

***In Vitro* Investigation of Multi-domain Fragments
of Squalestatin Tetraketide Synthase**

Von der Naturwissenschaftlichen Fakultät der
Gottfried Wilhelm Leibniz Universität Hannover

zur Erlangung des Grades

Doktor der Naturwissenschaften (Dr. rer. nat.)

genehmigte Dissertation

von

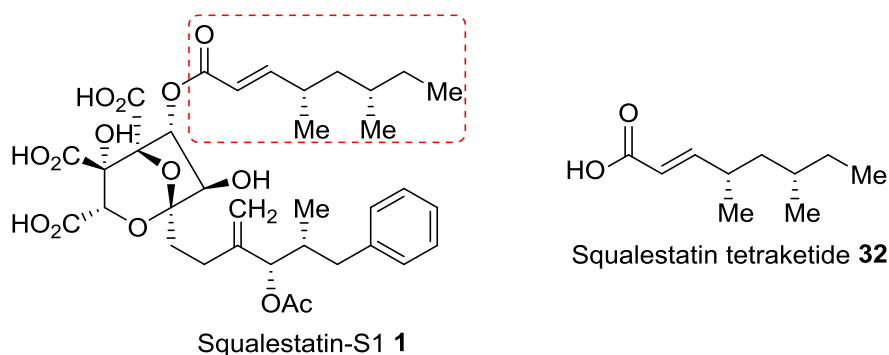
Hao Yao, Master of Science (China)

2019

Referent: Prof. Dr. Russell John Cox
Korreferent: Prof. Dr. Kürşad Turgay
Tag der Promotion: 07. Januar 2019

Abstract

Squalestatin tetraketide synthase (SQTKS) is a fungal iterative highly-reducing polyketide synthase (HR-PKS) that catalyzes the biosynthesis of the tetraketide side chain of squalestatin-S1 **1** which is a potent squalene synthase inhibitor and can be potentially used to treat serum cholesterol related diseases. The SQTKS protein is one of the simplest iterative type I HR-PKS as all the iterative β -modification domains are present in an active state and in which a degree of programming occurs. To investigate the programming of the HR-PKS, detailed *in vitro* and stereochemical studies are fundamental.



In this thesis, we have cloned and overexpressed several large multi-domain fragments of SQTKS in *E. coli*. These fragments have been isolated under a rational purification design and analyzed biochemically and biophysically.

Using pantetheine substrates (analogs of the true ACP-bound substrate in SQTKS), kinetic studies of enoyl reduction show that the ER of DH-KR tetradomain is capable of transforming various substrates and it performs the programming function by its inability to reduce the final tetraketide substrate. These results are in accord with the kinetic studies of the isolated SQTKS ER monodomain.^[75]

As an effect of domain-domain interaction, different stereoselectivities in enoyl reduction have been found by *in vitro* assays with the SQTKS ER monodomain^[75] and with a multidomain construct. With the tetradomain protein, the stereopreference of keto reduction has been studied. From the comparison between the mammalian fatty acid synthase (mFAS)

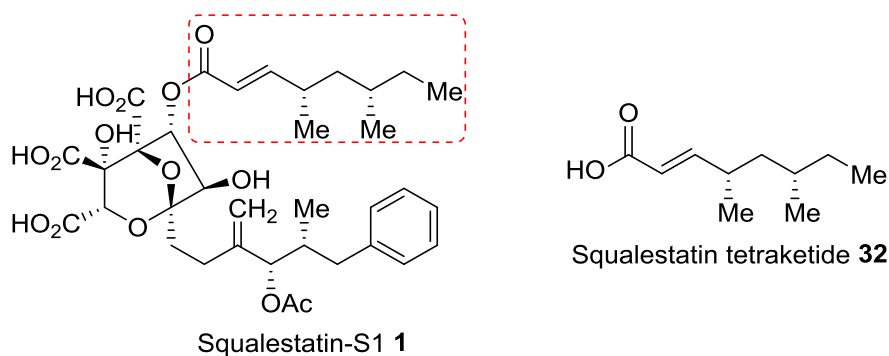
and the SQTKS in stereoselectivity, they are identical, which reinforces the idea that the HR-PKS and mFAS evolved from a common ancestor.

To investigate the methyltransfer mechanism, *in vitro* assays with isolated SQTKS C-methyl transferase (CMeT) monodomain have been studied. In addition, the effectiveness of using the pantetheine substrate as analogs of the original ACP-bound substrate was investigated in this research. These studies indicate that the full methyltransfer activity of SQTKS CMeT may need the cooperation of the acyl carrier protein (ACP).

Keywords: squalestatin tetraketide synthase, multi-domain fragments, *in vitro*, stereoselectivity.

Kurzzusammenfassung

Squalestatin Tetraketid Synthase (SQTKS) ist eine mykotische, iterative hoch-reduzierende Polyketid-Synthase (HR-PKS), die die Biosynthese der Tetraketid-Seitenkette des Squalestatin S1 **1** katalysiert. Squalestatin S1 ist ein potenzieller Squalensynthese Inhibitor und ermöglicht die Behandlung gegen Serumcholesterin ausgelöste Krankheiten. Das SQTKS Protein gehört zu den einfachsten iterativen Typ I HR-PKS, da zum einem in diesem alle iterativen β -Modifikations-Domänen im aktiven Zustand vorzufinden sind und zum anderen grundlegende Programmierungen der Domänen zu beobachten sind. Wichtig zur Erforschung der Programmierung von HR-PKS sind detaillierte *in vitro* und stereochemische Untersuchungen.



In dieser Arbeit wurden verschiedene große Multidomän-Fragmente von SQTKS in *E. coli* kloniert und anschließend überexprimiert. Diese Fragmente wurden mittels geeigneter Aufreinigungsbedingungen isoliert und anschließend biochemisch als auch biophysikalisch analysiert.

Durch die Verwendung von Pantethin Substraten (Analoge des ACP gebundenen Substrates in SQTKS) können kinetischen Studien zur Reduzierung der Enoyl-Gruppe durch die ER Domäne des DH-KR Tetradomänen-Fragmentes durchgeführt werden. Diese zeigen, dass die Domäne im Stande ist, verschiedene Substrate umzusetzen und die Programmierung besitzt, wodurch die Reduzierung des Tetraketide Substrates nicht stattfinden kann.

Aufgrund von Domän-Domän Interaktionen im Multi-Domänen Fragment sowie in der einzelnen SQTKS ER Domäne^[75], können verschiedene stereoselektive Reduzierungen der Enoyl-Gruppe im *in-vitro* Assay beobachtet werden. Anhand des Tetradomänen Proteins wurde die bevorzugte Stereochemie der Keto Reduktion untersucht. Hierbei zeigt der Vergleich zwischen der mammalia Fettsäure Synthase (mFAS) und der SQTKS in Bezug auf die Stereoselektivität, dass diese identisch ist. Dies verstärkt die Hypothese, dass sich HR-PKS und mFAS von einem gemeinsamen Vorfahren ausgehend entwickelt haben.

Zur Untersuchung des Methyltransfer Mechanismus wurden *in vitro* Assays der isolierten SQTKS C-Methyltransferase (CMeT) Monodomäne vorgenommen. Des Weiteren wurde in dieser Arbeit die Effektivität der Verwendung von Panthetine Substraten als Analoga zum originalen ACP gebundenen Substrat untersucht. Diese Untersuchungen zeigen, dass für eine vollständige Aktivität des Methyltransfers von der SQTKS cMeT Domäne vermutlich weiterhin die Wechselwirkung mit dem Acyl-Carrier-Protein (ACP) vonnöten ist.

Acknowledgment

I would like to thank Prof. Russell Cox for his timely support and supervision throughout last four years. He is the best teacher I have ever met in my life. He inspires me how to do and think scientifically, always with great patience and concern. Thank you.

I would like to thank the Hans-Peter Braun group, especially Dr. Jennifer Senkler for all the protein analysis. She is always nice to me even though I push her quite often for the results.

I would like to thank the Teresa Carlomagno group, especially Dr. Luca Codutti for the MALS, assistance with ITC and the Thermo shift assays. I also thank Neha Dhimole for answering my questions about protein.

I would like to thank PD Carsten Zeilinger for his help in experiment and patient safety instruction.

I would like to thank the analytical department and the media kitchen staff, especially Dr. Jörg Fohrer, Monika Rettstadt, Dagmar Körtje, Katja Körner and Tjoven Ostermeier. You are always helpful and dedicated to offer us convenient research environment.

I would like to thank all past and current Cox group members. In particular, I thank Dr. Nina Duensing for welcoming me in Hanover and instructing me lab techniques; Dr. Elizabeth Skellam, for all your help and concern; Dr. Xiaolong Yang, for letting me know the traditions; Dr. Raissar Schor, for letting me know the German rules; Francesco Trenti, for his jokes and music teaching; Dr. Julianne Buschmann, Oliver Piech and Vjaceslavs Hrupins, for all your ideas in our protein group; Karen Lebe and Verena Hantke, for your friendly accompany in the office and lab.

Finally, I would like to thank my wife Shufang, my father Weidong, my mother Yaping and my sister Jing for their selfless real love and support.

Abbreviations and Units

ACP	acyl carrier protein	NAC	<i>N</i> -acetylcysteamine
AT	acyl transferase	Ni-NTA	nickel-charged affinity resin
CMeT	<i>C</i> -methyltransferase	NMR	nuclear magnetic resonance
CoA	Coenzyme A	NR	non-reducing
Da	Dalton	NRPS	non-ribosomal peptide synthetase
DH	dehydratase	PAGE	polyacrylamide gel electrophoresis
DNA	deoxyribonucleic acid	PAM	protospacer adjacent motif
EDTA	ethylenediaminetetraacetic acid	PANT	panthetheine
ER	enoyl reductase	PCR	polymerase chain reaction
ESI	electron spray ionisation	PKS	polyketide synthase
ELSD	evaporative light scattering detector	PR	partially reducing
EIX	equilibrium isotope exchange	PT	product template
FAS	fatty acid synthase	Q-TOF	quadrupole time-of-flight
HR	highly reducing	R	reductive release domain
KS	β -ketosynthase	SAM	<i>S</i> -adenosyl methionine
kb	kilo base pairs	SAH	<i>S</i> -adenosyl-L-homocysteine
K_M	Michaelis-Menten constant	SAT	starter unit acyl carrier protein transferase
KR	β -ketoreductase	SAXS	Small-angle X-ray scattering
LCMS	liquid chromatography mass spectrometry	SEC	size exclusion chromatography
m/z	mass to charge ratio	Sfp	4'-phosphopantetheinyl transferase
MALS	Multi-Angle Light Scattering	SQTKS	squalestatin tetraketide synthase
MS	mass spectrometry	TAE	tris-acetate-EDTA
MTA	5'-methylthioadenosine	TE	thiolesterase
MTAN	5'-methylthioadenosine/ <i>S</i> -adenosyl -L-homocysteine nucleosidase	TSA	thermo shift assay
NADPH	nicotinamide adenine dinucleotide phosphate	UV	ultra violet

Contents

1	Introduction	13
1.1	Polyketides.....	13
1.2	Biosynthesis of Polyketides	13
1.3	Polyketide Synthases	15
1.3.1	Modular Type I Polyketide Synthases	16
1.3.2	Iterative Type I Polyketide Synthase	18
1.3.3	Programming of Iterative PKS.....	20
1.4	Relationship between Mammalian FAS and HR-PKS.....	24
1.5	Stereoselectivity of PKS	27
1.5.1	Stereoselectivity of Acyl Transferase (AT).....	29
1.5.2	Stereoselectivity of Keto Synthase (KS).....	30
1.5.3	Stereoselectivity of Keto Reductase (KR).....	32
1.5.4	Stereoselectivity of C-methyltransferase (CMeT).....	36
1.5.5	Stereoselectivity of Dehydratase (DH)	38
1.5.6	Stereoselectivity of Enoyl Reductase (ER)	39
1.6	Aims of This Project.....	41
2	Strategy and Gene Cloning of SQTKS Multi-domain Fragments	44
2.1	Introduction	44
2.1.1	Strategy and Aims	44
2.1.2	Previous Multi-domain Expression of Iterative Type I PKS.....	45
2.1.3	Gene Cloning.....	47
2.2	Domain Boundary Determination.....	49
2.2.1	SQTKS CMeT	49
2.2.2	SQTKS Multi-domain Fragments.....	53
2.3	Codon Optimization	55
2.4	Expression Constructs.....	57

2.5	Conclusion.....	61
3	Expression and Purification of SQTCS Multi-domain Fragments.....	62
3.1	Introduction	62
3.1.1	Expression	62
3.1.2	Identification of Target Protein.....	65
3.1.3	Purification.....	66
3.2	Expression and Purification of SQTCS Fragments	67
3.2.1	SQTCS DH-ER Tridomain.....	67
3.2.2	DH-KR Tetradomain	75
3.2.3	Structure of SQTCS DH-KR Tetradomain by SAXS	80
3.2.4	CMeT Monodomain	83
3.2.5	<i>E. coli</i> Codon Sub-optimized DH-KR Tetradomain.....	84
3.2.6	DH-ACP Pentadomain	86
3.3	Optimization of Purification.....	87
3.3.1	Storage Buffer Optimization	88
	3.3.1.1 SQTCS DH-KR Tetradomain.....	88
	3.3.1.2 SQTCS CMeT.....	95
3.3.2	On-column SQTCS DH-KR Tetradomain Denaturation	96
3.3.3	Further Purification	99
	3.3.3.1 Impurities Analysis.....	99
	3.3.3.2 Methods to Remove Impurities	100
	3.3.3.3 The Structural Form of SQTCS DH-KR in Solution.....	103
3.4	Summary and Future Work	105
4	<i>In Vitro</i> Assay of SQTCS Multi-domain Fragments.....	108
4.1	Introduction	108
4.1.1	Previous <i>In Vitro</i> Studies of the SQTCS ER Monodomain	109

4.1.2	Previous <i>In Vitro</i> Studies of the SQTCS DH Monodomain	113
4.2	Aims	114
4.3	Stability of Different Substrates	115
4.4	Kinetic Studies of Enoyl Reduction	121
4.4.1	SQTCS DH-ER Tridomain.....	121
4.4.2	SQTCS DH-KR Tetradomain	122
4.5	Studies of Stereoselectivity in Enoyl Reduction by SQTCS DH-KR Tetradomain.....	124
4.6	Studies of Stereoselectivity in Keto Reduction by SQTCS DH-KR Tetradomain.....	127
4.7	Monitoring the β -processing Modification by LCMS of DH-KR Tetradomain.....	131
4.8	Studies of SQTCS CMeT Monodomain	133
4.8.1	Methyltransfer Observed by LCMS	134
	4.8.1.1 PANT Substrate –SQTCS CMeT Monodomain and DH-KR Tetradomain.....	135
	4.8.1.2 NAC Substrate - SQTCS CMeT Monodomain and DH-KR Tetradomain.....	137
	4.8.1.3 5'-methylthioadenosine/S-adenosyl-L-homocysteine Nucleosidase (MTAN)	138
4.8.2	Isothermal Titration Calorimetry (ITC) of Cofactor Binding	139
4.8.3	Rate of Epimerization	141
4.9	Conclusion.....	145
4.10	Future Work	147
4.10.1	To Improve Reactivity of Methyltransfer with Ferric Ion	147
4.10.2	The Importance of Acyl Carrier Protein (ACP)	148
5	Summary and Outlook.....	151
5.1	Summary	151
5.2	To Investigate the Remaining Stereochemical Questions in SQTCS.....	155

6	Experimental	158
6.1	Materials and General Protocols.....	158
6.2	Experimental for Chapter 2.....	163
6.2.1	Amplification of the Target Gene.....	163
6.2.2	Cloning.....	164
6.3	Experimental for Chapter 3.....	167
6.3.1	Expression Cell Transformation	167
6.3.2	Expression and Purification of SQTCS Fragments.....	168
6.3.3	Thermal Shift Assay (TSA).....	170
6.3.4	Protein Identification	171
6.4	Experimental for Chapter 4.....	171
6.4.1	Substrate Synthesis.....	171
6.4.2	Preparation of 4'R-[4'- ² H]-NADPD for KR reaction.....	171
6.4.3	Kinetic Studies of Enoyl Reduction Catalyzed by SQTCS DH-KR Tetradomain.....	172
6.5	Appendix	172
6.5.1	Protein Identification by ESI-QTOF-MS.....	172
6.5.2	Multiple Sequence Alignment.....	175
6.5.3	Synthesized <i>E. coli</i> codon Optimized Gene	181
6.6	Bibliography.....	186

1 Introduction

1.1 Polyketides

Polyketide natural products are secondary metabolites that are derived from the condensation of short-chain carboxylic acids. They are produced by a wide variety of organisms ranging from bacteria to plants, and from fungi to marine organisms. Polyketides feature great structural diversity, ranging from aliphatic, cyclic, alicyclic, aromatic components to macrocyclic lactones. The various structures of natural polyketide arise from the programming of polyketide synthases which use simple substrates originated from primary metabolites. Polyketides display useful activities as cholesterol-lowering drugs (such as Squalestatin-S1 **1** and Lovastatin **2**), antiparasitic agents (such as Avermectin **3**) and antibiotics (such as Erythromycin A **4**, Figure 1.1).^[1-2]

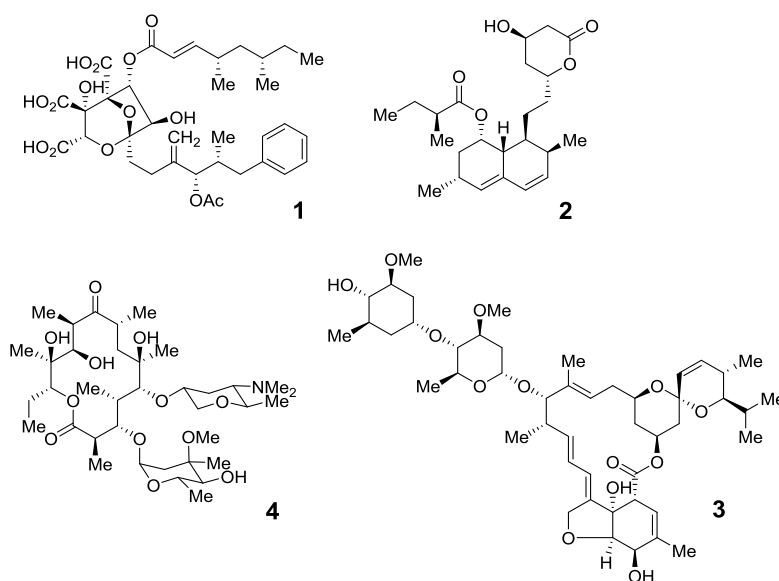


Figure 1.1. Some typical polyketide metabolites.

1.2 Biosynthesis of Polyketides

Polyketides are synthesized by proteins or protein complexes known as polyketide synthases (PKS). A PKS consists of several enzyme activities, including an acyl transferase (AT), a ketosynthase (KS) and an acyl carrier protein (ACP), which are essential. Polyketide

biosynthesis starts with the loading of the starter unit, such as the typical acetyl CoA **5**, to the AT domain as an ester. Afterwards, the acyl group of the starter unit transfers onto a cysteine residue of the KS domain as a thioester (Figure 1.2).

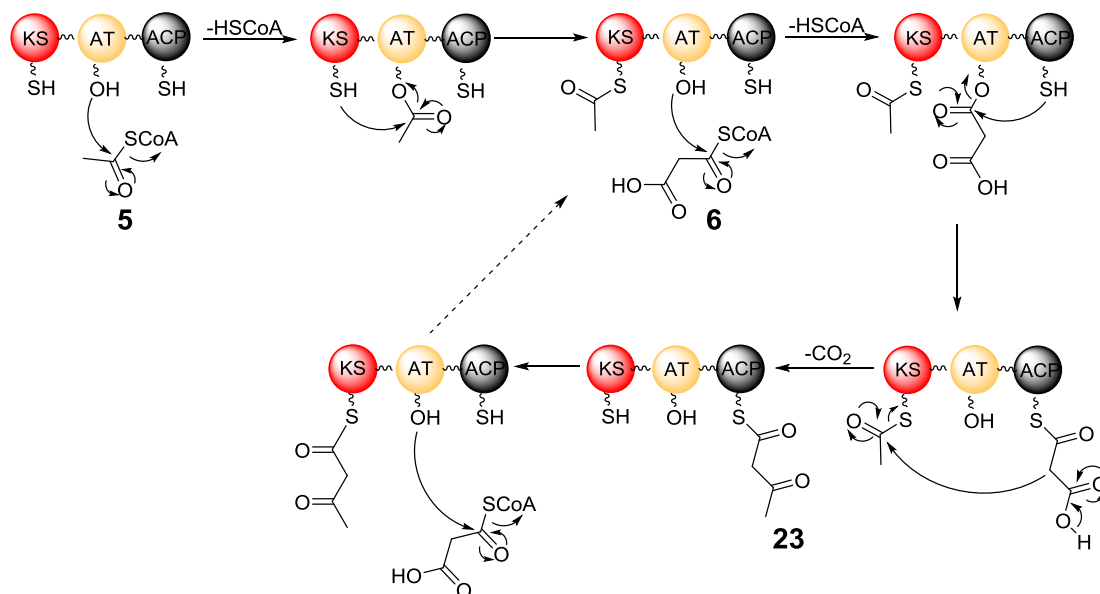


Figure 1.2. Chain elongation stage of polyketide biosynthesis.

Afterwards, an extender unit such as the typical malonyl CoA **6** comes into the reactive sites. Similar to the mechanism of acyl transfer, the malonyl group is linked to the ACP. The malonate then undergoes a decarboxylative Claisen condensation that gives an acetoacetyl unit **9** linked by the phosphopantetheine to the ACP.

After the chain extension, further modifications are performed by β -modification domains, such as *C*-methyl transferase (CMeT), ketoreductase (KR), dehydratase (DH), enoyl reductase (ER) and. *S*-adenosyl methionine **7** (SAM) is usually needed as the methyl donor in CMeT catalysis, and nicotinamide adenine dinucleotide phosphate **8** (NADPH, Figure 1.4) is needed as the hydride donor in the KR and ER catalysis. CMeT introduces a methyl group to the α position of **9**, KR reduces the β -ketone of **10** to form a hydroxyl group in **11**, which is then dehydrated by the DH domain. Finally ER reduces the α,β -unsaturated thioester **12** (Figure 1.3).

multi-domain megasynthases, that can be further divided into iterative type I PKS and modular type I PKS. Type II PKS consist of mono catalytic enzymes. The synthesized polyketides normally have a chain length of 16 to 24 carbons. Type II PKS are only found in bacteria so far. Type III PKS are simple homodimers of KS and do not use AT or ACP during the biosynthesis. Both Type II and Type III PKS are iterative.

1.3.1 Modular Type I Polyketide Synthases

The modular Type I PKS occur in bacteria and form assembly lines made from several active modules. Individual modules contain the essential extension domains (AT, KS and ACP) and variable manifests of β -processing domains. All the catalytic domains in each module are normally active. Subclasses of Modular PKS can be classified as *cis*-AT and *trans*-AT. The *cis*-AT modular PKS contains AT domains in every module. While in *trans*-AT, the AT domain is not present within the PKS multienzymes, but is an isolated protein which acts iteratively in *trans* to deliver the extender unit for polyketide elongation.

One of the most investigated *cis*-AT modular PKS is 6-Deoxyerythronolide B **108** synthase (DEBS), which synthesizes the backbone of erythromycin A **4**, an important polyketide used as an antibiotic. The biosynthetic route catalyzed by DEBS^[5] is shown in Figure 1.5. DEBS has three large proteins that can each be divided into several modules. In *cis*-AT modular PKS side chain α -alkyl group are installed by use of an α -alkyl extender such as methyl or ethyl malonyl CoA. In DEBS, the starter unit is propionyl CoA **43** and the α -alkyl extender is methyl malonyl CoA **44**. The elongation chain's structure diversity is formed by the different β -processing domains in different modules. After being cleaved and lactonized by a thiolesterase from PKS, the ring structure of 6-Deoxyerythronolid B **108** is formed. Then post-PKS-modifications take place by hydroxylases, methyl-transferases and glycosyl transferases to form the final compound **4**.

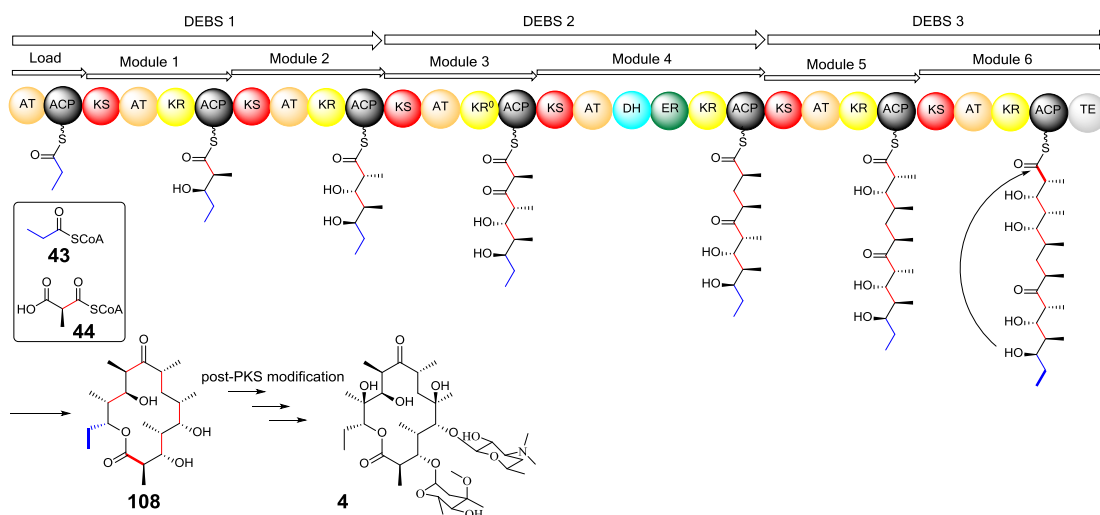


Figure 1.5. Biosynthesis of 6-Deoxyerythronolid B **108** by DEBS.

Compared to the *cis*-AT modular PKS, the *trans*-AT modular PKS are more complicated in domain organization and variation of extender units. *Trans*-AT PKS which make compounds such as mupirocin **67**^[6] and virginiamycin M **68**^[7] (Figure 1.6), contain modules which lack individual AT domains and the whole enzyme uses one or more *trans*-acting ATs. The α -alkyl substituents can be provided by either alkyl extender units or by methylation from SAM. Therefore, some *trans*-PKS contain active CMeT domains. In addition, some modules of this type of PKS also contain ACP-ACP di- or tri-domains or inactive KS-ACP domains, the functions of which are still under investigation.^[8]

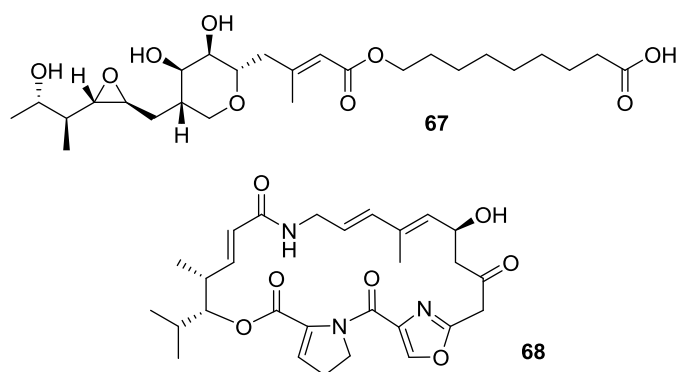


Figure 1.6. Mupirocin **67** and virginiamycin M **68**.

1.3.2 Iterative Type I Polyketide Synthases

Compared to the modular PKS, the iterative PKS contains only one module of catalytic domains which can be used repeatedly in several cycles during chain extension and β -processing. As a way to obtain the product of high diversity, every single β -processing domain can be “active” or “inactive” in different cycles. To better understand the behavior of the iterative type I PKS, the iterative type I PKS have been mainly classified into non-reducing PKS (NR-PKS), partially reducing PKS (PR-RKS) and highly reducing PKS (HR-PKS).^[3]

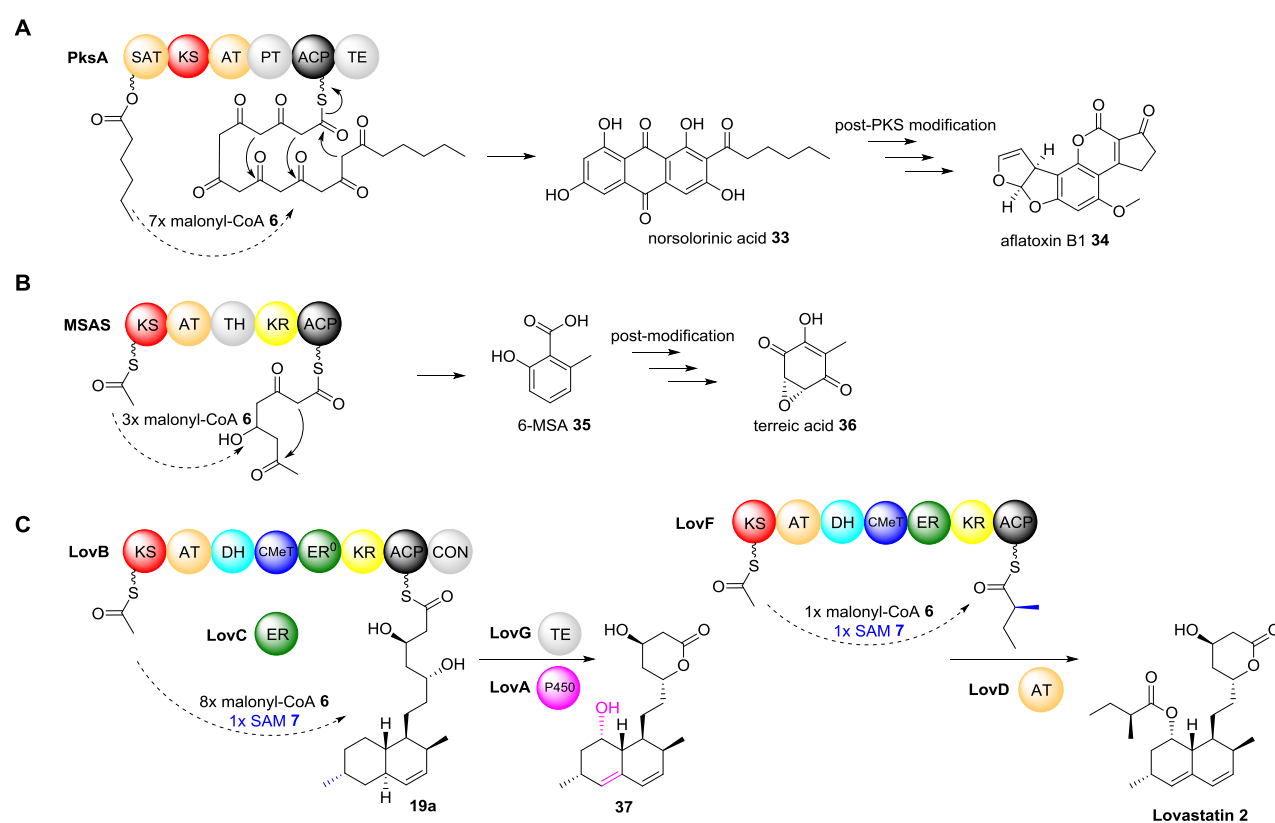


Figure 1.7. The biosynthesis catalyzed by NR-PKS, PR-PKS and HR-PKS.

The NR-PKS has no β -processing domains and catalyzes the synthesis of long unreduced polyketide chains. The polyketide chain is then cyclized through Claisen cyclizations catalyzed by the specific product template domain (PT).^[9] The polyaromatic product is released from the system, usually by a thioesterase (TE) domain or by a reductive domain (R). In the NR-PKS, ‘unusual’ starter units are generally incorporated by an additional SAT domain (starter acyl transferase). For example, during the biosynthesis of norsolorinic acid **33** (Figure 1.7 A), a precursor to aflatoxin B1 **34** (a common toxic contaminant in food), the norsolorinic acid synthase (NSAS)^[10] uses hexanoyl-CoA as the starter unit. The PKS then extends the hexanoyl unit by seven malonyl extender units, loaded by the AT domain. The formed polyketide is then cyclized through three Claisen cyclizations by the PT domain and released as norsolorinic acid **33** by the TE domain.

The PR-PKS generally contain a single KR domain to produce a partially reduced polyketide. It also contains a core domain. For example, the biosynthesis of 6-methylsalicylic acid **35** (6-MSA), the precursor to terreic acid **36** (a kinase inhibitor), is catalyzed by a PR-PKS named MSAS (Figure 1.6 B). MSAS contains a core domain which was previously thought to function as a DH and also as a TE domain. However, recent evidence has shown that no dehydration happens and the domain directly hydrolyzes a thioester bond of the tetraketide intermediate bound to ACP to release 6-MSA. Therefore, the core domain in MSAS was renamed thioester hydrolase (TH).^[11] This TH domain has been also found in other PR-PKSs such as the (*R*)-mellein PKS.^[12-13]

The HR-PKS generally contain all β -modification domains (KR, DH and ER) required to produce fully reduced polyketides. In addition they also often contain CMeT domains. The gene encodes HR-PKS are often found in gene clusters that also encode other iterative PKS. The ACP-linked product of HR-PKS can be used by other PKS as an advanced starter unit,^[14-15] which may explain why many HR-PKS lack a release domain. One common phenomenon in HR-PKS is the use of a *trans*-acting ER, such as found in the lovastatin nonaketide synthase (LNKS),^[17] tenellin synthetase (TENS) and desmethylbassianin synthetase (DMBS).^[19] Their own ER domains are inactive while a *trans*-acting ER works

well as a substitute. However, some HR-PKS still have active ER domains, such as lovastatin diketide synthase (LDKS)^[17] and squalenone synthase (SQTKS).^[17] For example, the synthesis of lovastatin **2** (Figure 1.6 C), a well-known cholesterol-lowering drug, involves two highly reducing PKS - the *trans*-ER PKS encoded by *lovB* (LNKS) and the *cis*-ER PKS encoded by *lovF* (LDKS). The LovB PKS is unusual as it contains parts of a nonribosomal peptide synthetase (NRPS) module including a condensation (CON) domain and part of an adenylation (A) domain (catalytic domain missing). The function of the CON domain here is not clear but important for the whole LNKS activity.^[16-17] LovB catalyzes the biosynthesis of the nonaketide **19a** which is cyclized by an intramolecular Diels-Alder reaction. Afterwards, an individual thioesterase (encoded by *lovG*) catalyzes the release of **19a**, followed by ring closure by lactonization. Then a P450 (*lovA*) catalyzes a double hydroxylation, after which one of the hydroxyls is eliminated to obtain the conjugated diene **37**.^[18] The simple LovF catalyzes the synthesis of an ACP-bound methylated diketide which is then transferred onto the diene **37** by an AT (encoded by *lovD*) to form lovastatin **2**.

1.3.3 Programming of Iterative PKS

PKS are programmed, which involves starter selection, chain-length control and control of cyclisation, and in addition involves the fact that not all the β -processing domains are always active after the chain elongation. This brings abundant diversity into the final product. Modular PKS use several modules with different set of domains to fulfill the programming. In comparison, the mechanism of programming of iterative PKS is still unknown.

To understand the programming of iterative PKS, domain swapping has been used among iterative PKS. According to the rational domain swaps experiment between the TENS and DMBS, the Cox group^[19] recently reported the first research into the programming of HR-PKS (Figure 1.7). TENS and DMBS are highly similar (~97%) in the functional domain regions. They are same in domain arrangement and both use *tran*-acting ER enzymes. Their products, pretenellin A **38** (TENS) and preDMB A **39** (DMBS), are highly similar in structure units, except from the difference of side chain length and methylation pattern. In addition to

the HR-PKS, TENS and DMBS also contain an NRPS which introduces a tyrosine into the product of HR-PKS to create a pyrrolidinone in the structure of **38** and **39**. Initial tests have shown that the NPRS and the *trans*-acting ER between the TENS and DMBS are interchangeable and have no effect on the structure of their products.

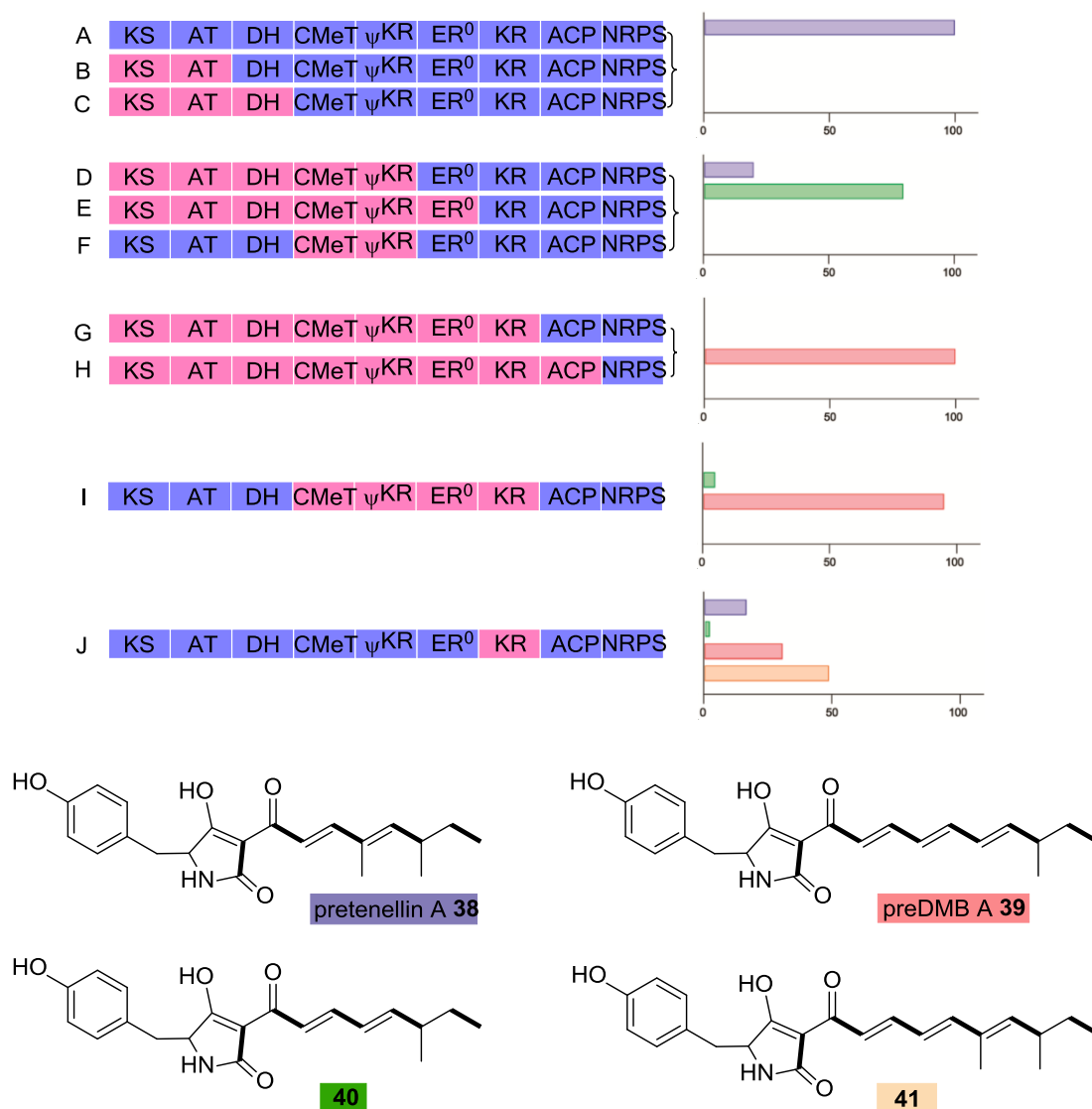


Figure 1.8. Rational domain swapping between TENS and DMBS. The *trans*-acting ER (not shown in the figure) and the NPRS are interchangeable with no programming effect.

By comparing domain swaps A-F (Figure 1.8) and corresponding products petenellin A **38** and desmethyl petenellin A **40**, it is easily presumed that the methylation pattern is controlled by the CMeT domain. Furthermore, by comparing domain swaps E and G-I (Figure

1.8) and corresponding products preDMB A **39** and desmethyl pretenellin A **40**, the chain length of product appears to be controlled by the KR domain. However, if swapping single KR domain from DMBS to TENS, separated from ER⁰, as shown in J (Figure 1.8), the result is complicated (new product **41** forms). It is assumed that the connection between KR and ER⁰ is important to perform reasonable activity. This is similar to the phenomenon observed in PR-PKS,^[20] any disruption between domain-domain interface may cause a loss of function. Meanwhile, in this research, ACP, KS, DH, AT and inactive ER⁰ are not related to the formation of methyl group and determination of chain length. The KS is thought to control the chain length in NR PKS.^[21] While in HR PKS, at least for TENS and DMBS, the chain length appears to be controlled by the KR domain.

An alternative strategy for investigating the programming of PKS enzymes is to study them *in vitro*. This involves protein isolation and monitoring the enzyme activity towards different substrates. In this way the detailed selectivities of specific functional domains of the PKS can be studied.

Recently, Vederas and Tang^[22] suggested that an interesting kinetically controlled mechanism evolved in the programming of HR-PKS. 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 CMeT → KR → DH → ER. In the case of the HR-PKS, LNKS encoded by *lovB*, which partners with a *trans*-acting ER encoded by *lovC*, the CMeT only recognizes substrate **13** while excluding all of the other substrates completely (Figure 1.8 B). They suggested that the outcome of the modifying steps is determined by the relative activity of each domain towards specific substrates. For the substrate **13**, the CMeT domain shows higher activity relative to the KR domain, therefore the CMeT domain functions first, and the KR domain acts later. However, for substrate **17**, the KR domain's activity is higher than the activity of the CMeT domain, therefore substrate **17** is first reduced by the KR domain, meaning that subsequent methylation is impossible. To certify the assumption, comparisons between individual rates of isolated CMeT and KR domains toward the different substrates have been made (Figure 1.8 A).

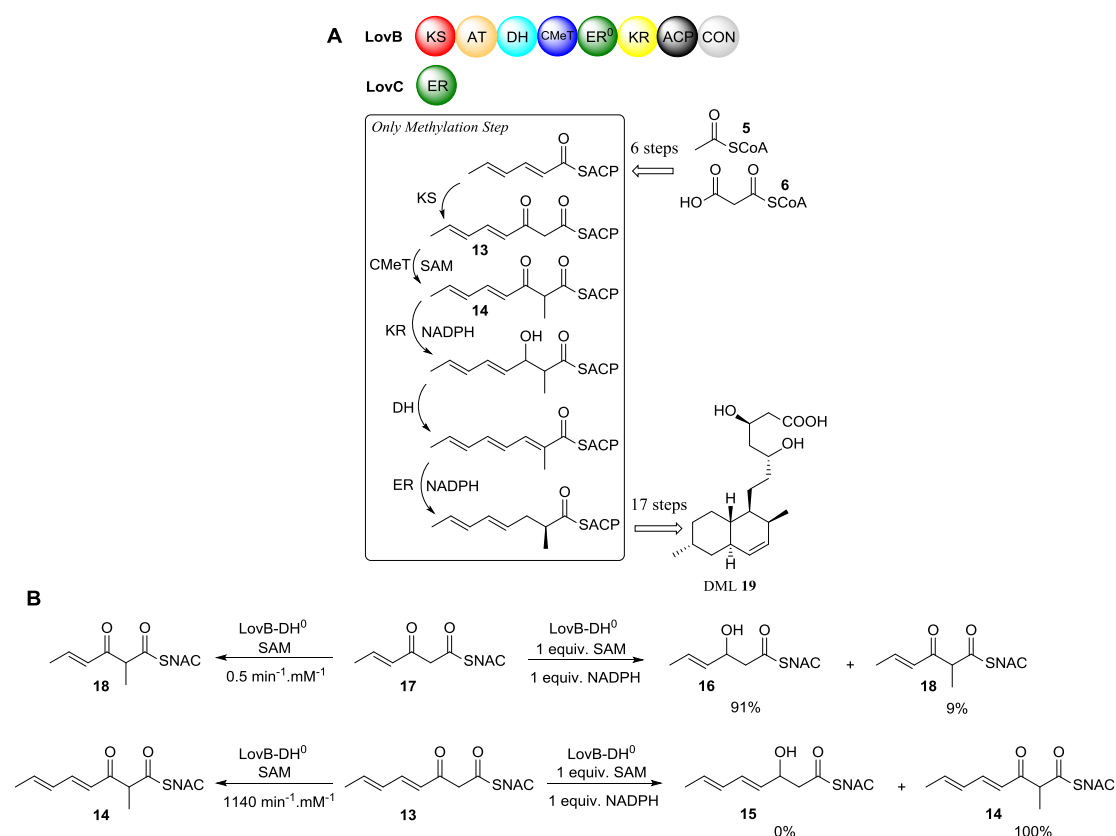


Figure 1.8: The suggested kinetically controlled programming of the HR-PKS LNKS. *lovB* encodes the HR-PKS and *lovC* encodes the *trans*-acting enoylreductase. CON refers to the NRPS-like condensation domain. **A**, the only methylation step of LNKS in the synthesis of dihydromonacolin L (DML **19**); **B**, kinetic analysis of the LNKS CMeT domain towards different ketoacyl-SNAC substrates. Compounds **13** and **17** are both natural product analogs.

To allow for 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 of LNKS to yield LNKS-DH⁰. The group also synthesized several acyl-*N*-acetylcysteamine (NAC) substrates which were analogous to the natural substrates, where the SNAC group acts as a substitute for the ACP domain. By expressing LNKS 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 CMeT domain activity to substrate **13** is more than 2000 times higher than substrate **17** (Figure 1.8 B). However, the KR domain exhibits far less

selectivity in reactivity to these substrates. This may suggest an important role of the CMeT domain as a gatekeeper domain in the programming of LNKS.

To determine if there is indeed competitive catalysis between the KR and CMeT domains towards the 3-ketoacyl substrates, Tang and co-workers performed a combined CMeT/KR assay. Both the natural substrates **13** and **17** were added to LNKS (with the inactive DH domain) in the presence of the same molar concentration of SAM and NADPH, and the yields of the methylation product and the ketoreduction product were compared (Figure 1.8 B). For substrate **17**, they observed a 10:1 ratio of KR-catalyzed product **16** to CMeT-catalyzed product **18**. However, for substrate **13**, only the methylated product **14** was observed and no reduced product **15** could be detected, thereby confirming the much higher catalytic efficiency of CMeT domain toward substrate **13** compared to that of 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 **14**, 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.

1.4 Relationship between Mammalian FAS and HR-PKS

Currently there is no available crystal structure of a HR-PKS meaning the three dimensional organization of catalytic domains and their linker regions is unknown. For efficient domain isolation and *in vitro* research into these complex biosynthetic machines, an accurate structural model is required. The mammalian fatty acid synthase (mFAS) is highly similar to HR-PKS in terms of the iterative mode of catalysis and the homologous domain organization, despite sharing only ~20% sequence identity. Phylogenetic analysis indicates that the relationship between mFAS and HR-PKS is closer than the relationship between mFAS and modular PKS.^[23] Both HR-PKS and mFAS (Figure 1.10) use the same starter unit (acetyl-CoA) and extender unit (malonyl-CoA) for chain initiation and extension and perform similar modification steps. The major differences are first, mFAS (and all other types of FAS)

have no programming, meaning that all the β -modification domains (*e.g.* KR, ER, DH) are always active. Second, there is an inactive pseudo-CMeT (ψ CMeT) domain in mFAS while most HR-PKS have an active CMeT domain. Third, the releasing mechanism in mFAS, catalyzed by the thioesterase (TE) domain is obvious, while the release mechanism is not always clear in HR-PKS.

The structure of mFAS (PDB 2VZ8)^[24] has been resolved as an X-shaped homodimer with dimensions of approx. 200 Å × 170 Å × 110 Å (3.2 Å resolution, Figure 1.9 A). The positions of the ACP and TE domains were not resolved in this structure possibly due to their high flexibility and mobility.^[25] The structure consists of two regions: the “condensing region” (KS-AT) and the “modifying region” (DH- ψ CMeT- ψ KR-ER-KR). These are connected between the AT domain and the DH domain by a post-AT linker which winds around the KS domain (Figure 1.9 B). The N-terminal KS domain is linked to the AT domain *via* a small linker domain (LD). The dimeric DH domain is mainly stabilized by the dimeric ER domain. Close to the DH domain, the monomeric KR domain is stabilized by a non-catalytic domain called the pseudo-KR (ψ KR) domain. At the outer edge of the modifying region there is a ψ CMeT connected to the DH domain by a post-DH linker which winds around the KR domain.^[26] For the whole structure of mFAS, the condensing region and the modifying region can swing, rotate and stretch around the central connection according to the EM analysis. During the stretching movement, little structure variation was observed in the condensing region, while considerably large variation was observed in the modifying region. The relationship between the enzyme catalysis and the conformational variation is still unknown, which may affect substrate transportation by the ACP domain around the modification domains.

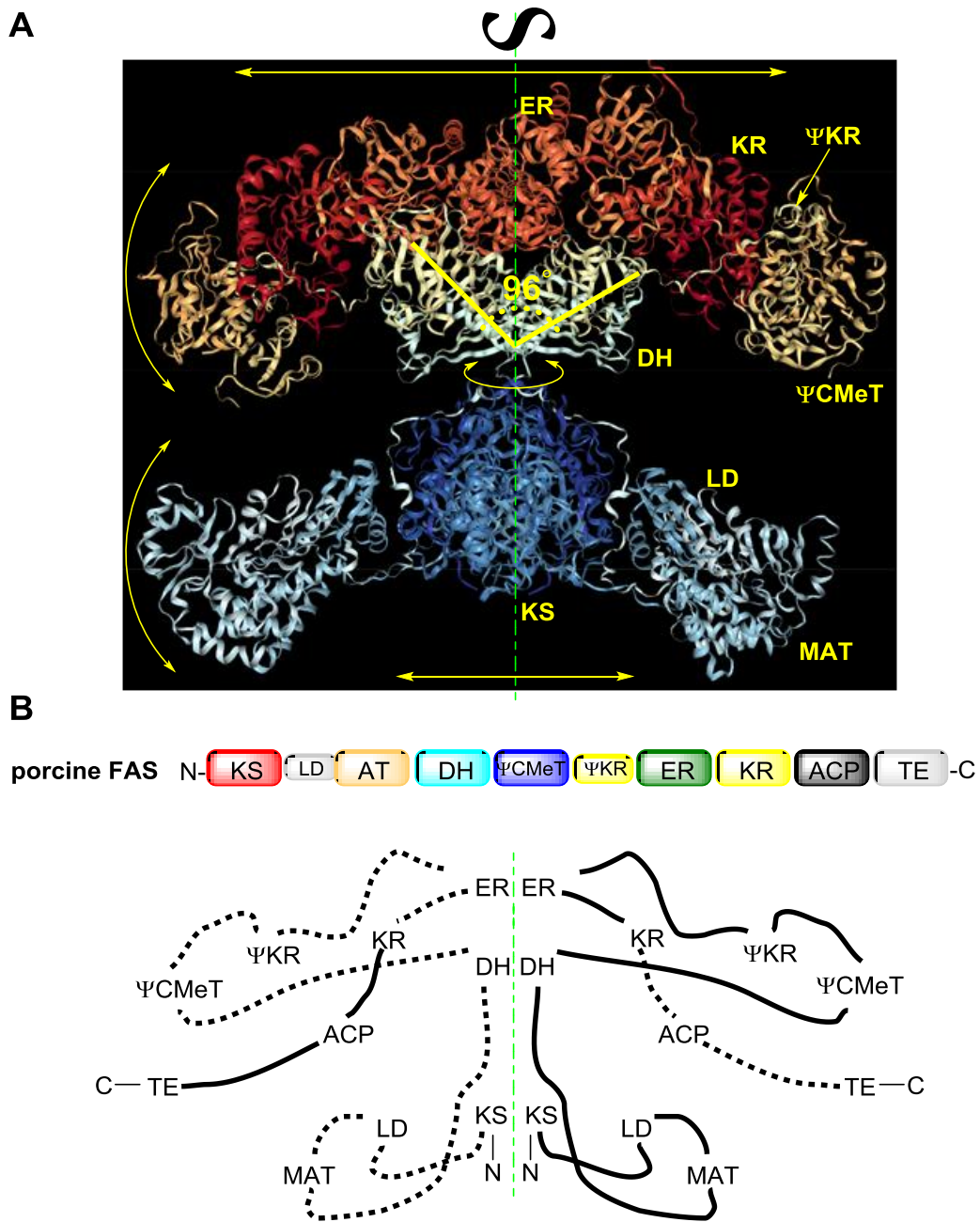


Figure 1.9. Structural overview of mFAS: **A**, cartoon^[68] representation demonstrates the “X” shape in the front view and the “S” shape from the top view. The movements of the architecture are shown as yellow double ended arrows; **B**, linear sequence organization of mFAS and the rough two-dimensional connections among domains.

In the modifying region of mFAS, the two DH domains are arranged in a V-shape with an angular separation of 96 °, whereas in the modular PKS DEBS and other dimeric PKS the DH domains are separated with angles ranging between 170-220 °.^[27-31] In many PKS an additional N-terminal β - α - β - α extension of the ψ KR is observed, however this is not present in mFAS.^[31-34] These comparisons of mFAS and PKS indicate that the modifying region architecture is not highly conserved. In contrast, the structure and active sites of the condensing region is highly conserved between mFAS and the reported structures of modular PKS^[35-36] (except for the PikAIII^[37]), MAS-like PKS (mycocerosic acid synthase)^[31] and NR-PKS.^[38] Among them, the iterative-acting MAS-like PKS^[39] contains the fully reducing modifying region which is similar to mFAS in domain organization and crystal structure.^[31] However, these catalytic domains are more closely related to modular PKS (~27-35% sequence identity) than to mFAS (~17-19% sequence identity) or HR-PKS (~20-22% sequence identity).^[31] In addition, unlike mFAS and HR-PKS, MAS-like PKS lack the ψ CMeT/CMeT domains as they are able to use methylmalonyl extender units and large starter units.^[40] It is not even sure what chemistry is actually catalyzed by MAS, whereas mFAS is very well understood at the chemical level. These features make the crystal structure of the MAS-like PKS less valuable as a reference structure for HR-PKS. Therefore, the structure of mFAS is still the most enlightening model to study the architecture of HR-PKS.

1.5 Stereoselectivity of PKS

The biosynthesis of polyketides proceeds with exquisite stereocontrol. 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 genetic engineering. Due to the close relationship between PKS and mFAS, the stereopreferences in mFAS are instructive. During the 1990s, Vernon Anderson and Gordon Hammes systematically studied the stereochemical control of each domain in mFAS (Figure 1.10) by using specific isotope labeling methods and enzymes.^[41] It was observed that the stereo-chemistries in fatty acids are strictly controlled during the biosynthetic process

catalyzed by mFAS. For example, the stereochemistry at C-2 of the malonyl-CoA **6** inverts during chain extension; the KR uses the 4'-*pro-S* hydride of NADPH to attack the 3-*Si* face of **23**; and *syn*-elimination of **20** occurs in the DH domain, forming the enone product **21** in the *E*-configuration.^[41] Similarly, the ER domain uses the 4'-*pro-R* hydride of NADPH to attack the 3-*Re* face of **21** and a proton, provided by water or protein residues, attacks from the 2-*Si* face to form the fully reduced product **22**.^[41]

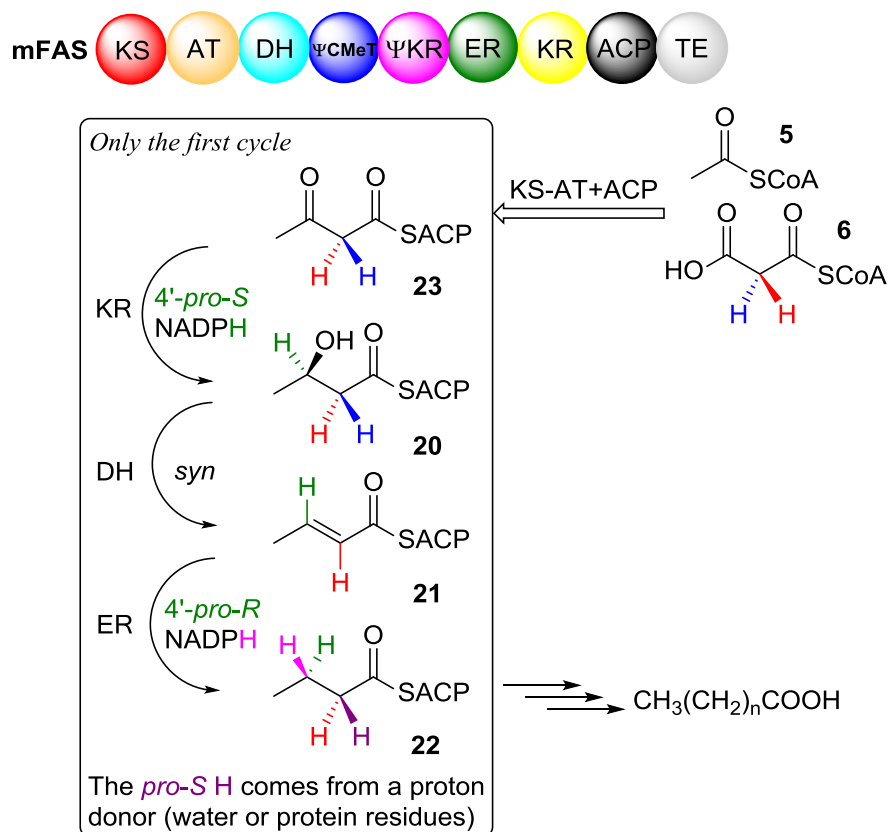


Figure 1.10. The strict stereochemical control in mammalian fatty acid synthase.

The following sections describe the current state of knowledge in stereocontrol during the key steps of the biosynthetic pathways of polyketides. In this area, far more research has been conducted into the Type I modular PKS than Type I iterative PKS. However, methods used for determining the stereo-preference of each domain in modular PKS provide inspiration for the analogous iterative domains in iterative PKS.

1.5.1 Stereoselectivity of Acyl Transferase (AT)

In modular PKS evidence has been gathered that shows that AT domains can select the extender unit with specific stereo-configuration. For example, AT domains in DEBS, the typical *cis*-AT PKS, only use 2*S*-methylmalonyl-CoA **44** as the extender unit. This was confirmed by *in vitro* studies of the recombinant DEBS1-TE protein.^[42] The protein was created by joining the terminal TE domain to the end of DEBS1 (Figure 1.11). In this way, the polyketide can be released at the triketide stage to form the lactone **42** which is experimentally traceable. In the presence of propionyl-CoA **43** (starter unit), 2*S*-methylmalonyl-CoA **44** (extender unit) and NADPH (KR cofactor), the product **42** was obtained at a satisfactory rate. However, when 2*R*-methylmalonyl-CoA was provided instead, no product was observed. This means that the system only uses 2*S*-methylmalonyl-CoA **44** as the extender unit and the methyl stereo-configurations in the product are not related to the corresponding extender units. A crystal structure of AT-5 from the DEBS PKS^[43] has shown that steric hinderance may be the reason why 2*R*-methylmalonyl-CoA cannot be used as the extender unit. Due to the high level of homology among many AT domains from *cis*-AT PKS, it is possible that all such acyl transferases exhibit the same stereoselectivity.^[44] For iterative PKS, the AT domain generally uses malonyl-CoA as the extender unit. Therefore, no C-2 stereochemistry is involved in iPKS AT domains.

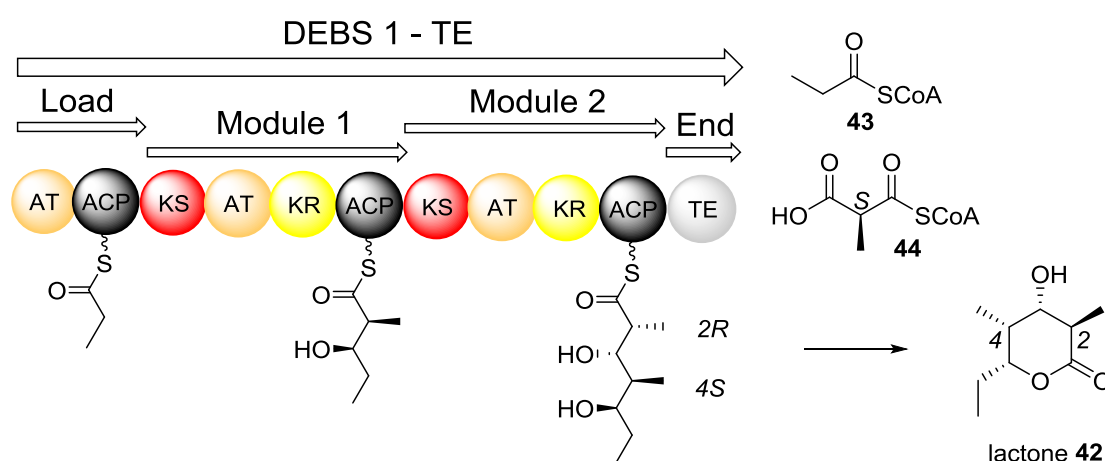


Figure 1.11. Creation of DEBS1-TE and the stereochemistry of the triketide lactone **42** can be observed.

1.5.2 Stereoselectivity of Keto Synthase (KS)

The extender unit selected by the AT domain is transferred to the ACP domain and forms an enolate generated by decarboxylation (Figure 1.13 A). The enolate attacks the starter unit, which is attached to the active site cysteine of the KS domain, and an extended polyketide is formed. The face of the enolate used for the nucleophilic attack determines whether the reaction occurs with retention or inversion of the methyl group configuration at the C-2 center. In animal FAS, the reaction results in the inversion of the stereochemistry at the C-2 center of the extender unit. ^[45]

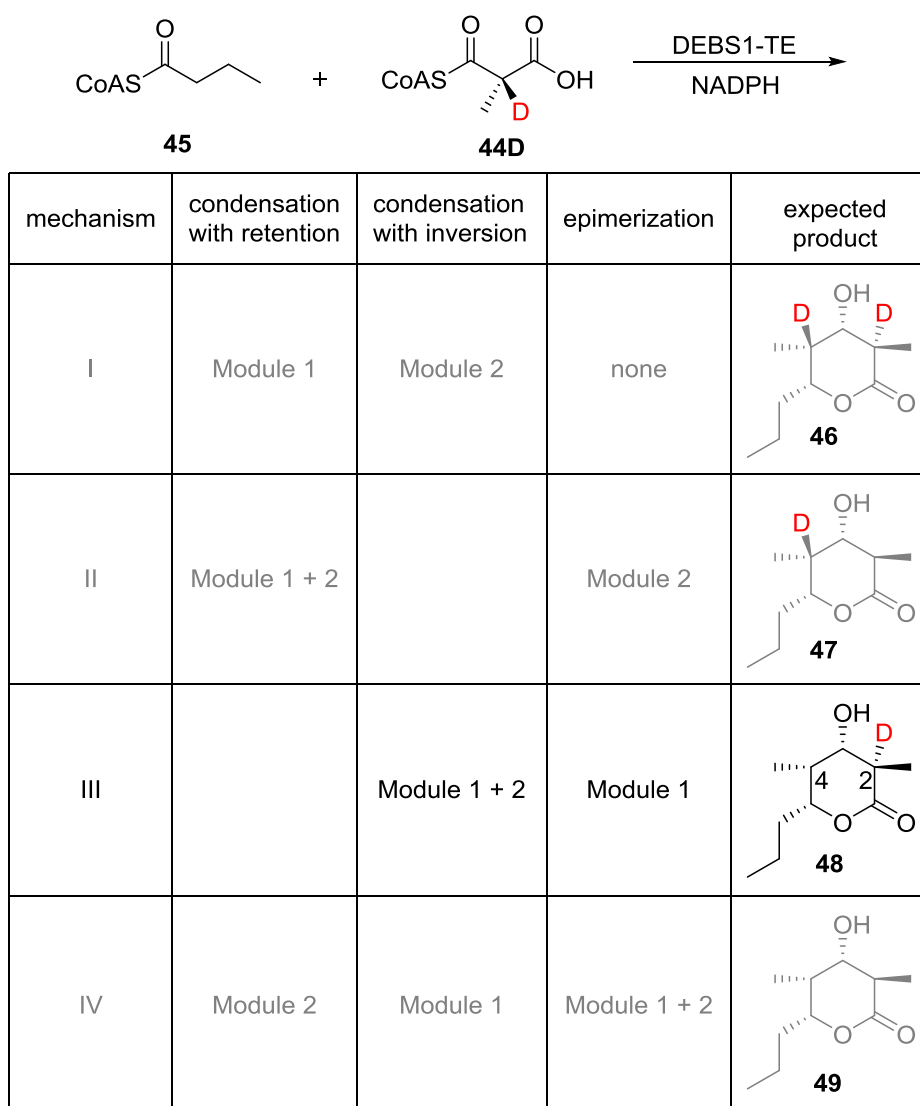


Figure 1.12. The relationship between the mechanism (retention or reversion of the stereochemistry of extender unit and methyl group epimerization) and the corresponding product.

