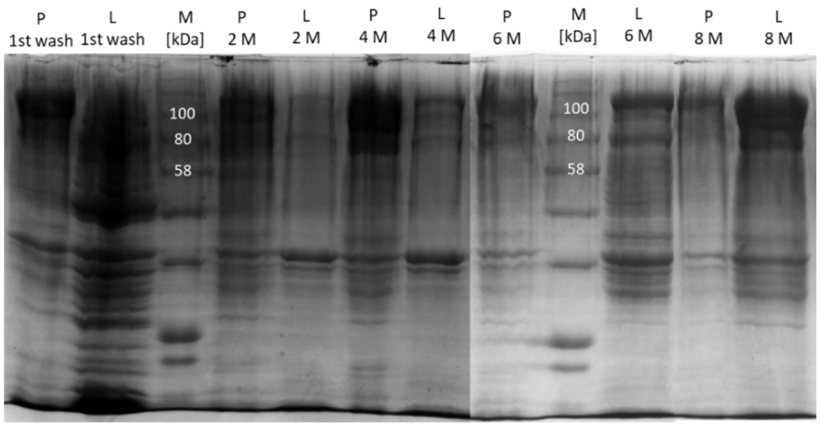


Figure 2.10: Solubilization of SQTKS KS/AT protein obtained from *E. coli* BL21 DE3 using urea at different concentrations.

Since the urea could not be eliminated using dialysis due to precipitations, it was tested to separate the urea *via* manual Ni-NTA-agarose-column. The SQTKS KS/AT protein which was solubilized with 6 M urea was incubated for 1 h with the Ni-NTA-agarose beads to bind the protein *via* the His-tag. The beads were washed with buffers containing a decreasing concentration of urea from 6 M, over 4 M, to 2 M and finally urea free buffer (section 10.4.7.2). The denaturant can be removed with this method while the protein is bound to the Ni-matrix. After the washing steps an elution buffer was used containing 500 mM imidazole to elute the protein from the column. Soluble protein was obtained (Figure 2.11). Unfortunately, the protein did not bind to the column, as the Ni-NTA-void (was obtained after incubating the solubilized protein with the Ni-NTA beads) still contains the protein. The remainder of the soluble unbound SQTKS KS/AT protein was washed away in the first washing step. No protein was detected in the elution fractions.

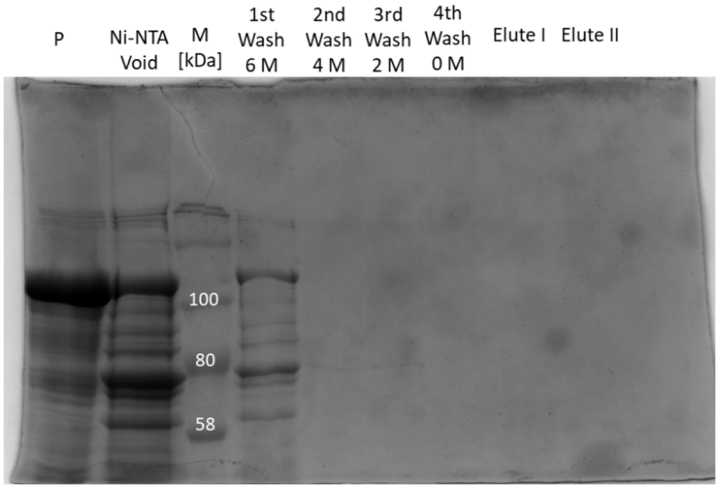


Figure 2.11: Refolding approach using manual Ni-NTA agarose column with four washing steps (decreasing urea concentration) and elution with 500 mM imidazole.

As the protein-containing fractions are still contaminated with 6 M urea and β-mercaptoethanol which need to be removed, the fractions were transferred into a 30 K Amicon Ultra-15 Centrifugal Filter device. The mixture was washed five times with urea free buffer and centrifuged subsequently. After the second washing step the protein started to precipitate. A sample was collected and analyzed by SDS-PAGE.

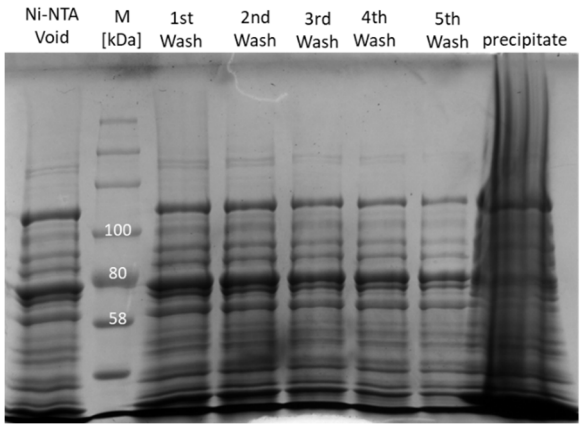


Figure 2.12: Removal of the urea by a centrifugal filter device after failed binding to Ni-NTA column.

The appearance of the precipitate correlates with the desired protein size of the SQTKS KS/AT didomain (Figure 2.12). These experiments show, that the protein starts to fold incorrect and forms inclusion bodies during removal of urea and β-mercaptoethanol.

2.3.5 Truncated SQTKS KS/AT Constructs

In previous work David Ivison tried to express the SQTKS KS/AT didomain for *in vitro* analysis from an unoptimized fungal sequence. Two constructs of SQTKS KS/AT were created and expressed. The boundaries of the first construct lay between amino acids M1 and V940, and this protein was found to be completely insoluble. The boundaries of the second construct tried by David Ivison were at A45 and V940. This construct gave minor amounts of soluble protein but not enough for *in vitro* analysis.[96] This may reflect the bioinformatics prediction of section 2.3.1 in which the N-terminus of the SQTKS KS/AT may be unstructured, and the cause of protein misfolding.

As already mentioned the current gene was synthetic and obtained in a codon optimized sequence. This results in very good protein production (section 2.3.2) but does not ensure correct folding. Therefore, different domain boundaries were selected with the aim of removing potentially unstructured regions at the N- and C-termini. Three forward and three reverse primers were chosen based on the peptide sequence alignment with fatty

acid synthase (FAS) of pig and mouse (Figure 2.13). David Ivison's work has shown that the starting point could have an influence on the solubility of the KS/AT didomain. The first 30 amino acids have low sequence similarity to both fatty acid synthases (rat and pig) but could still play an important role for the protein folding.

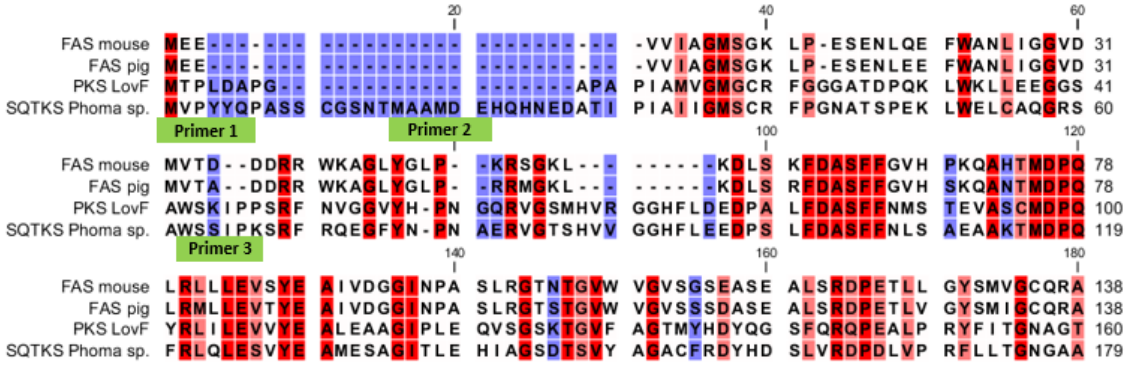


Figure 2.13: Protein alignment with pig FAS(UniProt A5YV76), mouse FAS(UniProt P19096), LovF PKS (UniProt: Q9Y7D5) and SQTKS in the N-terminal region including the forward primer positions.

Forward primer 1 starts directly at the N-terminus of the gene sequence. Forward primers 2 and 3 exclude the first residues with low sequence similarity (Figure 2.13). The starting points are at M1, M16 and W62 (Figure 2.13, Primer 1-3). All forward primers include *Nde*I restriction sites. The choice of the endpoint of the KS/AT didomain construct should have a higher effect on the solubility and activity of the resulting protein. The last 100 residues include important areas in which the endpoint of the AT domain, the linker domain and the beginning of the DH domain are located. mFAS which has a high sequence similarity to the SQTKS, has a linker domain, lying between the AT and the DH domain, consisting of a region with several hydrophilic residues, visible in the model coloured in magenta and red (Figure 2.14) and also in the alignment marked with a red box (Figure 2.15).

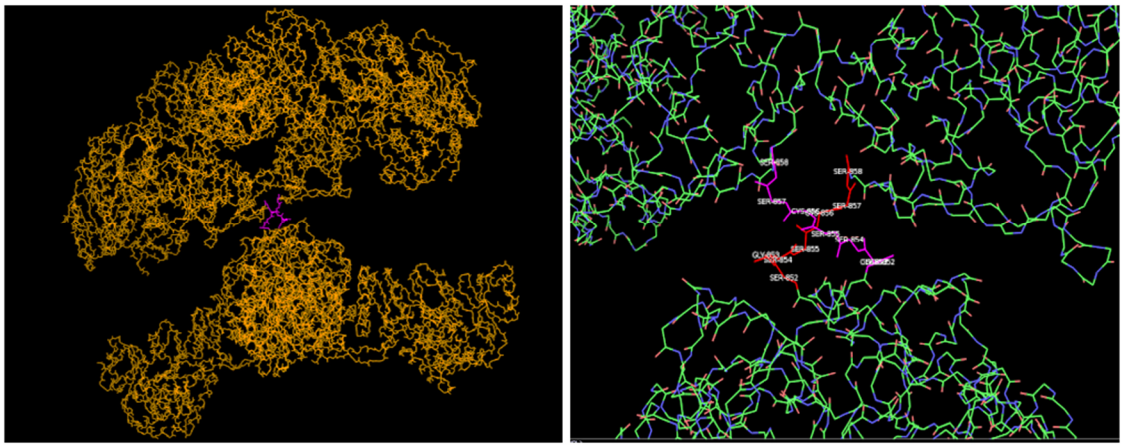


Figure 2.14: Structure of mFAS and the linker region between AT and DH domain.

This particular region, primer binding properties and also highly conserved regions of the alignment with fatty acid synthases and the PKS lovF were considered for the choice of the reverse primers (Figure 2.15). All reverse primers include stop codons, *Xho*I restriction sites and are positioned at M910 (primer 4), L927 (primer 5), A941 (primer 6) and G964 (primer7).

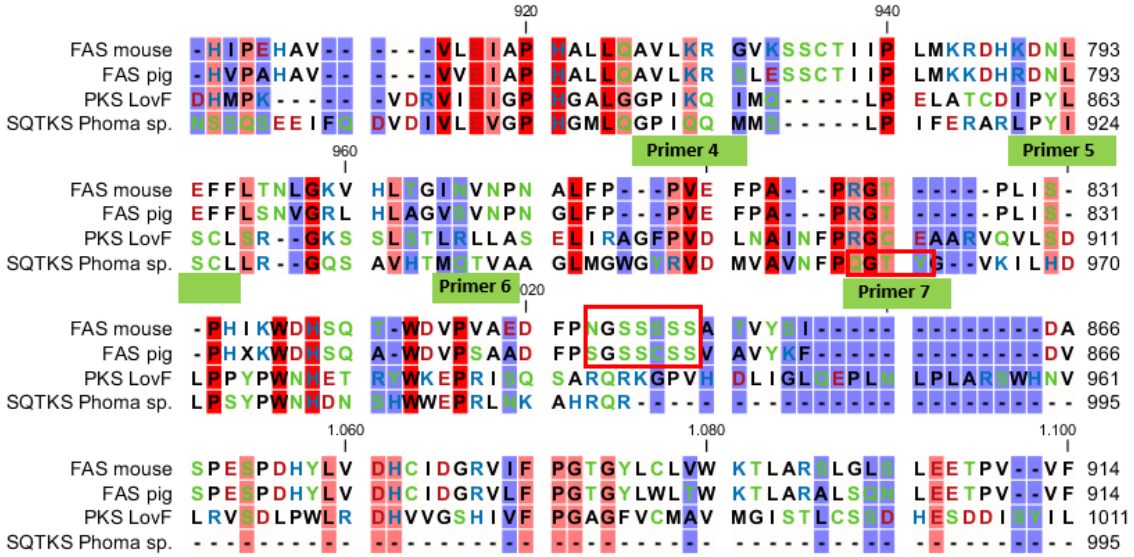


Figure 2.15: Protein alignment of SQTKS (region AT-DH) with mFAS of pig and mouse and LovF PKS. Hydrophilic regions are highlighted with red boxes.

Subsequently to the primer design several constructs were created using different combinations of primers. The chosen combinations are presented in table 2.3 with increasing length.

Table 2.3: Summarized SQTKS KS/AT constructs including length and primer combinations.

Construct	*Primer pair*	*Length [bp]*
0	Primer 3 /Primer 6	2640
1	Primer 2 / Primer 4	2685
2	Primer 2 / Primer 5	2736
3	Primer 1 / Primer 5	2781
4	Primer 2 / Primer 7	2847
5	Primer 1 / Primer 7	2992

The SQTKS KS/AT sequence (M1 – H1000) described in section 2.3.2 was used as the template for the PCR. Prior cloning the annealing temperatures were optimized for all primers. In the first step an amplification of the genes with the appropriate primers (Table 2.3) was performed by a PCR using Q5 polymerase. The resulting genes were cut by *Nde*I and *Xho*I in a double digest approach, causing sticky ended products. The pET28a (+)

was treated in the same way. The genes were ligated by a T4 DNA ligase. Constructs 0, 1, 2 and 5 were cloned succesfully and were transformed into *E. coli* Top10 and BL21. The cloning for constructs 3 and 4 were not successful.

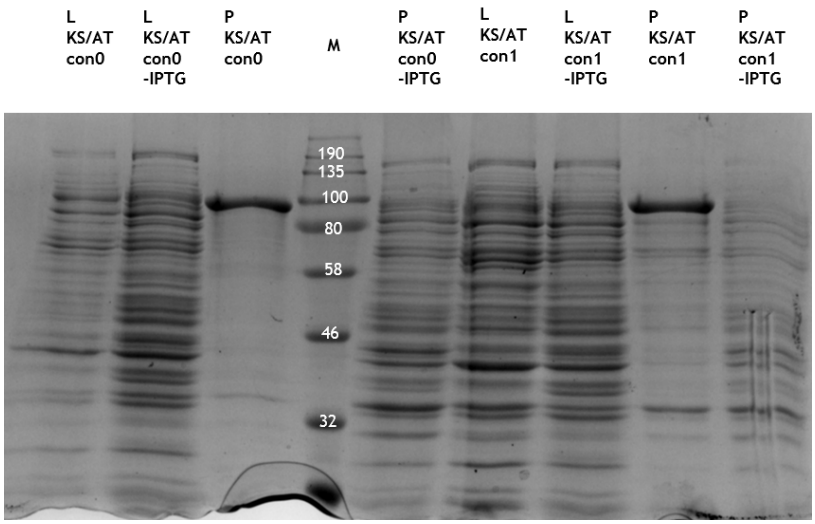


Figure 2.16: Expression of SQTKS KS/AT constructs 0 and 1 in *E. coli* BL21 DE3 for 22 h at 16 °C with soluble and insoluble parts and negative controls (-IPTG).

All successfully cloned constructs were expressed in *E. coli* BL21 for 22 h at 16 °C and 1 mM IPTG. The soluble and insoluble fractions were collected for SDS-PAGE analysis. Every expression was performed with a negative control excluding the IPTG induction. Both shortest constructs 0 and 1 had compared to negative controls intense bands in insoluble fractions. As expected there is no overexpression in the negative controls without the inducer IPTG. The soluble fraction of construct 1 showed a low intensity band of the expected size (98.2 kDa) compared to the control. The same was observed for soluble fraction of construct 0 (Figure 2.16).

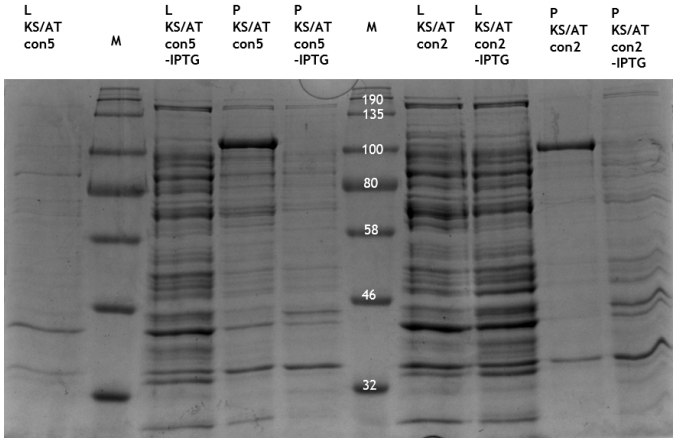


Figure 2.17: Expression of SQTKS KS/AT constructs 0 and 1 in *E. coli* BL21 DE3 for 22 h at 16 °C with soluble and insoluble parts and negative controls (-IPTG).

Constructs 2 and 5 were expressed in the same way as the previous ones. In both cases the desired protein was obtained in the insoluble fraction (Figure 2.17). All in all, gave construct 0 the best result in case of solubility and a first purification was set up. A culture of 500 ml was induced with 1 mM IPTG at 16 °C over night. After purification on the FPLC using a Ni-NTA and a size exclusion column the concentrated protein was and analyzed by SDS-PAGE (Figure 2.18).

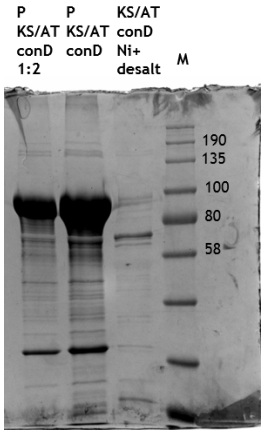


Figure 2.18: Purification of SQTKS KS/AT construct 0 by Ni-NTA and size exclusion column showing the insoluble fraction (P) and the purified protein (Ni+desalt).

The largest part of the protein is still in the insoluble fraction. There is only a tiny double band at the expected size and a band around 70 kDa which may be related to the SQTKS KS/AT construct. Unfortunately, the amount of soluble protein is not enough for *in vitro* experiments.

2.3.6 Variation of Media and use of a Solubility Tag

Growth of *E. coli* in a different medium and the subsequent expression can have an influence on the solubility of the protein. One example for improved solubility by media additions is the recombinant expression in *E. coli* Bl21 DE3 of bovine sex determining region Y protein. The addition of arginine and sorbitol increased the solubility of the protein.[116]

The expression of KS/AT construct 0 was performed in two different media. For testing if there is a difference in solubility a rich medium (TB) was used for expression and analyzed together with the standard medium (2TY) on the SDS gel. No significant difference could be observed and the main part of KS/AT stayed insoluble (Figure 2.19).

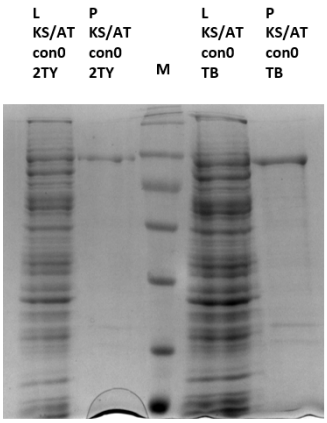


Figure 2.19: Expression of SQTKS KS/AT construct 0 in E. coli BL21 DE3; culturing in 2TY and TB medium for 22 h at 16 °C with 1 mM IPTG

LB medium was used for an auto induction expression. During the auto induction the production of the protein begins slowly in comparison to an IPTG induction. The production starts when the glucose concentration drops (which represses the lac operon) to a minimum and lactose (which induces the lac operon) has to be used as the energy source.

An influence on the solubility of a protein by changing from IPTG induction to lactose induction was achieved by the group of Kenealy.[117] They showed an increased solubility of the *Ricinus communis* stearoyl acyl carrier protein desaturase during a fed batch cultivation using lactose as the inducer. The variation of induction for the SQTKS KS/AT did not give any improvements in terms of protein solubility.

Increase of solubility can also be obtained with the help of specific tags. The SUMO fusion tag combines an improved solubility and a possibility to purify with Ni-NTA column. SUMO proteins are post-translational modifications in eukaryotic cells and play an important role in various cellular processes. This tag enhances the protein production in prokaryotic and eukaryotic cells and improves protein stability and solubility. The SUMO tag can be cleaved with a specific protease after the purification.[118] A study compared the SUMO tag to commonly used tags considering proteins which are difficult to express due to solubility issues.[119] The SUMO tag was compared to the maltose-binding protein (MBP), the glutathione S-transferase (GST), thioredoxin (TRX), NUS A and ubiquitin (Ub). SUMO and NUS A fusion tags enhanced the expression and solubility of recombinant proteins in *E. coli* most dramatically.

*Age*I and *Xho*I (both restriction sites introduced by primers) were used for cloning the SQTKS KS/AT gene (M1 – A941) into the pETM11-SUMO3GFP vector

with a restriction ligation approach. The construct includes a His-tag and SUMO tag at the N-terminus when expressed (aminoacid sequence in the appendix). The expression of the SUMO tagged protein led in 2TY and in TB medium to an insoluble protein of the expected size (113 kDa) (Figure 2.20).

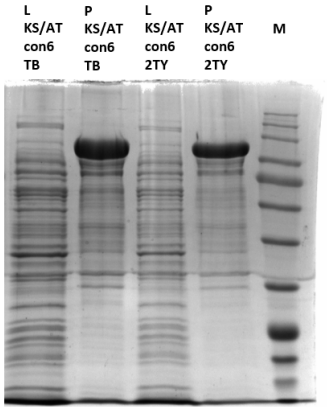


Figure 2.20: Expression of SUMO tagged SQTKS KS/AT in *E. coli* BL21 DE3 for 22 h at 16 °C. with soluble (L) and insoluble (P) parts.

2.4 Discussion

Escherichia coli is the simplest organism for protein expression with advantages of speed, low cost and ease of use. Even if it is the most convenient used organism, less than 15% of nonbacterial proteins can be solubly expressed in *E. coli*.[120] Another point is that proteins become more difficult to express in a soluble form as their molecular weight increases.[121] The KS/AT didomain is a large protein of about 1000 amino acids and derived from a fungus. The *E. coli* codon optimization may have an influence on the yield of the obtained protein but can have a negative effect on the correct folding and solubility during expression. Codon optimization leads to a faster expression causing a higher yield but also can bring misfoldings. Cloning of different constructs varying in length were successful and also the expression showed that the KS/AT protein was overexpressed. Different *E. coli* strains were tested, culture and expression conditions were modified, different culture media were tested, solubility tag (SUMO) was used, but all the attempts led to an insoluble protein. In this particular case the bioinformatics predictions proved to be correct and confirmed that the probability to obtain KS/AT soluble in *E. coli* cells is unrealistic. Creating different constructs with a variation of starting and endpoints gave only a minor improvement in terms of solubility but there is still the main part insoluble. The choice of the correct domain boundaries is very difficult and hard to predict. Only a few amino acids more or less can have a tremendous effect on the correct folding and in

consequence an increased solubility. Expression of an AT domain in an iterative type I PKS DynE8 (section 1.5.2) had shown that the linker regions have a high influence on the solubility of the expressed protein. In this case only three of twelve constructs led to a soluble expression.[76] Only a few residues difference can play an important role. Creating this great number of constructs shows that the methods for predicting domain boundaries are very limited. Also the position of the tag, N-terminal or C-terminal, can play a significant role and can be tested in a future project.

An access to a crystal structure of the SQTKS, which is still awaited, would simplify this process. For the future work, yeast or other fungal systems should be used to express KS/AT. The expression in these organisms should be more similar to the one of *Phoma* sp., KS/AT originated from. The main problem will be the decreased concentration of recombinant protein. The best expression conditions should be worked out and a large scale expression will be necessary to generate sufficient recombinant protein for *in vitro* experiments.

3. SQTKS and *B. fulva* PKS Acyl Carrier Proteins

3.1 Introduction

The Acyl Carrier Protein has a key role during polyketide biosynthesis. The small protein binds substrates and intermediates and transfers them to the extension and β-processing domains. Obtaining ACPs of two different pathways (SQTKS and Byssochlamic acid pathway) in an active form would make the *in vitro* analysis of the interactions with other already expressed and active catalytic domains possible.

The first expression trials of SQTKS ACP were done by David Ivison.[96] He tried to express a standalone ACP, as well as multidomain constructs including ACP. SQTKS ACP as a single stand-alone domain could not be expressed successfully. Two multidomain constructs including ACP (KR-ACP and ER-KR-ACP) were expressed, but protein was observed in low concentrations and no attempt was made to phosphopantetheinylate the ACP domain. No *in vitro* experiments were done with these constructs due to the low level expression. For *B. fulva* PKS (bfPKS) ACP no expression trials have been reported.

One successful example of an ACP expression and a subsequent characterization was reported by the group of Crump.[122] They solved the first structure of a fungal type I non-reducing PKS ACP. The structure of norsolorinic acid (NSAS) ACP was elucidated by NMR showing a four-helix bundle with an unusual hydrophobic N-terminus of helix III. Helix III is a shorter helix compared to the other three helices and lies perpendicular to helices II and IV. Compared to type II FAS and PKS ACPs, helix III of NSAS ACP has more bulky and hydrophobic amino acids. This hydrophobicity may cause inflexibility in Helix III compared to type II FAS and PKS ACPs. HSQC experiments with hexanoyl-NSAS ACP showed only minor structural changes compared to the *holo* species and strengthens the assumption, together with the mammalian type I FAS ACP, that the type I PKS ACPs do not show significant affinity for hydrophobic chain assembly intermediates. The NSAS ACP was expressed in *E. coli* BL21 (DE3) cells using LB medium and an IPTG concentration of 0.1 - 1 mM. The induction was performed for 3 - 4 h at 25 - 37 °C. The expressed protein had an N-terminal His-tag and and S-tag. The correctly folded state of the protein was checked by *in vitro* phosphopantetheinylation reaction using *S. coelicolor* ACPS and Coenzyme A.[123]

The Crump group has also demonstrated another example of expression and characterization of a functional ACP. The ACP from the type II iterative actinorhodin PKS, was expressed in *E. coli* K38 pGP1-2. Protein production was induced by heating to 42 °C. Expression was performed at 30 °C for 2 h in LB medium. The purification of the tag-free ACP was done by initial ammonium sulphate precipitation followed by desalting and Q-Sepharose anion exchange chromatography. The correct folding of the protein was checked *in vivo* by co-expression of ACP with ACPS which gave mainly *holo* ACP.[124] NMR structures for both *apo* and *holo* forms were solved[125] as well as seven thiolester and thiol-ether derivatives to gain insight into ACP intermediate interactions.[126] Interactions of ACPs with other modifying domains are described in section 1.6.

3.2 Aims

The main aim for this part of the project is to obtain soluble and functional ACP. To achieve this goal a successful cloning followed by an expression in *E. coli* will be attempted. ACPs are required from the SQTKS and Byssochlamic acid PKS (bfPKS) for *in vitro* assays with other components of the PKS. First of all, the borders of the Type I ACPs (SQTKS and bfPKS) should be determined using bioinformatic methods and comparisons to other successfully expressed ACP proteins. It is necessary to determine the likely boundaries of the ACP for two reasons: first, different N-terminal and C-terminal positions may affect solubility; and second, the structural integrity of the domain should be achieved in order to gain reliable and meaningful results. A combination of bioinformatic techniques and trial and error was used for designing constructs for heterologous expression.

Next, genes will be cloned into suitable *E. coli* expression vectors, potentially carrying tag sequences. Protein production will be performed in *E. coli* using previously described methods. Since *holo*-protein will be required for *in vitro* assays, *holo*-ACP will be produced *in vitro* using the PPTase reaction (section 3.3.5). The observation of mass differences between *apo* and *holo* ACP will be achieved by mass spectrometry (MS). Suitable MS methods will have to be established first. Development of protein-MS assays will allow observation of *apo* to *holo* conversions, as well as substrate and linker loading reactions.

3.3 Results

3.3.1 Determination of ACP Boundaries

Initially, the sequences of the SQTKS ACP region (A2497 - A2603) and bfPKS ACP region (S2511 – E2618) were aligned with structurally well characterized ACPs such as *Vibrio harveyi* ACP (PDB: 2L0Q), Erythronolide synthase ACP (PDB: 2JU1) and *S. coelicolor* actinorhodin ACP (PDB: 2K0Y, Figure 3.1 **A**). This gave a first overview for highly conserved regions. Most significant is the highly conserved DSL motif which can be found in all ACPs. The serine of this motif is the attachment point for the essential phosphopantetheine.

In the second step a fully automated protein structure homology-modelling server (SWISS-MODEL) was used for the approximation to the "real" domain borders of the SQTKS ACP.[127] S2491 was chosen as the N-terminal end of the analysed sequence to ensure that the ACP starting point is included. A2603, which is the C-terminus of the whole PKS sequence was selected as the C-terminus for the ACP modelling experiments. With this in hand, SWISS-MODEL searched for most similar experimental protein structures (templates). The best template was the ACP from module 2 of the erythronolide synthase (PDB: 2JU1) with the highest sequence identity of 24% (Figure 3.1 **B**). A second template was chosen, a type I modular polyketide synthase (PDB: 5HVC) from module 9 of the mycolactone PKS (23%). Two three-dimensional protein structure models were calculated by the modelling server using one template for each model and the SQTKS amino acid sequence (S2491 – A2603). SWISS-MODEL predicted the SQTKS ACP boundaries at A2504 and K2592. Primer positions of forward and reverse primers (Primer-1 at A2497, Primer-2 at M2510, Primer-3 at S2593 and Primer-4 at V2602) were chosen close but not exactly at the border prediction of SWISS-MODEL due to better primer binding properties (Melting temperature, GC content, primer length, avoidance of dimerization *etc*).

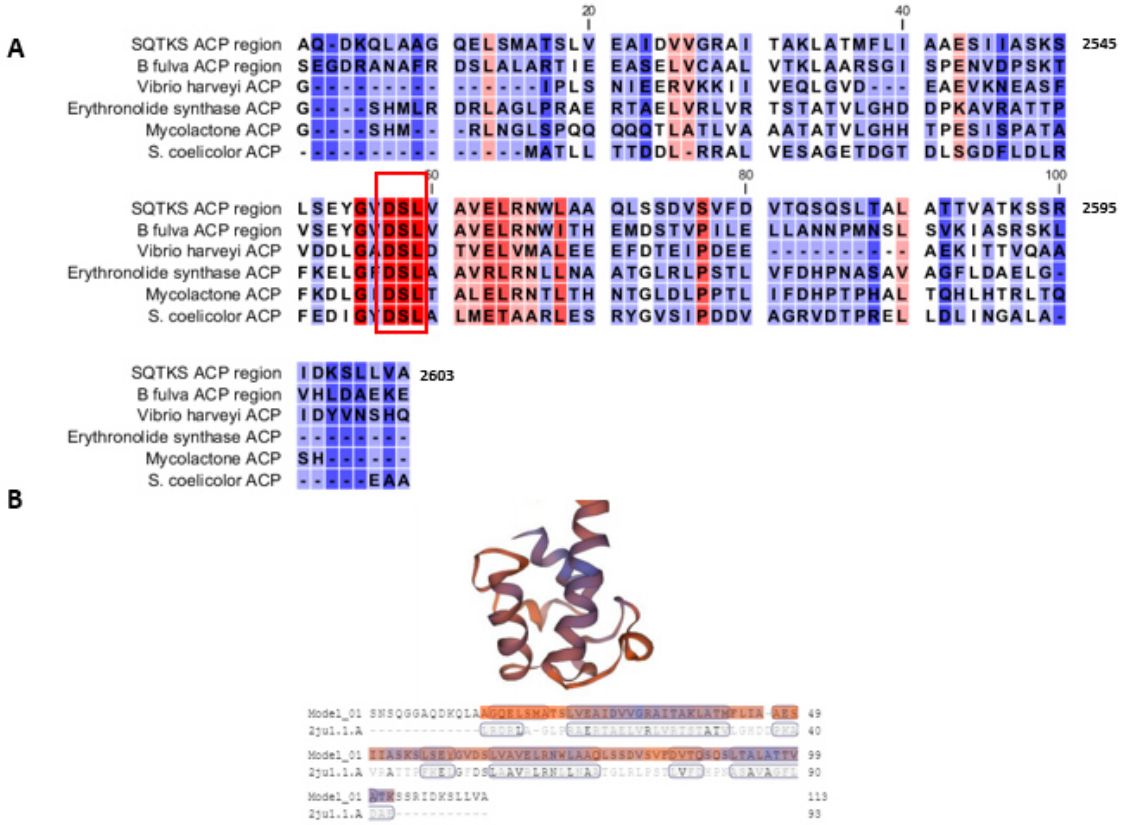


Figure 3.1: Border prediction for SQTKS ACP: **A**, Protein alignment of SQTKS ACP region with bfPKS ACP region, *Vibrio harveyi* ACP (PDB: 2L0Q), Erythronolide Synthase ACP (PDB: 2JU1), mycolactone ACP (PDB: 5HVC) and *S. coelicolor* ACP (PDB: 2K0Y) with highly conserved regions (red), less conserved regions (light red) and unconserved regions (blue). The characteristic ACP DSL region is marked with a red box. **B**, SWISS-MODEL of SQTKS ACP with Erythronolide synthase as template and the corresponding alignment.

The border prediction for bfPKS ACP was done similarly (Figure 3.1 **A**) but including more sequence either side of the core DSL motif (S2511 – E2618). With this strategy, shortening (N- and/or C-terminally) of the sequence would be later possible by cloning, if necessary. The N-terminal border for bfPKS ACP was set at S2511 and the C-terminal end at E2618 which is the end of the PKS sequence.

Protein-Sol was then used to analyse the likely solubility of the constructs in the same way as for SQTKS KS/AT (section 2.3.1). In addition to the solubility value, a prediction for folded and unfolded regions of a protein can be performed. This might give a hint why an in *E. coli* expressed protein is inactive or insoluble. Actinorhodin ACP was chosen to compare to current ACPs as it was often used as a model system for *in vitro* PKS studies and is already known to be highly soluble and properly folded.[126,128,129] The sequences of

Actinorhodin ACP (PDB: 2K0Y), bfPKS ACP (S2511 - E2618) and SQTKS ACP (A2497 - A2603) were analyzed one by one by Protein-Sol. Figure 3.2 displays fold propensity per residue calculated over a sliding, overlapping 21 amino acid window. The dark yellow and blue colours represent the positive and negative fold propensity scores per amino acid.

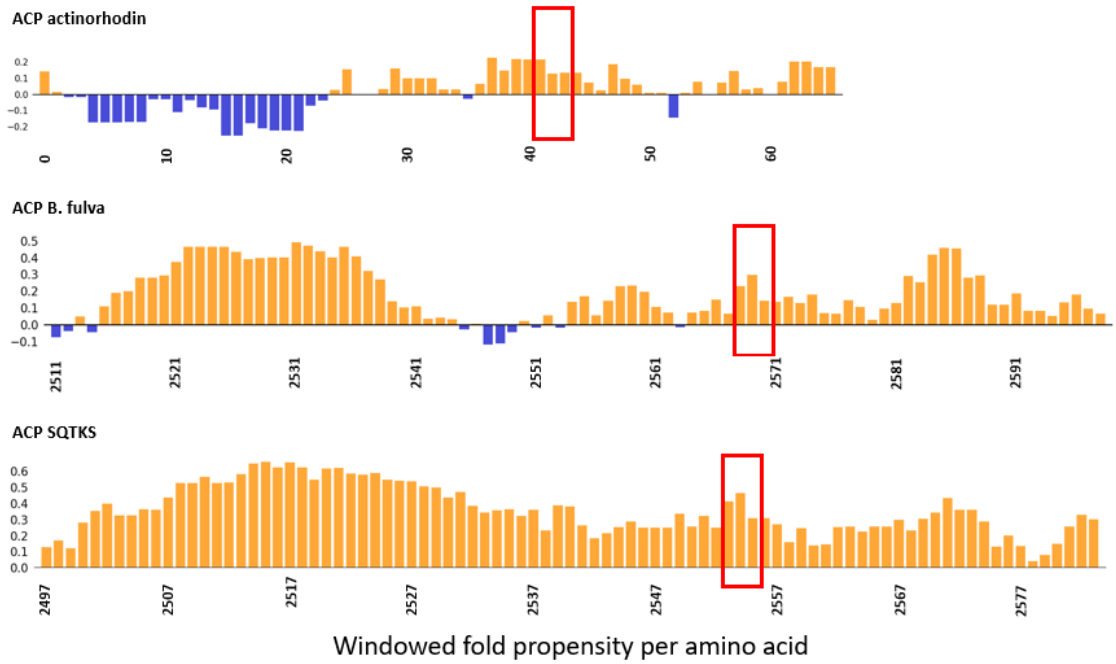


Figure 3.2: Windowed fold propensity per amino acid illustrated by Protein-Sol. ACPs of actinorhodin, SQTKS and *B. fulva* were used for the analysis. The characteristic DSL region is marked with a red box.

Actinorhodin ACP shows a predicted unfolded region (blue) in the first 25 amino acids followed by a folded region. This may reflect the known high flexibility of this Type II ACP. bfPKS ACP shows in the N-terminus a short unfolded region which might be a part of the linker region between KR and ACP. A second unfolded region lies around amino acid 2550. The SQTKS ACP is different from the other two proteins. This protein is predicted to be highly folded in each segment. No regions with a negative fold propensity can be found.

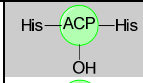
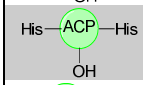
3.3.2 Cloning and expression of SQTKS and bfPKS ACP

An *E. coli* codon optimized SQTKS full PKS gene sequence was used as a template for amplification of the ACP gene. This was provided by colleague Hao Yao. Four different constructs varying in primer positions and tag pattern were designed as listed in table 3.1. In the first step an amplification of the gene for construct 3A (Con3A, A2497 - S2593)

was performed by a PCR using Q5 polymerase. Then the C-terminal end of the amplified gene was cut by *Nde*I (restriction site was introduced by forward primer) and the N-terminal end cut by *Xho*I (restriction site was introduced by reverse primer) in a double digest approach causing a sticky ended product. The cut and purified (NucleoSpin Gel and PCR Clean-up Kit) gene was ligated into a pET28a (+) which was treated before with the same restriction enzymes. For the ligation T4 DNA ligase was used. In the next step, the ligation product was transformed into *E. coli* Top10 cells to amplify Con3A. Cells which incorporated the successfully cloned plasmid survived on a LB agar plate containing Kanamycin. Colony PCR was performed to double check the successful cloning. In the last step Con3A was transformed into *E. coli* BL21 cells for first expression experiments.

Con3B and Con1 were cloned with the same methods as Con3A with the difference of not using a stop codon at the C-terminal end of the reverse primer leading to a double His-tagged protein. For Con1 also different primer combinations were used (M2510 - S2593).

Table 3.1: Summarized SQTKS ACP constructs, their primer positions, sizes, symbols and the required primer pairs.

Name	Symbol with Tag pattern	Start Position	Stop Position	Size in kDa	Primer pair
Con1		M2510	S2593	12.1	Primer 2 Primer 3
Con3A		A2497	S2593	12.4	Primer 1 Primer 3
Con3B		A2497	S2593	13.4	Primer 1 Primer 3
Con3C		A2497	S2593	11.6	Primer 1 Primer 3

Con3C was cloned in the same way as Con3B except with the difference of cutting the N-terminal His tag post expression as described later in this section.

For bfPKS ACP no cloning was necessary. A synthetic gene (S2511 – E2618), inserted into a pET28a (+) vector was ordered at the company Baseclear. The sequence was *E. coli* codon optimized by the company. A stop codon at the end of the gene and two restriction sites (*Nde*I, *Xho*I) were included. The protein had an N-terminal His tag when expressed. Transformations into *E. coli* Top10 and BL21 cells were performed as described for SQTKS ACP Con3A.

The first expression tests were performed at small scale with Con3A (A2497 - S2593). The expression was done in *E. coli* BL21, using 2TY medium, shaking at 200 rpm and a temperature of 37 °C until an optical density (OD_{600}) of 0.6 was reached. In the next step the temperature was lowered to 16 °C and the expression was induced with 1 mM IPTG. Negative controls were carried out in parallel with the same conditions with the single difference of no IPTG addition. No significant protein band could be observed at the expected size after different time points (1, 5, 22 h) of expression (Figure 3.3). An expression test at 37 °C also gave no soluble protein at expected size.

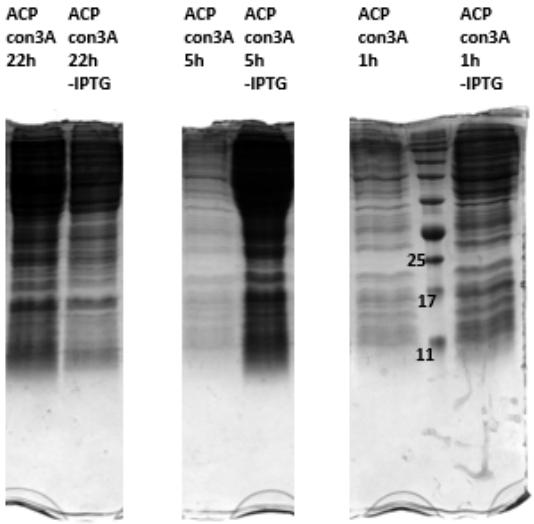
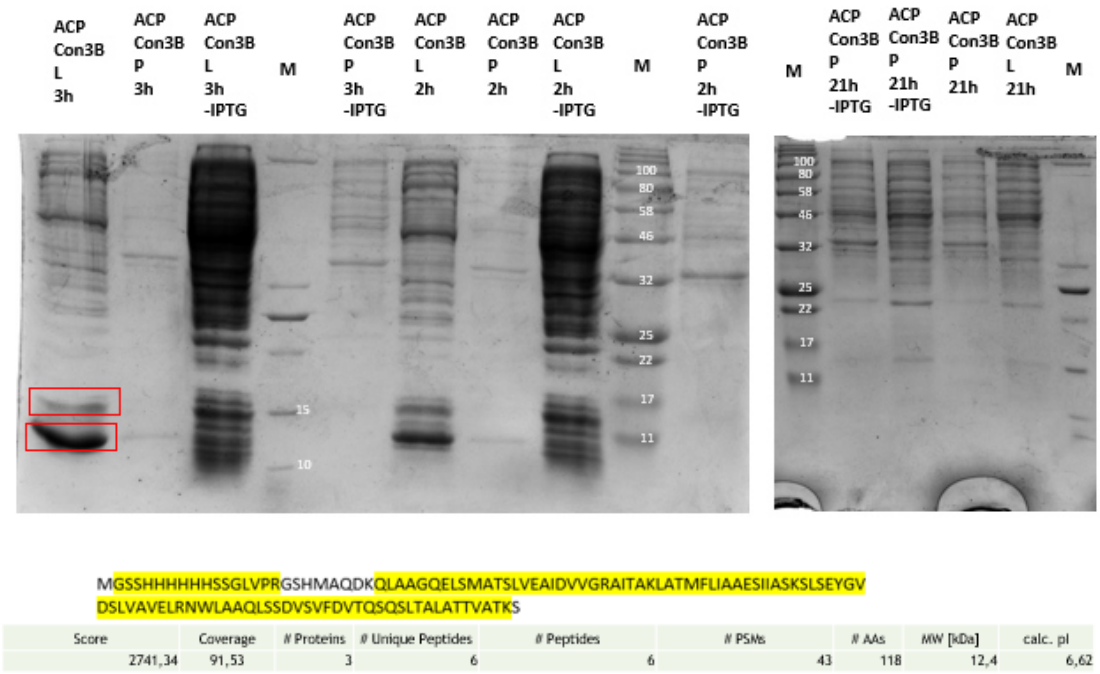


Figure 3.3: Expression of Con3A SQTKS ACP induced with 1 mM IPTG at 16 °C in 2TY medium. Soluble fractions of different induction times and negative controls without inducer (-IPTG) are shown.

Due to poor results with Con3A the SQTKS ACP protein was expressed with two His-tags (C-terminal and N-terminal). Con3B was expressed at 37 °C for 2, 3 and 21 h. At 37 °C (2 and 3 h induction) an intense band was observed at the expected size of 13.4 kDa in the soluble fraction (Figure 3.4). After 21 h of induction no soluble protein of the expected size was observed and the protein might be digested. The best result was obtained at 37 °C and 3 hours' induction time. The protein sequence was analyzed by MALDI-TOF and confirmed (coverage 91%) that the intense band on the SDS-gel corresponds to the desired protein.



MGSSHHHHHHSSGLVPRGSHMAQDKQLAAGQELSMATSLVEAIDVVGRAITAKLATMFLIAAESIIASKSLSEYGV
DSLVAVELRNWLAAQLSSDVSVFDVTQSQSLTALATTVATKS

Score	Coverage	# Proteins	# Unique Peptides	# Peptides	# PSMs	# AAs	MW [kDa]	calc. pI
2741,34	91,53	3	6	6	43	118	12,4	6,62

Figure 3.4: Expression of SQTKS ACP Con3B, induced with 1 mM IPTG at 37 °C for 2 h, 3 h and 21 h in 2TY medium with soluble (L) and insoluble (P) fractions. Protein fragments which were found in MALDI-TOF are marked in yellow.

Subsequently a large-scale expression was performed using 1 L 2TY medium and the best conditions from the small scale expression. A total of 17 mg ACP (Con3B) was purified from a 1 L fermentation. To achieve this goal, cells were lysed by sonication in the first step. The soluble and insoluble fractions were separated by centrifugation. In the next step the soluble fraction was filtered and purified by a two-step purification using a Ni-NTA column and the Äkta Pure system in the first step. In the second step a size exclusion column was used to eliminate the imidazole. The purification process was monitored by SDS-PAGE.

A band can be seen on the SDS-PAGE at approx. 11 kDa which is close to the expected 13.4 kDa. ACP proteins have a higher negative charge compared to average proteins, are very compact in their structure and run normally faster than regular proteins. On an SDS gel, ACP proteins are therefore usually found in a lower molecular range than expected. The insoluble fraction, purified not concentrated protein and concentrated protein are shown on the SDS-PAGE (Figure 3.5 **A**)

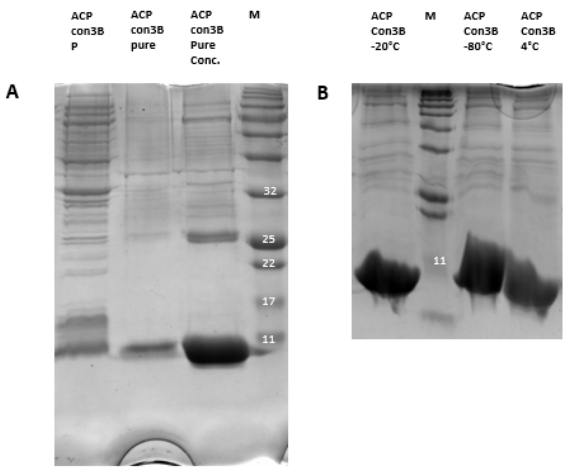


Figure 3.5: Ni-NTA + size exclusion column purified SQTKS ACP Con3B with pure (unconcentrated) protein and pure conc. (concentrated) with centrifugal concentrators: **A**, Ni-NTA+size exclusion column purified SQTKS ACP Con3B; **B**, protein stored at temperatures of 4, -20 and -80 °C for ten weeks.

