Engineering of Modular Polyketide Synthases

Dissertation
zur Erlangung des Doktorgrades
der Naturwissenschaften

vorgelegt beim
Fachbereich Biochemie, Chemie und Pharmazie (FB 14)
der Johann Wolfgang Goethe-Universität
in Frankfurt am Main

von
Maja Klaus
aus Melle

Frankfurt 2019
(D30)

vom Fachbereich Biochemie, Chemie und Pharmazie (FB 14) der Johann Wolfgang Goethe-Universität als Dissertation angenommen.

Dekan: Prof. Dr. Clemens Glaubitz

1. Gutachter: Prof. Dr. Martin Grininger
2. Gutachter: Prof. Dr. Chaitan Khosla

Datum der Disputation:

Zusammenfassung

Polyketidsynthasen (PKS) sind eine faszinierende Klasse von sekundärmetaboliten–produzierenden Megaenzymen mit Molekulargewichten im Megadaltonbereich. PKS kommen in Bakterien, Pilzen und Pflanzen vor und sind für die namensgebende Synthese von Polyketid–Naturstoffen verantwortlich, strukturell hochkomplexer Moleküle zu denen viele Verbindungen mit starker Bioaktivität wie Antibiotika, Anti–Cholesterin–Verbindungen und Immunsuppressiva gehören. Trotz der Komplexität des Endprodukts basiert die Polyketidbiosynthese auf vergleichsweise simplen Reaktionsschritten, dem wiederholten Kondensieren von Acetyl–Coenyzm–A–Molekülen. Das molekulare Engineering von PKS wird daher als vielversprechende Möglichkeit gesehen, neue Polyketide mit veränderten Eigenschaften zu kreieren. Hierbei ist besonders die Klasse der modularen PKS von großem biotechnologischen Interesse, in welcher die einzelnen Kondensationsreaktionen auf separaten Proteinen durchgeführt werden, welche durch nicht–kovalente Protein–Interaktionen zu einem Megaenyzm assemblieren (Abbildung 1A). Durch die modulare Architektur dieser PKS sollte es möglich sein, neue, komplexe Naturstoffe über den Austausch einzelner Module oder Domänen, wodurch chimäre PKS entstehen, herzustellen.

Trotz der Möglichkeit zum Engineering, welche die modularen PKS–Systeme auf den ersten Blick aufweisen, zeigt sich jedoch immer deutlicher, dass ihre Aktivität auf dem effizienten Zusammenspiel ihrer Module über spezifische Protein–Protein–Wechselwirkungen sowie den Substrat–Selektivitäten ihrer katalytischen Domänen beruht. In zahlreichen Engineering–Studien wurden hierbei die gegenwärtigen Grenzen der Entwicklung von chimären PKS aufgezeigt, die sich sowohl in verringerten Umsatzraten als auch geringeren Produktausbeuten im Vergleich zum Wildtyp–Enzym widerspiegeln. Um das Engineering–Potential von PKS effizient ausschöpfen zu können, ist daher ein grundlegenderes Verständnis der spezifischen Protein–Protein– und Protein–Substrat–Wechselwirkungen von Nöten.
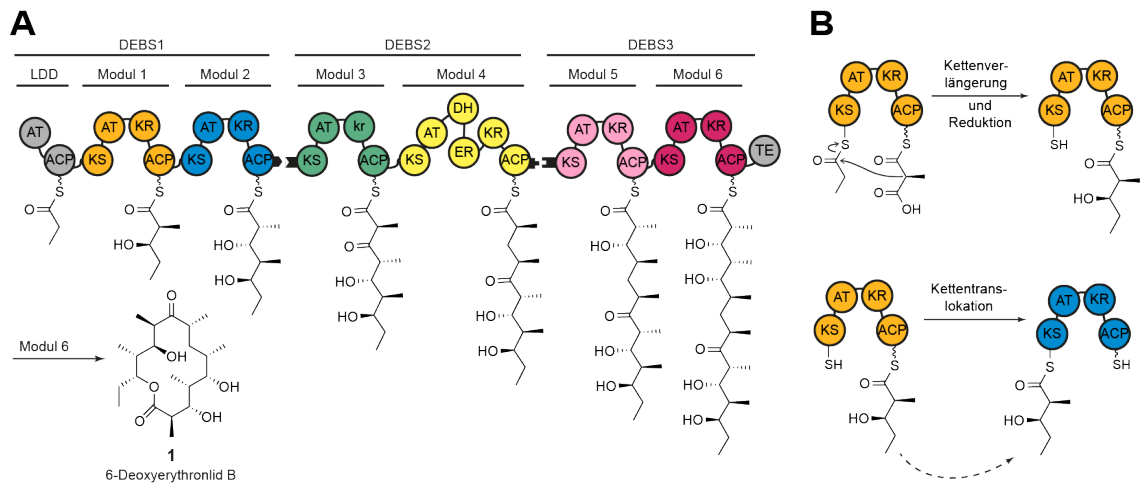
Figure 1 Beispiel einer modularen Polyketidsynthase und wichtiger Reaktionsschritte. (A) Die 6–Deoxyerythronolid B Synthase besteht aus drei Polypeptiden (DEBS1–3), welche jeweils zwei Module enthalten, und produziert 6–Deoxyerythronolid B (**1**), den Makrolidvorläufer das Antibiotikums Erythromycin. Domänenbennenung: LDD – Ladedomäne, AT – Acyltransferase, KS – Ketosynthase, ACP – Acyl–Carrier–Protein, KR – Ketoreduktase, DH –Dehydratase, ER – Enoylreduktase und TE – Thioesterase. Docking–Domänen sorgen für eine nicht–kovalente Verknüpfung der einzelnen Proteine (schwarze Domänen) (B) Die Kettenverlängerung und die Kettentranslokation sind wichtige Reaktionsschritte, welche auf der Interaktion zwischen einem Acyl–Carrier–Protein (ACP) und einer Ketosynthase (KS) beruhen. Sie finden entweder innerhalb eines Moduls (Verlängerung) oder zwischen zwei PKS–Modulen (Translokation) statt.

Um dies zu erreichen bedarf es einer umfassenden biochemischen Analyse des Reaktionsmechanismus sowie der Aufklärung struktureller Eigenschaften des Enyzmkomplexes, darüber hinaus aber auch neuer *In–silico*–Ansätze zur Modellierung transienter Domänen–Domänen–Wechselwirkungen.

Das Ziel dieser Dissertation war es, das grundlegende Verständnis über Funktion und Struktur von modularen PKS zu erweitern, als auch auf Basis der gewonnenen Erkentnisse neue Strategien für die Entwicklung chimärer PKS zu entwickeln, mit dem ultimativen Ziel neuartige Polyketide herzustellen. Im Speziellen sollten Domänen–Domänen–Wechselwirkungen charakterisiert und Engpässe in der Aktivität von chimären PKS identifiziert werden. Zur Unterstützung der Engineering–Ansätze wurde die strukturelle Charakterisierung eines intakten PKS–Moduls und seiner transienten Domänen–Domänen–Wechselwirkungen durchgeführt.

Dass der kombinatorische Zusammenbau von Modulen aus verschiedenen PKS prinzipiell möglich ist, wurde bereits zuvor in *In–vivo*–Studien aufgezeigt. Die

resultierenden chimären Systeme wiesen jedoch häufig verminderte Produktausbeuten auf, die relativen Beiträge von Protein–Protein–Wechselwirkungen und Enzym–Substrat–Erkennung zum Aktivitätsverlust blieben jedoch unklar. Das erste Projekt meiner Doktorarbeit bildete daher der Abschluss eines während meiner Masterarbeit begonnenen Projektes, in welchem ich eine Vielzahl chimärer bimodularer und trimodularer PKS charakterisiert habe, um den relativen Einfluss der Protein–Protein– und Enzym–Substrat–Erkennung auf den Umsatz chimärer PKS zu bestimmen. Während die Analyse der Protein–Protein–Wechselwirkungen weitgehend in der Masterarbeit abgeschlossen wurde, wurden zu Beginn der Doktorarbeit Experimente zur Charakterisierung der Substratspezifitäten durchgeführt. Ähnlich wie in vorangegangenen *In–vivo*–Studien wiesen die in diesem Projekt analysierten chimären PKS stark verringerte Umsatzraten auf. Die Analyse unserer Daten erlaubte uns allerdings den Grund für die verringerte Aktivität zu bestimmen und ergab, dass die Kettentranslokation über die heterologe Schnittstelle der geschwindigkeitsbestimmende Schritt in chimären PKS ist (Beispiel für eine Kettentranslokationsreaktion in Abbildung 1B). Darüber hinaus haben wir mithilfe einer Kombination aus zielgerichteter Mutagenese und Analyse der Substrattoleranz festgestellt, dass speziell die Protein–Protein–Wechselwirkung an der heterologen Grenzfläche, nämlich die Wechselwirkung des Acyl–Carrier–Proteins (ACP) mit der Ketosynthase (KS), die Aktivität von chimären PKS maßgeblich beeinflusst. Unsere Ergebnisse geben Aufschluss darüber, wie Engineering–Versuche zur Rekombination von intakten PKS–Modulen angegangen werden sollten, und zwar durch erstens Überwindung der Geschwindigkeitsbeschränkung, die durch eine beeinträchtigte ACP:KS–Erkennung entsteht, und zweitens der Optimierung der KS–Substratspezifität. Diese Einsicht diente als Ausgangspunkt für die folgenden Arbeiten in dieser Dissertation.

Der Fokus des nächsten Projektes lag demnach auch in der Optimierung einer spezifischen ACP:KS–Schnittstelle mittels Experimenten zur gerichteten Evolution. Da selbst die natürlichen ACP:KS–Wechselwirkungen schwach sind, ist eine noch

schwächere Wechselwirkung in einem chimären, nicht aufeinader abgestimmten System zu erwarten. Das Ziel war es daher, die Affinität eines ACPs für eine nicht–native KS zu erhöhen, in der Hoffnung, dass eine erhöhte Affinität mit einer Erhöhung der Kettentranslokationsrate korreliert, und dadurch mit einer Steigerung der Gesamtenzymaktivität. Hierzu etablierte ich Phagen–Display–Experimente, welche es mir ermöglichten aus Bibliotheken von mutierten ACPs, welche an Phagen gebunden waren, ACP–Mutanten mit erhöhter Affinität zu einer heterologen KS zu selektieren. Im Rahmen dieses Projekts konnten einige ACP–Mutanten mit erhöhter Bindungsaffinität zur KS angereichert werden, von denen allerdings nur zwei einen erhöhte Aktivität im Kontext eines chimären PKS–Systems zeigten. Dissoziationskonstanten für die ACP:KS–Wechselwirkung konnten nicht ermittelt werden, aus Schätzungen der Affinitäten aus ELISA–Experimenten konnte jedoch keine Korrelation zwischen erhöhter Affinität und verbesserter Aktivität festgestellt werden. Obwohl die Idee vielversprechend erscheint, Methoden zur gerichteten Evolution zu verwenden um chimäre ACP:Domänen–Schnittstellen zu optimieren, ist sie zum gegenwärtigen Zeitpunkt nicht ohne Weiteres realisierbar. Eines der Hauptprobleme scheint die inhärent schwache Bindungsaffinität zwischen dem ACP und den katalytischen Domänen zu sein, welche zu Sensitivitäts–Problemen während der Biopanning–Experimente führt.

Das Ziel eines weiteren Projektes war es daher, den Einsatz stabil interagierender, synthetischer Interaktionsdomänen (SYNZIP–Domänen) zur Generierung von chimären Systemen zu etablieren. Hierzu verglich ich verschiedene Designs von chimären PKS in denen SYNZIP–Domänen entweder zwischen Modulen oder innerhalb eines Moduls installiert wurden. So wurden Chimäre erstellt die entweder wie zuvor eine nicht–natürliche Kettentranslokations–, oder aber eine chimäre Kettenverlängerungs–Schnittstelle aufwiesen. Während die Kettentranslokations–Schnittstelle im PKS–Engineering gut etabliert ist, war die Möglichkeit der Nutzung von Docking–Domänen an der Grenzfläche zur Kettenverlängerung unbekannt. Meine Analysen haben gezeigt, dass SYNZIP–Domänen

ein beträchtliches Potenzial zur Nutzung im PKS–Engineering bergen, da es durch sie möglich wird, Module mit hoher Affinität zu verknüpfen und sie damit als Ersatz für kovalente Linker genutzt werden können. Der Vergleich der zwei im SYNZIP–Projekt analysierten Schnittstellen zeigte, dass beide gleichermaßen zur Generierung von chimären PKS genutzt werden können, jedoch stets in Verbindung mit Aktivitätseinbußen. Insbesondere zeigt sich im Vergleich beider Schnittstellen die Notwendigkeit zur individuellen Optimierung eines gewünschten PKS–Systems, da z. B. in Fällen mit einer nativen Protein–Protein–Schnittstelle in der Kettentranslokation Aktivitätseinbußen festgestellt werden konnten, welche auf eine Inhibierung durch die Substrat–Spezifität der Akzeptor–KS–Domäne hindeutet.

Infolgedessen, wurden als letzter Engineering–Ansatz, im Zuge einer von mir betreuten Masterarbeit, Möglichkeiten zur Erhöhung der Substrat–Toleranz von KS–Domänen mittels mehrerer gleichzeitiger Mutationen im aktiven Zentrum der KS untersucht. Ziel war es, die Substratspezifität der KS zu erweitern, um die Umsatzraten von kinetisch beeinträchtigten chimären PKS zu erhöhen. Der Ansatz zur Verwendung von Mehrfach–Mutationen wurde gewählt, um das aktive Zentrum der KS möglichst großflächig abzudecken. Zur Auswahl der Mutationen benutzten wir den FuncLib–Server, welcher es ermöglicht, eine Vielzahl verschiedener Mehrfach–Mutanten unter Berücksichtigung des Erhalts stabiler Interaktionen im aktiven Zentrum zu generieren. Am Beispiel von zwei KS–Domänen wurden Mutanten mit bis zu fünf gleichzeitig mutierten Resten im aktiven Zentrum hergestellt. Alle Mutanten wiesen ähnliche Aufreinigungseigenschaften wie das Wildtyp–Modul auf und zeigten keine Änderung in der Stabilität und in ihrem Oligomerisierungsverhalten. Dies deutet auf einen erfolgreichen Einsatz von FuncLib bei der Berechnung stabiler Mutanten im aktiven Zentrum auf der Basis einer Proteinstruktur hin. Die Analyse der bimodularen chimären PKS unter Verwendung der mutierten Module ergab eine weite Bandbreite veränderter Umsatzraten, welche von keinem Effekt bis zu mehrfach erhöhten Ak-

tivitäten reichen. Unsere Ergebnisse unterstreichen damit die Relevanz der KS–Substratspezifität für das Design von chimären PKS und etablieren die zielgerichtete Mehrfach–Mutagenese als Methode zur Erhöhung der Substrattoleranz von KS–Domänen.

Die unterschiedlichen Engineering–Strategien welche in dieser Arbeit verfolgt wurden, zeigen, dass die Generierung von chimären PKS, basierend auf dem derzeitigen Kenntnisstand, nicht mittels verallgemeinerbarer Ansätze durchführbar ist. Um zukünftige Engineering–Ansätze zu erleichtern sind bessere Strukturmodelle und *In–silico*–Modelle von Domänen–Domänen–Wechselwirkungen und von Substrat–Domänen–Wechselwirkungen erforderlich.

Der abschließende Fokus dieser Arbeit lag in der strukturellen Charakterisierung eines PKS–Moduls, um Einblicke in die übergeordnete Architektur des Moduls aber auch spezifischer Domänen–Domänen–Wechselwirkungen zu erhalten. Hierzu verwendeten wir eine Strategie, welche es uns ermöglicht, stabile und transiente Domänen–Domänen–Wechselwirkungen zu erfassen und diese mithilfe von computergestützten Simulationen zusammenzusetzen. Innerhalb eines Kooperationsprojektes kombinierten wir *small–angle X–ray scattering*, *cross–link mass spectrometry* und Strukturmodellierung, um ein PKS–Modul in Lösung zu analysieren. Unsere vorläufigen Modelle weisen auf eine große Konformationsflexibilität hin, welche jedoch noch unter Zuhilfenahme der *Cross–link*–Daten evaluiert werden muss.

Abstract

Polyketide synthases (PKSs) are large megaenzymes that occur in bacteria, fungi, and plants and produce polyketides, a class of secondary metabolites. Many polyketide natural products exhibit high biological activities e.g. as antibiotics or anti–fungal compounds. The modular architecture of assembly line PKSs makes them exciting targets for engineering approaches via the exchange of whole modules or single domains. Although many engineering attempts have been pursued over the last three decades, the resulting chimeric PKSs often exhibit decreased turnover rates or diminished product yields.

In this thesis, new approaches to engineer chimeric PKSs were explored, each targeting a different aspect of the chimeric system: First the relative contribution of protein–protein and protein–substrate recognition on the turnover of chimeric PKS was assessed, revealing the importance of protein–protein interactions between the acyl carrier protein (ACP) and the ketosynthase (KS) domain in the chain translocation step. Directed evolution experiments followed to optimize the protein–protein interaction across a chimeric interface. Additionally, different junction sites for the generation of chimeric PKSs were compared, showing the ability for recombination without interfering with the chain translocation reaction, and highlighting the use of SYNZIP domains to bridge PKS modules. To optimize chimeric PKSs even further, multipoint mutagenesis of KS domains was established, with positive effects on the activity of chimeric systems.

To support engineering attempts, several structure elucidation techniques were combined with *in silico* modeling to characterize the architecture of a PKS module and the domain–domain interactions within it. Preliminary results show a strong conformational flexibility of the PKS module and the great potential of these techniques to define the multitude of transient interactions in PKS modules.

Abbreviations

6–dEB	6–deoxyerythronolide B
AlphaLISA	Amplified Luminescent Proximity Homogenous Assay
ACP	acyl carrier protein
AT	acyltransferase
ATP	adenosine–5′–triphosphate
AVES	avermectin synthase
B. subtilis	*Bacillus subtilis*
BORS	borrelidin synthase
CoA	coenzyme A
cryo–EM	cryo–electron microscopy
DE	dimerization element
DEBS	6–deoxyerythronolide B synthase
DH	dehydratase
DSZS	disorazole synthase
EDK	enantiomeric diketide
EPOS	epothilone synthase
ER	enoylreductase
E. coli	*Escherichia coli*
FDA	US Food and Drug Administration
FPLC	fast protein liquid chromatography
IPTG	isopropyl–β–D–1–thiogalactopyranoside
ITC	isothermal titration calometry
KR	ketoreductase
KS	ketosynthase
LC–MS	liquid chromatography–mass spectrometry
LDD	loading didomain
LIPS	lipomycin synthase
K_D	dissociation constant
KIRS	kirromycin synthase
MatB	malonyl–CoA synthetase from *Streptomyces coelicolor*
MONENS	monensin synthase
MYCL/MYCS	mycolactone synthase
NADPH	β–nicotinamide adenine dinucleotide 2′–phosphate
NDK	natural diketide
NIDS	niddamycin synthase
NMR	nuclear magnetic sesonance
OLEAS	oleandomycin synthase

ori	origin of replication
PCR	polymerase chain reaction
PIKS	picromycin synthase
PKS	polyketide synthase
PKSE DynE8	enediyne synthase
PrpE	propionyl–CoA synthetase from *Escherichia coli*
PFU	plaque–forming units
RAPS	rapamycin synthase
R_g	radius of gyration
RIFS	rifamycin synthase
SAXS	small–angle X–ray scattering
SCME	methylmalonyl–CoA epimerase from *Streptomyces coelicolor*
SEC	size exclusion chromatography
S. erythraea	*Saccharopolyspora erythraea*
SDS–PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
Sfp	phosphopantheinyl tranferase from *Bacillus subtilis*
SNAC	*N*–acetylcysteamine thioester
SPNS	spinomycin synthase
TCEP	tris(2–carboxyethyl)phosphine
TYLS	tylactone synthase
TE	thioesterase
UV–Vis	ultraviolet–visible spectroscopy
VINS	vicenistatin synthase
XL–MS	cross–link mass spectrometry

Contents

1 Introduction

1.1 Natural Products for Drug Development

Natural products, such as polyketides, non–ribosomal peptides, terpenes, and alkaloids, are secondary metabolites produced by many bacteria, fungi, and plants.[1] They are characterized by their high biological activity and their potential as leads in drug discovery.[2] Important examples of natural product–derived drugs are the anti–inflammatory drug acetylsalicylic acid, the opiate morphine, anti–malaria drugs quinine and artemisinin, antibiotics such as penicillin or erythromycin, anti–cancer drugs such as paclitaxel or doxorubicin, or the immunosuppressant rapamycin.[1,3] In 2005, natural products or natural product–derived compounds made up 50% of the drugs in clinical usage.[4] Their relevance as lead structures is furthermore evident from the large number of newly approved compounds and drugs in clinical trials: E.g. from 1981–2010 34% of the US Food and Drug Administration (FDA) approved small molecule drugs were natural products or derivatives thereof,[5] and in the five year span from 2008 to 2013, 25 new natural product–based drugs were approved with another 100 candidates in clinical trials at the end of 2013.[6]

The inherent advantage of natural products is their evolutionary developed ability to interfere with critical cellular functions in their target organism(s), often accompanied by cellular permeability due to them being substrates for cellular transporter systems.[5] In contrast to many chemically synthesized compounds,

they are characterized by an increased number of sp^3–hybridized carbons and chiral centers, fewer aromatic rings, large macrocyclic aliphatic rings, low nitrogen and high oxygen content, which overall leads to more complex three dimensional structures.[3] This allows natural products to interact more efficiently with biological targets than less challenging compounds.[7] They are often referred to as "privileged structures" due to their evolutionary optimized ability to bind biomolecules and thus high potential to serve as promising starting points in drug development.[8]

Despite these major benefits, many pharmaceutical companies have decreased the use of natural products as lead structures for drug discovery and instead favor synthetic compounds.[3,4,9] The main disadvantages of natural products are their limited availability, as only small quantities can often be purified,[4] the challenging purification procedure involved,[3] and the high likelihood of re–identifying known compounds from novel bioactive extracts.[5] As pharmaceutical companies started to focus on high–throughput screening of large libraries, natural products became less interesting targets.[5] In contrast to the industrial trend, basic scientific studies on natural product biosynthesis have increased in number, extending our knowledge about the mechanism and engineering potential of natural product biosynthesis. Additionally, the recent explosion in genome sequencing revealed many unknown gene clusters with biosynthetic potential.[3] Besides fragment–based library design, which aims at developing new lead structures from basic fragments using chemical synthesis,[8,10,11] natural products and their derivatives are becoming more relevant again in drug discovery due to the greater usage of functional assays and phenotypic screens.[5,7]

One of the main driving forces of this development is the increasing number of resistances against common antibiotics,[12,13] which necessitates the discovery or development of new antibiotic drugs. As many polyketide products exhibit antibiotic activity, research on mechanism, structure, and engineering–potential of polyketide biosynthesis is of great interest in this regard. While discovery of

novel enzymatic pathways via genome mining is a valuable approach, it involves intensive groundwork to produce and purify the enzyme cluster and establish enzyme functionality. Engineering of existing biosynthetic pathways on the other hand allows for the generation of novel natural compounds by imposing some of the aspects of synthetic chemistry onto a known polyketide producing system. The newly generated compounds still comprise the properties of being produced by natural systems but sample a larger chemical space than the original product. This PhD thesis presents new insights into the feasibility of engineering strategies targeting an important model polyketide synthase, the 6–deoxyerythronolide B synthase (DEBS) with the goal of changing its product spectrum. The following section will introduce core concepts of PKSs and their intriguing complexity, before moving to aspects and challenges of PKS engineering.

1.2 Polyketide Biosynthesis

Polyketide synthases occur in bacteria, fungi, and plants and are responsible for the synthesis of polyketide natural products; among them many compounds with high bioactivity, such as antibiotics (e.g. erythromycin, rifamycin, and tetracyclin), anti–cholesterol compounds (e.g. lovastatin), environmental toxins (e.g. aflatoxin), immunosuppressants (e.g. rapamycin), apoptosis inducer (e.g. apoptolidin), and antineoplastics (e.g. daunorubicin; Figure 2).[14–16] PKSs use simple acyl–CoA building blocks and assemble them into complex compounds with molecular weights of up to several kilodaltons (kDa). It is fascinating that molecules of such high complexity as polyketides, which are very challenging to synthesize via organic synthesis in the lab, are generated by a remarkably simple process *in vivo*; the controlled stepwise condensation of small acyl–CoA substrates, and their subsequent chemical modification.
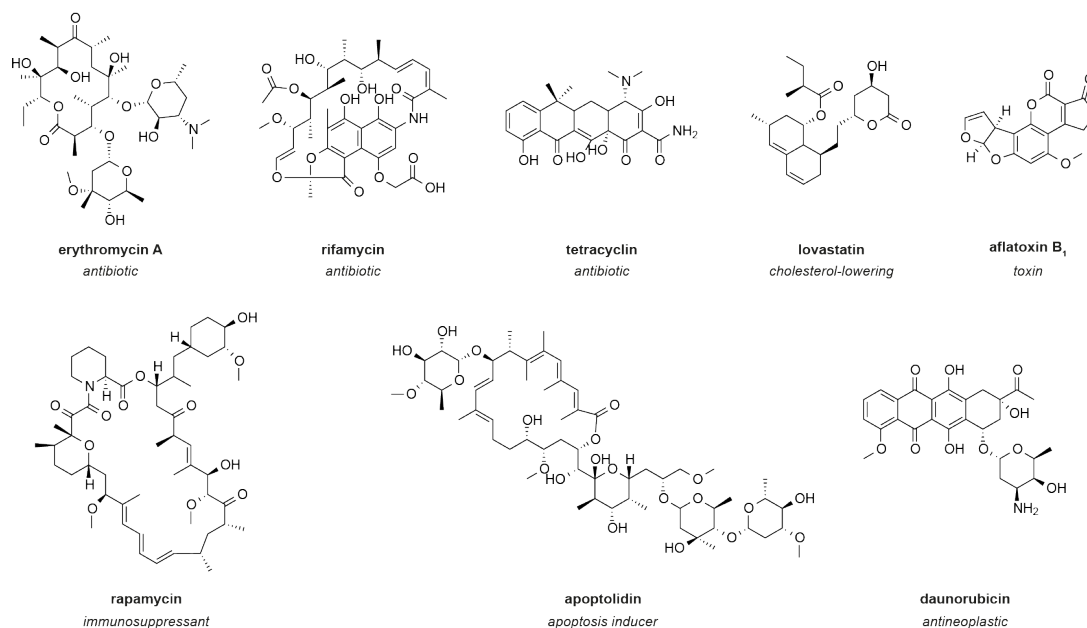
Figure 2 Examples of polyketide secondary metabolites and their biological function.

Based on their architecture and mode of action, PKSs can be classified into three different types.[14,17] PKSs can occur as type I systems (with similarity to type I fatty acid synthases (FASs)), in which catalytic domains are covalently fused to form large multidomain proteins, or as type II systems in which each catalytic domain is present on a separate protein (similar to type II FAS). Both type I and type II enzymes rely on acyl carrier proteins (ACPs) to shuttle the polyketide intermediate to the catalytic domains. Additional, multifunctional type III PKSs, also known as chalcone synthases, exist which act independently of ACPs.[18]

Type I PKSs can be further divided into iterative and non–iterative PKSs. Iterative type I PKSs perform synthesis in a recursive manner, during which the catalytic domains of a single polypeptide repeatedly condense acyl–CoA precursor units until the specific length of the compound is reached (e.g. lovastatin synthase). In contrast, non–iterative type I PKSs (or modular PKSs) occur as large assembly line–like complexes and successively condense precursor building blocks to the final natural compound. Furthermore, modular type I PKSs occur as *cis*–AT and *trans*–AT PKSs.[19] The focus of this thesis lies on non–iterative/modular type I

systems, therefore, if not indicated otherwise, the term PKSs is used for referring to type I *cis*–AT PKSs in this thesis

Since its genetic discovery almost three decades ago,[20,21] the 6–deoxyerythronolide B synthase (DEBS, Figure 3A), which produces 6–deoxyerythronolide B, the aglycone precursor of the antibiotic erythromycin, has served as the prototypical example to study a large pool of naturally occurring assembly line PKSs. DEBS consists of three polypeptides (DESBS1–3), each harboring two modules for chain elongation, with the addition of a loading didomain (LDD) and thioesterase (TE) at the N–/C–terminus of DEBS1/DEBS3. Each module comprises a set of essential domains for C–C bond formation, namely the catalytic domains ketoacyl synthase (KS), acyl transferase (AT), and the non–catalytic ACP domain. Within a module the collaborative reactions of the AT, ACP, and KS domains result in a non–reduced polyketide chain. Optionally, further processing domains (ketoreductase (KR), dehydratase (DH), and enoylreductase (ER)) are employed to catalyze the stepwise reduction of the β–keto group (Figure 3B). The flexible ACP domain shuttles the polyketide intermediate both within a module and to the next (downstream) module. Notably, the ACP reacts with both its cognate KS domain (chain elongation, Figure 3B, step II) and the downstream KS domain (chain translocation, Figure 3B, step IV). Physical linkage between individual modules is either achieved through covalent ACP–KS linkers or non–covalent linker domains (docking domains) at their C– and N–termini.[22,23]

These simple acyl–CoA condensation reactions, and the strict correlation of the assembly line architecture of PKSs and the final polyketide structure, stimulate efforts to engineer PKSs, preferentially by swapping entire modules or single domains, to generate new polyketides.[28–30] Recently, a broadly applicable mix–and–match strategy for engineering the related family of non–ribosomal peptide synthetases (NRPSs) has been presented,[31] which uses defined exchange units as interchangeable synthetic modules. However, attempts to establish similarly
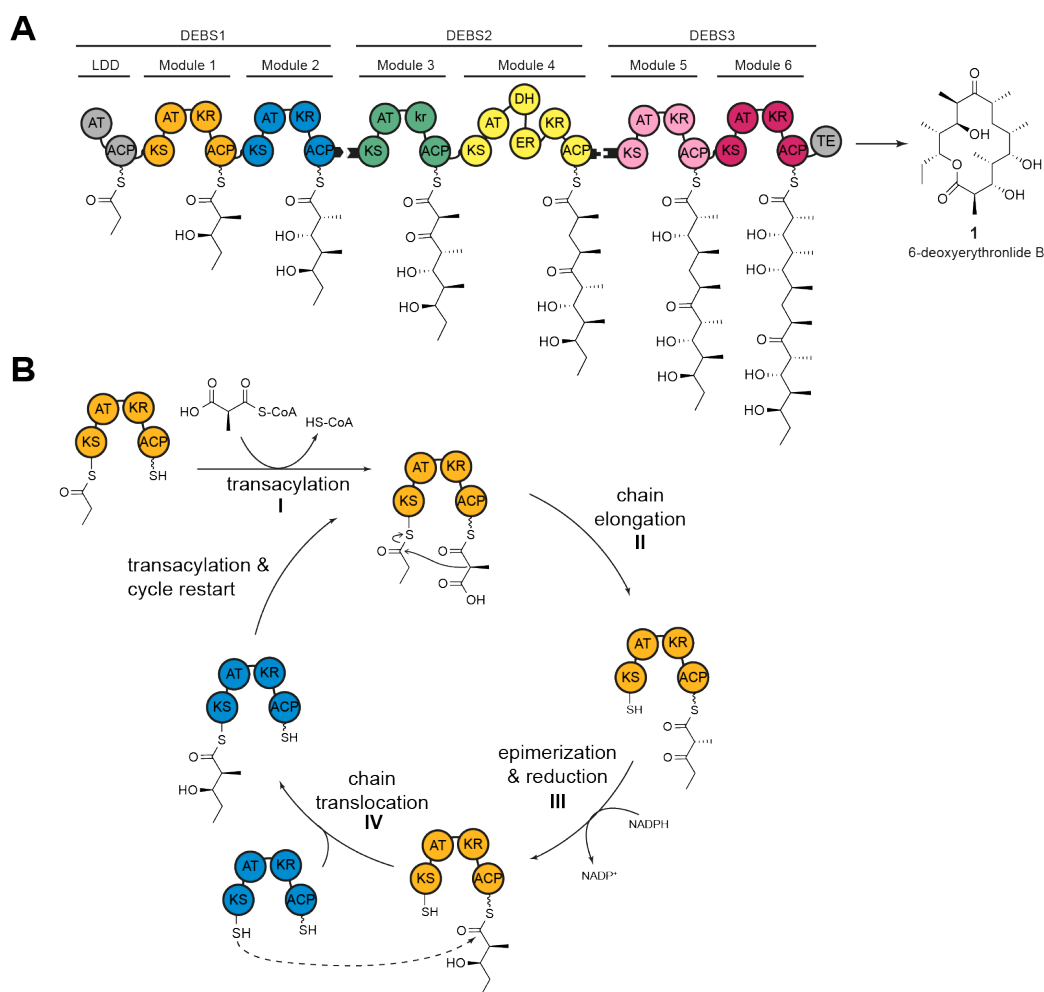
Figure 3   Schematic architecture of DEBS and its catalytic cycle. (A) The three polypeptides (DEBS1–3), the encoded modules (M1–M6), as well as the loading didomain (LDD), the thioesterase domain (TE), and the final product (**1**) are depicted. Polyketide intermediates are shown as attached to the respective acyl carrier protein (ACP). Black tabs depict docking domains. Domain annotations are as follows: LDD – loading didomain, AT – acyltransferase, KS – ketosynthase, ACP – acyl carrier protein, KR – ketoreductase, DH – dehydratase, ER – enoylreductase, TE – thioesterase. Module 3 has a ketoreductase–like domain (denoted in lowercase) that lacks NADPH–dependent oxidoreductase activity but harbors C–2 epimerase activity.[24,25] (B) Catalytic cycle on the example of M1 and M2. (I) AT–catalyzed transacylation using (2*S*)–methylmalonyl–CoA, followed by KS–mediated decarboxylative Claisen condensation (chain elongation, II) to give the (2*R*)–diketide[26]. In M1 the KR catalyzes the epimerization of the C2 methyl group, followed by diastereospecific reduction (III) to give the (2*S*,3*R*)–2–methyl–3–hydroxy–diketide condensation product. The elongated diketide is translocated to the KS domain of the downstream module (chain translocation, IV), from which the next round of reaction can start. Note that in DEBS only KR1 and KR3 catalyze epimerization of the C2 methyl group.[27]

broad strategies to harness the modularity of PKSs in engineering have been unsuccessful. The extensive scaffold of modular PKSs comprises a wealth of permanent domain–domain and module–module interactions and further samples a multitude of transient domain–domain interactions during ACP–mediated substrate shuttling. This complexity and our limited understanding of structural and functional features limits the use of PKSs in protein engineering.

In the following section, I will present a brief review of our current knowledge about the structural organization of a PKS module, the underlying protein–protein interactions, and different engineering strategies with regards to their ability to retain the structural scaffold.

1.3 The Structure of Modular Polyketide Synthases

While our understanding of the architecture of whole PKS modules and their arrangement into assembly lines is still limited, we have insight into structures of individual domains, didomains, and evolutionary related proteins. Recent progress in the characterization of the transient protein–protein interactions during the catalytic cycle of PKSs and an improved understanding of the vectorial synthetic progress further helps to guide precise engineering approaches.

1.3.1 The PKS Structural Scaffold

Analysis of the structural arrangement of PKS modules was enabled by the early structural characterization of the individual subdomains. The earliest X–ray structural information on PKS domains and didomains were obtained from studies on the DEBS KS–AT condensing unit,[32,33], KR,[34], DH,[35], ACP,[36] and docking domains.[22] These studies significantly advanced our structural understanding (and thus engineering approaches) by defining domain boundaries and possible

cleavage sites between domains.[37] As most domain–domain interactions were not captured however, the need remained to understand how these domains interact to form the scaffold of an intact PKS module.

As PKSs are evolutionary related to the mammalian fatty acid synthase (mFAS),[38,39] structural parallels can be drawn between these two enzyme families. Whereas in PKSs the minimal domains KS, AT, and ACP are sufficient to produce a polyketide product, in mFAS additional processing domains (KR, DH, and ER) are present to fully reduce the β–keto group to yield a saturated fatty acid product. These additional processing domains are optional in PKS modules and give rise to the more diverse biosynthetic potential as they allow for the formation of non–reduced, partially reduced, dehydrated, and fully reduced moieties at the β–position. As such a fully reducing PKS module (including the domains KS, AT, KR, DH, ER, and ACP) can be considered a close relative to mFAS.

Not suprisingly, when the X–ray crystal structure on the porcine mFAS was revealed in 2008, it gave new insights into mechanisms and spurred significant interest in the field of modular PKS (Figure 4A, mFAS).[40] Many subregions from mFAS and modular PKSs displayed structural consensus, including the non–catalytic KS–AT linker domains (KS–AT linker and post–AT linker) and KR (the non–catalytic ψKR) domain. This comparison led to an appealing model that assumed that mFAS is simply a fully reducing PKS, which can be transformed to partially reducing and non–reducing PKSs via deletions of modifying domains within the processing wing. Recent studies however call this interpretation in question, pointing out important structural differences between the processing parts of PKSs and mFAS despite the generally tight structural relationship between the two enzyme classes. Particularly, the first almost complete X–ray structural model of the iterative *Mycobacterium smegmatis* mycocerosic acid synthase (MAS)–like PKS gave important insight,[41] disagreeing with earlier models based on separate proteins and the KR–ER didomain from the spinomycin synthase (SPNS) which
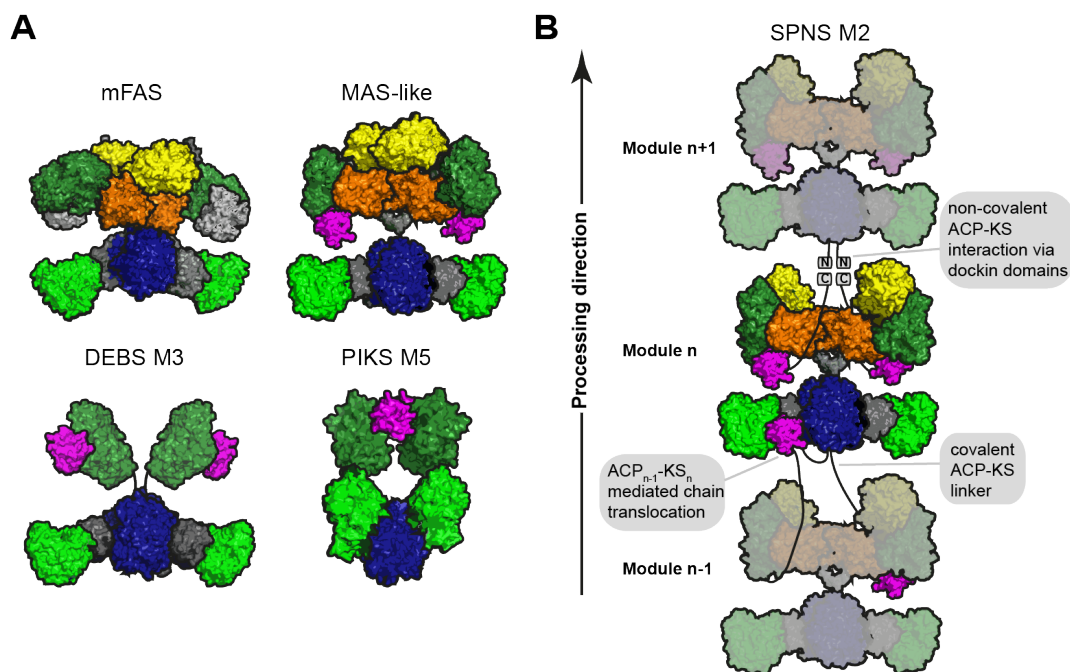
Figure 4 Different models for assembly line PKSs in comparison to mFAS. (A) Schematic representation of mFAS (based on PDB 2VZ8),[45] a MAS–like PKS (based on PDB 5BP1 and 5BP4),[41] a model for the partially reducing DEBS M3 based on SAXS analysis,[46] and a model of the partially reducing pikromycin synthase module 5 (PIKS M5) based on cryo–EM analysis.[47] (B) Graphical representation of a PKS module assembly using the example of the proposed model for the spinomycin synthase module 2 (SPNS M2).[42] The SPNS M2 model was assembled from PDBs 5BP1 (KS–AT), 3SLK (ER–KR), and 5BP4 (DH) with KR of 3SLK superimposed on KR of 5BP4. PKS modules can either be connected through covalent ACP–KS linkers or through non–covalently interacting docking domains at the C– and N–termini. Additionally, chain translocation is mediated through specific interactions between the upstream ACP (ACP_{n-1}) and the downstream (KS_n), shown for one of the two ACPs in module$_{n-1}$. Domain coloring: KS (blue), AT (light green), KR (dark green), DH (orange), ER (yellow), ACP (magenta), linker regions such as the KS–AT linker (dark gray), and only for mFAS non–catalytic pseudo–methyltransferase domain (light gray).

provided major contributions to the structural understanding of PKSs.[34,35,42–44] Based on these studies a MAS–like and a SPNS model emerged for the appearance of a PKS module in solution. The difference and significance of these two models are discussed below.

The DH domain is most important in the different organizations of the processing part of mFAS and PKS. In contrast to a V–shaped arrangement in mFAS, DHs of PKSs form an overall elongated dimeric arrangement inducing a different relative positioning of the modifying domains.[35,41] The embedment of the ER domain in the PKSs fold is currently disputed. The ER is dimeric in the iterative MAS–like PKS, resting upon the extended β–sheet of the DH dimer by forming a small

and variable interface (Figure 4A, MAS–like model). The relevance of a dimeric ER arrangement in modular PKSs is unclear,[48] since the ER domains from PKSs, structurally characterized by X–ray crystallography as individual proteins[44] and as part of a KR–ER didomain construct from SPNS[42] appeared to be monomeric. The ER domain may occur in different oligomeric states in PKSs, i.e. as a monomeric ER in modular PKSs and dimeric ER in iterative PKSs, consistent with ER–KR linkers of modular PKSs being generally shorter (usually <8 amino acids) than those of iterative PKSs (>17). A short linker may restrain ER conformational variability in modular PKSs and prevent ER dimerization at the twofold axis.[48] SAXS data have been reported, which support both dimeric and monomeric ER in modular PKSs.[41,42] Accordingly, two models for the structural appearance of a fully reducing PKS module exist today: A MAS–like PKS model with the ER domains dimerizing at the C2—axis (Figure 4A), and a SPNS M2 derived model, in which monomeric ERs are swung out and thereby free space at the protein's central region (Figure 4B). In order to reach consensus on the appearance of PKS modules, structural characterization of a variety of intact modules is required. With the MAS–like model, detailed structural information on a module from an iterative PKS is available, but since the SPNS model, the model for multimodular PKSs, is assembled from separate domains, the necessity remains to structurally characterize intact modules from multimodular assembly line PKSs. Furthermore, characterization of a variety of PKS modules might shed light on their inherent engineering potential.

Until today, detailed structural information on a complete non–iterative/modular PKS module is missing with the exception of medium resolution (7.3–9.5 Å resolution) cryo–electron microscopic (cryo–EM) data on the partially reducing module 5 of the picromycin synthase (PIKS M5). The PIKS M5 structure revealed an arched conformation,[47] with a relative arrangement of the KS and AT domains that contradicts X–ray crystallographic studies on the mammalian FAS[45,49,50] and PKS KS–AT structures.[32,33,41,51] In all these structures the AT domains were consistently found

to be embedded as bulge of the KS fold framed by a N–terminal KS–AT linker that constitutes most of the KS–AT interface, and a post–AT linker that wraps back to interact with the KS domain.[32] In the PIKS M5 model, this structural integrity is dissolved inducing the overall arched appearance. The PIKS M5 model provides a structural basis that can explain PKS function;[52] however, before the PIKS M5 model can be accepted as a relevant model for the functional mode of modular PKSs, several of its structural features need yet to be reconciled with the large amount of data collected over the last decades.[53,54] A low–resolution model of partially reducing PKS modules was additionally derived from SAXS analysis of DEBS.[46] Herein, a similar architecture of the KS–AT fold was observed as found in mFAS and PKSs. Thus, two different models for partially reducing modular PKS modules can be derived today based on the analysis of PIKS M5 and DEBS modules (Figure 4A).

While several models can be generated for the overall scaffold of a standalone PKS module, data on the architecture of a whole PKS assembly, namely the linkage of several modules together to build the entire PKS systems (e.g. DEBS, Figure 3A), is still missing with the only low–resolution example coming from SAXS analysis of DEBS.[46] In general, in assembly line PKSs, modules can be connected by covalent ACP–KS linkers or through non–covalently interacting, α–helical docking domains (Figure 4B). Several excised docking domains were structurally solved,[22,55–57] but no high resolution structure of a complete module–module interface was obtained so far.

1.3.2 Interactions within Modular PKSs

Given the modular nature of PKSs, controlled protein–protein interaction between domains determine the functional order and therefore specificity in these systems. ACP domains play a pivotal role in this regard, as they are responsible for substrate

shuttling in PKSs and interact with all catalytic domains within the PKS module, but also across module boundaries.

ACPs are generally loosely attached (via unstructured linkers) to the catalytic body of (type I) PKSs and interact transiently with the catalytic domains. Owing to the high conformational variability, ACP domains often remain unresolved in structural studies leading to paucity of details to the process of ACP–mediated substrate shuttling. The primary insight into the mode of substrate shuttling in PKSs was again provided by the homologous mFASs. For example, it was shown that swiveling and swinging motions of the condensing and processing part of mFAS occur and presumably assist the mobile ACP domain in substrate shuttling.[58–60]

Despite being loosely attached, the ACP domains are differently constrained in their conformational space in iterative and modular proteins. While in mFAS and iterative PKSs the ACP is free to move, solely restricted by terminating domains such as TE, the linkage of ACP to the KS of the downstream module constrains its conformational variability in modular PKS. In covalently connected modules the average length of an ACP–KS linker is 18 residues in length (in ER containing modules).[48] For anticipating a mode of action of ACP in assembly line PKSs, the ER arrangement is therefore of decisive impact (see above for the discussion about PKS scaffolds). A monomeric ER would allow a central positioning of ACP and grant intuitive, easy access to each of the integral catalytic domains, whereas an ER dimer would crowd the C2–axis and force the ACPs toward the peripheral lining of the assembly line (Figure 4B).

The pivotal steps for catalytic progress along the PKS assembly line are the ACP:KS mediated chain elongation and chain translocation reactions (Figure 3B, step II and IV). For a successful chain translocation reaction, two adjacent modules need to interact to bring the upstream ACP and the downstream KS into close proximity. While part of this interaction is mediated through covalent linkers or docking

domains, specificity is conferred through the interaction of the ACP and the KS–AT fold itself. Using DEBS as a model system, the ACP:KS interactions were analyzed in a comprehensive study combining *in vitro* analysis of chimeric ACPs and *in silico* docking simulations. Generation of different chimeras revealed that during chain elongation ACP_A (of protomer A) docks with loop I to the KS_A–AT_A linker, while engaging with KS_B (of protomer B) for chain elongation.[61,62] Two distinct recognition sites on the KS–AT linker were found to interact with a minimal epitope in ACP loop I. In a similar manner, the upstream ACP_A docks with the first ten N–terminal residues of ACP helix I to KS_A–AT_A linker to mediate chain translocation with KS_B.[62,63] Although the ACPs dock into the same deep cleft of the KS–AT fragment the position and orientation is distinct in the event of chain translocation and chain elongation.

The example of the ACP:KS interactions highlights the degree of fine tuning involved in ACP:domain interactions and the necessity of preserving these interfaces in protein engineering attempts to ensure catalytic activity of the engineered modules.

1.4 Engineering of Polyketide Synthases

Engineering of PKS systems has been performed at both the module and the domain level. Swapping of entire PKS modules (module–module exchange/mix–and–match approach) is an enticing possibility to create novel PKS assembly lines, while domain–domain exchanges focus on single domains e.g. AT domains or KR domains to specifically alter the product spectrum of a given PKS. In both cases, the resulting non–native protein interactions often entail lower structural stability and concurrent decrease of enzyme activity. Understanding and optimizing protein–protein interactions within the chimeric systems is therefore of prime importance to successfully engineer chimeric PKS systems. The following sections will give an

overview of the two strategies, their successes as well as structural considerations important for engineering of PKSs.

1.4.1 Module–Module Exchanges

The discovery of assembly line PKSs inspired the chemical biological community to harness their modularity for the rapid generation of new natural products.[64,65] One of the biggest challenges in combining intact PKS modules to chimeric PKS assembly lines is to preserve their catalytic integrity. Early on, the necessity of preserving linker regions between adjacent modules to retain functional proteins was noticed.[66–69] Although preservation of native interaction sites helped in generating productive assembly lines, the overall production rates were usually lowered in the engineered systems.

The first truly combinatorial approach of generating bimodular chimeric PKSs, consisting of intact modules from different PKS sources, was carried out in an *in vivo* study recombining 14 modules of 8 PKS clusters.[70] Using docking domains derived from DEBS, a total of 154 bimodular chimeric PKSs were assembled. About 50% of the chimeric PKSs yielded detectable product albeit with lower product yields as their natural bimodular reference system.[70] Later on this approach was extended towards chimeric trimodular PKSs, although with similar results.[71]

Until now, chimeric PKSs have been constructed either by genetic fusions of heterologous modules or by non–covalently connecting individual modules with docking domains derived from natural PKS sources. *In vitro* characterization of chimeric bimodular PKSs using fusion proteins, in which the non–naturally interacting modules were connected by a covalent linker, revealed similar k_{cat} values for wild–type and chimeric PKSs.[72] A more thorough analysis showed that linker regions and specific ACP:KS interactions play equal roles in influencing the turnover rate of chimeric PKSs.[73] As intermodular communication is necessary to generate

productive chimeric PKS assembly lines, questions arise about the affinity of these interactions in native proteins. So far only a few studies have addressed this issue. The dissociation constant of docking domains or modules connected by docking domains was found to be $K_D \approx 70$–$130\ \mu M$[56] and $K_D \approx 1$–$2\ \mu M$[74], respectively. Titration of different DEBS polypeptides and following Michaelis–Menten fitting resulted in K_{50} values 2.5–4 μM, indicating that the interaction efficiency of the naturally interacting polypeptides is rather weak.[75] Overall, efforts in engineering PKSs via module replacement revealed the importance of linker regions (covalent linkers or matching docking domains) to mediate communication between heterologous modules. The nature of the non–native ACP:KS interaction during chain translocation appears as another factor influencing turnover in chimeric PKSs. A mere module–module exchange employing docking domains to facilitate communication across heterologous modules has had limited success. Covalent linking of modules seems to allow the design of catalytically more active PKS chimera, but is done at the expense of modularity.[72] Based on the finding that distinct epitopes on the ACP are responsible for KS–AT recognition during chain elongation and chain translocation,[61,63] mix–and–match strategies may be best realized by adapting both the docking domains and chimeric ACP:KS interfaces for productive chain translocation.

1.4.2 Domain–Domain Exchanges

Another strategy of engineering PKSs for producing novel compounds is the exchange of individual domains. One of the most common ways of altering the final polyketide product is to insert AT domains with altered specificity to incorporate a different starter or extender unit and thus change the entity at the α–carbon. Altering the stereochemistry of the final product can also result in novel biological activities making KR domain exchanges or reductive loop swaps

of particular interest. The other domains (KS, ACP, and DH) were only rarely employed in domain exchanges.

AT Domain Exchanges in Loading Modules

One of the easiest way of inserting a new chemical moiety into a polyketide product is to exchange or mutate the AT domain of the loading module to incorporate a different starting unit. For example, the AT of the loading didomain of DEBS has been exchanged with the respective AT of the avermectin synthase (AVES) loading didomain, which increased the diversity of erythromycin analogs produced *in vivo*.[76] In another study, the complete exchange of loading didomains of the same type (AT–ACP form) of the tylactone synthase (TYLS) into the platenolide synthase was reported.[77] It was even possible to exchange loading domains of different architecture (AT–ACP from DEBS vs. KS_Q–AT–ACP from oleandomycin synthase (OLEAS)).[78] While exchanges of whole loading modules interfere with the ACP:KS interaction during translocation and thus may decrease product yields,[61] a mere AT exchange affects the specific interaction of AT:ACP during transacylation.[79] As both interactions are important for proper turnover, an adaption of the non–native module–module or domain–domain interfaces seems necessary to avoid kinetic penalties.

AT Domain Exchanges in Elongating Modules

The exchange of single extender AT domains within a given PKS system is the most commonly used engineering strategy to alter the polyketide product at a given position. The first AT exchange was carried out in 1996 and similar approaches followed in these early years after PKS discovery.[80–84] In most of these studies, the novel polyketide product was formed, albeit at lower yields. For receiving an initial understanding about the structural integration of extender AT domains

in the overall PKS fold, limited proteolysis experiments were conducted. Those experiments indicated that linker regions especially downstream of the AT domain are important for a successful AT exchange.[85] While it was shown that the post–AT linker is neither required for acylation of the AT domain nor for transacylation from the AT to the ACP domain, its presence is required for KS:ACP catalyzed chain elongation.[37] The impact of linker regions on the overall protein stability became apparent upon obtaining the first high–resolution structures.[32,33] A conserved sequence in the C–terminal part of the post–AT linker folds back onto the KS and an additional interaction between the KS–AT linker and the post–AT linker is observed (Figure 5).[86] Recently, a first systematic analysis of junction sites was conducted by exchanging ATs of DEBS as well as of lipomycin synthase (LIPS).[87] Based on sequence alignments of different AT domains, optimal fusion sites for AT domain exchanges were derived. Best tolerated AT exchanges were achieved by swapping the AT with the adjacent KS–AT linker (KAL) and the N–terminal part of the post–AT linker (PAL1; construct KAL–AT–PAL1).[87] In agreement with previous studies, this strategy is successful, as it preserves the unit of KS with the conserved N–terminal part of the post–AT linker (PAL2$_{cons.}$). Thus, there is good evidence that the best fusion sites for AT domain exchanges lie in the KS–AT and post–AT linker regions. Overall, AT domain exchanges are a good example for how structural information can guide PKS engineering in defining appropriate exchange sites.

Implication of ACP:AT Interactions on PKS Engineering

Only few studies have addressed the impact of non–native ACP:AT interactions on PKS turnover. As the interaction of the ACP with the AT is transient and presumably weak, mapping of the ACP:AT interface has been difficult in the past. Recently, the first X–ray structural atomic model of an ACP:AT complex from the *trans*–AT vicenistatin synthase (VINS) revealed an interaction of ACP helix II
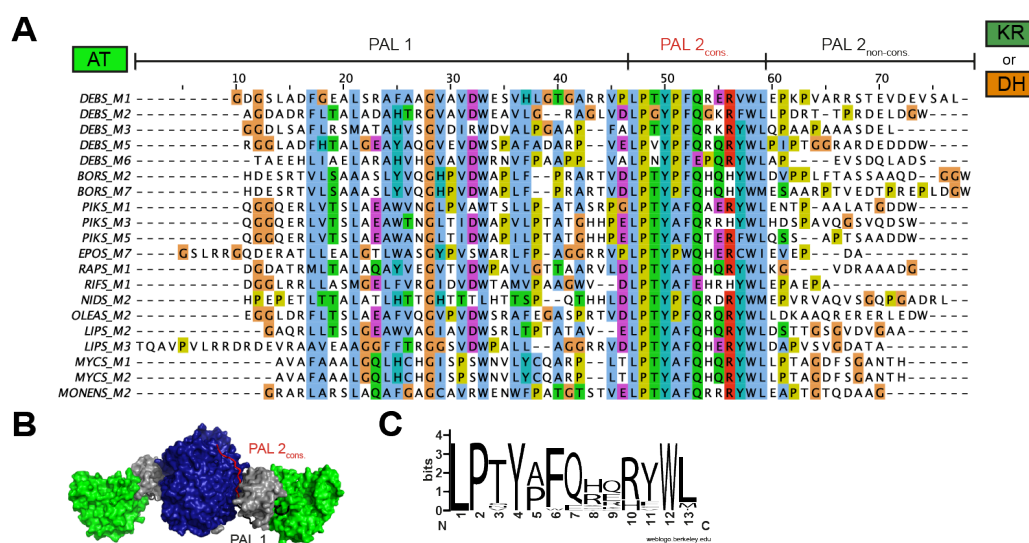
Figure 5 Important interfaces for engineering AT domain exchanges. (A) Sequence alignment of post–AT linkers from partially and fully reducing PKS modules. Important segments of the linker sequence are highlighted. Sequences derived from: DEBS; 6–deoxyerythronolide B synthase, BORS; borrelidin synthase, PIKS; pikromycin synthase, EPOS; epothilone synthase, RAPS; rapamycin synthase, RIFS; rifamycin synthase, NIDS; niddamycin synthase, OLEAS; oleandomycin synthase, LIPS; lipomycin synthase, MYCS; mycolactone synthase, and MONENS; monensin synthase. (B) Structure of DEBS KS3–AT3 (PDB 2QO3) with the post–AT linker highlighted. PAL1 (∼N–terminal 35–45 residues, black) interacts with the AT and the KS–AT linker. The conserved PAL2 sequence (∼13 residues, red) interacts with KS surface residues. The non–conserved part of PAL2 is not solved in structure. Domain coloring as in Figure 4. (C) Sequence logo of PAL2$_{cons.}$ consensus sequence.[88]

with the AT.[89] While this study provided the first high–resolution 3D–structure, previous studies already employed docking simulation and alanine scanning mutagenesis to map the respective ACP–AT interaction interface. Studies of the ACP:AT interaction in the *cis*–AT PKS DEBS,[90] the *trans*–AT disorazole synthase (DSZS),[91] and the iterative acting enediyne synthase (PKSE DynE8)[92] revealed overlapping – yet not identical – interfaces, in which the ACP interacts with the AT via regions on helix I and helix II (and additional loop regions depending on the different models). A combination of docking simulation and alanine–mutagenesis identified residues responsible for the ACP:AT interaction of the *trans*–AT kirromycin synthase (KIRS).[93] As was known from previous studies, residues of ACP helix I, loop I, and helix II were predicted to be involved in AT recognition and based on this knowledge a non–native AT:ACP interaction between the *trans*–AT KirCII and DEBS ACP6 was engineered towards improved transacylation.