In vitro experiments with DEBS1-TE have provided evidence that this inversion of stereochemistry also occurs in PKS. ^[46] In the study the extender unit (2*S*)-(2-²H)-methylmalonyl-CoA **44D**, the starter unit butyryl-CoA **45** and NADPH were provided to DEBS1-TE and only the lactone **48** having a single deuterium labeled at C-2 was produced (Figure 1.12). Only one mechanism can explain the formation of the methyl stereo-configuration at the C-2 position of **48**: inversion of the stereochemistry of the extender unit catalyzed by the KS domain in both module 1 and module 2. This must be followed by an additional epimerization in module 1 to explain the methyl stereo-configuration and the loss of deuterium at the C-4 position of **48**. Other potential mechanisms cannot explain the formation of the observed products (shown grey in Figure 1.12).

In vitro studies of the recombinant DEBS diketide synthase where the KS domain from module 2 was replaced by the KS domain from module 1 confirmed that the KS domain in module 1 has no epimerization activity. ^[47] The protein was created by fusing the DEBS1 loading module (AT-1 + ACP-1) and the KS₁ domain with catalytic domains from the second module of DEBS1 (AT-2 + KR-2 + ACP-2) with the terminal TE domain (Figure 1.13 B). In the presence of the starter unit propionyl-CoA **43**, the extender unit 2*S*-methylmalonyl-CoA **44** and the KR cofactor NADPH, the diketide **50A1** was obtained. The KS-1 domain can invert the C-2 stereochemistry of the extender unit; if it could also epimerize the C-2 methyl configuration, the recombinant DEBS1 diketide synthase should produce a diketide with *S*-configuration at the C-2 center of the diketide. However, the *R*-configuration was observed in the diketide product of this experiment. This means the epimerization does not arise from the action of the KS-1 domain, and so must be a result of the other domains within the DEBS protein. The epimerization activity of KR domains will be discussed in the next section.

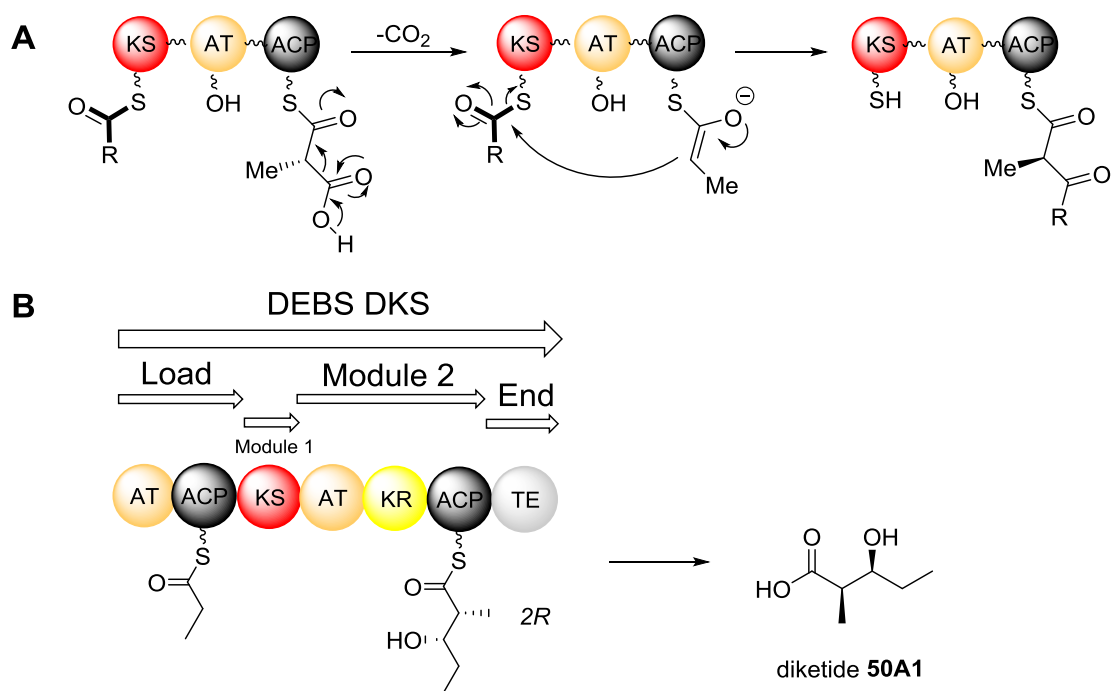


Figure 1.13. **A**, mechanism of the chain extension catalyzed by the KS domain; **B**, creation of DEBS DKS and the stereochemistry of the diketide product **50A1** can be observed.

For iterative PKS, no studies into the stereochemistry of KS domains have been reported. Due to the high similarity between mFAS and HR-PKS, it is generally assumed that the KS domains of HR-PKS probably also have the ability to invert the C-2 configuration of malonyl-CoA **1** during the chain elongation.

1.5.3 Stereoselectivity of Keto Reductase (KR)

KR domains catalyze the stereoselective reduction of C-3-ketone groups to form specific stereoisomers of the hydroxyl groups in the polyketide chain. By incubation of chirally deuterated NADPH molecules with different modules in DEBS and analysis of the resultant products by MS, it was shown that all the KR domains in DEBS transfer the 4'-*pro-S* hydride of the NADPH cofactor, as observed in fatty acid biosynthesis.^[47-48] All these KR domains possess the active site tetrad^[49] of Tyr, Ser, Lys and Asn, which is the characteristic motif of the short-chain dehydrogenase/reductase (SDR) superfamily.^[50-51] The 4'-*pro-S* hydride of

the NADPH can attack the C-3-carbonyl from its *Re* face or its *Si* face to form A-type products and B-type products respectively (Figure 1.14 A). In this way, the corresponding KR domains can be separated as A-type KR- and B-type KR domains.^[52]

In modular PKS, the KR domains also influence the 2-methyl stereo-configuration of the polyketide. The first direct proof was obtained by studies on reconstituted modules consisting of combinations of individually purified KS-AT didomains, ACPs and KR domains.^[53] A mixture of KS-AT didomains, KR domains, propionyl-SNAC, *holo* ACP and (2*S*)-methylmalonyl-CoA **44** gave the diketide products (Figure 1.14 B). The products were hydrolyzed from the ACP domains and their stereochemistries were determined by chiral GC-MS. Reconstituted DEBS modules 1 and 6 gave predominantly the expected products in the biosynthesis of 6-deoxyerythronolide B **32a**; a B-type diketide with 2*S*-methyl configuration for module-1 and an A-type diketide with 2*R*-methyl configuration for module 6. Mixing the KS-AT and ACP domains from DEBS module-6 with the KR domain from module-1 KR yielded the B-type diketide **50B2** with 2*S*-methyl configuration. This proves that the KR domain from module-1 is a B-type KR and it also has the epimerization activity which the KS domain lacks (section 1.4.2). Conversely, mixing the KS-AT and ACP domains from DEBS module-1 with the KR domain from module 6 yielded an A-type diketide **50A1** with *R*-methyl configuration. This indicates the KR domain from module-6 is an A-type KR and it may have no epimerization activity.

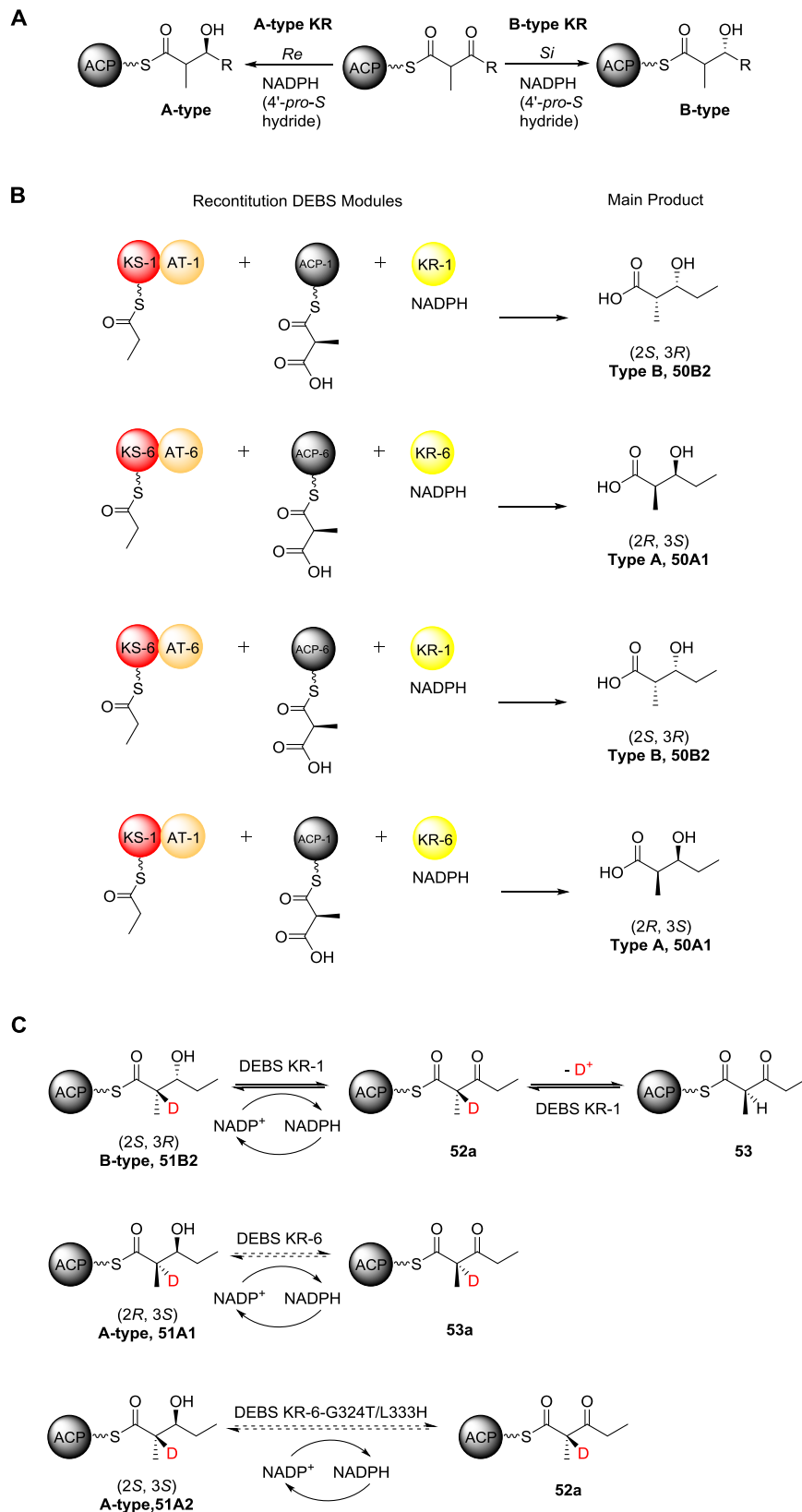


Figure 1.14. *In vitro* research investigating the function of the KR domain: **A**, A-type KR and B-type KR; **B**, reconstitution assay; **C**, assay of equilibrium isotope exchange (EIX).

Further direct proof of the epimerization activity of the KR domain comes from a series of so-called “equilibrium isotope exchange (EIX)” experiments^[54] (Figure 1.14 C). In this assay, stereo-configurationally stable KR substrates with a C-2 deuterium label were synthesized (*i.e.*, [2-²H]-2-methyl-3-hydroxypentanoate), and the substrate was linked to a DEBS ACP domain to form the ACP-bound substrate **51**. By incubating the ACP-bound substrate with the KR protein and NADP⁺, if the KR domain is capable of epimerization, then the redox reaction can process in reverse and form an equilibrium between the oxidized **52** and reduced diketide **51**. This can be reflected by observing a first-order time-dependent washout of the deuterium by LC-MS analysis. It was observed from this assay that the KR domain from DEBS module 1 is capable of epimerization (Equilibrium between **52a** and **53**), while the KR domain from DEBS module 6 is not capable of epimerization (no washout of deuterium). These results are in accordance with the previous reconstitution experiment (Figure 1.14 B).

Therefore, the KR domain can control the stereochemistry, not only of the C-3 hydroxyl, but also of the C-2 methyl configuration. A possible explanation is that the KR domain can epimerize the diketide to form two C-2 methyl isomers and only one isomer can be reduced by a specific KR domain.^[55] Even though the detailed mechanism of the relationship between reduction and epimerization is still not clear, a double mutant DEBS-KR-6-G324T/L333H has been created to revert its native reduction selectivity from (2*R*)- to (2*S*)-2-methyl-3-ketoacylthiolester substrates (**53a** to **52a**) and produce the (2*S*, 3*S*)-2-methyl-3-hydroxylacyl ACP **51A2**.^[56] In addition, the redox-inactive DEBS KR⁰-3 still has the ability of epimerization according to the EIX experiment.^[170]

For iterative PKS, research into the SQTKS DH domain^[57] has inferred that the KR domain may be a B-type KR and produce only (2*R*, 3*R*)-2-methyl-3-ketoacylthiolester, which is the same stereopreference at the C-3 centre as the mFAS KR domain.^[41] It is still unknown which 4'-hydride of the cofactor NADPH is transferred by the KR domain. In addition, the single KR of an iterative PKS must process several different substrates of differing chain-length, and few studies have probed this level of selectivity.

1.5.4 Stereoselectivity of C-methyltransferase (CMeT)

The stereopreference of CMeT domains has been rarely studied, partly since modular PKS do not have CMeT domains. It is difficult to infer the stereopreference of the CMeT from the structure of the final polyketide products since the configuration of the initially formed 2-methyl-3-ketoacyl ACP intermediates are obscured by spontaneous epimerization or by subsequent chain modifications such as the KR-catalyzed epimerization and the DH-catalyzed dehydration.

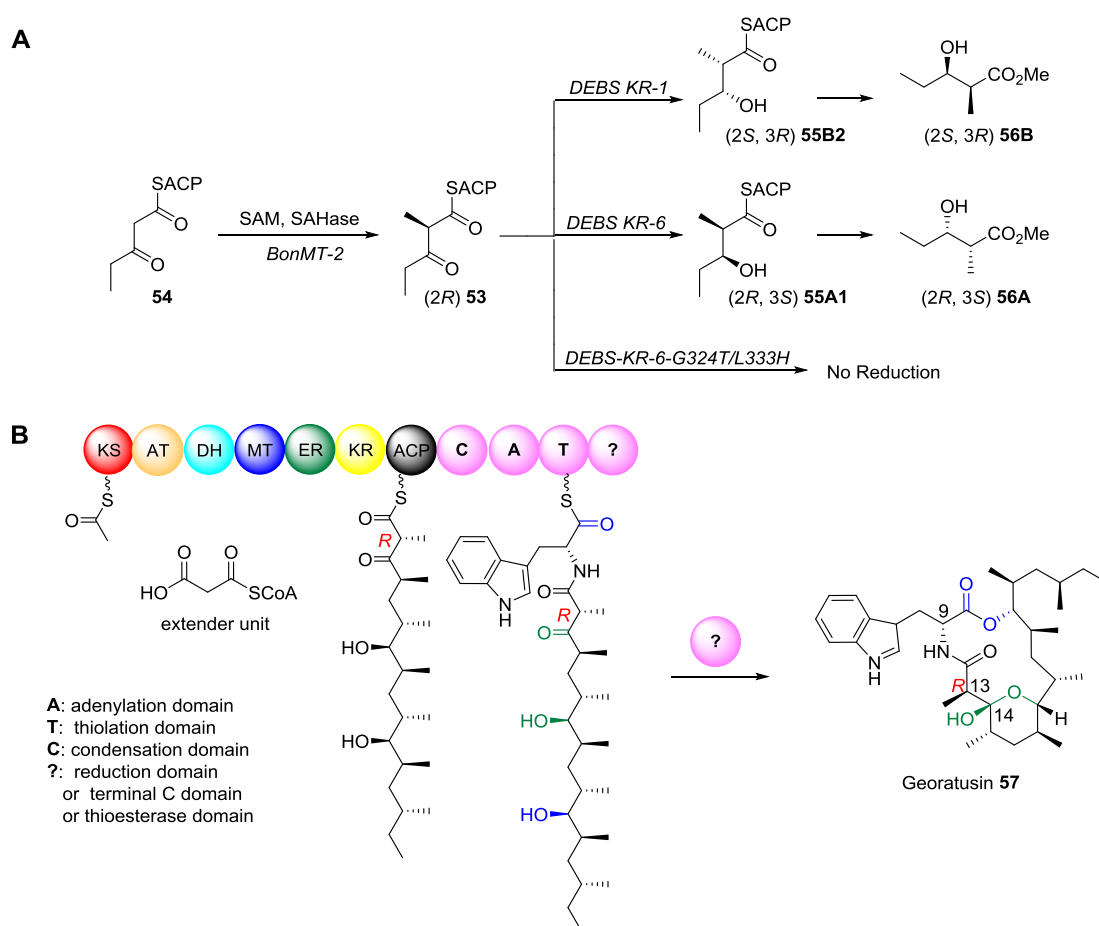


Figure 1.15. **A**, detecting the stereopreference of CMeT in modular PKS by *in vitro* assay with the combination of CMeT and epimerase-active/inactive KR; **B**, proposed biosynthesis of georatusin 57.

The existence of epimerase-active KR and epimerase-inactive KR domains was previously discussed (section 1.5.3).^[54] An *in vitro* assay with combinations of CMeT and KR domains revealed that some CMeT domains from *trans*-AT modular PKS are able to catalyze the formation of the (2*R*)-2-methyl-3-ketoacyl-ACP product **53** exclusively (Figure 1.15 A).^[58] For example, BonMT-2 from module 2 of the bongkreki acid PKS was isolated and was incubated with 3-ketopentanoyl-BonACP-2 **54**, SAM and SAH nucleosidase to produce the methylated product **53**.^[58] After 15 minutes incubation at room temperature, the reaction mixture was added to a solution of epimerase-active KR- or epimerase-inactive KR domains and NADPH. The resulting reduced product was hydrolyzed and converted into 2-methyl-3-hydroxypentanoate **56** with trimethylsilyldiazomethane and the stereochemistry of the resulting ester at C-2 and C-3 was confirmed by chiral GC-MS analysis. The results obtained indicated that, reduction with DEBS KR-6, an epimerase-inactive, B-type KR domain exclusively synthesized (2*R*,3*S*)-2-methyl-3-hydroxypentanoate **56**, while DEBS KR-1, an epimerase-active, A-type KR domain gave the corresponding (2*S*,3*R*)-2-methyl-3-hydroxypentanoate **55**. In contrast, the epimerase-inactive and (2*S*)-2-methyl-3-ketoacyl-ACP-specific mutant ketoreductase, DEBS-KR-6-G324T/L333H, could not produce 2-methyl-3-hydroxypentanoate. These results are in accordance with the product specificity of KR^[54] and demonstrate that BonMT-2 is stereoselective for (2*R*)-2-methyl-3-ketoacyl-ACP **53**.

For iterative PKS, no research into the preferred stereochemistry of CMeT domains has been reported to our knowledge. Recently, experimental results support the idea that the CMeT domain from iterative PKS are also stereoselective for (2*R*)-2-methyl-3-ketoacyl-ACP.^[59] The predicted biosynthesis of georatusin **57** is proposed to be catalyzed by an iterative highly reducing polyketide-peptide hybrid (Figure 1.15 B). The *R* stereoselectivity corresponds to the stereochemistry of the methyl group in the C-13 position of the georatusin. However, more research into the genome, identification of the biosynthetic gene cluster for **57** and investigation into the biosynthetic proteins are needed to confirm the proposed biosynthesis.

1.5.5 Stereoselectivity of Dehydratase (DH)

DH domains catalyze the elimination of water from polyketides containing alcohol groups introduced by KR domains forming alkenes with *E* (or, rarely *Z*) configuration. DH domains in mFAS produce olefins exclusively with an *E* configuration, which proceeds via *syn* elimination of a *pro*-2*S* hydrogen and a 3*R*-hydroxyl group.^[41] A mechanism was suggested in which a single histidine residue plays the role of an acid and a base.^[71-72] Performing a similar *syn* elimination, most PKS DH domains can transform B-type (2*R*,3*R*)-2-methyl-3-hydroxyl intermediates **58B1** to alkenes with *E* configuration, while some PKS DH domains, *e.g.*, from the curacin modular PKS, transform A-type (2*R*,3*S*) intermediates such as **58A1** to alkenes with *Z* configuration.^[60] For example, to check the stereopreference of the DEBS DH-4 domain,^[71] four diastereoisomers of the 2-methyl-3-hydroxyl diketide **58** were synthesized, *e.g.* (2*R*,3*R*), (2*S*,3*S*), (2*S*,3*R*), (2*R*,3*S*), and linked to the ACP domain. Only the (2*R*,3*R*)-2-methyl-3-hydroxyl intermediate **58B1** was recognized by the DEBS DH-4 domain and transformed into the corresponding alkene with an *E* configuration. Furthermore, some DH domains can also recognize and transform different diastereoisomers when different ACP domains are included (Figure 1.16 B). For example, the DH-10 domain from the modular rifamycin (Rif) PKS^[60] can only transform (2*S*,3*S*)-2-methyl-3-hydroxyl diketide **59A2** linked to the native Rif ACP-10 domain into the alkene with *E* configuration through *syn* elimination. However, the diastereospecificity of the dehydration was completely reverted when the diketide was linked to the DEBS ACP-6 domain, which only transforms the (2*R*,3*R*) **61B1** isomer into an *E* alkene through *syn* elimination. As dehydration using the 2*S* isomer is very rare, the function of ACP domain in this context is mysterious. Another interesting phenomenon recently observed is that some DH domains are also capable of epimerizing the methyl group in 2-methyl-3-ketoacyl-ACP as reported using “equilibrium isotope exchange (EIX)” experiments.^[73] However, the epimerase

activity of DH domains is more cryptic than that of KR domains and is still unclear how exactly it arises.

For iterative PKS, research into the SQTKS DH domain^[57] has shown that the DH domain only dehydrates the (2*R*, 3*R*)-2-methyl-3-hydroxylacyl ester (as opposed to the other stereoisomers) and the reaction only forms the *E* alkene through *syn* elimination, which is the same stereopreference as the mFAS DH domain (3*R*, *E*, *syn*).^[41]

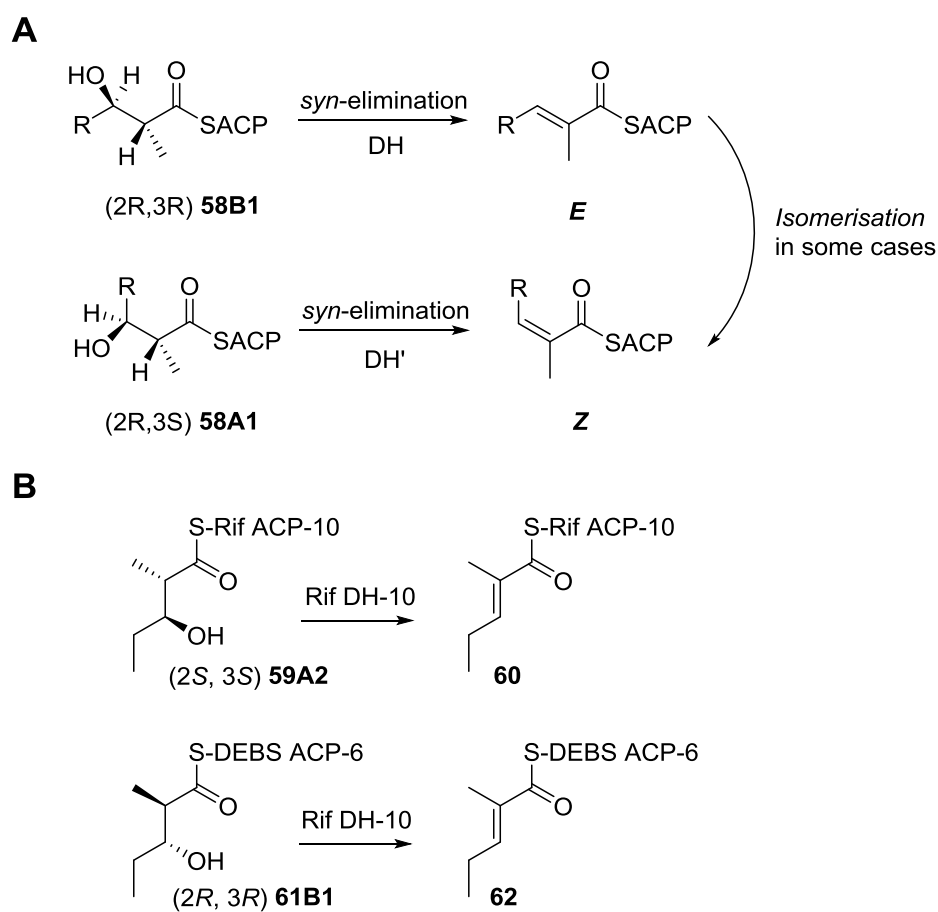


Figure 1.16. **A**, Typical *syn*-elimination catalyzed by DH domains forms alkenes with an *E* or *Z* configuration; **B**, Different ACPs affect the DH stereoselectivity.

1.5.6 Stereoselectivity of Enoyl Reductase (ER)

ER domains catalyze the reduction of α , β -unsaturated intermediates. In fatty acid biosynthesis, the reaction proceeds with attack of the 4'-*pro-R* hydride of NADPH on the

3-*Re* face of an alkene intermediate and performs overall *syn*-addition.^[41] Using the same hydride, in the presence of a C-2 methyl, the reduction can produce 2*R*- or 2*S*-products depending on which side of the double bond is protonated. According to the protein alignments and crystal structures,^[74] obtained from modular PKS, when there is a tyrosine (Y) residue which donates the proton, the 2*S*-product forms; whereas when there is a valine (V), alanine (A), or phenylalanine (F) residue which cannot donate a proton, a lysine residue can donate a proton from the *opposite* direction and create the 2*R*-product (Figure 1.17). The function of these residues in stereocontrol was partly confirmed *in vivo* by site-directed mutagenesis of DEBS ER-4. In the recombinant protein DEBS1-TE, the KR domain in DEBS module 2 was replaced by the DH-ER-KR tridomain from DEBS module 4 to create the DEBS Hybrids (Figure 1.17 A), which produces the triketide lactone **63S** with a methyl group in the 2*S*-configuration. When the conserved tyrosine (Y) residue in the ER-4 domain of DEBS Hybrids was replaced by a valine (V) residue, a lactone **63R** with 2*R*-methyl was produced exclusively.

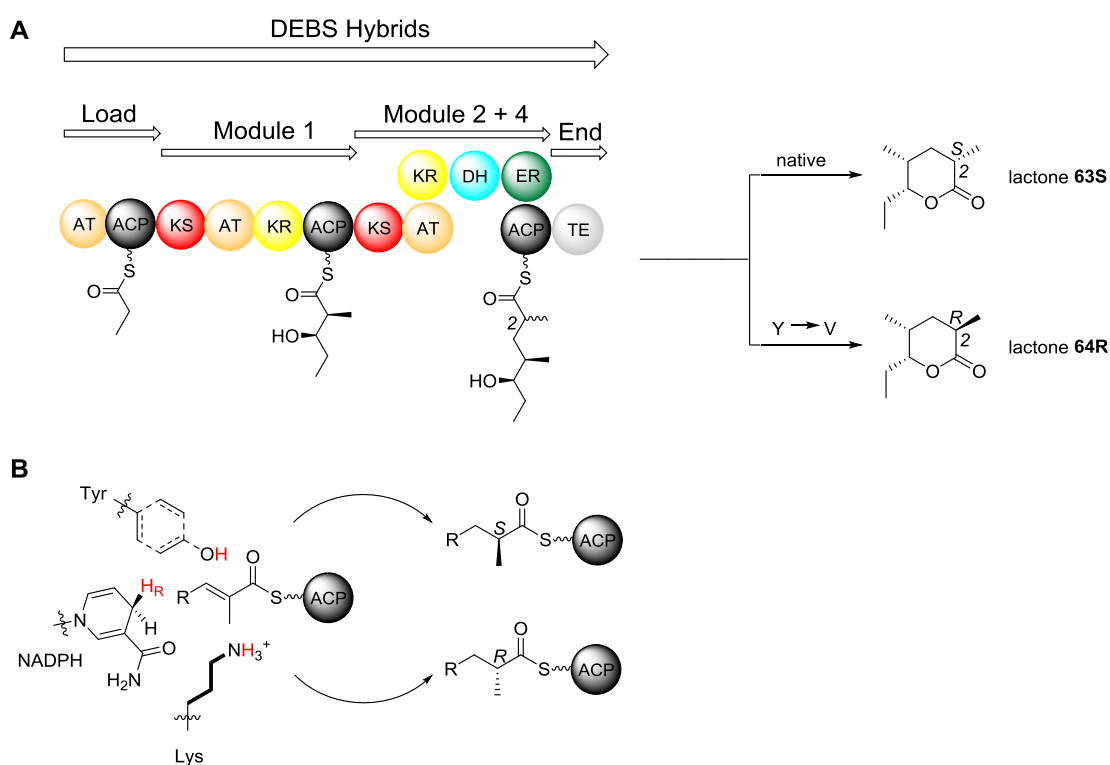


Figure 1.17. A, the function of the tyrosine residue in the active site of ER domain was confirmed *in vivo*;

B, the mechanism of stereocontrol by ER domain.

For iterative PKS, research into the SQTKS ER domain^[75] has shown that the 4'-*pro-R* hydride of NADPH is the donor hydride which attacks the 3-*Re* face of the alkene, resulting in the same stereopreference as the mFAS ER domain,^[41] However, the isolated ER monodomain cannot control the stereochemistry at the C-2 center, released α -racemic product, and the responsible proton donor is still not clear. Further details about this research will be discussed in section 4.1.1.

1.6 Aims of This Project

Squalestatin-S1 **1** (Figure 1.18) is a natural product produced by fungi. It is a potent squalene synthase inhibitor and can be potentially used to treat serum cholesterol related diseases.^[76] Squalestatin tetraketide synthase (SQTKS) is an iterative programmed HR-PKS that catalyzes the biosynthesis of the tetraketide side chain of squalestatin-S1 and is encoded by the *phpks1* gene in *Phoma sp.* C2932. Heterologous expression of *phpks1* in *Aspergillus oryzae* resulted in the production of squalestatin tetraketide **27**.^[77] SQTKS protein contains an active ER domain and this makes SQTKS one of the simplest iterative type I HR-PKS as all the iterative β -modification domains are present in an active state and in which a degree of programming occurs.

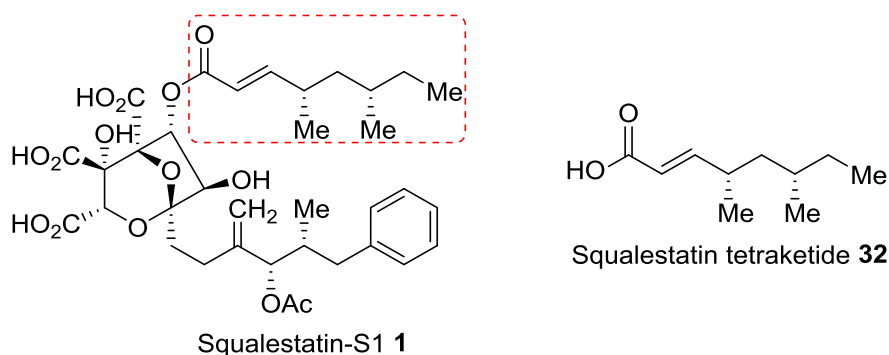


Figure 1.18. Squalestatin-S1 and squalestatin tetraketide.

The proposed biosynthesis of squalestatin tetraketide is shown in Figure 1.19. During the biosynthesis of **32**, acetyl CoA **5** is used as the starter unit and malonyl CoA **6** is used as the extender unit by SQTCS. There are three cycles of chain extension (C1, C2 and C3 in Figure 1.19) in the biosynthesis and in the β -processing modification stages, all of the modification domains are active and are used iteratively. In the first and second cycles of modification, all β -modification domains are active. However, in the third cycle, the CMeT and ER domains are inactive. Finally, compound **31** is formed and released from the system to form the product squalestatin tetraketide **32** by an unknown mechanism.

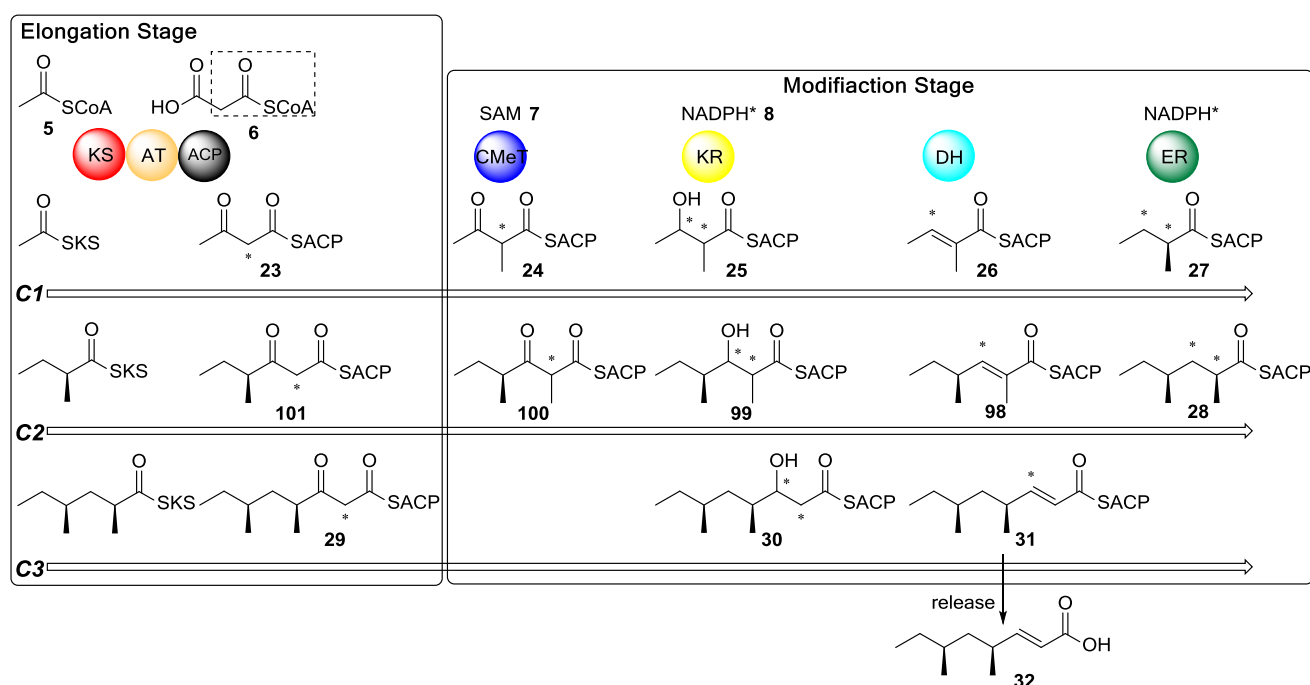


Figure 1.19. Proposed biosynthetic process of **32** by SQTCS (a Type I iterative PKS). All the stereochemistries involved are noted with *.

In this thesis, we have cloned and overexpressed multi-domain fragments of SQTCS. These fragments have been isolated under a rational purification design. As an effect of domain-domain interaction, different stereoselectivities have been found by *in vitro* assays with the isolated SQTCS ER monodomain ^[75] and with the multi-domain

DH-CMeT- ψ KR-ER-KR. With the multi-domain protein, the stereopreference of keto-reduction has been studied. To investigate the methylation mechanism, *in vitro* assays with isolated SQTKS CMeT have been studied. In addition, the effectiveness of using the pantetheine substrate as an analog of native ACP-bound substrate in PKS has been investigated.

2 Strategy and Gene Cloning of SQTKS Multi-domain Fragments

2.1 Introduction

2.1.1 Strategy and Aims

Compared to *in vivo* studies of enzymes, *in vitro* studies can be more controllable and more accurate. *In vivo* experiments generally involve feeding the starting material into a modified strain transformed with target genes to create the new final products. By comparing the new product to the original product, the relationship between the modified gene and the function of protein can be revealed to a certain extent. However, many things happen *in vivo* from DNA to protein and to its catalyzed biosynthesis, which makes tracing the formation of the product difficult. *In vitro*, we only need to pay attention to a relatively simpler relationship between isolated protein and its catalyzed specific reaction. In this way, detailed studies into the catalytic process are possible, such as determination of the stereochemistry during the catalytic process.

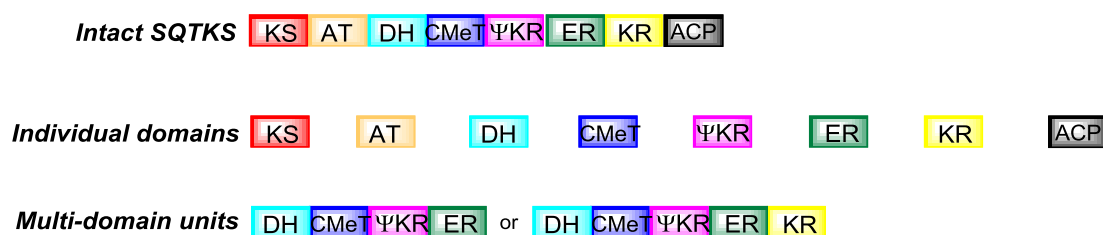


Figure 2.1. Strategy of *in vitro* enzyme assay

For *in vitro* investigation, it is necessary to obtain soluble protein. Several choices (Figure 2.1) can be considered. One reasonable choice is to express the whole SQTKS. However, expression of a soluble and active SQTKS is difficult due to the large size of the protein - 284 KDa per monomer. In our group, efforts have been made to express the

complete protein in *E. coli*, yeast and insect cells, but no active protein has been isolated successfully. The second choice is to isolate every single domain of SQTKS to check the function of each of them. DH^[57] and ER^[75] monodomains of SQTKS have been successfully isolated as soluble proteins recently in our group. One problem of this choice is that the isolated single domains lack connections to other domains, which may cause incomplete activity and unexpected phenomena in *in vitro* assays (see chapter 4). The final choice is to express di-domain, tri-domain or other multi-domain units. With this choice, domains of interest and connections among them can be both considered, as the size of the multi-domain protein is not too large to allow correct protein folding in *E. coli*. This project is focused on *E. coli* expression of a tetradomain (DH-KR) and a CMeT monodomain, while the expressions of the tridomain DH-ER and the penta-domain DH-ACP are also investigated and discussed.

2.1.2 Previous Multi-domain Expression of Iterative Type I PKS

In addition to domains of SQTKS, domains of other iterative type I PKS have been isolated for *in vitro* research over the past years (Table 2.1). However, in numbers, the isolated domains are far less than those of modular type I PKS.

Among them, structures of SAT,^[86] ACP (in solution),^[97] CMeT,^[79] PT^[89] and TE^[90] have been resolved for NR-PKS. Several intact PKS have been overexpressed in *E. coli* and yeast. Many multi-domain fragments of PKS have been overexpressed in *E. coli*. These studies have shed a great deal of light on the mechanisms and selectivities of individual steps that occur during polyketide biosynthesis. However, the programming problem of iterative PKS remains largely unknown. By examination of the selectivity and specificity of several PKS domains from a single system, such as SQTKS, more insights into the programming of polyketide biosynthesis by type I iterative PKS may be gained. As domains of most iterative PKSs have been successfully over-expressed using *E. coli* expression host, in this thesis we will focus on expressing the fragments of SQTKS using *E. coli* system.

Table 2.1. Isolated domains in typical iterative type I PKSs until August 2018 (*crystal structure is available, **solution structure is available).

Type	Overexpressed in <i>E. coli</i>	Overexpressed in Yeast
HR-PKS	SQTKS (squalestatin) DH ^[57] , ER ^[75] LovB (lovastatin) ^[87] KS-AT-DH-MT-ER ⁰ -KR-ACP-CON	Dhc3 (10,11-dehydrocurvularin) ^[85] KS-AT-DH-ψMT-ψKR-ER-KR-ACP CazF (cazaldehyde) ^[86] KS-AT-DH-MT-ER-KR-ACP
PR-PKS	MSAS (6-methylsalicylic acid) ^[111] TH, KS-AT-TH-KR-ACP SACE5532 ((<i>R</i>)-mellein) ^[113] KR, KS-AT-TH-KR-ACP NcsB (2-hydroxyl-5-methyl- naphthoic acid) ^[113] KR AziB (5-methyl-naphthoic acid) ^[112] KS-AT-TH-KR-ACP	
NR-PKS	PksA (norsoloric acid) ^[78] SAT-KS-AT, ACP, SAT ^[83] , TE* ^[90] , PT* ^[89] Pks4 (pre-bikaverin) ^[80] SAT-KS-AT, PT, ACP, TE, KS-AT ^[96] , SAT-KS-AT-PT-ACP-TE/CLC ^[94] , SAT-KS-AT-PT-ACP ^[95] Pks1 (1,3,6,8-tetrahydroxynaphthalene) ^[80] SAT-KS-AT, PT, ACP, TE PKsCT (citrinin) ^[79] CMeT* CTB1 (<i>nor</i> -toralactone) ^[80] PT, ACP, TE ACAS (atrochrysone carboxylic acid) ^[80] PT, ACP** ^[97] wA (YWA1) ^[80] PT, ACP, TE CazM (cazaldehyde) ^[86] SAT* AptA (asperthecin) ^[91] PT	Dhc5 (10,11-dehydrocurvularin) ^[85] SAT-KS-AT-PT-ACP-TE CazM (cazaldehyde) ^[86] SAT-KS-AT-PT-ACP-MT-R

2.1.3 Gene Cloning

To express the protein in *E. coli*, high-throughput construction of expression vectors is important. Various cloning methods^[98] have been developed to make the process simple and accurate. The methods can be mainly classified as restriction enzyme (RE)-based cloning^[100-104] (since 1980s), recombination-based cloning^[105-110] and ligation-independent cloning^[111-121] (LIC), since 2000. In addition it is common to use *E. coli* optimized codon sequences to ensure good levels of protein production.

For the RE-based cloning (Figure 2.2 A1), two specific RE sites are added to each end of the target gene by polymerase chain reaction (PCR) and then digested by two specific restriction enzymes (Step 1). The expression vector, which contains an antibiotic gene for selection, is digested by the same enzymes. Subsequently (Step 2), the target gene and vector are ligated and transformed into competent expression cells. In RE-based cloning, the concentration of PCR amplified target genes is low, which sometimes makes the RE-based ligation between the insert and vector difficult. To increase the concentration of the target gene, the highly efficient TOPO TA cloning (Figure 2.2 A2) is commonly used as an additional step. In this method the PCR product is efficiently ligated into an “activated” T vector and the resulting plasmid can be replicated infinitely in cells to obtain more target genes. TOPO TA cloning uses topoisomerase to both unwind and ligate DNA. The linearised vector with topoisomerase I bound to the overhanging deoxythymidine phosphate in each strand is called “activated” T vector, which is commercially available. First (Step 1), the PCR target gene product is added with a single, overhanging deoxyadenosine to each 3´ end by Taq polymerase, because this polymerase has a nontemplate-dependent terminal transferase activity and lack 5´-3´ proofreading activity;^[122] Second (Step 2), the resulting A-tailed PCR product can be easily ligated into the “activated” complementary T vector on each strand with the help of releasing topoisomerase from the phosphate by the SN2 attack of the 5´ hydroxyl group of the target gene;^[123] Finally (Step 3), large amount of target gene can be obtained from the replicated TOPO clones cut with REs.

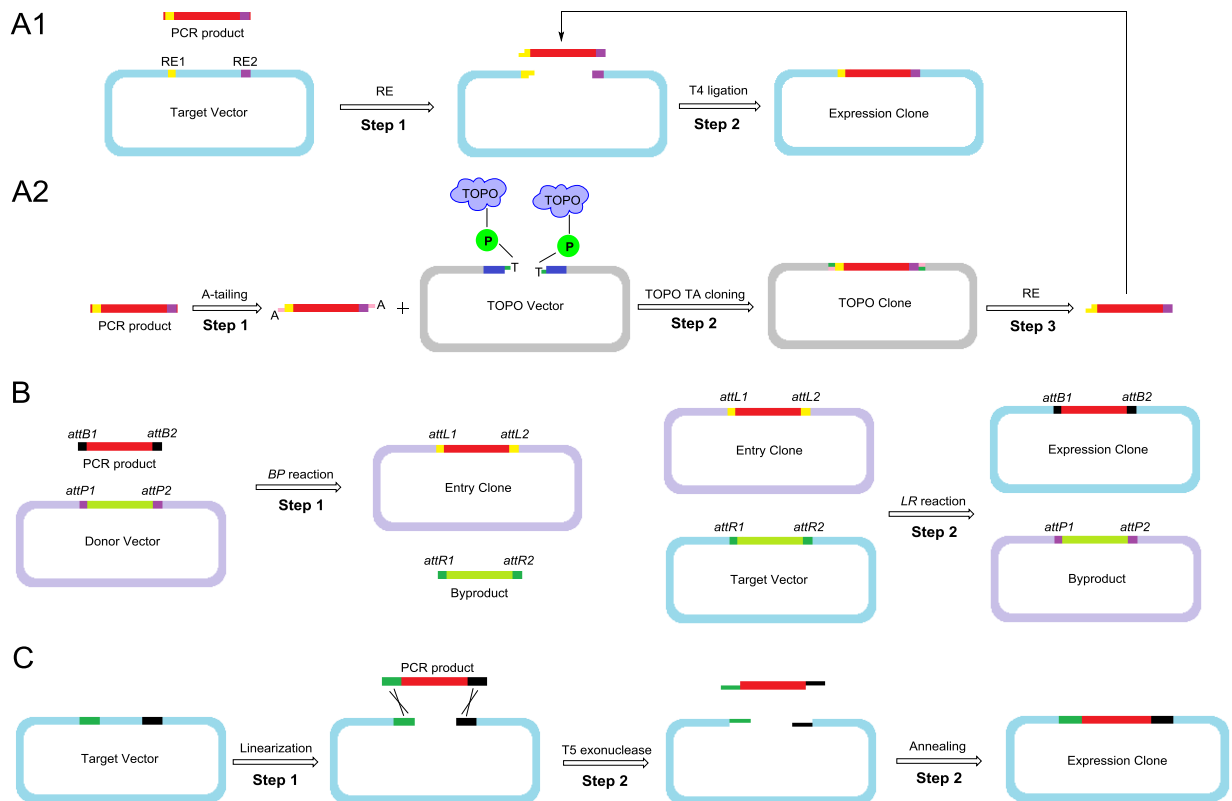


Figure 2.2. Subcloning methods: **A**, RE-based cloning (**A1**) and TOPO TA cloning (**A2**); **B**, recombination-based cloning; **C**, ligation-independent cloning (LIC).

For recombination-based cloning (Figure 2.2 B), a site-specific recombinase is used to construct the required recombinant vector without using any RE or ligase. For example, the Gateway cloning system uses bacteriophage λ to shuttle sequences between plasmids bearing flanking-compatible recombination attachment (*att*) sites. Initially, Gateway *attB1* and *attB2* are respectively added to the 5' and 3' end of the target gene by PCR. Afterwards (Step 1), the *attB*-sequence-containing PCR products are recombined into a donor vector with *attP* to create an “Entry Clone” with *attL* recombination site, which is catalyzed by the “BP Clonase” enzyme mixture. Finally (Step 2), the *attL*-sequence-containing target gene can be recombined into any expression vector which contains *attR* recombination sites, which is catalyzed by the “LR Clonase” enzyme mixture.

Ligation-independent cloning (LIC) (Figure 2.2 C) normally uses T5 exonuclease to remove portions of the 5' ends of insert and vector to generate single-stranded complementary

overhangs (more than 15 base pairs), which are joined together covalently by fusion DNA polymerase and Taq DNA ligase. First (Step 1), the vector is linearised by REs and the PCR product of target gene is generated with at least 15 bp extensions complementary to the vector ends. Second (Step 2), the insert and vector are mixed in a solution of T5 exonuclease, fusion DNA polymerase and Taq DNA ligase to create the expression plasmid. At the same time, the ligation-dependent cloning conveniently allows multiple gene fragments ligation.

All the cloning methods can work well. Decisions about using which cloning methods in a lab normally depend on the balance of cost and efficiency.

2.2 Domain Boundary Determination

2.2.1 SQTKS CMeT

The HR-PKS SQTKS is encoded by the gene *phpks1*.^[124] SQTKS consists of the functional domains KS-AT-DH-CMeT- ψ KR-ER-KR-ACP, where ψ KR is a structural part of the KR domain. For the multi-domain and single domain expression, domain boundary determination affects the function and solubility of the target protein, which is essential for *in vitro* research. By alignment and structural research regarding proteins of similar function, possible boundaries of target protein can be predicted. Meanwhile, a few changes of amino acids around the possible boundaries may considerably affect the protein solubility. Therefore, precisely defining domain boundaries is both important and challenging. In this section, methods to determine the boundaries of the SQTKS CMeT are used as a detailed example.

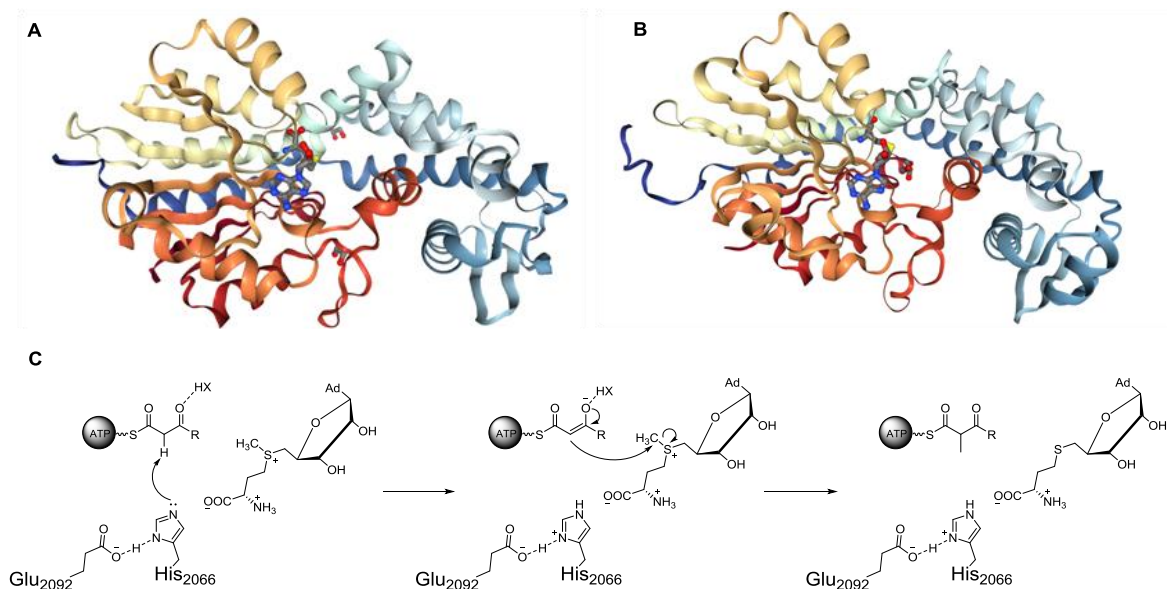


Figure 2.3. Crystal structures of Citrinin CMeT (**A**, 5MPT) and CurJ CMeT (**B**, 5THZ) with binding of SAH.

^[126-127] **C**, proposed mechanism of methyl transfer from the deprotonation effect by the His-Glu dyad (completion of the catalytic cycle by loss of the proton to the solvent is not shown).

Initially, domain boundaries of the SQTKS CMeT were identified by alignment (Figure 2.4) with CMeT domains of other type I PKS and the ψ CMeT of mFAS which is known from crystal structures. The TENS CMeT comes from the iterative highly reducing PKS, Citrinin CMeT comes from an iterative non-reducing PKS, CurJ CMeT comes from a modular PKS. Among them, crystal structures of Citrinin CMeT and CurJ CMeT are known and the methyltransfer activity sites have been identified (Figure 2.3 A, B), which are highly similar. This will help us understand the main difference among these alignment constructs. The active sites are noted as * for CurJ^[125] and Citrinin,^[79] which nearly 100 % matches with each other (Figure 2.4). For Citrinin CMeT, His2066 (Citrinin PKS numbering) is essential for methyl transfer by acting as a catalytic base (Figure 2.3 C). His2066 forms a catalytic dyad with Glu2092 (Citrinin PKS numbering) and deprotonates the α carbon to generate a stabilized enol(ate) nucleophile which enables S_N2-like attack at the methyl donor. The His-Glu dyad (Highlight in red box of Figure 2.4) is also conserved in CurJ CMeT

(His280-Glu306, CurJ numbering). In addition, the secondary structures (xxxx noted as the alpha helix, //// noted as the beta strand) are also similar between Citrinin CMeT (blue) and CurJ CMeT (green). As comparisons, alignments of TENS CMeT and SQTKS CMeT around the active sites are very good and the essential His-Glu dyads are also conserved in both. This indicates the quality of the alignment among the functional CMeTs of type I PKS is good.

However, compared to Citrinin CMeT and CurJ CMeT, in the same positions (Figure 2.4), many alpha helices disappear and more beta strands form in mFAS ψ CMeT (black) which also does not contain the His-Glu dyad in the active site. As the mFAS ψ CMeT has no methyl transfer activity, the considerable alignment difference between the PKS CMeT and mFAS ψ CMeT in turn indicates that the high similarity among the CMeT domains in type I PKS. The mFAS ψ CMeT probably fulfils only a structural role in mFAS. Without active residues, the mFAS ψ CMeT is not a good model for catalysis, but it probably preserves intra-domain interactions and linkers.

Secondly, with a qualified sequence alignment, we can predict the possible boundaries of SQTKS CMeT. A simple idea is, the range of SQTKS CMeT should include all the secondary structures and active sites. In addition, solubility prediction (SolPro) and primer design for gene cloning have been considered as tools for narrowing the range of the boundaries. In this way we designed four N-terminal positions and one C-terminal position for the CMeT expression experiments (Figure 2.4). The starting positions of four constructs MT1, MT2, MT3, MT4 are all located at the C-terminus of the DH domain which boundaries of which have been determined by David Ivison previously.^[128]

<i>Residue</i>	<i>Foreground</i>	<i>Property</i>
AVFPMILW	RED	Small (small+ hydrophobic)
DE	BLUE	Acidic
RK	MAGENTA	Basic
STYHCNGQ	GREEN	Hydroxyl + sulfhydryl + amine
Others	Grey	Unusual amino/imino acids

	MT4	MT3	MT2	MT1	End of SQTKS DH	
	xxxxxxxxxxxx					
		xxxxxxxxxxxx				
		//	////	////		//// /
SQTKS	-----FTVNVDVSSKADSEHTPVLEIKGLRNQSVGQMAPQPGDSSNNDLCFKLDWAPDIS					1334
TENS	---SSLTGNINVD--AESGRALIQVEGFVRAVG---EPDASKDRLLFYETVWGRDIS					1329
Citrinin	LPKSAILDAFRIAKEA-----					1804
CurJ	LPPDFLLDPVEVSQQLAPSLTELVT-----					1293
mFAS	-----QA---ADVVDNRNLNTVVAG-----GALFLGAHS					1101
		xxxx	xxxxxxxxxxxxxxxxxxxx			//
		xxxxxxxxxxxxxxxxxxxx				/
	//		////	///	///	
SQTKS	SVKQERLKEKFG--FPLDPTEAALIMGLRQACIHFHHSLSQSLTAPDRDQLDWHQ-----					1387
TENS	IMGLSDPIRD---ETS DAMVHNLSEATIEVSLFYVRLMGELSTADRRQANWYH----					1380
Citrinin	-----TDDFILNGQLGTYNNEVMPRSTRLCVAHIVAFEQLGCPIRSAAAYQRL----					1853
CurJ	-----LLDNARTSETGTQLEKLSVDYIVVGLL-----QMGWSYQPTES					1331
mFAS	SVAPRRPQ-----EHLKPILEKFCETPHV-----					1126
		:	:	:		
		Xxxxxxxx	x	xxxxx	////	////
	//xxxxxxx	xxxxxxxxxxx	xxx	////	////	xxx
	///					
SQTKS	-----KRFYDWMVLQIQLAEDRLAPNS-SAWLQCS-----SS					1419
TENS	-----TRMLAAFADYHLAKVHEETHLHLR-PEWLADD-----WA					1412
Citrinin	-----ERVYPYLPKHERFMNLI-Y--GLLEEARLIDINGSSEITRTSVPVST---KSVET					1900
CurJ	FDDLAAAQCLGVVPTQVRLFERRLLQ--ILAEVGILOSNQ-QQWQVQRTAQ-K---VNPSK					1384
mFAS	-----SGCLA--GNTA-----LQELQLCRGL-AQALQTKVAQQGLKMPVPG					1166
	Xxxxxxxx	xxxxxxxxxxxx	xxxxxx		*xxx*	xxxxxxxx**x
	Xxxxxxxx	xxxxxxxxxxxx	xxxxxx		x*xxx*	xxxx**x
SQTKS	DEQKLEIVRASVNGQIVVHVGKSLAIIRHIA----PLELRLQKLLY----RYYT					1470
TENS	VIQ-TIDRAYPDAVELQNLHAVGQNVADVIRGKKH----LLEVLRVNLNLD----RLYT					1462
Citrinin	MLEELLHEPLHAAEHKTSLTGSKADCTGKED----GLQLIFGPEGREIVTDVYA					1955
CurJ	QSQSLLSYDPDEATLTALERCASQSGVIRGEID----PVQLVFPQGLT-TATQLYK					1438
mFAS	DGAQAPRMAPQQLPRLAAACQLQNGNQLQLGQVLAQERPLCDPLLS----GLL-					1221
	:	:	:	:	:	:
	X**xxxxxxxxxxxx			////	****	xxxxxxxx
	X x**	xxxxxxxxxxxx		////	****	xxxx
		Xxxxxxx		////		xxxx
SQTKS	DAIKW----DRSYQQDQLKLAHAKCP---TAKITEIGAGTGGCTRAVLDAISNOGIAR					1523
TENS	EDKGM----HMANLFLANAEIEITFKFP---RCKIEIGAGTGATTWAALSAIGE----A					1511
Citrinin	KSPINAVWIQQAEFFEQQLKRLPNT---GEPLRIEIGAGTGGTTVKMLPLERLGV--					2010
CurJ	OSAVA---KVMNTIEKVMKAMEKLPSPGIRLLEIGAGTGGTTSYILPHNP----N					1490
mFAS	DAPAL-----KACDTAEN-----MASPKMKVVENLAGDQQLYRIPALINTQPV-M					1268
	.	:	:	:::	** *	:
	////	**	xxxxxxxx	////	*	////
	////	**	xxxxxxxx	////	*	////
	////		xxx	////		////
SQTKS	CAQYDFTDVSSGFFFAAQKFAAFDDVIRFQKLDIEKDIEMQGFECGSYDVIATSQVTHA					1583
TENS	FDTYTYTDLVSGFFENAVEKPSAFRHRMVFRALDIEKDPASQSFDLNSYDIIATNVVHA					1571
Citrinin	PVEYTMTDLSSSLIAARKKPKKYPF-MKFKVVNIESPPDPQL--VHSQHIIATNCVHA					2067
CurJ	QTEYIFTDIGALFTSKAQEKQDYRF-LGYQTLDIEVDPSSQGFESHRYDVIILANVHA					1549
mFAS	DLDYTATDRNPQALEAAQAKLEQLHVTQGG--W-DPANPAPG--SLGKADLLVONCALAT					1323
	*	**	.	*	:	:

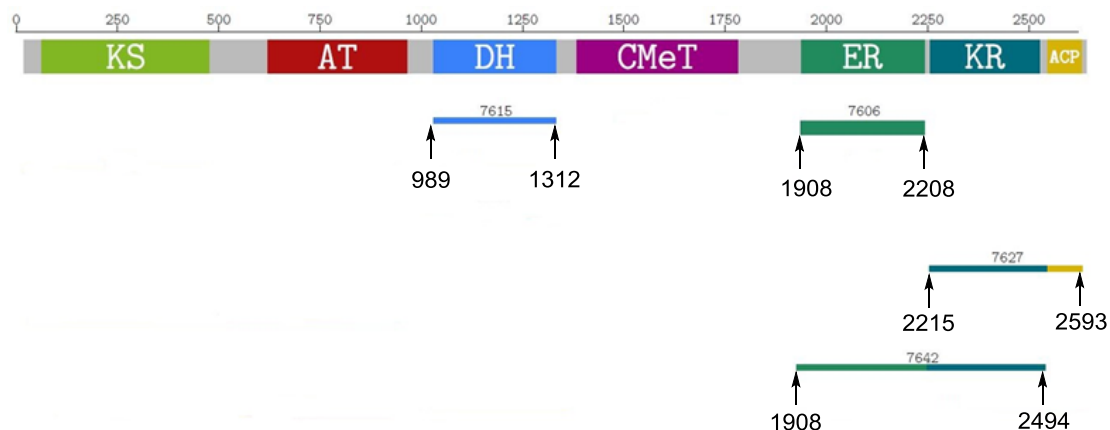


Figure 2.5. Domain boundaries of SQTCS DH-ER tridomain, DH-KR tetradomain, DH-ACP pentadomain used in previous work from David Ivison.

The isolation of SQTCS ER and DH monodomains has already succeeded in the Cox group^[57, 75] based on the determined boundaries (Figure 2.5). Therefore, the fragment of DH-CMeT-ψKR-ER (~135 kDa) is worthwhile to be studied as a multi-domain unit using the already used N-terminal and C-terminal boundaries. In addition, based on the similarity between Type I iterative PKS and mammalian FAS (section 1.3 and alignments in chapter 6), the structure of SQTCS can be modeled (Figure 2.6). The dimeric architecture includes a movable ACP and the X-shape construct is connected by few amino acids between DH-KR (top) and KS-AT (bottom). Therefore, compared to the DH-ER tridomain (~135 kDa), the longer DH-KR tetradomain (~166 kDa) has much less hydrophobic surface. This may increase the possibility of obtaining a soluble multi-domain protein. On this basis, the SQTCS DH-ACP pentadoman (~176 kDa) may also be possible to isolate as a soluble protein.

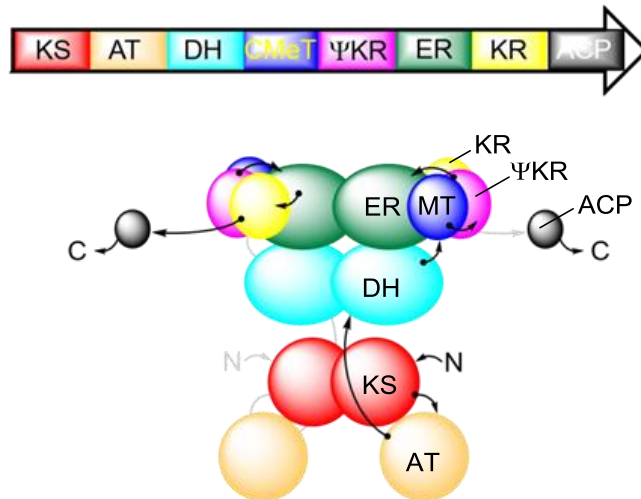


Figure 2.6. Dimeric structure cartoon of SQTKS, N to C terminals direction indicated, which shows the likely domain architecture based on that of mFAS.

2.3 Codon Optimization

The *phpks1* gene which encodes SQTKS was isolated from fungi of *Phoma* species. Our aim is to heterologously express it in *E. coli*. Many amino acids are coded by more than one codon and these codons are synonymous. During the transcription process, these synonymous codons are transferred by different tRNAs onto the ribosome. Genes normally display a non-random usage of synonymous codons for specific amino acids. Therefore, the codon usage regulates the ribosome elongation rate. Frequently used codons speed the rate up, while rare codons slow it down. However, codon usage is different between fungi (Figure 2.7 A) and *E. coli* (Figure 2.7 B). When *E. coli* does not have enough rare-codon tRNAs which are abundant in fungi, the transcription process may be ceased and the immature protein forms. This non-intact protein is fragile and tends to form insoluble inclusion bodies. This is illustrated in Figure 2.7 A and B which show the relative synonymous codon usage (RSCU) in the *Phoma* native host and *E. coli* for the same region of the *phpks1* gene. Codon frequency of lower than 10 % in some cases would be expected to severely slow translation on the ribosome.

One method to solve this problem is to optimize the original sequence by gene synthesis to adapt to the codon usage in the expression host. Another method is to express the original gene sequence in the expression host which supplies rare-codon related tRNAs. As the artificial gene synthesis is well-developed and it is getting cheaper and cheaper nowadays, we used synthetic genes in this work. In this way, to increase the chance of obtaining soluble and fully active SQTKS DH-KR, two methods of gene optimization have been performed: *E. coli* full codon optimization; and *E. coli* codon sub-optimization.