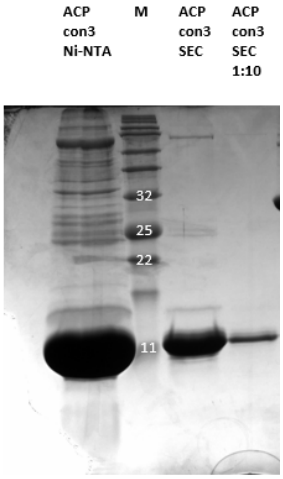


Figure 3.6: First purification step (Ni-NTA) of SQTKS ACP Con3B followed by size exclusion column 26/600 (SEC) and a 1:10 dilution.

SQTKS ACP Con3B was stored at different temperatures for ten weeks and analyzed after this time. Protein stored at -80 °C, -20 °C and 4 °C was still present in high concentration and not visibly degraded after this period of time (Figure 3.5 **B**).

Impurities, which are present in higher molecular range, even after Ni-NTA column and 26/10 desalting column purification, had to be removed. For this a size exclusion superdex column (26/600) was used (Figure 3.6). The usage of a different SEC column improved the purity of the protein compared to previous purifications.

Later in this chapter an activity problem of the SQTKS ACP Con3B is described. For this reason, another construct was tried. A small scale expression test with the double His-tagged ACP Con1 (M2510 - S2593) was performed. *E. coli* BL21 containing Con1 was grown in 2TY medium and induced with 1 mM IPTG. After 2.5 h, 5 h and 22 h of induction at 16 °C aliquots were taken, cells disrupted by sonication and the soluble and insoluble fractions were analyzed by SDS-PAGE. No significant band of the protein could be seen in the soluble fraction at the expected size of 12 kDa (Figure 3.7).

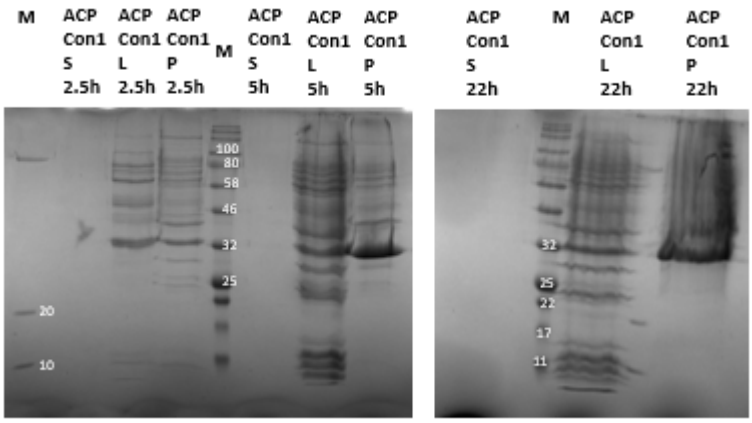


Figure 3.7: Expression of SQTKS ACP Con1 induced with 1 mM IPTG at 16 °C in 2TY medium with soluble (L) and insoluble (P) fractions and the supernatant (S) of the cell culture.

Expression of Con1 was also attempted at a temperature of 37 °C but still no soluble protein of the expected size was observed. N-terminally tagged SQTKS ACP (Con3A) could not be expressed successfully and the double tagged ACP (Con3B) was inactive (Section 3.3.5). It was assumed that the two tags prevent protein degradation but also have a negative influence on the activity of the protein. The final possible tag combination is the C-terminal His-tagged ACP (Con3C). This ACP construct was expressed in the same way as Con3B to give soluble protein. The N-terminal tag was cut with thrombin after expression. After finding the optimal conditions for the thrombin reaction the cut ACP was purified by size exclusion chromatography (Figure 3.8).

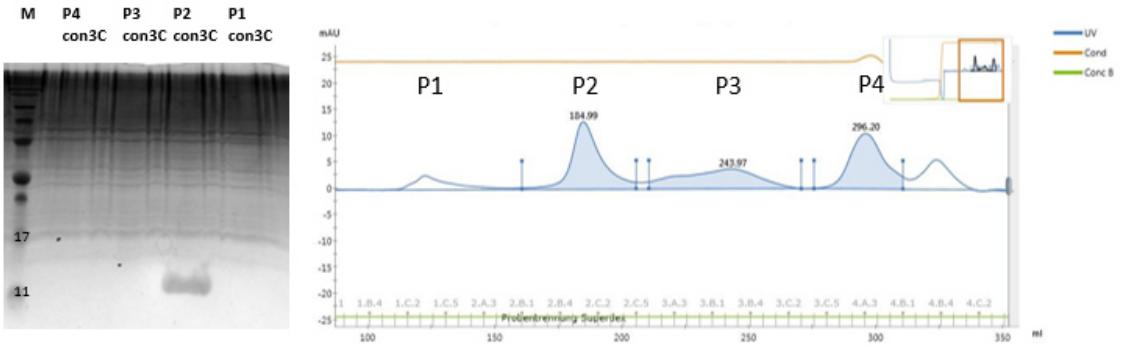


Figure 3.8: SEC purification of SQTKS ACP Con3C. Combined fractions of the UV peaks of the SEC run (P1-P4) were analyzed by SDS-PAGE showing ACP in P2.

To summarize, the SQTKS ACP expression trials included four different constructs. Con3A + Con1 could not be expressed successfully, but Con3B + Con3C were expressed and purified successfully.

Compared to the SQTKS ACP, the expression of bfPKS ACP was unproblematic. The pET28a (+) vector including the desired gene (*Nde*I and *Xho*I restriction sites included) was synthesized by Baseclear (*E. coli* codon optimized) and transformed into *E. coli* Top10 and BL21 (DE3) cells. First, small scale expression was performed at 37 °C for 2 and 4 h in 2TY and TB medium with 1 mM IPTG. Cell lysis (sonication) was followed by separation of the soluble and insoluble parts of the culture. SDS-PAGE analysis showed most of the protein in the insoluble fraction (Figure 3.9).

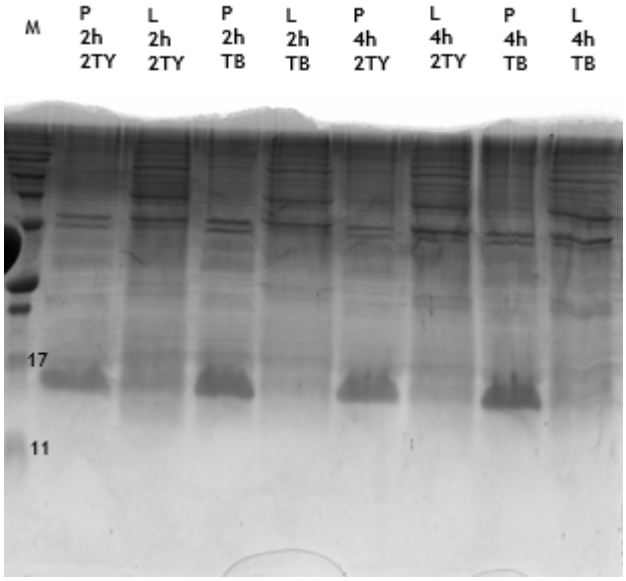


Figure 3.9: bfPKS ACP expression at 37 °C in two different media (2TY and TB) with insoluble (P) and soluble fractions (L).

Next, expression attempt was executed with reduced temperature to slow down the expression rate. The temperature was set to 16 °C for 5 h and 22 h of induction time. For the induction 0.1 mM and 1 mM IPTG were used. The reduced temperature gave an improvement in protein solubility. The best conditions for expression were found to be 22 h of induction time and 1 mM IPTG. Under these conditions a band in the soluble fraction close to expected size of the protein was observed (14 kDa, Figure 3.10 **A**).

In the next step a large scale expression followed by a purification was set up using the best expression conditions mentioned before. 800 mL culture was lysed after expression. The soluble part was purified by Ni-NTA column (Figure 3.10 **B**) using an Äkta Pure FPLC. Fractions of the Ni-NTA run containing the desired protein were combined and used for the size exclusion chromatography (Figure 3.10 **C**). Combined fractions of the SEC run were concentrated and analysed by SDS-PAGE. Fractions 1A4

to 1B1 contained the expected protein (Figure 3.10 **D**). The purified bfPKS ACP protein could be stored at 4 °C and also at -20 °C without precipitations. -20 °C samples were used for more than 6 months without loss of activity (Figure 3.10 **E**).

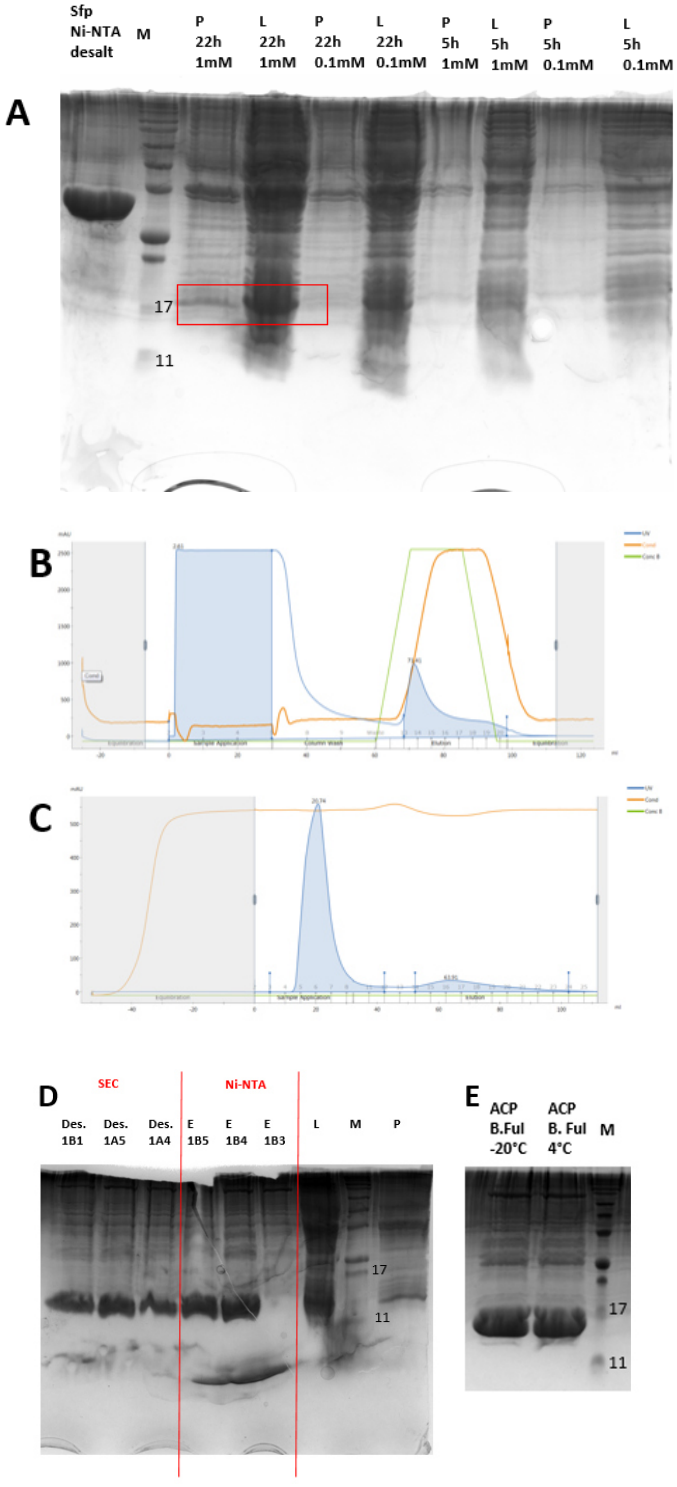


Figure 3.10: bfPKS ACP expression and purification attempts: **A**, expression at 16 °C with 1 and 0.1 mM IPTG and different induction times; **B**, purification by Ni-NTA column with an Äkta Pure system; **C**, following purification by size exclusion column; **E**, bfPKS ACP after storage at 4 °C and -20 °C for several weeks.

3.3.3 Development of Mass Spectrometry Methods for ACP

Elelctrospray ionisation (ESI) mass spectrometry was available in the working group and is known as a powerful method for the analysis of intact proteins. However, the method of measuring proteins has to be established first to get accurate mass spectra with a high signal to noise ratio. Mass spectra can become complicated due to spectral overlap in a mixture of several proteins. For this reason, it is important to have spectra as clear as possible.

ESI of proteins creates a distribution of ions in high charge states. There are different factors affecting these charge states including solvent, instrumental factors and the protein itself.[130] Protein conformation plays an important role in the charge state distribution (CSD). In the denatured form of the protein more charges are exposed at the surface of the protein towards the solvent than in the native form. Additionally, removal of internal charge neutralization interactions and the minimization of the Coulombic repulsion play an important role in a denatured protein.[131,132]

Horse heart myoglobin (HHMG) has been extensively used as a calibration standard in the past for optimizing protein MS methods.[133,134] HHMG is a cheap commercially available protein with a lack of glycosylation (no heterogeneity), good solubility and a strong ionization in ESI MS. The mass of HHMG is 16.950 kDa.[134]

We made the first measurements by direct injection of a solution of HHMG (*ca* 0.1 mg/ml) in a 50/50 mix of water/acetonitrile (ACN) into the mass spectrometer at 0.2 ml/min (Figure 3.11 **B**). No chromatography column for separation was installed. Positive and negative ionization modes were tested but only the positive mode gave clear spectra and this was used for all following measurements. MaxEnt® was used to deconvolute the spectrum. Typically, a protein generates multiply-charged positive ions leading to a series of peaks known as a *mass envelope*. Each of this series of peaks is transformed to single peak on a molecular weight scale to give a hypothetical zero-charge spectrum. The areas under the peaks of the spectrum are a measure of the sum of the intensities of the peaks in the original multiply-charged data leading to quantitative relative intensity data.[135]

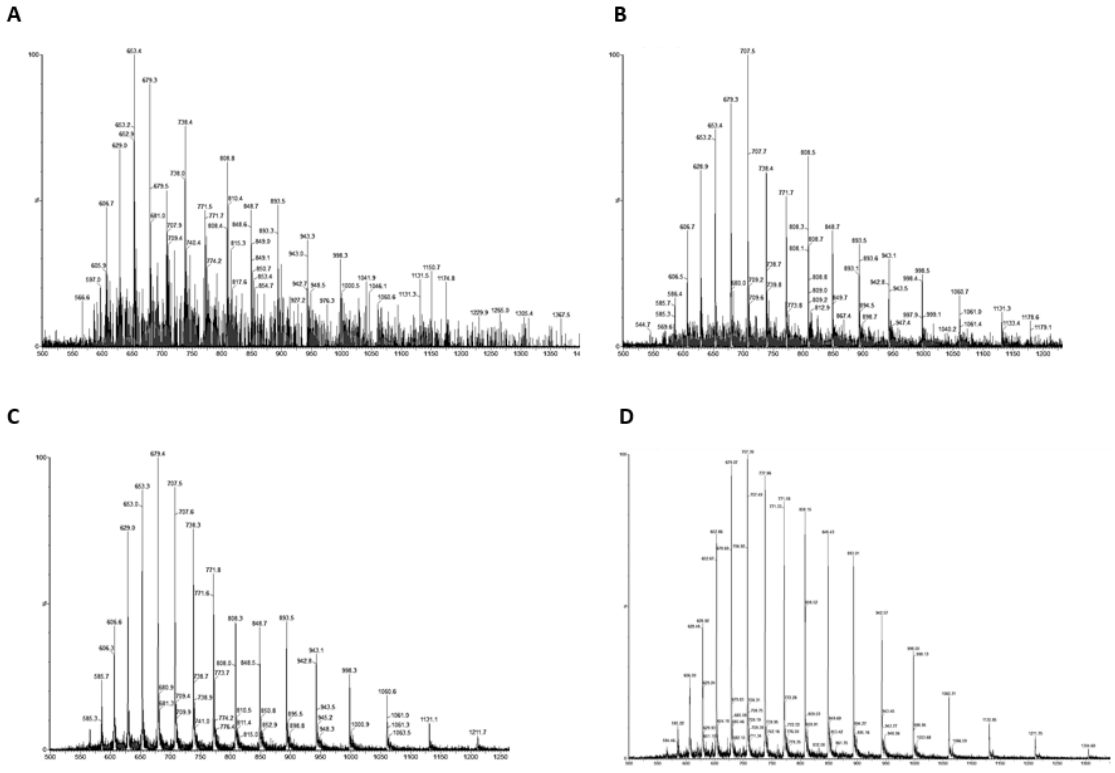


Figure 3.11: Spectrum of HHMG: **A**, running with 100% water; **B**, a mixture of 50% water and 50% acetonitrile; **C**, with 0.1% formic acid in the solvent; **D**, with 0.2% formic acid in the solvent.

An increase of acid in the solvent (from 0.1% to 0.2% formic acid) gave an additional improvement in signal to noise ratio (Figure 3.11 **C** and **D**). The acid increased the ionization of the protein, presumably by increasing protonation. No significant improvement could be observed for higher acid concentrations up to 0.5%.

After MaxEnt deconvolution, HHMG showed the expected mass of 16950 Da (Figure 3.12). These conditions (50/50 acetonitrile/water and 0.1% formic acid) were then used for most subsequent analyses.

Experiments were then conducted to determine whether injected proteins could be separated by chromatography before mass analysis. A chromatography column suitable for proteins was tested for several proteins. The column material has a pore size of 300 Å and a particle size of 5 µm, which is wide enough for larger molecules (*e.g.* proteins), to interact with. The C_5- material consists of pentyl as the functional ligand. The column is efficient for separation of hydrophobic biomolecules, such as proteins.[136]

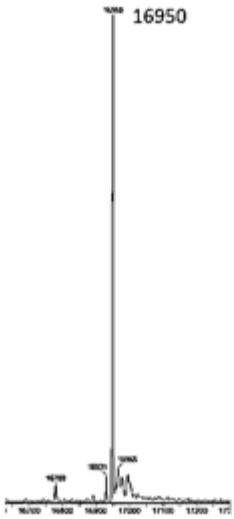


Figure 3.12: ESMS of Horse Heart Myoglobin with MaxEnt® deconvolution.

Advantages of a C_5 column are that protein samples are effectively desalted before entering the MS source which improves ionization, simplifies complex protein mixtures and clear spectra for each detected protein species.

The chromatography program was set to 30 min beginning with 10% acetonitrile (ACN) and 90% water, increasing linearly to 95% ACN after 24 min, then decreasing back to 10% ACN after 27 min. The flow rate was set to 0.5 ml/min leading directly to the MS without any split. HHMG was tested first with a concentration of 1 mg/ml and gave a sharp peak at 10 min. The corresponding MaxEnt deconvolution gave a mass of 16.953 kDa which is within the error tolerance of the calculated mass of HHMG (Figure 3.13 **A-C**).

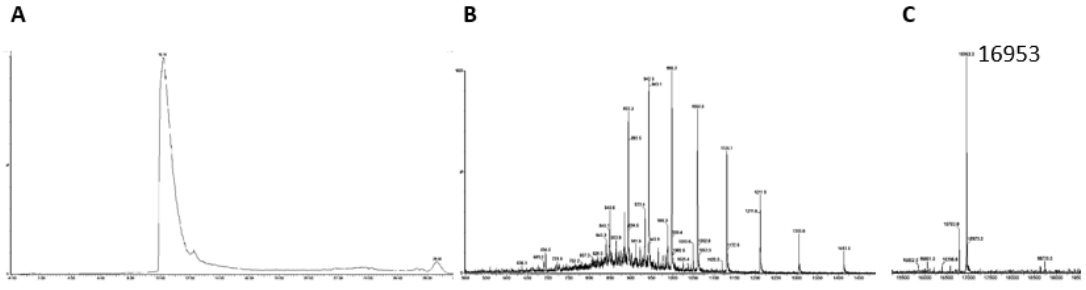


Figure 3.13: HHMG LCMS measurement using a C_5 column for protein separation: **A**, MS TIC; **B**, ES^+ spectrum; **C**, corresponding MaxEnt® deconvolution.

bfPKS ACP was analyzed with the same program settings resulting in a relatively clear spectrum and the expected mass of tag free bfPKS ACP (Figure 3.14 **A-C**). The biggest problem of this particular ACP is the strong binding to the C_5 column. Even after 6 runs of injection of water only (30 min each) there was ACP detectable in the eluent. Decreasing the ACP concentration led to a smaller peak with the same binding effect. This observation made the column for further measurements with this kind of protein impractical.

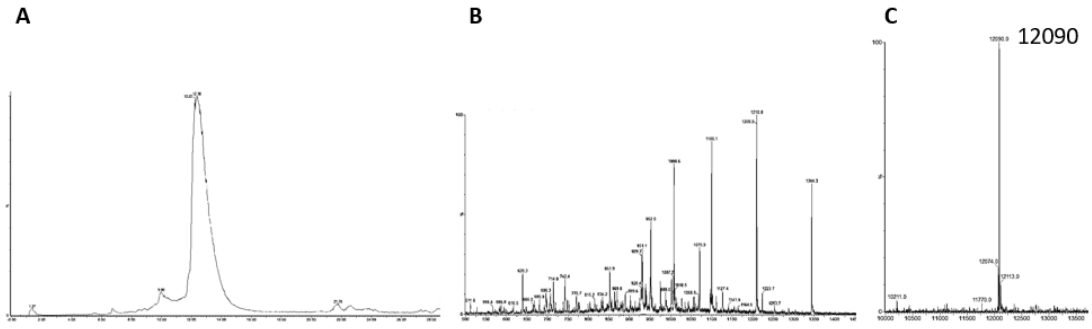


Figure 3.14: *B. fulva* LCMS measurement using a C_5 column for protein separation: **A**, MS TIC; **B**, ES^+ spectrum; **C**, corresponding MaxEnt® deconvolution.

However, use of the C_5 column was essential for a visualisation of the SQTKS *C*-methyltransferase (*C*-MeT) which was not detectable by direct injection. A measure of this protein with the C_5 column showed a clear spectrum (Figure 3.15, *C*-MeT experiments in section 6.3.2).

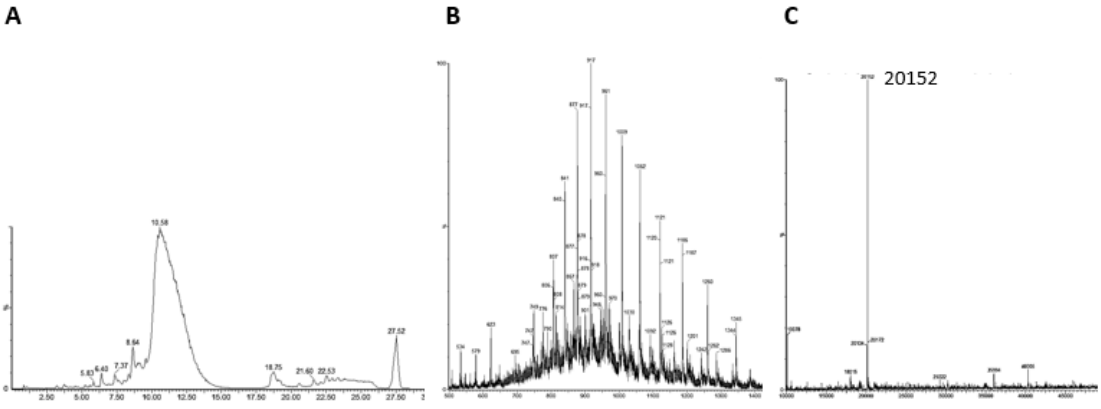


Figure 3.15: SQTKS *C*-MeT LCMS measurement using a C_5 column: **A**, MS TIC; **B**, ES^+ spectrum; **C**, corresponding MaxEnt® deconvolution.

3.3.4 Expression of Phosphopantetheine-Transferases

For the conversion of *apo*-ACP to *holo*-ACP a phosphopantetheine transferase (PPTase) is needed. The enzyme catalyzes the attachment of the phosphopantetheine moiety of coenzyme A to the hydroxyl group of a conserved serine (DSL region). Sfp is a bacterial (*B. subtilis)* PPTase and the gene of the enzyme was kindly provided by Franziska Hämmerling.[137] The amino acid sequence of the protein can be found in the appendix (section 12).

The pET28a(+) plasmid containing the Sfp gene was transformed into Top10 and BL21 DE3 cells and expressed in 2TY medium with 1 mM IPTG. The protein was expressed with a N-terminal His-tag and purified by a Ni-NTA followed by a Superdex 26/600 column. The purified protein is shown on the SDS-PAGE at the expected size of 28 kDa (Figure 3.16 **A**).

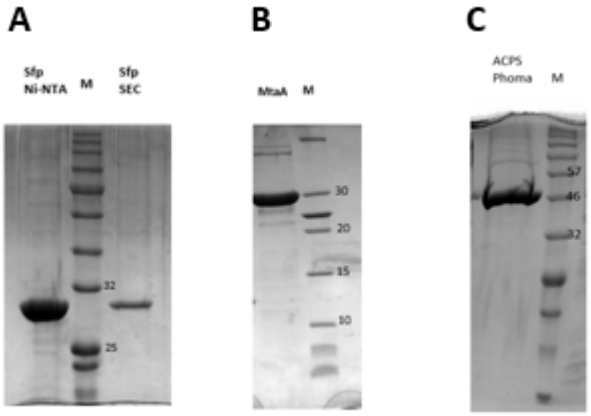


Figure 3.16: Expressed phosphopantetheine transferases and purified by Ni-NTA and SEC: **A**, Sfp; **B**, MtaA; **C**, ACPS *Phoma*.

The next PPTase which was expressed and purified was the MtaA protein. This PPTase is also a bacterial protein from *Stigmatella aurantiaca*. The gene (also provided by Franziska Hämmerling) was transformed into Top10 and BL21 cells, expressed and purified as explained for the previous protein. MtaA has a calculated size of 30.7 kDa and has also a N-terminal His-tag (amino acid sequence in the appendix). During the first Ni-NTA column run the protein precipitated when eluted. It was suggested that the imidazole concentration was too high (500 mM). In a second purification 250 mM imidazole were used for eluting the protein and as expected the protein did not precipitate. SDS-PAGE showed a successful purification obtaining the protein at the expected size of approx. 30 kDa (Figure 3.16 **B**).

The third PPTase to be investigated was the fungal *Phoma sp*. transferase. NpgA is a well known fungal PPTase from *Aspergillus fumigatus*. NpgA was used to search the genome of the MF5453 *Phoma* species (the native host of SQTKS) to find best homolog. The sequence similarity for this kind of proteins is generally low but there are 3 characteristic regions (Figure 3.17).[138]

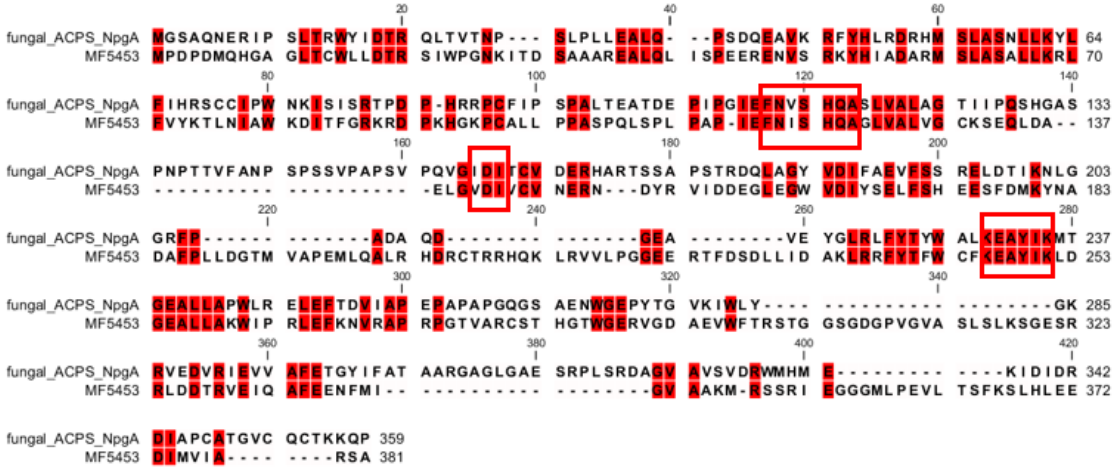


Figure 3.17: Sequence alignment of *Phoma sp*. MF5453 with npgA PPTase with characteristic regions marked with red boxes.

Considering these conserved regions, a gene was found in the full genome data of *Phoma* sp. MF5453 in GenDBE. The gene name in this system was MF5453_neug9815.t1. A synthetic gene was ordered (amino acid sequence in the appendix). The gene was *E. coli* codon optimized and synthesized by Baseclear in pET28a(+). The vector was transformed into *E. coli* Top10 and BL21, expressed and purified as described for the previous proteins. The SDS-PAGE showed an intense band at the expected size of 45.9 kDa (Figure 3.16 **C**). As for MtaA, 250 mM imidazole had to be used for elution.

Most of the following Phosphopantetheine transferase reactions were perfomed with Sfp. To be able to distinguish Sfp from other proteins in mixtures it was necessary to know the exact size of the protein by mass spectrometry. The MaxEnt transformed mass spectrum shows two main peaks of 28163 and 28137 (Figure 3.18). The expected size of Sfp with the loss of the first methionine is 28.134 kDa, which is close to the second main peak. The largest peak is probably a sodium adduct of Sfp.

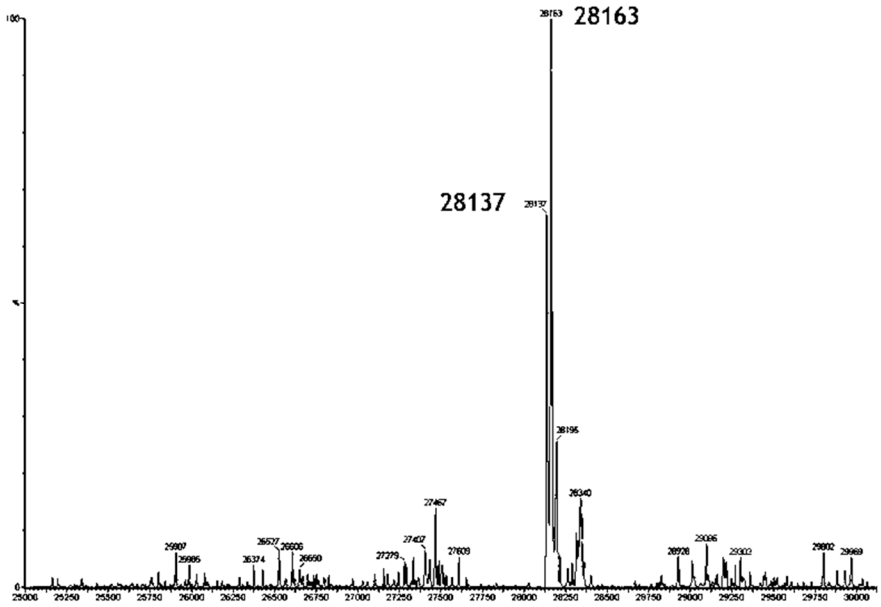


Figure 3.18: MaxEnt deconvolution for mass spectrum of Sfp.

3.3.5 Formation of *holo*-ACP

The main function of the ACP is to transfer substrates and intermediates, which are covalently bound to the protein, between active sites. The acyl carrier protein is a small domain that binds the growing polyketide chain at the 4'-phosphopantetheine (PPant) arm. A posttranslational modification transforms the ACP from the inactive *apo* form into the active *holo* form by tethering the PPant of Coenzyme A to a highly conserved serine. This reaction is catalyzed by a 4'-phosphopantetheine transferase (PPTase, Figure 3.19). The thiol at the free end of the PPant arm forms thiolester bonds with intermediates.

During the *holo* ACP reaction a mass increase of 340 Da, which corresponds to the mass of phosphopantetheine, is expected. The mass of SQTKS ACP (Con3B) was measured by ESMS before and after treatment with PPTase and CoA. The measured *apo* mass corresponds exactly to the calculated mass of the double His-tagged protein (13.317 kDa) The second intense peak has a 16 Da larger mass and might be the protein with an addition of an oxygen atom, an oxidized product. The *holo*-ACP reaction with Sfp and CoA was performed for SQTKS ACP Con3B but gave not the expected mass shift of 340 Da (Figure 3.20).

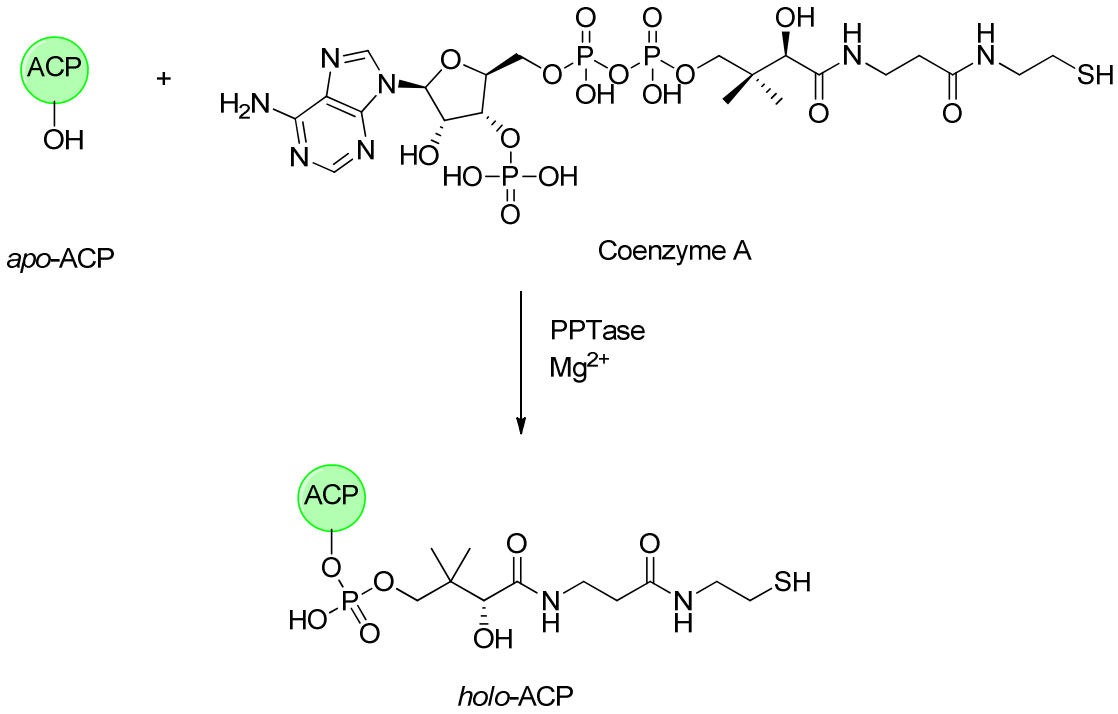


Figure 3.19: *holo* ACP reaction of *apo* ACP at the hydroxyl group of a strictly conserved serine.

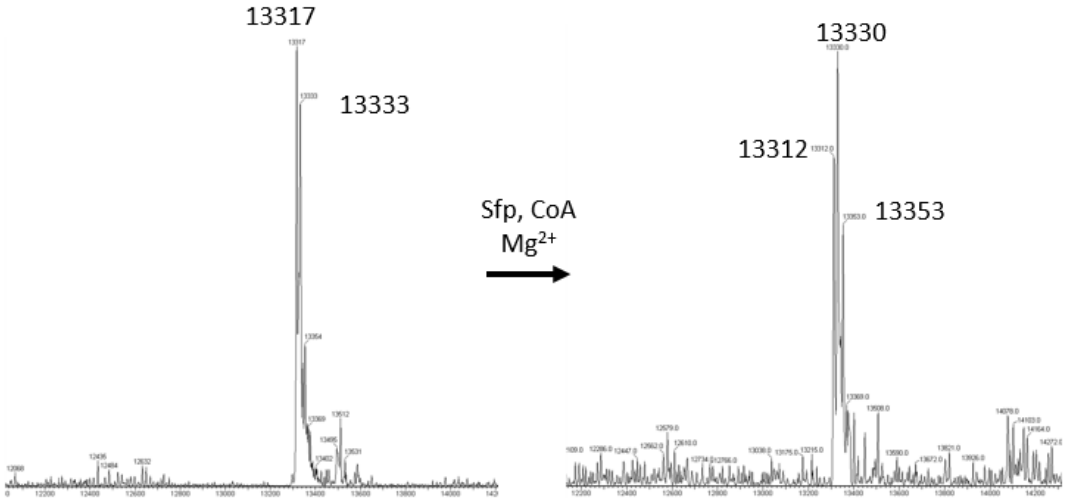


Figure 3.20: *holo* ACP reaction of SQTKS ACP Con3B.

In addition to standard protocols for the *holo*-ACP reaction,[98] divergent conditions were tested. Varying concentrations of ACP, PPTase, CoA as well as varying incubation times (30 min, 1 h, tried but all attempts did not lead to the desired formation of *holo*-ACP for Con3A.

The next tried construct for the *holo* reaction was Con3C, which has both N-terminal and C-terminal His-tags. The expected mass of the C-terminal tagged SQTKS

ACP (Con3C) was found in the spectrum at 11.570 kDa after a successful digestion with thrombin (calculated 11.566 kDa, Figure 3.21). The pET 28(a)+ vector has a thrombin cleavage site for the N-terminal His-tag but none for the C-terminal. This is the reason why only one His-tag is cleaved off. Again a *holo*-ACP reaction was performed with the single His-tagged SQTKS ACP but did not gave the desired mass shift. Obviously, the phosphopantetheinylation reaction was not successful.

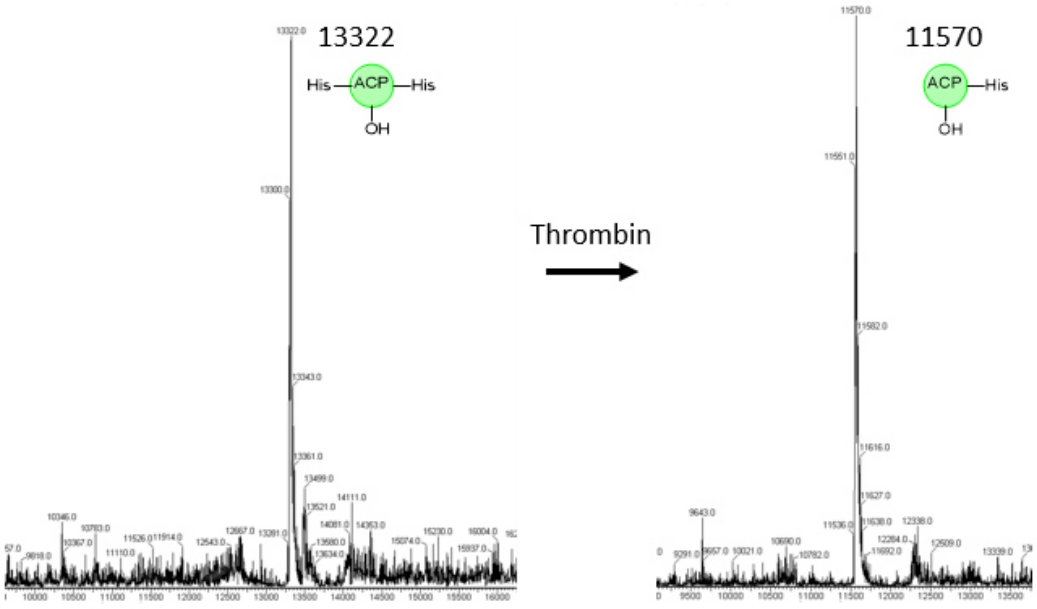


Figure 3.21: ESMS-MaxEnt deconvolution of SQTKS double tagged SQTKS ACP (Con3C) and C-terminal tagged ACP after thrombin cut.

Another reason for the failure of the reaction with two different SQTKS ACP constructs (Con3B and Con3C) could be the specificity of the PPTase. Although Sfp has a broad specificity, it is still a bacterial protein interacting with a fungal protein. Two alternatives were tried. MtaA, which is also a bacterial protein; and the *Phoma* PPTase. Both alternatives gave the same result as in the reaction with Sfp. No mass shift could be observed which means that no reaction took place.

A control, if all reagents are fine and enzymes active can be done with a different ACP. bfPKS ACP was used for the phosphopantetheinylation reaction. The mass spectrum of bfPKS *apo*-ACP showed two major peaks. The highest intensity peak belong to a protein of a size of 13.839 kDa (Figure 3.22, P1) which is correlated to the mass of N-terminal His-tagged bfPKS ACP (calculated mass: 13.841 kDa). The second protein (Figure 3.22, P2) has a 178 Da higher mass (14.017 kDa) than the first one.

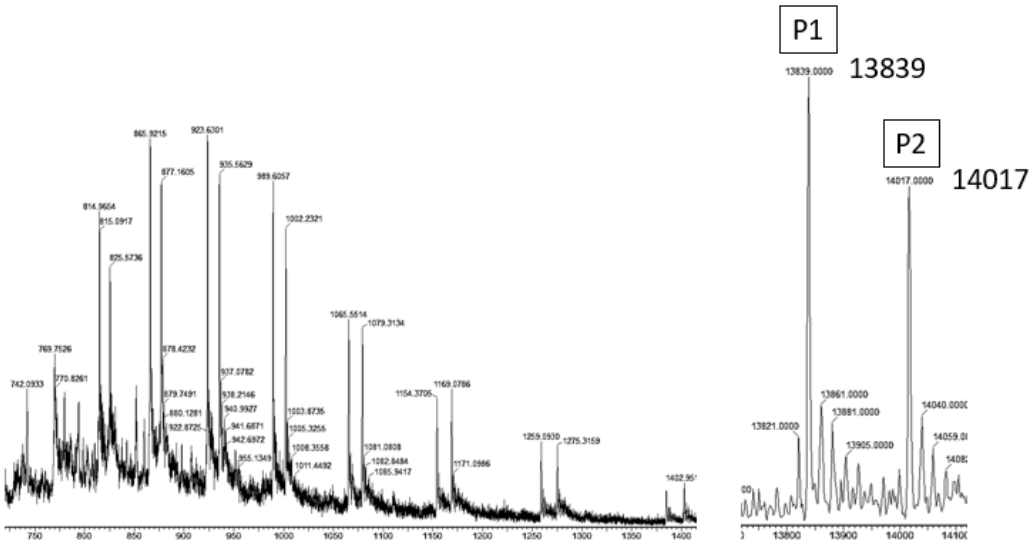


Figure 3.22: Mass spectrum and MaxEnt deconvolution of bfPKS ACP.

The bfPKS ACP phosphopantetheinylation reaction was successful. Both proteins (Figure 3.23, P1 and P2) showed a mass shift of 340 Da, corresponding to a phophopantetheine transfer, leading to *holo* proteins (Figure 3.23, P1' and P2') in the presence of Coenzyme A, Mg^{2+} and Sfp. This observation concludes that P2 is also a carrier protein including a modification.

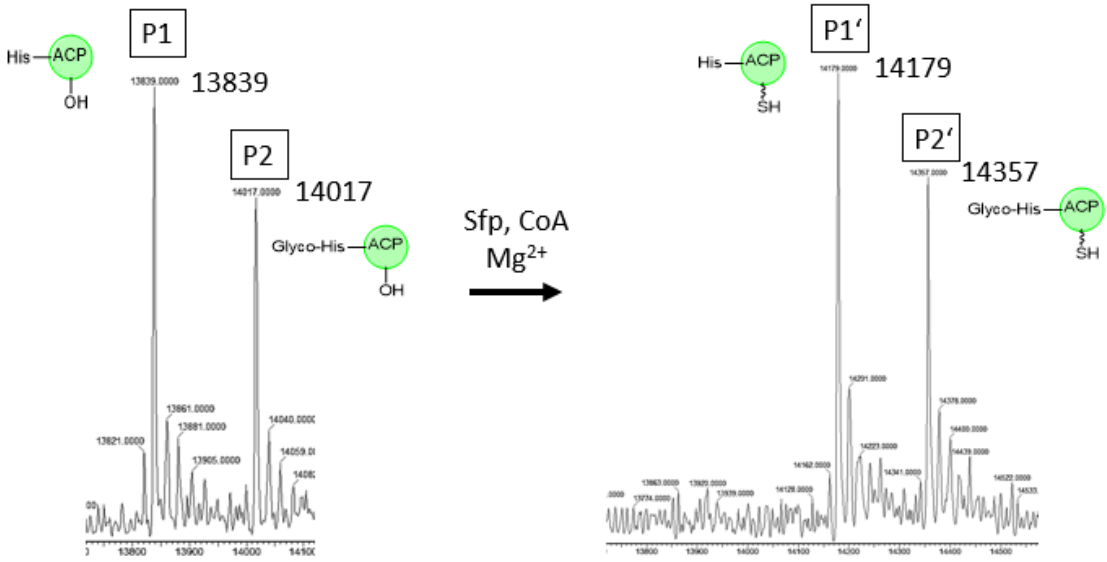


Figure 3.23: ESMS-MaxEnt deconvolutions for phosphopantetheinylation-reaction of bfPKS ACP with gluconoylated His-tag (Glyco-His) form of the protein.

A +178 Da modification of a protein is a known process during heterologous expression in *E. coli*. Proteins expressed in *E. coli* with the N-terminal $[His]_6$-Ser-Ser-Gly, included in addition to the expected mass of the protein, material with 178 Da of excess mass.[139] Gluconoylation on the N-terminal His-tag is causing the additional mass. In the case of bfPKS ACP the first residues after N-terminal His-tag are Ser-Glu-Gly. The second residue past the tag seems not to have a significant influence on this phenomenon described in the literature. Gluconoylation is a posttranslational glycosylation that can affect the quality of recombinant proteins such as lower the activity.[140] 6-phosphogluconolactone is responsible for the gluconoylation and can act as an electrophile. It is expected to be an *in vivo* nonenzymatic glycosylation reaction.[141]

A possibility to overcome the gluconoylation is to coexpress phosphogluconolactonase. This enzyme would convert 6-phosphogluconolactone by hydrolysis into 6-phosphogluconic acid and prevent the mass addition.[141] In this project a simpler method was used to solve this problem. A simplification of the two peaks ACP spectrum was obtained by cutting the N-terminal His-tag with thrombin. The additional peak of gluconoylated ACP was removed with this procedure. The best conditions for this reaction were already determined for SQTKS ACP and were used for the removal of N-terminal His-tag of bfPKS ACP.

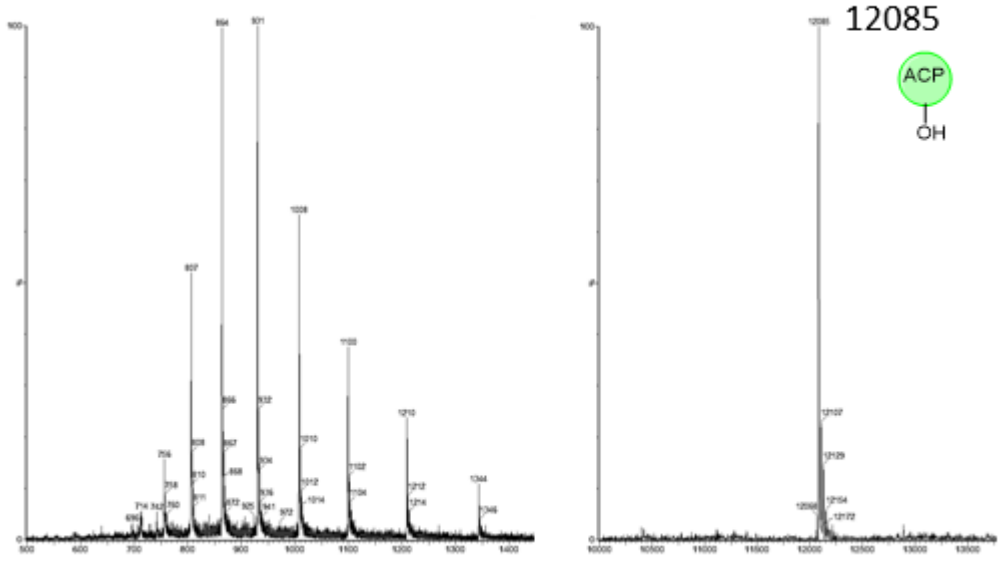


Figure 3.24: Spectrum and MaxEnt deconvolution for tag free *apo*-ACP of *B. fulva*.

