The Architecture of Polyketide Synthases

Inauguraldissertation

zur

Erlangung der Würde eines Doktors der Philosophie

vorgelegt der

Philosophisch-Naturwissenschaftlichen Fakultät

der Universität Basel

von

Yves Ulrich Tittes

Basel, 2023

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät

auf Antrag von

Prof. Dr. Maier, Timm

Prof. Dr. Hiller, Sebastian

Dr. Hartmann, Marcus

Basel, den 21.2.2023

Prof. Dr. Mayor, Marcel
Dekan

Abstract

Since the discovery of penicillin over a century ago, secondary metabolites from all kingdoms of life have proven to be of high medical value. One class of proteins prevalent in the production of secondary metabolites are polyketide synthases (PKSs). Their polyketide products are complex organic compounds based on carbon chains assembled from carboxylic acid precursors. Many polyketides are produced by their hosts with the primary purpose of gaining an advantage in their ecological niche. To contribute to such an advantage, a significant proportion of polyketides are active against pro- and eukaryotic microorganisms. Type I PKSs are giant multienzyme proteins employing an assembly line logic for the synthesis of the most complex polyketides. They are composed of one or more functional and structural modules, each capable of carrying out one step of precursor elongation during the formation of an extended polyketide product.

In this thesis, I address two fundamental and open questions in the biosynthesis of polyketides: First, what is the unique architecture underlying the assembly line logic of multimodular PKS assembly lines; and second, how is atomic accuracy achieved in cyclization and aromatic ring formation in the final step of PKS action.

The first aim is addressed in chapter two, which provides for the first time detailed structural insights into the organization of type I PKS multimodules. This is achieved by cryo-electron microscopic analysis of filamentous and non-filamentous forms of K3DAK4, a bimodular trans-acyltransferase (AT) PKS fragment from *Brevibacillus brevis*. Overall reconstructions are provided at an intermediate resolution of 7 Å, with detailed insights into individual domains at sub-3Å resolution from cryo-electron microscopy and X-ray crystallography. The bimodule core displays a vertical stacking of its two modules along the central dimer axis of all three enzymatic domains involved. Additionally, K3DAK4 oligomerizes into filaments horizontally via small scaffolding domains in a trans-AT PKS-specific manner.

In chapter three the second aim is tackled, as I visualize an intermediate of the enigmatic targeted cyclization and aromatic ring formation in the product template domain (PT) of the aflatoxin-producing PksA at 2.7 Å resolution using X-ray crystallography. To this end a substrate-analogue mimicking the transient intermediate after the first of two cyclization steps facilitated by the enzyme is covalently crosslinked to the active site. The positioning of the ligand relative to previously known ligands representing the pre-and post-cyclization states indicate an outward movement of the substrate throughout the process and a substantial effect of progressing cyclization on the meticulous positioning of the intermediates.

The work provides detailed insights into core aspects of PKS biology from the atomistic picture of guided product modification to the giant overall assembly line architecture. In chapter four, both of these levels are put into context with current advances in the analysis of modular structure and dynamics of PKSs, such as recent structural models of cis-AT PKS modules and iterative PKSs. Furthermore, it addresses currently open questions, such as the interaction of trans-AT PKS with their cognate trans-acting enzymes. Altogether, the current progress in mechanistic understanding of PKS systems makes systematic and structure-guided efforts to unleash the full potential of PKS bioengineering ever more achievable.