Reductive Loop Swaps/KR Exchanges

Ketoreductases are responsible for the NADPH–dependent reduction of the β–keto group and if present also the epimerization of the α–substituent.[94] The specificity of KR domains for their cognate ACPs has been suggested to be relatively low,[95] therefore KR domains may be more tolerant in domain swaps than AT domains.

The KR domain has been successfully exchanged in several cases in partly reducing β–modules, particularly for studying the stereochemistry of the reduction reaction that configures the β–hydroxy functionality,[96,97] and the α–alkyl group in case other units than malonyl–CoA are accepted for elongation.[98] Although KR domains can be generally classified by their reduction and epimerization activity, most KR exchanges do, intriguingly, not achieve to retain both activities in the heterologous context.[98] These engineering successes can be attributed to early availability of high resolution 3D–structures,[34] which allowed for a better understanding of the

KR from a structural point of view. More recently, a dimerization element (DE) was identified that is placed N–terminally to 50% of PKS modules that contain KR as the only processing domain.[43,99a] The DE turned out to be responsible for the stability of the respective KR domains. As a current rule for KR engineering in partially reducing PKSs (KS–AT–KR–ACP), DE domains of the host PKSs should be preserved in KR exchanges, while a DE–KR unit should replace KR in PKSs not natively carrying a DE structural motif.[99,100]

Swaps of complete processing wings (or reductive loops) were among the earliest achievement of PKS engineering, owing the structural integrity of this unit. A similar strategy has led to the successful production of adipic acid. Here, the processing wing of the borrelidin synthase module 1 (BORS M1), which contains only a KR domain, has been replaced with the fully reducing processing wing of the second module of SPNS. The DH domain turned out to be problematic owing to restricted substrate specificity, but when swapped with the DH–domain of BORS M2 the expected product was produced.[101] Not surprisingly, the definition of junction sites for reductive loop insertions underlay the same rules as AT exchanges pointing again at the importance of preserving the KS–PAL2$_{cons.}$ interface.

1.4.3 Non–Rational Engineering Strategies

Besides the aforementioned rational engineering attempts, non–rational engineering can also result in the generation of new polyketides. In nature, homologous recombination is an often used strategy for the generation of diversity. Based on homologous recombination in *S. cerevisiae*, chimeras of PIKS M5 and DEBS3 were successfully generated *in vivo*, of which some had measurable *in vitro* activity indicative of a stable protein fold.[102] Although it was possible to rapidly generate a library of chimeric PKSs, stability of the resulting folds could not be ensured

---

[a]Note that no DEs were found in any of the DEBS modules.

under the experimental conditions. Recently, a similar approach used homologous recombination within RAPS to generate several new rapamycin derivatives with altered biological activity.[103] Interestingly, the best chimeras revealed junction sites in the KS or AT domain, or in the linker upstream of the ACP domain. The approaches use homologous recombination as a tool to mimic evolutionary pressure for receiving chimeric PKSs with sufficient catalytic fitness. Such strategies can identify junction sites that may eventually support rational engineering strategies.

1.5 Aim of the Thesis

The goal of this thesis was to establish novel strategies to engineer chimeric PKSs for the generation of novel natural products. As outlined above, despite the apparent modularity of assembly line PKSs, most engineering attempts have had limited success. Poor understanding of domain–domain interactions has hindered many engineering approaches as chimeric domain–domain interfaces could not be adapted to functionally interact. In addition, the relative contribution of protein–protein and protein–substrate interaction on the turnover rate of chimeric PKSs remained unclear.

Thus as a first goal and as a continuation of my master's thesis project, I analyzed the relative influence of the two aforementioned interactions on the turnover rate of a library of chimeric PKSs (Chapter 2.1). Our results were published in the Journal of Biological Chemistry and major parts of this chapter were adapted from the publication.[104] Based on our findings, I then sought to deepen our understanding of, and to optimize, a chimeric ACP:KS interface by exploring the feasibility of directed evolution experiments to improve a chimeric interface (Chapter 2.2). Another focus of my work was to test the usefulness of synthetic docking domains in the generation of chimeric PKSs, both in replacing traditionally used PKS–derived

docking domains to connect modules and by combining modules via a novel domain–domain interface (Chapter 2.3). Our main results were published in the Journal ACS Chemical Biology and major parts of this chapter were prepared from the publication.[105] As a continuation of this line of experiments, I also attempted to broaden the substrate specificity of KS domains by multipoint site–directed mutagenesis (Chapter 2.4).

As discussed in this introduction, the lack of structural information on overall PKS module organization and shape has hindered many engineering approaches, which is why I employed a combination of structure elucidation techniques (SAXS, cross–link mass spectrometry (XL–MS), and structural modeling) to sample different conformations of a PKS module in solution and gain deeper insight into underlying domain–domain interactions (Chapter 2.5).

1.6 Milestones and Collaborations

This thesis was conducted in close collaboration between Prof. Martin Grininger's laboratory at Goethe University Frankfurt and Prof. Chaitan Khosla's laboratory at Stanford University. While many of the functional studies presented in this thesis were carried out at Stanford University, a focus on structure elucidation attempts was pursued at Goethe University.

The structure–focused project presented in Chapter 2.5 was carried out in collaboration with the Urlaub group at the MPI for Biophysical Chemistry in Göttingen and the Hummer group at the MPI for Biophysics in Frankfurt.

Portions of the text and data in the results section were either already published, prepared for publication, and/or generated by collaboration partners or students under my supervision. Please refer to Chapter 6.1 for a more detailed list of people involved in the work and their contributions. A timeline of milestones accomplished in this thesis is given in Figure 6.
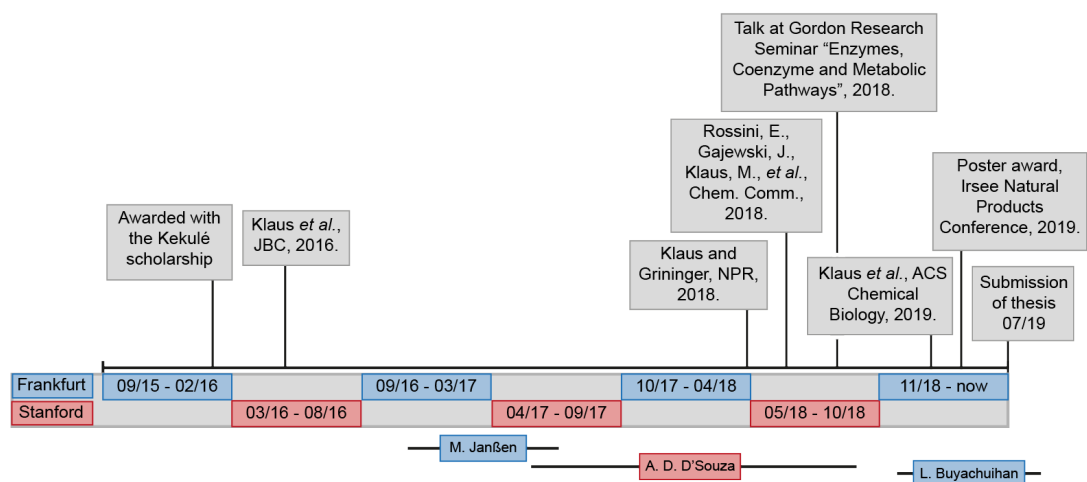
Figure 6 Milestones of this thesis. Publications,[104–107] awards, and talks are indicated, as well as my stays at Goethe University (blue) and Stanford University (red), and the supervision times of students whose work directly contributed to this thesis.

2 Results and Discussion

Over the course of my doctoral thesis I approached the characterization and engineering of PKS modules from different angles with the ultimate goal of creating a functional chimeric PKS system. In particular, my studies provide a detailed analysis of the influence of domain–domain interactions as well as substrate–domain interactions on the activity of chimeric PKSs. In the following chapter I will present the results of these projects and discuss how findings from one project informed our focus and experimental design in others. Early experiments that identified bottlenecks within chimeric PKSs, generated by mixing intact heterologous modules, had a major influence on the shape of this thesis. I started this project during my master's thesis and continued working on it during the early stages of my doctoral thesis (Chapter 2.1, published in JBC). On the basis of these results I explored several distinct engineering strategies to overcome the existing limitations (Chapters 2.2–2.4, the project in Chapter 2.3 published in ACS Chemical Biology) and performed structural characterizations of crucial domain–domain interactions and the overall PKS scaffold to guide engineering attempts (Chapter 2.5).

2.1 Protein–Protein Interactions, not Substrate Recognition, Dominates the Turnover of Chimeric Assembly Line Polyketide Synthases

During my master's thesis I characterized a library of chimeric bimodular and trimodular PKSs to assess the relative influence of protein–protein and enzyme–substrate recognition on the turnover of chimeric PKSs.[108] While the analysis of protein–protein interfaces was mostly complete, experiments to characterize the relative contribution of substrate recognition were conducted during my doctoral thesis. Our results, which I will summarize below, were published in the Journal of Biological Chemistry and served as a starting point for additional engineering attempts pursued over the course of my doctoral thesis work.[104]

Engineering of PKSs via the exchange of intact modules has been a longstanding goal, yet despite numerous efforts, most chimeric PKSs showed significantly decreased product yields in *in vivo* studies.[66,68,70,71] The recent breakthrough in the *in vitro* reconstitution of DEBS,[75] together with a better understanding of the protein–protein recognition interfaces between ACP and KS domains,[61,63] and of KR:ACP specificities, enabled us to quantitatively investigate, *in vitro*, the influence of both protein–protein and protein–substrate recognition on the turnover of chimeric PKSs and to identify bottlenecks along the assembly line.

2.1.1 Activity of Chimeric Bimodular and Trimodular PKSs

In this project, I investigated the turnover efficiency of chimeric PKSs that were generated via *in vitro*–recombination of intact modules from heterologous PKS systems. The libraries of both bimodular and trimodular PKSs were constructed similar to previously published *in vivo* characterizations.[70,71]
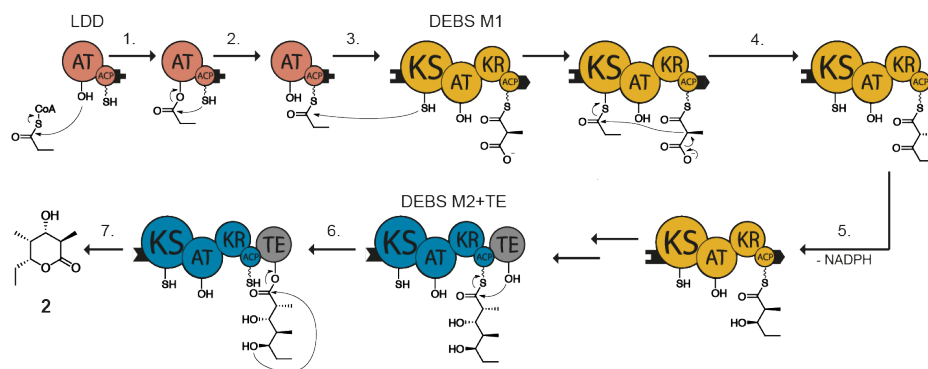
Figure 7 Catalytic cycle of the bimodular DEBS derivative frequently referred to in this thesis. This mini assembly line is comprised of three proteins: the LDD (shown in red), DEBS module 1 (M1, shown in yellow), and DEBS module 2 fused to the TE domain (M2+TE, shown in blue/gray). The acyltransferase (AT) of the LDD specifically transfers the propionyl moiety of propionyl–CoA (step 1) to the terminal thiol of the phosphopantetheinylated ACP (step 2) of the LDD. This primer unit is then translocated by acylation of the active site Cys–SH of the KS domain of DEBS M1 (step 3). Meanwhile, the ACP of DEBS M1 is loaded with a methylmalonyl extender unit by the action of the acyltransferase domain of DEBSM1. KS–catalyzed chain elongation by decarboxylative Claisen condensation yields an ACP–bound β–ketoacyl–diketide intermediate (step 4). In DEBS M1, the KR domain then catalyzes an epimerization of the C–2 methyl group followed by diastereospecific reduction (step 5) to give the mature (2*S*,3*R*)–diketide, which is then translocated to DEBS M2 via a thioester to thiol transacylation. There it undergoes another round of chain elongation and KR–catalyzed reduction (without epimerization), followed by TE–catalyzed release and lactonization (steps 6 and 7).

Eleven chimeric bimodular derivatives of the DEBS–derived bimodular PKS, consisting of LDD(4), (5)M1(2), and (3)M2+TE (mechanism in Figure 7),[b] were constructed in which DEBS M2 was replaced by modules from DEBS that usually do not interact with DEBS M1 or alternatively by modules from the rifamycin synthase (RIFS)[109] or the rapamycin synthase (RAPS[110]; Figure 8A). To facilitate intermodular interactions between heterologous modules, docking domains were installed at the chimeric interface. These α–helical domains mediate weak but specific non–covalent interactions between donor and acceptor modules ($K_D \sim 100$ μM).[56] In a similar manner eight trimodular chimeric PKSs were constructed based on the DEBS–derived trimodular PKS, consisting of LDD(4), (5)M1(2), (3)M2(2), and (3)M3+TE, in which DEBS M2 was replaced by heterolgous (3)Module(2) constructs.[104]

---

[b]Docking domains are denoted as numbers in brackets and indicate the origin of the matching domains, e.g. LDD is fused to the docking domain of module 4 ("LDD(4)") which interacts with the docking domain from module 5 "(5)".
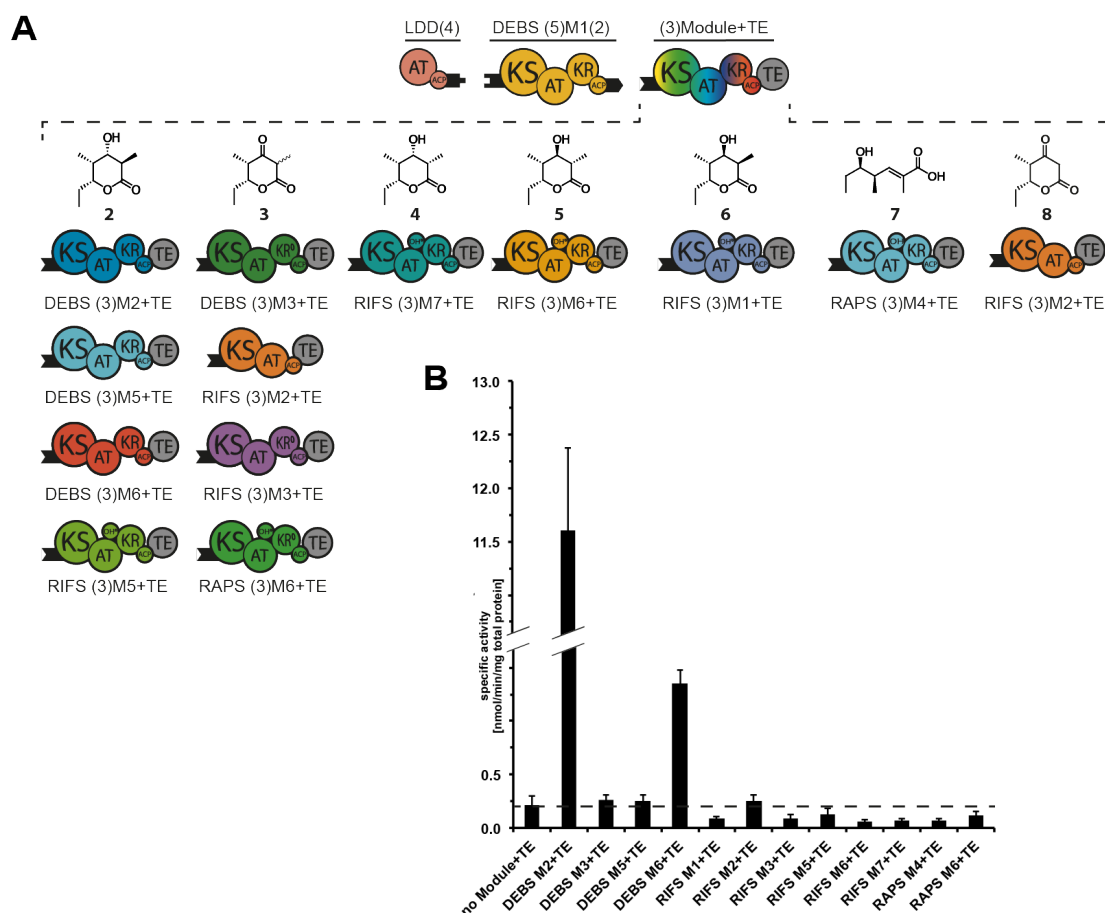
Figure 8 Design and turnover analysis of chimeric bimodular PKSs. (A) Design of chimeric bimodular PKSs. Each PKS included LDD(4) and DEBS (5)M1(2) in combination with (3)Module+TE as the variable acceptor. Acceptor modules were derived from DEBS, rifamycin synthase (RIFS), or rapamycin synthase (RAPS). Compatible docking domains from DEBS (depicted as black tabs or numbers in parentheses) were fused to the corresponding C– and N–terminal ends of the respective donor and acceptor modules to enhance the specificity and efficiency of intermodular chain translocation. The predicted triketide products **2–8** generated in the presence of propionyl–CoA, methylmalonyl–CoA (and malonyl–CoA, in the case of RIFS M2+TE), and NADPH, are shown for each chimeric module pair. All acceptor modules are specific for methylmalonyl-CoA, except for RIFS M2, which prefers malonyl–CoA but can also accept methylmalonyl-CoA. (B) Turnover rates of chimeric bimodular PKSs. All initial rate data were obtained at individual PKS protein concentrations of 4 µM and non–limiting concentrations of propionyl–CoA, methylmalonyl–CoA, and NADPH. In assays containing RIFS M2+TE, malonyl–CoA was also included because this module prefers malonyl extender units, although exclusion of malonyl–CoA did not affect the turnover rate of this system. Dashed lines indicate the threshold rate of NADPH consumption in the absence of the chimeric module. Error bars indicate averages of two measurements (each performed in triplicate) obtained with independent protein preparations.

Analysis of the turnover rate of these chimeric systems was performed using a previously developed UV_{340} spectrophotometric assay, which measures the polyketide formation rate by quantifying the amount of consumed NADPH.[75] For the bimodular PKSs, the reference, using DEBS M2+TE, showed a steady–state rate of triketide formation rate of 11.6 nmol/min/mg total protein, in good agreement with that previously reported for this system.[75] All the other chimeric bimodular assembly lines exhibited significantly reduced rates of turnover, the highest being the one using DEBS M6 as the surrogate acceptor module (1.4 nmol/min/mg total protein; Figure 8B). Background NADPH consumption was measured in a control reaction containing only LDD(4) and (5)DEBS M1(2) in the absence of a downstream acceptor module (Figure 8B, dashed line). Only the chimeras derived from DEBS M3, DEBS M5, DEBS M6, and RIFS M2 showed detectable activity above this threshold background. A similar decrease in polyketide synthase activity was observed for the chimeric trimodular PKSs.[104]

LC–MS product analysis after overnight incubation revealed that 6 of the 11 chimeric bimodular PKSs synthesized detectable quantities of their expected product. Likewise, 4 of 8 chimeric trimodular PKSs yielded detectable amounts of the expected tetraketide.[104] Indicating that they were in principle capable of performing the enzymatic reaction, but with severely diminished activities.

2.1.2 Role of ACP:KS Interactions at the Fusion Site in Chimeric PKSs

Given the greatly reduced activity of the chimeric systems, we sought to identify the specific step during the biosynthesis at which the growing polyketide chain might have become stalled. For this purpose, radio–SDS–PAGE analysis was performed using [^{14}C]–propionyl–CoA to prime the LDD. By quantifying the extent of labeling of each protein module in the presence of methylmalonyl–CoA and NADPH, the site of polyketide accumulation could be deduced. To avoid

hydrolysis of the polyketide product, the TE domain of each acceptor module was inactivated by active site Ser to Ala mutations.

As listed in Table 2, the acceptor module of the reference system (DEBS M2+TE^0) was rapidly labeled, reaching saturation within 3 min. In contrast, the single turnover occupancy level of the acceptor modules of representative chimeric bimodular PKSs was low, which correlated with their catalytic turnover determined earlier (compare Figure 8B and Table 2). Only DEBS M6+TE^0 showed significant labeling, reaching 42% of the reference system after 3 min. Other acceptor modules (DEBS M3+TE^0, RIFS M2+TE^0, RAPS M4+TE^0 , and RAPS M6+TE^0) showed negligible levels of diketide occupancy on this time scale. This finding indicated a rate–limiting effect in the chain translocation step, between ACP1 and the chimeric, downstream KS domain. To ascertain the cause–effect relationship between ineffective chain translocation and turnover, we sought to enhance the rate of chain translocation in a representative chimeric PKS via site–directed mutagenesis of the ACP domain of DEBS M1.

Table 2 Quantification of the occupancy of acceptor modules in chimeric bimodular PKSs. All bimodular PKSs included DEBS LDD and M1 plus the designated acceptor module. Only labeling data for the acceptor module are shown, since the upstream proteins are the same in all cases. The TE domain of each acceptor module was inactivated by site–directed mutagenesis in order to abolish TE–catalyzed triketide chain release and consequent multiple turnover from this module. Labeling of each acceptor module (2 µM) was quantified at 1 min and 3 min in the presence of both LDD (2 µM) and M1 (2 µM). Each measurement was performed in duplicate. The entire experiment was repeated with independently prepared proteins to verify the above data trends.

	Normalized area counts	
Acceptor module	after 1 min	after 3 min
DEBS M2+TE^0	0.83 ± 0.11	1.00 ± 0.08
DEBS M3+TE^0	< 0.02	< 0.02
DEBS M6+TE^0	0.21 ± 0.07	0.42 ± 0.17
RIFS M2+TE^0	0.04 ± 0.01	0.08 ± 0.00
RAPS M4+TE^0	< 0.02	< 0.02
RAPS M6+TE^0	< 0.02	0.05 ± 0.00

Previously, an epitope on ACP helix I was suggested to play an important role in interacting with a downstream KS during chain translocation.[63] Based on this

observation, we predicted that chain translocation from DEBS M1 to M3 in a chimeric bimodular PKS would be enhanced by introducing the E23K mutation (corresponding to E1424K, based on whole module numbering) into helix I of ACP1, because a cationic residue is also found at the corresponding position of ACP2. Indeed, turnover of the E23K mutant of DEBS M1+M3+TE was increased more than 2–fold over the corresponding bimodular PKS derived exclusively from wild–type modules (0.6 ± 0.05 nmol/min/mg total protein, compared with 0.3 ± 0.05 nmol/min/mg total protein).

Taken together, these results underscore the pivotal role of ACP:KS interactions during translocation of the growing polyketide chain between heterologous PKS modules.

2.1.3 Role of Substrate–KS Recognition at Fusion Junctions in Chimeric PKSs

Having examined the influence of ACP:KS interactions on chain translocation at chimeric PKS junctions, we sought to interrogate the relative contribution of substrate–KS recognition on the turnover efficiency of these systems. To this end we rationalized that presentation of the enantiomer of the native (2*S*,3*R*)–2–methyl–3–hydroxy–diketide–ACP product of DEBS M1 to the downstream modules of selected chimeric bimodular PKSs depicted in Figure 8 would enable a controlled comparison of the relative importance of specific protein–protein and protein–substrate recognition. As it was recently shown that the KR domain of DEBS M2 has comparable specificity for ACP1 as for its cognate ACP2 domain,[95] we engineered DEBS M1 by replacing its endogenous KR domain with the paralogous KR2 from DEBS M2. The resulting hybrid module, DEBS (5)M1–KR2(2), was produced as a soluble protein (2 mg/liter in *E. coli*) with purification behavior on an anion exchange chromatography column that was indistinguishable from most wild–type PKS modules.
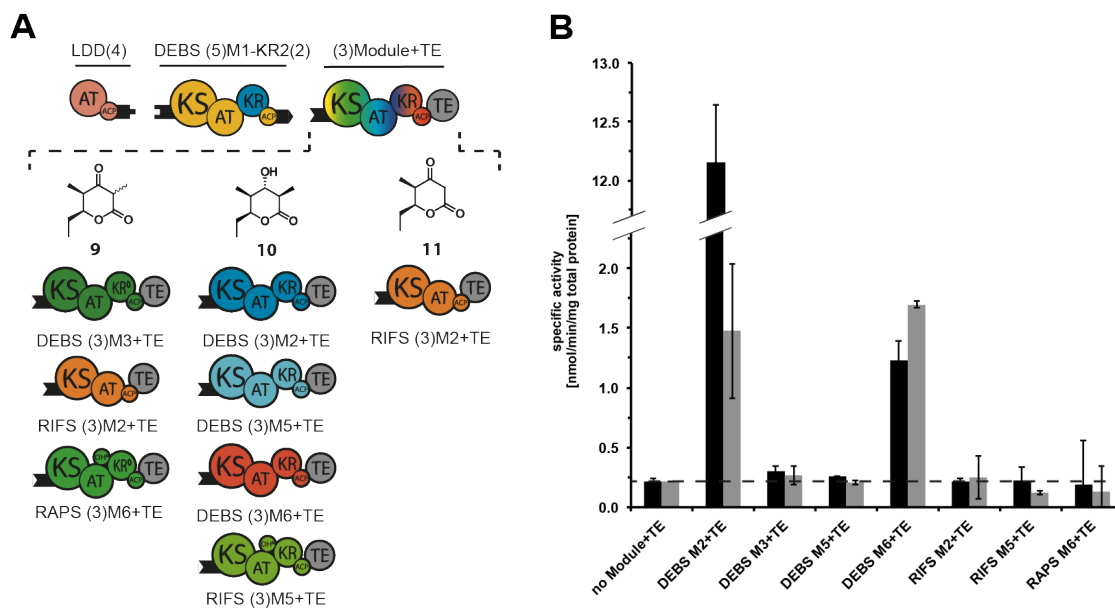
Figure 9 Influence of substrate–KS recognition on chimeric PKSs. (A) Chimeric bimodular PKSs harboring the DEBS M1–KR2 mutant. Each PKS included LDD(4) and DEBS (5)M1–KR2(2) in combination with (3)Module+TE as the variable acceptor. The predicted triketide products **9–11** generated in the presence of propionyl–CoA, methylmalonyl–CoA (and malonyl–CoA, in the case of RIFS M2+TE), and NADPH, are shown for each chimeric module pair.[c] (B) Turnover rates of bimodular constructs consisting of DEBS LDD, DEBS M1 plus the designated acceptor module (black bars) and turnover rates of the same systems with DEBS M1–KR2 mutant (gray bars) in place of DEBS M1. All initial rate data were obtained at individual PKS protein concentrations of 4 µM and non–limiting concentrations of propionyl–CoA, methylmalonyl–CoA (and malonyl–CoA, in the case of RIFS M2+TE), and NADPH. The dashed line indicates the threshold rate of NADPH consumption in the absence of the chimeric module. Measurements were performed in triplicate.

In addition to evaluating the turnover efficiency of this hybrid module in the presence of DEBS LDD(4) and (3)M2+TE, we also measured turnover in the presence of six other acceptor PKS modules (Figure 9). Consistent with the previously established 7–fold preference of DEBS M2 for its natural (2*S*,3*R*)–diketide substrate (NDK) over the enantiomeric (2*R*,3*S*)–diketide analog (EDK),[72] the bimodular construct harboring the hybrid donor module showed a comparable 7–fold reduction in turnover rate compared with its wild–type bimodular counterpart (Figure 9B). Remarkably, none of the other bimodular PKSs containing heterologous downstream modules was significantly affected by the change in the configuration of the diketide intermediate (Figure 9B), indicating that enzyme–substrate recognition is not as important as specific ACP:KS protein–protein recognition during intermodular polyketide chain translocation.

2.1.4 Conclusion — Chain Translocation is a Major Rate–Limiting Step in Chimeric PKSs and Affected by ACP:KS specificities

Combinatorial assembly of modules from different assembly line PKSs is a powerful strategy for complex molecule biosynthesis. The relative importance of protein–protein interactions and enzyme–substrate recognition in such engineered systems however remained unclear. During this project, I addressed these issues for the first time in a systematic, carefully controlled manner.

Similar to other *in vivo* studies,[70,71] the libraries of both chimeric bimodular and trimodular PKS, analyzed in this project, revealed severely diminished turnover rates. Through my analysis, chain translocation across the heterologous interface emerged as the major rate–limiting step in chimeric PKSs (Table 2). Furthermore, using a combination of site–directed mutagenesis and analysis of substrate–tolerance, we determined that the protein–protein interaction at the heterologous interface, namely the ACP:KS interaction, plays a pivotal role in influencing the activity of chimeric PKSs.

Our results shed light on how engineering attempts via the recombination intact PKS modules should be approached, namely by first overcoming rate–limitations imposed by impaired ACP:KS recognition during chain translocation at the junction of heterologous modules, and by secondly (if necessary) optimizing the KS substrate tolerance. In addition it became apparent that site–directed mutagenesis can be used to optimize an ACP:KS interface, although screening of a library of mutants requires more sophisticated techniques.

This insight has served as a starting point for the novel engineering strategies pursued in Chapter 2.2–2.4, and as the ACP:KS interactions are not structurally characterized prompted us to start the structure elucidation project outlined in Chapter 2.5.

2.2 Engineering of a Chimeric ACP:KS Interface via Directed Evolution

As outlined in the introduction, numerous studies have tried to engineer chimeric PKSs via mix–and–match approaches.[66,68,70,71,104] The resulting PKSs often either show decreased turnover rates *in vitro* or impaired product formation rates *in vivo*.[70,71,104] As summarized in Chapter 2.1, biochemical analysis of a representative set of chimeric bimodular PKSs revealed a major bottleneck within the ACP:KS interface across the non–native module boundaries and prompted me to attempt different engineering strategies to ameliorate this issue.[104] Since the rate–limiting effect was caused by non–natural protein–protein interactions,[104] I sought to optimize the problematic chimeric ACP:KS interface via directed evolution experiments using phage display.

2.2.1 Introduction to Phage Display

Display technologies enable the presentation of libraries of peptides/proteins on various surfaces in ways that mimic the process of natural evolution.[111] They can be used in screening processes for affinity maturation of a protein/peptide for a ligand, or of one protein to another target protein. Phage display was the first display technology invented and is still the most widely used.[112] In phage display the DNA encoding a target protein is genetically fused to the DNA encoding for a phage coat protein, resulting in phage particles that display the target protein on their surface while also harboring the corresponding DNA. The direct coupling of genotype and phenotype allows for the generation of large libraries ($10^9 - 10^{10}$) of encoded peptides or proteins that can be easily generated via mutagenesis techniques.[113,114]

While a variety of bacteriophages can be employed in phage display,[115] the most commonly used filamentous M13 bacteriophage was used in this study.[116] M13 is

composed of five coat proteins and contains a single–stranded DNA genome.[117] For most applications proteins are either displayed on the minor coat protein P3 or the major coat protein P8.[118] Usage of P8 will result in the display of numerous copies of target protein per phage (100–200 per phage), while fusion to P3 will result on average in one copy per phage.[119] Typically, a phagemid vector is used which encodes the gene of the target protein fused to P3 (or a truncated version missing the first N–terminal domain, which is not necessary for infection),[117] an antibiotic resistance gene for selection and propagation, and the origin of replication (ori) for phage and its host.[118] A helper phage (e.g. M13KO7, containing a defective origin of replication, but providing all other viral proteins) is used in conjunction with the phagemid to allow for packing of the M13 particle.[118,120]

Thus far, phage display has been successfully used on a variety of differently sized proteins and peptides, ranging from 4–41 kDa.[117] While phage display is mainly used in the affinity maturation of antibody fragments,[118,121] it has been also used in peptide library screens,[122] and for display of entire proteins.[123,124] A typical enrichment factor per round lies between factor 10^2–10^4.[111]

In this project I aimed to improve the interaction between an ACP domain and a non–cognate KS domain across an intermodular interface. Due to the large number of possible exchanges and lack of information to narrow the range of potentially useful exchanges, introduction of single mutations to improve the ACP:KS interaction would be inefficient and time consuming. This is why I aimed to evolve the ACP:KS interaction in a directed evolution approach using phage display. As the natural ACP:KS interaction is already weak,[56,74] we expected an even weaker interaction in a chimeric system. My goal was to increase the affinity of an ACP for a non–cognate KS domain, with the hope that an increased affinity would correlate with an increase in the chain translocation rate, one of the major rate determining steps of chimeric PKSs.[104]

2.2.2 Goal and System Design

As a test system I selected the bimodular chimeric PKS comprised of LDD(4), (5)M1(2), and (3)M3–TE. Experiments described earlier in this thesis (Chapter 2.1) revealed severely diminished turnover of this chimeric PKS compared to the DEBS–derived bimodular reference system (LDD(4), (5)M1(2), and (3)M2–TE, Figure 8B). In addition, previous studies by our lab indicated that the ACP, which shuttles the polyketide intermediate across catalytic domains, uses distinct interfaces to interact with its cognate KS domain during chain elongation versus the downstream KS domain during chain translocation.[61,63] It was thus tested whether replacement of ACP1 by ACP2, or those parts of ACP2 involved in the chain translocation reaction, would increase the turnover rate of the bimodular PKS using M3–TE as the acceptor, since ACP2 is the natural interaction partner of M3 during chain translocation. And indeed, an approximately 5–fold higher turnover rate was measured when M3–TE was presented with ACP2 or parts of ACP2 as the upstream ACP, highlighting the possibility of engineering the chain translocation reaction in chimeric PKSs by optimizing the ACP:KS interface.[125] However, this approach has a few serious disadvantages: As the interactions in a PKS module are finely tuned, replacement of the whole ACP domain by a non–natural ACP domain to improve one protein–protein interface will alter all other ACP:domain interactions with unpredictable effects. Replacing only parts of the ACP domain instead might circumvent this problem, but is in turn more likely to affect the domain folding of the ACP itself with the consequence of compromising the entire protein fold. Site–directed mutagenesis of surface residues on ACP1 might overcome both limitations and could therefore have superior effects on the turnover rate of chimeric PKSs.

2.2.3 Proof of Principle Experiments to Ascertain the Biological Relevance of ACPs Presented on Phage

To ensure that ACPs on phage surfaces are biologically active, phagemids presenting different ACPs were constructed (eight in total), tested for stable expression of the fusion protein, and two of them used in an enzymatic proof–of–principle assay before setting up phage display screens. The architecture of the ACP fusions presented on the phagemid were similar in all cases: The ACP was C–terminally attached to a docking domain (in some cases the docking domain was omitted), followed by a truncated version of the minor M13 coat protein P3, which leads to an average of one fusion protein displayed per phage.[119] A FLAG–tag was inserted at the N–terminus and a secretion signal (StII) was located upstream of the whole fusion protein (Figure 10A). The stable representation of the fusion proteins on phages was confirmed using Western blotting (Figure 10B). Despite low signal strength and the appearance of potential cleavage products, significant size differences could be observed between those constructs harboring a docking domain (9 kDa) or those without, indicative of correct fusion of the ACPs to P3 and presentation of the full–length protein in a significant fraction of the phagemids.

To further confirm activity of ACPs even after transport through the periplasm, the ability of ACP_L–phage to translocate a polyketide chain to its cognate downstream KS1 (Figure 10C) was measured. Phage–fused ACP_L (Figure 10C, magenta curve) was compared to a positive control with soluble ACP_L (Figure 10C, green curve), background activity controls without ACP (Figure 10C, gray&black curves), as well as ACP2–phage, which is a non–cognate, phage–bound ACP that should not bind the downstream KS and therefore is not expected to perform chain translocation (Figure 10C, blue curve). All ACPs were activated via incubation with Sfp (the phosphopantetheine transferase from *Bacillus subtilis*) to allow for complete phosphopantetheinylation. After removal of Sfp, the ACPs were added to a reaction including $LDD–ACP^0$ (for loading of the ACP with the substrate via
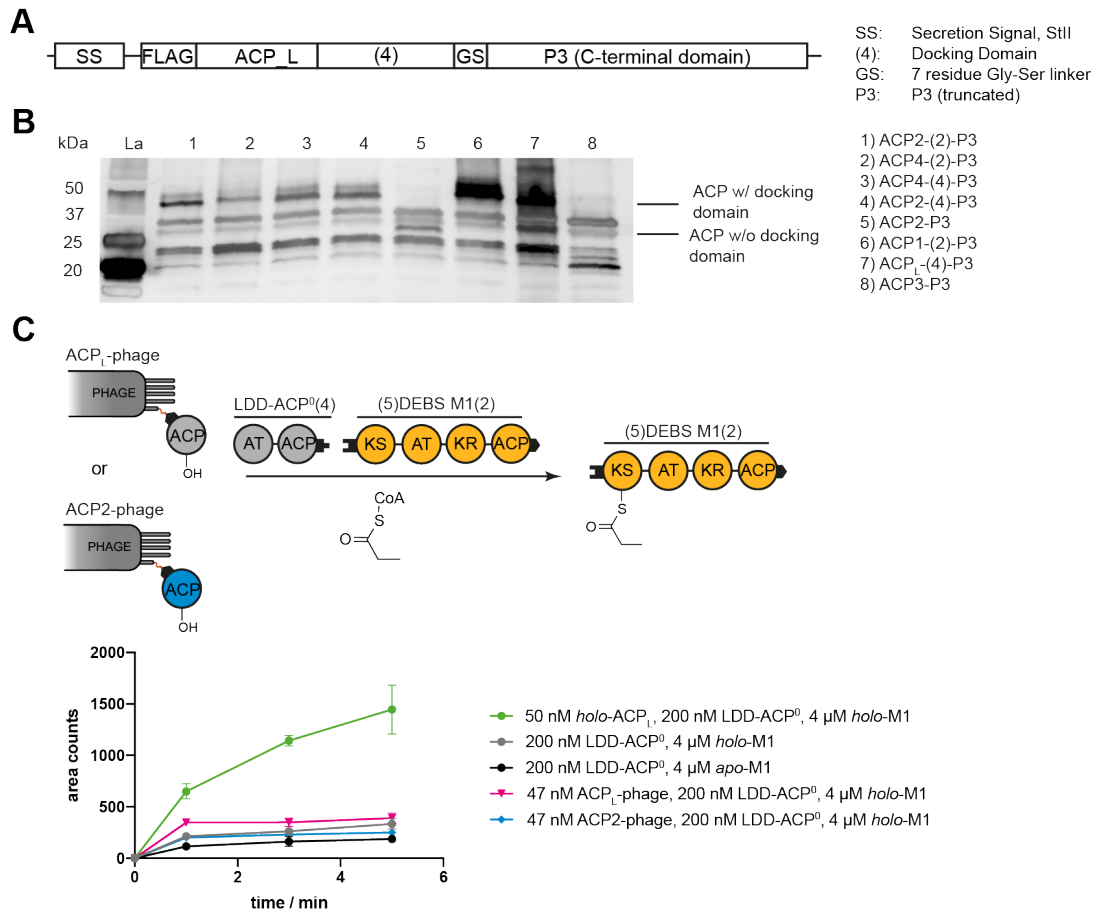
Figure 10 Analysis of biochemical properties of ACPs on phage. (A) Schematic representation of the phagemid architecture, abbreviations as indicated. (B) Western blot analysis to confirm presence of ACPs on phage. 10^{12} PFU/mL of phage were applied. First antibody: rabbit anti–FLAG, secondary antibody donkey anti–rabbit–HRP. (C) Chain translocation from ACP on phage to M1; assay design and results. ACP_L–phage/ACP2–phage were purified to highest possible titers (6.5×10^{12} resp. 8.7×10^{12} PFU/mL), and acylated via incubation with the phosphopantetheinyl transferase Sfp. After removal of Sfp chain translocation to M1 was measured using [^{14}C]–propionyl–CoA, LDD–ACP^0 (with an inactivated ACP) to load ACP–phage with the starter substrate. The reaction was quenched after 1,3, and 5 min and labeling on M1 was analyzed. For a detailed assay description see Section 4.5.2.

AT_L), [^{14}C]–propionyl–CoA, and M1. The reaction was quenched after certain time points and the amount of labeling on M1 as result of chain translocation was measured using an established radioisotopic SDS–PAGE labeling assay (Section 4.5.2).

As expected, labeling on M1 increased rapidly over a time course of 5 min in the positive control, indicating the ability of AT_L to load an ACP in *trans*. Negative controls in which only LDD–ACP^0 was incubated with either *holo*– or *apo*–M1 revealed only marginal increase in labeling over time (Figure 10C, gray&black curves). Of the two phage–bound ACPs used in this assay, only the cognate protein ACP_L (magenta) showed chain translocation activity over background, which indicated that phage–bound ACPs are indeed active and retain specificity for their targets, albeit at apparently greatly reduced activity compared to soluble ACP_L.

Several factors could contribute to this drop in apparent enzyme activity, either those directly affecting enzyme activity such as sterical hindrance and activity loss due to the multistep purification of the ACP–phage after Sfp incubation, or from overestimating the enzyme concentration in the assay. Since ACP concentration was indirectly determined using phage titer, differences in valency (active ACPs per phage) will greatly influence the calculated ACP–phage concentrations. However, other than decreasing the signal strength, these issues do not inherently interfere with the phage display screening procedure. The proof–of–principle experiments therefore indicated that the prerequisites for phage display were met. Phage–bound ACPs are sufficiently stably produced and able to bind their cognate enzymes via protein–protein interactions to make it possible to enrich them, which prompted me to set up phage display screens.

2.2.4 Library Preparation and Analysis of First Generation Mutants

To begin the first round of directed evolution experiments, a phage library of ACP1 variants in which five residues were randomized in the putative chain translocation epitope[61] was generated (ACP1–Lib1). These residues were chosen as likely candidates to alter protein:protein interaction specificities because they are located on the surface of ACP in the chain translocation epitope and not conserved between ACPs with different interaction partners (Figure 11A&B). Mutations were introduced by Kunkel mutagenesis.[126] The architecture of the ACP–fusions were similar to those in the proof–of–principle experiments, however now using docking domain 2 from DEBS (Figure 11C). The resulting phage library, presenting approximately one copy of ACP1 per phage was designed to be targeted against (3)KS3–AT3 (scheme in Figure 11D) and had a theoretical diversity of 3.12×10^6 different mutants. After transformation of *E. coli* SS320 cells with the library the practical diversity was determined to be 6.48×10^9 PFU/mL, indicating that each mutant was present at about 1000 copies. Sequencing of 96 clones from the library revealed that 60% of the clones harbored randomized sequences, whereas the remaining clones were wild–type ACP1.
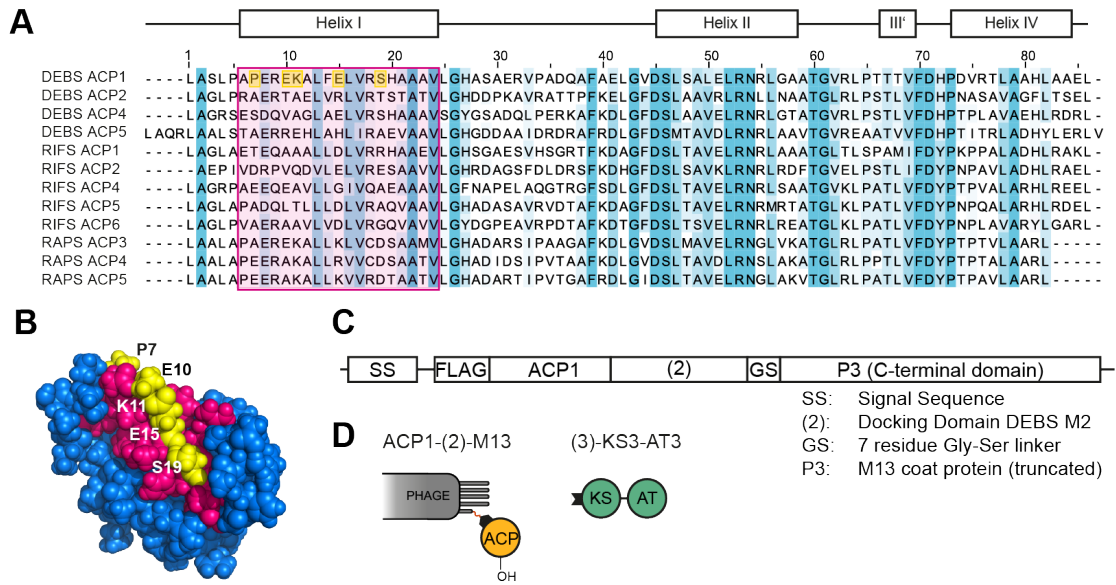
Figure 11 Design of the first ACP1–phagemid library. (A) Sequence alignment of ACPs. Helix I is highlighted in pink, position randomized for library generation are shown in yellow. Sequence identity highlighted in shades of blue (B) Homology model of ACP1. Surface residues selected for mutagenesis are highlighted. Color scheme as in (A). (C) Schematic representation of the phagemid architecture used for the library generation. Abbreviations as indicated. (D) Architecture of the ACP1 on phage and its target protein (3)KS3–AT3.

With the purified library a directed evolution experiment was performed against (3)KS3–AT3 as the target protein. Over five rounds of selection, ACP1 mutants with higher affinity towards (3)KS3–AT3 were enriched. Stringency was increased by increasing the number of wash steps performed in each round. As presented in Table 3, enrichment was first observed in round three. Thus only clones from elution round three, four and five were analyzed further for specific binding.

Table 3 Phage titers and enrichment factors during the directed evolution experiments of ACP1-Library1. Comparison of input and elution titers in each of the selection rounds and the corresponding enrichment factor. Enrichment factor calculated as change in the ratio of elution/input titers across two consecutive rounds of panning. N/A: not applicable.

Selection round	Input titer / PFU/mL	Elution titer / PFU/mL	Enrichment factor
1	2.32×10^{13}	9.67×10^5	N/A
2	5.60×10^{13}	4.17×10^5	0.18
3	3.04×10^{13}	3.50×10^6	15.47
4	5.60×10^{13}	6.67×10^5	0.10
5	1.25×10^{14}	3.83×10^5	0.26

In an initial ELISA, $\sim$100 clones from each of the selected elution rounds were analyzed for specific/increased binding towards (3)KS3–AT3. To assess non–specific binding, binding of the mutants to BSA was used as a negative control. Due to the intrinsically weak interaction of PKS proteins, the signal obtained for each mutant was compared to the signal obtained from wild–type ACP1 and ACP2 presented on phage. Only mutants which exhibited a stronger signal against (3)KS3–AT3 than wild–type ACP1, while also showing a lesser degree of BSA binding, were deemed valuable candidates and were further analyzed by sequencing. To reduce the number of mutants only those mutants harboring mutations in the selected positions (no frame shift causing mutations such as deletions or insertion) where subjected to another ELISA to confirm reproducibility of the observed binding (Figure 12A). The number of selected mutants was further decreased by excluding all mutants carrying mutations towards glycine or proline in the position 2–5 (located on ACP1 helix I) as glycine and proline are known to break α–helices.[127] This step was necessary as phage display only selects for proteins with a higher affinity, but as we were interested in enzymatic activity, conservation of the structural fold necessary for enzymatic activity had to be ensured. By applying these criteria the number of selected, enriched mutants was reduced to a total of six, two from each analyzed elution round.

The obtained mutations were cloned into a soluble, full–length M1 construct to assess the *in vitro* activity of the resulting chimeric PKSs. All six mutants showed similar purification behavior as wild–type M1. Using a UV_{340} spectrophotometric assay, the turnover rates of chimeric bimodular PKSs consisting of LDD(4), (5)M1(2), (3)M3–TE was determined using either wild–type M1 or any of the six mutants. The DEBS derived bimodular system using the acceptor (3)M2–TE was used as a reference system (Figure 12B). Of the six DEBS M1 mutants Lib1–Mut3 and Lib1–Mut5 exhibited the strongest effects on the chimeric system measured in a two–fold increase of the turnover rate. Nevertheless, none of the mutants was able to increase the activity of the chimeric system to that of the reference system.
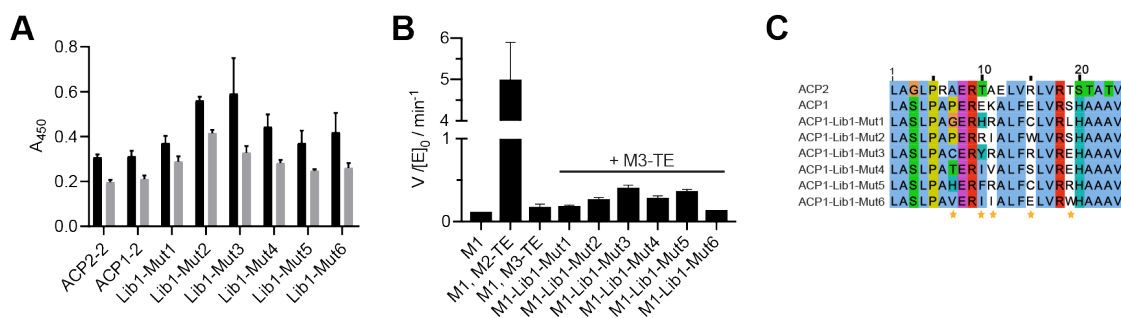
Figure 12 Analysis of enriched mutants from ACP1–Library1. (A) ELISA of ACPs presented on the phage surface. Results for selected mutants (Lib1–MutX) and wild–type ACPs are shown. Error bars represent signals from four individually grown phage cultures. Signal was obtained in (3)KS3–AT3 coated wells (black bars) and the degree of unspecific binding was assessed by comparing it to the signal in BSA coated wells (gray bars). (B) Turnover rates of wild–type and chimeric bimodular PKSs harboring the mutations enriched in the ACP1–Lib1 biopanning experiments. All bimodular PKS consisted of LDD(4), (5)M1(2), (3)ModuleX–TE. Either wild–type DEBS M1 was used as the first module (M1) or one of the six mutants obtained through the first directed evolution experiments (M1–Lib1–MutX). Initial rate data was obtained at individual PKS protein concentrations of 4 µM and non–limiting concentrations of propionyl–CoA, methylmalonyl–CoA, and NADPH. Measurements were performed in triplicate. (C) Sequence alignment of obtained mutants compared to wild–type ACP1 and ACP2. Randomized positions are indicated with a yellow asterisk.

Sequence alignment of the obtained mutants revealed no consensus sequence in any of the mutated positions (Figure 12C).

2.2.5 Analysis of Second Generation ACP1 Libraries

Based on the two most active mutants from ACP1–Lib1 (Lib1–Mut3 and Lib1–Mut5), two novel, deeper ACP1 libraries were generated. Both libraries introduced additional mutations at six other residues on helix I (ACP1–library 2 (ACP1–Lib2) was based on Lib1–Mut3 and ACP1–library 3 (ACP1–Lib3) was based on Lib1–Mut5). Similar to the design of library 1, the new positions chosen for randomization were not conserved across several ACP domains (Figure 13A) and located on the surface of the ACP (Figure 13B). Using the same protocol established for biopanning of ACP1–Lib1, both libraries were purified and subjected to the biopanning protocol. The theoretical diversity for both libraries was 6.40×10^7 PFU/mL and the practical diversity exceeded the theoretical diversity with 9.72×10^8 PFU/mL (ACP1–Lib2) and 3.02×10^8 PFU/mL (ACP1–Lib3) in both

cases. Significant enrichment of mutants from ACP1–Lib2 was observed in round 4 (Table S2), respectively round 5 for ACP1–Lib3 (Table S3).

The analysis of enriched clones from ACP1–Lib2 and ACP1–Lib3 was performed in the same way as for ACP1–Lib1. In an initial ELISA, 96 clones from each library were analyzed for specific/increased binding towards (3)KS3–AT3 compared to the negative control BSA. Binding of mutants from ACP1–Lib2 was compared to the signal obtained from Lib1–Mut3 (Figure 13C) and ACP1–Lib3 was compared to Lib1–Mut5 (Figure 13D). After applying the same criteria as before, only three mutants from ACP1–Lib2 and six mutants from ACP1–Lib3 were selected for analysis in the chimeric PKS system. It is noteworthy that most mutants exhibited a high degree of BSA binding, maybe due to the high prevalence of hydrophobic residues. Similar to ACP1–Lib1 no consensus sequence was observed among the enriched mutants (Figure 13E). The mutations corresponding to the nine selected mutants were cloned into the full–length M1 construct. All mutants were purified as soluble proteins with yields ranging from 0.5 – 2 mg/L of *E. coli* culture.