His-tag free bfPKS *apo*-ACP has a 1.754 kDa smaller mass than N-terminal His-tagged protein. The observed mass of 12.085 kDa is close to the calculated mass of a tag free

ACP (12.090 kDa) (Figure 3.24). No gluconoylated ACP species can be observed anymore. The protein is still active and shows *holo* conversion when treated with Coenzyme A, Mg^{2+} and Sfp (Figure 3.25). The mass difference between *holo* (12.425 kDa) and *apo* (12.085 kDa) is 340 Da which is the mass of phosphopantetheine.

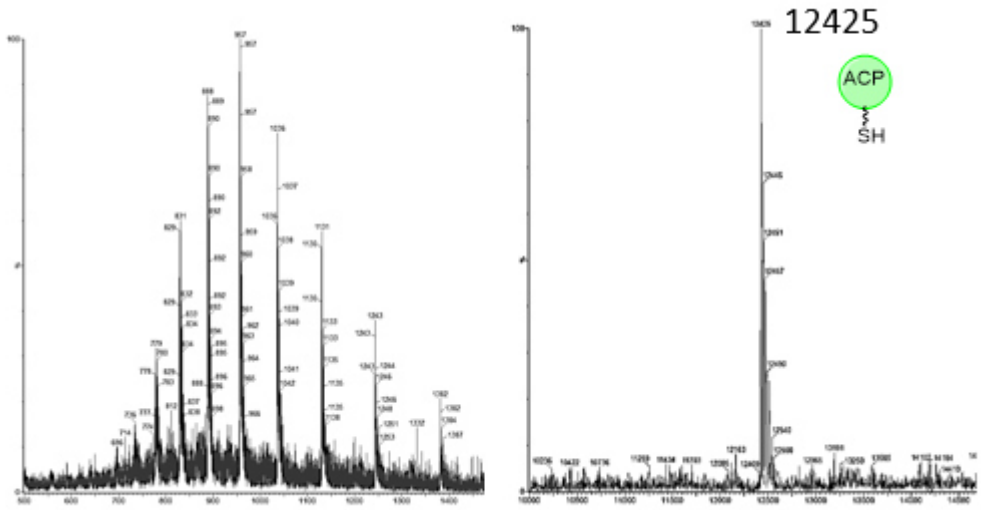


Figure 3.25: Spectrum and MaxEnt deconvolution for tag free *holo*-ACP of *B. fulva*.

3.3.6 Substrate Loading to ACP by a Chemical Reaction

After we had active phosphopantetheinylated ACP in our hands, the next step was to attach different substrates to the protein for *in vitro* investigation of corresponding domains. The first approach to attach substrates to the phosphopantetheinylated bfPKS ACP was a chemical reaction with thiol reactive compounds, which react with the thiol of the phosphopantetheine arm. The amino acid sequence of the bfPKS ACP contains a cysteine residue which might result in a double addition of a cysteine reactive compound. For this reason, the thiol group of the cysteine was blocked by iodoacetamide **77** (Figure 3.26). The blockage was performed after the Ni-NTA purification step of freshly expressed bfPKS *apo* ACP **76** with 50 µM **77** (excess) for 30 min at 4 °C. After the incubation time a SEC was performed. The blocked ACP **78** was then phosphopantetheinylated as described in section 3.3.5.

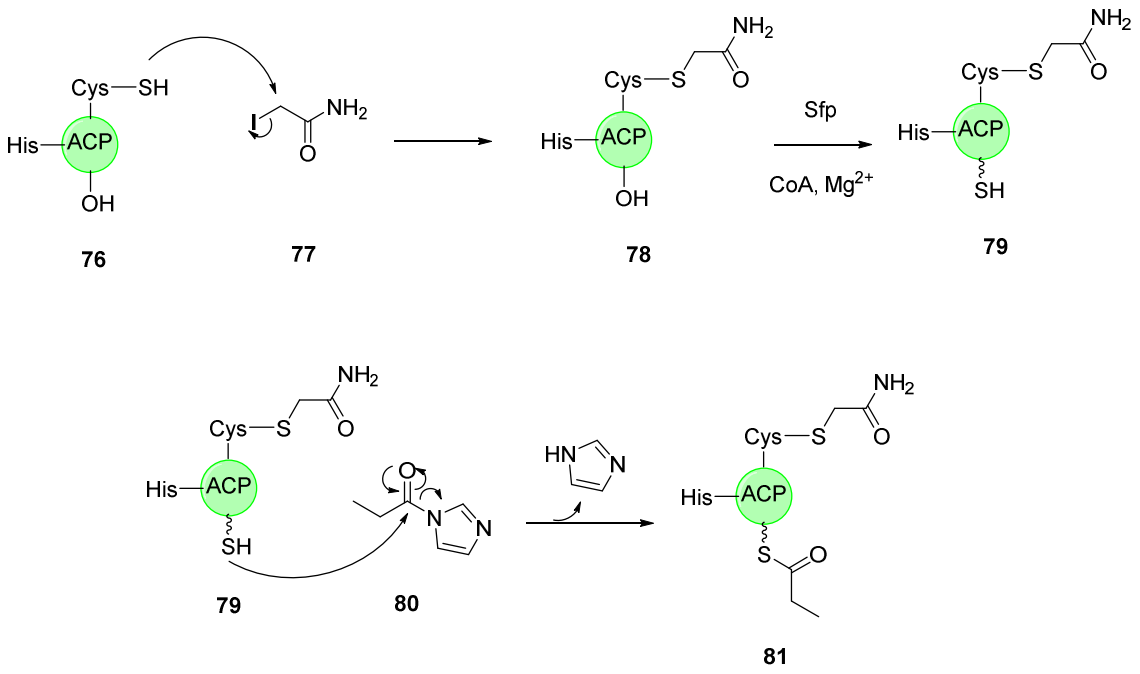


Figure 3.26: Blockage of the thiol group of a cysteine in bfPKS ACP with iodoacetamide, phosphopantetheinylation and loading reaction with a thiol reactive compound.

The *holo*-ACP **79** was then used for the first reactions with the commercially available thiol reactive compound **80** to obtain the loaded ACP **81**. For this reaction different concentrations were tested. ACP **79** (1 mg/ml) was incubated with 100 µM, 10 µM, 5 µM and 2 µM of **80** for 30 min at room temperature. After the incubation time the protein was measured by direct injection MS. The highest concentrations (100 µM and 10 µM) resulted in a spectrum in which no protein could be detected, assuming that the substrate concentration led to a precipitation of the ACP. The incubation of 5 µM substrate **80** led to a one-fold (calculated: 14.295 kDa), two-fold (calculated: 14.352 kDa) and three-fold (calculated: 14.409 kDa) addition to the ACP (Figure 3.27 **A**) even though there was no free thiol on the bfPKS ACP, except the phosphopantheine one. In the 2 µM incubation the one-fold addition was favored but the two- and three-fold additions were still detectable (Figure 3.27 **B**).

The issue, that even a simple substrate as **80** attaches multiple times to the bfPKS ACP makes this approach of loading substrates unattractive. Additionally, enzymatic loading reactions (section 5.3.2) showed satisfying results without side reactions. For this reason, this method was not used for following experiments.

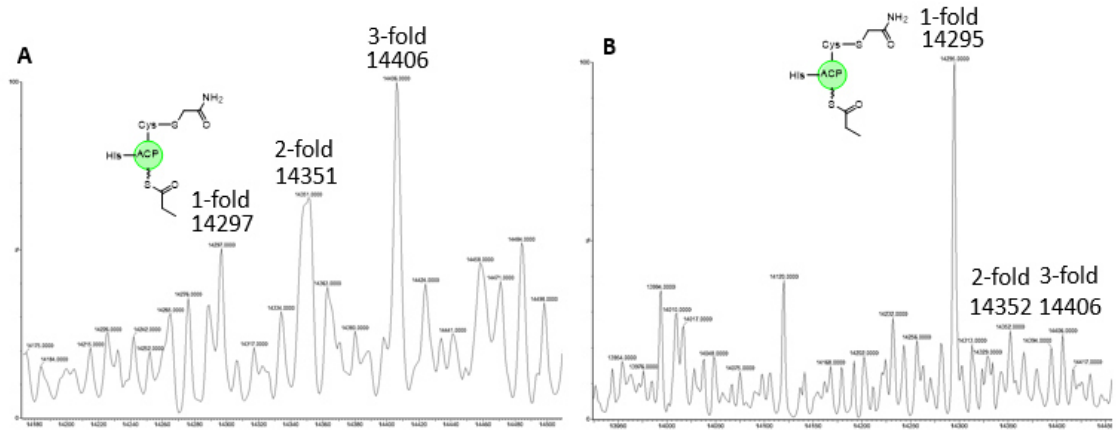


Figure 3.27: Substrate **80** loading to bfPKS ACP **79**: **A**, 5 µM substrate; **B**, 2 µM substrate.

3.4 Discussion

The expression of heterologous proteins in *E. coli* is most widely and routinely used. Besides the advantages of using bacterial hosts, there are also many limitations using *E. coli* as the heterologous host. These limitations include: post-translational modifications; limited ability for disulfide bond formation; varying expression rates leading to protein misfolding; aggregation; and protein degradation.[142] Obtaining soluble and active protein is still a challenge in protein chemistry. Depending on the folding during the expression the protein can be active or inactive. The unsuccessful attempts to obtain the active *holo*-form of the SQTKS ACP could be caused by the presence of two His-tags. These tags have combined a mass of 3.246 kDa (10.220 kDa is the mass for construct 3) which is 32% of the mass of Con3B ACP. The high proportion of tag is due to this double tagged construct and could have a big influence on the behavior of the ACP. It is very likely that the protein is folding in a disfavored manner and is not recognized by the PPTase. Post expression removal of the N-terminal His-tag causing a C-terminal tagged protein did not lead to a different folding and recognition by the PPTase. An expression of only N-terminal His-tagged ACP did not even lead to a successful expression. The protein seems to be digested or could be toxic for the cells. The same problem occurred with Con1. No expression could be observed for this case. It seems like the double tagged version of the SQTKS ACP protects it from digestion but leads to an inactive protein. Three different PPTases were tried, including one from the same species like the tested ACP. The reagents for the phosphopantetheinylation reaction were tested with a functional ACP, showing that the reaction conditions and reagents were fine.

Bioinformatic analysis show that the Type I SQTKS ACP (A2497 - A2603) has a very high folding propensity when expressed in *E. coli* compared to average proteins and the Type II actinorhodin ACP. Another hypothesis for the inactivity of the protein is that the expression in *E. coli* cells led to a highly folded protein in which the active site DSL region is not reachable for the PPTase. Expression of SQTKS ACP in eukaryotic systems might give active protein and should be tried in future projects.

A tag free expression in *E. coli* followed by a traditional non-affinity purification sequence as explained in section 3.1 for actinorhodin ACP might be worth a try for SQTKS ACP in future projects. Even if the purification of a protein without a tag may be more complicated, expressing a tag free ACP might fold the protein in a different way and could protect it from digestion. Another promising option could be expression in a eukaryotic system such as yeast. This organism is much more similar to the native host *Phoma sp.* when compared to the *E. coli* bacterial system.

In comparison to SQTKS ACP, the bfPKS ACP was unproblematic in terms of protein production, purification and *in vitro* activity. The *holo*-ACP reaction was successful and worked as expected. A posttranslational modification could be detected and identified as gluconoylation of the tag sequence. The gluconoylation problem was solved with removal of the N-terminal His-tag. This process simplified the mass spectrum, showing only one peak for the tag free ACP instead of two. ESMS was successfully optimized for protein measurements varying organic solvent to water ratio and acid concentration. C_5 column chromatography was tested for proteins and gave clear chromatograms for proteins like HHMG and SQTKS *C*-MeT but was not very useful for the main and most important protein for this project, the ACP. ACP had a strong binding towards the column which was too strong, leading to a long term elution of the protein. The column was not used for further experiments which were carried out by direct injection.

The substrate loading reaction with a thiol reactive compound showed multiple additions to the bfPKS ACP. Even at lower substrate concentration there were multiple attachments detectable, which made this approach of loading substrates unattractive.

4. Expression of SQTKS and *B. fulva* Dehydratase Domains

4.1 Introduction

In general, FAS and PKS dehydratase domains (DHs) use conserved His-Asp residues to catalyse a *syn*-*β*-elimination reaction to form an *E*-2,3 unsaturated product. There are several structures of DHs, which have been solved, such as: Curacin CurH and CurJ; FAS FabA; and erythromycin DHs (Figure 4.1).[143–145] All of them have a characteristic double-hotdog fold. The hotdog fold consists of two central *α*-helices ("sausages") wrapped in several *β*-strands ("bun"). This kind of characterization for a DH domain was firstly observed by the group of Leesong in 1996 where he described the hotdog fold of a DH from the *E. coli* type II fatty acid synthase FabA.[145]

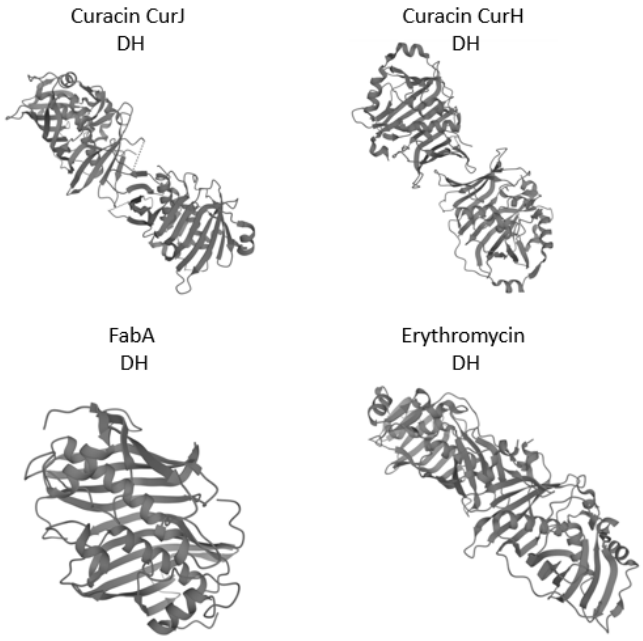
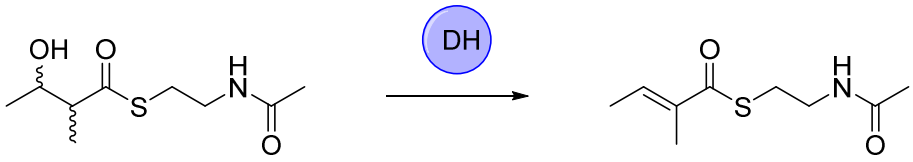


Figure 4.1: Crystal structures of Curacin CurJ DH (PDB: 3KG8), Curacin CurH DH (PDB: 3KG7), FabA DH (PDB: 1MKA) and erythromycin DH (PDB: 3EL6).

The structures of Curacin CurH and CurJ DHs show a dimeric interface which is typical for PKS DHs compared to FabA DH which is monomeric and typical for FAS DH domains. The Curacin DHs and the erythromycin DH domains all have similarly extended conformations.[143] The structure of the *ery* module 4 DH, showed how the *α*-hydrogen and the *β*-hydroxyl group of a substrate interact with the histidine and aspartate of the active site before elongation.[144]

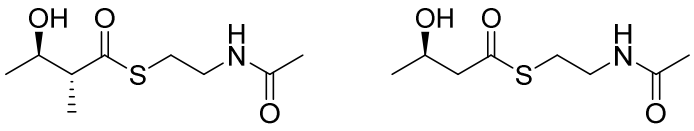
The first chemical investigations on the standalone SQTKS DH domain were done by David Ivison.[96] In his work the DH domain is catalytically active when removed from the context of the multidomain PKS protein (although the isolated protein is fairly unstable). He also showed that the DH substrate mimic 3-hydroxy-2-methylbutyryl-SNAC **82** is exclusively converted into the *E*-isomer **83** out of the mixture of all 4 diastereomers (Figure 4.2).



82 **83**

Figure 4.2: SQTKS DH catalysis out of 4 diastereomers of **82** resulted in the *E*-isomer **83**.

An optimization of expression and purification conditions for the SQTKS DH domain was achieved by Emma Liddle.[98,146] She showed that the optimized purification lead to a prolonged storage of the DH at 4 °C for up to 2 months without precipitation.[98] SQTKS DH substrate selectivity tests were also performed. Natural and unnatural substrate mimics were synthesized and tested *in vitro* with the DH standalone domain. The domain catalyses reaction of only the 2*R*,3*R*-**84** and the 3*R*-**85** substrate mimics (Figure 4.3) and has the same stereopreference as the mFAS DH domain.[98,147,148] The DH is highly programmed to only recognise a substrate with a 3*R*-hydroxyl group and tolerates only non-methylated or 2*R*-methylated substrates. All DH domains which have been analysed so far show a *syn*-elimination during catalysis. It is assumed that the elimination of water during the SQTKS DH catalysis is also performed by a *syn*-elimination.[149]



84 **85**

Figure 4.3: Structures mentioned in the text.

4.2 Aims

The main aim of this particular project is to express SQTKS and bfPKS DH catalytic domains and check their correct folding by a substrate conversion test. The overall aim is to crosslink the DH domains to an ACP domain and generate structural information of it. The expression of the SQTKS DH domain will be performed as described in previous work (section 10.6). BfPKS DH domain will be expressed for the first time. To do so, standard expression conditions will be used as for bfPKS ACP and the SQTKS DH purification protocols. An activity test with a potential substrate of the domains will be performed in the last step. The DH domains will be further used for crosslink experiments with the bfPKS ACP domain. Crosslinking between DH and ACP domain had already been performed in bacterial PKS.[150]

4.3 Results

4.3.1 Domain Boundary Prediction for the *B. fulva* Dehydratase Domain

In the first step of the boundary prediction for the bfPKS DH the complete *B. fulva* PKS sequence was inserted into HMMER to have an idea of the approximate boundaries for the DH. HMMER is a web server for biosequence similarity searches using profile hidden Markov Models.[151] The web server predicted the domain borders of the PKS (Figure 4.4 **A**). The DH domain was predicted in the region of F1010 - F1320. Making sure to not miss any parts of the DH due to prediction errors, an additional 40 residues at the N- and C-termini were included for further bioinformatic analysis (N970 – N1360).

The sequence (N970 – N1360) was analysed in the next step by SWISS MODEL using CurK (best hit after template search) as a template to create a 3D model. SWISS MODEL predicted the boundaries at R1012 – E1334 (Figure 4.4 **B**), similar to the HMMER prediction. A third bioinformatic method was used to approximate to the "real" domain boundaries of the DH. The protein sequence of the approximate region (N970 – N1360) was aligned with well characterized DH domains of a type I modular PKS (CurK 3KG9 and CurF 3KG6) and the SQTKS DH (N989 – A1312), a type I iterative PKS. The first (N-terminal, Figure 4.5 **A**) and the last (C-terminal, Figure 4.5 **D**) conserved residue in the alignment were chosen as borders. Similar to the bfPKS ACP boundary prediction this strategy enables shortening (N- and/or C-terminally) of the sequence by cloning, if

necessary. SWISS MODEL prediction (Figure 4.5 **C** and **D**) was the second choice for possible boundaries. The sequence region S1002 – A1357 was chosen for the gene synthesis and was performed by Baseclear. The gene was *E. coli* codon optimized and restriction sites (*Nde*I and *Xho*I) were included. The gene was finally subcloned into pET28a (+). The protein has an N-terminal His tag when expressed.

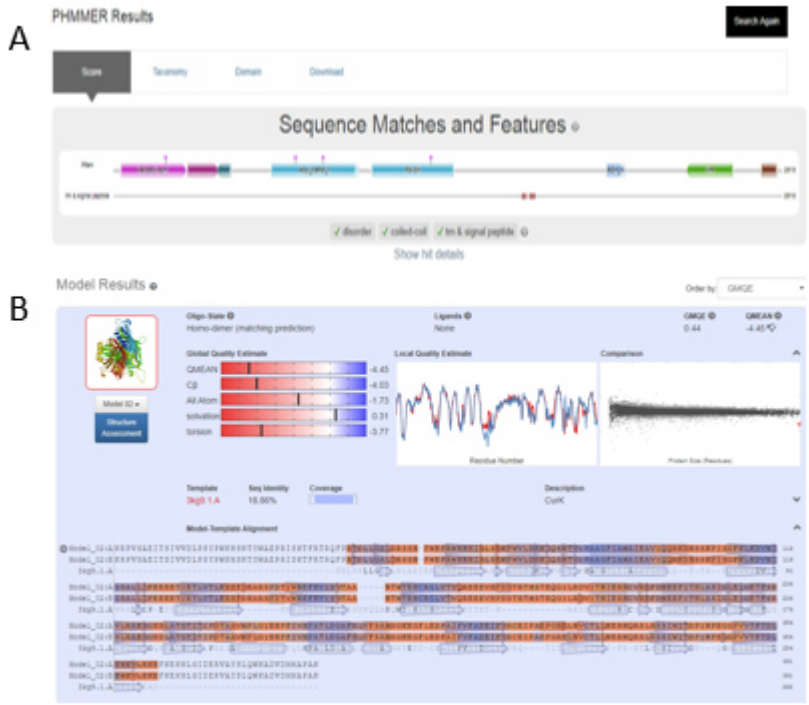


Figure 4.4: Bioinformatic analysis of the bfPKS DH region: **A**, result of protein HMMER analysis for the approximate region; **B**, analysis of the sequence with SWISS MODEL.

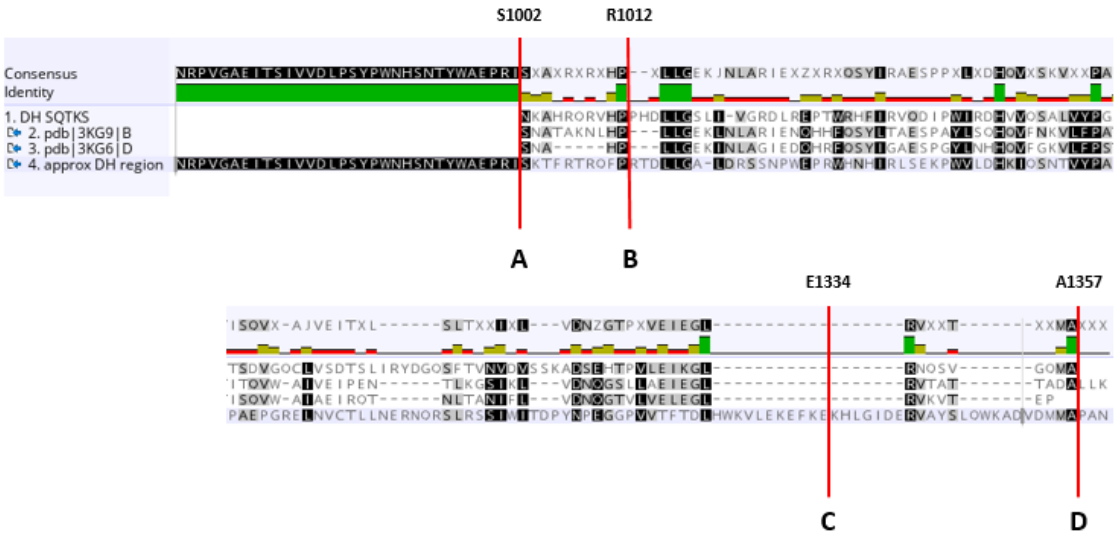


Figure 4.5: Sequence alignment of the bfPKS DH region with the sequences of SQTKS DH, CurF and CurK: **A**, start position based on the alignment; **B**, start position based on SWISS MODEL; **C**, end position based on SWISS MODEL; **D**, end position based on the alignment.

4.3.2 Expression of the SQTKS and the *B. fulva* Dehydratase Domains

As already mentioned in the introduction there were several group members in the past working with the SQTKS DH. The origin of the DH containing plasmid was from David Ivison (and later Christoph Bartel). The SQTKS Dehydratase domain was expressed as described by Emma Liddle in *E. coli* BL21 DE3.[98] The starter culture was inoculated in LB medium containing 50 µg/mL carbenicillin with a single colony. The cells were grown over night at 37 °C. Main culture (800 mL LB medium in eight 500 mL flasks each 100 mL) was inoculated with the starter culture at 1:100 dilution and incubated at 37 °C till an OD_{600} of 0.4 was reached. The temperature was lowered to 16 °C and the protein production was induced by 0.4 mM IPTG for overnight (approx. 21 h). In the next day the cells were lysed by sonication and the soluble part was separated from the insoluble part by centrifugation (10000 rpm for 1 h). After filtering the soluble part, the protein was purified by Ni-NTA and size exclusion column using an Äkta pure system. Modified buffers were used (section 10.6.2). The purified SQTKS DH protein was analysed by SDS PAGE and ESMS (Figure 4.6).

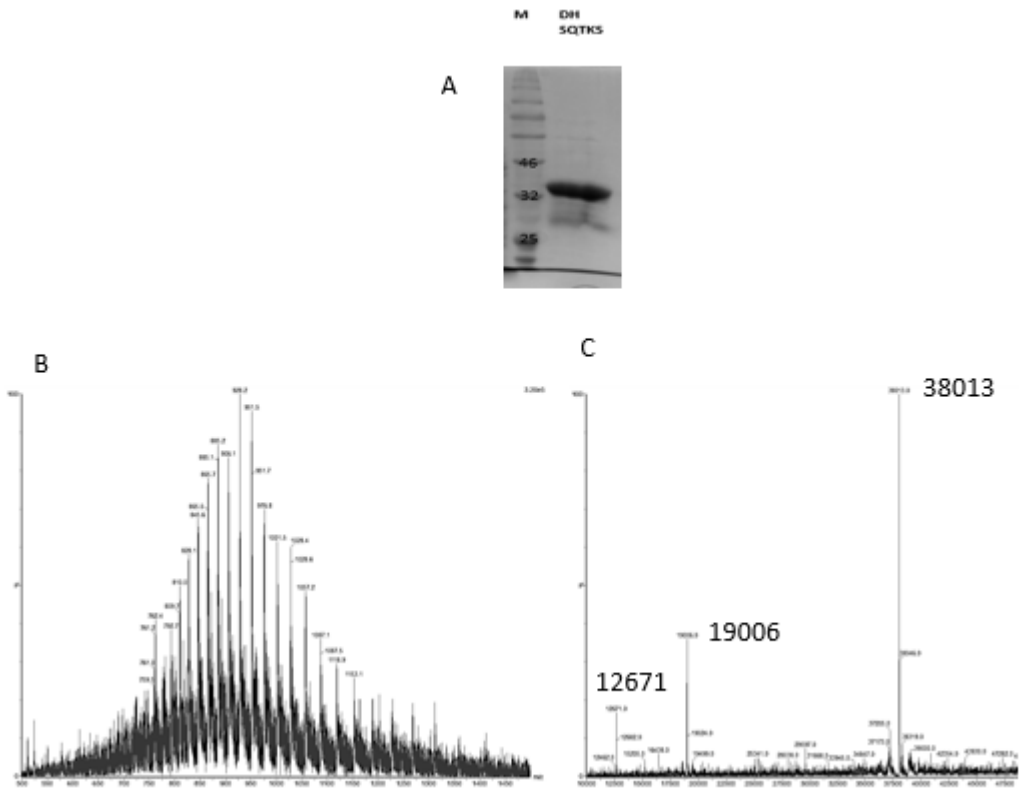


Figure 4.6: Heterologously expressed SQTKS DH: **A**, SDS PAGE after Ni-NTA and SEC purificatrion; **B**, ES^+ protein spectrum; **C**, deconvoluted spectrum.

The SDS PAGE analysis shows a protein band in the expected range (Figure 4.6 **A**). A light impurity or degradation of the protein can be seen in lower molecular range. The ES^+ spectrum shows a good ionization of the protein. An addition of 0.5% acetic acid was necessary for the improved ionization (Figure 4.6 **B**). The mass of 38.013 kDa (Figure 4.6 **C**) corresponds to the calculated mass of the SQTKS DH without the first 3 residues (38.014 kDa) of the N-terminal His tag (MGS). The calculated mass of the complete His-tagged protein is 38.307 kDa. Two other peaks in the deconvoluted spectrum have exactly the half (19.006 kDa) or one third (12.671 kDa) of the mass of the protein and are calculation artefacts.

The correct folding of the protein was ensured with a substrate conversion experiment. The SQTKS DH (0.1 mM) was incubated with a racemic mixture of **82** (1 mM) for 19 h at 25 °C and the resulting mixture analysed by LCMS. A negative control was performed in parallel using buffer instead of the DH. Dr. Daowan Lai synthesized the 3-hydroxy-2-methylbutyryl-SNAC **82** and also the expected product **83**. Enoyl SNAC **83** was used as a standard for the experiment.

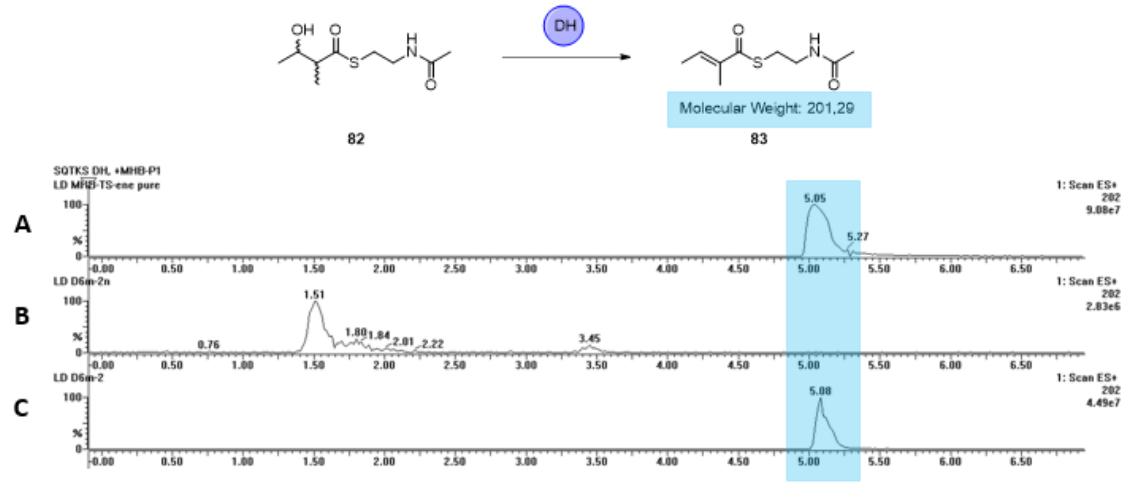


Figure 4.7: Activity test for the SQTKS DH domain: **A**, EIC of standard control of **83**; **B**, EIC of the negative control; **C**, EIC of the DH assay.

The SQTKS DH domain converted the substrate **82** to the expected *E*-alkene product **83** (Figure 4.7 **B**). The product **83** was detected in the mass spectrum at the same retention time as the positive control.

The first expression trial with the bfPKS DH was performed analogous to bfPKS ACP. The protein (S1002 – A1357) was expressed heterologously in *E. coli* BL21 DE3 using 2TY medium. The cells were grown at 37 °C, shaking at 200 rpm until they reached an OD_{600} of 0.6 and were then cooled down to 16 °C. The induction was performed with 1 mM IPTG over night. The cells were then lysed by sonication and the soluble fraction was separated from the insoluble fraction by centrifugation. The expression was controlled by the SDS-PAGE (Figure 4.8)

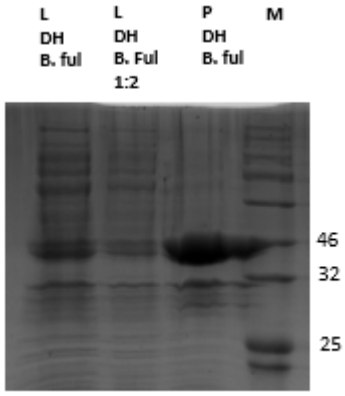


Figure 4.8: Expression of the bfPKS DH with soluble (L) and insoluble part (P).

A band of the expected size (41 kDa) was observed in the insoluble as well as in the soluble fractions. It was decided that the amount of soluble protein is enough for a

purification. A large scale expression was performed with the conditions described before. The expressed bfPKS DH was purified by a Ni-NTA column in the first step and by the size exclusion chromatography in the second step. The purification was performed with the same buffers as for SQTKS DH. The successfully purified protein was monitored by SDS-PAGE and mass spectrometry (Figure 4.8).

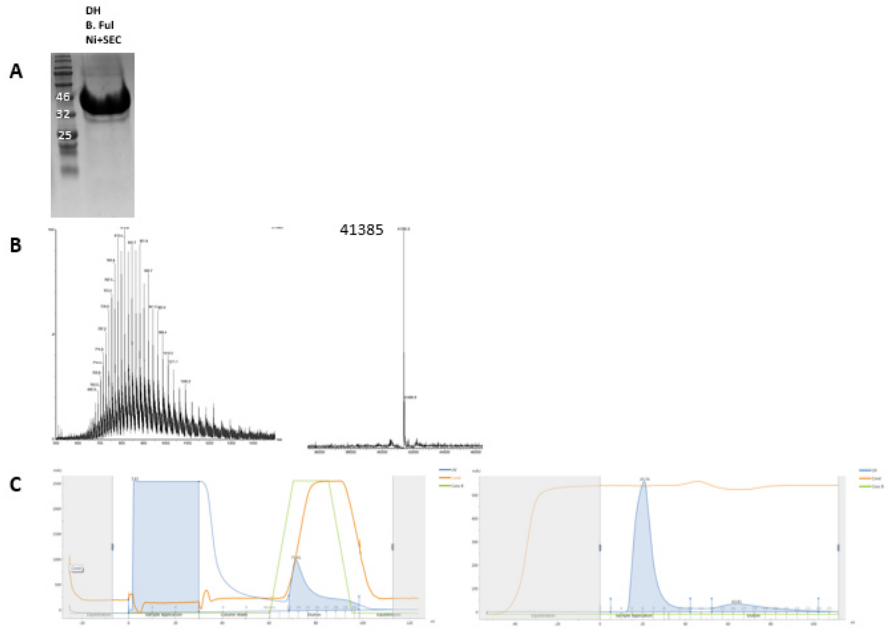


Figure 4.9: Purification and analysis of bfPKS DH: **A**, SDS-PAGE of Ni-NTA and SEC purified protein; **B**, protein mass spectrum; **C**, chromatograms of Ni-NTA and SEC run.

Next, the bfPKS DH was tested for its catalytic activity. The DH (0.1 mM) was incubated with the substrate **82** (1 mM) for 19 h at 25 °C. A negative control contained boiled DH and was treated in the same way. LCMS analysis of the product mixture showed that the expected product **83** was formed in the presence of enzyme, but not in the presence of the boiled control (Figure 4.9). The bfPKS DH is catalytically active.

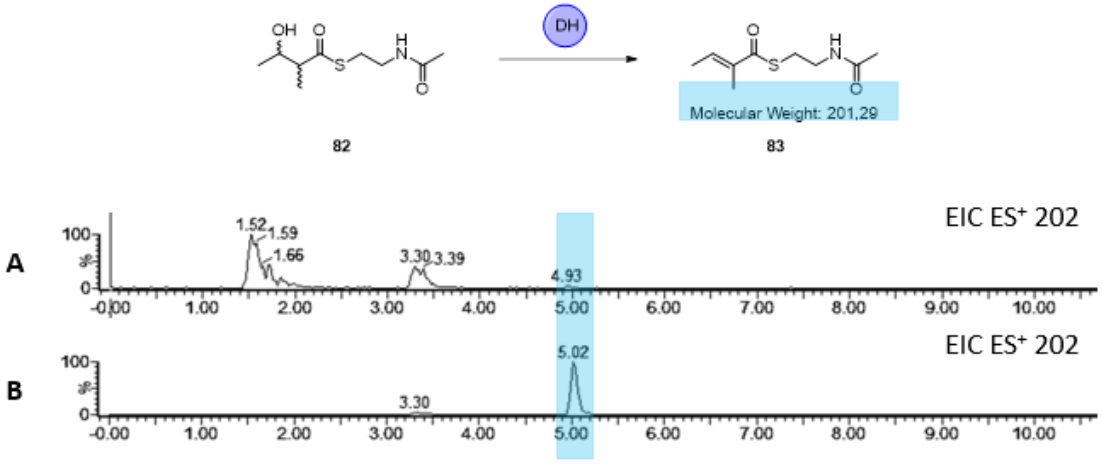


Figure 4.9: Activity test for the bfPKS DH domain: **A**, EIC of negative control; **B**, EIC of the DH assay.

A third DH domain was heterologously expressed and purified by Dr. Daowan Lai. The DH of the strobilurin pathway (stPKS) was also tested positive for its catalytic activity and was kindly provided for crosslinking experiments (section 7.3.1). The mass spectrum of the stPKS DH showed a protein of 31.809 kDa (Figure 4.10). The expected mass of the N-terminal His tagged DH is 36.621 kDa. The protein seems to have lost several amino acids, which have no influence on catalytic activity.

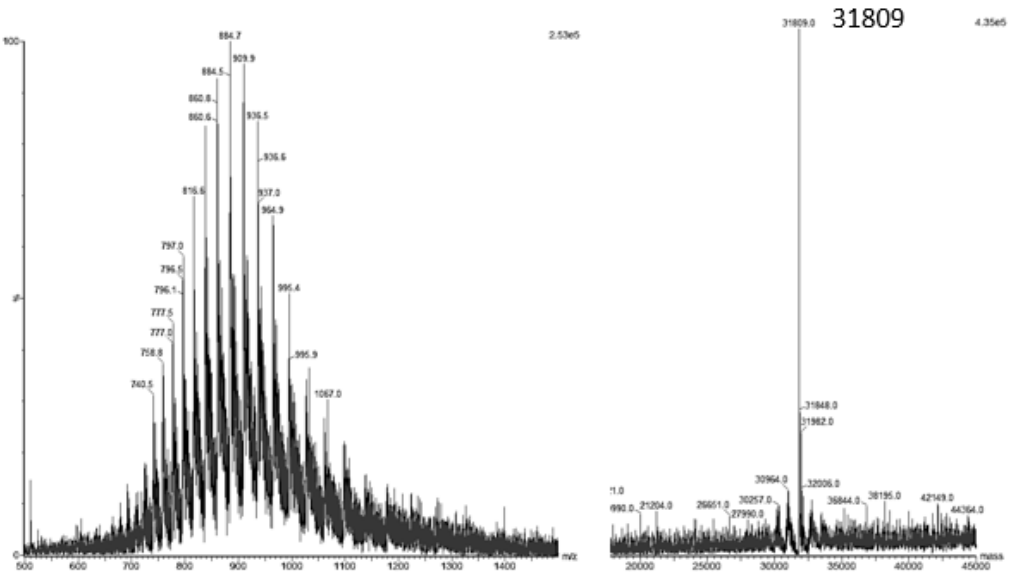


Figure 4.10: Mass spectrum of the strobilurin DH.

4.4 Discussion

The main aim of this part of the project was to heterologously express catalytically active bfPKS DH and SQTKS DH.

The bfPKS DH (S1002 – A1357) was expressed for the first time. The borders were predicted using bioinformatics tools and a synthetic gene was used for the expression. The protein was expressed in *E. coli* cells successfully. The same expression conditions were used as for bfPKS ACP. The purification was performed by Ni-NTA and SEC. The same buffer composition was used as for the SQTKS DH. The expressed bfPKS DH showed, compared to the SQTKS DH, a much cleaner SDS-PAGE with less impurities in lower molecular range. The bfPKS DH seems to be more stable compared to the SQTKS DH. The catalytic activity of the bfPKS DH was tested with a simple diketide SNAC substrate. The protein converted the substrate to the expected species. Experiments need to be performed in future to check for how long the activity can be maintained. Also conversion tests with natural and unnatural substrates can be perfomed in future to characterize the bfPKS DH domain in terms of selectivity.

The SQTKS DH was expressed and purified by an already existing protocol and showed also substrate conversion. As already described in a previous work, the SQTKS DH is degraded partially during the expression process and shows several additional bands on the SDS-PAGE. The SQTKS DH stays active for three to four days, when stored at 4 °C.

The stPKS DH was available in the group and showed catalytical activity with the same substrate as bfPKS DH and SQTKS DH. All three DHs showed clear mass spectra and can be used for crosslinking reactions.

5. ACP Loading with CoA Analogues

5.1 Introduction

Many primary and secondary metabolic processes require Coenzyme A and its thiolesters as substrates. During polyketide biosynthesis the AT domain binds CoA substrates and transfers them to the KS domain and ACP. *In vitro* investigations of PKS domains with CoA substrates or ACP bound substrates are closer to reality than using corresponding pantetheine or SNAC analogues, which are more simple and easier to synthesize. An effective method reported by Wright and colleagues was reported, using enzymatic reactions to generate CoA analogues from pantetheine derivatives.[152]

Pantothenate kinase (PanK), phosphopantotheine adenylyltransferase (PPAT) and dephosphocoenzyme A kinase (DPCK) are biosynthetic enzymes from *E. coli*. Chemically synthesized pantetheine analogues can be transformed in three steps or in a one pot reaction to the corresponding CoA analogues. A phosphorylation reaction followed by an adenylate attachment and finalized by a second phosphorylation are the required reactions for the biosynthesis (Figure 5.1). All three steps require ATP. Since we already have ready synthetic access to a range of acyl pantetheines (section 5.3.2), this methodology could be applied to create the corresponding CoAs. And since acyl CoAs are known to be loaded onto *apo*-ACP synthetic pantetheines could also then be used to create acylated ACPs.[152]

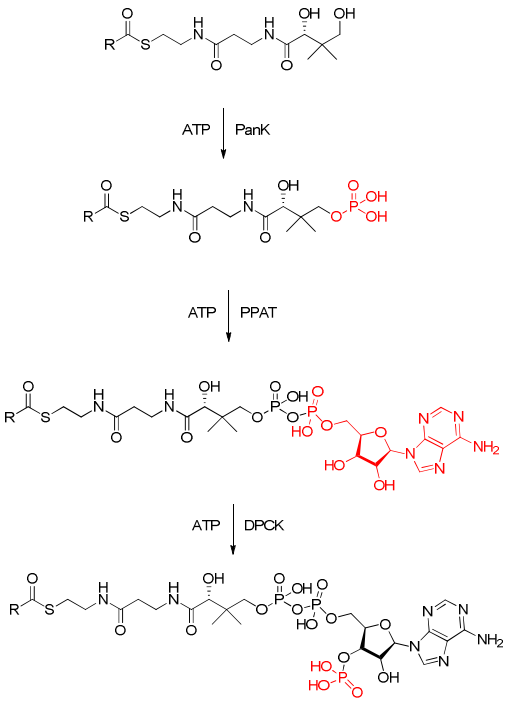


Figure 5.1: Biosynthetic reactions catalyzed by PanK, PPAT and DPCK.

5.2 Aims

The main goal is to synthesize acyl CoA analogues from acyl pantetheine precursors and load them onto bfPKS ACP. A biosynthetic approach will be used to achieve this goal, using three enzymatic reactions leading to the desired CoA analogues. First of all, the biosynthetic enzymes PanK, PPAT and DPCK will be expressed in *E. coli* and purified in the next step. All three catalytic steps, from a pantetheine to a CoA analogue, will be followed by LCMS. If the stepwise reaction works as expected, a one pot reaction will be set up for the following analogues containing PanK, PPAT, DPCK and also Sfp and ACP to load the produced CoA analogue to the *apo*-ACP directly.

5.3 Results

5.3.1 Expression of PanK, PPAT and DPCK

The pET28a (+) plasmids containing the PanK and PPAT genes were received from Dr. Juliane Buschmann. The DPCK gene was cloned from *E. coli* genomic DNA using the primer pair listed in table 10.8 (section 10.7.2). After amplification by PCR the gene and the pET28a (+) vector were cut with *Nde*I and *Xho*I in a double digest approach and ligated with the T4 DNA ligase as already described for SQTKS ACP (Section 3.3.2). PanK, PPAT and DPCK were expressed in 2TY medium using 50 µg/ml kanamycin and 1 mM IPTG for induction. The cells were induced for 4 h at 30 °C, harvested by centrifugation (5000 rpm) and lysed by sonication. The supernatant was separated from the pellet by centrifugation (10000 rpm for 1 h) and filtered before purification. In the first purification step a Ni-NTA column was used with 250 mM imidazole for the elution. The second purification and desalting step was performed by size exclusion chromatography. Both columns were installed in an Äkta pure system. The proteins were concentrated by centrifugal concentrators and stored at 4 °C and -20 °C.