Table of Contents

Introduction

Proteins

Proteins are the functional units of life. All known life relies on them as the basis for most of their functionality. Cells consist of a membrane, which encapsulates them and forms a barrier between the inside of the cell and its surroundings. Inside are intricate networks of different proteins with vastly different functions from simple scaffolding via enabling the uptake of nutrients from their surroundings to even the production of these same proteins [1]. Proteins are made up of amino acids (AAs). Each protein is a chain of covalently linked AAs arranged in a genetically encoded sequence. 20 different types of AAs are commonly found in most organisms. AAs themselves are commonly divided into their backbone, which is the same across all AAs, and their sidechain, which varies. In a protein the AAs backbones are covalently linked from the preceding AAs C-terminus to the succeeding ones N-terminus. Therefore, proteins have one AA at each end of the chain with a remaining N-terminus, at the start, and a remaining C-terminus at the end. This also gives the chain a directionality, which aligns with their synthesis [1]. Proteins follow a hierarchical organization. The first layer, namely the order of AAs in a protein is commonly referred to as its primary structure. Parts of the AA chain may form interactions within the chain with neighboring AAs backbones, forming arrangements such as alpha-helices or beta-strands. These highly common ways of folding of a protein chain are referred to as secondary structure elements. Secondary structure elements may combine again to form super secondary structures which combine multiple secondary structure elements. Those include more complex formations of the protein chain, typically 10-40 AAs in length, such as helix-bundles or hot-dog folds. The overall fold of a protein is known as its tertiary structure, typically containing both alpha-helices and beta-strands [1] as well as forming much larger structures like beta-barrels. This tertiary structure is crucial to enable the function of individual proteins, as well as being unique to each protein, although certain super secondary motifs may reoccur between different proteins. Lastly, multiple proteins may come together to form a complex, an ensemble of various proteins that are not necessarily covalently bound but are essential to its function, or serve a specific purpose in *i.e.* regulation of the proteins activity. Assembly into multiprotein complexes is referred to as quaternary structure. Given the relevance to function, protein folding and complex assembly are closely regulated in any living organism [1], and failure in systems supervising folding is a known cause of a variety of diseases, such as cystic fibrosis [2].

All organisms contain desoxyribonucleic acids (DNA), which contains the complete genetic information, termed its genotype. Not all information contained in the genotype is expressed and therefore displayed by a given organism at all times. Its observable characteristics are termed the phenotype, which can vary between organisms with identical genotypes [1]. Variation in phenotype can be influenced by for example the organism's environment or its current metabolic state. Processes known as transcription, a process wherein parts of an organism's DNA is copied to the more short lived ribonucleic acid (RNA), and translation, where the RNA is used as a blueprint to produce proteins, transform this information into the essential building blocks of each organism. DNA forms a double-stranded helix, where each strand consists of polynucleotides of just four molecules: adenine (A), guanine (G), thymine

(T) and cytosine (C). In the DNA double helix, the two strands interact with each other via always linking A on one strand to T on the other and the same for C and G [1]. This linking of both the strands ensures that the information on both strands is always the same – albeit reversed. For that reason, we call one strand the sense strand and the other the antisense strand. During transcription the double strand is partially opened, its contents are read out and assembled on a new strand. This new strand is called RNA and has three distinct differences to DNA. First it is single stranded, secondly it exchanges T for Uracil (U), and thirdly it is built from nucleotides instead of deoxynucleotides [1]. As it is a direct readout of a specific piece of DNA it contains the exact same information. During translation the ribosome, a complex consisting of both RNA and protein molecules, reads through the translated RNA strand. The readout is spaced in triplets of RNA bases termed a codon. Each codon is matched to a reverse matching piece, termed the anticodon, of tRNA with the corresponding AA attached. If codon and anticodon match the growing protein chain is extended by the attached AA and the process repeats until a so-called stop codon interrupts the chain, upon which the finished protein is released. Stop codons are base triplets whose anticodon RNA does not have an AA attached leading to an interruption of the polypeptide chain. Reading through the RNA this way allows for a defined sequence of AAs to be produced.

Enzymes

While proteins fulfill a variety of functions, one of the best studied class of proteins are enzymes. Enzymes are protein catalysts that enable chemical reactions to occur under the narrow aqueous condition of cellular life that would otherwise not progress, or very slowly only. As catalysts their main action is the lowering of activation energy for a given reaction or to change the kinetics of a reaction. Six principal classes of enzymes exist: oxidoreductases, transferases, hydrolases, lyases, isomerases and ligases [1,3]. Known functions include the production, modification and regulation of DNA and proteins, as well as the production of both DNA and protein building blocks.

