Evolution and Mechanism of Fatty Acid Synthase Multienzymes

Inauguraldissertation

zur

Erlangung der Würde eines Doktors der Philosophie

vorgelegt der

Philosophisch-Naturwissenschaftlichen Fakultät

der Universität Basel

von

Syed Habib Tahir Bukhari

aus Pakistan

Basel, 2015

Genehmigt von der Philosophisch-Naturwissenschaften Fakultät auf Antrag von

Prof. Dr. Timm Maier
Prof. Dr. Roderick Lim

Basel, den 23 Juni 2015

Prof.Dr. Jörg Schibler
Dekan

Summary

Fatty acids are central components of biological membranes, serve as energy storage compounds, and act as second messengers or as covalent modifiers governing protein localization. Biosynthesis of fatty acids uses a conserved mechanism across all species and is carried out in repeated cycles of reactions. In Eukaryotes, these reactions are catalyzed by type I fatty acid synthases (FAS), large architecturally diverse, multienzyme complexes that integrate all steps of fatty acid synthesis into complex biosynthetic assemblies. Two strikingly different types of FAS have emerged in fungi and in animals. The fungal FAS is a rigid, 2.6-MDa barrel-shaped structure with its 48 functional domains embedded in a matrix of scaffolding elements, which comprises almost 50% of the total sequence and determines the emergent multienzymes properties of fFAS. All functional core domains of fFAS are derived from monofunctional bacterial enzymes, but the evolutionary origin of the scaffolding elements remains enigmatic. In the first part of the thesis using a combined phylogenetic and structural biology approach we have identified two bacterial protein families of non-canonical fatty acid biosynthesis starter enzymes and trans-acting polyketide enoyl reductases (ER) as potential ancestors of core scaffolding regions in fFAS. The architectures of both protein families are revealed by representative crystal structures of the starter enzyme FabY and DfnA-ER. In both families, a striking structural conservation of insertions to scaffolding elements in fFAS is observed, despite marginal sequence identity. The combined phylogenetic and structural data provide first insights into the evolutionary origins of the complex multienzyme architecture of fFAS.

In contrast structural and evolutionarily analysis revealed that animal FAS is related to polyketide synthase type I (PKS I), which is utilized by bacteria to synthesize a broad spectrum of secondary metabolites. Animal FAS is

an open X-shaped structure with catalytic domains not interrupted by the insertion of scaffolding elements but connected to each other via short not conserved linker sequences. Crystallographic data together with biochemical and electron microscopy (EM) analysis indicate that animal FAS displays an extraordinary degree of flexibility to ensure productive interactions between the active sites during the reaction cycle. Conformational changes most likely result from a combination of internal domain flexibility in the linker regions, which connects individual domains in the animal FAS. The second part of the thesis is thus dedicated to investigating how intra domain linking influences catalytic properties and conformational crosstalk between domains. This was achieved by generating more then 40 different constructs with various linker lenths. Combined structural and kinetic data from purified constructs helped us to better understand the emergent properties of the megasynthase system. A long-term goal is to use these insights for the construction of artificial multienzymes incorporating complete and complex molecular pathways.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ACC	Acetyl-CoA carboxylase
ACP	Acyl carrier protein
AcPS	holoACP-synthase
ATP	Adenosine triphosphate
C14	14-carbon fatty acid
C16	16-carbon fatty acid
C18	18-carbon fatty acid
C20	20-carbon fatty acid
CMN-FAS	*Corynebacteria*, *Mycobacteria*, and *Nocardia* Fatty acid synthase
CoA	Coenzyme- A
DfnA	Difficidin biosynthesis cluster A
DfnA-ER	The enoyl reductase domain of DfnA
DH	Dehydratase
E.coli	*Escherichia coli*
EM	Electron microscopy
ER	Enoylreductase
FAS	Fatty acid synthase
fFAS	Fungal FAS
FMN	Flavin mononucleotide
G3P	Glycerol-3-phosphate
Hex A/B	Hexanoic acid synthase
KR	β-ketoreductase
KS	β-ketoacyl synthase
LFCA	Last fungal common ancestor
LPA	Lysophosphatidic acid
MAT	Malonyl-/acetyl-transferase
mFAS	mammalian Fatty Acid Synthase
MPT	Malonyl/palmitoyl transferase domain
mtFAS	Mitochondrial Fatty Acid Synthase
P-PAN	4-phosphopantetheine
PA	Phosphatidic acid
pKR	Pseudo-ketoreductase
PKS I	Polyketide synthase type I
pME	Pseudo-methyltransferase
PPT	Phosphopantetheine transferase

Introduction

In the beginning of the early 20th century it was considered that fatty acids have only two functions- serve as a source of calories and as building blocks for membranes[1]. In 1929, George and Mildred Burr published two papers, where they demonstrated that fatty acids were an essential dietary constituent[2,3]. In their experiment they kept rats on strict diet and noticed that if fatty acids were omitted from the food, a deficiency syndrome ensued that often led to death[4]. After this many other research groups were able to show that fatty acids and their metabolites possess very unique biological roles that is distinctive from its function as a source of energy or as a simple construction unit[5-7]. A wide range of cellular processes are dependent on fatty acids, from the biosynthesis of essential cellular structural components (membrane phospholipids, lipoproteins, and lipoglycans) and cofactors (lipoate and biotin) to energy storage reserves[8-10]. Fatty acids participate as components of signal transduction pathways and as docking sites for cytoplasmic signaling proteins such as kinases[11]. Polyunsaturated fatty acids containing two or more carbon–carbon double-bonds are important as constituents of the phospholipids, where they appear to confer distinctive properties to the membranes, in particular by decreasing their rigidity[12].

Storage lipids, such as triacylglycerols, are deposited as fat droplets in large amounts in vertebrate fat cells[13]. These droplets are surrounded by a protective monolayer of phospholipids and biologically active hydrophobic proteins. Triacylglycerols are released when required by hydrolysis reactions catalyzed by lipases under the influence of hormones[14]. Subsequent oxidation of triacylglycerols produces more than twice the energy (9 kcal/g) as the oxidation of carbohydrates (4 kcal/g)[15].

Biosynthesis of fatty acids uses a conserved mechanism across all species and it is carried out in repeated cycles of reactions. In Eukaryotes, these reactions are catalyzed by type I fatty acid synthase (FAS), a large architecturally diverse, multienzyme complexes that integrate all steps of fatty acid synthesis into complex biosynthetic assemblies[16,17]. In contrast, in dissociated type II FAS system, proteins are all expressed as individual polypeptides from separate genes, these systems are found mostly in bacteria but also in eukaryotic organelles such as mitochondria and plastids[18]. In the following section I will compare the fundamentally distinct organization of different FAS systems and examine the structural and chemical principals of enzyme reactions.

Bacterial FAS

The biosynthesis of fatty acids is the first step in the formation of membrane lipids and it is essential for all bacterial cells. It involves more than ten separately expressed genes and proteins, which are abundant in the bacterial cytosol. Central to this process is the acyl carrier protein (ACP), a cofactor protein that covalently binds all fatty acyl intermediates[19]. ACP is one of the most expressed protein in *E. coli* and is converted to its active holo-form by holoACP-synthase (AcpS) which transfers the 4-phosphopantetheine (P-PAN) prosthetic group from CoA to apo-ACP[20,21] (Figure 1). Activated holo-ACP then enters fatty acid biosynthesis cycle, which consist of four stages; Initiation, chain elongation, chain reduction and termination.

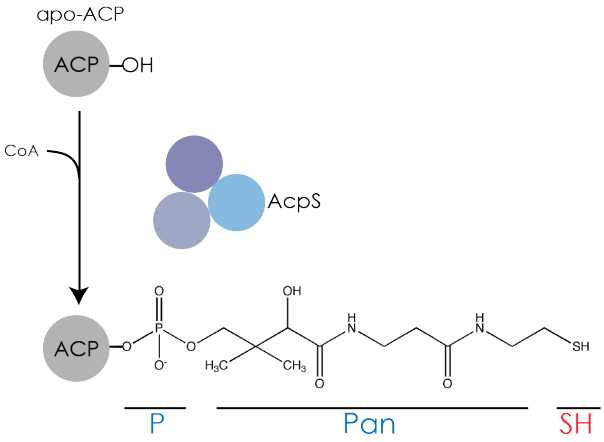


Figure 1. Activiation of ACP by AcpS. Trimeric AcpS attaches phosphopantetheine group from CoA on to a serine residue of ACP in a conserved Asp-Ser-Leu motif. The resulting terminal sulfhydryl group of the phosphopantetheine arm is then used to bind all the growing fatty acid intermediate in a covalent high-energy thioester bond.

Initiation of fatty acid biosynthesis. The first step in fatty acid biosynthesis is the carboxylation of acetyl-CoA by acetyl-CoA carboxylase (ACC) to form the universal extender unit malonyl-CoA **(Figure 2)**. The overall ACC reaction requires a biotin cofactor, adenosine triphosphate (ATP) and the coordinated action of four gene products, AccA, AccB, AccC, and AccD[22].

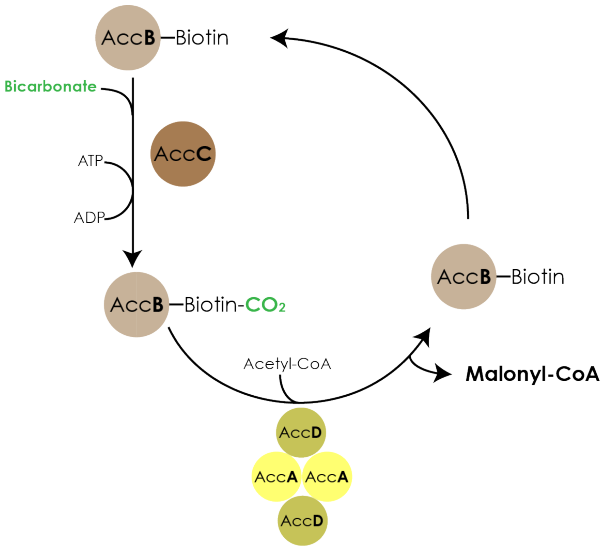


Figure 2. Biosynthesis of Malonyl-CoA by ACC. AccC catalyzes the ATP-dependent transfer of CO2 from bicarbonate to biotin attached to AccB. AccB then shuttles the carboxy-biotin intermediate to the tran-scarboxylase enzyme composed of AccA and AccD subunits where carboxyl group is transferred from biotin to acetyl CoA to form malonyl CoA.

Chain elongation. The elongation step is initiated by the Claisen condensation of malonyl- ACP with an acyl-CoA, catalyzed by the condensing enzyme, the ß-ketoacyl-ACP synthase III or FabH, to form ß-ketoacyl-ACP[23] (Figure 3 B). FabH of *E. coli*, produces mainly linear fatty acids, because it has specificity only for acetyl-CoA, in contrast Gram-positive bacteria utilize special FabH enzymes which can choose as a first building block larger branched-chain substrates[24].

Chain reduction. The Chain reduction cycle consists of three core enzyme activities that progressively reduce the acyl chain attached to ACP through each round (Figure 3 C). First, the NADPH-dependent ß-ketoacyl-ACP reductase, or FabG, reduces the ß-keto group to a ß-hydroxyl intermediate[25]. Second, two isoforms, FabA and FabZ, catalyze the dehydration of β -hydroxyacyl-ACPs, *albeit* with different substrate specificities [26]. The third step involves the reduction of the enoyl chain by the NADH-dependent FabI[27]. Gramm-negative bacteria utilize alternative enoyl-ACP reductase — the flavoprotein FabK[28]. The fully reduced acyl-ACP chain functions as a starter substrate for the next round of

elongation, which is initiated by an elongation condensing enzyme: FabF or FabB (Figure 3C). The FabF isoform is universally expressed, but some bacteria utilize the FabB enzyme, which is used for condensing unsaturated fatty acids[26,29].

During each round of the condensation reactions, the acyl chain is detached from ACP and binds to the cysteine residue in the active sites of FabH, FabB or FabF (Figure 3)[18]. An extender malonyl-ACP then enters the active site and the acyl chain is added to the carboxyl end of the malonyl unit, which loses a CO_2 group in the process. Therefore, the acyl chain is constructed 'inside out' as the additional carbon groups are added to the base of the acyl chain. The cycle is repeated until the acyl chain reaches 16–18 carbon groups in length, at which point the vast majority of acyl-ACPs are utilized in membrane biosynthesis (Figure 3D)[30].

Transfer of fatty acids to the membrane. The most ubiquitous system is the PlsX–PlsY pathway, which is found in all but one family of proteobacteria[30]. First, PlsX, a peripheral membrane protein, transfers the acyl group from the long-chain acyl-ACP end product of the elongation pathway (Figure 3, blue arrow) to inorganic phosphate to form a reactive acyl-phosphate intermediate (Figure 4D). This is then attached to glycerol-3-phosphate (G3P) to form acyl-glycerol-3- phosphate (LPA; lysophosphatidic acid) by the acyltransferase membrane protein PlsY. Another acyltransferase, PlsC then adds a second acyl chain to the 2-position of LPA to form phospatidic acid (PA)[31]. PA represents the fundamental building block

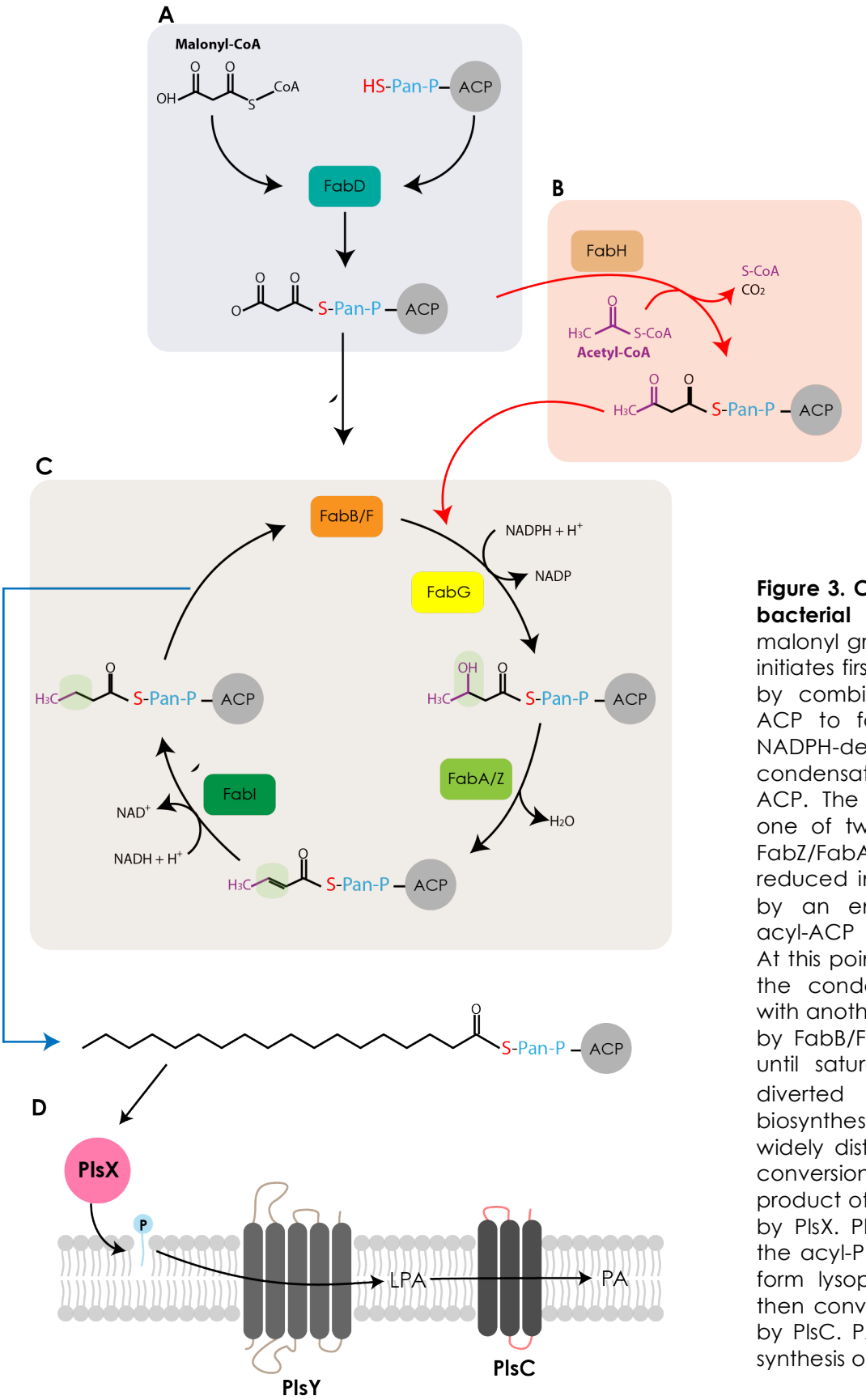


Figure 3. Catalytic reaction cycle of type II bacterial FAS. (**A**) FabD transfers the malonyl group from CoA to ACP. (**B**) FabH initiates first cycles of fatty acid elongation by combining acetyl-CoA with malonyl-ACP to form acetoacetyl-ACP. (**C**) The NADPH-dependent FabG reduces the condensation product to β-hydroxyacyl-ACP. The hydroxyl group is removed by one of two β- hydroxyacyl dehydratases FabZ/FabA. The double bond is then reduced in an NADH-dependent reaction by an enoylreductase FabI generating acyl-ACP extended by two carbon units. At this point the cycle starts again through the condensation reaction of acyl-ACP with another malonyl-ACP group catalyzed by FabB/F. This is repeated multiple times until saturated C_{16} or C_{18} acyl-ACP is diverted for utilization in membrane biosynthesis (blue arrow). (**D**) The most widely distributed pathway starts with the conversion of a long-chain acyl-ACP end product of fatty acid synthesis to an acyl-P by PlsX. PlsY transfers the fatty acid from the acyl-P to glycerol-3-phosphate G3P to form lysophosphatidic acid (LPA). LPA is then converted to phosphatidic acid (PA) by PlsC. PA is the key intermediate in the synthesis of all membrane glycerolipids.

from which the phospholipids are derived, giving rise to phosphatidylserine, phosphatidylethanolamine and phosphatidyl-glycerol.

Regulation of fatty acid synthesis. The primary pathway for the regulation of fatty acid synthesis in *E. coli* is through feedback inhibition by long-chain acyl-ACPs, which affects three enzymes: ACC, FabH and FabI. Inhibition of ACC limits the supply of malonate groups for chain initiation and elongation (Figure 2)[32]. Regulation of FabH prevents the initiation of new acyl chains and limits the total number of fatty acids that are produced (Figure 3B)[33]. Finally, FabI catalyzes the reduction of enoyl-ACP, which is critical for the completion of the acyl chain elongation cycle; a reduction in FabI activity slows the rate of fatty acid elongation (Figure 3C)[34].

Mitochondrial FAS

The mitochondrial FAS (mtFAS) produces short-chain fatty acids, which are essential for the structure, dynamics and enzymatic function of inner and outer mitochondrial membranes[35-37]. Any irregularity in enzymes, responsible for biosynthesis of mitochondrial fatty acids in eukaryotes results in respiratory incompetence, abnormal morphology and cell deaths[38-41].

mtFAS diverges from the cytosolic FAS as it is of the type II dissociated organization and many of the proteins are highly homologous to their bacterial counterparts, but nevertheless the eukaryotic type II systems do have three distinguishing features[36]. First, the prokaryotic systems utilize three different β-ketoacyl synthases (FabH, FabB and FabF) with divergent substrate specificities that range from 2- to 16-carbon; the mtFAS only has one KS and predominantly produces fatty acids of 14 or less carbon atoms[18,42]. Second, the bacterial type II system has a FabD, dedicated enzyme which directly utilizes acetyl-CoA in the initiation step[43]. In contrast, the mitochondrial type II system appears to generate the acetyl primer by decarboxylation of malonyl moieties at the β-ketoacyl synthase[35,42]. Third, the mtFAS proteins that catalyze the final two steps of the fatty acid elongation cycle, Htd2 and 2-enoyl-ACP reductase (Etr1), do not share clear sequence similarities to prokaryotic FAS type II enzymes and structurally belong to different protein classes[42].

Animal FAS

Eukaryotic type I fatty acid synthases (FAS) are giant multifunctional proteins. Various evolutionary processes such as gene duplication and gene fusion led to the emergence of the type I FASs[45]. In animal cells fatty acid synthesis is catalyzed by a single 540kDa homodimeric multienzyme with a characteristic X-shape (Figure 4)[17,46,47]. Based on its X-ray crystal structure, mammalian Fatty Acid Synthase (mFAS) is divided into a lower condensing portion containing the β-ketoacyl synthase (KS), malonyl-/acetyl-transferase (MAT) domains and an upper β-carbon modification section, consisting of the enzymatic dehydratase (DH), enoylreductase (ER), and β-ketoreductase (KR) domains[17] (Figure 4B and C). The upper part also possesses two additional non-enzymatic domains, a pseudo-ketoreductase (pKR) and a pseudo-methyltransferase (pME), according to their structural homology with active KR and ME enzymes[48]. All reaction intermediates, like in the bacterial FAS system, are covalently bound to an ACP, which translocates between the active sites during catalysis. Reaction products are released from the ACP as free fatty acids by the thioesterase (TE) domain (Figure 4D). Interestingly, both ACP and TE domains were found disordered in the crystal structure and could not be visualized due to their flexibility[48]. Structurally the KS, KR and MAT domains are homologs of their bacterial functional counterparts FabB, FabD and FabG[48-50]. The DH domain adopts a pseudo-dimeric fold, distantly resembling the bacterial homo-dimeric FabA[48].

A key feature distinguishing the type I FASs and type II counterparts is the presence of discrete connecting regions between the active domains. In the porcine FAS 9% of total sequence is invested in

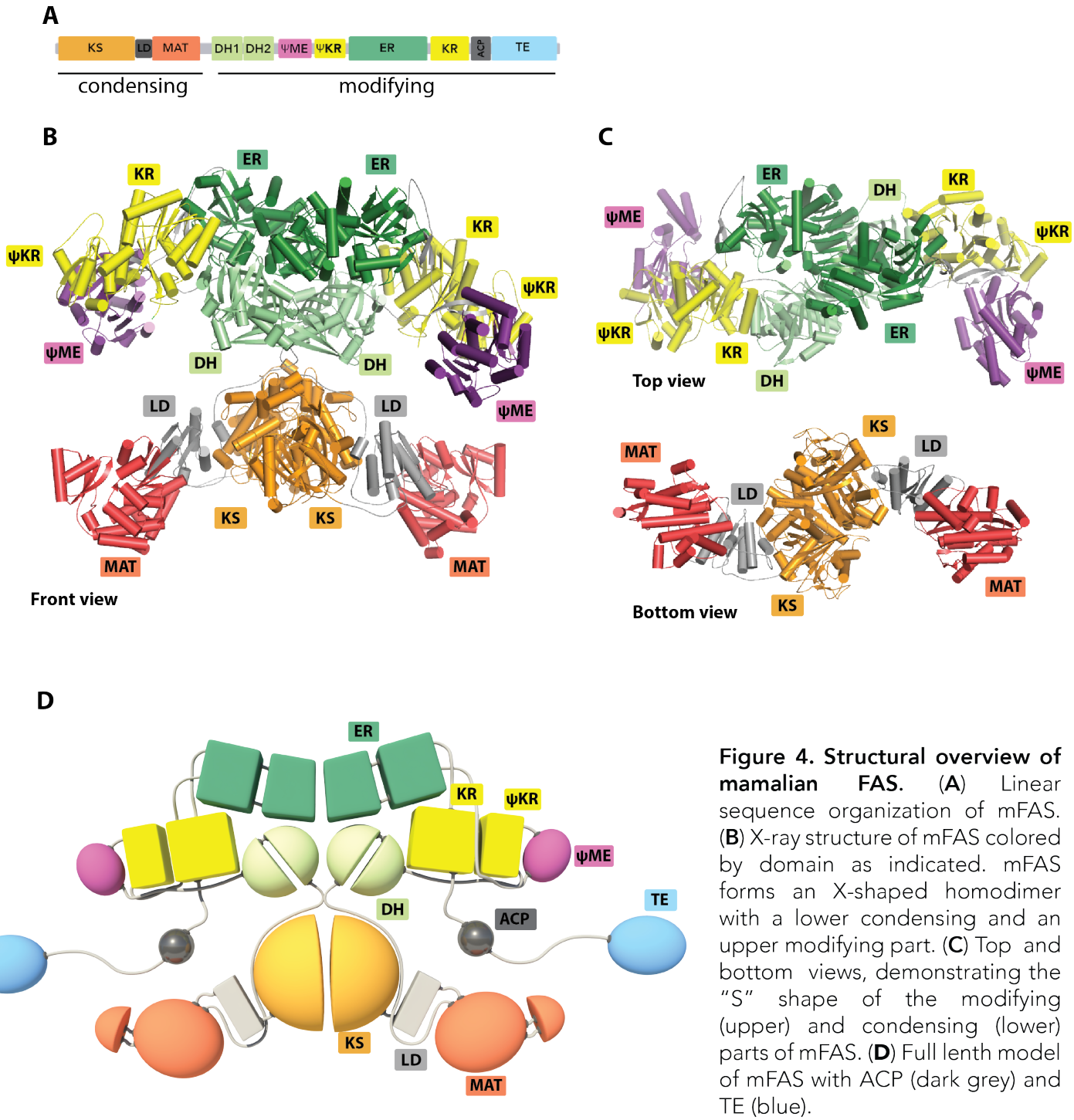


Figure 4. Structural overview of mamalian FAS. (**A**) Linear sequence organization of mFAS. (**B**) X-ray structure of mFAS colored by domain as indicated. mFAS forms an X-shaped homodimer with a lower condensing and an upper modifying part. (**C**) Top and bottom views, demonstrating the "S" shape of the modifying (upper) and condensing (lower) parts of mFAS. (**D**) Full lenth model of mFAS with ACP (dark grey) and TE (blue).

the form of solvent exposed linkers[48,51]. **The importances of the linking regions are discussed in part II.** Despite differences in the overall

organization of the mFAS, the enzymatic reactions and mechanism of *de novo* fatty acid synthesis are essentially identical to dissociative bacterial FAS system (Figure 5). For example, exactly the same intermediates and reactions are present in the elongation cycle. However, some enzymatic differences do occur in the stage of initiation, condensation and termination.

The MAT domain transfers acetyl- and malonyl-CoA. The Bacterial FAS system utilizes two dedicated enzymes, FabH and FabB, for transferring acetyl or malonyl-CoA [18,52] (Figure 5B, step 1 and 4). In contrast the animal FAS contain a single MAT enzyme for loading both the Acetyl and Malonyl-CoA units on to the ACP[46]. The choice of substrate loaded is entirely random acetyl and malonyl moieties are rapidly exchanged between CoA and FAS. If the applicable substrate is loaded, then a productive reaction can follow, otherwise the inappropriately loaded substrate is transferred back to CoA, which must be present at all times during the FAS reaction to ensure efficient substrate sorting[4]. Scavenging of CoA from the assay incubation mixture halts fatty acid synthesis[53,54].

TE and KS determine the chain length of fatty acid products. Two enzymatic domains determine the chain length in mFAS. First, during the elongation cycle ACP transfers growing fatty acid chains to the active site cysteine of the KS domain. For fatty acids containing up to 16 carbons this transfer is very rapid and occur in less then 1 second, however for chains containing 18, 20 or 22 C atoms this process requires several minutes[55]. Secondly, TE has very limited activity toward substrates with less than 16 carbon atoms [56]. Thus the specificities of the chain elongation and chain termination steps complement each other perfectly

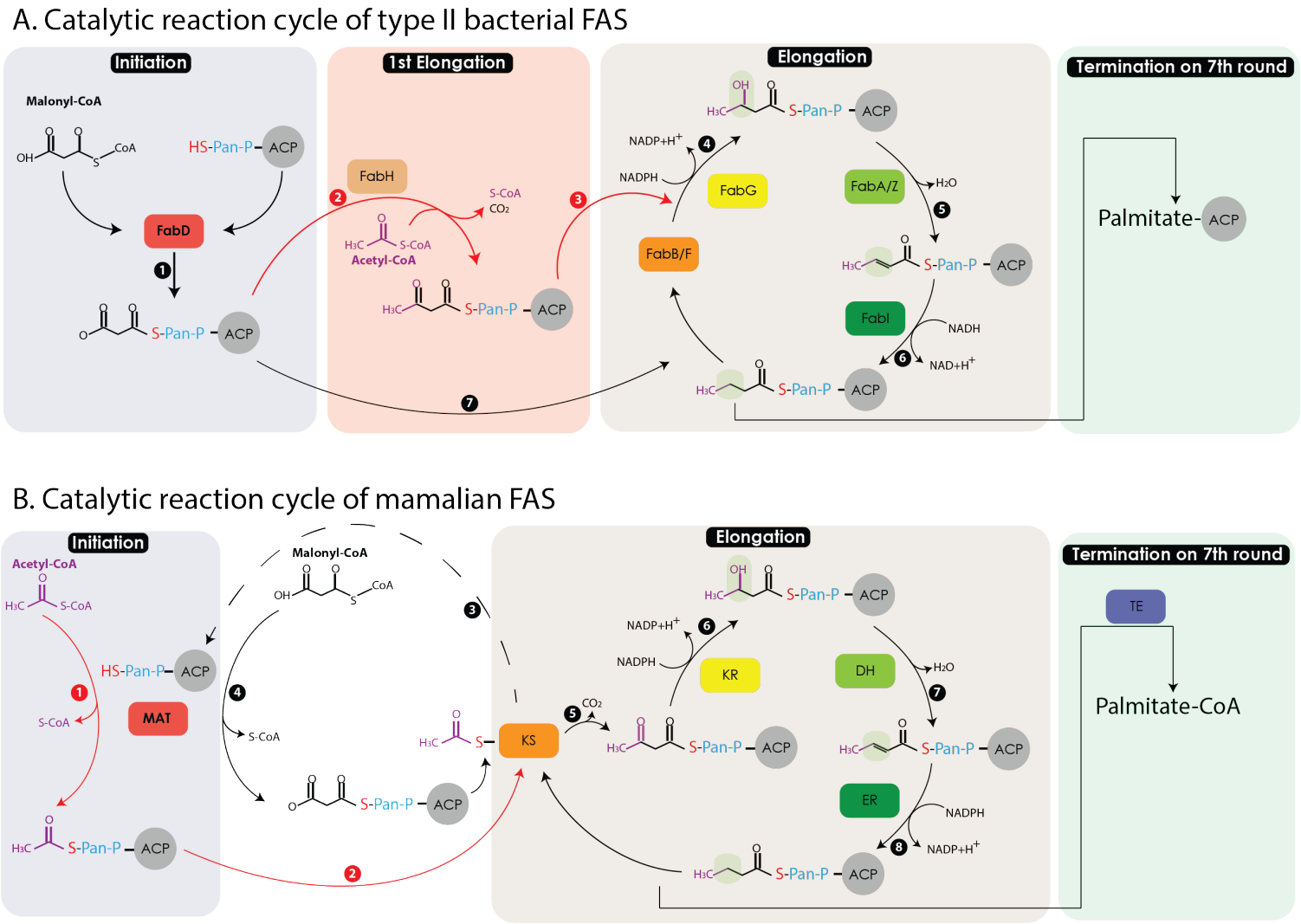


Figure 5. Comparison of the catalytic reaction cycle of type II bacterial FAS and mammalian FAS. Red arrows indicate steps that are repeated only once; black arrow shows reactions that are redone multiple times. Proteins sharing enzymatic functions in animal and type II FAS are colored in identical color. (**A**) Catalytic cycle of bacterial FAS. Bacteria utilize a specialized enzyme (FabH) for the initiation step in acyl chain formation and uses an acyl-CoA as a primer to condense with malonyl-ACP (step 2 and 3). For more details refer to **Figure 3**. (**B**) Catalytic reaction cycle of mammalian FAS. The reaction cycle of FAS is initiated by the transfer of the acyl moiety of the starter substrate acetyl-CoA to the ACP (step 1) catalyzed by the dual-specific malonyl/acetyl transferase (MAT). ACP then transfers the acetyl group to the active cysteine on the KS (step 2). In the next step the elongation unit malonyl-CoA is loaded onto ACP by MAT (step 3 and 4). The ß-ketoacyl synthase (KS) catalyzes the decarboxylative condensation of the acyl intermediate with malonyl-ACP (step 5). The product is further modified at the ß- carbon position by β-ketoreductase (KR) (step 6), dehydratase (DH) (step 7) and enoyl reductase (ER) (step 8) to yield a four carbon acyl substrate for further cyclic elongation with two-carbon units derived from malonyl-CoA (step 4). After seven rounds of elongation, the end product is released from the enzyme as free fatty acid by a thiosterase (TE)

to ensure that the main product released from the FAS is the 16 C atom fatty acid.

Fungal FAS

Yeast FAS, a member of the fungal type I FAS family, contains six copies of eight independent functional domains in an α6ß6 molecular complex of 2.6 MDa[57] (Figure 6). Each of the α and ß subunits accommodates four functional domains. The ß-chain carries the AT, ER, DH domain, and the largest part of the malonyl/palmitoyl transferase domain (MPT). The remaining half of the MPT, the double-tethered ACP, the KR, the KS and the phosphopantetheine transferase (PPT) are encoded by the α-chain[16,58,59] (Figure 6, A and B). These eight functional domains catalyze all reactions required for synthesis of fatty acids in yeast: activation, priming, multiple cycles of elongation, and termination[60]. The assembled fFAS adopts a barrel-shaped formation with two domes separated by a central wheel structure[16,44,57,58,60] (Figure 6 D). Each dome contains three full sets of enzymatic domains and three double-tethered ACP domains for substrate transfer (Figure 7A)[60,61].

Fungal FAS invests nearly 50% of its absolute sequence length into building scaffolding elements, which are mainly inserted of conserved enzymatic domains[16,62] (Figure 7B). These sequences are not directly involved in catalysis but instead dictate the architectural interactions and define the arrangement of the catalytic domains [16,58,60]. Despite considerable differences in the overall organization of fungal FAS, the enzymatic reactions and mechanism of *de novo* fatty acid synthesis are essentially identical to dissociative bacterial and X shaped mammalian FAS systems[18,50,57]. However, some differences occur in the activation of ACP, elongation and termination stages.

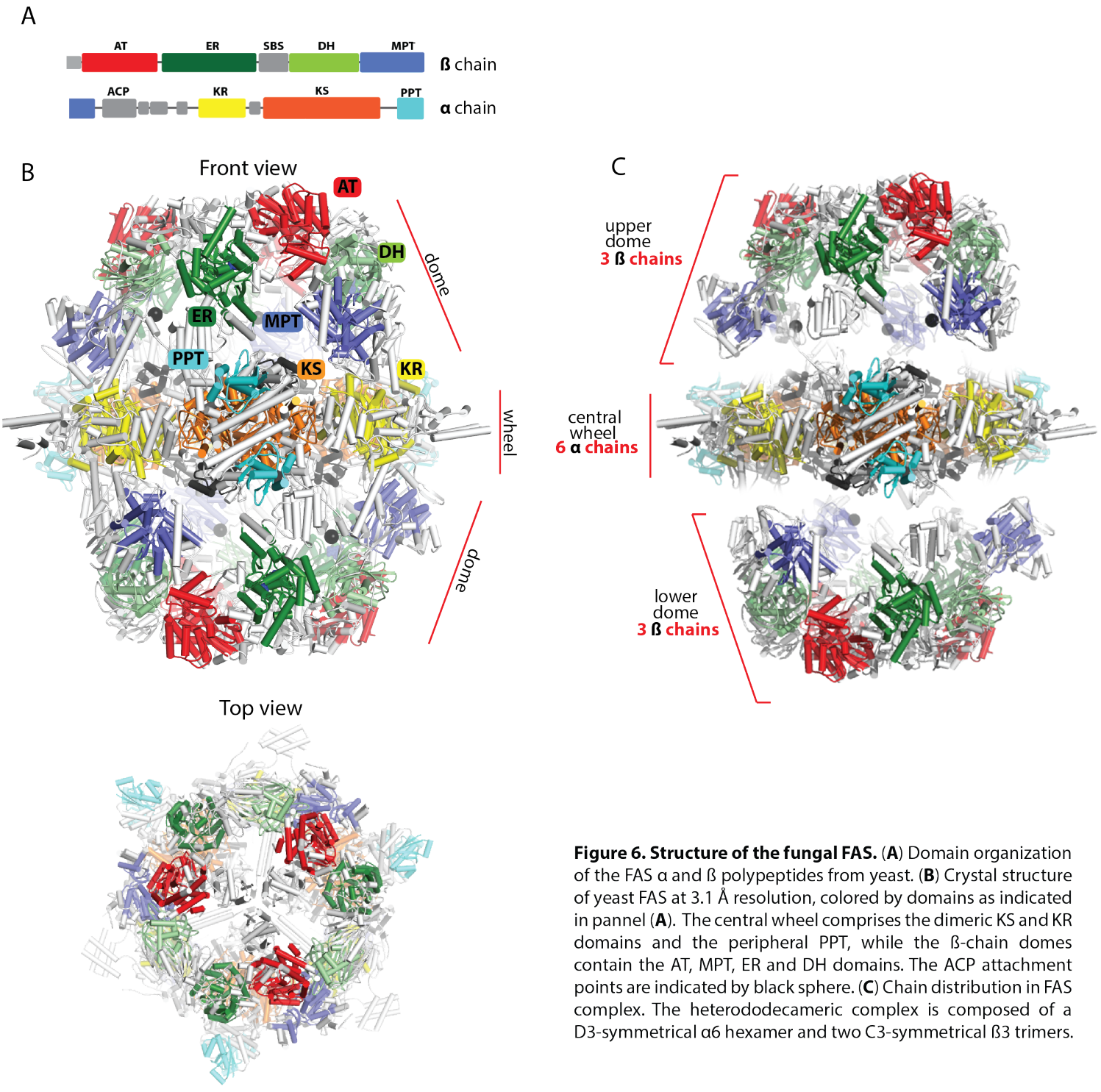


Figure 6. Structure of the fungal FAS. (**A**) Domain organization of the FAS α and ß polypeptides from yeast. (**B**) Crystal structure of yeast FAS at 3.1 Å resolution, colored by domains as indicated in pannel (**A**). The central wheel comprises the dimeric KS and KR domains and the peripheral PPT, while the ß-chain domes contain the AT, MPT, ER and DH domains. The ACP attachment points are indicated by black sphere. (**C**) Chain distribution in FAS complex. The heterododecameric complex is composed of a D3-symmetrical α6 hexamer and two C3-symmetrical ß3 trimers.