After the purification all three proteins were analysed by SDS PAGE and mass spectrometry. The calculated mass of N-terminal His tagged PanK is 38.673 kDa. The measured mass of the protein is 39.183 kDa (Figure 5.2 **B**) which is 510 kDa larger than the expected size. Some unknown modifications occuring after expression are presumably the reason for the increased mass

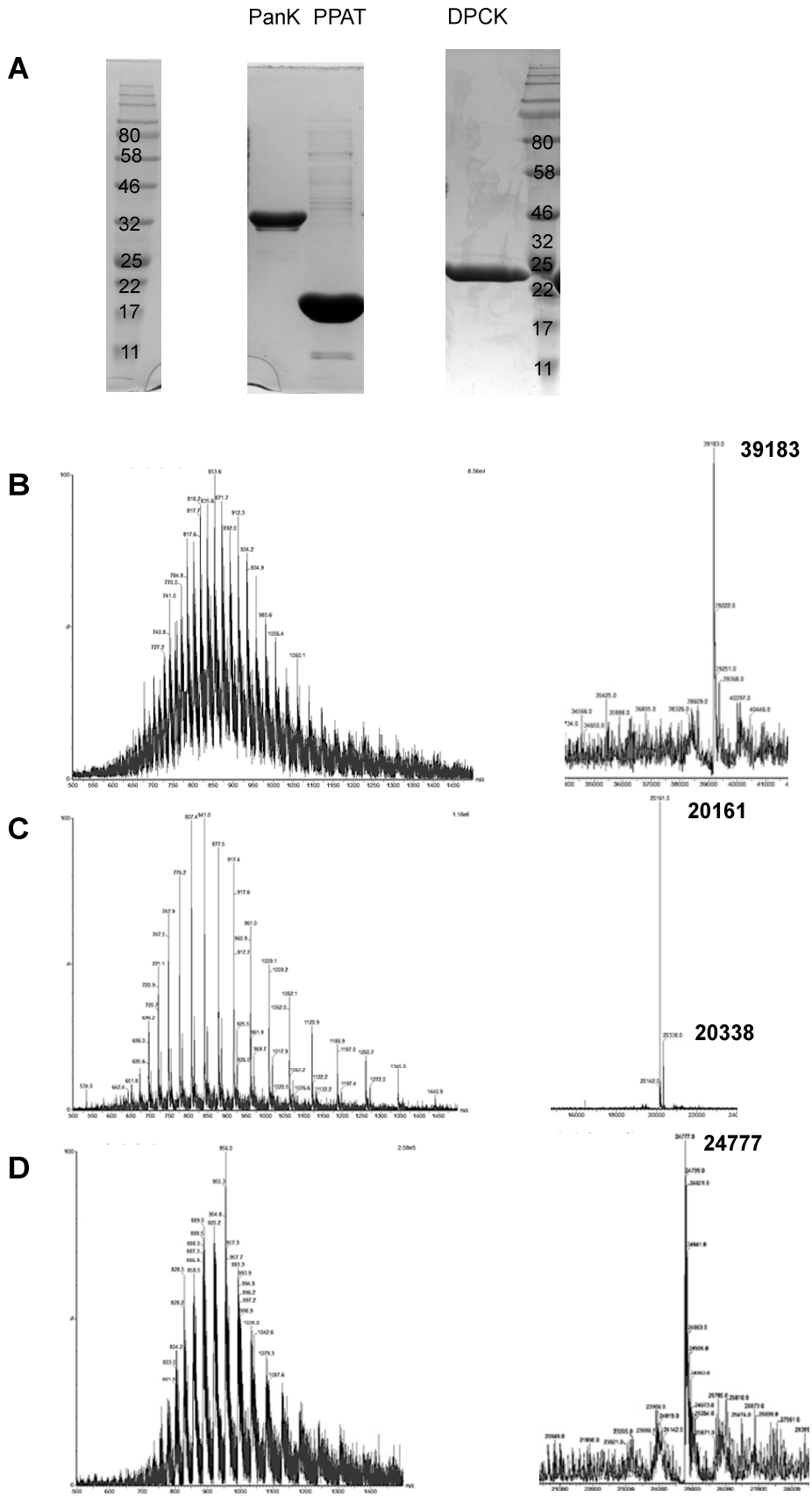


Figure 5.2: Biophysical analysis of purified PanK, PPAT and DPCK: **A**, SDS PAGE of purified PanK,PPATand DPCK; **B**, untransformed and transformed MS spectrum of PanK; **C**, untransformed and transformed MS spectrum of PPAT; **D**, untransformed and transformed MS spectrum of DPCK.

The SDS PAGE of the protein shows an intense band in the expected size without visible impurities (Figure 5.2 **A**). The calculated mass of PPAT is 20.150 kDa which is close to the measured mass of 20.161 kDa. The second peak at 20.338 kDa belongs to the gluconoylated species (Figure 5.2 **C**). The SDS PAGE of PPAT shows an intense band in the expected size and minor impurities (Figure 5.2 **A**). The SDS PAGE of the heterologously expressed and purified DPCK showed an intense band in the expected size of 25.8 kDa (Figure 5.2 **A**). DPCK gave a clear spectrum and showed a mass of 24.277 kDa (Figure 5.2 **D**) which corresponds to the mass of DPCK with the loss of the first six residues (24.280 kDa). The mass of His tagged DPCK without the first methionine is 24.785 kDa.

5.3.2 Preparation of CoA Analogues and Loading to bfPKS ACP

In the following two sections synthetic linkers (Figure 5.3) will be converted to CoA analogues and loaded to the bfPKS apo ACP. For this approach the biosynthetic enzymes PanK, PPAT, DPCK and Sfp were necessary.

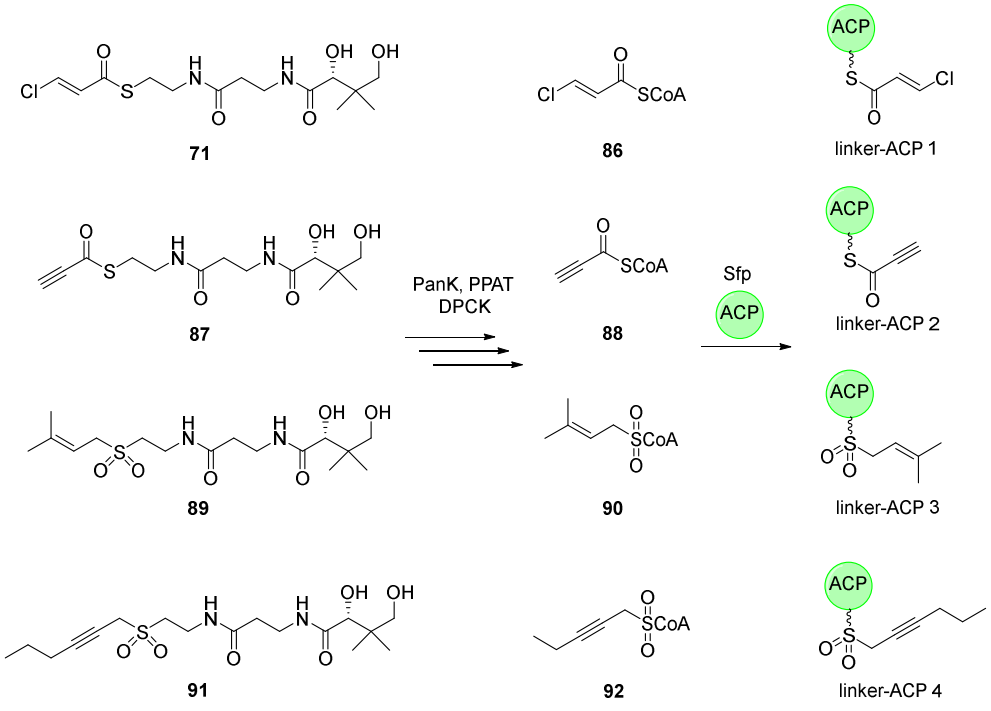


Figure 5.3: Overview of available linker pantetheines, their conversion into CoA analogues and loading to the *apo* ACP.

The sulfone linkers **89** and **91** were synthesized by Tanja Lau (Kirschning Group) and were provided in a protected form. Deprotection of the cyclic acetal was performed with $InCl_3$ in CH_3CN/H_2O leading to the pantetheine analogues **89** and **91**.[153] For the reactions 0.5 mmol (1 eq.) **93** (or **94**), 1 mmol $InCl_3$ (2 eq.) were dissolved in 2 mmol water (4 eq.) and 2.5 mL acetonitrile. The mixture was stirred for 3 h at room temperature. After evaporating the solvent, the products were purified by flash column chromatography (CH_2Cl_2 : MeOH 9 : 1). The success of the reaction was monitored by LCMS (Figure 5.4). The reactions were successful and the pantetheines **89** and **91** were used for the biosynthetic reactions. Pantetheine linkers **71** and **87** were synthetized and provided by Janina Meyer.

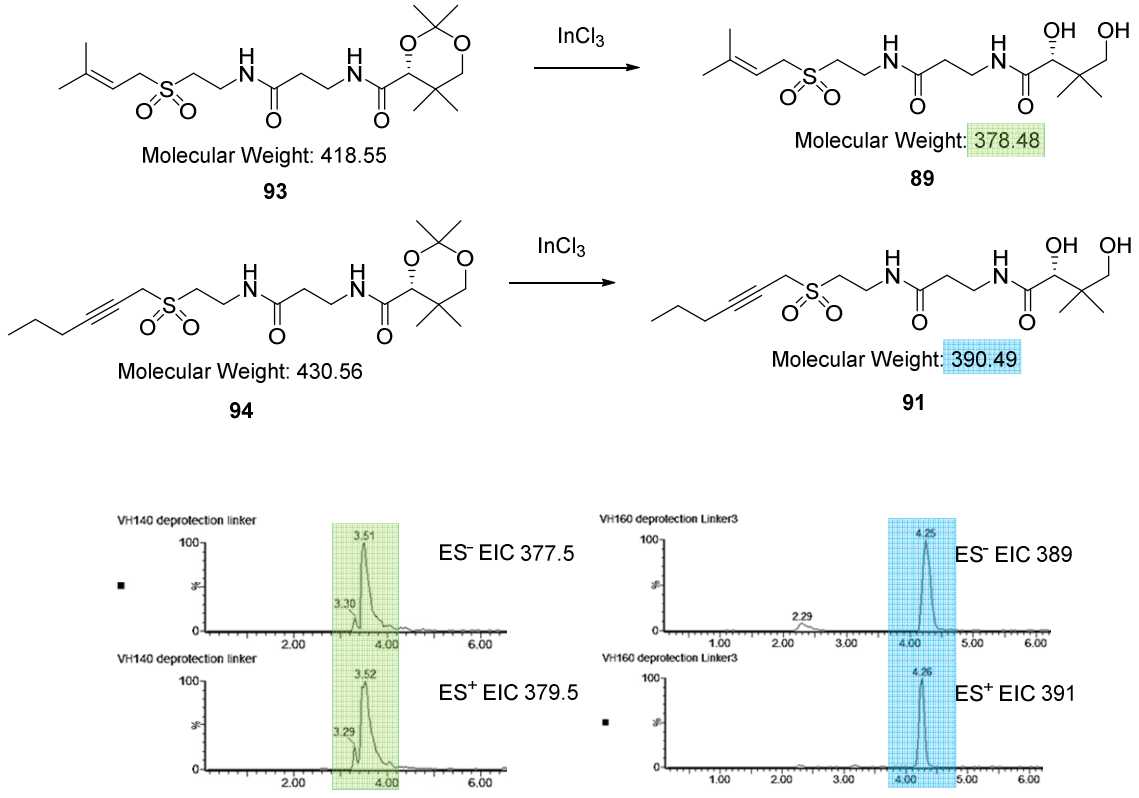


Figure 5.4: Acetal cleavage reactions of the linkers **93** and **94** with ES^+ and ES^- EIC for both products.

The first tested pantetheine analogue was **71**. Each of the three biosynthetic reactions (PanK, PPAT and DPCK) were followed by LCMS. The first reaction contained 50 mM Tris (pH 7.5), 20 mM KCl, 10 mM $MgCl_2$, 5 mM ATP, 0.2 mg/mL PanK and 2 mg/mL **71**. The mixture was incubated for 30 min at 25 °C and analysed by LCMS. The reaction worked as expected and the product was observed in positive as well as negative ionization modes at the expected mass of the phosphorylated product **95** (Figure 5.5 **A**).

The mixture was then used for the second reaction without further purification. 5 mM ATP and 0.2 mg/mL PPAT were added and incubated again for 30 min at 25 °C. The expected mass of the adenylate **96** was detected in both ionization modes (Figure 5.5 **B**). A smaller peak with the mass of 447.1 Da in the positive mode belongs to the substrate **95**, which was also detected, showing that the substrate was not fully converted at this time point.

For the third reaction 5 mM ATP and 0.2 mg/ml DPCK were added to the mixture and incubated again for 30 min at 25 °C. Positive as well as negative ionization modes showed the mass of the final CoA product **86** (Figure 5.5 **C**). The largest peak in the ES^- spectrum at 505.9 Da is the mass of ATP which was added in all three steps. All three reactions worked as expected, showing that all three biosynthetic enzymes are active, leading to the desired CoA linker.

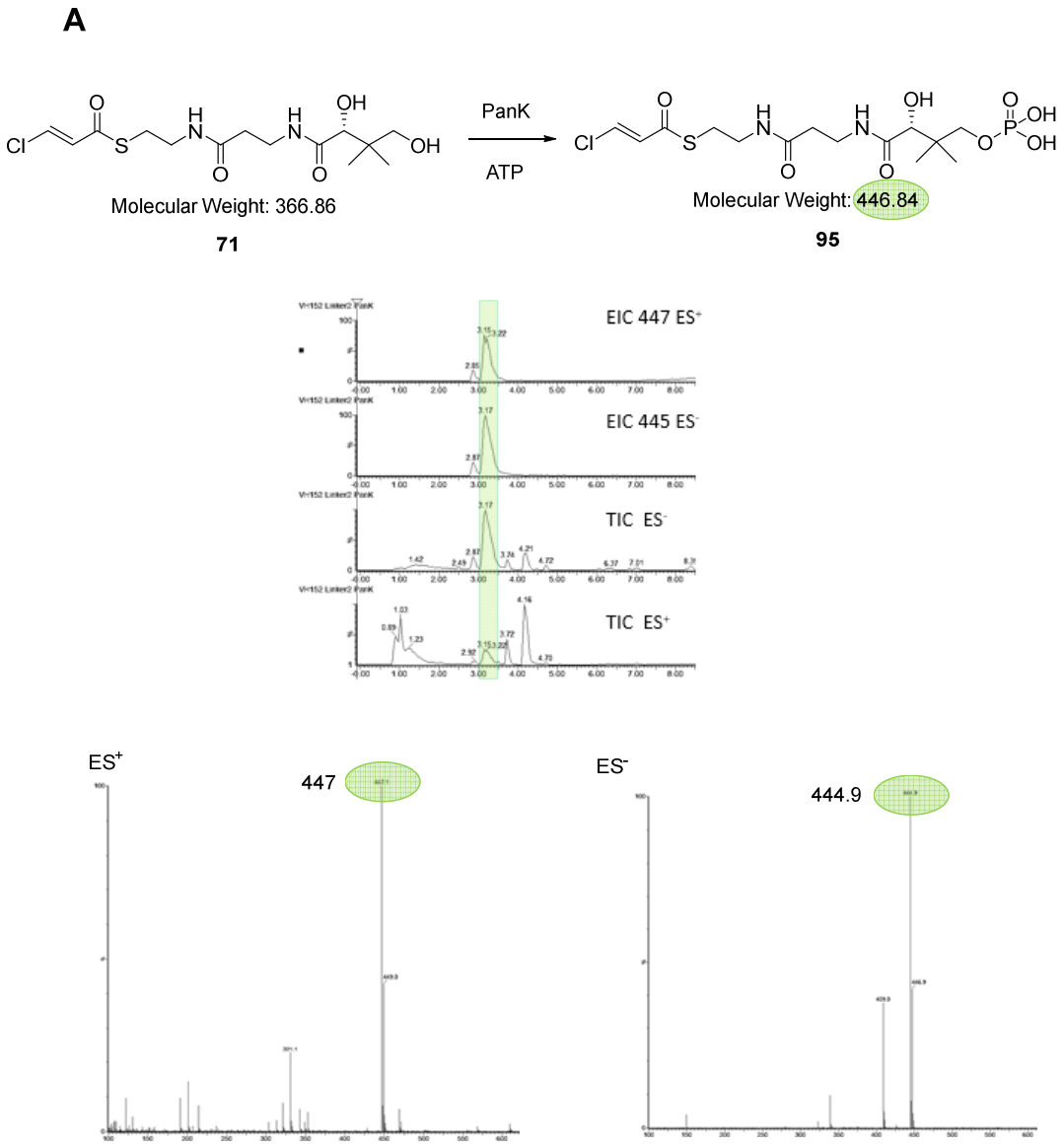


Figure 5.5: Biosynthetic reactions of the linker 71 monitored by LCMS: **A**, reaction catalyzed by PanK.

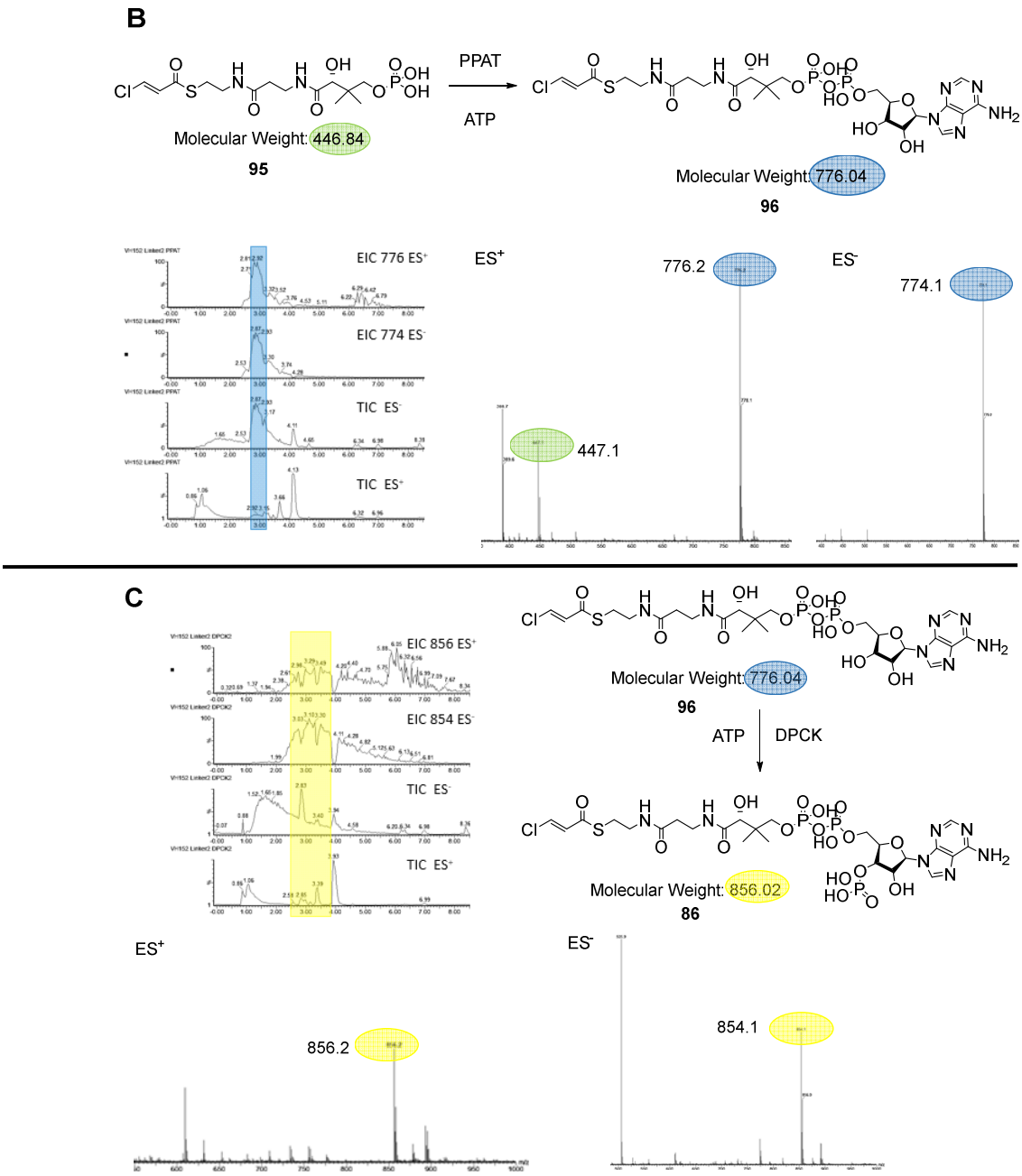


Figure 5.5: Biosynthetic reactions of the linker **71** monitored by LCMS: **B**, reaction catalyzed by PPAT; **C**, reaction catalyzed by DPCK.

In the next step the CoA linker **86** was loaded to the bfPKS *apo* ACP. For this 1.5 mg/ml *apo* ACP, 0.01 mg/ml Sfp and approx. 0.2 mg/ml CoA linker **86** were incubated at 30 °C for 1 h. Unfortunately, the linker reacted with the ACP during the loading in an unexpected way. The mass spectrum of the product (Figure 5.6) shows three main species are present. The expected mass for linker-ACP 1 is 12.520 kDa. The mass (12.521 kDa) was found after an incubation of 1 h at 30 °C as the lower concentrated species (Figure 5.6 **A**). The main peak has a measured mass of 12.482 kDa, which is 39 Da lower than

expected. The loss of 39 mass units may correspond to the loss of HCl (38 mass units for the heaviest isotopomer) The third peak belongs to *holo*-ACP. An experiment with a lower incubation temperature (22 °C and the same concentrations as previously) and shorter incubation time (20 min) showed that the amount of linker-ACP 1 is higher compared to the amount of the degraded species (Figure 5.6 **B**). It appears that crosslinking reactions need long incubation times, and for this reason a stable linker is necessary. The 12.482 kDa species might correspond to a propynoyl phosphopantetheine, or it could also arise by cross-linking to an ACP-derived nucleophile and loss of HCl. Lack of certainty meant that further use of this species was not followed further.

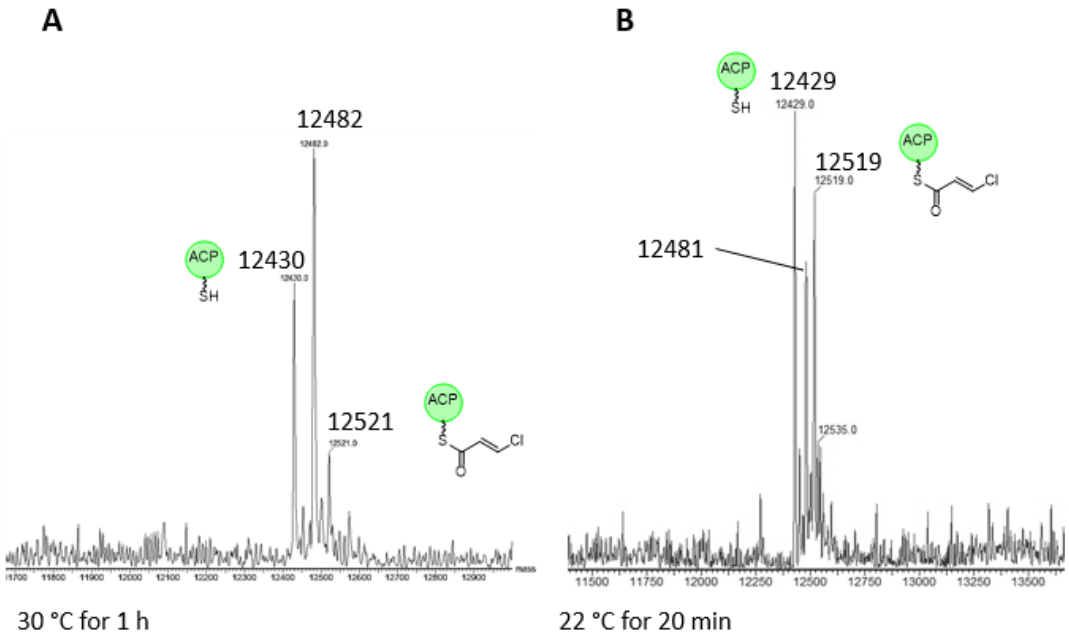


Figure 5.6: Deconvoluted mass spectra from the loading reaction of bfPKS *apo* ACP with linker **86** at different conditions: **A**, 1 h and 30 °C; **B**, 20 min at 22 °C.

Incubations of linker **86** with *apo* ACP only, and Sfp only, did not show any loss of HCl by mass spectrometry. The mass change happened only during the loading reaction. Probably the linker **86** is forced to be in a position where a reaction with neighbouring residues takes place.

The next three linkers **87**, **89** and **91** were also converted to CoA substrates and loaded onto the bfPKS *apo* ACP in one step. A one pot reaction was set up leading to the loaded ACP species. 50 mM Tris (pH 7.5), 20 mM KCl, 10 mM $MgCl_2$, 15 mM ATP, 0.2 mg/mL PanK, 0.2 mg/ml PPAT, 0.2 mg/ml DPCK, 1.5 mg/ml bfPKS *apo* ACP, 0.01 mg/ml Sfp and 1 mg/ml of linker **87** were mixed and incubated for 3 – 4 h at 30 °C.

The loading reaction and consequently the CoA forming reactions were successful (Figure 5.7). A repurification of the linker-ACP species was not performed. Higher protein and linker concentrations were necessary for a purification of the loaded protein, which was not possible due to precipitations of the biosynthetic enzymes. The deconvoluted protein mass spectrum shows a main peak with the mass of 12.482 kDa (calculated: 12.484 kDa) which corresponds to the expected linker-ACP 2. Two lower concentrated protein species (12.504 kDa and 12.524 kDa) are sodium adducts of the main protein species.

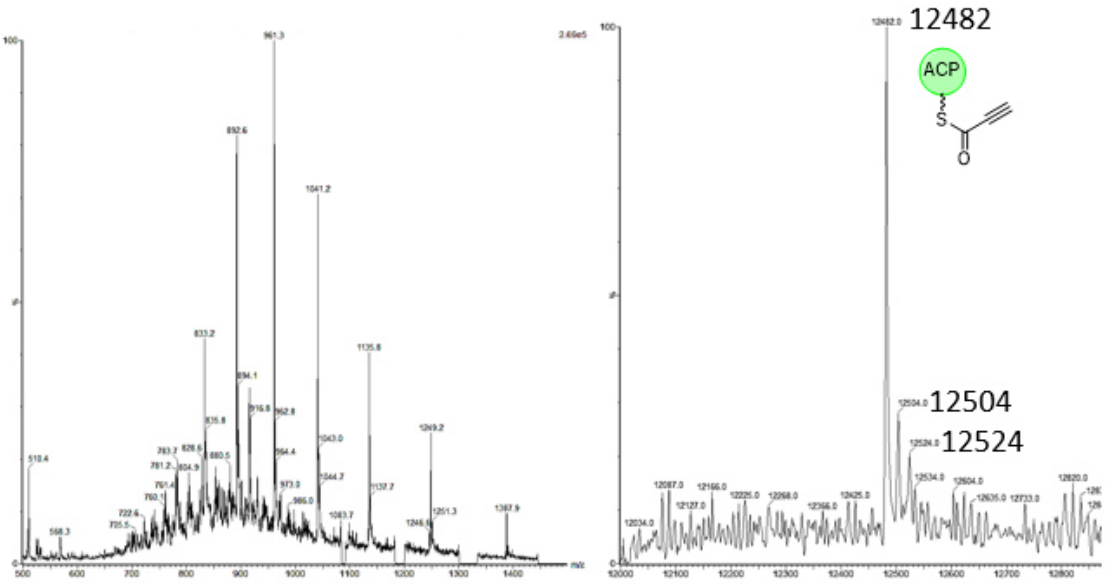


Figure 5.7: Mass spectrum after the one pot reaction with linker **87** resulting in linker-ACP 2.

For both sulfone linkers **89** and **91** the concentration of 1 mg/ml pantetheine substrate was too high and led to the precipitation of PPAT, and in consequence to an unsuccessful loading reaction. For the one pot reaction all components were used in the same concentration as mentioned above except of the substrate concentration. The linker **89** gave at the concentration of 0.1 mg/ml a successful loading reaction (Figure 5.8). The mass spectrum shows a protein species with the highest concentration at 12.531 kDa which corresponds to the expected mass of linker-ACP 3 (calculated: 12.530 kDa). The second largest peak at 12.554 kDa belongs to a sodium adduct of linker-ACP 3. The mass of 12.091 kDa is the mass of unreacted *apo*-ACP.

The one pot reaction of linker **91** needed even more trials to find out a concentration in which no protein precipitation took place. As for the previous linker the PPAT protein was the one which precipitated at higher substrate concentrations. Linker

91 concentration of 0.05 mg/ml led to a successful loading in the one pot reaction (Figure 5.9). The calculated size of the linker-ACP 4 is 12.544 kDa and was observed in the spectrum as the highest peak (12.543 kDa). The second peak of 12.090 kDa belongs to unreacted *apo* ACP (calculated: 12.090 kDa).

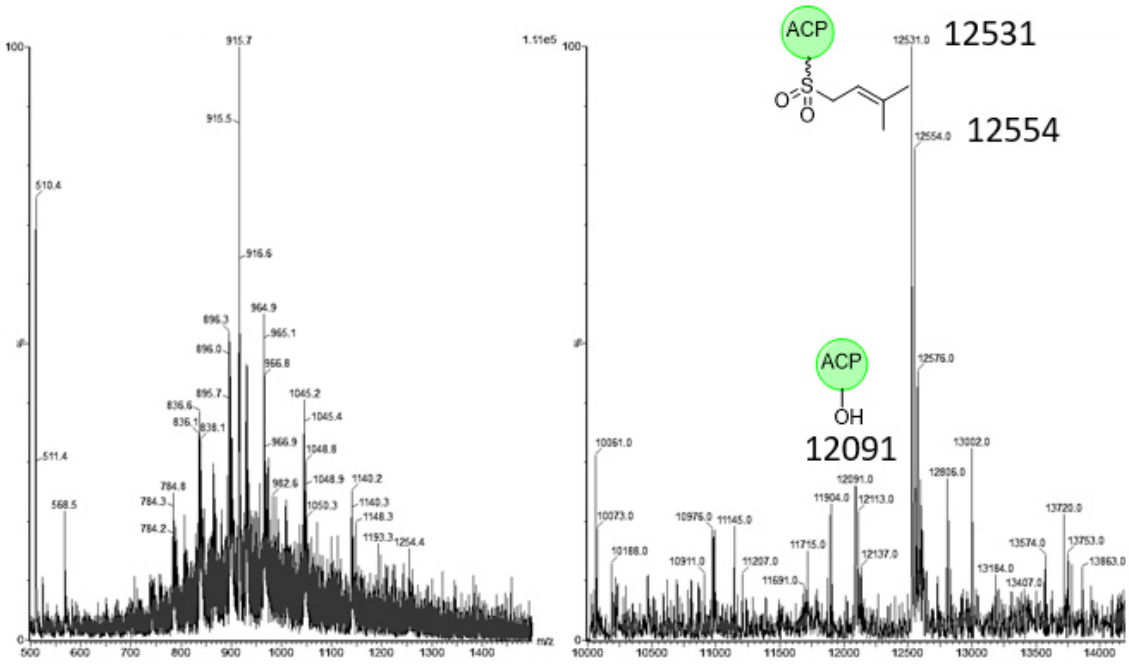


Figure 5.8: Mass spectrum after the one pot reaction with linker **89** resulting in linker-ACP 3.

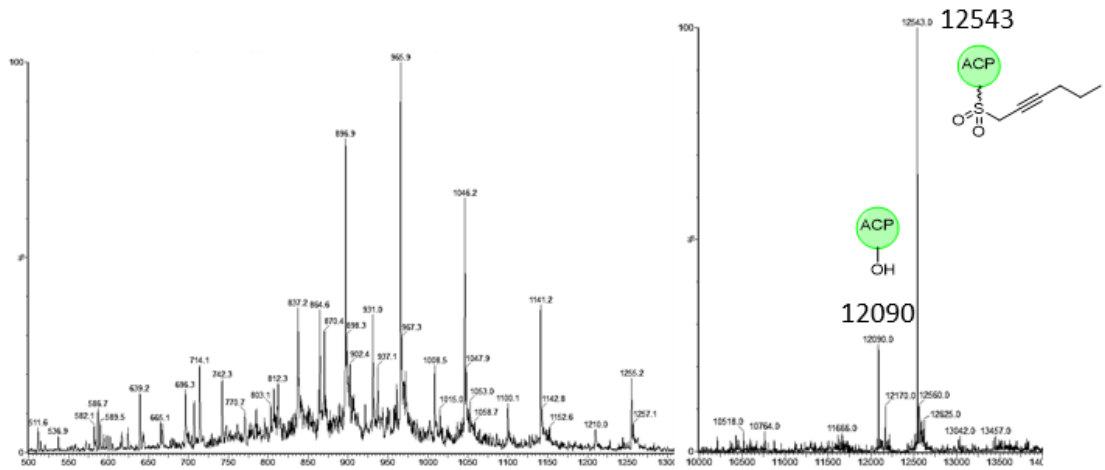


Figure 5.9: Mass spectrum after the one pot reaction with linker 91 resulting in linker-ACP 4.

5.4 Discussion

The three biosynthetic enzymes PanK, PPAT and DPCK were expressed and purified successfully and were characterized by SDS-PAGE and mass spectrometry. The proteins were pure and showed clean spectra. To check their catalytic activity, the first available linker **71** was used. The linker **71** was converted stepwise from an acyl pantetheine to an acyl CoA product. Each step was followed by LCMS. The expected masses were detected in positive and negative ionization modes and confirmed the activity of the recombinant PanK, PPAT and DPCK proteins. All three enzymes showed catalytic activity for more than two months when stored at -20 °C. The obtained CoA linker **86** was loaded using Sfp to the bfPKS ACP in a separate reaction. During the loading reaction the mass of the linker-ACP 1 decreased by an amount corresponding to HCl for an unknown reason. The mass change happened only after the loading reaction and not with the proteins separately. This leads to the assumption that the linker reacts with some surface residues in close proximity. However, uncertainty about the precise structure of this species meant that further work with it was not done.

The linkers **87**, **89** and **91** were also converted from pantetheine to CoA analogues and loaded to the ACP in one step using a one pot reaction. The one pot reaction contains all four biosynthetic enzymes (PanK, PPAT, DPCK and Sfp) to convert the pantetheine linker to the CoA linker and load it to the bfPKS ACP simultaneously. First, for the one pot reaction the same reaction conditions were used as described in literature by the group of Wright.[152] Relatively simple pantetheine analogues were transferred in this work with the help of PanK, PPAT and DPCK into CoA analogues. A transfer to an ACP was not performed in the previously published work. These conditions needed to be adjusted due to protein precipitations in some circumstances. Pantetheine linker concentrations were lowered to avoid these precipitations. Especially the enzyme PPAT was vulnerable to precipitations when higher linker concentrations were used. All of the linkers were successfully loaded to the bfPKS ACP and were available for crosslinking reactions. All aims for this part of the project were achieved.

6. SQTKS *C*-Methyltransferase domain

6.1 Introduction

During the biosynthesis of the squalestatin tetraketide **49** (SQTK) the *C*-Methyltransferase (*C*-MeT) domain performs its first methylation at the α-carbon atom (Figure 6.1 **A**) after the carbon-carbon bond forming reaction by the KS domain. A second methylation takes place in the next elongation cycle (Figure 6.1 **B**). In the third elongation cycle the *C*-MeT domain is inactive and no methylation takes place. This special "programmed behaviour" of the domain is a controlling step during the biosynthesis. For a full understanding of the SQTK biosynthesis it is crucial to analyse the *C*-MeT domain, and its interaction with ACP-bound intermediates in as much detail as possible.

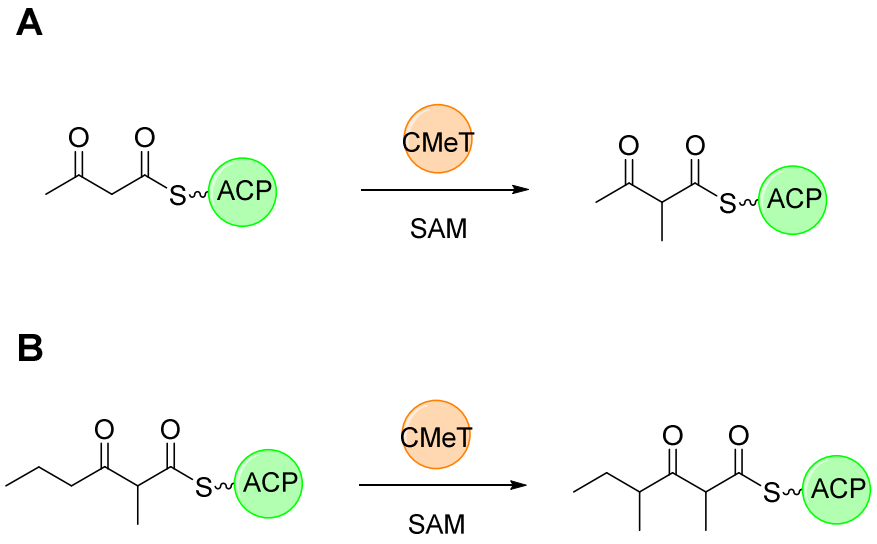
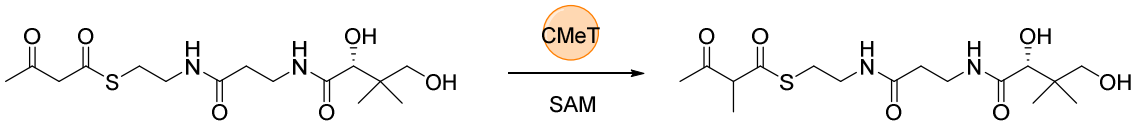


Figure 6.1: SQTKS *C*-MeT catalytic activity during the SQTK biosynthesis: **A**, first cycle; **B**, second cycle.

Previous work of Hao Yao showed that the SQTKS *C*-MeT domain had no or minor activity as an isolated single domain protein as well as in multidomain constructs.[100] The substrates used were SNACs and pantetheines. The *C*-MeT domain was expressed as an isolated domain also included in multidomain constructs including the DH-*C*-MeT-ER tridomain and DH-*C*-MeT-ER-KR tetradomain. The isolated *C*-MeT and the DH-*C*-MeT-ER-KR tetradomain showed a low level methylation of the pantetheine diketide **97** after a long incubation time of 2 days (Figure 6.2). However, SNAC-bound substrates were not converted to methylated products and the isolated *C*-MeT and the DH-*C*-MeT-ER-KR tetradomain were inactive with these substrates.



97 98

Figure 6.2: SQTKS *C*-MeT catalytic reaction with diketide-pantetheine substrate **97** to form the methylated product **98**.

A possible explanation for the low activity might be the missing ACP. In experiments performed by Hao Yao the natural substrates of *C*-MeT were offered as SNAC or pantetheine derivatives. It is possible, that the *C*-MeT domain has a high substrate selectivity and accepts only ACP bound substrates. However, the inactivity of the SQTKS ACP has prevented further work in this direction.

Concerning the stereoselectivity of the SQTKS *C*-MeT domain there is still no clear evidence of its preference. Due to the spontaneous epimerisation of the *C*-MeT product it is hard to elucidate the absolute configuration of the initial product. However, Hao Yao showed that this spontaneous epimerization reaction is slow under physiological conditions. Therefore, if the *C*-MeT reaction could be observed over a short time period it might also be possible to determine the stereoselectivity of the *C*-MeT.

6.2 Aims

The main aim of this project is to incubate an ACP bound β-ketothiolester with the *C*-MeT domain and check if the activity of the catalytic domain is increased compared to previous experiments with SNAC and pantetheine analogues. In the first step a pantetheine derivative will be synthesized. Next, the pantetheine derivative will be converted to a CoA analogue using PanK, PPAT and DPCK. In the third step the CoA analogue will be loaded to tag free bfPKS ACP and incubated in the last step with the SQTKS *C*-MeT domain to check their activity.

6.3 Results

6.3.1 Substrate Synthesis and ACP Loading Reaction

The SQTKS *C*-MeT domain uses acetoacetyl-ACP as the substrate for the first methylation. During the reaction the α-carbon is methylated (Figure 6.3)

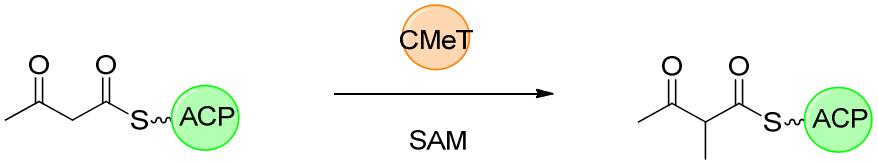


Figure 6.3: Methylation reaction of ACP bound acetoacetyl moiety by the SQTKS *C*-MeT domain.

The substrate acetoacetyl-pantetheine **97** was synthesized by Oliver Piech. The compound was solubilized in the reaction buffer containing 20 mM KCl, 10 mM $MgCl_2$, 50 mM Tris-Cl (pH 7.5) and 5 mM ATP. The mixture was incubated with PanK at 30 °C for 1 h and the reaction was followed by LCMS. Best results were obtained with 0.6 mg/ml **97**. Higher substrate **97** concentration (1 mg/ml) led to precipitation of PanK and poor conversion. The calculated mass (442 Da) of **99** was observed in positive as well as negative ionization mode (Figure 6.4 **A**). The negative ionization mode shows, additionally to the product **99**, a peak at 357 Da. This is the phosphopantetheine resulting from the reaction of pantetheine with ATP and catalysed by PanK. Pantetheine is an impurity from the chemical synthesis and could not be separated from acetoacetyl-pantetheine.

The second enzyme, PPAT, adds an adenyl moiety to acetoacetyl-phosphopantetheine **99**. For this reaction a mass shift of 329 Da is expected. **99** was incubated with PPAT and ATP at 30 °C for 1 h and analysed by LCMS. Positive and negative ionization modes confirmed the formation of the expected product **100** (Figure 6.4 **B**). The third reaction is performed by DPCK. This enzyme adds the final phosphate group at the adenyl moiety and forms the desired CoA product **101**. The third reaction was performed under the same conditions as the previous. The reaction worked as expected and the product mass was again observed in both positive and negative ionization modes (Figure 6.4 **C**). The large peak at 505.9 Da in the negative ionization mode belongs to ATP which was added in each reaction step.

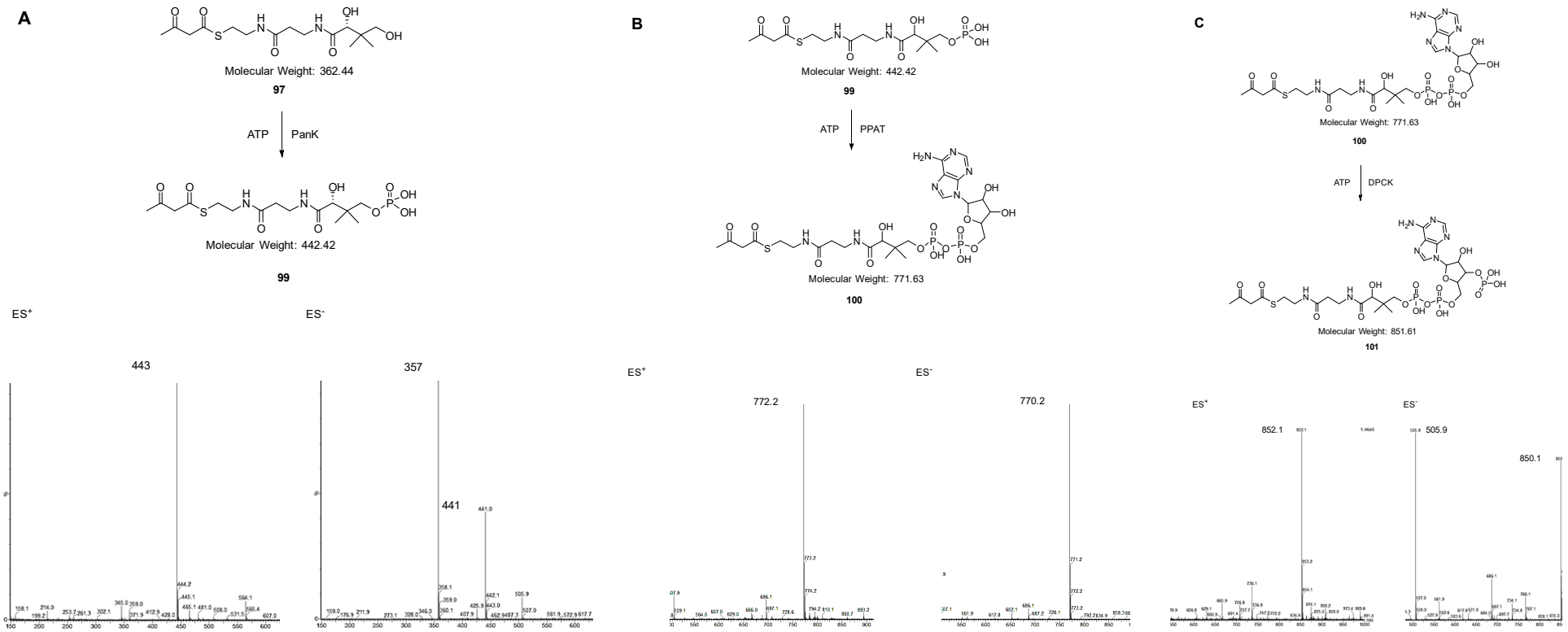
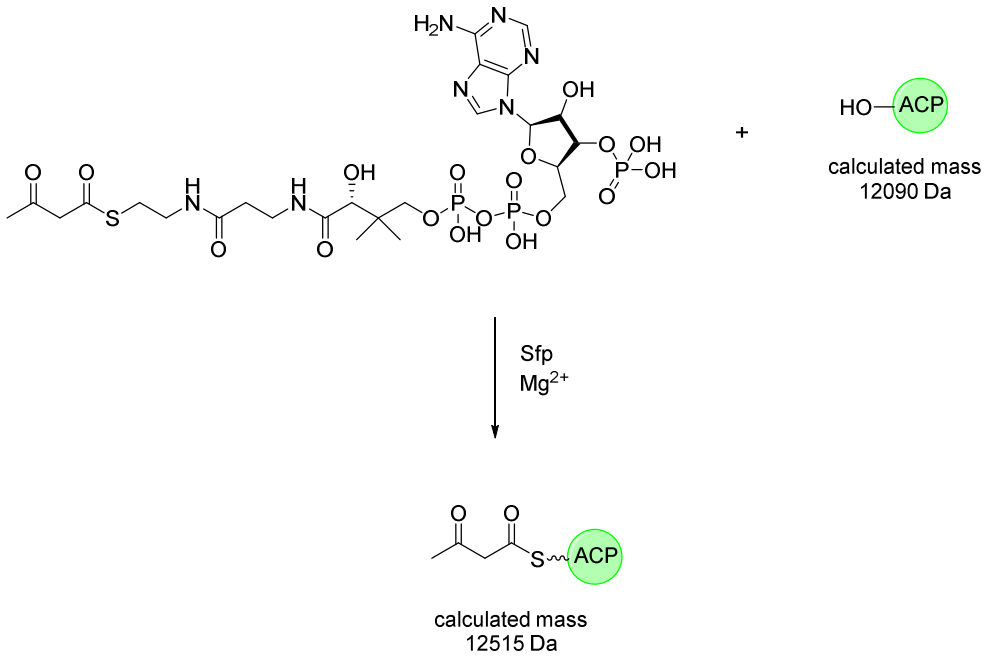


Figure 6.4: Transformation of acetoacetyl-pantetheine **97** to the CoA analogue **101** stepwise: **A**, PanK reaction; **B**, PPAT reaction; **C**, DPCK reaction.

With acetoacetyl-CoA **101** in hand, the loading to the tag-free bfPKS ACP was performed. For this 1.5 mg/ml ACP were used and incubated at 30 °C for 30 min with acetoacetyl-CoA **101** and Sfp (0.01 mg/ml). The transfer of the acetoacetyl moiety was successful and complete (Figure 6.5) as analysed by MS. No *apo*-ACP was detected (12.090 kDa). A new mass of 12.511 kDa (calculated: 12.515 kDa) was observed which corresponds to acetoacetyl-ACP and a minor amount of *holo*-ACP at 12.429 kDa (calculated: 12.430 kDa). *Holo*-ACP might be formed *via* hydrolysis of acetoacetyl-ACP or *via* loading of CoA which might be formed through the impurity with pantetheine mentioned before. This protein was then used for the activity test of the SQTKS *C*-MeT catalytic domain.



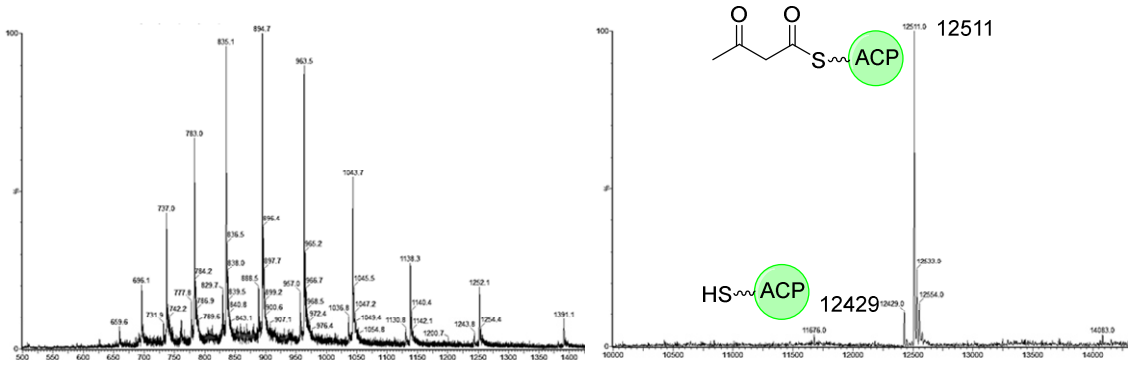


Figure 6.5: bfPKS ACP loading with acetoacetyl-CoA catalyzed by Sfp.