Many enzymes are highly specialized, and only act on a specific substrate [3]. To achieve their catalytic action the enzyme contains an active site, which catalyzes one reaction. Typically, enzymatic domains only contain one active site, where the entry of substrate and the release of products is diffusion based. The active site itself is commonly hidden inside the protein within a tunnel or cavity to allow for a highly controlled environment and spatial positioning of key residues, termed the active site residues. Given that most enzymes only catalyze one reaction, it is common to find multiple enzymes work in concert to achieve a synthesis of a final product[1]. To facilitate the concerted action of multiple enzymes two concepts are common. Either the involved enzymes form a complex, with interfaces between the individual proteins dictating positioning as well as strength of the interaction, or multiple enzymes are encoded on a single protein, therefore tethering the different enzymatic functions to one another in different domains of a continuous protein chain. The latter ones are termed multienzymes (Fig 1.1).

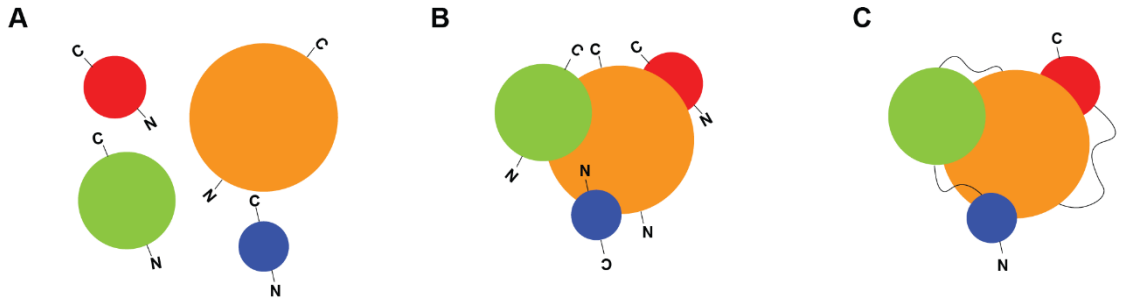


Figure 1.1: Interaction of proteins in solution. *N-and C depict each proteins termini.* ***(A)*** *Individual proteins, with no direct contact between one another.* ***(B)*** *A protein complex, consisting of four proteins interacting with one another.* ***(C)*** *A single multidomain protein with each of its domains interacting. If this protein contains enzymatic domains, it qualifies as a multienzyme.*

Biosynthetic enzymes are involved in the production of many components required for cellular function, such as AAs, proteins [4], nucleic acids or fatty acids [1,5]. Biosynthetic enzymes can fall under all three categories. The production of complex chemical molecules often requires multiple concerted steps to occur. Depending on the system they therefore rely on diffusion in the case of individual domains, while both the complex and the multienzyme keep the enzymatic domains in close proximity, potentially allowing for higher efficiency in product synthesis. Additionally, some biosynthetic enzymes produce chemicals which are not essential but typically yield an advantage to the producing organism in its ecological niche [6]. Such chemicals are termed secondary metabolites. This also means such secondary metabolites are not always produced, but their synthesis is highly regulated to only become activated when it is beneficial to the organism [7-9]. Many secondary metabolites inhibit the growth or critical cellular functions of competing organisms in their ecological niche [10-14]. They are being studied for decades now for their potential in medicine with use cases ranging from antibiotics, such as erythromycin [12], to anti-cancer drugs, *i.e.* epothilone B [15]. Among the enzymes producing such secondary metabolites are polyketide synthases (PKSs) and non-ribosomal peptide synthetases (NRPSs) [15]. Despite some differences, both follow the common concept of polymerizing either simple carboxylic (PKS) [16] or AA (NRPS)[17] building blocks into a final product [18]. In fact, the two systems are known to co-exist and occasionally also interact with one another to form a product containing both a polyketide and a polypeptide part [18,19].