The activity of the newly selected mutants was measured according to established protocols. Strikingly, none of the novel mutants showed higher turnover rates compared to mutants M1–Lib1–Mut3 and M1–Lib1–Mut5, but rather decreased turnover rates (Figure 13F). While Lib1–Mut3 and Lib1–Mut5 contain fairly different residues, some similarities both among them and in comparison to ACP2 can be observed (Figure 13E, bottom). Both mutants carry the K11R mutation and an additional positive charge is found with either E15R (Lib1–Mut3) or S19R (Lib1–Mut5). As ACP2 also contains an arginine in position 15, one can speculate that the additional charge in Lib1–Mut3/Mut5 has a positive impact on the activity. However, the effect of individual mutations in the context of a multipoint variant are hard to predict due to non–additivity of functional effects, a mechanism called mutational epistasis.[128,129]
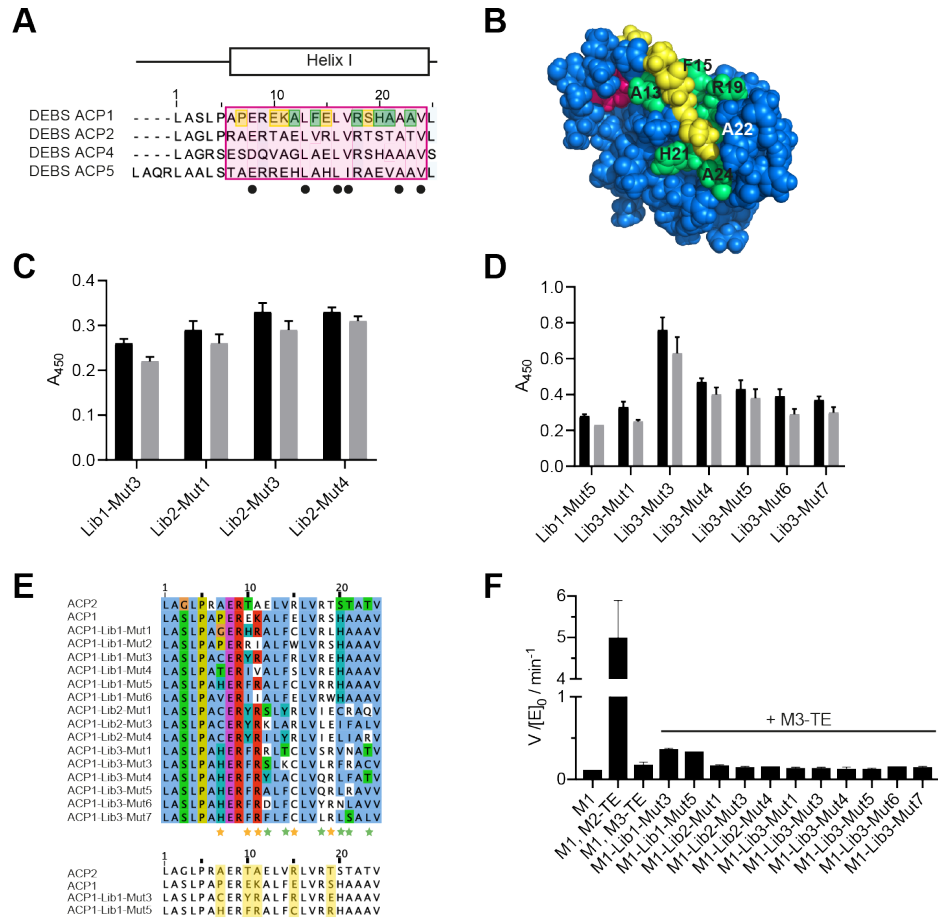
Figure 13 Design and analysis of two second generation ACP1 libraries. (A) Sequence alignment of ACPs highlighting positions selected for mutagenesis. Helix I in pink, position randomized for ACP1–Lib1 in yellow, and positions selected for a second round of directed evolution experiments are shown in green. Conserved positions are indicated with a dot (see alignment Figure 11A). (B) Homology model of ACP1. Surface residues selected for mutagenesis are highlighted. Color scheme as in (A). ELISA of ACP1–Lib2 mutants (Lib2–MutX) compared to Lib1–Mut3 (C) and ACP1–Lib3 mutants (Lib3–MutX) compared to Lib1–Mut5 (D). Error bars represent signals from four individually grown phage cultures. Signal was obtained in (3)KS3–AT3 coated wells (black bars) and the degree of unspecific binding was assessed by comparing it to the signal in BSA coated wells (gray bars). (E) Sequence alignment of obtained mutants compared to wild–type ACP1 and ACP2. Randomized positions are indicated with a yellow (ACP1–Lib1) or green (ACP1–Lib2/3) asterisk. Bottom panel shows the alignment of the wild–type ACP2 and ACP1 compared to the two best mutants from the directed evolution experiments; altered positions highlighted in yellow. (F) Turnover rates of wild–type and chimeric bimodular PKSs comparing all ACP1 mutations. All bimodular PKS consisted of LDD(4), (5)M1(2), (3)ModuleX–TE. All initial rate data was obtained at individual PKS protein concentrations of 4 µM and non–limiting concentrations of propionyl–CoA, methylmalonyl–CoA, and NADPH. Measurements were performed in triplicate. Either wild–type M1 was used as the first module (M1) or one of the ACP1 mutants enriched in the directed evolution experiments (M1–LibX–MutX).
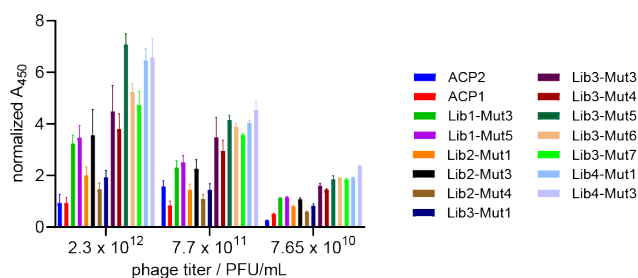
Figure 14 Titration ELISA of wild–type and mutant ACPs presented on the phage surface. Error bars represent measurements from two individual plates coated with (3)KS3–AT3. The ELISA signal was normalized to the amount of applied phage based on OD_{268} measurement.

To further evaluate the degree of increased binding of the various mutants a titration ELISA was conducted using three different phage concentrations. In contrast to previous ELISA measurements, the titration ELISA was performed with purified phage solutions instead of culture supernatant (for more details refer to Methods section 4.5.5 and 4.5.6). The concentration of each phage stock was measured by OD_{268} measurement and the obtained ELISA signal normalized to the amount of used phage. In contrast to the initial ELISA using phages from culture supernatants, four the selected mutants did not exhibit higher affinities than Lib1–Mut3/5 (Figure 14, Lib2–Mut1/3/4 and Lib3–Mut1). In summary the titration ELISA indicated that only mutants Lib1–Mut3/5 and Lib3–Mut3/4/5/6/7 can be treated as higher affinity binders compared to wild–type ACP1 (Figure 14).

2.2.6 Conclusion — Lessons to Learn from Directed Evolution Experiments

The use of phage display to engineer chimeric PKSs was based on two assumption, the one first being that helix I of the ACP plays a major role in interacting with the KS domain during chain translocation,[63] with the idea that optimization of the ACP:KS interaction during this reaction will increase the overall turnover rate of a chimeric PKS system. This assumption was supported by earlier experiments regarding the exchange of helix I for engineering of chimeric PKSs.[61,125] The second assumption was based on the idea that a non–native ACP:KS interaction

is weaker than a native interaction, and that improvement of the binding affinity between two heterlogous domains can lead to an increase in activity, implicating a correlation between affinity and activity.

Based on these assumptions I established phage display as a directed evolution approach to mutate surface residue on the ACP and thus increase the binding affinity of the ACP domain towards a non–cognate KS domain. All of the targeted positions were located on the ACP surface and non–conserved across ACP domains. We chose to work with a fusion to the minor coat protein of the M13 bacteriophage to avoid over–representation of the ACP domain. Despite extensive screening, only few of the enriched mutants exhibited specific binding towards the target protein (3)KS3–AT3. This might be explained with a common problem of phage display experiments, as sometimes phages with no actual affinity towards the target protein might be retained due to their ability to bind to the microtiter plate (plastic–binders) or due to propagation advantages.[113] Such phenomena can be found if many hydrophobic residues are located on the surface or if the protein in partly denatured.[113] And indeed in our experiments, many enriched mutants exhibited hydrophobic surface residues (Figure 13E). In addition, the high degree of unspecific binding might also be a result of the intrinsically weak affinity between PKS domains. Despite optimization of the biopanning protocol it is possible that I was not able to capture the weak ACP1:KS3 interaction as the k_{off} rate might be too fast to implement a suitable washing protocol (for reference, docking domains exhibit a dissociation constant (K_D) of $K_D \approx 20–100$ μM),[56,74,130] which is also observed in the lack of convergence of the mutants towards a consensus sequence. Finally, although the titration ELISA showed a tighter binding of mutants Lib1–Mut3/5 and Lib3–Mut3/4/5/6/7 compared to ACP1 (Figure 14), only the use of Lib1–Mut3 and Lib1–Mut5 resulted in a slight increase in the turnover rate of the chimeric system (Figure 13F). This data indicates that our second assumption about a correlation between affinity and activity, might either not be valid or only applicable within a certain range of affinities/activities.

Alternatively, the chosen surface residues, although participating the interaction, are not solely responsible for a productive ACP:KS interaction.

To validate my approach I wanted to measure the dissociation constant of the wild–type ACP1:KS3 interaction and of the mutant ACPs towards KS3. In a master thesis supervised by me, Melanie Janßen, compared different methods for the measurement of the dissociation constants between an ACP and a PKS module.[131] Both ITC (Isothermal Titration Calometry) and the AlphaLISA assay (Amplified Luminescent Proximity Homogenous Assay) were established to measure the native interaction between ACP2(2) and (3)M3–TE and ACP4(4) and (5)M5–TE. Both assays revealed dissociation constants in the µM–range ($K_D \approx 1$–20 µM) which is in the range of published values.[56,74,75] As ITC requires a high amount of purified protein for each measurement, the AlphaLISA assay was used for additional measurements. Unfortunately, and despite extensive troubleshooting, the results generated during the master's thesis could not be reproduced. This was mainly due to a bad signal–to–noise ratio of the AlphaLISA signal both in the presence and absence of the target protein. Due to our inability to obtain dissociation constants between all our mutants ACPs and KS3, we could not establish the degree of correlation between measured affinities and activities at the ACP:KS interface.

In summary, I established phage display as a new method in our laboratories for directed evolution experiments of the ACP:KS interface. In the course of this work some ACP mutants with increased binding affinities towards the KS domain were enriched, of which only two showed a rate–improving effect in the context of a chimeric PKS system. Dissociation constants for the ACP:KS interaction could not be established, yet an estimation of the affinity from titration ELISA experiments (Figure 14) argues against a tight correlation between affinity and activity as many of the enriched mutants showed increased binding with no effect on the activity (Figure 13F). The idea to use directed evolution approaches to optimize a chimeric

ACP:KS interface seems promising, but selectivity issues that emerged during the experiment combined with a lack of knowledge about crucial parameters of the ACP:KS interaction have to be overcome before phage display can turn into a promising option in directed evolution of PKS systems.

2.3 Engineering of Chimeric Polyketide Synthases using SYNZIP Docking Domains

Another aim of this thesis was to expand the existing toolbox to engineer chimeric PKSs. My goal was to establish a robust domain–domain interface that has minimal impact on enzyme activity to use in mix–and–match approaches, and to introduce synthetic docking domains as a new tool to bridge PKS modules and domains. The results of this study are mostly published in the Journal ACS Chemical Biology and large parts of the text and figures from this chapter are taken from the prepared manuscript.[105]

As outlined in the introduction, several studies have focused on the exchange of intact PKS modules to retain the inherent modularity of PKSs.[70,71,104] To mediate communication between separate modules, PKS–derived docking domains were installed at the N– and C–termini of heterologous modules. These docking domains are often relatively large and interact with weak affinities ($K_D \approx 20$–100 µM).[56,74,130] Chimeric PKSs harboring intact heterologous modules retain native domain–domain interactions within individual modules, but introduce a non–native ACP–KS interface in the chain translocation reaction (intermodular chain translocation interface),[63,72] and therefore suffer from low product titers *in vivo*[70] and decreased turnover rates *in vitro*.[104] Alternatively, structurally stable chimeric PKSs could also be engineered to preserve the chain translocation interface while modifying the intramodular chain elongation interface, by fusing the N–terminal condensing part (comprised of the KS and AT domains) of one module to the C–terminal processing part (harboring auxiliary enzyme(s) and ACP domain) of another. However, such possibilities may be constrained by KS–ACP specificity during polyketide chain elongation.[61,132] There remains a need for a general method to access catalytically efficient chimeric assembly line PKSs.[106]

In this project I explored the utility of heterospecific, high–affinity ($K_D \approx 10$ nM) coiled–coil interaction domains termed SYNZIPs,[133,134] for engineering chimeric PKS modules comprised of an N–terminal condensing part (KS–AT) and a C–terminal (DH–ER–KR)–ACP processing part derived from different modules. The non–covalent, SYNZIP domain–mediated interface was anticipated to be minimally invasive to a PKS module's function and as such could be harnessed for PKS engineering. This assumption was mainly built on structural data of PKSs and the related fatty acid synthases (FASs), where minimal contact is observed between the condensing and processing parts,[41,45,51] thereby allowing extensive rotational movement of these two parts.[58,60] Furthermore, genetic analysis suggested an evolutionary relationship between the upstream (DH–ER–KR)–ACP parts and the downstream KS domains, indicating that introduction of an intramodular chimeric PKS interface between the AT domain and the downstream processing part may be beneficial as it preserves the evolutionary conserved intermodular interface.[135,136] This in turn would enhance the modularity of the multienzyme system, allowing for numerous combinations in the generation of chimeric PKSs.

2.3.1 The Use of SYNZIP Domains for Engineering Chimeric PKS Modules

As a test case for the utility of SYNZIP domains, we used a previously characterized bimodular PKS derived from DEBS (Figure 3A) that is comprised of three proteins: LDD(4), (5)M1(2), (3)M2–TE (also employed in Chapter 2.1, see Figure 7).[75] In this system, the numbers in parenthesis denote the origins of matching docking domains that are fused to the N– or C–termini of these three proteins. For example, LDD(4) refers to the loading didomain to which the docking domain from the C–terminus of DEBS Module 4 has been fused. Previous studies have shown that the AT–KR linker can be generally cleaved at a well–defined site without structural perturbations to either the N–terminal or C–terminal domains,[37,87] but with a significant kinetic penalty to the chain elongation step.[132] Accordingly,

DEBS M1 was genetically cleaved at this site (Figure 15A). The choice of SYNZIP domains to introduce at this junction was guided by the length and orientation of the resulting coiled–coil. Among the best characterized orthogonal SYNZIP pairs are SYNZIP1+SYNZIP2 and SYNZIP3+SYNZIP4 (hereafter annotated as SZ1+SZ2 and SZ3+SZ4) both interacting in a parallel orientation. While both pairs form highly stable heterodimers (k_{off}, SZ1+SZ2 $\approx$ 7.8 x 10^{-4} s^{-1}, k_{off}, SZ3+SZ4 $\approx$ 1.4 x 10^{-2} s^{-1}),[137] we chose the SZ3+SZ4 pair, as the SZ1+SZ2 coiled–coil is predicted to be longer.[134] Throughout this study, SZ3 was fused to the C–terminus of one target protein, while SZ4 was fused to the N–terminus of another target protein. Being approximately 60 Å long, the SZ3+SZ4 coiled coil is predicted to be of the same length as the entire KS dimer (Figure 15B). Because we used a (GGSG)$_2$–linker to connect the SYNZIP domains to the PKS module fragments, we assumed the conformational flexibility of the ACP would be sufficient to allow for interaction with the KS and AT domain of the reconstituted module.

Using a UV$_{340}$ spectrophotometric assay, we measured the turnover rate of the bimodular PKS in the presence of either intact M1, split M1 bridged by SYNZIP domains, split M1 without SYNZIP domains, or split M1 in which only one of the SYNZIP domains needed to form the heterodimer was present (Figure 15C). Remarkably, the turnover rates of the PKSs harboring intact M1 or SYNZIP–bridged M1 did not differ significantly (Figure 15C, first three columns), while omission of either or both SYNZIP domains led to a predicted large drop in turnover rates (Figure 15C, last four columns). These results indicate that a SYNZIP–bridged module can achieve comparable catalytic activity to an intact one. By introducing a non–covalent SYNZIP pair at the junction site of the condensing and processing part of a module, it is thus possible to achieve a high effective molarity of the α–carboxyacyl–ACP substrate for the KS–catalyzed chain elongation step.

As the cleavage at the AT–KR junction site most deleteriously affects the KS:ACP interaction during chain elongation,[37,132] we next sought to compare the rates of
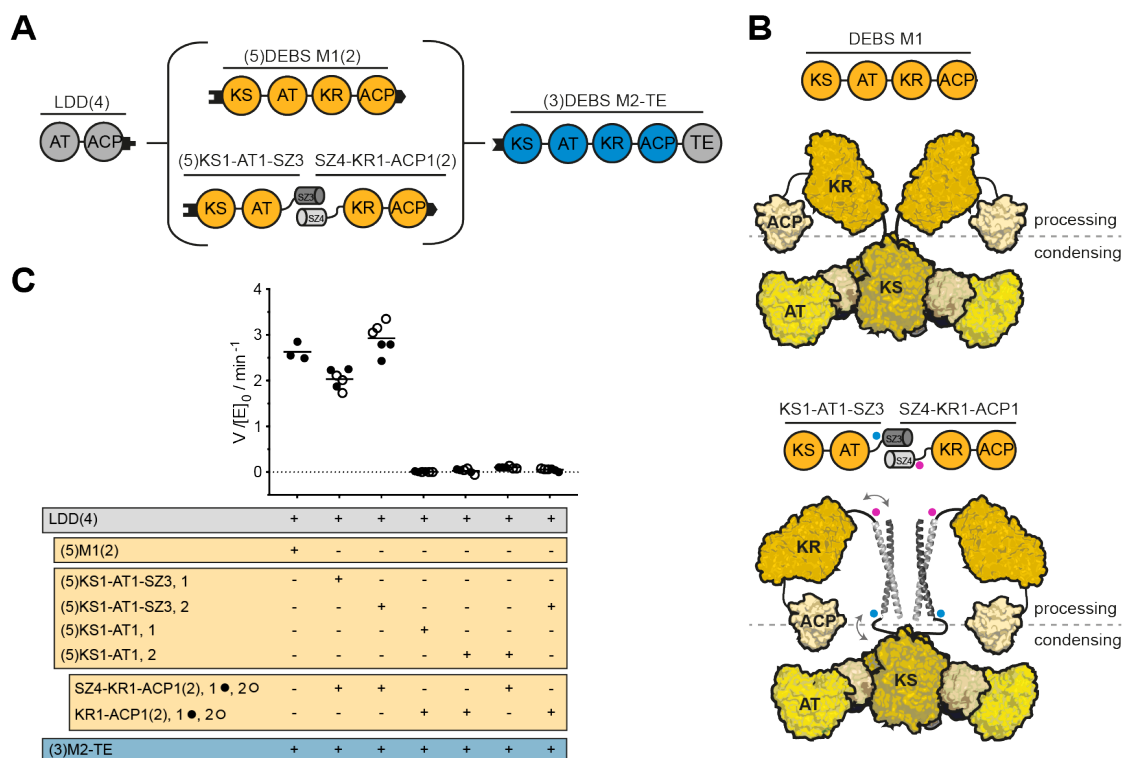
Figure 15 Use of SYNZIP domains in the design of catalytically efficient PKSs harboring a split module. (A) Design of a bimodular DEBS derivative comprised of LDD(4), intact (5)M1(2) or a split version thereof, and (3)M2–TE. (B) Model of DEBS M1 with the natural AT–KR linker (top) and the SYNZIP–containing variant (bottom). This model was built based on SAXS analysis of DEBS M3.[46] Fusion sites of the SYNZIP domains to either the AT or KR are indicated by blue/pink dots, illustrating the parallel orientation of the heterospecific coiled–coil. An eight–residue flexible Gly–Ser linker is used to connect the SYNZIP domain to the PKS protein (for protein sequences see Table S2 in ref[105]). (C) Turnover rates of bimodular PKSs employing M1 or a split M1. Except for intact M1, at least two independently purified protein preparations were evaluated (1 or 2): each preparation of the N–terminal part is shown in a separate column, whereas preparations of the C–terminal part are indicated as black and white dots. All initial rate data was obtained at 2 µM enzyme concentration and non–limiting concentrations of propionyl–CoA, methylmalonyl–CoA, and NADPH. Measurements were performed in triplicate and the grand mean is indicated. Protein quality was confirmed via SDS–PAGE and SEC analysis (Figures S1, S2, and S3).

chain elongation between an intact M1 and a split M1 with or without SYNZIP domains. For this purpose, we used a radioisotope labeling assay in which [1–^{14}C]–propionyl–CoA is used to prime the LDD, thereby allowing measurement of the occupancy of the N–terminal and C–terminal fragments of Module 1 with radiolabeled polyketide intermediates. First, we performed a set of assays in the absence of methylmalonyl–CoA, the co–substrate required for chain elongation. As shown in Figure 16A, the N–terminal (KS–AT) fragment of M1 was rapidly labeled (within 1 min) regardless of the presence or absence of a SYNZIP domain. Presumably this labeling was due to efficient translocation of the propionyl moiety from LDD to the KS active site. As expected, in the absence of SYNZIP domains, no labeling of the ACP–containing C–terminal protein was observed, nor did the labeling intensity of the intact M1 protein exceed an average occupancy of one equivalent[125] at any time. However, in the presence of SYNZIP domains, labeling of the C–terminal protein increased steadily over 10 min. The absence of methylmalonyl–CoA prevents chain elongation and transfer of the labeled growing polyketide chain to the C–terminal protein, which suggests that this labeling is indicative of transacylation of the propionyl moiety from the KS to the downstream ACP. Previously, it was proposed that the interaction site of the upstream ACP with the downstream KS during chain translocation is distinct from the one of the cognate ACP:KS pair during chain elongation.[61,63] The changed geometry in a SYNZIP–interfaced module provides the ACP with greater flexibility compared to an intact module, potentially allowing the ACP access to the KS domain in an orientation that is precluded in a native module and thus resulting in the proposed transacylation of the propionyl moiety. This transacylation reaction might be more pronounced in the absence of the co–substrate for chain elongation.

While chain translocation from LDD to the KS domain of a split module was unaffected by the presence of a SYNZIP interface (Figure 16A), chain elongation showed strong dependence on the SYNZIP domains upon the addition of methylmalonyl–CoA, (Figure 16B). The ACP–containing C–terminal protein was
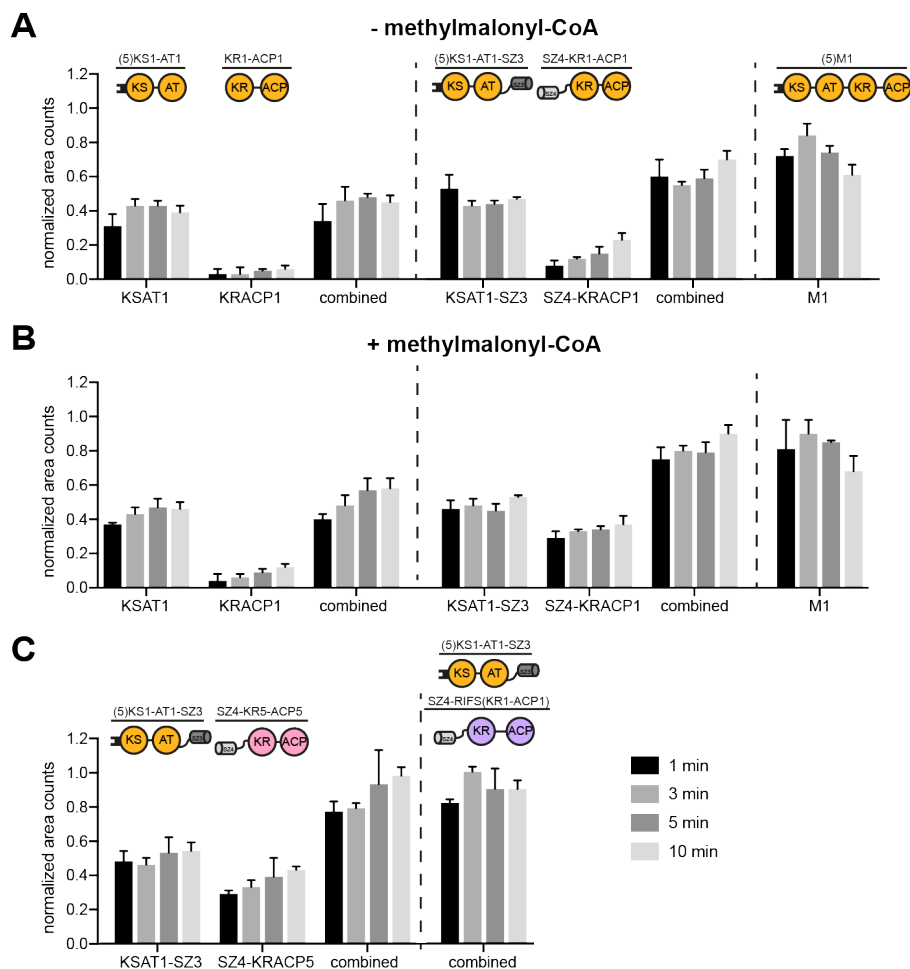
Figure 16 Comparison of chain translocation and elongation rates of intact versus split modules. The occupancy of individual proteins by growing polyketide chain precursors was measured in the absence (A) or presence (B) of methylmalonyl–CoA, and in modules containing chimeric KS:ACP interfaces (C, with methylmalonyl–CoA). All measurements included LDD(4) plus the designated split/intact module combinations. Labeling of the different proteins was quantified at different time points (1, 3, 5, and 10 min) and the resulting counts were normalized based on the maximal occupancy. Combined counts of both the KS and ACP containing proteins are also presented for split–module systems. In the case of RIFS Module 1, the N– and C–terminal fragments could not be separated by SDS–PAGE gel; hence only combined counts are reported (panel C). All measurements were performed in triplicate using 2 µM enzyme concentration and non–limiting concentrations of propionyl–CoA, methylmalonyl–CoA, and NADPH. Protein quality was confirmed via SDS–PAGE and SEC analysis (Figures S1, S2, and S3).

rapidly labeled (within one 1 min) and the combined labeling intensity of the two SYNZIP–fused proteins was similar to that observed for intact M1, suggesting that the effective molarity of the α–carboxyacyl–ACP species was comparable in both cases. A turnstile model has recently been proposed to account for the ability of an unoccupied KS active site to discriminate between the occupied versus unoccupied states of its partner ACP domain.[125] Assuming that module M1 saturates at an average occupancy of one radiolabeled acyl chain per module, our data argues that the turnstile mechanism has been preserved in SYNZIP–bridged M1, notwithstanding significant differences in module geometry and flexibility in the two cases (Figure 15B).

Encouraged by the results described above, we sought to examine whether SYNZIP domains could facilitate labeling between the KS–AT fragment of Module 1 and the ACP domain of a non–cognate module, indicative of efficient chain elongation across a chimeric interface. Specifically, we measured the occupancy of the SZ4–fused C–terminal (KR–ACP) fragments of DEBS Module 5 and rifamycin synthase (RIFS) Module 1 (Figure 16C). In both cases the combined counts of the N– and C–terminal proteins were comparable to those of the reference intact module. In light of the proposed ability of a KS–bound polyketide chain to transacylate to the ACP in a split–module system harboring SYNZIP domains (Figure 16A), it was not possible to conclude from this data alone whether chain transacylation or chain elongation kinetics were being enhanced by the SYNZIP interface in these chimeric modules. However, given that the apparent condensation rate of the reference system (Figure 16B) was much faster than the transacylation rate (Figure 16A), the strong labeling of the chimeric C–terminal proteins was assumed to originate mostly from elongation rather than transacylation.

Taken together, our data highlights the potential utility of SYNZIP domains as tools to engineer non–covalently recombined chimeric modules at the AT:KR interface.

2.3.2 Design and Activity of Chimeric PKSs using Different Domain–Domain Interfaces

Following this initial assessment, we extended our evaluation of SYNZIP domain–mediated engineering to a larger set of bimodular chimeric PKSs in which the non–cognate domain interface corresponds to the intermodular chain transloca-tion step of the PKS catalytic cycle (Figure 3B). This set of PKSs had previously been studied with intact heterologous modules interfaced by non–covalent dock-ing domains.[70,71,104] We constructed three types of bimodular PKSs: using non–covalent docking domains (Figure 17A), a SYNZIP–interfaced system (Figure 17B), and a combination of covalent fusions and SYNZIP domains (Figure 17C). Earlier analysis of a subset of analogous covalently fused chimeric PKSs had revealed quantitatively distinct properties compared to non–covalently fused systems.[72] As detailed below, careful examination of our findings can reconcile these quantitative differences.

As shown in Chapter 2.1, when non–covalent docking domains were engineered at the intermodular interface between DEBS M1 and various downstream modules, only the wild–type bimodular PKS with DEBS M2 as the acceptor achieved high turnover rates. All bimodular chimeric PKSs exhibited significantly decreased turnover rates (Figure 17A). When SYNZIP domains were installed at the inter-modular interface, the cognate DEBS M1–M2 system achieved similar turnover rates as the respective docking domain–interfaced one (Figure 17B). In addition, the turnover rate of the system harboring DEBS M6 as an acceptor increased sig-nificantly in the presence of SYNZIP domains, while the turnover rates of chimeric PKSs harboring DEBS M3 or DEBS M5 did not change appreciably (Figure 17B). In the third type of bimodular PKSs, the split Module 1 is bridged by SYNZIP do-mains and the downstream acceptor module is covalently fused to the C–terminal fragment of Module 1 (Figure 17C). Similarly to the second design, the cognate DEBS M1–M2 system was unaffected. However, chimeric systems harboring DEBS
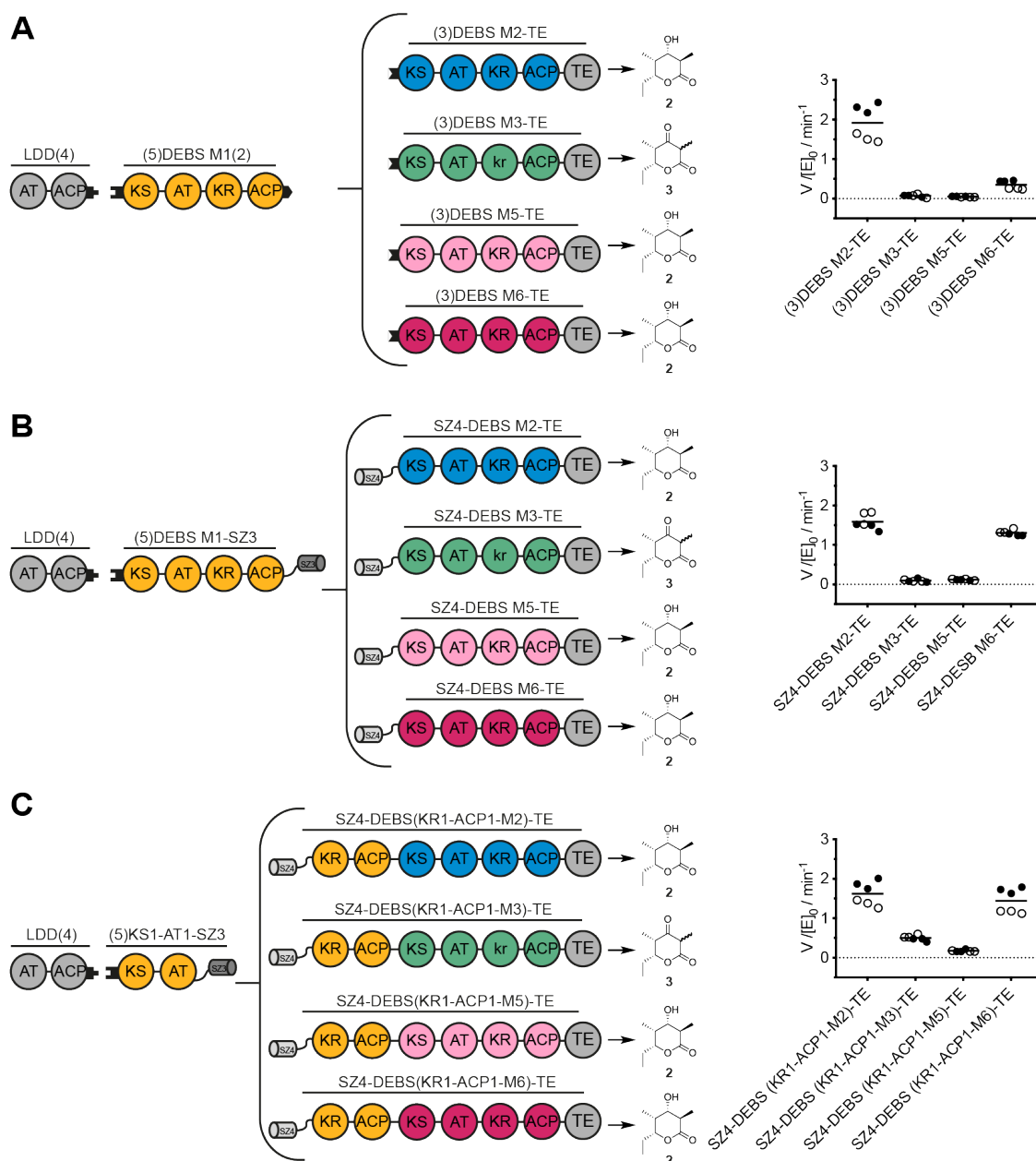
Figure 17 Turnover rates of bimodular chimeric PKSs harboring a chimeric chain translocation interface. Design and turnover rates of bimodular chimeric PKSs. LDD(4) was always used as a separate protein, whereas DEBS M1 and variable acceptor modules were interfaced using either non–covalent docking domains (A), SYNZIP domains (B) or a combination of covalent fusions and SYNZIP domains (C). Construct design and the predicted triketide lactone products are shown next to the measured turnover rate of each system. All initial rate data was obtained at 2 μM enzyme concentration and non–limiting concentrations of propionyl–CoA, methylmalonyl–CoA, and NADPH. In all cases, two independently purified protein preparations were evaluated (data shown as black and white dots). Each set of measurements was performed in triplicate and the grand mean is indicated. LC–MS analysis of bimodular PKSs newly generated in this study (B, C) was performed after overnight incubation to verify product identity (Figure S4). The expected products were detected for all reactions except for the bimodular PKS with SZ4–M3–TE as the acceptor. For product analysis of the reference system refer to Klaus *et al.*.[104] Protein quality was confirmed via SDS–PAGE and SEC analysis (Figures S1, S2, and S3).
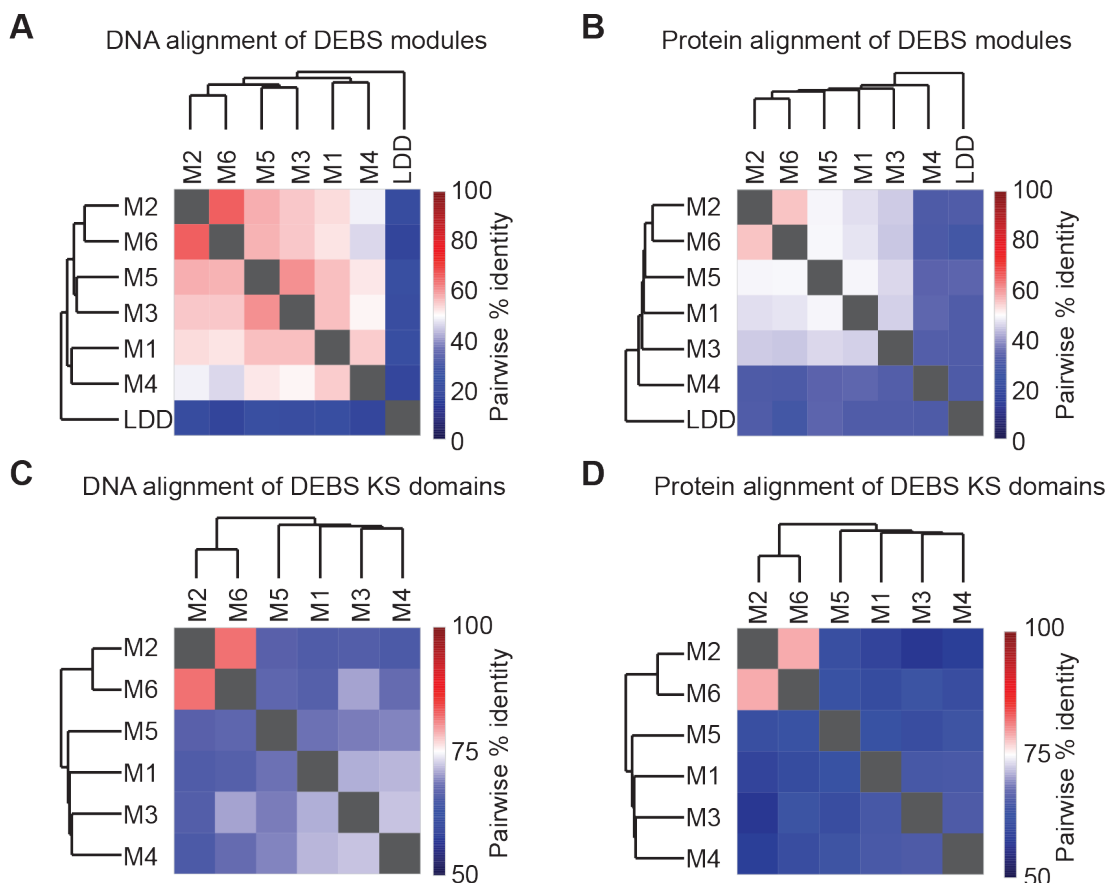
Figure 18 Sequence similarity analysis of DEBS modules and KS domains. Pairwise alignment of DNA and protein sequences of DEBS modules (A–B) and DEBS KS domains (C–D) are depicted as heatmaps. Analysis and figure generation by co–author Dr. Aleksandra Nivina.

M3 and M6 showed increased turnover rates compared to the analogous docking domain–interfaced systems (Figure 17C).

Of all modules, DEBS M6 benefits most from close proximity to the upstream module M1. Sequence similarity analysis across all DEBS modules and standalone KS domains revealed the strongest similarity between DEBS M2 and M6, specifically between KS2 and KS6, both on the DNA and the protein level (Figure 18). As such, in chimeric bimodular PKSs, M6 seems to be a suitable replacement for M2 with regards to both the ACP:KS mediated chain translocation reaction, as well as substrate recognition and processing of the diketide by the KS. Other heterologous modules (e.g. M3, Figure 17C) might also benefit from a tight connection to the upstream module, by e.g. increasing the effective molarity of the
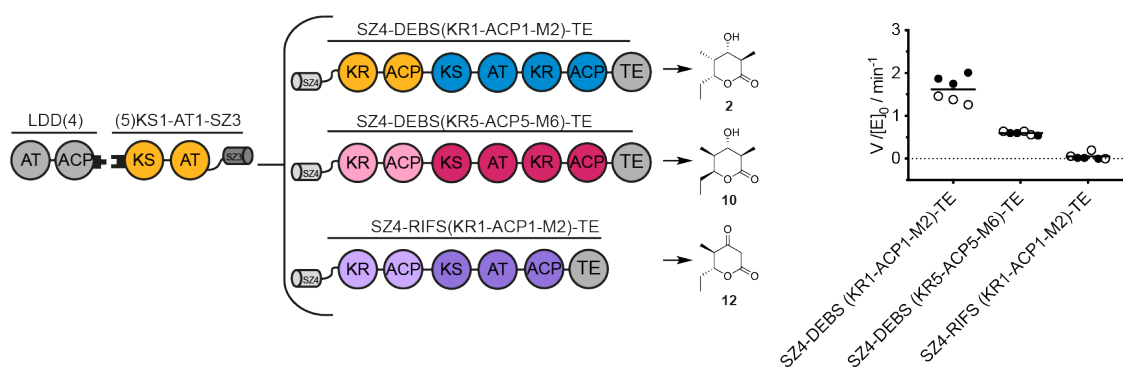
Figure 19 Turnover rates of bimodular chimeric PKSs harboring a chimeric chain elongation interface. Bimodular PKSs comprised of LDD(4), (5)KS1–AT1–SZ3 and variable SZ4–KR_n–ACP$_n$–Module$_{n+1}$–TE constructs. Construct design and the predicted triketide lactone products are shown next to the measured turnover rate of each system. All initial rate data was obtained at 2 μM enzyme concentration and non–limiting concentrations of propionyl–CoA, methylmalonyl–CoA, malonyl–CoA (RIFS M2 is malonyl specific), and NADPH. Turnover analysis was performed on two individually purified proteins per construct (back and white dots). Measurements were performed in triplicate and the grand mean is indicated. Protein quality was confirmed via SDS–PAGE and SEC analysis (Figures S1, S2, and S3).

bimodular complex during chain translocation,[104] yet other intrinsic specificities such as protein–protein interactions and protein–substrate recognition remain as possible sources of impaired activities.

Data on bimodular systems (see Figure 15 and Figure 17) suggests that SYNZIP–interfaced systems can enhance the turnover rates of some chimeric PKSs. SYNZIP interfaces appear to be comparable to covalent linkages in this respect, with the added benefit of mitigating the need to express and purify large proteins such as intact bimodules.[75] Because bridging the split Module 1 with SYNZIP domains is minimally deleterious to protein structure (see Figure 17C), and because chimeric modules showed elongation kinetics approaching that of the wild–type system (see Figure 16C), a new and superior strategy for engineering chimeric PKSs may be to install non–cognate interfaces within modules while in turn preserving the cognate module–module interfaces.

In order to test this hypothesis, the PKS design outlined in Figure 19 was utilized. The N–terminal fragment is identical to the one used in the previous assays and is bridged to the C–terminal fragment by SYNZIP domains (Figure 17C). However, the location of the chimeric interface in the C–terminal fragment has been shifted

from elongation to translocation: the processing parts of DEBS M5 and RIFS M1 are fused to their cognate downstream modules DEBS M6 and RIFS M2, respectively. Kinetic analysis of both chimeras revealed significantly attenuated turnover rates (Figure 19). LC–MS analysis after overnight incubation revealed the correct masses for the expected products of the reference system and the PKS with DEBS modules 5 and 6, yet not with RIFS modules 1 and 2 (Figure S4). The precise mechanism for these functional impairments was not established, but previous studies suggest that attenuated rates may be caused by impaired substrate recognition imposed by binding pocket or domain–domain interface specificity.[107,138] Alternatively, intramodular transacylation of the polyketide intermediate from the KS to the ACP without chain elongation could also present a barrier for polyketide chain growth by inhibiting the enzyme with a non–processable polyketide chain; this could be especially deleterious for chimeric modules with inherently low turnover rates. This side reaction might be suppressed by redesigning the SYNZIP system to limit the conformational variability of ACP, for example by shortening the coiled coil, rigidifying Gly–Ser linkers, or using lower–affinity docking domains.

2.3.3 Influence of different SYNZIP Domains on PKS Turnover

To explore the influence of SYNZIP domains on PKS geometry and turnover, we analyzed different combinations of SYNZIP domains in the context of a bimodular PKS. As depicted in Figure 15B, insertion of SYNZIP domains at the AT:KR interface likely alters the geometry of the PKS module. Although no structural information is available, one can predict that the length of the chosen SYNZIP coiled–coil can affect PKS activity by influencing the orientation of modules towards each other. In order to test this hypothesis we compared the activity of the bimodular DEBS reference system in which M1 was bridged at the AT:KR interface by different SYNZIP combinations (Figure 20A). Besides the previously introduced combination of SZ3+SZ4, we also analyzed SZ1+SZ2, and SZ3 in combination with
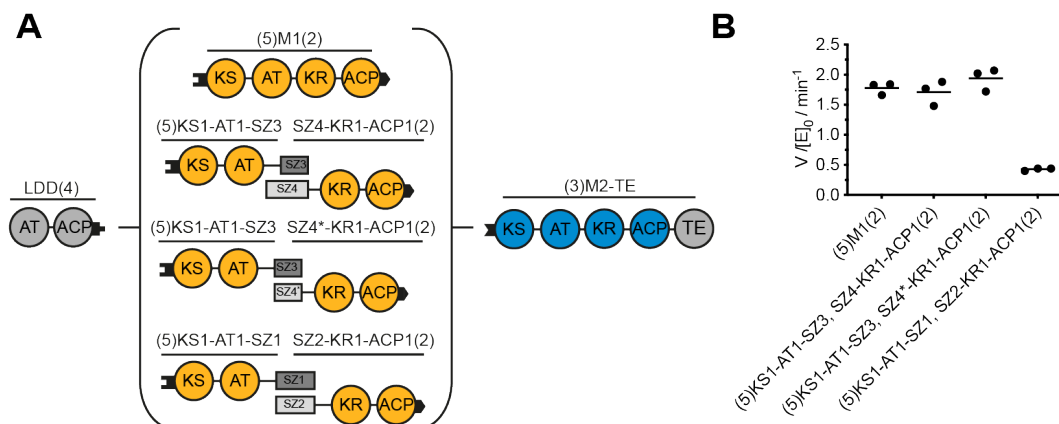
Figure 20 The influence of different SYNZIP domains on the turnover rate of bimodular PKSs. (A) Design of DEBS–derived bimodular PKSs harboring using different SYNZIP domains. SZ4* denotes a truncated version of SYNZIP4 missing the first ten N–terminal residues which are predicted not to be involved in the coiled–coil formation. (B) Turnover rates of bimodular PKSs employing M1 or a split M1. All initial rate data was obtained at 2 µM enzyme concentration and non–limiting concentrations of propionyl–CoA, methylmalonyl–CoA, and NADPH. Measurements were performed in triplicate and the grand mean is indicated. For proteins sequences see Table S2 in ref[105] and Table S4.

a truncated version of SZ4 (annotated as SZ4*), as the first ten N–terminal residues of SYNZIP4 are not predicted to participate in coiled–coil formation.[134] We also aimed at analyzing SYNZIP17+SYNZIP18 (hereafter annotated as SZ17+SZ18) as this pair is predicted to form an antiparallel coiled–coil, which might result in a more constraint PKS module geometry compared to SZ3+SZ4. While we were able to purify KS1–AT1–SZ17, purification attempts of SZ18–KR1–ACP1 only resulted in truncated protein preventing any further analysis of this SYNZIP combination. In contrast, KS1–AT1–SZ1, SZ2–KR1–ACP1, and SZ4*–KR1–ACP1 purified similar to previously analyzed KS–AT and KR–ACP fragments.

When analyzed in the context of a bimodular PKS (Figure 20A), the turnover rate of SZ3+SZ4 and SZ3+SZ4* was similar to that of intact M1, while the SZ1+SZ2 pair showed a significantly decreased turnover rate (Figure 20B). This data is a first indication that the choice of SYNZIP domains both with regard to the length of the formed coiled–coil, but also its stability, will affect the turnover rate of the bridged PKS module and should be considered in the context of individual experiments.

2.3.4 Conclusion — SYNZIP Docking Domains as a Tool for Engineering PKSs

Several different chimeric PKS designs involving SYNZIP domains installed at either a non–native chain translocation or chain elongation interface were systematically compared in this study. Whereas the chain translocation interface is well established, the benefit of docking domains at the chain elongation interface had not been investigated prior to this study. Our findings suggest that SYNZIP domains have considerable potential, especially in instances where proximity generated by a high–affinity linkers can improve ACP:KS interactions. In contrast, other chimeric PKSs cannot be optimized by trading one non–native protein–protein interface for another in a straightforward manner.[139] We conclude that both interfaces are equivalently suited for PKS engineering, but that the resulting chimeric PKSs are often kinetically impaired, requiring further optimization. Overall, our findings showcase the utility of SYNZIP domains as a new tool in PKS engineering that can replace traditionally used PKS docking domains while also mitigating the need for expressing and purifying excessively large PKS proteins. Their small size, high affinity interaction, and the multitude of existing orthogonal pairs makes them ideal candidates to bring non–natively interacting PKS modules/domains into close proximity. As they alter the geometry of a PKS module, they could also be used to study orientation, flexibility and conformational restraints of domains.

2.4 Engineering the Ketosynthase Substrate Specificity in Chimeric Polyketide Synthases

Active site mutagenesis is an alternative way to increase the product spectrum and activity of assembly line PKSs. While tight domain–domain interactions have been identified as crucial for chimeric PKS function in the previous chapter, inefficient turnover of non–natural substrates delivered to active sites creates another set of bottlenecks in chimeric PKSs. A clear example of a chimeric PKS's activity being restricted by a non–optimal substrate rather than inefficient domain interfaces can indeed be found in an earlier chapter: A bimodular chimeric PKS using M6–TE as the final acceptor was shown to be more active when a non–native upstream ACP delivered its preferred substrate than when the native ACP delivered an enantiomeric substrate. The design of the two bimodular PKSs is depicted in Figure 21, while the turnover rates are shown in Figure 17C&19.

In one case, KR1–ACP1 was fused as the upstream donor resulting in a chimeric ACP1:KS6 interface, with ACP1 harboring the DEBS derived natural diketide (NDK, (2*S*,3*R*)–2–methyl–3–hydroxy–diketide; Figure 17C), and in another setup, KR5–ACP5 was used as the upstream fragment, resulting in the natural ACP5:KS6 interface, but with ACP5 bearing the enantiomeric diketide (EDK, (2*R*,3*S*)–2– methyl–3–hydroxy–diketide; Figure 19). The ∼2–fold higher turnover rate correlated well with the preference of M6 for NDK over EDK as measured in k_{cat}/K_M,[140] indicating a rate–limiting effect due to the substrate specificity of M6–TE. This M6–TE PKS module, together with a second module (M3–TE) was thus used as a model in the following chapter to test a targeted approach of increasing the substrate tolerance of the enzyme in a chimeric context.

Since PKS systems have multiple active sites within a single module, narrowing down the possibly rate–limiting enzymatic reaction(s) is an important first step to allow targeted mutagenesis. In case of our model PKS, three reactions can in prin-
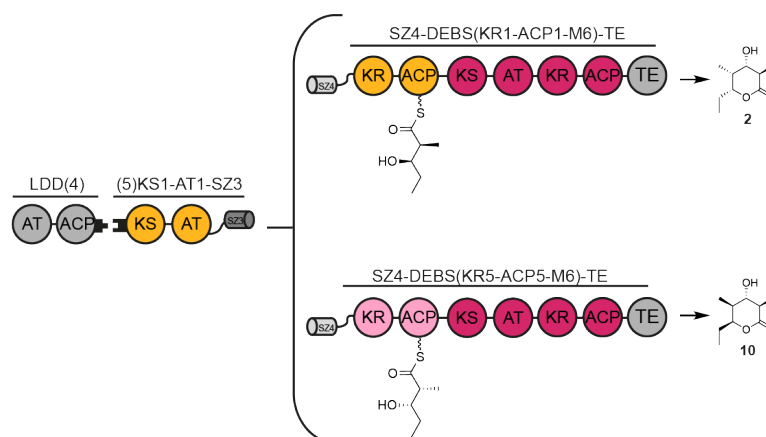
Figure 21 Two bimodular PKSs using M6–TE as the acceptor presented by either the natural (top) or the enantiomeric diketide (bottom). The turnover rates of both systems are presented in Figure 17C&19.

ciple account for substrate specificity of M6–TE and cause the above–mentioned phenomenon: either the KS6–catalyzed condensation reaction, or the following NADPH–dependent reduction of the β–ketoester by KR6, or the TE–catalyzed lactonization to the triketide product. While no information is available on the steady–state kinetics of each reaction step within M6–TE, previous analysis of the hydrolysis rate of EDK/NDK–SNAC[d] by the TE domain,[141] comparison of the reduction rates of two KR domains,[142] and kinetic analysis of the KS domain of DEBS,[140,143] have led to the conclusion that the KS–catalyzed decarboxylative condensation limits the rate of a given module.[143,144] In another project of my thesis I therefore targeted the KS active site to broaden the substrate tolerance of M6–TE

2.4.1 KS Active Site Mutagenesis — Previous Study and Current Approach

Active site mutagenesis to increase the turnover rate of PKSs, has been reported several times in the literature.[138,145] In one study, the substrate specificity of DEBS KS3 was engineered by introduction of single point mutants within the active site. Based on the observation that the KS domains of the mycolactone synthase

---

[d]SNAC; *N*–acetylcysteamine thioester

(MYCL) share a >97% sequence identity, despite accepting substrates of different length and chemical modification, single point mutations were introduced in DEBS KS3 replacing the DEBS residue with its counterpart from MYCL.[138] One mutant (A154W) turned out to be especially promiscuous and exhibited a 4.5–fold higher turnover rate with butyryl–SNAC compared to wild–type KS3. As A154 is located on the "dimer interface loop" of KS3, which shows no sequence conservation in most PKS domains (except MYCL), the authors proposed a potential role of this loop in substrate selection.[138]

The next logical step would be to increase the activity of the enzyme even further by screening sets of additional mutations in the single point–mutated PKS background. Directed evolution experiments revealed that on average 10 mutations are necessary to improve the activity of an enzyme by factor 1000,[146] indicating that several mutations will likely be necessary to increase turnover of non–native substrates by PKSs. Performing multipoint mutagenesis experiments in a random fashion in such a complex system however is tedious and of uncertain success as multipoint mutants show non–additive behavior of individual mutations, an effect termed mutational epistasis.[128,129] We therefore tried a more targeted approach to guide multipoint mutagenesis of KS, using the FuncLib server, a program that generates a variety of diverse multipoint mutants, by calculating stable networks of interacting active site residues.[147] FuncLib uses phylogenetic information to reveal possible, coupled mutations within a protein, by working on two levels: First, it uses phylogenetic information to suggest amino acid exchanges at selected positions, and, second, it ranks the mutated proteins by calculating their stability via the Rosetta program suite. In this process, mutations that are predicted to destabilize the protein fold are discarded. The FuncLib approach does not target individual substrates specificities, but delivers a set of stable variants that can than be screened for the specificities and activities of interest.

During a master thesis project supervised by me, Lynn Buyachuihan analyzed the applicability of FuncLib to broaden the substrate specificity of two DEBS–derived KS domains, whereby increasing the turnover rates of kinetically impaired bimodular chimeric PKSs.

2.4.2 Design of Multipoint KS Mutants

To test the feasibility of KS engineering to broaden the substrate specificity of chimeric PKSs via multipoint mutagenesis we decided to analyze two different bimodular PKSs using either M3–TE or M6–TE as the final acceptor (Figure 22). Similar to the bimodular PKSs shown in Figure 17 and Figure 19, we employed SYNZIP3+SYNZIP4 at the interface between the donor and acceptor module (Figure 22). To test the impact of different substrates on the engineered KS domains we used each acceptor module in the context of two different donor modules, which allowed us to query both the impact of different chimeric protein interfaces, as well as the effect of using NDK or EDK as the incoming substrate. To analyze SZ4–M3–TE or SZ4–M6–TE in the context of a chimeric chain translocation interface we used M1–SZ3 as the donor module, which naturally produces NDK on ACP1 and, in when used in combination with the two downstream modules, results in a chimeric ACP1:KS3/ACP1:KS6 interface (Figure 22A). In contrast, the chimeric chain elongation setup was tested using (5)KS1–AT1–KR2–ACP2–SZ3 as donor for M3 (Figure 22A) and (5)KS1–AT1–KR5–ACP5–SZ3 for M6 (Figure 22B). Both fusion modules produce EDK on ACP2 and ACP5 respectively. The only difference between these chain elongation chimeras and the previous setup (Figure 19) lies in the site of the SYNZIP domain. To retain maximum modularity we chose to install the SYNZIP domain at the ACP:KS rather than the AT:KR interface.

The choice of M3 and M6 was guided by their differing turnover rates with M1 (Figure 17), in case of M6, the previously observed substrate restrictions, and the
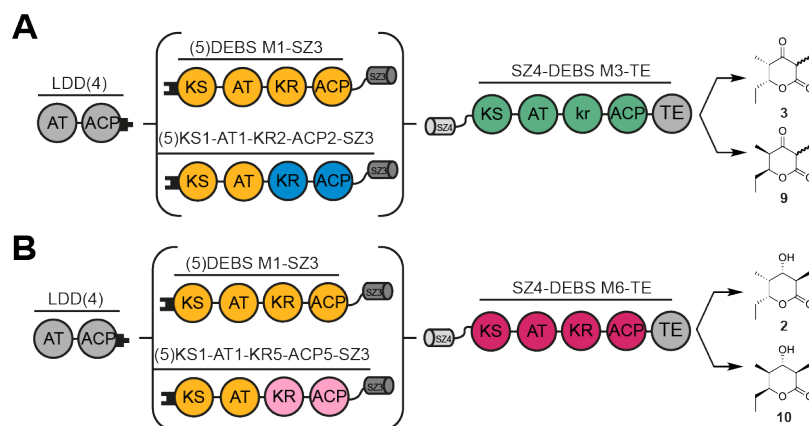
Figure 22 Design of bimodular chimeric PKSs to test the influence of KS substrate specificity on the turnover rate of chimeric PKSs. (A) Bimodular chimeric PKSs using M3 as the target module. Either (5)M1(2) was used as the donor module to create PKSs harboring a chimeric chain translocation interface and producing the (2*S*,3*R*)–2–methyl–3–hydroxy–diketide (NDK) leading to triketide lactone **3**, or (5)KS1–AT1–KR2–ACP2–SZ3 to create a chimeric chain elongation interface and producing the (2*R*,3*S*)–2–methyl–3–hydroxy–diketide (EDK) leading to triketide lactone **9**. (B) Bimodular chimeric PKSs using M6 as the target module. Either (5)M1(2) was used as the donor module to create PKSs harboring a chimeric chain translocation interface and producing NDK, leading to triketide lactone **2**, or (5)KS1–AT1–KR5–ACP5–SZ3 to create a chimeric chain elongation interface and producing EDK and leading to triketide lactone **10**. Protein sequences of donor new modules in Table S4.

availability of structural data. While X–ray structural data exists for KS3–AT3,[33,51] a homology model had to be generated for KS6–AT6. This gave us the opportunity to test the ability of the FuncLib server to calculate stable active sites both on the basis of a X–ray structure as well as an homology model. Since a minimal sequence identity of 40% is recommended for a homology model to be used in FuncLib, M6 with a sequence homology of 55.5% to KS3–AT3[33] and 55.2% to KS5–AT5[32] was a suitable candidate. In the following only the FuncLib results for M3 are presented, as the M6 mutants were still analyzed at the point of submission of this thesis. Preliminary data on M6 includes the analysis of wild–type M6 and the A154W point mutant with both donor modules.