The *E. coli* full codon optimization uses the most frequently used synonymous codons for each amino acid (Figure 2.7 D). In this way the ribosome elongation keeps in a high rate throughout transcription and intact protein should form smoothly. Recently, evidence^[129] has been shown that codon usage also affects the rate of co-translational protein folding and that use of non-preferred codons is generally related to unstructured protein regions. Therefore, it is possible that the rate fluctuation during the ribosome elongation in the original strain (Figure 2.7 A, from left to right), which is affected by the codon usage, is also necessary for proper protein folding. Therefore, based on the *E. coli* codon usage, a codon sub-optimization has been designed by imitating the RSCU (the ratio between the number of the tRNA related to the synonymous codon and that to the most frequently used synonymous codon) fluctuation in *Phoma* C2932 (Figure 2.7 C, from left to right). For example, the original RSCU of each codon of DH-KR in *Phoma* C2932 can be listed (Figure 2.7 A) and the original RSCU of the codon CUA (encoding leucine) is 38 %. In *E. coli*, the synonymous codons encoding leucine are CUA, CUU, CUC, UUA, UUG, CUG. Their RSCUs are ~ 7 %, ~ 24 %, ~ 27 %, ~ 31 %, ~ 33 % and 100 %, respectively. To imitate the 38% RSCU in *Phoma* C2932, the UUG (~ 33 %) is the most similar RSCU. Therefore, the original CUA should be changed to UUG to reach the similar RSCU of *Phoma* C2932 in *E. coli*.

In this way, codon (*Phoma* C2932-like *E. coli* codon) sub-optimization of DH-KR is possible (Figure 2.7 C). Compared to the original RSCU fluctuation of DH-KR in *E. coli* (Figure 2.7 B) with that of the fully *E. coli* codon optimization (Figure 2.7 D), the RSCU

fluctuation of the sub-codon optimization (Figure 2.7 C) is seen to be similar to that of the original RSCU in *Phoma C2932*.

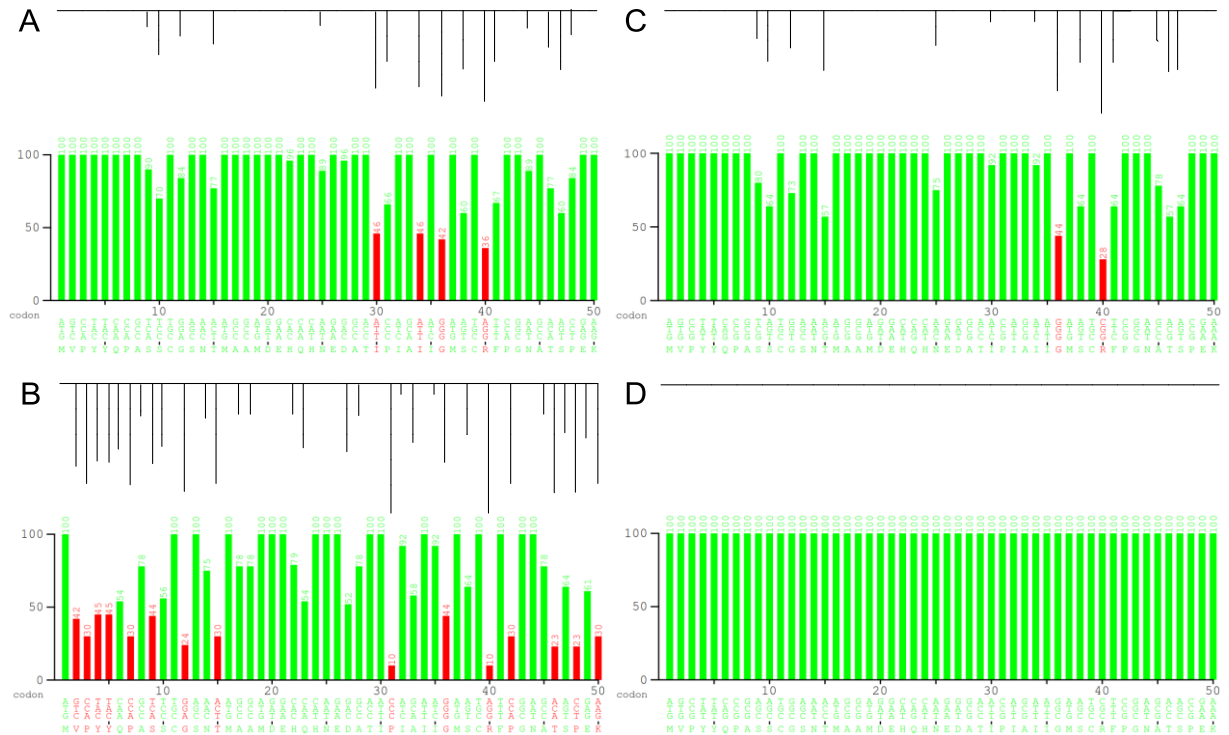


Figure 2.7. Relative Synonymous Codon Usage (RSCU) of the typical fragments of SQTGS DH-KR (Y axis refers to RSCU): **A**, RSCU in *Phoma C2932*; **B**, RSCU in *E. coli*; **C**, *E. coli* codon sub-optimization by imitating the RSCU fluctuation in **A**; **D**, *E. coli* full codon optimization.

2.4 Expression Constructs

The *E. coli* codon optimized genes of the whole SQTGS DH-ACP pentadomain and the codon sub-optimized SQTGS DH-ACP pentadomain were synthesized by Thermo Scientific and confirmed by gene sequencing (Chapter 6). Based on the boundary determination, gene fragments of domains were amplified by PCR. Finally, by using different gene cloning methods, 9 expression plasmids (Table 2.2) were constructed to produce DH-ER tridomain, DH-KR tetradomain, DH-ACP pentadomain and a CMeT monodomain.

Table 2.2. Successful constructed expression plasmids in this thesis.

Sequence	Vector	Codon Optimization	Cloning Method	Tag Type
DH-ER 2.8B	pET28a	full	RE + TOPO TA	N-His ₆
DH-KR 2.9A	pET28a	full	RE + TOPO TA	N-His ₆
MT1 2.8A	pET28a	full	RE + TOPO TA	N-His ₆
MT2 2.8A	pET28a	full	RE + TOPO TA	N-His ₆
MT3 2.8A	pET28a	full	RE + TOPO TA	N-His ₆
DH-ACP 2.8C	pET28a	full	LIC	N-His ₆
sub DH-KR 2.9B	pET28a	sub	LIC	N-His ₆
sub DH-KR 2.9C	pETM22	sub	LIC	C-His ₆ N-His ₆ TrxA (to remove)
sub DH-KR 2.9D	pETM11-SUMO3GFP	sub	LIC	N-His ₆ SUMO (to remove)

The CMeT clones were made from the fully *E. coli* codon optimized SQTKS. As shown in section 2.2, different boundaries of the SQTKS CMeT have been predicted. They are MT1 (1311-1707), MT2 (1298-1707), MT3 (1288-1707) and MT4 (1280-1707). The four CMeT fragments were cloned into expression vectors separately by using TOPO TA cloning and RE-based cloning (see experiment procedures in chapter 6). As a result, pET28a with MT2, MT3 and MT4 were successfully constructed (Figure 2.8 A). To successfully transform the pET28a with target gene into the expression host, only the cells with the plasmid containing kanamycin resistance gene can survive on kanamycin media. In *E. coli* expression host, under induction (see section 3.1), the RNA polymerase starts the transcription process at T7 promoter and ends at T7 termination. Afterwards, ribosome binds to the RBS site of the mRNA and the translation processes from the start codon to the stop codon. As there is an N-terminal His₆ tag before the target gene in pET28a, the final protein will be tagged with 6 histidines at the original N terminus, which is convenient for protein purification by Ni-NTA chromatography. The gene encoded DH-ER and DH-ACP also come from the *E. coli* full

codon optimized SQTKS. By using the same strategy, corresponding expression vector has been constructed (DH-ACP by Ligation-independent cloning, see chapter 6) to produce N-His DH-ER (Figure 2.8 B) and N-His DH-ACP (Figure 2.8 C).

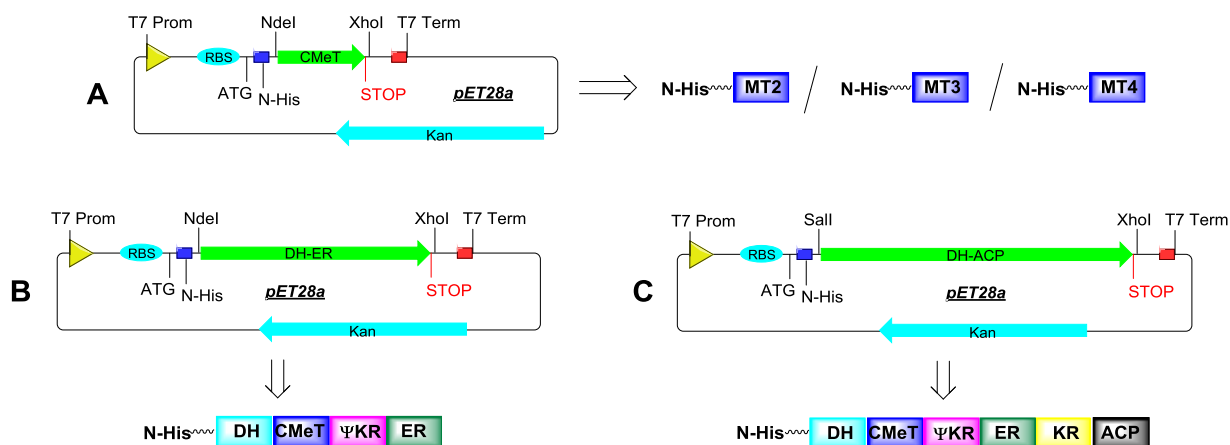


Figure 2.8. Constructed expression plasmids with SQTK CMeT, DH-ER and DH-ACP in pET28a: **A**, three plasmids with fully codon optimized CMeT of different boundaries (MT2 (1298-1707), MT3 (1288-1707), MT4 (1280-1707)) were constructed to produce proteins of MT2, MT3, MT4 with N terminal histidines; **B**, pET28a inserted with fully codon optimized DH-ER was constructed to produce protein of DH-ER with N terminal histidines; **C**, pET28a inserted with fully codon optimized DH-ACP was constructed to produce protein of DH-ACP with N terminal histidines.

For the SQTKS DH-KR, fully codon-optimized gene (Figure 2.9 A) and sub codon-optimized gene (Figure 2.9 B) have been inserted into pET28a expression vector by gene cloning (the former by TOPO TA cloning and RE-based cloning, the latter by Ligation-independent cloning) to produce N-His DH-KR respectively. To increase the chance of obtaining soluble DH-KR protein, expression plasmids with solubility-enhancing fusion tags, such as pETM22 with TrxA^[130] (Thioredoxin A, 11.7 kDa) and pETM11 with SUMO^[131] (small ubiquitin-like modifier, ~12 kDa), have been constructed for sub codon-optimized DH-KR. In this way, N-His-TrxA tagged DH-KR with C-terminal histidines (Figure 2.9 C)

and N-His-SUMO tagged DH-KR (Figure 2.9 C) will be produced. These solubility-enhancing fusion tags may help protein expression and folding, but it may also restrict the protein activity. Therefore, after obtaining the overexpressed proteins, these tags should be removed by specific proteases. As the specific protease site (provided with pETM22 and pETM11) is located at the position between the fusion tag and the DH-KR, addition of 3C protease will remove the N-His-TrxA and addition of TEV protease will remove the N-His-SUMO. As a result, DH-KR with C-terminal histidines and DH-KR without histidines are supposed to be isolated. This further increases the chance to obtain a stable and active DH-KR protein.

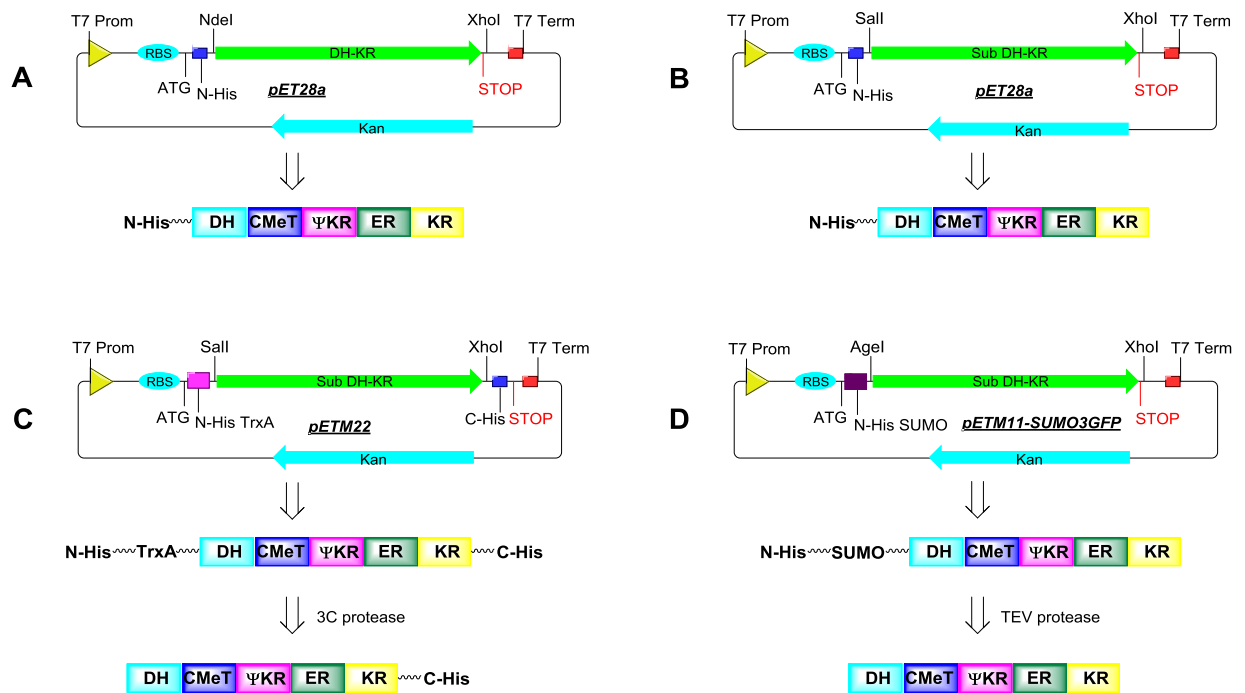


Figure 2.9. Constructed expression plasmids with the SQT DH-KR: **A**, pET28a inserted with fully codon-optimized DH-KR was constructed to produce protein of DH-ER with N terminal histidines; **B**, pET28a inserted with sub codon-optimized DH-KR was constructed to produce protein of DH-KR with N terminal histidines; **C**, pETM22 (with Trx fusion tag) inserted with sub codon-optimized DH-KR was constructed to produce protein of DH-KR with C terminal histidines; **D**, pETM11-SUMO3GFP (with SUMO fusion tag) inserted with sub codon-optimized DH-KR was constructed to produce protein of DH-KR with no terminal histidines.

2.5 Conclusion

To overexpress multi-domain fragments of SQTKS in *E. coli*, synthetic *E. coli* full codon optimized gene of SQTKS was used. In addition, to increase the chance of obtaining soluble and active proteins, special sub-optimization method was used. The synthetic genes have been certified by gene sequencing. Determination of domain boundaries was based on gene alignments. Corresponding genes of interest were successfully cloned into different expression vectors. The expression and purification of these SQTKS fragments will be discussed in chapter 3.

3 Expression and Purification of SQTKS Multi-domain Fragments

3.1 Introduction

3.1.1 Expression

In this thesis, the SQTKS DH-ER tridomain (~135 kDa), the DH-KR tetradomain (~166 kDa), DH-ACP pentadomain (~176 kDa) and the monodomain CMeT (~48 kDa) have been the subject of overexpression experiments in *E. coli*. Most of these proteins are considerably large for expression in *E. coli*. The chance of getting soluble protein in *E. coli* falls as the size of the protein increases. Therefore, different expression methods, strains, culture media and vectors (with different fusion tags to help protein folding, discussed in chapter 2) have been tried.

The *lac* operon is one of the most commonly used systems for creating recombinant proteins in *E. coli*. The operon (Figure 3.1) consists of: a promoter; an operator *lacO*; *lacZ*, which encodes β -galactosidase; *lacY*, which encodes β -galactoside permease; *lacA*, which encodes β -galactoside transacylase; and a terminator. The *lac* operon is controlled by the availability of lactose in the culture medium. As monosaccharides are more easily metabolized by cells than disaccharides, the presence of glucose (a monosaccharide) inhibits the use of lactose (a disaccharide). If the culture medium contains both glucose and lactose, lactose can only be used after the glucose is depleted. LacY (which pumps lactose into cell) and LacZ (which mainly hydrolyse lactose to galactose and glucose) are also known to be necessary for lactose catabolism.

In the absence of lactose, the tetrameric repressor protein encoded by *lacI* binds to the operator gene (*lacO*) and bends its DNA by 40 degrees. This represses access of RNA polymerase to the promoter site and thus prevents transcription. In this situation, only very little expression can happen.

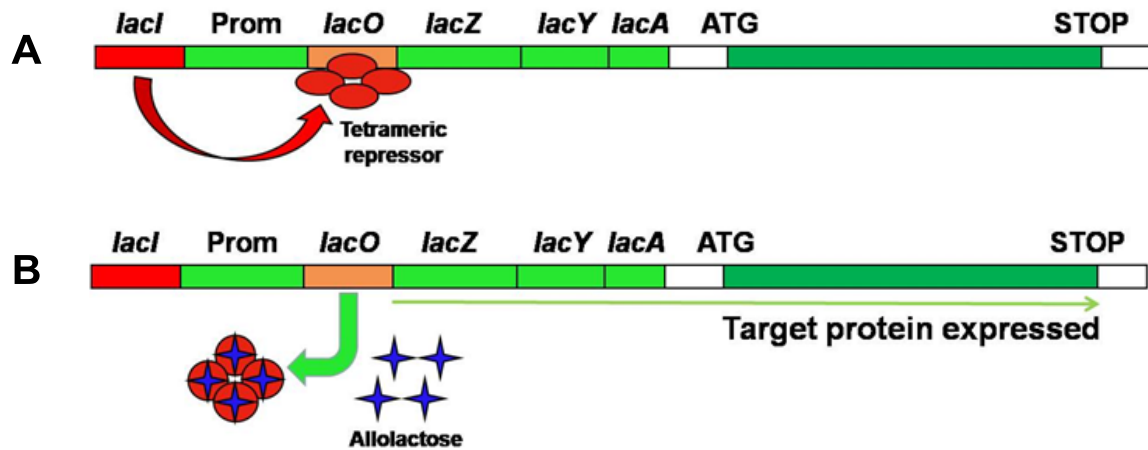


Figure 3.1. Negative control by the *lac* operon: **A**, situation without lactose. The *lac* operon is repressed by the tetrameric repressor protein encoded by *lacI*, which prevents the transcription by RNA polymerase; **B**, situation with available lactose. The *lac* operon is derepressed by unbinding of the repressor whose structure has been changed by the combination of allolactose, which paves the way for transcription.

In the presence of lactose, once the glucose is depleted, lactose can be pumped into the cell by β -galactoside permease encoded by *lacY* (low-level expression due to the repression of tetrameric repressor protein). Then some lactose molecules are transformed into allolactose by β -galactosidase encoded by *lacZ* (low-level expression). Then allolactose binds to the tetrameric repressor and induces a conformational change in the protein structure that renders it incapable of binding to the operator DNA sequence. In this way, repressor proteins begin to detach from the promoter sequence so that RNA polymerase can bind to the promoter and start the transcription. As a result, more β -galactoside permease and β -galactosidase are expressed and more allolactose is produced for derepression, which induces high-level protein expression. This process is called auto-induction, the over-expression method based on the use of lactose.

Isopropylthiogalactoside (IPTG), a structural mimic of allolactose, can also bind to the repressor protein and induce a similar conformational change that greatly reduces its affinity for operator DNA sequence. Unlike lactose, IPTG is not part of any metabolic pathways and so will not be broken down or used by the cell. This ensures that the concentration of IPTG added remains constant, making it a more controllable inducer of the *lac* operon than lactose itself.

However, more and more evidence^[132] has been shown that IPTG has some potential toxicity which is harmful for the cell growth thereby affecting protein expression. Compared to commonly used IPTG induction, auto-induction (*i.e.*, use of native lactose concentrations in the medium) can result in higher cell densities and higher protein production. In addition auto-induction is often slow and thereby encouraging folding of larger proteins. Previous work in our laboratory^[133] showed that the use of auto-induction led to the successful production of a soluble isolated AT domain from the norsolorinic acid synthase (NSAS), a type I iterative NR-PKS, where IPTG induction had failed.

To increase the chance of obtaining plenty of soluble protein, IPTG induction and auto-induction were both tested in this project. In addition, *phoma pks1*, which encodes SQTKS, was codon optimized by gene synthesis to adapt to the codon preference in *E. coli*. Without the necessity of trying any special *E. coli* strain offering rare codon tRNA, two strains were mainly used: the work-horse *E. coli* BL21 (DE3) and *E. coli* ArcticExpress (DE3), both of which contain the λ DE3 lysogen that carries the gene for T7 RNA polymerase. The T7 RNA polymerase is extremely T7 promoter-specific and transcribes only the recombinant target gene initiated by the T7 promoter (under the control of *lac* operon in pET vector). Under this system the recombinant target gene can be overexpressed specifically to improve the production of target protein. Derived from *E. coli* Gold cells, the ArcticExpress (DE3) strain also co-expresses the cold-adapted chaperonins Cpn10 and Cpn60 from the psychrophilic bacterium, *Oleispira Antarctica*.^[134] In this way, the cell enables controllable target gene overexpression at temperature much lower than that of other *E. coli* strains to achieve slow protein folding for increasing the recovery of soluble protein.

Several media have been used in cell culture. Yeast extract and tryptone are the main components. Yeast extract is the hydrolysate of yeasts made by removing the cell wall and other insoluble portions after digestion of its protein components. It provides the main elements, trace nutrients, vitamin B complex and other important growth factors, which are essential for the growth of diverse microorganisms. Tryptone is a pancreatic digest of casein, which provides a mixture of amino acids, including essential amino acids and larger peptides.

In the order of lysogeny broth (LB), 2 tryptone yeast medium (2TY), terrific broth (TB), the medium is getting richer which means more materials are available for healthier cell growth and higher possibility to obtain more soluble over-expressed protein. LBE5052 is an optimized medium for auto-induction.^[136] With the same amount of yeast extract and tryptone as LB, LBE5052 provides lactose for auto-induction, and it also provides additions to optimize the condition for cell growth. For example, NH_4^+ and SO_4^{2-} supplies more nitrogen and sulfur; Mg^{2+} helps to bridge the highly negatively charged Lipopolysaccharides (LPS) molecules in the outer membrane of *E. coli*, which is important to the structural integrity of most Gram-negative bacteria and makes the cell membrane function well;^[135] addition of trace metals can considerably increase cell saturation density; a glucose concentration of 0.05 % will not be depleted until the culture has grown to moderate density before auto-induction.^[136]

3.1.2 Identification of Target Protein

We can find out whether the target protein has been over-expressed by observing its molecular weight in Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophores (SDS-PAGE) which is a commonly used protein separation technique based on protein molecular weight. With the help of boiling and a reducing agent (normally DTT or β -mercaptoethanol to break down protein-protein disulphide bonds), the detergent SDS disrupts the tertiary structure of proteins and converts the natively folded proteins to unfolded linear molecules. With a uniform ratio of negative charge to mass, SDS coats the linear form of the protein and forms linear micelles which are of the same width with length proportional to their molecular weight. Therefore, under a supplied electrical field, with the same ratio of charge to mass, the migration of SDS-coated proteins will be only determined by their size (mobility in the gel) which is related to molecular weight. In this way, low molecular-weight proteins migrate faster than high molecular-weight proteins, which leads to the separation of proteins in SDS-PAGE.

From the SDS-PAGE, we can determine an over-expressed protein of approximate target molecular weight from thousands of host cell proteins. However, this method is not accurate enough to certify the over-expressed protein as the target protein, especially when the

expression level of the target protein is low or the protein has been lysed into several pieces. In order to get more information from the SDS-PAGE, protein identification by Electrospray-ionisation quadrupole time-of-flight (ESI-QTOF) mass spectrometry has been used in this thesis.

Mass spectrometry is an ideal method for protein mass determination. The two primary methods for ionization of proteins are electrospray ionization (ESI) and matrix-assisted laser desorption/ionization (MALDI). In ESI ionization proteins are ionized in solution and carry multiple charge state. The advantage of using ESI-QTOF analysis for protein mass determination is that due to the high charge state of proteins their m/z measurements is typically less than 2000 and the TOF detector has a very good resolution and mass accuracy in this scan range. This results in more accurate mass measurements for protein identification. The sample proteins can be obtained directly from slices cut from an SDS-PAGE analysis. The process involves using tryptic digestion to obtain masses of individual peptides derived from the protein. Subsequently, these peptides are introduced into the mass spectrometer and identified by tandem mass spectrometry. Then the masses are compared against the NCBI database, and probability-based scoring systems are used to determine the closest peptide matches and the protein to which they belong. In this way, by comparing the identified peptides to the sequence of protein, every protein in SDS-PAGE is knowable. This can help us further understand the situation of target protein to design better protein expression and purification methods. In this thesis we collaborated with the group of Prof. Haus-Peter Braun (LUH) who did the ESI-QTOF-MS analysis.

3.1.3 Purification

Sequential chromatographic methods, Immobilized Metal Ion Affinity Chromatography (IMAC) which works as the crude capture of the target protein, Ion Exchange Chromatography (IEX) which is an intermediate purification to remove bulk impurities) and

Size Exclusion Chromatography (SEC) which works as a polishing step to achieve final high-level purity, have been used for protein purification.

IMAC is based on the interaction of proteins with histidine residues (or Trp and Cys) on their surface with divalent metal ions (*e.g.*, Ni²⁺, Cu²⁺, Zn²⁺, Co²⁺) immobilized *via* a chelating ligand. Among them, histidine-tagged proteins have an extra high affinity in Ni-NTA which is made by the chelating ligand nitrilotriacetic acid (NTA) coupled to a cross-linked 6% agarose resin. The 6×His residues are usually the strongest binder among all the proteins in a crude cell extract, while other cellular proteins will not bind or bind weakly. Then addition of high concentration of imidazole elutes all the bound target protein out of the column. IEX separates proteins with different isoelectric point (pI) and is based on the reversible interaction between a charged protein and an oppositely charged chromatography medium. Elution is usually performed by increasing salt concentration or changing pH in a gradient. SEC, also called gel filtration (GF), is a non-bonding method which allows separation of proteins with differences in molecular size under mild conditions. The column is tightly packed with small porous polymer beads designed to have pores of the different size. Large molecules that cannot enter the pores of chromatography beads elute fast, while molecules of middle size that can use a fraction of the pore volume of the beads elute slower. Salt or other low-molecular-weight substances that can use the entire pore volume of the beads elute the slowest. In addition to the separation, the native size of protein can be estimated by calibration and this method is particularly useful for estimating the quaternary structure of the target protein.

3.2 Expression and Purification of SQTCS Fragments

3.2.1 SQTCS DH-ER Tridomain

Soluble and active single-domain constructs of SQTCS DH^[57] and ER^[75] have previously been produced and analyzed. This gives confidence that the domain boundaries used in these cases are effective for the production of soluble, active protein. The DH N-terminal and ER

C-terminal boundaries were used to define a three-domain fragment including DH, CMeT and ER domains, known as SQTKS DH-ER. The following paragraphs describe the process of expression and purification of this protein.

PCR was used to amplify the DH-ER tridomain sequence from the codon-optimized template. This was ligated into pET28a, giving an expression construct with an N-terminal His₆ tag (section 2.4). The plasmid (Figure 2.8) was transferred into *E. coli* BL21 (DE3). With the condition of 1 mM IPTG and 4 hours' incubation at 37 °C (180 rpm), the over-expressed insoluble protein of almost desired size was obtained (Figure 3.2). Being identified by ESI-QTOF-MS (chapter 6), the insoluble protein was proved to be N-His₆-DH-ER. These insoluble proteins, known as inclusion bodies, are commonly found in forced over-expression of a heterologous protein in *E. coli*. They are usually caused by incorrect protein folding. In comparison, no expression of soluble protein of the expected size was found.

To obtain more soluble protein, less IPTG and lower temperature incubation are commonly used to slow the expression process for better protein folding. With BL21 (DE3), 0.1 mM IPTG and 16 °C for 24 h was tested, but no improvement and similar expression level was found.

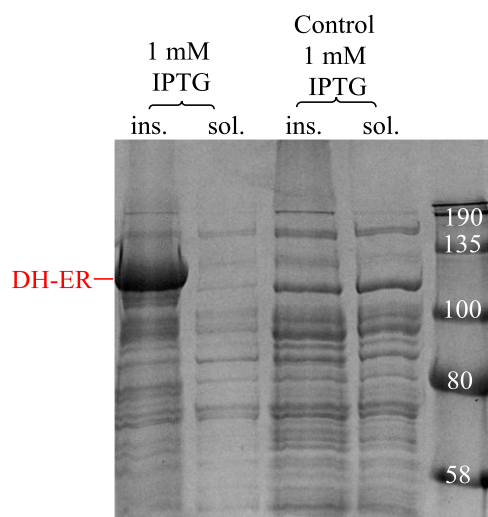


Figure 3.2. Expression of SQTKS DH-ER tridomain (135 kDa). Condition: *E. coli* BL21 (DE3), 1 mM IPTG, 37 °C, 180 rpm, 4 h, LB medium. Empty pET28a without target gene inserted was used as the control. Protein sample noted in red has been identified as DH-ER by ESI-QTOF-MS.

With the condition of incubation at 16 °C (250 rpm) for 36 h, auto-induction expression gives a modest improvement in the amount of soluble desired protein. However, no target protein was found after being purified by Ni-NTA (Figure 3.3). Most protein of desired size directly passed through the column without binding to the Ni-NTA. It is possible that the target protein may be unstable and precipitated in the column during the purification. For this reason, the column was further washed with 8 M Urea, but no further protein was eluted (Figure 3.3). This indicates the the His tag might be buried inside the target protein which made the Ni-NTA purification ineffective. In this case, expression of DH-ER at lower temperature to help slow protein folding might work.

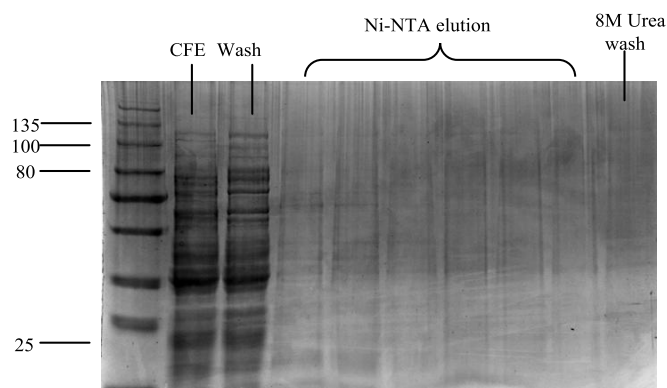


Figure 3.3. Ni-NTA of SQTKS DH-ER tridomain (135 kDa). Condition : *E. coli* BL21 (DE3), auto induction, 16 °C, 36 h, 250 rpm, auto-induction media LBE5052.

With the cold-adapted chaperonins, the expression of ArcticExpress cells can perform at a very low temperature. In this test, 10 °C was used as the standard incubation temperature. To avoid further inhibition of cell growth, no antibiotic was added into the expression culture at this temperature.

IPTG induction expression was first tested in ActiCExpress (DE3). Considerable insoluble protein of the expected target size was expressed. However, even by lowering the concentration of IPTG to 0.1 mM, no significant expression of soluble target protein was found (Figure 3.4).

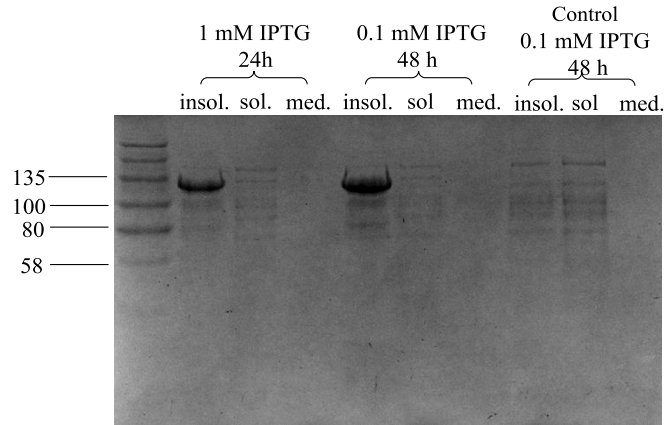


Figure 3.4. Expression of SQT KS DH-ER tridomain (135 kDa). Condition: *E. coli* ArcticExpress (DE3), IPTG, 10 °C, 250 rpm, LB medium. No protein was secreted into the medium (med.). Empty pET28a without the target gene inserted was used as the control.

In contrast to the use of IPTG, auto-induction in ArcticExpress (DE3) gave more soluble target protein when performed at 10 °C (250 rpm) for two days without kanamycin (Figure 3.5). The soluble and insoluble forms of the protein were both identified as the DH-ER tridomain of SQT KS by ESI-QTOF-MS (chapter 6).

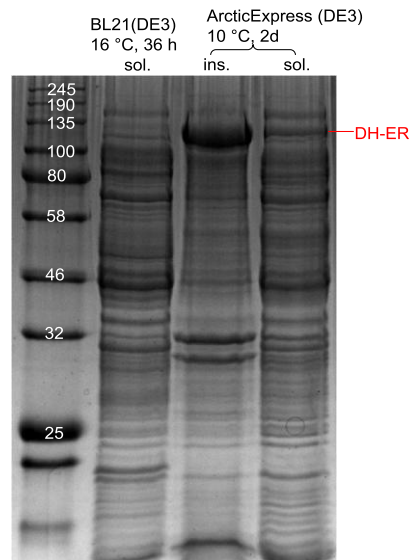


Figure 3.5. Expression of SQT KS DH-ER tridomain (135 kDa) in *E. coli* ArcticExpress (DE3) compared to that in *E. coli* BL21 (DE3). General condition: auto induction, 250 rpm, medium LBE 5052. The soluble form and insoluble form of the protein were both identified as the DH-ER tridomain of SQT KS by ESI-QTOF-MS.

As comparisons, different temperatures and incubation times were then tested for autoinduction (Figure 3.6). Incubation at 16 °C for 1 d gave a large amount of insoluble proteins. Compared to incubation at 10 °C for 2 d, incubation for 1 d gave much less expression. Therefore, time and temperature influence the expression significantly. Incubation at 10 °C for 2 d is the best condition.

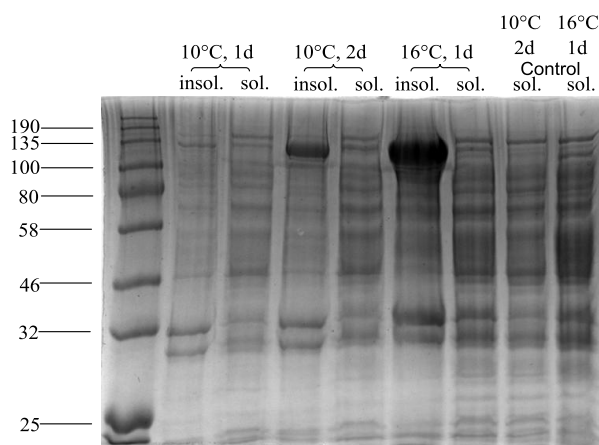


Figure 3.6. Time and temperature influence the expression of SQTCS DH-ER (135 kDa). General condition: *E. coli* ArcticExpress (DE3), auto induction, 250 rpm, medium LBE 5052. Empty pET28a without target gene inserted was used as the control, the SDS-PAGE comparison between 10 °C for one day and 16 °C for one day, comparison between 10 °C for one day and 10 °C for two days.

For the purification of the soluble protein, Ni-NTA chromatography was first used to capture the main his-tagged target protein. In addition to plenty of target protein, many other proteins were also contained in the elution (Figure 3.7 A), especially a protein of 58~80 kDa. As the Ni-NTA binds the His-tagged protein quite specifically, it is expected the elution should be purer. For this problem, there are two reasonable explanations. One common reason is that other impurities bind tightly with our target protein, for example the host cell proteins and DNA binding proteins which cannot be efficiently removed by Ni-NTA. Another possible reason is, our target protein is unstable and it has been lysed into several pieces during the purification. It is interesting that elongation of incubation time of the expression cell, such as incubation at 10 °C for 4 d (Figure 3.7 B), leads to a decrease of the amount of soluble target

protein and an increase of the amount of protein sized 58~80 kDa. To find out the correct explanation, identification of the protein (58~80 kDa) is needed.

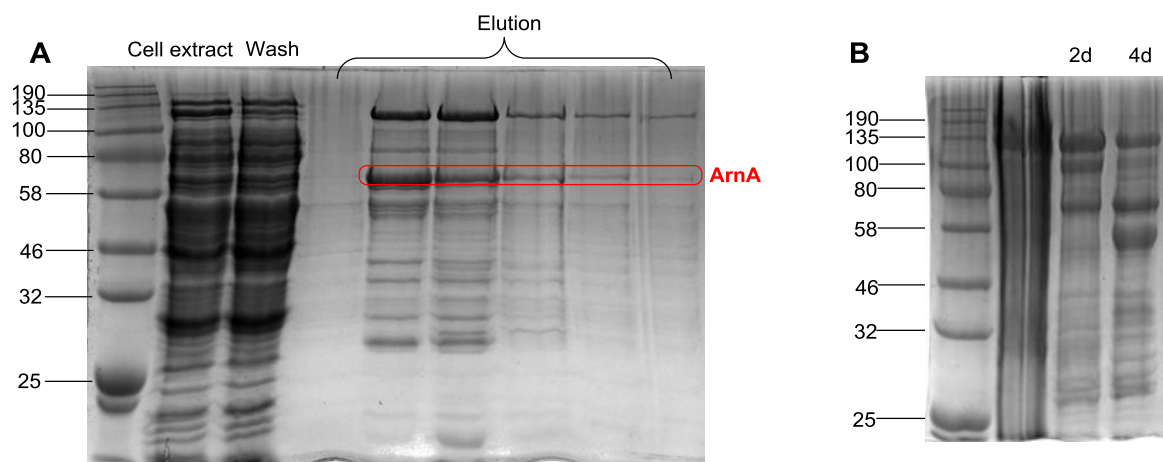


Figure 3.7. Ni-NTA purification of the SQTGS DH-ER tridomain (135 kDa): **A**, purification of DH-ER from incubation at 10 °C for 2 d (General condition: *E. coli* ArcticExpress (DE3), auto induction, medium LBE5052); **B**, comparison of purified DH-ER after Ni-NTA between the expression from incubation at 10 °C for 2 d and the expression from four incubation at 10 °C for 4 d. (General condition: *E. coli* ArcticExpress (DE3), auto induction, medium LBE5052).

ESI-QTOF mass spectroscopy identified the contaminating protein as an *E. coli* bifunctional polymyxin resistance protein ArnA (see section 3.3.3), not a part of our target protein or the cold-adapted chaperonin (Cpn60, size of 57 kDa) used in *E. coli* ArcticExpress (DE3). To remove the contaminant proteins, size exclusion chromatography was used as a further purification method. For different types of gel filtration column, the separation abilities are different. For a specific column, the separable size of protein is between a maximum size and a minimum size, any protein of the size bigger than the maximum size or lower than the minimum size cannot be separated. The SEC column used in this research is a HiLoad 26/600 Superdex 200 pg, of which the separable protein range of 10 - 600 kDa.

Before using the size exclusion chromatography, calibration of the relationship between protein size and elution volume is needed. The calibration is not accurate, as not only the molar mass of the protein can influence the elution volume, but also the shape of the protein. However, it can be used as a reference to know what the native size of a separated protein is in solution, which can give more information about the quaternary structure of the protein.

Here, the calibration (Figure 3.8) was performed under the same buffer and flow velocity used in purification of the SQTGS DH-ER. The void volume (V_0) was measured using by Blue Dextran (more than 2000 kDa). If an elution volume of a protein under investigation is the same as the void volume, the size of this protein is beyond the maximum size of protein the column can separate. The relationship between protein size and elution volume (V_e) was certified by a series of commercial protein with determined size (Ferritin: 443 kDa, Bovine Serum Albumin (BSA): 66 kDa, Carbonic Anhydrase: 29 kDa.)

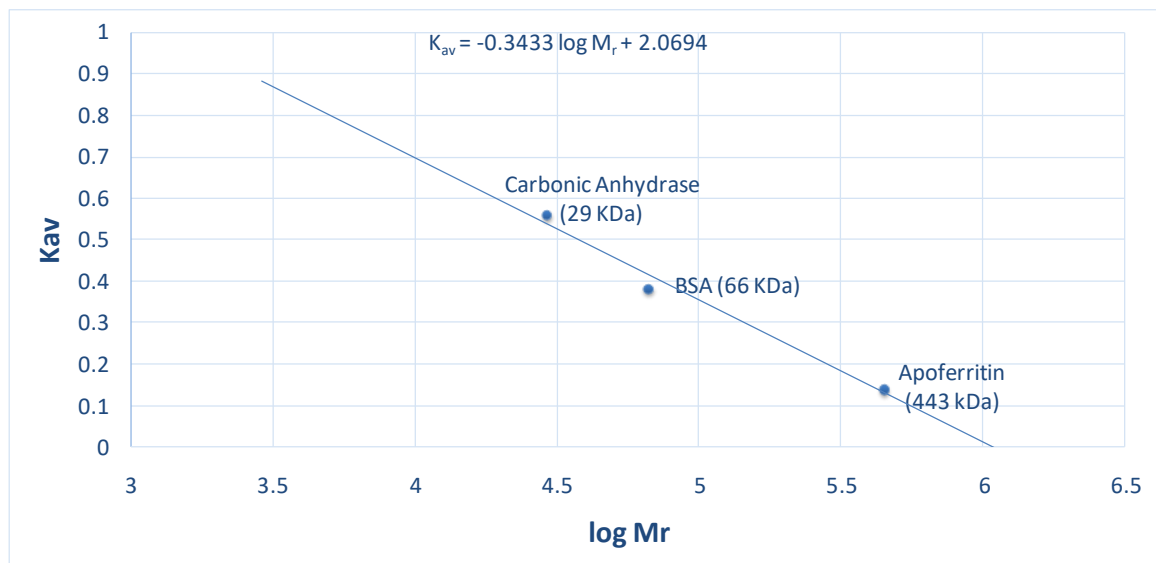


Figure 3.8. Calibration for GE HiLoad 26/600 Superdex 200 pg, (Velocity: 1mL/min. Buffer: pH 8.0, 50 mM Tris, 150 mM NaCl. 20 % Glycerol). Partition coefficient: $K_{av} = (V_e - V_0)/(V_c - V_0)$. V_e (Elution volume), V_0 (void volume 114.8 mL by Blue Dextran), V_c (geometric column volume, 320 mL). The calibration results is, $(V_e - 114.80)/205.2 = 2.0694 - 0.3433 \log M_r$.

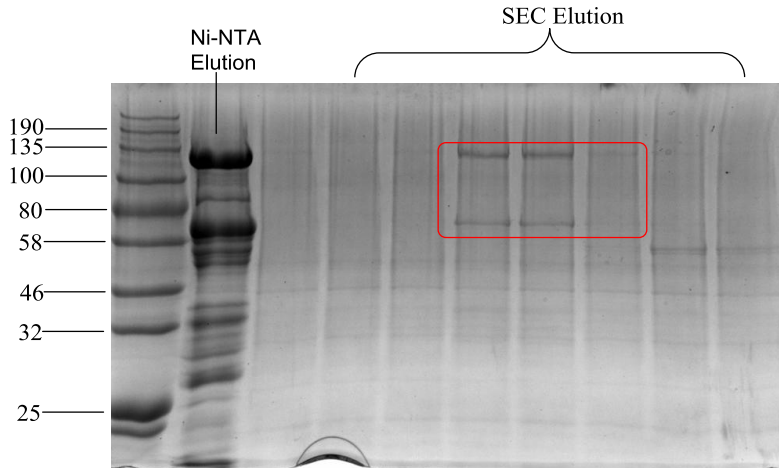


Figure 3.9. Size exclusion chromatography of the SQTGS DH-ER tridomain (135 kDa). Proteins in red box are the main products and cannot be separated.

According to the result of size exclusion chromatography (Figure 3.9), the extra protein ArnA was still combined with DH-ER. Surprisingly, the elution volume of the combined protein is 114.5 mL which is almost the same as the void volume (114.8 mL). This means the molecular weight of the combined protein mixture almost exceeds the size exclusion limit (600 kDa) of the packing material. However, the expected monomer size of the DH-ER tridomain is 135 kDa and the monomer size of ArnA is 74 kDa. It is thought that the DH-ER should form a dimer (270 kDa) in solutions according to the modelling structure of SQTGS (Figure 2.6). However, the results of this chromatography suggest that a multimeric $(\text{DH-ER})_m(\text{ArnA})_n$ species may be formed such that $135m+74n \geq 600$.

The isolation of soluble SQTGS DH-ER tridomain in *E. coli* is therefore partially successful. Co-expression with chaperonins, low-temperature expression, the use of ArcticExpress (DE3) and auto-induction were all required to obtain soluble DH-ER protein. However, at the same time, the *E. coli* protein ArnA was also expressed in the system and combined tightly with DH-ER.

A method to isolate the DH-ER tridomain from the contaminant protein ArnA (section 3.3.3) is adding more Mg^{2+} . However, addition of 10 mM of MgCl_2 into the culture medium

does not suppress the formation of the protein sized 58~80 kDa. The protein can still be found after Ni-NTA purification (Figure 3.10). To remove the protein, additional protein modification may be needed. The difficulty in purification may also be due to the fact that the DH-ER is not stable and it shows low activity, which will be discussed in section 4.4.1. For performing robust *in vitro* research, more stable and active proteins are needed.

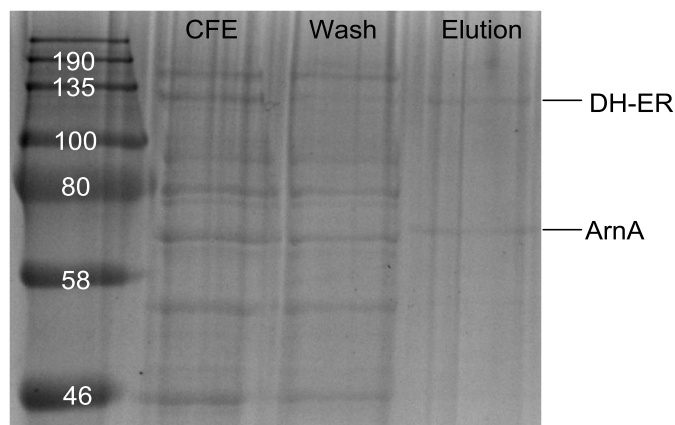


Figure 3.10. Ni-NTA purification of the SQTCS DH-ER (135 kDa) expressed in medium LBE5052 with 10 mM MgCl₂. CFE refers to cell free extract.

3.2.2 DH-KR Tetradomain

The isolated DH-ER tridomain of SQTCS is not stable enough for further *in vitro* research, therefore we decided to express a larger tetra-domain section, DH-CMeT-ER-KR (DH-KR). DH-KR is the whole upper segment of the SQTCS, containing all the β -modification domains (chapter 2). Compared to that of the DH-ER tridomain (135 kDa), it has less exposed hydrophobic surface and this is expected to make the protein more stable after isolation. However, with a size of 166 kDa, the chance to get a soluble and active protein is less.

Guided by the expression methods used for the DH-ER tridomain (section 3.2.1), the same four expression conditions (ArcticExpress (DE3) + IPTG (0.1 mM, 16 °C); BL21 (DE3) + IPTG (0.1 mM, 16 °C); ArcticExpress (DE3) + Auto-induction (10 °C); BL21 (DE3) + Auto-induction (16 °C)) were tested first (**Figure 3.11**). The control was *E. coli* BL21 (DE3)

transformed with empty pET28a. We observed that plenty of soluble protein of desired size (166 kDa) was produced by BL21 (DE3) by auto induction at 16 °C (Figure 3.11) for 24 h. The protein in the red box was cut from the SDS gel and identified as N-His₆-DH-KR by ESI-QTOF-MS (Figure 3.25).

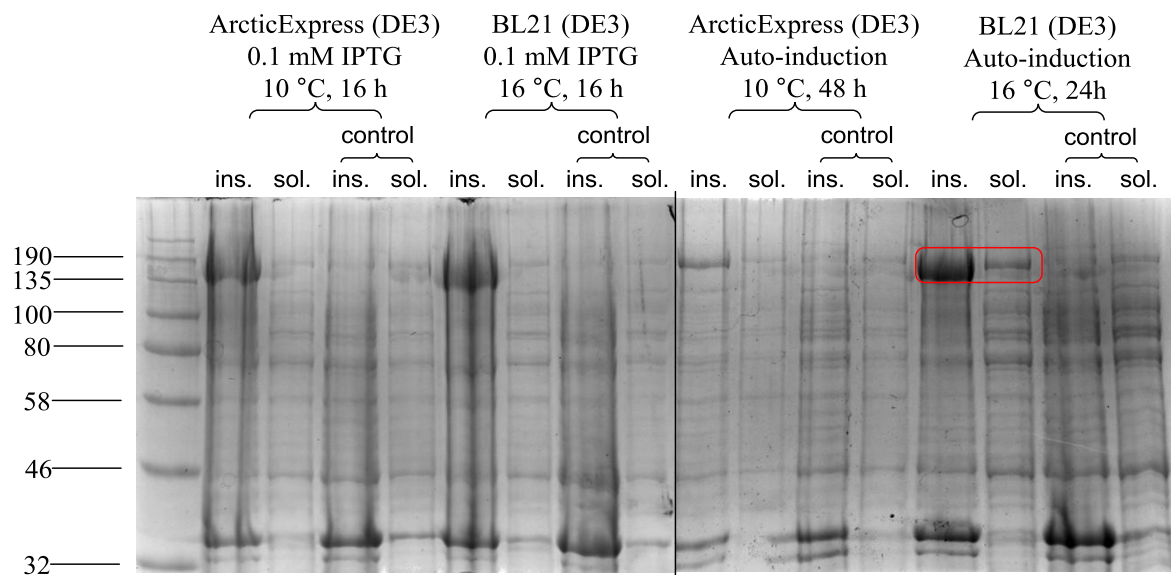


Figure 3.11. Expression of SQT KS DH-KR (166 kDa) by four different expression conditions. General condition: 200 rpm, IPTG (LB medium), auto induction (medium LBE 5052).

After the Ni-NTA purification, soluble DH-KR was found still with many impurities (Figure 3.12). Interestingly, incubation time affects the protein expression and further affects the purity of the DH-KR tetradomain. With the extension of cell culture time, more and more target protein formed in accord with the increase of host cell protein. This was confirmed by extracting the Ni-NTA purified proteins from the same-amount sample taken from cell culture for 6 h, for 12 h and for 24 h. The ratio between the amount of target protein and that of impurities increased considerably over time. This indicates that the majority of impurities may not be the DH-KR-bound host cell proteins, the amount of which should increase as the amount of DH-KR increases. The impurities may come from the DH-KR itself. A possible explanation is that large-protein folding might take several hours.^[155] At the early stage of

expression, the folding of DH-KR is not complete which makes the protein fragile and easily digested into fragments such as the protein sized ~80 kDa. As the folding completes, less and less ratio of fragmentation to mature protein is observed.

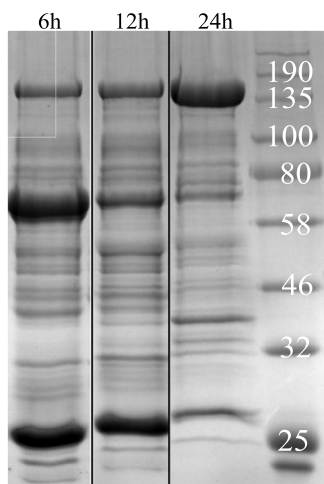


Figure 3.12. Incubation time affects the protein (soluble) expression and further affects the purity of SQTKS DH-KR tetradomain (166 kDa) after Ni-NTA chromatography.

To increase the chance of getting protein of higher purity, intermediate purification was performed by IEX chromatography. Diethylaminoethyl (DEAE) sepharose was used, which is a weak anion exchanger. This material, in contrast to a strong ion exchanger, displays pH-dependent function and can deliver optimal performance over a small pH range. In pH 8 storage buffer (section 3.3.3.2), the DH-KR tetradomain (pI 5.86) is negative charged, and should bind to the positively charged resin. During the sample application, some weakly bound proteins were washed through the column (Figure 3.13 B 1). By increasing the concentration of NaCl from 50 mM to 1000 mM, more strongly bound proteins were gradually eluted. However, even with low concentration of inject sample, that is, injecting 0.2 mg of DH-KR purified by Ni-NTA into 5 mL DEAE IEX column, the separation was poor. From SDS-PAGE, two fragments of eluted proteins 1 and 2 (Figure 3.13 B) were compared to the starting protein purified by Ni-NTA. No considerable improvement of purity of DH-KR was observed (Figure 3.13).

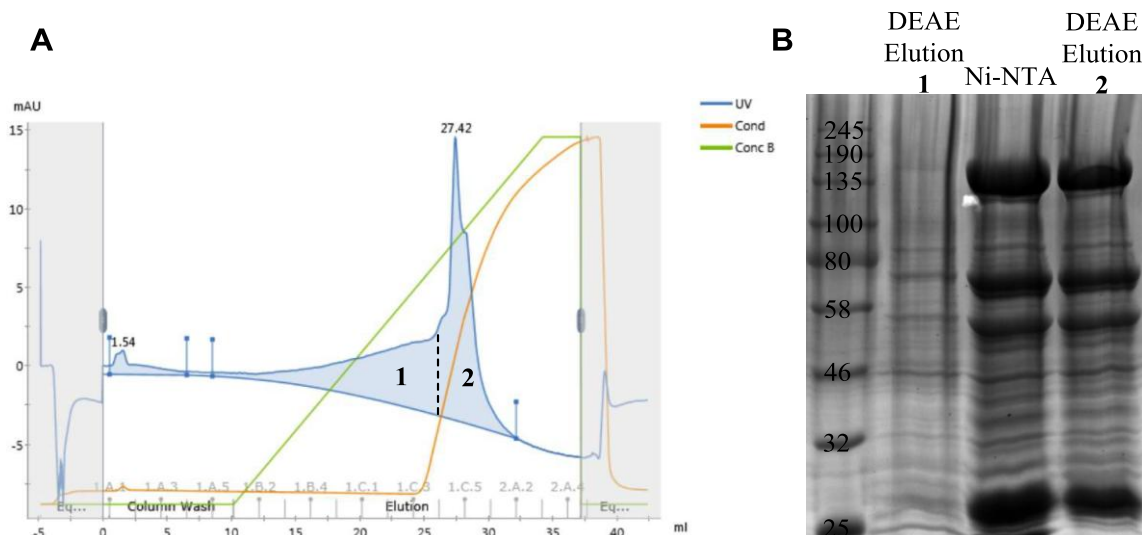


Figure 3.13. Ion exchange chromatography of the SQTKS DH-KR tetradomain: **A**, the chromatogram of IEC by DEAE-Sepharose. The concentration of the elution buffer (storage buffer with 1.2 M NaCl) increases from 0 % to 100 % (green) with the change of conductivity (orange); **B**, A-related protein elution.

Afterwards, size exclusion chromatography was used for polishing purification. The DH-KR tetradomain cannot be separated from the protein sized ~80 kDa (Figure 3.14 B). It is observed that the ratio between these two proteins changes in different fractions of elution (from left to right – elution order). It is possible that there is a balance between these two proteins, which makes the protein sized ~80 kDa difficult to be removed. Supporting evidence comes from the size exclusion calibration. The size of the DH-KR tetradomain in solution is predicted as 350~500 kDa (Figure 3.14 A), which means the native DH-KR (monomer 166 kDa) may exist as a dimer (expected 332 kDa) or a more complicated structure which may also bind the protein sized ~80 kDa.

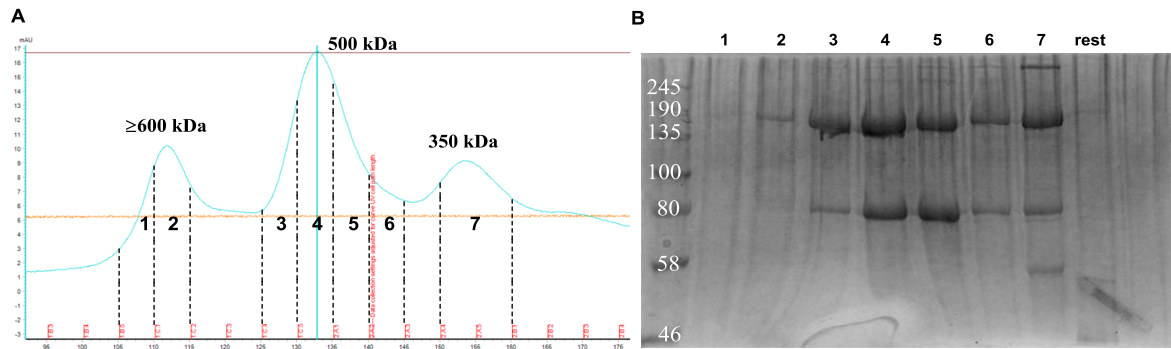


Figure 3.14. Purification of the SQTCS DH-KR tetradomain by size exclusion chromatography: **A**, the chromatogram of SEC; **B**, A-related protein elution (left to right).

To identify the protein sized ~80 kDa, ESI-QTOF-MS analysis is essential (shown in Figure 3.29). However, before that, the effect of cell lysis and host cell protease should be eliminated first. Considering that high sonication power during the cell lysis may break the protein, different sonication amplitudes (Am) were used (Figure 3.15). From the result, the effect of sonication power is similar from 10 % to 50 %, no significant change in the amount of the protein sized 58~80 kDa was observed. Even without the sonication treatment (0 % Am), the protein contamination still exists. This means the cell can be easily lysed and sonication is highly effective, and sonication is not the reason for the formation of the protein sized 58~80 kDa.

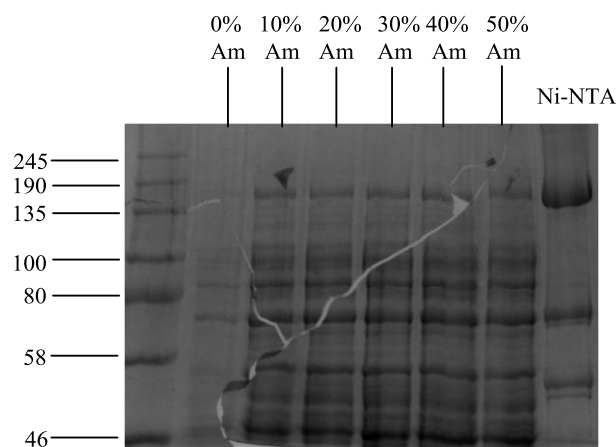


Figure 3.15. Sonication power effect in preparing cell free extraction (CFE) of the SQTCS DH-KR tetradomain (166 kDa).

In case of any *E. coli* protease can break down DH-KR, EDTA-free protease inhibitors (AEBSF - inhibits serine protease; bestatin - inhibits arginyl aminopeptidase; E-64 - inhibits cysteine protease; pepstatin A - inhibits aspartyl proteases; phosphoramidon - inhibits thermolysin; leupeptin - inhibits cysteine, serine and threonine peptidases; and aprotinin - inhibits trypsin, chymotrypsin and plasmin) were added before breaking the cell wall of *E. coli* by sonication. However, under this treatment, the production of DH-KR decreased while the amount of the contaminant protein 58~80 kDa increased (Figure 3.16), compared to the previous DH-KR purified by Ni-NTA without any protease inhibitor (control in Figure 3.16). The addition of protease inhibitors makes the protein unstable that almost no DH-KR (166 kDa) was found after purification by size exclusion chromatography. So far we are not able to explain this odd phenomenon.

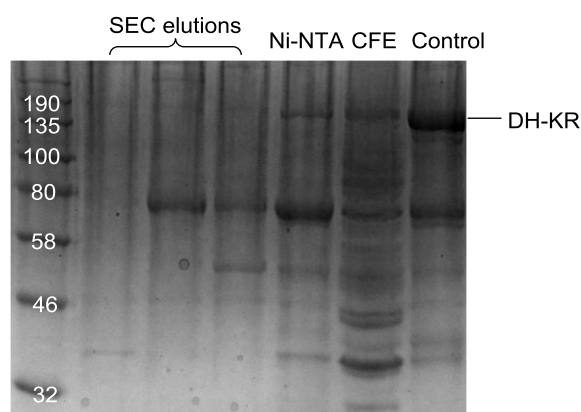


Figure 3.16. The purification process of DH-KR after addition of protease inhibitor cocktail: from cell free extract (CFE), to the protein purified by Ni-NTA, to the protein purified by size exclusion chromatography. Control refers to the previous protein purified by Ni-NTA without addition of protease inhibitors.

3.2.3 Structure of SQTKS DH-KR Tetradomain by SAXS

Small-angle X-ray scattering (SAXS) is a biophysical method used to study the overall size (kDa ~ GDa) and shape of biological macromolecules in solution at low resolution (10 Å ~ 50 Å). A solution of particles is illuminated by a collimated monochromatic X-ray beam, the

intensity of the scattered X-rays (with a scattering angle of 2θ) is recorded by an X-ray detector. Collaborating with the Weissman group at the University of Lorraine, high brilliance X-ray was obtained by ESRF synchrotron and used for this SAXS experiment to optimize the scattering intensity. The scattering pattern of the pure solvent is collected and subtracted from the sample solution scattering leaving only the signal from the particles of interest. The resulting isotropic scattering pattern is related to the overall shape and size of the particles under investigation.^[137]

The scattering intensity I is represented as a function of momentum transfer s . The value of $s = 4\pi\sin\theta/\lambda$ and λ is the beam wavelength. For very small scattering angles ($s < 1/R_g$) the scattering intensity can be described by Guinier approximation. The Guinier plot of $\log(I(s))$ against s^2 will give a straight line from which R_g and $I(0)$ ($s=0$) can be estimated. R_g , the radius of gyration, is the average root-mean-square distanced to the centre of density in the molecule weighted by the scattering length density. $I(0)$ is proportional to the molecular weight and the concentration of protein. If the Guinier plot cannot give a straight line, the protein aggregates. Moreover, with more detailed scattering intensity data available within the scattering angle 2θ , the shape of the protein can be described by the Pair-Distribution Function $P(R)$. The $P(R)$ function is a Fourier transform of the scattering curve $I(s)$, which is a histogram of distances between all possible pairs of atoms within a protein. With this function, more reliable R_g and $I(0)$ can be calculated. In addition, the maximum linear dimension D_{max} is obtained when $P(R)=0$. With these data and analysis, a low resolution model can be built by *ab initio* and rigid body modeling. This will find the best theoretical scattering curve based on the model to fit the experimentally scattering plot $I(s)$.

Good-quality data were obtained by SAXS (Figure 3.17). As pure SQTCS DH-KR tetradomain was impossible to obtain at that time, the protein purified by size exclusion chromatography with less than satisfactory purity was directly used as the sample for SAXS. The Guinier plot gave a straight line which indicates the protein is stable in the test environment and no aggregate forms. From the Pair-Distribution Function, R_g was determined as 65 Å and D_{max} was 285 Å. A low resolution modeling structure was obtained as a “turtle”

shape. However, the molecular weight of the protein was determined as 400 ~ 600 kDa which means the protein forms a trimer or even more complex structure (the monomer is 166 kDa). This makes the location of domains of SQTCS DH-KR in the modeling structure difficult. To settle this problem, other purified domains are needed to isolate and analyzed by SAXS as a comparison. Another issue, we still cannot be certain whether the protein is monodispersed or polydispersed, which is important to the reliability of all the result obtained by SAXS. Therefore, more detailed understanding of this protein is needed. This will be discussed in section 3.3.

