Structural Studies of Enzymes in Polyketide and Lipid Biosynthesis:

Trans-AT Polyketide Synthases and Eukaryotic Lipins

Inauguraldissertation

zur

Erlangung der Würde eines Doktors der Philosophie

vorgelegt der

Philosophisch-Naturwissenschaftlichen Fakultät

der Universität Basel

von

Solène MARTIN

aus Frankreich

Basel, 2015

Originaldokument gespeichert auf dem Dokumentenserver der Universität Basel

edoc.unibas.ch

Dieses Werk ist unter dem Vertrag „Creative Commons Namensnennung-Keine kommerzielle Nutzung-Keine Bearbeitung 2.5 Schweiz“ lizenziert.

Die vollständige Lizenz kann unter

creativecommons.org/licences/by-nc-nd/2.5/ch

eingesehen werden.

Genehmigt von der Philosophisch-Naturwissenschaften Fakultät auf Antrag von
Prof. Timm MAIER
Prof. Petr BROZ

Basel, den 21. April 2015

Prof. Dr. Jörg Schibler
Dekan

SUMMARY

Lipids are essential constituants for living organisms. They play various roles as energetic molecules, chemical messengers, membrane constituants and hormones. Several cascades of enzymatic reactions allow the conversion of glucose to acetyl-CoA, a substrate of the fatty acid synthase and the diversification of fatty acid to triacylglycerol, glycerophospholipids, sphingolipids and saccharolipids.

The first part of the thesis focuses on a homologue of the fatty acid synthase, which possesses the same structure and catalyses similar chemical reaction: the polyketide synthase (PKS). Polyketide are predominantly from microbial and plants origin. They are structurally complex compounds compared to the linear hydrocarbon chain of fatty acid. These molecules are interesting for their anti-microbial, anti-parasitic and anti-cancer properties. Understanding the synthesis mechanism of PKS is of main interest for the discovery of new therapeutical molecules. The family of *trans*-acyltransferase (AT) PKS are particularly attractive as they use an external AT for loading the acetyl-CoA to the PKS. This AT, which is not integrated into the assembly line of the PKS, may be easily switched to another AT. Incorporation of non-natural substrates to the polyketide backbone may generate new biologically active polyketides. The research project presented in this thesis investigates the mechanism of substrate delivery from the external AT to the *trans*-AT PKS. In canonical (called cis-AT) PKS, AT is attached by a rigid linker domain to the ketosynthase (KS) domain responsible for the chain elongation reaction. Here, the crystal structure of the ketosynthase domain from a *Brevibacillus brevis* (*B.brevis*) *trans*-AT PKS shows the conservation of the linker domain with a potential docking interface at the corresponding position of the AT in *cis*-AT PKS. From structure analysis and sequence alignment of 236 C-terminal linkers of KS domain from *trans*-AT PKS, four classes of docking interface were identified. The docking interface of the *B. brevis* KS-LD belonging to the class I presents a central hydrophobic groove and surrounding charge residues which form contact with neighbouring KS-LD in the crystal structure suggesting that this docking interface could be potentially used by a *trans*-AT. The structural organization of the *trans*-AT protein was further investigated in collaboration with the Dr. Roman Jakob. The *trans*-AT can be coupled to an enoyl-reductase (ER) in certain *trans*-AT PKS. The crystal structure of a *trans*-ER and the small-angle X-ray scattering (SAXS) analysis of a *trans*-AT-ER reveal the dimerization of the *trans*-AT-ER protein at the ER interface. From those results, a model of the organization of the *trans*-AT PKS could be suggested. In the *trans*-AT-ER proteins, the ER dimerizes and the AT may bind to linker domain of the KS.

The second part focuses on another protein of lipid synthesis cascade. The lipin catalyses the penultimate step of triacylglycerol synthesis by dephosphorylating the phosphatidate into diacylglycerol. The lipin is also transcription coactivator of lipid metabolism regulating genes. Role of lipin in the development of fatty liver dystrophy in mice has been shown and there are growing evidences that mutations and polymorphisms are responsible for metabolic and inflammatory disorders in human. The X-ray crystallographic and biochemical studies are of main interest to explore the effect of mutations on the lipin function and structure. Here, the bioinformatics analysis of lipin from several organisms helps mapping the structurally conserved

and highly disordered regions and allows the selection of lipin candidate susceptible to crystallise. This thesis reports a protocol optimized for the expression and purification of the *Kluyveromyces lactis* lipin homologue PAH1 construct (Klarg, Met1 to Arg527). Initial biochemical studies reveal the monomeric and oligomeric states of Klarg. These states are not in equilibrium, independent of phosphorylation and, for the oligomeric form, resistant to reducing conditions. Activity assay and crystallisation test have been started but they require further optimization.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

A	Adenylation domain
aa	Amino acid
ACC	Acetyl-Coa carboxylase
ACP	Acyl carrier protein
AFLD	Alcoholic fatty liver disease
AT	Acyltransferase
AUC	Analytical ultracentrifugation
B. brevis	*Brevibacillus brevis*
Bb	*Brevibacillus brevis*
C	Condensation domain
C16	16-carbon fatty acid
C20	20-carbon fatty acid
CoA	Coenzyme A
Ct	*Chaetomium thermophilum*
CTDNEP1	C-terminal domain nuclear envelope phosphatase 1
DAG	Diacylglycerol
DEBS	6-deoxyerythronolide b synthase
DH	Dehydratase
Dm	Drosophila melanogaster
E. coli	*Escherichia coli*
EM	Electronic microscopy
ER	Enoyl reductase
ER membrane	Endoplasmic reticulum membrane
FA	Fatty acid
FAS	Fatty acid Synthase
fld	Fatty liver dystrophy
FMN	Flavin mononucleotide
GF	Gel filtration
HAL	HIV-associated lipodystrophy
HMF	N-term-10His-Myc-Flag-TEV tag
KASI	B-ketoacyl-(acyl-carrier-protein) synthase I
Kl	*Kluyveromyces lactis*
KR	Ketoreductase
KS	Ketosynthase
LD	Linker domain
MAT	Malonyl- acetyl transferase
MEF2	Myocyte enhancer factor 2
mFAS	Mammalian fatty acid synthase
MGM	Malachite green molybdate
MST	Microscale thermophoresis
mTORC	Mammalian target of rapamycin complex
NEP1-R1	Nuclear envelope phosphatase 1-regulatory subunit 1
NFATc4	Nuclear factor of activated T-cells c4
NLS	Nuclear localization signal
NMR	Nuclear magnetic resonnance

NRPS	Non-ribosomal peptides synthetase
PA	Phosphatidate
PAP	Phosphatidate phosphatase
PBD	Polybasic domain
PC	Phosphatidylcholine
PE	Phosphatidylethanolamine
PGC-1 α	PPARγ coactivator 1 α
PI	Phosphoinositide
PKS	Polyketide synthase
PP-1cγ	Ser/Thr protein phosphatase 1 catalytic unit γ
PPARγ	Peroxisome proliferator-activated receptor γ
PPARα	Peroxisome proliferator-activated receptor α
PS	Phosphatidylserine
PUFAs	Polyunsaturated fatty acids
rmsd	Root-mean-square deviation
SAXS	Small-angle X-ray scattering
Sc	*Saccharomyces cerevisiae*
SCD	Stearoyl-Coa desaturase
SEC-MALS	Size-exclusion chromatography coupled to multi angle light scattering
SPR	Surface plasmon resonance
SRD	Serine rich domain
SREBP1	Sterol response element binding protein 1
TAD	Transcriptional activation domain
TAG	Triacylglycerol
TE	Thioesterase

INTRODUCTION

Bioassembly Of Storage Lipids

Lipids are essential for life as they provide substrates for energy metabolism, membrane synthesis, and play a critical role as secondary messengers in cellular signaling pathways [1]. Excess of free fatty acids (FA) can lead to cell dysfunction [2]. Disorders in TAG synthesis and storage occur in obesity and lipodystrophy and increase susceptibility to cardiovascular disease, diabetes mellitus, and dyslipidemia [1]. Therefore lipid storage in the form of neutral triacylglycerol (TAG) in lipid droplets is essential [2, 3].

Lipids are synthesized from the conversion of nutrient-derived carbons through a cascade of enzymatic reactions. Glucose or glutamine-derived citrate is converted to acetyl-bound Coenzyme A (acetyl-CoA), the first building block of FA (Figure 1). The acetyl-CoA carboxylase (ACC) converts the acetyl-CoA into malonyl-CoA, the second building block of FA. The repeated condensation of acetyl-CoA and malonyl-CoA by the multifunctional enzyme Fatty acid Synthase (FAS) leads to the generation of palmitic acid, a fully saturated 16-carbon (C16) FA. Palmitic acid can be further elongated and desaturated to generate a diverse spectrum of saturated antransd unsaturated FA. For example, the stearoyl-CoA desaturase (SCD) introduces a double bond at the 9th position from the carboxyl end of palmitic and stearic acid to generate monounsaturated FA. Interestingly, humans are not able to generate FA that are unsaturated at the 3rd and 6th position from the end of the carbon chain of the acyl chain. These FA named α-linolenic acid and linoleic acid are called “essential FA” and must be obtained from the diet.

FA are used to generate many different types of lipids of critical cellular function. Condensation of palmitic acid with serine followed by enzymatic reduction forms dihydrosphingosine, a precursor of the sphingolipids family. [4]. Besides playing structural roles in cellular membranes, sphingolipids such as ceramide, sphingosine, and sphingosine-1-phosphate, are bioactive signaling molecules involved in the regulation of cell growth, differentiation, senescence and apoptosis. The arachidonic acid (C20) is a precursor of eicosanoids such as prostaglandin, thromboxanes and leukotriene. Eicosanoids play an important role in inflammation by regulating vasodilatation, vascular permeability, pain and leukocytes recruitement [5].

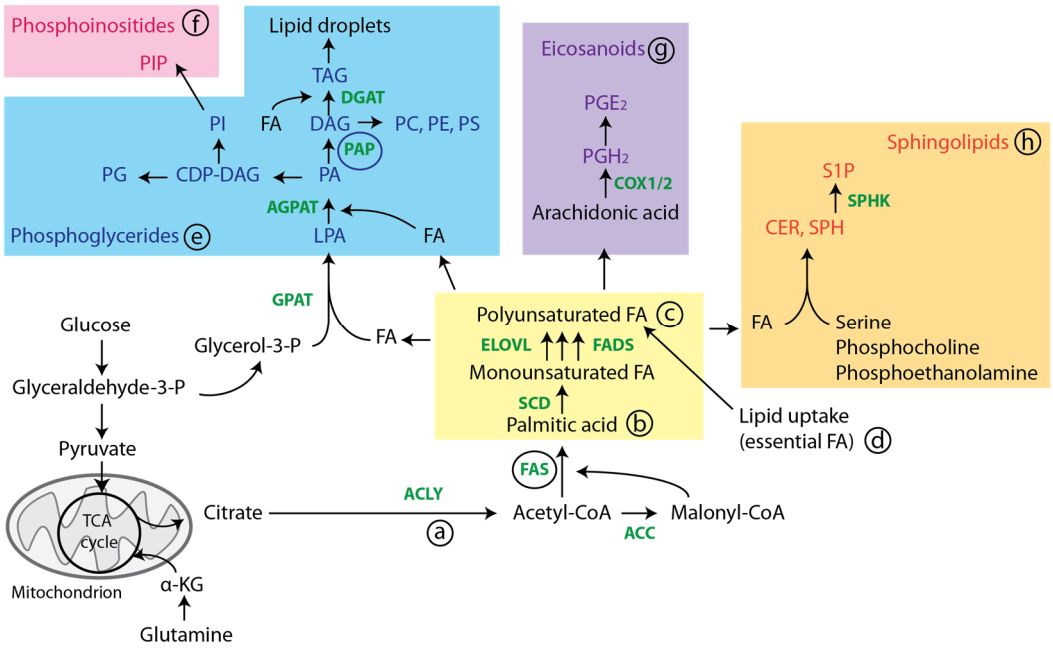


Figure 1: Lipid biosynthesis (adapted from Baenke, *et al.* 2006 [6]). Enzymes of the lipid biosynthetic pathways are indicated in green. (a) Glucose- or glutamine-derived citrate is first converted to acetyl-CoA by ACLY. (b) For FA biosynthesis, acetyl-CoA is converted into malonyl-CoA. The repeated condensation of acetyl-CoA and malonyl-CoA by the multifunctional enzyme FAS leads to the generation of palmitic acid, a fully saturated 16-carbon FA. The introduction of a double bond in the Δ9 position of the acyl chain by SCD generates mono-unsaturated FAs. (c) Subsequent elongation and further desaturation produces the repertoire of FAs with different saturation levels. (d) Essential FAs (ω3 and ω6 FAs) cannot be synthesised by human cells and need to be provided from dietary sources. (e,f) Saturated and unsaturated FAs are combined with glycerol-3-phosphate (glycerol-3-P) to generate (e) phosphoglycerides and (f) phosphoinositides. (g) Arachidonic acid, a long-chain polyunsaturated FA, is used for the synthesis of eicosanoids. (h) Sphingolipids contain acyl chains and polar head groups derived from serine, phosphocholine or phosphoethanolamine. Enzyme abbreviations: ACAT, acetyl-CoA acetyltransferase; ACC, acetyl-CoA carboxylase; ACLY, ATP citrate lyase; AGPAT, 1-acylglycerol-3-phosphate O- acyltransferase; COX1/2, prostaglandin-endoperoxide synthase (PTGS); DGAT, diacylglycerol O-acyltransferase; ELOVL, fatty acid elongase; FADS, fatty acid desaturase; FAS, fatty acid synthase; GPAT, glycerol-3-phosphate acyltransferase; PAP, phosphatidic acid phosphatase; SCD, stearoyl-CoA desaturase; SPHK, sphingosine-1-kinase. Metabolite abbreviations: α-KG, α-ketoglutarate; CDP-DAG, cytidine diphosphate-diacylglycerol; CER, ceramide; DAG, diacylglycerol; FA, fatty acid; LPA, lysophosphatidic acid; PA, phosphatidic acid; PC, phosphatidylcholine; PE, phosphatidylethanolamine; PG, phosphatidylglycerol; PGE_2, prostaglandin E_2; PGH_2, prostaglandin H_2; PI, phosphatidylinositol; PIP, phosphatidylinositol phosphate; PS, phosphatidylserine; S1P, sphingosine-1-phosphate; SPH, sphingosine; TAG, triacylglyceride. TCA: tricarboxylic acid cycle.

FA are also incorporated into the phosphoglyceride pathway, which produces on one hand the TAG for FA storage in lipid droplets and on the other hand phosphoglycerides and phosphoinositides. In this pathway, two FA are sequentially coupled to glycerol-3-Phosphate to form phosphatidate. Phosphatidate (PA) can be dephosphorylated by a phosphatidate phosphatase to diacylglycerol. Diacylglycerol is then used for the synthesis of major structural components of biological membranes like phosphatidylcholine (PC), phosphatidylethanolamine (PE) and phosphatidylserine (PS) or is further coupled to one FA chain to generate TAG. The branch of the phosphoglyceride pathway converts PA to PS, PE, PC or to phosphoinositide (PI). PI, like phosphatidylinositol (3,4,5)-trisphosphate, is highly specific binding platform for recruitment of effector protein to specific membrane compartment.

Considering such a large panel of activities, it is important that excessive FA are stored into TAG pool to protect against lipotoxicity. This thesis focuses on two effectors of the lipid synthesis and storage pathway. The first part focuses on a megasynthase presenting similar structural and functional characteristics as the FAS (black circle in Figure 1). This multidomain

megaenzyme, called Polyketide Synthase (PKS), uses similar precursors, similar biosynthetic chemistry and similar overall architecture as FAS. But instead of producing long linear chain of saturated FA, PKS are able to synthetize many structurally diverse natural products that include plant flavonoids, fungal aflatoxins and important pharmaceutical molecules that exhibit antibacterial, antifungal, immunosuppressive and antitumor properties (Figure 2) [7]. Studies on PKS megasynthase are highly relevant for the development of new therapeutical products. The family of trans-acyltransferase (AT) PKS for which a stand alone acyltransferase protein delivers the substrate to the PKS megasynthase has became more popular for its possibility to load the PKS with non-natural substrates and therefore to allow the synthesis of new biologically active polyketides. The first part of the thesis focuses on the structural organization of the trans-AT PKS and the substrate delivery mechanism.

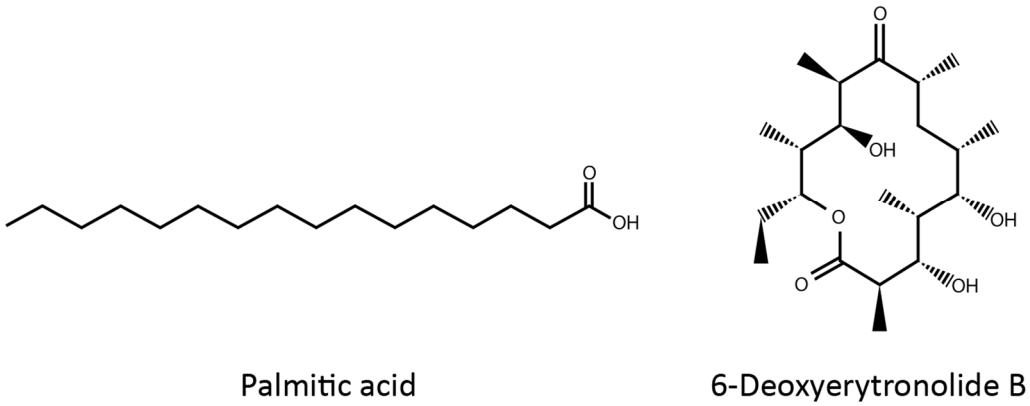


Figure 2: Chemical structure of the fatty acid Palmitic acid, a common product of the FAS, and the polyketide 6-Deoxyerythronolide, synthesized by PKS from *Saccharopolyspora erythreae* and precursor of the Erythromycine antibiotic [8].

The second part of the thesis focus on another effector of the lipid storage pathway. The phosphatidate phosphatase (dark blue circle in Figure 1), also called Lipin in eukaryotes, is responsible for the dephosphorylation of phosphatidate into diacylglycerol, the penultimate intermediate for TAG biosynthesis. Lipin is a key regulator in lipid metabolism as it also acts as transcriptional coactivators of lipid-oxidation regulating gene. More and more studies associate lipin with insulin resistance disorder, myoglobinury, inflammatory syndrome and alcoholic fatty liver disease [9, 10]. Several functional domains and many interactions with other protein partners have been identified; however the three-dimensional structure of the protein is still missing, to understand the functioning of this complex protein at a molecular level. The second part of the thesis presents the protein expression and purification strategy to produce soluble protein for biophysical and structural studies at large scale.

PART I: Structural Organization Of *trans*-AT Polyketide Synthase

Type I Polyketide Synthase

Polyketides are natural products found in bacteria, fungi and plants, which are studied for their therapeutic benefits such as antibiotic, immunosuppressing and antitumor activity (Figure 3) [11]. These complex compounds are synthesized by polyketide synthases (PKS) from acyl-CoA precursors [12]. Current research on polyketides aims to understand the biological activity of these commercially and therapeutically valuable compounds as well as the remarkable biosynthetic mechanism and malleability of PKS, enabling the generation of a large diversity of polyketides.

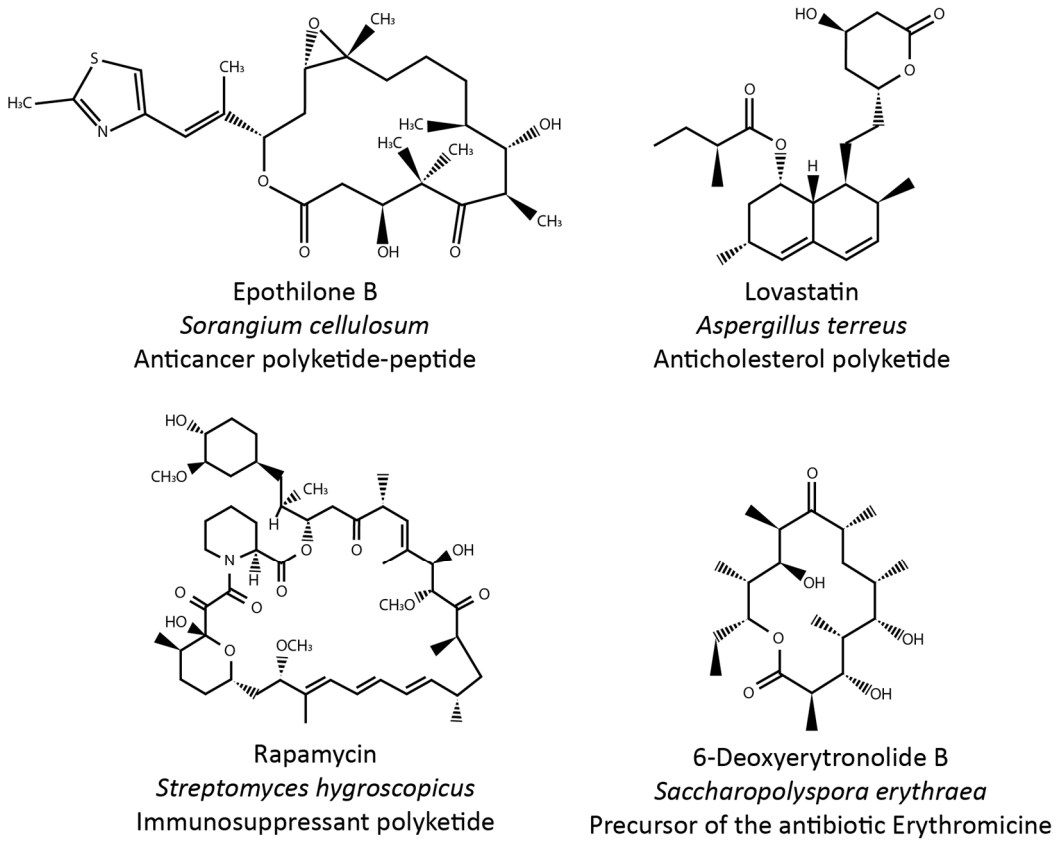


Figure 3: Structures, origins and biological activities of natural polyketides.

Bacterial PKS have been classified in three classes [12] (Figure 4). Type I PKS are multifunctional enzymes organized into modules. Each module harbors a set of distinct enzymatic domains and is responsible for the catalysis of one cycle of polyketide chain elongation and modification. They are predominantly type II PKS are multienzyme complexes, which catalyses several cycles iteratively. Each domain is reused in a cyclic fashion. Type I and II PKS use acyl-carrier proteins (ACP) to shuttle the nascent acyl-CoA intermediates from one enzymatic domain to the next. Type III PKS, also known as chalcone synthase-like PKS, are essentially iterative and do not use ACP but directly use Coenzyme A. Despite structural and mechanistic

differences, all types of PKS have a Ketosynthase (KS) activity, which catalyse a decarboxylative condensation for polyketide chain elongation.

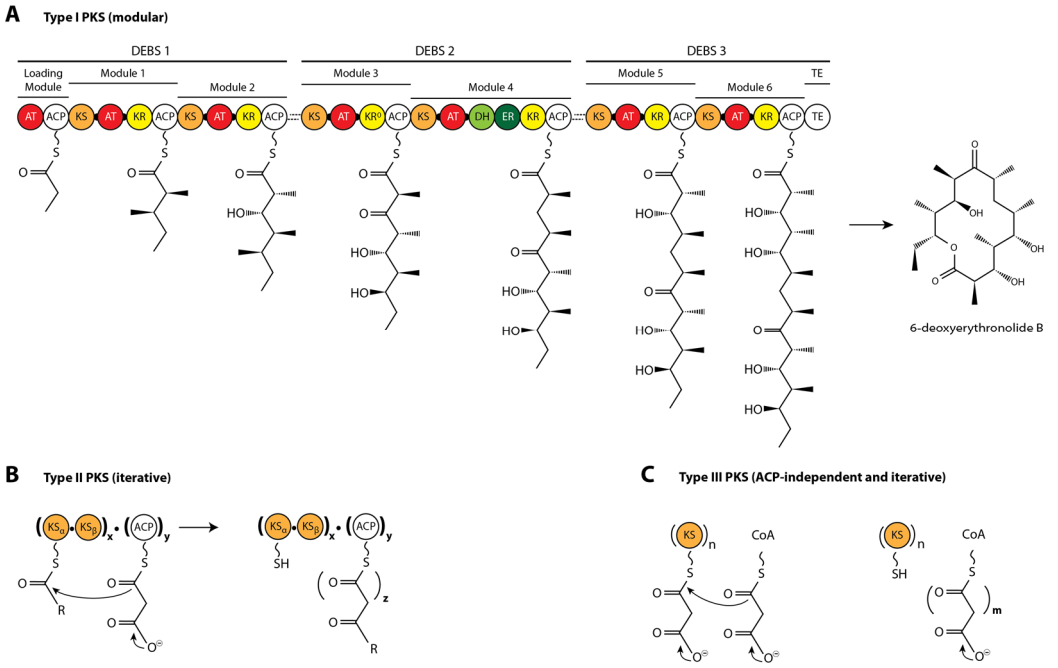


Figure 4: Domain organization and mechanism of bacterial PKS (adapted from Shen 2003 [12]). (A) Biosynthetic scheme of 6 deoxyerythronolide B, precursor of the antibiotic Erythromycin, synthesized by the modular type I PKS 6-deoxyerythronolide B Synthase (DEBS). Polyketide intermediate is transferred from one module to the next for chain elongation and modification. (B) Iterative Type II PKS. The KS domain is reused in cyclic fashion. (C) ACP-independent and iterative Type III PKS. The KS domain is directly loaded by acyl-CoA substrates.

The biosynthetic mechanism and structural organization of modular Type I PKS organization has been well described for the 6-deoxyerythronolide B Synthase (DEBS), which is responsible for the synthesis of the 6-deoxyerythronolide B, the precursor of the antibiotic erythromycin [8] (Figure 4). The DEBS is formed by three PKS megasynthases, each containing two to three modules. The minimal module for polyketide chain extension contains three functional domains: a ketosynthase (KS), an acyltransferase (AT) and an acyl-carrier protein (ACP). The minimal module can be complemented by a ketoreductase (KR), dehydratase (DH) and enoylreductase (ER) domains for the β-carbon modification of the intermediate polyketides. Three first structures of the complete mammalian fatty acid synthase (mFAS) [13] and the KS-AT domains from the DEBS module 3 and 5 [14, 15] supported by several other structures [16, 17] highlight the similar domain arrangement and chemical reaction logic of these two megasynthases.

Similarities between mFAS and Type I PKS

The crystal structure at 3.2Å resolution of the mFAS reveals the three-dimensional architecture of the domains [13] (Figure 5A). The mFAS is a homodimer in head-to-head arrangement connected by the malonyl-acetyltransferase (MAT)-DH linker domain. The condensing region composed of KS and MAT is covered by the modifying domains DH, ER and KR as well as the swinging ACP and the terminal Thioesterase (TE) domain. The malonyl-CoA and

acetyl-CoA substrates are sequentially loaded by the MAT onto the ACP and transfered to the KS domain (Figure 5B). The KS catalyses the decarboxylative condensation between the malonyl and the acetyl to produce a β-keto intermediate. The ACP transfers the β-keto intermediate to the modifying region for reduction and dehydratation into butyryl-ACP. The butyryl-ACP is shuttle back to the KS for an additional decarboxylative condensation with a malonyl substrate. The cycle of elongation and modification is repeated until the TE domain releases a fully saturated palmitic acid.

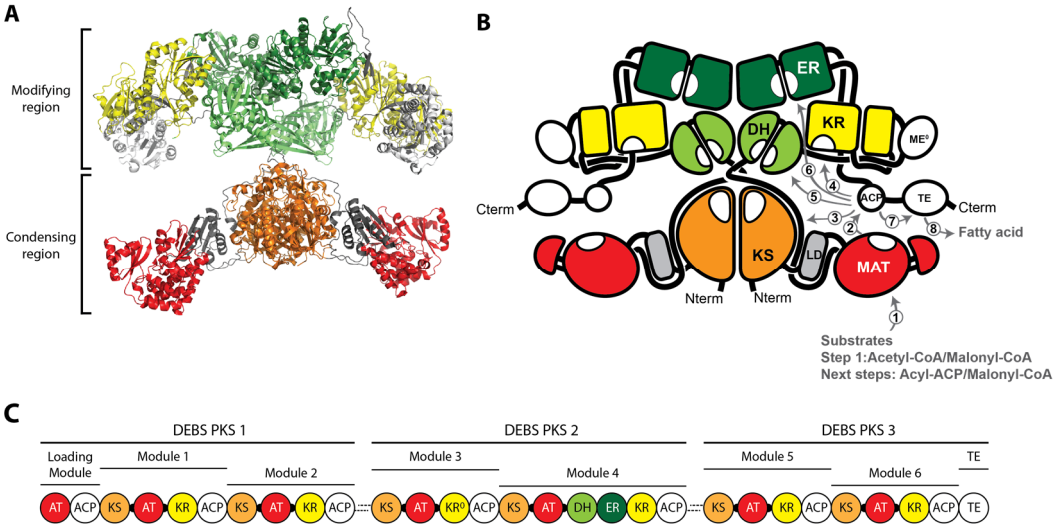


Figure 5: Main characteristics of the mFAS and the standard type I PKS. (A) Cartoon representation of the mFAS, colored by domains as shown in (B). The mFAS is a homodimer composed of two identical polypeptides connected by the MAT-DH linker domain (PDB 2vz9). The condensation and the modifications regions are indicated. (B) Schematic representation of the structure of mFAS. The root of the ACP is indicated by grey arrows and numbering. The domains are used for several cycles of acyl elongation and modification. LD: Linker Domain, ME0: inactive Methyltransferase. (C) Scheme of the organization of the standard type I PKS 6-Deoxyerythronolide B Synthase (DEBS) responsible for the biosynthesis of Erythromycin. The intermediate is transfered from one module to the next. The AT loads a substrate to the ACP of the same module.

The structure of the KS- AT homodimeric fragment from module 3 and 5 of the Type I PKS DEBS [14], [15] confirms the resemblance with mFAS (Figure 6). The thiolase fold KS domain is linked at the C-terminal to an α/β linker domain, which docks on the side of the α/β hydrolase-like AT domain. In both megasynthase, the catalytic machinery of the KS domain is comprised of a Cys-His-His tryad [7] but they present different substrate specificity. The mFAS KS is highly specific for saturated acyl chains whereas the DEBS KS accepts β-branched intermediates [7]. A number of residues lining the substrate-binding tunnel of mFAS are substituted with smaller residues in all DEBS PKS modules, resulting in a wider, more permissive tunnel in the KS of DEBS [18]. The crystal structures of AT and MAT exhibit an α/β hydrolase fold with an appended ferredoxin-like subdomain. Both MAT and AT have a catalytic Ser-His dyad located in a groove between the two subdomains. The mFAS MAT displays a broad specificity for malonyl-CoA derivatives (e.g., propionyl-CoA and methylmalonyl-CoA). MAT are distinguished from AT by the presence of a conserved active site arginine, which forms a bidentate salt bridge with the malonyl carboxylate. Mutation of this arginine to alanine or lysine reduced malonyl transacylase activity and increase selectivity for acetyl-CoA, transforming the MAT into an AT [19]. The α/β

linker domain does not have any catalytic activities but maintains the spatial organization between KS and AT in order to be reachable for the substrate-binding ACP domain.

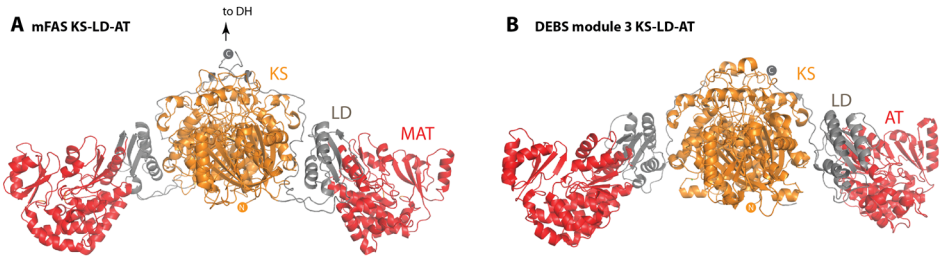


Figure 6: Structure of the crystallized dimeric KS-LD-AT domain from mFAS and DEBS module 3. (A) Structure of KS-LD-AT of the mFAS (PDB 2vz9). The domain annotation belong to the same monomer. (B) Structure of KS-LD-AT of the DEBS module 3 (PDB code 2HG4).

In contrast to mFAS, modular type I PKS do not use iterative elongation, but instead, the polyketide intermediate is transferred from one module to the next one for chain elongation. In certain modular type I PKS, modules can be used in iterative manner or can be skipped during polyketide elongation [20]. The KS is remarkably tolerant to a diverse array of extender units, while the AT likely discriminates between extender units that are native to the producing organism [21]. The substrate specificity of AT and the variation in the domain composition of modular type I PKS results in the biosynthesis of very diversified products. As the modular type I PKS polypeptide is encoded genomically in line, the structure of the product is already encoded in the gene. Mutation at the gene level of the catalytic site of PKS domain would lead to the synthesis of a modified polyketide. This represents a non-negligible advantage for PKS engineering in the pharmaceutical field and the production of new bioactive polyketide.

Engineering Type I PKS to produce new biologically active polyketide

Currently, most of the available polyketide drugs are from microbial origin [22]. Polyketide-drug discovery efforts have traditionally taken two forms: isolation and identification of new compounds from the biosphere, or the semi-synthetic modification of polyketide scaffold by chemistry [23]. There is evidence that bacteria, terrestrial and marine invertebrates, including insects, molluscs and sponges are true sources for new polyketide but unculturable microorganisms remains a serious constrain [23, 24]. An alternative strategy has also been used to produce new polyketides from genetically engineering PKS. Engineering of PKS assembly lines by site-directed mutagenesis or combinatorial biosynthesis are attractive approaches to generate novel therapeutic polyketides.

The AT is responsible for specifying and loading the appropriate CoA extender unit onto the ACP (e.g. malonyl-CoA, methylmalonyl-CoA or methoxymalonyl-CoA) [25]. The substrate specificity of the AT can be altered via site-directed mutagenesis. Site-directed mutagenesis of important residues has the potential to be minimally invasive, allowing the AT domain to remain in its native environment and causing minimal perturbation to important protein–protein interactions [26]. Mutation from the methylmalonyl-CoA specific YASH motif to the malonyl-CoA specific HAFH motif in AT of the DEBS module 1, 4 and 6 led to AT domains capable of incorporating both extender units [27-29]. Mutations outside of this motif have also allowed to a

lesser extent incorporation of the non-native extender unit. *In vitro* analysis of a YASH to HAFH motif mutant in AT of DEBS module 3 revealed that the activity of these mutants is drastically attenuated, implying that mutation of the catalytic site and incorporation of non-natural extender units results in diminished AT catalytic activity [30].

Combinatorial biosynthesis results in the construction of hybrid PKS modules by deletion, addition or swapping of domains. The general applicability of this approach was demonstrated by the synthesis of 14 different modules from eight different PKS systems and their rearrangement into 154 different bi-modular combinations, suitable for *E. coli* expression [31]. Half of the bi-modules generated the predicted triketide lactone product. However, each of the recombined modules demonstrated catalytic competence by processing unnatural substrates [31]. This suggests that the lack of activity displayed by some bi-module combinations is most likely due to faulty communication between the upstream ACP and the downstream KS of interacting modules [25]. Another approach replaced the AT domains in six of the seven geldanamycin PKS modules with a malonyl-specific AT from the rapamycin biosynthetic cluster. Four of the six hybrid modules successfully produced the predicted geldanamycin analogs, but in much lower yield relative to geldanamycin in the producing strain [32]. The use of inaccurate domain boundaries presumably destabilizes the modular architecture, resulting in reduced catalytic efficiency of the hybrid multifunctional protein.

AT domain specificity and inter-domain connections seem to be the major limiting components for successful engineering of novel and functional PKS systems. An attractive target to bypass these two limiting factors is the family of *trans*-AT PKS, which exhibit an AT activity in *trans*, compared to the previously described *cis*-AT PKS. [33-36].

Trans-AT PKS, an attractive alternative for PKS engineering

Trans-AT PKS are non-canonical PKS with the same biochemical reaction logic as modular PKS but using external proteins (commonly AT or ER) for individual reaction steps. Trans-AT PKS modules lack the internal AT domain of cis-AT PKS [33, 34, 36-39] and receive their acyl building blocks from free-standing AT enzymes acting in trans (Figure 7). In trans-AT PKS clusters, one to three AT proteins are encoded either on individual genes, fused as tandem ATs, or fused with a trans-acting ER domain into a trans-AT-ER protein (Piel 2010). In Bacillus subtilis, the trans-AT-ER PksE co-localizes with PKS megasynthases to form a megadalton-sized PKS biosynthetic structure at the cell membrane [40]. This finding suggests that trans-proteins directly interact with other PKS components. Each trans-AT protein may functionally serve several or all ACP domains of a PKS assembly line. Selection of an appropriate trans-AT for combinatorial biosynthesis could yield diversified polyketide products. Trans-AT PKS are also of particular interest for PKS engineering as they often integrate domains of the non-ribosomal peptides synthetase (NRPS) family in their assembly line. The adenylation domain (A) and the condensation (C) domain of the *Brevibacillus brevis* (*B. brevis*) trans-AT PKS incorporate an amino acid in the newly synthesized polyketide (Figure 7).

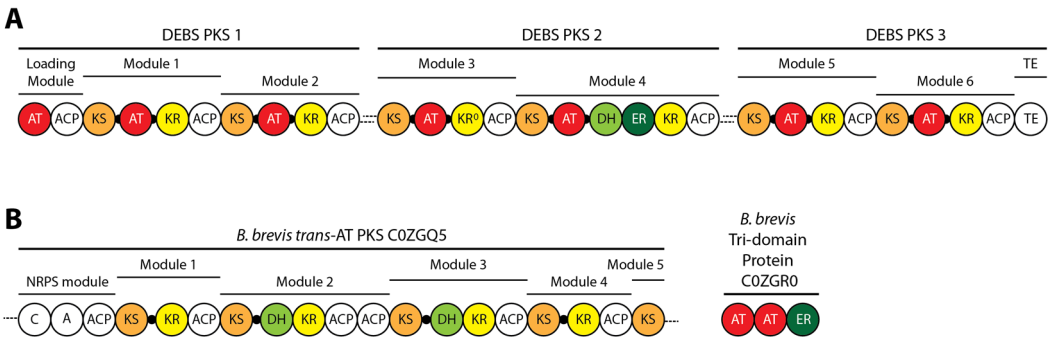


Figure 7: Schematic view of the cis-AT PKS DEBS and the hybrid NRPS/trans-AT PKS PKS C0ZGQ5 from *B. brevis*. C, Condensation domain, A, Adenylation domain.

Aim of Research: Insight into the mechanism of substrate delivery of trans-AT PKS

Trans-AT PKS are an attractive target for protein engineering and synthesis of new bioactive polyketide compounds. External proteins might be easier to modify or exchange than the intact PKS modules. However, the transfer of substrate between the *trans*-AT-ER proteins and the core KS-containing multienzyme component of *trans*-AT PKS may require specific interaction interfaces, which have not been elucidated so far. The aim of my research project was to study the interaction interface between a *trans*-AT KS domain and a *trans*-AT-AT-ER protein. I worked on *B. brevis trans*-AT PKS and used structural and biophysical approaches. X-ray crystallography was used to determine the crystal structure of the KS domain with its N-terminal linker from the module 2 of *B. brevis trans*-AT PKS and to study the potential docking interface for the *trans*-AT-AT-ER. KS and linker domains from trans-AT PKS were analysed by sequence alignment to determine component important for the structural organisation of the *trans*-AT PKS and interaction site with the *trans*-AT-AT-ER protein. Surface Plasmon Resonnance, Microscale thermophoresis and Analytical Ultracentrifugation were used to study protein-protein interaction between wild type/mutated KS domain and several constructs of the *B. brevis trans*-AT-AT-ER. Oligomeric state of the *trans*-AT-AT-ER protein was characterized by size-exclusion chromatography coupled to multi-angle light scattering (SEC-MALS) to give information of the organization of *trans*-AT-AT-ER protein from *trans*-AT PKS.

Part II: Lipin, A Key Regulator In Lipid Metabolism

The lipin family

Lipin is a bi-functional protein involved in lipid metabolism and adipogenesis. Lipin acts as phosphatidate phosphatase (PAP) in the triacylglycerol synthesis pathway and as transcriptional coactivator of the lipid-oxidation regulating gene [41]. The triacylglycerol synthesis or Kennedy pathway consists of sequential reactions starting from glycerol-3-phosphate and ending with triacylglycerol (TAG). Lipin catalyses the penultimate step by converting phosphatidic acid (PA) to diacylglycerol (DAG) (Figure 8). The phosphatidate phosphatase (PAP) activity of Lipin plays a crucial role in lipid homeostasis by controlling the relative proportions of its substrate PA and products. PA is a potent activator of MAP kinase/Erk, an inhibitor of mTORC2 (mammalian target of rapamycin complex 2) activity, a repressor of the UASino regulated genes and it regulates inflammatory signaling pathways [1], whereas DAG level regulates several signal transduction pathways through the activation of protein kinase C and protein kinase D. DAG is then used for the synthesis of the most abundant phospholipids phosphatidylethanolamine (PE) and phosphatidylcholine (PC) and for the synthesis of triacylglycerol (TAG) [41]. TAG is critical for long-term energy storage in adipose tissue and provides a source of fatty acids for oxidation in skeletal muscle and cardiac muscle. In the other branch of the Kennedy Pathway, PA is converted to cytidine di-phosphate diacylglycerol (CDP-DAG) that gives rise to phosphatidylinositol (PI), phosphatidylserine (PS), PE and PC. The regulation of lipin activity determines the distribution of the glycerol backbone between the two branches of the pathway [41].

