Functional Architectures of Polyketide Synthases

Inauguraldissertation

zur

Erlangung der Würde eines Doktors der Philosophie

vorgelegt der

Philosophisch-Naturwissenschaftlichen Fakultät

der Universität Basel

von

Dominik Alexander Herbst

aus Hessisch Oldendorf, Germany

Basel, 2018

Genehmigt von der Philosophisch-Naturwissenschaften Fakultät auf Antrag von

Prof. Dr. Timm Maier

Prof Dr. Henning Stahlberg

Basel, den 20.06.2017

Prof. Dr. Martin Spiess

Dekan

I ABSTRACT

Microbial polyketide synthases (PKS) are biological factories for the production of potent natural products, which include clinically relevant antibiotics, anti-cancer drugs, statins and more. The exceptional chemical diversity generated by PKSs is encoded in a modular architecture for precursor extension. The domains required for one step of precursor elongation and modification are combined into a functional polypeptide module, which is segregated into a mandatory condensing region for elongation and an optional and variable part for intermediate modification. PKS modules contain integral acyl carrier protein (ACP) domains, flanked by flexible peptide regions. ACPs are used to load substrates and to tether intermediates throughout ongoing synthesis, by linking them as thioesters to a covalently attached phosphopantetheine cofactor. PKS modules can either act iteratively (iPKS) or in a linearly organized assembly line of multiple modules (modPKS), where the nascent polyketide is handed over from one to the next module. The collinearity between synthesis and protein sequence in modPKS holds promise for rational re-engineering in order to produce novel bioactive compounds. Despite their cyclic mode of action, iPKS may employ specific reaction programs, which introduces different substitutions in each iteration by selective use of individual catalytic domains.

At the beginning of the thesis, the architecture of PKS modules as a basis for their modular organization and programmed biosynthesis was unknown. This thesis was focused on structural studies of the architecture of PKS modules, intramodular crosstalk and functional programming. Chapter one provides a comprehensive introduction into the molecular biology of PKS function.

Chapter two provides a hybrid crystallographic model of an iPKS module and demonstrates its relevance also for modPKS. Overlapping crystal structures of a condensing and a complete modifying region provided the first atomic model of a PKS module with a total of 10 catalytic domains. Multiple crystallogrpahically independent copies observed in the 3.75 Å structure of the dimeric modifying region provided snapshots of a variable linker-based architecture with implications for PKS evolution and conformational coupling of reaction steps in the dimeric synthase. Comparative small angle X-ray scattering demonstrates that the iPKS architecture is also representative for tested modPKSs.

Chapter three reports the crystal structure of a programming C-methyltransferase (CMeT) domain at 1.65 Å resolution. The structure reveals a novel N-terminal fold and a substrate binding cavity that accommodates intermediates of various length during iterative biosynthesis. Structural and phylogenetic analysis demonstrates conservation of CMeT domains in PKS as well as homology to an inactive pseudo-CMeT (ΨCMeT) remnant in mammalian fatty acid synthase (mFAS). The data suggest an involvement of the core elongating ketosynthase (KS) domain in PKS programming.

Chapter four provides a visualization of substrate loading in iPKS. A 2.8 Å resolution crystal structure provided detailed insights into an intertwined linker-mediated integration of substrate-loading starter-unit acyltransferase (SAT) domains into an iPKS condensing region. The post-loading state was trapped by mechanism-based crosslinking. Visualization by cryo electron microscopy at 7.1 Å resolution revealed asymmetry of ACP-KS interactions and depicts conformational coupling across the dimeric PKS for coordinated synthesis.

Chapter five integrates the results into the current structural and biological context and discusses current opinions and future perspectives in the field. The results of this thesis reflect the relevance of linker-based connections rather than stable domain-domain interfaces for PKS architecture. This work also highlights mechanisms for conformational coupling for synthesis and substrate channeling in dimeric, but asymmetric, PKS. These insights will support re-engineering iPKS and modPKS assembly lines for the production of novel bioactive compounds, in particular for drug discovery.

II ACKNOWLEDGEMENTS

First of all, I would like to thank Timm for the great opportunity to work on PKS in his lab, his support and being a great mentor. Thanks for the Coffee, your time and being inspiring!

Many thanks Henning for joining my committee and running an awesome EM facility.

Vielen Dank an meine Familie und vor allem meine Eltern Christine und Klaus Herbst, die mir alles in meinem Leben ermöglicht und mich immer unterstützt haben! Ohne euch wäre ich nicht dort wo ich jetzt bin! Vielen Dank!

Grosser Dank gilt auch meinem verstorbenen Onkel Ulrich Cruel, der mich von klein auf für Naturwissenschaften interessierte und mich dahingehend prägte.

I would like to thank Roman Jakob for his support and help throughout the years!
Thanks for the sweets and being my postman ;-)

Special thanks go to all my friends, who have been particularly patient during the last few month:

Particularly, I would like to thank Anne for always trying to animate and cheer me up, Ricardo & Ana for coming to Basel ☺ and all the EM discussions, Fab for always being there to have a beer ☺, Yusuke for being an awesome cook, Chee and Diana for great dinners at your place and Martino & Francesca for a great time.

Thanks all my friends at home! Thanks to Rebecca and Jens for being here today!

Thanks all former and present members of the lab! Thanks Moritz and Stefan for occasional Apèros; Anna, Eddi and Roman for taking care of all the instruments, and all others for creating a great environment: Freddy, Francesca, Leonie, Yana, Janine, Shubham, Yves, Matthias and Alexandra.

Thanks Yves for being a great master student, doing a great job and being patient during the last months.

Many thanks everybody who kept and keeps the great infrastructure at the Biozentrum running: Mohamed, Kenny, Ariane and Bill for keeping the microscopes running and providing us opportunities to use them; Barbara und Beat for managing the floor;

Particularly, I would like to thank all people who kept the IT running and are normally only contacted if something does not work… Guys, your doing a great job!
Thanks all people from Scicore, particularly Martin and Pablo and to everybody from Biophit, particularly Manuela and (initially) Jan!

Thanks all PhD Reps! We had a great time!

Thanks everybody else for the great time!

Thanks for funding by
the Werner Siemens Foundation / Biozentrum Fellowships For Excellence

III TABLE OF CONTENTS

IV LIST OF FIGURES & TABLES

Figures

Tables

V ABBREVIATIONS

aa	Amino acids
ACC	Acetyl-CoA carboxylase
ACP	Acyl carrier protein
AcSNAC	Acetyl-*S*-*N*-acetylcysteamine
ADP	Adenosine diphosphate
AH	Acylhydrolase
AMPCPP	α,β-Methyleneadenosine 5'-triphosphate
AT	Acetyltransferase
ATP	Adenosine triphosphate
AUC	Analytical untracentrifugation
AVES	Avermectin polyketide synthase
B	Branching domain
BCCP	Biotin carboxyl carrier protein
BIS-TRIS	Bis(2-hydroxyethyl)amino-tris(hydroxymethyl)methane
BLAST	Basic Local Alignment Search Tool
CC	Cross correlation
CMeT	C-methyl-transferase
CoA	Coenzyme A
cryo-EM	cryo electron microscopy
CTB1	Cercosporin biosynthesis gene 1
CV	Column volume
DEBS	6-Deoxyerythronolide B synthase
DH	Dehydratase
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
E.Coli	*Escherichia coli*
e.g.	Exempli gratia
EC	Enzyme commission
EDTA	Ethylenediaminetetraacetic acid
EM	Electron microscopy
ER	Enoylreductase
ESI	Electrospray ionization
FAS	Fatty acid synthase
FD	Ferredoxin-like subdomain
FMN	Flavin mononucleotide
GPC	Gel permeation chromatography
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HPLC	High-performance liquid chromatography
HR-PKS	Highly reducing iterative polyketide synthase
ID	Inter domain region
id	Identity
iPKS	Iterative polyketide synthase
IPTG	Isopropyl β-D-1-thiogalactopyranoside
kB	kilo bases
kDa	Kilo dalton
KR	Ketoreductase
KS	Ketosynthase
LD	Linker domain
LDD	Loading didomain

LINKS	Laterally-INteracting Ketosynthase Sequence
LOS	Lipooligosaccharide
LSSR	Local structural similarity restraints
MalSNAC	Malonyl-*S*-*N*-acetylcysteamine
MAS	Mycocerosic acid synthase
MAT	Malonyl-transferase (Malonyl-acetyl-transferase in the context of FAS)
MB	Mega base
MDa	Mega Dalton
MDR	Medium-chain dehydrogenase/reductase
MES	2-(N-morpholino)ethanesulfonic acid
mFAS	Mammalian FAS
miRNA	micro-RNA
modPKS	Modular polyketide synthase
MPD	2-methyl-2,4-pentanediol
MS	Mass spectrometry
MSAS	6-methylsalicylic acid synthase
Msl	MAS-like (members of the MAS-like PKS family)
MTAN	Methylthioadenosine nucleosidase
MTases	methyltransferases
NADH	Nicotinamide adenine dinucleotide
NADPH	Nicotinamide adenine dinucleotide phosphate
NMR	Nuclear magnetic resonance
NR-PKS	Non-reducing iterative polyketide synthase
NRPS	Nonribosomal peptide synthetase
ORF	Open reading frame
PAGE	Polyacrylamide gel electrophoresis
Pap	Polyketide associated protein
P-body	processing body
PCP	Peptidyl carrier protein
PDB	Protein data bank
PDH	pyruvate dehydrogenase
PEG	Polyethylene glycol
PEG MME	Polyethylene glycol monomethyl ether
PK	Polyketide
PKS	Polyketide synthases
PLP	Pyridoxal phosphate (active form of Vitamin B6)
Ppant	Phosphopantetheinyl
PPT	Phosphopantetheinyl transferases
PR-PKS	Partially reducing iterative polyketide synthase
PSI	Paul Scherrer Institute
PT	Product-template domain
PTM	Polycyclic tetramate macrolactam synthase
PUFA	Polyunsaturated fatty acid synthase
R	Reductase domain (functionally similar to TE, but involves reduction)
rmsd / r.m.s.d.	Root-mean-square deviation of atomic positions
RNA	Ribonucleic acid
SAH	S-adenosylhomocysteine
SAM	S-adenosylmethionine
SAS	Solvent accessible surface
SAT	Starter-acyltransferase
SAXS	Small-angle X-ray scattering
SDR	Short-chain dehydrogenase/reductase

SDS	Sodium dodecyl sulfate
Sfp	Surfactin NRPS phosphopantetheinyl transferases
SLS	Swiss Light Source
SNAC	*S*-*N*-acetylcysteamine
SNF	Swiss National Science Foundation
TCEP	Tris(2-carboxyethyl)phosphine hydrochloride
TE	Thioesterase
TE/CLC	Thioesterase / Claisen-cyclase
TEC	The transcription elongation complex
TEV	Tobacco etch virus
TIM	Triosephosphase isomerase
TLS	Translation/Libration/Screw
TPP	Thiamine pyrophosphate (active form of Vitamin B1)
TRIS	Tris(hydroxymethyl)aminomethane
UMA	Udwary-Merski algorithm
UPLC	Ultra Performance Liquid Chromatography
WHO	World health organization
yFAS	Yeast fatty acid synthase
ΨCMeT	Pseudo-C-methyl-transferase
ΨKR	Pseudo-ketoreductase

1 INTRODUCTION

1.1 Enzymes in biological pathways

Proteins are the main carrier of biological function. They are polymers of 21 different amino acids (aa, also called residues) with a typical length of 80-10,000 aa that are linked via peptide bonds and therefore are also called polypeptides. Every amino acid contributes one side chain of defined chemical properties to the polypeptide and their sequence is encoded in genetic information, which is translated into proteins by ribosomes. Once a protein is produced, its polypeptide folds and adopts a higher ordered structure, which is hierarchically classified into the primary (aa sequence), secondary (local arrangement into helices or sheets consisting of strands), tertiary (topology of secondary structure elements) and quartanery structure (complexes of various proteins). Regions that fold independently of others are called domains. A human cell contains approx. 20,000 different proteins whereas bacteria such as *Escherichia coli* (*E.Coli*) contain approx. 6,000 different proteins[1], that are involved in all cellular processes, including replication and translation of genetic information, cellular organization, signaling and motion. A key function of proteins is to catalyze chemical reaction, that are too inefficient to occur without catalysis under cellular conditions[2]. Biological catalysts are called enzymes and comprise 21 % (human) and 42 % (*E.Coli*), respectively of all annotated proteins as of May 2017[3].

Enzymes are biological catalysts that accelerate chemical reactions by reducing activation energies[2]. The site of an enzyme that binds a substrate and provides the chemical environment to catalyze a chemical reaction is called active site. This environment includes catalytic side chains of the protein, cosubstrates, ions, water and cofactors, respectively. Cofactors can be derived from vitamins and assist catalysis by direct or indirect involvement in the chemical reaction. While cofactors such as nicotinamide adenine dinucleotide phosphate (NADPH) are cosubstrates and consumed in reduction reactions, other cofactors such as Coenzyme A (CoA) are regenerated during the catalytic cycle and therefore are also called coenzymes. Cofactors that are covalently attached to the enzyme are called prosthetic groups, such as the pantothenic acid (vitamin B_5) derived phosphopantetheinyl (Ppant) cofactor.

In metabolism, multi-step chemical reactions are catalyzed by sequential action of enzymes in a biological pathway. All enzymes operate at equilibrium conditions and can accelerate reactions in both directions. However, all pathways have a direction, like the degradation of glucose in glycolysis. Under steady state conditions this direction is dictated by individual enzymes that catalyze irreversible steps, *e.g.* the loss of carbon dioxide or hydrolysis of energy equivalents such as adenosine triphosphate (ATP). Enzymes that catalyze such irreversible reaction in a pathway are key switches for regulation[2] and can stop an entire pathway upon inhibition.

Regulation of pathways also occurs at the gene level, which is the case if a pathway is not regularly used. Enzymes of such pathways are encoded in gene clusters under the control of a promotor that is activated by transcription factors as response to a signal (*e.g.* the lac operon)[4]. In microorganisms this regulation is often used for pathways that are not necessarily required for survival (secondary metabolite pathways), such as biosynthetic pathways for the production of natural products that are used in symbiosis, protection or chemical warfare with other organisms.

1.2 Substrate tethering to prosthetic groups and carrier proteins

Substrates and their intermediates can be covalently tethered to cofactors in order to alter chemical properties or stabilize chemical bonds. Biotin is used for the activation of carbonic acids in carboxylation reactions, whereas lipoyl is an oxidation agent for the production of energy-rich thioesters. Phosphopantetheinyl (Ppant) cofactors are derived from CoA and tether substrates via thioesters that can be utilized to form and cleave carbon-carbon bonds, respectively. In their active form, all three cofactors are covalently tethered as prosthetic groups to 8-10 kDa (70-80 aa) large carrier proteins (Figure 1.1). Lipoyl and biotinyl carboxylase carrier proteins (BCCP) share a common fold of two parallel arranged β-sheets and connect their cofactors via an amide bond to a lysine at the distal end of the protein relative to their N- and C-termini (Figure 1.1a, b). Ppant cofactors are found in acyl- and peptidyl carrier proteins (ACP/PCP) (Figure 1.1c, d), respectively, that are only distinguished by the kind of substrate they tether to the Ppant. These proteins consist of a conserved four α-helical bundle and connect the Ppant via a phosphoric ester to a serine side chain at the N-terminus of their second helix (helix II). The cofactor itself resembles a flexible arm with a length of 16-20 Å and can dock to the active site of a target enzyme for substrate delivery. The carrier proteins therefore serve a transport function[5-7].

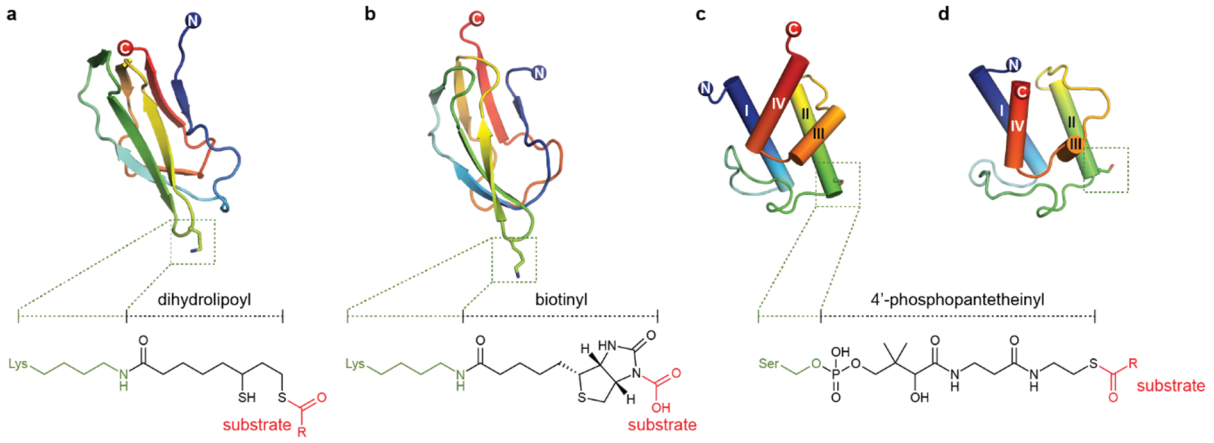


Figure 1.1 | Carrier domains and their substrate tethered cofactors.
Structures of representative apo-carrier proteins, their cofactor attachment points and substrate loaded states are shown. **a**, Lipoyl carrier domain from the human pyruvate dehydrogenase complex (PDB: 1FYC)[8]. **b**, Biotin carboxyl carrier protein (BCCP) domain from yeast acetyl-CoA carboxylase (PDB: 5CSA)[9]. **c**, Peptidyl carrier protein (PCP) domain seven of the teicoplanin nonribosomal peptide synthetase (NRPS) of *Actinoplanes teichomyceticus* (PDB: 2MR7)[10]. **d**, Acyl carrier protein (ACP) domain of the non-reducing polyketide synthase (NR-PKS) PksA from *Aspergillus parasiticus* (PDB: 2KR5)[11]. All carrier proteins tether their substrates covalently via prosthetic groups. Lipoyl (a) and BCCP (b) carrier protein domains share a similar β-sheet topology and connect their cofactors via amid bonds to a lysine. ACP and PCP domains share a conserved four helix bundle topology and both tether their substrates via Ppant, which is attached via a phosphoric ester bond to a conserved serine on helix II. Substrate loaded states are indicated in red. The polypeptide backbones are rainbow colored from the N-(blue) to the C-terminus (red).

ACPs in bacterial fatty acid biosynthesis sequester their substrate in a cavity between helix II and III[12] and can expose it upon binding to target enzymes by an allosteric "switch blade" mechanism[13] (chapter 1.6.3). This provides an efficient protection mechanism of highly reactive intermediates during cytosolic transport. However, in the context of multienzymes (chapter 1.4) in fatty acid[14] and polyketide[11] biosynthesis, sequestration has not been observed.

Besides substrate protection, tethering to carrier proteins can have the advantage of increased affinity and specificity with target domains associated to a specific pathway. If freely diffusing substrates would be turned over from all enzymes that are potentially able to catalyze reactions, different pathways could interfere and compete for them causing a metabolic imbalance. Therefore, carrier proteins tether their substrates not just to themselves, but also to a specific pathway. This strategy requires a tight cooperation between a carrier protein and its associated enzymatic domain for efficient turnover (Figure 1.2). Indeed, many pathways carry similar or even identical substrates, but their carrier proteins are not fully cross-compatible as exemplified by the fatty acid synthase ACP and its equivalent in the actinorhodin polyketide synthase (PKS) (both) from *S. coelicolor* A3[15,16]. In PKS and non-ribosomal peptide synthetase (NRPS) pathways, often several synthase and synthetase units, respectively, can be found that usually do not interfere. Recent biochemical and structural studies of ACPs with their respective target enzymes indicated key residues for interactions[13,17-35]. Nevertheless, a universal specificity conferring code between enzymes and carrier proteins could not be identified, which might be obscured by the fact that one carrier protein transports its substrate to a set of different enzymes in a sequential order. All observed interactions usually involve protein-protein, enzyme-substrate, and protein-co-factor interactions.

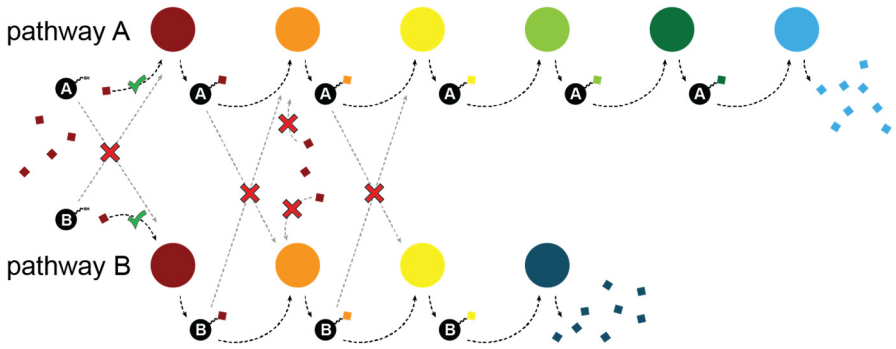


Figure 1.2 | Carrier proteins tether a substrate to a specific pathway.
Substrate delivery scheme of two pathways dependent carrier proteins (black). Enzymes of these pathways recognize substrates that are tethered by their own carrier proteins with higher specificity than carrier proteins of other pathways and untethered substrates, respectively. Enzymes catalyzing the same reaction are shown as colored spheres. Substrates are shown as squares and colored according to the enzyme that produced it. The first enzyme in each pathway (dark red) loads the carrier protein.

1.3 Post-translational modification of carrier protein domains

All carrier proteins have to be converted from their inactive *apo*- (lacking the cofactor) to their active *holo*-form by post-translational modification reactions. Phosphopantetheinylation of ACP and PCP domains is achieved by transfer of Ppant from CoA to the conserved active site serine of helix II (Figure 1.3a, d). This reaction is catalyzed by phosphopantetheinyl transferases (PPT) such as the bacterial AcpS and Sfp. AcpS is responsible for loading the ACP of bacterial fatty acid biosynthesis (Figure 1.3b), whereas Sfp is the PPT of the secondary metabolism pathway of the surfactin NRPS from *Bacillus subtilis*, which produces the antibiotic lipopeptide surfactin (Figure 1.3c, d). It was shown that usually every pathway has its own PPT and therefore the specificity with regard to their cognate carrier proteins is relatively high[36]. However, Sfp proved to have a much broader specificity and is capable of loading carrier proteins of the primary metabolism *in vitro*. This feature makes Sfp to a useful tool for *in vitro* loading reactions of recombinant expressed carrier proteins[37].

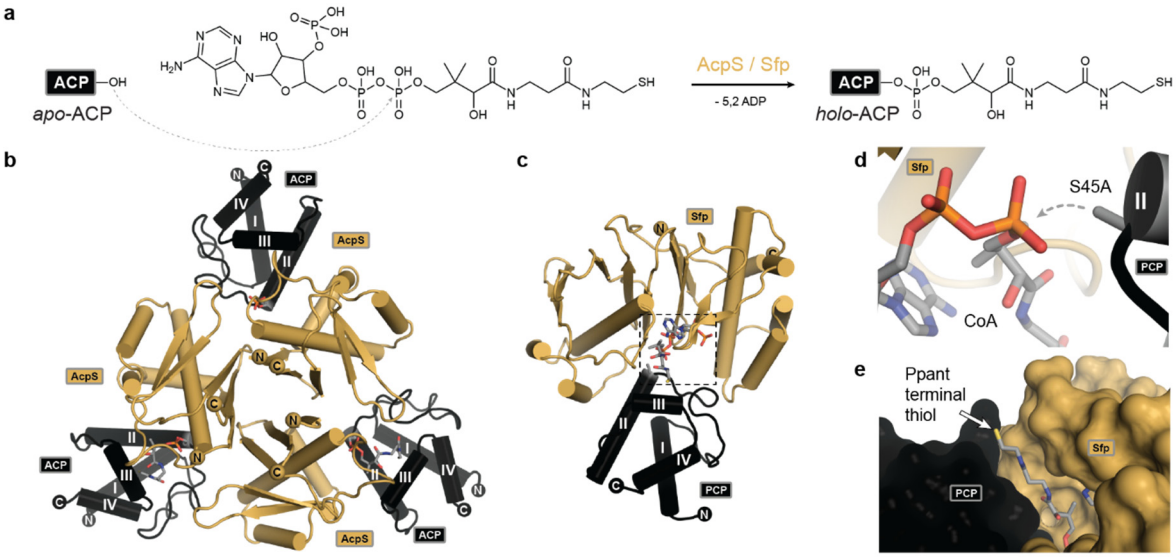


Figure 1.3 | The phosphopantetheinylation reaction catalyzed by AcpS and Sfp.
a, Both, ACP and PCP rely on post-translational modification by PPTs like AcpS and Sfp. These enzymes transfer Ppant from CoA onto a conserved serine residue in helix II. **b**, Hetero-hexameric complex of AcpS with ACP from *Bacillus subtilis*[38] responsible for phosphopantetheinylation of the ACP in fatty acid biosynthesis (PDB: 1F80). **c**, Complex structure of *Bacillus subtilis* Sfp with the PCP domain of module 3 (TycC) from the Tyrocidine NRPS of *Brevibacillus parabrevis*[39] (PDB: 4MRT). Sfp has a much broader substrate tolerance than AcpS and is often used in *in vitro* reactions to activate ACPs and PCPs of different pathways. **d**, Close-up view (indicated by a box in panel c) on the active site showing the transfer mechanism. The structure was solved with an alanine mutant of the reactive PCP serine. The nucleophilic attack of the native serine is indicated. **e**, The terminal thiol group of the CoA Ppant moiety is exposed to the outside. CoAs and Ppant cofactors are shown in stick representation.

All PPTs are active as oligomers and bind CoA as well as one ACP at their interfaces, as exemplified by the group of trimeric AcpS (Figure 1.3b). Sfp contains two PPT domains that are fused in a pseudo dimer with only one active site (Figure 1.3c). Interestingly, the Sfp-PCP complex buries the pyrophosphate moiety of CoA and exposes the terminal thiol group to the outside. This feature provides an opportunity for loading substrate charged Ppant co-factors for biochemical and structural studies (chapter 4, Supplementary Figure 4.10)[39].

1.4 Multi-subunit complexes vs Multienzymes

Enzymes can occur either as monomers, oligomers as part of a multi-subunit complex or as multienzymes. The discrimination between the latter two is critical: In multi-subunit enzyme complexes each subunit may be a different protein (or RNAs) that contributes one or several domains or enzymatic activities to the complex (Figure 1.4a), *e.g.* the *E.Coli* acetyl-CoA carboxylase (ACC)[40]. In contrast, multienzymes contain several enzymatic domains on a single polypeptide that are connected by long linkers (Figure 1.4b) that have evolved by gene fusion events[41]. In many cases multienzymes form higher oligomers that are called multienzyme complexes or megaenzymes. Representatives of this multienzyme family are type I (see chapter 1.5) PKS and fatty acid synthase (FAS), NRPS, and the eukaryotic ACC.

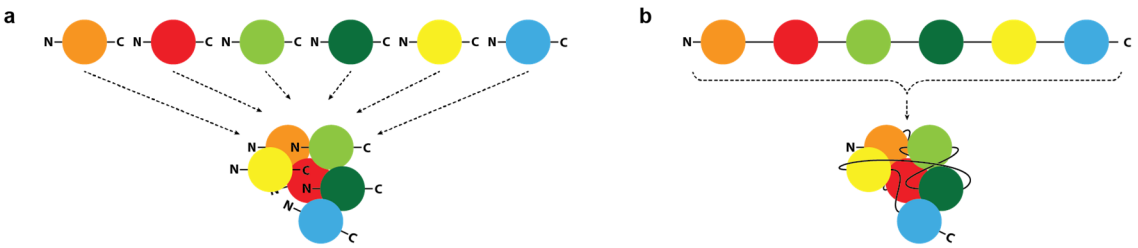


Figure 1.4 | The difference between multi-subunit complexes and multienzymes.
a, The domains of a multi-subunit complex are individual proteins. **b**, The domains of a multienzyme are part of the same polypeptide and are connected via long linkers. Domains are shown as colored circles.

Multi-subunit and multienzymes have in common that they provide short active site distances for efficient substrate channeling. However, the domains in multi-subunit complexes are involved in dissociation equilibrium reactions and therefore require strong interfaces on all domains that permit permanent interactions. In higher oligomeric multienzymes individual domain interfaces are relatively weak, because the presence of several interfaces in a single polypeptide provides avidity and thus might reduce the evolutional pressure on maintaining strong interfaces. This can become a problem for structural studies, because multienzymes (particularly PKS) are mostly studied as excised fragments, which can cause dissociation[42-47] (for more details see chapter 1.6.2).

The advantage of having a carrier protein as described in chapter 1.2 comes with the disadvantage of reduced turnover rates caused by lower diffusions coefficients of the larger carrier proteins compared to isolated substrates[15]. In a multienzyme this can be compensated, because a carrier protein is usually part of the polypeptide chain, which limits diffusion to the length of its connecting linkers. The linkers are approx. 20-40 aa long and usually double tethered with N- and C-termini close to each other (Figure 1.1). In many cases more than one carrier domain can be found in sequence, which is believed to increase the net-flux of substrates through the multienzyme[48]. The timescale of product biosynthesis by multienzymes is relatively low with respect to the amount of reactions and active site translocations they include. The mammalian fatty acid synthase (mFAS) catalyzes the production of C_{16}-fatty acids with six active sites from acetyl- and malonyl-CoA in approx. 1.25 sec (*in vitro*)[49], while the even more complex PKS 6-deoxyerythronolide B synthase (DEBS) with 23 active sites requires approx. two minutes (*in vitro*)[50], which will be described next.

1.5 Polyketide synthases (PKS)

Polyketide synthases are large dimeric multienzymes responsible for the production of natural products with biological activities such as antibiotics, anticancer drugs, toxins, statins, cell envelop components of mycobacteria and many more[51] (Figure 1.5). By the beginning of the millennium more than 60 % of all anticancer and 75 % of all anti-infectious disease drugs were from natural origin involving PKS and NRPS pathways[52]. In 2014 the WHO reported a growing number of antibiotic resistant strains, whereas no new classes of antimicrobial substances have been reported since the 1980s[53]. Recently, the polyketide avermectin received lots of attention for its discovery and application in parasitic worm treatments, which was awarded with the Nobel-Prize in Medicine and Physiology in 2015[54]. This shows that a better understanding of PKS can be crucial for the development of new drugs and help to fight multi-drug resistant pathogens.

Polyketides are chemically complex structures, but share some common properties. All polyketides have a continuous poly carbon chain and substitution patterns derived from carbonyl- and hydroxyl

groups, double bonds, and fully saturated carbon-carbon bonds. The name derives from their core catalytic activity, which elongates a carbon chain in Claisen-like condensation reactions and produces β-di-carbonyl functions that are subject to further modifications. Additionally, many pathways and producer strains contain tailoring enzymes that introduce further modification to the products like methylation, oxidation or the transfer of amino acids and sugars to increase solubility.

PKS are found in microorganism (bacteria, fungi, marine organism, protists) with a very specific ecological niche and limited resources or specialized adaption to their environment. Naturally, these organisms produce polyketides for chemical warfare with competing organism, symbiosis, parasitosis, or for the production of compounds that help to resist environmental conditions[55-57]. Remarkably, the larger the genome of a certain microorganism, the higher the probability to find one or multiple PKS clusters. Bacterial genomes with less than 2 Mbp lack PKS, whereas 88 % of bacteria with more than 7 Mbp have PKS[41].

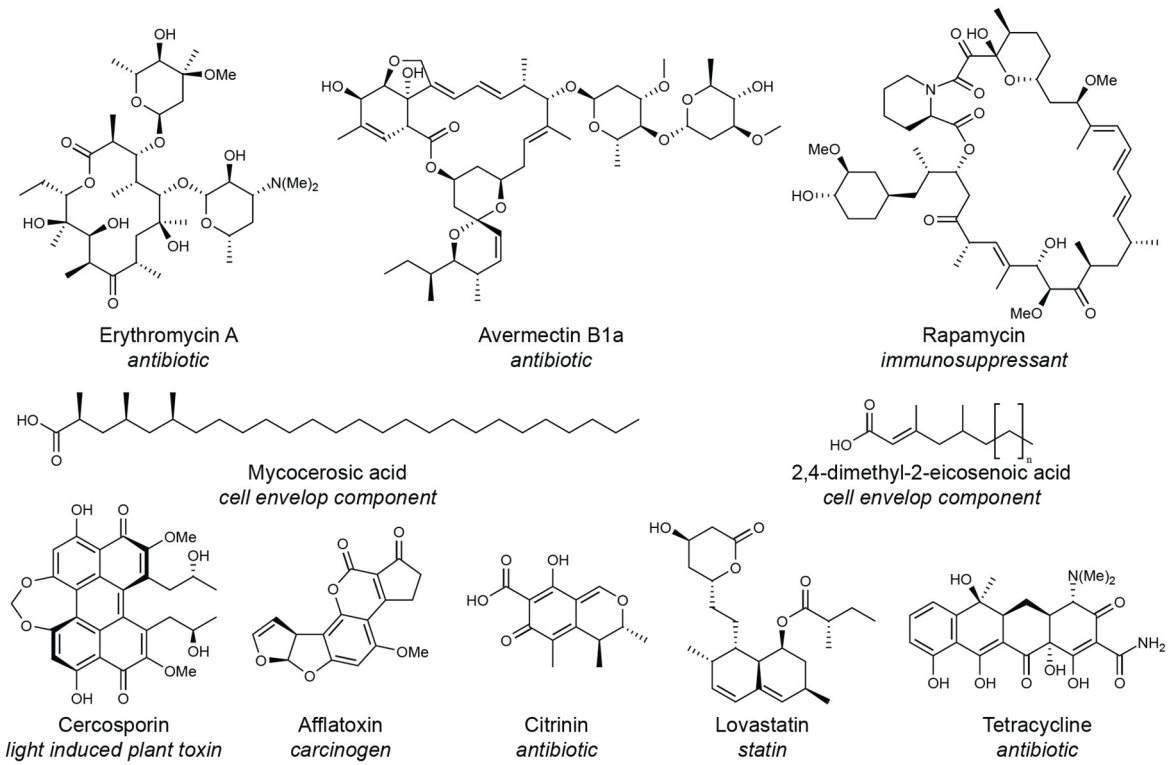


Figure 1.5 | Polyketide natural products.
PKS can produce a large variety of chemically divergent bioactive compounds including antibiotics, immunosuppressants, cell envelop components, statins (anti-cholesterol), toxins and many more. Names and functions (italic) are indicated. Except for tetracycline all shown products are produced by type I PKS.

PKS are classified to three superfamilies: Type I PKS are large ACP dependent multienzymes as introduced previously in chapter 1.4 and will be discussed in detail below. Type II PKS are also ACP dependent, but consist of distinct monofunctional enzymes that produce aromatic compounds like tetracycline (Figure 1.5)[58]. Type III PKS (also known as chalcone and stilbene synthases) are specialized individual enzymes and mainly found in plants[55,58]. All reactions (condensation and aromatization) take place in a single active site on a CoA tethered substrate. Since the substrate has not to be transported to other enzymes these PKS are independent of carrier proteins. The discrimination in type I and II superfamilies does also exist for FASs. However, all FASs including type II FAS produce fully saturated fatty acids, while type II PKS produce aromatic polyketides. Although these three superfamilies are

clearly defined, their definition sometimes becomes blurred by the identification of increasing numbers of hybrid pathways[55]. PKS as well as FAS use a common set of homolog enzyme (see chapter 1.6.1), which makes it easy to identify a PKS by homology methods. The growing number of available microbial genomes in GenBank (currently more than 13,700) led to an explosion of newly identified PKSs by genome mining[59,60]. However, the identification of products encoded by a PKS cluster is extremely difficult, because often the natural producer cannot be cultured (*e.g.* endosymbionts of deep sea sponges), does not produce a sufficient amount to study the products, or the promoter of the PKS cluster is silent. The common strategy in such cases involves cloning of the cluster (approx. 10-100 kbp) by advanced cloning methods[61-63] and heterologous expression under the control of an inducible promoter in a designated producer strain, that still might not provide all required precursors or exhibits a different tailoring apparatus[64]. In many cases this represents a daunting, if not impossible, task and alternatives for the identification of new drug candidates have to be explored. The most obvious approach is PKS engineering, which will be reviewed in chapter 1.5.4.

All following chapters will focus on type I PKS multienzymes.

1.5.1 Function and organization of enzymatic domains in type I PKS

All type I PKS require a minimum of two catalytic domains and one ACP. In a first step the synthase has to be primed with a substrate that subsequently will be elongated by the core catalytic ketosynthase (KS) domain in an irreversible decarboxylative Claisen-like condensation (Figure 1.6a). Loading of the priming substrate is controlled by a gatekeeping acetyltransferase (AT), which loads an acetyl-CoA derivative onto the Ppant cofactor of the ACP by transesterification. AT domains are also responsible for loading of all other extender units, which can be accomplished by either the same priming AT or by a separate one. If a designated priming AT domain is present, this domain is often called starter-unit AT (SAT) (see chapter 1.5.2) or loading AT (see chapter 1.5.3). Extender units are β-carboxy thioesters (malonyl-CoA derivatives) that are produced by biotin dependent caboxylases and can be predicted based on AT sequence motifs[65]. In literature for NR-PKS (chapter 1.5.2) and FAS (chapter 1.6.1) the extender unit specificity of AT domains is often reflected by their name, while the nomenclature is not consistent (*e.g.* MAT in NR-PKS described a malonyl-CoA dependent AT, while MAT in FAS refers to malonyl-acetyl transferase).

Once a priming substrate has been loaded to the ACP, it translocates to the KS domain and transfers the substrate to the active site cysteine of the KS. Afterwards the ACP translocates back to the/an AT, is loaded with an extender unit and translocates to the KS again, where it binds and inserts the Ppant cofactor with the tethered substrate into the active site tunnel. The condensation occurs by decarboxylation of the β-carboxy ester by forming a carbon nucleophile that attacks the thioester of the priming substrate with the KS active site cysteine (Figure 1.6a). The product is an ACP tethered β-keto thioester, which can be transferred back to the KS active site for further rounds of chain elongation or to optional modifying domains that catalyze reductions, water eliminations, isomerizations and methylations on the β-carbon atom (Figure 1.6b). This separates PKS into two functional units consisting of a mandatory condensing region (KS-AT, ACP) and an optional modifying region. The repertoire of domains in the modifying region determines the substitution pattern of the product (Figure 1.6c).

The typical modifying region in reducing PKS contains three to four different domains. Usually the first modification is an NADPH dependent reduction of the β-carbonyl to a hydroxyl group via a ketoreductase (KR) domain (Figure 1.6b). Subsequently, a dehydratase domain (DH) can form a double

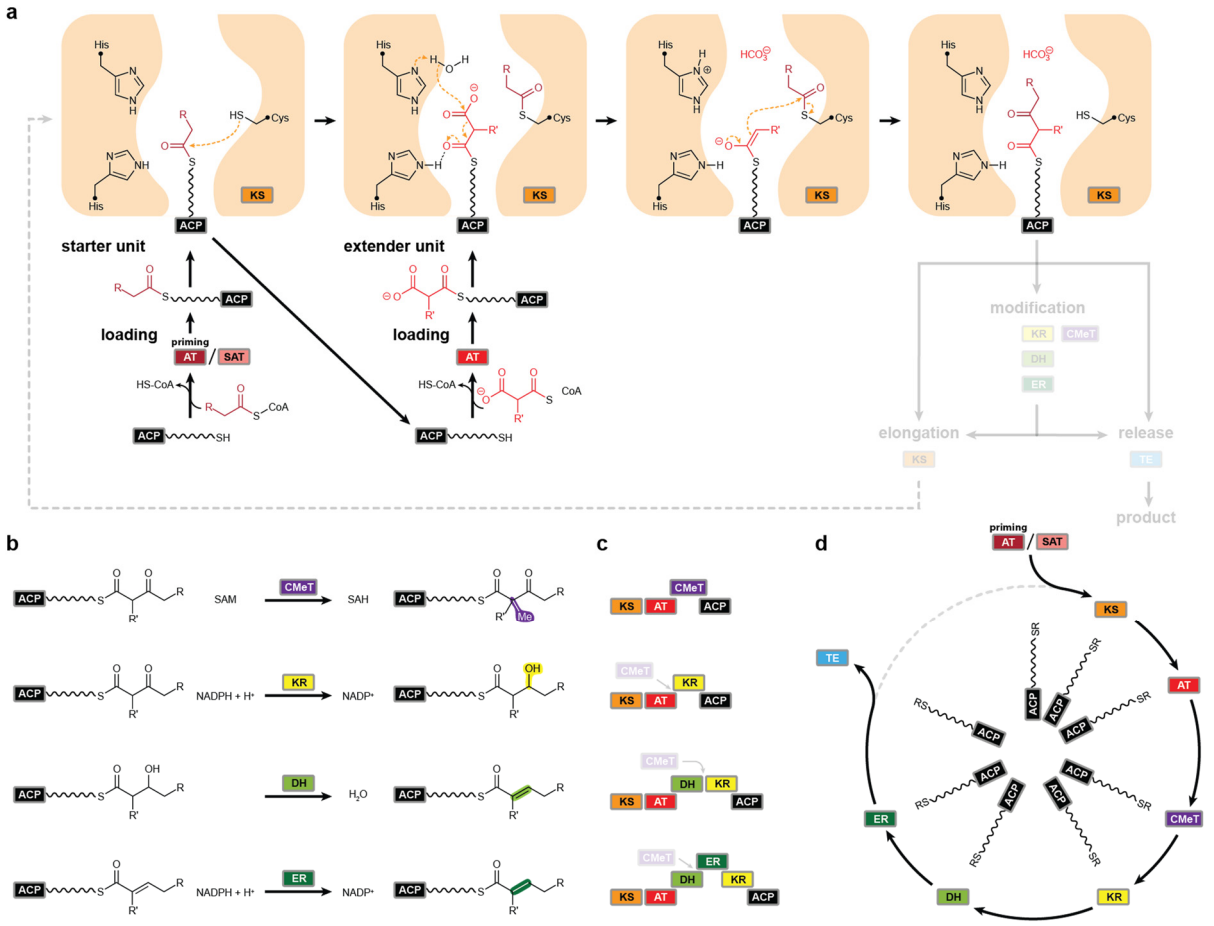


Figure 1.6 | Condensing and modifying reactions in canonical type I PKS.
a, Schematic representation of a condensing reaction catalyzed by KS, AT(s) and ACP. For initiation of polyketide biosynthesis, a priming substrate has to be loaded onto ACP by an SAT or loading AT, which is subsequently transferred to the KS active site. In the same way all other extender units are loaded, but a different AT might be used. Once both substrates are enzyme tethered and bound to the KS, chain elongation is catalyzed by a decarboxylative Claisen-like condensation. The product can either be modified, released or transferred back to the KS active site cysteine for another round of elongation. Transition states of the KS reaction are indicated in orange boxes. The opening at the top indicates an extended substrate tunnel, which can accommodate the growing polyketide in later cycles. Electron density shifts are indicated as dotted arrows (orange). Substrates are colored according to their originating AT. Dots in the KS active site residues indicate C_α atoms of the protein. The Ppant cofactor of ACP is shown as curly line. **b**, Main reactions catalyzed in PKS modifying regions are shown. The reactions are sorted in sequential order. **c**, Domain organization (KS at the N-terminus; ACP at the C-terminus) of modules producing the substitution patterns of (b), ranging from non-reducing (top) to fully reducing (bottom). All modules have a mandatory condensing region (baseline) and differ in the optional modifying region (indented). CMeT domains are rarely observed, but can exist in all combinations with other modifying domains. Corresponding insertion sites are indicated (transparent). **d**, Schematic representation of a reaction cycle containing all domains described above. The central ACP domain transports the substrates to all active sites. Initiation (priming AT / SAT) and termination (TE) steps are indicated. Reactions can occur iteratively.

bond by water elimination followed by an NADPH dependent reduction by an Enoylreductase (ER) domain. The product is a fully saturated carbon-carbon bond that can either be elongated in another cycle of condensation and modification or released from the ACP by a thioesterase (TE), a TE/Claisen-cyclase (TE/CLC), or a reductase (R) (Figure 1.6d). The substitution pattern is mainly controlled by the presence of these domains, thus a modifying region containing a KR produces a β-hydroxy group, a DH-KR an α-β double bond, and a DH-ER-KR a fully saturated carbon-carbon bond. If none of these domains is present, the product stays an unreduced (poly-)ketone. The fourth domain is a C-methyl-

transferase (CMeT), which is capable of SAM dependent methylation at the α-position, but only before ketoreduction[66-68]. Although this domain can occur in all of the previously mentioned combinations, it is rarely found in PKS containing AT domains as integral part of the multienzyme (*cis*-AT PKS, chapter 1.5.3), because methyl branches are usually incorporated via branched extender units like methyl-malonyl-CoA[69]. If a *cis*-AT PKS contains a CMeT domain, it is often inactive (pseudo-CMeT, ΨCMeT) or serves a regulatory function in programming substitution patterns in fungal iterative polyketide biosynthesis[70].

All condensation and modifications are under precise stereocontrol[71]. If an α-substituted extender unit is loaded, such as methyl-malonyl CoA, the configuration will be inverted during the condensing reaction[72]. In the modifying region the domains either produce stereoisomers as part of the main reaction[66,73] or designated isomerases alter stereochemistry. Studying stereocontrol in PKS can be challenging, because often it is unclear which domain creates a stereocenter and sometimes the center gets lost in the process of sequential modifications. Since many KR and DH domains did not contribute to the expected substitution pattern of the product, they were considered as inactive for a long time (KR^0, DH^0). Meanwhile, it was shown that these "inactive" domains actually are isomerases and epimerases, respectively[74,75], and work on intermediate products. For DH domains further variations are known, such as B domains (branching) and PT domains (product template, chapter 1.5.2), which have a similar modular integration and a common fold with DH domains.

A selection of these domains in one giant polypeptide creates a functional unit for the production of one polyketide, named module (Figure 1.6c). Each module contains at least one ACP, which shuttles the substrates to all catalytic domains. In some cases, more than one ACP can be found in series, which is supposed to increase the net-flux of substrates[48]. A module can either catalyze condensing and modifying reactions iteratively (chapter 1.5.2) or in cooperation with many other modules in a molecular assembly line, named modular PKS (modPKS, chapter 1.5.3).

1.5.2 Iterative type I polyketide synthases (iPKS)

Representatives of all PKS from type I to III make repeated use of the same domains (Figure 1.6)[55]. While all known members of type II and III superfamilies are iPKS, type I splits into seven groups, whereof five have an iterative mode of action[76]: Polyunsaturated fatty acid synthases (PUFAs)[77], endiyne synthases[78,79], polycyclic tetramate macrolactam synthases (PTMs)[80,81], mycocerosic acid synthase like PKS (MAS-like PKS) and fungal iPKS. The latter group comprises the best studied family and is described in detail next.

Fungal iPKS are differentiated by the degree of reductive domains in their modifying regions in highly reducing (HR-PKS), partially reducing (PR-PKS), and non-reducing (NR-PKS) (Figure 1.7a). Their name might suggest a uniform substitution pattern of the products, like the production of fatty acids by HR-PKS, but the products reveal non-uniform substitution patterns. This is caused by selective modifications in every iteration in a programmed manner. The precise mechanism of programming is still enigmatic, but recent publications on HR-PKS support a kinetic substrate competition model[70,82] and an involvement of a CMeT domain. For the CMeT of the lovastatin (Figure 1.5) HR-PKS LovB a gatekeeping function in programming was suggested that specifically binds and methylates only the natural intermediate of a central iteration cycle[70]. In case of a wrong modification the domain would not methylate the substrate and cause off-loading reactions.

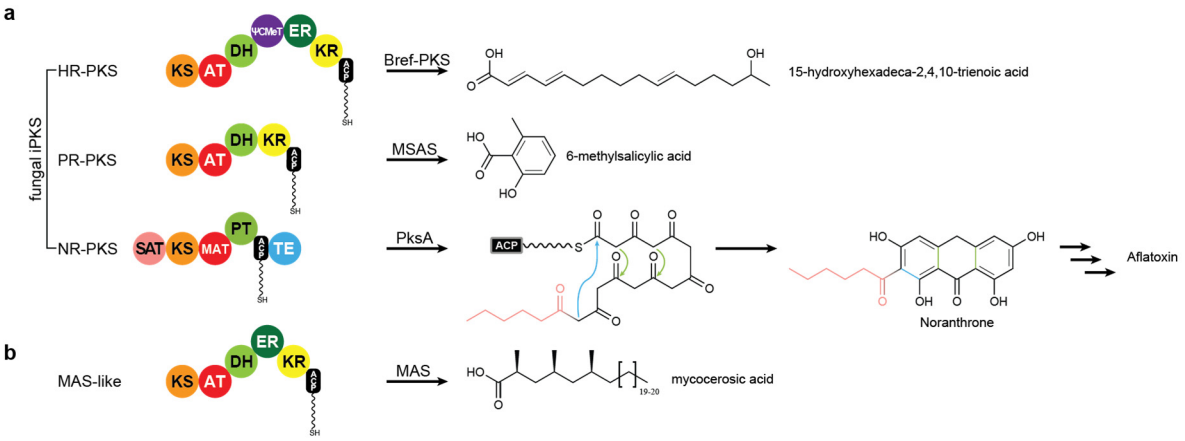


Figure 1.7 | Organization and examples for fungal iPKSs and MAS-like PKS.
a, Fungal iPKS with the modifying domains DH, (Ψ)CMeT, ER or KR reveal programming. Bref-PKS: PKS of the Brefeldin A (protein transport-inhibitor) biosynthesis pathway[83]; MSAS: 6-methylsalicylic acid synthase[84]; PksA: PKS of the aflatoxin (toxin) biosynthesis pathway[85]. For PksA cyclization reactions are indicated with arrows colored according to the corresponding catalytic domain. SAT and PT domains are unique for NR-PKS. The starter substrate moiety is colored according to the SAT domain. **b**, Fully reducing mycobacterial MAS-like PKS. Domains of modifying regions are indicated by indentations.

HR-PKS contain a full set of reductive domains and often include a CMeT or ΨCMeT domain (KS-AT-DH-[Ψ]CMeT-ER-KR-ACP. PR-PKS are the least studied class. They lack an ER domain (KS-AT-DH-KR-ACP), reveal a similar programming mechanism and are proposed to form homotetramers[64,84], like the 6-methylsalicylic acid synthase (MSAS)[86]. NR-PKS produce aromatic products often containing several anellated rings systems. They have no canonic modifying domain, except for an optional CMeT. Instead, they have additional SAT and PT domains that are unique for this class of iPKS (SAT-KS-MAT-PT-ACP-[CMeT]-TE/CLC). While the MAT specifically loads malonyl-CoA as extender units, the SAT loads a specific starter, which can be a CoA ester or the ACP tethered product of a designated FAS[87-89] or (HR-)iPKS[90]. The latter one represents a remarkable mechanism, because it creates an assembly line of two independent iterative multienzymes and represents a new “relay race” like concept of modularity in analogy to canonical modPKSs (chapter 1.5.3). The loading/condensing region consisting of SAT-KS-MAT is fully active *in vitro* and produces unreduced polyketides with a chain length that is determined by the KS[91]. These polyketides are highly reactive and capable of spontaneous cyclization and aromatization. In complete NR-PKS this is controlled by the PT domain, which provides a cavity for folding of the polyketide and its subsequent cyclization[92]. In some cases, two instead of one ACP can be found, which have been shown to be both fully active in the NR-PKS WA[93]. Interestingly, the ACPs as well as the entire loading/condensing region (SAT-KS-MAT) can be functionally swapped with other NR-PKS[48,91] in deconstruction experiments[85], indicating common interaction mechanisms. TE/CLC domains control the substrate release and serve an editing function that off-loads stalled products[48,94]. Representatives of this class are the aflatoxin (PksA) (Figure 1.7a), cercosporin (cersposporin biosynthesis gene 1, CTB1, chapter 4) and the Citrinin (PksCT, chapter 3) NR-PKS.

Another group of iPKS that are phylogenetically closely related with fully reducing modular PKS (chapter 1.5.3) (KS-AT-DH-ER-KR-ACP), but distant to fungal HR-PKS[44], comprise mycobacterial MAS-like PKSs[95](Figure 1.7b). Mycobacteria are known for the production of very long and branched chain fatty acids as part of their outer cell envelop[51,96] and stand out beyond microorganism for their collection of almost any kind of ACP/PCP-dependend multienzymes[97]. Most of the complex mycobacterial fatty acids are produced by fully reducing PKS, whereof the prototype is the mycocerosic

acid synthase (MAS). MAS produces mycocerosic acids (Figure 1.5), which are the main component of dimycocerosate and contribute to cell wall impermeability[95,98], inhibit native immune response[99-101], and determine lipid-specific replication of Mycobacterium tuberculosis in lung tissue[102]. The group of MAS-like PKS comprises seven PKS, which are involved in different and not yet completely characterized pathways. High sequence identities of approx. 60 % indicate a high structural conservation, but also cause false annotations in databases, as exemplified by the MAS-like PKS5, which is involved in lipooligosaccharide (LOS) biosynthesis[103,104] (chapter 2.9).

1.5.3 Modular Polyketide synthases (modPKS)

While some iPKS modules can work in collaboration with other iPKS or FAS modules (chapter 1.5.2), modPKS bring this to perfection. These PKS are molecular assembly lines consisting of several modules that literally work like Fords assembly line[105]. Every module catalyzes one cycle of condensation and modification before passing on the substrate to the next module in a highly coordinated fashion (Figure 1.8a). In many cases several modules are part of the same polypeptide, which can be as large as 25,572 aa (2.7 MDa) containing 11 modules and 71 domains per monomer (largest in Uniprot: Q49HL2). In most cases, one polypeptide contains between one and six modules and extents the assembly line by connecting additional modules via docking domains[46,106-111] (Figure 1.8, chapter 5.5). A programming as observed in fungal iPKS is usually not observed, which creates colinearity between the modules in the assembly line and the catalyzed reactions. In many cases, this colinearity is even extended to the genome level with open reading frames (ORF) organized collinear to the docked assembly line, as for instance observed for the 6-deoxyerythronolide B synthase (DEBS)[112,113], the prototype of all modPKS (Figure 1.8a, Figure 1.5). This property allows prediction of the polyketide substitution pattern on the DNA level, because every module catalyzes exactly one elongation and modification cycle (Figure 1.6). As soon as chemists realized this remarkable property, the field of combinatorial biosynthesis (chapter 1.5.4) was born and scientists all around the world tried to modify these machines. However, nature proved to be more complicated and most of the attempts failed, mostly due to the lack of structural information on the architecture, dynamics and substrate transfer mechanism.

Later, exceptions to the “rule of colinearity” were discovered, although they represent a minority of all modPKS. For instance the discovery of the gene clusters of the rapamycin PKS[114] (Figure 1.5) and the PKS producing avermectin (AVES)[115,116] (Figure 1.5) show that colinear organization of the ORFs in the genome is no requirement. More challenging to understand are the loss of colinearity by aberrant (*e.g*. in the epothilones modPKS[116,117]) and programmed module skipping (*e.g*. in the pikromycin modPKS[118]) or modules that catalyze more than one cycle[116,119-121] as shown by module five of the borrelidin PKS, which catalyzes three cycles of fully reducing polyketide extension[122]. Even for DEBS could be shown that aberrant iterations of the fully reducing module four (Figure 1.8a) are possible[123], which was called stuttering. Later, domain and module truncation experiments showed that the downstream KS of an iterative module can select for correct substrates[124] and swapping of an N-terminal part of helix I from DEBS ACP3 to DEBS ACP2 can induce iteration of module three[26]. Furthermore, it was shown that the Pikromycin PKS module five (PikAIII) can iterate in *in vitro* experiments, if it lacks its assembly line environment[118]. Recently, a “turnstyle” mechanism was identified, which prevents loading of a module’s KS from an upstream ACP, before the product is transferred to a downstream module in DEBS[125]. In summary, this creates a picture of kinetic control mechanisms that usually ensure compliance to the commonly observed colinearity and indicates that there are no fundamental differences between fully reducing modPKS and *e.g.* MAS-like PKS.

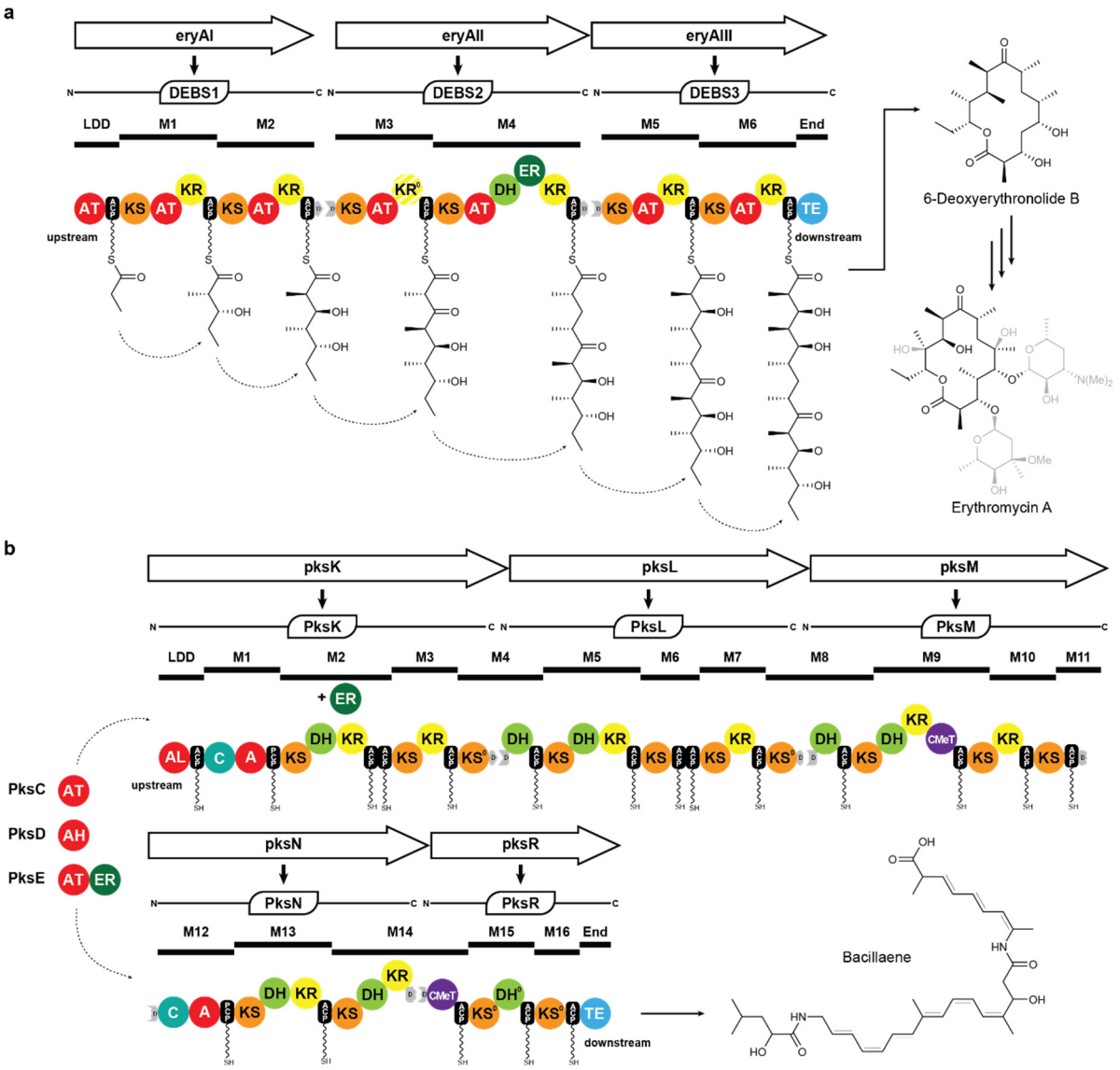


Figure 1.8 | Colinear reaction sequences in *cis*- and *trans*-AT modPKS.
a, Schematic representation of the *cis*-AT PKS 6-deoxyerythronolide B synthase (DEBS) showing a modular organization colinear from the genes in the genome (large outlined arrows) to the biosynthetic reactions sequence. The modPKS is encoded by three sequential ORFs in the genome of *Saccharopolyspora erythraea*. Once expressed, the proteins DEBS1-3 dock in the same order as encoded in the genome via docking domains (D/D) and form an assembly line with six modules (M1-6). The product is sequentially synthesized and passed from module to module until it is released. All modules in *cis*-AT PKS have one AT domain per module. Domains of modifying regions are indicated by indentations. Products of modules are shown tethered to ACPs. (LDD: loading didomain; KR^0: inactive KR with epimerase activity). **b**, Same schematic representation of the *trans*-AT PKS producing bacillaene in *Bacillus subtilis*. In contrast to *cis*-AT PKS, these assembly lines lack AT domains as integral components of each module. Extender units are loaded by two discrete AT domains (PksC and PksE). Loading is proofread by an AT with predominant hydrolase activity (AH: Acylhydrolase). The bacillaene modPKS is a hybrid of NRPS and PKS modules (see also chapter 5.5). Module one and twelve load an amino acid (A: NRPS adenylation domain) and condense (C: NRPS condensing domain) it with the polyketide. For reasons of space the ACP tethered intermediates are not shown. (AL: acyl-ligase; DH^0: inactive DH with isomerase activity; KS^0: inactive KS)

In analogy to SATs in NR-PKS, modPKS load a priming substrate via a designated loading module, which involves an AT-ACP didomain and in case of DEBS is called loading didomain (LDD). In some cases, these modules are equipped with specialized KS domains with a glutamine (or serine) mutation of the active site cysteine (KS^Q) that only decarboxylates a malonyl-CoA derivative that was previously loaded via

the AT. In some cases, these domains can occur as discrete domains and sometimes even NRPS adenylation domains are involved in loading an amino acid primer on an ACP[126]. The last module usually contains a TE, which releases the product or transfers it to tailoring enzymes. The modules in the middle most likely have evolved by gene duplication and recombination events[41] and recently it was shown that the KS domain coevolves with its upstream modifying region[127]. Based on this observation, a new evolutional module definition was proposed in which a module consists of AT-[modifying region]-ACP-KS instead of its [condensing region / KS-AT]-[modifying region]-ACP (conventional definition). Nevertheless, this definition might match evolution parameters, but is not applicable to other stand-alone PKS such as iPKS, which is why the proposed evolutional definition of a module will not be used in this thesis.

ModPKS that contain AT domains as integral part of every module (*cis*-AT PKS) have been extensively studied for decades. More recently, a second group of modPKS has been described that lack AT domains as part of their polypeptides (Figure 1.8). These assembly lines have one or few discrete AT domains that load substrates to all modules in *trans* and therefore are named *trans*-AT PKS[69,128]. AT domains of this group show high substrate specificity[129] and sometimes a second AT domain serves a proofreading function (AH, acylhydrolase)[130], which causes reduced extender unit flexibility. While *cis*-AT PKS use AT domains to incorporate acyl branches, *trans*-AT PKS compensate this with CMeT domains as part of their modifying regions[69]. Compared to *cis*-AT domains, *trans*-AT PKS appear more often as hybrids with NRPS and reveal a much larger repertoire of modifying domains, including Michael branching domains (B)[131], O-/N-methyl transferases and oxidases[69].

1.5.4 Combinatorial Biosynthesis

The discovery of PKSs and particularly their modular representatives generated huge excitement in the field of natural product research. PKS are doing not only incredible chemistry at high efficiency, but also use a strategy analog to Henry Ford's assembly line[105]. Enzymatic modules stagger in an assembly line and produce a specific polyketide, which is mostly defined by their order and domain composition. In principle this should not only allow prediction of products, but also their rational design by genetic manipulation of the corresponding PKS. Manipulation strategies include domain, module and modifying/condensing region swaps, module shuffling, precursor feeding, mutasynthesis, directed evolution, and pathway level combinatorial biosynthesis. Some of these attempts were successful, such as AT and modifying region swaps of DEBS with the rapamycin PKS[132] or SAT swaps in NR-PKS yielding new aromatic compounds[133]. Co-expression of random pairs of sequentially acting fungal iPKS[134,135] created new functional quasi-modPKS and novel products[136,137]. Various forms of so called mini-DEBS have been engineered[138-140] and remarkable high yields could be obtained by TE domain swaps and relocations in the assembly line[141]. Mostly, these approaches were not of pharmacological relevance, but they were of high value for functional studies and our understanding of PKS function. To the best of my knowledge, there is only one drug which was derived via combinatorial biosynthesis and made it to the market: Doramectin (Dectomax®) is the most effective antiparasitic drug in veterinary medicine and was generated by swapping the highly unspecific LDD of the avermectin modPKS (AVES1) with the cyclohexan carboxylic acid specific LDD of the phoslactomycin modPKS[141,142].

High throughput screening methods involve mutasynthesis and directed evolution. In mutasynthesis, precursor producing enzymes are inactivated by mutation and substrate analogs are screened in feeding experiments[143]. Directed evolution combines swapping or deletion experiments with random

mutagenesis and *in vivo* screening for novel products[144]. Beside all approaches that target PKS directly, other approaches have been developed that target the PKS products. Many polyketides have poor bioactivity, are insoluble or cytotoxic. To overcome this problem, many natural producers employ pathways that apply post processing to the polyketide product (*e.g.* glycosylation) (Figure 1.8a). This is addressed by pathway level combinatorial biosynthesis and includes approaches like coexpression of "sugar plasmids"[145] or chlorination biosynthetic machineries[146].

However, the progress in this field is relatively slow and proofed to be more challenging than initially thought. The main challenges in combinatorial biosynthesis are protein insolubility, impaired activity, reduced yields[141], the lack of appropriate expression hosts, size limitations of classical cloning strategies, limited knowledge about tailoring systems, and an incomplete understanding of protein folding and kinetics.