6.3.2 *C*-Methyltransferase Activity Check

As already mentioned, the SQTKS *C*-MeT domain was mostly inactive in previous experiments using single and multidomain constructs and SNAC and pantetheine substrates. The missing carrier protein might be the reason for not recognizing the diketide as a substrate. For the catalytic reaction the following components are necessary: acetoacetyl-ACP, SAM and the *C*-MeT domain. The isolated *C*-MeT was heterologously expressed by Oliver Piech. The protein had several impurities (Figure 6.6).

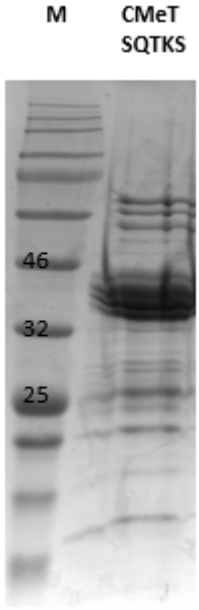


Figure 6.6: Heterologously expressed isolated SQTKS *C*-MeT.

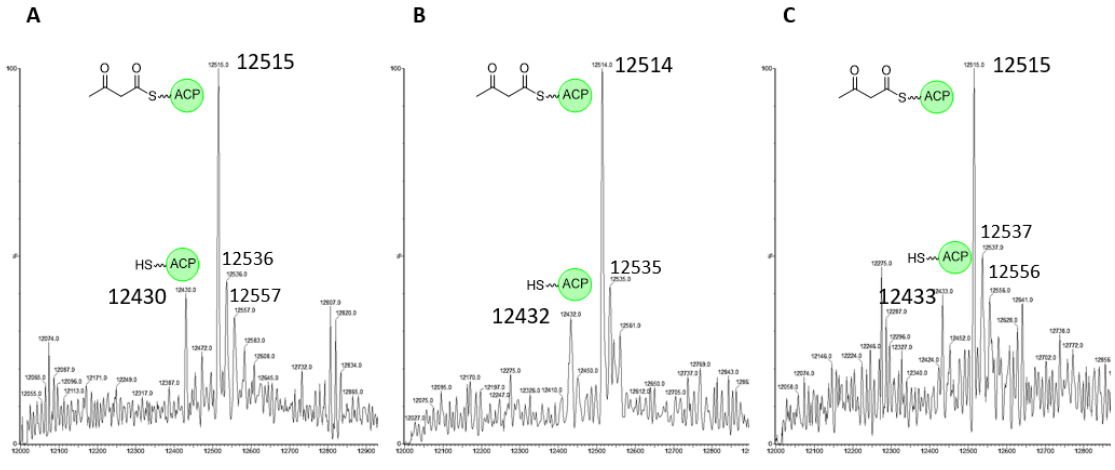


Figure 6.7: SQTKS *C*-MeT activity test including controls, all incubated for 1 h at 30 °C: **A**, acetoacetyl-ACP; **B**, acetoacetyl-ACP + SAM; **C**, acetoacetyl-ACP + SAM + *C*-MeT.

Three experiments were performed in parallel. All of them were incubated at 30 °C for 1 h. The first one contained acetoacetyl-ACP only (Figure 6.7 **A**). The second contained

acetoacetyl-ACP and SAM (Figure 6.7 **B**) and the third one acetoacetyl-ACP, SAM and *C*-MeT (Figure 6.7 **C**). With an active *C*-MeT domain we would expect an addition of 15 Da for the methyl group leading to a calculated mass of 12.530 kDa. All three experiments were analysed by ESMS. In all experiments there is a similar result showing acetoacetyl-ACP (calculated: 12.515 kDa) as the main peak and minor amounts of *holo*-ACP. Next, peaks which are closest to the mass of the expected species are sodium adducts of acetoacetyl-ACP. No methylation of the ACP bound substrate was observed. A long term incubation of *C*-MeT with acetoacetyl-ACP, like it was performed with pantetheine substrates in previous experiments, was not possible. The background noise of the MS protein measurements increased in a long term incubation significantly and made an analysis impossible.

6.4 Discussion

The synthesis of the desired CoA substrate for the *C*-MeT domain was performed successfully using *E. coli* CoA biosynthetic enzymes. Starting with acetoacetyl-pantetheine **97** the compound was converted in three steps using PanK, PPAT and DPCK into the CoA analogue. Each of the three steps leading to the CoA product were followed by LCMS. All of the expected masses were observed by positive and negative ionization modes. The acetoacetyl-CoA was then loaded on the tag free bfPKS ACP, which was catalysed by Sfp. The succes of the loading step was confirmed by protein mass spectrometry. With the acetoacetyl-ACP in hand, which is the natural substrate of the *C*-MeT, a substrate conversion test was set up. Methylation of the ACP bound substrate was not observed by protein mass spectrometry and there was no significant difference compared to the negative controls. The *C*-MeT domain is for an unknown reason still inactive.

The collegue Hao Yao showed that SNAC-derived and pantetheine-derived substrates were not methylated (except long term incubation) by the SQTKS *C*-MeT standalone domain as well as by a *C*-MeT in a multidomain environment. It was assumed that this domain has a high substrate selectivity and may need an ACP bound substrate for a proper recognition. We have shown that also an ACP-bound substrate did not improve the activity of the SQTKS *C*-MeT domain and it is unlikely that a high substrate selectivity could explain the inactivity. There might be an incorrect folding of the *C*-MeT

as a standalone domain and also as a multidomain protein during expression in *E. coli*, which could explain the inactivity.

The ACP which is used for the experiments belongs, like the SQTKS, to a fungal highly reducing iterative PKS. We would expect that the carrier protein is similar enough to the SQTKS ACP to show at least low activity in the assay. The usage of the SQTKS ACP, which belongs to the same pathway as the *C*-MeT is coming from, would be optimal but could not be expressed active as described in chapter 3.

For future projects it would be beneficial to use the SQTKS ACP for offering the loaded substrate from the corresponding protein. As described in chapter 3 several problems need to be fixed to make this happen. Another aim for a future project will be the expression the SQTKS *C*-MeT monodomain and SQTKS multidomain constructs in fungal systems. The origin of the SQTKS lies in a fungal system and the expression in a similar system might influence the folding of the domains which could lead to an increased activity of the SQTKS *C*-MeT. It will be very difficult to have high protein amounts, in a fungal heterologous expression, for purification and *in vitro* experiments.

7. Crosslinking of the bfPKS Acyl Carrier Protein with a Dehydratase Domain

7.1 Introduction

Due to the transient nature of interactions between PKS catalytic domains and the ACP it is very challenging to gain structural information of these interactions. In this section two examples of successful linkings between an ACP and a DH domain are presented. The group around Burkhart have reported the successful linkage between the ACP (AcpP) and the DH (FabA) domain from the *E. coli* Type II FAS.[36,150] The pantetheine linker **102** was chemically synthesized and converted to a CoA substrate by the biosynthetic enzymes PanK, PPAT and DPCK. Finally, the CoA substrate was loaded to the ACP by a PPTase (Figure 7.1 **A)** and crosslinked to the DH domain.

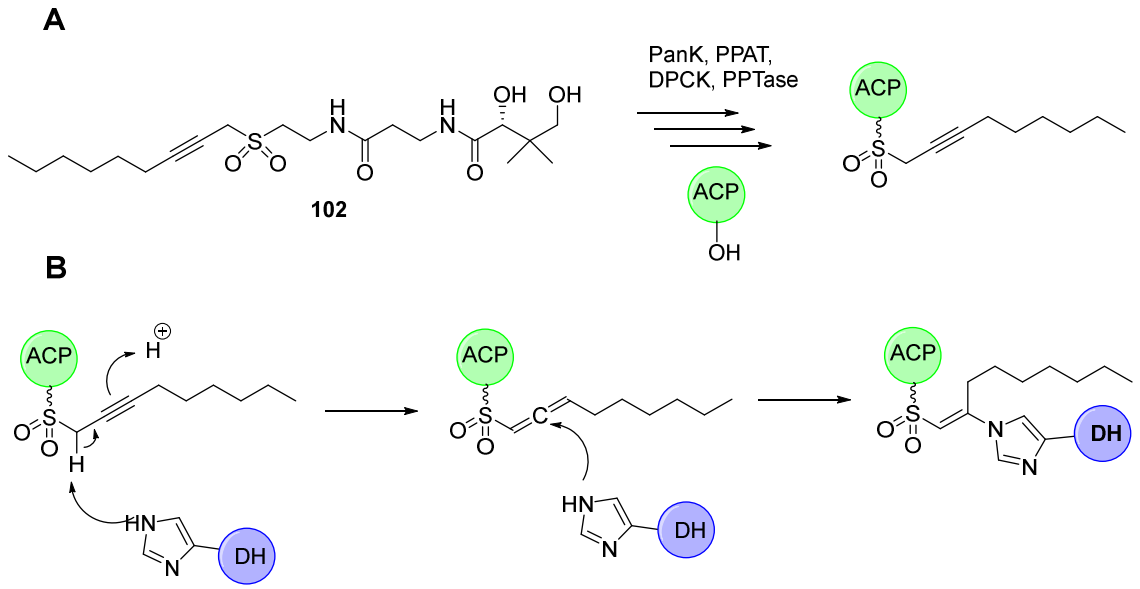


Figure 7.1: Biosynthetic steps to reach crosslinking between FabA and AcpP: **A**, transformation of the linker **102** to a CoA analogue and the subsequent loading to AcpP; **B**, proposed mechanism for the crosslinking reaction.[150]

The ACP bound crosslinker covalently and irreversibly connects to a highly conserved histidine of the active site of FabA *via* an allene intermediate (Figure 7.1 **B**). The Burkhart group characterized the FabA protein using the crosslinker reaction as a tool. They synthesized linkers with varying acyl chain length mimicking different fatty acid intermediates to probe for altered substrate specificities of the FabA mutants. The mutants were classified in three regions. The first region contained many positive charges and the reversal of them led to a large decrease in FabA activity. The second region affected the

length of the substrate tunnel and the third region contained “gating residues”, which had an influence on the discrimination between varying acyl chain length.[150]

Structural information was also obtained using the same linker method. A 1.9 Å crystal structure of crosslinked AcpP-FabA was reported as a homodimer in which AcpP showed two different conformations, representing probable snapshots of ACP in action (Figure 7.2). Solution nuclear magnetic resonance techniques gave additional information about residues at the interface.

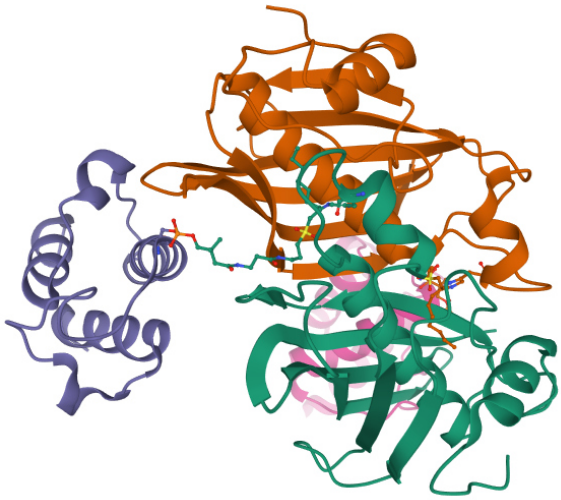
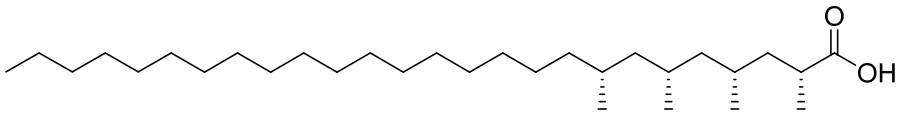


Figure 7.2: Structure of AcpP (PDB: 4KEH) crosslinked with FabA.

In a second successful crosslinking the Burkhart group linked an ACP together with a DH domain of a fully reducing iterative PKS, the mycoserosic acid PKS (MAS). The MAS PKS is closely related to modular PKS, with a complete modifying region, elongating linear C_{12} to C_{20} starter fatty acids to produce mycocerosic acid **103** (Figure 7.3).[154]



103

Figure 7.3: Structure of mycocerosic acid.

The group developed a system that can load the *apo*-ACP to its corresponding loaded species with an unreactive masked warhead and then reveal the warhead when appropriate.[155] The silylcyanohydrin is, compared to the sulfone, less reactive due to the

less acidic α-protons. This protects the reactive group from unwanted reactions prior crosslinking.

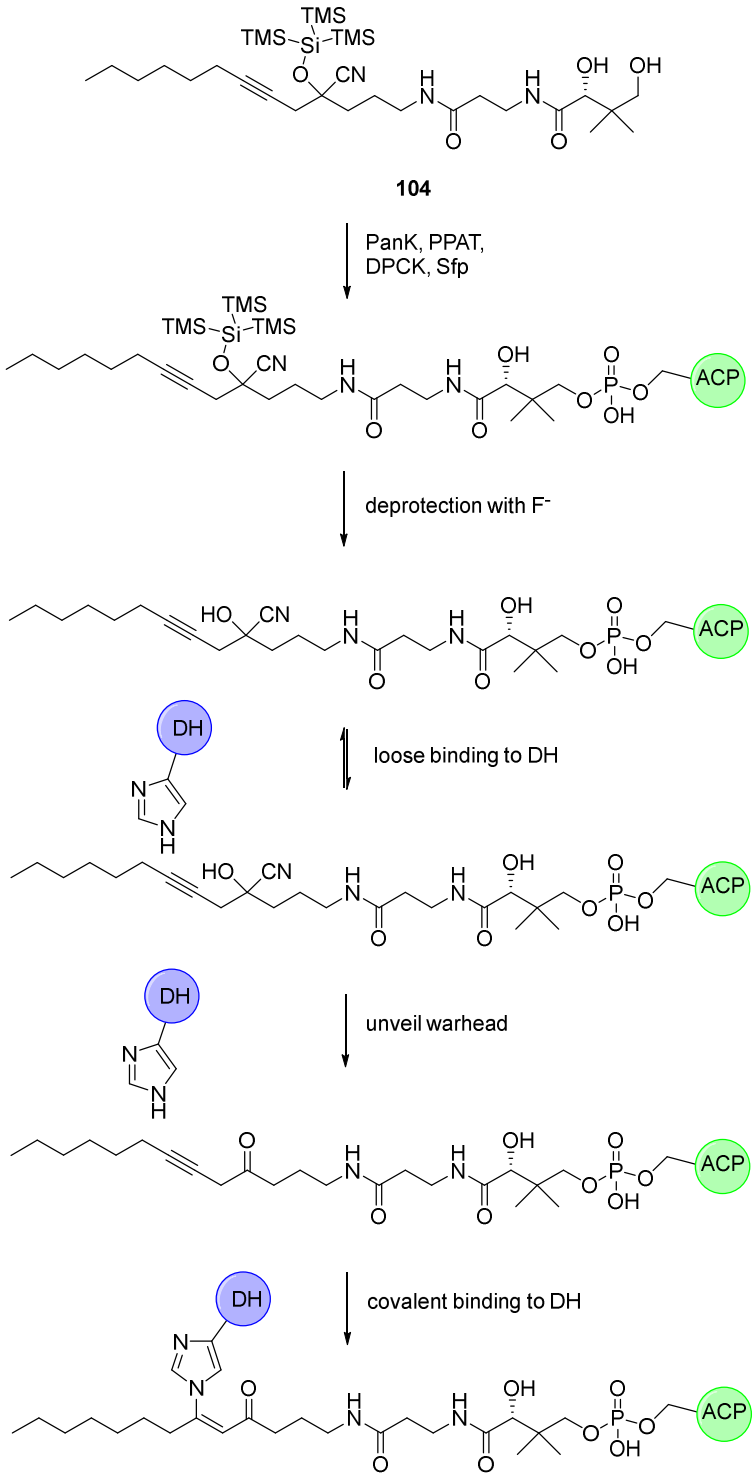


Figure 7.4: Crosslinking reaction with a trimethylsilyl-cyanohydrin linker **104**.[155]

After synthesis of the silylcyanohydrin linker **105** it was loaded to the MAS ACP using the one pot reaction described before. The silyl moiety was eliminated by addition of

fluoride leading to the cyanohydrin in the presence of the DH. The loss of HCN finally leads to the reactive species (acidic protons) which connects covalently to the DH (Figure 7.4).

7.2 Aims

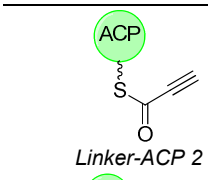
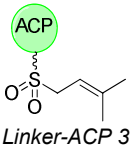
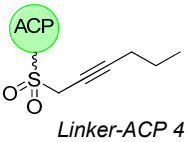
The aim for this project is to connect the bfPKS ACP to a DH domain for generating structural information by SAXS in collaboration with the Weismann group. There are several linker ACP constructs available and also DH domains which were expressed and tested for their activity (section 4.3.2). The crosslink reaction will be detected by mass spectrometry and SDS PAGE. A mass, which is the sum of the DH mass and the linker-ACP mass, is expected.

7.3 Results

7.3.1 Crosslinking Attempts of bfPKS ACP and DH Domains of Different Origins

Three of the four ACP (tag-free bfPKS ACP) linker constructs were synthesized successfully and were available for crosslinking reactions. As crosslinking "partners" three different DH domains were available. The expression of the SQTKS and bfPKS DH is described in chapter 4. The strobilurin DH domain was expressed and tested for correct folding by Dr. Daowan Lai. All three DH domains were catalytically active (showed conversion of the substrate, section 4.3.2). The combinations of ACP-linker constructs and DH domains, which were tested, are listed in table 7.1.

Table 7.1: Tested crosslinking reactions between different DH domains and linker-ACP constructs.

—	*bfPKS DH*	*SQTKS DH*	*St*PKS *DH*
 *Linker-ACP 2*	✓	✓	–
 *Linker-ACP 3*	✓	✓	✓
 *Linker-ACP 4*	✓	–	–

The first combination which was tested was the reaction between linker-ACP 2 and SQTKS DH. Approx. same concentration (10 µM) of linker-ACP 2 and SQTKS DH were incubated for 3 – 4 h at 25 °C. The protein mixture was analysed by mass spectrometry and SDS PAGE. The deconvoluted mass spectrum (Figure 7.5 **B**) showed the mass of linker ACP 2 at 12.482 kDa (calculated: 12.484 kDa) and also the mass of the SQTKS DH at 38.006 kDa (calculated: 38.014 kDa). The sum of both measured masses is 50.488 kDa. No protein species was detected in this mass range showing that the crosslinking reaction was not successful. The SDS PAGE (Figure 7.5 **C**) confirmed the MS measurement. Linker-ACP 2 and DH were observed on the gel separately and also in combination but no larger band was seen. On the SDS PAGE as well as on following gels a degradation product of the DH was observed which was already described by Emma Liddle.[98] This does not affect the activity of the DH.

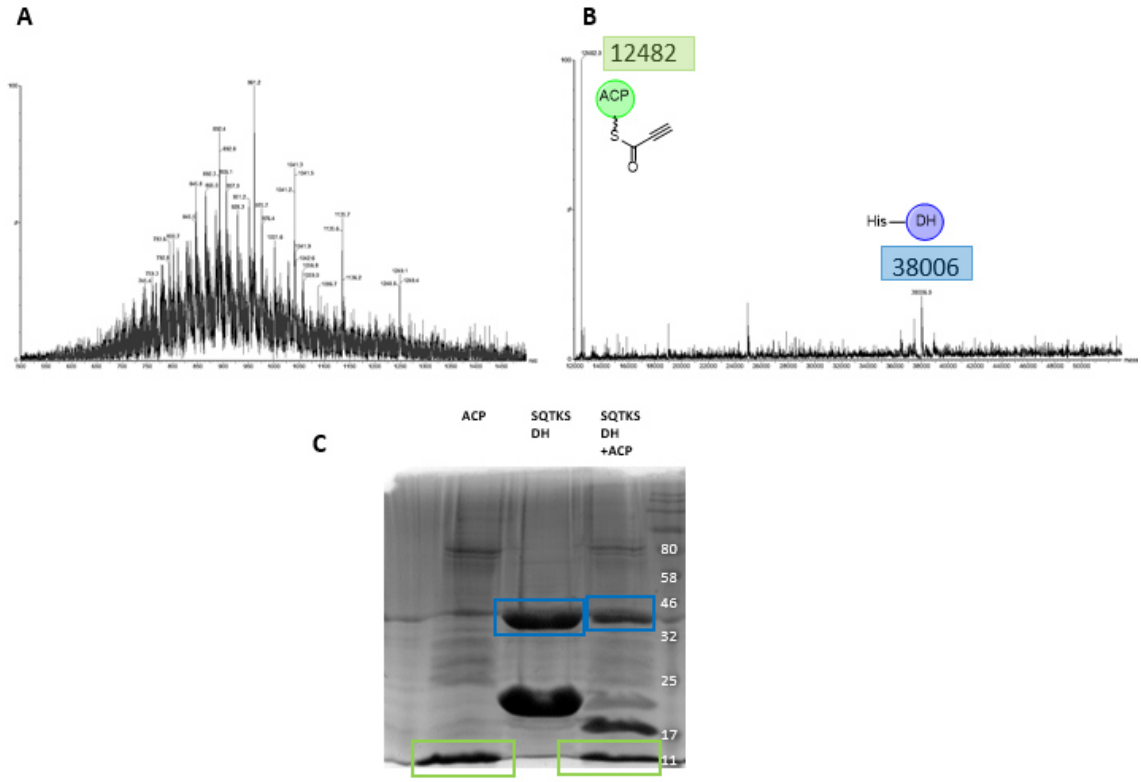


Figure 7.5: The crosslink reaction between linker-ACP 2 and SQTKS DH: **A**, mass spectrum; **B**, deconvoluted spectrum; **C**, SDS PAGE.

In the next experiment the SQTKS DH domain was incubated with linker-ACP 3. In the deconvoluted mass spectrum (Figure 7.6 **B**) there are two species of the SQTKS DH visible. One has the mass of 38.005 kDa and belongs to the DH including the His tag and the second species has the mass of 36.424 kDa which corresponds to the mass of the tag free DH. As already described in section 3.3.5 the ACP His tag was cut with thrombin

after the first purification step and a second purification followed. Apparently there is still active thrombin in the ACP sample. During the long incubation times of 3 – 4 h there is enough time to cut the His tag of the SQTKS protein partially. The spectrum showed the mass of linker-ACP 3 (12.530 kDa) as well as several calculation artefacts of the linker-ACP 3 and the DH. 18.212 kDa and 12.141 kDa are exactly the half and one third of the mass of tag free DH. 25.060 kDa is exactly the twofold size of linker-ACP 3 and the mass of 19.003 kDa is the half size of His tag DH. The mass of the crosslinked protein is 48.954 kDa for the tag free protein and 50.535 kDa for the DH including the His tag. No crosslinked protein was found in the mass spectrum and this was confirmed as in the SDS PAGE analysis (Figure 7.6 **C**).

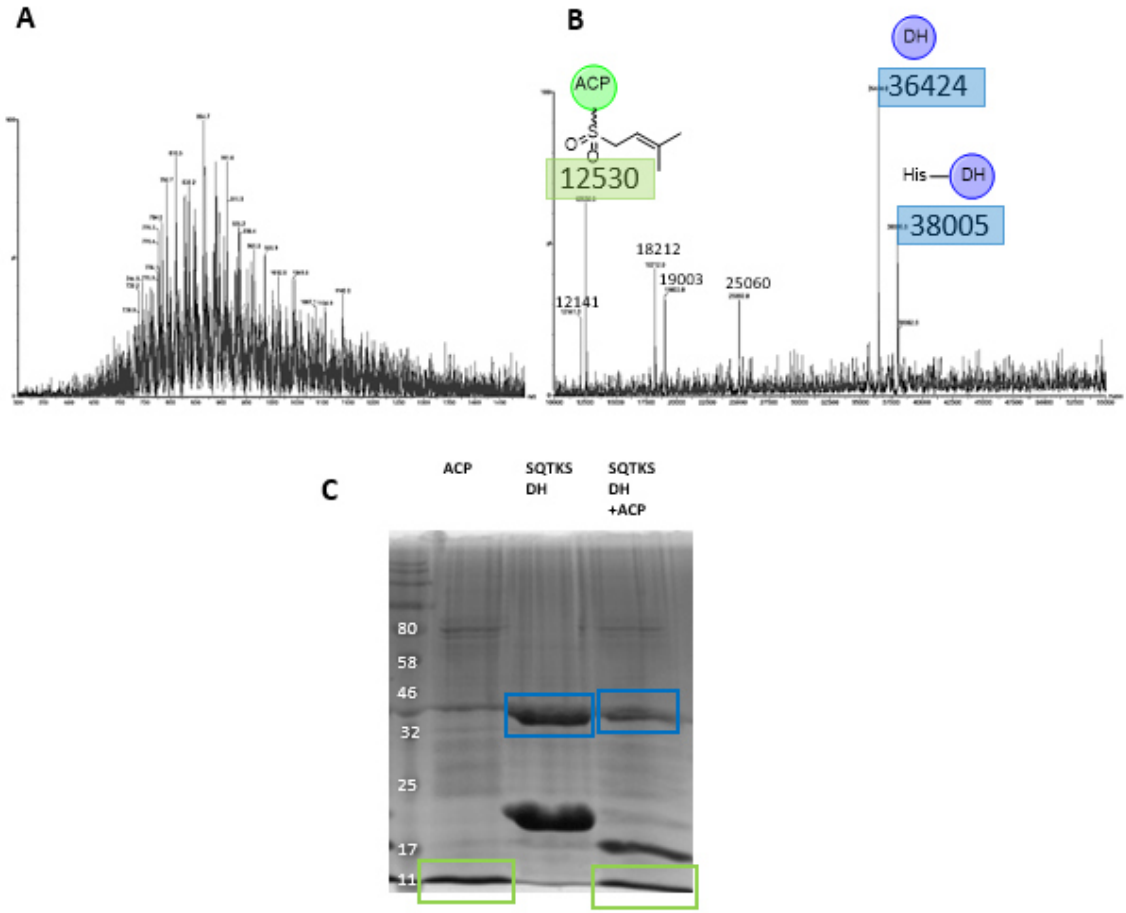


Figure 7.6: The crosslink reaction between linker-ACP 3 and SQTKS DH: **A**, mass spectrum; **B**, deconvoluted spectrum; **C**, SDS PAGE.

Next, the linker-ACP 3 was incubated with the stPKS DH. The deconvoluted mass spectrum (Figure 7.7 **B**) shows the mass of the tag free DH domain (30.057 kDa) and the mass of linker-ACP 3 (12.531 kDa). The combined mass of both is 42.588 kDa. Again no crosslinking reaction was observed, confirmed by the SDS PAGE analysis

(Figure 7.7 **C**). Three other peaks in the MS spectrum are calculation artefacts of the two proteins (6.266 kDa = ACP/2; 10.019 kDa = DH/3; 15.028 kDa = DH/2).

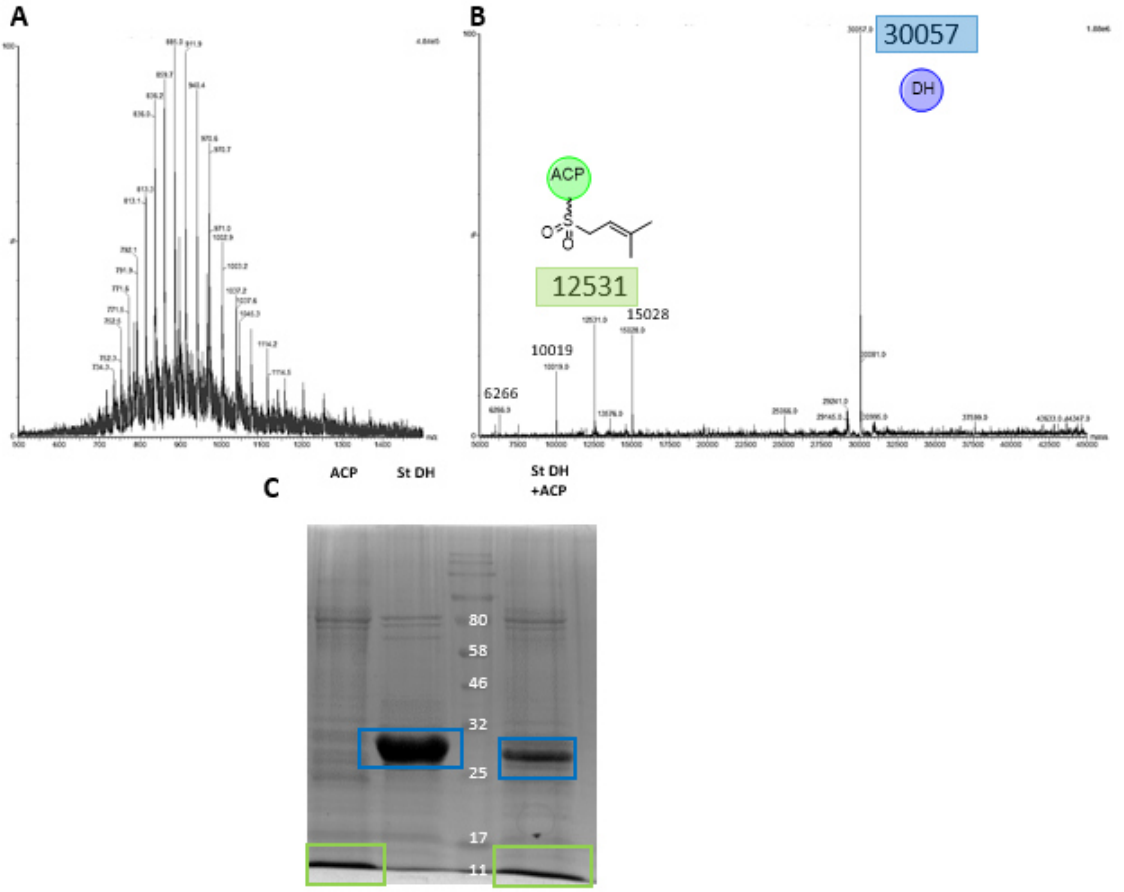


Figure 7.7: The crosslink reaction between linker-ACP 3 and stPKS DH: **A**, mass spectrum; **B**, deconvoluted spectrum; **C**, SDS PAGE.

The ACP and DH domains of the last three experiments were from different biosynthetic pathways. The ACP comes from the byssochlamic acid pathway and the DHs from the squalestatin and strobilurin pathways. Even if they are all from fungal highly reducing iterative systems they might not interact with each other.

For this reason, we tested a DH domain which comes from the same PKS as the ACP. The crosslinking reaction between linker-ACP 2 and the tag free bfPKS DH was tested under the same conditions as previously. Linker-ACP 2 (12.482 kDa) and the bfPKS DH (39.634 kDa) were observed in the deconvoluted mass spectra (Figure 7.8 **B**). Two calculation artefacts of the DH are also present (13.212 kDa = DH/3; 19.817 kDa = DH/2). The combined mass of both proteins (52.116 kDa) was found neither in the mass spectrum nor in the SDS PAGE analysis (Figure 7.8 **C**).

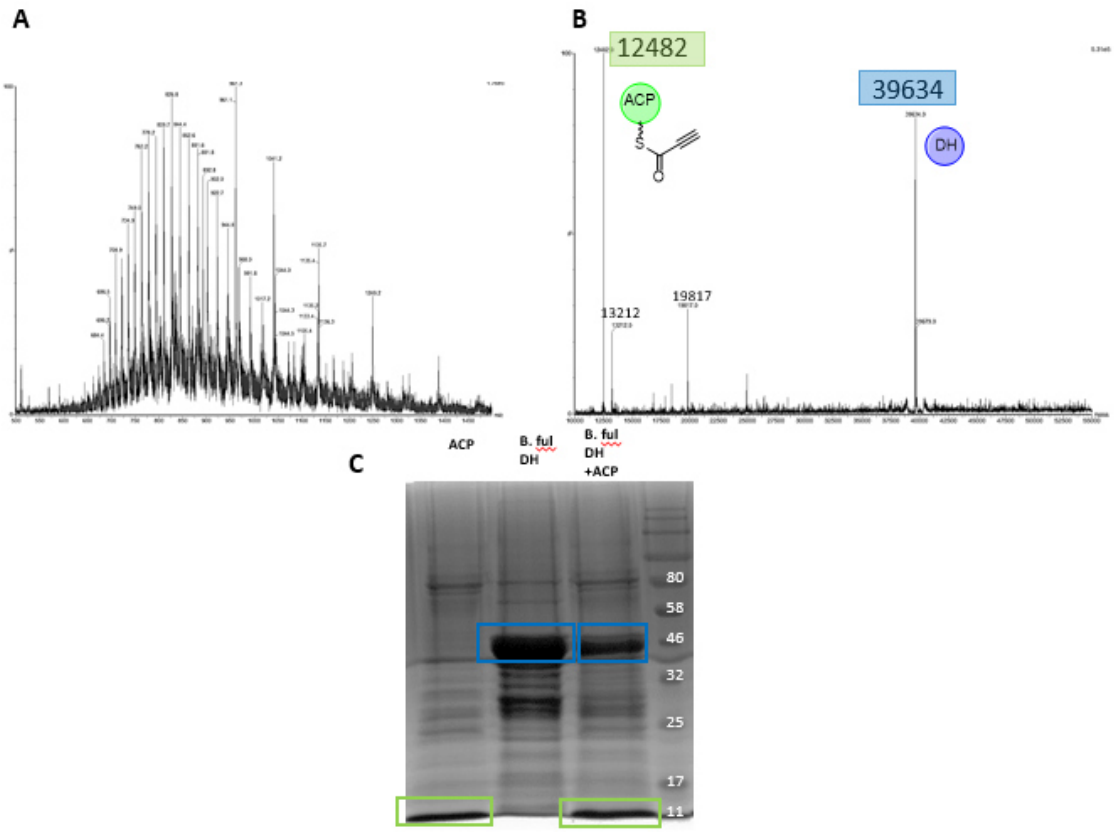


Figure 7.8: The crosslink reaction between linker-ACP 2 and bfPKS DH: **A**, mass spectrum; **B**, deconvoluted spectrum; **C**, SDS PAGE.

Also the next combination was not successful and did not show crosslinking. Linker-ACP 3 and bfPKS DH were incubated as previously described and were observed in the deconvoluted mass spectrum (Figure 7.9 **B**) at 12.530 kDa and 39.633 kDa. The combined mass of both proteins is 52.163 kDa, which was not present in the MS spectrum. The SDS PAGE confirmed the result (Figure 7.9 **C**).

The last tested combination was between linker-ACP 4 and bfPKS DH. Both proteins were observed in the mass spectrum (Figure 7.10 **B**) at 12.544 kDa and 39.634 kDa. The combined mass of 52.178 kDa was not observed. Also no band at the expected size was seen in the SDS-PAGE analysis (Figure 7.10 **C**).

Most of the experiments were also tried at long incubation times (24 h) but no protein species were detected due to precipitations. Unfortunately, the attempt to generate structural information with crosslinked proteins failed.

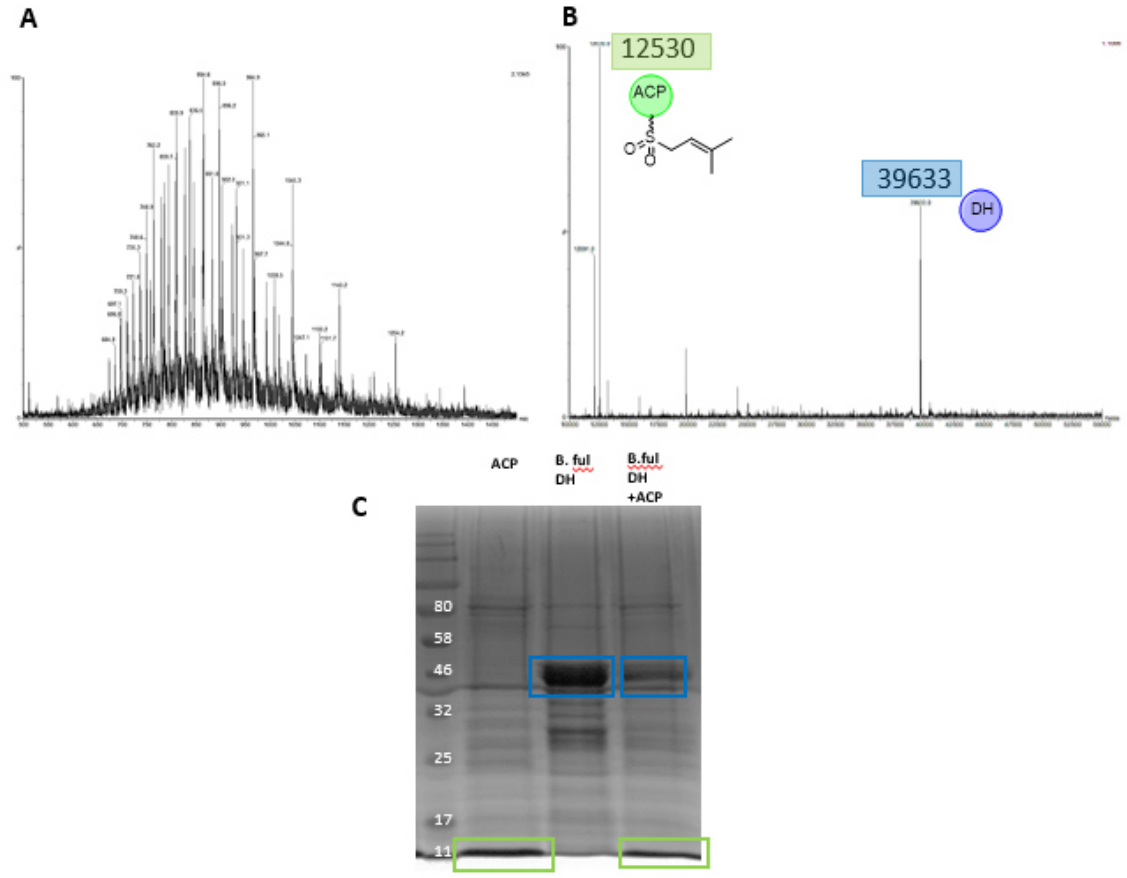


Figure 7.9: The crosslink reaction between linker-ACP 3 and bfPKS DH: **A**, mass spectrum; **B**, deconvoluted spectrum; **C**, SDS PAGE.

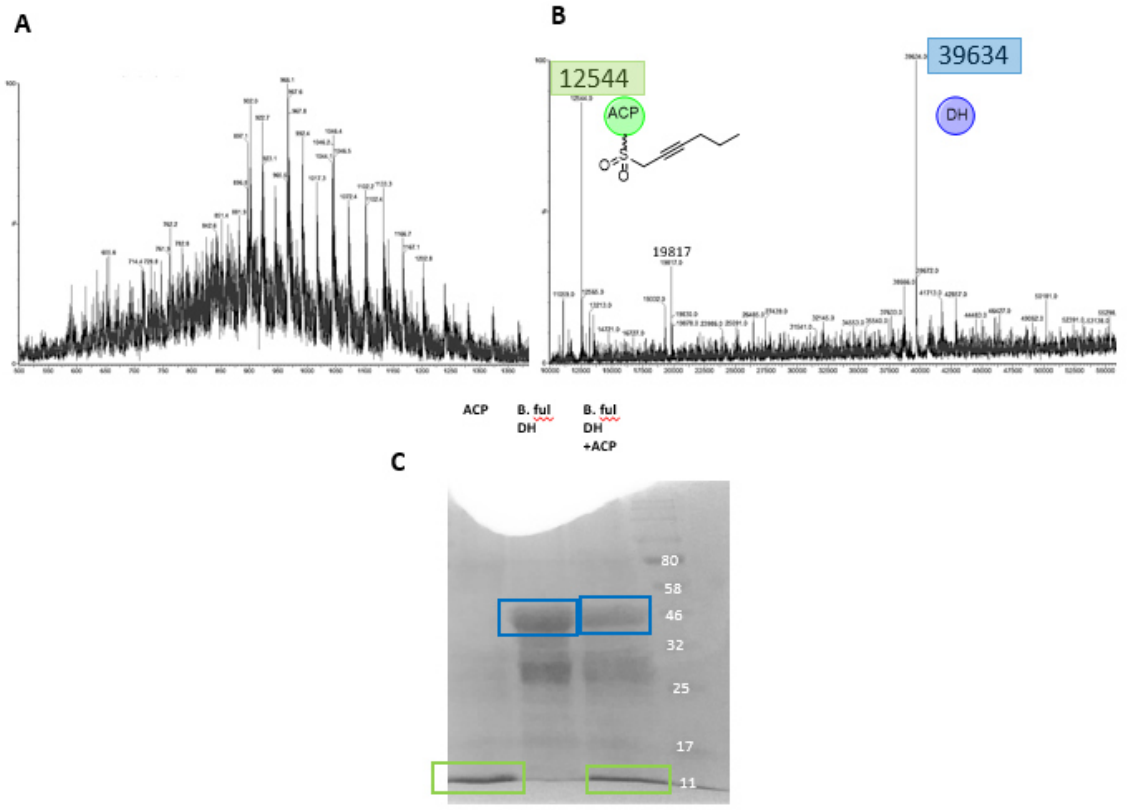


Figure 7.10: The crosslink reaction between linker-ACP 4 and bfPKS DH: **A**, mass spectrum; **B**, deconvoluted spectrum; **C**, SDS PAGE.

7.4 Discussion

The main aim of this part of the project was to crosslink bfPKS ACP with DH domains of different origin. Three different linkers were attached to the bfPKS ACP (constructs: linker-ACP 2, 3 and 4) and shown to be correctly formed by MS. These were then tested with bfPKS, stPKS and SQTKS DH domains. The DH domains were previously shown to be catalytically active with SNAC substrates. The crosslink reactions were checked by protein mass spectrometry and SDS-PAGE. All protein species, the bfPKS ACP including different linkers and the DH domains of stPKS, SQTKS and bfPKS were detected clearly in the deconvoluted mass spectra and also on the gels, showing that the analytical methods were working as expected. Six combinations, varying the linker and/or the DH domain, were tried but none of them lead to the desired result of a crosslinked species which can be used for structural elucidation. Even the combination of bfPKS DH with bfPKS ACP, i.e. two protein components from the same PKS, did not lead to cross-linking. The increased induction time did not solve the problem. The main aim of crosslink the bfPKS ACP with a DH domain was not achieved and for this reason no structural information could be generated.

Crosslinking with bacterial proteins showed success in the past. Type II bacterial ACP (AcpP) and the corresponding DH (FabZ) were crosslinked successfully.[37] The crosslink formation was confirmed by SDS-PAGE and the protein complex was characterized by crystallography. Similar concentrations (same ratio) were used for our experiments with the Type I fungal PKS proteins. The incubation time of the FabZ with the linker-ACP, which was used for the bacterial proteins (8 h) led to precipitation of the fungal proteins in our case. The incubation time for more than 5 h was not possible under these conditions due to more sensitive proteins. Maybe buffer optimization experiments could lead in future to a stabilization of these proteins. Anyway, the incubation time may improve the yield of the crosslinked species, and we would expect at least small amounts of crosslinked protein if the reaction works. The link between AcpP and FabZ was performed using a crosslinker which has same functional groups and mechanism as linker-ACP 4 in our work but for some reason the crosslinking did not work out for us with the fungal proteins of the iterative PKS.

A different strategy can be used by protecting the reactive α-protons of the linker-ACP 4 from unwanted reactions as it was described with the silylcyanohydrin warhead.[155] With this strategy the reative α-protons can be activated directly before crosslinking and

there is a reduced risk of side reactions prior the desired reaction. This experiment can be tested in a future project.

Another improvement which could be made is to implement a positive control. An ACP and a DH domain of an already described successful crosslink can be repeated to check if the methodology works fine.

8. *In vitro* Analysis of the Early Steps of Byssochlamic Acid Biosynthesis

8.1 Introduction

Byssochlamic acid (BA, **106**) has been of interest to synthetic and biosynthetic chemists since its structure was determined in 1962.[156] BA is bioactive and belongs to compounds that are known as maleidrides, which are a family of secondary metabolites consisting of 7, 8, or 9 membered carbon rings connected with one or two maleic anhydride (**105)** moieties (Figure 8.1).[157] Nine membered carbon ring maleidrides are called *nonadrides*, the group to which BA belongs.

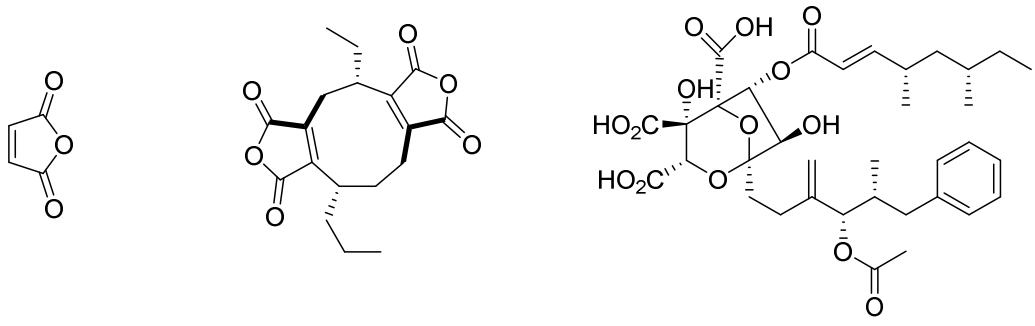


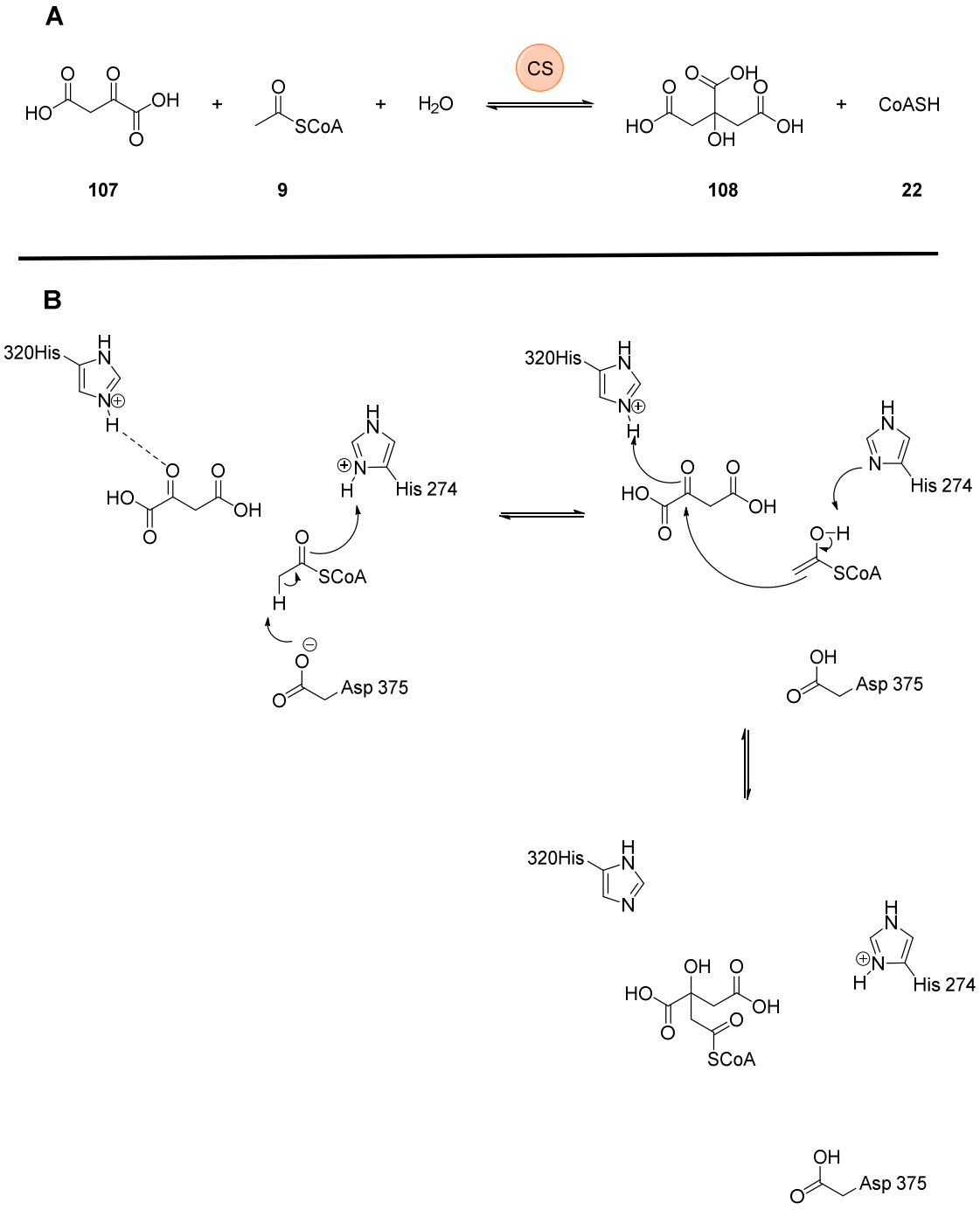
Figure 8.1: Structures mentioned in the text.