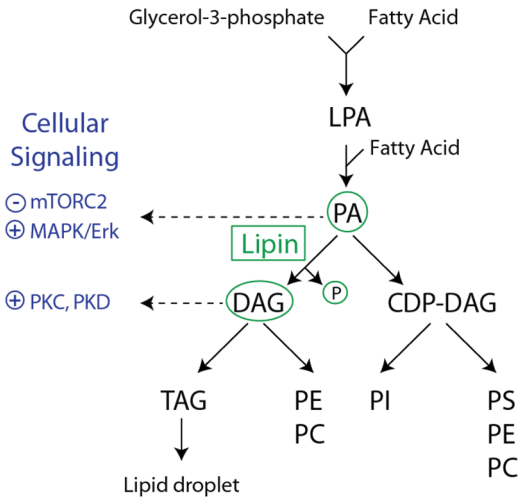
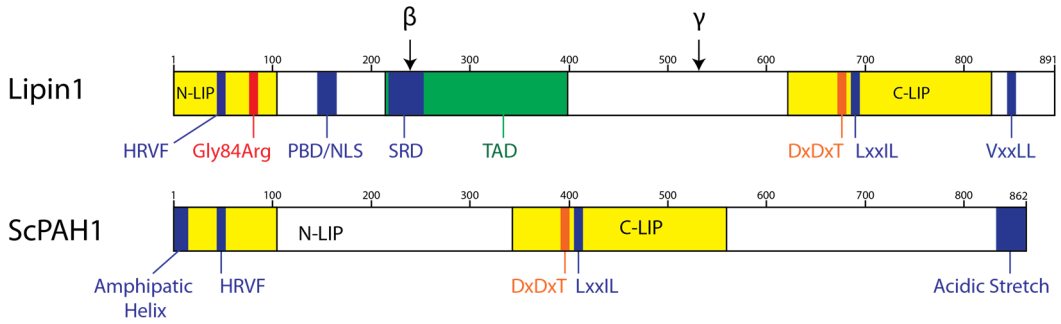


Figure 8: Simplified scheme of the role of lipins in the de novo phospholipid biosynthetic pathways. Glycerol 3-phosphate is converted through a series of acylations with fatty acids into lysophophatidic acid (LPA) and phosphatidic acid (PA). PA is a key precursor that is used, through its conversion to diacylglycerol (DAG), for the synthesis of triacylglycerol (TAG) and the major phospholipids phosphatidylethanolamine (PE) and phosphatidylcholine (PC). In the other branch of the pathway, PA is converted to cytidine di-phosphate diacylglycerol (CDP-DAG) which then gives rise to phosphatidylinositol (PI), and Phosphatidylserine (PS), PE and PC. PA is involved in cellular signaling by inhibiting mTORC2 and activating MAPK/Erk pathways. DAG activates protein kinase C (PKC) and protein kinase D (PKD).

Lipin expression has been shown in human [42], mouse [43], plants [44], insects [45], parasitic protozoan [46] and is referred to PAH1 for the yeast homologue [47]. All lipin members

contain the DxDxT motif for the phosphatidate phosphatase activity (PAP) as well as two extended stretches of highly conserved amino acid sequences at the amino- and carboxy-terminus, called as N-LIP and C-LIP (Figure 9). The yeast PAH1 presents the C-LIP in a central position.



Amino acid motifs

Gly84Arg	Fatty liver dystrophy phenotype in mouse
HRVF	Binding site for PP-1cγ
DxDxT	Catalytic site for phosphatidate phosphatase activity
LxxIL	Binding site for PPARα/PGC1α
VxxLL	Binding site for PPARγ

Domains

PBD/NLS	Polybasic domain,binding site for phosphatidate/Nuclear Localization Signal
SRD	Serine Rich Domain, binding site for 14-3-3 protein
TAD	Transcriptional activation domain, activation of PPARγ
Amphipatic Helix	Recrutement of ScPAH1 to the nuclear/ER membrane after dephosphorylation
Acidic Stretch	Binding site for Nem1p-Spo7p, dephosphorylation of ScPAH1
N-LIP and C-LIP	N-terminal and C-terminal conserved domain

Figure 9 : Mapping of the functional domains on Lipin 1 and ScPAH1. The conserved N-LIP and C-LIP domains are in yellow, the DxDxT catalytic site for PAP activity is in orange, the binding sites for protein are in blue and the transcriptional activation domain in green. The position of exon insertion for the β and γ isoforms are indicated on lipin 1.

Human and mouse encode three isoforms lipin 1, 2 and 3 [48]. Gene-expression studies have shown that these isoforms have particular expression profiles depending on the tissues [1] (Figure 10). Different lipins within the same tissue can compensate for each other in some settings but not others [1, 49]. In mouse, a broad set of tissues express lipin 1, but the enzyme is predominantly found in muscle and adipose tissue. Lipin 2 is expressed at highest level in liver and brain, and at lower level in the gastrointestinal tract and bone. Lipin 3 expression is generally much lower with the highest expression in small intestine and bones [1]. The three lipin possess PAP activity, but the intrinsic enzymatic activity varies dramatically (maximal velocity Vmax of lipin 1 >> lipin 2 > lipin 3) [50].

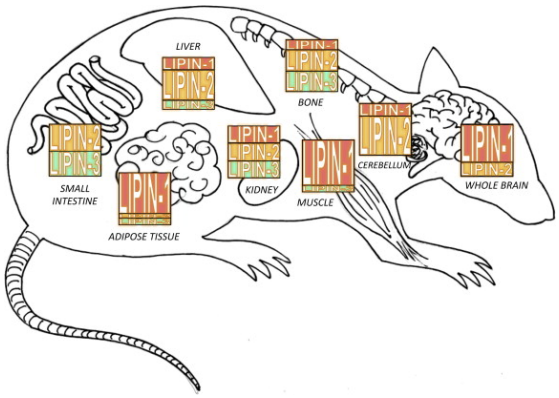


Figure 10: Mouse lipin expression profile (adapted from (Csaki et al., 2013)). The diagram indicates the relative levels of lipin 1, lipin 2 and lipin 3 mRNA expression in mouse tissues.

Lipin 1 can form hetero-oligomers in combination with lipin 2 or lipin 3 [48]. The oligomerization state of the lipin1 does not influence the PAP activity. Both the N- and C-terminal regions of lipin 1 mediate its oligomerization in a head-to-head/tail-to-tail manner. Size-exclusion chromatography suggests the tretramerization of lipin 1 [48] whereas atomic force microscopy on a lipid bilayer surface suggests an assembly from monomeric lipin to larger dodecameric complexes [51].

Lipin 1 is the most biochemically and pathophysiologically characterized member of the lipin family. Alternative splicing of the lipin 1 transcript gives rise to three distinct lipin 1 proteins (lipin 1α, lipin 1β and lipin 1γ). Lipin 1β contains 36 additional amino acids downstream the N-LIP while lipin 1γ has 26 additional residues upstream the C-LIP. Each splicing isoform has a unique cellular localization and intrinsic PAP activity [42, 52]. Lipin 1α appears to be predominantly nuclear, whereas lipin 1β resides mostly in the cytoplasm suggesting distinct cellular functions [52]. Lipin 1α is required for adipocyte differentiation whereas the major effect of lipin 1β expression is the induction of lipogenic genes [52]. The PAP activity of purified lipin 1α and lipin 1β is comparable, but their relative activity differs between two published studies that used different methods to purify lipin 1 [42, 50]. The PAP activity of lipin 1γ is almost negligible [42] but it has been shown to be the main lipin 1 isoform expressed in human brain suggesting a specialized role in regulating brain lipid metabolism [53]. Lipin 1β co-immunoprecipitates with lipin 1β, 1α, 2 and 3 [48] indicating the formation of homo- and hetero-oligomers.

Lipin, a phosphatidate phosphatase

Lipin is specific for PA and requires Mg^{2+} for PAP activity [42]. PA substrates containing 1,2-di-unsaturated and 1-saturated-2-unsaturated fatty acyl groups are equally good substrates for lipin 1 [42]. Among the substrates with unsaturated fatty acyl groups, the chain length (from C16 to C22) and the degree of unsaturation (from one to six double bonds) had little effect on the activities of lipin 1α, β and γ, but PA molecules with only saturated fatty acyl groups were poor substrates [42]. Lipin 1 catalyzes the PAP reaction based on a DxDxT motif within a haloacid dehalogenase-like domain [42]. The PAP activities of the ExDxT and DxExT mutant enzymes in yeast PAH1 were reduced by more than 99.9% [54].

Cells lacking the PAH1-encoded PA phosphatase activity contain an elevated level of PA and exhibit the derepression of the inositol-sensitive upstream-activated sequence (UASINO)-

containing phospholipid synthesis genes. When the levels of PA are high, the inhibitor Opi1p is excluded from the nucleus by binding to PA on the ER membrane and Ino2p-Ino4p activates the transcription of several phospholipid synthesis genes [55]. The yeast DAG kinase 1 (DGK1) counterbalances the function of PAH1 by phosphorylating DAG to PA (Figure 11). Cells bearing a DGK1 mutation do not exhibit any remarkable phenotypes under standard growth conditions. However, like the pah1 mutation, overexpression of DGK1 causes the accumulation of PA at the nuclear/ER membrane and the derepression of UASINO-containing genes [56].

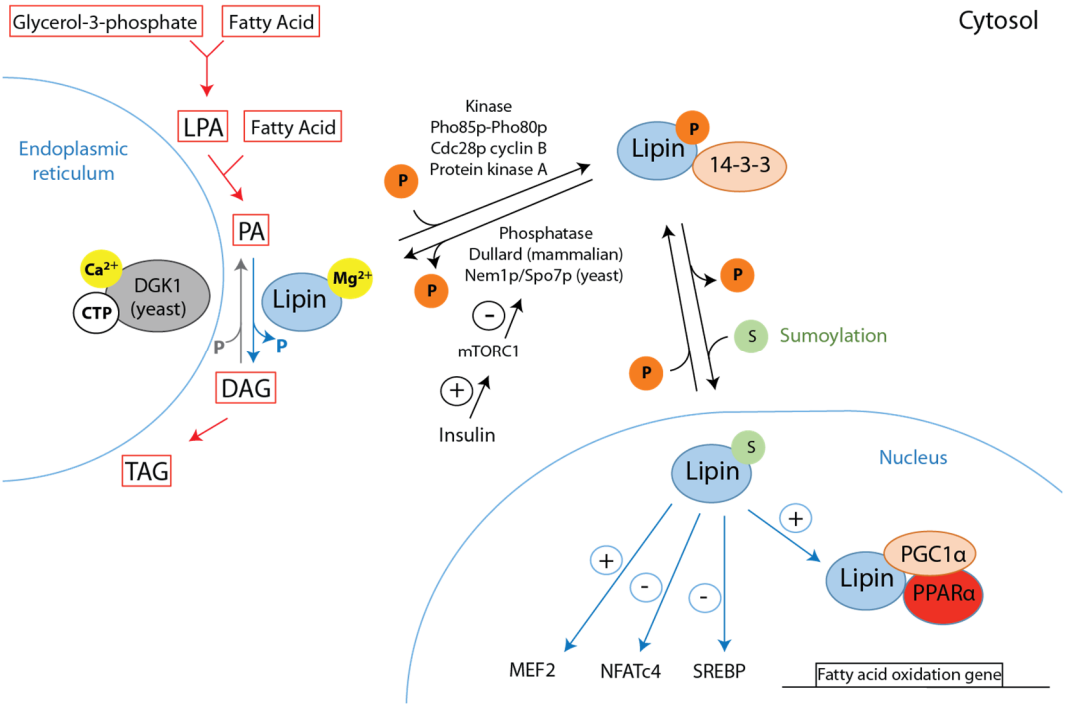


Figure 11: Scheme of the role, regulation and cellular localization of lipin 1. Lipin1 translocate from the cytosol to the ER membrane and the nucleus depending on its phosphorylation and sumoylation state. In the ER membrane, lipin 1 dephosphorylates phosphatidate to diacylglycerol whereas in the nucleus, lipin activates several transcription factors. 14-3-3: 14-3-3 protein, CTP: Cytidine triphosphate, DAG: Diacyglycerol, DGK1: Diacyglycerol kinase 1, LPA: Lysophosphatidate, MEF2: Myocyte enhancer factor-2, NFATc4: Nuclear factor of activated T cells, P: Phosphorylation, PA: Phosphatidate, PGC1α: PPARα coactivator 1α, PPARα: peroxisome proliferator-activated receptor α, S: Sumoylation, SREBP1: Sterol regulatory element binding protein, TAG: triacylglycerol, +: activation, - : inhibition.

To exert its PAP activity, lipin transiently translocates from the cytosol to the ER membrane where it dephosphorylates PA to DAG. Membrane association of lipin 1 is dependent of phospholipid composition and pH [57]. Phosphatidylethanolamine increases activity and binding of Lipin 1 to PA-containing liposomes presumably via electrostatic attraction. Lipin 1 preferentially binds di-anionic PA through a nine amino-acids KKRRKRRRK polybasic domain (PBD) [57]. In *Saccharomyces cerevisiae* conserves a KKKEK region in a similar position but the *Saccharomyces pombe* homolog does not contain more than two consecutive basic amino acids. The *Arabidopsis thaliana* homolog also lacks a PBD and yet functions similarly to yeast PAH1 [58]. Evolutionarily, the nine-amino acid PBD found in mammals did not occur until *Ciona intestinalis* and is largely conserved in all descendant species [57]. Instead, the yeast ScPAH1 contains an amino-terminal amphipathic helix (amino acid 1 to 18) that is required for membrane recruitement and PAP activity after dephosphorylation by Nem1p-Spo7p [59]. The helical

arrangement brings six residues in contact forming a hydrophobic phase (Met1, Val4, Leu8, Val11, Try15, Iso18). Deletion of these 18 amino acids as well as the mutation to alanine of four bulky residues (Met1, Val4, Leu8 and Try15) abolishes recruitment to the membrane. Mammalian lipin contains predicted amphipathic helices at their amino terminus, but whether the helix is required for membrane association has not been determined [57]. The PBD motif of mammalian lipin may be specific for PA association, whereas the amphipathic helix of yeat PAH1 is more likely involved in general membrane binding [60]. Interestingly, deletion of the KKRRKRRRK motif in lipin 1 impairs its membrane association and its nuclear import; identifying the role of the same sequence in both membrane binding and nuclear localization [61, 62].

Lipin, a transcriptional coactivator

The nuclear localization signal (NLS) is required for the translocation of lipin 1 to the nucleus where it interacts with a large set of transcription factors [62] (Figure 9). Lipin 1 activates mitochondrial fatty acid oxidation via activation of peroxisome proliferator-activated receptor α (PPARα) and peroxisome proliferator-activated receptor γ (PPARγ) coactivator 1 α (PGC-1 α) [63] (Figure 11). Lipin 1 physically interacts through its C-terminal LxxIL motif with the C-terminus of PPARα [63]. Mutation to LxxFF of lipin1 severely attenuated its interaction with PPARα. Lipin1 also regulates the transcriptional activity of PPARγ [64], a transcription factor involved in adipogenesis through the regulation of adipocyte-specific gene expression. The C-terminal VxxLL motif of Lipin 1 is critical for PPARγ binding while the central transcriptional activation domain (TAD, residues 217 to 399), is critical for PPARγ activation. The TAD includes the region corresponding to the additional exon 7 of the Lipin 1 β [64]. Interestingly, lipin1β showed more potent PPARγ activating capacity than lipin 1α. Lipin 2 and 3 conserve the LxxIL for PPARα binding [65] and both isoforms also conserve the VxxLL motif in C-LIP region for PPARγ-binding. However, they do not have any sequence homology to the TAD [64]. Accordingly, Lipin 2 is able to activate PPARα similarly to lipin1 but did not affect PPARγ activity [64]. The VxxLL motif by itself may be not sufficient to make lipin 2 and 3 an activator of PPARγ. This finding suggests that lipin1 has a unique function of activating PPARγ transcriptional activity.

Lipin 1 also regulates several others factor like the myocyte enhancer factor 2 (MEF2), the nuclear factor of activated T-cells c4 (NFATc4) and the sterol response element binding protein 1 (SREBP1) [1]. Therefore, Lipin 1 influences important metabolic processes involved in activation of fatty acid oxidation genes, adipocytes differentiation and adipocyte inflammatory gene expression.

Regulation and subcellular localization of Lipins

Lipin translocates from the cytosol to the ER membrane or the nucleus to exert its functions. Translocation and activity of lipin is regulated by several mechanisms involving phosphorylation, sumoylation and interaction with proteins like 14-3-3 protein and PP-1cγ.

Phosphorylation of lipin causes its relocation from the ER membrane and nucleus to the cytosol and attenuates its PAP and transcriptional cofactor activities. The yeast PAH1 is phosphorylated on seven sites (Ser110, Ser114, Ser168, Ser602, Thr723, Ser744, and Ser748) by the proline-directed protein kinases Pho85p-Pho80p [66], of which three sites (Ser602, Thr723, Ser744) are also the targets for the Cdc28p-cyclin B [67]. Protein kinase A also phosphorylates

PAH1 on Ser10, Ser677, Ser773, Ser774 and Ser788 [68]. Phosphorylation by Pho85p-Pho80p results in a reduced catalytic efficiency of PAP [66], whereas activity is not affected by phosphorylation via Cdc28p-cyclin B [67]. The protein kinase A-mediated phosphorylation of PAH1 inhibits its PAP activity by decreasing catalytic efficiency, albeit to a lesser extent compared to Pho85p-Pho80p [68]. A seven alanine mutant enzyme, where the seven (Ser/Thr)-Pro target sites of Pho85p-Pho80p are mutated to nonphosphorylatable alanine, exhibits elevated PAP activity. Simultaneous mutation of the three Cdc28p-cyclin B phosphorylation sites (3A) only partially mimics the physiological consequences of the seven alanine mutations [67]. The protein kinase A-mediated phosphorylation of Ser10 functions in conjunction with the phosphorylations mediated by Pho85p-Pho80p and Cdc28p-cyclin B. The Ser10Ala mutation enhances the physiological effects caused by the seven alanine mutations, whereas the Ser10Asp mutation, which mimicks phosphorylation, attenuates the effects of the seven alanine mutations. Phosphorylation of PAH1 by other kinase like protein kinase C and casein kinase II decreases its interaction with model membranes, further suggesting an inhibitory effect of phosphorylation on PAP activity [69]. Taken together, lipin undergoes a complex mechanism of phosphorylation modulating its PAP activity and cellular localization.

Cytosolic lipin is hyperphosphorylated. Hyperphosphorylated lipin associates in the cytosol with the 14-3-3 protein, which blocks its translocation to other compartment [70]. 14-3-3 protein characteristically binds to target proteins possessing recognition motifs consisting of phosphoserine followed by proline in the +2 position. Lipin 1 contains several of these motifs in a Serine Rich Domain (SRD amino acid 218-260). Structurally complex interactions between 14-3-3 and lipin 1 involve multiple phosphorylated and non-phosphorylated residues within the SRD.

Dephosphorylation of lipin enhances its activities by allowing translocation to the ER membrane and the nucleus. The human lipin is dephosphorylated by the C-terminal domain nuclear envelope phosphatase 1 (CTDNEP1, formerly “dullard”) in complex with nuclear envelope phosphatase 1-regulatory subunit 1 (NEP1-R1) [71] while the yeast PAH1 is dephosphorylated by the transmembrane protein phosphatase complex Nem1p-Spo7p [72]. Only dephosphorylated Lipin 1 can bind to the Ser/Thr protein phosphatase 1 catalytic unit γ (PP-1cγ) through a HRVF motif [73]. A mutation of the HRVF binding motif in lipin 1 to HARA abrogates the interaction with PP-1cγ as well as its nuclear translocation and the phosphatidate phosphatase activity. Lipin 2, 3 and PAH1 also contain the HRVF motif [73]. Sumoylation of lipin is also necessary for the translocation of Lipin to the nucleus. Mutating the two sumoylated residues (Lys566 and Lys596) of lipin 1 impairs its nuclear localization [74]. In conclusion, different levels of regulation may act on the same enzyme to modulate its activity.

Recent studies placed Lipin 1 under the control of insulin. Insulin increases lipin 1 phosphorylation on multiple sites and the ratio of cytosolic lipin [75]. However, insulin does not affect the PAP-specific activity measured in immune complexes isolated from insulin-stimulated fibroblasts cells using lipin antibodies [75]. Insulin controls lipin primilarly by changing localization through phosphorylation rather than intrinsic PAP activity. A complementary study shows that lipin is phosphorylated in a rapamycin-sensitive manner in response to insulin in rat adipocytes, indicating that lipin is under the control of the mammalian target of rapamycin complex 1 (mTORC1) [76]. More precisely, mTORC1 restrains Nem1p function by favoring the dephosphorylated state of Ser195 in Nem1p and potentially other unidentified residues. Nem1p

in complex with Spo7p is responsible for PAH1 dephosphorylation in yeast. By restraining Nem1p, mTORC1 favors the phosphorylated form of lipin. Finally, a study shows that insulin stimulated-phosphorylation enhances interaction of lipin 1 with 14-3-3 protein and promotes its cytoplasmic localization [70].

Lipin 1 is subject to a high degree of phosphorylation-mediated regulation. However, the mechanism of regulation is also not shared by all members of the lipin family. Lipin 2 preferentially binds di-anionic PA but has a lower rate of catalysis [77]. Like lipin 1, lipin 2 is highly phosphorylated. However, unlike lipin 1, phosphorylation of lipin 2 is not induced by insulin signaling nor is it sensitive to inhibition of the mammalian target of rapamycin [77]. Phosphorylation of lipin 2 does not negatively regulate either membrane binding or PAP activity suggesting that lipin 2 functions as a constitutively active PA phosphatase in stark contrast to the high degree of regulation for lipin 1 [77]. This knowledge of lipin 2 highlights the complexity of the function and regulation of the lipin family.

Lipin mutations, polymorphisms and associated diseases

An increasing amount of studies shows an involvement of lipin 1 in disease development. Mutation of lipin can affect the multiple functions of this protein but can also create a nice environment for pathogen development.

Originally, lipin 1 has been identified after a spontaneous complex gene rearrangement in laboratory mice [78]. These mice presented lipodystrophy, insulin resistance, peripheral neuropathy and neonatal fatty liver referred as fatty liver dystrophy (fld) (Reue, 2009). A single mutation Gly84Arg in mouse lipin 1 is able to produce the exact same phenotype as the fld mouse [43]. Humans carrying loss-of-function mutations in lipin1 do not exhibit the extreme lipodystrophy observed in fld mice. However, recent studies demonstrated that perturbations in lipin1 expression in humans result in rhabdomyolysis, weakness and myoglobinuria in children while polymorphisms studied in multiple human cohorts have been associated with body mass index disorder and insulin insensitivity [1, 79].

Mutation S734L in lipin 2 induces a different phenotype called Majeed syndrome, which is characterized by osteomyelitic lesions and dyserythropoeitic anemia [1, 65]. The lipin isoforms have different expression profiles, which may explain the different phenotype induced by lipin 1 and lipin 2 mutations [1].

Lipin 1 has been identified to play a role in the mechanisms causing HIV-associated lipodystrophy (HAL) [80] as well as alcoholic fatty liver disease (AFLD) [10]. The deletion of the PAH1 in yeast causes expanded ER membranes providing enhanced subcellular environment that may be exploited for replication of positive-RNA viruses [81].

Aim of the research work

Lipin plays a complex role in lipid metabolism. It catalyses the transformation of PA to DAG at the ER membrane but it also interacts with several transcriptional factors to regulate lipid metabolism gene expression. So far, the structure of lipin is still missing to understand at a molecular level the interaction interface with its protein partners and the influence of mutations leading to disease development. The aim of my research project was to establish a purification

protocol for the lipin and the fungi PAH1. The purified protein enabled initial studies of structure and function using biophysic methods.

PART I

THE STRUCTURAL ORGANIZATION OF *TRANS*-AT POLYKETIDE SYNTHASES: KETOACYL SYNTHASE AND *TRANS*-ACTING ENOYL REDUCTASE

Solène F. Martin[1], Roman P. Jakob[1], Dominik A. Herbst, Timm Maier*

Biozentrum, Universität Basel, Klingelbergstrasse 50/70, 4056 Basel (Switzerland)
[1] These authors contributed equally

Correspondence should be addressed to T.M. (timm.maier@unibas.ch)
E-mail addresses: solene.martin@unibas.ch, roman.jakob@unibas.ch, dominik.herbst@unibas.ch.

Keywords
Polyketide, Ketosynthase, Enoylreductase, *Trans*-AT polyketide synthase, NRPS/PKS, TIM barrel, Polyketide synthase evolution, X-ray crystallography, Small-angle X-ray scattering

STATEMENT OF MY OWN CONTRIBUTION

I cloned, expressed, purified and crystallized the KS2-LD fragment from *Brevibacillus brevis*. I determined the crystal structure with the help of Timm Maier and Dominik A. Herbst.

Roman Jakob cloned, expressed, purified, crystallized and determined the crystal structure of the MlnA-ER. He perfomed and analysed the SAXS data on MlnA-ER.

I cloned, expressed and purified several constructs of the *trans*-AT-AT-ER from Brevibacillus brevis. I tested these constructs for protein interaction experiments with wild type and LD-mutant BbKS2-LD using Microscale thermophoresis, Surface plasmon Resonnance and Analytical Ultracentrifugation.

I prepared and analysed the oligomeric states of several contructs from *trans*-AT PKS by SEC-MALS.

I analysed BbKS2-LD structure by structural comparison and sequence alignment with other KS-LD structure from FAS, *cis*-AT PKS and *trans*-AT PKS as presenting in the Results & Discussion.

The manuscript was written by myself, Roman Jakob and Timm Maier.

ABSTRACT

Trans-acyltransferase (*trans*-AT) modular polyketide synthases (PKS) catalyze the biosynthesis of a wide range of bioactive natural products. Each PKS module possesses a ketosynthase (KS) to catalyze chain elongation and the substrate to each module is delivered by free-standing *trans*-acyltransferase and enoyl reductase containing enzymes (AT-ER). To understand the interplay between the PKS megaenzymes with *trans*-AT-ER proteins, we report a structural and biophysical characterization of the KS from the second module of a *trans*-AT PKS of *Brevibacillus brevis* (Bb-KS2) and the ER domain from the MlnA *trans*-AT-ER protein (MlnA-ER) of the Macrolactin biosynthesis cluster of *Bacillus amyloliquefaciens*. The structure of Bb-KS2 revealed in addition to catalytic active thiolase domain a flanking linker domain (LD) with a characteristic helical triangle organization. Comparison with to ketosynthases identified distinct classes of the linker domain in *trans*-AT PKS. The structure of MlnA-ER demonstrates the dimeric architecture of *trans*-AT ER and revealed large expansion segments. The small-angle x-ray scattering analysis of AT-ER proteins suggests an arch-like molecule, with the ER dimer building the center and the AT the ends of the arch.

INTRODUCTION

Modular type I polyketide synthases (PKS) are microbial biosynthetic multienzymes producing potent natural products such as antibiotics [82, 83], the immunosuppressant rapamycin [84], and the antitumor drug epothilone B [85]. Type I PKS are organized into modules, responsible for one precursor elongation step [12, 83, 86]. Modules are integrated into a dimeric polypeptide, and several of these are tethered via linking domains to form one giant assembly line [87, 88]. The individual modules are variations of a common set of domains shared with animal fatty acid synthase [7, 18, 89]: An acyl transferase (AT) loads the acyl-moiety of acyl-coenzyme A (CoA) derivatives onto the acyl carrier protein (ACP). The ACP-bound intermediate is then shuttled to the ketoacyl synthase (KS) domain for chain elongation via decarboxylative condensation followed by β-carbon modification steps, catalyzed by ketoreductase (KR), dehydratase (DH) and enoyl reductase (ER) domains. The functional organization of such canonical or *cis*-AT PKS is represented also at the architectural level: The dimeric KS and the AT domain are connected by an α/β-fold linker domain (LD) and together built the well-conserved condensing part of a module, which is flexibly linked to a more variable part containing the β-carbon modification domains.

The chemical composition of a polyketide product is directly encoded in the primary structure of the PKS: individual modules carry out a single or occasionally a defined number of precursor elongation step before passing the product to the next module in sequence [7]. Product diversity is obtained via varying substrate specificities of AT domains and by altering the set of β-carbon modifying domains in each module [90, 91]. The minimal module consists of KS-AT-ACP domains and produces a β-keto intermediate, while KS-AT-KR-ACP, KS-AT-DH-KR-ACP and KS-AT-DH-ER-KR-ACP modules yield hydroxyl-, enoyl- and fully saturated acyl products in their respective elongation steps [90, 92].

Currently, most of the available polyketide drugs are from microbial origin. Re-engineering of PKS assembly lines [22, 25, 93] can yield novel therapeutic polyketides [23, 24, 30, 31, 94-98]. However, the success of engineering artificial PKS systems has been limited so far; likely due to defective inter-domain connections and non-matched domain specificities.

Trans-AT PKS are an attractive alternative target for combinatorial biosynthetic approaches [33, 34, 36, 99]. Here, PKS modules receive their acyl building blocks by dissociated AT proteins *in trans*; they lack the internal AT domain of *cis*-AT PKS [33, 34, 36-39] and may have evolved independently of *cis*-AT PKS [38]. Typically, in *trans*-AT PKS one to three AT proteins are encoded per *trans*-AT PKS gene cluster, either on individual genes, fused as tandem ATs, or most commonly fused with a *trans*-acting ER domain into a *trans*-AT-ER protein (Piel 2010). Each *trans*-AT protein may functionally serve several or all ACP domains of a PKS assembly line. Thus, selection of an appropriate *trans*-AT for combinatorial biosynthesis could yield diversified polyketide products. *Trans*-AT proteins may directly interact with other PKS components, as demonstrated for PksE, a *trans*-AT-ER protein of *Bacillus subtilis*, which co-localizes with PKS megasynthases to form a megadalton-sized PKS biosynthetic structure at the cell membrane [40].

Trans-AT proteins are monomeric approx. 35kDa proteins as exemplified by the representative structure of DszD/DisD from the Disorazol biosynthesis cluster [100] and resemble the internal AT domains of *cis*-AT PKS. In contrast, the *trans*-ER proteins adopt a completely different fold than *cis*-AT PKS internal ER domains: *trans*-ER are distantly related to the non-

canonical FMN-containing TIM-barrel ER FabK from bacterial fatty acid biosynthesis [101], while *cis*-AT PKS ER domains belong to the MDR family of short-chain dehydrogenases [102]. Compared to FabK, *trans*-AT-ER domains are about 50% larger, due to insertions to the core fold of unknown structure and function. Thus, the most common type of AT-ER proteins consists of a N-terminal 300 amino acids (aa) AT domain, which is connected via a linking region to a 450aa C-terminal ER domain.

In an effort to study the interplay between the core KS-containing multienzyme component of *trans*-AT PKS with *trans*-AT-ER proteins, we report here a structural and biophysical characterization of the KS and linking region from the second module of a *trans*-AT PKS of *Brevibacillus brevis* (Bb-KS2) and the ER domain from the MlnA *trans*-AT-ER protein (MlnA-ER) of the Macrolactin biosynthesis cluster of *Bacillus amyloliquefaciens*. Bb-KS2 is the first structurally characterized KS domain of a *trans*-AT PKS pipeline, which precedes a DH domain. Comparison with other *trans*-AT KS structures reveals a unique organization of the lateral linking region, and identifies distinct classes of *trans*-AT KS based on the conservation of core fold extensions. The dimeric structure of MlnA-ER, a representative of *trans*-AT-ER ER domains revealed extensions to the core fold positioned for supporting higher order protein-protein interactions in *trans*-AT PKS. Together with Small-angle X-ray scattering (SAXS) characterization we can deduce a plausible solution model for *trans*-AT-ER proteins.

RESULTS & DISCUSSION

To analyze the structural basis for the interplay between core *trans*-AT PKS and external *trans*-acting AT-ER proteins, we have cloned, expressed and purified core KS-domains including adjacent linker regions as well as AT-ER proteins from a series of microbial *trans*-AT assembly lines. The purified proteins were subjected to structural studies and interaction assays.

Bb-KS2-LD, a trans-AT KS region of a large Brevibacillus brevis NRPS/PKS

Soil-living gram-positive *Brevibacillus brevis (B. brevis)* is a producer of several nonribosomal peptides, including the antibiotics gramicidin and tyrocidine [103, 104]. In addition to several NRPS clusters, *B. brevis* contains one mixed NRPS/PKS cluster of about 99 kbp in size, for which no product has been described, yet. This cluster contains six megasynthase polypeptides, comprising a total of two NRPS and fourteen PKS modules (Figure S 1). The largest megasynthase of the cluster (Uniprot ID: C0ZGQ5) is 7146 amino acids in length and harbors an N-terminal NRPS module followed by five *trans*-AT PKS modules. The KS containing region (aa. 2314-2940) of module 2 of C0ZGQ5 was recombinantly expressed in *Escherichia coli* (*E. coli*), purified and crystallized (Figure 12 and Figure S 1). The structure was solved by molecular replacement using the KS domain from DEBS module 5 (DEBS-KS5) [15] and refined at 2.0 Å resolution (Table S 1). The crystallized fragment (Bb-KS2-LD) comprises the KS domain (Bb-KS2, aa. 2321-2760) and a large C-terminal extension (Bb-LD), with a compact α/β-fold (aa. 2761-2933), which links it in the intact PKS to the downstream DH domain (Figure 13).

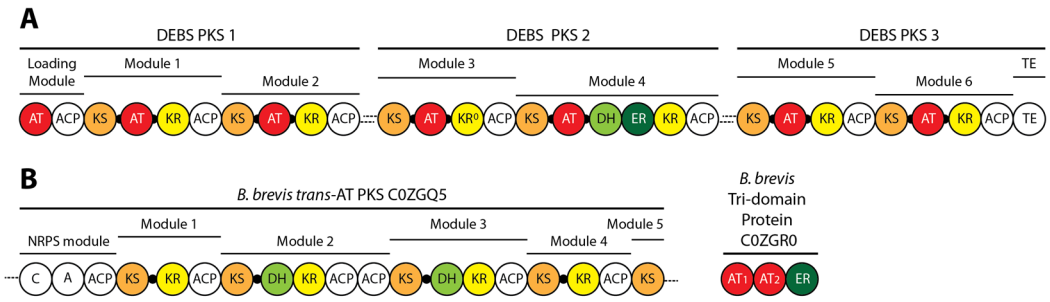


Figure 12: Schematic view of the domain organization of the DEBS *cis*-AT PKS and the *B. brevis trans*-AT PKS. (A) Domain organization of the *cis*-AT PKS from the 6-deoxyerythronolide (DEBS) operon in Erythromycin. The three multidomain proteins DEBS1, DEBS2 and DEBS3 associate together by N- and C-terminal linker domains, indicated by the dashed line. (B) Domain organization of the *trans*-AT PKS C0ZGQ5 from *B. brevis* and the external substrate loading and enoyl-reductase protein AT-AT-ER C0ZGR0. The crystal structure of KS domain of module 2 of C0ZGQ5 (shown in grey frame) was determined. Black circles indicate the position of the conserved KS-AT linker. AT, Acyltransferase; ACP, Acyl Carrier Protein; KS, Ketosynthase; DH, Dehydratase; ER, Enoyl reductase; KR, Ketoreductase; TE, Thioesterase; C, Condensation domain; A, Adenylation domain.

Bb-KS2 adopts a thiolase fold observed in all KS domains from fatty acid and polyketide synthases (Figure 13 and Figure S 2) with a canonical catalytic Cys-His-His (Cys2496, His2631 and His2671) triad (Figure S 3A) [7] and belongs to clade VIII of KS domains [103] (Table S 2 and Table S 3). Methionine 2495 narrows the substrate binding cleft and renders Bb-KS2 specific for β-unbranched substrate like malonyl-CoA [38, 105, 106] (Figure S 3B and D). Sequence-based methods predict the precursor transferred from the preceding module 1 to Bb-KS2 as a β-hydroxyacyl moiety [103] connected to an alanine that is introduced by the NRPS starter module [107] (Table S 3). The length and composition of the alanine-containing-β-hydroxyacyl precursor is unknown, as this PKS cluster from *B. brevis* is not annotated. C0ZGQ5 PKS may receive a polyketide intermediate from other PKS of the same cluster (Figure S 1). Comparison of Bb-KS2 with the co-crystal structure of the β-ketoacyl-(acyl-carrier-protein) synthase I (KASI) from *E. coli* in complex with a C12 substrate [108] demonstrates that Bb-KS2 could accommodate even larger than C12 substrates (Figure S 3C and D).

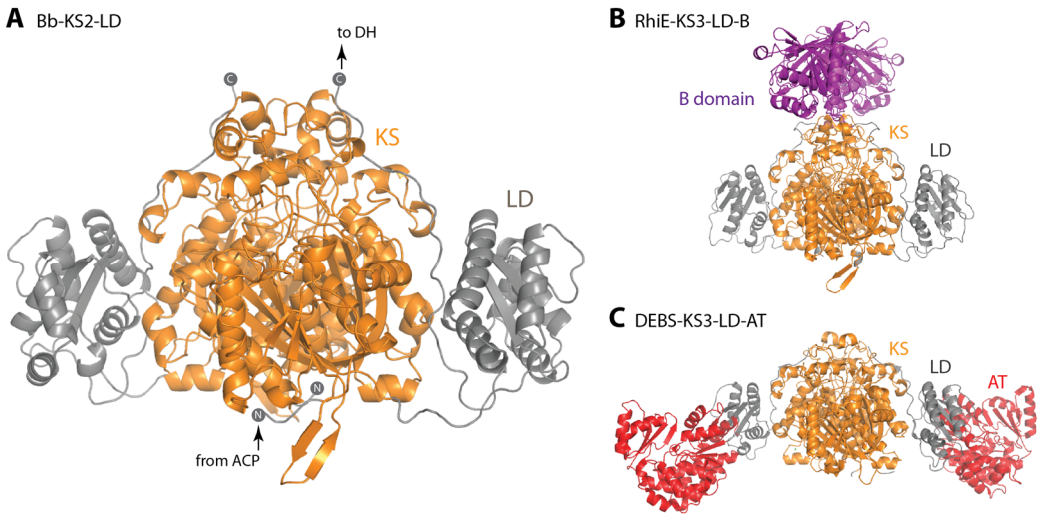


Figure 13: Overall structure of the dimeric *B. brevis* KS2 and comparison to other PKS ketosynthases. (A) Structure of the *trans*-AT Bb-KS2. N- and C-terminus are indicated as circles. The annotation of the KS and LD domains and the position of the preceding ACP and following DH domains belong to the same monomer. (B) Structure of the KS, LD and B domain of the dimeric *trans*-AT PKS RhiE (PDB code 4KC5). (C) Structure of the KS, LD and AT domains of the dimeric *cis*-PKS DEBS module 3 (PDB code 2HG4). KS, LD, AT, and B domains are annotated and colored in orange, grey, red, and purple, respectively.

Extensions of the KS core fold in Bb-KS2-LD

Bb-KS2 adopts the canonical dimeric KS structure characterized by a central ten-stranded sheets. They are formed by antiparallel β-strands donated from both protomers (Figure S 4A). A cap region composed of three α-helices, A, B & C (Figure 14A) additionally contributes to dimerization and provides 30 % of the total dimer interface. Helix A of the cap region is highly conserved in all PKS, FAS and KAS I & II ketosynthase, while helix B is present in all systems except in DEBS (Figure 14 and Figure S 2). In Bb-KS2, the N-terminal ends of helices B restrict the substrate binding cavities (Figure S 3B), while residues at the C-terminal end of these helices, Asp2458 and Lys2461, form two interchain salt-bridges across the dimer interface (Figure S 4B). The highly conserved dimeric interface of ketosynthases in *trans*- and *cis*-AT PKS [14, 15], demonstrates that overall dimeric architecture of the PKS assembly line is conserved in both families.

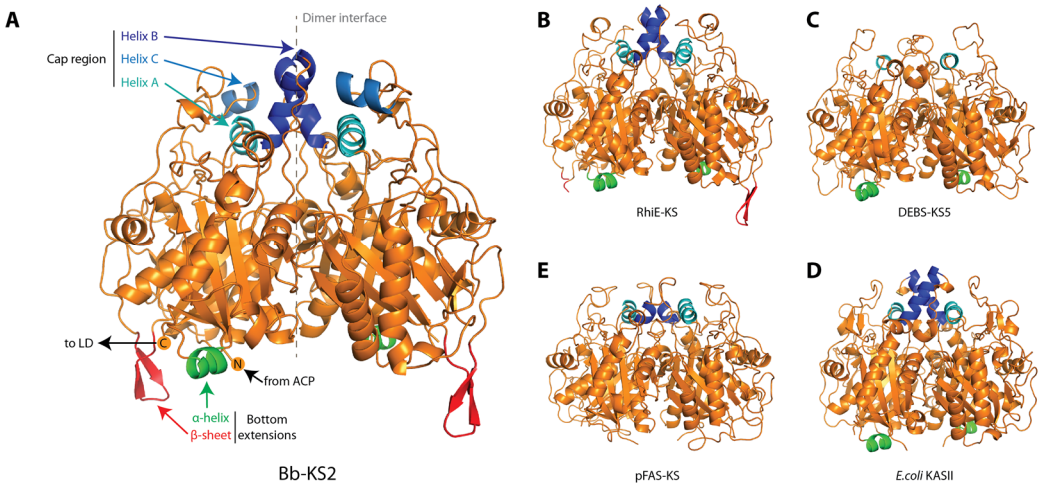


Figure 14: Comparison of the extensions of the core fold of Bb-KS2 with PKS, FAS and KAS II. (A) Characteristic extensions of Bb-KS2. The three α-helices of the cap domain are colored in different shade blue and annotated helix A (aa. 2529-2538), B (aa. 2452-2462) and C (aa. 2370-2376). The α-helix and the β-sheet extensions localized at the vicinity of the N-terminus (indicated by N in orange circle) of the KS domain are colored in green and red respectively. One α-helix extension of one monomer is localized at the back of Bb-KS2 structure. The dimer interface is indicated by a grey dash line and the domain annotations correspond to the same monomer. (B) Extensions of the KS dimer from the *trans*-AT PKS RhiE (PDB code 4KC5). One β-sheet extension of the RhiE was not built because of low electron density. (C) Extensions of the KS dimer from the module 5 of the *cis*-AT PKS DEBS (PDB code 2HG4). (D) Extensions of the KS dimer from the pFAS (PDB code 2VZ9). (E) Extensions of the KS dimer from the *E. coli* KAS II (PDB code 1KAS).

Helices A and B are structurally conserved in the recently determined crystal structures of the *trans*-AT PKS KS domain RhiE [109] and PksK [103], but helix C is specific for Bb-KS2 (Figure S 2). A hint towards the function of helix C is provided by comparison to RhiE: The corresponding sequence region of RhiE forms a loop, which is involved in interactions with the adjacent B domain (Figure S 5A). This B domain is a homologue of the DH domain, which follows Bb-KS2 in the intact assembly line. Thus helix C, together with an adjacent loop, might form a docking platform for the upstream DH domain in the full length PKS, presumably via a set of accessible charged residues (Asp2363, Glu2367, Asp 2369, Lys2372, Glu2376 and Lys 2378) (Figure S 5).