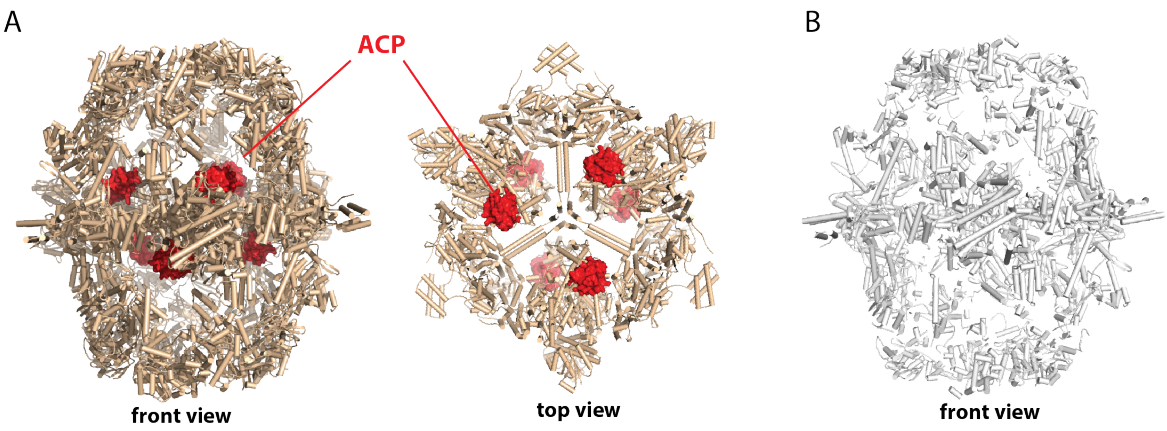


Figure 7. Location the ACP domains. (**A**) The ACP domains, shown as red surfaces, are located inside the fungal FAS barrel (golden color). The active site clefts of the enzymatic domains participating in the fatty acid elongation oriented to interact with the ACP. (**B**) Fungal FAS scaffolding elements (grey) without core enzymatic domains.

Fungal FAS requires a specific activation mechanism. Before the ACP can start to deliver its substrates, it has to be posttranslationally modified by the addition of a P-Pan[63,64]. This activation of the fungal FAS is performed by a specific PPT very similar to bacterial AcpS that covalently attaches the phosphopantetheine moiety of coenzyme A (CoA) onto a conserved serine residue of the ACP[65]. One of the major differences of fungal and mammalian FAS is the mechanism of the posttranslational modification of the ACP domain. A separately expressed PPT enzyme performs the activation of the mammalian FAS, in contrast the fungal FAS PPT domain is fused to the C terminal end of the α chain[66]. This PPT domain is located outside of the barrel, spatially separated from the ACP (Figure 8)[16,44,57-60]. Therefore it is currently not clear how fungal PPT could activate ACP. One possibility is that the fungal FAS auto-activates during the folding events prior to the closure of the reaction chambers[67].

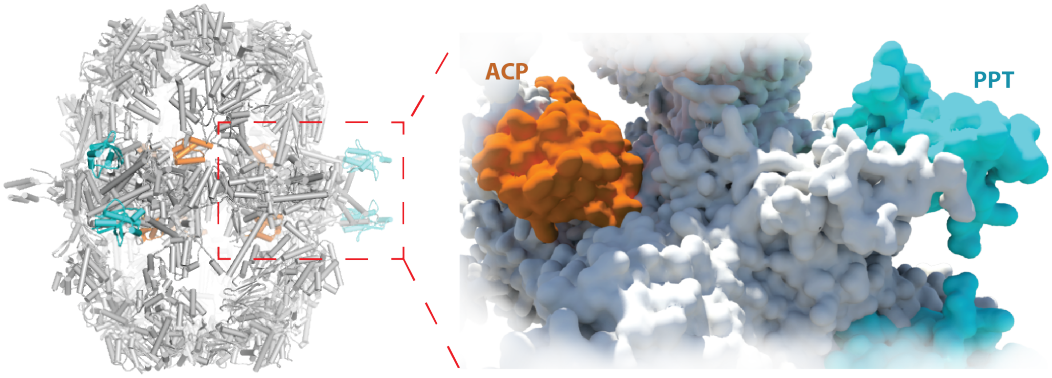


Figure 8. Activation of fungal ACP. Front view of the fungal FAS (right panel) and close-up view of ACP (orange) and PPT (cyan) locations.

Fungal FAS uses a bi-functional MPT for loading and termination. The fFAS harbors no TE domain like in animal FAS, but contains a bi-functional MPT domain instead, which transfers malonyl moieties used for the chain elongation from CoA to ACP and back-transfers saturated C16/C18 products from ACP to CoA[58,60,68] (Figure 9). Two factors determine this unique property of the fungal MPT. First MPT contain a deep hydrophobic pocket, which is optimally suited for binding the hydrophobic C16 tail of palmitate[58,60]. Second during the elongation cycle malonyl- CoA preferentially binds to the active site of MPT. But as soon as growing acyl chain is long enough the binding affinity to the MPT will be strong enough to displace malonyl CoA and allow transferring the mature fatty acid to free CoA[57,68-70].

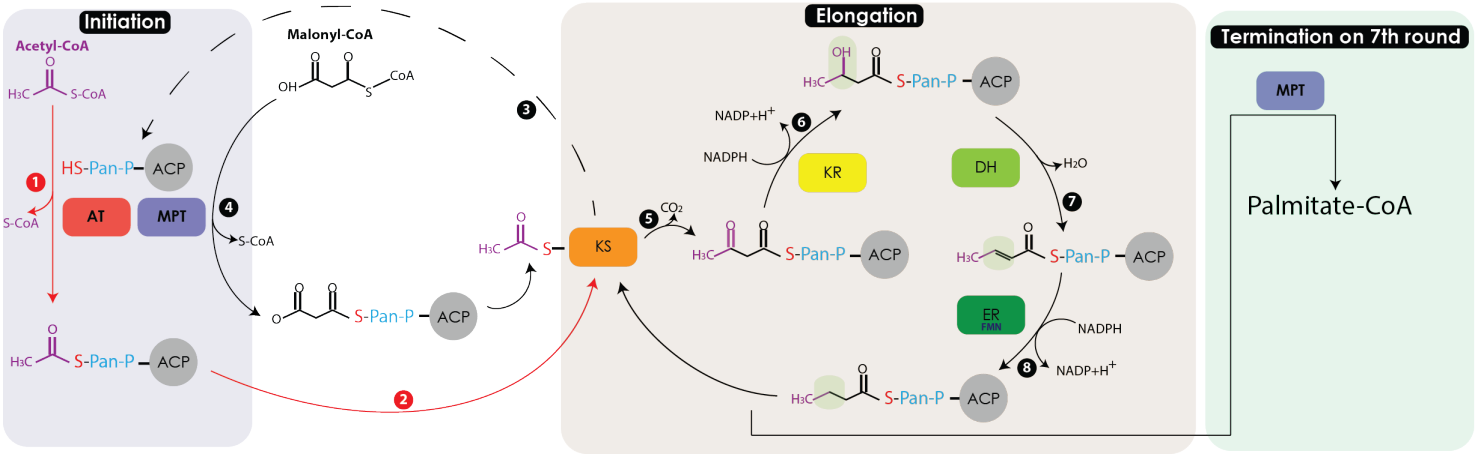


Figure 9. Catalytic reaction cycle of the fungal FAS. Fungal FAS utilizes a dedicated AT domain, which is located on ß-chain and has unique specificity for the priming substrate, acetyl-CoA (step 1). In contrast, mammalian FAS use the MAT domain to load both the priming and the elongating substrate onto ACP (Figure 5**B**, step 1 and 4). In bacterial type II FAS systems, the acetate primer is directly transferred from acetyl-CoA to the ß -ketoacyl-ACP synthase III (FabH) that catalyzes the first condensation reaction in the chain elongation cycle (Figyre 5**A**, step 2). Fungal FAS adopts bi-functional MPT for choosing elongation (step 4) substrate and terminating reaction by transferring 18 carbons fatty acid back to CoA.

Aim of the thesis

Nature developed three types of FAS enzymes, built on completely different architectural principles, but catalyzing highly related series of chemical reaction[18,57]. In the type II FAS, reaction intermediates are covalently attached to the ACP that shuttles substrates between the dissociated enzymatic components[18]. In the multifunctional eukaryotic FAS, ACP forms an integral part of the catalytic machinery resulting in minimized diffusion distances and higher catalytic efficiency[48,60]. The 2.6 MDa barrel shaped fungal FAS integrates all active domains in the rigid scaffolding matrix which comprises almost 50% of the total sequence[71]. Inside the barrel the concentration of ACP and all other active sites is approximately 1 mM ensuring that none of the enzymatic reaction steps are rate-limiting[58]. Substrate shuttling within the fungal FAS happens entirely by 2D diffusion of the double-tethered ACP, without a requirement for large overall conformational changes[61]. All functional domains of fFAS are derived from monofunctional bacterial enzymes, but the evolutionary origin of the scaffolding elements remains enigmatic. The first part of the thesis is therefore focused on finding out the evolutionary origins of scaffolding elements using combined phylogenetic and structural biology approach to better understand the evolutionary process, which led to the development of the fungal FAS.

Structural and evolutionarily analysis revealed that animal FAS is related to polyketide synthase type I (PKS I), which is utilized by bacteria to synthesize a broad spectrum of secondary metabolites [50,72,73]. Animal FAS is an open X-shaped structure with catalytic domains connected to each other via short not conserved linker sequences[17,48]. Crystallographic data together with biochemical and EM analysis indicate that animal FAS displays an extraordinary degree of flexibility to ensure productive interactions between the ACP and the active sites during the reaction cycle (Figure

10)[17,48,50,51]. The nature and dynamic aspects of the substrate shuttling mechanisms in animal FAS are not entirely understood. The second part of the thesis is thus dedicated to investigating how inter-domain linking influences catalytic properties and conformational crosstalk between domains. This will be done by generating more then three dozens of different constructs with systematically increasing or decreasing linker lengths in different areas of animal FAS. Combined structural and kinetic data from purified constructs will help us to better understand the emergent properties of the megasynthase system. A long-term goal is to use these insights for the construction of artificial multienzymes incorporating complete and complex molecular pathways.

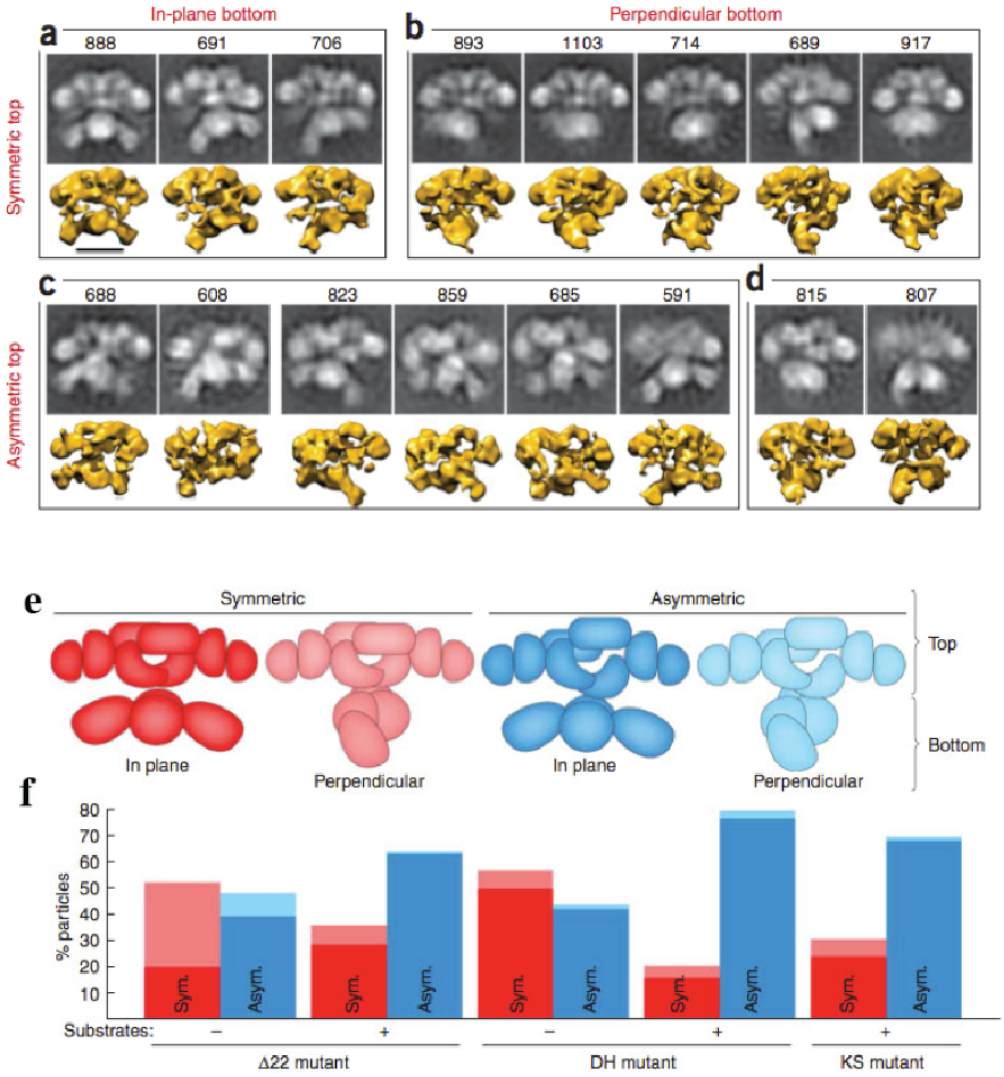


Figure 10. Distribution of animal FAS conformations (adopted from Brignole et al., 2009). (a-d) Class average of single particle images (black and white) with calculated three-dimensional structures (yellow). (e) Cartoon of different FAS arrangements in upper and lower part, red (asymmetric) and blue (symmetric) in the upper ß-carbon-processing, faded color represent perpendicular or in plane conformations of the lower part FAS. (f) Representation % of particles in different conformations, bars are colored according to conformations in (e)

Part I

Evolutionary origins of the multienzyme architecture of giant fungal fatty acid synthase

Habib S.T. Bukhari, Roman P. Jakob and Timm Maier

Author contributions

Timm Maier, Roman P. Jakob and I designed the project. Roman P. Jakob produced DfnA protein, obtained crystals and carried out crystallographic analyses.

I established purification, obtained crystals and determined crystal structure of FabY.

I performed bioinformatic analysis of FabY and DfnA. Timm Maier supervised the project, contributed to crystallographic work and analyzed the results.

Roman P. Jakob, I and Timm Maier wrote and revised manuscript. Manuscript was published in journal Structure (cell press), volume 22, number 12, December 2, 2014

Summary

Fungal fatty acid synthase (fFAS) is a key paradigm for the evolution of complex multienzymes. Its 48 functional domains are embedded in a matrix of scaffolding elements, which comprises almost 50% of the total sequence and determines the emergent multienzymes properties of fFAS. All functional domains of fFAS are derived from monofunctional bacterial enzymes, but the evolutionary origin of the scaffolding elements remains enigmatic. Here, we identify two bacterial protein families of non-canonical fatty acid biosynthesis starter enzymes and trans-acting polyketide enoyl reductases (ER) as potential ancestors of core scaffolding regions in fFAS. The architectures of both protein families are revealed by representative crystal structures of the starter enzyme FabY and DfnA-ER. In both families, a striking structural conservation of insertions to scaffolding elements in fFAS is observed, despite marginal sequence identity. The combined phylogenetic and structural data provide first insights into the evolutionary origins of the complex multienzyme architecture of fFAS.

Introduction

Fatty acids are central components of biological membranes, serve as energy storage compounds, and act as second messengers or as covalent modifiers governing protein localization. In most eukaryotes, their biosynthesis is catalyzed by giant multifunctional enzymes, the fatty acid synthases (Type I FASs) [46,57,74], while bacteria and plants employ a series of monofunctional enzymes (Type II FAS) [18,75,76]. All FAS systems are built upon a conserved set of chemical reactions and enzymatic activities: An acetyl primer and malonyl elongation substrates are loaded from coenzyme A (CoA) to the phosphopantetheinylated acyl carrier protein (ACP) by acetyl- and malonyl-transferases (AT and MT) and are condensed to acetoacetyl-ACP under decarboxylation by ketoacyl synthase (KS). In three subsequent reaction steps, the β-carbon group is processed by ketoacyl reductase (KR), dehydratase (DH), and enoyl-reductase (ER) to yield a saturated acyl-ACP elongated by a two-carbon unit. This product serves as a primer for the next round of elongation and the elongation cycle continues until a chain length of C_{16} or C_{18} is reached. In a terminating step, fatty acids are back-transferred to CoA or released by thioesterase (TE). The eukaryotic Type I FAS integrates all these enzymatic activities required for *de novo* fatty acid biosynthesis into unique protein assemblies catalysing more than 40 reaction steps. These Type I FASs are prototypic paradigms for the general trend in eukaryotes towards the formation of larger multidomain proteins, which minimize unspecific interactions and permit advanced regulation of localization, activity and degradation [44].

Two strikingly distinct Type I FAS have evolved in eukaryotes, the metazoan and the fungal FAS (fFAS). The metazoan FAS is a 540 kDa homodimer with two complete sets of functional domains and a versatile architecture defined by a minimal amount of scaffolding elements [17,48]. Structural

analysis of metazoan FAS and bacterial polyketide synthases (PKS) revealed a common architecture [50,72,73], which is used in PKS to synthesize a broad spectrum of secondary metabolites [44,77]. A fully methylating, iterative bacterial PKS was identified as a common evolutionary ancestor of metazoan FAS and modular PKS based on the presence of an evolutionary remnant of a methyltransferase domain in the metazoan FAS structure [45,48].

Fungal FAS forms a 2.6-megadalton assembly comprising 48 functional domains, as exemplified by the yeast $\alpha_6\beta_6$-heterododecameric FAS. In addition to the enzymatic activities for fatty acid elongation, it may also incorporate a phosphopantetheinyl transferase (PPT) domain, for cofactor attachment to the ACP. All domains are embedded into a scaffolding matrix that comprises nearly 50 % of the total mass and mediates the majority of architectural interactions determining the spatial arrangement of catalytic centres [16,58,60-62]. fFAS adopts a unique barrel-shape structure with two domes enclosing two reaction chambers, each housing three sets of functional domains, separated by a central wheel structure (Figure 1A). This architecture is shared with the more recently described, closely related **CMN**-FAS systems in ***C****orynebacteria*, ***M****ycobacteria*, and ***N****ocardia* [74,78], which have a slightly lower number of scaffolding expansions and lack an internal PPT [79,80]. The CMN- and fFAS are amongst the most complex biosynthetic protein machineries known [62]. Still, the evolutionary appearance of the hallmark scaffolding matrix for integrating functional domains, which defines the architecture of fFAS (and CMN-FAS) remains enigmatic, as no intermediate steps of assembly formation have been identified so far.

Extension to core conserved folds are notoriously difficult targets for the analysis of homology and phylogeny as well as for structure prediction, because overall sequence conservation in these regions is

extremely weak and strictly conserved motifs, e.g. representing catalytic sites, are absent. Thus, we use a hybrid approach of bioinformatic analysis guided by and combined with experimental structure determination as a gold standard for the analysis of the evolution of the fFAS scaffolding matrix [81,82]. Bioinformatically, we identify potential evolutionary ancestors of fFAS by searching for homologues of fFAS domains that carry insertions to their core folds in equivalent positions as their fFAS relatives. Crystal structures of candidate proteins reveal their structural organization and unambiguously demonstrate the fFAS-like organization of the respective insertion elements.

Materials and Methods

Sequence data retrieval, alignment and phylogenetic analyses

The amino acid sequences of all proteins were retrieved from GenBank (http://www.ncbi.nlm.nih.gov/sutils/genom_table.cgi). A BlastP search was performed using the protein sequence of FabK, fFAS ER, FabF and fFAS KS as the query sequence against completed bacterial and fungal genomes. A total of 47 ER domains and 80 KS domains derived from the complete genome survey were subjected to a phylogenetic analysis (Table S1). Sequences from each enzyme family were selected to have 40-80 % sequence identity to each other. Alignments were created using ClustalW and adjusted manually based on structural alignments using Geneious version 6.0. Rooted phylogenetic trees were generated by the Neighbor-Joining method using a Jukes-Cantor distance model. Bootstrapping was done using 100,000 random seeds, which were replicated 10,000 times.

Cloning, expression and purification of DfnA-ER

The enoyl reductase domain of DfnA (A7Z6E3, res. 300-752) from *Bacillus amyloliquefaciens* FZB42 (DSM 23117) [83,84] was cloned from genomic DNA and cloned into the expression plasmid pNIC28-Bsa4 [85]. Here the protein is linked to a N-terminally hexa-His-tag followed by a TEV-protease cleavage site. The protein was overproduced in *E. coli* BL21(DE3) pRIL pL1SL2 [86]. Cells were lysed by sonication in 50 mM Hepes/NaOH, 500 mM NaCl, pH 7.4, 20 mM imidazol and the supernatant was cleared by centrifugation. DfnA-ER was purified by immobilized metal-affinity chromatography on a Ni-NTA column (elution with 250 mM imidazole), with His-Tagged TEV-protease digested [87], followed by a Ni-NTA column step and then 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).

Cloning, expression and purification of FabY

FabY (PA5174) was PCR-amplified from *Pseudomonas aeruginosa* PAO1 genomic DNA and cloned into the expression plasmid pNIC28-Bsa4 [85]. The protein was overproduced in *E. coli* BL21(DE3) pRIL pL1SL2 [86]. Cells were lysed by sonication in 50 mM Hepes/NaOH, 500 mM NaCl, pH 7.4, 40 mM imidazol and the supernatant cleared by centrifugation, FabY was purified by metal-affinity chromatography on a Ni-NTA column (elution with 250 mM imidazole), and then 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

FabY crystals grew in sitting drop setups at 4 °C at a protein concentration of 8-12 mg/ml using 0.2 M Li_2SO_4, 0.1 M Bis Tris pH 5.5, and 15 % polyethylene glycol 3350. Crystals were flash frozen in liquid nitrogen after addition ethylene glycol to 25 % (v/v). DfnA-ER crystals were obtained at room temperature in 15 % *PEG3350,* 0.1 M sodium malonate, 0.1M Bis Tris at pH 6. Crystals were flash frozen after gradually increasing the ethylene glycol concentration to 20 % (v/v) over 2h. Data were collected at beamlines PXI and PXIII of the Swiss Light Source (Paul Scherrer Institut, Villigen, Switzerland) and processed using XDS [88,89]. FabY crystals belong to space group $C222_1$ with unit cell parameters of a= 99.4 Å, b = 123.3 Å and c= 100.6 Å and two molecules per asymmetric unit. Structure determination was performed by molecular replacement with the FabF

crystal structure (PDB ID: 1KAS) [90]. The final FabY model includes all residues. DfnA-ER crystallized in space group $P2_12_12_1$ with cell dimensions *a* = 80.8 Å, *b* = 94.0 Å, and *c* = 144.5 Å and two molecules per asymmetric unit. The final model comprises residues 304 to 751. Residues 300-303, 496-509 and 752 could not be build in the electron density map. Structure determination was performed by molecular replacement with FabK ER as a search model (PDB ID: 2Z6I) [28]. Model building and structure refinement were performed for both structures with Coot [91], PHENIX [92] and Buster-TNT [93] (Table 1).

Data deposition

The atomic coordinates for FabY and DfnA-ER have been deposited in the RCSB Protein Data Bank under the accession code 4cw4 and 4cw5.

Supplemental Information

Supplemental information includes one table and 10 figures

Acknowledgments

We thank the staff at the Swiss Light Source (Villigen, Switzerland) for outstanding support for crystallographic data collection, Prof. Peter Leadlay for providing pL1SL2 and Tina Jaeger for the genomic DNA of *Pseudomonas aeruginosa* PAO1. This work was supported by the Swiss National Science Foundation R'Equip and Project Funding grants 3106030_145023 and 31003A_138262, respectively.

Results

We hypothesized that scaffolding elements defining the multienzyme structure of fFAS might already occur in bacterial proteins not involved in the formation of assemblies with fFAS-like complexity. To test this hypothesis we focussed on two distinct functional domains of fFAS, the KS and ER. The KS domain is the defining unit of fatty acid synthases responsible for the decarboxylative condensation reaction. Together with the KR, it forms the fFAS α-chain central wheel (Figure 1*A*). The ER domain of fFAS is located in the central region of the β-chain, which forms the capping domes (Figure 1*A*). Direct interactions between insertion elements of KS and ER connect the dome and central wheel regions and determine the overall organization of fFAS.

Identification of extended TIM-barrel ERs in *trans*-AT PKS and PUFA synthases

ER domains display an unusual diversity among FAS systems [57,94]: The canonical bacterial ER and the metazoan FAS ER are NADPH-dependent Rossmann-fold enzymes, but the ER of fFAS is a ~550 aa domain comprising a TIM-barrel with a permanently bound flavin mononucleotide (FMN) and a large α-helical insertion. Its only distant structural neighbor is the non-canonical bacterial ER FabK [28], a 320 aa dimeric protein, which contains a conserved TIM-barrel but lacks all extension to the barrel observed in fFAS ER (Figure 1*B*).

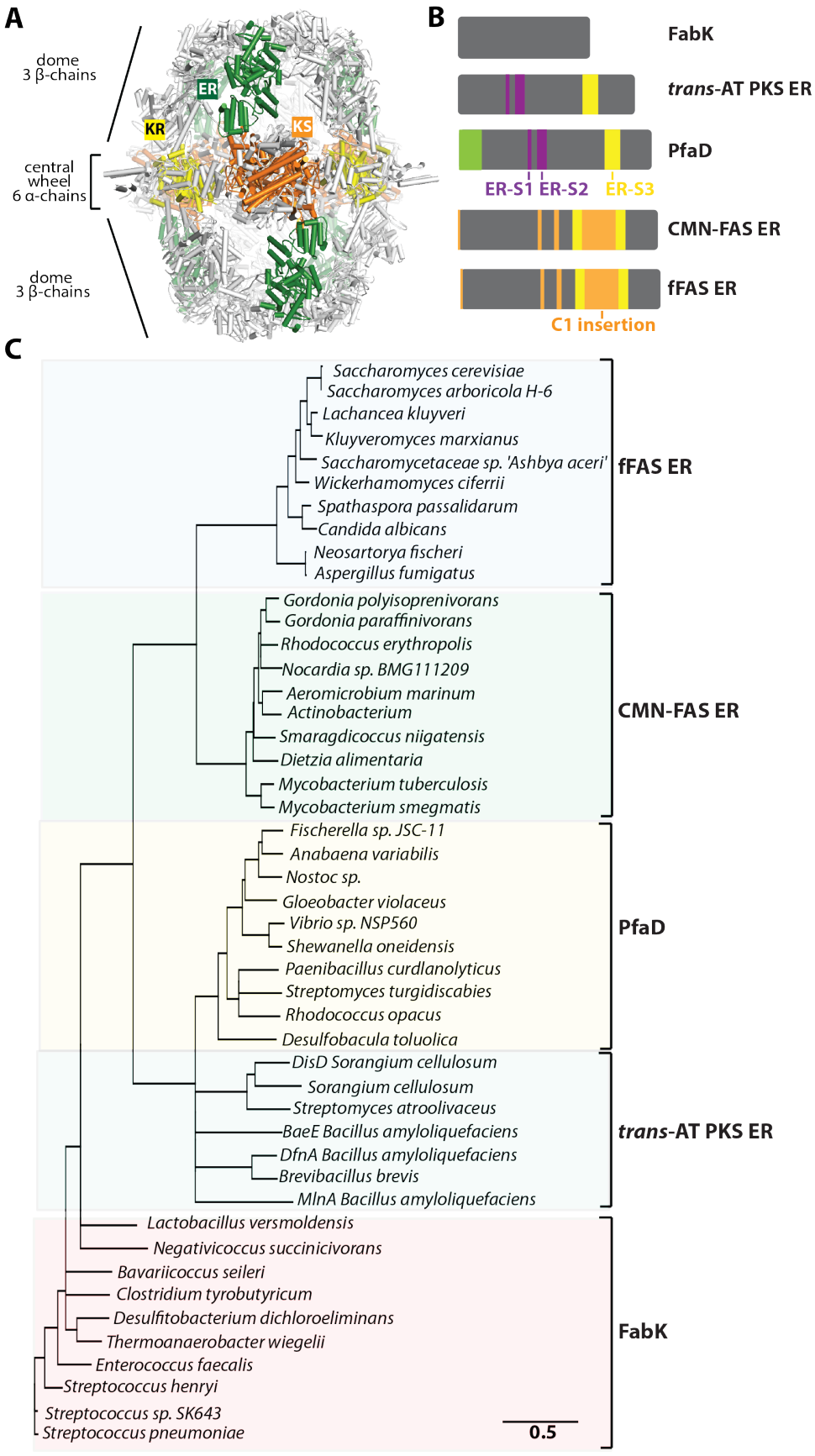


Fig. 1. Comparison of bacterial and fungal TIM-barrel-fold ERs. (**A**) Crystal structure of yeast FAS (PDB ID: 2UV8). The KS, ER and KR domains are colored in orange, green and yellow respectively. (**B**) Sequence organization of the ER domains of FabK, *trans*-AT PKS, PUFA, CMN-FAS and fFAS, at approximate sequence scale. CMN-FAS, fFAS, *trans*-AT PKS and PUFA specific sequences are colored in orange and purple, respectively. A PUFA-ER specific N-terminal extension is shown in green. Insertions present in *trans*-AT PKS, CMN-FAS and fFAS are shown in yellow. The same color code is used throughout. (**C**) Phylogenetic tree and distribution of TIM barrel ER proteins among bacteria and fungi. The tree is drawn to scale, with branch lengths in the units of number of amino acid substitutions per site. The five main groups FabK, *trans*-AT PKS, PUFA, CMN-FAS and fFAS are indicated.

We have used the minimal TIM-barrel ER FabK for the identification of TIM-barrel enoyl reductases in sequence similarity searches and reconstructed a phylogenetic tree for this family using a neighbor-joining algorithm (Figure 1*B* and 1C). This analysis identifies two families of ER domains that are more closely related to fFAS ER than FabK: ER from *trans*-AT PKS and the PfaD family of ERs in marine polyunsaturated fatty acid synthases (PUFA). PfaD homologues are standalone enzymes [95-97] and share a distinct ~80 aa N-terminal extension, unique to PUFA ERs (Figure 1*B*; green). *Trans*-AT PKS ER domains may occur as *trans*-acting isolated proteins (e.g. PedB, EtnA), but in most PKS systems they are attached to one or two AT domains to form a *trans*-acting AT-ER protein [98]. Members of both protein families have not been characterized structurally, but with 450 to 480 residues they are about 40 % larger than FabK (Figure 1*B* and Figure S1); the average sequence identities between *trans*-AT PKS ER and PUFA ER domains with fFAS ER are ~15 %. Sequence alignments identify three insertion sites in *trans*-AT PKS/PUFA ER as compared to FabK: Two adjacent insertions (ER-S1/S2) are specific for *trans*-AT PKS/PUFA ER domains (Figure 1*B*; magenta), whereas a large insertion (Figure 1*B*; orange) (ER-S3) overlaps with a major insertion element in fFAS ER (Figure S1).

The *trans*-AT PKS DfnA is a dimeric FMN-dependent enoyl reductase

To reveal the structural organization of PUFA- and *trans*-AT PKS ER, we crystallized the representative ER domain of DfnA (aa. 300-752), a *trans*-acting AT-ER protein involved in difficidin biosynthesis (Figure S2) [83,84]. The crystal structure was solved by molecular replacement using FabK (PDB ID: 2Z6I) [28] as search model and refined to R_{work}/R_{free} of 20.7/23.3 % at a resolution of 2.3 Å (Table 1)

Table 1. Statistics on diffraction data and refinement of FabY and DfnA-ER

	FabY	DfnA-ER
Wavelength (Å)	0.99997	1.0003
Resolution range (Å)	49.7 – 1.35 (1.40 - 1.35)*	47 - 2.30 (2.38 - 2.30)*
Space group	C 2 2 21	P 21 21 21
Unit cell	99.4 123.3 100.6	80.9 94.0 144.5
α, β, γ (°)	90 90 90	90 90 90
Total reflections	875960 (78609)	350174 (33349)
Unique reflections	134985 (13224)	52945 (5162)
Multiplicity	6.5 (5.9)	6.6 (6.5)
Completeness (%)	99.85 (98.80)	99.93 (99.71)
Mean I/sigma(I)	17.30 (1.59)	17.99 (1.38)
Wilson B-factor	12.97	46.72
R-merge	0.060 (1.005)	0.071 (1.256)
R-meas	0.065	0.077
CC1/2	0.999 (0.60)	0.999 (0.54)
CC*	1 (0.87)	1 (0.84)
R-work	0.148 (0.281)	0.207 (0.311)
R-free	0.177 (0.301)	0.233 (0.333)
Number of atoms	5934	7122
macromolecules	5067	6802
ligands	6	62
water	861	258
Protein residues	638	875
RMS(bonds)	0.009	0.007
RMS(angles)	1.28	1.00
Ramachandran favored (%)	97.0	98.0
Ramachandran outliers (%)	0.15	0.23
Clashscore	4.25	1.74

- Values in parentheses are for highest resolution shell.
- **Table 1. Statistics on diffraction data and refinement of FabY and DfnA-ER**

with two virtually identical molecules per asymmetric unit. The DfnA-ER monomer consists of a $(\beta/\alpha)_8$ TIM-barrel domain (aa. 305-603 and 701-752) with a bound FMN cofactor and an inserted α-helical substrate-binding domain (aa. 604-700) (Figure 2A). The $(\beta/\alpha)_8$ TIM-barrel domain closely resembles FabK with an r.m.s.d. of 1.7 Å over 321 matching Cα atoms. DfnA-ER dimerizes via a large 1920 $Å^2$ interface formed by its TIM-barrel domain and an extension of a C-terminal α-helix (dimerization tip) as observed for FabK. The structural analysis reveals that the ER-S1/S2 insertions expand the TIM-barrel domain relative to FabK by forming a small subdomain opposite to the dimerization interface and located 30 Å away from the active site (Figure 2A, Figure S3): Insertion ER-S1 forms a β-hairpin (aa. 432-442) which pack against the two helices (aa. 463-488) of insertion ER-S2.

The major insertion in DfnA is structurally conserved in fungal FAS

DfnA-ER exhibits very low sequence conservation (12.7 % identity) to the ER domain (Figure S1) of the fFAS β-chain (aa. 583-1109 in yeast FAS). Nevertheless, the DfnA-ER crystal structure reveals a close relationsship to the fFAS counterpart with an r.m.s.d. of 1.97 Å over 360 matching Cα atoms (Figure 2B) [99]. Short insertions in the TIM barrel domain relative to the FabK core fold are either specific to DfnA (insertions ER-S1, ER-S2, see above) or fFAS (aa. 784-794, 827-840), where they are involved in contacts to the neighboring AT domain (Figure 3).

The larger ER-S3 (aa. 628-674) insertion of DfnA-ER uses exactly the same insertion site as the major extension of CMN- and fFAS-ER, the C1 insertion [61] comprising residues 879-1024 in yeast FAS (Figure 2B). The expansion regions aa. 628-650 and 651-674 in DfnA match in a structural superimposition aa. 879-904 and 997-1024 in yeast FAS (Figure S4).

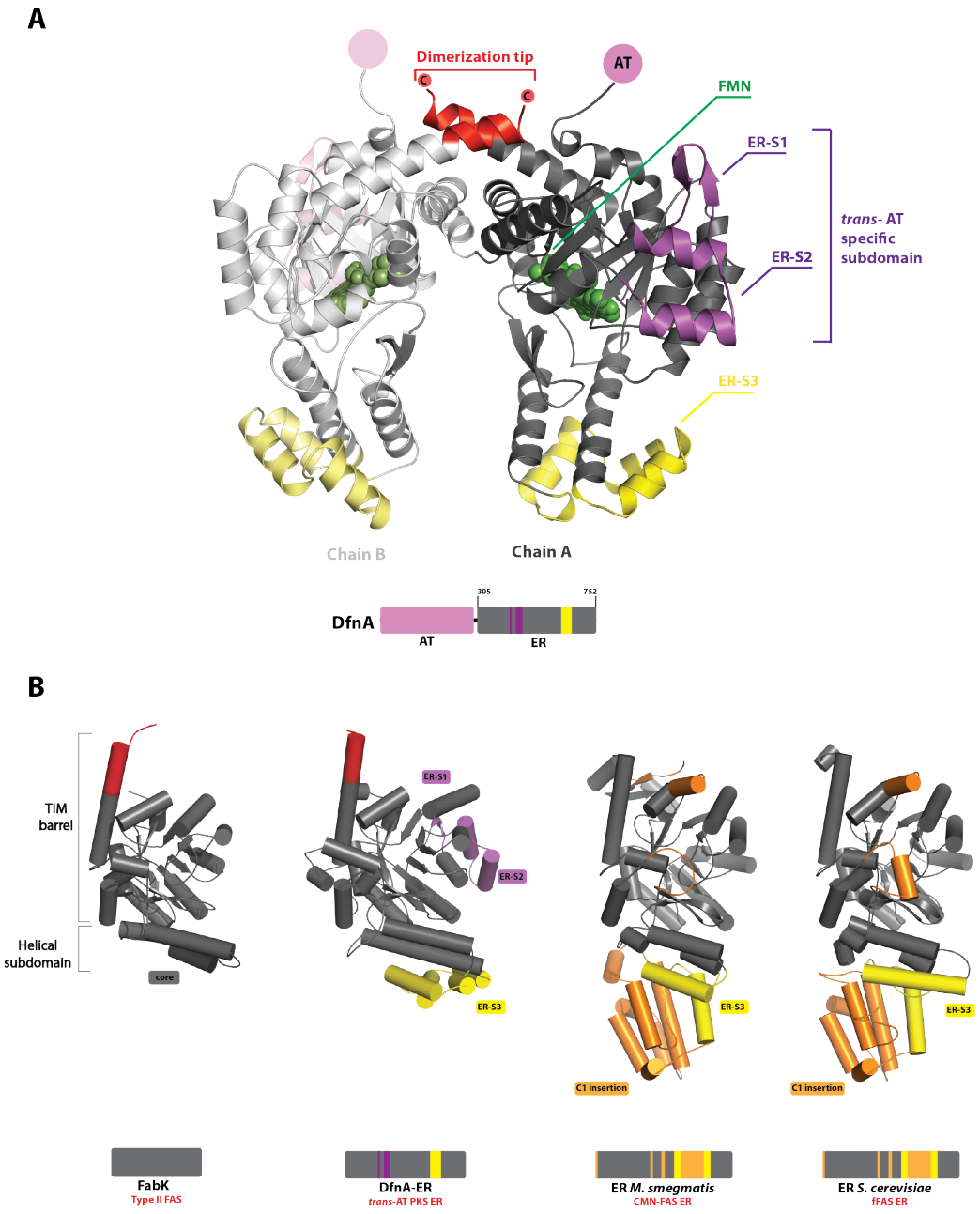


Fig. 2. Structure of DfnA-ER and comparison to bacterial and fungal homologues. (**A**) Cartoon representation of DfnA-ER. The FMN cofactor is shown in green, the dimerization tips of the C-terminal helix are shown in red, extensions ER-S1/S2 (magenta) and ER-S3 (yellow) are indicated. Anchor points for the N-terminally attached AT domain in full-length DfnA are indicated in pink. (**B**) Extension of the ER core fold of FabK in *trans*-AT PKS, CMN-FAS and fFAS. Structures of FabK (PDB ID: 2Z6I), DfnA-ER and fFAS ER from *S. cerevesiae* (PDB ID: 2UVA) and CMN ER of *S. smegmatis* (PDB ID: 3ZEN) are shown in cartoon representation.