BA was first isolated by Raistrick from the ascomycete *Byssochlamys fulva* in 1933.[158] Many decades later a total synthesis of racemic BA was established in 1972 by the group of Tabak.[159] The enantioselective synthesis of the compound was achieved in 2000 by Drapela and co-workers. They showed that both enantiomers of BA can be synthesized by a [2 + 2] photoaddition-cycloreversion pathway.[160]

The first biosynthetic analyses of BA were performed by Barton and Sutherland. They proved their hypothesis that **106** appears to derive from oxaloacetate and hexanoate by carbon-labelled feeding experiments in 1968.[161,162] These results lead to the suggestion, that a citrate synthase like enzyme has to be involved in the formation of **106**.

Citrate synthase (CS) is a well-studied enzyme from primary metabolism. The enzyme is a component of all living cells and plays a key role in the citric acid cycle. Within this cycle the CS catalyses the Claisen condensation reaction of acetyl-CoA **9** with oxaloacetate **107** to form citrate **108** (Scheme 8.1). The acetyl group of acetyl-CoA undergoes a *si*-face attack to the keto group of oxaloacetate under inversion of configuration of the methyl group.[163,164] Experiments by various groups showed that the

CS has a high substrate selectivity, allowing only minor changes in substrates.[165–167] In 1990 Remington and colleagues proposed a mechanism for the CS after solving structures of pig heart and chicken heart CSs.[168–170]



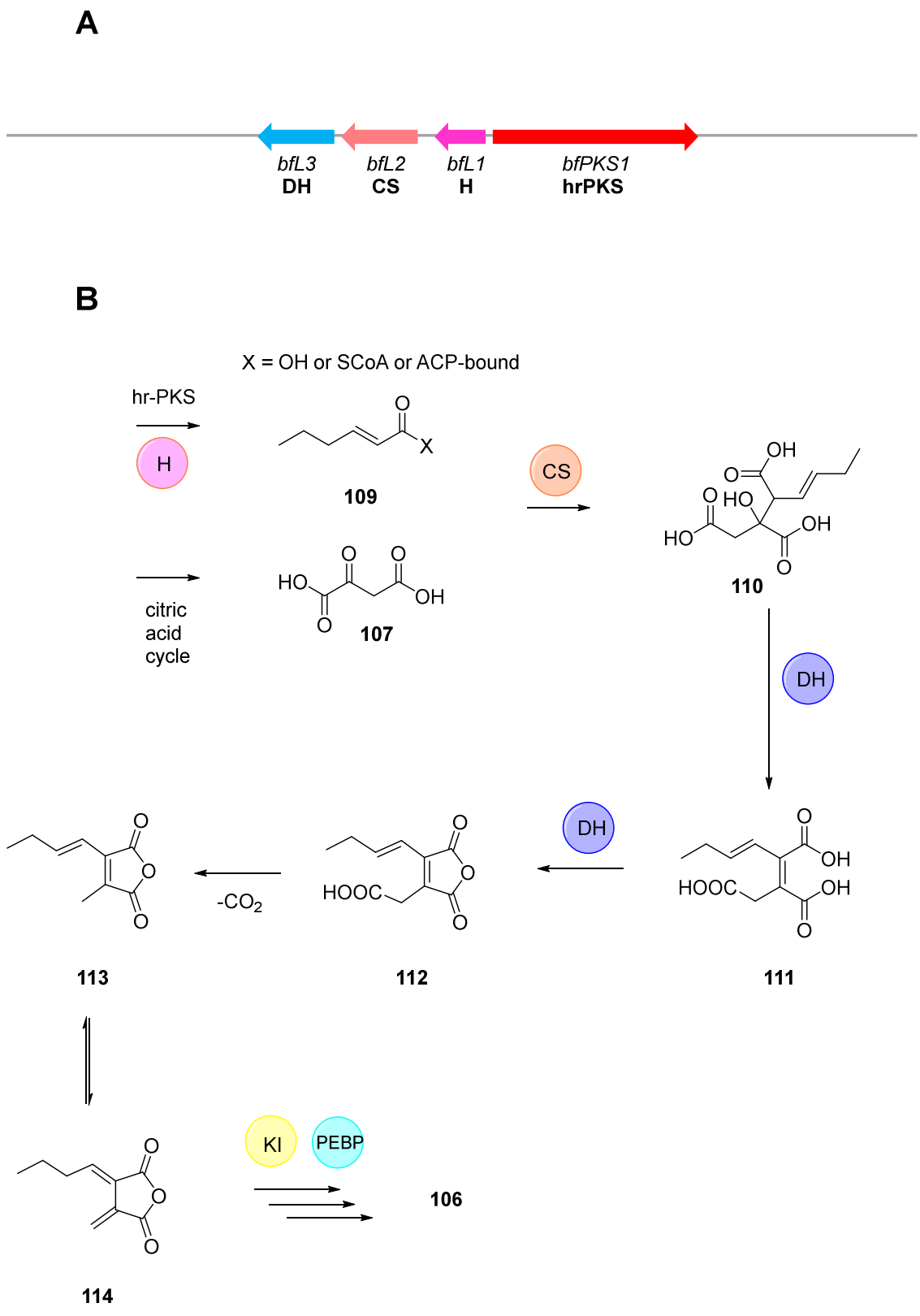
Scheme 8.1: Catalysis by the primary metabolism citrate synthase: **A**, catalyzed reaction; **B**, the proposed mechanism.[170]

Based on this structural information, the group proposed a mechanism for the CS reaction (Scheme 8.1 **B**). Three residues of the active site have a key role during catalysis (Asp 375, His-274 and His-320). In the first step a deprotonation of the acetyl-CoA methyl

group is performed by Asp-375. The enol is formed by receive of a proton from His-274. The carbon-carbon bond forming reaction between oxaloacetate and the enol is supported by the proton transfer of His-274 and His-320 leading to citryl-CoA. It is believed that the hydrolysis reaction, to form the final product citrate, is catalysed by the CS as well.

Later in 2016 Cox and co-workers revealed the biosynthesis of BA through gene disruption and heterologous expression experiments.[171] They sequenced the genomes of two *B. fulva* strains and identified a maleidride related biosynthetic gene cluster (BGC) in each of them, containing a highly reducing polyketide synthase (hrPKS) and a citrate synthase like enzyme (CS). A knockout of the *bfpks1* gene stopped production of BA **106**, showing that the gene is essential. Four genes including the PKS (*bfpks1*), the citrate synthase like enzyme (*bfL2*), the methylcitrate dehydratase (*bfL3*) and the hydrolase (*bfL1*) were hetrelogously expressed in *A. oryzae* and showed in further experiments, that they are involved in the production of monomers **113** and **114** (Scheme 8.2). The hydrolase has a homology to Type II thiolesterases, such as RifR[172] and LovG,[173] which are involved in the release of ACP-bound polyketide intermediates. Based on these investigations of the byssochlamic acid pathway,[174] the results are summarized in a proposed BA biosynthesis for the early steps (Scheme 8.2 B).

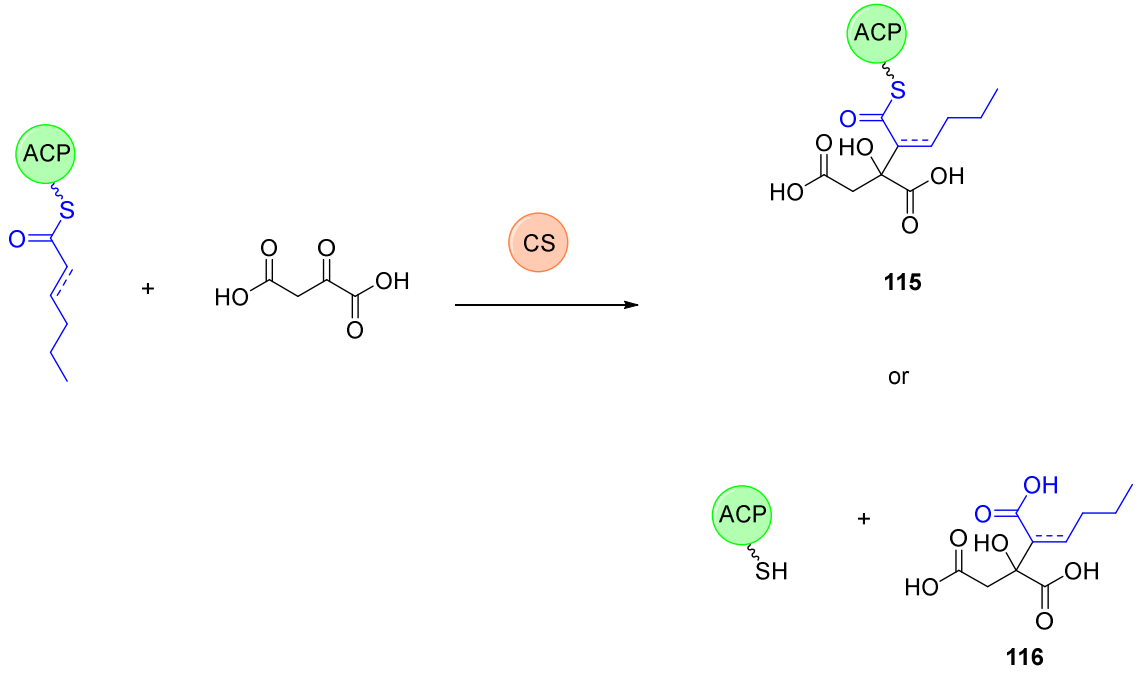
The hrPKS is proposed to produce the first intermediate, the hexaketide **109**, which might be released by the hydrolase (H). It is not clear if this intermediate is released as a CoA-thiolester, in an ACP-bound form or as a free acid. The second intermediate is oxaloacetate **107**, which probably comes from the citric acid cycle.[161,162] The citrate synthase homologue (CS) may then catalyse the reaction between **107** and **109** to form **110**. In the next step the 2-methyl citrate dehydratase (2MCDH) catalyses the elimination of the hydroxyl-group, resulting in **111**. Presumably, the next reaction, which forms the anhydride **112** is also catalysed by the DH, but may also be spontaneous. Previous work has shown that **112** is unstable and decarboxylates spontaneously to volatile products **113** and **114**. Late steps of the biosynthesis are catalysed by ketosteroid isomer like protein (KI) and phosphatidylethanol amine binding proteins (PEBP) resulting in the final product byssochlamis acid **106**.



Scheme 8.2: Early biosynthetic steps of BA: **A**, BGC including for the project relevant genes (other genes not shown); **B**, the proposed biosynthesis.[171]

The substrate preference of the CS concerning the polyketide is not known. There are three possible options: free acid, CoA bound or ACP bound polyketide. The group of

Tang reported a BGC which encodes the biosynthesis of squalestatin S1 (**27**, homologous to byssochlamic acid BGC) that contain genes for a citrate synthase like protein (CS) and a hydrolase (H).[51] Heterologous expression experiments in *A. nidulans* showed that only a hrPKS, a citrate synthase and a hydrolase are needed to produce squalestatin S1. In one of the constructs without the citrate synthase gene they could not detect any trace of the free polyketide, which shows the significance of the citrate synthase for the SQS1 biosynthesis. The group proposed an ACP coupled intermediate past the H- and CS-catalysed reactions. An analogous reaction in BA pathway would generate the intermediate **115** (Figure 8.3). Another possible reaction including the participation of ACP is the attachment of oxaloacetate followed by a hydrolysis, resulting in *holo* ACP and **116**.



Scheme 8.3: Proposed catalysis of the *B. fulva* CS with an ACP bound substrate.

8.2 Aims

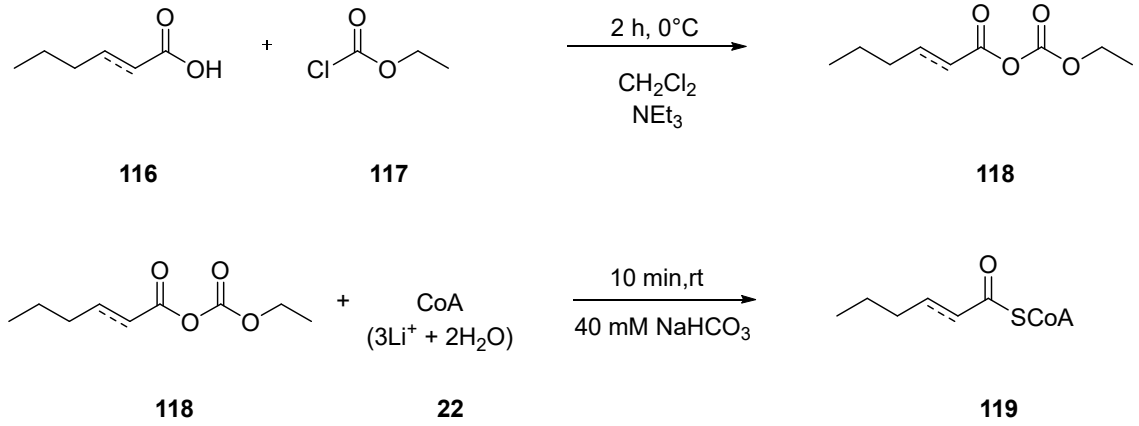
In this project the focus lies on the early steps of BA **106** biosynthesis. In particular, the interaction of the ACP of the *B. fulva* hrPKS with the hydrolase (H) and the citrate synthase (CS) will be examined *in vitro*. The role of the hydrolase is unclear and its precise substrate, which may be ACP-linked, is unknown. Hex-(an/en)-oyl-ACP will be incubated with the hydrolase and analysed by MS on protein- and small molecule level to check if the ACP-linked polyketide gets hydrolysed.

The CS is of special interest for this project. We want to determine, if the CS accepts ACP bound polyketide substrates, free acid substrates, CoA-bound substrates or if it has a broad selectivity for different substrates. Firstly, the polyketide substrate hex-(an/en)-oyl-CoA will be synthesized. Secondly, the hex-(an/en)-oyl-CoA will be loaded to bfPKS *apo* ACP using methods which were described previously. Thirdly, the hex-(an/ene)-oyl-ACP will be purified to make sure that no hex-(an/en)-oyl-CoA is left over. Lastly, the pure hex-(an/en)-oyl-ACP will be presented to the CS in the presence of oxaloacetate and analysed by MS on protein- and small molecule-level. Experiments will be performed with the saturated and also the unsaturated polyketide. For simplification reasons only the saturated version is shown and the results for both are the same if not other stated.

8.3 Results

8.3.1 Production of Acyl-CoAs and *in vitro* investigations of the *B. fulva* Hydrolase

The hydrolase encoded by *BfL1* was heterologously expressed in *E. coli* BL21 and purified by Ni-NTA and size exclusion chromatography by Dr. Steffen Friedrich. Hexanoyl-CoA **120** was synthesised by Dr. Steffen Friedrich using Coenzyme A and the anhydride **118** formed by prior reaction of hexanoic acid **116** with ethyl chloroformate **117**.[175] The same synthesis was used to produce hex-2-enoyl CoA **119** (Scheme 8.4).



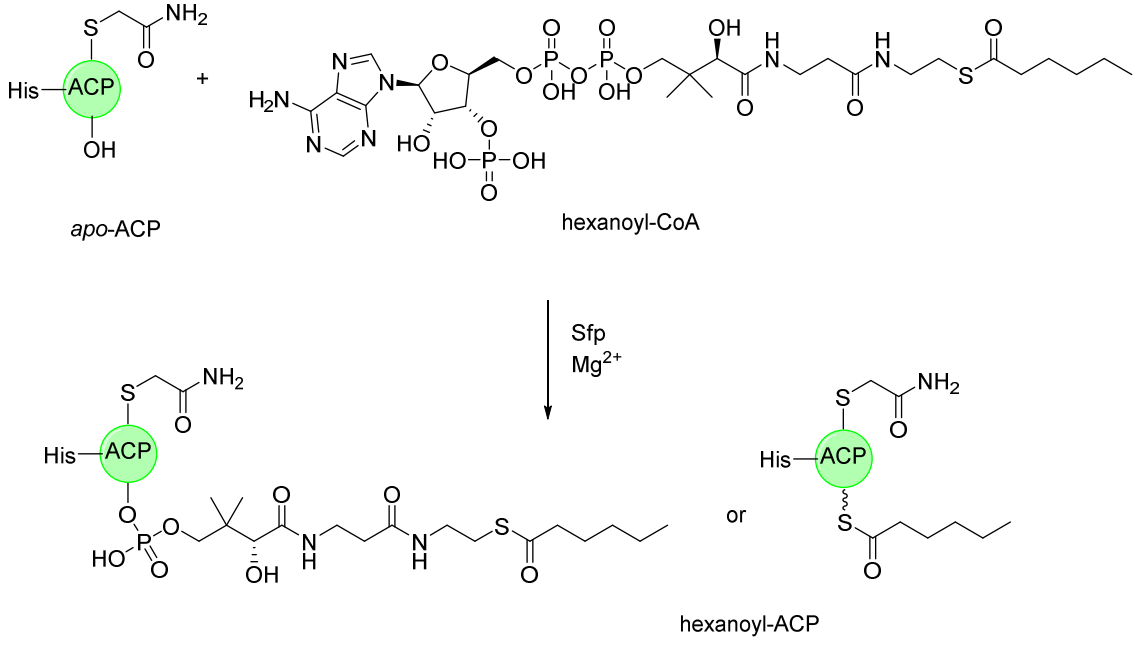
Scheme 8.4: Synthesis of acyl-CoA substrates for the CS.

At this time point a modified version of bfPKS ACP was used for the experiments. The modification was an addition of iodoacetamide to the thiol group of a cysteine within the

ACP (section 3.3.6). For this reason, the mass of *apo*-ACP is larger by 57 Da than described in previous chapters.

After successful synthesis of hexanoyl-CoA the loading to ACP was performed (Scheme 8.5). Incubation of N-terminal His-tagged bfPKS ACP (1.25 mg/ml) with Hexanoyl-CoA (25 µM) and Sfp (0.01 µg/µl) for 30 min at 30 °C led to a successful loading of the hexanoyl moiety to the *apo* ACP (Figure 8.2 **A**) as observed by ESIMS. The ACP loading step is performed in an analogous manner as the *holo* reaction described in section 3.3.5. The only difference is in the usage of hexanoyl-CoA instead of coenzyme A.

The calculated mass of *apo* ACP is 13.898 kDa, which was not found after the loading reaction, showing that the conversion was complete. The calculated mass of hexanoyl-ACP is 14.337 kDa. The main peak (Figure 8.2 **A)** has a mass of 14.336 kDa which is within the error range of the expected modified protein. Two additional protein species in lower concentrations were also detected. The first one has a mass of 14.236 kDa which is close to the calculated mass of *holo* ACP (14.238 kDa). Hydrolysis of hexanoyl-ACP or a contamination of substrate with Coenzyme A could explain the observation of *holo*-ACP. The third detected protein species has a mass of 14.514 kDa, which is 178 Da larger than the mass of hexanoyl-ACP. This is the result of gluconoylation which was described in section 3.3.5.



Scheme 8.5: Loading reaction of bfPKS *apo* ACP with hexanoyl-CoA.

To make sure that no free hexanoyl-CoA substrate is left in the mixture a gel filtration chromatography was performed (Superdex 10/300 column, Figure 8.2 **B**) to remove small molecules. The same buffer was used for the purification as the storage buffer for ACPs and KS/AT (section 10.5.10). The successful purification was confirmed by protein mass spectrometry. The gel filtration purification removed the small molecules and had no influence on the *holo*-ACP: hexanoyl-ACP: glyconoylated hexanoyl-ACP ratio.

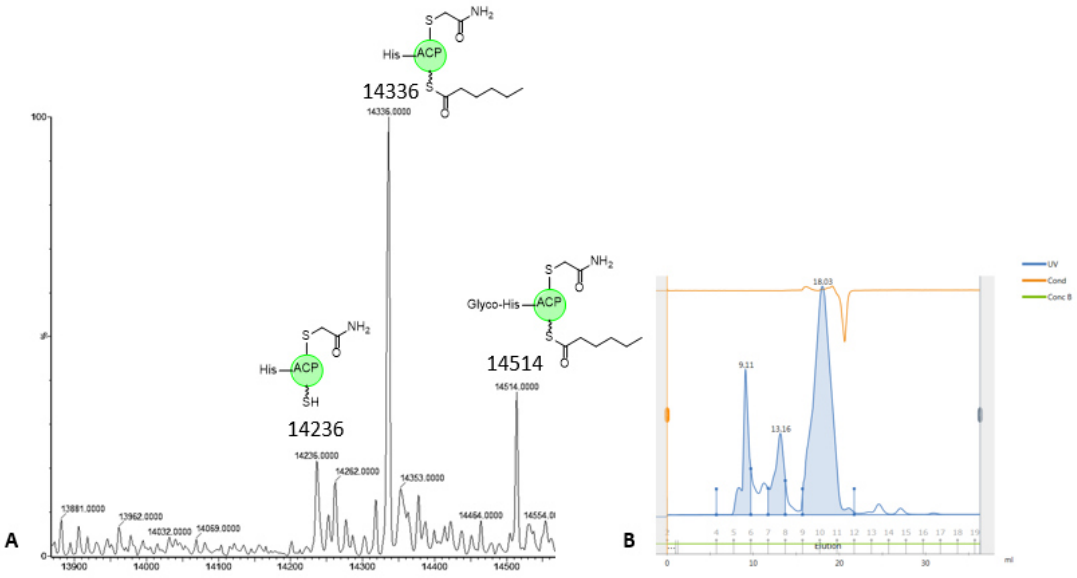


Figure 8.2: Reaction of modified bfPKS *apo* ACP with hexanoyl-CoA resulting in three protein species and their purification: **A**, deconvoluted mass spectrum; **B**, SEC of hexanoyl-ACP (middle peak) using a Superdex 10/300 column and the Äkta Pure system.

Finally, hexanoyl-ACP (1 mg/ml) was incubated with the hydrolase for 30 min at 30 °C. At the end of the reaction the reaction mixture was analysed by ESIMS directly (Figure 8.3). Full conversion of acyl-ACP with the mass of 14.279 kDa (calc.:14.280 kDa) to *holo* ACP with the mass of 14.181 kDa (calc.: 14.181 kDa) was observed. As described previously gluconoylated species are also present. Small molecules of the experiment were analysed by MS (Dr. Steffen Friedrich) showing the formation of hexanoic acid. This experiment confirms, that the hydrolase can cleave off the hexanoyl group from the ACP as expected, resulting in hexanoic acid and *holo* ACP.

8.3.2 Testing of CS Substrate Specificity

The purified hexanoyl-ACP (8.3.1) was used for testing the substrate selectivity of the citrate synthase, which was also heterologously expressed in *E. coli* BL21 and purified

by Ni-NTA and size exclusion chromatography (Dr. Steffen Friedrich). Hexanoyl-ACP (approx. 1 mg/ml) was incubated with oxaloacetate (1 mM) and CS for 3 h at 30 °C (Figure 8.4 **A**). A negative control, which was performed in parallel, contained boiled CS instead of native protein (Figure 8.4 **B**). At the end of the reaction the reaction mixture was analysed by ESIMS directly.

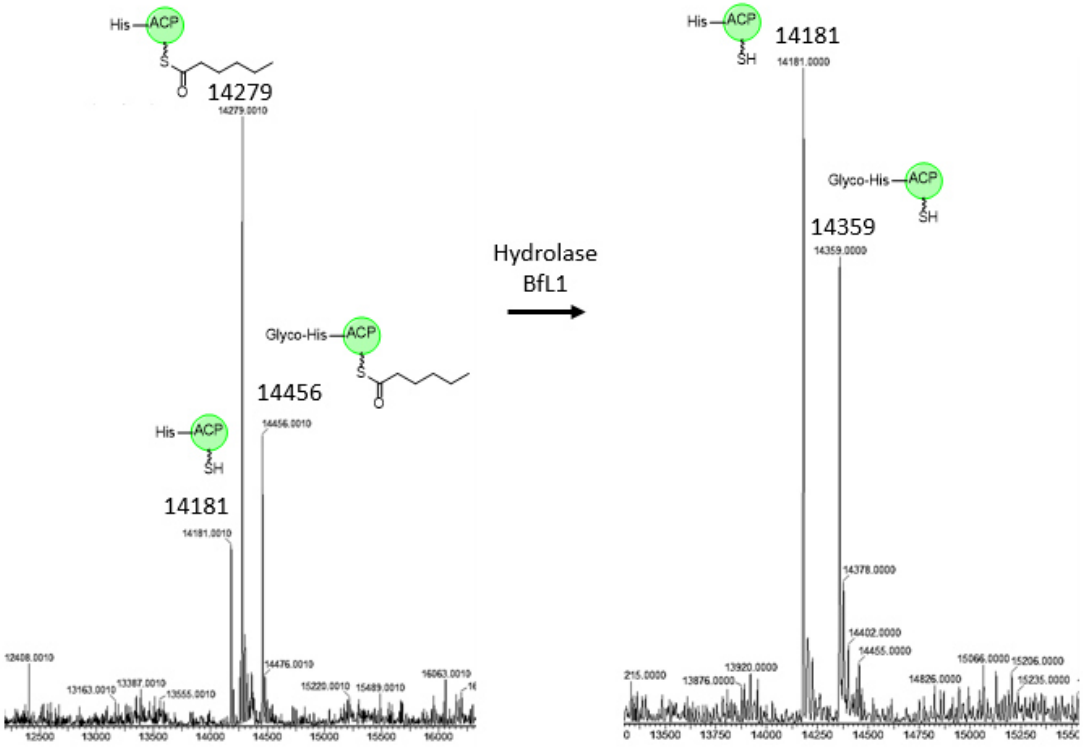


Figure 8.3: Reaction of *B. fulva* hexanoyl-ACP with *B. fulva* hydrolase.

In a parallel reaction the proteins were precipitated by addition of 1 volume of acetonitrile and removed by centrifugation. Small molecules were analysed by LCMS by Dr. Steffen Friedrich. No expected citrate **110** (Scheme 8.6) was formed during the reaction of CS with hexanoyl-CoA and oxaloacetate. At the protein level no reaction was observed either. Negative control showed the same unreacted hexanoyl-ACP species as the reaction with the active CS. The results show that the CS does not accept ACP bound substrate. In parallel reactions Dr. Steffen Friedrich showed that the CS enzyme is fully active with CoA substrates.

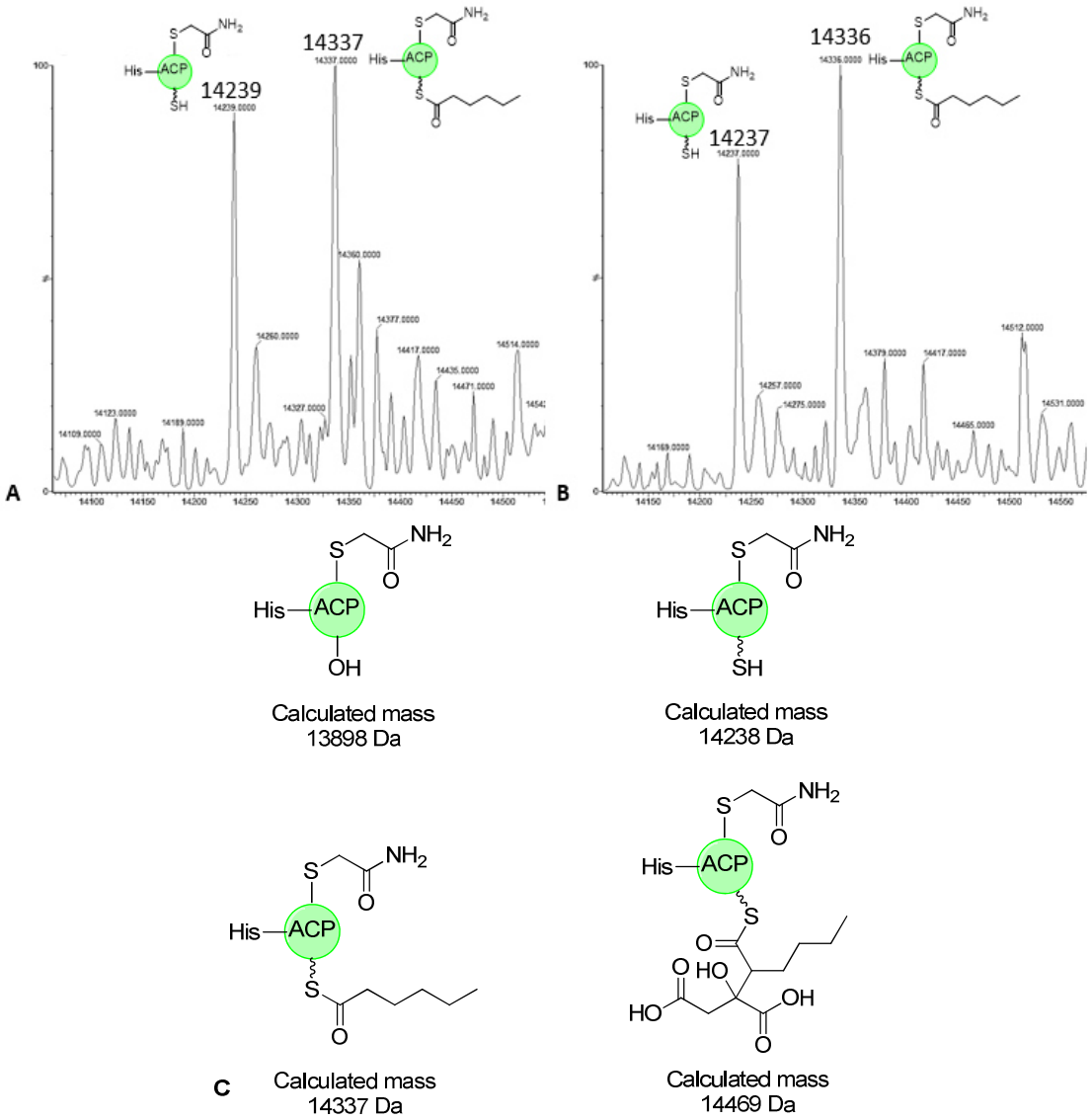
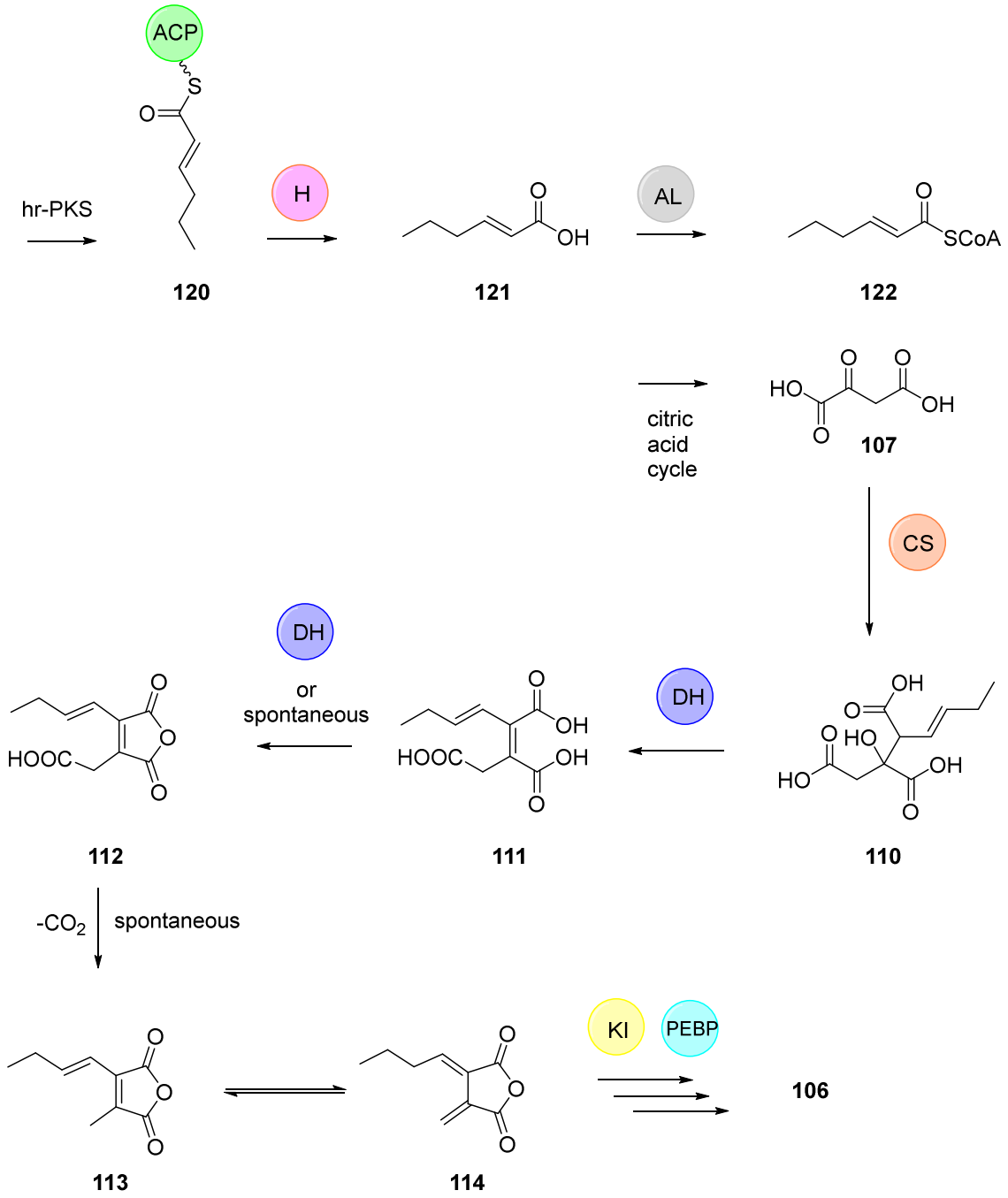


Figure 8.4: Substrate test for the CS: **A**, reaction of hexanoyl-ACP with oxaloacetate and CS; **B**, Negative control containing same chemicals except that CS was boiled prior reaction; **C**, most important ACP species and their calculated masses.

Taking together the newest results and considering the previously proposed biosynthesis of BA for the early steps, we created an updated scheme. It is now known, that the polyketide is released by the hydrolase resulting in a carboxylic acid species. We also know, that the citrate synthase does not accept SNAC, pantetheine, free acid derivatives or ACP-bound species. Whereas CoA bound polyketides are converted with oxaloacetate **107** to **110**. This conclusions bring the necessity of one or several enzymes converting the free acid polyketide to the CoA analogue. There may be an acyl-CoA ligase involved, which transforms the free acid **121** to CoA bound compound **122**. However, the first steps of the BA biosynthesis are now clearer but still need more research to be done for a full elucidation.



Scheme 8.6: New proposed biosynthesis of the early steps of BA (**106**) considering latest results.

8.4 Discussion

The Cox group had already done a lot of research to illuminate the biosynthesis of BA **106**. The special interest of this project lies on the early steps of the synthesis. The citrate synthase (CS), the hydrolase (H) and also the ACP of the hrPKS have all key roles for the first synthetic steps. Since the *B. fulva* PKS ACP was already available (chapter 2) it was

easy to assess whether ACP-bound species were substrates of the CS enzyme as proposed by Tang and coworkers.[175]

The experiments showed that CS can convert saturated and unsaturated hexaketides, but only in CoA bound form. Interestingly, the CS does not accept ACP bound substrates, free acid derivatives, SNACs and pantetheine derivatives. The CS seems to be selective in the CoA binding region and less selective in the polyketide area due to the acceptance of saturated as well as unsaturated substrates. It would be interesting to know how broad the substrate acceptance really is. Modifying the chain length and saturation pattern of the polyketide would give more understanding of the substrate selectivity.

Our experiments showed that the hydrolase *Bfl1* was able to release hex-(an/ene)-oic acid from the bfPKS ACP resulting in the corresponding carboxylic acid as a product.

Covering all latest results, a new more detailed proposal for the biosynthesis of the BA **106** was made. The first reaction past the polyketide synthesis is the hydrolysis of the CoA bound substrate by the hydrolase *Bfl1*. A new enzyme is therefore required to form a CoA analogue **122** out of the free acid **121**. The *bf* BGC does not encode such a protein, and may be that a primary metabolism-CoA synthetase is able to perform this step. More research need to be done to clear which protein/s are catalyzing the CoA activiation. In the third step of the pathway, the CS then reacts the polyketide CoA **122** with oxaloacetate **107**. This is in agreement with all known primary metabolism CS enzymes which use acetyl or propionyl CoA substrates.

9. Overall Conclusions

In this thesis protein domains from two fungal highly reducing iterative polyketide synthases were analysed *in vitro*. The focus lay on the ACP and its central role of offering substrates to corresponding domains. The ACPs of the squalestatin tetraketide synthase and of *B. fulva* (byssochlamic acid pathway) were expressed heterologously in *E. coli* and were purified. Three different PPTases (Sfp, MtaA and *Phoma* PPTase) were expressed successfully in the same way and were used for the phosphopantetheinylation of the inactive *apo*-ACPs to active *holo*-ACPs. The SQTKS ACP was expressed in a soluble form with two His tags on both ends (N-, and C-terminal), but the phosphopantetheinylation reaction was unsuccessful. The variation of the His-tag pattern, the usage of the three PPTases and the variation of sequence length did not solve the inactivity problem. In contrast, the bfPKS ACP was converted successfully to *holo*-ACP by using Sfp as the PPTase (Chapter 3). All *in vitro* investigations were carried out with the bfPKS ACP.

A protein mass spectrometry method was established and improved using a protein standard. Small changes on protein level like *apo-holo* conversions or substrate and linker loading reactions can be observed using this method.

A main aim of this thesis was to crosslink bfPKS ACP to DH domains of different origin to gain structural information. To do so, biosynthetic enzymes (PanK, PPAT, DPCK) were heterologously expressed in *E. coli*, purified and characterized by mass spectrometry and SDS-PAGE. These enzymes were able to convert pantetheine analogues into CoA analogues. Chemically synthesized pantetheine crosslinkers were transformed in stepwise reactions as well as in a one pot reaction to the CoA analogues using these biosynthetic enzymes. DH domains of SQTKS and *B. fulva* were heterologously expressed and purified successfully. The isolated DH domain of the strobilurin pathway was provided. All DH domains showed substrate conversion and were catalytically active. All linkers were attached successfully to the bfPKS ACP using Sfp. Different linker and DH combinations were tried but no crosslinking was observed (Chapters 5 and 7). More experiments need to be performed for a successful crosslinking. As shown by the group of Burkhart, many optimization experiments and suitable linkers need to be established first. Strategies of protecting sensible parts of the linker need to be tested, optimization of the protein stability against precipitation or using simpler bacterial PKS domains first to check if the overall experiment conditions and set up are fine before

trying to link more problematic fungal proteins. It seems like not only in expression experiments but also in *in vitro* experiments the fungal proteins are harder to handle compared to bacterial ones.

The bfPKS ACP was also used in the investigation of the citrate synthase (byssochlamic acid pathway) substrate preference. The citrate synthase (CS) substrate hexanoyl-CoA was synthsized by the biosynthetic enzymes PanK, PPAT and DPCK in a stepwise reaction and confirmed by LCMS analysis. The CoA analogue was loaded to the ACP using Sfp, which was observed by protein mass spectrometry, and was offered to the CS. Together with oxaloacetate the product **115** should be formed if the CS accepts the ACP bound substrate. The results showed that the CS did not convert the ACP bound substrate and prefers the CoA bound substrate (Chapter 8). If the CoA substrate has its origin in primary or in secondary metabolism needs to be figured out in future projects.

A problematic protein was the *C*-MeT of the SQTKS. The transferase showed in the past minor or no significant activity to pantetheine and SNAC substrates. It was assumed that the *C*-MeT needs ACP bound substrate for an increased activity. The *C*-MeT substrate acetoacetyl-CoA was synthesized from acetoacetyl pantetheine using PanK, PPAT and DPCK and was transferred to the bfPKS ACP using Sfp. The ACP loaded substrate was offered to the *C*-MeT but did not show methylation. The inactivity issue of the *C*-MeT domain remains still unclear (Chapter 6). The expression of the SQTKS *C*-MeT domain in a fungal system seems to be the only reasonable experiment to solve the inactivity issue.

The KS/AT didomain was the most problematic protein. The didomain was heterologously expressed in different *E. coli* strains (BL21 DE3, TaKaRa, Arctic express) resulting in an insoluble protein. Various expression conditions (temperature, inducer concentration, media) were tried, a SUMO solubility tag was integrated but did not gave any improvement. Solubilization and refolding attempts with urea and N-lauroylsarcosin were unsuccessful. Different truncated constructs were also tried but showed no significant difference to the full length construct. Again like the *C*-MeT domain this protein may need a fungal host to have a proper folding during the expression process. The expression in bacterial systems has many benefits but in this case the disadvantages predominate.

In general the work with fungal proteins remains challenging and their investigation takes more time and patience than bacterial proteins. Especially for SQTKS

mono and multi domains the expression in *E. coli* reaches its limits and needs to be expanded to eukaryotic systems.

10. Experimental

10.1 Materials and Equipment

All media, buffers and solutions were prepared with Millipore water (GenPure Pro Millipore device, Thermo Scientific) or distilled water. Media were sterilized at 121 °C for 15 min (Autoclave 2100 Classic, Prestige Medical).

Table 10.1: Media and Agar used in this work.

Media /Agar	Composition in % (*w/v*)	Ingredients
LB medium	0.5 0.5 1	Yeast extract NaCl Tryptone
2TY medium	1 0.5 1.6	Yeast extract NaCl Tryptone
SOC medium	0.5 0.06 2 0.02 25 mM 1	Yeast extract NaCl Tryptone KCl $MgCl_2$ x 6 H_2O D(+)-Glucose
TB medium	2.4 1.2 0.4 10	Yeast extract Tryptone Glycerol KPI buffer (sterile)
KPI buffer	2.31 12.54	KH_2PO_4 K_2HPO_4
LBE5052	0.5 1 0.5 0.05 0.2 0.07 0.25	Yeast extract Tryptone Glycerol D(+)-Glucose Monohydrate Lactose Monohydrate Na_2SO_4 NH_4Cl To be added (to 900 mL LBE5052): 1 mL $MgSO_4$ x $6H_2O$ (2 M) 1 mL Metals Mix (1000x) 100 mL Potassium Phasphate Mix (50 mM Phosphate)
Metals Mix 1000x	1.62	$FeCl_3$ To be added: 1 mL $MnCl_2$ x $4H_2O$ (1 M) 1 mL $ZnSO_4$ x $7H_2O$ (1 M) 1 mL $CoCl_2$ x $6H_2O$ (0.2 M) 1 mL $NiCl_2$ x $6H_2O$ (0.2 M) 46 mL pure water Filter sterilized
Potassium Phosphate Mix 50 mM Phosphate		10 mL KH_2PO_4 (1 M) 40 mL K_2HPO_4 (1 M) 50 mL pure water
LB agar	0.5 0.5 1 1.5	Yeast extract NaCl Tryptone Agar

Two antibiotics were used in this work. Stock solutions were prepared in distilled water with 50 mg/mL of carbenicillin and kanamycin. Stocks were diluted 1:1000 in media

when used. The final concentration was 50 µg/mL. *E. coli* strains which were used in this work are summarized below.

Table 10.2: Bacterial strains used in this work.

Bacterial strain	Origin
E. coli BL21 DE3	Thermo Fisher Scientific
E. coli Top 10	Thermo Fisher Scientific
E. coli TaKaRa	takarabio
E. coli Arctic Express	Agilent Technologies

10.2 Software

Following software was used in this work:

Table 10.3: Softwares used in this work.

Software	Developer
Geneious 7.1.9	Biomatters (Auckland 1010, New Zealand)
Image Lab	Bio-Rad Labaratories Ltd., Informatics Division (München, Germany)
MassLynx	Waters, Milford, MA, USA
Molecular Imager Gel doc XR+	Bio-rad Labaratories, Inc. (München, Germany)
CLC Sequence Viewer 6.8.1	CLC bio A/S (DK-8200 Aarhus N, Denmark)
SWISS-MODEL (web server)	Biozentrum University of Basel
HmmerWeb version 3	EMBL-EBI
Protein-Sol web server	Warwicker and Curtis Groups

10.3 General Techniques

10.3.1 *E. coli* Storage

Bacterial cultures were stored as 20 % glycerol suspensions. Colonies were stored on solid LB-media at 4 °C.

10.3.2 Agarose Gel Electrophoresis

Agarose gels were run on a BioRad horizontal gel electrophoresis equipment at an applied voltage of 130 V for 20 min using 0.5 TAE buffer (50X TAE: 2 M Tris-HCl, 1 M acetic acid, 50 mM EDTA). The samples were mixed with 6X DNA Loading Dye and loaded on 1 % (w/v) agarose gels with an addition of RedSafe DNA Stain. Depending on the expected size 1 kb or 100 bp DNA Ladder from NEB was used as a size comparison.

DNA bands were observed under UV light (254 nm) and digital images were recorded using a UVP camera.

10.3.3 Protein Concentration Measurement and Identification

Protein concentrations were measured determined using the Bradford assay. The protein samples were diluted with the storage buffer to a final volume of 100 µL and mixed with 900 µL of the Bradford reagent (Roti Nanoquant from Roth). The mixture was incubated for 15 min. Bovine Serum Albumin (BSA) was used in a range of 0.1 mg/mL to 1.5 mg/mL for the generation of a standard curve. The concentration of the protein sample was determined photometrically at 595 nm and calculated to the standard curve.

