

# **Computational and *In Vitro* Study of Isolated Domains from Fungal Polyketide Synthases**

Von der Naturwissenschaftlichen Fakultät  
der Gottfried Wilhelm Leibniz Universität Hannover

zur Erlangung des Grades

**Doktor der Naturwissenschaften (Dr. rer. nat.)**

genehmigte Dissertation

von

**Oliver Piech, M.Sc.**

**2020**

Referent: Prof. Dr. Russell Cox  
Korreferent: Prof Dr. Plettenburg  
Tag der Promotion: 29.05.2020

***Vi veri universum vivus vici***

## Abstract

Keywords: Polyketides, Enzyme engineering, Squalestatin tetraketide Synthase, Tenellin

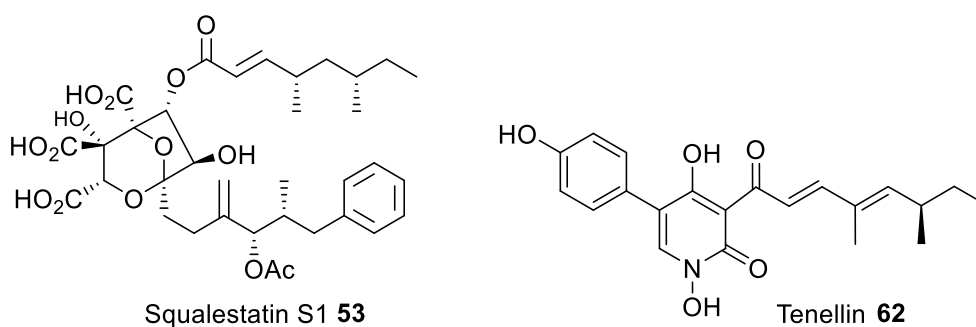
Diverse approaches have been explored to generate new polyketides by engineering polyketide synthases (PKS). Although it has been proven possible to produce new compounds by designed PKS, engineering strategies failed to make polyketides available via widely applicable rules and protocols.

The aim of this work was the first rational engineering of an iterative highly-reducing polyketide synthase (HR-PKS). This approach was performed on the Squalestatin Tetraketide Synthase (SQTKS), which catalyses the biosynthesis of the tetraketide side chain of squalestatin-S1 **53**, which is a potent squalene synthase inhibitor and can be potentially used to treat serum cholesterol related diseases. Second, tenellin **62** was investigated, which is the product of the iterative Type I polyketide synthase non ribosomal peptide synthetase (PKS-NRPS) TENS.

Using a combination of different *in silico* methods, structural models of the enoyl reductase (ER) domain of SQTKS were obtained and validated. With the generated protein models different rational engineering experiments *in silico* were performed, in which amino acids for the mutagenesis approach *in vitro* were identified.

The subsequent *in vitro* experiments revealed that it was possible to rationally engineer the ER domain of SQTKS. In addition, the different integrated mutations showed different effects on the intrinsic programming of the ER domain. Further, the chemical selectivity and kinetic parameters of the tested di-, tri-, tetra- and heptaketide substrate were influenced in a specific way through the different mutated ER domains.

In addition, the structural-biological foundations and analysis for the domain swaps between Pretenellin A Synthetase (TENS), Predesmethylbassianin A Synthetase (DMBS) and Premilitarinone C Synthetase (MILS) were investigated and validated. Through different *in silico* structural analyses it was possible to consider the effects of swaps on protein structure and to understand the effect of the swaps at the structural level. Additionally, the *in silico* analysis helped to clarify the influence of extrinsic and intrinsic programming factors.





## Kurzzusammenfassung

Schlagwörter: Polyketide, Enzym engineering, Squalestatin tetraketide Synthase, Tenellin

Verschiedene Ansätze wurden erforscht, um neue Polyketide durch engeeniering von Polyketide Synthasen (PKS) zu erzeugen. Obwohl es sich als möglich erwiesen hat neue Verbindungen durch künstliche PKS herzustellen, erwies es sich trotzdem nicht als möglich über allgemein anwendbare Regeln und Protokolle neue Polyketide herzustellen.

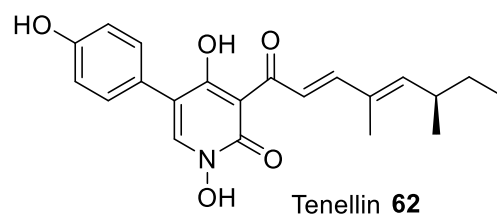
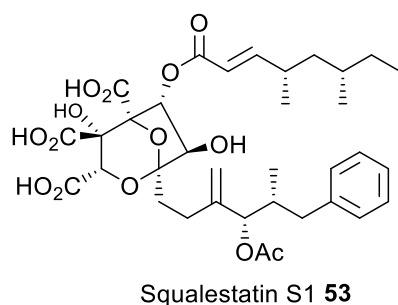
Ziel dieser Arbeit war das erste rationale Engineering einer iterativen, hochreduzierenden Polyketid Synthase (HR-PKS). Dieser Ansatz wurde an der Squalestatin-Tetraketid-Synthase (SQTKS) durchgeführt, die die Biosynthese der Tetraketid-Seitenkette von Squalestatin-S1 **53** katalysiert, einem wirksamen Inhibitor der Squalen-Synthase, der möglicherweise zur Behandlung von Erkrankungen im Zusammenhang mit Serumcholesterin eingesetzt werden kann. Desweiteren wurde Tenellin **62** untersucht, welches das Produkt der iterativen nicht-ribosomalen Peptid Synthetase (PKS-NRPS) TENS des Typ I ist.

Unter Verwendung einer Kombination verschiedener *in silico* Methoden wurde ein Strukturmodell der Enoylreduktase (ER) Domäne von SQTKS erzeugt und validiert. Mit den generierten Proteinmodellen wurden verschiedene rationale Engineering Experimente *in silico* durchgeführt, wobei unter anderem Aminosäuren für den *in vitro* Mutageneseansatz identifiziert wurden.

Die anschließenden *in-vitro* Experimente zeigten, dass es möglich war, die ER-Domäne von SQTKS rational zu engeenieren. Darüber hinaus zeigten die verschiedenen generierten Mutationen unterschiedliche Auswirkungen auf die intrinsische Programmierung der ER-Domäne. Darüber hinaus wurden die chemische Selektivität und die kinetischen Parameter des getesteten Di-, Tri-, Tetra- und Heptaketisubstrate auf spezifische Weise durch die verschiedenen mutierten ER-Domänen beeinflusst.

Darüber hinaus wurden die strukturbioologischen Grundlagen und Analysen für die Domain-Swaps zwischen Pretenellin-A-Synthetase (TENS), Predesmethylbassianin-A-Synthetase (DMBS) und Premilitarinon-C-Synthetase (MILS) untersucht und validiert

Durch verschiedene *in silico* Strukturanalysen konnten die Auswirkungen von Swaps auf die Proteinstruktur berücksichtigt und die Auswirkungen der Swaps auf die Strukturebene verstanden werden. Darüber hinaus hat die *in silico*-Analyse dazu beigetragen, den Einfluss extrinsischer und intrinsischer Faktoren der Programmierung zu klären.



## Acknowledgements

I would like to thank Prof. Russell Cox for providing me with such an interesting project to work on, and for his thorough and dedicated supervision throughout the course of my PhD. His support and encouragement was highly appreciated.

Especially, I would like to thank Prof Plettenburg taking the time to be the co-referee and chair of the examination board of my PhD thesis. No, less I thank Prof Preller for being examiner.

I would like to thank all past and current Cox group members, in particular, Verena, Karen, Haili, Francesco, Eric, Katja, Lukas, Carsten and my master student Yingwen Wu. Beate Bonsch and Christoph Bartels, who give me an affectionate reception into the group. At last, my office member Vjaceslavs Hrupins, who made the time enjoyable. In addition, once more Carsten and Lukas for all the time, we spent playing billard or were drinking beer and Eric, which whom I spend a great time in China on a holiday.

Furthermore, from the other groups of the OCI/BMWZ Maik Siebke, for all the Döner we eat together. Wiebke Hutwelker and Simon Blazy whom became good friends over the time and made the time very enjoyable. Lukas Schaaf whom also became a good friend and for all the time we drank whiskey together.

I want to thank Michael Ringel, who helped me a great deal with the SPR measurements. Also Tjorven Ostermeier and Luca Codutti, for the help with the ITC measurements. I also want to thank the members of the OCI and BMWZ. Especially the members of the NMR and MS department and the members of the media kitchen. Katja Körner and Anika Shaperjahn, who also did a great deal of work and without this contribution my research, would be not this fast.

I want to thank my friends outside the OCI and BMWZ in Hannover, Nina Bornemann, Philipp Müller and Sandra Dienemann. Sabrina Czelusteck, for the last beautiful years.

Finally, my family for their help and support during my studies and that you are there if I need you.

## Table of contents

<b>Abstract</b> .....	iv
<b>Kurzzusammenfassung</b> .....	v
<b>Acknowledgements</b> .....	vi
<b>List of Abbreviations and Units</b> .....	xi
<b>1 Introduction</b> .....	1
1.1 Natural Products.....	1
1.2 Fatty Acid Biosynthesis .....	2
1.3 Enzymes of Fatty Acid Biosynthesis.....	3
1.4 Architecture of Fatty Acid Synthases.....	7
1.5 From Fatty Acid Synthases (FAS) to Polyketide Synthases (PKS).....	8
1.6 Biosynthesis of Polyketides.....	9
1.6.1 Mechanism of the C-MeT domain in PKS.....	11
1.7 Polyketide Synthases .....	11
1.8 Modular Type I Polyketide Synthases.....	12
1.9 Iterative Type I Polyketide Synthases .....	14
1.10 Highly Reducing Iterative Polyketide Synthase (HR-PKS).....	15
1.11 Stereochemical course in HR-PKS.....	17
1.11.1 Excursus into KR domains .....	19
1.12 Biosynthesis of Squalestatin Tetraketide Synthase .....	21
1.13 Biosynthesis of Tenellin .....	23
1.14 Engineering of Polyketide Synthases.....	25
1.15 Engineering of C-MeT and ER domain from fungal HR-PKS .....	27
1.16 Rational Domain Swaps between the HR-PKS TENS and DMBS.....	28
1.17 Programming in Iterative Type I HR-PKS.....	29
1.18 Overall Project Aims .....	34
<b>2 Modelling Studies of the SQTGS ER Domain</b> .....	35
2.1 Introduction .....	35
2.1.1 Swiss-Model .....	37
2.1.2 Molecular Docking (AutoDock Vina).....	38
2.1.3 YASARA .....	41
2.2 Previous Studies .....	43
2.3 Aims of the Project .....	44
2.4 <i>In Silico</i> Studies of the SQTGS ER Domain .....	45

2.5	Integration of the cofactor into the modelled ER domain of SQTKS .....	47
2.6	Development of a Substrate and Cofactor Docked Model for the ER of SQTKS .....	49
<b>3</b>	<b><i>In Silico</i> Mutagenesis Studies of the SQTCS ER Domain.....</b>	<b>56</b>
3.1	Introduction .....	56
3.2	Aim of the Project.....	56
3.3	Identification of Residues Potentially Involved in Substrate Selectivity in the ER domain of SQTCS .....	56
3.4	<i>In Silico</i> Studies Concerning the Conversion of the Squalestatin Tetraketide-Pantetheine Stereoisomers 93a-d (ER domain of SQTCS) .....	59
3.4.1	Investigation of tetraketide stereoisomers.....	59
3.4.2	<i>In silico</i> mutation experiments.....	63
3.5	<i>In Silico</i> Studies Concerning Conversion of Longer Substrates (ER domain SQTCS)....	65
3.6	Conclusion.....	70
<b>4</b>	<b><i>In Vitro</i> Engineering of the SQTCS ER domain .....</b>	<b>72</b>
4.1	Introduction .....	72
4.2	Aim of the Project.....	72
4.3	Substrate Synthesis.....	73
4.4	Site Directed Mutagenesis .....	75
4.5	<i>In Vitro</i> Enzymatic Investigation of the ER domain .....	80
4.5.1	Stereoselectivity of the Mutated ER Domains .....	89
4.6	Discussion and Outlook.....	91
<b>5</b>	<b>Modelling Studies of the TENS C-MeT and KR Domains.....</b>	<b>95</b>
5.1	Introduction .....	95
5.2	Aims of the Project .....	96
5.3	<i>In Silico</i> Studies of the TENS C-MeT Domain .....	97
5.4	Integration of the cofactor into the modelled C-MeT domain of TENS.....	100
5.5	Development of Ligand and Cofactor Docked Models for the C-MeT of TENS.....	102
5.5.1	Docked Model of the Ketone Substrate 128 and Cofactor SAM 35 .....	103
5.5.2	Docked Model of the Enol Intermediate 129 and Cofactor SAM 35 .....	104
5.5.3	Docked Model of the Products 130a, 130b and Cofactor SAH 39.....	105
5.5.4	Overall assessment of C-MeT modelling and docking.....	106
5.6	<i>In Silico</i> Studies of the TENS KR Domain .....	107
5.7	Integration of the cofactor into the modelled KR domain of TENS.....	111
5.8	Development of a Substrate and Cofactor Docked Model for the KR of TENS.....	112

5.9	Conclusion.....	115
<b>6</b>	<b>Development of a chimeric C-MeT-ΨKR-KR sub-structure of TENS to Understand the Molecular Basis of Methylation and Chain-Length Programming <i>in-silico</i>.....</b>	<b>117</b>
6.1	Introduction .....	117
6.2	Aims of the Project .....	117
6.3	Development of a Chimeric Multidomain Model of TENS.....	117
6.4	Overview of the Molecular-Biological <i>in vivo</i> work.....	122
6.5	Structural Analysis of the Chimeric Multidomain Model of TENS.....	125
6.6	<i>In silico</i> Studies with the Single KR Model of TENS.....	127
6.7	<i>In silico</i> swap experiments with the KR domain.....	130
6.8	Discussion and Outlook.....	132
<b>7</b>	<b>Overall Conclusion.....</b>	<b>137</b>
<b>8</b>	<b>Experimental Section.....</b>	<b>139</b>
8.1	Equipment.....	139
8.2	Buffer, antibiotics, media and solutions .....	141
8.3	Synthesis of Pantetheine Substrates .....	143
8.3.1	Preparation of precursor substrates .....	143
8.3.2	Preparation of panthetine dimethyl ketal substrates .....	148
8.3.4	Synthesis of panthetine substrates first method <sup>137</sup> .....	155
8.3.5	Synthesis of panthetine substrates second method .....	158
8.4	SQTKS ER Domain .....	162
8.5	Colony Polymerase-chain-reaction.....	163
8.6	Bradford assay.....	164
8.7	SQTKS-ER Domain LCMS Assay.....	164
8.8	SQTKS-ER-Domain Enzyme Assay .....	164
8.9	SQTKS-ER-Domain Stereochemistry Assay.....	165
8.10	Mutagenesis.....	165
8.11	Statistical evaluations .....	167
8.12	Bioinformatics .....	167
<b>9</b>	<b>References.....</b>	<b>171</b>

<b>10 Appendix</b> .....	179
10.1 Protein pdb Data File List (attached CD).....	179
10.2 Multiple Alignment of $\beta$ -Processing Domains of TENS, DMBS, mFAS pig and mFAS rat	
181	

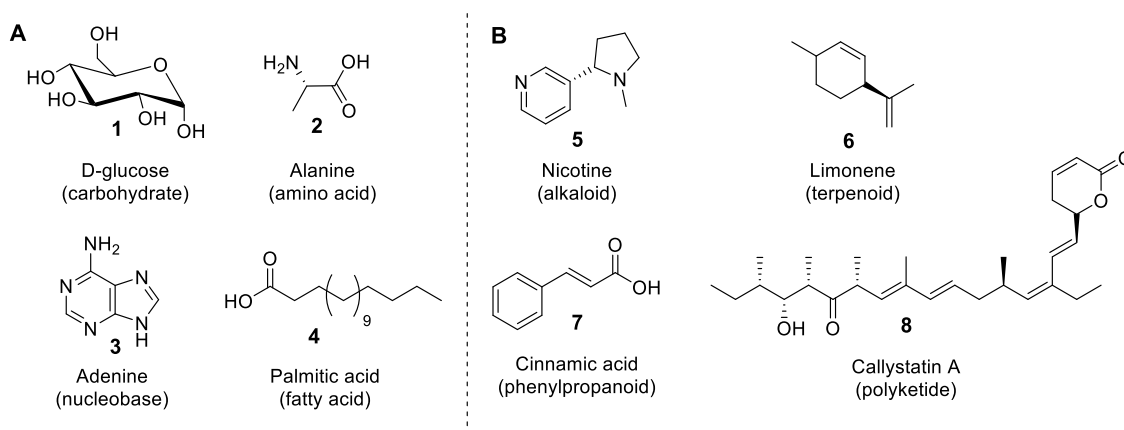
## List of Abbreviations and Units

ACP	Acyl carrier protein	TE	Thiolesterase
AT	Acyl-transferase	TENS	Pretenellin A Synthetase
CDI	<i>N, N'</i> -carbonyl diimidazole	TFA	Trifluoro acetic acid
DCM	Dichloromethane	THF	Tetrahydrofuran
C-MeT	C-Methyl transferase	UV	Ultraviolet
CoA	Coenzyme A	$V_{Max}$	Maximum rate of a reaction
Da	Dalton		
DH	Dehydratase		
DMAP	4-Dimethylaminopyridine		
DMBS	Predesmethylbassianin A Synthetase		
EDCI	1-Ethyl-3-(3-dimethylaminopropyl) carbodiimide		
ER	Enoyl reductase		
FAS	Fatty acid synthase		
His-tag	Histidine tag		
HPLC	High performance liquid chromatography		
HRMS	High resolution mass spectrometry		
HR-PKS	Highly reducing polyketide synthase		
IPTG	Isopropyl- $\beta$ -D-1-thiogalactopyranoside		
$k_{cat}$	Turnover number		
$K_M$	Michaelis-Menten constant		
KR	$\beta$ -Ketoacyl reductase		
KS	Keto synthase		
LC	Liquid chromatography		
LCMS	Liquid chromatography mass spectrometry		
MS	Mass spectrometry		
MeT	Methyl transferase		
MILS	Premilitarinone C Synthetase		
NADPH	Nicotinamide adenine dinucleotide phosphate		
NMR	Nuclear magnetic resonance		
NR-PKS	Non reducing Polyketide Synthase		
PANT	Pantetheine		
PCC	Pyridinium chlorochromate		
PCR	Polymerase chain reaction		
PKS	Polyketide synthase		
PR-PKS	Partly reducing polyketide synthase		
ppm	Parts per million		
RMSD	Root mean square derivation		
SUMO Tag	Small ubiquitin-related modifier		
SQTKS	Squalestatintetraketide synthase		

# 1 Introduction

## 1.1 Natural Products

In general, a natural product is a chemical compound or substance produced by a living organism.<sup>1</sup> In the organism the production of these natural products is accomplished by enzymes. Hence, the repertoire of secondary metabolites is evolved, like the proteome in the organisms is evolved to its specific physical and biological environmental requirements.<sup>2</sup> Natural products can be classified according to their biological function, biosynthetic pathway or source.



**Figure 1:** Examples of some natural products: **A**, Examples of natural products of the primary metabolite classes; **B**, Examples of natural products of different secondary metabolite classes.

The simplest classification of natural products is based on the biological source from which the naturally derived metabolites were isolated. For example prokaryotes like bacteria or archaea or eukaryotes like plants, animals and fungi.<sup>3</sup>

In addition, natural products are classified in two major classes, the primary and secondary metabolites (Figure 1). Primary metabolites, like **1-4**, have an intrinsic function that is *essential* to the survival of the organism that produces them. For example palmitic acid **4** (Figure 1), which is the most common saturated fatty acid in animals, plants and microorganism.<sup>4</sup> In contrast, secondary metabolites, like **5-8**, have a function that mainly affects other organisms. These compounds are not *essential* for survival but may give an organism an evolutionary advantage. Secondary metabolites are often useful to humans because of these effects. For example, callistatin A **8** has potent antitumor properties.<sup>5</sup> Secondary metabolites are further sub-classified according to their biosynthesis. Major

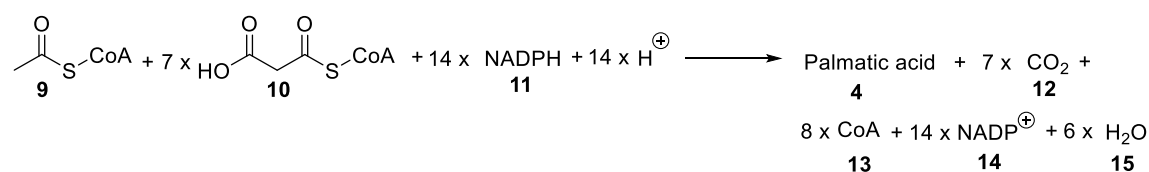


families include polyketides and fatty acids, peptides, terpenes, alkaloids and hybrid metabolites. The focus of this thesis is on polyketide and fatty acid metabolites.

## 1.2 Fatty Acid Biosynthesis

Fatty acids like palmitic acid **4**, are made using highly conserved enzymatic reactions.<sup>6</sup> The enzymes which catalyze fatty acid synthesis are called fatty acid synthases (FAS).<sup>6-8</sup> FAS are multi-enzyme proteins. Hence, it is not a single enzyme but a multifunctional enzymatic system, in which substrates are passed from one functional domain to the next.<sup>6-8</sup> FAS proteins differ in structure and are divided into Type I and II systems (Section 1.4).<sup>6</sup>

The overall reaction sequence of the biosynthesis of fatty acids is the sequential extension of an acyl chain by two carbons in each elongation cycle, by a series of decarboxylative condensation reactions that are summarized in Scheme 1.<sup>8</sup>

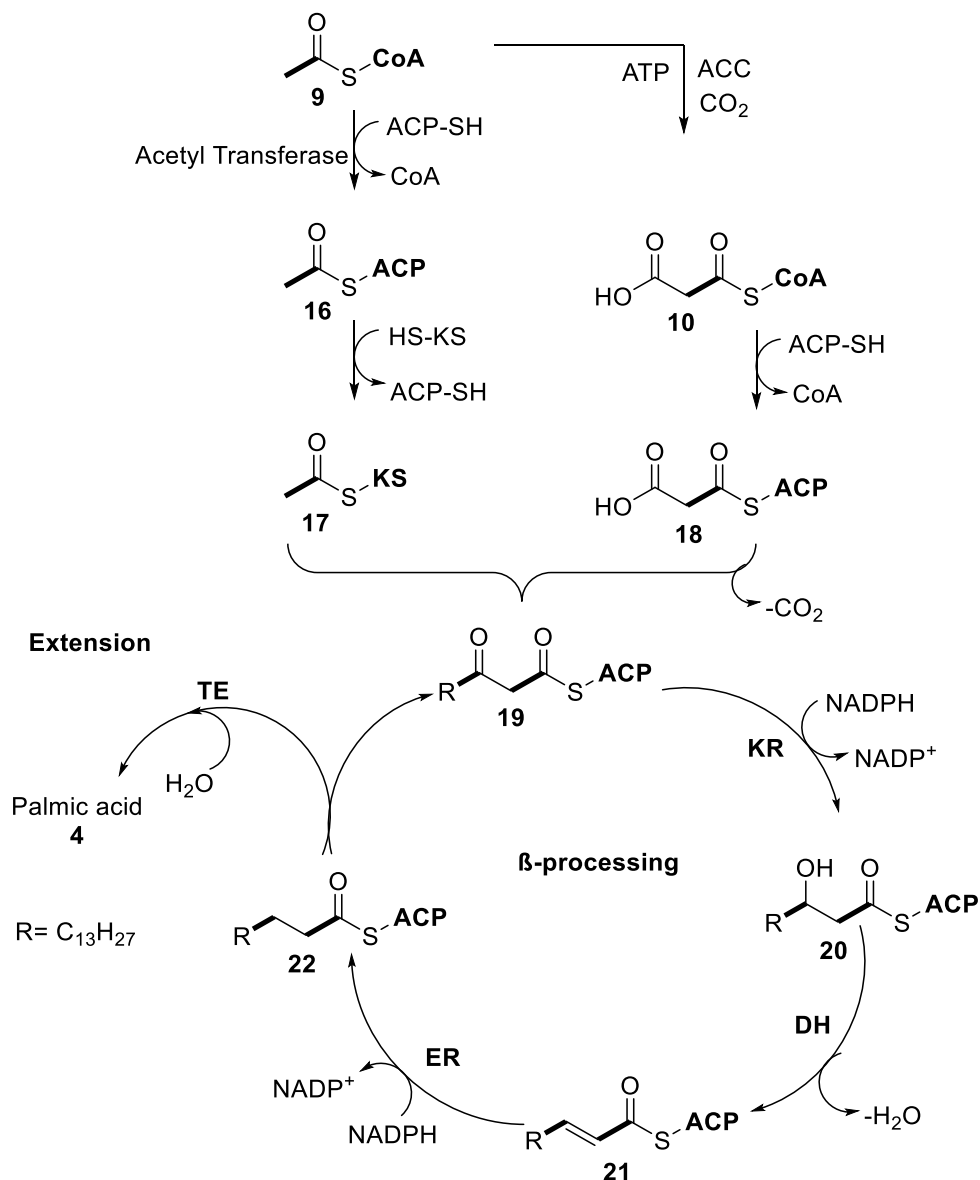


**Scheme 1:** Overall reactions sequence of the biosynthesis of palmitic acid **4**.<sup>8</sup>

The biosynthesis is initiated by the transfer of a starter unit, normally an acetyl moiety **9**, from CoA thiolester to the nucleophilic serine residue of an acyl transferase (AT). The AT then transfers the acyl group to the thiol of an acyl carrier protein (ACP) domain and finally to the active site cysteine residue of the  $\beta$ -ketoacyl synthase (KS, Scheme 2).<sup>8</sup>

Acetyl-CoA **9** is converted to malonyl-CoA **10** by carboxylation catalysed by the enzyme acetyl-CoA carboxylase (ACC).<sup>8,9</sup> The malonyl group is then transferred from its CoA thiolester to the FAS-ACP domain to form malonyl-ACP **18**. The new fatty acid C–C bond is formed by condensation of the acetyl-KS **17** and malonyl-ACP **18** moieties to form acetoacetyl ACP **19**. The driving force of the reaction is an energetically and entropically favorable decarboxylation. From this point forward, in each round of elongation the beta keto group is reduced to the fully saturated carbon chain by the sequential action of a ketoreductase (KR), to form a  $\beta$ -alcohol **20**, dehydratase (DH) to form an  $\alpha\beta$ -unsaturated thiolester **21** and enoyl reductase (ER) to form a fully saturated chain **22**. At this stage, short chains are passed back to the KS and extended again. At the end of chain-building, when the chain has reached its required length, the product is

released by the action of a thiolesterase (TE) upon reaching a carbon chain length of 16 (palmitic acid **4**, Scheme 2).<sup>8</sup>



**Scheme. 2:** Example of the reaction sequence for *de novo* biosynthesis of fatty acids by the animal FAS. After each round of elongation, the beta keto group is reduced to the fully saturated carbon chain by the sequential action of a ketoreductase (KR), dehydratase (DH), and enoyl reductase (ER). The chain is released by the action of a thiolesterase (TE) upon reaching a carbon chain length of 16 (palmitic acid **4**)

### 1.3 Enzymes of Fatty Acid Biosynthesis

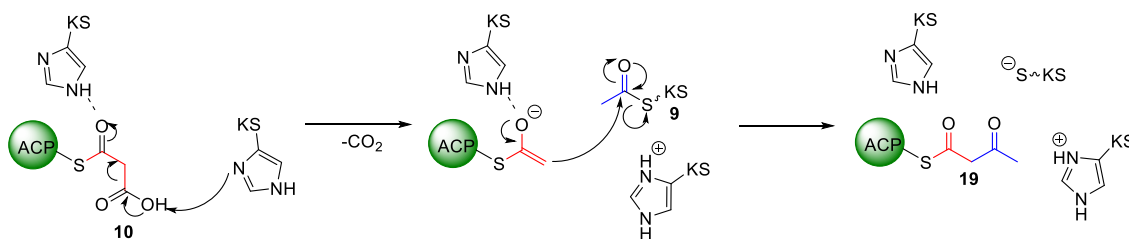
The starting point of the series of condensation reactions leading to the production of palmitic acid **4** is the translocation of one acetyl **9** and malonyl **10** moieties, from CoA thiolester to the phosphopantetheine thiol of the ACP domain.<sup>8</sup> ACP is a universal and highly conserved carrier of acyl intermediates. In yeast and mammals, ACP exists as a

domain within a large multifunctional fatty acid synthase polyprotein (type I FAS), whereas it is a small monomeric protein in bacteria and the plastids of plants (type II FAS).<sup>10</sup>

The next step that follows is the chain elongation. Type I FAS contains only one type of KS domain, whereas two or three KSs with different chain-length selectivity, are found in type II FAS.<sup>6</sup> The overall condensation reaction in KS domains can be further divided into three discrete steps: (1) transfer of saturated acyl moieties from acyl-ACP thiolester form to the active-site cysteine residue, (2) binding and decarboxylation of a malonyl-ACP thiolester moiety to yield a reactive carbanion and (3) formation of a new carbon-carbon bond.<sup>8</sup>

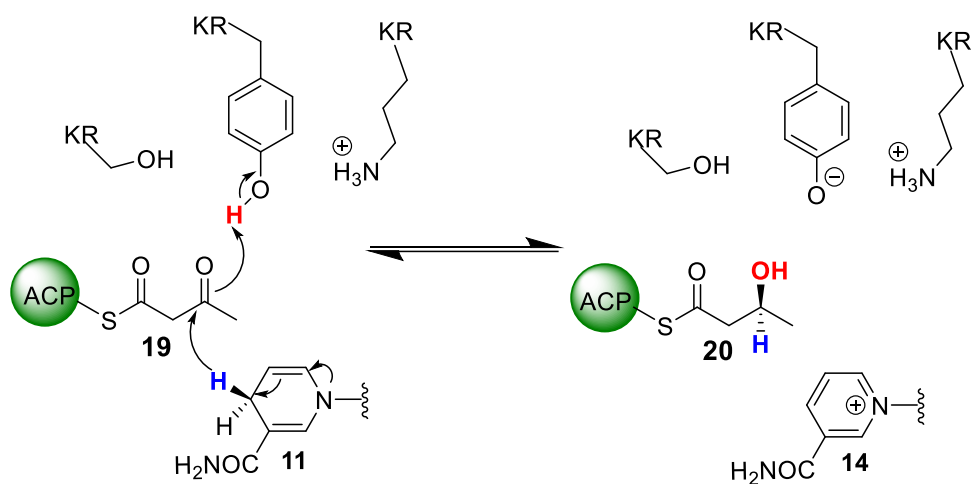
KS domains have been shown to have invariant hydrophobic residues that line their active site pockets to facilitate binding of the acyl chain. A catalytic triad lies at the bottom of the pocket and the acyl chain folds into an extended U-shaped conformation. KS domains usually have a Cys-His-His active site triad (Scheme 3), although other variants with a catalytic triad that consists of Cys-His-Asn are known.<sup>11,12</sup>

In the Cys-His-His active site triad, one histidine is proposed to abstract a proton from the carboxylic acid of the malonyl **10** residue, while the other histidine accepts a hydrogen bond and then makes contact with the thiolester oxo group of the malonyl residue inducing a partial positive charge on that group (Scheme 3). These two processes prompt the decarboxylation reaction, which is followed by condensation.<sup>11,13</sup>



**Scheme 3** Mechanism, with the Cys-His-His triad, of the ketosynthase domain.

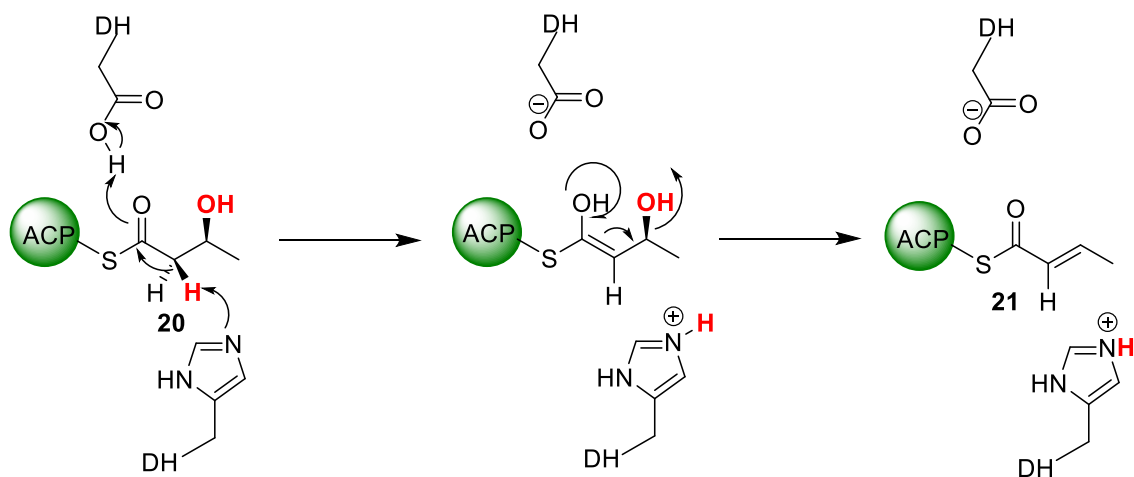
The carbon-carbon bond-formation is followed by  $\beta$ -processing steps during which the beta keto group is converted to the fully saturated carbon. The  $\beta$ -processing starts with the KR domain, which belongs to the short chain dehydrogenase/reductase (SDR) superfamily. KR reduces a  $\beta$ -ketone into an alcohol of *S*- or *R*-stereochemistry using NADPH **11** (Scheme.4).<sup>14</sup>



**Scheme 4:** Mechanism of the reduction of the keto reductase domain.

The next chemical step after the KR is catalysed by the DH domain (Scheme 5). The DH domain catalyzes the elimination of water from the  $\beta$ -alcohol product **19** of the KR domain.<sup>15</sup> This reaction is independent of the presence of an alkyl substituent at the alpha position. The *syn*-elimination process eliminates the alpha hydrogen atom (Scheme 5, red hydrogen atom) and the beta hydroxy group (Scheme 5, red marked) as water.

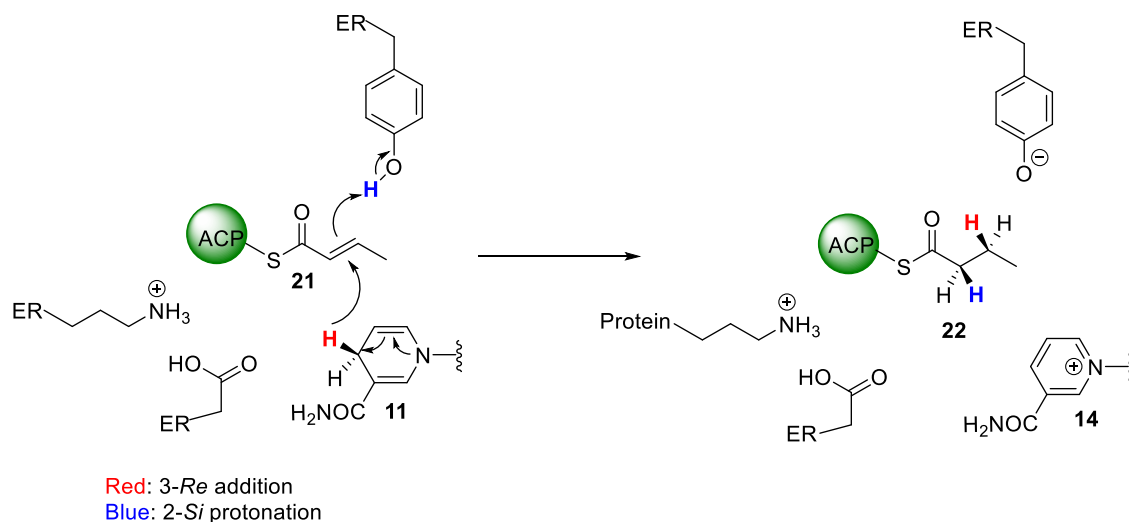
The catalysing process works by a His-Asp dyad mechanism in which a hydroxyl in the  $\beta$ -position is eliminated through an E1cb mechanism.<sup>15–17</sup> At the same time, aspartic acid protonates the  $\beta$ -hydroxyl group and forms water as a good leaving group.



**Scheme 5:** Mechanism of the dehydratase domain.

The ER domain belongs to the medium chain NADPH-dependent dehydrogenase/reductase family (MDR). ER domains reduce  $\alpha\beta$ -unsaturated thioester **21** by an addition of the hydride from NADPH **11** or NADH to the C-3 position and protonation at C-2, which determines the stereochemistry at the  $\beta$ -position (Scheme 6).

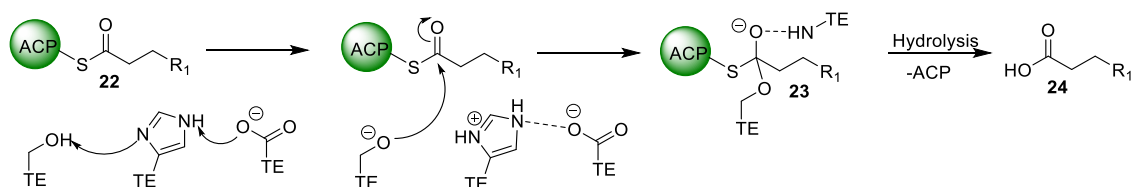
The ER of animal FAS, catalyses attack by the 4'-*pro-R* hydride of NADPH **11** at the 3-*Re* face of **21**, with subsequent protonation at the 2-*Si* face of **21**, giving an overall *syn* of H<sub>2</sub> addition (Scheme 6). The ER of yeast FAS catalyses attack by the 4'-*pro-S* hydride of NADPH **11** at the 3-*Si* face, with protonation occurring at the 2-*Si* face, an overall *anti* addition. In the type II FAS of *E. coli*, the ER catalyses attack by the 4'-*pro-S* hydride of NADH **11** at the 3-*Si* face of the enoyl intermediate with protonation at the 2-*Re* face to give an overall *syn* addition.<sup>18,19</sup>



**Scheme 6:** Mechanism of the enoyl reductase domain.

After the iterative chain elongation cycle, fatty acid biosynthesis is terminated either by offloading the fatty acid from the ACP by an acyl-ACP thioesterase releasing a free fatty acid or by transesterification onto a suitable nucleophile by an acyltransferase. Most often, prokaryotes utilize acyltransferases and eukaryotes utilize thioesterases.<sup>20</sup>

The  $\alpha/\beta$  hydrolase thioesterases have a conserved catalytic triad: a nucleophile-histidine-acid triad, with the nucleophile commonly being a serine residue (Scheme 7). The acid stabilizes the basic histidine, which accepts a proton from the nucleophile. The nucleophilic serine forms a tetrahedral intermediate **23** with the substrate before it is attacked by water resulting in product **22**.<sup>20,21</sup> The active site residues in TEs are on the surface of the protein, rather than buried in an acyl-binding pocket as seen with other FAS domains. The hotdog-fold TEs lack defined non-solvated binding pockets and conserved catalytic residues, and therefore a variety of catalytic residues and mechanisms exist.<sup>20,22</sup>



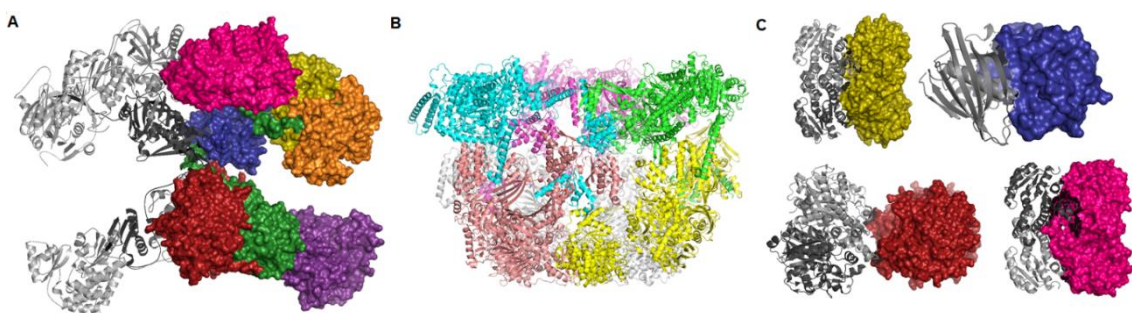
**Scheme 7:** Mechanism of the thioesterase domain.

## 1.4 Architecture of Fatty Acid Synthases

All known FAS produce fatty acids, like palmitic acid **4**, in an iterative process. This means the individual C-C bond forming and  $\beta$ -processing domains are reused in each elongation step until the full length fatty acid is produced.<sup>23</sup> Hence, the fatty acid synthase is a key paradigm for carrier protein-dependent multi enzymes.<sup>24</sup> As mentioned before (section 1.2) FAS proteins can differ in structure and organization and are thus divided into type I and II systems (Figure 2).

Type I FAS are large multifunctional proteins containing all domain activities for fatty acid biosynthesis in one mega synthase protein (Figure 2A).<sup>23/25</sup> All of the enzyme activities are covalently linked to form a large multifunctional protein. They are found in animals and fungi. Even so, the structural arrangement of fungal and animal synthases differ.

In contrast, Type II systems are found in archaea, bacteria and the plastids of plants and are characterized by the use of discrete, monofunctional enzymes, which most likely form a non-covalent complex (Figure 2C).



**Figure 2:** **A**, Homodimer crystal structure of mammalian FAS. PDB entry 2VZ8.<sup>23</sup> In purple the AT domain, In red the KS domain, in blue the DH domain, in violet the ER domain, in yellow the KR domain, in orange the  $\psi$ KR and  $\psi$ C-MeT domain; **B**, Crystal structure of the yeast fatty acid synthase. PDB entry 2UV8;<sup>26</sup> **C**, Crystal structure of KR domain (yellow), ER domain (violet), KS domain (red) and DH domain (blue) from *E. coli* of the Type II FAS system. PDB entry's: 1Q7B, 1MKB, 1DFI and 2VB9).<sup>27-30</sup>

The mammalian FAS (mFAS, Figure 2A) structurally consists of a homodimer of two identical protein subunits, in which the catalytic domains in the N-terminal section, the

keto synthase (KS) and malonyl/acetyltransferase (AT), are separated by a linker region from the C-terminal domains. There are the dehydratase (DH), enoyl reductase (ER), ketoacyl reductase (KR), acyl carrier protein (ACP) and thiolesterase (TE, Figure 2A). The FAS I structure also includes a  $\psi$ C-MeT domain and a  $\psi$ KR domain. The  $\psi$ C-MeT is an inactive (and mostly degraded) methyltransferase domain. The  $\psi$ KR domain is a structural (supporting) domain of the KR. In contrast, the yeast FAS (yFAS) assembles into a barrel-shaped structure (Figure 2B).<sup>24</sup> Catalytic domains of yFAS are embedded in an extensive scaffolding matrix and arranged around two enclosed reaction chambers.<sup>24</sup> Architecturally yFAS and mFAS are unrelated.

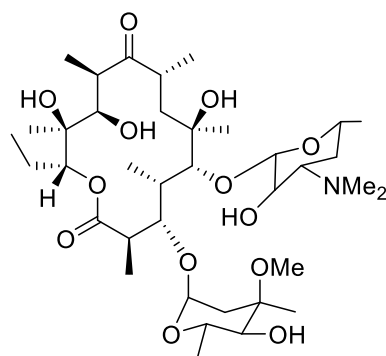
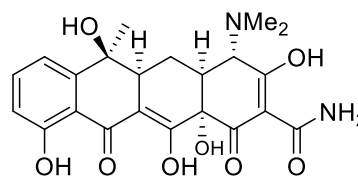
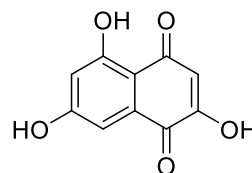
### 1.5 From Fatty Acid Synthases (FAS) to Polyketide Synthases (PKS)

While fatty acid synthases are optimized molecular machines specially for the production of fatty acids, like palmitic acid **4**, many multienzymes with an overlapping set of functional domains and a related synthetic scheme are known which are not involved in primary fatty acid metabolism.<sup>24</sup> In contrast, they are involved in secondary metabolism.

These secondary metabolism machines do not necessarily produce fully reduced products. They can possess modifications at the C- $\alpha$  as well as C- $\beta$  positions, which originate from incomplete reduction cycles caused by the absence of specific modifying domains, by catalytically inactive domains, or by programmed domain skipping. Because of their chemically diverse products, these synthases are named according to their primary condensation product: polyketide synthases (PKS).<sup>24</sup>

The catalytic domains of PKS are closely related to those of mFAS. The enzymes are also independently folding protein domains, which like the FAS, construct polymeric chains and tailor their functionalities.<sup>31</sup> They include KS, DH, ER, KR as well as active C-methyl transferase (C-MeT) domains. Substrate loading is catalyzed by acyltransferases, which are specific for the particular substrate of the PKS and are commonly abbreviated as AT.<sup>24</sup>

PKS are found in bacteria, fungi and plants. PKS produce a highly diverse group of natural products. This includes aliphatic, cyclic, acyclic and aromatic components and macrocyclic lactones. The activities of the polyketides reach from antibacterial and immunosuppressive to anticancer to just name a few.<sup>32</sup> Examples of microbial polyketides include erythromycin **25** from the bacterium *Saccaropolyspora erythraea* and tetracycline **26** from *Streptomyces rimosus*, which are used as antibiotics (Figure 3).<sup>33</sup>

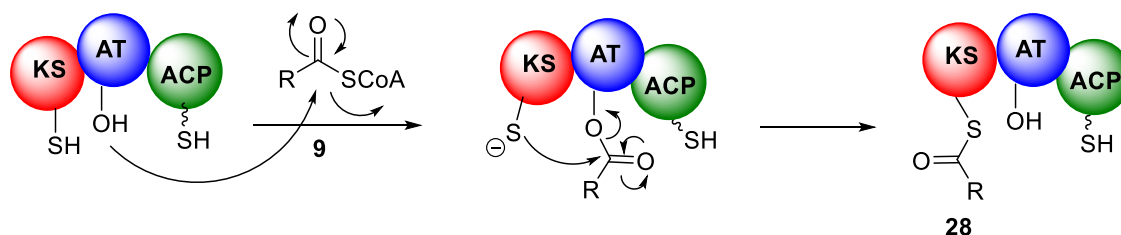
**Type I - modular**Erythromycin A **25****Type II - iterative**Tetracycline **26****Type III - iterative****27**

Flavolin

**Figure 3:** Some examples for natural products of the PKS produced by fungi and bacteria.**1.6 Biosynthesis of Polyketides**

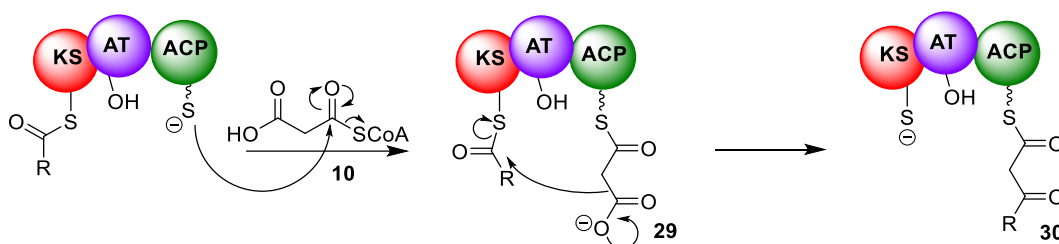
Polyketide synthases (PKS), like the FAS, are responsible for forming a backbone of carbon atoms out of different carbon units. The biosynthesis usually starts with a simple starter unit such as acetyl-CoA **9**. There is also the possibility to use other more unusual starter units for example propionyl-CoA.<sup>34</sup> Furthermore, a PKS consists of several catalytic domains. However, for each PKS the acyl transferase (AT), a ketosynthase (KS) and the acyl carrier protein (ACP) are obligatory. Other domains like keto reductase (KR), dehydratase (DH), enoyl reductases (ER), methylation domains, oxidase domains and more are all facultative and are not found or active in every PKS.

The first step, the AT domain “selects” the starting unit, which is normally an acyl CoA **9** and transfers this through a nucleophilic attack of the sulfur of the KS to the carbon on the acetyl-CoA to the KS **28** (Scheme 8).

**Scheme 8:** Loading of the KS domain with the starter unit.



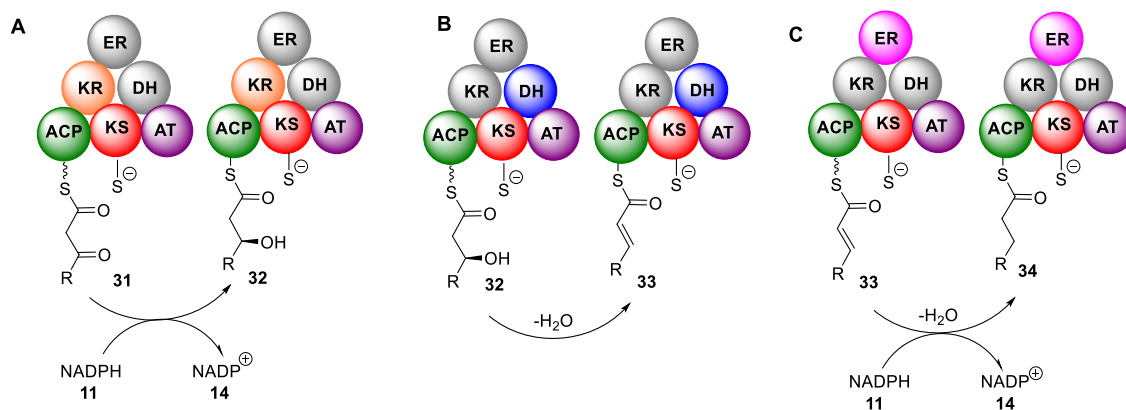
A typical extender unit for a PKS is malonyl CoA **10**, but many bacterial systems can use methylmalonyl CoA.<sup>35</sup> The ACP domain carries the polyketide chain, which is transferred through auto acylation to the KS domain. The malonyl-building block will bind to the ACP domain and will be decarboxylated. The generated enolate reacts similar to a Claisen condensation with the molecule on the KS. Now on the ACP domain is the elongated polyketide chain **30** with four C-atoms (Scheme. 9).



**Scheme 9:** Elongation of the polyketide chain without any modifications.

After the chain extension, further modifications of the developing polyketide chains are possible (Scheme. 10A-C). A different catalytic domain is responsible for each so-called  $\beta$ -processing step. The modifying domains are the ketoreductase (KR), resulting in  $\beta$ -alcohol **32** (Scheme. 10A), dehydratase (DH) resulting in alkene **33** (Scheme 10B) and the enoyl reductase (ER) resulting in alkane **34** (Scheme 10C). Each domain can be either active or inactive in each round of chain extension, so that a following modification might or might not take place. Fungal PKS can also include a functional C-MeT domain.

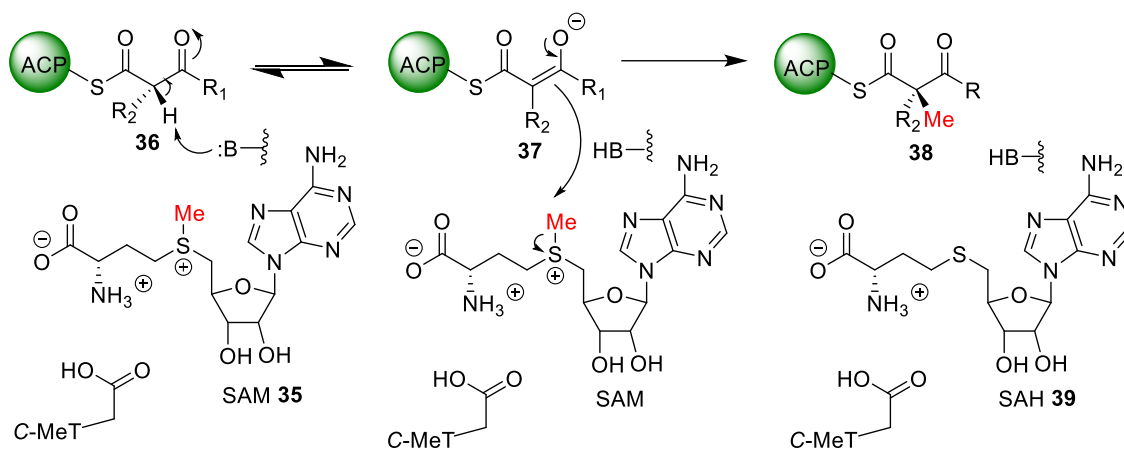
The end of the polyketide synthase process is generally determined through the thiolesterase (TE). They release the polyketide chain from the PKS complex. Often, a macrolactonisation can happen. In addition to that, after release, other post modifications through tailoring enzyme can take place.



**Scheme 10:** Elongation of the polyketide chain with different  $\beta$ -processing steps: **A**, Only the KR domain is active in the  $\beta$ -processing; **B**, The KR and DH domain are active in the  $\beta$ -processing; **C**, the KR, DH and ER domain are all active in the  $\beta$ -processing.

### 1.6.1 Mechanism of the C-MeT domain in PKS

In contrast to FAS, PKS can possess an active C-methyl transferase domain (C-MeT). The C-MeT can transfer a methyl group to the  $\alpha$ -position of the polyketide **36** (Scheme.11). The cofactor for this reaction is S-adenosyl methionine (SAM) **35**. The S-methyl group of **35** is a potent electrophile.<sup>15</sup> The hypothesized mechanism is through an attack by a nucleophilic carbon of **37** on the electrophilic SAM **35** methyl group resulting in product **38** (Scheme 11). Fingerprint residues (D/E) xGxGxG and neighbouring D/E) can be found in many C-MeT domains.



**Scheme 11:** Mechanism of the methyl transfer of the C-Methyl transferase domain.

## 1.7 Polyketide Synthases

PKS can be divided in three different groups. Based on the architecture of the protein that is responsible for the final compound.

### Type I PKS

Type I PKS are found in bacteria and fungi. They are like Type I FAS and consist of covalently linked enzymes. These PKS are further divided in modular (section 1.8) and iterative (section 1.9) types. Subclasses of the iterative Type I PKS are non-reducing (NR-PKS), partially reducing (PR-PKS, section 1.9) and highly reducing PKS (HR-PKS, section 1.10).<sup>36</sup>

### Type II PKS

Type II PKS are mono domain enzymes that come together for several catalyzing processes and form a complex. This complex consists of five to ten individual enzymes.

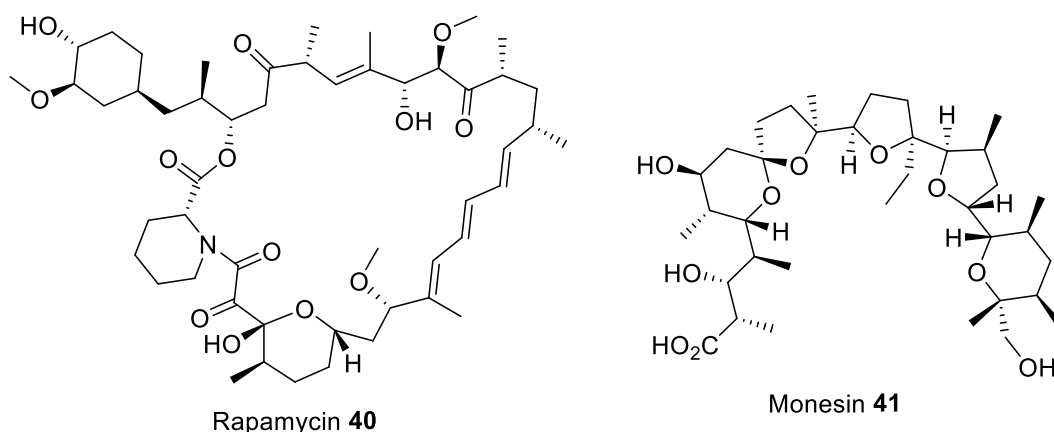
The synthesized polyketides normally have a chain length of 16 to 24 carbons. Type II PKS systems are typically found in bacteria.<sup>37</sup>

### Type III PKS

Type III PKS are found in plants, bacteria and fungi. They are simple homodimers of keto synthases. These systems do not use an ACP and they have no  $\beta$ -processing domains. They react directly with the extender substrates in a decarboxylative Claisen condensation. Type III PKS are iterative and produce small mono- and bicyclic aromatic metabolites like such as flavonoids **27**.

## 1.8 Modular Type I Polyketide Synthases

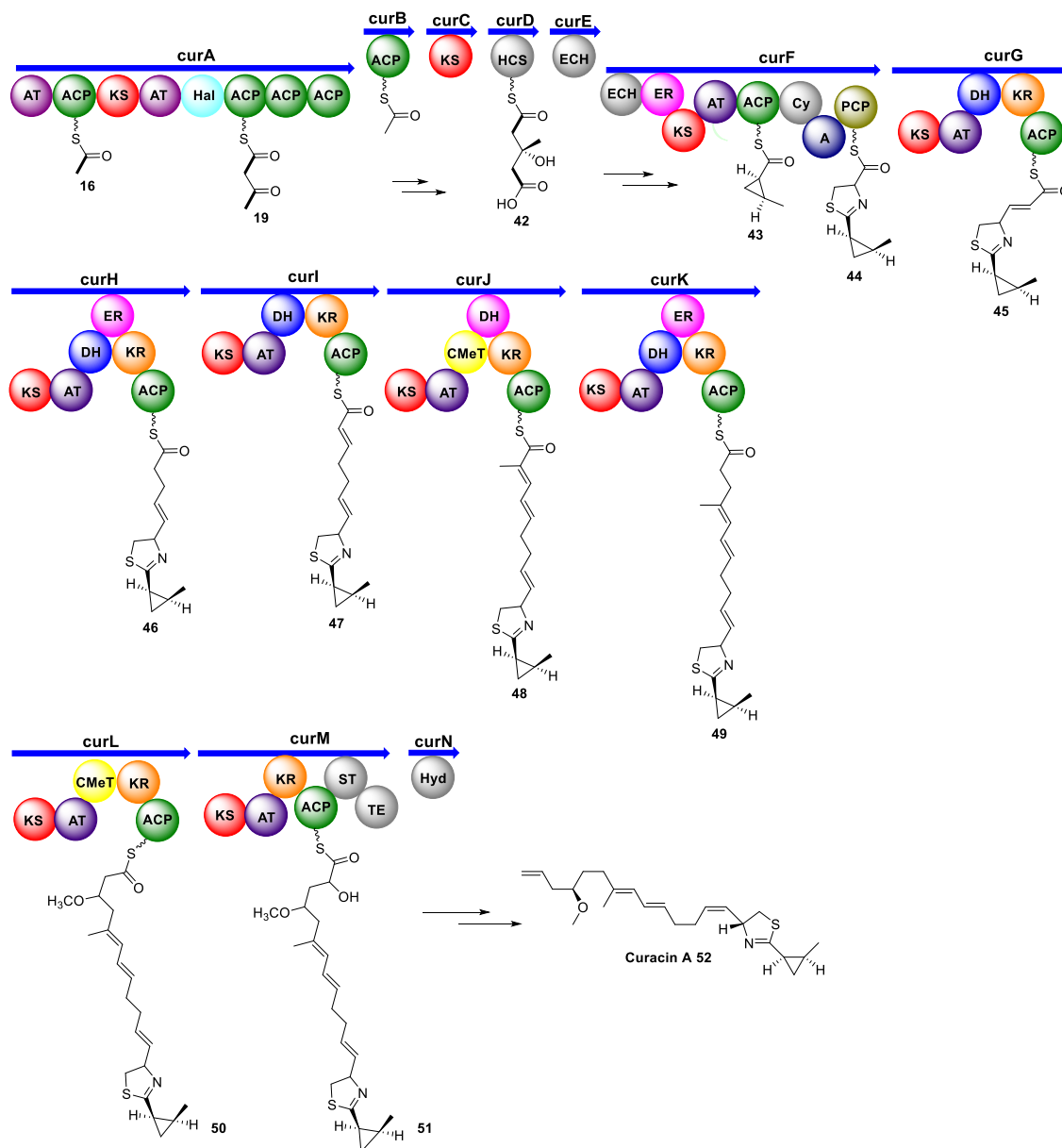
Modular type I PKS occur in bacteria and consist of ordered assembly lines of multi-domain modules. Examples of polyketides, which are synthesised by modular type I PKS, are shown in Figure 4. One module consists of KS, AT and ACP and optional KR, DH and ER domains. The PKS consist of two or more modules each extending and modifying a specific chain-elongation intermediate before transfer to the next module for further processing. Each module typically acts only once in the synthesis of a polyketide. The structure of the product can often be predicted by the domain sequence of the enzyme.<sup>38,39</sup>



**Figure 4:** Examples of polyketides synthesised by modular type I PKS<sup>40</sup>

One of the most investigated modular PKS is that which synthesizes Curacin A **52**. This system is a hybrid between a PKS and non-ribosomal peptide synthetase (NRPS) components from the cyanobacterium *Lyngbya majuscula*.<sup>41</sup> Curacin A **52** contains several interesting structural features. These includes the terminal alkene, a thiazoline ring and a unique cyclopropyl moiety, which is essential to the compound's biological

activity.<sup>41,42</sup> A gene cluster encoding 14 open reading frames (ORF) encodes the enzymes, which are responsible for the synthesis of Curacin A **52** (Scheme 12).



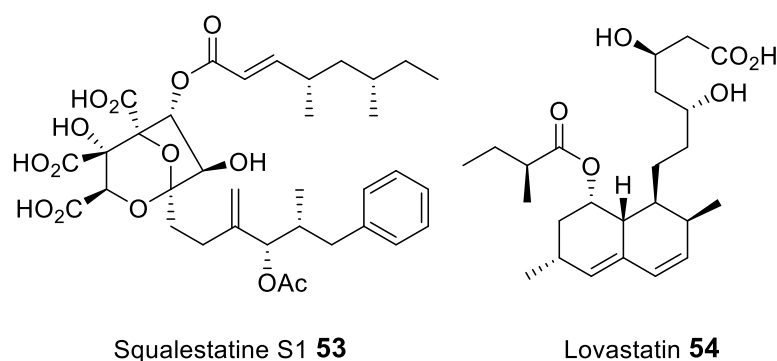
**Scheme 12:** Biosynthesis of Curacin A **52** of *Lyngbya majuscula*.<sup>41</sup>

The biosynthesis includes different unique features. This includes an HMG-CoA synthase cassette (HCS) located on CurD (Scheme 12). Further, CurA contains a unique GCN5-related *N*-acetyltransferase (GNAT) loading domain and an associated ACP.<sup>42</sup> The loading module passes an acetyl group **16** to the ACP that then condenses with one of three tandem ACP present in the module of CurA (Scheme 12).<sup>41–43</sup> A hydroxymethylglutaryl-CoA synthase cassette (mevalonate pathway) catalyses the formation of hydroxymethylglutaryl acid **42** by the addition of a malonyl-CoA unit to the terminal ketide **19** of the aceto-acetyl-ACP moiety of ACP1, ACP2, or ACP3.<sup>42,43</sup>

Subsequent enzymes, including a unique heme independent halogenase (Hal) catalyses the formation of a cyclopropyl ring **42**.<sup>3,42,43</sup> A cysteine specific NRPS module located on CurF follows after cyclopropyl ring formation, and due to the activity of a cyclizing condensation domain, forms a thiazole ring attached to the cyclopropyl moiety, resulting in thiazoline **44**.<sup>41-43</sup> Seven PKS (curG-CurM) modules follow to extend the growing polyketide chain. The final curacin synthase module, with intermediate **51**, employs a rare offloading strategy involving a sulfotransferase. The sulfotransferase sulfates the hydroxyl group of carbon 15, which activates the molecule for decarboxylation and terminal alkene formation, resulting in Curacin A **52** (Scheme 12).

### 1.9 Iterative Type I Polyketide Synthases

Most fungal PKS are iterative type I synthases consisting of one single module. Examples of polyketides, which are synthesised by iterative type I PKS, are shown in Figure 5. These enzymes repetitively use a single set of catalytic domains to assemble complex metabolites. The growing polyketide is cycled in each round of elongation. The most complex and interesting fact is that the extent of  $\beta$ -processing can have a large variation between the different extensions.



**Figure 5:** Example for products of type I iterative PKS systems<sup>39</sup>

Fungal iterative type I PKS are further divided into three functional classes based on the presence or absence of particular catalytic domains. These sub-classes are highly reducing (HR-PKS), partial reducing (PR-PKS) and non-reducing polyketide synthases (NR-PKS).

HR-PKS possess the full set of modifying domains (KR, ER, and DH) and usually a C-MeT. NR-PKS lack the whole set of  $\beta$ -processing domains (KR, ER, DH) but can possess a C-MeT. In contrast, they possess some unique domains. For example the *N*-terminal *starter unit acyl-carrier transacylase* (SAT).<sup>14,44</sup> Furthermore, another unique domain, which controls the cyclisation of the backbone and in addition the chain length

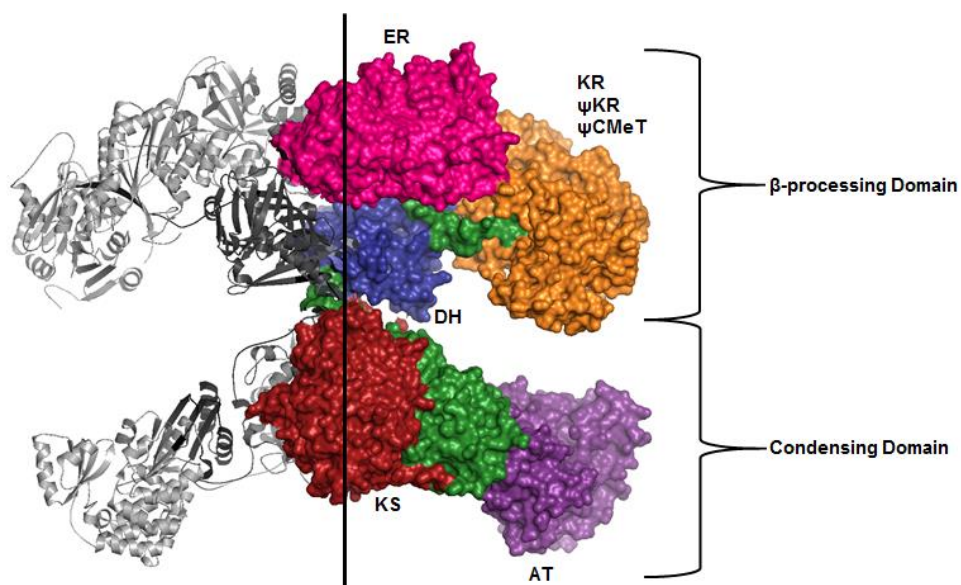
of the product. This domain is known as a *product template* (PT) domain. PT domains are required because poly-  $\beta$ -keto intermediates have acidic protons and keto functions that make the polyketide intermediate very unstable.<sup>14,44</sup> Hence, spontaneous cyclisation would be possible and the PT domain controls reactions.<sup>45</sup>

PR-PKS possess an incomplete set of modifying domains and often no known off-loading domain.<sup>46/45</sup> The PKS also consists of a KS, AT, DH, so called core domain, KR and ACP domain.<sup>47</sup> In PR-PKS the most important domain is the ketoreductase (KR) domain. Here, the KR domain selects which intermediate will be reduced and controls, through the size of substrate binding pocket, the chain length. However, how the KR domain achieves the selective keto-reduction is not yet fully understood.<sup>45</sup>

HR-PKS will be discussed in detail because the focus in this work is on HR-PKS.

### 1.10 Highly Reducing Iterative Polyketide Synthase (HR-PKS)

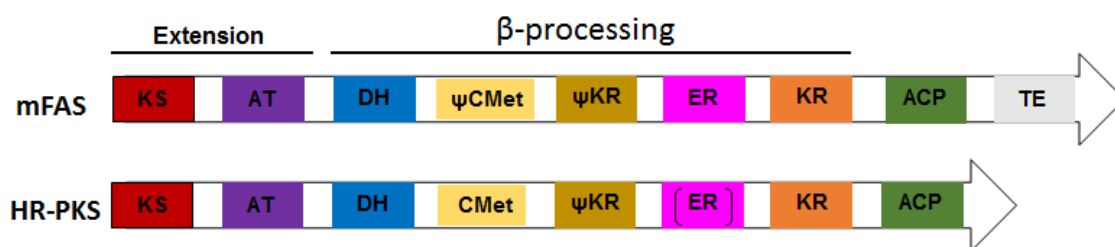
HR-PKS are a group of multi domain enzymes which act in an iterative way.<sup>48,49</sup> They also have a high sequence (and probable high structural similarity) to mammalian FAS. The strongest evidence for a common PKS-FAS architecture comes from their homologous domain organization (Figure 7) and sequence homology.<sup>40</sup> In addition, another evidence for a common architecture is the crystal structures of KS-AT di-domains excised from HR-PKS modules, which are strikingly similar to the KS-AT of Type I FAS.<sup>50</sup>



**Figure 6:** Homodimer crystal structure of mammalian FAS. PDB entry 2VZ8. Displayed in violet the AT domain, in green linker regions, in red the KS domain, in blue the DH domain, in pink the ER domain and in orange the KR,  $\psi$ KR and  $\psi$ C-MeT domain.

Based on the mFAS structure it is thought that HR-PKS structurally consists of a homodimer of two identical protein subunits, in which the catalytic domains in the N-terminal section (KS) and malonyl/acetyltransferase (AT) are separated by a short linker of extended polypeptides from the C-terminal domains. These are the dehydratase (DH), enoyl reductase (ER), -ketoacyl reductase (KR) and acyl carrier protein (ACP, Figure 6). In contrast to mFAS, may HR-PKS also have a full and active C-MeT domain (Figure 7).<sup>51/52</sup>

The KS, AT and ACP are active in each extension cycle, but the other domains can be active or inactive during  $\beta$ -processing cycle and this gives rise to the observed programming of the HR-PKS. Hence, the difference in the programming of the HR-PKS results in structurally complex compounds, such as Lovastatin **54** or Squalestatin S1 **53**.<sup>53</sup>

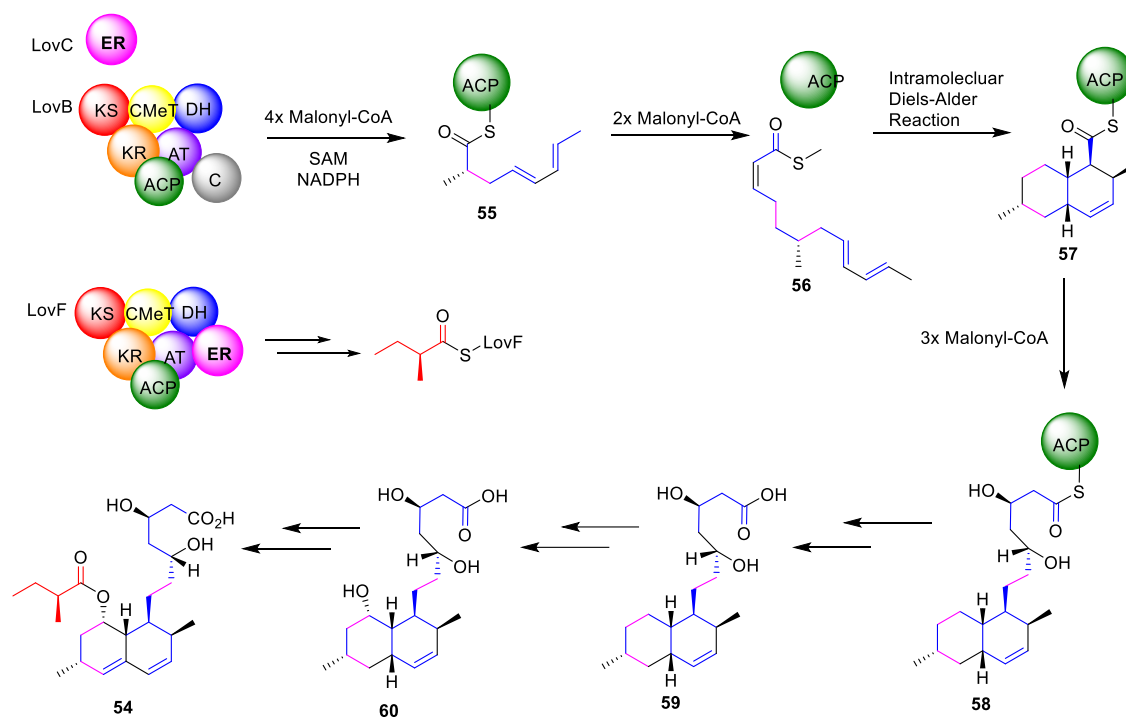


**Figure 7:** General domain architecture of the HR-PKS in comparison with the mFAS: KS and AT domains are responsible for the chain elongation while DH, C-MeT, ER and KR domains are  $\beta$ -processing domains. HR-PKS do not always possess an active ER domain that is why it is in brackets. HR-PKS often terminates with an ACP domain.<sup>39</sup>

One example for such a product of an iterative PKS is Lovastatin **54**. Lovastatin **54** is a polyketide metabolite produced by the filamentous fungus *Aspergillus terreus*. Among the enzymes that biosynthesize lovastatin are two polyketide synthases (PKS) and numerous accessory enzymes (Scheme 13).<sup>54,55</sup> The two-mega synthases are the lovastatin nonaketide synthase (LovB), which catalyses the assembly of the decalin core (Scheme 13, marked in blue) and lovastatin diketide synthase (LovF), which catalyses the assembly of the 2-methylbutyrate side chain (Scheme 13, marked in red), respectively.

LovB possess an inactive enoyl reductase (ER<sup>0</sup>). A *trans*-acting enoyl reductases, LovC, is encoded in the biosynthesis cluster. This *trans*-acting ER catalyses some reduction during the biosynthesis of Lovastatin (Scheme 13, marked in violet). LovC specifically interacts with only LovB, but not LovF, and accepts only three out of eight possible LovB intermediates as its substrates (tetra-, penta-, and heptaketides). Similar *trans*-acting ER have also been reported in other fungi and in particular the Tenellin PKS-NRPS (section 1.13).

When LovC and LovB are coupled, they iteratively catalyses more than 30 precisely synchronized reactions to yield a 19-carbon intermediate dihydromonacolin L **59** (Scheme 13). In comparison, LovF only catalyses one round of Claisen condensation to produce a 2-methylbutyryl intermediate, which is attached to the C-8 hydroxyl of monacolin J **60** to produce lovastatin **54** (Scheme 13).<sup>54,55</sup>

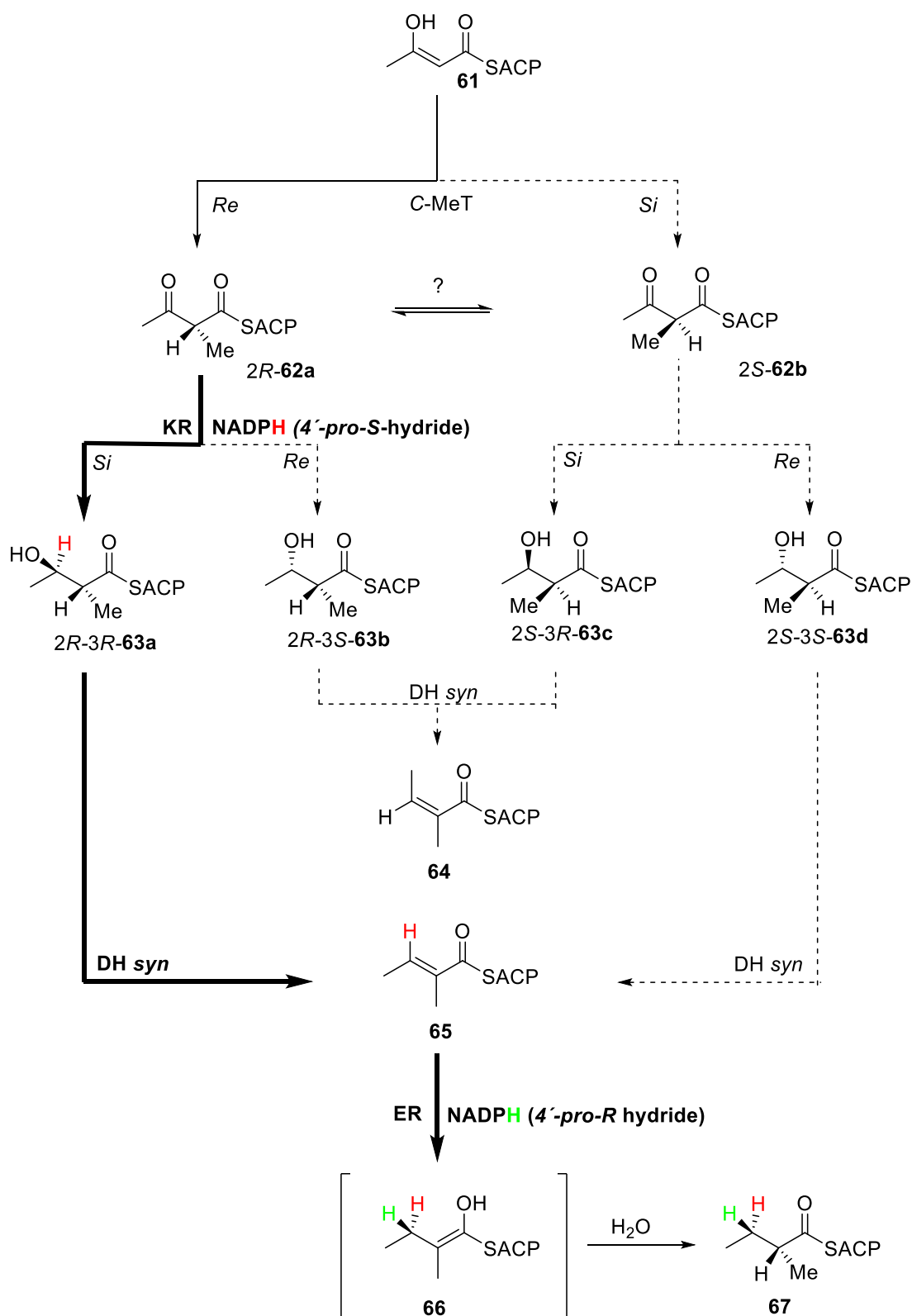


**Scheme 13:** The biosynthesis of Lovastatin **54** as an example of an iterative PKS.<sup>54,55</sup>

### 1.11 Stereochemical course in HR-PKS

The biosynthesis of polyketides proceeds under a very high level of stereocontrol.<sup>56</sup> As the stereochemistry of various functional groups affects the bioactivity of polyketides, the origins of stereochemical control are of significant interest for creating derivatives of these compounds by various methods such as genetic engineering.<sup>56</sup> Different *in vitro* studies have been able to provide a near complete stereochemical description of the first cycle of beta-modification reactions of a fungal highly reducing polyketide synthase (HR-PKS).<sup>56–58</sup> The following section describes the current state of knowledge in the stereochemical course of HR-PKS. As an example, for the stereochemical course of HR-PKS, the stereochemical course of the HR-PKS SQTKS is displayed in Scheme 14.