Two additional extensions localize to the vicinity of the N-terminus, the anchor point for the ACP domain of the preceding module in the intact assembly line (Figure 14). First, a short α-helix (aa. 2571-2577, green in Figure 14) is conserved in PKS, bacterial KS and fungal FAS KS, but not in animal FAS (Figure S 2). Sequence alignment of this helical insertion reveals a LxxAxxD motif shared by 79 out of 99 KS domains from *trans*-AT PKS (Figure S 6) (Table S 3 and Table S 4). The second extension in the vicinity of the N-terminus is a β-hairpin (aa. 2725-2737, red in Figure 14), which is absent in the *cis*-PKS DEBS, the eukaryotic and bacterial FAS (Figure S 2). The occurrence of a characteristic PxxxxxGxxxxxP sequence motif suggest that many, but not all other *trans*-AT PKS harbor this extension (Figure S 7) (Table S 3 and Table S 4); it is observed in the *trans*-AT PKS RhiE but replaced by a loop in the *trans*-AT KS PksK (Figure S 2). The β-hairpin is neither approaching the catalytic center nor contributing to the dimer interface; its occurrence strictly in *trans*-AT PKS suggests a role in *trans*-AT specific transient interactions.

A conserved linker domain fold with a novel cap organization in Bb-KS2-LD

Bb-KS2 is followed by a small domain comprising a three-stranded β-sheet flanked by three α-helices with a C-terminal extension that wraps around the KS domain and connects to the downstream DH (Figure 15A). This α/β domain resembles the linker domain that attaches AT to KS in *cis*-AT PKS (Figure 15B) and is thus also termed linker domain (Bb-LD). The crystal structures of *trans*-AT PKS ketosynthases PksK and RhiE also contain a linker domain, which connect in these cases the KS domain to a KR domain (PksK-LD) and to a B domain (RhiE-LD), respectively (Figure 15C and D). Comparison of Bb-LD to 99 *trans*-AT PKS ketosynthases demonstrates that 84 harbor LD of similar length (Figure S 8A). Six KS domains have an ACP domain inserted in LD and in nine KS domain the LD is absent (Figure S 8B) (Table S 3).

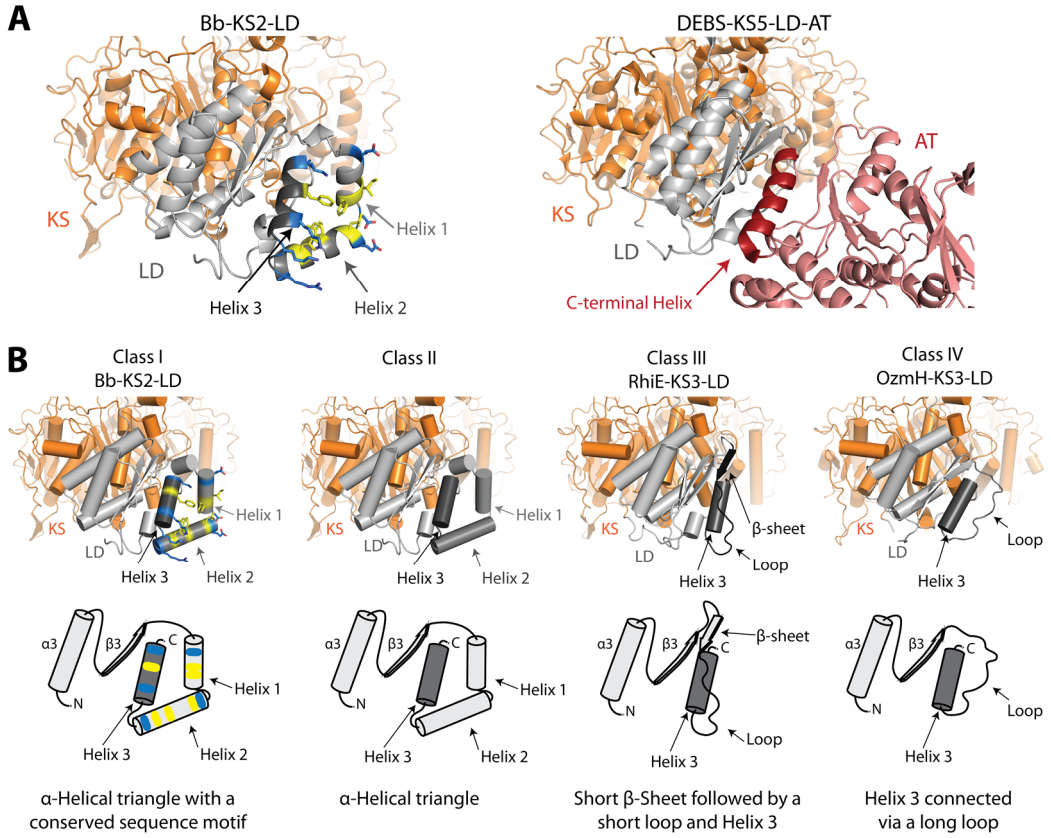


Figure 15: Comparison of the linker domain from *trans* and *cis*-AT PKS. (A) Structure of the linker domain from *B. brevis trans*-AT PKS and the linker domain of of module 5 from the DEBS operon (PDB code 2HG4). KS, LD and AT domains are colored in orange, light grey and salmon respectively. The three α- helices of Bb-LS2-LD located at the equivalent position of the AT domain in DEBS module 5 are colored by different shade of grey and numbered. Amino acids colored in yellow form a hydrophobic pocket and amino acids in blue form interaction in the crystal structure. The C-terminal helix of the AT domain of DEBS-KS5-LD-AT overlaps to the Helix 3 of Bb-LD (dark red). (B) Classification and schematic representation of the linker domain from *trans*-AT PKS. Classification was based the alignment of 212 ketosynthases sequences and PSIPRED [110] was used for secondary structure prediction. Representative structures for class I (Bb-KS2-LD; α-helical triangle with a characteristic sequence motif), class III (RhiE-KS3-LD; β-sheet - short loop - Helix 3 motif) (PDB code 4KC5) and class IV (loop - Helix3 motif; OzmH-KS3-LD) (PDB code 4OPE) are available. A model for class II (α-helical triangle without sequence conservation) was modeled [111] based on BryB-KS2 protein sequence.

Bb-LD (aa. 2864-2899), contains three α-helices arranged in a triangular shape (Figure 15A), where AT is attached to in *cis*-AT PKS (Figure 15A). This α-helical triangle has a hydrophobic surface and in the middle presents a characteristic pocket lined with hydrophobic residues

(Figure 15A, Figure S 9B and C). This region makes main crystal contacts to the neighboring LD (Figure S 9A). Key residues in Bb-LD involved in crystal packing interactions are conserved in 25 out of 90 *trans*-AT PKS including PksK-LD (Figure 15B and Figure S 9D), which constitutes one class of KS-LD found in *trans*-AT PKS. A second class of KS-LD (30 out of 90) is also predicted to harbor a helical triangle but does not contain characteristic sequence motifs found for class I.

Instead of a helical triangle, the LD of RhiE (aa. 3593-3616) contains a small β-sheet followed by a loop and a single α-helix, that corresponds to Helix 3 of Bb-LD triangle (aa. 2887-2899) (Figure 15B). Six out of 90 KS-LD belong to class III. In class IV of LD (15 of 90 proteins), represented by the LD of OzmH-KS3 (Figure 15B), only Helix 3 and a connecting loop of the helical triangle is present. The majority of the KS-LD protein can be classified into these four classes (76 of 90). For 14 out of 90 LD domains we were not able to unambiguously assign a class. From secondary sequence prediction eleven of those remaining 14 LD are supposed be exclusively α-helical in the helical triangle region, as class I and II, but the overall sequence length of the helical triangle is reduced by about ten amino acids and it is unclear if these proteins contain three very short or two longer helices (Figure 15B). OzmQ-KS-LD is the only linker domain among 236 linker sequences (Table S 3), which contains an insertion of 120 amino acids between β-sheet 3 and Helix 3, comprising a complex α-helix/β-sheet organization (PDB code 4OQJ). For the remaining two KS sequences, including the well-studied ketosynthase from module 4 of VirA [112], a LD domain with Helix 3 is present but the helical triangle region is reduced to a minimum; about 25-30 residues shorter then for class I and II and ten amino acids shorter then LD class IV, probably due to a deletion of the β-sheet 3 preceding the helical triangle (Figure 15B and Figure S 9).

In *cis*-AT PKS LD are structurally highly conserved [14-16] and probably constitute a single class. In contrast, KS-LD domains in *trans*-AT PKS show a higher variability in sequence and structure. We identified at least four different classes of KS-LD that are distributed to all *trans*-AT PKS clusters from Gram-positive and Gram-negative bacteria. A comprehensive characterization and classification of 236 KS proteins from 20 *trans*-AT PKS clusters are summarized in Table S 4. The analysis showed that Helix 3 is conserved throughout all LD from *trans*-AT PKS and about 50 % of sequences contain a LxxxWxxG motif (Figure S 8). Helix 3 structurally overlaps to the C-terminal α-helix of AT in *cis*-AT PKS (aa. 864-878 in DEBS-AT3 and aa. 856-870 in DEBS-AT5) (Figure 15B) [14, 15]. LDs in *cis*-AT PKs are thought to bind and orient the AT domain and in addition interact with the acyl-carrier protein for efficient substrate loading [8, 14]. Whether docking sites for discrete ATs and Acyl-Carrier proteins exist for *trans*-AT PKS KSs is unknown. It is tempting to speculate that the conserved Helix 3 might comprise a common general recognition motif in *trans*-AT PKS, whereas other parts of LD are used for specific interactions.

Trans-AT PKS enoyl reductases are FMN-dependent TIM-barrel proteins

In *trans*-AT PKS, AT-ER fusion proteins are stand alone enzymes and central in the biosynthetic pathway as the AT domain loads substrates to all ACP of the PKS cluster and the ER is responsible for most reduction steps during polyketide biosynthesis. For the AT domain of AT-ER proteins structural and functional characterization is available [30, 105, 113], whereas for ER domains we have only limited information. To understand the biological function of ER we have purified homologous ER domains from a series of AT-ER or AT-AT-ER (abbreviated as AT_x-ER) proteins derived from different *trans*-AT PKS assembly lines. MlnA from the macrolactin cluster

of *B. amyloliquefaciens* consists of an AT (aa. 1-300) and an ER (aa. 301-768) domain (MlnA-ER) (Figure 16A and Figure S 10). The isolated ER domain MlnA-ER was crystalized in space group $C222_1$ with two molecules per asymmetric unit. The structure was determined by molecular replacement using the structure of FabK [101] as a search model and a final structural model was refined to R_{work}/R_{free} of 24.5 %/26.3 %, respectively at 2.8 Å resolution (Table S 1).

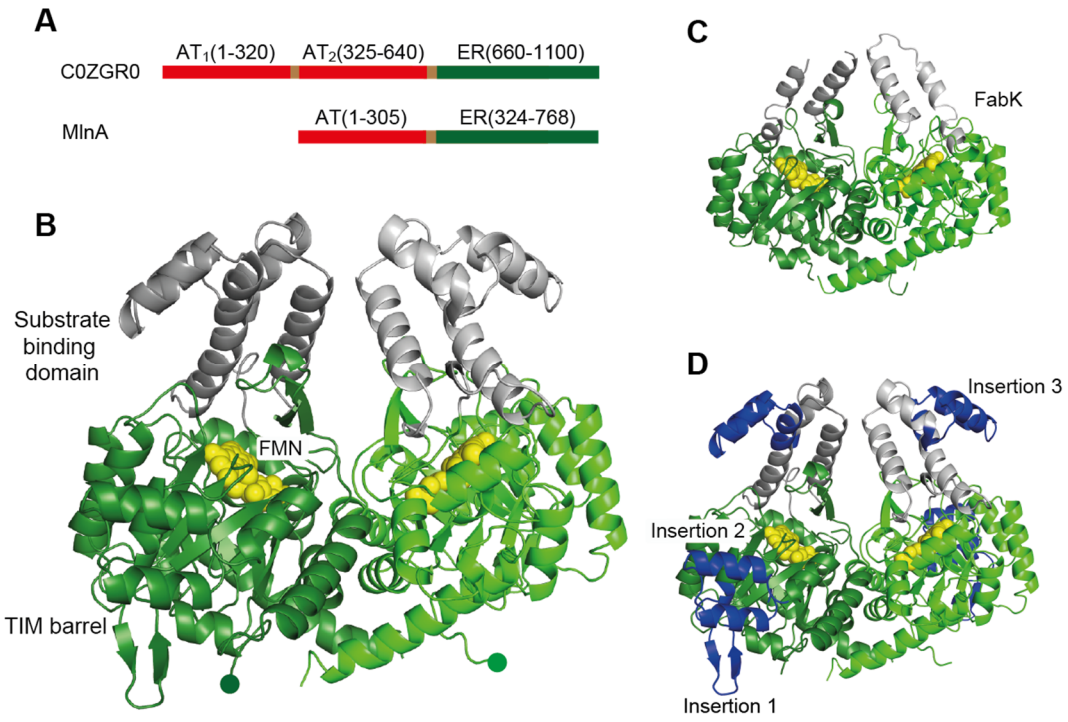


Figure 16: Overall structure of MlnA-ER and comparison to other enoyl reductases. (A) Sequence organization of MlnA (AT-ER) and C0ZGR0, the AT-AT-ER protein from the PKS cluster *B. brevis*. The AT domain, the ER domain and the linker in-between are colored in green, red and light brown, respectively. (B) Overall structure of MlnA dimer. Molecule A and B are colored in forest and green, and their substrate binding domains are in dark and light grey, respectively. The FMN is shown in yellow sphere representation. The N-terminal residue connecting to the AT-domain is shown as a ball. (C) Crystal structure of the enoyl reductase FabK (PDB code 2Z6I), colored as in (B). (D) Structure of MlnA: Domains homologous to Fabk are colored green, *trans*-AT ER specific Insertions 1 to 3 are colored in blue.

The MlnA-ER monomer comprises a $(\beta/\alpha)_8$ TIM-barrel fold (aa. 424-619 and 718-768) with a large α-helical substrate-binding domain (aa. 620-717) and binds one flavin mononucleotide (FMN) cofactor (Figure 16B). The two MlnA-ER molecules are highly similar to each other, with a root-mean-square deviation (rmsd) of 0.4 Å for all Cα atoms. The closest structural homologue of MlnA-ER is FabK (Figure 16C) with an rmsd of 1.8 Å over 300 matching Cα atoms and a sequence identity of 17 %. The two MlnA-ER monomers form a large dimer interface (Figure S 11), comprising 52 aa and interface area of 1860 Å2. The substrate binding sites are located about 17 Å apart from the dimer interface, shielded by several secondary structural elements. Dimerization arranges the exit of the active center between the TIM barrel and the substrate-binding domain in opposing direction to the general dimer interface. The attachment points of the linkers leading to the AT domain at Pro324 are about 25 Å apart from each other and do not fall in the close vicinity of the two-fold dimeric axes.

The FMN cofactor is tightly bound in the TIM barrel by several highly conserved residues (Figure S 12). The FMN phosphate interacts with four main chain amides (Gly554, Gly555, Gly576 and Ser577) and the side chains of Ser577 and Ser723 in a general phosphate binding motif also observed in FabK and other FMN-containing TIM barrel oxidoreductases [114]. The isoalloxazine ring of FMN contacts Met348, His349, Ser429 and Gln580. His349 serves as an H-bond donor with its main chain amide to the N5 of the isoalloxazine ring and modulates the redox potential of the cofactor [115]. A loop (aa. 510-520: EADCGSNTEQ) close to the active site has week electron density and was not modeled. The corresponding loop in FabK and 2-nitropropane dioxygenase contain a strictly conserved Histidine, proposed to be involved in proton shuttling [101, 116]. In *trans*-AT ER this histidine is only present in 50 % of the sequences (Figure S 13), pointing to a less important role in catalysis.

Core fold extensions of MlnA-ER are conserved in trans-AT PKS and PUFA synthases

MlnA-ER is significantly larger than its closest homologue FabK, mainly due to three insertions (Figure 16D). The first and second insertions are at the outside of the TIM barrel and pack against each other. They are comprised of residues building two β-strands (Insertion 1: aa. 441-456) and inserted between α-helix4 and β-sheet5 of the TIM barrel, and two α-helices (Insertion 2: residues 476-505) placed between α-helix7 and β-sheet8 (Figure 16D). The β-hairpin Insertion 1 is remarkably similar to the β-hairpin extension observed in Bb-KS2 (aa. 2725-2737, red in Figure 14). Both Insertions 1 and 2 are 25 Å apart from the active side of MlnA are presumably not involved in the enzymatic activity. Insertion 3 in MlnA-ER extends the substrate-binding domain by two α-helices (aa. 658-692) (Figure 16). As Insertion 1 and 2, the extension of the lid domain is probably not required for general stabilization of the core fold or a productive active site conformation, as evidenced from its absence in FabK. All extensions are conserved in *trans*-AT ER with Insertion 1 and 3 are reduced in OzmM (Figure S 13).

Phylogenetically, *trans*-AT PKS ER proteins build a distinct subclass to FabK and show a close relationship to PfaD proteins [117]. PfaD are standalone enoyl reductases of about 550 amino acids, catalyzing all reduction steps in the biosynthesis of long-chain polyunsaturated fatty acids (PUFAs) [118]. This biosynthetic pathway involves a multidomain enzyme complex related to polyketide synthases [119-121] and is present in several species of sea bacteria and recently also discovered in some terrestrial organisms [122]. PfaD homologous harbor all three insertions present in *trans*-AT ER, and additionally contain an 80 aa long N-terminal domain of unknown function (Figure S 14) [117]. The conservation of the expansion segments in *trans*-acting AT_x-ER proteins and PUFA ER suggests a conserved function, possibly the formation of interdomain or transient intermolecular protein interactions.

ER domains contribute to the dimerization of trans-acting AT_x-ER proteins/Models of AT_x-ER derived from SEC-MALS and SAXS experiments

No three-dimensional crystallographic structure of an intact AT_X-ER is available. We thus aimed to use size-exclusion chromatography coupled to multi angle light scattering (SEC-MALS) and Small-Angle X-ray Scattering (SAXS) [112, 123, 124] to determine the oligomeric state and obtain shape information, respectively.

The structural analysis of isolated domains of AT_X-ER indicated that AT is monomeric [30] and ER dimeric proteins. For SEC-MALS analysis we used the AT_X-ER protein C0ZGR0 (Figure S 1) from the single NRPS/PKS cluster of *B. brevis*. C0ZGR0 is the only protein in the cluster that shows similarity to AT and ER proteins and probably delivers substrates to all proteins in cluster. It consists of two AT and one ER protein (AT_1-AT_2-ER) (Figure 16) and shows a high similarity to MmpC from the Mupirocin [33, 125, 126] biosynthesis pathway: AT_1 shows high similarity to proofreading AT domains [127], whereas AT_2 clusters to malonyl-CoA specific loading enzymes [87, 127]. Unfortunately, full length C0ZGR0 protein could not be recombinantly expressed. Thus, we produced two overlapping didomain proteins: The first construct includes both AT domains and about half of the subsequent linker to ER (AT_1-AT_2; aa. 1-648). The second construct consists of AT_2 and ER (AT_2-ER; aa. 320-1100): AT_1-AT_2 is a monomeric protein (Figure S 16A) and AT_2-ER protein is mainly dimeric (Figure S 16B), as well as MlnA-ER (Figure S 16C). This suggests that the ER domain builds the dimeric core of AT_X-ER proteins, and monomeric AT is located at the exterior.

The SEC-MALS experiments give information about the oligomeric state of a protein, but do not provide insights into the relative domain arrangement and flexibility of AT and ER in AT_x-ER proteins. In AT_X-ER proteins, the AT_1 and AT_2 are connected via short 10-15 aa linking sequences and the AT to ER [37] is significantly longer (20 to 40 residues) and shows an increased helical propensity (Figure S 15) [128], pointing to different flexibility of the linking region.

As representative protein constructs for SAXS analysis we used AT_1-AT_2 from *B. brevis* and DfnA AT-ER from Difficidine [129, 130] biosynthesis cluster of *Bacillus amyloliquefaciens* (Figure 17A and B). Guinier's plot of the SAXS data of the AT_1-AT_2 gave an R_g of 38.9 ± 1.5 Å (Table S 5), while calculation of the distance distribution function (P(r)) using GNOM yielded a maximum particle diameter of (D_{max}) of 115 Å (Figure S 17A), in agreement with a monomeric protein. A Kratky analysis of AT1-AT2 shows a bell shaped curve (Figure S 17B), consistent with a globular particle where both AT domains have a preferred relative orientation and low flexibility. *Ab initio* modeling using DAMMIF [131] for 20 independent runs revealed a kidney shaped particle (Figure 17D).

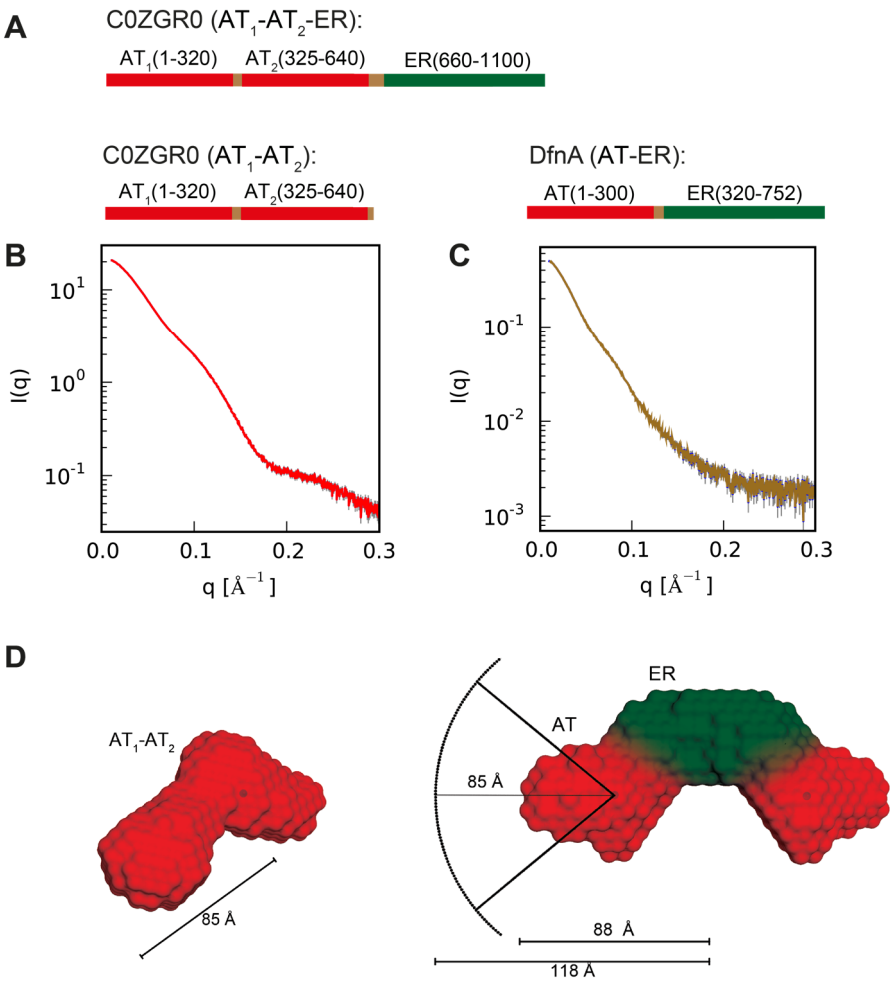


Figure 17: In solution small-angle X-ray scattering analysis of AT_x-ER proteins. (A) Constructs used for SAXS experiments. AT, ER, linker domains are colored, red, green, light brown, respectively. Experimental scattering curves of C0ZGR0 AT_1-AT_2 (B) and DfnA AT-ER (C). (D) Ab initio envelope of AT_1-AT_2 (left, red) and DfnA AT-ER (right) calculated from DAMMIF. The center of the first AT domain is indicated with a red dot. From this center the maximal distance to the second AT is 85 Å. The SAXS shape of AT-ER show a two-fold symmetry consistent with a central ER dimer (colored in green). The monomeric AT would be located at the outside of the central ER dimer. A radius of 85 Å around the center of AT in AT-ER indicates the potential boundary of the second AT domain in AT-AT-ER proteins.

For DfnA AT-ER, the determined R_g of 49.9 ± 0.2 Å and maximum particle diameter (D_{max}) of 176 Å (Figure S 17C) is consistent with a dimeric protein. The Kratky analysis of AT-ER showed an increase in the high q-range (Figure 17B), suggesting a higher domain flexibility then AT_1-AT_2. An *ab initio* modeling with *P*2 symmetry using DAMMIF [131] revealed similar arch-like shapes for 20 independent runs (Figure 17D). The shape suggests, that the dimeric ER sits at the center of the arch, and the AT domain would be located at the outside of AT-ER arch (Figure 17D).

The *ab initio* three-dimensional envelops of AT-AT and AT-ER can be combined to get insights into AT-AT-ER proteins. For AT_1-AT_2, the center of the first AT_1 has a maximal distance to the second AT_2 about 85 Å. The projection of 85 Å from the center of AT in AT-ER to the first AT_1 suggests an extended arch-like structure of AT_1-AT_2-ER with a total arch radius of approx. 120 Å (Figure 17D).

Conceptual model of trans-AT PKS assembly line

Classical *cis*-AT PKS assembly lines are characterized by their multimodular architecture where all catalytic necessary domains are integrated in giant multienzymes. This is in contrast to *trans*-AT PKS, where AT and ER proteins are standalone proteins, that transiently attach to the PKS machinery [40]. Analysis of the structural organization and characterization of the functional interplay between the PKS multienzyme and trans-AT-ER proteins is thus essential to understand trans-AT PKS biosynthesis.

Here, we report the crystal structure of ketosynthase Bb-KS2, which is the first representative of ketosynthase clade VIII, including a Met-Cys active site motif. The amino acid immediately preceding the active site cysteine is important for substrate specificity [132] and the Bb-KS2 structure helps to understand the factors dictating specificity and selectivity of KS domains, which can be used as a model for potential PKS engineering approaches. In addition to the catalytic thiolase domain, Bb-KS2 contains a linker domain LD, which is conserved throughout *trans*-AT PKS ketosynthases (Table S 4). Based on bioinformatical analysis we identified four LD classes, that differ significantly in sequence, sequence length and structure, with highly divergent LD surfaces (Figure 15). Common to all *trans*-AT PKS LD classes is Helix 3, that is central in the AT-to-KS-interaction in *cis*-AT PKS, suggesting a similar role in *trans*-AT PKS.

The structure of MlnA-ER from the macrolactin biosynthesis operon from *Bacillus amyloliquefaciens* revealed a TIM barrel fold with large characteristic extension, conserved throughout *trans*-AT PKS PUFA biosynthesis (Figure S 14). These Insertions do not contribute to the formation of the active site or substrate binding, suggesting a role in protein-protein interactions. Such interactions are indeed confirmed for PksE, a MlnA homolog of the Bacillaene biosynthesis cluster of *B. subtilis*, which colocalizes to a bimodular PKS megaenzyme, PksR, and other PKS [40].

AT-ER proteins are arch-like molecules, with the dimeric ER at the center and AT flexible attached to it (Figure 17). This dimerization might be important for efficient interplay with the dimeric PKS megaenzymes.

The *trans*-AT PKS operons of Chivosazol [133, 134] and Disorazol [133, 135] biosynthesis contain a single AT-ER. But an enoyl-reduction step appears to be unnecessary for their respective polyketide biosynthesis [133-135]. One might speculate that ER is only present in both biosynthetic pathways due to an important structural function. As stated above, in *cis*-AT PKS, KS, AT, DH, ER and KR domains are integrated in the giant PKS assembly line (Figure 18A) and existing structural and functional information can be combined in a PKS model: The overall PKS structure is homodimeric, with the ketosynthase and dehydratase domains comprising the dimer interface. The AT domain is associated to KS via the LD domain and both together form the condensing part. The ER domain together with DH and KR build the modifying part of a PKS module (Figure 18A). The presented structural and biophysical information on KS and ER in combination with recent structural information on *trans*-AT PKS [100, 109, 136] suggest a hypothetical model for the architecture of a complete *trans*-AT PKS module: The multienzyme core is formed by the dimeric KS, with the LD domain facing to the outside (Figure 18B). A DH dimer sits on top of KS and is attached to KR. The acyl-transferase and enoyl reductase activities are provided in *trans*, by the dimeric AT-ER (Figure 18B). In analogy to *cis*-AT PKS, the AT domain might interact with

the LD of KS, and additionally, the extensions of ER may contribute to interaction to the multienzyme. Such a temporary association to the PKS assembly line would support efficient substrate loading and modification and avoid slowdown of PKS biosynthesis. This interaction must be highly dynamic, as the AT-domain of AT-ER has to load substrates to all ACPs of the PKS cluster.

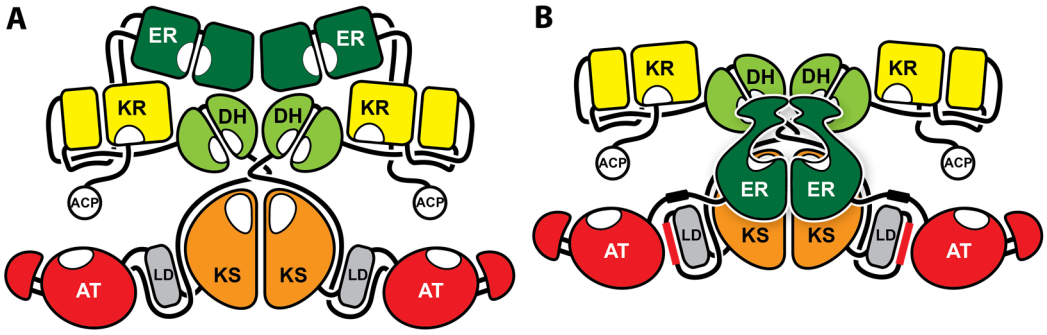


Figure 18: Model of *cis*- and *trans*-AT PKS organization. (A) Structural model of a complete *cis*-AT PKS module (KS-AT-DH-ER-KR-ACP) based on the similar domain organization to pFAS [8, 137]. The LD of the dimeric KS domain serves as docking station for AT domain. (B) Structural model of a complete *trans*-AT module (KS-AT-DH-KR-ACP) and the corresponding standalone *trans*-AT-ER protein. It may dock to the conserved LD domain of KS, in analogy to *cis*-AT PKS. The AT domain is covalently linked to the ER domain by a short linker (dark line) probably containing a α-helix (black tube). The *trans*-ER is dimeric and the positioning of AT is limited by length of the AT-to-ER linker and the large extension of ER may contribute to binding. AT, Acyltransferase; ACP, Acyl Carrier Protein; KS, Ketosynthase; LD, Linker; DH, Dehydratase; ER, Enoylreductase; KR, Ketoreductase.

To test this experimentally, we examined the interaction of Bb-KS2 with AT_1, AT_2, AT_1-AT_2 and AT_2-ER from C0ZGR0 of the corresponding *B. brevis* PKS cluster (Figure S 1) using analytical ultracentrifugation (AUC), microscale thermophoresis (MST) and surface plasmon resonance (SPR). But we could not detect any high affinity interaction (K_D < 40 µM). To test if additional domains are necessary for protein-protein interaction, we produced the complete module 2 from C0ZGQ5, including DHKR (KS2-DHKR) and examined binding to the different AT_1-AT_2-ER constructs. Additionally, KS-DH-KR and AT-ER protein constructs from the Difficidin and Macrolactin operons were produced and biophysical tested for interaction. But as well, we could not detect any interaction with K_D < 40 µM in MST, SPR and AUC. Our experiments do not exclude possible transient interactions, as for most multi-domain protein constructs solubility was limited to 50-70 µM.

In summary, the reported structural and functional analysis provides insights into the central enzymes of *trans*-AT PKS assembly lines. The developed extended classification of *trans*-AT ketosynthases is the basis for further studies to rationalize LD function and the revealed dimeric architecture of ER with characteristic extensions, will guide further enzymatic and functional studies. While KR, KS, DH and AT domains of *cis*-AT and *trans*-AT PKS share similar folds and catalytic mechanisms, there are considerable structural variations in ER: *cis*-AT ER has a medium-chain dehydrogenase/reductase (MDR) fold and *trans*-AT ER belongs to the TIM barrel enzyme family. Interestingly, this is analogous to type I fatty acid synthases (FAS) [137, 138], where metazoan FAS ER and fungal FAS ER are member of the MDR and TIM barrel fold family, respectively. Furthermore, fungal FAS and *trans*-AT ER share a common evolutionary ancestor [117]. The Insertion 3 of *trans*-AT ER, is also found in fungal FAS and used for intersubunit

connections. Our findings broaden the scope of modular assembly lines, and create new opportunities for advanced combinatorial biosynthesis of novel polyketides.

METHODS

Cloning, Expression and purification of KS-LD and MlnA-ER

Bb-KS2 (Uniprot ID: C0ZGQ5, aa. 2310-2940) and AT_1-AT_2 (Uniprot ID: C0ZGR0, aa. 1-648) and AT_2-ER (Uniprot ID: C0ZGR0, aa. 320-1100) were PCR-amplified from *Brevibacillus brevis* NBRC 100599 genomic DNA. The enoyl reductase domain of MlnA (Uniprot ID: A7Z470, aa. 300-768) was cloned from *Bacillus amyloliquefaciens* FZB42 (DSM 23117) [129, 139] genomic DNA. All constructs were cloned into the expression plasmid pNIC28-Bsa4 [140], where they are linked to a N-terminal hexa-Histag followed by a TEV-protease cleavage site. The proteins were overproduced in *E. coli* BL21(DE3) pRIL pL1SL2 [141].

After lysis of the cells in 50 mM Hepes/NaOH, 500 mM NaCl, pH 7.4, 40 mM imidazol with a sonicator and centrifugation, the proteins were purified by immobilized metal-affinity chromatography on a Ni-NTA column (elution with 250 mM imidazole), TEV digested [142], followed by a Ni-NTA column step. After concentrating the proteins were subjected to size-exclusion chromatography in 20 mM Hepes/NaOH pH 7.4, 250 mM NaCl, 5 % Glycerol and 5 mM DTT on a Superdex S200 column (GE Healthcare). The protein-containing fractions were pooled and concentrated in Amicon Ultra units (Millipore).

Protein crystallization and structure determination

Bb-KS2 crystals grew in vapor-diffusion sitting drop setups at 20 °C at a protein concentration of 10 mg/ml using 0.1 M sodium malonate, pH 7, and 15% polyethylene glycol (PEG) 3350. Rod shaped crystal were flash frozen in liquid nitrogen using Perfluoropolyether Cryo Oil. MlnA-ER crystals were obtained at 20 °C by sitting-drop vapor diffusion in 15 % PEG3350*,* 0.2 M sodium malonate, pH 7 and diffracting to 3.5 Å resolution. Dehydration of the crystals by gradually increasing the PEG3350 and ethylene glycol concentration to 30% and 20% (v/v) over 2h, respectively, improved the diffraction to 2.8 Å resolution. Data were collected at beamlines X06SA and X06DA of the Swiss Light Source (Paul Scherrer Institut, Villigen, Switzerland and processed using XDS [143, 144]. Bb-KS2 crystals belong to space group C2 with unit cell parameters of a= 113.3 Å, b = 85.8 Å and c=89.3 Å, with one molecule in the asymmetric unit. Structure determination was performed by molecular replacement with KS structure of DEBS5 as search model (Protein Data Bank accession code 2HG4, [15]). The final model includes residues 2318 to 2934. MlnA-ER crystals belong to space group $C222_1$ with cell dimensions a = 83.2 Å, b = 104.7 Å, and c = 216.1 Å and contain two molecules per asymmetric unit. Structure determination was performed by molecular replacement with FabK ER structure as search model (Protein Data Bank accession code 2Z6I, [101]. The final model comprises residues 325 to 767, the first 23 residues of the expressed protein, which would connect the ER domain to the AT domain in the native protein are disordered. Model building and structure refinement were performed for both structures with Coot [145], PHENIX [146] and Buster-TNT [147] (Tab. S1).

SEC-MALS

For size-exclusion chromatography coupled with multi-angle light scattering (SEC-MALS) measurements, 20 µl or 100 µl samples of 1-4 mg/ml protein were applied to a GE Healthcare Superdex 200 5/150 GL SEC column equilibrated overnight in either 20 mM Hepes pH 7.4, 250 mM NaCl, 0.5 mM TCEP, 5% glycerol or without glycerol at 5 °C, using an Agilent 1100 series HPLC system. Light-scattering and differential refractive index measurements were made using Wyatt miniDawn TriStar detector and a Wyatt Optilab rRex detector, respectively. The inter-detector delay volumes, band broadening, and the light-scattering detector normalization, were calibrated according to the manufacturer's protocol using a 2 mg/ml BSA solution (Thermo Pierce) run in the same buffer. The absolute refractive index of the buffer was measured using the refractive index detector. The data were collected and processed using the Wyatt Astra 5 software. The molar mass was calculated from a global fit of the light scattering signals from three detectors at different angles, and the differential refractive index signal, using algorithms in the Astra 5 software.

Small angle X-ray scattering

Small angle X-ray scattering (SAXS) data were collected at the coherent SAXS beamline X12SA at the swiss light source in Villigen Switzerland. All samples were dialyzed into freshly prepared 20 mM Hepes/NaOH pH 7.4, 250 mM NaCl, 5 % Glycerol and 5 mM DTT, diluted to concentrations between 3-12 mg/ml and constantly centrifuged (13,000 xg, 8 °C) until measurement. Glass capillaries (1 mm inner diameter) were mounted to a temperature-controlled holder (12 °C). Data acquisition was performed using a Pilatus 2M detector, a detector distance of 2.14 m and an energy of 12 keV. 10 individual capillary positions and different data collection strategies were conducted (Tab. S2). New capillaries were used for every protein concentration and sample. Frames with artefacts (*e.g.* caused by air bubbles) were identified using beamline software (SAXS_inspect2) and excluded from the buffer and protein data sets. Radial averages were calculated using beamline software and exported to individual ascii files, containing the scattering vector q, the intensities, and the intensity errors. The scattering vector ranged from 0.005 to 0.7 $Å^{-1}$ and is defined as $q = 4\pi/\lambda sin\theta$, were θ is half the scattering angle. If not more than 30 % of DfnA scattering profiles of the same capillary position and scan repeat were rejected, DATAVER [148] was used for profile averaging. Buffer profiles were subtracted using DATOP [148]. Scaling factors and p-values of a students-t test were analyzed using DATMERGE [148] and DATCMP [148], respectively, and plotted as function of capillary position and scan repeat [148]. The last 60 % of the data sets indicated a significant increase in scaling factors and a drop in p-values and therefore were excluded from further processing due to radiation damage. All remaining profiles were merged using DATMERGE [148], if no capillary position dependent fluctuation in scaling factors was observed. Artificial low resolution points, caused by the beamstop and micro aggregate, were removed in PRIMUS [149]. The radius of gyration (R_g) and zero angle intensity (I(0)) was calculated from the Guinier approximation using AUTORG [148] and is consistent with the values determined via the atomic distance distribution p(r) using DATGNOM. Scattering profiles at different concentrations were only merged using DATMERGE [148], if a noise reduction at medium and high scattering vectors could be obtained.

Ab initio shapes were calculated with the program DAMMIF [131] assuming prolate anisometry and two fold symmetry for DfnA. Shapes with low normalized special discrepancy were selected using DAMSEL [150], aligned with DAMSUP [150], averaged and filtered with DAMAVER [150]. Shapes were visualized using Pymol [151].

Data deposition

The atomic coordinates for Bb-KS2 and MlnA-ER have been deposited in the RCSB Protein Data Bank and are available under the accession code 4Z37 and 4Z38, respectively.

ACKNOWLEDGEMENTS

We thank the staff at the Swiss Light Source (Villigen, Switzerland) for their support for data collection and Prof. Peter Leadlay for providing pL1SL2. The authors thank the staff of PXIII and cSAXS beamlines from the Swiss Light Source, Paul Scherrer Institute, Villigen, Switzerland and Tim Sharpe (Biophysics Facility of the Biozentrum, University of Basel) for support with the SEC-MALS experiments.

FOOTNOTES

*This work was supported by the Swiss National Science Foundation (Grants 125357, 138262 and R'EQUIP 145023).

ABBREVIATIONS

AT, acetyl transferase; DH, dehydratase; ER, enoyl reductase; KR, ketoacyl reductase; KS, ketoacyl synthase; PPT, phosphopantetheinyl transferase; ACP, acyl carrier protein; LD, linker domain; TIM, triose phosphate isomerase; PKS, polyketide synthase; SEC-MALS, size-exclusion chromatography coupled to multi angle light scattering; SAXS, Small-angle X-ray scattering;

SUPPLEMENTARY INFORMATION

THE STRUCTURAL ORGANIZATION OF *TRANS*-AT POLYKETIDE SYNTHASES KETOACYL SYNTHASE AND *TRANS*-ACTING ENOYL REDUCTASE

Solène F. Martin[1], Roman P. Jakob[1], Dominik A. Herbst, Timm Maier*

Biozentrum, Universität Basel, Klingelbergstrasse 50/70, 4056 Basel (Switzerland)
[1] These authors contributed equally

Correspondence should be addressed to T.M. (timm.maier@unibas.ch)
E-mail addresses: solene.martin@unibas.ch, roman.jakob@unibas.ch,
dominik.herbst@unibas.ch.

Table S 1: Data collection and refinement statistics of Bb-KS2-LD and MlnA-ER. Values in parentheses correspond to the highest resolution shell. Datasets were collected at beamline X06DA at the Swiss Light Source in Villigen (Switzerland).