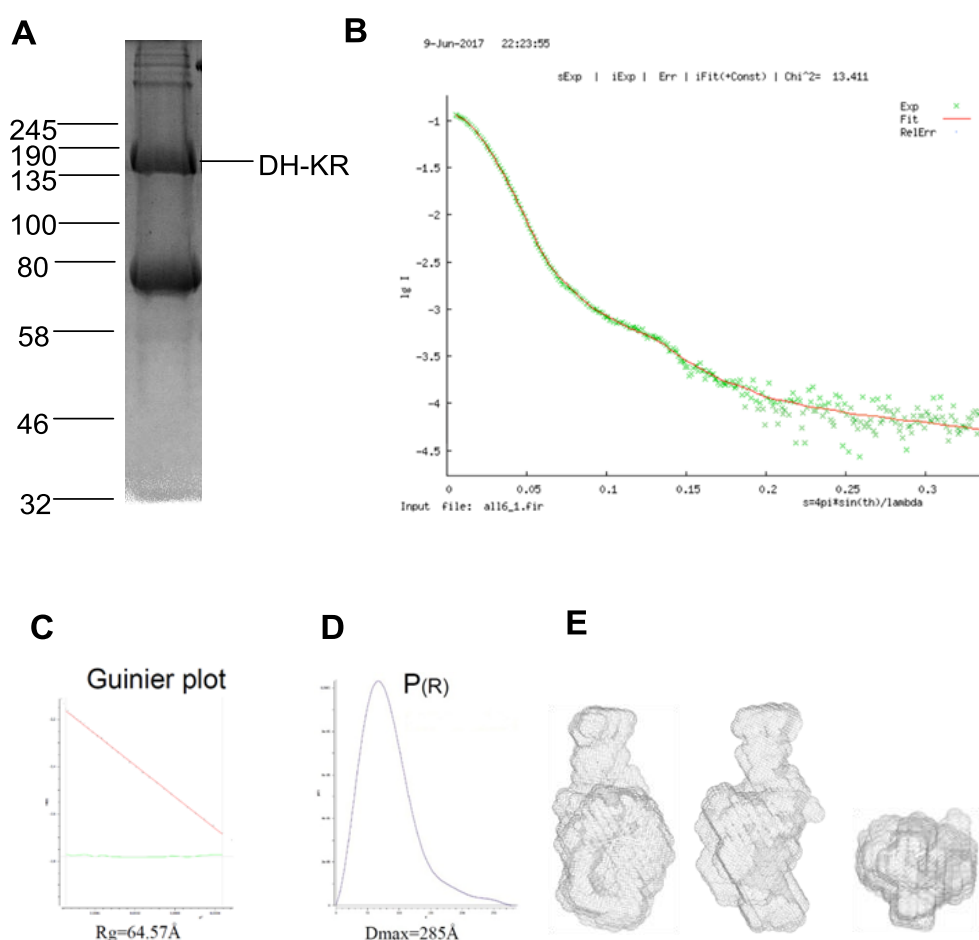


Figure 3.17. SAXS analysis of SQTCS DH-KR tetradomain (166 kDa): **A**, quality of sample protein (puried by size exclusion chromatography); **B**, original plot of $\log(I(s))$ against s by SAXS (green dots), the red line is the fitting curve adjusted by structure modeling; **C**, the Guinier plot of $\log(I(s))$ against s^2 ; **D**, The plot of Pair-Distribution Function $P(R)$; **E**, the final model of SQTCS DH-KR in different views. Data were provided by Weissman group.

3.2.4 CMeT Monodomain

As described in chapter 1, the SQTKS CMeT (48 kDa) plays an important role in the programming of the biosynthesis of squalestatin tetraketide. It is interesting to isolate the single CMeT domain and study its methylation activity in comparison with the activity of CMeT in the tetradomain DH-CMeT-ER-KR (DH-KR). Four His₆-CMeTs with different boundaries (MT1, MT2, MT3 and MT4) have been built (chapter 2), but only MT3 was successfully isolated as soluble protein.

In addition to the usual four expression conditions (ArcticExpress (DE3) + IPTG (0.1 mM, 16 °C); BL21 (DE3) + IPTG (0.1 mM, 16 °C); ArcticExpress (DE3) + Auto-induction (10 °C); BL21 (DE3) + Auto-induction (16 °C)) which have been used to improve the chances of obtaining soluble proteins, varied temperature and media were used for IPTG-induced expression.

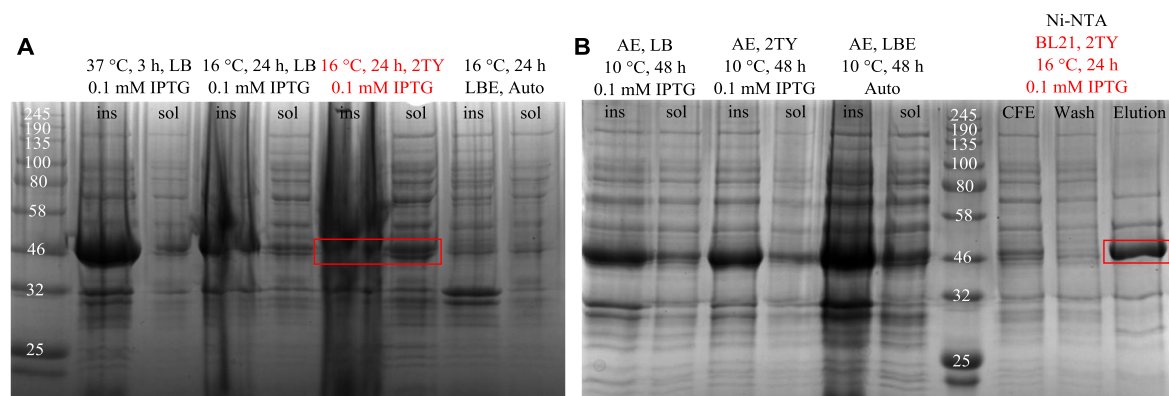


Figure 3.18. Expression and purification of SQTKS CMeT (48 kDa): **A**, expression in BL21 (DE3); **B**, expression in ArcticExpress (DE3) and Ni-NTA purification of the CMeT expressed in BL21 (DE3) with a condition of 2 TY, 0.1 mM IPTG, 16 degree, 200 rpm for 24 h. General condition: 200 rpm.

By comparing the amount of protein of target size between insoluble fractions and soluble fractions after cell lysis, IPTG-induced expression in 2 TY medium by *E. coli* BL21 (DE3) was selected as the best condition for production of soluble target protein (Figure 3.18

A, B). In comparison, other expression conditions produced more insoluble forms of the target protein.

The target protein was purified by Ni-NTA chromatography (Figure 3.18 B). The protein in the red box was identified as SQTKS CMeT by MALDI-TOF-MS (chapter 6). Just like the DH-KR tetradomain, there are still many impurities along with CMeT. Attempts to purify the protein by size exclusion chromatography failed. Very little protein was obtained without improvement in purity (Figure 3.19). This means the protein is unstable in the column and further purification after Ni-NTA is not recommended. The molecular weight of CMeT in the buffer was estimated as 140 kDa by calibration (see chapter 6), which means the protein may form a trimer in solution. In case that CMeT has been lysed by host cell proteases, EDTA-free protease inhibitor cocktail was added before breaking the cell lysis. However, same as SQTKS DH-KR, an unexpected decrease of purity was observed. Further understanding and purification of the protein will be discussed in section 3.3.

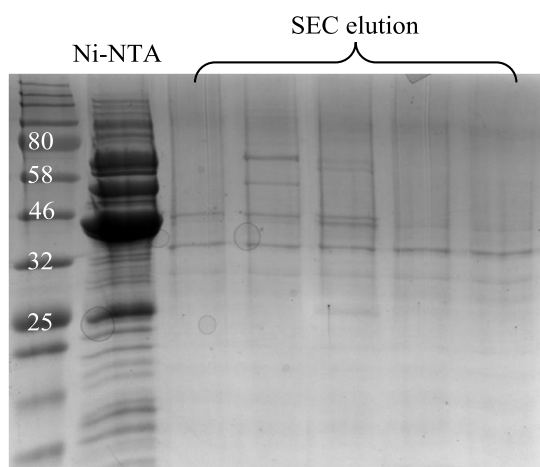


Figure 3.19. Purification of SQTKS CMeT by size exclusion chromatography. Elution order: from left to right. Elution buffer: pH 8.0 50 mM Tris, 150 mM NaCl, 20 % glycerol, 1 mM EDTA

3.2.5 *E. coli* Codon Sub-optimized DH-KR Tetradomain

The purity of the SQTK DH-KR tetradomain is compromised, which may affect its activity. As shown in Chapter 2, by adjusting the gene sequence to mimic the codon usage in the

original host cell where SQTKS forms, the folding of DH-KR may be enhanced. By using the same expression conditions (BL21 (DE3) + pET28a + auto-induction) used for the expression of the *E. coli* full codon optimized DH-KR tetradomain, no over-expressed target protein has been observed either in soluble form or as inclusion bodies. To increase the chance of getting soluble protein, expression vectors with different soluble fusion partners, such as pETM 22 with thioredoxin A (TrxA) and pETM 11 with SUMO3, have been built in chapter 2. Much effort has been spent. In addition to basic four expression conditions (ArcticExpress (DE3) + IPTG (0.1 mM); BL21 (DE3) + IPTG (0.1 mM); ArcticExpress (DE3) + Auto-induction; BL21 (DE3) + Auto-induction) which have been used before, more varied temperature and media were used for expression.

As a result, pETM22 with a TrxA fusion tag in both strains can only produce trace amount of soluble SQTKS DH-KR tetradomain with a great amount of the inclusion body. As thioredoxin can enhance the formation of disulfide bonds in protein, this phenomenon indirectly proves that the folding of DH-KR cannot be further improved by promoting disulfide bond formation. Therefore, efforts to improve protein folding by using *E. coli* strains like Shuffle^[168] or Origami^[169] which can promote disulfide bond formation is less likely to be helpful. This also indicates that the SQTKS DH-KR is produced in the periplasmic space in *E. coli* and the formation of disulfide bond is good in this environment.

Finally we found two good conditions (Figure 3.20). The first one is auto-induction from the vector pETM 11 (SUMO). The second one is IPTG induction in TB medium from the vector pET28a. Both can produce enough soluble SQTKS DH-KR tetradomain from *E. coli* BL21 (DE3) at 16 °C for 24 h. N-His₆-SUMO-DH-KR and N-His₆-DH-KR were purified by Ni-NTA (Figure 3.20). Cut by SUMO protease, DH-KR without any tag has been separated. However, compared to N-His₆-DH-KR, the purity of DH-KR is much worse. In addition, the purity of N-His₆-DH-KR expressed from sub-codon optimized SQTKS and of the same protein expressed from fully codon optimized SQTKS is of no considerable difference. From the protein identification by ESI-QTOF-MS, the component of the main protein sized ~80 kDa (Figure 3.29) and the protein sized 46~58 kDa are almost the same and both contain

fragments of DH-KR. Therefore, the fragmentation of DH-KR still exists and the purity of the protein is still compromised.

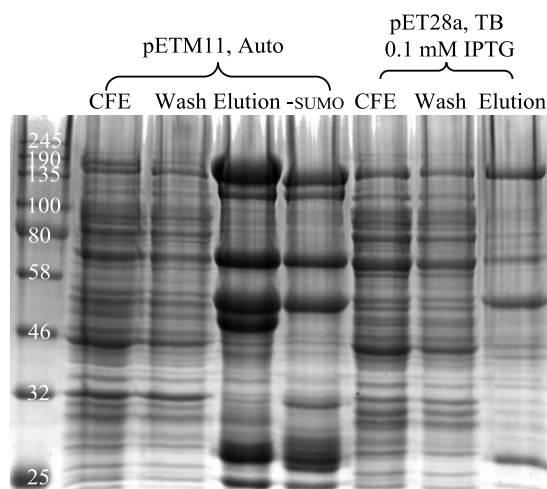


Figure 3.20. Expression and purification of sub-codon optimized SQTCS DH-KR (166 kDa). General condition: BL21 (DE3), 16 °C, 24 h. Elution buffer: pH 8.0 50 mM Tris, 20% Glycerol, 150 mM NaCl, 1 mM EDTA.

3.2.6 DH-ACP Pentadomain

ACP plays an important role in transferring the substrate among different domains. ACP may be also essential for the stability of the full SQTCS. Attempt to isolate the DH-CMeT-ER-KR-ACP pentadomain (DH-ACP, 176 kDa) has been done. However, no target soluble protein has been observed so far.

Guided by previously successful expression conditions, the DH-ACP gene fragment was only inserted into pET28a and expressed in *E. coli* BL21 (DE3) with auto-induction in LBE medium (Figure 3.21 B). In addition, IPTG-induced expression in different media such as LB, 2TY (Figure 3.21 A) and TB (Figure 3.21 B) was attempted. Large amount of the insoluble protein formed. Decreasing the induction temperature did not bring any improvement to the amount of related soluble protein. No His-tagged protein was observed after Ni-NTA purification.

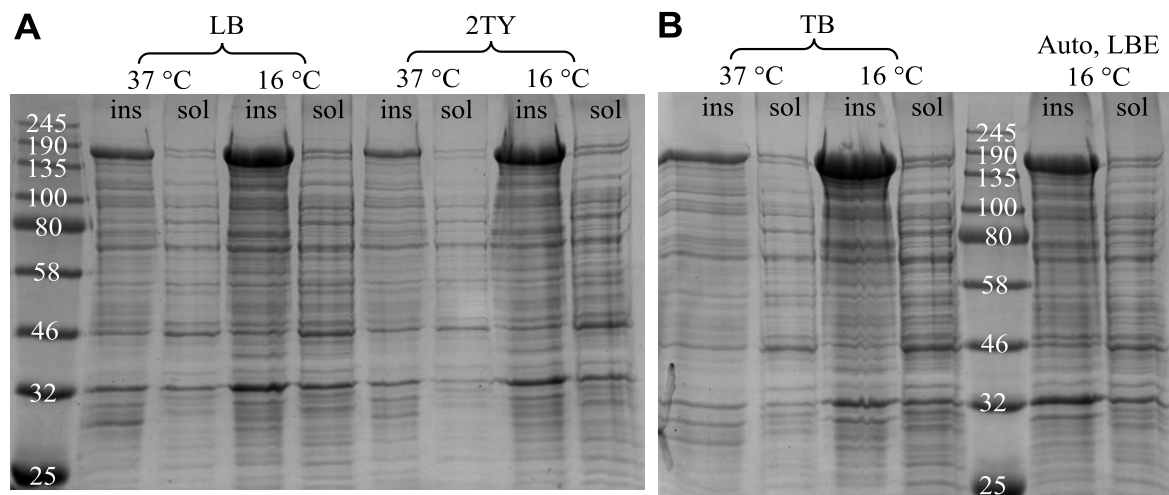


Figure 3.21. Expression of DH-ACP pentadomain (176 kDa) in *E. coli* BL21 (DE3): **A**, in LB medium and 2TY medium, 0.1 mM IPTG; **B**, in TB medium (0.1 mM IPTG) and LBE (auto induction). General condition: 200 rpm, 37 °C for 3h, 16 °C for 24h.

3.3 Optimization of Purification

Although different *E. coli* strains, expression vectors with different fusion tags, various expression conditions and different chromatography have been performed, the purity of isolated proteins such as the SQTKS DH-KR tetradomian and the SQTKS CMeT monodomain is still less than satisfactory. Further optimization can focus on understanding the component of impurities and finding solutions to remove these impurities. In addition, the storage buffer of these proteins should be optimized to balance the stability and the activity. Moreover, attempts to obtain structural information have been made to better understand the behavior of these proteins.

3.3.1 Storage Buffer Optimization

3.3.1.1 SQTKS DH-KR Tetradomain

To make purified proteins stable, suitable storage buffer is a vital factor. In our previous studies of SQTKS ER^[75] and DH^[57] monodomains, 20 % glycerol was used to help stabilize the ER and polar amino acids were used to stabilize the DH. These amphiphilic molecules were selected to improve interactions between exposed hydrophobic surfaces and the aqueous buffer. It is suggested that reactivity and thermo-stability of the purified protein should be considered at the same time, as strong reactivity does not always mean strong stability, and *vice versa*. Due to this, a systematic selection of buffers was tested.

Firstly, different additives were used for overcoming the exposure of protein hydrophobic surfaces. Based on the previous SQTKS ER studies^[75] (section 4.11), we can monitor the ER reactivity of the SQTKS DH-KR tetradomain by UV change (340 nm) directly (Figure 3.22). Additives were added to the reaction buffer to test any reactivity change. The activity assay was set up to contain 10 μ L of 2-methylhexenoyl pantetheine **65** (0.125 mM) and 10 μ L of NADPH (0.25 mM) in 370 μ L of buffer containing different additives. The reaction was initiated by adding 10 μ L of SQTKS DH-KR tetradomain (5 mg/mL stock) purified by Ni-NTA and the product **66** was formed with a decrease of UV absorption at 340 nm. Initial rates of the reaction were taken to reflect the reactivity. Compared to reaction buffer without additives, we only choose the additives which can make the ER reactivity increase.

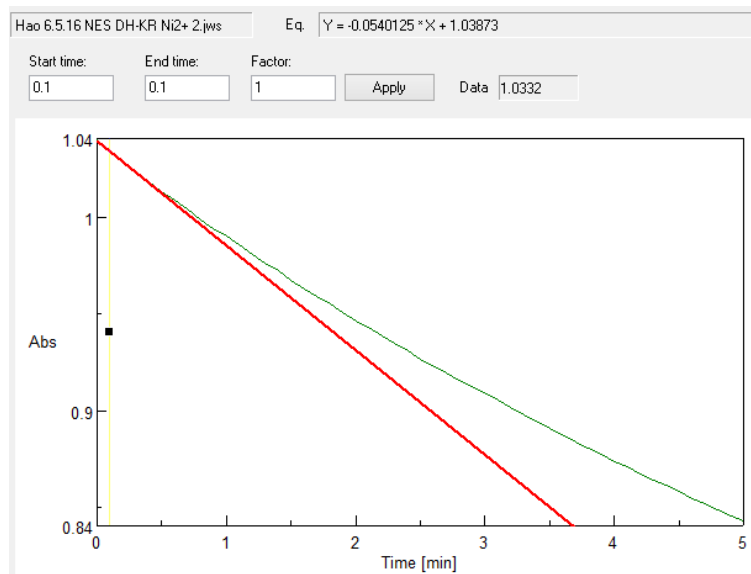
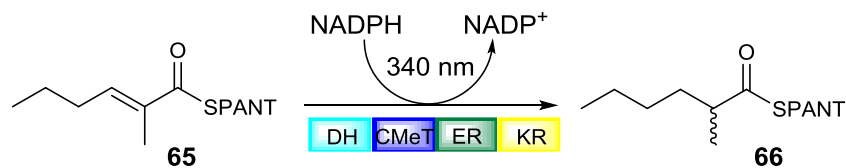


Figure 3.22. The UV assay used in the enoyl reduction catalyzed by SQTCS DH-KR.

The additives are detergents, metal chelators, reducing agents, salts and osmolytes. When protein aggregation arises in hydrophobic interactions, one option for disruption of these interactions is a non-denaturing detergent (such as tween 20),^[138-139] which can also protect proteins from sudden change of pressure when getting through a syringe. Metal chelators (such as EDTA) can be used for reducing oxidation damage and chelating metal ions. Reducing agents such as 2-Mercaptoethanol and tris(2-carboxyethyl)phosphine (TCEP) can also be used to reduce oxidation damage. It is advisable to use salts at low concentrations, because most bacterial and mammalian enzymes function under low salt conditions and are inhibited by high salt condition. Under high salt concentrations, proteins precipitate by competing for water molecules such that the hydration shell around the protein cannot be maintained. This process is called salting-out, which is commonly used in protein isolation and in protein crystallization.^[140-141] Under low salt concentrations, protein solubility increases, which is called salting-in. Divalent cations (such as Mg²⁺) can exert a powerful

effect on protein solubility. In general, large ions of high valence interact with water molecules strongly and are strongly hydrated, such as citrate, sulfates, acetate, and phosphates, which are better at solubilizing proteins than small ions of low valence ions such as chloride and nitrates. All organisms except halobacteria have evolved a response to denaturing stresses such as high temperature, desiccation, the presence of denaturants, and high osmotic pressure. The response involves intracellular production and accumulation of high levels of low molecular weight organic compounds called osmolytes ^[142-143] which can stabilize macromolecules and conserve their biological activity. Osmolytes can be polyols such as glycerol, sugars (sucrose, trehalose), polysaccharides, DMSO, neutral polymers, amino acids and their derivatives, and large dipolar molecules such as trimethylamine *N*-oxide (TMAO). ^[144] For example, glucose can gradually form direct contacts with the protein hydrophobic surface, which stabilizes the protein by preferential hydration. ^[145] Small neutral amino acids such as proline, glycine and alanine exhibit a concentration-independent degree of preferential hydration. ^[146] Studies have shown specific lipid (such as asolectin) are necessary for structural integrity and proper function of membrane proteins. Effects and types of protein–lipid interactions are diverse. Most of them have a chaperonin-like function, similar as the function of bovine serum albumin (BSA) in protein folding, or directly involve in the molecular mechanism such as enzymatic activity or transport processes across the membrane. ^[147]

As a control, the ER reaction was started with buffer (50 mM Tris, pH 8.0, 150 mM NaCl, 20 % Glycerol) without any additive, then continued with buffer which contained different types of additives (Figure 3.23). In all, most additives decrease the reactivity while addition of proline or EDTA can improve the reactivity slightly. We found 2-mercaptoethanol significantly slows down the reaction. One of reason is, the thioester of the substrate might be fragile when exposed to the strongly nucleophilic 2-mercaptoethanol. Therefore, we have to use reducing agent without strong nucleophilicity such as TCEP in the future. The adding of MgSO₄ also decreases the enzyme reactivity greatly while EDTA increases the reactivity. This clearly indicates that Mg²⁺ deactivates the enzyme. Addition of proline can also increase

the reactivity. However, increasing the amount of EDTA can further speed up the reaction while increasing the amount of proline cannot. Other additives, such as BSA, DMSO, Asolectin (lipid), Tween 20 and glucose, are not able to influence the reactivity of the enzyme effectively.

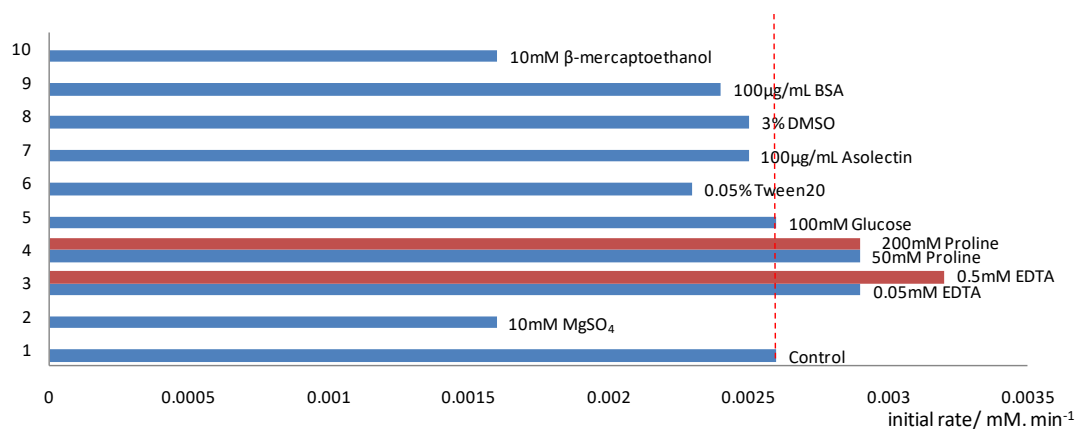


Figure 3.23. Additives influence the enoyl reduction reactivity of the SQTKS DH-KR tetradomain (control buffer: 50 mM Tris-HCl, pH 8.0, 150 mM NaCl, 20 % Glycerol). UV assay: 20 °C, 370 µL of Buffer (50 mM Tris-HCl pH 8.0, 150 mM NaCl, 20 % Glycerol, additive), 10 µL of NADPH (0.25 mM), 10 µL of 2-methylhexenoyl pantetheine (0.125 mM), 10 µL of DH-KR (5 mg/mL stock, 40 %purified by Ni-NTA).

After considering the effects of additives towards the enoyl reduction catalyzed by the SQTKS DH-KR tetradomain, suitable additives such as EDTA was selected to perform the thermo-stability test for buffer optimization. A thermal shift assay (TSA)^[148] was then used to evaluate the thermo stability of the protein. In this assay SYPRO Orange dye binds nonspecifically to hydrophobic surfaces of the protein, and water strongly quenches its fluorescence (Figure 3.24). When temperature increases, the protein unfolds, the exposed hydrophobic surfaces bind the dye, resulting in an increase in fluorescence by excluding water. The highest unfolding rate corresponds to the melting temperature (T_m). This effect can be used to study the unfolding of soluble proteins by measuring the increase in fluorescence with increasing temperature. To some extent, higher T_m means higher stability of the protein.

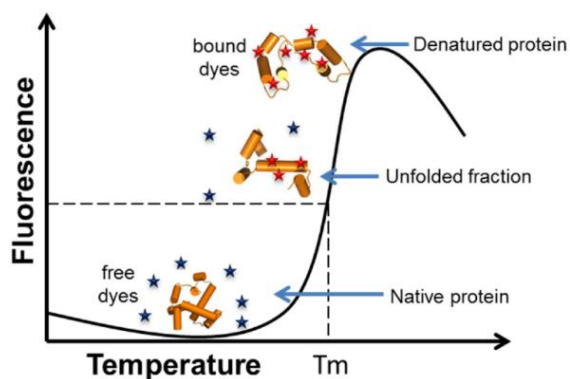
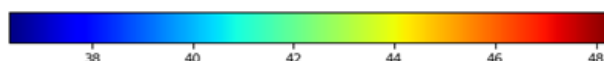
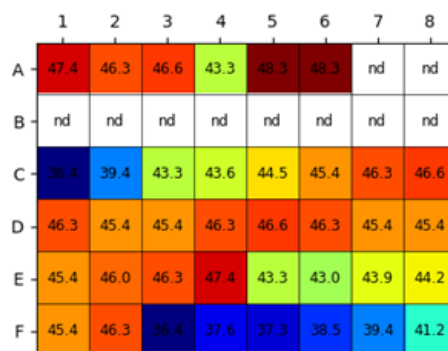


Figure 3.24. Relationship between fluorescence and protein melting temperature

In this assay (see details in section 6.3.3), 21 μL of filtered buffers with different additives were added into the 48 well microplate placed on ice. In each well 2 μL of SYPRO Orange and then 2 μL of the protein (final concentration of 1.6 – 8.0 μM) were dispensed. Then the plate was covered with adhesive film and spun down at 5000 rpm for 5 min at 4 $^{\circ}\text{C}$ to remove any air bubble in the samples. Finally the plate was put inside the RT-PCR for the thermal denaturation analysis at 569 nm by using a temperature gradient of 1 $^{\circ}\text{C}/\text{min}$ from 5 $^{\circ}\text{C}$ to 95 $^{\circ}\text{C}$.

Two screens were performed to optimize the storage buffer of the protein. Firstly, we screened different pH, different concentration of glycerol, addition of EDTA and addition of imidazole. To observe the influence of pH to the T_m under a certain condition, CHC buffer (Citric acid: HEPES: CHES = 2: 3: 4) was used. This buffer permits a wide range of pH value (pH 4-10). According to the test, the pH arrange between 7 ~ 8 (Figure 3.22 D1-E4) has the highest average T_m . The calculated isoelectric point of SQTKS DH-KR is pH 4 ~ 5 and protein precipitates within such pH value (Figure 3.22 A7-C2) where no signal or significantly low signal has been detected (nd). At the same pH, increase of the concentration of NaCl (from 0 mM, 83 mM, 208 mM, 417 mM, 833 mM to 1670 mM) generally improves the stability of the protein. While as a dilemma, too much salt will cause salting out and make the protein inactive. In addition, we found 20 % glycerol ($T_m = 47.4$ $^{\circ}\text{C}$) is slightly more stabilizing than 10 % glycerol (46.3 $^{\circ}\text{C}$) and 5 % glycerol (46.6 $^{\circ}\text{C}$) (Figure 3.25 A1, A2, A3).

The addition of 1 mM EDTA obviously increases the T_m (48.3 °C) (Figure 3.25 A5) while imidazole significantly decreases the T_m (43.3 °C) (Figure 3.25 A4). Therefore, imidazole should be removed as soon as possible after Ni-NTA chromatography.

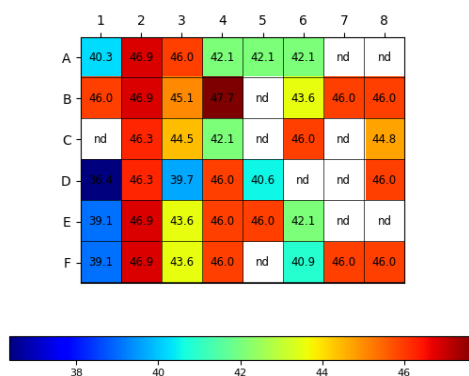


	1	2	3	4	5	6	7	8
A	Tris, pH 8 150 mM NaCl 20 % Glycerol	Tris, pH 8 150 mM NaCl 10 % Glycerol	Tris, pH 8 150 mM NaCl 5 % Glycerol	Tris, pH 8 150 mM NaCl 20 % Glycerol 500 mM Imidazole	Tris, pH 8 150 mM NaCl 20 % Glycerol 1 mM EDTA	Tris, pH 8 150 mM KCl 20 % Glycerol 1 mM EDTA	CHC, pH 4	CHC, pH 4 83 mM NaCl
B	CHC, pH 4 208 mM NaCl	CHC, pH 4 417 mM NaCl	CHC, pH 4 833 mM NaCl	CHC, pH 4 1670 mM NaCl	CHC, pH 5	CHC, pH 5 83 mM NaCl	CHC, pH 5 208 mM NaCl	CHC, pH 5 417 mM NaCl
C	CHC, pH 5 833 mM NaCl	CHC, pH 5 1670 mM NaCl	CHC, pH 6	CHC, pH 6 83 mM NaCl	CHC, pH 6 208 mM NaCl	CHC, pH 6 417 mM NaCl	CHC, pH 6 833 mM NaCl	CHC, pH 6 1670 mM NaCl
D	CHC, pH 7	CHC, pH 7 83 mM NaCl	CHC, pH 7 208 mM NaCl	CHC, pH 7 417 mM NaCl	CHC, pH 7 833 mM NaCl	CHC, pH 7 1670 mM NaCl	CHC, pH 8	CHC, pH 8 83 mM NaCl
E	CHC, pH 8 208 mM NaCl	CHC, pH 8 417 mM NaCl	CHC, pH 8 833 mM NaCl	CHC, pH 8 1670 mM NaCl	CHC, pH 9	CHC, pH 9 83 mM NaCl	CHC, pH 9 208 mM NaCl	CHC, pH 9 417 mM NaCl
F	CHC, pH 9 833 mM NaCl	CHC, pH 9 1670 mM NaCl	CHC, pH 10	CHC, pH 10 83 mM NaCl	CHC, pH 10 208 mM NaCl	CHC, pH 10 417 mM NaCl	CHC, pH 10 833 mM NaCl	CHC, pH 10 1670 mM NaCl

Figure 3.25. Thermal shift assay of SQTs DH-KR (pH 4.0 ~ 10.0): **A1-A6**, influence of glycerol, imidazole EDTA; **A7-B8**, protein precipitates; **A7-F8**, different pH in CHC buffer (Citric acid: HEPES: CHES = 2: 3: 4) with different concentration of NaCl.

Secondly, different buffering systems also influence the T_m . In this test, there are several unexplainable results such as the same buffer A3 and B4 having different T_m , which may be

caused by the quality of self-made buffers and the remaining tiny bubbles inside the buffer. However, we can still find some general conclusions. By comparing the average T_m in Tris buffer (Figure 3.26 B), phosphate buffer (C), HEPES (D), TES (E) and MOPS (F) at pH 8, Tris buffer has been proved as the buffer with the highest average buffer T_m . The addition of TCEP can decrease the T_m , such as A3 and A4. 20 % glycerol is generally better than 10 % glycerol, such as D3 and D4. In addition, we also found addition of 1 mM of EDTA is enough for keeping good T_m while more EDTA does not bring better T_m , such as B4 and B6.



	1	2	3	4	5	6	7	8
A	Tris, 20% Glycerol, 1mM EDTA	Tris, 20% Glycerol, 1mM EDTA, 50mM NaCl	Tris, 20% Glycerol, 1mM EDTA, 150mM NaCl	Tris, 20% Glycerol, 1mM EDTA, 150mM NaCl, 10mM TCEP	NaH ₂ PO ₄ , 20% Glycerol, 1mM EDTA, 150mM NaCl, 10mM TCEP	HEPES, 20% Glycerol, 1mM EDTA, 150mM NaCl, 10mM TCEP	TES, 20% Glycerol, 1mM EDTA, 150mM NaCl, 10mM TCEP	MOPS, 20% Glycerol, 1mM EDTA, 150mM NaCl, 10mM TCEP
B	Tris, 10% Glycerol, 150mM NaCl	Tris, 20% Glycerol, 150mM NaCl	Tris, 10% Glycerol, 150mM NaCl, 1mM EDTA	Tris, 20% Glycerol, 150mM NaCl, 1mM EDTA	Tris, 10% Glycerol, 150mM NaCl, 5mM EDTA	Tris, 20% Glycerol, 150mM NaCl	Tris, 10% Glycerol, 150mM NaCl, 25mM EDTA	Tris, 20% Glycerol, 150mM NaCl, 25mM EDTA
C	NaH ₂ PO ₄ , 10% Glycerol, 150mM NaCl	NaH ₂ PO ₄ , 20% Glycerol, 150mM NaCl	NaH ₂ PO ₄ , 10% Glycerol, 150mM NaCl, 1mM EDTA	NaH ₂ PO ₄ , 20% Glycerol, 150mM NaCl, 1mM EDTA	NaH ₂ PO ₄ , 10% Glycerol, 150mM NaCl, 5mM EDTA	NaH ₂ PO ₄ , 20% Glycerol, 150mM NaCl, 5mM EDTA	NaH ₂ PO ₄ , 10% Glycerol, 150mM NaCl, 25mM EDTA	NaH ₂ PO ₄ , 20% Glycerol, 150mM NaCl, 25mM EDTA
D	HEPES, 10% Glycerol, 150mM NaCl	HEPES, 20% Glycerol, 150mM NaCl	HEPES, 10% Glycerol, 150mM NaCl, 1mM EDTA	HEPES, 20% Glycerol, 150mM NaCl, 1mM EDTA	HEPES, 10% Glycerol, 150mM NaCl, 5mM EDTA	HEPES, 20% Glycerol, 150mM NaCl, 5mM EDTA	HEPES, 10% Glycerol, 150mM NaCl, 25mM EDTA	HEPES, 20% Glycerol, 150mM NaCl, 25mM EDTA
E	TES, 10% Glycerol, 150mM NaCl	TES, 20% Glycerol, 150mM NaCl	TES, 10% Glycerol, 150mM NaCl, 1mM EDTA	TES, 20% Glycerol, 150mM NaCl, 1mM EDTA	TES, 10% Glycerol, 150mM NaCl, 5mM EDTA	TES, 20% Glycerol, 150mM NaCl, 5mM EDTA	TES, 10% Glycerol, 150mM NaCl, 25mM EDTA	TES, 20% Glycerol, 150mM NaCl, 25mM EDTA
F	MOPS, 10% Glycerol, 150mM NaCl	MOPS, 20% Glycerol, 150mM NaCl	MOPS, 10% Glycerol, 150mM NaCl, 1mM EDTA	MOPS, 20% Glycerol, 150mM NaCl, 1mM EDTA	MOPS, 10% Glycerol, 150mM NaCl, 5mM EDTA	MOPS, 20% Glycerol, 150mM NaCl, 5mM EDTA	MOPS, 10% Glycerol, 150mM NaCl, 25mM EDTA	MOPS, 20% Glycerol, 150mM NaCl, 25mM EDTA

Figure 3.26. Thermal shift assay of SQTGS DH-KR tetradomain in more different conditions, such as different buffers, concentration of EDTA and glycerol, addition of TCEP. The pH is 8.0.

Overall, we found the buffer (50 mM Tris, pH 8, 150 mM NaCl, 20% Glycerol, 1mM EDTA) with highest T_m (48 °C). The optimized buffer significantly improves the stability of the SQTKS DH-KR tetradomain. By storing at 4 °C, the protein purified by Ni-NTA chromatography remains stable for at least one month with no reactivity change observed in UV assay (enoyl reduction activity in DH-KR) and with no constitution change observed in SDS-PAGE (Figure 3.27). Therefore, the thermal shift assay was successful and its certified buffer has been used as the standard storage buffer during all processes of protein purification.

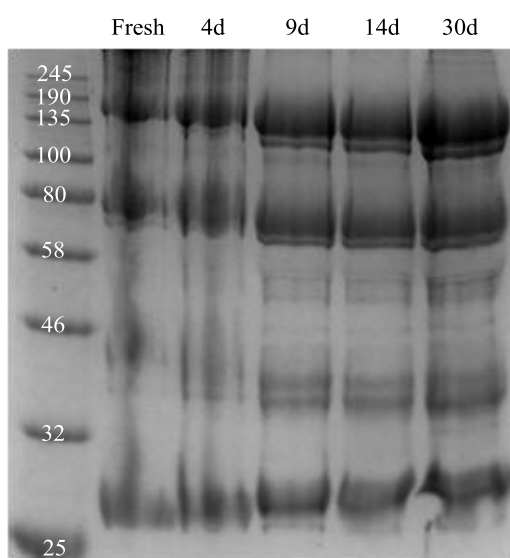


Figure 3.27. Long-term stability of SQTKS DH-KR tetradomain (166 kDa) at 4 °C (purified by Ni-NTA) in optimized buffer (50 mM Tris, pH 8.0, 150 mM NaCl, 20 % glycerol, 1 mM EDTA).

3.3.1.2 SQTKS CMeT

Due to the large exposed hydrophobic area of the isolated CMeT domain according to the modeling structure of SQTKS (chapter 2), the stability of SQTKS CMeT is much less than that of the SQTKS DH-KR tetradomain. In order to improve its stability, a great number of thermal shift assays have been performed to find a suitable storage buffer. However, this time we found the relationship between the T_m and protein stability is not always related.

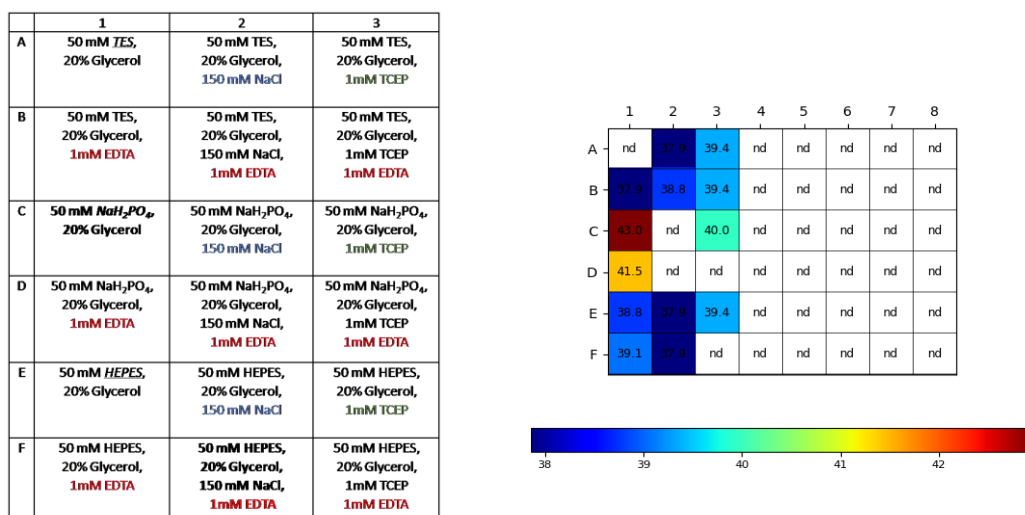


Figure 3.28. Thermal shift assay of SQTCS CMeT, the pH is 8.0.

The average value of T_m of SQTCS CMeT in different conditions is generally lower than that of SQTCS DH-KR (Figure 3.28). In consideration of different buffer with different amount of NaCl, TCEP and EDTA, the buffer (50 mM phosphate buffer, pH 8, 20 % glycerol) with highest T_m (43 °C) has been selected as the storage buffer of CMeT. However, this buffer makes CMeT unstable and it precipitates over several hours at 4 °C, while buffer (F2) with lowest T_m or the standard DH-KR storage buffer can stabilize the CMeT for much longer (4 °C for one week). Therefore, high T_m and high stability of the protein are not related for the isolated CMeT. The T_m reflects the binding between the dye and hydrophobic area of unfolding protein. If the CMeT probably does not unfold and it tends to form aggregated multimers due to the association among protein particles when temperature increases, then the thermal shift assay cannot be used effectively in this situation.

3.3.2 On-column SQTCS DH-KR Tetradomain Denaturation

With a good buffer in hand, we continued to try to improve the purity of the SQTCS DH-KR. According to the peptides identified by ESI-QTOF-MS, one part of the protein sized 58~80 kDa was certified as fragments of DH-KR (Figure 3.29). Domain boundaries of DH, CMeT, ER and KR have been noted as four long bars with different colour (Figure 3.29). Below these

long bars, the red small bars represent the identified peptides of DH-KR and the purple small bars represent the identified peptides of the protein sized ~80 kDa. The interesting phenomenon is, the identified peptides of the protein sized ~80 kDa, nearly half smaller than SQTCS DH-KR, covers areas of every single domain. One possible explanation is that some of DH-KR breaks in the middle and separates into two nearly equal sized fragments which can be both observed by ESI-QTOF-MS. These digested “halves” of the DH-KR tetradomain may retain significant structure and bind together so that the digested protein still co-elutes with undigested protein. Because the single band elutes at ~80 kDa, approximately half of the 166 kDa full size, it seems that the digestion site is in the center of the construct, corresponding to the ψ KR region.

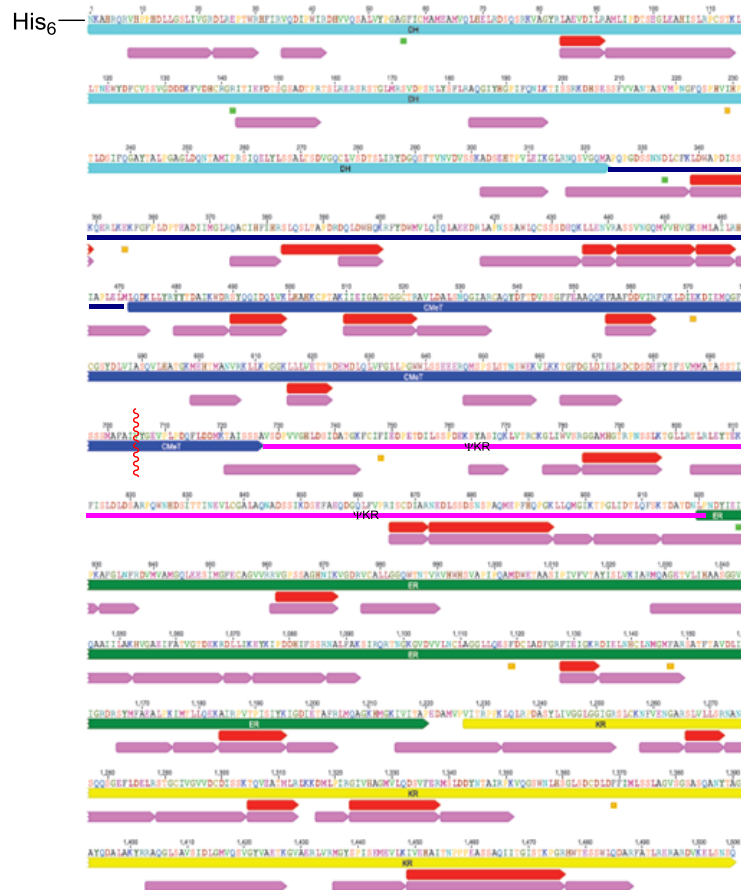


Figure 3.29. SQTCS DH-KR peptides identified by ESI-QTOF-MS. Identified peptides of the protein sized ~80 kDa are shown as **small red bars**. Identified peptides of the DH-KR tetradomain are shown as **small purple bars**. The area of each domain is shown as long bars in colour, in the order of DH (sky blue), CMET (deep blue), ψ KR (pink), ER (dark green), KR (yellow).

To investigate whether the explanation is correct, on-column denaturation was tried. The basic strategy is (Figure 3.30 A), with the N-His₆-DH-KR tetradomin attached to the Ni-NTA column, addition of denaturation agent guanidine hydrochloride (GuHCl) into the wash buffer will denature the protein on column. The protein was predicted to be fragile in the middle area where CMeT domain is located. Therefore, denaturation by GuHCl, the protein may separate into two similar-sized fragments containing one piece of CMeT (MTa or MTb) respectively. As a result, the C-terminal fragment without histidine tags are washed out and the N-terminal fragment with histidine tags will stay on the column. Afterwards, the fragment can be eluted by high concentration of imidazole. Finally, the eluted protein was quickly changed into the standard storage buffer by desalting column to remove GuHCl.

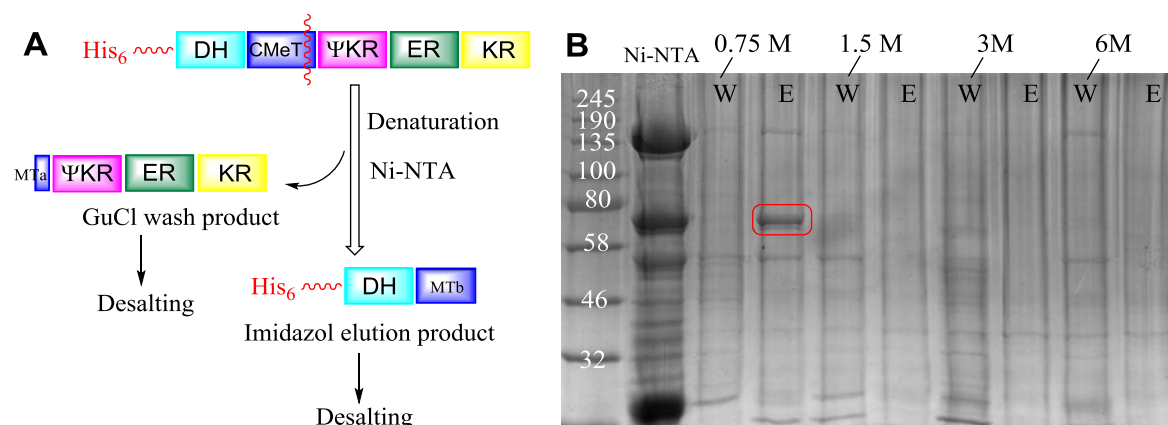


Figure 3.30. On column denaturation of SQTGS DH-KR: **A**, strategy; **B**, the analysis of the product in the wash and elution. Different concentrations of GuHCl (0.75 M, 1.5 M, 3 M, 6 M) in wash buffer (W) were used for denaturation. The elution buffer (E) contains 500 mM of Imidazole.

To remove most host cell proteins, DH-KR was purified by Ni-NTA first. After removing the imidazole, the purified DH-KR was again sent through the Ni-NTA for the on-column denaturation. Different concentrations of guanidine hydrochloride (GuHCl) were used for on-column denaturation (Figure 3.30 B). However, almost no target protein has been found in the wash and elution. As an exception, a protein sized ~80 remained on the Ni-NTA

column after being washed by 0.75 M of GuHCl. However, the protein is identified as ArnA, a host cell protein also observed in the purification of the SQTKS DH-ER tridomain. Further decrease in the concentration of GuHCl cannot denature the protein effectively.

The on-column denaturation of SQTKS DH-KR failed. The repeatability of this experiment was unsatisfactory. A possible explanation is, before detecting the existence of target protein by SDS-PAGE, the remaining hard-to-remove GuHCl can still denature the target protein and the unstable unfolded protein may be easily degraded into small pieces.

3.3.3 Further Purification

3.3.3.1 Impurities Analysis

In order to further purify the SQTKS DH-KR tetradomain and the CMeT monodomain, ESI-QTOF-MS amino acid analysis of the main impurities is necessary. In all cases the contaminating proteins were cut from the gel, digested and submitted for ESI-QTOF-MS analysis. In the case of CMeT (Figure 3.31 A), many fragments of *E. coli* proteins were observed in the digested protein sample sized ~58 kDa. Among them, two *E. coli* proteins show high relative reliabilities (Figure 3.31 Score): the bifunctional polymyxin resistance protein ArnA (section 3.3.3.2, fused UDP-L-Ara4N formyltransferase and UDP-GlcA C-4'-decarboxylase) which is involved in lipopolysaccharide biosynthesis; and the ATP-dependent chaperone Hsp70 (DnaK) which interacts with chaperone DnaJ to prevent the aggregation of stress-denatured proteins. In the case of DH-ER (Figure 3.31 B), the main contaminating protein ArnA was observed in the digested protein sample sized 58~80 kDa. In the case of the DH-KR tetradomain (Figure 3.31 C), both ArnA and Hsp70 were observed as the main contaminating proteins in the digested protein sample sized 58~80 kDa. These host cell proteins have been reported as having strong protein-protein interactions with some target proteins during ion exchange chromatography and size exclusion chromatography.^[149-150] Fortunately, these host cell proteins are removable.

A. CMeT contaminant around ~58 kDa

Description	Score	Coverage	MW [kDa]	calc. pI
fused UDP-L-Ara4N formyltransferase and UDP-GlcA C-4'-decarboxylase [Escherichia coli str. K12 substr. W3110]	2968.46	44.24	74.2	6.87
chaperone Hsp70, co-chaperone with DnaJ [Escherichia coli str. K12 substr. W3110]	2817.17	42.63	69.1	4.97
succinyl-CoA synthetase, beta subunit [Escherichia coli str. K12 substr. W3110]	83.98	2.84	41.4	5.52
F1 sector of membrane-bound ATP synthase, beta subunit [Escherichia coli str. K12 substr. W3110]	60.72	5.43	50.3	5.01
Cpn60 chaperonin GroEL, large subunit of GroESL [Escherichia coli str. K12 substr. W3110]	56.71	2.01	57.3	4.94
protein chain elongation factor EF-Tu [Escherichia coli str. K12 substr. W3110]	45.17	3.81	43.3	5.45
L-glutamine:D-fructose-6-phosphate aminotransferase [Escherichia coli str. K12 substr. W3110]	30.93	1.64	66.9	5.87
isocitrate dehydrogenase, specific for NADP+ [Escherichia coli str. K12 substr. W3110]	28.18	2.88	45.7	5.26
30S ribosomal subunit protein S1 [Escherichia coli str. K12 substr. W3110]	26.04	2.51	61.1	4.98
succinyl-CoA synthetase, NAD(P)-binding, alpha subunit [Escherichia coli str. K12 substr. W3110]	24.83	3.11	29.8	6.79
polynucleotide phosphorylase/polyadenylase [Escherichia coli str. K12 substr. W3110]	23.02	1.50	79.8	5.53
conserved hypothetical protein [Escherichia coli str. K12 substr. W3110]	20.79	0.91	94.9	6.09

B. DH-ER contaminant around 58-80 kDa

Description	Score	Coverage	MW [kDa]	calc. pI
fused UDP-L-Ara4N formyltransferase and UDP-GlcA C-4'-decarboxylase [Escherichia coli str. K12 substr. W3110]	4336.00	47.42	74.2	6.87
protein chain elongation factor EF-Tu [Escherichia coli str. K12 substr. W3110]	73.20	6.09	43.3	5.45
Cpn60 chaperonin GroEL, large subunit of GroESL [Escherichia coli str. K12 substr. W3110]	36.30	5.11	57.3	4.94

C. DH-KR contaminant around 58-80 kDa

Description	Score	Coverage	MW [kDa]	calc. pI
chaperone Hsp70, co-chaperone with DnaJ [Escherichia coli str. K12 substr. W3110]	1221.46	30.72	69.1	4.97
fused UDP-L-Ara4N formyltransferase and UDP-GlcA C-4'-decarboxylase [Escherichia coli str. K12 substr. W3110]	902.48	23.79	74.2	6.87
molecular chaperone HSP90 family [Escherichia coli str. K12 substr. W3110]	234.56	16.35	71.4	5.21
L-glutamine:D-fructose-6-phosphate aminotransferase [Escherichia coli str. K12 substr. W3110]	118.92	7.22	66.9	5.87
protein chain elongation factor EF-Tu [Escherichia coli str. K12 substr. W3110]	50.89	6.09	43.3	5.45
30S ribosomal subunit protein S1 [Escherichia coli str. K12 substr. W3110]	50.47	4.13	61.1	4.98
thioredoxin 1 [Escherichia coli str. K12 substr. W3110]	40.45	11.01	11.8	4.88
threonyl-tRNA synthetase [Escherichia coli str. K12 substr. W3110]	29.43	1.40	74.0	6.19
succinyl-CoA synthetase, beta subunit [Escherichia coli str. K12 substr. W3110]	29.31	2.84	41.4	5.52
isocitrate dehydrogenase, specific for NADP+ [Escherichia coli str. K12 substr. W3110]	28.96	2.88	45.7	5.26
transketolase 1, thiamin-binding [Escherichia coli str. K12 substr. W3110]	24.68	1.21	72.2	5.67
3-oxoacyl-[acyl-carrier-protein] synthase II [Escherichia coli str. K12 substr. W3110]	21.43	5.57	43.0	6.09

Figure 3.31. The ESI-QTOF-MS amino acid analysis of the main “contaminating” proteins sized 58~80 kDa of DH-KR CMeT monodomain, DH-ER tetradomain and DH-KR tetradomain. The score indicates the relative reliability of the data. The coverage indicates the percentage content of amino acids that have been identified in the recognized protein from the database.

3.3.3.2 Methods to Remove Impurities

From the analysis of the component of impurities, we may be able to explain the reason why the pure SQTGS DH-KR tetradomain and CMeT monodomain are difficult to be isolated: the impurities, including ArnA, molecular chaperones and other host cell proteins which are

difficult to be removed by chromatography, may attract target proteins together and form large-sized aggregates by strong protein-protein interaction (Figure 3.32).

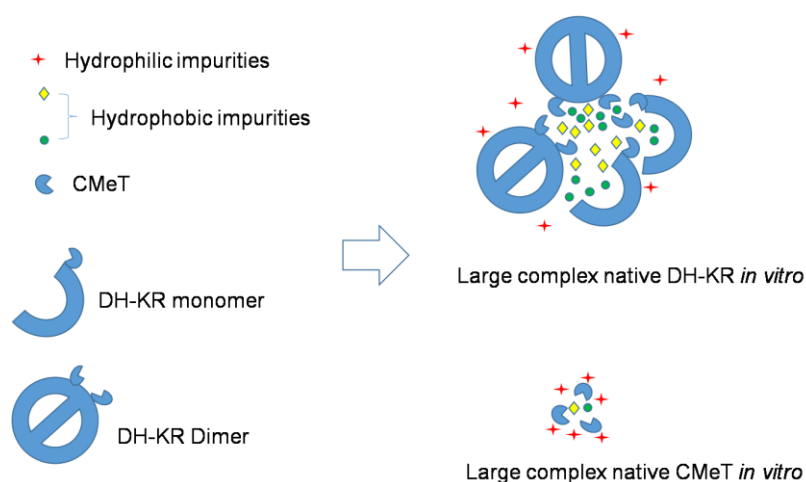


Figure 3.32. Proposed composition of isolated SQTCS DH-KR and CMeT in solution. Hydrophilic impurities could be DNA and DNA binding proteins which are difficult to be removed by Ni-NTA; hydrophobic impurities could be host cell proteins such as ArnA, molecular chaperones, etc. These impurities attract monomer form or dimer form of target proteins together to form a complex, which makes the purification difficult.

To remove these host cell impurities, an optimized purification plan was devised (Figure 3.33). The expression of ArnA in *E. coli* is known to be regulated by the PhoP-PhoQ system which is activated by low concentrations of Mg^{2+} and repressed by high concentrations of Mg^{2+} .^[151] Therefore, to remove the main contaminant ArnA, addition of Mg^{2+} into the culture medium may work. During cell lysis, a large amount of DNA and RNA will be released and make the lysis buffer sticky, which brings the filtration problem before any chromatography. DNA and RNA may also bind our target proteins and attract other DNA and RNA binding host cell proteins. To remove DNA and RNA as much as possible, we incubated the cell lysate with excess DNase and RNase, which are Mg^{2+}/Ca^{2+} dependent. As RNA degrades much easier than DNA and addition of extra RNase brought no considerable improvement on the purity and activity of our target proteins, only DNase was determined as a standard

reagent in the optimized purification method. To help release molecular chaperones, ATP was used.^[149] Binding of ATP by Hsp70 triggers a conformational shift at the N-terminal of the protein from an “open” form to a “close” form. Therefore, chaperons can be released from our target proteins by several ATP wash steps in the Ni-NTA column.

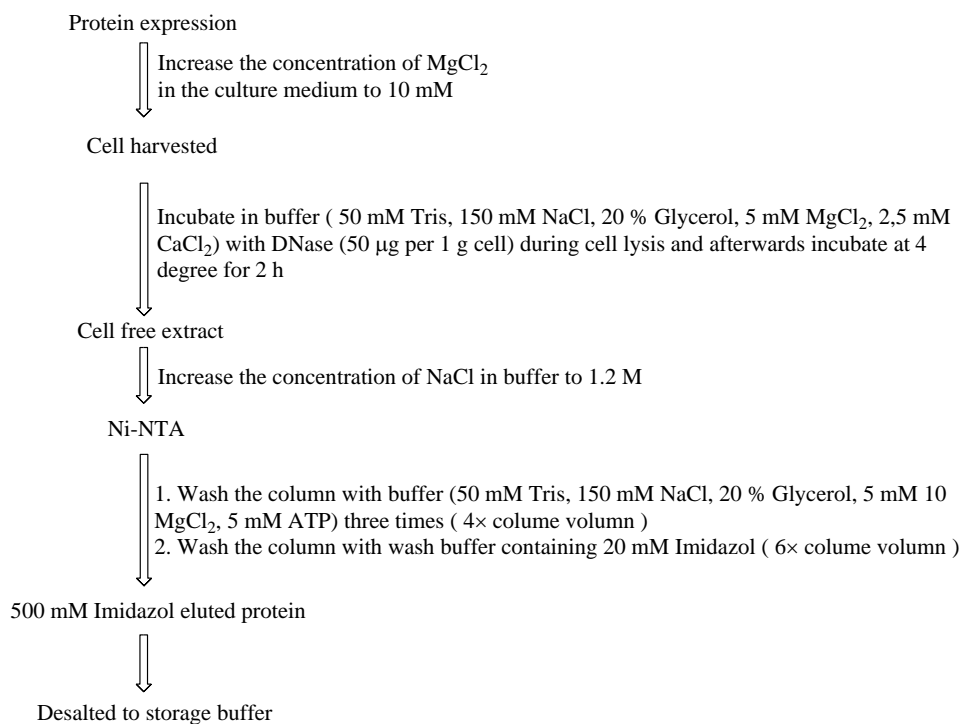


Figure 3.33. The optimized plan for purification of the SQTCS DH-KR tetradomain and CMeT monodomain.

As a result of using the new methods, compared to the normal purification without optimization (control in Figure 3.34), the purity of SQTCS DH-KR (Figure 3.34 A) and CMeT (Figure 3.34 B) after Ni-NTA were clearly improved, especially between 46 kDa ~ 80 kDa. This purification effect is similar to that of the genetically tailored expression host NiCo21(DE3) recently developed for better removal of host cell proteins.^[152] In comparison, our method is more specific and simpler.

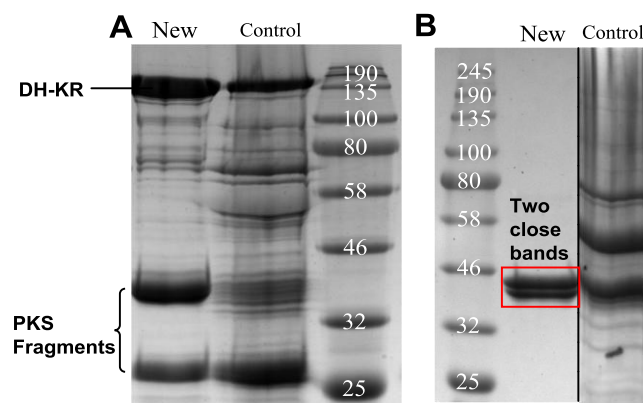


Figure 3.34. Ni-NTA purity of SQTCS DH-KR (166 kDa, **A**) and CMeT (48 kDa, **B**) obtained after optimized purification. Controls are the related protein without optimized purification.

For the SQTCS CMeT, two proteins of similar size were obtained after the optimized purification (Figure 3.34 B). ESI-QTOF-MS identified these two proteins as both fragments of CMeT. A possible explanation is the CMeT partially digested at either its N or C terminus to a smaller protein.

For the SQTCS DH-KR tetradomain (Figure 3.34 A), an interesting phenomenon is, the amount of protein sized 58~80 kDa decreased considerably while a protein sized ~40 kDa increased equally. Identified by ESI-QTOF-MS, the main “contaminant” proteins sized ~40 kDa and ~25 kDa both contain fragments of DH-KR (section 6.5.1). Therefore, a possible explanation is that, the previous ~80 kDa fragments of DH-KR may further degrade into smaller pieces during the optimized purification process. Fortunately, these small “contaminant” proteins can be finally removed by size exclusion chromatography (Figure 3.35 A GF3).

3.3.3.3 The Structural Form of SQTCS DH-KR in Solution

During the purification of the SQTCS DH-KR tetradomain by size exclusion chromatography, three main protein complexes eluted from the column (Figure 3.35A). The first (GF1) had an estimated Mw of 340 kDa corresponding to dimeric DH-KR tetradomain. The second (GF2) had an estimated Mw of 170 kDa corresponding to monomeric DH-KR tetradomain. The

third (GF3) comprised the 40 kDa fragment and eluted with an apparent Mw of 25 kDa, corresponding to a complex of 42 kDa in solution. To certify the structural form of SQT KS DH-KR in buffer, the GF1 and GF2 fractions were further investigated by Multi-Angle Light Scattering (MALS).^[156]

Similar to SAXS, a well collimated, single frequency, polarized laser beam is used in MALS to illuminate a solution containing a suspension of the macromolecules of interest. According to the Zimm Equation,^[153] the amount of light scattered is directly proportional to the product of the molar mass and the molecular concentration, and the variation of scattered light with scattering angle is proportional to the average size of the scattering molecules.

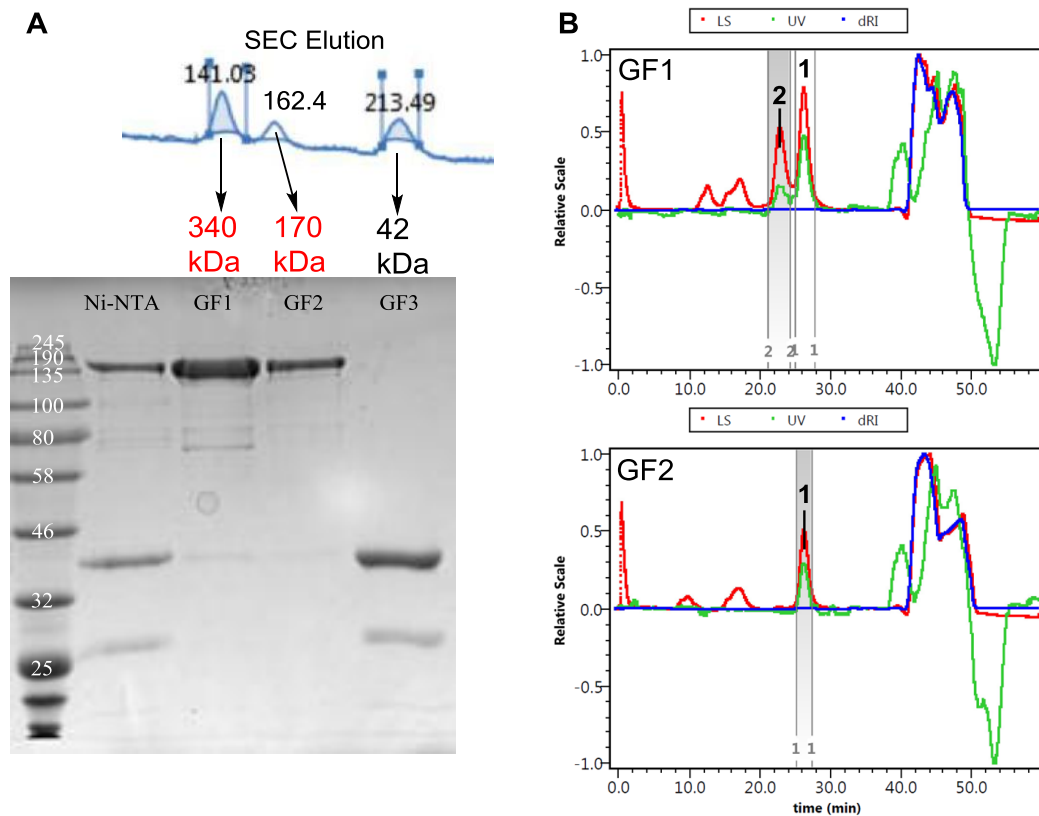


Figure 3.35. Relationship between GF1 and GF2 of SQT KS DH-KR revealed by size exclusion chromatography and MALS: **A**, the SDS-PAGE shows the purified proteins by size exclusion chromatography; **B**, the observation of GF1 and GF2 in SEC-MALS. The green line (UV) refers to the signal of protein. The red line (Light Scattering) is related to the determination of protein molar mass. The blue line (dRI) refers to differential refraction index. The running buffer in MALS: 50 mM Tris, pH 8.0, 150 mM NaCl.

In this experiment, a small amount of sample was sent through a high resolution size exclusion chromatography and the eluted fractions were then detected by MALS one by one. In the end of MALS spectrum (Figure 3.35 B), the peak with high UV and sudden increase of differential refraction index come from the effect of some EDTA and glycerol (storage buffer) in the injecting protein sample. Interestingly, the previous GF1 separated into two close peaks 1 and 2 (Figure 3.35 B), while peak 1 was only found in GF2. MALS certifies the molar mass of the peak 2 is 340 kDa and peak 1 is 170 kDa. Therefore, the peak 2 refers to dimer DH-KR and the peak 1 refers to the monomer form of DH-KR, which is similar to the calibration results by size exclusion chromatography. An explanation for what happens to GF2 is: an equilibrium exists between the dimer form and monomer form of DH-KR; an instability factor still exists, which can break the equilibrium to form more monomer DH-KR in solution. Therefore, the protein solution is polydispersed and the previous SAXS analysis is not able to deliver a correct modeling structure of SQTKS DH-KR. This may affect the activity of SQTKS DH-KR. In chapter 4, detailed *in vitro* assays of the protein will further illustrate this problem.