To determine which residues of KS3 to target for mutagenesis, all those residues within 12 Å of the active site cysteine C175 were considered (Figure 23). Based on a multiple sequence alignment, residues that were non–conserved across a variety of KS domains from different PKSs were identified (alignment Figure 24, 12 Å–range in orange; Figure 23 non–conserved residues highlighted in green).
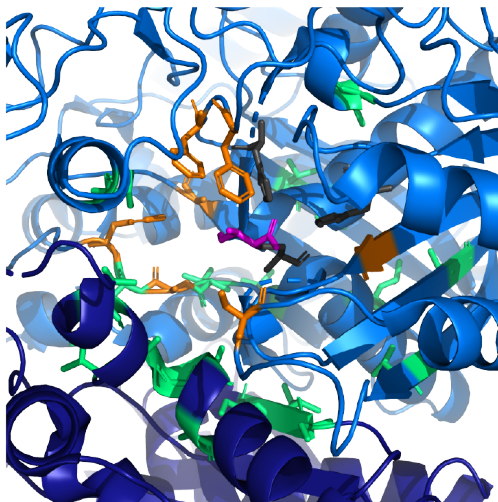
Figure 23 Active site analysis of KS3. Chain A/B in shades of blue, catalytic triad in gray, KS inhibitor cerulenin in black, non–conserved residues within 12 Å of the active site cysteine are highlighted in green, and residues selected for multipoint mutagenesis are shown in orange; PDB 2OQ3. Figure adapted from Lynn Buyachuihan.

Among those non–conserved residues within 12 Å of C175, seven residues were selected for mutagenesis (Figure 23, orange residues). The selection criteria was that residues were close to C175, pointed towards the center of the active site, were located on different parts of the active site, and showed promising results in the study by Murphy *et al.*.[138] The selected residues were A154, F156, A230, F263, F265, S306, and A441 based on the numbering in PDB 2OQ3 (Figure 24), which corresponds to A124, F126, A203, F236, F238, S279, and A416 if the entire KS sequence is considered (Figure 24).

Submission to FuncLib (with inclusion of the A154W mutation) resulted in a sequence space of 152,826 designs, harboring a minimal of three and a maximum number of five mutations. Table S5 lists the possible amino acids at each position. From the >150,000 designs, 1011 were calculated by Rosetta atomistic modeling to result in a more stable protein than the wild–type. From the 50 highest ranking designs, we chose 12 designs for subsequent analysis based on their large variability in mutated residues, and their range of different stabilities (pLB003 – pLB1015,

Table 4 Selection of M3 mutants based on FuncLib score. Mutants were selected to exhibit great variability and include some of the MYCL residues. In addition pLB001 carrying the A154W mutations was created based on previous results.[138] N/A: not applicable. The sequence space for each position is listed in Table S5.

plasmid	pos. 154	pos. 156	pos. 230	pos. 263	pos. 265	pos. 306	pos. 441	total score
pMK149 (wild–type)	A	F	A	F	F	S	A	-4682.351
pLB003	Q	Y	S	L	F	S	A	-4695.912
pLB004	L	Y	S	L	Y	S	A	-4692.154
pLB005	E	Y	T	Y	W	S	A	-4692.119
pLB006	Q	Y	T	L	M	S	A	-4692.054
pLB007	Q	K	A	F	W	T	S	-4691.481
pLB008	T	Q	A	L	F	P	A	-4691.025
pLB009	Q	N	A	L	F	R	A	-4690.721
pLB010	A	Y	T	A	F	T	S	-4690.53
pLB011	Q	Y	C	L	F	P	A	-4690.44
pLB013	Q	Y	S	F	F	P	A	-4690.227
pLB014	W	M	A	L	Y	S	A	-4689.953
pLB015	E	Y	A	L	Y	S	A	-4689.806
pLB001	W	F	A	F	F	S	A	N/A

Table 4). In addition, the previously reported point mutant A154W was included (pLB001, Table 4).

2.4.3 Generation of Multipoint KS Mutants and Turnover Analysis

From the 12 selected designs seven carried four mutations and five carried five mutations. A structural alignment of each calculated homology model with the wild–type structure predicted only minor structural rearrangements for most mutants (Figure 25; e.g. LB003, LB004, and LB010), with a few showing significantly altered active sites (Figure 25; LB005, LB009, and LB014).

Despite these differences *in silico*, all mutant proteins exhibited a similar purification behavior. Compared to the wild–type construct MK149, all mutant proteins could be purified at similar or higher yields (Table S6). Protein oligomerization, as measured by SEC, showed that besides Mut07 and Mut10 all proteins eluted in a
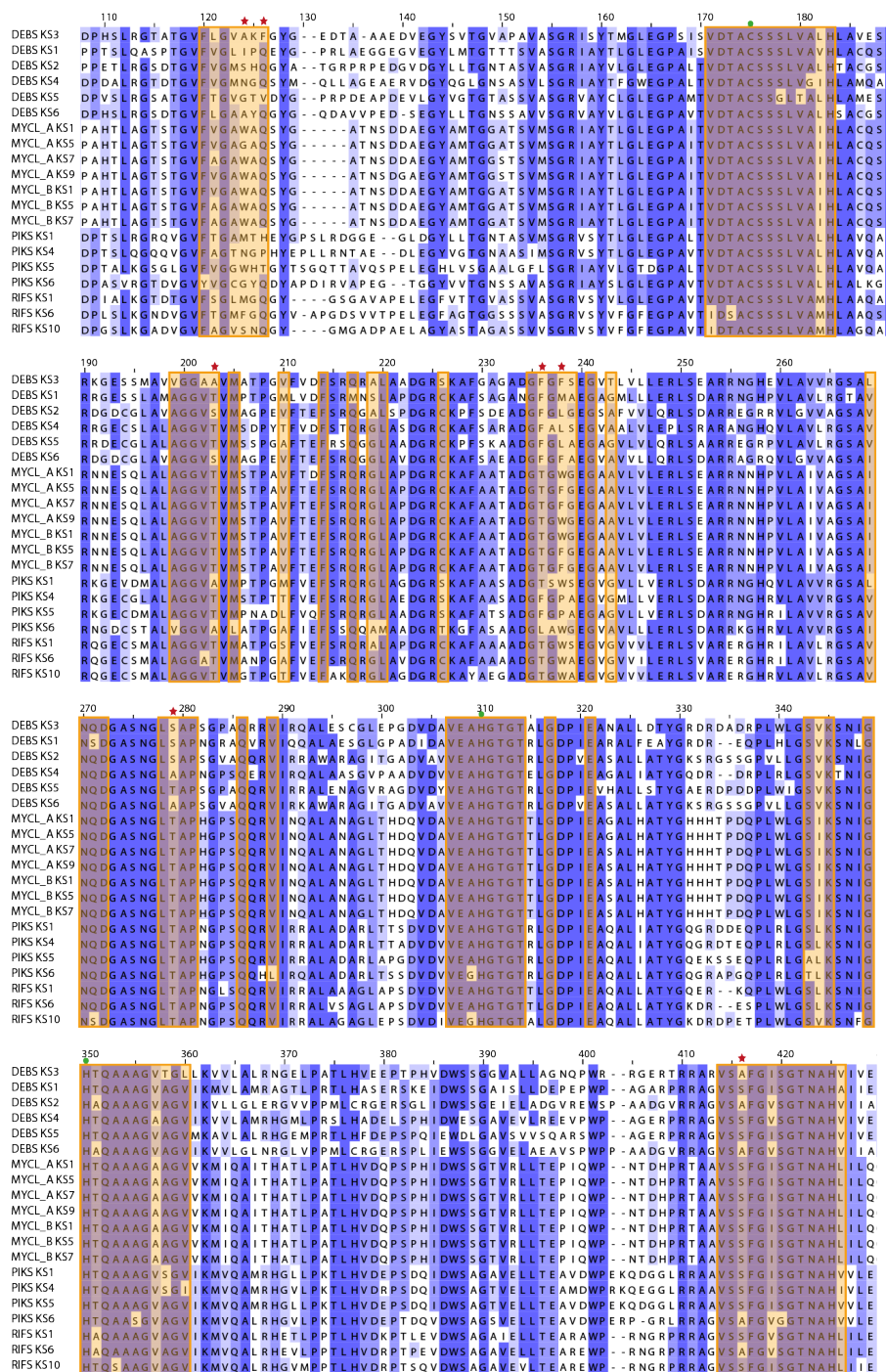
Figure 24 Sequence alignment of KS domains to identify residues for site–directed mutagenesis of DEBS KS3. Sequences were obtained from DEBS, mycolactone synthase (MYCL), PIKS, and RIFS. Residues within 12 Å of the active site cysteine of DEBS KS3 (C175) are highlighted in orange. Green circles; catalytic triad, red stars; residues selected for multipoint mutagenesis (A124, F126, A203, F236, F238, S279, and A416). Nomenclature of selected residues according to PDB 2OQ3: A154, F156, A230, F263, F265, S306, and A441. Figure adapted from Lynn Buyachuihan.

single peak (Figure 26). In addition, measurement of the thermal stability of each mutant indicated no significant difference between wild–type and mutant M3–TE (Table S7).

The turnover rate of all M3 mutants was first assessed in the bimodular PKS consisting of LDD(4), (5)M1–SZ3, and SZ4–M3–TE, a system harboring a chimeric ACP1:KS3 interface and presenting KS3 with NDK (Figure 27A). As measured before, the reference system using SZ4–M2–TE as the acceptor exhibited a ∼25–fold higher activity than wild–type M3. Compared to wild–type M3, the single point mutant Mut01 had a roughly 2–fold increased activity. Among the multipoint mutants, Mut06, Mut07, Mut10, Mut14, and Mut15 showed no improvement, or even loss of activity (Mut07, Figure 27A). A similar increase as for Mut01 was observed for Mut03 and Mut05, and an almost 3–fold increase in turnover rate was measured with Mut04, Mut08, Mut09, Mut11, and Mut13 (Figure 27A). In a similar way, the system using (5)KS1–AT1–KR2–ACP2–SZ3 as the donor module and thus harboring a chimeric chain elongation interface and presenting M3 with EDK, was analyzed (Figure 27B). Interestingly, the rate of wild–type M3 when using ACP2–EDK was slightly higher compared to the system using ACP1–NDK (Figure 27A&B, second columns), indicating the importance of ACP2:KS3 recognition for chain translocation, as M3–TE was previously shown to prefer NDK over EDK.[104,140] In this setup, mutants Mut03, Mut04, Mut06, and Mut10 showed no difference in turnover compared to wild–type M3, and the activity was abolished/decreased for Mut05 and Mut15 (Figure 27B). The multipoint mutant Mut14 which included the A154W mutation showed a 2–fold increase compared to wild–type, while the single point mutant without any additional mutation (Mut01) showed an even higher activity, emphasizing the effect of epistasis. The multipoint mutants Mut07, Mut08, Mut11, Mut13 exhibited the largest increase in turnover rate (by ∼3.5–fold), narrowing the gap towards the reference system to a 10–fold difference (Figure 27B).
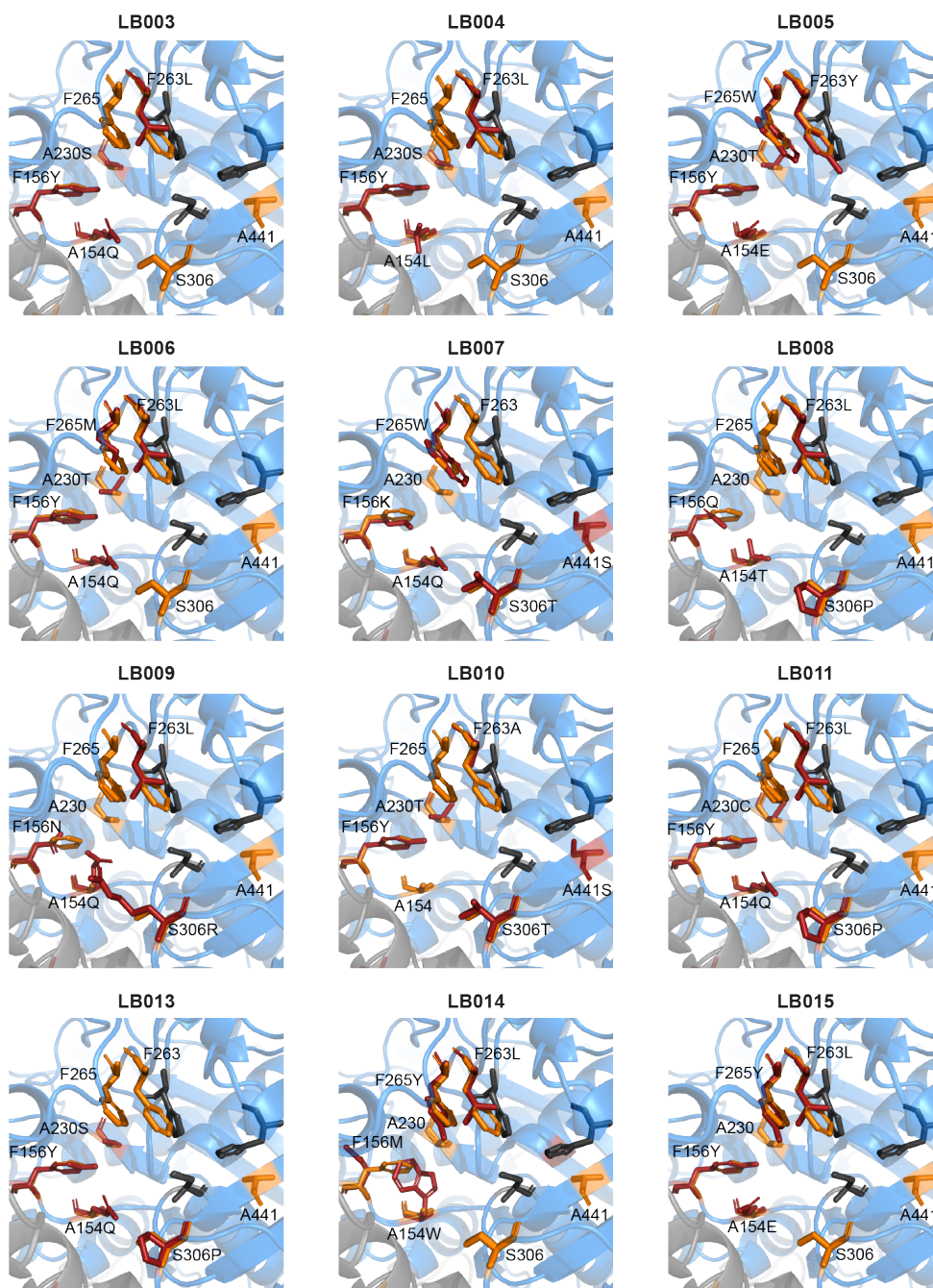
Figure 25 Alignment of active site residues from KS3 and the 12 chosen mutant designs. Catalytic triad in gray, residues selected for mutagenesis in orange, and mutated residues in each design in red. Wild–type PDB 2OQ3, and homology models provided by the FuncLib server for the mutants. Figure adapted from Lynn Buyachuihan.
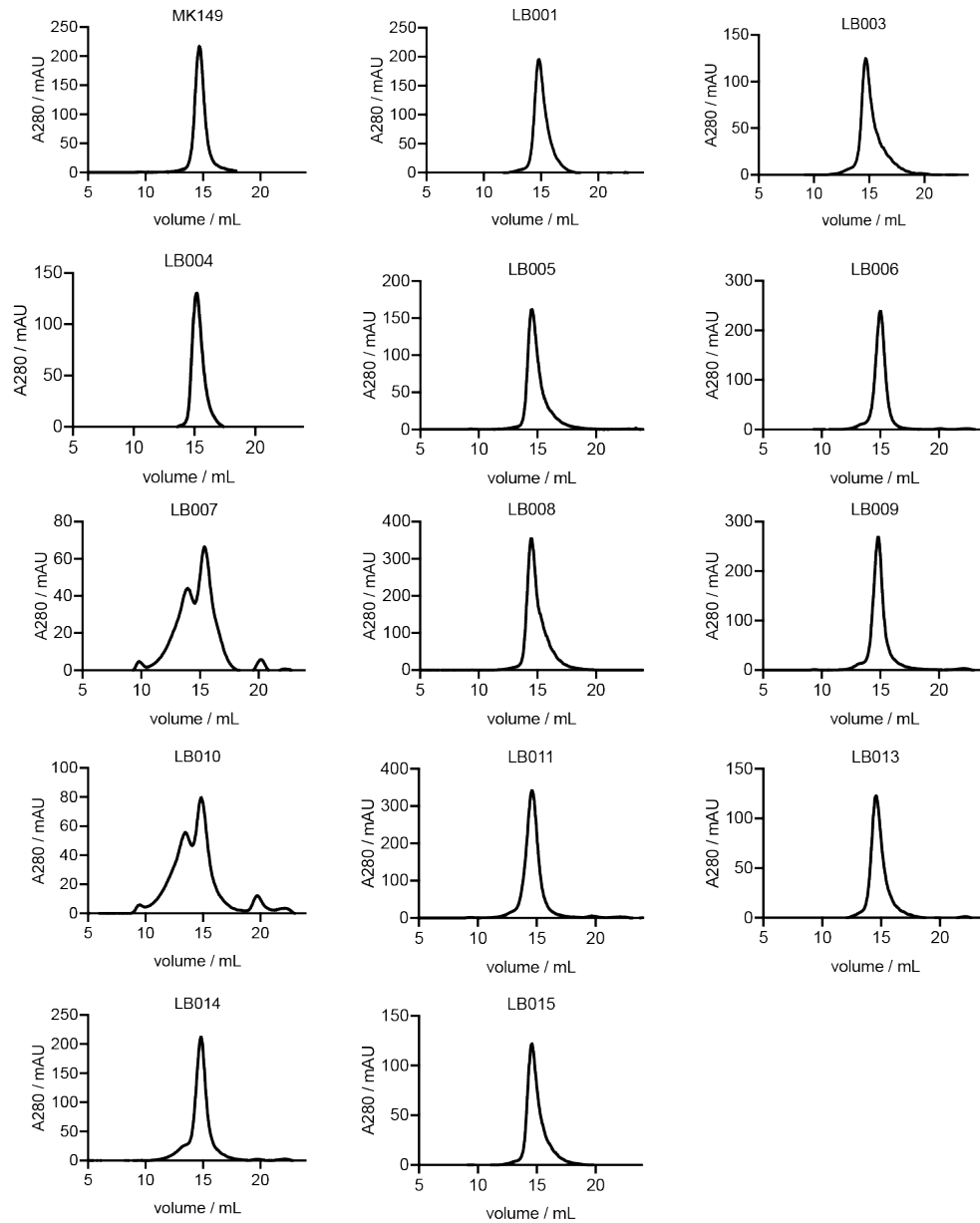
Figure 26 Quality of SZ4–M3–TE and its mutants analyzed by SEC. All proteins (except for LB007 and LB010) eluted in a predominantly single peak from SEC. MK149 is wild–type SZ4–M3–TE and LB001–LB015 correspond to SZ4–M3–TE_Mut01-Mut15. Figure adapted from Lynn Buyachuihan.
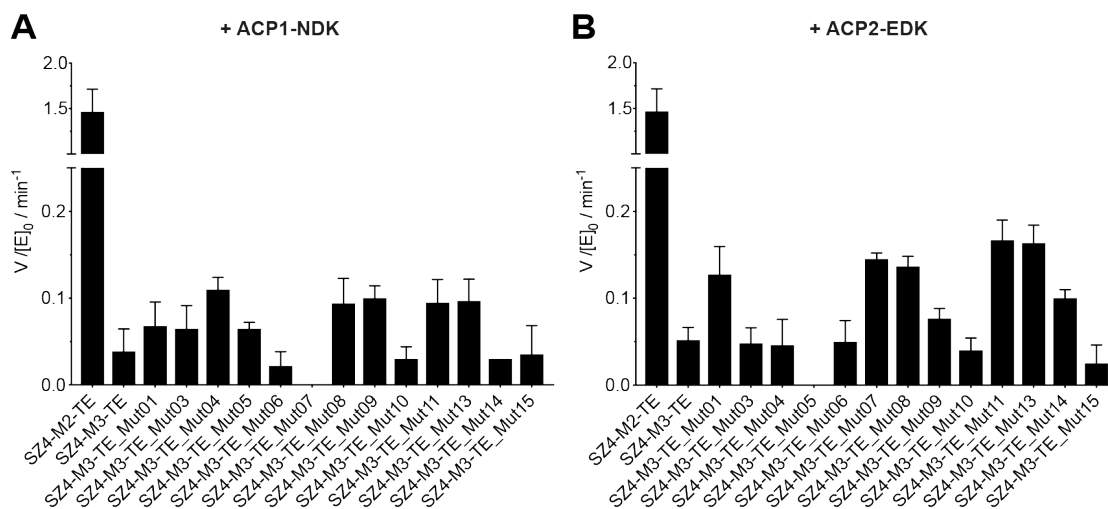
Figure 27 Turnover rates of bimodular PKSs using M3 multipoint mutants. (A) Turnover rates of bimodular PKSs harboring a chimeric chain translocation interface; consisting of LDD(4), (5)M1–SZ3, and SZ4–M3–TE (or variants thereof). (B) Turnover rates of bimodular PKSs harboring a chimeric chain elongation interface; consisting of LDD(4), (5)KS1–AT1–KR2–ACP2–SZ3, and SZ4–M3–TE (or mutants thereof). In both cases, the substrate identity (NDK/EDK) and the presenting ACP (ACP1/ACP2) are indicated. The turnover rate of the bimodular PKS LDD(4), (5)M1–Z3, and SZ4–M2–TE was used as a reference system in both cases. All initial rate data was obtained at 4 µM enzyme concentration and non–limiting concentrations of propionyl–CoA, methylmalonyl–CoA, and NADPH. Data was compiled from 2 to 9 measurements. Most data generated by Lynn Buyachuihan.

Among all mutants, Mut08, Mut11, and Mut13 are the most universal, as they improve the turnover rate of the bimodular PKS under usage of both NDK and EDK. While it has to be pointed out that the observed rates are a result of additive effects of all mutations, it is striking that from the 12 chosen designs, Mut08, Mut11, and Mut13 all include the S306P mutation (Table 4, Figure 25). S306 and F263 are located at the entrance of the active site tunnel (Figure 23), but showed different behavior in terms of their contributions to protein stability in FuncLib: Only minor changes were tolerated at F263 (mutation towards A,S, and Y), while S306 was highly diversified in the 1011 computed, stable designs (mutation towards A, E, G, K, M, N, P, Q, R, S, T, and V).

Analysis of M6–TE in the presence of either (5)M1–SZ3 or (5)KS1–AT1–KR5–ACP5–SZ3 showed similar results as before (compare Figure 28 to Figure 17B&19), confirming the preference of KS6 for NDK over EDK.[140] Additionally, both junction sites to generate PKSs with a chimeric chain elongation interface, namely by either
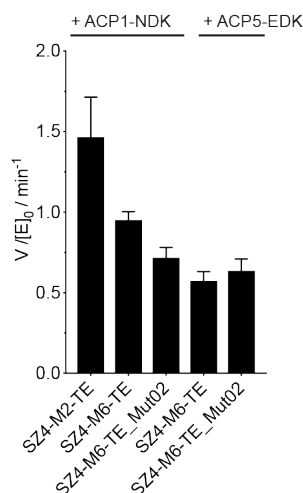
Figure 28   Turnover rates of bimodular PKSs using M6–TE. The turnover rates of bimodular PKSs consisting of LDD(4), (5)M1–SZ3, and SZ4–M6–TE/SZ4–M6–TE_A154W ("+ACP1–NDK") and using (5)KS1–AT1–KR5–ACP5–SZ3 ("+ACP5–EDK") as the donor module. The turnover rate of the bimodular PKSs LDD(4), (5)M1–Z3, and SZ4–M2–TE was used as a reference system. All initial rate data was obtained at 4 µM enzyme concentration and non–limiting concentrations of propionyl–CoA, methylmalonyl–CoA, and NADPH. Measurements were performed in triplicate. Data generated by Lynn Buyachuihan.

employing SYNZIP domains at the AT:KR interface (Figure 19) or by using a fusion protein such as (5)KS1–AT1–KR5–ACP5–SZ3 (Figure 28), produce similar results, emphasizing the possibility to use SYNZIP domains at different interfaces. Introduction of the A154W mutant, diminished the turnover rate of the system using (5)M1–SZ3 and did not significantly change the activity with (5)KS1–AT1–KR5–ACP5–SZ3 (Figure 28). The preference of KS6 for NDK, can be justified by the strong sequence similarity between KS2 and KS6 (Figure 18). As such it is unlikely that insertion of multipoint mutations can significantly improve the activity of M6 with NDK. In contrast, multipoint mutation of KS6 might facilitate reaction with ACP5–EDK.

2.4.4 Conclusion — Opportunities and Limitations of Multipoint Mutagenesis to Engineer KS Substrate Specificity

In this work, 12 multipoint mutants of KS3 were generated and tested for their ability to productively process two different substrates, presented by either the

natural upstream ACP or a chimeric ACP. Despite strong difference in active site residues, all mutants could be purified as stable proteins and with high yields, showing the success of FuncLib to calculate stable active site mutants on the basis of a protein structure. The calculated sequence space (Table S5) was retained in the 1011 stable designs, highlighting the large amount of possible active site residues that lead to the formation of a stable active site as determined by FuncLib. From the sequence space it can also be concluded that while some positions (e.g. A441/A230) do not seem amenable to site–directed mutagenesis, other positions harbor a higher potential for diversification (e.g. A154, F156, and S306, Table S5). Note that in our case the sequence space is mostly based on phylogenetic analysis as suggested by FuncLib. If desired, amino acids can be manually added within the Rosetta calculation of FuncLib, potentially leading to other stable exchanges at evolutionary conserved positions (an example is the inclusion of the A154W exchange, which was not suggested in FuncLib's initial sequence space). Based on the results obtained for KS3 it will be interesting to analyze whether the mutants based on the homology model of M6 will exhibit the same degree of stability.

Analysis of the turnover rate of bimodular PKSs using the mutant M3 acceptors revealed rate improvement or impairment for different mutants, which also differed in whether the upstream module harbored the chimeric ACP1 or the native ACP2, and/or NDK or EDK. A first interesting observation is that wild–type M3 showed a higher degree of turnover with ACP2–EDK than with ACP1–NDK, although is naturally prefers the NDK substrate.[140] The low turnover rates with ACP1–NDK as the donor have been previously observed to stem from a rate–limiting effect based on the non–native ACP1:KS3 interface (Chapter 2.1). Usage of ACP2–EDK overcomes these rate limitations, but introduces substrate recognition problems within KS3. While it was not possible with the current setup to assess the mutants ability for turnover with ACP2–NDK, it can be assumed that the rate improvements would be even greater with ACP2–NDK as the donor. Since different mutants showed different effects in the two analyzed systems (Figure

27), no generalized conclusion can be drawn from all mutants. Mut08, Mut11, and Mut13 exhibited strong rate improvements in both analyzed systems and thus far no higher turnover rate of a bimodular PKS using M3–TE has been measured. With the current setup the difference of the reference acceptor M2–TE and the best mutant M3–TE (Mut08, Mut11, and Mut13) has been decreased from a ∼25–fold higher activity to a 10–fold difference. As a next step, the identity of the product will be verified by LC–MS. Taken together, the data in Figure 27 suggests that the chimeric bimodular PKSs using M3–TE (i) benefits from retaining the native ACP2:KS3 interface, (ii) can be engineered towards higher turnover rates via KS mutagenesis, and (iii) that different mutants have differing effects based on the employed substrate.

The preliminary results on M6 confirm that mutagenesis of KS6 will most likely not improve activity of M6 presented with NDK by ACP1 (Figure 28), due to the inherent substrate preference of KS6 for NDK.[140] Yet, engineering the rate of turnover of M6 with NDK might be achieved by presenting NDK bound to the native upstream ACP. This could be achieved by creating a system consisting of LDD(4), (5)KS1–AT1–KR1–ACP5–SZ3, and SZ4–M6–TE. On the contrary, the activity of M6 with ACP5–EDK shows a strong potential for improvement by active site mutagenesis (Figure 28).

Taken together our results highlight the potential of KS active site mutagenesis in broadening the substrate tolerance of a PKS module, as well as the benefit of multipoint mutagenesis over single point mutants. By screening a set of only 12 mutants, we were able to identify five mutants with higher activity towards ACP1–NDK and four mutants with improved turnover with ACP2–EDK.

2.5 Describing the Structure of a Polyketide Synthase Module in Solution

As the last goal of this thesis, I wanted to gain a deeper structural understanding of domain–domain interactions in PKS modules and the degree of conformational flexibility of individual domains. By combining *in vitro* and *in silico* techniques, namely SEC–SAXS (size exclusion chromatography coupled to small–angle X–ray scattering), cross–link mass spectrometry (XL–MS), and structural modeling, we aimed at describing different conformations of a PKS module in solution. This study was carried out in collaboration with the Urlaub group at the MPI for Biophysical Chemistry in Göttingen (with Andreas Linden) and the Hummer group at the MPI for Biophysics in Frankfurt (with Dr. Emanuele Rossini). Most of the text in this part has been prepared for future publication.

Thus far engineering of PKSs has often proven difficult due to the limited structural understanding of domain–domain and module–module interactions as well as protein–substrate interaction with PKS modules.[106] As outlined in the introduction, PKS structure elucidation by common techniques such as X–ray crystallography, cryo–EM, SAXS or NMR has only been possible in limited cases due to the inherent flexibility of PKS modules and their large size.[54] Until today it has not been possible to obtain crystals from an intact PKS module, and so far only one PKS module could be analyzed by cryo–EM.[47] While X–ray crystallographic data is highly valuable due to the atomic–level resolution, it only provides static complex information and generally lacks information on dynamics and conformational changes. Cryo–EM on the other hand only allows for structure elucidation at lower resolution (> 3Å)[148]. Both X–ray crystallography and cryo–EM can give insight into protein conformational dynamics by merging several static conformations into a dynamic picture. SAXS can also capture the dynamic behavior of a protein but at a resolution too low to allow *de novo* structure elucidation (resolution >10Å).[149]

In addition, cryo–EM and SAXS data can be combined with high resolution X–ray structures to build a new model. In the field of PKS structure elucidation, NMR was only used for small domains such as ACPs[36,150] and docking domains,[22,23] and due to its size limitation cannot be applied to elucidate the structure of an intact PKS module.[151] As outlined in the introduction two different models for the appearance of a partially reducing PKS module are discussed in the literature. A DEBS–derived model was generated via SAXS, in which KS–AT adopts an extended conformation,[46] which is contradicted by a model of a PIKS module, in which KS–AT adopts an arched conformation,[47] highlighting the necessity of additional structural information to understand the interplay of PKS domains and to guide future PKS engineering. A detailed comparison of both models is given in Section 1.3.1.

2.5.1 The Use of XL–MS to Analyze Stable and Transient Protein Complexes

In this study, we employed XL–MS to obtain distance information on interacting domains and the conformational flexibility within PKS modules. XL–MS has not been used in the field of PKS structure elucidation prior to this thesis, but due to its ability to capture flexible and transient interactions it harbors great potential to improve our understanding of interactions within PKS modules and PKS assembly lines.

In an XL–MS experiment a bifunctional cross–linking reagent is used to covalently connect residues in close spatial proximity, whereby providing distance informa-tion of interacting residues. Cross–linkers exist as homo– or heterobifunctional reagents, reacting either with specific functional groups such as amines, carboxylic acids, and sulfhydryls groups, or unspecifically in the case of photoactivatable cross–linkers.[152] Alternatively, photoreactive cross–linkers can also be encoded genetically to study the spatial environment of a residue.[153] Most XL–MS ex-periments use cross–linkers that target the primary amino group of lysines via

N–hydroxysuccinimide (NHS) activated esters,[154] due to the high prevalence of lysine residues ($\sim$6% of all residues) and the high specificity of the cross–linking reaction.[155] Another consideration in the selection of the cross–linking reagent is the spacer length between the two reactive groups, specifying the distance of interacting residues. Due to the unspecific reactivity of photoreactive groups such as diazarines, these reagents allow for sampling of greater surface area,[156] but might complicate database searching.[157] XL–MS can provide information about flexible, transient, and stable interactions within proteins and protein complexes. The major challenges lie in the complex tandem mass–spectrometric analysis of fragmented peptides, enrichment of cross–linked peptides from complex peptide mixtures, and the subsequent database search that requires potent algorithms to identify the desired peptides from complex mixtures.[157] As such, since its first introduction in 2000,[158] the method has undergone severe development[159,160] until the first major breakthrough in 2010 when the 670 kDa large RNA Polymerase II–transcription factor initiation complex (RNA Pol II–TFIIF) was revealed using XL–MS.[161] Later, XL–MS studies have been employed to study a variety of large protein complexes such as the 26S Proteasome,[162–165] the nuclear pore complex,[166–169] the human spliceosome,[170–172] or individual RNA polymerases or RNA polymerase complexes.[173–178] In all cases XL–MS was applied in addition to other structure elucidation techniques such as X–ray crystallography, cryo–EM or SAXS as *de novo* computational prediction still remains a challenge.[179]

The lack of structures capturing the interaction between PKS domains prompted us to combine SAXS and XL–MS measurements to gain information on the overall conformational flexibility of a PKS module as well as on transient and stable domain–domain interactions within. In addition, we hoped that our analysis would shed light on the propensity of a PKS module to adopt the extended vs. arched conformation, and as such contribute to the ongoing efforts of structure elucidation of an intact PKS modules.

As neither SAXS, nor XL–MS alone are sufficient for *de novo* structure elucidation, we employed structural modeling to sample a variety of conformations that would agree with both experimental data sets. Our *in silico* approach harbors the additional advantage of enabling us to characterize transient domain–domain interactions. This approach is exemplified by a recent study in which we employed structural modeling to gain a deeper understanding of the relation between ACP:KS domain interactions and fatty acid chain length in the *Corynebacterium ammoniagenes* fatty acid synthase. By analyzing the product spectrum and enzymatic activity of FAS with differing KS surface mutations, in conjunction with calculating binding affinities between ACP and the different mutant KS domains, we were able to determine a correlation between binding affinities and FAS activities.[107] Our results highlight the potential of *in silico* studies to contribute towards a better understanding of domain–domain interactions.

2.5.2 Optimization of the Protein Purification Protocol via Codon Harmonization

To obtain high quality data in both SAXS and XL–MS measurements optimization of the protein purification procedure was required.

Initially the idea behind this project was to use XL–MS to map residues that interact within one module but also across module boundaries. We therefore aimed at studying the interaction between DEBS M2 and DEBS M3, in addition to each individual module, specifically a M2(2) and a (3)M3–TE construct. The M3–TE construct (RSG34) was purified many times with yields around 9 mg/L of culture, while various M2 construct were purified at much lower yields of 0.5–1 mg/L of culture.[66,75,104,105] To optimize the protein purification protocol for M2, we altered the gene sequence encoding for M2 in a process termed "codon harmonization" to improve protein folding and stability. Codon harmonization is used to optimize translation by introduction of synonymous codons that have a similar codon usage frequency in the native and target host.[180,181]

In our case the DEBS gene cluster was obtained from *Saccharopolyspora erythraea*,[182] but the protein was recombinantly purified from *E. coli*. As the codon usage between these two organisms differs drastically (compare Tables S8 and S9), any effect of co–translational folding due to transient ribosomal attenuation will be lost if the *S. erythraea* gene sequence is transcribed and translated in *E. coli*. The influence of codon usage on translation efficiency, translation elongation, mRNA stability, and protein folding has been intensively discussed in the literature.[183,184] In brief, variation in translation rates is observed between frequently and rarely used codons,[181,185,186] due to a variability in the abundance of the respective tRNA levels.[187] It has been shown that discontinuous translation has an influence on translational regulation[188] and can facilitate co–translational folding of proteins.[189–193] The effect of cotranslational folding on PKS polypeptides has not been discussed previously, but the multidomain architecture of PKS modules, their large size, and frequent occurance as bimodules makes them suitable targets for the application of gene harmonization to improve protein purification via (potentially) improved protein folding.

In collaboration with the Schwalbe group (Goethe University Frankfurt) the gene sequence of M2 was harmonized to reflect the codon usage rates from *S. erythraea* in *E. coli* (Figure 29). Comparison of the translation rates profile before and after harmonization (Figure 29B&C) revealed a strong difference in the translational rate within and between domains (Figure 29A). As the harmonization resulted in the generation of some especially slow translating stretches ($\sim$ residue 200 and 550; Figure 29C), the sequence was manually adjusted to avoid harsh spikes in the translation rate profile (Figure 29D).
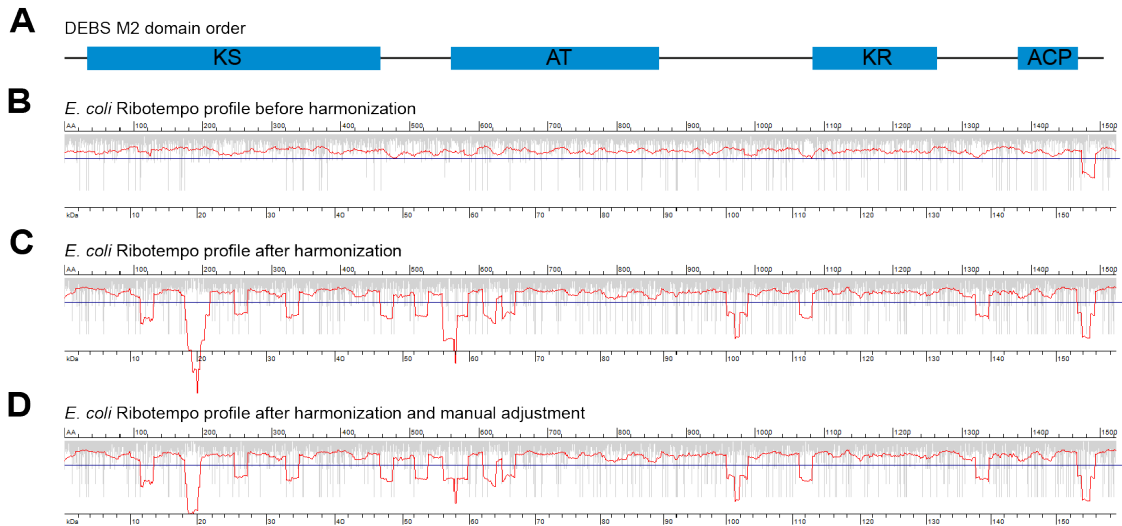
Figure 29 The effect of codon harmonization on the translation rate of M2. (A) M2 domain architecture. Prediction of the translation rate in *E. coli* before harmonization (B), after harmonization (C), and after manual adjustment of the harmonized sequence (D). Translation rates were predicted using the *Ribotempo* web server.[191] Manual adjustment of the harmonized sequence by Prof. Zoya Ignatova (University Hamburg). For gene sequences of original and harmonized gene see Table S10.

To test the success of this approach, a side–by–side purification of two M2 constructs differing only in the M2 gene sequence was conducted using two individually grown cultures (data generated by Ilka Siebels). For the KR1–ACP1–M2–TE construct (ADD03) an average yield of 0.52 mg/L of culture was obtained, whereas for KR1–ACP1–M2$_{harm}$–TE (ADD01) the overall yield was 0.75 mg/L, corresponding to an almost 40% increase in protein yield. Subsequently, different M2$_{harm}$ constructs and (3)M3–TE were purified to high quality for SEC–SAXS analysis (Figure 30).
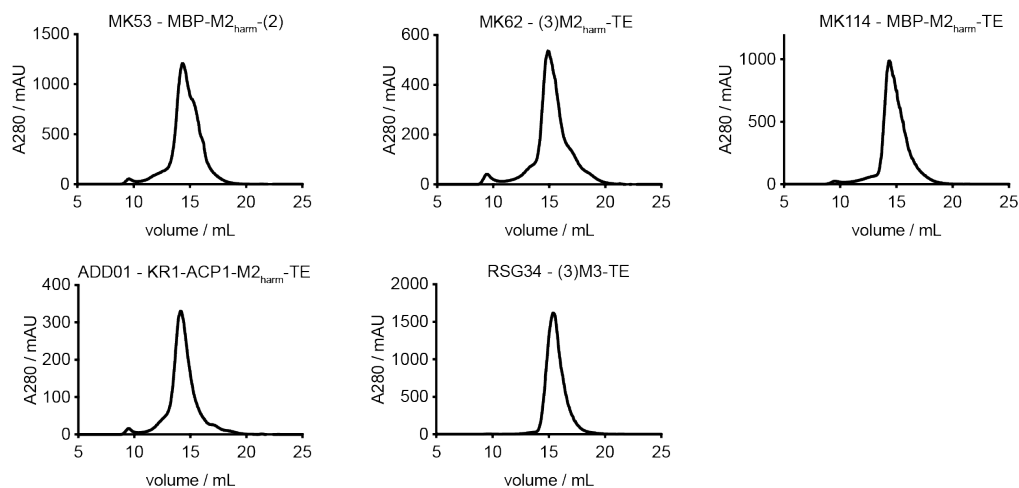
Figure 30 Purity of proteins submitted for SAXS analysis. Protein abbreviations and constructs are indicated. All proteins are pure and eluted in a predominately single peak from SEC. Molecular weight of analyzed constructs: MK53 – 201.5 kDa, MK62 – 185.4 kDa, MK114 – 223.5 kDa, ADD01 – 242.2 kDa, and RSG34 – 186.0 kDa.

2.5.3 SAXS Analysis of DEBS Constructs

Tandem SEC–SAXS analysis was performed on a set of four different $M2_{harm}$ constructs and the previously analyzed M3–TE construct.[46] All proteins eluted in a single peak of the expected homodimeric mass (Figure S5). Analysis of the scattering intensity and Kratky plots (Figure S6) showed similar, yet not identical curve patterns for the different constructs.

For M3–TE, the derived R_g value (6.4 nm) matches well with the previously published R_g value of 6.1 nm.[46] In good agreement with the data on (3)M3–TE, the R_g value of an architecturally similar construct based on M2 ((3)M2–TE) was calculated to be 6.44 nm (Table 5). The similarity in both R_g and D_{max} suggests a similar architecture of the two constructs which would be expected as both consists of similar domains. The larger constructs showed an interesting trend in which the two constructs harboring the TE domain (MBP–M2–TE and KR1–ACP1–M2–TE) exhibited similar R_g and D_{max} values (R_g (MBP–M2–TE) = 7.5 nm, R_g (KR1–ACP1–M2–TE) = 7.57 nm, D_{max} (MBP–M2–TE) = 23.0 nm, and D_{max} (KR1–ACP1–M2–TE) = 23.5 nm) revealing a higher degree of compactness than MBP–M2 (MK53) for

Table 5 SAXS data collection and scattering derived parameters for different M2$_{harm}$–constructs and M3–TE

	MBP– M2$_{harm}$(2) (MK53)	(3)M2$_{harm}$- TE (MK62)	MBP– M2$_{harm}$–TE (MK114)	KR1–ACP1– M2$_{harm}$–TE (ADD01)	(3)M3–TE (RSG34)
Data collection parameters					
Beamline	BM29	BM29	BM29	BM29	BM29
Type of experiment	SEC–SAXS	SEC–SAXS	SEC–SAXS	SEC–SAXS	SEC–SAXS
Wavelength /Å	0.992	0.992	0.992	0.992	0.992
Detector distance /m	2.87	2.87	2.87	2.87	2.87
q range /Å^{-1}	0.005 - 0.028	0.007 - 0.04	0.006 - 0.028	0.006- 0.028	0.0106 - 0.038
Exposure time /s	1	1	1	1	1
Temperature /°C	20	20	20	20	20
SEC parameters					
SEC column	Superose 6 increase 10/300	Superose 6 increase 10/300	Superose 6 increase 10/300	Superose 6 increase 10/300	Superose 6 increase 10/300
Amount loaded /nmol	2.54	1.68	2.00	2.02	2.69
Flow rate /mL/min	0.6	0.6	0.6	0.6	0.6
Structural parameters					
$I(0)$ from $P(r)$	85.6	70.4	91.7	89.27	102
R_g from $P(r)$ /nm	8.0	6.44	7.58	7.66	6.4
$I(0)$ from Guinier	85.6	70.1	92	89.23	102
R_g from Guinier /nm	8.0	6.35	7.5	7.57	6.4
D_{max} from $P(r)$ /nm	28.3	21.9	23.0	23.5	22.4
Software					
Data processing	PRIMUS	PRIMUS	PRIMUS	PRIMUS	PRIMUS

Abbreviations: R$_g$, radius of gyration; D_{max}, maximal particle dimension

which both the R_g (8.0 m) and D_{max} (28.3 nm) where considerably higher. The smaller D_{max} of the TE–containing constructs might reflect a higher degree of conformational restraints of the KR domain, caused by dimerization of the TE domain (homodimer MW = 66 kDa).[194] On the other hand, the flexibility of the KR domain in the MBP–M2(2)construct might be less constraint by the dimerization of the smaller docking domain (homodimer MW = 17 kDa).[22] In summary, we were able to obtain high quality SEC–SAXS data for all our constructs, which we next used to validate *in silico* models.

2.5.4 XL–MS Analysis and Preliminary Structural Modeling of DEBS M2

At the beginning of the cross–linking experiment we wanted to test both the homo-bifunctional cross–linkers bis(sulfosuccinimidyl)suberate (BS^3) and disuccinimidyl suberate (DSS), as well as the heterobifunctional reagent NHS–Diazirine/succinimi-dyl 4,4–azipentanoate (SDA; Figure 31A). Cross–linking efficiency of BS^3 was tested on MBP–M2(2), (3)M3–TE, and on a mixture of M2+M3 (Figure 31B). The highest cross–linking efficiency (especially for the M2+M3 mixture) was observed at 2 mM BS^3, a concentration that was used in all future cross–linking experiments. Despite the high cross–linking efficiency as judged by the in–gel analysis, only few cross–linked peptides were observed for either BS^3 or DSS when used on M2 and no hits were observed for M3. This was ascribed to the low number of lysine residues present in the DEBS proteins. To increase the number of obtained cross–linked peptides we employed the photoactivatable reagent SDA, which can both react with primary amino groups of e.g. lysines and insert under radical formation into C–C and C–heteroatom bonds.[195] By taking into consideration the lysine chain length, the 3.9 Å spacer, the chain length of long amino acids (e.g. lysine and arginine), and conformational flexibility, a maximum C_α–C_α distance of 20 Å is achieved.[196] While we obtained significantly more hits for MBP–M2(2) cross–linked with SDA, again very few hits were obtained for M3–TE. We thus

decided to abandon any attempts to cross–link M3 or the M2–M3 complex and focused solely on MBP–M2(2).
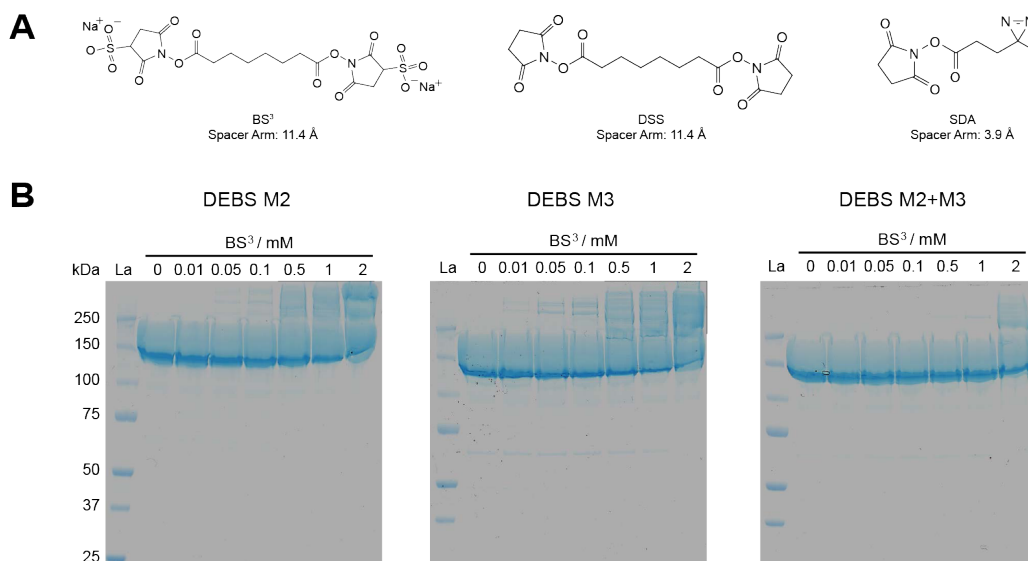
Figure 31 Cross–linking reagents and SDS–PAGE analysis of cross–linking efficiency of DEBS modules M2 and M3. (A) Cross–linking reagents used in this thesis. BS^3 – bis(sulfosuccinimidyl)suberate, DSS – disuccinimidyl suberate, SDA – NHS–Diazirine/succinimidyl 4,4–azipentanoate. (B) SDS–PAGE analysis of crosslinking efficiency: 30 μg of protein (or equimolar mixtures thereof) were cross–linked for 30 min with different amounts of BS^3 and afterwards analyzed via SDS–PAGE.

In four independent experiments we cross–linked MBP–M2(2) either at a high concentration of 14 mg/mL or at a lower concentration of 1 mg/mL with 2 mM SDA. After manual evaluation of each obtained spectrum, we sought to filter out all hits that showed a strong overlap in all four experiments. To this end, we assembled all inter–domain hits that appeared multiple times in multiple experiments and assigned them a strong relevance. In addition we searched for hits that cluster with "strong relevance" hits or among themselves and assigned them a "medium relevance". While "strong relevance" hits were significant due to their reproducible strong single–pair signal in our experiments, "medium relevance" hits denote important clusters of interacting residues rather than specific interaction pairs. Our results are divided into those hits stemming from lysine residues of DEBS M2 (Table S12) and those from lysine residues of MBP (Table S13). In addition, many more intra–domain cross–links were obtained that cannot be used

to evaluate the overall domain organization of the PKS module, but highlight the high prevalence of cross–links within one domain (Tables S14 and S15). It is noteworthy that no correlation was observed between the number of hits and the employed M2 concentration, indicating that no artifacts were introduced by crowding effects. Encouragingly, while we observed a large number of cross–links between the flexible ACP domain and catalytic domains, we also obtained a number of cross–links between the catalytic, less flexible domains, e.g. between KS:AT and KS:KR (lysine on KS surface K727), between AT:KS–AT linker and AT:KR (lysine on AT surface K1071), and between K1273, a lysine located on the post–AT linker interacting with the KS–AT linker, AT and KR (Table S12). The latter inter–domain cross–links are especially useful in validating structural models, as they provide distance information between those domains that define the scaffold of the PKS module. In addition to MBP–M2(2), we also analyzed a KR1–ACP1–M2– TE (ADD01) construct via XL–MS. In this case only two independent experiments were performed giving rise to a small set of inter–domain hits (Table S16).

Based on the experimental data, structural modeling of MBP–M2(2) was per- formed. While X–ray crystallographic data exists for MBP[197] and a NMR structure of ACP2[36], homology models had to be generated for KS2, the KS2–AT2 linker (LD), AT2, KR2, and docking domain 2 (DD2). Based on a sequence alignment generated by MAFFT,[198] the homology models were generated by MODELLER.[199] On the basis of the homology models, two initial models of MBP–M2(2) were generated, one starting from the extended conformation of KS–AT and one from the arched conformation. Generation of the initial arched model was done by fitting individual domains into the previously reported cryo–EM densities.[47] For the extended conformation, SAXS envelopes were obtained through the *ab initio* bead modeling program DAMMIN.[200] For both conformations, *in silico* scattering intensities calculated via FoXS[201] were derived from coarse–grained simulations using COMPLEXES++.[202] Finally, sample refinement was done with BioEn.[203]
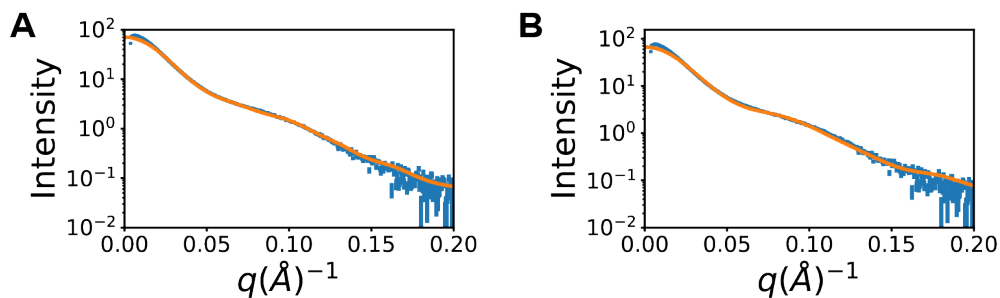
Figure 32 Comparison of scattering intensity plots log(*I*) *vs q* of experimental data and *in silico* computed data. Scattering curves based on models adopting the extended conformation (A) and the arched conformation(B). Experimental data (blue), *in silico* computed curves (orange). Analysis and figure generation by Dr. Emanuele Rossini.

To obtain a starting model for the extended conformation, all domains were fitted individually into the densities, with KS2–LD2–AT2 adopting the extended confirmation as in the crystal structures.[32,33] In this case, KS2–LD2–AT2 was treated as a single, rigid domain while all other domains were free to move within the constraints imposed by the inter–domain linker sequences. In contrast, for the starting model of the arched conformation, KS2, LD2, and AT2 were treated as individual domains, as this model is not supported by previous X–ray crystallographic data. The aim of the coarse–grained simulation was to define conformations for MBP–M2(2), for which *in silico* computed scattering curves agreed with the experimental data.

While we were able to identify conformations that largely agreed with the scattering intensities, the quality of the fit degenerated systematically at the Guinier region (large angles of the scattering curve, Figure 32; difference of orange (computed) and blue (experimental) curves), indicative of a small amount of aggregate in the protein sample. For both the extended and the arched conformation, BioEn identified two configurations, that as an ensemble (two configuration for the extended and separately two for the arched configuration) resulted in the *in silico* SAXS intensity plots and preliminary models (extended conformation, Figure 32A&33A and arched conformation, Figure 32B&33B). As such, based on the SAXS data alone, models could be generated that fulfill both the extended and
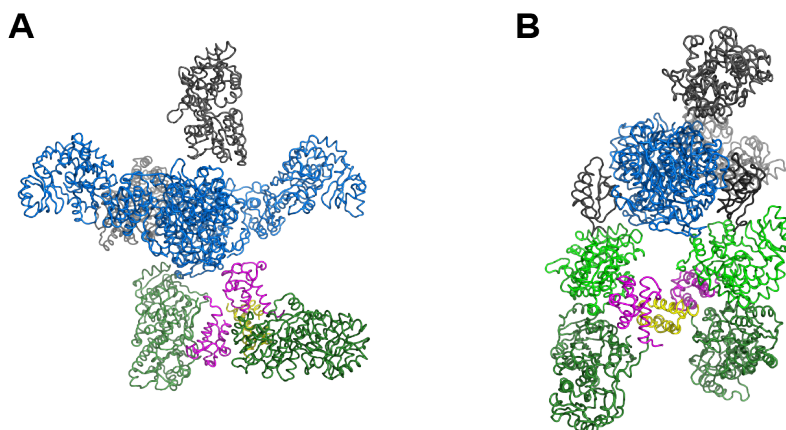
Figure 33 Preliminary coarse–grained models of MBP–M2(2). Model based on the extended conformation (A) and arched conformation (B). Domain coloring: MBP; light gray, KS; blue, LD, light gray, AT; green, KR; dark geen, ACP; magenta, DD; yellow. In case of extended conformation KS–LD–AT in blue. Note that the models are not validated by XL–MS hits, nor was symmetrization performed. Models generated by Dr. Emanuele Rossini.

the arched confirmation (Figure 33). For the extended conformation an R_g value of 69.8 Å and for the arched conformation an R_g value of 67.7 Å was computed. Preliminary attempts to account *in silico* for a small amount of aggregation in our sample preparation resulted in R_g values of 75.3 Å and 74.3 Å for the extended and arched conformation respectively, approaching the experimentally obtained R_g value (Table 5).

2.5.5 Outlook — Continuation of our Modeling Efforts

In this collaborative project we sought to combine SEC–SAXS, XL–MS, and structural modeling to deepen our understanding of domain–domain interactions in PKS modules as well as their overall appearance. Based on a codon harmonized M2 construct, we were able to obtain high quality SAXS data, agreeing with previously reported scattering data (Table 5).[46,51] XL–MS analysis was performed using a photoreactive cross–linker on two M2 constructs (MBP–M2(2) and KR1–ACP1–M2–TE), yielding a set of reliable inter–domain cross–links thereby providing for the first time distance information of specific residues within a PKS module (Table S12).