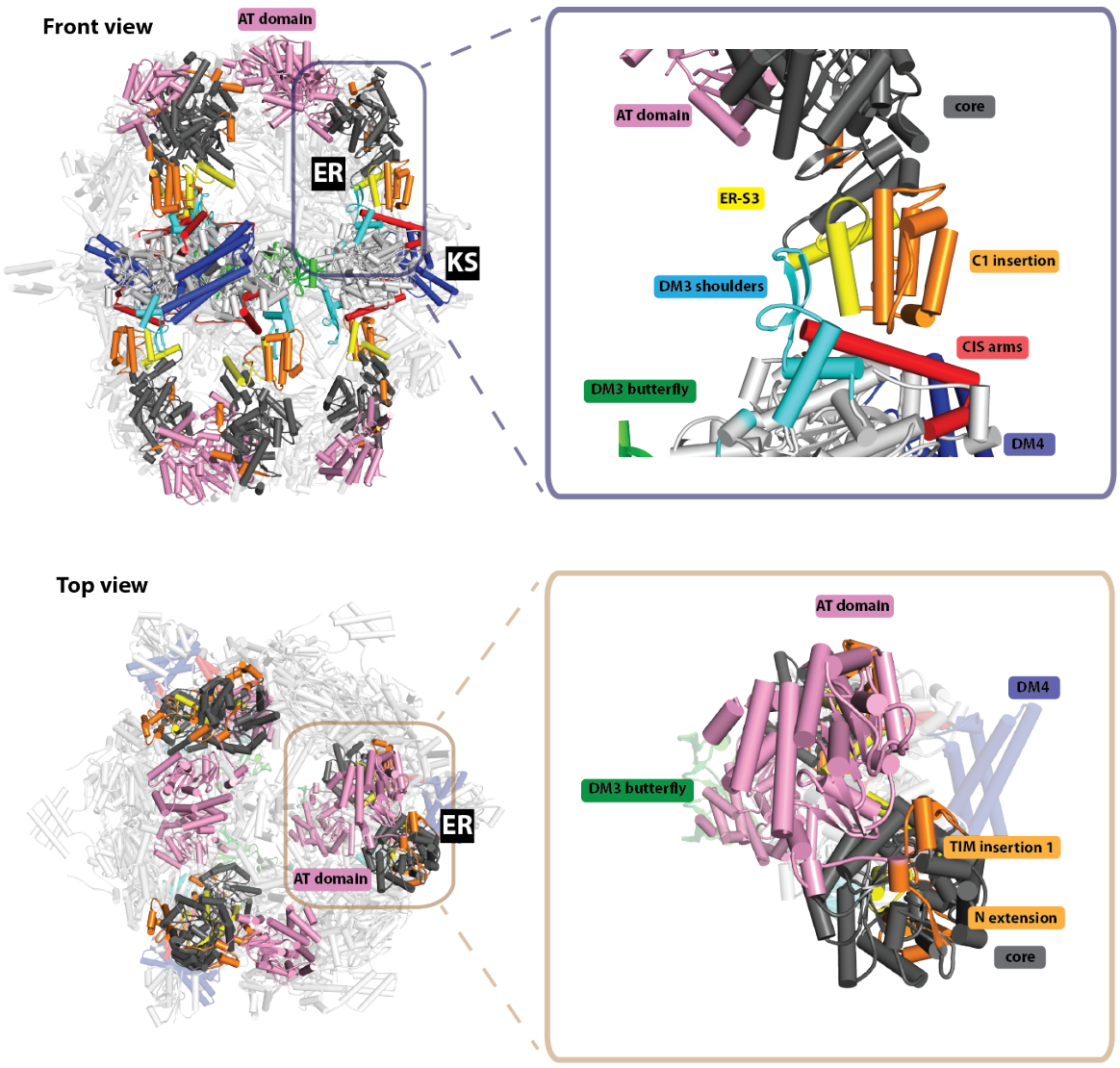


Fig. 3. Location and interactions of the ER domain in fFAS. Cartoon representation of *S. cerevesiae* FAS (PDB ID: 2UV8) in front (**top**) and top view (**bottom**). The core of the ER domain is colored in dark grey, ER expansion segments present in *trans*-AT PKS, CMN-FAS and fFAS are shown in yellow and FAS specific helical insertions are colored orange. The AT domain is colored in pale pink. The extension of the KS core fold are highlighted DM3/butterfly (green), DM3/shoulders (cyan), CIS/arms (red) and DM4 (dark blue). Close-up views of the inter-subunit interactions mediated by ER (β-chain) and KS (α-chain) in yeast FAS are shown on the right side.

651-674 in DfnA match in a structural superimposition aa. 879-904 and 997-1024 in yeast FAS (Figure S4). Inbetween these conserved segments, a further five-helix bundle (aa. 905-996) is inserted in fFAS. As evidenced by their absence from FabK, the ER-S3 insertion is not required for a general stabilization of the core fold or a productive active site conformation. In DfnA-ER it also has apparently no relevance at the level of the isolated ER-domain: It is neither involved in dimerization nor in DfnA-specific adaptations of the active site. Its conservation specifically in *trans*-acting $(AT)_xER$ proteins and PUFA rather suggests an involvement in the formation of interdomain or transient intermolecular interactions in PKS assembly lines, an analogy to the role of intersubunit connection C1 for bridging α- and β-subunits in fFAS (Figure 3).

FabY is the closest monofunctional relative of the CMN- and fFAS ketosynthase

The fungal FAS KS domain with a length of ~720 residues is much larger than its monofunctional bacterial counterparts FabH (~ 320 aa) [52] and FabB/F (~400 - 420 aa) [29]. This is mainly due to three large insertions (Figure 4): the dimerization modules 3 (DM3; green and cyan) and 4 (DM4; dark blue) and a C-terminal insertion (CIS; red) [61].

DM3 forms a core part of the central wheel and is involved in ACP binding of KS [60,61]. DM4 is located at the periphery of the KS dimer and provides the attachment point for the PPT domain, wheras CIS is involved in interactions with the ER domain in fFAS (Figure 3).

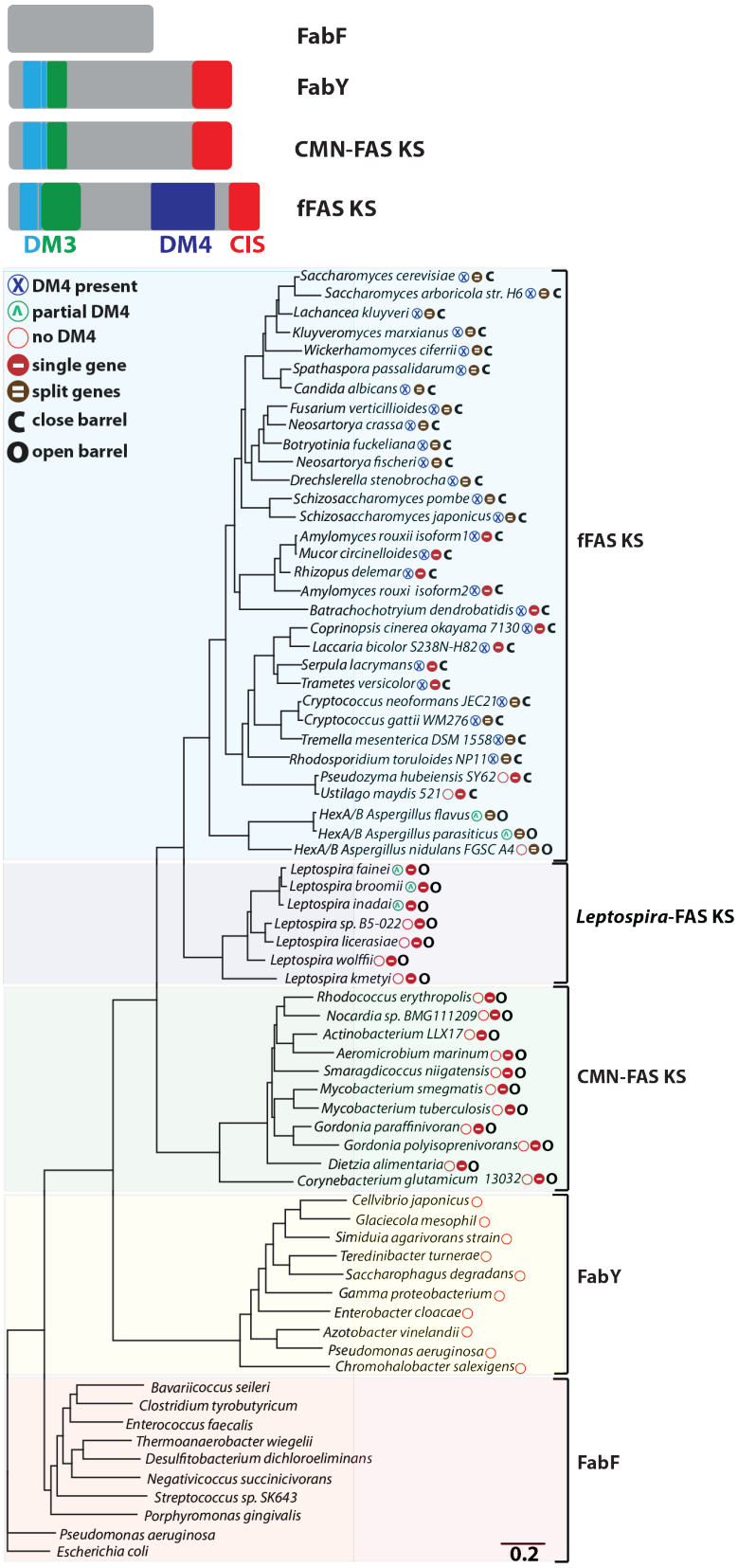


Fig. 4. Comparison of bacterial and fungal KS proteins and domains. (top) Linear sequence organization of the KS proteins FabF and FabY and the KS domains of CMN-FAS and fFAS, at approximate sequence scale. The DM3 domain is colored cyan and green, the DM4 and CIS insertions are shown in blue and red, respectively. (bottom) Phylogenetic tree and distribution of KS I/II domain-containing proteins among bacteria and fungi. The tree is drawn to scale, with branch lengths in the units of number of amino acid substitutions per site.

In a bioinformatic search we identified proteins of the FabY family as the most extended homologues of the complete fFAS KS domain (Figure 4). FabY acts as the starter enzyme for fatty acid biosynthesis in *P. aeruginosa* by catalyzing the condensation of acetyl moieties from acetyl-CoA to malonyl-ACP. Its deletion affects growth, siderophore secretion, quorum-sensing signaling and lipopolysaccharide synthesis [100,101]. Members of this recently described family of ketosynthases [102] are ~ 630 residues in size, about 200 aa larger than other monofunctional bacterial KSs from the FabB/F family (Figure 4). Our phylogenetic analysis shows that FabY shares a common evolutionary ancestor with the KS domain of CMN- and fFAS (Figure 4). FabY has only 19 % overall sequence identity to fFAS KS, but the fFAS insertions DM3 and CIS have correspondences in FabY (Figure S5). FabY lacks the DM4 insertion of fFAS, which is involved in PPT attachment. Interestingly, DM4 is also absent in bacterial CMN-FAS, which utilize *trans*-acting PPTs instead of integrated ones, as exemplified by mycobacterial FAS [79,80]. As a result, FabY is strikingly similar over its full length to the KS domain of mycobacterial type I FAS (26 % identity over 600 residues).

Large core-fold extensions in the non-canonical *P. aeruginosa* starter KS FabY

To analyze the similarity of fFAS KS and FabY at a structural level, we crystallized the dimeric 140 kDa *P. aeruginosa* FabY yielding crystals in space group $C222_1$, with one monomer per asymmetric unit and the dimeric assembly generated by crystallographic twofold-symmetry. The structure was solved by molecular replacement using FabF (PDB ID: 2GFW) and refined to R_{work}/R_{free} 14.8/17.7 % at 1.35 Å resolution. FabY is a member of the α/β-hydrolase superfamily (Figure 5A) [103,104]. Structural superimpositions of the FabY core fold only with those of the three

bacterial ketosynthase families reveals a close match (for 370 aligned residues) to both FabB (r.m.s.d. 1.6 Å) [105] and FabF (r.m.s.d. 1.8 Å) [90] (Figure S6), whereas FabH is structurally more distantly related (r.m.s.d. 2.6 Å). In line with the core fold similarity, FabY uses a Cys-His-His (Cys281, His434 and His474) triad typical for elongating KSs (FabB/F) enzymes (Figure *5B*), whereas the functional orthologs of FabY, the FabH starter condensing enzyme, is characterized by a His-Asn-Cys catalytic triad (Yuan, Sachdeva et al. 2012) (Figure *5B*). In the high resolution crystal structure, alternate conformations are observed for the active site cysteine residue (Cys281), which are also detected in the spatially adjacent loop 532-535. The active site residues of FabY have similar orientations as in the elongating KS homologues FabB/F (Figure *5B*), whereas the acyl pocket is much shorter and resembles the starter condensing enzyme FabH (Figure *5B*). This is consistent with the finding that FabY utilizes only short chain acyl-CoA as substrates [101].

While the bacterial KSs FabB/F closely resemble the catalytic core of FabY, its overall closest relative are the KS domains of CMN- and fFAS, to which the entire FabY superimposes with an r.m.s.d. of 1.9 Å and 2.0 Å over 570 matching residues, respectively [95]. In comparison to the bacterial canonical ketosynthases, FabY has three noticeable expansion segments, *Shoulder*, *Arms* and *Butterfly* (Figure 5). The *Shoulder* region (aa. 33-83) is inserted close to the N-terminus and comprises two α-helices and a small three-stranded β-sheet laterally positioned away from the two-fold symmetry axis of dimeric FabY. The *Butterfly* (aa. 98-158), follows only ten residues later and consists of a shaft

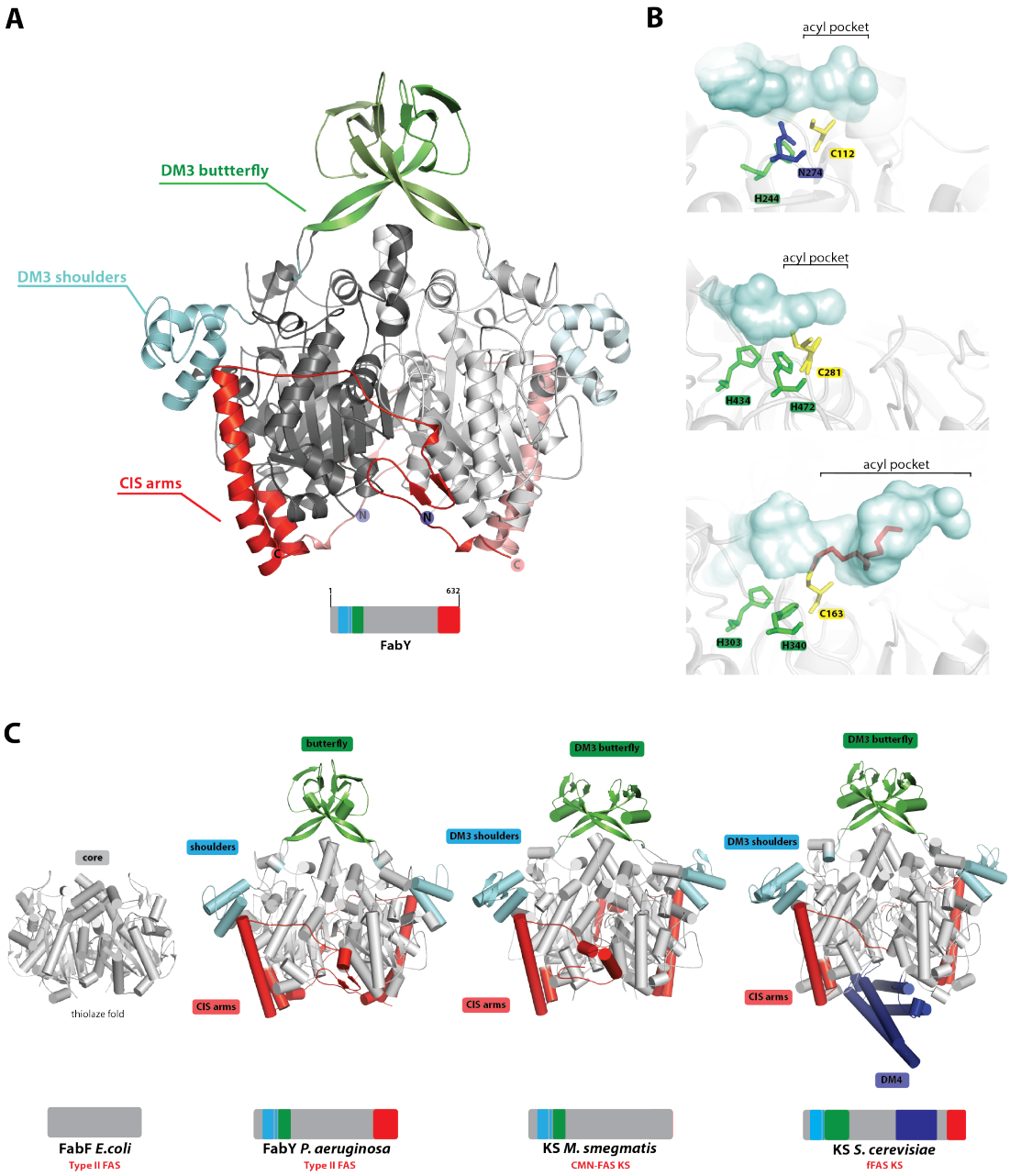


Fig. 5. Structural analysis of FabY. (**A**) Cartoon representation of dimeric (dark and light colours) FabY (**B**) Active site of FabH (top) (PDB ID: 1NHJ) [100], FabY (middle) and FabF (bottom) (PDB ID: 2GFY)[101]. The substrate entry and acyl pocket of the KS domains are oriented to the left and right, respectively. Active site residues and the dodecanoic acid (red) bound to FabF are shown in stick representation. Dual conformations of the active site Cys281 in FabY are indicated. The acyl pocket in the starter KS FabH and FabY are significant shorter than in the elongation KS FabF. (**C**) Extension of the KS core fold of FabF in FabY, CMN- and fFAS. Structures of FabF (PDB ID: 1GFW), FabY and the KS domains of *S. cerevesiae* (PDB ID: 2UV9) and *M. smegmatis* (PDB ID: 3ZEN) FAS are shown in ribbon representation. The insertion of butterfly (green), shoulders (cyan) and arm (red) in FabY are structurally conserved in CMN- and fFAS. The fFAS specific DM4 is shown in dark blue.

formed by two two-stranded anti-parallel β-sheets, capped by another three-stranded β-sheet. The *Arms* are formed by the 80 C-terminal residues of FabY (aa. 548-635). They originate from the terminal β-strand of the core fold and comprise a short and a very long helix spanning the height of the core domain. The helices are followed by an extensive linker (aa. 586-606) without regular secondary structure elements that protrude all across to the second protomer. The terminal region of the *Arms* forms a small three-stranded antiparallel β-sheet, before ending in a long loop (Figure S7).

All expansion elements are located on the periphery of FabY distant from the active site. While the *Shoulders* do not contribute to dimerization, both the *Arms* and the *Butterfly* expansion contribute considerably to the overall dimerization interface of FabY via contacts to the core fold or the synonomous expansion regions of the second protomer, respectively.

The three expansion elements strikingly resemble the structure of insertion elements observed in fungal and mycobacterial FAS multienzymes (Figure 5C). The *Arms* closely match the CIS insertion comprising residues 3012-3089 and 1659-1711, respectively, in CMN- and yeast FAS. The *Butterfly* and *Shoulder* region together resemble an insertion element designated as DM3 in fungal (aa. 1118-1179) [16] and mycobacterial FAS (aa. 2553-2615) [79]. DM3 in fFAS is a component of the central wheel structure and provides part of the binding interface for the ACP-KS interaction [60,61]. Based on sequence analysis and the conserved connecting region in between (Figure 4 and Figure S5), we suggest that the two expansions, *Butterfly* and *Shoulder*, resulted from independent insertion events and may have separate functions. The functional relevance of the insertion regions remains to be uncovered: *Butterfly* and *Arms* contribute to dimerization, however, an equivalent extended

dimerization interface is not required in other bacterial ketosynthases with a conserved dimeric structure.

Discussion

Here, we identify and characterize two families of bacterial enzymes, which carry extensions characteristic of the fFAS multienzyme architecture. DfnA is a two-domain, *trans*-acting AT-ER protein of the polyketide assembly line for difficidin A biosynthesis in *Bacillus amyloliquefaciens*. The structure of its ER domain is the first representative of a class of *trans*-acting ER domains in polyketide and PUFA biosynthesis and demonstrates their dimeric oligomerisation state.

FabY defines a family of non-canonical starter KS with a narrow and non-regular distribution in *Enterobacteriaceae* and *Pseudomonaceae*, including pathogenic *P. aeruginosa*. FabY inhibition increases hypersensitivity to antibiotics, but its function can be bypassed in infection by shunting of external C_8-fatty acids [100]. The structural analysis demonstrates that adaptions in FabY to its role as a starter enzyme are limited to the active site and substrate binding pocket and unexpectedly do not involve the ~200 aa inserted to the core fold [101].

The structures of DfnA-ER and FabY demonstrate a striking similarity in extensions to the enzymatic core domains as compared to scaffolding elements in the CMN- and fFAS multienzymes. The entire FabY with its insertion regions closely resembles the KS domains of fungal and particularly of CMN-FAS. In FAS, this domain is located in the central wheel, in a key position for nucleating complex assembly. The conserved DM3/butterfly extension of the fFAS α-chain KS domain plays a crucial role in the organization of the central hub region (Figure 6, DM3/butterfly) and contributes to the ACP binding interface for KS interaction. The CIS/arm interlinks the two protomers in the dimer (Figure S7) and orients the DM3/shoulder extension for the core contacts with the ER domain of the β chain (Figure 3, DM3/shoulders).

DfnA-ER carries an extension to its helical subdomain, which uses

identical anchor points and overlaps with the C1 insertion to the fFAS ER. The dimer interface of DfnA-ER has small overlap with the interdomain interface between the AT and ER domain in fFAS (Figure S8). For DfnA-ER and FabY, the conserved extensions do not contribute to the formation of active sites or substrate binding pockets suggesting a role in stabilization, interdomain or even intermolecular interactions. Such higher-order interactions are confirmed for PksE, a DfnA-homologue of the Bacillaene biosynthesis cluster of *B. subtilis*, which co-localizes with a bimodular PKS megaenzyme, PksR, and other PKS components to build a giant membrane associated megacomplex [106].

Our structural characterization of bacterial monofunctional proteins implementing parts of the CMN-/fFAS scaffolding elements provides first evidence of an evolutionary route to the development of the fFAS multienzyme architecture (Figure 7). Bacterial proteins of fatty acid, or polyketide and PUFA biosynthesis acquired initial extension elements, which (i) contributed to the stability of homooligomeric interactions, (ii) provided interfaces for interdomain interactions in multidomain proteins and (iii) contributed to protein-protein interactions. Gene fusions of several of these proteins formed larger gene products, possibly a KS□AT-ER complex, which already used the extensions identified in FabY and DfnA. There are currently no indications for the origin of the KR, MPT and ACP domains, which may have evolved from monofunctional FAS enzymes. Interestingly, the DH region of fFAS and DH proteins in PUFA biosynthesis are the only DH variants containing more than two succesive hotdog folds [107] (Figure 7).

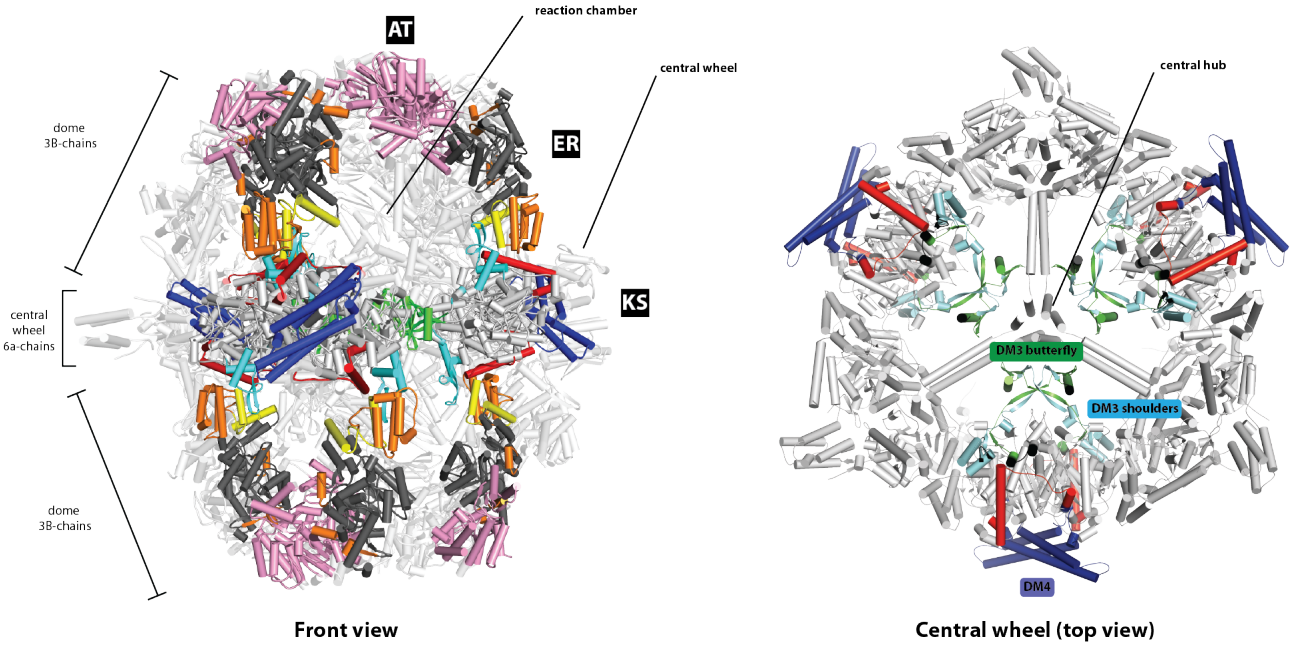


Fig. 6. Interactions of the KS domain in fFAS. Front view on *S. cerevisiae* FAS (PDB ID: 2UV8) (left) and a top view onto the isolated α-chain central wheel (right). The core of the ER domain is colored in dark grey, ER insertions present in *trans*-AT PKS as well as in CMN- and fFAS are shown in yellow and FAS specific ER insertions are colored orange. The extension of the KS core domain are indicated: DM3/butterfly (green), DM3/shoulders (cyan), CIS/arms (red) and specific dimerization module 4 (DM4) (dark blue).

Domain integration continued until a complete fatty acid synthase encoded by a single gene emerged as an evolutionary stable system in the CMN-group bacteria. Remarkably, a closely related FAS gene product, which already contains a partial DM4 extension, used in fFAS for PPT attachment, is found in the distant bacterial family of helically coiled *Leptospira* (Figure S4;S9 and supporting online text).

A fungal-type FAS has been described in genomes of fungi from all clades (Figure S10). Thus it was likely already encoded in the last fungal common ancestor (LFCA) [108]. The fungi closest to LFCA, exemplified by the euchytrid *Spizellomyces punctatus* and *Batrachochytrium dendrobatidis*, are characterized by single-gene encoded multienzyme FASs, which in contrast to bacterial multienzyme FAS is closed by the trimeric capping

insertion and comprises an internal PPT domain (Figure S9). The splitting into two genes in various positions occurs at larger evolutionary distance from LFCA around the development of *Glomeromycota, Basidiomycota* and *Ascomycota* (Figure S10). Despite the general presence of PPT, the DM4 insertion used for PPT attachment in yeast FAS is absent in *Ustilagomyceteae*-FAS, likely due to a secondary loss.

Our results clearly identify bacterial monofunctional proteins that implement scaffolding elements to found one of the most complex eukaryotic multienzymes, the fungal FAS. Careful phylogenetic and structural characterization links fungal FAS evolution to bacterial fatty acid and polyketide metabolism. The data enable targeted studies on the role of expansion elements in monofunctional and multifunctional enzyme systems and contribute to a functional dissection of the fungal FAS multienzyme architecture. This work is an important milestone towards the rational tailoring and synthetic construction of multienzyme-inspired molecular factories. Our approach of combining structural and phylogenetic data may well serve as a blueprint for the analysis of other eukaryotic multienzymes.

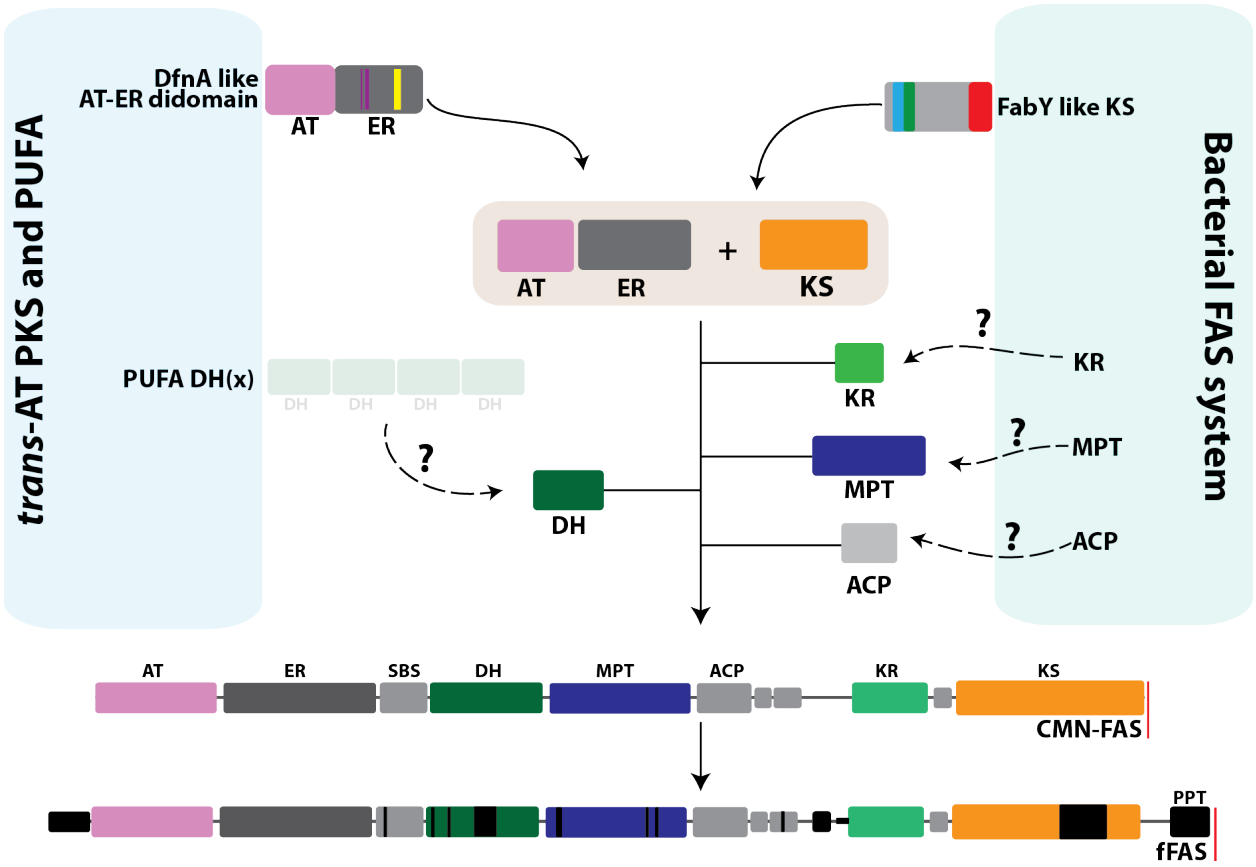


Fig. 7. Origin and development of the fFAS multienzymes architecture. AT-ER proteins from *trans*-AT PKS and monofunctional ketosynthase FabY already contain key interact segments to mediate first transient contacts. These transient interactions were further strengthened by additional expansion segments, gene fusion and additional enzymatic domains were implemented. Possibly DH domains, as found in PUFA, and KR, MPT and ACP domains from bacterial type II FAS system build a stable open bacterial CMN-FAS architecture, which was transferred to and developed (e.g. additional expansion segments shown as black boxes) in fungi to the closed fFAS architecture. The two only known DH families containing multiple hot-dog folds are fFAS (triplet) and PUFA (quadruple) and possibly share a common ancestor.

References

Adams, P.D., Grosse-Kunstleve, R.W., Hung, L.W., Ioerger, T.R., McCoy, A.J., Moriarty, N.W., Read, R.J., Sacchettini, J.C., Sauter, N.K., and Terwilliger, T.C. (2002). PHENIX: building new software for automated crystallographic structure determination. Acta Crystallogr D Biol Crystallogr *58*, 1948-1954.

Allen, E.E., and Bartlett, D.H. (2002). Structure and regulation of the omega-3 polyunsaturated fatty acid synthase genes from the deep-sea bacterium Photobacterium profundum strain SS9. Microbiology *148*, 1903-1913.

Betancor, L., Fernandez, M.J., Weissman, K.J., and Leadlay, P.F. (2008). Improved catalytic activity of a purified multienzyme from a modular polyketide synthase after coexpression with Streptomyces chaperonins in Escherichia coli. Chembiochem 9, 2962-2966.

Blanc, E., Roversi, P., Vonrhein, C., Flensburg, C., Lea, S.M., and Bricogne, G. (2004). Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT. Acta Crystallogr D Biol Crystallogr *60*, 2210-2221.

Boehringer, D., Ban, N., and Leibundgut, M. (2013). 7.5-A cryo-em structure of the mycobacterial fatty acid synthase. Journal of molecular biology *425*, 841-849.

Campbell, J.W., and Cronan, J.E., Jr. (2001). Bacterial fatty acid biosynthesis: targets for antibacterial drug discovery. Annu Rev Microbiol *55*, 305-332.

Castillo, Y.P., and Perez, M.A. (2008). Bacterial beta-ketoacyl-acyl carrier protein synthase III (FabH): an attractive target for the design of new broad-spectrum antimicrobial agents. Mini Rev Med Chem *8*, 36-45.

Chen, X.H., Koumoutsi, A., Scholz, R., Eisenreich, A., Schneider, K., Heinemeyer, I., Morgenstern, B., Voss, B., Hess, W.R., Reva, O.*, et al.* (2007). Comparative analysis of the complete genome sequence of the plant growth-promoting bacterium Bacillus amyloliquefaciens FZB42. Nat Biotechnol *25*, 1007-1014.

Chen, X.H., Vater, J., Piel, J., Franke, P., Scholz, R., Schneider, K., Koumoutsi, A., Hitzeroth, G., Grammel, N., Strittmatter, A.W.*, et al.* (2006). Structural and functional characterization of three polyketide synthase gene clusters in Bacillus amyloliquefaciens FZB 42. J Bacteriol *188*, 4024-4036.

Chung, S.Y., and Subbiah, S. (1996). A structural explanation for the twilight zone of protein sequence homology. Structure *4*, 1123-1127.

Ciccarelli, L., Connell, S.R., Enderle, M., Mills, D.J., Vonck, J., and Grininger, M. (2013). Structure and conformational variability of the mycobacterium tuberculosis fatty acid synthase multienzyme complex. Structure *21*, 1251-1257.

Dawe, J.H., Porter, C.T., Thornton, J.M., and Tabor, A.B. (2003). A template search reveals mechanistic similarities and differences in beta-ketoacyl synthases (KAS) and related enzymes. Proteins *52*, 427-435.
Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr *60*, 2126-2132.
Gago, G., Diacovich, L., Arabolaza, A., Tsai, S.C., and Gramajo, H. (2011). Fatty acid biosynthesis in actinomycetes. FEMS Microbiol Rev *35*, 475-497.
Gemperlein, K., Rachid, S., Garcia, R.O., Wenzel, S.C., and Muller, R. (2014). Polyunsaturated fatty acid biosynthesis in myxobacteria: different PUFA synthases and their product diversity. Chemical Science *5*, 1733-1741.
Grininger, M. (2014). Perspectives on the evolution, assembly and conformational dynamics of fatty acid synthase type I (FAS I) systems. Curr Opin Struct Biol *25C*, 49-56.
Haapalainen, A.M., Merilainen, G., and Wierenga, R.K. (2006). The thiolase superfamily: condensing enzymes with diverse reaction specificities. Trends Biochem Sci *31*, 64-71.
Heath, R.J., White, S.W., and Rock, C.O. (2002). Inhibitors of fatty acid synthesis as antimicrobial chemotherapeutics. Appl Microbiol Biotechnol *58*, 695-703.
Huang, W., Jia, J., Edwards, P., Dehesh, K., Schneider, G., and Lindqvist, Y. (1998). Crystal structure of beta-ketoacyl-acyl carrier protein synthase II from E.coli reveals the molecular architecture of condensing enzymes. The EMBO journal *17*, 1183-1191.
James, T.Y., Kauff, F., Schoch, C.L., Matheny, P.B., Hofstetter, V., Cox, C.J., Celio, G., Gueidan, C., Fraker, E., Miadlikowska, J., *et al.* (2006). Reconstructing the early evolution of Fungi using a six-gene phylogeny. Nature *443*, 818-822.
Jenke-Kodama, H., Sandmann, A., Muller, R., and Dittmann, E. (2005). Evolutionary implications of bacterial polyketide synthases. Mol Biol Evol *22*, 2027-2039.
Jenni, S., Leibundgut, M., Boehringer, D., Frick, C., Mikolasek, B., and Ban, N. (2007). Structure of fungal fatty acid synthase and implications for iterative substrate shuttling. Science *316*, 254-261.
Jenni, S., Leibundgut, M., Maier, T., and Ban, N. (2006). Architecture of a fungal fatty acid synthase at 5 A resolution. Science *311*, 1263-1267.
Kabsch, W. (2010a). Integration, scaling, space-group assignment and post-refinement. Acta Crystallogr D Biol Crystallogr 66, 133-144.
Kabsch, W. (2010b). Xds. Acta Crystallogr D Biol Crystallogr 66, 125-132.
Kapust, R.B., Tozser, J., Fox, J.D., Anderson, D.E., Cherry, S., Copeland, T.D., and Waugh, D.S. (2001). Tobacco etch virus protease: mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency. Protein Eng *14*, 993-1000.