3.4 Summary and Future Work

As a summary, the SQTKS DH-KR tetradomain (166 kDa) and the CMeT monodoamin (48 kDa) have been finally isolated at a purity of 90 % - 95 %. The DH-KR was inserted into the pET28a and expressed in *E. coli* BL21 (DE3) under a condition of 0.1 mM IPTG, 16 °C and 2 TY medium for 24 h. The CMeT was inserted into the pET28a and expressed in *E. coli* BL21 (DE3) under a condition of auto induction, 16 °C and LBE medium for 24 h. The isolation of the SQTKS DH-ER tridomain and DH-ACP pentadomain failed. The isolated DH-ER was little and extremely unstable while no soluble DH-ACP has been observed to date.

As a special way to promote the protein folding, the isolated SQTKS DH-KR coded by sub-codon-optimized gene showed no significant difference versus the protein coded by fully codon-optimized gene. Besides of low-temperature expression (to 10 °C) by *E. coli*

ArcticExpress (DE3), plasmids with different fusion tags such as Thioredoxin and SUMO cannot bring any significant improvement to the protein folding. These studies reveal that the protein folding is possibly not the reason which brings instability issue to our proteins.

To keep the proteins stable in solution, the best storage buffer (50 mM Tris, pH 8.0, 150 mM NaCl, 20 % Glycerol, 1mM EDTA) has been selected by UV assay and thermo shift assay (section 3.3.1). By using the purification optimization method (Figure 3.30) and size exclusion chromatography, the purity of DH-KR increased from 40 % (without purification optimization) to 90 %. As the more hydrophobic area exposed, the isolated SQTCS CMeT is less stable than DH-KR. Therefore, less chromatographic purifications have been done on CMeT. Even though, by using the purification optimization method (Figure 3.30), the purity of CMeT also increased from 40 % (without purification optimization) to 90 %.

The purified DH-KR tetradomain has been analyzed by MALS (Figure 3.32) which indicates the protein forms a balance between dimer and monomer and an unknown instability factor can make the dimer form of SQTCS DH-KR separate into monomer form. For the CMeT, the abnormal thermal shift and size calibration by size exclusion chromatography indicated that the protein is less stable than DH-KR and may form multimers in buffer. The instability issue of SQTCS DH-KR and CMeT may derive from the limitation of *E. coli* which lacks the function of post-translational modification. As SQTCS is originally produced by fungi, yeast expression or insect expression may be a better choice to produce stable and fully active SQTCS domains. In addition to change the expression system, some other aspects should be considered in the future to optimize the expression and purification of these proteins.

For the cell culture, rich medium was considered as a necessity to produce more soluble over-expressed target proteins. There are high concentrations of phosphate combined with amino acids and other nutrients in rich media. An important issue is, systematic tests revealed that the higher the concentration of phosphate, the higher the concentration of antibiotics needed to prevent growth of cells that have lost the plasmid.^[136] That means, for this project,

a normal kanamycin concentration of 50 µg/ml might be not enough in rich medium culture. In the future, 100 ~ 400 µg/ml of kanamycin could be tested to reduce the background of host cell proteins.^[136]

For the buffer optimization and protein purification by chromatography, determination of a suitable condition always takes heavy lab work. To find the best condition which generates the target protein of the highest purity in a reproducible way, there are many parameters (*e.g.*, concentration of salt and glycerol, pH, loading amount and velocity of chromatography, *etc.*) needed to be considered. If we check the influence of each parameter individually with specific controls, it will take huge effort and cost. Besides from many cryptic conclusions, the great risk of such one-factor-at-a-time experimental setup is that the true best condition cannot be identified, especially when some parameters are related to each other. For example, the concentration of salt in buffer will change when pH is adjusted by adding acid or base, the incubation time should be adjusted when temperature or rotate speed is changed in cell culture, ligation time and temperature should be adjusted when size of the insert is changed, *etc.* In fact, design a feasible experiment in the lab sometimes relies on experiences.

Design of Experiments (DoE) offers a systematic variance analysis to guide the experiment in this case, which has been widely used in industry. DoE allows us to use a minimum number of experiments, in which we systematically vary several experimental parameters simultaneously to obtain sufficient information. Based on the obtained data, a mathematical model of the studied process is created by software such as Minitab. In this way we can evaluate the influence of parameters from the outcome and predict the best condition efficiently. Therefore, it is highly recommended that DoE should be used in the daily lab work. For example, Unicorn software with ÄKTA FPLC offers function of DoE to optimize the protein purification.^[154]

4 *In Vitro* Assay of SQTCS Multi-domain Fragments

4.1 Introduction

For the *in vitro* assays, the choice of substrate is as important as the isolation of active enzyme. In the β -modification stage of the biosynthesis of squalestatin tetraketide, substrates are linked to ACP. ACP carries the substrate to meet the active pockets of each modification domain after each extension. It is suggested ACP can move among domains, and the connection between ACP and each domain is important for the movement and transformation of substrate from pocket to pocket. However, SQTCS ACP has never been isolated as an active soluble protein despite many efforts in the Cox group. Even if the isolation had succeeded, methods of substrate-linkage, detection and purification of the protein-linked products from *in vitro* assay would be a major challenge. To make it simple, several small-molecule analog structures (Figure 4.1) have been designed based on the linkage between ACP and substrates. Acyl CoA **71** should be considered first, as CoA is a natural product and acyl CoA is widely used as the substrate donor during the biosynthesis of polyketides. The challenge is, due to the strong polarity of CoA, it will also bring purification problems. Among the analogous structures, the simplest *N*-acetylcysteamine substrate **69** (NAC) has been widely and successfully used both in *in vivo* and *in vitro* enzyme assays. However, the NAC structure, compared to the original ACP linkage **72**, is too simple and is not well recognized by some enzymes, which may result in a low reaction rate and incorrect performance. By using acyl pantetheine **70** as the substrate for the ER domain, the rate of reduction was observed to be at least 12 times higher than the corresponding acyl-NAC, which makes the determination of k_{cat}/K_M more feasible and accurate. This has previously been shown to be the case for the SQTCS ER^[75] and DH^[57] domains.

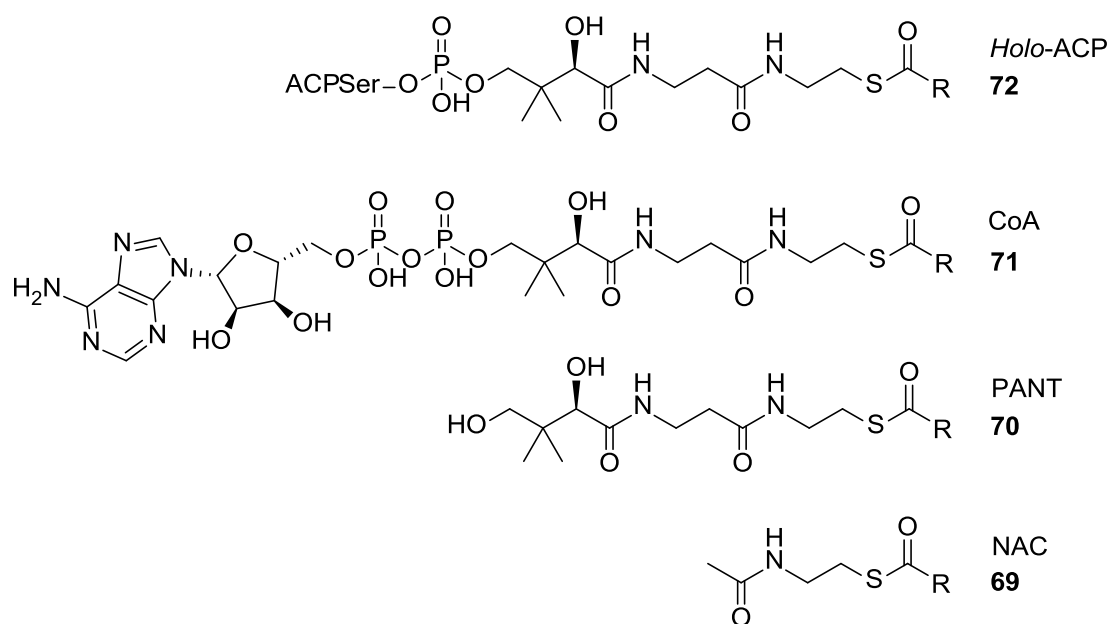


Figure 4.1. The 4'-phosphopantetheine ACP prosthetic group (top) and typical small-molecule substrate analogs. R refers to variable part of different polyketide substrates.

With the *in vitro* assay, we can study the function of individual domains with specific substrate analogs. Determination of the stereochemistry involved in the enzyme catalyzed reaction can help us understand the detailed relationship between protein structure and its functionality. In the *in vitro* assay, we can also monitor the reaction by observing the formation of product or consumption of cofactor involved in the reaction. Several initial rates of the reaction with different concentration of substrate are enough for obtaining the Michaeli-Menton saturation curve to determine the value of k_{cat}/K_M which reveals selectivity between specific substrate and protein.

4.1.1 Previous *In Vitro* Studies of the SQTGS ER Monodomain

Soluble SQTGS ER monodomain has been previously successfully isolated by the Cox group.^[75] When using acyl-NAC as substrate, it was difficult to determine the K_M value due to the slow rate of the reduction catalyzed by the ER domain. The Cox group have recently

synthesized acyl pantetheines^[75] as the substrates for this reaction. They found their rates of reduction are generally higher than those of acyl-NACs, which makes the determination of k_{cat}/K_M more feasible and accurate. By observing the UV change of NADPH, real-time monitoring of the reaction process is feasible (Figure 4.2).

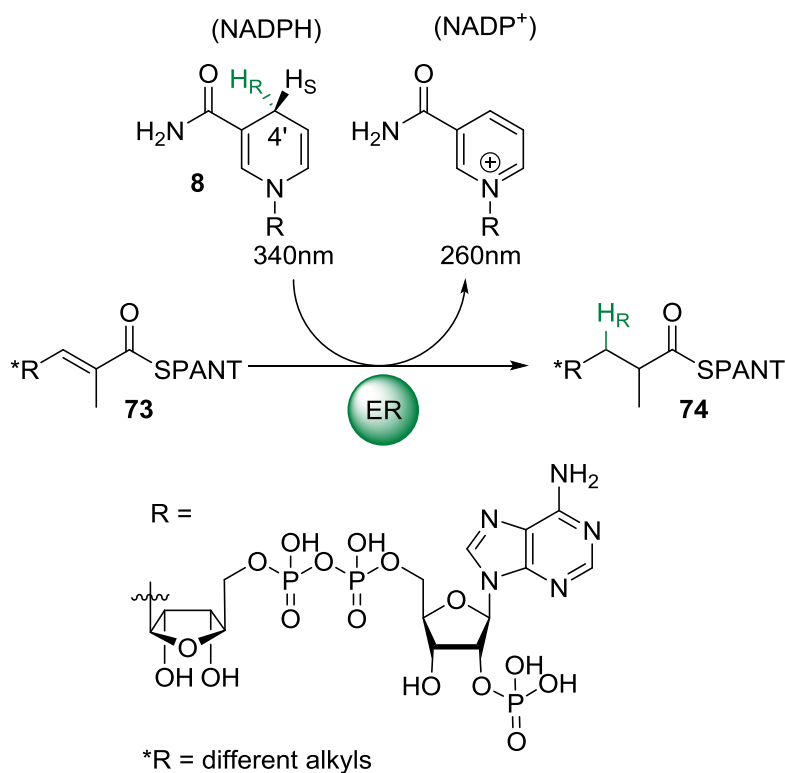


Figure 4.2. Real-time monitoring the UV change of NADPH for *in vitro* ER assay.

According to the Michaelis-Menten equation, the value of k_{cat}/K_M can be used as a measurement of the specific enzyme catalytic reactivity to a specific substrate. In this way, by testing the initial rate with different concentrations of substrates, Michaelis-Menten saturation curves were obtained for various substrates. Comparison of the value of k_{cat}/K_M shows that the ER has varying selectivities for different possible substrates (Figure 4.3).

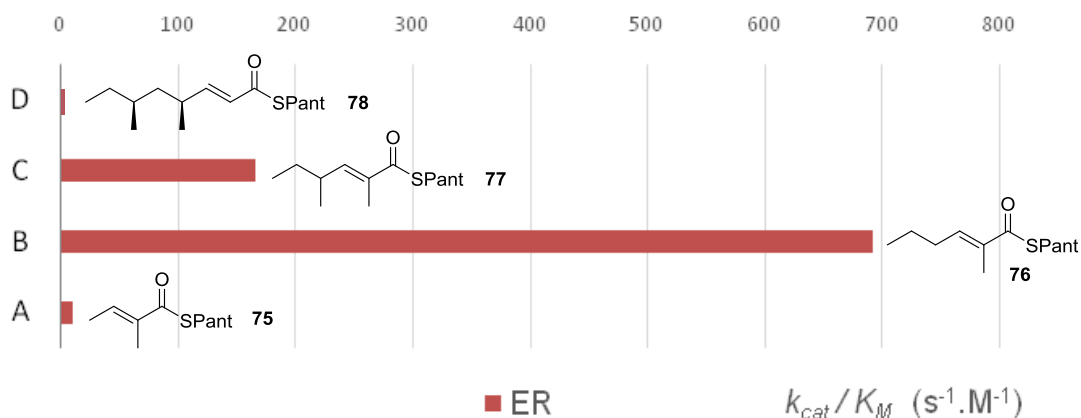


Figure 4.3. Kinetic studies of enoyl reduction by SQTKS ER^[75] with different substrates: **A**, Tigloyl pantetheine **75**; **B**, 2-methylhexenoyl pantetheine **76**; **C**, racemic 2,4-dimethylhexenoyl pantetheine **77**; **D**, 4S,6S-dimethyl-octenoyl pantetheine **78**.

ER substrate recognition appears to be controlled by chain length of the substrate^[75] (Figure 4.3). Tigloyl pantetheine **75**, 2,4-dimethylhexenoyl pantetheine **77** and 4S,6S-dimethyl-octenoyl pantetheine **78** are analogs of natural intermediates in the biosynthesis of squalstatin tetraketide **27**. However, with the highest value of k_{cat}/K_M , the best substrate is the unnatural 2-methylhexenoyl pantetheine **76**, which means the ER domain can act as a general catalyst. While for 4S,6S-dimethyl-octenoyl pantetheine **78**, the final product analog of SQTk, the reaction rate is too slow for observation. This means the ER domain shows its programming function by its *inability* to reduce the final tetraketide substrate. In fact, squalstatin tetraketide analog **78** itself can act as an inhibitor of the ER domain, it is observed to reduce the turnover of tigloyl substrate **75**. This may be due to the space occupation in the active pocket of the ER even though the squalstatin tetraketide is not reduced, which prevents the entry of other substrates.

SQTKS ER appears to be a general catalyst and it shows a low level of selectivity for substrates. Many unnatural substrates can enter the ER and be reduced. This supports the idea that the programming mechanism of highly reducing iterative PKS is based on kinetic

competition among catalytic domains for individual substrates. Because of this, similar substrates may form different products. For example, tigloyl substrate is reduced by ER rather than passed directly to KS for chain-extension by the AT, which is because the ER reacts faster than AT. ^[22]

The 4'-*pro-R* hydrogen of NADPH was proved to be the donor of hydride for the reduction at the β position of tigloyl SNAC from the *Re* face. ^[75] Based on the structure model of isolated ER domain (Figure 4.4 A) by using the crystal structure of mammalian fatty acid synthase (mFAS, PDB 2vz9) as the template, the 4'-*pro-S* hydrogen of NADPH points to the wall of reaction pocket while the 4'-*pro-R* hydrogen is freely available in the pocket for the reduction (Figure 4.4 B). ^[75] Specific contacts of NADPH were observed (SQTKS numbering), with residues S2072, K2055 (nucleotide 2' phosphate), G2029 (diphosphate), I2119 and V2144 (nicotinamide amide). All these residues are conserved in the mFAS ER domain which also uses the 4'-*pro-R* hydrogen of NADPH as the donor of hydride for enoyl reduction.

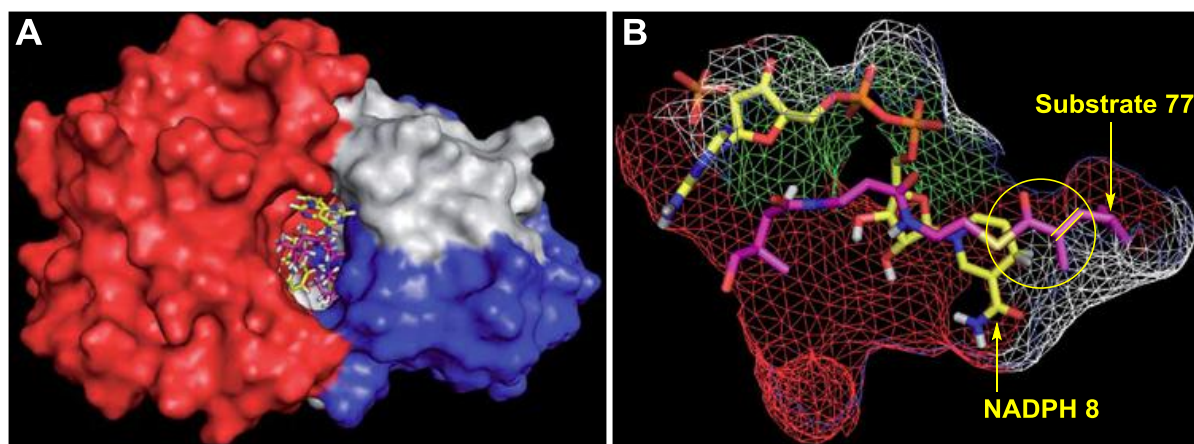


Figure 4.4. **A**, view of the overall modeling structure of SQTKS ER; **B**, view of the active site of SQTKS ER model with substrate **77** and NADPH docked. The reduction happens in the area noted with yellow ring.

Fully constituted SQTCS is able to produce enantiomerically pure products, but the isolated ER can only produce products which are racemic at the α -position. The result of the isolated ER assay shows the loss of the stereoselectivity of the protonation at the α -position of the substrate (Figure 4.5).^[75] Based on a modelled ER structure and related research, there is no suitable proton donor from amino acid residues in the active site for the α -position reprotonation (within 5 Å). For example, Ban and coworkers^[157] suggested K1771 of mFAS is involved in reprotonation. K1771 (mFAS numbering) is conserved in the SQTCS ER as K2121 (SQTCS numbering), but its ϵ -amino group is 9.7 Å distant from the reacting α carbon, which is too far to form a hydrogen bond. Therefore we suggest the proton donor is probably a water molecule. Due to the fact that the active site of the isolated ER monodomain is probably more flexible than that in intact SQTCS, water molecules can ingress from both sides of substrates, which leads to the formation of racemic products.

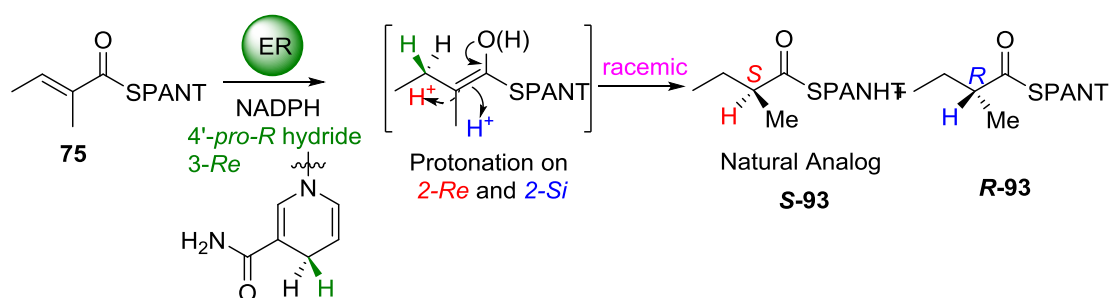


Figure 4.5. Stereoselectivity of the SQTCS ER monodomain.

4.1.2 Previous *In Vitro* Studies of the SQTCS DH Monodomain

A soluble DH mono domain of SQTCS has also been successfully isolated by the Cox group.^[57] The four stereoisomers of 3-hydroxy-2-methylbutyric *N*-acetylcysteamine thioesters were synthesized as the substrates for the dehydration reaction. It was found that the DH domain only dehydrates the (*2R,3R*) substrate **82** and the reaction only forms *E* product **81** through *syn* elimination (Figure 4.6). Here, the stereopreference of DH is also identical between SQTCS and mFAS. Based on the model structure,^[57] the active site residues involved are

H1034 (α -H), D1225 (hydroxyl), Y1041 and P1042 (SQTKS numbering, bound to NAC), which are conserved in mFAS. The mechanism is suggested as an *E1cb* process.

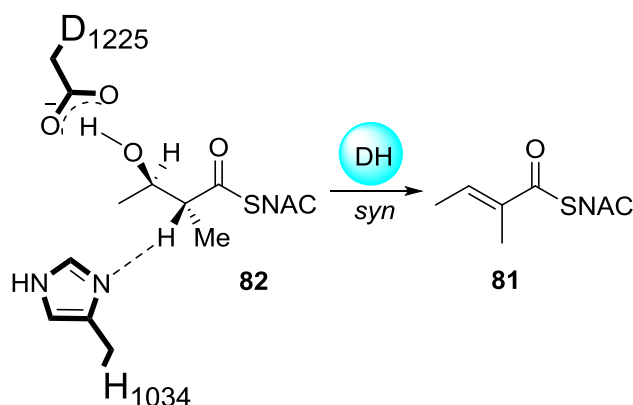


Figure 4.6. Stereochemical course of KR, DH and ER domains of SQTKS.

(2*R*,3*R*)-3-hydroxy-2-methylbutyric NAC **82** is the substrate for SQTKS DH, so the 2*R*,3*R* stereochemistry must also be the product of SQTKS KR which then also shows the same stereochemical preference as mFAS KR.^[41] The remaining question is which hydride of the cofactor NADPH is transferred by KR. In addition, the stereochemistry of SQTKS CMeT is still unknown due to formation of the planar intermediate **81** by the DH domain during β -processing.

4.2 Aims

From the previous *in vitro* assay of isolated ER and DH domains, we have mainly studied their stereochemistry in the first cycle (Figure 1.19 R1) of biosynthesis of squalestatin tetraketide.^[57] The aims of the work described next are: whether a multi-domain protein can partially or fully revive the stereopreference of the enoylreductase; which hydride of the cofactor NADPH is transferred by the SQTKS ketoreductase; and whether the methyltransferase has a stereopreference towards substrates. To answer these questions we need to be able to produce (chapter 3) and assay soluble multi-domain fragments of SQTKS, and also design new assays (for example for the CMeT domain), and test the stability of

substrates and products under the assay conditions. Particularly, the SQTKS DH-KR tetradomain will allow a full β -processing cycle to be studied *in vitro* for the first time for a fungal PKS, and it is interesting to observe whether the domain selectivities observed for the isolated domains are preserved or changed in the multi-domain proteins versus the monodomains already studied.

4.3 Stability of Different Substrates

In addition to the activity and accessibility of substrates, substrate stability is another important issue which must be considered, especially when one enzyme shows low *in vitro* activity (meaning long assay times) towards analogous substrates. Substrates involved in the polyketide synthesis are thioesters which might be fragile in the presence of nucleophiles. Thioesters are sensitive to acid and base catalyzed hydrolysis which makes substrates and products potentially unstable. However, hydrolysis is also important to the programming of the biosynthesis. According to the proposed kinetic competition mechanism^[22] (section 1.3.3), it is possible that only when the rate of extension or modification reaction is lower than the rate of hydrolysis, squalestatin tetraketide can finally be released from the biosynthetic protein. Therefore, ensuring the stability of substrates and products in solution will be important during the kinetic study of SQTKS catalysts.

Acyl-NAC substrates are widely used in *in vitro* assays of PKS. In order to qualitatively test the stability of acetoacetyl NAC **83a** (the first intermediate of PKS) in different buffers, 5 mM of **83a** was dissolved in H₂O (Figure 4.7A), 50 mM of pH 7.4 phosphate buffer (commonly used buffer with weak nucleophilicity) (**B**), or the same buffer containing NaCl (150 mM) (**C**) or both NaCl (150 mM) and glycerol (20 %) (**D**). In these buffers acetoacetyl NAC **83a** were incubated at 25 °C and monitored by LCMS after 24 h. The results showed that addition of glycerol (Figure 4.7 D), a nucleophile, accelerates the hydrolysis. This was reflected by the decreased amount of substrate. Without glycerol, in conditions of **A**, **B** and **C** the substrate **83a** remained stable for 24 h.

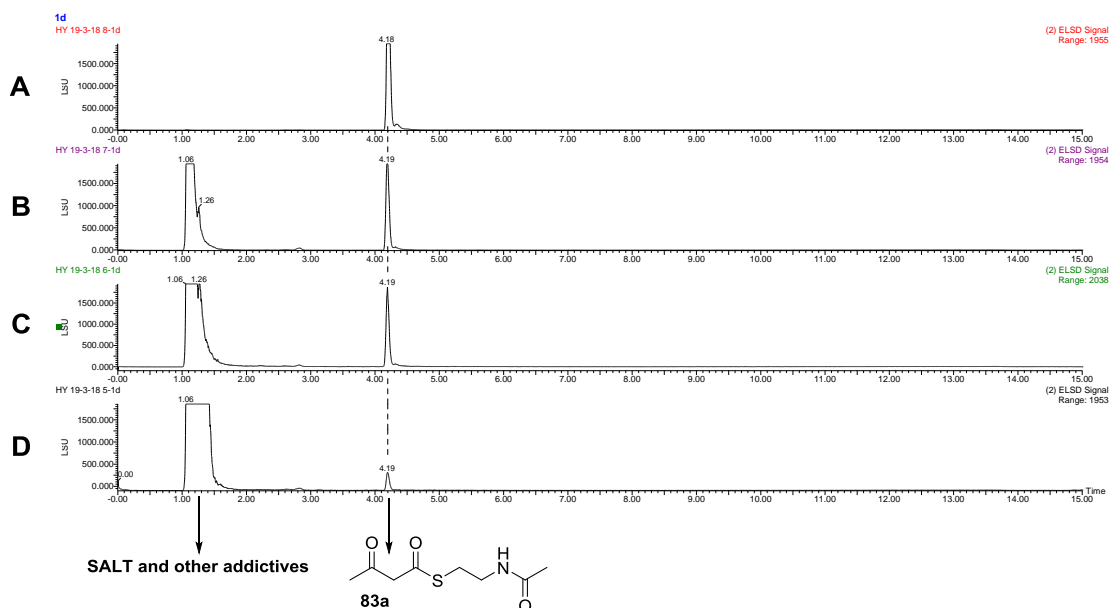


Figure 4.7. LCMS analysis of incubation of **83a** (5 mM) at 25 °C for 24 h in the presence of various additives: **A**, H₂O; **B**, 50 mM phosphate buffer (pH 7.4); **C**, 50 mM phosphate buffer (pH 7.4) containing 150 mM NaCl; **D**, 50 mM phosphate buffer (pH 7.4) containing 150 mM NaCl and 20 % glycerol.

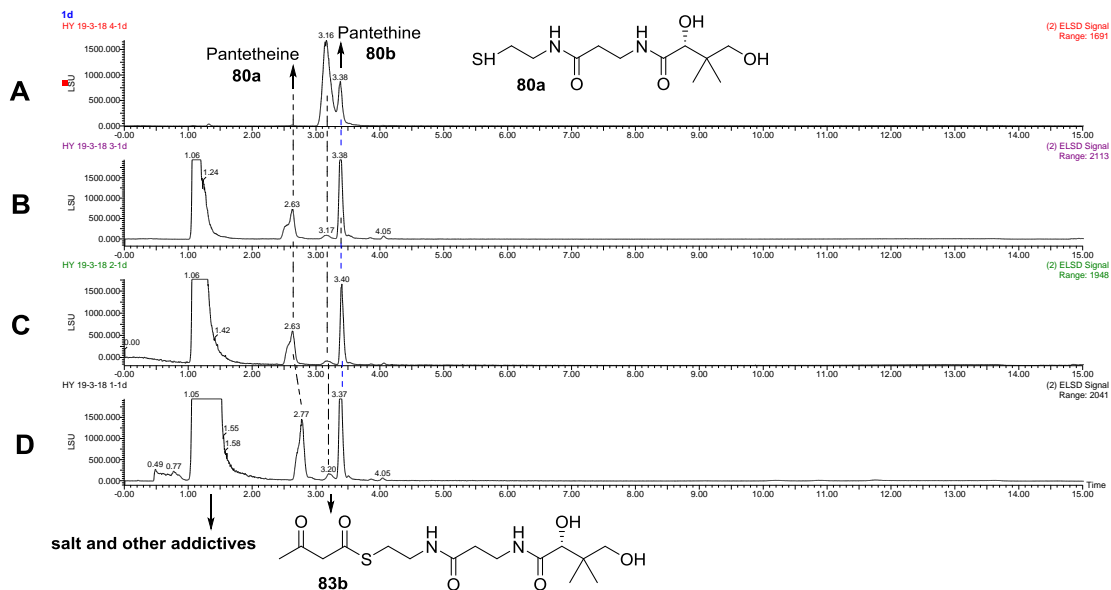


Figure 4.8. LCMS analysis (ELSD) of incubation of **83b** (5 mM) at 25 °C for 24 h in the presence of various additives: **A**, H₂O; **B**, 50 mM phosphate buffer (pH 7.4); **C**, 50 mM phosphate buffer (pH 7.4) containing 150 mM NaCl; **D**, 50 mM phosphate buffer (pH 7.4) containing 150 mM NaCl and 20 % glycerol.

We next examined the stability of acyl pantetheine substrate in these buffers. Acetoacetyl pantetheine **83b** was prepared and incubated in the same four conditions tested for the related SNAC substrate **83a**. After the incubation for 24 h, nearly all the acetoacetyl pantetheine **83b** has been hydrolyzed in phosphate buffer with different additives (Figure 4.8 B, C, D). In fact, serious hydrolysis of the pantetheine substrate can be observed after just 1 hour (Figure 4.9). The hydrolysis, indicated by the formation of pantethine **80b** (the oxidized disulfide dimer of **80a**) increases in the order of water (Figure 4.9 A), phosphate buffer (**B**) and phosphate buffer containing 20 % glycerol (**D**). Addition of *sodium* chloride in buffer (Figure 4.9 C) shows no significant effect toward the extent of hydrolysis.

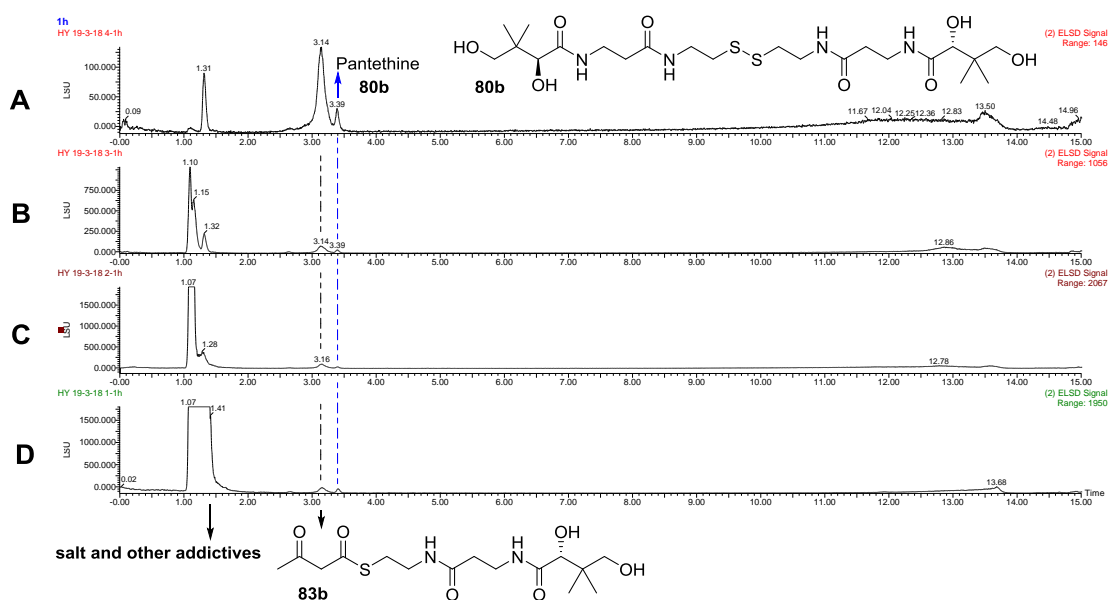


Figure 4.9. LCMS analysis (ELSD) of incubation of **83b** (5 mM) at 25 °C for 1 h in the presence of various additives: **A**, H₂O; **B**, 50 mM phosphate buffer (pH 7.4); **C**, 50 mM phosphate buffer (pH 7.4) containing 150 mM NaCl; **D**, 50 mM phosphate buffer (pH 7.4) containing 150 mM NaCl and 20 % glycerol.

In comparison with Figure 4.9 A and B, much slower hydrolysis happens in water without phosphate ions. Therefore, when pantetheine substrate such as **83b** exists, phosphate ions may play a role in accelerating the hydrolysis of the thiolester. It is reported that phosphate buffer can catalyze the hydrolysis of 2, 4-dinitrophenyl benzoate esters.^[158] The HPO_4^{2-} ion is the active catalytic species and the mechanism of catalysis is nucleophilic. However, this catalysis of hydrolysis by phosphate in our assay must be in collaboration with the pantetheine structure, as much less hydrolysis occurs when only acyl-NAC substrate such as **83a** is used (Figure 4.7). Therefore, the main reason why serious hydrolysis of acyl pantetheine substrate happens may be due to the reactivity of the acyl pantetheine substrate itself. As the thiolester is less stable than the carboxylate ester, the carboxylate ester can be formed by intramolecular ester exchange reactions with the hydroxyl groups at the end of the molecule (Figure 4.10).

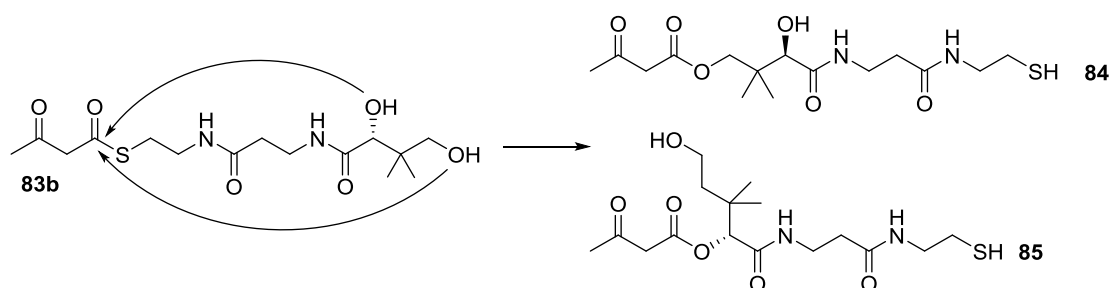


Figure 4.10. Intra-molecular ester exchange between acyl pantetheine **83b**.

The formation of structural isomer **84** or **85** can be observed by LCMS (Figure 4.11). The amount of isomers in phosphate buffer containing glycerol (Figure 4.11 C and D) is much more than that in water (Figure 4.11 A and B), which indicates the nucleophilicity of buffer accelerates the intramolecular ester exchange. The product of intermolecular ester exchange can also be observed in the LCMS data based on their molecular weight (Figure 4.12), while no purification has been done towards these products for further identification.

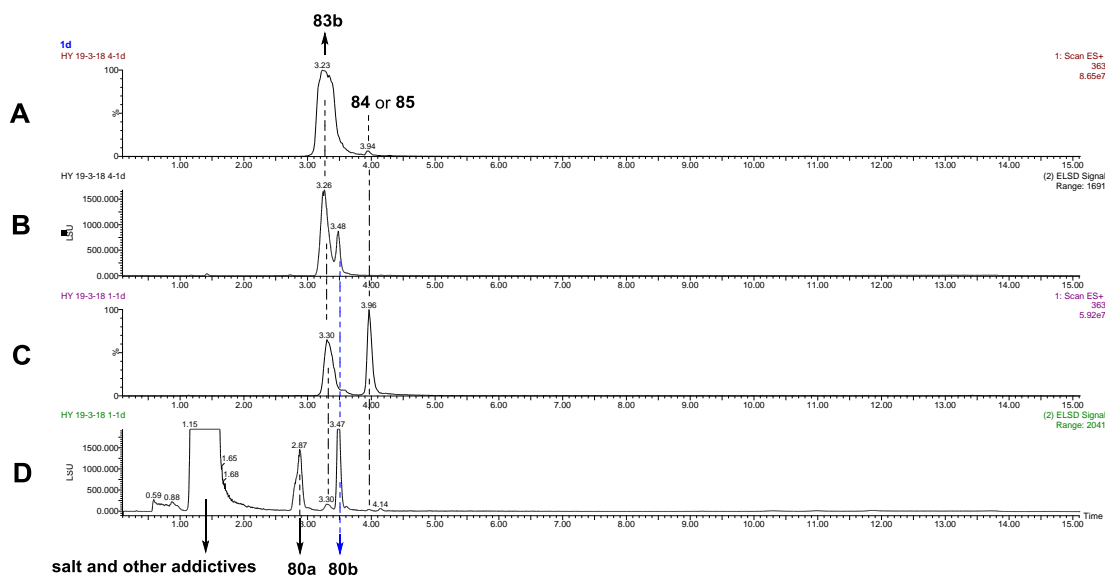


Figure 4.11. Observation of **84** or **85** by LCMS analysis (ELSD) of incubation of **83b** (5 mM) at 25 °C for 24 h in the presence of various additives: **A**, EIC signal in H₂O; **B**, ELSD signal in H₂O; **C**, EIC signal in 50 mM of phosphate buffer (pH 7.4) containing 150 mM of NaCl; **D**, ELSD signal in 50 mM of phosphate buffer (pH 7.4) containing 150 mM of NaCl and 20 % glycerol. EIC signal: Extracted Ion Chromatogram (EIC) in ES⁺ mode for $m/z=363$ **83b** correspond to isomers **84** or **85**.

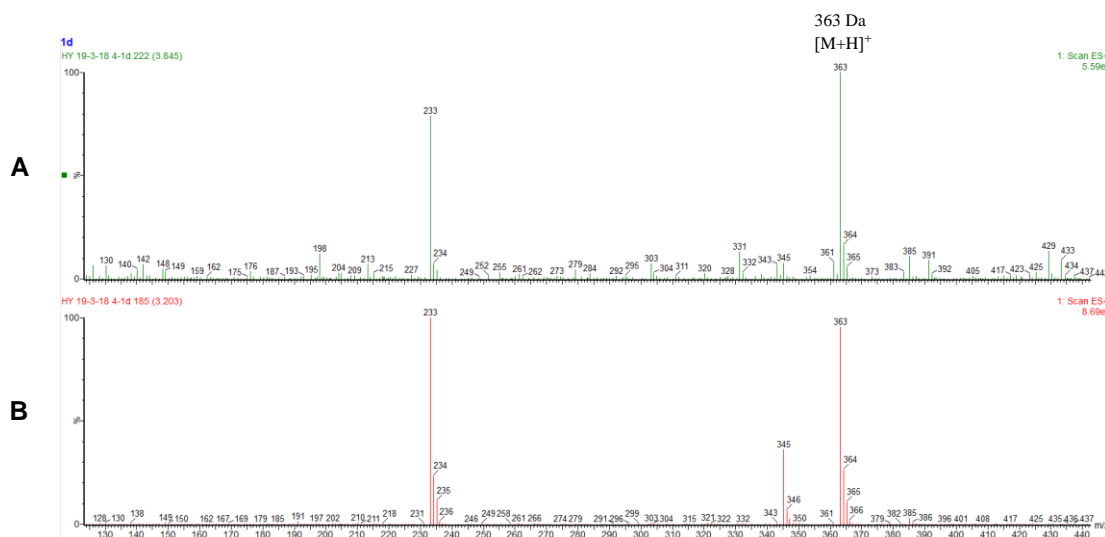


Figure 4.12. Mass spectra of peaks from Figure 4.11: **A**, MS of 3.9 min peak (**84** or **85**); **B**, MS of 3.2 min peak (**83b**).

The rates of hydrolysis towards different acyl pantetheine substrates are different. Although precise kinetic measurements were not carried out, we estimate $t_{1/2}$ for hydrolysis of acetoacetyl pantetheine **83b** as around 6 hours. However, the hydrolysis of 2-methylated acyl

panetheine substrate **88** is considerably slower than non-2-methylated acyl pantetheine substrate (Figure 4.13), because the steric hinderance of the methyl group slows down the ester exchange.

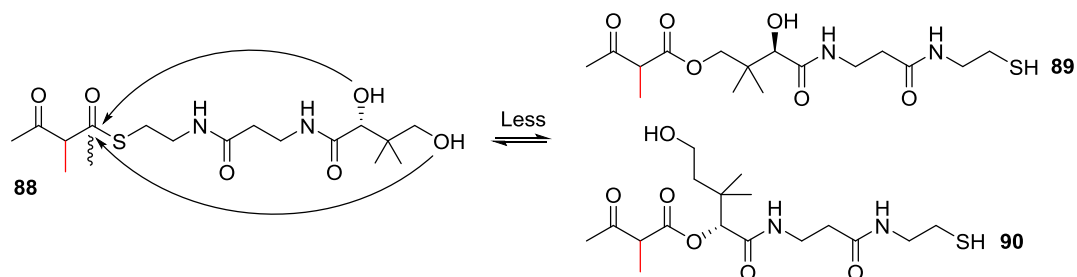


Figure 4.13. Inter- or intra-molecular ester exchange of 2-methylated acyl pantetheine **88**.

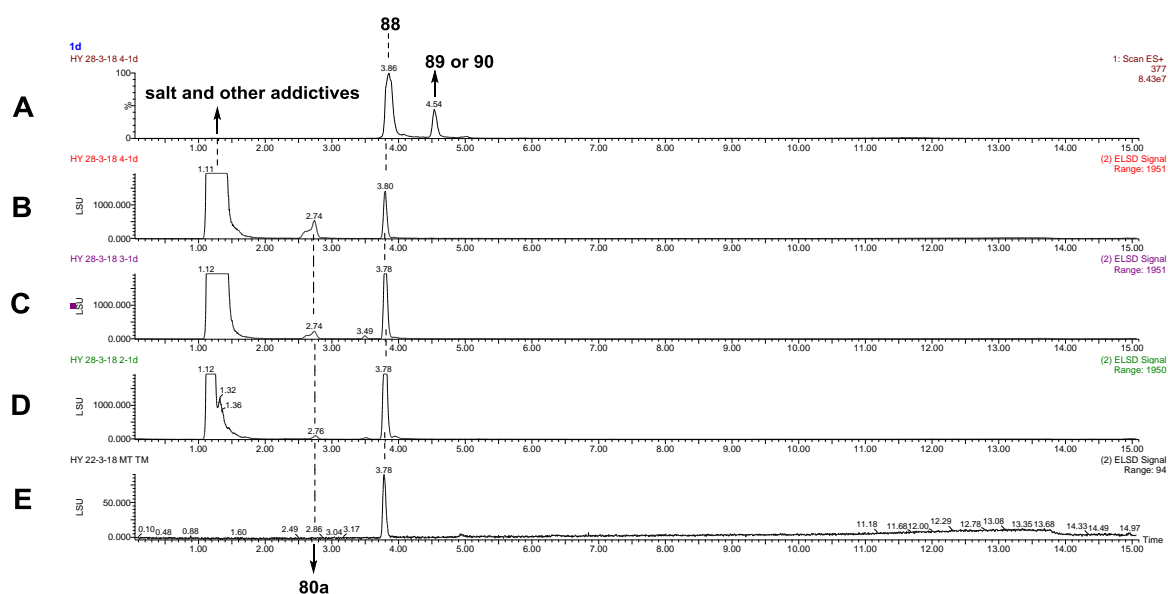


Figure 4.14. Observation of **89** or **90** by LCMS analysis (ELSD) of incubation of **88** (5 mM) at 25 °C for 24 h in the presence of various additives: **A**, EIC signal in Tris buffer (pH 8) containing 150 mM NaCl, 20 % glycerol and 1 mM EDTA; **B**, ELSD signal in Tris buffer (pH 8) containing 150 mM NaCl, 20 % glycerol and 1 mM EDTA; **C**, ELSD signal in 50 mM phosphate buffer (pH 7.4) containing 150 mM NaCl and 20 % glycerol; **D**, ELSD signal in 50 mM phosphate buffer (pH 7.4) containing 150 mM NaCl; **E**, ELSD signal in H₂O.

Most 2-methyl acetoacetyl pantetheine **88** still exists after incubation at 25 °C for 24 h (Figure 4.14). With the nucleophilic effect of the buffer and pH increases, the hydrolysis

accelerates (Figure 4.14 B-C-D-E). Therefore, for the future assay, to minimize the hydrolysis, 50 mM phosphate buffer (pH 7.4) containing 150 mM NaCl will be used as reaction buffer.

Although the stability of acyl pantetheine substrate is less than the stability of acyl-NAC substrate in buffer, the reactivity of acyl pantetheine substrate is still higher than that of acyl-NAC substrate according to the ER research.^[75] Therefore, for the following *in vitro* assay, with acyl pantetheine substrates, real time UV is used to obtain the initial rate of the reaction as soon as possible for determination of the kinetic constant. While LCMS, which takes long time, can only be used to detect the formation of desired products without analyzing the kinetics.

4.4 Kinetic Studies of Enoyl Reduction

4.4.1 SQTGS DH-ER Tridomain

Although the soluble SQTGS DH-ER tridomain protein is not pure (section 3.2.1), *in vitro* research can still be performed if the contaminant protein is of no influence. However, the DH-ER tridomain is quite unstable and only a trace of reduction product can be observed by LCMS after 24 h incubation at 25 °C in the phosphate buffer (pH 7.4) containing 150 mM of NaCl (Figure 4.15).

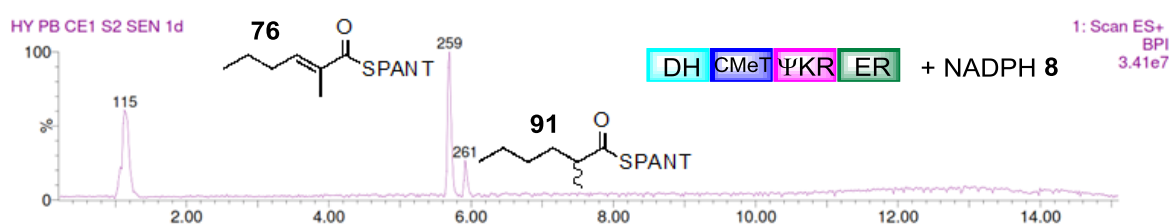


Figure 4.15. LCMS analysis of a 24 h incubation of **76** (2 mM) in the presence of SQTGS DH-ER tridomain (0.6 mg) and NADPH (5 mM) at 25 °C in 50 mM phosphate buffer (0.2 mL, pH 7.4, with 150 mM NaCl).

As the purity of protein increases during the purification process, the specific activity units per mg of total protein should increase due to the removal of impurities. By observing

the enoyl reduction reaction through UV change of NADPH (Figure 4.2), we measured the reactivity units per mg of protein in the isolated ER monodomain and in the DH-ER tridomain. For DH-ER tridomain, we observed an obvious loss of reactivity during purification (Blue in Figure 4.16). This indicates the DH-ER tridomain is unstable. Addition of protease inhibitor cocktail and reducing agent could not make it stable. In contrast the isolated ER increased its specific activity during purification as expected (Red in Figure 4.16).

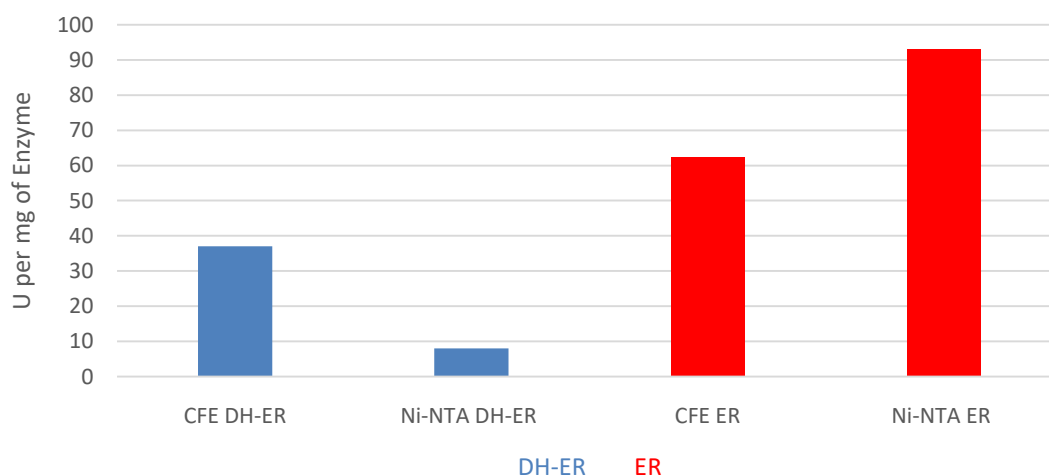


Figure 4.16. Comparison of stability between DH-ER tridomain and ER monodomain during purification. Reaction condition: 25 °C, 360 μ L of buffer (50 mM Tris-HCl pH 8.0, 150 mM NaCl, 20% Glycerol), 10 μ L of NADPH (0.25 mM), 20 μ L of enzyme (original protein sample volume and concentration: CFE DH-ER, 67 mL, 4.1 $\text{mg}\cdot\text{mL}^{-1}$; Ni-NTA DH-ER, 1.25 mL, 2.6 $\text{mg}\cdot\text{mL}^{-1}$; CFE ER, 40 mL, 0.8 $\text{mg}\cdot\text{mL}^{-1}$; Ni-NTA ER, 1.4 mL, 5.4 $\text{mg}\cdot\text{mL}^{-1}$), 10 μ L of 2-methylhexenoyl pantetheine **41** (0.125 mM). One specific activity unit (U) is 0.006 $\text{mM}\cdot\text{min}^{-1}$ (consumption of NADPH at 340 nm).

4.4.2 SQTGS DH-KR Tetradomain

Plenty of tetradomain protein has been purified by nickel affinity chromatography (section 3.3.3.2). Even though the purified DH-KR (by purification optimization) is not pure (50 % purity), it is stable and active enough for further research. By monitoring the consumption of NADPH at 340 nm, the initial rate of enoyl reduction can be tested. Under a certain condition,

by testing the initial rate at different concentrations of the substrate, we can draw the curve (Figure 4.1) to see the relation between initial rate and concentration. The relation follows the Michaelis-Menten equation ($V_0 = V_{\max}[S]/(K_M+[S])$, $V_{\max} = k_{cat}[E_0]$). The value of k_{cat}/K_M , the initial slope of this curve, can be used as a measurement of the specific enzyme catalytic reactivity to a specific substrate.

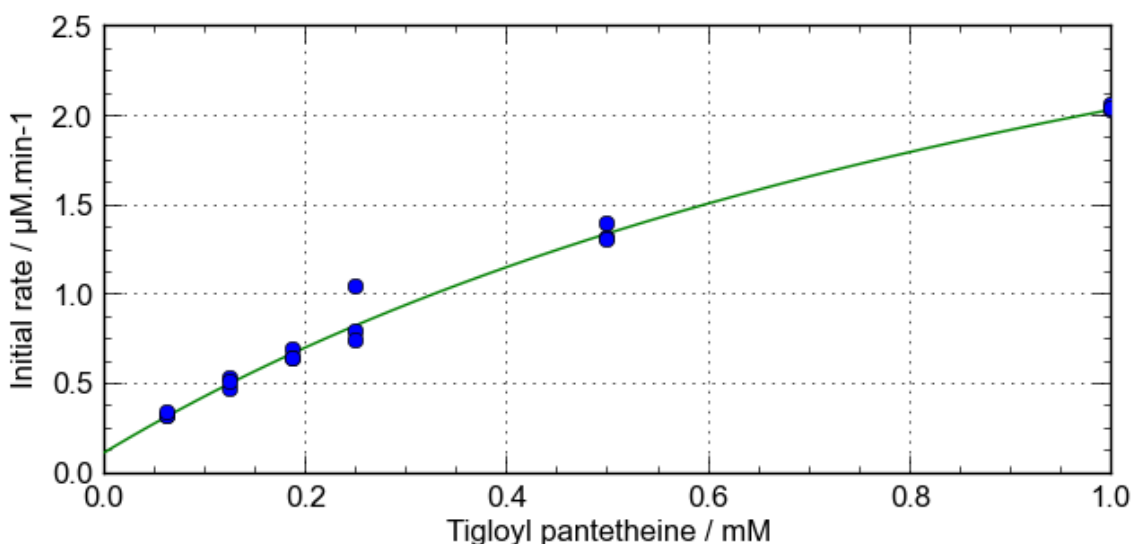


Figure 4.17. Michaelis-Menten saturation curve of tigloyl pantetheine substrate **75** with the SQTGS DH-KR tetradomain. Reaction condition: 25 °C, 340 nm, buffer (50 mM Tris-HCl, pH 8.0, 150 mM NaCl, 20% Glycerol), 10 μL of 0.25 mM NADPH, 20 μL of 0.15 nM DH-KR (40% purity after Ni-NTA), and addition of 10, 20, 30, 40, 80, 160 μL of 2.5 mM **75**, respectively, to make a total 400 μL of reaction system including the buffer. Each condition tested three times.

The DH-KR protein was tested with same four substrates (**75**, **76**, **77**, **78**) used in the previous *in vitro* ER assay (Figure 4.3). The results show that the substrate preference between ER and DH-KR are similar (Figure 4.18). Like ER domain, DH-KR can act as a general catalyst and the protein shows its programming function by its inability to reduce the final tetraketide substrate. The rate of enoyl reduction catalyzed by the ER monodomain and the DH-KR tetradomain is of no significant difference. In detail, the DH-KR protein seems

slightly more active for monomethylated substrate **76** than that of ER protein, while dimethylated substrate **77** appears similar. The difference may be due to the active site of the ER in the DH-KR tetradomain being more native than that of ER monodomain.

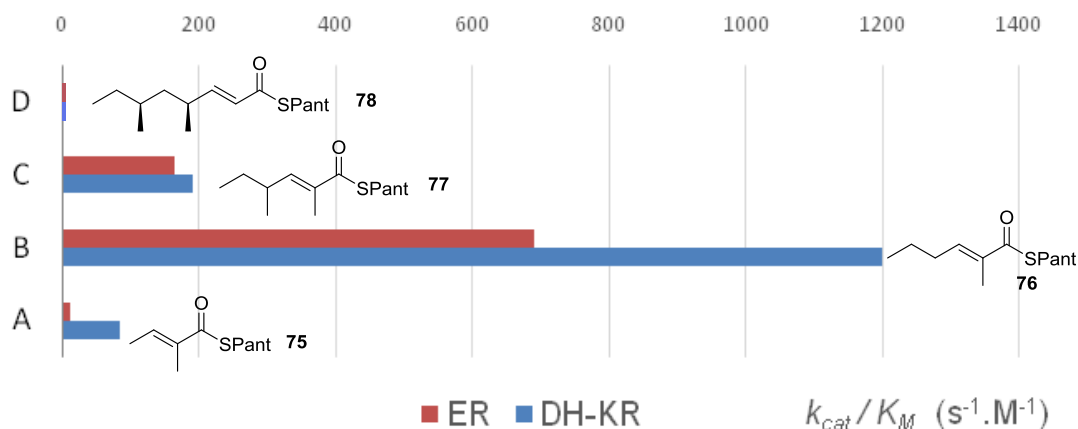


Figure 4.18. Kinetic studies of enoyl reduction between the ER monodomain (Red) and the DH-KR tetradomain (Blue). Substrates involved are Tigloyl pantetheine **75** (A), 2-methylhexenoyl pantetheine **76** (B), 2, 4-dimethylhexenoyl pantetheine **77** (C), 4S, 6S-dimethyl-octenoyl pantetheine **78** (D). See detailed information in section 6.4.3.

4.5 Studies of Stereoselectivity in Enoyl Reduction by SQTGS DH-KR Tetradomain

The SQTGS DH-KR is a multi-domain protein containing four individual catalytic domains and all connections among them. Compared to the isolated ER, the ER in DH-KR is structurally more similar to the ER in intact SQTGS. We therefore decided to test whether the stereoselectivity of the ER-catalyzed reaction is different between the multi-domain protein and the isolated ER.

To measure the stereochemistry of DH-KR at the substrate α -position, we used the method developed in single ER research.^[75] After incubation at 25 °C for 2 days (Figure 4.19)

with DH-KR (40 % purity after Ni-NTA), we hydrolyzed the reduction product of tigloyl pantetheine **75** to get the 2-methylbutyric acid (see chapter 6). 10 μmol of 2-methyl butyric acid **60** (produced from tigloyl **75** or angelic **92** pantetheine with the catalysis of DH-KR) was dissolved in 0.6 mL of CDCl_3 . (1*R*,2*R*)-(+)-1,2-diphenylethylenediamine **95**, the chiral auxiliary was added later to form the diastereoisomers *R*-**96** and *S*-**96** which can be observed by proton NMR (Figure 4.19 A). As controls, we treated commercially available *S*-2-methyl butyric acid *S*-**94** and racemic 2-methyl butyric acid **94** with the same chiral auxiliary **95**. In the absence of the chiral auxiliary, ^1H NMR resonance of racemic **94** in the 0.9~1.2 ppm region of the spectrum showed the expected triplet (β -methyl protons) and doublet (α -methyl protons). In the presence of the chiral auxiliary the signals were doubled in the expected 1:1 ratio (Figure 4.19 A). Enantiomerically pure *S*-**94** gave a single set of peaks in the presence of the chiral auxiliary (Figure 4.19 B).

Angelic pantetheine **92** was incubated with SQTGS DH-KR protein. The product of this reaction showed a 1:7 mixture of stereoisomers (Figure 4.19 C). Similar to this, tigloyl pantetheine substrate **75** gave a 1: 6 mixture of stereoisomers (Figure 4.19 D). In order to know which configuration is *R* and which configuration is *S*, we added into the 1:6 mixture two equivalents of commercial *S*-2-methyl butyric acid *S*-**94**. The ratio raised from 1:6 to 1:18 (Figure 4.19 E), so we knew the larger peak corresponds to *S*-**94** and the small peak corresponds to *R*-**94**. Since no racemization had been found during the hydrolysis of acyl-SNAC thiolester,^[75] no α -racemization should happen to its related pantetheine thiolester during the hydrolysis. As a result, by using DH-KR, both tigloyl pantetheine and angelic pantetheine give similar result with an *ee* value of 71% ~ 75% (*S*:*R* = 6:1 ~ 7:1) compared to the racemic results by using the ER monodomain.

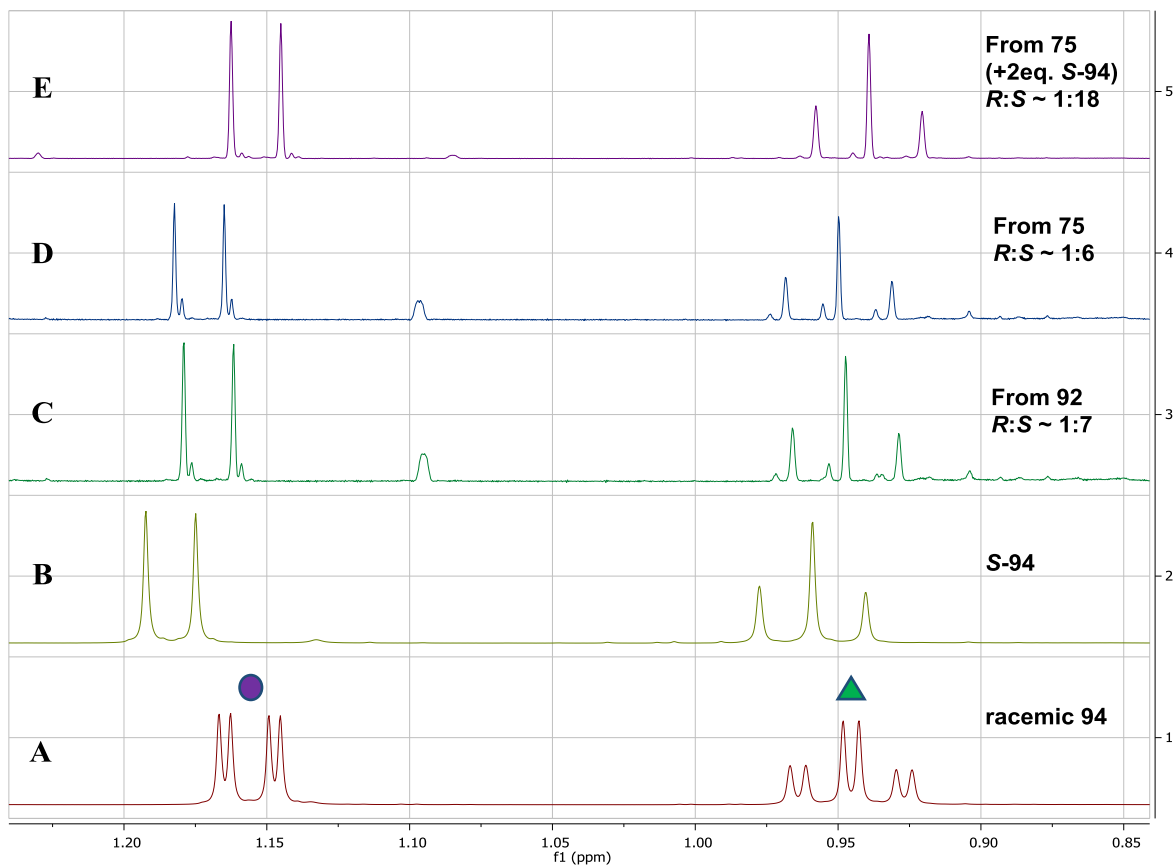
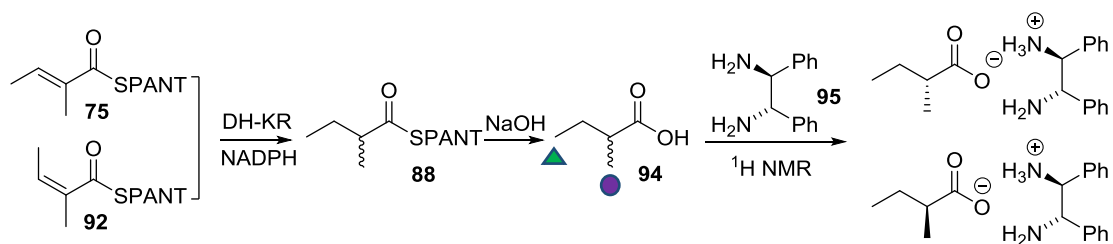


Figure 4.19. ¹H NMR analysis of the stereochemical outcome of enoyl reduction catalyzed by the SQTKS DH-KR tetradomain protein: **A**, 10 μmol of racemic 2-methylbutyric acid **88** with 10 μmol of (1*R*,2*R*)-(+)-1,2-diphenylethylenediamine **95**; **B**, 10 μmol of commercial *S*-2-methylbutyric acid **S-94** with 10 μmol of **95**; **C**, 10 μmol of angelic pantetheine **92** produced 2-methylbutyric acid **94** (*R:S* = 1:7) with 14 μmol of **95**; **D**, 10 μmol of tigolyl pantetheine **75** produced 2-methylbutyric acid **94** (*R:S* = 1:6) with 14 μmol of **95**; **E**, 10 μmol of tigolyl pantetheine **75** produced 2-methylbutyric acid **94** (*R:S* = 1:18) with 42 μmol of **95** plus 20 μmol of commercial *S*-**94**.

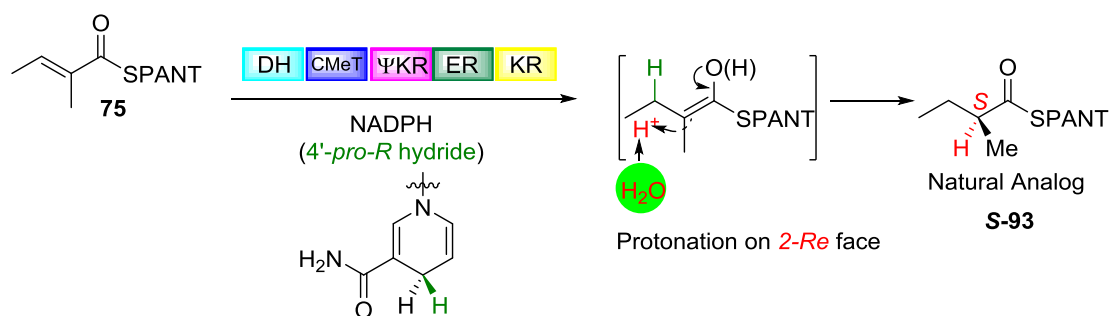


Figure 4.20. ER stereoselectivity of DH-KR tetradomain at α -position.