Preliminary coarse–grained modeling resulted in the generation of two different models for MBP–M2(2), one for each the extended and arched conformation (Figure 33). Interestingly, both conformations were able to explain the measured experimental scattering intensities (Figure 32). Neither of the two models has been refined by symmetrization (see non–symmetrical KR orientations in Figure 33A) nor has the cross–linking data been included for evaluation. Inclusion of the cross–linking hits in each model will be a crucial layer of information to further narrow down potential conformations of M2(2), potentially excluding one conformation, or revealing an equal likelihood for both conformations. As our models were set up in an unbiased manner, our analysis will enable a site–by–site comparison of both models based on a single experimental data set.

3 Conclusion and Outlook

Modular polyketide synthases are a fascinating class of mega–enzymes that produce polyketides, which are characterized by complex structures and high biological activity. The modular architecture of assembly line PKSs harbors the potential to engineer new, complex natural compounds via the exchange of individual modules or single domains. However, despite the engineering potential modular PKS systems exhibit at first glance, it has become more and more apparent that their activity relies on the efficient interplay of their modules via fine–tuned, specific and controlled protein–protein interactions at module interfaces and the substrate selectivity of their catalytic domains. Numerous engineering efforts have shown the current limitations of engineering chimeric PKSs, highlighted by decreased turnover rates and diminished product yields compared to the respective wild–type enzymes. Instead of simply swapping modules or domains, a detailed understanding of specificity determining protein–protein and protein–substrate interactions is thus necessary to fully harness the engineering potential of modular PKSs.

This is where the work presented in my dissertation started, aiming at analyzing the relative influence of protein–protein and enzyme–substrate recognition on the turnover of chimeric PKSs (Chapter 2.1), and, as they were found to be crucial in limiting the chain translocation reaction, examining the important ACP:KS interactions (Chapter 2.2), but also aiming to explore alternative protein interfaces for module exchanges and new tools to combine PKS modules or fragments (Chapter 2.3). As substrate selectivity also plays an important role in determining

the catalytic activity of chimeric PKSs, a combination approach was undertaken to analyze the relative contribution of protein–protein interactions and substrate promiscuity, via KS multipoint mutagenesis (Chapter 2.4). The different engineering approaches targeted different levels of polyketide biosynthesis, namely specific protein–protein interactions, traditional and novel interfaces to combine modules, and substrate promiscuity of KS domains. All engineering attempts were guided by structural consideration such as the impact on the PKS fold or active site stability. Thus, in addition, due to the lack of structural understanding of domain–domain interactions within and of the overall architecture of a PKS module, a collaborative project was set up to investigate different conformations of a PKS module in solution and defining ACP:domain interfaces for future engineering projects (Chapter 2.5).

An overarching focus of this thesis was on the KS domain, which was found to be crucial to chimeric PKS activity. ACP:KS interactions are finely tuned and often impact engineering attempts. They can impair the catalytic activity of a PKS module either as chimeric ACP:KS interactions in the chain translocation reaction (Chapter 2.1), but also at chimeric KS:ACP interfaces in the chain elongation reaction. From my analysis of different interfaces for combining PKS modules, the chain elongation interface emerged as more amenable to engineer than the chain translocation reaction (Chapter 2.3). It also became apparent that the chain translocation reaction is difficult to improve, as ACP:KS interfaces are not well enough defined and exhibit only weak affinities which hinder any attempts for affinity maturation via directed evolution experiments (Chapter 2.2). In addition, the generation of a non–native chain translocation interface is often accompanied by the introduction of a non–native substrate, convoluting any effects of the non–native ACP:KS interface with those of the substrate specificity of the KS domain. This mixture of rate–influencing effects causes difficulties in analyzing the relative contribution of each effect. During the course of my thesis, a KS mutagenesis study was published that analyzed the effect of different active site mutations on the sub-

strate tolerance of a KS domain.[138] This inspired me to design a KS mutagenesis project that, based on my previous findings on the generation of fusion proteins and the applicability of SYNZIP domains, allowed us to probe the effect of both the non–native substrate and the chimeric vs. native ACP:KS protein–protein interface (Chapter 2.4). Through these different engineering approaches a picture emerged in which engineering of chimeric PKSs is, at least on the basis of our current knowledge, not feasible by generalizable approaches but warrants a detailed analysis of the specific chimeric protein–protein interface and the substrate promiscuity of catalytic domains. To facilitate future engineering approaches, *in silico* models of domain–domain interactions and of substrate–domain interactions are required. As we have exemplarily shown for a bacterial FAS,[107] structural modeling can contribute a better understanding of transient domain–domain interactions and provide information about substrate shuttling in megasynthases.

While the analysis of different chimeric PKSs, helps to define these important factors, structure elucidation of the overall PKS scaffold and domain–domain interfaces also plays a crucial role in optimizing engineering strategies. To this day, structural insight into PKS module organization has been difficult due to the multitude of transient interactions. We employed a strategy that allowed us to capture stable and transient domain–domain interactions and assembled them via molecular dynamics simulations to generate a model of PKS module in solution, thereby capturing its conformational flexibility (Chapter 2.5). Based on two currently disputed models, we employed SAXS, XL–MS, and *in silico* modeling on a partially reducing PKS module. Our preliminary models indicate a large conformational flexibility which will be further refined by integration of the cross–link results.

Taken together, the work in this thesis has shed light on the feasibility of different engineering strategies for the generation of chimeric PKSs, namely directed evolution experiments, use of synthetic docking domains, and multipoint active

site mutagenesis. While the limitations of directed evolution experiments were emphasized and demand a more detailed understanding of protein–protein interfaces, the results also show the success of using SYNZIP domains to combine PKS modules, and the positive impact of multipoint mutagenesis. Both can be build upon to engineer novel chimeric PKSs and, in combination, can help in defining crucial bottlenecks. In addition, novel structure elucidation techniques to characterize PKS architecture have been introduced, and while our analysis is not yet completed, it already shows the great potential of these techniques to define the multitude of transient interactions in PKS modules.

4 Experimental Procedures

4.1 Materials

CloneAmp HiFI PCR Premix was from Clontech. T4 DNA ligase was from Invitrogen. Restriction enzymes were from New England Biolabs. All primers were synthesized by Elim Biopharm or Sigma Aldrich. For DNA purification the Gene-JET Plasmid Miniprep Kit and the GeneJET Gel Extraction Kit were purchased from Thermo Scientific. BCA Protein Assay Kit was from Thermo Scientific.

For phage display protocols,[204] T4 polynucleotide kinase, T7 DNA polymerase, T4 DNA polymerase, 40 mM dNTP mix, and M13K07 helper phage were from NEB. The QIAprep Spin M13 Kit and the QIAquick Gel Extraction Kit were from Qiagen. Mouse monoclonal anti–M13–HRP antibody (horseradish peroxidase conjugated) was from GE Healthcare 3,3',5,5'–Tetramethylbenzidine (TMB) Liquid Substrate System was from Sigma Aldrich.

Competent cells were purchased from Agilent Technologies, Clontech, and Lucigen (Table 6). All chemicals for buffer preparations were from Sigma Aldrich or CarlRoth.

For cell growth and protein purification chemicals isopropyl–β–D–1–thiogalactopyranoside (IPTG), kanamycin sulfate and carbenicillin were from Gold Biotechnology, Inc. LB–Miller Broth and 2xYT media for cell cultures were from Fischer Scientific. Ni–NTA affinity resin was from MC Lab; or from CarlRoth.

The HiTrapQ column, Superose 6 Increase 10/300 GL column were from GE Healthcare. For all SDS–PAGE analyses SDS–PAGE Mini Protean TGX precast gels were purchased from Bio–Rad (4%–20% and 7.5%) or cast from Acrylamide/Bis–acrylamide, 30% solution (Sigma), N,N,N',N'–Tetramethylethylenediamine (TE-MED, CarlRoth), Ammoniumpersulfat (APS, Sigma) and Tris/HCl (CarlRoth). Amicon Ultra centrifugal filters were from Merck Millipore. StrepTactin resin/Strep-Tactin column (5 mL) or StrepTactinXT coated microtiter plates were from IBA Life Sciences. Amylose resin was from NEB.

For PKS enzymatic assays coenzyme A (CoA), reduced β–nicotinamide adenine dinucleotide 2'–phosphate (NADPH), sodium propionate, propionic acid, methyl-malonic acid and magenesium chloride hexahydrate were purchased from Sigma Aldrich. Adenosine–5'–triphosphate (ATP) was from Teknova/Sigma. Reducing agent tris(2–carboxyethyl)phosphine (TCEP) was purchased from Thermo Scientific. UVette cuvettes (2 mm x 10 mm path) were from Eppendorf and UV–Star half area microtiter plates were from Greiner.

For ^{14}C–radioistotopic labeling assay [1–^{14}C]–propionic acid sodium salt, was from Moravek Biochemicals and Mini Protean TGX precast gels (7.5 % polyacrylamide) as well as filter papers used for gel drying (Filter Paper Backing) were from Bio–Rad.

4.2 Genetic Methods

4.2.1 Bacterial Strains

E. coli strains used for cloning of novel plasmids or as heterologous expression host are listed in Table 6.

Table 6 *E. coli* strains used in this thesis

Name	Reference	Used for
BL21–Gold (DE3)	Agilent Technologies	Expression of *apo*–ACPs
BAP1	Pfeifer *et al.*[205]	Expression of *holo*–ACPs
Stellar™ competent cells	Clontech	Cloning
SS320	Lucigen	Phage Amplification
XL1 Blue	Agilent Technologies	Phage Selection
CJ236	Lucigen	Generation of $^+$ung DNA

4.2.2 Cloning of Plasmids

All plasmids used in this thesis are listed in Table S1. Plasmids in this study where either assembled via Infusion Cloning or generated by site–directed mutagenesis. Codon optimized gene sequences were ordered from IDT or from Thermo Fisher.

4.2.3 Transformation of Chemically Competent *E. coli* Cells

E. coli Stellar, BL21 and XL1 blue cells were transformed according to established protocols. Briefly 40 µL cells were chilled on ice, the DNA was added, and the mixture incubated on ice for 30 min. The amount of used DNA was either 0.6 µL for plasmid DNA, 2.5 µL for infusion reaction mixtures, or 3 µL for linear site–directed mutagenesis products. After incubation on ice, the cells were heat shocked for 20 s to 45 s at 42 °C and immediately incubated on ice for 3 min. Afterwards the cells were rescued via addition of 500 µL LB/SOC media and incubated for 1 h at 37 °C, 300 rpm. Finally, the cells were concentrated for 3 min at 1000 x g and plated on the appropriate antibiotic containing LB–agar plate.

4.2.4 Transformation of Electro–competent *E. coli* Cells

E. coli BL21, BAP1, SS320, and CJ236 cells were transformed according to established protocols. Briefly 50 µL cells were chilled on ice, 0.6 µL DNA was added,

and the mixture transferred to a GenePulser^{TM}(Bio–Rad) electroporation cuvette. Electroporation was performed at 1.8 kV for 5 msec (2.5 kV field strength, 125 Ω resistance, 50 µF capacitance for 350 µL aliquots of SS320 cells) and the cells rescued via addition of 400 µL LB media and incubated for 1 h at 37 °C, 300 rpm. Finally, the cells were concentrated for 3 min at 1000 x g and plated on the appropriate antibiotic containing LB–agar plate.

4.3 Bacterial Cell Culture and Protein Purification

4.3.1 Large Scale Cell Growth and Lysis

All proteins were produced and purified using similar protocols.

For *holo*–proteins (where the ACP domain is post-translationally modified with a phosphopantetheine arm) *E. coli* BAP1 cells[205] were used as the host. In the absence of BAP1 cells, *E. coli* BL21 cells were cotransformed with the plasmid pAR357 (encoding for Sfp from *B. subtilis*). For *apo*–proteins (no phosphopantetheine arm on the ACP) *E. coli* BL21 cells were used for protein production. All proteins carried a tag for affinity chromatographic purification (His$_{6/8}$–tag ,and/or a Strep–tagII, and/or MBP–tag).

Cell culture growth for the protein production was carried out on a 1–2 L scale of LB–media/2xYT–media supplemented with 1% of the corresponding antibiotic(s) and 0.1% glucose. Cultures were inoculated with 10–20 mL overnight culture and grown at 37 °C to an OD$_{600}$ of 0.3, whereupon the temperature was adjusted to 18 °C. At OD$_{600}$ of 0.6 protein production was induced with 0.1 mM IPTG, and the cells were grown for another 18 h. Cells were harvested by centrifugation at 5000 x g for 8 min and lysed by sonication (or by French press) in a buffer consisting of 50 mM sodium phosphate, 10 mM imidazole, 450 mM NaCl, and 10% glycerol, pH 7.6. The cell debris was removed by centrifugation at 25,000 x g

for 1 h (later the protocol was optimized to 50,000 x g for 45–60 min) and the supernatant was added to the respective affinity chromatography resin.

4.3.1.1 Standard Purification Protcol using Ni-NTA Affinity Chromatography and Anion Exchange Chromatography

For $His_{6/8}$–tagged proteins the supernatant was incubated on Ni–NTA agarose resin (2 mL resin per liter of culture) at 4 °C for 1 h. Afterwards, the mixture was applied to a gravity flow column and washed with the above lysis buffer (10 column volumes). Additional washing was performed with 10 column volumes wash buffer (50 mM sodium phosphate, 25 mM imidazole, 300 mM NaCl, 10% glycerol, pH 7.6). Proteins were eluted with 6 column volumes 50 mM sodium phosphate, 500 mM 10% glycerol, pH 7.6. Using a HitrapQ column, additional purification was performed by anion exchange chromatography (AEX) on an ÄKTA FPLC system. Buffer A consisted of 50 mM sodium phosphate, 10% glycerol, pH 7.6, whereas buffer B contained 50 mM sodium phosphate, 500 mM NaCl, 10% glycerol, pH 7.6. Enzymes MatB, SCME and PrpE were purified as described before.[75,125] Protein concentrations were determined with the BCA Protein Assay Kit. Samples were stored as aliquots at −80 °C until further use.

4.3.1.2 StrepTactin and Amylose Affinity Chromatography

For strep–tagged proteins the lysate was applied to a 5 mL StrepTactin column and the column was washed with 5 column volumes of lysis buffer prior to elution with 3 column volumes lysis buffer containing 2.5 mM destiobiotin.

MBP–tagged proteins were purified using 5 mL amylose resin per 2 L of culture. After applying the lysate, the column was washed with 12 column volumes of lysis buffer prior to elution with 6 column volumes of lysis buffer containing 10 mM maltose.

4.3.1.3 Size Exclusion Chromatography

To determine the purity of the purified protein after anion exchange chromatography, samples of each protein were analyzed on an ÄKTA FPLC system using a Superose 6 Increase 10/300 GL column with a buffer containing 50 mM sodium phosphate, 500 mM NaCl, 10% glycerol, pH 7.6.

4.4 Enzymatic Assays

4.4.1 PKS Turnover Assay

As the PKS turnover directly correlates with the consumption of NADPH, a UV assay procedure can be employed to monitor PKS turnover.[75] Reactions were performed on a 70 µL scale, and contained 400 mM sodium phosphate (pH 7.2), 5 mM TCEP, 10 mM MgCl2, 1 mM CoA, and 8 mM ATP. Methylmalonate/malonate (1 mM) was converted to racemic methylmalonyl–CoA using the enzymes MatB (2/4 µM) and methylmalonyl–CoA epimerase (4/8 µM).[206] Propionyl–CoA was synthesized from propionate using PrpE (1/2 µM).[125] The concentration of these enzymes and cofactors was selected to assure that the acyl–CoA supply was not limiting the rate of product formation. Reactions were started by the addition of PKS proteins (4 µM each) along with a cocktail of propionate (0.5 mM), methyl-malonate/ malonate (1 mM), and NADPH (0.5 mM). The rate was monitored at 340 nm over 20 min either in UVette cuvettes (Eppendorf) or UV–half area plates (Greiner). Method description adapted from ref[104].

4.4.2 ^{14}C–radioisotopic labeling assay with purified DEBS proteins

[1–^{14}C]–Propionate was used to interrogate intramodular chain elongation within intact DEBS M1, broken variants thereof, and chain elongation between chimeric

KS–AT and KR–ACP proteins. A 10x substrate mix was generated by mixing 0.4 mM [1–^{14}C]–propionate (Moravek Biochemicals), 2 mM methylmalonate, 2.4 mM CoA, 3.5 μM PrpE, 2 μM MatB, and 4 μM SCME in a reaction containing 400 mM sodium phosphate, pH 7.2, 5 mM TCEP, 10 mM MgCl$_2$, and 8 mM ATP and incubated for 1 h. To quantify the efficiency of channeling of the labeled propionyl group successively from the LDD to either intact M1 or broken variants thereof, the substrate mixture was diluted ten–fold into individual reaction mixtures containing 2 μM LDD(4) and 2 μM of either intact M1 of 2 μM of each a KS1–AT1 construct and a KR–ACP construct in the presence of 400 mM sodium phosphate, pH 7.2, 5 mM TCEP, and 0.75 mM NADPH. By minimizing the amount of labeled propionyl–CoA in the reaction mixture, non–specific transfer of the radiolabel directly from the LDD to an acceptor module was minimized. At specified time–points, reactions were quenched by addition of Laemmli buffer, and samples were separated via 7.5% SDS–PAGE (Mini Protean TGX precast gels, Bio–Rad) at 200 V for 44 min. The gel was washed with water for 5 min and stained with SimplyBlue SafeStain (Invitrogen) for 20 min. Following destaining with water for 5 min, the gel was mounted on a filter paper (Bio–Rad) and dried *in vacuo* for 2 h using a Bio–Rad 543 Gel Dryer. The dried gel was imaged for 20 min lane by lane to quantify ^{14}C on individual protein bands using a Rita Star TLC Analyzer (Raytest). Peaks were integrated to quantify the radiolabel bound to each protein. Method description adapted from ref[105].

4.4.3 LC–MS Analysis

Dried samples were reconstituted in 100 μL methanol, separated on a ZORBAX Extend C18 column (Agilent, 1.8 um, 2.1 x 50 mm), connected to an Agilent Infinity 1290 II HPLC over a 6 min linear gradient of acetonitrile from 5% to 95% in water, and subsequently injected into a 6545 QTOF mass spectrometer. Reduced and unreduced triketide products were located by searching for the theoretical m/z

for the $[M+Na]^+$- and $[M+H]^+$-ion. Unreduced triketides: $[M+Na]^+$ = 193.080 and $[M+H]^+$ = 171.098, reduced triketides $[M+Na]^+$ = 195.100 and $[M+H]^+$ =173.118. Method description adapted from ref[105].

4.5 Phage Display

4.5.1 Western Blotting of ACP–phage

Western blotting was performed using the Bio–Rad transblotting turbo kit with a PVDF membrane. The membrane was blocked with blocking buffer (1xPBS, 5%BSA, 0.1% Tween20) for 1 h at RT and afterwards incubated for 1 h at RT with the primary antibody (rabbit anti–FLAG, Sigma, 1:1000 dilution in blocking buffer). Washing was performed three times for 5 min with PBST (1xPBS, 0.1% Tween20) and the membrane incubated with the secondary antibody (donkey anti–rabbit HRP conjugate (Sigma), 1:2000 dilution) for 1 h at RT. After another washing procedure, the membrane was developed using the Pierce ECL2 Western Blotting Substrate (Thermo Scientific), quenched with water, and the fluorescence measured at 473 nm.

4.5.2 ^{14}C–radioisotopic labeling assay to measure chain translocation from ACP–phage to M1

$[1–^{14}C]$–Propionate was used to interrogate chain translocation from ACP–phage to DEBS M1. A 20x substrate mix was generated by mixing 0.2 mM $[1–^{14}C]$–propionate (Moravek Biochemicals), 0.1 mM CoA, and 3.0 µM PrpE in a reaction containing 400 mM sodium phosphate, pH 7.2, 10 mM $MgCl_2$, and 4 mM ATP and incubated for 1 h. Subsequently, 5 µM of the substrate mix were combined with 3 µM Sfp.

In parallel, concentrated ACP–phage samples ($\sim 5 \times 10^{12}$ PFU/mL; ACP_L– and ACP2–phage, total amount $\sim$ 1.4–2.8 pmol) were precipitated with a solution (1:5 dilution) containing 20% PEG8000 (w/v) and 2.5 M NaCl for 15 min at room temperature. After centrifugation, the pellet was resuspended in 50 µL Sfp–buffer (400 mM sodium phosphate, pH 7.2, 10 mM $MgCl_2$, and 5 mM TCEP). To allow for activation of the ACP, 7 µL of the Sfp–substrate mixture were added to the prepared phage samples and incubated overnight, resulting in an estimated *holo*–ACP–phage concentration of 50 nM. The concentration was calculated from the phage titer under the assumption that one ACP was presented per phage.

On the next day, Sfp was removed and the ACP–phage transferred to a new buffer system lacking $MgCl_2$. Towards this end, the activated phage sample was supplemented with a final concentration of 20 mM EDTA, precipitated again via addition of the PEG/NaCl solution, and after centrifugation resuspended in 200 µL transfer buffer (400 mM sodium phosphate, pH 7.2 and 5 mM TCEP). An additional wash step was performed via another round of precipitation and centrifugation and pellet resuspended in 15 µL transfer buffer to yield the activated and concentrated ACP–phage with and approx. concentration of 50 nM.

To assess the chain translocation ability of *holo*–ACP–phage to M1. A reaction was set up with 50 nM *holo*–ACP–phage, 200 nM of LDD–ACP^0 (MK29), 4 µM of *holo*–M1 and 10 µM substrate mix in a total volume of 30 µL transfer buffer and incubated for specific time points (1, 3, and 5 min). At specified time points, reactions were quenched by addition of Laemmli buffer, and samples were separated via 7.5% SDS–PAGE (Mini Protean TGX precast gels, Bio–Rad) at 200 V for 44 min. The gel was washed with water for 5 min and stained with SimplyBlue SafeStain (Invitrogen) for 20 min. Following destaining with water for 5 min, the gel was mounted on a filter paper (Bio–Rad) and dried *in vacuo* for 2 h using a Bio–Rad 543 Gel Dryer. The dried gel was imaged for 20 min lane by lane to quantify ^{14}C

on individual protein bands using a Rita Star TLC Analyzer (Raytest). Peaks were integrated to quantify the radiolabel bound to M1.

4.5.3 Library Generation

A library of ACP1(2) mutants fused to the N–terminus of the minor coat protein P3 of the M13 bacteriophage was generated as described before.[204] A primer for randomization of five amino acids in the chain translocation epitope of ACP1 was designed using NNK at the target positions (P-MK162: 5′ CTG GCG TCG CTG CCC GCG NNK GAG CGC NNK NNK GCG CTG TTC NNK CTC GTG CGC NNK CAC GCG GCC GCC GTC CTC 3′ (N: A/C/G/T K: G/T). Based on the two best mutants gained from ACP–Lib1 (Lib1–Mut3, Lib1–Mut5), two deeper libraries were designed in which 6 other residues of helix I of ACP1 were randomized. Primer P-MK263 was used to generate ACP1-Lib2 based on the previous mutation from Lib1–Mut3 and primer P-MK264 was used to create ACP1–Lib3 out of the mutant Lib1–Mut5. (P-MK263: 5′ GCG TGT GAG CGC TAT CGG NNK CTG NNK AGG CTC GTG NNK GAG NNK NNK GCC NNK GTC CTC GGC CAC GCC TC 3′ and P-MK264: 5′ CTG CCC GCG CAT GAG CGC TTT CGT NNK CTG NNK TGT CTC GTG NNK AGG NNK NNK GCC NNK GTC CTC GGC CAC GCC TCG GC 3′). *E. coli* SS320 were transformed with the Kunkel mutagenesis product and the phage library was harvested from 0.5 L of culture. Phage harvest was done using standard protocols and the final product was resuspended in a buffer consisting of 50 mM sodium phosphate, 150 mM NaCl, pH 7.55. The titer was calculated to determine the practical diversity of the library.

4.5.4 Biopanning

For all panning steps a buffer containing 50 mM sodium phosphate, 150 mM NaCl, pH 7.55 was used. All incubations were done at room temperature while

shaking at 200 rpm. Panning was performed in a StrepTactinXT coated microtiter plate. As a first step, four wells were coated with (3)KS3–AT3–Strep (100 µL at 5 µg/mL) for 2 h. Afterwards, the plate was blocked using 200 µL of blocking buffer (50 mM sodium phosphate, 150 mM NaCl, 0.5% BSA, 0.05% Tween20, pH 7.55). At the same time the purified phage library was also blocked in the same buffer (procedure as described as before[204]). The blocked phage library (100 µL per well) was incubated for 1 h on the coated and blocked wells, prior to washing with 200 µL of wash buffer (50 mM sodium phosphate, 150 mM NaCl, 0.05% Tween20, pH 7.55). Wash steps were increased each round over the course of the panning protocol (10x, 12x, 14x, 16x, 18x). Elution of bound phage was done with 100 µL of 100 mM hydrochloric acid for 20 min. The eluted phage were collected and the solution neutralized by adding 1/10 volume of 1 M tris base pH 8.0 and 1/10 volume of blocking buffer. The elution was propagated and the titer was determined. For propagation 2 mL of log phase *E. coli* SS320 (OD600 0.8) were added to 200 µL of eluted phage and infection was allowed for 30 min and 37 °C, 200 rpm whereupon 20 µL M13K07 helper phage ($1.0x10^{12}$ PFU/mL) were added. Following incubation for 1 h at 37 °C, 200 rpm 45 mL of 2xYT/carb/kan were added to the culture and cells continued to grow over night. The next day, phages were harvested by centrifugation at 5000 x g for 10 min to remove the cells. Phage particles remained in the supernatant and were precipitated with a solution containing 20% PEG8000 (w/v) and 2.5 M NaCl for 15 min at room temperature. Subsequently, phages were pellet at 13000 x g for 10 min. The pellet containing the phage was resuspended in 1 mL of blocking buffer. Another spin for 5 min at 15000 x g removed the remaining cells. Phages were either stored at 4 °C or used directly. Of all five rounds cells containing the eluted phage (from the titering microtiter plate) were plated on LB/carb plates to obtain colonies for future analysis.

4.5.5 Specificity ELISA

To test the eluted phage for increased and specific binding a specificity ELISA was performed. Colonies form the target elution rounds were grown in a minitube rack containing 400 µL 2xYT/carb/K07 medium (supplemented with 10^{10} PFU/mL M13K07). A MaxiSorp plate was coated for 2 h at room temperature with 100 µL of 2 µg/mL target protein (3)KS3–AT3 and as a negative control with BSA. Afterwards the plate was blocked for 1 h with 200 µL of blocking buffer (50 mM sodium phosphate, 150 mM NaCl, 0.5% BSA, 0.05% Tween20, pH 7.55), followed by incubation of 100 µl of phage containing supernatant from the minitube rack for 1h. The plate was washed twice with 200 µL of wash buffer (50 mM sodium phosphate, 150 mM NaCl, 0.05% Tween20, pH 7.55), before incubating 100 uL of a 1:5000 anti–M13–HRP antibody solution in blocking buffer for 1 h. Following antibody incubation the plate was washed three times before development with 100 uL TMB substrate and quenching upon color change with 1 M phosphoric acid. Read out was carried out at 450 nm on a Tecan Synergy HT plate reader.

4.5.6 Titration ELISA

To further confirm the relative binding intensity of the newly enriched mutants, a titration ELISA using purified phage was used. Phages were purified out of a 45 mL culture using described protocols.[204] The OD_{268} was used as a measure to determine phage amounts (OD_{268} = $5x10^{12}$ PFU/mL). The phage stock was serial diluted in ELISA blocking buffer and applied to the ELISA plate using the same protocol as in Section 4.5.5. The advantage of this assay is that defined phage amounts can be used and the signal can be normalized across different phage preparations.

4.6 Structure Elucidation Techniques

4.6.1 Tandem Size Exclusion Chromatography and Small Angle X–Ray Scattering

SEC–SAXS analysis was performed on the Bio–SAXS beamline BM29 at the European Synchrotron Radiation Facility (ESRF).[207] All measurements were performed in a buffer consisting of 50 mM sodium phosphate, 500 mM NaCl, 5% glycerol, pH 7.55 using a Superose 6 Increase 10/300 GL column. Details of SAXS data collection and analysis are listed in Table 5.

4.7 Cross–link Mass Spectrometric Analysis

All XL–MS experiments were carried out by Andreas Linden (Urlaub group at the MPI for Biophysical Chemistry in Göttingen), who also provided the methods description below.

4.7.1 SDA Cross–linking

150 to 175 µg of M2, M3 and ADD01 at a final protein concentration of 1 µg/µL or less were incubated with 2 mM SDA (100 mM stock in DMSO) for 30 min at room temperature, in a buffer consisting of 50 mM sodium phosphate, 500 mM NaCl, 10% glycerol, pH 7.55 (MK53) or 200 mM sodium phosphate, 200 mM NaCl, 10% glycerol, pH 7.55 (ADD01, second experiment). The cross–linking reactions were quenched with 50 mM Tris–HCl. The samples were dialyzed against reconstitution buffer via a membrane filter (MF Membrane Filters, 0.025 µm VSWP, Merck). Afterwards, samples were irradiated with UV light (365 nm) for 5 min at 4 °C.

4.7.2 Protein Digestion and Enrichment of Cross–linked Peptides

Proteins of cross-linked samples were reduced and alkylated with 10 mM DTT and 40 mM iodoacetamide, respectively. Proteins were digested by trypsin in an enzyme–to–protein ratio of 1:50 at 37 °C overnight at a final concentration of 1 M urea. Peptides were acidified with trifluoroacetic acid (TFA) to a final concentration of 0.5% (v/v), desalted on MicroSpin Colums (Harvard Apparatus) following manufacturer's instructions and vacuum dried. Peptides were resuspended in 50 μL 30% acetonitrile (v/v)/0.1% TFA (v/v) to enrich cross–linked peptide species by peptide size exclusion chromatography (SuperdexPeptide 3.2/300 column, GE Healthcare). Fractions of 50 μL were collected at a flow rate of 50 μL/min and those that eluted first and contain the cross–linked peptide pairs were subjected to LC–MS/MS analysis.

4.7.3 LC–MS/MS Analysis

Cross–linked peptides were measured in technical duplicates on an Orbitrap Fusion Tribrid Mass Spectrometer or on a Q Exactive HF–X coupled to a Dionex UltiMate 3000 UHPLC system (Thermo Fisher Scientific) equipped with an in house–packed C18 column (ReproSil–Pur 120 C18-AQ, 1.9 μm pore size, 75 μm inner diameter, 30 cm length, Dr. Maisch GmbH). Samples were separated applying the following gradient: mobile phase A consisted of 0.1% formic acid (FA, v/v), mobile phase B of 80% ACN/0.08% FA (v/v). The gradient started at 5% B, increasing to 12, 15 or 20% B within 3 min (according to fraction), followed by a continuous increase to 46% B within 45 min, then keeping B constant at 90% for 8 min. After each gradient the column was again equilibrated to 5% B for 2 min. The flow rate was set to 300 nL/min. MS1 survey scans were acquired in the orbitrap (OT) with a resolution of 120,000, an injection time (IT) of 60 ms (50 ms on HF–X) and an automatic gain control (AGC) target of 5×105. Dynamic exclusion

was set to 10 s (30 s on HF–X) and only charge states between +3 and +8 were considered for fragmentation. MS2 spectra were acquired in the OT of the 20 (25 on HF–X) most abundant precursor ions, resolution 30,000, IT 128 ms and AGC target $5×10^4$. Fragmentation was enforced by higher–energy dissociation (HCD) at 30%. collisional

4.7.4 Data Analysis

ProteomeDiscoverer 1.4 (Thermo Fisher Scientific) was used for converting raw files into .mgf format (signal–to–noise ratio 1.5, 1000—10000 Da precursor mass). The generated .mgf files were subjected to pLink v. 1.23 (pFind group)[208] to identify cross–linked peptides. Here, default settings were applied with carbamidomethylation of cysteines as fixed and oxidation of methionines as variable modification. FDR was set to 0.01. SDA was selected as cross–linker. All spectra were evaluated manually.

4.8 Bioinformatical Analysis

4.8.1 Sequence Alignments

Sequence alignments were generated with ClustalW as implemented in Jalview 2.[209]

4.8.2 Homology Models

Homology models of PKS domains were either generated with I–TASSER[210–212] or SWISS–MODEL.[213–217]

4.8.3 Gene Harmonization

Gene harmonization of DEBS M2 was done by Florian Buhr from the Schwalbe group (Goethe University Frankfurt) based on the codon usage rates of *E. coli* and *S. erythraea* as implemented in the Codon Usage Database (Table S8&S9) and as published before.[190] The translation rates were predicted using the *RiboTempo* web server,[191] and slow translating stretches were manually adjusted by Prof. Zoya Ignatova (University Hamburg).

4.8.4 Calculation of Multi–point Mutants with FuncLib

The FuncLib webserver was used to calculate diverse multipoint mutations within the KS active site.[147] The calculation for KS3 was based on the X–ray structure of the KS3–AT3 didomain.[33] The calculation for DEBS KS6 was based on the homology model generated with SWISS–MODEL based on the X–ray structure of KS5–AT5.[32] In the first step, where the possible sequence space is calculated, chains A+B were included for KS3. To decrease the calculation effort only chain A was included for KS6. Within the parameter selection 7 residues were selected for mutation for KS3 and 9 positions for KS6. In addition, residues of the catalytic tri-ade (Cys202, H337 and H377) were chosen not be altered during simulations. The default parameters were used for generation of the multiple sequence alignment (Min ID 0.3, Max targets 3000, Coverage 0.6 and E value 0.0001) for all calculations. In the second calculation step (design step), the parameters were selected to yield between 500 – 500,000 designs. Parameters are listed in Table 7. For KS3 152,826 and for KS6 285,822 design were selected for Rosetta calculation.

4.8.5 Molecular Dynamics Simulation of DEBS modules

Homology models of DEBS KS2, the KS2–AT2 linker (LD), AT2, KR2, and docking domain 2 (DD2) were generated by MODELLER[199] and were based on a multiple

Table 7 Settings chosen in FuncLib during the design step

Parameter	selection for KS3 and KS6
Minimal number of mutations per design	3
Maximum number of mutations per design	5
Minimal PSSM threshold	-1
$\Delta\Delta G$	6
Difference between clustered variants	2

sequence alignment (MAFFT).[198] On the basis of the homology models, two initial models of MBP–M2(2) were generated, starting from either the extended or arched conformation of KS–AT. The initial arched model was generated by fitting individual domains into the previously reported cryo–EM densities.[47] For the extended conformation, DAMMIN was used to calculate SAXS envelopes.[200] For both conformations, *in silico* scattering intensities calculated via FoXS,[201] were derived from coarse–grained simulations using COMPLEXES++.[202] BioEn was used for sample refinement.[203]

5 Appendix

5.1 Supplementary Tables

Table S1   Plasmids generated in this thesis

Plamsmid
pMK01_TwinStrep_DEBS2_H8_pET22b_carb
pMK02_TwinStrep_DEBSM2-(2)_H8_pET22b_carb
pMK03_DEBSM2-(2)_H8_pET22b_carb
pMK04_SUMO3_DEBSM2-(2)_H8_pET22b_carb
pMK05_MBP_DEBSM2-(2)_H8_pET22b_carb
pMK06_TwinStrep_SUMO3_DEBSM2-(2)_H8_pET22b_carb
pMK07_GST_DEBSM2-(2)_H8_pET22b_carb
pMK08_DEBSACP2-(2)_H8_pET22b_carb
pMK09_DEBSACP2-(2)_pET22b_carb
pMK10_DEBSACP2-(4)_H8_pET22b_carb
pMK11_DEBSACP2-(4)_pET22b_carb
pMK12_DEBSACP4-(4)_pET22b_carb
pMK13_DEBSACP4-(2)_pET22b_carb
pMK14_TwinStrep_DEBSM2-(2)_pET22b_carb
pMK15_TwinStrep_SUMO3_DEBSM2-(2)_pET22b_carb
pMK16_GST_DEBSM2-(2)_pET22b
pMK17_DEBSACP2-(2)_phagemid
pMK18_DEBSACP4-(2)_phagemid
pMK19_DEBSACP4-(4)_phagemid
pMK20_DEBSACP2-(4)_phagemid
pMK21_DEBSACP2_phagemid
pMK22_DEBSACP1-(2)_phagemid
pMK24_TwinStrep_DEBSACP1-(2)_phagemid
pMK25_FLAG_DEBSACPL-(4)_phagemid
pMK26_FLAG_DEBSACP3-(0)_phagemid
pMK28_DEBSACP1-(2)_H8_pET22b_carb
pMK29_H6_LDD(4)_ACP inactiviation SXXA_pET28_kan
pMK31_H6_linker aa_DEBS M2_(2)_pET28_Kan
pMK32_TwinStrep_linker aa_DEBSM2-(2)_pET22b_carb
pMK33_H6_linker aa_DEBS M4_(4)_pET28_Kan
pMK34_TwinStrep_linker aa_DEBSM4-(4)_pET22b_carb
pMK35_(5)-DEBS M5_H8_pET22b_carb
pMK36_H6-DEBSACP2-(2)_pET28_kan
pMK38_H6-harm_DEBSM2-(2)_pET28_kan
pMK39_TwinStrep_harm_DEBSM2-(2)_pET22b_carb

Table S1 – continued from previous page

Plamsmid

pMK40_TwinStrep_harm_DEBS KS2-AT2_EcoRI_pET22b_carb
pMK41_TwinStrep_non-harm_DEBS KS2-AT2_EcoRI_pET22b_carb
pMK42_(5)-DEBS M5+TE-H6_pET22_Carb
pMK44_TwinStrep_linkeraa_DEBSM4-(4)_H8_pET22b_carb
pMK45_(3)-KS3-AT3-TwinStrep_pET22b_carb
pMK46_TwinStrep_harm_DEBS KS2-AT2-complete linker_EcoRI_pET22b_carb
pMK47_TwinStrep_harm_DEBS KS2-AT2-linker up to YR_EcoRI_pET22b_carb
pMK48_TwinStrep_DEBSACP2-(2)_pET22b_carb
pMK49_TwinStrep_DEBSACP4-(4)_pET22b_carb
pMK50_TwinStrep_(5) harm DEBSM2-(2)_pET22b_carb
pMK51_TwinStrep_GST_harm_DEBSM2-(2)_pET22b_carb
pMK52_TwinStrep_MBP_harm_DEBSM2-(2)_pET22b_carb
pMK53_MBP_harm_DEBSM2-(2)_pET22b
pMK54_(3)KS6-AT6-H6_pET22b_carb
pMK55_TwinStrep_DEBSACP2_pET22b_carb
pMK56_TwinStrep_DEBSACP4_pET22b_carb
pMK57_TwinStrep_DEBSACP2-(4)_pET22b_carb
pMK58_(5)-KS5-AT5-H6_pET22_Carb
pMK59_TwinStrp-FactorXa_DEBSACP1-(2)_pET22b_carb
pMK60_MBP_DEBSACP1-(2)_pET22b_carb
pMK61_MBP_DEBSACP2-(4)_pET22b_carb
pMK62- (3)harmDEBSM2+TE-H6_pET22_carb
pMK63_MBP_ACP2_pET22b_carb
pMK64_MBP_ACP4_pET22b_carb
pMK65_MBP-ACP5-(4)_pET22b_carb
pMK66_MBP-ACP5_pET22b_carb
pMK67_MBP-ACP3_pET22b_carb
pMK68_MBP-ACP3-2_pET22b_carb
pMK70_TwinStrep_MBP_DEBSACP2-(2)_pET22b_carb
pMK71_TwinStrep_MBP_DEBSACP4-(4)_pET22b_carb
pMK72_TwinStrep_MBP_DEBSACP1-(2)_pET22b_carb
pMK73_ (5)DEBS M1-Mut1(2)-H6_pET22b_carb
pMK74_ (5)DEBS M1-Mut2(2)-H6_pET22b_carb
pMK75_ (5)DEBS M1-Mut3(2)-H6_pET22b_carb
pMK76_ (5)DEBS M1-Mut4(2)-H6_pET22b_carb
pMK77_ (5)DEBS M1-Mut5(2)-H6_pET22b_carb
pMK78_ (5)DEBS M1-Mut6(2)-H6_pET22b_carb
pMK79_ACP1-harmM2+TE-H6_pET22b_carb
pMK80_ACP1-M5+TE-H6_pET22b_carb
pMK81_ACP1-M6+TE-H6_pET22b_carb
pMK83_ACP1-non-harmM2+TE-H6_pET22b_carb
pMK84_(3)harmDEBSM2+TE-H6_as BL16_pET22_carb
pMK91_(5)KSAT1-H6_pET22b_carb
pMK92_(5)KSAT1+KR1-H6_with RE sites_pET22b_carb
pADD01_KR1-ACP1-harmM2+TE-H6_pET22b_carb
pADD02_KR1-ACP1-M5+TE-H6_pET22b_carb
pADD03_KR1-ACP1-non-harmM2+TE-H6_pET22b_carb
pMK93_KR1-ACP1-M6+TE
pMK96_KR1-ACP1-DD2-H6_carb
pMK97_StrepI-FactorXa_KR1-ACP1-(2)_pET22b_carb
pMK98_StrepI_KR1ACP1-linker-harmKS2-AT2-linker up to YR_EcoRI_pET22b_carb
pMK99_StrepI_KR5ACP5-linker-KS6-AT6-linker part_pET22b_carb
pMK100_KR1-ACP1-H6_pET22b_carb
pMK105_KR1-ACP1-non-harmM2+TE(null)-H6_pET22b_carb_(pADD03 TEnull)

Table S1 – continued from previous page

Plamsmid

pMK106_(3)DEBSM5+TE(null)-H6_carb_(pBL17 TEnull)
pMK107_KR1-ACP1-M5+TE-H6_pET22b_carb_(pADD02 TEnull)
pMK108_KR1-ACP1-M6+TE-H6_pET22b_carb_(pMK93 TEnull)
pMK109_(5)KSAT1-SZ3-H6_pET22b_carb
pMK110_SZ4-KR1-ACP1-DD2-H6_pET22_carb
pMK111_(5)KSAT1-GSG-SZ3-H6_pET22b_carb
pMK112_SZ4-GGSG-KR1-ACP1-DD2-H6_pET22_carb
pMK113_(5)KSAT1-_PAL2-H6_pET22b_carb_pMK91 shorter
pMK114_MBP-harmDEBSM2+TE-H6_pET22_carb
pMK115_H6-(AT-ACP)LDD-KS1AT1-PAL_pET28_kan
pMK116_(5)KSAT1-GGSG2-SZ1-H6_pET22b_carb
pMK117_SZ2-GGSG2-KR1-ACP1-DD2-H6_pET22_carb
pMK118_(5)KSAT1-GGSG2-SZ16-H6_pET22b_carb
pMK119_SZ21-GGSG2-KR1-ACP1-DD2-H6_pET22_carb
pMK120 DEBSM2+TE NULL AGAIN (same as pBL60)
pMK121- KR1-ACP1-harmDEBSM2+TE(TE null)-H6_pET22_carb
pMK122- KR1-ACP1-harmDEBSM2 (KS0)+TE-H6_pET22_carb
pMK123- KR1-ACP1-harmDEBSM2 (KS2null and ACP2null)+TE-H6_pET22_carb
pMK125_KR1-ACP1(ACP1 null)-harmDEBSM2+TE(TE null)-H6_pET22_carb
pADD05_(5)M1-linker-harmDEBSM2+TE-H6_pET22_carb_based on pADD01 = pMK126
pADD06_(5)M1-linker-M5+TE-H6_pET22b_Carb_based on pADD02 = pMK127
pADD07_(5)M1-linker-nonharmM2+TE-h6_pET22b_carb_based on pADD03 = pMK128
pADD08_(5)M1-linker-M6+TE-H6_pET22b_carb_based on pMK93 = pMK129
pMK131_pADD06+D_(5)M1-ACP1-linker-M5+TE-H6_pET22b_Carb
pMK132_pADD08+D_(5)M1-linker-M6+TE-H6_pET22b_carb
pMK133_pADD03+D_KR1-ACP1-non-harmM2+TE-H6_pET22b_carb
pMK134_pADD02+D_KR1-ACP1-M5+TE-H6_pET22b_carb
pMK135_pMK93+D_KR1-ACP1-M6+TE_pET22b_carb
pMK136_SZ4-L-KR1-ACP1-non-harmM2+TE-H6_pET22b_carb_based on pMK133
pMK137_SZ4-L-KR5-ACP5-M6+TE-H6_pET22b_carb
pMK138_SZ4-L-RIFS(KR1-ACP1-L-M2)-TE-H6_pET22b_carb
pMK140_TwinStrep_MBP_KS1-AT1-L-SZ3_pET22b_carb
pMK141_SZ4-L-KR1-ACP1-linker-M5+TE-H6_pET22b_Carb_based on pMK134
pMK142_SZ4-L-KR1-ACP1-L-DEBS M6+TE-H6_pET22b_carb_QC_based on pMK135
pMK143_(3)-KS3-AT3-L-SZ3-H6_pET22b_carb
pMK144_KS3-AT3-L-SZ3-H6_pET22b_carb
pMK146_SZ4-GGSG-HindIII-H6_pET22_carb
pMK147_SZ4-GGSG-HindIII-M6-TE-HindIII-H6_pET22_carb
pMK148_SZ4-GGSG-HindIII-M2-TE-HindIII-H6_pET22_carb
pMK149_SZ4-GGSG-HindIII-M3-TE-HindIII-H6_pET22_carb
pMK150_(5)M1-GSG-SZ3-H6_pET22b_carb
pMK151_SZ4-L-KR1-ACP1-(GGGS)4-DEBS M6+TE-H6_pET22b_carb
pMK152-L12_SZ4-L-KR1-ACP1-(L12)-DEBS M3+TE-H6_pET22b_carb
pMK153_SZ4-GGSG-HindIII-KR3-HindIII-H6_pET22_carb
pMK156_StrepII_MCS_H8_pET22b_carb
pMK157_StrepII_RS_(D_dd4)-RS-H8_pET22b_carb
pMK158_StrepII_RS_(D_dd2)-RS-H8_pET22b_carb
pMK159_StrepII_RS_(D_dd5)-RS-(D_dd2)-RS-H8_pET22b_carb
pMK160_StrepII_RS-(D_dd3)-MCS_H8_pET22b_carb
pMK161_H6-TYLS LDD_pETDuet-1_carb
pMK162_H6-TYLS M1_pETDuet-1_carb
pMK163_H6-TYLS M2_pETDuet-1_carb
pMK164_StrepII_TYLS M3_H8_pET22b_carb
pMK165_StrepII_TYLS LDD_(D_dd4)-H8_pET22b_carb

Table S1 – continued from previous page

Plamsmid
pMK166_StrepII_(D_dd5)-TYLS M1-(D_dd2)-H8_pET22b_carb
pMK167_StrepII_RS-(D_dd3)-TYLS M2-(T-DD2)_H8_pET22b_carb
pMK168_SZ4-L-KR5-ACP5-DD2-H6_pET22_carb
pMK169_SZ4-L-RIFS(KR1-ACP1)-DD2-H6_pET22_carb
pMK170_SZ4-GGSG-HindIII-M5-TE-HindIII-H6_pET22_carb
pMK171_SZ4-GGSG-KR1-ACP1-TE-H6_pET22_carb
pMK172_SZ4-L-KR5-ACP5-TE-H6_pET22_carb
pMK173_RIFS-(ACP1)-H6_pET22_carb
pMK174_SZ4-L-RIFS(KR1-ACP1)-TE-H6_pET22_carb
pMK175_SZ4-L-KR2-ACP2-TE-H6_pET22b_carb
pMK176_KR5-ACP5-TE(null)-H6_pET22_carb
pMK177_RIFS(KR1-ACP1)-TE(null)-H6_pET22_carb
pMK178_(5)KS1-AT1-KR2-ACP2-GSG-SZ3-H6_pET22b_carb
pMK179_(5)KS1-AT1-KR5-ACP5-GSG-SZ3-H6_pET22b_carb
pMK180_SZ4_short-GGSG-KR1-ACP1-DD2-H6_pET22_carb
pMK181_(5)KSAT1-GSG-SZ17-H6_pET22b_carb
pMK182_SZ18-GGSG-KR1-ACP1-DD2-H6_pET22_carb
pLB001_SZ4-GGSG-M3-TE-H6_A189W_pET22_carb
pLB002_SZ4-GGSG-M6-TE-H6_A189W_pET22_carb
pLB003_SZ4-GGSG-M3-TE-H6_A189Q_F191Y_A265S_F298L_pET22_carb
pLB004_SZ4-GGSG-M3-TE-H6_A189L_F191Y_A265S_F298L_F300Y_pET22_carb
pLB005_SZ4-GGSG-M3-TE-H6_A19E_F191Y_A165T_F298Y_F300W_pET22_carb
pLB006_SZ4-GGSG-M3-TE-H6_A189Q_F191Y_A265T_F298L_F300M_pET22_carb
pLB007_SZ4-GGSG-M3-TE-H6_A189Q_F191K_F298W_S341T_A476S_pET22_carb
pLB008_SZ4-GGSG-M3-TE-H6_A189T_F191Q_F298L_S306P_pET22_carb
pLB009_SZ4-GGSG-M3-TE-H6_A189Q_F191N_F298L_S341R_pET22_carb
pLB010_SZ4-GGSG-M3-TE-H6_F191Y_A265T_F298A_S341T_A476S_pET22_carb
pLB011_SZ4-GGSG-M3-TE-H6_A189Q_F191Y_A265C_F298L_S341P_pET22_carb
pLB013_SZ4-GGSG-M3-TE-H6_A189Q_F156Y_A265S_S341P_pET22_carb
pLB014_SZ4-GGSG-M3-TE-H6_A189W_F191M_F298L_F300Y_pET22_carb
pLB015_SZ4-GGSG-M3-TE-H6_A189E_F191Y_F298L_F300Y_pET22_carb
pLB016_SZ4-GGSG-HindIII-M6-TE-HindIII-H6_A189S_Q191H_S239A_V272I_A341S_pET22_carb
pLB017_SZ4-GGSG-HindIII-M6-TE-HindIII-H6_A189W_Q191S_S239A_S265A_A341R_pET22_carb
pLB018_SZ4-GGSG-HindIII-M6-TE-HindIII-H6_A189W_Q191H_S239A_S265T_A341T_pET22_carb
pLB019_SZ4-GGSG-HindIII-M6-TE-HindIII-H6_A189S_Q191H_S239A_S265H_A341S_pET22_carb
pLB021_SZ4-GGSG-HindIII-M6-TE-HindIII-H6_A189W_Q191S_S265A_A341T_pET22_carb
pLB022_SZ4-GGSG-HindIII-M6-TE-HindIII-H6_Q191S_S239A_V272L_F300M_A341E_pET22_carb
pLB023_SZ4-GGSG-HindIII-M6-TE-HindIII-H6_A189G_Q191H_S239A_S265H_A341Q_pET22_carb
pLB024_SZ4-GGSG-HindIII-M6-TE-HindIII-H6_A189W_S239A_S265T_A341S_A477S_pET22_carb
pLB026_SZ4-GGSG-HindIII-M6-TE-HindIII-H6_S239A_V272L_F300M_pET22_carb
pLB027_SZ4-GGSG-HindIII-M6-TE-HindIII-H6_A189T_S239A_S265H_V272M_A341S_pET22_carb

Plamids were either generated by Maja Klaus "pMK", by Alicia D. D'Souza "pADD", or Lynn Buyachuihan "pLB".

Table S2 Phage titers and enrichment factors during the directed evolution experiments of ACP1–Library 2. Comparison of input and elution titers in each of the selection rounds and the corresponding enrichment factor. Enrichment factor calculated as change in the ratio of elution/input titers across two consecutive rounds of panning. N/A: not applicable.

Selection round	Input titer / PFU/mL	Enput titer / PFU/mL	Enrichment factor
1	3.20×10^{12}	4.40×10^{5}	N/A
2	1.31×10^{13}	2.60×10^{5}	0.14
3	2.40×10^{13}	1.02×10^{6}	2.14
4	2.40×10^{13}	1.06×10^{7}	10.39
5	4.48×10^{13}	5.60×10^{6}	0.28

Table S3 Phage titers and enrichment factors during the directed evolution experiments of ACP1–Library 3. Comparison of input and elution titers in each of the selection rounds and the corresponding enrichment factor. Enrichment factor calculated as change in the ratio of elution/input titers across two consecutive rounds of panning. N/A: not applicable.