1.6 The architectures of PKSs

1.6.1 Mammalian FAS as prototype for PKS architecture

The discovery of PKS is closely related to that of FAS. While polyketide chemistry was discovered more than a hundred years ago[147], most of the insights in their basic biosynthetic strategies were discovered in the field of fatty acid biosynthesis between 1950 and 1980[148]. The catalytic activities in FAS and PKS are highly related and even the concept of type I and II synthases was originally adapted from FAS[148,149]. Shortly after the raise of genetic methods in the 1980s the first genes of PKS were identified[150] in 1984, followed by the characterization of DEBS[112](Figure 1.8a). The domain organization of mammalian FAS (mFAS) closely matches the organization of the fully reducing DEBS module four (Figure 1.9a) with a sequence identity of 23 % and all other modules appear to be smaller variants of mFAS with domain deletions in the modifying region. mFAS is an active dimer, which can dissociate into monomers once exposed to low temperatures and low salt conditions[151,152]. This feature was employed for functional studies in mutant complementation assays for both enzyme families[153-155], although the temperature and salt dependency was not documented in experimental procedures for PKS. In 2001 the age of structural biology on FAS and PKS begun with an initial cryo electron microscopy (cryo-EM) reconstruction of the human FAS, which was misinterpreted as antiparallel dimer in a head-to-tail organization[156]. Later crosslinking experiments showed that mFAS dimerizes via the N-terminal KS domain and approved the correct head-to-head model[157] (Figure 1.9b), which roughly agreed with a parallel double helical model that was proposed for DEBS eight years earlier[158]. Shortly afterwards cryo-EM reconstructions provided first architectural shapes of mFAS in agreement with the head-to-head model[159]. In the following year, the crystal structure of mFAS was determined at 4.5 Å resolution, revealing detailed insights in the domain architecture and indications for a highly dynamic structure[160] in agreement with previous mutant complementation assays[153](Figure 1.9c). However, linkers and the identity of an inter domain (ID) region of approx. 650 aa between AT and ER could not be resolved. One month later, the structure of the KR domain of DEBS module one revealed a non-functional KR domain[161], which was later named pseudo-KS (ΨKR)[162]. This indicated that the ER is integrated between a ΨKR and KR domain (ΨKR-ER-KR) in both enzymes. Afterwards the structure of the condensing region of DEBS module five[163] and three[164] were solved, that closely matched the mFAS architecture. The structures showed that the KS and AT domains are bridged by a small 63 aa long linker domain (LD) and a structural highly integrated AT-DH (post-AT) linker that entangles with the LD and is embedded in a surface groove of the KS domain. In 2008, the crystal structure of mFAS could be improved to 3.2 Å resolution and finally revealed the complete architecture of mFAS, without the flexibly attached ACP

and TE, including all linkers and a characterization of the ID region[162](Figure 1.9c, d). Besides a similar ΨKR domain as observed in DEBS KR1, this region contained another inactive domain that was identified as partially deleted pseudo CMeT domain (ΨCMeT). This domain and its location in sequence (DH-ΨCMeT-ΨKR-ER-KR) represents yet another similarity between mFAS and *cis*-AT (i)PKS, while CMeT domains in PKS were mostly identified after the structural characterization of mFAS. Inactive ΨCMeT domains in PKS remain largely unrecognized due to very low sequence conservation, deletions and the lack of a functionally active PKS reference structure that does not contain large scale deletions. The previous observed parallel low resolution H-like shape of mFAS turned out to be X-shaped with central dimer interfaces on KS, DH and ER. The two principle regions comprise the condensing and modifying region that are connected via a supposedly flexible linker, which highlights that the previous functional separation is also reflected in architecture.

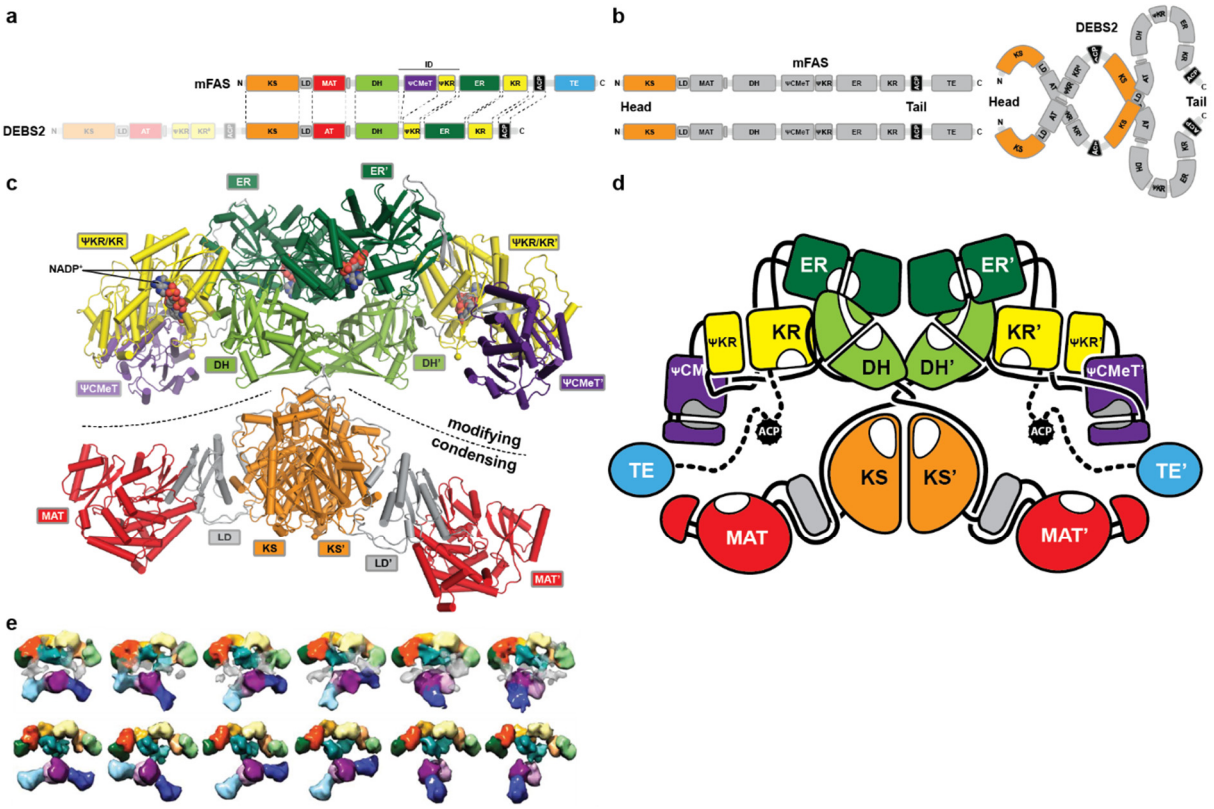


Figure 1.9 | mFAS is the prototype for PKS architecture.
a, mFAS and PKS modules (here DEBS module three and four) share a similar domain organization. PKS modules can have less domains per module, but still maintain the same arrangement in sequence. **b**, mFAS has a head-to-head arrangement. A similar double helical model with equivalent head-to-head orientation was proposed already a few years before the mFAS model was proposed. KS and ACP domains are colored in order to indicate the direction. **c**, Crystal structure of mFAS at 3.2 Å resolution (2008, PDB: 2VZ9). The structure reveals an asymmetric X-shaped dimer with clear separation in the condensing and modifying region. The modifying region reveals direct domain-domain interfaces. ACP and TE domains remained unresolved. **d**, Schematic representation showing the domain connectivity and the location of the unresolved ACP and TE in a resting position. **e**, Random conical tilt reconstructions (upper row) and simulated band-pass filtered (21-330 Å) map of fitted domains show high conformational variability. The condensing region can rotate by up to 90° (distinguishable by the model) and swivel laterally by up to 25°. (violet: KS, blue: MAT, turquoise: DH, red: ΨKR/KR, green: ΨCMeT, yellow: ER). Panel reprinted from Brignole et al. (2009)[165], Copyright 2009, with permission from Elsevier.

On the individual domain level, structural differences to their type II counterparts are most evident with regard to the oligomeric state. While the oligomeric state of KS and AT is equivalent to that found

in the conserved condensing region of mFAS, domains of the modifying region are different. Functional DH domains in type II FAS are dimers consisting of two hot-dog folds and two active sites. DH domains in mFAS are also dimers, but each domain consists of two (double) hot-dog folds with one active site. Interestingly, a similar fusion is also observed in Sfp, a PPT of another multienzyme pathway (chapter 1.3). Type II ER domains are either NADPH-dependent tetrameric short-chain dehydrogenases/reductases (SDRs) or FMN dependend dimeric a TIM-barrel (triosephosphate isomerase fold) proteins, whereas the mFAS ER has an NADPH-dependent medium-chain dehydrogenase/oxidoreductase (MDR) fold. All KR domains are NADPH-dependent SDRs and tetramers in type II FAS, but pseudo dimers of ΨKR/KR in mFAS. Notably, the dimeric unit of a type II tetramer shares the same interface with the ΨKR/KR pseudo dimer and represents again another example for domain fusion in multienzymes. In comparison with type II FAS, this shows that most variations are observed for the domains of the modifying region that determines the substitution pattern, while the domains of the condensing region reveal higher topological conservation. Structural differences to PKS domains will be discussed in detail in chapter 1.6.2.

Based on the conformational differences in the mFAS structure, disordered ACPs and TEs, and active site distances too far apart from each other to be overcome by exclusive motions of the ACP Ppant cofactor ("swinging arm", chapter 1.2) a highly dynamic architecture was proposed. An earlier single particle cryo-EM study in 2005 confirmed that the conformational heterogeneity of mFAS could be reduced by addition of substrates and active site mutations of KS, KR and TE[159]. In order to analyze the conformational variability in more detail a similar approach was chosen by Brignole and Asturias in 2009[165]. Two active site (KS and DH) as well as a deletion mutant in the ACP-TE linker were generated and analyzed by random-conical tilt as well as single particle cryo-EM. Multivariate statistical analysis as well as hierarchical ascendant classification yielded classes in defined conformations for each mutant with and without substrate[166]. The reconstructions delineated the conformational variability during a reaction cycle (Figure 1.9e) and demonstrated flexibility of the mFAS architecture with a hinge between the modifying and condensing region. Interestingly, the data agreed with previous mutant complementation assays that could only be explained by 180° rotations of these two regions[153]. However, the relevance and extent of these conformational states for PKS remain unknown, since similar complementation assays of a mini-DEBS (DEBS module 1+2+TE, chapter 1.5.4) did not agree with rotations of up to 180°, but also did not exclude large scale conformational changes in this region[154].

In summary, mFAS and PKS are two families of multienzymes that share principles of their structural organization and biosynthetic logic. mFAS serves as a prototype of PKS organization as long as structures of complete PKS modules remain unknown.

At this point it is noteworthy, that a second form of type I FAS exists in fungi and mycobacteria. The first to be characterized was the *saccharomyces cerevisiae* FAS (yeast FAS, yFAS) and forms a 2.4 MDa heterododecameric assembly with a barrel like shape that accommodates three ACPs in the inside of each half of the barrel[167]. This barrel like structure is constructed by up to 50 % of protein scaffold as compared to only 9% in mFAS (Figure 1.10). PKS with a similar organization have not been observed, although it is known that some yFAS analogs can deliver precursors to NR-PKS as exemplified by HexS (HexA/HexB), which delivers hexanoic acid to the SAT of the NR-PKS PksA[87-89,168] (Figure 1.7).

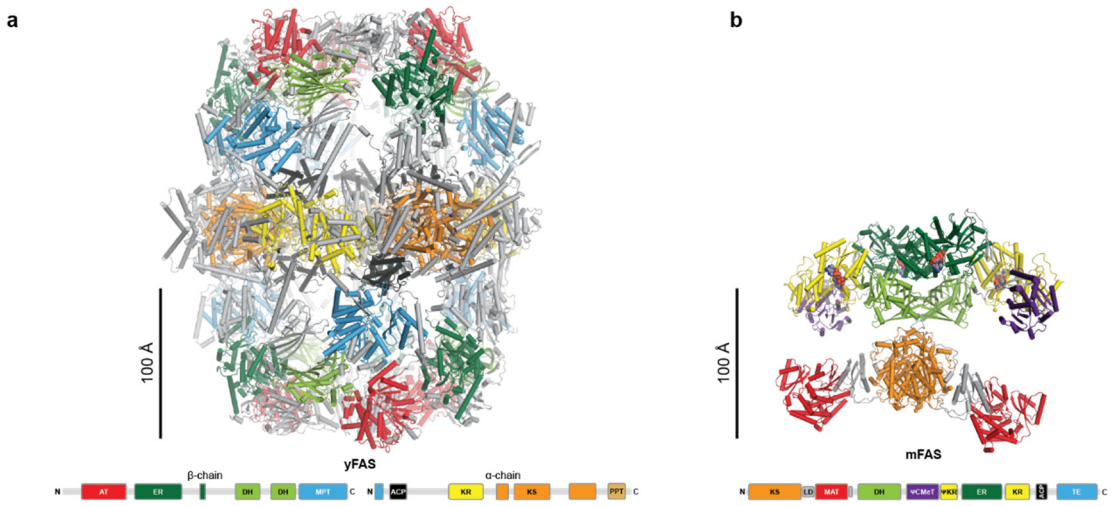


Figure 1.10 | Structural comparison of yFAS with mFAS.
a, Heterododecameric structure of yFAS. **b**, Dimeric structure of mFAS. Sequence representations are shown below the structures. Domains are colored according to the sequence representation. In yFAS a malonyl/palmitoyl transferase (MPT) is responsible for extender unit loading and off-loading of fatty acid products as CoA ester. yFAS architectures are unknown for to occur in PKS.

1.6.2 Structures of PKS

Already before the structure of mFAS was solved, structures of PKS KS (type II), TE, ACP and a docking domain[106] (DD; later three classes were characterized[46,106-109,111]; chapter 5.5) were known. TE domains of mFAS and NR-PKS are monomers, whereas modPKS TEs have been crystallized as dimers[17,169-171], which was a first sign that PKS and FAS might have different architectures apart from the conserved condensing region.

At the beginning of this thesis three *cis*-AT PKS[46,163,164] and two mFAS[162,172] condensing regions were determined that share virtually the same organization (Figure 1.11a). Furthermore, low resolution small-angle X-ray scattering (SAXS) envelops of DEBS module six and a DEBS3 (bimodule) were in agreement with this organization[43]. However, during manuscript preparation of chapter 2, cryo-EM structures of the pikromycin PKS module 5 (PikAIII), containing the smallest modifying region with only one KR, were solved at 7.3-9.5 Å resolution by cryo-EM[42,45] (Figure 1.11b). Compared to the PKS condensing regions solved by X-ray crystallography (approx. 50 % sequence identity) and almost the same linker length (diverging by 1-3 aa), the PikAIII structure has an entirely different domain organization. The AT and LD domains are rotated by more than 90° relative to other PKSs condensing regions and reveal a symmetric organization of inverted handedness (Figure 1.11), which was confirmed by tilt pair validation[42]. Fitted models of DEBS condensing region structures reveal severe clashes and overlaps, respectively, and the connecting linkers are too short to explain the EM map. Therefore, the authors suggested refolding events. Surprisingly, also an alternative active site entrance could be identified, which is only accessible if the post-AT linker dislocates from its highly embedded surface groove on the KS and LD entanglement in order to connect with the KR. However, the limited resolution precluded the identification of linkers. Furthermore, 180° flips of the KRs and ACPs in proximity to all domains could be visualized, although the ACP-domain interactions are most likely not catalytically relevant[173] (see chapter 4). The structure of PikAIII is controversially discussed in the PKS field and its relevance for PKS architecture remains unclear[174].

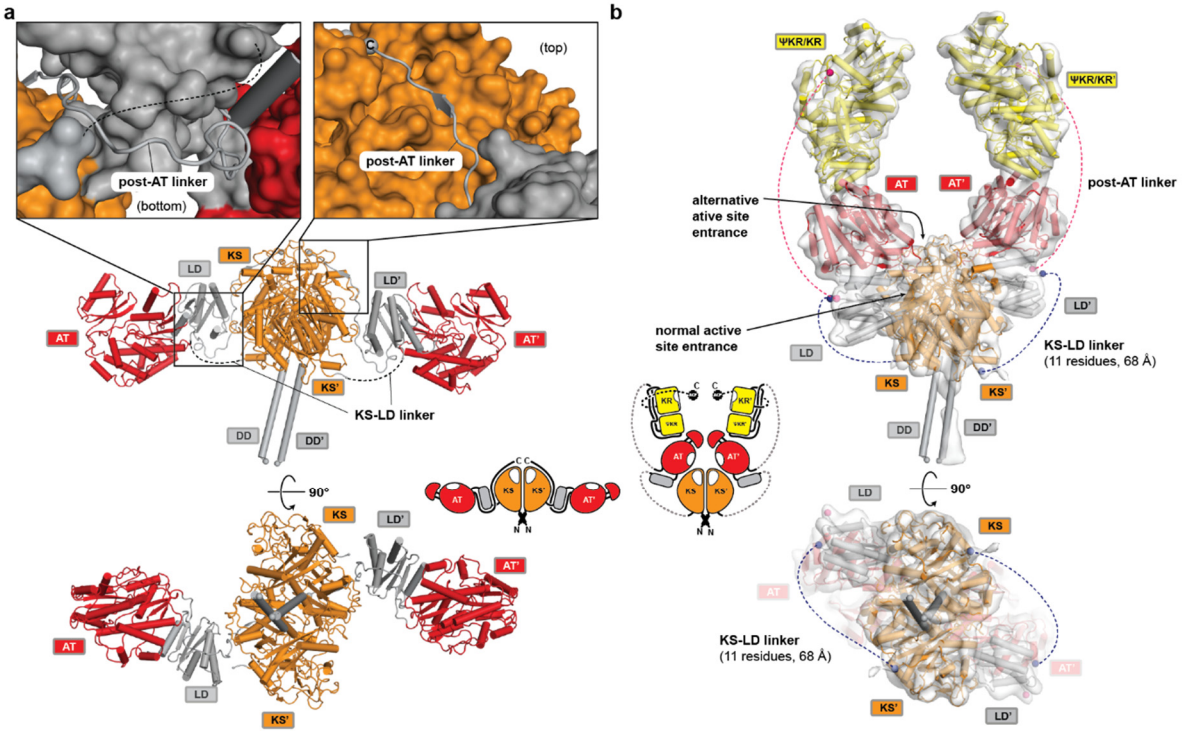


Figure 1.11 | Diverging models for PKS condensing region architecture.
The crystal structure of the DEBS module five condensing region (PDB: 2HG4) represents the conserved architecture of PKS and FAS condensing regions. Side and bottom views are shown. Insets at the top show the tight entanglement of the post-AT linker with the LD and KS. **b**, Cryo-EM structure of PikAIII (EMDB: 5649) at 7.8 Å in side and bottom view as in (a). Domains of DEBS module five (a) and DEBS KR1 (module one, PDB: 2FR0) were fitted into the map (contoured at 5.0 σ) as described by Dutta *et al.*[42]. The structures share sequence identity of 47 % and connecting linkers were not resolved. Relative to (a), the ATs reveal an upward rotation of larger than 90° and the LDs connect to the central KS dimer on the opposite side, which blocks the canonical KS active site entrance. The corresponding KS-LD linker contains 11 residues and must bridge a distance of 68 Å. In order to connect to the ΨKR/KR, the post-AT linker must adopt an entirely different connectivity, free of any KS or LD entanglement as shown in (a). This led to the proposition of an alternative KS active site entrance at the top of the KS, which is blocked by the post-AT linker in (a). Linker and domain connectivity of both architectures are shown as schemes in the center. DD: Docking domain.

More in agreement with existing PKS structures is the observed dissociation into monomers upon deletion of a C-terminal fragment. Similar observations have been made for other PKS condensing regions[43,44,46]. In the context of multiple domain interactions in multienzymes, this provides less obvious, but crucial implications for structural studies. Excision of fragments from larger multi-domain proteins is a common approach to study macromolecular structures. However, the interpretation of the results must always reflect these circumstances, particularly if oligomeric properties are transferred to the full length protein. All of the structures mentioned above are part of multienzymes, which are known to be in a monomer-dimer equilibrium[151-155], and contain interfaces on several domains as part of one polypeptide. In comparison to discrete oligomeric type II PKS and FAS, the domains of type I PKS and FAS are less exposed to evolutional pressure for maintaining stable interfaces, because their oligomeric state relies on avidity. Consequently, the observation of a dimeric fragment in solution might be indicative for the full length assembly, whereas no information can be derived from a monomer.

In contrast to the conserved condensing region of PKS and FAS, more architectural differences are expected for the modifying region due to varying domain compositions. This became clear with the crystal structures of DEBS DH4[175] (module four) and several other PKS DH domains[44,176-179] that adopted

a rather linear dimer than a V-shaped pseudo-dimer as observed in FAS (Figure 1.12a). Later a ΨKR-ER-KR fragment of the spinosyn PKS module two (SpnB ER-ΨKR/KR) was crystallized[47] as partially unfolded domain swapped monomer and analyzed by SAXS. The structure revealed a 180° rotation of the ER relative to the ER in mFAS (Figure 1.12b) and remodeling on mFAS was proposed to be impossible due to short linkers. In combination with a monomeric ER derived from limited proteolysis of DEBS module four[158] and a monomeric crystal structure of a *trans*-acting ER[180], an alternative architecture for fully reducing modifying regions was proposed (Figure 1.12c). However, ER and KR domains never occur without DH domain (Figure 1.6), which contribute another dimer interface and complete the functional and structural unit of a fully reducing modifying region.

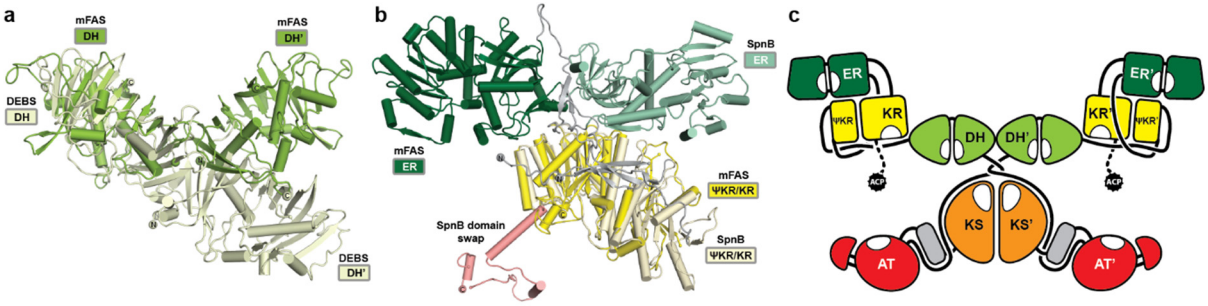


Figure 1.12 | Proposed model for fully reducing PKS modules.
a, The superposition of mFAS DH (V-shape) and DEBS DH4 (linear) illustrates that mFAS and PKS modifying regions must have divergent architectures. **b**, KR superposition of the monomeric SpnB-ER-ΨKR/KR fragment on mFAS ER-KR shows that the ERs are rotated by approx. 180°. The SpnB-ER-ΨKR/KR fragment was crystallized with a domain swap which is shown in red. **c**, Proposed architecture of fully reducing PKS modules according to Zheng *et al.* (2012)[47].

Structural studies of NR-PKS were limited to the individual domains PT[92], ACP[11], and TE[181] of PksA. Recently, the first structure of an SAT domain was published[182], which revealed high structural similarity to regular AT domains. All of these structures were crystallized as monomers, except for the PT domain, which is a dimer and shares structural homology with DH domains. So far, no structure of a multidomain fragment of NR-PKS is known, but it is reasonable to assume that NR-PKS have a similar condensing region as mFAS and modPKS. In reviews, the SAT domain is often assumed to be flexibly tethered to the N-terminus of the KS via a long linker[17], but there is no evidence for that.

1.6.3 Substrate transfer by the acyl carrier protein domain (ACP)

Substrate transport by carrier proteins requires interactions stable enough to engage specifically with a target domain, but also weak enough to release the carrier protein again. Because of this transient nature structures of such complexes are difficult to study. However, a variety of ACP interactions with carrier proteins could be crystallized that provide insights in their interactions with target domains[13,19,25,28,30,31,34,38,39,173,183-186], including the previously discussed AcpS and Sfp (Figure 1.3). Despite the large variety of different structures only the structure of yFAS[186] shows an ACP in complex with an intact multienzyme. Two structures of PKS domains (3-hydroxymethylglutaryl synthase[185]; *trans* acting AT[173]) are available, both belong to a type I pathways, but are crystalized with ACP in isolation. The other structures belong to type II FAS or related pathways. All structures have in common that ACP binding is primarily mediated by charge-charge interactions with positive charges on the target and negative charges on the ACP domain. All structures reveal key residues for ACP interaction, but no universal sequence motif could be identified.

Nevertheless, some of these complexes provide remarkable mechanistic insights in ACP binding. The hexameric type II FAS DH domain FabZ[34] was crystallized as trimer of dimers, whereof each dimer binds exactly one ACP, although there are twice that much active sites available. Based on this structure and various in-solution measurements, a seesaw model was proposed, which mediates ACP release as soon as another ACP binds to the other side of the dimer. Another mechanistically interesting ACP interaction could be shown by site specific crosslinking[187-193] of an ACP to another type II FAS DH (FabA)[13]. Here, a CoA homolog with a reactive head group was synthesized, loaded on the ACP and mixed with FabA for crosslinking. The crystal structure showed that the ACP cavity between helix II and III was closed and NMR titration of ACP revealed that FabA binding mediates the closure and expels the sequestered Ppant with substrate (here the crosslinker) into the active size in a "switch blade" mechanism. For type II ACPs this is a plausible mechanism, because substrate loading experiments via AcpS of various CoA-fatty acid intermediates, revealed substrate dependent conformational changes of ACP helix III.

However, much less structural data is available on type I PKS and mFAS that do not sequester substrates in "switch blade" mechanism[11,14]. The yFAS was crystallized with its ACP bound to the KS, but mostly it was not considered as relevant for PKS and mFAS, because yFAS has a completely different architecture, the ACP contains a four helix extension at the N-terminus and it contains a cavity that might be used for substrate sequestration[186]. Besides yFAS, the cryo-EM structures of PikAIII contain ACPs in various positions, but besides of one (with the KR, EMDB: 5664) none fulfills the distance requirements for a catalytically relevant interaction[42,45]. Besides these structures, most details about ACP interactions in PKS modules were derived from functional assays of DEBS and NR-PKS. For DEBS it was shown that the directional transport along the assembly line is partially mediated by alternating charge-charge interactions of ACP helix I (ACP_n) with the downstream condensing region ($KS\text{-}AT_{n+1}$). Swapping the N-terminus of DEBS ACP3 helix I with DEBS ACP2 induces iteration of the module[26]. Furthermore, loading and site specific crosslinking experiments revealed a "turnstile" mechanism, which prevents that a substrate is transferred from ACP_{n-1} to KS_n, before the substrate of the module was transferred from ACP_n to the downstream KS_{n+1}, which implies that the throughput of an assembly line is kinetically regulated from its end[125]. The last domain in modPKS as well as NR-PKS is usually a TE, which was already shown to regulate non-reducing polyketide biosynthesis and determines success or failure in deconstruction experiments[48]. ACP interactions in NR-PKS appear to be highly conserved, because ACPs, PTs and entire condensing/loading regions can be functionally swapped with other NR-PKS. Based on homology models, mutations and site specific crosslinking, interactions of NR-PKS ACPs have been probed with PT[194] and KS[21] domains. Most interestingly, crosslinking with KS of NR-PKS loading/condensing regions showed a crosslinking ratio of 50 % as observed for the ACP-FabZ interaction.

1.7 Aims of the thesis

Research on PKS is a highly competitive field, due to its high potential impact on drug development and application in synthesis of chemically complex molecules. So far, most functional as well as structural studies relied on dissected fragments that provided first mechanistic and structural insights in PKS architectures. Besides the structure of CMeT and initially SAT domains, all canonical domains of PKS have been structurally characterized.

Aim I: *Elucidate the structure of a fully reducing PKS module in order to retrieve a conceptual model for PKS architecture (chapter 2)*. For this task homologs of DEBS module four have been identified via a "Basic Local Alignment Search Tool" (BLAST) search and a MAS-like PKS with a sequence identity of 31 % was selected for structural studies. Due to high sequence identities of all MAS-like PKS (chapter 1.5.2), this PKS was initially annotated as mycocerosic acid synthase (MAS) in sequence databases, which was later corrected to MAS-like PKS5[195] (chapter 2.9). During the manuscript preparation of chapter 2, the cryo-EM structure of PikAIII[42,45] and a very low resolution SAXS envelop of the DEBS3 bimodule[43] were published, but both did not provide enough detail to interpret a modular architecture based on its linker connectivity.

Aim II: *Reveal the structural basis for programmed C-methylation in NR-PKS (chapter 3)*. C-methylation by CMeT domains was shown to be directly involved in programming modifying reactions in fungal iPKS[70]. The simplest construct to investigate programming involves a fungal iPKS that lacks all modifying domains (KR, DH, ER) except for a CMeT domain and uses methylation to program the length of the polyketide, as exemplified by the citrinin NR-PKS (PksCT).

Aim III: *Visualize an integrated substrate loading architecture of an NR-PKS (chapter 4).* SAT domains are connected to the N-terminus of NR-PKS condensing regions, but their structural integration in the PKS were entirely unknown. Previous studies have shown that the loading/condensing regions (SAT-KS-MAT) and ACPs of NR-PKS can be functionally exchanged (chapter 1.5.2 and 1.6.3), which indicated a conserved architecture and interactions. Furthermore, it was shown that ACPs interact in a 1:2 ratio with KS domains of loading/condensing regions (chapter 1.6.3)[21], which indicated a functional crosstalk. The CTB1 is a representative of these NR-PKS and was selected for structural and functional analysis.

2 MYCOCEROSIC ACID SYNTHASE EXEMPLIFIES THE ARCHITECTURE OF REDUCING POLYKETIDE SYNTHASES

This research was originally published in Nature[44].

Dominik A. Herbst[*], Roman P. Jakob[*], Franziska Zähringer & Timm Maier.

[*]Contributed equally

Nature. 2016 Mar 24;531(7595):533-7.

Herbst, D. A., Jakob, R. P., Zahringer, F. & Maier, T. Mycocerosic acid synthase exemplifies the architecture of reducing polyketide synthases. Nature 531, 533-537, doi:10.1038/nature16993 (2016),

2.1 Letter

Polyketide synthases (PKSs) are biosynthetic factories that produce natural products with important biological and pharmacological activities[55,196,197]. Their exceptional product diversity is encoded in a modular architecture. Modular PKSs (modPKSs) catalyze reactions colinear to the order of modules in an assembly line[55], whereas iterative PKS (iPKSs) use a single module iteratively as exemplified by fungal iPKSs (fiPKSs)[55]. However, in some cases non-colinear iterative action is also observed for modPKSs modules and is controlled by the assembly line environment[116,119]. PKSs feature a structural and functional separation into a condensing and a modifying region as observed for fatty acid synthases (FASs)[162]. Despite PKSs' outstanding relevance, the detailed organization of PKSs with complete fully-reducing modifying regions remains elusive. Here, we report a hybrid crystal structure of *Mycobacterium smegmatis* mycocerosic acid synthase (MAS) based on structures of its condensing and modifying regions. MAS is a fully reducing iPKS, closely related to modPKSs, and the prototype of mycobacterial MAS-like (Msl-)[51,95] PKSs. It is involved in the biosynthesis of C_{20}-C_{28} branched-chain fatty acids, which are important virulence factors of mycobacteria[100]. Our structural data reveal a dimeric linker-based organization of the modifying region and visualize dynamics and conformational coupling in PKSs. Based on comparative small angle X-ray scattering (SAXS), the observed modifying region architecture may be common also in modPKSs. The linker based organization provides a rationale for the characteristic variability of PKS modules as a main contributor to product diversity. The comprehensive arechitectural model enables functional dissection and re-engineering of PKSs.

Each homodimeric PKSs module sequentially elongates acyl-carrier protein (ACP) tethered precursors by the sequential action of an acyltransferase (AT) and a ketosynthase (KS), organized in the essential condensing region (KS-AT). The product can further be sequentially modified by a ketoreductase (ΨKR/KR), a dehydratase (DH), and an enoylreductase (ER)[55,196]. These optional domains form the variable modifying region of PKSs. MAS is a fully-reducing PKS with a complete modifying region (DH-ΨKR-ER-KR). It iteratively elongates linear C_{12}-C_{20} starter fatty acids in one to four rounds with methyl-malonyl-CoA extender units[51] to produce mycocerosic acids. These MAS products form the core of phenolic glycolipids and phthiocerol dimycocerosates, key lipids of the mycobacterial cell envelope[51]. The condensing and modifying regions of MAS are centrally connected by non-conserved linkers, which permit large-scale relative motions in related systems[165]. To obtain a high-quality hybrid model, we divided MAS into its condensing and modifying region, and excluded the flexibly tethered ACP (Figure 2.1a).

Three constructs of staggered C-terminal length were employed to define the length of the condensing region (see Methods). All variants crystallized under the same condition; structure determination mapped the last ordered residue to Glu887. The structure of the most extended variant (1-892) was refined at 2.3 Å resolution (Extended Data Table 2.1a). MAS KS-AT comprises an α/β-fold linker domain (LD) connecting AT to KS (Figure 2.1b). The monomeric condensing region closely resembles those of other PKSs and fatty acid synthases (FASs)[46,162,163], the closest structural homologue at individual domain level is module 5 of the 6-deoxyerythronolide B synthase (DEBS) PKS (Extended Data Table 2.2a). Compared to previous KS-AT didomain structures, the AT domain is slightly rotated towards the C-terminal post-AT linker.

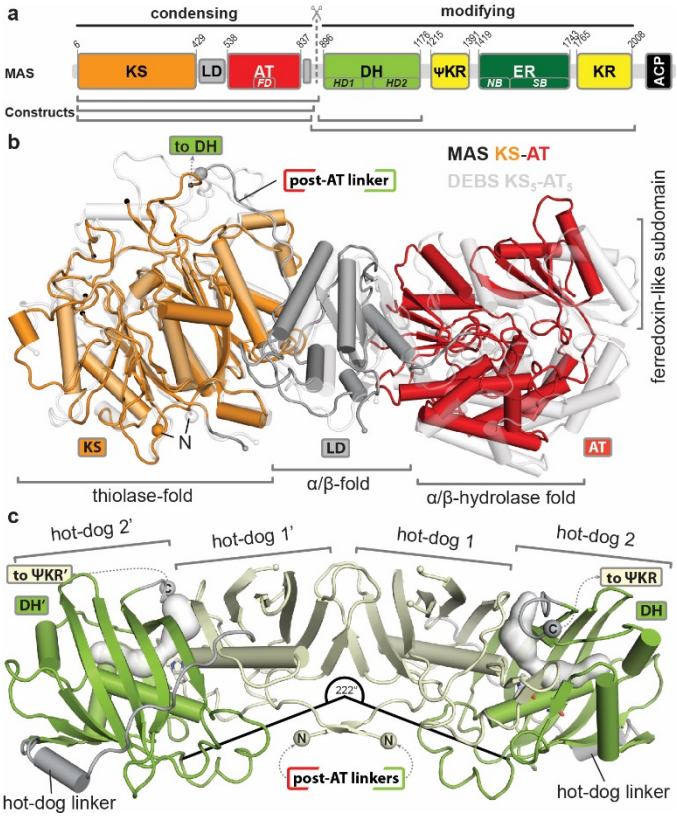


Figure 2.1 | Domain organization, condensing region, and dimeric DH domain of MAS.
a, MAS is organized in a condensing (KS: ketosynthase, LD: linker domain, AT: acyltransferase, AT_{FD}: ferredoxin-like AT subdomain) and a modifying region (DH: dehydratase, $DH_{HD1/2}$: DH hot-dog fold 1 or fold 2, KR: ketoreductase, ΨKR: non-catalytic pseudo-KR domain, ER: enoylreductase, $ER_{NB/SB}$: nucleotide/substrate binding ER subdomain), followed by a flexibly-tethered acyl carrier protein domain (ACP). Crystallized constructs are indicated. **b**, Monomeric condensing region crystal structure. The AT position corresponds to a rotation around a hinge in the LD relative to DEBS KS_5-AT_5^{163} (white). Black spheres indicate ends of disordered segments (aa 47-65, 132-151, 211-220, 277-283). **c**, Crystal structure of the dimeric DH. Each monomer comprises two hot-dog folds connected by a 20 aa hot-dog linker (grey). The DH active site tunnel (white) has two openings and the dimer is bent with an interdomain angle of 222°.

Isolated MAS KS-AT crystallized as a monomer lacking the canonical KS-based dimerization[46,162,163], but is in monomer-dimer equilibrium in solution with a K_d of 0.4 mM as determined by analytical ultracentrifugation (AUC). It is the first condensing region crystallized as monomer, but dissociation has been observed for other condensing region fragments in the absence of dimeric partner domains[42,43,46]. Differences to canonical dimeric KS, as exemplified by DEBS KS_5^{163} or CurL KS[46], are observed around the dimer interface, presumably due to the absence of stabilizing dimer interactions: The interface-spanning active site tunnel is incomplete and the loop containing the catalytic cysteine (Cys178) is bent outwards by 9 Å into a non-productive conformation, while the active site histidines (His313,349) are at expected positions (Extended Data Figure 2.5a-c). Four interface segments of 6-19 amino acid (aa) length are disordered (Figure 2.1b), while equivalent regions are ordered in dimeric KS domains.

A single mode of dimerization based on canonical KS organization was identified by automated sequence-based methods (see Methods) and homology-based modeling of dimeric MAS KS-AT restores the active site tunnel and a productive conformation of Cys178 (Extended Data Figure 2.5b, c). The KS-AT dimer adopts a linear shape due to the rotation of AT relative to KS (Extended Data Figure 2.5d). The C-terminal post-AT linkers of the condensing region, which connect to the modifying region,

are proximal to the two-fold dimer axis above the KS active site, as observed in previous condensing region structures[46,162,163].

The DHs connect the modifying region to the post-AT linkers of the condensing region. We solved crystal structures of a MAS DH construct (aa 884-1186), which overlaps in sequence with the crystallized KS-AT, in two crystal forms with a total of six protomers arranged into almost identical dimers (Extended Data Table 2.1a). The DH protomer is composed of two hot-dog folds connected by a 20 aa hot-dog linker (Figure 2.1c). A hydrophobic substrate binding tunnel extends over both hot-dog folds with entrances near the C-terminus and at the distal end of hot-dog fold 2. Active site residues are contributed by both hot-dog folds and are located close to the C-terminus (Extended Data Figure 2.5e). The nearest structural homologs of DH protomers are modPKS DH domains (Extended Data Table 2.2a). In the DH dimer, the two protomers arrange with their lateral ends bent towards the post-AT linkers with an interdomain angle of 222° (Figure 2.1c). The MAS DH dimer is distinct from the V-shaped DH arrangement in FAS[162], which lacks a dimerization interface and is bent into the opposite direction at an angle of 96°. MAS DH rather resembles linear DH dimers of modPKSs with interdomain angles of 167-203°[175,176,179] and a common mode of dimerization via "handshake" interactions between β-strands of the N-terminal hot-dog folds (Extended Data Figure 2.5f-h).

To obtain an authentic representation of the MAS modifying region, we crystallized in presence of $NADP^+$ the complete DH-ΨKR-ER-KR segment, which is dimeric in solution based on AUC. Based on SAXS, ACP deletion is not affecting the overall structure of this region (Extended Data Figure 2.6a-c). The crystallographic asymmetric unit reveals a complex packing of nine dimers related by non-crystallographic symmetry (NCS). The corresponding 18 polypeptide chains comprise 20,502 aa (2.2 MDa protein mass), of which 17,680 are modeled. Real-space NCS averaging and NCS-restrained refinement led to a high-quality model (R_{work}/R_{free} = 0.23/0.24) at 3.75 Å resolution (Figure 2.2, Extended Data Table 2.1a, Extended Data Figure 2.6d-f). The modifying region dimerizes along an extended interface formed by DH and ER (Extended Data Table 2.2b); the ΨKR/KR is laterally connected to DH and ER. MAS, as well as most reducing modPKSs, lacks a non-catalytic pseudo-methyltransferase domain (ΨME), which is a characteristic of FASs and fiPKSs. The DH in the modifying region adopts the same dimeric structure as in the isolated form (Extended Data Table 2.2a), demonstrating the intrinsic nature of DH dimerization and its role in organizing the modifying region. The ER domain is characterized by a large active site tunnel and a well-ordered $NADP^+$ cofactor (Extended Data Figure 2.7a, b). The ER dimerizes via pseudo-continuous β-sheet formation between the nucleotide binding subdomains (ER_{NB}) and provides the largest contribution to the modifying region dimer interface. Its closest structural neighbors are the isolated modPKS ERs from *Lyngbya majuscula*[198] and the SpnB ER-ΨKR/KR didomain[47] (Extended Data Table 2.2a), even though these ERs are monomeric. The dimerization mode of MAS ER closely resembles those of the ER_{NB} subdomain of the PpsC modPKS and the ER of FAS[162,199] (Extended Data Figure 2.7c). The split ΨKR/KR resembles modPKSs ΨKR/KR (Extended Data Table 2a)[73]; as in related B-type KR domains[200], a flexible lid region (aa 1948-1960) remains disordered in the absence of ligand, and concomitantly, the nicotinamide moiety of $NADP^+$ is disordered (Extended Data Figure 2.7d). The MAS ΨKR exhibits an N-terminal β-α-β-α extension, which is commonly observed in modPKSs, but not in FASs[162,201]; this extension exhibits increased flexibility as indicated by temperature factor distributions (Extended Data Figure 2.7e, f).

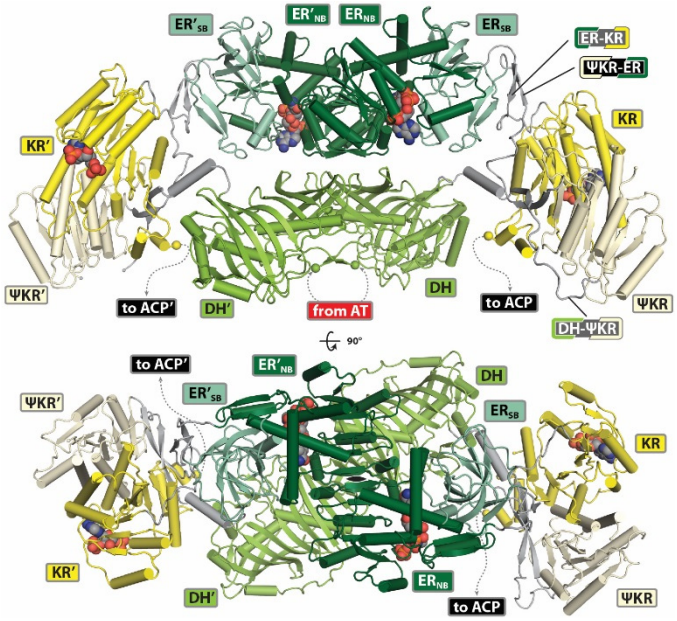


Figure 2.2 | Crystal structure of the dimeric MAS modifying region.
The MAS modifying region is organized by the dimerization of the central DH (light green) and ER (darker greens) domains (upper panel: front view, lower panel: top view). The DH dimer reveals virtually the same bent organization as observed in the crystal structures of the isolated DH domains. The ΨKR/KR (yellow) domains are laterally tethered, share no direct interface with any other catalytic domain, and their positioning is the most variable of all domains. Bound cofactors are shown in sphere representation colored by element type. A two fold dimer axis is indicated in the lower panel.

Previously, modifying region architecture was discussed based on domain interfaces in FAS and PKSs fragments[196]. However, the current analysis of the MAS modifying region reveals a striking absence of stable interfaces between the different domains: The ER dimer rests on a platform formed by the DH dimer, but the interface between the two is small and variable (345-638 Å^2) (Extended Data Table 2.2b, Supplementary Video 1). The ΨKR/KR does not contact its neighboring domains at all and is the region of highest structural variability. Instead, the architecture of the modifying region is based on three linkers interconnecting the ΨKR/KR, DH, and ER domains, which act as spacers as well as interaction partners amongst each other and with catalytic domains (Figure 2.3a, b, Extended Data Figure 2.6d): (i) The 27-aa ΨKR-ER linker plays a central organizing role by forming extended interfaces to ΨKR/KR (975±28 Å^2) and ER (353±20 Å^2). Moreover, it interacts with each of the other two linkers via two double-stranded, antiparallel β-sheets. The β-sheet formed between the ΨKR-ER and DH-ΨKR linkers (B1 in Figure 2.3) is embedded in a surface groove of the ΨKR/KR. It partially extends the Rossman-fold of the KR and is conserved in PKSs (Extended Data Figure 2.8). The β-sheet between the ER-KR and ΨKR-ER linker (B2 in Figure 2.3) mostly interacts with the ER and establishes a gap between the ER and KR. (ii) The 38-aa DH-ΨKR linker comprises an N-terminal 10-aa α-helix ($\alpha_{D\Psi}$ Figure 2.3) followed by the β-strand paired to ΨKR-ER linker and an irregular segment (S1 in Figure 2.3a), which wraps around the ΨKR. Helix $\alpha_{D\Psi}$ separates DH and ΨKR/KR; fragments of it are also observed in structures of isolated DH domains from the Curacin PKS (CurH, K, J)[176]. (iii) The 20-aa ER-KR linker consists of a terminal irregular segment (S3 in Figure 2.3) and the central β-strand paired to the ΨKR-ER linker. It contacts ER and KR via interfaces of 432±24 Å^2 and 547±14 Å^2, respectively, and together with the ΨKR-ER linker forms a continuous connection layer between these domains.

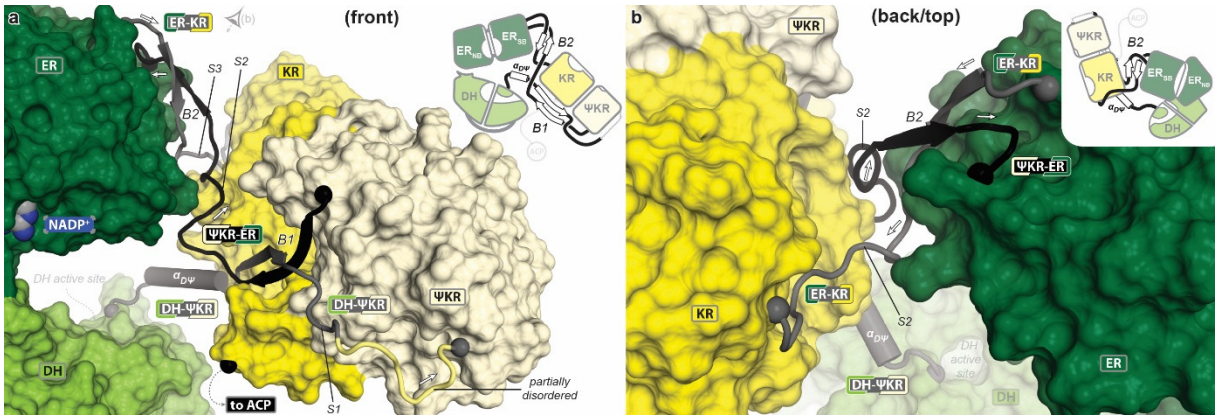


Figure 2.3 | Linker-based organization of the MAS modifying region.
The DH and ER lack direct interdomain contacts to the ΨKR/KR domains. The modifying region is established by an interplay of irregular and helical linker segments with two double-stranded antiparallel linker β-sheets (B1, B2), which interact with the ΨKR/KR and ER domain, respectively. **a**, The DH-ΨKR linker (aa 1177-1214) provides helix $\alpha_{D\Psi}$ as a spacer between DH and β-sheet B1 on the surface of the ΨKR/KR. The DH-ΨKR linker continues into segment S1 and ends in a partially disordered loop (light yellow), which was traced only in one chain. The central ΨKR-ER linker (aa 1392-1418) engages in both β-sheets (B1 and B2); the stretch S2 between sheet B1 and B2 adopts two alternate conformations among different chains. **b**, The ER-KR linker (aa 1744-1764) contains an irregular stretch (S3, aa 1753-1764), which is considerably longer than required to bridge the interdomain gap.

To obtain a MAS hybrid model we connected the overlapping modifying and condensing region fragments *in silico* (Figure 2.4a). We assume that the condensing region adopts a canonical dimeric state upon tethering to the dimeric modifying domain. The relative orientation of the condensing and modifying regions is not defined by the two structures and was chosen in accordance to intact FAS[162]. As in FAS, the two fragments connect without secondary contacts outside the linking region. Based on multiple modes of motion around the central linkage observed in FAS[165], the selected orientation may represent only one out of an ensemble of states in both multienzymes. Helix formation of the sequence segment linking modifying and condensing domain was observed at the N-terminus of four protomers in the crystallized modifying region under stabilization by crystal contacts. The central connection in the hybrid model consequently was modelled with an α-helix (Extended Data Figure 2.9a, b), in contrast to an irregular linker in FAS. Notably, short helices in equivalent sequence positions are observed in modPKS DEBS DH_4[175] and Rif DH_{10}[179] as well as in RhiE KS-B[131] (Extended Data Figure 2.9c, d), suggesting a more general conservation of helical linkers in modPKSs.

Conformational dynamics are a key component of multi-enzyme action. They have been visualized by EM for FAS[165] and PikAIII[42,45], but not at resolutions required for mechanistic dissection. The crystallographic visualization of 18 instances of the modifying regions now provides an opportunity to analyze conformational variability in MAS. The central DH and ER dimers each behave as rigid bodies, but the ERs move in a screw motion with a translation of up to 8.5 Å and a rotation by 14° on the DH platform (Figure 2.4b, Extended Data Figure 2.10a, b). The ERs are conformationally coupled to the ΨKR/KRs (Supplementary Video 1): Owing to the tethering of ΨKR/KR to both, the DH and ER, the screw motion of the ERs is transduced into a rotation of the ΨKR/KRs by up to 40° via a pivot in the linkers (Extended Data Figure 2.10c, d). Even larger motions may occur in solution, as indicated by pronounced disorder of some ΨKR/KRs in the crystal. Importantly, conformational coupling via relative DH-ER motions provides crosstalk between the two lateral clefts of MAS. Although a mechanism for reading out active site states remains unknown, this coupling could transmit reaction states across the MAS dimer. Notably, the mobile ACP is tethered to the most flexible catalytic domain (ΨKR/KR), creating a hierarchic network of gradually increasing domain flexibility.

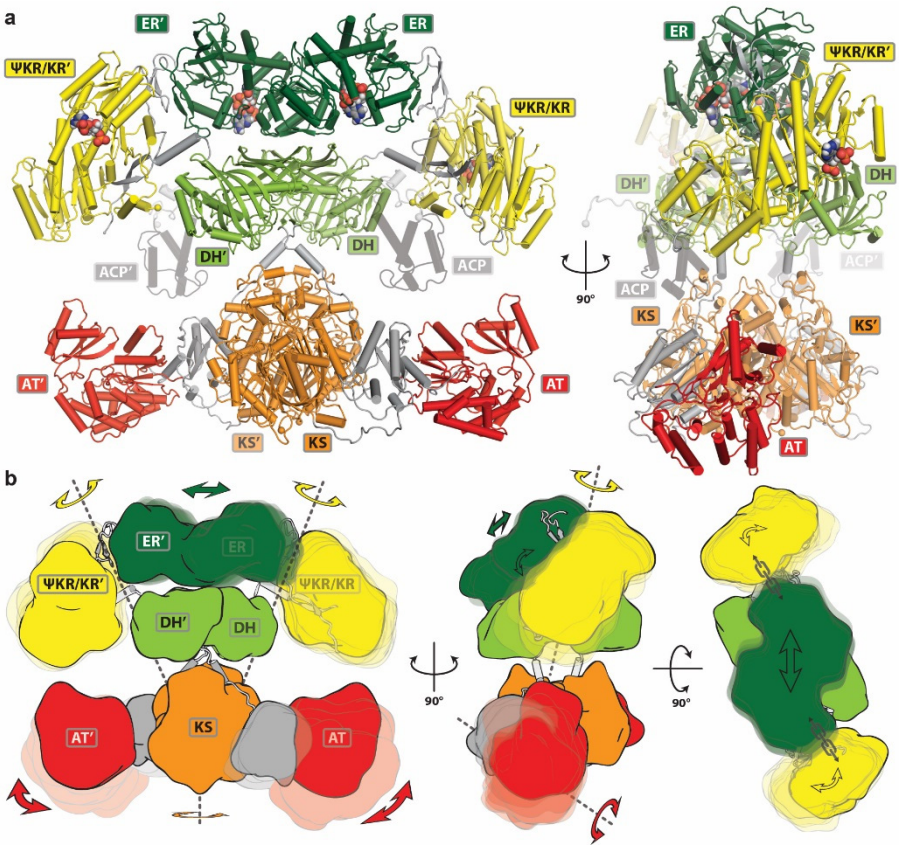


Figure 2.4 | Hybrid model of a dynamic MAS dimer.
a, A hybrid MAS model was assembled by linking the condensing and modifying region structures. Central helical linkers connect the two regions without secondary interactions. The orientation around the linkage is presumably flexibly and has been modeled according to the FAS structure. A homology model of mobile ACP is indicated (transparent grey) in a resting position without domain interactions. **b**, Conformational variability based on a comparison of 18 MAS modifying region chains and five homologous condensing region (Extended Data Figure 2.10e, f) structures in combination with normal-mode analysis are shown. The lateral ER dimer motion on the DH platform is coupled to a rotation of both double-tethered ΨKR/KRs by up to 40.4° (Supplementary Video 1). The positions of the AT relative to KS in different condensing region structures correspond to a rotation of up to 43° between the most linear (MAS) and the most bent (human FAS) variant.

Only one condensing region instance has been visualized here, but it extends the previously observed range of KS-AT conformations[46,162,163] (Extended Data Figure 2.10e, f). MAS KS-AT features the most linear conformation, which results in narrowing the gap to the modifying region and shortening of the AT-ACP anchor distances. Variations between condensing regions correspond to a hinge-bending motion of AT around a pivot in LD (Supplementary Video 2, Figure 2.4b). Although experimental evidence of flexibility in each system is lacking, normal-mode analysis indicates a conservation of this hinge in all KS-AT didomains. In the EM reconstruction of PikAIII the AT domain is rotated by approx. 90° relative to MAS and remains a clear outlier to the set of KS-AT regions depicted by crystallography[46,162,163], EM[165], and SAXS[43].

The MAS hybrid model is a prototype for Msl-PKS organization[95]. Moreover, our structural data reinforce the sequence-based conclusion that MAS also serves as a paradigm for modPKSs. Despite its iterative mode of action, MAS is clearly assigned phylogenetically to modPKSs (27-35 % sequence identity) rather than fiPKSs (20-22 % id.) or FASs (19 % id.) (Extended Data Figure 2.11). Structurally, the closest neighbors of all individual MAS domains are from modPKSs. The absence of a ΨME domain and the presence of a ΨKR β-α-β-α extension distinguish MAS and most modPKSs from FASs and fiPKSs.

"Handshake" interactions of isolated dimeric DHs are observed only in modPKSs, but not in FAS. Our structural data reinforce the earlier hypothesis that modPKSs are fundamentally similar to non-colinear iPKSs such as MAS, and presumably evolved by kinetic coupling of modules[124]. Indeed, several modPKS modules act in an iterative mode as part of an assembly line (*e.g.* BorA5[122], AurA[120]). Other modPKS modules can be converted into a non-colinear mode of action by mutation, *e.g.* DEBS module 3[26], or by isolation from their assembly line environment, *e.g.* PikAIII[118].

The analysis of the hybrid MAS structure depicts a unique PKS architecture. It agrees with previous biochemical and structural data on modPKSs fragments, with the exception of the monomeric state of some isolated ER domains[198] or in the domain-swapped crystal structure of the excised ER-ΨKR/KR didomain of the fully-reducing modPKS SpnB[47]. Based on the structure and the monomeric solution state of SpnB-ER-ΨKR/KR, as well as shorter ER-KR linker in modPKSs, a divergent architecture of modPKSs modifying regions based on a dimeric DH arrangement with laterally positioned monomeric KRs and ERs was proposed[47]. On the contrary, the MAS modifying region retains a central dimeric ER as observed in FAS and in a fragment of the modPKS PpsC ER (PDB: #1PQW) (Extended Data Figure 2.7c). Importantly, MAS reveals a dynamic linker-based organization, which (in contrast to FAS) could also accommodate the typical range of ER-KR linker lengths (5-22 aa) observed in modPKSs (Extended Data Figure 2.8) by slight adaptions of the ΨKR/KR position.

Models of PKS modifying regions based on SpnB-ER-ΨKR/KR and MAS are clearly distinct on a macromolecular scale and can be experimentally distinguished via SAXS distance distributions. We selected two well-expressed modifying regions from modPKSs bimodules, EryA of *Gamma proteobacterium* HdN1 (GpEryA) and "Pks" (Uniprot: Q3L885) from *Mycobacterium smegmatis* (MsPks) for comparative SAXS analysis. Calculated SAXS curves for SpnB-ER-ΨKR/KR and MAS-like models were compared with experimental SAXS data of MAS, GpEryA, and MsPks. The derived distance distributions closely match those calculated from a MAS-like model, but not those based on SpnB-ER-ΨKR/KR (Extended Data Figure 2.12). The SAXS analysis of GpEryA and MsPks clearly supports a wider relevance of the MAS architecture for modPKSs.

Our structural analysis not only provides detailed insights into MAS, a mycobacterial drug target, but also establishes a new paradigm for the organization of PKSs modules. It reveals a unique, dynamic structure of the modifying region based on dimeric DH and ER domains and provides insights into conformational variability and coupling in fully-reducing PKS modifying regions. The linker-based architecture supports modularity of the modifying region by requiring only the adaptation of variable linker regions for evolutionary domain shuffling. It thus rationalizes an important aspect of the outstanding success of the PKS architecture in the generation of chemical diversity. Our results highlight the relevance of matching linker-, rather than domain-domain interactions in PKS engineering. They contribute to the fundamental understanding of PKS architecture, as well as to the functional dissection and re-engineering of related synthases including relevant drug targets and important producers of bioactive compounds.

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

2.2 Acknowledgements

We acknowledge Friedrich Widdel and Johannes Zedelius for providing *Gamma proteobacterium* HdN1, Peter Leadlay and Lorena Betancor for providing plasmid pETcoco-2A-L1SL2, and EMBL Heidelberg for providing the pETG-10A vector; John Missimer (Paul-Scherrer Institute, Villigen, Switzerland (PSI) and Andreas Menzel (PSI) for outstanding support in SAXS data acquisition and raw data processing; Timothy Sharpe (Biophysics Facility, Biozentrum, University of Basel) for analytical ultracentrifugation, Adam Mazur (Research IT, Biozentrum) for SAXS refinement and Martino Bertoni (Swiss Institute of Bioinformatics and Biozenturm) for support of the homology-based assignment of the oligomeric state of MAS KS-AT. Data were collected at beamlines PXI, PXIII and cSAXS of PSI; we acknowledge excellent support from the beamline teams. This work was supported by the Swiss National Science Foundation (SNF) project grants 125357, 138262, 159696 and R'equip grant 14502. D.A.H. acknowledges a fellowship by the Werner-Siemens Foundation.

2.3 Author Contributions

R.P.J. expressed, purified and crystallized MAS, obtained the crystal structure of the condensing region, collected SAXS data and cloned constructs. F.Z cloned constructs and purified MAS, GpEryA and MsPks. D.A.H. purified MAS, optimized MAS crystallization, determined the structure of the isolated DH domains and the modifying region, collected SAXS data, analyzed the data, performed homology modelling, cloned constructs, and wrote the manuscript. TM designed and guided research, analyzed data, contributed to crystallographic analysis and wrote the manuscript.

2.4 Author Information

Address for all authors: Department Biozentrum, University of Basel, Klingelbergstrasse 50/70, 4056 Basel. E-mail addresses: dominik.herbst@unibas.ch, roman.jakob@unibas.ch, franziska.zaehringer@googlemail.com, timm.maier@unibas.ch

2.5 Accession Numbers

Atomic coordinates and structure factors for the reported crystal structures have been deposited with the Protein Data Bank under accession code 5BP1, 5BP2, 5BP3, 5BP4

Correspondence and requests for materials should be addressed to timm.maier@unibas.ch.

The authors declare no competing interests.

2.6 Methods

2.6.1 Cloning, Expression and Purification

Mycobacterium smegmatis (ATCC® 700084) was cultured according to ATCC recommendations. Cells were pelleted and washed with TE buffer. A cell pellet of G*amma proteobacterium HdN1* was provided by Dr. J. Zedelius (Max Planck Institute, Bremen, Germany). Cells were resuspended in lysis buffer

(0.1 M Tris pH 8.0, 0.2 M NaCl, 5 mM EDTA, 0.2 mg ml^{-1} lysozyme), incubated for 6 h at 37 °C, subsequently supplemented with 0.5 % SDS and 0.2 mg ml^{-1} proteinase K and incubated at 65 °C for 24 h. DNA was purified by phenol-chloroform extraction and dissolved in TE buffer. The MAS KS-AT constructs (Uniprot: A0R1E8, aa 1-884, 1-887, 1-892), MAS DH (A0R1E8, 884-1186), MAS DH-ΨKR-ER-KR (A0R1E8, 884-2020), "Pks" DH-ΨKR-ER-KR (Q3L885, 2450-3580) were cloned into pNIC28a-Bsa vectors; GpEryA DH-ΨKR-ER-KR (E1VID6, 2420-3575) constructs were cloned into a Gateway® compatible pETG-10A destination vector (provided by EMBL Heidelberg). MAS DH-ΨKR-ER-KR-ACP (A0R1E8, 884-2111) was cloned by codon optimized gene synthesis of ACP (GenScript) and restriction cloning (BsrGI/HindIII) into pNIC28a-Bsa-MAS DH-ΨKR-ER-KR (884-2020). All constructs were designed as N-terminal tobacco etch virus (TEV) protease cleavable hexa-histidine (His6) fusion constructs and co-expressed with *Streptomyces* chaperonins[202] (pETcoco-2A-L1SL2 plasmid) in BL21(DE3) and Rosetta(DE3) pLysS (GpEryA). Cells were cultured in 2xYT media, supplemented with 0.5 % glycerol, NPS (25 mM $(NH4)_2SO_4$, 50 mM KH_2PO_4, 50 mM Na_2HPO_4), kanamycin (100 µg/ml), chloramphenicol (34 µg/ml), and ampicillin (100 µg/ml). An expression culture (1.5 L) was inoculated (1:20), grown at 37 °C for 2 h, cooled to 20 °C, and induced with isopropyl-β-D-thiogalactopyranosid (0.1 mM) at an OD_{600} of 1.0. Cells were harvested after 12 h by centrifugation (7,000 x g) and resuspended in lysis buffer (50 mM HEPES pH 7.4, 20 mM imidazole, 0.5 M NaCl, 5 mM $MgCl_2$, 10 % glycerol (v/v), 2.5 mM β-mercaptoethanol), supplemented with protease inhibitors (200 µM phenylmethylsulfonyl fluoride, 20 µM bestatin, 4 µM E64, 2 µM pepstatin A, 20 µM phenantrolin, 2 µM phosphoramidon) as well as DNase, RNase, and lysozyme. Cells were placed on ice and lysed by sonication. The lysate was cleared by centrifugation (100,000 x g, 30 min) and the supernatant was loaded onto a 5 ml Ni-affinity column (GenScript) pre-equilibrated with lysis buffer. Unbound protein was eluted with 4 alternating wash cycles of 5 column volumes (CV) lysis buffer and HisA buffer (50 mM HEPES pH 7.4, 20 mM imidazole, 50 mM NaCl, 5 mM $MgCl_2$, 10 % glycerol (v/v), 2.5 mM β-mercaptoethanol, inhibitors), until a stable baseline (A280) was reached. The sample was eluted with 2 CV HisB buffer (50 mM HEPES pH 7.4, 250 mM imidazole, 50 mM NaCl, 10 % glycerol (v/v), 2.5 mM β-mercapto ethanol, inhibitors) and diluted (1:10) with AIC-A buffer (50 mM Tris-HCl pH 7.4, 20 mM KCl, 10 % (v/v) glycerol, 2.5 mM β-mercaptoethanol). The sample was loaded on a 6.5 ml anion exchange column (PL-SAX 4,000 Å, 10 µm) and washed with 20 CV. The samples were eluted with a stepped gradient to 100 % AIC-B (50 mM Tris-HCl pH 7.4, 1 M NaCl, 10 % (v/v) glycerol, 2.5 mM β-mercaptoethanol). For DH-ΨKR-ER-KR the gradient was held at a conductivity of 15 mS/cm until a stable baseline (A280) was obtained in order to elute *Streptomyces* chaperonins. DH-ΨKR-ER-KR eluted at 17-20 mS/cm. Pure fractions were pooled, supplemented with TEV protease (1 mg protease per 100 mg tagged protein) and incubated for 10 h at 4 °C. Uncleaved protein, as well as the cleaved His6-tag was removed by passing the solution through a 5 ml orthogonal Ni-affinity column (GenScript). The flow-through was pooled, concentrated and subjected to gel permeation chromatography (Superdex 200 16/60, GE Healthcare) using GPC buffer (20 mM HEPES pH 7.4, 250 mM NaCl, 5 % glycerol (v/v), 5 mM dithiothreitol). Pure fractions were pooled, and monodispersity was monitored by dynamic light scattering at 1 mg ml^{-1}. Related purification protocols were applied to MAS DH-ΨKR-ER-KR-ACP, "Pks" DH-ΨKR-ER-KR (HiTrap CaptoQ column), GpEryA DH-ΨKR-ER-KR (both no TEV protease cleavage and orthogonal Ni-affinity column), and MAS DH (no anion exchange chromatography).

2.6.2 Crystallization

All crystallization experiments were performed using a robotic setup applying the sitting drop vapor diffusion method.

MAS KS-AT bipyramidal crystals were grown at 4 °C by mixing 0.2 µl of protein in GPC buffer (38 mg ml^{-1}) with 0.2 µl reservoir solution (0.1 M MES/imidazole pH 6.5, 0.1 M $MgCl_2$, 0.1 M $CaCl_2$, 12.5 % (v/v) polyethylene glycol 1,000 (v/v), 7.5 % (w/v) polyethylene glycol (PEG) 3,350, 12.5 % 2-methyl-2,4-pentanediol (MPD)). Crystals grew to a final size of 0.8x0.4x0.2 mm^3 within one week and were flash frozen in liquid nitrogen.

The MAS DH domain was crystallized in space group $P2_1$ at 18 °C by mixing 0.2 µl of protein in GPC buffer (38 mg/ml) with 0.1 µl reservoir solution (0.1 M bis-Tris pH 6.5, 0.2 M $MgCl_2$, 25 % (v/v) PEG 3,350) and grew to a final size of 0.4x0.2x0.1 mm^3 within one week. Crystals in space group $P2_12_12$ appeared after 30 days at 18 °C by mixing 1 µl of protein in GPC buffer (38 mg/ml) with 2 µl reservoir solution (0.25 M di-sodium malonate, 24 % (w/v) PEG 3,350) and grew to a final size of 1x0.4x0.3 mm^3. Prior to harvesting all crystals of MAS DH were cryo protected (25 % (v/v) ethylene glycol) and flash frozen in liquid nitrogen.

Needle shaped crystals of MAS DH-ΨKR-ER-KR were obtained by mixing protein solution at 18.4 mg ml^{-1} (GPC buffer, 1.5 mM $NADP^+$) and reservoir solution (0.03 M $MgCl_2$, 0.03 M $CaCl_2$, 20 % ethylene glycol, 10 % PEG 8000, 0.1 M MES/imidazole pH 6.5) at 4 °C. Crystallization was optimized by exchanging PEG 8,000 by PEG 3,350, decreasing the PEG 3,350 concentration to 7-13% (w/v) and by carefully monitored microseeding. Subsequent optimization was performed using automated robotic setup and seeding at 4 °C. Final crystals (1.0x0.3x0.2 mm^3) were obtained after mixing 1 µl protein (20.3 mg ml^{-1} in GPC buffer incl. 1.5 mM $NADP^+$) with 1 µl of reservoir solution (5.25 % (w/v) PEG 3,350, 20 % (v/v) ethylene glycol, 0.1 M MES pH 7.0, 52 mM $MgCl_2$, 52 mM $CaCl_2$) and 0.2 µl seed stock. Diffraction properties were optimized by crystal dehydration: Over a period of 4 h crystals were transferred to a dehydration solution (0.05 M MES pH 7.0, 25 % ethylene glycol, 25 % PEG 3,350, 56 mM $MgCl_2$, 56 mM $CaCl_2$, 1.5 mM $NADP^+$) by a step-wise exchange of the drop solution. All crystals were harvested and plunge frozen in liquid nitrogen. Integrity of the protein in final crystals was examined by denaturing polyacrylamide gel electrophoresis.