Keatinge-Clay, A.T. (2012). The structures of type I polyketide synthases. Nat Prod Rep *29*, 1050-1073.
Krissinel, E., and Henrick, K. (2004). Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. Acta Crystallogr D Biol Crystallogr *60*, 2256-2268.
Leibundgut, M., Jenni, S., Frick, C., and Ban, N. (2007). Structural basis for substrate delivery by acyl carrier protein in the yeast fatty acid synthase. Science *316*, 288-290.
Leibundgut, M., Maier, T., Jenni, S., and Ban, N. (2008). The multienzyme architecture of eukaryotic fatty acid synthases. Curr Opin Struct Biol *18*, 714-725.
Lomakin, I.B., Xiong, Y., and Steitz, T.A. (2007). The crystal structure of yeast fatty acid synthase, a cellular machine with eight active sites working together. Cell *129*, 319-332.
Maier, T., Jenni, S., and Ban, N. (2006). Architecture of mammalian fatty acid synthase at 4.5 A resolution. Science *311*, 1258-1262.
Maier, T., Leibundgut, M., and Ban, N. (2008). The crystal structure of a mammalian fatty acid synthase. Science *321*, 1315-1322.
Maier, T., Leibundgut, M., Boehringer, D., and Ban, N. (2010). Structure and function of eukaryotic fatty acid synthases. Q Rev Biophys *43*, 373-422.
Massengo-Tiasse, R.P., and Cronan, J.E. (2009). Diversity in enoyl-acyl carrier protein reductases. Cell Mol Life Sci *66*, 1507-1517.
Okuyama, H., Orikasa, Y., Nishida, T., Watanabe, K., and Morita, N. (2007). Bacterial genes responsible for the biosynthesis of eicosapentaenoic and docosahexaenoic acids and their heterologous expression. Appl Environ Microbiol *73*, 665-670.
Olsen, J.G., Kadziola, A., von Wettstein-Knowles, P., Siggaard-Andersen, M., and Larsen, S. (2001). Structures of beta-ketoacyl-acyl carrier protein synthase I complexed with fatty acids elucidate its catalytic machinery. Structure *9*, 233-243.
Oyola-Robles, D., Gay, D.C., Trujillo, U., Sanchez-Pares, J.M., Bermudez, M.L., Rivera-Diaz, M., Carballeira, N.M., and Baerga-Ortiz, A. (2013). Identification of novel protein domains required for the expression of an active dehydratase fragment from a polyunsaturated fatty acid synthase. Protein science : a publication of the Protein Society *22*, 954-963.
Piel, J. (2010). Biosynthesis of polyketides by trans-AT polyketide synthases. Nat Prod Rep *27*, 996-1047.
Ponting, C.P., and Russell, R.R. (2002). The natural history of protein domains. Annu Rev Biophys Biomol Struct *31*, 45-71.
Rock, C.O., and Jackowski, S. (2002). Forty years of bacterial fatty acid synthesis. Biochem Biophys Res Commun *292*, 1155-1166.
Saito, J., Yamada, M., Watanabe, T., Iida, M., Kitagawa, H., Takahata, S., Ozawa, T., Takeuchi, Y., and Ohsawa, F. (2008). Crystal structure of enoyl-

acyl carrier protein reductase (FabK) from Streptococcus pneumoniae reveals the binding mode of an inhibitor. Protein Sci *17*, 691-699.
Savitsky, P., Bray, J., Cooper, C.D., Marsden, B.D., Mahajan, P., Burgess-Brown, N.A., and Gileadi, O. (2010). High-throughput production of human proteins for crystallization: the SGC experience. J Struct Biol *172*, 3-13.
Schweizer, E., and Hofmann, J. (2004). Microbial type I fatty acid synthases (FAS): major players in a network of cellular FAS systems. Microbiol Mol Biol Rev *68*, 501-517, table of contents.
Six, D.A., Yuan, Y., Leeds, J.A., and Meredith, T.C. (2014). Deletion of the beta-acetoacetyl synthase FabY in Pseudomonas aeruginosa induces hypoacylation of lipopolysaccharide and increases antimicrobial susceptibility. Antimicrob Agents Chemother *58*, 153-161.
Smith, J.L., and Sherman, D.H. (2008). Biochemistry. An enzyme assembly line. Science *321*, 1304-1305.
Smith, S., and Tsai, S.C. (2007). The type I fatty acid and polyketide synthases: a tale of two megasynthases. Nat Prod Rep *24*, 1041-1072.
Smith, S., Witkowski, A., and Joshi, A.K. (2003). Structural and functional organization of the animal fatty acid synthase. Progress in lipid research *42*, 289-317.
Staunton, J., and Weissman, K.J. (2001). Polyketide biosynthesis: a millennium review. Nat Prod Rep *18*, 380-416.
Straight, P.D., Fischbach, M.A., Walsh, C.T., Rudner, D.Z., and Kolter, R. (2007). A singular enzymatic megacomplex from Bacillus subtilis. Proc Natl Acad Sci U S A *104*, 305-310.
White, S.W., Zheng, J., Zhang, Y.M., and Rock (2005). The structural biology of type II fatty acid biosynthesis. Annual review of biochemistry *74*, 791-831.
Yuan, Y., Sachdeva, M., Leeds, J.A., and Meredith, T.C. (2012). Fatty acid biosynthesis in Pseudomonas aeruginosa is initiated by the FabY class of beta-ketoacyl acyl carrier protein synthases. J Bacteriol *194*, 5171-5184.
Zhang, Y.M., and Rock, C.O. (2012). Will the initiator of fatty acid synthesis in Pseudomonas aeruginosa please stand up? J Bacteriol *194*, 5159-5161.

Supplementary Information

Content

Supporting online text

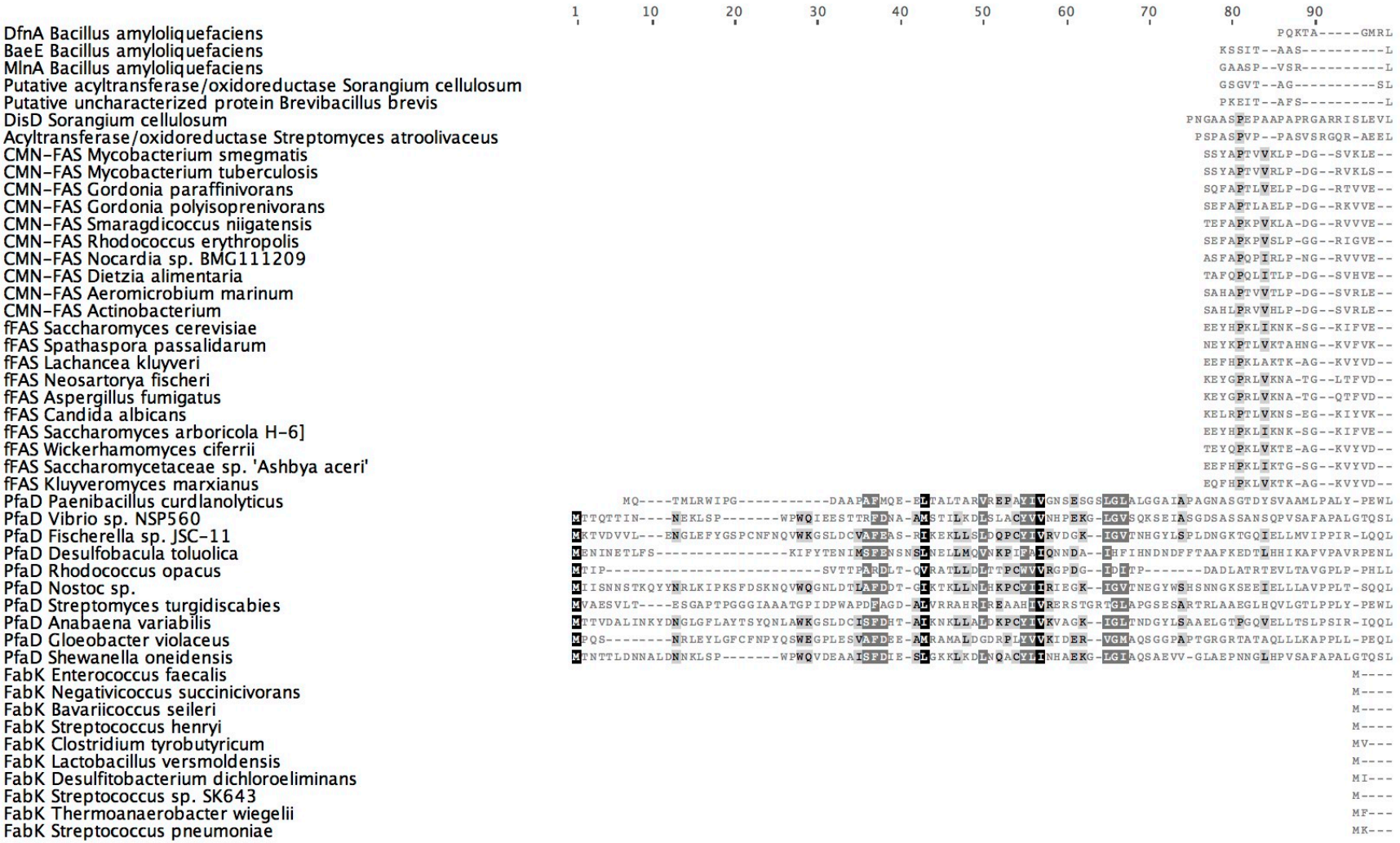
All primary metabolism fFAS differ from bacterial CMN-FAS by the presence of the barrel-closing trimeric insertion and an integral PPT. However, the secondary metabolism HexA/B hexanoic acid synthase in *Aspergillus sp.* contains an open structure resembling CMN-FAS (Figure S7A-*B*). Indeed, *Aspergillus* HexA/B forms an independent group in the KS-based multienzyme FAS phylogeny (fig. 4) and although it is a split-gene FAS it is the closest evolutionary neighbor to the bacterial *Leptospira* FAS in disagreement with the general phylogeny. The presence of an integrated PPT and the split-gene organization still argue against a rare horizontal gene transfer from bacteria [45] to *Aspergillus*. HexA/B thus likely followed a different evolutionary route dominated by its particular functional requirement for substrate transfer to interacting proteins.

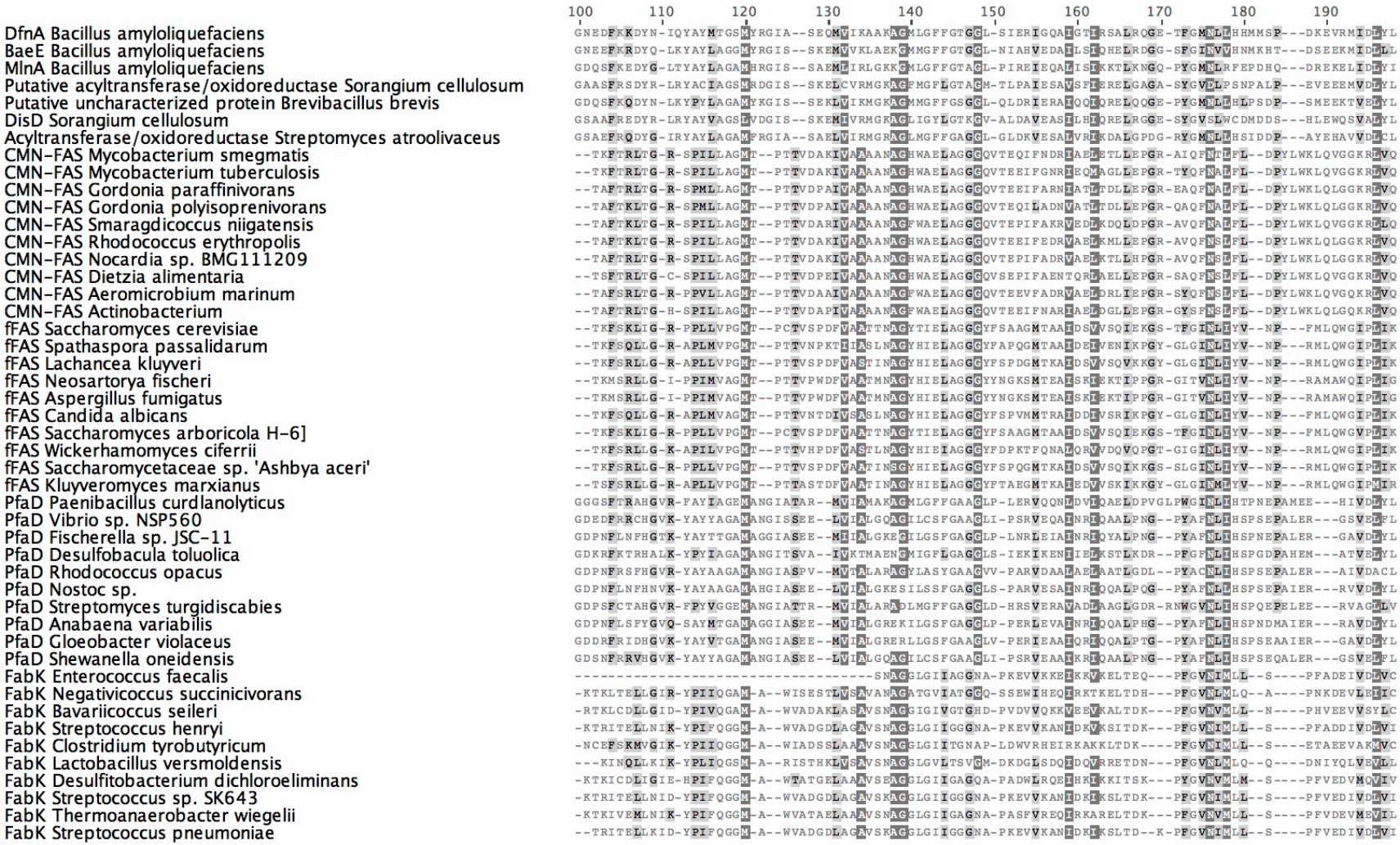
Table S1. Protein sequences used for protein alignment and phylogenetic analysis

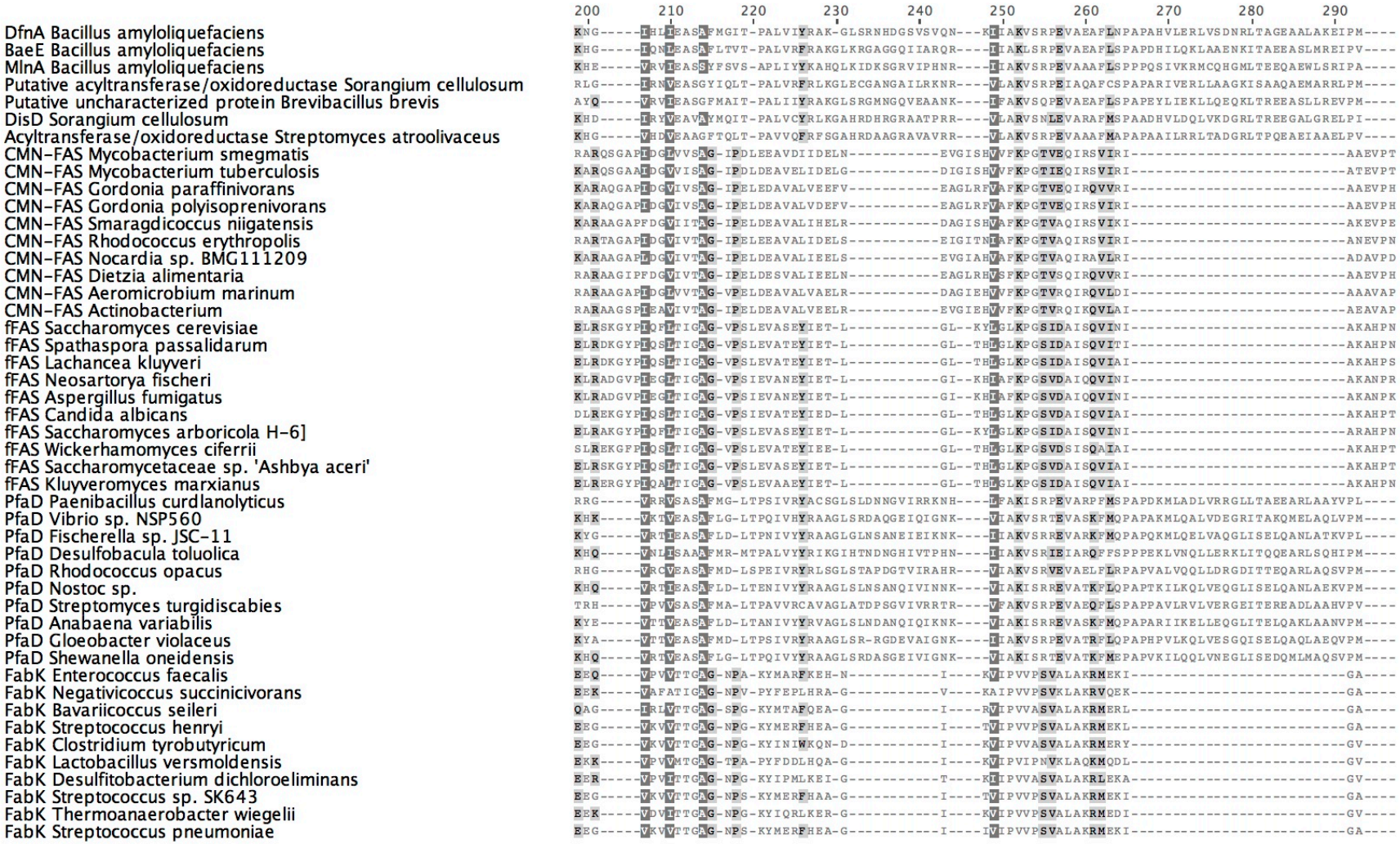
Protein name	**NCBI entry**
FabK Streptococcus pneumonia	NP_344942.1
FabK Enterococcus faecalis	ETJ09071.1
FabK Negativicoccus succinicivorans	ETI85168.1
FabK Bavariicoccus seileri	WP_022795425.1
FabK Streptococcus henryi	WP_018163773.1
FabK Clostridium tyrobutyricum	WP_017751433.1
FabK Lactobacillus versmoldensis	WP_010623791.1
FabK Desulfitobacterium dichloroeliminans	YP_007221491.1
FabK Streptococcus sp. SK643	WP_000857444.1
FabK Thermoanaerobacter wiegelii	YP_004820135.1
BaeE Bacillus amyloliquefaciens	YP_001421287.1
MlnA Bacillus amyloliquefaciens	YP_001421027.1
Putative acyltransferase/oxidoreductase Sorangium cellulosum	YP_001614781.1
Putative uncharacterized protein Brevibacillus brevis	YP_002773473.1
DisD Sorangium cellulosum	AAY32968.1
Acyltransferase/oxidoreductase Streptomyces atroolivaceus	AAN85520.1
PfaD Paenibacillus curdlanolyticus	WP_006039949.1
PfaD Anabaena variabilis	YP_323104.1
PfaD Desulfobacula toluolica	YP_006762305.1
PfaD Fischerella sp. JSC-11	WP_009756804.1
PfaD Gloeobacter violaceus	NP_925777.1
PfaD Nostoc sp.	YP_007073795.1
PfaD Rhodococcus opacus	WP_005263192.1
PfaD Shewanella oneidensis	NP_717210.1
PfaD Streptomyces turgidiscabies	WP_006378357.1
PfaD Vibrio sp. NSP560	CCA30304.1
CMN FAS Actinobacterium	WP_023645376.1
CMN FAS Aeromicrobium marinum	WP_007078357.1
CMN FAS Gordonia paraffinivorans	WP_006899551.1
CMN FAS Gordonia polyisoprenivorans	WP_020171601.1
CMN FAS Mycobacterium smegmatis	YP_889015.1
CMN FAS Mycobacterium tuberculosis	WP_003917018.1
CMN FAS Nocardia sp. BMG111209	WP_019929056.1
CMN FAS Rhodococcus erythropolis	WP_019749144.1
CMN FAS Smaragdicoccus niigatensis	WP_018162296.1
CMN FAS Dietzia alimentaria	WP_010541927.1

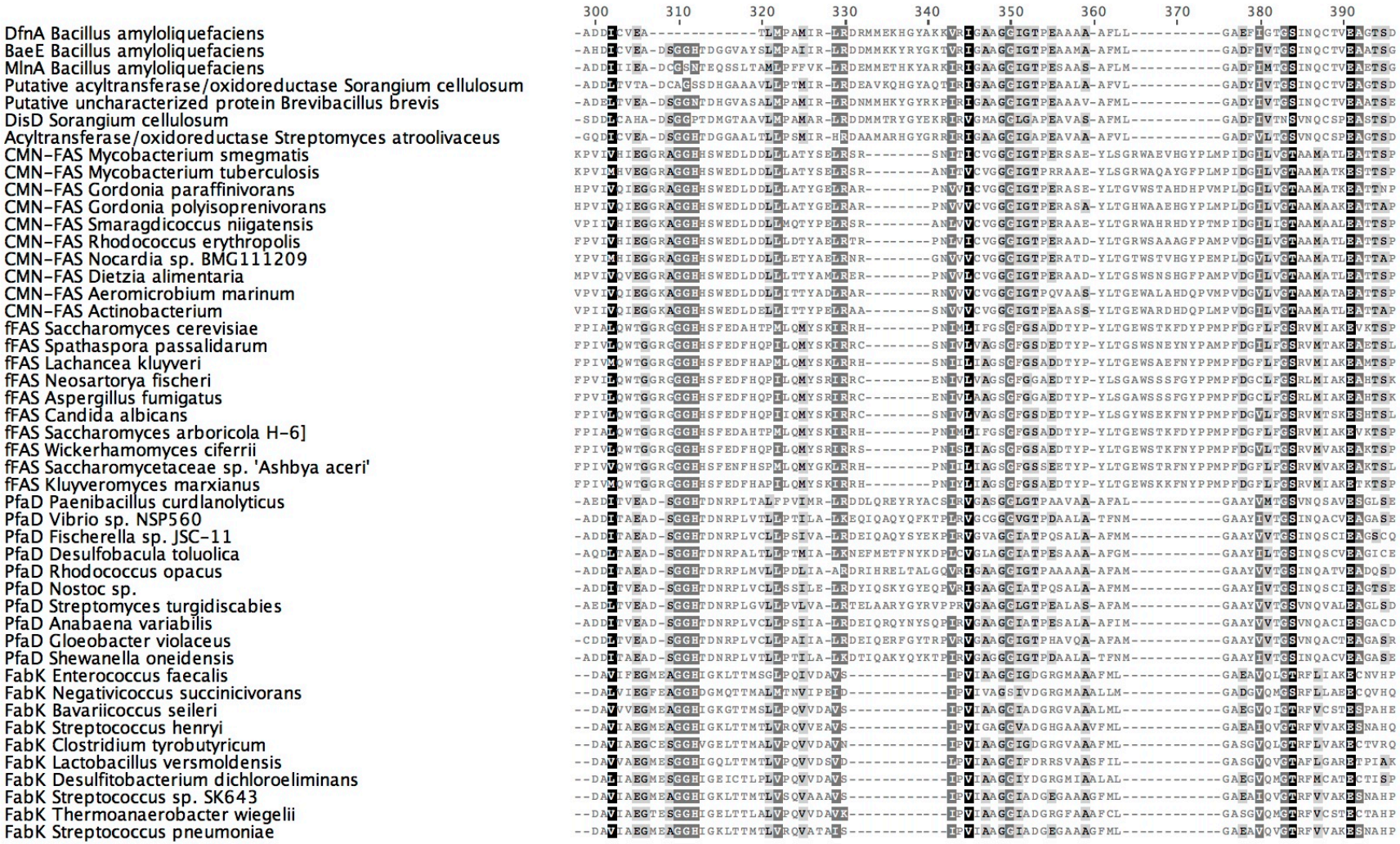
Leptospira sp. B5-022	WP_020769903
Leptospira licerasiae	WP_008590590
Leptospira wolffii	WP_016546093
Leptospira inadai	WP_020989101
Leptospira fainei	WP_016550436
Leptospira broomii	WP_010568982
fFAS Trametes versicolor	EIW57289
fFAS Mucor circinelloides	EPB92410
fFAS Drechslerella stenobrocha 248	EWC4471
fFAS Batrachochotryium dendrobatidis	XP_001547465
fFAS Botryotinia fuckeliana	XP_001547461
fFAS Fusarium verticillioides 7600 (modified)	EWG42433
fFAS Neurospora crassa OR74A (modified)	XP_962466
fFAS Leptospira kmetyi	WP_020985483
fFAS Batrachochytrium dendrobatidis	BDEG_05610.1
fFASCoprinopsis cinerea okayama7#130	XP_001836417
fFASLaccaria bicolor S238N-H82	XP_001880844
fFAS Serpula lacrymans var. lacrymans S7.3	EGN98830
fFAS Aspergillus parasiticus	AAS66003
fFAS Aspergillus nidulans FGSC A4	XP_682677
fFAS Aspergillus flavus	AAS90085.1
fFAS Rhizopus delemar RA 99-880	EIE91460
fFAS Mucor circinelloides f. circinelloides 1006PhL	EPB87701
fFAS Ustilago maydis 521	XP_759118
fFAS Pseudozyma hubeiensis SY62	GAC97557
fFAS Amylomyces rouxii iso 1	ADN94479.1
fFAS Amylomyces rouxii iso 2	ADN94478.1
fFAS Tremella mesenterica DSM 1558	EIW67374
fFAS Cryptococcus gattii WM276	XP_003194424
fFAS Cryptococcus neoformans var. neoformans JEC21	XP_571100
fFAS Trametes versicolor FP-101664 SS1	EIW57289
fFAS Serpula lacrymans var. lacrymans S7.3	EGN98830
fFAS Laccaria bicolor S238N-H82	XP_001880844
fFAS Coprinopsis cinerea okayama7#130	XP_001836417
fFAS Rhodosporidium toruloides NP11	EMS21161
fFAS Aspergillus fumigatus	EDP53206.1
fFAS Candida albicans	EEQ46070.1
fFAS Kluyveromyces marxianus	BAO40550.1
fFAS Lachancea kluyveri	BAB62141.1
fFAS Neosartorya fischeri	XP_001259180.1

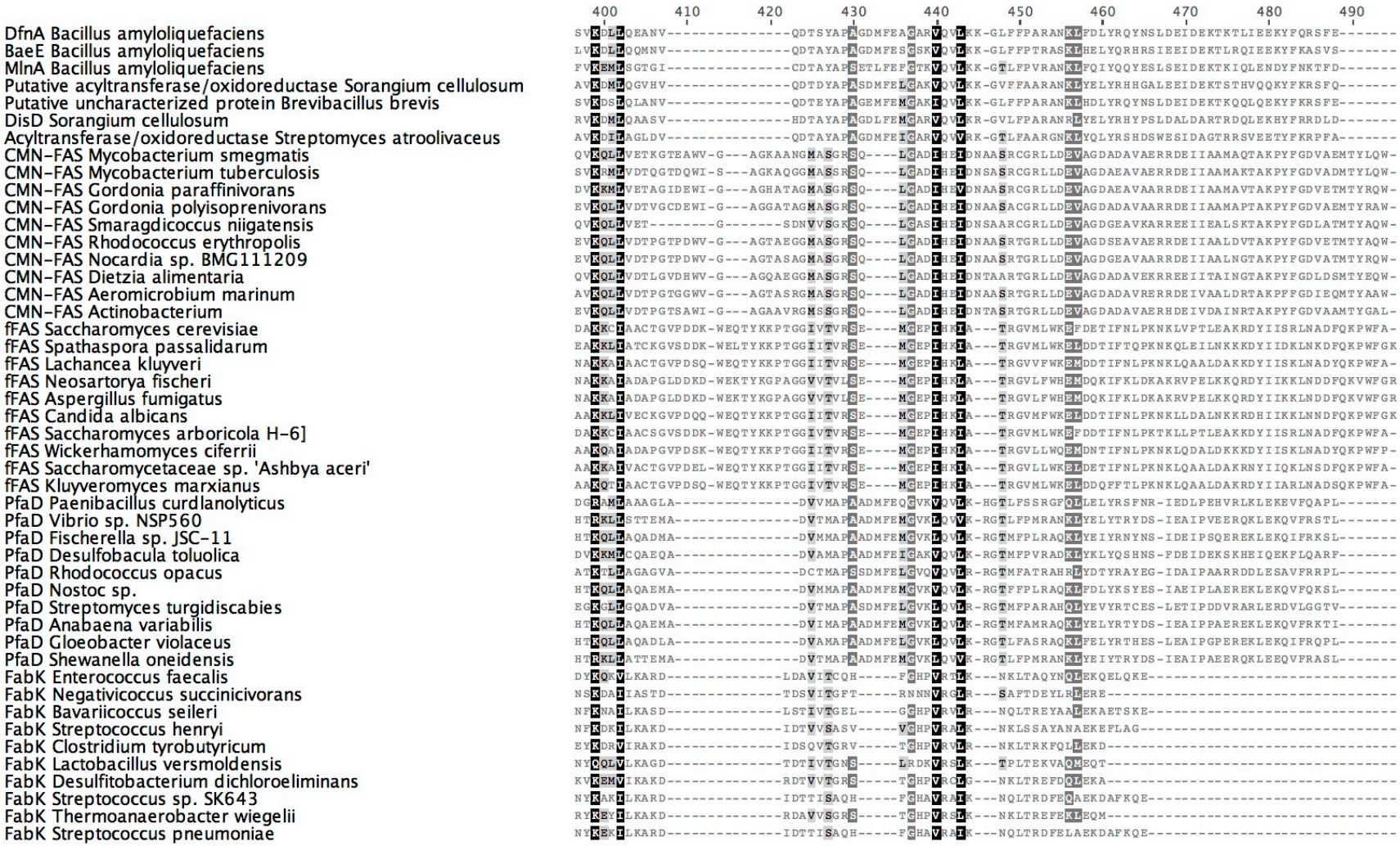
fFAS Saccharomyces arboricola H-6	EJS42996.1
fFAS Saccharomyces cerevisiae	AAA34602.1
fFAS Saccharomycetaceae sp. 'Ashbya aceri'	AGO12437.1
fFAS Spathaspora passalidarum	XP_007375193.1
fFAS Wickerhamomyces ciferrii	CCH45960.1
FabF Bavariicoccus seileri	WP_022795422.1
FabF Clostridium tyrobutyricum	WP_017751436.1
FabF Desulfitobacterium dichloroeliminans LMG P-21439	YP_007221486.1
FabF Enterococcus faecalis	NP_814075.1
FabF Negativicoccus succinicivorans	ETI84545.1
FabF Porphyromonas gingivalis	NP_905866.1
FabF Pseudomonas aeruginosa	NP_251655.1
FabF Streptococcus sp. SK643	WP_000774057.1
FabF Thermoanaerobacter wiegelii	YP_004820131.1
FabF Escherichia coli	2GFV_A
FabY Pseudomonas aeruginosa	NP_253861.1
FabY Azotobacter vinelandii CA6	YP_002797779.1
FabY Cellvibrio japonicus strain Ueda107	YP_001981921.1
FabY Chromohalobacter salexigens strain DSM 3043	YP_574978.1
FabY Enterobacter cloacae BWH 31	WP_023310156.1
FabY Gamma proteobacterium	YP_003810577.1
FabY Glaciecola mesophila KMM 241	WP_006991739.1
FabY Saccharophagus degradans strain 2-40	YP_527077.1
FabY Simiduia agarivorans strain DSM 21679	YP_006915075.1
FabY Teredinibacter turnerae strain ATCC 39867	YP_003073149.1

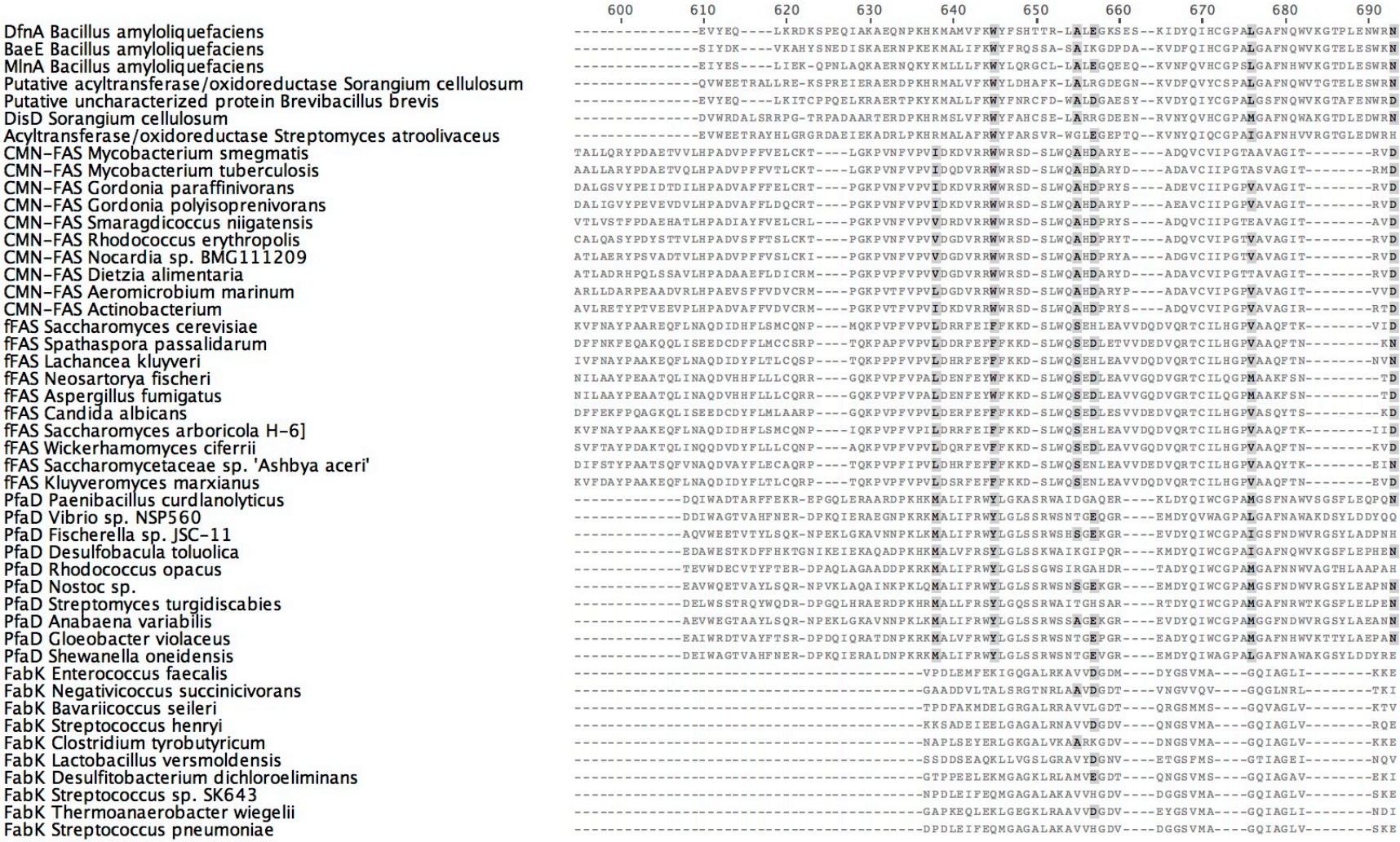

1 10 20 30 40 50 60 70 80 90
DfnA Bacillus amyloliquefaciens
BaeE Bacillus amyloliquefaciens
MlnA Bacillus amyloliquefaciens
Putative acyltransferase/oxidoreductase Sorangium cellulosum
Putative uncharacterized protein Brevibacillus brevis
DisD Sorangium cellulosum
Acyltransferase/oxidoreductase Streptomyces atroolivaceus
CMN-FAS Mycobacterium smegmatis
CMN-FAS Mycobacterium tuberculosis
CMN-FAS Gordonia paraffinivorans
CMN-FAS Gordonia polyisoprenivorans
CMN-FAS Smaragdicoccus niigatensis
CMN-FAS Rhodococcus erythropolis
CMN-FAS Nocardia sp. BMG111209
CMN-FAS Dietzia alimentaria
CMN-FAS Aeromicrobium marinum
CMN-FAS Actinobacterium
fFAS Saccharomyces cerevisiae
fFAS Spathaspora passalidarum
fFAS Lachancea kluyveri
fFAS Neosartorya fischeri
fFAS Aspergillus fumigatus
fFAS Candida albicans
fFAS Saccharomyces arboricola H-6]
fFAS Wickerhamomyces ciferrii
fFAS Saccharomycetaceae sp. 'Ashbya aceri'
fFAS Kluyveromyces marxianus
PfaD Paenibacillus curdlanolyticus
PfaD Vibrio sp. NSP560
PfaD Fischerella sp. JSC-11
PfaD Desulfobacula toluolica
PfaD Rhodococcus opacus
PfaD Nostoc sp.
PfaD Streptomyces turgidiscabies
PfaD Anabaena variabilis
PfaD Gloeobacter violaceus
PfaD Shewanella oneidensis
FabK Enterococcus faecalis
FabK Negativicoccus succinicivorans
FabK Bavariicoccus seileri
FabK Streptococcus henryi
FabK Clostridium tyrobutyricum
FabK Lactobacillus versmoldensis
FabK Desulfitobacterium dichloroeliminans
FabK Streptococcus sp. SK643
FabK Thermoanaerobacter wiegelii
FabK Streptococcus pneumoniae


100
110
120
130
140
150
160
170
180
190
DfnA Bacillus amyloliquefaciens
BaeE Bacillus amyloliquefaciens
MlnA Bacillus amyloliquefaciens
Putative acyltransferase/oxidoreductase Sorangium cellulosum
Putative uncharacterized protein Brevibacillus brevis
DisD Sorangium cellulosum
Acyltransferase/oxidoreductase Streptomyces atroolivaceus
CMN-FAS Mycobacterium smegmatis
CMN-FAS Mycobacterium tuberculosis
CMN-FAS Gordonia paraffinivorans
CMN-FAS Gordonia polyisoprenivorans
CMN-FAS Smaragdicoccus niigatensis
CMN-FAS Rhodococcus erythropolis
CMN-FAS Nocardia sp. BMG111209
CMN-FAS Dietzia alimentaria
CMN-FAS Aeromicrobium marinum
CMN-FAS Actinobacterium
fFAS Saccharomyces cerevisiae
fFAS Spathaspora passalidarum
fFAS Lachancea kluyveri
fFAS Neosartorya fischeri
fFAS Aspergillus fumigatus
fFAS Candida albicans
fFAS Saccharomyces arboricola H-6]
fFAS Wickerhamomyces ciferrii
fFAS Saccharomycetaceae sp. 'Ashbya aceri'
fFAS Kluyveromyces marxianus
PfaD Paenibacillus curdlanolyticus
PfaD Vibrio sp. NSP560
PfaD Fischerella sp. JSC-11
PfaD Desulfobacula toluolica
PfaD Rhodococcus opacus
PfaD Nostoc sp.
PfaD Streptomyces turgidiscabies
PfaD Anabaena variabilis
PfaD Gloeobacter violaceus
PfaD Shewanella oneidensis
FabK Enterococcus faecalis
FabK Negativicoccus succinicivorans
FabK Bavariicoccus seileri
FabK Streptococcus henryi
FabK Clostridium tyrobutyricum
FabK Lactobacillus versmoldensis
FabK Desulfitobacterium dichloroeliminans
FabK Streptococcus sp. SK643
FabK Thermoanaerobacter wiegelii
FabK Streptococcus pneumoniae


200
210
220
230
240
250
260
270
280
290
DfnA Bacillus amyloliquefaciens
BaeE Bacillus amyloliquefaciens
MlnA Bacillus amyloliquefaciens
Putative acyltransferase/oxidoreductase Sorangium cellulosum
Putative uncharacterized protein Brevibacillus brevis
DisD Sorangium cellulosum
Acyltransferase/oxidoreductase Streptomyces atroolivaceus
CMN-FAS Mycobacterium smegmatis
CMN-FAS Mycobacterium tuberculosis
CMN-FAS Gordonia paraffinivorans
CMN-FAS Gordonia polyisoprenivorans
CMN-FAS Smaragdicoccus niigatensis
CMN-FAS Rhodococcus erythropolis
CMN-FAS Nocardia sp. BMG111209
CMN-FAS Dietzia alimentaria
CMN-FAS Aeromicrobium marinum
CMN-FAS Actinobacterium
fFAS Saccharomyces cerevisiae
fFAS Spathaspora passalidarum
fFAS Lachancea kluyveri
fFAS Neosartorya fischeri
fFAS Aspergillus fumigatus
fFAS Candida albicans
fFAS Saccharomyces arboricola H-6]
fFAS Wickerhamomyces ciferrii
fFAS Saccharomycetaceae sp. 'Ashbya aceri'
fFAS Kluyveromyces marxianus
PfaD Paenibacillus curdlanolyticus
PfaD Vibrio sp. NSP560
PfaD Fischerella sp. JSC-11
PfaD Desulfobacula toluolica
PfaD Rhodococcus opacus
PfaD Nostoc sp.
PfaD Streptomyces turgidiscabies
PfaD Anabaena variabilis
PfaD Gloeobacter violaceus
PfaD Shewanella oneidensis
FabK Enterococcus faecalis
FabK Negativicoccus succinicivorans
FabK Bavariicoccus seileri
FabK Streptococcus henryi
FabK Clostridium tyrobutyricum
FabK Lactobacillus versmoldensis
FabK Desulfitobacterium dichloroeliminans
FabK Streptococcus sp. SK643
FabK Thermoanaerobacter wiegelii
FabK Streptococcus pneumoniae