Polyketide synthases

One class of biosynthetic enzymes producing a vast array of secondary metabolites are the PKSs [16]. They exist in a large array of organisms, ranging from bacteria over fungi to plants [20-23]. PKSs produce their products in a stepwise fashion where in each step the growing substrate is first extended by an acyl-moiety and subsequently modified in a distinct pattern[16]. Repeating this pattern for a programmed number of times leads to a highly specific product being synthesized reliably.

Overall, there are three classes of PKSs following this basic principle. Type I PKSs, occurring mainly in fungi and bacteria, are big multienzymes which constitute all the enzymes for at least one step of polyketide chain extension on one protein[23,24]. Type II PKSs, prevalent in bacteria, have each enzyme as a standalone protein, typically as homo- or pseudo dimers[22]. Lastly, type III PKSs are commonly found in plants

and to a lesser extent in bacteria [21]. They employ individual standalone enzymes[25] as well. Despite the differences in the organization of the biosynthetic enzymes, the basic pattern of stepwise polyketide extension and optional modification holds true for all three classes. Additionally, the enzymatic domains involved are highly similar both in fold and active sites involved[16].

PKS enzymatic domains can be grouped into those involved in condensation, and therefore chain extension and those involved in the optional modification of the chain. All of the individual enzymatic domains are well studied both on a structural and a chemical level[16,26-31] (Fig. 1.2).