2.6.3 Data collection and structure determination

All data sets were collected at the Swiss Light Source (SLS, Villigen, Switzerland) at a temperature of 100 K. Data sets of DH crystals were collected at beamline X06DA ($P2_1$: λ= 0.999870 Å, T= 100 K; $P2_12_12$: λ= 0.97626 Å). All other data sets were collected at beamline X06SA (KS-AT: λ=0.97940 Å, DH-ΨKR-ER-KR: λ= 0.97626 Å). Data reduction was performed using XDS[203] and XSCALE[203], datasets were analyzed with phenix.xtriage[204]. All structures were solved with PHASER[205] using molecular replacement.

Crystals of all KS-AT didomain variants of MAS are isomorphic in space group $P4_12_12$. The KS and AT domains of DEBS KS_5-AT_5[163] were used as molecular replacement templates and initial rebuilding was achieved by BUCCANEER[206]. All three crystal structures were virtually identical except for the identity of the last ordered C-terminal residue. The construct with the most extended C-terminus (1-892) revealed aa 887 as last ordered residue, which is overlapping in sequence with the modifying region. Thus we continued refinement only for crystals of this variant (aa 1-892) with unit cell constants of a= 77.5 Å, b= 77.5 Å, c= 371.2 Å and a solvent content of 56 %. A final model was obtained after iterative cycles of real space model building in COOT[207] and TLS refinement in Phenix[204] and was refined to R_{work}/R_{free} values of 0.21/0.23 at 2.3 Å resolution with excellent geometry (Ramachandran favored/outliers: 97.8 %/0.2 %) (Extended Data Table 2.1a).

Crystals of the DH domain of MAS belong to space group $P2_1$ (a= 59.7 Å, b= 162.4 Å, c= 66.6 Å, β= 91.4°) and $P2_12_12$ (a= 67.1 Å, b= 162.2 Å, c= 59.5 Å) with a solvent content of 49 % and 51 %, respectively. A molecular replacement search model was based on CurK DH[176]. Initial maps were improved by density modification and NCS averaging with PARROT[208], followed by automated rebuilding with BUCCANEER[206]. Final models were obtained after iterative cycles of model building in COOT[207], and refinement in BUSTER[209] ($P2_1$) and Phenix[204] ($P2_12_12$), yielding excellent geometry (Ramachandran favored/outliers: $P2_1$=98.2 %/0.0 %; $P2_12_12$=98.2 %/0.2 %) and R_{work}/R_{free} values of 0.18/0.20 ($P2_1$) and 0.15/0.18 ($P2_12_12$) (Extended Data Table 1a).

Crystals of MAS DH-ΨKR-ER-KR in space group P1 (a= 151.4 Å, b= 190.4 Å, c= 270.8 Å, α= 95.6°, β= 91.9°, γ= 103.7°) diffracted to a maximum resolution of 3.75 Å. The asymmetric unit contained 18 protomers in nine dimers with 20,502 amino acids and a molecular mass of 2.2 MDa at 65 % solvent content. Data were collected at four different positions of a single crystal and combined to obtain a complete high-quality dataset. The resolution cutoff was determined by $CC_{1/2}$ criterion[210]. Self-rotation functions revealed non-crystallographic rotational symmetry (NCS) and the native patterson function indicated translational NCS.

Initially, a partial molecular replacement solution was obtained for the ER dimer using the ER domain of porcine FAS (pFAS)[162]. Other known structures of homologous domains did not provide efficient search models. The structure of the isolated MAS DH domain, determined here independently, yielded equivalent solutions in agreement with the pFAS ER based solution. For final structure determination both models where used in subsequent rounds of molecular replacement. Start models for building further regions were generated by homology modelling using Swiss Model[211]. In order to allow unbiased refinement in real and reciprocal space, phenix.reflection_tools[204] was used to define a thin-resolution shell-based test set[212], and test set reflections were excluded from calculating maps, which were used for real-space refinement. Initial refinement cycles included rigid body refinement and restrained refinement. The impact of various low-resolution restraint formulations on refinement were tested carefully. Local NCS is particularly well-suited for MAS DH-ΨKR-ER-KR intermediate resolution refinement due to the high degree of NCS and the fact that using local NCS restrains avoids any external standard restraints based on assumptions on secondary structure or homologous peptide structures. Thus local structural similarity restraints (LSSR)[213] were combined only with reference model restraints to the authentic DH domain structure using autopruning in BUSTER[209]. After every round of refinement, bias-reduced, solvent flattened and NCS-averaged maps were calculated using DM[214] without applying phase combination. Sharpened NCS-average maps were generated by applying a sharpening B-factor to the structure factor amplitudes prior to averaging. Initially, real-space rigid body fitting of individual secondary structure elements was applied for instances of every domain type (DH, ΨKR, ER_{NB}, ER_{SB}, KR) followed by symmetry expansion and rigid body fitting for entire domains. Best defined regions of the electron density maps were used for rebuilding of every domain type using Coot[207] and O[215], respectively, symmetry expanded, and recombined into 18 chains. At this point, unambiguous difference electron density indicated the connecting linkers, which were manually built into the maps and refined without symmetry expansion (Extended Data Fig. 2d, e). Later refinement cycles included TLS refinement, using one group per domain and linker, individual B-factor refinement and automated weight factor determination. During rebuilding B-sharpening, NCS average and density modification as well as feature enhanced maps[204] were used. Overall, the use of 18-fold-domain-wise NCS averaging results in highly accurate and unbiased phase determination irrespective of details of the atomic model. The combined use of NCS-averaging and B-factor sharpening led to an exceptional

map quality typical for maps at considerable higher resolution; (Extended Data Figure 2.6f). Bound $NADP^+$ cofactors were added for final refinement cycles. $NADP^+$ is well ordered in the ER domain, while the nicotinamide moieties are disordered in the KR domains and were not included in the final model. A total of five KR and four ΨKR domains, which lack stabilization by crystal contacts, were either disordered or present in multiple orientations, and not included in the final model, despite significant positive difference density. The KR domain in chain L shows a significantly more tilted orientation as observed in all other instances of the KR domains, which however agrees with the identified hinge regions. A single model was placed for this domain, which achieved the largest improvement of R-factors and was characterized by the lowest B-factors after refinement, although a secondary alternate conformation might be present. The refinement of the final model (excluding disordered regions (chains): 883-895 (E-R), 1206-1213, 1283-1287, 1948-1960, ΨKR(I/L/O/Q-R) , KR(I/O/Q-R)) was completed with R_{work}/R_{free} values of 0.23/0.24 and very good geometry for the resolution range (Ramachandran favored/outliers: 91.6 % / 1.8 %).

2.6.4 Analytical Ultracentrifugation

To determine oligomeric states in solution, sedimentation equilibrium analytical ultracentrifugation experiments were performed for MAS DH-ΨKR-ER-KR and MAS KS-AT. 140 µl columns containing proteins at concentrations of 3.5-4.5 mg ml^{-1} in GPC buffer were subjected to centrifugation at 4,800 and 7,800 rpm at 12 °C, with detection by radial absorbance scanning at 305 nm. At each speed, centrifugation was allowed to proceed until sedimentation equilibrium was attained, as judged by pairwise comparison of scans using the approach to equilibrium function in SEDFIT (https://sedfitsedphat.nibib.nih.gov). Buffer density (1.0277 g/ml) and viscosity (1.5306 cP) were measured at 12 °C using an Anton Paar DMA4500M densitometer and an AMVn viscometer, respectively. Molar extinction coefficients at 305 nm were calculated for each protein from the ratio of observed absorbance at various wavelengths in spectra at different dilutions and calculated molar extinction coefficients. The partial specific volume for each protein was calculated from sequence in SEDFIT. The radial absorbance scans at equilibrium for the two speeds were globally fitted to the "single species of interacting system" mode in SEDPHAT[216] to determine the apparent molecular mass of the protein in solution. If the obtained molecular mass was intermediate between the value expected for a monomer and a dimer, the data were globally fitted to the monomer-dimer association model in SEDPHAT, with the molecular mass of the monomer fixed to the value calculated from the sequence. In both cases data were fitted using a fixed meniscus position, a floating bottom position, mass conservation constraints, a floating baseline and fitting radially-independent noise components. Confidence intervals on single-species masses or dissociation constants were obtained by the Monte-Carlo method implemented in SEDPHAT.

2.6.5 Small angle X-ray scattering

Small angle X-ray scattering (SAXS) data were collected at the beamline X12SA of SLS. Samples were dialyzed into GPC buffer, diluted to concentrations between 3-10 mg ml^{-1} and centrifuged at 13,000 x g and 8 °C until measurement. Glass capillaries (1 mm inner diameter) were mounted on a temperature-controlled holder at 12 °C. Data collection was performed using a Pilatus 2M detector at a distance of 2.14 m and a wavelength of 1.000 Å. Data were collected in eight repetitive scans each including ten 40 ms acquisitions at ten capillary positions yielding a total of 800 frames per buffer and protein, respectively. Frames with artefacts *e.g.* from air bubbles, were identified using SLS/PSI software (SAXS_inspect2) and excluded from the data sets. Radial averages were calculated and

exported using beamline software for scattering vectors from 0.005 to 0.7 $Å^{-1}$ defined as $q = 4\pi/\lambda sin\theta$. Scattering curves were averaged using DATAVER[217]; buffer profiles were subtracted using DATOP[217]. Scaling factors and p-values of a Students-T test were analyzed using DATMERGE[217] and DATCMP[217], respectively. Later frames were affected by increasing radiation damage and were excluded from further processing. Final scattering curves for each sample concentration were thus obtained from 300 individual profiles. The radius of gyration (R_g) and zero angle intensity (I(0)) was calculated from the Guinier approximation using AUTORG[217] and is consistent with values obtained from atomic distance distributions p(r) using DATGNOM[217] (Extended Data Table 2.1b). Scattering profiles at different concentrations were only combined if a noise reduction at medium and high scattering vectors could be obtained.

Modifying regions bear an intrinsic flexibility, which requires a flexible fitting approach in order to sample the full conformational space of the structures. Some approaches for flexible SAXS fitting have been described[218,219], but none was able to refine an individual structure while maintaining two-fold symmetry. Therefore, we combined dynamic elastic network restraints from CNS[220] with SAXS-target refinement and twofold symmetry averaging in XPLOR-NIH[221] for the refinement of individual structures by simulated annealing. SAXS scattering curves of atomic models, fits with experimental data, and distance distributions were calculated using CRYSOL[217] and DATGNOM[217]. All SAXS curves were plotted using Python Matplotlib.

For comparing calculated and experimental SAXS scattering curves, three models for the architecture of modifying regions were generated based on the crystal structure of the domain-swapped SpnB fragment (ER-KR/ΨKR). The first model was obtained according to the original publication[47] by superposing the monomeric ER-KR/ΨKR domain on the KR domain of pFAS[162]. A linear homology model of SpnB DH[211] was placed into the position of pFAS DH and the domain swap in SpnB ER-ΨKR/KR was replaced with the corresponding region from DEBS KR_1[161]. The second model was constructed in the same way via a superposition on MAS KR. The relative domain arrangement of SpnB ER-ΨKR/KR was not altered in these two models, only the domain swap was corrected. The third more generalized modPKS model was constructed in order to verify if shorter ER-KR linkers are in contradiction with the architecture of MAS. As a representative for short ER-KR linkers, the structure of SpnB ER-ΨKR/KR (6 aa) and the corresponding DH homology model were modeled as individual domains on MAS, while the linear DH dimer was maintained. ΨKR-ER linkers could be readily reconnected and regularized, whereas the ER-KR linker required a tilt of the ΨKR/KR domain. The tilt maintained a reasonable distance between the C-terminus of the DH and N-terminus of the ΨKR domain and yielded a linker architecture of a modPKS in agreement with MAS without stable direct interdomain contacts. SAXS curves and distance distributions were calculated of all models and compared to experimental SAXS scattering curves of MAS and two modPKS modifying regions with short ER-KR linkers (GpEryA: 9 aa; MsPks: 8 aa).

2.6.6 Structure analysis and visualization

Related structures were identified using PDBeFold[222] and interfaces were analyzed using QtPISA[223]. Transformations and coordinate manipulations were carried out using CCP4[224] tools, MODTRAFO (T. Schirmer, Biozentrum Basel, http://www.biozentrum.unibas.ch) and MOLEMAN[225]. The automated Oligo algorithm[226] as implemented in Swiss Model unambiguously detected and predicted a single mode of dimerization of MAS KS-AT based on sequence homology. Initially, the dimeric form of KS-AT was assembled by least squares fitting of secondary structure elements on DEBS KS_5[163]. Then, all

residues in a radius of 7.5 Å to the dimer interface were deleted and multi-template homology modelling using modeller 9.15[227] was used to construct a full-length dimeric homology model based on 20 homodimeric PKSs/FASs KS structures and the interface deleted MAS KS-AT structure. Remodeled regions (excluding all crystallographically defined regions beyond the radial cutoff) were geometry minimized using phenix.geometry_minimization[204]. The position where the post-AT linker becomes disordered was located by crystallization of KS-AT didomains with three different linker lengths (1-884, 1-887, 1-892). Normal mode analyses was carried out using the Bio3D[228] library in "R". Hinge bending analysis was carried out by pre-aligning all structures to a reference substructure using LSQKAB[229], followed by a MODTRAFO (T. Schirmer, Biozentrum Basel, http://www.biozentrum.unibas.ch) analysis of the moving substructure. Principle screw axes were determined by averaging the direction vectors of the screw axes using Python Numpy and locating a central hinge point from the position of all screw axes. Active site distances were calculated using BIOPYTHON[230]. All axes were visualized using PYMOL[231]. Interdomain angles of DH dimers were calculated by pre-aligning all DH dimers to one DH domain of MAS DH, followed by calculating the angle between the first principle component vector of the secondary structure elements of both domains. The angles were visualized using PYMOL[231]. Bias-removal for F_{obs}-F_{calc} omit maps was achieved by applying a random perturbation to coordinates (Δ0.2 Å) and B-factors (Δ20 % of the mean overall B-factor) using MOLEMAN2[225] prior to refinement. Figures, movies and active site tunnels were generated using PYMOL[231], LSQMAN[232], and CAVER 3.0[233].

2.6.7 Sequence analysis

55 sequences containing fully reducing modifying regions were selected from FASs, fiPKSs, Msl-, one *trans*-AT and 36 modPKSs modules. Structure-based sequence alignments of all PKSs/FASs type I domain structures were generated using PDBefold[222] and used as reference for the alignment of individual domains using ClustalW2[234]. Linkers were aligned without reference, assembled with the individual domain alignments and manually corrected in Geneious v7.1.7[235]. Phylogenetic trees were generated using the neighboring joining algorithm in Geneious v7.1.7[235].

2.7 Extended Data

Extended Data Table 2.1 | X-ray data collection and processing table.

a

	KS-AT 1-892	DH 884-1189	DH 884-1189	DH-ΨKR-ER-KR 884-2020
Data collection				
Space group	P 4_12_12	$P2_1$	$P2_12_12$	P_1
Cell dimensions				
a, b, c (Å)	77.53, 77.53, 371.22	59.65, 162.40, 66.62	67.06, 162.20, 59.49	151.38, 190.37, 270.84
α, β, γ (°)	90.0, 90.0, 90.0	90.0, 91.4, 90.0	90.0, 90.0, 90.0	95.6, 91.9, 103.7
Resolution (Å)	92.81 - 2.20	66.65 – 1.75	47.99 - 1.45	78.62 – 3.75
R_{merge} (%)*	8.7 (135.8)	3.6 (123.2)	4.3 (134.5)	25.0 (315.9)
$I/\sigma I$*	17.98 (2.26)	12.32 (1.13)	19.07 (1.36)	8.95 (1.00)
$CC_{1/2}$*	99.9 (67.7)	99.9 (46.2)	100.0 (76.6)	99.6 (44.9)
Completeness (%)*	99.6 (99.2)	96.8 (91.6)	99.3 (97.0)	99.0 (98.9)
Redundancy*	12.9 (13.2)	3.9 (3.8)	6.5 (6.5)	9.0 (9.3)
Unique reflections*	58,436 (9,008)	123,118 (8,582)	114,826 (18,878)	296,164 (21,835)
Refinement				
Protomers	1	4	2	18
Resolution (Å)	54.82 - 2.20	66.6 – 1.75	47.97 - 1.45	78.62 – 3.75
R_{work}/ R_{free}	0.21 / 0.23	0.18 / 0.20	0.15 / 0.18	0.23 / 0.24
No. atoms	12,279	18,419	9,282	262,498
Protein	11,984	17,273	8,725	260,724
Ligand/ion	--	408	60	1774
Water	295	738	497	--
B-factors	81.91	57.63	35.29	171.34
Protein ($Å^2$)	82.38	56.90	34.55	171.44
Ligand/ion ($Å^2$)	--	76.57	52.12	157.33
Water ($Å^2$)	63.04	64.27	46.27	--
R.m.s deviations				
Bond lengths (Å)	0.003	0.010	0.011	0.008
Bond angles (°)	0.73	1.10	1.31	0.97

b

	MsMAS DH-ΨKR-ER-KR	MsMAS DH-ΨKR-ER-KR-ACP	MsPks DH-ΨKR-ER-KR	GpEryA DH-ΨKR-ER-KR
Beamline	X12SA (SLS)	X12SA (SLS)	X12SA (SLS)	X12SA (SLS)
Wavelength (Å)	1.00000	1.00000	1.00000	1.00000
Detector distance (m)	2.14	2.14	2.14	2.14
q range ($Å^{-1}$)	0.005-0.707	0.005-0.707	0.005-0.707	0.005-0.707
Capillary diameter (mm)	0.1	0.1	0.1	0.1
Scan lengths / step size (mm)	4.5 / 0.5	4.5 / 0.5	4.5 / 0.5	4.5 / 0.5
Positions / acquisitions	10 / 10	10 / 10	10 / 10	10 / 10
Scan repeats	8	8	8	8
Exposure time (sec)	0.04	0.04	0.04	0.04
Concentration (mg ml^{-1})	3, 6	5, 10	6	3, 6
Temperature (K)	285	285	285	285
I(0) ($Å^{-1}$) [from P(r)]	3.00 ± 0.00	2.26 ± 0.00	2.08 ±0.00	2.00 ± 0.00
Rg (Å) [from P(r)]	58.25 ± 0.09	66.90 ± 0.19	55.05 ± 0.09	56.54 ± 0.09
I(0) ($Å^{-1}$) [from Guinier]	3.01 ± 0.01	2.19 ± 0.01	2.1 ± 0.01	1.98 ± 0.01
Rg (Å) [from Guinier]	57.2 ± 0.30	62.23 ± 0.63	54.7 ± 0.29	54.9 ± 0.23
Dmax (Å)	201	250	191	192
Model fit (χ)	1.79	N/A	N/A	N/A

a, Crystallographic data collection and refinement statistics. The resolution cutoff was determined by $CC_{1/2}$ criterion (Karplus and Diederichs, 2012). **b**, Small angle X-ray scattering data collection and processing. *, Highest resolution shell is shown in parenthesis.

Extended Data Table 2.2 | Structural comparison and interface analysis.

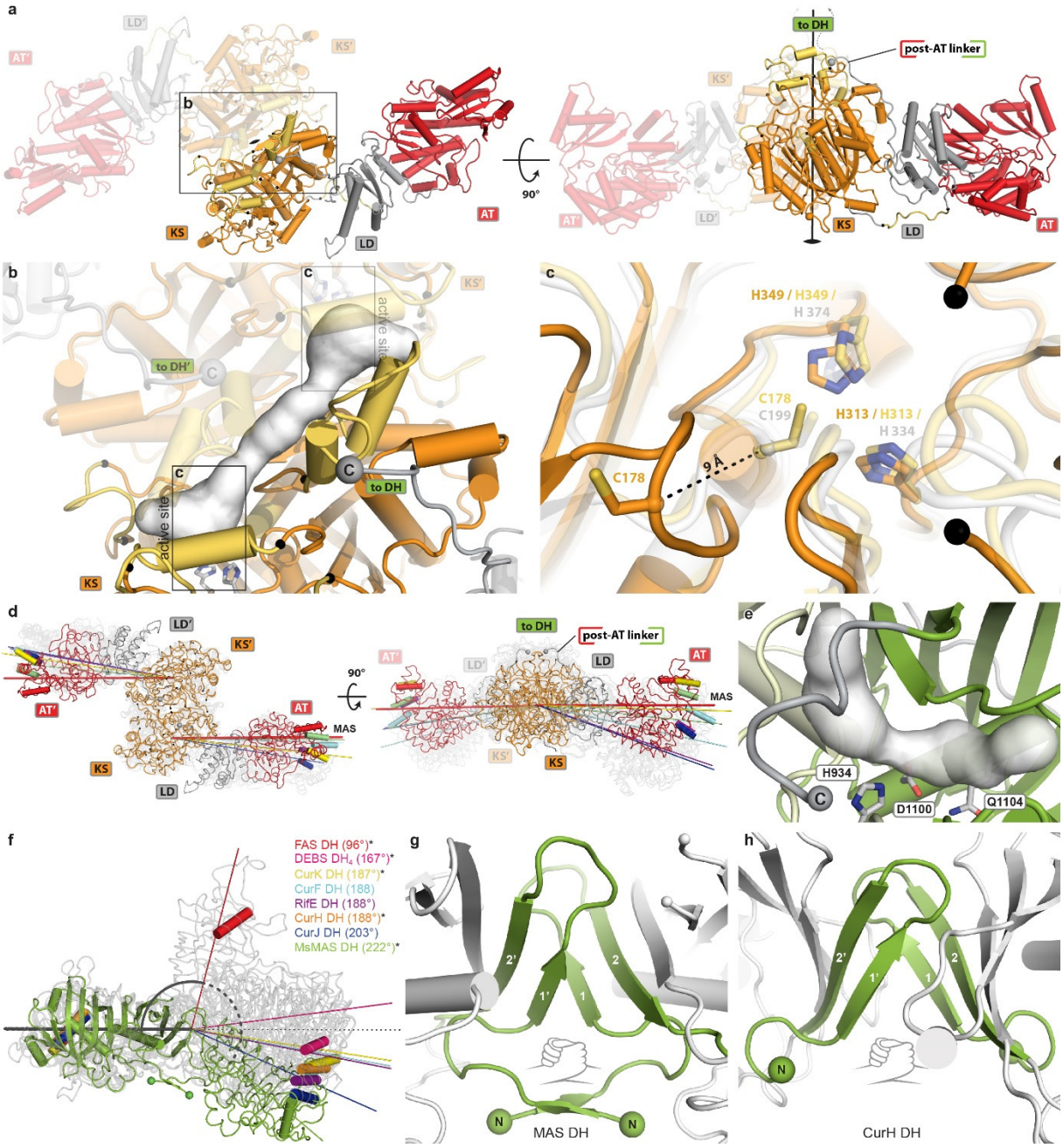
a

Structure 1	Structure 2	C_α r.m.s.d. [Å]	Aligned residues
MAS KS	DEBS KS_5	1.17	348
MAS KS	DEBS KS_3	1.21	349
MAS AT	DEBS KS_5	1.77	294
MAS AT	DEBS KS_3	1.21	287
MAS DH	Rif DH_{10}*	1.86	249
MAS DH	CurF DH*	1.61	259
MAS DH	CurJ DH*	1.88	252
MAS DH	CurK DH	1.68	257
MAS DH	CurH DH	1.93	258
MAS DH	DEBS DH_4	1.82	244
MAS DH	MAS DH $P2_12_12$ / $P2_1$	0.70	284
MAS DH $P2_1$	MAS DH $P2_12_12$	0.40	284
MAS ER	SpnB ER	1.98	298
MAS ER_{NB}	PpsC ER_{NB}†	1.13	170
MAS ER	JamJ ER*	1.63	308
MAS ER	CurF ER*	1.72	309
MAS ER	CurK ER	2.27	306
MAS ER	mFAS ER	2.15	299
MAS ΨKR/KR	Tyl ΨKR/KR_{10}*	1.83	380

b

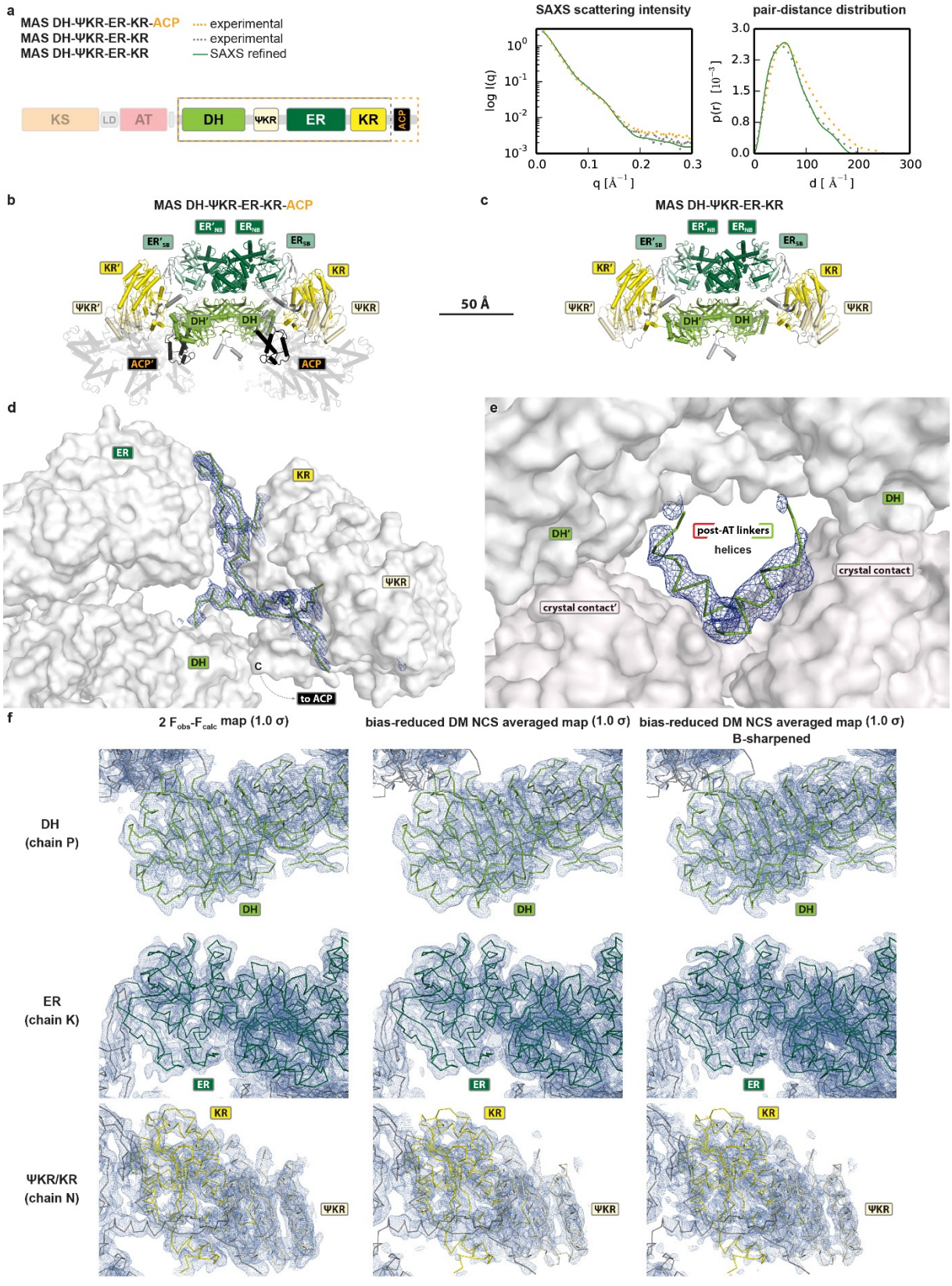
Interface 1	Interface 2	Area [Å^2]	SDV [Å^2]	Min [Å^2]	Max [Å^2]
KS	KS‡	1,759			
	post-AT linker	770			
LD	KS	521			
	AT	236			
DH ($P2_1$)	DH ($P2_1$)	979	59	938	1,021
DH ($P2_12_12$)	DH ($P2_12_12$)	999			
DH (P_1)	DH (P_1)	961	13	943	978
ER	ER	1,424	20	1,397	1,449
DH-ΨKR-ER-KR	DH-ΨKR-ER-KR	2,688	112	2,547	2,904
DH[d]	ER[d]	532	105	345	638
DH-ΨKR linker	ΨKR/KR	1,117	24	1,077	1,172
	ΨKR-ER linker	695	9	672	711
ΨKR-ER linker	ΨKR/KR	975	28	926	1,027
	ER-KR linker	471	24	426	517
	ER	353	20	330	411
ER-KR linker	ER	432	24	399	480
	KR	547	14	527	574
ΨKR-ER / ER-KR linkers	KR	1,111	22	1,076	1,162
	ΨKR/KR	1,488	25	1,449	1,544
	ER	683	30	637	735

a, C_α r.m.s. deviations obtained for structural comparison of MAS domains with their closest structural neighbors. *, not part of a fully-reductive modifying region, †, PDB entry 1pqw (unpublished). **b**, Interfaces in the crystal structures of MAS variants. Standard deviations (SDV) and minima/maxima are given for structures containing more than one interface.‡, by direct superposition of the monomeric KS-AT structure on the DEBS KS5 dimer. In the KS-AT dimer with restored interface (by homology modelling), the total area increases to 2.289 Å^2. d, dimer.



Extended Data Figure 2.5 | Reconstruction of the dimeric KS-AT didomain and DH dimer organization.
a, The condensing region dimer was reconstructed by least square fitting on DEBS KS_5[163] and multi-template homology modeling of disordered segments and the active site loop (gold). Termini of the remodelled segments are indicated by black spheres. A pseudo-continuous β-sheet is formed across the dimer interface. The post-AT linker terminates close to the dimer axis. **b**, Close-up view on the reconstructed KS dimer with an active site tunnel spanning both protomers (white), which is enclosed by four remodelled segments (gold). **c,** The active site loop containing the catalytic Cys178 is dislocated in the monomeric (orange) form of MAS KS-AT, whereas the active site His313 and His349 occupy the same position as in the dimeric DEBS KS_5-AT_5 structure (white-transparent). The canonical conformation of Cys178 observed in dimeric KS domains is restored in the dimeric KS-AT model (gold-transparent) **d**, MAS KS-AT (colored, red line) reveals the most linear overall structure (right panel) of all PKSs/FAS condensing region structures[46,162-164,172] (corresponding to Extended Data 6e, f). **e**, The DH active site residues are located at the interface of the two hot-dog folds (light and dark green; active site tunnel in white). **f**, Interdomain angles in DH dimers[162,175,176,179]. Dimers were superposed onto one protomer (left) of MAS, and the angles between two protomers are compared. For clarity, only MAS DH is shown in green, for other DH domains only one equivalent helix is highlighted in color. The FAS pseudo-dimeric DH domains (red helix) adopt a "V"-shaped structure (interdomain angle: 96°), while PKS DH dimers (various colors) are almost linear (167° - 203°). The MAS DH dimer (green) is bent to the opposite direction relative to FAS, and exhibits the largest

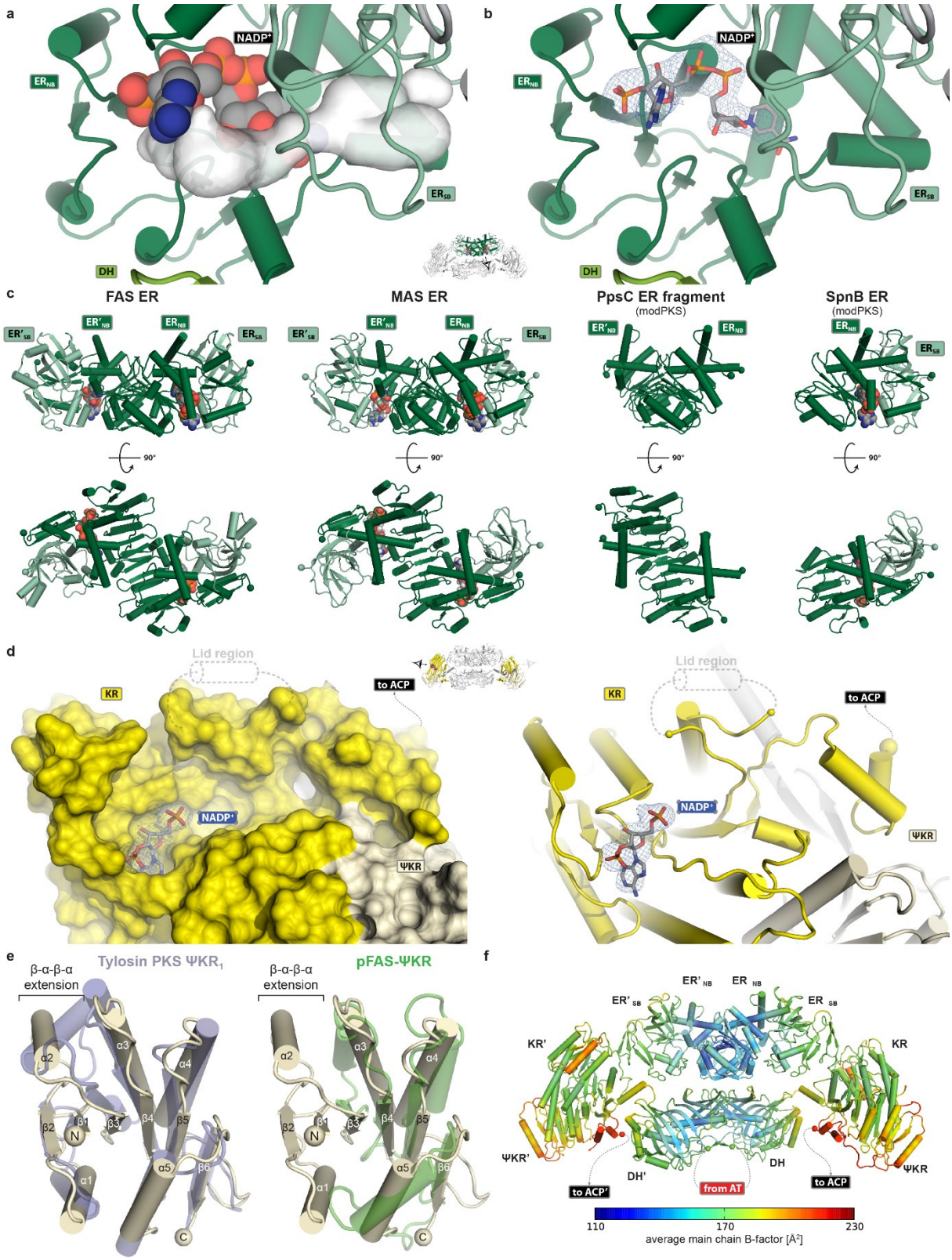
interdomain angle (222°) (asterisks indicates DHs that are part of fully-reducing modifying regions). **g**, Dimer interface of MAS DH, and **h,** dimer interface of the isolated DH of the CurH^{176} modPKS. Dimerization of MAS and CurH DH are mediated by "handshake" interactions of the N-terminal hot-dog folds. In MAS DH, an N-terminal β-strand extension further contributes to dimerization.



Extended Data Figure 2.6 | Effect of ACP deletion and electron density maps of the MAS modifying region crystal structure.

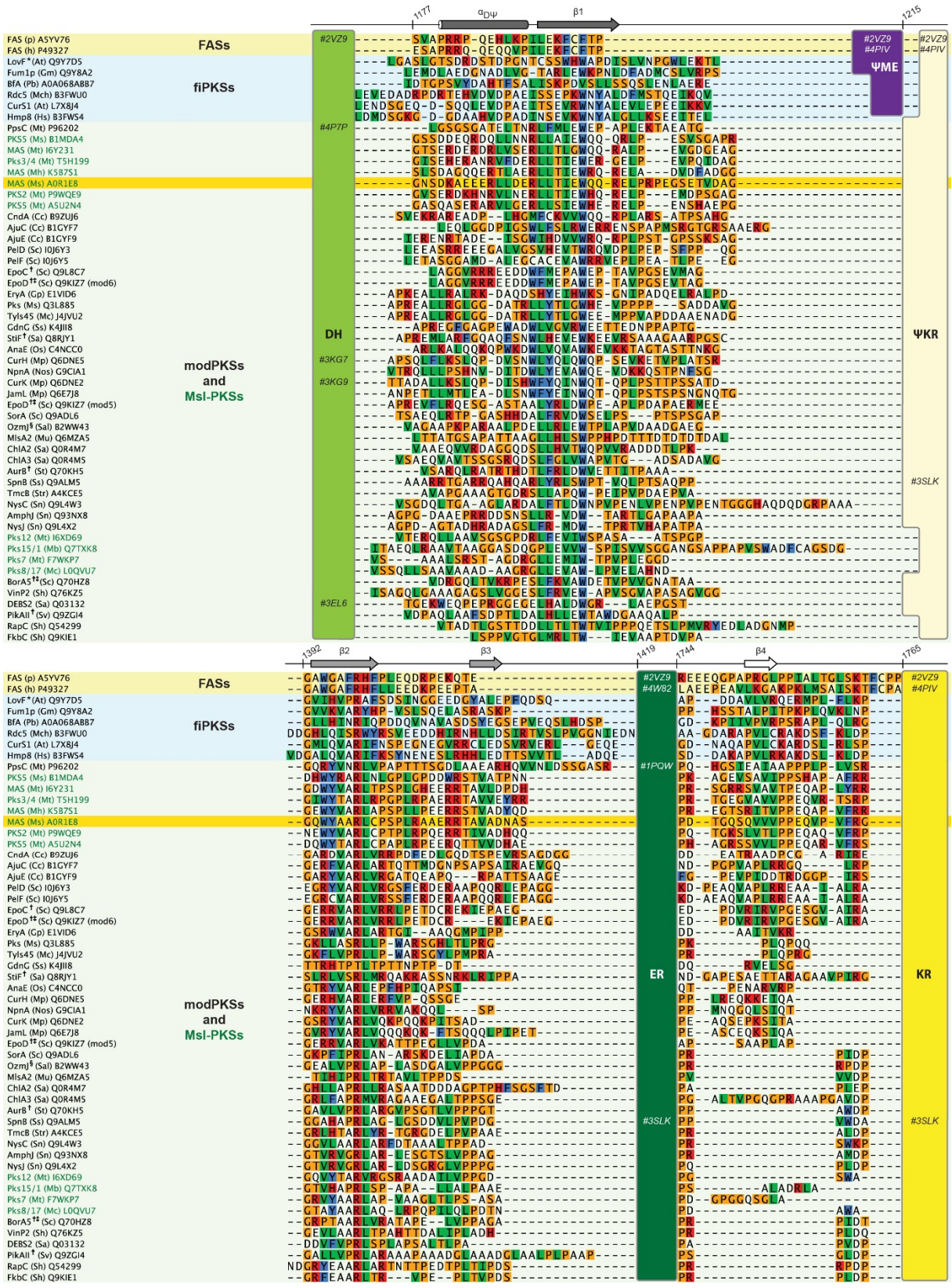
a, SAXS experiments reveal conserved scattering profiles for the modifying region with ACP (dotted orange) and without ACP (dotted green), which resemble the scattering curve of the SAXS-refined X-ray structure (green). **b**, **c**, The experimentally determined interatomic distance distributions are in agreement with the maximum extends of the modifying domain with (**b**) and without (**c**) ACP, 250 Å and 201 Å, respectively. In **b** a set of plausible ACP positions is shown (transparent), based on the length of the KR-ACP linker. **d,** Unbiased F_{obs}-F_{calc}

omit difference map of the modifying region linkers in chain B (contoured at 2.5 σ) is shown. **e**, Unbiased F_{obs}-F_{calc} omit difference map of the post-AT linker helices in chain A and B (contoured at 2.5 σ); The helices could be modeled due to stabilizing crystal contacts. **f,** Electron density maps covering the three different domain types as indicated (left: 2 F_{obs}-F_{calc} at 1.0 σ, middle: bias-reduced density modified NCS average map at 1.0 σ, right: bias-reduced density modified NCS average map at 1.0 σ, with additional details revealed by applying a B-sharpening factor of -80 $Å^2$).



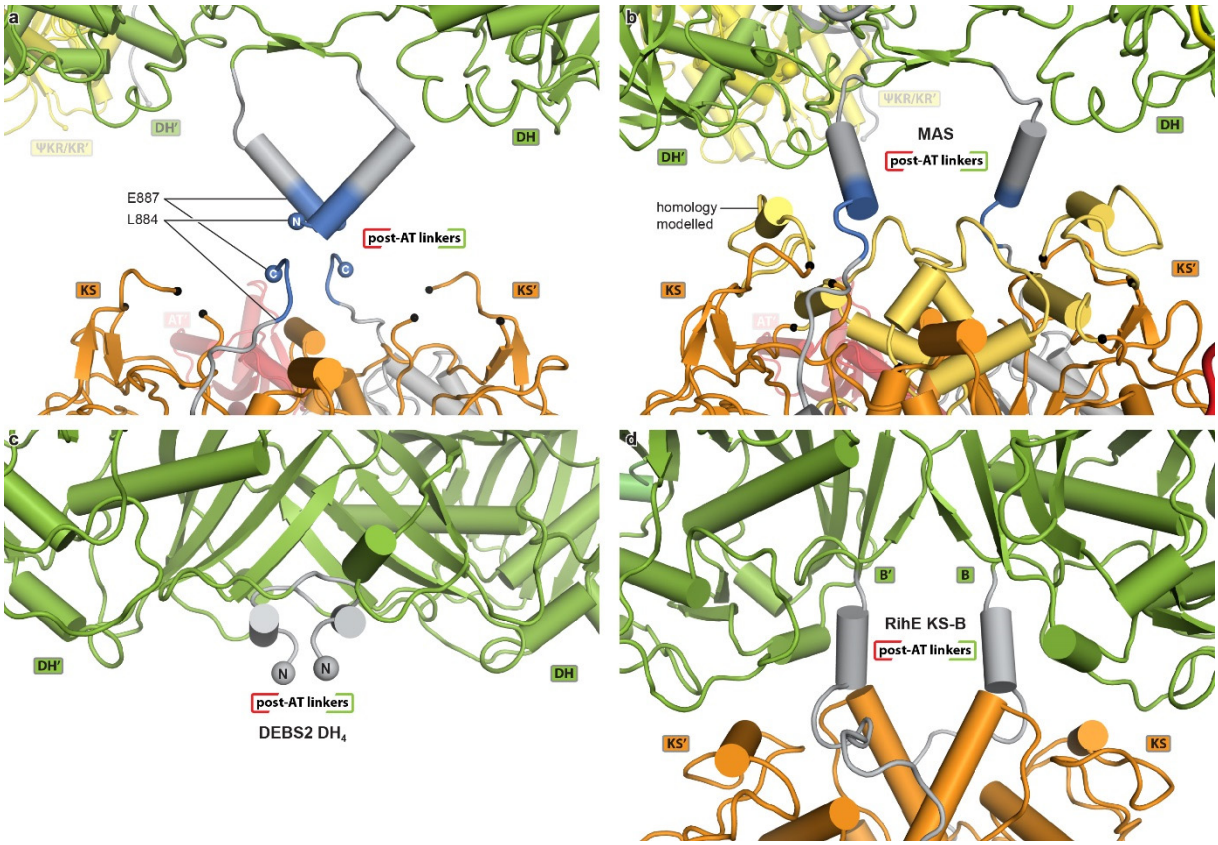
Extended Data Figure 2.7 | Active site and structural comparison of the MAS ER and ΨKR/KR domains.
a, The MAS ER active site tunnel (white) is lined by an NADP$^+$ cofactor. **b**, An F_{obs}-F_{calc} shaked omit map (contoured at 3.0 σ) is shown for the NADP$^+$ cofactor in chain J. **c**, The ER domains of FAS[162], MAS, and the modPKS PpsC dimerize via continuous β-sheet formation between the nucleotide binding subdomains (ER_{NB}), whereas the SpnB ER was crystallized as monomer and represents a group of isolated ER domains[47,198]. **d**, The active site of ΨKR/KR locates to an elongated surface groove, which partially extends to the ΨKR domain and is presumably closed upon ligand binding by a disordered lid region (aa 1948-1960). An F_{obs}-F_{calc} omit map (contoured at 3.0 σ) is shown for the partially ordered NADP$^+$ cofactor. Left: surface, right: cartoon representation. **e**, MAS (pale yellow)

features an N-terminal β_1-α_1-β_2-α_2 extension of the ΨKR Rossmann-fold, which is commonly found in PKSs (violet: Tylosin PKS ΨKR_1[73]), but absent in FASs (green: porcine FAS (pFAS) ΨKR[162]). Secondary structure labels refer to MAS ΨKR. **f**, Average main chain B-factors across all chains reveal distally increasing flexibility with highest B-factors for the ΨKR domain, in particular its β-α-β-α extension, and the C-terminal ACP anchor.

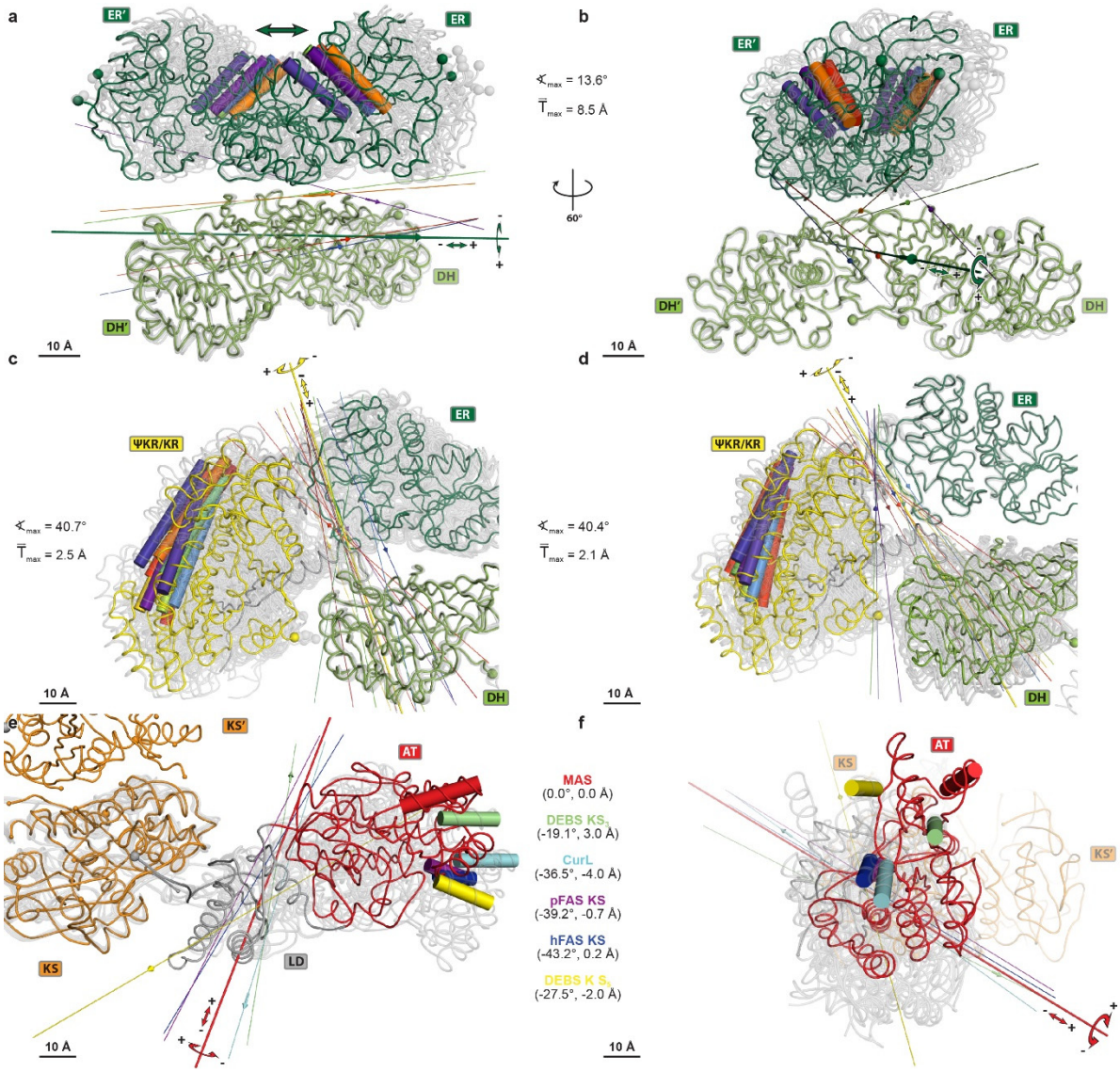


Extended Data Figure 2.8 | Alignment of linker regions of 55 fully reducing modifying regions of PKSs and FASs. The alignment reveals sequence conservation of the β-sheet B1 (β1 and β2), which is inserted in a surface groove of the ΨKR/KR domain. In MAS, strands β3 and β4 form the second antiparallel β-sheet B2. The ER-KR linker is considerably shorter in a subgroup of modPKSs. Sequence numbers and secondary structure elements correspond to *M. smegmatis* MAS (MAS (Ms) highlighted in orange). All modules are labeled as: Protein name (organism abbr.) Uniprot number. Modules of Msl-PKSs (green text), modPKSs (light green), fiPKSs (blue), and FASs (yellow) are grouped by phylogeny (for details and color coding see Extended Data Figure 2.11). PDB

identifiers are indicated in the boxes representing the corresponding domains. Amino acids are shown in Clustal colors. (*, diketide synthase; †, PKS cluster contains non-colinear iterative modules; ‡, modular non-colinear iPKS module ; §, *trans*-AT PKS)

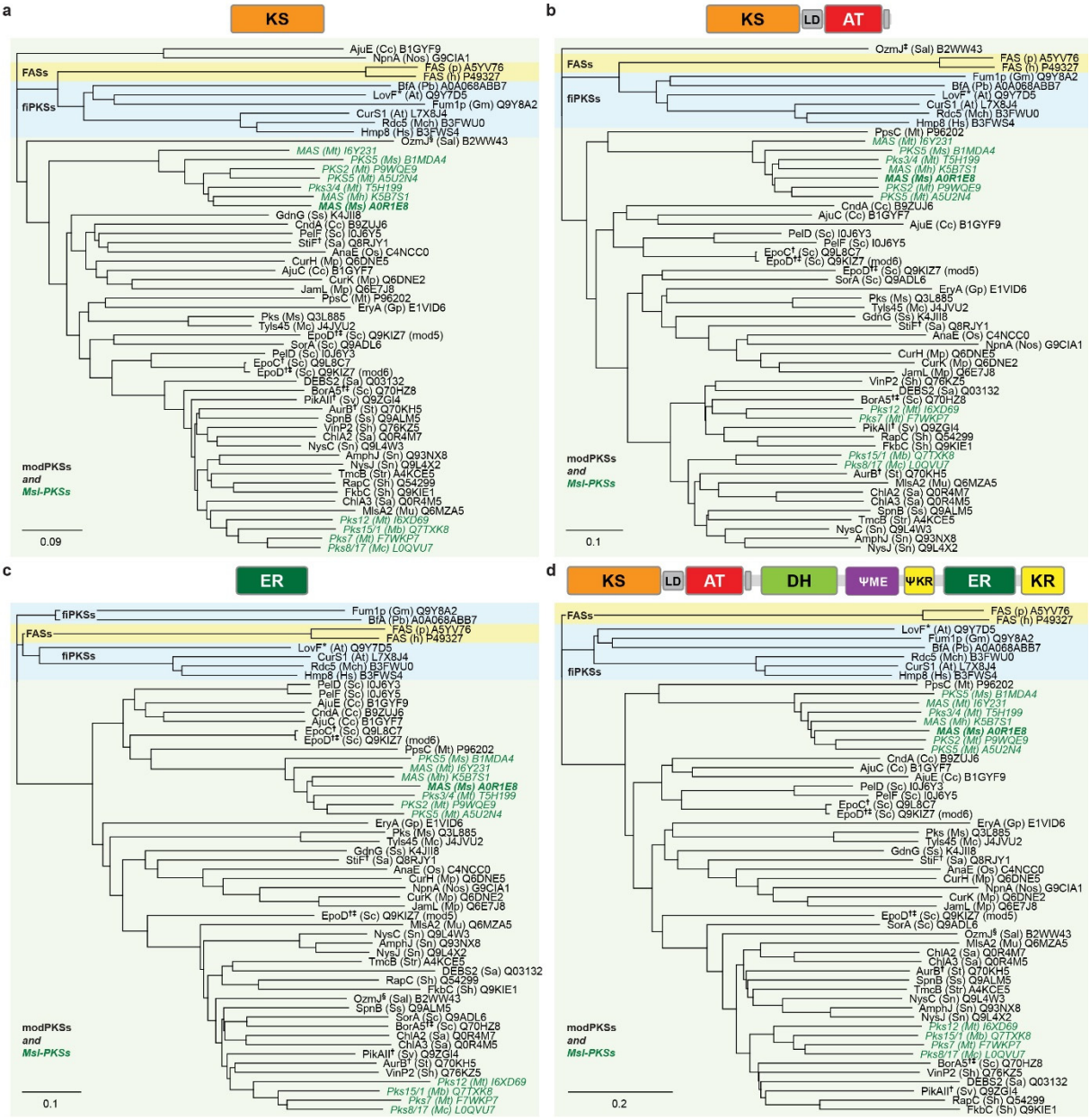


Extended Data Figure 2.9 | Helical organization of central linking segments in MAS and modPKSs.
a, Assembly of the MAS central linking region from authentic crystal structures of the condensing and modifying regions. The two structures overlap in sequence by four residues (blue). **b**, Hybrid model based on the homology completed KS dimer and reconnected helical linkers. Ends of loops defined by the KS-AT crystal structure are indicated by black spheres. Disordered segments in the dimeric condensing region are reconstructed by multi-template homology modelling (gold); color coding is as in a. **c**, **d,** Helix formation in sequence regions corresponding to central linkers are also observed in the isolated crystal structure of the modPKS DH domain of the fully-reducing DEBS module 4[175] (**c**), $RifDH_{10}$[179] (not shown) and in the crystal structure of the RhiE KS-B didomain[131] (**d**), where a KS domain is connected directly to a DH homologous domain, the B domain.



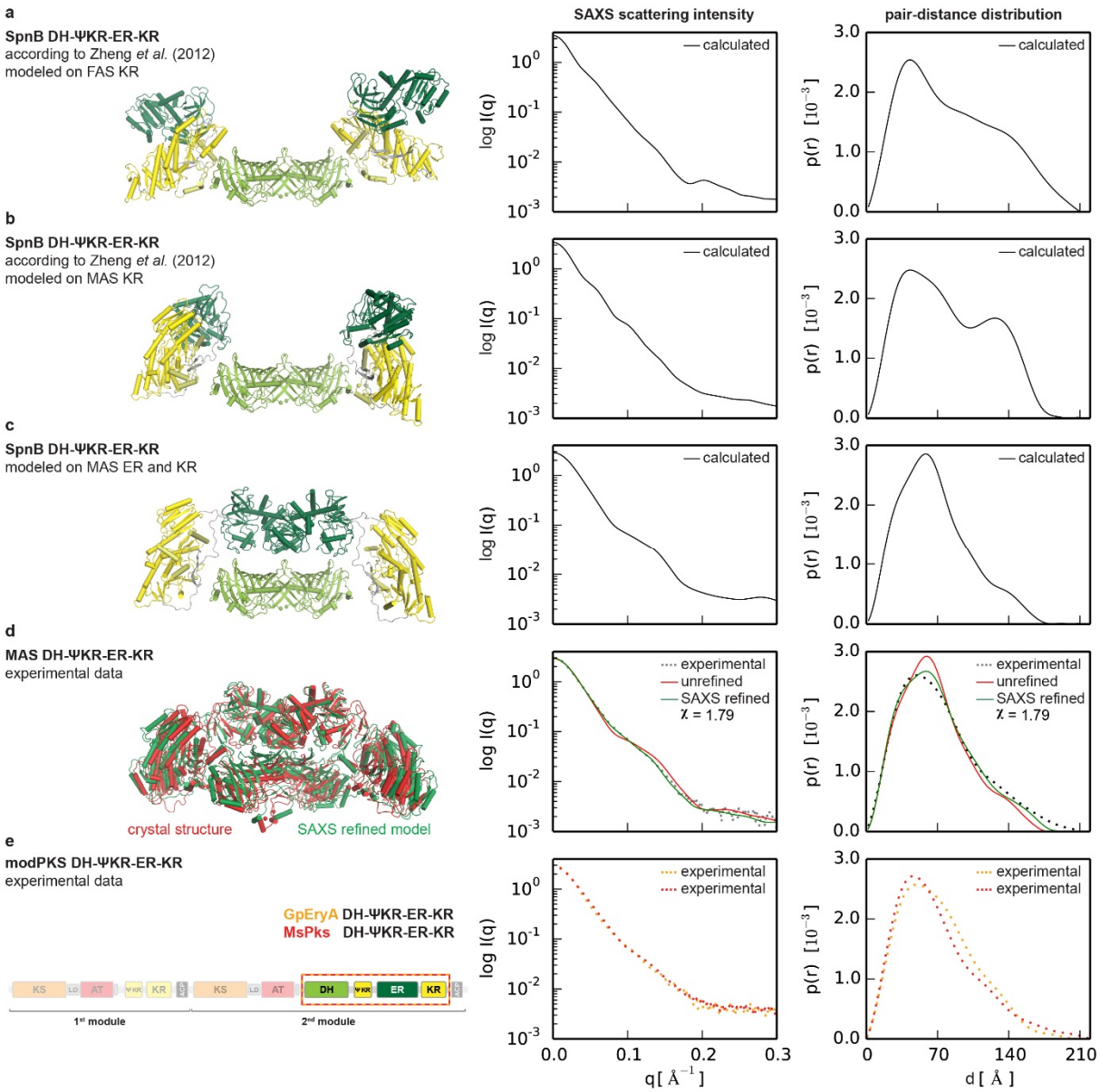
Extended Data Figure 2.10 | Analysis of structural variability in the modifying and condensing regions of MAS and related multienzymes.

a-d, Analysis of interdomain conformational variability between the 18 protein chains in the MAS modifying region crystal structure. **a, b**, Variability of ER positioning relative to DH from two perspectives reveals a screw axis motion combining translation of up to 8.5 Å with rotation of up to 13.6°. **c, d**, Variability of ΨKR/KR domain orientation relative to DH (**c**) and ER (**d**), respectively, reveals a hinge located in the interdomain linker region. **e**, **f**, Top and front view of six overlayed KS-AT didomain structures[46,162-164,172] as indicated and the derived rotational distance of AT positioning around a common hinge in the LD. **a-f**, Relative locations of individual structures are highlighted by representative colored helices. Translational components are indicated with an arrow on the rotation axes with signs indicated on the principle axis (thick, colored according to the moving domain). All structures are aligned to a MAS reference domain (colored ribbon). Rotation axes are shown for rotations larger than 6° and arrows are shown for translations larger than 1 Å.



Extended Data Figure 2.11 | A comprehensive phylogenetic analysis classifies MAS into the branch of modPKSs.

Phylogenetic tree for 55 fully reducing MASs/PKSs/FASs modules were constructed based on only KS domains (**a**), complete condensing regions (**b**), the ER domain (**c**), or all catalytic domains (**d**). *M. smegmatis* MAS (MAS (Ms), bold, italic) and Msl-PKSs (italic) are more closely related to modPKSs (light green) and distinct to fiPKSs (blue) and animal FASs (yellow). All modules are labeled as: Protein name (organism abbr.) Uniprot number. Units are given as amino acid substitutions per site. Indices correspond to Extended Data Figure 2.8Extended Data Figure 2.9.



Extended Data Figure 2.12 | SAXS analysis supports a MAS-like organization of PKS modifying regions. Models (left) of modifying region organization and their respective theoretical and experimental scattering curves as well as pair-distance distributions (right) are shown. **a, b,** As proposed by Zheng *et al.*, the intact SpnB modifying region was modeled based on the domain-swapped SpnB ER-ΨKR/KR structure[47], using either the structure of FAS (**a**) or of the MAS modifying region (**b**) as a guide for positioning KR relative to DH. The SpnB DH structure was generated by homology modelling. **c**, Model of the intact SpnB modifying region with dimeric DH and ER based on the structure of the intact MAS modifying region. **d,** Crystal structure of MAS before and after fitting to experimental SAXS data. A good fit (χ=1.79) is obtained by fitting SAXS data with a single model corresponding to an average conformation of the MAS structure. **e,** Sequence organization of two authentic modPKS modifying regions of similar ER-KR linker length to SpnB (left), together with experimental SAXS scattering data (right). The data closely match calculated scattering curves for a MAS-like architecture, but disagree with models based on a monomeric ER as suggested for SpnB.

2.8 Supplementary Information

Supplementary Video 1 | Conformational variability and coupling in the MAS modifying region. All dimeric modifying regions of the MAS crystal structure were aligned to the DH domains and combined into one animation. The ER dimer moves in a screw motion with a lateral translation of up to 8.5Å and a rotation of up to 13.6° on the dimeric DH platform. The ER motion is linked to a rotation of the laterally double-tethered ΨKR/KR domains and couples the conformations of the ΨKR/KR across the MAS modifying region dimer. The maximum observed rotation of the ΨKR/KR domains (40.4°) causes an active site distance shift of 10 Å (euclidean space) relative to the DH domain.

http://www.nature.com/nature/journal/v531/n7595/full/nature16993.html#videos

Supplementary Video 2 | Condensing region conformations in crystal structures of PKSs and FASs. Structures of the homologous condensing regions are aligned and animated from MAS over PDB: 2QO3, 2HG4, 4MZ0, 2VZ9 to 3HHD. The superposition indicates a common hinge for rotational mapping of the AT domain positions located in the linker domain (grey). MAS adopts the most linear arrangement of all AT domains, which differs by a rotation of 43.2° around the common hinge from those observed in the crystal structure of human FAS (PDB: 3HHD).

http://www.nature.com/nature/journal/v531/n7595/full/nature16993.html#videos

2.9 Corrigendum: Mycocerosic acid synthase exemplifies the architecture of reducing polyketide synthases

This corrigendum was originally published in Nature[195].

Dominik A. Herbst*, Roman P. Jakob*, Franziska Zähringer & Timm Maier.

*Contributed equally

Nature. 2016 Aug 18;536(7616):360

Herbst, D. A., Jakob, R. P., Zahringer, F. & Maier, T. Corrigendum: Mycocerosic acid synthase exemplifies the architecture of reducing polyketide synthases. Nature 536, 360, doi:10.1038/nature18281 (2016).,

In this Letter, we studied the three-dimensional structure of a protein from *Mycobacterium smegmatis* assigned as mycocerosic acid synthase (MAS) in sequence databases as A0R1E8 in Uniprot (http://www.uniprot.org/uniprot/A0R1E8) and YP_888986.1 in NCBI (https://www.ncbi.nlm.nih.gov/protein/118473069). In conclusion, we provided a template structure of MAS-like polyketide synthases (PKSs) and a first example of reducing PKS architecture. However, we now note that Etienne *et al.*[104] provided a biochemical characterization of a deletion strain of the corresponding gene MSMEG_4727 (https://www.ncbi.nlm.nih.gov/gene/4534621), which indicated a physiological role of the protein in the production of 2,4-dimethyl-2-eicosenoic acid, a lipid component of lipooligosaccharides, rather than mycocerosic acids, via a reaction closely related to those of MAS. Until comprehensive characterization at the protein level is available, the protein we studied should therefore be referred to as a 'mycocerosic-acid synthase like-PKS' or 'MAS-like PKS'; the database records will be updated accordingly. We thank the authors of ref.[104] for drawing our attention to this publication. The main scientific conclusions of our manuscript remain unchanged.

3 FUNCTIONAL AND STRUCTURAL ANALYSIS OF PROGRAMMED C-METHYLATION IN THE BIOSYNTHESIS OF THE FUNGAL POLYKETIDE CITRININ

This research was originally published in Cell Chemical Biology[236].

Philip A. Storm, Dominik A. Herbst, Timm Maier, and Craig A. Townsend

Cell Chem Biol. 2017 Mar 16;24(3):316-325.

Reprinted from

Storm, P. A., Herbst, D. A., Maier, T. & Townsend, C. A. Functional and Structural Analysis of Programmed C-Methylation in the Biosynthesis of the Fungal Polyketide Citrinin. Cell Chem Biol 24, 316-325, doi:10.1016/j.chembiol.2017.01.008 (2017),

3.10 Article

3.10.8 Summary

Fungal polyketide synthases (PKSs) are large, multi-domain enzymes that biosynthesize a wide range of natural products. A hallmark of these megasynthases is the iterative use of catalytic domains to extend and modify a series of enzyme-bound intermediates. A subset of these iterative PKSs (iPKSs) contains a *C*-methyltransferase (CMeT) domain that adds one or more *S*-adenosylmethionine (SAM)-derived methyl groups to the carbon framework. Neither the basis by which only specific positions on the growing intermediate are methylated ("programming") nor the mechanism of methylation are well understood. Domain dissection and reconstitution of PksCT, the fungal non-reducing PKS (NR-PKS) responsible for the first isolable intermediate in citrinin biosynthesis, demonstrates the role of CMeT-catalyzed methylation in precursor elongation and pentaketide formation. The crystal structure of the *S*-adenosyl-homocysteine (SAH) coproduct-bound PksCT CMeT domain reveals a two-subdomain organization with a novel N-terminal subdomain characteristic of PKS C-methyltransferase domains and provides insights into cofactor and ligand recognition.

3.10.9 Introduction

PKSs and fatty acid synthases (FASs) utilize common strategies to produce an enormous number of natural products with highly diverse chemical scaffolds, including pigments, environmental toxins, and pharmacologically active substances[143,237]. Simple precursor acyl substrates are tethered to the acyl carrier protein (ACP) by thioesterification of a post-translational phosphopantetheine arm. They are then extended by the ketosynthase (KS) through a determined number of decarboxylative condensation reactions with ACP-bound malonyl extender units supplied by the malonyl-CoA:ACP transacylase (MAT). In many cases, the newly extended intermediate is delivered to additional tailoring domains that perform programmed modifications at the α- or β-carbon. Modular PKSs (modPKSs) generally have a distinct module for each round of extension and modification, whereas iPKSs use the same domains repeatedly a defined number of times. How iPKS domains control which tailoring domains are used after a given round of extension is still largely unanswered and poses a significant challenge to engineering efforts.

Perhaps the simplest case of programmed tailoring among PKSs occurs during iterative *C*-methylation of fungal non-reduced polyketides, exemplified by the group VI and VII, or Clade III, families where a tetra- or pentaketide is *C*-methylated one or more times[238,239]. Each round of extension produces a potentially enolizable nucleophilic α-carbon, but only some such positions are methylated. A more complete understanding of how methylation is programmed and the structural basis by which methylation occurs will be essential for future efforts to rationally influence product methylation patterns. Site-selective polyketide *C*-methylation would be a powerful tool to engineer new biosynthetic pathways, and many methyltransferases can accept SAM analogs with larger alkyl donors, further expanding potential scaffolds[240].

We selected PksCT as an initial candidate to explore non-reducing PKS (NR-PKS) *C*-methylation[241]. Responsible for the first isolable intermediate in citrinin (**1**) biosynthesis[242], PksCT produces a triply methylated pentaketide that is C2-C7 cyclized by the product template (PT) domain and released as the benzaldehyde (**2**) by the *C*-terminal reductase domain (R) (Figure 3.1). Previous efforts to understand NR-PKS programming have benefited greatly from a deconstruction approach that allows for the extension domains to be separated from the tailoring domains[48,85,91]. A growing number of

excised PKS domains have been structurally characterized, providing the molecular details of catalysis and substrate selection[243]. We sought to expand this experimental approach to PksCT in order to isolate CMeT domain activity and explore the structural basis for NR-PKS *C*-methylation programming.