	Bb-KS2-LD	**MlnA-ER**
PDB accession number		
Data collection statistics		
Spacegroup	C2	$C222_1$
Cell dimensions and angle	a = 113.29 Å	a = 83.17 Å
	b = 85.80 Å	b = 104.67 Å
	c = 89.25Å	c = 216.10 Å
	β = 121.5°	
Wavelength λ	1.00000 Å	1.00000 Å
Resolution	42.9 to 2.0 Å	108.0 – 2.8 Å
Reflections	146,525 (22,323),	155,632 (15,978)
Unique reflections	48,935 (7,705)	23,587 (2,324)
Completeness	98.7% (97.2%)	99.7% (99.0%)
Multiplicity	3.0 (2.9)	6.6 (6.9)
R_{meas}	8.9% (79.3%)	7.4% (70.4%)
$CC_{1/2}$	99.8 (72.7)	99.9 (81.8)
I/σI	11.9 (1.7)	18.9 (1.7)
Wilson B-factor	38.0 Å^2	80.0 Å^2
Mosaicity	0.136°	0.218°
Refinement statistics		
R_{work} / R_{free}	16.2% / 19.8%	24.6% / 26.3%
Atoms	5,270	7,159
Macromolecules	4,844	6,984
Ligand		62
Solvent molecules	426	113
Protein residues	607	876
Mean isotropic B-factor		
Macromolecules	46.0 Å^2	120.6 Å^2
Ligand		91.0 Å^2
Solvent molecules	46.0 Å^2	73.1 Å^2
R.m.s.d. bond length	0.005 Å	0.013 Å
R.m.s.d. bond angle	0.91°	1.58°
Ramachandran angles		
Favored / portion	98.0%	96.0%
Allowed / portion	1.7%	4.0%
Outlier / portion	0.3%	0.0%

Table S 2: Classification of the KS domain according to the precursor specificity and the residues localized within 15Å around the reactive cysteine [136]. Amino acid numbering corresponds to Bb-KS2.Bb-KS2 belongs to the clade VIII receiving a β-hydroxyacyl precursor as substrate for decarboxylative condensation.

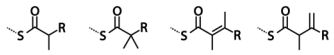
BbKS2	2471	2495	2535	2540	2557	2600	2601	2637	2641	2673	2674	2676	2747	2748	
CLADE VIII	A	A	L	M	F	T	V	L	I	E	S	A	A	G	
CLADE I	A	A	M	M	F	T	A	L	I	A	A/T	A	F	S	
CLADE II	H	A	M	M	L	T	A	L	V/I	F	A	S	M	S	
CLADE III	A	A/Q/M	L	Q	H	A	A	V	A	E	S	S	A/F	G	KS*, mixed substrates
CLADE IV	S	A	A/L/F	M	Y	T	A	L	I	E	P	A	F	G	
CLADE V	S	M	L	F	Y	T	V	L	I	E	S	A	A	G	
CLADE VI	V	A	F	M	Y	T	A	L	I	E	L	A	F	G	 and related starters
CLADE VII	A	A	F	M	Y	T	Y	K	I	E	A	A	F	G	
CLADE VIII	A	A	L	M	F	T	V	L	I	E	S	A	A	G	
CLADE IX	S	M	L	M	F	T	V	L	I	E	A	A	A	G	
CLADE X	S	A	L	L	Y/W	G	A/V	L	I	E	A	S	A	G	KS*, 
CLADE XI	N	V	Y	M	Y	T	V	L	I	E	S	A	A	S/T	After clade XIV module
CLADE XII	D	A	L/M	A/M	Y	T	V	L	I	E	G/S	A	A	G	
CLADE XIII	L	A	L	F	Y	T	V	L	V/I	E	A	A	A	G	
CLADE XIV	G	A	F	L	V	A/S	S/T	V	L	L	C	E	D	G	KS*+DH, 
CLADE XV	G	A	L	A	F	T	T	I	I	L	S	A	F	G	 generated by DH-less module
CLADE XVI	Q	N	H	N	M	Y	A	L	I	D	T	A	L	G	

Table S 3: Classification and characteristics of ketosynthase from trans-AT PKS. The table lists the *trans*-AT PKS from *B. brevis* cluster and nineteen other *trans*-AT PKS [37]. Uniprot codes are in parentheses. Inactive KS domains are indicated by an asterix. The clade of the KS domain has been determined according to the precursor and the motif of residues localized within 15Å around the catalytic cysteine [136]. Differences in the motif are written in small letter. The amino acid preceding the catalytic cysteine is mentioned, as it appears to affect the size of the catalytic pocket [105]. All KS domain harbors the catalytic triad [87] unless otherwise mentioned by the single letter code of the corresponding amino acid. Inactive KS domains present an amino acid modification for one of the two catalytic histidines. However, the EtnI-KS4 annotated as inactive [37] harbors the catalytic triad Cys-His-His and the VirA-KS3 annotated as active [37] lacks the full KSNIGH motif. The presence of the motif of the β-sheet and the α-helix extensions observed in Bb-KS2 are indicated by a tick (✓). Variation in the motif are mentioned. All KS domains harbor a similar linker domain as Bb-KS2-LD unless otherwise mentioned. Nineteen linkers out of 236 present an inserted ACP [136] localized after the corresponding first α-helix of the Bb-LD (aa. 2786-2803). Those ACPs lack the catalytic serine. KS domains of split module have linker shorter than 90 amino acids. ChiB-KS1, KirAVI-KS1 and KirAVI-KS2 are followed by an AT domain and present a linker similar to the KS-LD-AT-LD found in the *cis*-AT PKS DEBS. Certain linkers present a shorter linker connecting directly to the following domain. The linker annotated by a tick (✓) present at the same position as the docking domain of Bb-LD, similar residues involved in crystal contact. A, Adenylation domain, AL, Acyl-CoA Ligase, ACP, Acyl-Carrier Protein, B, Branching domain, C, Condensation domain, CY, Cyclisation, DH, Dehydratase, ECH, Enoyl-CoA Dehydratase, ER, Enoyl Reductase, F, Formylation, FeeH, FeeH-like domain, FkbH, FkbH-like domain, GNAT, Acetyl-loading AT of the GCN5-related N-acetyl transferase superfamily, KR, Ketoreductase, KS, Ketosynthase, MT, Methyltransferase, OMT, O-methyltransferase, OX, Oxydation domain, OXY, oxygenase, PS, Pyran Synthase, TE, Thioesterase.

Organism/Pathway	PKS	Module organization			KS domain classification		Amino acid preceding the catalytic cystein	Catalytic triad Cys-His-His		Extensions of the KS domain		Linker domain	
					Residues within 15Å of the reactive cysteine.	Clade		All KS harbor the catalytic cysteine		Presence of PxxxxxGxxxxxP motif of the β-sheet extension	Presence of LxxAxxD motif of the helical extension	Characteristics of the linker domain	Docking domain classification
								Absence of the catalytic histidine from EA**H**GTG motif	Absence of the catalytic histidine of KSNIG**H**T motif				
Brevibacillus brevis	(C0ZGR0)	AT-AT-ER											
	(C0ZGQ2)	C-A-MT-ACP-MT											
	(C0ZGQ3)	MT-ACP-	KS1*	-MT-ACP-ACP-	SMLlfmtivESsgG	IX	M	S			YxxAxxD		
			KS2	-ACP	dAMMiTvLvtAAFS	I	A				✓		I
	(C0ZGQ4)	ACP-	KS1	-ACP-DH-KR-ACP-	NVYMYSVLIESAAS	XI	V			SxxxxxGxxxxxP	✓	Insertion of an inactive ACP	I
			KS2	-DH-KR-ACP	sALMFTVLIESAAG	VII	A			✓	✓		I
	(C0ZGQ5)	C-A-ACP-	KS1	-KR-ACP-	QNyNMYALIDTALG	XVI	N				AxxAxxD		II
			KS2 (Bb-KS2)	-DH-KR-ACP-ACP-	SMLMFTVLIEAAAG	VIII	M			✓	✓		I
			KS3	DH-KR-ACP-	SMLFYTVLIESAAG	IX	M			PxxxGxxxP	✓		I
			KS4	-KR-ACP-	SMLFYnVLIESAAG	IX	M			PxxxGxxxP	✓		I
			KS5		GAFilATVLLCEDG	XIV	A				VxxAxxD	Short linker, split module followed by C0ZGQ4	
	(C0ZGQ6)	MT-ACP-	KS1	-DH-ACP-	tALMFTVLIESAAG	VII	A			✓	✓		I
			KS2	-DH-KR-ACP-ACP-	SMLFYTVLIESAAG	IX	M			PxxxxxGxxxxxA	✓		I
			KS3*	-ACP	SALmYsvLIEASAG	X	A	A		✓	✓		II
	(C0ZGQ7)	ACP-ACP-	KS1	-MT-ACP-	EASVMTALITMSSN	No corresponding motif	A				VxxAxxD		II
			KS2	-DH-KR	MAAMFTALIQIAAS	No corresponding motif	A				LxxAxxC		II
Xanthomonas albilineans Albicidin	AlbI (Q70C47)	AL-ACP-	KS1	-KR-MT-ACP-ACP-	hAsMfrALIEaAFG	VI	A				✓		IV
			KS2	-ACP-C-A-ACP-C-?-ACP-C-A-ACP-C	rALMFssLvESAAG	VII	A				✓		I
	AlbIV (Q70C44)	A-FeeH-ACP									✓		
	AlbIX (Q70C52)	A-ACP-C-A-ACP-TE									✓		
Candidatus Endobugula sertula (*Bugula neritina* bacterial symbiont) Bryostatin	BryP (A2CLL7)	AT-AT											
	BryA (A2CLL5)	DH-KR-FkbH-ACP-	KS1	-KR-ACP-	dAyMtTVLIESAAG	VII	A			✓	✓		II
			KS2	-KR-ACP-	HtlMmTALIFASiS	II	T				✓		II
			KS3	-OMT-ACP	SMLMFTVLIEAAAG	VIII	M			✓	✓		I
	BryB (A2CLL1)	ER-	KS1	-MT-ACP-	nAlMFTALIgvAFS	I	A				MxxAxxD		I
			KS2	-DH-KR-ACP-	AAlMiTALItmAFS	I	A				✓		II
			KS3	-KR-ACP-	SMLiYnVLIESAAG	IX	M				✓		I
			KS4	-OMT-ACP	HAMMiTALIFASMS	II	A				✓		II
	BryC (A2CLL2)	ER-	KS1	-DH-PS-KR-ACP-	nAlMFmALIgvAvS	I	A				MxxAxxD		II
			KS2	-MT-ACP-	fAgMYTaLIESAiG	XII	A				✓		I
			KS3	-KR-ACP-	AAlMiTALItmAFS	I	A				✓		II
			KS4*	-ACP	SmmvfGVmIEASAG	X	M	A			LxxAxxG		
	BryD (A2CLL4)	DH-KR-ACP-	KS1	-KR-ACP-	HAMimTALIFASiS	II	A				✓		II
			KS2*	-ACP-C	SmLifmgLIEsSAG	X	M	A		LxxxxxGxxxxxP	✓		II
	BryX (A2CLL3)	GNAT-MT-ACP-	KS1	-DH-ACP-	MAYMYTVLVEASGG	No corresponding motif	A				✓		I
			KS2*	-ACP-GNAT-TE	SMIlYSVLIESAAG	IX	M		N		✓		I
			KS3		TASLTTALIFSAAT	No corresponding motif	A				✓		I

Organism/Pathway	PKS	Module organization			KS domain classification		Amino acid preceding the catalytic cystein	Catalytic triad Cys-His-His		Extensions of the KS domain		Linker domain	
					Residues within 15Å of the reactive cysteine.	Clade		All KS harbor the catalytic cysteine		Presence of PxxxxxGxxxxxP motif of the β-sheet extension	Presence of LxxAxxD motif of the helical extension	Characteristics of the linker domain	Docking domain classification
								Absence of the catalytic histidine from EA**H**GTG motif	Absence of the catalytic histidine of KSNIG**H**T motif				
Sorangium cellulosum (strain So ce56)	ChiA (A9EWA3)	AT-ER											
Chivosazol	ChiB (A9EWA0)	ACP-	KS1	-AT-AT-KR-ACP-	SAmvYTALvEaAyG	IV	A				✓	KS followed by AT. *cis*-AT-PKS-like linker	
			KS2	-ACP-KR-MT-ACP-	SALMYTtLIEtAFG	IV	A			AxxxxxGxxxxxP	✓	Insertion of an inactive ACP	II
			KS3	-DH-KR-MT-ACP	dAwMFsALIATAig	I	A				✓		II
	ChiC (A9EW96)		KS1	-KR-ACP-ACP-	AAlMFTALIlAAFS	I	A				✓		II
			KS2	-DH-ACP-ACP-KR-	GALALTTIILSAFG	XV	A				✓	Short linker followed by DH domain	
			KS3	-DH-KR-ACP-	SMLlYTVLIESAAG	IX	M				✓		II
			KS4	-KR-ACP-	sALMFTVLIESAAG	VII	A				✓	Insertion of an inactive ACP	
			KS5	-KR-MT-ACP-	nAAifTALvEPAFG	IV	A				✓		II
			KS6*	-ACP	SqLafmALaEgaAG	X	Q	Q		PxxxxxGxxxxxT	✓		II
	ChiD (A9EW91)	CY-CY-A-ACP-ACP-OX-	KS1*	-ACP-	lsLmWGmisEsSAt	X	S	A			AxxAxxS		IV
			KS2	-KR-ACP-	glhmFTVLIEaAAG	IX	L				✓		II
			KS3		GALAFsTIILSAFG	XV	A				✓	Short linker, split module	
	ChiE (A9EW87)	DH-ACP-KR-	KS1	-ACP-KR-MT-ACP-	lMLlYTVLIEaAAG	IX	M			✓	✓	Insertion of an inactive ACP	II
			KS2	-DH-KR-MT-ACP	dAaMFTALIAlsmS	I	A				✓		II
	ChiF (A9EW83)		KS1	-KR-ACP-	aALmFTvlIeaAaG	XV	A				✓		
			KS2	-DH-ACP-KR-	aALAFTTIILSAFG	XV	A				✓	Short linker followed by DH domain	
			KS3	-ACP-DH-KR-MT-ACP-	iMLmYTVLIEaAAG	IX	M			PxxxxxQxxxxxP	✓	Insertion of an inactive ACP	II
			KS4*	-ACP-TE	pAfLitVLIEASAG	X	A	A			✓		II
	ChiG (A9EW76)	MT											
Bacillus amyloliquefaciens (strain FZB42)	DfnA (A7Z6E3)	AT-ER											
Difficidin	DfnD (A7Z6D8)	ACP-	KS1	-DH-KR-MT-ACP-ACP-	LALFYTVLIEAAAG	XIII	A				✓		IV
			KS2	-DH-KR-ACP-	sALMFmVLIESAAG	VII	A			✓	✓		I
			KS3		VAFMYTsLIELAFG	V	A				✓		II
	DfnE (A7Z6D7)	DH-ACP-	KS1	-KR-ACP-	SMLMFTVLIEAAAG	VIII	M			✓	✓		I
			KS2*		GAFLiATVLLCEDG	XIV	A	N			AxxAxxD	Short linker, split module	
	DfnF (A7Z6D6)	DH-ACP-	KS	-DH-KR	NVYMYTVLIESAAT	XI	V			✓	✓	Insertion of an inactive ACP	I
	DfnG (A7Z6D5)	MT-ACP-ACP-	KS1	-DH-KR-ACP-	hALMFTVLIESAAG	VII	A			✓	✓		
			KS2	-DH-KR-ACP-	HAiMmTALVFASMS	II	A				✓		II
			KS3	-KR-ACP-	VAFMYTALIELAFG	V	A				✓		I
			KS4		GAmitTTIILSAlG	XV	A				✓	Short linker, split module	
	DfnH (A7Z6D4)	DH-ACP-KR-	KS1	-KR-ACP-	tMLFYTVLIEgAAs	IX	M				VxxAxxM		III
			KS2*		GAFiegSivllsDG	XIV	A	N			✓	Short linker, split module	
	DfnI (A7Z6D3)	DH-DH-	KS	-DH-KR	gMLFYTVLIESAAG	IX	M				✓		I
	DfnJ (A7Z6D2)	MT-ACP-ER-	KS	-ACP-ACP-TE	hAsMFTALIAAAFS	I	A				✓		II

Organism/Pathway	PKS	Module organization			KS domain classification		Amino acid preceding the catalytic cystein	Catalytic triad Cys-His-His		Extensions of the KS domain		Linker domain	
					Residues within 15Å of the reactive cysteine.	Clade		All KS harbor the catalytic cysteine		Presence of PxxxxxGxxxxxP motif of the β-sheet extension	Presence of LxxAxxD motif of the helical extension	Characteristics of the linker domain	Docking domain classification
								Absence of the catalytic histidine from EA**H**GTG motif	Absence of the catalytic histidine of KSNIG**H**T motif				
Sorangium cellulosum	DszD (Q4U443)	AT-ER											
Disorazol	DszA (Q4U447)		KS1	-?-DH-KR-ACP-	yAFMYTALIEaAFG	V	A				LxxAxxA	No sequence alignment of the linker, followed by domain of unknown function	
			KS2	-KR-MT-ACP-ACP-	LALFYTVLVEAAAG	XII	A				✓		I
			KS3	-KR-ACP-	tAlMlTALIlAsmS	I	A				✓		II
			KS4	-DH	pAiMYTALIEaAvG	IV	A				✓		II
	DszB (Q4U446)	KR-ACP-	KS1	-DH-KR-ACP-	DAMAYTVLIEGAAG	XII	A			PxxxxxDxxxxxP	✓		II
			KS2	-DH-KR-ACP-	DALMYTVLIESAAG	XII	A			✓	✓	Insertion of an inactive ACP	II
			KS3	-DH-KR-ACP-	SMLFYTVLIESAAG	IX	M				✓		II
			KS4*	-ACP	atLfmtALIEgSAG	X	T	A			✓		
	DszC (Q4U445)	CY-CY-A-ACP-OX-	KS1*	-ACP-	fsLLWGmisEsSAt	X	S	A			LxxAxxR		III
			KS2*	-ACP-TE	SALLYGVLIEASAG	X	A	A		✓	✓		II
Sorangium cellulosum (strain So ce56)	EtnK (A9GJ47)	AT-AT											
Etnangien	EtnB AT (A9GJ23)												
	EtnA (A9GJ18)	ER											
	EtnD(A9GJ30)		KS1	-ACP-ACP-ACP-	SAFMYTvLIEaAyG	IV	A				✓		
			KS2	-KR-ACP-	sAgMFTALIlAAFS	I	A				✓		II
			KS3	-DH-KR-ACP-	SMLMfTVLIESAAG	IX	M			✓	✓		I
			KS4	-DH-KR	nMLFYTVLIESAAG	IX	M			✓	✓		I
	EtnE (A9GJ33)	ACP-	KS1	-DH-KR-ACP-	rMLFYSVLIESAAG	IX	M			✓	✓		I
			KS2	-ACP-ACP-	TMLFYTVLIESAAG	IX	M			✓	✓		I
			KS3	-KR	sAMMFTALItaAFS	I	A				✓		III
	EtnF (A9GJ35)	MT-ACP-	KS1	-KR-ACP-	AAaMFTALvAAAFS	I	A				✓		II
			KS2	-KR-ACP-	SMLMyTVLIEsAAG	VIII	M			LxxxxxGxxxxxP	✓		II
			KS3		HAlMLTALVFASMS	II	A				✓		
	EtnG (A9GJ37)	DH-KR-ACP-	KS1	-KR-ACP-	VAFMYTALvEvAFG	V	A				✓	Insertion of an inactive ACP	II
			KS2	-OMT-ACP-	HAlMmTALvFASMS	II	A				✓		III
			KS3	-KR-ACP-	dAMMiTALvtasFS	I	A				LxxAxxE		IV
			KS4		GALiiTTIILSAlG	XV	A				✓	Short linker, split module	
	EtnH (A9GJ40)	DH-ACP-	KS1	-KR-ACP-	YALVYYSVIESSIG	no corresponding motif	A				✓		
			KS2	-DH-KR-MT-ACP	nMLiYTVLIESAAG	IX	M				✓		II
	EtnI (A9GJ42)	ER-	KS1	-KR-ACP-	AAaMmTALIvTAFS	I	A				LxxAxxN		II
			KS2	-DH-KR-MT-ACP-	SMLMmTVLIEAAAG	VIII	M			✓	✓		I
			KS3	-KR-ACP-	SMLMmlVLIEAAAG	VIII	M			✓	✓		I
			KS4*	-ACP-TE	HAlMmTALIFASMS	II	A	LG**H**SVG	KPVIG**H**		✓		IV

Organism/Pathway	PKS	Module organization			KS domain classification		Amino acid preceding the catalytic cystein	Catalytic triad Cys-His-His		Extensions of the KS domain		Linker domain	
					Residues within 15Å of the reactive cysteine.	Clade		All KS harbor the catalytic cysteine					
								Absence of the catalytic histidine from EA**H**GTG motif	Absence of the catalytic histidine of KSNIG**H**T motif	Presence of PxxxxxGxxxxxP motif of the β-sheet extension	Presence of LxxAxxD motif of the helical extension	Characteristics of the linker domain	Docking domain classification
Streptomyces collinus Kirromycin	KirCI (B0B510)	AT-AT-ER											
	KirAI (B0B502)	?-ACP-	KS1	-KR-ACP-	SlyvYTALIEaAmG	IV	L				✓		
			KS2		nALAFTTIILSAFG	XV	A				✓	Short linker, split module	
	KirAII (B0B503)	DH-ACP-KR-	KS1	-KR-MT-ACP-	fMLFYyVLvESAAG	IX	M				✓	Insertion of an inactive ACP	IV
			KS2	-KR-ACP-	AgLMFTaLvESAyG	VII	G				FxxAxxH		
			KS3	-ACP	aAmMiTALvEaAlG	IV	A				✓		II
	KirAIII (B0B504)	C-A-ACP-	KS		VAlMYTvLIEaAmG	V	A				✓	Short linker, split module	
	KirAIV (B0B505)	DH-KR-ACP-	KS1	-DH-KR-MT-ACP-	SMLFYTVLvESsAG	IX	M				✓		IV
			KS2	-ACP-	gmLfmaVLaESAAG	VII	M				✓	Insertion of an inactive ACP	IV
			KS3	-KR-ACP-	aALMYTAiaEPAyG	IV	A				✓		IV
			KS4	-KR-ACP-	aAgvYTsLvEPAFG	IV	A				✓		
			KS5		aAmAiTTImLSAFG	XV	A				AxxAxxD	Short linker, split module	
	KirAV (B0B506)	DH-ACP-KR-	KS	-DH-KR-ACP	fMLFYSVLIESAAG	IX	M				✓	Insertion of an inactive ACP	IV
	KirAVI (B0B507)		KS1	-AT-DH-KR-ACP-	gAfgtTALvqAAvS	I	A				RxxAxxA	KS followed by AT. *cis*-AT-PKS-like linker	
			KS2	-AT-ACP	gAfgtTALIqAAiS	I	A				LxxAxxH	KS followed by AT. *cis*-AT-PKS-like linker	
	KirB (B0B509)	C-A-ACP											
Streptomyces rochei Lankacidin	LkcD (Q83X89)	AT											
	LkcA (Q83X86)	C-A-ACP-	KS		SALfYTALIEPAFG	IV	A				✓		IV
	LkcB (Q83X87)	DH									✓		
	LkcC (Q83X88)	KR-MT-ACP-ACP-	KS	-DH	AALMFTvLIESAAG	VII	A				AxxAxxD		IV
	LkcF(Q83X91)	KR-ACP-	KS1	-KR-ACP-	SMLFYTVLIESAAG	IX	M				✓		
			KS2		aAAvYTALIEPAFG	IV	A				✓		III
	LkcG (Q83X92)		KS	-ACP-TE	AAaMFmALIlvsFt	I	A				HxxAxxR		IV
Streptomyces atroolivaceus Leinamycin	LnmG (Q8GGP5)	ATER											
	LnmQ (Q8GGN5)	A									✓		
	LnmP (Q8GGN6)	ACP									✓		
	LnmI (Q8GGP3)	CY-CY-A-ACP-	KS1*	-ACP-	tsLLWGAiaEASAt	X	S	A			AxxAxxA		IV
			KS2	-KR-ACP-	aMqmfmVLvEaAAG	IX	M				✓		
			KS3		dAfAFTTIILSAlG	XV	A				✓	Short linker, split module	
	LnmJ (Q8GGP2)	DH-ACP-KR-	KS1	-ACP-	laLFYTVLIESAAG	IX	A				✓	Insertion of an inactive ACP	
			KS2	-DH-KR-ACP-MT-ACP-	tAaMtTALItmAFS	I	A				✓		II
			KS3	-DH-KR-ACP-	AALMFTVLIEgAAG	VII	A			✓	✓		
			KS4	-ACP-ACP-?-?-TE	aAFMYTALIEtsFG	V (IV)	A				✓		IV

Organism/Pathway	PKS	Module organization			KS domain classification		Amino acid preceding the catalytic cystein	Catalytic triad Cys-His-His		Extensions of the KS domain		Linker domain	
					Residues within 15Å of the reactive cysteine.	Clade		All KS harbor the catalytic cysteine					
								Absence of the catalytic histidine from EA**H**GTG motif	Absence of the catalytic histidine of KSNIG**H**T motif	Presence of PxxxxxGxxxxxP motif of the β-sheet extension	Presence of LxxAxxD motif of the helical extension	Characteristics of the linker domain	Docking domain classification
Bacillus amyloliquefaciens (strain FZB42)	MlnA (A7Z470)	AT-ER											
Macrolactin	MlnB (A7Z471)		KS1	-?-KR-ACP-	fAFMYTalIEAAFG	VI	A				LxxAxxA	No sequence alignment of the linker, followed by a domain of unknown function	
			KS2	-DH-KR-ACP-ACP-	FARMFTVLIEAAAG	VIII	A				✓		III
			KS3	-DH	nAFMYTALIEtAFG	V	A				✓		
	MlnC (A7Z472)	KR-ACP-	KS	-ACP	fMLFYTVLIESAAG	IX	M				✓		II
	MlnD (A7Z473)	KR-ACP-	KS1	-DH-KR-ACP-	SMLFYTVLIEaAAG	IX	M				VxxAxxD		
			KS2		SAtiYTALIEPsFG	IV	A				✓	Insertion of an inactive ACP	II
	MlnE (A7Z474)	KR-ACP-	KS1	-KR-ACP-	SMLFYTVLIEaAAG	IX	M				✓		III
			KS2		GALAFTTIImSAFG	XV	A				LxxAxxE	Short linker, split module	
	MlnF (A7Z475)	DH-ACP-ACP-KR-	KS	-DH	vMLFYsiLIESAAG	IX	M				✓		IV
	MlnG (A7Z476)	KR-ACP-	KS1	-KR-ACP-	pMLfyTVLIEAAAG	VIII	M			✓	AxxAxxD	Insertion of an inactive ACP	II
			KS2		sALAFTTIILSAFG	XV	A				✓	Short linker, split module	
	MlnH (A7Z477)	DH-ACP-ACP-KR-TE											
Pseudomonas fluorescens	MmpC (Q8RL73)	AT-AT-ER											
Mupirocin	MmpD (Q8RL72)		KS1	-DH-KR-MT-ACP-	AAFMYTYKIEAAFG	VI	A				✓		II
			KS2	-DH-KR-ACP-	AAaMFTALIATAFS	I	A				✓		III
			KS3	-DH-KR-MT-ACP-	SMLFYTVLIESAAG	IX	M			RxxxxxGxxxxxP	✓		I
			KS4	-KR-ACP	AAaMFTALIvtAFS	I	A				✓		II
	MmpA (Q8RL76)		KS1*	-ACP-	AMLQyAAVAhaSAG	III	M	Q			✓		IV
			KS2	-KR-ACP-	SafvFTVLIEAAAG	VIII	A				✓		II
			KS3	-ACP-ACP	SMLfyaVLIEsAAG	VIII	M				✓		II
	MmpB (Q8RL74)		KS	-DH-KR-ACP-ACP-ACP-TE	lMhFYTVLIESAAG	IX	M				✓		
Symbiont bacterium of *Theonella swinhoei* Onnamide	OnnB (Q5MP07)	GNAT-ACP-	KS1	-KR-MT-ACP-	AAyMYTvlIEAAsG	VI	A			✓	✓		
			KS2	-ECH-ECH-ACP-ACP-ACP-	SMLMFTVLIEsAAG	VIII	M				✓		I
			KS3	-ACP	vAMMFTALIATAFS	I	A				YxxAxxD		II
	OnnI (Q5MP00)		KS1*	-ACP-C-A-ACP-	hQLQHAAlgESSiG	III	Q	Q			✓		II
			KS2	-KR-ACP-	QNHNlYALIDTAiG	XVI	N				AxxAxxD		I
			KS3	-KR-MT-ACP-ACP	SMLMFTVLIEAAAG	VIII	M			✓	✓		

Organism/Pathway	PKS	Module organization			KS domain classification		Amino acid preceding the catalytic cystein	Catalytic triad Cys-His-His		Extensions of the KS domain		Linker domain	
					Residues within 15Å of the reactive cysteine.	Clade		All KS harbor the catalytic cysteine		Presence of PxxxxxGxxxxxP motif of the β-sheet extension	Presence of LxxAxxD motif of the helical extension	Characteristics of the linker domain	Docking domain classification
								Absence of the catalytic histidine from EA**H**GTG motif	Absence of the catalytic histidine of KSNIG**H**T motif				
Streptomyces albus Oxazolomycin	OzmM (B2WW46)	AT-AT-ER											
	OzmO (B2WW48)	F-A-ACP											
	OzmQ (B2WW50)		KS	-ACP	daHitYtLIDTAiG	XVI	A				✓		
	OzmN (B2WW47)		KS1	-DH-KR-ACP-	DAYITAALVDAAIG	no corresponding motif	A				LxxAxxA		IV
			KS2	-DH-KR-ACP-	SMLFYTVLIESAAG	IX	M				✓		IV
			KS3	-DH-KR-MT-ACP	nALmYTVLIEGAAG	XII	A			✓	✓		II
	OzmH (B2WW42)		KS1	-KR-MT-ACP-C-A-ACP-	AALMFTVLIESAgG	VII	A			✓	✓	Insertion of an inactive ACP	
			KS2	-DH-KR-ACP-	aALMYTsLIEaAFG	IV	A				HxxAxxD		II
			KS3	-DH-ACP-MT-KR-ACP-	SMLFYsVLvESAAG	IX	M				RxxAxxD		IV
			KS4*-		GAFLfAtiiqCEDG	XIV	A	N			LxxAxxA	Short linker, inactive KS directly followed by an active KS	
			KS5	-KR-ACP	NlYMYsVLvESAAT	XI	L				✓		
	OzmJ (B2WW43)		KS1	-ACP-	SALATTALVGAAHS	no corresponding motif	A				✓	Insertion of an inactive ACP	I
			KS2*	-DH-ER-KR-ACP-OMT	SALLITSIVEAAGG	no corresponding motif	A	T			✓		IV
	OzmK (B2WW44)	ACP-	KS	-MT-ACP	DAMTMTSLIMGAYS	no corresponding motif	A				✓		IV
	OzmL (B2WW45)	C-A-MT-ACP-C											
Symbiont bacterium of *Paederus fuscipes* Pederin	PedB (Q6VT99)	ER											
	PedC (Q6VT98)	AT											
	PedD (Q6VT97)	AT											
	PedI (Q6TAB6)	GNAT-ACP-	KS1	-KR-MT-ACP-	sAyMYTvlIEAAsG	VI	A				LxxAxxS		III
			KS2	-ECH-ECH-ACP-	SMLMFTVLIEAAAG	VIII	M			PXxxxxxGxxxxxT	✓		II
			KS3	-?-KR-ACP	sAMMFTALIATAFS	I	A				✓		III
	PedF (Q6VT95)		KS1*	-ACP-C-A-ACP-	qQLQHAAlvESSmG	III	Q	Q			✓		III
			KS2	-KR-ACP-	QNyNlYALvDTAiG	XVI	N				AxxAxxD		III
			KS3	-KR-MT-ACP-	SMLMFTVLIEAAAG	VIII	M			TxxxxxGxxxxxP	✓		I
			KS4	-DH-PS-KR-ACP-	sglMmTALvssshS	I	G				✓		II
			KS5	-KR-ACP-	fAsMYTALvEaAvG	IV	A				✓		I
			KS6	-DH	HAMMLTALIFASMS	II	A				✓		II
	PedG (Q6VT94)	OXY									✓		
	PedH (Q6VT93)	ACP-	KS1	-DH-KR-ACP-	gAyMYTALIEtAaG	V	A				✓		II
			KS2	-DH-ACP-	VAFMYTALIELAFG	V	A				LxxAxxA	Insertion of an inactive ACP	I
			KS3	-KR-ACP-	fMLFYsVLIESAAG	IX	M				✓		
			KS4*	-ACP-C-A-ACP-TE	SmLiYfVLIEASAG	X	M	A		PxxxXxxGxxxxxP	✓		II

Organism/Pathway	PKS	Module organization			KS domain classification		Amino acid preceding the catalytic cystein	Catalytic triad Cys-His-His		Extensions of the KS domain		Linker domain	
								All KS harbor the catalytic cysteine					
					Residues within 15Å of the reactive cysteine.	Clade		Absence of the catalytic histidine from EA**H**GTG motif	Absence of the catalytic histidine of KSNIG**H**T motif	Presence of PxxxxxGxxxxxP motif of the β-sheet extension	Presence of LxxAxxD motif of the helical extension	Characteristics of the linker domain	Docking domain classification
Bacillus subtilis (strain 168) Bacillaene	PksC (O34825)	AT											
	PksD (O34877)	AT											
	PksE (O34787)	AT-ER											
	PksK (P40806)	AL-ACP-C-A-ACP-	KS1	-DH-KR-ACP-ACP-	QNHNMYALIDTALG	XVI	N				AxxAxxD		I
			KS2	-KR-ACP-	iAyMYTsLIELAiG	V	A				✓		I
			KS3*		GAFiVaTVLLCEDG	XIV	A	N			VxxAxxD	Short linker, split module	
	PksL (Q05470)	DH-ACP-	KS1	-DH-KR-ACP-	NVYMYsVLIESAAT	XI	V			✓	✓		
			KS2	-ACP-ACP-	nMLFYsVLIESAAG	IX	M			✓	✓		I
			KS3	-KR-ACP-	sAMMFTAqIsAAyS	I	A				✓		I
			KS4*		GAFLVSSVLLCEDG	XIV	A	N			VxxAxxD	Short linker, split module	
	PksM (P40872)	DH-ACP-	KS1	-DH-KR-MT-ACP-	NVYMYTVLIEaAAS	XI	V			✓	✓		II
			KS2	-**KR-**ACP**-**	gALMFTVLIESAAG	VII	A				✓		I
			KS3	-ACP	qAlMLTAllFASiS	II	A				✓		II
	PksN (O31782)	C-A-ACP-	KS1	-DH-KR-ACP-	QNHNMYALIDTALG	XVI	N				AxxAxxD		II
			KS2	-DH-KR-ACP-	sAFafTtLvsLAiG	V	A				LxxAxxA		I
			KS3	-DH-KR	sAyafTtLvsLAiG	V	A				✓		I
	PksR (O31784)	MT-ACP-	KS1*	-DH-ACP-	gALLltaLIEASAG	X	A	A			✓		IV
			KS2*	-ACP-TE	vAfyfHAliElSfG	III	A				✓		II
Burkholderia rhizoxinica Rhizoxin	RhiG (A1KQR5)	AT-AT											
	RhiA (A1KQR6)	GNAT-ACP-	KS	-ACP-CY-CY	SAyfYTALIElfFG	IV	A				✓		
	RhiB (A1KQR7)	A-ACP-OX-	KS1*	-ACP-	FSLLWGALAESSAS	X	S	A			✓		IV
			KS2	-DH-KR-MT-ACP-	SiqmfTVLIESAsG	IX	I			✓	✓		I
			KS3	-DH-KR-ACP-	sALMFTVLIESAAG	VII	A			RxxxxxGxxxxxK	✓		II
			KS4	-**DH-**KR	cMLFYsVLIEaAAG	IX	M				✓		II
	RhiC (A1KQR8)	MT-ACP-	KS1	-KR-MT-ACP-	nALMFTVLIESAAG	VII	A				✓		II
			KS2	-KR-ACP-	sglMtTALVAAAhS	I	G				✓		
			KS3	-KR-MT-ACP-ACP	SMLMFTVLIEgAAG	VIII	M				✓		I
	RhiD (A1KQR9)		KS1	-DH-KR-ACP-	sglMiTALvsAAhS	I	G				✓		III
			KS2	-DH-KR-MT-ACP-	VAFMYTALIELAFG	V	A				RxxAxxD		I
			KS3*		tALymsaLIEASAG	X	A	A			✓		IV
	RhiE (A1KQS1)	DH-ACP-	KS1	-KR-ACP-	SAFaYTALIEgAFs	IV	A				✓		IV
			KS2	-DH-KR-ACP-	HAMMmTALVFASMS	II	A				✓		II
			KS3	-B-ACP	GMTFTNVLILASAG	No corresponding motif	M			✓	✓		III
	RhiF (A1KQS2)		KS1	-DH-KR-ACP-	aAAMYTALIEsAFG	IV	A				YxxAxxD		IV
			KS2*	-ACP-TE	SmLvYsmmIEASAG	X	M	A		✓	✓		II
	RhiH (A1KQS0)	Oxidoreductase											
	RhiI (A1KQR4)	Methyltransferase											

Organism/Pathway	PKS	Module organization			KS domain classification		Amino acid preceding the catalytic cystein	Catalytic triad Cys-His-His		Extensions of the KS domain		Linker domain	
					Residues within 15Å of the reactive cysteine.	Clade		All KS harbor the catalytic cysteine		Presence of PxxxxxGxxxxxP motif of the β-sheet extension	Presence of LxxAxxD motif of the helical extension	Characteristics of the linker domain	Docking domain classification
								Absence of the catalytic histidine from EA**H**GTG motif	Absence of the catalytic histidine of KSNIG**H**T motif				
Myxococcus xanthus (strain DK 1622)	TaV (Q1D5F1)	AT-AT											
Myxovirescin	Ta1 (Q1D5F8)	C-A-ACP-	KS1	-KR-ACP-	QNHNMYALvDTALG	XVI	N				AxxAxxD		IV
			KS2	-ECH-KR-ACP-	SMLMFTVLIEAAAG	VIII	M			✓	✓		II
			KS3*	-ACP-	AMLQHAAVAESSAG	III	M	Q			LxxAxxE		
			KS4	-DH-KR-ACP-	taLMFTVLIEAAAG	VIII	A				✓		III
			KS5	-ACP-	VAFMYTSLvELAFG	V	A				✓		II
			KS6	-DH-KR-ACP-	sAMMFTALIsAAFS	I	A				✓		II
			KS7	-ACP	nMLFYsVLvESAAG	IX	M			✓	✓		I
	TaO (Q1D5G0)	ER-	KS1	-DH-KR-ACP-	AAsMFTALImTAFS	I	A				✓		I
			KS2	-DH-ACP-	VAFMYTALvELAFG	V	A				✓		I
			KS3	-DH-KR-ACP-	VAFMYTALIELAiG	V	A				✓		II
			KS4		VAFMYTALvELAFG	V	A				✓		II
	TaP (Q1D5G1)	MT-ACP-	KS	-DH-KR-MT-ACP-ER-TE	AAsMFTALIATAFS	I	A				✓		II
Burkholderia thailandensis (E264)	TaiC (Q2T4N0)	AT-AT-ER											
Thailandamide	TaiA (Q2T4M8)	ACP											
	TaiD (Q2T4N1)		KS1	-KR-ACP-	AAFMmTYKIEsAFa	VI	A				✓		II
			KS2	-DH-KR-ACP-	HAMMfTALVlASMS	II	A				✓		II
			KS3	-DH-KR-MT-ACP-	SMLFYTVLIESAAG	IX	M				✓		
			KS4	-ACP	AAfffAViElSFG	III	A				✓	Short linker followed by ACP	
	TaiE (Q2T4N2)	C-A-ACP-	KS1	-DH-KR-ACP-	QNHNMYAiIDTAvG	XVI	N				VxxAxxD		IV
			KS2	-ACP-ACP	SMLFYTVLIESAAG	IX	M			PxxxXXxxGxxXxxxP	✓		II
	TaiK (Q2T4N8)		KS1	-KR-ACP-	AAMMFTALIsAAlS	I	A				✓		III
			KS2*		GAFLtSSVLLCqDG	XIV	A	N			LxxAxxH	Short linker, split module	
	TaiL (Q2T4N9)	DH-ACP-	KS1	-ACP-	NacMYmaLIEaAAS	XI	A				✓	Insertion of an inactive ACP	III
			KS2	-KR-ACP-	qALfYavLvEAaAG	X	A				✓		II
			KS3	-DH-KR-ACP-ACP-	SMLMFTVLIEAAAG	VIII	M				✓		I
			KS4	-ACP-ACP	nMLFYTVLIESAAG	IX	M				✓		
	TaiM (Q2T4P0)		KS1	-DH-KR-ACP-	HAMMfTALIsAafS	I	A				LxxAxxH		III
			KS2	-DH-KR-MT-ACP-	nMLFYTVLIESAAG	IX	M				✓		IV
			KS3		rALMFTVLIESAAG	VII	A				✓		II
	TaiN (Q2T4P1)	MT-ACP-	KS*	-OMT-ACP-TE	SMLFFTAMIEASGG	No corresponding motif	M	A			✓		II
Streptomyces virginiae	VirI (A4PHM2)	AT											
Virginiamycin	VirA (A4PHN0)		KS1	-DH-KR-MT-ACP-ACP-	FAYMFTALIEGSFG	No corresponding motif	A				✓		IV
			KS2	-DH-KR-ACP-C-A-ACP-	aMLMFTVLIEAAAG	VIII	M				✓		
			KS3	-DH-KR-ACP-	DAALTHALITPPIG	No corresponding motif	A		no KSNIGH motif		✓		
			KS4	-ACP-ACP	AALFYTVLIESAAG	IX or VII	A				✓		
	VirF (A4PHM5)		KS		sAMMiTALIsAAFS	I	A				✓		II
	VirG (A4PHM4)	KR-ACP-	KS	-ACP	HALFFTALVESAAG	VII or IX	A				✓		II
	VirH (A4PHM3)	CY-CY-A-ACP-ACP-	KS*	-ACP	tgLLWGiLLEASgG	X	G	A			✓		IV

Table S 4: *Trans*-AT PKS sequences used for sequence alignment shown on the figures S6, S7, S8 and S9.