**Scheme 14:** Stereochemical course of KR, DH and ER domain of SQTCS. Experimental evidence supports the pathway shown with bold arrows.

The starting point is the ACP bound diketide **61**, which in the first step is methylated by the *C*-MeT domain to give the *C*-2 methylated 3-oxo diketide *2R*-**62a** (Scheme14). Afterwards, the *2R*, *3R* -diketide is created by reduction of the 3-oxo group by the KR

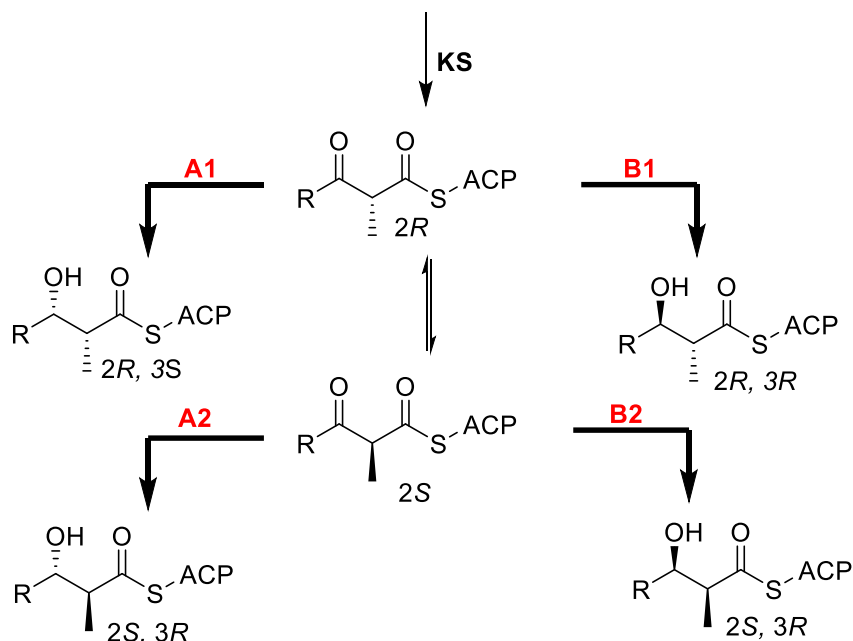
domain using NADPH as the cofactor (Scheme 14). Studies in the Cox group showed that the SQTCS KR releases a 3*R* substrate, and thus it must reduce the 3-oxo group of its substrate **62a** by 3-*Si* addition of the hydride. Isotopic labelling studies have shown that the 4'-*pro-S* hydride of NADPH is transferred in this step.<sup>56-58</sup> Since racemization at the 2-position of the diketide is strongly disfavored after reduction of the 3-oxo group, this observation also suggests that the KR accepts and releases 2*R*-methylated diketides (*e.g.* 2*R*-**63a**, Scheme 14).<sup>56-58</sup> Subsequently the DH domain dehydrates 2*R*, 3*R* substrate **63a** to give E-products **65** by *syn* elimination. Cox. *et. al.* showed that the SQTCS DH has exactly the same stereochemical selectivity as the mFAS DH.<sup>57</sup> Even though the SQTCS substrate is methylated at the 2-position, the 2*R* stereochemistry ensures that the 2-*pro-S* proton is removed during DH reaction.<sup>56-58</sup> In addition, Cox *et. al* elucidated the stereoselectivity of the ER domain.<sup>58</sup> They showed that the stereochemical preferences at the β-carbon are identical for SQTCS ER and mFAS ER, in terms of both the cofactor itself (transfer of 4'-*pro-R* hydrogen) and the substrate (addition of hydride to the 3-*Re* face, Scheme 14).<sup>56-58</sup>

Overall, the *in vitro* studies concerning the stereochemical course show that SQTCS shares not only the sequence homology and domain organization with mFAS (Chapter 1.10) but also that its fundamental mechanisms for substrate reaction and stereoselectivity are preserved.<sup>56-58</sup> Furthermore, other HR-PKS, for example such as TENS, DMBS and MILS should have the some stereochemical course, since they show significant sequence homology.

### 1.11.1 Excursus into KR domains

The KR domain determines the stereochemistry of the processed alcohol **63a** and the stereochemistry of substituents at the β-position. KR domains are classified into different stereochemical types (Scheme 15), often distinguishable through fingerprints in their amino acid sequence.<sup>59</sup> There is a distinction between A1-Type, A2-Type, B1-Type, B2 type and C-types (Latter will be not explained in detail). The residues, which are unique to each type of KR were identified through multiple sequence alignments.<sup>59</sup> A leucine-aspartate-aspartate (LDD) motif is consistently present in the B-Type KR, whereas a conserved tryptophan is present in A-type KR domains.<sup>59</sup> A2-Type KR domains can often be distinguished from A1-Type KR domains through the presence of a histidine three residues N-terminal to the catalytic tyrosine B2-Type KR domains can often be

distinguished from B1-type KR domains through the presence of a proline two residues C-terminal to the catalytic tyrosine.<sup>59</sup>



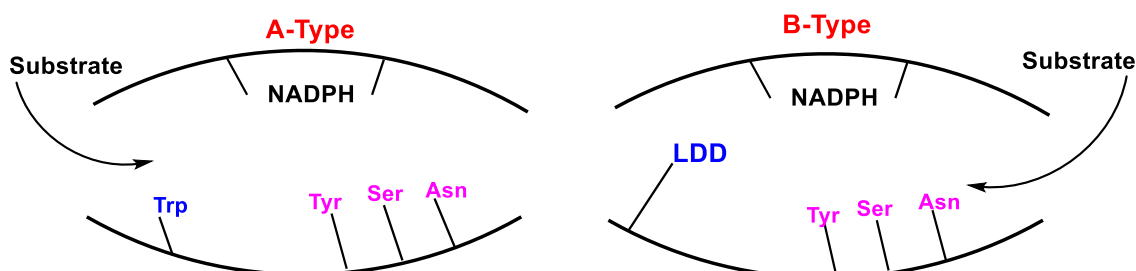
**Scheme 15:** KR stereocontrol on natural substrate.

The KR domains can stereoselectively generate either an 3*S*-hydroxyl group (A-type KR domains, Scheme 14 **63b** and **63d**) or a 3*R*-hydroxyl group (B-type KR domains, Scheme 14 **63a** and **63c**). Furthermore, they can also be stereoselective in reducing a substrate with either a 2*R*-substituent (A1- or B1-type KR) or a 2*S*-substituent (A2- or B2-type KR).<sup>15,60–62</sup>

The crystal structures of the EryKR1 and TylKR1 domains have revealed that the conserved catalytic residues of KR domains cooperate as observed in other short-chain dehydrogenase/reductase (SDR) enzymes.<sup>15,60–62</sup> The substrate  $\beta$ -carbonyl is positioned adjacent to the nicotinamide ring by a tyrosine and a serine so that the 4-*pro*-S hydride can attack the  $\beta$ -carbonyl (Scheme 16). As anticipated, the A- and B-type sequence motifs reside on opposite sides of the catalytic groove. Keatinge-Clay and Raid *et. al.* hypothesized that the LDD motif (The leucine in this motif can be occasionally replaced by valine or isoleucine; several residues can replace the first aspartate; the second aspartate is strictly conserved) mediate the entrance of the phosphopantetheinyl-bound  $\beta$ -ketoacyl thiolester substrate from the right side of the catalytic groove (the side closer to the NADPH binding site), resulting in formation of a 3-*R* hydroxyl group (Scheme 16). The conserved tryptophan is hypothesized to guide the substrate into the groove from the left side, thereby presenting the opposite face of the  $\beta$ -keto group to NADPH to produce a 3-*S*

hydroxyl group.<sup>15,60–62</sup> However, the orientations of the catalytic residues change little between KR-types (Scheme 16).<sup>59</sup>

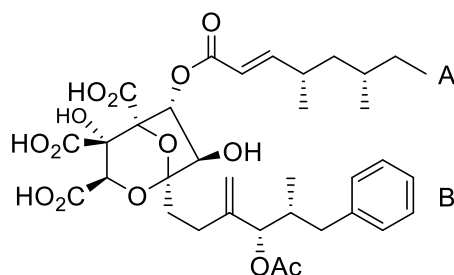
In addition, the nicotinamide coenzyme binds to both A- and B-type KRs in the same location and conformation. NADPH occupies a characteristic groove in the Rossmann fold of KR and presents its *pro-4-S* hydride to the active site, consistent with the observation that both A- and B-type KRs transfer the NADPH *pro-4 S* hydride to substrates (Scheme 16).<sup>59</sup> By this analysis the KRs of mFAS and HR-PKS are B1-types.



**Scheme 16:** Display of the active pocket with the bound cofactor **11** and different catalytic and structural amino acids.

### 1.12 Biosynthesis of Squalestatin Tetraketide Synthase

Squalestatins, also known as zaragozic acids, are polyketides isolated from the fungus *Phoma sp.* C2932, among others.<sup>59</sup> They were identified by research groups of Glaxo and Merck.<sup>63</sup> They are potent inhibitors of cholesterol biosynthesis and act by inhibiting squalene synthase probably by mimicking the natural substrate, presqualene pyrophosphate. All squalestatins contain the common 4,6,7-trihydroxy-2,8-dioxabicyclo[3.2.1]octane-3,4,5-tricarboxylic acid core and hydrophobic side chains.<sup>39</sup> The side chains have a high diversity in the chain length, methylation pattern and methylation positions. In addition, the reduction pattern of the side chains can differ.



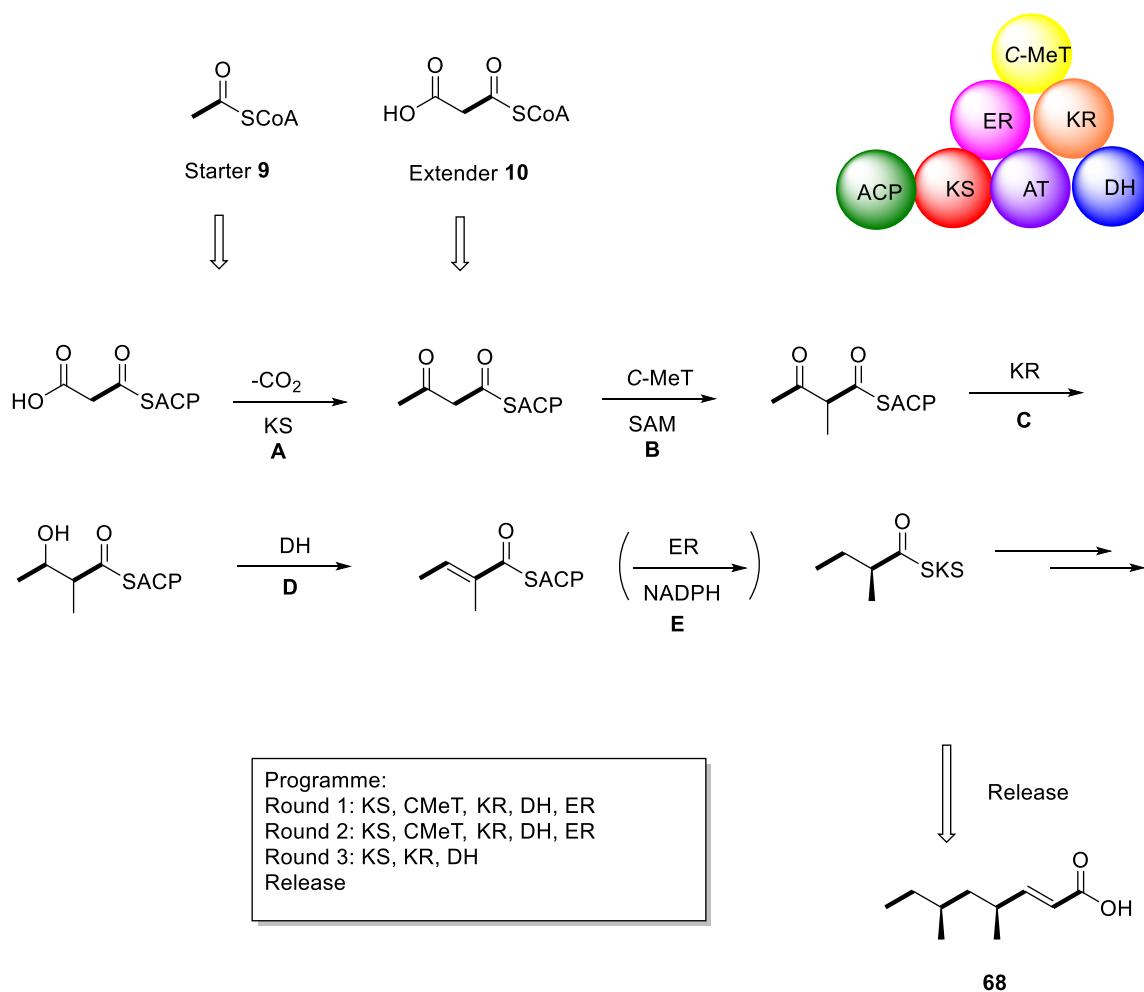
Squalestatin S1 **53**

**Figure8:** Structure of Squalestatin S1 **53**.

The biosynthesis of squalestatin S1 **53** (Figure 8) has been observed by several feeding experiments using isotopes. Chain A is a tetraketide with acetate as starter and extender

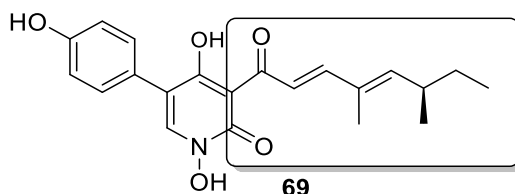
units. The genes which encoded the enzymes responsible for the tetraketide chain were identified by Cox et al.<sup>64</sup> Chain A is produced by a single PKS known as squalestatin tetraketide synthase (SQTKS).<sup>63</sup> SQTKS consists of KS, AT, DH, C-MeT, ER, KR and ACP domains (Scheme 17). The biosynthesis comprises three rounds and starts with the binding of acetyl-CoA **9** to the AT. The AT then transfers the acetate to the KS and binds malonyl-CoA **10**. After that, the ACP takes the malonyl group from the AT *via* trans-esterification. After decarboxylation of the bound malonyl group, the chain at the ACP is extended by the addition of the acetate from the KS (Scheme 17A). The ACP carries the  $\beta$ -keto compound and the C-MeT adds a methyl group at the  $\alpha$ -position (Scheme 17B).<sup>39</sup> Then the remaining  $\beta$ -processing steps follow (Scheme 17 A-E).

SQTKS, like other characterised fungal HR-PKS, lacks a TE or any other obvious releasing domain. Most characterised highly reduced polyketides do not exist as free polyketide chains but are extensively modified by post-PKS enzymes and linked to other moieties. During the synthesis of SqualestatinS1 **53** the tetraketide **68** becomes linked to the squalestatin core.



**Scheme 17:** Biosynthesis of the side chain of Squalestatin S1 through the iterative working of SQTKS.

### 1.13 Biosynthesis of Tenellin

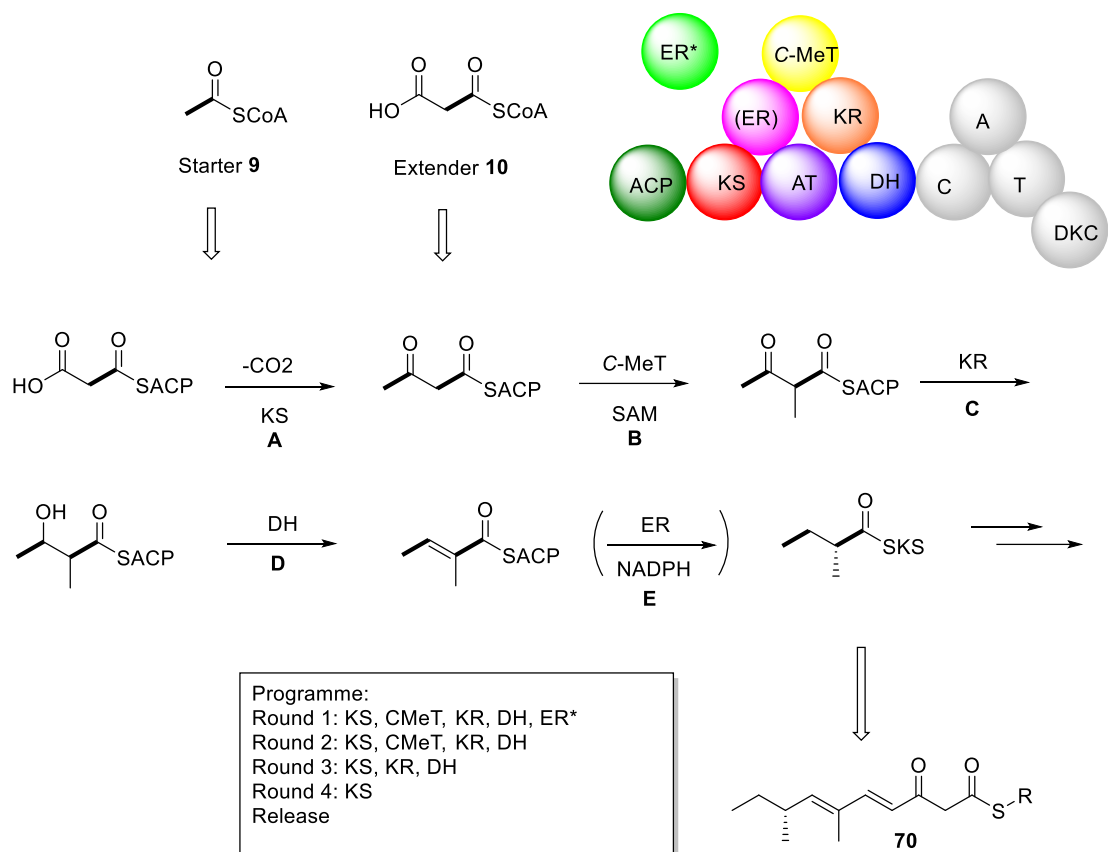


**Figure9:** Structure of Tenellin **69**

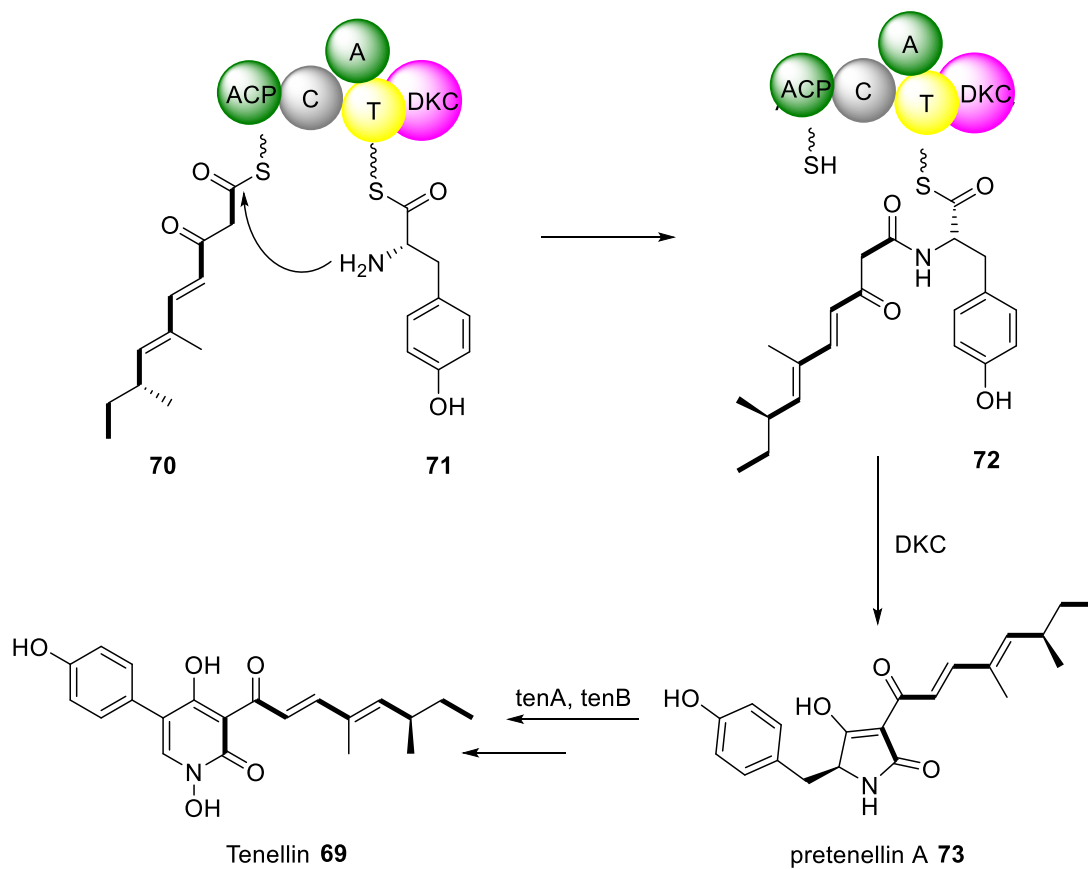
A hybrid PKS-NRPS system biosynthesizes the fungal metabolite Tenellin **69**.<sup>65</sup> The PKS which builds the polyketide part of tenellin, is an HR-PKS. The special thing about this PKS is that the ER domain is inactive and a *trans*-acting ER domain catalyses the reaction.

The biosynthesis of pretenellin A **73** starts with the building block acetyl CoA **9**, which is elongated four times with malonyl CoA **10** (Scheme 18). After the first elongation, all  $\beta$ -processing enzymes including the C-MeT are active and the enoyl reduction is performed by the TenC *trans*-ER. This *trans*-ER is only active in the first elongation round so an unsaturated carbon chain is formed in the next elongations steps (Scheme 18). In the first two rounds of the programming, the methylation (Scheme 18B) takes place, similarly to other PKS systems like Squalestatin. The KR (Scheme 18C) is functional after the first three rounds of extension, but inactive after the final elongation step.

Afterwards the NRPS adds a tyrosine to the resultant polyketide chain and catalyses a Dieckmann cyclisation (Scheme 19), which acts as the release step. After release, two cytochrome P450 enzymes oxidize pretenellin A **73**. TenA catalyses an oxidative ring expansion to form the pyridone. Finally TenB hydroxylates the pyridone nitrogen.<sup>48,65</sup>



**Scheme 18:** Biosynthesis of the PKS part of Tenellin through the iterative working of TenS.



**Scheme 19:** Part of the biosynthesis of Tenellin **69**.

## 1.14 Engineering of Polyketide Synthases

The practical usefulness and chemical diversity of natural products stands against some difficulties, such as limited accessibility through total synthesis, structural complexity, solubility, bioavailability, exposure, stability, and metabolism.<sup>66</sup> This explains the long standing interest in engineering PKS enzymes to produce novel polyketides in a predictable manner. There is a vast potential for producing pharmaceutically relevant compounds and in addition, it would be a simple way to get these products.

Theoretically, different changes to HR-PKS or the tailoring enzymes should be possible (Figure 10). Focused on the PKS, the most obvious starting point is manipulating individual HR-PKS domains. For example, by specific modification of active site residues, the domains could be inactivated or their substrate or stereochemical specificity could be adjusted.<sup>67</sup> Also, there is the possibility to switch one domain for another from the same or a different HR-PKS in order to alter these same properties. This is a reason why they are attractive targets for genetic engineering. Modifications also include the removal or addition of specific tailoring steps or changes in the regiospecificity or stereospecificity of the reactions.<sup>68</sup>

To date, different engineering approaches have been undertaken in modular PKS systems, from which a few examples will be described in the following.

The AT domain has been a key target for polyketide engineering in modular PKS, and a number of studies focus on understanding and engineering building block selection by this domain.<sup>69</sup> Mainly, typical starter units, substrates different to malonyl-CoA **10** or methylmalonyl-CoA are used. For example, an AT domain from the modular monensin PKS4 and the loading AT from the avermectin PKS5, both with relaxed substrate specificities, were subjected to computational modelling and structural analysis, respectively, to enable the identification of residues responsible for substrate binding and specificity.<sup>69</sup>

In addition, the KR domain of modular PKS has also been a target for engineering. The specificity can be modified through site-directed mutagenesis. As an example, the mutation of the catalytic tyrosine of the DEBS module 6 KR. The residue was identified through homology modeling of KR domains to the SDR superfamily of enzymes. Afterwards, this was tested *in vitro* through the point mutation of this residue. The *in vivo* production resulted in the expected ketone, illustrating that even through deactivating of domain it is still possible to generate new polyketides.<sup>60</sup> In addition to site-directed

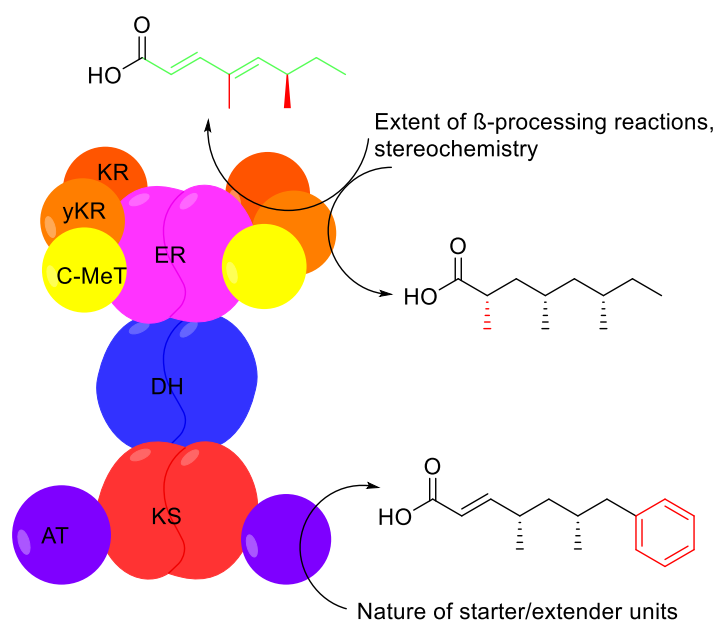


methods for engineering KR domains, numerous experiments have illustrated the viability of full KR domain swaps to modulate the stereochemistry and oxidation state of a given polyketide in modular systems.<sup>70</sup>

Engineering the DH domain has not been as extensively pursued as the engineering of other PKS domains, but even so several examples of both site-directed and swap-based engineering exist. For example a DH domain from the modular DEBS was inserted into a module of the avermectin synthase in *S. avermitilis*, resulting in the exclusive production of C22,23-unsaturated avermectins.<sup>71</sup>

One of the earliest examples of mutasynthesis targeted the avermectin PKS. A strain of *Streptomyces avermitilis* was generated wherein the enzymes required for generating the precursors 2-methylbutyryl-CoA and isobutyryl-CoA were inactivated. Out of 800 potential precursors, which were tested, over 40 starter unit analogs were tolerated by the avermectin PKS.<sup>72</sup> Through this method, a cyclohexyl-containing avermectin derivative (Doramectin) was generated that exhibited increased anti parasitic activity against veterinary pathogens.<sup>73</sup>

Another approach to understand and engineer PKS is through iterative type I PKS. However, due to the complexity of programming of Type I PKS, very few reports of rational engineering exist. Numerous bioactive polyketides are synthesized by iterative type I PKS. These fungal PKS have only one module that is used iteratively. The related programming rules regarding substrate selection, catalytic domain utilization in each elongation cycle, regiospecific modification, polyketide chain length, control, chain release and transfer, *etc.* are only just beginning to be understood.<sup>69</sup> A more simple approach is through non-reducing iterative type I PKS (NR-PKS) that are involved in aromatic polyketide synthesis. The starter unit ACP transacylase (SAT) domain has been the focus of several recent studies including domain swapping to generate new aromatic polyketides<sup>69</sup> and structural analysis to identify the basis for acyl unit selection, since unnatural starter units often are processed properly by the rest of the catalytic domains of NR-PKS.<sup>69</sup>



**Figure 10:** Schematic overview over different possibilities for rational changes of the programming of PKS.

### 1.15 Engineering of C-MeT and ER domain from fungal HR-PKS

Very few engineering approaches have been reported with iterative C-MeT domains from PKS. One rare example is the approach from Cox *et al.* where the C-MeT domain of the SQTKS S1 was engineered *in vivo* in order to investigate the programme of SQTKS.<sup>74</sup> Their aim was to disable the methylation of the tetraketide by blocking the co-factor SAM binding to alter the methylation pattern. This was done through site-directed mutations of potential residues involved in SAM binding.<sup>74</sup> However, the results of LCMS analysis of the mutant protein showed that mutations of the C-MeT domain from SQTKS did not give the desired effect with certainty. The mutations overall might prevent the SAM binding, but the linear alkene product was rapidly metabolised by growing *A. oryzae* M-2-3. From these results they could only propose that the methylation of the tetraketide prevents  $\beta$ -oxidation.<sup>74</sup> Furthermore, they suggested that the inhibition of the methylation in the early steps of the elongation cycle may prevent further processing and during the later steps, it might possibly block the PKS. Overall the programming of HR-PKS was not able to be revealed or changed though these *in vivo* expression and mutagenesis experiments.<sup>74</sup>

Even less engineering has been done with ER domains. Some early mutagenesis studies of ER domains in DEBS module 4 targeted the conserved "HAAAGGVGMA" NADPH binding motif for engineering.<sup>70</sup> Changing this sequence to "HAAASPVGMA",

based on the NADPH binding motif of the inactive KR domain from the same module, resulted in production of primarily D-6,7-anhydroerythromycin C.<sup>70</sup>

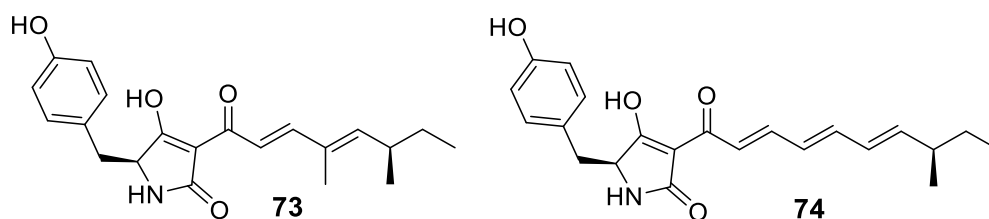
A different approach of PKS engineering, which is not at the amino acid, domain, and module level, is the precursor-directed biosynthesis (PDB). By feeding analogues of their natural building blocks that are likely tolerated by the native biosynthetic PKS in the producing host. The efficiency of this technique can be enhanced with the previously mentioned mutasynthesis, wherein the naturally occurring precursor pathways are inactivated, thus removing competition from natural precursors.<sup>70</sup> In addition to precursor supplementation through feeding experiments, metabolic pathways that produce precursor analogs can be introduced heterologously to replace the deleted pathways.<sup>66</sup>

As polyketide, engineering continues to progress, the development of tools and strategies, to understand the underlying PKS programming rules and exploit them for the diversification and overproduction of new polyketide compounds is of importance. Thereby, *in vitro* biochemical analysis using purified enzymes continues to serve as the most important method for studying PKS enzymology, particularly in understanding the precise function and substrate specificity of catalytic domains, reaction mechanisms, and the internal kinetics of the catalytic program.

### 1.16 Rational Domain Swaps between the HR-PKS TENS and DMBS

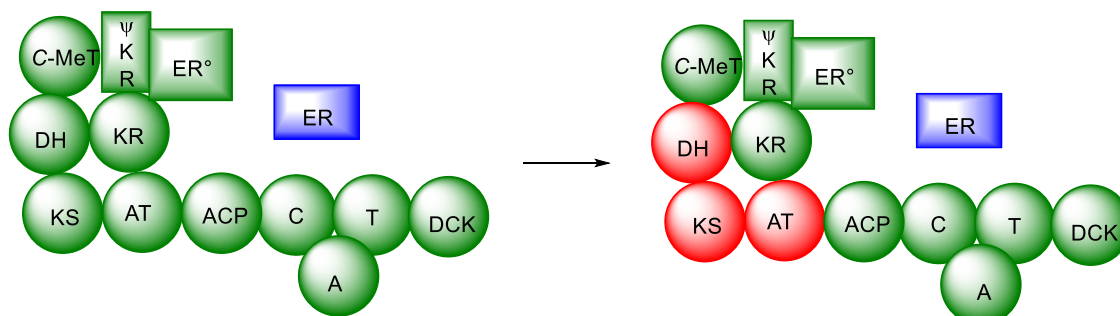
One of the very rare examples of engineering of HR-PKS is through domain swaps. In previous studies, Cox *et. al* could successfully swap different domains of the HR-PKS TENS with the corresponding sequence from the HR-PKS DMBS.<sup>48,75-77</sup> The TENS and DMBS polyketide synthases are 86% identical, but produce polyketides which differ in chain-length and methylation pattern.<sup>76</sup> The product of TENS is pretenellin A **73** and of DMBS predesmethylobassianin A **74** (Figure 11).

It was shown that the exchange of entire functional domains between these closely related (but differently programmed) HR-PKS, could lead to the creation of new polyketides (Figure 13).<sup>76</sup>



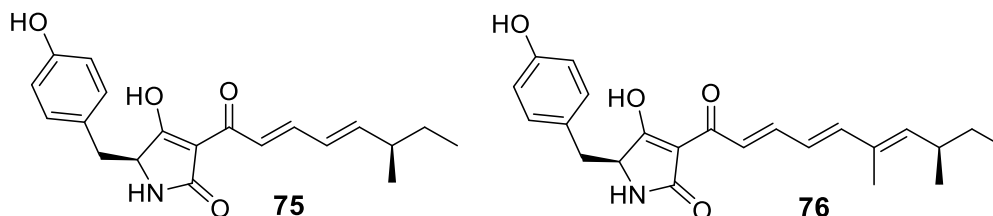
**Figure11:** Structure of pretenellin **73** and predesmethylobassianin A **74**.

The first domain swaps between TENS and DMBS included the KS, AT and DH of TENS with the ones of DMBS (Fig 12). This resulted in no change in the polyketide chain. The product was still pretenellin A **73**.



**Figure12:** Example of a swap experiment. The DH, KS and AT domain from TENS (green) was swapped with the KS AT and DH domain from DMBS (red).

However, if the C-MeT and KR domain from DMBS were swapped into TENS, this reprogrammed the PKS and resulted in two new products. Desmethyl-pretellin A **75** and prebassianin A **76** were isolated as the major and minor product (Figure 13).



**Figure13:** Structure of desmethyl-pretellin A **75** and prebassianin A **76**.

Overall, the domain swap experiments are, on the one hand, another example for the engineering approach of HR-PKS and on the other hand give a first insight into the programming of HR-PKS (section 1.17).

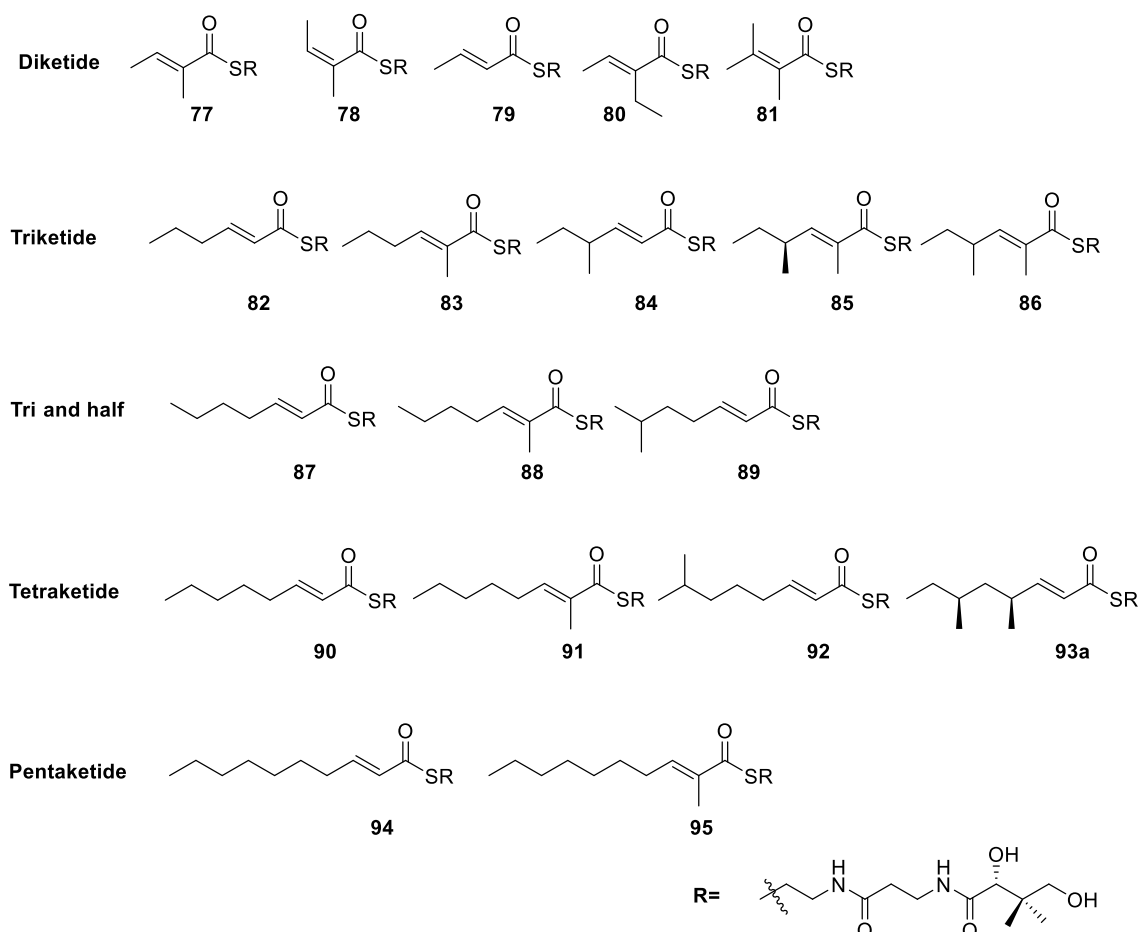
### 1.17 Programming in Iterative Type I HR-PKS

HR-PKS synthesize complex products using a single set of domains in a highly programmed, iterative fashion. Although many examples of type I HR-PKS are known in the literature, the mechanism that controls the iteration processes in these PKS has not been clearly understood.<sup>78</sup> This stands in contrast to modular Type I PKS. In modular PKS the programming is simply controlled through the assembly of the modules and the

availability of the specific domains in each respective module (for example see Curacin A **52**, section 1.8). However, a limited number of *in vitro* and *in vivo* investigations of HR-PKS catalytic domains have been discussed in the literature, which gives first evidence for a possible programming mechanism.

The catalytic domains of HR-PKS have two types of selectivity. First, they have a chemical selectivity, whereby they only act on substrates, which are chemically competent. For example, the DH domain can only act on ACP-bound  $\beta$ -alcohols and the AT can only supply malonyl extender units.<sup>79</sup>

The second level of selectivity is the structural selectivity of each domain. For example, Cox *et al.* recently showed *in vitro* that the ER domain from Squalstatin Tetraketide Synthase has low substrate selectivity and is able to effectively reduce a wide range of enoyl-pantetheines, even including unnatural isomers such as **83** and stereoisomers such as **78** (Figure 14).<sup>58</sup>



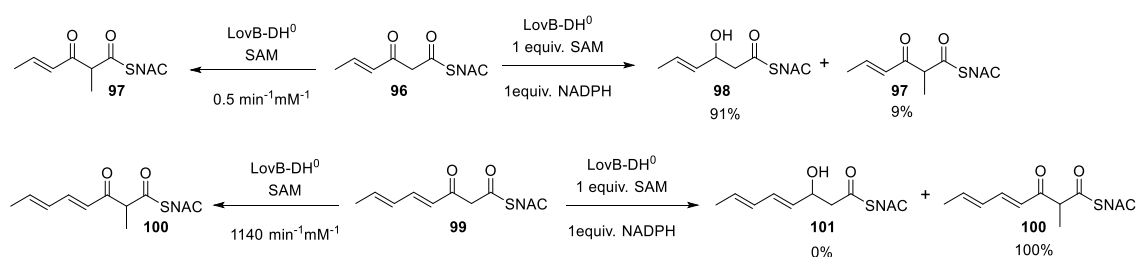
**Figure 14:** Synthesized pantetheine substrates for SQTKS ER assays from the previous studies.

Furthermore, Vederas *et. al.* could show *in vitro* that the C-MeT domain from the lovastatin nonaketide synthase displays tight selectivity for its substrate, while the KR in the same system is less selective.<sup>80</sup>

For the clarification, comparisons between individual rates of isolated C-MeT and KR domains toward the different substrates were investigated. For the subsequent quantification, and to prevent further tailoring reactions of the KR products in the assay, a point mutation was introduced in the active site of the DH domain to yield LNKS-DH<sup>0</sup>.<sup>80</sup>

By expressing LovB with an inactive DH domain and assaying these synthetic substrates, they recorded the Michaelis-Menten saturation kinetics (comparing  $k_{cat}/K_M$ ) and found that the C-MeT domain activity to substrate **99** is more than 2000 times higher than substrate **96** (Scheme 20). However, the KR domain exhibits far less selectivity in reactivity to these substrates. This may suggest an important role of the C-MeT domain as a gatekeeper domain in the programming of LNKS.<sup>56,80</sup>

To determine if there is indeed competitive catalysis between the KR and C-MeT domains towards the 3-ketoacyl substrates, Tang and co-workers performed a combined C-MeT/KR assay (Scheme 20). Both the natural substrates **96** and **99** were added to LNKS-DH<sup>0</sup> with the same equivalent of SAM and NADPH, and the yields of the methylation product and the ketoreduction product were compared (Scheme 20). Thereby, Tang and Vederas could observe for substrate **96** a 10:1 ratio of KR-catalysed product **98** to C-MeT-catalysed product **97**. However, for substrate **99**, only the methylated product **100** was observed and no reduced product **101** could be detected, thereby confirming the much higher catalytic efficiency of C-MeT domain toward substrate **93** compared to that of the KR domain. These experiments verified the assumption that the outcome of the modifying steps is determined by the relative activity of each domain towards specific substrates. However, one unexpected result is that the KR domain cannot reduce compound **99**, although it is the natural substrate for the KR domain in this biosynthesis. One possible explanation is that the recognition of this substrate needs interactions with the ACP domain.<sup>56,80</sup>



**Scheme 20:** Kinetic analysis of the LNKS CMeT domain towards different ketoacyl-SNAC substrates. Compounds **90** and **93** are both natural product analogues.

In addition, the stereoselectivity of each domain is fixed. For example, in the KR domain the stereochemical course of each  $\beta$ -ketoacyl-ACP reduction is programmed and is independent of either modular context or substrate structure, including chain length and substitution pattern.<sup>81</sup>

Overall, Vederas and Tang hypothesised that the programming mechanism of HR-PKS is kinetically controlled. In the modification stage, HR-PKS may adopt an assembly line-like model in which each substrate is passed through the catalytic domains sequentially in the order of  $C$ -MeT  $\rightarrow$  KR  $\rightarrow$  DH  $\rightarrow$  ER. In the case of the HR-PKS, LNKS encoded by *lovB*, which partners with a *trans*-acting ER encoded by *lovC*, the  $C$ -MeT only recognizes substrate **99** while excluding all of the other substrates completely (Scheme 20). They suggested that the outcome of the modifying steps is determined by the relative activity of each domain towards specific substrates.<sup>56,80</sup>

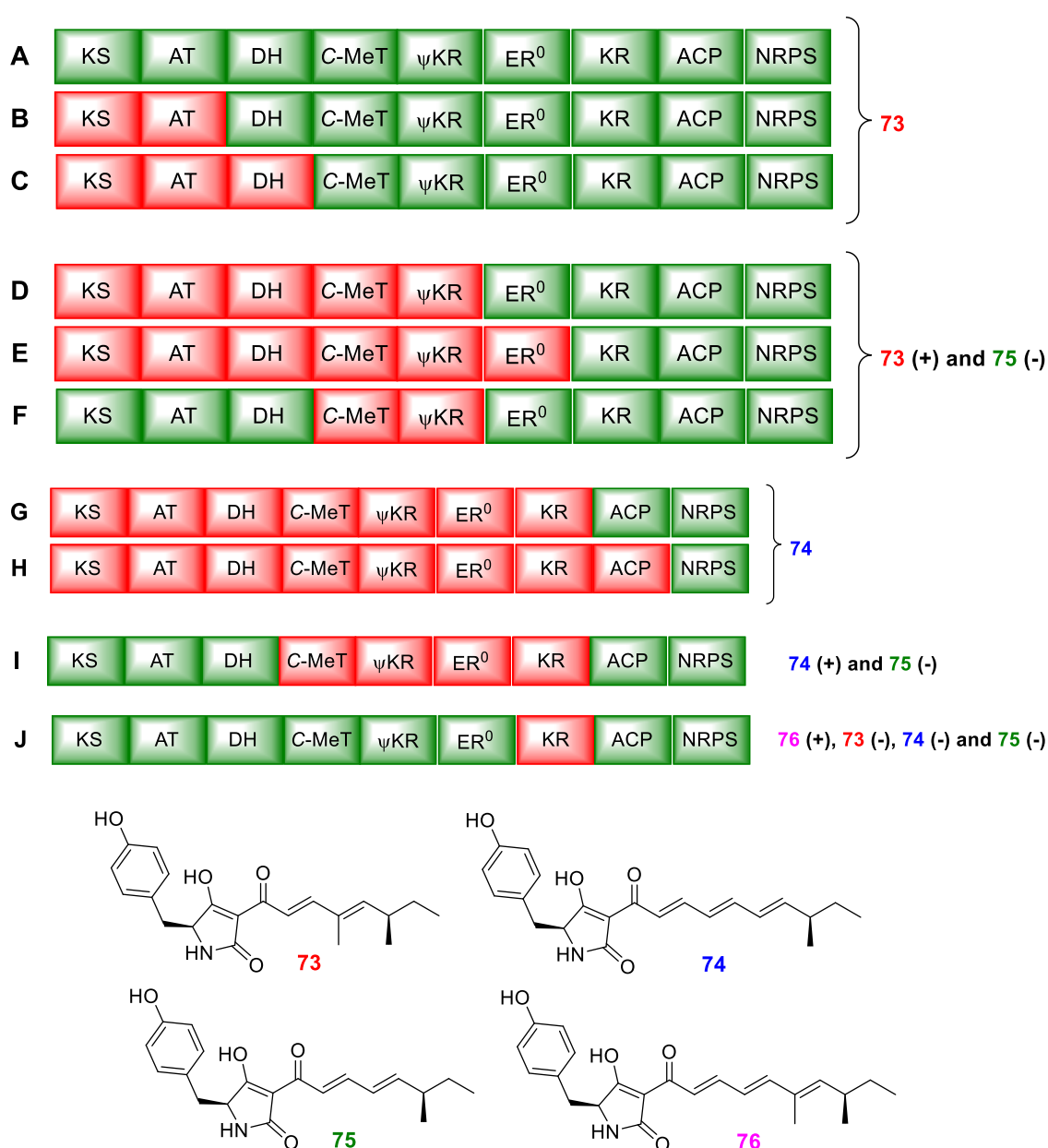
These *in vitro* experiments are supplemented through the *in vivo* swap experiments with the HR-PKS of TENS and DMBS of Cox *et. al* (section 1.16). The swap experiments could successfully elucidate a part of the programming mechanism in the HR-PKS TENS.<sup>48,75–77</sup>

In different swap experiments with different domains such as KS, AT and DH of TENS with the ones of DMBS, no change in the polyketide chain was observed (Figure 15A-C), hence, the product was still pretenellin **67**. Therefore, it was concluded that the chain length seems not to be programmed by these three domains (Figure 15A-C).

In addition, in the case of both TENS and DMBS the PKS proteins are terminated with a non-ribosomal peptide synthetase (NRPS) module which catalyzes the transfer of the polyketide *via* an amide linkage to tyrosine. A Dieckmann cyclisation domain (DKC) then effects chain-release by intramolecular cyclisation - the NRPS also does not influence the programming of the PKS.<sup>79</sup> Further the *trans*-acting enoyl reductase proteins can also be exchanged without affecting programming.<sup>79</sup> However, programming and fidelity changes are observed in these systems when the KR and  $C$ -MeT domains are swapped, and when the *trans*-acting ER is absent (Figure 15D-J). This swap reprogrammed the PKS and resulted in two new products with different chain length and methylation pattern (Figure 15D-J).

In combination with other results, it can be suggested that the methylation pattern is controlled by the  $C$ -MeT- $\psi$ KR. Hence, the  $C$ -MeT domain is programmed on its own. Whereas the DMBS  $C$ -MeT seems only to recognize diketides, the TENS  $C$ -MeT can recognise both diketides and triketides.

Overall, the results hypothesize that the programming mechanism of HR-PKS is kinetically controlled. Therefore, the next level of programming is on the level of the single domains. They have on the one hand chemical selectivity, whereby they only act on substrates which are chemically competent and on the other hand a structural selectivity. The single domains can also be further divided into domains which play a part in the programming mechanism of the chain length or methylation pattern (*e.g.* KR, C-MeT and  $\psi$ KR) or don't play a part in the programming mechanism of the chain length and methylation pattern (*e.g.* KS, AT, DH). The last effect on the programming is not on the single domains, but on the interaction of the different domains with each other. In other words: Domain-domain interactions.



**Figure 15:** Rational domain swapping between TENS and DMBS. The trans-acting ER (not shown in the figure) and the NPRS are interchangeable with no programming effect.



## 1.18 Overall Project Aims

Diverse approaches have been explored to generate new polyketides by engineering polyketide synthases (PKS). Although it has proven possible to produce new compounds by designed PKS, engineering strategies have so-far failed to make polyketides available *via* widely applicable rules and protocols and engineering of HR-PKS is barely described.<sup>82</sup>

In this study, the aim is to engineer the ER domain of the HR-PKS Squalestatin tetraketide synthase (SQTKS) in a rational approach. SQTKS is one of the simplest iterative type I HR-PKS that has all the iterative  $\beta$ -modification domains present in an active state and in which a degree of programming occurs.

There is no existing structural information, for *complete* fungal iterative HR-PKS, such as SQTKS. However, if site-directed mutagenesis of these systems is to be attempted it will be necessary in a first step to obtain *useful* structural models of the specific catalytic domains.

With the generated protein models, different docking studies with modern modelling procedures will be used. The best-generated homology models will then be used as the basis for the design of different rational engineering experiments *in silico*. Afterwards, the applied *in silico* modifications shall be applied on the ER domain of SQTKS and effects of the mutagenesis verified through suitable *in vitro* assays.

Overall, the experiments should reveal if it is possible to engineer an HR-PKS. In addition, the effect of the different mutations on the intrinsic programming of the respective domain shall be investigated, especially how the chemical selectivity and kinetics of the different substrates are influenced by the changes introduced into the respective domain.

In the second project, the aim is to provide in parallel to the molecular-biological work, the structural-biological foundations and analysis for the domain swaps between TENS, DMBS and MILS. Sen Yin, Dr. Katherine Williams, Dr. Xiao-Long Yang and Dr. Steffen Friedrich performed the actual domain swaps and the molecular-biological work.

The purpose of the *in silico* structural analyses is to consider the effects of swaps on protein structure and to understand the effect of the swaps on the structural level. Additionally, the *in silico* analysis helps to clarify the influence of extrinsic and intrinsic programming factors. Subsequently, this analysis should provide the knowledge for targeted mutations and rational enzyme engineering.

## 2 Modelling Studies of the SQTCS ER Domain

### 2.1 Introduction

Protein structures provide valuable information which is important for understanding the molecular basis of their function and reengineering.<sup>83</sup> Currently, more than 130 000 experimental protein structures are deposited in the Protein Data Bank (PDB: Status 2019). Even so, the number of structurally characterized proteins are low compared to the number of known protein sequences. Hence, no structural information is available for the majority of protein sequences.<sup>83</sup> Methods such as X-ray crystallography, nuclear magnetic resonance spectroscopy (NMR) or cryo-electron microscopy (cryo-EM) are available to determine the structure of proteins and their complexes. However, all these techniques have the drawback that they are time-consuming processes without guaranteed success.<sup>40,83</sup>

Other techniques for the quick calculation of protein structures are therefore necessary. In order to solve the problem, it is understood that proteins are subjected to the process of evolution. Different results indicate that three-dimensional protein structure is evolutionarily more conserved than expected due to sequence conservation. Hence evolutionarily related proteins have similar sequences and naturally occurring homologous proteins have similar protein structures.<sup>84,85</sup> So-called homology modeling, or comparative modeling, is currently the only method that can reliably generate a 3D model of a protein (target) from its amino acid sequence.<sup>83,86,87</sup> Successful model building requires that at least one experimentally solved 3D structure (template), which has a significant amino acid sequence similarity to the target sequence, is available. In general homology modeling consist of five main steps:<sup>88-90</sup>

1. Identification of evolutionarily related proteins with experimentally solved structures that can be used as template(s) for modeling the target protein of interest;
2. Mapping corresponding residues of the target sequence and the template structure(s) by means of sequence alignment methods and manual adjustment;
3. Building a three-dimensional model on the basis of the alignment;
4. Evaluation of the quality of the resulting model;
5. Repetition of this procedure until a satisfactory model is obtained.

Potential structural templates are identified using a search for homologous proteins in available databases, for example the PDB databank. From the resulting list of possible candidate structures, a template structure is chosen based on its suitability according to various criteria such as the level of similarity between the query and template sequences, the experimental quality of the solved structures, the presence of ligands or cofactors, *etc.* In the best case, a wide range of the protein sequence should be matched by a single high-quality template.<sup>90</sup>

Estimating the accuracy of a protein structural model is an important step in the modelling process. The quality of the homology model is dependent on the quality of the sequence alignment and template structure. For example, the modelling could be complicated through the presence of alignment gaps, indicating a structural region, which is present in the target but not in the template. On the other side, this structure gap could be present in the template. Finally, the model quality can also decline with decreasing sequence identity.<sup>83,90</sup>

The three-dimensional models, which are built on possible templates, can then be further investigated on their quality. Differences in protein backbone structures are quantified by the root mean square deviation (RMSD) of the positions of alpha carbons ( $C_{\alpha}$ ). A model can be considered 'accurate' when its RMSD is within the spread of deviations observed for experimental structures displaying a similar sequence identity level as the target and template sequences.<sup>91-95</sup>

As a general rule, the core  $C_{\alpha}$  atoms of protein models sharing 50% sequence identity with their templates will deviate by  $\sim 1.0 \text{ \AA}$  RMSD from experimentally elucidated structures. Although the atomic coordinates of the three-dimensional model, for regions of the target protein aligned to the template, can be modeled on the basis of the information provided by template structure,<sup>62,96</sup> regions that are not aligned with a template (insertions/deletions) require specialized approaches.<sup>90,97-99</sup> Unaligned regions of the target such as loops that are modeled using *de novo* techniques will on average be less accurate than structurally conserved regions of the model on the basis of information derived directly from the template.<sup>90</sup>

If the percentage identity falls below  $\sim 30\%$  (so-called 'twilight zone'), the model quality estimation on the basis of sequence identity can become unreliable, as the relationship between sequence and structure similarity gets increasingly dispersed.<sup>90,100</sup> With decreasing sequence identity, alignment errors and the incorrect modeling of large insertions become the major source of inaccuracies.<sup>90</sup>

Overall, the two most common and large-scale sources of error in homology modeling arise from poor template selection and inaccuracies in target-template sequence alignment.<sup>101</sup> These problems can be minimized by the use of multiple templates, but the method is complicated by both the templates differing local structures around the gap and by the likelihood that a missing region in one experimental structure is also missing in other structures of the same protein family. Missing regions are most common in loops where high local flexibility increases the difficulty of resolving the region by structure-determination methods.<sup>87</sup>

Even bearing these problems in mind, homology modelling is a good method to determine possible structure of a protein of unknown structure. The only requirement for this method is the availability of adequate templates. Currently, completely automated methods for homology modelling and assessment are available online. These methods have been extensively validated and are being increasingly used to build useful protein models, which allow deeper understanding and engineering protein structures in cases where X-ray and NMR structures are not easily available. Such methods include, Swiss-Model, Modeller and ROSETTA.<sup>83 102 103</sup>

### **2.1.1 Swiss-Model**

Swiss-Model is an automated system for modelling the 3D structure of a protein from its amino acid sequence using homology modelling techniques.<sup>83,92,95</sup> Since its original development it has been continuously improved.<sup>90,104</sup> Today, Swiss-Model is one of the most widely used structure modelling web servers worldwide, with more than 0.9 million requests for protein models annually.

Like most modelling programs, the Swiss-Model workflow consist of four main steps that are involved in building a homology model of a given protein structure. First, identification of structural template(s) is performed. Therefore, BLAST and HHblits are used to identify templates.<sup>105,106</sup> In the second step, templates are stored in the Swiss-Model Template Library (SMTL), which is derived from the PDB (Protein Data Bank).<sup>105</sup> Afterwards the alignment of target sequence and template structure(s) is performed.<sup>105</sup> The GMQE score (Global Model Quality Estimation) is a quality estimation, which combines properties from the target–template alignment and the template search method.<sup>105</sup> The resulting GMQE score is expressed as a number between 0 and 1, reflecting the expected accuracy of a model built with that alignment and template and the coverage of the target. Numbers closer to 1.0 indicate higher reliability. The GMQE

score gives a first impression of the overall the quality of the model.<sup>105</sup> The last two steps consist of energy minimization and assessment of the overall quality of the generated model.<sup>83,92,95</sup>

Model quality estimation is an essential component of protein structure predictions, as the accuracy of a model determines its usefulness for practical applications. Swiss-Model provides model quality estimates based on a QMEAN potential, specifically re-parameterized for models built by Swiss-Model.<sup>83,92,95</sup> In addition, the accuracy of the Swiss-Model server is independently evaluated in comparison with other state-of-the-art methods by the CAMEO project. It is based on target sequences weekly pre-released by the Protein Data Bank (PDB).<sup>83,92,95</sup>

QMEAN is a composite estimator based on different geometrical properties and provides both global (*i.e.* for the entire structure) and local (*i.e.* per residue) absolute quality estimates based on one single model.<sup>107</sup> The QMEAN Z-score provides an estimate of the ‘degree of nativeness’ of the structural features observed in the model on a global scale. It indicates whether the QMEAN score of the model is comparable to what one would expect from experimental structures of similar size. QMEAN Z-scores around zero indicate good agreement between the model structure and experimental structures of similar size. Scores below -4.0 are an indication of models with low quality.<sup>90,104</sup>

The reason why Swiss-Model was used in these studies is the advantage that the template search and the validation of the model are already integrated in the software and therefore did not have to be done separately.

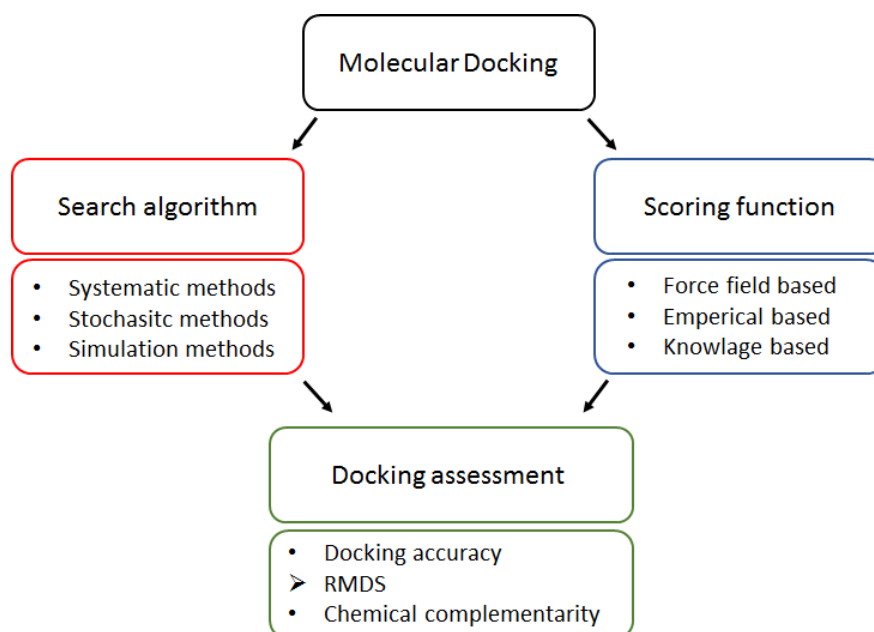
### **2.1.2 Molecular Docking (AutoDock Vina)**

Molecular docking is a method to predict the structure of intermolecular complexes formed between two or more constituent molecules.<sup>108</sup> In particular, the complex formed between a protein and a ligand such as a cofactor or substrate.

Many methods for molecular docking and virtual screening have been developed to date, such as *AutoDock*, *AutoDock Vina*, DOCK, Flex, Glide, GOLD, RosettaDock, SLIDE and Surflex.<sup>109–118</sup> Docking protocols can be described as a combination of a search algorithm and a scoring function (Scheme 21).<sup>108</sup> However, these methods introduce various approximations to simplify the problem; otherwise the calculation would include too many variables and would be nearly impossible to solve.

The search algorithm should allow several degrees of freedom of the protein–ligand system to be sampled sufficiently to include the true binding modes. Thereby, the

success of a docking algorithm in predicting a ligand binding pose is normally measured in terms of RMSD between the experimentally observed heavy-atom positions of the ligands and the one(s) predicted by the algorithm.<sup>119</sup>



**Scheme 21:** Overview over the docking process and available methods.

Different levels of approximation can be performed in the search algorithm. The earliest and most basic one is the rigid-body approximation.<sup>119,120</sup> This is valid for rigid ligands in which one or a limited set of ligand conformations is possible. However, most ligands are highly flexible. A more common approach nowadays is to model the ligand flexibility while assuming a rigid protein receptor.<sup>119,121</sup> Three general categories of algorithms are devised to treat ligand flexibility: systematic methods; random or stochastic methods; and simulation methods. Furthermore, three basic types of methods are based on random algorithms: Monte Carlo methods (MC); Genetic Algorithm methods (GA); and Tabu Search methods. Of these, the genetic algorithms apply ideas derived from genetics and the theory of biological evolution to docking.<sup>119</sup>

GA methods start from an initial population of different conformations of the ligand with respect to the protein. Each conformation is defined by a set of state variables (defined as genes) that describe aspects like the translation, orientation, and conformation of the ligand in relation to the protein receptor. The full set of the ligand's state variables is defined as the genotype, whereas the atomic coordinates refer to the phenotype. Genetic operators (mutations, crossovers, and migrations) are applied to the population to sample the conformational space, until a final population that optimizes a predefined fitness

function is reached.<sup>119</sup> AutoDock and AutoDock Vina include this type of search algorithm, for example.<sup>110,122,123</sup>

The scoring function should represent the thermodynamics of interaction of the protein–ligand system adequately so as to distinguish the true binding modes from all the others explored, and to rank them accordingly.<sup>119</sup> Different assumptions must be made for the searching algorithms and the scoring function to minimize the computational intensity. However, these assumptions result in less accuracy concerning the docking procedure. For this reason, the lack of a suitable scoring function both in terms of speed and accuracy is the major bottleneck in docking.<sup>119,124</sup>

Scoring functions can be divided into three major classes: force field-based, empirical, and knowledge- based scoring functions.<sup>119</sup> Standard force fields generally quantify the sum of two energies: the interaction energy between the receptor and the ligand, as well as the internal energy of the ligand.<sup>119</sup> The energies are normally accounted through a combination of van der Waals with an electrostatic energy terms. A Lennard-Jones potential is used to describe the van der Waals energy term, whereas the electrostatic term is given by a Coulombic formulation with a distance-dependent dielectric function that lessens the contribution from charge–charge interactions.<sup>119</sup> The limitations of force field scoring functions include the absence of solvation and entropic terms, and the inaccurate treatment of the long-range effects involved in binding. Several force field scoring functions such as D-Score, G-Score, Gold-Score or AMBER are known.<sup>115,125–127</sup>

As mentioned previously, AutoDock is one of the many protein–ligand docking programs available. The first version was described in 1990 and has been improved during the following years.<sup>109,128</sup> The AutoDock scoring function is based on the molecular mechanics force field AMBER with two additional terms: one to model the desolvation free energy change on binding, which is based on atomic solvation parameters and one empirical term to model the loss of conformational entropy on binding.<sup>122,129,130</sup> Furthermore AutoDock has been used in the discovery of several drugs.<sup>131,132</sup> In addition, it is the most cited docking program in previous studies.<sup>119,128,133</sup>

AutoDock Vina was developed in 2010 by the same group as AutoDock, intending to improve accuracy and performance.<sup>128</sup> AutoDock Vina achieves an approximately two orders of magnitude increase in speed compared to the molecular docking software previously developed in AutoDock 4, while also significantly improving the accuracy of the binding mode predictions. Furthermore, multi-core processors are now common on desktop PCs which also accelerates the process.

AutoDock Vina automatically calculates the grid maps and clusters the results in a way transparent to the user.<sup>110</sup>

Two main criteria are considered during the evaluation of the results of docking experiments: First, how well does the postulated binding mode agree with known structural data; and second, how well did the docking rank the ligands or rather, if the method's scoring function is designed to predict binding affinities, how high is the agreement.<sup>122</sup>

The first evaluation method is usually the RMSD between the docked position and the crystallographically observed binding position of the ligand, where success is typically regarded as being less than 2 Å.<sup>119,122</sup>

In addition, if stochastic methods are performed, the experiment could be run several times with different initial conditions. The similarity of the predicted binding modes can be assessed by computing a matrix of pairwise RMSD values, and clustering docked conformations according to an RMSD threshold, typically 2 Å. If all of the docking-clusters fall into one family, this indicates that the search parameters were sufficient for each docking to converge.<sup>119,122</sup>

Further, if the docked conformation with the lowest energy would always correspond to the crystallographically observed binding mode, it would be concluded that there are no bad contacts in the crystal structure. This is not always the case, and sometimes a different binding mode is observed significantly more often than the lowest energy-binding mode. Furthermore, current docking methods will tend to find the binding mode with the lowest possible interaction energy for a given ligand: this score does not necessarily indicate whether the ligand even binds.<sup>119,122</sup>

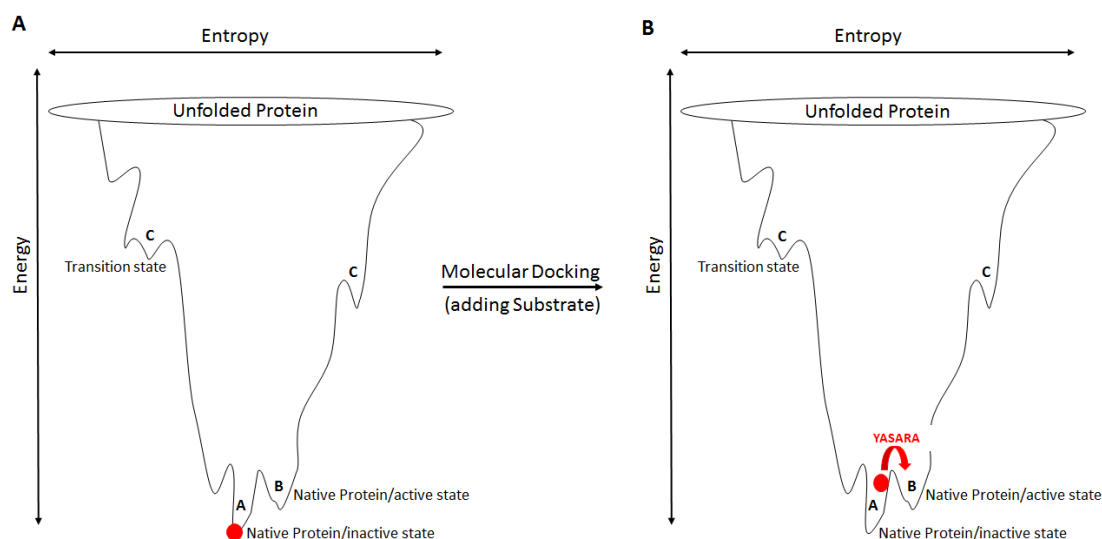
### 2.1.3 YASARA

Molecular dynamics (MD) is a computer simulation method for studying the physical movements of atoms and molecules. It is widely used as a powerful simulation method in many fields of molecular modelling. In the context of docking, by moving each atom separately in the field of the rest of the atoms, MD simulation represents the flexibility of *both the ligand and protein* more effectively than other algorithms.<sup>134</sup> However, the disadvantage of MD simulations is that they progress in very small steps and thus have difficulties in stepping over high energy conformational barriers, which may lead to inadequate sampling. On the other hand, MD simulations are often efficient at local



optimization (Figure 16). Thus, the most common strategy for simulated ligand docking involves docking experiments with another software (such as AutoDock Vina), followed by MD simulations for further refinement after docking (Figure 16).<sup>134</sup> Different software packages are available for this approach, such as ORCAR, Lee-Server, ROSETTA, Undertaker or YASARA.<sup>135</sup>

YASARA, is a software package that runs molecular dynamics simulations of models in explicit solvent, using a new partly knowledge-based all-atom force field derived from AMBER, the parameters of which have been optimized to minimize the damage done to protein crystal structures. The YASARA force field addresses these issues by combining the AMBER all-atom force field equation with multi-dimensional knowledge-based torsional potentials and with a consistent set of force field parameters to maximize the accuracy.<sup>135,136</sup> This is achieved by making a random change to one or more parameters (*e.g.* a certain van der Waals radius, a charge, or the weight of a knowledge-based potential). To ensure that all forces responsible for the experimentally observed structure are considered, minimizations are done in crystal space, using complete unit cells. As a result, one obtains a force field that has stable energy minima as close as possible to native structures. This is essentially equivalent to a force field that moves models closer to native structures during a simulation.<sup>135,136</sup>



**Figure 16:** Schematic example for the applications of YASARA. **A;** Native protein in the inactive state, hence without a substrate. Positioned in the local minimum A. **B;** Protein obtained a higher energy level, after the substrate was docked inside through a molecular docking software package (still located in the local minimum A). Indicated with the red arrow: YASARA would refine the model and “move” the protein with the substrate into the local minimum B. Which also represents a native state, but an active conformation.

## 2.2 Previous Studies

Iterative highly reducing PKS (HR-PKS) show distinct and often complex programming. The level of complexity of fungal polyketides synthesized by HR-PKS, like SQTKS, is on a similar level of structural complexity as polyketides which are synthesized by bacterial modular PKS. However, through the use of a single iterative PKS module, such fungal PKS systems are highly efficient and the programming is much more complicated (Chapter 1, section 1.17). Furthermore, through a variety of variables (amino acid sequence, structure of the modules, module-module interactions, structure of the substrate, *etc.*) the mechanisms of programming of these types of PKS are cryptic. For modular systems, the number of modules, and the content of  $\beta$ -processing domains in each module, controls the program. Hence the presence or absence of  $\beta$ -processing domains within modules (*i.e.* extent of functionalization at each  $\beta$ -position); and individual domain selectivity, for example by the AT domains, which controls the selection of extender units (*i.e.* the methylation pattern). Furthermore, modular PKS can control the stereochemistry of each  $\alpha$ - and  $\beta$ -position, and this is usually controlled by the selectivity of the KR domain ( $\beta$ -hydroxyl positions) and the ER domain ( $\alpha$ -alkyl positions).

In the previous studies of Cox *et. al.* the fundamental work for the programming of SQTKS could be laid and serves here as the fundamental basis for enzyme engineering (Chapter 1, section 1.12).<sup>58,137</sup> In addition, the expression and analysis of the ER domain is based on primary works of Cox *et. al.*<sup>58</sup>

### ER Domain of SQTKS

The biosynthesis and programming of SQTKS was elucidated before (Chapter 1, Section 1.12, 1.17), hence will here not further be discussed. In previous studies, it was shown that the isolated ER domain of SQTKS possesses a broad substrate selectivity *in vitro*. It accepts both natural and unnatural enoyl species as substrates. The substrates included likely intermediates as well as others which are improbable intermediates (including *Z*-isomers) and those which cannot be intermediates (*e.g.* ethyl-substituted, incorrectly methylated, and compounds with odd carbon chain-lengths, Figure 14, Section 1.17)<sup>58,137</sup>

Further, Cox *et. al.* could show in *in vitro* assays that the isolated SQTKS ER domain is not able to control the stereochemistry at the  $\alpha$ -position. Hence, all substrates were produced as  $\alpha$ -racemates.<sup>58</sup> Neither possible product enantiomer was preferred

because reprotonation of an enol(ate) intermediate could occur from either face (Scheme 6). In contrast, the stereochemical outcome at the  $\beta$ -position could be predicted and the transfer of hydride from NADPH was shown to be highly stereoselective, indicating a rigid arrangement of the cofactor and substrate in the ER active site. Other studies showed that multi-domain fragments of SQTks, in which specific fragments from the DH-KR domains are included, are better able to control the stereochemistry of the  $\alpha$ -position compared to isolated domains, suggesting that reprotonation of the  $\alpha$ -position is probably achieved by a water molecule rather than a specifically located amino acid residue.<sup>56</sup>

In addition, a first model structure of the ER domain was obtained using the crystal structure of mammalian fatty acid synthase (mFAS, 2vz9) as the template. The 3D model of SQTks ER was consistent with numerous experimental observations. For example, the structural domain organization is consistent with that observed for other PKS and FAS ER proteins and docking of NADPH showed interactions with known cofactor binding residues and the correct 4'-hydrogen of the cofactor was exposed for reaction. Likewise, docking of substrate pantetheines gave structures consistent with the observed stereochemistry of reduction at the substrate  $\beta$ -carbon. Finally, it was shown that the ER may sequester its final substrate to prevent further chain extension.<sup>58</sup>

### 2.3 Aims of the Project

There is no existing structural information, for *complete* fungal iterative HR-PKS, such as SQTks. However, if site-directed mutagenesis of these systems is to be attempted it will be necessary to obtain *useful* structural model of the specific catalytic domain. The enzyme domain, which is of interest for engineering and thus modelling, is the ER domain of SQTks. Information on isolated proteins would be useful, but it would also be useful to generate multi-domain models, which could help show how the domains interact. The modeling should be done through homology modeling.

As mentioned before, a critical requirement for homology modeling is the availability of a suitable template. Sequence alignments show that the mammalian fatty acid synthase (mFAS) has an appropriate similarity and the domains are arranged in the same order (Chapter 1.10). However, the mFAS crystal structure was obtained at relatively low resolution (3.2 Å) and parts of the structure are missing.<sup>25</sup> In addition, mFAS has no functional C-MeT domain. Hence, the first steps will be to find appropriate

templates for the specific domains and use these for homology modelling. After the modeling the quality of the modelled structure should be validated.

With the generated protein model in hand, different docking studies with AutoDock Vina will be performed. On the one hand, this will be done to obtain a substrate and cofactor docked model of the ER domain, which could be used as a template for the mutagenesis *in silico*. On the other hand, this will be done in order to determine the range of values for different parameters, such as C-H distances and carbonyl position, *etc.* which could result in productive or non-productive conversion of the substrate by the ER domain. The best-generated homology model will then be used as the basis for the design and validation of different rational engineering experiments *in silico* (Chapter 3).

## 2.4 *In Silico* Studies of the SQTCS ER Domain

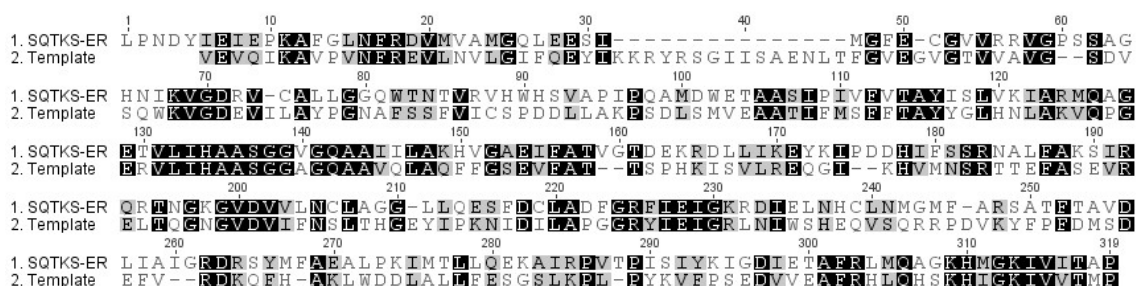
The first domain to be modelled was the enoyl reductase from SQTCS. However, before homology modelling could be done the domain boundaries of the ER domain of SQTCS should be determined. Therefore, the sequence of SQTCS was examined with BLAST (basic local alignment search tool).<sup>138</sup> Subsequently, a conserved domain search (CD-Search) was performed to determine the domain boundaries.<sup>139-142</sup> The search determined the boundaries for the ER domain from L1908 as start point and P2208 as an end point of the ER domain.

Afterwards, homology modeling was done using Swiss-Model.<sup>83,92,95,105</sup> This selected 50 possible templates including Mycoerolic acid synthase (PDB 5bp4),<sup>24,143</sup> CurK (PDB 5dp1),<sup>144</sup> JamJ (PDB 5doz1),<sup>144</sup> mFAS (PDB 2vz9)<sup>25</sup> and CurF (PDB 5dp2).<sup>144</sup> For the selection as template the GMQE score (section 2.1.1), sequence identity and the resolution of the templates were considered and compared. A low GMQE score, level of sequence identity or resolution of the template, would result in a low quality model. From the proposed templates, the CurF ER from the curacin A biosynthetic pathway (PDB 5dp2) was selected as the most suitable template for the modelling.<sup>144</sup> The CurF PKS module is an unusual modular cyanobacterial PKS, which contains functional  $\beta$ -processing domains (Chapter 1, section 1.8).<sup>145</sup>

The CurF ER was chosen for various reasons. On the one hand, the structures have high sequence identity of 37 %, which is also displayed in the sequence alignment of the template with the SQTCS ER domain (Figure 17). Further, the template had a high GMQE score of 0.71. The final reason, which was an important difference to the other possible templates, was the X-Ray resolution. The other templates were obtained at lower

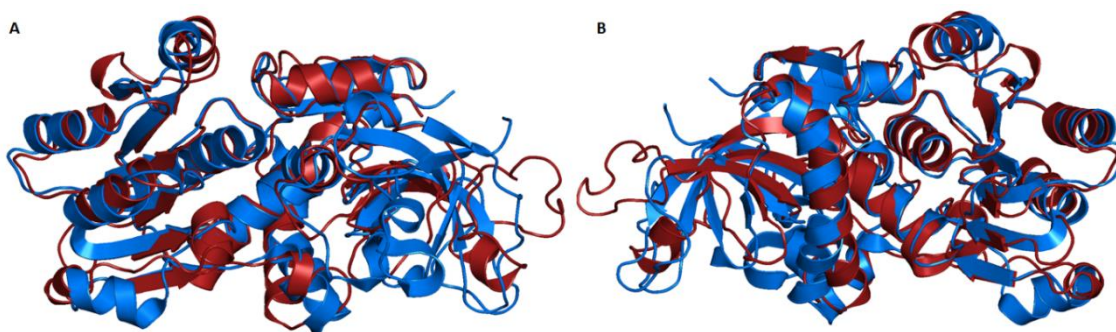
X-ray resolutions between 1.8 - 3.8 Å. In contrast, the X-Ray resolution of CurF was 1.0 Å. In addition, the PKS CurJ might be closer to the actual structure of SQTCS, than mFAS.