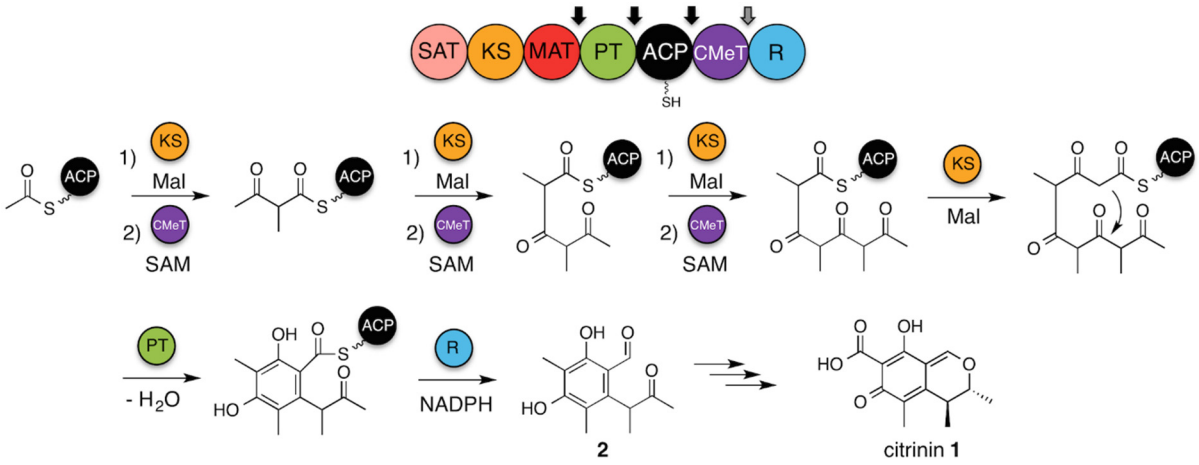


Figure 3.1 | Proposed biosynthesis of 2 by PksCT.
The domain architecture of PksCT is shown above, with arrows indicating points of domain dissection. Mal indicates a unit of ACP-bound malonyl used for decarboxylative condensation prior to methylation. Necessary domains and substrates are shown above and below the reaction arrows. See also Figure S3.6 and Item S1.

3.10.10 Results

Domain deconstruction and reconstitution of PksCT

Using the Udwary-Merski algorithm[244] to recognize interdomain linker regions in a set of related proteins, a domain deconstruction approach was applied to PksCT and used to generate mono- and multidomain fragments (Figure S3.6A and Supplemental Procedures). All constructs gave soluble protein except for those containing the SAT domain in initial expressions. When multiple domain boundaries did not improve expression, we reassessed the reported exons using the FGENESH exon prediction suite[245]. These results suggested alternate exon boundaries, which would include a conserved Trp195 in the SAT domain that was not present in the original annotated sequence. Total RNA was isolated from wild-type Monascus purpureus grown under citrinin producing conditions[241,246]. The cDNA was synthesized, and a region containing the exon boundaries was amplified and subcloned. Five clones were sequenced and all verified that the revised prediction generated the correct exon boundaries (Figure S3.6B and Item S1). We propose an update to GenBank: AB167465, joining gene fragments 1197-1835 and 1892-9034. PksCT residue numbering used hereafter reflects this revision.

Earlier reports speculated that PksCT could use several potential starter units[247], but heterologous *in vivo* extracts suggested an acetyl starter[242]. We confirmed this mode of initiation *in vitro* using a radiochemical assay[87]. PksCT SAT was capable of loading an acetyl starter unit from [1-^{14}C]-acetyl-CoA and transferring it to PksCT *holo*-ACP (Figure S3.6C). We then conducted a series of reconstitution reactions to identify the activity of the PksCT CMeT domain (Figure 3.2 and Figure S3.7, Table S3.2 and Table S3.4). Each individually expressed and purified fragment was combined as indicated with all of the expected substrates. Acetyl and malonyl units were included as the *N*-acetylcysteamine thioesters (SNAC), which serve as substitutes for the CoA thioesters used *in vivo*[48]. Stock solutions of SAM and NADPH were freshly prepared.

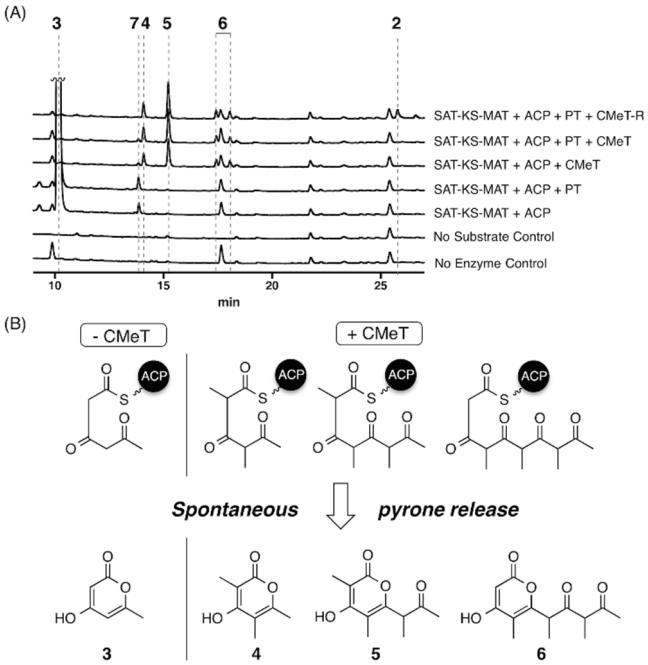


Figure 3.2 | *In vitro* reconstitution of PksCT.
(A) HPLC absorbance traces are shown at 280 nm for the combination of domains indicated to the right. Traces are vertically offset and peak **3** is truncated for clarity. See also Figure S3.6and Table S3.2. (B) Spontaneous pyrone release of intermediates as tri-, tetra-, and pentaketide pyrones. High resolution UPLC-ESI-MS and UV-Vis data can be found in Table S3.4 and Figure S3.7.

The minimal PKS, SAT–KS–MAT and *holo*-ACP, with acetyl and malonyl substrates primarily generated triketide **3**. Upon addition of PT, the product profile was not substantially different and no C2-C7 cyclized product was observed, even after attempts to chemically release any potentially cyclized intermediate with base or cysteamine[248]. Addition of CMeT and SAM to the minimal PKS, however, resulted in the loss of **3** but the appearance of four new species corresponding to spontaneous pyrone release of methylated tri-, tetra-, and pentaketides **4**, **5**, and **6**. The pentaketide **6** eluted as two peaks, identical by mass, that appear to be diastereomers as each collected peak equilibrates to the same two components upon isolation and reinjection. The minimal PKS reactions also contained small amounts of **7**, having a mass consistent with a singly methylated form of **3**. The disappearance of **7** in reactions containing CMeT suggests that it results from an on-pathway intermediate to **4**, **5**, and **6**, and is discussed in detail below.

Full extension of the polyketide chain appears to rely upon the correct methylation of the growing linear intermediate, a feature seen previously in studies of the highly-reduced polyketide lovastatin[70,249,250]. This series of intermediates indicates that methylation occurs at the thioester α-carbon iteratively following each extension and not processively on a fully elongated pentaketide. These methylated derailment products are observed in the presence of the PT and R domains as well. With all domains, AcSNAC, MalSNAC, SAM, and NADPH present, the expected post-PKS aldehyde **2** was observed also, albeit in low yield.

Several efforts to improve the efficiency of reconstituted PksCT were attempted, including addition of methylthioadenosine nucleosidase (MTAN)[251] to hydrolyze the potentially inhibitory SAH co-product or inclusion of Mg^{2+}, a cofactor used by some MTases[252]. Previous work has demonstrated that NR-PKS deconstruction reduces the yield of post-PKS product and increases the abundance of derailment products[48,253]. In the case of deconstructed PksCT, it is possible that transfer of methylated polyketides

from the KS active site to the CMeT active site allows for racemization of the methyl groups. Whether the KS, PT, or R domain active sites are structured for a specific methyl stereochemistry, and whether polyketides with scrambled methyl stereochemistry are inhibitory, is currently unknown. Additionally, *in vivo* experiments of citrinin biosynthesis indicate that co-expression of an adjacent gene, encoding the hydrolase CitA, significantly improves titres of **2**[242], though the specific role of CitA is still unclear.

Structure of PksCT CMeT in complex with SAH

The 1.65 Å crystal structure of the CMeT domain reveals an organization of two subdomains with the active site located at their interface (Figure 3.3A). PksCT CMeT was co-crystallized with the stable co-product S-adenosylhomocysteine (SAH) in rhombohedral space group H3 containing one protomer per asymmetric unit (Table S3), the structure was solved by SAD phasing and refined to R/R_{free}: 0.19/0.22.

The 202 amino acid (aa) *C*-terminal subdomain (residues 1959-2160) reveals a class I methyltransferase fold[254], which is the most common methyltransferase fold in natural product biosynthetic enzymes, despite diverse acceptor substrates and low sequence conservation (Liscombe et al., 2012) (Figure 3.3B). The closest structural homologs are the SAM-dependent methyltransferase from *Aquifex aeolicus* (unpublished, PDB:3DH0), the methyltransferase domain of bacterial-AvHen1-CN (PDB: 3JWJ), a putative methyltransferase from *Sulfolobus solfataricus* (unpublished, PDB: 3I9F), NodS from *Bradyrhizobium japonicum* WM9 (PDB: 3OFK) and the methyltransferase domain of bacterial-CtHen1-C (PDB: 3JWG) with a C_α rmsd of 2.0-2.4 Å and 160-195 aligned residues. The catalytically inactive pseudo-methyltransferase (ΨCMeT) of the human fatty acid synthase (FAS) aligns with a C_α rmsd of 2.5 Å (158 aligned aa)[201] and is the closest structural homolog in carrier protein-dependent multienzymes. The *C*-terminal subdomain integrates into the *N*-terminal subdomain through a region around helix 15 that forms the base of the ligand binding tunnel, resembling an open palm ("palm helix region"), and is structurally conserved in mammalian FAS[162,201], but deleted in insect FAS and some highly reducing-PKSs (HR-PKS) (Figure S3.8).

The 144 aa *N*-terminal CMeT subdomain (residues 1815-1958) shares an interface of 1,515 $Å^2$ with the *C*-terminal subdomain and adopts an uncharacterized helical fold (Figure 3.3B). Such *N*-terminal subdomains of Class I methyltransferases that cap the SAH/SAM-binding pocket often serve a role in acceptor substrate selection[255]. In the related mammalian FAS multienzymes, which comprise catalytically inactive ΨCMeTs domains and provide the only structural depiction of CMeT integration into a multienzyme, this region is considerably truncated and disordered[162,201]. Nevertheless, the *N*-terminal CMeT subdomains in FASs reveal helices at equivalent positions to helix 4, 7 and 9/10 with a C_α rmsd of 2.5 Å over 42 aligned aa, indicating a common evolutionary origin (Figure S3.9A). The *N*-terminal subdomain is connected to the preceding protein regions in PksCT by a 30 aa long linker containing two helices and wraps around the *C*-terminal subdomain in a groove with an interface area of 1,461 $Å^2$. In mammalian FASs[162,201] the *N*-terminal linker leading into ΨCMeT contacts this domain by appending a β-strand to the central β-sheet of the C-terminal subdomain (Figure S3.9B). In PksCT, however, the *N*-terminal linker (residues 1785-1814) wraps around the *C*-terminal subdomain in a prominent surface groove (Figure S3.9C). Consequently, in contrast to mammalian FAS, the *N*- and *C*-terminal linker ends are located in close proximity to each other (Figure S3.9C), indicating an alternate integration of CMeT in the PksCT multienzyme.

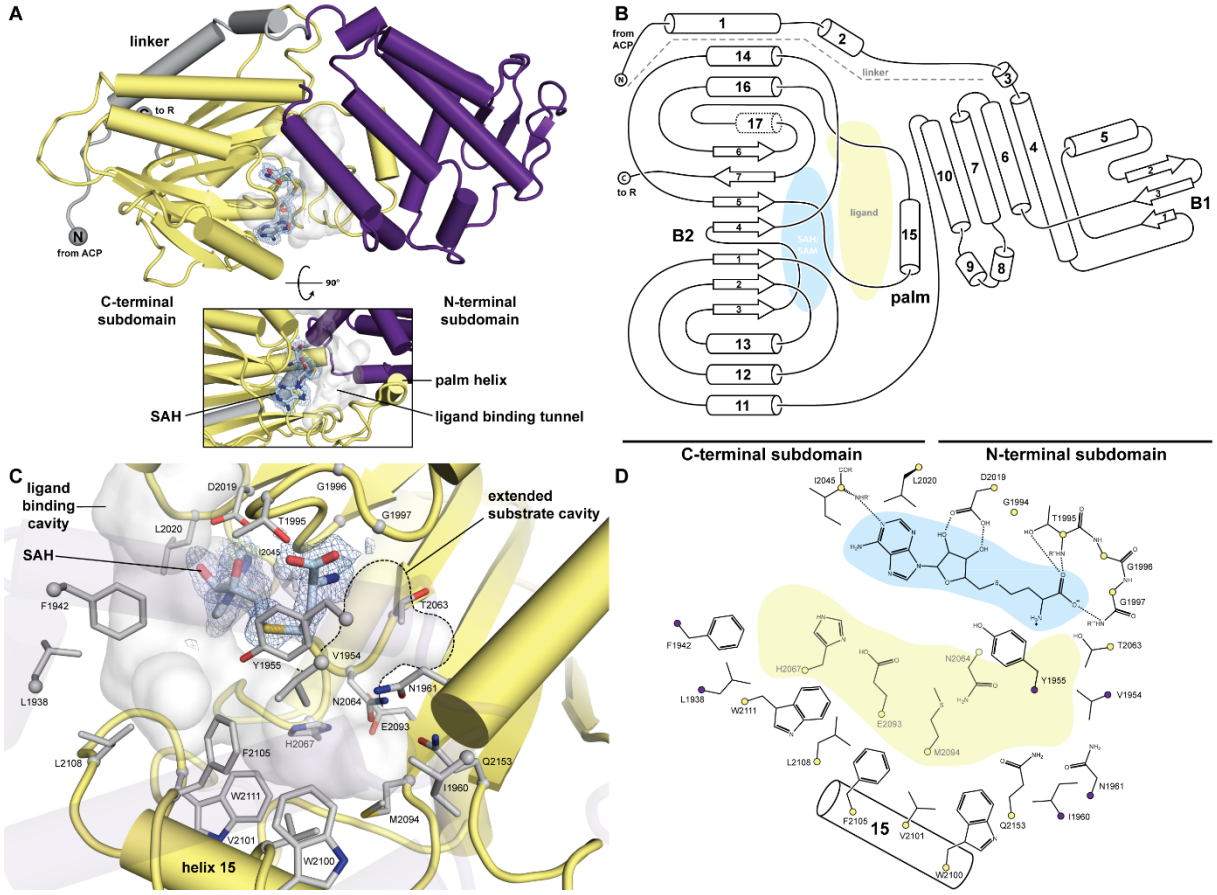


Figure 3.3 | Crystal structure and ligand binding site of CMeT.
(**A**) CMeT is organized into an N-terminal linker (grey), an N-terminal subdomain (violet) and a C-terminal subdomain (yellow). The active site is located at the subdomain interface, Fo-Fc omit difference density at 2.5 σ level is shown for SAH. See also Table S3.3 and Figure S3.9. (**B**) Topology of CMeT highlighting domain organization and substrate binding sites. α-Helices are numbered, β-strands are numbered relative to their position in the respective β-sheet B1 or B2. Helix 17 (dotted) is a 3_{10} helix. (**C**) Substrate interactions and ligand binding tunnel. The C-terminal subdomain laterally binds SAH and forms an active site tunnel (grey surface) along SAH. The back of this tunnel is lined with hydrophobic residues of the palm helix region and is closed by the N-terminal subdomain. The invariant residues Tyr1955 as well as His2067 together with Glu2093 face the ligand binding tunnel from opposite sides and are involved in catalysis. SAH contributes to the formation of an extended cavity for binding larger substrates. Difference density is depicted as in A. See also Figure S3.10. (**D**) Schematic active site representation. Hydrogen bonds are indicated by dotted lines. C_α atoms are shown as spheres in the color of their respective subdomain.

The active site of PksCT CMeT is located at the interface of the *N*- and *C*-terminal subdomains. Most of the active site is formed by the *C*-terminal subdomain, which binds the SAH parallel to the ligand binding tunnel, which is lined with hydrophobic residues of the palm helix region. Helix 9 and 10 of the N-terminal subdomain contributes four residues to the distal region of the ligand binding tunnel to form the opposite face (Figure 3.3C and D), and the highly conserved Tyr1955 splits the active site into a tunnel for the SAM co-substrate and the polyketide substrate. Remarkably, the homocysteine moiety of the SAH contributes to the formation of a pocket for extended substrates, which could accommodate varying polyketide chain lengths during elongation cycles 1-3.

The SAH is tightly bound at an interface of 356 Å^2 along one side of the co-substrate binding pocket via highly conserved residues: A GxGxGG (residues 1992-1997) motif forms hydrogen bonds with the homocysteine moiety, Asp2019 binds the ribose through two hydrogen bonds and the adenine moiety

is positioned in a hydrophobic pocket formed by Leu2020 and Ile2045 with a backbone hydrogen bond to Ile2045.

The active site tunnel reveals a funnel like shape and branches off into the extended substrate cavity at its inner end, providing a central channel for Cα substrate alignment with regard to different polyketide chain length (Figure 3.3C and Figure S3.10). The central channel is surrounded by a patch of hydrophobic residues (Figure 3.3C and D); the highly conserved Phe1942 as well as Leu1938 line the substrate entrance region and the conserved residues Leu2108 and Trp2111 and residues Trp2100, Val2101, Phe2105 of the palm helix would likely interact with the linear polyketide substrates (Figure S3.8). Together, these residues form a hydrophobic patch along which the hydrophobic backbone of substrates at various elongation states might slide into the active site. Interestingly, the residues involved in the hydrophobic patch are highly conserved among other NR-PKS CMeT domains – such as AusA, PkeA, MpaC, and DtbA – that install different methylation patterns than PksCT[238{Regueira, 2011 #1948,256] and are functionally conserved in both *cis*- and *trans*-AT PKSs (Figure S3.8).

Although no obvious delineation amongst the NR-PKS CMeTs correlated with observed product methylation pattern, the number of available sequence–product pairs may be insufficient to identify specific programming motifs. Furthermore, a phylogenetic analysis of CMeT domains across multiple PKS/FAS families clearly demonstrated a branching between active CMeT and inactive ΨCMeTs (Figure 3.4). Active CMeTs clustered into three groups: *trans*-AT PKSs, *cis*-AT and HR-PKSs, and NR-PKSs. Active HR-PKS CMeT domains are nested among *cis*-AT CMeTs, perhaps reflecting similar embedding amongst reductive domains not present in NR-PKSs.

Catalysis by CMeT

For catalysis, the α-carbon of the acceptor substrate presumably must be deprotonated to generate the enolate nucleophile for S_N2-like attack at the methyl donor. The invariant residues Tyr1955 and His2067 together with Glu2093 face the ligand binding tunnel from opposite sides and are well positioned to act as catalytic residues. His2067 forms a hydrogen bridge with Glu2093 and might act as catalytic dyad for enolization, as previously proposed for the homocitrate synthase[257] (Figure 3.5B).

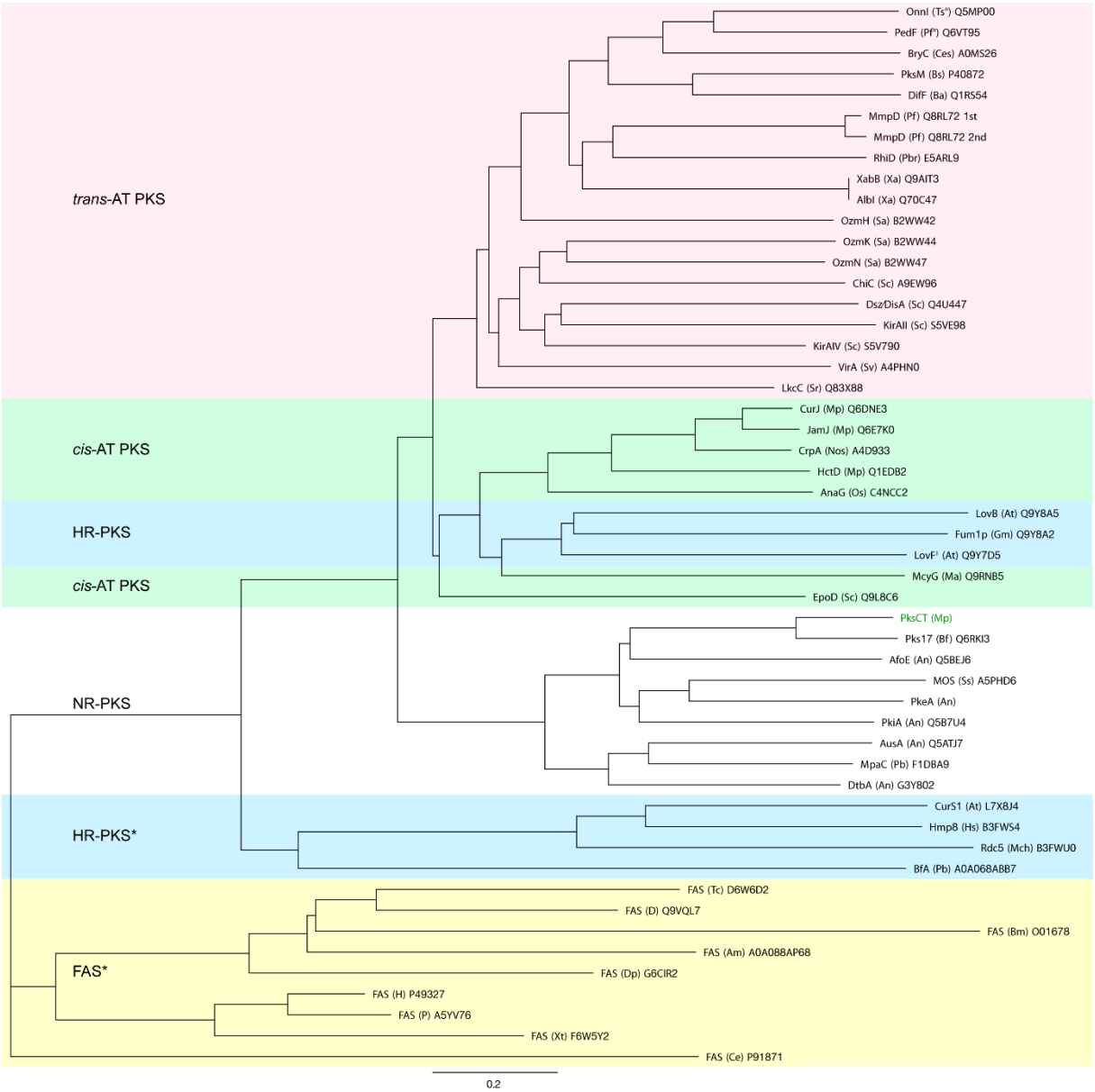


Figure 3.4 | Phylogenetic analysis of 51 CMeT domains of PKS and FAS.
38 active CMeT domains of PKS and 13 inactive ΨCMeT domains of HR-PKS and FAS were aligned and phylogenetically analyzed (see Figure S3.8). Multienzyme family classifications are indicated in colored groups. Units are given as amino-acid substitutions per site. All sequences are labelled as "protein name (organism abbreviation) Uniprot number". The sequence of PksCT corresponds to Item S1 (°, endosymbiont of this organism; ‡, diketide synthase; *, inactive ΨCMeT domain)

Reconstitution of the minimal PksCT with the CMeT His2067Ala mutant primarily gave triketide **3**, matching the profile of the minimal PKS without CMeT (Figure 3.5A). This experimental observation demonstrated the essential role of His2067 for efficient methyl transfer and supports its likely role as the catalytic base. The more conservative His2067Gln mutant[258,259] also produced **3** as a major product in addition to smaller amounts of two different triketides **7** and **8** identical by mass, but singly methylated at C-4 and C-2, respectively. Tyr1955Phe and Tyr1955Ala mutations appeared to have a lesser effect on catalysis. Neither produced significant amounts of **3**, but instead yielded small amounts each of **4**, **5**, **7**, and **8**. The reduced amounts of **5** and the absence of **6**, compared to wild-type, suggest that Tyr1955 may be especially important for methylation of tetra- and pentaketide substrates.

Because **7** was observed also in the absence of CMeT, but appears to be on-path to **4**, **5**, and **6**, we speculate that some SAT–KS–MAT may be purified with a methylacetoacetyl unit bound to Cys555 in the KS active site. Such a substrate could arise from direct loading from methylacetoacetyl-CoA, an intermediate in isoleucine catabolism[260], or perhaps homologation of KS-bound acetyl by endogenous methylmalonyl-CoA during heterologous expression [for closely related biochemistry see:[261,262]]. A single round of extension of methylacetoacetyl, followed by spontaneous pyrone formation would result in **7**, methylated at C-4. The decreased activity of the CMeT His2067Gln mutant could allow for a relative increase in **7** and the appearance of **8**, methylated at C-2, where the first potential methylation is skipped but the second methylation is successful. It is known that polyketide extension is exceedingly rapid in NR-PKSs[85,94]. Kinetic competition between chain elongation and programmed methylation where the latter is impaired in the His2067Gln mutant could lead to expected alkylation at C-4 to be missed.

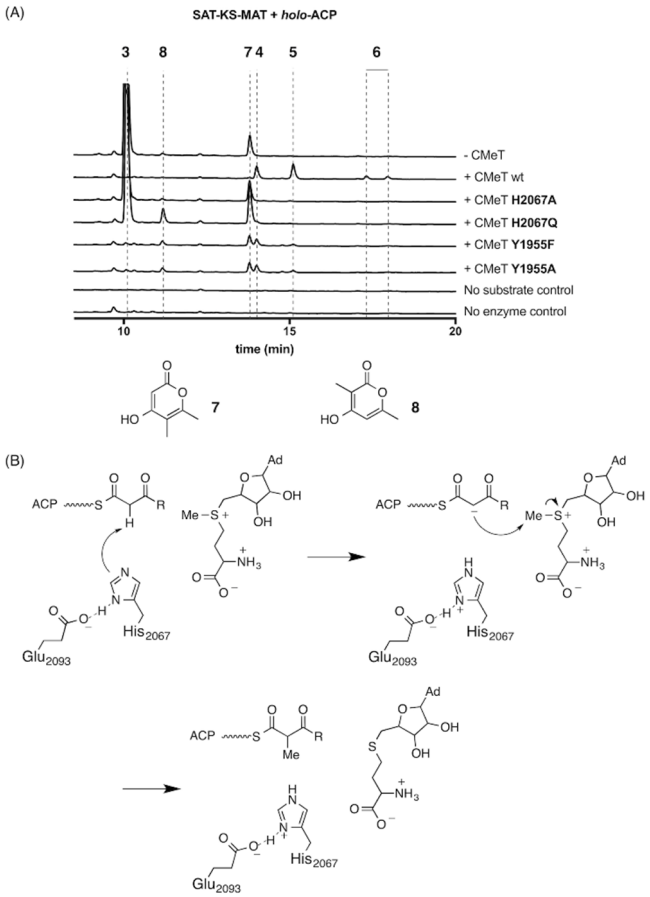


Figure 3.5 | PksCT CMeT His2067 is essential for methyl transfer and positioned to act as the catalytic base. (**A**) *In vitro* reconstitution of PksCT SAT–KS–MAT and *holo*-ACP with CMeT mutants show that methylated products **4**, **5**, and **6** are not produced by His2067 mutants. Tyr1955 mutants are capable of generating methylated triketides, but not pentaketides, suggesting a role for this residue in acceptor substrate binding. Absorbance traces are shown at 280 nm for the CMeT variant indicated on the right and vertically offset; peak **3** is truncated for clairty. For mass and UV-Vis absorption data, see Table S3.4 and Figure S3.7. (**B**) Proposed mechanism for methyl transfer. His2067 forms a catalytic dyad with Glu2093 and deprotonates the α-carbon to generate an enolate nucleophile capable of S_N2-like attack at the methyl donor. Completion of the catalytic cycle by loss of the removed proton to solvent is not explicitly shown. R indicates the potential chain lengths described in Figure 3.1.

3.10.11 Discussion

Methylation is a common and powerful strategy to diversify the structures of microbial natural products. iPKSs incorporate *C*-methyl groups into their product scaffolds early in the biosynthetic pathway because the pK_a of the methylated position is generally never lower or more accessible to enzymatic chemistry than immediately following extension to the β-ketothioester. Early indications of the interplay between KS and CMeT domains were observed in LovB, where the lack of SAM or CMeT resulted in the accumulation of truncated intermediates[70,249]. Recently, it was shown that the LovB CMeT domain outcompetes the LovB KR activity by virtue of its higher catalytic efficiency (Cacho et al, 2015). Non-cognate substrates, due to incorrect modifications in previous cycles, caused significant attenuation of the highly specific CMeT leading to off-loading reactions. This finding suggested a role of CMeT as gatekeeper, which has to be passed after three of six cycles of polyketide chain extension. Unlike for LovB CMeT, PksCT CMeT is not competing with a KR domain and methylates multiple substrates. Therefore, PksCT CMeT appears to have a similar function by adding methyl groups as check-point tags, which are recognized by PksCT KS, such that a lack of methylation causes release of immature products at the triketide stage. Abortive release of an off-pathway intermediate may be especially important for NR-PKSs that lack a *C*-terminal thioesterase domain to serve an editing role[94].

Any future efforts to modulate the pattern of *C*-methylation in iPKSs may require perturbation of both CMeT and KS substrate binding sites. The highly conserved residues of the acceptor binding pocket also indicate that methylation programming is determined by other features of the active site or acceptor:CMeT complex. That programmed methylation is retained even after domain dissection, however, points towards control residing in intrinsic features of the CMeT. Additional studies will be needed to identify and manipulate these features to control programmed methylation. The existence of *gem*-dimethylating CMeTs in *trans*-AT PKSs[69] combined with the iterative action of NR-PKS CMeTs show that considerable flexibility is possible, potentially allowing for engineered products.

Sequence motifs have been identified in other iPKS modification domains that allow for prediction of the stereochemical outcome, notably the A- and B-type ketoreductase (KR) domains[243]. A similar dichotomy for CMeTs, pro-*R* or pro-*S*, has not been identified. The stereochemistry of methylation could not be inferred from our data, and in the context of NR-PKSs this information is lost upon aromatization. Future work should provide greater detail about the methyl transfer mechanism.

Structure determination of the PksCT CMeT domain revealed a characteristic organization of the *N*-terminal subdomain in PKS C-Met domains and a detailed description of residues lining the substrate binding site. Structural alignments to mammalian FAS and sequence comparison to modular *cis*- and *trans*-AT PKS demonstrate a conserved organization of inactive ΨCMeT domains, commonly observed in HR-PKS and FAS and active CMeT domains in all systems. The PksCT CMeT structure thus provides a blueprint for the systematic analysis and variation of CMeT functions across all PKS systems. Despite a high degree of conservation at the level of subdomain organization and active site assembly, the organization of *N*-terminal linkers in the CMeT domain of the NR-PKS PksCT clearly deviates from those observed for the ΨCMeT domain of fully-reducing mammalian FAS, indicating distinct modes of integration in the two systems. The exact modes of domain integration (and the extent of their conservation) of CMeT domains in systems other than FAS, namely iterative and modular *cis*- and *trans*-AT PKS, remains to be explored. First indications may be obtained from a careful comparison of flanking regions in excised CMeT domains.

In connection with these considerations, while this manuscript was under review, a structure of an excised CMeT from a modular PKS was determined and closely agrees with the present data on a comparable domain from an iterative PKS[68].

3.10.12 Significance

Iterative PKSs reuse in a "programmed" manner their complement of catalytic domains to selectively synthesize a defined product. The non-reducing iterative PKSs have no tailoring domains that carry out the reduction and dehydration steps during polyketide chain extension characteristic of FASs and other PKS subclasses. In the absence of tailoring domains, the number of cycles of ketide homologation (chain length control) is dictated principally by the ketosynthase domain. *C*-methyltransferase domains distinguish the Clade III NR-PKSs and afford the simplest systems to study programmed events occurring during iterative polyketide extension. Revealed clearly here is a defined interplay between correctly patterned methylations, which take place only during certain chain elongation cycles, and successful ketide extension. The strong interdependence of ketosynthase and *C*-methyltransferase behavior will be a constraint in the re-programming of iterative PKSs to new synthetic tasks. To guide these efforts and to help understand how methylation impacts substrate elongation and programmed product formation, we obtained the crystal structure of a non-reducing PKS *C*-methyltransferase domain in complex with SAH, providing a structural foothold for analyzing programmed methyl transfer. A previously uncharacterized *N*-terminal subdomain is involved in substrate binding, and sequence comparisons show this subdomain is an indicator of an active *C*-methyltransferase. Several conserved features are present across functionally and biologically diverse PKS families, including an active site histidine necessary for activity and aromatic residues lining the substrate binding pocket. Systematic variation of *C*-methyltransferase domains may be explored for the bioproduction of unnatural polyketides with programmable methylation or alternate alkyl donors.

3.11 Experimental Procedures

3.11.13 Materials

Reagents were purchased from Sigma Aldrich (St. Louis, MO) unless stated otherwise. Standard molecular biology procedures were used for nucleic acid isolation and assembly of expression constructs. Sequencing was done by the Johns Hopkins University Synthesis and Sequencing Facility (Baltimore, MD). *M. purpureus* NRRL 1596 was received from the ARS, USDA (Peoria, IL) and cultured on PDA plates. Genomic DNA was extracted from mycelia using the DNeasy Plant Mini kit (Qiagen, Hilden, Germany). Target sequences were amplified by polymerase chain reaction using the primers listed in the Supplemental Information (Table S3.1).

3.11.14 Sequence analysis, cloning, and protein production

To verify the mRNA sequence of *pksCT*, *M. purpureus* NRRL 1596 was grown in MC medium following reported methods for citrinin production[241]. After 7 days of growth at 28 °C, the mycelia were harvested by centrifugation and lyophilized. The mycelial powder was ground under liquid nitrogen and total RNA was purified using the RNeasy Plant Mini kit (Qiagen, Hilden, Germany) following the manufacturer's protocol with on-column DNA digestion. A cDNA library was prepared using 1 µg total RNA with 0.5 µg random hexamer (Applied Biosystems, Foster City, CA) and M-MLV reverse transcriptase (Promega, Madison, WI) at 37 °C for 1 h. Negative control reactions lacked any reverse

transcriptase. PCR amplification of the desired exon boundaries was done using the cDNA library products directly (2 µL) with primers MpPksCTex1-5 and MpPksCT-SAT-3 and Phusion polymerase (New England Biolabs, Ipswich, MA). Primers for the constitutively expressed actin gene MpAct5 and MpAct3, based on a previous report[241], were used as positive controls. The PCR reactions were separated by 1 % agarose gel and the lone product band was ligated into pCR-blunt (Life Technologies, Carlsbad, CA) and transformed into DH5α by heat shock treatment. Five clones were sequenced using M13F and M13R primers.

Domain boundaries and expression constructs are detailed in Table S3.2. Each construct was transformed to *E. coli* BL21(DE3) and grown in LB media at 37 °C to OD_{600} of 0.6. The cultures were cooled in an ice bath for 30-60 min, and expression was induced with 1 mM IPTG (GoldBio, St. Louis, MO) overnight at 18 °C. Cell pellets were harvested by centrifugation for 15 min, at 4,000 x g, 4 °C. Pellets were either flash frozen in $N_2(l)$ and stored at -80 °C until use, or resuspended immediately in 5 mL/g cell pellet in buffer A (50 mM potassium phosphate, 300 mM NaCl, 10 % glycerol, pH 7.6). The slurry was sonicated on ice for 10 x 10 s at 40 % amplitude (Vibra-Cell Ultrasonic Processor, Sonics & Materials, Inc., Newtown, CT) and the lysate was cleared by centrifugation for 25 min at 27,000 x g, 4 °C. The cleared lysate was batch bound to Co^{2+}-TALON (Clontech, Mountain View, CA) at 4 °C for 1 h. Purification was performed by gravity column and the applied resin was washed with buffer A (10 CV) and buffer A + 2 mM imidazole (5 CV) before eluting in buffer A + 100 mM imidazole (5 CV). Each domain fragment was dialyzed into reaction buffer (100 mM potassium phosphate, 10 % glycerol, pH 7.0) and used immediately or flash frozen in $N_2(l)$ and stored at -80 °C until use.

For crystallization, PksCT CMeT was expressed and initially purified as described above, but dialyzed into 25 mM Tris, 5 % glycerol, pH 7.5 (buffer B). A secondary purification was performed by applying 60 mg PksCT CMeT to 25 mL Q-Sepharose Fast-flow resin pre-equilibrated in buffer B. A series of 2 CV step-wise washes of buffer B + 0, 10, 50, 100, 150, 200, 300, and 500 mM KCl were then applied to the column, with the CMeT eluting in the 50 and 100 mM KCl fractions. These fractions were combined, dialyzed into 2 x 1 L buffer B, and concentrated to 10 mg/mL.

Selenomethionine-labeled (SeMet) CMeT was prepared by growth in M9 minimal media to OD_{600} = 0.6, at which point the media was supplemented with the following L-amino acids per 1 L: 100 mg each lysine, phenylalanine, and threonine; 50 mg each isoleucine, leucine, and valine; 30 mg selenomethionine[263]. The cultures were allowed to continue to grow at 37 °C for 15 min before being cooled and inoculated as decribed above. Purification of SeMet CMeT for crystallization was identical to that of the native CMeT.

3.11.15 *In vitro* reconstitution of PksCT and analysis of product formation

Protein concentration was determined by Bradford assay (BioRad, Hercules, CA) in duplicate using bovine serum albumin as a standard. Prior to *in vitro* reactions, ACP was activated by Sfp with CoASH and $MgCl_2$ in reaction buffer for 1 h at room temperature as previously reported[48]. PksCT reconstitution reactions contained 10 µM of each included domain with 0.5 mM AcSNAC, 2 mM MalSNAC, 2 mM SAM, 1 mM NADPH, and 1 mM TCEP in reaction buffer totaling 250 µL. AcSNAC was prepared synthetically and MalSNAC was purified from MatB reactions[94]. After 4 h at room temperature, the reactions were quenched with 5 µL concentrated HCl and extracted into ethyl acetate 3 x 250 µL. The combined organic extracts were dried to a residue and resuspended in 250 µL of 20 % aqueous acetonitrile. Extracts were analyzed on an Agilent 1200 HPLC with autosampler by injecting 100 µL

onto a Prodigy ODS3 column (4.5 x 250 mm, 5µ, Phenomenex, Torrence, CA) with 5-85 % MeCN/H_2O, with 0.1 % formic acid, over 40 min at 1 mL/min. Mass spectrometric analysis was done using a Waters Acquity Xevo G-2 UPLC-ESI-MS in positive ion mode, with 5 µL injected on a BEHC C18 column and 10-90 % MeCN/H_2O, with 0.1 % formic acid, over 10 min.

Radiolabel transfer assays were carried out with 10 µM each protein in the above reaction buffer. Reactions were initiated with the addition of [1-^{14}C]-acetyl-CoA (American Radiolabeled Chemicals, St. Louis, MO) to 50 µM at room temperature for 5 min and quenched with the addition of 5X SDS loading buffer. The samples were separated by 16 % SDS-PAGE and dried. The dried gel was exposed to a phosphorimager screen (Amersham, Piscataway, NJ). Data was collected on a Typhoon 9410 Variable Mode imager (Amersham, Piscataway, NJ) and analyzed using ImageJ[264].

3.11.16 CMeT crystallization

Protein stock solutions were dialyzed (4x, 1:2000) against crystallization buffer (25 mM Tris-HCl, 5 % (v/v) glycerol, 2 mM DTT, pH 7.0). All crystallization experiments were performed using a robotic setup applying the sitting drop vapor diffusion method. Crystals in space group H3 were grown at 17 °C by mixing 0.5 µL of protein at 14.9 mg/mL (supplemented with 2 mM SAH) in crystallization buffer with 0.5 µL reservoir solution (0.1 M BIS-TRIS pH 6.57, 25% (w/v) PEG MME 2K) and grew to a final size of 0.4x0.2x0.1 mm^3 within three days. Prior to harvesting, crystals were cryo-protected by slowly exchanging the drop solution (0.1 M BIS-TRIS 8.0 pH, 25 % (w/v) PEG MME 2K, 10 mM acetoacetyl-CoA, 20 % (v/v) ethylene glycol) followed by an incubation of 5 min and flash freezing in N_2(*l*). SeMet CMet was crystallized in space group H3 at 17 °C by mixing 0.5 µL of protein at 14.3 mg/mL (supplemented with 2 mM SAH) in crystallization buffer with 0.5 µL reservoir solution (0.1 M BIS-TRIS pH 6.21, 21.4% (w/v) PEG MME 2K). Crystals were harvested after four days at a size of 0.4x0.3x0.05 mm^3. Prior to harvesting and flash freezing in N_2(*l*), crystals were dehydrated and cryo-protected (0.1 M BIS-TRIS pH 6.5, 30.0 % (w/v) PEG MME 2K, 2 mM SAH, 25 % (v/v) ethylene glycol) by carefully exchanging the drop solution over a period of 30-60 min.

3.11.17 Data collection, structure determination and analysis

Data sets were collected at the Swiss Light Source (SLS, Villigen, Switzerland) at beamline X06DA (native: λ = 0.99998 Å and 1.90747 Å; SeMet: λ = 0.97929 Å) and a temperature of 100 K. Data reduction was performed using DIALS[265] and XDS[203], datasets were analyzed with Phenix[204]. Resolution cutoffs were determined by $CC_{1/2}$ criterion[210]. The structure of SeMet CMeT was solved by SAD phasing using SHELX[266]. The asymmetric unit contained one protomer, which was used for initial phasing of native crystals diffracting to higher resolution using molecular replacement in PHASER[205]. Initial maps were improved by density modification using PARROT[208], followed by automated rebuilding with BUCCANEER[206]. The final model was obtained after iterative cycles of model building in COOT[207] and TLS refinement in Phenix[204], yielding excellent geometry (Ramachandran favored/outliers: 99.48 % / 0.00 %) and R_{work}/R_{free} values of 0.19 / 0.22 (Table S3.3). The final model covers residues 1785-2163. All structures of native and SeMet-labeled protein revealed weak, but significant Fo-Fc difference density for a partially occupied ligand in the ligand binding tunnel (Figure S3.10), which could not be displaced by acetoacetyl-CoA soaks (cryo solutions + 15 mM acetoacetyl-CoA). Anomalous data collection at λ = 1.90747 Å did not reveal significant anomalous signal as expected for a fully occupied sulfur or phosphorus atom as part of the ligand (Figure S3.10B). The observation was verified by two protein purifications and crystal treatments in the presence or absence of acetoacetyl-CoA. Despite

the featured shape of the density, the nature and origin of the ligand, either from the crystallization conditions or protein purification, could not be unambiguously determined.

Related folds were identified using the protein structure comparison service PDBeFold at European Bioinformatics Institute (http://www.ebi.ac.uk/msd-srv/ssm), authored by E. Krissinel and K. Henrick[222], using 40 % (query) and 70 % (target) as matching parameters, and interfaces were analyzed using QtPISA[223]. Bias-removal for Fo-Fc omit maps was achieved by applying a random perturbation to coordinates (Δ0.2 Å) and B-factors (Δ20 % of the mean overall B-factor) using MOLEMAN2[267] prior to refinement. Figures were generated using PYMOL[268].

3.11.18 Phylogenetic sequence analysis

51 sequences of CMeT domains from NR-, HR-, *cis*-AT- and *trans*-AT PKS as well as FAS were selected to analyze phylogenetic distances. Sequences of inactive pseudo CMeT domains were included for FAS and HR-PKS. Alignments were generated using Clustal Omega[269] and phylogenetic trees were generated using the neighbor joining algorithm implemented in Geneious version 8.1.6 (http://www.geneious.com)[235].

3.12 Author Contributions

PAS carried out all molecular biology, protein production, *in vitro* reconstitution and functional analysis; DAH crystallized and determined the structure of CMeT and analyzed the structure and phylogeny; TM contributed to structure determination and analysis of CMeT; CAT designed the work and analyzed data. All authors contributed to manuscript preparation.

3.13 Acknowledgements

Data were collected at beamlines PXIII of PSI; we acknowledge excellent support from the beamline teams. This work was supported by the Swiss National Science Foundation (SNF) project grants 159696 and R'equip grant 145023, as well as NIH grant ES001670. DAH acknowledges a fellowship by the Werner-Siemens Foundation. PAS thanks Anna Vagstad for providing the MTAN expression plasmid, as well as Adam Newman, Dan Marous, and Callie Huitt-Roehl for helpful discussions.

3.14 Accession Number

The atomic coordinates and structure factors have been deposited in the PDB

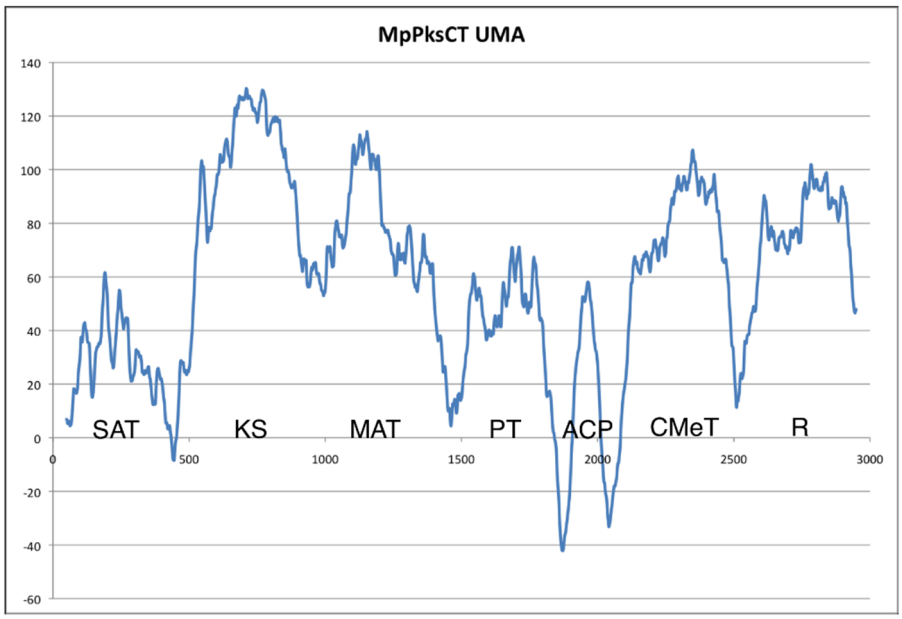
(http://wwpdb.org) under the accession code PDB: 5MPT.

3.15 Supplemental Information

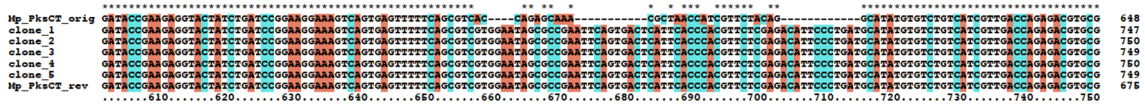
Table S3.1 | Primers used in this work.

Primer Name	Sequence
MpPksCT-g5	ATGATTGACTCAACTTCGCACTCAAATCTGAG
MpPksCT-g3	TTAATCTAGAAATCCCATGGTCTTCCATGC
MpPksCT-ex1-5	GCATATTAATGATTGACTCAACTTCGCACTC
MpPksCT-ex1-3.2	GTCAACGATGACAGACACATATGCATCAGGGAATGTCTCGAGAACG
MpPksCT-ex2-5.2	CTCGAGACATTCCCTGATGCATATGTGTCTGTCATCGTTGAC
MpPksCT-ex2-3	GCATGCGGCCGCATCTAGAAATCCCATGGTCTTCCATG
MpPksCT-SAT-3	GCATGCGGCCGCTCCACCGAGAAGCTGCCC
MpPksCT-SKM3-3	GCATGCGGCCGCAGAGCGCTGGGATGCGTCTTG
MpPksCT-PT-5	GCATATTAATGACAGCAATTATCGAGGCACC
MpPksCT-PT-3	GCATGCGGCCGCGCCGCTTACCGCAGTAGACG
MpPksCT-ACP-5	GCATATTAATGACACCTCCGACAAAGAAAGCGC
MpPksCT-ACP-3	GCATGCGGCCGCACCGTTACCATTAACGACACCATTAGTGC
MpPksCT-CMeT-5	GCATATTAATGGTCTTGTTTCCTGAGCTAGGGGG
MpPksCT-CMeT-3	GCATGCGGCCGCGTCTGTCAATGGGACCTGCACC
MpPksCT-CMeTAseI-5	CGAGGAGGCCAGACTGATTGATATCAACG
MpPksCT-CMeTAseI-3	CGTTGATATCAATCAGTCTGGCCTCCTCG
MpAct-5	GGAATTCTGCAGATTCTACAACGAACTCCG
MpAct-3	GGAATTCTGCAGTCAGGGAGTTCATAGGAC
MpPksCT-CMeT-Y1955F-5	GTCACTGATGTCTTTGCCAAATCTCC
MpPksCT-CMeT-Y1955F-3	GGAGATTTGGCAAAGACATCAGTGAC
MpPksCT-CMeT-Y1955A-5	GTCACTGATGTCGCGGCCAAATCTCC
MpPksCT-CMeT-Y1955A-3	GGAGATTTGGCCGCGACATCAGTGAC
MpPksCT-CMeT-H2067A-5	CAAATTGTGTGGCGGCGACCCGG
MpPksCT-CMeT-H2067A-3	CCGGGTCGCCGCCACACAATTTG
MpPksCT-CMeT-H2067Q-5	CAAATTGTGTGCAGGCGACCCGG
MpPksCT-CMeT-H2067Q-3	CCGGGTCGCCTGCACACAATTTG

Relates to Experimental Procedures.



B)



C)

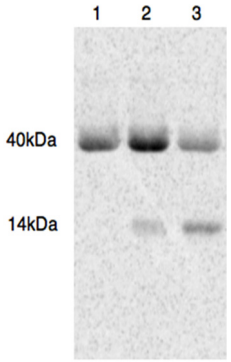


Figure S3.6 | Domain deconstruction, exon revision, and starter unit verification for PksCT.
Relates to Figure 3.1 and Figure 3.2. **A**) Sequence conservation, predicted secondary structure, and local hydrophobicity are combined to generate an UMA score for each position. Low scores indicate likely linker regions and were used to guide the domain boundaries for mono- or multi-domain expression constructs. For details on input sequences and algorithm settings, please see Supplemental Experimental Procedures. **B**) Five independent clones amplified from a cDNA library generated from total RNA under citrinin producing conditions. Each clone exactly matched the FGENESH prediction for the exon 1:exon 2 boundary. **C**) Radiolabel assay of acetyl starter unit transfer in PksCT. Lane 1: PksCT SAT, Lane 2: PksCT SAT + ACP without Sfp activation, Lane 3: PksCT SAT + ACP with Sfp activation. NR-PKS ACP domains are often partially phosphopantetheinylated during expression, accounting for the low ACP signal in lane 2.

Table S3.2 | Plasmids used in this work.

PksCT domain(s)	Sequence	Plasmid name	Parent vector	Tag	MW (kDa)
SAT	M1-G380	pEMpPksCT-SAT	pET-24a	C-His6	41.8
SAT-KS-MAT	M1-I1288	pEMpPksCT-SKM3	pET-24a	C-His6	144.0
PT	T1285-G1651	pEMpPksCT-PT	pET-24a	C-His6	41.5
ACP	T1652-G1779	pEMpPksCT-ACP	pET-24a	C-His6	15.0
CMeT	V1780-D2182	pEMpPksCT-CMeT	pET-24a	C-His6	46.6
CMeT-Y1955A	V1780-D2182	pEMpPksCT-CMeT-Y1955A	pET-24a	C-His6	46.6
CMeT-Y1955F	V1780-D2182	pEMpPksCT-CMeT-Y1955F	pET-24a	C-His6	46.6
CMeT-H2067A	V1780-D2182	pEMpPksCT-CMeT-H2067A	pET-24a	C-His6	46.6
CMeT-H2067Q	V1780-D2182	pEMpPksCT-CMeT-H2067Q	pET-24a	C-His6	46.6
CMeT-R	V1780-D2593	pEMpPksCT-CMeT-R	pET-24a	C-His6	92.7

Relates to Figure 3.2.

Table S3.3 | Crystallographic data collection and refinement statistics.

	CMeT native		CMeT SeMet
Data collection			
Space group	H3	H3[†]	H3[†]
Cell dimensions			
a, b, c (Å)	98.37, 98.37, 133.43	98.52, 98.52, 133.20	98.13, 98.13, 124.45
a, b, g (°)	90.0, 90.0, 120.0	90.0, 90.0, 120.0	90.0, 90.0, 120.0
Wavelength (Å)	0.99998	1.90747	0.97929
Swipes	1	2	2
Resolution (Å)	71.80 – 1.65[‡]	71.85 - 2.05	50.21 – 1.85
R_{merge} (%)*	8.0 (159.6)	8.4 (102.9)	13.5 (274.2)
$I/\sigma I$*	10.61 (1.37)	15.62 (1.38)	13.75 (1.12)
$CC_{1/2}$ (%)*	99.9 (76.0)	99.8 (71.1)	99.9 (57.1)
Completeness (%)*	98.9 (94.3)	97.9 (81.7)	99.8 (99.7)
Redundancy*	6.6 (6.6)	8.5 (2.6)	10.4 (10.2)
Unique reflections*	57,453 (4,045)	59,253 (3,685)	76,184 (5,646)
Refinement			
Resolution (Å)	71.80 – 1.65		
R_{work}/ R_{free}	0.19 / 0.22		
Ramachandran favored / outlier (%)	99.48 / 0.00		
No. atoms	6,512		
Protein	6,173		
Ligand	45		
Solvent	294		
B-factors	36.22		
Protein (Å^2)	36.29		
Ligand (Å^2)	26.84		
Solvent (Å^2)	36.11		
R.m.s deviations			
Bond lengths (Å)	0.009		
Bond angles (º)	0.975		

Relates to Figure 3.3. The resolution cutoff was determined by $CC_{1/2}$ criterion (Karplus and Diederichs, 2012). *, Highest resolution shell is shown in parenthesis. Datasets of native CMeT were acquired from the same crystal. [†]Anomalous data processing was applied. [‡]A deprecated conservative resolution cutoff at $I/\sigma I$=2.91 and R_{merge}= 69.8% is reached at a resolution of 1.84 Å.

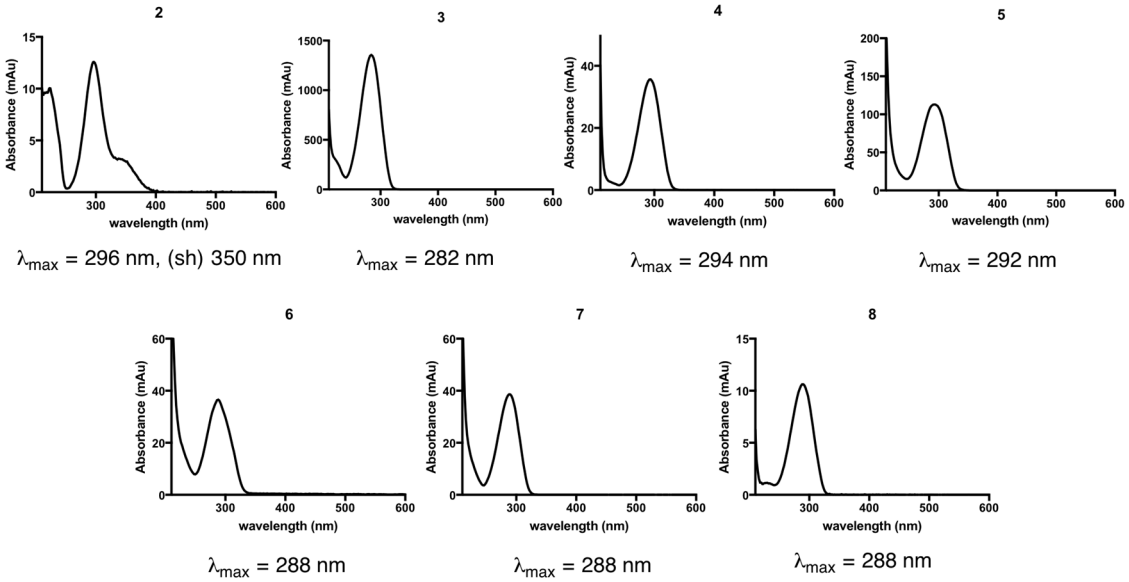


Figure S3.7 | UV-Vis spectra for compounds 2-8.
Relates to Figure 3.2 and Figure 3.5A.

Table S3.4 | Detected masses for compounds 2-8.

#	Calculated [MH+] or [MNa+] m/z	Detected m/z (ppm)	Formula
2	237.1127	237.1121(2.53)	$C_{13}H_{17}O_4$
3	127.0395	127.0391 (3.15)	$C_6H_7O_3$
4	155.0708	155.0704 (2.58)	$C_8H_{11}O_3$
5	211.0970	211.0970 (0.00)	$C_{11}H_{15}O_4$
	233.0787	233.0790 (1.29)	$C_{11}H_{14}O_4Na$
	169.0865	169.0859 (2.37)	$C_9H_{13}O_3[-C_2H_2O]$
6	253.1075	253.1076 (0.40)	$C_{13}H_{17}O_5$
	275.0895	275.0893 (0.73)	$C_{13}H_{16}O_5Na$
	211.0970	211.0965 (2.73)	$C_{11}H_{15}O_4[-C_2H_2O]$
	155.0708	155.0701 (4.51)	$C_8H_{11}O_3[-C_5H_6O_2]$
7	141.0552	141.0553 (0.71)	$C_7H_9O_3$
8	141.0552	141.0552 (0)	$C_7H_9O_3$

Relates to Figure 3.2 and Figure 3.5A.


active
inactive
NR-PKS
trans-AT PKS
cis-AT PKS
HR-PKS
FAS
Sequence Logo
secondary structure:
active site residues:
PksCT (Mp)
linker
N-terminal subdomain
C-terminal subdomain
Nte
Onnl (Ts*) Q5MP00
PedF (Pf*) Q6VT95
BryC (Ces) A0MS26
PksM (Bs) P40872
DifF (Ba) Q1RS54
MmpD (Pf) Q8RL72 1st
MmpD (Pf) Q8RL72 2nd
RhiD (Pbr) E5ARL9
XabB (Xa) Q9AIT3
AlbI (Xa) Q70C47
OzmH (Sa) B2WW42
OzmK (Sa) B2WW44
OzmN (Sa) B2WW47
ChiC (Sc) A9EW96
Dsz/DisA (Sc) Q4U447
KirAII (Sc) S5YE98
KirAIV (Sc) S5V790
VirA (Sv) A4PHN0
LkcC (Sr) Q83X88
CurJ (Mp) Q6DNE3
JamJ (Mp) Q6E7K0
CrpA (Nos) A4D933
HctD (Mp) Q1EDB2
AnaG (Os) C4NCC2
LovB (At) Q9Y8A5
Fum1p (Gm) Q9Y8A2
LovF*(At) Q9Y7D5
McyG (Ma) Q9RNB5
EpoD (Sc) Q9L8C6
Pks17 (Bf) Q6RKI3
AfoE (An) Q5BEJ6
MOS (Ss) A5PHD6
PkeA (An)
PkiA (An) Q5B7U4
AusA (An) Q5ATJ7
MpaC (Pb) F1DBA9
DtbA (An) G3Y802
CurS1 (At) L7X8J4
HmpB (Hs) B3FWS4
Rdc5 (Mch) B3FWU0
BfA (Pb) A0A068AB87
FAS (Tc) D6W6D2
FAS (D) Q9VQL7
FAS (Bm) O01678
FAS (Am) A0A088AP68
FAS (H) P49327
FAS (P) A5YY76
FAS (Xt) F6W5Y2
FAS (Ce) P91871

Figure S3.8 | Alignment of 51 CMeT domains from PKSs and FASs.
Relates to Figure 3.4. Alignment of 51 CMeT domains from PKSs and FASs. The alignment reveals a high sequence conservation of active CMeT domains, whereas inactive ΨCMeT domains of FAS and some HR-PKS reveal deletions predominantly in the *N*-terminal subdomain and a low sequence conservation of active site residues. Sequence numbers, secondary structure elements, and active site residues correspond to PksCT. All sequences are labeled as protein name (organism abbreviation) Uniprot number and sorted according to their phylogenetic distance depicted in Figure 3.4. The sequence of PksCT corresponds to Supplemental Item 1. Multienzyme family classifications are indicated in colored groups. (°, endosymbiont of this organism; ‡, diketide synthase)

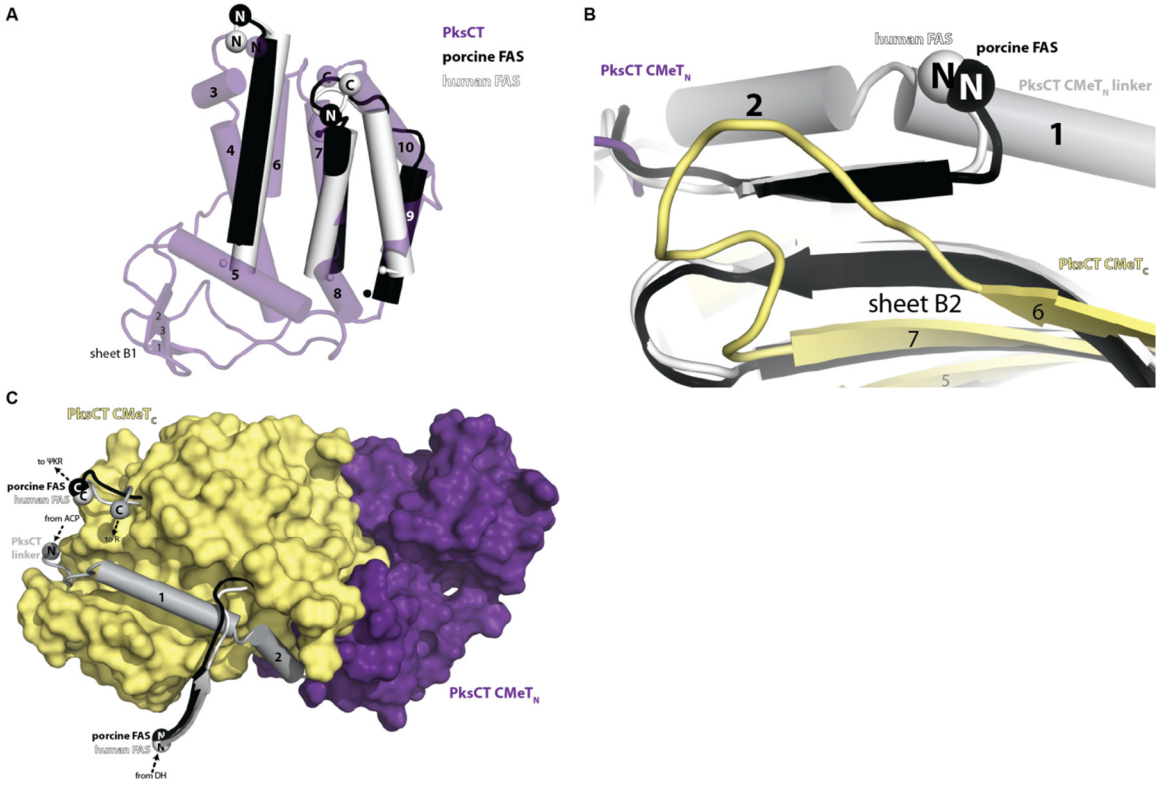


Figure S3.9 | Structural comparison of PksCT CMeT with mammalian FAS ΨCMeT domains.
Relates to Figure 3.3. (**A**) A superposition of the truncated and largely disordered *N*-terminal subdomain of the animal FAS ΨCMeT (porcine: black; human: white) with PksCT CMeT indicates a structural conservation of three helices, overlapping with PksCT CMeT helices 4, 7 and 9/10. Ends of disordered regions are indicated by small spheres. Secondary structures are assigned according to Figure 3.3B. (**B**) In FAS, the linker connecting to the *N*-terminal subdomain contains a β-strand that extents the central β-sheet of the *C*-terminal subdomain. In PksCT CMeT, the position of this β-strand is occupied by helix 2 of the *N*-terminal linker leading into CMeT. (**C**) In PksCT the *N*-terminal linker wraps in a surface groove around the *C*-terminal subdomain and positions the *N*- and *C*-terminal linker termini in close proximity to each other, whereas in FASs the *N*-terminal linker contacts this domain in a different location distant to the *C*-terminal linker terminus.

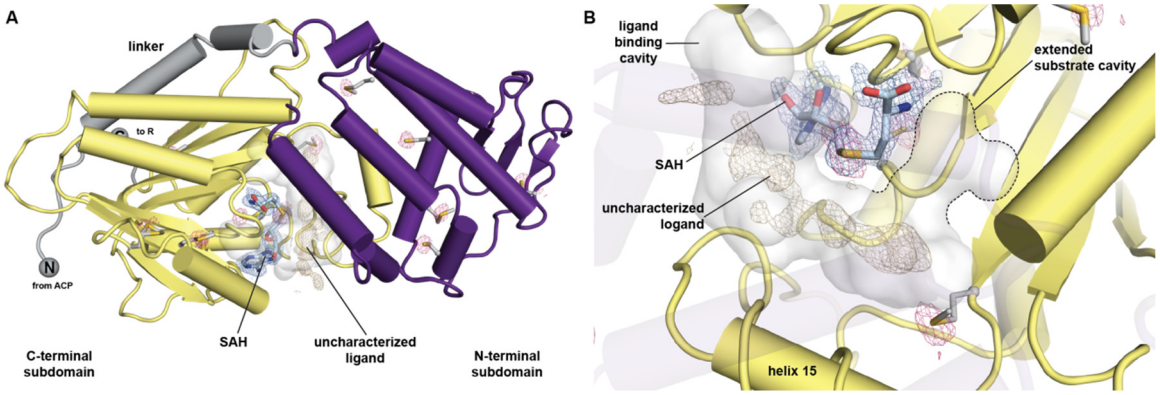


Figure S3.10 | Uncharacterized ligand and anomalous difference density.
Relates to Figure 3.3A and Figure 3.3C. (**A**) The active site tunnel reveals significant F_O-F_C omit difference density at 2.5 σ (brown) with a featured shape, which does not match components of the crystallization condition. Anomalous difference density at 3.7 σ (pink), indicates the presence of sulfur in methionine and cysteine residues as well as in SAH, but no significant difference density as part of the ligand. (**B**) The close up view on the active site reveals the expansion of the ligand's difference density along the central tunnel with its end in the junction to the extended substrate cavity. The same contour level of the SAH difference density (blue) suggests either a partial occupancy or various conformations of the less defined ligand. The extent of difference density corresponds approximately to an atomistic binding interface of 500 $Å^2$.