Organism	Pathway	PKS	Uniprot number
Candidatus Endobugula sertula (*Bugula neritina* bacterial symbiont)	Bryostatin	BryA	A2CLL5
		BryB	A2CLL1
		BryC	A2CLL2
		BryD	A2CLL4
		BryX	A2CLL3
Bacillus amyloliquefaciens (strain FZB42)	Difficidin	DfnD	A7Z6D8
		DfnE	A7Z6D7
		DfnF	A7Z6D6
		DfnG	A7Z6D5
		DfnH	A7Z6D4
		DfnI	A7Z6D3
		DfnJ	A7Z6D2
Streptomyces atroolivaceus	Leinamycin	LnmI	Q8GGP3
		LnmJ	Q8GGP2
Bacillus amyloliquefaciens (strain FZB42)	Macrolactin	MlnB	A7Z471
		MlnC	A7Z472
		MlnD	A7Z473
		MlnE	A7Z474
		MlnF	A7Z475
		MlnG	A7Z476
Pseudomonas fluorescens	Mupirocin	MmpA	Q8RL76
		MmpB	Q8RL74
		MmpD	Q8RL72
Streptomyces albus	Oxazolomycin	OzmH	B2WW42
		OzmJ	B2WW43
		OzmK	B2WW44
		OzmN	B2WW47
		OzmQ	B2WW50
Bacillus subtilis (strain 168)	Bacillaene	PksK	P40806
		PksL	Q05470
		PksM	P40872
		PksN	O31782
		PksR	O31784
Burkholderia rhizoxinica	Rhizoxin	RhiA	A1KQR6
		RhiB	A1KQR7
		RhiC	A1KQR8
		RhiD	A1KQR9
		RhiE	A1KQS1
		RhiF	A1KQS2

Table S 5: Data collection and scattering-derived parameters for AT1AT2 (C0ZGR0) and AT-ER (DfnA). All data were collected at the Swiss Light Source in Villigen (Switzerland). ‡Reported for X mg/ml

	AT_1-AT_2 (C0ZGR0)	**AT-ER (DfnA)**
Data-collection parameters		
Beamline	X12SA (SLS)	X12SA (SLS)
Wavelength (Å)	1.00000	1.00000
Detector distance (m)	2.14	2.14
q range ($Å^{-1}$)	0.005-0.707	0.005-0.707
Capillary diameter (mm)	0.1	0.1
Scan lengths / step size (mm)	4.5 / 0.5	4.5 / 0.5
Positions / acquisitions	10 / 1	10 / 10
Scan repeats	20	8
Exposure time (sec)	0.5	0.04
Concentration (mg ml^{-1})	8	3
Temperature (K)	285	285
Structural Parameters		
I(0) ($Å^{-1}$) [from *P(r)*]	21.44 ± 0.02	0.55 ± 0.01
R_g (Å) [from *P(r)*]	37.62 ± 0.05	51.21 ± 0.14
I(0) ($Å^{-1}$) [from Guinier]	22.09 ± 0.04	0.55 ± 0.00
R_g (Å) [from Guinier]	38.94 ± 1.47	49.90 ±0.17
D_{max} (Å)	115	176
Porod volume estimate ($Å^3$)	114,514	219,499
Monomeric dry volume calculated from sequence ($Å^3$)	91,665 (incl. Tag)	102,916
Software employed		
Primary data reduction	Beamline software,	Beamline software
Sorting for radiation damage	In-house Software, DATCMP, DATMERGE	In-house Software, DATCMP, DATMERGE
Data processing	DATOP, DATAVER, GROM, AUTORG, DATGNOM, DATPOROD	DATOP, DATAVER, GROM, AUTORG, DATGNOM, DATPOROD
Ab *initio* modelling	DAMMIF	DAMMIF

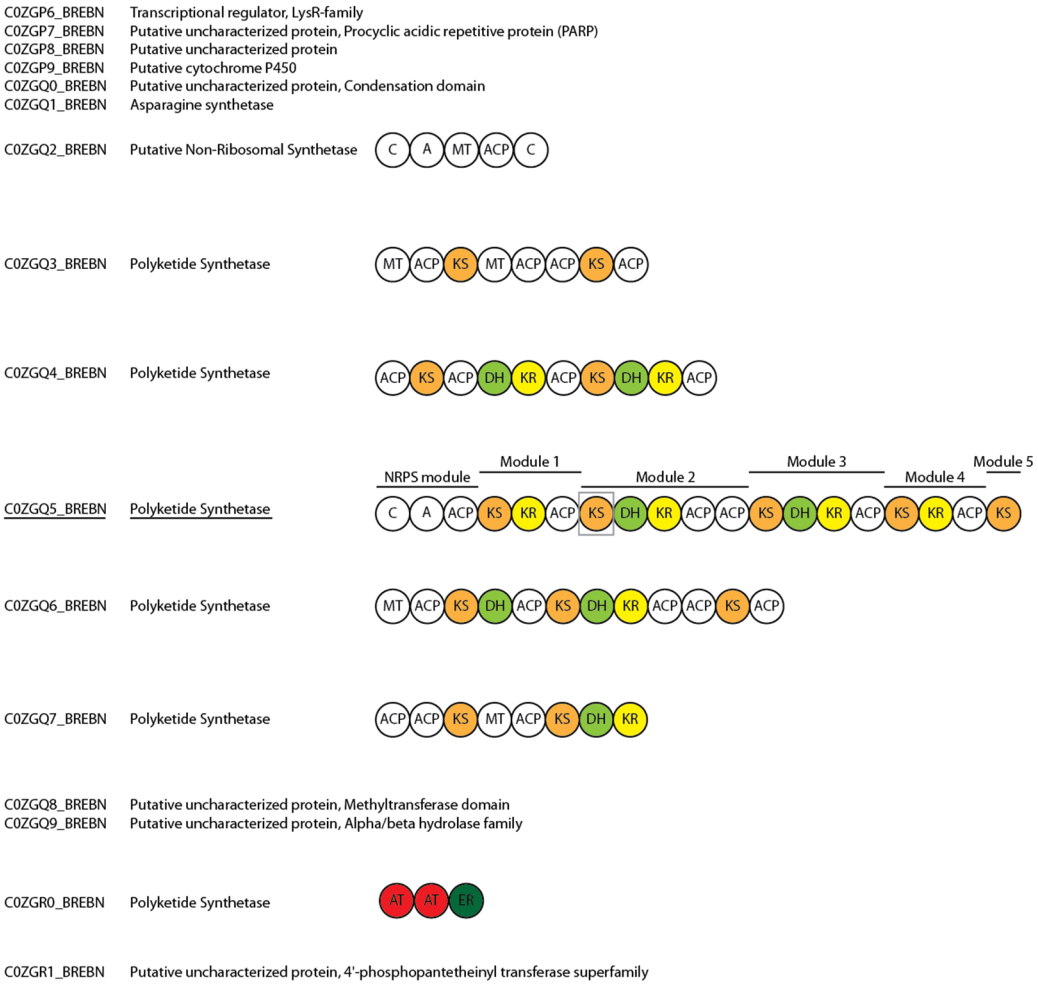


Figure S 1: Schematic view of the hybrid NRPS/PKS cluster from *Brevibacillus brevis*. The crystallized KS domain (grey frame) belongs to the module 2 of the PKS C0ZGQ5. The UniProt accession number is given on the left. AT: Acyltransferase, ACP: Acyl Carrier Protein, KS, Ketosynthase, DH, Dehydratase, ER, Enoylreductase, KR: Ketoreductase, TE, Thioesterase, C, Condensation domain, A, Adenylation domain, MT, Methyltransferase domain.

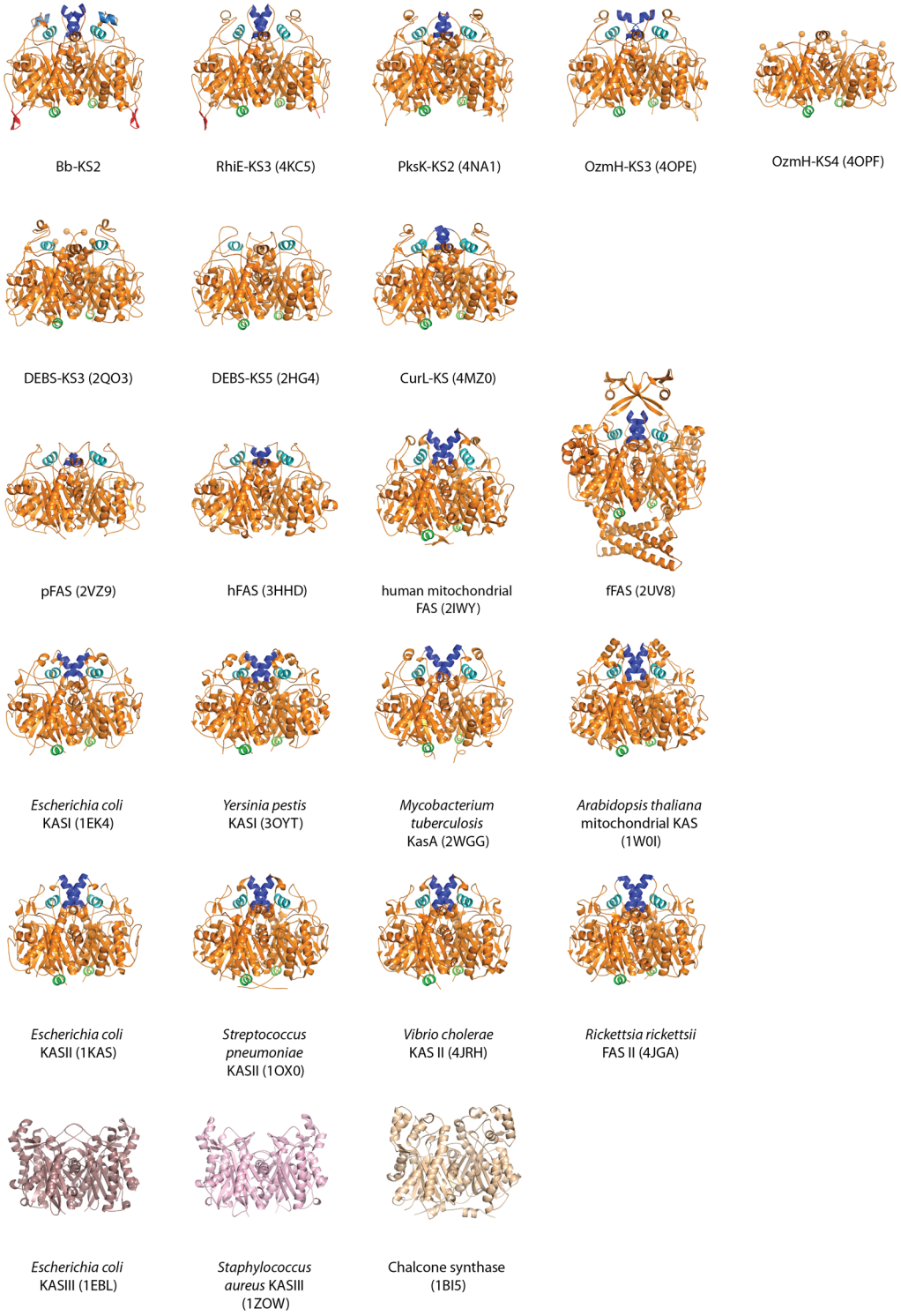


Figure S 2: Comparison of core-fold extensions in thiolase fold-like enzymes. The PDB codes are indicated in parentheses. The external β-sheet shared by the *trans*-AT PKS Bb-KS2 and RhiE-KS3 are colored in red. The pFAS and hFAS miss the green α-helix observed in PKS, fFAS and KAS. The three α-helices of the cap domain of Bb-KS2 and the equivalent α-helices of the others PKS, FAS and KAS are colored in different shade blue. The orange balls on DEBS-KS3 and OzmH-KS4 indicate absence of defined structure. The bacterial KASIII and the chalcone synthase exhibited several unrelated extension additional to the α/β thiolase core fold.

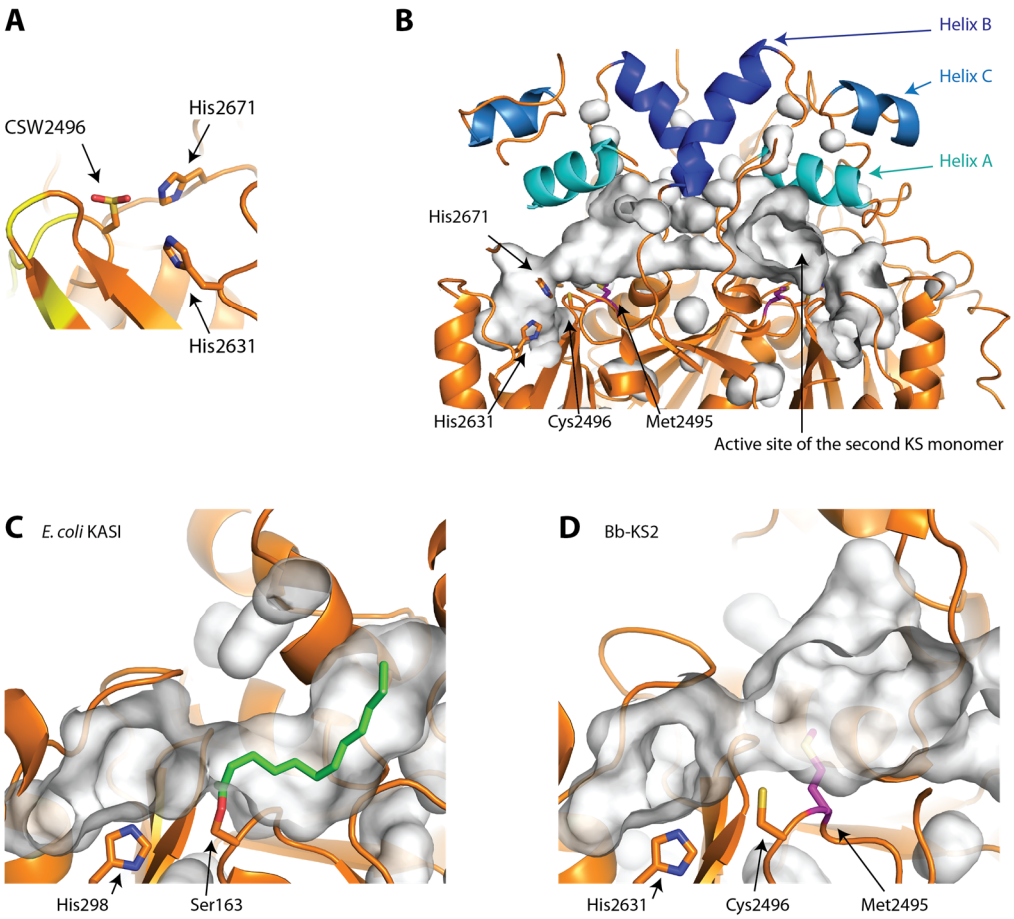


Figure S 3: Active site and substrate binding pocket of Bb-KS2. (A) Catalytic triad Cys2496, His2631 and His2671 of Bb-KS2. The electron density of the catalytic cysteine 2496 indicates oxidation to cysteine-S-dioxide. (B) Substrate binding pocket of the dimeric Bb-KS2. The three helix A, B, C of the cap domain are colored in different shades of blue. The amino acid of the catalytic triad Cys-His-His and the bulky methionine 2495 are shown as sticks and annotated for one monomer. The position of the active site of the second KS monomer is pointed by a black arrow. The binding pocket (white surface) of both KS monomer are narrowed by the Helix B. (C) Binding pocket of the *E. coli* KASI (PDB code 1EK4). The catalytic cysteine 163 has been mutated to a serine to covalently irreversibly attach a C12 fatty acid in the binding pocket. One of the histidine of the catalytic triad is shown by a stick. (D) Binding pocket of Bb-KS2. The equivalent view of the binding pocket of *E. coli* KASI is shown for Bb-KS2 with the catalytic His2631 and Cys2496 annotated. The binding pocket of Bb-KS2 is larger and the bulky methionine 2495 preceding the catalytic cysteine is colored in magenta.

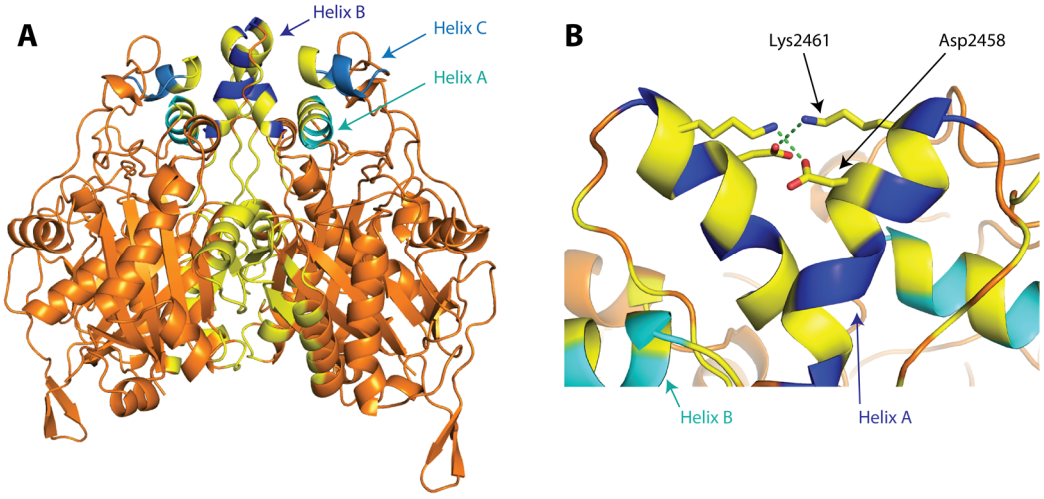


Figure S 4: Dimer interface of Bb-KS2. (A) Residues involved in the dimer interface. Residues of the dimer interface were defined by PISA [152] and are colored in yellow. The three α-helices A, B, C of the cap domain are colored in different shades of blue. (B) Salt bridges connecting the helix B of each monomer in Bb-KS2. The Lys2461 and Asp2458 from each monomer form a 2.9 Å salt bridge. Annotated helix A and helix B belongs to the same monomer.

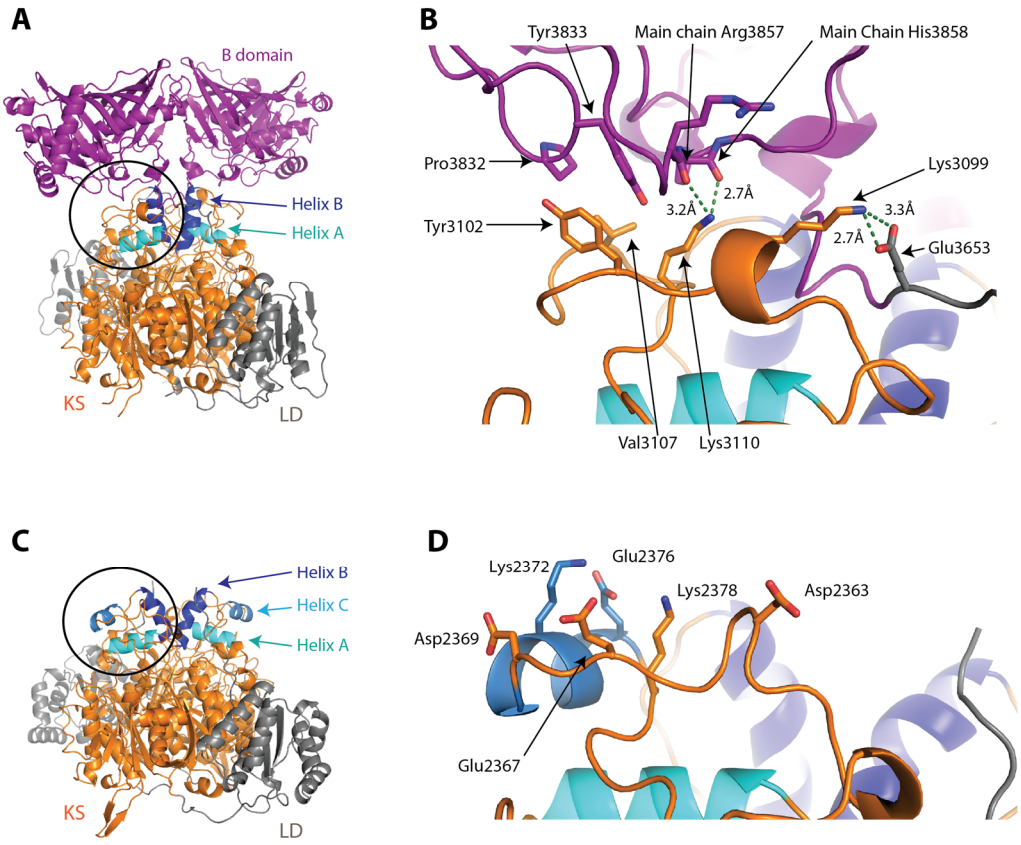


Figure S 5: Interface of KS with the downstream domain. (A) Structure of the KS-LD-B domain of the *trans*-AT PKS RhiE. The interface between the KS and the B domain is indicated by a black circle. The helix A and B from the KS dimeric interface are colored in cyan and blue respectively. (B) Close-up view of the KS-B domain interface in RhiE. A hydrophobic interface is created by the Tyr3102, Val3107, Pro3832 and Tyr3833 from the KS and B domain (sticks representation). The Lys3099 forms hydrogen bonds with the Glu3653 of the LD preceding the B domain while the Lys3110 interacts with the main chain of the B domain at the position of the amino acid Arg3857 and His3858. (C) Crystal structure of the Bb-KS2-LD. The circle indicates the equivalent position of the KS-B domain interface from RhiE. The helix A, B and C from KS dimeric interface are colored in cyan, blue and skyblue respectively. (D) Close-up view of the interface position from Bb-KS2-LD. Charged amino acids at the interface of the KS with the subsequent DH domain are annotated. The Lys 2372 and the Glu2376 belong the Helix C of the cap domain.

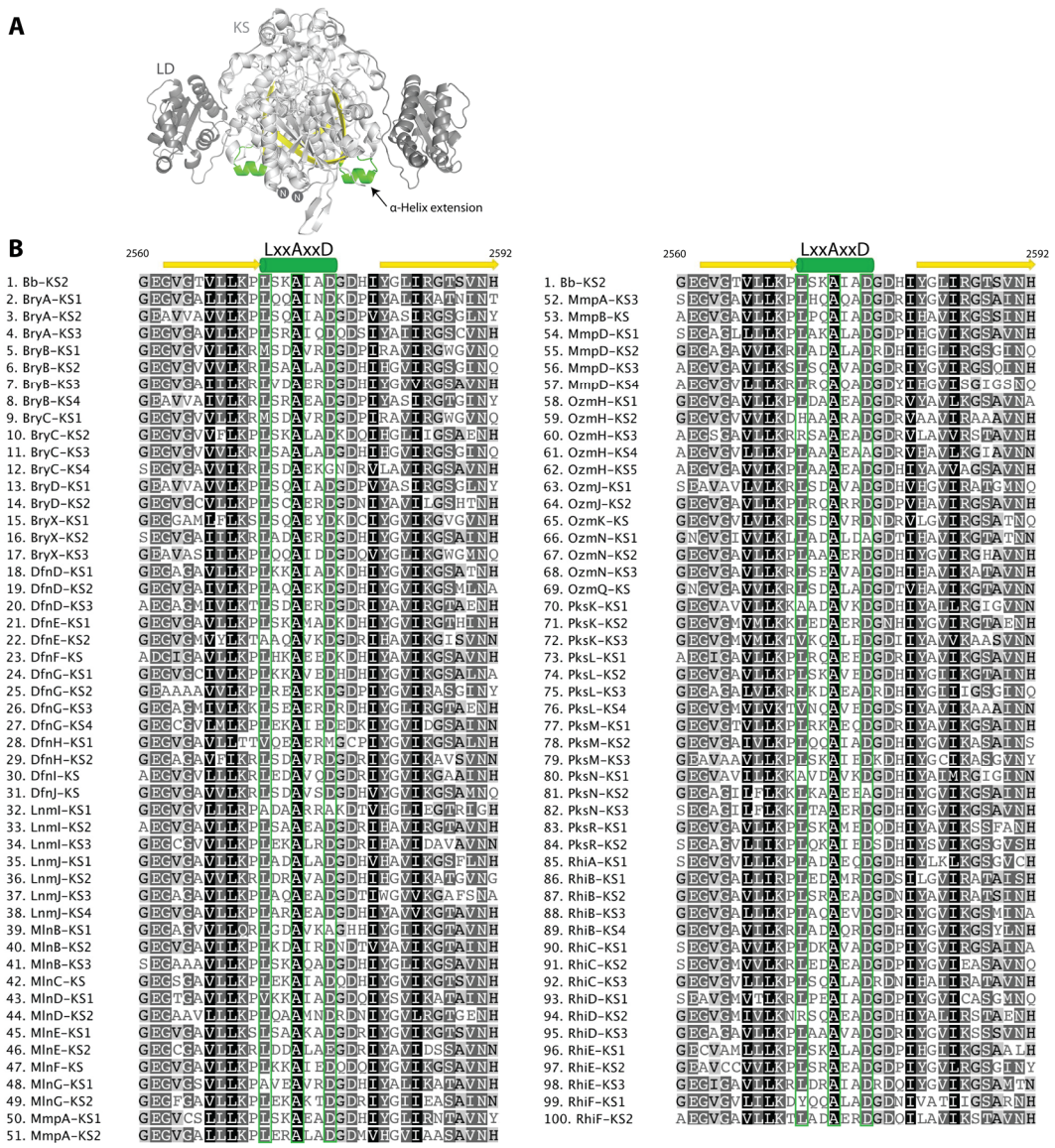


Figure S 6: Multiple sequence alignment of the short α-helical extension (aa. 2571-2577) from Bb-KS2 with KS from *trans*-AT PKS. (A) Position of the short α-helix (green) and the connected β-sheets (yellow) on Bb-KS2-LD dimer structure. (B) Sequence conservation around the α-helical extension from Bb-KS2. Sequences were aligned using CLUSTALW. The final figure was generated by Geneious [153]. Highly conserved residues are shown white in black boxes (>80% conservation). Residues with sequence conservation >60% are in grey boxes. The sequence alignment shows the conservation of the LxxAxxD motif (green frames) in the short α-helix localized at the vicinity of the N-terminus in Bb-KS2-LD structure. Amino acid numbering and position of the α-helix (green tube) and β-sheets (yellow arrow) are indicated for Bb-KS2.

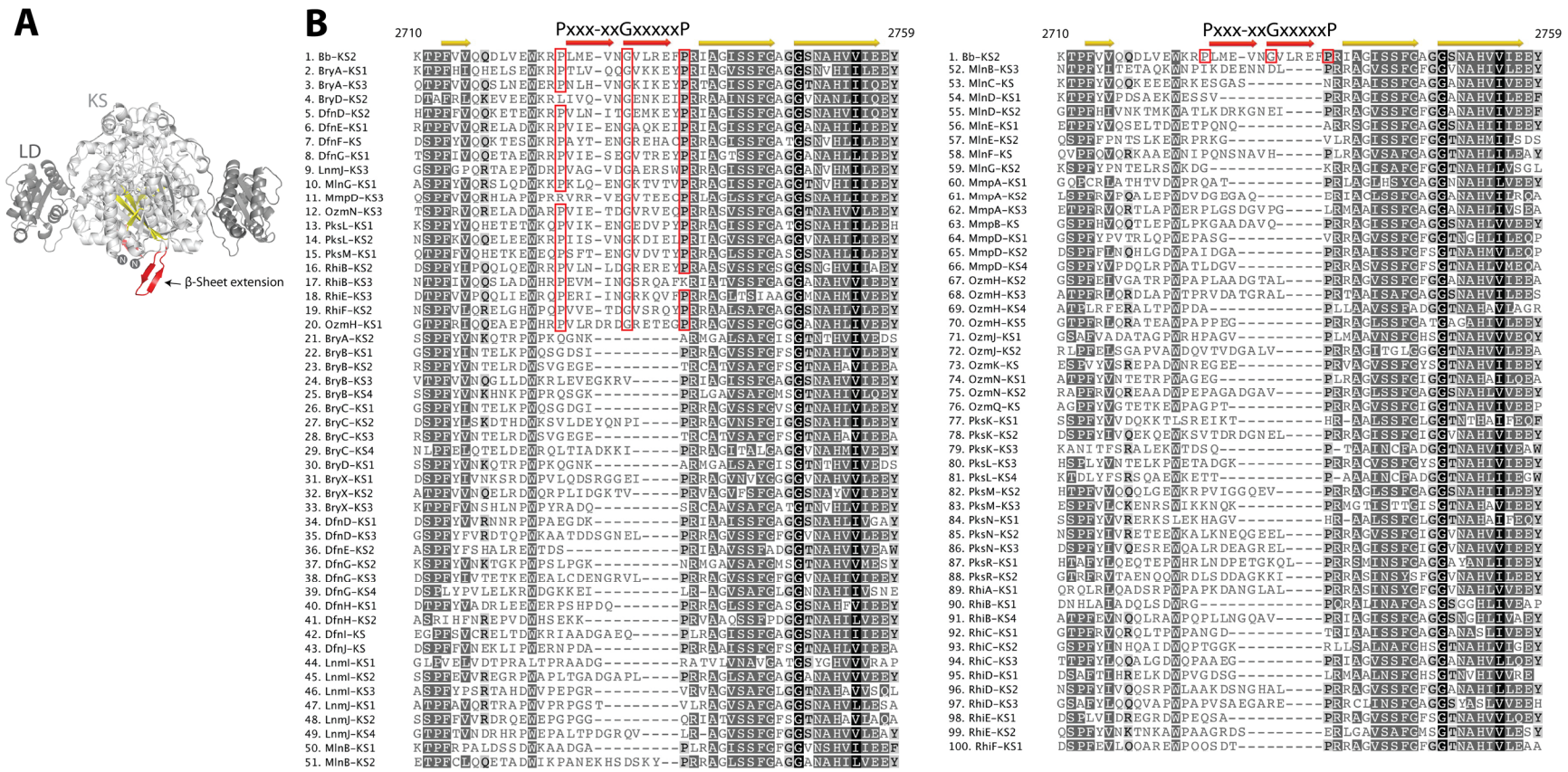


Figure S 7: Multiple sequence alignment of the β-sheet extension (aa. 2725-2737) from Bb-KS2 with several *trans*-AT PKS. Position of β-sheet extension (red) and the connected β-sheet (yellow) on Bb-KS2-LD dimer structure. (B) Sequence alignment of the region of the β-sheet extension from Bb-KS2 with several *trans*-AT PKS. Sequences were aligned using CLUSTALW. The final figure was generated using Geneious[153]. Highly conserved residues are shown white in black boxes (>80% conservation). Residues with sequence conservation >60% are in grey boxes. The β-sheet extension of Bb-KS2 is indicated on top by red arrows. The PxxxxxGxxxxxP motif is shown in red frames and conserved in 20 out 100 sequences from *trans*-AT PKS.

Figure S 8: Sequence alignment of the linker domain from *trans*-AT PKS. (A) Sequences were aligned using CLUSTALW. The final figure was generated using Geneious [153]. Highly conserved residues are shown white in black boxes (>80% conservation). Residues with sequence conservation >60% are in grey boxes. Amino acid numbering and position of the α-helix (magenta tube) and β-sheet (yellow arrow) are indicated for Bb-KS2 linker. The three α-helices of Bb-LD corresponding to the figure 4 are annotated with their number. The conserved LxxxWxxG motif at the position of the third α-helix of Bb-LD triangle is shown in magenta frames. (B) Sequence alignment of *trans*-AT PKS harboring shorter linker domains. The inactive OzmH-KS4 is directly linked to an active KS domain whereas the other linker domains are located at the C-terminal end of split-module. (C) Position of the secondary structure on Bb-LD. α-helix and β-sheet are colored in purple and yellow respectively. The three α-helices of the potential docking site for *trans*-AT are annotated by number.

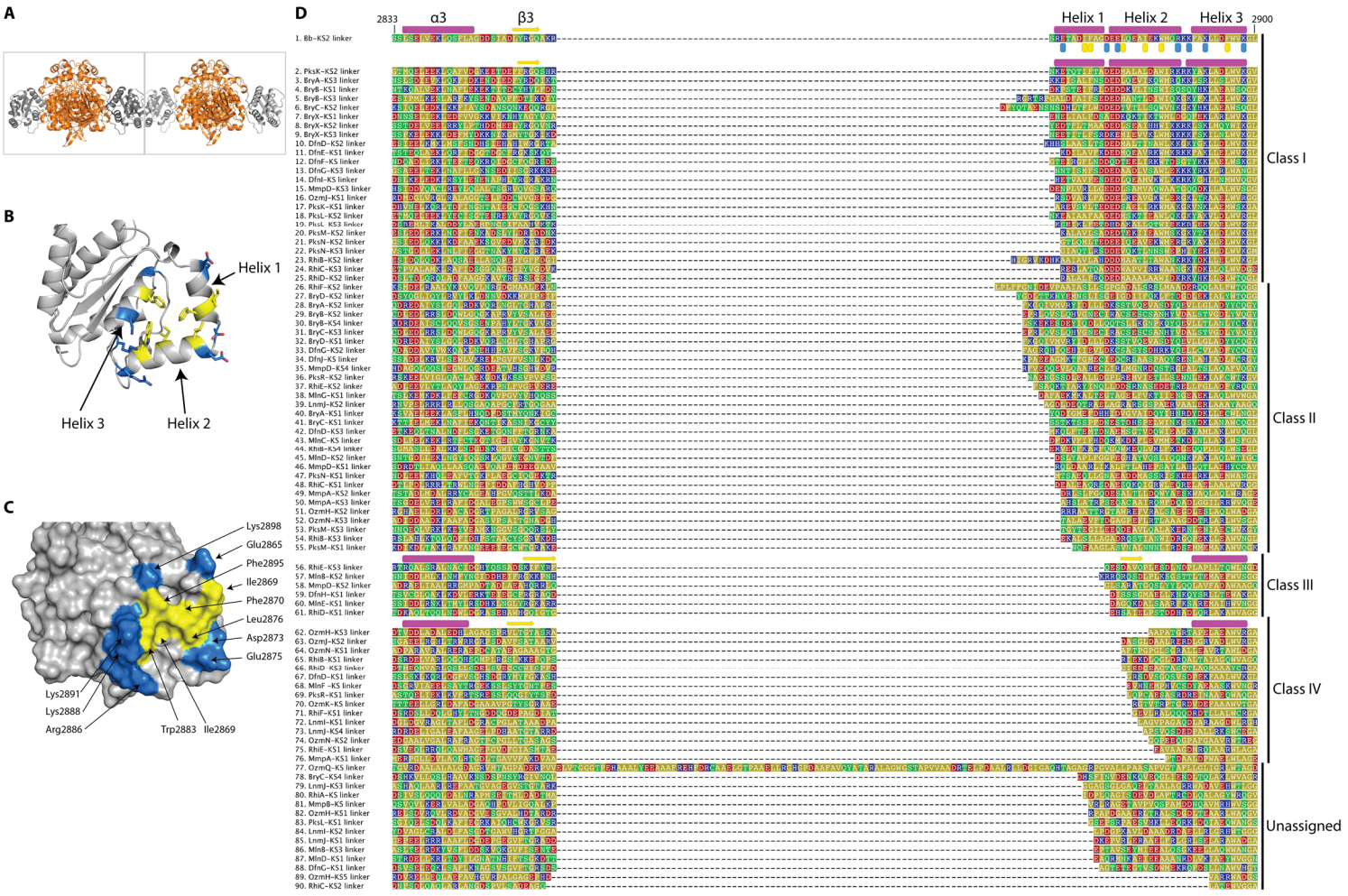


Figure S 9: Crystal contacts at the docking domain of Bb-KS2 linker. (A) Structure of two Bb-KS2-linker proteins interacting at the docking domain in the crystal. (B) Docking domain of one protein with amino acid involved in the interaction. Amino acids colored in yellow form the hydrophobic pocket. Amino acids represented by blue sticks form interaction between the two proteins in the crystal structure. (C) Surface representation of Bb-LD (grey). Residues lining a hydrophobic pocket are shown in yellow. Amino acids in blue are involved in crystal contact (D) Sequence alignment of the three α-helical triangle region from the linker domain. Sequences were aligned using CLUSTALW. The final figure was generated using Geneious[153]. Amino acids are colored by polarity. Classes of the linker domains LD are indicated on the right. The 25 first linker domain present amino acid with similar polarities at the same position as Bb-KS2 linker and PksK-KS2 linker. In Bb-KS2-LD crystal structure, proteins interact at the docking domain with hydrogen bonds between the two proteins (amino acids labeled by blue squares). A hydrophobic pocket (amino acids labelled by yellow squares) is formed at the middle of the docking domain. α-Helix and β-sheet are indicated by pink tubes and yellow arrows, respectively.

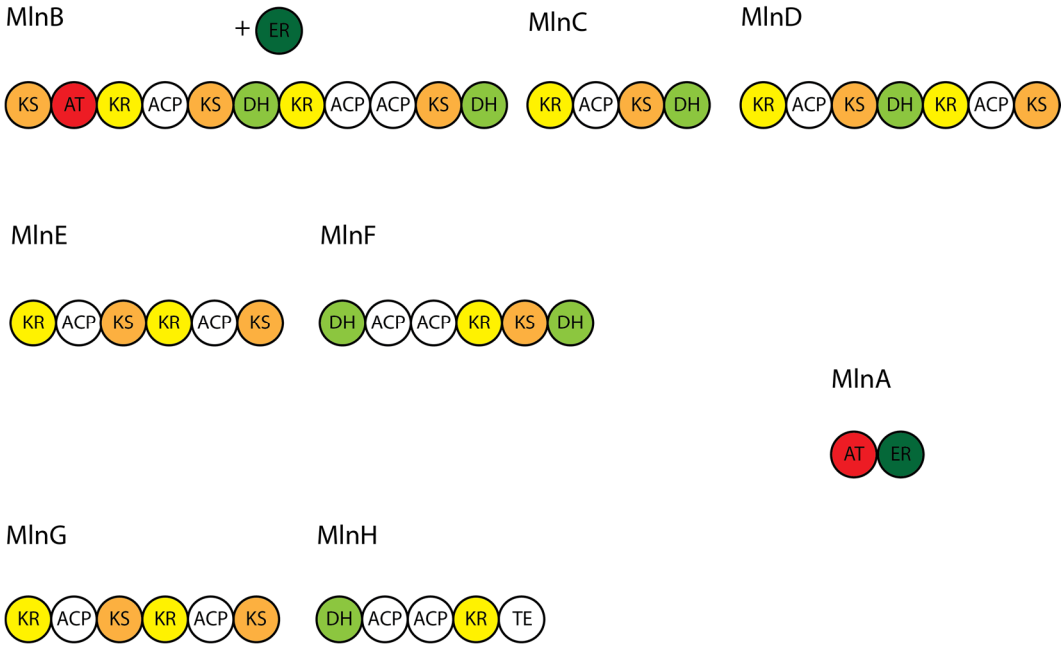


Figure S 10: Schematic view of the macrolactin cluster from *B. amyloliquefaciens* FZB42. Based on the chemical structure of Macrolactin the enoyl reductase domain of MlnA is supposed to act after second KS domain of MlnB. KS, Ketosynthase; AT, acyl-CoA transferase, KR, ketoreductase, ACP, acyl carrier protein, DH, dehydratase, KR, ketoreductase, ER, enoyl reductase, TE, thioesterase.

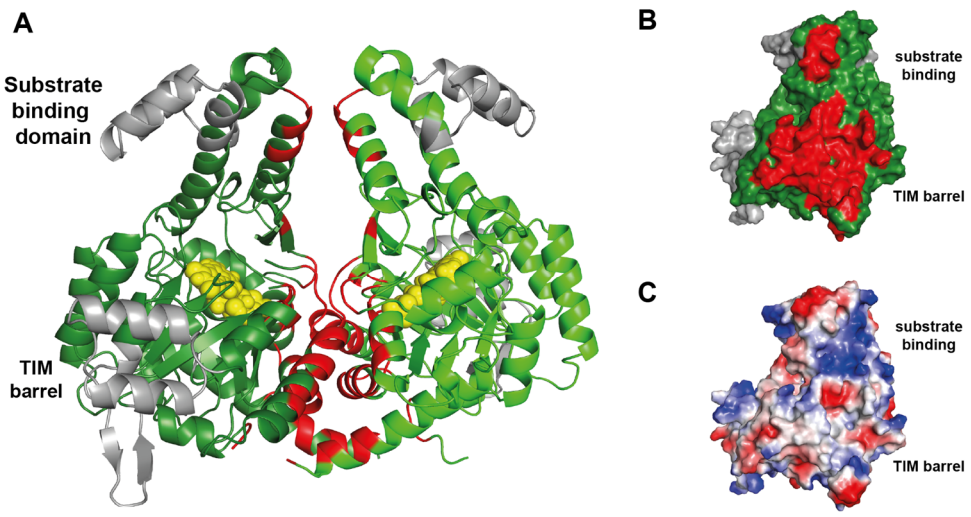


Figure S 11: Dimer Interface of MlnA. (A) Overall structure of MlnA dimer. Molecule A and B are colored in forest and green, respectively. Extra protein sequences additional to the FabK structure are shown in gray, the dimer interface is colored in red. The FMN is shown in yellow sphere representation. (B) Side view of the dimer interface of one monomer MlnA shown in surface representation. (C) Electrostatic potential of one monomer MlnA, in the same view as in (B).

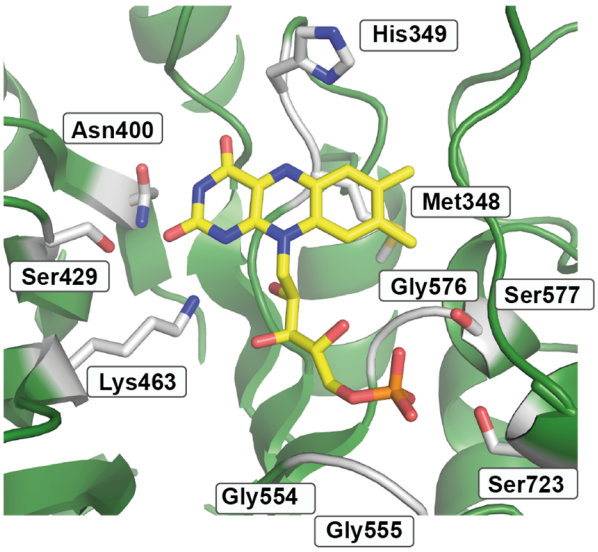


Figure S 12: Active site of MlnA. All residues are displayed in stick mode with carbon atoms of the apoprotein in grey, carbon atoms of the FMN in yellow, nitrogen in blue, oxygen in red, and phosphorus in orange.