300
310
320
330
340
350
360
370
380
390
DfnA Bacillus amyloliquefaciens
BaeE Bacillus amyloliquefaciens
MlnA Bacillus amyloliquefaciens
Putative acyltransferase/oxidoreductase Sorangium cellulosum
Putative uncharacterized protein Brevibacillus brevis
DisD Sorangium cellulosum
Acyltransferase/oxidoreductase Streptomyces atroolivaceus
CMN-FAS Mycobacterium smegmatis
CMN-FAS Mycobacterium tuberculosis
CMN-FAS Gordonia paraffinivorans
CMN-FAS Gordonia polyisoprenivorans
CMN-FAS Smaragdicoccus niigatensis
CMN-FAS Rhodococcus erythropolis
CMN-FAS Nocardia sp. BMG111209
CMN-FAS Dietzia alimentaria
CMN-FAS Aeromicrobium marinum
CMN-FAS Actinobacterium
fFAS Saccharomyces cerevisiae
fFAS Spathaspora passalidarum
fFAS Lachancea kluyveri
fFAS Neosartorya fischeri
fFAS Aspergillus fumigatus
fFAS Candida albicans
fFAS Saccharomyces arboricola H-6]
fFAS Wickerhamomyces ciferrii
fFAS Saccharomycetaceae sp. 'Ashbya aceri'
fFAS Kluyveromyces marxianus
PfaD Paenibacillus curdlanolyticus
PfaD Vibrio sp. NSP560
PfaD Fischerella sp. JSC-11
PfaD Desulfobacula toluolica
PfaD Rhodococcus opacus
PfaD Nostoc sp.
PfaD Streptomyces turgidiscabies
PfaD Anabaena variabilis
PfaD Gloeobacter violaceus
PfaD Shewanella oneidensis
FabK Enterococcus faecalis
FabK Negativicoccus succinicivorans
FabK Bavariicoccus seileri
FabK Streptococcus henryi
FabK Clostridium tyrobutyricum
FabK Lactobacillus versmoldensis
FabK Desulfitobacterium dichloroeliminans
FabK Streptococcus sp. SK643
FabK Thermoanaerobacter wiegelii
FabK Streptococcus pneumoniae


400
410
420
430
440
450
460
470
480
490
DfnA Bacillus amyloliquefaciens
BaeE Bacillus amyloliquefaciens
MlnA Bacillus amyloliquefaciens
Putative acyltransferase/oxidoreductase Sorangium cellulosum
Putative uncharacterized protein Brevibacillus brevis
DisD Sorangium cellulosum
Acyltransferase/oxidoreductase Streptomyces atroolivaceus
CMN-FAS Mycobacterium smegmatis
CMN-FAS Mycobacterium tuberculosis
CMN-FAS Gordonia paraffinivorans
CMN-FAS Gordonia polyisoprenivorans
CMN-FAS Smaragdicoccus niigatensis
CMN-FAS Rhodococcus erythropolis
CMN-FAS Nocardia sp. BMG111209
CMN-FAS Dietzia alimentaria
CMN-FAS Aeromicrobium marinum
CMN-FAS Actinobacterium
fFAS Saccharomyces cerevisiae
fFAS Spathaspora passalidarum
fFAS Lachancea kluyveri
fFAS Neosartorya fischeri
fFAS Aspergillus fumigatus
fFAS Candida albicans
fFAS Saccharomyces arboricola H-6]
fFAS Wickerhamomyces ciferrii
fFAS Saccharomycetaceae sp. 'Ashbya aceri'
fFAS Kluyveromyces marxianus
PfaD Paenibacillus curdlanolyticus
PfaD Vibrio sp. NSP560
PfaD Fischerella sp. JSC-11
PfaD Desulfobacula toluolica
PfaD Rhodococcus opacus
PfaD Nostoc sp.
PfaD Streptomyces turgidiscabies
PfaD Anabaena variabilis
PfaD Gloeobacter violaceus
PfaD Shewanella oneidensis
FabK Enterococcus faecalis
FabK Negativicoccus succinicivorans
FabK Bavariicoccus seileri
FabK Streptococcus henryi
FabK Clostridium tyrobutyricum
FabK Lactobacillus versmoldensis
FabK Desulfitobacterium dichloroeliminans
FabK Streptococcus sp. SK643
FabK Thermoanaerobacter wiegelii
FabK Streptococcus pneumoniae


600 610 620 630 640 650 660 670 680 690
DfnA Bacillus amyloliquefaciens
BaeE Bacillus amyloliquefaciens
MlnA Bacillus amyloliquefaciens
Putative acyltransferase/oxidoreductase Sorangium cellulosum
Putative uncharacterized protein Brevibacillus brevis
DisD Sorangium cellulosum
Acyltransferase/oxidoreductase Streptomyces atroolivaceus
CMN-FAS Mycobacterium smegmatis
CMN-FAS Mycobacterium tuberculosis
CMN-FAS Gordonia paraffinivorans
CMN-FAS Gordonia polyisoprenivorans
CMN-FAS Smaragdicoccus niigatensis
CMN-FAS Rhodococcus erythropolis
CMN-FAS Nocardia sp. BMG111209
CMN-FAS Dietzia alimentaria
CMN-FAS Aeromicrobium marinum
CMN-FAS Actinobacterium
fFAS Saccharomyces cerevisiae
fFAS Spathaspora passalidarum
fFAS Lachancea kluyveri
fFAS Neosartorya fischeri
fFAS Aspergillus fumigatus
fFAS Candida albicans
fFAS Saccharomyces arboricola H-6]
fFAS Wickerhamomyces ciferrii
fFAS Saccharomycetaceae sp. 'Ashbya aceri'
fFAS Kluyveromyces marxianus
PfaD Paenibacillus curdlanolyticus
PfaD Vibrio sp. NSP560
PfaD Fischerella sp. JSC-11
PfaD Desulfobacula toluolica
PfaD Rhodococcus opacus
PfaD Nostoc sp.
PfaD Streptomyces turgidiscabies
PfaD Anabaena variabilis
PfaD Gloeobacter violaceus
PfaD Shewanella oneidensis
FabK Enterococcus faecalis
FabK Negativicoccus succinicivorans
FabK Bavariicoccus seileri
FabK Streptococcus henryi
FabK Clostridium tyrobutyricum
FabK Lactobacillus versmoldensis
FabK Desulfitobacterium dichloroeliminans
FabK Streptococcus sp. SK643
FabK Thermoanaerobacter wiegelii
FabK Streptococcus pneumoniae

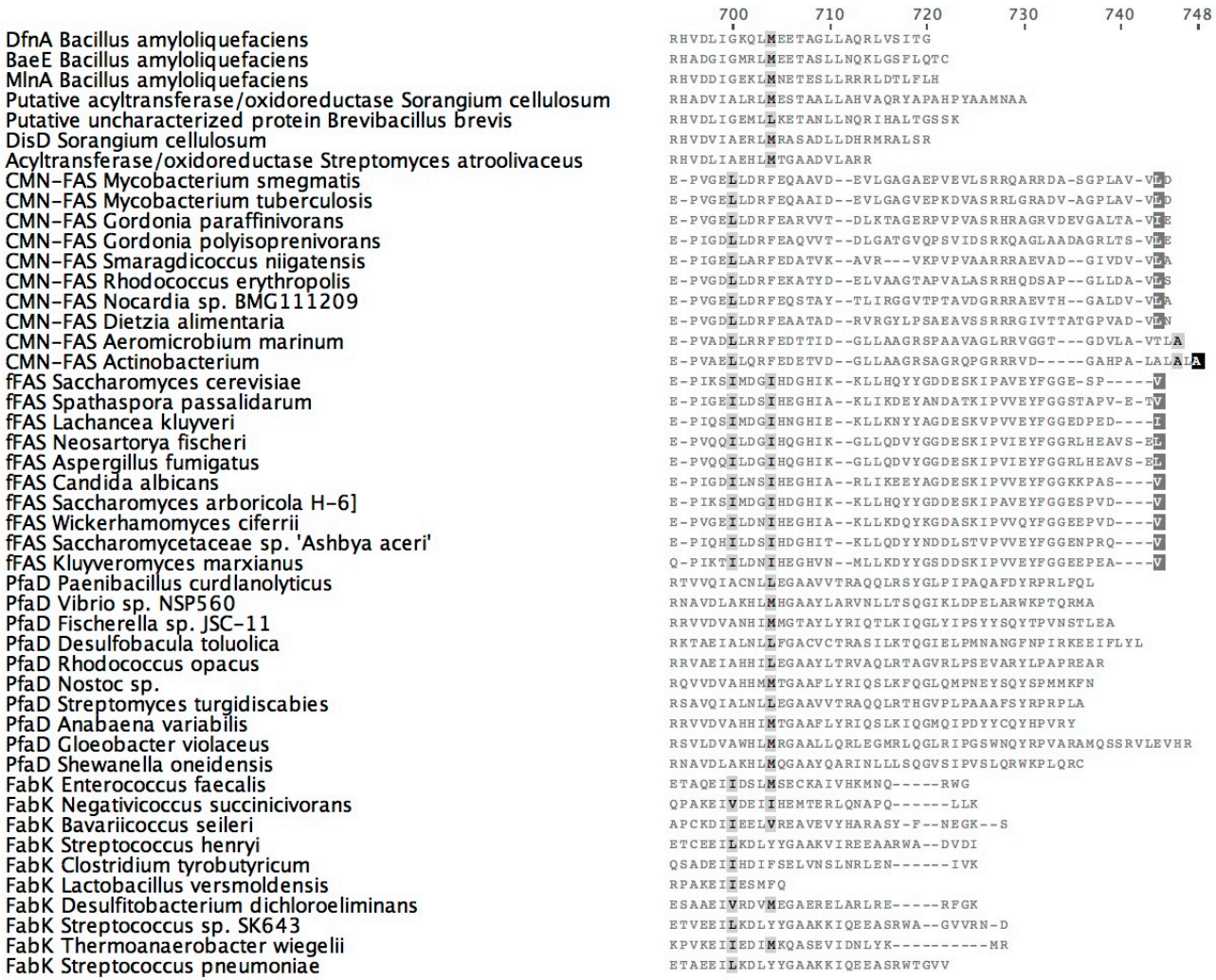

700
710
720
730
740
748
DfnA Bacillus amyloliquefaciens
BaeE Bacillus amyloliquefaciens
MlnA Bacillus amyloliquefaciens
Putative acyltransferase/oxidoreductase Sorangium cellulosum
Putative uncharacterized protein Brevibacillus brevis
DisD Sorangium cellulosum
Acyltransferase/oxidoreductase Streptomyces atroolivaceus
CMN-FAS Mycobacterium smegmatis
CMN-FAS Mycobacterium tuberculosis
CMN-FAS Gordonia paraffinivorans
CMN-FAS Gordonia polyisoprenivorans
CMN-FAS Smaragdicoccus niigatensis
CMN-FAS Rhodococcus erythropolis
CMN-FAS Nocardia sp. BMG111209
CMN-FAS Dietzia alimentaria
CMN-FAS Aeromicrobium marinum
CMN-FAS Actinobacterium
fFAS Saccharomyces cerevisiae
fFAS Spathaspora passalidarum
fFAS Lachancea kluyveri
fFAS Neosartorya fischeri
fFAS Aspergillus fumigatus
fFAS Candida albicans
fFAS Saccharomyces arboricola H-6]
fFAS Wickerhamomyces ciferrii
fFAS Saccharomycetaceae sp. 'Ashbya aceri'
fFAS Kluyveromyces marxianus
PfaD Paenibacillus curdlanolyticus
PfaD Vibrio sp. NSP560
PfaD Fischerella sp. JSC-11
PfaD Desulfobacula toluolica
PfaD Rhodococcus opacus
PfaD Nostoc sp.
PfaD Streptomyces turgidiscabies
PfaD Anabaena variabilis
PfaD Gloeobacter violaceus
PfaD Shewanella oneidensis
FabK Enterococcus faecalis
FabK Negativicoccus succinicivorans
FabK Bavariicoccus seileri
FabK Streptococcus henryi
FabK Clostridium tyrobutyricum
FabK Lactobacillus versmoldensis
FabK Desulfitobacterium dichloroeliminans
FabK Streptococcus sp. SK643
FabK Thermoanaerobacter wiegelii
FabK Streptococcus pneumoniae

Figure S1 (related to Figure 1). Sequence alignments of DfnA/FabK/ER homologues. The sequence alignment shows the conserved insertion points in FabK and DfnA-ER for the extension elements. Insertions are colored according to figure 1. All sequences used are given in supplementary table S1.

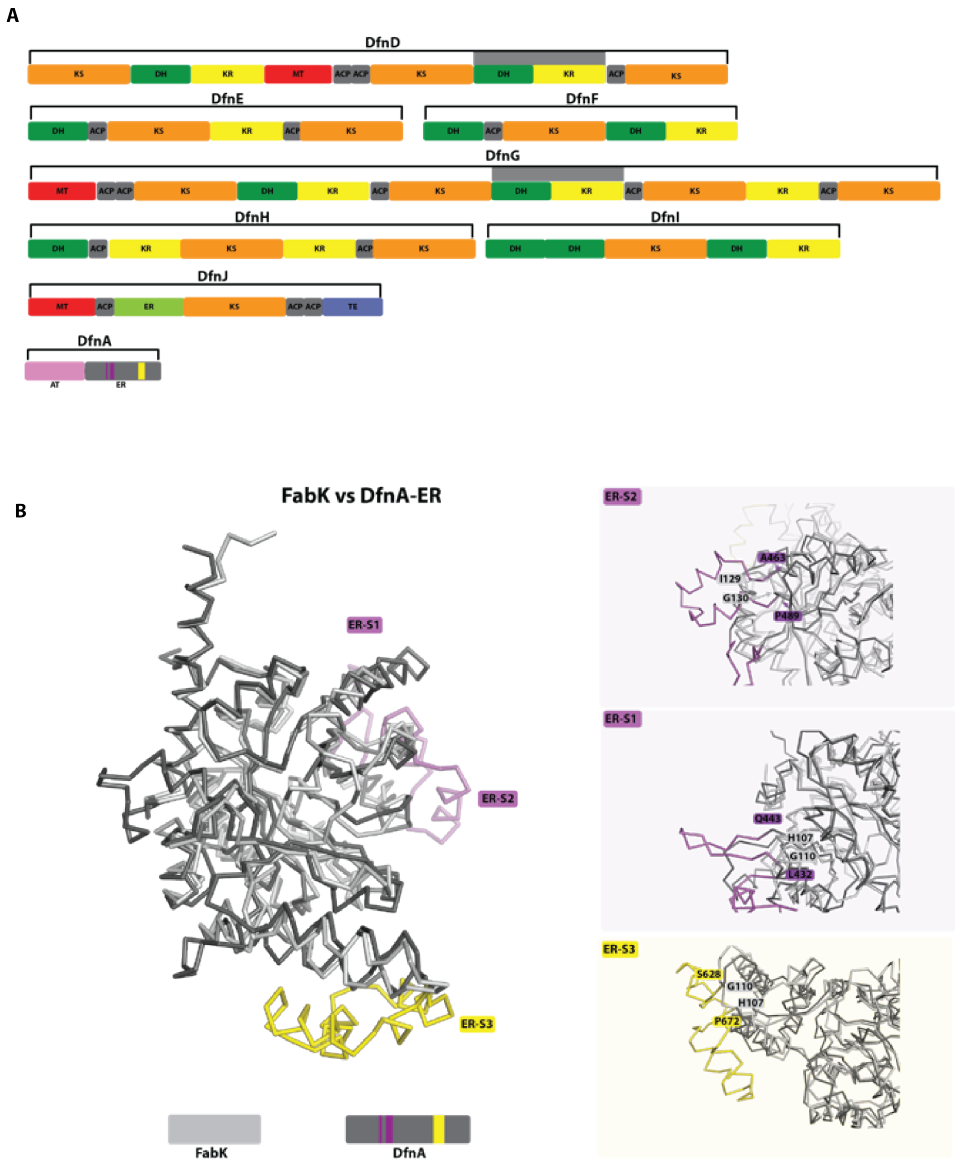


Figure S2 (related to Figure 2). Domain organization of the difficidin cluster from *B. amyloliquefaciens* FZB42 and comparison of FabK and DfnA-ER. ***(****A****)*** Schematic view of the difficidin cluster from *B. amyloliquefaciens* FZB42. KS, ketosynthase (orange); AT, acyl-CoA transferase (pink); KR, ketoreductase (yellow); ACP, acyl carrier protein (grey); DH, dehydratase (green); KR, ketoreductase (yellow); ER, enoyl reductase (drak grey); TE, thioesterase (dark blue). The location of the DHKR domain in DfnD and DfnG, where the DfnA-ER is supposed to act on the growing polyketide chain is indicated by a grey bar above the sequence. *(B)* Comparison of FabK and DfnA-ER and location of DfnA-ER specific insertion elements. (left) Superimposition of FabK (light grey) and DfnA-ER (dark grey). *Trans*-AT PKS specific elements (ER-S1 and ER-S2) are colored purple. The helical insertion in the (ER-S3) is shown in yellow. (right) Close-up view of insertion sites in DfnA-ER.

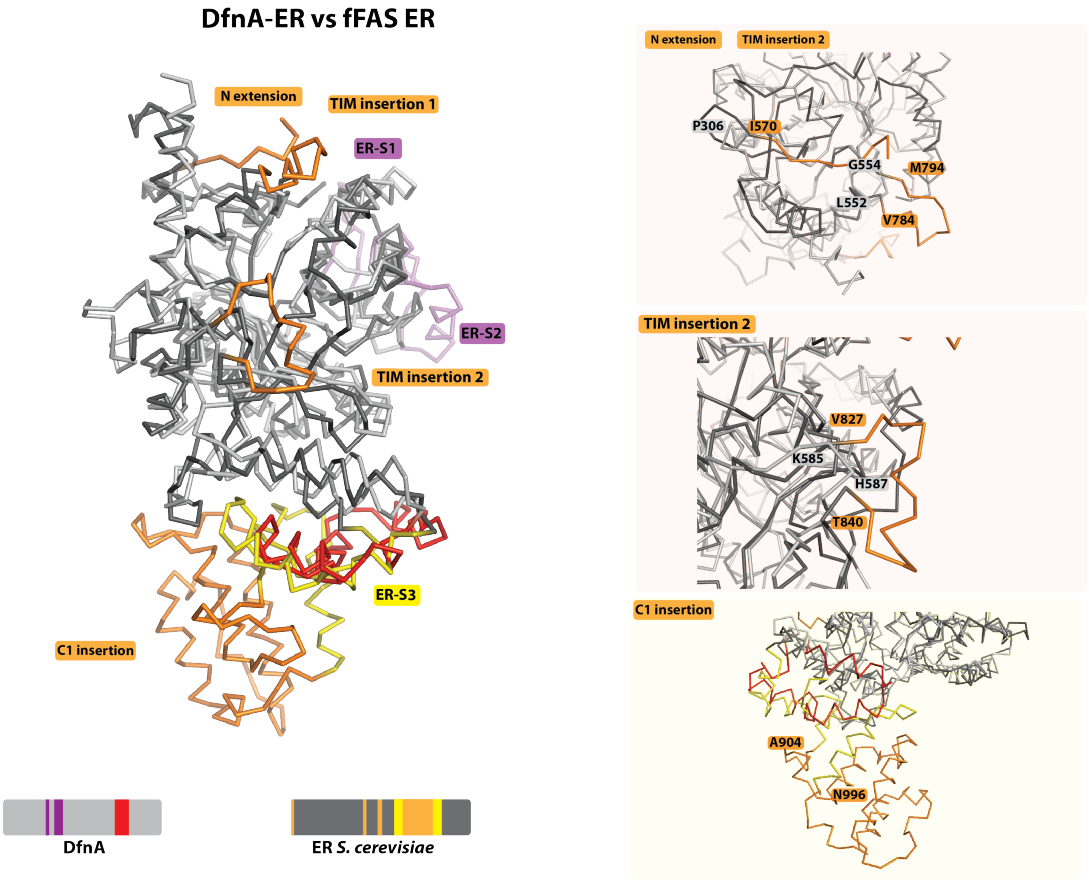
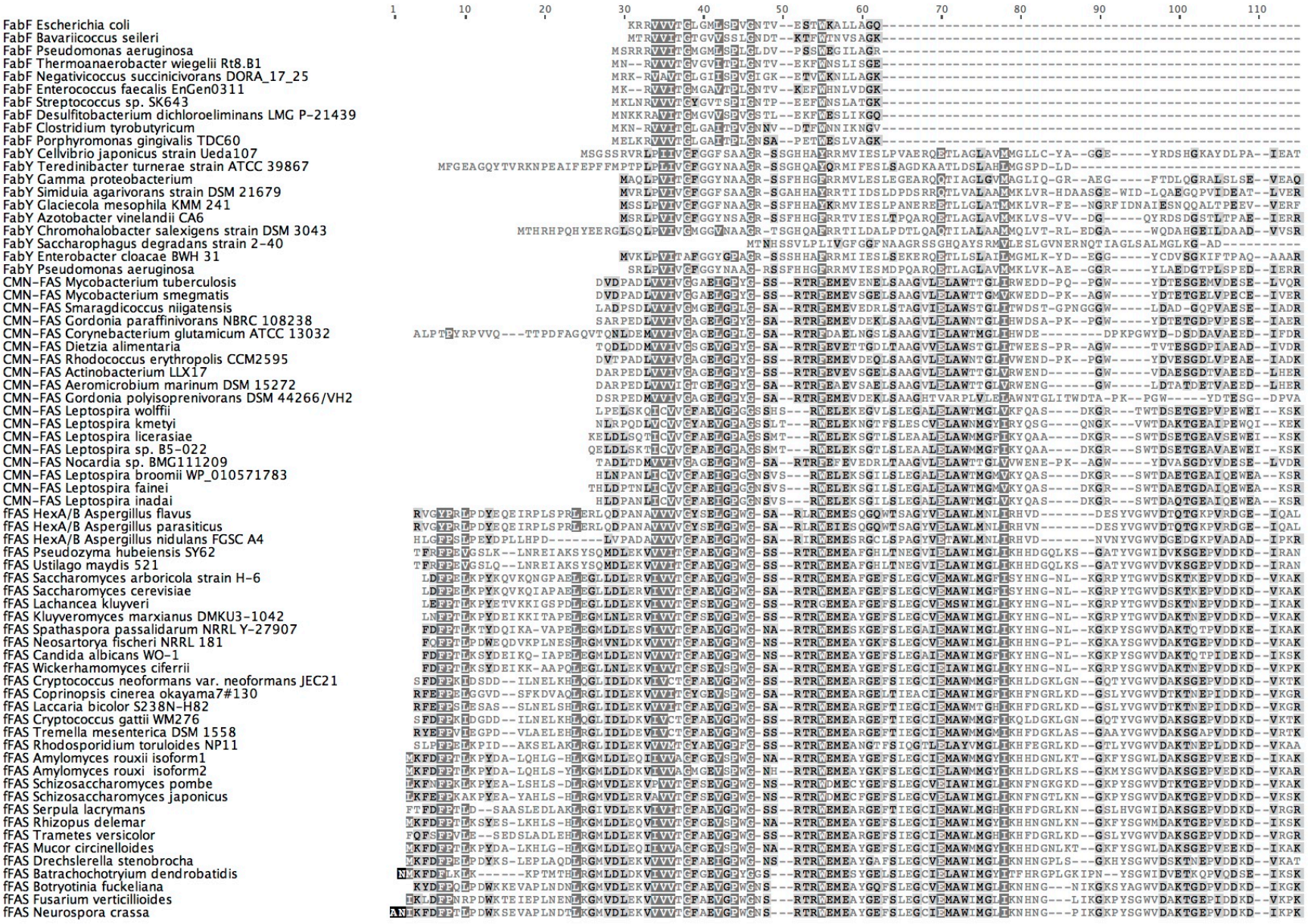
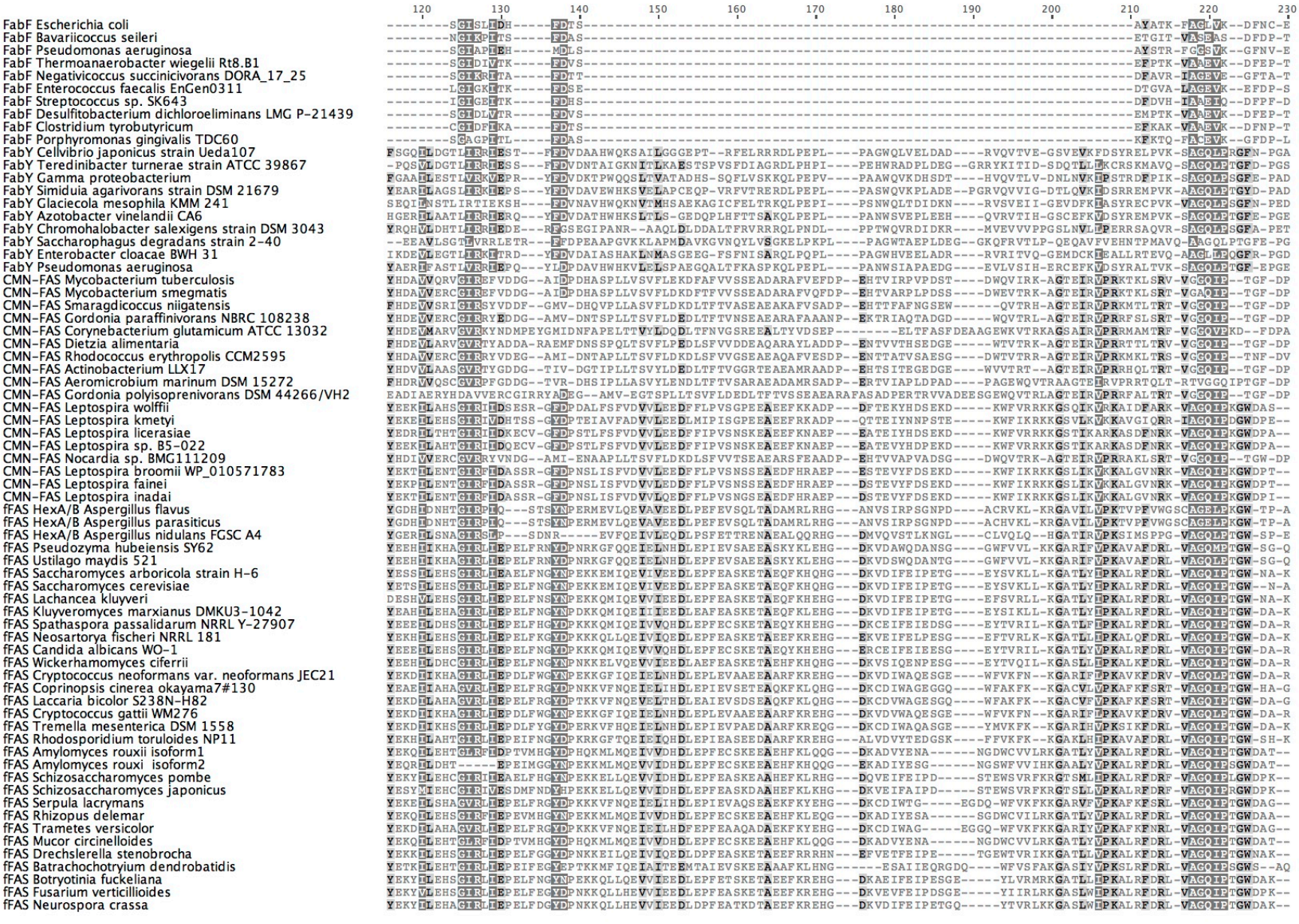
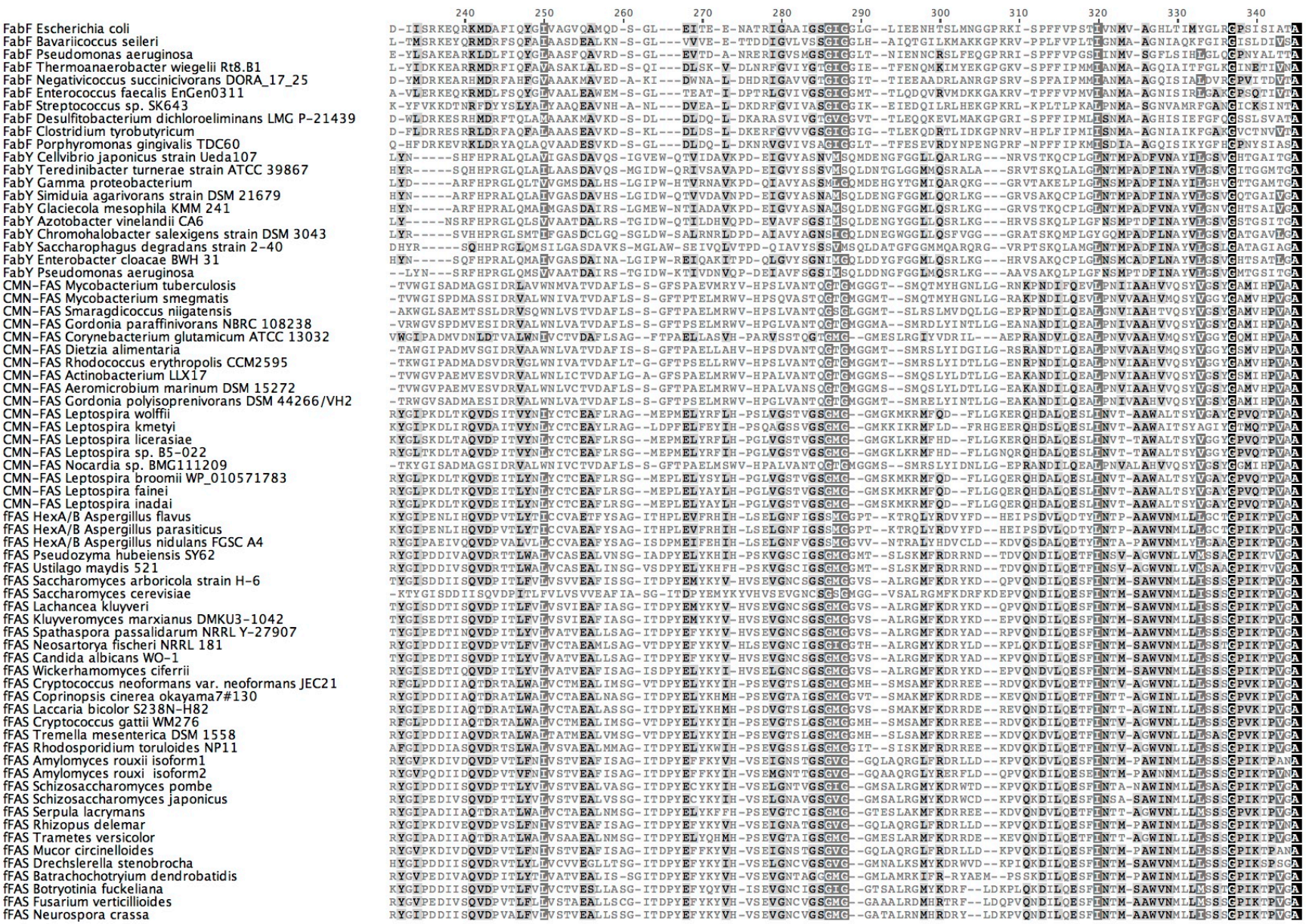
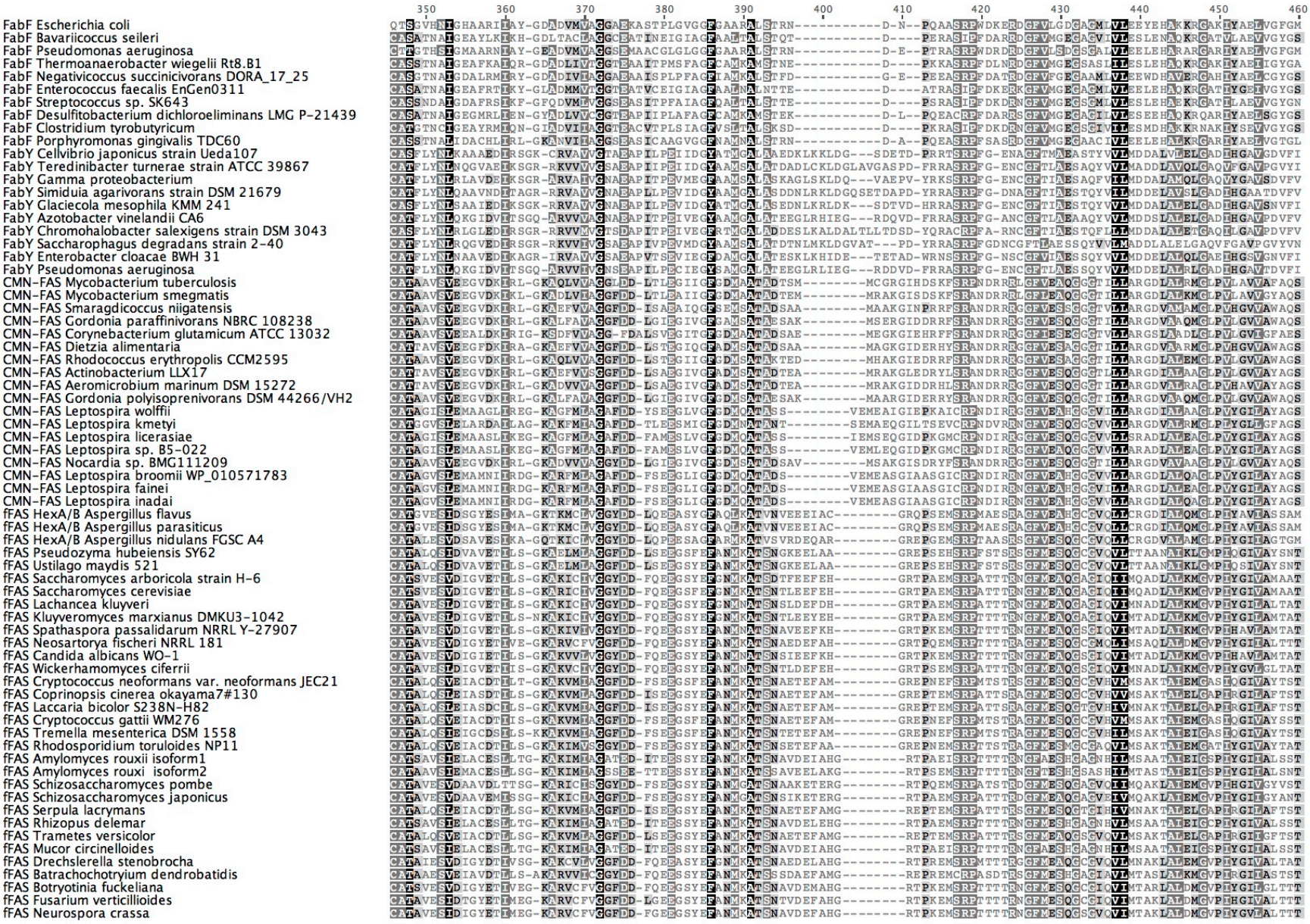


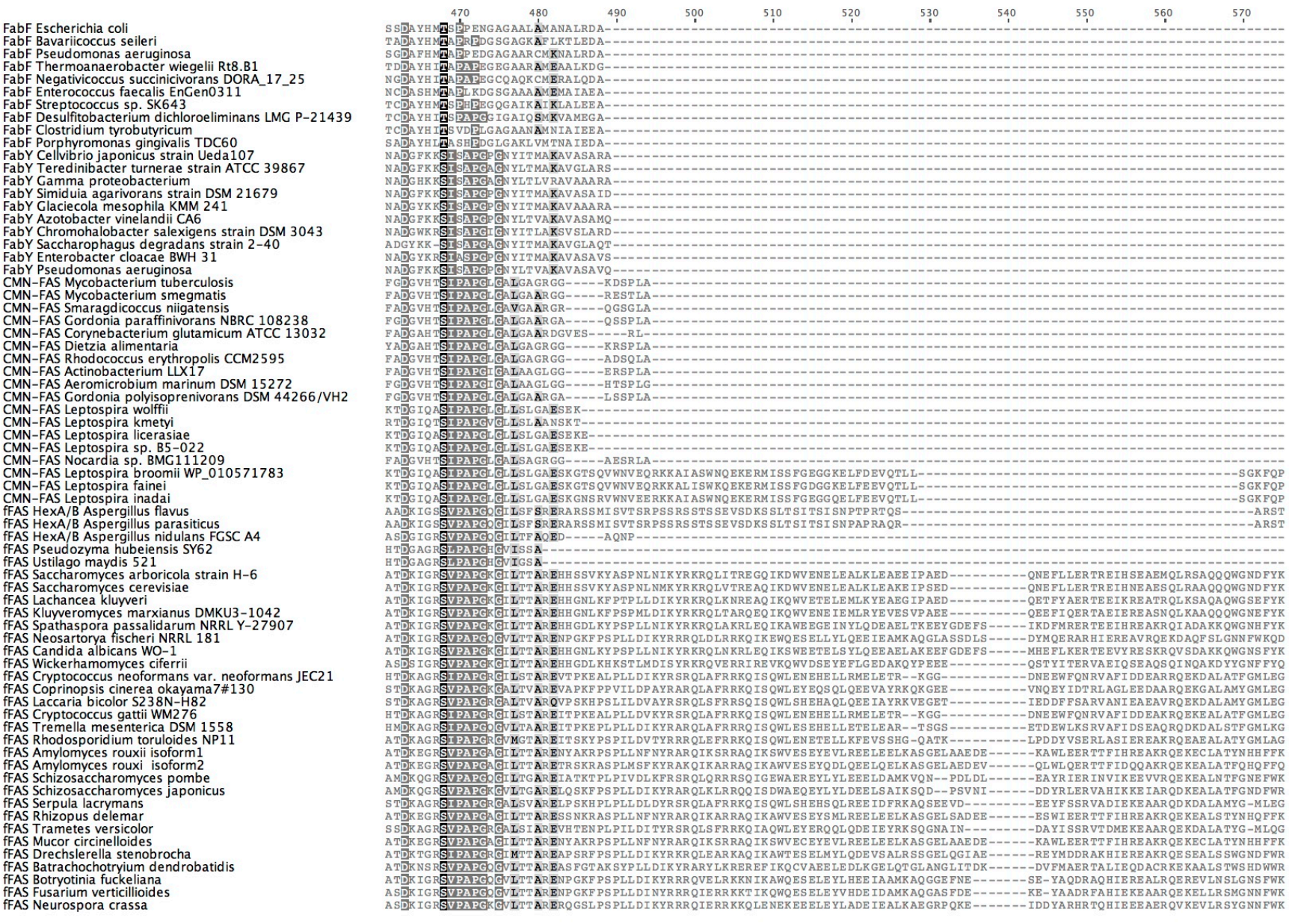
Figure S3 (related to Figure 2). Comparison of DfnA-ER and the fFAS ER. (left) Superposition of DfnA-ER (light grey) and fFAS ER (dark grey). *Trans*-AT PKS specific elements are colored purple. The α-helical insertion colored in red (DfnA ER) and yellow (fFAS ER) is conserved in both proteins. CMN- and fFAS specific helical insertions are colored orange. (right) Close-up view of fFAS specific insertions and their insertion sites.