Selection round	Input titer / PFU/mL	Enput titer / PFU/mL	Enrichment factor
1	5.33×10^{11}	6.40×10^{4}	N/A
2	4.53×10^{12}	1.04×10^{6}	1.91
3	2.67×10^{13}	1.80×10^{6}	0.29
4	2.67×10^{13}	4.00×10^{5}	0.22
5	4.51×10^{13}	1.16×10^{7}	17.16

Table S4   Amino acid sequences of proteins used in Chapters 2.3 and 2.4

Constructs	Amino acid sequence
MK180 SZ4*–KR1– ACP1(2)	MLNRNEQLKNKVEELKNRNAYLKNELATLENEVARLENDVAEGGSGGGSGDEVSALRYRIEWRPTGAGEPA RLDGTWLVAKYAGTADETSTAAREALESAGARVRELVVDARCGRDELAERLRSVGEVAGVLSLLAVDEAEP EEAPLALASLADTLSLVQAMVSAELGCPLWTVTESAVATGPFERVRNAAHGALWGVGRVIALENPAVWGGL VDVPAGSVAELARHLAAVVSGGAGEDQLALRADGVYGRRWVRAAAPATDDEWKPTGTVLVTGGTGGVGGQI ARWLARRGAPHLLLVSRSGPDADGAGELVAELEALGARTTVAACDVTDRESVRELLGGIGDDVPLSAVFHA AATLDDGTVDTLTGERIERASRAKVLGARNLHELTRELDLTAFVLFSSFASAFGAPGLGGYAPGNAYLDGL AQQRRSDGLPATAVAWGTWAGSGMAEGPVADRFRRHGVIEMPPETACRALQNALDRAEVCPIVIDVRWDRF LLAYTAQRPTRLFDEIDDARRAAPQAAAEPRVGALASLPAPEREKALFELVRSHAAAVLGHASAERVPADQ AFAELGVDSLSALELRNRLGAATGVRLPTTTVFDHPDVRTLAAHLAAELGTEVRGEAPSALAGLDALEAAL PEVPATEREELVQRLERMLAALRPVAQAADASGTGANPSGDDLGEAGVDELLEALGRELDGDPNSSSVDKL AAALEHHHHHH
MK116 (5)KS1–AT1– SZ1	MSGDNGMTEEKLRRYLKRTVTELDSVTARLREVEHRAGEPVAVVAMACRLPGGVSTPEEFWELLSEGRDAV AGLPTDRGWDLDSLFHPDPTRSGTAHQRGGGFLTEATAFDPAFFGMSPREALAVDPQQRLMLELSWEVLER AGIPPTSLQASPTGVFVGLIPQEYGPRLAEGGEGVEGYLMTGTTTSVASGRIAYTLGLEGPAISVDTACSS SLVAVHLACQSLRRGESSLAMAGGVTVMPTPGMLVDFSRMNSLAPDGRCKAFSAGANGFGMAEGAGMLLLE RLSDARRNGHPVLAVLRGTAVNSDGASNGLSAPNGRAQVRVIQQALAESGLGPADIDAVEAHGTGTRLGDP IEARALFEAYGRDREQPLHLGSVKSNLGHTQAAAGVAGVIKMVLAMRAGTLPRTLHASERSKEIDWSSGAI SLLDEPEPWPAGARPRRAGVSSFGISGTNAHAIIEEAPQVVEGERVEAGDVVAPWVLSASSAEGLRAQAAR LAAHLREHPGQDPRDIAYSLATGRAALPHRAAFAPVDESAALRVLDGLATGNADGAAVGTSRAQQRAVFVF PGQGWQPWAGMAVDLLDTSPVFAAALRECADALEPHLDFEVIPFLRAEAARREQDAALSTERVDVVQPVMFA VMVSLASMWRAHGVEPAAVIGHSQGEIAAACVAGALSLDDAARVVALRSRVIATMPGNKGMASIAAPAGEV RARIGDRVEIAAVNGPRSVVVAGDSDELDRLVASCTTECIRAKRLAVDYASHSSHVETIRDALHAELGEDF HPLPGFVPFFSTVTGRWTQPDELDAGYWYRNLRRTVRFADAVRALAEQGYRTFLEVSAHPILTAAIEEIGD GSGADLSAIHSLRRGDGSLADFGEALSRAFAAGVAVDWESVHLGTGARRVPLPTYPFQRERVWLGGSGGGS GNLVAQLENEVASLENENETLKKKNLHKKDLIAYLEKEIANLRKKIEELEHHHHHH
MK117 SZ2–KR1– ACP1(2)	MARNAYLRKKIARLKKDNLQLERDEQNLEKIIANLRDEIARLENEVASHEQGGSGGGSGDEVSALRYRIEW RPTGAGEPARLDGTWLVAKYAGTADETSTAAREALESAGARVRELVVDARCGRDELAERLRSVGEVAGVLS LLAVDEAEPEEAPLALASLADTLSLVQAMVSAELGCPLWTVTESAVATGPFERVRNAAHGALWGVGRVIAL ENPAVWGGLVDVPAGSVAELARHLAAVVSGGAGEDQLALRADGVYGRRWVRAAAPATDDEWKPTGTVLVTG GTGGVGGQIARWLARRGAPHLLLVSRSGPDADGAGELVAELEALGARTTVAACDVTDRESVRELLGGIGDD VPLSAVFHAAATLDDGTVDTLTGERIERASRAKVLGARNLHELTRELDLTAFVLFSSFASAFGAPGLGGYA PGNAYLDGLAQQRRSDGLPATAVAWGTWAGSGMAEGPVADRFRRHGVIEMPPETACRALQNALDRAEVCPI VIDVRWDRFLLAYTAQRPTRLFDEIDDARRAAPQAAAEPRVGALASLPAPEREKALFELVRSHAAAVLGHA SAERVPADQAFAELGVDSLSALELRNRLGAATGVRLPTTTVFDHPDVRTLAAHLAAELGTEVRGEAPSALA GLDALEAALPEVPATEREELVQRLERMLAALRPVAQAADASGTGANPSGDDLGEAGVDELLEALGRELDGD PNSSSVDKLAAALEHHHHHH
MK178 (5)KS1–AT1– KR2–ACP2–SZ3	MSGDNGMTEEKLRRYLKRTVTELDSVTARLREVEHRAGEPVAVVAMACRLPGGVSTPEEFWELLSEGRDAV AGLPTDRGWDLDSLFHPDPTRSGTAHQRGGGFLTEATAFDPAFFGMSPREALAVDPQQRLMLELSWEVLER AGIPPTSLQASPTGVFVGLIPQEYGPRLAEGGEGVEGYLMTGTTTSVASGRIAYTLGLEGPAISVDTACSS SLVAVHLACQSLRRGESSLAMAGGVTVMPTPGMLVDFSRMNSLAPDGRCKAFSAGANGFGMAEGAGMLLLE RLSDARRNGHPVLAVLRGTAVNSDGASNGLSAPNGRAQVRVIQQALAESGLGPADIDAVEAHGTGTRLGDP IEARALFEAYGRDREQPLHLGSVKSNLGHTQAAAGVAGVIKMVLAMRAGTLPRTLHASERSKEIDWSSGAI SLLDEPEPWPAGARPRRAGVSSFGISGTNAHAIIEEAPQVVEGERVEAGDVVAPWVLSASSAEGLRAQAAR LAAHLREHPGQDPRDIAYSLATGRAALPHRAAFAPVDESAALRVLDGLATGNADGAAVGTSRAQQRAVFVF PGQGWQPWAGMAVDLLDTSPVFAAALRECADALEPHLDFEVIPFLRAEAARREQDAALSTERVDVVQPVMFA VMVSLASMWRAHGVEPAAVIGHSQGEIAAACVAGALSLDDAARVVALRSRVIATMPGNKGMASIAAPAGEV RARIGDRVEIAAVNGPRSVVVAGDSDELDRLVASCTTECIRAKRLAVDYASHSSHVETIRDALHAELGEDF HPLPGFVPFFSTVTGRWTQPDELDAGYWYRNLRRTVRFADAVRALAEQGYRTFLEVSAHPILTAAIEEIGD GSGADLSAIHSLRRGDGSLADFGEALSRAFAAGVAVDWESVHLGTGARRVPLPTYPFQRERVWLLPDRTTP RDELDGWFYRVDWTEVPRSEPAALRGRWLVVVPEGHEEDGWTVEVRSALAEAGAEPEVTRGVGGLVGDCAG VVSLLALEGDGAVQTLVLVRELDAEGIDAPLWTVTFGAVDAGSPVARPDQAKLWGLGQVASLERGPRWTGL VDLPHMPDPELRGRLTAVLAGSEDQVAVRADAVRARRLSPAHVTATSEYAVPGGTILVTGGTAGLGAEVAR WLAGRGAEHLALVSRRGPDTEGVGDLTAELTRLGARVSVHACDVSSREPVRELVHGLIEQGDVVRGVVHAA GLPQQVAINDMDEAAFDEVVAAKAGGAVHLDELCSDAELFLLFSSGAGVWGSARQGAYAAGNAFLDAFARH RRGRGLPATSVAWGLWAAGGMTGDEEAVSFLRERGVRAMPVPRALAALDRVLASGETAVVVTDVDWPAFAE SYTAARPRPLLDRIVTTAPSERAGEPETESLRDRLAGLPRAERTAELVRLVRTSTATVLGHDDPKAVRATT PFKELGFDSLAAVRLRNLLNAATGLRLPSTLVFDHPNASAVAGFLTSELGGGSGGGSGNEVTTLENDAAFI ENENAYLEKEIARLRKEKAALRNRLAHKKLEHHHHHH

Constructs	Amino acid sequence
MK179 (5)KS1–AT1–KR5–ACP5–SZ3	MSGDNGMTEEKLRRYLKRTVTELDSVTARLREVEHRAGEPVAVVAMACRLPGGVSTPEEFWELLSEGRDAV AGLPTDRGWDLDSLFHPDPTRSGTAHQRGGGFLTEATAFDPAFFGMSPREALAVDPQQRLMLELSWEVLER AGIPPTSLQASPTGVFVGLIPQEYGPRLAEGGEGVEGYLMTGTTTSVASGRIAYTLGLEGPAISVDTACSS SLVAVHLACQSLRRGESSLAMAGGVTVMPTPGMLVDFSRMNSLAPDGRCKAFSAGANGFGMAEGAGMLLLE RLSDARRNGHPVLAVLRGTAVNSDGASNGLSAPNGRAQVRVIQQALAESGLGPADIDAVEAHGTGTRLGDP IEARALFEAYGRDREQPLHLGSVKSNLGHTQAAAGVAGVIKMVLAMRAGTLPRTLHASERSKEIDWSSGAI SLLDEPEPWPAGARPRRAGVSSFGISGTNAHAIIEEAPQVVEGERVEAGDVVAPWVLSASSAEGLRAQAAR LAAHLREHPGQDPRDIAYSLATGRAALPHRAAFAPVDESAALRVLDGLATGNADGAAVGTSRAQQRAVFVF PGQGWQWAGMAVDLLDTSPVFAAALRECADALEPHLDFEVIPFLRAEAARREQDAALSTERVDVVQPVMFA VMVSLASMWRAHGVEPAAVIGHSQGEIAAACVAGALSLDDAARVVALRSRVIATMPGNKGMASIAAPAGEV RARIGDRVEIAAVNGPRSVVVAGDSDELDRLVASCTTECIRAKRLAVDYASHSSHVETIRDALHAELGEDF HPLPGFVPFFSTVTGRWTQPDELDAGYWYRNLRRTVRFADAVRALAEQGYRTFLEVSAHPILTAAIEEIGD GSGADLSAIHSLRRGDGSLADFGEALSRAFAAGVAVDWESVHLGTGARRVPLPTYPFQRERVWLPIPTGGR ARDEDDDWRYQVVWREAEWESASLAGRVLLVTGPGVPSELSDAIRSGLEQSGATVLTCDVESRSTIGTALE AADTDALSTVVSLLSRDGEAVDPSLDALALVQALGAAGVEAPLWVLTRNAVQVADGELVDPAQAMVGGLGR VVGIEQPGRWGGLVDLVDADAASIRSLAAVLADPRGEEQVAIRADGIKVARLVPAPARAARTRWSPRGTVL VTGGTGGIGAHVARWLARSGAEHLVLLGRRGADAPGASELREELTALGTGVTIAACDVADRARLEAVLAAE RAEGRTVSAVMHAAGVSTSTPLDDLTEAEFTEIADVKVRGTVNLDELCPDLDAFVLFSSNAGVWGSPGLAS YAAANAFLDGFARRRRSEGAPVTSIAWGLWAGQNMAGDEGGEYLRSQGLRAMDPDRAVEELHITLDHGQTS VSVVDMDRRRFVELFTAARHRPLFDEIAGARAEARQSEEGPALAQRLAALSTAERREHLAHLIRAEVAAVL GHGDDAAIDRDRAFRDLGFDSMTAVDLRNRLAAVTGVREAATVVFDHPTITRLADHYLERLVGGSGGGSGN EVTTLENDAAFIENENAYLEKEIARLRKEKAALRNRLAHKKLEHHHHHH

Table S5 Sequence space of selected KS3 active site residues during FuncLib calculation

posistion	theoretical sequence spaces
154	ADEGIKLMNQSTVW
156	FADEGHKMNQSTY
230	ACST
263	FALSY
265	FMWY
306	SAEGKMNPQRTV
441	AS

Table S6   Yields of proteins used in section 2.4. For sequence of newly generate MK178 and MK179 see Table S4. Typical yields are presented.

Construct	Protein	Yield / mg/L of culture
MK147	SZ4–M6–TE	6.31
MK148	SZ4–M2–TE	2.45
MK149	SZ4–M3–TE	5.89
MK178	(5)KS1–AT1–KR2–ACP2–SZ3	4.13
MK179	(5)KS1–AT1–KR5–ACP5–SZ3	5.4
LB001	SZ4–M3–TE_Mut01	7.87
LB002	SZ4–M6–TE_Mut02	5.45
LB003	SZ4–M3–TE_Mut03	10.58
LB004	SZ4–M3–TE_Mut04	5.41
LB005	SZ4–M3–TE_Mut05	7.24
LB006	SZ4–M3–TE_Mut06	8.54
LB007	SZ4–M3–TE_Mut07	8.36
LB008	SZ4–M3–TE_Mut08	7.94
LB009	SZ4–M3–TE_Mut09	8.62
LB010	SZ4–M3–TE_Mut10	11.1
LB011	SZ4–M3–TE_Mut11	17.35
LB013	SZ4–M3–TE_Mut13	5.7
LB014	SZ4–M3–TE_Mut14	10.18
LB015	SZ4–M3–TE_Mut15	6.85

Table S7   Melting temperatures of SZ4–M3–TE and its mutants

Construct	T_m / °C
MK149	41.0±0.6
LB001	40.3±0.6
LB003	41.1±0.5
LB004	39.9±1.5
LB005	40.1±0.5
LB006	42.3±0.3
LB007	44.3±0.4
LB008	40.5±0.4
LB009	43.0±0.0
LB010	44.5±0.0
LB011	40.5±0.0
LB013	39.8±0.4
LB014	45.5±0.0
LB015	41.5±0.5

Table S8 Codon usage in *E. coli.* Data obtained from the Codon Usage Database (https://www.kazusa.or.jp/codon/). Frequencies per thousand.

codon	amino acid	frequency	codon	amino acid	frequency	codon	amino acid	frequency	codon	amino acid	frequency
UUU	F	24.4	UCU	S	13.1	UAU	Y	21.6	UGU	C	5.9
UUC	F	13.9	UCC	S	9.7	UAC	Y	11.7	UGC	C	5.5
UUA	L	17.4	UCA	S	13.1	UAA	*	2.0	UGA	*	1.1
UUG	L	12.9	UCG	S	8.2	UAG	*	0.3	UGG	W	13.4
CUU	L	14.5	CCU	P	9.5	CAU	H	12.4	CGU	R	15.9
CUC	L	9.5	CCC	P	6.2	CAC	H	7.3	CGC	R	14.0
CUA	L	5.6	CCA	P	9.1	CAA	Q	14.4	CGA	R	4.8
CUG	L	37.4	CCG	P	14.5	CAG	Q	26.7	CGG	R	7.9
AUU	I	29.6	ACU	T	13.1	AAU	N	29.3	AGU	S	13.2
AUC	I	19.4	ACC	T	18.9	AAC	N	20.3	AGC	S	14.3
AUA	I	13.3	ACA	T	15.1	AAA	K	37.2	AGA	R	7.1
AUG	M	23.7	ACG	T	13.6	AAG	K	15.3	AGG	R	4.0
GUU	V	21.6	GCU	A	18.9	GAU	D	33.7	GGU	G	23.7
GUC	V	13.1	GCC	A	21.6	GAC	D	17.9	GGC	G	20.6
GUA	V	13.1	GCA	A	23.0	GAA	E	35.1	GGA	G	13.6
GUG	V	19.9	GCG	A	21.1	GAG	E	19.4	GGG	G	12.3

Table S9 Codon usage in *S. erythraea*. Data obtained from the Codon Usage Database (https://www.kazusa.or.jp/codon/). Frequencies per thousand.

codon	amino acid	frequency	codon	amino acid	frequency	codon	amino acid	frequency	codon	amino acid	frequency
UUU	F	0.4	UCU	S	0.6	UAU	Y	0.8	UGU	C	0.4
UUC	F	28.1	UCC	S	14.9	UAC	Y	18.0	UGC	C	7.9
UUA	L	0.1	UCA	S	1.0	UAA	*	0.1	UGA	*	2.5
UUG	L	5.1	UCG	S	20.0	UAG	*	0.5	UGG	W	15.2
CUU	L	1.9	CCU	P	1.8	CAU	H	1.4	CGU	R	4.0
CUC	L	30.7	CCC	P	18.5	CAC	H	22.1	CGC	R	40.6
CUA	L	0.5	CCA	P	1.7	CAA	Q	1.8	CGA	R	3.1
CUG	L	65.3	CCG	P	37.0	CAG	Q	26.8	CGG	R	30.5
AUU	I	0.6	ACU	T	1.2	AAU	N	0.6	AGU	S	1.1
AUC	I	30.5	ACC	T	36.4	AAC	N	17.2	AGC	S	15.0
AUA	I	0.3	ACA	T	1.4	AAA	K	1.1	AGA	R	0.6
AUG	M	16.6	ACG	T	15.9	AAG	K	16.5	AGG	R	5.1
GUU	V	2.3	GCU	A	4.2	GAU	D	4.7	GGU	G	11.8
GUC	V	44.2	GCC	A	61.9	GAC	D	55.2	GGC	G	56.1
GUA	V	1.1	GCA	A	5.6	GAA	E	10.0	GGA	G	7.3
GUG	V	43.3	GCG	A	59.2	GAG	E	50.8	GGG	G	18.7

Table S10   Gene sequence of M2 before and after harmonization

Construct	DNA sequence
M2 original sequence	GAGCCGATCGCGATCGTCGGCATGGCGTGCCGGCTGCCCGGGGAGGTCGACTCCCCGGAGCGGCTGTGGGAGCTGATCACC TCGGGACGCGACTCCGCGGCGGAGGTCCCCGATGACCGGGGCTGGGTCCCCGACGAGCTGATGGCCTCCGACGCGGCGGGA ACCCGCCGCGCCCACGGCAACTTCATGGCGGGCGCCGGTGACTTCGACGCGGCGTTCTTCGGGATCTCGCCGCGCGAGGCG CTGGCGATGGACCCGCAGCAGCGCCAGGCGCTGGAGACGACGTGGGAGGCGCTGGAAAGCGCGGGCATCCCCACCGGAGACG TTGCGCGGCAGCGACACCGGCGTGTTCGTCGGCATGTCCCACCAGGGCTACGCGACCGGGCGTCCGCGCCCGGAGGACGGC GTCGACGGGTACCTGCTCACCGGCAACACCGCGAGCGTCGCGTCGGGACGCATCGCCTACGTGCTGGGGCTGGAAGGTCCC GCGCTGACGGTGGACACGGCGTGTTCGTCGTCGTTGGTGGCGTTGCACACGCGTGTGGGTCGTTGCGTGACGGTGACTGC GGTCTTGCGGTGGCCGGTGGTGTGTCGGTGATGGCGGGTCCGGAGGTGTTCACCGAGTTCTCCCGCCAGGGCGCGCTCTCG CCGGACGGCCGGTGCAAGCCCTTCTCGGACGAGGCCGACGGATTCGGTCTCGGGGAGGGTTCGGCGTTCGTCGTGCTCCAG CGGTTGTCCGACGCCAGGCGGGAGGGCCGCCGCGTGCTCGGCGTGGTGGCCGGGTCCGCGGTGAACCAGGACGGCGCGAGC AACGGGCTCTCCGCTCCGAGCGGCGTCGCGCAGCAGCGGGTCATCCGCCGGGCGTGGGCGCGTGCGGGGATCACGGGCGCG GATGTGGCCGTGGTGGAGGCGCATGGGACCGGTACGCGGCTGGGCGATCCGGTGGAGGCGTCGGCGTTGCTGGCTACTTAC GGCAAGTCGCGCGGGTCGTCGGGCCCGGTGCTGCTGGGTTCGGTGAAGTCGAACATCGGTCACGCGCAGGCGGCCGCGGGT GTCGCGGGCGTGATCAAGGTGCTGCTCGGCCTGGAACGCGGTGTGGTGCCCCCGATGCTGTGCCGGGGCGAGAGGTCGGGC CTCATCGACTGGTCCTCCGGCGAGATCGAGCTCGCAGACGGCGTGCGGGAGTGGTCGCCCGCCGCGGACGGGGTGCGCCGG GCAGGTGTGTCGGCGTTCGGGGTGAGCGGGACGAACGCGCACGTGATCATCGCCGAGCCGCCGGAACCGGAGCCCGTGCCG CAACCGCGACGCATGCTGCCCCGCGACCGGGGTGGTGCCGGTCGTGCTGTCGGCCAGGACCGGGGCGGCGTTGCGGGCGCAG GCCGGCAGGCTCGCCGACCACCTCGCCGCGCATCCCGGGATCGCCACCGGCCGACGTGAGCTGGACGATGGCGCGGGCCCGC CAGCACTTCGAGGAGCGGGCCGCGGTGCTCGCCGCCGACACCGCCGAGGCCGTGCACCGGTTGCGGGCGGTGGCCGACGGC GCGGTGGTTCCCGGTGTTGTCACCGGCAGTGCCTCCGACGGTGGTTCAGTGTTCGTCTTCCCTGGGCAGGGTGCCCAGTGG GAAGGCATGGCGCGGGAGTTGTTGCCGGTTCCCGTCTTCGCCGAGTCGATCGCCGAGTGCGATGCGGTGTTGTCGGAGGTG GCCGGATTCTCGGTGTCCGAGGTGCTGGAGCCACGTCCGGACGCGCCGTCGCTGGAGCGGGTCGACGTGGTGCAGCCGGTG CTGTTCGCGGTGATGGTGTCGCTGGCGCGGTTGTGGCGTGCCTGCGGTGCCGTTCCTTCGGCCGTCATAGGGCACTCGCAG GGTGAGATCGCCGCCGCGGTGGTGGCGGGAGCGTTGTCGCTGGAGGACGGCATGCGCGTCGTCGCCCGCCGGTCGAGGGCG GTGCGTGCGGTCGCGGGCCGGGGGAGCATGCTCTCGGTGCGCGGCGGCCGCTCCGACGTCGAGAAGCTGCTCGCCGACGAC AGCTGGACCGGCAGGCTGGAGGTCGCCGCGGTCAACGGCCCCGACGCCGTGGTGGTGGCCGGTGACGCCCAGGCGGCGCGC GAGTTCCTGGAGTACTGCGAGGGCGTGGGCATCCGCGCCCGCGCGATCCCGGTGGACTACGCCTCGCACACCGCGCACGTC GAGCCCGTGCGCGACGAACTGGTCCAGGCGCTGGCCGGGATCACCCCGCGACGGGCCGAGGTGCCGTTCTTCTCCACCCTG ACCGGCGACTTCCTCGACGGCACCGAGCTGGACGCGGGCTACTGGTACCGCAACCTGCGTCACCCGGTGGAGTTCCACTCC GCCGTGCAGGCGCTGACCGACCAGGGATACGCGACGTTCATCGAGGTCAGCCCGCACCCGGTGCTGGCGTCGAGCGTCCAG GAGACCCTCGACGACGCCGAGTCGGACGCGGCGGTGCTCGGGACGCTGGAACGCGACGCGGGCGACGCCGACCGCTTCCTC ACGGCACTCGCCGACGCGCACACGCGCGGTGTCGCGGTCGACTGGGAAGCGGTGCTCGGCCGGGCCGGACTGGTCGACCTG CCGGGTTATCCTTTCCAGGGCAAGCGGTTCTGGCTGCTGCCGGACCGCACCACCCCTCGTGACGAGCTCGACGGCTGGTTC TACCGGGTCGACTGGACCGAGGTGCCGCGCTCCGAACCTGCCGCGCTGCGCGGCCGTTGGCTCGTGGTGGTGCCCGAGGGG CACGAGGAGGACGGCTGGACCGTCGAGGTGCGGTCCGCGCTCGCCGAGGCCGGCGCCGAACCGGAGGTCACGCGCGGCGTC GGCGGGCTGGTCGGTGACTGCGCGGGCGTGGTGTCGTTGCTCGCCCTCGAGGGCGATGGTGCGGTGCAAACCCTTGTGCTG GTGCGGGAACTCGACGCCGAGGGCATCGACGCGCCACTGTGGACGGTCACCTTCGGCGCGGTCGACGGTGACGCGGGCAGTCCGGTG GCCCGCCCGGACCAGGCGAAGCTGTGGGGGCTGGGCCAGGTCGCGTCCCTGGAACGCGGGCCCCGCTGGACCGGCCTCGTC GACCTGCCGCACATGCCGGACCCGGAACTGCGAGGCCGTCTCACCGCGGTGCTGGCCGGCTCGGAGGACCAGGTCGCGGTG CGCGCCGACGCCGTGCGTGCGCGGCGGCTTTCCCCCGCCCACGTCACCGCCACCTCGGAGTACGCGGTGCCGGGCGGCACA ATCCTGGTCACCGGTGGCACCGCCGGCCTGGGCGCGGAGGTGGCCCGGTGGCTCGCCGGTCGCGGCGCCGAACACCTCGCG CTGGTCAGCAGGCGAGGCCCGGACACCGAGGGCGTCGGCGACCTGACCGCCGAGCTGACCCGGCTCGGCGCGCGGGTGTCG GTGCACGCGTGCGACGTCAGCAGCCGCGAACCGGTGAGGGAACTCGTGCACGGCCTGATCGAGCAGGGCGACGTCGTCCGC GGTGTGGTGCACGCGGCGGGACTGCCGCAGCAGGTCGCGATCAACGACATGGACGAGGCCGCCTTCGACGAGGTGGTCGCG GCCAAGGCCGGGGGCGCGGTGCACCTGGACGAGCTGTGCTCGGACGCCGAGCTGTTCCTGCTGTTCTCCTCCGGGGCCGGG GTGTGGGGAAGCGCCCGCCAGGGCGCCTACGCCGCGGGCAACGCGTTCCTGGACGCCTTCGCCCGGCACCGCCGGGGCCGC GGCCTGCCCGCCACGTCGGTGGCGTGGGGGCTGTGGGCGGCGGGCGGCATGACCGGCGACGAGGAGGCCGTGTCGTTCCTG CGCGAGCGCGGTGTGCGGGCGATGCCCGTACCGCGCGCCCTCGCCGCCCTGGACAGGGTGCTGGCCTCCGGGGAGACGGCG GTGGTCGTGACGGACGTGGACTGGCCCGCCTTCGCCGAGTCCTACACCGCCGCCCGGCCCCGGCCGTTGCTCGACCGCATC GTCACGACCGCGCCGAGCGAGCGGGCCGGAGAACCGGAGACGGAGAGCCTGCGCGACCGGCTGGCGGGTCTGCCGCGTGCC GAGCGGACGGCGGAGCTGGTGCGCCTGGTCCGCACCAGCACCGCGACCGTGCTGGGCCACGACGACCCGAAGGCGGTGCGC GCGACCACGCCGTTCAAGGAGCTCGGGTTCGACTCGCTGGCGGCCGTCCGGCTGCGCAACCTGCTCAACGCGGCCACCGGG CTCCGCCTGCCGTCGACGCTGGTCTTCGACCACCCGAACGCCTCCGCGGTCGCCGGTTTCCTCGACGCCGAGCTC

Table S10 – continued from previous page

Construct	DNA sequence
M2 har-monized sequence	GAACCGATTGCAATTGTTGGTATGGCATGTCGTCTGCCGGGCGAAGTTGATAGCCCGGAACGTCTGTGGGAACTGATTACC AGCGGGCGTGATAGCGCAGCAGAAGTTCCGGACGATCGTGGTTGGGTTCCGGATGAACTGATGGCAAGCGATGCAGCAGGG ACCCGTCGTGCACATGGTAACTTTATGGCAGGTGCAGGGGATTTTGATGCAGCATTTTTTGGCATTAGCCCGCGTGAAGCA CTGGCAATGGATCCGCAGCAGCGTCAGGCACTGGAAACAACATGGGAAGCACTGGAGAGCGCAGGTATTCCCCCGGAAACA CTACGTGGTAGCGATACCGGTGTTTTTGTTGGTATGAGCCATCAGGGTTATGCAACCGGCAGGCCGCGTCCGGAAGATGGT GTTGATGGCTATCTGCTGACCGGTAACACCGCAAGCGTTGCAAGCGGGCGTATTGCCTATGTTCTGGGCGTGGAGGGGCCG GCACTGACAGTTGATACAGCAtgtAGCAGCAGCCTAGTTGCACTACATACAGCATGCGGCAGCctcAGGGATGGGGATTGT GGGctcGCAGTTGCAGGGGGGGTTAGCGTTATGGCAGGGCCGGAAGTTTTTACCGAATTTAGCCGTCAGGGTGCACTGAGC CCGGATGGTCGTTGTAAGCCGTTTAGCGATGAAGCAGATGGGTTTGGGCTGGGCGAAGGGAGCGCATTTGTTGTTCTGCAG CGTCTAAGCGATGCACGACGTGAAGGTCGTCGTGTTCTGGGTGTTGTTGCAGGCAGCGCAGTTAACCAGGATGGTGCAAGC AACGGCCTGAGCGCTCCGAGCGGTGTTGCACAGCAGCGTGTTATTCGTCGTGCATGGGCAAGGGCAGGCATTACAGGTGCA GACGTTGCAGTTGTTGAAGCACACGGCACCGGGACACGTCTGGGTGACCCGGTTGAAGCAAGCGCACTACTGGCTACTTAT GGTAAGAGCCGTGGCAGCAGCGGTCCGGTTCTGCTGGGGAGCGTTAAGAGCAACATTGGGCATGCACAGGCAGCAGCAGGG GTTGCAGGTGTTATTAAGGTTCTGCTGGGTCTGGAGCGTGGGGTTGTTCCGCCGATGCTGTGTCGTGGTGAACGAAGCGGT CTGATTGATTGGAGCAGCGGTGAAATTGAACTGGCTGATGGTGTTCGTGAATGGAGCCCGGCAGCAGATGGCGTTCGTCGT GCTGGGGTTAGCGCATTTGGCGTTAGCGGCACAAACGCACATGTTATTATTGCAGAACCGCCGGAGCCGGAACCGGTTCCG CAACCGAGGCGTATGCTGCCGGCAACCGGCGTTGTTCCGGTTGTTCTGAGCGCACGAACCGGCGCAGCACTACGTGCACAG GCAGGTCGACTGGCAGATCATCTGGCAGCACACCCGGGCATTGCTCCGGCAGATGTTAGCTGGACAATGGCACGTGCACGT CAGCATTTTGAAGAACGTGCAGCAGTTCTGGCAGCAGATACCGCAGAAGCAGTTCATCGTCTACGTGCAGTTGCAGATGGT GCAGTTGTACCGGGGGTAGTTACCGGTTCGGCAAGCGATGGGGGGTCGGTTTTTGTTTTTCCCGGCCAGGGGGCACAGTGG GAGGGTATGGCACGTGAActcCTACCGGTACCGGTTTTTGCAGAAAGCATTGCAGAATGTGACGCAGTTCTAAGCGAAGTT GCAGGGTTTAGCGTTAGCGAAGTTCTGGAACCCAGGCCGGATGCACCGAGCCTGGAACGTGTTGATGTTGTTCAGCCGGTT CTGTTTGCAGTTATGGTTAGCCTGGCACGTCTATGGAGGGCATGTGGGGCAGTACCCAGCGCAGTTATAGGCCATAGCCAG GGGGAAATTGCAGCAGCAGTTGTTGCAGGGGCACTAAGCCTGGAAGATGGTATGCGTGTTGTTGCACGTCGTAGCCGAGCA GTTAGGGCAGTTGCAGGTCGTGGCAGCATGCTGAGCGTTCGTGGTGGTCGTAGCGATGTTGAAAAGCTGCTGGCAGATGAT AGCTGGACCGGTCGACTGGAAGTTGCAGCAGTTAACGGTCCGGATGCAGTTGTTGTTGCAGGGGATGCACAGGCAGCACGT GAATTTCTGGAATATTGTGAAGGTGTTGGTATTCGTGCACGTGCAATTCCGGTTGATTATGCAAGCCATACCGCACATGTT GAACCGGTTCGTGATGAGCTGGTTCAGGCACTGGCAGGCATTACCCCGAGGCGTGCAGAAGTTCCGTTTTTTAGCACCCTG ACCGGTGATTTTCTGGATGGTACCGAACTGGATGCAGGTTATTGGTATCGTAACCTGAGGCATCCGGTTGAATTTCATAGC GCAGTTCAGGCACTGACCGATCAGGGGTATGCAACATTTATTGAAGTTAGCCCGCATCCGGTTCTGGCAAGCAGCGTTCAG GAAACCCTGGATGATGCAGAAAGCGATGCAGCAGTTCTGGGCACACTGGAGCGTGATGCAGGTGATGCAGATCGTTTTCTG ACAGCTCTGGCAGATGCACATACACGTGGGGTTGCAGTTGATTGGGAGGCAGTTCTGGGTCGTGCAGGGCTGGTTGATCTG CCGGGGTACCCCTTTCAGGGTAAGCGTTTTTGGCTGCTGCCGGATCGTACCACCCCCAGGGATGAACTGGATGGTTGGTTT TATCGTGTTGATTGGACCGAAGTTCCGCGTAGCGAGCCCGCAGCACTGCGTGGTAGGTGGCTGGTTGTTGTTCCGGAAGGC CATGAAGAAGATGGTTGGACCGTTGAAGTTCGTAGCGCACTGGCAGAAGCAGGTGCAGAGCCGGAAGTTACACGTGGTGTT GGTGGCCTGGTTGGGGATTGTGCAGGTGTTGTTAGCCTACTGGCACTGGAAGGTGACGGGGCAGTTCAAACCCTAGTTCTG GTTCGTGAGCTGGATGCAGAAGGTATTGATGCACCCCTGTGGACAGTTACCTTTGGTGCAGTTGATGCAGGTTCGCCGGTT GCACGTCCGGATCAGGCAAAGCTGTGGGGCCTGGGTCAGGTTGCAAGCCTGGAGCGTGGCCCGCGTTGGACCGGTCTGGTT GATCTGCCGCACATGCCGGATCCGGAGCTGAGGGGTAGGCTGACCGCAGTTCTGGCAGGTAGCGAAGATCAGGTTGCAGTT CGTGCAGATGCAGTTAGGGCACGTCGTCTAAGCCCGGCACATGTTACCGCAACCAGCGAATATGCAGTTCCGGGTGGTACT ATTCTGGTTACCGGGGGTACCGCAGGTCTGGGTGCAGAAGTTGCACGTTGGCTGGCAGGGCGTGGTGCAGAGCATCTGGCA CTGGTTAGCCGAAGGGGTCCGGATACCGAAGGTGTTGGTGATCTGACCGCAGAACTGACCCGTCTGGGTGCACGTGTTAGC GTTCATGCATGTGATGTTAGCAGCCGTGAGCCGGTTCGAGAGCTGGTTCATGGTCTGATTGAACAGGGTGATGTTGTTCGT GGGGTTGTTCATGCAGCAGGGCTGCCGCAGCAGGTTGCAATTAACGATATGGATGAAGCAGCATTTGATGAAGTTGTTGCA GCAAAGGCAGGCGGTGCAGTTCATCTGGATGAACTGTGTAGCGATGCAGAACTGTTTCTGCTGTTTAGCAGCGGCGCAGGC GTTTGGGGGAGCGCACGTCAGGGTGCCTATGCAGCAGGTAACGCATTTCTGGATGCATTTGCACGTCATCGTCGTGGTCGT GGTCTGCCGGCAACAAGCGTTGCATGGGGCCTGTGGGCAGCAGGTGGTATGACCGGTGATGAAGAAGCAGTTAGCTTTCTG CGTGAACGTGGGGTTCGTGCAATGCCGGTACCGCGTGCACTGGCAGCACTGGATCGAGTTCTGGCAAGCGGCGAAACAGCA GTTGTTGTTACAGATGTTGATTGGCCGGCATTTGCAGAAAGCTATACCGCAGCACGTCCGCGTCCGCTACTGGATCGTATT GTTACAACCGCACCGAGCGAACGTGCAGGGGAGCCGGAAACAGAAAGCCTGCGTGATCGTCTGGCAGGGCTGCCGAGGGCA GAACGTACAGCAGAACTGGTTCGTCTGGTTCGTACCAGCACCGCAACCGTTCTGGGTCATGATGATCCGAAGGCAGTTCGT GCAACCACACCGTTTAAGGAACTGGGCTTTGATAGCCTGGCAGCAGTTCGTCTGCGTAACCTGCTGAACGCAGCAACCGGC CTGCGTCTGCCGAGCACACTGGTTTTTGATCATCCGAACGCAAGCGCAGTTGCAGGGTTTCTGGATGCAGAACTG

Table S11 Amino acid sequence of proteins used in Chapter 2.5

Constructs	Amino acid sequence
MK53 MBP–M2harm(2)	MKIEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFPQVAATGDGPDIIFWAHDRFGGY AQSGLLAEITPDKAFQDKLYPFTWDAVRYNGKLIAYPIAVEALSLIYNKDLLPNPPKTWEEIPALDKELKA KGKSALMFNLQEPYFTWPLIAADGGYAFKYENGKYDIKDVGVDNAGAKAGLTFLVDLIKNKHMNADTDYSI AEAAFNKGETAMTINGPWAWSNIDTSKVNYGVTVLPTFKGQPSKPFVGVLSAGINAASPNKELAKEFLENY LLTDEGLEAVNKDKPLGAVALKSYEEELVKDPRIAATMENAQKGEIMPNIPQMSAFWYAVRTAVINAASGR QTVDEALKDAQTNSSSNNNNNNNNNLGPGAAHYGSAAPATTAPVDEPIAIVGMACRLPGEVDSPERLWEL ITSGRDSAAEVPDDRGWVPDELMASDAAGTRRAHGNFMAGAGDFDAAFFGISPREALAMDPQQRQALETTW EALESAGIPPETLRGSDTGVFVGMSHQGYATGRPRPEDGVDGYLLTGNTASVASGRIAYVLGLEGPALTVD TACSSSLVALHTACGSLRDGDCGLAVAGGVSVMAGPEVFTEFSRQGALSPDGRCKPFSDEADGFGLGEGSA FVVLQRLSDARREGRRVLGVVAGSAVNQDGASNGLSAPSGVAQQRVIRRAWARAGITGADVAVVEAHGTGT RLGDPVEASALLATYGKSRGSSGPVLLGSVKSNIGHAQAAAGVAGVIKVLLGLERGVVPPMLCRGERSGLI DWSSGEIELADGVREWSPAADGVRRAGVSAFGVSGTNAHVIIAEPPEPEPVPQPRRMLPATGVVPVVLSAR TGAALRAQAGRLADHLAAHPGIAPADVSWTMARARQHFEERAAVLAADTAEAVHRLRAVADGAVVPGVVTG SASDGGSVFVFPGQGAQWEGMARELLPVPVFAESIAECDAVLSEVAGFSVSEVLEPRPDAPSLERVDVVQP VLFAVMVSLARLWRACGAVPSAVIGHSQGEIAAAVVAGALSLEDGMRVVARRSRAVRAVAGRGSMLSVRGG RSDVEKLLADDSWTGRLEVAAVNGPDAVVVAGDAQAAREFLEYCEGVGIRARAIPVDYASHTAHVEPVRDE LVQALAGITPRRAEVPFFSTLTGDFLDGTELDAGYWYRNLRHPVEFHSAVQALTDQGYATFIEVSPHPVLA SSVQETLDDAESDAAVLGTLERDAGDADRFLTALADAHTRGVAVDWEAVLGRAGLVDLPGYPFQGKRFWLL PDRTTPRDELDGWFYRVDWTEVPRSEPAALRGRWLVVVPEGHEEDGWTVEVRSALAEAGAEPEVTRGVGGL VGDCAGVVSLLALEGDGAVQTLVLVRELDAEGIDAPLWTVTFGAVDAGSPVARPDQAKLWGLGQVASLERG PRWTGLVDLPHMPDPELRGRLTAVLAGSEDQVAVRADAVRARRLSPAHVTATSEYAVPGGTILVTGGTAGL GAEVARWLAGRGAEHLALVSRRGPDTEGVGDLTAELTRLGARVSVHACDVSSREPVRELVHGLIEQGDVVR GVVHAAGLPQQVAINDMDEAAFDEVVAAKAGGAVHLDELCSDAELFLLFSSGAGVWGSARQGAYAAGNAFL DAFARHRRGRGLPATSVAWGLWAAGGMTGDEEAVSFLRERGVRAMPVPRALAALDRVLASGETAVVVTDVD WPAFAESYTAARPRPLLDRIVTTAPSERAGEPETESLRDRLAGLPRAERTAELVRLVRTSTATVLGHDDPK AVRATTPFKELGFDSLAAVRLRNLLNAATGLRLPSTLVFDHPNASAVAGFLDAELGTEVRGEAPSALAGLD ALEAAALPEVPATEREELVQRLERMLAALRPVAQAADASGTGANPSGDDLGEAGVDELLEALGRELDGD
MK62 (3)M2harm–TE	MASTDSEKVAEYLRRATLDLRAARQRIRELESAAPATTAPVDEPIAIVGMACRLPGEVDSPERLWELITSG RDSAAEVPDDRGWVPDELMASDAAGTRRAHGNFMAGAGDFDAAFFGISPREALAMDPQQRQALETTWEALE SAGIPPETLRGSDTGVFVGMSHQGYATGRPRPEDGVDGYLLTGNTASVASGRIAYVLGLEGPALTVDTACS SSLVALHTACGSLRDGDCGLAVAGGVSVMAGPEVFTEFSRQGALSPDGRCKPFSDEADGFGLGEGSAFVVL QRLSDARREGRRVLGVVAGSAVNQDGASNGLSAPSGVAQQRVIRRAWARAGITGADVAVVEAHGTGTRLGD PVEASALLATYGKSRGSSGPVLLGSVKSNIGHAQAAAGVAGVIKVLLGLERGVVPPMLCRGERSGLIDWSS GEIELADGVREWSPAADGVRRAGVSAFGVSGTNAHVIIAEPPEPEPVPQPRRMLPATGVVPVVLSARTGAA LRAQAGRLADHLAAHPGIAPADVSWTMARARQHFEERAAVLAADTAEAVHRLRAVADGAVVPGVVTGSASD GGSVFVFPGQGAQWEGMARELLPVPVFAESIAECDAVLSEVAGFSVSEVLEPRPDAPSLERVDVVQPVLFA VMVSLARLWRACGAVPSAVIGHSQGEIAAAVVAGALSLEDGMRVVARRSRAVRAVAGRGSMLSVRGGRSDV EKLLADDSWTGRLEVAAVNGPDAVVVAGDAQAAREFLEYCEGVGIRARAIPVDYASHTAHVEPVRDELVQA LAGITPRRAEVPFFSTLTGDFLDGTELDAGYWYRNLRHPVEFHSAVQALTDQGYATFIEVSPHPVLASSVQ ETLDDAESDAAVLGTLERDAGDADRFLTALADAHTRGVAVDWEAVLGRAGLVDLPGYPFQGKRFWLLPDRT TPRDELDGWFYRVDWTEVPRSEPAALRGRWLVVVPEGHEEDGWTVEVRSALAEAGAEPEVTRGVGGLVGDC AGVVSLLALEGDGAVQTLVLVRELDAEGIDAPLWTVTFGAVDAGSPVARPDQAKLWGLGQVASLERGPRWT GLVDLPHMPDPELRGRLTAVLAGSEDQVAVRADAVRARRLSPAHVTATSEYAVPGGTILVTGGTAGLGAEV ARWLAGRGAEHLALVSRRGPDTEGVGDLTAELTRLGARVSVHACDVSSREPVRELVHGLIEQGDVVRGVVH AAGLPQQVAINDMDEAAFDEVVAAKAGGAVHLDELCSDAELFLLFSSGAGVWGSARQGAYAAGNAFLDAFA RHRRGRGLPATSVAWGLWAAGGMTGDEEAVSFLRERGVRAMPVPRALAALDRVLASGETAVVVTDVDWPAF AESYTAARPRPLLDRIVTTAPSERAGEPETESLRDRLAGLPRAERTAELVRLVRTSTATVLGHDDPKAVRA TTPFKELGFDSLAAVRLRNLLNAATGLRLPSTLVFDHPNASAVAGFLDAELSGTPAREASSALRDGYRQAG VSGRVRSYLDLLAGLSDFREHFDGSDGFSLDLVDMADGPGEVTVICCAGTAAISGPHEFTRLAGALRGIAP VRAVPQPGYEEGEPLPSSMAAVAAVQADAVIRTQGDKPFVVAGHSAGALMAYALATELLDRGHPPRGVVLI DVYPPGHQDAMNAWLEELTATLFDRETVRMDDTRLTALGAYDRLTGQWRPRETGLPTLLVSAGEPMGPWPD DSWKPTWPFEHDTVAVPGDHFTMVQEHADAIARHIDAWLGGGNSSSVDKLAAALEHHHHHH

Table S11 – continued from previous page

Constructs	Amino acid sequence
MK114 MBP–M2harm–TE	MKIEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFPQVAATGDGPDIIFWAHDRFGGY AQSGLLAEITPDKAFQDKLYPFTWDAVRYNGKLIAYPIAVEALSLIYNKDLLPNPPKTWEEIPALDKELKA KGKSALMFNLQEPYFTWPLIAADGGYAFKYENGKYDIKDVGVDNAGAKAGLTFLVDLIKNKHMNADTDYSI AEAAFNKGETAMTINGPWAWSNIDTSKVNYGVTVLPTFKGQPSKPFVGVLSAGINAASPNKELAKEFLENY LLTDEGLEAVNKDKPLGAVALKSYEEELVKDPRIAATMENAQKGEIMPNIPQMSAFWYAVRTAVINAASGR QTVDEALKDAQTNSSSNNNNNNNNNNLGPGAAHYGSEPIAIVGMACRLPGEVDSPERLWELITSGRDSAAE VPDDRGWVPDELMASDAAGTRRAHGNFMAGAGDFDAAFFGISPREALAMDPQQRQALETTWEALESAGIPP ETLRGSDTGVFVGMSHQGYATGRPRPEDGVDGYLLTGNTASVASGRIAYVLGLEGPALTVDTACSSSLVAL HTACGSLRDGDCGLAVAGGVSVMAGPEVFTEFSRQGALSPDGRCKPFSDEADGFGLGEGSAFVVLQRLSDA RREGRRVLGVVAGSAVNQDGASNGLSAPSGVAQQRVIRRAWARAGITGADVAVVEAHGTGTRLGDPVEASA LLATYGKSRGSSGPVLLGSVKSNIGHAQAAAGVAGVIKVLLGLERGVVPPMLCRGERSGLIDWSSGEIELA DGVREWSPAADGVRRAGVSAFGVSGTNAHVIIAEPPEPEPVPQPRRMLPATGVVPVVLSARTGAALRAQAG RLADHLAAHPGIAPADVSWTMARARQHFEERAAVLAADTAEAVHRLRAVADGAVVPGVVTGSASDGGSVFV FPGQGAQWEGMARELLPVPVFAESIAECDAVLSEVAGFSVSEVLEPRPDAPSLERVDVVQPVLFAVMVSLA RLWRACGAVPSAVIGHSQGEIAAAVVAGALSLEDGMRVVARRSRAVRAVAGRGSMLSVRGGRSDVEKLLAD DSWTGRLEVAAVNGPDAVVVAGDAQAAREFLEYCEGVGIRARAIPVDYASHTAHVEPVRDELVQALAGITP RRAEVPFFSTLTGDFLDGTELDAGYWYRNLRHPVEFHSAVQALTDQGYATFIEVSPHPVLASSVQETLDDA ESDAAVLGTLERDAGDADRFLTALADAHTRGVAVDWEAVLGRAGLVDLPGYPFQGKRFWLLPDRTTPRDEL DGWFYRVDWTEVPRSEPAALRGRWLVVVPEGHEEDGWTVEVRSALAEAGAEPEVTRGVGGLVGDCAGVVSL LALEGDGAVQTLVLVRELDAEGIDAPLWTVTFGAVDAGSPVARPDQAKLWGLGQVASLERGPRWTGLVDLP HMPDPELRGRLTAVLAGSEDQVAVRADAVRARRLSPAHVTATSEYAVPGGTILVTGGTAGLGAEVARWLAG RGAEHLALVSRRGPDTEGVGDLTAELTRLGARVSVHACDVSSREPVRELVHGLIEQGDVVRGVVHAAGLPQ QVAINDMDEAAFDEVVAAKAGGAVHLDELCSDAELFLLFSSGAGVWGSARQGAYAAGNAFLDAFARHRRGR GLPATSVAWGLWAAGGMTGDEEAVSFLRERGVRAMPVPRALAALDRVLASGETAVVVTDVDWPAFAESYTA ARPRPLLDRIVTTAPSERAGEPETESLRDRLAGLPRAERTAELVRLVRTSTATVLGHDDPKAVRATTPFKE LGFDSLAAVRLRNLLNAATGLRLPSTLVFDHPNASAVAGFLDAELSGTPAREASSALRDGYRQAGVSGRVR SYLDLLAGLSDFREHFDGSDGFSLDLVDMADGPGEVTVICCAGTAAISGPHEFTRLAGALRGIAPVRAVPQ PGYEEGEPLPSSMAAVAAVQADAVIRTQGDKPFVVAGHSAGALMAYALATELLDRGHPPRGVVLIDVYPPG HQDAMNAWLEELTATLFDRETVRMDDTRLTALGAYDRLTGQWRPRETGLPTLLVSAGEPMGPWPDDSWKPT WPFEHDTVAVPGDHFTMVQEHADAIARHIDAWLGGGNSSSVDKLAAALEHHHHHH
ADD01 KR1–ACP1–M2harm– TE	MDEVSALRYRIEWRPTGAGEPARLDGTWLVAKYAGTADETSTAAREALESAGARVRELVVDARCGRDELAE RLRSVGEVAGVLSLLAVDEAEPEEAPLALASLADTLSLVQAMVSAELGCPLWTVTESAVATGPFERVRNAA HGALWGVGRVIALENPAVWGGLVDVPAGSVAELARHLAAVVSGGAGEDQLALRADGVYGRRWVRAAAPATD DEWKPTGTVLVTGGTGVGGQIARWLARRGAPHLLLVSRSGPDADGAGELVAELEALGARTTVAACDVTDR ESVRELLGGIGDDVPLSAVFHAAATLDDGTVDTLTGERIERASRAKVLGARNLHELTRELDLTAFVLFSSF ASAFGAPGLGGYAPGNAYLDGLAQQRRSDGLPATAVAWGTWAGSGMAEGPVADRFRRHGVIEMPPETACRA LQNALDRAEVCPIVIDVRWDRFLLAYTAQRPTRLFDEIDDARRAAPQAAAEPRVGAHMLASLPAPEREKAL FELVRSHAAAVLGHASAERVPADQAFAELGVDSLSALELRNRLGAATGVRLPTTTVFDHPDVRTLAAHLAA ELGGATGAEQAAPATTAPVDEPIAIVGMACRLPGEVDSPERLWELITSGRDSAAEVPDDRGWVPDELMASD AAGTRRAHGNFMAGAGDFDAAFFGISPREALAMDPQQRQALETTWEALESAGIPPETLRGSDTGVFVGMSH QGYATGRPRPEDGVDGYLLTGNTASVASGRIAYVLGLEGPALTVDTACSSSLVALHTACGSLRDGDCGLAV AGGVSVMAGPEVFTEFSRQGALSPDGRCKPFSDEADGFGLGEGSAFVVLQRLSDARREGRRVLGVVAGSAV NQDGASNGLSAPSGVAQQRVIRRAWARAGITGADVAVVEAHGTGTRLGDPVEASALLATYGKSRGSSGPVL LGSVKSNIGHAQAAAGVAGVIKVLLGLERGVVPPMLCRGERSGLIDWSSGEIELADGVREWSPAADGVRRA GVSAFGVSGTNAHVIIAEPPEPEPVPQPRRMLPATGVVPVVLSARTGAALRAQAGRLADHLAAHPGIAPAD VSWTMARARQHFEERAAVLAADTAEAVHRLRAVADGAVVPGVVTGSASDGGSVFVFPGQGAQWEGMARELL PVPVFAESIAECDAVLSEVAGFSVSEVLEPRPDAPSLERVDVVQPVLFAVMVSLARLWRACGAVPSAVIGH SQGEIAAAVVAGALSLEDGMRVVARRSRAVRAVAGRGSMLSVRGGRSDVEKLLADDSWTGRLEVAAVNGPD AVVVAGDAQAAREFLEYCEGVGIRARAIPVDYASHTAHVEPVRDELVQALAGITPRRAEVPFFSTLTGDFL DGTELDAGYWYRNLRHPVEFHSAVQALTDQGYATFIEVSPHPVLASSVQETLDDAESDAAVLGTLERDAGD ADRFLTALADAHTRGVAVDWEAVLGRAGLVDLPGYPFQGKRFWLLPDRTTPRDELDGWFYRVDWTEVPRSE PAALRGRWLVVVPEGHEEDGWTVEVRSALAEAGAEPEVTRGVGGLVGDCAGVVSLLALEGDGAVQTLVLVR ELDAEGIDAPLWTVTFGAVDAGSPVARPDQAKLWGLGQVASLERGPRWTGLVDLPHMPDPELRGRLTAVLA GSEDQVAVRADAVRARRLSPAHVTATSEYAVPGGTILVTGGTAGLGAEVARWLAGRGAEHLALVSRRGPDT EGVGDLTAELTRLGARVSVHACDVSSREPVRELVHGLIEQGDVVRGVVHAAGLPQQVAINDMDEAAFDEVV AAKAGGAVHLDELCSDAELFLLFSSGAGVWGSARQGAYAAGNAFLDAFARHRRGRGLPATSVAWGLWAAGG MTGDEEAVSFLRERGVRAMPVPRALAALDRVLASGETAVVVTDVDWPAFAESYTAARPRPLLDRIVTTAPS ERAGEPETESLRDRLAGLPRAERTAELVRLVRTSTATVLGHDDPKAVRATTPFKELGFDSLAAVRLRNLLN AATGLRLPSTLVFDHPNASAVAGFLDAELSGTPAREASSALRDGYRQAGVSGRVRSYLDLLAGLSDFREHF DGSDGFSLDLVDMADGPGEVTVICCAGTAAISGPHEFTRLAGALRGIAPVRAVPQPGYEEGEPLPSSMAAV AAVQADAVIRTQGDKPFVVAGHSAGALMAYALATELLDRGHPPRGVVLIDVYPPGHQDAMNAWLEELTATL FDRETVRMDDTRLTALGAYDRLTGQWRPRETGLPTLLVSAGEPMGPWPDDSWKPTWPFEHDTVAVPGDHFT MVQEHADAIARHIDAWLGGGNSSSVDKLAAALEHHHHHH

Table S11 – continued from previous page

Constructs	Amino acid sequence
RSG34 (3)M3–TE	MASTDSEKVAEYLRRATLDLRAARQRIRELESDPIAIVSMACRLPGGVNTPQRLWELLREGGETLSGFPTD RGWDLARLHHPDPDNPGTSYVDKGGFLDDAAGFDAEFFGVSPREAAAMDPQQRLLLETSWELVENAGIDPH SLRGTATGVFLGVAKFGYGEDTAAAEDVEGYSVTGVAPAVASGRISYTMGLEGPSISVDTACSSSLVALHL AVESLRKGESSMAVVGGAAVMATPGVFVDFSRQRALAADGRSKAFGAGADGFGFSEGVTLVLLERLSEARR NGHEVLAVVRGSALNQDGASNGLSAPSGPAQRRVIRQALESCGLEPGDVDAVEAHGTGTALGDPIEANALL DTYGRDRDADRPLWLGSVKSNIGHTQAAAGVTGLLKVVLALRNGELPATLHVEEPTPHVDWSSGGVALLAG NQPWRRGERTRRARVSAFGISGTNAHVIVEEAPEREHRETTAHDGRPVPLVVSARTTAALRAQAAQIAELL ERPDADLAGVGLGLATTRARHEHRAAVVASTREEAVRGLREIAAGAATADAVVEGVTEVDGRNVVFLFPGQ GSQWAGMGAELLSSSPVFAGKIRACDESMAPMQDWKVSDVLRQAPGAPGLDRVDVVQPVLFAVMVSLAELW RSYGVEPAAVVGHSQGEIAAAHVAGALTLEDAAKLVVGRSRLMRSLSGEGGMAAVALGEAAVRERLRPWQD RLSVAAVNGPRSVVVSGEPGALRAFSEDCAAEGIRVRDIDVDYASHSPQIERVREELLETTGDIAPRPARV TFHSTVESRSMDGTELDARYWYRNLRETVRFADAVTRLAESGYDAFIEVSPHPVVVQAVEEAVEEADGAED AVVVGSLHRDGGDLSAFLRSMATAHVSGVDIRWDVALPGAAPFALPTYPFQRKRYWLQPAAPAAASDELAY RVSWTPIEKPESGNLDGDWLVVTPLISPEWTEMLCEAINANGGRALRCEVDTSASRTEMAQAVAQAGTGFR GVLSLLSSDESACRPGVPAGAVGLLTLVQALGDAGVDAPVWCLTQGAVRTPADDDLARPAQTTAHGFAQVA GLELPGRWGGVVDLPESVDDAALRLLVAVLRGGGRAEDHLAVRDGRLHGRRVVRASLPQSGSRSWTPHGTV LVTGAASPVGDQLVRWLADRGAERLVLAGACPGDDLLAAVEEAGASAVVCAQDAAALREALGDEPVTALVH AGTLTNFGSISEVAPEEFAETIAAKTALLAVLDEVLGDRAVEREVYCSSVAGIWGGAGMAAYAAGSAYLDA LAEHHRARGRSCTSVAWTPWALPGGAVDDGYLRERGLRSLSADRAMRTWERVLAAGPVSVAVADVDWPVLS EGFAATRPTALFAELAGRGGQAEAEPDSGPTGEPAQRLAGLSPDEQQENLLELVANAVAEVLGHESAAEIN VRRAFSELGLDSLNAMALRKRLSASTGLRLPASLVFDHPTVTALAQHTSQLDSGTPAREASSALRDGYRQA GVSGRVRSYLDLLAGLSDFREHFDGSDGFSLDLVDMADGPGEVTVICCAGTAAISGPHEFTRLAGALRGIA PVRAVPQPGYEEGEPLPSSMAAVAAVQADAVIRTQGDKPFVVAGHSAGALMAYALATELLDRGHPPRGVVL IDVYPPGHQDAMNAWLEELTATLFDRETVRMDDTRLTALGAYDRLTGQWRPRETGLPTLLVSAGEPMGPWP DDSWKPTWPFEHDTVAVPGDHFTMVQEHADAIARHIDAWLGGGNSSSVDKLAAALEHHHHHH

Table S12 Inter–domain cross–links from DEBS lysines within MBP–M2(2)

res 1	res 2	Ex.#1	Ex.#2	Ex.#3	Ex.#4	pos 1	pos 2	priority	type	relevance
727K	279E	3	4	2	8	KS2	MBP	1st	inter	strong
727K	281L	1	6	-	-	KS2	MBP	2nd (1st)	inter	strong
727K	285L	-	4	3	2	KS2	MBP	1st	inter	strong
727K	286L	-	11	4	1	KS2	MBP	1st	inter	strong
727K	390G	5	1	6	1	KS2	MBP	1st	inter	strong
727K	397T	10	2	-	-	KS2	linker	1st	inter	strong
727K	1123Y	1	-	3	-	KS2	AT2	2nd (1st)	inter	strong
727K	1470T	-	3	1	-	KS2	KR2	2nd (1st)	inter	strong
1071K	48F	-	5	3	1	AT2	MBP	1st	inter	strong
1071K	894A	4	-	7	1	AT2	KAL	1st	inter	strong
1071K	1623Q	-	5	4	2	AT2	KR2	1st	inter	strong
1273K	1474E	1	6	-	-	PAL	KR2	2nd (1st)	inter	strong
1273K	1475Y	1	2	2	-	PAL	KR2	1st	inter	strong
1273K	1476A	1	2	1	2	PAL	KR2	1st	inter	strong
1273K	1482I	3	2	-	-	PAL	KR2	1st	inter	strong
1273K	1503G	5	4	-	-	PAL	KR2	1st	inter	strong
1273K	1683A	-	1	1	7	PAL	KR2	2nd (1st)	inter	strong
1775K	1295V	-	-	7	1	ACP2	KR2	2nd (1st)	inter	strong
1775K	1303S	3	4	2	3	ACP2	KR2	1st	inter	strong
1775K	1304E	-	-	3	1	ACP2	KR2	2nd (1st)	inter	strong
1775K	1733A	-	7	7	1	ACP2	KR2	1st	inter	strong
1775K	1734G	2	3	-	-	ACP2	KR2	1st	inter	strong
1775K	1861E	-	-	4	2	ACP2	DD	1st	inter	strong
1784K	539G	-	-	2	2	ACP2	KS2	1st	inter	strong
1784K	554I	1	2	7	-	ACP2	KS2	1st	inter	strong
1784K	556Y	6	5	-	-	ACP2	KS2	1st	inter	strong
1784K	560L	7	1	1	1	ACP2	KS2	2nd (1st)	inter	strong
1784K	894A	6	-	1	-	ACP2	KAL	2nd (1st)	inter	strong
1784K	1303S	3	3	-	-	ACP2	KR2	1st	inter	strong
1784K	1468H	4	3	1	-	ACP2	KR2	1st	inter	strong
1784K	1469V	1	1	6	-	ACP2	KR2	2nd (1st)	inter	strong
1784K	1473S	1	1	2	-	ACP2	KR2	2nd (1st)	inter	strong
1784K	1474E	-	2	2	-	ACP2	KR2	1st	inter	strong
1784K	1733A	3	-	1	-	ACP2	KR2	2nd (1st)	inter	strong
1784K	1845L	2	4	-	-	ACP2	DD2	1st	inter	strong
727K	370S	-	-	3	-	KS2	MBP	2nd	inter	medium
727K	371S	-	-	2	-	KS2	MBP	2nd	inter	medium
727K	373N	-	-	2	-	KS2	MBP	2nd	inter	medium
727K	388H	-	-	8	-	KS2	MBP	2nd	inter	medium
727K	392A	3	-	-	-	KS2	linker	2nd	inter	medium
727K	393A	1	1	-	-	KS2	linker	3rd	inter	medium
727K	1124A	1	1	1	-	KS2	AT2	3rd	inter	medium
1071K	838M	-	6	-	-	AT2	KAL	2nd	inter	medium
1071K	902A	2	-	-	-	AT2	KAL	2nd	inter	medium
1071K	903E	-	-	3	-	AT2	KAL	2nd	inter	medium
1273K	871H	-	1	2	-	PAL	KAL	2nd	inter	medium
1273K	872P	-	-	2	-	PAL	KAL	2nd	inter	medium
1273K	1123Y	-	-	3	-	PAL	AT2	2nd	inter	medium
1273K	1125S	2	-	-	-	PAL	AT2	2nd	inter	medium
1273K	1464L	1	1	-	-	PAL	KR2	3rd	inter	medium
1273K	1465S	1	2	-	-	PAL	KR2	2nd	inter	medium
1273K	1471A	1	-	1	-	PAL	KR2	3rd	inter	medium

Table S12 – continued from previous page

res 1	res 2	Ex.#1	Ex.#2	Ex.#3	Ex.#4	pos 1	pos 2	priority	type	relevance
1273K	1473S	-	2	-	-	PAL	KR2	2nd	inter	medium
1273K	1484V	-	3	-	-	PAL	KR2	2nd	inter	medium
1775K	538D	-	-	1	-	ACP2	KS2	no	inter	medium
1775K	539G	-	-	2	-	ACP2	KS2	2nd	inter	medium
1775K	540Y	-	-	6	-	ACP2	KS2	2nd	inter	medium
1775K	554I	-	-	1	-	ACP2	KS2	no	inter	medium
1775K	556Y	1	-	7	-	ACP2	KS2	2nd	inter	medium
1775K	1732R	-	-	1	-	ACP2	KR2	no	inter	medium
1775K	1735E	2	1	-	-	ACP2	KR2	2nd	inter	medium
1784K	540Y	-	-	3	-	ACP2	KS2	2nd	inter	medium
1784K	598V	-	4	-	-	ACP2	KS2	2nd	inter	medium
1784K	599S	-	3	-	-	ACP2	KS2	2nd	inter	medium
1784K	600V	-	4	-	-	ACP2	KS2	2nd	inter	medium
1784K	1476A	-	1	1	-	ACP2	KR2	3rd	inter	medium
1784K	1833E	-	1	1	-	ACP2	DD2	3rd	inter	medium
1784K	1836G	-	3	-	-	ACP2	DD2	2nd	inter	medium
1784K	1853P	2	-	1	-	ACP2	DD2	2nd	inter	medium
1784K	1854E	3	1	-	-	ACP2	DD2	2nd	inter	medium

Cross–links from four individual experiments are shown (Ex.#1–4), the position within the construct is indicated (KAL, KS–AT linker; PAL, AT–KR linker). Priority (1st, 2nd, 3rd, or no) based on number of counts for each hit. 1st – multiple hits in multiple hits, 2nd (1st) – at least one single– and one multi–count hit, 2nd – one multi–count hit, 3rd – min. two single hits, and "no" – a single hit. Type – inter– and intra–domain hits are indicated. Relevance for modeling was based on clustering of hits. All "1st" and "2nd (1st)" hits have a strong relevancy, additionally all hits that cluster within a range of two residues or close to strong relevancy hit were selected to have a medium relevancy. Experiments #1&2 were performed at 14 mg/mL M2 concentration and Ex. #3&4 at 1 mg/mL.