Overall, a structure model of the ER of SQTCS with a QMEAN (Chapter 2.1.1) value of -2.37 was obtained after the modeling (Figure 19), which indicates, that the quality of the generated structure model was good enough to proceed.<sup>91-95</sup>



**Figure 17:** Sequence alignment of the ER sequence of SQTCS with the CurF ER template sequence (5dp2).

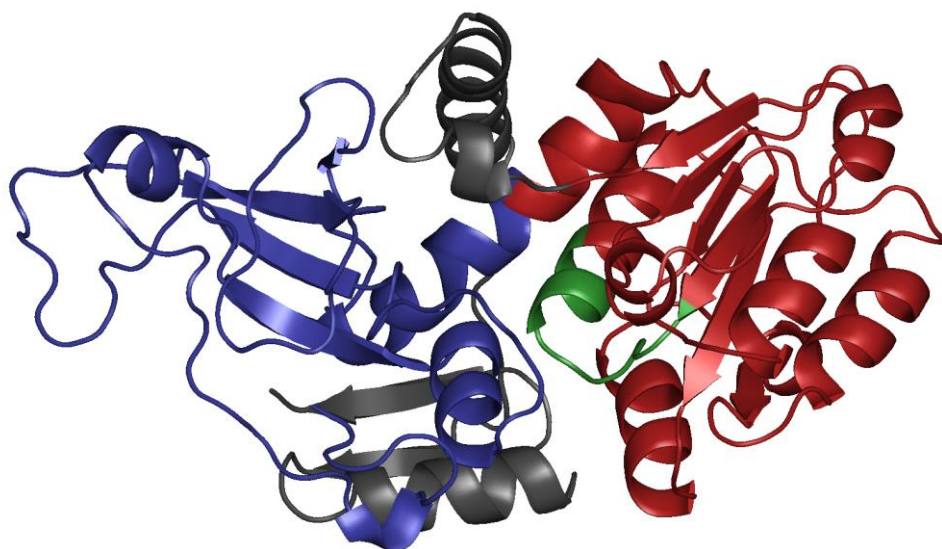
In the next step, the homology model and the template were aligned in PyMOL (Figure 18). The alignment resulted in C $\alpha$ -RMSD value of 1.34 Å. This C $\alpha$ -RMSD value is good. Even so, some structural differences in the loops of both structures were observed. The core  $\alpha$ -helices and  $\beta$ -sheets overlay very well (Figure 18); the crystallization of loop regions is not always very accurate, because of their high flexibility.



**Figure 18:** Alignment of the SQTCS ER model (red) with the template CurF (PDB: 5dp2; blue) displayed in PyMOL. **A**, Front view; **B**, Back view.

The generated model itself consists of three main structural features (Figure 19).<sup>58</sup> The N-terminus (L1908-I2001, blue in Figure 19) forms a globular domain, which is involved in contacting the acyl-pantetheine substrate. The central sequence V2002-V2144 (red/green, Figure 19) forms the cofactor-binding domain and includes a Rossmann-fold (green, Figure 19). Finally, the C-terminal sequence (D2145-P2208 grey, Figure 19) forms

a link between the cofactor and substrate binding domains as well as a part of a capping region above the active site.

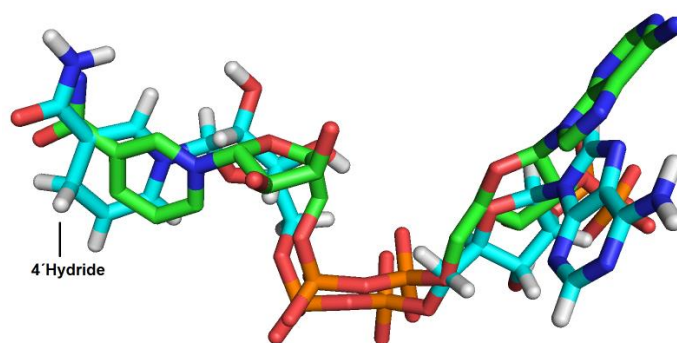


**Figure 19:** Model of the ER domain of the SQTGS displayed in PyMOL. Shown are the structural features: Red, cofactor binding domain; blue, substrate binding domain; green, Rossmann fold; grey, C-terminal sequence.

## 2.5 Integration of the cofactor into the modelled ER domain of SQTGS

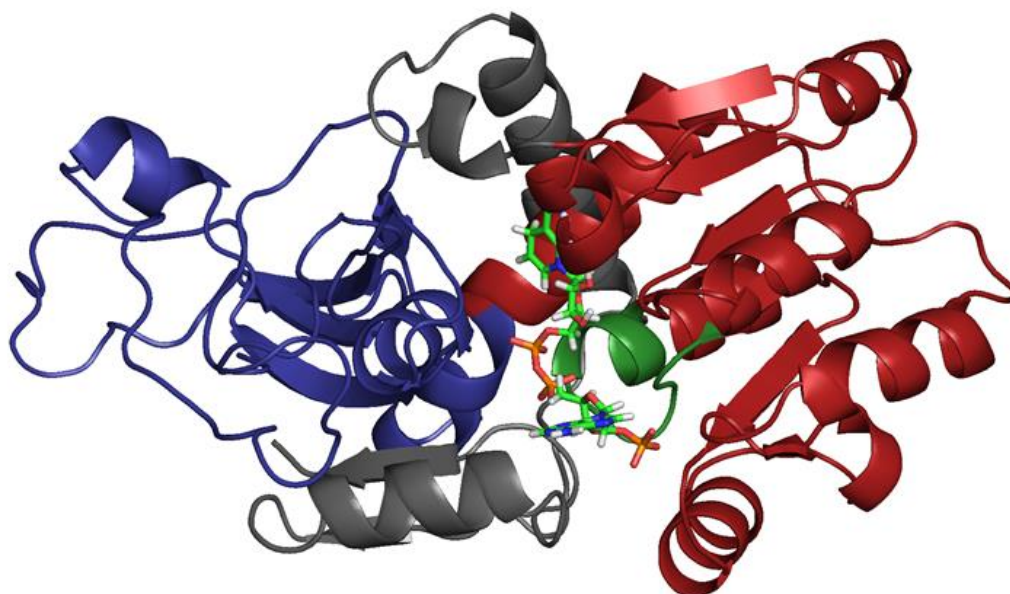
The next step was to add the cofactor NADPH **11** into the active side of the ER domain (Figure 21). This was done through the alignment of the homology model with the template in PyMOL (Figure 18).<sup>146,147</sup> The cofactor was extracted from the template and then manually integrated into the structural model of the SQTGS ER. Afterwards, the generated ER domain plus cofactor was minimized in YASARA, to refine the protein-cofactor interaction (section 2.1.3, Figure 16).<sup>135</sup> Then the extracted and refined cofactor **11** was aligned with the cofactor **11** from the template (Figure 20). The alignment resulted in a RMSD of 2.01 Å. The highest variability in the alignment is at the purine moiety. However, the nicotinamide moiety and the diphosphate “bridge” overlay well. Hence, the cofactor in the ER SQTGS model should have the correct position and orientation. Further, the nicotinamide 4'-*pro-R* hydrogen is exposed for reaction in agreement with experimental results.<sup>58</sup>





**Figure 20:** Alignment of the cofactor **11** from the SQTKS ER model (green) with the cofactor from the CurF (blue)

Furthermore, NADPH **11** contacts one side of the tunnel made up from the N-terminal domain of the ER, making specific contacts with residues S2072 and K2055, G2029 (diphosphate), I2119 and V2144 (nicotinamide). All these residues are conserved in other PKS and mFAS ER domains (Sequence alignment of the ER, Table 4). In addition, the interaction of the conserved binding NADPH binding motif “HAASGGVGQA” with the NADPH cofactor **11** was observed. Overall, the QMEAN and the structural features of the model generated for the SQTKS ER indicate that the quality should be suitable for further *in silico* studies.

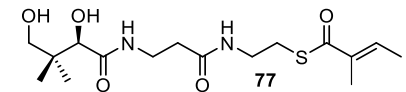
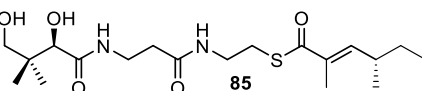
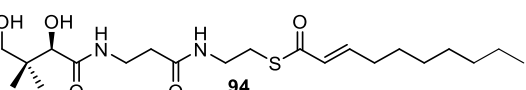
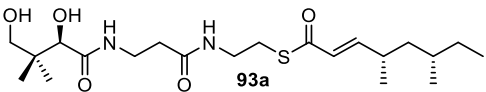


**Figure 21:** Model of the SQTKS ER domain displayed in PyMOL. Shown are the structural features: red, cofactor binding domain; blue, substrate binding domain; green, Rossmann fold; grey, C-terminal sequence. The cofactor **11** as sticks with the nicotinamide near the top of the image.

## 2.6 Development of a Substrate and Cofactor Docked Model for the ER of SQTKS

In the next step, different substrates were docked into the active pocket of the ER SQTKS (table 1). On the one hand, this was done to obtain a substrate and cofactor docked model of the ER domain, which could be used as a template for the mutagenesis *in silico*. On the other hand this was done in order to determine the range of values for different parameters, such as C-H distances and carbonyl position, *etc.* which could be correlated with productive or non-productive conversion of the substrate by the ER domain. Hence, in the following *in silico* experiments we aimed to correlate results from *in vitro* kinetic experiments with the *in silico* docking results to attempt to determine a range of *productive* and *non-productive* geometric substrate poses from which predictive geometric parameters could be extracted.<sup>58,137</sup> Therefore data was generated which included three-dimensional parameters of the substrates known to be active or inactive with the SQTKS ER.<sup>58</sup>

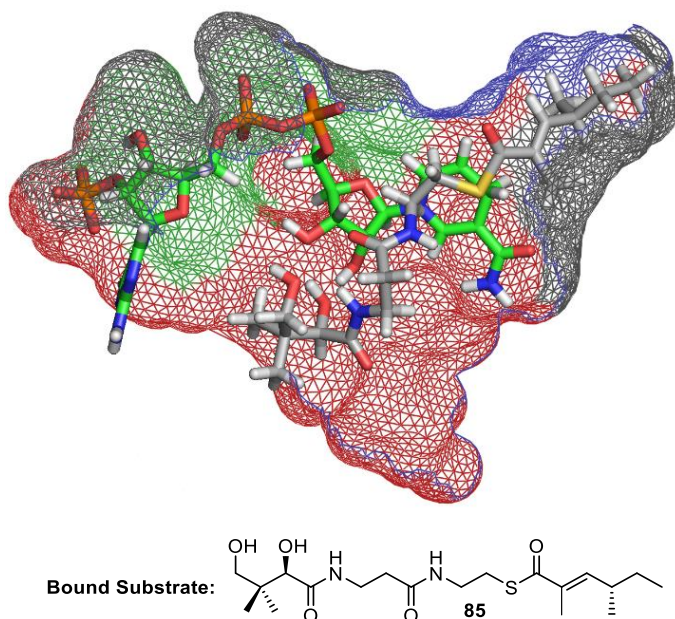
**Table 1:** Model substrates for the determination of the validation parameters (ER SQTKS).

	Substrate	Observed Substrate Specificity ( $k_{cat}/K_M$ ) <sup>58</sup>
A	 77	Okay (12.0)
B	 85	Good (119.0)
C	 94	Okay (23.6)
C	 93a	Bad (0.0)

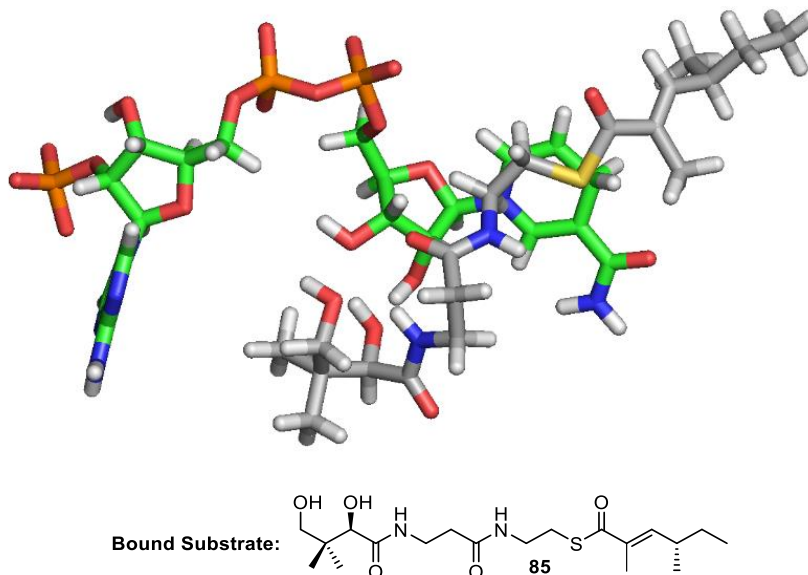
The docking was performed by manually overlaying the substrate in the SQTKS ER active site using PyMOL. This was done to simplify the docking in the next step. This method minimized the so-called Grid Box, which has a critical role in the speed of the docking calculations. Molecular docking was done using AutoDock Vina (Chapter 2, section 2.1.2).<sup>110,147,148</sup> The model was then refined by YASARA. The visualization of the different models, after the refinement step was always done in PyMOL. An example is shown with squalestatin triketide pantetheine **85** (*E*-4*S*-2,4-dimethylhex-2-enoyl-



pantetheine, Figure 22 and Figure 23) which is known to be both a natural substrate and a good substrate *in vitro*.



**Figure 22:** Active site of the ER domain with a mesh surface. In green the cofactor NADPH 11 and in white the modeled substrate squalestatin triketide pantetheine (*E*-4S-2,4-dimethylhex-2-enoyl-pantetheine) 85.



**Figure 23:** Active site of the ER domain. In green the Cofactor NADPH 11 and in white the modeled substrate squalestatin triketide pantetheine (*E*-4S-2,4-dimethylhex-2-enoyl-pantetheine) 85.

Since no structural data for the SQTGS ER domain was available, it was not possible to perform a direct validation of the molecular docking of the substrates. For example, the RMSD between the docked substrate and the structural available substrate could not be compared. In addition, validation of the docked substrates according to the lowest energy

conformation is not very reliable (section 2.1.2). The last validation option that could be used would be a cluster of RMSD analysis. Even so, this method is very time consuming.

Hence, a visual validation of the docking was performed based on different criteria, before the values for the different parameters were determined. The criteria of the visual validation included amongst other things the orientation of the substrate towards the cofactor **11**.




In the following, the determination of the values for the C-H bond, distance to the active hydride, *etc.* are shown for squalestatin triketide pantetheine **85** (*E*-4*S*-2,4-dimethylhex-2-enoyl-pantetheine, Fig 24 and 25) and **93a** squalestatin tetraketide (6*S*,4*S*-2*E*-dimethyloct-2-enoylpantetheine, Fig 26 and 27) which is known not to be a substrate. The best docking result, which could be obtained with optimization of different parameters of the Grid Box in the docking procedure with AutoDock Vina (section 2.1.2), are shown in Figures 24-27.

In the model of the squalestatin triketide pantetheine **85**, the catalytic nicotinamide moiety of the NADPH cofactor **11** is located inside the ER-domain. NADPH **11** is in contact with its binding site, consisting of highly conserved residues (Sequence alignment of the ER, Table 2). The squalestatin triketide pantetheine **85** extends into the protein. The pantetheine part of the substrate extends parallel to the adenine diphosphate locating the thiolester and the  $\beta$ -carbon adjacent to the nicotinamide. The  $\alpha/\beta$ -unsaturated carbonyl of the substrate adopts an *s-cis* conformation (Figure 25), which places the reactive  $\beta$ -carbon 3.4 Å away from the cofactor's correct/observed reactive 4'-*pro-R* hydrogen (Fig 24). The Burgi-Dunitz angle substrate-cofactor complex is 61.7° and the dihedral angle 46.6° (table 2). The Burgi-Dunitz angle and the dihedral angle might differ from the optimal angles from literature (*e.g.* Burgi-Dunitz angle: 107°),<sup>149,150</sup> however, enzymes are flexible systems and in crystals and the generated model only snapshot of a certain state is displayed. Hence, in nature the enzyme might change its structural conformation until the right pose with the necessary angle is occupied. In addition, other angles than the optimal angle of 107° have been observed in enzymatic reactions, since the angle in the enzymatic conversion in the SQTGS ER might be also different.<sup>151</sup>

**Table 2:** Sequence alignment of the ER (wild type) domain from SQTKS with other fungal ER domains.

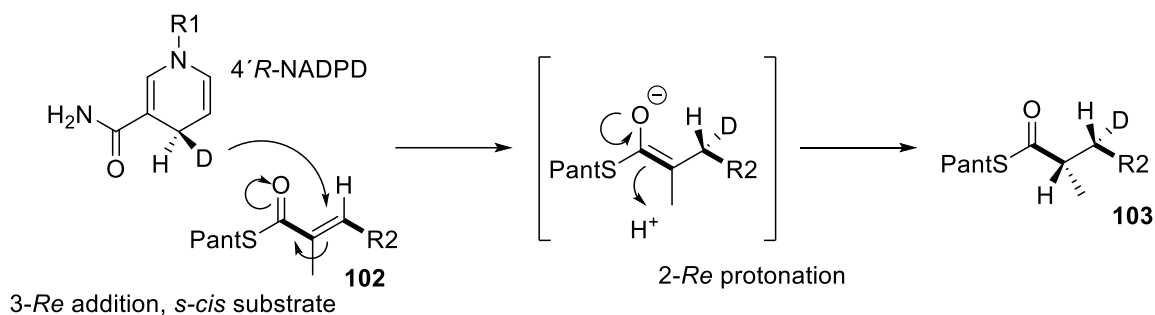
			SSS	S S	S S		
Fumonisin	(1809)	NFRD	VLLAMG	IVEANN	LGIG	LEGS	VITDVGAGV----
Zearalenone	(1688)	NFRD	VMASMA	LVPVK--	GLGQ	EASG	IVLRTGRDA----
Alternapyrone	(1185)	NFKD	VLVALG	NLAEN-K-	LGVD	ASGI	VTRVGS AV----
<b>Squalestatin</b>	(1923)	NFRD	VMAMG	QLEES--	IMGEE	CG	GVVRRVGPSS--
Asperfuranone	(1853)	NFRD	VMAMG	QLKER-V-	MGLE	CAGV	ITRVGAE A-A-AQGF
							CSC CC
Fumonisin	(1853)	YLDD	NCFSTR	ITMSAM	CAKIF	SFSL	SYEEAA
Zearalenone	(1706)	TLDM	GTHATV	MRADHR	VTVKI	PDAM	SFEEAA
Alternapyrone	(1927)	TASCD	TFATY	VRFPAK	GAIGV	PTGM	SFEEAA
<b>Squalestatin</b>	(1966)	ALLG	GQWTN	TVRVH	WHSV	APIP	QAMD
Asperfuranone	(1897)	ALLL	GPFS	SRARV	SWHGV	ASMP	AGMG
							CCCCCCCC
Fumonisin	(1901)	GLQS	GQSV	LIHSAC	GGIG	IAAI	NV
Zearalenone	(1758)	KLQR	GQSV	LIHAAA	GGVQ	QAAL	QLAN-
Alternapyrone	(1975)	GIVAG	EKVL	IHAAA	GGVQ	QAAL	IMTAQ-
<b>Squalestatin</b>	(2015)	RMQA	GETV	LIHAAS	GGVQ	QAAL	IILAK-
Asperfuranone	(1945)	RLSQ	GQTV	LIHAAA	GGVQ	QAAL	VILAKE
							CCCC C
Fumonisin	(1954)	TFNI	PRASI	FNSR	DTSF	FRED	VLAH
Zearalenone	(1810)	TYQV	SEDHI	FNSR	DASF	AKGI	MIRV
Alternapyrone	(2027)	QYGI	PEDHI	FSSR	DTSF	VKGV	LRAT
<b>Squalestatin</b>	(2067)	EYKI	PDDHI	FSSRN	ALFA	KSI	RQRT
Asperfuranone	(1998)	EYGI	PDDHI	FNSRD	SSE	FAPA	ALAA
							CC
Fumonisin	(2006)	CVAP	YGKM	LEIG	KRDF	IGKAK	LSMDI
Zearalenone	(1862)	CLAT	FCTF	VEI	GLRDI	TNNM	LLDM
Alternapyrone	(2080)	CLAK	FGRF	LEIG	KADL	FANT	GLDM
<b>Squalestatin</b>	(2119)	CLAD	FGRF	LEIG	KRDI	ELNH	CLNM
Asperfuranone	(2038)	VLAP	FGHF	VEI	GLRDL	EQNS	LLEM
							CSSS
Fumonisin	(2006)	CVAP	YGKM	LEIG	KRDF	IGKAK	LSMDI
Zearalenone	(1862)	CLAT	FCTF	VEI	GLRDI	TNNM	LLDM
Alternapyrone	(2080)	CLAK	FGRF	LEIG	KADL	FANT	GLDM
<b>Squalestatin</b>	(2119)	CLAD	FGRF	LEIG	KRDI	ELNH	CLNM
Asperfuranone	(2038)	VLAP	FGHF	VEI	GLRDL	EQNS	LLEM
							CSSS
Fumonisin	(2056)	CHPL	LTRT	VQML	EAGHI	KPIA	PR
Zearalenone	(1915)	LGDI	LEEV	FKLL	GGGIL	QTPS	EM
Alternapyrone	(2133)	AVAL	WHDT	TAKMI	HDGAI	KPIA	PL
<b>Squalestatin</b>	(2157)	FAEA	LPKI	MILL	QEKAI	RPV	TP
Asperfuranone	(2091)	AHRV	LSEL	ARLAG	QGIV	KPVH	EV

**Legend:**

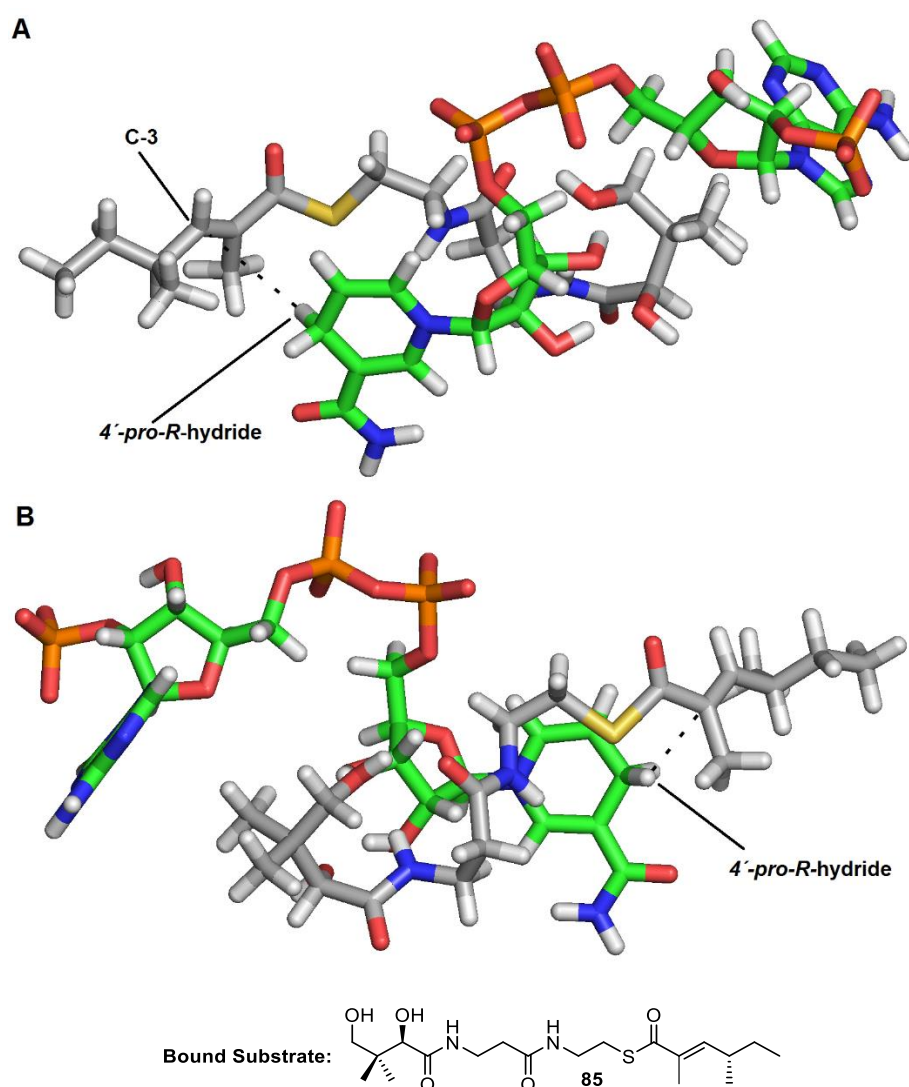
	Identical within HR-PKS ER domains		Highly conserved within HR-PKS ER domains
	Mutagenesis residue	C = Cofactor binding; S = Substrate binding	

In addition, in Scheme 20 the stereochemical course for the reduction of **79** catalyzed by the isolated SQTKS ER domain is shown (Chapter 1.11).<sup>58</sup> Previous studies demonstrated that for the highly stereoselective transfer of the 4'-*pro-R* hydrogen of NADPH **11**, the cofactor must be rigidly located in the active site. Furthermore, the transfer of the hydride to the 3-carbon of the substrate is also highly stereoselective, indicating that the substrate must take a single conformation relative to NADPH, which was in the previous studies

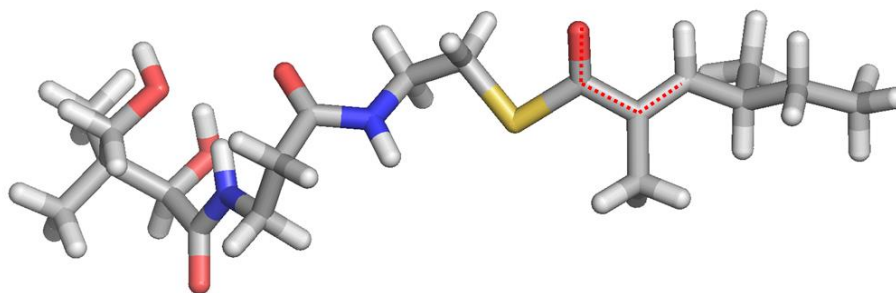
determined for the substrate as the *s-cis* conformation.<sup>58</sup> Our calculations for the good substrate are similar to the results found in previous studies.<sup>58</sup>



**Scheme 22:** Stereochemical course of the reduction catalyzed by the SQTCS isolated ER domain.

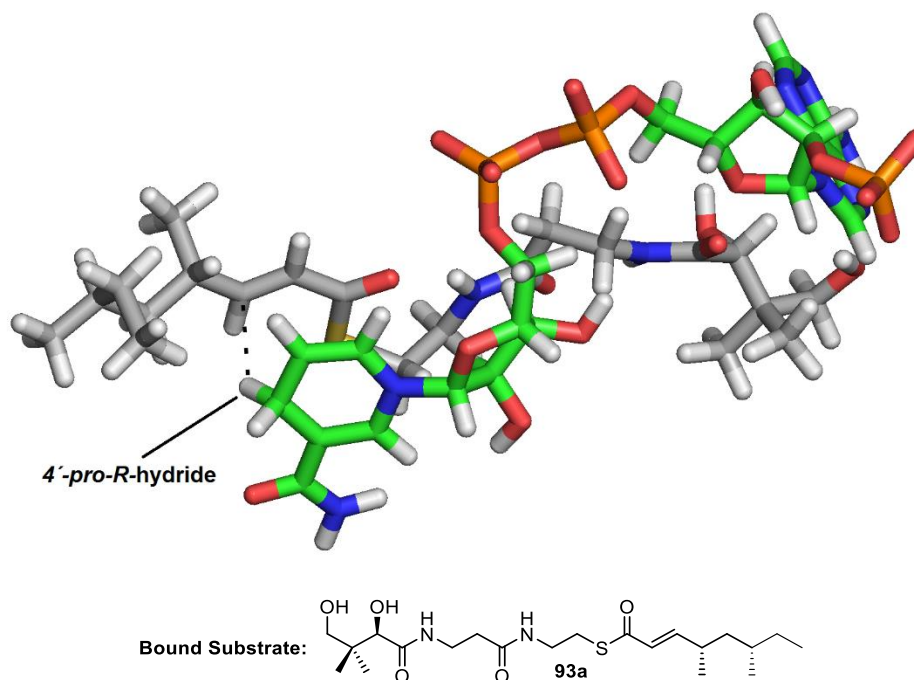


**Figure 24:** Active site of the ER domain. Cofactor NADPH **11** (green) and the modeled substrate squalastatin tetraketide (4*S*-2*E*-dimethylhex-2-enoylpantetheine) **85** (white). The hydride at the cofactor are marked. **A**, Back view; **B**, Front view.

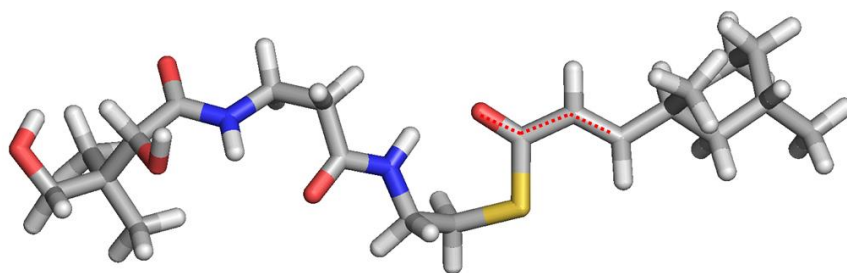


**Figure 25:** Modeled substrate squalestatin tetraketide (6S,4S-2E-dimethyloct-2-enoylpantetheine) **85** (white). Shown in red the *s-cis* geometry of the  $\alpha/\beta$ -unsaturated carbonyl.

In the model of the squalestatin tetraketide pantetheine **93a** the catalytic nicotinamide moiety of the NADPH cofactor **11** is also located inside the ER-domain. NADPH **11** is in contact with its conserved binding site, consisting of highly conserved residues (see Sequence alignment of the ER, Table 2). The squalestatin tetraketide pantetheine **93a** extends with the side-chain into the protein. The pantetheine part of the substrate extends parallel to the adenine diphosphate locating the thiolester and the  $\beta$ -carbon adjacent to the nicotinamide. The  $\alpha/\beta$ -unsaturated carbonyl of the substrate adopts a different pose to the triketide **85**. An *s-trans* conformation is observed (Figure 27), which places the reactive  $\beta$ -carbon 3.2 Å away from the cofactor's reactive 4'-*pro-R* hydrogen (Fig 26). The Burgi-Duniz angle substrate-cofactor complex is 100° and the dihedral angle at 72.3° is also higher than for the triketide **85** (table 2).



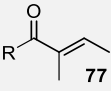
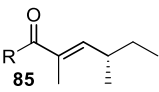
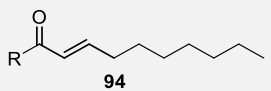
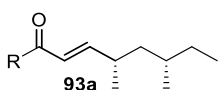
**Figure 26:** Active site of the ER domain. Cofactor NADPH **11** (green) and the modeled inhibitor squalestatin tetraketide (6S,4S-2E-dimethyloct-2-enoylpantetheine) **93a** (white). The hydride at the cofactor are marked.



**Figure 27:** Modeled inhibitor squalastatin tetraketide (6S,4S-2E-dimethyloct-2-enoylpantetheine) **93a** (white). Shown in red the *s-trans* geometry of the  $\alpha/\beta$ -unsaturated carbonyl.

These parameters: distance of the C-2 position to the NADPH hydride; orientation of the substrate towards the NADPH; the geometry of the  $\alpha/\beta$ -unsaturated carbonyl; the Burgi-Dunitz angle and the dihedral angle were reconsidered for the determination of whether the substrate is, or is not, held in a productive conformation. This is summarized for all three substrates in Table 3. Overall, a geometry of the  $\alpha/\beta$ -unsaturated carbonyl in *s-cis* conformation and a dihedral angle with  $<65^\circ$  are the best predictors for a good substrate.

**Table 3:** Summary of the validation parameters for the different model substrates.

	C-H distance	Si/Re-Face	Geometry - $\alpha/\beta$ -unsaturated carbonyl	Burgi-Dunitz angle	Dihedral angle
	3.5 Å	<i>Re</i> -Face	<i>s-cis</i>	46.6°	63.8°
	3.4 Å	<i>Re</i> -Face	<i>s-cis</i>	61.7°	46.6°
	4.7 Å	<i>Re</i> -Face	<i>s-cis</i>	151°	42.8°
	3.2 Å	<i>Si</i> -Face	<i>s-trans</i>	100°	72.3°

### **3 *In Silico* Mutagenesis Studies of the SQTCS ER Domain**

#### **3.1 Introduction**

The generation of the model of SQTCS ER was described in chapter 2. Validation of this model protein by various methods suggested that the binding of cofactor and substrates is chemically reasonable. We therefore considered that these models could form a valid basis for the design of further experiments with the aim of engineering the substrate selectivity of the ER domain of HR-PKS. There have been almost no reports of the successful rational engineering of HR-PKS, but based on the results from several *in vitro* studies of isolated HR-PKS domains, combined with the modelling described in the previous chapter, we considered that such experiments should now be possible.

#### **3.2 Aim of the Project**

The best-generated homology models of the ER domain of SQTCS will be used for different rational engineering experiments *in silico*. For example, different site-directed mutants will be generated *in silico* and then docking studies with different potential substrates using AutoDock Vina will be performed. In combination with sequence alignments, these studies will be used to investigate the effect of the site-directed changes of the ER domain *in silico*. The validation of mutagenesis and the docking studies will be done using the parameters previously determined for the functional conversion of the substrate-mutant complex. Hence, we should be able to assess if the substrate is, or is not, held in a productive pose. Overall, the generated models should be suitable for the rational design of different potential site-directed mutants for later *in-vitro* or *in-vivo* studies.

#### **3.3 Identification of Residues Potentially Involved in Substrate Selectivity in the ER domain of SQTCS**

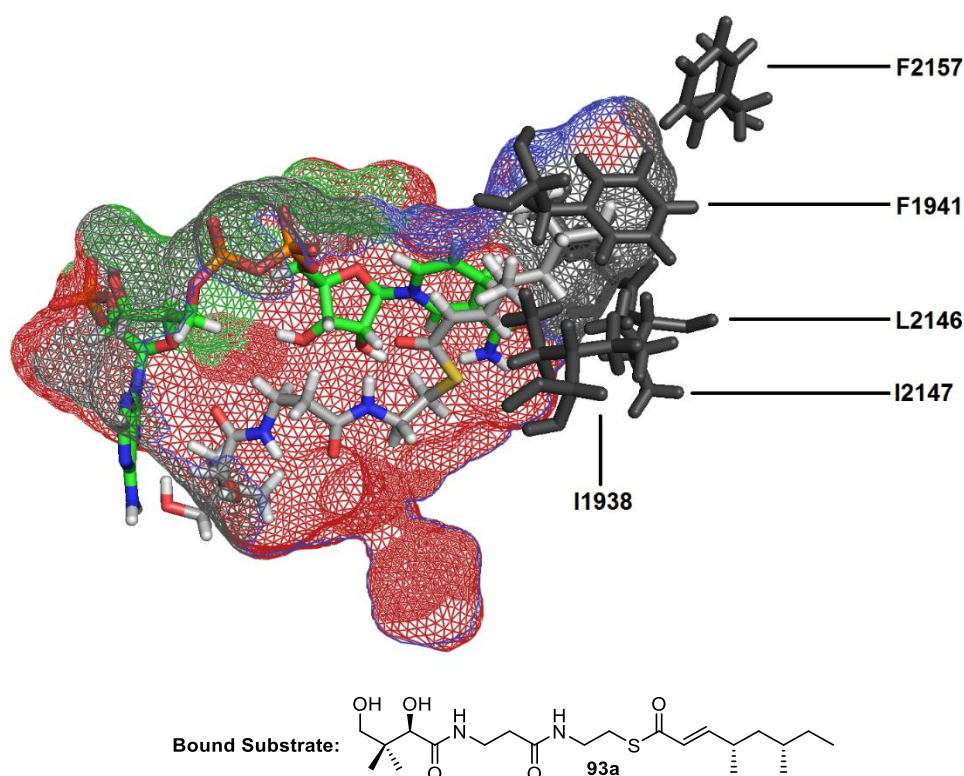
We hypothesized that the ER enacts its substrate selectivity *via* a fixed pocket length, which does not allow longer, or more methylated chains to bind productively for reduction (Figure 28). Therefore, increasing the length and volume of the substrate-binding pocket could enable larger or more methylated substrates to bind productively. For this purpose, amino acids, which bind the thiolester or the pantetheine, should not be mutated, but amino acids, which line the alkyl-binding part of the pocket, could be considered for mutation. The selection of the amino acids was done in PyMOL after the



docking of the substrates **77**, **85**, **93a** and **94** into the active pocket (Chapter 2, section 2.6).

In the first example squalestatin tetraketide pantetheine **93a** (6*S*,4*S*-2*E*-dimethyloct-2-enoyl-pantetheine, Figure 28 and 29) was used as a test-case. In particular, large hydrophobic residues lining the substrate-binding pocket (grey) were targeted (*e.g.* I1938, F1941, L2146, I2147 and F2157).

Additional to the docking studies, a sequence alignment was performed in Geneious. Only ER domain sequences from other fungal HR-PKS were chosen for the alignment (table 2). Sequence alignment of the ER of SQTKS with other HR-PKS pointed out highly conserved residues and variable residues in the ER sequence. This information was considered in selecting possible mutation residues. Since highly conserved residues are most likely important for an intact structure and function of the protein, these residues were not selected for mutation.



**Figure 28:** Active site of the ER domain with a mesh surface. In green the Cofactor NADPH **11** and in white the modeled substrate squalestatin tetraketide (6*S*,4*S*-2*E*-dimethyloct-2-enoylpantetheine) **93a**. Possible mutation residues in dark grey for the conversion of longer substrates and for the conversion of the squalestatintetraketide.

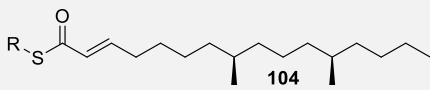
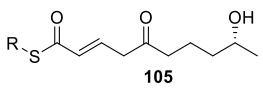
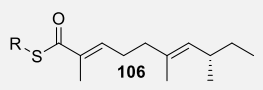
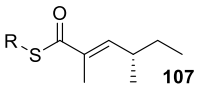
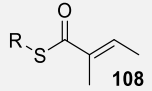
In addition, the structures of the longest intermediate, which is able to be reduced by the ER domain of each respective PKS, are shown in Table 4. However, the sequence alignment did not reveal any correlation between maximal chain length of the substrate

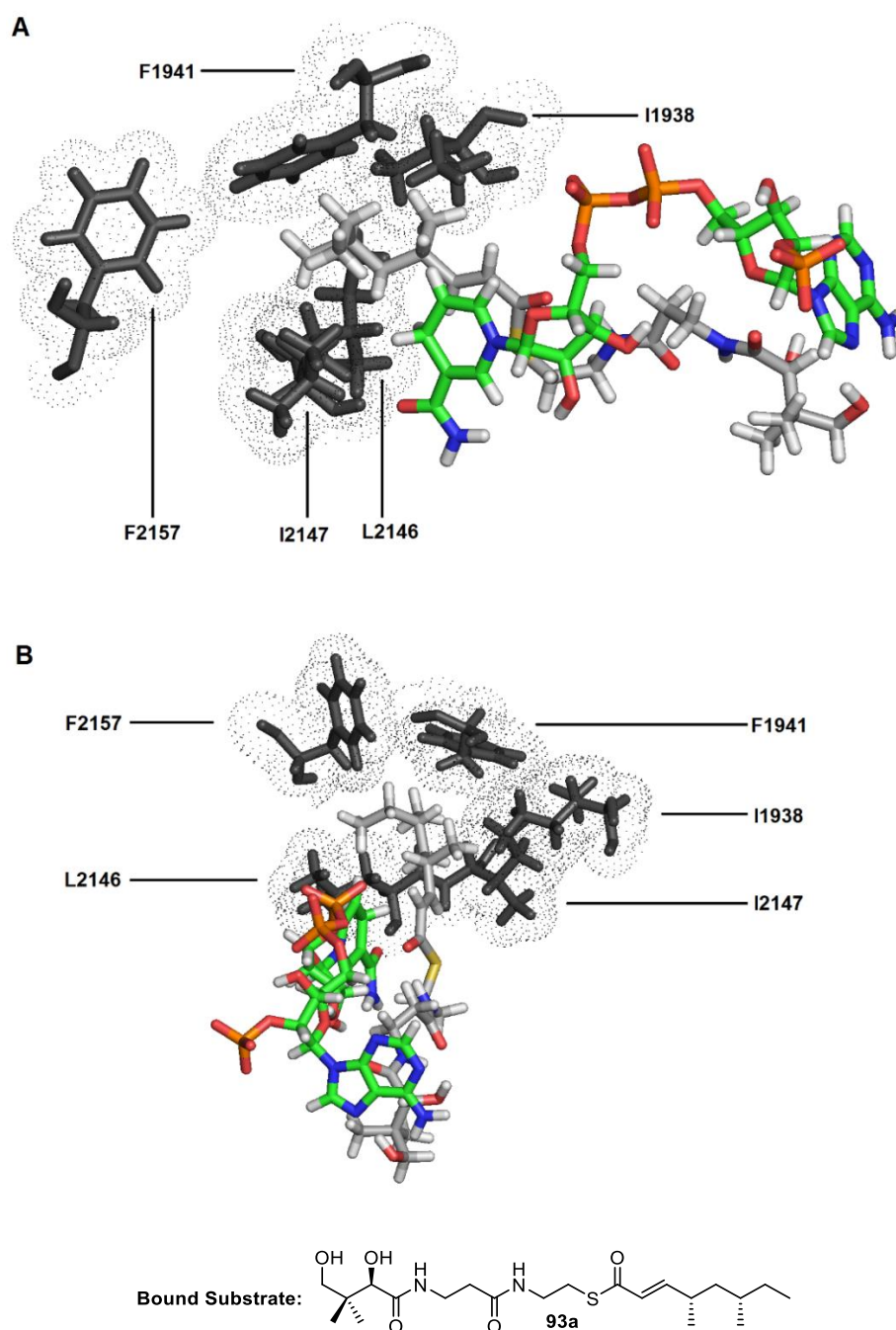


and the amino acid sequence of the ER-domain. As an example: the ER of FUM1 can reduce a C<sub>20</sub> chain.

A large and hydrophobic amino acid (F2157) is positioned at the end of the closed cavity of the ER SQTKS active site (Figure 29). Concerning the sequence alignment, this position is highly variable throughout fungal PKS (table 2). Overall, it is difficult to reason from the sequence alignment onto the chain length of the product, without further information, such as a 3D structure of each individual protein. However, the combined model and sequence analysis suggests it would be worthwhile investigating these identified residues further.

**Table 4:** Structure of the longest intermediate of the specific PKS, which still could be reduced by the respective ER domain

Compound	Structure	Length (acelates)
Fumonisin <sup>152</sup>		8
Zearalenone <sup>153</sup>		5
Alternapyrone <sup>154</sup>		5
Squalestatin tetraketide <sup>39</sup>		3
Asperfuranone <sup>155</sup>		2



**Figure 29:** Active site of the ER domain with a mesh surface. In green the Cofactor NADPH **11** and in white the modeled substrate squalestatin tetraketide (6S,4S-2E-dimethyloct-2-enoylpantetheine) **93a** possible mutation residues in grey for the conversion of longer substrates and for the conversion of the squalestatin tetraketide with their mesh surface. **A**, Back view; **B**, Side view.

### 3.4 *In Silico* Studies Concerning the Conversion of the Squalestatin Tetraketide-Pantetheine Stereoisomers 93a-d (ER domain of SQTKS)

#### 3.4.1 Investigation of tetraketide stereoisomers

In the previous section (section 3.3), various residues of the SQTKS ER domain were identified for possible mutations by docking studies. From previous studies in the Cox

group<sup>58,137</sup> it was known that a racemic mixture of diastereomers of the squalostatin tetraketide-pantetheine **93a-d** can be converted by the isolated ER domain *in vitro* although the single 4*S*,6*S* stereoisomer is not converted.<sup>58</sup> Thus, at least one of the three remaining stereoisomers **93b-d** should be converted by the ER domain of SQTGS. The conversion of any of the three diastereomers **93b-d** should depend on its orientation towards one or more residues within the active site. It therefore might be that not all diastereomers will be converted.

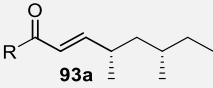
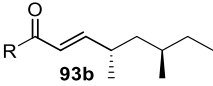
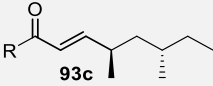
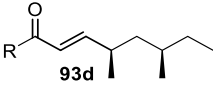
Hence, all four stereoisomers of **93** were docked, one by one, into the active site of SQTGS to visualize their orientation (Figure 30). This was achieved by starting with the already docked **93a** model (chapter 2.6). This was used as a template to generate the other stereoisomers (6*R*,4*S*)-**93b** (Figure 30A blue), (6*S*,4*R*)-**93c** (Figure 30B pink) and (6*R*,4*R*)-**93d** (Figure 30C yellow). The coordinates of the corresponding substrates were directly switched in the raw PDB-files. Afterwards, the identical Grid Box parameters for the calculations in AutoDock Vina were used. Subsequently each substrate-protein complex was minimized with YASARA. These calculations resulted in four different models, one for each diastereomer (Figure 30A-C). It was observed that the different backbones of the diastereomers take up similar poses in the active site of the ER domain (Figure 30A-C).

Furthermore, by comparing the diastereomers (6*S*, 4*S*) **93a** (Grey) with (6*S*, 4*R*, Pink, **93c**) and (6*R*, 4*R*, Yellow, **93d**) in Figure 32, it is observed that the CH<sub>3</sub>-groups at the 4 and 6 position of the substrates always take up the same pose in the active site and interact with the same amino acids (Table 5). Only the stereoisomer (6*R*, 4*S*, **93b**) differs from the other isomers (Figure 30B, table 5). The backbone of the chain overlaps very well with the natural stereoisomer **93a**. However, the CH<sub>3</sub>-group at position 6 has a different orientation in the active site of the ER domain. The CH<sub>3</sub>-group is oriented downwards with respect to the residue F1941 in the active site. This differs in comparison to the other stereoisomers **93c** and **93d**, where the CH<sub>3</sub>-group is orientated towards the residue F1941 (Table 5).

The  $\alpha/\beta$ -unsaturated carbonyls of the stereoisomers **93a**, **93c** and **93d** respectively adopt an *s-trans* conformation, which places the reactive  $\beta$ -carbon 3.2, 3.0 and 3.5 Å away from the cofactor's reactive 4'-*pro-R* hydrogen (Figure 30). In contrast, the  $\alpha/\beta$ -unsaturated carbonyl of the stereoisomer **93b** adopts an *s-cis* conformation, which places the reactive  $\beta$ -carbon 3.4 Å away from the cofactor's reactive 4'-*pro-R* hydrogen (Figure 30). As previously determined, the *s-cis* conformation is also observed in docked substrates known to be reduced by the ER.

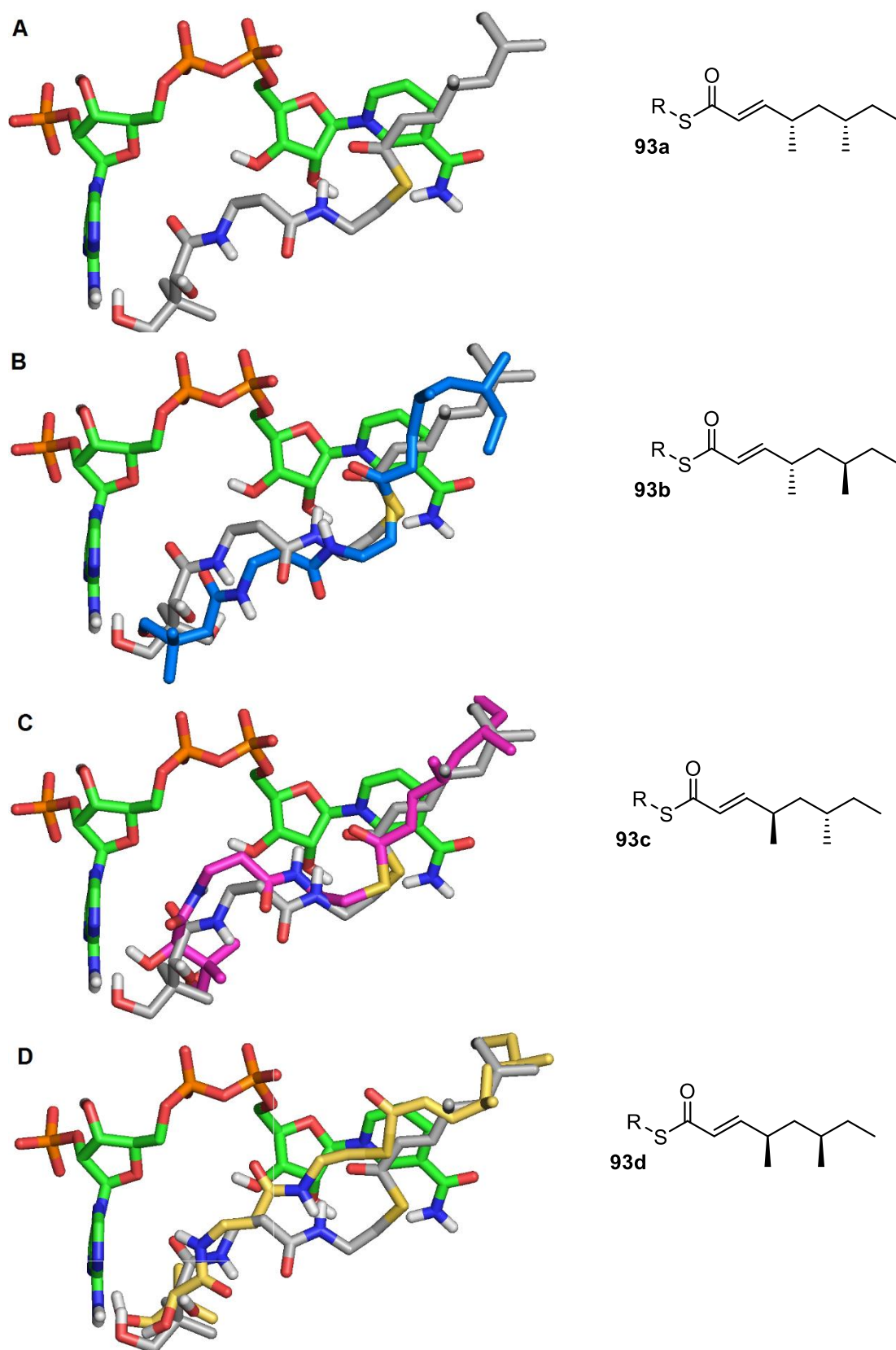
Hence, the *6R,4S*-stereoisomer has some parameters, such as C-H distances and carbonyl position, *but especially the s-cis conformation* which could result in productive conversion of the substrate by the ER domain (chapter 2, section 2.6). Overall, this information suggests that diastereomer **93b** can take up a productive conformation in the active site of WT SQTGS ER and **93b** may therefore be the diastereomer shown previously to be reduced *in vitro*. The residue F1941 may be the critical residue, which prevents conversion of diastereomers **93a**, **93c** and **93d** by the ER domain. F1941 appears to induce a steric control of the substrate through the interaction with the *S*-orientated CH<sub>3</sub>-group at the 6-position. Mutation of this residue could therefore lead to a different conformation of **93a** in the active site, which should be similar to the specific isomer **93b** and potentially allow the reduction reaction to occur.

**Table 5:** Overview of the different validation parameters for the poses of the different substrates.

	C-H distance	Si/Re-Face	Geometry - $\alpha/\beta$ -unsaturated carbonyl	Burgi-Dunitz angle	Dihedral angle	Interaction with other amino acids
	3.2 Å	Si-Face	<i>s-trans</i>	100°	72.3°	I9138(M), F1941 (B), L2146 (B), I2147 (B), F2157 (B)
	3.4 Å	Re-Face	<i>s-cis</i>	76.9°	75.6°	I9138(M), F1941 (M), L2146 (B), I2147 (B), F2157 (B)
	3.0 Å	Si-Face	<i>s-trans</i>	76.9°	74.3°	I9138(M), F1941 (B), L2164 (B), I2146 (B), F2157 (B)
	3.5 Å	Si-Face	<i>s-trans</i>	106.4°	44.1°	I9138(M), F1941 (B), L2146 (B), I2147 (B), F2157 (B)

Legend: (M) – Interaction with methyl group of the substrate

(B) – Interaction with the backbone of the substrate



**Figure 30:** Display of the active site of the ER domain with the cofactor in (green). In addition to the squalestatin tetraketide **93a** (6S,4S) are displayed the other docked diastereomers (grey). Blue (6R, 4S) **93b**, pink (6S, 4R) **93c** and in yellow (6R, 4R) **93d**.

The following docking experiments (Figure 31) should investigate if mutation of residues I1938 and F1941 to alanine would influence the orientation of the squalenstatin tetraketide-pantetheine **93a** to a similar orientation in the active site, like the stereoisomer **93b** Figure 31.

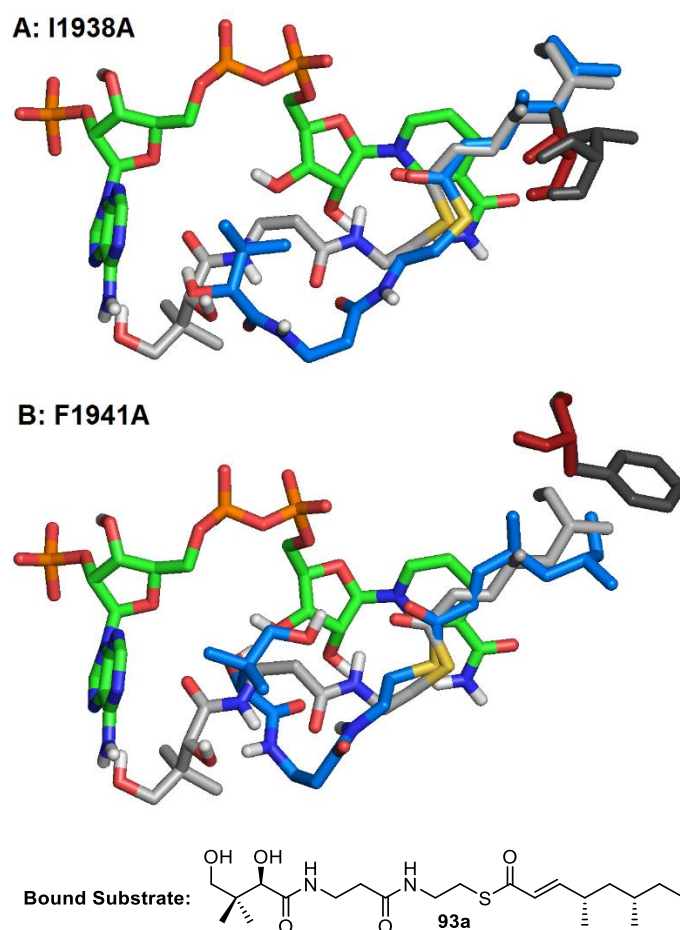
### 3.4.2 *In silico* mutation experiments

Therefore, residues I1938 and F1941 were individually changed in the *holo*-ER (ER domain with cofactor **11**) in PyMOL from their original amino acid to alanine *in silico*. The reason for the change to alanine was to make sure that a significant size change was performed. Afterwards, the mutated *holo*-ER model was submitted to YASARA to refine the respective model. Overall, two mutated *holo*-ER domains were obtained (I1938A and F1941A). The volume increase of the active sites for the different mutated ER domains *in silico* were investigated. The pocket volume calculations for the WT and mutated ER domains were predicted through an open software program (3V web server, Table 6).<sup>156</sup>

Subsequently the tetraketide **93a** was docked in the two mutated *holo*-ER domains. The molecular docking was performed with AutoDock Vina, without changing the GridBox parameters. Validation of the docking was done visually. The best-obtained substrate-protein complex was then minimized with YASARA. Through this *in silico* experiment two mutated *holo*-ER domains, with the docked tetraketide **93a** were obtained.

In the I1938A mutant it was observed that the orientation of the CH<sub>3</sub>-groups at positions 4 and 6 of **93a**, are unchanged in comparison to the original conformations (Figure 31A). The  $\alpha/\beta$ -unsaturated carbonyl of the I1938A mutated domains all adopt *s-trans* conformations (Table 6). However, in the F1941A mutant a new pose of squalenstatin tetraketide-pantetheine **93a** in the active pocket was observed (Figure 31B). In this pose the  $\alpha/\beta$ -unsaturated carbonyl of the docked substrate **93a** adopts an *s-cis* conformation. In addition, the CH<sub>3</sub>-group at position 6 possess a different orientation in the active site of the ER domain (Table 6). Hence, in the F1941A mutant the tetraketide **93a** can adopt a conformation which could result in productive reduction by the ER domain (chapter 2, section 2.6).

Overall, these *in silico* docking experiments suggest that the residue F1941 plays an important role in blocking the conversion of squalenstatin tetraketide-pantetheine **93a**. Based on the analysis it was hypothesised that the mutation F1941A could possibly result in conversion of **93a** by the ER of SQTKS.



**Figure 31:** Active site of the ER domain. In green the Cofactor NADPH **11**. Displayed are the docked substrates **87a** of the wild type (white) compared to the docked substrates after the change of a specific amino acid in the active side (blue) **93a**.

**Table 6:** Overview of the different validation parameters for the poses of the different mutated ER domains with docked **93a**.

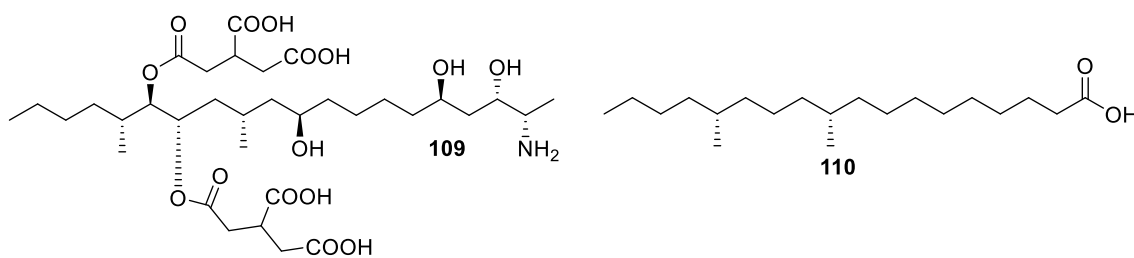
ER domain	Volume of active site / Å <sup>3</sup>	C-H distance	Si/Re-Face	Interaction with other amino acids	Geometry - $\alpha/\beta$ -unsaturated carbonyl
WT	1383	3.2 Å	Si-Face	I1938(M), F1941 (B), L2164 (B), I2147 (B), F2157 (B)	<i>s-trans</i>
I1938A	1480	3.0 Å	Si-Face	A1938(M), F1941 (B), L2164 (B), I2147 (B), F2157 (B)	<i>s-trans</i>
F1941A	1405	2.7 Å	Re-Face	I1938(M), A1941 (M), L2164 (B), I2147 (B), F2157 (B)	<i>s-cis</i>

Legend: (M) – Interaction with methyl group of the substrate (B) – Interaction with the backbone of the substrate

### 3.5 *In Silico* Studies Concerning Conversion of Longer Substrates (ER domain SQTKS)

So far, the docking experiments only investigated which possible residues could lead to an engineered enzyme which might convert squalstatin tetraketide-pantetheine **93a**. This analysis identified residue F1941 (section 3.4) as potentially blocking conversion of this substrate. We were also interested in engineering the ER to accept longer chains – for example penta- or hexa-ketides. Some HR-PKS ER domains must be able to accept longer chains, for example the ER involved in the biosynthesis of the nonaketide fumonisin **110** which must accept an octaketide.

Therefore, the structural features between SQTKS ER and FUM1, of the fumonisin **109** biosynthesis, were compared (Figure 33). FUM1 is a HR-PKS and the product of this PKS is **109** (Figure 32).<sup>80,152</sup> **109** is a highly reduced molecule, with even a longer carbon backbone than SQTKS. Hence, it is a perfect structure to compare the ER of SQTKS, which reduces a shorter carbon chain.

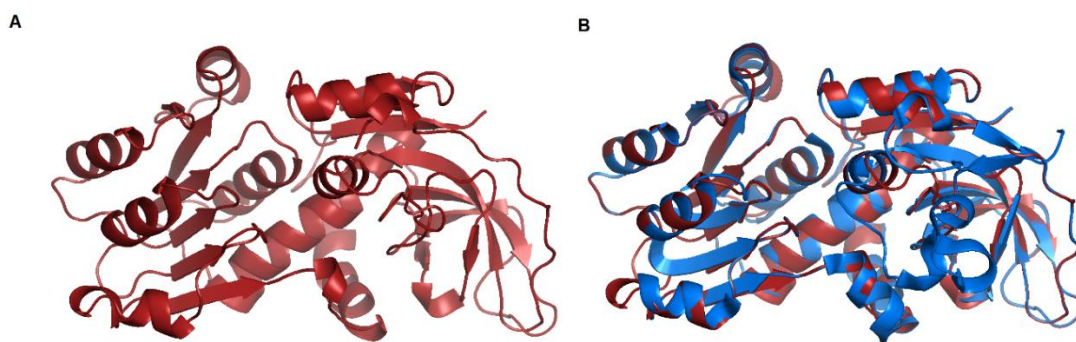


**Figure 32:** Structures of Fumosnin **109** and the product of FUM1 of the Fumonisin biosynthesis **110**.<sup>80,152</sup>

The domain boundaries of the FUM1 ER domain were determined using a conserved domain search (CD-Search).<sup>139–142</sup> Homology modeling was then performed in Swiss-Model<sup>83,92,95,105</sup> using the same CurF ER template (PDB 5dp2) as used for the modelling of the SQTKS ER domain. Validation of the model was done in the same way as the SQTKS ER (chapter 2, section 2.4). Overall, a structure model of the ER of FUM1 with a QMEAN value of -1.87 and an RMSD of 1.07 Å when compared to the template (section 2.1.1) was obtained after the modeling (Figure 33, red), which indicates, that the quality of the generated structure model was good.<sup>91–95</sup>

The alignment of the ER domain of SQTKS (Figure 33, blue) and with the ER domain of FUM1 (Figure 33, red) is displayed in Figure 33. This shows that the backbone overlays very well, which is also displayed in the backbone RMSD 1.37 Å. In addition, the surfaces of the pockets are highlighted.

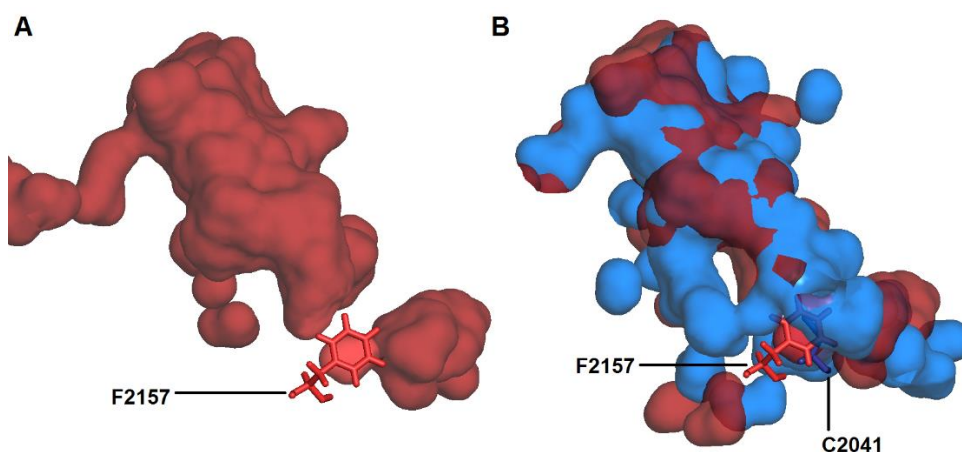




**Figure 33:** **A**, Model of the generated FUM1 ER; **B**, Alignment of the FUM1 ER (red) with the template CurF ER (blue, PDB 5dp2).

The volume of the active site pockets were estimated using the open software program 3V web server.<sup>156</sup> The calculations predicted that the volume of the active site channel of the FUM1 ER (1456 Å<sup>3</sup>) is increased compared to the active site channel of the SQTGS ER (1383 Å<sup>3</sup>). An increase in pocket volume of the SQTGS ER active site channel could facilitate conversion of longer substrates.

The main differences between the two structures appears to be at position F2157 which is a large hydrophobic residue located at the end of the active pocket the SQTGS ER (section 3.3). This is changed to a smaller residue (cysteine) in the FUM1 structure (sequence alignment, table 4). This seems to open up an extra space beyond the end of the SQTGS ER channel.



**Figure 34: Overlay of ER homology models:** **A**, in red the pocket volume of SQTGS Er with the residue F2157 at the end of the pocket displayed; **B**, in red the pocket volume of SQTGS ER, in blue pocket volume of FUM1 ER. In addition are the residues F2157 of the SQTGS and C2041 of FUM1 at the end of the pocket displayed. Interior surfaces of the proposed substrate binding pockets are shown.

For the *in silico* mutagenesis of the SQTGS ER, residues L2146, I2147 and F2157 were changed in PyMOL to alanine or valine. The mutation to alanine was done to perform a significant size change. On the other hand, the mutation to valine was done to perform a

lesser size change which might cause less structural damage to the protein. In addition, two mutated ER domains were generated, which also included the mutation F1941A, since this might be able to convert the squalestatin tetraketide **93a**. Afterwards, the mutated *holo*-ER model was submitted to YASARA to refine the respective model. Overall, eight mutated *holo*-ER domains were generated (Table 7).

The volume increase of the active side for the different mutated SQTKS ER domains *in silico* were investigated. The calculations for the WT and mutated ER domains were predicted through an open software program (3V web server, Table 7).<sup>156</sup> It was observed that some mutations increased the pocket volume (*e.g.* F2157A and F1941A/F2157A) while other changes decreased the pocket volume (*e.g.* I2147A and I2147A/F2157V, *etc.*) which was rather against expectations. Even so, these amino acids might have a critical effect on the general scaffold of the active pocket, effectively holding it open. Hence, through the mutagenesis to smaller residues the active site could partially collapse, which decreases in the volume of the active site.

The next step was to perform molecular docking in AutoDock Vina with the triketide **85** and pentaketide **94** in the same way as previously described (Chapter 2, section 2.1.2).<sup>110,147,148</sup> The models were then refined by YASARA. The visualization of the different models after the refinement step was always in PyMOL. As an example, the complex between pentaketide substrate **94** in the F2157A mutated ER domain is shown in Figure 35.

The residue F2157 is an ideal candidate for mutation experiments because it fulfils the following criteria. The residue F2157 is orientated in the end of the active site pocket (Figure 35). In addition, the residue is a large hydrophobic amino acid. Hence, possible interactions, such as hydrogen bond or other polar interactions could be neglected. Further, its mutation in the case of F2157A (61 Å<sup>3</sup>) and less-so in the case of F2157V (49 Å<sup>3</sup>) leads to an apparent pocket volume increase (Table 7). From these docked substrate-protein complexes parameters, such as C-H distances and carbonyl positions were determined (Table 7). Subsequently, these parameters were analyzed to characterize if the docked substrate-protein complex could result in productive conversion of the substrate by the ER domain.

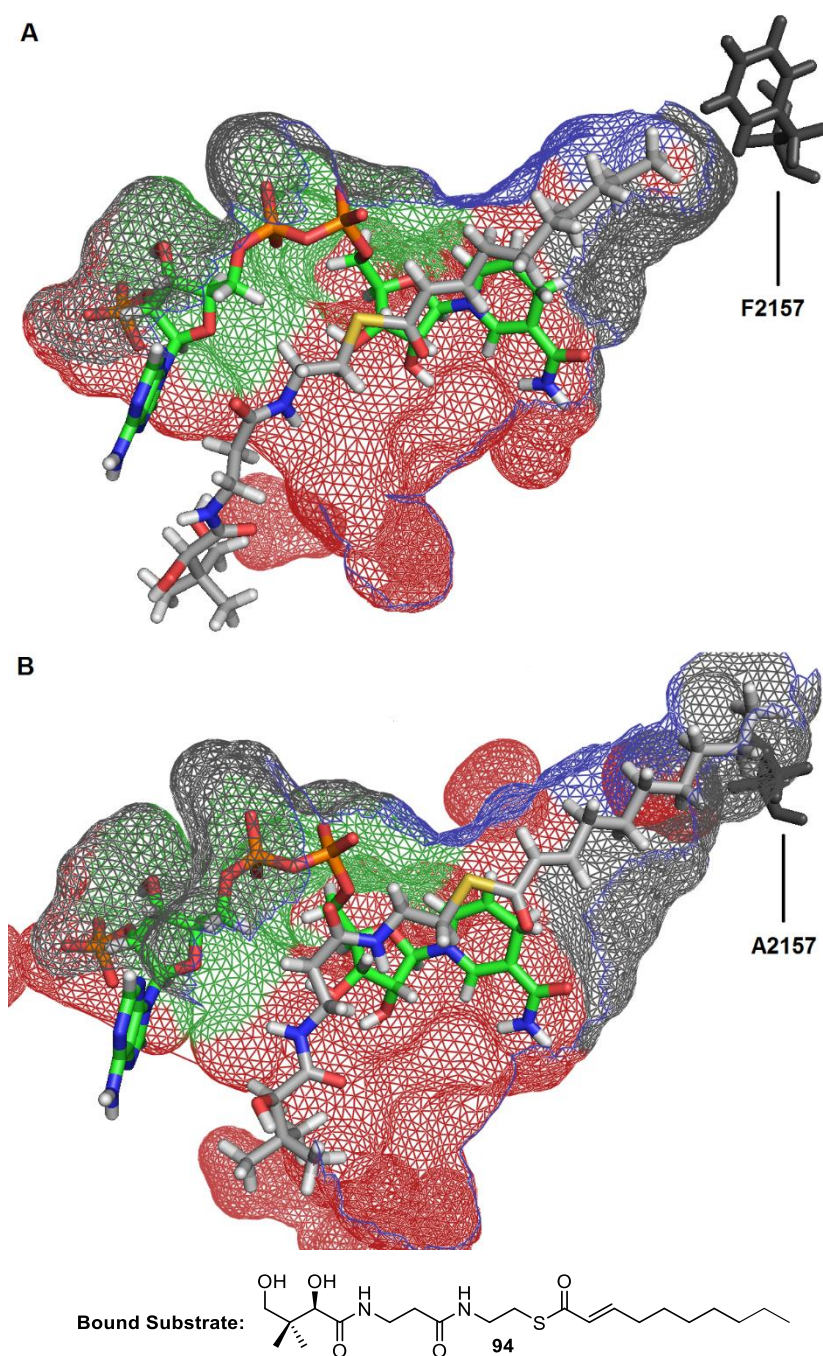
If the orientation of the docked substrate **94** in the non-mutated ER domain (Figure 35A) is compared to the docked substrate **94** in the mutated ER (Fig 35B) it could be observed that the mutation elongates the active site and gives the substrate **94** more space. The substrate **88** can fit deeper into the active site. This correlates with the volume

increase of the mutant compared to the WT (Table 7). Further, the thiolester and the  $\beta$ -carbon are located closer to the nicotinamide moiety of the cofactor. In the WT ER domain, the distance between the  $\beta$ -carbon and the catalytic hydrogen of the nicotinamide is 4.7 Å. In the mutated ER domain, the distance is reduced to 3.2 Å. For good substrates (Triketide **85**) an average distance of 3.4 Å was observed (Table 7). Hence, the distance in the mutated ER domain now corresponds better to a substrate-protein complex which could result in a productive conversion.

**Table 7:** Overview of the different validation parameters for the poses of the triketide **85** and pentaketide **94** in the different mutated ER domains *in silico*. Green figures show an increase in pocket volume; red figures show a decrease in pocket volume.

ER domain	Volume / Å <sup>3</sup>	C-H distance Triketide 85 Å	Si/Re-Face	C-H distance Pentaketide 94 Å	Si/Re-Face	Geometry Triketide 85	Geometry Pentaketide 94
ER (WT)	1383	3.4	Re	4.7	Si	s-cis	s-cis
L2146A	1367	3.5	Re	4.6	Si	s-cis	s-cis
L2146V	1376	3.4	Re	4.7	Re	s-cis	s-cis
I2147A	1307	3.0	Re	4.8	Re	s-cis	s-cis
F2157A	1444	2.8	Re	3.2	Si	s-cis	s-cis
L2146A/ I2147A	1269	3.0	Re	4.7	Si	s-cis	s-cis
I2147A/ F2157V	1317	2.4	Re	4.0	Si	s-cis	s-cis
F1941A/ F2157A	1467	2.8	Re	3.3	Re	s-cis	s-cis
F1941A/ I2147A/ F2157V	1392	2.5	Re	3.9	Si	s-cis	s-cis

Furthermore, when comparing the distances for the triketide, it was observed that mutations of F2157A in combination with a mutation of I2147A greatly reduced the distance between the substrate and hydride. In combination with the volume decrease of the active pocket, the substrate might need less time to take up the correct pose; hence, shorter chains might bind more productively for reduction through the volume decrease of the active pocket.



**Figure 35:** Active site of the SQTKS ER domain with a mesh surface. In green the Cofactor NADPH **11** and in white the modeled substrate **94**. Modelled are: **A**, the wild type and **B**, the mutated ER (F2157A).

### 3.6 Conclusion

To date, different approaches have been explored to generate new polyketides by engineering HR-PKS.<sup>82</sup> Most efforts to reprogram HR-PKS have been focussed on swaps of domains or other parts of the specific PKS.<sup>82</sup> A different approach is the modification of active sites to induce different substrate specificity or higher substrate variability.<sup>82</sup> Such an approach could preserve the structural and conformational properties of the complete PKS as well as the protein-protein interactions. Hence, only the intrinsic programming of the protein would be influenced, whereby the extrinsic programming would not be changed. However, no examples of HR-PKS engineered by site-directed mutagenesis are known in the literature to date.<sup>69,157–159</sup> The major reason for this is that engineering of PKS requires detailed knowledge about structural arrangements, the active site and the entire PKS itself, and to-date this information has not been available because of the lack of structural data for HR-PKS domains.

Hence, our first aim was to build and validate a model of the SQTGS ER domain which had a high enough quality to perform *in silico* studies. The validation of the generated model of the ER domain was done through different parameters (Chapter 2.4-2.6). One of the validation parameters was the QMEAN score of Swiss-Model. The QMEAN score verified that it was possible to create a detailed model of the ER domain, which was suitable for docking studies. However, it should be always kept in mind, that homolog modelling is always based on the structure of another protein. It cannot reflect the exact structure of the modeled PKS or PKS domain, like crystal structures. Even so, a crystal structure also does not display the exact reality, because it is only a snapshot of a certain state of the enzyme. However, enzymes are flexible structures and can have different conformations. Hence, the protein or the active site could have different states or structural conformations, which could not be displayed in the crystal structure. However, in addition to the good QMEAN value, the generated model showed that after different minimization steps no change in its overall structure could be observed. This indicates that the ER domain, or rather the active pocket of the ER domain, is not highly flexible. We also docked the cofactor NADPH **11** and showed that this docks in a sensible way, preserving known protein-cofactor contacts and exposing the correct 4'-*pro-R* hydrogen known to be transferred during the reduction reaction. In addition, the template that was chosen for the modeling had a high resolution of 1.0 Å. Taking everything together: the QMEAN, the inflexible structure and the good template resolution, it can be concluded that the generated ER model of the SQTGS is good enough for further study.

Afterwards, docking studies were performed with AutoDock Vina, which indicated possible residues for a rational engineering approach. Two aims were reconsidered doing the docking studies. The mutations should lead to a conversion of longer substrates, which should be achieved through an elongation of the pocket or/and an increase of the pocket volume. In addition, the mutations should lead to a conversion of the squalestatin tetraketide-pantetheine **93a**.

The docking experiments of the ER domain SQTGS then indicated different amino acids for a mutagenesis approach. The amino acids, which were selected for the mutagenesis included F194I, for a conversion of the squalestatin tetraketide pantetheine **93a**. In addition, mutation of the amino acid F2157 in different combinations should allow longer chains to bind productively for reduction. Further, a decrease of the active pocket was observed, if certain amino acids were mutated, for example I2147A or L2146A. Hence, these amino acids might have a critical influence on the scaffold of the active pocket. Even so, shorter chains might bind more productively for reduction through the volume decrease of the active pocket.

Overall, the models were suitable to rationally design different potential site-directed mutants *in silico*. These amino acids shall be subsequently mutated and investigated in the *in-vitro* studies.

## 4 *In Vitro* Engineering of the SQTCS ER domain

### 4.1 Introduction

In chapters 2 and 3 (section 2.4-2.6 and 3.2-3.5) fundamental *in silico* studies for the *in vitro* engineering of the ER domain from SQTCS were performed. This included the creation of a homology model of the ER domain. The quality of the model was verified by various means (chapters 2 and 3). The modelled ER domain was then used as a scaffold for further *in silico* studies, which included, amongst other things, different docking experiments which indicated possible residues for a rational engineering approach (Chapter 3, section 3.3 and 3.4). Two aims were considered during the docking studies. The mutations should lead to a conversion of longer substrates, which should be achieved through an elongation of the pocket and/or an increase of the pocket volume. In addition, the mutations should lead to a conversion of the squalestatin tetraketide-pantetheine **93a**. For this, hydrophobic residues were considered. The selected residues were mutated *in silico*, and the results of further docking studies suggested different effects on the behavior of substrates in the active site. For example, the mutation F1941A indicated that the tetraketide intermediate **93a** could become a substrate. For the F2591A mutation, the docking studies suggested that the active site is expanded and that pentaketide substrates could possibly be reduced.