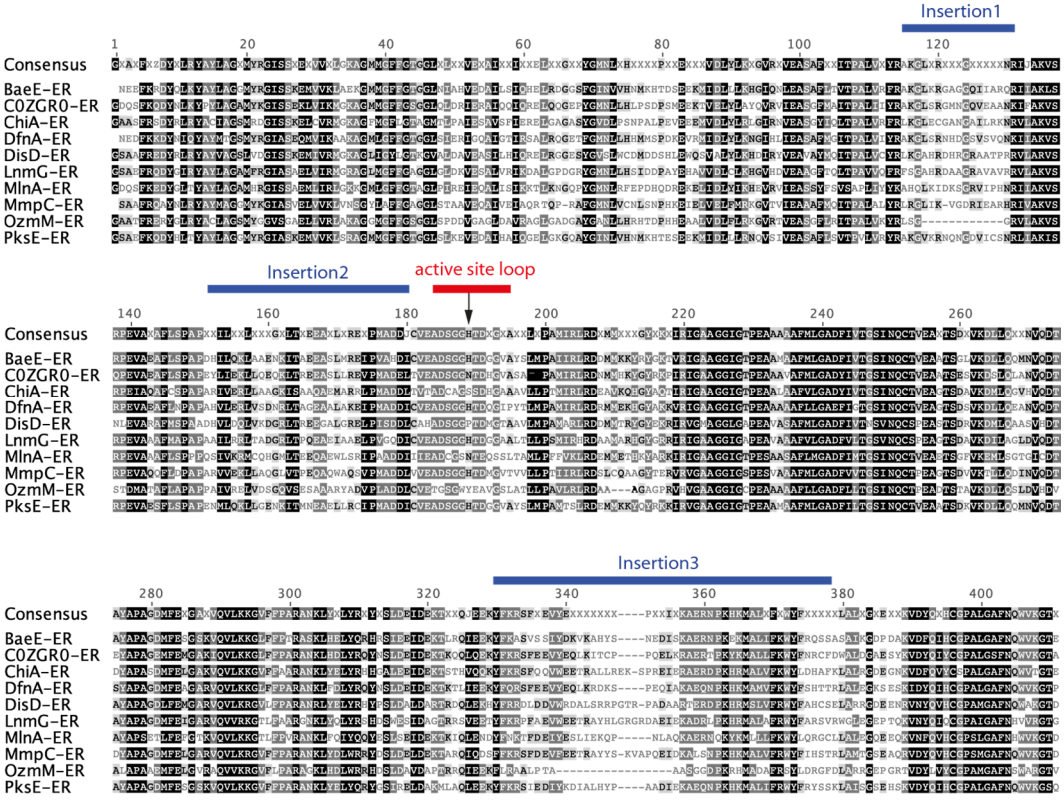


Figure S 13: Multiple sequence alignment of the AT-ER proteins from several species. Highly conserved residues are shown white in black boxes (>80% conservation). Residues with sequence conservation >60% are in grey boxes. Sequences of representative ER proteins were retrieved from the UniProt database [154] (http://uniprot.org) and aligned using CLUSTALW. The final figure was generated using Geneious [153]. Protein abbreviations and UniProt accession numbers are BaeE, *B. amyloliquefaciens* FZB42 (A7Z4Y0); C0ZGR0, *B. brevis* (C0ZGR0); ChiA., *S. cellulosum* So ce12 (A9EWA3); DfnA, *B. amyloliquefaciens* FZB42 (A7Z6E3); DisD, *S. cellulosum* So ce12 (Q4U443); LnmG, *Streptomyces sp* (Q8GGP5); MlnA, *B. amyloliquefaciens* FZB42 (A7Z470); MmpC, *P. fluorescens* (Q8RL73); PksE, *B. subtilis* 168 (O34787).

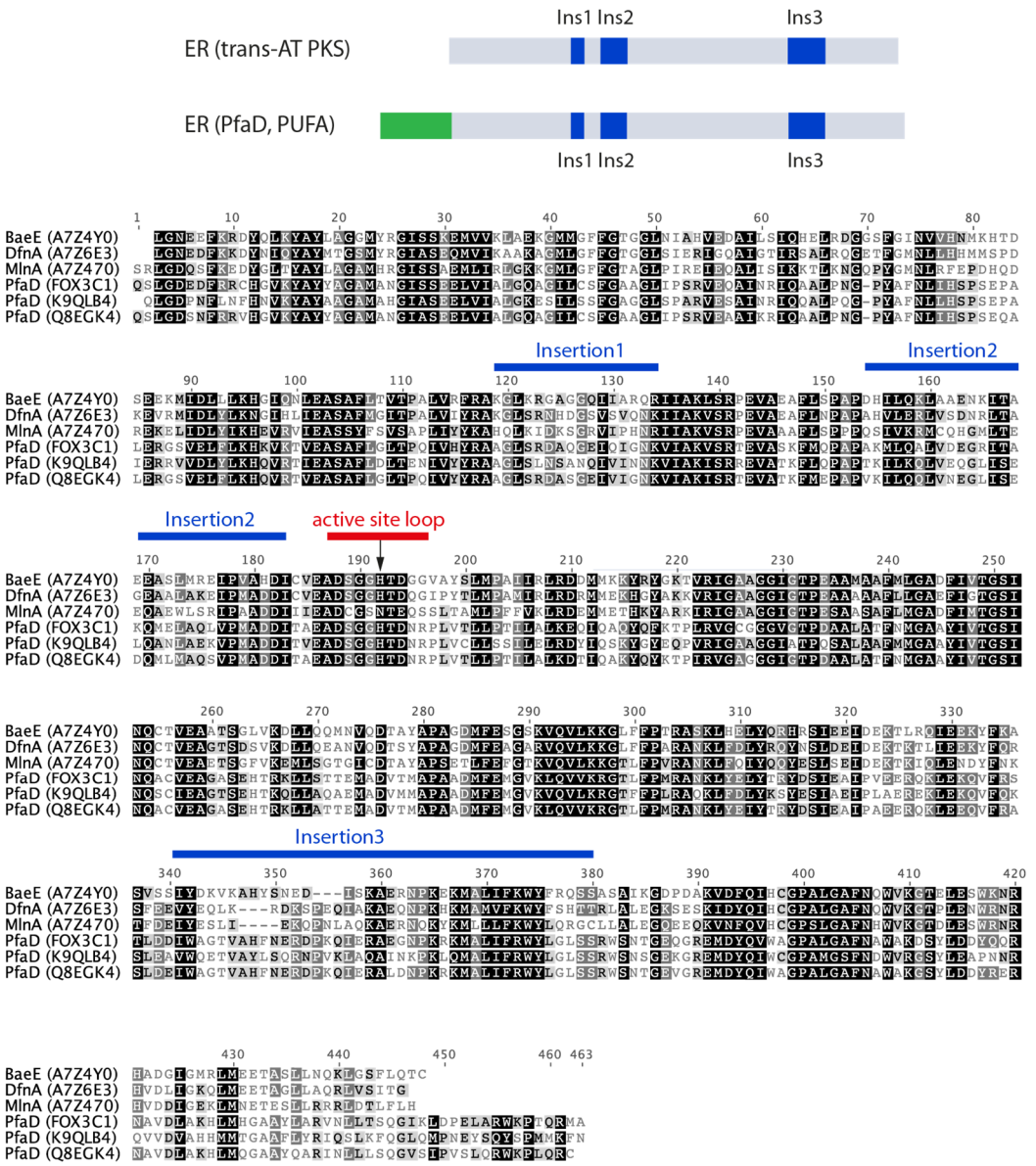


Figure S 14: Multiple sequence alignment of the ER proteins from trans-AT and PUFA biosynthesis pathways. (Top) Sequence organization of *trans*-AT ER and PfaD ER protein sequences. Insertion 1,2,3 (Ins1, 2, 3) are colored in blue. The PUFA specific N-terminal extension (~80 aa) is shown in green. Highly conserved residues are shown white in black boxes (>80% conservation). (Bottom) Sequence Alignment Residues with sequence conservation >60% are in grey boxes. Sequences of representative ER proteins were retrieved from the UniProt database [154] (http://uniprot.org) and aligned using CLUSTALW. The final figure was generated using Geneious [153]. Protein abbreviations and UniProt accession numbers are BaeE, *B. amyloliquefaciens* FZB42 (A7Z4Y0); DfnA, *B. amyloliquefaciens* FZB42 (A7Z6E3); MlnA, *B. amyloliquefaciens* FZB42 (A7Z470); PfaD, *Vibrio* sp. NSP560 (FOX3C1); PfaD, *Nostoc* sp. (strain ATCC 29411/PCC 7524 (K9QLB4); PfaD, *Shewanella oneidensis* (strain MR-1) (Q8EGK4).

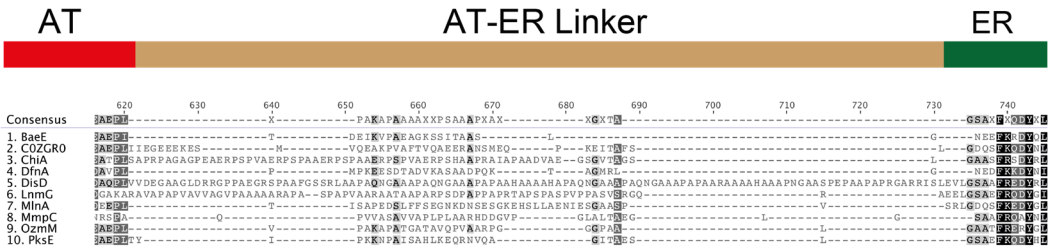


Figure S 15: Multiple sequence alignment of the AT-to-ER linker from several species. Highly conserved residues are shown white in black boxes (>80% conservation). Residues with sequence conservation >60% are in grey boxes. Sequences of the AT-ER proteins were retrieved from the UniProt database [154] (http://uniprot.org) and aligned using CLUSTALW[155]. The final figure was generated using Geneious [153]. Protein abbreviations and UniProt accession numbers are BaeE, *B. amyloliquefaciens* FZB42 (A7Z4Y0); C0ZGR0, *B. brevis* (C0ZGR0); ChiA., *S. cellulosum* So ce12 (A9EWA3); DfnA, *B. amyloliquefaciens* FZB42 (A7Z6E3); DisD, *S. cellulosum* So ce12 (Q4U443); LnmG, *Streptomyces sp* (Q8GGP5); MlnA, *B. amyloliquefaciens* FZB42 (A7Z470); MmpC, *P. fluorescens* (Q8RL73); PksE, *B. subtilis* 168 (O34787).

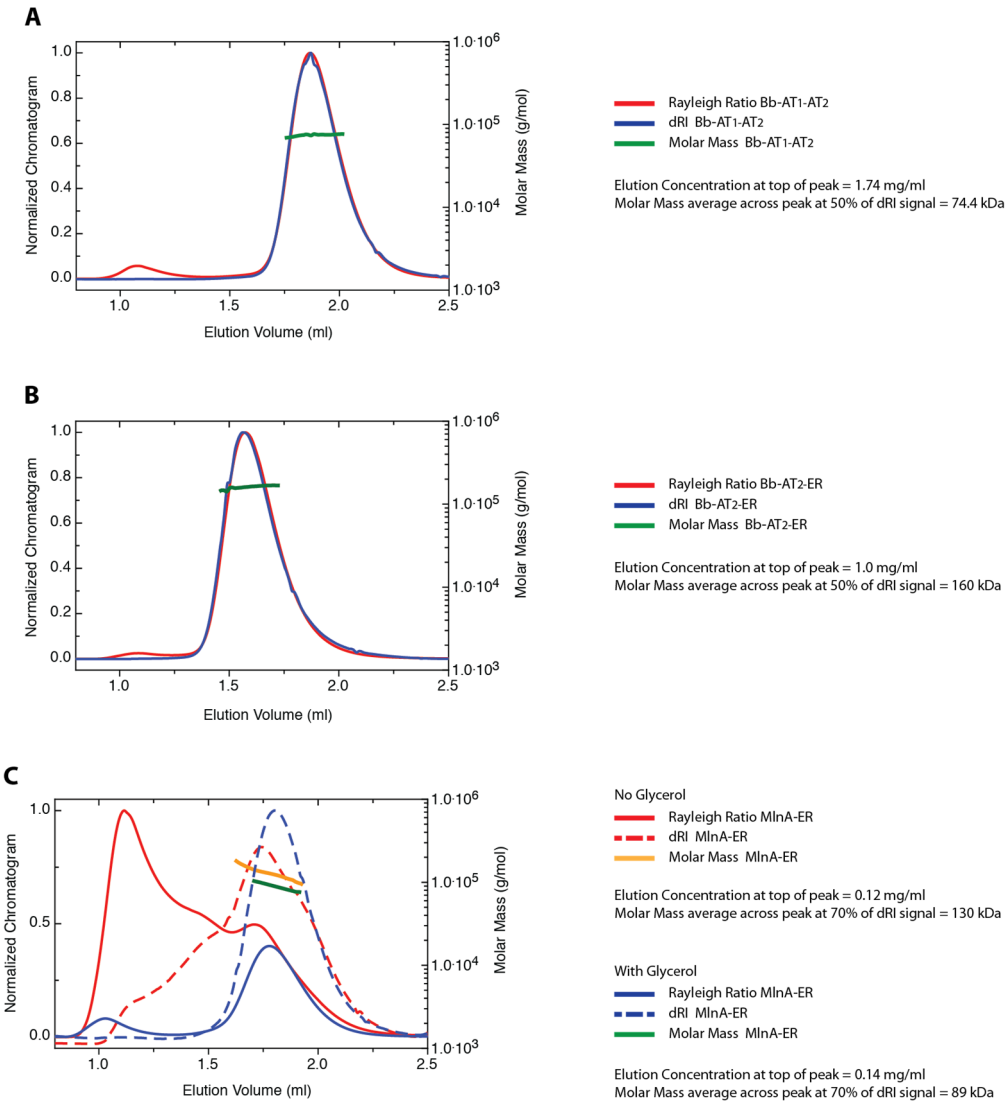


Figure S 16: Molecular weight determination using size-exclusion chromatography coupled to multi-angle light scattering (SEC-MALS). (A) SEC-MALS chromatogram of the AT_1-AT_2 protein from *B. brevis.* The N-terminal His and TEV tagged Bb-AT_1-AT_2 has a molecular weight of 75.9 kDa. (B) SEC-MALLS chromatogram of AT_2-ER from *B. brevis* (Bb-AT_2-ER). The N-terminal His and TEV tagged Bb-AT_2-ER has molecular weight of 90.1 kDa (or 180.2kDa for a dimer. (C) SEC-MALLS chromatogram of the MlnA-ER protein from *B. brevis.* The N-terminal His and TEV-cleavage-site tagged MlnA-ER has molecular weight of 55.6 kDa (or 111.2 kDa for a dimer.

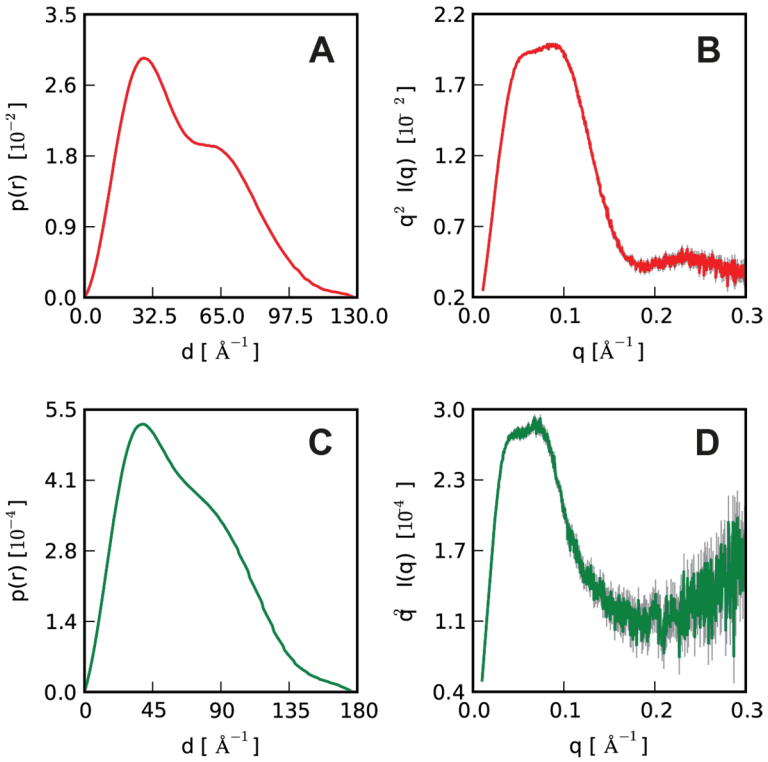


Figure S 17: Distance distribution (A, C) and Kratky analysis (B, D) of AT-AT and AT-ER didomain proteins. Distance distribution function derived of AT_1-AT_2 from C0ZGQ0 (A) and DfnA AT-ER (C) calculated with GNOM. Kratky plot of AT_1-AT_2 from C0ZGQ0 (B) and DfnA AT-ER (D).

PART II

EXPRESSION, PURIFICATION AND BIOPHYSICAL CHARACTERIZATION OF LIPIN, A KEY REGULATOR OF LIPID METABOLISM

Solène F. Martin, Timm Maier*

Biozentrum, Universität Basel, Klingelbergstrasse 50/70, 4056 Basel (Switzerland)

Correspondence should be addressed to T.M. (timm.maier@unibas.ch)
E-mail addresses: solene.martin@unibas.ch

Keywords
Lipin, PAH1, phosphatidate phosphatase, diacylglycerol, lipid metabolism, protein expression, insect cells, protein purification.

STATEMENT OF MY OWN CONTRIBUTION

I designed, performed and analysed all experiments mentioned, in discussion with other lab members. I studied the lipin domain organization using bioinformatics tool as presented in Results & Discussion and wrote the manuscript.

ABSTRACT

Lipins are key regulators of eukaryotic lipid metabolism with a unique dual function as phosphatidate phosphatase enzyme at the endoplasmic reticulum membrane and as transcriptional coactivator in the nucleus. Several functional and structural domains have been characterized for the mammalian lipin and the yeast homologue PAH1 but the three dimensional structure and the structural basis underlying its dual role have not been identified. We have selected candidate lipin homologues for large-scale production and structural studies based on sequence and homology analysis. Recombinant lipins have been screened for expression and purification from *E.coli* and insect cells. Here, we report the successful large-scale expression and purification of a recombinant variant of the *Kluyveromyces lactis* PAH1 defined by limited proteolysis. We identify monomeric and oligomeric forms of PAH1, and analyse the impact of phosphorylation and oxidation on this protein construct.

INTRODUCTION

Lipin is a bi-functionnal protein that acts as phosphatidate phosphatase (PAP) in the triacylglycerol biosynthesis pathway and as transcriptional coactivator of fatty acid oxidation genes [41].

A single isoform of lipin is observed in yeast and nematodes whereas human and murine lipin occurs in three isoforms named lipin 1, 2 and 3 [48]. Lipin 1 has been identified initially after a spontaneous gene rearrangement in laboratory mice [78], which resulted in lipodystrophy, insulin resistance, peripheral neuropathy and neonatal fatty liver [9]. This phenotype, referred as fatty liver dystrophy (fld), can be caused by a single Gly84Arg mutation in the lipin gene [43]. In human, mutations in the lipin 1 gene have been associated with rhabdomyolysis, weakness and myoglobinuria; further polymorphisms in the gene are associated with body mass index and insulin sensitivity disorders [1]. Mutations in the human lipin 2 gene induce a different phenotype called Majeed syndrome, which is characterized by osteomyelitic lesions and dyserythropoeitic anemia [1]. Gene-expression revealed tissue-specific expression profiles of lipin isoforms [50]. Lipin 1 is broadly expressed but predominantly in muscle and adipose tissues. Lipin 2 and 3 have lower expression with lipin 2 localized into the brain, liver and gastrointestinal tract and lipin 3 expressed at even lower level mainly in small intestine [1]. Different lipin isoforms within the same tissue can only partially compensate for each other [50].

Lipins contribute to the balance between phosphatidate (PA) and diacylglycerol (DAG) by catalyzing via a DxDxT motif the dephoshorylation of PA to DAG in the triacylglycerol biosynthesis pathway [42]. PA and DAG serve as building blocks for triacylglycerol and glycerolipids and as signalling molecules in various pathways controlling cell growth, cell differentiation, insulin resistance and expression of genes linked to lipid metabolism [1]. In the nucleus, lipin 1 interacts via a LxxIL motif with the peroxisome proliferator-activated receptor α (PPARα) in complex with the PPARα coactivator 1α (PGC1α) in order to activate fatty acid oxidation genes [63, 65]. Lipin 1 also activates PPARγ via a VxxLL binding motif [64] and a transcriptional activation domain (TAD) [64]. Other factors are regulated by lipin 1 such as the myocyte enhancer factor 2 (MEF2), the nuclear factor of activated T-cells c4 (NFATc4) and the sterol response element binding protein 1 (SREBP1) [1].

Lipin is regulated by several mechanisms including phosphorylation, sumoylation and protein-protein interaction. The yeast homologue PAH1 can be phosphorylated on multiple sites and by several kinases including Pho85p-Pho80p [66], Cdc28p-cyclin B [67], protein kinase A [68, 69], protein kinase C and casein kinase II [69]. Hyperphosphorylated lipin associates in the cytosol via its Serine Rich Domain (SRD) with the 14-3-3 protein, which blocks its translocation to other cellular compartments [70]. Dephosphorylation of the yeast PAH1 is mediated by the transmembrane protein phosphatase complex Nem1p-Spo7p [72] and is required for the binding of lipin via a HRVF motif to the Ser/Thr protein phosphatase 1 catalytic unit γ (PP-1cγ) [73]. Binding of lipin 1 to PP-1cγ allows its translocation to the endoplasmic reticulum (ER) membrane and the nucleus. At the ER membrane, lipin 1 preferentially binds di-anionic PA via a Polybasic domain/Nuclear localization signal (PBD/NLS) and exert its PAP activity [57]. The PBD/NLS domain [57] and the sumoylation of two lysines [74] are required for the translocation of lipin 1 to the nucleus where it acts a transcriptional coactivator.

Lipin is a key regulator in lipid homeostasis. Several functional domains have been associated with catalytic activity and different protein-protein interactions. However, the structure of lipin as a basis to understand the effect of mutations and polymorphisms inducing human diseases remains unknown. Production of stable and monodisperse lipins is a prerequisite for structural studies and has not been previously achieved. Lipins with shortest disordered regions were selected by bioinformatics analysis. Lipin constructs were screened for expression in *E. coli* and insect cells and short stable fragments were defined by limited proteolysis. Here, we report a protocol for large-scale expression and purification of the recombinant variant PAH1 from *Kluyveromyces lactis*. We characterized the oligomeric states of PAH1 by size-exclusion chromatography coupled to multi-angle light scattering (SEC-MALS) and we determined the phosphorylation status and redox resistance property from initial biophysical studies.

RESULTS & DISCUSSION

Sequence-based definition of target sequences for lipin expression

Sequence alignment and secondary structure prediction of lipin from different organisms reveal the presence of structurally conserved regions. Knowing the distribution of these regions is interesting to get an estimation of the region stability and the susceptibility for successful structural studies like x-ray crystallography, nuclear magnetic resonance (NMR), small-angle X-ray scattering (SAXS) and Electron Mycroscopy (EM). Three human lipin isoforms 1, 2, 3, three mouse isoforms 1, 2, 3, three fungal PAH1 and the *Drosophila melanogaster* lipin (Dm lipin) were studied by sequence alignment, Psipred [110] and Disopred [156]. Human lipin 1, 2, 3, mouse lipin 1, 2, 3 and *Saccharomyces cerevisiae* PAH1 (ScPAH1) are 886 ± 38 amino acids long. Dm lipin is the longest with 1089 amino acids whereas *Kluyveromyces lactis* PAH1 (KlPAH1) and *Chaetomium thermophilum* (CtPAH1) have 796 and 748 amino acids respectively. Human lipin 1, 2, 3 are closely related with their respective mouse homologue 1, 2 and 3 (89%, 90% and 80% sequence identity respectively). Lipin isoform 1 shares 51% and 49% sequence identity with lipin 2 and 3 respectively. The fungal PAH1 and the Dm lipin are more distant with only 29% and 35% sequence identity to the human lipin 1. Between the fungal PAH1, ScPAH1 has 50% sequence similarity with KlPAH1 and 37% with the CtPAH1 (Table S 1, Figure 19 and Figure S 18).

All lipin present the N-terminal N-LIP domain (approximately 110 amino acids) and C-LIP domain (approximately 210 amino acids) localized at the C-terminus for human, mouse and Dm and centrally in the fungal PAH1 (Figure 19 and Figure S 18). Secondary structure prediction of the N-LIP reveals a well-structured domain with one N-terminal helix followed by six β-sheets. The C-LIP domain conserves the predicted α-helices/β-sheets architecture from fungi to human lipin (Figure 19). The N-terminal helix of the N-LIP corresponds to the amphipathic helix of

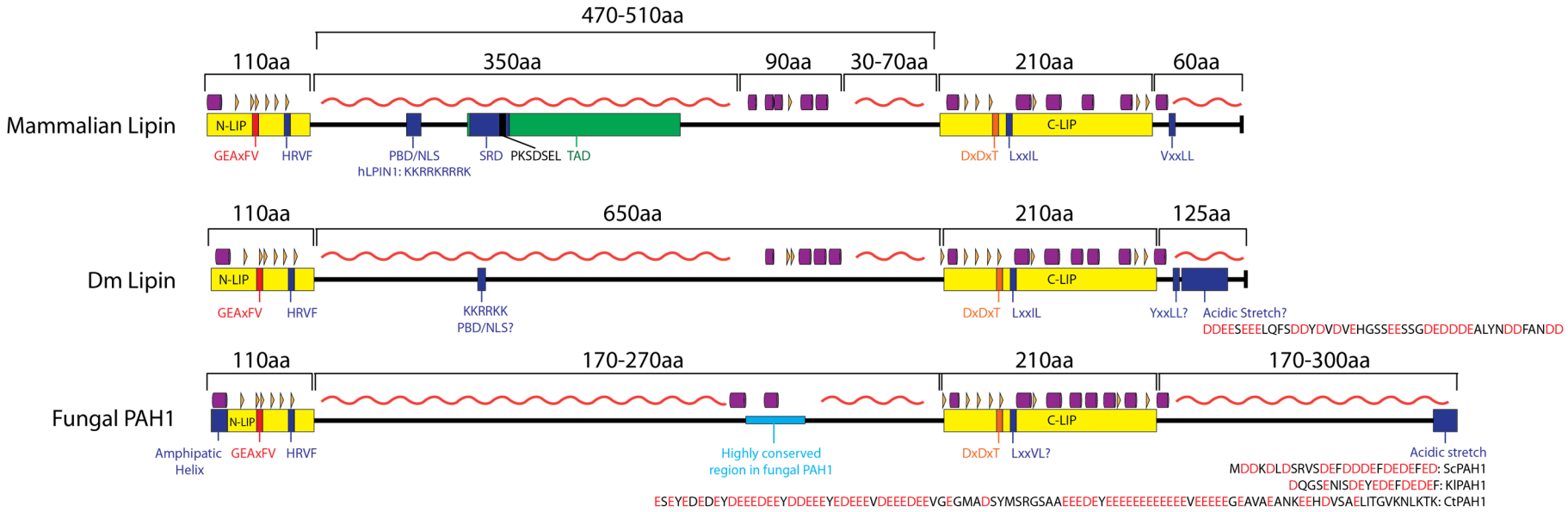


Figure 19: Characteristics of the mammalian lipin, Dm lipin and fungal PAH1 observed by sequence alignment and secondary structure prediction. Sequence alignment, secondary structures (purple α-helices and yellow β-sheets) and disordered regions (orange waves) gave information about the architecture of the lipin. The mammalian lipin refers to the human and mouse lipin isoforms 1, 2 and 3. The fungal PAH1 refers to ScPAH1, KlPAH1 and CtPAH1. Motifs and domain followed by a question mark (KKRRKK PBD/NLS?, YxxLL?, Acidic Stretch?, LxxVL?) have been identified by sequence alignment but their functionality has not been studied. The sequence of acidic stretch for ScPAH1, KlPAH1, CtPAH1 and Dm lipin is written with aspartate and glutamate colored in red.

ScPAH1 (amino acids 1 to 18), which is required for membrane recruitment after dephosphorylation by Nem1p-Spo7p [59]. The lipin from human and mouse have similar N-terminal sequence. However, mammalian lipin uses the PBD/NLS for PA binding. It is unknown if the N-terminal helix of mammalian lipin plays also a role in membrane recruitment. All lipins from fungi to human contains in their N-LIP, the HRVF motif for the binding of PP-1cγ, which is required for activation of the lipin, PAP activity and nuclear localization. All lipins also harbor the Gly84 of the mouse lipin 1 in a GEAxFV motif. Mutation of Gly to Arg is responsible for the fatty liver dystrophy phenotype in mice. The GEAxFV motif is conserved in 75 fungal PAH1 based on sequence alignment.

The N-LIP of the human and mouse lipin isoforms is followed by a disordered region of approximately 350 amino acids. This region of lipin 1 contains three functional domains: 1) the PBD/NLS domain allowing either the binding of PA at the ER membrane or serving as nuclear localization signal, 2) the SRD domain for binding the 14-3-3 protein in the cytosol, 3) and the overlapping TAD for PPARγ activation in the nucleus. The PBD/NLS is present in all mammalian lipin isoforms. The SRD and TAD do not present high sequence similarity except for a PKSDSEL motif containing charged residues. This motif overlaps the SRD and the TAD but it has not been tested if this motif has functional relevance in 14-3-3 protein binding or PPARγ activation. A 90 residues region following this disordered region does not have any identified functional domains but is structurally conserved by the presence of predicted α-helices and β-sheet. A second 30 to 70 residues disordered region precedes the N-LIP.

The intermediate region between the N-LIP and the C-LIP is different in the Dm Lipin and the fungal PAH1. The Dm Lipin present a longer intermediate region of 650 amino acids between the C-LIP and N-LIP compared to 470-510 amino acids in mammalian lipin isoforms. The intermediate region of Dm Lipin conserves the polybasic sequence KKRRKK, but for which the role of PBD/NLS has not been investigated, and conserves also the predicted α-helices/β-sheets containing region. In contrast, the intermediate region in fungi is much shorter with only 170 amino acids for KlPAH1, 240 amino acids for ScPAH1 and 270 amino acids for CtPAH1. Fungal PAH1 lack the functional domains observed in mammalian lipins but they present the same architecture with the succession of disordered region/predicted α-helices/β-sheets region/disordered region.

All lipins possess in the well-structured C-LIP the DxDxT motif for PAP activity. The mammalian and the Dm lipins present the LxxIL motifs for PPAR α binding. The fungal PAH1 have instead a LxxVL motif but it has not been investigated if this motif also interact with a fungal transcription factor. Modelisation of the human lipin 1 was tested with the Swiss Model [111]. Only very low quality models covering the amino acids 671 to 847 were generated. This region corresponds to the C-LIP and was built using templates from Haloacid dehalogenase-like hydrolase (PDB code 4eze) and deoxy-D-mannose-octulosonate 8-phosphate phosphatase (PDB code 1k1e). The region of the C-LIP seems to have a well-conserved architecture.

The C-LIP is followed in the mammalian lipin by 60 amino acids sequence containing the VxxLL binding site for PPARγ. The Dm lipin presents at the same position a YxxLL motif, for which binding of PPARγ homologue has not been studied. In fungi the sequence following the C-LIP is more extended and represent 1/3 of the full length PAH1. The PAH1 C-terminus is mainly disordered. The 25 amino-terminal residues of ScPAH1 form an acidic tail, which is required for

binding and dephosphorylation of PAH1 by the Nem1p-Spo7p complex. Similar aspartate/glutamate-rich sequences are observed at the C-terminus of KlPAH1 and CtPAH1. Interestingly, the Dm lipin contains such acid stretch at its C-terminus but no insect phosphatase has been identified so far.

Selection of fungal PAH1 for protein expression

Protein Blast of ScPAH1 revealed more precisely the regions, which are functionally and structurally important (Figure S 19). Among 75 sequences of fungal PAH1 obtained from pBlast, all conserved the catalytic DxDxT motif and the glycine of the GEAxFV motif corresponding to the Gly84 of the fld mice (Figure 20). All fungal PAH1 harbor in the C-terminal a tail aspartate/glutamate-rich sequence, which is required for binding and dephosphorylation of ScPAH1 by the Nem1p-Spo7p complex (Figure S 20). Two PAH1 out of 75 do not have any sequence similarities for the N-terminal amphiphatic helix involved in membrane recruitment. Four fungal PAH1 do not have the HVRF motif for PP1cγ binding and four present a mutation to HIRF motif. PAH1 with short disordered region can be identified and selected for protein expression (Figure S 19). Among 75 fungal PAH1 sequence, four sequences are longer than ScPAH1 (862 aa). The other PAH1 are shorter up to 618 amino acids. The KlPAH1 (794 aa) and the CtPAH1 (747 aa) are good candidates for protein expression. KlPAH1 has the same length of C-terminal disordered tail after the C-LIP than ScPAH1 but it has in general shorter internal disordered region between the N-LIP and the C-LIP (81 aa belonging to the disordered region) compared to ScPAH1 (143 aa belonging to the disordered region). Thermophilic fungi like *Chaetomium thermophilum* may produce protein thermostable. CtPAH1 has longer internal disordered region (163 aa between the N-LIP and C-LIP belongs to the disordered region) and shorter C-terminal tail (148 aa compared to 278 amino acid for ScPAH1).

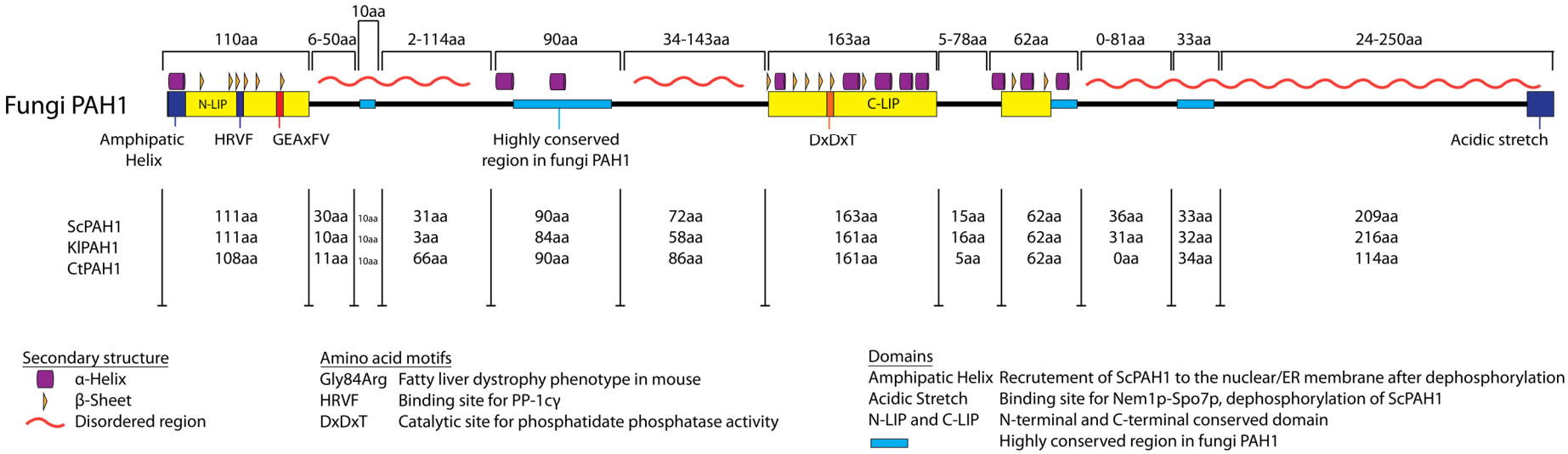


Figure 20: Characteristics of the fungal PAH1 identified by sequence alignment of 75 PAH1 obtained from pBlast with ScPAH1. Sequence alignment, secondary structures (purple α-helices and yellow β-sheets) and disordered regions (orange waves) identified the well-known conserved domains (C-LIP and N-LIP in yellow) and highly conserved region in fungal PAH1 (light blue square).

Expression screening of eukaryotic lipins in E. coli

So far, lipin has been recombinantly expressed for immunofluorescence microscopy, immunoblotting, immunoprecipitation and PAP activity assay but literature does not mention any large-scale purification of lipin with high purity and absence of degradation, as it is required for crystallization. *E. coli* is a well-established and fast sytem for protein expression and expression of lipin in such system has been reported [42]. *E. coli* was used as first strategy to express recombinant Lipin. A first trial of protein purification on a 5ml Ni-NTA column showed that the C-term 6 His-tagged ScPAH1 and KlPAH1 did not bind to the column. The C-term His tag might not be accessible for binding to Ni-NTA material. A N-term tag has also been reported in literature for the lipin [51, 64] and was used as next strategy. Monitoring of the protein expression over time and a first step of protein purification on 5ml Ni-NTA column of the full length, the C-LIP and the N-LIP constructs of the ScPAH1 and hLPIN1 revealed that after lysis and fractionation, proteins remained mostly in the insoluble fraction (Table S 7). The full-length ScPAH1 and hLPIN1 showed C-terminal degradation judged from western blot against N-terminal His-tag. Both proteins were toxic for *E. coli*, the leaky pellet was an indication of cell lysis and cell density decreased overtime during protein expression. The shortened constructs of hLPIN1 and ScPAH1 did not affect bacterial growth but proteins remained mainly insoluble and the expression did not increase overtime even when *E.coli* density was quadrupling (Table S 7). On the overall, *E. coli* was not an appropriate expression system due to the toxicity, the insolubility and the low expression yields of the protein of interest. Expression of recombinant lipin has also been reported in insect cells [48] and was therefore used as second strategy for lipin expression.

Expression screening of eukaryotic lipins in insect cells

Gateway cloning technology was used to clone into pABG2-N-term-10His-Myc-Flag-TEV (HMF) tag-containing plasmid, several full-length lipin constructs from different organisms: human, mouse, *Drosophila melanogaster*, *Saccharomyces cerevisiae*, *Kluyveromyces lactis* and *Chaetomium thermophilum.* Lipin-containing pABG2 were transformed into DH10 EmbacY *E. coli* for baculovirus production. Positive bacmids controled by sequencing were transfected into insect cells Sf21 using Escort IV reagent (Sigma), to produce the initial virus V0. EmbacY bacmid contains an YFP gene to follow the infection by fluorescence microscopy of insect cells. V0 was amplified by successive infection of Sf21 cells, generating V1 for expression test and V2 as virus template for large protein production.

Successful expression of lipin was challenged by the absence or very low expression of all lipin constructs observed at V1. Lipin-encoding-plasmid pABG2 and bacmids were cloned and sequenced again to overcome this problem. However, the critical point to observe protein expression was the V0 step. Initially, insect cells were transfected with EmbacY bacmid and Escort IV transfection reagent for four days and the supernatant V0 was used to generate V1 and to test the protein expression. In the modified protocol, cell medium was exchange after four days and infected cells were cultured for 3 more days. Insect cells fluorescence increased from 1 to 50% from day 4 to day 7 and protein expression was positive after generating V1 based on anti-His western blot of cell lysis. Based on bioinformatics analysis and protein expression, KlPAH1 was used for purification test and optimization.

Strategy for the purification of recombinant lipin from insect cells

A first strategy composed of three purification steps was designed for the full-length KlPAH1 construct: 1) affinity chromatography, 2) anion-exchange chromatography, 3) size-exclusion chromatography. Affinity chromatography on nickel-charged resin retain successfully His-tagged KlPAH1 but required several steps of buffer optimization to get rid of the strongly bound contaminants. Full-length KlPAH1 was prone to degradation and C-terminal degradation fragments observed from western blot against N-terminal His-tag could not be eliminated from the three purification steps. Limited proteolysis was carried out to identify stable fragments. A 527 amino acids fragment of KlPAH1 (Klarg: Methionine 1 to Arginine 527) was obtained by clostripain proteolytic digest and identified on SDS-PAGE followed by mass spectrometry analysis (Figure S 21). The cutting site of the Clostripain is localized 36 amino acids after the C-LIP domain. Klarg construct expressed without degradation fragments and in a good amount. From SDS-PAGE (Figure S 21), Klarg (61kDa) has a low electrophoretic mobility and runs at 73kDa. This electrophoretic behavior has been also described for the human lipin 1 [42].

Two purification strategies were tested for the Klarg construct: 1) affinity chromatography on Ni-NTA column followed by anion-exchange chromatography and gel filtration, or 2) affinity chromatography on a Ni-NTA column followed by tag-cleavage, orthogonal nickel-affinity chromatography and gel filtration. The SDS-PAGE of eluted fractions from gel filtration on Superdex 200 16/600 shows that both approaches yield protein at a purity above 95% (Figure 21). Less contaminants are observed for untagged-Klarg from the strategy 2, due to high salt with (500mM NaCl) on the first step of Ni-NTA chromatography, which successfully removes various contaminants and was not applied for the strategy 1 performed previously (Figure S 22). A total amount of 10 mg of Klarg could be purified by each method from 25 g of cell pellet.