Item S1 | Revised M. purpureus PksCT sequence. Relates to Figure 3.1. Residues encoded by alternate exon boundary are underlined:

MIDSTSHSNLRSKAFIFGPQDLSFDVRSFNKLHSQLQNHQWVLDALASLPKLWDNFAASDQKVQQSNTGKLLENL
NAWISSGVAPEEAFPLPNVLLSPLVVIGQLVEYMTFLKAAFPDLGKKHDLPISIKEDTETFGLCTGTLCAFAVACSSNI
ADIQHYGAVAARLAMLVGAIVDTEEVLSDPEGKSVSFSASWNSAEFSDSFTHVLETFPDAYVSVIVDQRRATLTASK
KTAPAIIERLKQEGAHVTSIALSGRFHWKKHQDAVSSLIQFCGLDPDLQLADATKMLLPSRSSSDGQYITTGKLHELA
LRAILLEQSEWYKTCRISYLSKFIMDDAAVICFGPERCMPPTLARKLGPRLTYVSEIDISSSRVPGQLLGGTQKLNLTDL
PDERIAVIGMACRLPGAEDHEGFWEILKTGQSQHREVPEDRFGMATAWREADKRKWYGNFIDNYDTFDHKFFKK
SPREMASTDPQHRLMLQVAYQAVEQSGYFRNNGTNRRIGCFMGVGNVDYEDNIACYPANAYSATGNLKSFLAGK
ISHHFGWTGPSLTLDTACSSSSVAIHQACRSILSGECNGALAGGVNVITSPNWYHNLAGASFLSPTGQCKPFDAKGD
GYCRGEGVGAVFLKRLSSAIADGDQVFGVIASTKVYQNQNCTAITVPNAISLSELFTDVVRQARLEPKDITLVEAHGT
GTAVGDPAEYDGIRAVFGGPIRSDVLSLGSVKGLVGHTECASGVVSLIKTLLMIQQGFIPPQASFSSINPSLNAKAEEK
IEISTRLKPWDAPFRAALINNYGASGSNASMVVTQPPNLTETPSTPLPGKSYPFWISAFDQQSLQSYVRRLRQFLEKH
AADKNLSVANLSFQVACQSNWSLPQALVFSASTKEELNRALASFEKGSTDFPSVQLPDPKPVILCFGGQVSTYVGLD
QEVYNSTAILRHYLDQCDAMCLSLGLQSIYPAIFQRSPIEDIVQLQTALFAMQYSCAKAWIDSGLKVASVVGHSFGEL
IALCVSNAVSLKDAVKMISGRARLIKERWGADKGSMIAVEADLSDVEALLAKVKSQMGSETGLAIACYNASKSFTLA
GPTKDVDHAENLLKNDPDFSGIRYKRLNVTNAFHSVLVDALIDDLESLGQGIRFKEPTIKLERATEQESTSTLNANYVA
THMRKPVFFAQAVKRLSDKFPVAIWLEAGSNSTITAMASRALGTSNSSFQAVNITSEGAFRFLCDTTVKLWKEGQK
VSFWAHHRLQTPMYTPVLLPPYQFEKSRHWMDLKVPPKPEASVQVAEQTAIIEAPKGLTTFVGYQDASQRSVRFR
VNVTTEKFNRLLSGHIMANAAAVCPGMFQVEIALDALTSLRPEFQARSFIPELHDLRHYQPLVRDESRAVWIEAHCP
NAEGLVWNWKLTASDDKGSGSVTHTSGTITFQAADSVQVKSEFEKLRRLIGRKRCLQLLDSNVADDILQGRNIYRAF
TEVIDYKEIYRHVTKIAGRDNESAGRVIKTYDGETWLDTVLTDCFCQVAGIFVNLMTTKIDLSERGIFICDGIDRWLRA
PNAGSNNTPSQVYEVFALHHCESDSKYLSDVFAFDAREGSLVEVALGISYQKVSISGIRRVLSKGMPAGLQPQVPTS
PAAVAAIKTVSPPPVADSPLVDGSSTAVSGTPPTKKAPKAPSVDITGKMREIICNLSGLEPDEVKDDSDLVELGIDSL
MSMELAREVDLAFKTTIDVTQLIDVTDFRSLVECMQRILGIDNQEDNTYLAEGLNGHEGVVTNGNAYHVNGTNGV
VNGNGVLFPELGGSILPKSAILDAFRIAKEATDDFILNGQLGTYYNEVMPRSTELCVAHIVNAFEQLGCPIRSAAAYQ

RLERVPYLPKHERFMNLIYGLLEEARLIDINGSEITRTSVPVSTKSVETMLEELLHDEPLHAAEHKLTSLTGSKFADCITG
KEDGLQLIFGSPEGREIVTDVYAKSPINAVWIQQAEFFLEQLVKRLPNTGEPLRILEMGAGTGGTTVKMLPLLERLGV
PVEYTMTDLSSSLIAAARKRFKKYPFMKFKVVNIESPPDPQLVHSQHIILATNCVHATRNLEISTRNIHRILRPDGFLLL
LEMTEQVPWVDFIFGLLEGWWLFEDGRRHALQPATHWKKILTSVGYGHVDWTEGTRPEANIQRLIIALASEPRYD
HTPQSLQPPVQVPLTDIAGRQEIIDTYIREYTEDFRALPIPGIQQAVMPAPTGHCVLVTGATGSLGSHVVGYLSRLPN
VHTVVCLNRRSTVPATIRQEEALKVRGISLDDNSRSKLVVLEVETAKPLLGLPVETYQKLVNTATHIVHSAWPMSLTR
PIRGYESQFKVMQNLINLAREVAAWRPVPFKFSFQFISSIGVVGYYPLRYGEIIAPEETMTADSVLPVGYAEAKLVCER
MLDETLHQYPDRFRPMAVRIAQIAGSTSNGHWNPVEHFAFLIKSSQTLKALPDFDGSLSWCPVDDVSATLGELLIS
NTTPYSIYHIENPSRQQWRKMVKTLAQSLDIPRDGIIPFDQWIERVRNSSASINDNPARQLLEFFDQHFIRMSCGNLI
LDTTKTREHSATLRERGPVGPGLVEKYISAWKTMGFLD

Supplemental Procedures

UMA-guided domain dissection

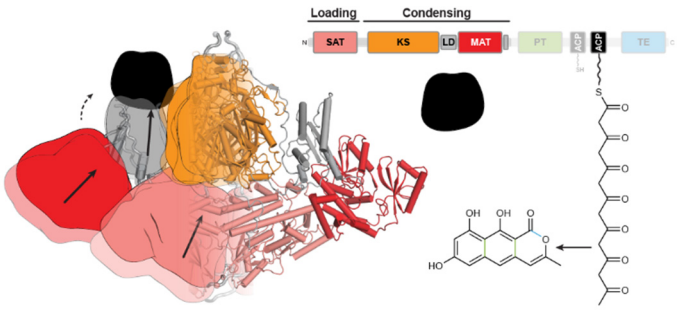
UMAv2.18 was used with the BLOUSUM62 matrix. Secondary structure predictions were generated with ProteinPredict ((Yachdav et al., 2014), https://www.predictprotein.org/). Comparative sequences were curated from NCBI using the Conserved Domain Architecture Retrieval Tool (Geer et al., 2002) with PksCT as the query and included the following entries: XP_001217248, ANID_07903, ANID_00523, CAN87161, XP_001273475, XP_002381902, EHA28237, XP_001818926, XP_001393501, CAK40124, XP_660990.1, XP_658638.1, AAR90253.1, BAE66025.1, XP_001243185.1, CAK48487.1, XP_001559289.1, XP_002149737.1, XP_002567553.1, XP_002487778.1, XP_002340070.1, XP_002384396.1, XP_002848394.1, XP_003070229.1, XP_003010590.1, XP_001395291.2, EHA25844.1, EHA28237.1, and GAA92425.1.

Supplemental References

Geer, L.Y., Domrachev, M., Lipman, D.J., and Bryant, S.H. (2002). CDART: protein homology by domain architecture. Genome Res 12, 1619-1623.

Yachdav, G., Kloppmann, E., Kajan, L., Hecht, M., Goldberg, T., Hamp, T., Honigschmid, P., Schafferhans, A., Roos, M., Bernhofer, M., et al. (2014). PredictProtein-an open resource for online prediction of protein structural and functional features. Nucleic Acids Research 42, W337-W343.

4 THE STRUCTURAL ORGANIZATION OF SUBSTRATE LOADING IN ITERATIVE POLYKETIDE SYNTHASES



Dominik A. Herbst, Callie R. Huitt-Roehl, Roman P. Jakob, Jacob M. Kravetz, Philip A. Storm, Jamie R. Alley, Craig A. Townsend & Timm Maier

Manuscript in publication

4.1 Abstract

Polyketide synthases (PKSs) are microbial multienzymes for the biosynthesis of biologically potent secondary metabolites. Polyketide production is initiated by loading of a starter unit onto an integral acyl carrier protein (ACP) and its subsequent transfer to the ketosynthase (KS). Initial substrate loading is achieved either by multi-domain loading modules or by the integration of designated loading domains, such as starter unit acyl-transferases (SAT), of which the structural integration into PKS remains unresolved. A crystal structure of the loading/condensing region of the non-reducing PKS CTB1 demonstrates the ordered insertion of a pseudo-dimeric SAT into the condensing region aided by the SAT-KS linker. Cryo-electron microscopy of the post-loading state trapped by mechanism-based crosslinking of ACP to KS reveals asymmetry across the CTB1 loading/condensing region in accord with preferential 1:2 binding stoichiometry. These results are critical for re-engineering the loading step in polyketide biosynthesis and support functional relevance of asymmetric conformations of PKSs.

4.2 Introduction

Polyketide synthases (PKSs) are multienzymes that underlie the biosynthesis of potent bioactive compounds such as antibiotics or toxins[55]. Their product diversity is generated by the sequential action of PKS modules, each responsible for polyketide elongation by condensation of acyl building blocks and subsequent reduction/modification. Homodimeric PKS modules are organized into a mandatory condensing and an optional modifying region[44,162,163,270] and integrate acyl carrier proteins (ACPs) for shuttling of covalently tethered intermediates. Condensing regions consist of an acyl-transferase (AT) and a ketosynthase (KS) domain. The AT is responsible for loading acyl-CoA derived extender units onto the phosphopantetheine (Ppant) modified *holo*-ACP, while the KS catalyzes the elongation of a polyketide precursor by decarboxylative Claisen condensation with the ACP-bound acyl extender unit. Modifying regions contain up to three canonical types of domains, the ketoreductase (KR), dehydratase (DH), and enoyl-reductase (ER), which sequentially modify the ß-keto-positon of the KS condensation product. In the presence of all three canonical modifying domains (KR, DH, ER) a fully saturated product is generated, while in non- or partially-reducing modules, various combinations of the modifying domains can maintain the β-keto group or optionally reduce/dehydrate it to a hydroxyl, alkene, or alkane. PKS modules either act iteratively (iPKS) or sequentially with directed substrate transfer between linked modules in multi-modular assembly lines (modPKS). Specialized loading modules or domains act as gatekeepers and initiate polyketide synthesis by starter unit loading[271].

Iterative non-reducing (NR)-PKSs, having no modifying domains, elongate successive poly-β-ketone intermediates until a final length determined by the KS is reached[91]. The non-reduced polyketide is cyclized by a product template (PT) domain to form an aromatic or fused aromatic ring system that is released from the ACP by a thioesterase (TE). In NR-PKSs, biosynthesis is initiated by starter unit acyl-transferase (SAT) domains, which load the ACP with acyl starter units obtained from coenzyme A (CoA) thioesters or from ACPs of other multienzymes[87,168]. The ACP transfers the starter unit to the active site cysteine of the KS and then is charged with a malonyl-CoA extender unit by a malonyl-CoA-specific acyl-transferase (MAT) in all subsequent elongation cycles. NR-PKSs share a common architecture and conserved interdomain interactions as demonstrated by functional domain swaps of ACPs, PTs, and loading/condensing regions (SAT-KS-MAT) of homologous enzymes in deconstruction experiments[48,91]. In some NR-PKSs, ACPs occur as functionally equivalent tandem didomains[48,93].

PKSs are dimeric multienzymes with structural segregation of condensing and modifying regions[42,44], similar to the mammalian fatty acid synthase (mFAS)[162]. However, the structural integration of dedicated loading modules or of specialized substrate loading domains, as exemplified by the NR-PKS SAT-KS-MAT core region, remains unknown.

The KS-MAT didomains of NR-PKSs are homologues of the corresponding condensing regions in other PKSs and mFAS, which have been structurally characterized as isolated didomains[44,46,162-164,172]. However, in NR-PKSs, soluble expression of KS-MAT is only achieved by including the SAT domain in a tridomain construct[85,94], indicating that SAT, KS, and MAT form an integrated loading/condensing region architecture. SATs can be successfully isolated[182] and swapped in the context of the condensing region environment[272]. Together with an ACP, the SAT-KS-MAT core region of NR-PKSs is catalytically competent for synthesizing the complete polyketide backbone *in vitro*[48,85,91]. Early crosslinking experiments demonstrated binding of NR-PKS ACP domains to the dimeric KS preferentially with a 1:2 stoichiometry[21].

To study PKS substrate loading architectures and the organization of NR-PKSs, we focused on the SAT-

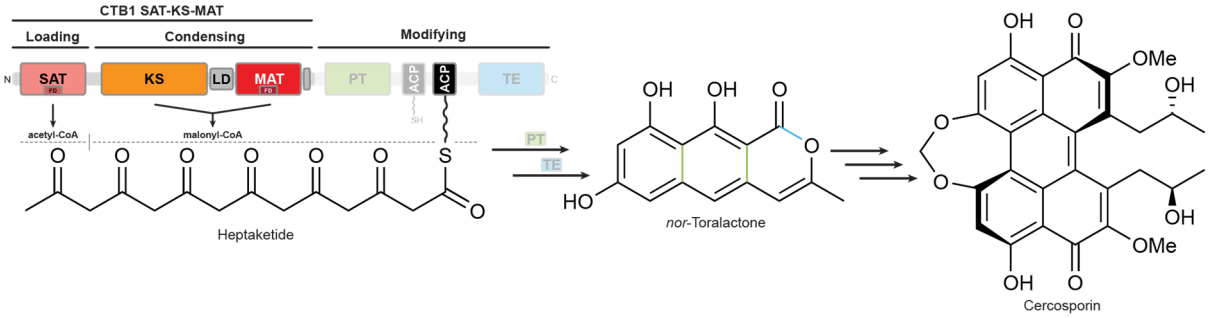


Figure 4.1 | Domain organization and catalytic scheme of the cercosporin PKS CTB1.
The domain organization is shown schematically with functional regions indicated. The CTB1 SAT-KS-MAT and ACP2 constructs used for structural studies are shown in full color, other domains in faded colors. The iterative synthesis of *nor*-toralactone involves initial loading of the starter unit acetyl-CoA by the SAT and six cycles of condensation by the KS using malonyl-CoA loaded by the MAT. The linear product is cyclized and released by the PT and TE, respectively, as indicated by colored bonds. Downstream processing of *nor*-toralactone yields cercosporin. FD: Ferredoxin-like subdomain.

KS-MAT loading/condensing region of the NR-PKS CTB1. CTB1 produces *nor*-toralactone, the aromatic polyketide precursor to cercosporin in the fungus *Cercospora nicotianae*[273,274] (Figure 4.1). Cercosporin is a light-activated plant toxin and protein kinase C inhibitor essential for *C. nicotianae* infectivity[275]. CTB1 contains five enzymatic domains as well as tandem ACPs (ACP_1-ACP_2). For initiation of CTB1 biosynthesis, acetyl-CoA is loaded via the SAT[271] onto an ACP and transferred to the KS. The iterative action of the KS-MAT condensing region produces an ACP-tethered heptaketide (C_{14}), which is subsequently cyclized by the PT and released as a pyrone by the TE[276] (Figure 4.1).

Here, we report the crystal structure of the CTB1 SAT-KS-MAT region, which reveals a pseudo-dimeric arrangement of SAT domains nested in a cleft formed by the KS and MAT domains. To specifically resolve the transient interactions in the substrate loading step, we employed mechanism-based crosslinking[13] to trap a functional state of ACP-KS interaction for cryo-electron microscopy (EM) structure determination.

4.3 Results

4.3.19 Crystal structure of an integrated loading architecture

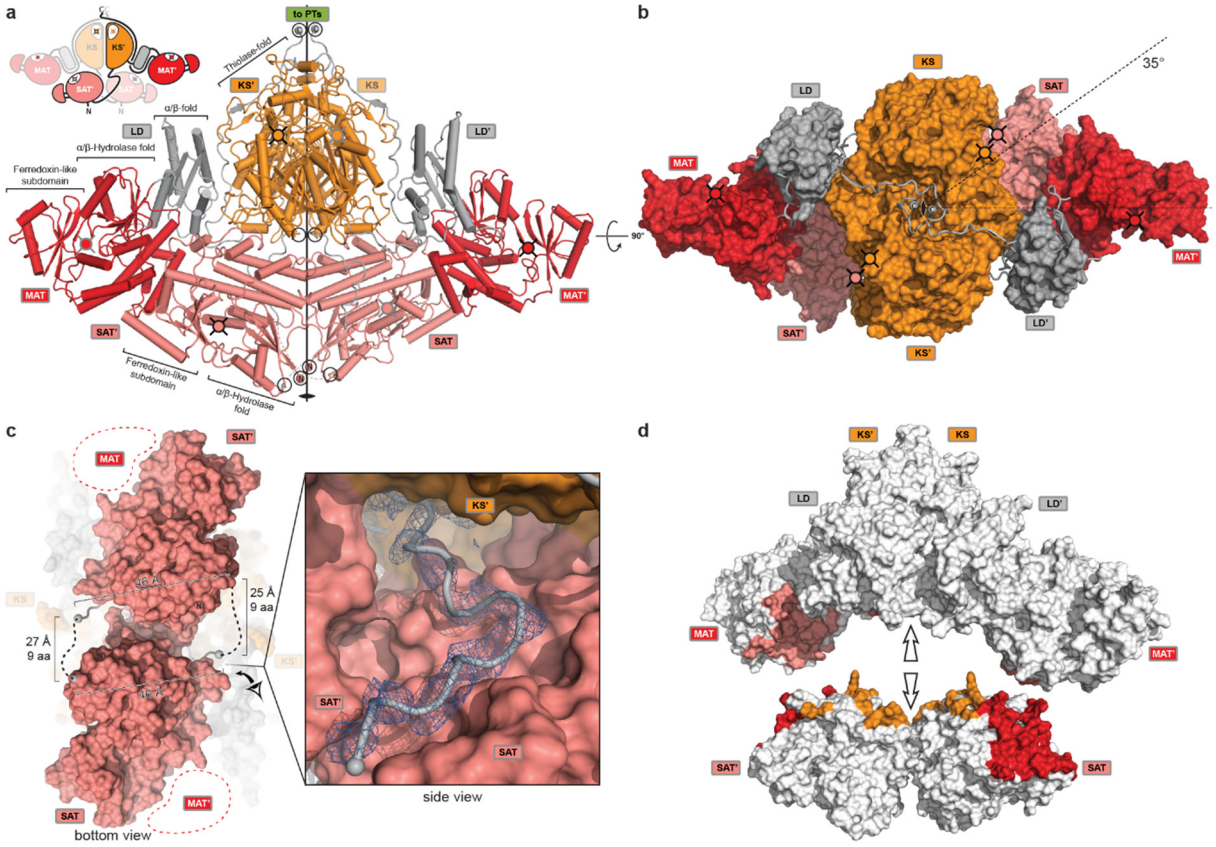


Figure 4.2 | Crystal structure and interdomain interactions in CTB1 SAT-KS-MAT.
(a) Cartoon representation of the SAT-KS-MAT crystal structure (front view); the intertwined domain connectivity is indicated in the inset. Active sites are marked by cross-hatched circles, termini of SAT and KS aligned along the twofold axis are indicated by circles, *N*- and *C*-termini of SAT-KS-MAT are additionally labelled. **(b)** Top view of SAT-KS-MAT, domains are shown in surface representation, linkers as cartoon. Active sites, termini and twofold axis are indicated as in **(a)**, the angle between the longest axes of the KS-LD-MAT and SAT (pseudo-) dimers is indicated. **(c)** The SAT-KS linker. Bottom view of the crystal structure as in **(b)** highlighting the connectivity of the SAT-KS linker and euclidean distances. The inset provides a close-up view from the side of the *C*-terminal region of the SAT-KS linker involved in contacts to KS' and SAT. An F_o-F_c shaked omit difference map is shown at 2.5σ **(d)**, Interdomain interfaces between the SAT domains and the KS-LD-MAT region of CTB1 are mapped onto a split-surface representation of the SAT-KS-MAT structure by coloring according to the interaction partner. The SAT domains have been separated from the KS-LD-MAT for representation.

The 2.8 Å resolution crystal structure of SAT-KS-MAT reveals a compact dimer of rhomboid shape (Figure 4.2, Supplementary Table 4.1a, Supplementary Figure 4.5). The two KS domains dimerize and each monomer of the KS laterally connects to one MAT through a small linker domain (LD). The individual domain structures (Supplementary Table 4.2a) and the overall arrangement of the KS-LD-MAT domains (Figure 4.2a,b) resemble condensing region structures from mFAS[162,172], fully-reducing iPKS[44] and several modPKS[46,163,164], but differ from the architecture of the PikAIII modPKS resolved at intermediate resolution by cryo-EM[42,45] (Supplementary Figure 4.6). The SAT has structural homology with other ATs, with root mean square deviations of C_α atom positions (r.m.s.d.) between 2.5-3.2 Å for 62-75 % aligned amino acids (aa) (Supplementary Table 2a) and an active site located in a cleft between the α/β-hydrolase core fold and a ferredoxin-like subdomain. The *C*-terminus of the SAT domain is connected via a 26 aa linker to the *N*-terminus of the KS domain (Figure 4.2c). Nine aa of this linker remain disordered in the crystal structure, but the connectivity is unambiguously determined based

on maximal linker length and spatial restraints (Figure 4.2c). In an intertwined arrangement, the SAT domain linked to the KS of the first protomer is nested by the KS-LD-MAT domains of the second protomer (Figure 4.2a). The clefts formed between the MAT and KS domains, which have been suggested to guide the interactions of the upstream ACP with the KS in modPKS[26], are partially occupied by the SAT domains. The active site entrance of each SAT domain points into this cleft toward the other protomer's KS active site, which is located at a distance of 56 Å.

The arrangement of the two SAT domains is stabilized by interfaces with surrounding domains (Figure 4.2d, Supplementary Figure 4.7a). The two SAT domains form a direct pseudo-dimer interface of 324 Å^2 devoid of hydrogen bonds or salt bridges (Supplementary Table 4.2b, Supplementary Figure 4.7b,c). The *C*-terminal end of the SAT-KS linker adds an average interface area of 493 Å^2 by binding to a surface groove of the opposite SAT (Figure 4.2c, Supplementary Table 4.2b). The largest interface with highest predicted stability is formed between the SAT and MAT of opposite chains, covering on average 957 Å^2 (Figure 4.2d, Supplementary Figure 4.7d,e) and a network of at least two salt bridges and five hydrogen bonds, while the interface to the other protomer's KS domain is considerably smaller (184 Å^2) (Supplementary Figure 4.7f,g). Notably, the SAT contacts the KS and MAT of the other protomer, without interfacing with the connecting LD.

The SAT pseudo-dimer is arranged at a 35° angle relative to the condensing region KS-MAT dimer (Figure 4.2b), but the twofold symmetry axes of KS and SAT are aligned. The *N*- and *C*-termini of the KS and SAT domains are arranged in close proximity to the twofold axis (Figure 4.2a). Although CTB1 is an iPKS, the location of *N*- and *C*-termini at opposite ends of the molecule would be compatible with extended connectivity in an assembly line. The only other known structure of a NR-PKS SAT, which was excised as an isolated domain from CazM[182], reveals a *C*-terminal linker element that ends in a similar position to that of the SAT-KS in CTB1, despite a lack of sequence conservation in the linker (Supplementary Figure 4.8a,b). In contrast, the excised loading AT domain of the avermectin modPKS (AVES1)[277], which features completely different connectivity via an ACP to the downstream KS domain, has a distinct linker architecture resembling the post-MAT linker in PKS and FAS condensing regions (Supplementary Figure 4.8c,d). Together, these structural comparisons, in agreement with functional domain swaps[272], support a similar organization among NR-PKS loading domains. The active sites of the loading/condensing region reveal high structural conservation with their closest structural and functional neighbours (Supplementary Table 4.2a). The SAT has a conserved Cys-His dyad (Supplementary Figure 4.9a), which resembles the catalytic Ser-His dyad in AT domains of PKSs and mFASs. Similarly, the active sites of the MAT (Supplementary Figure 4.9b) and KS (Supplementary Figure 4.9c) are conserved. The KS active sites are connected by a hydrophobic and relatively narrow active site tunnel (Supplementary Figure 4.9d), as compared with other PKSs and mFASs, which may be involved in chain length determination.

4.3.20 Cryo-EM analysis of a trapped ACP=KS post-loading state

To analyze the impact of substrate loading, we trapped CTB1 SAT-KS-MAT in an initial post-loading state by mechanism-based crosslinking[13,21] and analyzed it by cryo-EM (Supplementary Figure 4.10). In CTB1, the SAT mediates loading of acetyl or, less efficiently, propionyl moieties onto ACP[271] for subsequent transfer to the KS active site cysteine. A Ppant analog linked to a selectively reactive α-bromopropionyl moiety was synthesized (see Supplementary Note 4.1) and enzymatically loaded onto isolated CTB1 ACP2 (Supplementary Figure 4.10a), which is functionally equivalent to CTB1 ACP1

(Supplementary Figure 4.10b) and shares a higher sequence identity (ACP2: 40% *vs*. ACP1: 30%) with the previously characterized PksA ACP[11,21]. For crosslinking, the loaded ACP2 was added to a variant of CTB1 SAT-KS-MAT, in which the SAT and MAT active sites were mutated to strictly direct crosslinking to the KS (denoted as SAT°-KS-MAT°). Loading of only a single ACP2 to the dimeric SAT°-KS-MAT° was kinetically favored over two-site loading, as previously observed for related NR-PKSs that contain only a single ACP[21] (Supplementary Figure 4.10c). Using an excess of ACP2 and an extended reaction time, >90% crosslinking was achieved. Negative stain EM revealed mostly dimeric particles, while cryo-EM grid preparation led to substantial monomerization of SAT°-KS-MAT°=ACP2 (= symbolizes chemical cross link; Supplementary Figure 4.10d, Supplementary Figure 4.11). Under optimized conditions, 60% of all particles were observed as intact dimers (Supplementary Figure 4.10e,f).

Single particle analysis of dimeric SAT°-KS-MAT°=ACP2 provided a map at 7.1 Å overall resolution, which was interpreted by rigid body and real space refinement using an all-atom model based on the crystal structure of SAT-KS-MAT and a homology model of ACP2 based on the NMR structure of PksA ACP[11]. Due to the limited resolution, and the absence of an authentic high-resolution structure, the analysis for ACP2 is restricted to conserved structural elements (Supplementary Table 4.1b, for details see online Methods). Single-particle processing including reference-free maximum likelihood classification provided a single, predominant conformation of SAT°-KS-MAT°=ACP2 (for details see online Methods). Although the sample used for grid preparation was >90% crosslinked to ACP2, only a single ACP position was clearly resolved in an overall asymmetric structure (Figure 4.3a-c). ACP2 is bound in a cleft between the LD and the KS/KS'-dimer. The *N*-terminal end of helix II in ACP2, which contains the Ppant analog-bearing DSL motif, points directly towards the KS active site entrance. For ACP interactions in modPKSs two different interaction sites have been proposed for the transthioesterification from the upstream ACP and elongation with the canonical ACP in modPKSs[26]. Both proposed sites are buried in the KS-LD-AT clefts of the DEBS PKS module five, while the ACP in CTB1 rather sits on top of the KS-LD-MAT cleft and therefore diverges from both proposed models.

At the second KS active site entrance, where no ACP is visualized, the surroundings of the substrate entry tunnel are more disordered (Supplementary Figure 4.12). Presuming that ACP2 is still crosslinked to the second site, binding of ACP2 to one KS active site may prevent ordered and competent ACP binding to the second active site by allosteric effects. Such negative feedback is supported by activity assays of 50% crosslinked samples of the related Pks4, which contains a single ACP and exhibits only 20%, instead of the expected 50%, activity relative to a non-crosslinked sample[21].

For the SAT-KS-MAT region, amino acids involved in ACP2 interactions can be identified based on the authentic higher resolution crystal structure, while the limited resolution of the EM reconstruction precludes assignment of interface amino acids in ACP2. The LD core around residue R879 and the LD-

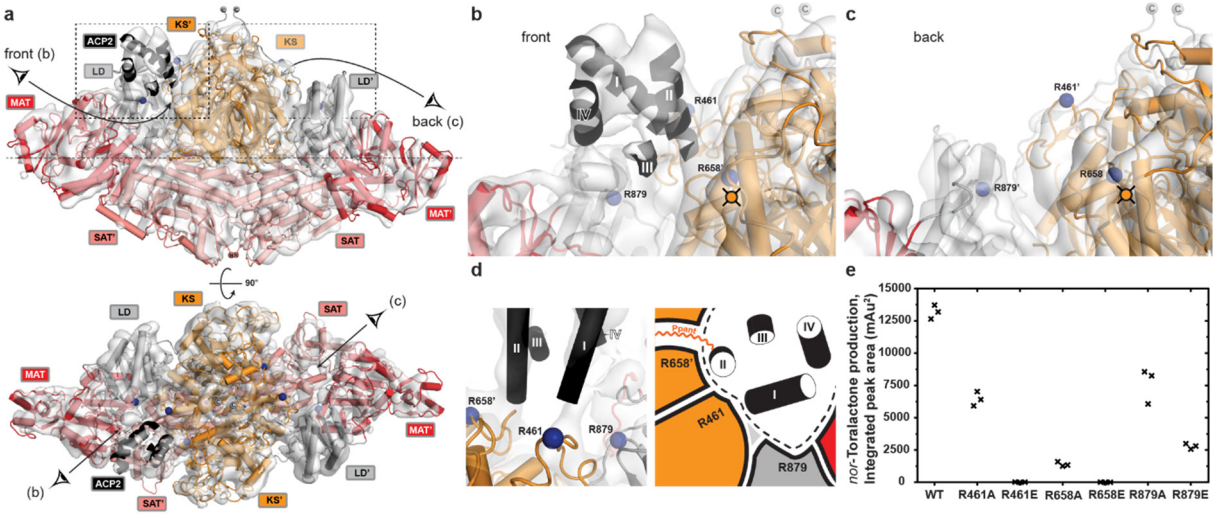


Figure 4.3 | Asymmetric cryo-EM structure of CTB1 SAT°-KS-MAT°=ACP2.
(a) Reconstructed volume and model of SAT°-KS-MAT°=ACP2 in front and top view. In the asymmetric structure, ACP2 is only resolved at one of the two active sites in dimeric CTB1 SAT°-KS-MAT°. The canonical four helix bundle of CTB1 ACP2 is shown. Termini, and views of **(b)** and **(c)** are indicated. All maps are contoured at 7.1σ. **(b, c)** Close-up of the active site (indicated) regions of the two KS domains with **(b)** or without **(c)** resolved ACP2. **(d)** Location of residues on SAT°-KS-MAT° interacting with ACP2 as structure and schematic representation. In the scheme, the approximate position of the KS active site tunnel and the crosslinker-modified serine in ACP2 are indicated. **(e)** Functional analysis of CTB1 SAT-KS-MAT mutants of ACP2 interacting residues. *Nor*-toralactone production was quantified by integrated peak areas of HPLC chromatograms at 280 nm. Each mutant was assayed in triplicate, and data from each replicate are presented.

MAT linker interact with the loop between helix I and II in ACP2. In accord with this proposed interaction, mutations of R879 to alanine or glutamate reduce activity in deconstruction experiments for the complete CTB1 PKS (Figure 4.3d,e, Supplementary Figure 4.13), but do not completely abolish *nor*-toralactone production. ACP2 interactions at the KS/KS' active site entrance (Figure 4.3b,c) involve two positively charged residues: R461 is strictly conserved in NR-PKSs, while R658' is conserved as a positively charged arginine or histidine in the CTB1 clade (Supplementary Figure 4.14a,b). The mutations R461A or R658A reduce activity considerably while the corresponding charge-inversion mutations to glutamate completely abolish *nor*-toralactone production, strongly suggesting that the two arginines mediate critical interactions with negatively charged groups in ACP2 (Figure 4.3e, Supplementary Figure 4.13).

4.4 Discussion

Interactions of an ACP with a multienzyme KS domain have previously been resolved in the crystal structure of yeast FAS (yFAS) at 3.1 Å resolution[186] and in the cryo-EM reconstruction of the PikAIII modPKS at 8.6 Å and 7.3 Å resolution[42](Supplementary Figure 4.15a-e). Both structures reveal strictly symmetric ACP interactions, but only the heterododecameric yFAS fulfills the distance restraints for productive ACP interaction with solvent accessible surface (SAS) distances of the ACP Ppant attachment point to the KS active site cysteine of approximately 27 Å (Supplementary Figure 4.15c,f-i). The position and orientation of ACP relative to KS is conserved between yFAS and the CTB1 NR-PKS, despite a completely different integration of the KS domain in a barrel-like scaffold and the presence of an unusual four-helical extension of the canonical ACP domain in yFAS. The corresponding SAS distance in SAT°-KS-MAT°=ACP2 is 28 Å, in agreement with a covalently trapped functional interaction (Supplementary Figure 4.15c,f). The ACP-KS interaction observed in PikAIII, which is non-productive

based on SAS distances (Supplementary Figure 4.15h,i), however, is considerably different. The ACP docking surface observed in CTB1, and verified by mutant activity assays in deconstruction experiments, is partially blocked in PikAIII due to the divergent location of the LD domain (Supplementary Figure 4.15d,e).

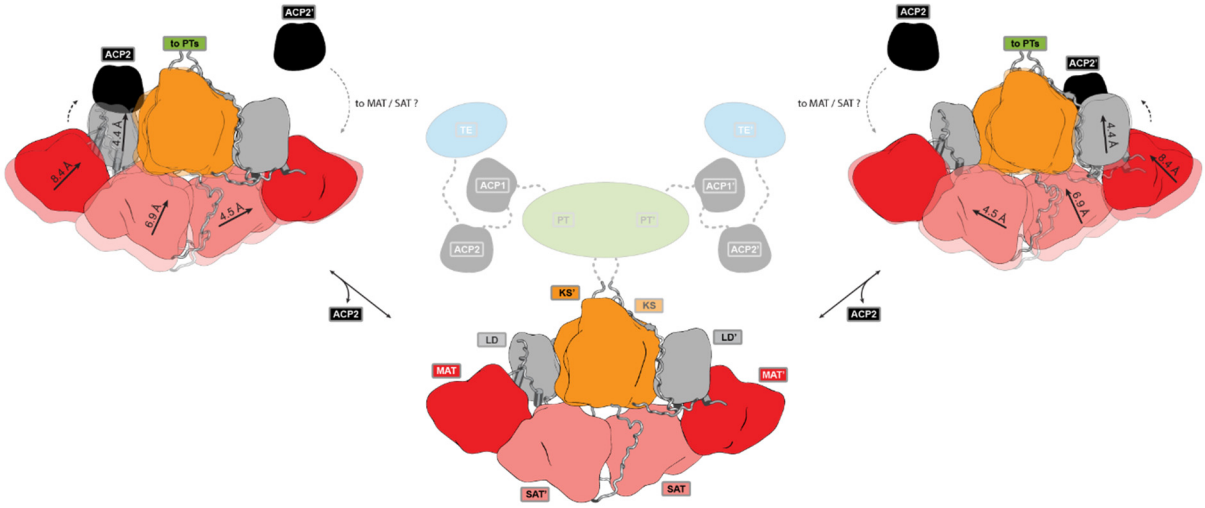


Figure 4.4 | Schematic illustration of suggested modes of conformational coupling in CTB1.
ACP2 preferentially binds to a single KS active site. ACP2 binding results in coupled conformational changes across the dimeric CTB1, disfavoring productive ACP2-KS interaction in the other protomer and re-aligning active sites for ACP2 interaction and substrate shuttling. Such conformational coupling resembles a V-twin engine mechanism for alternating KS-ACP2 interactions.

The cryo-EM solution structure of SAT°-KS-MAT°=ACP2 reveals considerably greater asymmetry than the crystal structure without ACP2, strongly suggesting that the predominant asymmetric conformation derives from ACP2 binding by either an induced fit or conformational selection mechanism[278]. Relative to the crystal structure, the SATs are translocated individually, not as a rigid dimer, towards the central KS domains (Supplementary Figure 4.16a,b). ACP2 binding caused an upward movement of the proximal LD-MAT region by 4.4 Å and 8.4 Å, respectively, (Figure 4.4, Supplementary Movie 4.1) resulting in an opening of the interface between the SAT and MAT domains (Figure 4.2d, Supplementary Figure 4.16c). Coordinated with this separation, the proximal SAT domain moved toward the KS by 6.9 Å, substantially enlarging the SAT-KS interface, while the second SAT is pushed (by 4.5 Å) against the second pair of KS and MAT domains (Supplementary Figure 4.16d). Since large fragments and domains of various NR-PKSs can be functionally interchanged in deconstruction experiments, coupling of structural changes and catalysis might be representative for many NR-PKSs[48,91].

Conformational asymmetry and coupling have previously been observed in iteratively acting fully-reducing iPKS[44] and mFASs[160,162,165]. The asymmetric conformational states observed in these systems, as well as in SAT°-KS-MAT°=ACP2 are derived by local hinge-bending or shear motions, maintaining the individual domain structures, key conserved interfaces, and the overall interdomain architecture. For non-iteratively acting modPKSs, however, currently no information is available on functional asymmetry. In contrast, symmetric large scale conformational transitions with the requirement for partial refolding of conserved domains have been visualized by EM analysis of distinct functional states of the modPKS PikAIII on the basis of a divergent condensing region architecture[42,45]. Functional asymmetry in the interaction of ACP with dimeric enzymatic domains has previously been suggested in the form of a seesaw-type conformational coupling for a bacterial hexameric dehydratase (FabZ)

involved in fatty acid biosynthesis[34], which binds ACPs in a 6:3 ratio. However, a stoichiometrically ACP-crosslinked structure of the bacterial dimeric dehydratase FabA demonstrates that asymmetry is not a general feature in related dehydratases, even though this structure also revealed increased disorder for the second ACP, as indicated by the crystallographic B-factors[13]. A distinct type of persistent structural asymmetry is observed in the iterative actinorhodin type II PKS, where asymmetry is implemented by heterodimerization of one active KS protomer with a catalytically inactive KS homolog, the chain length factor (CLF)[279]. Iterative elongation is catalyzed only at the KS active site, while the CLF determines maximum chain length by restricting the available volume of the substrate binding tunnel[279].

Here, the crosslinked interaction of an ACP induces conformational asymmetry across the CTB1 SAT-KS-MAT loading/condensing region. Binding of an early elongation intermediate, mimicked here by the α-bromopropionyl moiety, may cause local asymmetry around the KS active sites. However, the observed overall asymmetry is more plausibly explained by binding of the ACP, which directly connects the KS active site to the surrounding domains via the LD, which is part of the ACP binding interface.

Substrate elongation in NR-PKSs occurs rapidly, and reactive intermediates remain bound in the active site tunnel of the KS, as suggested by the absence of detectable ACP-bound chain elongation intermediates in PksA[85]. After initial loading, the ACP, however, shuttles between the KS and the MAT active site, for delivery of malonyl-CoA extender-units. Presuming that ACP binding is linked to overall asymmetry in the SAT-KS-MAT region, productive binding of an extender-unit-charged ACP (malonyl-ACP) to one KS active site would disfavor binding to the other KS active site in every elongation cycle. Product elongation by the dimeric KS would occur with alternating asymmetric states analogous to a V-twin engine, favoring substrate loading by the MAT on one side and elongation by the KS on the other side of the dimer (Figure 4.4, Supplementary Movie 4.1). Considering that the ACP binding sites are located proximal to the attachment points for the downstream modification (PT) and release (TE) domains, ACP could plausibly contribute to a coordination of substrate shuttling and overall product formation analogous to conformational coupling[44] and large scale rearrangements[42,45] observed in modifying regions of other PKSs. However, further analysis will be required to dissect the relative contributions of substrate loading and ACP interaction on asymmetry in iterative and modular PKSs.

Understanding the architectures and dynamic properties of PKSs remains a central challenge in natural product research. Although all canonical types of PKS domains have been studied in detail in isolation, understanding substrate loading and transfer, as well as the functional integration of individual domains into large multienzymes or even assembly lines remains a challenge. Our analysis of CTB1 SAT-KS-MAT now reveals the organization of an integrated loading architecture in a PKS, and suggests a mechanism for conformational coupling between the two protomers of the dimeric PKS promoted by substrate loading to KS. These results provide an important step towards a comprehensive understanding of conformational coupling in PKS multienzyme function and afford critical insights for the dissection and re-engineering of initial substrate-loading for polyketide production.

4.5 Acknowledgements

We thank Peter Leadlay and Lorena Betancor for providing plasmid pETcoco-2A-L1SL2. X-ray diffraction data were collected at beamline PXI of the Paul Scherrer Institute, Villigen, Switzerland, and cryo-EM data were collected at the BioEM facility of the University of Basel; we acknowledge excellent support by teams of both facilities. This work was supported by the Swiss National Science Foundation (SNF)

project grants 138262, 159696, SNF R'equip grants 145023 and 164074, and the National Institutes of Health (ES001670). D.A.H. acknowledges a fellowship from the Werner-Siemens Foundation.

4.6 Author Contributions

R.P.J. expressed, purified and crystallized CTB1 SAT-KS-MAT. D.A.H., R.P.J., and T.M. solved the crystal structure. D.A.H. performed cryo-EM, data processing, modelling, refinement and analysis of all structural data. C.H.R. optimized and prepared crosslinked CTB1 SAT°-KS-MAT°=ACP2 for structural analysis and performed mutational experiments for structural validation. J.M.K. synthesized the α-bromopropionyl crosslinker. P.A.S. and J.R.A. performed initial exploratory experiments, and JRA prepared the CTB1 SAT°-KS-MAT° construct for crosslinking. C.A.T. and T.M. designed research. The manuscript was written by D.A.H., T.M., C.H.R., and C.A.T..

4.7 Competing financial interests

Authors declare no competing financial interests.

4.8 Methods

Methods, including statements of data availability and any associated accession codes and references, are available in the online version of the paper.

4.9 Additional information.

Any supplementary information, chemical compound information and source data are available in the online version of the paper. Reprints and permissions information is available online at http://www.nature.com/reprints/index.html. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to T.M.

4.10 Online Methods

4.10.1 Cloning.

Plasmids used in this study were assembled using standard protocols. CTB1 SAT°-KS-MAT° was generated by site directed introduction of the mutations C119A and S1010A. All plasmids used in this study are summarized in Supplementary Table 4.3. The plasmids for expression of wild-type CTB1 SAT-KS-MAT (pECTB1-NKA6), PT (pECTB1-PT), and TE (pECTB1-TE) have been previously described[276]. Primers used for assembly of new plasmids in this study are detailed in Supplementary Table 4.4. All expression constructs were prepared and maintained in *E. coli* BL21(DE3). Cut sites for CTB1 ACP2 deconstruction were chosen based on the PksA ACP solution structure[11]. CTB1 ACP2 was cloned as an *N*-terminal Thrombin-cleavable His_6-tagged construct to allow for differential purification throughout the crosslinking process. The desired gene sequence was amplified by PCR from pECTB1-ACP (tandem ACP, previously described)[276] using CTB1 ACP2-5 and CTB1 ACP2-stop-3, and ligated into pET-28a at NdeI and NotI sites using T4 DNA ligase. Active-site mutations for selective crosslinking were introduced into pECTB1-NKA6 by Gibson assembly[62] of PCR-amplified fragments using standard protocol, giving pECTB1-SKM-C119A-S1010A. Mutations to CTB1 SAT-KS-MAT for interface validation

were introduced via Gibson assembly in the same manner. Assembled plasmids were screened by restriction digestion and sequences confirmed by the Johns Hopkins Synthesis and Sequencing Facility (Baltimore, MD). Sequencing revealed the point mutation T321A (compared to Uniprot Q6DQW3) in all SAT containing constructs.

4.10.2 Protein expression and purification for crosslinking and interface validation

Proteins were expressed as His_6-tagged constructs in *E. coli* BL21(DE3). One liter Terrific Broth cultures supplemented with 25 µg/mL kanamycin (GoldBio) were inoculated with overnight starter cultures, and grown at 37 °C with shaking at 250 rpm until OD_{600} reached 0.7. Cultures were cold-shocked in ice water for 1 h prior to induction with 0.5 mM isopropyl-β-D-thiogalactopyranoside (IPTG) (GoldBio). Expression was carried out at 19 °C with shaking at 250 rpm for approximately 16 h. Cells were harvested by centrifugation at 4,000 × g for 15 min and flash frozen in liquid nitrogen for storage at -80 °C.

Cell pellets were thawed in lysis buffer A (50 mM potassium phosphate, pH 8, 300 mM NaCl, 10% (v/v) glycerol) and lysed by sonication. Lysate was cleared by centrifugation at 27,000 x g for 25 min. The resulting supernatant was batch bound to Co^{2+}-TALON resin (Clontech) at 4 °C, typically for 1 h. The protein-bound resin was applied to a gravity-flow column and washed and eluted with lysis buffer A containing increasing concentrations of imidazole (0-100 mM), as directed by the manufacturer. Fractions containing the protein of interest were pooled and dialyzed against the appropriate buffer (50 mM Tris pH 7.5, 5% (v/v) glycerol unless otherwise indicated) at 4 °C. For dialysis of CTB1 SAT°-KS-MAT°, 1 mM DTT was included to maintain the KS active-site cysteine in its reduced form. If necessary, the protein was concentrated in Amicon Ultra centrifuge filters (Millipore). Protein concentration was quantified based on absorbance at 280 nm on a Cary 50 UV-Vis spectrophotometer and extinction coefficients calculated using Expasy ProtParam (Supplementary Table 4.3). CTB1 ACP1 concentration was determined in triplicate by Bradford assay (Bio-Rad) using bovine serum albumin as a standard.

4.10.3 Proteolysis of tag from CTB1 ACP2-His

To facilitate separation of ACP2 from both crosslinker loading reactions and crosslinking reactions, the *N*-terminal His_6-tag was removed by thrombin proteolysis using a Thrombin CleanCleave Kit (Sigma Aldrich). The thrombin agarose resin was washed with reaction buffer (50 mM Tris pH 8, 10 mM $CaCl_2$) and conditions were optimized as recommended by the manufacturer. Complete cleavage was achieved using approximately 100 µL thrombin agarose resin per 50 mg CTB1 ACP2 at 2 mg/mL CTB1 ACP2 in reaction buffer. Reactions were rotated for 3 h at 25 °C. Cleavage was monitored by SDS-PAGE and confirmed by MALDI-TOF (Bruker Auto Flex III). Cleaved CTB1 ACP2 was isolated by brief centrifugation at 500 × g to pellet the thrombin resin and dialyzed in 50 mM Tris pH 7.5, 5% (v/v) glycerol at 4 °C.

4.10.4 Expression and purification of CTB1 SAT-KS-MAT for crystallization

C-terminal hexa-histidine (His_6) tagged CTB1 SAT-KS-MAT (pECTB1-NKA6, aa 1-1293) was overexpressed with *Streptomyces* chaperonins (pETcoco-2A-L1SL2 plasmid)[202] in *E. coli* BL21(DE3) pRIL. Cells were cultured in 2×YT medium, supplemented with 0.5 % (v/v) glycerol, NPS (25 mM $(NH_4)_2SO_4$, 50 mM KH_2PO_4, 50 mM Na_2HPO_4), kanamycin (100 µg/mL), chloramphenicol (34 µg/mL), and ampicillin (100 µg/mL). An expression culture (1.5 L) was inoculated (1:20), grown at 37 °C for 2 h, cooled to 20 °C, and induced with IPTG (0.1 mM) at an OD_{600} of 1.0. Cells were harvested after 12 h by centrifugation (7,000 × g) and resuspended in lysis buffer B (50 mM HEPES pH 7.4, 200 mM NaCl, 2.5mM $MgCl_2$, 40

mM imidazole, 10% (v/v) glycerol, 5 mM β-mercaptoethanol), supplemented with protease inhibitors (200 µM phenylmethylsulfonyl fluoride, 20 µM bestatin, 4 µM E64, 2 µM pepstatin A, 20 µM phenantrolin, 2 µM phosphoramidon) as well as DNase I, RNase A, and lysozyme. Cells were placed on ice and lysed by sonication. The lysate was cleared by centrifugation (100,000 × g, 30 min) and immobilized by metal-affinity chromatography on a 25 mL Ni-affinity column (GenScript) pre-equilibrated with lysis buffer B. Unbound protein was eluted with five column volumes (CV) of HisA buffer (50 mM HEPES pH 7.4, 200 mM NaCl, 2.5mM $MgCl_2$, 40 mM imidazole, 10% (v/v) glycerol, 5 mM β-mercaptoethanol, protease inhibitors). The sample was eluted with a linear gradient to 100% HisB buffer (50 mM HEPES pH 7.4, 500 mM imidazole, 50 mM NaCl, 10%(v/v) glycerol, 2.5 mM β-mercaptoethanol, and protease inhibitors) and directly loaded on a strong anion exchange column (PL-SAX 4,000 Å, 10 µm, Agilent). The protein was eluted with a linear gradient from 0% AIC-A (50 mM HEPES pH 7.4, 50 mM NaCl, 10% (v/v) glycerol, 2.5 mM β-mercaptoethanol) to 100% AIC-B (50 mM HEPES pH 7.4, 1 M NaCl, 10% (v/v) glycerol, 2.5 mM β-mercaptoethanol). The protein was concentrated with Amicon Ultra centrifuge filters (Millipore) and subjected to size exclusion chromatography (SEC) (Superdex 200 16/60, GE Healthcare) using 20 mM HEPES pH 7.4, 250 mM NaCl, 5% glycerol (v/v), 1 mM dithiothreitol. Protein-containing fractions were pooled and concentrated in Amicon Ultra centrifuge filters (Millipore) to 10-15 mg/mL and frozen in liquid nitrogen.

4.10.5 Synthesis of α-bromopropionyl aminopantetheine.

For full synthetic details and compound characterization see Supplementary Note 4.1.

4.10.6 Phosphorylation of α-bromopropionyl aminopantetheine.

α-bromopropionyl aminopantetheine was phosphorylated enzymatically under the following conditions: 0.25 mg/mL PanK (pantothenate kinase), 2 mM α-bromopropionyl aminopantetheine, 5 mM ATP, 20 mM KCl, 10 mM $MgCl_2$, 50 mM Tris pH 7.5 at room temperature for approximately 20 h. Complete phosphorylation was confirmed by UPLC-ESI-MS (Waters Acquity/Xevo-G2, negative ion mode). All protein was removed from the reaction with a 3 kDa MWCO Amicon Ultra centrifuge filter (Millipore). The flow-through was flash frozen in liquid nitrogen and lyophilized to dryness. The resulting white powder was resuspended in 5% aqueous acetonitrile. α-bromopropionyl aminopantetheine was HPLC purified on an Agilent 1100 equipped with a Kinetex C18 semi-prep column (250 x 10 mm, 5 µ, Phenomenex) using the following method at 4 mL/min: hold 5% solvent A/95% solvent B for 3 minutes, 5-35% solvent A over 17 minutes, 35-95% solvent A over 5 minutes, followed by re-equilibration to 5% solvent A, where solvent A is acetonitrile + 0.1% trifluoroacetic acid and solvent B is 0.1% trifluoroacetic acid. Purity and identity of the isolated compound was confirmed by UPLC-ESI-MS as above, and the pooled fractions were lyophilized to dryness.

4.10.7 Loading of ACP2 with α-bromoacyl crosslinker.

α-bromopropionyl aminophosphopantetheine was adenylated and subsequently loaded onto CTB1 ACP2 in a one-pot enzymatic reaction under the following conditions: 0.5 mM α-bromopropionyl aminophosphopantetheine, 0.5 mg/mL PPAT (phosphoribosyl pyrophosphate amidotransferase), 5 µM Sfp (surfactin 4'-phosphopantetheinyl transferase), 0.5 mM CTB1 ACP2, 5 mM ATP, 10 mM $MgCl_2$, 20 mM KCl, 50 mM Tris pH 7.5, 10% glycerol. Reactions were run at room temperature and monitored by MALDI-TOF (Bruker AutoFlex III). Once loading was complete, the His-tagged Sfp and PPAT were separated from untagged CTB1 ACP2 by passing the reaction solution through an equilibrated Co^{2+}-TALON column (Clontech) and recovering the flow-through. Additionally, any unreacted crosslinker

was removed by repeated concentration and dilution with a 3 kDa MWCO Amicon Ultra centrifuge filter (Millipore). Complete removal of free crosslinker was confirmed by UPLC-ESI-MS analysis.

4.10.8 Crosslinking of CTB1 SAT°-KS-MAT° to ACP2.

CTB1 SAT°-KS-MAT° was combined with 20-fold excess CTB1 ACP2 bearing the α-bromoacyl crosslinker in 50 mM Tris pH 7.5, 10% glycerol (v/v) at room temperature for 1 h. Crosslinking was monitored by SDS-PAGE, and found to be approximately 95% complete after 1 h. Crosslinked SAT°-KS-MAT°=ACP2 was separated from remaining free SAT°-KS-MAT° by anion exchange chromatography. Q-Sepharose Fast Flow resin (GE Healthcare) was equilibrated with 50 mM Tris pH 7.5, 10% glycerol, and the crosslinking reaction mixture was applied to the column. SAT°-KS-MAT° and SAT°-KS-MAT°=ACP2 were separated using a gradient of potassium chloride from 0-500 mM. To remove free ACP2 from the crosslinked protein, fractions containing SAT°-KS-MAT°=ACP2 were pooled and applied to an equilibrated Co^{2+}-TALON column. ACP2 was removed in the flow-through, and pure SAT°-KS-MAT°=ACP2 was eluted with a gradient of imidazole from 2-100 mM.

4.10.9 *In vitro* enzymatic reactions.

CTB1 SAT-KS-MAT (wild-type and mutants of proposed interface residues), PT, ACP1, ACP2, and TE were purified as described above, and dialyzed into 100 mM potassium phosphate pH 7, 10% glycerol. ACP1 and ACP2 were activated enzymatically with CoA for 1 h at 25 °C under the following conditions: 2 μM Sfp, 0.5 mM CoA, 10 mM $MgCl_2$, 200 μM ACP2, 100 mM potassium phosphate pH 7, 10% glycerol.

To compare the biosynthetic competency of ACP1 and ACP2, specifically the ability of the ACP to interact with KS, minimal PKS reactions containing only SAT-KS-MAT and ACP were performed. ACP1 and ACP2 concentrations were determined in triplicate by Bradford assay using bovine serum albumin as a standard. 10 μM SAT-KS-MAT was combined with 10 μM ACP1, 10 μM ACP2, or 5 μM of each ACP in assay buffer (100 mM potassium phosphate pH 7, 10% glycerol, 1 mM TCEP). Reactions were initiated by addition of 0.5 mM acetyl- *S*-*N*-acetylcysteamine (SNAC) and 2 mM malonyl-SNAC, and incubated for 4 h at 25 °C. The 250 μL reactions were quenched with 10 μL hydrochloric acid and extracted thrice with ethyl acetate. Extracts were pooled, dried, and resuspended in 250 μL 20% aqueous acetonitrile for HPLC analysis.

Product profiles of each enzyme mutant were analyzed on an Agilent 1200 equipped with a Prodigy ODS3 analytical column (4.6 x 250 mm, 5 μ, Phenomenex). 100 μL injections were separated by a bisolvent method at 1 mL/min: 5-85% solvent A over 40 min, 85-95% solvent A over 10 min, re-equilibrate to 5% solvent A over 10 min, where solvent A is acetonitrile + 0.1% formic acid, and solvent B is 0.1% formic acid. Chromatograms were recorded at 280 nm, and UV-vis spectra were recorded over a range of 200-800 nm. High resolution mass data were collected on a Waters Acquity/Xevo-G2 in positive ion mode. Previously characterized CTB1 derailment products were identified by their UV-Vis spectrum and exact mass[276].

For CTB1 interface validation, the full PKS was reconstituted with each SAT-KS-MAT mutant under the following conditions: 10 μM CTB1-SAT-KS-MAT, 10 μM PT, 10 μM ACP2, and 10 μM TE were combined in assay buffer. Reactions were initiated by addition of 0.5 mM acetyl-SNAC and 2 mM malonyl-SNAC, and incubated for 4 h at 25 °C. Reactions were extracted and analyzed by HPLC as above.

4.10.10 Crystallization, data collection and structure determination of CTB1 SAT-KS-MAT.

CTB1 SAT-KS-MAT plate-like crystals were obtained by the sitting drop vapor diffusion method in 0.2 M $MgCl_2$, 0.1 M Bis-Tris Propane pH 6.5, 18% (v/v) PEG3350 at 4 °C, a drop ratio of 1.0 µL + 0.5 µL and further optimized by seeding. Crystals grew to a final size of 0.4×0.04×0.04 mm^3 within one week. The crystals were dehydrated and cryo-preserved by successively increasing the concentration to 0.5 M $MgCl_2$, 0.1 M Bis-Tris Propane pH 6.5, 25% PEG3350, 22% ethylene glycol. Diffraction data of crystals in space group $C222_1$ (a= 108.1 Å, b= 230.2 Å, c= 253.8 Å) and 56% solvent content were collected at the Swiss Light Source (SLS, Villigen, Switzerland) at beamline X06SA, a temperature of 100 K and a wavelength of 1.000 Å. Data reduction was performed using XDS and XSCALE[203] and data were analyzed using phenix.xtriage[204]. The structure was solved with PHASER[205] using molecular replacement and the KS and AT domains of DEBS module five[163] as well as the AT domain of CurL[46] as search models. An initial model was obtained in iterative cycles of rebuilding with BUCCANEER[206] and density modification using solvent flattening and histogram matching in PARROT[208]. Iterative cycles of manual model building and real space refinement in Coot[207] and TLS refinement in Phenix[204] yielded a high quality model with R_{work}/R_{free} values of 0.21/0.24 at 2.8 Å resolution with excellent geometry (Ramachandran favored/outliers: 97.07%/0.00%) (Supplementary Table 4.1a, Supplementary Figure 4.5).

4.10.11 EM sample preparation and data collection.

Initial sample quality was evaluated by conventional negative-stain EM. Initial cryo-EM analysis revealed a monomer-dimer equilibrium, which has previously been observed for PKS and FAS fragments[42-44,46,159]. Buffer conditions were optimized by analyzing SAT-KS-MAT by analytical size-exclusion chromatography (SEC) (4.6/250 Superose 6 Increase, GE Healthcare)at a flow rate of 0.1 mL/min followed by grid preparation and testing the maximal achievable particle concentration. The results indicated a fast equilibrium and a mostly dimeric species in low salt buffer (20 mM Tris pH 7.4, 50 mM NaCl, 2.5 mM tris(2-carboxyethyl)phosphine (TCEP)) (Supplementary Figure 4.10e). For cryo-EM, crosslinked sample was concentrated to 15-16 mg/mL, applied to analytical SEC and collected in fractions of 50 µL. For grid preparation 4 µL of individual fractions were applied to glow discharged (30 s) lacey carbon grids (300 mesh, copper), immediately blotted for 2 s and plunge-frozen in liquid ethane using a Vitrobot (FEI, Vitrobot III). An incubation step was omitted due to fast adsorption of the particles to the carbon. The integrity of the crosslinked sample was analyzed by denaturing polyacrylamide gel electrophoresis (PAGE) of the remaining sample after grid preparation (Supplementary Figure 4.10f).

Grid quality was analyzed using a Philips CM200 FEG cryo-transmission electron microscope, operated at an acceleration voltage of 200 kV. Optimal conditions were identified for peak fractions diluted to 270 µg/mL, in a trade-off between low particle density (approx. 70 particles per micrograph), monomer formation and particle adsorption to carbon. Final data was collected using a Titan Krios electron microscope (FEI), operated at 300 kV, a GIF Quantum LS imaging filter (Gatan) and a K2 Summit (Gatan) operating in counting mode. Images were acquired at 0.8-4.5 µm defocus and a nominal magnification of 105,000x, corresponding to a pixel size of 1.326 Å (Supplementary Table 4.1b). Movies were collected with a total dose of approx. 90 $e^-/Å^2$ per 18 s exposure, fractionated over 60 frames.

4.10.12 EM data processing and analysis.

Drift correction of dose fractionated frames[280] was performed using Zorro[281] and global contrast transfer function (CTF) parameters were determined using CTFFIND4.1[282]. Particles were picked from aligned dose-filtered averages using the swarm semi-automated procedure as implemented in e2boxer.py[283], followed by extraction from unfiltered averages using Relion-2[284] and local CTF parameter refinement using Gctf v1.06[285]. Single particle analysis included several rounds of reference-free maximum *a posteriori*-2D classification[284] to remove junk particles and monomers, which revealed a clearly distinguishable shape and could be reconstructed at low resolution (Supplementary Figure 4.10d, Supplementary Figure 4.11a). For reconstruction of the dimeric, crosslinked SAT°-KS-MAT°=ACP, an initial volume was generated using e2initialmodel.py[283] and low-pass filtered to 50 Å. Exhaustive global followed by local reference-free 3D classification and alignment without symmetry were applied, yielding highly similar classes that were combined in a consensus refinement prior to movie refinement and particle polishing as implemented in Relion-2[284]. Parameters for B-factor fitting as function of movie frames have been carefully examined as well as the different number of frames in the particle movie stacks. At this stage all 60 frames were included in the weighted particle movie stacks as no significant difference between including different amounts of frames per movie could be identified after applying weighting. Another round of 2D and 3D classification yielded three almost identical volumes that were refined as individual classes and in all possible combinations of classes. The combination of two classes (Supplementary Figure 4.11a) resulted in a slightly improved model at 7.3 Å resolution (25,107 particles) as compared to all three classes at 7.5 Å (44,859 particles). For the refinement of the final map at 7.1 Å resolution weighted particle averages of 27 movie frames were used, corresponding to a total dose of 41 $e^-/Å^2$. All reported resolution estimates are based on the 0.143 threshold criterion of the Fourier Shell Correlation (FSC)[286] between two halves of the dataset refined independently in RELION-2, after accounting for correlations introduced by masking[287]. Throughout all refinements in Relion-2[284], particles were grouped by k-mean clustering of their CTF parameters with a minimum of 50 particles per group. The alignment was not influenced by a preferred orientation[288] (Supplementary Figure 4.11b,c), which was assessed by reducing the number of particles per angular sampling group to the mean of all groups by random subset deletion in four independent refinements. Later refinements were carried out using a soft mask and solvent corrected FSC during the refinement[287]. Final maps were sharpened with a sharpening B-factor of -350 $Å^2$. Local resolution was calculated with ResMap[289] and indicated higher resolution for secondary structure elements in the center of the particle (Supplementary Figure 4.11d,e). Therefore, masked 3D classification and refinement after signal subtraction[290] was carried out, resulting in slightly improved resolution, but without significant differences in the electron density map. Map generation from coordinates as well as filtering and file type conversion was done using EMAN2 tools[283]. For analysis, map transformations were applied using the CCP4 tool maprot[224]. Maps were resampled using UCSF Chimera[291] and difference maps were calculated Python Numpy and MRCZ[292].

4.10.13 Cryo-EM structure refinement and modelling.

Modelling of the cryo-EM structure was started by rigid body fitting of individual domains of the CTB1 SAT-KS-MAT crystal as well as of an ACP2 homology model in Coot[207]. The orientation of ACP2 could be unambiguously determined based on steric requirements of the Ppant attachment site (DSL motif), a continuously connected density between ACP2 helix II and the KS active site entrance, and additional density features which can be explained by loop regions of the homology model. The model for CTB1 ACP2 was generated using SwissModel[211] based on PksA ACP[11], which shares 40% sequence identity

and can be functionally replaced with the CTB1 tandem ACPs in *in vitro* assays[48,91]. Initial B-factors of the model were scaled relative to B-factors of atomic displacement parameter (ADP) refined crystal structures in the PDB at similar resolution. Prior to map-based real-space refinement, the model was protonated using phenix.reduce and geometry minimization with secondary structure restraints and reference coordinate restraints was applied using phenix.geometry_minimization[204]. Afterwards global minimization and ADP

refinement with reference structure restraint was carried out against the unsharpened cryo-EM map in phenix.real_space_refine[204]. Reference structure and nonbonded distance restraint weights were carefully monitored and optimized. As the local resolution around the resolved ACP2 is not sufficient to unambiguously interpret possible conformational changes of loop regions relative to the isolated PksA ACP NMR structure, the loop regions were excluded from the final model after all atom refinement, without affecting model statistics (Supplementary Table 4.1b). The resulting model contained the regions corresponding to the canonical ACP four helix bundle (residues: 1782-1796, 1816-1828, 1839-1842, 1846-1854) and was validated using phenix.molprobity[204] with Ramachandran plot statistics of "favored/outliers": 96.47%/0.12%. A position for a second crosslinked ACP2 could not be undoubtedly resolved in any intermediate or final EM map. The sample has been analyzed after grid preparation and is >90% crosslinked (Supplementary Figure 4.10f).

Due to residual noise around the putative second ACP2 binding site the presence of a second ACP2 in a mostly disordered state without a solid interface, tethered only by covalent crosslinking (Supplementary Figure 4.12a,b) cannot be strictly excluded. However, we cannot exclude the possibility that effects of selective adsorption of partially or fully crosslinked SAT°-KS-MAT°=ACP2 or of selective destabilization under grid preparation also contribute to selective imaging of a single-crosslinked state with low population in the sample used for grid preparation.

4.10.14 Structure analysis and visualization.

Related structures were identified using PDBeFold (20% query / 10% target)[222] and interfaces were analyzed using QtPISA v2.1.0[224]. Transformations and coordinate manipulations were carried out using CCP4 tools and MOLEMAN[225]. Bias-removal for crystallographic F_o-F_c omit maps was achieved by applying a random perturbation to coordinates (Δ0.2 Å) and B-factors (Δ20% of the mean overall B-factor) with MOLEMAN2[225] prior to refinement. Cα-r.m.s.d. between domains of the crystal and cryo-EM structure were calculated using LSQMAN[232]. Structures of PikAIII were modelled according to Dutta *et al.*[42] using Chimera[291]. Solvent accessible surface distances (SAS) between C_α residues of ACP phosphopantetheinylation and KS active site residues were calculated using Xwalk[293]. Figures, movies and plots were generated using Pymol (Schrodinger, v1.8.6.0) and Python.

4.10.15 Sequence analysis.

Sequence alignments were generated using Clustal Omega[269]. A Phylogenetic tree was generated using the neighboring joining algorithm in Geneious (Biomatters Limited, v8.1.6).

4.10.16 Data availability.

Atomic coordinates, structure factors for the reported crystal structure, and cryo-EM map have been deposited with the Protein Data Bank under accession codes 6FIJ, 6FIK and EMDB accession code EMD-4266.

4.10.17 Code availability.

Custom computer code is available upon request.