This result is consistent with our ideas developed from observation of the stereoselectivity of the single ER research.^[75] A water molecule is probably the source of the α -protonation (Figure 4.20). The ER in the DH-KR construct may be narrower and less flexible than that of single ER domain, which allows less water molecule into the active site, and more selective positions on *Re* face. In the tetradomain construct the stereocontrol of ER reprotonation is better, but still not perfect. It is likely that ER in the whole SQTKS construct is even more rigid, which can achieve the best stereocontrol at α -position of the substrate.

4.6 Studies of Stereoselectivity in Keto Reduction by SQTKS DH-KR Tetradomain

To understand which 4'-hydride of NADPH is responsible for the reduction catalyzed by the KR domain, we used the same method developed for the single ER.^[75] 4'*R*-[4'-²H]-NADPD was prepared from a mixture of *Thermoanaerobium brockii* alcohol dehydrogenase (TbADH), NADP⁺ and *isopropanol-d₈* in pH 9 tris buffer (section 6.4.2). This solution was incubated at 43 °C and the progress of the reaction was monitored by UV to observe the formation of NADPD. When no further change in absorption (340 nm) was observed (90% complete in 1 h), the formed acetone and the remaining *isopropanol-d₈* was removed *in vacuo*. The quality of 4'*R*-[4'-²H]-NADPD **8a** (747.2 [M+H]⁺) was confirmed by MS (Figure 4.21 B) with the comparison to NADPH **8** (746.2 [M+H]⁺, Figure 4.21 A).

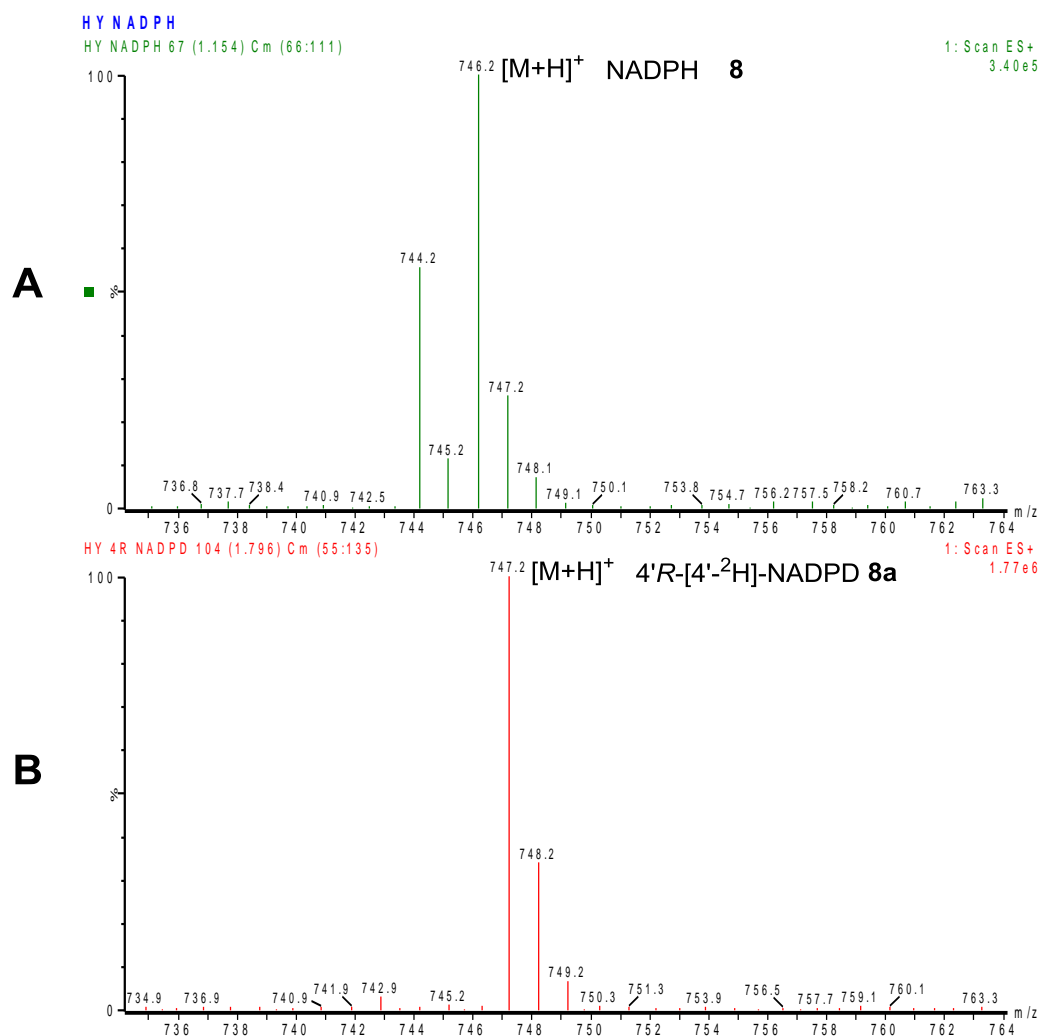


Figure 4.21. Quality of prepared 4'R-[4'-²H]-NADPD **8a** compared to commercial NADPH **8**: A, **8**; B, **8a**.

Based on the previous DH *in vitro* studies,^[57] the DH only takes (2*R*, 3*R*)-2-methyl-3-hydroxyl butanoyl pantetheine **95** as the substrate. Therefore, the **95** should be the product of keto reduction (Figure 4.22). As the KR domain may have the ability of epimerization (section 1.4.3), either *S*-**88** or *R*-**88** could be the potential substrate, which can be shown as an equilibrium might exist via an enol intermediate (Figure 4.22). If the hydride comes from 4'-*pro-R*, the deuterium will attack the 3-carbonyl to form a deuterated alcohol product. On the contrary, if the proton comes from 4'-*pro-S*, no deuterated alcohol product will be found.

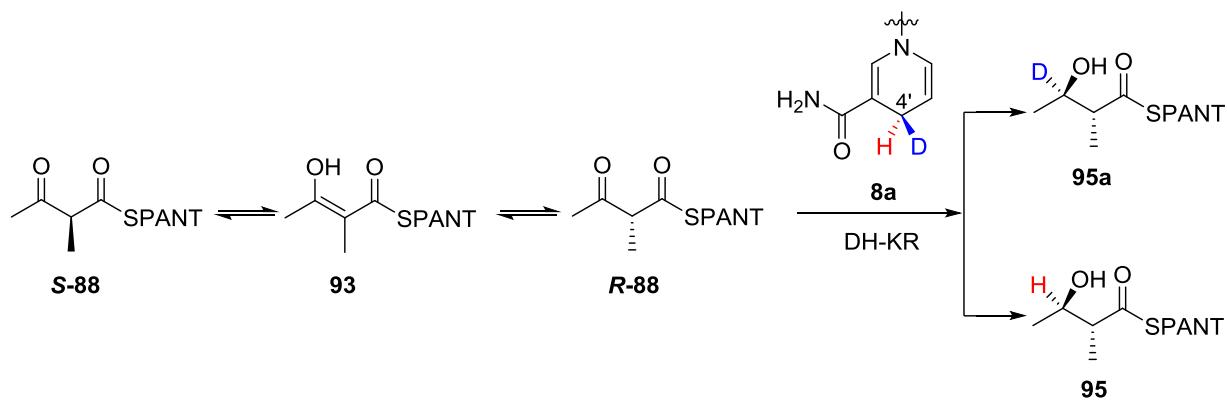


Figure 4.22. Predicted different reduction results by 4'*R*-[4'-²H]-NADPD **8a** in keto-reduction.

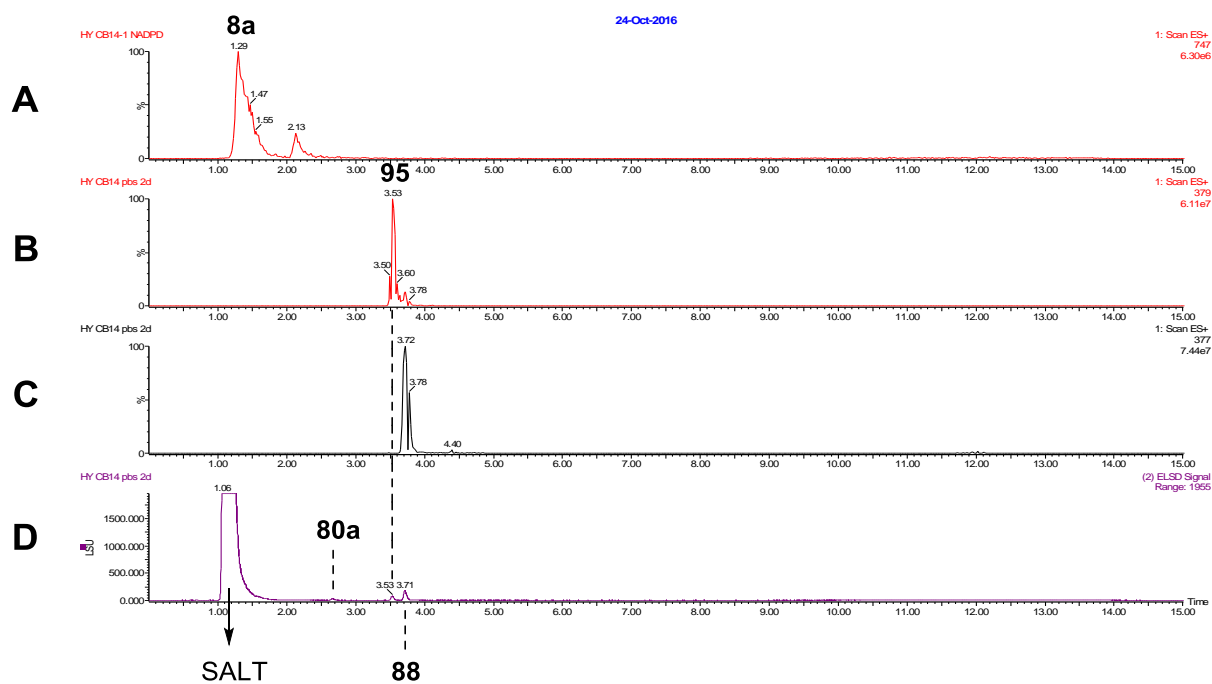


Figure 4.23. LCMS detection of formation of alcohol product from the 4'*pro*-S proton of NADPH: **A**, EIC signal of prepared 4'*R*-[4'-²H]-NADPD **8a**; **B**, EIC signal of produced alcohol product **95** (379 [M+H]⁺); **C**, EIC signal of starting material **88** (377 [M+H]⁺); **D**, ELSD signal of keto-reduction reaction after two days.

In this experiment, 2 mM racemic 2-methyl acetoacetyl pantetheine **88** was incubated with 5 mM 4'*R*-[4'-²H]-NADPD **8a** and 10 μM DH-KR (40 % purity after Ni-NTA) in 0.2 mL phosphate buffer (pH 7.4, 150 mM NaCl), at 25 °C for 2d. Due to the secondary kinetic isotope effect, the reaction proceeded slowly, although we can still observe the formation of alcohol product (Figure 4.23). By measuring the mass of the alcohol product **95**, we observed no incorporation of deuterium, which means the hydride used in the reduction catalyzed by DH-KR is 4'-*pro-S* hydride of NADPH. This result is in accord with the KR stereochemistry of mFAS, which also uses the 4'-*pro-S* hydride.

There are many identical features between fungal HR-PKS and mFAS (section 1.3). In our recent research of the single ER^[75] domain and single DH^[57] domain of SQTKS, we found the stereochemical preferences are identical between SQTKS ER and mFAS ER, and also between SQTKS DH and mFAS DH: In ER reaction, they both transfer 4'-*pro-R* hydride of NADPH onto the 3-*Re* face of α,β -unsaturated substrate; in DH domain, the next β -processing reaction, they both only dehydrate 3*R*-hydroxyl substrate and give *E*- α,β -unsaturated product by *syn* elimination. Now, same stereochemical preference occurs between SQTKS KR and mFAS KR: they both transfer 4'-*pro-S* hydrogen of NADPH onto the 3-*Si* face of their β -carbonyl substrates. Therefore, the fundamental mechanisms of reaction of KR, DH and ER, particularly in stereoselectivity, are identical between SQTKS and mFAS, which reinforces the idea that fungal HR-PKS and mFAS evolved from a common ancestor.

4.7 Monitoring the β -processing Modification by LCMS of DH-KR Tetradomain

In order to know if all the domains of the SQTKS DH-KR tetradomain are active, we used LCMS to monitor the formation of product over a period of hours. We started with racemic 2-methyl diketide substrate **88** (Figure 4.24 Scheme) to observe whether KR, DH and ER of the DH-KR tetradomain can all work together. With the mixture of 2 mM substrate **88**, 5 mM NADPH and 10 μ M DH-KR (50 % purity after Ni-NTA by purification optimization) in 0.2 mL phosphate buffer (pH 7.4, 150 mM NaCl), we found all related products can be discovered after 6 hours' incubation at 25 $^{\circ}$ C (Figure 4.24 and Figure 4.25). Interestingly, very little dehydrated product **75** can be found (4.9 min Figure 4.25 A). At the same time, we observed LCMS peaks corresponding to pantetheine **80a** and its oxidized disulfide dimer **80b**, which means some substrates or products have been hydrolyzed as a side reaction.

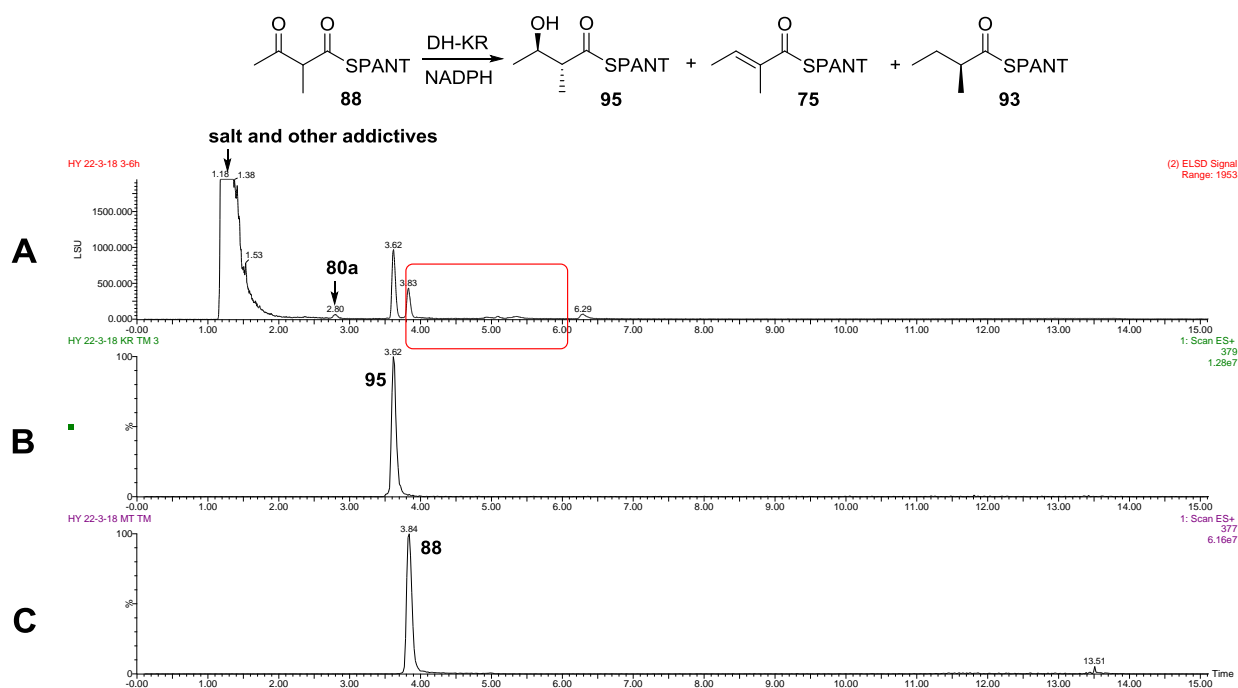


Figure 4.24. LCMS analysis of products of overall β -processing modification after incubation for 6 h: **A**, ELSD signal of reaction; **B** and **C** are EIC signal of **95** and **88**.

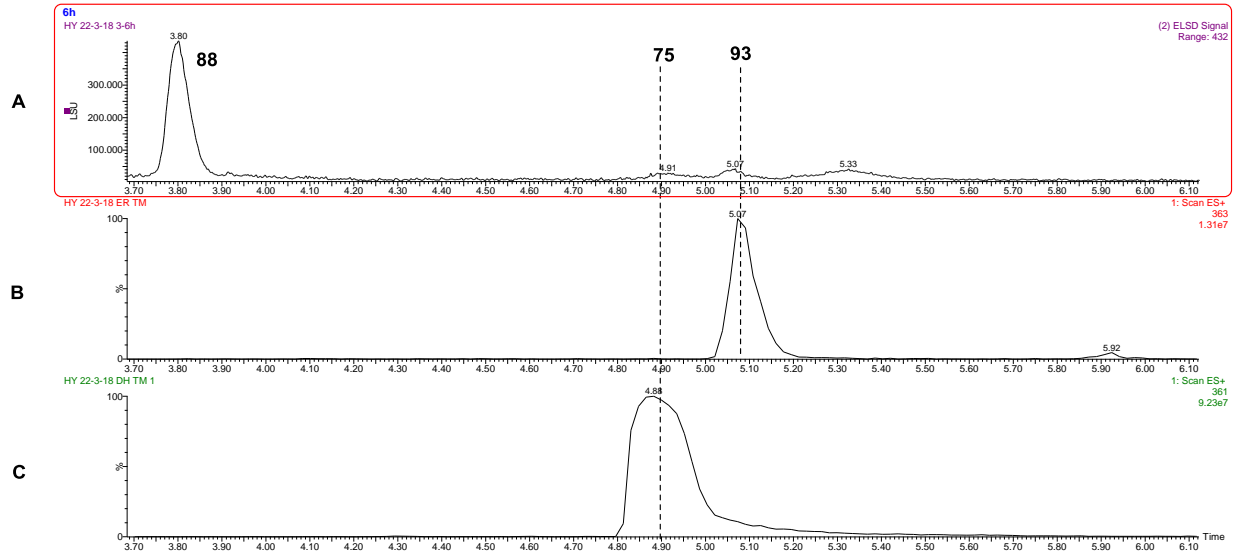


Figure 4.25. Expansion of the red box in Figure 4.24 A - very little dehydrated product can be observed:

A, ELSD signal of reaction; B and C are EIC signal of 93 and 75, respectively.

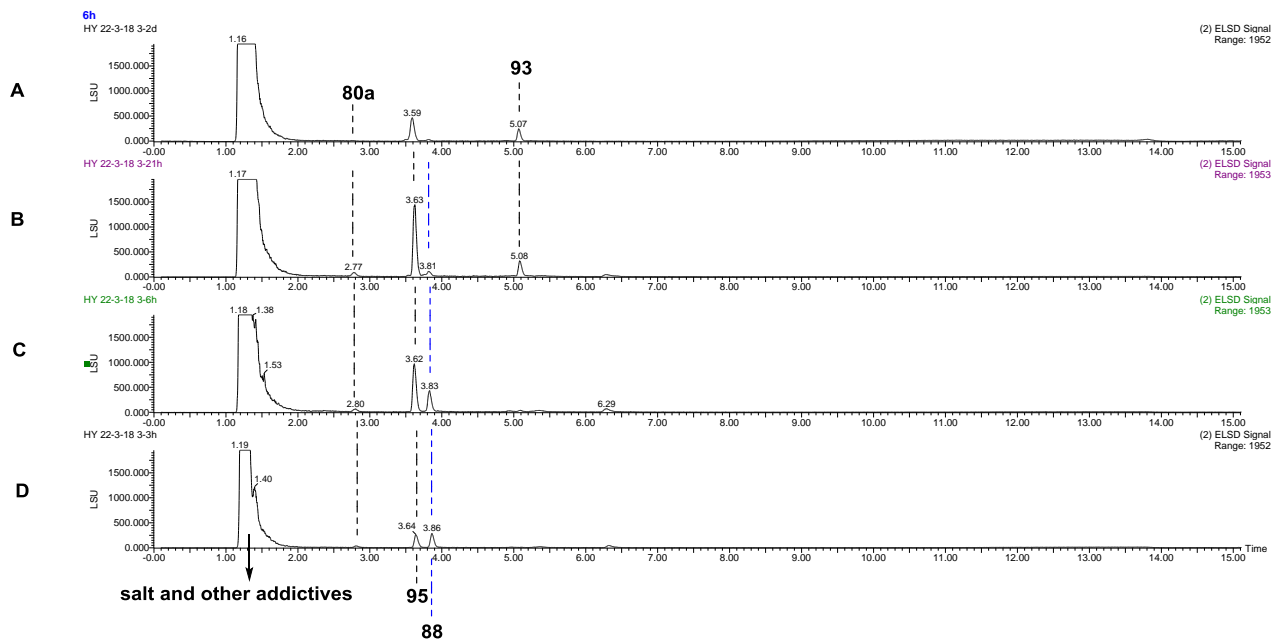


Figure 4.26. LCMS analysis of the process of overall β -processing modification by SQTCS DH-KR

tetradomain: A, after 3 h; B, after 6 h; C, after 21 h; D, after 48 h.

In all, from Figure 4.26, the products of keto reduction, dehydration and enoyl reduction can be observed one by one after 6 hours of incubation. After that, the starting substrate **88** (3.8 min) gradually disappeared and the final enoyl reduction product **93** (5.1 min) formed considerably. However, the reaction was not complete after two days with plenty of keto reduction product **95** (3.6 min) left. Interestingly, very little dehydration product **75** (4.9 min) can be found all the time, which indicates the ER domain of DH-KR can catalyze the enoyl reduction with a low K_M value. The remaining KR product **95** may indicate relatively low activity of the DH domain, possibly because of inhibition, or poor activity of pantetheine substrates compared with the native ACP-bound substrates.

4.8 Studies of SQTKS CMeT Monodomain

Before the β -processing cycle catalyzed by SQTKS, α -methylation can happen first. In the view of kinetic competition,^[22] when methyltransfer is slower than keto reduction and enoyl reduction is slower than hydrolysis, the biosynthesis ends. In the first two cycles of modification catalyzed by the full SQTKS, CMeT works well and so do all the other modification domains. While at the third cycle, CMeT appears to stop working well and passes the non-methylated substrate to the KR. After the keto reduction and dehydration, the biosynthesis stops and the product is released without being reduced by the ER (section 1.5). From this consequence, CMeT works like a controller of the programming process. For this reason, understanding the behavior of the methyltransferase is important. If considering the competitive mechanism, the α -methylation should be faster than keto reduction in the first cycle of modification stage (Figure 1.19). The reaction shown in Figure 4.27 is the analogous methylation in the first cycle of biosynthesis of squalestatin tetraketide, which will be mainly discussed in this section.

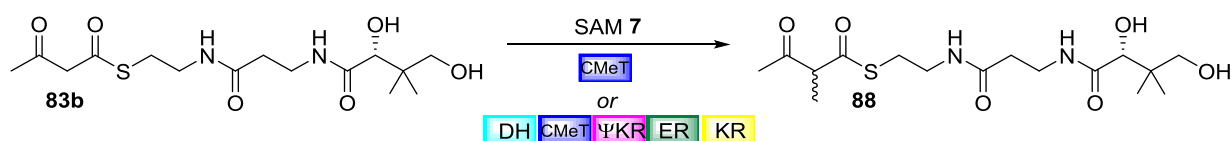


Figure 4.27. Methylation by single CMeT monodomain or DH-KR tetradomain.

4.8.1 Methyltransfer Observed by LCMS

S-Adenosyl methionine **7** (SAM) is used in the methylation reaction as the methyl donor and it transforms to SAH **97** afterwards. The main unstable factor of SAM comes from intramolecular decomposition which produces 5'-methylthioadenosine **96** (MTA) as the side product (Figure 4.28).

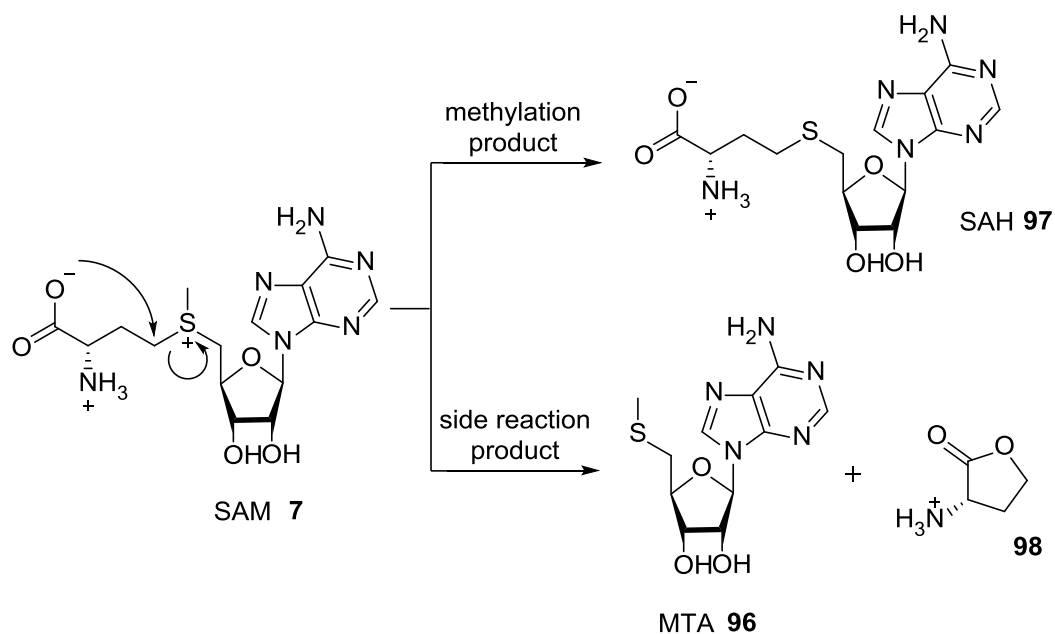


Figure 4.28. Main decomposition reaction of SAM **7** produces MTA **96**.

To check the stability of SAM **7**, it was incubated in water at 25 °C for two days. The result (Figure 4.29) shows SAM is quite stable in water, almost no MTA **96** was observed. It is predictable that if incubating SAM **7** in the reaction buffer, the decomposition will accelerate. However, the ELSD signal of salt appears at the same place where SAM **7** is located and this makes the determination of the stability of SAM in reaction buffer difficult.

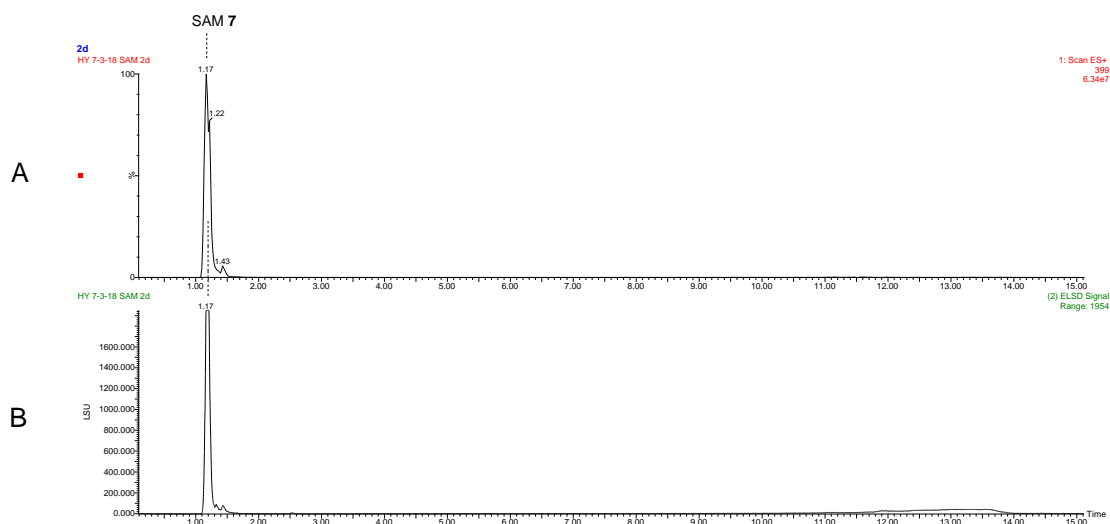


Figure 4.29. LCMS analysis of the stability of SAM 7 in water after 2 d: **A**, EIC signal, **B**, ELSD signal.

Condition: 2.5 mM SAM 7 in H₂O at 25 °C.

4.8.1.1 PANT Substrate –SQTCS CMeT Monodomain and DH-KR Tetradomain

As a standard *in vitro* assay to investigate the methylation, 2 mM acetoacetyl pantetheine **83b**, 2.5 mM SAM 7, and 25 μM isolated CMeT (90 % purity) or DH-KR (50 % purity) after purification optimization (after Ni-NTA) (section 3.3.3.2) was added into 0.2 mL of 50 mM sodium phosphate buffer (pH 7.4, 150 mM of NaCl), the reaction was incubated at 25 °C.

With the CMeT monodomain, considerable methylated product **88** (3.8 min) was observed after 6 hours' incubation (Figure 4.30 C) and the amount of product increased after 2 days (Figure 4.30 B). Hydrolysis of substrate (pantetheine **80a** and pantethine **80b**) and intramolecular decomposition of SAM (MTA **96**) were also observed, as expected under these conditions. However, the reaction cannot complete, which indicates the activity of CMeT is still less than satisfactory even after purification optimization.

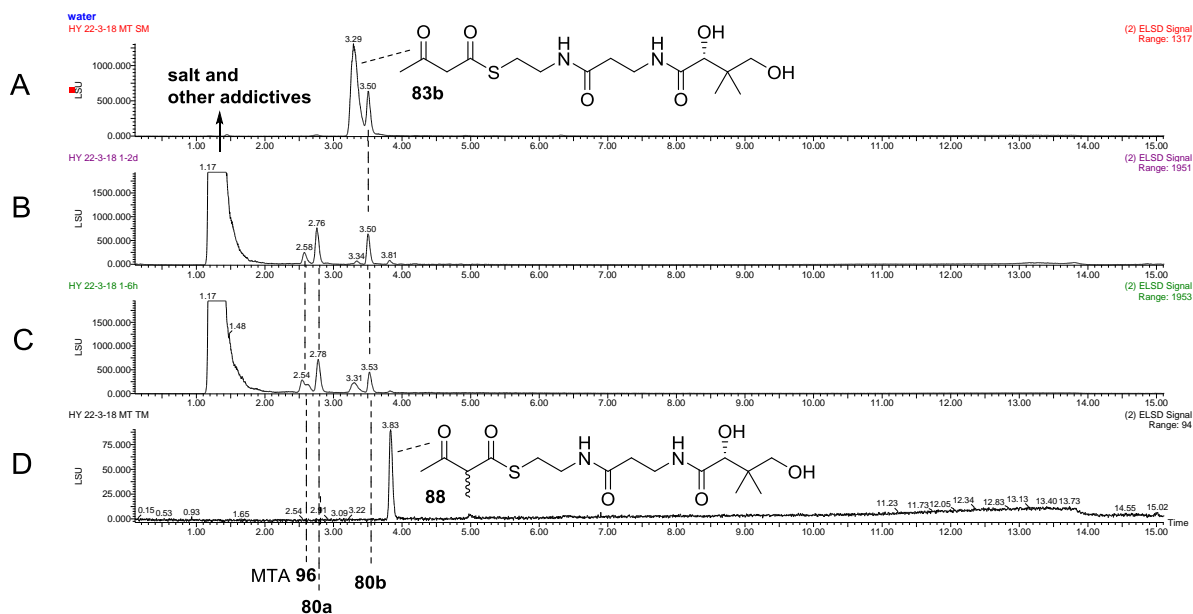


Figure 4.30. LCMS analysis of methylation by SQTCS CMeT with pantetheine substrate: **A**, substrate **83b**; **B**, reaction of methylation after 2 d; **C**, reaction of methylation after 6 h; **D**, the methylated product **88**. Condition: 2 mM substrate **83b**, 2.5 mM SAM **7**, 25 μ M CMeT in 0.2 mL of reaction buffer at 25 $^{\circ}$ C.

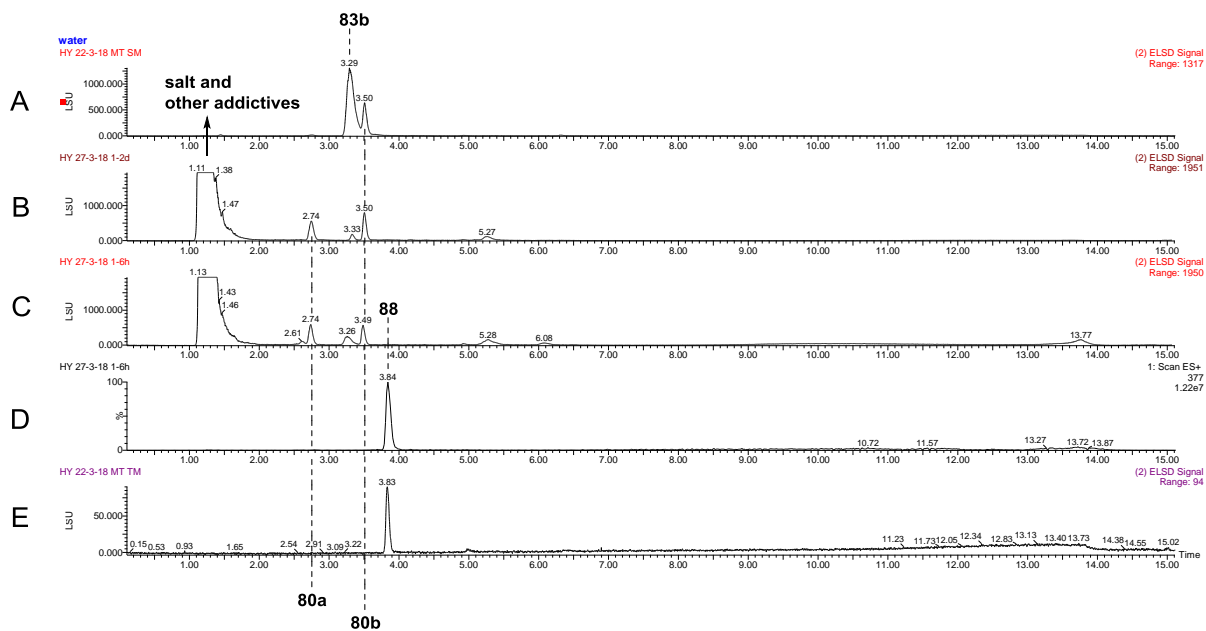


Figure 4.31. LCMS analysis of methylation by SQTCS DH-KR tetradomain with pantetheine substrate: **A**, substrate **83b**; **B**, reaction of methylation after 2 d; **C**, reaction of methylation after 6 h; **D**, EIC signal of trace methylated product **88** from **C**; **E**, the peak of **88**. Condition: 2 mM substrate **83a**, 2.5 mM SAM **7**, 25 μ M DH-KR in 0.2 mL of reaction buffer at 25 $^{\circ}$ C.

With DH-KR, only a trace of methylated product **88** (3.8 min) can be observed under the same conditions after 6 h (Figure 4.31 C) and no increase after 2 d (Figure 4.31 B), which indicates DH-KR tetradomain is less active than CMeT monodomain towards methylation. DH-KR activity site of methylation might be less exposed than that of CMeT.

4.8.1.2 NAC Substrate - SQTCS CMeT Monodomain and DH-KR Tetradomain

With 2 mM acetoacetyl SNAC **83a**, the previous standard *in vitro* assay was used (4.8.1.1). As a result, no methylated product can be observed, either by the isolated CMeT monodomain or by the DH-KR tetradomain (Figure 4.32). The observed large amount of MTA reveals SAM is less stable in buffer than in water.

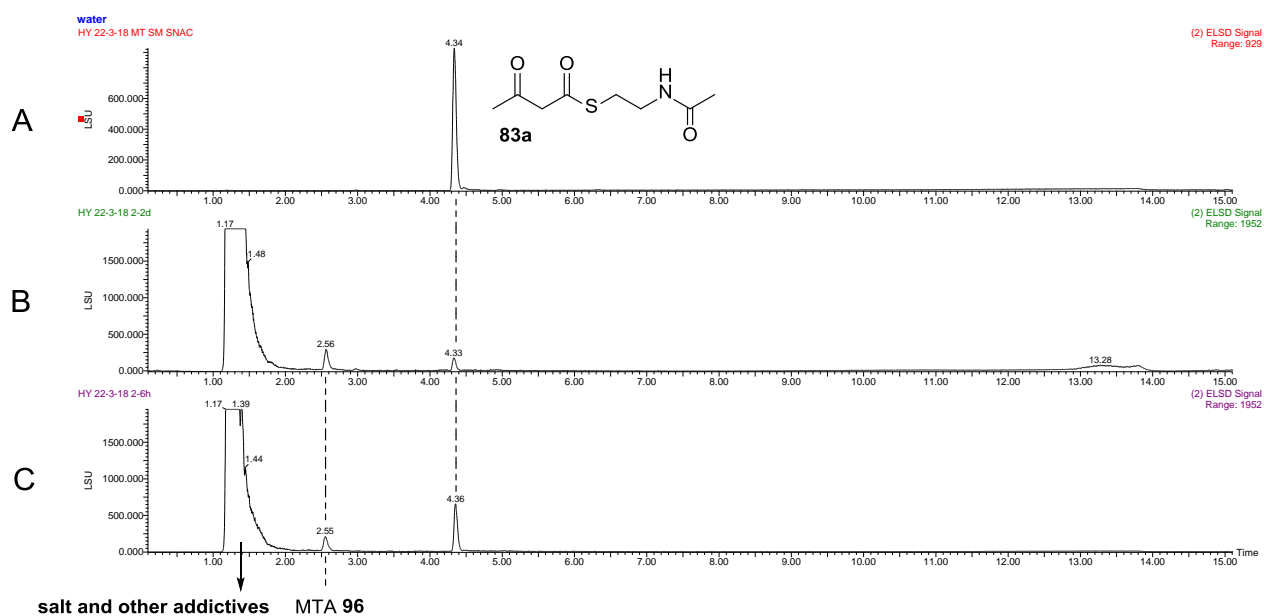


Figure 4.32. LCMS analysis of methylation by SQTCS CMeT with SNAC substrate: **A**, substrate **47a**; **B**, reaction of methylation after 2 d; **C**, reaction of methylation after 6 h. Condition: 2 mM substrate **47a**, 2.5 mM SAM **7**, 25 μ M CMeT in 0.2 mL of reaction buffer at 25 $^{\circ}$ C.

4.8.1.3 5'-methylthioadenosine/S-adenosyl-L-homocysteine Nucleosidase (MTAN)

S-adenosyl-L-homocysteine (SAH **97**), the product of methylation from SAM **7**, is an effective feedback inhibitor of various SAM-dependent methylation reactions.^[159] As the chemical structures of SAM and SAH are nearly identical, SAH can bind to the active site of methyltransferase enzymes and act as a potent competitive inhibitor of those enzymes. 5'-Methylthioadenosine/S-adenosyl-L-homocysteine nucleosidase (MTAN) can catalyze the hydrolysis of MTA **96** and SAH **97**. To check if addition of MTAN can improve the methylation by removing SAH from reaction buffer, the protein has been isolated according to the reported literature.^[160] The *E. coli* MTAN inserted into pET28a was expressed in *E. coli* BL21 (DE3) at 20 degree for 24 h in LB medium with 1 mM IPTG. After purification by Ni-NTA, pure and large-scale protein was produced (Figure 4.33).

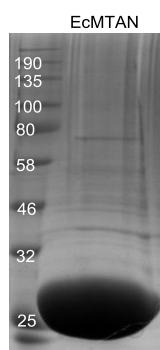


Figure 4.33. Isolation of EcMTAN (26.5 kDa). Storage buffer: 50 mM sodium phosphate buffer, 150 mM NaCl, pH 7.5.

However, in the standard reaction buffer, different proportions of MTAN to DH-KR, 1:4, 1:2 and 1:1 (Figure 4.34 C, B and A) were added, while still no improvement of methylation has been observed by LCMS.

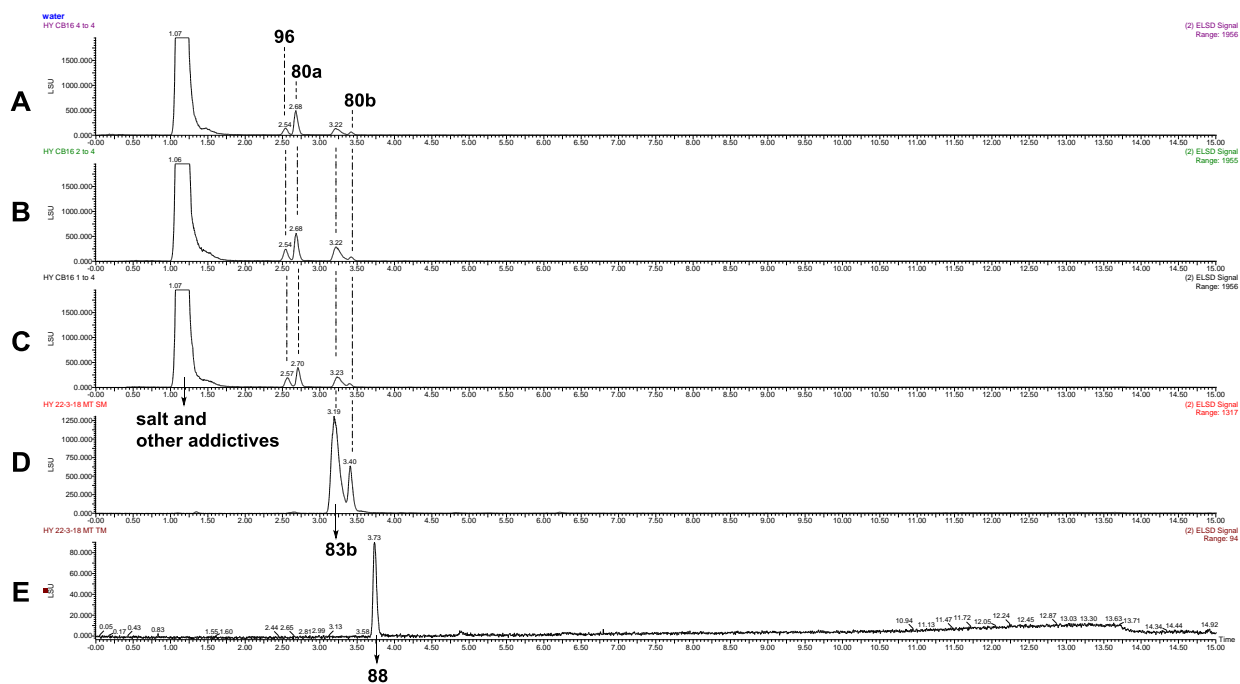


Figure 4.34. LCMS analysis of methylation by SQTKS DH-KR tetradomain with MTAN after 2 days: **A**, DH-KR:MTAN= 1:1; **B**, DH-KR:MTAN= 1:2; **C**, DH-KR:MTAN= 1:4; **D**, substrate **83b**; **E**, promising product **88**. Condition: 2 mM substrate **83a**, 2.5 mM SAM **7**, 25 μ M DH-KR in reaction buffer at 25 $^{\circ}$ C.

4.8.2 Isothermal Titration Calorimetry (ITC) of Cofactor Binding

The level of activity of CMeT in the DH-KR tetradomain protein is much lower than that of the single CMeT, this is unusual as it has been shown that the enoyl reduction activity of DH-KR and single ER is similar (see 4.4.2). One possible reason is that the purification of DH-KR is more difficult than that of CMeT monodomain, which makes the active site of CMeT in DH-KR tetradomain less exposed than that of CMeT monodomain (see 3.3.3.2). As a way to investigate, isothermal titration calorimetry (ITC) was used to check the binding between the cofactor SAM and the two different protein constructs.

ITC is a physical technique used to determine the thermodynamic parameters of interactions in solution. The ITC instrument contains two cells surrounded by an adiabatic jacket: one cell is for sample (in buffer); and the other one is for reference (water). Sensitive

thermocouple circuits are used to detect temperature differences between the two cells. During the experiment, ligand is titrated into the sample cell in precisely known aliquots. As the first injection is made, all injected ligand is bound to the target macromolecule. The binding enthalpy makes the temperature of the sample cell and reference cell different and causes a peak in the signal. The microcalorimeter compensates for the small heat change which causes the signal returns to baseline before the next injection. As a second injection is made, again all injected ligand becomes bound to the target. Signal again returns to baseline before the next injection. As the injections continue, the target becomes saturated with ligand, so less binding occurs and the heat change starts to decrease. When the macromolecule is saturated with ligand, no more binding occurs, and only heat of dilution is observed.

The area of each peak is then integrated and plotted versus the molar ratio of ligand to protein. The resulting isotherm can be fitted to a binding model from which the affinity (K_D) can be determined. At the same time, reaction stoichiometry and enthalpy (ΔH , amount of heat released per Mole of ligand bound) are also derived.

As the ITC is very sensitive, these thermodynamic data are only valuable when pure and stable proteins are used. However, both DH-KR tetradomain (40 % purity after Ni-NTA) and CMeT monodomain (40 % purity after Ni-NTA) used in this assay were of fairly poor purity that had not been optimized (section 3.3.3.2). Therefore, in this ITC assay, we only observe qualitatively whether binding between SAM and protein happens, and compare the relative binding effect towards SAM between the two different constructs - CMeT monodomain and DH-KR tetradomain.

With 2 μL of ligand for each injection and 22 injections with intervals every 300 s at 20 $^\circ\text{C}$, the green signal (Figure 4.35) is the control by injecting 50 μL of 2 mM SAM into 300 μL of the HEPES buffer, the red signal is sample by injecting 50 μL of 2 mM SAM into 300 μL of 0.07 mM Protein in HEPES buffer with protein inside (Figure 4.35). Subtraction between the values of the two signals gives the real signal of binding between SAM and protein. As can be seen, dissolution of SAM in buffer is seriously exothermic (Figure 4.35 green, as the control) which considerably affects the sample signal (Figure 4.35 red, as the

sample). This is because the structure of SAM is similar to ATP and the dissolution of ATP is a great exothermic reaction.

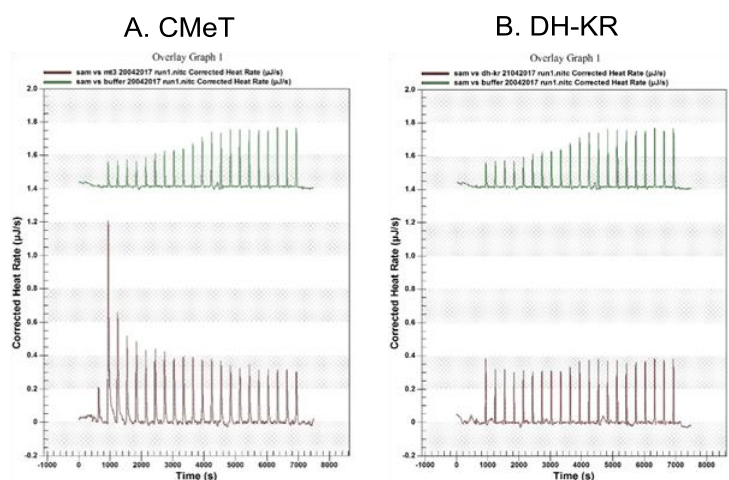


Figure 4.35. ITC investigation of binding between SAM and protein: **A**, CMeT monodomain (red); **B**, DH-KR tetradomain (red). Control: binding between SAM and buffer (green). Condition: 50 μ L of 2 mM Ligand (SAM), 2 μ L of ligand for each inject, 22 injects with intervals every 300 s, 300 μ L of 0.07 mM Protein, Buffer: 50 mM HEPES, 150 mM NaCl, 20 % Glycerol, 1 mM EDTA, T: 20 $^{\circ}$ C.

By comparison, from the height of first several peaks, it is clear that the binding between SAM and the CMeT monodomain (Figure 4.35 A) is much higher than that between SAM and the DH-KR tetradomain (Figure 4.35 B). The binding between SAM and DH-KR tetradomain is very weak, which supports the suggestion that the active site of CMeT in the DH-KR protein is less exposed than that of CMeT monodomain. Large amount of protein is needed in ITC test, therefore only poor purity protein was used in this assay at present. However, it is expected that better data of cofactor-binding and data of substrate-binding can be obtained by using the purification optimized SQTCS DH-KR and CMeT.

4.8.3 Rate of Epimerization

To understand the stereoselectivity of CMeT, one obstacle is whether the 2-methyl product **88** maintains its stereochemistry at its α position long enough to be measured. As chemical epimerization (Figure 4.36 B) may occur due to the low pK_a of the α -proton, this will bring

difficulty in determination of the chirality at the α position of the product **88**. According to the proposed mechanism of methyltransfer^[79] and the gene alignment (section 2.2), the catalytic dyad (His1582 and Glu1608, SQTKS numbering) can interact with the substrate **83b** and catalyze the methyltransfer (Figure 4.36 A). In addition, this catalytic dyad may also have the ability to accelerate the epimerization of the methylated product **88** (Figure 4.36 C). In this way the CMeT could catalyze α -H exchange as well as methyltransfer. Compared to the chemical epimerization, if it does happen and thereby accelerating the epimerization (enzyme epimerization), then we could show the interaction between the product **88** and CMeT. However, if it does not happen and the chemical epimerization of **88** is also slow, then this interaction does not exist and we might have a chance to observe the stereoselectivity of the methyltransfer. To test whether the enzyme can catalyze epimerization, real-time ¹H NMR was used to monitor the rate of epimerization of **88** with, and without, CMeT monodomain.

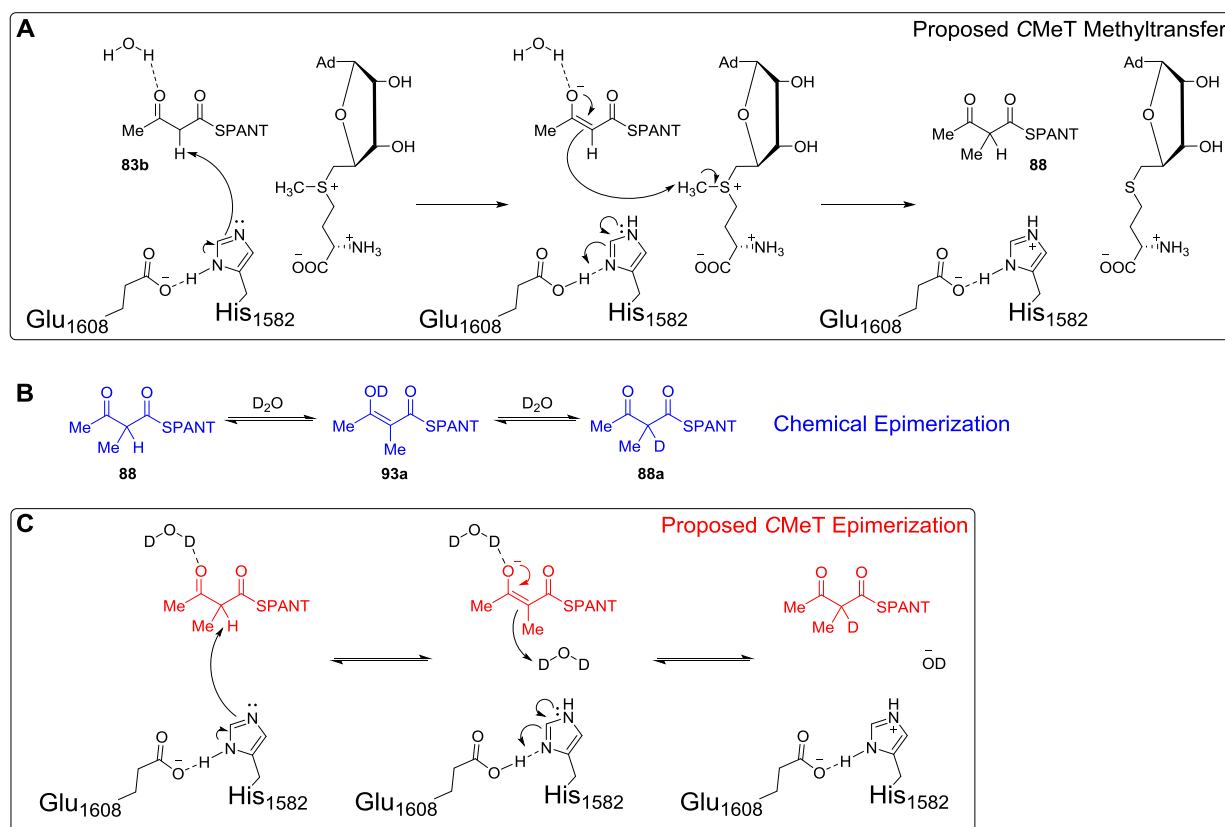


Figure 4.36. **A**, the proposed methyltransfer catalyzed by SQTKS CMeT; **B**, the natural chemical epimerization of **88**; **C**, the proposed epimerization of **88** catalyzed by SQTKS CMeT.

We dissolved 2 mg of 2-methyl acetoacetyl panthetheine **88** in deuterium oxide with 20 μL of phosphate buffer (pH 7.4), put the solution into an NMR tube and obtained ^1H NMR data in every 10 min over 100 min. If the α -position is deuterated, its ^1H NMR resonance will disappear and the signal of the α -methyl group will become a singlet. The phenomenon was observed and the peak changed gradually over time (Figure 4.37). After 1 hour, the α -proton became fully deuterated. The assay was then repeated in the presence of CMeT monodomain in an attempt to observe any change in rate of epimerization.

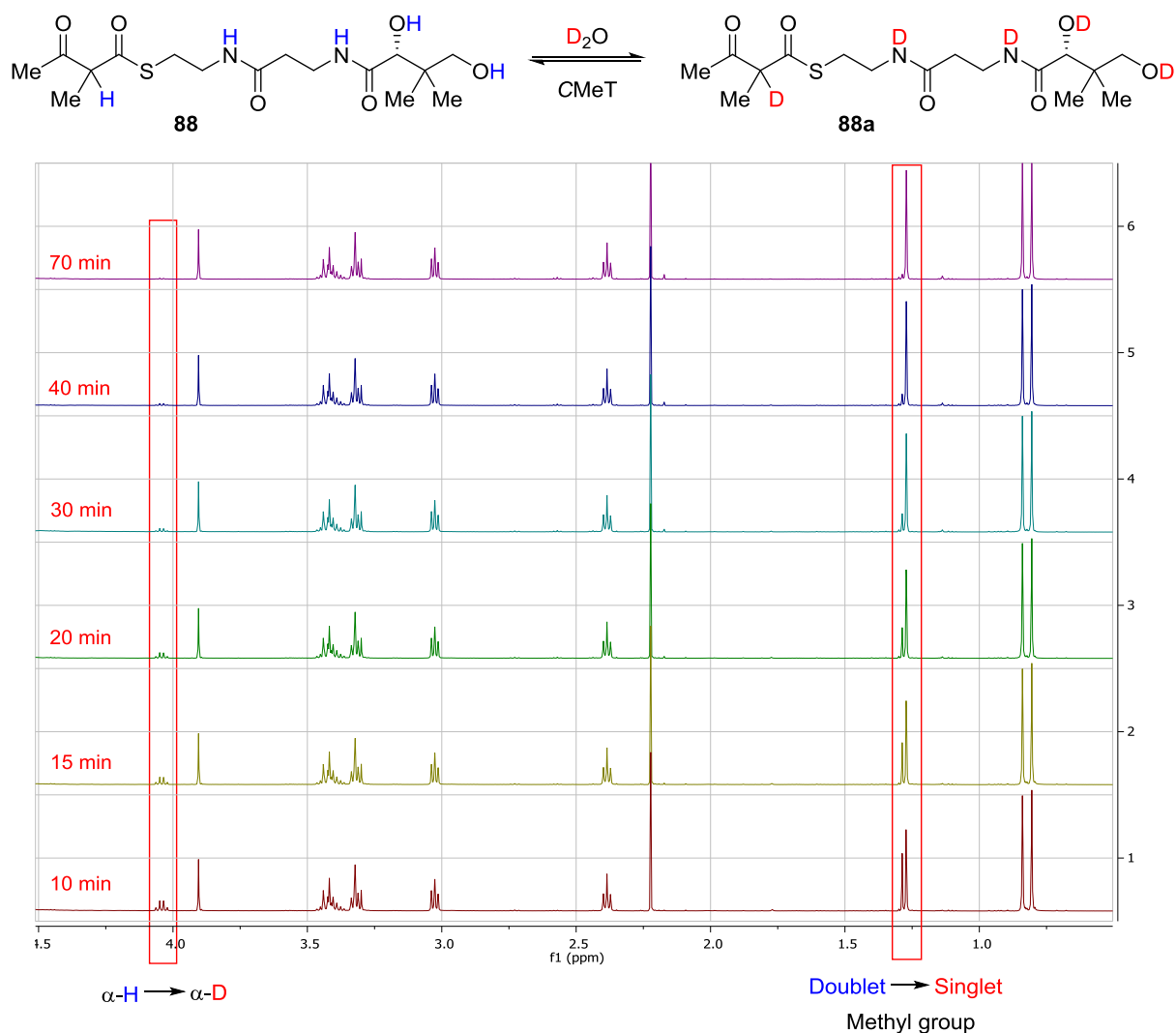


Figure 4.37. Real-time monitoring epimerization of **88** by ^1H NMR. Condition: add 2 mg substrate and 20 μL of PBS buffer with CMeT (50 mM phosphate salt, 150 mM NaCl, pH 7.4) into 0.5 mL D_2O at 20 $^\circ\text{C}$.

We integrated the proportion of α -proton in nmr spectrum and express its relationship with time (Figure 4.38). The control (without CMeT) is shown in blue. Addition of 0.4 μ M CMeT makes the curve kink which may due to some random errors. While from the initial rate, the rate of epimerization is similar to that of control. Addition of 4 μ M CMeT obtained a nicer curve, however, the initial rate of epimerization had not increased tenfold as the concentration of CMeT increased. Therefore, the enzyme epimerization may not be significant and the interaction between **88** and CMeT may not exist. This may be the reason why the methylation activity is less than satisfactory in previous CMeT *in vitro* assay (section 4.8.1). As the compound **88** is not the original substrate **83b** for the methylation reaction and these pantetheine substrates are not attached to the original ACP protein, this real interaction between the original ACP-bound substrate and CMeT is not clear. However, the chemical epimerization was found to be slow with a half-time life ($t_{1/2}$) of pantetheine substrate **88** around 15 min. The epimerization could be even slower when the pantetheine substrate bounds to the ACP protein, which can decrease the pKa (see section 4.10.2), this phenomenon makes observation of the stereoselectivity of SQTGS CMeT possible.

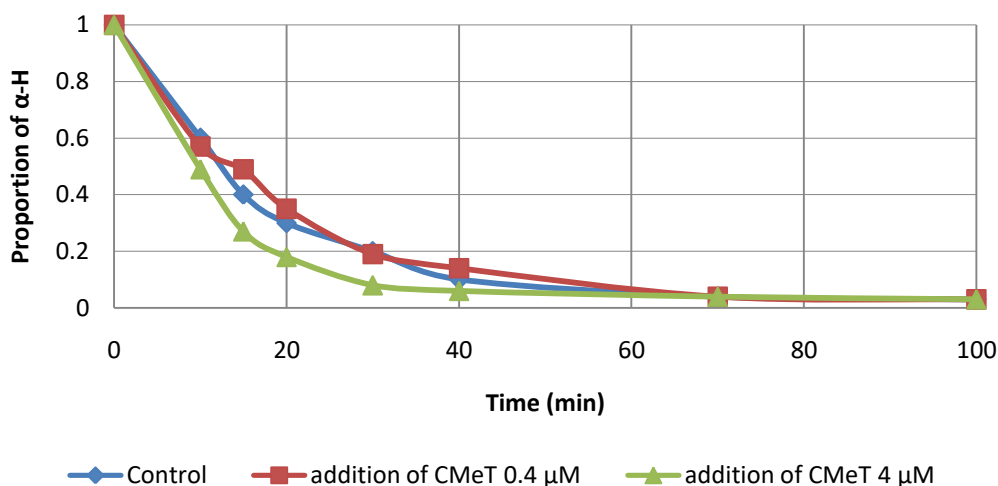


Figure 4.38. Rate of epimerization affected by addition of SQTGS CMeT monodomain. Control: chemical epimerization without CMeT.

4.9 Conclusion

The DH-KR tetradomain and CMeT monodomain proteins are active even though the purities of these proteins are compromised (section 3.3). By measuring the K_M value of different analog substrates for the ER reaction, we found the ER in the DH-KR tetradomain can be used as a general catalyst, while it still shows a programming effect by its inability to reduce the squalestation tetraketide. This result is similar to that of the isolated ER monodomain research previously in our group.^[75] In addition, by using LCMS analysis, we also found all the products of DH, KR and ER. It was not easy to observe the formation of DH product, because the ER can quickly transform it into the final product in very low concentration.

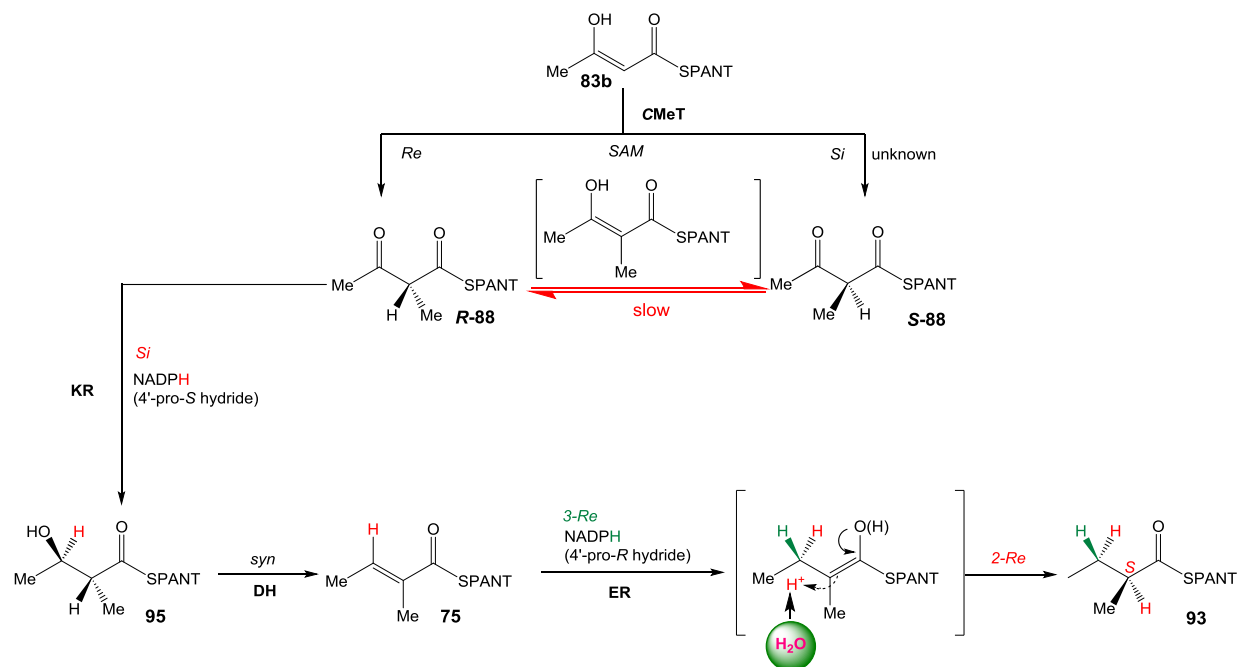


Figure 4.39. *In vitro* assay reveals stereochemistry in the first cycle biosynthesis of squalestatin tetraketide.

We also tested the stereoselectivity of ER, KR and CMeT domains by using simple diketide panthethine substrates (Figure 4.39). As the analog of native product of the first cycle of biosynthesis of squalestatin, *S*-2-methyl butanoyl panthethine **93** is proved to be the stereo-preferred product from enoyl reduction by ER of DH-KR tetradomain. This contrasts to the isolated ER monodomain which gives racemic products. The ER active site in the DH-KR

tetradomain construct may be narrower and less flexible than that of the isolated ER monodomain, which allows fewer water molecules into the active site (most gather on *Re* face) and thereby producing more *S* products.

The hydride donor of the KR reaction is 4'-*pro-S* proton of NADPH. The activity of the isolated CMeT monodomain is compromised and its stereoselectivity remains currently unknown by using pantetheine substrate. So far, all the stereopreferences which have been found in the KR, DH and ER of SQTKS are identical between SQTKS and mFAS, which further reinforces the idea that fungal highly reducing PKS and mFAS evolved from a common ancestor.