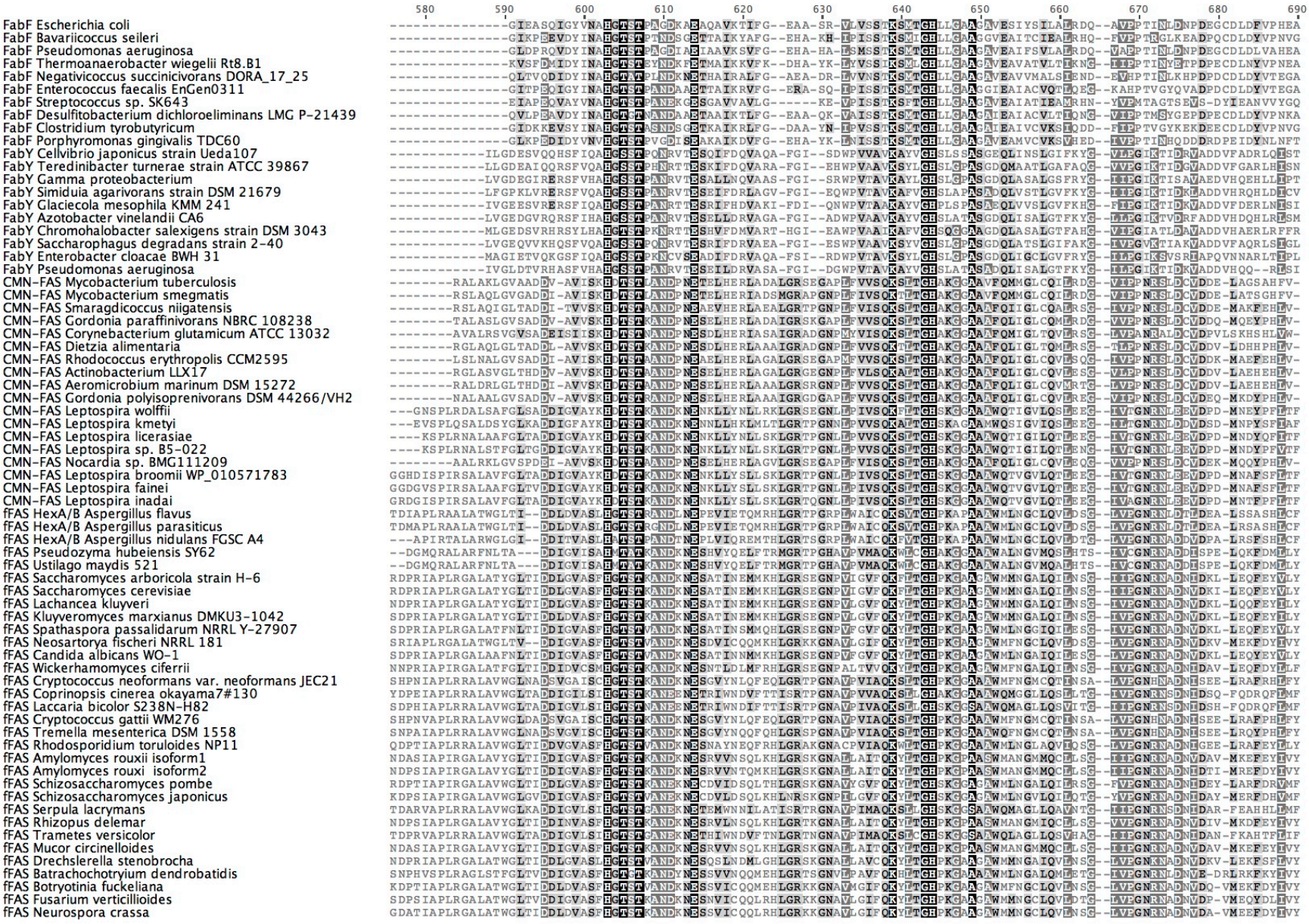

FabF Escherichia coli
FabF Bavariicoccus seileri
FabF Pseudomonas aeruginosa
FabF Thermoanaerobacter wiegelii Rt8.B1
FabF Negativicoccus succinicivorans DORA_17_25
FabF Enterococcus faecalis EnGen0311
FabF Streptococcus sp. SK643
FabF Desulfitobacterium dichloroeliminans LMG P-21439
FabF Clostridium tyrobutyricum
FabF Porphyromonas gingivalis TDC60
FabY Cellvibrio japonicus strain Ueda107
FabY Teredinibacter turnerae strain ATCC 39867
FabY Gamma proteobacterium
FabY Simiduia agarivorans strain DSM 21679
FabY Glaciecola mesophila KMM 241
FabY Azotobacter vinelandii CA6
FabY Chromohalobacter salexigens strain DSM 3043
FabY Saccharophagus degradans strain 2-40
FabY Enterobacter cloacae BWH 31
FabY Pseudomonas aeruginosa
CMN-FAS Mycobacterium tuberculosis
CMN-FAS Mycobacterium smegmatis
CMN-FAS Smaragdicoccus niigatensis
CMN-FAS Gordonia paraffinivorans NBRC 108238
CMN-FAS Corynebacterium glutamicum ATCC 13032
CMN-FAS Dietzia alimentaria
CMN-FAS Rhodococcus erythropolis CCM2595
CMN-FAS Actinobacterium LLX17
CMN-FAS Aeromicrobium marinum DSM 15272
CMN-FAS Gordonia polyisoprenivorans DSM 44266/VH2
CMN-FAS Leptospira wolffii
CMN-FAS Leptospira kmetyi
CMN-FAS Leptospira licerasiae
CMN-FAS Leptospira sp. B5-022
CMN-FAS Nocardia sp. BMG111209
CMN-FAS Leptospira broomii WP_010571783
CMN-FAS Leptospira fainei
CMN-FAS Leptospira inadai
fFAS HexA/B Aspergillus flavus
fFAS HexA/B Aspergillus parasiticus
fFAS HexA/B Aspergillus nidulans FGSC A4
fFAS Pseudozyma hubeiensis SY62
fFAS Ustilago maydis 521
fFAS Saccharomyces arboricola strain H-6
fFAS Saccharomyces cerevisiae
fFAS Lachancea kluyveri
fFAS Kluyveromyces marxianus DMKU3-1042
fFAS Spathaspora passalidarum NRRL Y-27907
fFAS Neosartorya fischeri NRRL 181
fFAS Candida albicans WO-1
fFAS Wickerhamomyces ciferrii
fFAS Cryptococcus neoformans var. neoformans JEC21
fFAS Coprinopsis cinerea okayama7#130
fFAS Laccaria bicolor S238N-H82
fFAS Cryptococcus gattii WM276
fFAS Tremella mesenterica DSM 1558
fFAS Rhodosporidium toruloides NP11
fFAS Amylomyces rouxii isoform1
fFAS Amylomyces rouxi isoform2
fFAS Schizosaccharomyces pombe
fFAS Schizosaccharomyces japonicus
fFAS Serpula lacrymans
fFAS Rhizopus delemar
fFAS Trametes versicolor
fFAS Mucor circinelloides
fFAS Drechslerella stenobrocha
fFAS Batrachochotryium dendrobatidis
fFAS Botryotinia fuckeliana
fFAS Fusarium verticillioides
fFAS Neurospora crassa

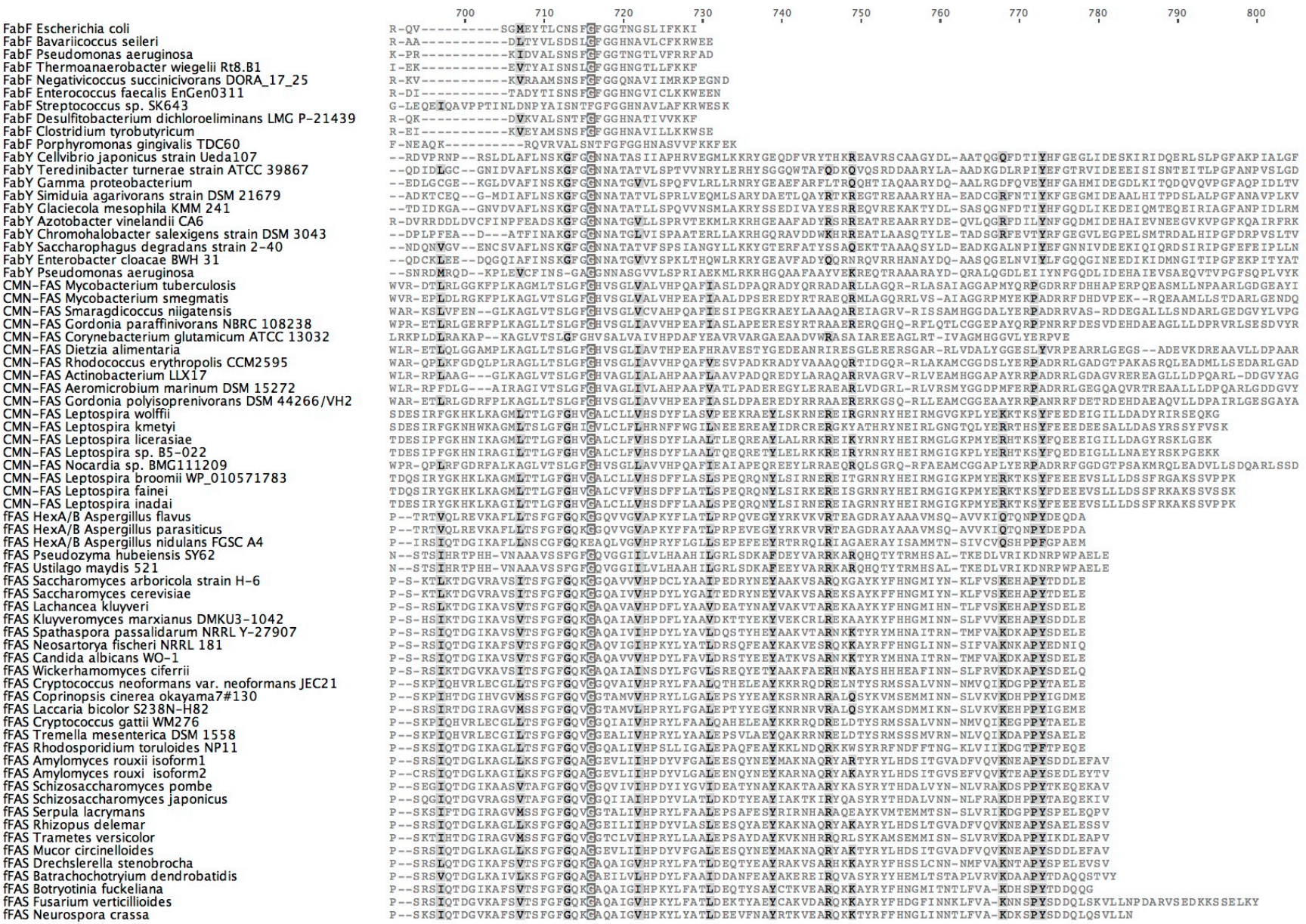

120 130 140 150 160 170 180 190 200 210 220 230
FabF Escherichia coli
FabF Bavariicoccus seileri
FabF Pseudomonas aeruginosa
FabF Thermoanaerobacter wiegelii Rt8.B1
FabF Negativicoccus succinicivorans DORA_17_25
FabF Enterococcus faecalis EnGen0311
FabF Streptococcus sp. SK643
FabF Desulfitobacterium dichloroeliminans LMG P-21439
FabF Clostridium tyrobutyricum
FabF Porphyromonas gingivalis TDC60
FabY Cellvibrio japonicus strain Ueda107
FabY Teredinibacter turnerae strain ATCC 39867
FabY Gamma proteobacterium
FabY Simiduia agarivorans strain DSM 21679
FabY Glaciecola mesophila KMM 241
FabY Azotobacter vinelandii CA6
FabY Chromohalobacter salexigens strain DSM 3043
FabY Saccharophagus degradans strain 2-40
FabY Enterobacter cloacae BWH 31
FabY Pseudomonas aeruginosa
CMN-FAS Mycobacterium tuberculosis
CMN-FAS Mycobacterium smegmatis
CMN-FAS Smaragdicoccus niigatensis
CMN-FAS Gordonia paraffinivorans NBRC 108238
CMN-FAS Corynebacterium glutamicum ATCC 13032
CMN-FAS Dietzia alimentaria
CMN-FAS Rhodococcus erythropolis CCM2595
CMN-FAS Actinobacterium LLX17
CMN-FAS Aeromicrobium marinum DSM 15272
CMN-FAS Gordonia polyisoprenivorans DSM 44266/VH2
CMN-FAS Leptospira wolffii
CMN-FAS Leptospira kmetyi
CMN-FAS Leptospira licerasiae
CMN-FAS Leptospira sp. B5-022
CMN-FAS Nocardia sp. BMG111209
CMN-FAS Leptospira broomii WP_010571783
CMN-FAS Leptospira fainei
CMN-FAS Leptospira inadai
fFAS HexA/B Aspergillus flavus
fFAS HexA/B Aspergillus parasiticus
fFAS HexA/B Aspergillus nidulans FGSC A4
fFAS Pseudozyma hubeiensis SY62
fFAS Ustilago maydis 521
fFAS Saccharomyces arboricola strain H-6
fFAS Saccharomyces cerevisiae
fFAS Lachancea kluyveri
fFAS Kluyveromyces marxianus DMKU3-1042
fFAS Spathaspora passalidarum NRRL Y-27907
fFAS Neosartorya fischeri NRRL 181
fFAS Candida albicans WO-1
fFAS Wickerhamomyces ciferrii
fFAS Cryptococcus neoformans var. neoformans JEC21
fFAS Coprinopsis cinerea okayama7#130
fFAS Laccaria bicolor S238N-H82
fFAS Cryptococcus gattii WM276
fFAS Tremella mesenterica DSM 1558
fFAS Rhodosporidium toruloides NP11
fFAS Amylomyces rouxii isoform1
fFAS Amylomyces rouxi isoform2
fFAS Schizosaccharomyces pombe
fFAS Schizosaccharomyces japonicus
fFAS Serpula lacrymans
fFAS Rhizopus delemar
fFAS Trametes versicolor
fFAS Mucor circinelloides
fFAS Drechslerella stenobrocha
fFAS Batrachochotryium dendrobatidis
fFAS Botryotinia fuckeliana
fFAS Fusarium verticillioides
fFAS Neurospora crassa


240 250 260 270 280 290 300 310 320 330 340
FabF Escherichia coli
FabF Bavariicoccus seileri
FabF Pseudomonas aeruginosa
FabF Thermoanaerobacter wiegelii Rt8.B1
FabF Negativicoccus succinicivorans DORA_17_25
FabF Enterococcus faecalis EnGen0311
FabF Streptococcus sp. SK643
FabF Desulfitobacterium dichloroeliminans LMG P-21439
FabF Clostridium tyrobutyricum
FabF Porphyromonas gingivalis TDC60
FabY Cellvibrio japonicus strain Ueda107
FabY Teredinibacter turnerae strain ATCC 39867
FabY Gamma proteobacterium
FabY Simiduia agarivorans strain DSM 21679
FabY Glaciecola mesophila KMM 241
FabY Azotobacter vinelandii CA6
FabY Chromohalobacter salexigens strain DSM 3043
FabY Saccharophagus degradans strain 2-40
FabY Enterobacter cloacae BWH 31
FabY Pseudomonas aeruginosa
CMN-FAS Mycobacterium tuberculosis
CMN-FAS Mycobacterium smegmatis
CMN-FAS Smaragdicoccus niigatensis
CMN-FAS Gordonia paraffinivorans NBRC 108238
CMN-FAS Corynebacterium glutamicum ATCC 13032
CMN-FAS Dietzia alimentaria
CMN-FAS Rhodococcus erythropolis CCM2595
CMN-FAS Actinobacterium LLX17
CMN-FAS Aeromicrobium marinum DSM 15272
CMN-FAS Gordonia polyisoprenivorans DSM 44266/VH2
CMN-FAS Leptospira wolffii
CMN-FAS Leptospira kmetyi
CMN-FAS Leptospira licerasiae
CMN-FAS Leptospira sp. B5-022
CMN-FAS Nocardia sp. BMG111209
CMN-FAS Leptospira broomii WP_010571783
CMN-FAS Leptospira fainei
CMN-FAS Leptospira inadai
fFAS HexA/B Aspergillus flavus
fFAS HexA/B Aspergillus parasiticus
fFAS HexA/B Aspergillus nidulans FGSC A4
fFAS Pseudozyma hubeiensis SY62
fFAS Ustilago maydis 521
fFAS Saccharomyces arboricola strain H-6
fFAS Saccharomyces cerevisiae
fFAS Lachancea kluyveri
fFAS Kluyveromyces marxianus DMKU3-1042
fFAS Spathaspora passalidarum NRRL Y-27907
fFAS Neosartorya fischeri NRRL 181
fFAS Candida albicans WO-1
fFAS Wickerhamomyces ciferrii
fFAS Cryptococcus neoformans var. neoformans JEC21
fFAS Coprinopsis cinerea okayama7#130
fFAS Laccaria bicolor S238N-H82
fFAS Cryptococcus gattii WM276
fFAS Tremella mesenterica DSM 1558
fFAS Rhodosporidium toruloides NP11
fFAS Amylomyces rouxii isoform1
fFAS Amylomyces rouxi isoform2
fFAS Schizosaccharomyces pombe
fFAS Schizosaccharomyces japonicus
fFAS Serpula lacrymans
fFAS Rhizopus delemar
fFAS Trametes versicolor
fFAS Mucor circinelloides
fFAS Drechslerella stenobrocha
fFAS Batrachochotryium dendrobatidis
fFAS Botryotinia fuckeliana
fFAS Fusarium verticillioides
fFAS Neurospora crassa


350
360
370
380
390
400
410
420
430
440
450
460
FabF Escherichia coli
FabF Bavariicoccus seileri
FabF Pseudomonas aeruginosa
FabF Thermoanaerobacter wiegelii Rt8.B1
FabF Negativicoccus succinicivorans DORA_17_25
FabF Enterococcus faecalis EnGen0311
FabF Streptococcus sp. SK643
FabF Desulfitobacterium dichloroeliminans LMG P-21439
FabF Clostridium tyrobutyricum
FabF Porphyromonas gingivalis TDC60
FabY Cellvibrio japonicus strain Ueda107
FabY Teredinibacter turnerae strain ATCC 39867
FabY Gamma proteobacterium
FabY Simiduia agarivorans strain DSM 21679
FabY Glaciecola mesophila KMM 241
FabY Azotobacter vinelandii CA6
FabY Chromohalobacter salexigens strain DSM 3043
FabY Saccharophagus degradans strain 2-40
FabY Enterobacter cloacae BWH 31
FabY Pseudomonas aeruginosa
CMN-FAS Mycobacterium tuberculosis
CMN-FAS Mycobacterium smegmatis
CMN-FAS Smaragdicoccus niigatensis
CMN-FAS Gordonia paraffinivorans NBRC 108238
CMN-FAS Corynebacterium glutamicum ATCC 13032
CMN-FAS Dietzia alimentaria
CMN-FAS Rhodococcus erythropolis CCM2595
CMN-FAS Actinobacterium LLX17
CMN-FAS Aeromicrobium marinum DSM 15272
CMN-FAS Gordonia polyisoprenivorans DSM 44266/VH2
CMN-FAS Leptospira wolffii
CMN-FAS Leptospira kmetyi
CMN-FAS Leptospira licerasiae
CMN-FAS Leptospira sp. B5-022
CMN-FAS Nocardia sp. BMG111209
CMN-FAS Leptospira broomii WP_010571783
CMN-FAS Leptospira fainei
CMN-FAS Leptospira inadai
fFAS HexA/B Aspergillus flavus
fFAS HexA/B Aspergillus parasiticus
fFAS HexA/B Aspergillus nidulans FGSC A4
fFAS Pseudozyma hubeiensis SY62
fFAS Ustilago maydis 521
fFAS Saccharomyces arboricola strain H-6
fFAS Saccharomyces cerevisiae
fFAS Lachancea kluyveri
fFAS Kluyveromyces marxianus DMKU3-1042
fFAS Spathaspora passalidarum NRRL Y-27907
fFAS Neosartorya fischeri NRRL 181
fFAS Candida albicans WO-1
fFAS Wickerhamomyces ciferrii
fFAS Cryptococcus neoformans var. neoformans JEC21
fFAS Coprinopsis cinerea okayama7#130
fFAS Laccaria bicolor S238N-H82
fFAS Cryptococcus gattii WM276
fFAS Tremella mesenterica DSM 1558
fFAS Rhodosporidium toruloides NP11
fFAS Amylomyces rouxii isoform1
fFAS Amylomyces rouxii isoform2
fFAS Schizosaccharomyces pombe
fFAS Schizosaccharomyces japonicus
fFAS Serpula lacrymans
fFAS Rhizopus delemar
fFAS Trametes versicolor
fFAS Mucor circinelloides
fFAS Drechslerella stenobrocha
fFAS Batrachochotryium dendrobatidis
fFAS Botryotinia fuckeliana
fFAS Fusarium verticillioides
fFAS Neurospora crassa


470
480
490
500
510
520
530
540
550
560
570
FabF Escherichia coli
FabF Bavariicoccus seileri
FabF Pseudomonas aeruginosa
FabF Thermoanaerobacter wiegelii Rt8.B1
FabF Negativicoccus succinicivorans DORA_17_25
FabF Enterococcus faecalis EnGen0311
FabF Streptococcus sp. SK643
FabF Desulfitobacterium dichloroeliminans LMG P-21439
FabF Clostridium tyrobutyricum
FabF Porphyromonas gingivalis TDC60
FabY Cellvibrio japonicus strain Ueda107
FabY Teredinibacter turnerae strain ATCC 39867
FabY Gamma proteobacterium
FabY Simiduia agarivorans strain DSM 21679
FabY Glaciecola mesophila KMM 241
FabY Azotobacter vinelandii CA6
FabY Chromohalobacter salexigens strain DSM 3043
FabY Saccharophagus degradans strain 2-40
FabY Enterobacter cloacae BWH 31
FabY Pseudomonas aeruginosa
CMN-FAS Mycobacterium tuberculosis
CMN-FAS Mycobacterium smegmatis
CMN-FAS Smaragdicoccus niigatensis
CMN-FAS Gordonia paraffinivorans NBRC 108238
CMN-FAS Corynebacterium glutamicum ATCC 13032
CMN-FAS Dietzia alimentaria
CMN-FAS Rhodococcus erythropolis CCM2595
CMN-FAS Actinobacterium LLX17
CMN-FAS Aeromicrobium marinum DSM 15272
CMN-FAS Gordonia polyisoprenivorans DSM 44266/VH2
CMN-FAS Leptospira wolffii
CMN-FAS Leptospira kmetyi
CMN-FAS Leptospira licerasiae
CMN-FAS Leptospira sp. B5-022
CMN-FAS Nocardia sp. BMG111209
CMN-FAS Leptospira broomii WP_010571783
CMN-FAS Leptospira fainei
CMN-FAS Leptospira inadai
fFAS HexA/B Aspergillus flavus
fFAS HexA/B Aspergillus parasiticus
fFAS HexA/B Aspergillus nidulans FGSC A4
fFAS Pseudozyma hubeiensis SY62
fFAS Ustilago maydis 521
fFAS Saccharomyces arboricola strain H-6
fFAS Saccharomyces cerevisiae
fFAS Lachancea kluyveri
fFAS Kluyveromyces marxianus DMKU3-1042
fFAS Spathaspora passalidarum NRRL Y-27907
fFAS Neosartorya fischeri NRRL 181
fFAS Candida albicans WO-1
fFAS Wickerhamomyces ciferrii
fFAS Cryptococcus neoformans var. neoformans JEC21
fFAS Coprinopsis cinerea okayama7#130
fFAS Laccaria bicolor S238N-H82
fFAS Cryptococcus gattii WM276
fFAS Tremella mesenterica DSM 1558
fFAS Rhodosporidium toruloides NP11
fFAS Amylomyces rouxii isoform1
fFAS Amylomyces rouxii isoform2
fFAS Schizosaccharomyces pombe
fFAS Schizosaccharomyces japonicus
fFAS Serpula lacrymans
fFAS Rhizopus delemar
fFAS Trametes versicolor
fFAS Mucor circinelloides
fFAS Drechslerella stenobrocha
fFAS Batrachochotryium dendrobatidis
fFAS Botryotinia fuckeliana
fFAS Fusarium verticillioides
fFAS Neurospora crassa


580 590 600 610 620 630 640 650 660 670 680 690
FabF Escherichia coli
FabF Bavariicoccus seileri
FabF Pseudomonas aeruginosa
FabF Thermoanaerobacter wiegelii Rt8.B1
FabF Negativicoccus succinicivorans DORA_17_25
FabF Enterococcus faecalis EnGen0311
FabF Streptococcus sp. SK643
FabF Desulfitobacterium dichloroeliminans LMG P-21439
FabF Clostridium tyrobutyricum
FabF Porphyromonas gingivalis TDC60
FabY Cellvibrio japonicus strain Ueda107
FabY Teredinibacter turnerae strain ATCC 39867
FabY Gamma proteobacterium
FabY Simiduia agarivorans strain DSM 21679
FabY Glaciecola mesophila KMM 241
FabY Azotobacter vinelandii CA6
FabY Chromohalobacter salexigens strain DSM 3043
FabY Saccharophagus degradans strain 2-40
FabY Enterobacter cloacae BWH 31
FabY Pseudomonas aeruginosa
CMN-FAS Mycobacterium tuberculosis
CMN-FAS Mycobacterium smegmatis
CMN-FAS Smaragdicoccus niigatensis
CMN-FAS Gordonia paraffinivorans NBRC 108238
CMN-FAS Corynebacterium glutamicum ATCC 13032
CMN-FAS Dietzia alimentaria
CMN-FAS Rhodococcus erythropolis CCM2595
CMN-FAS Actinobacterium LLX17
CMN-FAS Aeromicrobium marinum DSM 15272
CMN-FAS Gordonia polyisoprenivorans DSM 44266/VH2
CMN-FAS Leptospira wolffii
CMN-FAS Leptospira kmetyi
CMN-FAS Leptospira licerasiae
CMN-FAS Leptospira sp. B5-022
CMN-FAS Nocardia sp. BMG111209
CMN-FAS Leptospira broomii WP_010571783
CMN-FAS Leptospira fainei
CMN-FAS Leptospira inadai
fFAS HexA/B Aspergillus flavus
fFAS HexA/B Aspergillus parasiticus
fFAS HexA/B Aspergillus nidulans FGSC A4
fFAS Pseudozyma hubeiensis SY62
fFAS Ustilago maydis 521
fFAS Saccharomyces arboricola strain H-6
fFAS Saccharomyces cerevisiae
fFAS Lachancea kluyveri
fFAS Kluyveromyces marxianus DMKU3-1042
fFAS Spathaspora passalidarum NRRL Y-27907
fFAS Neosartorya fischeri NRRL 181
fFAS Candida albicans WO-1
fFAS Wickerhamomyces ciferrii
fFAS Cryptococcus neoformans var. neoformans JEC21
fFAS Coprinopsis cinerea okayama7#130
fFAS Laccaria bicolor S238N-H82
fFAS Cryptococcus gattii WM276
fFAS Tremella mesenterica DSM 1558
fFAS Rhodosporidium toruloides NP11
fFAS Amylomyces rouxii isoform1
fFAS Amylomyces rouxii isoform2
fFAS Schizosaccharomyces pombe
fFAS Schizosaccharomyces japonicus
fFAS Serpula lacrymans
fFAS Rhizopus delemar
fFAS Trametes versicolor
fFAS Mucor circinelloides
fFAS Drechslerella stenobrocha
fFAS Batrachochotryium dendrobatidis
fFAS Botryotinia fuckeliana
fFAS Fusarium verticillioides
fFAS Neurospora crassa


700
710
720
730
740
750
760
770
780
790
800
FabF Escherichia coli
FabF Bavariicoccus seileri
FabF Pseudomonas aeruginosa
FabF Thermoanaerobacter wiegelii Rt8.B1
FabF Negativicoccus succinicivorans DORA_17_25
FabF Enterococcus faecalis EnGen0311
FabF Streptococcus sp. SK643
FabF Desulfitobacterium dichloroeliminans LMG P-21439
FabF Clostridium tyrobutyricum
FabF Porphyromonas gingivalis TDC60
FabY Cellvibrio japonicus strain Ueda107
FabY Teredinibacter turnerae strain ATCC 39867
FabY Gamma proteobacterium
FabY Simiduia agarivorans strain DSM 21679
FabY Glaciecola mesophila KMM 241
FabY Azotobacter vinelandii CA6
FabY Chromohalobacter salexigens strain DSM 3043
FabY Saccharophagus degradans strain 2-40
FabY Enterobacter cloacae BWH 31
FabY Pseudomonas aeruginosa
CMN-FAS Mycobacterium tuberculosis
CMN-FAS Mycobacterium smegmatis
CMN-FAS Smaragdicoccus niigatensis
CMN-FAS Gordonia paraffinivorans NBRC 108238
CMN-FAS Corynebacterium glutamicum ATCC 13032
CMN-FAS Dietzia alimentaria
CMN-FAS Rhodococcus erythropolis CCM2595
CMN-FAS Actinobacterium LLX17
CMN-FAS Aeromicrobium marinum DSM 15272
CMN-FAS Gordonia polyisoprenivorans DSM 44266/VH2
CMN-FAS Leptospira wolffii
CMN-FAS Leptospira kmetyi
CMN-FAS Leptospira licerasiae
CMN-FAS Leptospira sp. B5-022
CMN-FAS Nocardia sp. BMG111209
CMN-FAS Leptospira broomii WP_010571783
CMN-FAS Leptospira fainei
CMN-FAS Leptospira inadai
fFAS HexA/B Aspergillus flavus
fFAS HexA/B Aspergillus parasiticus
fFAS HexA/B Aspergillus nidulans FGSC A4
fFAS Pseudozyma hubeiensis SY62
fFAS Ustilago maydis 521
fFAS Saccharomyces arboricola strain H-6
fFAS Saccharomyces cerevisiae
fFAS Lachancea kluyveri
fFAS Kluyveromyces marxianus DMKU3-1042
fFAS Spathaspora passalidarum NRRL Y-27907
fFAS Neosartorya fischeri NRRL 181
fFAS Candida albicans WO-1
fFAS Wickerhamomyces ciferrii
fFAS Cryptococcus neoformans var. neoformans JEC21
fFAS Coprinopsis cinerea okayama7#130
fFAS Laccaria bicolor S238N-H82
fFAS Cryptococcus gattii WM276
fFAS Tremella mesenterica DSM 1558
fFAS Rhodosporidium toruloides NP11
fFAS Amylomyces rouxii isoform1
fFAS Amylomyces rouxi isoform2
fFAS Schizosaccharomyces pombe
fFAS Schizosaccharomyces japonicus
fFAS Serpula lacrymans
fFAS Rhizopus delemar
fFAS Trametes versicolor
fFAS Mucor circinelloides
fFAS Drechslerella stenobrocha
fFAS Batrachochotryium dendrobatidis
fFAS Botryotinia fuckeliana
fFAS Fusarium verticillioides
fFAS Neurospora crassa

```
                                                    810       821
FabF Escherichia coli
FabF Bavariicoccus seileri
FabF Pseudomonas aeruginosa
FabF Thermoanaerobacter wiegelii Rt8.B1
FabF Negativicoccus succinicivorans DORA_17_25
FabF Enterococcus faecalis EnGen0311
FabF Streptococcus sp. SK643
FabF Desulfitobacterium dichloroeliminans LMG P-21439
FabF Clostridium tyrobutyricum
FabF Porphyromonas gingivalis TDC60
FabY Cellvibrio japonicus strain Ueda107             SNPYADMFD
FabY Teredinibacter turnerae strain ATCC 39867      NEGY-DF
FabY Gamma proteobacterium                          VSEFADMLRSPCER
FabY Simiduia agarivorans strain DSM 21679          KSPYKDLVK
FabY Glaciecola mesophila KMM 241                   KNPFEDMI
FabY Azotobacter vinelandii CA6                     DERFGDMLD
FabY Chromohalobacter salexigens strain DSM 3043    DNPYGQL
FabY Saccharophagus degradans strain 2-40           DEGYGEF
FabY Enterobacter cloacae BWH 31                    DKQYPDF
FabY Pseudomonas aeruginosa                         KDARFSDMLDA
CMN-FAS Mycobacterium tuberculosis                  G
CMN-FAS Mycobacterium smegmatis                     YVL
CMN-FAS Smaragdicoccus niigatensis                  AGCR
CMN-FAS Gordonia paraffinivorans NBRC 108238        PADAVGADERS
CMN-FAS Corynebacterium glutamicum ATCC 13032
CMN-FAS Dietzia alimentaria                         INDDGVLASSTGAGDE
CMN-FAS Rhodococcus erythropolis CCM2595            LVYRSN----LPGCK
CMN-FAS Actinobacterium LLX17                       PACR
CMN-FAS Aeromicrobium marinum DSM 15272             AGPACR
CMN-FAS Gordonia polyisoprenivorans DSM 44266/VH2   DAAQDVSSSSSAATI
CMN-FAS Leptospira wolffii
CMN-FAS Leptospira kmetyi
CMN-FAS Leptospira licerasiae
CMN-FAS Leptospira sp. B5-022
CMN-FAS Nocardia sp. BMG111209                      GVYIPTGCTTTPSAK
CMN-FAS Leptospira broomii WP_010571783
CMN-FAS Leptospira fainei
CMN-FAS Leptospira inadai
fFAS HexA/B Aspergillus flavus
fFAS HexA/B Aspergillus parasiticus
fFAS HexA/B Aspergillus nidulans FGSC A4
fFAS Pseudozyma hubeiensis SY62
fFAS Ustilago maydis 521
fFAS Saccharomyces arboricola strain H-6
fFAS Saccharomyces cerevisiae
fFAS Lachancea kluyveri
fFAS Kluyveromyces marxianus DMKU3-1042
fFAS Spathaspora passalidarum NRRL Y-27907
fFAS Neosartorya fischeri NRRL 181
fFAS Candida albicans WO-1
fFAS Wickerhamomyces ciferrii
fFAS Cryptococcus neoformans var. neoformans JEC21
fFAS Coprinopsis cinerea okayama7#130
fFAS Laccaria bicolor S238N-H82
fFAS Cryptococcus gattii WM276
fFAS Tremella mesenterica DSM 1558
fFAS Rhodosporidium toruloides NP11
fFAS Amylomyces rouxii isoform1
fFAS Amylomyces rouxi  isoform2
fFAS Schizosaccharomyces pombe
fFAS Schizosaccharomyces japonicus
fFAS Serpula lacrymans
fFAS Rhizopus delemar
fFAS Trametes versicolor
fFAS Mucor circinelloides
fFAS Drechslerella stenobrocha
fFAS Batrachochotryium dendrobatidis
fFAS Botryotinia fuckeliana
fFAS Fusarium verticillioides
fFAS Neurospora crassa
```

Figure S4 (related to Figure 4). Sequence alignments of FabY homologues, insertions are colored like in figure 4. All sequences used are given in supplementary table S1.

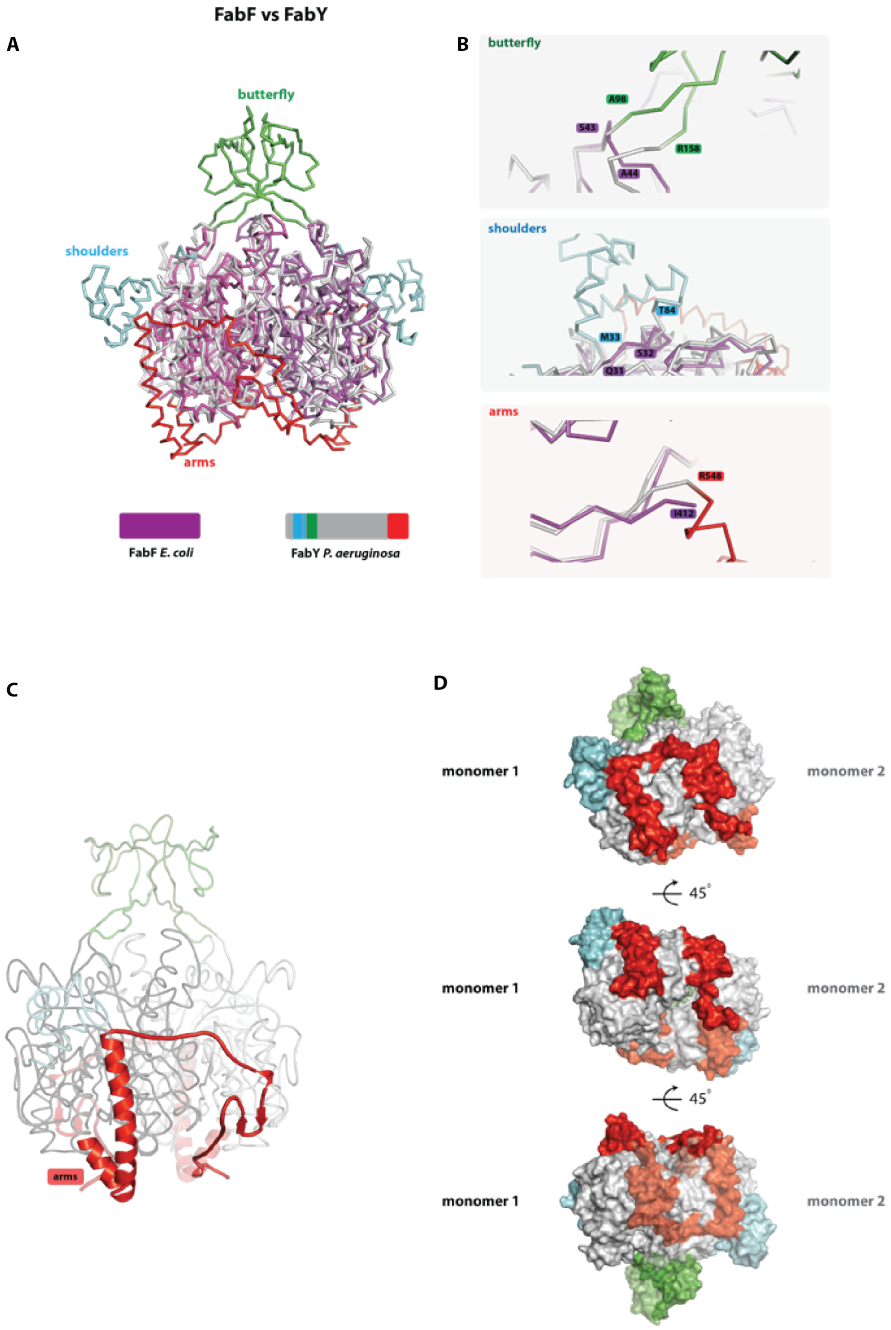


Figure S5 (related to Figure 5). Comparison of FabF and FabY. *(A)* Superposition of FabF (2GFW.pdb) (purple) and FabY (dark grey). The butterfly and shoulder insertions in FabY are shown in green and cyan, respectively. The C-terminal extension (arms) is colored red. *(B)* Close-up view of insertion sites in FabY. The arm expansion segment extends over the full surface of FabY. *(C)* The KS core domain, the butterfly, shoulder and arm extension are colored in light grey, light green, light cyan and red. *(D)* Rotated views of a surface representation highlighting the arm extension in FabY

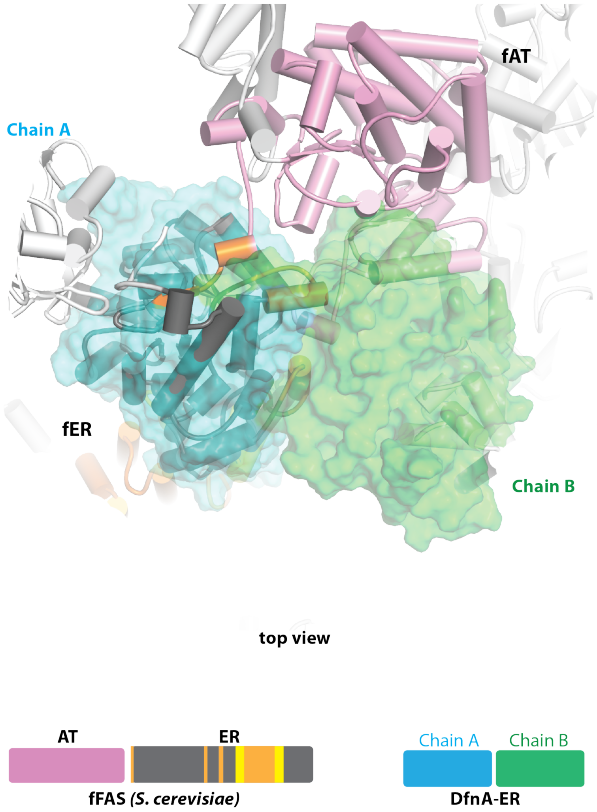


Figure S6 (related to Figure 6). DfnA-ER has small overlap with the interdomain interface between the AT and ER domain in fFAS. DfnA-ER is superimposed on fFAS (2UV9.pdb). DfnA-ER molecule A and B are shown in blue and green, respectively. The AT, the core of the fungal ER and fungal ER insertions are shown in pink, dark grey and orange, respectively. (left) The fFAS is in cartoon and the DfnA-ER in surface representation.