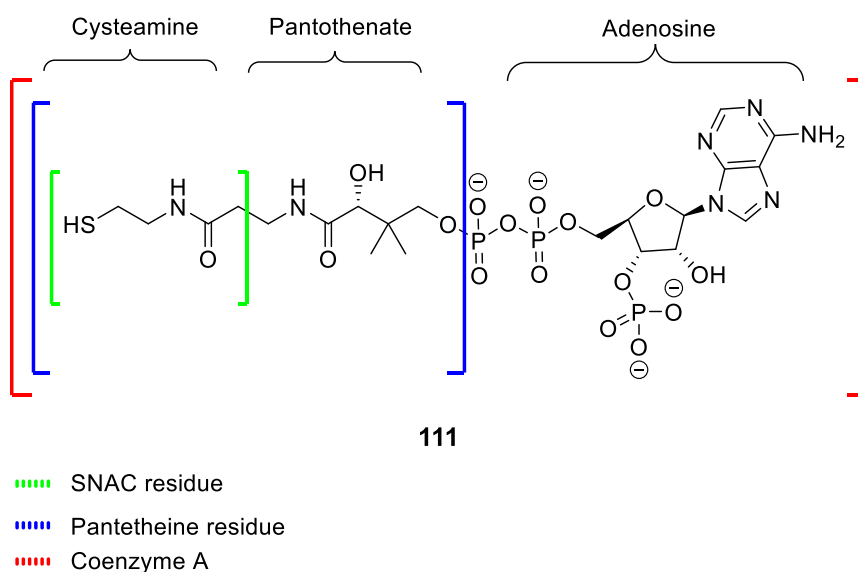
### 4.2 Aim of the Project

The aim of this project was to rationally engineer HR-PKS by introducing a small number of residue changes, rather than the much larger domain-swap which have been previously reported. Previously, the isolated ER domain of SQTCS was analyzed *in silico* and a plan for the rational redesign was proposed (chapter. 3, section 3.2-5). This shall be applied on the ER domain of SQTCS and effects of the mutagenesis verified through suitable *in vitro* assays. For this purpose, amino acids that had been selected in the docking experiments (Chapter 3, section 3.2-5), were mutated in the ER domain through molecular biological methods. Afterwards, the ER protein was expressed in *E. coli BL21* and purified. In parallel, a substrate library for the ER domain, varying in chain length and methylation pattern, was synthesized. These natural and unnatural substrates were used together with the expressed enzymes in either UV-assays or LCMS-assays to characterize the influence of the introduced mutations on the enzyme activity.

Overall, the experiments should reveal if it is possible to rationally engineer an HR-PKS or a sub-domain. In addition, the effect of the different mutations on the intrinsic programming of the ER domain will be investigated, especially how the chemical selectivity and kinetics of the different substrates are influenced by the changes introduced into the ER domain.

### 4.3 Substrate Synthesis

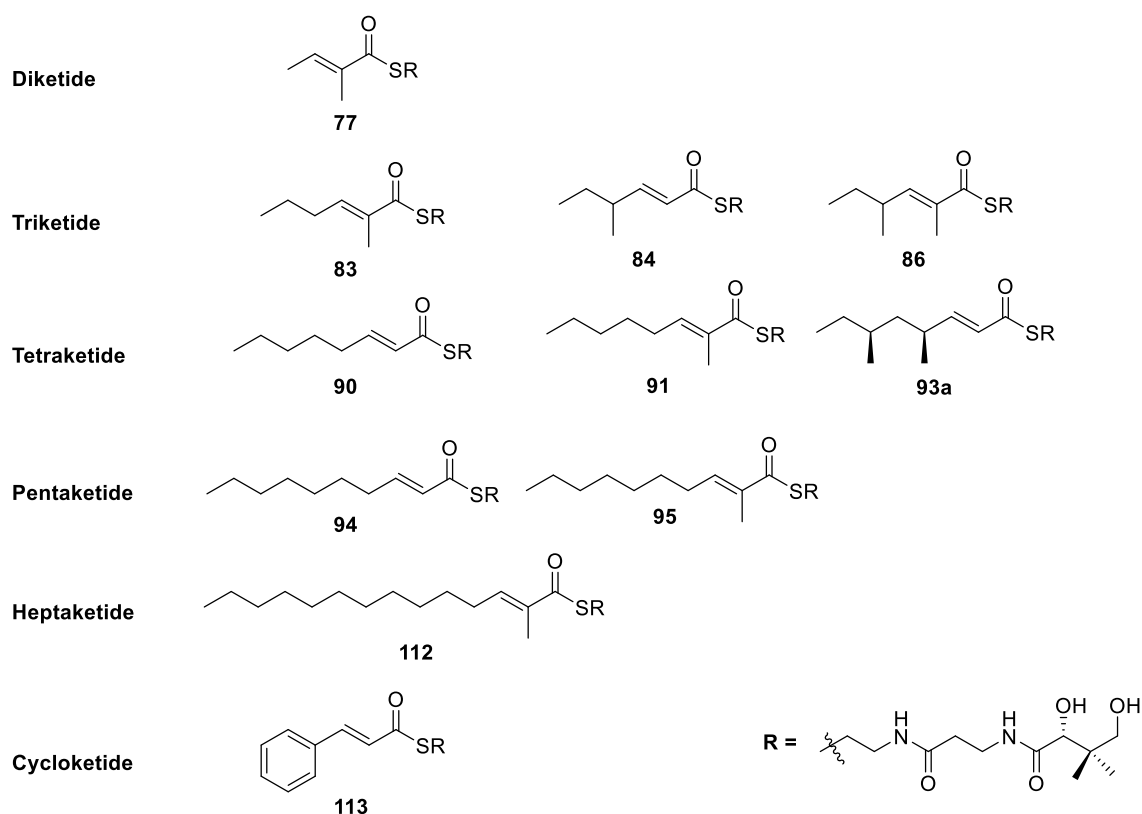
For the kinetic assays, as in the previous studies, acyl pantetheines were used (Figure 36).<sup>58,137</sup> Since ACP-bound substrates are hard to produce and to analyse, acyl-ACP mimics such as SNACS and pantetheines are widely used.<sup>58,137</sup> Previous work in the Cox group had shown that pantetheine-bound substrates have higher specificity ( $k_{cat}/K_M$ ) than SNACS for the SQTGS ER domain.<sup>137</sup> Additionally, compared to SNAC substrates pantetheines are more water soluble which is especially important for longer (and hence more hydrophobic) acyl groups.



**Figure 36:** Structure of Coenzyme A

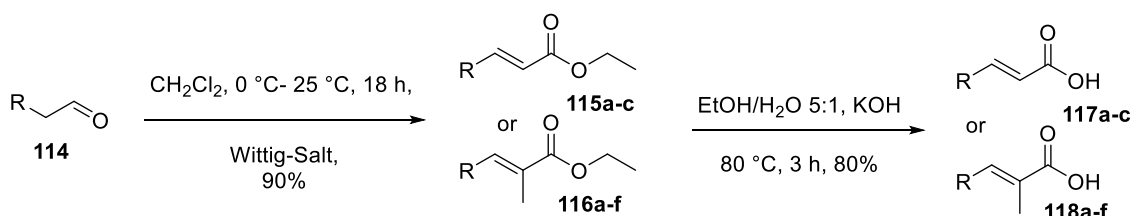
In addition, the synthesis of the library was focused on specific substrates to compare them to previous studies.<sup>137</sup> In the previous studies, the main goal was to test a wide range of different pantetheine substrates, from diketides to pentaketides. The library which was generated here also includes natural substrates (**77** and **86**) and unnatural substrates, *e.g.* with modifications in the chain length (*e.g.* **94**, **95** and **112**); the methylation pattern (*e.g.* **90** and **91**) and the methylation position. Additionally, a long chain substrate (**112**) and a cinnamic acid derivative (**113**, Figure 37) were included.





**Figure 37:** Synthesized pantetheine substrates for the enzyme assays.

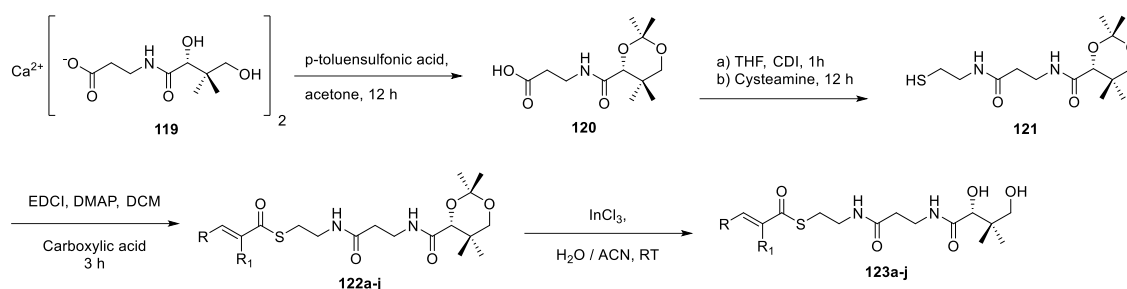
Wittig reactions were mostly used to create the  $\alpha$ -methylated or non-methylated unsaturated carboxylic acids (Scheme 23). Therefore, different aldehydes were used as the starting materials for elongation by Wittig salts. Both starting materials were stirred in  $\text{CH}_2\text{Cl}_2$  at RT for 15-18 hours and the olefinic products were purified by flash column chromatography. The resulting esters were cleaved by potassium hydroxide in a mixture of ethanol and water (5:1). After three hours under reflux, the mixture was acidified and extracted to isolate the final acids (Scheme 23).



**Scheme 23:** Synthesis route of  $\alpha$ -methylated and non-methylated carboxylic acids.

In parallel, protected pantetheine **121** was produced following a previously published protocol (Scheme 24).<sup>160</sup> The pantothenic acid hemi-calcium salt **119** was protected in acetone at 25 °C in the presence of *p*-toluene sulfonic acid to give the acetal **120**. After that, a coupling with CDI and cysteamine in THF was used to generate the pantetheine

dimethyl ketal **112** (Scheme 23). This was then coupled to the previously synthesised carboxylic acids through an EDCI/DMAP coupling protocol (Scheme 24) to give **122a-j**. The yields of 80-84% for the reaction are similar to the previous studies.<sup>58,137</sup>



**Scheme 24:** Synthesis route of pantetheine substrates

The introduced protection group changes the solubility of the pantetheine so that **122a-j** are soluble in  $\text{CH}_2\text{Cl}_2$ , EtOAc and MeCN. In the previous studies, the deprotection of **122a-j** was done with aqueous TFA.<sup>137</sup> However, this method causes partial cleavage of the thiolester, which results in lower yields. A new deprotection method of the pantetheine dimethyl ketal substrates **122a-j** was therefore established using  $\text{InCl}_3$  catalysis in aqueous  $\text{CH}_3\text{CN}$  which avoided the thiolester hydrolysis problem (Scheme 24). In total, eleven pantetheine substrates **77-95** and **112-113** were successfully synthesized which varied in the methylation pattern, chain length and the methylation position (Figure 37). The exception is substrate **93a**, which was produced by hydrolysis from squalestatin S1 itself followed by coupling to **121** and standard deprotection.<sup>137</sup> All compounds were fully characterised by NMR and MS. This substrate library of diketides, triketides, tetraketides and pentaketides, heptaketides and cycloketides completes the chemical part of the project.

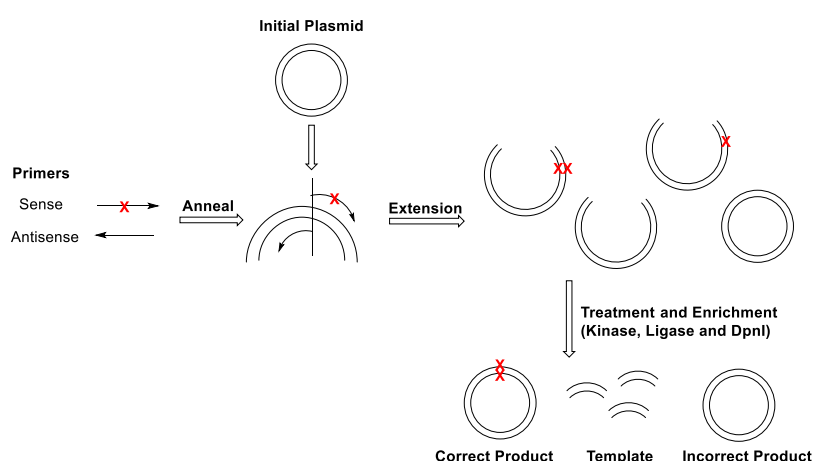
#### 4.4 Site Directed Mutagenesis

Besides synthesizing a suitable substrate library, the desired mutations (chapter 3.2-3.5) should be inserted into the isolated ER domain. Then the mutated proteins should be expressed in a suitable host and be purified using Ni-affinity chromatography.

The docking studies in the *in silico* analysis (chapter 3.2-3.5) revealed that the mutation of the amino acids, which are displayed in table 8, had an effect on either the volume of the active site of the ER domain or the orientation of the substrates in its binding pocket. The SQTGS ER domain has been previously cloned into an expression vector (pET28a) from an *E. coli* optimized sequence.<sup>161</sup> This vector was used as the

template for the introduction of mutations. For the insertion of mutations into proteins there are different methods known in literature.<sup>162</sup> In this project, a site directed mutagenesis approach, was chosen. The template DNA of the isolated ER was contained on a pET28a (+) vector with an N-terminal his<sub>6</sub>-tag. This vector was available in the stocks of the Cox group and was generated by a previous PhD-student in the lab (David Ivison).<sup>163</sup> The vector also contains a kanamycin resistance gene and the expression was regulated through the Lac-Operon (Figure 38).

Different, non-overlapping, short DNA primers were generated which contained the required mutations on the sense primer and no mutations on the antisense primer. These synthetic primers were complementary to the template DNA around the mutation site, so they can hybridize with the DNA sequence of the gene of interest (Scheme 25). Afterwards, the primer was extended using a high-fidelity DNA polymerase Q5, which amplifies the rest of the plasmid and results in a blunt ended PCR product (Scheme 25).<sup>162</sup> The amplified DNA sequence then contains the mutated gene. After the elongation, the amplified PCR product was closed through a ligation step using Quick Ligation™ Kit (New England Biolabs, Inc.) to ligate the blunt ends. Next, the template DNA was removed through a *DpnI* digestion which digests methylated template DNA, but not unmethylated PCR products (Scheme 25). The plasmid was then transformed into *E.coli* TOP10. Clones were selected and isolated and miniprep-DNA was prepared. Finally, mutants were selected by DNA sequencing to check if they contained the desired mutation.



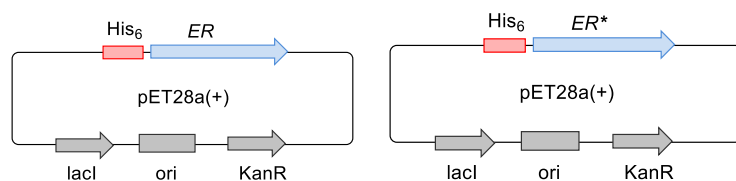
**Scheme 25:** Schematic overview over the individual steps of the site directed mutagenesis protocol <sup>162</sup>

Overall, nine theoretical mutations should be inserted into the ER domain (table 8). However, only five mutated ER proteins, which contained one or more mutated amino acids, were successfully generated. For the mutation I2147A, it was not even possible to

insert the desired mutation into the vector (Table 8). All sequencing results were negative and showed that the base pair sequence in the specific ER domain was not changed. Variation of the conditions for the site directed mutagenesis did not lead to an insertion in the vector (Table 8). The variation included different primers and annealing temperatures. There are different possible explanations for the failed creation of these mutants. The most obvious reason could be the high GC content in this area of the DNA. This could prevent the correct binding of the primer to the DNA. Another reason could be that the primer still contained secondary structures such as hairpins and even higher temperatures could not unfold these secondary structures.

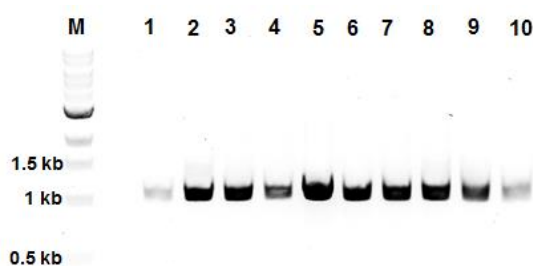
**Table 8: Overview of generated mutants of the ER domain from SQTGS:** Red text indicates mutants predicted to reduce active site volume; green text indicates mutations predicted to increase active site volume.

	<b>Mutated amino acids</b>	<b>Stability of the mutated ER protein</b>	<b>Predicted effect on the substrate or the ER domain (chapter 3.2-3.5)</b>
<b>1</b>	F1941A	Stable	Orientation Volume 1405 Å <sup>3</sup>
<b>2</b>	L2146A	Unstable protein	Volume 1367 Å <sup>3</sup>
<b>3</b>	L2146V	Unstable protein	Volume 1376 Å <sup>3</sup>
<b>4</b>	I2147A	Mutation was not possible	Volume 1375 Å <sup>3</sup>
<b>5</b>	F2157A	Stable	Volume 1444 Å <sup>3</sup>
<b>6</b>	L2146A I2147A	Unstable protein	Volume 1269 Å <sup>3</sup>
<b>7</b>	I2147A F2157V	Stable	Volume 1317 Å <sup>3</sup>
<b>8</b>	F1941A F2157A	Stable	Orientation Volume 1467 Å <sup>3</sup>
<b>9</b>	F1941A I2147A F2157V	Stable	Orientation Volume 1392 Å <sup>3</sup>



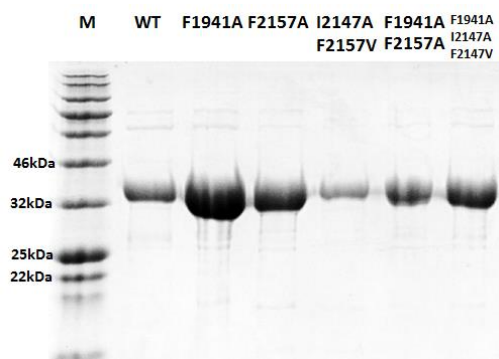
**Figure 38:** Plasmid in pET28a (+) \_ER and ER\* (mutated) which was transformed into in *E. coli*.

Successful transformation of the plasmid into *E. coli* was confirmed by colony PCR, amplifying a *ca.* 1kb DNA fragment specific of the ER domain. An example is shown in Figure 39 for the F1941A mutant ER. The expression of the wild type and mutated ER protein was achieved by heterologous expression of the plasmid in *E. coli* BL21.



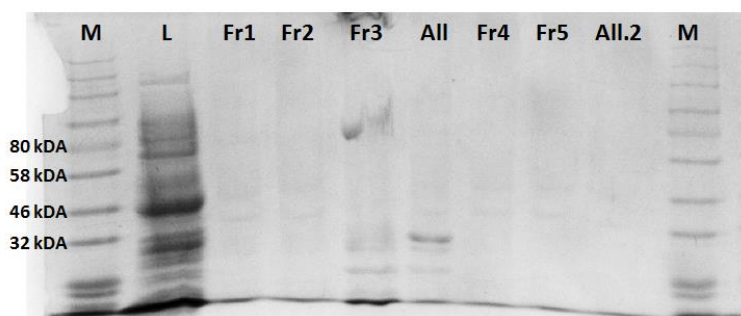
**Figure 39:** Colony PCR of different *E. coli* transformants (1-10) for the F1941A mutant ER. Marker (M) displayed on the left, on the right different analyzed colonies numbered from 1-10. A band corresponding to the desired ER DNA fragment was observed at 1.2 kb

Positive strains were grown in a starter culture of 2TY-medium. This starter culture was incubated overnight. The next day 100 ml batches of 2TY- media were inoculated with the starter culture. The protein production was induced with *iso*-propyl  $\beta$ -D-1-thiogalactopyranoside (IPTG, 0.1 mM) at an OD<sub>600</sub> of 0.6 and at 16 °C. After incubation overnight, the media was centrifuged and the pellet was collected. The cells were lysed through sonification and the purification of the ER protein was accomplished through Ni-affinity chromatography. The separation of the ER protein from the lysate was done using the N-terminal His<sub>6</sub>-tag with the corresponding columns (Chapter 8, section 8.4). Overall, five of the seven mutated proteins and the wild type ER were successfully expressed without any contaminations (Figure 40).



**Figure 40:** SDS Gel of the different expressed mutated ER domains. M = Marker, WT = Wild type, F1941A-F2157A/V = different mutants. Selected marker bands are highlighted in kDa.

Expression of the mutated ER domains (L2146A, L2146V and L2146A/I2147A) resulted in no production of protein or rather no corresponding band could be observed in the SDS-PAGE Gel. As an example, the SDS-PAGE gel of the different obtained fractions after the Nickel and Superdex-columns for the ER mutant (L2146A / I2147A) is shown in Figure 41. In the lysate fraction there is still a band to the corresponding mass of 32 kDa. However, this band cannot be observed in the following fractions of the next purification steps. Possibilities are that the protein was misfolded because of the mutation of the amino acid residues. If the protein would then form inclusion bodies, these could not be observed in the lysate fraction, but only in the pellet. Hence, the inclusion bodies are insoluble. Another possible reason could be that the N-terminal His<sub>6</sub>-tag was destroyed or impaired; hence, the protein cannot interact any more with the Nickel-column for the purification. Even so, most likely is that the protein got unstable or misfolded through the mutation of L2146. Miss folding could result in quicker degradation by the proteasome or decomposition into smaller parts. This would explain the degradation over time and the reason that no protein was observed in the later fractions of the purifications steps.



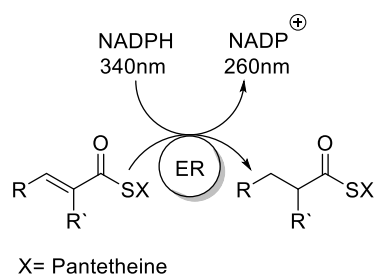
**Figure 41:** SDS gel of the different fraction of the Nickel-column and size exclusion column chromatography for the purification of the ER(L2146A / I2147A). M: Marker: not purified Lysate, Fr1-F3: Fractions from the Nickel column. All: united Nickel-columns; Fr4-F5: Fractions from the exclusion column chromatography; All.2 united Fractions from the exclusion column chromatography.

The instability or mis-folding of the single mutated protein and double mutated proteins could be explained through residue L2146 and its significant influence on protein stability. The sequence alignment for the ER domain from SQTGS (Chapter 3, section 3.3, table 6) showed that this position was highly conserved in the ER domain from different organisms. It is also conserved in mouse and pig FAS ER. In combination with the observation that every mutant in which L2146 was mutated, no protein could be purified, it could be assumed that this amino acid plays a role in the overall structure stability of the ER domain. These observations also correspond with the observed collapse of the substrate-binding pocket when L2146 and I2147 were mutated *in silico* (Chapter 3, section 3.5). Hence, if this amino acid, which is buried deep into the protein active site, is changed to something smaller, the change influences an overall critical change in the protein structure and it is degraded or broken down.

Overall, five stable mutant proteins were expressed in *E. coli BL21* to successfully give soluble protein. These were then used for the enzymatic investigations. In the enzyme assays, the substrate range and the kinetic conversion of wild type ER domain and mutated ER domains will be compared.

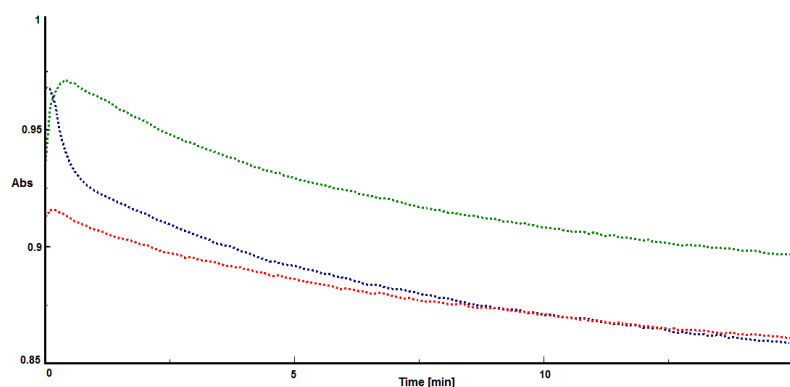
#### 4.5 *In Vitro* Enzymatic Investigation of the ER domain

Kinetic investigations of the wild type ER and the mutated ER domains were performed with the synthesized substrates to compare the influence of the mutation on the kinetic values, in particular the specificity constant ( $k_{cat}/K_M$ ). Since  $V_{MAX} = k_{cat} \cdot [\text{enzyme concentration}]$  it is usually more convenient to measure  $V_{MAX}$  at constant enzyme concentration and the specificity constant becomes  $V_{MAX} / K_M$ . The specificity constant is a measure of how efficiently an enzyme converts substrates into products. A comparison of specificity constants can be used as a measure of the preference of an enzyme for different substrates (*i.e.* its substrate specificity). The higher the specificity constant, the more the enzyme “prefers” that substrate.<sup>164</sup>



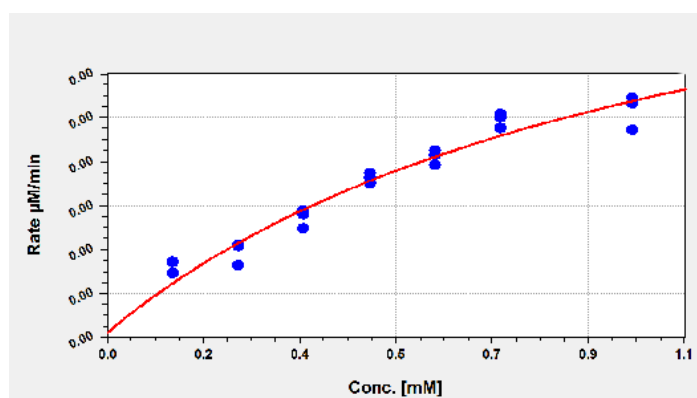
**Scheme 26:** Reaction of the enoyl reductase in the presence of NADPH

A continuous assay was performed. This method is based on the absorption wavelength of the NADPH cofactor (340 nm). The ER oxidizes NADPH during the conversion of the substrate and therefore its concentration decreases over time in the assay, which is measured continuously at 340 nm (Scheme 26). Since the reaction stoichiometry is 1:1, the change in NADPH concentration equals the change in substrate (or product) concentration. For standardisation, the enzyme and cofactor concentration are the same in every performed assay. Only the substrate concentration varies in the assays.



**Figure 42:** Absorption curves (green, blue, red) of the three replicates of tigloyl pantetheine **77** at 1 mM.

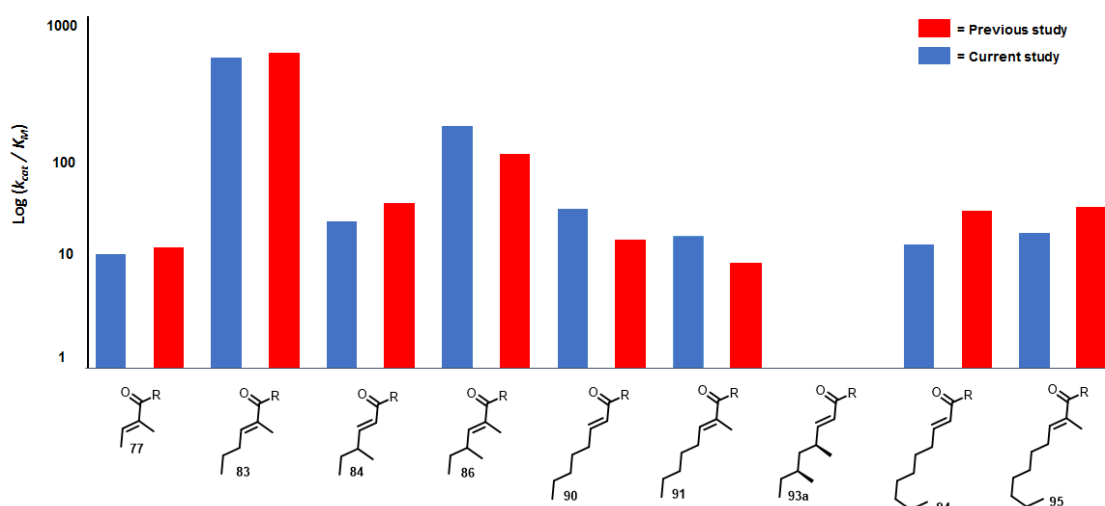
This data is then used to obtain the rate of reaction for different substrates at different concentrations using Microsoft EXCEL (Figure 42). The rate data were plotted with the software *Curve Expert* (Figure 43).<sup>58,137</sup> In that way, it was possible to generate a nonlinear fit of initial rates in triplicate by direct fit to the Michaelis Menten equation (Figure 43).<sup>164</sup> In all cases, these data are of good quality, as shown with the simplest substrate tigloyl pantetheine **77** (Figure 42 and 43). With this, it was possible to determine the Michaelis Menten constant  $K_M$  and the rate of reaction  $V_{MAX}$ . Further calculations gave the value of  $V_{MAX}/K_M$  (the specificity constant, Figure 44), which is the gradient of the curve at  $[S] = 0$ .



**Figure 43:** Michaelis Menten curve with tigloyl **77** with the mutant F2571A as a general example, substrate concentration against the rate.



Before the specificity constants of the mutants could be compared with the constants of the wild type (WT) ER, first the specificity constants of the WT ER had to be measured. In addition, the obtained values should be compared to the specificity constant of the WT ER for the specific substrate from the previously published studies<sup>137</sup> (Figure 44). This should ensure that the assays are reproducible and show whether a deviation in the values for the respective kinetics could be observed. The results are shown in a logarithmic scale in Figure 44. It can be observed that the experiments were reproducible and lead to nearly the same specificity constants. Hence, the mutated proteins can be measured and compared with the wild type, without apprehension that deviations occur through the measurement.

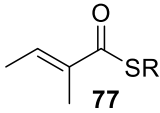
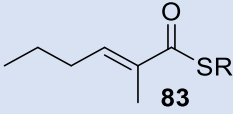
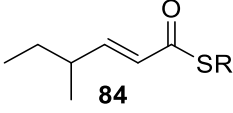
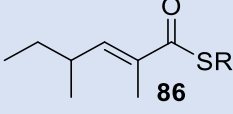
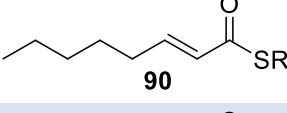
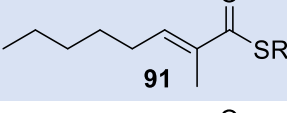
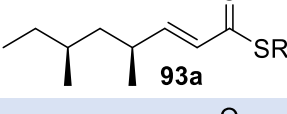
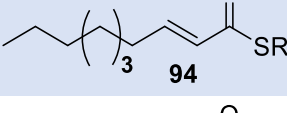
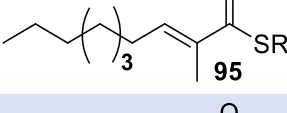
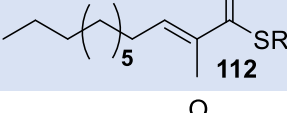
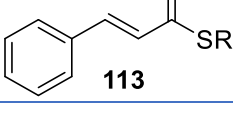


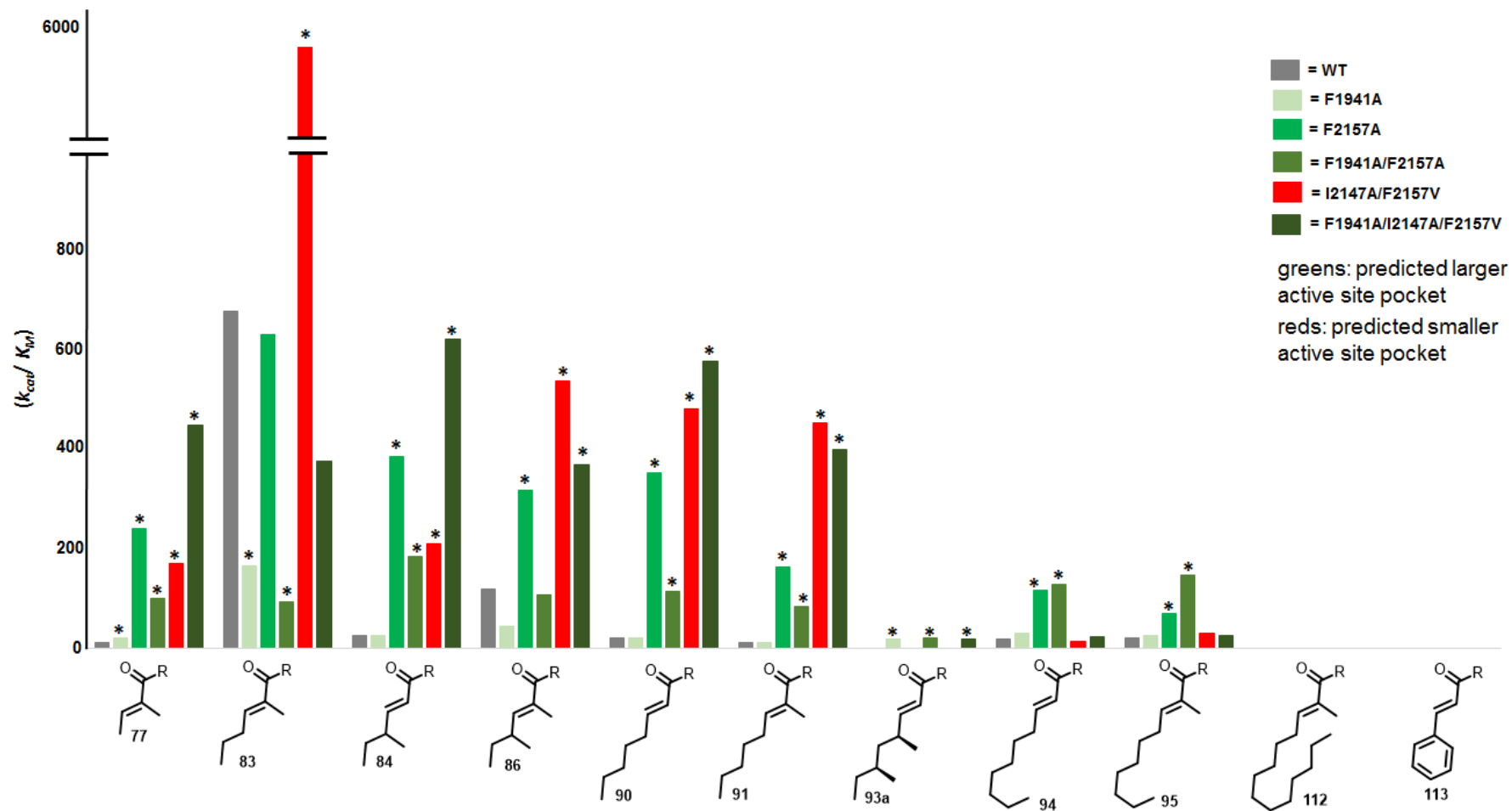
**Figure 44:** Graphical display of the specificity constants of the comparison of the wild type in this study (blue) and the wild type of the previous study<sup>137</sup>(red). Logarithmical scale.

The results of the different mutant ER proteins and the wild type ER protein for the different measured substrates are shown in Figure 45. Statistical evaluations were done with Sigma Plot (version 12.00). The paired t-test was performed to test the influence of the mutation on the specificity constant for the different substrates compared to the WT ER.<sup>165</sup> Where a significant change in the specificity constant was observed after the mutation this is marked by an asterisk (\*  $p < 0.05$ ) in Figure 45. The test for normality was performed by the Shapiro-Wilk test. Groups with (n = 3) were examined.<sup>166</sup>

Overall, all substrates were converted by the mutant isolated ER domains. The exceptions are the heptaketide **112** and the cinnamic acid derivative **113**, which were not converted by either the WT or any mutant enzymes under the conditions tested.

**Table 9:** Overview over the specificity constants for the ER wild type and the different ER mutants. R = CH<sub>2</sub>CH<sub>2</sub>NHCOCH<sub>2</sub>CH<sub>2</sub>COCOHC(CH<sub>2</sub>)<sub>2</sub>CHOH

Substrate	WT	F1941A	F2157A	F1941A / I2147A / F2157V		
				F2157A	F2157V	F2157V
<b>Specificity constant (<math>V_{max}/k_m</math>) [<math>M^{-1}s^{-1}</math>]</b>						
	11.2	22.7	257.8	107.6	128.0	483.5
	730.2	179.0	679.1	101.0	5483.0	403.5
	27.6	28.7	415.3	197.3	226.0	669.5
	128.6	48.5	342.6	115.3	577.0	397.2
	22.0	21.4	380.0	123.4	518.0	621.6
	12.7	12.9	176.6	90.5	487.5	430.4
	0.0	19.7	0.0	23.4	0.0	20.1
	20.7	32.7	126.7	137.3	13.5	25.5
	23.6	28.4	76.4	157.4	17.2	27.5
	0.0	0.0	0.0	0.0	0.0	0.0
	0.0	0.0	0.0	0.0	0.0	0.0



**Figure 45:** Specificity constants for the different tested substrates displayed for the wild type ER and the different mutated ER domains. The paired t-test was performed as investigation to compare the wild type to the mutated enzymes. The asterisk marks a significant change in the specificity constant.

Figure 45 shows the various specificity constants for the investigated substrate analog with the WT ER and the various mutant ER domains. Different effects of the mutations on the specificity constants of the respective substrate analogs can be observed.

Starting with diketide tigloyl substrate **77**, all introduced mutations increase the specificity constant significantly. The smallest increase in the specificity constant is observed in the F1941A mutation. The triple-mutant F1941A/I2147A/F2157V shows the greatest effect (Figure 45). This is probably because the  $\alpha/\beta$ -unsaturated carbonyl of the substrate has the shortest distance to the reactive  $\beta$ -carbon from the cofactor's reactive 4'-*pro-R* hydrogen compared of all mutants (chapter 3.5).

For the triketide substrate **83**, each of the mutated ERs F1941A, F1941A/F2157A, and I2147A/F2157V show a significant influence on the specificity constant (Figure 45). This substrate **83** is contrary to the other substrates. Hence, increasing the volume of the product decreases the specificity constant, but decreasing the pocket volume increases the specificity constant for this substrate.

The mutations also influence the specificity constants for the triketide methylation isomer **84**. The only exception is the F1941A mutation, which shows no significant difference for the specificity constant compared to the specificity constant of the wild type. In general, F1941 makes, no significant effect for most substrates, but it allows the turnover of **93a** as designed (Chapter 3.4). The highest specificity constant is again for the triply mutated ER F1941A / I2147A / F2157V.

In the case of the dimethylated triketide substrate **86**, it could be observed that the F1941A and F1941A/F2157A mutations have no significant difference in their specificity constants compared to the wild type ER. The F1941A/I2147A/F2157V mutant shows a significantly greater specificity constant than the wild type ER (Figure 45).

Interestingly, there is no significant difference between the double mutated ER I2147A/F2157V and the triple mutated ER F1941A/I2147A/F2157V for substrate **86**. The only difference between these mutants was the additional mutation of the amino acid F1941. This mutation, however, seems to lose its influence on the specificity constant with increasing methylation and chain length (Figure 45). The same observation can be made for the non-methylated tetraketide substrate **90** and the monomethylated tetraketide **91**. The specificity constant for the mutated ERs I2147A/F2157V and F1941A/I2147A/F2157V is still significantly different compared to the WT ER (Figure 45). Likewise, the specificity constants of the mutated ERs F2157A and F1941A/F2157A are increased for the tetraketide substrates **90** and **91** compared to the wild type (Figure 45). This is in agreement with the docking experiments (Chapter 3.5), where the *in silico*

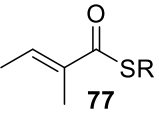
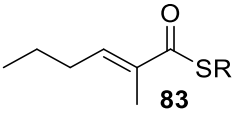
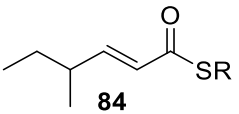
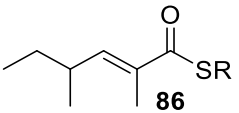
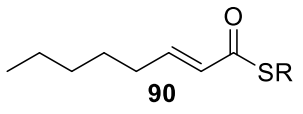
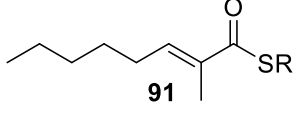
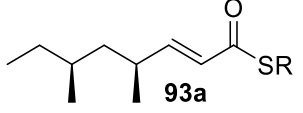
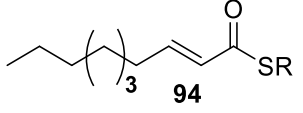
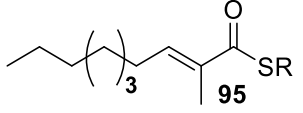
analysis indicated that mutation of the position F2157 would lead to an elongation of the active pocket. Hence, a more productive pose for the conversion of longer substrates could be obtained or rather an increased specificity constant. This effect could be observed both for the mutation to valine at this position and for the mutation to alanine.

The observations for the substrate **93a** are interesting. In every ER containing the F1941A mutation, a conversion of the substrate could be observed (Figure 45). The conversion was not observed in the WT ER. This confirms the hypothesis that this position in the ER domain of SQTKS prevents conversion of **93a**, as suggested by the *in-silico* experiments (chapter 3.4). This appears to be controlled *via* an interaction of the  $\gamma$ -CH<sub>3</sub> group of the substrate with the amino acid F1941 (Chapter 3.4). However, after the mutation of this position to alanine, additional volume is generated in the active pocket at this position, hence, the CH<sub>3</sub> group has more space in the pocket. From this, follows that the substrate may be able to adopt a productive *s-cis* reaction conformation in relation to the cofactor as the docking experiments suggested (Chapter 3.4). Hence, it seems not important if an additional volume increase, like in the mutated ER F1941A/I2147A/F2157V or an additional elongation of the pocket, like in the mutated ER F1941A/F2157A was additionally integrated into the ER domain. Only the mutation of the residue F1941 determines whether the substrate **93a** can be converted or not.

The last two substrates **94** and **95**, which were observed in the UV assay, only with the mutated ERs F2157A and F1941A/F2157A, are significantly different in their specificity constants compared to the wild type ER (Figure 45). In this case, it was observed that the elongation of the active pocket, through the mutation F2157A has a significant influence on the conversion of substrates with longer chain length (Figure 45). This agrees with the *in silico* calculations (Chapter 3.5). Further, significant differences could be observed, depending on whether the position was mutated to alanine or valine. Finally, conversion of the heptaketide substrate **112** and the aromatic **113** could not be observed in the UV-assay in either the wt or any mutant.

Table 10 shows the change in the specificity constant of the mutated ER domains for the specific substrate compared to the wt activity, as a logarithmic representation.

**Table 10:** Change in the specificity constant of the mutated ER domains, as a logarithmic representation. Color code: 0.0 (grey) no change, -0.4--0.0 (light blue) small decrease, -1.9- -0.5 (blue) decrease, 0.1-0.4 (green) nearly no increase, 0.5-0.9 (orange) small increase, 1.0-1.5 (red) increase, 1.6< (dark red) strong increase in the specificity constant. R = CH<sub>2</sub>CH<sub>2</sub>NHCOCH<sub>2</sub>CH<sub>2</sub>COCOHC(CH<sub>2</sub>)<sub>2</sub>CHOH

Mutant/ Substrate	F1941A	F2157A	F1941A / F2157A	I2147A / F2157V	F1941A / I2147A / F2157V
 77	0.3	1.3	0.9	1.0	1.6
 83	-0.6	-0.1	-0.9	0.9	-0.3
 84	0.0	1.2	0.9	0.9	1.4
 86	-0.4	0.4	0.0	0.7	0.5
 90	0.0	1.2	0.7	1.4	1.5
 91	0.0	1.2	0.9	1.6	1.6
 93a	1.7	0.0	1.7	0.0	1.7
 94	-0.8	0.8	0.8	-0.2	0.1
 95	0.1	0.5	0.8	-0.1	0.1

For the F1941A mutant, the specificity constant for di- **77**, tetra- **90**, **91** and pentaketide **95** were nearly the same as in the WT ER. Only for the substrate **93a** was an increase of the specificity constant observed (Table 10). The only exception are both methylated triketides **83** and **86** at the  $\alpha$ -position. For these substrates, a decrease of the specificity constant was observed. This decrease was also observed by every other mutant with a mutation of the position F1941 (*e.g.* F1941A/F2157A and F1941A/I2147A/F2571A). Especially was this observed for the tetraketide **83**. Otherwise, every mutant with the inserted mutation F1941A converted the squalestatin tetraketide **93a**, which agrees with the *in silico* studies (chapter 3.4).

Further, the ER mutants F2157A and F1941A/F2157A show a general increase of the specificity constants for the di-, tri-, tetra- and pentaketides (Table 10). The exception again were the tetraketides, which are methylated at the  $\alpha$ -position. However, these mutants are the only mutated ER domains, which showed an increase of the specificity constants for longer substrates, such as pentaketides. Through the increase and elongation of the active pocket, which agrees with the *in silico* studies (chapter 3.5).

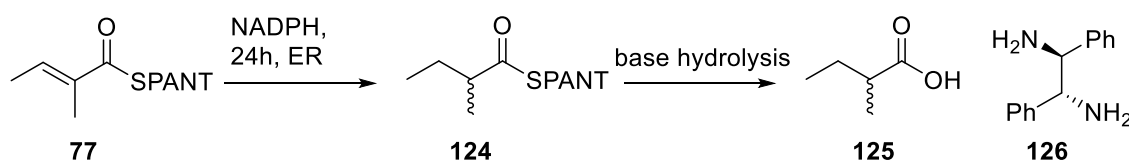
On the other hand, the mutated ER domains I2147A/F2157V and F1941A/I2147A/F2157V showed especially for substrates with shorter chain length (*e.g.* di-, tri- and tetraketides) an increase of the specificity constant (Table 10). This increase is particularly pronounced in contrast to the other mutated ER domains by the triketide, which has a methylation at the  $\alpha$ - and  $\gamma$ - position and tetraketides, and even stronger than for the mutations F2157A and F1941A/F2157A (Table 10). The exception again were the mutated ER domain F1941A/I2147A/F2157V, where the specificity constant for the triketide **83** was decreased.

Overall, the *in-vitro* assay showed three different effects. First, mutated ER domains with the residue F1941A were able to convert the squalestatin tetraketide **93a**. Second, mutated ER domains with a mutation of F2157A possess a higher specificity constant for longer substrates (*eg.* pentaketides). Third, mutated ER domains of the residue I2157A, F2157V and in addition with or without F1941A possess a higher specificity constant for shorter substrates (*e.g.* di-, tri- and tetraketides).

#### 4.5.1 Stereoselectivity of the Mutated ER Domains

The stereoselectivity of the isolated SQTGS-ER domain was investigated in previous work by Doug Roberts (Cox Group) using Parker's *in situ* NMR method for the determination of chirality of  $\alpha$ -substituted carboxylic acids.<sup>161</sup> The *in vitro* assays showed that the isolated ER domain is not able to control the stereochemistry of the  $\alpha$ -position. Hence, all substrates were produced as racemates, although transfer of hydride to the beta position is highly stereoselective suggesting that the substrate takes up a single defined conformation in the active site.

The ER product 2-methylpantetheine **124** was used to investigate the stereoselectivity. By hydrolysing and adding 1*R*,2*R*-1,2-diphenylethylenediamine **126** to the NMR sample, resonances for the methyl groups of **125** are shifted enantioselectively, meaning the *R* and *S* enantiomers of **125** are resolved in the <sup>1</sup>H NMR spectrum.



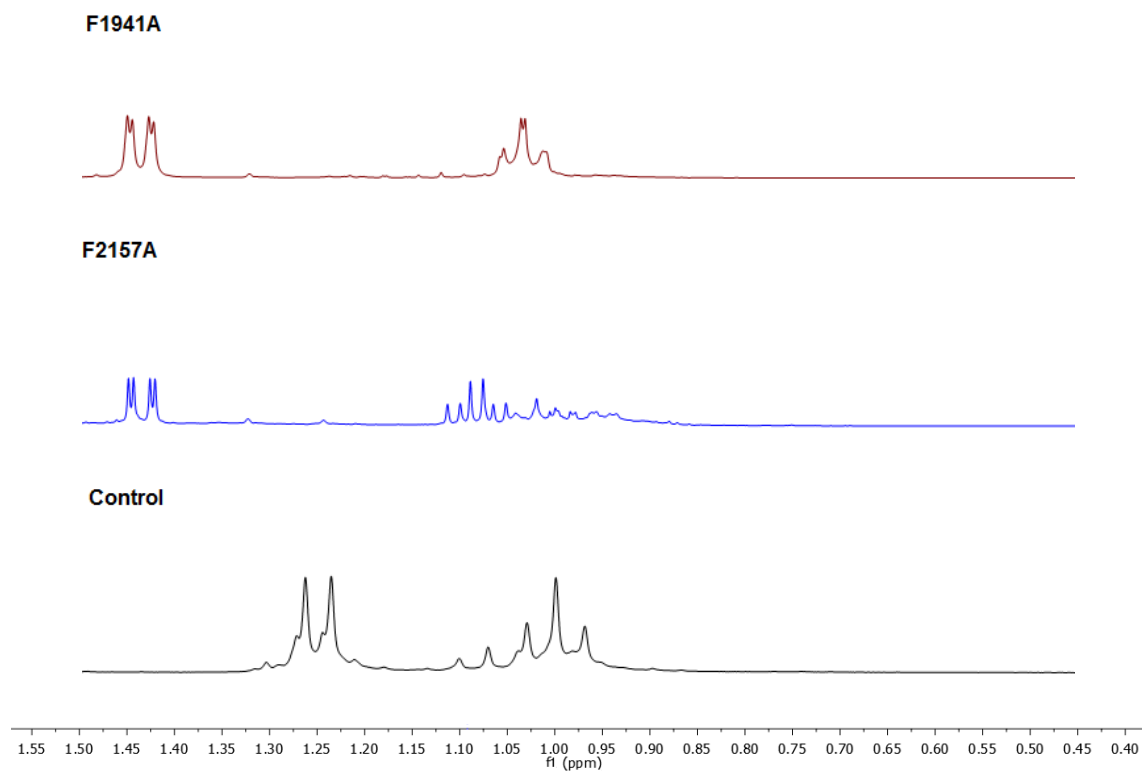
**Scheme 27:** Investigation of C2-position stereochemistry.

Results are shown in Figure 46 for products obtained from the F1941A and F2157A ER mutants. The <sup>1</sup>H NMR spectrum of racemic 2-methylbutyric acid (-)-**126** was measured first (Figure 46, control) at 500 MHz in CDCl<sub>3</sub>. In the presence of 1*R*,2*R*-1,2-diphenylethylenediamine **126** the methyl resonances of **125** are shifted to higher field and the *R* and *S* enantiomers are resolved.<sup>58,161</sup> Next, the isolated mutated ER domains were incubated with tigloyl pantetheine **77** and an excess of NADPH for 24 h. The reaction was acidified to pH 3, extracted directly into CDCl<sub>3</sub> and two equivalents of 1*R*,2*R*-1,2-diphenyl-ethylenediamine **126** was added before examination by <sup>1</sup>H NMR. In mutants, the <sup>1</sup>H NMR indicated that racemic product <sup>1</sup>H was again produced (Figure 46, shown for ER F1914A and F2157A).

In the isolated mutated ER domains, it is also not possible to predict the stereochemistry of the  $\alpha$ -position. The test substrate **125** was produced as racemate. This is similar to the results found in previous studies<sup>58,137</sup> and shows that the inserted mutations did not have any influence on the outcome of the stereochemistry for the isolated ER domain. However further studies by Hao Yao with a multi domain construct, which included the ER domain, build up from the DH to KR showed that bigger pieces control the stereochemistry at the  $\alpha$ -position better (data not shown).<sup>56</sup> It was suggested,



that for the optimal control of the stereochemistry in the ER domain all interactions with the other domains from the SQTGS have to be present. Hence, the extrinsic programming was in the isolated ER domain, through the missing interactions with the other domains, not given.



**Figure 46:** Examples of two <sup>1</sup>H NMR spectra of substrate **125** with **126** and the control (-)-2-methylbutanoic acid.

## 4.6 Discussion and Outlook

The aim of this project was to rationally engineer HR-PKS by introducing a small number of residue changes, rather than the much larger domain-swaps previously used to change the selectivity of these types of HR-PKS domains. Therefore, one or more amino acid mutations were inserted into the ER mono domain simultaneously. This combination of different mutations should reveal if the combination of two or more mutations at the same time would have further effects on the intrinsic programming of the ER mono domain.

First, the results show that it is possible to generate functional structures of SQTKS ER even with a lack of structural data for complete HR-PKS. The validation of the generated SQTKS model showed further that a variety of features of the model correlate well with measured parameters. These include the stereochemistry of the reduction reaction and the conformation and distance of the substrate towards to the cofactor for example.

Subsequently it was possible to generate and investigate nine mutations *in silico*. However, it was possible to produce only seven mutated ER expression vectors. These contained one or more mutated amino acids. From the seven generated mutant vectors, five proteins were successfully expressed and purified and were stable enough for *in vitro* study.

Through the subsequent *in-vitro* assays, it was then possible to show three different aspects concerning the *in-silico* calculations. First, that it was possible to achieve the aim to rationally engineer the ER domain to convert the SQTKS product **93a**, which cannot be reduced by the WT enzyme. In addition, rational engineering of the ER domains enabled the significantly faster conversion of substrates with longer chains (*e.g.* pentaketides). Hence, a higher specificity constant for the specific substrates was observed. Third, to engineer ER domains, which possess a higher specificity constant for shorter substrates (*e.g.* di-, tri- and tetraketides) which is even dependent on the methylation pattern of the substrate.

The results confirm the hypothesis that the position F1941 in the ER domain of SQTKS prevents conversion of SQTKS, as suggested by the *in-silico* experiments (Chapter 3.4). This is probably mediated *via* an interaction of the 4*S*-CH<sub>3</sub> group of the substrate with the amino acid F1941. After the mutation of this position to alanine, additional volume was generated in the active pocket at this position, hence, the CH<sub>3</sub> group has more space in the pocket. From this, follows that the substrate may be able to adopt a conformation in

relation to the cofactor **11**, in which it can be converted, as the docking experiments (Chapter 3.4) suggest.

No similar studies on the ER domain from HR-PKS are known in the literature until today. Hence, generally it cannot be assumed that this position has a similar influence in other ER domains of HR-PKS.

Furthermore, it could be shown that it is possible to change specific amino acids, which play a key role for the chemical selectivity of individual HR-PKS catalytic domains and to change the substrate selectivity of SQTKS. Hence, it could be shown that a decrease in the volume, like mutation I2147A/F2157V, and/or an elongation and volume increase, like mutants F2157A and F1941A/I2147A/F2157V, of the active pocket have significant effects on the specificity constants for various substrates, hence how fast the enzyme converts them. This agrees with the prediction of the *in silico* studies (Chapter 3.5).

Interestingly, both volume expansion and volume reduction have a similar effect for shorter substrates (*e.g.* di-, tri-, and tetraketide). The results showed that the specificity constant is mostly increased. Hence, the substrates are converted faster. This may be because due to the increase in volume, the substrates can more quickly accommodate their correct poses in the active pocket. In contrast, for the mutant I2147A/F2157V in particular, the volume loss has been observed. However, the scaffold break could lead to the substrates already being forced to come into their active pockets in their correct conformation. Which then also increase the specificity constant in comparison to the WT.

Even so, the results in this study give a further insight into the intrinsic programming of this domain, especially in the SQTKS. Nevertheless, it could be assumed that in other HR-PKS the chemical selectivity is also in a large part regulated through specific amino acids. Overall, it was possible to change the substrate specificity of the ER domain *in vitro* and therefore to change the intrinsic programming of the ER domain of SQTKS.

Concerning the reaction stereochemistry, it could be shown, that isolated mutant ER domains are not able to control the stereochemistry of the  $\alpha$ -position. Hence, the test substrate **77** was reduced to give a racemate. This agrees with the previous studies of the wild type ER, where the isolated ER domain of SQTKS was also not able to control the correct stereochemistry.<sup>56,58</sup>

Overall, it was possible for the first time to rationally re-design an ER domain of an iterative HR-PKS. Thereby, it could be shown that the mutations significantly influenced the kinetic conversion, hence the specificity constant for different substrates, were altered in different ways. Mostly a faster conversion of the substrates was observed. Furthermore, it was possible to create mutants which accept the natural product of the SQTks, as a substrate, and thus enhance the substrate variability. This, may further extend the  $\beta$ -processing reactions. The results expand the few examples of other ER engineering approaches, which are found in literature so far. In these studies, only the stereochemistry of the ER domains was targeted. For example, the ER domain in DEBS module 4 was targeted the conserved “HAAAGGVGMA” NADPH binding motif for engineering.<sup>70</sup> Resulting in a different stereochemistry in the product.

Hypothetically, it should also be possible to decrease the active site of the ER domain further. For example, if the residue F2157 is mutated to some more voluminous amino acids, like tryptophan or arginine in combination with the mutation of the residue I2147. Overall, these studies lead to a better understanding of the programming of PKS modules and PKS in general. As a long-term aim, it might be possible to rationally engineer PKS in such a way that new metabolites could be generated. If one or more PKS modules could be rationally engineered like the ER domain it would be possible to get access to a multitude of different compounds. For example, the ER domain in a PKS could be engineered together with the C-MeT or KR domain. This would give access to new metabolites, which differ in the chain length, methylation pattern and/ or in the saturation of the chain.

Finally, the work described here shows that modelling of HR-PKS domains can be a successful method for the generation of hypotheses for the rational re-engineering of HR-PKS. Despite the fact that many assumptions had to be made, the fact that many levels of modelling were applied and the fact that there are almost no existing X-ray or NMR structures of HR-PKS catalytic domains, it proved possible to generate and validate models of the SQTks ER which could be successfully used to both understand and re-engineer the substrate specificity of the domain. This gives significant hope that similar methods could be applied to other HR-PKS with the aim of rationally engineering other aspects of their programming, without the loss of fidelity and titer which often accompanies engineering based on domain swaps.

In the future, the generated mutants should be inserted into the whole SQTks gene on a suitable expression vector. However, mutations will not be able to be inserted *via* SDM, as it was done in this study, on such a large gene (*ca* 8 Kb) or vector containing

the whole PKS gene without the risk of introducing unwanted mutations. The reason for this is, that the SDM is limited to small vectors.<sup>167,168</sup> However, the mutations could be inserted by using homologous recombination in the yeast *Saccharomyces cerevisiae*. This method is used with a DNA fragment which is homologous to the digest a plasmid *via* 30-50 bp homology at both ends.<sup>169,170</sup> Subsequently, generated mutants should be expressed in the producing strain *Aspergillus oryzae* NSARI as host. The grown culture should be extracted and the investigated by LCMS for the production of new compounds.

This would first show that the successfully mutated ER domains *in-vitro* could be applied to the *in vivo* level. This could give information if the extrinsic programming would be disturbed by the mutations. Alternatively, new derivatives of SQTKS could be generated dependent on the specific mutant. If this would be successful not only the ER domain could be mutated *in-vivo*, but combined with the KR and C-MeT domain, as these are still domains that underlie programming in PKS. Such further studies would then give the understanding about not only intrinsic but also extrinsic programming of PKS. In addition, the possibilities for the rational engineering of PKS and thus the access to new metabolites can be investigated.

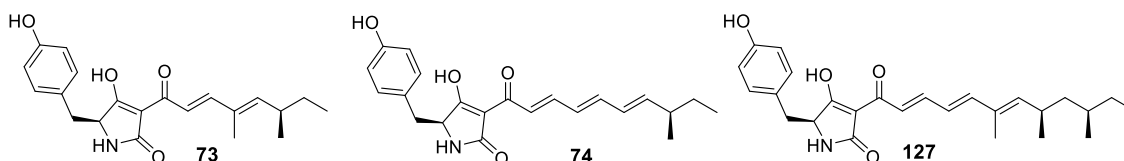
## 5 Modelling Studies of the TENS C-MeT and KR Domains

### 5.1 Introduction

Discovering the mechanism of the programming of HR-PKS is an important goal in order to rationally engineer these systems. This aim is difficult to achieve if no structural data for partial or complete fungal HR-PKS are available. However, since fungal HR-PKS are closely related to mammalian FAS at the peptide sequence level, for which structures are known, mammalian FAS can be used as an overall template for the structure organisation of HR-PKS.<sup>79</sup> In addition, other modules from known PKS can be used as a template for homology modelling of single or multiple domains.

The TENS and DMBS polyketide synthases are 86% identical, but produce polyketides which differ in chain-length and methylation pattern.<sup>76</sup> The products of TENS and DMBS are pretenellin A **73** and predesmethylbassianin A **74**, respectively (Figure 47). In previous experiments, a part of the programming mechanism in the HR-PKS TENS was successfully elucidated (Chapter 1, section 1.17).<sup>48,75-77</sup> It was shown that the exchange of entire functional domains between closely related (but differently programmed) HR-PKS, in this case TENS and DMBS, could lead to the creation of new polyketides (Chapter 1, section 1.16 and 1.17).<sup>76</sup>

In parallel to this study, Dr. Katherine Williams, Dr. Xiao-Long Yang, Dr. Steffen Friedrich and Sen Yin (members of the Cox group) performed a large number of swaps of smaller domain fragments on the KR and C-MeT domains of TENS.<sup>79</sup> In total, the C-MeT, ΨKR and KR domains of TENS were divided into 14 sub-fragments and 31 different swaps were achieved with the corresponding fragments of DMBS. Additionally, a swap using a KR domain from the PKS involved in the biosynthesis of the heptaketide militarinone C **127** (MILS) was created. Most of these experiments resulted in functional hybrid PKS systems and examination of the structures of the respective products led to conclusions about the mechanism of programming within such PKS and the prospects for their rational engineering.<sup>79</sup>



**Figure 47:** Structure of pretenellin **73**, predesmethylbassianin A **74** and premilitarone **127**.

## 5.2 Aims of the Project

The aim of this project was, in parallel to the molecular-biological work, to provide the structural-biological foundations and analysis for the domain swaps between TENS, DMBS and MILS. Subsequently, this analysis should provide the knowledge for future targeted mutations and rational enzyme engineering. There is no existing structural information for *complete* fungal iterative HR-PKS such as TENS. However, in order to understand the programming mechanism and to perform rational site-directed mutagenesis of these systems, it will be necessary to obtain *useful* structural models of the specific catalytic domains.

As mentioned before, there are significant homologies between mFAS and HR-PKS (Chapter 1, section 1.5, 1.10 and 1.11) which suggest that mFAS is an appropriate model for fungal HR-PKS. While the mFAS structure gives good information regarding the likely quaternary arrangement and interactions of the HR-PKS catalytic domains, it was obtained at relatively low resolution (3.2 Å). The mFAS ΨC-MeT domain, in particular, is represented poorly, and significant deletions and sequence variations from the PKS C-MeT domains are indicated by the sequence alignment (Chapter 10.2). In order to obtain better quality information, the individual domains of TENS should be modelled.

The enzyme domains that are of interest for engineering, and thus modelling, are the C-MeT and KR domain of TENS, because changes in these domains resulted in changes in programming.<sup>76</sup> Information on individual domains would be useful to identify the active site and thus to investigate the influence of the domain-swap on the active site, but it would also be useful to generate multi-domain models, which could help to show how the domains interact. The modelling performed in this work should be done by homology modelling. As mentioned before, a critical requirement for homology modelling is the availability of a suitable template. Hence, the first steps will be to find appropriate templates for the individual domains and use these for homology modelling. After the modeling, the quality of the modelled structure will be validated. With the generated protein model in hand, the substrate-binding pocket should be identified. Afterwards, different docking studies with AutoDock Vina will be performed and validated considering the known stereochemistry of the KR and C-MeT domains using substrates and cofactors (Chapter 1.11).

The best generated homology models will then be used as the basis for the design of a chimeric model structure using TENS and the mammalian FAS (Chapter 6). The respective chimeric model and its modelled single domains will then serve as the basis

for further *in silico* experiments considering the domain swap experiments (Chapter 6). Subsequently, this knowledge will be used to apply a first rational engineering approach to the  $\beta$ -processing domains of TENS.

### 5.3 *In Silico* Studies of the TENS C-MeT Domain

The first step was to model the TENS C-MeT domain. A search in the available structural databases revealed that, to date, no crystal of any HR-PKS is available, although PKS C-MeT domains from fungal Type I NR-PKS and bacterial modular systems are known, as well as from bacterial Type II iterative PKS systems. Therefore, homology modeling was performed using the open free software Swiss-Model.<sup>91–95</sup> However, prior to homology modelling the domain boundaries of the C-MeT domain of TENS had to be determined. Therefore, two methods were used: first, the protein sequence of TENS was examined with BLAST (basic local alignment search tool)<sup>138</sup> and subsequently a conserved domain search (CD-Search) was performed to determine the domain boundaries.<sup>139–142</sup> Second, a multiple alignment between mFAS, TENS, DMBS and MILS was performed (Chapter 10.2). The CD-search determined the likely boundaries for the C-MeT domain from H1347 to Q1669 and the multiple alignment confirmed these boundaries through the inclusion of known C-MeT sequence motifs. Swiss-Model selected 50 possible templates based on the selected sequence. These included templates such as the CurJ C-MeT, (Type I modular, PDB 5thy),<sup>171</sup> citrinin polyketide synthase (Type I NR-PKS, PDB 5mpt),<sup>172</sup> AprA C-methyl transferase (Type I modular NR-PKS, PDB 6d6y),<sup>173</sup> DisA (Type I modular, PDB 6cca)<sup>174</sup> and protein RdmB (Type II, PDB 1xds).<sup>175</sup>

For the selection as template the GMQE score (section 2.1.1), sequence identity and the resolution of the templates were considered and compared. Among the proposed templates, the SeMet-substituted CurJ C-MeT from the curacin A biosynthetic pathway (PDB 5thy) was selected as the template. The CurJ PKS module is an unusual cyanobacterial PKS, which contains all functional  $\beta$ -processing domains (Chapter 1, section 1.8).<sup>145</sup> The CurJ and TENS C-MeT domains have the highest sequence identity (35%) of all templates (Table 11). Furthermore, the CurJ template had the highest GMQE score (0.65) of all possible templates. In addition, the CurJ C-MeT has a good X-Ray resolution of 2.1 Å. Overall, modelling created a structural model of the C-MeT of TENS with a QMEAN value of -3.16, which indicates, that the quality of the generated structural model was good enough to proceed.<sup>91–95</sup>

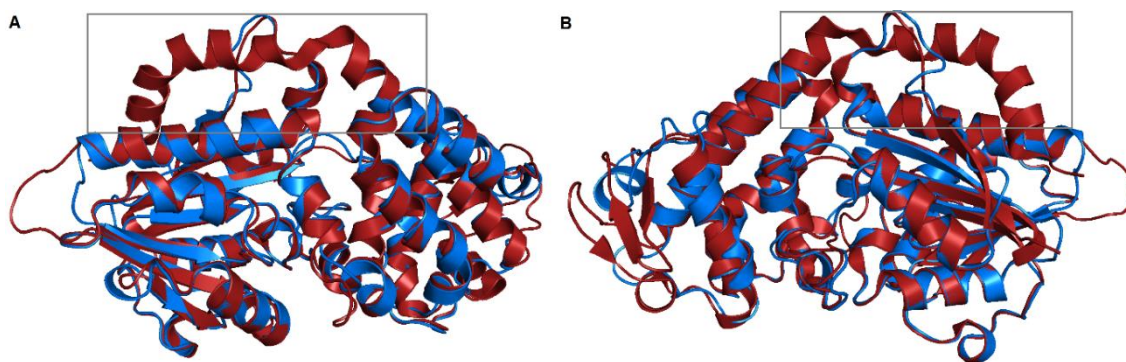


**Table 11:** Sequence alignment of the C-MeT sequence of TENS with the CurJ C-MeT template sequence (5thy); yellow, conserved residues; green, catalytic residues; blue, conserved cofactor motif. \* His-Glu diad.

Template	(1)	MSSGVDLGTENLYFQSNALPPDFLLDPVEVSQQLAPSLITELVTLIDNARTSE
TENS C-MeT	(1347)	HNLSEAIERVSLFYVRQLMGELSTADRRQANWYHTRMLAAFDYHLAKVHEET
Template	(53)	IGTQLEELSVDYIVQGLLQMGWSYQPTESFDLDAAAQCLGVVPTQVRLFERL
TENS C-MeT	(1399)	HLHLRPEWLADDWAVIQTIDEAYPDAVELQMLHAVGQNVADV-----
Template	(105)	LLQILAEVGILQSNQQQWQVQKTAQKVNPSKQSQSLLSQYPDEAATLTLLER
TENS C-MeT	(----)	-----
Template	(157)	CASQLSGVLRGEIDPVQLVFPQGLTTATQLYKDYAVAKVMNTIVEKVMIMKA
TENS C-MeT	(1441)	-----IRGKKHLLLEVLRVDNLLDRLYTEDKGMHMANLFLANAL
Template	(209)	MEKLPSPRGIRLLEIGAGTGGTTSYIILPHLNPNQTEYIFTDIGALFTSKAQE
TENS C-MeT	(1479)	EEITFKFPRCKILEIGAGTGATTWAALSAIGEAFDITYTYTDLVSGFFENAVE
Template	(261)	KFQDYRF-LGYQTLDIEVDPSSQGFESHRYDVIIAANVLHATTSLKQTLSHV
TENS C-MeT	(1531)	RFSAFRHRMVFRALDIEKDPASQSFDLNSYDII IATNVLHATRNLGVTLGNV
		*
Template	(313)	RQLLAPGGILVLYEATTRSRWVDLIFGLLEGWVKFTDYELRPDYPLLNREQW
TENS C-MeT	(1583)	RALLKPGGYLLLNKTPGESLRATFNFGGLEGWLLAEKERQLSPLMSPDGW
		*
Template	(385)	KKVLSETGFTQVVTLPEVEGMAEALSQQTIVIVAQAAS
TENS C-MeT	(1635)	DAQLQKASFSGVDHI--VHDVQEDQQDKQQNSMIMSQ

In the next step, the homology model and the template were aligned in PyMOL (Fig. 48). The alignment resulted in C $\alpha$ -RMSD value of 0.62 Å, which is very good. However, it was observed that a structural  $\alpha$ -helix, which is present in the template, is missing in the TENS C-MeT model (Figure 48, marked with the grey box). In the model of TENS, this region is replaced by a short loop region. This correlates with the sequence alignment between the C-Met sequence of TENS and CurJ in which a gap in the alignment was observed at this position (Table 11).

Skiba *et al.* crystallized the template CurJ as a single protein and determined its domain organization and active site architecture.<sup>171</sup> The missing  $\alpha$ -helix is a structural feature without any obvious catalytic relevance; hence, on the one hand, this structural part could have degenerated in TENS over time, or it might represent a special structural feature of the C-MeT domain of CurJ. The second hypothesis is corroborated by the fact that a similar gap can also be found in the sequence alignment between the SQTKS C-MeT and CurJ (data not shown). However, the missing helix should not influence the overall quality of the generated model as the remaining structural features are similar.

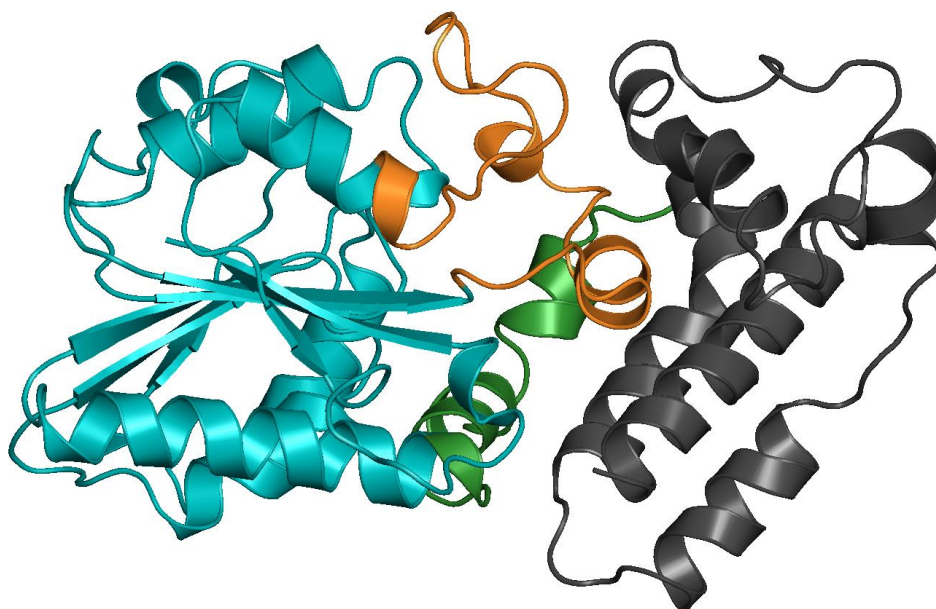


**Figure 48:** Alignment of the TENS C-MeT model (blue) with the template CurJ (PDB 5thy, red) displayed in PyMOL. **A**, Front view; **B**, Back view.

Skiba *et. al.* further identified four main structural features for the CurJ C-MeT domain. Here, these structural features were assigned to the generated model of the C-MeT domain of TENS. The *lid* is displayed in grey (H1347-K1444, Fig. 49). The *core structure* is displayed in cyan (M1467-K1597 and W1634-Q1669, Fig. 49). Displayed in orange, above the core (T1598-G1633), is the so-called *core insertion*. The final structural feature in green is the *junction* (K1445-G1466), located between the lid and the core insertion. Overall, the QMEAN of the model together with the conserved structural features indicate that the generated model of the C-MeT from TENS should be suitable for further *in silico* studies.

Some catalytic residues were already determined amongst others by Skiba *et. al.* and Kishimoto *et. al.*, such as the His-Glu dyad which is invariant in all bacterial and fungal PKS C-MeTs.<sup>171,176</sup> In TENS C-MeT these are residues H1570 and E1596 (Table 11). It is proposed that the imidazole acts as a catalytic base to deprotonate the substrate  $\alpha$ -carbon and form a carbanion nucleophile to facilitate the C-MeT reaction.<sup>171,176,177</sup> Kishimoto *et. al.* and Skiba *et. al.* showed through site directed mutagenesis experiments *in vitro* that the histidine and glutamic acid of the dyad are required for catalysis by the respective C-MeT since in the mutants a significant loss of activity was observed.<sup>171,174</sup> The distance of the imidazole to the  $\alpha$ -carbon can vary between 2.5 - 4 Å.<sup>171,176</sup> This catalysis may be enhanced by the surrounding hydrophobic environment (Template 5thz: Ala281, Phe318, Trp 324 and Pro337).<sup>171</sup> Additionally, they described a conserved tyrosine between 3 - 4 Å away from the substrate which may facilitate the methyl transfer through carbonyl polarization *via* hydrogen bonding.<sup>171,176</sup> The mutation of this residue *in vitro* by Kishimoto *et. al.* and Skiba *et. al.* showed that in the respective mutants also a significant loss of activity was observed.

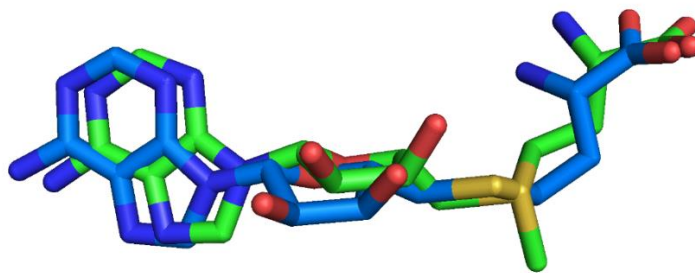
These conserved residues were also found in the model of TENS near the active site pocket (Figure 51) as well as in the sequence alignment of TENS with the template (Table 11). For example the catalytic Tyr1461, the conserved His-Glu dyad His1570 and Glu1596 (Chapter 10.2).



**Figure 49:** Model of the C-MeT domain of TENS. Shown are the structural features. Cyan, core structure; orange, core insertion; green, junction; grey, lid.

#### 5.4 Integration of the cofactor into the modelled C-MeT domain of TENS

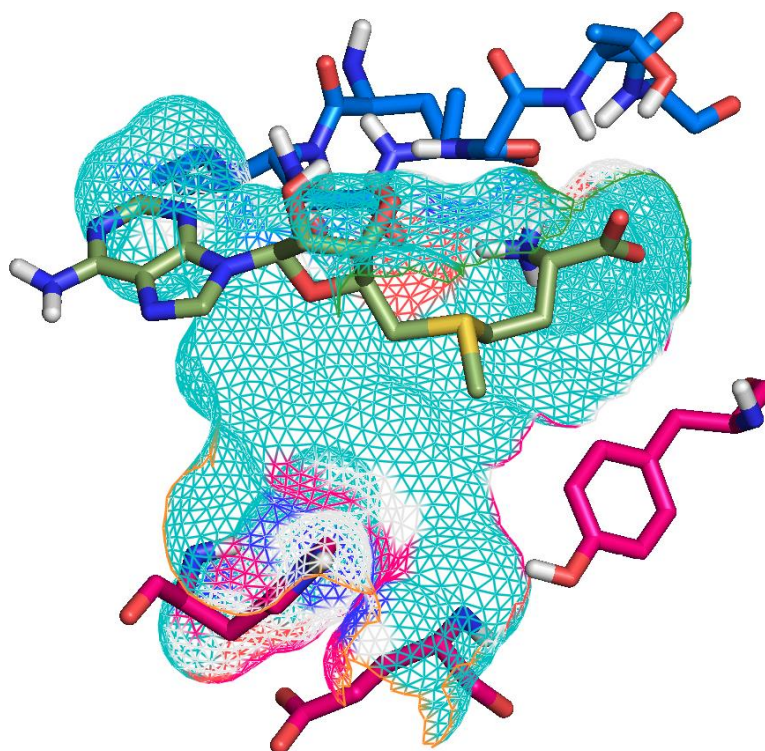
The generated model did not have the *S*-adenosyl methione (SAM) **35** integrated in the active site. Hence, the next step was to integrate the cofactor into the active site of the model (Figure 52), before further docking studies were performed. Therefore, a PDB file of SAM **35** was created. Then, SAM **35** was aligned with the cofactor *S*-adenosyl homocysteine (SAH) **39** from the CurJ template in PyMOL and subsequently the cofactor SAM **35** was integrated into the model.<sup>146,147</sup> Then, the generated C-MeT domain plus cofactor was minimized in YASARA to refine the protein-cofactor interaction (section 2.1.3, Fig. 16).<sup>135</sup> Finally, the extracted and refined cofactor **35** was aligned with the SAH cofactor **39** from the template (Fig. 50). The alignment resulted in an RMSD of 1.4 Å.