Table S13 Inter–domain cross–links from MBP lysines within MBP–M2(2). Same nomenclature as in Table S12.

res 1	res 2	Ex.#1	Ex.#2	Ex.#3	Ex.#4	pos 1	pos 2	priority	type	relevance
47K	1408L	-	-	3	3	MBP	KR2	1st	inter	strong
143K	1408L	-	-	2	-	MBP	KR2	2nd	inter	strong
145K	1408L	1	-	1	5	MBP	KR2	2nd (1st)	inter	strong
180K	667Q	-	3	-	-	MBP	KS2	2nd	inter	strong
180K	668D	-	3	-	-	MBP	KS2	2nd	inter	strong
180K	672N	-	5	1	-	MBP	KS2	2nd (1st)	inter	strong
180K	1088N	-	5	2	-	MBP	AT2	1st	inter	strong
180K	1104E	-	2	2	-	MBP	AT2	1st	inter	strong
84K	496T	-	9	-	-	MBP	KS2	2nd	inter	medium
84K	1701T	-	6	-	-	MBP	KR2	2nd	inter	medium
84K	1703V	-	3	-	-	MBP	KR2	2nd	inter	medium
203K	668D	-	3	-	-	MBP	KS2	2nd	inter	medium
203K	672N	-	4	-	-	MBP	KS2	2nd	inter	medium
240K	522S	-	-	-	2	MBP	KS2	2nd	inter	medium
240K	525G	-	-	-	2	MBP	KS2	2nd	inter	medium
240K	531P	-	-	-	2	MBP	KS2	2nd	inter	medium
240K	532R	-	-	-	2	MBP	KS2	2nd	inter	medium

Table S14 Intra–domain cross–links from DEBS lysines within MBP–M2(2). Same nomenclature as in Table S12.

res 1	res 2	Ex.#1	Ex.#2	Ex.#3	Ex.#4	pos 1	pos 2	priority	type	relevance
727K	693A	2	9	7	6	KS2	KS2	1st	intra	strong
727K	699D	7	3	2	2	KS2	KS2	1st	intra	strong
727K	703V	1	2	1	4	KS2	KS2	1st	intra	strong
727K	778S	3	10	3	2	KS2	KS2	1st	intra	strong
727K	782D	1	-	1	2	KS2	KS2	2nd (1st)	intra	strong
727K	784S	9	11	10	14	KS2	KS2	1st	intra	strong
727K	785S	8	6	1	-	KS2	KS2	1st	intra	strong
727K	786G	4	4	2	2	KS2	KS2	1st	intra	strong
727K	787E	2	7	-	-	KS2	KS2	1st	intra	strong
727K	796E	2	-	2	4	KS2	KS2	1st	intra	strong
727K	816G	1	1	2	2	KS2	KS2	1st	intra	strong
741K	702V	3	-	-	2	KS2	KS2	1st	intra	strong
1071K	1104E	19	8	43	26	AT2	AT2	1st	intra	strong
1071K	1106L	4	2	3	7	AT2	AT2	1st	intra	strong
1071K	1107E	28	11	24	24	AT2	AT2	1st	intra	strong
1071K	1108Y	53	68	40	28	AT2	AT2	1st	intra	strong
1071K	1423W	-	2	1	1	AT2	KR2	2nd (1st)	intra	strong
1273K	1251V	-	-	5	3	PAL	PAL	1st	intra	strong
1273K	1254E	-	2	-	2	PAL	PAL	1st	intra	strong
1273K	1286D	3	4	2	-	PAL	PAL	1st	intra	strong
1407K	1396A	1	-	5	3	KR2	KR2	1st	intra	strong
1407K	1397G	10	1	2	-	KR2	KR2	1st	intra	strong
1407K	1399P	3	1	-	-	KR2	KR2	2nd (1st)	intra	strong
1407K	1401A	52	4	4	5	KR2	KR2	1st	intra	strong
1407K	1576I	1	-	4	1	KR2	KR2	2nd (1st)	intra	strong
1407K	1578D	3	2	7	10	KR2	KR2	1st	intra	strong
1407K	1579M	5	4	24	5	KR2	KR2	1st	intra	strong
1407K	1580D	2	-	6	-	KR2	KR2	1st	intra	strong
1407K	1583A	1	1	3	-	KR2	KR2	2nd (1st)	intra	strong
1407K	1584F	-	1	7	1	KR2	KR2	2nd (1st)	intra	strong
1407K	1585D	16	7	6	2	KR2	KR2	1st	intra	strong
1407K	1586E	8	2	6	5	KR2	KR2	1st	intra	strong
1407K	1587V	12	3	-	-	KR2	KR2	1st	intra	strong
1775K	1759R	1	-	10	1	ACP2	ACP2	2nd (1st)	intra	strong
1784K	1808L	-	1	6	-	ACP2	ACP2	2nd (1st)	intra	strong
1784K	1812L	2	2	-	-	ACP2	ACP2	1st	intra	strong
1784K	1814F	-	1	4	-	ACP2	ACP2	2nd (1st)	intra	strong
1784K	1815D	6	3	3	-	ACP2	ACP2	1st	intra	strong
1784K	1816H	1	5	-	-	ACP2	ACP2	2nd (1st)	intra	strong
1784K	1818N	1	3	-	-	ACP2	ACP2	2nd (1st)	intra	strong
1784K	1819A	2	2	2	-	ACP2	ACP2	1st	intra	strong
1784K	1820S	6	9	7	-	ACP2	ACP2	1st	intra	strong
1784K	1821A	4	4	1	-	ACP2	ACP2	1st	intra	strong
1784K	1822V	6	3	2	-	ACP2	ACP2	1st	intra	strong
1784K	1829E	1	2	2	-	ACP2	ACP2	1st	intra	strong
623K	525G	-	-	2	-	KS2	KS2	2nd	intra	medium
623K	526Y	1	-	-	-	KS2	KS2	no	intra	medium
623K	539G	-	-	4	-	KS2	KS2	2nd	intra	medium
623K	540Y	-	-	1	-	KS2	KS2	no	intra	medium
623K	598V	-	1	-	-	KS2	KS2	no	intra	medium
623K	599S	-	3	-	-	KS2	KS2	2nd	intra	medium

Table S14 – continued from previous page

res 1	res 2	Ex.#1	Ex.#2	Ex.#3	Ex.#4	pos 1	pos 2	priority	type	relevance
623K	623K	-	-	2	-	KS2	KS2	2nd	intra	medium
623K	624P	-	-	1	-	KS2	KS2	no	intra	medium
727K	696T	1	1	-	-	KS2	KS2	3rd	intra	medium
727K	697G	1	1	-	-	KS2	KS2	3rd	intra	medium
727K	698A	2	-	-	-	KS2	KS2	2nd	intra	medium
727K	700V	1	-	2	-	KS2	KS2	2nd	intra	medium
727K	781I	-	2	-	-	KS2	KS2	2nd	intra	medium
727K	815S	2	-	1	-	KS2	KS2	2nd	intra	medium
727K	822I	-	2	-	-	KS2	KS2	2nd	intra	medium
727K	823I	1	-	2	-	KS2	KS2	2nd	intra	medium
741K	699D	1	-	-	-	KS2	KS2	no	intra	medium
741K	700V	1	-	-	-	KS2	KS2	no	intra	medium
741K	703V	2	-	-	-	KS2	KS2	2nd	intra	medium
741K	712L	3	-	1	-	KS2	KS2	2nd	intra	medium
741K	716V	2	-	-	-	KS2	KS2	2nd	intra	medium
741K	717E	1	-	-	-	KS2	KS2	no	intra	medium
741K	721L	-	-	2	-	KS2	KS2	2nd	intra	medium
741K	722L	1	-	-	-	KS2	KS2	no	intra	medium
758K	702V	1	-	-	-	KS2	KS2	no	intra	medium
758K	703V	1	-	1	-	KS2	KS2	3rd	intra	medium
758K	704E	2	-	-	-	KS2	KS2	2nd	intra	medium
1071K	940Q	-	-	-	1	AT2	AT2	no	intra	medium
1071K	941W	-	-	-	1	AT2	AT2	no	intra	medium
1071K	1082L	-	-	1	-	AT2	AT2	no	intra	medium
1071K	1084V	-	-	-	1	AT2	AT2	no	intra	medium
1071K	1123Y	-	-	1	-	AT2	AT2	no	intra	medium
1071K	1124A	-	-	3	-	AT2	AT2	2nd	intra	medium
1071K	1126H	-	-	1	-	AT2	AT2	no	intra	medium
1071K	1128A	-	-	1	-	AT2	AT2	no	intra	medium
1071K	1129H	-	1	-	-	AT2	AT2	no	intra	medium
1273K	1252D	1	1	-	-	PAL	PAL	3rd	intra	medium
1407K	1387W	2	-	-	-	KR2	KR2	2nd	intra	medium
1407K	1388T	5	-	-	-	KR2	KR2	2nd	intra	medium
1407K	1389V	5	-	-	-	KR2	KR2	2nd	intra	medium
1407K	1390T	1	-	-	-	KR2	KR2	no	intra	medium
1407K	1394V	3	-	-	-	KR2	KR2	2nd	intra	medium
1407K	1398S	2	-	-	-	KR2	KR2	2nd	intra	medium
1407K	1426L	-	-	1	-	KR2	KR2	no	intra	medium
1407K	1427V	-	-	1	-	KR2	KR2	no	intra	medium
1407K	1474E	1	-	-	-	KR2	KR2	no	intra	medium
1407K	1475Y	1	-	1	-	KR2	KR2	3rd	intra	medium
1407K	1577N	-	-	1	1	KR2	KR2	3rd	intra	medium
1407K	1581E	-	-	2	-	KR2	KR2	2nd	intra	medium
1407K	1582A	-	-	1	1	KR2	KR2	3rd	intra	medium
1407K	1623Q	1	-	-	-	KR2	KR2	no	intra	medium
1407K	1624G	1	-	-	-	KR2	KR2	no	intra	medium
1407K	1626Y	-	-	4	-	KR2	KR2	2nd	intra	medium
1407K	1627A	-	-	1	-	KR2	KR2	no	intra	medium
1784K	1813V	1	1	-	-	ACP2	ACP2	3rd	intra	medium
1784K	1817P	-	1	-	-	ACP2	ACP2	no	intra	medium
1784K	1823A	-	1	1	-	ACP2	ACP2	3rd	intra	medium
1784K	1824G	1	-	-	-	ACP2	ACP2	no	intra	medium
1784K	1825F	-	1	-	-	ACP2	ACP2	no	intra	medium

Table S14 – continued from previous page

res 1	res 2	Ex.#1	Ex.#2	Ex.#3	Ex.#4	pos 1	pos 2	priority	type	relevance
1784K	1826L	-	1	1	-	ACP2	ACP2	3rd	intra	medium
1784K	1828A	-	2	-	-	ACP2	ACP2	2nd	intra	medium

Table S15 Intra–domain cross–links from MBP lysines within MBP–M2(2). Same nomenclature as in Table S12.

res 1	res 2	Ex.#1	Ex.#2	Ex.#3	Ex.#4	pos 1	pos 2	priority	type	relevance
26K	30K	-	3	1	-	MBP	MBP	2nd (1st)	intra	strong
27K	23E	-	-	2	1	MBP	MBP	2nd	intra	strong
27K	25G	5	6	25	12	MBP	MBP	1st	intra	strong
30K	17G	1	-	9	11	MBP	MBP	1st	intra	strong
30K	19N	-	-	13	-	MBP	MBP	2nd	intra	strong
30K	21L	-	-	3	8	MBP	MBP	1st	intra	strong
30K	22A	3	1	3	-	MBP	MBP	1st	intra	strong
30K	23E	6	11	2	3	MBP	MBP	1st	intra	strong
30K	26K	-	3	1	-	MBP	MBP	2nd (1st)	intra	strong
43K	48F	10	-	-	6	MBP	MBP	1st	intra	strong
47K	41P	4	-	-	-	MBP	MBP	2nd	intra	strong
47K	42D	13	-	-	5	MBP	MBP	1st	intra	strong
47K	48F	-	-	13	14	MBP	MBP	1st	intra	strong
84K	90L	-	2	12	1	MBP	MBP	1st	intra	strong
89K	82P	1	4	-	-	MBP	MBP	2nd (1st)	intra	strong
89K	83D	1	4	-	-	MBP	MBP	2nd (1st)	intra	strong
89K	283N	-	5	3	-	MBP	MBP	1st	intra	strong
138K	279E	-	3	4	1	MBP	MBP	1st	intra	strong
180K	172Y	-	-	2	4	MBP	MBP	1st	intra	strong
201K	202N	-	-	4	5	MBP	MBP	1st	intra	strong
201K	204H	3	-	-	-	MBP	MBP	2nd	intra	strong
201K	210D	-	-	-	2	MBP	MBP	2nd	intra	strong
240K	48F	1	-	1	3	MBP	MBP	2nd (1st)	intra	strong
252K	258P	8	-	1	-	MBP	MBP	2nd (1st)	intra	strong
252K	260V	5	1	-	-	MBP	MBP	2nd (1st)	intra	strong
257K	250T	1	-	-	3	MBP	MBP	2nd (1st)	intra	strong
257K	251F	4	1	-	-	MBP	MBP	2nd (1st)	intra	strong
274K	279E	7	-	-	16	MBP	MBP	1st	intra	strong
296K	307S	-	-	25	11	MBP	MBP	1st	intra	strong
296K	309E	-	-	4	5	MBP	MBP	1st	intra	strong
296K	310E	-	-	1	4	MBP	MBP	2nd (1st)	intra	strong
296K	311E	-	-	1	3	MBP	MBP	2nd (1st)	intra	strong
296K	314K	-	1	1	4	MBP	MBP	2nd (1st)	intra	strong
298K	295N	1	-	1	3	MBP	MBP	2nd (1st)	intra	strong
298K	307S	-	1	12	2	MBP	MBP	2nd (1st)	intra	strong
306K	294V	2	-	-	5	MBP	MBP	1st	intra	strong
306K	297D	-	-	1	4	MBP	MBP	2nd (1st)	intra	strong
314K	90L	-	3	17	2	MBP	MBP	1st	intra	strong
314K	91Y	-	14	7	3	MBP	MBP	1st	intra	strong
314K	287T	-	-	6	4	MBP	MBP	1st	intra	strong
314K	291L	-	-	2	4	MBP	MBP	1st	intra	strong
314K	296K	-	1	1	4	MBP	MBP	2nd (1st)	intra	strong
314K	318I	-	-	12	6	MBP	MBP	1st	intra	strong
314K	323E	-	3	-	1	MBP	MBP	2nd (1st)	intra	strong
16K	39E	2	-	-	-	MBP	MBP	2nd	intra	medium

Table S15 – continued from previous page

res 1	res 2	Ex.#1	Ex.#2	Ex.#3	Ex.#4	pos 1	pos 2	priority	type	relevance
16K	40H	10	-	-	-	MBP	MBP	2nd	intra	medium
47K	18Y	2	-	-	-	MBP	MBP	2nd	intra	medium
47K	19N	2	-	-	-	MBP	MBP	2nd	intra	medium
47K	71Y	-	-	4	-	MBP	MBP	2nd	intra	medium
47K	73Q	-	-	2	-	MBP	MBP	2nd	intra	medium
47K	239S	-	-	-	2	MBP	MBP	2nd	intra	medium
47K	241V	-	-	-	2	MBP	MBP	2nd	intra	medium
201K	207A	1	-	-	4	MBP	MBP	2nd (1st)	intra	medium
201K	209T	-	-	-	9	MBP	MBP	2nd	intra	medium
240K	45E	-	-	1	2	MBP	MBP	2nd	intra	medium
240K	55G	-	-	-	2	MBP	MBP	2nd	intra	medium
240K	56D	-	-	-	4	MBP	MBP	2nd	intra	medium
252K	157F	2	-	-	-	MBP	MBP	2nd	intra	medium
252K	158T	2	-	-	-	MBP	MBP	2nd	intra	medium
278K	271S	2	-	-	1	MBP	MBP	2nd	intra	medium
278K	272P	3	-	-	-	MBP	MBP	2nd	intra	medium
296K	17G	-	-	2	-	MBP	MBP	2nd	intra	medium
296K	19N	-	1	1	-	MBP	MBP	3rd	intra	medium
298K	294V	-	-	-	2	MBP	MBP	2nd	intra	medium
298K	308Y	-	-	4	-	MBP	MBP	2nd	intra	medium
314K	281L	-	-	1	2	MBP	MBP	2nd	intra	medium
314K	284Y	-	-	3	-	MBP	MBP	2nd	intra	medium
314K	295N	-	-	2	-	MBP	MBP	2nd	intra	medium

Table S16   Inter–domain cross–links from DEBS lysines within KR1–ACP1–M2–TE

res 1	res 2	Ex.#1	Ex.#2	pos 1	pos 2	priority	type	relevance
217K	579A	2		KR1	KS2	2nd	inter	medium
217K	580A	1		KR1	KS2	no	inter	medium
217K	1274V		1	KR1	AT2	no	inter	medium
217K	1275N		2	KR1	AT2	2nd	inter	medium
217K	1484W	1		KR1	AT2	no	inter	weak
217K	1485T	1		KR1	AT2	no	inter	weak
495K	477E	1		ACP1	KR1	no	inter	weak
495K	479R		1	ACP1	KR1	no	inter	weak
810K	259G	1		KS2	KR1	no	inter	medium
810K	261G	1		KS2	KR1	no	inter	medium
810K	263L	3		KS2	KR1	2nd	inter	medium
914K	391V		1	KS2	KR1	no	inter	medium
914K	393W		4	KS2	KR1	2nd	inter	medium
914K	395T	2		KS2	KR1	2nd	inter	medium
914K	396W	4		KS2	KR1	2nd	inter	medium
914K	398G	1		KS2	KR1	no	inter	medium
914K	399S		1	KS2	KR1	no	inter	medium
914K	400G		2	KS2	KR1	2nd	inter	medium
914K	578Q	2	1	KS2	ACP-KS linker	2nd (1st)	inter	strong
914K	580A		1	KS2	ACP-KS linker	no	inter	medium
914K	1087D	1		KS2	KAL	no	inter	weak
914K	1091A	1		KS2	KAL	no	inter	weak
914K	1116G	1		KS2	KAL	no	inter	weak
914K	1120V	1		KS2	AT2	no	inter	weak
1258K	394G	1		AT2	KR1	no	inter	weak

Table S16 – continued from previous page

res 1	res 2	Ex.#1	Ex.#2	pos 1	pos 2	priority	type	relevance
1258K	395T	1		AT2	KR1	no	inter	weak
1258K	396W	1		AT2	KR1	no	inter	weak
1258K	400G	1		AT2	KR1	no	inter	weak
1258K	819F	1		AT2	KS2	no	inter	medium
1258K	820G	2		AT2	KS2	2nd	inter	medium
1258K	821L	2		AT2	KS2	2nd	inter	medium
1460K	969D	1		PAL	KS2	no	inter	weak
1460K	970W	1		PAL	KS2	no	inter	weak
1460K	971S	1		PAL	KS2	no	inter	weak
1460K	1310Y	1		PAL	AT2	no	inter	weak
1460K	1311A	1		PAL	AT2	no	inter	weak
1460K	1660S		1	PAL	KR2	no	inter	medium
1460K	1661E	2		PAL	KR2	2nd	inter	medium
1460K	1662Y	2		PAL	KR2	2nd	inter	medium
1460K	1663A	6		PAL	KR2	2nd	inter	medium
1460K	1664V	5		PAL	KR2	2nd	inter	medium
1460K	1669I	1		PAL	KR2	no	inter	weak
1460K	1671V	1		PAL	KR2	no	inter	weak
1460K	1790S	1		PAL	KR2	no	inter	weak
1460K	1793E	1		PAL	KR2	no	inter	weak
1594K	1120V	1		KR2	AT2	no	inter	medium
1594K	1121F	2		KR2	AT2	2nd	inter	medium
1594K	1143E		1	KR2	AT2	no	inter	weak
1594K	1145I		1	KR2	AT2	no	inter	weak
1594K	1584G		1	KR2	AT2	no	inter	weak
1962K	186G		1	ACP2	KR1	no	inter	weak
1962K	1490S	1	2	ACP2	KR2	2nd (1st)	inter	strong
1962K	1920A		2	ACP2	KR2-ACP2 linker	2nd	inter	medium
1962K	1921G	2		ACP2	KR2-ACP2 linker	2nd	inter	medium
1971K	2119E		1	ACP2	TE	no	inter	weak
1971K	2122P	1		ACP2	TE	no	inter	weak
1971K	2123L	1		ACP2	TE	no	inter	weak
2254K	572G	2	1	TE	ACP1-KS2 linker	2nd (1st)	inter	strong
2254K	574T	1		TE	ACP1-KS2 linker	no	inter	medium

Cross–links from two individual experiments are shown (Ex.#1–2), the position within the construct is indicated (KAL, KS–AT linker; PAL, AT–KR linker). Priority (1st, 2nd, 3rd, or no) based on number of counts for each hit. 1st – multiple hits in multiple hits, 2nd (1st) – at least one single– and one multi–count hit, 2nd – one multi–count hit, 3rd – min. two single hits, and "no" – a single hit. Type – inter– and intra–domain hits are indicated. Relevance for modeling was based on clustering of hits. All "1st" and "2nd (1st)" hits have a strong relevancy, additionally all hits that cluster within a range of two residues or close to strong relevancy hit were selected to have a medium relevancy, and "weak" relevancy denotes single measurements, that show some clustering. All experiments were performed at a protein concentration of 1 mg/mL.

5.2 Supplementary Figures

BL16 - (3)M2-TE, 184 kDa
RSG34 - (3)M3-TE, 185 kDa
BL17 - (3)M5-TE, 185 kDa
BL18 - (3)M6-TE, 181 kDa

MK148 - SZ4-M2-TE, 187 kDa
MK149 - SZ4-M3-TE, 189 kDa
MK170 - SZ4-M5-TE, 188 kDa
MK147 - SZ4-M6-TE, 184 kDa
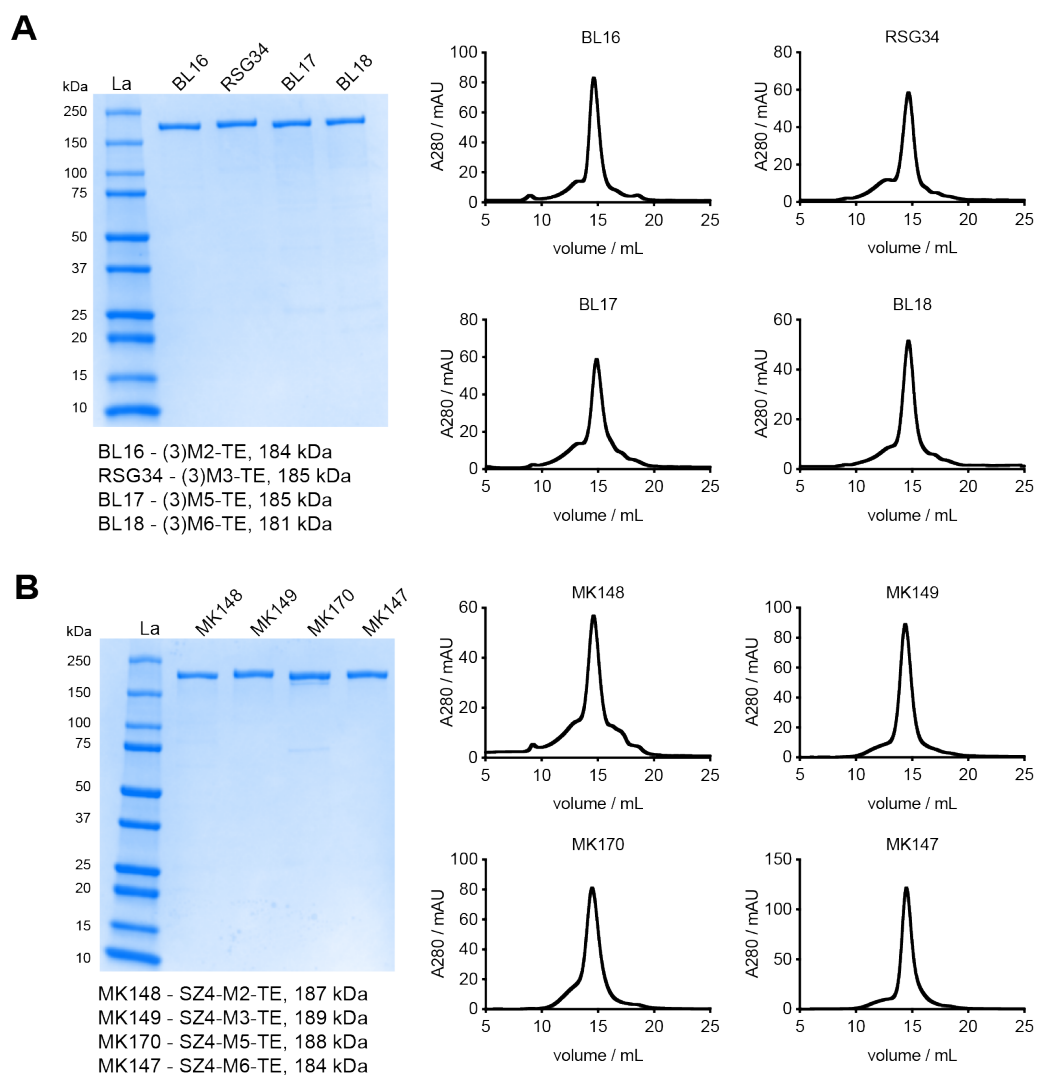
Figure S1 Analysis of protein purity by SDS–PAGE and SEC — single module acceptors. Acceptor proteins harboring a DEBS docking domain (A), and in comparison, a SZ domain (B). Protein abbreviations and their molecular weights (MW) are indicated. All proteins are pure and eluted in a predominately single peak from SEC.

MK136 - SZ4-DEBS(KR1-ACP1-M2)-TE, 249 kDa
MK152 - SZ4-DEBS(KR1-ACP1-M3)-TE, 250 kDa
MK141 - SZ4-DEBS(KR1-ACP1-M5)-TE, 249 kDa
MK142 - SZ4-DEBS(KR1-ACP1-M6)-TE, 245 kDa
MK137 - SZ4-DEBS(KR5-ACP5-M6)-TE, 246 kDa
MK138 - SZ4-RIFS(KR1-ACP1-M2)-TE, 213 kDa
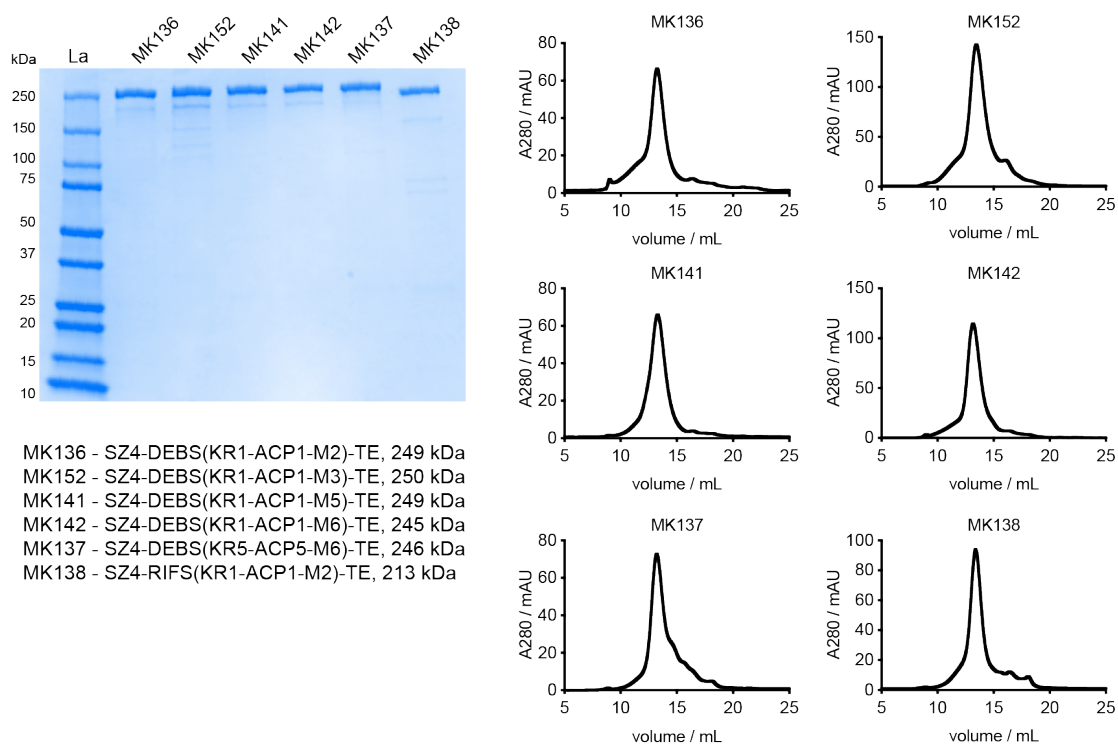
Figure S2 Analysis of protein purity by SDS–PAGE and SEC — covalent fusion acceptors. Acceptor proteins in which a KR–ACP fragment was fused to the acceptor module. Protein abbreviations and their MW are indicated. All proteins are pure and eluted in a predominately single peak from SEC.

BL13 - (5)M1(2), 168 kDa
MK150 - (5)M1-SZ3, 164 kDa
MK113 - (5)KS1-AT1, 97 kDa
MK111 - (5)KS1-AT1-SZ3, 103 kDa
MK96 - KR1-ACP1(2), 70 kDa
MK112 - SZ4-KR1-ACP1(2), 77 kDa
MK168 - SZ4-KR5-ACP5(2), 77 kDa
MK169 - SZ4-RIFS(KR1-ACP1)(2), 90 kDa
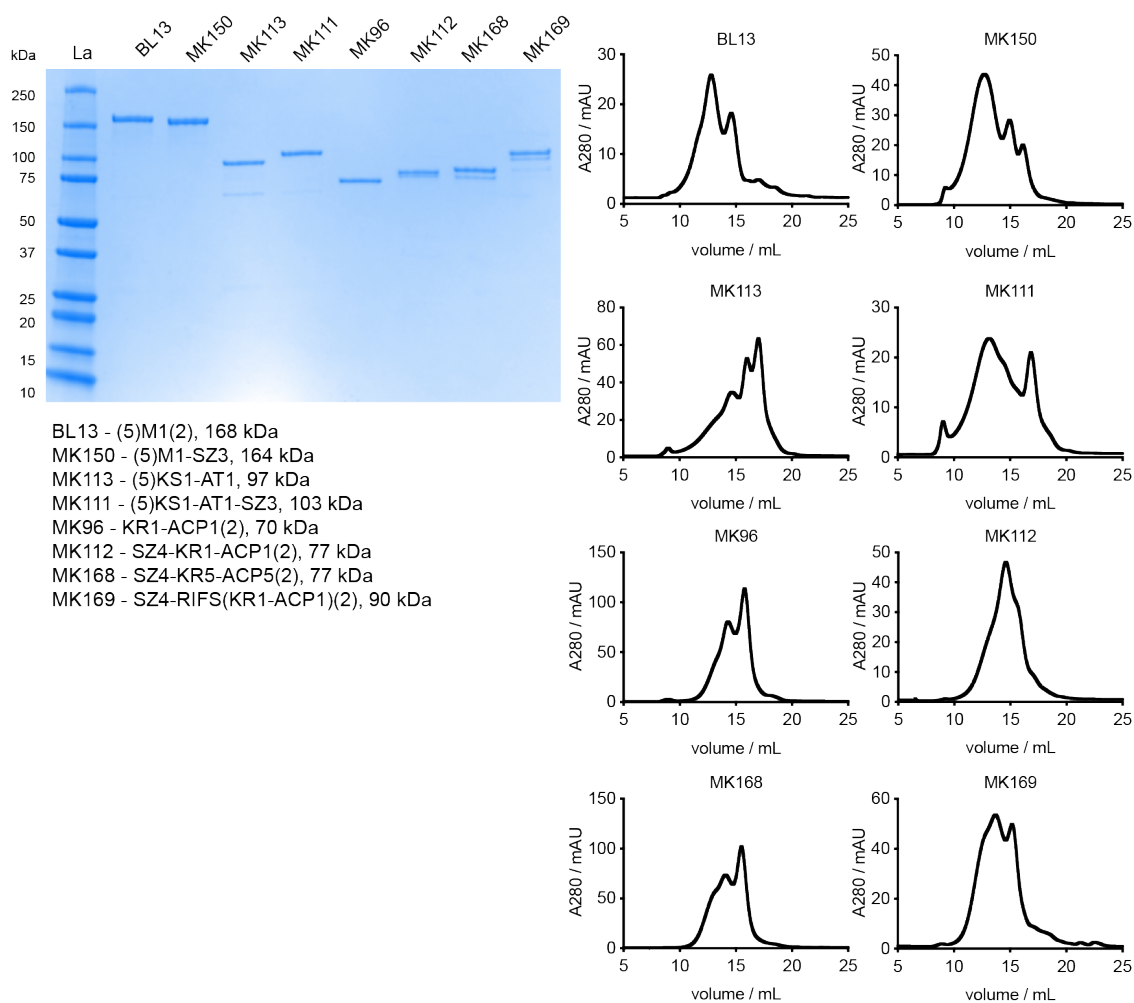
Figure S3 Analysis of protein purity by SDS–PAGE and SEC — DEBS M1 and its derivatives. DEBS M1 and its derivatives/fragments used in this study. If no indication is given the PKS domain/modules are derived from DEBS. Protein abbreviations and their MW are indicated. All proteins are pure, yet they appear as multiple oligomeric species based on SEC.
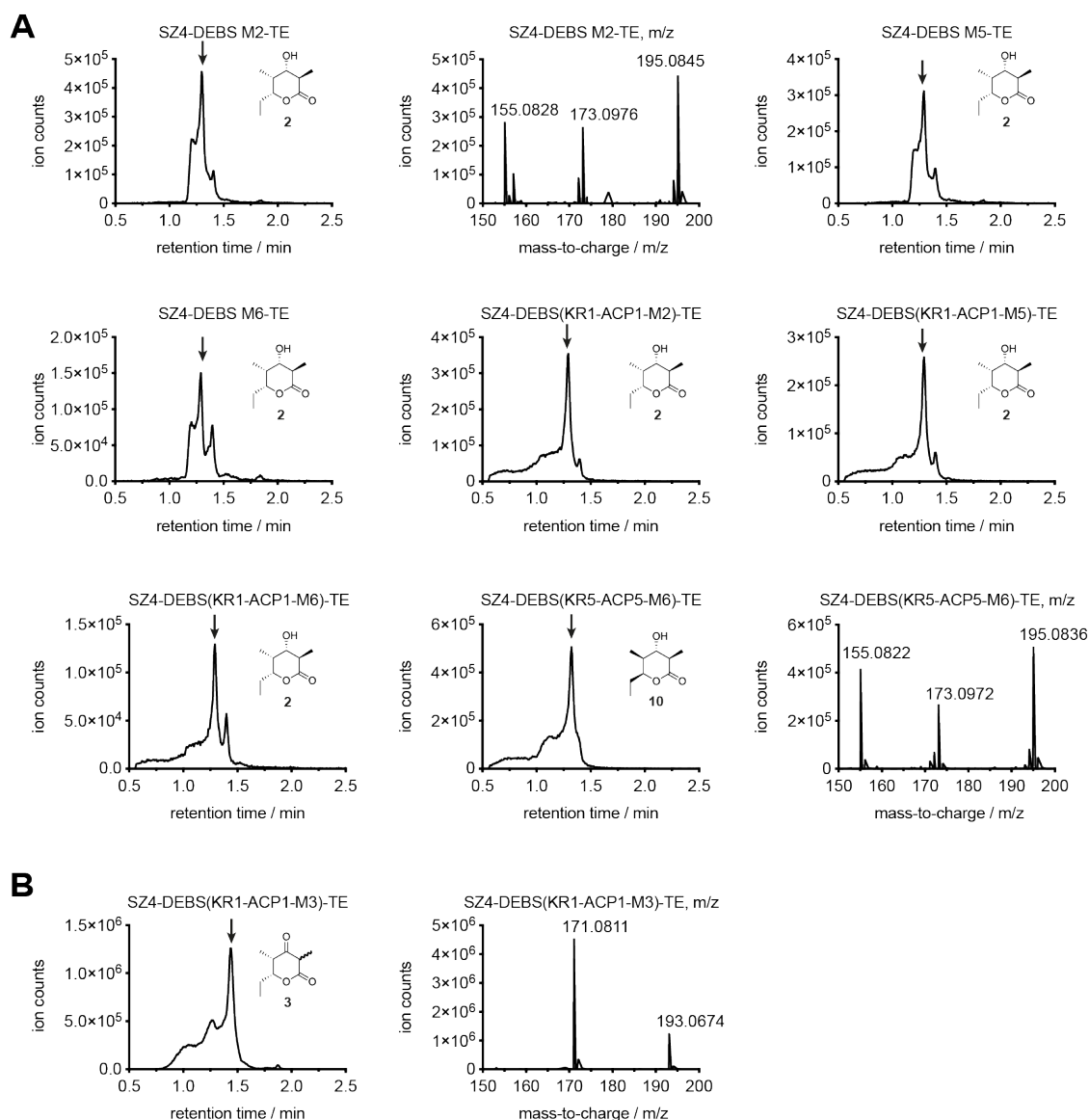
Figure S4   LC–MS analysis of products produced by bimodular chimeric PKSs newly generated in this study. (A) Reduced triketide lactone products **2** and **10** ($C_9H_{16}O_3$, calculated MW 172.110) were detected in reaction mixtures containing SZ4–M2–TE, SZ4–M5–TE, SZ4–M6–TE, SZ4–KR1–ACP1–M2–TE, SZ4–KR1–ACP1–M5–TE, SZ4–KR1–ACP1–M5–TE, and SZ4–KR5–ACP5–M6–TE. (B) Ketolactone **3** ($C_9H_{14}O_3$, calculated MW 170.090) was detected in reaction mixtures containing SZ4–KR1–ACP1–M3–TE. All PKS domain/modules are derived from DEBS. For all systems the extracted ion chromatograms were obtained by extraction of the $[M+Na]^+$ species and one chromatogram per compound is shown as an example. Labeled peaks from left to right correspond to $[M+H-H2O]^+$, $[M+H]^+$, and $[M+Na]^+$ ions. The peak of interest is marked with an arrow based on its mass spectrum.
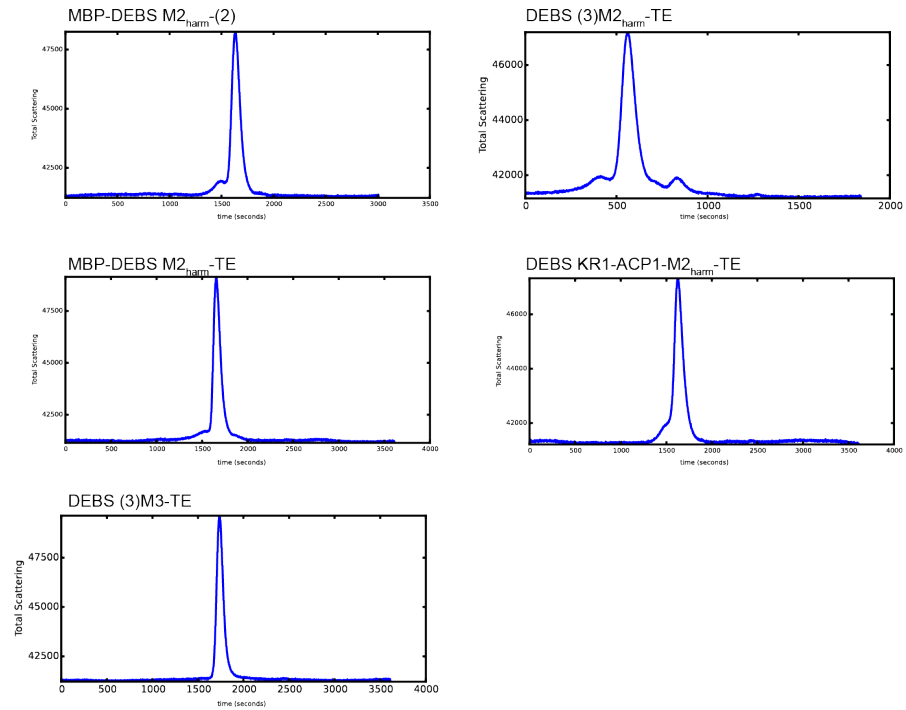
Figure S5 SEC–SAXS analysis of DEBS proteins — HPLC traces. HPLC light scattering intensity for all analyzed proteins is shown.
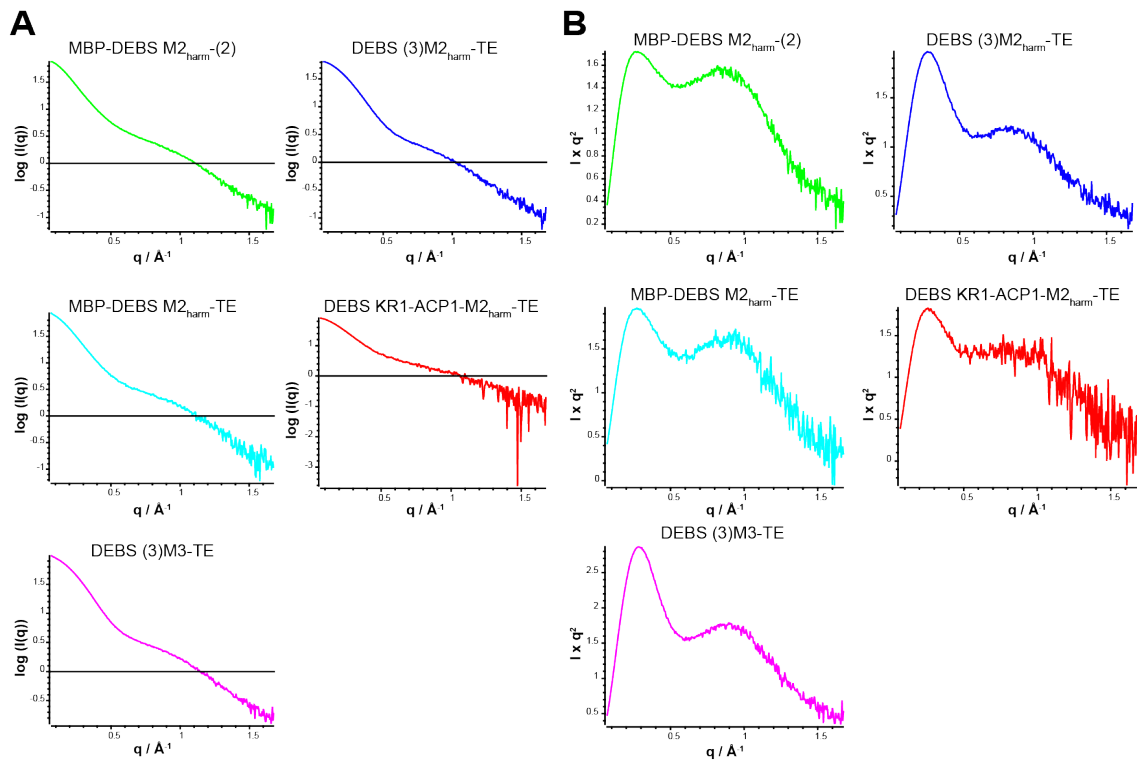
Figure S6 Raw data of tandem SEC–SAXS analysis of DEBS proteins. (A) Scattering intensity log(*I*) *vs q* and Kratky–plots (B) of all measured constructs. Plots were derived from the center of the scattering peak (for HPLC traces see Figure S5)

List of Figures

List of Tables

References

1. Dias, D. A., Urban, S., and Roessner, U. (2012): A historical overview of natural products in drug discovery. *Metabolites* 2(2), 303–36.

2. Newman, D. J. and Cragg, G. M. (2016): Natural Products as Sources of New Drugs from 1981 to 2014. *Journal of Natural Products* 79(3), 629–661.

3. Wright, G. D. (2019): Unlocking the potential of natural products in drug discovery. *Microbial Biotechnology* 12(1), 55–57.

4. Paterson, I. and Anderson, E. A. (2005): Products as Drug Candidates. *Science* 310, 451–453.

5. Harvey, A. L., Edrada-Ebel, R., and Quinn, R. J. (2015): The re-emergence of natural products for drug discovery in the genomics era. *Nature Reviews Drug Discovery* 14(2), 111–129.

6. Butler, M. S., Robertson, A. A., and Cooper, M. A. (2014): Natural product and natural product derived drugs in clinical trials. *Natural Product Reports* 31(11), 1612–1661.

7. Rodrigues, T., Reker, D., Schneider, P., and Schneider, G. (2016): Counting on natural products for drug design. *Nature Chemistry* 8(6), 531–541.

8. Breinbauer, R., Vetter, I. R., and Waldmann, H. (2002): From protein domains to drug candidates - Natural products as guiding principles in the design and synthesis of compound libraries. *Angewandte Chemie - International Edition* 41(16), 2878–2890.

9. Harvey, A. L. (2008): Natural products in drug discovery. *Drug Discovery Today* 13(19-20), 894–901.

10. Pahl, A., Waldmann, H., and Kumar, K. (2017): Exploring Natural Product Fragments for Drug and Probe Discovery. *CHIMIA International Journal for Chemistry* 71(10), 653–660.

11. Karageorgis, G. and Waldmann, H. (2019): Guided by Evolution: Biology-Oriented Synthesis of Bioactive Compound Classes. *Synthesis* 51(1), 55–66.

12. Spellberg, B., Guidos, R., Gilbert, D., Bradley, J., Boucher, H. W., Scheld, W. M., Bartlett, J. G., and Edwards, J. (2008): The epidemic of antibiotic-resistant infections: a call to action for the medical community from the Infectious Diseases Society of America. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 46(2), 155–164.

13. Zaman, S. B., Hussain, M. A., Nye, R., Mehta, V., Mamun, K. T., and Hossain, N. (2017): A Review on Antibiotic Resistance: Alarm Bells are Ringing. *Cureus* 9(6).

14. Hertweck, C. (2009): The biosynthetic logic of polyketide diversity. *Angewandte Chemie - International Edition* 48(26), 4688–4716.

15. Staunton, J. and Weissman, K. J. (2001): Polyketide biosynthesis: a millennium review. *Natural Product Reports* 18(4), 380–416.

16. Weissman, K. J. (2015a): The structural biology of biosynthetic megaenzymes. *Nature Chemical Biology* 11(9), 660–670.

17. Hopwood, D. A. (1997): Genetic Contributions to Understanding Polyketide Synthases. *Chemical Reviews* 97(7), 2465–2498.

18. Funa, N., Ohnishi, Y., Fujii, I., Shibuya, M., Ebizuka, Y., and Horinouchi, S. (1999): A new pathway for polyketide synthesis in microorganisms. *Nature* 400(6747), 897–899.

19. Piel, J. (2010): Biosynthesis of polyketides by trans-AT polyketide synthases. *Natural product reports* 27(7), 996–1047.

20. Cortes, J., Haydock, S. F., Roberts, G. A., Bevitt, D. J., and Leadlay, P. F. (1990): An unusually large multifunctional polypeptide in the erythromycin-producing polyketide synthase of Saccharopolyspora erythraea. *Nature* 348(6297), 176–178.

21. Donadio, S., Staver, M. J., McAlpine, J. B., Swanson, S. J., and Katz, L. (1991): Modular Organization of Genes Required for Complex Polyketide Biosynthesis. *Science* 252, 675–679.

22. Broadhurst, R., Nietlispach, D., Wheatcroft, M. P., Leadlay, P. F., and Weissman, K. J. (2003): The Structure of Docking Domains in Modular Polyketide Synthases. *Chemistry & Biology* 10(8), 723–731.

23. Dorival, J., Annaval, T., Risser, F., Collin, S., Roblin, P., Jacob, C., Gruez, A., Chagot, B., and Weissman, K. J. (2016): Characterization of Intersubunit Communication in the Virginiamycin trans -Acyl Transferase Polyketide Synthase. *Journal of the American Chemical Society* 138, 4155–4167.

24. Castonguay, R., He, W., Chen, A. Y., Khosla, C., and Cane, D. E. (2007): Stereospecificity of ketoreductase domains of the 6-deoxyerythronolide B synthase. *Journal of the American Chemical Society* 129(44), 13758–13769.

25. Valenzano, C. R., Lawson, R. J., Chen, A. Y., Khosla, C., and Cane, D. E. (2009): The biochemical basis for stereochemical control in polyketide biosynthesis. *Journal of the American Chemical Society* 131(51), 18501–18511.

26. Cane, D. E. (2010): Programming of Erythromycin Biosynthesis by a Modular Polyketide Synthase. *Journal of Biological Chemistry* 285(36), 27517–27523.

27. Zheng, J. and Keatinge-Clay, A. T. (2013): The status of type I polyketide synthase ketoreductases. *Med. Chem. Commun.* 4(1), 34–40.

28. Leadlay, P. F. (1997): Combinatorial approaches to polyketide biosynthesis. *Current Opinion in Chemical Biology* 1(2), 162–168.

29. Khosla, C. and Zawada, R. J. (1996): Generation of polyketide libraries via combinatorial biosynthesis. *Trends in Biotechnology* 14(9), 335–341.

30. McDaniel, R., Welch, M., and Hutchinson, C. R. (2005): Genetic approaches to polyketide antibiotics. *Chemical Reviews* 105(2), 543–558.

31. Bozhüyük, K. A. J., Fleischhacker, F., Linck, A., Wesche, F., Tietze, A., Niesert, C.-P., and Bode, H. B. (2017): De novo design and engineering of non-ribosomal peptide synthetases. *Nature Chemistry* 10(3), 275–281.

32. Tang, Y., Kim, C.-y., Mathews, I. I., Cane, D. E., and Khosla, C. (2006): The 2 .7-Å crystal structure of a 194-kDa homodimeric fragment of the 6-deoxyerythronolide B synthase. *Proceedings of the National Academy of Sciences* 103(30), 11124–11129.

33. Tang, Y., Chen, A. Y., Kim, C.-Y., Cane, D. E., and Khosla, C. (2007): Structural and mechanistic analysis of protein interactions in module 3 of the 6-deoxyerythronolide B synthase. *Chemistry & Biology* 14(8), 931–943.