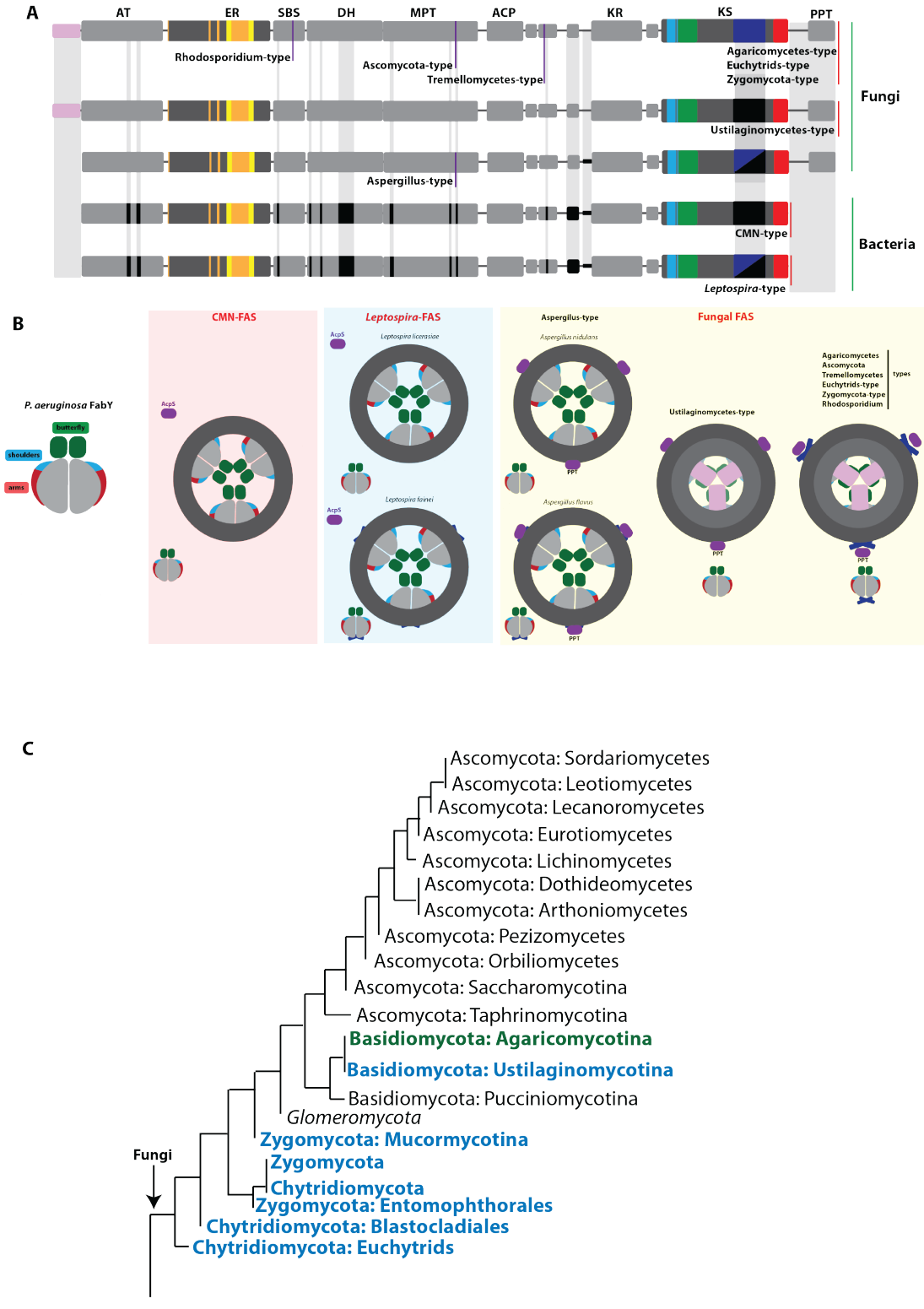

A
AT
ER
SBS
DH
MPT
ACP
KR
KS
PPT
Rhodosporidium-type
Ascomycota-type
Tremellomycetes-type
Agaricomycetes-type
Euchytrids-type
Zygomycota-type
Fungi
Ustilaginomycetes-type
Aspergillus-type
CMN-type
Bacteria
Leptospira-type
B
P. aeruginosa FabY
butterfly
shoulders
arms
CMN-FAS
Leptospira-FAS
AcpS
Leptospira licerasiae
Leptospira fainei
Aspergilus-type
Aspergillus nidulans
Aspergillus flavus
PPT
Fungal FAS
Ustilaginomycetes-type
Agaricomycetes
Ascomycota
Tremellomycetes
Euchytrids-type
Zygomycota-type
Rhodosporidium
types
C
Ascomycota: Sordariomycetes
Ascomycota: Leotiomycetes
Ascomycota: Lecanoromycetes
Ascomycota: Eurotiomycetes
Ascomycota: Lichinomycetes
Ascomycota: Dothideomycetes
Ascomycota: Arthoniomycetes
Ascomycota: Pezizomycetes
Ascomycota: Orbiliomycetes
Ascomycota: Saccharomycotina
Ascomycota: Taphrinomycotina
Basidiomycota: Agaricomycotina
Basidiomycota: Ustilaginomycotina
Basidiomycota: Pucciniomycotina
Glomeromycota
Zygomycota: Mucormycotina
Zygomycota
Chytridiomycota
Zygomycota: Entomophthorales
Chytridiomycota: Blastocladiales
Chytridiomycota: Euchytrids
Fungi

Figure S7 (related to Figure 7). Schematic comparison of fungal FAS I subtypes and distribution of single and split genes of fFAS in the fungal kingdom. (*A*) Comparison of domain structures of type I FAS from bacteria and fungi. A full-length sequence carrying all currently known insertions (e.g. *Ascomycota*) is given on top, and deletions in the respective FAS types are shown in black. Purple lines for the different fungal species indicate the partitioning in two chains. Insertions conserved between in *trans*-AT PKS/PUFA-ER and Type I FAS ER are shown in yellow and CMN-/fFAS specific insertions in orange. An N-terminal extension of AT is shown in pink. The DM3 domain is colored cyan and green; the DM4 and CIS insertions of KS are shown in blue and red, respectively. In *Aspergillus*- and *Leptospira*-type either no or a smaller DM4 insertion (50 % sequence length) is present (half filled blue/black). (*B*) Variations in KS of FabY and CMN-/fFAS together with an analysis of 'closed' and 'open' barrel structures (left) Schematic representation of FabY. The conserved KS core is shown in light gray, the insertion of butterfly shoulders and arms are shown in green, cyan and red, respectively. KS extensions of each FAS subtype are shown as an inset together with a schematic representation of the overall architecture. In bacteria, the 4'-phosphopantetheinyl transferase (PPT) activity is provided *in trans* by the standalone AcpS protein (violet) whereas the PPT is always part of the multienzyme in fFAS. In *Aspergillus* and *Ustilaginomycetes* species, the PPT is directly attached to the KS, where as in *Agaricomycetes/Ascomycoty/ Tremellocytes/Rhodospordium/Euchytrids /Zygomycota* (right site) the PPT is anchored to the DM4 region (dark blue). This type of fFAS, as well as the *Ustilaginomycetes* FAS are characterized by an N-terminal extension (pink) to the AT domain, which closes the FAS barrel. This overview demonstrates that closing of the FAS barrel is not coupled to the presence of the DM4 domain. *(C)* Distribution of single and split genes of fFAS in the fungal kingdom. A schematic representation of fungal phylogeny is adapted from [108]. Fungal phyla are colored based on the occurrence of FAS variants: Single gene FAS (blue), Split (two) gene FAS (black) or both variants (single/split gene) FAS (green). Phyla with no DNA sequencing information for fFAS are shown in italic. The most ancient fungal phyla are characterized by single FAS genes and splitting occurred later in evolution around the development of *Glomeromycota*, *Basidiomyceta* and *Ascomyceta*.

Supplementary References

James, T.Y., Kauff, F., Schoch, C.L., Matheny, P.B., Hofstetter, V., Cox, C.J., Celio, G., Gueidan, C., Fraker, E., Miadlikowska, J.*, et al.* (2006). Reconstructing the early evolution of Fungi using a six-gene phylogeny. Nature *443*, 818-822.

Jenke-Kodama, H., Sandmann, A., Muller, R., and Dittmann, E. (2005). Evolutionary implications of bacterial polyketide synthases. Molecular biology and evolution *22*, 2027-2039.

Part II

Emergent properties of Dynamic Multienzymes: The influence of interdomain linking on animal FAS multienzyme kinetics

Author contributions

Timm Maier, and I designed the project.

I established cloning strategies, expressed and purified all constructs of the porcine FAS and linkers.

I performed kinetic studies of the linker constructs. Timm Maier supervised the project and contributed to the analysis of results.

Summary

In this project we altered mechanical and physical properties of linker regions in the animal FAS to better characterize how intradomain linking influences catalytic properties and conformational crosstalk between domains. This was achieved by generating more then 40 different constructs with varied linker sequences. Combined structural and kinetic data from purified constructs helped us to better understand the emergent properties of the megasynthase system. A long-term goal is to use these insights for the construction of artificial multienzymes incorporating complete and complex molecular pathways.

Introduction

The *de novo* synthesis of long chain fatty acids from malonyl-CoA is an iterative process that requires seven catalytic activities. In most bacteria and plants the associated enzymes are discrete monofunctional polypeptides[18]. Two strikingly different types of fatty acid synthase (FAS) systems have evolved in fungi and in animals[57]. Fungal FAS is a rigid, 2.6-MDa barrel-shaped α6β6-heterododecamer with 50% of its polypeptide chains providing scaffolding elements[16,58]. Contrastingly, the animal FAS consists of a 270-kDa polypeptide chain comprising all seven enzymatic activities, that assemble into an intertwined, X-shaped 540kDa homodimeric multienzyme[17] (Figure 1). Mammalian FAS (mFAS) is divided into a condensing portion containing the β-ketoacyl synthase (KS), malonyl-/acetyl- transferase (MAT) domains and an β-carbon modification section, consisting of the enzymatic dehydratase (DH), enoylreductase (ER) and β-ketoreductase (KR) domains as well as two non-enzymatic domains, the pseudo-KR (pKR) and the pseudo-methyltransferase (pME)[48] (Figure 1 A). All reaction intermediates are covalently bound to an ACP and translocate between the active sites during the catalysis[46] .

The fatty acid production cycle in animal FAS is initiated by the transfer of the acyl moiety of the starter substrate acetyl-CoA onto the terminal thiol of the phosphopantheteine (P-Pant) cofactor of the ACP, catalyzed by the double-specific MAT[46]. Then, ACP passes the acetyl group to the active cysteine on the KS. In the next step the elongation unit malonyl-CoA is loaded onto ACP by MAT. The KS catalyzes the decarboxylative condensation of the acyl intermediate with malonyl-ACP. KR, DH and ER further modify the product at the ß- carbon position yielding a fully saturated acyl chain elongated by two carbon units. This acyl chain

functions as a starter substrate for the next round of extension, until the growing fatty acid reaches a length of 16 to 18 carbon atoms and is released by thioesterase (TE) from ACP.

Prior to each round of chain extension, the growing acyl chain is detached from the ACP and transferred to the active-site cysteine of the KS[50]. This transfer requires the P-Pant moiety to thread the acyl chain into the KS active site. Interaction with the ß -carbon processing domains does not require detachment of the acyl chain from the hydrophobic core of the ACP[46]. In animal FAS ACP is connected via flexible linker (approx. 14 aa) to the KR domain[17,48]. In the X ray structure of porcine FAS, both ACP and the following TE domains were not visible due to their flexibility[17,48]. Modelling an ACP to the porcine FAS structure revealed that distances between the P-Pant thiol of the one ACP are not sufficient to reach active sites in one reaction chamber[17,48]. However, early biochemical evidence, including mutant-complementation analyses and site-specific crosslinking, indicated that the ACP domains can functionally contact to the KS and MAT domains of both subunits[109]. In the crystal structure it was observed that two reaction chambers do not appear in identical conformations; the distances between active sites associated with the lower and upper parts on the left cleft are different to the right cleft[17,48]. Therefore, it was proposed that efficient substrate shuttling and catalysis requires conformational changes, which could result from the flexibility in the linker regions[16,58].

These predictions were later confirmed by electron microscopy (EM) analysis that revealed the high structural flexibility of animal FAS: animal FAS can adopt more then 16 different conformational states[51]. In the absence of substrate, 2D image analysis revealed class averages with an identical organization of the upper β -carbon processing part, but the

lower parts of FAS rotated by about 90 degree relative to upper section of the molecule making it nearly perpendicular[51]. It is also possible that this conformation could represent a full 180-degree rotation of the upper relative to the lower part. Once FAS was substrate-loaded the majority of the molecule showed asymmetric arrangement in the upper part most likely resulting from internal flexibility in the β-carbon processing domains[51].

The conformational flexibility in animal FAS originates from the linking regions which make approximately 9% of overall amino acid sequence[48]. Some linker regions in the FAS appear to be conserved and well structured. For example, the linker region between KS and AT domains is composed of two short α-helices and a three-stranded antiparallel β-sheet which restricts the relative movement of the KS and AT domains[17]. Other linker regions such as KR-ACP or DH-MAT do not appear to have special folds or secondary structure elements. They are rich in alanine and serine residues, similar to the inter-domain segments of the dihydrolipoyl transacetylase polypeptide chains of pyruvate dehydrogenase, that provide conformational flexibility and the same time avoid collapsed conformations[110]. Until now, only the ACP to TE linker in animal FAS was examined[111]. Increasing the linker length by 13 amino acids did not alter overall FAS activity, whereas, decreasing the linker length by 22 residues reduced fatty acid synthesis by 28%[111].

Currently, it is unclear if and how intra-domain linking can influence catalytic properties and conformational domain crosstalks in anima FAS. It seems that some linker's function as flexible tethers, while other may function more like structured hinges. In order to address these questions we have generated more then 40 different FAS variants with increased or decreased linker lengths in two areas of porcine FAS; linker which links

ACP to the KR domain and linker between MAT-DH domains that connects upper and lower part of the animal FAS. Combined molecular dynamics and kinetic data from modified linker constructs enable us to better understand complex dynamics of animal FAS.

Materials and Methods

Cloning of full-length pFAS.

Full-length pFAS (FL pFAS) was synthetically synthesized (gene script). Linear fragment containing FL pFAS was cloned in to the pIDK vector (MultiBac) using Red ET recombination and sequentially transferred into the EmBacY following standart MultiBac protocol.

Cloning of the linker constructs.

Mutagenesis on the original pFAS insert and vector was problematic because of the large size of the construct. Therefore two-step approach was developed. In the first step small fragment usually 1-2 KB in size was isolated and mutated using PCR. In the next step this fragment was recombined with digested original plasmid using Red E/T recombination. For the KR-ACP linkers the length of the isolated fragment was 1.5 kb which was amplified using standard P1_KR_ACP_frw (AACGCATCAAGTC TGGACGGT) and P1_KR_ACP_rev (CATGCTATGCATCAGCTGCT) primers. This primers were used in combination with other primers containing deletion (table 1) and insertion (table 2) sequences. Original plasmid containing FL pFAS was digested with ApaI and MfeI (NEB) and later recombined with PCR amplified fragments carrying appropriate deletion or insertion sequences. DH-MAT linker constructs were cloned as mentioned above, using standard P2_DH_MAT_frw (TCTGCTGCTGAGTACT GACGAAG) and P2_DH_MAT_rev (CAGTGTGTACAGCTTCTGACGGT). Linker deletion and insertion primers are listed in the table 3 and 4. Original plasmid was digested using AatII and NheI (NEB).

Table 1. Primers sequence for the KR-ACP deletion constructs	
Primer name	**Primer sequence**
KR_ACP_DEL_1_AA_frw	AAGAAAGCGGCTCCTCGCGACGGAAGCT
KR_ACP_DEL_1_AA_rev	CGAGGAGCCGCTTTCTTCTCCGCCAG
KR_ACP_DEL_3_AA_frw	GAGAAGAAACCTCGCGACGGAAGCTCG
KR_ACP_DEL_3_AA_rev	GTCGCGAGGTTTCTTCTCCGCCAGAAC
KR_ACP_DEL_5_AA_frw	GCGGAGAAGCGCGACGGAAGCTCGCAGAA
KR_ACP_DEL_5_AA_rev	TTCCGTCGCGCTTCTCCGCCAGAACAAATG
KR_ACP_DEL_7_AA_frw	CTGGCGGAGGACGGAAGCTCGCAGAAAG
KR_ACP_DEL_7_AA_rev	CTTCCGTCCTCCGCCAGAACAAATGATG
KR_ACP_DEL_10_AA_frw	GTTCTGGCGAGCTCGCAGAAAGACCTGGTCA
KR_ACP_DEL_10_AA_rev	CTGCGAGCTCGCCAGAACAAATGATGAC
KR_ACP_DEL_14_AA_frw	TTTGTTCTGGACCTGGTCAAGGCGGTTGCC
KR_ACP_DEL_14_AA_rev	ACCGCCTTGACCAGGTCCAGAACAAATGATGACAGCACAGG

Table 2. Primers sequence for the KR-ACP insertion constructs	
Primer name	**Primer sequence**
KR_ACP_INS_1_AA_frw	GCGGAGAAGAAACTGGCGGCAGCTCCTCGCGACGGAAGCT
KR_ACP_INS_1_AA_rev	AGGAGCTGCCGCCAGTTTCTTCTCCGCCAGAACAAATG
KR_ACP_INS_3_AA_frw	TGGCGGAGAAGAAACTGGCTGAACTGGCTGAAGCGGCAGCTCCTC GCGACGGAA
KR_ACP_INS_3_AA_rev	TCGCGAGGAGCTGCCGCTTCAGCCAGTTCAGCCAGTTTCTTCTCCGC CAGAACAAA
KR_ACP_INS_5_AA_frw	GAAAGCGGCAGCTCTGGCTGAAAAGGCGCCTCGCGACGGAAGCT CGCAG
KR_ACP_INS_5_AA_rev	GTCGCGAGGCGCCTTTTCAGCCAGAGCTGCCGCTTTCTTCTCCGCC
KR_ACP_INS_7_AA_frw	GCGGAGAAGAAACTGGCTGAAAAGGCGCTGGCGGCGGCAGCTC CTCGCGACGGAAG
KR_ACP_INS_7_AA_rev	GCGAGGAGCTGCCGCCGCCAGCGCCTTTTCAGCCAGTTTCTTCTCCG CCAGAACAAA
KR_ACP_INS_10_AA_frw	CGGAGAAGAAACTGGCTGAAAAGGCGCTGGCGAAAGAGAAGGC GGCAGCTCCTCGCGACGGAAG
KR_ACP_INS_10_AA_rev	AGGAGCTGCCGCCTTCTCTTTCGCCAGCGCCTTTTCAGCCAGTTTCTTCT CCGCCAGAACAAAT
KR_ACP_INS_14_AA_frw	CGGAGAAGAAACTGGCTGAAAAGGCGCTGGCGAAAGAGAAGAGT GACTTTTCGGCGGCAGCTCCTCGCGACGGAAG
KR_ACP_INS_14_AA_rev	AGGAGCTGCCGCCGAAAAGTCACTCTTCTCTTTCGCCAGCGCCTTTTC AGCCAGTTTCTTCTCCGCCAGAACAAAT

Table 3. Primers sequence for MAT-DH deletion constructs	
Primer name	**Primer sequence**
MAT_DH_DEL_1_AA_frw	CGGTAGCTGTAGCTCTGTTGCGGTT
MAT_DH_DEL_1_AA_rev	AACCGCAACAGAGCTACAGCTACCGCTCGGGAAATCCG
MAT_DH_DEL_2_AA_frw	ATTTCCCGAGCGGTAGCAGCTCTGTTGCGGTTTATAAG
MAT_DH_DEL_2_AA_rev	ACCGCAACAGAGCTGCTACCGCTCGGGAAATCCGC
MAT_DH_DEL_3_AA_frw	CCCGAGCGGTAGCTCTGTTGCGGTTTATAAG
MAT_DH_DEL_3_AA_rev	ACCGCAACAGAGCTACCGCTCGGGAAATCCGCCGCGGAT
MAT_DH_DEL_4_AA_frw	GATTTCCCGAGCAGCTCTGTTGCGGTTTATAAG
MAT_DH_DEL_4_AA_rev	AACAGAGCTGCTCGGGAAATCCGCCGCGGA
MAT_DH_DEL_5_AA_frw	CGGCGGATTTCCCGAGCTCTGTTGCGGTTTATAAG
MAT_DH_DEL_5_AA_rev	ACCGCAACAGAGCTCGGGAAATCCGCCGCGGAT
MAT_DH_DEL_6_AA_frw	TCTGTTGCGGTTTATAAGTTTGATG
MAT_DH_DEL_6_AA_rev	CTTATAAACCGCAACAGACGGGAAATCCGCCGCGGATG

Table 3. Primers sequence for MAT-DH insertion constructs	
Primer name	**Primer sequence**
MAT_DH_INS_1_AA_frw	GCGGTAGCAGCAGTTGTAGCTCTGTTGCGG
MAT_DH_INS_1_AA_rev	CAACTGCTGCTACCGCTCGGGAAATCCGC
MAT_DH_INS_2_AA_frw	TAGCGGCAGCAGTTGTAGCTCTGTTGCGG
MAT_DH_INS_2_AA_rev	CTACAACTGCTGCCGCTACCGCTCGGGAAATCCG
MAT_DH_INS_3_AA_frw	GTAGCAGTGGCAGCAGTTGTAGCTCTGTTGCGG
MAT_DH_INS_3_AA_rev	ACTGCTGCCACTGCTACCGCTCGGGAAATCCG
MAT_DH_INS_6_AA_frw	TAGCGGAAGCGGTAGTGGCAGCAGTTGTAGCTCTGTTGCGG
MAT_DH_INS_6_AA_rev	GCTGCCACTACCGCTTCCGCTACCGCTCGGGAAATCCG
MAT_DH_INS_8_AA_frw	CGGTGGAAGCAGCGGTAGTGGCAGCAGTTGTAGCTCTGTTGCGG
MAT_DH_INS_8_AA_rev	CTGCTGCCACTACCGCTGCTTCCACCGCTACCGCTCGGGAAATCCG
MAT_DH_INS_10_AA_frw	AGCGGCGGTGGAAGCAGCGGTAGTGGCAGCAGTTGTAGCTCTGTTGCGG
MAT_DH_INS_10_AA_rev	GCTGCCACTACCGCTGCTTCCACCGCCGCTGCTACCGCTCGGGAAATCCG

Expression and purification of pFAS and linker mutant constructs.

pFAS constructs with a C-terminal 10 HIS tag were expressed in Spodoptera frugiperda (Sf21) cells using a MultiBac baculovirus vector. Cells were lysed in 25 mM HEPES, 150 mM NaCl and 20 mM imidazole at pH 7.5 using a Sonicator. The lysate was centrifuged for 40min at 35 000 rpm in a TI70 rotor at 4 °C and then loaded on a 5-ml HisTrap FF column (GE Healthcare) equilibrated in 25 mM HEPES, 150 mM NaCl and 20 mM imidazole, pH 7.5. The column was washed with 25 mM HEPES, 40 mM NaCl and 20 mM imidazole, pH 7.5 and then the FAS protein was eluted with a gradient up to 300 mM imidazole in 25 mM HEPES and 150 mM NaCl, pH 7.5. The fractions containing FAS were pulled and concentrated in a Milipore with a 100-kDa cutoff membrane; applied to a 120-ml Superdex 200 column equilibrated in 50 mM HEPES and 200 mM NaCl at pH 7.4 with 5 mM DTT and 10% glycerol. The FAS fractions were pooled and concentrated with a Millipore 100-kDa cutoff membrane to yield the purified pFAS.

Assay of overall pFAS activities with coenzyme A substrates.

The activity of the overall FAS reaction were measured continuously at 28C temperature in 96-well plates using the decrease in 340 nm absorbance due to NADPH oxidation. Kinetic values were determined in the presence of 40 μM Acetyl-CoA, 150 μM NADPH and varied concentrations of Malonyl-CoA (1-75 uM), for the overall reaction at 15 nM pFAS for wild type and linker constructs. The total reaction volume was 80μl containing 50 mM sodium phosphate, pH 7.0. To correct for non-enzymatic oxidation of NADPH, controls were conducted in the absence of pFAS. BioTek hybrid 1 reader was used to monitor absorbance changes.

Assay of the partial pFAS activity.

The activity of the KR domain was measured using trans-1-decalone at 28C in 96-well plates using the decrease in 340 nm absorbance. The total reaction volume 80 ul containing 50 mM sodium phosphate pH 7.5, 4nM wild type pFAS or linker constructs, 150 uM NADPH and varied concentrations of trans-1-decalone (0.01 to 8 mM from 1M stock solution diluted in DMSO). To correct for non-enzymatic oxidation of NADPH, controls were conducted in the absence of pFAS. BioTek hybrid 1 reader was used to monitor absorbance changes.

Data analysis.

Values for the kinetic parameters and their standard errors were obtained by fitting data to the appropriate equations using the nonlinear regression function of Prism (GraphPAD Software).

Results

Effect of KR-ACP linker extension on catalytic activity of pFAS.

Animal FAS has two active chambers. The distances between the individual active sites are up to 85 Å, which means that the ACP on its 14 residue flexible linker in combination with intra-domain flexibility of FAS is just enough to reach all necessary enzymes for condensation and B-processing reactions. The KR-ACP linker does not have notable conserved amino acid (aa) sequence and stable secondary structure elements. Interestingly the length of this linker is very well conserved across the species. Investigation of potential importance of the linker length for the substrate shuttling during the catalysis and overall FAS activity was initiated by making insertions of 3,5,7,10 and 14 aa in to the KR-ACP linker region. All 6 constructs demonstrated good expression (Figure 1E) and could be easily purified. First we tested activity of the single domain using trans-1-decalone as a model substrate that interacts directly with the active site of the KR domain and does not require involvement of the ACP domain. These results revealed that activity of the KR was not affected by increasing length of the linkers (Table 1). Next, we measured overall FAS activity of purified linker constructs, which are summarized in table 2. Increasing linker length up to 5 aa has little effect on FAS activity (Table 1). Interestingly further increasing length of the linker by 7,10 and 14 aa resulted in 12, 36 and 38% drop of the overall activity compared to the w/t pFAS.

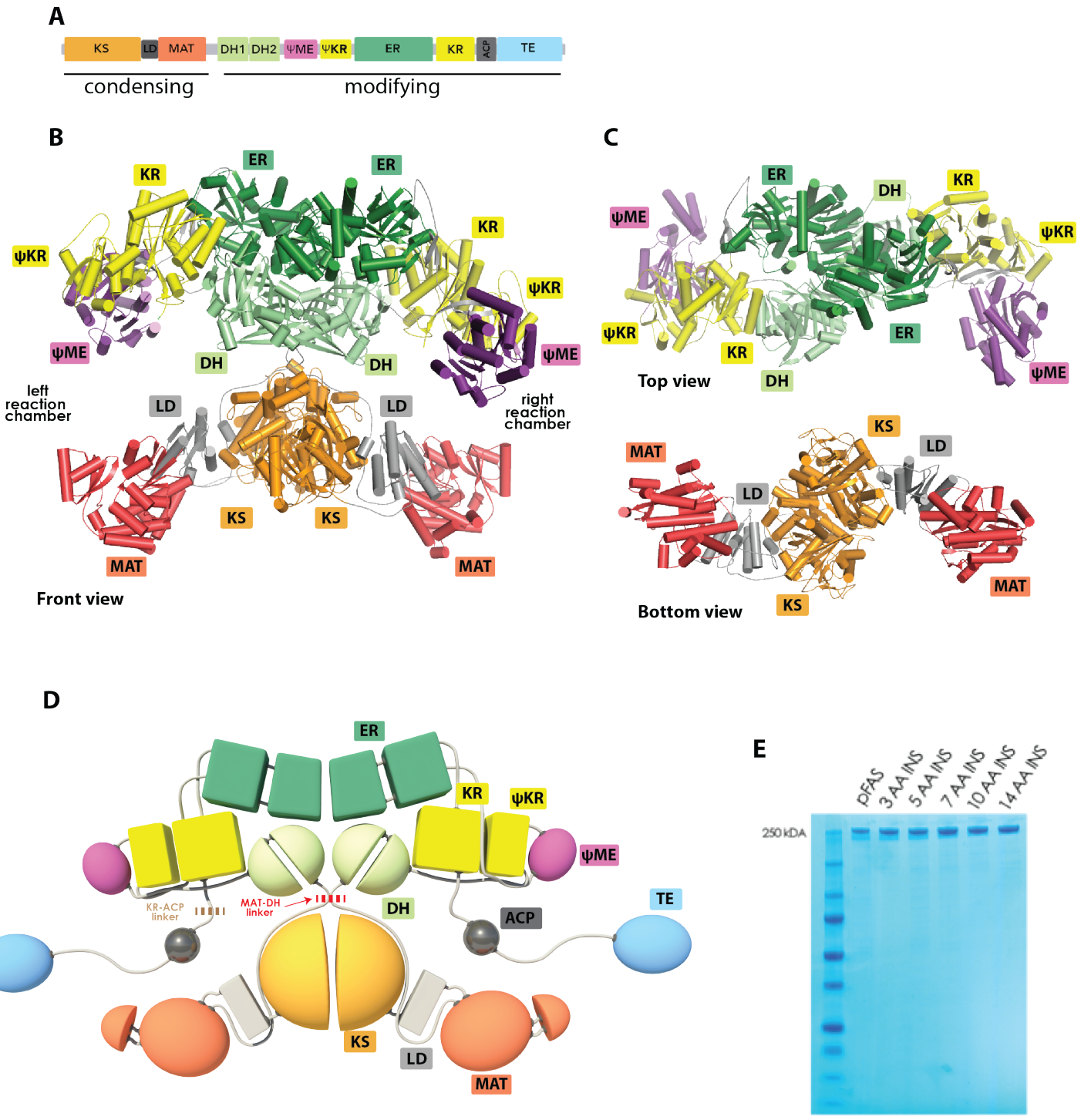


Figure 1. Structural overview of mamalian FAS. (**A**) Linear sequence organization of mFAS. (**B**) X-ray structure of mFAS colored by domain as indicated. mFAS forms an X-shaped homodimer with a lower condensing and an upper modifying part. (**C**) Top and bottom views, demonstrating the "S" shape of the modifying (upper) and condensing (lower) parts of mFAS. (**D**) Full lenth model of mFAS with ACP (dark grey) and TE (blue). Dashed red line indecent MAT-DH linker region, Dashed brown line represent KR-ACP linker boundary. (**E**) SDS gel of purified KR-ACP linker constructs.

Construct	KR(%)	FAS (%)
w/t pFAS	100 (+/- 1.13)	100(+/- 1.4)
KR-ACP 3 AA EXT	96.4(+/- 0.93)	94(+/- 2)
KR-ACP 5 AA EXT	105.2(+/- 0.89)	95(+/- 2.3)
KR-ACP 7 AA EXT	109.2(+/- 1.33)	88(+/- 1.94)
KR-ACP 10 AA EXT	99.3(+/- 2.06)	64(+/- 1.08)
KR-ACP 14 AA EXT	112.4(+/- 0.77)	62(+/- 0.9)

Table 1. The K_{cat} activity of the wt FAS (100% value) was 120 +/- 1.9 n**M** of NADPH oxidized nM^{-1} •min^{-1} and 606 +/- 6.9 nM^{-1} •min^{-1} of trans-1-decalone hydrolyzed in the ketoreductase assay.

Effect of MAT-DH modifications on catalytic activity of pFAS (ongoing).

This linker region 854 to 858 connects B-carbon processing upper part and condensing bottom part (Figure 1D). Due to the extreme flexibility of this "hinge" region FAS can make to distinct motions: swinging of the MAT-KS2-MAT module from side to side, and swiveling motion perpendicular to the upper portion of the structure. In order to better evaluate importance of the six aa linker region in this area, we cloned twelve constructs with systematically increasing and decreasing linker length (table 2). Currently all constructs are in the stage of virus production.

MAT- DH linker insertion constructs	MAT- DH linker deletion constructs
1 aa insertion	1 aa deletion
2 aa insertion	2 aa deletion
3 aa insertion	3 aa deletion
6 aa insertion	4 aa deletion
8 aa insertion	5 aa deletion
10 aa insertion	6 aa deletion

Table 2. Summary of the all cloned linker constructs for the MAT-DH region.

Conclusion

The catalytic properties of multienzymes are not only determined by its enzymatic domains, but to a large extend also by the specifics of how these domains are coupled. In the animals FAS individual catalytic domains are connected via small linkers. In order to understand how linking influences catalysis and facilitates efficient communication between adjusted domains we systematically varied linker regions and correlated linker properties to the resulting enzymatic properties. Our experiments revealed that the FAS can tolerate substantial changes in length of the KR-ACP linker without influencing serious effects on the activity; more than 10 aa had to be inserted to the linker region before any significant outcome on functioning of the FAS was observed, which are very distant relative of modern animal FAS, contain long KR ACP linkers (up to 21 aa). Further experiment needed to be conducted in order to understand exact reason why longer linkers (more then 10 aa) lowered activity only by 40% and what are effects of reducing length of KR-ACP linker on the catalytic properties of FAS. Furthermore we are currently investigating one more linker region, which connects the ß-carbon processing upper part and the condensing bottom part. The information which we will gain from this experiments will help us better understand complex dynamics of FAS, moreover it will enable us as to construct guidelines for individual linker regions which can be used in future for building artificial multienzymes incorporating unique enzymatic domains.

References:

1 White, S. W., Zheng, J., Zhang, Y. M. & Rock. The structural biology of type II fatty acid biosynthesis. *Annual review of biochemistry* **74**, 791-831, doi:10.1146/annurev.biochem.74.082803.133524 (2005).

2 Maier, T., Leibundgut, M., Boehringer, D. & Ban, N. Structure and function of eukaryotic fatty acid synthases. *Q Rev Biophys* **43**, 373-422, doi:10.1017/S0033583510000156 (2010).

3 Jenni, S. *et al.* Structure of fungal fatty acid synthase and implications for iterative substrate shuttling. *Science* **316**, 254-261, doi:10.1126/science.1138248 (2007).

4 Jenni, S., Leibundgut, M., Maier, T. & Ban, N. Architecture of a fungal fatty acid synthase at 5 A resolution. *Science* **311**, 1263-1267, doi:10.1126/science.1123251 (2006).

5 Maier, T., Jenni, S. & Ban, N. Architecture of mammalian fatty acid synthase at 4.5 A resolution. *Science* **311**, 1258-1262, doi:10.1126/science.1123248 (2006).

6 Maier, T., Leibundgut, M. & Ban, N. The crystal structure of a mammalian fatty acid synthase. *Science* **321**, 1315-1322, doi:10.1126/science.1161269 (2008).

7 Smith, S., Witkowski, A. & Joshi, A. K. Structural and functional organization of the animal fatty acid synthase. *Progress in lipid research* **42**, 289-317 (2003).

8 Smith, S. & Tsai, S. C. The type I fatty acid and polyketide synthases: a tale of two megasynthases. *Natural product reports* **24**, 1041-1072, doi:10.1039/b603600g (2007).

9 Witkowski, A. *et al.* Head-to-head coiled arrangement of the subunits of the animal fatty acid synthase. *Chemistry & biology* **11**, 1667-1676, doi:10.1016/j.chembiol.2004.09.016 (2004).

10 Brignole, E. J., Smith, S. & Asturias, F. J. Conformational flexibility of metazoan fatty acid synthase enables catalysis. *Nature structural & molecular biology* **16**, 190-197, doi:10.1038/nsmb.1532 (2009).

11 Perham, R. N. Swinging arms and swinging domains in multifunctional enzymes: catalytic machines for multistep reactions. *Annual review of biochemistry* **69**, 961-1004, doi:10.1146/annurev.biochem.69.1.961 (2000).

12 Gokhale, R. S., Hunziker, D., Cane, D. E. & Khosla, C. Mechanism and specificity of the terminal thioesterase domain from the erythromycin polyketide synthase. *Chemistry & biology* **6**, 117-125, doi:10.1016/S1074-5521(99)80008-8 (1999).

Summary and outlook

Multienzyme complexes confer a kinetic advantage by shuttling reaction intermediates between consecutive enzymes and reducing the diffusion time. The emergent properties of multienzymes are not only determined by composition of enzymatic domains, but to a large extend by how these domains are linked. Multienzymes can utilize two types of linkage methods: one being flexible linkers connecting the individual domains, whereas the other consists of rigid scaffolding elements, which fixes the domains in one position[50,72,110]. Both types of linkage offer advantages such as prevention of side reactions, enhanced local concentration of active sites and the coordinated regulation of expression[57]. The only example of such multienzymes that utilized both type of linkage is eukaryotic fatty acid synthase (FAS)[16,17,48,57,58]. Two strikingly different types of FAS have emerged in fungi and in animals[57]. The fungal FAS is a rigid, 2.6-MDa barrel-shaped structure with 50% of its polypeptide chain providing supportive scaffolding elements[16,58]. Inside the barrel, the concentration of all active sites is approximately 1 mM, which ensures high catalytic efficiency[58]. In the fungal FAS 48 functional domains are embedded into a scaffolding matrix that mediates the majority of architectural interactions determining the spatial arrangement of catalytic centers[71]. The sequential and structural homology of fFAS is distantly shared with the more recently described CMN-FAS systems in *Corynebacteria*, *Mycobacteria*, and *Nocardia* that have a slightly lower number of scaffolding expansions[112]. Using combined phylogenetic and structural biology approach we have identified two freestanding bacterial proteins FabY and DfnA-ER as potential ancestors of scaffolding regions in fFAS. The entire FabY with its insertion regions closely resembles the KS domains of fungal FAS and particularly of CMN-FAS. The conserved extension of the fFAS KS plays a crucial role in the organization of the

central wheel structure and provides binding sites for cooperating enzymes.