For the identification of the protein, a SDS-PAGE band was cut and enzymatically digested by trypsin. The peptide fragments were further analysed by ESI Q-TOF (Dr. Jennifer Senkler, working group of Prof. Braun, Institute of Plant Genetics, Leibniz University of Hanover).

10.3.4 SDS-PAGE

For analyzing protein samples the SDS-PAGE was used. The acrylamide (40 %) comprised of a 19:1 and a 37.5:1 ratio of acrylamide:bis-acrylamide. Polymerization of the gels were induced by the addition of ammonium persulphate (AMPS) and N, N, N', N' tetramethylethyethylene deaminase (TEMED). Protein samples were dissolved in Lämmli buffer and heated for 5 min at 95 °C. Pockets of the gel were loaded with 10 µl of sample and 3 µl of a marker (Color Prestained Protein Standard Broad Range, 11-245 kDa, NEB; Page Ruler Unstained Low Range Protein Ladder, 3.4-100 kDa, Thermo Fisher). The electrophoresis was performed in the beginning at 20 mA (2 gels) and was increased to 30 mA and ran for 45-60 min.

Table 10.4: Composition of SDS-PAGE.

Additions	5 % stacking gel (ml)	12 % separation gel (ml)	15 % separation gel (ml)
Acrylamide	0.268	1.5	1.875
ddH_2O	0.85	1.225	0.85
1.5 M Tris-HCl, pH8.8	-	1.9	0.95
1.5 M Tris-HCl, pH6.8	0.125	-	-
10 % (w/v) SDS	0.01	0.0375	0.0375
10 % (w/v) APS	0.01	0.0375	0.0375
TEMED	0.001	0.0015	0.0015

After electrophoresis run, the gels were stained in Coomassie staining solution (Acetic acid, Isopropanol, Coomassie) overnight and destained with a bleach solution, consisting of Acetic acid and isopropanol in a ratio 2.5:1, for 1 h. Destained gels were scanned using the Molecular Imager Gel doc XR+ (Bio-Rad) system.

10.3.5 Analytical Liquid Chromatography/ Mass Spectrometry (LCMS)

LCMS data was generated by a waters Quattro API mass spectrometer (operating in ES^+ and ES^- modes between 100 *m/z* and 1500 *m/z*) with a waters 2797 separation module and a waters 2545 pump system. For separation a Phenomenex column (C18, 100 Å, 4.6 x 100 mm) was used equipped with a Phenomenex Security Guard precolumn (Luna C_5 300 Å). The detection was performed by Waters 2998 diode array detector in the range of 200 to 600 nm and a Waters 2424 ELSD. Solvents were HPLC-grade H_2O containing 0.05% formic acid and HPLC-grade CH_3CN containing also 0.05% formic acid. Following gradient was used: 0 min 10% CH_3CN; 10 min 90% CH_3CN; 13 min 10% CH_3CN; 15 min 10% CH_3CN.

10.3.6 Protein Mass Spectrometry

Protein measurements were performed using a waters Quattro API mass spectrometer with a waters 2797 separation module and a waters 2545 pump system. Proteins were measured by direct injection (no column installed) of a protein solution (approx. 1 mg/ml) in a 50/50 mix of water/acetonitrile (0.05 % formic acid) at 0.2 ml/min. The mass

spectrometer scanned over a range of 500 – 1500 m/z. Raw data was then transformed using MaxEnt® to give the molecular weight.

10.4 Experimental for Chapter 2

10.4.1 Synthetic Gene for SQTKS KS/AT

A synthetic gene for SQTKS KS/AT was ordered at Baseclear. The gene was *E. coli* codon optimized and was delivered in a pET28a (+) vector. Restriction sites were included: *Nde*I (N-terminally) and *Xho*I (C-terminally). A stop codon at the end of the gene sequence was also included. The His-tag is encoded N-terminally. Same order was done for the Phoma PPTase gene.

10.4.2 Transformation of pET28a(+) vector into Competent *E. coli* Top10 Cells

50 µL of competent cells were thawed on ice. 1-3 µL of pure plasmid were added and incubated for 30 min on ice. The cells were heat shocked for 10 s at 42 °C and cooled down for 1 min on ice. After addition of 300 µL SOC medium the cells were shaken at 350 rpm and 37 °C for 1 h. The cells (50-150 µL) were streaked out on LB agar plates containing the appropriate antibiotic. The plates were incubated over night at 37 °C.

10.4.3 Transformation of pET28a(+) vector into Competent *E. coli* BL21 Cells

50 µL of competent cells were thawed on ice. 1-3 µL of pure plasmid were added and incubated for 30 min on ice. The cells were heat shocked for 30 s at 42 °C and cooled down for 1 min on ice. After addition of 800 µL SOC medium the cells were shaken at 350 rpm and 37 °C for 1 h. The cells (50-150 µL) were streaked out on LB agar plates containing the appropriate antibiotic. The plates were incubated over night at 37 °C.

10.4.4 Expression in *E. coli* BL21 DE3

A single colony was picked from a LB agar plate and inoculated in 5 mL (preculture) LB media with 50 µg/mL Kanamycin. The culture was incubated over night at 37 °C and 200 rpm on a bench top incubator. 1 mL of the preculture was used to inoculate 100 ml 2TY media with 50 µg/mL Kanamycin. The culture was grown at 37 °C to the OD of 0.4-0.6. One half of the culture stayed uninduced (negative control) whereas the other half was induced with 1 mM IPTG. The temperature was set to 16 °C and was incubated for 22 h.

The cells were harvested by centrifugation at 5000 g and lysed by sonication. The supernatant and the pellet were separated by centrifugation at 10000 g and resuspended in Lämmli buffer for the SDS-PAGE analysis.

10.4.5 Expression in *E. coli* Arctic Express® (DE3)

A single colony was picked from a LB agar plate and inoculated in 5 mL (preculture) LB media with 50 µg/mL Kanamycin. The culture was incubated over night at 37 °C and 200 rpm on a bench top incubator. 1 mL of the preculture was used to inoculate 100 ml 2TY media with 50 µg/mL Kanamycin. The culture was grown at 30 °C to the OD of 0.4-0.6 The temperature was set to 10 °C and 0.1/ 1 mM IPTG were added for the induction (21 h). The cells were harvested by centrifugation at 5000 g and lysed by sonication. The supernatant and the pellet were separated and resuspended in Lämmli buffer for the SDS-PAGE analysis.

10.4.6 Expression in *E. coli* TaKaRa®

A single colony was picked from a LB agar plate and inoculated in 5 mL (preculture) LB media with 50 µg/mL Kanamycin. The culture was incubated over night at 37 °C and 200 rpm on a bench top incubator. 1 mL of the preculture was used to inoculate 100 ml 2TY media with 50 µg/mL Kanamycin. The main culture was grown with L-arabinose (c=0.5-4 mg/mL) and tetracycline (c= 1-10 ng/mL) at 37 °C until the OD of 0.6 was reached. The culture was incubated for 30 min at 15 °C before induction with 1 mM IPTG. After 24 h of induction the cells were harvested by centrifugation at 5000 g and lysed by sonication. The supernatant and the pellet were separated and resuspended in Lämmli buffer for the SDS-PAGE analysis.

10.4.7 Protein Solubilisation and Refolding Approaches

10.4.7.1 N-Lauroylsarcosin

The cells were separated from the medium using centrifugation at 5000 g and were resuspended in a buffer with following composition: 50 mM Tris pH 8, 150 mM NaCl, 10% glycerol. The cells were lysed by sonication and the soluble part was separated from the insoluble part by centrifugation at 10000 g for 1 h. The insoluble part was resuspended in a buffer with the following composition: 50 mM Tris pH 8, 150 mM NaCl, 10% glycerol 1% N-Lauroylsarcosine. The suspension was stirred for 1 h and analysed by SDS-PAGE.

10.4.7.2 Urea

The cells were lysed in buffer I (50 mM Tris-HCl pH 8.5, 150 mM NaCl, 1 mM EDTA). The insoluble part was separated from the soluble part by centrifugation at 5000 g. The pellet was divided in four parts (aliquots) and all of them were dissolved in buffer II (50 mM Tris-HCl pH 8.5, 150 mM NaCl, 20 mM β-Mercaptoethanol) with different concentration of urea (2, 4, 6 and 8 M urea). The mixtures were incubated for 1 h at 40 °C and at room temperature. The samples were analysed by SDS-PAGE.

10.4.7.3 Refolding by Dialysis

The dissolved protein was transferred into a dialysis membrane. 2 L of buffer III (50 mM Tris-HCl pH 8.5, 150 mM NaCl, 4 M urea) were used for the dialysis in the first 24 h. The membrane was stirred in the buffer at 4 °C. In the next 24 h the protein sample was treated in the same way with buffer IV (50 mM Tris-HCl pH 8.5, 150 mM NaCl, 2 M urea) and in the final 24 h with buffer V (50 mM Tris-HCl pH 8.5, 150 mM NaCl). The protein sample was then analysed by SDS-PAGE.

10.4.7.4 Refolding by Manual Ni-NTA Column

The solubilized protein was incubated for 1 h with Ni-NTA agarose beads. Next, the beads were transferred into a column and washed with buffers III, IV and V (decreasing urea concentration). The flow thru was collected after each step and analysed by SDS-PAGE. In the last step, the bound protein was eluted using the elution buffer (50 mM Tris pH 8.5, 150 mM NaCl, 500 mM imidazole) and also analysed by SDS-PAGE.

10.4.8 Cloning of Truncated Constructs

The synthetic SQTKS KS/AT (in a pET28(a)+ vector) gene was used as the template for cloning the SQTKS KS/AT constructs into a pET28a(+) vector.

10.4.9 Polymerase Chain Reaction (PCR)

DNA was amplified using oligonucleotide primers (Table 10.5) by the polymerase chain reaction (PCR). For the cloning the Q5® High-Fidelity 2X Master Mix from NEB was used. The manufacturer's instructions were followed to perform the PCR.

Table 10.5: Primers used for the SQTKS KS/AT contructs.

Name	Sequence (5'-3')
Primer 1 (KS/AT)	AAACATATG ATGGTGCCGTATTATCAGCC
Primer 2 (KS/AT)	AAACATATG ATGGCGGCGATGGATGAA
Primer 3 (KS/AT)	AAACATATG TGGAGCAGCATTCCGAAAAG
Primer 4 (KS/AT)	TCATCTCTCGAGTTA CATCTGCTGAATCGGGCC
Primer 5 (KS/AT)	TCATCTCTCGAGTTA CAGGCAGCTAATATACGGCAG
Primer 6 (KS/AT)	TCATCTCTCGAGTTA CGCCACGGTCTGCATGG
Primer 7 (KS/AT)	TCATCTCTCGAGTTA GCCATAGGTGCCCTGCG

10.4.10 Restriction Digest and DNA Purification

The pET28a(+) vector and the PCR products were digested with appropriate restriction enzymes (*Nde*I, *Xho*I) simultaneously for 3 h at 37 °C. Buffer and concentrations were used from the protocol of NEB. To prevent self-ligation the DNA was treated for 30 min with alkaline phosphatase (SAP). After incubation the enzymes were deactivated for 10 min at 65 °C. The resulting fragments were purified by NucleoSpin Gel and PCR Clean-up kit from Macherey-Nagel using the manufacturer's protocol.

10.4.11 Ligation

T4 DNA Ligase was used to perform the ligations. The reactions contained DNA (Vector to insert in a 1:3 ratio), T4 ligase and ligation buffer. The reaction mixture was incubated over night at 16 °C and inactivated at 65 °C for 10 min.

10.4.12 Auto Induction in *E. coli* BL21 DE3

For the auto induction expression LBE5052 medium was prepared as described in table 10.1. Expression was performed in the same way as described in 10.4.4 with the difference that no manual induction was necessary.

10.4.13 SUMO-Tag Construct

The SQTKS KS/AT region (M1 – A941) was amplified by PCR as described in section 10.4.9 using the following two primers.

Table 10.6: Primers used for the SQTKS KS/AT SUMO construct.

Name	Sequence (5'-3')
Primer 1 (KS/AT) SUMO	AAAACCGGT ATGGTGCCGTATTATCAGCC
Primer 6 (KS/AT) SUMO	TCATCTCTCGAGTTA CGCCACGGTCTGCATGG

The gene as well as the plasmid pETM11-SUMO3GFP were digested as described in section 10.4.10. The ligation and transformations were performed as described in sections 10.4.11, 10.4.2 and 10.4.3. The expression was performed as described in section 10.4.4.

10.5 Experimental for Chapter 3

10.5.1 Cloning of SQTKS ACP

An SQTKS (in a pET28(a)+ vector) gene was used as the template for cloning the SQTKS ACP region into a pET28a(+) vector.

10.5.2 Polymerase Chain Reaction (PCR)

DNA was amplified using oligonucleotide primers (column 10.7) by the polymerase chain reaction (PCR). For the cloning the Q5® High-Fidelity 2X Master Mix from NEB was used. The manufacturer's instructions were followed to perform the PCR.

Table 10.7: Primers for SQTKS ACP.

Name	Sequence (5'-3')
Primer 1 (SQTKS ACP)	AAACATATG GCGCAGGATAAGCAGTTA
Primer 2 (SQTKS ACP)	AAACATATG ATGGCCACTTCTCTCGTG
Primer 3 (SQTKS ACP)	CTCGAG CGATTTAGTGGCTACTGTGGT
Primer 3' (SQTKS ACP)	CTCGAGTTA CGATTTAGTGGCTACTGTGGT

10.5.3 Restriction Digest and DNA Purification

Same procedure as used in section 10.4.10.

10.5.4 Ligation

Same procedure as used in section 10.4.11.

10.5.5 Synthetic Gene for bfPKS ACP and Phoma PPTase

A synthetic gene for bfPKS ACP was ordered at Baseclear. The gene was *E. coli* codon optimized (optimization performed by the company) and was delivered in a pET28a (+) vector. Restriction sites were included: *Nde*I (N-terminally) and *Xho*I (C-terminally). A stop codon at the end of the gene sequence was also included. The His-tag is encoded N-terminally. Same order was done for the Phoma PPTase gene.

10.5.6 Transformation of pET28a(+) vector into Competent *E. coli* Top10 Cells

Same transformation procedure was performed as for SQTKS KS/AT (section 10.4.2).

10.5.7 Transformation of pET28a(+) vector into Competent *E. coli* BL21 Cells

Same transformation procedure was performed as for SQTKS KS/AT (section 10.4.3).

10.5.8 Best expression conditions for the SQTKS ACP in *E. coli* BL21 DE3

A single colony was picked from a LB agar plate and inoculated in 5 mL (preculture) LB media with the 50 µg/ml Kanamycin. The culture was incubated over night at 37 °C and 200 rpm on a bench top incubator. 8 mL of the preculture were used to inoculate 800 ml 2TY media with 50 µg/ml Kanamycin. The culture was grown at 37 °C to the OD of 0.4-0.6. For induction 1 mM IPTG was added and incubated for 3 h 37 °C. The cells were separated by centrifugation at 5000 g from the medium, resuspended in the loading buffer (50 mM Tris pH 8, 150 mM NaCl, 20% glycerol (*v/v*), 20 mM imidazole) and lysed by sonication. The supernatant was separated from the pellet by centrifugation at 10000 g and was used for further purification steps.

10.5.9 Best expression conditions for the bfPKS ACP

A single colony was picked from a LB agar plate and inoculated in 5 mL (preculture) LB media with the 50 µg/ml Kanamycin. The culture was incubated over night at 37 °C and 200 rpm on a bench top incubator. 8 mL of the preculture were used to inoculate 800 ml 2TY media with 50 µg/ml Kanamycin. The culture was grown at 37 °C to the OD of 0.4-0.6. Before the induction with 1 mM IPTG the temperature was lowered to 16 °C and incubated for 22 h. The cells were separated by centrifugation at 5000 g from the medium, resuspended in the loading buffer (50 mM Tris pH 8, 150 mM NaCl, 20% glycerol (*v/v*), 20 mM imidazole) and lysed by sonication. The supernatant was separated from the pellet by centrifugation at 10000 g and was used for further purification steps.

10.5.10 Purification of Proteins by Ni-NTA and Size Exclusion Chromatography (SEC)

The used system was an ÄKTA Pure FPLC which consisted of a UV monitor (U9-L: 280 nm), conductivity reader, 2 system pumps, fraction collector, valves (inlet, column, outlet) and a mixer. The system was controlled by the Unicorn® software. Prior injection

protein samples were filtered with syringe filters (CA, pore size 0.45 µm). Depending on the purification step different columns were used: HisTrap® FF 5ml, HiPrep® 26/10 Desalting, HiLoad® 26/600 Superdex 200 pg. All columns were provided by GE Healthcare. In the first step the proteins were loaded onto the HisTrap FF 5ml column (proteins were dissolved in loading buffer:50 mM Tris pH 8, 150 mM NaCl, 20% glycerol (*v/v*), 20 mM imidazole) and washed with 60 mL of loading buffer to discard unbound proteins. The His-tagged protein was eluted from the column using the elution buffer (50 mM Tris pH 8, 150 mM NaCl, 20%glycerol (*v/v*), 500 mM imidazole) in a linear gradient (from 20 mM imidazole to 500 mM imidazole in 10 mL). The elution was collected in 2 mL fractions. The fractions were checked by SDS-PAGE and the protein of interest containing fractions were combined and used for the size exclusion chromatography (SEC). The protein was loaded onto a size exclusion column (HiPrep® 26/10 Desalting or HiLoad® or 26/600 Superdex 200 pg) and the storage buffer (50 mM Tris pH 8, 150 mM NaCl, 20% glycerol (*v/v*)) was used for the SEC run. The 2 mL fractions containing the protein of interest were combined and concentrated using centrifugal concentrators (Merck) with a cut off 3000 Da, 10000 Da and 30000 Da, depending on the protein size. The purified protein was stored at 4 °C and -20 °C.

10.5.11 Thrombin Digestion for SQTKS ACP

The digestion was performed after the Ni-NTA purification run. The protein containing fractions were combined, the protein concentration was measured and thrombin was added following the manufacturers protocol (Merck). The mixture was incubated for 1 h at 4 C and subsequently purified by the SEC (26/600 Superdex 200 pg). The protein was analysed by mass spectrometry.

10.5.12 Expression and Purification of Sfp, MtaA and Phoma PPTase

The same expression and purification conditions were used as for the bfPKS ACP. The only difference was the imidazol concentration in the elution buffer (250 mM).

10.5.13 Phosphopantetheinylation of ACP

For the reaction 50 mM Tris buffer (pH 8.8), 10 mM $MgCl_2$, 1 mg/ml *apo*-ACP, 0.01 mg/ml Sfp and 1 mM CoA were incubated for 30 min at 30 °C. The protein was analysed by mass spectrometry.

10.5.14 Protein Modification by Iodoacetamide

bfPKS ACP was incubated after the Ni-NTA purification step with 50 µM of the thiol reactive species iodoacetamide. The mixture was incubated for 30 min at 4 °C and was purified subsequently by SEC. The protein was analysed by mass spectrometry

10.6 Experimental for Chapter 4

10.6.1 Transformation into *E. coli* BL21 DE

Same transformation procedure was performed for SQTKS DH and bfPKS DH (both in pET28a(+))as for SQTKS KS/AT (section 10.4.4).

10.6.2 Expression of SQTKS DH

A single colony was picked from a LB agar plate and inoculated in 5 mL (preculture) LB media with the 50 µg/ml carbenicillin. The culture was incubated over night at 37 °C and 200 rpm on a bench top incubator. 8 mL of the preculture were used to inoculate 800 ml 2TY media with 50 µg/ml carbenicillin. The culture was grown at 37 °C to the OD of 0.4. Before the induction with 0.4 mM IPTG the temperature was lowered to 16 °C and incubated for 21 h. The cells were separated by centrifugation at 5000 g from the medium, resuspended in the modified loading buffer (20 mM Tris-HCl pH 8, 150 mM NaCl, 5 mM imidazole and 10 % glycerol) and lysed by sonication. The supernatant was separated from the pellet by centrifugation at 10000 g and was used for further purification steps.

10.6.3 Expression of bfPKS DH

The expression was performed as described in section 10.6.2.

10.6.4 Purification of SQTKS DH and bfPKS DH

The purification was performed as described in 10.5.10 but using modified buffers: elution buffer (20 mM Tris-HCl pH 8, 150 mM NaCl, 400 mM imidazole and 10 % glycerol), storage buffer (20 mM Tris-HCl pH 8, 150 mM NaCl, 10 % glycerol 100 mM L-Arginine and L-Glutamic acid).

10.6.5 DH Domain Enzyme Activity Assay

The DH activity test contained the following: 1 M Tris-HCl buffer pH 8, 30% glycerol, 0.1 mM DH, 1 mM substrate mimic. The mixture was incubated at 25 °C for 19 h. 20 µL

were taken and mixed with 20 µl CH_3CN, centrifuged to remove the precipitated protein and analysed by LCMS.

10.7 Exerimental for Chapter 5

10.7.1 Cloning of DPCK

E. coli genomic DNA was used as the template for cloning the DPCK gene into a pET28a(+) vector.

10.7.2 Polymerase Chain Reaction (PCR)

DNA was amplified using oligonucleotide primers (column 10.8) by the polymerase chain reaction (PCR). For the cloning the Q5® High-Fidelity 2X Master Mix from NEB was used. The manufacturer's instructions were followed to perform the PCR.

Table 10.8: Primers used for DPCK cloning.

Name	Sequence (5'-3')
F DPCK NdeI	AAACATATG ATGAGGTATATAGTTGCCTTAACGGG
R DPCK XhoI	TCATCTCTCGAGTTA CGGTTTTTCCTGTGAGACAAACTG

10.7.3 Restriction Digest and DNA Purification

Same procedure as used in section 10.4.10.

10.7.4 Ligation

Same procedure as used in section 10.4.11.

10.7.5 Transformation of pET28a(+) vector into Competent *E. coli* Top10 Cells

Same transformation procedure was performed in section 10.4.2.

10.7.6 Transformation of pET28a(+) vector into Competent *E. coli* BL21 Cells

Same transformation procedure was performed in section 10.4.3.

10.7.7 Expression of PanK, PPAT and DPCK

A single colony was picked from a LB agar plate and inoculated in 5 mL (preculture) LB media with the 50 µg/ml Kanamycin. The culture was incubated over night at 37 °C and

200 rpm on a bench top incubator. 8 mL of the preculture were used to inoculate 800 ml 2TY media with 50 µg/ml Kanamycin. The culture was grown at 37 °C to the OD of 0.4-0.6. Before the induction with 1 mM IPTG the temperature was lowered to 30 °C and incubated for 4 h. The cells were separated by centrifugation at 5000 g from the medium, resuspended in the loading buffer (50 mM Tris pH 8, 150 mM NaCl, 20% glycerol (*v/v*), 20 mM imidazole) and lysed by sonication. The supernatant was separated from the pellet by centrifugation at 10000 g and was used for further purification steps.

10.7.8 Purification of PanK, PPAT and DPCK

The purification was performed as described in 10.5.10 with the only difference of using 250 mM imidazole in the elution buffer.

10.7.9 Pantetheine Linker Synthesis

All linkers in this work were synthesized by Tanja Lau and Janina Meyer.

10.7.10 Preparation of Linker-CoA Analogues in Stepwise Enzymatic Reactions

The reaction mixture contained 50 mM Tris (pH 7.5), 20 mM KCl, 10 mM $MgCl_2$, 5 mM ATP, 0.2 mg/mL PanK and 0.05 mg/ml – 2 mg/ml linker CoA analogue. The mixture was incubated for 30 min at 25 °C and analysed by LCMS. For the second reaction 0.2 mg/mL PPAT and 5 mM ATP were added, incubated again for 30 min at 25 °C and controlled by LCMS. For the third reaction 0.2 mg/ml DPCK and 5 mM ATP were added and incubated for 30 min at 25 °C. The final product was analysed by LCMS.

10.7.11 Loading of Linker-CoA Analogues to ACP

For the reaction 50 mM Tris (pH 7.5), 20 mM KCl, 10 mM $MgCl_2$, 1.5 mg/ml *apo* ACP, 0.01 mg/ml Sfp and approx. 0.2 mg/ml CoA linker were mixed and incubated for 1 h at 30 °C. The protein was analysed by mass spectrometry.

10.7.12 Preparation of Linker-CoA Analogues and ACP Loading in a One-Pot Reaction

The reaction mixture contained 50 mM Tris buffer (pH 7.5), 20 mM KCl, 10 mM $MgCl_2$, 15 mM ATP, 0.2 mg/mL PanK, 0.2 mg/ml PPAT, 0.2 mg/ml DPCK, 1.5 mg/ml *apo*-ACP, 0.01 mg/ml Sfp and 0.05 mg/ml – 2 mg/ml linker CoA analogue. The mixture was incubated for 3 – 4 h at 30 °C. The success of the reaction was checked by mass spectrometry.

10.8 Experimental for Chapter 6

10.8.1 Synthesis of Acetoacetyl-CoA

Acetoacetyl-Pantetheine was synthesized by Oliver Piech. The stepwise CoA transformation was performed as described in section 10.7.10.

10.8.2 Loading of acetoacetyl-CoA to the ACP

50 mM Tris (pH 7.5), 20 mM KCl, 10 mM $MgCl_2$, 1.5 mg/ml *apo* ACP, 0.01 mg/ml Sfp and approx. 0.2 mg/ml acetoacetyl-CoA were incubated at 30 °C for 30 min. The protein was analysed by mass spectrometry.

10.8.3 SQTKS Methyltransferase Activity Assay

2.5 mM SAM, appox. 1 mg/mL acetoacetyl-ACP and 1 mg/mL *C*-MeT were incubated for 1 h at 30 °C. The protein was analysed by mass spectrometry.

10.9 Experimental for Chapter 7

10.9.1 Crosslinking Reaction

For the croslink approx. 10 µM of Linker-ACP and of a DH domain were incubated for 3 -4 h at 25 °C. The proteins were analysed by mass spectrometry and SDS-PAGE.

10.10 Experimental for Chapter 8

10.10.1 Loading of Hexanoyl-CoA to bfPKS ACP

For the loading reaction of Hexanoyl-CoA to bfPKS ACP following components were mixed: 50 mM Tris buffer (pH 8.8), 10 mM $MgCl_2$, 0.01 mg/mL Sfp, 1.25 mg/ml *apo* bfPKS ACP and 25 µM Hexanoyl-CoA. The mixture was incubated for 30 min at 30 °C. The protein was analysed by mass spectrometry.

10.10.2 Purification of Hexanoyl-CoA

The purification was performed with an ÄKTA Pure FPLC. The protein was loaded onto a size exclusion column (26/600 Superdex 200 pg) and the storage buffer (50 mM Tris pH 8, 150 mM NaCl, 20% glycerol (*v/v*)) was used for the SEC run. The 2 mL fractions containing the protein of interest were combined and concentrated using centrifugal concentrators (Merck) with a cut off 3000 Da.

10.10.3 Hydrolase Activity with Hexanoyl-ACP

1 mg/mL of Hexanoyl-ACP were incubated with 15 µM of *Bfl1* Hydrolase for 30 min at 30 °C. The protein was analysed by mass spectrometry.

10.10.4 Citrate Synthase Assay

1 mg/mL Hexanoyl-ACP, 1 mM oxaloacetate and 20 µM citrate synthase were incubated for 3 h at 30 °C. Protein and small molecules were analysed by mass spectrometry.

11 References

(1) Pye, C. R.; Bertin, M. J.; Lokey, R. S.; Gerwick, W. H.; Linington, R. G. Retrospective analysis of natural products provides insights for future discovery trends. *Proceedings of the National Academy of Sciences of the United States of America* **2017**, *114*, 5601–5606.

(2) Williams, D. H.; Stone, M. J.; Hauck, P. R.; Rahman, S. K. Why are secondary metabolites (natural products) biosynthesized? *Journal of natural products* **1989**, *52*, 1189–1208.

(3) Bennett, R. N.; WALLSGROVE, R. M. Secondary metabolites in plant defence mechanisms. *New Phytol* **1994**, *127*, 617–633.

(4) Wink, M. Evolution of secondary metabolites from an ecological and molecular phylogenetic perspective. *Phytochemistry* **2003**, *64*, 3–19.

(5) Beutler, J. A. Natural Products as a Foundation for Drug Discovery. *Current Protocols in Pharmacology* **2009**, *46*.

(6) Wani, M. C.; Taylor, H. L.; Wall, M. E.; Coggon, P.; McPhail, A. T. Plant antitumor agents. VI. The isolation and structure of taxol, a novel antileukemic and antitumor agent from Taxus brevifolia. *J. Am. Chem. Soc.* **1971**, *93*, 2325–2327.

(7) Weaver, B. A. How Taxol/paclitaxel kills cancer cells. *Molecular biology of the cell* **2014**, *25*, 2677–2681.

(8) Debnath, B.; Singh, W. S.; Das, M.; Goswami, S.; Singh, M. K.; Maiti, D.; Manna, K. Role of plant alkaloids on human health: A review of biological activities. *Materials Today Chemistry* **2018**, *9*, 56–72.

(9) Plowman, T. The Ethnobotany of Coca (Erythroxylum spp., Erythroxylaceae). *Advances in Economic Botany* **1984**, 62–111.

(10) Alberti, F.; Foster, G. D.; Bailey, A. M. Natural products from filamentous fungi and production by heterologous expression. *Appl Microbiol Biotechnol* **2017**, *101*, 493–500.

(11) Hamad, B. The antibiotics market. *Nature reviews. Drug discovery* **2010**, *9*, 675–676.

(12) Miller, E. L. THE PENICILLINS: A REVIEW AND UPDATE. *Journal of Midwifery & Women's Health* **2002**, *47*, 426–434.

(13) Sirtori, C. R. The pharmacology of statins. *Pharmacological research* **2014**, *88*, 3–11.

(14) Campbell, C. D.; Vederas, J. C. Biosynthesis of lovastatin and related metabolites formed by fungal iterative PKS enzymes. *Biopolymers* **2010**, *93*, 755–763.

(15) Tola, M.; Kebede, B. Occurrence, importance and control of mycotoxins: A review. *Cogent Food & Agriculture* **2016**, *2*.

(16) Ayofemi Olalekan Adeyeye, S. Aflatoxigenic fungi and mycotoxins in food: a review. *Critical reviews in food science and nutrition* **2020**, *60*, 709–721.

(17) Rushing, B. R.; Selim, M. I. Aflatoxin B1: A review on metabolism, toxicity, occurrence in food, occupational exposure, and detoxification methods. *Food and chemical toxicology : an international journal published for the British Industrial Biological Research Association* **2019**, *124*, 81–100.

(18) Alcazar-Fuoli, L.; Mellado, E. Ergosterol biosynthesis in Aspergillus fumigatus: its relevance as an antifungal target and role in antifungal drug resistance. *Frontiers in microbiology* **2012**, *3*, 439.

(19) Dittmann, J.; Wenger, R. M.; Kleinkauf, H.; Lawen, A. Mechanism of cyclosporin A biosynthesis. Evidence for synthesis via a single linear undecapeptide precursor. *Journal of Biological Chemistry* **1994**, *269*, 2841–2846.

(20) Xu, W.; Gavia, D. J.; Tang, Y. Biosynthesis of fungal indole alkaloids. *Natural product reports* **2014**, *31*, 1474–1487.

(21) Brady, R. O.; Bradley, R. M.; Trams, E. G. Biosynthesis of Fatty Acids. *Journal of Biological Chemistry* **1960**, *235*, 3093–3098.

(22) Fujii, I.; Mori, Y.; Watanabe, A.; Kubo, Y.; Tsuji, G.; Ebizuka, Y. Enzymatic synthesis of 1,3,6,8-tetrahydroxynaphthalene solely from malonyl coenzyme A by a fungal iterative type I polyketide synthase PKS1. *Biochemistry* **2000**, *39*, 8853–8858.

(23) Smith, S.; Tsai, S.-C. The type I fatty acid and polyketide synthases: a tale of two megasynthases. *Natural product reports* **2007**, *24*, 1041–1072.

(24) Cox, R. J. Polyketides, proteins and genes in fungi: programmed nano-machines begin to reveal their secrets. *Organic & biomolecular chemistry* **2007**, *5*, 2010–2026.

(25) Janßen, H. J.; Steinbüchel, A. Fatty acid synthesis in Escherichia coli and its applications towards the production of fatty acid based biofuels. *Biotechnology for biofuels* **2014**, *7*, 7.

(26) Maier, T.; Leibundgut, M.; Boehringer, D.; Ban, N. Structure and function of eukaryotic fatty acid synthases. *Quarterly reviews of biophysics* **2010**, *43*, 373–422.

(27) Lomakin, I. B.; Xiong, Y.; Steitz, T. A. The crystal structure of yeast fatty acid synthase, a cellular machine with eight active sites working together. *Cell* **2007**, *129*, 319–332.

(28) Leibundgut, M.; Maier, T.; Jenni, S.; Ban, N. The multienzyme architecture of eukaryotic fatty acid synthases. *Current opinion in structural biology* **2008**, *18*, 714–725.

(29) Jenni, S.; Leibundgut, M.; Maier, T.; Ban, N. Architecture of a fungal fatty acid synthase at 5 A resolution. *Science (New York, N.Y.)* **2006**, *311*, 1263–1267.

(30) Maier, T.; Leibundgut, M.; Ban, N. The crystal structure of a mammalian fatty acid synthase. *Science (New York, N.Y.)* **2008**, *321*, 1315–1322.

(31) McKinney, D. C.; Eyermann, C. J.; Gu, R.-F.; Hu, J.; Kazmirski, S. L.; Lahiri, S. D.; McKenzie, A. R.; Shapiro, A. B.; Breault, G. Antibacterial FabH Inhibitors with Mode of Action Validated in Haemophilus influenzae by in Vitro Resistance Mutation Mapping. *ACS infectious diseases* **2016**, *2*, 456–464.

(32) Cukier, C. D.; Hope, A. G.; Elamin, A. A.; Moynie, L.; Schnell, R.; Schach, S.; Kneuper, H.; Singh, M.; Naismith, J. H.; Lindqvist, Y.*; et al.* Discovery of an allosteric inhibitor binding site in 3-Oxo-acyl-ACP reductase from Pseudomonas aeruginosa. *ACS chemical biology* **2013**, *8*, 2518–2527.

(33) Misson, L. E.; Mindrebo, J. T.; Davis, T. D.; Patel, A.; McCammon, J. A.; Noel, J. P.; Burkart, M. D. Interfacial plasticity facilitates high reaction rate of E. coli FAS malonyl-CoA:ACP transacylase, FabD. *Proceedings of the National Academy of Sciences of the United States of America* **2020**, *117*, 24224–24233.

(34) Milligan, J. C.; Lee, D. J.; Jackson, D. R.; Schaub, A. J.; Beld, J.; Barajas, J. F.; Hale, J. J.; Luo, R.; Burkart, M. D.; Tsai, S.-C. Molecular basis for interactions between an acyl carrier protein and a ketosynthase. *Nature chemical biology* **2019**, *15*, 669–671.

(35) Mindrebo, J. T.; Patel, A.; Kim, W. E.; Davis, T. D.; Chen, A.; Bartholow, T. G.; La Clair, J. J.; McCammon, J. A.; Noel, J. P.; Burkart, M. D. Gating mechanism of elongating β-ketoacyl-ACP synthases. *Nature communications* **2020**, *11*, 1727.

(36) Nguyen, C.; Haushalter, R. W.; Lee, D. J.; Markwick, P. R. L.; Bruegger, J.; Caldara-Festin, G.; Finzel, K.; Jackson, D. R.; Ishikawa, F.; O'Dowd, B.*; et al.* Trapping the dynamic acyl carrier protein in fatty acid biosynthesis. *Nature* **2014**, *505*, 427–431.

(37) Dodge, G. J.; Patel, A.; Jaremko, K. L.; McCammon, J. A.; Smith, J. L.; Burkart, M. D. Structural and dynamical rationale for fatty acid unsaturation in Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America* **2019**, *116*, 6775–6783.

(38) Zhang, L.; Xiao, J.; Xu, J.; Fu, T.; Cao, Z.; Zhu, L.; Chen, H.-Z.; Shen, X.; Jiang, H.; Zhang, L. Crystal structure of FabZ-ACP complex reveals a dynamic seesaw-like catalytic mechanism of dehydratase in fatty acid biosynthesis. *Cell research* **2016**, *26*, 1330–1344.

(39) Wakil, S. J.; Stoops, J. K.; Joshi, V. C. Fatty acid synthesis and its regulation. *Annual review of biochemistry* **1983**, *52*, 537–579.

(40) Plate, C. A.; Joshi, V. C.; Wakil, S. J. Studies on the Mechanism of Fatty Acid Synthesis. *Journal of Biological Chemistry* **1970**, *245*, 2868–2875.

(41) Rock, C. O.; Cronan, J. E. Escherichia coli as a model for the regulation of dissociable (type II) fatty acid biosynthesis. *Biochimica et Biophysica Acta (BBA) - Lipids and Lipid Metabolism* **1996**, *1302*, 1–16.

(42) Lynen, F. On the structure of fatty acid synthetase of yeast. *European journal of biochemistry* **1980**, *112*, 431–442.

(43) Volpe, J. J.; Vagelos, P. R. Mechanisms and regulation of biosynthesis of saturated fatty acids. *Physiological reviews* **1976**, *56*, 339–417.

(44) Schweizer, E.; Hofmann, J. Microbial type I fatty acid synthases (FAS): major players in a network of cellular FAS systems. *Microbiology and molecular biology reviews : MMBR* **2004**, *68*, 501-17, table of contents.

(45) Cragg, G. M.; Newman, D. J. Natural products: a continuing source of novel drug leads. *Biochimica et biophysica acta* **2013**, *1830*, 3670–3695.

(46) Marcella, A. M.; Barb, A. W. Acyl-coenzyme A:(holo-acyl carrier protein) transacylase enzymes as templates for engineering. *Applied microbiology and biotechnology* **2018**, *102*, 6333–6341.

(47) Hertweck, C. The biosynthetic logic of polyketide diversity. *Angewandte Chemie (International ed. in English)* **2009**, *48*, 4688–4716.

(48) Staunton, J.; Weissman, K. J. Polyketide biosynthesis: a millennium review. *Natural product reports* **2001**, *18*, 380–416.

(49) Cox, R. J.; Glod, F.; Hurley, D.; Lazarus, C. M.; Nicholson, T. P.; Rudd, B. A. M.; Simpson, T. J.; Wilkinson, B.; Zhang, Y. Rapid cloning and expression of a fungal polyketide synthase gene involved in squalestatin biosynthesis. *Chemical communications (Cambridge, England)* **2004**, 2260–2261.

(50) Cannell, R. J.; Dawson, M. J.; Hale, R. S.; Hall, R. M.; Noble, D.; Lynn, S.; Taylor, N. L. The squalestatins, novel inhibitors of squalene synthase produced by a species of Phoma. IV. Preparation of fluorinated squalestatins by directed biosynthesis. *The Journal of antibiotics* **1993**, *46*, 1381–1389.

(51) Liu, N.; Hung, Y.-S.; Gao, S.-S.; Hang, L.; Zou, Y.; Chooi, Y.-H.; Tang, Y. Identification and Heterologous Production of a Benzoyl-Primed Tricarboxylic Acid Polyketide Intermediate from the Zaragozic Acid A Biosynthetic Pathway. *Organic letters* **2017**, *19*, 3560–3563.

(52) Bonsch, B.; Belt, V.; Bartel, C.; Duensing, N.; Koziol, M.; Lazarus, C. M.; Bailey, A. M.; Simpson, T. J.; Cox, R. J. Identification of genes encoding squalestatin S1 biosynthesis and in vitro production of new squalestatin analogues. *Chemical communications (Cambridge, England)* **2016**, *52*, 6777–6780.

(53) Nerud, F.; Sedmera, P.; Zouchova, Z.; Musilek, V.; Vondracek, M. Biosynthesis of mucidin, an antifungal antibiotic from basidomycete. *Collection Czechoslovak Chem. Commun.* **1982**.

(54) Iqbal, Z.; Han, L.-C.; Soares-Sello, A. M.; Nofiani, R.; Thormann, G.; Zeeck, A.; Cox, R. J.; Willis, C. L.; Simpson, T. J. Investigations into the biosynthesis of the antifungal strobilurins. *Organic & biomolecular chemistry* **2018**, *16*, 5524–5532.

(55) Nofiani, R.; Mattos-Shipley, K. de; Lebe, K. E.; Han, L.-C.; Iqbal, Z.; Bailey, A. M.; Willis, C. L.; Simpson, T. J.; Cox, R. J. Strobilurin biosynthesis in Basidiomycete fungi. *Nature communications* **2018**, *9*, 3940.

(56) Townsend, C. A.; Christensen, S. B.; Trautwein, K. Hexanoate as a starter unit in polyketide biosynthesis. *J. Am. Chem. Soc.* **1984**, *106*, 3868–3869.

(57) McKeown, D. S. J.; McNicholas, C.; Simpson, T. J.; Willett, N. J. Biosynthesis of norsolorinic acid and averufin: substrate specificity of norsolorinic acid synthase. *Chemical communications (Cambridge, England)* **1996**, 301.

(58) Brobst, S. W.; Townsend, C. A. The potential role of fatty acid initiation in the biosynthesis of the fungal aromatic polyketide aflatoxin B 1. *Can. J. Chem.* **1994**, *72*, 200–207.

(59) Crawford, J. M.; Townsend, C. A. New insights into the formation of fungal aromatic polyketides. *Nature reviews. Microbiology* **2010**, *8*, 879–889.

(60) Hertweck, C.; Luzhetskyy, A.; Rebets, Y.; Bechthold, A. Type II polyketide synthases: gaining a deeper insight into enzymatic teamwork. *Natural product reports* **2007**, *24*, 162–190.

(61) Austin, M. B.; Noel, J. P. The chalcone synthase superfamily of type III polyketide synthases. *Natural product reports* **2003**, *20*, 79–110.

(62) Liu, T.; Sanchez, J. F.; Chiang, Y.-M.; Oakley, B. R.; Wang, C. C. C. Rational domain swaps reveal insights about chain length control by ketosynthase domains in fungal nonreducing polyketide synthases. *Organic letters* **2014**, *16*, 1676–1679.

(63) He, Y.; Cox, R. J. The molecular steps of citrinin biosynthesis in fungi. *Chemical science* **2016**, *7*, 2119–2127.

(64) Beck, J.; Ripka, S.; Siegner, A.; Schiltz, E.; Schweizer, E. The multifunctional 6-methylsalicylic acid synthase gene of Penicillium patulum. Its gene structure relative to that of other polyketide synthases. *European journal of biochemistry* **1990**, *192*, 487–498.

(65) Abegaz, B. M.; Kinfe, H. H. Secondary metabolites, their structural diversity, bioactivity, and ecological functions: An overview. *Physical Sciences Reviews* **2019**, *4*.

(66) Hasumi, K.; Tachikawa, K.; Sakai, K.; Murakawa, S.; Yoshikawa, N.; Kumazawa, S.; Endo, A. Competitive inhibition of squalene synthetase by squalestatin 1. *The Journal of antibiotics* **1993**, *46*, 689–691.

(67) Jones, C. A.; Sidebottom, P. J.; Cannell, R. J.; Noble, D.; Rudd, B. A. The squalestatins, novel inhibitors of squalene synthase produced by a species of Phoma. III. Biosynthesis. *The Journal of antibiotics* **1992**, *45*, 1492–1498.

(68) Byrne, K. M.; Arison, B. H.; Nallin-Omstead, M.; Kaplan, L. Biosynthesis of the zaragozic acids. 1. Zaragozic acid A. *J. Org. Chem.* **1993**, *58*, 1019–1024.

(69) Lebe, K. E.; Cox, R. J. Oxidative steps during the biosynthesis of squalestatin S1. *Chemical science* **2019**, *10*, 1227–1231.

(70) Hantke, V.; Skellam, E. J.; Cox, R. J. Evidence for enzyme catalysed intramolecular 4+2 Diels-Alder cyclization during the biosynthesis of pyrichalasin H. *Chemical communications (Cambridge, England)* **2020**, *56*, 2925–2928.

(71) Robbins, T.; Kapilivsky, J.; Cane, D. E.; Khosla, C. Roles of Conserved Active Site Residues in the Ketosynthase Domain of an Assembly Line Polyketide Synthase. *Biochemistry* **2016**, *55*, 4476–4484.

(72) Arnstadt, K. I.; Schindlbeck, G.; Lynen, F. Zum Mechanismus der Kondensationsreaktion der Fettsäurebiosynthese. *European journal of biochemistry* **1975**, *55*, 561–571.

(73) Musiol-Kroll, E. M.; Wohlleben, W. Acyltransferases as Tools for Polyketide Synthase Engineering. *Antibiotics (Basel, Switzerland)* **2018**, *7*.

(74) Tang, Y.; Kim, C.-Y.; Mathews, I. I.; Cane, D. E.; Khosla, C. The 2.7-Angstrom crystal structure of a 194-kDa homodimeric fragment of the 6-deoxyerythronolide B synthase. *Proceedings of the National Academy of Sciences of the United States of America* **2006**, *103*, 11124–11129.

(75) Tang, Y.; Chen, A. Y.; Kim, C.-Y.; Cane, D. E.; Khosla, C. Structural and mechanistic analysis of protein interactions in module 3 of the 6-deoxyerythronolide B synthase. *Chemistry & biology* **2007**, *14*, 931–943.

(76) Liew, C. W.; Nilsson, M.; Chen, M. W.; Sun, H.; Cornvik, T.; Liang, Z.-X.; Lescar, J. Crystal structure of the acyltransferase domain of the iterative polyketide synthase in enediyne biosynthesis. *The Journal of biological chemistry* **2012**, *287*, 23203–23215.

(77) Chan, D. I.; Vogel, H. J. Current understanding of fatty acid biosynthesis and the acyl carrier protein. *The Biochemical journal* **2010**, *430*, 1–19.

(78) Byers, D. M.; Gong, H. Acyl carrier protein: structure-function relationships in a conserved multifunctional protein family. *Biochemistry and cell biology = Biochimie et biologie cellulaire* **2007**, *85*, 649–662.

(79) Holak, T. A.; Kearsley, S. K.; Kim, Y.; Prestegard, J. H. Three-dimensional structure of acyl carrier protein determined by NMR pseudoenergy and distance geometry calculations. *Biochemistry* **1988**, *27*, 6135–6142.

(80) Kim, Y.; Prestegard, J. H. A dynamic model for the structure of acyl carrier protein in solution. *Biochemistry* **1989**, *28*, 8792–8797.

(81) Zornetzer, G. A.; Fox, B. G.; Markley, J. L. Solution structures of spinach acyl carrier protein with decanoate and stearate. *Biochemistry* **2006**, *45*, 5217–5227.

(82) Gallagher, J. R.; Prigge, S. T. Plasmodium falciparum acyl carrier protein crystal structures in disulfide-linked and reduced states and their prevalence during blood stage growth. *Proteins* **2010**, *78*, 575–588.

(83) Roujeinikova, A.; Baldock, C.; Simon, W. J.; Gilroy, J.; Baker, P. J.; Stuitje, A. R.; Rice, D. W.; Slabas, A. R.; Rafferty, J. B. X-Ray Crystallographic Studies on Butyryl-ACP Reveal Flexibility of the Structure around a Putative Acyl Chain Binding Site. *Structure* **2002**, *10*, 825–835.