**Fig. 50:** Alignment of the cofactor from the TENS C-MeT model (green) with the cofactor from the CurJ (blue)

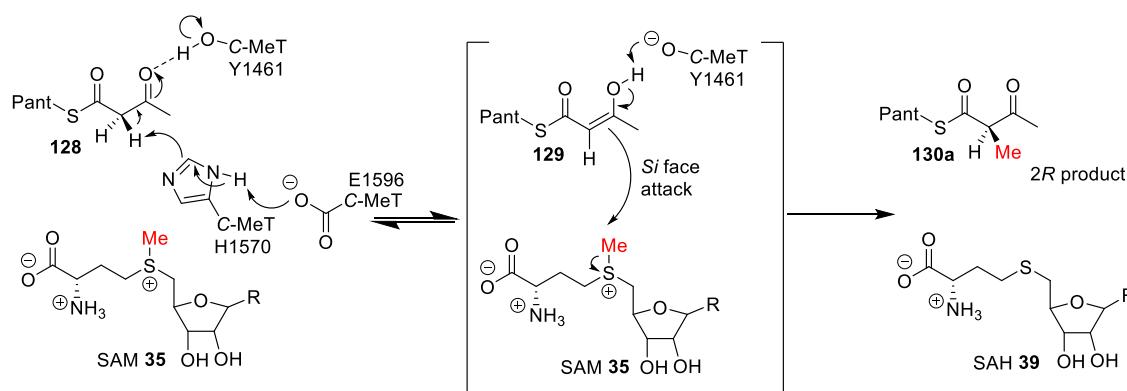
The cofactors overlay reasonably well. Hence, the cofactor in the C-MeT TENS model was in the correct position and orientation and is located in the active pocket. Different SAM cofactor binding motifs, depending on the class of SAM binding protein, are known in the literature. These were determined by sequence analysis and crystal structure analysis.<sup>178</sup>

In the TENS C-MeT, which is a Rossmann fold like protein, the expected interaction of a conserved binding motif “EIGAGTG” with the SAM cofactor **35** was observed.<sup>171,176</sup> Overall, the QMEAN and the structural features of the model generated for the TENS C-MeT indicated that the quality should be suitable for further *in silico* studies.



**Figure 51:** Model of the C-MeT domain of TENS with mesh surface. The cofactor SAM **35** is displayed in green. Red, catalytic residues Y1461, H1570, and E1596; blue, cofactor binding “EIGAGT” motif.

The proposed stereochemical course for the methylation of **128** catalyzed by the isolated C-MeT domain is shown in Scheme 28 (Chapter 1.11).<sup>57</sup> The hypothesized mechanism is through an S<sub>N</sub>2 attack by a nucleophilic carbon of **129** on the electrophilic SAM **35** methyl group, resulting in the C-2 methylated 2*R*-3-oxo diketide **130a**. Important is also that the S<sub>N</sub>2 mechanism follows a transition state in which a carbon under nucleophilic attack is penta-coordinated and approximately sp<sup>2</sup> hybridised. The nucleophile attacks the carbon at 180° to the sulfonium leaving group, since this provides the best overlap between the nucleophile's lone pair and the C–X σ\* antibonding orbital.



**Scheme 28:** Proposed stereochemical course of the methylation catalyzed by the isolated TENS C-MeT domain

## 5.5 Development of Ligand and Cofactor Docked Models for the C-MeT of TENS

To this point, it was possible to generate a structural model of the C-MeT domain of TENS and to integrate the cofactor SAM **35** or SAH **39** (data not shown) into the model. By means of the cofactor positioning, it was then possible to roughly estimate the position of the active pocket in PyMOL. Next, different ligands, such as the substrate 3-oxobutanoyl pantetheine **128**, the intermediate enol 3-hydroxybut-2-enolate pantetheine **129** and the proposed product of the methylation 2*R*-methyl-3-oxobutanoylpantetheine **130a**, and the other possible product stereoisomer 2*S*-methyl-3-oxobutanoylpantetheine **130b**, were docked into the estimated active pocket. In the case of the methylated products both enantiomers were docked. It was initially assumed that the C-MeT domain produces the 2*R* product **130a** which is the substrate for the later KR domain. However, it is also possible that the 2*S* product **130b** is produced and later a racemisation occurs before the KR domain reaction takes place.

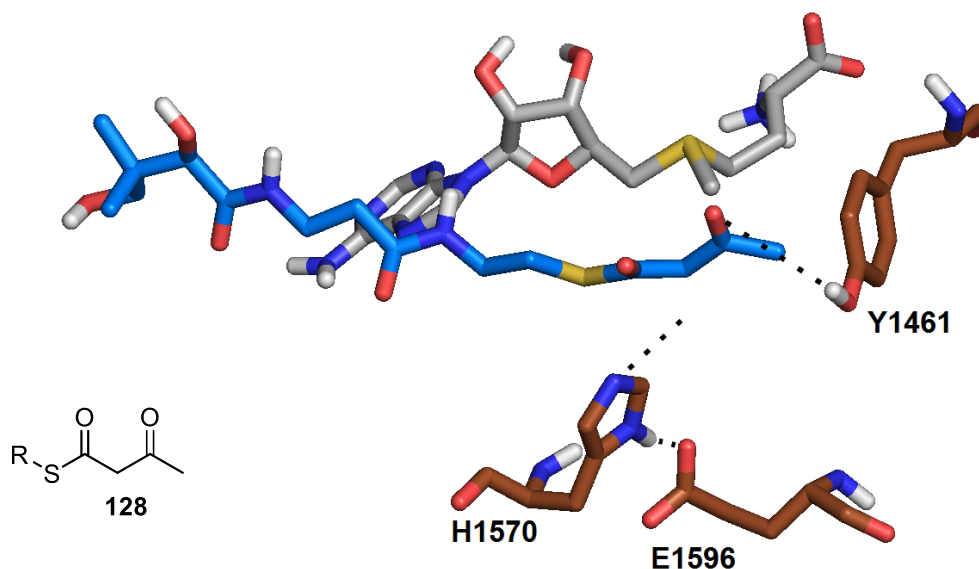
Overall, the docking experiments were done in order to further define the shape and catalytic residues of the C-MeT domain (Figure 52-54).



For the docking, as achieved for the ER domain of SQTCS (Chapter 2.6), PDB data files of the different ligands were generated. The PDB files included three-dimensional parameters of the substrate 3-oxobutanoyl pantetheine **128**, the intermediate enol **129** and the products 2*R/S*-methyl-3-oxobutanoylpantetheine **130a** and **130b**. The subsequent docking steps and quality validations follow the same standard operational protocol as used during the ER docking (Chapter 2.6) and will be not be described here in detail again. The best docking results, which were obtained with optimization of different parameters of the Grid Box in the docking procedure with AutoDock Vina (Chapter 2.1.2), are shown in Chapter 5.5.1-3.

### 5.5.1 Docked Model of the Ketone Substrate **128** and Cofactor SAM **35**

The first docked ligand is the substrate 3-oxo-diketide **128** and is shown in Figure 52. The conserved catalytic residues H1570, E1596 and Y1461, which are important for the catalytic deprotonation of the hydrogen of the substrate at the  $\alpha$ -carbon, are displayed in brown.



**Figure 52:** Active site of the C-MeT domain. In white the Cofactor SAM **35** and. In blue the possible substrate 3-oxobutanoylpantetheine **128**. In brown the catalytic residues which are involved deprotonation of the substrate. R = CH<sub>2</sub>CH<sub>2</sub>NHCOCH<sub>2</sub>CH<sub>2</sub>COCOHC(CH<sub>2</sub>)<sub>2</sub>CHOH

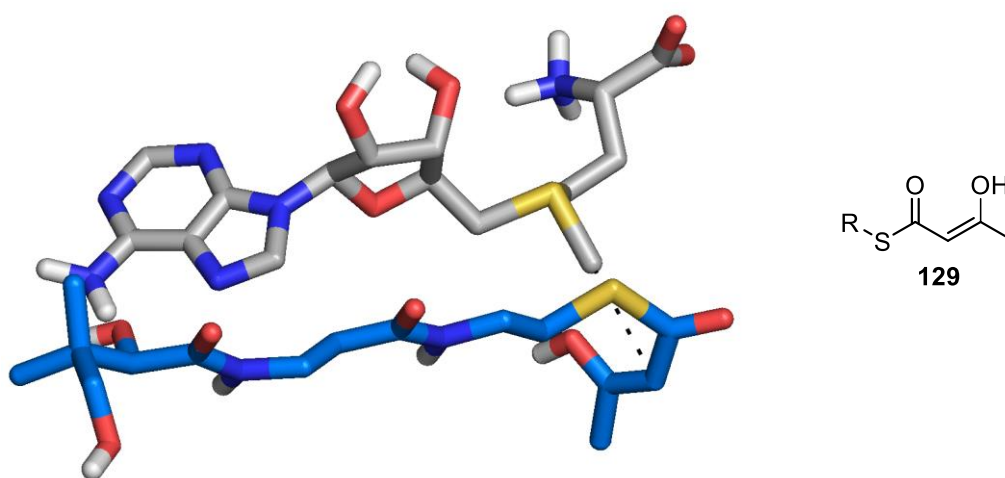
The imidazole, which is proposed to act as a catalytic base, is placed 3.1 Å away from the hydrogen of the  $\alpha$ -carbon in the model (Figure 52). This distance correlates to the observations known in literature (Section 5.3). In addition, the catalytic tyrosine Y1461 which is proposed to stabilize the ketone at the  $\beta$ -carbon is placed 4.0 Å away from the residue. Hence, this residue is also in the observed range known from the literature

(Section 5.3). The observed interaction geometries for these catalytic residues differ slightly from the optimal angles known from literature. For example, for the hydrogen bonding of Y1461 to the  $\beta$ -ketone an angle of  $120^\circ$  was expected (Figure 52).<sup>179</sup> In the model  $107^\circ$  was observed. The same applies to the catalytic His-Glu dyad. The angle for deprotonation between the His1570 and Glu1596 represents also not the optimum angle of  $180^\circ$  but is  $168^\circ$ . The biggest difference in the interaction geometries in the model was observed for the angle between the histidine  $\alpha$ -hydrogen-carbon deprotonation. A value of  $116^\circ$  was observed in the model, whereas an optimal angle of  $180^\circ$  was expected.

### 5.5.2 Docked Model of the Enol Intermediate **129** and Cofactor **SAM 35**

Next, the enol **129** was docked into the model, to observe the intermediate state of the methylation reaction *in silico* (Figure 53). The nucleophilic carbon of **129** is  $3.5 \text{ \AA}$  away from the electrophilic SAM **35** methyl group (Figure 53). This may be a good distance for the conversion of **129**, but the substrate is not in a position which would allow an  $S_N2$  attack to take place. The reason for this is that the optimal angle of the nucleophilic attack in the  $S_N2$  reaction would be  $180^\circ$ . However, in the docked model was an angle of  $160^\circ$  was observed. Furthermore, the angle of the attack between the plane and SAM **35** should be  $90^\circ$ , but only an angle of  $54^\circ$  was observed. Hence, in this case, the geometry of the substrate would have to change significantly with respect to the cofactor.

Overall, comparing the geometry of the transition state of the  $S_N2$  reaction with the geometry of the substrate **129** after the docking in the active site of C-MeT TENS reveals that it is not satisfactory (Figure 53).

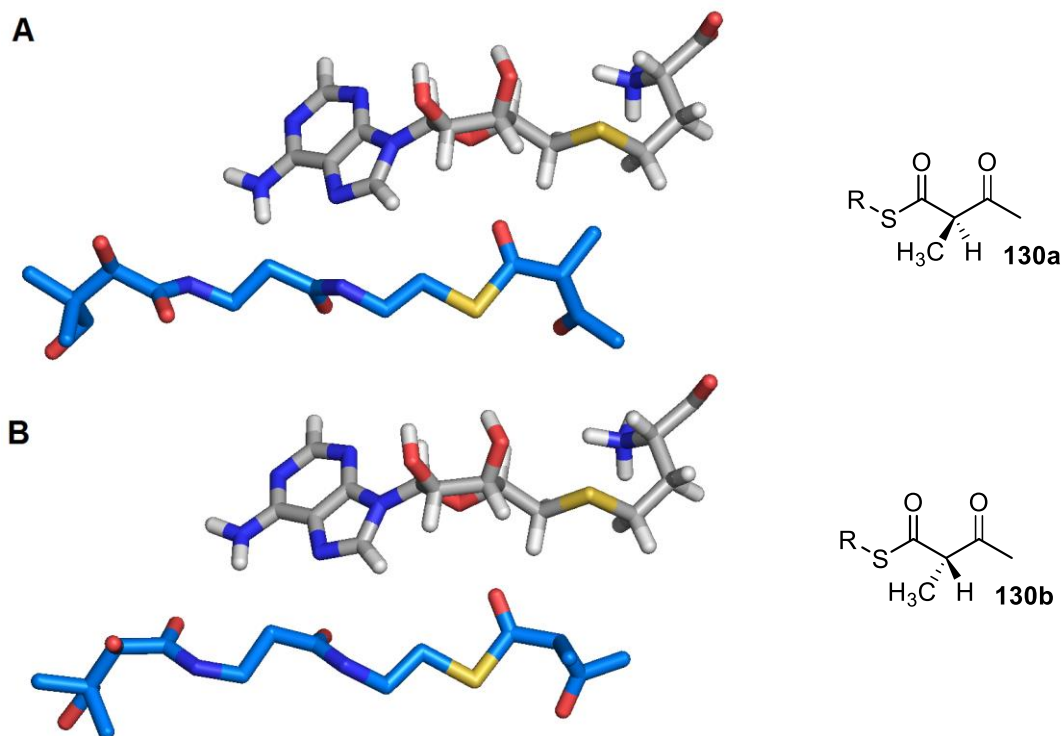


**Figure 53:** Active site of the C-MeT domain. In white the Cofactor **SAM 35** and. In blue the possible intermediate enol 3-hydroxybut-2-enolatepanthetine **129**. R =  $\text{CH}_2\text{CH}_2\text{NHCOCH}_2\text{CH}_2\text{COCOHC}(\text{CH}_2)_2\text{CHOH}$

### 5.5.3 Docked Model of the Products 130a, 130b and Cofactor SAH 39

Finally, the two possible stereoisomer products **130a** and **130b** were docked in the SAH-bound model (Figure 54). If the position between the methylated  $\alpha$ -carbon and the SAH **39** was observed, could be determined that the position did not match the expected position at the time after the nucleophilic attack. The expected angle right after the  $S_N2$  attack would be  $180^\circ$ . Even so, an angle of  $142^\circ$ , which is even lower than for the enol substrate **129**, but still in a suitable range, was observed after the docking. However, the distance of the  $\alpha$ -carbon to the cofactor SAH with  $3.4 \text{ \AA}$  was in the expected distance range.

The observation that the expected product does not fit well in the active pocket after the docking could also be found even more for the  $2S$ -stereoisomer of the methylated product **130b** (Figure 54B). The reason for is, that the  $2S$ -methylated product possesses an orientation in the active site, which places the  $\text{CH}_3$ -group of the  $\alpha$ -carbon away from the SAH cofactor. This orientation of the  $\text{CH}_3$  group after methylation is more than illogical, since, on the one hand, the angle for the  $S_N2$  reaction is not right and on the other hand, the cofactor would then have to be oriented differently to the product. Hence, it is unlikely that this stereoisomer is produced.



**Figure 54:** Active site of the C-MeT domain. In white the Cofactor SAM **35** and SAH **39**. In blue the model biosynthesis products  $2R$ -methyl-3-oxobutanolypanthetheine **130a** and  $2S$ -methyl-3-oxobutanolypanthetheine **130b**.  $R = \text{CH}_2\text{CH}_2\text{NHCOCH}_2\text{CH}_2\text{COCOHC}(\text{CH}_2)_2\text{CHOH}$



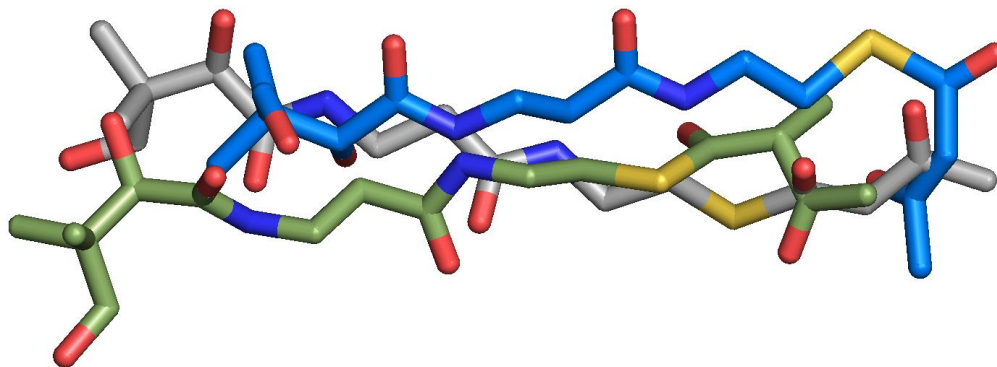
Furthermore, all docking or co-crystallization experiments in literature were only done with the corresponding substrate of the C-MeT domain and not with its corresponding product.<sup>171,176</sup> Therefore, it is difficult accurately to determine if there were expected product-domain or product-cofactor interactions at this point.

#### 5.5-4 Overall assessment of C-MeT modelling and docking

It was observed that in each individual docking experiment the distances and geometries for the different ligands were in a suitable parameter range (Fig. 52-54). However, the position of the respective ligand, especially the panthetine part are inconsistent if compared between each other (Figure 55). It would be expected that the panthetine, and -S-CO- of the thiolester would be very similarly positioned in each case. However, as Figure 55 shows, this was not observed.

The template structure contains the reacted cofactor SAH **39** (the product of the reaction) and does not contain a polyketide substrate or mimic. It may therefore represent a “closed” conformation of the C-MeT and may thus be unsuitable for modelling ligands.

In conclusion, it was not possible to generate high confidence representative ligand-cofactor docked models for the TENS C-MeT. However, the boundaries of the active site were roughly identified. Hence, with regards to domain swap experiments described in Chapter 6 it is possible to observe if a domain-swap probably resulted in an amino acid in the active site being changed or not. Even so, there is no possibility to accurately predict how any specific swap is going to effect the active pocket. In this respect, our results differ from the ER studies (Chapter 2-3), where docked substrates did appear to take up realistic conformations.

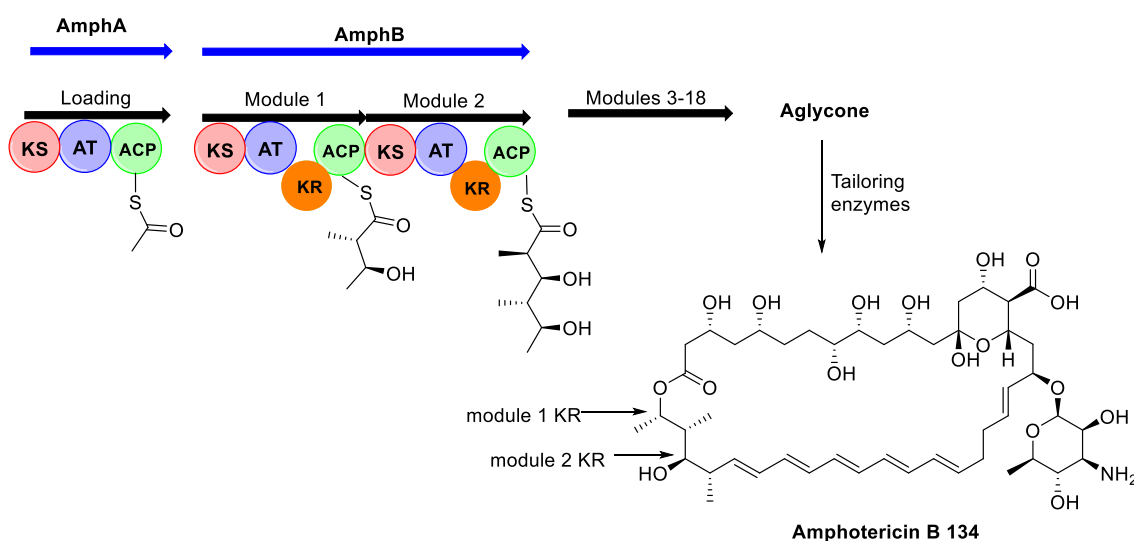


**Figure. 55:** Overlay of the different docked ligands. White, possible substrate 3-oxobutanoylpantetheine **128**; blue, possible intermediate enol 3-hydroxybut-2-enolatepanthetine **129**, yellow, products 2*R*-methyl-3-oxobutanoylpantetheine **130a**.

## 5.6 *In Silico* Studies of the TENS KR Domain

The second domain that was modelled was the KR domain of TENS. However, prior to homology modelling the domain boundaries of the KR domain of TENS had to be determined. Therefore, the sequence of TENS was analysed with BLAST.<sup>138</sup> Subsequently a conserved domain search (CD-Search) was performed to determine the domain boundaries.<sup>139–142</sup> In addition to the CD-Search a multiple alignment between mFAS, TENS, DMBS and MILS was performed in order to find the boundaries of the KR domain of TENS. The search determined the likely boundaries for the KR domain from L2203 to A2484.

Homology modelling was performed for TENS KR using Swiss-Model.<sup>83,92,95,105</sup> Swiss-Model selected 50 possible templates including the polyketide extender module 2 from the spinosin modular PKS (PDB 3slk),<sup>180</sup> AmphB from the amphotericin modular PKS (PDB 3mjs),<sup>181</sup> Plm1 from the phoslactomycin modular PKS (PDB 4hxy),<sup>182</sup> mFAS (PDB 2vz9)<sup>25</sup> and polyketide synthase PksJ (PDB 4jq1).<sup>183</sup> For the selection as a template the GMQE score (section 2.1.1), sequence identity and the resolution of the templates were considered and compared. A low GMQE score, identity or resolution of the template, would result in a low quality model. From the available templates, the AmphB 2 KR, which is an A-type KR domain (Chapter 1.11.1) from the amphotericin modular polyketide synthase (PDB 3mjs) was selected.<sup>181</sup> A part of the biosynthesis of Amphotericin B **134** is displayed in Figure 56.



**Figure 56:** Part of the biosynthesis of Amphotericin B **134**.

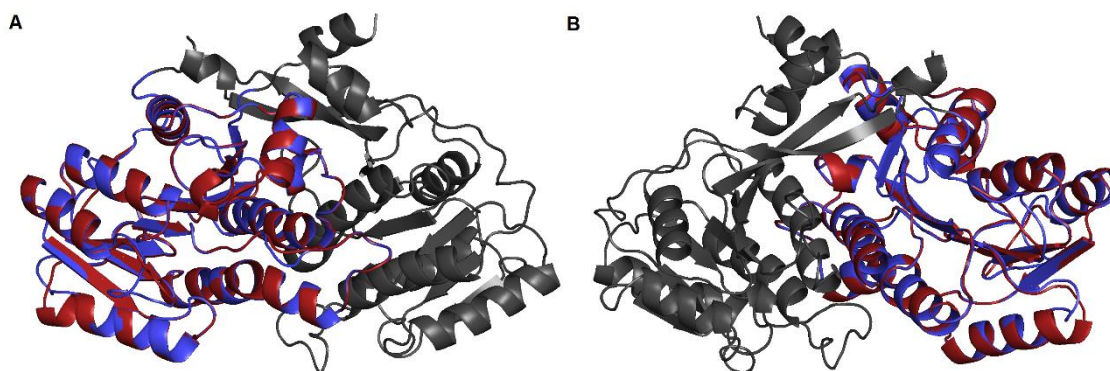
The AmphB module 2 KR was chosen for various reasons: on the one hand, the structures have a sequence identity of 27 % to the TENS KR, which is low, but amongst the best available (Table 12). Further, the template had a high GMQE score of 0.55. The crucial reason was the quality of its X-Ray resolution. The other templates were obtained at low X-ray resolutions between 1.7 - 3.8 Å. In contrast, the X-Ray resolution of AmphB KR is 1.4 Å. In addition, the PKS AmphB might be closer to the actual structure of TENS compared to mFAS KR. However, a disadvantage of AmphB is that its modules are missing other domains such as an ER domain (Fig. 56). The other disadvantage might be that SQTKS and mFAS KR produces the 2*R*,3*R* product **63a**. However, the AmphB KR2 produces the 2*R*,3*S* product **63b**, hence the stereochemistry might be not correct. Still, considering all parameters, AmphB is the best choice as a template. Hence, this should not influence the overall structure of the KR domain since three-dimensional protein structures are evolutionarily conserved. Hence, if the sequence similarity is high, the 3D structure should be also similar. Moreover, AmphB and KR TENS have the highest sequence similarity (Chapter 2.1).

Overall, a structural model of the KR of TENS with a QMEAN (section 2.1.1) value of -2.82 was obtained after the modeling, which indicates that the quality of the generated structure was good enough to proceed.<sup>91-95</sup>

**Table 12:** Sequence alignment of the KR sequence of TENS with the template sequence of 3mjs and a B1 type KR (2Z5L). Light blue, conserved residues for NADPH cofactor binding; green, conserved residues of the catalytical triad; red, B1 specific LDD (VDD) motif, yellow, conserved residues.

Template	(1)	---SVLVTGGTGGIGGRVARRLAEQGAHLVLTSRGADAPPGGAELRAELEQ
TENS KR	(2204)	FDRTYLMVGAAGGLGTSICRWMVRNGARHVVTSRNPKADP---EMLNEAER
B1 KR (2Z5L)	(1)	---TVLITGGMGAI GRRLARRLAAEGAERLVLTSRRGPEAPGAAELAEELRG
Template	(50)	LGVRVTIAACDAADREALAALLAELPEDA-PLTAVFHSAGVAHDDAPVADLT
TENS KR	(2251)	YGAAVQVVPMDACSKDSVQTVVDMIRATMPPPIAGVCNAAMVLRDKL-FLDMN
B1 KR (2Z5L)	(50)	HGCEVVAACDVAERDALAALVTAYPPNA-----VFHTAGILDDAVIDTLSP
Template	(101)	LGQLDALMRAKLTAAARHLHELTADLDDLDAFVLFSSGAAVFGSGGQPGYAAAN
TENS KR	(2303)	VDHMKDVLGPKMQGTEHLDSIFAQEPLDFFVLLSSSAAILNNTGQSNYHCAN
B1 KR (2Z5L)	(97)	ESFETVRGAKVCGAELLHQLTADIKGLDAFVLFSSSVTGTWGNAGQGAYAAAN
Template	(153)	AYLDALAEHRRSLGLTASSVAWG TWGEVGMATDPEVHDRLVLRQGLAMEPEH
TENS KR	(2358)	LYMDSLVTNRRSRGLAASIIHVGHVCDTGYVARLVDDTKVQMSLGTTTVMSV
B1 KR (2Z5L)	(149)	AALDALAERRRAAGLPATSVAWGLWGGGMAAGAGEEESLSRRGLRAMDPDAA
Template	(205)	ALGALDQMLENDDTA-----
TENS KR	(2410)	SETDVHHAFAEAVRGGQPDSRSGSHNIIMGIEPPTKPLDLTKRKPVWISDPR
B1 KR (2Z5L)	(201)	VDALLGAMRGRNDVCV-----

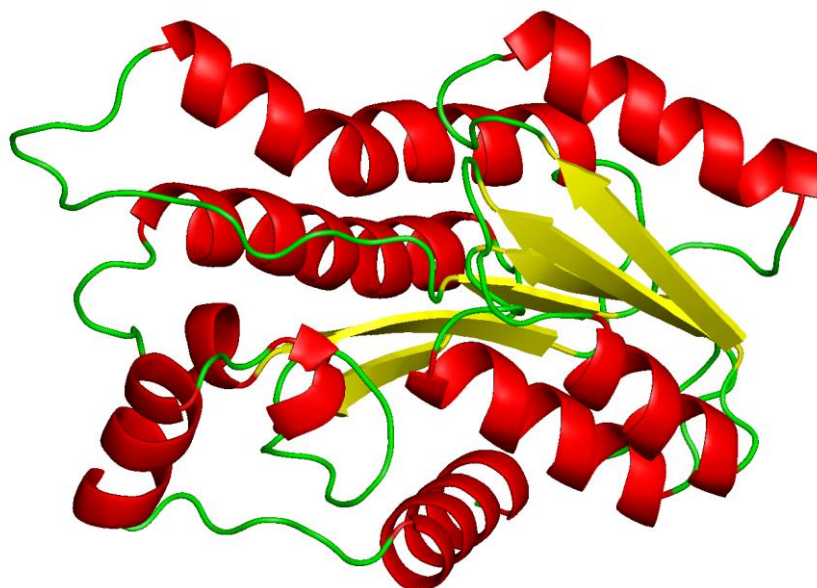
In the next step, the homology model and the template were aligned in PyMOL (Fig. 57). The alignment resulted in C $\alpha$ -RMSD value of 0.1 Å, which is very good, for the blue (catalytic) part of the structure.



**Figure 57:** Alignment of the TENS KR model (red) with the template AmphB (PDB 3mjs, blue, catalytic part; grey, structural part) displayed in PyMOL. **A**, Front view; **B**, Back view.

Zheng *et al.* crystallized the template AmphB as a single protein and determined its domain organization and active site architecture.<sup>181</sup> The overall structure of AmpB KR2 consists of two subdomains, each a variation of the Rossmann fold. The N-terminal 203 residues constitute the structural subdomain, which lacks the traditional SDR cofactor binding site and catalytic residues and apparently serves a structural role in stabilizing the catalytic subdomain.<sup>181</sup> The generated TENS KR domain only aligns with the catalytic C-terminal 210 residues of the AmphB KR domain (Figure 57). The structural subdomain is a particularity of the PKS which lack ER domains. The modelled TENS KR should not include this structural feature (Figure 57). In systems with an ER (such as TENS, mFAS and “full modules” of modular PKS), the ER is inserted into the KR between the structural and catalytic domains. In these systems the “structural” part is then known as  $\Psi$ KR and the catalytic part is known as KR.

However, the domain organization described Zeng *et al.* for the KR domain is very crude. The generated model of the whole TENS KR corresponds to the catalytic subdomain of AmphB KR. Hence, a simple color scheme was used in Figure 58. The  $\alpha$ -helices are colored in red,  $\beta$ -sheets in yellow and loops in green.

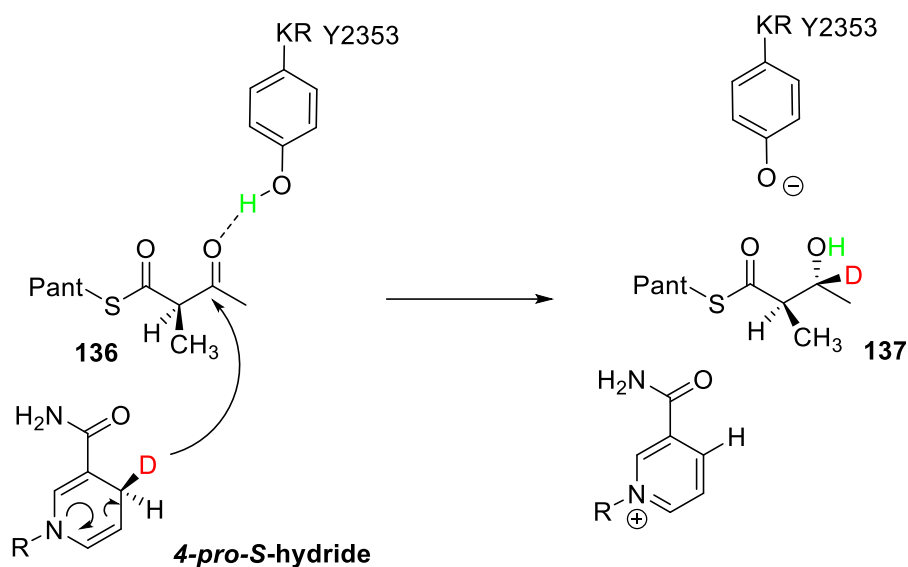


**Figure 58:** Model of the KR domain of TENS. In red the  $\alpha$ -helices; in yellow the  $\beta$ -sheets and loops in green.

In addition, Keatinge-Clay *et al.* and Caffrey and Reid *et al.* observed that the conserved catalytic residues of KR domains cooperate as observed in other short-chain dehydrogenase/reductase (SDR) enzymes.<sup>60,62,184</sup> Furthermore, through this knowledge and subsequent experiments the catalytic residues were identified: Reid and co-workers showed through structural analyses and SDM experiments that Tyr159, Ser146, and Lys163 form the catalytic triad in the tropinone KR II domain from *Datura stramonium* (PDB 2AE2).<sup>60</sup> It was hypothesised that the tyrosine is activated by the neighbouring lysine to donate its proton to the carbonyl oxygen during hydride transfer.<sup>60,183,62</sup>

This conserved catalytic triad was also found in the model of TENS KR near the active site pocket (Figure 61). For example, the catalytic Y2353, the conserved S2340 and N2357 (Table 12 and Chapter 10.2). In addition, the LDD motif, which is characteristic for B1-type KR domains, was also found (Chapter 1.11.1). In the TENS KR, this is represented with the residues L2292, R2293 and D2294 (Table 11 and Chapter 10.2).

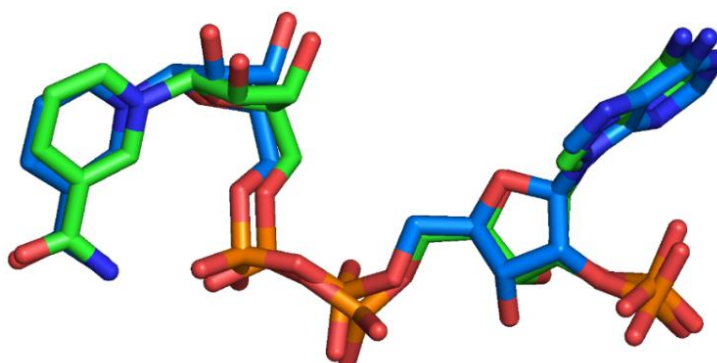
The proposed stereochemical course for the reduction of **136** catalyzed by the KR domain is shown in Scheme 29. Previous studies demonstrated that the cofactor must be rigidly located in the active site for the highly stereoselective transfer of the 4'-*pro-S* hydrogen of NADPH **11**. Furthermore, the transfer of the hydride to the 3-carbon of the substrate is also highly stereoselective, indicating that the substrate must take a single conformation relative to NADPH.



**Scheme 29:** Proposed Stereochemical course of the reduction catalyzed by the isolated TENS KR domain

### 5.7 Integration of the cofactor into the modelled KR domain of TENS

The next step was to add the cofactor NADPH **11** into the active site of the KR domain (Figure 60). This was done by alignment of the homology model with the template in PyMOL (Figure 57).<sup>146,147</sup> The cofactor was extracted from the template and then manually integrated into the structural model of TENS KR. The generated KR domain plus cofactor was then minimized in YASARA to refine the protein-cofactor interaction (section 2.1.3, Figure 16).<sup>135</sup> Subsequently the extracted and refined cofactor **11** was aligned with the cofactor **11** from the template (Figure 59). The alignment resulted in a RMSD of 0.3 Å. Hence, the cofactor in the TENS KR model is likely to have the correct position and orientation.

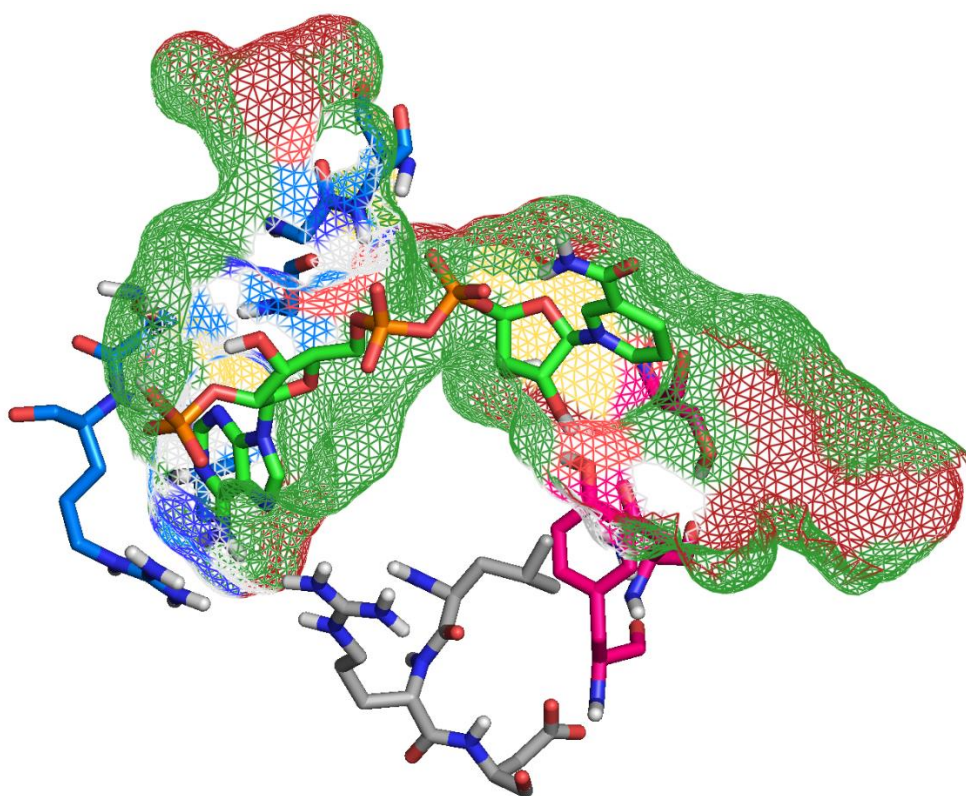


**Figure 59:** Alignment of the cofactor **11** from the TENS KR model (green) with the cofactor from the AmphB (blue)



The NADPH cofactor binding motif for B1-type KR domains are known in the literature (Chapter 1.11.1). These were determined by sequence analysis and crystal structure analysis.<sup>15,184,185</sup>

NADPH **11** contacts one side of the active site, making specific contacts with residues S2239, R2240 and A2265. In addition, the interaction of the conserved dinucleotide binding motif “GXGXXG” with the NADPH cofactor **11** was observed (Fig. 60 and Table 12).<sup>181</sup> Further, the expected nicotinamide 4'-*pro-S* hydrogen is exposed for reaction.<sup>56,57</sup> Overall, the QMEAN and the structural features of the model generated for the TENS KR indicate that the quality should be suitable for further *in silico* studies.



**Figure 60:** Model of the KR domain of TENS with the mesh surface. Green, the cofactor **11**; blue, conserved cofactor binding motif; grey, conserved LDD-motif (L2292, R2293, D2294); pink, catalytic residues (Y2353, S2340, N2357).

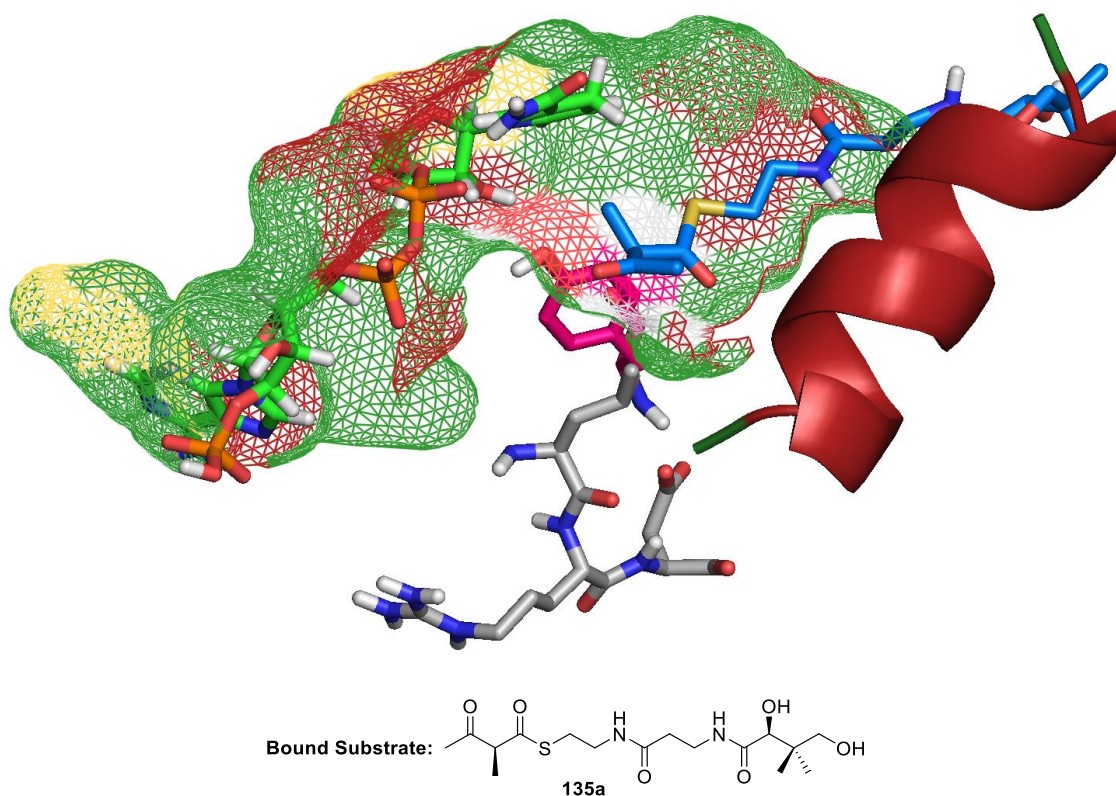
## 5.8 Development of a Substrate and Cofactor Docked Model for the KR of TENS

It was possible to generate the KR domain of TENS using Swiss-Model as well as to integrate the cofactor NADPH **11** into the model. With respect to the cofactor, it is now possible to roughly identify the active pocket in PyMOL. Therefore, a PDB file was generated which included three-dimensional parameters of the respective 2*R*-methyl-acetoacetylpanthetheine **135a**, which is known to be the correct substrate for the KR

domain of SQTKS and mFAS (Chapter 1.11). Since TENS is also a HR-PKS this stereoisomer was chosen for the docking experiments.

The subsequent docking steps and validation criteria followed the usual standard operation protocol (Chapter 2.6) and will not be described again here in detail. The best docking results were obtained by optimization of different parameters of the Grid Box in the docking procedure with AutoDock Vina (section 2.1.2, Figure 61 and 62).

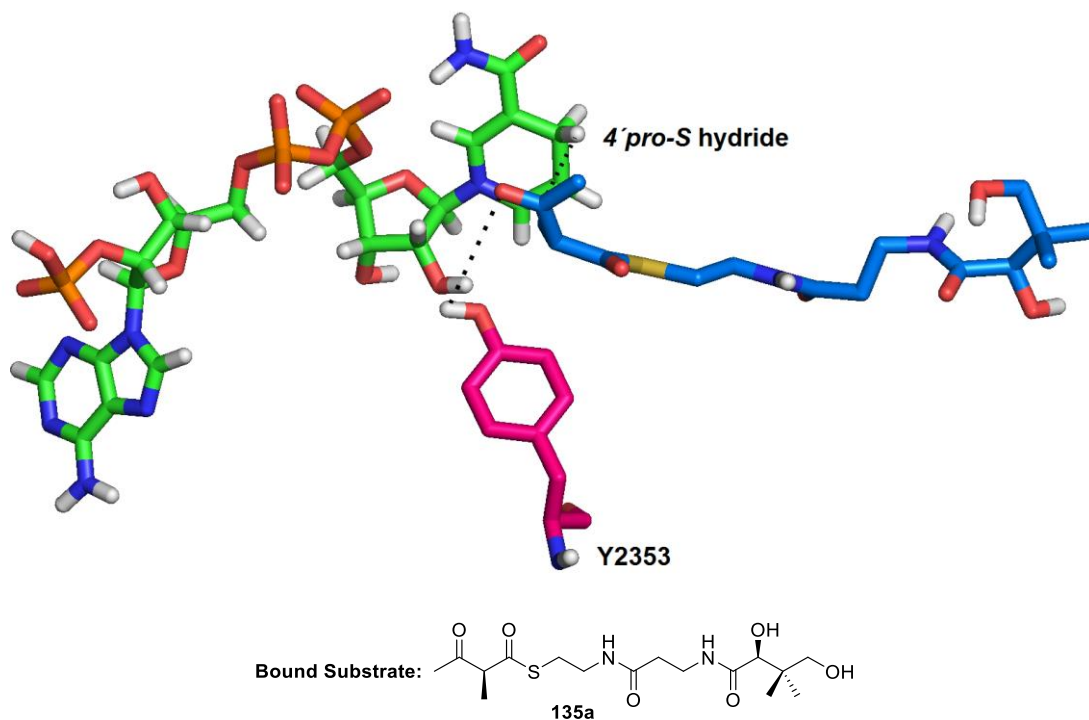
The docked model of the substrate **135a** is shown in figures 61 and 62. The active site of the TENS KR domain with a mesh surface is shown in Figure 61. In addition, the cofactor **11**, substrate **135a**, the catalytic tyrosine and the LDD-motif are displayed. KR domains control the stereochemistry of the  $\beta$ -hydroxyl group of a polyketide by the direction that the polyketide enters the active site in relation to the NADPH cofactor (Chapter 1.11.1).<sup>185</sup> Hence, the first structural feature which is conserved in KR domains is the entrance of the substrate into the domain. In B-type KR domains, such as SQTKS or mFAS, if the phosphopantetheine arm enters from the right side it will encounter the LDD motif (Figure 61, Chapter 1.11.1).<sup>185</sup> The LDD motif prevents the substrate from slipping behind the lid helix.<sup>185,62</sup>



**Figure 61:** Mesh surface of the KR domain with the Cofactor, substrate **135a**, the conserved LDD-motif and catalytic residue Y2353.



In Figure 62, for a better display only the cofactor **11** and the substrate **135a** are shown. In the model, the catalytic nicotinamide moiety of the NADPH cofactor **11** is located inside the KR-domain.



**Figure 62:** Active site of the KR domain; Green, the cofactor **11**; Blue, the substrate **135a**; Grey LLD-motif (L2292, R2293, D2294); Pink, catalytic residue Y2353.

The *2R*-methyl-acetoacetylpanetheine **135a** extends into the protein (Figure 62), from the right site and is located near the cofactor NADPH **11**. This places the reactive  $\beta$ -carbon 3.8 Å away from the cofactor's correct/observed reactive *4'*-*pro-S* hydrogen (Fig 62). The substrate is orientated towards the cofactor with the expected *Si*-face of the 3-oxo group facing the NADPH *4'*-*pro-S* hydrogen. The Burgi-Dunitz angle for the substrate-cofactor complex is 83° and the dihedral angle 37.5°.

The conserved LDD-motif was also observed. However, the substrate is not in direct contact with the LDD-motif (Figure 61) which was to be expected since the chosen diketide substrate is still an early intermediate of the biosynthesis. Later longer intermediates of the biosynthesis might interact with the LDD-motif.

Furthermore, an  $\alpha$ -helix, which is in contact with the substrate at the opening of the active site, was observed (Figure 61). This  $\alpha$ -helix will be referred to as *substrate binding helix* in the following and will be discussed in detail in sections 6.3 and 6.6-7.

Finally, the catalytical residues, which were determined by Reid and Keating-Clay *et. al.*, were observed near the substrate.<sup>60,185,62</sup> Y2353, from the catalytic triad, is located 4.8 Å away from the substrate 3-oxo group. Furthermore, an angle of 84.2° was observed

for the possible hydrogen bond between the tyrosine and the 3-oxo group. However, this angle differs from the optimal angle of  $120^\circ$ .<sup>179</sup> Keatinge-Clay *et. al.* did not determine an optimal bond length between the catalytic residues and the substrate. Hence, the only factor, which is not particularly good in the TENS model, is the angle of the hydrogen bond between the tyrosine and the substrate. In this case, the geometry of the substrate would have to change with respect to catalytic residue. However, the KR domains may also exist in different conformations. Therefore, *in vivo* it is quite possible that the KR domain would slightly change its conformation with could orient the substrate differently to reach the optimal angles.

The results of the *in silico* docking which placed the substrate **135a** towards the cofactor with the expected *Si*-face of the 3-oxo group facing the NADPH 4'-*pro-S* hydrogen, results in similar stereochemistry of the reduction as observed for other B1 type KR domains, such as SQTGS and corresponds with the known stereochemical course of mFAS (Chapter 1.11, 1.11.1 and Scheme 26).<sup>56,57</sup> The actual stereochemistry of the TENS KR domain is not known, but the results give a good evidence that it should be similar to other HR-PKS and mFAS. Overall, a representative substrate and cofactor-docked model of the KR of TENS was developed.

## 5.9 Conclusion

Structural information of the single or mutli-domains are necessary for understanding mechanism of the programming of HR-PKS and the subsequent rational engineering of these domains. Here, the enzyme domains that are of interest for engineering and thus modelling are the C-MeT and KR domain of TENS.

Our first aim was to build and validate models of the TENS KR and C-MeT domains that ultimately met with quality criteria to perform *in silico* studies. The validation of the generated models of the KR and C-MeT domain were done similarly to the ER domain of SQTGS (Chapter 5.3-5.7). One of the validation parameters was the QMEAN score of Swiss-Model. The QMEAN score verified that it was possible to create a detailed model of each respective domain, which was suitable for docking studies. Further, the respective cofactor NADPH **11** or SAM **35** was docked into the domain and showed that this docks in a sensible way preserving known protein-cofactor contacts. In addition, it was shown that the KR domain exposes the correct 4'-*pro-S* hydrogen known to be transferred during the reduction reaction.<sup>56,57</sup> The docking experiments placed the KR substrate in a potentially correct orientation within the active site towards the cofactor

with the expected *Si*-face of the 3-oxo group facing the NADPH 4'-*pro-S* hydrogen. The subsequent hydride transfer would result in the correct stereochemical product, which corresponds to the known stereochemical course in the mFAS and SQTGS KR domains.<sup>56</sup> It should be noted, however, that the actual stereochemical course of the TENS KR is not known. Hence, for the KR domain of TENS it was possible to predict the substrate-binding pocket and to perform docking experiments, which indicated that the docked substrate has the likely correct stereochemical orientation in the active site.

In the case of the C-MeT model, the distance of the cofactor to the substrate was in a suitable parameter range. It was not possible to generate a substrate-cofactor docked model for TENS C-MeT, which would be representative. However, it was possible to roughly predict the active site of the C-MeT domain.

The templates chosen for the modeling had a high resolution of 2.1 Å (C-MeT) and 1.4 Å (KR). Overall the QMEAN, the inflexible structure, the good substrate docking and the good template resolution, indicate that the models generated for KR and C-MeT of the TENS were good enough for further study.

Overall, the models should be suitable to model a chimeric model of TENS and mFAS *in silico*. For the KR domain, the model should be suitable to investigate the influence of the domain swaps on the active site. These models shall be used to understand the molecular basis of the methylation and chain-length programming *in silico*.

## **6 Development of a chimeric C-MeT-ΨKR-KR sub-structure of TENS to Understand the Molecular Basis of Methylation and Chain-Length Programming *in-silico*.**

### **6.1 Introduction**

The generation of models of the TENS KR and C-MeT was described in chapter 5. These model proteins were validated by various methods which suggested that the binding of cofactors is chemically reasonable in both cases but the binding of substrates is only chemically reasonable in the case of the KR domain. Therefore, it was considered that these models could form a valid basis for the design of a chimeric model of TENS and mFAS which could give some structural support to the results of domain-swap experiments.

### **6.2 Aims of the Project**

The first aim was to design a chimeric quaternary structural model of TENS by using the generated TENS KR and C-MeT domain with the mFAS as a scaffold. This generated chimeric model should then serve as the foundation for the design and understanding of the domain swaps. Further, the chimeric model structure in combination with the single domains and the multiple sequence alignment should be used to understand the molecular basis of the methylation and chain-length programming. Afterwards, this knowledge could be used to perform a first rational engineering approach on the KR domain of TENS.

### **6.3 Development of a Chimeric Multidomain Model of TENS**

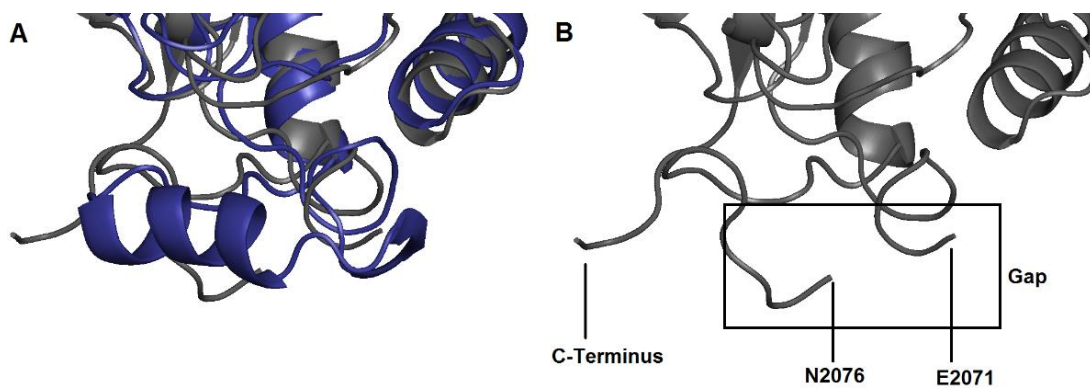
In the first step, a hybrid model for TENS based on mFAS was created (Figure 66), assuming that the general structure of HR-PKS is similar to mFAS (Chapter 1.10). The hybrid model only included the  $\beta$ -processing domains, which are the DH, ER, KR, ΨKR and C-MeT domain. Such hybrid models have been used successfully in other studies. For example, Skiba *et. al.* used this technique to generate a model of the quaternary structure of the CurJ PKS.<sup>171</sup>

The chimeric mFAS/TENS structure should include the DH and ER and ΨKR domains from mFAS and the C-MeT and KR domain from TENS, respectively. The DH and ER were not changed to the corresponding domains of TENS in the chimeric model, because in these particular domains no domain swap was performed. In addition, the ΨKR from TENS was not modelled and integrated into the chimeric model, because of the difficulty to model this domain. This is due to sparse literature information about ΨKR domains as well as their high structural variability. In addition, while the mFAS possesses an active ER domain, the ER<sup>0</sup> domain from TENS is present but inactive and its function is replaced by a *trans*-ER (Chapter 1.13 and 1.17). Yet, the ER<sup>0</sup> from TENS and the ER from mFAS are similar in terms of sequence and likely structure.

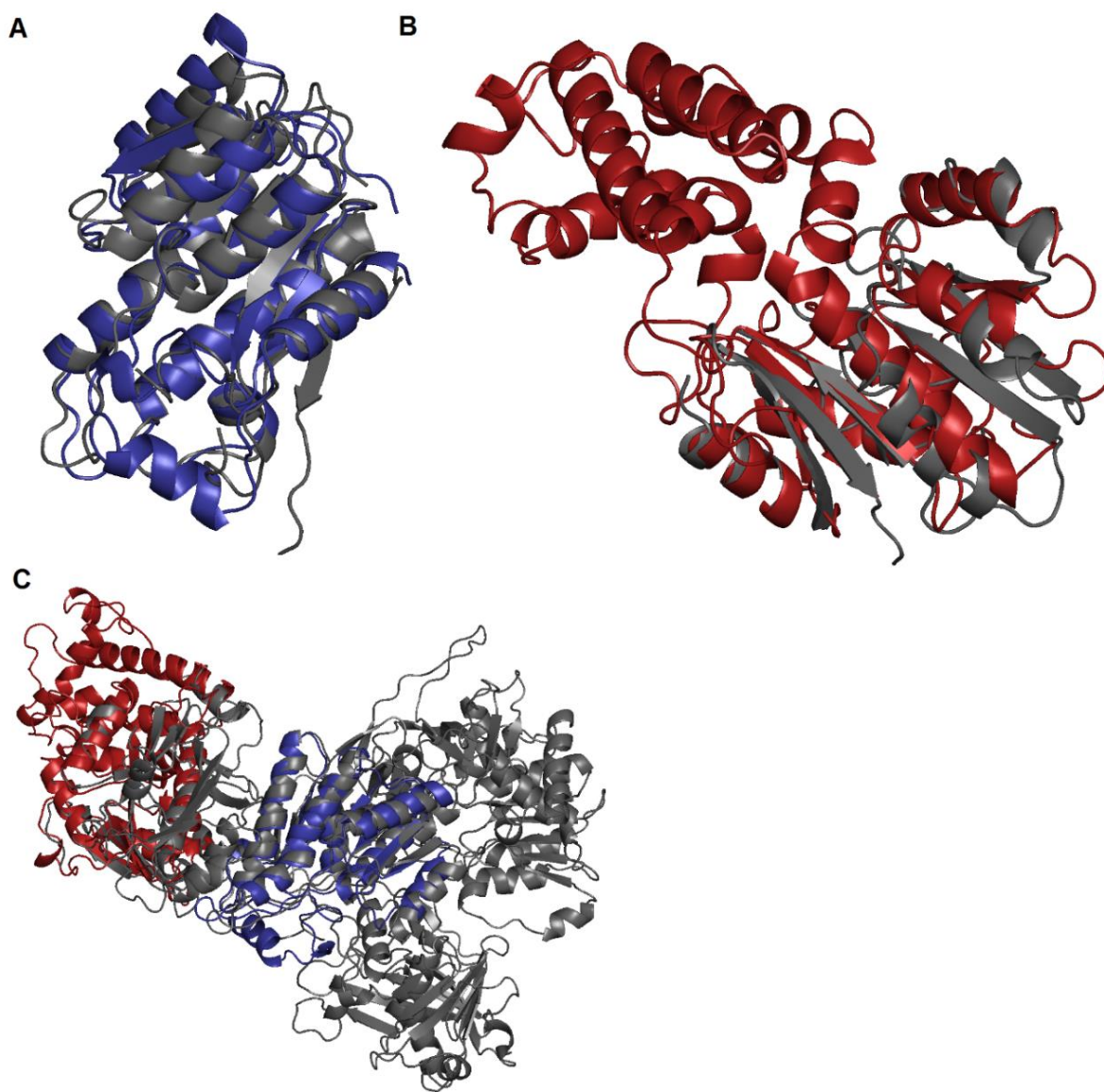
The C-MeT and KR domains in the chimeric model were based on the model structures built in the previous chapter, because domains swaps of these two domains were already performed between DMBS, TENS and in some cases MILS. Additionally, the mFAS ΨC-MeT domain is functionally inactive, degraded in terms of sequence and would therefore poorly represent the active C-MeT domain of TENS.

In order to generate the desired hybrid model, the model KR and C-MeT domains (Chapter 5) from TENS were aligned with the mFAS structure in PyMOL (Figure 64). The alignment resulted in an RMSD value for KR of 1.0 Å and for C-MeT of 2.7 Å, respectively. The RMSD value for the alignment of the C-MeT domain between TENS and FAS is relatively high, but the two structures overlay very well (Figure 64). The higher RMSD value for the C-MeT results from the missing structural part of the C-MeT domain of mFAS.

Furthermore, a gap in the loop structure in the KR domain of mFAS was observed (Figure 63B). In contrast to mFAS, this area in the TENS model KR contains a structured  $\alpha$ -helix which is in contact with the substrate (Chapter 6.6) and will be referred to as the *substrate binding helix* in the following. However, no gap in the protein sequence of the mFAS was observed. This gap can also be observed in an isolated crystallized KR domain of the mFAS with a better resolution of 2.3 Å (PDB 5c37).<sup>186</sup> However, it is most likely that this region is poorly resolved in the crystal, suggesting a highly flexible loop region, in mFAS which is replaced by a longer and more structured helix in the PKS KR domains.

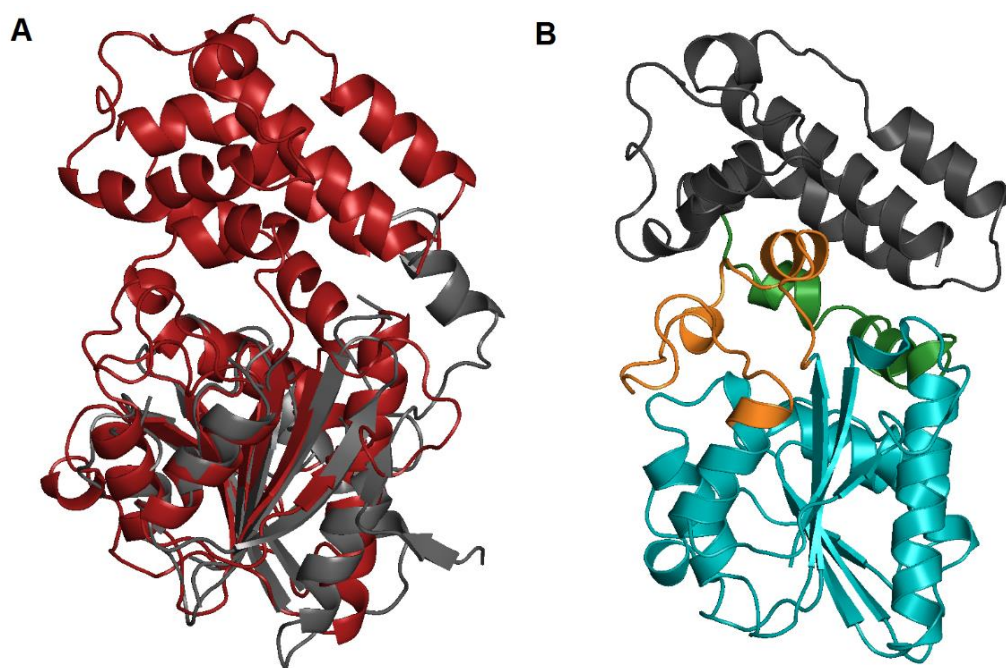


**Figure 63:** **A**, In blue, the KR domain of TENS aligned with the KR domain of mFAS in grey; **B**, part of the KR domain of mFAS.



**Figure 64:** Alignment of the generated mono domains of TENS KR and C-MeT with the corresponding domain models of mFAS; **A**, In blue, the KR domain of TENS aligned with the KR domain of mFAS in grey; **B**, In red the C-MeT domain of TENS aligned with the C-MeT domain of mFAS in grey; **C**, overall alignment of the C-MeT domain (red) and KR domain (blue) models of TENS with the DH, CMeT, ΨKR, ER and KR domain of mFAS (grey).

If the missing part of the *C*-MeT domain is observed in detail, with regards to the structural elements described by Skiba *et. al.*, it can be stated that these structural features are the so-called lid, core insertion and insertion (Figure 65, Chapter 5.4).<sup>171</sup> These structural features are important for stabilization (lid) and catalytic activity (core insertion and core) of the domain. Therefore, one can assume that the lid, core insertion and insertion structures have been lost during evolution of the  $\Psi$ -*C*-MeT of mFAS. Even so, the core structure of the TENS and mFAS  $\Psi$ -*C*-MeT domains overlap very well. In addition, if the alignment of the *C*-MeT domains is examined on a bigger structural scale (Figure 64C). The core structure is positioned directly next to the  $\Psi$ KR domain. Further, the interface of the *C*-MeT domain and the  $\Psi$ KR domain overlay very well. Hence, after the alignment of the TENS *C*-MeT domain core structure with the core structure of the mFAS *C*-MeT domain, the missing parts (*e.g.* lid, core insertion and insertion) should take up the correct 3D orientation.



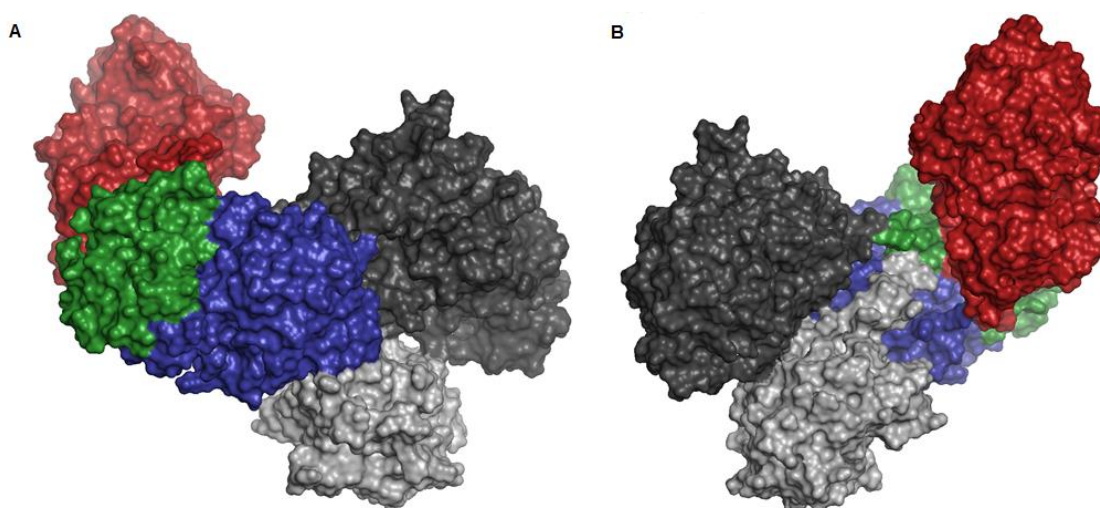
**Figure 65:** **A**, Alignment of the generated *C*-MeT domain of TENS (red) with the corresponding  $\Psi$ -*C*-MeT domain of mFAS (grey); **B**, *C*-MeT domain of TENS displayed with the structural features according to Skiba *et. al.*; cyan, core; orange, core insertion; green, insertion; grey, lid.

Based on the mFAS scaffold (Figure 64) the TENS KR and *C*-MeT domain were now in the correct orientation in PyMOL. The coordinates of the PDB files of the KR and *C*-MeT domains were adjusted considering the mFAS scaffold, since every atom of a protein can only have one specific coordinate in a PyMOL file.



The original mFAS C-MeT and KR domains were then deleted from the PyMOL file and both the remaining mFAS domains (DH, ER ΨKR ) and the modeled KR and C-MeT domain of TENS were combined in PyMOL to yield a new structure. The chimeric mFAS/TENS model was then submitted for refinement to YASARA. The resulting chimeric model of the mFAS/TENS is shown in Figure 66.

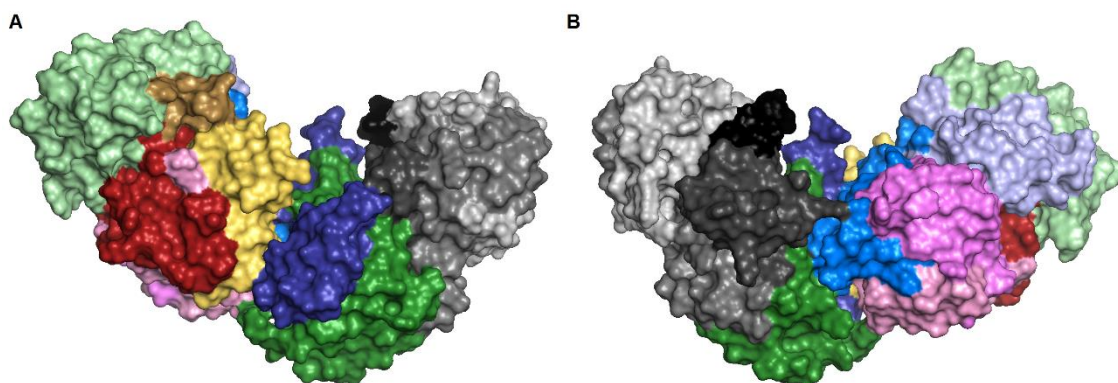
The hybrid model generated for TENS indicates how the ΨKR forms a linking domain between the KR and C-MeT domain and it shows the presumed quaternary structure of the β-processing region of TENS (Figure 66). In addition, it can be observed that the model C-MeT from TENS does not only overlay well with the ΨC-MeT domain of mFAS (Figure 64C), but also fits perfectly in the “old” position of the C-MeT domain in the quaternary structure.



**Figure 66:** Modelled structures of the TENS iterative HR-PKS. Focus on β-processing domains of TENS model. Colors: light grey, DH; dark grey, ER; red C-MeT; green, KR; blue, ΨKR; **A**, displayed from the front; **B**, Displayed from behind.<sup>79</sup>

Using the hybrid model, the TENS C-MeT-ΨKR region was then divided (including proposed linker regions) into ten sub fragments of between 27 and 80 residues (1A1-2B2, ΨKR1 and ΨKR2, Table 11 and Figure 67). The KR of TENS was also divided into four sub-regions of approximately 70 amino acids each (4A-5B, Table 11 and Figure 67).



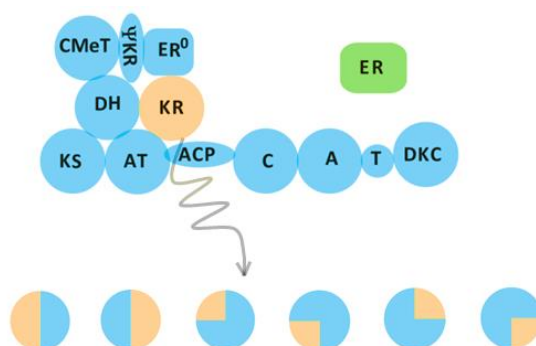


**Figure 67:** Modelled structures of the TENS iterative HR-PKS. Focus on KR,  $\Psi$ KR and C-MeT domains of TENS model. Colors: See table 12; **A**, displayed from the front; **B**, Displayed from behind.<sup>79</sup>

#### 6.4 Overview of the Molecular-Biological *in vivo* work.

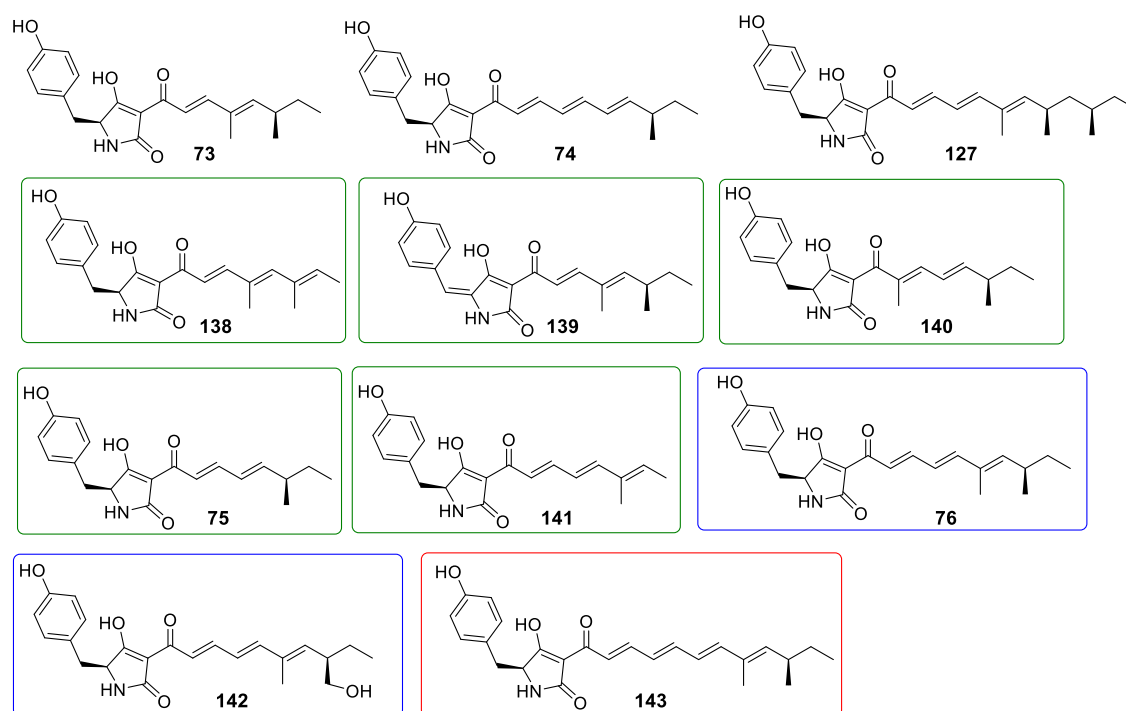
Molecular biology work was performed by K. Williams, S. Friedrich and S. Yin (who worked on the KR domain swaps) and X.-L. Yang (who worked on the C-MeT domain swaps) and shall be briefly discussed.

The fragment-exchange procedure was based on combining overlapping DNA fragments constructed from *tensS*, *dmbS* or *milS* templates with the linearized *tenS* cloning vector and transforming the mixture into yeast.<sup>79</sup> The assembled plasmids were transformed into *E. coli* and domain encoding gene cassettes were transferred by *in vitro* recombination methodology into the expression vectors, which contained *tenC*. These were then transformed into *A. oryzae*. After selection, multiple *A. oryzae* transformants were grown in media containing maltose, which induces expression of the hybrid *tenS* gene, and organic extracts were prepared and examined using a standardised extraction and LCMS protocol.



**Figure 68:** Schematic example for the different domain swaps which were performed in the KR domain of TENS. For example, in some domain swaps a quarter of the *tenS* KR-region was swapped with the corresponding region of *dmbS*, whereas in other swaps up to one half of the domain was switched.

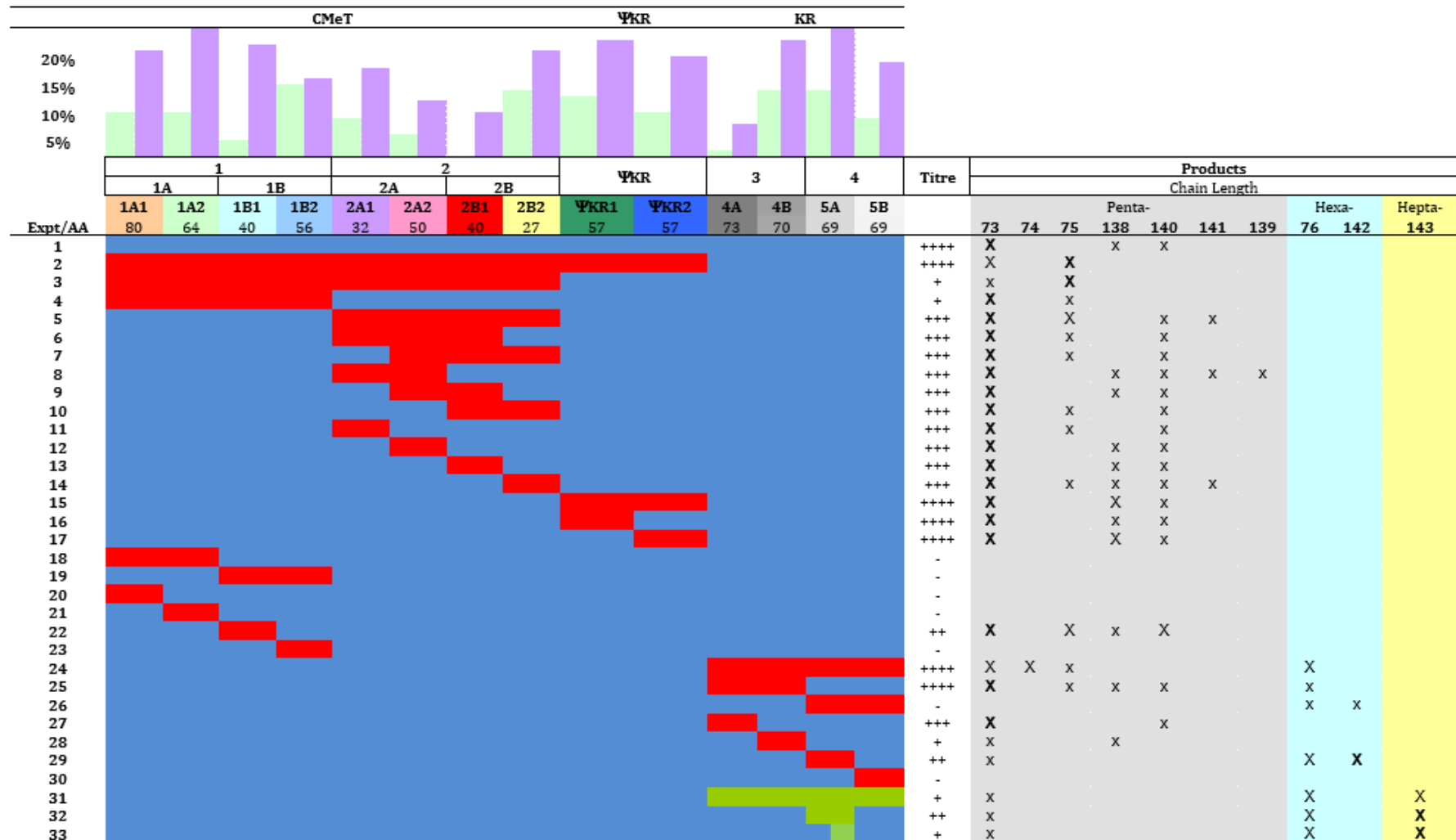
In total nine swaps within the KR domain were achieved, resulting in different products (Figure 69). The most characteristic result was obtained when the TENS KR domain was exchanged with the DMBS KR, which led to the formation of different *pentaketides*, for example **75**, **138**, **139**, **140** and **141** (green) and the *hexaketides* **76** and **142** (blue). When the TENS KR domain was swapped with the MILS KR *heptaketides* **143** (red) were produced.



**Figure 69:** Example of products observed after the domain swaps of TENS with DMBS and MILS. Pentaketides (green), hexaketides (blue) and heptaketides (red) were observed.

The swapped fragments are referred to by a combination of numbers and letters. Fragment I is the first half of the C-MeT, and is further divided into quarters 1A and 1B. In turn, these are further divided into eighths. For example, fragment 1A2 is the second eighth of the C-MeT. The labelling scheme is shown in full in Table 13. For consistency, these fragments are also given fixed colours, which are used, in the following sections.

**Table 13.** Summary of swaps made and LCMS results. Bar graph shows % difference in sequence (similarity green, identity blue), e.g. fragment 1A1 is 92%/81% similar/identical = 8%/19% expressed as a difference. Red bars show donated fragments from DMBS; green bar shows donated fragment from MILS.<sup>79</sup>



## 6.5 Structural Analysis of the Chimeric Multidomain Model of TENS

A detailed structural analysis of the models in comparison to the multiple sequence alignment (Chapter 10.2) was performed. This revealed that the swap fragments 1B1, 1B2, 2A1, 2A2, 2B1, 4A, 4B and 5A contained swapped residues contacting either the cofactor or the substrate (Table 14). Only the fragments 1B1 (C-MeT) and 5A (KR) contain *mutations* in these residues. All other fragments contain no changes in contact-residues and would not be expected to directly alter the active site upon exchange.