Contrastingly, the animal FAS is an open X-shaped structure with catalytic domains not interrupted by the insertion of scaffolding elements but connected to each other via short non-conserved linker sequences[48].

Many large multienzymes like biotin-dependent carboxylases, polyketide synthase type I (PKS I) and pyruvate dehydrogenase share very similar construction logic to animal FAS; individual enzymatic domains with unique chemical functions are connected to each other via non conserved linkers and each of this system utilizes small carrier proteins, which translocate substrates from one enzymatic domain to another[110]. In biotin-dependent carboxylases, the biotin-carboxyl carrier protein, which is responsible for shuttling the substrate intermediate during catalysis, is connected via linkers to the rest of the protein[113]. The linker sequence is not conserved and very rich in alanine and serine residues; as a result it provides important conformational flexibility to achieve high catalytic efficiency[113]. Interestingly similar linkers are found in the dihydrolipoyl transacetylase polypeptide chains of pyruvate dehydrogenase, where they provide conformational flexibility and at the same time avoid collapsed conformations[110]. Crystallographic data together with biochemical and EM analysis indicate that animal FAS displays an extraordinary degree of flexibility to ensure productive interactions between the ACP and the active sites during the reaction cycle[17,48,51]. Conformational changes most likely result from a combination of internal domain elasticity and flexibility in the linker regions, which connects individual domains in the animal FAS. In the present work we investigated how intra domain linking influences catalytic properties and conformational crosstalk between domains. This was achieved by generating more than 40 different constructs with systematically

increasing or decreasing linker lengths between KR-ACP and MAT-DH linkers. Initial kinetic data indicate that animal FAS can tolerate substantial changes in length of the KR-ACP linker without having serious effects on the activity. Our next goal is to kinetically characterize remaining KR-ACP and MAT-DH linkers constructs, which will enable us to better understand complex dynamic of the animal FAS.

We have used a combined approach of phylogenetic, structural and biochemical analysis to functionally dissect two distinct fatty acid machinaries. The information, which we gained, has a impact on the field of synthetic biology and therapeutics in the following ways:

1) Our structural and evolutionary analysis of the fungal FAS offers new insights for building artificial enzymes based on rigid scaffolds. We observed in FabY and DfnA that scaffolding elements are inserted outside of the core regions and provide critical surface for assembling the fungal FAS barrel. These scaffolding elements could be utilized for integration of different chemical enzymes to build highly processive multienzymes catalyzing unique chemical reactions in a coordinated manner. The use of so-called protein scaffolds for the generation of novel binding proteins via combinatorial engineering has recently emerged as a powerful alternative in the field of protein engineering and are used in for numerous applications in synthetic biology, diagnostics, therapeutics and imaging[114,115].

2) Polyketides are a structurally diverse but biosynthetically related family of natural products that display clinically relevant biological activities, including anticancer (e.g. calicheamicin and

bleomycin), immunosuppression (e.g. rapamycin), and antibacterial (e.g. erythromycin and vancomycin)[115-118]. Structural and evolutionarily analysis revealed that animal FAS is distantly related to PKS I[45]. The animal FASs polypeptides contain a single copy of all seven functional domains, whereas the type I PKSs consists of multiple FAS-like modules[48]. Each module contains specific combinations of catalytic domains that catalyze unique biochemical steps of chain elongation and processing. ACP domains in each model are not only responsible for shuffling reaction intermediates inside of the single module but also provide the functional link between modules as they transfer end product to the module immediately downstream, where the next chain of chemical reactions are conducted with the cooperation of the resident ACP domain of that module[50]. To generate analogs, traditionally, engineering has focused on substituting individual enzymatic domains or entire modules with ones of different building block specificity or deleting various enzymatic functions while disregarding the linkage region[119]. But unfortunately these attempts resulted in extremely poor yields or completely failed[120]. We showed in the structurally related animal FAS that intra-domain linking regions play a crucial role in the dynamics of this enzyme. As a result of these observations we aim to construct guidelines for individual linker regions that can be used for connecting multiple enzymes and maintaining high catalytic efficiency. Such guidelines could be used in re-engineering PKS with unique chemical functions. Moreover, recently, linking protein domains into novel, artificial polyproteins has resulted in new classes of high-affinity binder molecules as potential protein therapeutics and

accelerated elucidation of mechanisms governing protein folding by single-molecule techniques[121,122].

3) FAS is highly expressed in a number of cancers, with low expression observed in most normal tissues. Although normal tissues tend to obtain fatty acids though diet, tumor tissues rely on *de novo* fatty acid synthesis, making FAS an attractive target for the treatment of cancer[122]. The medical use of FAS inhibitors has been hampered by off-target activities[123]. Our mechanistic understanding of dynamics offers new ways to design inhibitors that relay on interfering with FAS dynamics e.g. trapping linkers. One of such strategy is to use scaffolds based inhibitor as a promising alternative to small molecules for therapeutics application[124]. Recently, such a scaffolding based blocker has been used in mechanically inhibiting aberrant MAPK regulation that can contribute to cancer and other human diseases[124].

Multienzyme technologies have recently gained prominence as particularly useful tools for synthetic biology[121,122]. Our structural and functional dissection of FAS, one of the most complex biosynthetic machineries, provides important insights into the complex dynamics of large carrier based proteins. Moreover it provides excellent tools for building novel multienzyme-based nanomachines for synthetizing unique chemical compounds.

Acknowledgements

I would like to thank Prof. Dr. Timm Maier, without his help advice, expertise and encouragements this research and dissertation would not have happened. I would also like to thank department of structural biology for giving me the opportunity to write an honors thesis. To my committee, Prof. Dr. Roderick Lim and Prof. Dr. Torsten Schwede, I am extremely grateful for your assistance and suggestions throughout my project.

I would like to thank each member of my group: Roman, Stefan, Moritz, Dominik, Anna, Fabian, Friederike, Edward, Evelyn and Rita for being good friends and creating excellent working environment.

I am also grateful to all, who directly or indirectly, were present with me, for fun or for work, during my PhD time.

References:

1 Osborne, T. B. Growth on diets poor in true fats. *The journal of biological chemistry* **45**, 145-152 (1920).

2 Burr, G. O. & Burr, M. M. Nutrition classics from The Journal of Biological Chemistry 82:345-67, 1929. A new deficiency disease produced by the rigid exclusion of fat from the diet. *Nutrition reviews* **31**, 248-249 (1973).

3 Burr, G. O. & Burr, M. M. On the nature and role of the fatty acids essential in nutrition. *J Biol Chem* **86**, 0587-0621 (1930).

4 Burr, G. O. The essential fatty acids fifty years ago. *Progress in lipid research* **20**, xxvii-xxix (1981).

5 Anderson, R. A., Boronenkov, I. V., Doughman, S. D., Kunz, J. & Loijens, J. C. Phosphatidylinositol phosphate kinases, a multifaceted family of signaling enzymes. *J Biol Chem* **274**, 9907-9910 (1999).

6 Tvrzicka, E., Kremmyda, L. S., Stankova, B. & Zak, A. Fatty acids as biocompounds: their role in human metabolism, health and disease--a review. Part 1: classification, dietary sources and biological functions. *Biomedical papers of the Medical Faculty of the University Palacky, Olomouc, Czechoslovakia* **155**, 117-130 (2011).

7 Kaur, N., Chugh, V. & Gupta, A. K. Essential fatty acids as functional components of foods- a review. *Journal of food science and technology* **51**, 2289-2303, doi:10.1007/s13197-012-0677-0 (2014).

8 Sohlenkamp, C. & Geiger, O. Bacterial membrane lipids: diversity in structures and pathways. *FEMS microbiology reviews*, doi:10.1093/femsre/fuv008 (2015).

9 Bogdanov, M., Dowhan, W. & Vitrac, H. Lipids and topological rules governing membrane protein assembly. *Biochimica et biophysica acta* **1843**, 1475-1488, doi:10.1016/j.bbamcr.2013.12.007 (2014).

10 De Castro, C., Holst, O., Lanzetta, R., Parrilli, M. & Molinaro, A. Bacterial lipopolysaccharides in plant and mammalian innate immunity. *Protein and peptide letters* **19**, 1040-1044 (2012).

11 Eyster, K. M. The membrane and lipids as integral participants in signal transduction: lipid signal transduction for the non-lipid biochemist. *Advances in physiology education* **31**, 5-16, doi:10.1152/advan.00088.2006 (2007).

12 Simons, K. & Sampaio, J. L. Membrane organization and lipid rafts. *Cold Spring Harbor perspectives in biology* **3**, a004697, doi:10.1101/cshperspect.a004697 (2011).

13 Watt, M. J. & Steinberg, G. R. Regulation and function of triacylglycerol lipases in cellular metabolism. *The Biochemical journal* **414**, 313-325, doi:10.1042/BJ20080305 (2008).
14 Gross, D. A. & Silver, D. L. Cytosolic lipid droplets: from mechanisms of fat storage to disease. *Critical reviews in biochemistry and molecular biology* **49**, 304-326, doi:10.3109/10409238.2014.931337 (2014).
15 Yamamoto, Y., Imori, Y. & Hara, S. Oxidation behavior of triacylglycerol containing conjugated linolenic acids in sn-1(3) or sn-2 position. *Journal of oleo science* **63**, 31-37 (2014).
16 Jenni, S. *et al.* Structure of fungal fatty acid synthase and implications for iterative substrate shuttling. *Science* **316**, 254-261, doi:10.1126/science.1138248 (2007).
17 Maier, T., Jenni, S. & Ban, N. Architecture of mammalian fatty acid synthase at 4.5 A resolution. *Science* **311**, 1258-1262, doi:10.1126/science.1123248 (2006).
18 White, S. W., Zheng, J., Zhang, Y. M. & Rock. The structural biology of type II fatty acid biosynthesis. *Annual review of biochemistry* **74**, 791-831, doi:10.1146/annurev.biochem.74.082803.133524 (2005).
19 Majerus, P. W. Acyl carrier protein: effects of acetylation and tryptic hydrolysis on function in fatty acid synthesis. *Science* **159**, 428-430 (1968).
20 Rawlings, M. & Cronan, J. E., Jr. The gene encoding Escherichia coli acyl carrier protein lies within a cluster of fatty acid biosynthetic genes. *J Biol Chem* **267**, 5751-5754 (1992).
21 Chirgadze, N. Y., Briggs, S. L., McAllister, K. A., Fischl, A. S. & Zhao, G. Crystal structure of Streptococcus pneumoniae acyl carrier protein synthase: an essential enzyme in bacterial fatty acid biosynthesis. *The EMBO journal* **19**, 5281-5287, doi:10.1093/emboj/19.20.5281 (2000).
22 Cronan, J. E., Jr. & Waldrop, G. L. Multi-subunit acetyl-CoA carboxylases. *Progress in lipid research* **41**, 407-435 (2002).
23 Gajiwala, K. S. *et al.* Crystal structures of bacterial FabH suggest a molecular basis for the substrate specificity of the enzyme. *FEBS letters* **583**, 2939-2946, doi:10.1016/j.febslet.2009.08.001 (2009).
24 Musayev, F., Sachdeva, S., Scarsdale, J. N., Reynolds, K. A. & Wright, H. T. Crystal structure of a substrate complex of Mycobacterium tuberculosis beta-ketoacyl-acyl carrier protein synthase III (FabH) with lauroyl-coenzyme A. *Journal of molecular biology* **346**, 1313-1321, doi:10.1016/j.jmb.2004.12.044 (2005).
25 Zhang, Y. M., Wu, B., Zheng, J. & Rock, C. O. Key residues responsible for acyl carrier protein and beta-ketoacyl-acyl carrier protein reductase (FabG) interaction. *J Biol Chem* **278**, 52935-52943, doi:10.1074/jbc.M309874200 (2003).

26 Heath, R. J. & Rock, C. O. Roles of the FabA and FabZ beta-hydroxyacyl-acyl carrier protein dehydratases in Escherichia coli fatty acid biosynthesis. *J Biol Chem* **271**, 27795-27801 (1996).
27 Heath, R. J., Yu, Y. T., Shapiro, M. A., Olson, E. & Rock, C. O. Broad spectrum antimicrobial biocides target the FabI component of fatty acid synthesis. *J Biol Chem* **273**, 30316-30320 (1998).
28 Saito, J. *et al.* Crystal structure of enoyl-acyl carrier protein reductase (FabK) from Streptococcus pneumoniae reveals the binding mode of an inhibitor. *Protein science : a publication of the Protein Society* **17**, 691-699, doi:10.1110/ps.073288808 (2008).
29 Heath, R. J., White, S. W. & Rock, C. O. Inhibitors of fatty acid synthesis as antimicrobial chemotherapeutics. *Appl Microbiol Biotechnol* **58**, 695-703, doi:10.1007/s00253-001-0918-z (2002).
30 Lu, Y. J. *et al.* Acyl-phosphates initiate membrane phospholipid synthesis in Gram-positive pathogens. *Molecular cell* **23**, 765-772, doi:10.1016/j.molcel.2006.06.030 (2006).
31 Coleman, J. Characterization of the Escherichia coli gene for 1-acyl-sn-glycerol-3-phosphate acyltransferase (plsC). *Molecular & general genetics : MGG* **232**, 295-303 (1992).
32 Davis, M. S. & Cronan, J. E., Jr. Inhibition of Escherichia coli acetyl coenzyme A carboxylase by acyl-acyl carrier protein. *Journal of bacteriology* **183**, 1499-1503, doi:10.1128/JB.183.4.1499-1503.2001 (2001).
33 Heath, R. J. & Rock, C. O. Inhibition of beta-ketoacyl-acyl carrier protein synthase III (FabH) by acyl-acyl carrier protein in Escherichia coli. *J Biol Chem* **271**, 10996-11000 (1996).
34 Heath, R. J. & Rock, C. O. Regulation of fatty acid elongation and initiation by acyl-acyl carrier protein in Escherichia coli. *J Biol Chem* **271**, 1833-1836 (1996).
35 Yasuno, R., von Wettstein-Knowles, P. & Wada, H. Identification and molecular characterization of the beta-ketoacyl-[acyl carrier protein] synthase component of the Arabidopsis mitochondrial fatty acid synthase. *J Biol Chem* **279**, 8242-8251, doi:10.1074/jbc.M308894200 (2004).
36 Hiltunen, J. K., Chen, Z., Haapalainen, A. M., Wierenga, R. K. & Kastaniotis, A. J. Mitochondrial fatty acid synthesis--an adopted set of enzymes making a pathway of major importance for the cellular metabolism. *Progress in lipid research* **49**, 27-45, doi:10.1016/j.plipres.2009.08.001 (2010).
37 Hiltunen, J. K. *et al.* Mitochondrial fatty acid synthesis and respiration. *Biochimica et biophysica acta* **1797**, 1195-1202, doi:10.1016/j.bbabio.2010.03.006 (2010).
38 Yoshida, I., Sweetman, L., Kulovich, S., Nyhan, W. L. & Robinson, B. H. Effect of lipoic acid in a patient with defective activity of pyruvate

dehydrogenase, 2-oxoglutarate dehydrogenase, and branched-chain keto acid dehydrogenase. *Pediatric research* **27**, 75-79, doi:10.1203/00006450-199001000-00020 (1990).
39 Seyda, A. *et al.* A novel syndrome affecting multiple mitochondrial functions, located by microcell-mediated transfer to chromosome 2p14-2p13. *American journal of human genetics* **68**, 386-396, doi:10.1086/318196 (2001).
40 Feng, D., Witkowski, A. & Smith, S. Down-regulation of mitochondrial acyl carrier protein in mammalian cells compromises protein lipoylation and respiratory complex I and results in cell death. *J Biol Chem* **284**, 11436-11445, doi:10.1074/jbc.M806991200 (2009).
41 Chen, Z. *et al.* Myocardial overexpression of Mecr, a gene of mitochondrial FAS II leads to cardiac dysfunction in mouse. *PloS one* **4**, e5589, doi:10.1371/journal.pone.0005589 (2009).
42 Christensen, C. E., Kragelund, B. B., von Wettstein-Knowles, P. & Henriksen, A. Structure of the human beta-ketoacyl [ACP] synthase from the mitochondrial type II fatty acid synthase. *Protein science : a publication of the Protein Society* **16**, 261-272, doi:10.1110/ps.062473707 (2007).
43 Tsay, J. T., Oh, W., Larson, T. J., Jackowski, S. & Rock, C. O. Isolation and characterization of the beta-ketoacyl-acyl carrier protein synthase III gene (fabH) from Escherichia coli K-12. *J Biol Chem* **267**, 6807-6814 (1992).
44 Leibundgut, M., Maier, T., Jenni, S. & Ban, N. The multienzyme architecture of eukaryotic fatty acid synthases. *Current opinion in structural biology* **18**, 714-725, doi:10.1016/j.sbi.2008.09.008 (2008).
45 Jenke-Kodama, H., Sandmann, A., Muller, R. & Dittmann, E. Evolutionary implications of bacterial polyketide synthases. *Mol Biol Evol* **22**, 2027-2039, doi:10.1093/molbev/msi193 (2005).
46 Smith, S., Witkowski, A. & Joshi, A. K. Structural and functional organization of the animal fatty acid synthase. *Progress in lipid research* **42**, 289-317 (2003).
47 Asturias, F. J. *et al.* Structure and molecular organization of mammalian fatty acid synthase. *Nat Struct Mol Biol* **12**, 225-232, doi:nsmb899 [pii]
10.1038/nsmb899 (2005).
48 Maier, T., Leibundgut, M. & Ban, N. The crystal structure of a mammalian fatty acid synthase. *Science* **321**, 1315-1322, doi:10.1126/science.1161269 (2008).
49 Serre, L., Verbree, E. C., Dauter, Z., Stuitje, A. R. & Derewenda, Z. S. The Escherichia coli malonyl-CoA:acyl carrier protein transacylase at 1.5-A resolution. Crystal structure of a fatty acid synthase component. *J Biol Chem* **270**, 12961-12964 (1995).

50 Smith, S. & Tsai, S. C. The type I fatty acid and polyketide synthases: a tale of two megasynthases. *Natural product reports* **24**, 1041-1072, doi:10.1039/b603600g (2007).
51 Brignole, E. J., Smith, S. & Asturias, F. J. Conformational flexibility of metazoan fatty acid synthase enables catalysis. *Nature structural & molecular biology* **16**, 190-197, doi:10.1038/nsmb.1532 (2009).
52 Castillo, Y. P. & Perez, M. A. Bacterial beta-ketoacyl-acyl carrier protein synthase III (FabH): an attractive target for the design of new broad-spectrum antimicrobial agents. *Mini Rev Med Chem* **8**, 36-45 (2008).
53 Soulie, J. M., Sheplock, G. J., Tian, W. X. & Hsu, R. Y. Transient kinetic studies of fatty acid synthetase. A kinetic self-editing mechanism for the loading of acetyl and malonyl residues and the role of coenzyme A. *J Biol Chem* **259**, 134-140 (1984).
54 Stern, A., Sedgwick, B. & Smith, S. The free coenzyme A requirement of animal fatty acid synthetase. Participation in the continuous exchange of acetyl and malonyl moieties between coenzyme a thioester and enzyme. *J Biol Chem* **257**, 799-803 (1982).
55 Witkowski, A., Joshi, A. K. & Smith, S. Characterization of the interthiol acyltransferase reaction catalyzed by the beta-ketoacyl synthase domain of the animal fatty acid synthase. *Biochemistry* **36**, 16338-16344, doi:10.1021/bi972242q (1997).
56 Libertini, L. J. & Smith, S. Synthesis of long chain acyl-enzyme thioesters by modified fatty acid synthetases and their hydrolysis by a mammary gland thioesterase. *Archives of biochemistry and biophysics* **192**, 47-60 (1979).
57 Maier, T., Leibundgut, M., Boehringer, D. & Ban, N. Structure and function of eukaryotic fatty acid synthases. *Q Rev Biophys* **43**, 373-422, doi:10.1017/S0033583510000156 (2010).
58 Jenni, S., Leibundgut, M., Maier, T. & Ban, N. Architecture of a fungal fatty acid synthase at 5 A resolution. *Science* **311**, 1263-1267, doi:10.1126/science.1123251 (2006).
59 Johansson, P. *et al.* Inhibition of the fungal fatty acid synthase type I multienzyme complex. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 12803-12808, doi:10.1073/pnas.0805827105 (2008).
60 Lomakin, I. B., Xiong, Y. & Steitz, T. A. The crystal structure of yeast fatty acid synthase, a cellular machine with eight active sites working together. *Cell* **129**, 319-332, doi:10.1016/j.cell.2007.03.013 (2007).
61 Leibundgut, M., Jenni, S., Frick, C. & Ban, N. Structural basis for substrate delivery by acyl carrier protein in the yeast fatty acid synthase. *Science* **316**, 288-290, doi:10.1126/science.1138249 (2007).

62 Grininger, M. Perspectives on the evolution, assembly and conformational dynamics of fatty acid synthase type I (FAS I) systems. *Current opinion in structural biology* **25C**, 49-56, doi:10.1016/j.sbi.2013.12.004 (2014).
63 Elovson, J. & Vagelos, P. R. Acyl carrier protein. X. Acyl carrier protein synthetase. *J Biol Chem* **243**, 3603-3611 (1968).
64 Lambalot, R. H. & Walsh, C. T. Cloning, overproduction, and characterization of the Escherichia coli holo-acyl carrier protein synthase. *J Biol Chem* **270**, 24658-24661 (1995).
65 Fichtlscherer, F., Wellein, C., Mittag, M. & Schweizer, E. A novel function of yeast fatty acid synthase. Subunit alpha is capable of self-pantetheinylation. *European journal of biochemistry / FEBS* **267**, 2666-2671 (2000).
66 Bunkoczi, G. *et al.* Mechanism and substrate recognition of human holo ACP synthase. *Chemistry & biology* **14**, 1243-1253, doi:10.1016/j.chembiol.2007.10.013 (2007).
67 Johansson, P. *et al.* Multimeric options for the auto-activation of the Saccharomyces cerevisiae FAS type I megasynthase. *Structure* **17**, 1063-1074, doi:10.1016/j.str.2009.06.014 (2009).
68 Engeser, H., Hubner, K., Straub, J. & Lynen, F. Identity of malonyl and palmitoyl transferase of fatty acid synthetase from yeast. 2. A comparison of active-site peptides. *European journal of biochemistry / FEBS* **101**, 413-422 (1979).
69 Pirson, W., Schuhmann, L. & Lynen, F. The specificity of yeast fatty-acid synthetase with respect to the "priming" substrate. Decanoyl-coA and derivatives as "primers" of fatty-acid synthesis in vitro. *European journal of biochemistry / FEBS* **36**, 16-24 (1973).
70 Schweizer, E., Kniep, B., Castorph, H. & Holzner, U. Pantetheine-free mutants of the yeast fatty-acid-synthetase complex. *European journal of biochemistry / FEBS* **39**, 353-362 (1973).
71 Grininger, M. Perspectives on the evolution, assembly and conformational dynamics of fatty acid synthase type I (FAS I) systems. *Current opinion in structural biology* **25**, 49-56, doi:10.1016/j.sbi.2013.12.004 (2014).
72 Smith, J. L. & Sherman, D. H. Biochemistry. An enzyme assembly line. *Science* **321**, 1304-1305, doi:10.1126/science.1163785 (2008).
73 Keatinge-Clay, A. T. The structures of type I polyketide synthases. *Natural product reports* **29**, 1050-1073, doi:10.1039/c2np20019h (2012).
74 Schweizer, E. & Hofmann, J. Microbial type I fatty acid synthases (FAS): major players in a network of cellular FAS systems. *Microbiol Mol Biol Rev* **68**, 501-517, table of contents, doi:10.1128/MMBR.68.3.501-517.2004 (2004).

75 Campbell, J. W. & Cronan, J. E., Jr. Bacterial fatty acid biosynthesis: targets for antibacterial drug discovery. *Annu Rev Microbiol* **55**, 305-332, doi:10.1146/annurev.micro.55.1.305 (2001).
76 Rock, C. O. & Jackowski, S. Forty years of bacterial fatty acid synthesis. *Biochem Biophys Res Commun* **292**, 1155-1166, doi:10.1006/bbrc.2001.2022 (2002).
77 Staunton, J. & Weissman, K. J. Polyketide biosynthesis: a millennium review. *Natural product reports* **18**, 380-416 (2001).
78 Gago, G., Diacovich, L., Arabolaza, A., Tsai, S. C. & Gramajo, H. Fatty acid biosynthesis in actinomycetes. *FEMS microbiology reviews* **35**, 475-497, doi:10.1111/j.1574-6976.2010.00259.x (2011).
79 Boehringer, D., Ban, N. & Leibundgut, M. 7.5-A cryo-em structure of the mycobacterial fatty acid synthase. *Journal of molecular biology* **425**, 841-849, doi:10.1016/j.jmb.2012.12.021 (2013).
80 Ciccarelli, L. *et al.* Structure and conformational variability of the mycobacterium tuberculosis fatty acid synthase multienzyme complex. *Structure* **21**, 1251-1257, doi:10.1016/j.str.2013.04.023 (2013).
81 Ponting, C. P. & Russell, R. R. The natural history of protein domains. *Annu Rev Biophys Biomol Struct* **31**, 45-71, doi:10.1146/annurev.biophys.31.082901.134314 (2002).
82 Chung, S. Y. & Subbiah, S. A structural explanation for the twilight zone of protein sequence homology. *Structure* **4**, 1123-1127 (1996).
83 Chen, X. H. *et al.* Comparative analysis of the complete genome sequence of the plant growth-promoting bacterium Bacillus amyloliquefaciens FZB42. *Nat Biotechnol* **25**, 1007-1014, doi:10.1038/nbt1325 (2007).
84 Chen, X. H. *et al.* Structural and functional characterization of three polyketide synthase gene clusters in Bacillus amyloliquefaciens FZB 42. *Journal of bacteriology* **188**, 4024-4036, doi:10.1128/JB.00052-06 (2006).
85 Savitsky, P. *et al.* High-throughput production of human proteins for crystallization: the SGC experience. *J Struct Biol* **172**, 3-13, doi:10.1016/j.jsb.2010.06.008 (2010).
86 Betancor, L., Fernandez, M. J., Weissman, K. J. & Leadlay, P. F. Improved catalytic activity of a purified multienzyme from a modular polyketide synthase after coexpression with Streptomyces chaperonins in Escherichia coli. *Chembiochem* **9**, 2962-2966, doi:10.1002/cbic.200800475 (2008).
87 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* **14**, 993-1000 (2001).

88 Kabsch, W. Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr D Biol Crystallogr* **66**, 133-144, doi:10.1107/S0907444909047374 (2010).
89 Kabsch, W. Xds. *Acta Crystallogr D Biol Crystallogr* **66**, 125-132, doi:10.1107/S0907444909047337 (2010).
90 Huang, W. *et al.* Crystal structure of beta-ketoacyl-acyl carrier protein synthase II from E.coli reveals the molecular architecture of condensing enzymes. *The EMBO journal* **17**, 1183-1191, doi:10.1093/emboj/17.5.1183 (1998).
91 Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-2132, doi:10.1107/S0907444904019158 (2004).
92 Adams, P. D. *et al.* PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr D Biol Crystallogr* **58**, 1948-1954 (2002).
93 Blanc, E. *et al.* Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT. *Acta Crystallogr D Biol Crystallogr* **60**, 2210-2221, doi:10.1107/S0907444904016427 (2004).
94 Massengo-Tiasse, R. P. & Cronan, J. E. Diversity in enoyl-acyl carrier protein reductases. *Cellular and molecular life sciences : CMLS* **66**, 1507-1517, doi:10.1007/s00018-009-8704-7 (2009).
95 Gemperlein, K., Rachid, S., Garcia, R. O., Wenzel, S. C. & Muller, R. Polyunsaturated fatty acid biosynthesis in myxobacteria: different PUFA synthases and their product diversity. *Chemical Science* **5**, 1733-1741, doi:Doi 10.1039/C3sc53163e (2014).
96 Allen, E. E. & Bartlett, D. H. Structure and regulation of the omega-3 polyunsaturated fatty acid synthase genes from the deep-sea bacterium Photobacterium profundum strain SS9. *Microbiology* **148**, 1903-1913 (2002).
97 Okuyama, H., Orikasa, Y., Nishida, T., Watanabe, K. & Morita, N. Bacterial genes responsible for the biosynthesis of eicosapentaenoic and docosahexaenoic acids and their heterologous expression. *Appl Environ Microbiol* **73**, 665-670, doi:10.1128/AEM.02270-06 (2007).
98 Piel, J. Biosynthesis of polyketides by trans-AT polyketide synthases. *Natural product reports* **27**, 996-1047, doi:10.1039/b816430b (2010).
99 Krissinel, E. & Henrick, K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr D Biol Crystallogr* **60**, 2256-2268, doi:10.1107/S0907444904026460 (2004).
100 Six, D. A., Yuan, Y., Leeds, J. A. & Meredith, T. C. Deletion of the beta-acetoacetyl synthase FabY in Pseudomonas aeruginosa induces hypoacylation of lipopolysaccharide and increases

antimicrobial susceptibility. *Antimicrob Agents Chemother* **58**, 153-161, doi:10.1128/AAC.01804-13 (2014).
101 Yuan, Y., Sachdeva, M., Leeds, J. A. & Meredith, T. C. Fatty acid biosynthesis in Pseudomonas aeruginosa is initiated by the FabY class of beta-ketoacyl acyl carrier protein synthases. *Journal of bacteriology* **194**, 5171-5184, doi:10.1128/JB.00792-12 (2012).
102 Zhang, Y. M. & Rock, C. O. Will the initiator of fatty acid synthesis in Pseudomonas aeruginosa please stand up? *Journal of bacteriology* **194**, 5159-5161, doi:10.1128/JB.01198-12 (2012).
103 Dawe, J. H., Porter, C. T., Thornton, J. M. & Tabor, A. B. A template search reveals mechanistic similarities and differences in beta-ketoacyl synthases (KAS) and related enzymes. *Proteins* **52**, 427-435, doi:10.1002/prot.10421 (2003).
104 Haapalainen, A. M., Merilainen, G. & Wierenga, R. K. The thiolase superfamily: condensing enzymes with diverse reaction specificities. *Trends Biochem Sci* **31**, 64-71, doi:10.1016/j.tibs.2005.11.011 (2006).
105 Olsen, J. G., Kadziola, A., von Wettstein-Knowles, P., Siggaard-Andersen, M. & Larsen, S. Structures of beta-ketoacyl-acyl carrier protein synthase I complexed with fatty acids elucidate its catalytic machinery. *Structure* **9**, 233-243 (2001).
106 Straight, P. D., Fischbach, M. A., Walsh, C. T., Rudner, D. Z. & Kolter, R. A singular enzymatic megacomplex from Bacillus subtilis. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 305-310, doi:10.1073/pnas.0609073103 (2007).
107 Oyola-Robles, D. *et al.* Identification of novel protein domains required for the expression of an active dehydratase fragment from a polyunsaturated fatty acid synthase. *Protein science : a publication of the Protein Society* **22**, 954-963, doi:10.1002/pro.2278 (2013).
108 James, T. Y. *et al.* Reconstructing the early evolution of Fungi using a six-gene phylogeny. *Nature* **443**, 818-822, doi:10.1038/nature05110 (2006).
109 Witkowski, A. *et al.* Head-to-head coiled arrangement of the subunits of the animal fatty acid synthase. *Chemistry & biology* **11**, 1667-1676, doi:10.1016/j.chembiol.2004.09.016 (2004).
110 Perham, R. N. Swinging arms and swinging domains in multifunctional enzymes: catalytic machines for multistep reactions. *Annual review of biochemistry* **69**, 961-1004, doi:10.1146/annurev.biochem.69.1.961 (2000).
111 Gokhale, R. S., Hunziker, D., Cane, D. E. & Khosla, C. Mechanism and specificity of the terminal thioesterase domain from the erythromycin polyketide synthase. *Chemistry & biology* **6**, 117-125, doi:10.1016/S1074-5521(99)80008-8 (1999).

112 Bukhari, H. S., Jakob, R. P. & Maier, T. Evolutionary origins of the multienzyme architecture of giant fungal fatty acid synthase. *Structure* **22**, 1775-1785, doi:10.1016/j.str.2014.09.016 (2014).
113 Tong, L. Structure and function of biotin-dependent carboxylases. *Cellular and molecular life sciences : CMLS* **70**, 863-891, doi:10.1007/s00018-012-1096-0 (2013).
114 Nygren, P. A. & Skerra, A. Binding proteins from alternative scaffolds. *Journal of immunological methods* **290**, 3-28, doi:10.1016/j.jim.2004.04.006 (2004).
115 Skerra, A. Imitating the humoral immune response. *Current opinion in chemical biology* **7**, 683-693 (2003).
116 Aparicio, J. F. *et al.* Organization of the biosynthetic gene cluster for rapamycin in Streptomyces hygroscopicus: analysis of the enzymatic domains in the modular polyketide synthase. *Gene* **169**, 9-16 (1996).
117 Donadio, S., Staver, M. J., McAlpine, J. B., Swanson, S. J. & Katz, L. Modular organization of genes required for complex polyketide biosynthesis. *Science* **252**, 675-679 (1991).
118 Solenberg, P. J. *et al.* Production of hybrid glycopeptide antibiotics in vitro and in Streptomyces toyocaensis. *Chemistry & biology* **4**, 195-202 (1997).
119 Kuhstoss, S., Huber, M., Turner, J. R., Paschal, J. W. & Rao, R. N. Production of a novel polyketide through the construction of a hybrid polyketide synthase. *Gene* **183**, 231-236 (1996).
120 Wong, F. T. & Khosla, C. Combinatorial biosynthesis of polyketides--a perspective. *Current opinion in chemical biology* **16**, 117-123, doi:10.1016/j.cbpa.2012.01.018 (2012).
121 Desmet, J. *et al.* Structural basis of IL-23 antagonism by an Alphabody protein scaffold. *Nature communications* **5**, 5237, doi:10.1038/ncomms6237 (2014).
122 Hoffmann, T. & Dougan, L. Single molecule force spectroscopy using polyproteins. *Chemical Society reviews* **41**, 4781-4796, doi:10.1039/c2cs35033e (2012).
123 Proulx, K., Cota, D., Woods, S. C. & Seeley, R. J. Fatty acid synthase inhibitors modulate energy balance via mammalian target of rapamycin complex 1 signaling in the central nervous system. *Diabetes* **57**, 3231-3238, doi:10.2337/db07-1690 (2008).
124 Crepin, T. *et al.* Polyproteins in structural biology. *Current opinion in structural biology* **32**, 139-146, doi:10.1016/j.sbi.2015.04.007 (2015).

Habib Bukhari, PhD
10/05/1988
Email: bukhari.habib@unibas.ch
Schwarzwalalle 173
4058 Basel Switzerland

Education

2006-2009	Bsc in Molecular Biology/AL-Farabi Kazakh National University/Almaty, Kazakhstan
2009-2011	Msc in Molecular Biotechnology/Technical University of Dresden/Dresden, Germany Master thesis supervisor: Jennifer Doudna (UC Berkeley) and Blake Wiedenheft (University of Montana)
2011-07.2014	PhD student in the lab of Prof. Dr. Timm Maier (Biozentrum Basel)
Since 07.2014	PostDoc in the lab of Prof.Dr. Timm Maier (Biozentrum Basel)

Professional Experience

2012-Current	PhD studentship working with Prof. Dr. Timm Maier (Biozentrum/Basel) on structural and functional analysis of large multienzymes.
03/2011-10/2011 (7 months)	Master Thesis research in Jennifer Doudna Lab (UC Berkeley) worked on biochemistry of the target recognition by the CRISPR RNA complex from *Pseudomonas aeruginosa*.
12/2010-2/2011 (3 months)	Worked in Max Planck Institute of Molecular Cell Biology and Genetics, as student assistant in Karla Neugebauer Lab (current Yale University). "Role of SR proteins in export of mRNA"
05/2010-10/2010 (5 months)	Part of iGEM team from TUD Biotec, under supervision of Andrew Oates, (current MRC UK). Developed bacterial based sensors for leukemia.
03/2010 -11/2010 (8 months)	Worked in the Center for Regenerative Therapies Dresden (CRTD)/Biotec as student assistant in Christopher Antos' Lab "Appendage and Organ Regeneration in zebrafish"
01/2008-05/2009 (16 months)	Worked in "General Genetics and Cytology Institute" as research assistant in the lab of Djansugurava L.B, Kazakstan Almaty. Studying Base Excision Repair pathways.

Honors and Awards

03/2011 Dresden/Germany	Gesellschaft von Freunden und Förderern der TU Dresden scholarship award
10/2010 USA/Boston/MIT	Gold medal in the iGEM jamboree, representing TUD/Biotec
11/2009 Germany/Bonn	DAAD scholarship award
05/2009 Kazakhstan/Almaty	Award, for being in the top 5 % of students

Publications

Mechanism of foreign DNA recognition by a CRISPR RNA-guided surveillance complex from *Pseudomonas aeruginosa*.
Rollins MF, Schuman JT, Paulus K, **Bukhari HS**, Wiedenheft B.
Nucleic Acids Res. 2015 Feb 8.

Evolutionary origins of the multienzyme architecture of giant fungal fatty acid synthase.
Bukhari HS, Jakob RP, Maier T.
Structure. 2014 Dec 2;22(12):1775-85

Manuscripts in preparation

Evolutionary origins of Type I Fatty Acid Synthase
Bukhari HS and Timm Maier
Review

Function characterization of intra linking regions in large multienzymes.
Bukhari HS and Timm Maier
Article

In this project we have altered mechanical and physical properties of linker regions in a large multienzyme to better characterize how intradomain linking could influence catalytic properties and conformational crosstalk between domains. This was achieved by generating more then three dozens of different constructs with various linker sequences. Combined structural and kinetic data from purified constructs helped us to better understand the emergent properties of the megasynthase system. A long-term goal is to use these insights for the construction of artificial multienzymes incorporating complete and complex molecular pathways.

Techniques

Cloning	-Familiar with most of the advanced and high throughput techniques specializing on cloning of very long and difficult DNA constructs. (Red ETg and Red a/b/y recombination). -Experience in precise and fluent manipulation of bacterial and fungal genomes.
Protein Production and purification	-Experience in expression and purification of large and difficult protein complexes in bacterial and eukaryotic hosts (HEK, yeast and insect cells). -HPLC/FPLC at GE ÄKTA and Biorad NGC systems;
Protein crystallization	Involved in x-ray crystallography work of 10-15 different proteins.
Biochemical Techniques	-Experience in working with radioactive labeled DNA and RNA -Enzymatic assay development for high throughput screening.
Bioinformatics	-Knowledge of a series of bioinformatics tools and features concerning phylogenetic analysis. -Programing in Python.

Further Knowledge

Languages	English (fluent), Russian (fluent), Urdu (fluent), Kazakh (Intermediate) and German (B1)
3D design	Maya Autodesk and Adobe Creative Suite (Advanced)