(84) Lim, J.; Sun, H.; Fan, J.-S.; Hameed, I. F.; Lescar, J.; Liang, Z.-X.; Yang, D. Rigidifying acyl carrier protein domain in iterative type I PKS CalE8 does not affect its function. *Biophysical journal* **2012**, *103*, 1037–1044.

(85) D'agnony, G.; Rosenfeld, I. S.; Awaya, J.; Ōmura, S.; Vagelos, P. Inhibition of fatty acid synthesis by the antibiotic cerulenin. *Biochimica et Biophysica Acta (BBA) - Lipids and Lipid Metabolism* **1973**, *326*, 155–166.

(86) Takeshima, H.; Kitao, C.; Omura, S. Inhibition of the biosynthesis of leucomycin, a macrolide antibiotic, by cerulenin. *Journal of biochemistry* **1977**, *81*, 1127–1132.

(87) Funabashi, H.; Kawaguchi, A.; Tomoda, H.; Omura, S.; Okuda, S.; Iwasaki, S. Binding site of cerulenin in fatty acid synthetase. *Journal of biochemistry* **1989**, *105*, 751–755.

(88) Worthington, A. S.; Rivera, H.; Torpey, J. W.; Alexander, M. D.; Burkart, M. D. Mechanism-based protein cross-linking probes to investigate carrier protein-mediated biosynthesis. *ACS chemical biology* **2006**, *1*, 687–691.

(89) Herbst, D. A.; Huitt-Roehl, C. R.; Jakob, R. P.; Kravetz, J. M.; Storm, P. A.; Alley, J. R.; Townsend, C. A.; Maier, T. The structural organization of substrate loading in iterative polyketide synthases. *Nature chemical biology* **2018**, *14*, 474–479.

(90) Bruegger, J.; Haushalter, R. W.; Haushalter, B.; Vagstad, A. L.; Vagstad, A.; Shakya, G.; Mih, N.; Townsend, C. A.; Burkart, M. D.; Tsai, S.-C. Probing the selectivity and protein·protein interactions of a nonreducing fungal polyketide synthase using mechanism-based crosslinkers. *Chemistry & biology* **2013**, *20*, 1135–1146.

(91) Miyanaga, A.; Iwasawa, S.; Shinohara, Y.; Kudo, F.; Eguchi, T. Structure-based analysis of the molecular interactions between acyltransferase and acyl carrier

protein in vicenistatin biosynthesis. *Proceedings of the National Academy of Sciences of the United States of America* **2016**, *113*, 1802–1807.

(92) Miyanaga, A.; Ouchi, R.; Ishikawa, F.; Goto, E.; Tanabe, G.; Kudo, F.; Eguchi, T. Structural Basis of Protein-Protein Interactions between a trans-Acting Acyltransferase and Acyl Carrier Protein in Polyketide Disorazole Biosynthesis. *J. Am. Chem. Soc.* **2018**, *140*, 7970–7978.

(93) Ma, S. M.; Tang, Y. Biochemical characterization of the minimal polyketide synthase domains in the lovastatin nonaketide synthase LovB. *The FEBS journal* **2007**, *274*, 2854–2864.

(94) Cochrane, R. V. K.; Gao, Z.; Lambkin, G. R.; Xu, W.; Winter, J. M.; Marcus, S. L.; Tang, Y.; Vederas, J. C. Comparison of 10,11-Dehydrocurvularin Polyketide Synthases from Alternaria cinerariae and Aspergillus terreus Highlights Key Structural Motifs. *Chembiochem : a European journal of chemical biology* **2015**, *16*, 2479–2483.

(95) Winter, J. M.; Cascio, D.; Dietrich, D.; Sato, M.; Watanabe, K.; Sawaya, M. R.; Vederas, J. C.; Tang, Y. Biochemical and Structural Basis for Controlling Chemical Modularity in Fungal Polyketide Biosynthesis. *J. Am. Chem. Soc.* **2015**, *137*, 9885–9893.

(96) David Ivison. *Investigating the Programming of Investigating the Programming of Type I Iterative Polyketide Synthase Enzymes*; PhD Thesis: Bristol, **2013**.

(97) Douglas Roberts. *Investigating the Programming of Type I Highly Reducing Iterative Polyketide Synthases*; PhD Thesis: Bristol, **2014**.

(98) Emma C. Liddle. *Kinetics and Selectivity of an Isolated Dehydratase Domain from a Fungal Polyketide Synthase*; PhD Thesis: Bristol, **2018**.

(99) Christoph Bartel. *Investigation of Isolated Domains of the Squalestatin Tetraketide Synthase and Tailoring Enzymes Involved in the Biosynthesis of Squalestatin S1*; PhD Thesis: Hannover, **2017**.

(100) Hao Yao. *In Vitro Investigation of Multi-domain Fragments of Squalestatin Tetraketide Synthase*; PhD Thesis: Hannover, **2019**.

(101) Oliver Piech. *Computational and In Vitro Study of Isolated Domains from Fungal Polyketide Synthases*; PhD Thesis: Hannover, **2020.**

(102) Hebditch, M.; Carballo-Amador, M. A.; Charonis, S.; Curtis, R.; Warwicker, J. Protein-Sol: a web tool for predicting protein solubility from sequence. *Bioinformatics (Oxford, England)* **2017**, *33*, 3098–3100.

(103) Chan, P.; Curtis, R. A.; Warwicker, J. Soluble expression of proteins correlates with a lack of positively-charged surface. *Scientific reports* **2013**, *3*, 3333.

(104) Fuhrmann, M.; Hausherr, A.; Ferbitz, L.; Schödl, T.; Heitzer, M.; Hegemann, P. Monitoring dynamic expression of nuclear genes in Chlamydomonas reinhardtii by using a synthetic luciferase reporter gene. *Plant Mol Biol* **2004**, *55*, 869–881.

(105) Puigbò, P.; Romeu, A.; Garcia-Vallvé, S. HEG-DB: a database of predicted highly expressed genes in prokaryotic complete genomes under translational selection. *Nucleic acids research* **2008**, *36*, D524-7.

(106) Puigbò, P.; Guzmán, E.; Romeu, A.; Garcia-Vallvé, S. OPTIMIZER: a web server for optimizing the codon usage of DNA sequences. *Nucleic acids research* **2007**, *35*, W126-31.

(107) Ferrer, M.; Chernikova, T. N.; Timmis, K. N.; Golyshin, P. N. Expression of a temperature-sensitive esterase in a novel chaperone-based Escherichia coli strain. *Applied and environmental microbiology* **2004**, *70*, 4499–4504.

(108) Schein, C. H. Production of Soluble Recombinant Proteins in Bacteria. *Nat Biotechnol* **1989**, *7*, 1141–1149.

(109) Saibil, H. Chaperone machines for protein folding, unfolding and disaggregation. *Nature reviews. Molecular cell biology* **2013**, *14*, 630–642.

(110) Carrió, M. M.; Villaverde, A. Localization of chaperones DnaK and GroEL in bacterial inclusion bodies. *Journal of bacteriology* **2005**, *187*, 3599–3601.

(111) Singh, A.; Upadhyay, V.; Upadhyay, A. K.; Singh, S. M.; Panda, A. K. Protein recovery from inclusion bodies of Escherichia coli using mild solubilization process. *Microbial cell factories* **2015**, *14*, 41.

(112) Sachdev, D.; Chirgwin, J. M. Solubility of proteins isolated from inclusion bodies is enhanced by fusion to maltose-binding protein or thioredoxin. *Protein expression and purification* **1998**, *12*, 122–132.

(113) An Overview of the Top Ten Detergents Used for Membrane Protein Crystallization. *Crystals* **2017**, *7*, 197.

(114) Anfinsen, C. B. Principles that govern the folding of protein chains. *Science (New York, N.Y.)* **1973**, *181*, 223–230.

(115) Cabrita, L. D.; Bottomley, S. P. Protein expression and refolding – A practical guide to getting the most out of inclusion bodies; Biotechnology Annual Review; Elsevier, 2004; pp 31–50.

(116) Soleymani, B.; Mostafaie, A. Analysis of Methods to Improve the Solubility of Recombinant Bovine Sex Determining Region Y Protein. *Reports of biochemistry & molecular biology* **2019**, *8*, 227–235.

(117) Hoffman, B. J.; Broadwater, J. A.; Johnson, P.; Harper, J.; Fox, B. G.; Kenealy, W. R. Lactose fed-batch overexpression of recombinant metalloproteins in Escherichia coli BL21 (DE3): process control yielding high levels of metal-incorporated, soluble protein. *Protein expression and purification* **1995**, *6*, 646–654.

(118) Ulrich, H. D. SUMO Protocols. Preface. *Methods in molecular biology (Clifton, N.J.)* **2009**, *497*, v–vi.

(119) Marblestone, J. G.; Edavettal, S. C.; Lim, Y.; Lim, P.; Zuo, X.; Butt, T. R. Comparison of SUMO fusion technology with traditional gene fusion systems: enhanced expression and solubility with SUMO. *Protein science : a publication of the Protein Society* **2006**, *15*, 182–189.

(120) Wang, T.; Badran, A. H.; Huang, T. P.; Liu, D. R. Continuous directed evolution of proteins with improved soluble expression. *Nature chemical biology* **2018**, *14*, 972–980.

(121) Braun, P.; LaBaer, J. High throughput protein production for functional proteomics. *Trends in Biotechnology* **2003**, *21*, 383–388.

(122) Wattana-amorn, P.; Williams, C.; Płoskoń, E.; Cox, R. J.; Simpson, T. J.; Crosby, J.; Crump, M. P. Solution structure of an acyl carrier protein domain from a fungal type I polyketide synthase. *Biochemistry* **2010**, *49*, 2186–2193.

(123) Ma, Y.; Smith, L. H.; Cox, R. J.; Beltran-Alvarez, P.; Arthur, C. J.; Simpson F R S, T. J. Catalytic relationships between type I and type II iterative polyketide synthases: The Aspergillus parasiticus norsolorinic acid synthase. *Chembiochem : a European journal of chemical biology* **2006**, *7*, 1951–1958.

(124) Cox, R. J.; Hitchman, T. S.; Byrom, K. J.; Findlow, I. S.; Tanner, J. A.; Crosby, J.; Simpson, T. J. Post-translational modification of heterologously expressed Streptomyces type II polyketide synthase acyl carrier proteins. *FEBS letters* **1997**, *405*, 267–272.

(125) Evans, S. E.; Williams, C.; Arthur, C. J.; Burston, S. G.; Simpson, T. J.; Crosby, J.; Crump, M. P. An ACP structural switch: conformational differences between the apo and holo forms of the actinorhodin polyketide synthase acyl carrier protein. *Chembiochem : a European journal of chemical biology* **2008**, *9*, 2424–2432.

(126) Evans, S. E.; Williams, C.; Arthur, C. J.; Płoskoń, E.; Wattana-amorn, P.; Cox, R. J.; Crosby, J.; Willis, C. L.; Simpson, T. J.; Crump, M. P. Probing the Interactions of early polyketide intermediates with the Actinorhodin ACP from S. coelicolor A3(2). *Journal of molecular biology* **2009**, *389*, 511–528.

(127) Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F. T.; Beer, T. A. P. de; Rempfer, C.; Bordoli, L.*; et al.* SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic acids research* **2018**, *46*, W296-W303.

(128) Crump, M. P.; Crosby, J.; Dempsey, C. E.; Parkinson, J. A.; Murray, M.; Hopwood, D. A.; Simpson, T. J. Solution structure of the actinorhodin polyketide synthase acyl carrier protein from Streptomyces coelicolor A3(2). *Biochemistry* **1997**, *36*, 6000–6008.

(129) Keatinge-Clay, A. T.; Maltby, D. A.; Medzihradszky, K. F.; Khosla, C.; Stroud, R. M. An antibiotic factory caught in action. *Nature structural & molecular biology* **2004**, *11*, 888–893.

(130) Iavarone, A. T.; Williams, E. R. Mechanism of charging and supercharging molecules in electrospray ionization. *J. Am. Chem. Soc.* **2003**, *125*, 2319–2327.

(131) Marchese, R.; Grandori, R.; Carloni, P.; Raugei, S. A computational model for protein ionization by electrospray based on gas-phase basicity. *Journal of the American Society for Mass Spectrometry* **2012**, *23*, 1903–1910.

(132) Schnier, P. D.; Gross, D. S.; Williams, E. R. On the maximum charge state and proton transfer reactivity of peptide and protein ions formed by electrospray ionization. *Journal of the American Society for Mass Spectrometry* **1995**, *6*, 1086–1097.

(133) Ogorzalek Loo, R. R.; Stevenson, T. I.; Mitchell, C.; Loo, J. A.; Andrews, P. C. Mass spectrometry of proteins directly from polyacrylamide gels. *Analytical chemistry* **1996**, *68*, 1910–1917.

(134) Zaia, J.; Annan, R. S.; Biemann, K. The correct molecular weight of myoglobin, a common calibrant for mass spectrometry. *Rapid communications in mass spectrometry : RCM* **1992**, *6*, 32–36.

(135) *MaxEnt: An Essential Maximum Entropy Based Tool for Interpreting Multiply-Charged Electrospray Data*.

(136) Separation Methods Technologies, Inc. C5: Pentyl Columns. https:// separationmethods.com/product-category/reversed-phase-columns/c5-pentyl-columns/ (accessed February 25, 2021).

(137) Reuter, K.; Mofid, M. R.; Marahiel, M. A.; Ficner, R. Crystal structure of the surfactin synthetase-activating enzyme sfp: a prototype of the 4'-phosphopantetheinyl transferase superfamily. *The EMBO journal* **1999**, *18*, 6823–6831.

(138) Keszenman-Pereyra, D.; Lawrence, S.; Twfieg, M.-E.; Price, J.; Turner, G. The npgA/ cfwA gene encodes a putative 4'-phosphopantetheinyl transferase which is essential for penicillin biosynthesis in Aspergillus nidulans. *Current genetics* **2003**, *43*, 186–190.

(139) Geoghegan, K. F.; Dixon, H. B.; Rosner, P. J.; Hoth, L. R.; Lanzetti, A. J.; Borzilleri, K. A.; Marr, E. S.; Pezzullo, L. H.; Martin, L. B.; LeMotte, P. K.*; et al.* Spontaneous alpha-N-6-phosphogluconoylation of a "His tag" in Escherichia coli: the cause of extra mass of 258 or 178 Da in fusion proteins. *Analytical biochemistry* **1999**, *267*, 169–184.

(140) Beránek, M.; Drsata, J.; Palicka, V. Inhibitory effect of glycation on catalytic activity of alanine aminotransferase. *Molecular and cellular biochemistry* **2001**, *218*, 35–39.

(141) Aon, J. C.; Caimi, R. J.; Taylor, A. H.; Lu, Q.; Oluboyede, F.; Dally, J.; Kessler, M. D.; Kerrigan, J. J.; Lewis, T. S.; Wysocki, L. A.*; et al.* Suppressing posttranslational gluconoylation of heterologous proteins by metabolic engineering of Escherichia coli. *Applied and environmental microbiology* **2008**, *74*, 950–958.

(142) Brown, T. *Gene cloning and DNA analysis: An introduction*, Sixth edition, [5th printing]; Wiley-Blackwell: Chichester, West Sussex, **2010**.

(143) Akey, D. L.; Razelun, J. R.; Tehranisa, J.; Sherman, D. H.; Gerwick, W. H.; Smith, J. L. Crystal structures of dehydratase domains from the curacin polyketide biosynthetic pathway. *Structure* **2010**, *18*, 94–105.

(144) Keatinge-Clay, A. Crystal structure of the erythromycin polyketide synthase dehydratase. *Journal of molecular biology* **2008**, *384*, 941–953.

(145) Leesong, M.; Henderson, B. S.; Gillig, J. R.; Schwab, J. M.; Smith, J. L. Structure of a dehydratase-isomerase from the bacterial pathway for biosynthesis of unsaturated fatty acids: two catalytic activities in one active site. *Structure* **1996**, *4*, 253–264.

(146) Liddle, E.; Scott, A.; Han, L.-C.; Ivison, D.; Simpson, T. J.; Willis, C. L.; Cox, R. J. In vitro kinetic study of the squalestatin tetraketide synthase dehydratase reveals the stereochemical course of a fungal highly reducing polyketide synthase. *Chemical communications (Cambridge, England)* **2017**, *53*, 1727–1730.

(147) Emma C. Liddle. *Studies on a Type II Polyketide Synthase: A Chain Initiation Factor*; PhD Thesis: Bristol, 2004.

(148) Hammes, G. G. Fatty acid synthase: elementary steps in catalysis and regulation. *Current topics in cellular regulation* **1985**, *26*, 311–324.

(149) Schwab, J. M.; Klassen, J. B.; Habib, A. Stereochemical course of the hydration reaction catalysed by β-hydroxydecanoylthioester dehydrase. *J. Chem. Soc., Chem. Commun.* **1986**, 357–358.

(150) Finzel, K.; Nguyen, C.; Jackson, D. R.; Gupta, A.; Tsai, S.-C.; Burkart, M. D. Probing the Substrate Specificity and Protein-Protein Interactions of the E. coli Fatty Acid Dehydratase, FabA. *Chemistry & biology* **2015**, *22*, 1453–1460.

(151) Finn, R. D.; Clements, J.; Eddy, S. R. HMMER web server: interactive sequence similarity searching. *Nucleic acids research* **2011**, *39*, W29-37.

(152) Nazi, I.; Koteva, K. P.; Wright, G. D. One-pot chemoenzymatic preparation of coenzyme A analogues. *Analytical biochemistry* **2004**, *324*, 100–105.

(153) Pfrengle, F.; Dekaris, V.; Schefzig, L.; Zimmer, R.; Reissig, H.-U. Indium Trichloride Mediated Cleavage of Acetonides in the Presence of Acid-Labile Functional Groups - Enhancing the Synthetic Utility of 1,3-Dioxolanyl-Substituted 1,2-Oxazines. *Synlett* **2008**, *2008*, 2965–2968.

(154) Herbst, D. A.; Jakob, R. P.; Zähringer, F.; Maier, T. Mycocerosic acid synthase exemplifies the architecture of reducing polyketide synthases. *Nature* **2016**, *531*, 533–537.

(155) Konno, S.; La Clair, J. J.; Burkart, M. D. Trapping the Complex Molecular Machinery of Polyketide and Fatty Acid Synthases with Tunable Silylcyanohydrin Crosslinkers. *Angewandte Chemie (International ed. in English)* **2018**, *57*, 17009–17013.

(156) HAMOR, T. A.; PAUL, I. C.; ROBERTSON, J. M.; SIM, G. A. The structure of byssochlamic acid. *Experientia* **1962**, *18*, 352–354.

(157) Mattos-Shipley, K. M. J. de; Spencer, C. E.; Greco, C.; Heard, D. M.; O'Flynn, D. E.; Dao, T. T.; Song, Z.; Mulholland, N. P.; Vincent, J. L.; Simpson, T. J.*; et al.* Uncovering biosynthetic relationships between antifungal nonadrides and octadrides. *Chemical science* **2020**, *11*, 11570–11578.

(158) Raistrick, H.; Smith, G. Studies in the biochemistry of micro-organisms: The metabolic products of Byssochlamys fulva Olliver and Smith. *The Biochemical journal* **1933**, *27*, 1814–1819.

(159) Stork, G.; Tabak, J. M.; Blount, J. F. Total synthesis of (+-)-byssochlamic acid. *J. Am. Chem. Soc.* **1972**, *94*, 4735–4737.

(160) White, J. D.; Kim, J.; Drapela, N. E. Enantiospecific Synthesis of (+)-Byssochlamic Acid, a Nonadride from the Ascomycete Byssochlamysfulva. *J. Am. Chem. Soc.* **2000**, *122*, 8665–8671.

(161) Barton, D. H. R.; Sutherland, J. K. 329. The nonadrides. Part I. Introduction and general survey. *J. Chem. Soc.* **1965**, 1769.

(162) Bloomer, J. L.; Moppett, C. E.; Sutherland, J. K. The biosynthesis of glauconic acid. *Chem. Commun. (London)* **1965**, 619.

(163) HANSON, K. R.; ROSE, I. A. THE ABSOLUTE STEREOCHEMICAL COURSE OF CITRIC ACID BIOSYNTHESIS. *Proceedings of the National Academy of Sciences of the United States of America* **1963**, *50*, 981–988.

(164) Eggerer, H.; Buckel, W.; Lenz, H.; Wunderwald, P.; Gottschalk, G.; Cornforth, J. W.; Donninger, C.; Mallaby, R.; Redmond, J. W. Stereochemistry of enzymic citrate synthesis and cleavage. *Nature* **1970**, *226*, 517–519.

(165) Weidman, S. W.; Drysdale, G. R.; Mildvan, A. S. Interaction of a spin-labeled analog of acetyl coenzyme A with citrate synthase. Paramagnetic resonance and proton relaxation rate studies of binary and ternary complexes. *Biochemistry* **1973**, *12*, 1874–1883.

(166) BRADY, R. O. Fluoroacetyl coenzyme A. *Journal of Biological Chemistry* **1955**, *217*, 213–224.

(167) Fanshier, D. W.; Gottwald, L. K.; Kun, E. Enzymatic Synthesis of Monofluorocitrate from β-Fluoro-oxaloacetate. *Journal of Biological Chemistry* **1962**, *237*, 3588–3596.

(168) Remington, S.; Wiegand, G.; Huber, R. Crystallographic refinement and atomic models of two different forms of citrate synthase at 2.7 and 1.7 A resolution. *Journal of molecular biology* **1982**, *158*, 111–152.

(169) Wiegand, G.; Remington, S. J. Citrate synthase: structure, control, and mechanism. *Annual review of biophysics and biophysical chemistry* **1986**, *15*, 97–117.

(170) Karpusas, M.; Branchaud, B.; Remington, S. J. Proposed mechanism for the condensation reaction of citrate synthase: 1.9-A structure of the ternary complex with oxaloacetate and carboxymethyl coenzyme A. *Biochemistry* **1990**, *29*, 2213–2219.

(171) Williams, K.; Szwalbe, A. J.; Mulholland, N. P.; Vincent, J. L.; Bailey, A. M.; Willis, C. L.; Simpson, T. J.; Cox, R. J. Heterologous Production of Fungal Maleidrides Reveals the Cryptic Cyclization Involved in their Biosynthesis. *Angewandte Chemie (International ed. in English)* **2016**, *55*, 6784–6788.

(172) Claxton, H. B.; Akey, D. L.; Silver, M. K.; Admiraal, S. J.; Smith, J. L. Structure and functional analysis of RifR, the type II thioesterase from the rifamycin biosynthetic pathway. *Journal of Biological Chemistry* **2009**, *284*, 5021–5029.

(173) Xu, W.; Chooi, Y.-H.; Choi, J. W.; Li, S.; Vederas, J. C.; Da Silva, N. A.; Tang, Y. LovG: the thioesterase required for dihydromonacolin L release and lovastatin nonaketide synthase turnover in lovastatin biosynthesis. *Angewandte Chemie (International ed. in English)* **2013**, *52*, 6472–6475.

(174) Fujii, R.; Matsu, Y.; Minami, A.; Nagamine, S.; Takeuchi, I.; Gomi, K.; Oikawa, H. Biosynthetic Study on Antihypercholesterolemic Agent Phomoidride: General Biogenesis of Fungal Dimeric Anhydrides. *Organic letters* **2015**, *17*, 5658–5661.

(175) Bond-Watts, B. B.; Weeks, A. M.; Chang, M. C. Y. Biochemical and structural characterization of the trans-enoyl-CoA reductase from Treponema denticola. *Biochemistry* **2012**, *51*, 6827–6837.

12. Appendix

KS/AT *E. coli* codon optimized sequence:

ATGGTGCCGTATTATCAGCCGGCGAGCAGCTGCGGCAGCAATACCATGGCG
GCGATGGATGAACATCAGCATAATGAAGATGCGACCATTCCGATTGCGATT
ATTGGCATGAGCTGCCGCTTTCCGGGCAATGCGACCAGCCCGGAAAAACTG
TGGGAACTGTGCGCGCAGGGCCGCAGCGCGTGGAGCAGCATTCCGAAAAGC
CGCTTTCGCCAGGAAGGCTTTTATAATCCGAATGCGGAACGCGTGGGCACC
AGCCATGTGGTGGGCGGCCATTTTCTGGAAGAAGATCCGAGCCTGTTTGAT
GCGAGCTTTTTTAATCTGAGCGCGGAAGCGGCGAAAACCATGGATCCGCAG
TTTCGCCTGCAGCTGGAAAGCGTGTATGAAGCGATGGAAAGCGCGGGCATT
ACCCTGGAACATATTGCGGGCAGCGATACCAGCGTGTATGCGGGCGCGTGC
TTTCGCGATTATCATGATAGCCTGGTGCGCGATCCGGATCTGGTGCCGCGCT
TTCTGCTGACCGGCAATGGCGCGGCGATGAGCAGCAATCGCGTGAGCCATT
TTTATGATCTGCGCGGCGCGAGCATGACCGTGGATACCGGCTGCAGCACCA
CCCTGACCGCGCTGCATCTGGCGTGCCAGGGCCTGCGCAATCGCGAAAGCA
AAACCAGCATTGTGACCGGCGCGAATGTGATTCTGAATCCGGATATGTTTGT
GACCATGAGCAGCCTGGGCCTGCTGGGCCCGGAAGGCAAAAGCCATACCTT
TGATGCGCGCGCGAATGGCTATGGCCGCGGCGAAGGCATTGCGACCGTGAT
TATTAAACGCCTGGATGATGCGCTGCGCGCGCAGGATCCGATTCGCTGCATT
ATTCGCGGCACCGCGCTGAATCAGGATGGCCGCACCGCGACCCTGACCAGC
CCGAGCCAGACCGCGCAGAGCGATCTGATTCGCGCGTGCTATCGCGCGGCG
GCGCTGGATCCGAATGATACCGCGTTTCTGGCGGCGCATGGCACCGGCACC
CGCACCGGCGATGCGGTGGAAATTGCGGCGGCGGCGGATGTGTTTGGCGAA
AAACGCAGCCCGGAACGCCCGCTGTGGATTGGCAGCGTGAAAACCAATATT
GGCCATAGCGAAGCGACCAGCGGCCTGGCGAGCGTGATTCAGGCGGCGCTG
GCGCTGGAAAAAGGCCTGATTCCGCCGAATATTAATTTTAAAGAACCGAAT
GAAAAACTGGGCCAGGTGAGCGCGGCGGTGCGCGTGCCGAGCAATCTGCA
GAAATGGCCGAGCGTGAGCGGCGTGCGCCGCGCGAGCGTGAATAATTTTGG
CTATGGCGGCGCGAATGCGCATGTGATTCTGGAAAGCGGCATTCCGGGCCA
TACCCCGATTGCGAATGGCAGCGGCCGCAGCAATGGCACCGGCAATGGCCA
TAATGGCGCGAATGGCACCACCAATGGCCATAATGGCACCAATGGCACCAC
CAATGGCCATTTTGATGCGACCCAGGCGACCAATGGCCATTATGGCACCGA

TGAAACCCCGGATTATGCGCCGCTGGATAGCTTTGTGATTAGCATTAGCGCG
AAAGAAGAAGCGAGCGCGCGCAGCATGGTGACCAATCTGGCGGATTATCTG
CGCACCCTGCAGGTTCAGGATGAAACCAAACATTTTAAAAGCATTGCGCAT
ACCCTGGGCAGCCATCGCAGCATGTTTAAATGGACCGCGGCGAAAAGCATT
ACCGGCCCGGAAGAACTGATTGCGGCGGCGGAAGGCGGCCAGTTTCAGGCG
AGCCGCGCGCTGGAACGCACCCGCCTGGGCTTTGTGTTTACCGGCCAGGGC
GCGCAGTGGTTTGCGATGGGCCGCGAACTGATTAATACCTATCCGGTGTTTC
GCCAGAGCCTGGATCGCGCGGATCGCTATCTGAAAGAATTTGGCTGCGAAT
GGAGCATTATTGATGAACTGAGCCGCGATGCGGAAAATAGCAATGTGAATG
ATATGACCCTGAGCCCGCCGCTGTGCACCGCGGTGCAGATTAGCCTGGTGC
AGCTGCTGGAAAGCTGGGGCATTGTGCCGACCGCGGTGACCGGCCATAGCA
GCGGCGAAATTGCGGCGGCGTATGCGGCGGGCGCGCTGGATTTTAAAAGCG
CGATGGCGGTGACCTATTTTCGCGGCGAAGTGGGCCTGGCGTGCCAGGATA
AAATTGTGGGCAAAGGCGGCATGATTGCGGTGGGCCTGGGCCCGGAAGATG
CGGAAGATCGCATTGCGCGCGTGCAGAGCGGCAAAATTGTGGTGGCGTGCA
TTAATAGCCAGAGCAGCGTGACCGTGAGCGGCGATCTGAGCGGCATTGTGG
AACTGGAAGATCTGCTGAAAGCGGAAGGCGTGTTTGCGCGCCGCGTGAAAG
TGCAGGCGGCGTATCATAGCCATCACATGCAGGTTATTGCGAATGGCTATCT
GACCAGCCTGAAAGATATGCTGAAACCGACCAAAAAATTTGGCAAAATTAT
TTATAGCAGCCCGACCACCGGCCGCCGCGAAACCAATGCGAAACTGATGGC
GAGCGCGCAGCATTGGGTGAATAATATGCTGAGCCCGGTGCGCTTTGCGGA
AAGCTTTCAGAATATGTGCTTTAGCAATCGCAATAGCAGCCAGAGCGAAGA
AATTTTTCAGGATGTGGATATTGTGCTGGAAGTGGGCCCGCATGGCATGCTG
CAGGGCCCGATTCAGCAGATGATGAGCCTGCCGATTTTTGAACGCGCGCGC
CTGCCGTATATTAGCTGCCTGCTGCGCGGCCAGAGCGCGGTGCATACCATGC
AGACCGTGGCGGCGGGCCTGATGGGCTGGGGCTATCGCGTGGATATGGTGG
CGGTGAATTTTCCGCAGGGCACCTATGGCGTGAAAATTCTGCATGATCTGCC
GAGCTATCCGTGGAATCATGATAATAGCCATTGGTGGGAACCGCGCCTGAA
TAAAGCGCATCGCCAGCGCGTGCATCCGCCGCATTAA

BfPKS ACP *E. coli* codon optimized sequence including the restriction sites:

CATATGAGTGAAGGTGATCGTGCGAACGCGTTCCGTGATTCTCTGGCTCTGG
CTCGTACCATCGAAGAAGCGTCTGAACTGGTTTGCGCGGCGCTGGTTACCA
AACTGGCGGCGCGTAGCGGTATCTCTCCGGAAAACGTTGATCCGTCTAAAA

CCGTTTCTGAATATGGTGTTGATTCTCTGGTTGCTGTTGAACTGCGTAACTG
GATCACCCACGAAATGGATTCTACCGTTCCGATCCTGGAACTGCTGGCTAAC
AACCCGATGAACAGCCTGAGCGTTAAAATCGCTTCTCGTTCTAAACTGGTTC
ACCTGGATGCTGAAAAAGAATAACTCGAG

BfPKS DH *E. coli* codon optimized sequence including the restriction sites:

CATATGTCTAAAACCTTCCGTACCCGTCAGTTCCCGCGTACCGATCTGCTGG
GTGCGCTGGATCGTTCCTCTAACCCGTGGGAACCGCGTTGGCACAACCACAT
CCGTCTGTCTGAAAAACCGTGGGTGCTGGATCACAAAATTCAGTCTAACAC
CGTGTACCCGGCGGCGGGCTTCATTGCGATGGCGATCGAAGCGGTTTACCA
GCAGCGTAAAGATAACAGCAGCAAACCGATCAGCGGTTTCAAACTGAAAG
ATGTTAACATCGGCAGCGCGCTGCTGATCCCGGAAGAAGAAGAAACTATCG
AAACTCTGGTTACCCTGAAACGTCACTTCGATCACGCGCGCAGCCCGGATA
CCTACTGGAACGAATTCCACGTTTACTCCGTGACCGCGGCGAACACCTGGA
CCGAACACTGCCGTGGCCTGGTTACCGTTCAGCGTGAAGAAGAAGTTCACC
CGGATGACACCCGTACCATGTCCACCGAACAGGGCTACCTGCGTGATGTGT
ACACCAACATCGAACGCAACTGCGTTCGTGAATTCGACATGGAAGAATTCT
ACACCCACCTGGCGTCTATCGGCCTGGAATACGGTGAAACCTTCAGCCGTGT
TCTGCGTGCGAAAAGCGGCCACGACCTGGCGACCGGCACCATTACCATCCC
GGATACCGCGGCGGTTATGCCGCTGGGCTACGAACACCCGTTCATCGTTCAC
CCGGCGACCCTGGATGGCGCGTTCCACCTGGTTTTCTCCGCGATGAGCGGTA
AAGAAGGTTTCCTGGAAGACCCGGCGATCCCGGTTTTCGCGGATGAAATCT
TCGTTAGCTGCGAAATCCCGGCTGAACCGGGTCGTGAACTGAACGTTTGCA
CCCTGCTGAACGAACGTAACCAGCGTAGCCTGCGTTCCTCTATTTGGATCAC
CGATCCGTACAACCCGGAAGGCGGCCCGGTTGTTACCTTCACCGATCTGCAC
TGGAAAGTTCTGGAAAAAGAATTTAAAGAAAAACACCTGGGCATCGATGAA
CGTGTTGCGTACAGCCTGCAGTGGAAAGCGGATGTTGATATGATGGCGTAA
CTCGAG

Phoma sp. PPTase *E. coli* codon optimized sequence including the restriction sites:

CATATGCCGGACCCGGACATGCAGCACGGTGCGGGTCTGACCTGCTGGCTG
CTGGATACCCGTTCTATCTGGCCGGGTAACAAAATCACTGACTCCGCTGCTG
CTCGTGAAGCGCTGCAGCTGATCTCTCCGGAAGAACGTGAAAACGTTTCTC
GTAAATACCACATCGCTGACGCGCGTATGTCTCTGGCTTCTGCTCTGCTGAA

ACGTCTGTTCGTTTACAAAACCCTGAACATCGCATGGAAAGACATCACCTTC
GGTCGTAAACGTGATCCGAAACACGGTAAACCGTGCGCTCTGCTGCCGCCG
GCGAGCCCGCAGCTGTCTCCGCTGCCGGCTCCGATCGAATTCAACATCTCTC
ACCAGGCGGGTCTGGTTGCGCTCGTGGGTTGCAAATCTGAACAGCTGGACG
CCGAACTGGGTGTTGACATCGTTTGCGTTAACGAACGTAACGATTACCGTGT
AATCGATGATGAAGGCCTGGAGGGCTGGGTTGATATCTACTCCGAACTGTT
CTCTCACGAAGAATCCTTCGACATGAAATACAACGCTGACGCATTCCCGCTG
CTGGACGGTACCATGGTGGCTCCGGAAATGCTGCAGGCACTGCGTCACGAT
CGCTGCACCCGTCGTCACCAGAAACTGCGTGTTGTGCTGCCGGGTGGCGAA
GAACGCACCTTCGACTCTGACCTGCTGATCGATGCGAAACTGCGTCGTTTCT
ACACCTTCTGGTGCTTCAAAGAAGCGTACATCAAACTGGACGGCGAAGCGC
TGCTGGCTAAATGGATCCCGCGTCTGGAATTCAAAAACGTTCGCGCACCGC
GTCCGGGCACCGTTGCGCGTTGCTCTACCCACGGTACCTGGGGTGAACGCGT
TGGTGACGCGGAAGTTTGGTTCACCCGTAGCACCGGCGGTAGCGGTGACGG
TCCGGTTGGTGTTGCGAGCCTGTCTCTGAAATCTGGTGAATCCCGTCGTCTG
GACGACACTCGTGTTGAAATCCAGGCTTTCGAAGAAAACTTCATGATCGGC
GTTGCTGCTAAAATGCGTTCTTCTCGTATCGAAGGTGGTGGTATGCTGCCGG
AAGTTCTGACCTCTTTCAAATCTCTGCACCTGGAAGAAGACATCATGGTTAT
CGCGCGTTCTGCGCTCGAG

KS/AT amino acid sequence:

MGSSHHHHHHSSGLVPRGSHMVPYYQPASSCGSNTMAAMDEHQHNEDATIPI
AIIGMSCRFPGNATSPEKLWELCAQGRSAWSSIPKSRFRQEGFYNPNAERVGTS
HVVGGHFLEEDPSLFDASFFNLSAEAAKTMDPQFRLQLESVYEAMESAGITLEH
IAGSDTSVYAGACFRDYHDSLVRDPDLVPRFLLTGNGAAMSSNRVSHFYDLRG
ASMTVDTGCSTTLTALHLACQGLRNRESKTSIVTGANVILNPDMFVTMSSLGLL
GPEGKSHTFDARANGYGRGEGIATVIIKRLDDALRAQDPIRCIIRGTALNQDGRT
ATLTSPSQTAQSDLIRACYRAAALDPNDTAFLAAHGTGTRTGDAVEIAAAADV
FGEKRSPERPLWIGSVKTNIGHSEATSGLASVIQAALALEKGLIPPNINFKEPNEK
LGQVSAAVRVPSNLQKWPSVSGVRRASVNNFGYGGANAHVILESGIPGHTPIA
NGSGRSNGTGNGHNGANGTTNGHNGTNGTTNGHFDATQATNGHYGTDETPD
YAPLDSFVISISAKEEASARSMVTNLADYLRTLQVQDETKHFKSIAHTLGSHRS
MFKWTAAKSITGPEELIAAAEGGQFQASRALERTRLGFVFTGQGAQWFAMGRE
LINTYPVFRQSLDRADRYLKEFGCEWSIIDELSRDAENSNVNDMTLSPPLCTAV
QISLVQLLESWGIVPTAVTGHSSGEIAAAYAAGALDFKSAMAVTYFRGEVGLA

CQDKIVGKGGMIAVGLGPEDAEDRIARVQSGKIVVACINSQSSVTVSGDLSGIV
ELEDLLKAEGVFARRVKVQAAYHSHHMQVIANGYLTSLKDMLKPTKKFGKIIY
SSPTTGRRETNAKLMASAQHWVNNMLSPVRFAESFQNMCFSNRNSSQSEEIFQ
DVDIVLEVGPHGMLQGPIQQMMSLPIFERARLPYISCLLRGQSAVHTMQTVAA
GLMGWGYRVDMVAVNFPQGTYGVKILHDLPSYPWNHDNSHWWEPRLNKAH
RQRVHPPH

KS/AT SUMO tagged protein sequence:

MKHHHHHHPMSDYDIPTTENLYFQGAMGNDHINLKVAGQDGSVVQFKIKRHT
PLSKLMKAYCERQGLSMRQIRFRFDGQPINETDTPAQLEMEDEDTIDVFQQQTG
MVPYYQPASSCGSNTMAAMDEHQHNEDATIPIAIIGMSCRFPGNATSPEKLWEL
CAQGRSAWSSIPKSRFRQEGFYNPNAERVGTSHVVGGHFLEEDPSLFDASFFNL
SAEAAKTMDPQFRLQLESVYEAMESAGITLEHIAGSDTSVYAGACFRDYHDSL
VRDPDLVPRFLLTGNGAAMSSNRVSHFYDLRGASMTVDTGCSTTLTALHLACQ
GLRNRESKTSIVTGANVILNPDMFVTMSSLGLLGPEGKSHTFDARANGYGRGE
GIATVIIKRLDDALRAQDPIRCIIRGTALNQDGRTATLTSPSQTAQSDLIRACYRA
AALDPNDTAFLAAHGTGTRTGDAVEIAAAADVFGEKRSPERPLWIGSVKTNIG
HSEATSGLASVIQAALALEKGLIPPNINFKEPNEKLGQVSAAVRVPSNLQKWPS
VSGVRRASVNNFGYGGANAHVILESGIPGHTPIANGSGRSNGTGNGHNGANGT
TNGHNGTNGTTNGHFDATQATNGHYGTDETPDYAPLDSFVISISAKEEASARS
MVTNLADYLRTLQVQDETKHFKSIAHTLGSHRSMFKWTAAKSITGPEELIAAA
EGGQFQASRALERTRLGFVFTGQGAQWFAMGRELINTYPVFRQSLDRADRYLK
EFGCEWSIIDELSRDAENSNVNDMTLSPPLCTAVQISLVQLLESWGIVPTAVTGH
SSGEIAAAYAAGALDFKSAMAVTYFRGEVGLACQDKIVGKGGMIAVGLGPED
AEDRIARVQSGKIVVACINSQSSVTVSGDLSGIVELEDLLKAEGVFARRVKVQA
AYHSHHMQVIANGYLTSLKDMLKPTKKFGKIIYSSPTTGRRETNAKLMASAQH
WVNNMLSPVRFAESFQNMCFSNRNSSQSEEIFQDVDIVLEVGPHGMLQGPIQQ
MMSLPIFERARLPYISCLLRGQSAVHTMQTVA

SQTKS ACP amino acid sequence:

MGSSHHHHHHSSGLVPRGSHMAQDKQLAAGQELSMATSLVEAIDVVGRAITA
KLATMFLIAAESIIASKSLSEYGVDSLVAVELRNWLAAQLSSDVSVFDVTQSQS
LTALATTVATKSLEHHHHHH

BfPKS ACP amino acid sequence:

MGSSHHHHHHSSGLVPRGSHMSEGDRANAFRDSLALARTIEEASELVCAALVT
KLAARSGISPENVDPSKTVSEYGVDSLVAVELRNWITHEMDSTVPILELLANNP
MNSLSVKIASRSKLVHLDAEKE

Sfp amino acid sequence:

MGSSHHHHHHSSGLVPRGSHMKIYGIYMDRPLSQEENERFMSFISPEKQEKCRR
FYHKEDAHRTLLGDVLVRSVISRQYQLDKADIRFSAQEYGKPCIPDLPDAHFNIS
HSGRWVICAFDSHPIGIDIEKMKPISLEIAKRFFSKTEYSDLLAKNKDEQTDYFY
HLWSMKESFIKQEGKGLSLPLDSFSVRLHQDGQVSIELPDSHTPCYIKTYEVDPG
YKMAVCAAHPDFPEDITMLSYEALL

MtaA amino acid sequence:

MGSSHHHHHHSSGLLPRGSHMPTSSPALPLLKLPPNEVHVWIVEPERITEPGLLE
SYRALLDPGERDKQHRFYFERHRLQYLVSHALVRLTLSRYAPVAPEAWSFSAN
QYGRPEIRGEEKPWLRFNLSHTDGMALCAVARDVDVGADVEDTERRGETVEIA
DSFFAPAEVASLPALPVSDQRERFFDYWTLKEAYIKARGMGLSLPLDQFAFQVS
QGLSTRISFDPRLVDEPSQWQYVLFLPLTSGIPSRWLCADSRTPTSWCVLQRTAS
PP

***Phoma* ACPS amino acid sequence:**

MGSSHHHHHHSSGLVPRGSHMPDPDMQHGAGLTCWLLDTRSIWPGNKITDSA
AAREALQLISPEERENVSRKYHIADARMSLASALLKRLFVYKTLNIAWKDITFG
RKRDPKHGKPCALLPPASPQLSPLPAPIEFNISHQAGLVALVGCKSEQLDAELGV
DIVCVNERNDYRVIDDEGLEGWVDIYSELFSHEESFDMKYNADAFPLLDGTMV
APEMLQALRHDRCTRRHQKLRVVLPGGEERTFDSDLLIDAKLRRFYTFWCFKE
AYIKLDGEALLAKWIPRLEFKNVRAPRPGTVARCSTHGTWGERVGDAEVWFT
RSTGGSGDGPVGVASLSLKSGESRRLDDTRVEIQAFEENFMIGVAAKMRSSRIE
GGGMLPEVLTSFKSLHLEEDIMVIARSALEHHHHHH

DPCK (*E. coli*) amino acid sequence:

MGSSHHHHHHSSGLVPRGSHMMRYIVALTGGIGSGKSTVANAFADLGINVIDA
DIIARQVVEPGAPALHAIADHFGANMIAADGTLQRRALRERIFANPEEKNWLNA
LLHPLIQQETQHQIQQATSPYVLWVVPLLVENSLYKKANRVLVVDVSPETQLK
RTMQRDDVTREHVEQILAAQATREARLAVADDVIDNNGAPDAIASDVARLHA
HYLQLASQFVSQEKP

BfPKS DH amino acid sequence:

MGSSHHHHHHSSGLVPRGSHMSKTFRTRQFPRTDLLGALDRSSNPWEPRWHN
HIRLSEKPWVLDHKIQSNTVYPAAGFIAMAIEAVYQQRKDNSSKPISGFKLKDV
NIGSALLIPEEEETIETLVTLKRHFDHARSPDTYWNEFHVYSVTAANTWTEHCR
GLVTVQREEEVHPDDTRTMSTEQGYLRDVYTNIERNCVREFDMEEFYTHLASI
GLEYGETFSRVLRAKSGHDLATGTITIPDTAAVMPLGYEHPFIVHPATLDGAFH
LVFSAMSGKEGFLEDPAIPVFADEIFVSCEIPAEPGRELNVCTLLNERNQRSLRSS
IWITDPYNPEGGPVVTFTDLHWKVLEKEFKEKHLGIDERVAYSLQWKADVDM
MA

SQTKS DH amino acid sequence:

MGSSHHHHHHSSGLVPRGSHMNKAHRQRVHPPHDLLGSLIVGRDLREPTWRH
FIRVQDIPWIRDHVVQSALVYPGAGFICMAMEAMVQLHELRDSQSRKVAGYRL
AEVDILRAMLIPDTSEGLEAHISLRPCSTKLLLTNEWYDFCVSSVGDDDKFVDH
CRGRITIEFDTSGSADTPRTSLRERSRSTGLMRSVDPSNLYSFLRAQGIYHGPIFQ
NLKTISSRKDHSESSFVVANTASVMPNGFQSPHVIHPTTLDSIFQGAYTALPGAG
LDQNTAMIPRSIQELYLSSALTSDVGQCLVSDTSLIRYDGQSFTVNVDVSSKADS
EHTPVLEIKGLRNQSVGQMA

Lebenslauf

02/2017 – 02/2021	Promotionsstudent an der Gottfried Wilhelm Leibniz Universität Hannover
10/2014 – 01/2017	Masterstudium Wirk- und Naturstoffchemie an der Gottfried Wilhelm Leibniz Universität Hannover
10/2010 – 09/2014	Bachelorstudium Biochemie an der Gottfried Wilhelm Leibniz Universität Hannover
10/2009 – 07/2010	Wehrdienst
2009	Abitur an der Käthe-Kollwitz Schule Hannover

Veröffentlichungen, Publikationen

1. Mohammadi-Ostad-Kalayeh, S.; Hrupins, V.; Helmsen, S.; Ahlbrecht, C.; Stahl, F.; Scheper, T.; Preller, M.; Surup, F.; Stadler, M.; Kirschning, A.*; et al.* Development of a microarray-based assay for efficient testing of new HSP70/DnaK inhibitors. *Bioorganic & medicinal chemistry* **2017**, *25*, 6345–6352.

2. Helmsen, S.; Ahlbrecht, C.; Hrupins, V.; Stahl, F.; Zeilinger, C. Entwicklung komplexer Screening verfahren am Beispiel von Hsp90. *Biospektrum* **2017**, *23*, 540–543.