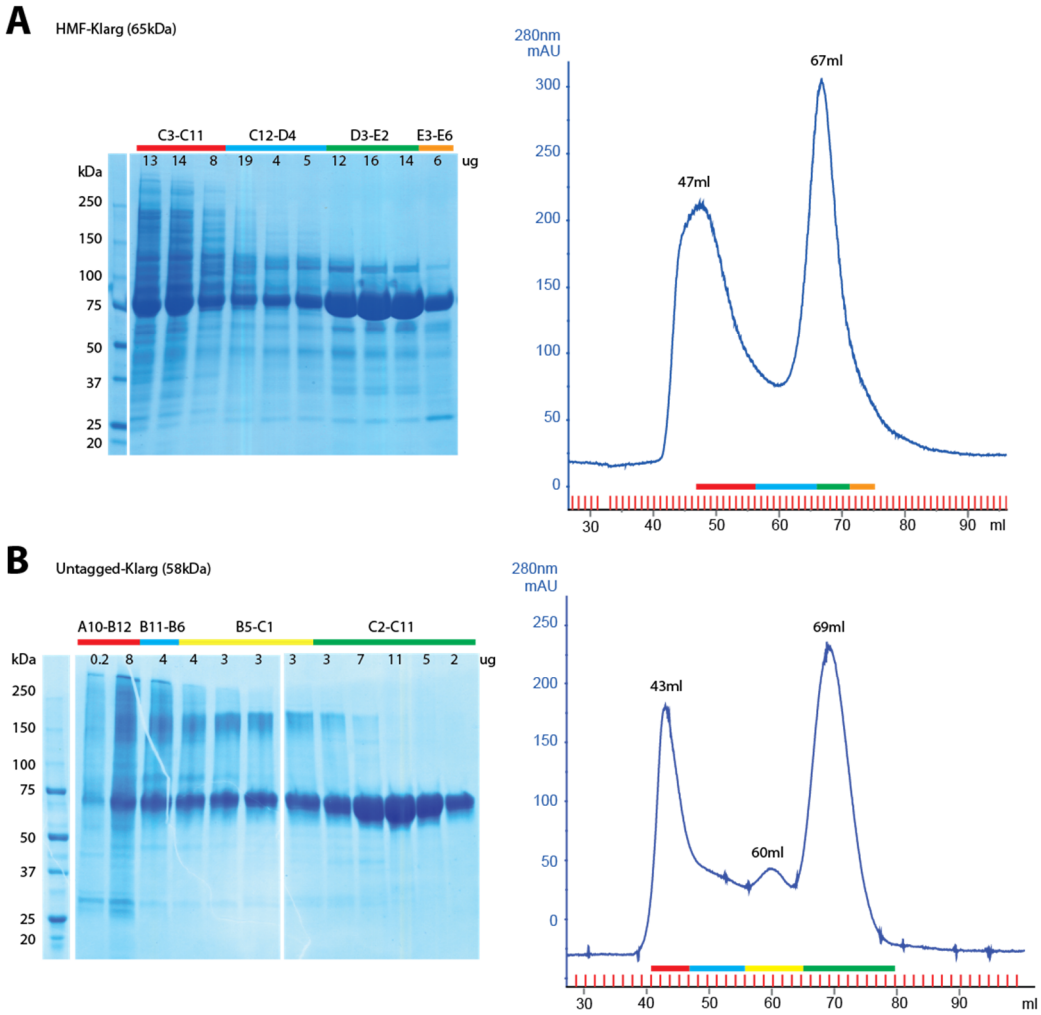


Figure 21: Last purification step of HMF-Klarg and untagged-Klarg on Superdex 200 16/600. Chromatogram and SDS-PAGE of the elution of the HMF-Klarg (obtained from the strategy 1) (A) and the untagged-Klarg (obtained from the strategy 2) (B) show the purity of the protein and the presence of species with different molecular weights. Amount in ug of protein loaded on the gel is indicated for each elution fraction.

Gel filtration on Superdex 200 16/600 revealed different oligomeric states of Klarg, which were further characterized by SEC-MALS. Two peaks at an elution volume of 47 and 67ml were observed for the HMF-Klarg (65kDa). The peak at 67ml was determined by MALS as the monomeric form of Klarg with a molecular weight of 68kDa, whereas the peak eluted at 47ml was identified as a complex mixture of species from 100 to 1000kDa (Figure 22). These peaks at 47 and 67ml were run separately during SEC-MALS and identified as separate species, indicating that no monomer-oligomer equilibrium or a slow equilibrium occurs in solution. So far, monomeric form of lipin eluted from gel filtration has not been mentioned in scientific literature. One study mentioned the elution of mouse lipin 1 (monomere 98.6kDa) at 500-600kDa on a Superose 6 gel filtration [48]. This 500-600kDa species were defined by Liu *et al.* (2010) [48] as lipin oligomers rather than aggregates as the PAP activity is conserved suggesting that the proteins are functional.

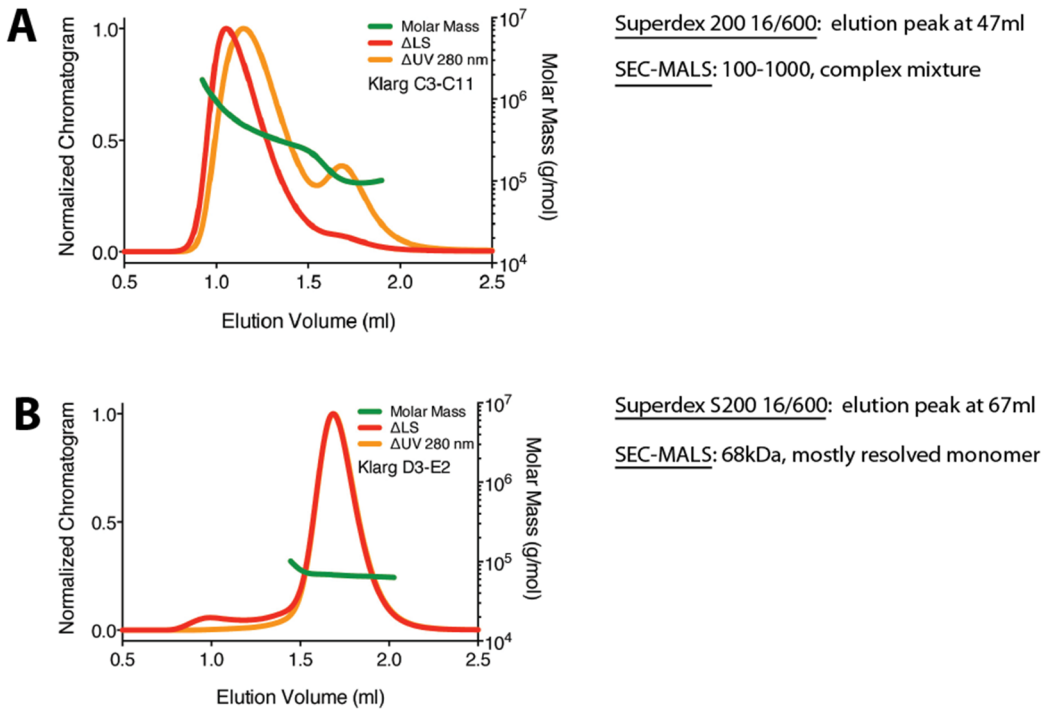


Figure 22: Determination of the oligomeric state of Klarg by SEC-MALS. (A) SEC-MALS chromatogram of the GF elution peak at 47ml. The molecular weight corresponds to complex mixture of species from 100-1000KDa. (D) SEC-MALS chromatogram of the GF elution peak at 67ml. The Klarg species eluted at 67ml have a molecular weight of 68kDa corresponding to the monomeric form of Klarg (65kDa)

The full length ScPAH1 expressed well and was tested for purification. Strategy 1 and 2 did not eliminate contaminants as well as for Klarg (Figure S 23). HMF-ScPAH1 (102kDa) also eluted in two peaks at 45.8 ml and 56ml from gel filtration (Figure S 23). The peak of elution at 56ml was identified by MALS as the mostly resolved monomer (100-130kDa) and the peak at 45.8ml was identified as oligomer/aggregate bigger than 1000kDa. Biophysical studies were pursued to determine if the oligomerization of Klarg and ScPAH1 was dependant of the redox condition from the buffer or the post-translational modifications such as phosphorylation.

Biophysical characterization of eukaryotic lipins

Sensitivity of Klarg oligomers to reducing agent was analysed by gel filtration on Superdex 200 5/150. Klarg oligomers eluted at 1.5ml whereas Klarg monomers eluted at 2.2ml. Addition of 25mM DTT to the Klarg oligomers showed an elution peak at 1.5ml, indicating that Klarg oligomers were redox-insensitive and did not involve formation of intermolecular disulfide bonds. Study from Liu et al. 2010 [48] also demonstrate the redox-insensitivity of lipin oligomers by coimmunoprecipitation of HA-tag lipin 1 and V5-lipin 1 in presence of DTT, NEM or H_2O_2.

Post-translational modifications of Klarg and ScPAH1 were also analysed to determine if they play a role in protein oligomerisation. Phosphorylation sites of the monomeric and oligomeric forms of Klarg and ScPAH1 were identified by mass spectrometry. Four sites at the C-terminus of ScPAH1 were significantly more phosphorylated on the monomeric form than on the oligomeric form (Figure S 25). However, Klarg does not contain this homologous portion and only one phosphorylation site was identified at the N-terminus on both monomeric and oligomeric form (Figure S 25). These results indicate that while oligomerization of ScPAH1 may depend on

phosphorylation, oligomerization of the Klarg lipin fragment does not. Co-immunoprecitation studies of HA-tag lipin 1 and V5-lipin 1 by Lui *et al* (2010) [48] shows similar behavior; co-immunopreciation of these two species was not affected by the addition of lamda-phosphatase. Other modifications were identified on Klarg like the acylation of six lysines (50, 55, 179, 210, 279, 289 and with a lower confidence 421) but with no difference between monomeric and oligomeric form. Oligomerization of Klarg is not influence by redox conditions and post-translational modification. Isolation of the monomeric and oligomeric forms of Klarg and ScPAH1 allow functional studies to determine whether they display equivalent enzymatic activities. Further biophysical studies were performed to determine a buffer composition appropriate for PAP activity assay.

Thermofluor assay was used to determine a buffer suitable for protein stability, storage and analysis (Figure S 24). Klarg was equally stable in all tested salt conditions with melting temperature (Tm) varying from 37 to 40°C (Table 1). The pH has a greater influence on Klarg stability, decreasing pH from 8 (Tm 35°C) to 6.5 (Tm 45°C) increases Tm by 10°C. Similar behavior for ScPAH1 was observed with a minor effect of salt type and concentration but a considerable impact of pH (Table 1 and Figure S 24). ScPAH1 has a higher Tm of 5°C in general compared to Klarg. Sodium phosphate pH 6.5 has the highest Tm for both Klarg and ScPAH1 but phosphate ions are problematic for activity assay (Table 1). The activity assay is based on the quantification of free phosphate released by the phosphatase activity of the lipin. Based on the result of differential scanning fluorimetry screening the following buffer was chosen for all protein storage and analysis: 150mM NaCl, 2mM DTT in 100mM MES buffer at pH 6.5 (Table 1).

Table 1: Melting temperatures of Klarg and ScPAH1 depending on buffer, pH and salt conditions. Buffer conditions colored in orange have the highest Tm. MES: 2-(N-morpholino)ethanesulfonic acid, MOPS: 3-(N-morpholino)propansulfonic acid

			Melting temperature (°C)	
			Klarg	ScPAH1
Buffer	Sodium acetate	pH 5	28.2	40.3
	Sodium citrate	pH 5.5	39.3	46.5
	MES	pH 6	43.4	46.3
	BisTris	pH 6.5	40.3	44.5
	Sodium Phosphate	pH 6.5	45.0	47.0
	Imidazole	pH 6.75	37.8	43.7
	Potassium phosphate	pH 7	42.5	45.8
	Hepes	pH 7	41.2	44.8
	MOPS	pH 7	39.6	44.3
	Sodium Phosphate	pH 7.5	38.8	43.5
	Potassium phosphate	pH 7.5	38.8	44.1
	Hepes	pH 7.5	37.9	42.9
	Tris	pH 7.5	37.5	42.7
	Hepes	pH 8	35.3	40.1
	Tris	pH 8.5	35.9	41.6
	Bicine	pH 9	33.7	38.8
Salt	Ammonium chloride		37.0	42.2
	Ammonium sulfate		40.5	46.2
	Potassium chloride		38.4	42.8
	Potassium acetate		40.0	45.0
	Sodium chloride		38.9	43.3
Salt concentration	0 mM		38.0	42.8
	125 mM		39.6	44.0
	250 mM		39.3	43.9
	375 mM		39.2	44.2
	500 mM		37.6	43.3

Activity assays using colometric method were performed to determine if both monomeric and oligomeric species of Klarg and ScPAH1 have PAP activity. PAP activity was measured by following the release of free phosphate from the dephosphorylation of L-α-phosphatidic acid by

PAH1. Reaction was stopped by addition of 0.1N HCl in methanol and free orthophosphates released by the PAP reaction were isolated from the hydrophobic L-α-phosphatidic acid using water/chloroform phase separation [42, 157]. In contact with malachite green molybdate reagent, free orthophosphate form a complex which absorbes at 620-640 nm [157]. PAP activity can be defined as the amount of enzyme that catalyzed the formation of one µmol of product per minute. So far, the enzymatic assay has been tested once but requires repetitive experiments to give confident results on Klarg and ScPAH1 PAP activity.

Discussion

Lipin is key regulator of lipid metabolism by its catalytic role in the triacylglycerol biosynthesis pathway and its transcriptional coactivator activity in lipid metabolism. Its role in lipid homeostasis is complex and highly regulated by several processes like phosphorylation, sumoylation and protein-protein interaction. Little is known about its structure, which would be of help to describe the protein-protein interaction interface and understand the effect of mutation involved in disease development. From bioinformatic analysis, we identify highly conserved domains, which seems to be important either for the structure, or the function of the lipin. The Gly84Arg responsible for the fld phenotype in mice is located in a GEAxF motif, which is conserved in all lipin from fungi to human. This mutation abrogates PAP activity [75] and blocks lipin-1 nuclear localization [43]. The gly84 is located 31 amino acids downstream the HRVF binding site for PP-1cγ. Mutation of this glycine may disrupt the structure of the N-LIP and the proper binding of PP-1cγ. The HRVF binding motif of PP-1cγ is conserved in mammalian lipin, Dm lipin and in 67 out of 75 PAH1 giving more evidence that PP1cγ binding is an important binding partner for the translocation of lipin to the ER membrane and the nucleus. It is interesting to see that the same domain PBD/NLS is used either for PA binding at the ER membrane or for translocation to the nucleus. A PKSDSEL motif overlaps the SRD and TAD domain in all mammalian lipins. The direct interaction of this motif to the 14-3-3 protein or to the PPARγ has not been investigated but a similar “ping-pong” binding mechanism can be hypothesized. This motif would serve either for 14-3-3 protein binding in the cytosol or activation of PPARγ in the nucleus. All lipin from fungi to human conserved a α-helices/β-sheets predicted region localised in the disordered part between the C-LIP and N-LIP. No function has been associated so far with this region but it seems to be important for the conservation of the structure of the protein. A tail enriched in aspartate/glutamate is present in all fungal PAH1. This acidic tail is required for the binding of Nem1p-Spo7p and the dephosphorylation of ScPAH1 [59]. The occurrence of the rich acidic sequence at the C-terminus of all fungal PAH1 indicates that this mechanism for PAH1 dephosphorylation is well-conserved. No such stretch is observed in mammalian lipin and so far, no binding site for the CTDNEP1/NEP1-R1 mammalian homologue has been identified. Interestingly, the Dm lipin has a rich aspartate/glutamate sequence at its C-terminal suggesting that the enzyme, which dephosphorylates Dm lipin may use a similar binding site as observed in fungal PAH1.

Given the importance of the structural and functional domains identified by sequence analysis, it is interesting to further explore their molecular interaction and function in the 3-dimensional structure of the protein. We established a protocol for the expression and purification of the recombinant Klarg (KlPAH1 Met1-Arg527). We observed that insect cells Sf21

were more adapted for the expression of lipin compared to *E. coli*. We optimized the protocol of EmbacY bacmid infection to get increased protein expression. Protein of good purity could be obtained from three purification steps: 1) affinity chromatography on Ni-NTA column, 2) Ortho-nickel chromatography, 3) and size exclusion chromatography. The full-length ScPAH1 was purified by the same protocol but various contaminants remained. From gel filtration and SEC-MALS, we could identify that Klarg is purified in the form of two separate species, a monomere and an oligomer. There is no equilibrium between these two species. Similar behavior was observed for ScPAH1, which was purified as monomer and oligomeric species. The oligomer is not dependant of phosphorylation and acylation and is resistant to reducing agent. It has been shown by previous studies that the oligomer were able to dephosphorylation of PA, indicated that they are not aggregation of unfolded proteins. We started initial PAP activity assay but the several improvements are required before giving any conclusion concerning the activity of the purified Klarg and ScPAH1. X-ray crystallography studies are also of main interest to study the structure of ScPAH1 and Klarg. So far, no crystal hits could be observed from the first crystallization trial and buffer optimization is required to identify the nucleation conditions. The expression and purification of Klarg presented herein is the first protocol for high-scale production of recombinant lipin. It is the first step, which allows further structural and biochemical investigations on the lipin family.

METHODS

Bioinformatics

Sequences were aligned using CLUSTALW and the final figure was generated by Geneious [153]. The secondary structure were predicted by Psipred [110] and the disordered region by Disopred [156].

Cloning and protein expression in insect cells Sf21

Lipin gene from *Drosophila melanogaster* (Uniprot Q8SXP0) and PAH1 from *Saccharomyces cerevisiae* (Uniprot P32567), *Kluyveromyces lactis* (Uniprot Q6CRD9) and *Chaetomium thermophilum* (Uniprot G0SF40) were PCR-amplified from genomic DNA of the corresponding organism. Lipin gene from human and mouse were purchased from ImaGenes GmbH and encode protein corresponding to the following Uniprot number: hLIPIN1α Q14693, hLPIN2 Q92539, hLPIN3 Q9BQK8, mLPIN1β Q80XT6, mLPIN2 Q99PI5, mLPIN3 Q99PI4. Recombinant baculoviruses were generated using the Bac-to-Bac baculovirus expression system (Invitrogen) according to the manufacturer's protocol. Briefly, lipin genes were amplified by PCR using AttB site encoding primers. Gene of interested was transferred by BP reaction into pdonor221 and further LR reaction into pABG2 containing a N-terminal 10 Histidine-Myc-Flag tag followed by a TEV-protease cleavage site. Bacmid was generated from transformation of lipin-encoded pABG2 into EmbacY *E. coli*. Plasmids and EmbacY bacmids were controlled by sequencing. In 6 wells plate, 2ml culture of Sf21 cells at 0.6 million cells/ml were transfected by lipin-encoded bacmid using Escort IV (Sigma) according to the manufacturer's protocol to generate V0. Insect-XPRESS™ Protein-free Insect Cell Medium with L-glutamine (Lonza) was used as growth medium and exchanged after 24 hours and 72 hours. At day 6, fluorescence of insect cells was observed under microscope to control transfection efficiency. Supernatant was

transferred to eppendorf tube and centrifuge for 5 minutes at 16000g. Supernatant corresponding to V0 was transferred to fresh tube and store in the dark at 4°C. Generation of V1 and V2 were done according to Fitzgerald et *al.* [158]. Dilution of V2 to generate high scale culture was adapted according to V2 infection capacity. Sf21 Insect cells at 0.8 million cells/ml were infected by 1/400 (v/v) V2. After 24 hours, cells were diluted to 1 million cells/ml and completed with the addition of 1/400 (v/v) V2. Cells were harvested at 72hours with controlling the following parameters: fluorescence around 99% and viability between 80 and 95%. Cells were centrifuged at 1200g for 30 minutes at 8°C and pellet was stored at -80°C.

Protein purification

Cell pellet were resuspended in lysis buffer (100 mM Hepes pH7, 150 mM NaCl, 30 mM imidazole, 7mM beta-mercaptoethanol) with addition of 4 µg/µl DNAse, a cocktail of protease inhibitor (PMSF, Bestatin, Pepstatin, Aprotinin, Phosphoamidon) and 33mM MgCl2. Cells were lysed using Microfluidizer®, centrifuged at 14200g, 45 min, at 4°C, and filter on 0.8um filter before loading on 5ml Ni-NTA column (Genscript). Contaminants were washed from the column by an extensive 20 CV column volume (CV) wash (100 mM Hepes pH7, 500 mM NaCl, 30 mM imidazole, 7mM beta-mercaptoethanol). Proteins were eluted with 500mM imidazole containing buffer. Proteins were TEV digested and dialysed overnight in dialysis buffer for ortho-nickel column (100 mM Hepes pH7, 150 mM NaCl, 7mM beta-mercaptoethanol). White turbidity was eliminated by filtering on 0.2µm filter before loading on ortho-Nickel column. Tag-cleaved protein were collected in the flowthrough and concentrating in Amicon Ultra units (Millipore) for size-exclusion chromatography. Proteins were eluted from Superdex200 16/600 (GE Healthcare) in gel filtration buffer (100mM MES pH6, 150mM NaCl, 2mM DTT) and concentrated in Amicon Ultra units (Millipore).

Proteolysis digest

Proti-Ace™ 2 kit (Hampton Research) was used according to the manual. Subtilisin, α-chymotrypsin, trypsin and elastase were purchased separately and added at the same concentration as the proteases in the kit. Samples were incubated overnight at 4°C before loading on SDS-PAGE for detection of proteolysis fragment. Identification of proteolysis fragment from SDS-PAGE bands was done by mass spectrometry in the Proteomics core facility of the Biozentrum.

ThermoFluor Assay

ThermoFuor Assay was to test thermal stability of lipin in different buffer conditions by measuring its melting temperature ™. In 96-wells and optically clear plate, 2µl of protein (20uM) were added to 8µl SYPRO® Orange dye (5x final concentration) and 10µl of condition screen (1x final concentration). Condition screen test several buffers and pH (Sodium acetate pH5, Sodium citrate pH5.5, MES pH6, BisTris pH6.5, Sodium phosphate pH6.5, Imidazole pH6.75, Potassium phosphate pH7, Hepes pH7, MOPS pH7, Sodium phosphate pH7.5, Potassium phosphate pH7.5, Hepes pH7.5, Tris pH7.5, Hepes pH8, Tris pH8.5, Bicine pH9) in combination with salts (Ammonium chloride, Ammonium sulphate, Potassium chloride, Potassium acetate, Sodium chloride) at concentration from 0 to 500mM. Changes in fluorescence at 520nm du to binding of

SYPRO® Orange dye to unfolded protein was measured on C1000™ Thermalcycler (BioRad) coupled to CFX96™ Real Time System (BioRad). Temperature was increased from 4 to 90°C with 0.5°C intervals every 40 secondes. The derivative of the fluorescence signal was plotted as function of temperature to identify Tm. Tm for each condition was determined () and an average for conditions containing the same buffer at same pH, the same salt or the same salt concentration was calculated (Table 1).

Activity assays

PAP activity was measured by following the release of water-soluble free phosphate from chloroform-soluble PA. L-α-phosphatidic acid (Avanti) was solubilized at a concentration of 10mM into 100mM Triton X-100. The reaction mixture contained 50 mM Tris-maleate pH7, 0.5 mM MgCl2, 10 mM β-mercaptoethanol, 0.2 mM L-α-phosphatidic acid (Avanti) in 2 mM Triton X-100 and 0.5ng/µl enzyme in a total volume of 50µl. Enzyme was added to other tubes at 5 minutes intervals. The reaction mixture was incubated at 37 °C and enzymatic reaction was stopped for all tubes at the time point 20 minutes by adding 250µl of 0.1N HCl in methanol. Phase separation was performed by adding 500µl of chloroform and 500µl of milli-Q water. Samples were shacked vigourously for one minute and centrifuge three minutes at 100g. After phase separation, 700µl of the aqueous (upper) phase was mixed with 100µl milli-Q water. 200ul of malachite green-molybdate reagent (MGM reagent, 2.64mM malachite green, 3.1mM H_2SO_4, 60.7mM $(NH_4)_6Mo_7O_{24}$, 10% Tween) was added at one minute interval in to all tubes and mixed by inversion several times. Absorbance at 650 nm was measured at exactly 10 minutes after addiction of MGM reagent. The production of phosphate was quantified with a phosphate calibration curve. A unit of enzymatic activity was defined as the amount of enzyme that 96atalysed the formation of 1 µmol of product/min

SUPPLEMENTARY INFORMATION

EXPRESSION, PURIFICATION AND BIOPHYSICAL CHARACTERIZATION OF LIPIN, A KEY REGULATOR OF LIPID METABOLISM

Solène F. Martin, Timm Maier*

Biozentrum, Universität Basel, Klingelbergstrasse 50/70, 4056 Basel (Switzerland)

Correspondence should be addressed to T.M. (timm.maier@unibas.ch)
E-mail addresses: solene.martin@unibas.ch

Table S 6: Sequence similarity between lipins based on sequence alignment.

		% sequence similarity
Human lipin 1 (890 aa)	**Mouse lipin 1** (891 aa)	86
Human lipin 2 (896 aa)	**Mouse lipin 2** (893 aa)	90
Human lipin 3 (851 aa)	**Mouse lipin 3** (848 aa)	80
Human lipin 1 (890 aa)	**Human lipin 2** (896 aa)	51
	Human lipin 3 (851 aa)	48
	Dm lipin (1089 aa)	35
	ScPAH1 (862 aa)	28
	KlPAH1 (794 aa)	31
	CtPAH1 (747 aa)	27
ScPAH1 (862 aa)	**KlPAH1** (794 aa)	50
	CtPAH1 (747 aa)	38

Table S 7: Expression of recombinant lipin in *E. coli*.

	Bacterial growth Monitoring of optical density at 590nm after IPTG induction of two *E. coli* cultures at 37°C for 5 hours and 17°C overnight.	Expression profiles Monitoring by western blot every hours of *E.coli* culture at 37°C for 5 hours
Full length human Lipin 1	Decrease of bacterial density during cell culture	Mostly insoluble Presence of C-terminal degraded fragments
C-terminal human Lipin 1 construct (Ser623-Ala890)	No effect on bacterial growth	Mostly insoluble Protein expression does not increase over time even though bacterial density triple
Full lenght ScPAH1	No bacterial growth	Mostly insoluble Presence of C-terminal degraded fragments
C-terminal ScPAH1 construct (Ser340-Ser655)	No effect on bacterial growth	Mostly insoluble Protein expression does not increase over time even though bacterial density quadruple

Figure S 18: Sequence alignment of hLPIN1, 2, 3 , mLPIN1, 2, 3, DmLPIN, KlPAH1, ScPAH1, CtPAH1. Domains are indicated by numbers or by their full name. The structurally conserved region in 75 fungi PAH1 is indicated by a light blue square. Domains mentioned with the question mark presents residues ressembling to the motifs but for which the function has not been studied.

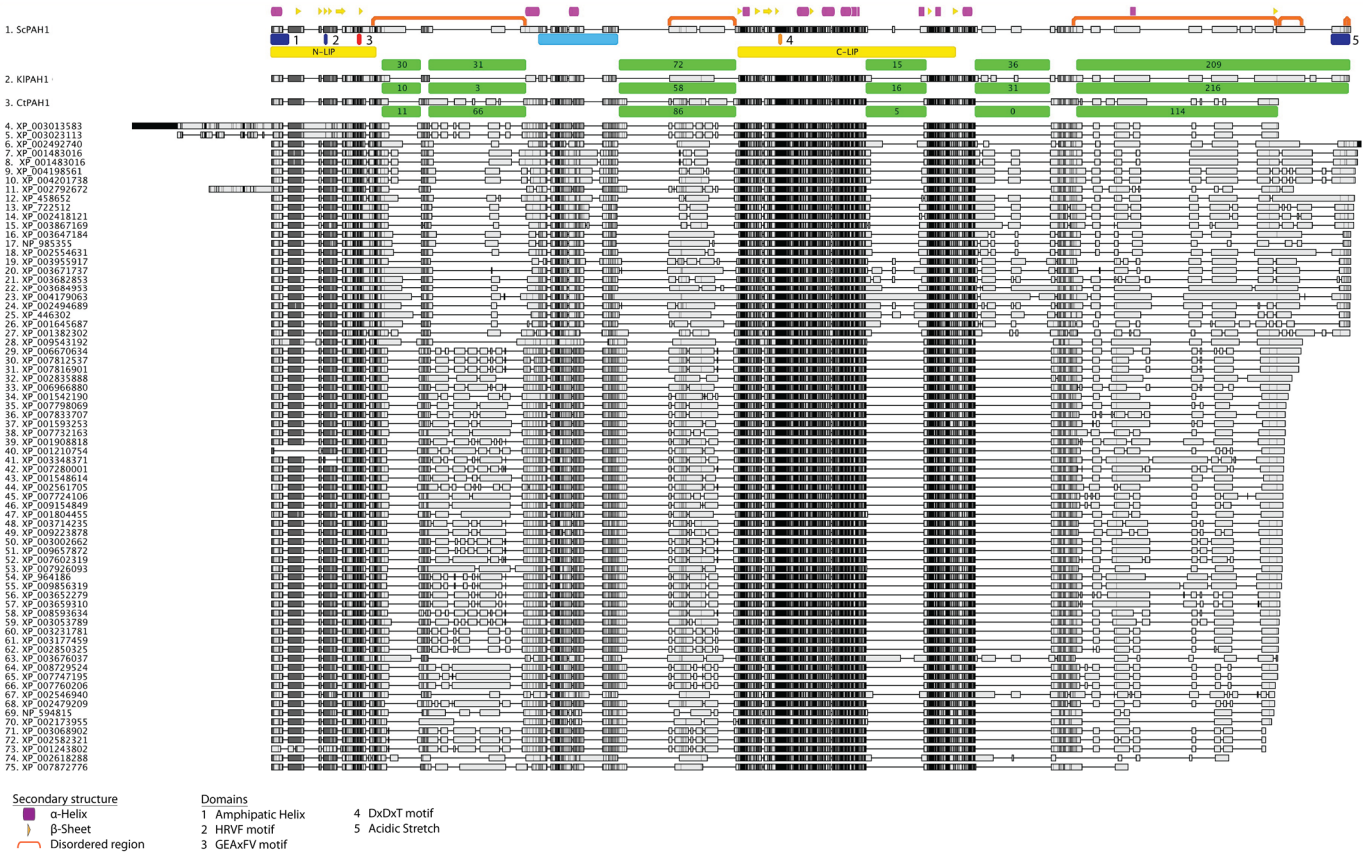


Figure S 19 : Sequence alignment of 75 sequences of fungal PAH1. Domains are indicated by numbers or by their full name. The structurally conserved region of the disordered part in 75 fungi PAH1 is indicated by a light blue square. The length of the disordered region for ScPAH1, KlPAH1 and CtPAH1 are indicated by number of residues in green square.

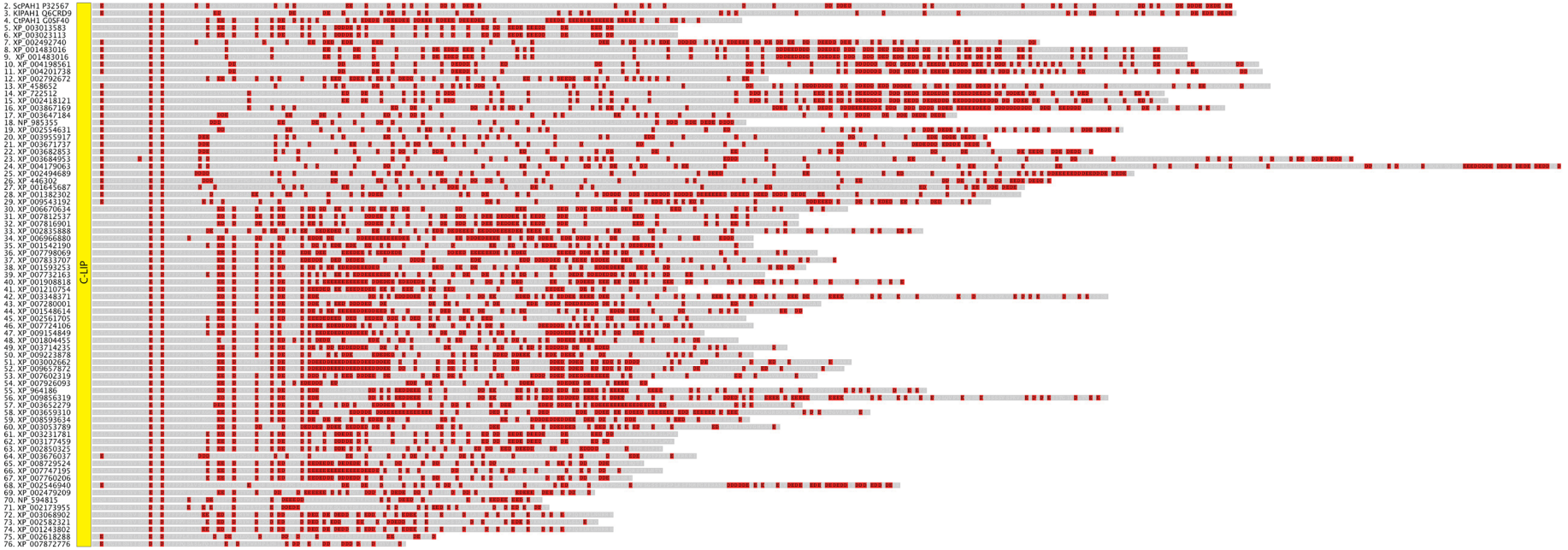


Figure S 20: C-terminal tail of 75 fungal PAH1. All fungal PAH1 contains enriched region of Aspartate/Glutamate (colored in red) residues, which serves in ScPAH1 for Nem1p/Spo7p binding.

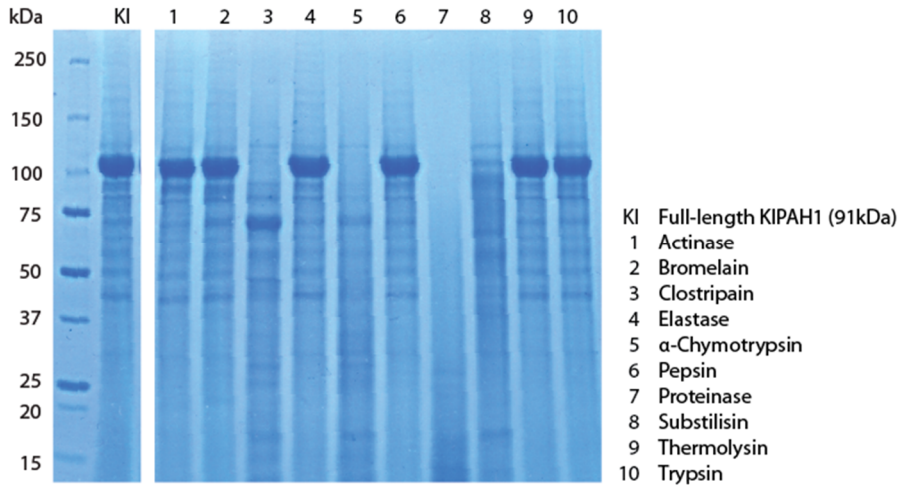


Figure S 21: Limited proteolysis of the full-length KlPAH1. A shorter fragment of KlPAH1 at 73kDa is obtained by Clostripain proteolytic digest (3) and was characterized by mass spectrometry as a KlPAH1 fragment from methionine 1 to arginine 527 (Klarg). Klarg (61kDa) runs at 73kDa on SDS-PAGE. This low mobility on electrophoretic gel has been described for the human lipin 1 [42].

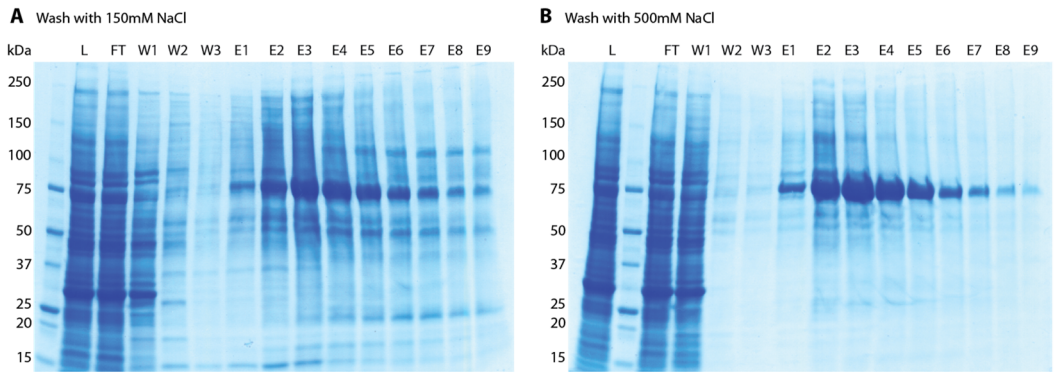


Figure S 22: Elution of the small variant of KlPAH1 (Klarg, Met1-Arg 527) from Ni-NTA chromatography column. Same volume of insect lysis expressing HMF-Klarg (65kDa) was loaded on two separate 5ml Ni-NTA columns. Wash with 500mM NaCl helps to eliminate contaminants. L: Load, FT, Flowthrough, W: wash, E: Elution.

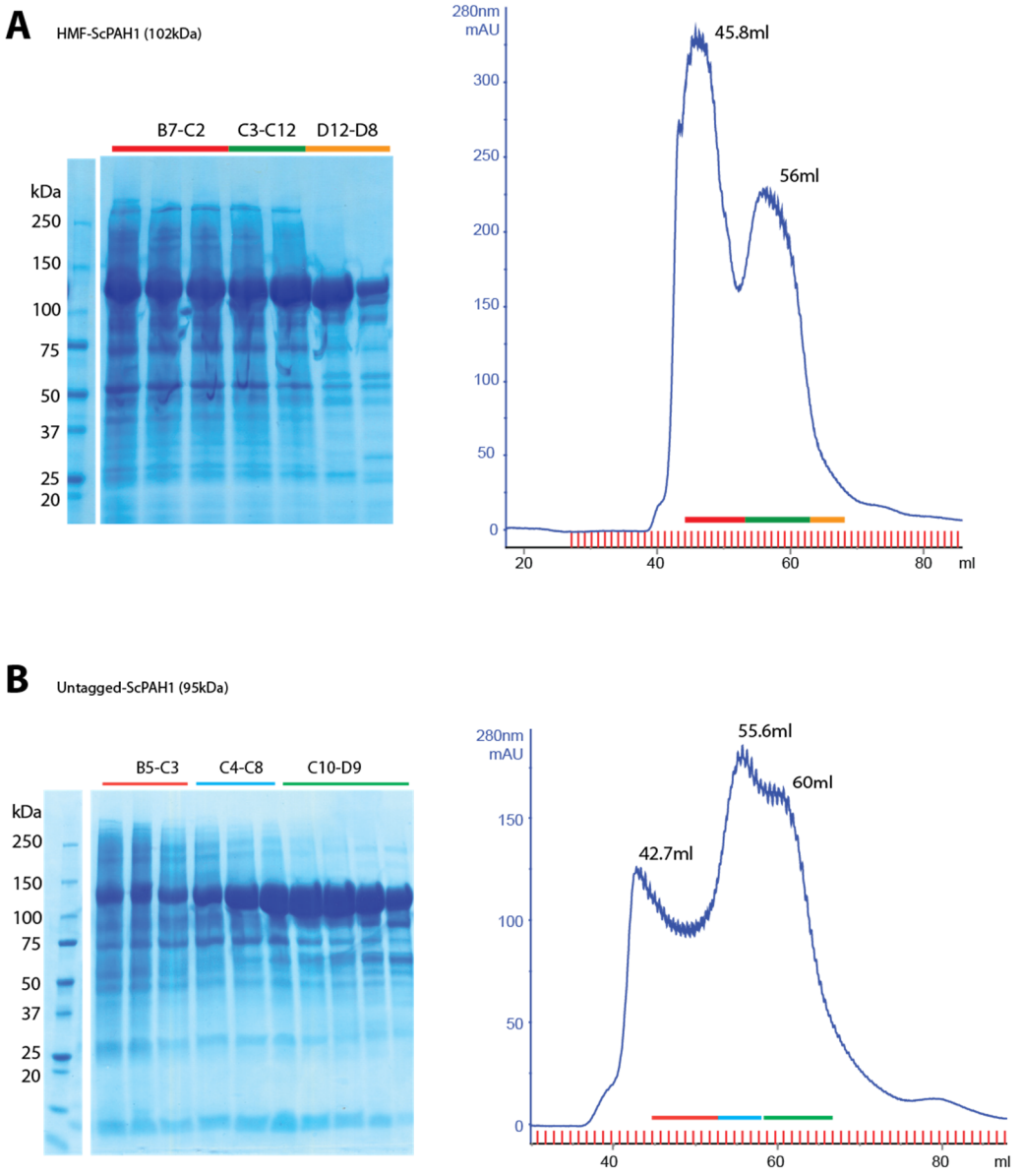


Figure S 23: Gel filtration of HMF-ScPAH1 and untagged-ScPAH1 on Superdex 200 16/600. Chromatogram and SDS-PAGE of the elution of the HMF-ScPAH1 (obtained from the strategy 1) (A) and the untagged-ScPAH1 (obtained from the strategy 2) (B) show the purity of the protein and the presence of species with different molecular weights. ScPAH1 was obtained with a lower purity compared to Klarg.