4.11 Supplementary Information

Supplementary Table 4.1 | X-ray and cryo-EM data collection and refinement statistics.

a

	CTB1 SAT-KS-MAT
Data collection	
Space group	$C222_1$
Cell dimensions	
a, b, c (Å)	108.05, 230.2, 253.8
α, β, γ (°)	90.0, 90.0, 90.0
Resolution (Å)	126.9 – 2.77
R_{merge} (%)*	21.2 (193.8)
$I/\sigma I$*	10.13 (1.03)
$CC_{1/2}$*	99.5 (43.3)
Completeness (%)*	99.2 (91.9)
Redundancy*	8.5 (5.6)
Refinement	
Resolution (Å)	63.45 – 2.77
R_{work}/ R_{free}	0.21 / 0.24
No. atoms	39,201
Protein	38,682
Ligand/ion	167
Water	352
B-factors	70.17
Protein (Å^2)	70.31
Ligand/ion (Å^2)	84.20
Water (Å^2)	48.78
R.m.s deviations	
Bond lengths (Å)	0.002
Bond angles (°)	0.496

b

	CTB1 SAT°-KS-MAT°=ACP2
Data collection and processing	
Nominal magnification	105,000x
Voltage (keV)	300
Electron exposure (e- Å^{-2}) initial / final	90 / 41
Defocus range(μm)	0.8-4.5
Pixel size (Å)	1.326
Symmetry imposed	C1
Movies	1,728
Frames per movie	60
Initial particles images	122,089
Final particle images	25,107
Refinements	
Initial model used	ab initio
Model resolution (Å)	7.1
FSC threshold	0.143
Model resolution range(Å)	4-10
Map sharpening B-factor (Å^2)	-350
Model composition	
Non-hydrogen atoms	19,628
Protein residues	2,591
Ligands	0
R.m.s deviations	
Bond lengths (Å)	0.010
Bond angles (°)	1.290
Validation	
MolProbity score	1.57
All-atom clashscore	6.09
Rotamer outliers (%)	0.00
Ramachandran plot	
Favored (%)	96.47
Allowed (%)	3.41
Disallowed (%)	0.12

(a) Crystallographic data collection and refinement statistics. The resolution cutoff was determined by $CC_{1/2}$ criterion (Karplus, P. A. & Diederichs, K. Science 336, 1030-1033 (2012)). *, Highest resolution shell is shown in parentheses. **(b)** Cryo-electron microscopic data collection and refinement statistics.

Supplementary Table 4.2 | Structural comparison and interface analysis of CTB1 SAT-KS-MAT.

a

Structure 1	Structure 2	C_α r.m.s.d. [Å]	Aligned residues
CTB1 SAT	CTB1 SAT (2nd chain)	0.34	350
	Closest* in PDB (3G87) no multienzyme	2.56	262
	human FAS MAT (3HHD)	2.59	249
	CazM SAT (4RPM) loading domain	2.63	272
	PksE AT (5DZ7)†	2.64	245
	procine FAS MAT (2VZ9)	2.65	244
	PKS13 AT (3TZX)	2.75	249
	PksC AT (5DZ6) †	2.78	251
	CTB1 MAT	2.81	255
	DEBS AT_5 (2HG4)	2.81	254
	CurL AT (4MZ0)	2.83	251
	DEBS AT_3, (2QO3)	2.86	261
	DisD AT (3RGI)†	2.87	258
	VinK (5CZD)	2.87	242
	AVES1 AT (4RL1) loading domain	2.89	256
	ZmaA (4QBU)	3.01	244
	DYNE8 AT (4AMP)	3.05	217
	MAS-like AT (5BP1)	3.18	259
	OzmQ partial AT (4OQJ)	3.22	92
CTB1 KS	CTB1 KS (2nd chain)	0.23	435
	PksL KS (5ENY)	1.40	392
	C0ZGQ5 KS (4Z37)	1.40	398
	CurL KS (4MZ0)	1.41	400
	MgsF KS (4TKT)†	1.41	401
	PksJ KS (4NA3)†	1.42	404
	human FAS KS (3HHD)	1.44	389
	procine FAS KS (2VZ9)	1.48	388
	DEBS KS_3, (2QO3)	1.49	406
	DEBS KS_5 (2HG4)	1.49	392
	MgsE KS (4QYR)†	1.52	400
	OzmH KS0 (4OPF)†	1.52	339
	MAS-like KS (5BP1)	1.55	347
	OzmN KS (4WKY)†	1.60	374
	RhiE KS (4KC5)†	1.61	399
	Closest* - FabF KS (4JB6) no multienzyme	1.64	378
	OzmQ (4OQJ)†	1.75	375
CTB1 MAT	CTB1 MAT (2nd chain)	0.23	304
	CurL AT (4MZ0)	1.72	281
	MAS-like AT (5BP1)	1.82	285
	PKS13 AT (3TZY)	1.83	290
	human FAS MAT (3HHD)	1.83	280
	procine FAS MAT (2VZ9)	1.85	279
	DYNE8 AT (4AMP)	1.86	231
	DEBS AT_3, (2QO3)	1.87	285
	DEBS AT_5 (2HG4)	1.89	290
	Closest* in PDB: AVES1 AT (4RL1) loading domain	1.92	284
	ZmaA (4QBU)	2.00	275
	DisD AT (3RGI)†	2.02	269
	PksC AT (5DZ6)†	2.06	266
	PksE AT (5DZ7)†	2.09	264
	VinK (5CZD)	2.26	261
	OzmQ partial AT (4OQJ)	2.50	89
	CazM SAT (4RPM) loading domain	2.68	252

b

Interface 1	Interface 2	Min area [Å^2]	Max area [Å^2]	Hydrogen bonds	Salt bridges
SAT	KS-LD-MAT-post MAT linker	1197.2	1230.1	6-8	3
SAT[§]	KS[§]	183.8	184.9	1-2	0
SAT[§]	MAT[§]	946.1	966.8	5-6	2-3
SAT	SAT	323.7	323.7	0	0
SAT dimer	KS dimer	659.9	659.9	3	0
SAT dimer	KS-LD-MAT-post MAT linker dimer	2718.5	2718.5	14	7
SAT[§]	SAT-KS linker[§]	486.7	500.1	1-2	0
KS	KS	2607.1	2607.1	10	1
KS	LD	415.9	416.5	4	2
KS	post-MAT linker	1460.0	1465.8	12-13	3-4
LD	post-MAT linker	1262.5	1277.4	8	0-1
MAT	post-MAT linker	908.5	911.5	4	1

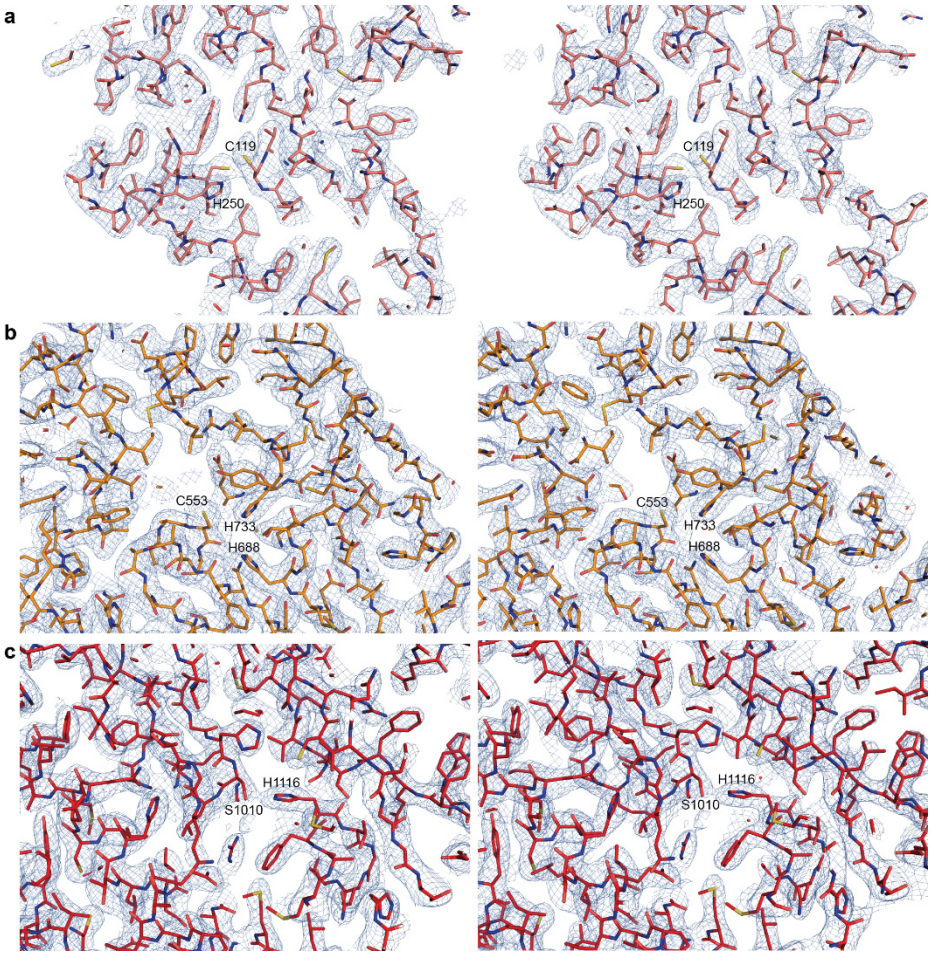
(a) C_α r.m.s. deviations obtained for structural comparison of crystallized CTB1 domains with their closest structural neighbors in the PDB and of multienzyme PKSs and FASs (to the best of our knowledge). For structures with several protomers only the best matches are reported. *,Q-score based, [†], *trans*-AT PKS. **(b)** Interfaces in the crystal structures of CTB1 SAT-KS-MAT. [§]Interface with the second protomer.

Supplementary Table 4.3 | Plasmids used in this study.

Plasmid	Protein	Vector	Tag	MW (g/mol)	ε ($M^{-1}cm^{-1}$)
pECTB1-NKA6	SAT-KS-MAT	pET-24a	*C*-His_6	140150	138200
pECTB1-SKM-C119A-S1010A	SAT°-KS-MAT°	pET-24a	*C*-His_6	140150	138200
p28CTB1-ACP1	ACP1	pET-28a	*N*-His_6	11160	N/A
p28CTB1-ACP2	ACP2	pET-28a	*N*-His_6	11970	5500
pECTB1-SKM-R461A	SAT-KS-MAT-R461A	pET-24a	*C*-His_6	140150	138200
pECTB1-SKM-R461E	SAT-KS-MAT-R461E	pET-24a	*C*-His_6	140150	138200
pECTB1-SKM-R658A	SAT-KS-MAT-R658A	pET-24a	*C*-His_6	140150	138200
pECTB1-SKM-R658E	SAT-KS-MAT-R658E	pET-24a	*C*-His_6	140150	138200
pECTB1-SKM-R879A	SAT-KS-MAT-R879A	pET-24a	*C*-His_6	140150	138200
pECTB1-SKM-R879E	SAT-KS-MAT-R879E	pET-24a	*C*-His_6	140150	138200
pECTB1-PT	PT	pET-24a	*C*-His_6	41240	26150
pECTB1-TE	TE	pET-28a	*N*-His_6	33630	38055

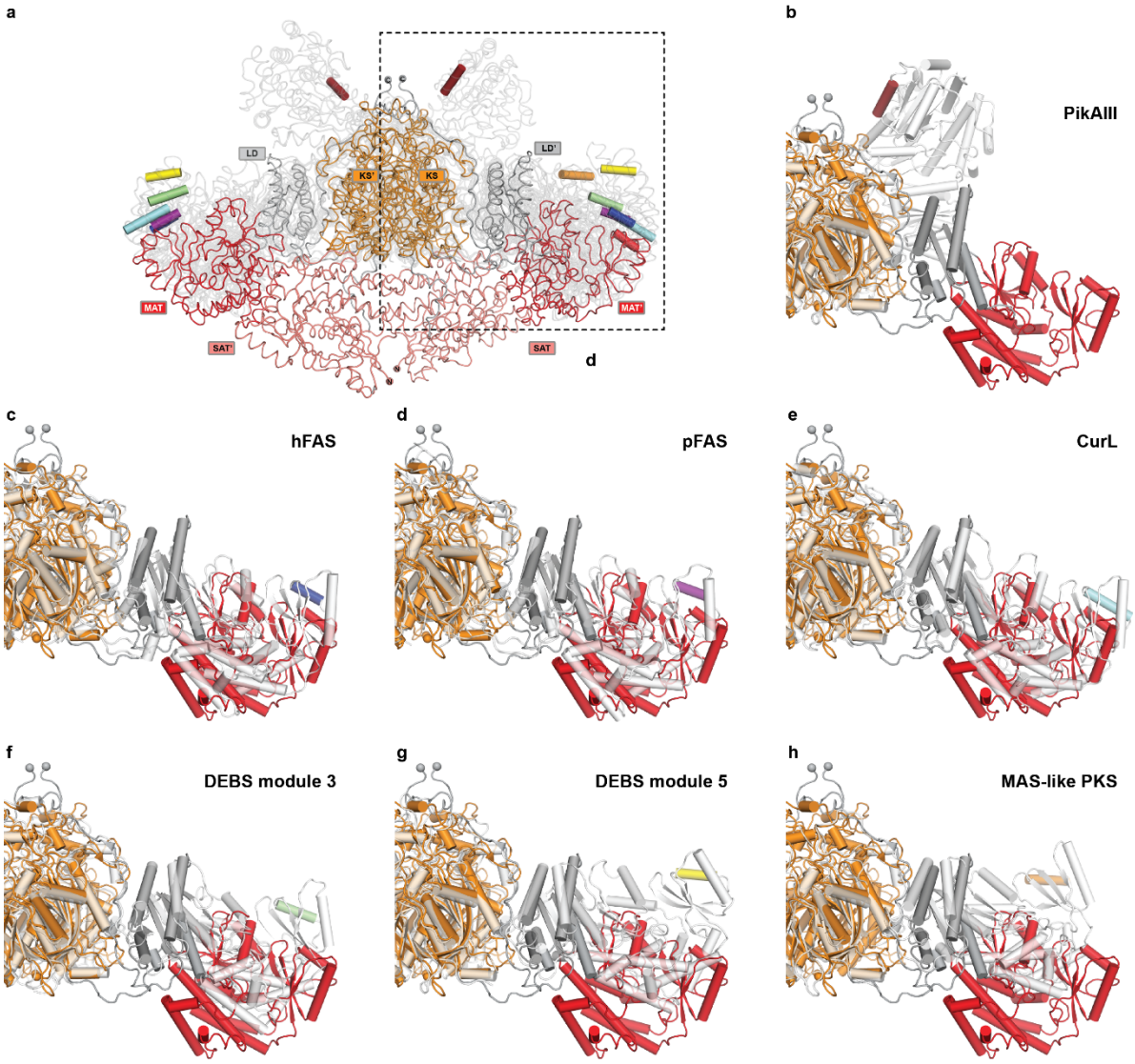
Supplementary Table 4.4 | Primers used in this study.

Primer	Sequence 5'-3'
CTB1-ACP1-5-A	ATTACATATGGCAACCCAAGTGACTCCGCAA
CTB1-ACP1-3-C	ATTAGCGGCCGCGATCTCGTTAGGGGATGGATCAGT
CTB1-ACP2-5	ATTACATATGGATCCATCCCCTAACGAGAT
CTB1-ACP2-stop-3	TAATGCGGCCGCCTATTCCGTTGACCCAGAGAACC
T7	TAATACGACTCACTATAGGG
T7-term	GCTAGTTATTGCTCAGCGG
CTB1-C119A-5	TGCATTACCGGCGTTGCAACCGGCGCA
CTB1-C119A-3	CGTCAATGCGCCGGTTGCAACGCCGGT
CTB1-S1010A-5	GGTAGTTGGCCACGCGTTGGGCGAGTATGC
CTB1-S1010A-3	GCATACTCGCCCAACGCGTGGCCAACTACC
CTB1-R461A-5	ACATGAGTCCGGCGGAAGCGCCGC
CTB1-R461A-3	GCGGCGCTTCCGCCGGACTCATGT
CTB1-R461E-5	ACATGAGTCCGGAAGAAGCGCCGC
CTB1-R461E-3	GCGGCGCTTCTTCCGGACTCATGT
CTB1-R658A-5	GCCTCTATCACTGCGCCTCATGCCGGAG
CTB1-R658A-3	CTCCGGCATGAGGCGCAGTGATAGAGGC
CTB1-R658E-5	GCCTCTATCACTGAACCTCATGCCGGAG
CTB1-R658E-3	CTCCGGCATGAGGTTCAGTGATAGAGGC
CTB1-R879A-5	GAGTGCACCATGCGCACAGAGCCGTAGC
CTB1-R879A-3	GCTACGGCTCTGTGCGCATGGTGCACTC
CTB1-R879E-5	GAGTGCACCATGAACACAGAGCCGTAGC
CTB1-R879E-3	GCTACGGCTCTGTGTTCATGGTGCACTC



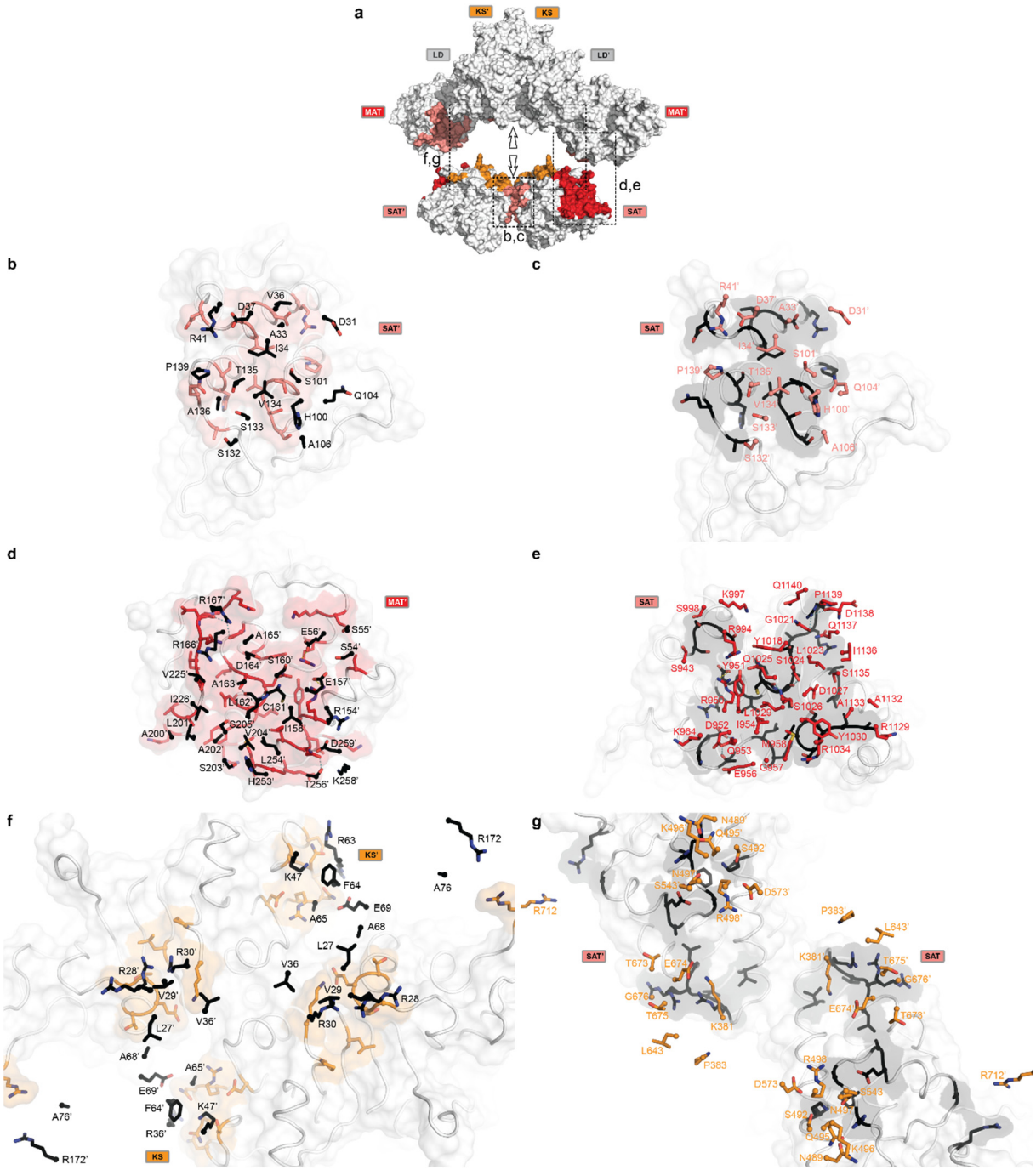
Supplementary Figure 4.5 | Stereo electron density quality assessment maps.

2Fo-Fc electron density maps at 1.0 σ of the SAT **(a)**, KS **(b)**, and MAT **(c)** active site regions are shown.

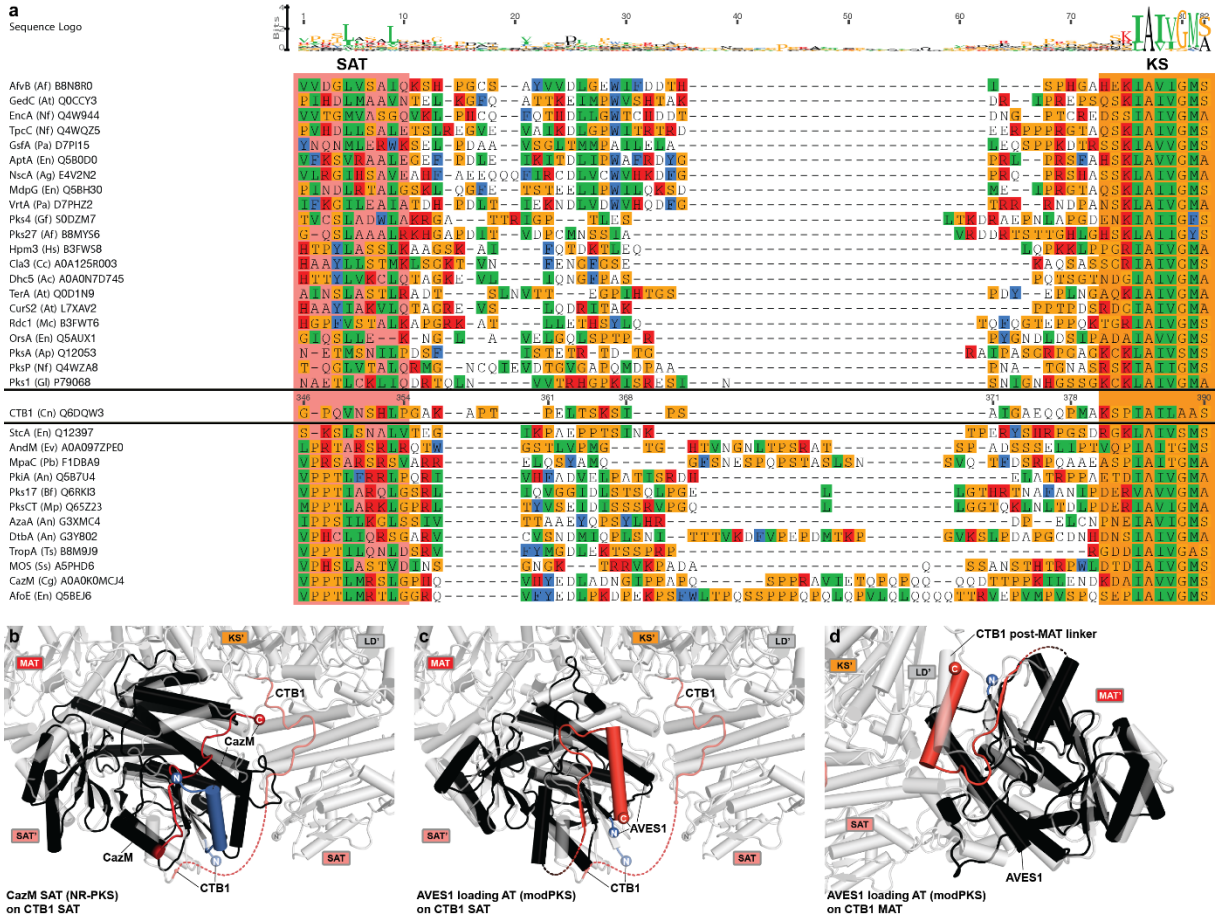


Supplementary Figure 4.6 | Comparison of PKS and FAS condensing region structures.

(a) Overview of all superposed condensing region structures on the CTB1 KS domain. Structural differences are visualized by hinge-bending motions of the (M)ATs as described by Herbst *et al.* 2016 (Nature 531, 533-537). Colored loop representation is shown for CTB1, all other structures are shown grey transparent. Equivalent cylindrical helix of the (M)AT domains are shown to indicate structural differences relative to CTB1 (light red) for PikAIII (dark red, **b**), human FAS (hFAS, dark blue, PDB:3hhd, **c**), porcine FAS (pFAS, PDB:2vz9, violet, **d**), CurL (cyan, PDB:4mz0, **e**), DEBS module 3 (green, PDB:2qo3, **f**), DEBS module 5 (yellow, PDB:2hg4, **g**) and a MAS-like PKS (orange, PDB:5bp1, **h**). **(b-h)** Inset of **(a)** in cartoon representation showing the individual superpositions. The model corresponding to PikAIII **(b)** was fitted according to Dutta *et al.* 2014 (Nature 510, 512-517) (EMDB:5649).

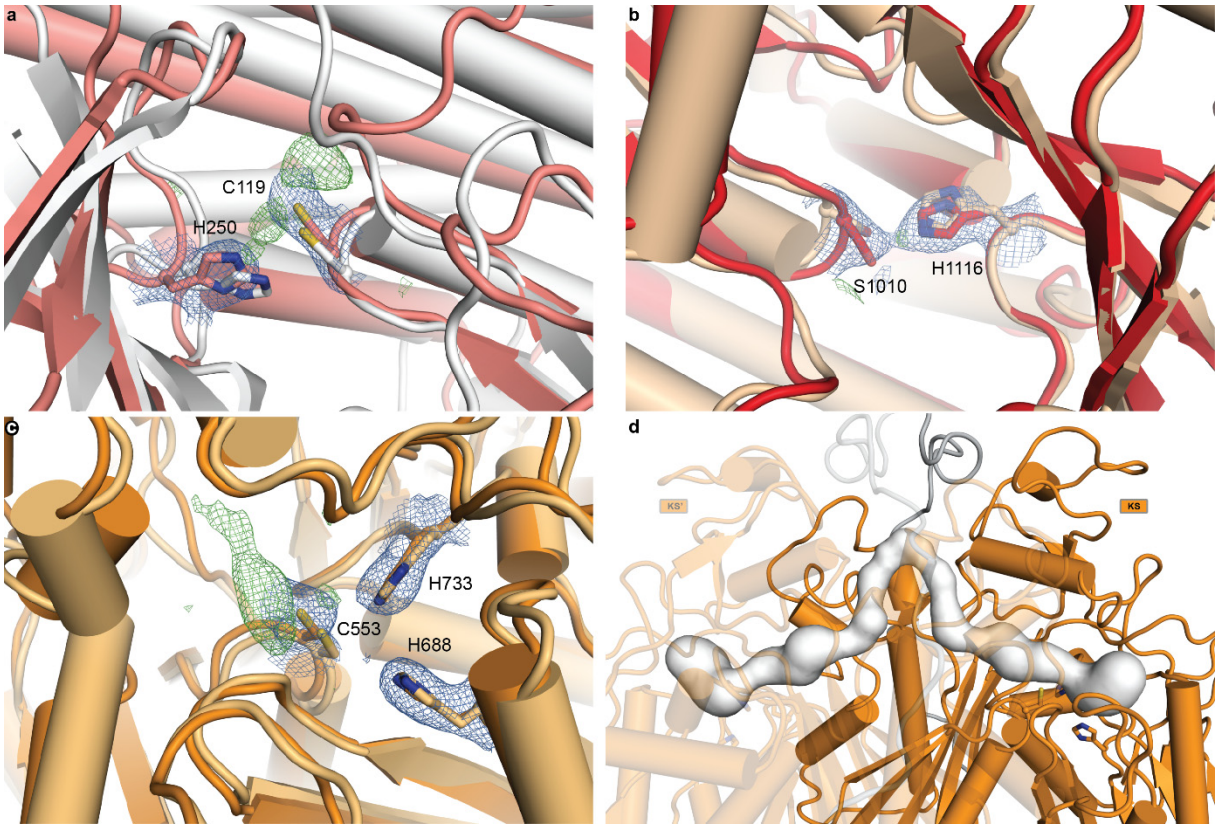


Supplementary Figure 4.7 | SAT interfaces in the loading/condensing region of CTB1.
(a) Overview of interfaces with insets showing the location of the SAT/SAT **(b,c)**, SAT/MAT **(d,e)**, and SAT/KS **(f,g)** interfaces. All interfaces are shown in a top view. Transparent surfaces are colored according to interfacing residues of the domain in the back. All interfaces are shown from both sides corresponding to a 180° rotation between the left and right panel. Residue numbers correspond to interfacing residues in the front shown in the same color. C_α positions are shown as spheres. Salt bridges (yellow) and hydrogen bonds (cyan) are indicated as dotted lines.



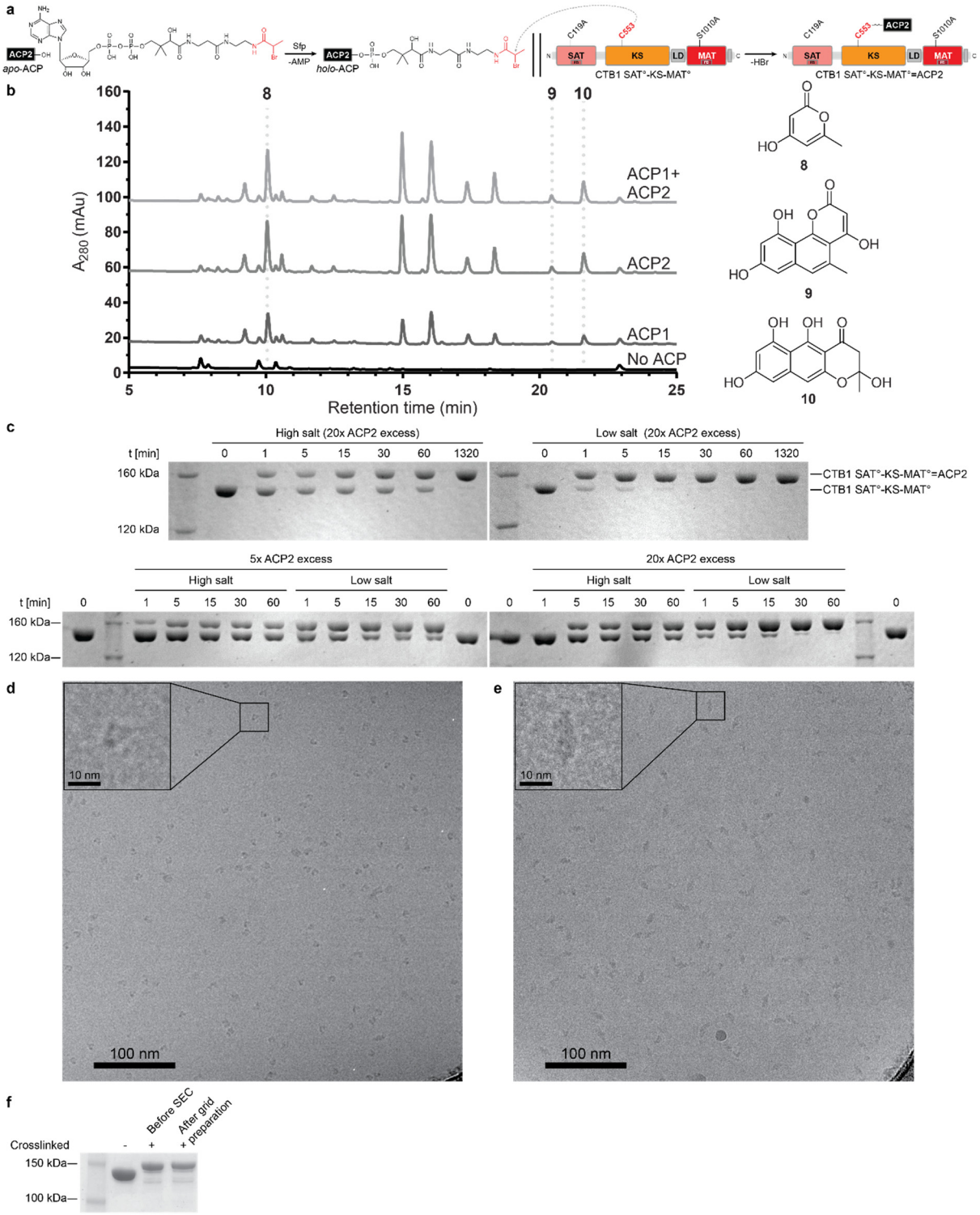
Supplementary Figure 4.8 | Integration of loading domains in PKS.

(a), Sequence alignment of 34 SAT-KS linker sequences of NR-PKS. All sequences are labelled as "protein name (organism abbreviation) Uniprot number". **(b)** Superposition of the CazM SAT (black) onto the SAT domain of CTB1 SAT-KS-MAT (light grey) shows a similar location of the SAT *C*-terminal linker ending (dark red/light red for CazM/CTB1), which points towards the *N*-terminus of the KS in CTB1. **(c)** Superposition of the loading domain of the AVES1 avermectin modPKS (black) onto the SAT domain of CTB1 SAT-KS-MAT reveals differences in C-terminal linker organization. **(d)** Superposition of the AVES1 loading domain onto CTB1 MAT shows that the AVES1 post-loading domain linker contains an α-helix that matches the linker architecture of MAT in CTB1 (and AT domains in modPKS).



Supplementary Figure 4.9 | The individual active sites of SAT, MAT and KS are structurally conserved.

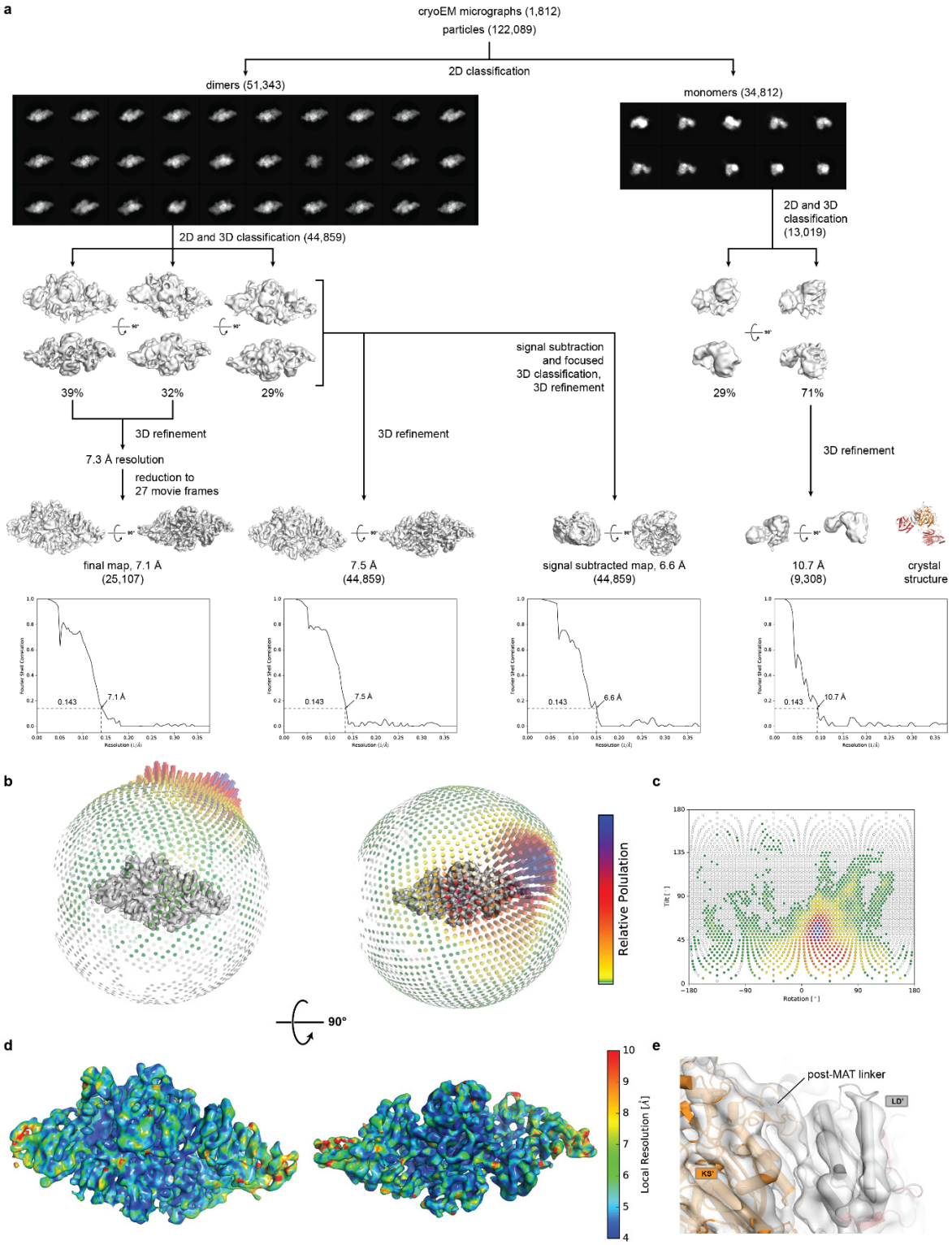
(a-c) Superpositions onto previously known structures (as listed in parenthesis) of the SAT (**a**, pink, with CazM SAT in white), MAT (**b**, red, with CurL AT in wheat), and KS (**c**, orange, with CurL KS in light orange) domains are shown in cartoon representation. Active site residues are shown in stick representation with residue numbers corresponding to CTB1. Electron density maps for CTB1 active site residues are shown with a blue ($2F_o$-F_c at 1.0 σ) and green (F_o-F_c at 3.0 σ) mesh. Unidentified difference density around the active site residues C119 (SAT) and C553 (KS) indicated partial acylation or oxidation. **(d)** A narrow and hydrophobic active site tunnel (white) connects the KS active sites.



Supplementary Figure 4.10 | ACP activity, crosslinking and cryo-EM sample preparation.

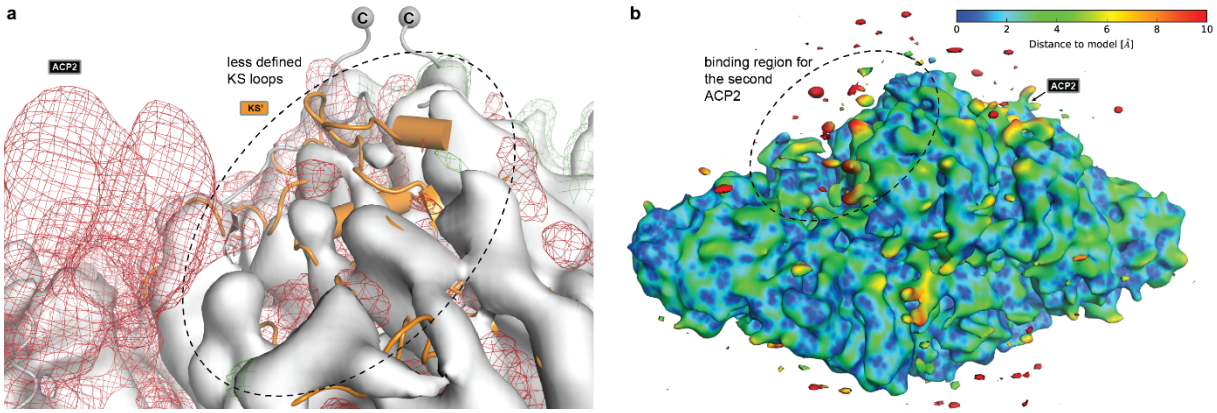
(a) Scheme for ACP2 loading with a dephospho-amino-CoA analog (black), containing the α-bromopropionyl crosslinker (red), and site specific crosslinking to the KS domain of CTB1 SAT°-KS-MAT°. **(b)** HPLC chromatograms (280 nm) of minimal *in vitro* CTB1 reconstitution reactions with SAT-KS-MAT and ACP1 (10 µM), ACP2 (10 µM), or both (5 µM each). Curves are off-set for visualization. Previously characterized CTB1 derailment products 8, 9, and 10, and negative controls lacking ACP are indicated. No increase in activity is observed when ACP1 and ACP2 are combined, suggesting they are biosynthetically equivalent and interchangeable. These experiments were repeated twice with similar results. **(c)** Denaturing SDS-PAGE analysis with Coomassie staining of crosslinking time course in high (50 mM Tris pH 7.5, 250 mM NaCl) and low (50 mM Tris pH 7.5) salt buffer as well as for two different CTB1 SAT°-KS-MAT°:ACP2 ratios. Crosslinking proceeds faster in low salt buffer and at higher excess of ACP2. Gels are cropped. These experiments were repeated more than

three times from independent protein preparations with similar results. **(d,e)** Drift corrected and dose filtered cryo-EM micrographs reveal clearly recognizable shapes for monomeric **(d)** and dimeric **(e)** particles. More than 10 grids have been prepared independently from more than three protein preparations with similar results. **(f)** Denaturating SDS-PAGE analysis with Coomassie staining of the sample used for cryo-EM grid preparation and data collection before SEC and after grid preparation.



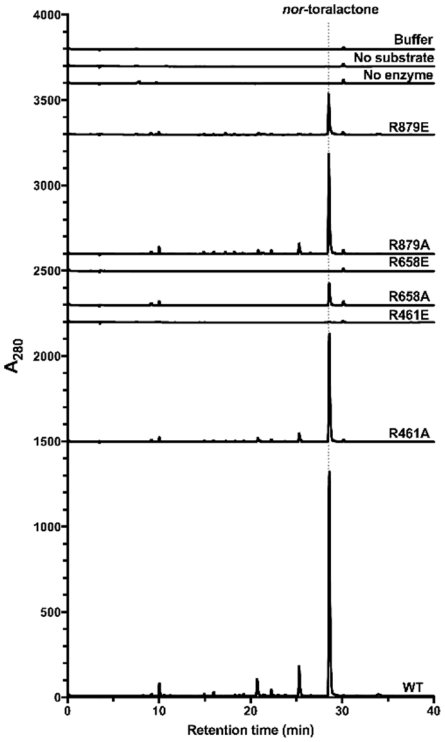
Supplementary Figure 4.11 | Cryo-EM data processing scheme.

(a) 2D- and 3D-classification and sorting scheme for reconstructions of CTB1 SAT°-KS-MAT°=ACP2 dimers as well as monomers. 3D-class distributions are indicated below the models. FSC plots (corrected for effects of the mask) used for determining resolution based on the 0.143 criterion (Rosenthal, P. B. & Henderson, R. J. Mol. Biol. 333, 721-745 (2003)) is shown below the final models. A single data set has been collected from one pre-screened protein and grid preparation. **(b, c)** 3D **(b)** and 2D **(c)** angular distribution plot based on the alignment used for the final reconstruction. **(d)** Local resolution map of the final reconstruction at 7.1 Å resolution shows a resolution below 5 Å in the central KS region, decreasing towards the lateral MAT domains to around 10 Å. **(e)** The final map reveals features of secondary structure elements as well as linkers. All maps are contoured at 7.1σ.



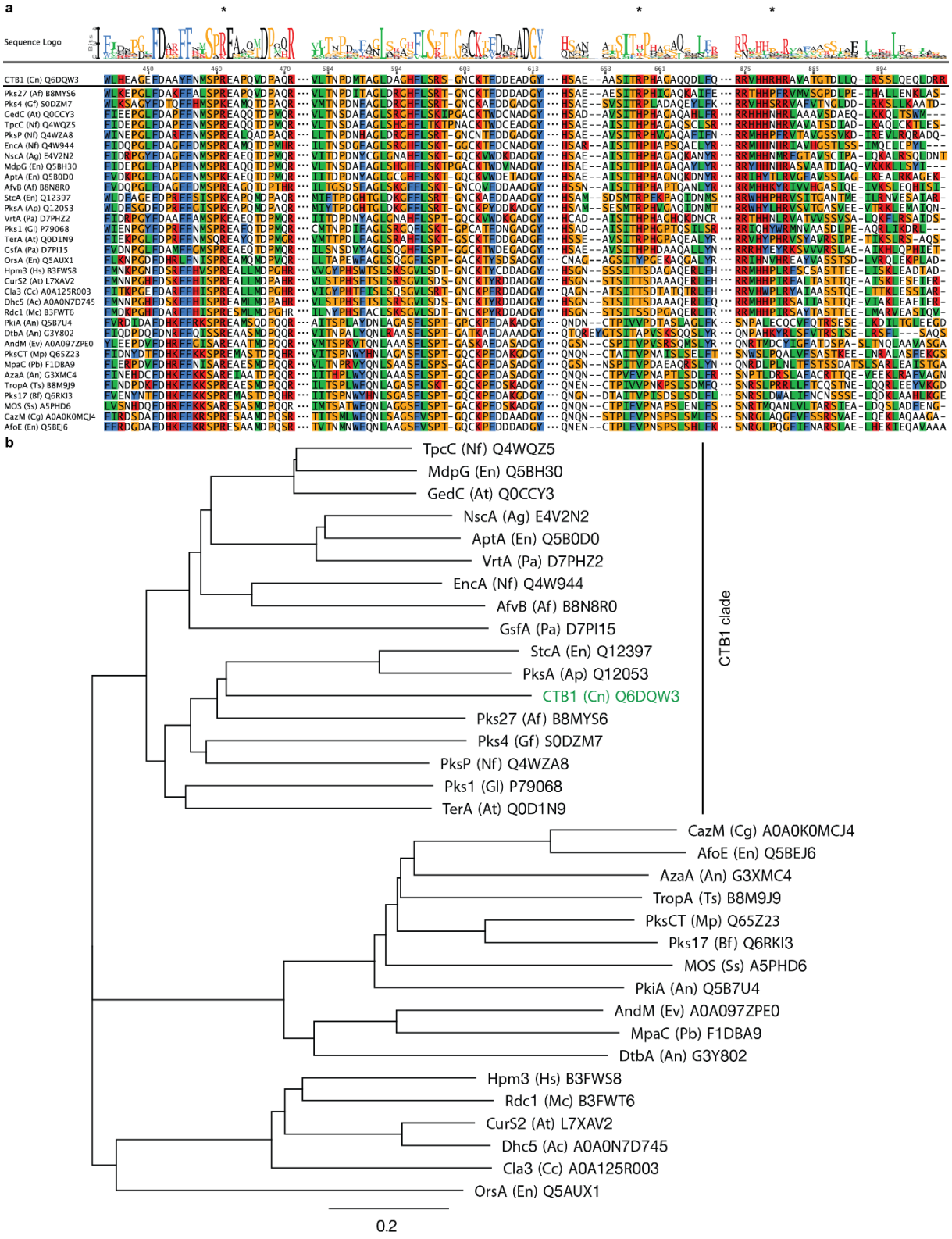
Supplementary Figure 4.12 | Conformational variability in the cryo-EM structure of CTB1 SAT^0-KS-MAT^0=ACP2.

(a, b) KS-based C2 symmetry differences map of the final reconstruction at 7.1 Å resolution, contoured at 7.1σ (green: positive, red: negative). The potential ACP2 binding region on the KS surface is less ordered on the side lacking ACP2 density. **(b)** An unmasked map of the final reconstruction contoured at 3σ and colored by distance to the atomic model shows enhanced noise around the second ACP binding site.



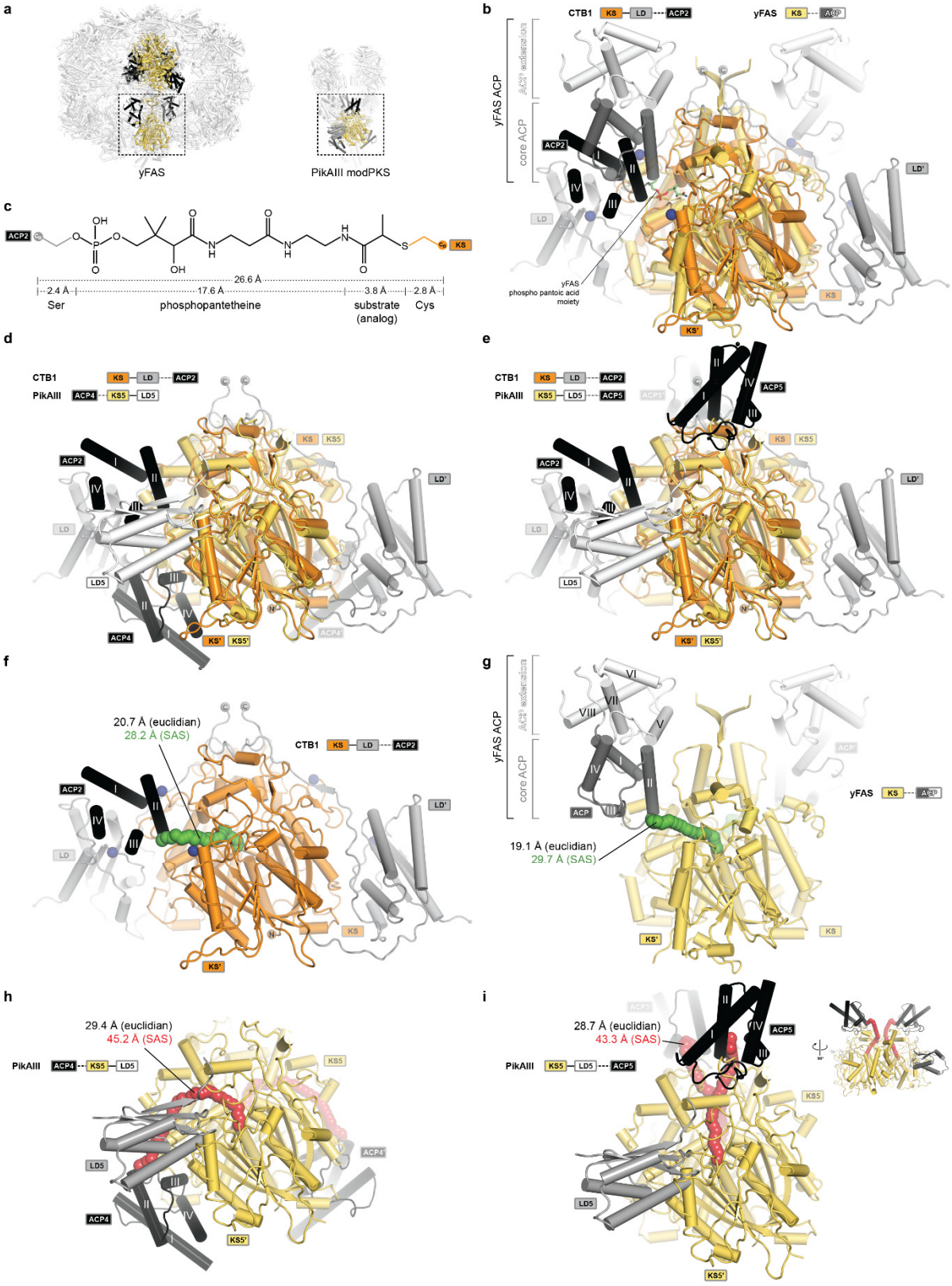
Supplementary Figure 4.13 | CTB1 mutant activity in deconstruction experiments and ACP2 interface validation.

Representative HPLC chromatograms (280 nm) of full *in vitro* CTB1 reconstitution reactions (SAT-KS-MAT + ACP2 + PT + TE) with SAT-KS-MAT interface mutants are shown. Curves are off-set for visualization. *Nor*-toralactone production is indicated by a peak at a retention time of 28.6 min. These experiments were repeated three times with similar results.



Supplementary Figure 4.14 | CTB1 SAT-KS-MAT alignment of regions interfacing ACP2 and phylogeny.

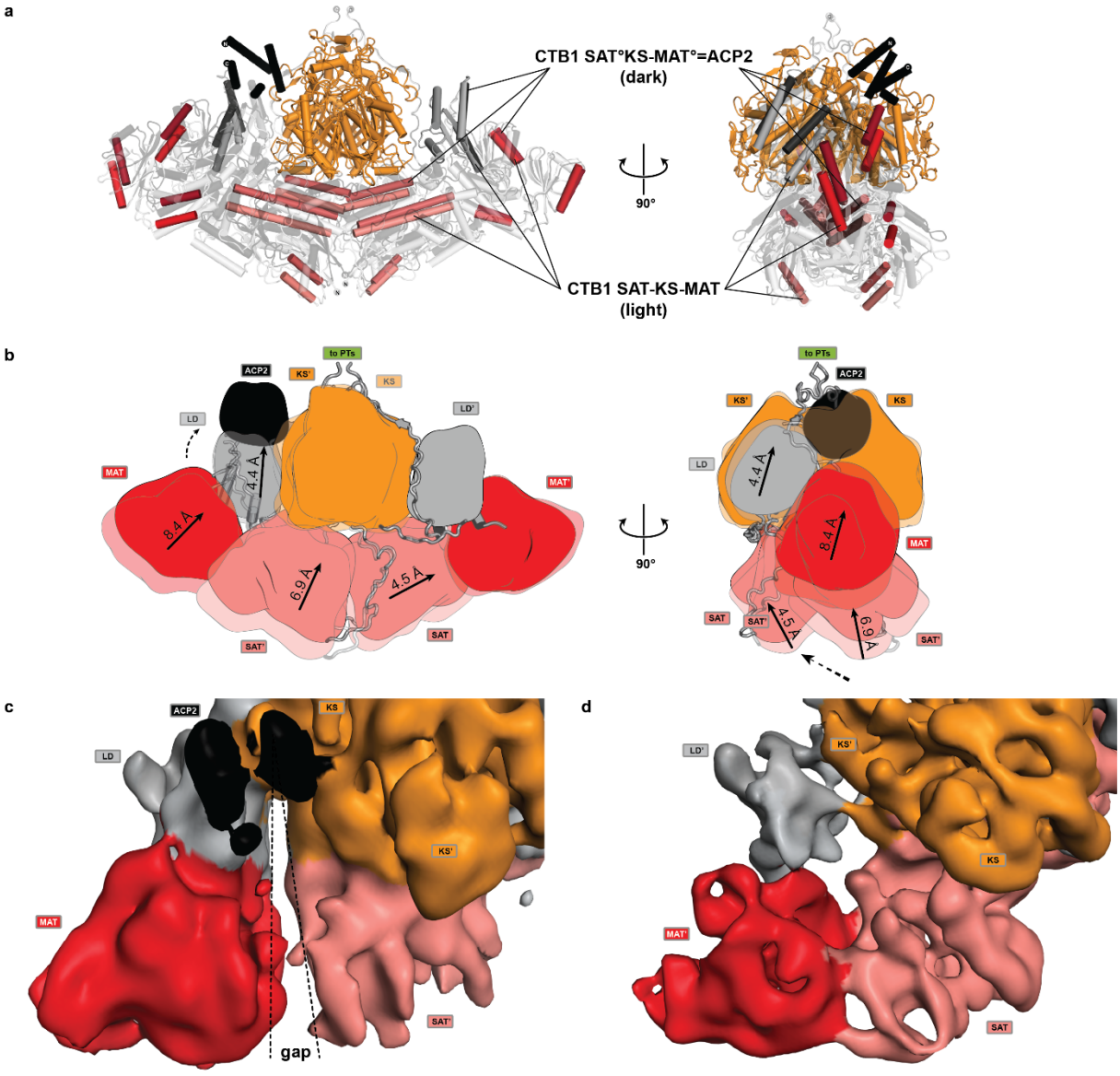
All sequences are labelled as "protein name (organism abbreviation) Uniprot number". **(a)** Sequence alignment of regions relevant for ACP2 interaction. *, indicates interface mutants that reduce activity. **(b)** Phylogenetic analysis of NR-PKS indicates three main clades with one clustering around CTB1. Distance unit is given as accepted amino-acid substitutions per site and indicated by scale bar.



Supplementary Figure 4.15 | Comparison of KS-ACP interactions in PKS and FAS.

(a) ACP interactions with multienzyme KS domains have been observed in the hetero-dodecameric yFAS and PikAIII modPKS. Superposition of the architecturally distinct yFAS-KS with CTB1 KS-LD°=ACP2. **(b)** The divergent ACP of yFAS consists of a *C*-terminal four-helix bundle extension (white) and the conserved four-helix bundle at the N-terminus, which interacts with the KS in a similar position as observed for ACP2 in CTB1. ACP-interfacing residues in CTB1 are indicated by blue spheres (see Figure 4.3) **(c)** Derivation of the length of a fully extended

crosslink between ACP2 and KS in CTB1. The defined length of the Ppant cofactor bridges a distance of approx. 27 Å between the ACP-Ser and the KS-Cys C_α atoms for functional ACP interaction with active sites. **(d, e)** Superpositions of PikAIII with ACP4 **(d)** and ACP5 **(e)** on CTB1 KS-LD=ACP2 show the relative locations of the ACP binding interfaces. The LD5 arrangement in PikAIII is inverse to CTB1. **(f-i)** C_α-C_α solvent accessible surface (SAS) distances between the ACP anchor and the KS active site cysteine in CTB1 **(f)**, yFAS **(g)**, and PikAIII with ACP4 **(h)** and ACP5 **(i)** are indicated. Solvent accessible surface paths are indicated by tubes of spheres and colored according to the agreement (green) or disagreement (red) with the length of the Ppant cofactor. The path for ACP5 to KS5 in PikAIII **(i)** is only accessible in the absence of the post-(M)AT linker, which was not resolved in the cryo-EM structure. An additional miniaturized-sideview is indicated. Corresponding euclidian distances are given in the panels. Domain colors are indicated in the panels; only KS, LD, and ACP domains are shown **(b-i)**.



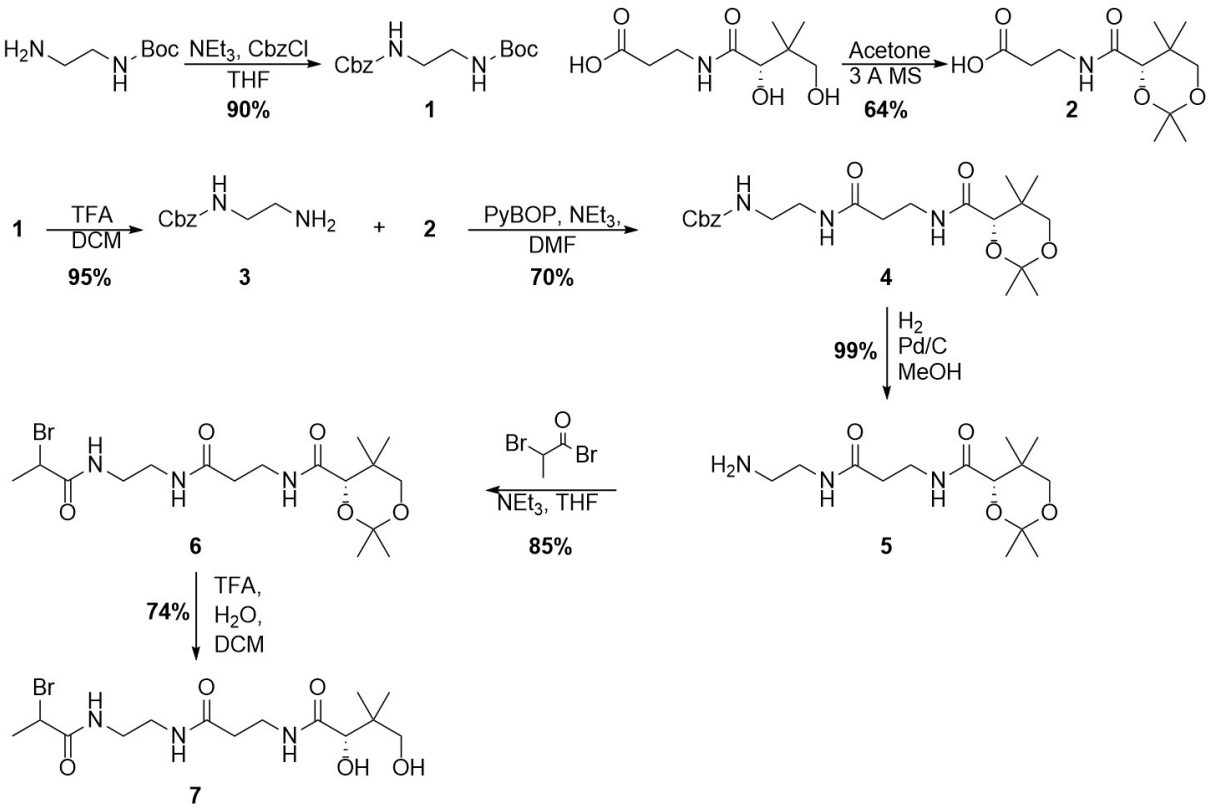
Supplementary Figure 4.16 | Structural comparison between CTB1 SAT-KS-MAT and SAT°-KS-AT°=ACP2.

(a) Superposition on the central KS dimer. Transparent cartoon representation of CTB1 is shown in white (SAT-KS-MAT) and black (SAT°-KS-MAT°=ACP2). Selected helices are shown in light (SAT-KS-MAT) and dark (SAT°-KS-MAT°=ACP2) domain colors. **(b)** Schematic representation of relative domain motions. Distances are derived by domain-wise C_α-r.m.s.d. calculations. **(c, d)** Structural differences at the MAT-SAT interface in the cryo-EM reconstruction are shown for the unsharpened map. While the side with visualized ACP2 **(c)** reveals a gap, the side with unresolved ACP2 shows a connected interface **(d)**.

Supplementary Movie 4.1 | ACP2 binding to the CTB1 loading/condensing region.

Structures of the CTB1 loading/condensing region with and without crosslinked ACP2 are morphed. In the ACP2-bound structure, ACP is visualized at a single binding-site in an overall structure with increased asymmetry. ACP2 is located in a cleft between the KS dimer and one LD. Around ACP2, LD and MAT are moved upwards, and the interface between the neighboring MAT and SAT domains is loosened. The SAT is translocated towards the KS and the second SAT is pushed against the KS and MAT in the opposite half of the dimer. (The supplementary movie is available online: www.nature.com/nchembio)

Supplementary Note 4.1 | Synthetic procedure for the synthesis of α-bromopropionyl-aminopantetheine (7)



Supplementary Figure 4.17 | ACP to SAT-KS-MAT crosslinker synthesis scheme

All reagents were purchased from Sigma-Aldrich (St. Louis, MO) unless otherwise indicated.

Benzyl *tert*-butyl ethane-1,2-diyldicarbamate (1): In a 500 mL round-bottomed flask, *N*-Boc-ethylenediamine (14 mL, 88.5 mmol) was stirred in anhyd. DCM (200 mL) under argon. Triethylamine (16 mL, 113.9 mmol) was added, followed by slow addition of Cbz-Cl (15 mL, 105 mmol). DMAP (1.1 g, 8.9 mmol) was added, and the solution was stirred at room temperature overnight. The reaction was quenched by the addition of a saturated aqueous solution of NH_4Cl (100 mL). The aqueous layer was washed with EtOAc (3×50 mL). The combined organic extracts were dried over anhyd. Na_2SO_4. Crystallization proceeded spontaneously upon concentration on a rotary evaporator and the resulting solid was filtered on a glass frit to give *N*-Boc,*N*'-Cbz-ethylenediamine **1** (23.1 g, 79.7 mmol, 90%) as a

white solid. **^{1}H NMR** (400 MHz, $CDCl_3$): δ 7.35-7.33 p.p.m. (m, 5H), 5.09 (s, 2H), 3.29-3.24 (m, 4H), 1.42 (s, 9H). **^{13}C{^{1}H} NMR** (101 MHz, $CDCl_3$): δ 136.41, 128.44, 128.04, 128.01, 79.46, 66.65, 45.71, 41.38, 40.52, 28.29, 8.51. **UPLC-HRMS** (ESI) calc'd for $C_{15}H_{22}N_2O_4Na^+$ 317.1472; found 317.1471 $[M+Na]^+$.

(*S*)-3-(2,2,5,5-tetramethyl-1,3-dioxane-4-carboxamido)propanoic acid (2): *D*-pantothenic acid hemicalcium salt (11 g, 46.2 mmol) and *p*-toluene sulfonic acid hydrate (17 g, 100 mmol) were added to acetone (200 mL) in a flame-dried, round-bottomed flask under argon. Molecular sieves (4 Å, 200 g) were added, and the reaction mixture was stirred vigorously enough to break up the sieves. The reaction was run overnight at room temperature. The sieves were filtered off through a bed of Celite. The solution was concentrated by rotary evaporation and the resulting syrup was redissolved in EtOAc (100 mL). The organic solution was washed with brine (2×50 mL) and dried with anhyd. Na_2SO_4. The solution was concentrated partially *in vacuo*, and hexanes were added. The ketal-protected pantothenic acid **2** (7.64 g, 29.4, 64%) precipitated as a white solid upon concentration to dryness. **^{1}H NMR** (400 MHz, $CDCl_3$): δ 7.06 p.p.m. (t, *J* = 5.6 Hz, 1H), 4.10 (s, 1H), 3.67 (d, *J* = 11.6 Hz, A of AB_q, 1H), 3.61-3.56 (m, 1H), 3.50-3.44 (m, 1H), 3.27 (d, *J* = 11.6 Hz, B of AB_q, 1H), 2.59 (t, *J* = 6.0 Hz, 2H), 1.39 (s, 3H), 1.37 (s, 3H), 0.96 (s, 3H), 0.91 (s, 3H). **^{13}C{^{1}H} NMR** (101 MHz, $CDCl_3$): δ 176.32, 170.22, 99.04, 77.00, 71.35, 34.08, 33.77, 32.92, 29.34, 21.96, 18.78, 18.66. **UPLC-HRMS** (ESI) calc'd for $C_{12}H_{21}NO_5Na^+$ 282.1312; found 282.1312 $[M+Na]^+$.

Benzyl (2-aminoethyl)carbamate (3): To a 250 mL round-bottomed flask containing DCM:TFA (30 mL:15 mL) was added **1** (10 g, 34 mmol). The solution was stirred for 2 h at room temperature. The reaction mixture was diluted with toluene, and concentrated by rotary evaporation three times to give Cbz-ethylenediamine (6.27 g, 32.3 mmol, 95%) as a yellow oil, which was taken directly to the next reaction. The observed spectral data matched those data reported previously[294].

Benzyl (*S*)-(2-(3-(2,2,5,5-tetramethyl-1,3-dioxane-4-carboxamido)propanamido) ethyl) carbamate (4): In a flame-dried, round-bottomed flask, **2** (8 g, 30.6 mmol), **3** (6.6 g, 34 mmol) and triethylamine (9.48 mL, 68 mmol) were dissolved in DCM (300 mL). The solution was cooled to 0 ℃ in an ice bath. PyBOP (19.45 g, 37.4 mmol) was added in one portion. The reaction mixture was stirred for 3 h, and allowed to warm to room temperature. The reaction was quenched by the addition of a saturated aqueous solution of NH_4Cl (150 mL). The aqueous layer was extracted with EtOAc (2×100 mL) and the combined organic extracts were dried over anhyd. Na_2SO_4, followed by concentration by rotary evaporation. A white solid precipitated from the solution and was filtered off. The remaining oil was fractionated by flash silica chromatography (100% EtOAc) to afford pure, diprotected Cbz-aminopantetheine **4** (9.32 g, 21.4 mmol, 70%) as a colorless solid, which was recrystallized from EtOAc/Hexanes. **^{1}H NMR** (400 MHz, $CDCl_3$): δ 7.25 p.p.m. (m, 5H), 7.05-7.02 (m, 2H), 5.92 (s, 1H), 5.00 (s, 2H), 3.98 (s, 1H), 3.57 (d, *J* = 11.7 Hz, A of AB_q, 1H), 3.42 (septet, *J* = 6.6 Hz 2H), 3.30-3.21 (m, 4H), 3.17 (d, *J* = 11.7 Hz, B of AB_q, 1H), 2.33 (t, *J* = 6.4 Hz, 2H), 1.38 (s, 3H), 1.32 (s, 3H), 0.92 (s, 3H), 0.88 (s, 3H). **^{13}C{^{1}H} NMR** (101 MHz, $CDCl_3$): δ 171.69, 170.39, 157.04, 136.37, 128.49, 128.14, 128.10, 99.09, 77.09, 71.32, 66.81, 40.81, 40.13, 36.12, 34.79, 32.90, 29.40, 22.05, 18.81, 18.64. **UPLC-HRMS** (ESI) calc'd for $C_{22}H_{34}N_3O_6^+$ 436.2442; found 436.2443 $[M+H]^+$.