4.10 Future Work

During our *in vitro* research into the function of the DH-KR tetradomain and the CMeT monodomain of SQTKS, several problems are still unsolved. An explanation for the difficulty of isolating the pure protein, either DH-KR or CMeT, has been proposed (section 3.3.3.2) based on the combination of impurities and the instability of these proteins themselves. Biophysics experiments showed that SQTKS DH-KR exists in solution as a mixture of monomer and dimer. The stability of CMeT is less than DH-KR, which can be attributed to the larger exposed hydrophobic area, but isolated CMeT exhibited higher methyltransfer activity from *in vitro* assay. In comparison, the SQTKS DH-KR tetradomain with much less exposed hydrophobic area based on the structure of mFAS, showed trace methyltransfer activity. However the tetradomain does show measureable KR, DH and ER activity with acyl pantetheine substrates. A possible explanation is that the methylation active site in the isolated DH-KR is still not well exposed compared to CMeT. Impurities difficult to remove may still exist in the DH-KR tetradomain and shield the methylation active site. For the future, we may solve this problem by considering the function of metal ions and using ACP-bound substrates.

4.10.1 To Improve Reactivity of Methyltransfer with Ferric Ion

The poor activity of CMeT monodomain and DH-KR tetradomain towards methyltransfer is difficult to be explained at this moment. It is possible that metal ions are needed for the full functionality of protein. It is reported a ferric ion SAM-dependent C-methyltransferase involved in the biosynthesis of β -methylphenylalanine, a building unit of glycopeptide antibiotic mannopeptimycin.^[161] For the modular polyketide synthase, it has been reported recently an unusual ferric ion dependent methyltransferase that acts in the initiation steps of apratoxin A biosynthesis (AprA MT1). Fe³⁺-containing AprA MT1 catalyzes one or two methyl transfer reactions on the substrate malonyl-ACP **99**, whereas Co²⁺, Fe²⁺, Mn²⁺, and Ni²⁺ support only a single methyl transfer.^[162] It is proposed that the metal facilitates catalysis by acting as a Lewis acid to lower the pK_a of the malonyl or methylmalonyl α -carbon, leading

to deprotonation, enolate formation, and attack of the SAM methyl (Figure 4.40). In this case, even though no metal dependent methyltransferase in iterative polyketide synthase has been reported, trivalent and more multivalent metal ions are worth trying.

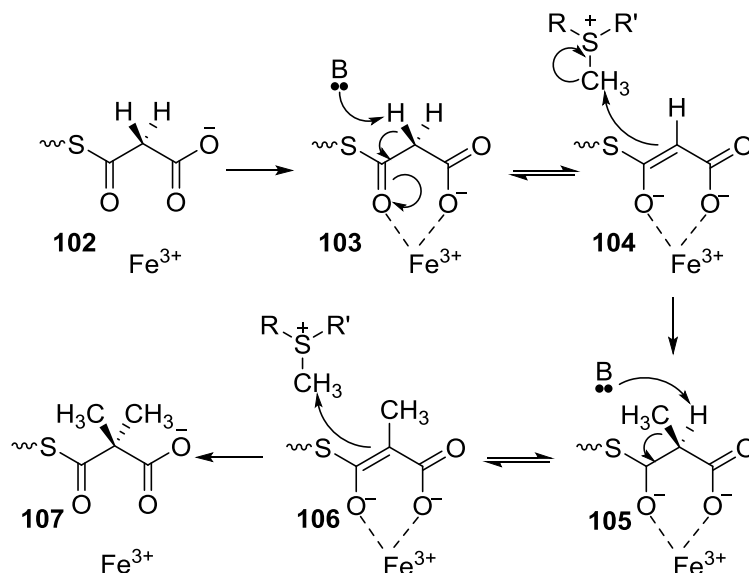


Figure 4.40. Fe³⁺ related AprA MT1 mechanism. R and R' refer to parts of SAM 7.

4.10.2 The Importance of Acyl Carrier Protein (ACP)

Acyl carrier proteins (ACP) are vital to polyketide and fatty acid biosynthesis and enable the movement of acyl polyketide intermediates between the active sites of the functional PKS and FAS domains. ACPs are small (~10 kDa) and often unusually acidic. From the crystal structure of mammalian FAS, we cannot see the structure of acyl carrier protein (ACP), which is because the movable ACP brings a problem on its structure analysis. ACP moves around to transfer the substrate it carries from domain to domain. It is unknown how ACP transfers the substrate from the active pocket of a domain to that of another domain. The movement among substrate and domains involves a series of action such as in, out, move to another domain. However, without ACP, by diffusion a pantetheine substrate can still meet the active pocket of each domain and transform itself through keto reduction, dehydration and enoyl reduction in the DH-KR discussed in section 4.6. The question is, what is the role of ACP?

As we have shown that the pantetheine substrate is not very stable in the phosphate buffer (section 4.3), it is possible that ACP has a function of stabilizing the pantetheine substrate. In this way, the hydrolysis and ester exchange will be prevented. Actually, it has already been discussed by David Cane and coworkers the stability function of DEBS ACP towards the substrate.^[163] The stability of the ACP-bound 3-oxo-2-methylacylthioester intermediates is remarkable, which retain their configurational integrity even after 1h at pH 7.2. Under similar condition, complete exchange of α -proton happens in 1h at pH 7.4 during the epimerization of 2-methyl acetoacetyl panthetheine **88** (see 4.7.3), which is one factor which thwarted our research into the stereopreference of the SQTKS CMeT. It is predictable that this problem might be solved by using ACP-bound substrate. The simplest explanation of the remarkable chemical and configurational stability of the ACP-bound polyketide is that this substrate does not swing freely in solution but is bound by the ACP itself. And, ACP-bound polyketide might prevent the β -ketone and the thioester carbonyl from becoming coplanar by forming a kink in the chain, which will considerably increase the pK_a of the α -proton. In this way, spontaneous lactonization, epimerization, ester exchange and other intramolecular reactions are suppressed. It is reported that *E. coli* ACP contains a cavity in which acyl chains 6, 7 or 10 carbon atoms in length can be observed interacting with hydrophobic amino acid residues, which makes this carbon atoms well shielded from the buffer.^[164] The crystal structure of *Saccharomyces cerevisiae* FAS shows the ACP-bound substrate well buried in a groove on the ACP with only the reactive portion of the substrate exposed.^[157]

In addition, ACP-bound substrate might also have other advantages. It has been reported recently^[165] that ACP-bound tigloyl substrate proves the stereospecificity of several modular ER domains by obtaining single enantiomer while the pantetheine substrate produces a racemic mixture. Therefore, the dependence of ER stereospecificity involves protein-protein interactions between ER and ACP domains. Crystal structures of an AT-ACP complex from the vicenistatin PKS and a curacin β -branching enzyme in complex with ACP indicate that the ACP surface of α II, loop II, and α III interacts with partner enzymes.^[166-167]

However, no crystal structure of ER-ACP is available to further understand the basis of ACP recognition for ER stereochemistry. Consideration of the different stereopreference between DH-KR and single ER domain of SQTKS by using pantetheine substrate, the dependence of ER stereospecificity might be different between modular and iterative polyketide synthases.

5 Summary and Outlook

5.1 Summary

In this thesis, a simple HR-PKS with an active ER has been studied. To instruct the domain isolation, a structure model of SQTGS was built based on the reported crystal structure of mFAS because of the considerable similarities between HR-PKS and mFAS. The determinations of domain boundaries are based on protein alignments and structural comparison with other similar PKS (chapter 2). Several multidomains, including DH-ER tridomain (~135 kDa), DH-KR tetradomain (~166 kDa), DH-ACP pentadomain (~176 kDa) and CMeT monodomain (~48 kDa), were over-expressed in *E. coli* which has been successfully used in domain isolation of many PKS. With rational gene optimizations, different cloning methods and expression conditions (various strains and vectors), only the DH-KR tetradomain and CMeT monodomain could be isolated as relatively stable soluble proteins, possibly due to fewer hydrophobic areas according to the modelled SQTGS structure. In addition, stepwise buffer optimization and purification optimization have been investigated as they affect the protein activity. We found that addition of more Mg^{2+} into the cell culture, addition of EDTA into the storage buffer, and addition of DNAase and ATP during the purification process can considerably improve the stability and purity of the DH-KR tetradomain and CMeT monodomain proteins. Biophysics studies of the more stable DH-KR tetradomain have been performed and revealed that the protein forms a balance between monomeric and dimeric species in solution, which brings difficulties for its structure determination by SAXS. The less stable CMeT forms a trimer or tetramer in solution and it tends to aggregate in unfavored environment (see chapter 3).

In chapter 4, detailed *in vitro* assays of these proteins are discussed. As there were difficulties in monitoring ACP-bound substrates, we used pantetheine substrate analogs. Compared to the widely used SNAC substrates, the pantetheine substrates react faster and bring convenience in kinetic studies which was previously shown for the ER domain by the Cox group.^[75] In the SQTGS DH-KR tetradomain, the KR, DH and ER showed observable activity towards the pantetheine substrate in LCMS. Among them, the kinetic studies of enoyl

reduction show that the ER of DH-KR tetradomain is capable of transforming various substrates and it controls the programming function of SQTCS by its inability to reduce the final tetraketide substrate. These results are in accord with the kinetic studies of the single SQTCS ER. However the CMeT domain showed very low activity with all substrates tested.

In regard of the stereochemistry, the isolated ER monodomain^[75] of SQTCS can control the stereoselectivity of β -hydride addition to the enone **26**, but it cannot control the stereoselectivity of α -protonation. However, the ER in the DH-KR tetradomain of SQTCS was shown here to control the α -protonation using NMR studies. As there is no reasonable protein residue that can act as the proton donor according to the modeling structure of the SQTCS ER,^[75] a water molecule is suggested to donate a proton only from the 3-*Re* face of **26** to create the (2*S*)-methylated fully reduced polyketide **27** (Figure 5.1). Compared to the ER monodomain, the ER in the DH-KR tetradomain may be narrower and less flexible, which allows less water into the active site which could achieve more rigid stereochemical control during α -protonation.

The hydride responsible for the KR reduction was shown here to be the 4'-*pro-S* hydride of the NADPH by LCMS studies. Unfortunately, due to the low activity of the SQTCS CMeT monodomain (only trace of methyltransfer activity of the CMeT in the DH-KR tetradomain was observed, which may due to the left impurities in the active site of its CMeT) towards a panthetheine substrate, the stereoselectivity of the CMeT could not be studied. However, through the NMR studies, we found that the rate of spontaneous epimerization at the α -methylated acetoacetyl pantetheine **88** is slow (complete epimerization happens in one hour in pH 7.4 phosphate buffer at 20 °C). As the ACP has the ability to stabilize the pantetheine substrate (section 4.10.2), it is predictable that the epimerization of an ACP-bound substrate should be very slow. Previous SQTCS DH studies by our group^[57] infer that the KR only selects (2*R*, 3*R*)-**25** as its substrate. In consideration of the low epimerization rate, and if the SQTCS KR and DH have no epimerization activity which has been found in some KR and DHs of modular PKS domains (see section 1.4.3 and 1.4.5), the SQTCS CMeT may have the 2*R* stereoselectivity for the product **24**. This 2*R* stereoselectivity has been found (see section

1.4.4) in a modular PKS (BonMT2)^[53] and predicted in an HR-PKS (Geroratusin PKS)^[59] (Figure 1.15). However, more direct proofs should be given in the future by *in vitro* studies to support this suggestion.

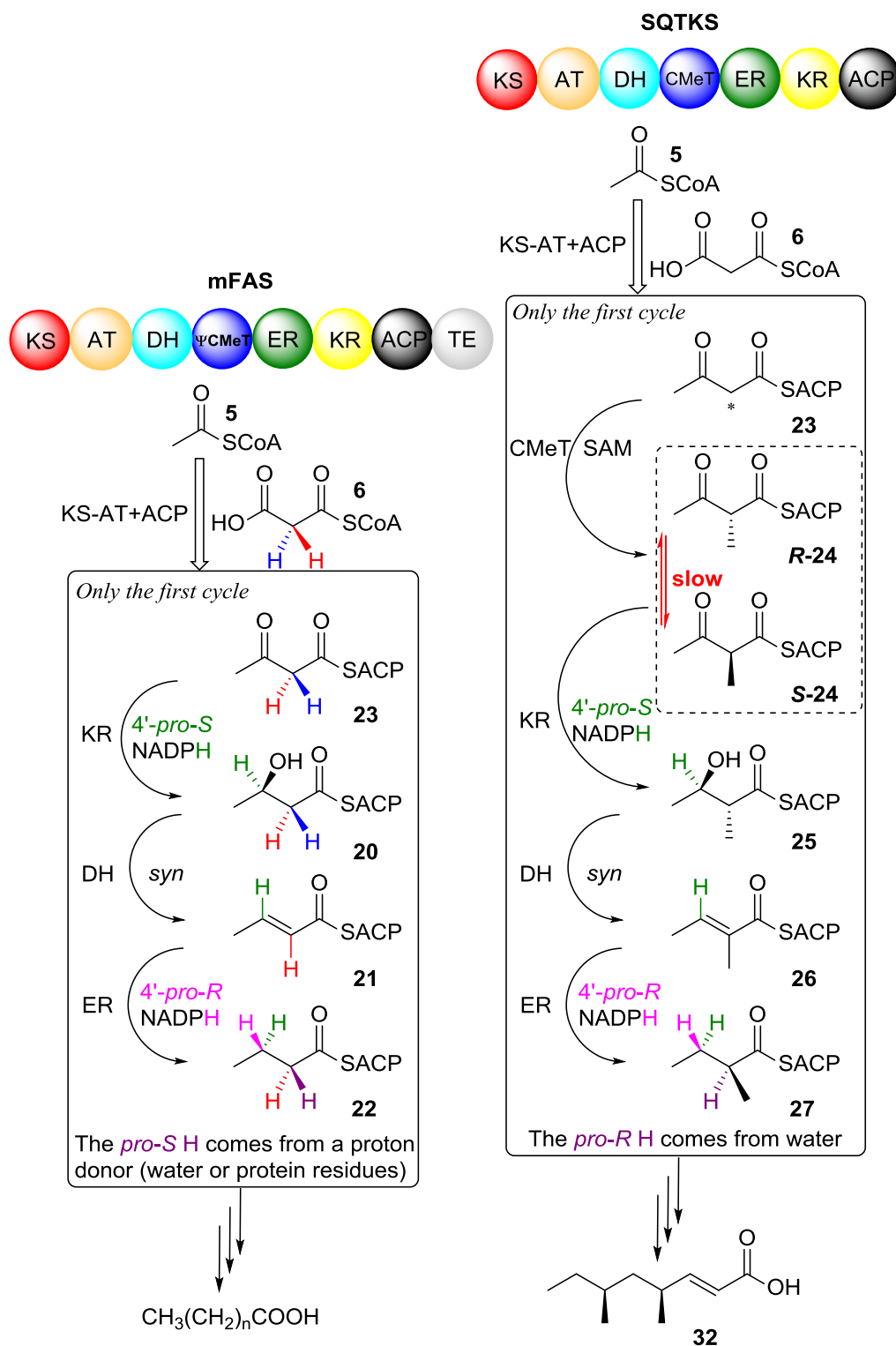


Figure 5.1. The comparison between the mFAS and the SQTKS in stereochemistry.

Comparison of the stereoselectivities of mFAS and SQTks shows they are identical (Figure 5.1), which reinforces the idea that the HR-PKS and mFAS evolved from a common ancestor. Specifically, the KR and DH domain of SQTks show identical stereopreferences compared to those of mFAS, those are: the KR uses the 4'-*pro-S* hydride of NADPH to attack the 3-*Si* face of **24**; then, *syn*-elimination of **25** is catalyzed by DH and the enone product **26** of *E*-configuration forms. The ER domain of SQTks also shows the same stereopreferences of hydride donor (4'-*pro-R* hydride of NADPH) and β -addition (the hydride attacks the 3-*Re* face of **26**) compared to that of mFAS, although it shows different stereoselectivity in C2 protonation: a proton donor attacks from the 3-*Re* face of **26** in SQTks ER while from the 3-*Si* face of **21** in mFAS ER. The inversion of the α stereochemistry happens in KS domain of mFAS, but this has not yet been verified in SQTks.

Evidence in this thesis has shown that pantetheine substrates are unstable if incubated for a long time in common buffers, such as phosphate buffer and tris buffer, due to the possible hydrolysis and inter/intra ester exchange observed by LCMS (section 4.3). In addition, the binding of the pantetheine substrate to the active site of CMeT has proven to be weak by an indirect conclusion from the epimerization experiment (section 4.8.3) and the ITC analysis (section 4.8.2). Therefore, the behavior of the pantetheine substrates is compromised, which can affect the catalytic reaction and the judgement of the protein activity.

The importance of the ACP in PKS has been discussed in section 4.10.2. The ACP is capable of stabilizing its bound pantetheine substrate during its interactions with the modifying domains. It also affects the stereoselectivities of some modular ERs (section 4.10.2) and some modular DHs (section 1.4.5). Therefore, incorporation of the ACP to form the original SQTks substrates should be explored in future research.

5.2 To Investigate the Remaining Stereochemical Questions in SQTKS

According to the discussion in section 1.4.3 and 1.4.5, the “equilibrium isotope exchange (EIX)” experiment^[54] proves that KR and DH in some modular PKS may exhibit activity of α -epimerization of 2-methyl-3-ketoacyl substrates. In SQTKS, we could also use the EIX experiment to check whether epimerization happens in SQTKS DH and KR.

As a plan for the future work, the stereo-configurationally stable KR substrate with C-2 deuterium label could be synthesized (*i.e.*, [2-²H]-2-methyl-3-hydroxypentanoate),^[54] and the substrate could be linked to CoA and transferred to a *holo*-ACP with the Sfp (4'-phosphopantetheinyl transferase), afterwards to form the deuterated SQTKS ACP-bound substrates **25A1a** and **25B2a** (Figure 5.2). The DEBS KR-6 does not have the epimerization activity while a tiny balance may still exist between compound **25A1a** and *R*-**24a** (section 1.4.3) by incubating **25A1a** with NADP⁺. By adding the previous isolated SQTKS DH^[57] and DEBS KR-6, if the first-order time-dependent washout of the deuterium occurs (Figure 5.2 A), as the DEBS KR-6 does not epimerize, then the SQTKS DH has more epimerization activity. If no washout of the deuterium can be observed, the SQTKS DH has also no epimerization activity.

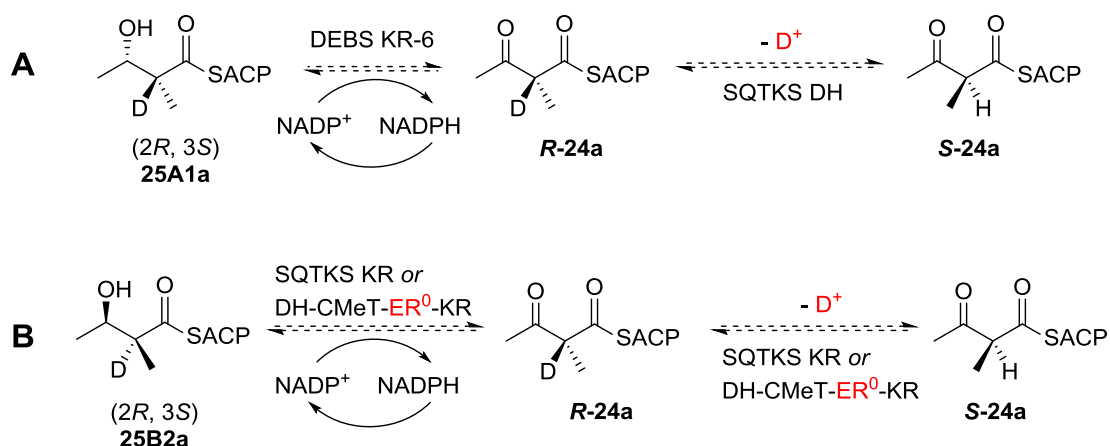


Figure 5.2. The proposed “equilibrium isotope exchange (EIX)” experiments of SQTKS DH and KR, which are based on assays in section 1.4.3 and 1.4.5.

For the KR in SQTks, the single SQTks KR (to be isolated) or DH-KR with mutated-nonfunctional ER could be incubated with **25B2a** and NADP⁺ to see if any *R*-**24a** forms and so that first-order time-dependent washout of the deuterium occurs (Figure 5.2 B). In this way, we could tell whether the SQTks KR has the epimerization activity.

The SQTks CMeT may have the *2R* stereoselectivity according to the discussion of the section 5.1, but so far no direct proof has been shown. Based on the existence of modular epimerase-active KR and epimerase-inactive KR, ^[54] *in vitro* assays with the combination of CMeT and these KR (such as Figure 1.15 A) could be used to certify the stereoselectivity of the SQTks CMeT. The isolated SQTks CMeT monodomain or DH-KR tetradomain (only when the KR and DH have no epimerization activity) could be used in this assay. According to the assay in section 1.4.4, the SQTks CMeT could be incubated with 3-ketobutanoyl-ACP **2** (generated *in situ* from acetyl-SNAC-incubated BonKS-2 and malonyl BonACP produced by BonACP, malonyl-CoA, Sfp), ^[53] SAM **7** and SAH **97** nucleosidase to produce the methyl product *R*-**24** (Figure 5.3). Afterwards, the product could be quickly reduced by the KR and transformed into the ester **66** which could be confirmed by chiral GC-MS analysis (stereochemistry at C-2 and C-3). ^[58] As a result, if the SQTks CMeT shows the *2R* stereoselectivity, reduction with EryKR-6 (epimerase-inactive, B-type) will exclusively form (*2R*, *3S*)-2-methyl-3-hydroxybutanoate **66** while the EryKR-1 (epimerase-active, A-type) forms the corresponding (*2S*, *3R*)-2-methyl-3-hydroxybutanoate **65**. In addition, the epimerase-inactive and (*2S*)-2-methyl-3-ketoacyl-ACP-specific mutant ketoreductase, DEBS KR-6-G324T/L333H, will not produce any product. If the SQTks CMeT shows the *2S* stereoselectivity, then the DEBS KR-6 will produce no product, the EryKR-1 will produce the (*2R*, *3R*) **66** and the mutated DEBS-KR6-G324T/L333H will produce the (*2R*, *3R*) **66** exclusively. In this way, the stereoselectivity of the SQTks CMeT could be determined.

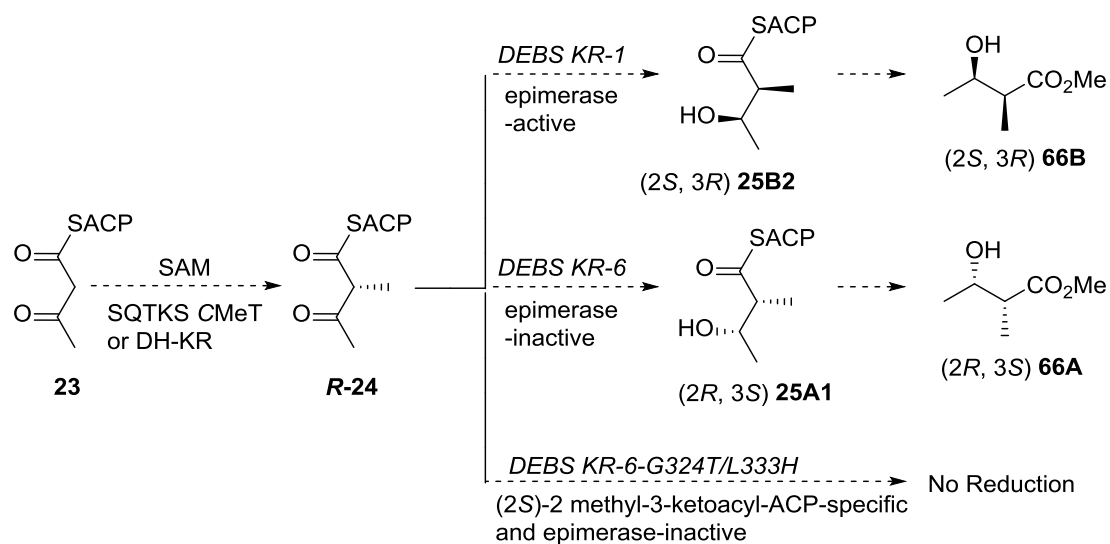


Figure 5.3. The proposed method to determine the stereoselectivity of the SQTCS CMeT, which is based on the assay in section 1.4.4.

In all, the studies in this thesis reveal the main stereopreferences of SQTCS in the first cycle (*RI*) of its catalyzed biosynthesis (Figure 1.19). From the gene alignments and crystal structures of similar PKSs, protein residues in the active pockets can be inferred (Chapter 6). However, the key protein residues which determine the stereoselectivity of each domain are not clear. Therefore, more crystal structures elucidating the binding between the substrate and its specific protein residues, in combination of targeted gene mutation, are needed in the future research. In addition, reprogramming of SQTCS should consider the function of these key stereoselective protein residues to create new polyketides with desired stereochemistries.

6 Experimental

6.1 Materials and General Protocols

Software

Genenius 7.0.6, Curve Expert 2.6.5, Microsoft Excel.

For codon optimization: <http://genomes.urv.es/OPTIMIZER/Form.php>

For viewing codon usage in different strains: <http://www.kazusa.or.jp/codon/>

For viewing the effect of codon optimization: http://gcu.schoedl.de/sequential_v2.html

For multiple sequence alignment: <https://www.ebi.ac.uk/Tools/msa/clustalo/>

For a prediction of the solubility of protein overexpressed in *E.coli*:
<http://scratch.proteomics.ics.uci.edu/index.html>

Media

Deionised water was further purified by a GenPure Pro UV/UF millipore device from Thermo Scientific and used to prepare all media and buffers. Media were sterilized at 120 °C for 15 min using a Systec VX150 or a Classic Prestige Medical 2100 autoclave, or filter sterilized using a 45 µM syringe filter with a cellulose acetate membrane.

Media	Component
Lysogeny broth (LB)	0.5 % (w/v) Yeast extract (Duchefa Biochimie), 1.0 % (w/v) Tryptone (Duchefa Biochimie), 0.5 % (w/v) Sodium chloride (VWR).
LB Agar:	LB 1.5 % (w/v) Agar (Duchefa Biochimie)
2 tryptone yeast medium (2TY)	1.0 % (w/v) Yeast extract (Duchefa Biochimie) 1.6 % (w/v) Tryptone (Duchefa Biochimie) 0.5 % (w/v) Sodium chloride (VWR)
TB	2.4 % (w/v) Yeast extract (Duchefa Biochimie) 1.2 % (w/v) Tryptone (Duchefa Biochimie) 0.4 % (w/v) Glycerol (Roth) Filtered sterile 10 % (v/v) KPI buffer (added after the autoclave)
KPI buffer	2.31 % (w/v) Monopotassium phosphate (Roth) 12.54 % (w/v) Dipotassium phosphate (Roth)

LBE5052	0.50 % (w/v) Yeast extract (Duchefa Biochemie) 1.00 % (w/v) Tryptone (Duchefa Biochemie) 0.50 % (w/v) Glycerol (Roth) 0.05 % (w/v) D(+)-Glucose Monohydrate (Roth) 0.20 % (w/v) Lactose Monohydrate (Roth) 0.07 % (w/v) Sodium sulfate (Honeywell) 0.25 % (w/v) Ammonium chloride (Alfa Aesar) To be added (to 900 mL LBE-5052): 1.0 mL Magnesium sulfate hexahydrate 2M (Sigma Aldrich) 1,0 mL Metals Mix (1000 ×) 100 mL Potassium Phosphate Mix (50 mM Phosphate) seperate autoclaved
Metals Mix 1000 ×	1,62 % (w/v) Iron-(III)-chloride (Fisher Scientific) 50 mL of 0.1 M Solution dissolved in 0.1 M HCl To be added: 1 mL Manganese-(II)-chloride tetrahydrate (Roth) 1M 1 mL Zinc sulfate heptahydrate (Acros) 1 M 1 mL Cobalt-(II)-chloride hexahydrate (Roth) 0.2 M 1 mL Nickel-(II)-chlorid hexahydrate 0.2 M 46 mL pure water Filter sterilize
Potassium Phosphate Mix (50 mM Phosphate)	10 mL Monopotassium phosphate (Roth) 1 M (1,36 g) 40 mL Dipotassium phosphate (Roth) 1 M (6,96 g) 50 mL pure Water
Super optimal broth with magnesium chloride and glucose (SOC)	0.50 % (w/v) Yeast extract (Duchefa Biochemie) 2.00 % (w/v) Tryptone (Duchefa Biochemie) 0.06 % (w/v) Sodium chloride (VWR) 0.02 % (w/v) Potassium chloride (Roth) 25 M final concentration Magnesium chloride hexahydrate 2M (Roth) 10 % final concentration D (+)-Glucose Monohydrate 20 % (Roth)

Gel Electrophoresis: BIO-RAD PowerPac Basic (300 V, 400 mA, 75 W), BIO-RAD Mini-PROTEAN Tetra System, BIO-RAD Molecular Imager Gel Doc XR⁺ with Image Lab Software.

For the agarose gel electrophoresis, 5:1 (DNA sample/6 × DNA Loading Dye, v/v) was loading into the agarose gel and the gel was run for 30-90 min at 110 V by using 0.5 × TAE buffer as the running buffer. 2 µL 1Kb DNA ladder (NEB N3232S) was used.

For the SDS polyacrylamide gel electrophoresis (SDS-PAGE), 12 μ L protein sample and 3 μ L 4 \times SDS loading buffer were denatured at 95 $^{\circ}$ C for 5 min. The mixture was centrifuged and the supernatant was loaded onto the SDS polyacrylamide gel. The gel was put into the 1 \times SDS Running Buffer (pH 8.3) and run at 20 mA for 60-90 min, and afterwards stained in Coomassie dye solution for 1 h and destained in Coomassie bleach overnight with the help of an agitator. 5 μ L Color Prestained Protein Standard, Broad Range (11–245 kDa, P7712) was used.

The buffers used in gel electrophoresis are listed as follows:

Use	Name	Component
Agarose Gel Electrophoresis	6 \times DNA Loading Dye	0.1 % (w/v) bromophenol blue 0.1 % (w/v) xylene cyanol 30 % (v/v) glycerol
	50 \times Tris-Acetate-EDTA (TAE) pH 8.3	2 M Tris-HCl 1 M Acetic acid 50 mM EDTA
	Agarose gel	Per 25 mL 0.5 \times TAE 0.8 % (w/v) Agarose 1 μ l Roti-GelStain (Roth)
SDS polyacrylamide gel electrophoresis (SDS-PAGE)	4 \times SDS Loading Buffer (pH 6.8)	250 mM Tris-HCl 8 % (w/v) sodium dodecyl sulfate (SDS) 40 % (v/v) glycerol 10 % (v/v) β -mercaptoethanol 2.5 mg/L bromophenol blue
	10x SDS Running Buffer (pH 8.3)	25 mM Tris-HCl 192 mM glycine 0.1 % (w/v) SDS
	SDS polyacrylamide gel - 5 % stacking gel (2.5 mL)	1.7 mL H ₂ O 0.535 mL Acrylamid (Rotiphorese Gel 30 (37.5 :1)) 0.25 mL Tris (0.5 M, pH 6.8) 0.02 mL SDS (10 %) 0.02 mL Ammonium persulfate (APS,10 %) 0.002 mL Tetramethyl-ethylenediamine (TEMED)

	SDS polyacrylamide gel - 10 % separating gel (7.5 mL)	2.95 mL H ₂ O 2.5 mL Acrylamid (Rotiphorese Gel 30 (37.5 :1)) 1.9 mL Tris (1.5 M, pH 8.8) 0.075 mL SDS (10%) 0.075 mL APS (10%) 0.003 mL TEMED
	Coomassie Dye Solution	25 % (v/v) acetic acid 10 % (v/v) isopropanol 0.1 % (w/v) Coomassie Brilliant Blue
	Coomassie Bleach	25 % (v/v) acetic acid 10 % (v/v) isopropanol

PCR: BIO-RAD T100 Thermo Cycler. The proof-reading Takara PrimerStar HS DNA Polymerase was used. 10~50 µL reaction was prepared according to the following components:

Component	10 µL Reaction
5 × PS buffer	2 µL
dNTP mix (2.5 mM each)	0.8 µL
10 µM forward primer	0.4 µL
10 µM revert primer	0.4 µL
PrimerStar Polymerase (2.5 units/µL)	0.2 µL
DNA template (<200 ng)	0.4 µL
H ₂ O	5.8 µL

Thermal cycler conditions were as follows:

Step	Temperature / °C	Time
Denaturation	95	3 min
34 cycles		
Denaturation	98	10 sec
Annealing	T _A	15 sec
Extension	72	1 min / kbp
Final extension	72	5 min
Hold	4	

Quantification: DeNovix DS-11 + Spectrophotometer, Molecular Devices SpectraMax Plus 384, FoodALYT Photometer, Bradford assay.

Clean bench: Thermo Scientific SAFE 2020.

Incubator: New Brunswick Scientific Innova 44.

Centrifuge: Sorvall LYNX 6000 Superspeed.

Sonication: BANDELIN SONOPULS HD 3100 with MS 73 Sonotrode. Put the sonotrode inside (2 cm deep) the cell free extract (30 mL) in a 50 mL centrifuge tube on ice. Set the Ampere at 45 % for 7 min with with pauses every 30s

ITC: TA Instruments NANO ITC

SEC-MALS: Mini DAWN TREOS Wyatt technology MALS.

qPCR: Applied Biosystems StepOne Real-Time PCR System with StepOne Software

FPLC: ÄKTA pure system from the company GE Healthcare combined with the software UNICORN 6.3.

Column chromatography: Ni-NTA (GE HisTrap FF, 5 mL), Desalting (GE HiPrep 26/10, 50 mL), Ion exchange (GE HiTrap DEAE FF, 1 mL), Size exclusion (GE HiLoad 26/600 Superdex 200pg, 320 mL).

LCMS: Analytical LC-MS data were obtained using a Waters LCMS system comprising of a Waters 2767 autosampler, Waters 2545 pump system, a Phenomenex Kinetex column (2.6 μ , C₁₈, 100 Å, 4.6 \times 100 mm) equipped with a Phenomenex Security Guard precolumn (Luna C₅ 300 Å) eluted at 1 mL/min. Detection was by Waters 2998 Diode Array detector between 200 - 600 nm. Waters 2424 ELSD and Waters SQD-2 mass detector measured between 100 - 1000 *m/z*. Solvents were: A, HPLC grade H₂O containing 0.05 % (v/v) formic acid; B, HPLC grade CH₃CN containing 0.05 % (v/v) formic acid. Gradients (H₂O-CH₃CN) were as follows: 0-1 min, 90-10 %; 1-10 min, 10-90 %; 10-12 min, 10-90 %; 12-13 min, 90-10 %; 13-15 min, 90-10 %.

UV-Analysis: Kinetic 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 Julabo F250 Recirculating Cooler at 25 °C. The processed data was transformed into the Michaelis-Menten curve by software Curve Expert.

NMR: Bruker DRX 500 MHz Spectrometer, 293 K. The software MestReNova 10.0 was used for analysis of the data.

6.2 Experimental for Chapter 2

6.2.1 Amplification of the Target Gene

The original gene source comes from the *E. coli* fully codon optimized SQTCS DH-ACP (in pET28a) and the sub codon optimized SQTCS DH-ACP (in PMX) which were synthesized by Thermo Scientific.

Primer list

Target gene	Target vector	Primer 5'-3' (F: forward primer; R: revert primer)
DH-ER	pET28a	F: GTCGACATGAACAAAGCTCATCGTCAGCGTGTACA R: CTCGAGTTATGGCGCGGTGATGACAATTTTGCC
DH-KR	pET28a	F: CATATGAACAAAGCTCATCGTCAGC R: CTCGAGTTATTGGGAGTTACTCAGT
CMeT	pET28a	General R: CTCGAGTTACATATCATCCAGGAAC MT1 F: CATATGATGGCACCGCAACC MT2 F: CATATGTTGGAAATTAAGGGGCTGC MT3 F: CATATGTCAAAGCGGATTCAGAAC MT4 F: CATATGTTTACGGTGAATGTGGACG
DH-ACP	pET28a	F1: GGTGGTGGTCTCGACTCGAGTTACGATTTAGTGGCTACTGTG R1: GTTTCTCCGTATATTCCAG F2: CTGGAATATACGGAGAAAC R2: GTGCCGCGCGGCAGCCATATGAACAAAGCTCATCGTCAG
Sub DH-KR		General R1: AGCCCGGATGAAAAAAGCT General F2: AGCCCGGATGAAAAAAGCT General R2: GTGGTGGTGGTGGTCTCGAGTTACTGCGAATTCGACAATTC
	pET28a	F1: CCGAATTGGAGCTCCGTCGACATAACAAAGCGCATCGCCAG
	pETM11	F1: GTGTTCCAGCAACAGACCGGTGGAAACAAAGCGCATCGCCAG
	pETM22	F1: CCGAATTGGAGCTCCGTCGACATAACAAAGCGCATCGCCAG

Target Gene PCR

Different annealing temperatures (T_A) have been tried in PCR in order to obtain the target gene of the best quality (see general condition in section 6.1 PCR).

Target gene	Selected T_A / °C
DH-KR	63
DH-ER	67
MT1	61
MT2	67
MT3	67
MT4	67
DH-ACP	54
Sub DH-KR	54

6.2.2 Cloning

For downstream processes, plasmids were extracted from a 5 mL *E. coli* overnight culture by standard miniprep procedures. NEB Restriction enzymes were used and the restriction hydrolysis (200 μ L System) was performed according to the manufacturer's protocol. DNA cut from the agarose gel was extracted by NucleoSpin Gel Clean-up kit (Macherey-Nagel) for TA cloning or further ligation step. The quality of purified DNA samples was determined on agarose gels and the concentration was measured on a spectrophotometer (DeNovix DS-11+ Spectrophotometer). Sequencing of DNA samples was carried out through Eurofins Genomics with custom oligonucleotides.

Topo TA cloning (DH-ER, DH-KR and CMeT)

7.8 μ L of PCR product, 1 μ L of 10X ThermoPol® Buffer, 1 μ L of 2.5 mM dATPs and 0.2 μ L Taq DNA Polymerase were mixed into a PCR tube on ice, then incubated the mixture at 68 °C for 10 min. Afterwards, 4 μ L of the A-tailing product mixture was taken into another PCR tube, followed by adding 1 μ L salt solution and 1 μ L PCR 2.1 vector. Incubate the new mixture at 20 °C for 1 hour to get Topo TA cloning product. Subsequently, take all the product for *E. coli* Top 10 transformation (Spread 40 μ L 40 mg/mL X-Gel on solid LB

medium containing 50 µg/mL Kanamycin). The succeed TA clone shows white in this white-blue plaque selection.

T4 ligation (DH-ER, DH-KR and CMeT)

The recombinant plasmid with target gene and empty expression vector were cut with same restriction enzymes to create the insert DNA and vector DNA for ligation. To prevent the cut expression vector from recircularising, Shrimp Alkaline Phosphatase (rSAP) was used for dephosphorylation. The molar ratio of the insert DNA to the vector DNA was determined by spectrophotometer and agarose gel electrophoresis.

Three ligation mixtures were then prepared (10 µL system):

Component	Amount			
	Ligation 1	Ligation 2	Control 1	Control 2
Vector DNA	20 ng (X µL)	60 ng (X µL)	60 ng (X µL)	60 ng (X µL)
Insert DNA	60 ng (Y µL)	20 ng (Y µL)	0	0
Ligase buffer	1 µL	1 µL	1 µL	0
T4 DNA ligase	1 µL	1 µL	1 µL	0
H ₂ O	8-X-Y µL	8-X-Y µL	8-X µL	10-X µL

These mixtures were incubated at 16 °C overnight, then 5 µL was transformed into *E. coli* TOP10 and spread onto agar plates containing 50 µg/mL Kanamycin. The background of recircularised vector not containing an insert was estimated by comparison of the number of colonies on the control plate and the ligation plate.

In-fusion cloning (DH-ACP and sub DH-KR)

Design PCR primers for target gene with 15-bp extensions (5') that are complementary to the ends of the linearized vector. In this experiment, the target gene was separated into two fragments of similar size. Therefore, four primers (F1, R1, F2 and R2) with 15-bp complementary extensions are designed. Among them, F1 and R2 are added with restriction enzyme site for ligating into expression vector. The correct PCR product was extracting from

agarose gel. The expression vector was linearized with the corresponding REs and dephosphorylated by rSAP. The resulted vector DNA was extracted from the agarose gel. Afterwards, 1 μ l 5x In-Fusion HD Enzyme Premix, 5 ng linearized vector (X μ l), 15 ng insert DNA (Y μ l) and H₂O (4-X-Y μ l) were mixed and incubated for 15 min at 50 °C. The mixture was placed on ice before *E. coli* Stellar competent cell transformation.

Transformation

50 μ l competent *E. coli* Top10 cells were thawed on ice for 30 min and followed by adding 1 μ l of the appropriate plasmid. Vibrate it gently, then incubate it on ice for 30 min. For the heat shock transformation, the cells were incubated at 42 °C for 35 s (*E. coli* Stellar cell for 45 s), then immediately chilled on ice for 2 min. Afterwards, 500 μ l SOC medium was added and the transformed cells were shaken (180 rpm) for 1 h at 37 °C. Positive clones were selected by plating 50 μ L and 500 μ l of the cells on solid LB medium containing 50 μ g/mL Kanamycin. As a control, cells with no plasmid being added and were treated with the same procedure. The cells were grown overnight at 37 °C. Plates were stored for up to 1 month at 4 °C. After confirmation the correct clones by PCR and RE digestion, stocks were prepared from 500 μ l overnight culture and 500 μ l 50 % glycerol and stored at -80 °C.

6.3 Experimental for Chapter 3

6.3.1 Expression Cell Transformation

***E. coli* BL21 (DE3)**

50 µl chemically competent *E. coli* BL 21 (DE3) cells were thawed on ice and incubated on ice for 30 min with 1 µl of pET28 inserted with DH to ER. For the heat shock transformation, 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 (180 rpm) for 1 h at 37 °C. Positive clones were selected by plating 50, 200, 600 µl of the cells on solid LB medium containing 50 µg/mL Kanamycin. As a control, cells with empty pET28a being added and were treated with the same procedure. The cells were grown over night at 37 °C and then stored at 4 °C.

***E. coli* ArcticExpress (DE3)**

100 µl chemically competent *E. coli* ArcticExpress (DE3) cells were thawed on ice and incubated on ice for 30 min with 2 µl of pET28 inserted with DH to ER. For the heat shock transformation, the cells were incubated at 42 °C for 20 s, then immediately chilled on ice for 2 min. 800 µl SOC medium were added. The transformed cells were shaken (180 rpm) for 1 h at 37 °C. Positive clones were selected by plating 50, 200, 600 µl of the cells on solid LB medium containing 50 µg/mL Kanamycin and 20 µg/mL of gentamycin. As a control, cells with empty pET 28a being added and were treated with the same procedure. The cells were grown over night at 37 °C and then stored at 4 °C.

Plates were stored for up to 2 weeks at 4 °C. After confirmation the correct clones by PCR and RE digestion, stocks were prepared from 500 µl overnight culture and 500 µl 50 % glycerol and stored at -80 °C.

6.3.2 Expression and Purification of SQTKS Fragments

Expression in *E. coli* BL21 (DE3)

A single colony of expression vector inserted with target gene in *E. coli* BL21 (DE3) was inoculated into 50 mL culture medium. The cells were grown in LB media containing 50 µg/mL Kanamycin and incubated at 37 °C/ 180 rpm overnight. Afterwards, 10 mL of the overnight culture was added to 1 L culture medium containing 50 µg/mL Kanamycin, and then the mixture were mixed and separated into five 500 mL sterile flasks (200 mL medium each). For IPTG induction expression, the cells were incubated at 37 °C/ 180 rpm to an OD₆₀₀ around 0.6, then the flask was cooled down to 16 °C and added with 0.1 mL 1M isopropyl β-D-1-thiogalactopyranoside (IPTG) solution. This solution was incubated for 24 h at 16 °C/ 200 rpm. For auto-induction expression, LBE 5052 medium containing 50 µg/mL Kanamycin was used. The cells were incubated for 2 h at 37 °C / 200 rpm. Afterwards, the temperature was set to 16 °C and the incubation was continued for 24 h at 200 rpm.

Expression in *E. coli* ArcticExpress (DE3)

A single colony of expression vector inserted with target gene in *E. coli* ArcticExpress (DE3) was inoculated into 50 mL LB medium overnight at 30 °C / 180 rpm (with 20 µg/mL of gentamycin and 50 µg/mL of Kanamycin). Afterwards, 10 mL of overnight culture was added into 1 L of culture medium containing no selection antibiotics, and then the mixture were mixed and separated into five 500 mL sterile flasks (200 mL medium each). For the IPTG induction expression, the cells were incubated at 30 °C / 200 rpm to an OD₆₀₀ around 0.6, then the flask was cooled down to 10 °C and added with 0.1 mL 1M IPTG solution. This solution was incubated for 24 h at 200 rpm. For auto-induction expression, LBE 5052 medium was used. The cells were incubated for 2 h at 30 °C / 200 rpm. Afterwards, the temperature was set to 10 °C and the incubation was continued for 48 h at 200 rpm.

Cell lysis

The culture medium containing the cells was centrifuged at 5000 rpm (Sorvall LYNX 6000 Superspeed) for 15 min at 4 °C. The cells could be used immediately or frozen at -20 °C. The cells were suspended in storage buffer (50 mM Tris pH 8, 150 mM NaCl, 20% glycerol (v/v), 1 mM EDTA) and sonicated. Afterwards, cell debris was removed by centrifugation (20000 g for 2 h) and the supernatant was filtered through 0.45 µm syringe layered filter before column chromatography.

Column chromatography

For Ni-NTA, the column loaded with protein sample was washed with 6 × sample column volume of loading buffer (50 mM Tris pH 8, 150 mM NaCl, 20% glycerol (v/v), 1 mM EDTA and 20 mM imidazole). The target protein was then eluted with a linear gradient of elution buffer (50 mM Tris pH 8, 150 mM NaCl, 20% glycerol (v/v), 1 mM EDTA and 0.5 M imidazole) with a flow rate of 4 mL/min. After the Ni-NTA, the elution buffer containing the target protein (peak in UV₂₈₀) was immediately changed to standard storage buffer by a desalting column with a flow rate of 10 mL/min. The purity of the protein was checked by SDS-PAGE and the concentration of protein was estimated by Bradford method.

For IEX, the protein was eluted over a gradient of 0 – 100% of the elution buffer (50 mM Tris pH 8, 1.2 M NaCl, 20% glycerol (v/v), 1 mM EDTA) in 50 mL with a flow rate of 1 mL/min. 2 mL fractions were collected and analyzed by SDS-PAGE.

For SEC, 1 mL of sample protein was injected into a 320 mL column and eluted with 480 mL storage buffer at a flow rate of 1 mL/min. 5 mL fractions were collected and analyzed by SDS-PAGE.

Bradford assay

The concentration of protein is determined by Bradford assay. Standard solutions of bovine serum albumin (0.1-2 mg/ml) in size exclusion buffer (50 mM Tris pH 8, 150 mM NaCl, 20% glycerol (v/v)) were prepared by serial dilution. 100 µl of the standards were mixed with

Bradford dye reagent (1 ml) and incubated for 15 mins at RT. The absorption of each sample was measured at 595 nm against a standard (storage buffer 100 μ l, Bradford dye reagent, 1 ml) to construct a standard concentration curve. A sample of the protein to be quantified (20 μ l) was diluted in size exclusion buffer (80 μ 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.

Purification optimization

For further purification of SQTKS CMeT and DH-KR, the protocol was strictly based on the Figure 3.33.

6.3.3 Thermal Shift Assay (TSA)

TF screen (thermofluor screen) was prepared in a 48 well microplate (MicroAmp[®] Fast Optical 48-Well Reaction plate, Catalog number: 4375928 in Thermo scientific).

SYPRO Orange (Invitrogen S6651 or SigmaAldrich 5000 \times concentrated solution in DMSO) was thaw at room temperature. 120 μ L of 62.5 \times SYPRO Orange was prepared by diluting 1.5 μ L of 5000 \times SYPRO in 118.5 μ L water (always freshly prepared). Filtered buffers and controls were added into the 48 well microplate and place the plate on ice. Into each well 2 μ L of 62.5 \times SYPRO Orange was dispensed (final working concentration of 5 \times). Then into each well 2 μ L of the protein was dispensed (final concentration of 1.6 - 8 μ M). Afterwards the plate was covered with the adhesive film (avoid bubbles between plate and film). the plate was spun down to get rid of air bubbles inside the samples (4 $^{\circ}$ C, 2500 g, 30 seconds). Place the plate in the RT-PCR machine and start the thermal denaturation analysis.

CHC buffer (Citric acid: HEPES: CHES = 2:3:4) in different pH and ionic strength was purchased from Jena Bioscience (JBScreen Thermofluor FUNDAMENT HTS, TSA for protein stability, cat-No.: CS-332). Other buffers were prepared by hand.

6.3.4 Protein Identification

The fragments of protein cut from SDS gel were digested with porcine trypsin and analyzed with ESI-QToF-MS. The system contains a micrOTOF Q II from Bruker coupled to an easy nLC (from Proxeon which was bought by Thermo) together with the software from Bruker and a MASCOT search engine for which an in-house server is available. The *E. coli* database was downloaded from NCBI and the protein sequences were added to this database. Detailed protocol can be found in <https://www.ncbi.nlm.nih.gov/pubmed/29731306>. The analysis was done by Dr. Jennifer Senkler in the group of Prof. Hans-Peter Brau in Plant Genetics of Leibniz Universität Hannover.

6.4 Experimental for Chapter 4

6.4.1 Substrate Synthesis

All the substrates were offered by Christoph Bartel and Oliver Piech in the Cox group, the synthesis methods are based on Christoph Bartel PhD thesis (2017): Investigation of Isolated Domains of the Squalstatin Tetraketide Synthase and Tailoring Enzymes Involved in the Biosynthesis of Squalstatin S1.

6.4.2 Preparation of 4'R-[4'-²H]-NADPD for KR Reaction

1 mg (3.26 U/mg) of alcohol dehydrogenase (NADP⁺ dependent from *thermoaneroobium brockii* (TbADH)) was dissolved in 25 μ L of 25 mM NaH₂PO₄ buffer (pH 8.0). The mixture was added into a solution of NADP⁺ salt (20.8 μ mol), 300 μ L of D₈-isopropanol (Sigma) and 7.5 μ L of 25 mM NaH₂PO₄ buffer (pH 8.0). This solution was incubated at 43 °C for 2 h (monitored by UV₃₄₀). When no further increase in absorption was observed, the remaining isopropanol and generated acetone was removed in vacuo. The product was confirmed by MS (747, [M] H⁺) and used later without further purification (Calculated concentration of 4'R-[4'-²H]-NADPD: 11.5 mM in 1.8 mL of pH 8.0, 100 mM phosphate buffer).

2-methylated acetoacetyl pantetheine **88** (1 μmol), 200 μL of phosphate buffer (as above-mentioned, containing 11.5 mM 4'-R-[4'- ^2H]-NADPD, 2.3 μmol), 50 μL of DH-KR tetradomain (1 mg/mL, 40 % purity after Ni-NTA) in storage Tris buffer. Reacted at 30 $^{\circ}\text{C}$ and monitored by analytic LCMS.

6.4.3 Kinetic Studies of Enoyl Reduction Catalyzed by SQTCS DH-KR Tetradomain

All the components were dissolved in storage buffer (50 mM Tris, pH 8.0, 150 mM NaCl, 20% glycerol (v/v), 1 mM EDTA). Assays were run in 500 μL quartz cuvettes with a path length of 1 cm and the absorption was measured at 340 nm against standard cuvette containing the storage buffer at 25 $^{\circ}\text{C}$. Initial rates were plotted vs substrate concentration.

Component	Volume / μL	Final concentration
Storage buffer	to final volume of 400	
10 mM NADPH	10	0.025 mM
1.0 mg/mL DH-KR (40 %)	10	0.15 μM
5 mM Substrate	160	2 mM
	80	1 mM
	40	0.5 mM
	20	0.25 mM
	10	0.125 mM

6.5 Appendix

6.5.1 Protein Identification by ESI-QTOF-MS

The peptides in red were identified.

DH-ER

NKAHRQRVHPPHDLLGSLIVGRDLREPTWRHFIRVQDIPWIRDHVVQSALVYPGAGFI
 CMAMEAMVQLHELDRDSQSRKVAGYRLAEVDILRAMLIPDTSEGLEAHISLRPCSTKL
 LLTNEWYDFCVSSVGDGDDKFDVHCRGRITIEFDTSGSADTPRTSLRERSRSTGLMRSV
 DPSNLYSFLRAQGIYHGPIFQNLKTISSRKDHSESSFVANTASVMPNGFQSPHVIHPT

TLDSIFQGA Y TALPGAGLDQNTAMIPRSIQELYLSSALTS DVGQCLVSDTSLIR YDGQS
FTVNVDVSSKADSEHTPVLEIKGLRNQSVGQMAPQPGDSSNNDLCFKLDWAPDISSV
KQERLKEKFGFPLDPT EADIIMGLRQACIHFHRSLSL TAPDRDQLDWHQKRFYDW
MVLQIQLAEEDRLAPNSSAWLQCSSSDEQKLENV RASSVNGQMVVHV GKSMLAIL
RHEIAPLEMLQDKLLYRYTDAIKWDRSYQQIDQLVKLHAHKCPTAKIIEIGAGTG
GCTRAVL DALSNQGIARCAQYDFTDVSSGFF EAAQQKFAAFDDVIRFQKLDIEKDIE
MQGFECGSYDLVIASQVLHATGKMEHTMANVRKLLKPGGK LLLVETTRDEMDLQL
VFGLLPGWWSSEEER QMSPSLSTNSWEKVLKKTGFDGLDIELRDCDSDEFYSFSVM
MATASSTIASSMAFAIVYGEVPLPDQFLDDMKTAISSSAVSDPVVGHLSIDATGKF
CIFIEDPETDILSSPDEKSYASIQKLVTRCKGLIWVSRGGAMHGTRPNSSLKTGLLRTL
RLEYTEKRFISLDLDSARPQWNHDSITTINEVLCGAL AQNADSSIKDSEFAEQDQGLF
VPRISCDIARNEDLSSDSNSPAQMEPFHQPGKLLQMGIKTPGLIDTLQFSKTDATDNLP
NDYIEIEPKAFGLNFRDVMVAMGQLEESIMGFECAGVVRVGPSSAGHNIKVGDRVC
ALLGGQWTNTVRVHWHSVSPIQAMDWETAASIPV FVTAYISLVKIARMQAGETVL
IHAASGGVQAAIILAKHVGAEIFATVGTDEKRDLLIKEYKIPDDHIFSSRNALFAKSIR
QRTNGKGV DVVLNCLAGGLLQESFDCLADFGRFIEIGKRDIELNHCLNMGMFARSAT
FTAVDLIAIGRDRSYMFAEALPKIMTLLQEKAIRPVTPI SIYKIGDIETAFRLMQAGKH
MGKIVITAP

ArnA

MKTVVFA YHDMGCLGIEALLAAGYEISAIFTHTDNPGEK AFYGSVARLAAERGIPVY
APDNVNHPLWVERIAQLSPDVIFS FYRHLIYDEILQLAPAGAFNLHGSLLPKYRGRA
PLNWVLVNGETETGVT LHRMVKRADAGAIVAQLRIAIAPDDIAITLHHKLCHAARQL
LEQTLPAIKHGNILEIAQRENEATCFGR RTPDDSFLEWHK PASVLHNMVRAVADPWP
GAFSYVGNQKFTVWSSRVHPHASKAQPGSVISVAPLLIACGDGALEIVTGQAGDGIT
MQGSQLAQTLGLVQGSRLNSQPACTARRRTRVLILGVNGFIGNHLTERLLREDHYEV
YGLDIGSDAISRFLNHPHFHFVEGDISIHSEWIEYHVKKCDVVLPLVAIATPIEYTRNPL
RVFELDFEENLRIIRYCVKYRKR IIFPSTSEVYGMCS DKYFDEDHSNLIVGPVNKPRWI
YSVSKQLLDRVIWAYGEKEGLQFTLFRPFNWMPRLDNLNAARIGSSRAITQLILNLV
EGSPIKLIDGGKQKRCFTDIRDGIEALYRIIENAGNRCDGEIINIGNPENEA SIEELGEML
LASF EKHLRHHFPPFAGFRVVESSSYGKGYQDVEHRKPSIRNAHRCLDWEPKIDM
QETIDETLDFFLRTVDLTDKPS

DH-KR

NKAHRQRVHPPHDLLGSLIVGRDLREPTWRHFIRVQDIPWIRDHV VQSALVYPGAGFI
CMAMEAMVQLHEL RDSQSRKVAGYRLAEVDILRAMLIPDTSEGLEAHISLRPCSTKL
LLTNEWYDFCVSSVGDDDKFVDHCRGRITIEFD TSGSADTPRTSLRERSRSTGLMRSV
DPSNLYSFLRAQGIYHGPIFQNLKTISSRKDHSESSFV VANTASVMPNGFQSPHVIHPT

TLDSIFQGAYTALPGAGLDQNTAMIPRSIQELYLSSALTSVDVGQCLVSDTSLIRYDGQS
FTVNVDVSSKADSEHTPVLEIKGLRNQSVGQMAPQPGDSSNNDLCKFDWAPDISSV
KQERLKEKFGFPLDPTADIIMGLRQACIHFHRSLSLQSLTAPDRDQLDWHQKRFYDW
MVLQIQLAEEDRLAPNSSAWLQCSSSDEQKLENNVRASSVNGQMVVHVHGKSMMLAIL
RHEIAPLEMLQDKLLYRYTDAIKWDRSYQQIDQLVKLHAHKCPTAKIIEIGAGTG
GCTRAVLDAKSNQGIARCAQYDFTDVSSGFFFAAQKFAAFDDVIRFQKLDIEKDIE
MQGFECGSYDLVIASQVLHATGKMEHTMANVRKLLKPGGKLLL VETTRDEM DLQL
VFGLLPGWWSSEEERQMSPSLSTNSWEKVLKKTGFDGLDIELRDCDSDEFYSFSVM
MATASSTIASSSMAFAIVYGEVPLPDQFLDDMKTAISSSAVSDPVVGHLSIDATGKF
CIFIEDPETDILSSPDEKSYASIQKLVTRCKGLIWVSRGGAMHGTRPNSSLKTGLLRTL
RLEYTEKRFISLDLDSARPQWNHDSITTINEVLCGALAQNADSSIKDSEFAEQDQQLF
VPRISCDIARNEDLSSDSNSPAQMEPFHQPGKLLQMGIKTPGLIDTLQFSKTDATDNL
NDYIEIEPKAFGLNFRDVMVAMGQLEESIMGFECAGVVRVGPSSAGHNIKVGDRVC
ALLGGQWTNTVRVHWHSVSPIPQAMDWETAASIPVFTAYISLVKIARMQAGETVL
IHAASGGVQAAAILAKHVGAEIFATVGTDEKRDLLIKEYKIPDDHIFSSRNALFAKSIR
QRTNGKGVDDVVLNCLAGGLLQESFDCLADFGRFIEIGKRDIELNHCLNMGMFARSAT
FTAVDLIAIGRDRSYMFAEALPKIMTLLQEKAIRPVTPISIYKIGDIETAFRLMQAGKH
MGKIVITAPEDAMVPVITRPPKLQLRPDASYLIVGGLGGIGRSLCKNFVENGARSLVL
LSRNANVSQQSGEFLDELSTGCIVGVVDCDISSKTQVEATMLRLKKDMLPIRGIVHA
GMVLQDSVFERMSLDDYNTAIRPKVQGSWNLHSGLSDCDLFFIMLSSLAGVSGSAS
QANYTAGGAYQDALAKYRRAQGLSAVSIDLMVQSVGYVAETK GVAERLVRMGY
SPISEMEVLKIVEHAITNPPPEASSAQIITGISTKPGRHWTESSWLQDARFATLRERARD
VKELSNSQ

~ 80 kDa fragment of DH-KR

NKAHRQRVHPPHDLLGSLIVGRDLREPTWRHFIRVQDIPWIRDHVVQSALVYPGAGFI
CMAMEAMVQLHELKDSQSRKVAGYRLAEVDILRAMLIPDTSEGLEAHISLRPCSTKL
LLTNEWYDFCVSSVGGDDDKFVDHCRGRITIEFDTSGSADTPRTSLRERSRSTGLMRSV
DPSNLYSFLRAQGIYHGPIFQNLKTISSRKDHSESSFVANTASVMPNGFQSPHVIHPT
TLDSIFQGAYTALPGAGLDQNTAMIPRSIQELYLSSALTSVDVGQCLVSDTSLIRYDGQS
FTVNVDVSSKADSEHTPVLEIKGLRNQSVGQMAPQPGDSSNNDLCKFDWAPDISSV
KQERLKEKFGFPLDPTADIIMGLRQACIHFHRSLSLQSLTAPDRDQLDWHQKRFYDW
MVLQIQLAEEDRLAPNSSAWLQCSSSDEQKLENNVRASSVNGQMVVHVHGKSMMLAIL
RHEIAPLEMLQDKLLYRYTDAIKWDRSYQQIDQLVKLHAHKCPTAKIIEIGAGTG
GCTRAVLDAKSNQGIARCAQYDFTDVSSGFFFAAQKFAAFDDVIRFQKLDIEKDIE
MQGFECGSYDLVIASQVLHATGKMEHTMANVRKLLKPGGKLLL VETTRDEM DLQL
VFGLLPGWWSSEEERQMSPSLSTNSWEKVLKKTGFDGLDIELRDCDSDEFYSFSVM
MATASSTIASSSMAFAIVYGEVPLPDQFLDDMKTAISSSAVSDPVVGHLSIDATGKF
CIFIEDPETDILSSPDEKSYASIQKLVTRCKGLIWVSRGGAMHGTRPNSSLKTGLLRTL
RLEYTEKRFISLDLDSARPQWNHDSITTINEVLCGALAQNADSSIKDSEFAEQDQQLF

VPRISCDIARNEDLSSDSNSPAQMEPFHQPGKLLQMGIKTPGLIDTLQFSKTDATDNLP
 NDYIEIEPKAFGLNFRDVMVAMGQLEESIMGFECAGVRRVGPSSAGHNIKVGDRVC
 ALLGGQWTNTVRVHWHSVSPIQAMDWETAASIPVFTAYISLVKIARMQAGETVL
 IHAASGGVGQAAILAKHVGAEIFATVGTDEKRDLLIKEYKIPDDHIFSSRNALFAKSIR
 QRTNGKGVDDVVLNCLAGGLLQESFDCLADFGRFIEIGKRDIELNHCLNMGMFARSAT
 FTAVDLIAIGRDRSYMFAEALPKIMTLLQEKAIRPVTPISIYKIGDIETAFRLMQAGKH
 MGKIVITAPEDAMVPVITRPPKLQLRPDASYLIVGGLGGIGRSLCKNFVENGARSLVL
 LSRNANVSQQSGEFLDELSTGCIVGVVDCDISSKTQVEATMLRLKKDMLPIRGIVHA
 GMVLQDSVFERMSLDDYNTAIRPKVQGSWNLHSGLSDCDLDFFIMLSSLAGVSGSAS
 QANYTAGGAYQDALAKYRRAQGLSAVSIDLGMVQSVGYVAETKGAERLVRMGY
 SPISEMEVLKIVEHAITNPPPEASSAQITGISTKPGRHWTSSWLQDARFATLRERARD
 VKELSNSQ

CMeT

SKADSEHTPVLEIKGLRNQSVGQMAPQPGDSSNNDLCFKLDWAPDISSVKQERLKEK
 FGFPLDPTADIIMGLRQACIHFHRSLSLTAPDRDQLDWHQKRFYDWMVLQIQLAE
 EDRLAPNSSAWLQCSSSDEQKLEENVRASSVNGQMVVHVVGKSMILAIRHEIAPLEL
 MLQDKLLYRYTDAIKWDRSYQQIDQLVKLHAHKCPTAKIIEIGAGTGGCTRAVLD
 ALSNQGICARCAQYDFTDVSSGFFEEAAQQKFAAFDDVIRFQKLDIEKDIEMQGFECGSY
 DLVIASQVLHATGKMEHTMANVRKLLKPGGKLLL VETTRDEMDLQLVFGLLPGW
 LSSEEERQMSPSLSTNSWEKVLKKTGFDGLDIELRDCDSDEFYSFSVMMATASSTIAS
 SSMAFAIVYGEVPLPDQFLDDM

6.5.2 Multiple Sequence Alignment

Alignment among mFAS, CurJ, TENS and SQTks (with predicted boundaries).

Identical amino acids are noted as yellow background (stars). Similar amino acids are noted as green background (dots and colons). * refers to possible reactivity site. The boundaries of CMeT of CurJ is noted in box.