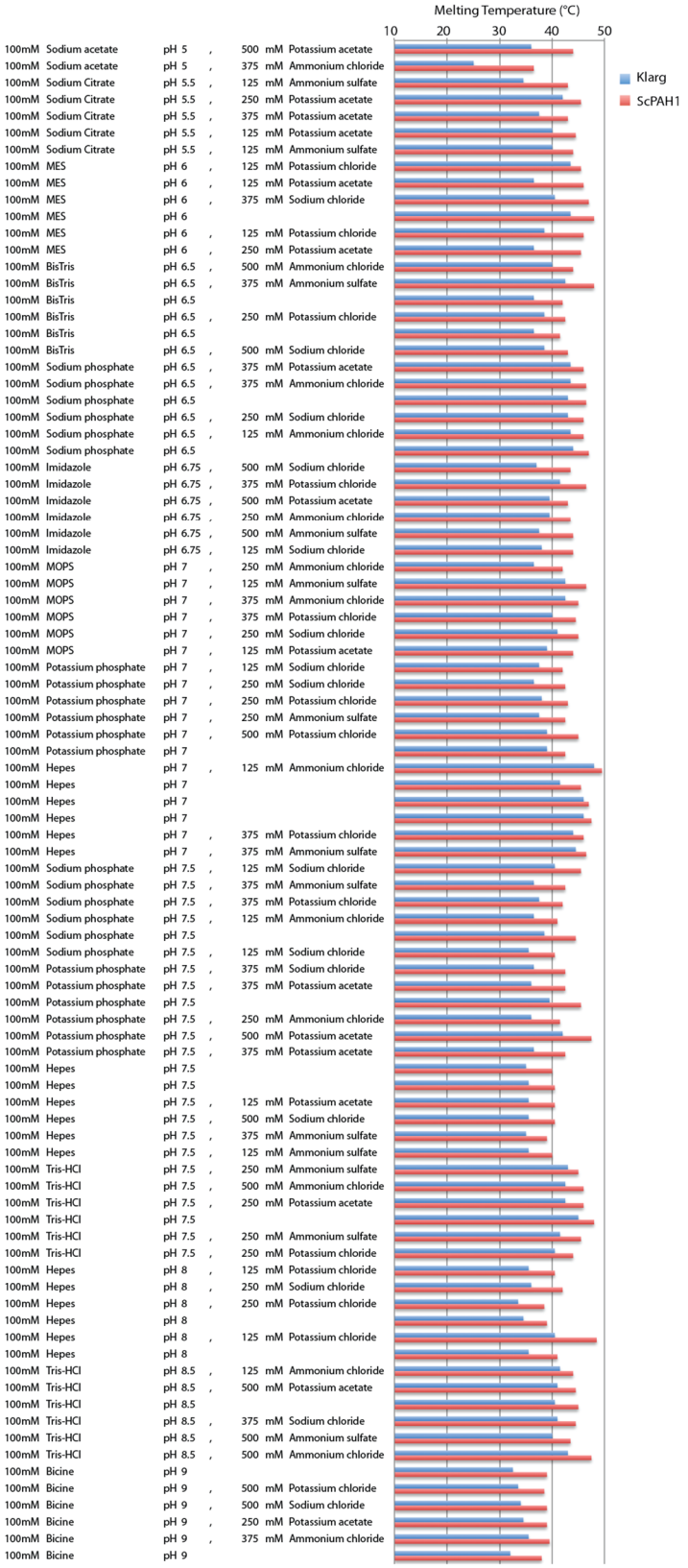


Figure S 24: Melting temperature of Klarg and ScPAH1 measure by thermofluorassay. Buffer, pH and salt conditions used to test thermal stability of Klarg and ScPAH1 are indicated on the left. Melting temperature is representing in the form of scale bar from 10 to 50°C colored in blue for Klarg and red for ScPAH1. MES: 2-(N-morpholino)ethanesulfonic acid, MOPS: 3-(N-morpholino)propansulfonic acid.

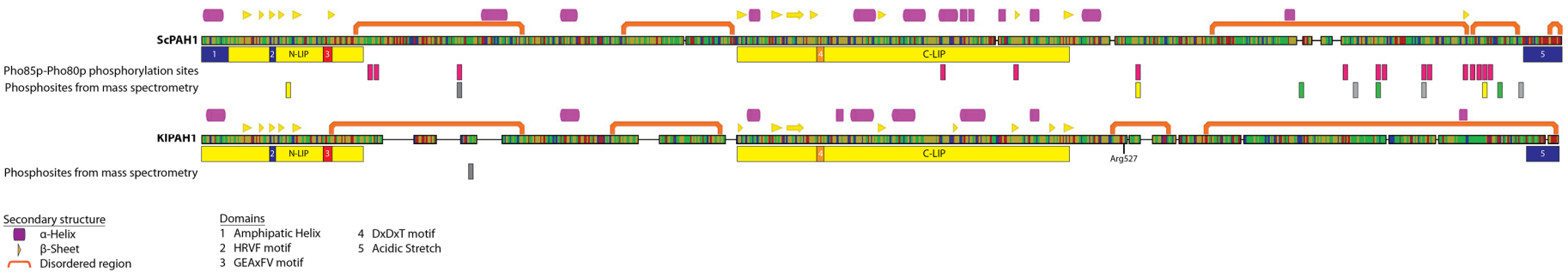


Figure S 25: Sequence alignment and mapping of phosphosite on ScPAH1 and KlPAH1. The phosphorylation sites in Pah1p were identified by O'Hara *et al.* [159]. Phosphosites identified by mass spectrometry are indicated by square. Residues predominantly phosphorylated in the monomeric form of ScPAH1 compared to the oligomeric form are colored in green. Residues predominantly phosphorylated in ScPAH1 but with less confidence are in yellow. Residues phosphorylated in both monomeric and oligomeric form of ScPAH1 and KlPAH1 are in grey. The arginine 527 corresponding to the last residues of the Klarg construct is indicated.

CONCLUSION & OUTLOOK

This thesis focuses on enzymes involved in polyketide and lipid biosynthesis. Polyketide synthases and eukaryotic lipins are of major interest in human health. Polyketide synthases are mainly from microbial origin and responsible for the synthesis of therapeutical compounds such as antibiotics, immunosuppressant and anti-tumoral drugs. Eukaryotic lipins are key regulator of lipin metabolism by their phosphatidate phosphatase activity and their transcriptional coactivator function. Mutations and polymorphisms in lipin gene are associated with metabolic and inflammatory disorders highlighting the important role of lipin in lipid metabolism homeostasis.

Structural organization of trans-AT-ER in complex with trans-AT PKS

In the present work, the structure of the KS domain with its linker domain from the *B. brevis trans*-AT PKS was determined by X-ray crystallography. The KS domain presents the thiolase fold and the catalytic triad Cys-His-His characteristic of the KS domain from *cis*-AT PKS and FAS. Helical and β-strand extensions absent in FAS and observed in *B. brevis* KS structure might be involved in domain interaction in modular *trans*-AT PKS. Structural comparisons and sequence alignment of 236 KS domains with their C-terminal linker from various *trans*-AT PKS highlight the presence of a structurally conserved linker domain, which in *cis*-AT PKS, docks the AT domain. In *trans*-AT PKS, the interface of the LD for potential docking of AT, could be classified into 4 groups depending on their structural organization. *B. brevis* LD presents a helical triangle interface with a central hydrophobic groove and peripheral charged residues making contact with other KS-LD protein in the crystal. However, no interaction between *B. brevis* KS-LD and didomain constructs from the related *trans*-AT-AT-ER could be detected in our experimental conditions but protein solubility limited the measurements and therefore cannot exclude the possibility of transient interactions.

Collaboration with the Dr. Roman Jakob brought information on the structural organization of *trans*-AT-ER protein. The crystal structure of the ER domain from the *trans*-AT-ER of the macrolactin cluster of *Bacillus amyloquefaciens* revealed the dimeric TIM-barrel structure with additional insertions, which might serve in modular *trans*-AT PKS assembly. SAXS and SEC-MALS studies showed the monomeric behaviour of the AT1-AT2 construct from *B. brevis trans*-AT1-AT2-ER while the *trans*-AT-ER from the difficidine cluster of *Bacillus amyloliquefaciens* dimerizes at the ER interface.

From these findings, we suggest a model of organization of the *trans*-AT PKS. The AT domain of the dimeric *trans*-AT-ER may interact with *trans*-AT PKS complex at the LD interface in order to be positioned like in *cis*-AT PKS. Extensions of the dimeric *trans*-ER may also contribute to the positioning of the *trans*-ER and the assembly with the *trans*-AT PKS. This model of *trans*-AT PKS in complex with *trans*-AT-ER bring new knowledge for the engineering of novel *trans*-AT PKS and combinatorial biosynthesis of novel polyketides. Redesign of biosynthetic pathways is often limited by lack of understanding of the structure, dynamics and interplay between the enzymatic domains. The present model increases the ability to identify interaction interface, which are

important for *trans*-AT PKS assembly. Maintenance of the inter-domain boundary is crucial for the transfer of the growing polyketide between successive enzymatic domains. In the design of hybrid PKS, non-natural *trans*-AT having appropriate interaction interfaces could be used for delivery of specific substrate to *trans*-AT PKS. Protein in *trans* are therefore interesting for domain swapping but they can also be engineered more easily compared to domain from the non-split *cis*-AT PKS assembly line. Mutagenesis of the active site from *trans*-AT and *trans*-ER can be used to synthesize specific products without destabilizing the *trans*-AT PKS complex. Engineering the great machinery of *trans*-AT PKS can produce large libraries of chemically complex coumpounds. There is significant interest in developing synthetic strategies that could be used to improve the pharmacological properties of polyketide. Complete synthesis of polyketide remains challenging by chemistry and semi-synthesis strategies that rely on diversifying the structures are limited by the polyketide scaffolds. Combinatorial biosynthesis based on hybrid PKS opens the route for the discovery of structurally and functionally diversified drugs.

Pure and stable eukaryotic lipins for biophysical and structural studies

The present work reports the challenging expression and purification of stable and monodisperse recombinant lipin. Successful protein expression was obtained in insect cells Sf21 after adapting the baculovirus infection protocol. A stable construct of the *Kluyveromyces lactis* PAH1 was identified by limited proteolysis. Buffers composition and purification strategy was set up to allow successful and reproducible purification of the Klarg construct. Protein of good purity was obtained by three purification steps composed of 1) affinity chromatography on Ni-NTA column, 2) orthogonal nickel-affinity chromatography, 3) and size-exclusion chromatography. These expression and purification protocols are great tools for the large-scale production of pure Klarg, and the starting point for biophysical and structural studies.

The presence of two separate species of Klarg was identified by gel filtration and SEC-MALS, one specie behaving as monomer and the other one as oligomer. The oligomeric form of Klarg was shown to be independent of phosphorylation state and resistant to reducing agent. Activity assays and crystallization trials were started but require optimization. The purity of the Klarg is optimal for other structural studies such as NMR, SAXS and EM and the purification protocols can be further adapted for lipins from other organisms.

Identifying the structural organization of lipin is of main interest since its role in metabolic disorders has been clearly identified in human and fld mice. Variations in lipin 1 expression levels and gene polymorphisms are associated with insulin sensitivity, diabetes, hypertension and myogloburinemia while mutations in lipin 2 result in the development of an inflammatory disorder in human patients. The conservation of one or several isoforms, in species ranging from yeast to invertebrates, plants and mammalians settle the fundamental role of lipin in cellular metabolism. By its phosphatidate phosphatase activity and its function as transcriptional coactivator, lipin play a paradoxical role in the lipid synthesis at the ER membrane and the activation of fatty oxidation gene in the nucleus. Exploring the structure of lipin will reveal the architecture of the catalytic site and the interaction interface with other transcription factors (PPARγ and PGC1α-PPARα) and regulatory proteins (14-3-3 protein and PP-1cγ). It can be

imagined that, based on the structural information, therapeutical molecules can be designed to modulate or mimic lipin functions in order restore the balance of lipids in metabolic diseases.

ACKNOWLEGMENTS

My first thanks go to Pr. Timm Maier, who gave me the opportunity to pursue my aspiration for science and biology. I was already very interested in molecular biology and understanding the relationships between proteins, lipids, DNA, which allows the functioning of the cell. X-ray crystallography is, for me, a great “microscope” to observe the structure of proteins and how ligands influence their dynamics and therefore the cell biochemistry. I am very grateful that I had the opportunity to learn so many lab techniques and to be part of the scientific world. I am very thankful to Pr. Timm Maier for sharing expertise and guidance for my project.

I would like to thank each member of my group: Roman, Evelyn, Edward, Friederike, Habib, Fabian, Anna, Dominik, Moritz, Stefan, Rita and Franziska, Corinne, Marlise and Nicole. They always took time to discuss with me and I thank them for all the great helpful tips and valuable experience that they shared with me.

I am very thankful for the encouragement during my PhD that I received from the members of my PhD committee, Pr. Timm Maier, Pr. Petr Broz and Pr. Roderick Lim.

I am also grateful to all, who directly or indirectly, were present with me, for fun or for work, during these three years.

REFERENCES

1. Csaki, L.S., et al., *Lipins, lipinopathies, and the modulation of cellular lipid storage and signaling*, in *Prog. Lipid Res.*2013. p. 305-316.
2. Bosma, M., et al., *Re-evaluating lipotoxic triggers in skeletal muscle: relating intramyocellular lipid metabolism to insulin sensitivity.* Prog Lipid Res, 2012. **51**(1): p. 36-49.
3. Listenberger, L.L., et al., *Triglyceride accumulation protects against fatty acid-induced lipotoxicity.* Proc Natl Acad Sci U S A, 2003. **100**(6): p. 3077-82.
4. Hannun, Y.A. and L.M. Obeid, *Principles of bioactive lipid signalling: lessons from sphingolipids.* Nat Rev Mol Cell Biol, 2008. **9**(2): p. 139-50.
5. Tilley, S.L., T.M. Coffman, and B.H. Koller, *Mixed messages: modulation of inflammation and immune responses by prostaglandins and thromboxanes.* J Clin Invest, 2001. **108**(1): p. 15-23.
6. Baenke, F., et al., *Hooked on fat: the role of lipid synthesis in cancer metabolism and tumour development.* Dis Model Mech, 2013. **6**(6): p. 1353-63.
7. Smith, S. and S.C. Tsai, *The type I fatty acid and polyketide synthases: a tale of two megasynthases.* Natural Product Reports, 2007. **24**(5): p. 1041-72.
8. Khosla, C., et al., *Structure and Mechanism of the 6-Deoxyerythronolide B Synthase*, in *Annu. Rev. Biochem.*2007. p. 195-221.
9. Reue, K., *The lipin family: mutations and metabolism.*, in *Curr. Opin. Lipidol.*2009. p. 165-170.
10. Bi, L., Z. Jiang, and J. Zhou, *The role of lipin-1 in the pathogenesis of alcoholic Fatty liver.* Alcohol Alcohol, 2015. **50**(2): p. 146-51.
11. Meier, J.L. and M.D. Burkart, *The chemical biology of modular biosynthetic enzymes*, in *Chem. Soc. Rev.*2009. p. 2012.
12. Shen, B., *Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms.* Current Opinion in Chemical Biology, 2003. **7**(2): p. 285-295.
13. Maier, T., M. Leibundgut, and N. Ban, *The Crystal Structure of a Mammalian Fatty Acid Synthase*, in *Science*2008. p. 1315-1322.
14. Tang, Y.Y., et al., *Structural and mechanistic analysis of protein interactions in module 3 of the 6-deoxyerythronolide B synthase.* Chemistry & Biology, 2007. **14**(8): p. 931-943.
15. Tang, Y.Y., et al., *The 2.7-angstrom crystal structure of a 194-kDa homodimeric fragment of the 6-deoxyerythronolide B synthase.* Proceedings of the National Academy of Sciences of the United States of America, 2006. **103**(30): p. 11124-11129.
16. Whicher, J.R., et al., *Cyanobacterial polyketide synthase docking domains: a tool for engineering natural product biosynthesis.* Chem Biol, 2013. **20**(11): p. 1340-51.
17. Pappenberger, G., et al., *Structure of the Human Fatty Acid Synthase KS-MAT Didomain as a Framework for Inhibitor Design*, in *J Mol Biol*2010. p. 12-12.
18. Maier, T., S. Jenni, and N. Ban, *Architecture of mammalian fatty acid synthase at 4.5 A resolution.* Science, 2006. **311**(5765): p. 1258-62.
19. Rangan, V.S. and S. Smith, *Alteration of the substrate specificity of the malonyl-CoA/adetyl-CoA:acyl carrier protein S-acyltransferase domain of the multifunctional fatty acid synthase by mutation of a single arginine residue.* Journal of Biological Chemistry, 1997. **272**(18): p. 11975-11978.
20. Moss, S.J., C.J. Martin, and B. Wilkinson, *Loss of co-linearity by modular polyketide synthases: a mechanism for the evolution of chemical diversity.* Natural Product Reports, 2004. **21**(5): p. 575-93.
21. Koryakina, I., et al., *Promiscuity of a modular polyketide synthase towards natural and non-natural extender units.* Org Biomol Chem, 2013. **11**(27): p. 4449-58.
22. Wilkinson, B. and J. Micklefield, *Mining and engineering natural-product biosynthetic pathways.* Nature Chemical Biology, 2007. **3**(7): p. 379-86.
23. Weissman, K.J. and P.F. Leadlay, *Combinatorial biosynthesis of reduced polyketides.* Nature Reviews Microbiology, 2005. **3**(12): p. 925-936.

24. Wong, F.T. and C. Khosla, *Combinatorial biosynthesis of polyketides--a perspective.* Current Opinion in Chemical Biology, 2012. **16**(1-2): p. 117-23.
25. Kittendorf, J.D. and D.H. Sherman, *Developing tools for engineering hybrid polyketide synthetic pathways.* Curr Opin Biotechnol, 2006. **17**(6): p. 597-605.
26. Dunn, B.J. and C. Khosla, *Engineering the acyltransferase substrate specificity of assembly line polyketide synthases.* J R Soc Interface, 2013. **10**(85): p. 20130297.
27. Reeves, C.D., et al., *Alteration of the substrate specificity of a modular polyketide synthase acyltransferase domain through site-specific mutations.* Biochemistry, 2001. **40**(51): p. 15464-15470.
28. Del Vecchio, F., et al., *Active-site residue, domain and module swaps in modular polyketide synthases.* Journal of Industrial Microbiology & Biotechnology, 2003. **30**(8): p. 489-494.
29. Sundermann, U., et al., *Enzyme-Directed Mutasynthesis: A Combined Experimental and Theoretical Approach to Substrate Recognition of a Polyketide Synthase.* Acs Chemical Biology, 2013. **8**(2): p. 443-450.
30. Dunn, B.J., D.E. Cane, and C. Khosla, *Mechanism and Specificity of an Acyltransferase Domain from a Modular Polyketide Synthase.* Biochemistry, 2013. **52**(11): p. 1839-1841.
31. Menzella, H.G., et al., *Combinatorial polyketide biosynthesis by de novo design and rearrangement of modular polyketide synthase genes.* Nature Biotechnology, 2005. **23**(9): p. 1171-1176.
32. Patel, K., et al., *Engineered biosynthesis of geldanamycin analogs for Hsp90 inhibition (vol 11, pg 1625, 2004).* Chemistry & Biology, 2006. **13**(3): p. 341-341.
33. El-Sayed, A.K., et al., *Characterization of the Mupirocin Biosynthesis Gene Cluster from Pseudomonas fluorescens NCIMB 10586*, in *Chemistry & Biology*2003. p. 419-430.
34. Cheng, Y.Q., G.L. Tang, and B. Shen, *Type I polyketide synthase requiring a discrete acyltransferase for polyketide biosynthesis.* Proc Natl Acad Sci U S A, 2003. **100**(6): p. 3149-54.
35. Piel, J., et al., *Targeting modular polyketide synthases with iteratively acting acyltransferases from metagenomes of uncultured bacterial consortia.* Environ Microbiol, 2004. **6**(9): p. 921-7.
36. Piel, J., *A polyketide synthase-peptide synthetase gene cluster from an uncultured bacterial symbiont of Paederus beetles.* Proc Natl Acad Sci U S A, 2002. **99**(22): p. 14002-7.
37. Piel, J., *Biosynthesis of polyketides by trans-AT polyketide synthases.* Natural Product Reports, 2010. **27**(7): p. 996-1047.
38. Nguyen, T., et al., *Exploiting the mosaic structure of trans-acyltransferase polyketide synthases for natural product discovery and pathway dissection*, in *Nat. Biotechnol.*2008. p. 225-233.
39. Calderone, C.T., et al., *Convergence of isoprene and polyketide biosynthetic machinery: isoprenyl-S-carrier proteins in the pksX pathway of Bacillus subtilis.* Proc Natl Acad Sci U S A, 2006. **103**(24): p. 8977-82.
40. Straight, P.D., et al., *A singular enzymatic megacomplex from Bacillus subtilis.* Proceedings of the National Academy of Sciences of the United States of America, 2007. **104**(1): p. 305-310.
41. Siniossoglou, S., *Phospholipid metabolism and nuclear function: Roles of the lipin family of phosphatidic acid phosphatases*, in *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids*2013. p. 575-581.
42. Han, G.S. and G.M. Carman, *Characterization of the Human LPIN1-encoded Phosphatidate Phosphatase Isoforms*, in *Journal of Biological Chemistry*2010. p. 14628-14638.
43. Péterfy, M., et al., *Lipodystrophy in the fld mouse results from mutation of a new gene encoding a nuclear protein, lipin.*, in *Nat. Genet.*2001. p. 121-124.
44. Mietkiewska, E., et al., *Lipins from plants are phosphatidate phosphatases that restore lipid synthesis in a pah1Δ mutant strain of Saccharomyces cerevisiae.*, in *FEBS J.*2011. p. 764-775.

45. Valente, V., et al., *Drosophila melanogaster lipins are tissue-regulated and developmentally regulated and present specific subcellular distributions.*, in *FEBS J.*2010. p. 4775-4788.
46. Pelletier, M., et al., *Identification of a novel lipin homologue from the parasitic protozoan Trypanosoma brucei.* BMC Microbiol, 2013. **13**: p. 101.
47. Han, G.S., W.I. Wu, and G.M. Carman, *The Saccharomyces cerevisiae Lipin homolog is a Mg2+-dependent phosphatidate phosphatase enzyme.* J Biol Chem, 2006. **281**(14): p. 9210-8.
48. Liu, G.-H., et al., *Lipin proteins form homo- and hetero-oligomers.*, in *Biochem. J.*2010. p. 65-76.
49. Dwyer, J.R., et al., *Mouse lipin-1 and lipin-2 cooperate to maintain glycerolipid homeostasis in liver and aging cerebellum.* Proc Natl Acad Sci U S A, 2012. **109**(37): p. E2486-95.
50. Donkor, J., et al., *Three mammalian lipins act as phosphatidate phosphatases with distinct tissue expression patterns.* J Biol Chem, 2007. **282**(6): p. 3450-7.
51. Creutz, C.E., J.M. Eaton, and T.E. Harris, *Assembly of High Molecular Weight Complexes of Lipin on a Supported Lipid Bilayer Observed by Atomic Force Microscopy.*, in *Biochemistry*2013.
52. Peterfy, M., J. Phan, and K. Reue, *Alternatively spliced lipin isoforms exhibit distinct expression pattern, subcellular localization, and role in adipogenesis.* Journal of Biological Chemistry, 2005. **280**(38): p. 32883-9.
53. Wang, H., et al., *Lipin-1gamma isoform is a novel lipid droplet-associated protein highly expressed in the brain.* FEBS Lett, 2011. **585**(12): p. 1979-84.
54. Han, G.S., S. Siniossoglou, and G.M. Carman, *The cellular functions of the yeast lipin homolog PAH1p are dependent on its phosphatidate phosphatase activity.* J Biol Chem, 2007. **282**(51): p. 37026-35.
55. Siniossoglou, S., *Lipins, lipids and nuclear envelope structure.*, in *Traffic*2009. p. 1181-1187.
56. Pascual, F. and G.M. Carman, *Phosphatidate phosphatase, a key regulator of lipid homeostasis.*, in *Biochim. Biophys. Acta*2013. p. 514-522.
57. Eaton, J.M., et al., *Phosphorylation of lipin 1 and charge on the phosphatidic acid head group control its phosphatidic acid phosphatase activity and membrane association.*, in *Journal of Biological Chemistry*2013. p. 9933-9945.
58. Eastmond, P.J., et al., *Phosphatidic acid phosphohydrolase 1 and 2 regulate phospholipid synthesis at the endoplasmic reticulum in Arabidopsis.* Plant Cell, 2010. **22**(8): p. 2796-811.
59. Karanasios, E., et al., *A phosphorylation-regulated amphipathic helix controls the membrane translocation and function of the yeast phosphatidate phosphatase.* Proc Natl Acad Sci U S A, 2010. **107**(41): p. 17539-44.
60. Drin, G., et al., *A general amphipathic alpha-helical motif for sensing membrane curvature.* Nat Struct Mol Biol, 2007. **14**(2): p. 138-46.
61. Ren, H., et al., *A phosphatidic acid binding/nuclear localization motif determines lipin1 function in lipid metabolism and adipogenesis.* Mol Biol Cell, 2010. **21**(18): p. 3171-81.
62. Bou Khalil, M., et al., *The level and compartmentalization of phosphatidate phosphatase-1 (lipin-1) control the assembly and secretion of hepatic VLDL.* J Lipid Res, 2009. **50**(1): p. 47-58.
63. Finck, B.N., et al., *Lipin 1 is an inducible amplifier of the hepatic PGC-1alpha/PPARalpha regulatory pathway.*, in *Cell Metab.*2006. p. 199-210.
64. Kim, H.E., et al., *Lipin1 regulates PPARγ transcriptional activity.*, in *Biochem. J.*2013. p. 49-60.
65. Donkor, J., et al., *A conserved serine residue is required for the phosphatidate phosphatase activity but not the transcriptional coactivator functions of lipin-1 and lipin-2.*, in *J. Biol. Chem.*2009. p. 29968-29978.

66. Choi, H.-S., et al., *Pho85p-Pho80p phosphorylation of yeast Pah1p phosphatidate phosphatase regulates its activity, location, abundance, and function in lipid metabolism.*, in *J. Biol. Chem.*2012. p. 11290-11301.
67. Choi, H.S., et al., *Phosphorylation of phosphatidate phosphatase regulates its membrane association and physiological functions in Saccharomyces cerevisiae: identification of SER(602), THR(723), AND SER(744) as the sites phosphorylated by CDC28 (CDK1)-encoded cyclin-dependent kinase.* J Biol Chem, 2011. **286**(2): p. 1486-98.
68. Su, W.M., et al., *Protein kinase A-mediated phosphorylation of Pah1p phosphatidate phosphatase functions in conjunction with the Pho85p-Pho80p and Cdc28p-cyclin B kinases to regulate lipid synthesis in yeast.* J Biol Chem, 2012. **287**(40): p. 33364-76.
69. Xu, Z., W.M. Su, and G.M. Carman, *Fluorescence spectroscopy measures yeast PAH1-encoded phosphatidate phosphatase interaction with liposome membranes.* J Lipid Res, 2012. **53**(3): p. 522-8.
70. Peterfy, M., et al., *Insulin-stimulated interaction with 14-3-3 promotes cytoplasmic localization of lipin-1 in adipocytes.* Journal of Biological Chemistry, 2010. **285**(6): p. 3857-64.
71. Han, S., et al., *Nuclear envelope phosphatase 1-regulatory subunit 1 (formerly TMEM188) is the metazoan Spo7p ortholog and functions in the lipin activation pathway.* J Biol Chem, 2012. **287**(5): p. 3123-37.
72. Santos-Rosa, H., et al., *The yeast lipin Smp2 couples phospholipid biosynthesis to nuclear membrane growth.* EMBO J, 2005. **24**(11): p. 1931-41.
73. Kok, B.P.C., et al., *Conserved Residues in the N Terminus of Lipin-1 Are Required for Binding to Protein Phosphatase-1c, Nuclear Translocation, and Phosphatidate Phosphatase Activity*, in *Journal of Biological Chemistry*2014. p. 10876-10886.
74. Liu, G.-H. and L. Gerace, *Sumoylation regulates nuclear localization of lipin-1alpha in neuronal cells.*, in *PLoS ONE*2009. p. e7031.
75. Harris, T.E., et al., *Insulin controls subcellular localization and multisite phosphorylation of the phosphatidic acid phosphatase, lipin 1.* J Biol Chem, 2007. **282**(1): p. 277-86.
76. Huffman, T.A., I. Mothe-Satney, and J.C. Lawrence, Jr., *Insulin-stimulated phosphorylation of lipin mediated by the mammalian target of rapamycin.* Proc Natl Acad Sci U S A, 2002. **99**(2): p. 1047-52.
77. Eaton, J.M., et al., *Lipin 2 binds phosphatidic acid by the electrostatic hydrogen bond switch mechanism independent of phosphorylation.* J Biol Chem, 2014. **289**(26): p. 18055-66.
78. Reue, K. and D.N. Brindley, *Thematic Review Series: Glycerolipids. Multiple roles for lipins/phosphatidate phosphatase enzymes in lipid metabolism*, in *J. Lipid Res.*2008. p. 2493-2503.
79. Scott, S.A., et al., *Chemical modulation of glycerolipid signaling and metabolic pathways.*, in *Biochim. Biophys. Acta*2014. p. 1060-1084.
80. Squillace, N., et al., *Changes in subcutaneous adipose tissue microRNA expression in HIV-infected patients.* J Antimicrob Chemother, 2014. **69**(11): p. 3067-75.
81. Chuang, C., et al., *Inactivation of the host lipin gene accelerates RNA virus replication through viral exploitation of the expanded endoplasmic reticulum membrane.* PLoS Pathog, 2014. **10**(2): p. e1003944.
82. Cortes, J., et al., *An unusually large multifunctional polypeptide in the erythromycin-producing polyketide synthase of Saccharopolyspora erythraea.* Nature, 1990. **348**(6297): p. 176-8.
83. Donadio, S., et al., *Modular Organization of Genes Required for Complex Polyketide Biosynthesis.* Science, 1991. **252**(5006): p. 675-679.
84. Schwecke, T., et al., *The Biosynthetic Gene-Cluster for the Polyketide Immunosuppressant Rapamycin.* Proceedings of the National Academy of Sciences of the United States of America, 1995. **92**(17): p. 7839-7843.
85. Molnar, I., et al., *The biosynthetic gene cluster for the microtubule-stabilizing agents epothilones A and B from Sorangium cellulosum So ce90.* Chemistry & Biology, 2000. **7**(2): p. 97-109.

86. Staunton, J. and K.J. Weissman, *Polyketide biosynthesis: a millennium review.* Natural Product Reports, 2001. **18**(4): p. 380-416.
87. Keatinge-Clay, A.T., *The structures of type I polyketide synthases*, in *Nat Prod Rep*2012. p. 1050.
88. Fischbach, M.A. and C.T. Walsh, *Assembly-line enzymology for polyketide and nonribosomal Peptide antibiotics: logic, machinery, and mechanisms.* Chemical Reviews, 2006. **106**(8): p. 3468-96.
89. Jenke-Kodama, H. and E. Dittmann, *Evolution of metabolic diversity: insights from microbial polyketide synthases.* Phytochemistry, 2009. **70**(15-16): p. 1858-66.
90. Hertweck, C., *The biosynthetic logic of polyketide diversity.* Angew Chem Int Ed Engl, 2009. **48**(26): p. 4688-716.
91. Gulder, T.A., M.F. Freeman, and J. Piel, *The Catalytic Diversity of Multimodular Polyketide Synthases: Natural Product Biosynthesis Beyond Textbook Assembly Rules.* Top Curr Chem, 2011.
92. Khosla, C., et al., *Tolerance and specificity of polyketide synthases.* Annual Review of Biochemistry, 1999. **68**: p. 219-53.
93. Watanabe, K. and H. Oikawa, *Robust platform for de novo production of heterologous polyketides and nonribosomal peptides in Escherichia coli.* Org Biomol Chem, 2007. **5**(4): p. 593-602.
94. McDaniel, R., et al., *Rational design of aromatic polyketide natural products by recombinant assembly of enzymatic subunits.* Nature, 1995. **375**(6532): p. 549-54.
95. Cane, D.E., C.T. Walsh, and C. Khosla, *Harnessing the biosynthetic code: combinations, permutations, and mutations.* Science, 1998. **282**(5386): p. 63-8.
96. Rix, U., et al., *Modification of post-PKS tailoring steps through combinatorial biosynthesis.* Natural Product Reports, 2002. **19**(5): p. 542-80.
97. McDaniel, R., et al., *Multiple genetic modifications of the erythromycin polyketide synthase to produce a library of novel "unnatural" natural products.* Proc Natl Acad Sci U S A, 1999. **96**(5): p. 1846-51.
98. Liou, G.F. and C. Khosla, *Building-block selectivity of polyketide synthases.* Current Opinion in Chemical Biology, 2003. **7**(2): p. 279-84.
99. Piel, J., et al., *Antitumor polyketide biosynthesis by an uncultivated bacterial symbiont of the marine sponge Theonella swinhoei.* Proc Natl Acad Sci U S A, 2004. **101**(46): p. 16222-7.
100. Wong, F.T., et al., *Structure and mechanism of the trans-acting acyltransferase from the disorazole synthase.* Biochemistry, 2011. **50**(30): p. 6539-48.
101. Saito, J., et al., *Crystal structure of enoyl-acyl carrier protein reductase (FabK) from Streptococcus pneumoniae reveals the binding mode of an inhibitor.* Protein Sci, 2008. **17**(4): p. 691-9.
102. Massengo-Tiasse, R.P. and J.E. Cronan, *Diversity in enoyl-acyl carrier protein reductases.* Cell Mol Life Sci, 2009. **66**(9): p. 1507-17.
103. Mootz, H.D. and M.A. Marahiel, *The tyrocidine biosynthesis operon of Bacillus brevis: complete nucleotide sequence and biochemical characterization of functional internal adenylation domains.* J Bacteriol, 1997. **179**(21): p. 6843-50.
104. Marahiel, M.A., M.M. Nakano, and P. Zuber, *Regulation of peptide antibiotic production in Bacillus.* Mol Microbiol, 1993. **7**(5): p. 631-6.
105. Jenner, M., et al., *Substrate Specificity in Ketosynthase Domains from trans-AT Polyketide Synthases.* Angewandte Chemie-International Edition, 2013. **52**(4): p. 1143-1147.
106. Tae, H., E.B. Kong, and K. Park, *ASMPKS: an analysis system for modular polyketide synthases.* BMC Bioinformatics, 2007. **8**: p. 327.
107. Bachmann, B.O. and J. Ravel, *Chapter 8. Methods for in silico prediction of microbial polyketide and nonribosomal peptide biosynthetic pathways from DNA sequence data.* Methods Enzymol, 2009. **458**: p. 181-217.
108. Olsen, J.G., et al., *Structures of beta-ketoacyl-acyl carrier protein synthase I complexed with fatty acids elucidate its catalytic machinery.* Structure, 2001. **9**(3): p. 233-43.

109. Bretschneider, T., et al., *Vinylogous chain branching catalysed by a dedicated polyketide synthase module.* Nature, 2013. **502**(7469): p. 124-8.
110. Jones, D.T., *Protein secondary structure prediction based on position-specific scoring matrices.* J Mol Biol, 1999. **292**(2): p. 195-202.
111. Arnold, K., et al., *The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling.* Bioinformatics, 2006. **22**(2): p. 195-201.
112. Davison, J., et al., *Insights into the function of trans-acyl transferase polyketide synthases from the SAXS structure of a complete module.* Chemical Science, 2014. **5**(8): p. 3081-3095.
113. Dunn, B.J., et al., *Comparative Analysis of the Substrate Specificity of trans- versus cis-Acyltransferases of Assembly Line Polyketide Synthases.* Biochemistry, 2014. **53**(23): p. 3796-3806.
114. Nagano, N., C.A. Orengo, and J.M. Thornton, *One fold with many functions: the evolutionary relationships between TIM barrel families based on their sequences, structures and functions.* Journal of Molecular Biology, 2002. **321**(5): p. 741-65.
115. Fraaije, M.W. and A. Mattevi, *Flavoenzymes: diverse catalysts with recurrent features.* Trends Biochem Sci, 2000. **25**(3): p. 126-32.
116. Ha, J.Y., et al., *Crystal structure of 2-nitropropane dioxygenase complexed with FMN and substrate. Identification of the catalytic base.* Journal of Biological Chemistry, 2006. **281**(27): p. 18660-7.
117. Bukhari, H.S., R.P. Jakob, and T. Maier, *Evolutionary origins of the multienzyme architecture of giant fungal fatty acid synthase.* Structure, 2014. **22**(12): p. 1775-85.
118. Bumpus, S.B., et al., *Polyunsaturated Fatty-Acid-Like Trans-Enoyl Reductases Utilized in Polyketide Biosynthesis*, in *J. Am. Chem. Soc.*2008. p. 11614-11616.
119. Allen, E.E. and D.H. Bartlett, *Structure and regulation of the omega-3 polyunsaturated fatty acid synthase genes from the deep-sea bacterium Photobacterium profundum strain SS9.* Microbiology, 2002. **148**(Pt 6): p. 1903-13.
120. Metz, J.G., et al., *Production of polyunsaturated fatty acids by polyketide synthases in both prokaryotes and eukaryotes.* Science, 2001. **293**(5528): p. 290-3.
121. Kaulmann, U. and C. Hertweck, *Biosynthesis of polyunsaturated fatty acids by polyketide synthases.* Angew Chem Int Ed Engl, 2002. **41**(11): p. 1866-9.
122. Gemperlein, K., et al., *Polyunsaturated fatty acid biosynthesis in myxobacteria: different PUFA synthases and their product diversity.* Chemical Science, 2014. **5**(5): p. 1733-1741.
123. Edwards, A.L., et al., *Architectures of whole-module and bimodular proteins from the 6-deoxyerythronolide B synthase.* J Mol Biol, 2014. **426**(11): p. 2229-45.
124. Weissman, K.J., *Uncovering the structures of modular polyketide synthases.* Nat Prod Rep, 2015. **32**(3): p. 436-53.
125. Gurney, R. and C.M. Thomas, *Mupirocin: biosynthesis, special features and applications of an antibiotic from a gram-negative bacterium.* Appl Microbiol Biotechnol, 2011. **90**(1): p. 11-21.
126. Thomas, C.M., et al., *Resistance to and synthesis of the antibiotic mupirocin*, in *Nature Publishing Group*2010, Nature Publishing Group. p. 281-289.
127. Jensen, K., et al., *Polyketide Proofreading by an Acyltransferase-like Enzyme.* Chemistry & Biology, 2012. **19**(3): p. 329-339.
128. McGuffin, L.J., K. Bryson, and D.T. Jones, *The PSIPRED protein structure prediction server.* Bioinformatics, 2000. **16**(4): p. 404-5.
129. Chen, X.H., et al., *Comparative analysis of the complete genome sequence of the plant growth-promoting bacterium Bacillus amyloliquefaciens FZB42.* Nature Biotechnology, 2007. **25**(9): p. 1007-14.
130. Chen, X.H., et al., *Structural and functional characterization of three polyketide synthase gene clusters in Bacillus amyloliquefaciens FZB 42.* Journal of Bacteriology, 2006. **188**(11): p. 4024-4036.
131. Franke, D. and D.I. Svergun, *DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering.* Journal of Applied Crystallography, 2009. **42**: p. 342-346.

132. Kohlhaas, C., et al., *Amino acid-accepting ketosynthase domain from a trans-AT polyketide synthase exhibits high selectivity for predicted intermediate.* Chemical Science, 2013. **4**(8): p. 3212-3217.
133. Perlova, O., et al., *Identification and analysis of the chivosazol biosynthetic gene cluster from the myxobacterial model strain Sorangium cellulosum So ce56.* J Biotechnol, 2006. **121**(2): p. 174-91.
134. Rachid, S., et al., *Deciphering regulatory mechanisms for secondary metabolite production in the myxobacterium Sorangium cellulosum So ce56.* Mol Microbiol, 2007. **63**(6): p. 1783-96.
135. Carvalho, R., et al., *The biosynthetic genes for disorazoles, potent cytotoxic compounds that disrupt microtubule formation.* Gene, 2005. **359**: p. 91-8.
136. Gay, D.C., et al., *A Close Look at a Ketosynthase from a Trans-Acyltransferase Modular Polyketide Synthase.* Structure, 2014.
137. Maier, T., et al., *Structure and function of eukaryotic fatty acid synthases*, in *Quart. Rev. Biophys.*2010. p. 373-422.
138. Leibundgut, M., et al., *The multienzyme architecture of eukaryotic fatty acid synthases.*, in *Curr. Opin. Struct. Biol.*2008. p. 714-725.
139. Chen, X.H., et al., *Genome analysis of Bacillus amyloliquefaciens FZB42 reveals its potential for biocontrol of plant pathogens.* J Biotechnol, 2009. **140**(1-2): p. 27-37.
140. Savitsky, P., et al., *High-throughput production of human proteins for crystallization: the SGC experience.* J Struct Biol, 2010. **172**(1): p. 3-13.
141. Betancor, L., et al., *Improved catalytic activity of a purified multienzyme from a modular polyketide synthase after coexpression with Streptomyces chaperonins in Escherichia coli.* Chembiochem, 2008. **9**(18): p. 2962-6.
142. Kapust, R.B., et al., *Tobacco etch virus protease: mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency.* Protein Eng, 2001. **14**(12): p. 993-1000.
143. Kabsch, W., *Integration, scaling, space-group assignment and post-refinement.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 2): p. 133-44.
144. Kabsch, W., *Xds.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 2): p. 125-32.
145. Emsley, P. and K. Cowtan, *Coot: model-building tools for molecular graphics.* Acta Crystallogr D Biol Crystallogr, 2004. **60**(Pt 12 Pt 1): p. 2126-32.
146. Adams, P.D., et al., *PHENIX: building new software for automated crystallographic structure determination.* Acta Crystallogr D Biol Crystallogr, 2002. **58**(Pt 11): p. 1948-54.
147. Blanc, E., et al., *Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT.* Acta Crystallogr D Biol Crystallogr, 2004. **60**(Pt 12 Pt 1): p. 2210-21.
148. Petoukhov, M.V., et al., *New developments in the ATSAS program package for small-angle scattering data analysis.* Journal of Applied Crystallography, 2012. **45**: p. 342-350.
149. Konarev, P.V., et al., *PRIMUS: a Windows PC-based system for small-angle scattering data analysis.* Journal of Applied Crystallography, 2003. **36**: p. 1277-1282.
150. Volkov, V.V. and D.I. Svergun, *Uniqueness of ab initio shape determination in small-angle scattering.* Journal of Applied Crystallography, 2003. **36**: p. 860-864.
151. Schrodinger, L., *The AxPyMOL Molecular Graphics Plugin for Microsoft PowerPoint, Version 1.0.* 2010.
152. Krissinel, E. and K. Henrick, *Inference of macromolecular assemblies from crystalline state.* Journal of Molecular Biology, 2007. **372**(3): p. 774-97.
153. Kearse, M., et al., *Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data.* Bioinformatics, 2012. **28**(12): p. 1647-9.
154. Jain, E., et al., *Infrastructure for the life sciences: design and implementation of the UniProt website.* BMC Bioinformatics, 2009. **10**: p. 136.
155. Goujon, M., et al., *A new bioinformatics analysis tools framework at EMBL-EBI.* Nucleic Acids Res, 2010. **38**(Web Server issue): p. W695-9.
156. Ward, J.J., et al., *The DISOPRED server for the prediction of protein disorder.* Bioinformatics, 2004. **20**(13): p. 2138-9.

157. Han, G.S. and G.M. Carman, *Assaying lipid phosphate phosphatase activities.* Methods Mol Biol, 2004. **284**: p. 209-16.
158. Fitzgerald, D.J., et al., *Protein complex expression by using multigene baculoviral vectors.* Nat Methods, 2006. **3**(12): p. 1021-32.
159. O'Hara, L., et al., *Control of phospholipid synthesis by phosphorylation of the yeast lipin Pah1p/Smp2p Mg2+-dependent phosphatidate phosphatase.*, in *J. Biol. Chem.*2006. p. 34537-34548.