(*S*)-*N*-(3-((2-aminoethyl)amino)-3-oxopropyl)-2,2,5,5-tetramethyl-1,3-dioxane-4-carboxamide (5): Compound **4** (7 g, 18.4 mmol) was added to MeOH (10 mL) in a 100 mL Parr bomb flask. The reaction mixture and flask were purged with argon. 10% Pd/C (700 mg) was added. The reaction vessel was placed under a H_2 atmosphere (50 bar), and shaken for 3 h. The reaction mixture was filtered over Celite to remove the catalyst and the Celite washed with MeOH (3×10 mL). The filtrate was

concentrated by rotary evaporation to afford ketal-protected aminopantetheine **5** (5.5 g, 18.3 mmol, 99%) as a pale yellow oil, which was taken directly to the next step. **^{1}H NMR** (400 MHz, $CDCl_3$): δ 7.07 p.p.m. (t, *J* = 8.0 Hz, 1H), 6.88 (s, 1H), 4.04 (s, 1H), 3.65 (d, *J* = 15.6 Hz, A of AB_q, 1H), 3.52-3.46 (m, 3H), 3.42 (s, 1H), 3.38-3.37 (m, 2H), 3.24 (d, *J* = 15.6 Hz, B of AB_q, 1H), 2.86 (t, *J* = 7.5 Hz, 2H), 2.42 (t, *J* = 8.7 Hz, 2H), 1.44 (s, 3H), 1.40 (s, 3H), 0.98 (s, 3H), 0.93 (s, 3H). **^{13}C{^{1}H} NMR** (101 MHz, $CDCl_3$): δ 171.51, 170.05, 98.90, 76.96, 71.19, 41.54, 41.02, 35.81, 34.86, 32.75 29.26, 21.95, 18.70, 18.51. **UPLC-HRMS** (ESI) calc'd for $C_{14}H_{28}N_3O_4^+$ 302.2074; found 302.2079 $[M+H]^+$.

(4*S*)-*N*-(3-((2-(2-bromopropanamido)ethyl)amino)-3-oxopropyl)-2,2,5,5-tetramethyl-1,3-dioxane-4-carboxamide (6): In a round-bottomed flask under argon, **5** (1.77 g, 5.87 mmol) was added to DCM (20 mL). To this solution was added triethylamine (1.82 mL, 12.9 mmol). The temperature of the solution was lowered to 0 ℃ in an ice bath. Bromopropionyl bromide (0.645 mL, 6.16 mmol) was added dropwise. The initially milky solution was stirred at 0 ℃ for 1 h and became clear. The reaction was quenched by the addition of a saturated aqueous solution of NH_4Cl (30 mL), and the organic layer was washed with saturated aqueous Na_2CO_3 (15 mL) and brine (15 mL). The organic layer was concentrated by rotary evaporation. The resulting off-white solid was purified by flash silica chromatography (1:9 MeOH:EtOAc) to give 2-bromopropionyl aminopantetheine **6** (2.13 g, 4.87 mmol, 83%) as a colorless solid. The isolated mixture of diastereomers gave the following analytical data: **^{1}H NMR** (400 MHz, $CDCl_3$): δ 7.11 p.p.m. (s, 1H), 7.02 (t, *J* = 5.6 Hz, 1H), 6.57 (br s, 1H), 4.39/4.38 (2×q, *J* = 7.0 Hz, 1H), 4.09/4.07 (2×s, 1H), 3.68 (d, *J* = 12.0 Hz, A of AB_q, 1H), 3.59-3.52 (m, 2H), 3.43-3.38 (m, 4H), 3.28 (d, *J* = 12.0 Hz, B of AB_q, 1H), 2.46 (t, *J* = 6.2 Hz, 2H), 1.85/1.84 (2×d, *J* = 7.0 Hz, 3H), 1.46 (s, 3H), 1.42 (s, 3H), 1.02 (s, 3H), 0.97/0.96 (2×s, 3H). **^{13}C{^{1}H} NMR** (101 MHz, $CDCl_3$): δ 172.10/172.07, 170.42, 170.10/170.07, 98.90/98.90, 76.95, 71.12, 43.81/43.80, 40.12/40.05, 38.93, 35.81/35.75, 34.85, 32.74/32.73, 29.26, 22.31/22.82, 21.96/21.95, 18.72, 18.53. **UPLC-HRMS** (ESI) calc'd for $C_{17}H_{31}BrN_3O_5^+$ 436.1442; found 436.1433 $[M+H]^+$.

(2*S*)-*N*-(3-((2-(2-bromopropanamido)ethyl)amino)-3-oxopropyl)-2,4-dihydroxy-3,3-dimethylbutanamide (7): To DCM (2 mL) stirring in a 25 mL round-bottomed flask, **6** (25 mg, 0.06 mmol) was added. 4 mL of deionized water, and 0.1 mL of TFA were added to the reaction mixture. The solution was allowed to stir at room temperature for 60 min. The reaction mixture was diluted with 1:1 H_2O:toluene and concentrated by rotary evaporation three times to remove residual TFA. The resulting pale yellow oil was purified by reverse-phase (C_{18}) flash silica chromatography (gradient: 5-95% ACN in H_2O) to afford 2-bromopropionyl aminopantetheine **7** (15 mg, 1.2 mmol, 63%) as a white solid after lyophilization. **^{1}H NMR** (400 MHz, MeOD): δ 4.47 p.p.m. (q, *J* = 6.8 Hz, 1H), 3.91 (s, 1H), 3.50-3.47 (m, 2H), 3.48 (d, *J* = 10.8 Hz, A of AB_q, 1H), 3.40 (d, *J* = 10.8 Hz, B of AB_q, 1H), 3.36-3.32 (m, 4H), 2.43 (t, *J* = 6.8 Hz, 2H), 1.77 (d, *J* = 6.8 Hz, 3H), 0.93 (s, 6H). **^{13}C{^{1}H} NMR** (101 MHz, MeOD): δ 176.00, 174.17, 172.76, 77.30, 70.28, 43.97, 40.32, 40.30, 39.70, 36.58, 36.38, 22.41, 21.35, 20.88. **UPLC-HRMS** (ESI) calc'd for $C_{14}H_{27}BrN_3O_5^+$ 396.1129; found 396.1120 $[M+H]^+$.

5 DISCUSSION & OUTLOOK

5.1 Summary of results

The functions of PKS are directly coupled to their architecture as exemplified by modPKS, where the product is determined by the enzymatic composition of modules in an assembly line. The dynamic nature of these megaenzymes represents a challenge for structural biology. Common domain compositions and the observation that some domains or even entire modifying regions[48,91,295,296] can be functionally swapped between different PKSs, indicate a common architecture and a framework that is permissive for insertions, duplications, and deletions of domains and even entire modules. The structure of mFAS revealed that the functional condensing and modifying regions are also structurally distinct units and are connected via flexible linkers[162,165]. Structures of PKS condensing regions revealed virtually the same connectivity and indicated that PKS modules share the structural segregation first observed in FAS.

The work presented in chapter 2 revealed the architecture of an iterative fully reducing PKS and the relevance of this architecture also for modPKS. The observed organization of the modifying region is different from all previous suggested models, but in fact integrates features of different models. The condensing region is similar in mFAS and PKS. However, the modifying region of PKSs differ from mFAS by their linker-based architecture without stable domain interfaces. This linker-centric organization provides the framework for evolutionary domain shuffling and rationalizes the observed diversity of PKS modifying region domain compositions. The alternative model for the architecture of fully reducing modPKS based on outward-facing ER domains is not supported by current data. However, existence of such states cannot strictly be ruled out across all other PKS. Modes of conformational coupling were visualized in a dynamic modifying region architecture, that permits variations of active site distances with consequences for the access options of ACP to active sites. Conformational coupling might directly result in cross-talk between catalytic steps in the two lateral reaction clefts.

Chapter 3 provides the structural characterization of a PKS CMeT domain and its involvement in programming of fungal iterative PKS. In PksCT, methylation is used as a product elongation check point. Off-loading reactions are triggered, if the product has not been methylated in the first three cycles. The structure of CMeT revealed a new N-terminal fold and shares structural homology with the inactive mFAS CMeT domain, which confirms a common evolutionary origin. The active site showed a tunnel with an extended substrate binding cavity that is capable of accommodating substrates of different length. The structure itself did not provide direct evidence for programming NR-PKS substrate formation, but the interplay of KS and CMeT suggest that the KS domain specifically recognizes the methylation pattern that has been introduced by the CMeT. Parallel to this work a structure of a CMeT domain of a modPKS was published[68], which revealed virtually the same architecture and indicates a conservation of CMeT domains across different PKS families. *Cis*-AT PKS rarely contain CMeT domains, because they usually incorporate methyl branches via AT domains, as exemplified by the canonical MAS, which uses methyl-malonyl CoA as extender units.

The structure of a loading/condensing region of the NR-PKS CTB1, described in chapter 4, provided the first visualization of a specialized substrate loading architecture in PKS. It revealed the intertwined integration of SAT loading domains into NR-PKS condensing regions. Functional asymmetry of ACP binding was observed biochemically and also in the accompanying cryo-EM structure of a covalently

trapped state of initial substrate loading. The EM structure revealed asymmetry and conformational coupling extending all across the dimeric PKS, which indicated a V-twin engine like model, where one ACP binds to the KS, while the other one loads another extender unit.

5.2 Architectures of PKS modules

Various architectural models have been described for PKS, including the initial FAS-like, MAS-like, PikAIII-like and recently a CTB1-like architecture. All models except for the PikAIII-based one, suggest a common condensing region architecture.

The MAS-like architecture is similar to those of mFAS, but diverges in the linker based architecture (Figure 5.1). While evolutionary domain shuffling contributes to the diversity and product spectrum of PKS, similar events in mFAS are potentially lethal, because a cell cannot survive without fatty acids and the relative evolutionary pressure on maintaining the integrity of the fully reducing modifying region must be high. Consequently, mFAS does not require a linker based architecture and might have lost it during evolution. In PKS, the fully reducing modifying region is not the most common one. Fully reduced polyketides are chemically inert fatty acids, which suited for isolating cellular compartments, as energy storage in animal cells, and as carbon storage in plants, but not for secondary chemical reactions or adapted protein binding as chemical agent. However, the fully reducing module is still the most interesting one (Figure 5.2a, b), because the shared linker connectivity between mFAS and MAS-like PKS shows that all modules might have evolved from a common fully reducing ancestor that also included a CMeT domain, which is still present as inactive remnant in mFAS (ΨCMeT, Figure 5.1b, d). Therefore, the fully reducing architecture serves as prototype, which contains a functional framework that is also used by smaller modules. Even the only larger module, containing a CMeT domain, can be derived from the fully reducing MAS-like architecture: All ΨKR/KR domains have a conserved N-terminal linker that wraps around the ΨKR/KR surface and end in a short unstructured loop N-terminally to the ΨKR. This loop is present in all modules with ΨKR/KR and can accommodate a CMeT domain (Figure 5.1c, Figure 5.2), which is exemplified by mFAS(Figure 5.1d).

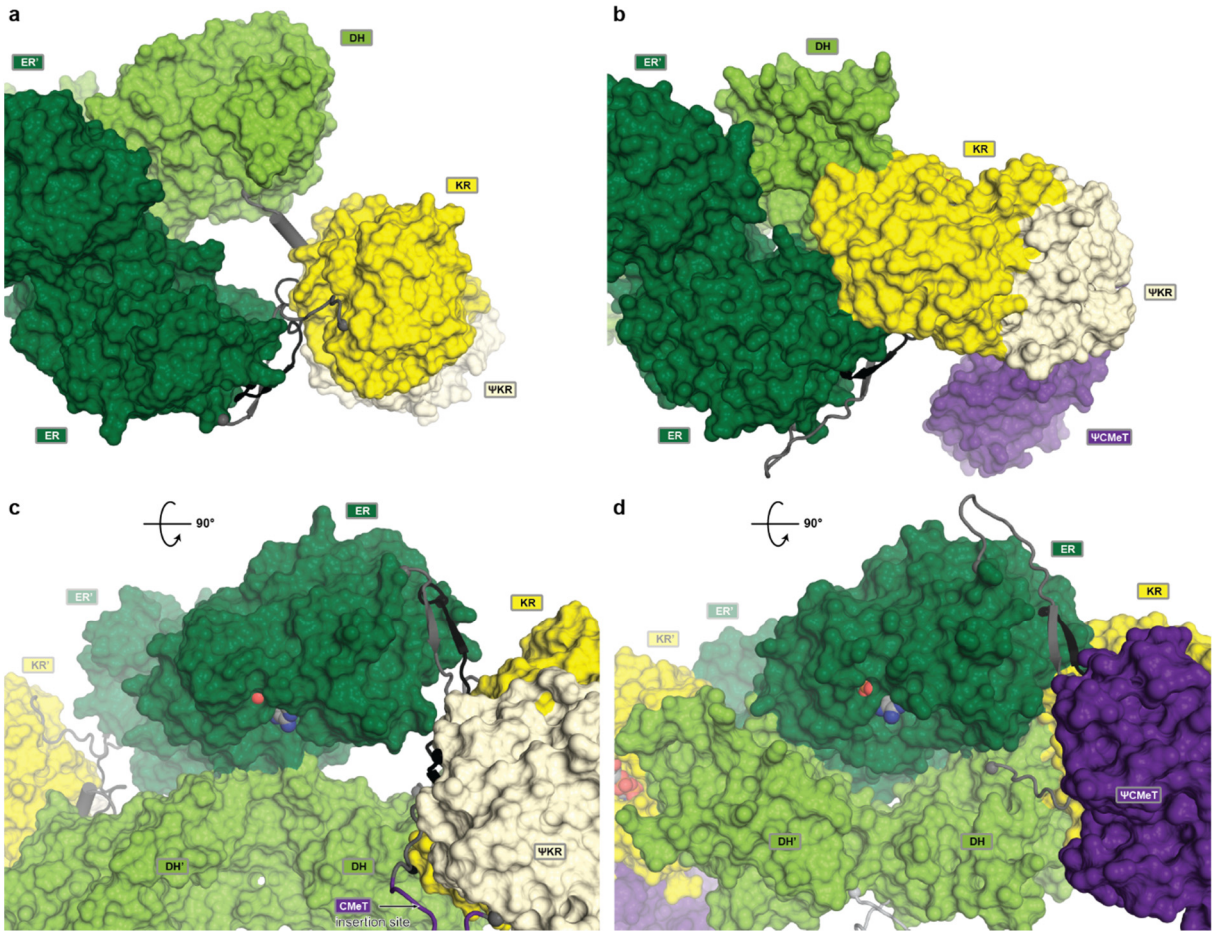


Figure 5.1 | Domain interactions in PKS and FAS fully reducing modifying regions.
a, **b**, Top views on the modifying region of a MAS-like PKS (a) and mFAS (b). In contrast to mFAS, the MAS-like PKS has a linker based architecture without direct domain interfaces of the ΨKR/KR with the DH or ER. **c**, **d**, Side view showing the ER integration into the modifying region. The ER domains in the MAS-like (c) modifying region sits on top of the DH dimer and forms a small and highly variable interface (compare Extended Data Figure 2.10). In mFAS the ER dimer is wedged between the V-shaped DH dimer. The insertion site of a CMeT domain in the PKS architecture is indicated (c). Domains are shown in surface and linkers in cartoon representation.

The next smaller reductive modifying region lacks the ER and produces double bonds. Since the ER is inserted in the loop that connects ΨKR and KR, a plausible architecture of a module with DH-ΨKR/KR can be easily derived by virtual ER removal (Figure 5.2c, d). Indeed, crystal structures of isolated DH[176,178,179] and ΨKR/KR[200] domains from such modules closely resemble the domains in the MAS-like architecture and also contain a short fragment of the DH-ΨKR linker helix[176,178].

The smallest reducing modifying region contains only the ΨKR/KR domain and produces hydroxyl groups. It is the most frequent modifying region, probably because it makes the product hydrophilic, susceptible for tailoring modifications (*e.g.* glycosylation or O-methylation), and prevents side reactions of non-reduced polyketides during transfer steps to downstream modules (see below). The architecture of such a modifying region can be reconstructed from two fragments: One crystal structure of a ΨKR/KR domain was solved containing an N-terminal helical linker that facilitates dimerization[297]. This element occupies a central position that is spatially close enough to reasonably connect a condensing region (Figure 5.2e). Two condensing regions of such modules (KS-AT-ΨKR/KR) have been crystallized[163,164] and reveal the same conserved connection as observed for all other condensing regions. Furthermore, all N-terminal ΨKR/KR linkers (in all modifying regions) contain a helix, which is another indication for the relationship of the three reducing modules and creates a

complete picture of related architectures with staggered complexity. However, this picture is disturbed by the cryo-EM structure of PikAIII (Figure 1.11). The architecture of PikAIII (Figure 5.2f) reveals differences with respect to almost every detail, including an inverted handedness of the LD/AT attachment site, linker distances that require unfolding of domains, no central linker connection to the modifying region, and a different KS active site entrance (chapter 1.6.2)[42,45]. If the structure represents a transition state as part of the normal reaction cycle or describes an alternative architecture is so far unclear and subject to controversial discussions[174].

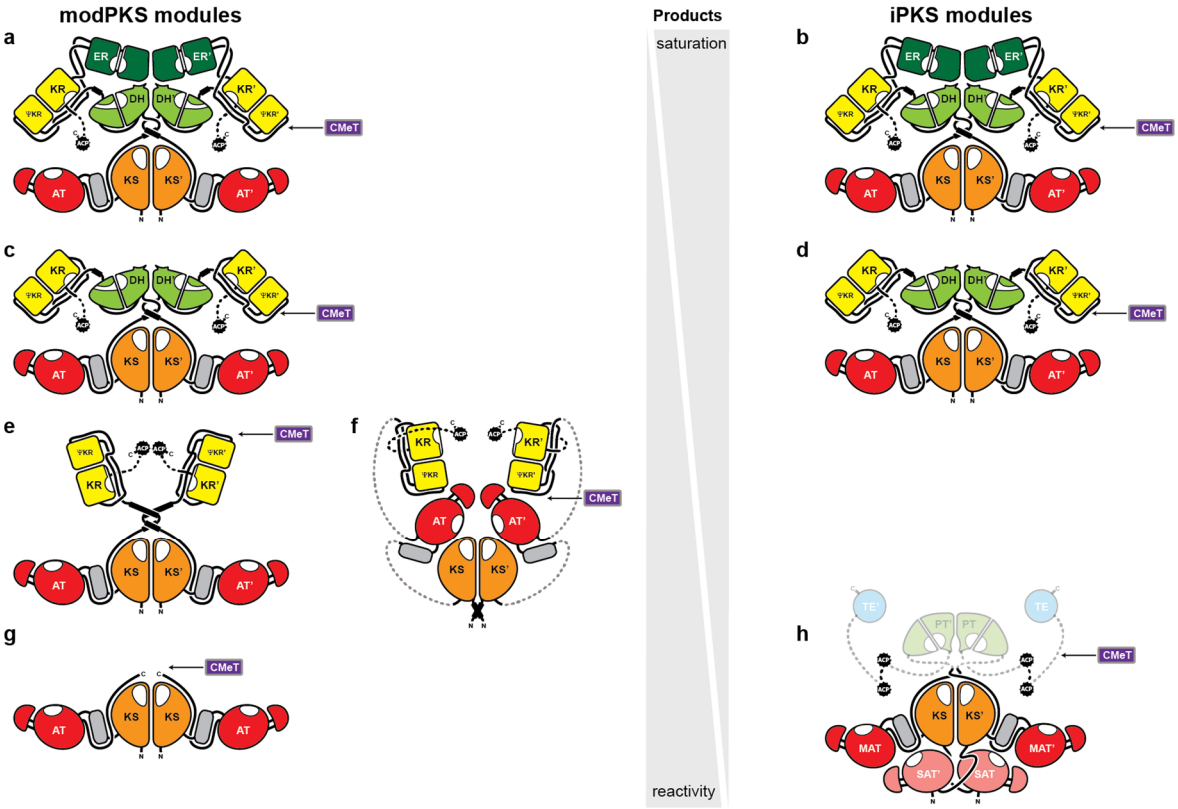


Figure 5.2 | Proposed module architectures in PKS.
PKS modules for the production of fully saturated (top), unsaturated, hydroxylated, and carbonyl substituted products (bottom) in modular (left) and iterative PKS (right). The central bar, describing saturation and reactivity of the products, corresponds to both modPKS and iPKS. Insertion sides for CMeT domains are indicated. Dotted lines indicate uncharacterized linkers. The most reactive PKS products are produced by non-reducing PKS modules and only rarely occur in modPKS. Fully reducing PKS modules produce fully saturated products (fatty acids) of chemically inert character. **a**, **b**, Highly/fully reducing architecture as prototype for modPKS and iPKS modules based on the MAS-like structure[44]. **c**, Partially reducing module bases on (a) and (b). **e**, Smallest reducing module, based on a ΨKR/KR crystal structure with central dimerization element[297] and two condensing region crystal structures[163,164]. **f**, Cryo-EM structure of PikAIII[42,45]. **g**, Least frequent occurring non-reducing module as part of a modPKS. The architecture is assumed to corresponds to an individual condensing region. **h**, NR-PKS, a highly specialized fungal iterative counterpart to (g). The greyed out region supports cyclization of the non-reduced polyketides. Structures of PT[92] and TE[181] are known. The architecture with two ACPs corresponds to CTB1.

A module, even smaller than this contains no reducing domains at all and produces poly-ketones (polyketides), as exemplified by NR-PKS (Figure 5.2g). In contrast to fully reducing modules (Figure 5.2a, b), non-reducing modules produce the most reactive polyketide species, which is so reactive that it can spontaneously cyclize into aromatic ring systems, even without chaperoning by a PT. Such modules are the least frequent of all modules in modPKS and at the same time one of the main representatives of iPKS (Figure 5.2h). The reason why this module is a main representative of iterative

and not modular PKS, might be based on the exorbitant reactivity of the product. While one mFAS ACP has to interact 44 times with catalytic domains of the same module to produce palmitic acid, DEBS uses seven ACPs sequentially to interact with 23 domains for the production of 6-deoxyerythronolide B. Despite the fact that mFAS virtually uses almost twice the amount of domains as DEBS, it requires only 1.25 sec[49] compared to approx. 2 min for DEBS[50]. This is a clear indication that the transfer steps between the modules are rate limiting, which was also suggested by *in vitro* reconstitution of various DEBS assembly lines[298]. For non-reducing modPKS modules this indicates that a highly reactive (poly-)ketone must survive for minutes without causing derailment products, which is particularly challenged by the observation that type I PKS and FAS most likely do not sequester substrates by the ACPs[11,14]. In turn, an NR-PKS module catalyzes no transfer reactions to other modules. Furthermore, the product of a KS condensing reaction does not necessarily have to leave the protective KS active site environment[94], as long as no modification is required as part of a programmed event (*e.g.* as imposed by a CMeT domain). In that case the nascent non-reduced polyketide chain would be transferred between the KS active site cysteine and the ACP Ppant (compare Figure 1.6) without leaving the active site. Only the ACP would leave the KS active site in order to load extender units. This shows that iterative NR-PKS are particularly designed to handle the most reactive form of all polyketides and that iteration can protect the product.

These examples from fully reducing to non-reducing modules show how differently modules are employed and how their function reflects the reactivity of the product.

5.3 Programmed iterations and inter/intramodular transport of substrates

As described above, iteration can be beneficial for the protection of non-reduced highly reactive polyketides. For KS domains in NR-PKS it was shown, that the KS domain determines the polyketide chain length[91] and therefore must bind different polyketides with a larger substrate promiscuity compared to modPKS. Assuming that individual polyketides are bound with different specificities, this can impose a programming, as exemplified by PksCT: If methylation is specifically required by the KS for a polyketide with a specific length, the polyketide has to leave the KS active site during the first three elongation cycles in order to become methylated. Once triple methylated, the binding constant might be so high, that it stays sequestered in the KS active site until it reaches its final length. In turn, no methylation might dramatically reduce affinity and cause a more efficient off-loading via the proofreading function of the R domain or spontaneous release by ACP thioester hydrolysis. A substrate competition mechanism has been shown for the CMeT and KR domains in the Lovastatin producing HR-PKS LovB[70]. Both domains compete for the initial non-reduced polyketide. While the CMeT domain requires C_α-acidity for nucleophilic attack of SAM, which is caused by the C_β-carbonyl, the KR domain reduces the C_β-carbonyl and eliminates C_α-acidity (Figure 5.3a). Consequently, the domain that binds the substrate first, determines the program. Very strong and selective binding has also been shown for a KR domain in a PR-PKS[299], which influences programming. This shows that always the domain determines the substitution pattern that dominates intermediate substrate binding and thereby eventually interrupts the normal substrate channeling (KS→KR→DH→ER) (Figure 5.3b). Via this mechanism fungal iPKS are able to use one module for reactions that require several modules in modPKS with presumably higher turnover. This increased speed presumably causes reduced exposure

of highly reactive intermediates to the environment. According to this theory, the program of an iPKS would correspond to the reaction trajectory along the most susceptible domains for each intermediate product.

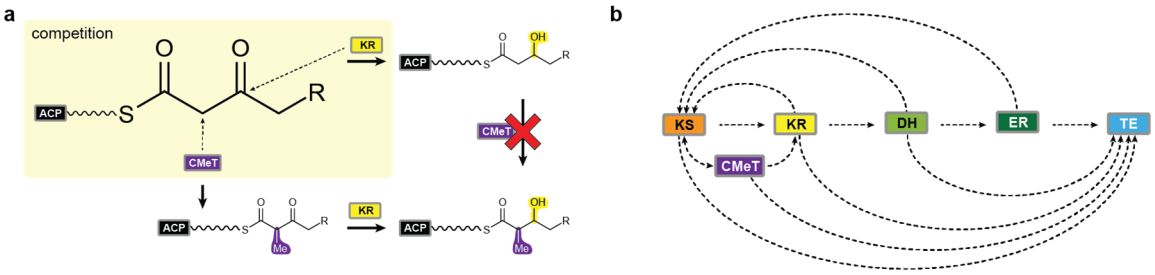


Figure 5.3 | Competing reactions in fungal iPKS.
a, CMeT and KR compete for the same substrate. CMeT can only methylate the C_α-carbon position as long as a KR did not reduce the C_β-carbonyl. **b**, Overview of potentially competing reactions in a HR-PKS, that contains all possible catalytic domains. The arrows indicate substrate channeling. Only the domain that binds the substrate more efficiently will turn over the intermediate. The program of a fungal iPKS describes the path through this scheme.

The same mechanism might also apply to modPKS, but with a much lower substrate promiscuity, which creates a program that is collinear to the organization of the domains in the polypeptide. This requires that every downstream KS must bind a substrate more efficiently than the KS domain of the same module in order to prevent iteration. Recent evolutional analysis revealed that KS domains co-evolve with the modifying region of the upstream module[127] and biochemical assays showed that downstream KS domains indeed specifically recognize substrates[119]. Interestingly, various modules start to iterate once extracted from the assembly line environment[118,119,123], which indicates that the best acceptor domain was removed from the trajectory (compare Figure 5.3b), causing transfer to the next best acceptor (the initial KS). An increasing amount of assembly lines is discovered that contain iterating modules[116,119,124], including a mutant strain of the prototypical DEBS[123]. As such events often produce derailment products and are not the main product of a modPKS, this was initially called “stuttering”. The substrate recognition by downstream modules does not only include the recognition of the substrate itself, but also of the ACP. In order to prevent back transfer to the initial KS, alternating constructive charge-charge interactions were proposed between an ACP and its downstream KS, while charge repulsion with the KS of the same module would hinder back transfer[26]. This was partially confirmed by swapping an N-terminal fragment of ACP helix I of DEBS ACP2 with ACP3, which induced iteration. Another proposed kinetic control mechanism includes a turnstyle, which is proposed to prevent substrate transfer from an upstream ACP as long as the downstream KS is not free[125].

A comparison of iterative and modular substrate transfer within a module reveals further similarities (Figure 5.4) besides iteration. Modules of both types process their substrates along structural regions that are arranged along the central dimer axis, starting with a region at the N-terminus and ending with regions close to the C-terminus: First a substrate is transferred and loaded, respectively, via an N-terminal region to the KS domain, which is the center of the condensing region. The condensing region catalyzes one or several elongation cycles and transfers the product to the modifying region, which comes next in sequence. In NR-PKS this mostly corresponds to a PT (or CMeT), which facilitates cyclization. Finally, the C-terminal ACP either transfers the product to the next module (modPKS) or to a TE (or R), which releases the product and is the most C-terminal domain.

Furthermore, NR-PKS can cooperate with other modules such HR-PKS or FAS, via a direct substrate transfer mechanism from their ACP to the SAT, representing an alternative form of modular transfer[87-89,168]. This transfer provides a possibility to further extend a polyketide with a highly reactive non-reduced/aromatic compound, which is most likely inaccessible to modPKS, because long transfer steps might cause side reactions (chapter 5.2).

In summary, iterative and modPKS modules share common functional and architectural features. They mainly differentiate in their programming, which follows colinearity in modPKS and an encrypted substrate promiscuity trajectory in fungal iPKS. Exceptions from the "rule of colinearity" in modPKS reveal common principles and highlight their close relationship with iPKS.

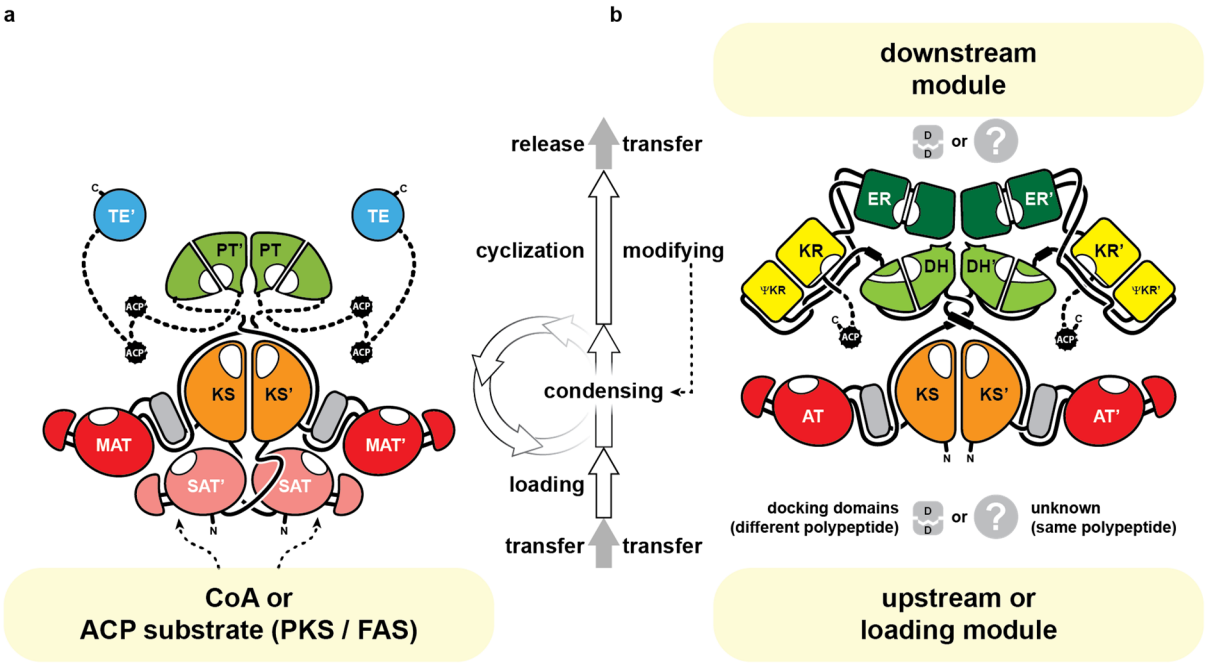


Figure 5.4 | Substrate processing within iPKS and modPKS modules.
Substrate processing follows a principle direction from the N- to the C-terminus (indicated by arrows in the center). Priming substrates of every module can be loaded/transferred from upstream modules. **a**, Architecture of an NR-PKS (CTB1), which uses SAT domains for loading the priming substrate. This can either be derived from CoA (in CTB1) or from another PKS or FAS module (e.g. in PksA). **b**, The architecture of a MAS-like PKS is shown as representative for fully reducing modules in modPKS. Priming substrates in modPKS are loaded via loading modules or transferred from upstream modules (bottom) at the N-terminus. The product is transferred to a downstream module (top). Very little is known about the inter-modular connection and transfer mechanism. If modules of different polypeptides are connected, docking domains establish a connection. Dotted domain connections are structurally not characterized. Representative structures of all domains have been analyzed.

5.4 Architectures designed to be flexible

The presented work revealed conformational coupling in two PKS modules, which creates asymmetry and might coordinate catalysis in the two reaction clefts. Previous data on full length or large fragments of FAS and PKS have already indicated that these synthases can be asymmetric[21,165], but still this intrinsic feature is not always considered.

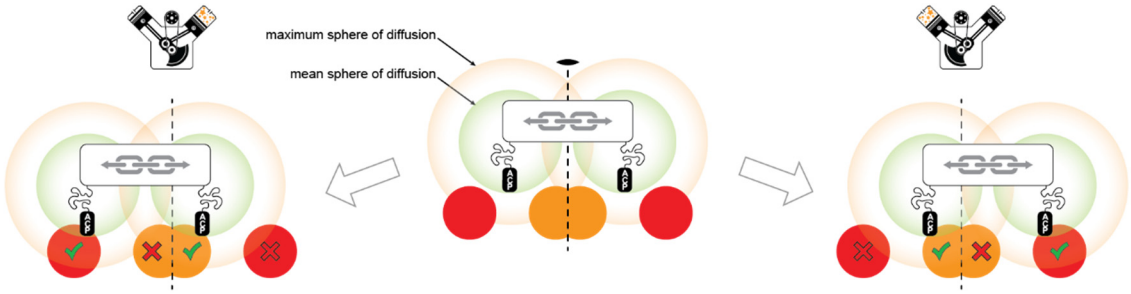


Figure 5.5 | Conformational coupling has the potential to modulate active site distances.
Asymmetry introduced by conformational coupling can guide reactions by moving active sites (red, orange) into the mean sphere of diffusion of an ACP (and vice versa), which is determined by a mean ACP linker length. Although all domains may be in an accessible distance for productive ACP interactions, the binding probability might be increased by architectural adaptations. A feature capable of supporting optimal ACP interactions with all domains in a multienzyme, regardless where it is located. An additional enzyme that selectively binds an ACP in a non-stoichiometric ratio might be able to synchronize the coupling by local structural changes, that either prevent binding of a second ACP or transmit asymmetry via conformational coupling through the multienzyme. A symmetric model with indicated two-fold axis is shown in the center, while the neighboring models show coupled asymmetric interactions and their influence on the mean sphere of diffusion range of the ACPs. The overall model resembles a V-twin engine. Red and orange circles indicate catalytic domains. The colors correspond to KS and (M)AT domains, but they may be any other.

Oligomeric enzymes that catalyze only one reaction and provide identical conditions at all active sites, are often substantially less flexible than PKS, exhibit high symmetry and can be studies at high resolution, like the β-galactosidase[300]. However, PKS and FAS contain plenty of different catalytic domains on one polypeptide and their architecture must provide appropriate conditions for all reactions, including active site distances that must be overcome by the ACPs. From a rational point of view, it appears unlikely that dimeric multienzymes with flexibly tethered carrier proteins and a linker based architecture always adopt symmetric conformations. Therefore, asymmetry of a dimeric multienzymes is not surprising. The question is, how does the asymmetric coupling contribute to the function of PKS? A plausible explanation might be that the observed conformational change increases the turnover by guiding the random walk of the ACP: Every ACP might have a specific mean sphere of diffusion, which keeps the ACP in an average distance for most of the time. Therefore, it might be that some domains beyond this range are reached with lower efficiency, although the absolute linker length allows productive interactions. Conformational changes of the architectural framework have the potential to shift the center of the average ACP location, which is brought to an extreme in mFAS where 180° rotation of the condensing region provide even access to domains far beyond the maximum linker length[165] (Figure 1.9e). Asymmetric coupling of such conformational changes provides an opportunity to move two catalytic domains, that were originally located at the perimeter of the mean sphere of diffusion of the two ACPs, closer to their center and might be most efficient for two domains located proximal (*e.g.* KS) and distal (*e.g.* (M)AT) to the central dimer axis (Figure 5.5).

Besides influences on global variations in active site distances, ACP binding events might also cause local conformational changes. This was shown for the cryo-EM structure of CTB1 SAT-KS-MAT, where ACP binding triggers conformational changes via LD-MAT and SATs (Figure 5.6a). However, it is unclear if this conformational change inhibits ACP binding to the second KS active site or if this inhibition is mediated by changes within the KS dimer, as indicated by enhanced disorder in the binding region of the not visualized second ACP (Supplementary Figure 4.12a). It might be that the conformational change only promotes ACP binding to the MAT. The conformational coupling in the MAS-like PKS modifying region might also reveal changes that are not related to global active site distances. Here,

the coupled ER dimer translation induces rotations of the KR domain, which can either block access to the DH active site entrance or clear it (Figure 5.6b).

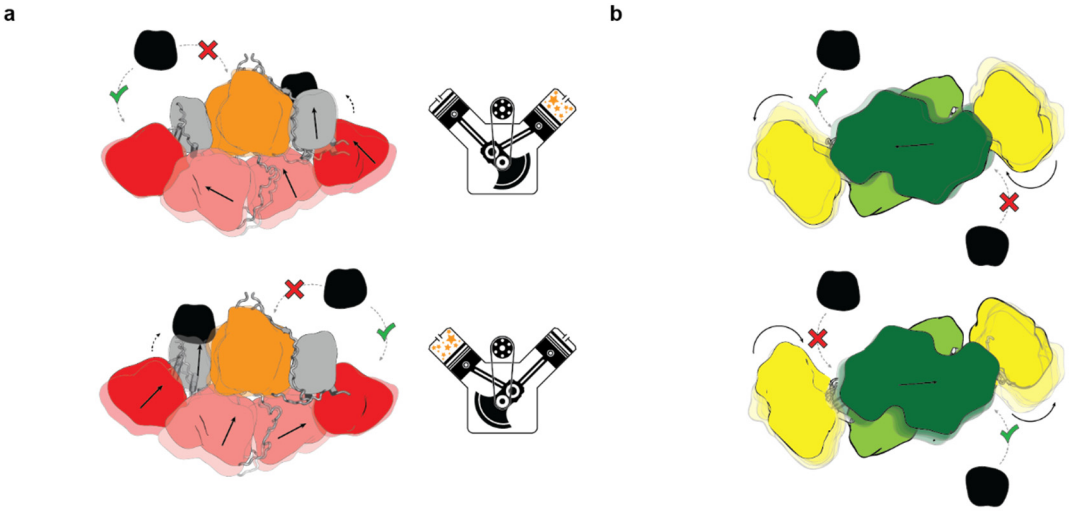


Figure 5.6 | Local variations by conformational coupling in PKS modules.
a, The conformational coupling in an NR-PKS loading/condensing region resembles a V-twin engine: While one ACP can bind to a KS domain the other one is free to be loaded by the MAT. **b**, Top view of the MAS-like PKS modifying region, showing that conformational coupling might have an influence on ACP binding. One putative example is the binding of ACP do the DH active site entrance, which is located in the cleft between DH, ER and ΨKR/KR. This cleft widens and narrows with horizontal translation of the ER dimer and converts this movement into inverted lateral rotations of the ΨKR/KR domains. In this way, access to the DH active site entrance can be exposed or blocked. Coupled domain movements are indicated by arrows. All domains are colored according to (Figure 5.4).

Such a mechanism might increase the turnover of PKS and reflects the mechanism of a V-twin engine where two pistons synergistically work together. In every V-twin engine, combustions have to be coordinated. In analogy, the KS dimer (at least in NR-PKS) fulfills such a synchronization function by selective binding of just one ACP. Notably, the KS domain catalyzes the irreversible step (V-twin engine: combustion; here: CO_2 release) and therefore is predestined to act as regulator of the pathway (chapter 1.1), while equivalent crosslinking experiments with the excised PksA PT domain did not reveal asymmetry[20]. So far, the V-twin engine model is based on individual structural snapshots and the functional half-side binding of ACP to the KS in NR-PKS could be further analyzed by coarse grained molecular dynamics simulations.

For DEBS a turnstile mechanism was proposed (chapter 5.3)[125], where the $KS_{(n)}$ enters a "closed" state and cannot be re-loaded again before it's $ACP_{(n)}$ was off-loaded. This model suggested an involvement of hydrocarbonate/CO_2, which is produced by the KS, but neither considers asymmetry, nor explains iteration of DEBS modules[26,123]. Large scale conformational changes were also observed in the cryo-EM structure of PikAIII, where the (Figure 1.11b, Figure 5.2f) ΨKR/KR was suggested to flip (180°) around its horizontal axis and the rotation and inverse arrangement of the LD-AT (Figure 1.11a, b) imply a transition from the canonical KS-LD-AT architecture (Figure 5.2e, f). However, conformational coupling was not analyzed, because C_2 symmetry was enforced throughout cryo-EM data processing[42,45].

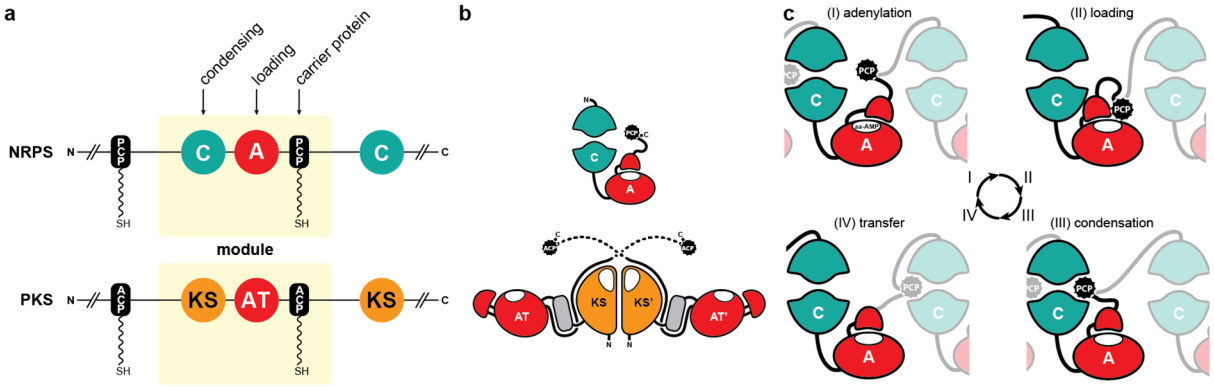


Figure 5.7 | Relationship between PKS and NRPS
a, Schematic comparison of NRPS and PKS modules. The functional organization of catalytic domains in sequence resembles a non-reducing PKS module. **b**, Architectural schemes of (a). NRPS are monomers and employ an adenylation domain (A domain) with a highly flexible C-terminal subdomain (A_{sub}), which is tethered to the PCP. The monomeric C domain is shown as central slice, showing a tunnel with PCP binding sites on each side. **c**, Catalytic cycle and conformational changes in an NRPS module.

Large scale conformational changes are also observed in NRPSs. These multienzymes resemble modPKSs in their assembly line organization and utilization of a 4-helix bundle Ppant-modified carrier protein, the PCP (Figure 1.1c)[301]. Instead of polyketides, NRPS produce polypeptides by modular action of an amino acid adenylation (A) and condensation domain (C). Every NRPS module contains these three domains in the order C-A-PCP, which resembles a PKS condensing region in function and sequence organization (Figure 5.7a). NRPSs show the same colinear organization and synthesis principle as modPKSs and their products can be predicted from sequence. Here, the recruited amino acid can be identified via a sequence motif of the core catalytic A domain ("Stachelhaus motif")[302]. Furthermore, both multienzymes employ TE or R domains in their last module for product release. The modular architecture of NRPS modules has been studies in detail[303-307] and is fundamentally different from PKSs (Figure 5.7b) primarily by the fact that NRPSs are monomers. In analogy to AT domains, an A domain is the gatekeeper of every NRPS module and selects a specific extender unit. The domain catalyzes an ATP dependent activation of an amino acid (adenylation) and its subsequent transesterification/loading to the PCP Ppant[308] (Figure 5.7c). During these reactions the small subunit of the A domain (A_{sub}) rotates by 140°[309] and opens the active site for the PCP distally from the modules C-domain[304,310]. This shows that the activity of the central domain of NRPS modules relies on strong conformational changes. Notably, the PCP is directly connected to A_{sub}, which has the potential to couple A domain activity to PCP motion. Indeed, structural comparison on α,β-methyleneadenosine 5'-triphosphate (AMPCPP)[311], AMP[312] and PCP[310] bound states indicates that the approach of the PCP to the A domain active site tilts A_{sub} in a lever-like movement, which triggers adenylation and subsequent PCP loading in one concerted motion. Afterwards, the loaded $PCP_{(n)}$ as well as the upstream $PCP_{(n-1)}$ translocate to the C domain, which catalyzes peptide bond formation between an amino acid group amine and a PCP thioester. In contrast to PKS, the nascent polypeptide always stays attached to a PCP and is not covalently transferred to the C domain, which prevents iteration. Furthermore, in PKSs the irreversible step (decarboxylation) is catalyzed by the condensing KS, while in NRPSs the loading A domain fulfils this function (ATP hydrolysis).

5.5 Modular assemblies and hybrid modules

Despite of the architectural differences between PKS and NRPS, both multienzymes can occur as hybrids (Figure 1.8b), which dramatically expands the chemical space these machines can explore. Such hybrid NRPS systems are not limited to modPKS, but can also be found for fungal iPKS, that can have a C-terminal NRPS module, which attaches an amino acid after iterative polyketide synthesis, such as the Equistatin (HIV-I integrase inhibitor) producing HR-PKS EqiS[313]. To the best of my knowledge, functional architectures of hybrid synthetases and particularly their modular interconnections are entirely unexplored.

Modular interconnections have only been studied for modules and domains, respectively, which are located at the N- and C-terminus of a polypeptide at the level of excised, small docking domains (DD), which mediate specific interactions with partner domain[46,106-109]. These 25-50 aa large DD regions adopt ordered structures in their respective complexes. Two different types have been characterized for *cis*-AT PKS that form tetrameric complexes, consisting of two homodimers from both ends of each module[46,106,107]. For the more diverse family of *trans*-AT PKS, that often contains split-modules (Figure 1.8b), a third type was identified, which forms hetero-dimers[108,109]. Hetero-dimeric docking domains have also been identified for NRPS and are located between PCP and C domains[110,111]. Depending on their type and specificity, these domains mediate connections either within PKS, NRPS or between both multienzyme families (Figure 1.8b, module 11 and 12) and form hybrid NRPS-PKS.

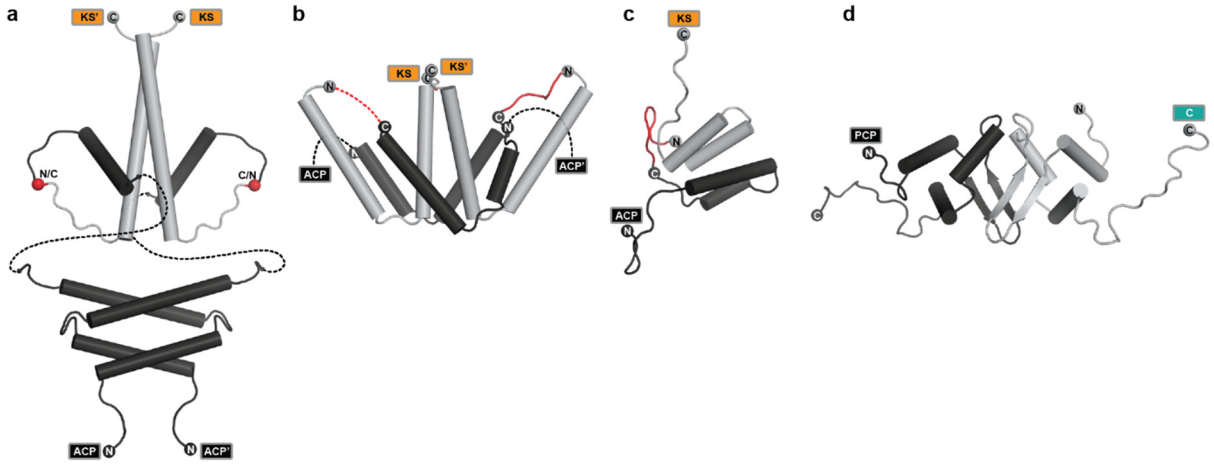


Figure 5.8 | Docking domains in PKS and NRPS
a, *Cis*-AT DD complex of DEBS (PDB: 1PZR, 1PZQ)[106]. The N-terminal part following the ACP often contains a four helix bundle (dimerization element). Structures of indicated regions were determined separately, but are connected by an unstructured linker. N- and C-termini of the construct (black and grey) in the upper part of the panel were fused for structural studies (red ball). **b**, Cis-AT DD complex of the curacin PKS (PDB: 4MYY)[46]. For structural studies the N- and C-termini of the constructs were fused by a glycine-serine linker (red). **c**, *Trans*-AT docking domain complex of the virginiamycin PKS[108]. The constructs were fused similarly as in (c). **d**, NRPS docking domain complex of the tubulysin hybrid NRPS-PKS[111]. The natural complexes of *cis*-AT PKS docking domains are hetero –tetramers, while those of *trans*-AT and NRPS are hetero-dimers.

However, in most cases modules are covalently connected via their continuous polypeptide chain; and only very limited information is available about these connections. It is also unclear if structured elements can be found in the interface regions that either interact with one of the neighboring domains or form an independent domain. Most likely, these regions are flexible and play a crucial role in intermodular transfer and organization. So far, no structure of a PKS bi-module was solved, and a recent negative stain EM reconstruction at 26 Å resolution of a monomeric NRPS bi-module revealed

huge conformational heterogeneity and suggested that NRPSs lack a uniform supermodular architecture[307]. For PKS a low resolution SAXS model of the DEBS3 bimodule was published[43] and revealed minor features. In contrast to NRPS, this might indicate that PKS have a supermodular architecture. Multiple modules as part of one polypeptide/subunit can transfer their products from one assembly line to another via docking domains. These domains were shown to associate *in vitro* with affinities between of 1-20 µM for PKS[46,106,108,109] and approx. 50 µM for NRPS[111]. Compared to an average protein concentration in *E. Coli* of 1 nM to 10 µM[314,315] such interactions might be classified as transient and, to the best of my knowledge, no association of assembly line polypeptides has been observed *in vitro*. This might indicate that these interactions are of transient nature or that the assembly lines must form high local concentrations *in vivo* in order to stably associate. For the PksR subunit of the bacillaene *trans*-AT PKS (Figure 1.8b)[316] from the *Bacillus subtilis*, quantitative Western blots indicated the presence of 50-150 copies per cell[317], which roughly corresponds to a concentration of 50-150 nM, relative to *E.Coli*[1]. For the stable association of assembly line subunits this concentration is by far too low. However, it was shown that the bacillaene assembly line subunits localizes to a membrane region and form regions of high local concentration with a mass of 10-100 MDa[317] (Figure 5.9a-d). The biological reason for the formation of such cellular foci as well as the function of bacillaene itself are unknown. Nevertheless, the clustering of such enzymes close to the membrane points towards a cellular factory that associates and secretes a natural product. This observation indicates that either cellular mechanisms exist that support localization to a single spot in the cell or that this PKS contains domains that facilitate clustering. A complete bacillaene PKS has a mass of approx. 2.5 MDa, therefore this association cannot be solely explained by docking domains. Based on the observation that various KS-LD domains of *trans*-AT PKS have been crystallized with lattice contacts involving LD-LD interactions, an organization in a two dimensional array[318] was proposed (Figure 5.9e). So far there is no experimental evidence for the formation of such arrays or filaments and the biological relevance of such arrays is unclear. Therefore, further investigations of PKS and hybrid NRPS-PKS assembly lines are required to unravel supermodular architecture and intermodular substrate transfer.

Interestingly, similar organizations with different functions can be observed in eukaryotic cells, such as processing-bodies (P-bodies) or stress granules. Such foci can fulfill different functions in mRNA degradation, storage, micro-RNA (miRNA) processing, and nonsense-mediated decay[319]. All these regions have in common that hey organize enzymes in a dynamic and reversible environment, that keeps local substrate and enzyme concentration high and provides short distances for diffusion.

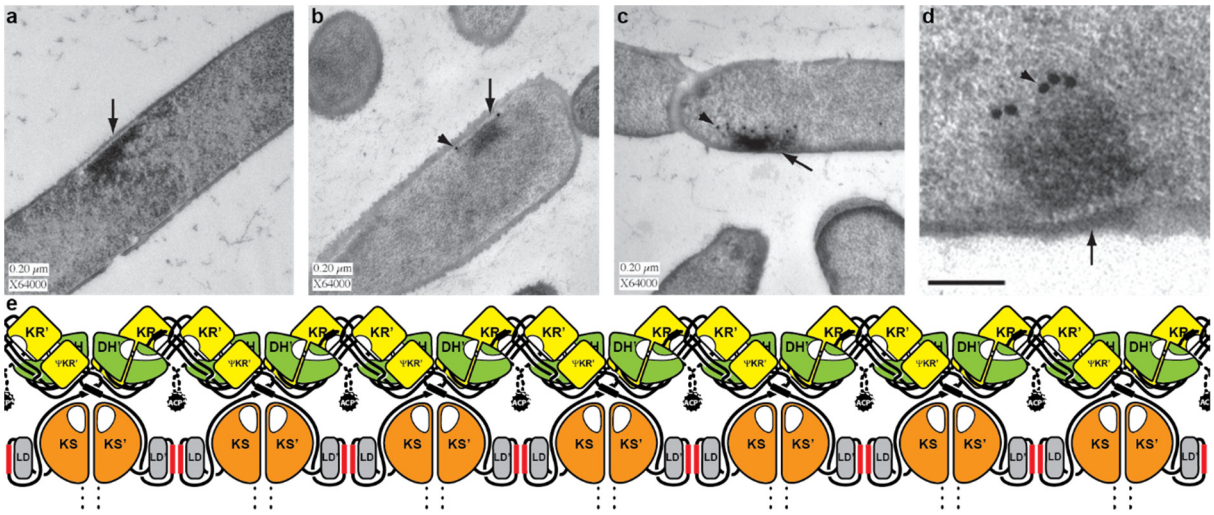


Figure 5.9 | Megacomplexes of *trans*-AT PKS
The bacillaene PKS (Figure 1.8b) in *bacillus subtilis* localizes to a membrane region and forms an organelle like organization in the cytosol. **a**, The negative stain EM micrographs after cryo-sectioning of wild-type *bacillus subtilis* and glutaraldehyde crosslinking shows a region of high density close to the membrane. The white box (scale bar) corresponds to 200 nm. The arrow indicates a region of high density. **b-d**, The PKS polypeptides PksE and PksR (Figure 1.8b) were tagged with fluorescence proteins and transgenic *B. subtilis* strains were generated. Additionally, to the sample preparation in (a), the samples were subjected to immunogold labelling with primary antibodies against the fluorescence protein tag. The micrographs show the localization of gold beads (arrow head) to the dense structure (arrow), corresponding to PksE (b) and PksR (c). **d**, PksR immunogold labelled at higher magnification (scale bar: 100 nm). Panels (a-d) were adapted from Straight P.D. *et al.*[317] Copyright 2006 National Academy of Sciences. **e**, Proposed megacomplex formation by Laterally-INteracting Ketosynthase Sequence (LINKS)[318]. In this model *trans*-AT PKS modules interact via their linker domains (LD) and form a two dimensional array. Only one horizontally arranged module is shown. Individual assembly lines are arranged along the vertical axis.

5.6 Outlook

Since the discovery of polyketide chemistry more than a century ago[147], the perspective on the producers of natural products has changed dramatically. Nobody would have thought that microorganism employ machines with the complexity and supramolecular size of organelles for the production of secondary metabolites. In fact, still very little is known about their higher-order organization and cellular integration. The biosynthetic potential of catalyzing multi-step chemical syntheses by applying assembly lines of catalytic modules and the vision of creating custom products by genetic engineering attracted great attention in natural product research and drug discovery. The incredible size of some PKS subunits suggests that they might start polyketide production, already during translation by the ribosome. A whole set of cellular components might be involved in maintenance and regulation, such as chaperones[202] and various regulatory factors. This shows that more effort will be required in understanding the cellular integration of PKS in order to pave the way for efficient engineering of these multienzymes.

Many enzymes that catalyze just a single reaction can be readily expressed in *E.Coli*, purified in bulk amounts and used in *in vitro* experiments as well as chemo-enzymatic applications. A large amount of PKS are from bacterial origin, but this does not mean that all of them can be expressed in *E.Coli*. Usually, the larger a biological machine, the more complex its regulation and integration in the cellular environment. Indeed, a major bottleneck in recombinant expression of polyketide pathways is the lack

of suitable host strains, which are capable of expressing large PKS proteins and catalyzing tailoring of products[141] (chapter 1.5.4). Besides special chaperoning systems[202], the transcription machinery must be able to generate transcripts of enormous size without premature termination. In *E.Coli* transcription is terminated either via Rho-dependent terminators that cause ATP dependent dissociation of the transcription elongation complex (TEC)[320] or intrinsic features such as GC-rich hairpins followed by uridine rich regions that cause halting of the complex and release of the transcript. Many PKS expressing bacterial strains have a remarkable high genomic GC content, *e.g.* Myxobacteria (70 %)[41], which causes stable RNA-hairpin formation and can result in massive stalling of the TEC as long as the PKS is not expressed in its natural environment. For the difficidin and macrolactin PKS clusters of *Bacillus amyloliquefaciens* it was recently shown that a processive antiterminator[321] of the NusG[322] family called "long operon associated protein" (LoaP) is encoded in close proximity to the gene clusters and prevents premature termination by suppressing transcriptional pause sites in the transcripts of the gene clusters[323]. Without LoaP it was shown that the transcription of the 70 kilo base (kB) polycistronic dificidin RNA is terminated after 10 kB, while the full length is produced in the presence of LoaP. In contrast to unknown transcription initiation mechanism of PKS pathways that can be readily circumvent by replacing promotor regions, factors like LoaP have to be known before they can be used in combinatorial biosynthesis (chapter 1.5.4). Another hidden bottleneck for PKS pathway engineering and expression might involve riboswitches[324], which are likely to occur based on the high GC content of PKS genes and the tendency to form stable hairpins in the RNA. These RNA tertiary structures can bind metabolites and regulate their own expression. Such factors are hard to predict and most likely essential in the natural producer. For combinatorial biosynthesis and recombinant expression, such hidden factors can be avoided by large scale gene synthesis and codon optimization. Unfortunately, synthesis of entire PKS clusters is neither economic nor feasible with current methods, but this technological challenge might be solved in the future.

Besides regulatory mechanism on the expression level, PKS might also be regulated on the protein or cell biological level. Machines like the ribosome can be actively regulated via their subunits and it might be that similar mechanism exist for PKS. However, to the best of my knowledge, none of the common regulatory mechanisms, such as phosphorylation or feedback regulation, have been observed for PKS. Presumably, PKS are not as strictly regulated as other large biosynthetic machines (*e.g.* ribosomes, spliceosomes, polymerase complexes or mFAS and ACC), because they are not part of the primary metabolism and therefore not strictly essential for survival of the organism. Nevertheless, regulation might be possible *e.g.* by phosphopantetheinylation, regulation of type II TEs that remove stalled intermediates, or acyl-CoA carboxylases that are believed to produce extender units for PKS. Cell biological regulation could be accomplished by *e.g.* localization of PKS subunits to a specific region in the cell, as compared to P-bodies (chapter 5.5), in order to order to increase local concentrations and promote transient interaction between docking domains.

The example of the bacillaene PKS (Figure 5.9) shows that PKS subunits locate to a specific spot close to the membrane, suggesting that some tethering mechanism exists. In mycobacteria, PKS are responsible for the synthesis of complex cell wall lipids, which have to be exported from the cell. The last subunit PpsE of the modPKS that produces phthiocerol dimycocerosates (PDIMs) in cooperation with MAS was shown to interact with a membrane transporter. Such interactions were suggested to be a general mechanism for polyketide export[325] and might also apply here. This analogy might indicate that bacillaene is actively secreted and the structural implications of a membrane associated megacomplex make *Bacillus subtilis* particularly interesting for studies on PKS biology. It is the ideal

model organism for PKS research with established genetics, low genomic GC content, LoaP independent transcription of the bacillaene PKS[323], good properties for recombinant protein expression[326], and is simple to culture. Furthermore, some structures of PKS[74,318] and NRPS[39,303] have already been determined from this organism. In addition, some modules of the bacillaene PKS provide optimal conditions for structural studies on bimodules by Cryo-EM or X-ray crystallography. One example is a bimodule of the PksR subunit with the unusual organization KS-DH^0-ACP-KS. Besides of the ACP, all domains should be dimeric and arranged along the central dimer axis without flexibly tethered AT or ΨKR/KR domains that might cause increased heterogeneity. Moreover, the entire megacomplex might be studied *in vivo* by Cryo-EM tomography with sub tomogram averaging and could provide first insights in the organization of a complete PKS assembly lines.

The results of this thesis on PKS architecture are likely to establish new paradigm for PKS organization. Since the work in chapter 2 was published, the structure has already been used to analyze possible integrations sites of a CMeT domain[68] and it might further be used to interpret functional data at lower resolution (*e.g.* bimodules), as exemplified by the crystal structure of mFAS[162] and the analysis of its conformational flexibility during fatty acid production by cryo-EM[165] (Figure 1.9e). For NR-PKSs no structure of larger fragments is known and therefore CTB1 might also become a paradigm for NR-PKS organization. The applied methods provide a roadmap for further structural studies on ACP complexes and will ease structure determination of complexes with SAT and MAT.

Moreover, the results might now be used to design new PKS modules either on the basis of MAS-like PKSs or homolog modPKSs, by guiding fragment exchanges in regions that are free of domain contacts and not involved in secondary structure elements. The most promising approach for producing new natural products might involve domain swaps in combination with directed evolution (chapter 1.5.4) to apply minor adaptations and obtain functional PKSs. The architecture of the NR-PKS loading/condensing region might be used to exchange SAT domains in order to load different starter-units. SATs have been shown to specifically accept starter-units from other HR-PKS and FAS, which provides a possibility to engineer substrate transfer between these multienzymes by specific SAT swaps. Another approach for drug development might not involve PKSs for drug production, but also as targets. Specific inhibitors that inhibit PKS pathways could be useful drugs against microorganism that employ PKS as virulence factors, such as *Mycobacterium tuberculosis*[100].

In a time of rising antibiotic resistances and a discovery void of antibiotics, the results of this thesis will help to better understand the biological machines for the production of life-saving natural products. They can guide engineering for the development of new drugs against the most common threads to human life, such as cancer and infectious diseases.

6 REFERENCES

1 Moran, U., Phillips, R. & Milo, R. SnapShot: key numbers in biology. *Cell* **141**, 1262-1262 e1261, doi:10.1016/j.cell.2010.06.019 (2010).

2 Metzler, D. E. & Metzler, C. M. *Biochemistry : the chemical reactions of living cells*. 2nd edn, (Harcourt/Academic Press, 2001).

3 The UniProt, C. UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* **45**, D158-D169, doi:10.1093/nar/gkw1099 (2017).

4 Watson, J. D. *Molecular biology of the gene*. Seventh edition edn, (Pearson, 2014).

5 Alberts, A. W., Majerus, P. W., Talamo, B. & Vagelos, P. R. Acyl-Carrier Protein. Ii. Intermediary Reactions of Fatty Acid Synthesis. *Biochemistry* **3**, 1563-1571 (1964).

6 Majerus, P. W., Alberts, A. W. & Vagelos, P. R. The Acyl Carrier Protein of Fatty Acid Synthesis: Purification, Physical Properties, and Substrate Binding Site. *Proc. Natl. Acad. Sci. U. S. A.* **51**, 1231-1238 (1964).

7 Majerus, P. W., Alberts, A. W. & Vagelos, P. R. Acyl Carrier Protein. Iv. The Identification of 4'-Phosphopantetheine as the Prosthetic Group of the Acyl Carrier Protein. *Proc. Natl. Acad. Sci. U. S. A.* **53**, 410-417 (1965).

8 Howard, M. J. *et al.* Three-dimensional structure of the major autoantigen in primary biliary cirrhosis. *Gastroenterology* **115**, 139-146 (1998).

9 Wei, J. & Tong, L. Crystal structure of the 500-kDa yeast acetyl-CoA carboxylase holoenzyme dimer. *Nature* **526**, 723-727, doi:10.1038/nature15375 (2015).

10 Haslinger, K., Redfield, C. & Cryle, M. J. Structure of the terminal PCP domain of the non-ribosomal peptide synthetase in teicoplanin biosynthesis. *Proteins* **83**, 711-721, doi:10.1002/prot.24758 (2015).

11 Wattana-amorn, P. *et al.* Solution structure of an acyl carrier protein domain from a fungal type I polyketide synthase. *Biochemistry* **49**, 2186-2193, doi:10.1021/bi902176v (2010).

12 Ploskon, E. *et al.* Recognition of intermediate functionality by acyl carrier protein over a complete cycle of fatty acid biosynthesis. *Chem. Biol.* **17**, 776-785, doi:10.1016/j.chembiol.2010.05.024 (2010).

13 Nguyen, C. *et al.* Trapping the dynamic acyl carrier protein in fatty acid biosynthesis. *Nature* **505**, 427-431, doi:10.1038/nature12810 (2014).

14 Ploskon, E. *et al.* A mammalian type I fatty acid synthase acyl carrier protein domain does not sequester acyl chains. *J. Biol. Chem.* **283**, 518-528, doi:10.1074/jbc.M703454200 (2008).

15 Perham, R. N. Swinging arms and swinging domains in multifunctional enzymes: catalytic machines for multistep reactions. *Annu. Rev. Biochem.* **69**, 961-1004, doi:10.1146/annurev.biochem.69.1.961 (2000).

16 Revill, W. P., Bibb, M. J. & Hopwood, D. A. Relationships between fatty acid and polyketide synthases from Streptomyces coelicolor A3(2): characterization of the fatty acid synthase acyl carrier protein. *J. Bacteriol.* **178**, 5660-5667 (1996).

17 Xu, W., Qiao, K. & Tang, Y. Structural analysis of protein-protein interactions in type I polyketide synthases. *Crit. Rev. Biochem. Mol. Biol.*, doi:10.3109/10409238.2012.745476 (2012).

18 Alekseyev, V. Y., Liu, C. W., Cane, D. E., Puglisi, J. D. & Khosla, C. Solution structure and proposed domain domain recognition interface of an acyl carrier protein domain from a modular polyketide synthase. *Protein Sci.* **16**, 2093-2107, doi:10.1110/ps.073011407 (2007).

19 Babu, M. *et al.* Structure of a SLC26 anion transporter STAS domain in complex with acyl carrier protein: implications for E. coli YchM in fatty acid metabolism. *Structure* **18**, 1450-1462, doi:10.1016/j.str.2010.08.015 (2010).