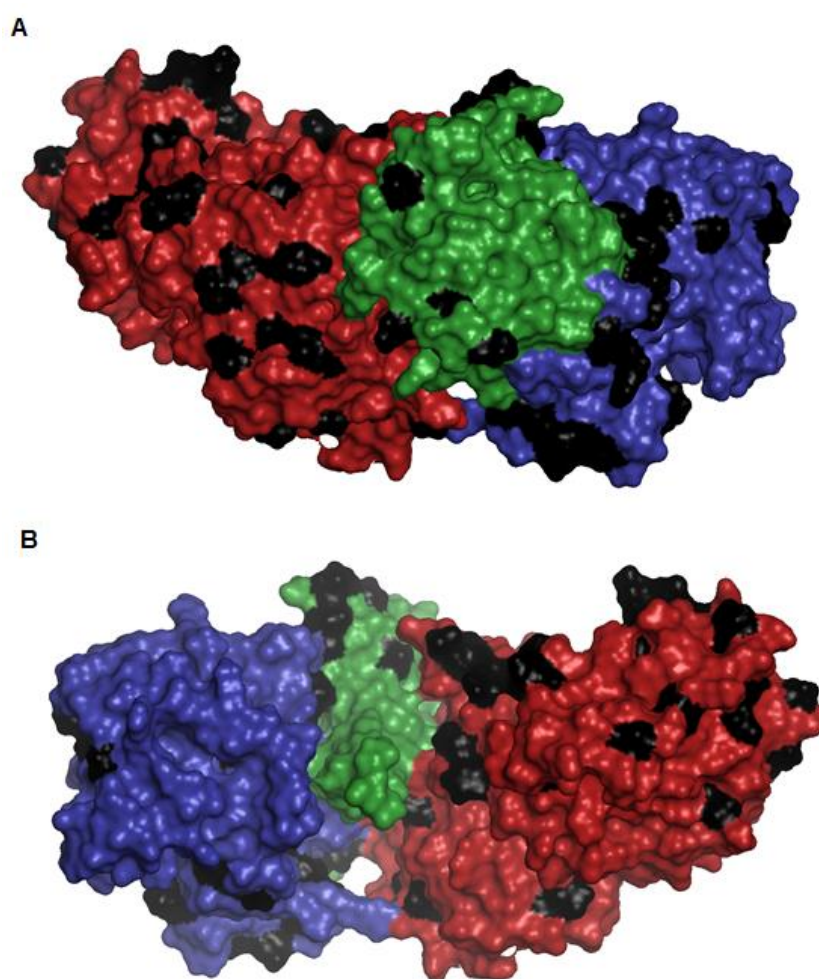
**Table 14:** Summary of the swap positions and structural features.

Fragment		Putative Active site Contact Residues				Other Non-Surface Change	Note	
		TENS/DMBS Identical		TENS/DMBS non-identical				
		Cofactor	Substrate	Cofactor	Substrate			
C-MeT	1	1A1	-	-	-	-	S1350A	Linker
		1A2	-	-	-	-	-	
		1B1	R1452, L1457, D1458,	Y1461	-	L1451M	V1433I, V1437M	
		1B2	1489-KILEIGAGTGAT-1500	-	-	-	A1504V	SAM
	2	2A1	D1519, L1520, S1521, F1524, L1595, D1595, I1596	E1597	-	-	-	
		2A2	T1566, N1567, V1568	H1569, A1570, T1571, K1597	-	-	-	
		2B1	-	A1605, F1609, L1612, E1613, G1614, W1615, L1617, P1627, L1628	-	-	-	
		2B2	-	-	-	-	-	
ΨKR	3	ΨKR1	-	-	-	-		
		ΨKR2	-	-	-	-	-	
KR	4	4A	2213-VGAAGGLG-2220, 2239-SRNPKA-2244 2264-MDAC-2267	-	-	-	I2223L	NADPH
		4B	2292-AAMV-2295, P2315, K2316, L2338, S2340,	L2296, A2342	S2341A	-	V2336I, S2339G	NADPH
	5	5A	Y2353 2379-GHV-2381, 2384-TGY-2386	L2345, N2346, N2347, Q2350, V2387	V2378I	V2397I, S2400N, L2401I, T2404M, V2406A	-	Substrate-binding helix
		5B	-	-	-	-	-	

For the fragment 1B1 in the C-MeT, the mutation is the highly conserved L1451M (Table 14). The swap fragment 1B1 alone leads to formation of pretenellin isomer **140**, which features a non-standard methylation pattern. The product **140** was also observed in other swaps not involving fragment 1B1 (*e.g.* experiments 5–17, Table 13). For fragment 4B in the KR, there is variation at position 2341, which contacts the cofactor only, and change of fragment 4B is not associated with a change in programming. In case of fragment 5A in the KR, there are five changes in the amino acid residues (Table 14, Figure 72), which

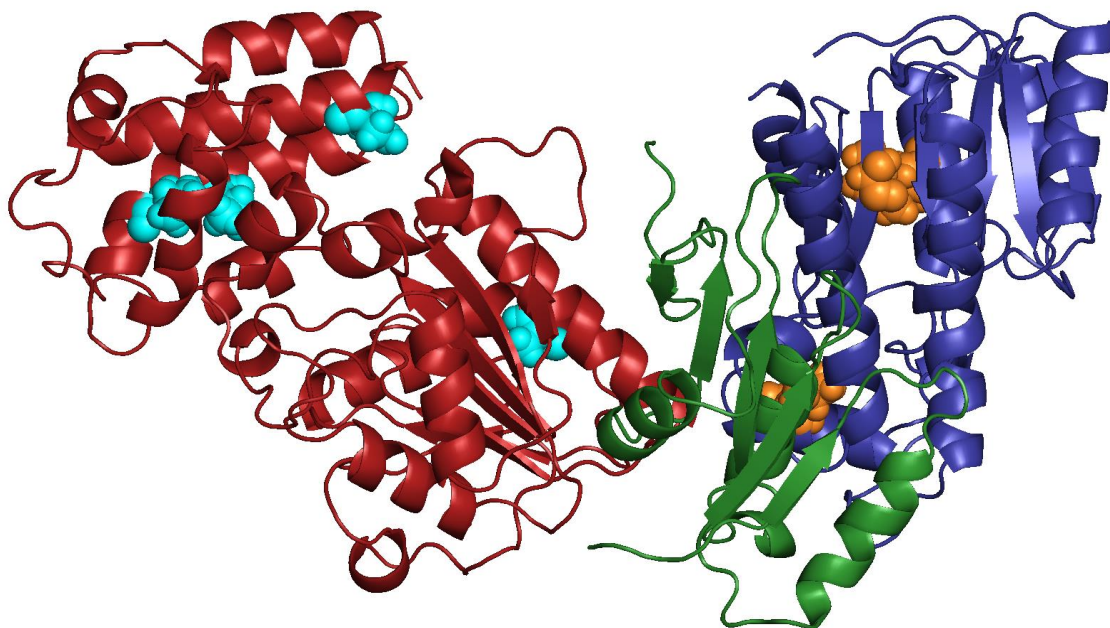
appear to contact the substrate in the structural model (Figure 72), compared to DMBS. These residues are clustered on the substrate contacting helix (Chapter 6.3) which is absent in mFAS (Figure 64, 66 and 72). The significance of the substrate contacting helix will be discussed later in more detail (section 6.6).

Positions which are different between the sequence of TENS and DMBS were mapped onto the structural model (Figure 70), revealing that most of the positions with different amino acids are located at solvent-exposed surfaces (Figure 70). Hence, they are unlikely to be involved in significant substrate selectivity effects.



**Figure 70:** Modelled structures of the TENS iterative HR-PKS. Displayed are the C-MeT, KR and  $\Psi$ KR of the TENS model. Colors: red, C-MeT; blue, KR; green,  $\Psi$ KR. Changed surface residues in comparison to DMBS are displayed in black. **A**, Front view; **B**, Back view.

However, some changed residues are not located at the surface, such as V1433I and V1437M in the C-MeT domain and I2223L and V2336I in the KR domain. Since these residues are in buried hydrophobic regions, they are also unlikely to be in contact with active sites (Figure 71).



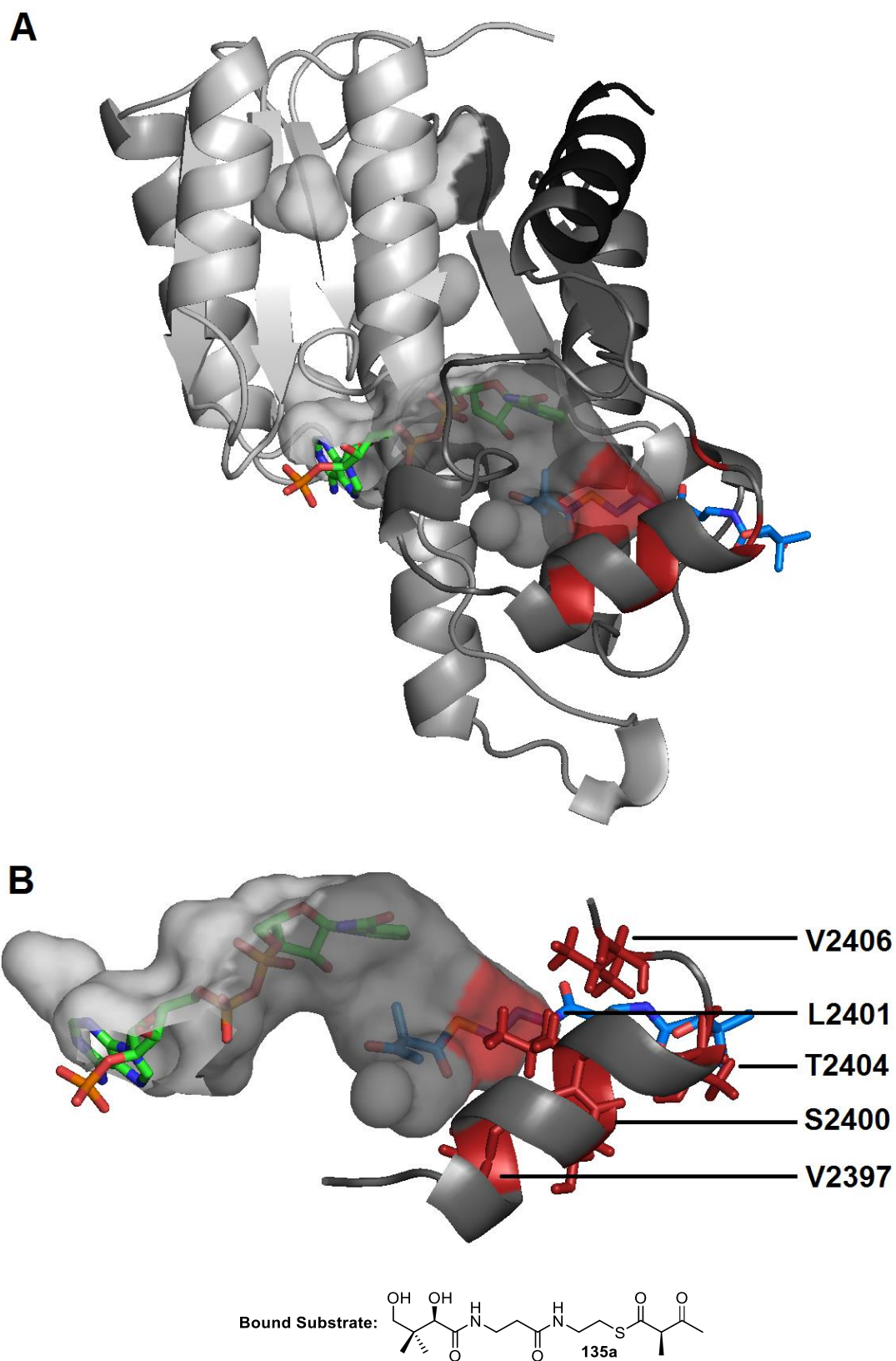
**Figure 71:** Modelled structure of the TENS HR-PKS. Focus on a part of the  $\beta$ -processing domains of TENS model. Colors: red; C-MeT; green, KR; blue,  $\Psi$ KR. Changed non-surface residues in comparison to DMBS are displayed in orange (KR domain) or cyan (C-MeT domain).

## 6.6 *In silico* Studies with the Single KR Model of TENS

In the following, the single KR domain of TENS (Figure 72A-B) and the importance of the substrate-binding helix for the chain length were investigated *in silico*. First, the TENS KR domain, which was modelled in the previous chapter (5.6-5.8), will be analyzed. Afterwards, the KR domains of DMBS and MILS should be modelled. This was done to compare KR domains of TENS, DMBS, MILS and mFAS with special consideration of the substrate-binding helix structures, which the domain swap experiments showed was important for chain-length programming.<sup>76</sup>

The KR domain of TENS with the swap fragments (4A-5B) is displayed in Figure 72A. In addition, the non-identical amino acids between TENS and DMBS (Table 14), which are involved in a possible substrate interaction, are shown. It was observed that all these residues are located on the substrate-binding  $\alpha$ -helix (Table 14). Further, the swap experiments showed that if the specific fragment 5A, which includes the substrate-binding  $\alpha$ -helix, is swapped to either DMBS or MILS sequences, the chain length of the product was influenced (Table 13).<sup>76</sup> Hence, this substrate-binding helix appears to have a critical role in the control of the chain length of the respective product.





**Fig 72:** Focus on the active pocket of the KR domain of TENS with bound NADPH 11 (green) and 2*R*-methyacetoylpanetheine **135a**. Red residues indicate positions of L2363, V2397, S2400, L2401, T2404 and V2406. Cartoon colors correspond to peptide colors according to table 12.

In order to determine the role of the substrate-binding helix in the programming mechanism of the chain length, the KR domains of TENS, DMBS, MILS and mFAS were compared. Therefore, the KR domains of DMBS and MILS were individually modelled *via* Swiss Model. The KR of mFAS was available in the protein data bank (PDB 2vz8). Hence, it was not necessary to model this domain.

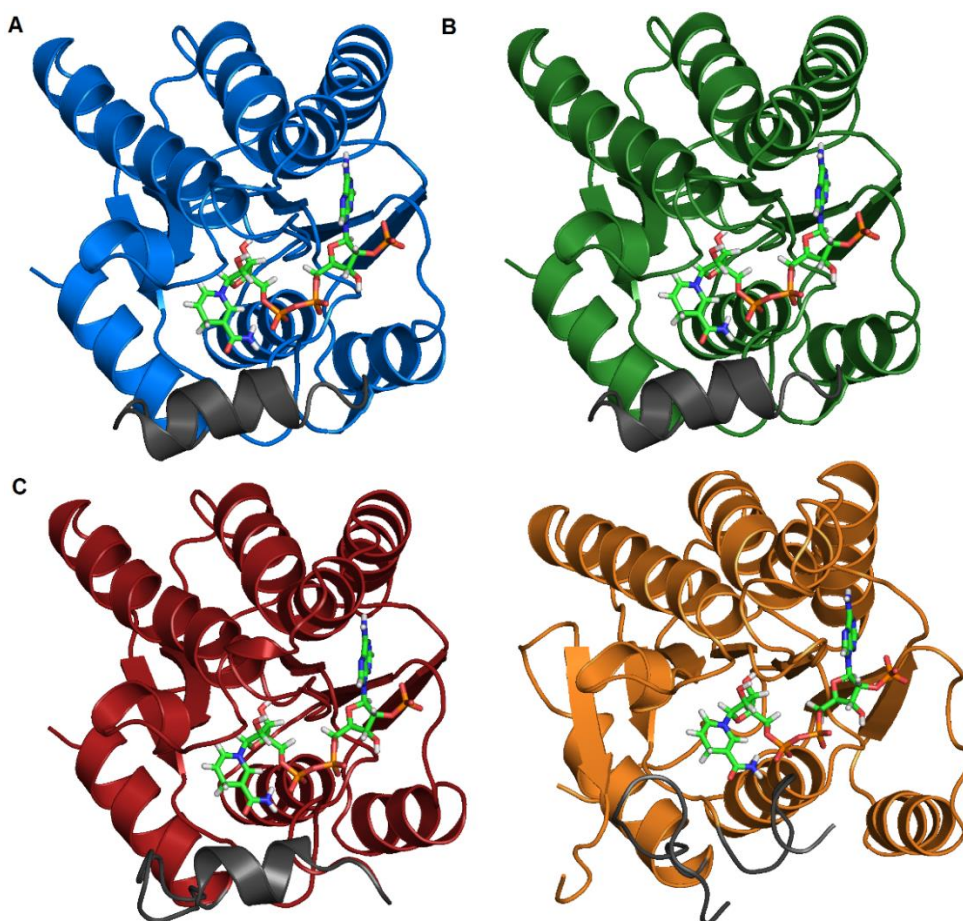
The homology modelling of the KR domain of DMBS and MILS was similar to the procedure for the KR domain of TENS (Chapter 5.6) and will be therefore not be described here in detail again. Prior to the actual modelling the domain boundaries had to be determined. This was done through a conserved domain search (CD-Search) and through an alignment with the TENS KR. Afterwards, AmphB KR was chosen as a template again. This was done for a better comparison of the generated models, since the TENS model was already modelled with AmphB KR as a template.

Overall, the homology modelling resulted in structural models for the DMBS KR with a QMEAN value of -2.35 (sequence similarity 27%) and for the MILS KR with a QMEAN value of -2.81 (sequence similarity 25 %). The QMEAN values indicate that the quality of the generated structural models were good enough to proceed. Afterwards, the cofactor NADPH **11** was integrated into the domains in PyMOL (Chapter 5.7).

The structural models for the KR of TENS (blue), DMBS (green), MILS (red) and mFAS (orange, PDB 2vz8) are displayed in Figure 74. In each domain, the substrate-binding helices are colored in grey (Figures 73A-D).

With an increase of the chain length of the product (TENS < DMBS < MILS < mFAS), it was observed that the substrate-binding helices lose their “structural scaffold displaying more and more of an unstructured loop. The TENS and DMBS substrate-binding regions are  $\alpha$ -helices, whereas in mFAS the structural motif of this region is “reduced” to a loop. In the swap experiments (Chapter 6.4-6.5, table 14) and in the modelling of the single KR domain (Chapter 6.6) it was elucidated that some residues (*e.g.* L2363, V2397, S2400, L2401, T2404 and V2406) on this substrate-binding helix differ between the different PKS. It was hypothesized that a specific combination of these residues would be important to build either a helix structure or a loop in this region. If the KR domain has at this region a structured  $\alpha$ -helix, only shorter substrates will be recognized. If there is a more flexible loop region, then longer substrates could be recognized by the KR domain. These observations are consistent with the result of the swap experiments (Chapter 6.4-6.5), showing that the structure of the substrate-binding helix is correlated to the observed control of the chain length of the respective product.





**Fig 73:** Modell structures of KR domains with different substrate selectivity. **A**, in blue KR domain of TENS; **B**, in green KR of DMBS; **C**, in red KR domain of MILS; **D**, in orange KR domain of mFAS. The substrate-binding region is marked in grey.

### 6.7 *In silico* swap experiments with the KR domain

The comparison between TENS, DMBS, MILS and mFAS KR domains (Chapter 6.6) indicated that the structural feature of the substrate-binding region controls a crucial role in the mechanism of the programming of the chain length. In the following, a simple *in silico* swap experiment was performed. On the one hand, this *in silico* experiment should elucidate the effect on the substrate-binding helix in the domain swaps *in vivo*; on the other hand, the experiment should provide first ideas for a rational engineering approach of the KR domain of TENS.

The specific amino acids (*e.g.* S2400, L2401, T2404 and V2406), which are part of the substrate binding helix and known to be different between TENS, DMBS and MILS (table 14, sequence alignment chapter 10.2), were swapped in the sequence of TENS with the specific amino acids of MILS (Table 15). If these residues were correlated to the structural building of the helix or loop scaffold, the switch of these amino acids from TENS to

MILS should reduce the helix structure of the TENS substrate-binding helix to a loop structure.

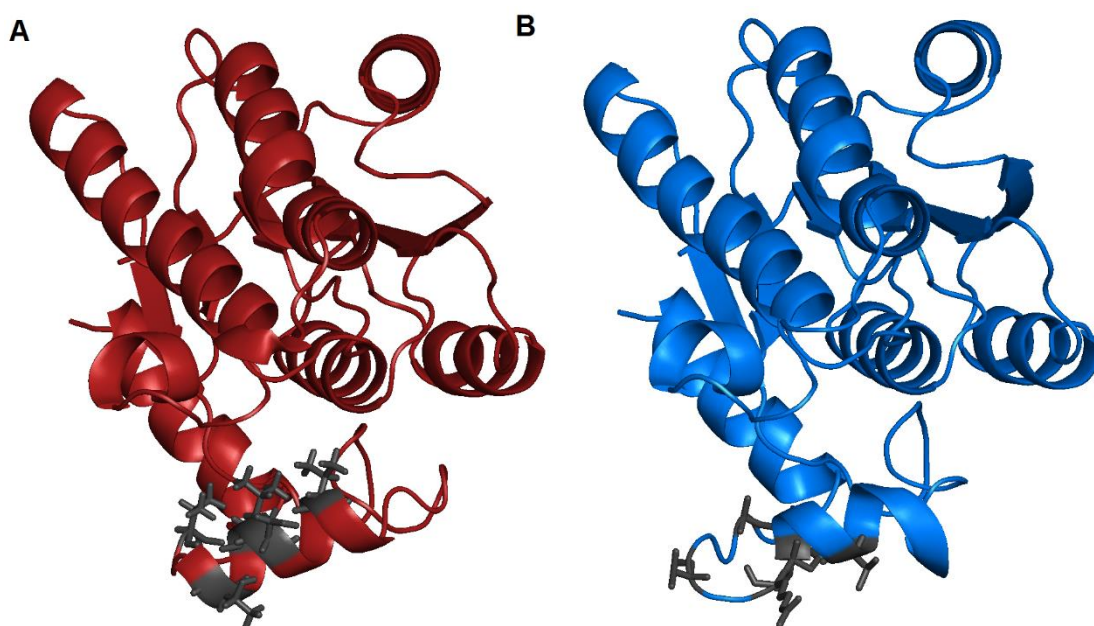
**Table 15:** Sequence alignment of a part of the TENS KR, DMBS KR, MILS KR and mutated TENS KR\*; Yellow, mutated amino acids in the TENS KR\* domain.

		5A	5A
TENS KR	(2343)	AILNNTGQSNYHCANLYMDSLVTNRRSRGLAASIIH	VGHVCDTGYVARLVDD
DMBS KR	(2338)	AILNNMGQSNYHCANLYMDSLVKHRRSRGLAASIIH	IGHVCDTGYVARMVDD
MILS KR	(2236)	TIANNIGQSNYHCANLYMDSLVAQRRSRGLAASIIH	IGYICDTGYVARLGDD
TENS KR*	(2343)	AILNNTGQSNYHCANLYMDSLVTNRRSRGLAASIIH	IGHVCDTGYVARLVDD
		5A	5A
TENS KR	(2395)	TKVQMSLGTTRVMSVSETDVHHAFAEAVRGGQPDSRSGSHNIIMGIEPPTKP	
DMBS KR	(2390)	NRIQSNIAATMRAMLSETDVHHAFAQAVRGGQLDSRSGSYNIIMGIEPPTKP	
MILS KR	(2288)	AKVHSNRDVMRATTLSETDVHHAFAEAVRGGSPGSPIGSYNIIMGIDPPTKS	
TENS KR*	(2395)	TKVQMNRGVTRAMSVSETDVHHAFAEAVRGGQPDSRSGSHNIIMGIEPPTKP	
		5A	5A
TENS KR	(2447)	LDLTKRKPVWISDPRLGCLPFSTLENQMMASEQA	
DMBS KR	(2442)	LDLTRRQAVWLSDPRLGHMLPYSTLENQMIASGQA	
MILS KR	(2440)	LDLTKRKPVWISDPRLGHMVPYSASADQAVTSEQA	
TENS KR*	(2447)	LDLTKRKPVWISDPRLGCLPFSTLENQMMASEQA	

The exchange of the amino acids (*e.g.* S2400N, L2401R, T2404V and V2406A) was performed manually in the Fasta file of TENS. Afterwards, this *mutated* TENS KR domain was again modelled with SwissModel. As template AmphB KR was chosen once more. Overall, the homology modelling resulted in structural models of the mutated TENS KR with a QMEAN value of -2.80 (sequence similarity 27%). Nearly the same validation parameters as for the WT TENS KR were observed (Chapter 5.6). Afterwards, the cofactor NADPH **11** was integrated into the model and minimized *via* YASARA (Chapter 5.7).

The result are displayed in Figure 74. It was observed that after exchange of the specific amino acids from TENS to MILS the *substrate-binding helix* (Figure 74A) “loses” its scaffold and takes up a loop structure, which is similar to the native structure in MILS (Figure 74B and Figure 72C). Hence, this mutated TENS KR domain should now be able to recognize longer substrates.

Overall, the exchange of these amino acids changed the scaffold of the substrate-binding region. Hence, the *in silico* swap experiment offers a possible structural explanation for the structural changes in the KR domain in the *in vivo* swap experiments. Furthermore, these amino acids could be further targeted by engineering experiments of the KR domain *in vitro* or *in vivo*, since the chain length seems to be programmed through this specific substrate binding helix.



**Fig 74:** **A**, KR domain of TENS with native amino acids; **B**, KR domain of TENS with swapped amino acids from MILS. In grey swapped amino acids (V2397I, S2400N, L2401R, T2404V and V2406A).

## 6.8 Discussion and Outlook

HR-PKS synthesize complex products using a single set of domains in a highly programmed, iterative fashion. Although many examples of type I HR-PKS are described in the literature, the mechanism that controls the iteration processes in these iPKS has not been clearly understood.<sup>78</sup> This stands in contrast to modular Type I PKS. In this type of PKS the programming is simply controlled through the assembly of the modules and the availability of the specific  $\beta$ -processing domains in the respective domain (for example Curacin A **52**, chapter 1.8). However, in the literature there is only a limited number of *in vitro* and *in vivo* investigations of HR-PKS catalytic domains and very little evidence for a possible mechanism of programming to be found.

The programming of the TENS, DMBS and MILS PKS systems appears to be controlled by four catalytic domains: KS, C-MeT, KR and the *trans*-acting ER (Scheme 30). The catalytic domains have a chemical selectivity, whereby they only act on substrates which are chemically competent. For example, the DH domain always acts on ACP-bound  $\beta$ -alcohols and the AT always supplies extender units.<sup>79</sup> This is supported by the limited number of reported *in vitro* investigations of HR-PKS catalytic domains.<sup>58,80,81</sup>

For example, Cox *et al.* recently showed that the functional ER domain from squalestatin tetraketide synthase (SQTKS) has low selectivity and is able to effectively reduce a wide range of enoyl-pantetheines including even unnatural isomers and stereoisomers.<sup>58</sup> Furthermore, Vederas *et al.* showed that the C-MeT domain from the

lovastatin nonaketide synthase displays tight selectivity for its substrate, while the KR in the same system is less selective.<sup>80</sup> In addition, in the KR domains the stereo-chemical course of each  $\beta$ -ketoacyl-ACP reduction is programmed and is independent of either modular context or substrate structure, including chain length and substitution pattern.<sup>81</sup>

Overall, the selectivity of the single domains can be referred to as *intrinsic* selectivity *e.g.* within their own active sites. This stands against the *extrinsic* factors such as domain-domain and protein-protein interactions, which may also be a significant determinant of PKS-programming.

Cox *et. al.* gave evidence of such an extrinsic protein-protein interaction between the catalytic domains in recent domain swap studies. The swaps appear to affect the rates at which the domains compete for ACP-bound intermediates without changing their active sites. This was also supported by previous domain-swap studies in which the cleanest change in selectivity was observed when the complete C-MeT- $\Psi$ KR-ER<sup>0</sup>-KR tetra-domain fragment was exchanged, consistent with the limited introduction of deleterious extrinsic protein-protein interactions.<sup>76</sup>

These observations are consistent with a programming mechanism which arises by *competition between* the selective domains for the ACP-bound substrate. A programming hypothesis, which is derived from these ideas, is shown in Scheme 30.

Synthesis starts with the KS domain, which catalyses the chain extension and is not selectively programmed, although it does not usually extend  $\beta$ -keto intermediates. Subsequently, a competition between the C-MeT domain, the KR domain and the chain release takes place. Depending on the intrinsic and extrinsic specification of the different domains, the kinetic parameters for the respective domains for the specific substrate at the specific time of the biosynthesis can differ (Scheme 30A). Hence, depending on how the kinetics for the domains for the specific substrate are currently, for example either methylation through the C-MeT domain or the  $\beta$ -reduction through the KR domain takes place.

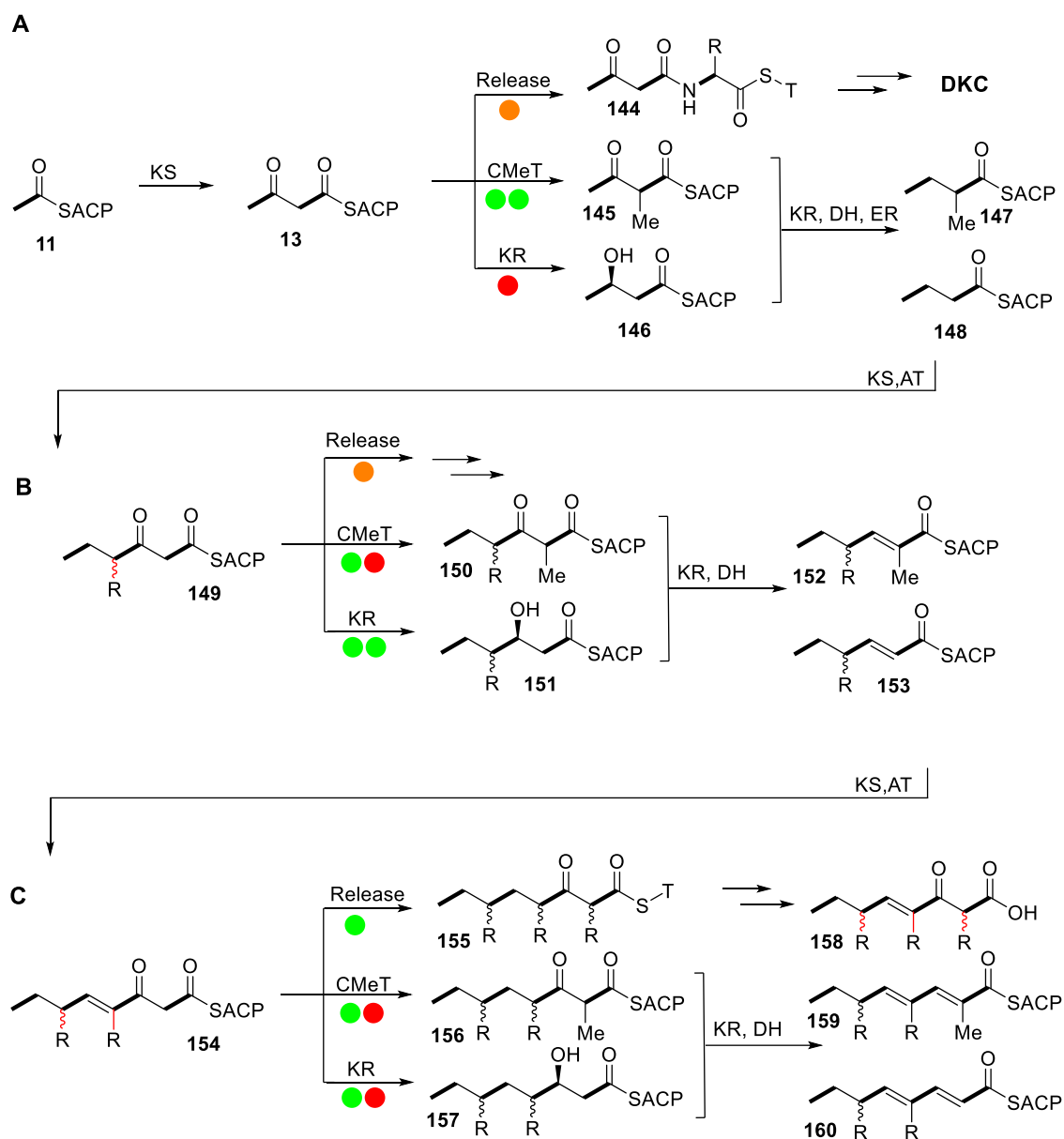
Chain elongation is always followed by the  $\beta$ -processing domains (C-MeT, KR, DH and ER), which can be active to a variable degree in HR-PKS. Subsequently, the KS domain extends the chain again. After the elongation (Scheme 30B), the C-MeT and the KR domain compete again with each other. Depending on the kinetic parameters for the respective substrate at the current time point in the biosynthesis, the intrinsic specificity and extrinsic specificity are different. Hence, either the C-MeT or the KR domain will be active. However, many iPKS products have fewer methylations with an increase of the chain length. Hence, the kinetic advantage could be more on the side of the KR domain,

than on the C-MeT domain (Scheme 30B). Depending on the PKS a termination of the elongation with subsequent modifications or cyclisation *etc* is possible as well. The release of the polyketide is marked by the orange ball in Scheme 30B. As displayed in Scheme 30C, methylation of the growing polyketide chain becomes less likely to occur with an increasing chain length. Even if the C-MeT theoretically could have the intrinsic specificity for the substrate, extrinsic factors in the programming are probably coming into play, ensuring that the kinetic advantage is on the side of the KR or the chain release.

Overall, depending on the PKS, this programming may vary in its intrinsic and extrinsic factors. As shown in Scheme 30C it is again possible that the chain release could occur, which is always more probable with increasing chain length. This may be the case because a longer elongation product or rather a product from a late step of the biosynthesis is more likely not to be a specific substrate for the individual domains any longer. Hence, the kinetic parameters will most likely favour reactions marked by the chain release.

Thus, a mechanism in which competition for substrates by a limited number of catalytic domains is the underlying principle, can explain the observed programming outcomes. The results also show that programming in these iterative PKS arises due to a *combination* of intrinsic and extrinsic factors rather than simply because of direct substrate-selectivity of each catalytic domain.

The clearest evidence for intrinsic selectivity of a catalytic domain for its substrate comes from the KR domain. Here, our modelled structure indicates that five key residues on a substrate-binding helix appear to form part of the substrate binding pocket, and the difference of these residues between TENS, DMBS and MILS appears to account for the observed change in KR substrate selectivity and hence chain-length control. The domain swap of fragment 5A for the corresponding fragment from DMBS gave hexaketides, while the same fragment from MILS specifies chain-lengths up to heptaketides. This also corresponds with additional variations in the MILS KR active site (*e.g.* L2345A). This result is confirmed by the observation that exchange of a 12-residue fragment corresponding to Q2398 - V2409 of the substrate binding helix of the TENS KR (Table 13) reprograms the system to produce hexa- and heptaketides.



**Scheme 30:** Course of polyketide biosynthesis: **A**, First condensation and subsequent selectivity; **B**, Second condensation and subsequent selectivity; **C**, Third condensation and subsequent selectivity. R = H or Me.

In turn, this swap in particular validates the modelling of the TENS PKS which predicts that these residues are most likely in contact with the substrate in the KR active site (Figure 69). This sequence is absent or of very low homology in the mFAS and AmphB KR sequences which correlates with the absence of programming in these systems - in both cases the mFAS and AmphB KR domains need only an unselective ability to reduce the  $\beta$ -keto thioesters supplied to them.<sup>79</sup>

This clear evidence for intrinsic selectivity in the KR domain, however, is lacking for the C-MeT domain where there appears to be no significant changes in the active site itself when comparing the TENS and DMBS sequences.

Evidence for the operation of extrinsic programming effects in the TENS PKS is widespread. First, programming changes are usually not limited to changing a single structural factor of the product. Second, unreduced compounds **138** and **141** must arise through inactivity of the *trans*-acting ER TenC, which was not changed in the expression experiments. This suggests that TenC does not interact with a 'free' substrate-bound ACP, but with a substrate-bound ACP intimately associated with the rest of the PKS. Swaps which result in failure of TenC to reduce correctly are associated with the C-MeT and ΨKR domains, although not with the KR, suggesting that the *trans*-acting ER associates with C-MeT and ΨKR.

Third, all active swaps still produce the native product **73**, either as the major product or as part of a mixture, again indicating that there is no simple control of overall selectivity. Finally, swaps within non-catalytic domains such as the ΨKR (experiments 15 - 17) also lead to changes in methylation pattern (*e.g.* compound **142**), although not to changes in chain-length.<sup>79</sup> All these factors suggest that extrinsic protein-protein interactions between the catalytic domains can affect the rates at which they compete for ACP-bound intermediates.

In further studies, a rational engineering of the KR domain of TENS could be performed. Therefore, the residues of the substrate-contacting-helix will be mutated to the corresponding residues of either DMBS or MILS. The first *in silico* experiments indicated that the chain length of the product might be programmed through this helix. Hence, it should be possible to perform a rational enzyme engineering as it was done for the ER domain from the HR-PKS SQTKS.

Afterwards, the generated mutants could be cloned into suitable expression vectors. The mutations will be inserted *via* homologous recombination in the yeast *Saccharomyces cerevisiae*. Finally, the mutated protein could be expressed in the fungal host strain *Aspergillus oryzae* NSAR1 and the transformants will be screened by LCMS for the production of new compounds.



## 7 Overall Conclusion

The first aim of this study was to engineer the ER domain of the HR-PKS SQTCS in a rational approach. Thereby, for the first time it was shown, that it is possible to use a *in silico* approach to generate and validate models of the SQTCS ER which could be successfully used to both understand and re-engineer the substrate specificity of the domain. Even with a lack of structural data, such as NMR or X-Ray, for the respective domain. In the subsequent *in vitro* assay, it was possible successfully to introduce the different mutation into an expression vector through site directed mutagenesis. Afterwards, expression and purification of different mutated SQTCS ER proteins was achieved.

The synthesis of the substrates for the SQTCS ER domain was already literature known, even so through a new established deprotection method using  $\text{InCl}_3$  catalysis in aqueous  $\text{CH}_3\text{CN}$  it was possible that the thioester hydrolysis problem was avoided.

In the enzyme assays, it was shown that the mutations have significant influence on the kinetic conversion, hence the specificity constant for different substrates, were altered in different ways. Mostly a faster conversion of the substrates was observed. Furthermore, it was possible to construct mutants, which accept the natural product of the SQTCS, as a substrate, and thus enhance the substrate variability.

In the future, the generated mutants should be inserted into the whole SQTCS gene on a suitable expression vector to investigate the influence of the mutants *in vivo*. Furthermore, to investigate if it would be possible to generate new metabolites. In addition, other domains, such as the KR domain or the C-MeT should be investigated through the rational engineering. The combination of the different engineered domains could give access to a completely new range of metabolites.

Overall, it was shown for the first time that modelling of HR-PKS domains can be a successful method for the generation of hypotheses for the rational re-engineering of HR-PKS.

In the second project, the aim is to provide in parallel to the molecular-biological work, the structural-biological foundations and analysis for the domain swaps between TENS, DMBS and MILS. On the one hand to understand the molecular basis of methylation and chain length programming *in silico*. On the other hand, to perform first *in silico* studies for the subsequent rational-engineering of the TENS KR domain.



Thereby, it was shown, that it was possible to model and validate models of the TENS KR and C-MeT domain *in silico*. Furthermore, a quaternary hybrid structure of the TENS and mFAS domains, including the DH, ER, ΨKR (mFAS) and C-MeT and KR (TENS) with the mFAS as a scaffold was generated. This generated chimeric model was successfully used to design and understand the domain swaps. Further, under consideration of the chimeric model structure in combination with the single domains and the multiple sequence alignment it was possible to understand the molecular basis of the methylation and chain-length programming. Hence, that programming in these iterative PKS arises due to a *combination* of intrinsic and extrinsic factors rather than simply because of direct substrate-selectivity of each catalytic domain.

In the case of the KR domain, it was possible to show for the first time the influence of the so-called *substrate-binding helix*. This helix forms part of the substrate binding pocket, and the difference of these residues between TENS, DMBS and MILS appears to account for the observed change in KR substrate selectivity and hence chain-length control. Furthermore, first *in silico studies* concerning the substrate-binding helix of the KR domain successfully showed that it should be possible to rationally-engineer the KR domain. In further studies, mutants of the TENS KR domain shall be constructed and cloned into suitable expression vector for the subsequent *in vivo* expression. Overall, these studies lead to a better understanding of the programming mechanism of PKS modules and PKS in general.

## 8 Experimental Section

### 8.1 Equipment

**NMR analysis:**  $^1\text{H}$ -NMR analysis was performed using BRUKER DPX 200, Avance 400, DPX 400 and DRX 500 instruments. Signals are determined in some cases with two dimensional NMR  $^1\text{H}$ ,  $^1\text{H}$ -COSY,  $^{13}\text{C}$ -HSQC and  $^1\text{H}$ ,  $^{13}\text{C}$ - $J_3$ -HMBC.  $^{13}\text{C}$ -NMR analysis was performed using BRUKER Avance 400, DPX 400 and DRX 500 instruments. Deuterated chloroform (ref. 7.26 ppm / 77.2 ppm)<sup>187</sup> and deuterated acetonitrile (ref. 1.94 ppm / 118.4 ppm) were used as solvents and served as internal references. All  $\delta$  values are reported in ppm. All  $J$  values are reported in Hz.

**Column Chromatography:** For column chromatography silicageol 60 (particle size 35-70 micron, Sigma-Aldrich or 40-63 micron, Macherey-Nagel) was used. Columns were packed wet under  $\text{N}_2$  pressure. Products were eluted with the indicated solvent mixtures. Purified fractions were analysed by TLC and combined if same  $R_f$  was observed. Final products were evaporated in *vacuo*.

**TLC:** TLC analysis was performed on TLC plates with a polyester backed 0.2 mm silica gel phase from Macherey and Nagel using the indicated solvent systems. Analysis of the plates were performed by ultraviolet light (254 nm) or with potassium permanganate (5 mmol) or *o*-anisaldehyde (anisaldehyde [15 g], EtOH [250 ml] and concentrated  $\text{H}_2\text{SO}_4$  [2.5 ml]) solution.

**Analytical LCMS:** Analytical LCMS data were obtained with using a Waters LCMS system comprising of a Waters 2767 autosampler, Waters 2545 pump system, a Phenomenex Kinetex column (2.6  $\mu$ ,  $\text{C}_{18}$ , 100  $\text{\AA}$ , 4.6  $\times$  100 mm) equipped with a Phenomenex Security Guard precolumn (Luna  $\text{C}_5$  300  $\text{\AA}$ ) eluted at 1 mL/min. Detection was by Waters 2998 Diode Array detector between 200 and 600 nm; Waters 2424 ELSD and Waters SQD-2 mass detector operating simultaneously in  $\text{ES}^+$  and  $\text{ES}^-$  modes between 100  $m/z$  and 650  $m/z$ . Solvents were: **A**, HPLC grade  $\text{H}_2\text{O}$  containing 0.05% formic acid; **B**, HPLC grade MeOH containing 0.045% formic acid; and **C**, HPLC grade  $\text{CH}_3\text{CN}$  containing 0.045% formic acid. Gradients were as follows. *Method 1*. Kinetex/ $\text{CH}_3\text{CN}$ : 0 min, 10% **C**; 10 min, 90% **C**; 12 min, 90% **C**; 13 min, 10% **C**; 15 min, 10% **C**.

**Preparative LCMS:** Purification of final compounds was generally achieved using a Waters mass-directed auto purification system comprising of a Waters 2767 autosampler, Waters 2545 pump system, a Phenomenex Kinetex Axia column (5 $\mu$ , C<sub>18</sub>, 100 Å, 21.2 × 250 mm) equipped with a Phenomenex Security Guard precolumn (Luna C<sub>5</sub> 300 Å) eluted at 20 mL/min at ambient temperature. Solvent **A**, HPLC grade H<sub>2</sub>O + 0.05% formic acid; Solvent **B**, HPLC grade CH<sub>3</sub>CN + 0.045% formic acid. The post-column flow was split (100:1) and the minority flow was made up with HPLC grade CH<sub>3</sub>CN + 0.045% formic acid to 1 mL·min<sup>-1</sup> for simultaneous analysis by diode array (Waters 2998), evaporative light scattering (Waters 2424) and ESI mass spectrometry in positive and negative modes (Waters SQD-2). Detected peaks were collected into glass test tubes. Combined tubes were evaporated (vacuum centrifuge), weighed, and residues dissolved directly in solvent for use or analysis.

**Protein purification:** The protein purification was performed with an FPLC ÄKTA pure system from the company GE Healthcare. For FPLC analysis a combination with the software UNICORN 7.0 and different columns (Nickel column Protino Ni-NTA Columns 5 mL, Size exclusion column- HiLoad 26/600 Superdex 200pg (GE Healthcare), 320 mL) was used.

**UV-Analysis:** UV assays were measured with a JASCO-V630-spectrophotometer in quartz glass cuvettes with a diameter of 10 mm. The temperature was controlled by the JASCO-V630-Spectrophotometer at 25 °C. The processed data (by JASCO/Spectramanager) was after that recalculated with Microsoft EXCEL.

## 8.2 Buffer, antibiotics, media and solutions

All enzymes used in this work were purchased from Thermo Fisher Scientific (Waltham, MA, USA), Takara Bio Inc. (Shiga, J), Invitrogen Life Technologies (Darmstadt, D), or Bioline (London, UK). All enzymes were used according to manufacturer's instructions with appropriate supplied buffers. Buffers and media used in this work were sterilized by autoclaving 15 min at 121 °C (Autoclave 2100 Classic, Prestige Medical) or by disposable sterile filter (0.45 µm pore size, Roth) and are summarized in tables 16 and 17.

**Table 16:** Media used in this work

Media	Composition [% (w/v)]	Ingredients
<b>2TY</b>	1	Yeast extract
	0.5	Sodium chloride
	1.6	Tryptone
<b>LB</b>	0.5	Yeast extract
	1	Tryptone
	0.5	Sodium chloride
<b>SOC</b>	0.5	Yeast extract
	2	Tryptone
	0.06	Sodium chloride
	0.02	Potassium chloride
	25 mM	Magnesium chloride x 6 H <sub>2</sub> O
	1	D(+)-Glucose

**Table 17:** Agar used in this work

Agar	Composition [% (w/v)]	Ingredients
<b>LB agar</b>	0.5	Yeast extract
	1	Tryptone
	0.5	Sodium chloride
	1.5	Agar

Antibiotics stock solutions were prepared in distilled water or ethanol. They were filter sterilized through 0.45  $\mu\text{m}$  syringe filter and stored at  $-20\text{ }^{\circ}\text{C}$ . Stock and working concentrations are listed in table 18

**Table 18:** Agar used in this work

Antibiotic	Solvent	Stock concentration [mg / ml]	Working concentration [ $\mu\text{g}$ / ml]
<b>Carbenicillin</b>	H <sub>2</sub> O	50	50
<b>Kanamycin</b>	H <sub>2</sub> O	50	50

Information about *E. coli* strains used in this work are summarized in table 19.

**Table 19:** *E. coli* strains used in this work

Bacteria	Reference
<b>BL21 (DE3)</b>	Thermo Fisher Scientific
<b>OneShot <i>ccdB</i> survival 2T1R</b>	Thermo Fisher Scientific
<b>OneShot Top10</b>	Thermo Fisher Scientific

### 8.3 Synthesis of Pantetheine Substrates

The used solvents were used without any purification or drying process. Tetrahydrofuran, diethyl ether and dichloromethane in the dry version and NADPH were bought from Carl Roth. Other chemicals were ordered from Sigma Aldrich or TCI Deutschland GmbH. The syntheses were done established protocols from Christoph Bartel,<sup>58,137</sup> where this was not the case it was recorded.

#### 8.3.1 Preparation of precursor substrates

##### General Methods<sup>58,137</sup>

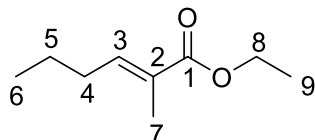
**I.** A solution of dichloromethane (5 mL) and the respective aldehyde (1.00 mmol) was stirred at 0 °C. (Carbethoxyethylidene)triphenyl phosphorane (0.72 g, 2.00 mmol) or Carbethoxymethylidene)triphenyl phosphorane (0.69 g, 2.00 mmol) was added to this solution warmed to 25 °C and stirred for 16-18 hours. Then the solvent was evaporated under a nitrogen flow. The crude product was purified by column chromatography.

**II.** To a solution of the corresponding ester in ethanol/water 5:1 (5 mL/ 1 mL) potassium hydroxide (22 mmol) was added to the solution. After stirring under reflux for 3 hours diethyl ether was added. The mixture was washed with NaHCO<sub>3</sub> (3 × 10 mL). Then the aqueous layer was acidified with 2 M HCl until pH 1 and extracted with ethyl acetate (3 x 10 ml). The organic layer was dried over MgSO<sub>4</sub> and concentrated *in vacuo*.

**III.** Malonic acid (2.5 g, 24.0 mmol) and aldehyde (38.0 mmol) were dissolved in pyridine (8 ml) and morpholine (35 µl). The solution was stirred for 17 hours at 25 °C and was heated to 115 °C and stirred for further 6 hours. The mixture was quenched with 1M NaOH (20 ml) and extracted with diethyl ether (3 x 15 ml). Then the water layer was acidified with 2M HCl and extracted with diethyl ether (6 x 15 ml). The organic layer was dried over MgSO<sub>4</sub> and concentrated *in vacuo*.

**IV.** To a solution of squalestatin S1 (1.00 g, 1.40 mmol) in ethanol/water 5:1 (5 mL/ 1 mL) potassium hydroxide (22.00 mmol) was added to the solution. After stirring under reflux for 3 hours, diethyl ether was added through the solution and washed 3 × with NaHCO<sub>3</sub> (3 × 10 mL). The aqueous layer with the (4*S*, 6*S*)-2-4-dimethyloct-2-enoic acid was acidified with 2 M HCl until pH 1 and extracted with ethyl acetate (3 × 10 mL). The organic layer was dried over MgSO<sub>4</sub> and concentrated *in vacuo*.

### ***E*-Ethyl 2-methylhex-2-enoate 161<sup>188</sup>**



The obtained product was a colorless oil (0,18 g, 1.15 mmol, 57.6 %)

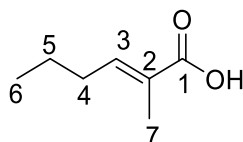
Rf 0.21(EtOAc / PE 1: 10)

<sup>1</sup>H-NMR (CDCl<sub>3</sub>, 400 MHz): δ 0.93 (t, 3H, 6-CH<sub>3</sub>); 1.29 (t, 3H, 9-CH<sub>3</sub>); 1.39-1.55 (m, 2H, 5-CH<sub>2</sub>); 1.82 (s, 3H, 7-CH<sub>3</sub>); 2.18 (q, 2H, 4-CH<sub>2</sub>); 4.18 (q, 2H, 8-CH<sub>2</sub>); 6.74 (t, 1H, 3-CH).

<sup>13</sup>C-NMR (CDCl<sub>3</sub>, 100 MHz): δ12.4 (6-CH<sub>3</sub>), 13.9 (9-CH<sub>3</sub>), 14.3 (7-CH<sub>3</sub>), 21.9 (5-CH<sub>2</sub>); 30.7 (4-CH<sub>2</sub>); 60.4 (8-CH<sub>2</sub>); 127.9 (2-C); 142.2 (3-CH); 168.4 (1-CO)

ESMS: *m/z*: 179 [M + Na]H<sup>+</sup>, 157 [M]H<sup>+</sup>, 129 [M - CH<sub>2</sub>CH<sub>3</sub>]H<sup>+</sup>.

### ***E*-Methylhex-2-enoic acid 162<sup>58,137</sup>**



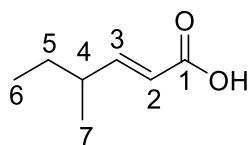
The obtained product was a red oil (0.11 g, 0.85 mmol, 74.7 %).

<sup>1</sup>H-NMR (CDCl<sub>3</sub>, 400 MHz): δ 0.91 (t, 3H, 6-CH<sub>3</sub>); 1.32-1.50 (m, 2H, 5-CH<sub>2</sub>); 1.85 (s, 3H, 7-CH<sub>3</sub>); 2.22 (q, 2H, 4-CH<sub>2</sub>); 6.93 (t, 1H, 3-CH).

<sup>13</sup>C-NMR (CDCl<sub>3</sub>, 100 MHz): δ12.0 (6-CH<sub>3</sub>), 13.8 (7-CH<sub>3</sub>), 21.7 (5-CH<sub>2</sub>), 30.6 (4-CH<sub>2</sub>), 127.0 (2-C), 145.2 (3-CH), 173.1 (1-CO).

ESMS: *m/z*: 152 [M + Na]H<sup>+</sup>, 129 [M]H<sup>+</sup>; 111 [M - H<sub>2</sub>O] H<sup>+</sup>.

### ***E*-4-Methylhexenoic acid 163<sup>58,137</sup>**



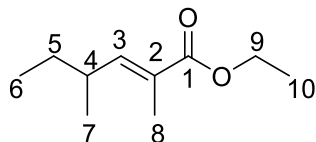
The obtained product was a yellow oil (1.43 g, 11.17 mmol, 47 %).

<sup>1</sup>H-NMR (CDCl<sub>3</sub>, 400 MHz): δ 0.88 (t, 3H, 6-CH<sub>3</sub>); 1.04 (d, 3H, 7-CH<sub>3</sub>); 1.36-1.44 (m, 2H, 5-CH<sub>2</sub>); 2.17-2.27 (m, 1H, 4-CH); 5.78 (d, 1H, 2-CH); 6.87 (t, 1H, 3-CH).

<sup>13</sup>C-NMR (CDCl<sub>3</sub>, 100 MHz): δ11.7 (6-CH<sub>3</sub>), 18.9 (7-CH<sub>3</sub>), 28.7 (5-CH<sub>2</sub>), 38.1 (4-CH), 119.3 (2-CH), 154.8 (3-CH), 167.3 (1-CO)

ESMS: *m/z*:152 [M + Na]H<sup>+</sup>, 129 [M]H<sup>+</sup>

### ***E*-Ethyl-2,4-dimethylhex-2-enoate 164<sup>188</sup>**



The obtained product was a colorless oil (0.29 g, 1.60 mmol, 85.3%).

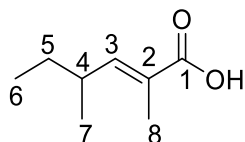
R<sub>f</sub>: 0.78 (EtOAc / PE 1: 10).

<sup>1</sup>H-NMR (CDCl<sub>3</sub>, 400 MHz): δ 0.85 (t, 3H, 6-CH<sub>3</sub>); 1.00 (d, 3H, 7-CH<sub>3</sub>); 1.28-1.55 (m, 2H, 5-CH<sub>2</sub>); 1.30 (t, 3H, 10-CH<sub>3</sub>); 1.82 (s, 3H, 8-CH<sub>3</sub>); 2.41 (m, 1H, 4-CH); 4.19 (q, 2H, 9-CH<sub>2</sub>); 6.52 (d, 1H, 3-CH).

<sup>13</sup>C-NMR (CDCl<sub>3</sub>, 100 MHz): δ 11.9 (6-CH<sub>3</sub>), 12.6 (7-CH<sub>3</sub>), 14.3 (10-CH<sub>3</sub>), 19.7 (8-CH<sub>3</sub>), 29.7 (5-CH<sub>2</sub>), 34.9 (4-CH), 60.4 (9-CH<sub>2</sub>), 126.6 (2-C), 147.9 (3-CH), 168.4 (1-CO)

ESMS: *m/z*: 193 [M + Na]H<sup>+</sup>, 171 [M]H<sup>+</sup>

### ***E*-2,4-Dimethylhex-2-enoic acid 165<sup>58,137</sup>**



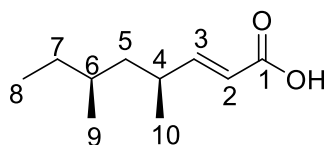
The obtained product was a red oil (0.17 g, 1.19 mmol, 70.8 %).

<sup>1</sup>H-NMR (CDCl<sub>3</sub>, 400 MHz): δ 0.86 (t, 3H, 1-CH<sub>3</sub>); 1.01 (d, 3H, 4-CH<sub>3</sub>); 1.24-1.49 (m, 2H, 2-CH<sub>2</sub>); 1.85 (s, 3H, 7-CH<sub>3</sub>); 2.38-2.49 (m, 1H, 3-CH); 6.68 (qq, 1H, 5-CH).

<sup>13</sup>C-NMR (CDCl<sub>3</sub>, 100 MHz): δ 11.8 (1-CH<sub>3</sub>), 12.2 (7-CH<sub>3</sub>), 19.5 (4-CH<sub>3</sub>), 29.5 (2-CH<sub>2</sub>), 35.1 (3-CH), 125.7 (6-C), 150.8 (5-CH), 173.3 (8-COOH).

ESMS: *m/z*: 184 [M + CH<sub>3</sub>CN]H<sup>+</sup>, 143 [M]H<sup>+</sup>

### ***E*-(4*S*,6*S*)-2-4-Dimethyloctenoic acid 166<sup>58,137</sup>**



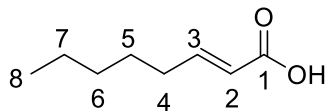
The obtained product was yellow solid (0.22 g, 1.29 mmol, 84 %)

<sup>1</sup>H-NMR (CDCl<sub>3</sub>, 400 MHz): δ 0.88 (t, *J* = 7.4, 6H, 8-9-CH<sub>3</sub>); 1.07 (d, *J* = 6.2, 3H, 10-CH<sub>3</sub>); 1.27-1.46 (m, 5H, 7-5-CH<sub>2</sub>-6-CH); 2.41-2.52 (m, 1H, 4-CH), 5.81 (d, *J* = 1.7, *J* = 16.0, 1H, 2-CH), 6.96 (dd, *J* = 8.8, *J* = 16.0, 1H, 3-CH).

<sup>13</sup>C-NMR (CDCl<sub>3</sub>, 100 MHz): δ 11.1 (8-CH<sub>3</sub>); 20.2 (9-CH<sub>3</sub>); 20.7 (10-CH<sub>3</sub>); 29.8 (7-CH<sub>2</sub>); 31.9 (6-CH); 34.4 (4-CH); 43.1 (5-CH<sub>2</sub>), 118.3 (2-CH); 164.6 (3-CH); 176.7 (1-CO).

ESMS: *m/z*: 193 [M + Na]H<sup>+</sup>, 171 [M]H<sup>+</sup>



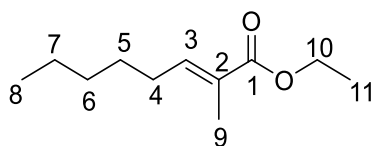
***E*-oct-2-enoic acid 167**<sup>58,137</sup>

The obtained product was a yellow oil. (1.69 g, 11.9, 49.5 %)

<sup>1</sup>H-NMR (CDCl<sub>3</sub>, 400 MHz): δ 0.92 (t, 3H, 8-CH<sub>3</sub>); 1.26-1.52 (m, 6H, 5-6-7-CH<sub>2</sub>); 2.22 (q, 2H, 4-CH<sub>2</sub>); 5.78 (dd, 1H, 2-CH); 6.94 (tq, 1H, 3-CH).

<sup>13</sup>C-NMR (CDCl<sub>3</sub>, 100 MHz): δ 12.0 (8-CH<sub>3</sub>), 22.5 (7-CH<sub>2</sub>), 28.1 (5-CH<sub>2</sub>), 28.8 (4-CH<sub>2</sub>), 31.5 (6-CH<sub>2</sub>), 126.8 (2-C), 145.5 (3-CH), 173.1 (1-CO).

ESMS: *m/z*: 143 [M-H]<sup>-</sup>

***E*-Ethyl-2-methyloct-2-enoate 168**<sup>188</sup>

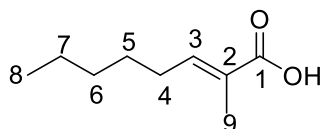
The obtained product was a colorless oil (0.23 g, 1.25 mmol, 59.5%).

R<sub>f</sub>: 0.65 (EtOAc / PE 1: 10).

<sup>1</sup>H-NMR (CDCl<sub>3</sub>, 400 MHz): δ 0.88 (t, 3H, 8-CH<sub>3</sub>); 1.25-1.55 (m, 9H, 5-7-CH<sub>2</sub>, 11-CH<sub>3</sub>); 1.82 (s, 3H, 9-CH<sub>3</sub>); 2.15 (q, 2H, 4-CH<sub>2</sub>); 4.18 (q, 2H, 10-CH<sub>2</sub>); 6.77 (t, 1H, 3-CH).

<sup>13</sup>C-NMR (CDCl<sub>3</sub>, 100 MHz): δ 12.3 (8-CH<sub>3</sub>), 14.0 (9-CH<sub>3</sub>), 14.3 (11-CH<sub>3</sub>), 22.5 (7-CH<sub>2</sub>), 28.3 (5-CH<sub>2</sub>), 28.7 (4-CH<sub>2</sub>), 31.6 (6-CH<sub>2</sub>), 60.4 (10-CH<sub>2</sub>), 127.7 (2-C), 142.5 (3-CH), 168.4 (1-C)

ESMS: *m/z*: 207 [M + Na]H<sup>+</sup>, 185 [M]H<sup>+</sup>.

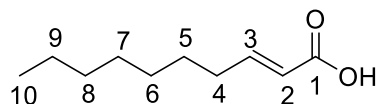
***E*-2-Methyloct-2-enoic acid 169**<sup>58,137</sup>

The obtained product was a red oil (0.11 g, 0.7 mmol, 56.4 %).

<sup>1</sup>H-NMR (CDCl<sub>3</sub>, 400 MHz): δ 0.92 (t, 3H, 8-CH<sub>3</sub>); 1.26-1.52 (m, 6H, 5-6-7-CH<sub>2</sub>); 1.86 (s, 3H, 9-CH<sub>3</sub>); 2.22 (q, 2H, 4-CH<sub>2</sub>); 6.94 (t, 1H, 3-CH).

<sup>13</sup>C-NMR (CDCl<sub>3</sub>, 100 MHz): δ 12.0 (8-CH<sub>3</sub>), 14.0 (9-CH<sub>3</sub>), 22.5 (7-CH<sub>2</sub>), 28.1 (5-CH<sub>2</sub>), 28.8 (4-CH<sub>2</sub>), 31.5 (6-CH<sub>2</sub>), 126.8 (2-C), 145.5 (3-CH), 173.1 (1-CO).

ESMS: *m/z*: 179 [M + Na]H<sup>+</sup>, 157 [M]H<sup>+</sup>

***E*-Dec-2-enoic acid 170**<sup>58,137</sup>

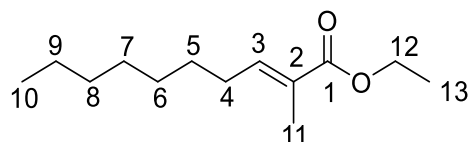
The obtained product was a yellow oil (0.17 g, 1 mmol, 62.9 %)

**<sup>1</sup>H-NMR** (CDCl<sub>3</sub>, 400 MHz): δ 0.90 (t, 3H, 10-CH<sub>3</sub>); 1.29-1.50 (m, 10H, 5-9-CH<sub>2</sub>); 2.25 (dq, 2H, 4-CH<sub>2</sub>); 5.84 (d, 1H, 2-CH); 7.10 (dt, 1H, 3-CH).

**<sup>13</sup>C-NMR** (CDCl<sub>3</sub>, 100 MHz): δ 14.1 (10-CH<sub>3</sub>), 22.6 (9-CH<sub>2</sub>); 27.8 (8-CH<sub>2</sub>); 29.0 (7-CH<sub>2</sub>); 29.1 (6-CH<sub>2</sub>); 31.7 (5-CH<sub>2</sub>); 32.3 (4-CH<sub>2</sub>); 120.5 (2-CH); 152.5 (3-CH); 171.9 (1-CO).

**ESMS:** *m/z*: 212.4 [M]H<sup>+</sup>, 193.2 [M - Na]H<sup>+</sup>, 153.5 [M - H<sub>2</sub>O]H<sup>+</sup>

### ***E*-Ethyl-2-methyldec-2-enoate 171<sup>188</sup>**



The obtained product was a colorless oil (0.27 g, 1.27 mmol, 63 %).

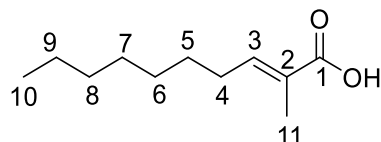
R<sub>f</sub>: 0.85 (EtOAc / PE 1:7).

**<sup>1</sup>H-NMR** (CDCl<sub>3</sub>, 400 MHz): δ 0.88 (t, 3H, 10-CH<sub>3</sub>); 1.25-1.58 (m, 13H, 5-9-CH<sub>2</sub>, 13-CH<sub>3</sub>); 1.82 (s, 3H, 11-CH<sub>3</sub>); 2.15 (q, 2H, 4-CH<sub>2</sub>); 4.18 (q, 2H, 12-CH<sub>2</sub>); 6.75 (t, 1H, 3-CH).

**<sup>13</sup>C-NMR** (CDCl<sub>3</sub>, 100 MHz): δ 12.3 (10-CH<sub>3</sub>), 14.1 (11-CH<sub>3</sub>), 14.3 (13-CH<sub>3</sub>), 22.6 (9-CH<sub>2</sub>), 28.6 (7-CH<sub>2</sub>), 28.7 (6-CH<sub>2</sub>), 29.1 (5-CH<sub>2</sub>), 29.4 (4-CH<sub>2</sub>), 31.8 (8-CH<sub>2</sub>), 60.4 (12-CH<sub>2</sub>), 127.6 (2-C), 142.5 (3-CH), 168.4 (1-CO)

**ESMS:** *m/z*: 254 [M + CH<sub>3</sub>CN]H<sup>+</sup>, 213 [M]H<sup>+</sup>.

### ***E*-2-Methyldec-2-enoic acid 172<sup>58,137</sup>**



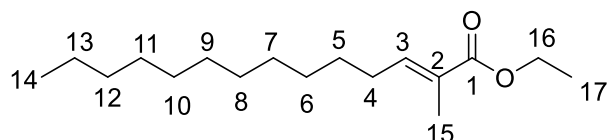
The obtained product was a red oil (0.17 g, 0.92 mmol, 62.9 %)

**<sup>1</sup>H-NMR** (CDCl<sub>3</sub>, 400 MHz): δ 0.87-0.90 (m, 3H, 10-CH<sub>3</sub>); 1.26-1.30 (m, 10H, 5-9-CH<sub>2</sub>); 1.83 (s, 3H, 11-CH<sub>3</sub>); 2.19 (q, 2H, 4-CH<sub>2</sub>); 6.90 (t, 1H, 3-CH).

**<sup>13</sup>C-NMR** (CDCl<sub>3</sub>, 100 MHz): δ 12.3 (10-CH<sub>3</sub>), 14.1 (11-CH<sub>3</sub>), 22.6 (9-CH<sub>2</sub>), 28.4 (7-CH<sub>2</sub>), 28.9 (6-CH<sub>2</sub>), 29.1 (5-CH<sub>2</sub>), 29.3 (4-CH<sub>2</sub>), 31.8 (8-CH<sub>2</sub>), 126.8 (2-C), 145.6 (3-CH), 172.8 (1-CO).

**ESMS:** *m/z*: 226 [M + CH<sub>3</sub>CN]H<sup>+</sup>, 167 [M - H<sub>2</sub>O]H<sup>+</sup>.

### ***E*-Ethyl-2-methyltetradec-2-enoate 173<sup>74</sup>**



The obtained product was a colorless oil (0.34 g, 1.27 mmol, 63.4 %)

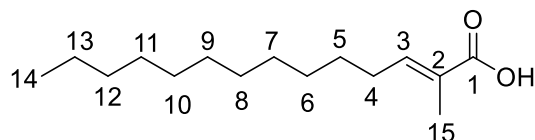
R<sub>f</sub>: 0.9 (EtOAc / PE 1:7).

<sup>1</sup>H-NMR (CDCl<sub>3</sub>, 400 MHz): δ 0.88 (t, 3H, 14-CH<sub>3</sub>), 1.24-1.36 (m, 18H, 5-13-CH<sub>2</sub>), 1.43 (t, 3H, 17-CH<sub>3</sub>), 2.16 (q, 2H, 4-CH<sub>2</sub>), 6.76 (d, 1H, 3-CH), 1.82 (s, 3H, 15-CH<sub>3</sub>), 4.19 (q, 2H, 16-CH<sub>2</sub>).

<sup>13</sup>C-NMR (CDCl<sub>3</sub>, 100 MHz): δ 13.2 (15-CH<sub>3</sub>), 14.0 (14-CH<sub>3</sub>), 14.0 (17-CH<sub>3</sub>), 22.5 (13-CH<sub>2</sub>), 28.1-31.8 (4-12-CH<sub>2</sub>), 61.1 (16-CH<sub>2</sub>), 142.8 (2-C), 126.5 (3-CH), 168.1 (1-CO).

ESMS: *m/z*: 291 [M+Na]H<sup>+</sup>

### ***E*-2-Methyltetradec-2-enoic acid 174<sup>74</sup>**



The obtained product was a red oil (0.29 g, 1.2 mmol, 85.3 %)

<sup>1</sup>H-NMR (CDCl<sub>3</sub>, 400 MHz): δ 0.88 (t, 3H, 14-CH<sub>3</sub>), 1.24-1.36 (m, 18H, 5-13-CH<sub>2</sub>), 2.16 (q, 2H, 4-CH<sub>2</sub>), 6.76 (d, 1H, 3-CH), 1.82 (s, 3H, 15-CH<sub>3</sub>).

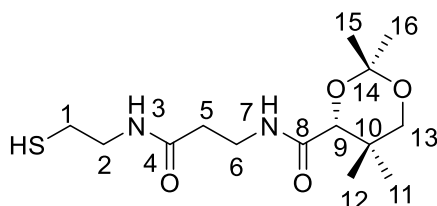
<sup>13</sup>C-NMR (CDCl<sub>3</sub>, 100 MHz): δ 13.2 (15-CH<sub>3</sub>), 14.0 (14-CH<sub>3</sub>), 22.5 (13-CH<sub>2</sub>), 28.1-31.8 (4-12-CH<sub>2</sub>), 142.8 (2-C), 126.5 (3-CH), 168.1 (1-CO).

ESMS: *m/z*: 263 [M + Na]H<sup>+</sup>, 223 [M - H<sub>2</sub>O]H<sup>+</sup>

## **8.3.2 Preparation of panthetine dimethyl ketal substrates**

### **Preparation of pantetheine dimethyl ketal 121<sup>189,58,137</sup>**

D-pantothenic acid hemicalcium salt (2.50 g, 10.50 mmol), p-toluensulfonic acid (2.30 g, 13.00 mmol) and 5 g molecular sieves were suspended in 125 mL dry acetone and stirred at 25 °C for 12 hours under a nitrogen atmosphere. The suspension was filtered with celite and washed with 200 ml acetone. The filtrate was concentrated to a colorless oil, redissolved in 200 ml ethyl acetate and washed two times with brine (25 ml) and dried over MgSO<sub>4</sub>. After that the ethyl acetate was removed under vacuum and hexane was added to the flask to get a white solid that was dried under high vacuum. The corresponding D-pantothenic dimethyl ketal (1.90 g, 7.00 mmol) was dissolved in 40 mL dry THF with CDI (1.70 g, 11.00 mmol) and stirred for one hour at 25 °C. Then cysteamine (1.30 g, 11.00 mmol) was added to the solution and stirred for 12 hours. The solution was concentrated under vacuum and dichloromethane was added. The organic layer were washed with NH<sub>4</sub>Cl (25 mL) and brine (25 mL), dried over MgSO<sub>4</sub> and concentrated *in vacuo*. After that, the colorless oil was purified by column chromatography.



The obtained product was a white solid (1.82 g, 5.74 mmol, 85 % over to steps)

R<sub>f</sub>: 0.2 (EtOAc).

**<sup>1</sup>H-NMR** (CDCl<sub>3</sub>, 400 MHz): δ [ppm]: 0.98 (s, 3H, 11-CH<sub>3</sub>); 1.05 (s, 3H, 12-CH<sub>3</sub>); 1.39 (t, 1H, SH); 1.43 (s, 3H, 15-CH<sub>3</sub>); 1.47 (s, 3H, 16-CH<sub>3</sub>); 2.40 (t, 2H, 5-CH<sub>2</sub>); 2.64-2.70 (m, 2H, 1-CH<sub>2</sub>); 3.29 (d, 1H, 13a-CH<sub>2</sub>); 3.37-3.63 (m, 4H, 2-6-CH<sub>2</sub>); 3.69 (d, 1H, 13b-CH<sub>2</sub>); 4.09 (s, 1H, 9-CH); 6.37 (bt, 1H, 3-NH); 7.03 (bt, 1H, 7-NH).

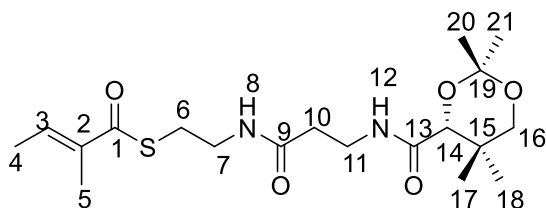
**<sup>13</sup>C-NMR** (CDCl<sub>3</sub>, 100 MHz): δ [ppm]: 18.7 (12-CH<sub>3</sub>); 18.9 (11-CH<sub>2</sub>); 22.1 (15-CH<sub>3</sub>); 24.6 (1-CH<sub>2</sub>); 29.5 (15-CH<sub>3</sub>); 33.0 (10-C); 34.9 (6-CH<sub>2</sub>); 36.2 (5-CH<sub>2</sub>); 42.4 (2-CH<sub>2</sub>); 71.4 (13-CH<sub>2</sub>); 77.2 (9-CH); 99.1 (14-C); 170.3 (4-CO); 171.1 (8-CO).

**ESMS:** *m/z*: 319 [M] H<sup>+</sup>, 261 [M - (CH<sub>3</sub>)<sub>2</sub>CO]H<sup>+</sup>

### General procedure pantetheine dimethyl ketal-compounds<sup>137</sup>

Acid (1.00 mmol) and pantetheine dimethyl ketal (0.32 g, 1.00 mmol,) were dissolved in dichloromethane (8 ml). The mixture was cooled to 0 °C. Then *N,N*-dimethyl aminopyridine (0.10 g, 0.80 mmol,) and *N*-(3-Diethylamino-propyl)-*N*-ethyl carbodiimide (0.38 g, 2.00 mmol) were added. The mixture was warmed to 25 °C and stirred for 4 hours. After that the mixture was quenched with 2M HCl (10 ml) and extracted with dichloromethane (3x 25 ml). The organic layer was washed with saturated NaHCO<sub>3</sub> (20 ml) and brine (20 ml). The product was dried over MgSO<sub>4</sub> and concentrated under vacuo. The crude product was purified by column chromatography

### Tigloyl pantetheine dimethyl ketal **175**<sup>189,58,137</sup>



The obtained product was a colorless oil (0.318 g, 0.795 mmol, 79.5 %)

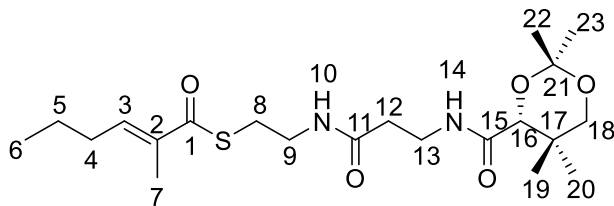
R<sub>f</sub>: 0.46 (EtOAc).

**<sup>1</sup>H-NMR** (CDCl<sub>3</sub>, 400 MHz): δ 0.99 (s, 3H, 17-CH<sub>3</sub>); 1.04 (s, 3H, 18-CH<sub>3</sub>); 1.44 (s, 3H, 21-CH<sub>3</sub>); 1.48 (s, 3H, 22-CH<sub>3</sub>); 1.85-1.87 (m, 6H, 4-5-CH<sub>3</sub>); 2.44 (t, *J* = 6.5, 2H, 10-CH<sub>2</sub>); 3.08 (t, *J* = 6.3, 2H, 6-CH<sub>2</sub>); 3.28 (d, *J* = 11.7, 1H, 16a-CHH); 3.41-3.62 (m, 4H, 7-11-CH<sub>2</sub>); 3.68 (d, *J* = 11.6, 1H, 16b-CHH); 4.08 (s, 1H, 14-CH); 6.13 (bt, *J* = 5.8, 1H, 8-NH); 6.93 (q, *J* = 1.4, 6.8, 1H, 3-CH); 7.03 (bt, *J* = 5.9, 1H, 12-NH).

**<sup>13</sup>C-NMR** (CDCl<sub>3</sub>, 100 MHz): δ 12.2 (4-CH<sub>3</sub>); 14.4 (5-CH<sub>3</sub>); 18.7 (17-CH<sub>3</sub>); 18.9 (18-CH<sub>3</sub>); 22.1 (21-CH<sub>3</sub>); 28.3 (6-CH<sub>2</sub>); 29.5 (22-CH<sub>3</sub>); 32.9 (15-C); 34.8 (10-CH<sub>2</sub>); 35.9 (11-CH<sub>2</sub>); 39.7 (7-CH<sub>2</sub>); 71.5 (16-CH<sub>2</sub>); 77.2 (14-CH); 99.1 (19-C); 136.8 (2-C); 136.9 (3-CH); 170.1 (9-CO); 171.1 (13-CO); 190.2 (1-CO)

**ESMS:** *m/z*: 423.3 [M + Na]H<sup>+</sup>, 401 [M]H<sup>+</sup>

### ***E*-2-Methylhex-2-enoyl pantetheine dimethyl ketal 176**<sup>179,58,137</sup>



The obtained product was a yellow oil (0.38 g, 0.88 mmol, 88 %).

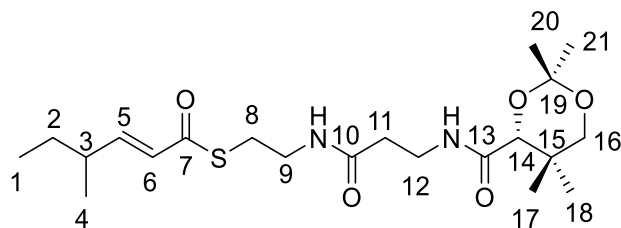
R<sub>f</sub>: 0.45 (EtOAc).

**<sup>1</sup>H-NMR** (CDCl<sub>3</sub>, 400 MHz): δ 0.96 (t, *J* = 7.6, 3H, 6-CH<sub>3</sub>); 0.97 (s, 3H, 19-CH<sub>3</sub>); 1.04 (s, 3H, 20-CH<sub>3</sub>); 1.44-1.55 (m, 2H, 5-CH<sub>2</sub>); 1.41 (s, 3H, 22-CH<sub>3</sub>); 1.46 (s, 3H, 23-CH<sub>3</sub>); 1.87 (s, 3H, 7-CH<sub>3</sub>); 2.20 (q, *J* = 7.4, 2H, 4-CH<sub>2</sub>); 2.42 (t, *J* = 6.1, 2H, 12-CH<sub>2</sub>); 3.05 (t, *J* = 6.5, 2H, 8-CH<sub>2</sub>); 3.27 (d, *J* = 11.2, 1H, 18a-CHH); 3.39-3.62 (m, 4H, 9-13-CH<sub>2</sub>); 3.68 (d, *J* = 11.7, 1H, 18b-CHH); 4.07 (s, 1H, 16-CH); 6.10 (bt, *J* = 5.7, 1H, 10-NH); 6.77 (t, *J* = 1.3, 6.9, 1H, 3-CH); 7.03 (bt, *J* = 5.7, 1H, 14-NH).