34. Keatinge-Clay, A. T. and Stroud, R. M. (2006): The Structure of a Ketoreductase Determines the Organization of the *β*-Carbon Processing Enzymes of Modular Polyketide Synthases. *Structure* 14(4), 737–748.

35. Keatinge-Clay, A. (2008): Crystal Structure of the Erythromycin Polyketide Synthase Dehydratase. *Journal of Molecular Biology* 384(4), 941–953.

36. Alekseyev, V. Y., Liu, C. W., Cane, D. E., Puglisi, J. D., and Khosla, C. (2007): Solution structure and proposed domain domain recognition interface of an acyl carrier protein domain from a modular polyketide synthase. *Protein Science* 16(10), 2093–2107.

37. Chen, A. Y., Cane, D. E., and Khosla, C. (2007): Structure-Based Dissociation of a Type I Polyketide Synthase Module. *Chemistry and Biology* 14(7), 784–792.

38. Jenke-Kodama, H., Sandmann, A., Müller, R., and Dittmann, E. (2005): Evolutionary implications of bacterial polyketide synthases. *Molecular Biology and Evolution* 22(10), 2027–2039.

39. Smith, S. and Tsai, S.-C. (2007): The type I fatty acid and polyketide synthases: a tale of two megasynthases. *Natural product reports* 24(5), 1041–1072.

40. Weissman, K. J. and Müller, R. (2008): Protein-protein interactions in multienzyme megasynthetases. *ChemBioChem* 9(6), 826–848.

41. Herbst, D. A., Jakob, R. P., Zähringer, F., and Maier, T. (2016): Mycocerosic acid synthase exemplifies the architecture of reducing polyketide synthases. *Nature* 531(7595), 533–537.

42. Zheng, J., Gay, D. C., Demeler, B., White, M. A., and Keatinge-Clay, A. T. (2012): Divergence of multimodular polyketide synthases revealed by a didomain structure. *Nature Chemical Biology* 8(7), 615–621.

43. Zheng, J., Fage, C. D., Demeler, B., Hoffman, D. W., and Keatinge-Clay, A. T. (2013a): The missing linker: A dimerization motif located within polyketide synthase modules. *ACS Chemical Biology* 8(6), 1263–1270.

44. Khare, D., Hale, W. A., Tripathi, A., Gu, L., Sherman, D. H., Gerwick, W. H., Håkansson, K., and Smith, J. L. (2015): Structural Basis for Cyclopropanation by a Unique Enoyl-Acyl Carrier Protein Reductase. *Structure* 23(12), 2213–2223.

45. Maier, T., Leibundgut, M., and Ban, N. (2008): The crystal structure of a mammalian fatty acid synthase. *Science* 321(5894), 1315–1322.

46. Edwards, A. L., Matsui, T., Weiss, T. M., and Khosla, C. (2014): Architectures of whole-module and bimodular proteins from the 6-deoxyerythronolide B synthase. *Journal of Molecular Biology* 426(11), 2229–2245.

47. Dutta, S., Whicher, J. R., Hansen, D. a., Hale, W. a., Chemler, J. a., Congdon, G. R., Narayan, A. R. H., Håkansson, K., Sherman, D. H., Smith, J. L., and Skiniotis, G. (2014): Structure of a modular polyketide synthase. *Nature* 510(7506), 512–517.

48. Keatinge-Clay, A. T. (2016): The Structural Relationship between Iterative and Modular PKSs. *Cell Chemical Biology* 23(5), 540–542.

49. Pappenberger, G., Benz, J., Gsell, B., Hennig, M., Ruf, A., Stihle, M., Thoma, R., and Rudolph, M. G. (2010): Structure of the Human Fatty Acid Synthase KS-MAT

Didomain as a Framework for Inhibitor Design. *Journal of Molecular Biology* 397(2), 508–519.

50. Rittner, A., Paithankar, K. S., Huu, K. V., and Grininger, M. (2018): Characterization of the Polyspecific Transferase of Murine Type I Fatty Acid Synthase (FAS) and Implications for Polyketide Synthase (PKS) Engineering. *ACS Chemical Biology* 13(3), 723–732.

51. Li, X., Sevillano, N., La Greca, F., Deis, L., Liu, Y. C., Deller, M. C., Mathews, I. I., Matsui, T., Cane, D. E., Craik, C. S., and Khosla, C. (2018): Structure-Function Analysis of the Extended Conformation of a Polyketide Synthase Module. *Journal of the American Chemical Society* 140(21), 6518–6521.

52. Whicher, J. R., Dutta, S., Hansen, D. a., Hale, W. a., Chemler, J. a., Dosey, A. M., Narayan, A. R. H., Håkansson, K., Sherman, D. H., Smith, J. L., and Skiniotis, G. (2014): Structural rearrangements of a polyketide synthase module during its catalytic cycle. *Nature* 510(7506), 560–564.

53. Rittner, A. and Grininger, M. (2014): Modular Polyketide Synthases (PKSs): A New Model Fits All? *ChemBioChem* 15(17), 2489–2493.

54. Weissman, K. J. (2015b): Uncovering the structures of modular polyketide synthases. *Natural Product Reports* 32, 436–453.

55. Richter, C. D., Nietlispach, D., Broadhurst, R. W., and Weissman, K. J. (2008): Multienzyme docking in hybrid megasynthetases. *Nature Chemical Biology* 4(1), 75–81.

56. Buchholz, T. J., Geders, T. W., Bartley III, F. E., Reynolds, K. a., Smith, J. L., and Sherman, D. H. (2009): Structural Basis for Binding Specificitybetween Subclasses of Modular Polyketide Synthase Docking Domains. *ACS Chemical Biology* 4(1), 41–52.

57. Whicher, J. R., Smaga, S. S., Hansen, D. a., Brown, W. C., Gerwick, W. H., Sherman, D. H., and Smith, J. L. (2013): Cyanobacterial polyketide synthase docking domains: A tool for engineering natural product biosynthesis. *Chemistry and Biology* 20(11), 1340–1351.

58. Brignole, E. J., Smith, S., and Asturias, F. J. (2009): Conformational flexibility of metazoan fatty acid synthase enables catalysis. *Nature Structural and Molecular Biology* 16(2), 190–197.

59. Joshi, A. K., Rangan, V. S., Witkowski, A., and Smith, S. (2003): Engineering of an Active Animal Fatty Acid Synthase Dimer with Only One Competent Subunit. *Chemistry & Biology* 10(2), 169–173.

60. Benning, F. M., Sakiyama, Y., Mazur, A., Bukhari, H. S., Lim, R. Y., and Maier, T. (2017): High-Speed Atomic Force Microscopy Visualization of the Dynamics of the Multienzyme Fatty Acid Synthase. *ACS Nano* 11(11), 10852–10859.

61. Kapur, S., Chen, A. Y., Cane, D. E., and Khosla, C. (2010): Molecular recognition between ketosynthase and acyl carrier protein domains of the 6-deoxyerythronolide B synthase. *Proceedings of the National Academy of Sciences* 107(51), 22066–22071.

62. Kao, C. M., Pieper, R., Cane, D. E., and Khosla, C. (1996): Evidence for two catalytically independent clusters of active sites in a functional modular polyketide synthase. *Biochemistry* 35(38), 12363–12368.

63. Kapur, S., Lowry, B., Yuzawa, S., Kenthirapalan, S., Chen, A. Y., Cane, D. E., and Khosla, C. (2012): Reprogramming a module of the 6-deoxyerythronolide B synthase for iterative chain elongation. *Proceedings of the National Academy of Sciences* 109(11), 4110–4115.

64. Donadio, S., McAlpine, J. B., Sheldon, P. J., Jackson, M., and Katz, L. (1993): An erythromycin analog produced by reprogramming of polyketide synthesis. *Proceedings of the National Academy of Sciences USA* 90(15), 7119–7123.

65. Katz, L. and Donadio, S. (1993): Polyketide Synthesis: Prospects for Hybrid Antibitics. *Annual Review of Microbiology* 47, 875–912.

66. Gokhale, R. S., Tsuji, S. Y., Cane, D. E., and Khosla, C. (1999a): Dissecting and exploiting intermodular communication in polyketide synthases. *Science* 284(5413), 482–485.

67. Ranganathan, A., Timoney, M., Bycroft, M., Cortés, J., Thomas, I. P., Wilkinson, B., Kellenberger, L., Hanefeld, U., Galloway, I. S., Staunton, J., and Leadlay, P. F. (1999): Knowledge-based design of bimodular and trimodular polyketide synthases based on domain and module swaps: A route to simple statin analogues. *Chemistry and Biology* 6(10), 731–741.

68. Tang, L., Fu, H., and McDaniel, R. (2000): Formation of functional heterologous complexes using subunits from the picromycin, erythromycin and oleandomycin polyketide synthases. *Chemistry & Biology* 7(2), 77–84.

69. Rowe, C. J., Böhm, I. U., Thomas, I. P., Wilkinson, B., Rudd, B. a. M., Foster, G., Blackaby, A. P., Sidebottom, P. J., Roddis, Y., Buss, A. D., Staunton, J., and Leadlay, P. F. (2001): Engineering a polyketide with a longer chain by insertion of an extra module into the erythromycin-producing polyketide synthase. *Chemistry and Biology* 8(5), 475–485.

70. Menzella, H. G., Reid, R., Carney, J. R., Chandran, S. S., Reisinger, S. J., Patel, K. G., Hopwood, D. A., and Santi, D. V. (2005): Combinatorial polyketide biosynthesis by de novo design and rearrangement of modular polyketide synthase genes. *Nature Biotechnology* 23(9), 1171–1176.

71. Menzella, H. G. and Reeves, C. D. (2007): Combinatorial biosynthesis for drug development. *Current Opinion in Microbiology* 10(3), 238–245.

72. Wu, N., Tsuji, S. Y., Cane, D. E., and Khosla, C. (2001): Assessing the balance between protein-protein interactions and enzyme-substrate interactions in the channeling of intermediates between polyketide synthase modules. *Journal of the American Chemical Society* 123(27), 6465–6474.

73. Wu, N., Cane, D. E., and Khosla, C. (2002): Quantitative Analysis of the relative contributions of donor acyl carrier proteins, acceptor ketosynthases, and linker regions to intermodular transfer of intermediates in hybrid polyketide synthases. *Biochemistry* 41(15), 5056–5066.

74. Tsuji, S. Y., Cane, D. E., and Khosla, C. (2001): Selective Protein-Protein Interactions Direct Channeling of Intermediates between Polyketide Synthase Modules. *Biochemistry* 40(8), 2326–2331.

75. Lowry, B., Robbins, T., Weng, C. H., O'Brien, R. V., Cane, D. E., and Khosla, C. (2013): In vitro reconstitution and analysis of the 6-deoxyerythronolide B synthase. *Journal of the American Chemical Society* 135(45), 16809–16812.

76. Marsden, A. F., Wilkinson, B., Cortés, J., Dunster, N. J., Staunton, J., and Leadlay, P. F. (1998): Engineering broader specificity into an antibiotic-producing polyketide synthase. *Science* 279(5348), 199–202.

77. Kuhstoss, S., Huber, M., Turner, J. R., Paschal, J. W., and Rao, R. N. (1996): Production of a novel polyketide through the construction of a hybrid polyketide synthase. *Gene* 183(1-2), 231–236.

78. Long, P. F., Wilkinson, C. J., Bisang, C. P., Cortés, J., Dunster, N., Oliynyk, M., McCormick, E., McArthur, H., Mendez, C., Salas, J. a., Staunton, J., and Leadlay, P. F. (2002): Engineering specificity of starter unit selection by the erythromycin-producing polyketide synthase. *Molecular Microbiology* 43, 1215–1225.

79. Dunn, B. J., Cane, D. E., and Khosla, C. (2013): Mechanism and Specificity of an Acyltransferase Domain from a Modular Polyketide Synthase. *Biochemistry* 52, 1839–1841.

80. Oliynyk, M., Brown, M. J., Cortés, J., Staunton, J., and Leadlay, P. F. (1996): A hybrid modular polyketide synthase obtained by domain swapping. *Chemistry & Biology* 3(10), 833–839.

81. Ruan, X., Pereda, A., Stassi, D. L., Zeidner, D., Summers, R. G., Jackson, M., Shivakumar, A., Kakavas, S., Staver, M. J., Donadio, S., and Katz, L. (1997): Acyltransferase domain substitutions in erythromycin polyketide synthase yield novel erythromycin derivatives. *Journal of Bacteriology* 179(20), 6416–6425.

82. Stassi, D. L., Kakavas, S. J., Reynolds, K. a., Gunawardana, G., Swanson, S., Zeidner, D., Jackson, M., Liu, H., Buko, a., and Katz, L. (1998): Ethyl-substituted erythromycin derivatives produced by directed metabolic engineering. *Proceedings of the National Academy of Sciences USA* 95(13), 7305–7309.

83. Lau, J., Fu, H., Cane, D. E., and Khosla, C. (1999): Dissecting the role of acyltransferase domains of modular polyketide synthases in the choice and stereochemical fate of extender units. *Biochemistry* 38(5), 1643–1651.

84. McDaniel, R., Thamchaipenet, a., Gustafsson, C., Fu, H., Betlach, M., and Ashley, G. (1999): Multiple genetic modifications of the erythromycin polyketide synthase to produce a library of novel "unnatural" natural products. *Proceedings of the National Academy of Sciences USA* 96(5), 1846–1851.

85. Kim, C. Y., Alekseyev, V. Y., Chen, A. Y., Tang, Y., Cane, D. E., and Khosla, C. (2004): Reconstituting modular activity from separated domains of 6-deoxyerythronolide B synthase. *Biochemistry* 43(44), 13892–13898.

86. Yuzawa, S., Kapur, S., Cane, D. E., and Khosla, C. (2012): Role of a Conserved Arginine Residue in Linkers between the Ketosynthase and Acyltransferase Domains of Multimodular Polyketide Synthases. *Biochemistry* 51(18), 3708–3710.

87. Yuzawa, S., Deng, K., Wang, G., Baidoo, E. E. K., Northen, T. R., Adams, P. D., Katz, L., and Keasling, J. D. (2017): SI_Comprehensive in Vitro Analysis of Acyltransferase Domain Exchanges in Modular Polyketide Synthases and Its Application for Short-Chain Ketone Production. *ACS synthetic biology* 6(1), 139–147.

88. Crooks, G., Hon, G., Chandonia, J., and Brenner, S. (2004): WebLogo: A Sequence Logo Generator. *Genome Research* 14, 1188–1190.

89. Miyanaga, A., Iwasawa, S., Shinohara, Y., Kudo, F., and Eguchi, T. (2016): Structure-based analysis of the molecular interactions between acyltransferase and acyl carrier protein in vicenistatin biosynthesis. *Proceedings of the National Academy of Sciences* 113(7), 1802–1807.

90. Wong, F. T., Chen, A. Y., Cane, D. E., and Khosla, C. (2010): Protein-protein recognition between acyltransferases and acyl carrier proteins in multimodular polyketide synthases. *Biochemistry* 49(1), 95–102.

91. Wong, F. T., Jin, X., Mathews, I. I., Cane, D. E., and Khosla, C. (2011): Structure and mechanism of the trans -Acting acyltransferase from the disorazole synthase. *Biochemistry* 50(30), 6539–6548.

92. Liew, C. W., Nilsson, M., Chen, M. W., Sun, H., Cornvik, T., Liang, Z. X., and Lescar, J. (2012): Crystal structure of the acyltransferase domain of the iterative polyketide synthase in enediyne biosynthesis. *Journal of Biological Chemistry* 287(27), 23203–23215.

93. Ye, Z., Musiol, E. M., Weber, T., and Williams, G. J. (2014): Reprogramming acyl carrier protein interactions of an acyl-CoA promiscuous trans-acyltransferase. *Chemistry and Biology* 21(5), 636–646.

94. Xie, X., Garg, A., Keatinge-Clay, A. T., Khosla, C., and Cane, D. E. (2016): Epimerase and Reductase Activities of Polyketide Synthase Ketoreductase Domains Utilize the Same Conserved Tyrosine and Serine Residues. *Biochemistry* 55(8), 1179–1186.

95. Ostrowski, M. P., Cane, D. E., and Khosla, C. (2016): Recognition of acyl carrier proteins by ketoreductases in assembly line polyketide synthases. *The Journal of Antibiotics* 69(7), 507–510.

96. Bedford, D., Jacobsen, J. R., Luo, G., Cane, D. E., and Khosla, C. (1996): A functional chimeric modular polyketide synthase generated via domain replacement. *Chemistry & Biology* 3(10), 827–831.

97. Kellenberger, L., Galloway, I. S., Sauter, G., Böhm, G., Hanefeld, U., Cortés, J., Staunton, J., and Leadlay, P. F. (2008): A polylinker approach to reductive loop swaps in modular polyketide synthases. *ChemBioChem* 9(16), 2740–2749.

98. Annaval, T., Paris, C., Leadlay, P. F., Jacob, C., and Weissman, K. J. (2015): Evaluating ketoreductase exchanges as a means of rationally altering polyketide stereochemistry. *ChemBioChem* 16(9), 1357–1364.

99. Zheng, J., Piasecki, S. K., and Keatinge-Clay, A. T. (2013b): Structural studies of an A2-type modular polyketide synthase ketoreductase reveal features controlling α-substituent stereochemistry. *ACS Chemical Biology* 8(9), 1964–1971.

100. Eng, C. H., Yuzawa, S., Wang, G., Baidoo, E. E. K., Katz, L., and Keasling, J. D. (2016): Alteration of Polyketide Stereochemistry from anti to syn by a Ketoreductase Domain Exchange in a Type i Modular Polyketide Synthase Subunit. *Biochemistry* 55(12), 1677–1680.

101. Hagen, A., Poust, S., De Rond, T., Fortman, J. L., Katz, L., Petzold, C. J., and Keasling, J. D. (2016): Engineering a Polyketide Synthase for in Vitro Production of Adipic Acid. *ACS Synthetic Biology* 5(1), 21–27.

102. Chemler, J. A., Tripathi, A., Hansen, D. A., Neil-johnson, M. O., Williams, R. B., Starks, C., Park, S. R., and Sherman, D. H. (2015): Evolution of Efficient Modular Polyketide Synthases by Homologous Recombination. *Journal of the American Chemical Society* 137, 10603–10609.

103. Wlodek, A., Kendrew, S. G., Coates, N. J., Hold, A., Pogwizd, J., Rudder, S., Sheehan, L. S., Higginbotham, S. J., Stanley-Smith, A. E., Warneck, T., Nur-E-Alam, M., Radzom, M., Martin, C. J., Overvoorde, L., Samborskyy, M., Alt, S., Heine, D., Carter, G. T., Graziani, E. I., Koehn, F. E., McDonald, L., Alanine, A., Rodríguez Sarmiento, R. M., Chao, S. K., Ratni, H., Steward, L., Norville, I. H., Sarkar-Tyson, M., Moss, S. J., Leadlay, P. F., Wilkinson, B., and Gregory, M. A. (2017): Diversity oriented biosynthesis via accelerated evolution of modular gene clusters. *Nature Communications* 8(1), 1206.

104. Klaus, M., Ostrowski, M. P., Austerjost, J., Robbins, T., Lowry, B., Cane, D. E., and Khosla, C. (2016): Protein Protein Interactions, not Substrate Recognition, Dominates the Turnover of Chimeric Assembly Line Polyketide Synthases. *Journal of Biological Chemistry* 291(31), 16404–16415.

105. Klaus, M., D'Souza, A. D., Nivina, A., Khosla, C., and Grininger, M. (2019): Engineering of Chimeric Polyketide Synthases Using SYNZIP Docking Domains. *ACS Chemical Biology* 14(3), 426–433.

106. Klaus, M. and Grininger, M. (2018): Engineering strategies for rational polyketide synthase design. *Natural Product Reports* 35(10), 1070–1081.

107. Rossini, E., Gajewski, J., Klaus, M., Hummer, G., and Grininger, M. (2018): Analysis and engineering of substrate shuttling by the acyl carrier protein (ACP) in fatty acid synthases (FASs). *Chemical Communications* 2622, 11606–11609.

108. Klaus, M. (2015): The production and in vitro characterization of non–natural bimodular and trimodular chimeric polyketide synthases. *Master's thesis*, 1–72.

109. Yu, T. W., Shen, Y., Doi-Katayama, Y., Tang, L., Park, C., Moore, B. S., Richard Hutchinson, C., and Floss, H. G. (1999): Direct evidence that the rifamycin polyketide synthase assembles polyketide chains processively. *Proceedings of the National Academy of Sciences* 96(16), 9051–9056.

110. Schwecke, T., Aparicio, J. F., Molnár, I., König, A., Khaw, L. E., Haydock, S. F., Oliynyk, M., Caffrey, P., Cortés, J., Lester, J. B., Böhm, G. A., Staunton, J., and Leadlay, P. F. (1995): The biosynthetic gene cluster for the polyketide immunosuppressant rapamycin. *Proceedings of the National Academy of Sciences* 92(17), 7839–7843.

111. Galán, A., Comor, L., Horvatić, A., Kuleš, J., Guillemin, N., Mrljak, V., and Bhide, M. (2016): Library-based display technologies: Where do we stand? *Molecular BioSystems* 12(8), 2342–2358.

112. Smith, G. (1985): Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science* 228(4705), 1315–1317.

113. Vodnik, M., Zager, U., Strukelj, B., and Lunder, M. (2011): Phage display: Selecting straws instead of a needle from a haystack. *Molecules* 16(1), 790–817.

114. Sidhu, S. S., Fairbrother, W. J., and Deshayes, K. (2003): Exploring protein-protein interactions with phage display. *ChemBioChem* 4(1), 14–25.

115. Li, W. and Caberoy, N. B. (2010): New perspective for phage display as an efficient and versatile technology of functional proteomics. *Applied microbiology and biotechnology* 85(4), 909–9019.

116. Ledsgaard, L., Kilstrup, M., Karatt-Vellatt, A., McCafferty, J., and Laustsen, A. (2018): Basics of Antibody Phage Display Technology. *Toxins* 10(6), 236.

117. Russel, M., Lowman, H., and Clackson, T. (2004): Introduction to phage biology and phage display. *Phage Display: A Practical Approach*, 1–23.

118. Bazan, J., Całkosiñski, I., and Gamian, A. (2012): Phage displaya powerful technique for immunotherapy: 1. Introduction and potential of therapeutic applications. *Human Vaccines and Immunotherapeutics* 8(12), 1817–1828.

119. Hoess, R. H. (2001): Protein design and phage display. *Chemical reviews* 101(10), 3205–3218.

120. Chasteen, L., Ayriss, J., Pavlik, P., and Bradbury, A. R. M. (2006): Eliminating helper phage from phage display. *Nucleic Acids Research* 34(21), 1–11.

121. Clackson, T., Hoogenboom, H. R., Griffiths, A. D., and Winter, G. (1991): Making antibody fragments using phage display libraries. *Nature* 352(6336), 624–628.

122. Scott, J. K. and Smith, G. P. (1990): Searching for peptide ligands with an epitope library. *Science* 249(4967), 386–390.

123. Swimmer, C., Lehar, S. M., McCafferty, J., Chiswell, D. J., Blättler, W. a., and Guild, B. C. (1992): Phage display of ricin B chain and its single binding domains: system for screening galactose-binding mutants. *Proceedings of the National Academy of Sciences USA* 89(9), 3756–3760.

124. Lowman, H. B. and Wells, J. A. (1993): Affinity maturation of human growth hormone by monovalent phage display. *Journal of Molecular Biology* 234(3), 564–578.

125. Lowry, B., Li, X., Robbins, T., Cane, D. E., and Khosla, C. (2016): A Turnstile Mechanism for the Controlled Growth of Biosynthetic Intermediates on Assembly Line Polyketide Synthases. *ACS Central Science* 2(1), 14–20.

126. Kunkel, T. A. (1985): Rapid and efficient site-specific mutagenesis without phenotypic selection. *Proceedings of the National Academy of Sciences of the USA* 82(2), 488–492.

127. López-Llano, J., Campos, L. A., and Sancho, J. (2006): Alpha-helix stabilization by alanine relative to glycine: roles of polar and apolar solvent exposures and of backbone entropy. *Proteins* 64(3), 769–778.

128. Miton, C. M. and Tokuriki, N. (2016): How mutational epistasis impairs predictability in protein evolution and design. *Protein Science*, 1260–1272.

129. Starr, T. N. and Thornton, J. W. (2016): Epistasis in protein evolution. *Protein Science*, 1204–1218.

130. Kumar, P., Li, Q., Cane, D. E., and Khosla, C. (2003): Intermodular communication in modular polyketide synthases: Structural and mutational analysis of linker mediated protein - protein recognition. *Journal of the American Chemical Society* 125(14), 4097–4102.

131. Janßen, M. (2017): Analyse von Domänen-Domänen-Interaktionen in modulare Polyketidsynthasen. *Master's thesis*, 1–63.

132. Chen, A. Y., Schnarr, N. a., Kim, C. Y., Cane, D. E., and Khosla, C. (2006): Extender unit and acyl carrier protein specificity of ketosynthase domains of the 6-deoxyerythronolide B synthase. *Journal of the American Chemical Society* 128(9), 3067–3074.

133. Reinke, A. W., Grant, R. A., and Keating, A. E. (2010): A synthetic coiled-coil interactome provides heterospecific modules for molecular engineering. *Journal of the American Chemical Society* 132(17), 6025–6031.

134. Thompson, K. E., Bashor, C. J., Lim, W. A., and Keating, A. E. (2012): SYNZIP Protein Interaction Toolbox: *in Vitro* and *in Vivo* Specifications of Heterospecific Coiled-Coil Interaction Domains. *ACS Synthetic Biology* 1(4), 118–129.

135. Zhang, L., Hashimoto, T., Qin, B., Hashimoto, J., Kozone, I., Kawahara, T., Okada, M., Awakawa, T., Ito, T., Asakawa, Y., Ueki, M., Takahashi, S., Osada, H., Wakimoto, T., Ikeda, H., Shin-ya, K., and Abe, I. (2017): Characterization of Giant Modular PKSs Provides Insight into Genetic Mechanism for Structural Diversification of Aminopolyol Polyketides. *Angewandte Chemie - International Edition* 56(7), 1740–1745.

136. Vander Wood, D. A. and Keatinge-Clay, A. T. (2018): The modules of trans-acyltransferase assembly lines redefined with a central acyl carrier protein. *Proteins* 86(6), 664–675.

137. Anderson, G. P., Shriver-Lake, L. C., Liu, J. L., and Goldman, E. R. (2018): Orthogonal Synthetic Zippers as Protein Scaffolds. *ACS Omega* 3(5), 4810–4815.

138. Murphy, A. C., Hong, H., Vance, S., Broadhurst, R. W., and Leadlay, P. F.: Broadening substrate specificity of a chain-extending ketosynthase through a single active-site mutation. *Chemical Communications*(54), 8373–8376.

139. Keatinge-Clay, A. T. (2017): Polyketide Synthase Modules Redefined. *Angewandte Chemie - International Edition* 56(17), 4658–4660.

140. Wu, N., Kudo, F., Cane, D. E., and Khosla, C. (2000): Analysis of the molecular recognition features of individual modules derived from the erythromycin polyketide synthase. *Journal of the American Chemical Society* 122(20), 4847–4852.

141. Gokhale, R. S., Hunziker, D., Cane, D. E., and Khosla, C. (1999b): Mechanism and specificity of the terminal thioesterase domain from the erythromycin polyketide synthase. *Chemistry and Biology* 6, 117–125.

142. Siskos, A. P., Baerga-Ortiz, A., Bali, S., Stein, V., Mamdani, H., Spiteller, D., Popovic, B., Spencer, J. B., Staunton, J., Weissman, K. J., and Leadlay, P. F. (2005): Molecular Basis of Celmer's Rules: Stereochemistry of Catalysis by Isolated Ketoreductase Domains from Modular Polyketide Synthases. *Chemistry & Biology* 12(10), 1145–1153.

143. Wu, J., Kinoshita, K., Khosla, C., and Cane, D. E. (2004): Biochemical analysis of the substrate specificity of the beta-ketoacyl-acyl carrier protein synthase domain of module 2 of the erythromycin polyketide synthase. *Biochemistry* 43(51), 16301–16310.

144. Khosla, C., Tang, Y., Chen, A. Y., Schnarr, N. A., and Cane, D. E. (2007): Structure and Mechanism of the 6-deoxyerythronolide B Synthase. *Annual review of biochemistry* 76, 195–221.

145. Robbins, T., Kapilivsky, J., Cane, D. E., and Khosla, C. (2016): Roles of Conserved Active Site Residues in the Ketosynthase Domain of an Assembly Line Polyketide Synthase. *Biochemistry* 55(32), 4476–4484.

146. Goldsmith, M. and Tawfik, D. S. (2017): Enzyme engineering: reaching the maximal catalytic efficiency peak. *Current Opinion in Structural Biology* 47, 140–150.

147. Khersonsky, O., Lipsh, R., Avizemer, Z., Ashani, Y., Goldsmith, M., Leader, H., Dym, O., Rogotner, S., Trudeau, D. L., Prilusky, J., Amengual-Rigo, P., Guallar, V., Tawfik, D. S., and Fleishman, S. J. (2018): Automated Design of Efficient and Functionally Diverse Enzyme Repertoires. *Molecular Cell* 72(1), 178–186.

148. Wlodawer, A., Li, M., and Dauter, Z. (2017): High-Resolution Cryo-EM Maps and Models: A Crystallographer's Perspective. *Structure* 25(10), 1589–1597.

149. Skou, S., Gillilan, R. E., and Ando, N. (2014): Synchrotron-based small-angle X-ray scattering of proteins in solution. *Nature Protocols* 9(7), 1727–1739.

150. Busche, A., Gottstein, D., Hein, C., Ripin, N., Pader, I., Tufar, P., Eisman, E. B., Gu, L., Walsh, C. T., Sherman, D. H., Löhr, F., Güntert, P., and Dötsch, V. (2012): Characterization of Molecular Interactions between ACP and Halogenase Domains in the Curacin A Polyketide Synthase. *ACS Chemical Biology* 7(2), 378–386.

151. Sugiki, T., Kobayashi, N., and Fujiwara, T. (2017): Modern Technologies of Solution Nuclear Magnetic Resonance Spectroscopy for Three-dimensional Structure Determination of Proteins Open Avenues for Life Scientists. *Computational and Structural Biotechnology Journal* 15, 328–339.

152. Back, J. W., De Jong, L., Muijsers, A. O., and De Koster, C. G. (2003): Chemical cross-linking and mass spectrometry for protein structural modeling. *Journal of Molecular Biology* 331(2), 303–313.

153. Dehling, E., Volkmann, G., Matern, J. C. J., Dörner, W., Alfermann, J., Diecker, J., and Mootz, H. D. (2016): Mapping of the Communication-Mediating Interface in Nonribosomal Peptide Synthetases Using a Genetically Encoded Photocrosslinker Supports an Upside-Down Helix-Hand Motif. *Journal of Molecular Biology* 428(21), 4345–4360.

154. Holding, A. N. (2015): XL-MS: Protein cross-linking coupled with mass spectrometry. *Methods* 89, 54–63.

155. Leitner, A., Walzthoeni, T., Kahraman, A., Herzog, F., Rinner, O., Beck, M., and Aebersold, R. (2010): Probing native protein structures by chemical cross-linking, mass spectrometry, and bioinformatics. *Molecular & cellular proteomics* 9(8), 1634–1649.

156. Belsom, A., Mudd, G., Giese, S., Auer, M., and Rappsilber, J. (2017): Complementary Benzophenone Cross-Linking/Mass Spectrometry Photochemistry. *Analytical Chemistry* 89(10), 5319–5324.

157. Yu, C. and Huang, L. (2018): Cross-Linking Mass Spectrometry: An Emerging Technology for Interactomics and Structural Biology. *Analytical Chemistry* 90(1), 144–165.

158. Rappsilber, J., Siniossoglou, S., Hurt, E. C., and Mann, M. (2000): A generic strategy to analyze the spatial organization of multi-protein complexes by cross-linking and mass spectrometry. *Analytical Chemistry* 72(2), 267–275.

159. Sinz, A. (2006): Chemical cross-linking and mass spectrometry to map three-dimensional protein structures and protein-protein interactions. *Mass Spectrometry Reviews* 25(4), 663–682.

160. Maiolica, A., Cittaro, D., Borsotti, D., Sennels, L., Ciferri, C., Tarricone, C., Musacchio, A., and Rappsilber, J. (2007): Structural Analysis of Multiprotein Complexes by Cross-linking, Mass Spectrometry, and Database Searching. *Molecular & Cellular Proteomics* 6(12), 2200–2211.

161. Chen, Z. A., Jawhari, A., Fischer, L., Buchen, C., Tahir, S., Kamenski, T., Rasmussen, M., Lariviere, L., Bukowski-Wills, J. C., Nilges, M., Cramer, P., and Rappsilber, J. (2010): Architecture of the RNA polymerase II-TFIIF complex revealed by cross-linking and mass spectrometry. *EMBO Journal* 29(4), 717–726.

162. Lasker, K., Forster, F., Bohn, S., Walzthoeni, T., Villa, E., Unverdorben, P., Beck, F., Aebersold, R., Sali, A., and Baumeister, W. (2012): Molecular architecture of the 26S proteasome holocomplex determined by an integrative approach. *Proceedings of the National Academy of Sciences* 109(5), 1380–1387.

163. Wang, X., Cimermancic, P., Yu, C., Schweitzer, A., Chopra, N., Engel, J. L., Greenberg, C., Huszagh, A. S., Beck, F., Sakata, E., Yang, Y., Novitsky, E. J., Leitner, A., Nanni, P., Kahraman, A., Guo, X., Dixon, J. E., Rychnovsky, S. D., Aebersold, R., Baumeister, W., Sali, A., and Huang, L. (2017): Molecular Details Underlying Dynamic Structures and Regulation of the Human 26S Proteasome. *Molecular & Cellular Proteomics* 16(5), 840–854.

164. Schweitzer, A., Aufderheide, A., Rudack, T., Beck, F., Pfeifer, G., Plitzko, J. M., Sakata, E., Schulten, K., Förster, F., and Baumeister, W. (2016): Structure of the human 26S proteasome at a resolution of 3.9 Å. *Proceedings of the National Academy of Sciences* 113(28), 7816–7821.

165. Aufderheide, A., Beck, F., Stengel, F., Hartwig, M., Schweitzer, A., Pfeifer, G., Goldberg, A. L., Sakata, E., Baumeister, W., and Förster, F. (2015): Structural characterization of the interaction of Ubp6 with the 26S proteasome. *Proceedings of the National Academy of Sciences* 112(28), 8626–8631.

166. Thierbach, K., Von Appen, A., Thoms, M., Beck, M., Flemming, D., and Hurt, E. (2013): Protein interfaces of the conserved Nup84 complex from chaetomium thermophilum shown by crosslinking mass spectrometry and electron microscopy. *Structure* 21(9), 1672–1682.

167. Kim, S. J., Fernandez-Martinez, J., Sampathkumar, P., Martel, A., Matsui, T., Tsuruta, H., Weiss, T. M., Shi, Y., Markina-Inarrairaegui, A., Bonanno, J. B., Sauder, J. M., Burley, S. K., Chait, B. T., Almo, S. C., Rout, M. P., and Sali, A. (2014): Integrative Structure-Function Mapping of the Nucleoporin Nup133 Suggests a Conserved Mechanism for Membrane Anchoring of the Nuclear Pore Complex. *Molecular & Cellular Proteomics* 13(11), 2911–2926.

168. Von Appen, A., Kosinski, J., Sparks, L., Ori, A., DiGuilio, A. L., Vollmer, B., Mackmull, M. T., Banterle, N., Parca, L., Kastritis, P., Buczak, K., Mosalaganti, S., Hagen, W., Andres-Pons, A., Lemke, E. A., Bork, P., Antonin, W., Glavy, J. S., Bui, K. H., and Beck, M. (2015): In situ structural analysis of the human nuclear pore complex. *Nature* 526(7571), 140–143.

169. Fernandez-Martinez, J., Kim, S. J., Shi, Y., Upla, P., Pellarin, R., Gagnon, M., Chem-mama, I. E., Wang, J., Nudelman, I., Zhang, W., Williams, R., Rice, W. J., Stokes, D. L., Zenklusen, D., Chait, B. T., Sali, A., and Rout, M. P. (2016): Structure and Function of the Nuclear Pore Complex Cytoplasmic mRNA Export Platform. *Cell* 167(5), 1215–1228.

170. Haselbach, D., Komarov, I., Agafonov, D. E., Hartmuth, K., Graf, B., Dybkov, O., Urlaub, H., Kastner, B., Lührmann, R., and Stark, H. (2018): Structure and Conformational Dynamics of the Human Spliceosomal Bact Complex. *Cell* 172(3), 454–464.

171. Brüning, L., Hackert, P., Martin, R., Davila Gallesio, J., Aquino, G. R. R., Urlaub, H., Sloan, K. E., and Bohnsack, M. T. (2018): RNA helicases mediate structural transitions and compositional changes in pre-ribosomal complexes. *Nature Communications* 9(1), 5383.

172. Bertram, K., Agafonov, D. E., Dybkov, O., Haselbach, D., Leelaram, M. N., Will, C. L., Urlaub, H., Kastner, B., Lührmann, R., and Stark, H. (2017): Cryo-EM Structure of a Pre-catalytic Human Spliceosome Primed for Activation. *Cell* 170(4), 701–713.

173. Jennebach, S., Herzog, F., Aebersold, R., and Cramer, P. (2012): Crosslinking-MS analysis reveals RNA polymerase I domain architecture and basis of rRNA cleavage. *Nucleic Acids Research* 40(12), 5591–5601.

174. Wu, C.-C., Herzog, F., Jennebach, S., Lin, Y.-C., Pai, C.-Y., Aebersold, R., Cramer, P., and Chen, H.-T. (2012): RNA polymerase III subunit architecture and implications for open promoter complex formation. *Proceedings of the National Academy of Sciences* 109(47), 19232–19237.

175. Martinez-Rucobo, F. W., Kohler, R., Waterbeemd, M. van de, Heck, A. J., Hemann, M., Herzog, F., Stark, H., and Cramer, P. (2015): Molecular Basis of Transcription-Coupled Pre-mRNA Capping. *Molecular Cell* 58(6), 1079–1089.

176. Schilbach, S., Hantsche, M., Tegunov, D., Dienemann, C., Wigge, C., Urlaub, H., and Cramer, P. (2017): Structures of transcription pre-initiation complex with TFIIH and Mediator. *Nature* 551(7679), 204–209.

177. Vos, S. M., Farnung, L., Boehning, M., Wigge, C., Linden, A., Urlaub, H., and Cramer, P. (2018a): Structure of activated transcription complex Pol II–DSIF–PAF–SPT6. *Nature* 560(7720), 607–612.

178. Vos, S. M., Farnung, L., Urlaub, H., and Cramer, P. (2018b): Structure of paused transcription complex Pol II–DSIF–NELF. *Nature* 560(7720), 601–606.

179. Hofmann, T., Fischer, A. W., Meiler, J., and Kalkhof, S. (2015): Protein structure prediction guided by crosslinking restraints - A systematic evaluation of the impact of the crosslinking spacer length. *Methods* 89, 79–90.

180. Angov, E., Hillier, C. J., Kincaid, R. L., and Lyon, J. A. (2008): Heterologous protein expression is enhanced by harmonizing the codon usage frequencies of the target gene with those of the expression host. *PLoS ONE* 3(5), 1–10.

181. Komar, A. A. (2009): A pause for thought along the co-translational folding pathway. *Trends in Biochemical Sciences* 34(1), 16–24.

182. Caffrey, P., Bevitt, D. J., Staunton, J., and Leadlay, P. F. (1992): Identification of DEBS 1, DEBS 2 and DEBS 3, the multienzyme polypeptides of the erythromycin-

producing polyketide synthase from Saccharopolyspora erythraea. *FEBS Letters* 304(2-3), 225–228.

183. Brule, C. E. and Grayhack, E. J. (2017): Synonymous Codons: Choose Wisely for Expression. *Trends in Genetics* 33(4), 283–297.

184. Hanson, G. and Coller, J. (2018): Translation and Protein Quality Control: Codon optimality, bias and usage in translation and mRNA decay. *Nature Reviews Molecular Cell Biology* 19(1), 20–30.

185. Dana, A. and Tuller, T. (2014): The effect of tRNA levels on decoding times of mRNA codons. *Nucleic Acids Research* 42(14), 9171–9181.

186. Gardin, J., Yeasmin, R., Yurovsky, A., Cai, Y., Skiena, S., and Futcher, B. (2014): Measurement of average decoding rates of the 61 sense codons in vivo. *eLife* 3, 1–20.

187. Lavner, Y. and Kotlar, D. (2005): Codon bias as a factor in regulating expression via translation rate in the human genome. *Gene* 345(1), 127–138.

188. Makhoul, C. H. and Trifonov, E. N. (2002): Distribution of Rare Triplets Along mRNA and Their Relation to Protein Folding. *Journal of Biomolecular Structure and Dynamics* 20(3), 413–420.

189. Purvis, I. J., Bettany, A. J., Santiago, T. C., Coggins, J. R., Duncan, K., Eason, R., and Brown, A. J. (1987): The efficiency of folding of some proteins is increased by controlled rates of translation in vivo. A hypothesis. *Journal of Molecular Biology* 193(2), 413–417.

190. Buhr, F., Jha, S., Thommen, M., Mittelstaet, J., Kutz, F., Schwalbe, H., Rodnina, M. V., and Komar, A. A. (2016): Synonymous Codons Direct Cotranslational Folding toward Different Protein Conformations. *Molecular Cell* 61(3), 341–351.

191. Zhang, G., Hubalewska, M., and Ignatova, Z. (2009): Transient ribosomal attenuation coordinates protein synthesis and co-translational folding. *Nature Structural & Molecular Biology* 16(3), 274–280.

192. Zhang, G. and Ignatova, Z. (2011): Folding at the birth of the nascent chain: Co-ordinating translation with co-translational folding. *Current Opinion in Structural Biology* 21(1), 25–31.

193. O'Brien, E. P., Ciryam, P., Vendruscolo, M., and Dobson, C. M. (2014): Understanding the influence of codon translation rates on cotranslational protein folding. *Accounts of Chemical Research* 47(5), 1536–1544.

194. Tsai, S. C., Miercke, L. J., Krucinski, J., Gokhale, R., Chen, J. C., Foster, P. G., Cane, D. E., Khosla, C., and Stroud, R. M. (2001): Crystal structure of the macrocycle-forming thioesterase domain of the erythromycin polyketide synthase: versatility from a unique substrate channel. *Proceedings of the National Academy of Sciences USA* 98, 14808–14813.

195. Tanaka, Y., Bond, M. R., and Kohler, J. J. (2008): Photocrosslinkers illuminate interactions in living cells. *Molecular BioSystems* 4(6), 473–480.

196. Belsom, A., Schneider, M., Fischer, L., Brock, O., and Rappsilber, J. (2016): Serum Albumin Domain Structures in Human Blood Serum by Mass Spectrometry and Computational Biology. *Molecular & Cellular Proteomics* 15(3), 1105–1116.

197. Rubin, S. M., Lee, S.-Y., Ruiz, E. J., Pines, A., and Wemmer, D. E. (2002): Detection and characterization of xenon-binding sites in proteins by 129Xe NMR spectroscopy. *Journal of molecular biology* 322(2), 425–440.

198. Katoh, K., Misawa, K., Kuma, K.-i., and Miyata, T. (2002): MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic acids research* 30(14), 3059–3066.

199. Fiser, A. and Sali, A. (2003): Modeller: generation and refinement of homology-based protein structure models. *Methods in enzymology* 374, 461–491.

200. Svergun, D. (1999): Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophysical journal* 76(June), 2879–2886.

201. Guttman, M., Weinkam, P., Sali, A., and Lee, K. K. (2013): All-atom ensemble modeling to analyze small-angle X-ray scattering of glycosylated proteins. *Structure* 21(3), 321–331.

202. Kim, Y. C. and Hummer, G. (2008): Coarse-grained models for simulations of multiprotein complexes: application to ubiquitin binding. *Journal of molecular biology* 375(5), 1416–1433.

203. Hummer, G. and Köfinger, J. (2015): Bayesian ensemble refinement by replica simulations and reweighting. *The Journal of Chemical Physics* 143(24), 243150.

204. Tonikian, R., Zhang, Y., Boone, C., and Sidhu, S. S. (2007): Identifying specificity profiles for peptide recognition modules from phage-displayed peptide libraries. *Nature protocols* 2(6), 1368–1386.

205. Pfeifer, B. A., Admiraal, S. J., Gramajo, H., Cane, D. E., and Khosla, C. (2001): Biosynthesis of complex polyketides in a metabolically engineered strain of E. coli. *Science* 291(5509), 1790–1792.

206. Hughes, A. J. and Keatinge-Clay, A. (2011): Enzymatic Extender Unit Generation for In Vitro Polyketide Synthase Reactions: Structural and Functional Showcasing of Streptomyces coelicolor MatB. *Chemistry & Biology* 18(2), 165–176.

207. Pernot, P., Round, A., Barrett, R., De Maria Antolinos, A., Gobbo, A., Gordon, E., Huet, J., Kieffer, J., Lentini, M., Mattenet, M., Morawe, C., Mueller-Dieckmann, C., Ohlsson, S., Schmid, W., Surr, J., Theveneau, P., Zerrad, L., and McSweeney, S. (2013): Upgraded ESRF BM29 beamline for SAXS on macromolecules in solution. *Journal of Synchrotron Radiation* 20(4), 660–664.

208. Yang, B., Wu, Y.-J., Zhu, M., Fan, S.-B., Lin, J., Zhang, K., Li, S., Chi, H., Li, Y.-X., Chen, H.-F., Luo, S.-K., Ding, Y.-H., Wang, L.-H., Hao, Z., Xiu, L.-Y., Chen, S., Ye, K., He, S.-M., and Dong, M.-Q. (2012): Identification of cross-linked peptides from complex samples. *Nature methods* 9(9), 904–906.

209. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., and Barton, G. J. (2009): Jalview Version 2–a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25(9), 1189–1191.

210. Zhang, Y. (2008): I-TASSER server for protein 3D structure prediction. *BMC bioinformatics* 9, 40.

211. Roy, A., Kucukural, A., and Zhang, Y. (2010): I-TASSER: a unified platform for automated protein structure and function prediction. *Nature protocols* 5(4), 725–738.

212. Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., and Zhang, Y. (2015): The I-TASSER Suite: protein structure and function prediction. *Nature methods* 12(1), 7–8.

213. Guex, N., Peitsch, M. C., and Schwede, T. (2009): Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: a historical perspective. *Electrophoresis* 30, S162–S173.

214. Benkert, P., Biasini, M., and Schwede, T. (2011): Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics* 27(3), 343–350.

215. Bertoni, M., Kiefer, F., Biasini, M., Bordoli, L., and Schwede, T. (2017): Modeling protein quaternary structure of homo- and hetero-oligomers beyond binary interactions by homology. *Scientific reports* 7(1), 10480.

216. Bienert, S., Waterhouse, A., Beer, T. A. P. de, Tauriello, G., Studer, G., Bordoli, L., and Schwede, T. (2017): The SWISS-MODEL Repository-new features and functionality. *Nucleic acids research* 45(45), D313–D319.

217. Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F. T., Beer, T. A. P. de, Rempfer, C., Bordoli, L., Lepore, R., and Schwede, T. (2018): SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic acids research* 46(W1), W296–W303.

6 Non–scientific supplementary

6.1 Statement of Personal Contribution

As parts of thesis are either already published, prepared for publication, and/or generated by collaboration partners or students under my supervision, a detailed statement of personal contribution is given below. The results in this thesis were intensively discussed with Prof. Martin Grininger and Prof. Chaitan Khosla and as such they contributed towards the direction of individual projects.

In 2018, Prof. Martin Grininger and I published a review on "Engineering strategies for rational polyketide synthase design".[106] Large parts of the introduction were prepared from this publication and Prof. Martin Grininger contributed especially to the overview of the PKS structural scaffold (Section 1.3.1).

Chapter 2.1 was a continuation of my master's thesis project. The data presented in Sections 2.1.1&2.1.2 and was mostly obtained during the master's thesis, while the results in Section 2.1.3 were gathered during the course of this dissertation in collaboration with Matthew Ostrowski. Matthew Ostroswki cloned and purified the DEBS M1–KR2 construct and performed some of the activity measurements. The text in this thesis was prepared from the published manuscript.[104] I wrote the manuscript together with Prof. Chaitan Khosla, but all authors contributed during the editing process. From the other co–authors, Jonas Austerjost constructed the plasmid library. All authors were involved in the design of this study. Thomas Robbins and Brian Lowry mentored the master's theses of Jonas Austerjost and me.

The work presented in Chapter 2.2 was conducted by me and I wrote the corresponding text. The initial phagemid vector as well as the anti–M13 antibody were a gift from Dr.

Andreas Ernst (Goethe University Frankfurt). During a master thesis supervised by me, Melanie Janßen established ITC and AlphaLISA measurements, but the results could not be used to determine dissociation constants.

The work presented in Chapter 2.3 is largely published and the text in this thesis was prepared from the published manuscript.[105] Within this work Alicia D'Souza, an undergraduate student, was supervised by me and performed some of the cloning and activity measurements presented in this section. Dr. Aleksandra Nivina performed the sequence similarity analysis. I wrote the entire manuscript, but in the editing process changes were made by Prof. Martin Grininger, Prof. Chaitan Khosla, and to a lesser extend Dr. Aleksandra Nivina. During supervision of students in the form of two week practical courses, some proteins were purified by students.

The project presented in Chapter 2.4 was designed by me and carried out by the master student Lynn Buyachuihan under my supervision. She performed the cloning, protein purification, and turnover analysis. I generated the figures presented in this part are based on her data. I analyzed the data and wrote the text.

For the collaborative work present in Chapter 2.5, Prof. Martin Grininger and I designed the project. Dr. Florian Buhr (Schwalbe lab, Goethe University Frankfurt) harmonized the gene sequence, which was manually adjusted by Prof. Zoya Ignatova (University Hamburg). I purified the proteins and together with Karthik Paithankar performed the SAXS analysis. XL–MS measurements and manual evaluation were performed by Andreas Linden (Urlaub lab, MPI for Biophysical Chemistry, Göttingen). I analyzed the obtained cross–links. Structural modeling was done by Dr. Emanuele Rossini (Hummer lab, MPI for Biophysics, Frankfurt), who also created some of the figures. Analysis of the modeling data was largely done by Dr. Emanuele Rossini, me, and Prof. Martin Grininger. I wrote the text in this thesis, from which portions are intended for publication.

I was also part of the *in silico* study metionned in Section 2.5.1, where we employed structural analysis to characterize domain–domain interactions in a bacterial fatty acid synthase.[107] This study showed us the potential of using structural modeling on PKSs and worked as a basis for the project present in Chapter 2.5. I am a co–author on this publication as I participated in the data analysis.

6.2 Acknowledgment

First of all, I would like to thank my two supervisors Prof. Martin Grininger and Prof. Chaitan Khosla. I am entirely grateful, to have had the opportunity to pursue my PhD in this cooperation and to be able to learn from both of you. Martin, thank you for accepting me in your lab, although knowing so little about me, for teaching me how to organize my own and collaboration projects, and for being a great example of a young group leader in science. Chaitan, thank you for making so many stays at Stanford possible for me, for teaching me how to critically reflect my experiments and projects, and for sparking my passion for science in the first place. Thank you both for your ongoing support.

I would also like to thank my collaboration partners outside of the Grininger/Khosla lab for, Dr. Emanuele Rossini and Andreas Linden for putting so much effort into our project, as well as Dr. Florian Buhr and Prof. Zoya Ignatova for assisting us with the gene harmonization.

Many thanks to all the members of the Grininger, Khosla, and Kim lab for supporting my frequent travels and always making leaving hard and at the same time coming back so enjoyable.

From Frankfurt, I want to especially thank Ilka, Manuel, Florian, and Karthik: Manuel for many early morning conversations, Karthik for IT–support and traveling with me to France for the SAXS measurements, Florian for his tireless technical support, and Ilka for being such a generous bench neighbor, for helping me with the harmonization experiments, and for organizing absolutely everything with an incredibly positive attitude.

From Stanford, I am especially thanking Tom for teaching me everything about PKSs, Karthik for countless scientific and non–scientific lunch discussions, Nielson for making sure I always had a bench to work at and for being such an entertaining bench neighbor, Arek for always finding a desk for me and being the best lab bully, and Matt for reminding me that life is not just about science. Last but not least, I want to thank Aleks and Alicia. It's been amazing to have worked and laughed with you and I am so happy to have met you. I admire both of your passion for science and your determination.

Many thanks to my students Melanie, Alicia, and Lynn for working so hard and making teaching so enjoyable. This work would not have been possible without the great support from Susan in organizing my Stanford stays, from the CheM–H Metabolic Chemistry Analysis Center and the Bode group in helping with the LC–MS analysis, and from the Frankfurt lab kitchen team.

Finally, I want to thank my family and friends for their support and assistance over the last four years, especially my best friends in science Karen and Ayse for not only sharing my passion for science, but also actually contributing with their ideas and different perspectives, my family for supporting me no matter how far apart we are, and Hendrik for always being at my side and possibly being the only one to read this entire thesis.

My PhD experience wouldn't have been the same without all of you!

6.3 Eidesstattliche Erklärung

Ich erkläre hiermit, dass ich die vorgelegte Dissertation selbständig angefertigt und mich anderer Hilfsmittel als der in ihr angegebenen nicht bedient habe, insbesondere, dass alle Entlehnungen aus anderen Schriften mit Angabe der betreffenden Schrift gekennzeichnet sind.

Ich versichere, die Grundsätze der guten wissenschaftlichen Praxis beachtet, und nicht die Hilfe einer kommerziellen Promotionsvermittlung in Anspruch genommen zu haben.

Ort, Datum (Maja Klaus)

6.4 *Curriculum vitae*

Current position

09/2015 – present	PhD student, Johann Wolfgang Goethe University Frankfurt Advisor, Prof. Martin Grininger, Institute of Organic Chemistry and Chemical Biology; Buchmann Institute for Molecular Life Sciences (BMLS) Dissertation: "Engineering of Modular Polyketide Synthases"

Education

10/2013 – 07/2015	Master of Science, Life Science, Gottfried Wilhelm Leibniz University Hanover, Germany Master's thesis with Prof. Chaitan Khosla at Stanford University (CA, US) on "The Production and *in vitro* characterization of non–natural bimodular and trimodular chimeric polyketide synthases"
10/2010 – 09/2013	Bachelor of Science, Life Science, Gottfried Wilhelm Leibniz University Hanover, Germany Bachelor's thesis: "New insights into the phage shock system of *Escherichia coli*", Institute of Microbiology, advisor Prof. Thomas Brüser
08/2001 – 06/2010	Abitur at Gymnasium am Markt Bünde, Germany

Awards

06/2019	Scholarship of the DAAD to attend the GRS/GRC on "Enzymes, Coenzymes, and Metabolic Pathways"
02/2019	Poster award at Irsee Natural Products Conference
02/2016 – 01/2018	Kekulé Mobility Fellowship of the German Chemical Industry Association (VCI).
04/2014 – 09/2015	Scholarship of the German National Academic Foundation
2014	Scholarship of the DAAD (ISAP program) for a nine month stay at Stanford University
1st, 3rd, 4th and 5th academic year	Scholarship of the Leibniz University Hanover

Publications

2019	**Klaus, M**., D'Souza, A.D., Nivina, A., Khosla, C., and Grininger, M. (2019). Engineering of Chimeric Polyketide Synthases Using SYNZIP Docking Domains. ACS Chem. Biol. 14, 426–433.
2018	**Klaus, M.**, and Grininger, M. (2018). Engineering strategies for rational polyketide synthase design. Nat. Prod. Rep. 35, 1070–1081.
2018	Rossini, E., Gajewski, J., **Klaus, M.**, Hummer, G., and Grininger, M. (2018). Analysis and engineering of substrate shuttling by the acyl carrier protein (ACP) in fatty acid synthases (FASs). Chem. Commun. 54, 11606–11609.
2016	**Klaus, M.**, Ostrowski, M.P., Austerjost, J., Robbins, T., Lowry, B., Cane, D.E., and Khosla, C. (2016). Protein Protein Interactions, not Substrate Recognition, Dominates the Turnover of Chimeric Assembly Line Polyketide Synthases. J. Biol. Chem. 291, 16404–16415.