20 Barajas, J. F. *et al.* Structural and Biochemical Analysis of Protein-Protein Interactions Between the Acyl-Carrier Protein and Product Template Domain. *Angew. Chem. Int. Ed. Engl.* **55**, 13005-13009, doi:10.1002/anie.201605401 (2016).

21 Bruegger, J. *et al.* Probing the selectivity and protein.protein interactions of a nonreducing fungal polyketide synthase using mechanism-based crosslinkers. *Chem. Biol.* **20**, 1135-1146, doi:10.1016/j.chembiol.2013.07.012 (2013).
22 Busche, A. *et al.* Characterization of molecular interactions between ACP and halogenase domains in the Curacin A polyketide synthase. *ACS Chem. Biol.* **7**, 378-386, doi:10.1021/cb200352q (2012).
23 Charkoudian, L. K. *et al.* Probing the interactions of an acyl carrier protein domain from the 6-deoxyerythronolide B synthase. *Protein Sci.* **20**, 1244-1255, doi:10.1002/pro.652 (2011).
24 Chen, A. Y., Schnarr, N. A., Kim, C. Y., Cane, D. E. & Khosla, C. Extender unit and acyl carrier protein specificity of ketosynthase domains of the 6-deoxyerythronolide B synthase. *J. Am. Chem. Soc.* **128**, 3067-3074, doi:10.1021/ja058093d (2006).
25 Cryle, M. J. & Schlichting, I. Structural insights from a P450 Carrier Protein complex reveal how specificity is achieved in the P450(BioI) ACP complex. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 15696-15701, doi:10.1073/pnas.0805983105 (2008).
26 Kapur, S. *et al.* Reprogramming a module of the 6-deoxyerythronolide B synthase for iterative chain elongation. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 4110-4115, doi:10.1073/pnas.1118734109 (2012).
27 Klaus, M. *et al.* Protein-Protein Interactions, Not Substrate Recognition, Dominate the Turnover of Chimeric Assembly Line Polyketide Synthases. *J. Biol. Chem.* **291**, 16404-16415, doi:10.1074/jbc.M116.730531 (2016).
28 Masoudi, A., Raetz, C. R., Zhou, P. & Pemble, C. W. t. Chasing acyl carrier protein through a catalytic cycle of lipid A production. *Nature* **505**, 422-426, doi:10.1038/nature12679 (2014).
29 Ostrowski, M. P., Cane, D. E. & Khosla, C. Recognition of acyl carrier proteins by ketoreductases in assembly line polyketide synthases. *J. Antibiot. (Tokyo)* **69**, 507-510, doi:10.1038/ja.2016.41 (2016).
30 Rafi, S. *et al.* Structure of acyl carrier protein bound to FabI, the FASII enoyl reductase from Escherichia coli. *J. Biol. Chem.* **281**, 39285-39293, doi:10.1074/jbc.M608758200 (2006).
31 Sundlov, J. A., Shi, C., Wilson, D. J., Aldrich, C. C. & Gulick, A. M. Structural and functional investigation of the intermolecular interaction between NRPS adenylation and carrier protein domains. *Chem. Biol.* **19**, 188-198, doi:10.1016/j.chembiol.2011.11.013 (2012).
32 Tallorin, L. *et al.* Trapping of the Enoyl-Acyl Carrier Protein Reductase-Acyl Carrier Protein Interaction. *J. Am. Chem. Soc.* **138**, 3962-3965, doi:10.1021/jacs.5b13456 (2016).
33 Wu, N., Cane, D. E. & Khosla, C. Quantitative analysis of the relative contributions of donor acyl carrier proteins, acceptor ketosynthases, and linker regions to intermodular transfer of intermediates in hybrid polyketide synthases. *Biochemistry* **41**, 5056-5066 (2002).
34 Zhang, L. *et al.* Crystal structure of FabZ-ACP complex reveals a dynamic seesaw-like catalytic mechanism of dehydratase in fatty acid biosynthesis. *Cell Res.* **26**, 1330-1344, doi:10.1038/cr.2016.136 (2016).
35 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).
36 Lambalot, R. H. *et al.* A new enzyme superfamily - the phosphopantetheinyl transferases. *Chem. Biol.* **3**, 923-936 (1996).
37 Quadri, L. E. *et al.* Characterization of Sfp, a Bacillus subtilis phosphopantetheinyl transferase for peptidyl carrier protein domains in peptide synthetases. *Biochemistry* **37**, 1585-1595, doi:10.1021/bi9719861 (1998).
38 Parris, K. D. *et al.* Crystal structures of substrate binding to Bacillus subtilis holo-(acyl carrier protein) synthase reveal a novel trimeric arrangement of molecules resulting in three active sites. *Structure* **8**, 883-895 (2000).
39 Tufar, P. *et al.* Crystal structure of a PCP/Sfp complex reveals the structural basis for carrier protein posttranslational modification. *Chem. Biol.* **21**, 552-562, doi:10.1016/j.chembiol.2014.02.014 (2014).

40 Tong, L. Structure and function of biotin-dependent carboxylases. *Cell. Mol. Life Sci.* **70**, 863-891, doi:10.1007/s00018-012-1096-0 (2013).
41 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).
42 Dutta, S. *et al.* Structure of a modular polyketide synthase. *Nature* **510**, 512-517, doi:10.1038/nature13423 (2014).
43 Edwards, A. L., Matsui, T., Weiss, T. M. & Khosla, C. Architectures of whole-module and bimodular proteins from the 6-deoxyerythronolide B synthase. *J. Mol. Biol.* **426**, 2229-2245, doi:10.1016/j.jmb.2014.03.015 (2014).
44 Herbst, D. A., Jakob, R. P., Zahringer, F. & Maier, T. Mycocerosic acid synthase exemplifies the architecture of reducing polyketide synthases. *Nature* **531**, 533-537, doi:10.1038/nature16993 (2016).
45 Whicher, J. R. *et al.* Structural rearrangements of a polyketide synthase module during its catalytic cycle. *Nature* **510**, 560-564, doi:10.1038/nature13409 (2014).
46 Whicher, J. R. *et al.* Cyanobacterial polyketide synthase docking domains: a tool for engineering natural product biosynthesis. *Chem. Biol.* **20**, 1340-1351, doi:10.1016/j.chembiol.2013.09.015 (2013).
47 Zheng, J., Gay, D. C., Demeler, B., White, M. A. & Keatinge-Clay, A. T. Divergence of multimodular polyketide synthases revealed by a didomain structure. *Nat. Chem. Biol.* **8**, 615-621, doi:10.1038/nchembio.964 (2012).
48 Newman, A. G., Vagstad, A. L., Storm, P. A. & Townsend, C. A. Systematic domain swaps of iterative, nonreducing polyketide synthases provide a mechanistic understanding and rationale for catalytic reprogramming. *J. Am. Chem. Soc.* **136**, 7348-7362, doi:10.1021/ja5007299 (2014).
49 Chang, S. I. & Hammes, G. G. Structure and mechanism of action of a multifunctional enzyme: fatty acid synthase. *Acc. Chem. Res.* **23**, 363-369, doi:10.1021/ar00179a003 (1990).
50 Pieper, R., Ebert-Khosla, S., Cane, D. & Khosla, C. Erythromycin biosynthesis: kinetic studies on a fully active modular polyketide synthase using natural and unnatural substrates. *Biochemistry* **35**, 2054-2060, doi:10.1021/bi952860b (1996).
51 Chopra, T. & Gokhale, R. S. Polyketide versatility in the biosynthesis of complex mycobacterial cell wall lipids. *Methods Enzymol.* **459**, 259-294, doi:10.1016/S0076-6879(09)04612-6 (2009).
52 Newman, D. J., Cragg, G. M. & Snader, K. M. Natural products as sources of new drugs over the period 1981-2002. *J. Nat. Prod.* **66**, 1022-1037, doi:10.1021/np030096l (2003).
53 World-Health-Organization. Antimicrobial resistance: global report on surveillance. *WHO Library Cataloguing-in-Publication Data* (2014).
54 Callaway, E. & Cyranoski, D. Anti-parasite drugs sweep Nobel prize in medicine 2015. *Nature* **526**, 174-175, doi:10.1038/nature.2015.18507 (2015).
55 Hertweck, C. The biosynthetic logic of polyketide diversity. *Angew. Chem. Int. Ed. Engl.* **48**, 4688-4716, doi:10.1002/anie.200806121 (2009).
56 Schmidt, E. W. Trading molecules and tracking targets in symbiotic interactions. *Nat. Chem. Biol.* **4**, 466-473, doi:10.1038/nchembio.101 (2008).
57 Piel, J. Metabolites from symbiotic bacteria. *Nat. Prod. Rep.* **26**, 338-362, doi:10.1039/b703499g (2009).
58 Weissman, K. J. Introduction to polyketide biosynthesis. *Methods Enzymol.* **459**, 3-16, doi:10.1016/S0076-6879(09)04601-1 (2009).
59 Nguyen, T. *et al.* Exploiting the mosaic structure of trans-acyltransferase polyketide synthases for natural product discovery and pathway dissection. *Nat. Biotechnol.* **26**, 225-233, doi:10.1038/nbt1379 (2008).
60 Teta, R. *et al.* Genome mining reveals trans-AT polyketide synthase directed antibiotic biosynthesis in the bacterial phylum bacteroidetes. *Chembiochem* **11**, 2506-2512, doi:10.1002/cbic.201000542 (2010).

61 Fu, J. *et al.* Full-length RecE enhances linear-linear homologous recombination and facilitates direct cloning for bioprospecting. *Nat. Biotechnol.* **30**, 440-446, doi:10.1038/nbt.2183 (2012).
62 Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* **6**, 343-345, doi:10.1038/nmeth.1318 (2009).
63 Zhou, Y. *et al.* Iterative Mechanism of Macrodiolide Formation in the Anticancer Compound Conglobatin. *Chem. Biol.* **22**, 745-754, doi:10.1016/j.chembiol.2015.05.010 (2015).
64 Staunton, J. & Weissman, K. J. Polyketide biosynthesis: a millennium review. *Nat. Prod. Rep.* **18**, 380-416 (2001).
65 Yadav, G., Gokhale, R. S. & Mohanty, D. Computational approach for prediction of domain organization and substrate specificity of modular polyketide synthases. *J. Mol. Biol.* **328**, 335-363 (2003).
66 Xie, X., Khosla, C. & Cane, D. E. Elucidation of the Stereospecificity of C-Methyltransferases from trans-AT Polyketide Synthases. *J. Am. Chem. Soc.* **139**, 6102-6105, doi:10.1021/jacs.7b02911 (2017).
67 Wagner, D. T. *et al.* alpha-Methylation follows condensation in the gephyronic acid modular polyketide synthase. *Chem. Commun. (Camb.)* **52**, 8822-8825, doi:10.1039/c6cc04418b (2016).
68 Skiba, M. A. *et al.* Domain Organization and Active Site Architecture of a Polyketide Synthase C-methyltransferase. *ACS Chem. Biol.* **11**, 3319-3327, doi:10.1021/acschembio.6b00759 (2016).
69 Helfrich, E. J. & Piel, J. Biosynthesis of polyketides by trans-AT polyketide synthases. *Nat. Prod. Rep.* **33**, 231-316, doi:10.1039/c5np00125k (2016).
70 Cacho, R. A. *et al.* Understanding Programming of Fungal Iterative Polyketide Synthases: The Biochemical Basis for Regioselectivity by the Methyltransferase Domain in the Lovastatin Megasynthase. *J. Am. Chem. Soc.* **137**, 15688-15691, doi:10.1021/jacs.5b11814 (2015).
71 Weissman, K. J. Polyketide stereocontrol: a study in chemical biology. *Beilstein J. Org. Chem.* **13**, 348-371, doi:10.3762/bjoc.13.39 (2017).
72 Keatinge-Clay, A. T. Stereocontrol within polyketide assembly lines. *Nat. Prod. Rep.* **33**, 141-149, doi:10.1039/c5np00092k (2016).
73 Keatinge-Clay, A. T. A tylosin ketoreductase reveals how chirality is determined in polyketides. *Chem. Biol.* **14**, 898-908, doi:10.1016/j.chembiol.2007.07.009 (2007).
74 Gay, D. C., Spear, P. J. & Keatinge-Clay, A. T. A double-hotdog with a new trick: structure and mechanism of the trans-acyltransferase polyketide synthase enoyl-isomerase. *ACS Chem. Biol.* **9**, 2374-2381, doi:10.1021/cb500459b (2014).
75 Garg, A., Xie, X., Keatinge-Clay, A., Khosla, C. & Cane, D. E. Elucidation of the cryptic epimerase activity of redox-inactive ketoreductase domains from modular polyketide synthases by tandem equilibrium isotope exchange. *J. Am. Chem. Soc.* **136**, 10190-10193, doi:10.1021/ja5056998 (2014).
76 Chen, H. & Du, L. Iterative polyketide biosynthesis by modular polyketide synthases in bacteria. *Appl. Microbiol. Biotechnol.* **100**, 541-557, doi:10.1007/s00253-015-7093-0 (2016).
77 Kaulmann, U. & Hertweck, C. Biosynthesis of polyunsaturated fatty acids by polyketide synthases. *Angew. Chem. Int. Ed. Engl.* **41**, 1866-1869 (2002).
78 Van Lanen, S. G., Lin, S. & Shen, B. Biosynthesis of the enediyne antitumor antibiotic C-1027 involves a new branching point in chorismate metabolism. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 494-499, doi:10.1073/pnas.0708750105 (2008).
79 Liu, W., Christenson, S. D., Standage, S. & Shen, B. Biosynthesis of the enediyne antitumor antibiotic C-1027. *Science* **297**, 1170-1173, doi:10.1126/science.1072110 (2002).
80 Blodgett, J. A. *et al.* Common biosynthetic origins for polycyclic tetramate macrolactams from phylogenetically diverse bacteria. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 11692-11697, doi:10.1073/pnas.1001513107 (2010).
81 Li, Y. *et al.* Iterative assembly of two separate polyketide chains by the same single-module bacterial polyketide synthase in the biosynthesis of HSAF. *Angew. Chem. Int. Ed. Engl.* **53**, 7524-7530, doi:10.1002/anie.201403500 (2014).

82 Roberts, D. M. *et al.* Substrate selectivity of an isolated enoyl reductase catalytic domain from an iterative highly reducing fungal polyketide synthase reveals key components of programming. *Chem. Sci.* **8**, 1116-1126, doi:10.1039/c6sc03496a (2017).
83 Zabala, A. O., Chooi, Y. H., Choi, M. S., Lin, H. C. & Tang, Y. Fungal polyketide synthase product chain-length control by partnering thiohydrolase. *ACS Chem. Biol.* **9**, 1576-1586, doi:10.1021/cb500284t (2014).
84 Cox, R. J. Polyketides, proteins and genes in fungi: programmed nano-machines begin to reveal their secrets. *Org. Biomol. Chem.* **5**, 2010-2026, doi:10.1039/b704420h (2007).
85 Crawford, J. M. *et al.* Deconstruction of iterative multidomain polyketide synthase function. *Science* **320**, 243-246, doi:10.1126/science.1154711 (2008).
86 Beck, J., Ripka, S., Siegner, A., Schiltz, E. & Schweizer, E. The multifunctional 6-methylsalicylic acid synthase gene of Penicillium patulum. Its gene structure relative to that of other polyketide synthases. *Eur. J. Biochem.* **192**, 487-498 (1990).
87 Foulke-Abel, J. & Townsend, C. A. Demonstration of starter unit interprotein transfer from a fatty acid synthase to a multidomain, nonreducing polyketide synthase. *Chembiochem* **13**, 1880-1884, doi:10.1002/cbic.201200267 (2012).
88 Brown, D. W., Adams, T. H. & Keller, N. P. Aspergillus has distinct fatty acid synthases for primary and secondary metabolism. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 14873-14877 (1996).
89 Hitchman, T. S. *et al.* Hexanoate synthase, a specialized type I fatty acid synthase in aflatoxin B1 biosynthesis. *Bioorg. Chem.* **29**, 293-307, doi:10.1006/bioo.2001.1216 (2001).
90 Cochrane, R. V. *et al.* Production of New Cladosporin Analogues by Reconstitution of the Polyketide Synthases Responsible for the Biosynthesis of this Antimalarial Agent. *Angew. Chem. Int. Ed. Engl.* **55**, 664-668, doi:10.1002/anie.201509345 (2016).
91 Vagstad, A. L. *et al.* Combinatorial domain swaps provide insights into the rules of fungal polyketide synthase programming and the rational synthesis of non-native aromatic products. *Angew. Chem. Int. Ed. Engl.* **52**, 1718-1721, doi:10.1002/anie.201208550 (2013).
92 Crawford, J. M. *et al.* Structural basis for biosynthetic programming of fungal aromatic polyketide cyclization. *Nature* **461**, 1139-1143, doi:10.1038/nature08475 (2009).
93 Fujii, I., Watanabe, A., Sankawa, U. & Ebizuka, Y. Identification of Claisen cyclase domain in fungal polyketide synthase WA, a naphthopyrone synthase of Aspergillus nidulans. *Chem. Biol.* **8**, 189-197 (2001).
94 Vagstad, A. L., Bumpus, S. B., Belecki, K., Kelleher, N. L. & Townsend, C. A. Interrogation of global active site occupancy of a fungal iterative polyketide synthase reveals strategies for maintaining biosynthetic fidelity. *J. Am. Chem. Soc.* **134**, 6865-6877, doi:10.1021/ja3016389 (2012).
95 Sirakova, T. D., Thirumala, A. K., Dubey, V. S., Sprecher, H. & Kolattukudy, P. E. The Mycobacterium tuberculosis pks2 gene encodes the synthase for the hepta- and octamethyl-branched fatty acids required for sulfolipid synthesis. *J. Biol. Chem.* **276**, 16833-16839, doi:10.1074/jbc.M011468200 (2001).
96 Brown, L., Wolf, J. M., Prados-Rosales, R. & Casadevall, A. Through the wall: extracellular vesicles in Gram-positive bacteria, mycobacteria and fungi. *Nat. Rev. Microbiol.* **13**, 620-630, doi:10.1038/nrmicro3480 (2015).
97 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).
98 Camacho, L. R. *et al.* Analysis of the phthiocerol dimycocerosate locus of Mycobacterium tuberculosis. Evidence that this lipid is involved in the cell wall permeability barrier. *J. Biol. Chem.* **276**, 19845-19854, doi:10.1074/jbc.M100662200 (2001).
99 Astarie-Dequeker, C. *et al.* Phthiocerol dimycocerosates of M. tuberculosis participate in macrophage invasion by inducing changes in the organization of plasma membrane lipids. *PLoS Pathog.* **5**, e1000289, doi:10.1371/journal.ppat.1000289 (2009).

100 Cambier, C. J. *et al.* Mycobacteria manipulate macrophage recruitment through coordinated use of membrane lipids. *Nature* **505**, 218-222, doi:10.1038/nature12799 (2014).
101 Reed, M. B. *et al.* A glycolipid of hypervirulent tuberculosis strains that inhibits the innate immune response. *Nature* **431**, 84-87, doi:10.1038/nature02837 (2004).
102 Cox, J. S., Chen, B., McNeil, M. & Jacobs, W. R., Jr. Complex lipid determines tissue-specific replication of Mycobacterium tuberculosis in mice. *Nature* **402**, 79-83, doi:10.1038/47042 (1999).
103 Boritsch, E. C. *et al.* pks5-recombination-mediated surface remodelling in Mycobacterium tuberculosis emergence. *Nat Microbiol* **1**, 15019, doi:10.1038/nmicrobiol.2015.19 (2016).
104 Etienne, G. *et al.* Identification of the polyketide synthase involved in the biosynthesis of the surface-exposed lipooligosaccharides in mycobacteria. *J. Bacteriol.* **191**, 2613-2621, doi:10.1128/JB.01235-08 (2009).
105 Nye, D. E. America's assembly line. *Choice: Current Reviews for Academic Libraries* **51**, 1062-1062 (2014).
106 Broadhurst, R. W., Nietlispach, D., Wheatcroft, M. P., Leadlay, P. F. & Weissman, K. J. The structure of docking domains in modular polyketide synthases. *Chem. Biol.* **10**, 723-731 (2003).
107 Buchholz, T. J. *et al.* Structural basis for binding specificity between subclasses of modular polyketide synthase docking domains. *ACS Chem. Biol.* **4**, 41-52, doi:10.1021/cb8002607 (2009).
108 Dorival, J. *et al.* Characterization of Intersubunit Communication in the Virginiamycin trans-Acyl Transferase Polyketide Synthase. *J. Am. Chem. Soc.* **138**, 4155-4167, doi:10.1021/jacs.5b13372 (2016).
109 Zeng, J. *et al.* Portability and Structure of the Four-Helix Bundle Docking Domains of trans-Acyltransferase Modular Polyketide Synthases. *ACS Chem. Biol.* **11**, 2466-2474, doi:10.1021/acschembio.6b00345 (2016).
110 Dowling, D. P. *et al.* Structural elements of an NRPS cyclization domain and its intermodule docking domain. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 12432-12437, doi:10.1073/pnas.1608615113 (2016).
111 Richter, C. D., Nietlispach, D., Broadhurst, R. W. & Weissman, K. J. Multienzyme docking in hybrid megasynthetases. *Nat. Chem. Biol.* **4**, 75-81, doi:10.1038/nchembio.2007.61 (2008).
112 Cortes, J., Haydock, S. F., Roberts, G. A., Bevitt, D. J. & Leadlay, P. F. An unusually large multifunctional polypeptide in the erythromycin-producing polyketide synthase of Saccharopolyspora erythraea. *Nature* **348**, 176-178, doi:10.1038/348176a0 (1990).
113 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).
114 Schwecke, T. *et al.* The biosynthetic gene cluster for the polyketide immunosuppressant rapamycin. *Proc. Natl. Acad. Sci. U. S. A.* **92**, 7839-7843 (1995).
115 Omura, S. *et al.* Genome sequence of an industrial microorganism Streptomyces avermitilis: deducing the ability of producing secondary metabolites. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 12215-12220, doi:10.1073/pnas.211433198 (2001).
116 Moss, S. J., Martin, C. J. & Wilkinson, B. Loss of co-linearity by modular polyketide synthases: a mechanism for the evolution of chemical diversity. *Nat. Prod. Rep.* **21**, 575-593, doi:10.1039/b315020h (2004).
117 Hardt, I. H. *et al.* New natural epothilones from Sorangium cellulosum, strains So ce90/B2 and So ce90/D13: isolation, structure elucidation, and SAR studies. *J. Nat. Prod.* **64**, 847-856 (2001).
118 Beck, B. J., Aldrich, C. C., Fecik, R. A., Reynolds, K. A. & Sherman, D. H. Iterative chain elongation by a pikromycin monomodular polyketide synthase. *J. Am. Chem. Soc.* **125**, 4682-4683, doi:10.1021/ja029974c (2003).
119 Busch, B. *et al.* Multifactorial control of iteration events in a modular polyketide assembly line. *Angew. Chem. Int. Ed. Engl.* **52**, 5285-5289, doi:10.1002/anie.201301322 (2013).
120 He, J. & Hertweck, C. Iteration as programmed event during polyketide assembly; molecular analysis of the aureothin biosynthesis gene cluster. *Chem. Biol.* **10**, 1225-1232 (2003).

121 Gaitatzis, N. *et al.* The biosynthesis of the aromatic myxobacterial electron transport inhibitor stigmatellin is directed by a novel type of modular polyketide synthase. *J. Biol. Chem.* **277**, 13082-13090, doi:10.1074/jbc.M111738200 (2002).
122 Olano, C. *et al.* Biosynthesis of the angiogenesis inhibitor borrelidin by Streptomyces parvulus Tu4055: cluster analysis and assignment of functions. *Chem. Biol.* **11**, 87-97, doi:10.1016/j.chembiol.2003.12.018 (2004).
123 Wilkinson, B. *et al.* Novel octaketide macrolides related to 6-deoxyerythronolide B provide evidence for iterative operation of the erythromycin polyketide synthase. *Chem. Biol.* **7**, 111-117 (2000).
124 Sugimoto, Y. *et al.* Freedom and Constraint in Engineered Noncolinear Polyketide Assembly Lines. *Chem. Biol.*, doi:10.1016/j.chembiol.2014.12.014 (2015).
125 Lowry, B., Li, X., Robbins, T., Cane, D. E. & Khosla, C. A Turnstile Mechanism for the Controlled Growth of Biosynthetic Intermediates on Assembly Line Polyketide Synthases. *ACS Cent Sci* **2**, 14-20, doi:10.1021/acscentsci.5b00321 (2016).
126 Keatinge-Clay, A. T. The Uncommon Enzymology of Cis-Acyltransferase Assembly Lines. *Chem. Rev.* **117**, 5334-5366, doi:10.1021/acs.chemrev.6b00683 (2017).
127 Zhang, L. *et al.* Characterization of Giant Modular PKSs Provides Insight into Genetic Mechanism for Structural Diversification of Aminopolyol Polyketides. *Angew. Chem. Int. Ed. Engl.* **56**, 1740-1745, doi:10.1002/anie.201611371 (2017).
128 Piel, J. Biosynthesis of polyketides by trans-AT polyketide synthases. *Nat. Prod. Rep.* **27**, 996-1047, doi:10.1039/b816430b (2010).
129 Dunn, B. J., Watts, K. R., Robbins, T., Cane, D. E. & Khosla, C. Comparative analysis of the substrate specificity of trans- versus cis-acyltransferases of assembly line polyketide synthases. *Biochemistry* **53**, 3796-3806, doi:10.1021/bi5004316 (2014).
130 Jensen, K. *et al.* Polyketide proofreading by an acyltransferase-like enzyme. *Chem. Biol.* **19**, 329-339, doi:10.1016/j.chembiol.2012.01.005 (2012).
131 Bretschneider, T. *et al.* Vinylogous chain branching catalysed by a dedicated polyketide synthase module. *Nature* **502**, 124-128, doi:10.1038/nature12588 (2013).
132 McDaniel, R. *et al.* Multiple genetic modifications of the erythromycin polyketide synthase to produce a library of novel "unnatural" natural products. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 1846-1851 (1999).
133 Liu, T., Sanchez, J. F., Chiang, Y. M., Oakley, B. R. & Wang, C. C. Rational domain swaps reveal insights about chain length control by ketosynthase domains in fungal nonreducing polyketide synthases. *Org Lett* **16**, 1676-1679, doi:10.1021/ol5003384 (2014).
134 Chooi, Y. H. & Tang, Y. Navigating the fungal polyketide chemical space: from genes to molecules. *J. Org. Chem.* **77**, 9933-9953, doi:10.1021/jo301592k (2012).
135 Xu, Y. *et al.* Characterization of the biosynthetic genes for 10,11-dehydrocurvularin, a heat shock response-modulating anticancer fungal polyketide from Aspergillus terreus. *Appl. Environ. Microbiol.* **79**, 2038-2047, doi:10.1128/AEM.03334-12 (2013).
136 Xu, Y. *et al.* Diversity-oriented combinatorial biosynthesis of benzenediol lactone scaffolds by subunit shuffling of fungal polyketide synthases. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 12354-12359, doi:10.1073/pnas.1406999111 (2014).
137 Sun, H., Liu, Z., Zhao, H. & Ang, E. L. Recent advances in combinatorial biosynthesis for drug discovery. *Drug Des. Devel. Ther.* **9**, 823-833, doi:10.2147/DDDT.S63023 (2015).
138 Cortes, J. *et al.* Repositioning of a domain in a modular polyketide synthase to promote specific chain cleavage. *Science* **268**, 1487-1489 (1995).
139 Bohm, I. *et al.* Engineering of a minimal modular polyketide synthase, and targeted alteration of the stereospecificity of polyketide chain extension. *Chem. Biol.* **5**, 407-412 (1998).
140 Kao, C. M., Luo, G., Katz, L., Cane, D. E. & Khosla, C. Manipulation of macrolide ring size by directed mutagenesis of a modular polyketide synthase. *J. Am. Chem. Soc.* **117**, 9105-9106, doi:10.1021/ja00140a043 (1995).

141 Weissman, K. J. Genetic engineering of modular PKSs: from combinatorial biosynthesis to synthetic biology. *Nat. Prod. Rep.* **33**, 203-230, doi:10.1039/c5np00109a (2016).
142 Goudie, A. C. *et al.* Doramectin--a potent novel endectocide. *Vet. Parasitol.* **49**, 5-15 (1993).
143 Weissman, K. J. & Leadlay, P. F. Combinatorial biosynthesis of reduced polyketides. *Nat. Rev. Microbiol.* **3**, 925-936, doi:10.1038/nrmicro1287 (2005).
144 Fischbach, M. A., Lai, J. R., Roche, E. D., Walsh, C. T. & Liu, D. R. Directed evolution can rapidly improve the activity of chimeric assembly-line enzymes. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 11951-11956, doi:10.1073/pnas.0705348104 (2007).
145 Olano, C., Méndez, C. & Salas, J. A. in *Natural Products* 317-339 (John Wiley & Sons, Inc., 2014).
146 Runguphan, W., Qu, X. & O'Connor, S. E. Integrating carbon-halogen bond formation into medicinal plant metabolism. *Nature* **468**, 461-464, doi:10.1038/nature09524 (2010).
147 Collie, N. & Myers, W. S. VII.-The formation of orcinol and other condensation products from dehydracetic acid. *Journal of the Chemical Society, Transactions* **63**, 122-128, doi:10.1039/CT8936300122 (1893).
148 Smith, S. & Tsai, S. C. The type I fatty acid and polyketide synthases: a tale of two megasynthases. *Nat. Prod. Rep.* **24**, 1041-1072, doi:10.1039/b603600g (2007).
149 Brindley, D. N., Matsumura, S. & Bloch, K. Mycobacterium phlei Fatty Acid Synthetase[mdash]A Bacterial Multienzyme Complex. *Nature* **224**, 666-669 (1969).
150 Malpartida, F. & Hopwood, D. A. Molecular cloning of the whole biosynthetic pathway of a Streptomyces antibiotic and its expression in a heterologous host. *Nature* **309**, 462-464 (1984).
151 Kumar, S., Dorsey, J. K. & Porter, J. W. Mechanism of dissociation of pigeon liver fatty acid synthetase complex into half-molecular weight subunits and their reassociation to enzymatically active complex. *Biochem. Biophys. Res. Commun.* **40**, 825-832 (1970).
152 Smith, S. & Abraham, S. Fatty acid synthetase from lactating rat mammary gland. 3. Dissociation and reassociation. *J. Biol. Chem.* **246**, 6428-6435 (1971).
153 Witkowski, A., Joshi, A. & Smith, S. Fatty acid synthase: in vitro complementation of inactive mutants. *Biochemistry* **35**, 10569-10575, doi:10.1021/bi960910m (1996).
154 Gokhale, R. S., Lau, J., Cane, D. E. & Khosla, C. Functional orientation of the acyltransferase domain in a module of the erythromycin polyketide synthase. *Biochemistry* **37**, 2524-2528, doi:10.1021/bi971887n (1998).
155 Kao, C. M., Pieper, R., Cane, D. E. & Khosla, C. Evidence for two catalytically independent clusters of active sites in a functional modular polyketide synthase. *Biochemistry* **35**, 12363-12368, doi:10.1021/bi9616312 (1996).
156 Brink, J. *et al.* Quaternary structure of human fatty acid synthase by electron cryomicroscopy. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 138-143, doi:10.1073/pnas.012589499 (2002).
157 Witkowski, A. *et al.* Head-to-head coiled arrangement of the subunits of the animal fatty acid synthase. *Chem. Biol.* **11**, 1667-1676, doi:10.1016/j.chembiol.2004.09.016 (2004).
158 Staunton, J. *et al.* Evidence for a double-helical structure for modular polyketide synthases. *Nat. Struct. Biol.* **3**, 188-192 (1996).
159 Asturias, F. J. *et al.* Structure and molecular organization of mammalian fatty acid synthase. *Nat. Struct. Mol. Biol.* **12**, 225-232, doi:10.1038/nsmb899 (2005).
160 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).
161 Keatinge-Clay, A. T. & Stroud, R. M. The structure of a ketoreductase determines the organization of the beta-carbon processing enzymes of modular polyketide synthases. *Structure* **14**, 737-748, doi:10.1016/j.str.2006.01.009 (2006).
162 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).
163 Tang, Y., Kim, C. Y., Mathews, II, Cane, D. E. & Khosla, C. The 2.7-Angstrom crystal structure of a 194-kDa homodimeric fragment of the 6-deoxyerythronolide B synthase. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 11124-11129, doi:10.1073/pnas.0601924103 (2006).

164 Tang, Y., Chen, A. Y., Kim, C. Y., Cane, D. E. & Khosla, C. Structural and mechanistic analysis of protein interactions in module 3 of the 6-deoxyerythronolide B synthase. *Chem. Biol.* **14**, 931-943, doi:10.1016/j.chembiol.2007.07.012 (2007).
165 Brignole, E. J., Smith, S. & Asturias, F. J. Conformational flexibility of metazoan fatty acid synthase enables catalysis. *Nat. Struct. Mol. Biol.* **16**, 190-197, doi:10.1038/nsmb.1532 (2009).
166 Brignole, E. J. & Asturias, F. Single-particle electron microscopy of animal fatty acid synthase describing macromolecular rearrangements that enable catalysis. *Methods Enzymol.* **483**, 179-202, doi:10.1016/S0076-6879(10)83009-5 (2010).
167 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).
168 Crawford, J. M., Dancy, B. C., Hill, E. A., Udwary, D. W. & Townsend, C. A. Identification of a starter unit acyl-carrier protein transacylase domain in an iterative type I polyketide synthase. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 16728-16733, doi:10.1073/pnas.0604112103 (2006).
169 Liew, C. W. *et al.* Induced-fit upon ligand binding revealed by crystal structures of the hot-dog fold thioesterase in dynemicin biosynthesis. *J. Mol. Biol.* **404**, 291-306, doi:10.1016/j.jmb.2010.09.041 (2010).
170 Tsai, S. C. *et al.* Crystal structure of the macrocycle-forming thioesterase domain of the erythromycin polyketide synthase: versatility from a unique substrate channel. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 14808-14813, doi:10.1073/pnas.011399198 (2001).
171 Tsai, S. C., Lu, H., Cane, D. E., Khosla, C. & Stroud, R. M. Insights into channel architecture and substrate specificity from crystal structures of two macrocycle-forming thioesterases of modular polyketide synthases. *Biochemistry* **41**, 12598-12606 (2002).
172 Pappenberger, G. *et al.* Structure of the human fatty acid synthase KS-MAT didomain as a framework for inhibitor design. *J. Mol. Biol.* **397**, 508-519, doi:10.1016/j.jmb.2010.01.066 (2010).
173 Miyanaga, A., Iwasawa, S., Shinohara, Y., Kudo, F. & Eguchi, T. Structure-based analysis of the molecular interactions between acyltransferase and acyl carrier protein in vicenistatin biosynthesis. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 1802-1807, doi:10.1073/pnas.1520042113 (2016).
174 Rittner, A. & Grininger, M. Modular polyketide synthases (PKSs): a new model fits all? *Chembiochem* **15**, 2489-2493, doi:10.1002/cbic.201402432 (2014).
175 Keatinge-Clay, A. Crystal structure of the erythromycin polyketide synthase dehydratase. *J. Mol. Biol.* **384**, 941-953, doi:10.1016/j.jmb.2008.09.084 (2008).
176 Akey, D. L. *et al.* Crystal structures of dehydratase domains from the curacin polyketide biosynthetic pathway. *Structure* **18**, 94-105, doi:10.1016/j.str.2009.10.018 (2010).
177 Faille, A. *et al.* Insights into Substrate Modification by Dehydratases from Type I Polyketide Synthases. *J. Mol. Biol.*, doi:10.1016/j.jmb.2017.03.026 (2017).
178 Fiers, W. D., Dodge, G. J., Sherman, D. H., Smith, J. L. & Aldrich, C. C. Vinylogous Dehydration by a Polyketide Dehydratase Domain in Curacin Biosynthesis. *J. Am. Chem. Soc.* **138**, 16024-16036, doi:10.1021/jacs.6b09748 (2016).
179 Gay, D., You, Y. O., Keatinge-Clay, A. & Cane, D. E. Structure and Stereospecificity of the Dehydratase Domain from the Terminal Module of the Rifamycin Polyketide Synthase. *Biochemistry*, doi:10.1021/bi400988t (2013).
180 Ames, B. D. *et al.* Crystal structure and biochemical studies of the trans-acting polyketide enoyl reductase LovC from lovastatin biosynthesis. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 11144-11149, doi:10.1073/pnas.1113029109 (2012).
181 Korman, T. P. *et al.* Structure and function of an iterative polyketide synthase thioesterase domain catalyzing Claisen cyclization in aflatoxin biosynthesis. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 6246-6251, doi:10.1073/pnas.0913531107 (2010).
182 Winter, J. M. *et al.* Biochemical and Structural Basis for Controlling Chemical Modularity in Fungal Polyketide Biosynthesis. *J. Am. Chem. Soc.* **137**, 9885-9893, doi:10.1021/jacs.5b04520 (2015).

183 Agarwal, V., Lin, S., Lukk, T., Nair, S. K. & Cronan, J. E. Structure of the enzyme-acyl carrier protein (ACP) substrate gatekeeper complex required for biotin synthesis. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 17406-17411, doi:10.1073/pnas.1207028109 (2012).
184 Guy, J. E. *et al.* Remote control of regioselectivity in acyl-acyl carrier protein-desaturases. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 16594-16599, doi:10.1073/pnas.1110221108 (2011).
185 Maloney, F. P., Gerwick, L., Gerwick, W. H., Sherman, D. H. & Smith, J. L. Anatomy of the beta-branching enzyme of polyketide biosynthesis and its interaction with an acyl-ACP substrate. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 10316-10321, doi:10.1073/pnas.1607210113 (2016).
186 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).
187 Worthington, A. S., Rivera, H., Torpey, J. W., Alexander, M. D. & Burkart, M. D. Mechanism-based protein cross-linking probes to investigate carrier protein-mediated biosynthesis. *ACS Chem. Biol.* **1**, 687-691, doi:10.1021/cb6003965 (2006).
188 Worthington, A. S., Porter, D. F. & Burkart, M. D. Mechanism-based crosslinking as a gauge for functional interaction of modular synthases. *Org. Biomol. Chem.* **8**, 1769-1772, doi:10.1039/b925966j (2010).
189 Worthington, A. S. *et al.* Probing the compatibility of type II ketosynthase-carrier protein partners. *Chembiochem* **9**, 2096-2103, doi:10.1002/cbic.200800198 (2008).
190 Worthington, A. S. & Burkart, M. D. One-pot chemo-enzymatic synthesis of reporter-modified proteins. *Org. Biomol. Chem.* **4**, 44-46, doi:10.1039/b512735a (2006).
191 Meier, J. L., Mercer, A. C., Rivera, H., Jr. & Burkart, M. D. Synthesis and evaluation of bioorthogonal pantetheine analogues for in vivo protein modification. *J. Am. Chem. Soc.* **128**, 12174-12184, doi:10.1021/ja063217n (2006).
192 Meier, J. L., Haushalter, R. W. & Burkart, M. D. A mechanism based protein crosslinker for acyl carrier protein dehydratases. *Bioorg. Med. Chem. Lett.* **20**, 4936-4939, doi:10.1016/j.bmcl.2010.06.028 (2010).
193 Haushalter, R. W., Worthington, A. S., Hur, G. H. & Burkart, M. D. An orthogonal purification strategy for isolating crosslinked domains of modular synthases. *Bioorg. Med. Chem. Lett.* **18**, 3039-3042, doi:10.1016/j.bmcl.2008.01.026 (2008).
194 Barajas, J. F. *et al.* Polyketide mimetics yield structural and mechanistic insights into product template domain function in nonreducing polyketide synthases. *Proc. Natl. Acad. Sci. U. S. A.*, doi:10.1073/pnas.1609001114 (2017).
195 Herbst, D. A., Jakob, R. P., Zahringer, F. & Maier, T. Corrigendum: Mycocerosic acid synthase exemplifies the architecture of reducing polyketide synthases. *Nature* **536**, 360, doi:10.1038/nature18281 (2016).
196 Weissman, K. J. Uncovering the structures of modular polyketide synthases. *Nat. Prod. Rep.*, doi:10.1039/c4np00098f (2014).
197 Wong, F. T. & Khosla, C. Combinatorial biosynthesis of polyketides--a perspective. *Curr. Opin. Chem. Biol.* **16**, 117-123, doi:10.1016/j.cbpa.2012.01.018 (2012).
198 Khare, D. *et al.* Structural Basis for Cyclopropanation by a Unique Enoyl-Acyl Carrier Protein Reductase. *Structure*, doi:10.1016/j.str.2015.09.013 (2015).
199 Sippel, K. H., Vyas, N. K., Zhang, W., Sankaran, B. & Quiocho, F. A. Crystal structure of the human fatty acid synthase enoyl-acyl carrier protein-reductase domain complexed with triclosan reveals allosteric protein-protein interface inhibition. *J. Biol. Chem.* **289**, 33287-33295, doi:10.1074/jbc.M114.608547 (2014).
200 Bonnett, S. A. *et al.* Structural and stereochemical analysis of a modular polyketide synthase ketoreductase domain required for the generation of a cis-alkene. *Chem. Biol.* **20**, 772-783, doi:10.1016/j.chembiol.2013.04.014 (2013).
201 Hardwicke, M. A. *et al.* A human fatty acid synthase inhibitor binds beta-ketoacyl reductase in the keto-substrate site. *Nat. Chem. Biol.* **10**, 774-779, doi:10.1038/nchembio.1603 (2014).

202 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).
203 Kabsch, W. XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125-132, doi:doi:10.1107/S0907444909047337 (2010).
204 Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213-221, doi:10.1107/S0907444909052925 (2010).
205 McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Cryst.* **40**, 658-674, doi:10.1107/S0021889807021206 (2007).
206 Cowtan, K. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 1002-1011, doi:doi:10.1107/S0907444906022116 (2006).
207 Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126-2132, doi:10.1107/S0907444904019158 (2004).
208 Cowtan, K. Recent developments in classical density modification. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 470-478, doi:10.1107/S090744490903947X (2010).
209 Bricogne G., B. E., Brandl M., Flensburg C., Keller P., Paciorek W., & Roversi P, S. A., Smart O.S., Vonrhein C., Womack T.O. BUSTER version 2.10.2. *Cambridge, United Kingdom: Global Phasing Ltd.* (2011).
210 Karplus, P. A. & Diederichs, K. Linking crystallographic model and data quality. *Science* **336**, 1030-1033, doi:10.1126/science.1218231 (2012).
211 Schwede, T., Kopp, J., Guex, N. & Peitsch, M. C. SWISS-MODEL: An automated protein homology-modeling server. *Nucleic Acids Res.* **31**, 3381-3385 (2003).
212 Fabiola, F., Korostelev, A. & Chapman, M. S. Bias in cross-validated free R factors: mitigation of the effects of non-crystallographic symmetry. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 227-238, doi:doi:10.1107/S0907444905040680 (2006).
213 Smart, O. S. *et al.* Exploiting structure similarity in refinement: automated NCS and target-structure restraints in BUSTER. *Acta Crystallogr. D Biol. Crystallogr.* **68**, 368-380, doi:doi:10.1107/S0907444911056058 (2012).
214 Cowtan, K. An automated procedure for phase improvement by density modification. *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography* **31**, 34-38 (1994).
215 Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved Methods for Building Protein Models in Electron-Density Maps and the Location of Errors in These Models. *Acta Crystallogr. A* **47**, 110-119, doi:Doi 10.1107/S0108767390010224 (1991).
216 Vistica, J. *et al.* Sedimentation equilibrium analysis of protein interactions with global implicit mass conservation constraints and systematic noise decomposition. *Anal. Biochem.* **326**, 234-256, doi:10.1016/j.ab.2003.12.014 (2004).
217 Petoukhov, M. V. *et al.* New developments in the ATSAS program package for small-angle scattering data analysis. *J. Appl. Cryst.* **45**, 342-350, doi:doi:10.1107/S0021889812007662 (2012).
218 Pelikan, M., Hura, G. L. & Hammel, M. Structure and flexibility within proteins as identified through small angle X-ray scattering. *Gen. Physiol. Biophys.* **28**, 174-189 (2009).
219 Zheng, W. & Tekpinar, M. Accurate flexible fitting of high-resolution protein structures to small-angle x-ray scattering data using a coarse-grained model with implicit hydration shell. *Biophys. J.* **101**, 2981-2991, doi:10.1016/j.bpj.2011.11.003 (2011).
220 Brunger, A. T. *et al.* Crystallography & NMR System: A New Software Suite for Macromolecular Structure Determination. *Acta Crystallogr. D Biol. Crystallogr.* **54**, 905-921, doi:doi:10.1107/S0907444998003254 (1998).
221 Schwieters, C. D., Kuszewski, J. J., Tjandra, N. & Clore, G. M. The Xplor-NIH NMR molecular structure determination package. *J. Magn. Reson.* **160**, 65-73 (2003).

222 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).
223 Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774-797, doi:10.1016/j.jmb.2007.05.022 (2007).
224 Collaborative Computational Project, N. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **50**, 760-763, doi:doi:10.1107/S0907444994003112 (1994).
225 Kleywegt, G. J. Validation of protein models from Calpha coordinates alone. *J. Mol. Biol.* **273**, 371-376, doi:10.1006/jmbi.1997.1309 (1997).
226 Biasini, M. *et al.* SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.* **42**, W252-258, doi:10.1093/nar/gku340 (2014).
227 Sali, A. & Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234**, 779-815, doi:10.1006/jmbi.1993.1626 (1993).
228 Skjaerven, L., Yao, X. Q., Scarabelli, G. & Grant, B. J. Integrating protein structural dynamics and evolutionary analysis with Bio3D. *BMC Bioinformatics* **15**, 399, doi:10.1186/s12859-014-0399-6 (2014).
229 Kabsch, W. A solution for the best rotation to relate two sets of vectors. *Acta Crystallogr. A* **32**, 922-923, doi:doi:10.1107/S0567739476001873 (1976).
230 Cock, P. J. *et al.* Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **25**, 1422-1423, doi:10.1093/bioinformatics/btp163 (2009).
231 Schrodinger, L. *The PyMOL Molecular Graphics System, Version 1.7.0.3* (2010).
232 Kleywegt, G. Use of Non-crystallographic Symmetry in Protein Structure Refinement. *Acta Crystallogr. D Biol. Crystallogr.* **52**, 842-857, doi:doi:10.1107/S0907444995016477 (1996).
233 Medek, P. B., P.; Sochor, J. Computation of tunnels in protein molecules using Delaunay triangulation. *Journal of WSCG, University of West Bohemia, Pilsen* **15(1-3)**, 107-114 (2007).
234 Larkin, M. A. *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947-2948, doi:10.1093/bioinformatics/btm404 (2007).
235 Kearse, M. *et al.* Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647-1649, doi:10.1093/bioinformatics/bts199 (2012).
236 Storm, P. A., Herbst, D. A., Maier, T. & Townsend, C. A. Functional and Structural Analysis of Programmed C-Methylation in the Biosynthesis of the Fungal Polyketide Citrinin. *Cell Chem Biol* **24**, 316-325, doi:10.1016/j.chembiol.2017.01.008 (2017).
237 Crawford, J. M. & Townsend, C. A. New insights into the formation of fungal aromatic polyketides. *Nat. Rev. Microbiol.* **8**, 879-889, doi:10.1038/nrmicro2465 (2010).
238 Ahuja, M. *et al.* Illuminating the diversity of aromatic polyketide synthases in Aspergillus nidulans. *J. Am. Chem. Soc.* **134**, 8212-8221, doi:10.1021/ja3016395 (2012).
239 Kroken, S., Glass, N. L., Taylor, J. W., Yoder, O. C. & Turgeon, B. G. Phylogenomic analysis of type I polyketide synthase genes in pathogenic and saprobic ascomycetes. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 15670-15675, doi:10.1073/pnas.2532165100 (2003).
240 Struck, A. W., Thompson, M. L., Wong, L. S. & Micklefield, J. S-adenosyl-methionine-dependent methyltransferases: highly versatile enzymes in biocatalysis, biosynthesis and other biotechnological applications. *Chembiochem* **13**, 2642-2655, doi:10.1002/cbic.201200556 (2012).
241 Shimizu, T. *et al.* Polyketide synthase gene responsible for citrinin biosynthesis in Monascus purpureus. *Appl. Environ. Microbiol.* **71**, 3453-3457, doi:10.1128/AEM.71.7.3453-3457.2005 (2005).
242 He, Y. & Cox, R. J. The molecular steps of citrinin biosynthesis in fungi. *Chemical Science* **7**, 2119-2127, doi:10.1039/C5SC04027B (2016).
243 Keatinge-Clay, A. T. The structures of type I polyketide synthases. *Nat. Prod. Rep.* **29**, 1050-1073, doi:10.1039/c2np20019h (2012).

244 Udwary, D. W., Merski, M. & Townsend, C. A. A method for prediction of the locations of linker regions within large multifunctional proteins, and application to a type I polyketide synthase. *J. Mol. Biol.* **323**, 585-598 (2002).
245 Solovyev, V., Kosarev, P., Seledsov, I. & Vorobyev, D. Automatic annotation of eukaryotic genes, pseudogenes and promoters. *Genome Biol.* **7 Suppl 1**, S10 11-12, doi:10.1186/gb-2006-7-s1-s10 (2006).
246 Sakai, K., Kinoshita, H., Shimizu, T. & Nihira, T. Construction of a citrinin gene cluster expression system in heterologous Aspergillus oryzae. *J Biosci Bioeng* **106**, 466-472, doi:10.1263/jbb.106.466 (2008).
247 Davison, J. *et al.* Genetic, molecular, and biochemical basis of fungal tropolone biosynthesis. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 7642-7647, doi:10.1073/pnas.1201469109 (2012).
248 Belecki, K. & Townsend, C. A. Biochemical determination of enzyme-bound metabolites: preferential accumulation of a programmed octaketide on the enediyne polyketide synthase CalE8. *J. Am. Chem. Soc.* **135**, 14339-14348, doi:10.1021/ja406697t (2013).
249 Ma, S. M. *et al.* Complete reconstitution of a highly reducing iterative polyketide synthase. *Science* **326**, 589-592, doi:10.1126/science.1175602 (2009).
250 Fisch, K. M. *et al.* Rational domain swaps decipher programming in fungal highly reducing polyketide synthases and resurrect an extinct metabolite. *J. Am. Chem. Soc.* **133**, 16635-16641, doi:10.1021/ja206914q (2011).
251 Kitagawa, M. *et al.* Complete set of ORF clones of Escherichia coli ASKA library (a complete set of E. coli K-12 ORF archive): unique resources for biological research. *DNA Res.* **12**, 291-299, doi:10.1093/dnares/dsi012 (2005).
252 Liscombe, D. K., Louie, G. V. & Noel, J. P. Architectures, mechanisms and molecular evolution of natural product methyltransferases. *Nat. Prod. Rep.* **29**, 1238-1250, doi:10.1039/c2np20029e (2012).
253 Huitt-Roehl, C. R. *et al.* Starter unit flexibility for engineered product synthesis by the nonreducing polyketide synthase PksA. *ACS Chem. Biol.* **10**, 1443-1449, doi:10.1021/acschembio.5b00005 (2015).
254 Martin, J. L. & McMillan, F. M. SAM (dependent) I AM: the S-adenosylmethionine-dependent methyltransferase fold. *Curr. Opin. Struct. Biol.* **12**, 783-793 (2002).
255 Peng, Y. *et al.* Structural basis of substrate recognition in human nicotinamide N-methyltransferase. *Biochemistry* **50**, 7800-7808, doi:10.1021/bi2007614 (2011).
256 Yeh, H. H. *et al.* Engineering fungal nonreducing polyketide synthase by heterologous expression and domain swapping. *Org Lett* **15**, 756-759, doi:10.1021/ol303328t (2013).
257 Qian, J., Khandogin, J., West, A. H. & Cook, P. F. Evidence for a catalytic dyad in the active site of homocitrate synthase from Saccharomyces cerevisiae. *Biochemistry* **47**, 6851-6858, doi:10.1021/bi800087k (2008).
258 Khaleeli, N., Busby, R. W. & Townsend, C. A. Site-directed mutagenesis and biochemical analysis of the endogenous ligands in the ferrous active site of clavaminate synthase. The His-3 variant of the 2-His-1-carboxylate model. *Biochemistry* **39**, 8666-8673 (2000).
259 Kyte, J. & Doolittle, R. F. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**, 105-132 (1982).
260 Conrad, R. S., Massey, L. K. & Sokatch, J. R. D- and L-isoleucine metabolism and regulation of their pathways in Pseudomonas putida. *J. Bacteriol.* **118**, 103-111 (1974).
261 Steyn, P. S. & Vleggaar, R. Biosynthesis of asteltoxin by cultures of Emericella variecolor. The role of propionate in the biosynthesis and evidence for a 1,2-bond migration in the formation of the bistetrahydrofuran moiety. *Journal of the Chemical Society, Chemical Communications*, 977-979, doi:10.1039/C39840000977 (1984).
262 Steyn, P. S., Vleggaar, R. & Wessels, P. L. Biosynthesis of the aurovertins B and D. The role of methionine and propionate in the simultaneous operation of two independent biosynthetic pathways. *Journal of the Chemical Society, Perkin Transactions 1*, 1298-1308, doi:10.1039/P19810001298 (1981).

263 Van Duyne, G. D., Standaert, R. F., Karplus, P. A., Schreiber, S. L. & Clardy, J. Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. *J. Mol. Biol.* **229**, 105-124, doi:10.1006/jmbi.1993.1012 (1993).
264 Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* **9**, 671-675 (2012).
265 Waterman, D. G., Winter, G., Parkhurst, J.M., Fuentes-Montero, L., Hattne, J., Brewster, A., Sauter, N.K., and Evans, G. The DIALS framework for integration software. *CCP4 Newsl. Protein Crystallogr.*, 16-19 (2013).
266 Sheldrick, G. M. A short history of SHELX. *Acta Crystallogr. A* **64**, 112-122, doi:10.1107/S0108767307043930 (2008).
267 Kleywegt, G. J. & Jones, T. A. Model building and refinement practice. *Methods Enzymol.* **277**, 208-230 (1997).
268 Schrodinger, L. *The PyMOL Molecular Graphics System, Version 1.8* (2015).
269 Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539, doi:10.1038/msb.2011.75 (2011).
270 Weissman, K. J. The structural biology of biosynthetic megaenzymes. *Nat. Chem. Biol.* **11**, 660-670, doi:10.1038/nchembio.1883 (2015).
271 Crawford, J. M., Vagstad, A. L., Whitworth, K. P., Ehrlich, K. C. & Townsend, C. A. Synthetic strategy of nonreducing iterative polyketide synthases and the origin of the classical "starter-unit effect". *Chembiochem* **9**, 1019-1023, doi:10.1002/cbic.200700702 (2008).
272 Liu, T., Chiang, Y. M., Somoza, A. D., Oakley, B. R. & Wang, C. C. Engineering of an "unnatural" natural product by swapping polyketide synthase domains in Aspergillus nidulans. *J. Am. Chem. Soc.* **133**, 13314-13316, doi:10.1021/ja205780g (2011).
273 Choquer, M. *et al.* The CTB1 gene encoding a fungal polyketide synthase is required for cercosporin biosynthesis and fungal virulence of Cercospora nicotianae. *Mol. Plant. Microbe Interact.* **18**, 468-476, doi:10.1094/MPMI-18-0468 (2005).
274 Chung, K. R., Ehrenshaft, M., Wetzel, D. K. & Daub, M. E. Cercosporin-deficient mutants by plasmid tagging in the asexual fungus Cercospora nicotianae. *Mol. Genet. Genomics* **270**, 103-113, doi:10.1007/s00438-003-0902-7 (2003).
275 Daub, M. E. & Ehrenshaft, M. THE PHOTOACTIVATED CERCOSPORA TOXIN CERCOSPORIN: Contributions to Plant Disease and Fundamental Biology. *Annu Rev Phytopathol* **38**, 461-490, doi:10.1146/annurev.phyto.38.1.461 (2000).
276 Newman, A. G., Vagstad, A. L., Belecki, K., Scheerer, J. R. & Townsend, C. A. Analysis of the cercosporin polyketide synthase CTB1 reveals a new fungal thioesterase function. *Chem. Commun. (Camb.)* **48**, 11772-11774, doi:10.1039/c2cc36010a (2012).
277 Wang, F. *et al.* Structural and Functional Analysis of the Loading Acyltransferase from Avermectin Modular Polyketide Synthase. *ACS Chem. Biol.* **10**, 1017-1025, doi:10.1021/cb500873k (2015).
278 Changeux, J. P. & Edelstein, S. Conformational selection or induced fit? 50 years of debate resolved. *F1000 Biol. Rep.* **3**, 19, doi:10.3410/B3-19 (2011).
279 Keatinge-Clay, A. T., Maltby, D. A., Medzihradszky, K. F., Khosla, C. & Stroud, R. M. An antibiotic factory caught in action. *Nat. Struct. Mol. Biol.* **11**, 888-893, doi:10.1038/nsmb808 (2004).
280 Li, X. *et al.* Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nat Methods* **10**, 584-590, doi:10.1038/nmeth.2472 (2013).
281 McLeod, R. A., Kowal, J., Ringler, P. & Stahlberg, H. Robust image alignment for cryogenic transmission electron microscopy. *J. Struct. Biol.* **197**, 279-293, doi:10.1016/j.jsb.2016.12.006 (2017).
282 Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216-221, doi:10.1016/j.jsb.2015.08.008 (2015).
283 Tang, G. *et al.* EMAN2: an extensible image processing suite for electron microscopy. *J. Struct. Biol.* **157**, 38-46, doi:10.1016/j.jsb.2006.05.009 (2007).

284 Kimanius, D., Forsberg, B. O., Scheres, S. H. & Lindahl, E. Accelerated cryo-EM structure determination with parallelisation using GPUs in RELION-2. *Elife* **5**, doi:10.7554/eLife.18722 (2016).
285 Zhang, K. Gctf: Real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1-12, doi:10.1016/j.jsb.2015.11.003 (2016).
286 Rosenthal, P. B. & Henderson, R. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J. Mol. Biol.* **333**, 721-745 (2003).
287 Chen, S. *et al.* High-resolution noise substitution to measure overfitting and validate resolution in 3D structure determination by single particle electron cryomicroscopy. *Ultramicroscopy* **135**, 24-35, doi:10.1016/j.ultramic.2013.06.004 (2013).
288 Urnavicius, L. *et al.* The structure of the dynactin complex and its interaction with dynein. *Science* **347**, 1441-1446, doi:10.1126/science.aaa4080 (2015).
289 Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. *Nat Methods* **11**, 63-65, doi:10.1038/nmeth.2727 (2014).
290 Bai, X. C., Rajendra, E., Yang, G., Shi, Y. & Scheres, S. H. Sampling the conformational space of the catalytic subunit of human gamma-secretase. *Elife* **4**, doi:10.7554/eLife.11182 (2015).
291 Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605-1612, doi:10.1002/jcc.20084 (2004).
292 McLeod, R. A., Diogo Righetto, R., Stewart, A. & Stahlberg, H. MRCZ - A file format for cryo-TEM data with fast compression. *J. Struct. Biol.*, doi:10.1016/j.jsb.2017.11.012 (2017).
293 Kahraman, A., Malmstrom, L. & Aebersold, R. Xwalk: computing and visualizing distances in cross-linking experiments. *Bioinformatics* **27**, 2163-2164, doi:10.1093/bioinformatics/btr348 (2011).
294 Barker, P. L., Gendler, P. L. & Rapoport, H. Acylation of dibasic compounds containing amino amidine and aminoguanidine functions. *The Journal of Organic Chemistry* **46**, 2455-2465, doi:10.1021/jo00325a006 (1981).
295 McDaniel, R., Kao, C. M., Hwang, S. J. & Khosla, C. Engineered intermodular and intramodular polyketide synthase fusions. *Chem. Biol.* **4**, 667-674 (1997).
296 Kao, C. M. *et al.* Gain of Function Mutagenesis of the Erythromycin Polyketide Synthase. 2. Engineered Biosynthesis of an Eight-Membered Ring Tetraketide Lactone. *J. Am. Chem. Soc.* **119**, 11339-11340, doi:10.1021/ja972609e (1997).
297 Zheng, J., Piasecki, S. K. & Keatinge-Clay, A. T. Structural studies of an A2-type modular polyketide synthase ketoreductase reveal features controlling alpha-substituent stereochemistry. *ACS Chem. Biol.* **8**, 1964-1971, doi:10.1021/cb400161g (2013).
298 Lowry, B. *et al.* In vitro reconstitution and analysis of the 6-deoxyerythronolide B synthase. *J. Am. Chem. Soc.* **135**, 16809-16812, doi:10.1021/ja409048k (2013).
299 Soehano, I. *et al.* Insights into the programmed ketoreduction of partially reducing polyketide synthases: stereo- and substrate-specificity of the ketoreductase domain. *Org. Biomol. Chem.* **12**, 8542-8549, doi:10.1039/c4ob01777c (2014).
300 Bartesaghi, A. *et al.* 2.2 A resolution cryo-EM structure of beta-galactosidase in complex with a cell-permeant inhibitor. *Science* **348**, 1147-1151, doi:10.1126/science.aab1576 (2015).
301 Marahiel, M. A. A structural model for multimodular NRPS assembly lines. *Nat. Prod. Rep.* **33**, 136-140, doi:10.1039/c5np00082c (2016).
302 Stachelhaus, T., Mootz, H. D. & Marahiel, M. A. The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem. Biol.* **6**, 493-505, doi:10.1016/S1074-5521(99)80082-9 (1999).
303 Tanovic, A., Samel, S. A., Essen, L. O. & Marahiel, M. A. Crystal structure of the termination module of a nonribosomal peptide synthetase. *Science* **321**, 659-663, doi:10.1126/science.1159850 (2008).
304 Drake, E. J. *et al.* Structures of two distinct conformations of holo-non-ribosomal peptide synthetases. *Nature* **529**, 235-238, doi:10.1038/nature16163 (2016).

305 Miller, B. R., Drake, E. J., Shi, C., Aldrich, C. C. & Gulick, A. M. Structures of a Nonribosomal Peptide Synthetase Module Bound to MbtH-like Proteins Support a Highly Dynamic Domain Architecture. *J. Biol. Chem.* **291**, 22559-22571, doi:10.1074/jbc.M116.746297 (2016).
306 Reimer, J. M., Aloise, M. N., Harrison, P. M. & Schmeing, T. M. Synthetic cycle of the initiation module of a formylating nonribosomal peptide synthetase. *Nature* **529**, 239-242, doi:10.1038/nature16503 (2016).
307 Tarry, M. J., Haque, A. S., Bui, K. H. & Schmeing, T. M. X-Ray Crystallography and Electron Microscopy of Cross- and Multi-Module Nonribosomal Peptide Synthetase Proteins Reveal a Flexible Architecture. *Structure* **25**, 783-793 e784, doi:10.1016/j.str.2017.03.014 (2017).
308 Marahiel, M. A. & Essen, L. O. Chapter 13. Nonribosomal peptide synthetases mechanistic and structural aspects of essential domains. *Methods Enzymol.* **458**, 337-351, doi:10.1016/S0076-6879(09)04813-7 (2009).
309 Reger, A. S., Wu, R., Dunaway-Mariano, D. & Gulick, A. M. Structural characterization of a 140 degrees domain movement in the two-step reaction catalyzed by 4-chlorobenzoate:CoA ligase. *Biochemistry* **47**, 8016-8025, doi:10.1021/bi800696y (2008).
310 Mitchell, C. A., Shi, C., Aldrich, C. C. & Gulick, A. M. Structure of PA1221, a Nonribosomal Peptide Synthetase Containing Adenylation and Peptidyl Carrier Protein Domains. *Biochemistry* **51**, 3252-3263, doi:10.1021/bi300112e (2012).
311 Herbst, D. A., Boll, B., Zocher, G., Stehle, T. & Heide, L. Structural basis of the interaction of MbtH-like proteins, putative regulators of nonribosomal peptide biosynthesis, with adenylating enzymes. *J. Biol. Chem.* **288**, 1991-2003, doi:10.1074/jbc.M112.420182 (2013).
312 Conti, E., Stachelhaus, T., Marahiel, M. A. & Brick, P. Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S. *EMBO J.* **16**, 4174-4183 (1997).
313 Campbell, C. D. & Vederas, J. C. Biosynthesis of lovastatin and related metabolites formed by fungal iterative PKS enzymes. *Biopolymers* **93**, 755-763, doi:10.1002/bip.21428 (2010).
314 Taniguchi, Y. *et al.* Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells. *Science* **329**, 533-538, doi:10.1126/science.1188308 (2010).
315 Davison, J. *et al.* Insights into the function of trans-acyl transferase polyketide synthases from the SAXS structure of a complete module. *Chemical Science* **5**, 3081-3095, doi:10.1039/c3sc53511h (2014).
316 Butcher, R. A. *et al.* The identification of bacillaene, the product of the PksX megacomplex in Bacillus subtilis. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 1506-1509, doi:10.1073/pnas.0610503104 (2007).
317 Straight, P. D., Fischbach, M. A., Walsh, C. T., Rudner, D. Z. & Kolter, R. A singular enzymatic megacomplex from Bacillus subtilis. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 305-310, doi:10.1073/pnas.0609073103 (2007).
318 Gay, D. C. *et al.* The LINKS motif zippers trans-acyltransferase polyketide synthase assembly lines into a biosynthetic megacomplex. *J. Struct. Biol.* **193**, 196-205, doi:10.1016/j.jsb.2015.12.011 (2016).
319 Eulalio, A., Behm-Ansmant, I. & Izaurralde, E. P bodies: at the crossroads of post-transcriptional pathways. *Nat. Rev. Mol. Cell Biol.* **8**, 9-22 (2007).
320 Ray-Soni, A., Bellecourt, M. J. & Landick, R. Mechanisms of Bacterial Transcription Termination: All Good Things Must End. *Annu. Rev. Biochem.* **85**, 319-347, doi:10.1146/annurev-biochem-060815-014844 (2016).
321 Santangelo, T. J. & Artsimovitch, I. Termination and antitermination: RNA polymerase runs a stop sign. *Nat. Rev. Microbiol.* **9**, 319-329, doi:10.1038/nrmicro2560 (2011).
322 Burova, E., Hung, S. C., Sagitov, V., Stitt, B. L. & Gottesman, M. E. Escherichia coli NusG protein stimulates transcription elongation rates in vivo and in vitro. *J. Bacteriol.* **177**, 1388-1392 (1995).

323 Goodson, J. R., Klupt, S., Zhang, C., Straight, P. & Winkler, W. C. LoaP is a broadly conserved antiterminator protein that regulates antibiotic gene clusters in Bacillus amyloliquefaciens. *Nat Microbiol* **2**, 17003, doi:10.1038/nmicrobiol.2017.3 (2017).
324 Breaker, R. R. Prospects for riboswitch discovery and analysis. *Mol. Cell* **43**, 867-879, doi:10.1016/j.molcel.2011.08.024 (2011).
325 Jain, M. & Cox, J. S. Interaction between polyketide synthase and transporter suggests coupled synthesis and export of virulence lipid in M. tuberculosis. *PLoS Pathog.* **1**, e2, doi:10.1371/journal.ppat.0010002 (2005).
326 Vavrova, L., Muchova, K. & Barak, I. Comparison of different Bacillus subtilis expression systems. *Res. Microbiol.* **161**, 791-797, doi:10.1016/j.resmic.2010.09.004 (2010).