**<sup>13</sup>C-NMR** (CDCl<sub>3</sub>, 100 MHz): δ 12.5 (6-CH<sub>3</sub>); 13.9 (7-CH<sub>3</sub>); 18.7 (19-CH<sub>3</sub>); 18.9 (20-CH<sub>3</sub>); 21.8 (5-CH<sub>2</sub>); 22.1 (22-CH<sub>3</sub>); 28.4 (8-CH<sub>2</sub>); 28.8 (4-CH<sub>2</sub>); 29.5 (23-CH<sub>3</sub>); 32.9 (17-C); 34.8 (12-CH<sub>2</sub>); 35.9 (13-CH<sub>2</sub>); 39.7 (9-CH<sub>2</sub>); 71.5 (18-CH<sub>2</sub>); 77.2 (16-CH); 99.1 (21-C); 135.9 (2-C); 142.1 (3-CH); 170.0 (11-CO); 171.2 (15-CO); 193.7 (1-CO).

**ES-MS**: *m/z*: 429.7 [M]<sup>+</sup>

### ***E*-4-Methylhex-2-enoyl pantetheine dimethyl ketal 177**<sup>179,58,137</sup>



The obtained product was a colorless oil (0.35 g, 0.82 mmol, 82 %).

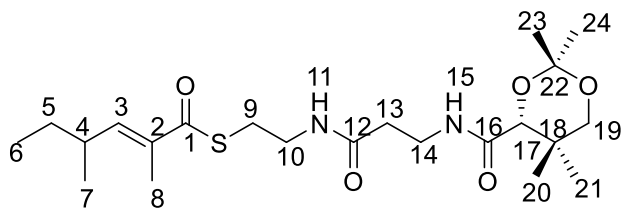
R<sub>f</sub>: 0.47 (EtOAc)

**<sup>1</sup>H-NMR** (CDCl<sub>3</sub>, 400 MHz): δ 0.88 (t, *J* = 7.2, 3H, 6-CH<sub>3</sub>); 0.97 (s, 3H, 19-CH<sub>3</sub>); 1.04 (s, 3H, 20-CH<sub>3</sub>); 1.06 (d, *J* = 6.7, 3H, 7-CH<sub>3</sub>); 1.40-1.45 (m, 2H, 5-CH<sub>2</sub>); 1.41 (s, 3H, 22-CH<sub>3</sub>); 1.46 (s, 3H, 23-CH<sub>3</sub>); 2.17-2.26 (m, 1H, 4-CH); 2.42 (t, *J* = 6.3, 2H, 12-CH<sub>2</sub>); 3.05 (t, *J* = 6.3, 2H, 8-CH<sub>2</sub>); 3.27 (d, *J* = 11.9, 1H, 18a-CHH); 3.40-3.62 (m, 4H, 9-13-CH<sub>2</sub>); 3.68 (d, *J* = 11.7, 1H, 18b-CHH); 4.07 (s, 1H, 16-CH); 6.08 (dd, *J* = 1.2, 15.6, 1H, 2-CH); 6.12 (bt, *J* = 5.1, 1H, 10-NH); 6.77 (dd, *J* = 7.5, 15.5, 1H, 3-CH); 7.03 (bt, *J* = 6.0, 1H, 14-NH).

**<sup>13</sup>C-NMR** (CDCl<sub>3</sub>, 100 MHz): δ 11.6 (6-CH<sub>3</sub>); 12.2 (7-CH<sub>3</sub>); 18.7 (19-CH<sub>3</sub>); 18.9 (20-CH<sub>3</sub>); 22.1 (22-CH<sub>3</sub>); 28.3 (5-CH<sub>2</sub>); 28.4 (8-CH<sub>2</sub>); 29.5 (23-CH<sub>3</sub>); 32.9 (17-C); 34.8 (12-CH<sub>2</sub>); 35.9 (13-CH<sub>2</sub>); 38.1 (4-CH); 39.7 (9-CH<sub>2</sub>); 71.5 (18-CH<sub>2</sub>); 77.2 (16-CH); 99.1 (21-C); 126.7 (2-CH); 151.7 (3-CH); 170.0 (11-CO); 171.2 (15-CO); 190.3 (1-CO)

**ESMS**: *m/z* (%): 429.7 [M]<sup>+</sup>, 452 [M + Na]<sup>+</sup>

### ***E*-2-4-Dimethylhex-2-enoyl pantetheine dimethyl ketal 178**<sup>179,58,137</sup>



The obtained product was a colorless oil (0.38 g, 0.89 mmol, 89%).

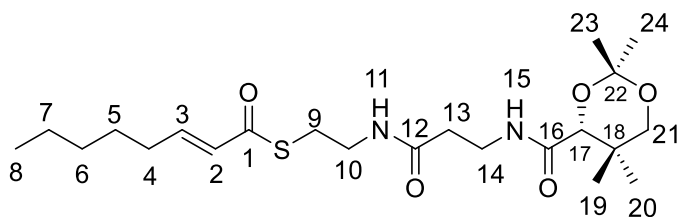
R<sub>f</sub>: 0.45 (EtOAc).

**<sup>1</sup>H-NMR** (CDCl<sub>3</sub>, 400 MHz): δ 0.87 (t, *J* = 7.9, 3H, 6-CH<sub>3</sub>); 0.97 (s, 3H, 20-CH<sub>3</sub>); 1.04 (s, 3H, 21-CH<sub>3</sub>); 1.03 (d, *J* = 7.5, 3H, 7-CH<sub>3</sub>), 1.24-1.49 (m, 3H, 4-CH, 5-CH<sub>2</sub>); 1.41 (s, 3H, 23-CH<sub>3</sub>); 1.46 (s, 3H, 24-CH<sub>3</sub>); 1.88 (s, 3H, 8-CH<sub>3</sub>); 2.42 (t, *J* = 5.9, 2H, 13-CH<sub>2</sub>); 3.05 (t, *J* = 6.4, 2H, 9-CH<sub>2</sub>); 3.27 (d, *J* = 11.4, 1H, 19a-CHH); 3.39-3.64 (m, 4H, 10-14-CH<sub>2</sub>); 3.68 (d, *J* = 11.7, 1H, 19b-CHH); 4.07 (s, 1H, 17-CH); 6.09 (bt, *J* = 5.4, 1H, 11-NH); 6.53 (dd, *J* = 1.4, 9.8, 1H, 3-CH); 7.03 (bt, *J* = 5.8, 1H, 15-NH).

**<sup>13</sup>C-NMR** (CDCl<sub>3</sub>, 100 MHz): δ 11.9 (6-CH<sub>3</sub>); 12.2 (7-CH<sub>3</sub>); 18.7 (20-CH<sub>3</sub>); 18.9 (21-CH<sub>3</sub>); 19.9 (8-CH<sub>3</sub>); 22.1 (23-CH<sub>3</sub>); 28.5 (9-CH<sub>2</sub>); 29.5 (24-CH<sub>3</sub>); 29.6 (5-CH<sub>2</sub>); 32.9 (18-C); 34.4 (13-CH<sub>2</sub>); 35.0 (4-CH); 35.9 (14-CH<sub>2</sub>); 39.7 (10-CH<sub>2</sub>); 71.5 (19-CH<sub>2</sub>); 77.2 (17-CH); 99.1 (22-C); 134.5 (2-C); 147.6 (3-CH); 170.0 (12-CO); 171.2 (16-CO); 193.9 (1-CO).

**ESMS**: *m/z*: 471 [M]<sup>+</sup>, 493 [M + Na]<sup>+</sup>

### ***E*-Oc-2-enoic pantetheine dimethyl ketal 179**<sup>179,58,137</sup>



The obtained product was a colorless oil (0.35 g, 0.79 mmol, 56.1 %)

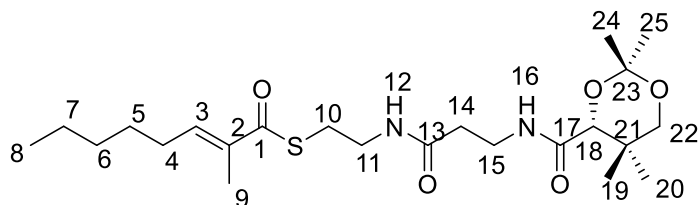
R<sub>f</sub>: 0.56 (EtOAc / PE, 9:1)

**<sup>1</sup>H-NMR** (CDCl<sub>3</sub>, 400 MHz): δ 0.90 (t, *J* = 7.4, 3H, 8-CH<sub>3</sub>); 0.97 (s, 3H, 19-CH<sub>3</sub>); 1.04 (s, 3H, 20-CH<sub>3</sub>); 1.31-1.51 (m, 6H, 5-6-7-CH<sub>2</sub>); 1.41 (s, 3H, 23-CH<sub>3</sub>); 1.46 (s, 3H, 24-CH<sub>3</sub>); 2.21 (g, *J* = 7.3, 2H, 4-CH<sub>2</sub>); 2.42 (t, *J* = 5.8, 2H, 13-CH<sub>2</sub>); 3.05 (t, *J* = 6.7, 2H, 9-CH<sub>2</sub>); 3.27 (d, *J* = 11.8, 1H, 21a-CHH); 3.39-3.62 (m, 4H, 10-14-CH<sub>2</sub>); 3.68 (d, *J* = 12.2, 1H, 21b-CHH); 4.07 (s, 1H, 17-CH); 6.08 (bt, *J* = 5.8, 1H, 11-NH); 6.92 (t, *J* = 1.5, 7.4, 1H, 3-CH); 7.03 (bt, *J* = 5.8, 1H, 15-NH).

**<sup>13</sup>C-NMR** (CDCl<sub>3</sub>, 100 MHz): δ 12.5 (8-CH<sub>3</sub>); 18.7 (19-CH<sub>3</sub>); 18.9 (20-CH<sub>3</sub>); 22.1 (24-CH<sub>3</sub>); 22.5 (7-CH<sub>2</sub>); 28.2 (10-CH<sub>2</sub>); 28.4 (5-CH<sub>2</sub>); 28.8 (4-CH<sub>2</sub>); 29.5 (23-CH<sub>3</sub>); 31.6 (6-CH<sub>2</sub>); 32.9 (18-C); 34.8 (13-CH<sub>2</sub>); 35.9 (14-CH<sub>2</sub>); 39.7 (9-CH<sub>2</sub>); 71.5 (21-CH<sub>2</sub>); 77.2 (17-CH); 99.1 (22-C); 135.7 (2-C); 142.4 (3-CH); 170.0 (12-CO); 171.1 (16-CO); 193.7 (1-CO).

**ESMS**: *m/z*: 443.62 [M]<sup>+</sup>

### ***E*-2-Methyloct-2-enoyl pantetheine dimethyl ketal 180**<sup>179,58,137</sup>



The obtained product was a colorless oil (0.27 g, 0.60 mmol, 60%).

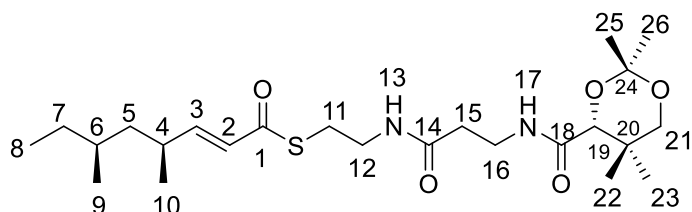
R<sub>f</sub>: 0.47 (EtOAc).

**<sup>1</sup>H-NMR** (CDCl<sub>3</sub>, 400 MHz): δ 0.90 (t, *J* = 7.4, 3H, 8-CH<sub>3</sub>); 0.97 (s, 3H, 19-CH<sub>3</sub>); 1.04 (s, 3H, 20-CH<sub>3</sub>); 1.31-1.51 (m, 6H, 5-6-7-CH<sub>2</sub>); 1.41 (s, 3H, 24-CH<sub>3</sub>); 1.46 (s, 3H, 25-CH<sub>3</sub>); 1.87 (s, 3H, 9-CH<sub>3</sub>); 2.21 (g, *J* = 7.3, 2H, 4-CH<sub>2</sub>); 2.42 (t, *J* = 5.8, 2H, 14-CH<sub>2</sub>); 3.05 (t, *J* = 6.7, 2H, 10-CH<sub>2</sub>); 3.27 (d, *J* = 11.8, 1H, 22a-CHH); 3.39-3.62 (m, 4H, 11-15-CH<sub>2</sub>); 3.68 (d, *J* = 12.2, 1H, 22b-CHH); 4.07 (s, 1H, 18-CH); 6.08 (bt, *J* = 5.8, 1H, 12-NH); 6.92 (t, *J* = 1.5, 7.4, 1H, 3-CH); 7.03 (bt, *J* = 5.8, 1H, 16-NH).

**<sup>13</sup>C-NMR** (CDCl<sub>3</sub>, 100 MHz): δ 12.5 (8-CH<sub>3</sub>); 13.9 (1-CH<sub>3</sub>); 18.7 (19-CH<sub>3</sub>); 18.9 (20-CH<sub>3</sub>); 22.1 (24-CH<sub>3</sub>); 22.5 (7-CH<sub>2</sub>); 28.2 (10-CH<sub>2</sub>); 28.4 (5-CH<sub>2</sub>); 28.8 (4-CH<sub>2</sub>); 29.5 (25-CH<sub>3</sub>); 31.6 (6-CH<sub>2</sub>); 32.9 (21-C); 34.8 (14-CH<sub>2</sub>); 35.9 (15-CH<sub>2</sub>); 39.7 (11-CH<sub>2</sub>); 71.5 (22-CH<sub>2</sub>); 77.2 (18-CH); 99.1 (23-C); 135.7 (2-C); 142.4 (3-CH); 170.0 (13-CO); 171.1 (17-CO); 193.7 (9-CO).

**ESMS**: *m/z*: 457.7 [M]<sup>+</sup>

### ***E*-4*S*-6*S*-Dimethyl-octenoyl-pantethine dimethyl ketal 181**<sup>179,58,137</sup>



The obtained product was colorless oil (0.22 g, 0.47 mmol, 36.2 %)

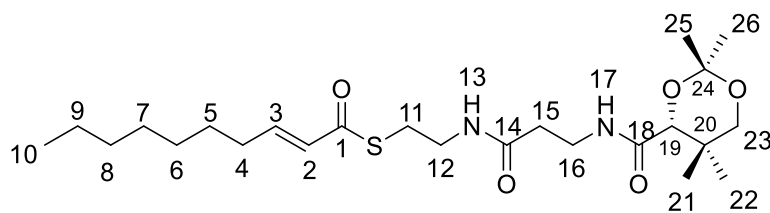
R<sub>f</sub>: 0.47 (EtOAc / PE, 9:1)

**<sup>1</sup>H-NMR** (CDCl<sub>3</sub>, 400 MHz): δ 0.87 (m, 6H, 8-9-CH<sub>3</sub>); 0.97 (s, 3H, 22-CH<sub>3</sub>); 1.04 (s, 3H, 23-CH<sub>3</sub>); 1.13-1.19 (m, 2H, 7-CH<sub>3</sub>), 1.26-1.49 (m, 3H, 5-6-CH<sub>2</sub>); 1.44 (s, 3H, 25-CH<sub>3</sub>); 1.48 (s, 3H, 26-CH<sub>3</sub>); 2.34 (m, 1H, 4-CH); 2.45 (t, *J* = 5.9, 1H, 14-CH); 3.10 (t, *J* = 6.5, 2H, 11-CH<sub>2</sub>); 3.27 (d, *J* = 11.7, 1H, 21a-CHH); 3.42-3.62 (m, 4H, 12-16-CH<sub>2</sub>); 3.70 (d, *J* = 11.6, 1H, 21b-CHH); 4.09 (s, 1H, 19-CH); 6.12 (dd, *J* = 1.0, 16.0, 1H, 2-CH); 6.28 (bt, *J* = 6.3, 1H, 13-NH); 6.80 (d, *J* = 8.2, 15.8, 1H, 3-CH); 7.06 (bt, *J* = 5.8, 1H, 17-NH).

**<sup>13</sup>C-NMR** (CDCl<sub>3</sub>, 100 MHz): δ 11.1 (8-CH<sub>3</sub>); 14.2 (9-CH<sub>3</sub>); 18.7 (22-CH<sub>3</sub>); 18.9 (23-CH<sub>3</sub>); 22.1 (25-CH<sub>3</sub>); 28.2 (11-CH<sub>2</sub>); 29.5 (26-CH<sub>3</sub>); 29.7 (7-CH<sub>2</sub>); 32.9 (6-CH); 32.9 (18-C); 34.4 (4-CH); 34.7 (15-CH<sub>2</sub>); 35.9 (16-CH<sub>2</sub>); 39.7 (12-CH<sub>2</sub>); 43-5 (5-CH) 71.5 (21-CH<sub>2</sub>); 77.2 (19-CH); 99.1 (24-C); 126.6 (2-CH); 152.1 (3-CH); 170.1 (14-CO); 171.1 (18-CO); 190.1 (1-CO).

**ESMS**: *m/z*: 471 [M]<sup>+</sup>, 493 [M + Na]<sup>+</sup>

### *E*-Dec-2-enoyl pantetheine dimethyl ketal **182**<sup>179,58,137</sup>



The obtained product was a colorless oil (0.18 g, 0.38 mmol, 38.3%)

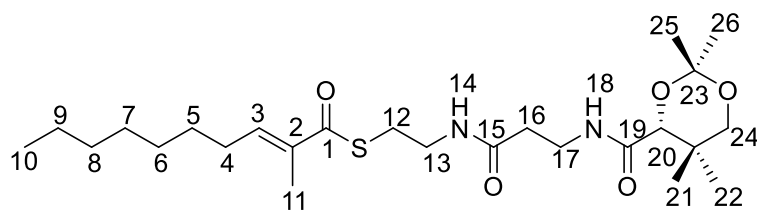
R<sub>f</sub>: 0.49 (EtAc / PE, 9:1)

**<sup>1</sup>H-NMR** (CDCl<sub>3</sub>, 400 MHz): δ 0.88 (t, *J* = 7.6, 3H, 10-CH<sub>3</sub>); 0.97 (s, 3H, 21-CH<sub>3</sub>); 1.04 (s, 3H, 22-CH<sub>3</sub>); 1.30-1.47 (m, 10H, 5-6-7-8-9-CH<sub>2</sub>); 1.41 (s, 3H, 25-CH<sub>3</sub>); 1.46 (s, 3H, 26-CH<sub>3</sub>); 2.21 (q, *J* = 7.1, 2H, 4-CH<sub>2</sub>); 2.42 (t, *J* = 6.2, 2H, 15-CH<sub>2</sub>); 3.05 (t, *J* = 6.2, 2H, 11-CH<sub>2</sub>); 3.27 (d, *J* = 12.2, 1H, 23a-CHH); 3.39-3.62 (m, 4H, 12-16-CH<sub>2</sub>); 3.68 (d, *J* = 12.2, 1H, 23b-CHH); 4.07 (s, 1H, 19-CH); 6.08 (bt, *J* = 5.0, 1H, 13-NH); 6.77 (t, *J* = 1.9, 6.9, 1H, 3-CH); 7.03 (bt, *J* = 5.5, 1H, 17-NH).

**<sup>13</sup>C-NMR** (CDCl<sub>3</sub>, 100 MHz): 14.0 (10-CH<sub>3</sub>); 18.7 (21-CH<sub>3</sub>); 18.9 (22-CH<sub>3</sub>); 22.1 (25-CH<sub>3</sub>); 22.6 (9-CH<sub>2</sub>); 28.4 (5-CH<sub>2</sub>); 28.5 (12-CH<sub>2</sub>); 28.8 (6-CH<sub>2</sub>); 29.1 (7-CH<sub>2</sub>); 29.5 (26-CH<sub>3</sub>); 31.8 (8-CH<sub>2</sub>); 33.2 (4-CH<sub>2</sub>); 34.8 (15-CH<sub>2</sub>); 35.9 (16-CH<sub>2</sub>); 39.7 (11-CH<sub>2</sub>); 71.5 (23-CH<sub>2</sub>); 77.2 (19-CH); 99.1 (20-C); 135.7 (2-C); 142.4 (3-CH); 170.0 (14-CO); 171.2 (18-CO); 193.7 (11-CO).

**ESMS**: *m/z*: 471.76 [M]<sup>+</sup>, 493 [M + Na]<sup>+</sup>

### *E*-2-Methyldec-2-enoyl pantetheine dimethyl ketal **183**<sup>189,58,137</sup>



The obtained product was a colorless oil (0.225 g, 0.53 mmol, 63 %).

R<sub>f</sub>: 0.45 (EtOAc).

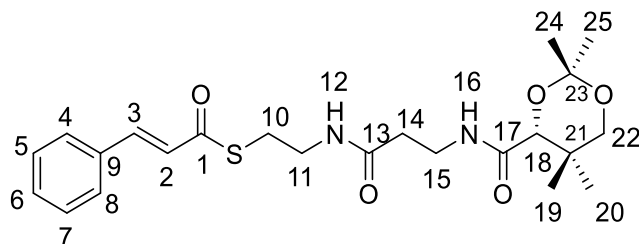
**<sup>1</sup>H-NMR** (CDCl<sub>3</sub>, 400 MHz): δ 0.87 (m, 6H, 8-9-CH<sub>3</sub>); 0.97 (s, 3H, 22-CH<sub>3</sub>); 1.04 (s, 3H, 23-CH<sub>3</sub>); 1.13-1.19 (m, 2H, 7-CH<sub>3</sub>); 1.26-1.49 (m, 3H, 5-6-CH<sub>2</sub>); 1.44 (s, 3H, 25-CH<sub>3</sub>); 1.48 (s, 3H, 26-CH<sub>3</sub>); 2.34 (m, 1H, 4-CH); 2.45 (t, *J* = 5.9, 1H, 14-CH); 3.10 (t, *J* = 6.5, 2H, 11-CH<sub>2</sub>); 3.27 (d, *J* = 11.7, 1H, 21a-CHH); 3.42-3.62 (m, 4H, 12-16-CH<sub>2</sub>); 3.70 (d, *J* = 11.6, 1H, 21b-CHH); 4.09 (s, 1H, 19-CH); 6.12 (dd, *J* = 1.0, 16.0, 1H, 2-CH); 6.28 (bt, *J* = 6.3, 1H, 13-NH); 6.80 (t, *J* = 8.2, 15.8, 1H, 3-CH); 7.06 (bt, *J* = 5.8, 1H, 17-NH).

**<sup>13</sup>C-NMR** (CDCl<sub>3</sub>, 100 MHz): δ 11.1 (8-CH<sub>3</sub>); 14.2 (9-CH<sub>3</sub>); 18.7 (22-CH<sub>3</sub>); 18.9 (23-CH<sub>3</sub>); 22.1 (25-CH<sub>3</sub>); 28.2 (11-CH<sub>2</sub>); 29.5 (26-CH<sub>3</sub>); 29.7 (7-CH<sub>2</sub>); 32.9 (6-CH); 32.9 (18-C); 34.4 (4-CH); 34.7 (15-CH<sub>2</sub>); 35.9 (16-CH<sub>2</sub>); 39.7 (12-CH<sub>2</sub>); 43.5 (5-CH); 71.5 (21-CH<sub>2</sub>); 77.2 (19-CH); 99.1 (24-C); 126.6 (2-CH); 152.1 (3-CH); 170.1 (14-CO); 171.1 (18-CO); 190.1 (1-CO).

**ESMS**: *m/z*: 485.7 [M]<sup>+</sup>



### ***E*-Cinnamon pantetheine dimethyl ketal 184**<sup>179,58,137</sup>



The obtained product was a white solid (0.45, 0.73 mmol, 48.8 %)

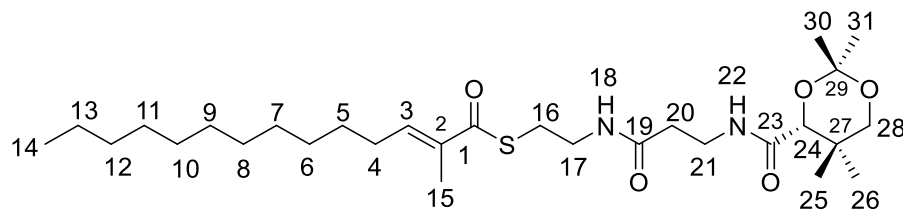
R<sub>f</sub>: 0.42 (EtOAc / PE, 9:1).

<sup>1</sup>H-NMR (CDCl<sub>3</sub>, 400 MHz): 0.97 (s, 3H, 19-CH<sub>3</sub>); 1.04 (s, 3H, 20-CH<sub>3</sub>); 1.41 (s, 3H, 24-CH<sub>3</sub>); 1.46 (s, 3H, 25-CH<sub>3</sub>); 2.42 (t, *J* = 5.8, 2H, 14-CH<sub>2</sub>); 3.05 (t, *J* = 6.7, 2H, 10-CH<sub>2</sub>); 3.27 (d, *J* = 11.8, 1H, 22a-CHH); 3.39-3.62 (m, 4H, 11-15-CH<sub>2</sub>); 3.68 (d, *J* = 12.2, 1H, 22b-CHH); 4.07 (s, 1H, 18-CH); 6.08 (bt, *J* = 5.8, 1H, 12-NH); 6.70 (d, 1H, 2-CH); 7.03 (bt, *J* = 5.8, 1H, 16-NH), 7.73 (dd, 3H, 5-,6-,7-CH); 7.52 (m, 2H, 4-,8-CH); 7.59 (d, *J* = 15.82, 1H, 3-CH)

<sup>13</sup>C-NMR (CDCl<sub>3</sub>, 100 MHz): δ 18.7 (19-CH<sub>3</sub>); 18.9 (20-CH<sub>3</sub>); 22.1 (24-CH<sub>3</sub>); 28.2 (10-CH<sub>2</sub>); 29.5 (25-CH<sub>3</sub>); 32.9 (21-C); 34.8 (14-CH<sub>2</sub>); 35.9 (15-CH<sub>2</sub>); 39.7 (11-CH<sub>2</sub>); 71.5 (22-CH<sub>2</sub>); 77.2 (18-CH); 99.1 (23-C); 123 (2-C) 127-128.5 (4-C, 5-C, 6-C, 7-C, 8-C) 135.7 (9-C); 143 (3-C), 170.0 (13-CO); 171.1 (17-CO); 193.7 (1-CO).

ESMS: *m/z*: 449 [M]H<sup>+</sup>, 472 [M + Na]H<sup>+</sup>

### ***E*-2-Methyltetradec-2-enoyl panthetine dimethyl ketal 185**<sup>179,58,137</sup>



The obtained product was white solid (0.12 g, 0.22 mmol, 18.5 %)

R<sub>f</sub>: 0.50 (EtOAc / PE, 9:1).

<sup>1</sup>H-NMR (CDCl<sub>3</sub>, 400 MHz): δ [ppm]: 0.88 (t, 3H, 14-CH<sub>3</sub>); 0.97 (s, 3H, 25-CH<sub>3</sub>); 1.04 (s, 3H, 26-CH<sub>3</sub>); 1.30-1.47 (m, 18H, 5-13-CH<sub>2</sub>); 1.41 (s, 3H, 30-CH<sub>3</sub>); 1.46 (s, 3H, 31-CH<sub>3</sub>); 1.87 (s, 3H, 15-CH<sub>3</sub>); 2.21 (q, 2H, 4-CH<sub>2</sub>); 2.42 (t, 2H, 20-CH<sub>2</sub>); 3.05 (t, *J* = 6.2 Hz, 2H, 16-CH<sub>2</sub>); 3.27 (d, *J* = 12.2 Hz, 1H, 28a-CH<sub>2</sub>); 3.39-3.62 (m, 4H, 17-21-CH<sub>2</sub>); 3.68 (d, *J* = 12.2 Hz, 1H, 28b-CH<sub>2</sub>); 4.07 (s, 1H, 24-CH); 6.08 (bt, *J* = 5.0 Hz, 1H, 18-NH); 6.77 (t, 1H, 3-CH); 7.03 (t, *J* = 5.5 Hz, 1H, 22-NH).

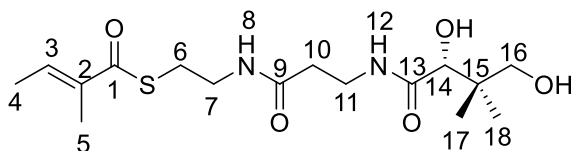
<sup>13</sup>C-NMR (CDCl<sub>3</sub>, 100 MHz): δ [ppm]: 12.5 (14-CH<sub>3</sub>); 14.0 (15-CH<sub>3</sub>); 18.7 (25-CH<sub>3</sub>); 18.9 (26-CH<sub>3</sub>); 22.1 (30-31-CH<sub>3</sub>); 22.6 (13-CH<sub>2</sub>); 28.5 (12-CH<sub>2</sub>); 29.1-29.9 (5-11-CH<sub>2</sub>); 31.6 (4-CH<sub>2</sub>); 34.8 (21-CH<sub>2</sub>); 39.7 (20-CH<sub>2</sub>); 71.5 (28-CH<sub>2</sub>); 77.2 (18-CH); 99.1 (29-C); 135.7 (29-C); 142.4 (2-CH); 170.0 (23-CO); 171.1 (19-CO); 193.7 (1-CO).

ESMS: *m/z*: 541.3 [M]H<sup>+</sup>, 573 [M + Na]H<sup>+</sup>

### 8.3.4 Synthesis of panthetine substrates first method<sup>137</sup>

The acyl panthetine dimethyl ketal was stirred in a mixture of acetonitrile and water (1:1) and 10% TFA for 20 minutes. The reaction was followed by TLC and LCMS. After that the solvents were liphophilized. The product was purified by preparative HPLC (acetonitrile).

#### Tigloyl-pantethine 77<sup>179,58,137</sup>



The obtained product was a colourless oil (0.167 g, 0.46 mmol, 90 %)

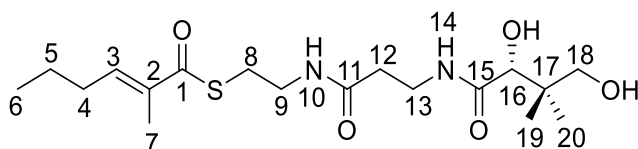
R<sub>f</sub>: 0.44 (DCM / MeOH).

<sup>1</sup>H-NMR (CDCl<sub>3</sub>, 400 MHz): δ 0.92 (s, 3H, 17-CH<sub>3</sub>); 0.92 (s, 3H, 18-CH<sub>3</sub>); 1.82-1.87 (m, 6H, 4-5-CH<sub>3</sub>); 2.41 (t, *J* = 5.9, 2H, 10-CH<sub>2</sub>); 3.01-3.10 (m, 2H, 6-CH<sub>2</sub>); 3.32-3.60 (m, 6H, 7-11-16-CH<sub>2</sub>); 3.99 (s, 1H, 14-CH); 6.33 (bt, *J* = 5.8, 1H, 8-NH); 6.86 (q, *J* = 1.3, 6.8, 1H, 3-CH); 7.40 (bt, *J* = 5.6, 1H, 12-NH).

<sup>13</sup>C-NMR (CDCl<sub>3</sub>, 100 MHz): δ 12.2 (5-CH<sub>3</sub>); 14.5 (4-CH<sub>3</sub>), 20.4 (17-CH<sub>3</sub>); 21.5 (18-CH<sub>3</sub>); 28.2 (6-CH<sub>2</sub>); 35.2 (11-CH<sub>2</sub>); 35.6 (10-CH<sub>2</sub>); 39.3 (15-C); 39.9 (7-CH<sub>2</sub>); 70.9 (16-CH<sub>2</sub>); 77.9 (14-CH); 136.8 (3-C); 137.3 (2-CH); 171.8 (9-CO); 173.5 (13-CO); 194.1 (1-COS).

ESMS: *m/z*: 383 [M + Na]H<sup>+</sup>, 361 [M]H<sup>+</sup>, 343 [M - H<sub>2</sub>O]H<sup>+</sup>

#### E-2-Methylhex-2-enoylpantethine 83<sup>179,58,137</sup>



The obtained product was a colourless oil (0.009 g, 0.02 mmol, 15 %)

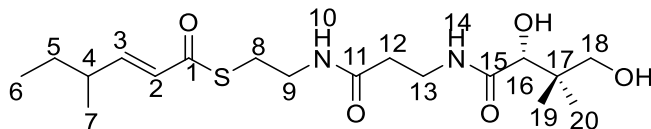
R<sub>f</sub>: 0.44 (DCM / MeOH).

<sup>1</sup>H-NMR (CDCl<sub>3</sub>, 400 MHz): δ 0.93 (q, 3H, 6-CH<sub>3</sub>); 0.95 (t, *J* = 7.4, 3H, 19-CH<sub>3</sub>); 1.02 (s, 3H, 20-CH<sub>3</sub>); 1.46-1.55 (m, 2H, 5-CH<sub>2</sub>); 1.88 (d, *J* = 1.4, 7-CH<sub>3</sub>); 2.20 (dq, *J* = 1.3, *J* = 7.0, 2H, 4-CH<sub>2</sub>); 2.41 (t, *J* = 5.9, 2H, 12-CH<sub>2</sub>); 3.00-3.13 (m, 2H, 8-CH<sub>2</sub>); 3.35-3.59 (m, 6H, 9-13-18-CH<sub>2</sub>); 3.99 (s, 1H, 16-CH); 6.23 (bt, *J* = 5.7, 1H, 10-NH); 6.77 (tq, *J* = 1, 4, *J* = 7.4, 1H, 3-CH); 7.38 (bt, *J* = 5.9, 1H, 14-NH).

<sup>13</sup>C-NMR (CDCl<sub>3</sub>, 100 MHz): δ 12.5 (6-CH<sub>3</sub>), 13.9 (7-CH<sub>3</sub>), 20.4 (19-CH<sub>3</sub>); 21.7 (20-CH<sub>3</sub>); 21.8 (5-CH<sub>2</sub>); 28.3 (8-CH<sub>2</sub>); 30.8 (4-CH<sub>2</sub>); 35.1 (12-CH<sub>2</sub>); 35.5 (13-CH<sub>2</sub>); 39.4 (17-C); 39.8 (9-CH<sub>2</sub>); 70.9 (18-CH<sub>2</sub>); 77.7 (16-CH); 135.7 (2-C); 142.5 (3-CH); 171.6 (11-CO); 173.4 (15-CO); 194.2 (1-COS).

ESMS: *m/z*: 411 [M + Na]H<sup>+</sup>, 389 [M]H<sup>+</sup>

### ***E*-4-Methylhex-2-enoyl pantetheine 84**<sup>179,58,137</sup>



The obtained product was a colourless oil (0.013 g, 0.03 mmol, 21 %)

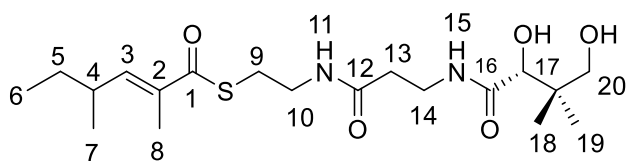
R<sub>f</sub>: 0.1 (DCM / MeOH).

<sup>1</sup>H-NMR (CDCl<sub>3</sub>, 400 MHz): δ 0.88 (t, *J* = 7.5, 3H, 6-CH<sub>3</sub>); 0.98 (s, 3H, 19-CH<sub>3</sub>); 1.01 (s, 3H, 20-CH<sub>3</sub>); 1.05 (d, *J* = 6.9, 7-CH<sub>3</sub>); 1.38-1.47 (m, 2H, 5-CH<sub>2</sub>); 2.19-2.27 (m, 1H, 4-CH); 2.41 (t, *J* = 5.8, 2H, 12-CH<sub>2</sub>); 2.94-3.15 (m, 2H, 8-CH<sub>2</sub>); 3.33-3.58 (m, 6H, 9-13-18-CH<sub>2</sub>); 4.00 (s, 1H, 16-CH); 6.08 (dd, *J* = 1.4, *J* = 15.4, 1H, 2-CH); 6.40 (bt, *J* = 5.4, 1H, 10-NH); 6.83 (dd, *J* = 7.4, *J* = 15.4, 1H, 3-CH); 7.40 (bt, *J* = 5.4, 1H, 14-NH).

<sup>13</sup>C-NMR (CDCl<sub>3</sub>, 100 MHz): δ 11.6 (6-CH<sub>3</sub>), 18.7 (7-CH<sub>3</sub>), 20.4 (19-CH<sub>3</sub>); 21.7 (20-CH<sub>3</sub>); 28.2 (5-CH<sub>2</sub>); 28.8 (8-CH<sub>2</sub>); 35.1 (12-CH<sub>2</sub>); 35.8 (13-CH<sub>2</sub>); 38.2 (4-CH); 39.4 (16-C); 39.7 (9-CH<sub>2</sub>); 70.8 (18-CH<sub>2</sub>); 77.6 (16-CH); 126.7 (2-CH); 152.1 (3-CH); 171.7 (11-CO); 173.6 (15-CO); 190.8 (1-COS).

ESMS: *m/z*: 411 [M + Na]H<sup>+</sup>, 389 [M]H<sup>+</sup>

### ***E*-2,4-Dimethylhex-2-enoylpantetheine 86**<sup>179,58,137</sup>



The obtained product was a colourless oil (0.012 g, 0.029 mmol, 20 %)

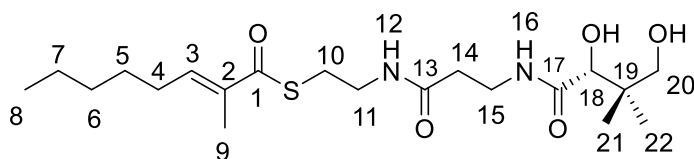
R<sub>f</sub>: 0.42 (DCM / MeOH).

<sup>1</sup>H-NMR (CDCl<sub>3</sub>, 400 MHz): δ 0.86 (t, *J* = 7.4, 3H, 6-CH<sub>3</sub>); 0.92 (s, 3H, 20-CH<sub>3</sub>); 1.03 (s, 3H, 21-CH<sub>3</sub>); 1.03 (d, *J* = 2.3, 3H, 7-CH<sub>3</sub>); 1.30-1.51 (m, 4H, 5-CH<sub>2</sub>); 1.88 (d, *J* = 1.1, 8-CH<sub>3</sub>), 2.41 (t, *J* = 5.6, 2H, 13-CH<sub>2</sub>); 2.40-2.49 (m, 1H, 4-CH); 2.99-3.14 (m, 2H, 9-CH<sub>2</sub>); 3.35-3.59 (m, 6H, 10-14-19-CH<sub>2</sub>); 3.99 (s, 1H, 17-CH); 6.19 (bt, *J* = 5.7, 1H, 11-NH); 6.54 (dq, *J* = 1, 3, *J* = 9.8, 1H, 3-CH); 7.36 (bt, *J* = 5.9, 1H, 15-NH).

<sup>13</sup>C-NMR (CDCl<sub>3</sub>, 100 MHz): δ 11.9 (6-CH<sub>3</sub>), 12.7 (7-CH<sub>3</sub>), 19.5 (8-CH<sub>3</sub>), 20.4 (20-CH<sub>3</sub>); 21.7 (21-CH<sub>3</sub>); 28.3 (9-CH<sub>2</sub>); 29.6 (5-CH); 35.1 (13-CH<sub>2</sub>); 35.6 (14-CH<sub>2</sub>); 39.4 (18-C); 39.8 (10-CH<sub>2</sub>); 71.0 (19-CH<sub>2</sub>); 77.7 (17-CH); 134.4 (2-C); 147.8 (3-CH); 171.4 (11-CO); 173.3 (15-CO); 194.5 (1-COS).

ESMS: *m/z*: 425 [M + Na]H<sup>+</sup>, 403 [M]H<sup>+</sup>, 385 [M - H<sub>2</sub>O]H<sup>+</sup>

### ***E*-2-Methyloct-2-enoylpantetheine 91**<sup>179,58,137</sup>



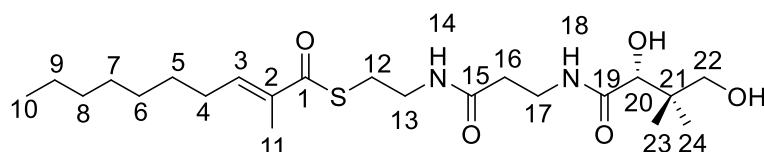
The obtained product was a colourless oil (0.01 g, 0.024 mmol, 17%)

R<sub>f</sub>: 0.45 (DCM / MeOH).

**<sup>1</sup>H-NMR** (CDCl<sub>3</sub>, 400 MHz): δ 0.90 (t, *J* = 6.9, 3H, 8-CH<sub>3</sub>); 0.91 (s, 3H, 21-CH<sub>3</sub>); 1.01 (s, 3H, 22-CH<sub>3</sub>); 1.25-1.36 (m, 4H, 6-7-CH<sub>2</sub>); 1.41-1.50 (m, 2H, 5-CH<sub>2</sub>); 1.85 (s, 3H, 9-CH<sub>3</sub>); 2.20 (q, *J* = 7.5, 2H, 4-CH<sub>2</sub>); 2.41 (t, *J* = 6.6, 2H, 14-CH<sub>2</sub>); 2.99-3.12 (m, 2H, 10-CH<sub>2</sub>); 3.35-3.58 (m, 6H, 11-15-20-CH<sub>2</sub>); 3.99 (s, 1H, 18-CH); 6.37 (bt, *J* = 5.7, 1H, 12-NH); 6.77 (t, *J* = 1.3, *J* = 7.6, 1H, 3-CH); 7.41 (bt, *J* = 6.2, 1H, 16-NH).  
**<sup>13</sup>C-NMR** (CDCl<sub>3</sub>, 100 MHz): δ 12.5 (9-CH<sub>3</sub>); 14.1 (8-CH<sub>3</sub>); 20.4 (21-CH<sub>3</sub>); 21.6 (22-CH<sub>3</sub>); 22.5 (7-CH<sub>2</sub>); 28.1 (5-CH<sub>2</sub>); 28.2 (4-CH<sub>2</sub>); 28.8 (10-CH<sub>2</sub>); 31.6 (6-CH<sub>2</sub>); 35.1 (14-CH<sub>2</sub>); 35.6 (15-CH<sub>2</sub>); 39.4 (19-C); 39.8 (11-CH<sub>2</sub>); 70.9 (20-CH<sub>2</sub>); 77.8 (18-CH); 135.6 (2-C); 142.7 (3-CH); 171.7 (13-CO); 173.2 (17-CO); 194.2 (1-COS).

**ESMS:** *m/z*: 439 [M + Na]H<sup>+</sup>, 417 [M]H<sup>+</sup>, 399 [M - H<sub>2</sub>O]H<sup>+</sup>

### ***E*-2-Methyldec-2-enoylpantetheine 95**<sup>179,58,137</sup>



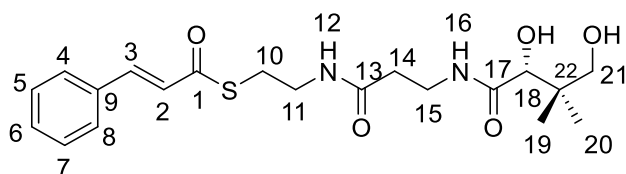
The obtained product was a colourless oil (0.01 g, 0.02 mmol, 17 %)

R<sub>f</sub>: 0.48 (DCM / MeOH).

**<sup>1</sup>H-NMR** (CDCl<sub>3</sub>, 400 MHz): δ 0.88 (q, *J* = 7.2, 3H, 10-CH<sub>3</sub>); 0.93 (s, 3H, 23-CH<sub>3</sub>); 1.04 (s, 3H, 24-CH<sub>3</sub>); 1.28-1.31 (m, 8H, 6-9-CH<sub>2</sub>); 1.43-1.48 (m, 2H, 5-CH<sub>2</sub>); 1.87 (s, 3H, 11-CH<sub>3</sub>); 2.22 (q, *J* = 7.2, 2H, 4-CH<sub>2</sub>); 2.41 (t, *J* = 6.2, 2H, 16-CH<sub>2</sub>); 3.00-3.14 (m, 2H, 12-CH<sub>2</sub>); 3.35-3.63 (m, 6H, 13-17-22-CH<sub>2</sub>); 3.99 (s, 1H, 20-CH); 6.09 (bt, *J* = 5.6, 1H, 14-NH); 6.78 (tq, *J* = 1.3, *J* = 7.4, 1H, 3-CH); 7.38 (bt, *J* = 6.6, 1H, 18-NH).  
**<sup>13</sup>C-NMR** (CDCl<sub>3</sub>, 100 MHz): δ 12.5 (11-CH<sub>3</sub>); 14.1 (10-CH<sub>3</sub>); 20.4 (23-CH<sub>3</sub>); 21.7 (24-CH<sub>3</sub>); 22.6 (9-CH<sub>2</sub>); 28.3 (5-CH<sub>2</sub>); 28.5 (12-CH<sub>2</sub>); 29.1 (7-CH<sub>2</sub>); 29.4 (6-CH<sub>2</sub>); 31.8 (8-CH<sub>2</sub>); 35.1 (16-CH<sub>2</sub>); 35.6 (17-CH<sub>2</sub>); 39.4 (21-C); 39.8 (13-CH<sub>2</sub>); 70.9 (22-CH<sub>2</sub>); 77.8 (19-CH); 135.6 (2-C); 142.8 (3-CH); 171.7 (15-CO); 173.2 (19-CO); 194.2 (1-COS).

**ESMS:** *m/z*: 467 [M + Na]H<sup>+</sup>, 445 [M]H<sup>+</sup>, 427 [M - H<sub>2</sub>O]H<sup>+</sup>

### ***E*-Cinnamon acid pantetheine 113**<sup>179,58,137</sup>



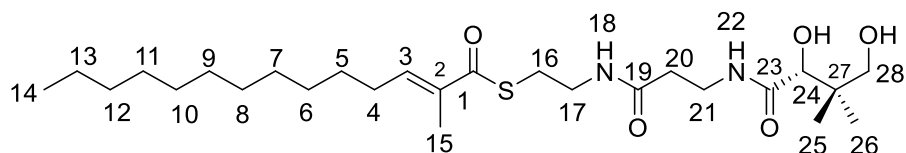
The obtained product was a colourless oil (0.03 g, 0.07 mmol, 26 %)

R<sub>f</sub>: 0.5 (DCM / MeOH).

**<sup>1</sup>H-NMR** (CDCl<sub>3</sub>, 400 MHz): 0.97 (s, 3H, 19-CH<sub>3</sub>); 1.04 (s, 3H, 20-CH<sub>3</sub>); 2.42 (t, *J* = 5.8, 2H, 14-CH<sub>2</sub>); 3.05 (t, *J* = 6.7, 2H, 10-CH<sub>2</sub>); 3.27 (d, *J* = 11.8, 1H, 21a-CHH); 3.39-3.62 (m, 4H, 11-15-CH<sub>2</sub>); 3.68 (d, *J* = 12.2, 1H, 21b-CHH); 4.07 (s, 1H, 18-CH); 6.08 (bt, *J* = 5.8, 1H, 12-NH); 6.70 (d, 1H, 2-CH); 7.03 (bt, *J* = 5.8, 1H, 16-NH), 7.73 (dd, 3H, 5-,6-,7-CH); 7.52 (m, 2H, 4-,8-CH); 7.59 (d, *J* = 15.82, 1H, 3-CH)  
**<sup>13</sup>C-NMR** (CDCl<sub>3</sub>, 100 MHz): δ 18.7 (19-CH<sub>3</sub>); 18.9 (20-CH<sub>3</sub>); 28.2 (10-CH<sub>2</sub>); 70.2 (21-C); 34.8 (14-CH<sub>2</sub>); 35.9 (15-CH<sub>2</sub>); 39.7 (11-CH<sub>2</sub>); 77.2 (18-CH); 48.1 (22-C); 123 (2-C) 127-128.5 (4-C, 5-C, 6-C, 7-C, 8-C) 135.7 (9-C); 143 (3-C), 170.0 (13-CO); 171.1 (17-CO); 193.7 (1-CO).

ESMS:  $m/z$ : 432 [M + Na]H<sup>+</sup>, 409 [M]H<sup>+</sup>

### ***E*-2-Methyltetradec-2-enoylpantetheine 112**<sup>179,58,137</sup>



The obtained product was a colourless oil (0.01 g, 0.02 mmol, 16 %)

R<sub>f</sub>: 0.3 (DCM / MeOH).

<sup>1</sup>H-NMR (CDCl<sub>3</sub>, 400 MHz): δ [ppm]: 0.88 (t, 3H, 14-CH<sub>3</sub>); 0.97 (s, 3H, 25-CH<sub>3</sub>); 1.04 (s, 3H, 26-CH<sub>3</sub>); 1.30-1.47 (m, 18H, 5-13-CH<sub>2</sub>); 1.87 (s, 3H, 15-CH<sub>3</sub>); 2.21 (q, 2H, 4-CH<sub>2</sub>); 2.42 (t, 2H, 20-CH<sub>2</sub>); 3.05 (t,  $J = 6.2$  Hz, 2H, 16-CH<sub>2</sub>); 3.27 (d,  $J = 12.2$  Hz, 1H, 28a-CH<sub>2</sub>); 3.39-3.62 (m, 4H, 17-21-CH<sub>2</sub>); 3.68 (d,  $J = 12.2$  Hz, 1H, 28-CH<sub>2</sub>); 4.07 (s, 1H, 24-CH); 6.08 (bt,  $J = 5.0$  Hz, 1H, 18-NH); 6.77 (t, 1H, 3-CH); 7.03 (t,  $J = 5.5$  Hz, 1H, 22-NH).

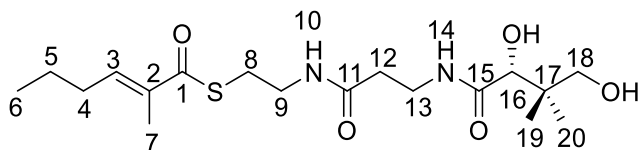
<sup>13</sup>C-NMR (CDCl<sub>3</sub>, 100 MHz): δ[ppm]: 12.5 (14-CH<sub>3</sub>); 14.0 (15-CH<sub>3</sub>); 18.7 (25-CH<sub>3</sub>); 18.9 (26-CH<sub>3</sub>); 22.1 (30-31-CH<sub>3</sub>); 22.6 (13-CH<sub>2</sub>); 28.5 (12-CH<sub>2</sub>); 29.1-29.9 (5-11-CH<sub>2</sub>); 31.6 (4-CH<sub>2</sub>); 34.8 (21-CH<sub>2</sub>); 39.7 (20-CH<sub>2</sub>); 71.5 (28-CH<sub>2</sub>); 77.2 (18-CH); 142.4 (2-CH); 170.0 (23-CO); 171.1 (19-CO); 193.7 (1-CO).

ESMS:  $m/z$ : 501.3 [M]H<sup>+</sup>, 523 [M + Na]H<sup>+</sup>

### **8.3.5 Synthesis of panthetine substrates second method**

The corresponding Acyl pantetheine dimethyl ketal was dissolved in 8 ml Acetonitrile. Indium chloride (2 eq.) and water (4 eq.) were added to the mixture. The reaction was stirred for 3 h by RT. After 3 h the reaction was controlled by TLC and LCMS. Afterwards was the Acetonitrile evaporated and purified by flash column chromatography (DCM / MeOH).

### ***E*-2-Methylhex-2-enoylpantetheine 83**<sup>74,190</sup>



The obtained product was a colourless oil (0.2 g, 0.51 mmol, 90 %)

R<sub>f</sub>: 0.44 (DCM / MeOH).

<sup>1</sup>H-NMR (CDCl<sub>3</sub>, 400 MHz): δ 0.93 (s, 3H, 1-CH<sub>3</sub>); 0.95 (t,  $J = 7.4$ , 3H, 19-CH<sub>3</sub>); 1.02 (s, 3H, 20-CH<sub>3</sub>); 1.46-1.55 (m, 2H, 5-CH<sub>2</sub>); 1.88 (d,  $J = 1.4$ , 7-CH<sub>3</sub>); 2.20 (dq,  $J = 1.3$ ,  $J = 7.0$ , 2H, 4-CH<sub>2</sub>); 2.41 (t,  $J = 5.9$ , 2H, 12-CH<sub>2</sub>); 3.00-3.13 (m, 2H, 8-CH<sub>2</sub>); 3.35-3.59 (m, 6H, 9-13-18-CH<sub>2</sub>); 3.99 (s, 1H, 16-CH); 6.23 (bt,  $J = 5.7$ , 1H, 10-NH); 6.77 (t,  $J = 1$ ,  $J = 7.4$ , 1H, 3-CH); 7.38 (bt,  $J = 5.9$ , 1H, 14-NH).

<sup>13</sup>C-NMR (CDCl<sub>3</sub>, 100 MHz): δ 12.5 (6-CH<sub>3</sub>), 13.9 (7-CH<sub>3</sub>), 20.4 (19-CH<sub>3</sub>); 21.7 (20-CH<sub>3</sub>); 21.8 (5-CH<sub>2</sub>); 28.3 (8-CH<sub>2</sub>); 30.8 (4-CH<sub>2</sub>); 35.1 (12-CH<sub>2</sub>); 35.5 (13-CH<sub>2</sub>); 39.4 (17-C); 39.8 (9-CH<sub>2</sub>); 70.9 (18-CH<sub>2</sub>); 77.7 (16-CH); 135.7 (2-C); 142.5 (3-CH); 171.6 (11-CO); 173.4 (15-CO); 194.2 (1-COS).

ESMS:  $m/z$ : 411 [M + Na]H<sup>+</sup>, 389 [M]H<sup>+</sup>

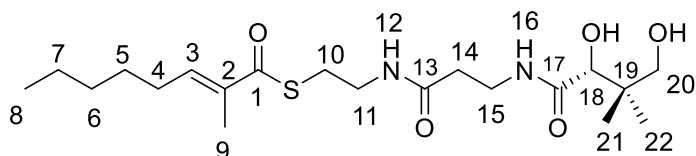


**<sup>1</sup>H-NMR** (CDCl<sub>3</sub>, 400 MHz): δ 0.91 (t, *J* = 6.7, 3H, 8-CH<sub>3</sub>); 0.95 (s, 3H, 20-CH<sub>3</sub>); 1.06 (s, 3H, 21-CH<sub>3</sub>); 1.28-1.61 (m, 6H, 5-6-7-CH<sub>2</sub>); 2.21 (dq, *J* = 1.4, *J* = 7.3, 2H, 4-CH<sub>2</sub>); 2.43 (t, *J* = 6.0, 2H, 13-CH<sub>2</sub>); 3.05-3.20 (m, 2H, 9-CH<sub>2</sub>); 3.38-3.62 (m, 6H, 10-14-19-CH<sub>2</sub>); 4.01 (s, 1H, 17-CH); 6.13 (t, *J* = 4.07, 1H, 2-CH); 6.17 (dt, *J* = 1.4, *J* = 15.4, 1H, 11-NH); 6.96 (dt, *J* = 6, 9, *J* = 15.4, 1H, 3-CH); 7.34 (bt, *J* = 5.82, 1H, 15-NH).

**<sup>13</sup>C-NMR** (CDCl<sub>3</sub>, 100 MHz): δ 13.9 (8-CH<sub>3</sub>), 20.4 (20-CH<sub>3</sub>); 21.7 (21-CH<sub>3</sub>); 22.4 (7-CH<sub>2</sub>); 27.6 (6-CH<sub>2</sub>); 28.1 (9-CH<sub>2</sub>); 31.7 (5-CH<sub>2</sub>); 32.3 (4-CH<sub>2</sub>); 35.1 (13-CH<sub>2</sub>); 35.5 (14-CH<sub>2</sub>); 39.3 (18-C); 39.9 (10-CH<sub>2</sub>); 70.9 (19-CH<sub>2</sub>); 77.8 (17-CH); 128.1 (2-CH); 147.3 (3-CH); 171.6 (12-CO); 173.2 (16-CO); 190.6 (1-CO).

**ESMS:** *m/z*: 425 [M + Na]H<sup>+</sup>, 403 [M]H<sup>+</sup>

### ***E*-2-Methyloct-2-enoylpantetheine 91**<sup>74,190</sup>



The obtained product was a yellow oil. (0.19 g; 91.2 %)

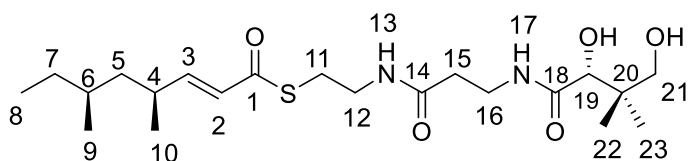
R<sub>f</sub>: 0.45 (DCM / MeOH).

**<sup>1</sup>H-NMR** (CDCl<sub>3</sub>, 400 MHz): δ 0.90 (t, *J* = 6.9, 3H, 8-CH<sub>3</sub>); 0.91 (s, 3H, 21-CH<sub>3</sub>); 1.01 (s, 3H, 22-CH<sub>3</sub>); 1.25-1.36 (m, 4H, 6-7-CH<sub>2</sub>); 1.41-1.50 (m, 2H, 5-CH<sub>2</sub>); 1.85 (s, 3H, 9-CH<sub>3</sub>); 2.20 (q, *J* = 7.5, 2H, 4-CH<sub>2</sub>); 2.41 (t, *J* = 6.6, 2H, 14-CH<sub>2</sub>); 2.99-3.12 (m, 2H, 10-CH<sub>2</sub>); 3.35-3.58 (m, 6H, 11-15-20-CH<sub>2</sub>); 3.99 (s, 1H, 18-CH); 6.37 (bt, *J* = 5.7, 1H, 12-NH); 6.77 (t, *J* = 1.3, *J* = 7.6, 1H, 3-CH); 7.41 (bt, *J* = 6.2, 1H, 16-NH).

**<sup>13</sup>C-NMR** (CDCl<sub>3</sub>, 100 MHz): δ 12.5 (9-CH<sub>3</sub>); 14.1 (8-CH<sub>3</sub>); 20.4 (21-CH<sub>3</sub>); 21.6 (22-CH<sub>3</sub>); 22.5 (7-CH<sub>2</sub>); 28.1 (5-CH<sub>2</sub>); 28.2 (4-CH<sub>2</sub>); 28.8 (10-CH<sub>2</sub>); 31.6 (6-CH<sub>2</sub>); 35.1 (14-CH<sub>2</sub>); 35.6 (15-CH<sub>2</sub>); 39.4 (19-C); 39.8 (11-CH<sub>2</sub>); 70.9 (20-CH<sub>2</sub>); 77.8 (18-CH); 135.6 (2-C); 142.7 (3-CH); 171.7 (13-CO); 173.2 (17-CO); 194.2 (1-COS).

**ESMS:** *m/z*: 439 [M + Na]H<sup>+</sup>, 417 [M]H<sup>+</sup>, 399 [M - H<sub>2</sub>O]H<sup>+</sup>

### **6*S*,4*S*-2*E*-Dimethyloct-2-enoylpantetheine 93a**<sup>74,190</sup>



The obtained product was a yellow oil (0.16 g, 0.37 mmol, 73 %)

R<sub>f</sub>: 0.42 (DCM / MeOH).

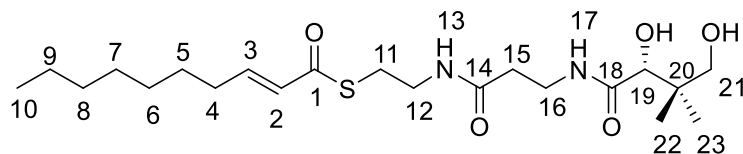
**<sup>1</sup>H-NMR** (CDCl<sub>3</sub>, 400 MHz): δ 0.81-0.90 (m, 6H, 8-9-CH<sub>3</sub>), 0.92 (s, 3H, 21-CH<sub>3</sub>), 1.03 (s, 3H, 22-CH<sub>3</sub>), 1.05 (m, 3H, 10-CH<sub>3</sub>), 1.10-1.44 (m, 6H, 5-7-CH<sub>2</sub>, 6-CH), 2.42 (t, 2H, *J* = 6.0, 15-CH<sub>2</sub>), 2.37-2.41 (m, 1H, 4-CH), 3.03-3.17 (m, 2H, 11-CH<sub>2</sub>), 3.36-3.58 (m, 6H, 12-16-21-CH<sub>2</sub>), 3.99 (s, 1H, 19-CH), 6.08 (dd, 1H, *J* = 1.4, *J* = 15.6, 2-CH), 6.23 (bt, *J* = 6.0, 1H, 13-NH), 6.79 (dd, *J* = 7.6, 1H, 3-CH); 7.37 (bt, *J* = 6.0, 1H, 17-NH).

**<sup>13</sup>C-NMR** (CDCl<sub>3</sub>, 100 MHz): 11.3 (8-CH<sub>3</sub>), 18.8 (9-CH<sub>3</sub>), 20.2 (10-CH<sub>3</sub>), 20.5 (21-CH<sub>3</sub>), 21.9 (22-CH<sub>3</sub>), 28.3 (11-CH<sub>2</sub>), 29.7 (7-CH<sub>2</sub>), 31.9 (6-CH<sub>2</sub>), 34.3 (4-CH), 35.1 (15-CH<sub>2</sub>), 35.9 (16-CH<sub>2</sub>), 39.5 (18-C), 39.8

(12-CH<sub>2</sub>), 43.3 (5-CH<sub>2</sub>), 71.1 (21-CH<sub>2</sub>), 77.9 (19-CH), 126.2 (2-CH), 152.5 (3-CH), 171.8 (14-CO), 173.5 (18-CO), 190.9 (1-CO).

**ESMS:** *m/z*: 453 [M + Na]H<sup>+</sup>, 431 [M]H<sup>+</sup>, 413 [M - H<sub>2</sub>O]H<sup>+</sup>

***E*-Dec-2-enoyl-pantetheine 94**<sup>74,190</sup>



The obtained product was a colourless oil (0.16 g, 0.37 mmol, 89 %)

*R*<sub>f</sub>: 0.48 (DCM / MeOH).

**<sup>1</sup>H-NMR** (CDCl<sub>3</sub>, 400 MHz): δ 0.91 (t, *J* = 7.1, 3H, 10-CH<sub>3</sub>); 0.95 (s, 3H, 22-CH<sub>3</sub>); 1.06 (s, 3H, 23-CH<sub>3</sub>); 1.28-1.33 (m, 8H, 6, 7, 8, 9-CH<sub>2</sub>); 1.45-1.53 (m, 2H, 5-CH<sub>2</sub>); 2.23 (dq, *J* = 1.6, *J* = 7.2, 2H, 4-CH<sub>2</sub>); 2.43 (t, *J* = 7.2, 2H, 15-CH<sub>2</sub>); 3.05-3.20 (m, 2H, 11-CH<sub>2</sub>); 3.37-3.62 (m, 6H, 12-16-21-CH<sub>2</sub>); 4.01 (s, 1H, 19-CH); 6.13 (m, 1H, 2-CH); 6.17 (dt, *J* = 1.5, *J* = 15.4, 1H, 13-NH); 6.96 (dt, *J* = 7, 0, *J* = 15.5, 1H, 3-CH); 7.34 (bt, *J* = 6.0, 1H, 17-NH).

**<sup>13</sup>C-NMR** (CDCl<sub>3</sub>, 100 MHz): δ 14.1 (10-CH<sub>3</sub>), 20.4 (22-CH<sub>3</sub>); 21.7 (23-CH<sub>3</sub>); 22.6 (9-CH<sub>2</sub>); 27.9 (5-CH<sub>2</sub>); 28.6 (11-CH<sub>2</sub>); 29.1 (7-CH<sub>2</sub>); 29.2 (6-CH<sub>2</sub>); 31.7 (8-CH<sub>2</sub>); 32.3 (4-CH<sub>2</sub>); 35.1 (15-CH<sub>2</sub>); 35.6 (16-CH<sub>2</sub>); 39.3 (20-C); 39.9 (12-CH<sub>2</sub>); 70.9 (21-CH<sub>2</sub>); 77.8 (19-CH); 128.1 (2-CH); 147.7 (3-CH); 171.6 (14-CO); 173.3 (18-CO); 190.9 (1-COS).

**ESMS:** *m/z*: 453 [M + Na]H<sup>+</sup>, 431 [M]H<sup>+</sup>



## 8.4 SQTKS ER Domain

### Transformation of pET28-SQTKS-ER into *E. coli* Top10

50 µl chemically competent *E. coli* Top10 cells were thawed on ice and incubated on ice for 30 min with 1 µl of the appropriate vector. Subsequently heat shock transformation was performed, therefore, the cells were incubated at 42 °C for 30 s, then immediately chilled on ice for 2 min. 250 µl SOC medium were added. The transformed cells were shaken (350 rpm) for 1 h at 37 °C. Positive clones were selected by plating 50 - 150 µl of the cells on solid LB medium containing the appropriate antibiotic. The cells were grown over night at 37 °C

### Transformation of pET28-SQTKS-ER into *E. coli* BL21

50 µl chemically competent *E. coli* BL 21 cells were thawed on ice and incubated on ice for 30 min with 1 µl of the SQTKS-ER-domain. subsequently heat shock transformation was performed, therefore, the cells were incubated at 42 °C for 10 s, then immediately chilled on ice for 2 min. 800 µl SOC medium were added. The transformed cells were shaken (350 rpm) for 1 h at 37 °C. Positive clones were selected by plating 50 - 150 µl of the cells on solid LB medium containing the appropriate antibiotic. The cells were grown over night at 37 °C.

### Protein expression and purification

A starter culture was prepared by scraping the surface of a glycerol stock of *E.coli* BL21 transformed with pET28a-ER. The cells were grown in LB media with kanamycin (50 mg/mL) and incubated at 37 °C, 180 rpm, overnight. 0.5 ml of the starter culture was added to 100 mL 2TY media. The cells were incubated to an OD<sub>600</sub> of 0.6, then the flask was cooled down to 16 °C and induced with 50 µl 1M IPTG solution. This solution was incubated over night at 16 °C, 180 rpm.

To isolate the protein, the media was centrifuged at 7000 rpm for 30 min and the pellet was collected. The cells could be used immediately or frozen at -20 °C. The cells were suspended in 50 ml of nickel column wash buffer (50 mM Tris pH 8, 150 mM NaCl, 10 % glycerol (v/v) and 20 mM imidazole) and sonicated on ice for 6.5 minutes. Every 30 second the sonicator was switched on and of. The rest of the solution was centrifuged at 8500 rpm for 30 min, filtered and purified by a nickel column\*. For this a linear gradient of elution buffer (50 mM Tris pH 8, 150 mM NaCl, 10% glycerol (v/v) and 0.5 M imidazole) was used. The fractions were checked by SDS gel and the fractions with

the correct mass were combined and concentrated. A second purification was done by a size exclusion column chromatography\*\*. The protein solution was loaded onto the column and eluted with size exclusion elution buffer (Buffer: 50 mM Tris pH 8, 150 mM NaCl, 20% glycerol (v/v)). The fractions were analyzed again by SDS gel and the protein with the correct mass and high purity was combined and concentrated. The concentration of protein was estimated by a calculated absorption coefficient  $\epsilon = 0.69$  at 280 nm. After that the protein was divided into several aliquots and stored at - 4 °C.

\*Nickel column Protino Ni-NTA Columns 5 mL

\*\*Size exclusion column- HiLoad 26/600 Superdex 200pg (GE Healthcare), 320 ml).

### **Protein expression with the Bioreactor**

A starter culture was prepared by scraping the surface of a glycerol stock of *E.coli* BL21 transformed with pET28-ER. The cells were grown in LB media with kanamycin (50 mg / mL) and incubated at 37 °C, 200 rpm, overnight. The reactor was filled with 8 L 2TY Medium and autoclaved. The pH was adjusted to pH 7 with 1 M HCl and 25 % ammonia solution and kept at this pH. 4 ml of the starter culture was added to 8 L 2TY media. The air pressure was adjusted to 10 L/min and the rotational speed to 400 rpm. The cells were incubated to an OD<sub>600</sub> of 0.8, then the reactor was cooled down to 20 °C and induced with 6 ml 1M IPTG solution. This solution was incubated over night at 20°C, 400 rpm. For the purification of the protein the same procedure like for the non-bioreactor protein expression was used.

### **8.5 Colony Polymerase-chain-reaction**

The successful transformation of BL21 cells was configured by Colony-PCR. For a Colony-PCR the template was added by transferring a small part of a single *E.coli* colony into the reaction mixture using a sterile toothpick. For each primer pair the specific annealing temperature was determined in a PCR temperature gradient. A typical Colony-PCR contained:

0.2  $\mu$ l Forward Primer  
0.2  $\mu$ l Reverse Primer  
- Picked DNA  
2.5  $\mu$ l One Taq Master Mix  
7.1  $\mu$ l Water

---

Total: 10  $\mu$ l

### **8.6 Bradford assay**

Standard solutions of bovine serum albumin (0.1-2 ml/ml) in size exclusion buffer [50 mM Tris pH 8, 150 mM NaCl, 20 % glycerol (v/v)] were prepared by serial dilution. 100  $\mu$ l of the standards were mixed with Bradford dye reagent (1 ml) and incubated for 15 min at RT. The absorption of each sample was measured at 595 nm against a standard (size exclusion buffer 100  $\mu$ l, Bradford dye reagent, 1 ml) to construct a standard concentration curve. A sample of the protein to be quantified (20  $\mu$ l) was diluted in size exclusion buffer (80  $\mu$ l) and treated with Bradford dye reagent (1ml). This was incubated at room temperature for 15 mins and then the absorption was measured at 595 nm. This was compared to the previously prepared concentration curve to calculate the amount of protein that had been produced.

### **8.7 SQTKS-ER Domain LCMS Assay**

A sample of reaction mixture (20  $\mu$ l) was mixed with acetonitrile (80  $\mu$ l) to precipitate the protein and then centrifuged (13,000 rpm, 1 min). The supernatant was analysed directly by LCMS using a standard LCMS profile.

### **8.8 SQTKS-ER-Domain Enzyme Assay**

The different pantetheine assays are performed in a total volume of 400  $\mu$ l. The concentration of NADPH (10  $\mu$ L of 10 mM stock) and SQTKS-ER-domain (20  $\mu$ L of def. 0.02 U) and the temperature at 25 °C were kept constant in all assays. The amount of buffer\* (290-365  $\mu$ L) and substrate (5-80  $\mu$ L of 5 mM Stock) varied between the assays. The assays are performed in cuvettes under measuring the UV absorption at 340 nm.

\* 50 mM Tris pH 8.0, 150 mM NaCl, 20 % Glycerol.

## 8.9 SQTGS-ER-Domain Stereochemistry Assay

NMR assays were performed with tigloyl panthetine (2.0 mg, 10 mmol), NADPH (8 mg, 10 mmol), buffer (50mM Tris, 150 mM NaCl, 20 % (v/v) glycerol and pH 8.0) and the isolated ER domain (380  $\mu$ U, 100  $\mu$ l of stock) in water (total volume 1 ml). The progress of the reaction was monitored by LCMS. After 24 h further NADPH (8 mg, 10 mmol) was added to the reaction. After a further 24 h the protein was precipitated with  $\text{CH}_2\text{Cl}_2$  (1 ml) and centrifuged (3000 rpm, 20 min).

The organic layer was removed, and the aqueous layer was further extracted with  $\text{CH}_2\text{Cl}_2$  ( $2 \times 1$  ml). The organic fractions were combined, and the solvent was removed under a flow of  $\text{N}_2$  and then dried *in vacuo* for a further 2 h. To the dried sample was added water (450  $\mu$ l) and aqueous NaOH (2 M, 50  $\mu$ l). The sample was left for 90 min at RT and then acidified with aqueous HCl (2 M, 100  $\mu$ l) to pH 3 and extracted with  $\text{CDCl}_3$  ( $4 \times 200$   $\mu$ l). The organics were combined, dried with  $\text{MgSO}_4$  and then filtered through  $\text{MgSO}_4$  directly into an NMR tube. (1*R*, 2*R*)-(+)-1, 2 diphenyl ethylenediamine was titrated (100 mM stock) to optimize resolution of the obtained spectra.

## 8.10 Mutagenesis

### Plasmid isolation

For the isolation of the plasmid first an overnight culture of *E-coli* Top10 cells with the desired plasmid were cultivated. The next day the cell culture was centrifuged in a 2 ml Eppendorf-tube at 13.000 g for 2 min and 4 °C. Afterwards the supernatant was discarded and the previous step was repeated with the remaining overnight culture, if necessary. The pellet was resuspended in 100  $\mu$ l Solution (5 mM Glucose, 25 mM Tris-HCl (pH 8.0) and 10 mM EDTA) and 5  $\mu$ l RNase. The sample was incubated for 5 min by RT. 200  $\mu$ l of Solution II (0.2 M NaOH and 1 % SDS) was added It was mixed carefully by inversion. Incubated 5 min by RT. 150  $\mu$ l Solution III was added (60 ml 5M Potassium acetate, 11.5 ml conc. acetic acid and 20.5 ml  $\text{H}_2\text{O}$ ) and mixed carefully by inversion. The mixture was centrifuged by 11 000 g for 5 min by 4°C. The supernatant was transferred to a new 1.5 ml Eppendorf-tube. Then 800  $\mu$ l EtOH (96 %, bio quality) and 45  $\mu$ l 3 M Na-Acetate were added, inverted and to precipitate the plasmid DNA 30 min by -20 °C incubated. Afterwards by 11 000 g for 5 min by 4 °C centrifuged. Then the supernatant was discarded and the pellet was washed with 70 % EtOH (bio quality) (100  $\mu$ l). Afterwards the suspension was centrifuged at max speed for 5 min at RT. Then the supernatant discarded

and pellet was dried by 37 °C for 15 min until it was clear. The pellet was resuspended in 20 µl Water and could be stored at -20 °C for further use.

### Site directed mutagenesis<sup>162</sup>

For the site directed mutagenesis, different Oligo-primers were ordered from Sigma Aldrich and used in different combinations. Hereby non-overlapping primer pairs were tested.

As a template the vector pET28a containing the ER sequence of SQTGS was chosen. Two different PCR approaches with different polymerases were tested. The first approach contained the Q5 High-Fidelity DNA polymerase the other the Phusion High-Fidelity DNA polymerase.

Q5 High-Fidelity DNA polymerase		Phusion	High-Fidelity DNA
12.5 µl	Q5 Master Mix	5 µl	Phusion Buffer (5x)
1.25 µl	Forward Primer	1.25 µl	Forward Primer
1.25 µl	Reverse Primer	1.25 µl	Reverse Primer
1 µl	Plasmid DNA	1 µl	Plasmid DNA
(6.5-9 µl)	Water	(9-11.25 µl)	Water
(0-2.5 µl)	MgCl <sub>2</sub> (50 nM)	(0-2.5 µl)	MgCl <sub>2</sub> (50 nM)
		5 µl	dNTP Mix (1.25 mM)
		0.25 µl	Polymerase
<hr/>		<hr/>	
Total: 25 µl		Total: 25 µl	

Different annealing temperatures were tested to find the best condition for the different primer combination. For efficient DNA amplification, 22 PCR cycles with an elongation time of 6.30 min were applied.

The products of the PCR were checked on an Agarose-Gel. Afterwards the blunt ends of the primers were ligated through the use of the Quick Ligation Kit (New England BioLabs). Hereby to 10 µl of the PCR product were 10 µl of the 2 x Quick Ligation Buffer added and mixed. 1 µl Quick T4 DNA Ligase were added and mixed. The mixture was incubated by RT for 15 min. Then 0.5 µl of DpnI were added to the sample and incubated for 2 h. The enzyme cuts methylated DNA, therefore the parental DNA string will be cut.

Afterwards 1  $\mu$ l of the mixture were transformed into TOP10 Cells (see 8.4). The transformation were checked via Colony-PCR.

### **Sequencing**

Sequencing was done by Eurofins Genomics. Plasmids were sent to sequencing at a concentration of 100-200 ng /  $\mu$ l.

Sequence analysis was done in Geneious (Version 7.1.9) through alignment with the original DNA sequence.

### **8.11 Statistical evaluations**

First, the paired t-test was performed to test the influence of the mutation on the specificity constant for the different substrates compared to non-mutated ER. Whereas a significant change in the specificity constant was observed after the mutation this is marked by the asterisk (\*  $p < 0.05$ ). The test for normality was performed by Shapiro-Wilk test. Groups with ( $n = 3$ ) were examined.

Second, the One Way ANOVA was performed test if there is a significant difference between the specificity constants between the mutated ER domains. Whereas a significant change in the specificity constant was observed after the mutation this is marked by the asterisk (\*  $p < 0.05$ ). The test for normality was performed by Shapiro-Wilk test. If the Test-Failed, a ANOVA on Ranks were performed. Hence the Kruskal-Wallis One Way Analysis of Variance on Ranks performed. Afterwards the was the multiple comparison procedure with the Student-Newman-Keuls method performed. Groups with ( $n = 3$ ) were examined.

### **8.12 Bioinformatics**

#### **Sequence alignment**

The sequence of the ER, C-MeT domains SQTKS and KR and C-MeT domains from TENS were available in our working group. A BLAST search was done with the respective sequence in the NCBI database. From the possible hits, only iterative PKS from fungi were considered. The final sequence alignment was done in Geneious (Version 7.1.9) through alignment with the original sequence.

## **Investigation of the pocket volume**

For the investigation of the pocket volumes of the different mutated domains from the ER and the ER domain an open software program 3V web server.<sup>156</sup> Therefore were the PDB Files transmitted to the Software and as search parameters an outer probe radius of 10 and an inner probe radius of 2 were determined.

## **Protein modeling**

Before homology modelling for the specific domain could be done the domain boundaries of the specific domain of SQTks or TENS were determined. Therefore, the sequence of SQTks or TENS was examined with BLAST (basic local alignment search tool).<sup>138</sup> Subsequently, a conserved domain search (CD-Search) was performed to determine the domain boundaries.<sup>139–142</sup> Afterwards, homology modeling was done using Swiss-Model.<sup>83,92,95,105</sup> Therefore was the protein sequence submitted to SwissModel.

Models are computed by the SWISS-MODEL server homology modelling pipeline which relies on ProMod3, an in-house comparative modelling engine based on OpenStructure<sup>83,92,95,105</sup>. ProMod3 extracts initial structural information from the template structure. Insertions and deletions, as defined by the sequence alignment, are resolved by first searching for viable candidates in a structural database. Final candidates are then selected using statistical potentials of mean force scoring methods. If no candidates can be found, a conformational space search is performed using Monte Carlo techniques. Non-conserved side chains are modelled using an in-house backbone-dependent rotamer library. The optimal configuration of rotamers is estimated using the graph based TreePack algorithm by minimising the SCWRL4 energy function. As a final step, small structural distortions, unfavourable interactions, or clashes introduced during the modelling process are resolved by energy minimisation. ProMod3 uses the OpenMM library to perform the computations and the CHARMM27 force field for parameterisation.<sup>83,92,95,105</sup>

Afterwards was the template selected and the modelling performed. Subsequently the model was validated.

Templates:

ER domain SQTks (Chapter 2): CurF (PDB: 5dp2)<sup>144</sup>  
CMeT domain TENS (Chapter 5): CurJ (PDB: 5thy)<sup>171</sup>  
KR domain TENS (Chapter 5): AmphB (PDB: 3slk)<sup>180</sup>

### **Model validation via Swiss-Model**

For the first validation of the generated homolog models Swiss-Model calculates to parameters, which were reconsidered.