<i>Residue</i>	<i>Foreground</i>	<i>Property</i>
AVFPMILW	RED	Small (small+ hydrophobic)
DE	BLUE	Acidic
RK	MAGENTA	Basic - H
STYHCNGQ	GREEN	Hydroxyl + sulfhydryl + amine
Others	Grey	Unusual amino/imino acids etc

mFAS -----MEEVVIAGMSSGKIPSEEN-LLEFWAN 25
 CurJ MEPTTNKEQLSLSKQMFALKQAEAKLEMMELAKSEPLAIIIGIICREPPGNANTPLSEFWQL 60
 TENS -----MSPMKQN-----EESHVSVEPLAIIIGSAYRFPGGCNTPSPKILWDL 40
 SQTCS MVPYYQPASSCGSNTMAPQPGDSSNAMHQHNEDATIPIAIIIGMSSGKIPGNATSPKILWEL 54
 :.* * . :.* . . :.*

mFAS LIIGVDMVT-ADDRRWKAGLGLER-----RMGKLKLLSRFDASFFGWSHKQA 72
 CurJ LANGEDGVREIPPERWNPDIYHPDPTPGKMYIR--HASLVQVLDQFEPFFFGSPREA 118
 TENS LRQPRDILKEIDPERLNLRRYHPDGETHGSTDVANKAYTLEHISRFDAFFGSPLEA 100
 SQTCS CAQGRSAWSSIPKSRFRQEGYNEAERVGTSHVVGGH-FLEHPSLFDASFFNLSAEAA 113
 . * . : * : . * : . * : . *

mFAS NIDDPQLRMLLEVTEAIVDEGINPASLRGISTGVVWGVSSSDASHALSRDPTLVGYSM 132
 CurJ HSLDPQQRFLLEVTEALERAGINPQOLENHTGVFLGMQNDYALGFNQLNLSPIYDA 178
 TENS ASMDPQQRFTLLEVVESTETAGIPLDKIRGSLTSVHVGVMTTDWAGMQRRTPTLPQYTA 160
 SQTCS KIDDPQFFLQLESVYEMESAGITLEHTAGSDTSVYAGACFRDYHLSLVRDPLPRFLL 173
 :.* * . * . :.* : . : * . * : . : . : . :

mFAS IGCQRAMMANRISFFDFKGPSITIDTACSSSLLALQSAIYQATEGGECSAAVVGGNLVLL 192
 CurJ TGNGFCAAGRSYFLGLQGPSLIDTACSSSIVAIHEACHSLRGEENLALAGGVQLIL 238
 TENS TGIASSIISNRISYIFDLKGASETIDTACSSSIVAIHNAARALQSGDCEKATVAGVNLIL 220
 SQTCS TGNGAARSNRSISHEYDLRGASMTVDTCSTLLTALHLACGLLNRESKTSITVGANVIL 233
 * . : . : * . : . : * . : . : * . : . : * . : . : * . : . : * . : . : *

mFAS LHHHTPNPEIPALQDGRLOVDR--PLPIR---GGNIGINSFQFGGSNVHVLQPNSRP 410
 CurJ LNKHPNPYLDWEN-LPLEVPTLTPWSSSSG--QRVAGVSSFMISGTNAHIVLEAPRE 464
 TENS LWFDKLNPEIARYY-GPLQPTPAIPWPKLAPGTPLRASVNSFQFGGTNAHALERYDAS 458
 SQTCS LNKHPNEKIGQVS-AAVRPSNLQKWPSVSG--VRRASVNNFQYGGANAHVLEISGIPG 459
 : . : * : . : . : . : . : . : . : . : . : . : . : . : . : . : . : . :

mFAS AP---PPAQHA-A-----LPRLLQASGRT---LEAV 434
 CurJ GNRQPPVSPQGGKATGNSE-----DCLERSI---N-LLTLSAKT---EIAL 504
 TENS QS---YCSQW--RRNMTEE-----KTIARTQNNESI--EIPVPLVLTAKT---GRAL 500
 SQTCS HT---PIANGSGRNSGTGNGHNGANGTTNGHNGTNGTTNGHFDATQATNGHYGTDETPD 515
 . : . : . : . : . : . : . : . : . : . : . : . : . : . : . : . : . : . : . :

mFAS QTLI---EQGLR-----SRDLAFVGMLEIAAVSP--V-A 464
 CurJ AELVLSYLVH---YLENHPELELEDVCYTA-----NTGRAHFVKHRLAVIAADRKELVEK 554
 TENS WRTVDAYAQHLR---QHPKLRVTNLSQFMHSRSTHVRVRSFSGA-----SREELVEN 550
 SQTCS YAPVDSFVISISAKEEASARSMVTNLADYLRTLQVQDETKEFKSIA-----HTLGSHR 569
 : . : . : . : . : . : . : . : . : . : . : . : . : . : . : . : . : . : . :

mFAS L--PF-----RG-YAVLGGEGSQEVQVPGKRPVWFTCSGMGQVQGMGLSML-DR 515
 CurJ LHY-----QETGEQAVGLFSGHASKLPPKIVFLFTGQGSQYVNMGRQIYETQPT 603
 TENS LAKFVQAHAADAKSPASQNRIGYSPLHIDPKIAPGLGVFTGQGNQMPAMGRDMVQSPL 610
 SQTCS LFKWTAAKSITGPEELIAAAEGGQFQASRALRTRIGFVFTGQGNQMFAMGREINTYPV 629
 : . : . : . : . : . : . : . : . : . : . : . : . : . : . : . : . : . : . :

mFAS FRDSILRSQALKPLGL-----RVSLLI-----LSTDEAVLDIVSDFVSLTSTQIALI 564
 CurJ FRKALEQCCQILSYLE-----YRLLEVLYPQNAPSSSSSLDQATYTPALFAIFAYALA 658
 TENS FRKTIADCESVLQALPAKDAPVWSSEEL---KKDASTSRLGFAEISQPLCTAVQLALV 666
 SQTCS FRKLDRAIRYLKEF---GCEWSLDEL---SRDAENSNVNIIMTLSPPLCTAVGISLV 681
 ** . : . : * : . : * : . : . : . : . : . : . : . : . : . : . : . : . : . :

ER * ER

mFAS	VRGRMQPGESVLIHSGSGGVGAAAIATLSRGC RVFTTVGSAEKRAYL-QARFPQLDETC	1719
CurJ	-----GIVISL-----	1812
TENS	QHMMQALDSAV---KRHGQGSTAL-----IYGADEELAKLTSERFVAVRESKV	2057
SQTKS	KIARMQAGETVLIHAASGGVGAAILLAKHVGAEIFATVGTDEK-RDLLIKEYKIPDDHI	2070

* :

ER *

mFAS	FANSRDTSFE-----QHVLRH TAGKGVLLVINSLAEEKLQASVRC LAQHGRFL	1767
CurJ	-----EPELNCVR-----	1821
TENS	YF-ASSRTFAPGDWLKVQPLLSKFALSQMI PADVVFIDCLGDTE---SFDACR TLQSC L	2113
SQTKS	FS-SRNALFA-----KSIRQRTNGKGVVVVINC LAGLLQESFDCLADFRFI	2117

. : : :

ER * ER

mFAS	E-IGKFDLSNNHALGMAVFLKRVTFH GILLDSLFEEGGATQV VSELLKAGIQEGVVQPL	1826
CurJ	-----IDLDPHOT-----IKKQADVLEKLIWSK-----	1844
TENS	STTRTVQHRLDACL LSQMSRCS-PDALVD-----AYSMA RTQSN-----	2151
SQTKS	E-IGKRDIELNHCLNMGVFARSATFTAVDLIAIGRDRSYMTALALPKIMTLLQEKAIRPV	2176

: . : :

ER ER KR KR

mFAS	K-CTVFPRTKVEAAF RYM-AQGRHIGKVVIOREEEQGPAPRGLPPIALTGLSKTFCPPH	1884
CurJ	DQEDQVAWRGD-----GRYVA---RIVASHHQ-----QT-DATTEQSLSFRED	1883
TENS	---AEFSWNGYVKTF TAAELAGL SHSLIHSYMTNWQKKDSILVT-VPPLQTRGLFKSD	2207
SQTKS	TPISIIYKIGDIE TAFRLM-QAGRHMG---KIVIT---APEDAMVP-VITRPPKLQLRPD	2227

* : :

KR*****KR

mFAS	KSYVITGSLGGGLQLAOWRERFGAKLVLTSGIR TGYQASQREWRROGVQVVLVST S	1944
CurJ	SYVITGSMGGGLLVARWVDRGAKHLVLSRAPDDA-ANQKTELEMAGA QVVVEKA	1942
TENS	RYLIMVGAAGGIGTSLCRWTVRNGARHVVTSTNPKAD---PIMNEAERYGA AVQVPM	2264
SQTKS	ASYLIVGELGGGRSLCKNFVENGARSVLVLSNANV SQQSGVFDLRLSTGCI VGVVDC	2287

: : : * . * : : : * . * * *

KR KR

mFAS	NASSLDGARSL-ITEATQLGPGGGVFNLMVLRDAVLENQTPFFQD VSKPKYSGIANLD	2003
CurJ	IVSDVESITRVLYKIEHSKIPLAGVIHAAGLSDGVLQNSWSEFQVMAPKVSGVWHLH	2002
TENS	IASSKDSVQTVVDMIRATMPPLAGVGNAAVLRDKLFLDMNVVHMKDVLGPKMGT EHL D	2324
SQTKS	IISKTQVEATMLRLK KDMLEIRGIVHAGMLQDSVBERMSLDYNTAIRPKVIGSWNLH	2347

: . . * : : . : * * : : . . : . * * . * : : *

KR KR

mFAS	RVTREACPELDYFVIFSSVSCGRGNAGQANVGFAN SAMERTCEKRRHDGLPCLAQWGAT	2063
CurJ	QFTQN--QPLDFVLFSSASLLGSPGQGNHSAANGFLDGLAHYRRGIGLPELSVHWGAV	2060
TENS	SIFAQ--EPLDFVLLSSSAAILNNTGQSNVHCANLYMDSLV TNRRSRGLASIVHWGHV	2382
SQTKS	SGLSD--CDLDFVLLSSSLGVSASQSNV TAGGAYQDALAKYRFAQGLSIVSVDLGMV	2405

: * : * : : * : . . * * : : : * * * . : . * :

KR KR

mFAS	GDDGVVLETMGTNDTVI--GGLPQRIASCLVLL---LFLSQPHPV-----	2105
CurJ	AQVGEAER GADV R GQQQMGMA----LSPTQVLSLELLSSTAKAQENNLSDIEVGVV	2115
TENS	CDTGYVAVLVDDTKV-QMSLGTTRVMSVSEIVVHAFAEARGGQPD SRSGSHNIIMGIE	2441
SQTKS	QSVGYVAETKGV AER-LVRMGYSP---ISEMVLKIVEHATTNPP--PEASSAQIITGIS	2459

. * . . * * * * : * : . :

KR KRACP ACP

mFAS	-----LSSFVLAEKRAAAPR DGSQK	2126
CurJ	PI-----EW--SAWQERVA---QWFFFTD WORTIQTTS EISKSEFL-LKL---EATAPS	2160
TENS	PPTKPLDLTKRKPVWISDPRLGCLPFSTLENCMMASEQAAASAVDSL AQQVSEATTDE	2501
SQTKS	TKPG---RHWT ESSLQDARFATL RERARDV KILSN-SGGQKQLAAGQELSMATSLV	2515

: : . : :

	ACP	*****	ACP	
mFAS	---DLVK---AVAHITGIRDVASINPDSTVTLGLDSLMEVENRQILERHDLVLS			2176
CurJ	KRRSLLLAHVRQL--ALVTGINHPESISLETGFTDLGNDSLTVVEFRNKQTFDCSP			2218
TENS	EAAVAALKGFATKLEGIILPLGSIGEDSAGRPTDLGTDSLVEVEFRNWLKQLRVDVP			2561
SQTKS	EAI DVVGRAITAKLATMFLAAS---IIASKSSEYGNDSLIVVEFRNWLAAALSSDMS			2572
	: :	: : * : * * * . * * . * :		
	ACP		ACP	
mFAS	MREVRQL-SLRKIQELSSKST---DADPATPTSHEDSP-----RQQAT			2217
CurJ	STIAFEYANRESITDYLENKIVILSSNSTAASEAETLDDMD--EGSLESEL---ESLA			2272
TENS	VMKILGGSIVGQLSALAA-KLA-----RQDAKKRAQLLEPSGNQPVALPSPPPKDKA			2612
SQTKS	VFDVTVQSSTALAT---LVA-----TKSSRIDKSLVA-----			2603
	: :	* :		
	ACP		ACP	
mFAS	LNLSLTLVNPEGPTLTRLNSVQSAERPLF			2246
CurJ	AEINQLSEDE--MDL-AVSQAVSQLDQLL			2298
TENS	GGLNKNKGS PKLPEIAQVDTVVERMEPLV			2641
SQTKS	-----			2603

6.5.3 Synthesized *E. coli* Codon Optimized Gene

E. coli fully codon optimized DH-ACP of SQTKS

2965-AACAAAGCTCATCGTCAGCGTGTACACCCACCGCACGATCTGTTAGGTTCCC
TGATCGTTGGTCGTGACCTCCGTGAACCGACTTGGCGTCACTTTATCCGTGTGCAA
GATATTCCCGTGGATTCCGCGATCATGTGGTGCAATCTGCCCTGGTATACCCCGGGG
CCGGTTTTATCTGTATGGCAATGGAAGCAATGGTCCAACCTCCATGAACTTCGGGA
TAGCCAGTCTCGGAAAGTGGCCGGCTATCGCCTGGCGGAAGTAGACATCCTGCGT
GCGATGTTAATCCCAGATACGTCGGAAGGTCTGGAAGCGCATATTTCACTCCGCC
CATGCAGTACCAAACCTTCTGCTGACGAACGAATGGTATGACTTTTGTGTCAGCAG
CGTCGGCGATGATGATAAATTCGTGGATCATTGCCGTGGTCGTATTACCATTGAG
TTTGATACGTCCGGTAGTGCGGACACTCCTCGCACGAGTCTGCGCGAACGTTCCC
GCAGTACGGGCTTAATGCGCTCAGTCGATCCGAGCAATTTGTATAGCTTTCTGCG
CGACAAGGCATCTATCATGGCCCGATCTTCCAGAATCTGAAAACCATCTCTAGT
CGTAAGGATCACTCTGAAAGCAGTTTCGTTGTGGCGAATACGGCGTCTGTTATGC
CGAACGGATTTCAAGTCGCCGCACGTCATTCACCCGACCACCCTTGATTGATTTTC
CAGGGTGCATATAACGCGTTACCAGGCGCGGGCCTGGATCAGAATACTGCGATG
ATCCTCGCTCAATCCAGGAGCTGTATCTGTCCAGCGCCTTAACATCCGACGTTG
GGCAGTGCTTGGTCAGCGATACGTCGCTCATTGCTACGATGGCCAATCGTTTAC
GGTGAATGTGGACGTTTCGTCAAAGCGGATTCAGAACATACCCAGTATTGGAA
ATTAAGGGGCTGCGCAACCAGTCAGTGGGACAGATGGCACCGCAACCGGGTGAT
TCCTCTAACAACGATCTGTGCTTCAAACCTGGATTGGGCACCCGATATTTCAAGCG
TGAAGCAAGAGCGCCTTAAAGAGAAGTTTGGCTTTCCTCTGGACCCTACCGAAGC
GGACATTATCATGGGTCTGCGCCAGGCTGTATTCATTCATCCATCGTTTCGTTAC
AGAGCCTTACAGCTCCTGATCGCGATCAATTGGATTGGCATCAGAAACGGTTCTA
CGATTGGATGGTTCTTCAAATCCAGTTGGCAGAGGAAGATCGCCTGGCGCCTAAC
TCTAGCGCATGGTTGCAGTGCAGCTCCTCGGACGAACAGAACTGCTTGAGAATG

TTCGCGCAAGCTCAGTGAATGGCCAGATGGTTGTTTCATGTGGGGAAATCTATGCT
CGCGATCTTACGTCATGAAATTGCCCCGTTAGAACTTATGCTGCAGGACAAACTG
CTGTACCGCTACTACACAGACGCGATCAAATGGGACCGTTCCTATCAGCAAATTG
ATCAGCTCGTGAAACTGCATGCGCACAAATGTCCAACCGCTAAAATTATTGAAAT
TGGCGCAGGAACTGGTGGTTGTACGCGCGCGGTCTCTGGATGCGCTGAGCAACCA
GGGAATTGCCCGTTGTGCCAGTATGACTTTACCGATGTGAGCTCGGGTTTCTTTG
AAGCCGCGCAGCAAAAAGTTCGCCGCATTTCGATGACGTCATTTCGTTTTCAAAGCT
GGACATTGAGAAAGACATCGAAATGCAGGGCTTTGAATGCGGCTCCTACGATCTC
GTCATCGCTAGCCAAGTTTTACACGCTACCGGGAAAATGGAACACACCATGGCG
AATGTCCGCAAGCTCTTGAAACCGGGTGGGAAATTACTGCTGGTTCGAGACAACTC
GGGATGAGATGGATCTGCAATTGGTGTTCGGTCTGCTGCCAGGCTGGTGGCTTAG
TTCTGAGGAAGAACGCCAGATGTCCCCGAGCTTGAGTACCAACAGCTGGGAGAA
AGTTCTGAAGAAAAGTGGCTTTGATGGCCTGGATATTGAGCTGCGCGATTGCGAC
TCCGATGAATTTTACTCGTTTAGTGTGATGATGGCCACCGCTTCTAGTACCATCGC
TTCATCTAGCATGGCTTTTGCCATTGTGTATGGCGAGGTACCCCTGCCGGATCAGT
TCCTGGATGATATGAAAACGGCGATTTCTTCCAGCGCCGTATCGGACCCGGTAGT
AGGCCATCTGGACTCGATTGACGCCACCGGAAAATTTTGCATCTTCATCGAGGAT
CCGGAAACCGATATTCTGAGCTCCCCGGATGAAAAGAGTTACGCGTCGATTCAA
AGCTGGTAACCCGCTGCAAAGGCCTGATTTGGGTGTCACGTGGTGGTGAATGCA
TGAACTCGGCCAATAGCAGCCTTAAAACGGGCTTGCTGCGGACCCCTCCGCCTG
GAATATACGGAGAAACGCTTCATCAGTCTGGATCTCGATTAGCTCGCCCTCAAT
GGAATCACGACTCGATTACCACCATCAATGAAGTATTATGTGGTGCCTCGCTCA
AATGCGGATTCTAGCATCAAAGACTCGGAATTCGCGGAACAGGATGGTCAGTT
GTTTCGTGCCCGTATTTTCATGTGACATTGCCCGGAATGAGGATCTGTCCAGCGAC
TCAAACCTCGCCGGCACAGATGGAACCCTTTCATCAGCCGGGGAAGCTGCTGCAG
ATGGGGATCAAACCCCGGGTCTGATTGACACCCTGCAGTTTAGCAAACGGAC
GCCACGGACAATCTGCCTAACGATTACATTGAGATTGAACCCAAAGCGTTTGGCC
TCAAACCTCCGTGATGTTCATGGTGGCAATGGGCCAGCTGGAAGAAAGCATTATGGG
GTTTGAATGCGCCGGGGTTGTGCGTCGTGTAGGCCCGTCAAGTGCGGGCCACAAC
ATCAAAGTGGGTGATCGCGTCTGTGCCTTGCTGGGTGGCCAATGGACCAACACTG
TCCGCGTGCATTGGCATTCCGTAGCCCCAATTCCACAGGCGATGGATTGGGAGAC
AGCGGCTAGTATTCCGATCGTCTTTGTGACGGCCTATATTTCTCTGGTCAAATCG
CGCGTATGCAAGCCGGCGAAACTGTCTGATTTCATGCGGCTAGTGGTGGGGTTGG
ACAGGCCGCCATCATTCTGGCCAAACACGTAGGGGCTGAAATCTTTGCCACCGTC
GGGACAGATGAGAAACGCGACTTACTGATTAAGGAATATAAGATTCCGGATGAT
CACATTTTCTCCTCGCGCAACGCACTGTTTGCGAAATCAATTCGTCAACGTACCA
ACGGTAAAGGTGTTGATGTGGTCTGAACTGTCTTGCGGGCGGTCTGTTACAGGA
GTCATTCGATTGCTTAGCGGACTTTGGCCGCTTCATTGAGATCGGCAAACGCGAC
ATCGAACTGAACCATTGCCTGAATATGGGCATGTTTCGCACGTTCTGCGACGTTA
CCGCGGTGATTTGATCGCTATTGGGCGCGACCGTTCGTATATGTTTGCCGAAGC
ACTGCCGAAAATCATGACCCTGCTTCAGGAGAAAGCGATTTCGTCCAGTTACACCG
ATTTCAATCTACAAAATCGGTGATATTGAAACCGCCTTCCGCTTAATGCAGGCGG
GTAACACATGGGCAAAATTGTCATCACCGCGCCAGAGGACGCCATGGTGCCAG

TAATCACGCGCCACCGAAACTGCAATTACGCCCTGATGCAAGCTATCTGATTGT
TGGCGGCTTGGGCGGTATTGGCCGCTCTCTTTGCAAGAACTTTGTTGAGAACGGT
GCACGTAGTTTGGTATTATTGAGTCGCAATGCAAATGTTTCACAACAGTCCGGAG
AATTTCTGGATGAACTCCGGTCTACCGGCTGTATCGTGGGTGTTGTGGATTGCGA
CATCTCAAGCAAACACAGGTGGAAGCAACCATGTTGCGTCTGAAGAAGGATAT
GCTCCCTATCCGCGGAATTGTCCACGCAGGAATGGTGCTGCAAGACAGCGTGTTT
GAACGCATGAGCCTGGATGATTACAACACCGCGATTCCGCCGAAAGTGCAGGGG
AGCTGGAATTTGCATTCTGGGCTGAGTGACTGCGATCTGGATTTCTTCATTATGCT
GTCGTGCTTGCCGGAGTTTCTGGCTCGGCTAGTCAGGCAAACCTATAACGGCTGGT
GGTGCGTATCAGGACGCACTTGCGAAATATCGTTCGTGCCAAGGTCTGAGCGCCG
TGAGCATCGATCTGGGTATGGTTCAGTCCGTTGGCTACGTTGCGGAGACAAAAGG
CGTTGCCGAACGGTTAGTGCGCATGGGTTATAGCCCGATCTCCGAAATGGAAGTG
CTGAAAATTGTTGAGCATGCGATTAATACTCCCGCCGGAAGCGTCGAGCGCAC
AGATTATCACGGGCATTTTCGACGAAACCGGGACGCCATTGGACTGAATCCTCCTG
GTTACAGGACGCCCGCTTTGCTACTCTGCGTGAACGCGCGCGCGATGTCAAAGAA
CTGAGTAACTCCCAAGGCGGCGCGCAGGATAAGCAGTTAGCTGCTGGCCAAGAG
CTGAGCATGGCCACTTCTCTCGTGGAGGCAATCGACGTTGTGGGTCTGTGCGATCA
CGGCTAAACTGGCAACCATGTTCTTGATCGCAGCTGAAAGCATTATCGCGAGTAA
AAGCCTTAGCGAATACGGCGTGGATAGCCTGGTTGCGGTAGAATTGCGCAATTGG
TTAGCTGCCAGCTCTCTAGTGATGTGTCCGTCTTTGATGTGACGCAATCTCAGTC
TCTTACCGCGCTGGCAACCACAGTAGCCACTAAATCG-7779

***E. coli* sub-codon optimized DH-ACP of SQTKS**

2965-AACAAAGCGCATCGCCAGCGCGTGCATCCGCCGCATGATTTGCTGGGCAGC
CTGATCGTGGGCGCGATCTGCGGAACCGACGTGGCGGCATTTTATTCGCGTGC
AGGATATCCCGTGGATTCGTGATCATGTGGTGCAGTCGGCGCTGGTGTATCCGGG
CGCGGGCTTTATTTGCATGGCGATGGAAGCGATGGTGCAGCTGCATGAATTGCGT
GATTCTCAGAGCCGAAAGTTGCGGGCTATCGCCTGGCCGAAGTGGATATCCTGC
GGGCGATGCTGATCCCGGATACGAGTGAAGGGCTGGAAGCGCATATTTGCTGC
GCCCCGTGCAGCACGAACTGCTGCTGACGAACGAATGGTATGATTTCTGCGTGTG
GTCGGTGGGGGATGATGATAAATTCGTGGATCATTGCCGCGGCCGTATCACGATT
GAATTCGATACGAGCGGCTCGGCGGATACGCCGCGGACCTCGCTGCGTGAACGT
AGCCGTTCTACGGGTCTGATGCGCAGCGTGGATCCGAGCAACCTGTATAGCTTTC
TGCGTGCGCAGGGGATCTATCATGGCCCGATCTTTCAGAACTTGAAAACGATTTT
GTCGCGGAAAGATCATTTCGAAAGTAGTTTTGTGGTTGCGAATACCGCGTCCGGT
ATGCCGAACGGCTTCCAGAGCCCGCATGTGATTCATCCGACGACCCTGGATTCGA
TCTTTCAGGGTTCGTATAACGGCGCTGCCGGGCGCGGGTCTGGATCAGAACCCGC
GATGATTCCGCGGTTCGATTCAGGAACTGTATTTGAGCAGCGCGCTGACCTCGGAT
GTTGGCCAGTGCCTGGTTTCGGATACGAGTCTGATTCGTTATGATGGCCAGAGCT
TTACGGTGAACGTTGATGTGAGTAGCAAAGCGGATAGCGAACATACCCGGTGC
TGAAATTAAGGCCTGCGTAATCAGAGCGTGGGCCAGATGGCGCCGCAGCCGG
GGGATAGTAGCAATAATGATCTGTGCTTCAAACCTGGATTGGGCGCCGGATATTAG

CTCGGTGAAACAGGAACGTCTGAAAGAAAAATTCGGTTTCCCGCTTGATCCGACC
GAAGCGGATATTATTATGGGCCTGCGTCAGGCGTGTATTCATTTTATTCATCGCA
GCCTGCAGAGCCTGACCGCGCCGGATCGTGATCAGTTGGATTGGCATCAGAAAC
GGTTCTATGATTGGATGGTGCTGCAGATCCAGCTGGCGGAAGAAGATCGTCTGGC
GCCGAACAGCAGTGCCTGGCTTCAGTGTAGCAGCAGCGATGAACAGAACTGCT
GGAAAACGTTTCGTGCGAGTTCTGTGAACGGCCAGATGGTGGTGCATGTGGGCAA
AAGCATGCTGGCGATTCTGCGGCATGAAATCGCGCCGTTGGAACCTGATGCTGCAG
GATAAACTGTTGTATCGCTATTATACGGATGCGATTAAATGGGATCGTAGTTATC
AGCAGATCGATCAGCTGGTGAAACTGCATGCGCATAAATGCCCGACCGCGAAAA
TTATTGAAATCGGTGCGGGCACCGGCGGCTGCACGCGGGCGGTGCTGGATGCGCT
TAGTAATCAGGGTATTGCGCGTTGTGCGCAGTATGATTTTACGGATGTGTCGAGC
GGCTTCTTTGAAGCGGCGCAGCAGAAATTCGCGGCGTTTGATGATGTTATCCGTT
TCCAGAAACTGGATATCGAAAAAGATATTGAAATGCAGGGCTTTGAATGTGGCA
GCTATGATCTGGTTATCGCGAGCCAGTTCTGCATGCGACCGGGAAAATGGAACA
TACCATGGCGAACGTTTCGGAAACTTCTGAAACCGGGCGGCAAACCTGCTGCTTGTG
GAAACGACCCGCGATGAAATGGATCTGCAGTTGGTGTTCGGTCTGCTGCCGGGCT
GGTGGCTGTCTAGTGAAGAAGAACGGCAGATGAGCCCGAGCCTGAGCACGAACA
GCTGGGAAAAAGTTCTGAAAAAAACGGGCTTTGATGGCCTGGATATTGAACTGC
GTGATTGCGATTCTGATGAATTTTATTCGTTTTTCGGTGTGATGATGGCGACCGCGTCT
TCGACGATCGCGTCGTCTTCGATGGCGTTTGCGATTGTTTATGGCGAAGTGCCGCT
GCCGGATCAGTTCCTGGATGATATGAAAACCGCGATTTTCGTCGAGCGCGGTTAGC
GATCCGGTTGTGGGGCATTGGATAGCATCGATGCGACGGGTAAATTTTGCATTT
TTATCGAAGATCCGGAAACGGATATCCTGAGTAGCCCGGATGAAAAAAGCTATG
CGAGTATTCAGAAATTGGTTACGCGTTGCAAAGGCCTGATTTGGGTGAGCCGTGG
CGGCGCGATGCATGGCACGCGCCCGAATTCGAGCCTGAAAACGGGGCTGCTGCG
CACGCTGCGGCTGGAATATACCGAAAAACGTTTTTATTTTCGCTTGATCTGGATTTCG
GCGCGGCCGAGTGAATCATGATAGCATTACCACGATTAATGAAGTTCTGTGTG
GTGCGCTGGCGCAGAACGCGGATAGCTCGATTAAAGATAGCGAATTTGCGGAAC
AGGATGGTCAGCTGTTTGTGCCGCGTATTAGTTGCGATATTGCGCGTAATGAAGA
TCTGTGAGCGATTTCGAACTCTCCGGCGCAGATGGAACCGTTTCATCAGCCGGGT
AAACTGTTGCAGATGGGCATTAACCGCCGGGCTGATTGATACGCTGCAGTTTA
GTAAAACCGATGCGACGGATAACCTGCCGAATGATTATATTGAAATTGAACCGA
AAGCGTTTGGGTTGAATTTTCGCGATGTGATGGTGGCGATGGGCCAGCTGGAAGA
AAGTATTATGGGTTTTGAATGTGCGGGTGTGGTGCCTGCTGTGGGTCCGAGTTCT
GCGGGCCATAATATTAAGTGGGCGATCGCGTGTGCGCGCTTCTGGGCGGCCAGT
GGACGAACACCGTGCGTGTGCATTGGCATAAGTGTGGCGCCGATTCCGCAGGCGAT
GGATTGGGAAACCGCGGCGTTCGATTCCGATTGTGTTTGTGACGGCGTATATTTTCG
CTGGTGAAAATCGCGCGTATGCAGGCGGGCGAAAACGGTGTGCTGATTCATGCGGCG
TCGGGTGGTGTGGGCCAGGCGGCGATTATTCTGGCGAAACATGTGGGCGCGGAA
ATTTTCGCGACGGTGGGCACGGATGAAAAACGGGATCTGCTGATTAAAGAATAT
AAAATCCCGGATGATCATATTTTCAGCTCTCGTAACGCGCTGTTTTCGAAAAGCA
TCCGTCAGCGGACGAATGGCAAAGGTGTGGATGTGGTGTGCTGAATTGCCTTGCGGG
TGGTCTGCTGCAGGAAAGCTTCGATTGCCTGGCGGATTTTGGGCGTTTTTATTGAA

ATTGGCAAACGCGATATTGAACTGAACCATTGCCTGAATATGGGCATGTTTGCGC
GTAGCGCGACCTTTACGGCGGTGGATCTGATTGCGATCGGCCGTGATCGGAGCTA
TATGTTTGCGGAAGCGCTGCCGAAAATTATGACGCTGCTGCAGGAAAAAGCGATT
CGGCCGGTGACGCCGATTAGCATTTATAAAATCGGCGATATCGAAACGGCGTTCC
GCCTGATGCAGGCGGGCAAACATATGGGCAAATCGTGATCACCGCGCCGGAAG
ATGCGATGGTGCCGGTTATTACGCGGCCGCCGAAACTGCAGCTGCGTCCGGATGC
GAGCTATCTGATTGTGGGCGGCTTGGGCGGCATTGGCCGTAGTCTGTGAAAAAT
TTCGTGGAAAACGGCGCGCGTTTCGCTGGTGCTGCTGAGCCGCAATGCGAATGTGA
GTCAGCAGTCTGGTGAATTCTTGGATGAACTGCGCAGCACCGGCTGTATTGTGGG
CGTGGTGGATTGTGATATTTTCGTCTAAAACGCAGGTGGAAGCGACCATGCTGCGC
CTGAAAAAAGATATGCTTCCGATCCGGGGCATTGTGCATGCGGGCATGGTGTTC
AGGATTCGGTTTTTCGAACGTATGTCTCTGGATGATTATAATACGGCGATCCGGCC
GAAAGTTCAGGGCAGCTGGAATCTGCATTCGGGCCTGAGCGATTGTGATCTGGAT
TTTTTTATTATGCTGAGTTCGCTTGCGGGCGTGTGCGGTAGCGCGTCTCAGGCGAA
TTATACGGCGGGTGGCGCGTATCAGGATGCGCTGGCGAAATATCGTCGCGCGCA
GGCCTGTGCGCGGTTTCGATCGATCTGGGCATGGTGCAGTCGGTGGGCTATGTG
GCGGAAACGAAAGGCGTGGCGGAACGCCTGGTGCGCATGGGCTATAGCCCGATT
AGCGAAATGGAAGTTCTGAAAATTGTGGAACATGCGATCACGAATCCGCCGCCG
GAAGCGAGCTCGGCGCAGATTATCACGGGTATTAGTACCAAACCGGGCCGTCATT
GGACGGAATCGTCGTGGCTGCAGGATGCGCGCTTCGCGACGCTGCGGGAACCGG
CGCGCGATGTGAAAGAATTGTCGAATTCGCAGGGCGGGCGCGCAGGATAAACAGC
TGCGGCGGGCCAGGAACTGTCGATGGCGACGTCGCTGGTGGAAAGCGATCGATG
TGGTGGGGCGGGCGATCACCGCGAAACTGGCGACCATGTTTCTGATTGCGGCGG
AATCGATTATTGCGAGCAAATCGCTGAGCGAATATGGCGTGGATTCTGCTGGTGGC
GGTGGAACTGCGTAATTGGCTGGCGGGCGCAGCTGAGCTCGGATGTGAGTGTTTTT
GATGTGACGCAGAGTCAGTCGCTGACGGCGCTGGCGACGACCGTGGCGACGAAA
AGC-7779

6.6 Bibliography

- [1] J. Staunton and K. J. Weissman, *Nat. Prod. Rep.*, **2001**, *18*, 380-416.
- [2] M. S. Butler, *Nat. Prod. Rep.*, **2008**, *25*, 475-516.
- [3] R. J. Cox, *Org. Biomol. Chem.*, **2007**, *5*, 2010-2026.
- [4] D. A. Herbst, R. P. Jakob, F. Zahring and T. Maier, *Nature*, **2016**, *531*, 533-537.
- [5] C. Khosla, Y. Tang, A. Y. Chen, N. A. Schnarr and D. E. Cane, *Annu. Rev. Biochem.*, **2007**, *76*, 195-221.
- [6] A. K. El-Sayed, J. Hothersall, S. M. Cooper, E. Stephens, T. J. Simpson and C. M. Thomas, *Chemistry & Biology*, **2003**, *10*, 419-430.
- [7] N. Pulsawat, S. Kitani and T. Nihira, *Gene*, **2007**, *393*, 31-42.
- [8] J. Davison, J. Dorival, H. Rabeharindranto, H. Mazon, B. Chagot, A. Gruez and K. J. Weissman, *Chem. Sci.*, **2014**, *5*, 3081-3095.
- [9] J. M. Crawford, P. M. Thomas, J. R. Scheerer, A. L. Vagstad, N. L. Kelleher and C. A. Townsend, *Science*, **2008**, *320*, 243-246.
- [10] C. R. Huitt-Roehl, E. A. Hill, M. M. Adams, A. L. Vagstad, J. W. Li and C. A. Townsend, *ACS Chem. Biol.*, **2015**, *10*, 1443-1449.
- [11] T. Moriguchi, Y. Kezuka, T. Nonaka, Y. Ebizuka and I. Fujii, *J. Biol. Chem.*, **2010**, *285*, 15637-15643.
- [12] S. Mori, D. Simkhada, H. Zhang, M. S. Erb, Y. Zhang, H. Williams, D. Fedoseyenko, W.K. Russell, D. Kim, N. Fleer, S. E. Ealick and C. M. H. Watanabe, *Biochemistry*, **2016**, *55*, 704-714.

- [13] H. Sun, C. L. Ho, F. Ding, I. Soehano, X. W. Liu and Z. X. Liang, *J. Am. Chem. Soc.*, **2012**, *134*, 11924-11927.
- [14] J. Foulke-Abel and C. A. Townsend, *ChemBioChem*, **2012**, *13*, 1880-1884.
- [15] J. M. Winter, D. Cascio, D. Dietrich, M. Sato, K. Watanabe, M. R. Sawaya, J. C. Vederas and Y. Tang, *J. Am. Chem. Soc.*, **2015**, *137*, 9885-9893.
- [16] J. Kennedy, K. Auclair, S. G. Kendrew, C. Park, J. C. Vederas and R. C. Hutchinson, *Science*, **1999**, *284*, 1368-1372.
- [17] C. D. Campbell and J. C. Vederas, *Biopolymers*, **2010**, *93*, 755-763.
- [18] J. Barriuso, D. T. Nguyen, J. W. H. Li, J. N. Roberts, G. MacNevin, J. L. Chaytor, S. L. Marcus, J. C. Vederas and D.K. Ro, *J. Am. Chem. Soc.*, **2011**, *133*, 8078-8081.
- [19] K. M. Fisch, W. Bakeer, A. A. Yakasai, Z. Song, J. Pedrick, Z. Wasil, A. M. Bailey, C. M. Lazarus, T. J. Simpson, R. J. Cox, *J. Am. Chem. Soc.* **2011**, *133*, 16635-16641.
- [20] I. Soehano, L. F. Yang, F. Q. Ding, H. H. Sun, Z. J. Low, X. W. Liu and Z. X. Liang, *Org. Biomol. Chem.*, **2014**, *12*, 8542-8549.
- [21] T. Liu, J. F. Sanchez, Y. M. Chiang, B. R. Oakley, C. C. C. Wang, *Org. Lett.* **2014**, *16*, 1676-1679.
- [22] R. A. Cacho, J. Thuss, W. Xu, R. Sanichar, Z. Gao, A. Nguyen, J. C. Vederas and Y. Tang, *J. Am. Chem. Soc.*, **2015**, *137*, 15688-15691.
- [23] D. A. Herbst, C. A. Townsend and T. Maier, *Nat. Prod. Rep.*, **2018**, *35*, 1046-1069.
- [24] T. Maier, M. Leibundgut and N. Ban, *Science*, **2008**, *321*, 1315-1322.
- [25] A. S. Rose, A. R. Bradley, Y. Valasatava, J. M. Duarte, A. Prlić and P. W. Rose, *Bioinformatics*, **2018**, *34*, 3755-3758.

- [26] E. J. Brignole, S. Smith and F. J. Asturias, *Nat. Struct. Mol. Biol.*, **2009**, *16*, 190-197.
- [27] A. Keatinge-Clay, *J. Mol. Biol.*, **2008**, *384*, 941-953.
- [28] D. L. Akey, J. R. Razelun, J. Tehranisa, D. H. Sherman, W. H. Gerwick and J. L. Smith, *Structure*, **2010**, *18*, 94-105.
- [29] W. D. Fiers, G. J. Dodge, D. H. Sherman, J. L. Smith and C. C. Aldrich, *J. Am. Chem. Soc.*, **2016**, *138*, 16024-16036; A. Faille, S. Gavaldà, N. Slama, C. Lherbet, L. Maveyraud, V. Guillet, F. Laval, A. Quemard, L. Mourey and J. D. Pedelacq, *J. Mol. Biol.*, **2017**, *429*, 1554-1569.
- [30] D. Gay, Y. O. You, A. Keatinge-Clay and D. E. Cane, *Biochemistry*, **2013**, *52*, 8916-8928.
- [31] D. A. Herbst, R. P. Jakob, F. Zahringer and T. Maier, *Nature*, **2016**, *531*, 533-537.
- [32] S. A. Bonnett, J. R. Whicher, K. Papireddy, G. Florova, J. L. Smith and K. A. Reynolds, *Chem. Biol.*, **2013**, *20*, 772-783; M. Schafer, C. E. Stevenson, B. Wilkinson, D. M. Lawson and M. J. Buttner, *Cell Chem. Biol.*, **2016**, *23*, 1091-1097.
- [33] A. T. Keatinge-Clay, *Chem. Biol.*, **2007**, *14*, 898-908; A. T. Keatinge-Clay and R. M. Stroud, *Structure* **2006**, *14*, 737-748; J. Zheng and A. T. Keatinge-Clay, *J. Mol. Biol.*, **2011**, *410*, 105-117; J. Zheng, S. K. Piasecki and A. T. Keatinge-Clay, *ACS Chem. Biol.*, **2013**, *8*, 1964-1971; J. Zheng, C. A. Taylor, S. K. Piasecki and A. T. Keatinge-Clay, *Structure* **2010**, *18*, 913-922.
- [34] J. Zheng, D. C. Gay, B. Demeler, M. A. White and A. T. Keatinge-Clay, *Nat. Chem. Biol.*, **2012**, *8*, 615-621.
- [35] J. R. Whicher, S. S. Smaga, D. A. Hansen, W. C. Brown, W. H. Gerwick, D. H. Sherman and J. L. Smith, *Chem. Biol.*, **2013**, *20*, 1340-1351.
- [36] Y. Tang, A. Y. Chen, C. Y. Kim, D. E. Cane and C. Khosla, *Chem. Biol.*, **2007**, *14*,

931-943; Y. Tang, C. Y. Kim, I. I. Mathews, D. E. Cane and C. Khosla, *Proc. Natl. Acad. Sci. U. S. A.*, **2006**, *103*, 11124-11129.

[37] S. Dutta, J. R. Whicher, D. A. Hansen, W. A. Hale, J. A. Chemler, G. R. Congdon, A. R. Narayan, K. Hakansson, D. H. Sherman, J. L. Smith and G. Skiniotis, *Nature*, **2014**, *510*, 512-517.

[38] D. A. Herbst, C. R. Huiitt-Roehl, R. P. Jakob, J. M. Kravetz, P. A. Storm, J. R. Alley, C. A. Townsend and T. Maier, *Nat. Chem. Biol.*, **2018**, *14*, 474-479.

[39] T. D. Sirakova, A. K. Thirumala, V. S. Dubey, H. Sprecher and P. E. Kolattukudy, *J. Biol. Chem.*, **2001**, *276*, 16833-16839.

[40] L. E. N. Quadri, *Crit. Rev. Biochem. Mol. Biol.*, **2014**, *49*, 179-211.

[41] V. Anderson and G. Hammes, *Biochemistry*, **1984**, *23*, 2088-2094.

[42] J. Cortes, K. E. Wiesmann, G. A. Roberts, M. J. Brown, J. Staunton and P. F. Leadlay, *Science*, **1995**, *268*, 1487-1489.

[43] Y. Tang, C. Y. Kim, I. I. Mathews, D. E. Cane, C. Khosla, *Proc. Natl. Acad. Sci. U. S. A.*, **2006**, *103*, 11124-11129.

[44] B. J. Dunn, C. Khosla, *J. R. Soc., Interface*, **2013**, *10*, 20130297.

[45] B. Sedgwick, S. J. French, J. W. Cornforth, R. T. Gray, E. Kelstrup and P. Willadsen, *FEBS J.*, **1977**, *75*, 481-495.

[46] K. J. Weissman, M. Timoney, M. Bycroft, P. Grice, U. Hanefeld, J. Staunton and P. F. Leadlay, *Biochemistry*, **1997**, *36*, 13849-13855.

[47] I. Böhm, I. E. Holzbaur, U. Hanefeld, J. Cortési, J. Staunton and P. F. Leadlay, *Cell Chem. Biol.*, **1998**, *5*, 407-412.

[47] O. Almarsson and T. C. Bruice, *J. Am. Chem. Soc.*, **1993**, *115*, 2125-2138.

- [48] A. Glasfeld, G. F. Leanz, S. A. Benner, *J. Biol. Chem.*, **1990**, *265*, 11692-11699.
- [49] R. Reid, M. Piagentini, E. Rodriguez, G. Ashley, N. Viswanathan, J. Carney, D. V. Santi, C. R. Hutchinson and R. McDaniel, *Biochemistry*, **2003**, *42*, 72-79.
- [50] C. Filling, K. D. Berndt, J. Benach, S. Knapp, T. Prozorovski, E. Nordling, R.; Jörnvall, H. Ladenstein and U. J. Oppermann, *Biol. Chem.*, **2002**, *277*, 25677-25684.
- [51] K. L. Kavanagh, H. Jörnvall, B. Persson, U. Oppermann, *Cell. Mol. Life Sci.*, **2008**, *65*, 3895-3906.
- [52] P. Caffrey, *ChemBioChem*, **2003**, *4*, 654-657.
- [53] C. R. Valenzano, R. J. Lawson, A. Y. Chen, C. Khosla and D. E. Cane, *J. Am. Chem. Soc.* **2009**, *131*, 18501-18511.
- [54] A. Garg, X. Q. Xie, A. Keatinge-Clay, C. Khosla and D. E. Cane, *J. Am. Chem. Soc.* **2014**, *136*, 10190-10193.
- [55] K. J. Weissman, *Beilstein J. Org. Chem.*, **2017**, *13*, 348-371.
- [56] X. Q. Xie, A. Garg, C. Khosla and D. E. Cane, *J. Am. Chem. Soc.*, **2017**, *139*, 3283-3292.
- [57] E. Liddle, A. Scott, L. C. Han, D. Ivison, T. J. Simpson, C. L. Willis and R. J. Cox, *Chem. Commun.*, **2017**, *53*, 1727-1730.
- [58] X. Q. Xie, C. Khosla and D. E. Cane, *J. Am. Chem. Soc.*, **2017**, *129*, 6102-6105.
- [59] Y. M. Shi and H. Bode, *Org. Lett.*, **2018**, *20*, 1563-1567.
- [60] D. Gay, Y. O. You, A. T. Keatinge-Clay and D. E. Cane, *Biochemistry* **2013**, *52*, 8916-8928.
- [61] B. Kusebauch, B. Busch, K. Scherlach, M. Roth, C. Hertweck, *Angew. Chem., Int. Ed.*, **2010**, *49*, 1460-1464.

- [62] F. Lohr, I. Jenniches, M. Frizler, M. J. Meehan, M. Sylvester, A. Schmitz, M. Gütschow, P. C. Dorrestein, G. M. König, T. F. Schüberle, *Chem. Sci.*, **2013**, *4*, 4175-4180.
- [63] D. C. Gay, P. J. Spear, A. T. Keatinge-Clay, *ACS Chem. Biol.*, **2014**, *9*, 2374-2381.
- [64] Palaniappan, N.; Alhamadsheh, M. M.; Reynolds, K. A. *J. Am. Chem. Soc.*, **2008**, *130*, 12236-12237.
- [65] O. Vergnolle, F. Hahn, A. Baerga-Ortiz, P. F. Leadlay, J. N. Andexer, *ChemBioChem*, **2011**, *12*, 1011-1014.
- [66] N. Kandziora, J. N. Andexer, S. J. Moss, B. Wilkinson, P. F. Leadlay and F. Hahn, *Chem. Sci.*, **2014**, *5*, 3563-3567.
- [67] H. Y. He, M. C. Tang, F. Zhang, G. L. Tang, *J. Am. Chem. Soc.*, **2014**, *136*, 4488-4491.
- [68] A. S. Rose, A. R. Bradley, Y. Valasatava, J. M. Duarte, A. Prlić and P.W Rose, *Bioinformatics*, **2018**, *34*, 3755–3758.
- [70] C. R. Valenzano, Y. O. You, A. Garg, A. Keatinge-Clay, C. Khosla, D. E. Cane, *J. Am. Chem. Soc.*, **2010**, *132*, 14697-14699.
- [71] D. L. Akey, J. R. Razelun, J. Tehranisa, D. H. Sherman, W. H. Gerwick, J. L. Smith, *Structure*, **2010**, *18*, 94-105.
- [72] J. W. Labonte and C. A. Townsend, *Chem. Rev.*, **2013**, *113*, 2182-2204.
- [73] X. Q. Xie, A. Garg, C. Khosla, D. E. Cane, *J. Am. Chem. Soc.*, **2017**, *139*, 9507-9510.
- [74] J. Zheng, D. C. Gay, B. Demeler, M. A. White and A. T. Keatinge-Clay, *Nat. Chem. Biol.*, **2012**, *8*, 615-621.
- [75] D. M. Roberts, C. Bartel, A. Scott, D. Ivison, T. J. Simpson and R. J. Cox, *Chem. Sci.*, **2017**, *8*, 1116-1126.

- [76] A. Baxter, B. J. Fitzgerald, J. L. Hutson, A. D. McCarthy, J. M. Motteram, B. C. Ross, M. Sapra, M. A. Snowden, N. S. Watson, R. J. William and C. Wright., *J. Biol. Chem.*, **1992**, 267, 11705-8.
- [77] R. J. Cox, F. Glod, D. Hurley, C. M. Lazarus, T. P. Nicholson, B. A. M. Rudd, T. J. Simpson, B. Wilkinsonc and Y. Zhang. *Chem. Commun.* **2004**, 0, 2260-2261.
- [78] J. M. Crawford, P. M. Thomas, J. R. Scheerer, A. L. Vagstad, N. L. Kelleher and C. A. Townsend, *Science*, **2008**, 320, 243.
- [79] P. A. Storm, D. A. Herbst, T. Maier, and C. A. Townsend, *Cell Chem. Biol.*, **2017**, 24, 316-325.
- [80] A. G. Newman, A. L. Vagstad, P. A. Storm, and C. A. Townsend, *J. Am. Chem Soc.*, **2014**, 136, 7348-7362.
- [81] K. Belecki and C. A. Townsend, *J. Am. Chem. Soc.*, **2013**, 135, 14339-14348.
- [82] N. A. Herman, S. J. Kim, J. S. Li, W. L. Cai, H. Koshino and W. J. Zhang, *Nature Communications*, **2017**, 8, 1514.
- [83] J. M. Crawford, B. C. R. Dancy, E. a. Hill, D. W. Udvary and C. a. Townsend, *Proc. Natl. Acad. Sci. U. S. A.*, **2006**, 103, 16728-16733.
- [84] C. W. Liew, M. Nilsson, M. W. Chen, H. Sun, T. Cornvik, Z.-X. Liang and J. Lescar, *J. Biol. Chem.*, **2012**, 287, 23203-23215.
- [85] R. V. K. Cochrane, Z. Z. Gao, G. R. Lambkin, W. Xu, J. M. Winter, S. L. Marcus, Y. Tang and J. C. Vederas, *Chembiochem*, **2015**, 16, 2479-2483.
- [86] J. M. Winter, D. Cascio, D. Dietrich, M. Sato, K. Watanabe, M. R. Sawaya, J. C. Vederas and Y. Tang, *J. Am. Chem. Soc.*, **2015**, 137, 9885-9893.

- [87] R. A. Cacho, J. Thuss, W. Xu, R. Sanichar, Z. Z. Gao, A. Nguyen, J. C. Vederas and Yi Tang, *J. Am. Chem. Soc.*, **2015**, *137*, 15688-15691.
- [88] H. M. Ge, T. T. Huang, J. D. Rudolf, J. R. Lohman, S. X. Huang, X. Guo and B. Shen, *Org. Lett.*, **2014**, *16*, 3958-3961.
- [89] J. M. Crawford, T. P. Korman, J. W. Labonte, A. L. Vagstad, E. a. Hill, O. Kamari-Bidkorpheh, S.-C. Tsai and C. A. Townsend, *Nature*, **2009**, *461*, 1139-1143.
- [90] T. P. Korman, J. M. Crawford, J. W. Labonte, A. G. Newman, J. Wong, C. A. Townsend and S.-C. Tsai, *Proc. Natl. Acad. Sci. U. S. A.*, **2010**, *107*, 6246-6251.
- [91] Y. Li, I. I. Image, W. Xu, I. Image and Y. Tang, *J. Biol. Chem.*, **2010**, *285*, 22764-22773.
- [92] T. AnnaVal, J. D. Rudolf, C.Y. Chang, J. R. Lohman, Y. C Kim, L. Bigelow, R. Jedrzejczak, G. Babnigg, A. Joachimiak, G. N. Phillips, Jr., and B. Shen, *ACS Omega*, **2017**, *2*, 5159-5169.
- [93] J. Lim, R. Kong, E. Murugan, C. L. Ho, Z.-X. Liang and D. Yang, *PLoS One* **2011**, *6*, e20549.
- [94] S. M. Ma, J. Zhan, K. Watanabe, X. Xie, W. Zhang, C. C. Wang and Y. Tang, *J. Am. Chem. Soc.*, **2007**, *129*, 10642-10643.
- [95] S.M. Ma, J. Zhan, X. Xie, K. Watanabe, Y. Tang and W. Zhang, *J. Am. Chem. Soc.*, **2008**, *130*, 38-39.
- [96] W. Zhang, Y. Li, and Y. Tang, *Proc. Natl. Acad. Sci. U.S.A.*, **2008**, *105*, 20683-20688.
- [97] P. Wattana-amorn, C. Williams, E. Pzosko, R. J. Cox, T. J. Simpson, J. Crosby and M. P. Crump, *Biochemistry*, **2010**, *49*, 2186-2193.
- [98] B. Jia and C. O. Jeon, *Open Biol.*, **2016**, *6*, 160-196.

- [99] P. G. Blommel, P. A. Martin, R. L. Wrobel, E. Steffen, B. G. Fox, *Protein Expr. Purif.*, **2006**, *47*, 562-570.
- [100] T. Nagase, H. Yamakawa, S. Tadokoro, D. Nakajima, S. Inoue, K. Yamaguchi, Y. Itokawa, R. F. Kikuno, H. Koga, O. Ohara, *DNA Res.*, **2008**, *15*, 137-149.
- [101] C. Engler, R. Kandzia and S. Marillonnet, *PLoS ONE*, **2008**, *3*, e3647.
- [102] L. Whitman, M. Gore, J. Ness, E. Theodorou, C. Gustafsson, and J. Minshull, *Genet. Eng. Biotechnol.*, 2013, *33*, 42.
- [103] W. H. Chen, Z. J. Qin, J. Wang and G. P. Zhao, *Nucleic Acids Res.*, **2013**, *41*, e93.
- [104] R. Chao, Y. Yuan and H. Zhao, *FEMS Yeast Res.*, **2015**, *15*, 1-9.
- [105] D. Esposito, L. A. Garvey, C. S. Chakiath, *Methods Mol. Biol.*, **2009**, *498*, 31-54.
- [106] D. L. Cheo, S. A. Titus, D. R. N. Byrd, J. L. Hartley, G. F. Temple and M. A. Brasch, *Genome Res.*, **2004**, *14*, 2111-2120.
- [107] Y. Zhang, U. Werling and W. Edelmann, *Nucleic Acids Res.*, **2012**, *40*, e55.
- [108] K. Motohashi, *BMC Biotechnol.*, **2015**, *15*, 47.
- [109] Y. Okegawa and K. Motohashi, *Anal. Biochem.*, **2015**, *486*, 51-53.
- [110] Y. Okegawa and K. Motohashi, *Biochem. Biophys. Rep.*, **2015**, *4*, 148-151.
- [111] C. Aslanidis, P. J. de Jong, *Nucleic Acids Res.*, **1990**, *18*, 6069-6074.
- [112] D. G. Gibson, L. Young, R. Y. Chuang, J. C. Venter, C. A. III. Hutchison and H. O. Smith, *Nat. Methods*, **2009**, *6*, 343-345.
- [113] C. R. Irwin, A. Farmer, D. O. Willer and D. H. Evans, *Methods Mol. Biol.*, **2012**, *890*, 23-35.

- [114] H. E. Klock, E. J. Koesema, M. W. Knuth and S. A. Lesley, *Proteins*, **2008**, *71*, 982-994.
- [115] M. Z. Li and S. J. Elledge, *Nat. Methods*, **2007**, *4*, 251-256.
- [116] A. V. Bryksin and I. Matsumura, *BioTechniques* **2010**, *48*, 463-465.
- [117] J. Stevenson, J. R. Krycer, L. Phan and A. J. Brown, *PLoS ONE*, **2013**, *8*, e83888.
- [118] S. Thomas, N. D. Maynard and J. Gill, *Nat. Methods*, **2015**, *12*, 11.
- [119] C. M. Camilo and I. Polikarpov, *Protein Expr. Purif.*, **2014**, *99*, 35-42.
- [120] J. L. Schmid-Burgk, T. Schmidt, V. Kaiser, K. Honing and V. Hornung, *Nat. Biotechnol.*, **2013**, *31*, 76-81.
- [121] H. Yuan, L. Peng, Z. Han, J. J. Xie and X. P. Liu, *Front. Microbiol.*, **2015**, *6*, 943.
- [122] S. Shuman, *Proc. Natl. Acad. Sci. USA*, **1991**, *88*, 10104-10108.
- [123] S. Shuman, *J. Biol. Chem.*, **1994**, *269*, 32678-32684.
- [124] R. J. Cox, F. Glod, D. Hurley, C. M. Lazarus, T. P. Nicholson, B. A. M. Rudd, T. J. Simpson, B. Wilkinsonc and Y. Zhang, *Chem. Commun.*, **2004**, *0*, 2260-2261.
- [125] M. A. Skiba, A. P. Sikkema, W. D. Fiers, W. H. Gerwick, D. H. Sherman, C. C. Aldrich and J. L. Smith, *ACS Chem. Biol.*, **2016**, *11*, 3319-3327.
- [126] A. S. Rose, A. R. Bradley, Y. Valasatava, J. M. Duarte, A. Prlić and P. W. Rose, *ACM Proceedings of the 21st International Conference on Web3D Technology (Web3D '16)*, **2016**, 185-186.
- [127] A. S. Rose and P. W. Hildebrand, *Nucleic Acids Res.*, **2015**, *43*, W576-W579.
- [128] D. Ivison, PhD thesis: Investigating the Programming of Type I Iterative Polyketide

Synthase Enzymes, **2013**.

[129] C. H. Yu, Y. K. Dang, Z. P. Zhou, C. Wu, F. Z. Zhao and Yi Liu, *Molecular Cell*, **2015**, *59*, 744-754.

[130] E. R. LaVallie, E. A. DiBlasio, S. Kovacic, K. L. Grant, P. F. Schendel and J. M. McCoy, *Biotechnology (NY)*, **1993**, *11*, 187-193.

[131] J. G. Marblestone, S. C. Edavettal, Y. Lim, P. Lim, X. Zuo, T. R. Butt, *Protein Sci.*, **2006**, *15*, 182-189.

[132] P. Dvorak, L. Chrast, P. I. Nikel, R. Fedr, K. Soucek, M. Sedlackova, R. Chaloupkova, V. de Lorenzo, Z. Prokop and J. Damborsky, *Microb Cell Fact.*, **2015**, *14* (201).

[133] Y. Ma, L. H. Smith, R. J. Cox, P. Beltran-Alvarez, C. J. Arthur and T. J. Simpson, *ChemBioChem*, **2006**, *7*, 1951-1958.

[134] M. Ferrer, T. N. Chernikova, M. Yakimov, P. N. Golyshin, and K. N. Timmis, *Nat Biotechnol.*, **2003**, *21*, 1266-1267.

[135] S. Wee and B. Wilkinson, *J. Bacteriol.*, **1988**, *170*, 3283-3286.

[136] F. W. Studier, *Protein Expr. Purif.*, **2005**, *41*, 207-234.

[137] A. G. Kikhney and D. I. Svergun, *FEBS Letters*, **2015**, *589*, 2570-2577.

[138] L. M. Hjelmeland and A. Chrambach, *Methods Enzymol.*, **1989**, *104*, 305.

[139] J. Neugebauer, *A Guide to the Properties and Uses of Detergents in Biology and Biochemistry*, Clarkson University, **1988**.

[140] A. McPherson, *Eur. J. Biochem.*, **1990**, *189*, 1-23.

[141] A. Ducruix and R. Gieg, *Crystallization of Nucleic Acids and Proteins: A Practical Approach*, Oxford University Press, **1992**.

- [142] S. N. Timasheff, *Annu. Rev. Biophys. Biomol. Struct.*, **1993**, *22*, 67-97.
- [143] E. H. Yancey, M. E. Clark, S. C. Hand, R. D. Bowlus and G. N. Somero, *Science*, **1982**, *217*, 1214-1222.
- [144] C. H. Schein, *Biotechnology*, **1990**, *8*, 308-317.
- [145] E. P. Pittz and S. N. Timasheff, *Biochemistry*, **1978**, *17*, 615-623.
- [146] S. N. Timasheff, in *Methods in Molecular Biology, Protein Stability and Folding: Theory and Practice*, ed. B. A. Shirley, Humana Press Inc., Totowa, **1995**, vol 40, ch. 11.
- [147] H. Palsdottir and C. Hunte, *Biochim Biophys Acta.*, **2004**, *1666*, 2-18.
- [148] P. Cimmerman, D. Matulis, in *Biophysical Approaches Determining Ligand Binding to Biomolecular Targets: Detection, Measurement and Modelling*, ed. A. Podjarny, A. P. Dejaegere and B. Kieffer, Royal Society of Chemistry, **2011**, ch 8, 247-274.
- [149] D. V. Rial and E. A. Ceccarelli, *Protein Expr Purif.*, **2002**, *25*, 503-507.
- [150] G. L. Rosano and E.A. Ceccarelli, *Front Microbiol.*, **2014**, *5*, 172.
- [151] K. B. Steinbuch and M. Fridman, *Med. Chem. Commun.*, **2016**, *7*, 86-102.
- [152] C. Robichon, J. Luo, T. B. Causey, J. S. Benner and J. C. Samuelson, *Appl. Environ. Microbiol.*, **2011**, *77*, 4634-4646.
- [153] B. H. Zimm, *J. Chem. Phys.*, **1948**, *16*, 1099-1116.
- [154] *Handbook of Design of Experiments in Protein Production and Purification*, GE Healthcare.
- [155] A. N. Naganathan and V. Muñoz, *J. Am. Chem. Soc.*, **2005**, *127*, 480-481.
- [156] P. J. Wyatt, *Anal Chim Acta.*, **1993**, *272*, 1-40.

- [157] M. Leibundgut, S. Jenni, C. Frick and N. Ban, *Science*, **2007**, *316*, 288–290.
- [158] O. A. E. Seoud, M. -F. Ruasseb and W. A. Rodriguesa, *J. Chem. Soc., Perkin Trans. 2*, 1053–1058
- [159] K. K. Siu, K. Asmus, A. N. Zhang, C. Horvatin, S. Li, T. Liu, B. Moffatt, V. L. Jr. Woods and P. L. Howell, *J. Struct. Biol.*, **2011**, *173*, 86–98.
- [160] J. Siegrist, S. Aschwanden, S. Mordhorst, L. Thöny-Meyer, M. Richter and J. N. Andexer, *ChemBioChem*, **2015**, *16*, 2576–2579.
- [161] X. W. Zou, Y. C. Liu, N. S. Hsu, C. J. Huang, S. Y. Lyu, H. C. Chan, C. Y. Chang, H. W. Yeh, K. H. Lin, C. J. Wu, M. D. Tsai and T. L. Li, *Acta Cryst.*, **2014**, *70*, 1549-1560.
- [162] M. A. Skiba, A. P. Sikkema, N. A. Moss, C. L. Tran, R. M. Sturgis, L. Gerwick, W. H. Gerwick, D. H. Sherman and J. L. Smith, *ACS Chem. Biol.*, **2017**, *12*, 3039–3048.
- [163] R. Castonguay, W. H. A. Y. Chen, C. Khosla and D. E. Cane, *J. Am. Chem. Soc.*, **2007**, *129*, 13758–13769.
- [164] A. Roujeinikova, W. J. Simon, J. Gilroy, D. W. Rice, J. B. Rafferty and A. R. Slabas, *J. Mol. Biol.*, 2007, *365*, 135–145.
- [165] L. Zhang, J. Ji, M. Yuan, Y. Feng, L. Wang, Z. Deng, L. Bai and J. Zheng, *ACS Chem Biol.*, 2018, *13*, 871-875.
- [166] M. P. Ostrowski, D. E. Cane and C. Khosla, *J. Antibiot.*, **2016**, *69*, 507–510.

- [167] A. Miyanaga, S. Iwasawa, Y. Shinohara, F. Kudo and T. Eguchi, *Proc. Natl. Acad. Sci. U. S. A.*, **2016**, *113*, 1802–1807.
- [168] J. Lobstein, C. A. Emrich, C. Jeans, M. Faulkner, P. Riggs and M. Berkmen, *Microb Cell Fact.*, **2012**, *11* (56).
- [169] A. Gąciarz, N. K. Khatri, M. L. Velez-Suberbie, M. J. Saaranen, Y. Uchida, E. Keshavarz-Moore and L. W. Ruddock, *Microb Cell Fact.*, **2017**, *16* (108).
- [170] A. Garg, X. Xie, A. Keatinge-Clay, C. Khosla and D. E. Cane, *J. Am. Chem. Soc.*, **2014**, *136*, 10190-10193.

Curriculum Vitae

Education

- 10/2014-** **Leibniz Universit ät Hannover, PhD student, Chemistry**
- 12/2018** Department of Organic Chemistry, Prof. Russell J. Cox
-
- 09/2011-** **Xiamen University, China, Master of Organic Chemistry**
- 09/2014** Department of Organic Chemistry, Prof. Eric Meggers
-
- 09/2007-** **Anqing Normal College, China, Bachelor of Chemistry**
- 06/2011** Department of Chemistry
-
- 09/2003-** **Anqing No.1 Middle School, China, High School**
- 06/2007** **Anqing No.2 Middle School, China, High School**

Publications

H. Yao, L. Zhang, P. Qiu, P. J. Carroll, L. Gong and E. Meggers, *Inorganica Chimica Acta*, **2014**, *421*, 489-495.

Results of this project will be published in due course.