**GMQE** (Global Model Quality Estimation) is a quality estimation, which combines properties from the target–template alignment and the template search method. The resulting GMQE score is expressed as a number between 0 and 1, reflecting the expected accuracy of a model built with that alignment and template and the coverage of the target. Higher numbers indicate higher reliability.

**QMEAN** is a composite estimator based on different geometrical properties and provides both global (i.e. for the entire structure) and local (i.e. per residue) absolute quality estimates based on one single model. The QMEAN Z-score provides an estimate of the ‘degree of nativeness’ of the structural features observed in the model on a global scale. It indicates whether the QMEAN score of the model is comparable to what one would expect from experimental structures of similar size. QMEAN Z-scores around zero indicate good agreement between the model structure and experimental structures of similar size. Scores of -4.0 or below are an indication of models with low quality.

### **Integration of the Cofactor**

The respective cofactor was integrated in the specific domain through the alignment of the homolog model with the template in PyMOL (DeLano Scientific LLC, Version 1.8.2.0).<sup>146,147</sup>

The cofactor was extracted from the template and then manually integrated into the structural model of the respective domain in PyMOL. Afterwards, the generated domain plus cofactor was minimized in YASARA, to refine the protein-cofactor interaction.<sup>135</sup> Then the extracted and refined cofactor was alignment with the cofactor from the template for validation.

### **Molecular Docking**

The docking was performed by roughly manually overlaying the specific substrate in active site of the protein using PyMOL. Thereby, the easiest method for the identification of the active site was by displaying the volume of the pockets of the protein. Further, if the cofactor was known, this was used as an orientation point. This was done to simplify the docking in the next step.

This method minimized the so-called Grid Box, which has critical role in the speed of the docking calculations. Molecular docking was done using Autodock Vina Vina (PyRx



0.8).<sup>110,147,148</sup> For performing the calculations with AutoDock Vina Python 3.6 was used as the underlying programming language

The protocol included the creation of the pdbqt files of the respective substrate and protein. Further, the degrees of freedom for the substrate were set. The degree of freedom as a default always were the maximum chosen. The coordinates of the Grid Box were customized for the specific protein-substrate complex. However, the size of the Grid Box that was used by default was 20 x 20 x 20 number of points. Afterwards, this information was saved in the corresponding configuration file. For the subsequent calculations, the AutoDock Vina file was used. Maximal 2 runs were performed. In addition, the used algorithm by Auto Dock Vina was the Broyden-Fletcher-Goldfarb-Shanno algorithm.

Docking results including the RMSD, the lowest binding energy and mean binding energy were obtained from the docking log (dlg) file. The respective, obtained docked substrate were then compared in the protein in PyMOL. The best docked substrates were chosen and separate saved in a new pdb file. Criteria for the selection was the RMSD value and the overall location in the active pocket. Afterwards, the specific substrates were integrated in the protein in PyMOL. The model was then refined by YASARA. The visualization of the different models, after the refinement step was always done in PyMOL.

### **YASSARA Minimization Server**

This server performs an energy minimization of your protein model with the YASARA force field. Or rather, YASARA, which runs molecular dynamics simulations of models in explicit solvent, using a new partly knowledge-based all atom force field derived from Amber, whose parameters have been optimized to minimize the damage done to protein crystal structures.<sup>135</sup>

## 9 References

1. Editorial, *Nature Chemical Biology*, **2007**, 3, 351.
2. G. Strobel and B. Daisy, *Microbiology and Molecular Biology Reviews*, **2003**, 67, 491.
3. L. Gu, B. Wang, A. Kulkarni, T. W. Geders, R. V. Grindberg, L. Gerwick, K. Håkansson, P. Wipf, J. L. Smith, W. H. Gerwick and D. H. Sherman, *Nature*, **2009**, 459, 731.
4. A. J. Dijkstra, F. D. Gunstone and J. L. Harwood, *The lipid handbook*, **2007**.
5. N. Murakami, W. Wang, M. Aoki, Y. Tsutsui, M. Sugimoto and M. Kobayashi, *Tetrahedron Letters*, **1998**, 39, 2349.
6. J. Beld, D. J. Lee and M. D. Burkart, *Molecular BioSystems*, **2015**, 11, 38.
7. A. W. Alberts, A. W. Strauss, S. Hennessy and P. R. Vagelos, *Proceedings of the National Academy of Sciences of the United States of America*, **1975**, 72, 3956.
8. S. Smith, A. Witkowski and A. K. Joshi, *Progress in Lipid Research*, **2003**, 42, 289.
9. K. Magnuson, S. Jackowski, C. O. Rock and J. E. Cronan, *Microbiological Reviews*, **1993**, 57, 522.
10. D. M. Byers and H. Gong, *Biochemistry and Cell Biology*, **2007**, 85, 649.
11. J. G. Olsen, A. Kadziola, P. von Wettstein-Knowles, M. Siggaard-Andersen and S. Larsen, *Structure*, **2001**, 9, 233.
12. C. Davies, R. J. Heath, S. W. White and C. O. Rock, *Structure*, **2000**, 8, 185.
13. W. Huang, J. Jia, P. Edwards, K. Dehesh, G. Schneider and Y. Lindqvist, *The EMBO journal*, **1998**, 17, 1183.
14. J. M. Crawford, B. C. R. Dancy, E. A. Hill, D. W. Udway and C. A. Townsend, *Proceedings of the National Academy of Sciences of the United States of America*, **2006**, 103, 16728.
15. A. T. Keatinge-Clay, *Natural Product Reports*, **2012**, 29, 1050.
16. M. Leesong, B. S. Henderson, J. R. Gillig, J. M. Schwab and J. L. Smith, *Structure*, **1996**, 4, 253.
17. D. Kostrewa, F. K. Winkler, G. Folkers, L. Scapozza and R. Perozzo, *Protein Science*, **2005**, 14, 1570.
18. D. H. Kwan and F. Schulz, *Molecules*, **2011**, 16, 6092.
19. D. H. Kwan, Y. Sun, F. Schulz, H. Hong, B. Popovic, J. C. C. Sim-Stark, S. F. Haydock and P. F. Leadlay, *Chemistry & Biology*, **2008**, 15, 1231.
20. K. Finzel, D. J. Lee and M. D. Burkart, *ChemBioChem*, **2015**, 16, 528.
21. D. L. Ollis, E. Cheah, M. Cygler, B. Dijkstra, F. Frolow, S. M. Franken, M. Harel, S. J. Remington, I. Silman and J. Schrag, *Protein Engineering*, **1992**, 5, 197.
22. Z. Zhuang, F. Song, H. Zhao, L. Li, J. Cao, E. Eisenstein, O. Herzberg and D. Dunaway-Mariano, *Biochemistry*, **2008**, 47, 2789.
23. T. Maier, M. Leibundgut and N. Ban, *Science*, **2008**, 321, 1315.

24. D. A. Herbst, C. A. Townsend and T. Maier, *Natural Product Reports*, **2018**, *35*, 1046.
25. T. Maier, S. Jenni and N. Ban, *Science*, **2006**, *311*, 1258.
26. M. Leibundgut, S. Jenni, C. Frick and N. Ban, *Science*, **2007**, *316*, 288.
27. M. Leesong, B.S. Henderson, J.F Gillig, J.M. Schwab, J.L. Smith, *Structure*, **1996**, *4*, 253-64.
28. A. C. Price, Y.-M. Zhang, C. O. Rock and S. M. White, *Structure*, **2004**, *12*, 417-428.
29. C. Baldock, J. B. Rafferty and D. W. Rice, *J. Mol. Biol.*, **1998**, *284*, 1529-46.
30. G. Pappenberger, T. Schulz-Gasch, J. Bailly and M. Hennig, *Acta Crystallorg.*, **2007**, *63*, 1208.
31. K. J. Weissman, *Nature Chemical Biology*, **2015**, *11*, 660.
32. A. Hagen, S. Poust, T. d. Rond, J. L. Fortman, L. Katz, C. J. Petzold and J. D. Keasling, *ACS Synthetic Biology*, **2016**, *5*, 21.
33. T. Dingermann, R. Hänsel and I. Zündorf, *Pharmazeutische Biologie: Molekulare Grundlagen und klinische Anwendung*, Springer, **2002**.
34. S.-C. Tsai, H. Lu, D. E. Cane, C. Khosla and R. M. Stroud, *Biochemistry*, **2002**, *41*, 12598.
35. C. Hertweck, *Angewandte Chemie*, **2009**, *48*, 4688.
36. T. P. Nicholson, B. A. Rudd, M. Dawson, C. M. Lazarus, T. J. Simpson and R. J. Cox, *Chemistry & Biology*, **2001**, *8*, 157.
37. M. E. Hillenmeyer, G. A. Vandova, E. E. Berlew and L. K. Charkoudian, *Proceedings of the National Academy of Sciences of the United States of America*, **2015**, *112*, 13952.
38. J. Schümann and C. Hertweck, *Journal of Biotechnology*, **2006**, *124*, 690.
39. R. J. Cox, *Organic & Biomolecular Chemistry*, **2007**, *5*, 2010.
40. J. Staunton and K. J. Weissman, *Nat. Prod. Rep.*, **2001**, *18*, 380.
41. Z. Chang, N. Sitachitta, J. V. Rossi, M. A. Roberts, P. M. Flatt, J. Jia, D. H. Sherman and W. H. Gerwick, *Journal of Natural Products*, **2004**, *67*, 1356.
42. L. Gu, T. W. Geders, B. Wang, W. H. Gerwick, K. Håkansson, J. L. Smith and D. H. Sherman, *Science*, **2007**, *318*, 970.
43. L. Gu, E. B. Eisman, S. Dutta, T. M. Franzmann, S. Walter, W. H. Gerwick, G. Skiniotis and D. H. Sherman, *Angewandte Chemie*, **2011**, *50*, 2795.
44. J. M. Crawford, A. L. Vagstad, K. P. Whitworth, C. A. Townsend and K. C. Ehrlich, *ChemBioChem*, **2008**, *9*, 1019.
45. I. Soehano, L. Yang, F. Ding, H. Sun, Z. J. Low, X. Liu and Z.-X. Liang, *Organic & Biomolecular Chemistry*, **2014**, *12*, 8542.
46. J. Beck, S. Ripka, A. Siegner, E. Schiltz and E. Schweizer, *European Journal of Biochemistry*, **1990**, *192*, 487.
47. T. Shimizu, H. Kinoshita, S. Ishihara, K. Sakai, S. Nagai and T. Nihira, *Applied and Environmental Microbiology*, **2005**, *71*, 3453.

48. K. L. Eley, L. M. Halo, Z. Song, H. Powles, R. J. Cox, A. M. Bailey, C. M. Lazarus and T. J. Simpson, *ChemBioChem*, **2007**, *8*, 289.
49. T. Ugai, A. Minami, R. Fujii, M. Tanaka, H. Oguri, K. Gomi and H. Oikawa, *Chemical Communications*, **2015**, *51*, 1878.
50. J. L. Smith, G. Skiniotis and D. H. Sherman, *Current Opinion in Structural Biology*, **2015**, *31*, 9.
51. S. S. Chirala, A. Jayakumar, Z. W. Gu and S. J. Wakil, *Proceedings of the National Academy of Sciences of the United States of America*, **2001**, *98*, 3104.
52. S. Smith, *FASEB*, **1994**, *8*, 1248.
53. J. S. Parascandolo, J. Havemann, H. K. Potter, F. Huang, E. Riva, J. Connolly, I. Wilkening, L. Song, P. F. Leadlay and M. Tosin, *Angewandte Chemie*, **2016**, *128*, 3524.
54. S. M. Ma and Y. Tang, *The FEBS Journal*, **2007**, *274*, 2854.
55. M. Sato, F. Yagishita, T. Mino, N. Uchiyama, A. Patel, Y.-H. Chooi, Y. Goda, W. Xu, H. Noguchi, T. Yamamoto, K. Hotta, K. N. Houk, Y. Tang and K. Watanabe, *ChemBioChem*, **2015**, *16*, 2294.
56. H. Yao, *In Vitro Investigation of Multi-Domain Fragments of Squalestatin Tetraketide Synthase*, PhD-Thesis, Hannover, **2018**.
57. E. Liddle, A. Scott, L.-C. Han, D. Ivison, T. J. Simpson, C. L. Willis and R. J. Cox, *Chemical Communications*, **2017**, *53*, 1727.
58. D. M. Roberts, C. Bartel, A. Scott, D. Ivison, T. J. Simpson and R. J. Cox, *Chem. Sci.*, **2017**, *8*, 1116.
59. J. Zheng and A. T. Keatinge-Clay, *Med. Chem. Commun.*, **2013**, *4*, 34.
60. R. Reid, M. Piagentini, E. Rodriguez, G. Ashley, N. Viswanathan, J. Carney, D. V. Santi, C. R. Hutchinson and R. McDaniel, *Biochemistry*, **2003**, *42*, 72.
61. A. T. Keatinge-Clay, *Chemistry & Biology*, **2007**, *14*, 898.
62. A. T. Keatinge-Clay, *Nat. Prod. Rep.*, **2016**, *33*, 141.
63. B. Bonsch, V. Belt, C. Bartel, N. Duensing, M. Koziol, C. M. Lazarus, A. M. Bailey, T. J. Simpson and R. J. Cox, *Chemical Communications*, **2016**, *52*, 6777.
64. R. J. Cox, F. Glod, D. Hurley, C. M. Lazarus, T. P. Nicholson, B. A. M. Rudd, T. J. Simpson, B. Wilkinson and Y. Zhang, *Chemical Communications*, **2004**, *20*, 2260.
65. M. N. Heneghan, A. A. Yakasai, K. Williams, K. A. Kadir, Z. Wasil, W. Bakeer, K. M. Fisch, A. M. Bailey, T. J. Simpson, R. J. Cox and C. M. Lazarus, *Chem. Sci.*, **2011**, *2*, 972.
66. J. Kennedy, *Natural Product Reports*, **2008**, *25*, 25.
67. Y. Shimizu, H. Ogata and S. Goto, *ChemBioChem*, **2017**, *18*, 1048.
68. K. J. Weissman, *Natural Product Reports*, **2016**, *33*, 203.
69. W. Zhang and J. Liu, *F1000Research*, **2016**, *5*.
70. J. F. Barajas, J. M. Blake-Hedges, C. B. Bailey, S. Curran and J. D. Keasling, *Synthetic and Systems Biotechnology*, **2017**, *2*, 147.

71. J.-H. Yong and W.-H. Byeon, *Journal of microbiology*, **2005**, *43*, 277.
72. C. J. Dutton, S. P. Gibson, A. C. Goudie, K. S. Holdom, M. S. Pacey, J. C. Ruddock, J. D. Bu'Lock and M. K. Richards, *The Journal of Antibiotics*, **1991**, *44*, 357.
73. H. Mehlhorn, H. L. Jones, A. J. Weatherley and B. Schumacher, *Parasitol Res*, **1993**, *79*, 603.
74. E. J. Skellam, D. Hurley, J. Davison, C. M. Lazarus, T. J. Simpson and R. J. Cox, *Molecular BioSystems*, **2010**, *6*, 680.
75. L. M. Halo, J. W. Marshall, A. A. Yakasai, Z. Song, C. P. Butts, M. P. Crump, M. Heneghan, A. M. Bailey, T. J. Simpson, C. M. Lazarus and R. J. Cox, *ChemBioChem*, **2008**, *9*, 585.
76. K. M. Fisch, W. Bakeer, A. A. Yakasai, Z. Song, J. Pedrick, Z. Wasil, A. M. Bailey, C. M. Lazarus, T. J. Simpson and R. J. Cox, *ACS*, **2011**, *133*, 16635.
77. A. A. Yakasai, J. Davison, Z. Wasil, L. M. Halo, C. P. Butts, C. M. Lazarus, A. M. Bailey, T. J. Simpson and R. J. Cox, *ACS*, **2011**, *133*, 10990.
78. B. Busch and C. Hertweck, *Phytochemistry*, **2009**, *70*, 1833.
79. X.-L. Yang, S. Friedrich, S. Yin, O. Piech, K. Williams, T. J. Simpson and R. J. Cox, *Chem. Sci.*, **2019**, *10*, 8478-8489.
80. R. A. Cacho, J. Thuss, W. Xu, R. Sanichar, Z. Gao, A. Nguyen, J. C. Vederas and Y. Tang, *ACS*, **2015**, *137*, 15688.
81. C. M. Kao, M. McPherson, R. N. McDaniel, H. Fu, D. E. Cane and C. Khosla, *ACS*, **1998**, *120*, 2478.
82. M. Klaus and M. Grininger, *Nat. Prod. Rep.*, **2018**, *35*, 1070.
83. T. Schwede, *Nucleic Acids Research*, **2003**, *31*, 3381.
84. S. Kaczanowski and P. Zielenkiewicz, *Theor Chem Acc*, **2010**, *125*, 643.
85. D. Baker and A. Sali, *Science*, **2001**, *294*, 93.
86. A. Tramontano, R. Leplae and V. Morea, *Proteins*, **2001**, *Suppl 5*, 22.
87. M. A. Martí-Renom, A. C. Stuart, A. Fiser, R. Sánchez, F. Melo and A. Sali, *Annual Review of Biophysics and Biomolecular Structure*, **2000**, *29*, 291.
88. T. J. Chuan and S. Ranganathan, eds., *Computer-Aided Baccine Design*, Witney, **2011**.
89. O. Sensoy, J. G. Almeida, J. Shabbir, I. S. Moreira and G. Morra, *Methods in Cell Biology*, **2017**, *142*, 205.
90. L. Bordoli, F. Kiefer, K. Arnold, P. Benkert, J. Battey and T. Schwede, *Nature Protocols*, **2009**, *4*, 1.
91. A. Waterhouse, M. Bertoni, S. Bienert, G. Studer, G. Tauriello, R. Gumienny, F. T. Heer, T. A. P. de Beer, C. Rempfer, L. Bordoli, R. Lepore and T. Schwede, *Nucleic Acids Research*, **2018**, *46*, W296-W303.
92. S. Bienert, A. Waterhouse, T. A. P. de Beer, G. Tauriello, G. Studer, L. Bordoli and T. Schwede, *Nucleic Acids Research*, **2017**, *45*, D313-D319.
93. N. Guex, M. C. Peitsch and T. Schwede, *Electrophoresis*, **2009**, *30 Suppl 1*, S162-73.

94. P. Benkert, M. Biasini and T. Schwede, *Bioinformatics*, **2011**, *27*, 343.
95. M. Bertoni, F. Kiefer, M. Biasini, L. Bordoli and T. Schwede, *Scientific Reports*, **2017**, *7*, 10480.
96. A. Sali and T. L. Blundell, *Journal of Molecular Biology*, **1993**, *234*, 779.
97. A. Fiser, R. K. Do and A. Sali, *Protein Science*, **2000**, *9*, 1753.
98. C. A. Rohl, C. E. M. Strauss, D. Chivian and D. Baker, *Proteins*, **2004**, *55*, 656.
99. C. S. Soto, M. Fasnacht, J. Zhu, L. Forrest and B. Honig, *Proteins*, **2008**, *70*, 834.
100. B. Rost, *Protein Engineering*, **1999**, *12*, 85.
101. K. Ginalski, *Current Opinion in Structural Biology*, **2006**, *16*, 172.
102. S. Ovchinnikov, H. Park, D. E. Kim, F. DiMaio and D. Baker, *Proteins*, **2018**, *86 Suppl 1*, 113.
103. A. Fiser and A. Šali, *Macromolecular Crystallography*, Elsevier, **2003**, p 461.
104. F. Kiefer, K. Arnold, M. Künzli, L. Bordoli and T. Schwede, *Nucleic Acids Research*, **2009**, *37*, D387-92.
105. M. Biasini, S. Bienert, A. Waterhouse, K. Arnold, G. Studer, T. Schmidt, F. Kiefer, T. Gallo Cassarino, M. Bertoni, L. Bordoli and T. Schwede, *Nucleic Acids Research*, **2014**, *42*, 252-8.
106. M. Remmert, A. Biegert, A. Hauser and J. Söding, *Nature Methods*, **2011**, *9*, 173.
107. P. Benkert, S. C. E. Tosatto and D. Schomburg, *Proteins*, **2008**, *71*, 261.
108. J. Biesiada, A. Porollo, P. Velayutham, M. Kouril and J. Meller, *Hum Genomics*, **2011**, *5*, 497.
109. D. S. Goodsell, G. M. Morris and A. J. Olson, *J. Mol. Recognit.*, **1996**, *9*, 1.
110. O. Trott and A. J. Olson, *Journal of Computational Chemistry*, **2010**, *31*, 455.
111. P. T. Lang, S. R. Brozell, S. Mukherjee, E. F. Pettersen, E. C. Meng, V. Thomas, R. C. Rizzo, D. A. Case, T. L. James and I. D. Kuntz, *RNA*, **2009**, *15*, 1219.
112. B. K. Shoichet, I. D. Kuntz and D. L. Bodian, *J. Comput. Chem.*, **1992**, *13*, 380.
113. H. Claussen, C. Buning, M. Rarey and T. Lengauer, *JMB*, **2001**, *308*, 377.
114. R. A. Friesner, J. L. Banks, R. B. Murphy, T. A. Halgren, J. J. Klicic, D. T. Mainz, M. P. Repasky, E. H. Knoll, M. Shelley, J. K. Perry, D. E. Shaw, P. Francis and P. S. Shenkin, *Journal of Medicinal Chemistry*, **2004**, *47*, 1739.
115. M. L. Verdonk, J. C. Cole, M. J. Hartshorn, C. W. Murray and R. D. Taylor, *Proteins*, **2003**, *52*, 609.
116. I. W. Davis and D. Baker, *JMB*, **2009**, *385*, 381.
117. M. I. Zavodszky, A. Rohatgi, J. R. van Voorst, H. Yan and L. A. Kuhn, *JMR*, **2009**, *22*, 280.
118. A. N. Jain, *Journal of Medicinal Chemistry*, **2003**, *46*, 499.
119. S. F. Sousa, P. A. Fernandes and M. J. Ramos, *Proteins*, **2006**, *65*, 15.
120. I. D. Kuntz, J. M. Blaney, S. J. Oatley, R. Langridge and T. E. Ferrin, *JMB*, **1982**, *161*, 269.

121. I. Muegge and M. Rarey in *Reviews in Computational Chemistry*, Volume 17, **2001**, p 1.
122. G. M. Morris and M. Lim-Wilby, *Methods in Molecular Biology*, **2008**, 443, 365.
123. G. M. Morris, D. S. Goodsell, R. S. Halliday, R. Huey, W. E. Hart, R. K. Belew and A. J. Olson, *J. Comput. Chem.*, **1998**, 19, 1639.
124. D. B. Kitchen, H. Decornez, J. R. Furr and J. Bajorath, *Nature Reviews*, **2004**, 3, 935.
125. B. Kramer, M. Rarey and T. Lengauer, *Proteins*, **1999**, 37, 228.
126. P. K. Weiner and P. A. Kollman, *J. Comput. Chem.*, **1981**, 2, 287.
127. S. J. Weiner, P. A. Kollman, D. A. Case, U. C. Singh, C. Ghio, G. Alagona, S. Profeta and P. Weiner, *ACS*, **1984**, 106, 765.
128. T. Gaillard, *Journal of Chemical Information and Modeling*, **2018**, 58, 1697.
129. R. Huey, G. M. Morris, A. J. Olson and D. S. Goodsell, *J. Comput. Chem.*, **2007**, 28, 1145.
130. W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell and P. A. Kollman, *ACS*, **1995**, 117, 5179.
131. J. R. Schames, R. H. Henchman, J. S. Siegel, C. A. Sotriffer, H. Ni and J. A. McCammon, *Journal of Medicinal Chemistry*, **2004**, 47, 1879.
132. R. E. Amaro, A. Schnaufer, H. Interthal, W. Hol, K. D. Stuart and J. A. McCammon, *Proceedings of the National Academy of Sciences of the United States of America*, **2008**, 105, 17278.
133. Y.-C. Chen, *Trends in Pharmacological Sciences*, **2015**, 36, 78.
134. X.-Y. Meng, H.-X. Zhang, M. Mezei and M. Cui, *Current Computer-Aided Drug Design*, **2011**, 7, 146.
135. E. Krieger, K. Joo, J. Lee, J. Lee, S. Raman, J. Thompson, M. Tyka, D. Baker and K. Karplus, *Proteins*, **2009**, 77, 114.
136. E. Krieger, G. Koraimann and G. Vriend, *Proteins*, **2002**, 47, 393.
137. C. Bartel, *Enzymology of Isolated functional Domains from Iterative Fungal Polyketide Synthases*, PhD-Thesis, Hannover, **2017**.
138. S. F. Altschul, T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller and D. J. Lipman, *Nucleic Acids Research*, **1997**, 25, 3389.
139. A. Marchler-Bauer and S. H. Bryant, *Nucleic Acids Research*, **2004**, 32, W327-31.
140. A. Marchler-Bauer, S. Lu, J. B. Anderson, F. Chitsaz, M. K. Derbyshire, C. DeWeese-Scott, J. H. Fong, L. Y. Geer, R. C. Geer, N. R. Gonzales, M. Gwadz, D. I. Hurwitz, J. D. Jackson, Z. Ke, C. J. Lanczycki, F. Lu, G. H. Marchler, M. Mullokandov, M. V. Omelchenko, C. L. Robertson, J. S. Song, N. Thanki, R. A. Yamashita, D. Zhang, N. Zhang, C. Zheng and S. H. Bryant, *Nucleic Acids Research*, **2011**, 39, 225-9.
141. A. Marchler-Bauer, M. K. Derbyshire, N. R. Gonzales, S. Lu, F. Chitsaz, L. Y. Geer, R. C. Geer, J. He, M. Gwadz, D. I. Hurwitz, C. J. Lanczycki, F. Lu, G. H. Marchler, J. S. Song, N. Thanki, Z. Wang, R. A. Yamashita, D. Zhang, C. Zheng and S. H. Bryant, *Nucleic Acids Research*, **2015**, 43, 222-6.
142. A. Marchler-Bauer, Y. Bo, L. Han, J. He, C. J. Lanczycki, S. Lu, F. Chitsaz, M. K. Derbyshire, R. C. Geer, N. R. Gonzales, M. Gwadz, D. I. Hurwitz, F. Lu, G. H. Marchler, J. S. Song, N. Thanki, Z. Wang, R. A.

- Yamashita, D. Zhang, C. Zheng, L. Y. Geer and S. H. Bryant, *Nucleic Acids Research*, **2017**, *45*, 200-D203.
143. D. A. Herbst, R. P. Jakob, F. Zähringer and T. Maier, *Nature*, **2016**, *531*, 533 EP -.
144. D. Khare, W. A. Hale, A. Tripathi, L. Gu, D. H. Sherman, W. H. Gerwick, K. Håkansson and J. L. Smith, *Structure*, **2015**, *23*, 2213.
145. W. D. Fiers, G. J. Dodge, D. H. Sherman, J. L. Smith and C. C. Aldrich, *ACS*, **2016**, *138*, 16024.
146. G. Janson, C. Zhang, M. G. Prado and A. Paiardini, *Bioinformatics*, **2017**, *33*, 444.
147. S. Forli, R. Huey, M. E. Pique, M. F. Sanner, D. S. Goodsell and A. J. Olson, *Nature Protocols*, **2016**, *11*, 905.
148. D. Seeliger and B. L. de Groot, *Journal of Computer-Aided Molecular Design*, **2010**, *24*, 417.
149. H. B. Burgi, J. D. Dunitz, J. M. Lehn and G. Wipff, *Tetrahedron*, **1974**, *30*, 1563.
150. H.-B. Bürgi and J. D. Dunitz, *Structure Correlation*, Wiley, **1994**.
151. S. H. Light, G. Minasov, M.-E. Duban and W. F. Anderson, *Acta crystal.*, **2014**, *70*, 544.
152. E. D. Caldas, K. Sadilkova, B. L. Ward, A. D. Jones, C. K. Winter and D. G. Gilchrist, *J. Agric. Food Chem.*, **1998**, *46*, 4734.
153. I. Gaffoor and F. Trail, *Applied and Environmental Microbiology*, **2006**, *72*, 1793.
154. I. Fujii, N. Yoshida, S. Shimomaki, H. Oikawa and Y. Ebizuka, *Chemistry & Biology*, **2005**, *12*, 1301.
155. Y.-M. Chiang, E. Szewczyk, A. D. Davidson, N. Keller, B. R. Oakley and C. C. C. Wang, *ACS*, **2009**, *131*, 2965.
156. N. R. Voss and M. Gerstein, *Nucleic Acids Research*, **2010**, *38*, 555-62.
157. A. C. Murphy, H. Hong, S. Vance, R. W. Broadhurst and P. F. Leadlay, *Chemical Communications*, **2016**, *52*, 8373.
158. F. Wang, Y. Wang, J. Ji, Z. Zhou, J. Yu, H. Zhu, Z. Su, L. Zhang and J. Zheng, *ACS Chem. Biol.*, **2015**, *10*, 1017.
159. D. H. Kwan and P. F. Leadlay, *ACS Chem. Biol.*, **2010**, *5*, 829.
160. A. G. Newman, A. L. Vagstad, P. A. Storm and C. A. Townsend, *ACS*, **2014**, *136*, 7348.
161. D. M. Roberts, *Investigating the Programming of Type I highly Reducing Iterative Polyketide Synthases*, PhD-Thesis, Bristol, **2014**.
162. L. Zheng, U. Baumann and J.-L. Reymond, *Nucleic Acids Research*, **2004**, *32*, 115.
163. David Ivison, *Investigating the Programming of Type I Iterative Polyketide Synthase Enzymes*, PhD-Thesis, Bristol, **2013**.
164. D. Voet, J. G. Voet and C. W. Pratt, *Principles of Biochemistry*, **2008**.
165. H. A. David and J. L. Gunnink, *The American Statistician*, **1997**, *51*, 9.
166. S. S. Shapiro and M. B. Wilk, *Biometrika*, **1965**, *52*, 591.



167. J. Chiu, P. E. March, R. Lee and D. Tillett, *Nucleic Acids Research*, **2004**, *32*, 174.
168. J. Chiu, D. Tillett, I. W. Dawes and P. E. March, *Journal of Microbiological Methods*, **2008**, *73*, 195.
169. R. M. Q. Shanks, N. C. Caiazza, S. M. Hinsa, C. M. Toutain and G. A. O'Toole, *Applied and Environmental Microbiology*, **2006**, *72*, 5027.
170. S. B. Hua, M. Qiu, E. Chan, L. Zhu and Y. Luo, *Plasmid*, **1997**, *38*, 91.
171. M. A. Skiba, A. P. Sikkema, W. D. Fiers, W. H. Gerwick, D. H. Sherman, C. C. Aldrich and J. L. Smith, *ACS Chemical Biology*, **2016**, *11*, 3319.
172. P. A. Storm, D. A. Herbst, T. Maier and C. A. Townsend, *Cell Chemical Biology*, **2017**, *24*, 316.
173. M. A. Skiba, A. P. Sikkema, N. A. Moss, A. N. Lowell, M. Su, R. M. Sturgis, L. Gerwick, W. H. Gerwick, D. H. Sherman and J. L. Smith, *ACS Chem. Biol.*, **2018**, *13*, 1640.
174. J. L. Meinke, M. R. Mehaffey, D. T. Wagner, N. Sun, Z. Zhang, J. S. Brodbelt and A. T. Keatinge-Clay, *ACS Chem. Biol.*, **2018**, *13*, 3306.
175. A. Jansson, H. Koskiniemi, A. Erola, J. Wang, P. Mäntsälä, G. Schneider and J. Niemi, *The Journal of Biological Chemistry*, **2005**, *280*, 3636.
176. S. Kishimoto, Y. Tsunematsu, T. Matsushita, K. Hara, H. Hashimoto, Y. Tang and K. Watanabe, *Biochemistry*, **2019**, *58*, 3933.
177. S. Horowitz, L. M. A. Dirk, J. D. Yesselman, J. S. Nimtz, U. Adhikari, R. A. Mehl, S. Scheiner, R. L. Houtz, H. M. Al-Hashimi and R. C. Trievel, *ACS*, **2013**, *135*, 15536.
178. P. Z. Kozbial and A. R. Mushegian, *BMC Structural Biology*, **2005**, *5*, 19.
179. A. C. Legon and D. J. Millen, *Chem. Soc. Rev.*, **1987**, *16*, 467.
180. J. Zheng, D. C. Gay, B. Demeler, M. A. White and A. T. Keatinge-Clay, *Nature Chemical Biology*, **2012**, *8*, 615.
181. J. Zheng, C. A. Taylor, S. K. Piasecki and A. T. Keatinge-Clay, *Structure*, **2010**, *18*, 913.
182. S. A. Bonnett, J. R. Whicher, K. Papireddy, G. Florova, J. L. Smith and K. A. Reynolds, *Chemistry & Biology*, **2013**, *20*, 772.
183. S. K. Piasecki, J. Zheng, A. J. Axelrod, M. E. Detelich and A. T. Keatinge-Clay, *Proteins*, **2014**, *82*, 2067.
184. P. Caffrey, *ChemBioChem*, **2003**, *4*, 654.
185. A. T. Keatinge-Clay and R. M. Stroud, *Structure*, **2006**, *14*, 737.
186. C. Schubert, C. M. Milligan, K. Vo and B. Grasberger, *To be published*, **2016**.
187. H. E. Gottlieb, V. Kotlyar and A. Nudelman, *J. Org. Chem.*, **1997**, *62*, 7512.
188. N. M. Gaudelli and C. A. Townsend, *J. Org. Chem.*, **2013**, *78*, 6412.
189. A. L. Vagstad, A. G. Newman, P. A. Storm, K. Belecki, J. M. Crawford and C. A. Townsend, *Angewandte Chemie*, **2013**, *52*, 1718.
190. F. Pfrenkle, V. Dekaris, L. Schefzig, R. Zimmer and H.-U. Reissig, *Synlett*, **2008**, *2008*, 2965.

## 10 Appendix

### 10.1 Protein pdb Data File List (attached CD)

Number	Name (pdb file)	Content/Docked Substrated
1	SQTKS-ER-Apo	ER (w/o) NADPH
2	SQTKS-ER-Holo	ER with NADPH
3	SQTKS-ER-Tetra-S-S	Tetraketide <b>93a</b>
4	SQTKS-ER-Tetra-S-R	Tetraketide <b>93b</b>
5	SQTKS-ER-Tetra-R-S	Tetraketide <b>93c</b>
6	SQTKS-ER-Tetra-R-R	Tetraketide <b>93d</b>
7	SQTKS-ER-Tigloyl	Tigloyl <b>77</b>
8	SQTKS-ER-Triketide	Triketide <b>85</b>
9	SQTKS-ER-Penta	Pentaketide <b>94</b>
10	SQTKS-ER-I1938A	Tetraketide <b>93a</b>
11	SQTKS-ER-F1941A	Tetraketide <b>93aa</b>
12	SQTKS-ER-L2146A	Triketide <b>85</b>
13	SQTKS-ER-L2146A	Pentaketide <b>94</b>
14	SQTKS-ER-L2146V	Triketide <b>85</b>
15	SQTKS-ER-L2146V	Pentaketide <b>94</b>
16	SQTKS-ER-I2147A	Triketide <b>85</b>
17	SQTKS-ER-I2147A	Pentaketide <b>94</b>
18	SQTKS-ER-F2157A	Triketide <b>85</b>
19	SQTKS-ER-F2157A	Pentaketide <b>94</b>
20	SQTKS-ER-L2146A/I2147A	Triketide <b>85</b>
21	SQTKS-ER L2146A/I2147A	Pentaketide <b>94</b>
22	SQTKS-ER-I2147A/F2157A	Triketide <b>85</b>
23	SQTKS-ER-I2147A/F2157A	Pentaketide <b>94</b>
24	SQTKS-ER-F1941A/F2157A	Triketide <b>85</b>
25	SQTKS-ER-F1941A/F2157A	Pentaketide <b>94</b>
26	SQTKS-ER- F1941/I2147A/F2157V	Triketide <b>85</b>
27	SQTKS-ER- F1941/I2147A/F2157V	Pentaketide <b>94</b>
28	TENS-CMeT-Apo	C-MeT (w/o) SAM
29	TENS-CMeT-Holo	C-MeT with SAM
30	TENS-CMeT-Sub	Substrate <b>128</b>

<b>31</b>	TENS-CMeT-Sub	Substrate <b>129</b>
<b>32</b>	TENS-CMeT-Sub	Substrate <b>130</b>
<b>33</b>	TENS-KR-Apo	KR (w/o) NADPH
<b>34</b>	TENS-KR-Holo	KR with NADPH
<b>35</b>	TENS-KR-Sub	Substrate <b>135a</b>
<b>36</b>	TENS/mFAS-Hybrid	Chimeric structure of TENS KR, C-MET domain and mFAS DH, ER and $\psi$ KR
<b>37</b>	DMBS-KR	KR (w/o) NADPH
<b>38</b>	MILS-KR	KR (w/o) NADPH
<b>39</b>	TENS-KR*(MILS)	Mutated KR domain of TENS to MILS

## 10.2 Multiple Alignment of $\beta$ -Processing Domains of TENS, DMBS, mFAS pig and mFAS rat

		1		50
.			KS	KS
TENS	(1)	MSPMKQNESESHSVSEPIAIIIGSAYRFPGGCNTPSKLDLRLRQPRDILKE		
DMBS	(1)	MSPMKQNESESHCVSEPIAIVGSAYRFPGGCNTPSKLDLRLRQPRDILKE		
SQTKS	(1)	MVPYYQPASCGSN-SEPIAIIIGMSCRFPGNATSPKLEWELCAQGRSAWSS		
mFAS pig	(1)	-----MEEVVIAGMSGKLPESLENLEEFWANLIG-----		
mFAS rat	(1)	-----MEEVVIAGMSGKLPESLENLQEFWANLIG-----		
		51		100
.		KS		KS
TENS	(51)	IDPERLNLRYYHPDGETHGSTDVANKAYTLEEDISRFDASFFGISPLEA		
DMBS	(51)	IDPERLNLRYYHPDGETHGSTDVSNRAYTLEEDISRFDASFFGISPLEA		
SQTKS	(65)	IPKSRFRQEGFYNPNAERVG-TSHVVGGHFLEEDPSLFDASFFNLSAEAA		
mFAS pig	(29)	-----GVDMVTADDRRWKAGLYGLPRRMGKLDLSRFDASFFGVHSKQA		
mFAS rat	(29)	-----GVDMVTDDDRRWKAGLYGLPKRSGKLDLSKFDASFFGVHFKQA		
		101		150
.		KS		KS
TENS	(101)	ASMDPQORTLLEVVESTETAGIPLDKLRGSLT SVHVGVMITDWAQMQR		
DMBS	(101)	AGMDPQORTLLEVVESTETAGIPLDKLRGSLT SVHVGVMITDWAQMQR		
SQTKS	(114)	KTMDPQFRLLQLESVYEAMESAGITLEHIAGSDTSVYAGACFRDYHDSLVR		
mFAS pig	(73)	NTMDPQLRMLLEVVEAIVDGGINPASLRGTSTGVWVGVSSEASEALSR		
mFAS rat	(73)	HTMDPQLRLLEVVEAIVDGGINPASLRGTNTGVWVGVSSEASEALSR		
		151		200
.		KS	*****	KS
TENS	(151)	DPETMPQYTATGIASSIISNRISYIFDLKGASETIDTACSSSLVALHNAA		
DMBS	(151)	DPETMPQYTATGIASSIISNRISYIFDLKGASETIDTACSSSLVALHNAA		
SQTKS	(165)	DPDLVPRFLLTGNGAAMSSNRVSHFYDLRGASMTVDTGCSTTLTALHLAC		
mFAS pig	(123)	DPETLVGYSMIGCORAMMANRLSFFDFKGPSTITIDTACSSSLVALQ SAY		
mFAS rat	(123)	DPETLLGYSMVGCORAMMANRLSFFDFKGPSTIALDTACSSSLVALQ NAY		
		201		250
.		KS		KS
TENS	(201)	RALQSGDCEKAIIVAGVNLILDPPFIYESKLIHMLSPDARSRMWDAAANGY		
DMBS	(201)	RALQSGDSEKAIIVAGVNLILDPPFIYESKLIHMLSPDSRSMWDAAANGY		
SQTKS	(215)	QGLRNRESKTSIVTGANVILNPD MFVTMSSLGLLGPEGKSHFTDARANGY		
mFAS pig	(173)	QAIRGGECSAAVVGGNLNVLKPNSSLQFMKLGMLSQDGTCRSFD AEGTGY		
mFAS rat	(173)	QAIRSGECPAATVGGINLLKPNSSVQFMKLGMLSQDGTCRSFD DSGNGY		
		251		300
.		KS		KS
TENS	(251)	ARREGAAAVLKLTLGHALRDGDRIEGVIRSTFVNSDGLSSGLTMPSSAAQ		
DMBS	(251)	ARREGAAAVLKLTLGHALRDGDQIEGVIRSTYVNSDGLSSGLTMPSSAAQ		
SQTKS	(265)	GRGEGIATV IIKRLDDALRAQDP IRCIIRGTALNQDGRATLISPSQTAQ		
mFAS pig	(223)	CRAEAVVAVLLTKKSLARRVYATILNAGTNTDGSKE---QGVTFPSGDVQ		
mFAS rat	(223)	CRAEAVVAVLLTKKSLARRVYATILNAGTNTDGSKE---QGVTFPSGEAQ		
		301		350
.		KS		KS
TENS	(301)	TALIRQTYRKAGLDPVRDRPQFFECHGTGTKAGDPVEARAISDAFLP PSH		
DMBS	(301)	TALIRQTYRKAGLDPVKDRPQFFECHGTGTKAGDPVEARAISDAFLP NHK		
SQTKS	(315)	SDLIRACYRAALDPNDTAFLLAAHGTGTRTGDAVEIAAAAD-DKRS PERP		
mFAS pig	(270)	EQLIRSLYAPA--GDPESLEYIEAHGTGTKVGDPELNGITVNALCATRR		
mFAS rat	(270)	EQLIRSLYQPG--GVAPESLEYIEAHGTGTKVGDPELNGITRSLCAFRQ		

```

351                                     400
.
TENS (351) RTNGAATVVDAPLYVGSIKTVVGHLEGCAGLAGLVKVLLSLKHGIIFPNL
DMBS (351) TKG--AATVDAPLYVGSIKTVVGHLEGCAGLAGVIKVLLSLKHGIIFPNL
SQTKS (367) LWIGSVKTNIGHSEATSGLASVIQAALALEKGLIPPINIFKEPNEKLGQV
mFAS pig (318) EPL-----LIGSTKSNMGHPEPASGVAALIKVLLSLEHGVAAPNL
mFAS rat (318) SPL-----LIGSTKSNMGHPEPASGLAALTKVLLSLENGVWAPNL

401                                     450
.
TENS (401) WFDKLNPEIARYYGPLQIPTKAIWPVKLAPGTPLRASVNSFGFGGTNAHA
DMBS (399) WFNKLNPEIARYYGPLQIPTTAIPWPELAPGTFFRASVNSFGFGGTNAHA
SQTKS (417) SAAVRVPSNLQKWPSVSGVRRASVNNFGYGGANAHVILES GIPGHTPIAN
mFAS pig (358) HYHTFNPEIPALQDGRLOVVDRLPIRGGN-----VGINSFGFGGSNVHV
mFAS rat (358) HFHNFNPEIPALLDGRLOVVDRLPVRRGGI-----VGINSFGFGGANVHV

451                                     500
.
TENS (451) IIERYDASQSYCSQWRRNMTEEKTIARTQNNESIEIPVPLVLTAKTGRAL
DMBS (449) IIERYDANQSYCSQWRRDMTEEQKTIIVRPQDEGNTNIPVPLLLTAKTGGAL
SQTKS (467) GSGRSNGTGNHNGANGTTNGHNGTNGT-NGHFDATQATNGHYGTDETPD
mFAS pig (403) ILQ-----PNSRPAPPQAHAALPRLIQASGRGLEAV
mFAS rat (403) ILQ-----PNTQQAPAPAPHAALPHLLHASGRTMEAV

501                                     550
TENS (501) WRTVDAYAQHRLRQHPKLRVTNLSQFMHSRRSTHRVRASFSGASREELVEN
DMBS (499) WRTVDAYAQHRLRQNPGLANLSKFMHSRRATHRVRASFSGASREELLEN
SQTKS (517) YAPLDSFVISISAKEEASARSMVTNLADYLRTLQVQDETKHFKSIAHTLG
mFAS pig (435) QTLLEQGLRHSDRLAFVGMLEIAAVSPVAMPFRGYAVLGGEGS-----
mFAS rat (435) QGLEQGRQHSQDLAFVSMLENDIATPTAAMPFRGYTVLGVLEGHV-----

551                                     600
.
TENS (551) MAKFVQAAHAADAKSPASQNRIGYSPLHIDPKKEAPGILGVFTGQGAQWPAM
DMBS (549) MAKFVQAAHAADAKSPASQNRIGYSPLLIDPKKEVPGILGVFTGQGAQWPAM
SQTKS (567) SHRSMFKWTAAKSITGPEELIAAAEGGQFQASRALERTRLGFVFTGQGAQ
mFAS pig (480) -----QEVQQVPGSKRPVWFICSGMGCAQWQGM
mFAS rat (480) -----QEVQQVPASQRPLWFLICSGMGTQWRGM

601                                     650
.
TENS (601) GRDMMHQSPFLFRKTIADCESVLQALPAKDAPVWSLSSEELKKDASTSRLGE
DMBS (599) GRDMMHQSPFLFRKTIADCESVLQALPSKDVPWSLSSEELKKDASTSRLGE
SQTKS (617) WFAMGRELINTYPVFRQSLDRADRYLKEFGCEWSIIDELSRDAENS NVND
mFAS pig (507) GLSLMRLDRFRDSILRSDQALKPLGLRVSDLLLSSTDEAVLDDIVSS----
mFAS rat (507) GLSLMRLDSFRESILRSDQALKPLGVKVS D L L L S T D E H T F D D I V H S ----

651                                     700
.
TENS (651) AEISQPLCTAVQLALVNVLLASGVHFDVAVGHSSGEIAATYASGIINLEA
DMBS (649) AEISQPLCTAVQLALVNVLTASGVHFDVAVGHSSGEIAATYASGIISLKG
SQTKS (667) MTLSPPLCTAVQISLVQLLESWIVPTAVTGHSSGEIAAAYAAGLDFKSA
mFAS pig (553) ----FVSLTSTIQIALIDLTLTSLGLQPDGII GHSLGEVACGYADGCLTQEE
mFAS rat (553) ----FVSLTAIQIALIDLTLTSMGLKPDGII GHSLGEVACGYADGCLSQRE

701                                     750
.
TENS (701) AMQIAYYRGLYAKLARGETDAAGGMMAAGLSMNDVAVKLCRLEFEFGRIHV
DMBS (699) AMQIAYYRGLYAKLARGKSDSGGMMAAGLSMNEAVKLCRLEFEFGRIQV
SQTKS (717) AMAVTYFRGEVGLACQDKIVGKGMIAVGLGPEDAE-DRIARVQSGKIVV
mFAS pig (599) AVLSYWRGYCIKEAN---VLPGAMA AVGLSWE ECKQRCPPGIVP-----
mFAS rat (599) AVLAYWRGQCICKAN---LPAGSMA AVGLSWE ECKQRCPPGVVP-----

```

```

751                                     800
.
TENS (751) AASNAPQSVTLSGDKEAIKAAKAKLDADGVFARELKVDIAYHSHHMLPCA
DMBS (749) AASNAPQSVTLSGDKEAIKAAKAMLDSDGVFARELKVDIAYHSHHMLPCA
SQTKS (767) ACINSQSSVTVSGDLSGIVELEDLLKAEGVFARRVKVQAAAYHSHHMLQVIA
mFAS pig (641) ACHNSKDTVTTISGPOAAMSEFLQQLKREDVVFVKEVRTGGIAFHHSYFMEST
mFAS rat (641) ACHNSEDVTVTISGPOAAVNEEFVEQLKQEGVFAKEVRTGGGLAFHSYFMEGI

801                                     850
.
TENS (801) EPYLLKALLACDIQVSAPTTTPGRKCMWSSSVRGDAELLRHDRNLDLTKGP
DMBS (799) EPYLLKALLACDIQVSAPTPG---KCMWSSSVRGDAELLRHDRNLDLTKGP
SQTKS (817) NGYLLTSLKDKMLKPTKKFGKIIYSSPTTGRRETNAKLMSAQHWVNNMLSP
mFAS pig (691) APTLLRQLRKVILDPKPRSK-----RWLSTSIPEAQWQSSLART
mFAS rat (691) APTLLQALKKVIREFRPSA-----RWLSTSIPEAQWQSSLART

851                                     900
.
TENS (851) YWVANMVQTVLFSRAVQSTIWHGGPFDLAEVGPHPALKGPTTEQTLKAVY
DMBS (846) YWVANMVQTVLFSRAVQSTIWHGGPFDLAEVGPHPALKGPTTEQTLKAVY
SQTKS (867) VRFAESFQNMCFNSNRNSSQSEEIFQDVIDIVLEVGPHGMLQGPQQMMSLP
mFAS pig (730) FSAEYSVNNLVSPVLFQEALQHVPAHAVVVEIAPHALLQAVLKRSLSSC
mFAS rat (730) SSAEYNVNNLVSPVLFQEALQHVPEHAVVLEIAPHALLQAVLKRGVKPSK

901                                     950
.
TENS (901) GSAPLYTGVLRSRGANDAVAFSTAIIGNIWSHLGPAFVDITGYQSIFFSSTICE
DMBS (896) GSTPLYTGVLRRRGANDAVAFSTAIIGNIWSHLGPAFVDMTGCQSIFFSGASE
SQTKS (917) IFERARLPYISCLLRGQSAVHT-----MQTVAAGLMGWGYRVDMAVN
mFAS pig (780) TIIPLMKKDHKDNLEFFLSNVGRLHLAGVSVNPNGLFPPVEFPFRGTPL
mFAS rat (780) TIIPLMKRDHKDNLEFFLTNLGKVHLTGIDINPNALEFPVEFPVFRGTPL

951                                     1000
.
TENS (951) GHGGGSAAPFISDLPLYPWDHDEEYWRRESRISRRHRTGKDESHELLGRRT
DMBS (946) GHGGSAAPFISDLPLYPWDHDEEYWRRESRISRRYRTGKDESHELLGRRT
SQTKS (967) FPQGTYGVKILHDLPSYPWNHDNSYWWEPRLNKAHRQVRVPPHELLGLSLI
mFAS pig (830) ISPH-----IKWDHSQAWDVPAAADFPSPSSCSSAVVYKFDVS
mFAS rat (830) ISPH-----IKWDHSQAWDVPAAADFPSPSSCSSAVVYKFDVS

1001                                    1050
.
TENS (1001) DH*****DH
DMBS (995) PDDNEREIRWRNLLKVSELPWTQGHRVLEEVLLPGAAYISMAIEAGRRIA
SQTKS (1016) VGRDLREPTWRHFIRVQDIPWIRDHVVSQALVYPGAGFICMAMEAMVQLH
mFAS pig (868) PESP-----DHYLVDHCIDGRVLFPGTGYLWLTWTKTLARAL
mFAS rat (868) SESS-----DHYLVDHCIDGRVLFPGTGYLYLVWTKTLARSL

1051                                    1100
.
TENS (1051) LDQGREARLLEVSDVDILRPVVADNKEGTETLFTVRLIDEYASTGKKSD
DMBS (1045) LDQGRQVCLEEVFDVDILRPVVADNKEGTETLFTVRLIDEHTVSAKKLD
SQTKS (1067) ELRDSQSRKVAGYRLAEVDILRAMLIPDTSEGLEAHISLRPCSTKLLLLTN
mFAS pig (904) SQN-----LEETPVVFFEDVTLHQATILPKTGT
mFAS rat (904) SLS-----LEETPVVFFENVTFHQATILPRTGT

1101                                    1150
.
TENS (1101) ELITASFSFYIYNPASTSIVHTCEGRIAVQLGAKLGSEAGANSMPQLPH
DMBS (1095) EIITASFSFYIHNSASTSVVHTCEGRMAVHLGAKLGSVGVANSMPQLPQ
SQTKS (1117) EWYDFCVSSVGGDDDKFVDHCRGTITIEFDTSGSADTPRTLRSRSTGLM
mFAS pig (931) VSLEVRLLLEASHAFEVSDSNGSLIASGKVYQWESPDPKLFDTRAAVDPAD
mFAS rat (931) VPLEVRLLLEASHAFEVSDS-GNLIIVSGKVYQWEDPDSKLFDHPEVPIPAE

1151                                    1200

```

```

.
.
TENS (1151) REPSISNLQQLDCEKLYSVFETIIGLEYSGAFFRIVSSRCLGHATATASW
DMBS (1145) RELSVSNLQPIDCEKLYSLFETIIGLEYSGAFFRAINSSSRRLGHATASASW
SQTKS (1181) RSVDPNSLYSFLRAQGIYHGPIFQNLKTISSRKDHSESSFVVANTASVMP
mFAS pig (981) STAEFRLSQGDVYKDLRLRGYDYGPFQLVLES DLEGNR-----
mFAS rat (980) SESVSRLTQGEVYKELRLRGYDYGPFHQGVYEATLEGEQ-----

1201 1250
.
.
TENS (1201) PTTDLNDCYLHPAILDVAFQTFVARAHPDSGQLS SALLPSRIERVRVV
DMBS (1195) ASL DLNDCYLHPAILDVAFQTFVARAHPDSGQLN SALLPSRIERVRVI
SQTKS (1218) NGFQSPHVIHPTTLD SIFQAYTALPGAGLDQNTAMIPRSIQELYLSSAL
mFAS pig (1020) -----
mFAS rat (1019) -----

1251 1300
.
.
.
TENS (1251) PSLAMGSKLQNNENFNAAIDSWALNQ TASSLTGNINVYDAESGRALIQVE
DMBS (1245) PSSAMESKLOSNENINAEIDSWVLNQ TVSSLTGLNVYD TDTGTIPLQVE
SQTKS (1268) TSDVQCLVSDTSLIRYDQSFVTVNVVDVSSKADSEHTPVLEIKGLRNQSV
mFAS pig (1020) -----GRLOWNDSWVSFLDAMLHMSILAPGQLGLYIIPTRFTSIRIDPVT
mFAS rat (1019) -----GKILWKDNWVTFMDTMLQISILGFSKQSLQLPTRVTAIYIDPAT

1301 1350
.
.
.
TENS (1301) GFEVRAVGEPDASKDRLLFYETVWGRDISIMGLSDPIRDETS DAMVHNL S
DMBS (1295) GFEVRAVGEPDASKDRLLFSETVWGRDISIMGLSDPIRNETT DAAVQSLAD
MILS (1341) QSLA
5thz (64) EELS
SQTKS (1318) GQMAPQPGDSSNNDLCKFLDWAPDISSVKQERLKEKFGFPLDPT EADIIM
mFAS pig (1064) HRQKLYTLQD TQAADVVDRLNLT VVAGGALFLGAHSSVAPRRPQEHLK
mFAS rat (1063) HLQKVYML EGD TQVADVTT SRCLGVTVSGGVYISRLQTTATSRRQQEQLV

1351 1400
.
.
.
TENS (1351) EAIERSVSLFYVRQLMGELSTADRRQANWYHTRMLA AFDYHLAKVHEETHL
DMBS (1345) EAIERSVSLFYVRQLMSELSTKDRREANWYHSRMLTAFEHHLARIHEDTHL
MILS (1345) EAGERVSLFYVRRLMSELTAEDRDQANWYHTRMLQAFDHHLTEVKNDTHL
5thz (68) VDYIVQGLLQMGWSYQPTESFDLD-CLGVVPTQVRLFERLLQILA EVGIL
SQTKS (1381) GLRQACIHF IHRSLQSLTAPDRDQLDWHOKRFYDWMVLQIQLA EEDRLSA
mFAS pig (1114) PILEKFCFTPHVESGCLAGNTALQEE LQLCRGLAQALQTKVAQQGLK MVV
mFAS rat (1113) PTLEK FVFTPHVEPECLSESA I LQKELQLCKGLAKALQTKATQQGLKMTV

1401 1450
.
.
.
TENS (1401) HLRPEWLADDWAVIQ TIDEAYPDAVE LQMLHAVGQNVADVIRGK KHLLEV
DMBS (1395) HVRQEWLSDDW SVIQIIDEAYPDTVELQMLHAI GQNMANVIRGEKHMLEV
MILS (1394) HLRREWLSDDWAAIHAIDEAYPDTVELQMLHAVGKNMVSVIKGEQHMLEV
5thz (123) -NQQWQVQKT-QSQSLLSQYPDE-TLTL LERCASQLSGVLRGEIDPVQL
SQTKS (1418) NSSAWLQCSSSDEQKLL ENVRASSVNGQMVVHVGVKSM LA I LRHEIAPLEL
mFAS pig (1164) PGLDGAQAPREAPQ QSLPRL LAAACQLQLNGNLQLELIGQVLAQE-----
mFAS rat (1163) PGLE-----DLPQHGLPRL LAAACQLQLNGNLQLELIGEVLAQE-----

```

```

1451                                     1500
. CMeT *****CMeT
. 1B1      1B11B2                                     1B2
. SC      CC S                                     CCCCCCCC CCC
TENS (1451) LRVNLLDRLYTEDKGMHMANLFLANALEEITFKFPRCKILEIGAGTGAT
DMBS (1445) MRVNNLLDRLYTEDKGMQGNHFLANALKEITFKFPRCKILEIGAGTGAT
MILS (1444) MRVDDMLDRFYADDKGMQOVNHFLAGALNEITFKFPRCNILEIGAGTGAT
5thz (180) VFPQGDIL-QLYKDSAVAKVMNTIVEKVMKAMEKLPSS-LLLEIGAGTGGT
SQTKS (1468) MLQDKLLYRYYTDAIKWDRSYQQIDQLVKLHAHKCPTAKIIEIGAGTGGC
mFAS pig (1208) ---RPLLCDPPLL SGLLDAPALKACVDTALENMAFPKMKVVEVLVAGDGGQL
mFAS rat (1201) ---RLLPEDPLI SGLLNSQALKACIDTALENLSTLKMVVEVLVAGEGHL

```

```

1501                                     1550
. CMeT CMeT
. 1B2      1B22A1                                     2A1
. 1B2B      CCC C                                     CCCS
TENS (1501) TWAALSAIGEAFTDYTYTDL SVGF FENAVERFSAFRHRMVFRAIDIEKDF
DMBS (1495) TWAVLSAIDETFDTYTYTDL SVGF FETAVERFSAFRHKMIFKALDIEKSP
MILS (1494) TSAVLNALDDAFDTYTYTDL SVGF FETAMERFSSFRHKMIFKALDVEKDV
5thz (236) TSYILPHLNPNQTEYIFTDL GALFTSKAQEKFDYRF-LGYQTL DIEVDP
SQTKS (1518) TRAVLDA-IARCAQYDFTDVSSGF FEAAQQKFAAFDDVIRFQKLDIEKDI
mFAS pig (1255) YSRIPALINTQPVMDLDYTATDRNPOALEAAQAKLEQLHVTVQGWDFAN-
mFAS rat (1248) YSHISALINTQVPLQLEYTATDRHPQALKDVOTKLVQHDVAQGWDFSG-

```

```

1551                                     1600
. CMeT CMeT
. 2A2                                     2A2
. CCC SSS S
TENS (1551) ASQSFDLNSYDIIIIATNVLHATRNLGVT LGNVRALLKPGGYLLILNEKTGP
DMBS (1545) AAQSFDLGSYDIIIIATNVLHATRNLDT LGNVRSLKPGGYLLILNEKTGP
MILS (1544) ATQGFDLGSYDIIIIAANVLHATRSLEVTLGNVRSLLKAGGYLLILNEKTGA
SQTKS (1572) EMQGFECGSYDLVIASQVVLHATGKMEHTMANVRKLLKPGGKLLLVETTRD
5thz (285) SSQGFESHRYDVIIAANVLHATTSLSKQTL SHVRQLIAPGGIILVLYEAT-T
mFAS pig (1304) -PAPGSLGKADLLVCNCAIATLGDPAVAVGNMAATLKEGGFLIILHTLLAG
mFAS rat (1297) -PAPTNLGALDLVCNCAIATLGDPALALDNMVAALKDGGFLIIMHTVLKG

```

```

1601                                     1650
. CMeT CMeT
. 2B1                                     2B12B2 2B2
. S S SSS S SS
TENS (1601) ESLRATFNFGGLEGWLLAEEKER-QLSPILMSPDGWDAQLQKASFSQVVDHI
DMBS (1595) ESLRATFNFGGLEGWLLAEEEDR-QLSPILMSPDGWDSQLQKTFQFSQVVDHV
MILS (1594) ESLRATFNFGGLQGWLLAEEEDR-QLSPILMSPDGWDAQLQRRARFSQIDHV
5thz (334) RSRWVDLIFGLLEGWW-TDYELR-PDYPLLNRQWKKVLSSETGFETQVV-T
SQTKS (1622) E-MDLQLVFGLLPGWLLSSEEEERQMSPSLSTNSWEKVLKKTGF DGLDIEL
mFAS pig (1353) HPLGEMVGF L TSPEQGGRRHLLSQDQWESLFGASLHLVALKRSFYGSVLF
mFAS rat (1346) HALGETLACLPSQEVQPGPSFLSQEWEWESLFSRKALHLVGLKKSFYGTALF

```

```

1651                                     1700
. CMeT CMeTYKR YKR
. 2B2      2B2YKR1 YKR1
.
TENS (1650) VHDVQEDQQDKQNSMIMSQA VDDTFYARLSP LSEMANLLPMNEPLLIIG
DMBS (1644) VHDVQEEGKQ--QNSMIMSQA VDDAFYARLSP LSEMASLLPTQEPPLLIIG
MILS (1643) VHDIPPEE-PKQQ
5thz (383) LPEVEGMAEALSQQT VIVAQAAS
SQTKS (1672) LRDCDSDEFYSFSVMATASSTIASSSMAFAIVYGEVPLPDQFLDDMKTA
mFAS pig (1403) LCRQQT PQDSP-----VFLSVEDTSFRWVDSLK DILADASSRPVWLM AV
mFAS rat (1396) LCRRLS PQDKP-----IFLPSVEDTSFQWVDSLK S ILATSSSQPVWLTAM

```



		1701		1750
		YKR		
		YKR1	YKR1YKR2	YKR2
TENS	(1700)	GQTTATLKMILKEIQKLLPRQWRHKVRLIASVDHVEAEGLPAHSDVICLOE		
DMBS	(1692)	GQNTTLRLIKEIQKQLPRKWRHKIRLIASVDQLEDEDLPAHSDVICVQE		
SQTKS	(1721)	ISSSAVSDPVVGHLDSDATGKFCIFIEDPETDILSSPDEKSYASIQKLV		
mFAS pig	(1447)	GCSTSGVVGMVNCLEKEPGGHRIRCVLVSNLSSTSPAPEMHFPSSSELOKVV		
mFAS rat	(1440)	NCPSTSGVVGLVNCLEKEPGGHRIRCVLVSNLSSTSHVPKLDFGSSSELOKVV		
		1751		1800
		YKR		
		YKR2	YKRER	ER
TENS	(1750)	LDRGLFTTAMTSKCLDALKTLFINTRNLLWVTNAQNSSSMTPRASMFRTGI		
DMBS	(1742)	LDRGLFTTAMTSKRLNALKSLFMNTKNLLWVTNAQNSSSMTPRASMFRTGI		
SQTKS	(1771)	TRCKGLIWVSRGGAMHGTRPNSSLKTGLLRTLRLEYTEKRFISLDLDSAR		
mFAS pig	(1497)	LQGDLMNVYRDGAWGAFRHFPLEQDRPEKQTEHAFVN-----VL		
mFAS rat	(1490)	LESGLMNVYRDGAWGAFRHFQLEQDKPEEQTAHAFVN-----VL		
		1801		1850
		ER		
TENS	(1800)	TRVLDGEVPHIRTQVLGIEPRETPSATARTLLEAFLRIRSDDGRHAGNVD		
DMBS	(1792)	TRVMDGEVPHIRTQILGIEPIGAPSTIARNLLEAFLRIRFDDTYQAATID		
SQTKS	(1821)	PQWNHDSITTINEVLCGALAQADSSIKDSEFAEQDQLFVPRISCDIAR		
mFAS pig	(1537)	SRGDLSSIRWVCSPLHYALP---ASCQDRLCSVYYTSLNFRDVMLATGKL		
mFAS rat	(1530)	TRGDLASIRWVSSPLKHMQPP--SSSGAQLCTVYYASLNFRDIMLATGKL		
		1851		1900
		ER		
TENS	(1850)	EDGADGSSQQVLWLHEPEAEALLSNGTMMVPRVKARKSLNDTYLASTRAIS		
DMBS	(1842)	GDGADGGSQQVLWSHEPEVDLLSSGTMMIPRVKLRKSLNDTYLASTRAIS		
SQTKS	(1871)	NEDLSSDSNSPAQMEPFHQPGKLLQMGIKTPGLIDTLQFSKTDATDNLPN		
mFAS pig	(1584)	SPDSIPG-----		
mFAS rat	(1578)	SPDAIPG-----		
		1901		1950
		ER		
TENS	(1900)	TTVDARCVSVQAVAGPAKMLLRPVEDFAGEHAISNQTSDSKVHIQVESTL		
DMBS	(1892)	TTVDARCVVPQAVAGPAKIMLRPVEDIAVDHEISSQTSDPKVHIQVEVTL		
SQTKS	(1921)	DYIEIEPKAFGLNFRDVMVAMQGLEESIMGFECAGVRRVGPSSAGHNIK		
mFAS pig	(1591)	KWLTRDCMLGMEFSGRDASGRVMGMVPAEGLATSVLLLQHATWEVPSW		
mFAS rat	(1585)	KWARSDCMLGMEFSGRDKCGRVMGLVPAEGLATSVLLSPDFLWDVPSW		
		1951		2000
		ER		
TENS	(1950)	HIPEALDGTCLYLVCGWTRTAET----SVPVIALSANNASMVAVESKAVVA		
DMBS	(1942)	HIPEALDGTCLYLVCGWTRPAEASDTS SVPVMALSTS NASLI AVEPKAVVA		
SQTKS	(1971)	VGDRVCALLGGQWTNTVRVHWS----VAPIPQAMDWETAASIPVFAVA		
mFAS pig	(1641)	TLEEAASVPIVYTTAYYSLVVRGRMQ-----PGESVL		
mFAS rat	(1635)	TLEEAASVPIVYTTAYYSLVVRGRIQ-----HGETVL		
		2001		2050
		ER		
TENS	(1996)	MIDEVDVKPETLLRVFQHMAMQALDSAVKRHGQCGQSTALIYGADDEELAKL		
DMBS	(1992)	MIDEVDLKPEALLRVFQHMAMQAVDSAVRRHGQRQRTALIYGADDEELAKL		
SQTKS	(2018)	YISLVKIAMQAGETVLIHAASGGVGQAAILAKHVGAEIFATVGTDEKR		
mFAS pig	(1673)	IHSGSGGVQAAIAIALSRGCRVFTTVGSAEKRAYLQARFPQLDETCFAN		
mFAS rat	(1667)	IHSGSGGVQAAIAIALSLGCRVFTTVGSAEKRAYLQARFPQLDDTSEAN		

		2051		2100
.		ER ER		
TENS	(2046)	TSERFVAVRESKVVYFASRTFAPGDWLVKVPQLLSKFALSQMIFADVEVFID		
DMBS	(2042)	TSKRCAVRESKIYFASSHSAAPGDWLVKVRHLLSSKFAMSQMVFSGVQVFID		
SQTKS	(2067)	DLLIKEYKIPDDHIFSSRNALFAKSIHQRTNGKGVVDVNLCLAGLLQES		
mFAS pig	(1723)	SRDTSFEQHVLRHTAGKGVLDLVNLSLAEKLIQASVRCVLAQHGGRFLEIGKF		
mFAS rat	(1717)	SRDTSFEQHVLLHTGGKGVLDLVNLSLAEKLIQASVRCVLAQHGGRFLEIGKF		
		2101		2150
.		ER ER		
TENS	(2096)	CLGDTESFDACRTLQSCLSVTRTVQHRLDACLLSQMSRCSPDALVDAYS		
DMBS	(2092)	CLGGTESFDACRTLQSCVLPVTTCTVHR-LDACLLSEMSQCSPDFLLDAYS		
SQTKS	(2117)	FDCLADFGRFIEIGKRDIELNHCLNMGMFARSATFTAVDLIAIGDRSYM		
mFAS pig	(1773)	DLSNNHALGMVAVFLKNVTFHGILLDSLFEEGATWQEVSELIKAGIQEGV		
mFAS rat	(1767)	DLSNNHPLGMAIFLKNVTFHGILLDALFEGANDSWREVAELLIKAGIRDGV		
		2151		2200
.		ER ER		
TENS	(2146)	AKTQSNAEFSWNGYVKTFTAAELAGKLSHSLIHSVYMTNWQKDSILVTV		
DMBS	(2141)	AQTQSNAGFSRSDNIKFTAAELAGKLSHSLINSMYITDWQKQDAILVTV		
SQTKS	(2166)	FAEALPKIMTLLQEKAIRPVTPISIIYKIGDIETAFRLMQAGKHMGIKIVIT		
mFAS pig	(1823)	VQPLKCTVFPRTKVEAAFRYMAQGKHIGKVVIVQVREEEQGPAPRGLPPIA		
mFAS rat	(1817)	VKPLKCTVFPKQAVEDAFRYMAQGKHIGKVLVQVREEEPEAMLPGAQPTL		
		2201		2250
.		ER KR ***** KR		
.		4A 3A		
.		CCCCCCCC B CCC CC		
TENS	(2196)	PPLQTRGLFKSDRTYLMVGAAGGLGTSICRWMVRNGARHVVVTSRN--PK		
DMBS	(2191)	PPLQTRGLFKSDRTYLMVGAAGGLGTSICRWMVRNGARHVVVTSRN--PK		
MILS	(2196)	LFQSDRTYLMVGAAGGVGTSICRWMVRHGARHVIIVTSRN--PK		
3mjv	(235)	-----RPPVHGSVLVTGGTGGIGGRVARRLAEQGAHLVLTSTRRGAD-		
SQTKS	(2118)	APEDAMVPVITRPPKLQLRPDASYLIVGGLGGIGRSLCKNFVENGA--RS		
mFAS pig	(1873)	LTGLSKTFCPPHKSYVITGGLGGFGLQLAQWLRIRGAKVLTSTRSGIRT		
mFAS rat	(1867)	ISAIKSTFCPEHKSYIITGGLGGFGLQLARWLVLRGAKVLTSTRSGIRT		
		2251		2300
.		KR KR		
.		4A 4A4B 3B		
.		C CCCC CC		
TENS	(2244)	ADPEMLNEAERYGAIVRVVPMDCNPKDSVQTVVDMIRATMPPIAGVCNAA		
DMBS	(2239)	ADPEMLNEAERYGAIVRVVPMDCNPKDSVQTVVDMIRATMPPIAGVCNAA		
MILS	(2237)	GDPTMLSEAKQYGATVRVVSMDVCDRRSVEAVVGMIRATMPPIACVCNAA		
3mjv	(279)	GAAELRAELEQLGVRVTIAACDAADREALAALLAEL-PEDAPLTAVFHSA		
SQTKS	(2265)	LVLLSRNANVSQQSGEFLDELSTGCIVGVVDCDISSKTQVEATMLRLKK		
mFAS pig	(1923)	GYQARQVREWRROGVQVVLVSTSNASSLDGARS LITEATQLGFPVGGVFNLA		
mFAS rat	(1917)	GYQAKHVREWRROGIVHLVSTSNVSSLEGARALIAEATKLGFPVGGVFNLA		
		2301		2350
.		KR KR		
.		4B 4B		
.		CCS CC B CBCS		
TENS	(2294)	MVLKDKLFLDMNVDMKDVLPKMQGTEHLDLSIFAQEP--LDFFVLLSSS		
DMBS	(2289)	MVLCDKLFLLDMDVDQMNNTLGPVVDGTEYLDLSIFAHEP--LDFFVLLSSA		
MILS	(2287)	MVLCDKLFLLDMDVDILNNTLGPVVDGTEYLDLSIFSEEA--LDFFVLLGST		
3mjv	(328)	GVAHDD-PVDLTGLQVDFALMRKLTAAARHLHELTADL--DLDAFVLFSSG		
SQTKS	(2317)	PIRGIVHAGMVLQSDVFERMSLDYNTAIRPKV-----SWNLHSGLS		
mFAS pig	(1973)	MVLKDAVLENQTPPEFFQDVSKPKYSGTANLDRVTREACPELDYFVIFSSV		
mFAS rat	(1967)	MVLKDAVLENQTPPELQDVNPKPKYNGTLNLDRAVREACPELDYFVAFSSV		

2351 2400

. KR KR

. 5A 5A

. S SSS S C CCCC CCCS C

TENS (2342) AAILNNTGQSNYHCANLYMDSLVTNRRSRGLAASIIHVGHVCDTGYVARI

DMBS (2337) AAILNNTMGQSNYHCANLYMDSLVKHRRSRGLAASIIHGHVCDTGYVARM

MILS (2235) ATIANNIGQSNYHCANLYMDSLVAQRRSRGLAASIIHIGYICDTGYVARL

3mjv AAVFGSGGQPGYAAANAYLDALAEHRRSLGLTASSVAWGTWGEVGMATDP

SQTKS (2363) CDLDFIFIMLSSLAGVSGSASQANYTAGGAYQDALAKYRRAQGLSAVSIIDL

mFAS pig (2023) SCGRGNAGQANYGFANSAMERICERKRRHDGLPGLAVQWGAIGDVGIVLET

mFAS rat (2017) SCGRGNAGQSNYGFANSTMERICEQRRHDGLPGLAVQWGAIGDVGIIIEA

2401 2450

. KR KR

. 5A 5A5B 5B

. C S SS<sup>B</sup> S S

TENS (2392) VDDTKVQMSLGTTRVMSVSETDVHHAFAEAVRGGQPDSRSGSHNIIMGIE

DMBS (2387) VDDNRIQSNIAITMRAMRLSETDVHHAFAQAVRGGQLDSRSGSYNIIMGIE

MILS (2385) GDDAKVHSNRDVMRATTLSETDVHHAFAEAVRGGSPGSPIGSYNIIMGID

3mjv (337) EVHDLRVRQGV LAMEPPEHALGALDQMLNDDTAAAPITMDWEMFAPAFTN

SQTKS (2413) GMVQSVGYVAETKGV AERLVRMGYSPISEMEVLKIVEHAITNPPPEASSA

mFAS pig (2073) MGTNDTVIGGTLPRIASCLEVLDLFLSQPHPVLS

mFAS rat (2067) MGTNDTVVGGTLPRISSCMEVLDLFLNQPHAVLS

2451 2500

. KR KR

. 5B 5B

TENS (2442) PPTKPLDLTKRKPVWISDPRLGCLPFSTLENQMMASEQAAAASAVDSLAA

DMBS (2437) PPTKPLDLTRRQAVWISDPRLGHMLPYSTLENQMIASGQAAAAS-ADSLA-

MILS (2435) PPTKSLDSSRRKALWISDPRLGHMVPYSASADQAVTSEQA

3mjv (478) RPSALLSTVPEAVSALSDE-----

SQTKS (2463) QIITGISTKPGRHWTSSWLQDARFATLRERARDVKELSNSQGAQ-DKQL

mFAS pig (2108) -----SFLVLAEKKAAPRDSGSSQK-----

mFAS rat (2102) -----SFLVLEKKAVAHGDGEAQR-----

2501 2550

. ACP \*\*\*\*\*

TENS (2492) QQVSEATTDEEA AVAALKGFATKLEGIILLPLGSI GEDSAGRPVTDLGLD

DMBS (2485) QQVSEATTDEEA TAAV LKGFATKLEGIILLPLPGSI GEDSAGRPVTDLGLD

SQTKS (2512) AAGQELSMATSLVEAIDVVGRAITAKLATMFLIAESI IASKSLSEYGV

mFAS pig (2127) -----DLVKAVAHILGIRDVASTINPDSTIIVDLGLD

mFAS rat (2121) -----DLVKAVAHILGIRDLAGINLDSLADLGLD

2551 2600

. \*\*\*\*\* ACP

TENS (2542) SILVAVEIRTWFLKQLRVDVPVMKILGGSTVGLSALAAKLARQDAKKRAQ

DMBS (2535) SILVAVEIRTWFLKQLRVDVPVMKILGGSTVGLSALAAKLARQDAKKQAQ

SQTKS (2563) SILVAVELRNWLAQAQLSSDVSVF-TQSQSLTALATTVATKSSRIDKSLVA

mFAS pig (2157) SLMGVEVRQILEREHDLVLSMREVRQLSLRKLQELSSKTSTDADPATPTS

mFAS rat (2151) SLMGVEVRQILEREHDLVLP IREVRQLTLRKLQEMSSKAGSDTELAAPK-

2601 2650

TENS (2592) LEEPSGNQPVALPSP PPKDKAGGLNKNKGS PKLPEIAQVDIVVERMEPLV

DMBS (2585) VEEASGNQHVALP -- PPKDKVG -PNTNGKAQDSPETAQVGTILIERMEPLV

SQTKS (----) -----

mFAS pig (2207) HEDSPVRQQATILN-----LSTLLVNPEGPTLTRLNSVQSAERPLF

mFAS rat (2200) SKNDTSLKQAQLN-----LSILLVNPEGPTLTRLNSVQSSERPLF

### Alignment colour coding

	<i>Foreground</i>	<i>Background</i>
Non-similar	black	white
Conservative	dark blue	cyan
Block of Similar	black	light green
Identical	red	yellow
Weakly Similar	dark green	white

C = Cofactor binding

\*\*\*\* = 'active site'

S = Substrate binding

## Wissenschaftlicher Werdegang

- 06/2016 – 11/2019** Leibniz University Hannover (LUH) /BMWZ, **PhD in Chemistry**
- 10/2015-05/2016** LUH, **M. Sc.**, Medical Natural Product Chemistry
- 05/2015 - 10/2015** Department of Proteomics, **Scientific assistant**, Helmholtz Centre for Environmental Research GmbH Leipzig Germany
- 10/2012 10/2014-** University Halle-Wittenberg, **M. Sc.**, Biochemistry
- 10/2009 - 09/2012** University Leipzig – **Bs. Sc.**, Biochemistry
- 2009** **Abitur** Ratsgymnasium Minden

## Veröffentlichungen, Publikationen

Yang X., Friedrich S., Yin S., **Piech O.**, Williams K., Simpson T.J. and Cox R.J., Molecular Basis of Methylation and Chain-Length Programming in a Fungal Iterative Highly Reducing Polyketide Synthase, *Chem. Sci.*; **2019**, *10*, 8478-8489