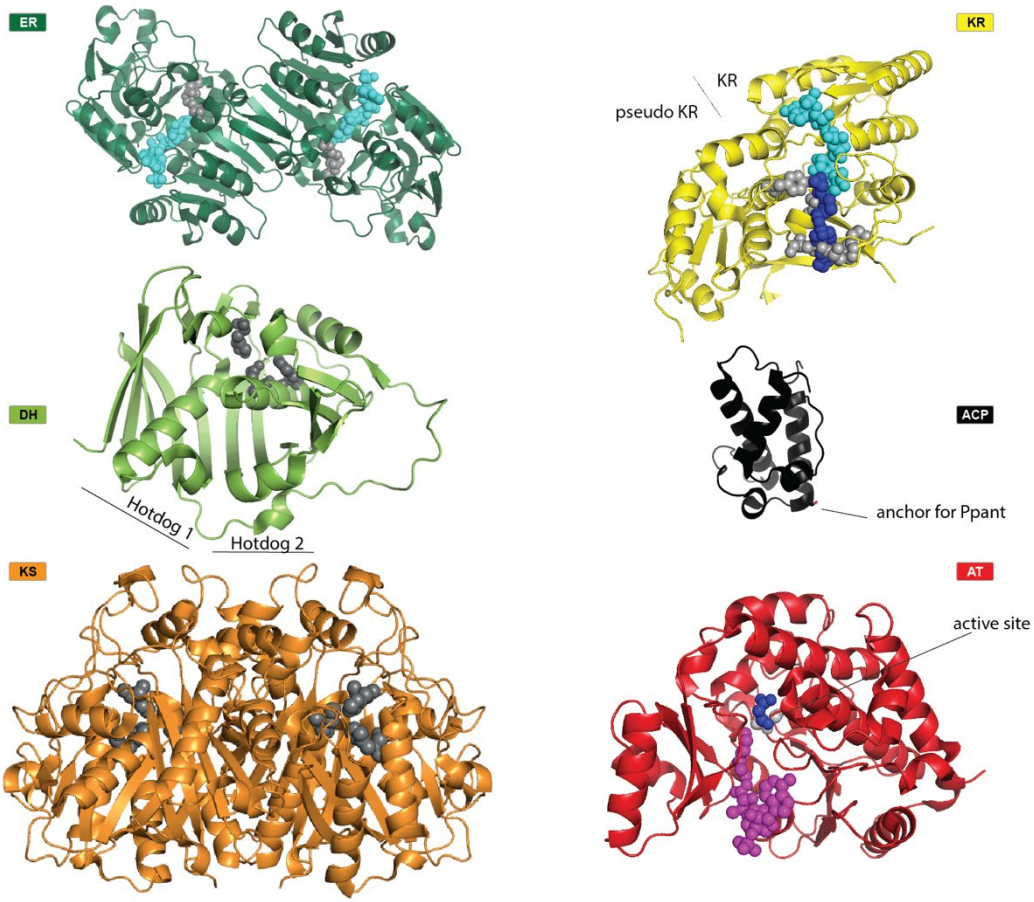


Figure 1.2: Core PKS and FAS enzymes, with their active site and cofactors displayed. *Active site residues displayed by balls for all domains in grey, with additional indication in the AT, due to low visibility. NADP displayed in cyan. Substrate in the AT and KR displayed in blue. The AT has a phosphopantetheine arm reaching into the active site displayed in purple. The DHs double hot dog fold relates to its two central helices surrounded by beta sheets, as indicated below the structure. The KRs split between the main KR and its scaffolding part is indicated. In addition to the here shown dimeric ER and KS, the DH can also dimerize. Depending on the system the KR can oligomerize, especially when existing as individual domains. The ACP and AT are not known to oligomerize. All active sites are in an active site cavity, which the phosphopantetheine arm on the ACP can access, as exemplified for the AT. All depictions made in pymol using PDB IDs 2VZ9 (ER, DH), 4PIV (KR), 3HDD (KS), 5MY0 (AT), 5HVC (ACP).*

The enzyme responsible for the Claisen-condensation which extends the growing polyketide chain is the ketosynthase (KS). KSs adopt a conserved thiolase fold, with their active site consisting of one cysteine and two histidines[32]. As with most other PKS enzymatic domains, their similarity to fatty acid synthase (FAS) KSs is striking albeit not surprising given the chemical closeness of substrates. As they are the core

enzyme responsible for polyketide chain extension, they occur in virtually all PKSs. Differences in substrate selectivity are mainly attributable to the size and makeup of their active site tunnels[33]. It homodimerizes with each protomer featuring an active site chamber in which it extends the growing polyketide. During biosynthesis the growing substrate is covalently bound to the KSs active site cysteine via a thioester bond, with the substrate being transferred back to the acyl-carrier-protein (ACP) in a Claisen-condensation step, which extends the chain by a ketoacyl-moiety[16].

The KS is supplied with substrates by the ACP the only non-enzymatic protein domain involved in polyketide synthesis. The ACP is a small domain (~15 kDa) consisting of a highly conserved four helix bundle [34]. On helix two it contains a conserved DSL (aspartic acid-serine-leucine) motif onto which a phosphopantetheine transferase transfers a phosphopantetheine arm from Coenzyme A (CoA) to the serine of the DSL motif forming a covalently bound phosphoester[35]. The phosphopantetheine arm contains a terminal thiol group which is essential in the ACPs function to carry all carboxylic acid PKS substrates covalently tethered to the phosphopantetheine arm as thioesters[16].

The other enzyme belonging to the condensing part of the PKS is the acyl-transferase (AT). The AT relies on a key serine active site residue. While its exact substrate varies depending on the PKS, it generally transfers carboxylic acid derivatives as building blocks from a CoA onto the ACPs phosphopantetheine arm [36]. Common building blocks include malonyl-CoA or acetyl-CoA[16]. The ACP transfers this building block to the KS, which extends the substrate via the Claisen-condensation and back-transfers its product onto the ACP. While the ACP is getting loaded by the AT with a new building block, the KS has the substrate covalently bound to its active site cysteine via a thioester. The extension process is repeated until the programmed polyketide length is achieved, giving PKSs their distinct stepwise synthesis. The programming of chain length is nuanced depending on the individual PKS but generally co-depends on each KSs limited active site cavity which restricts the maximal substrate size and form[37].

Additionally, in each step a series of optional modifications may occur to the acetyl moiety. These are carried out by the modifying enzymes, which typically consist of a keto-reductase (KR) reducing the carbonyl group to a hydroxy group [29], a double hot-dog fold domain (DH), typically functioning as a dehydratase [28] catalyzing a water elimination to form a double bond and finally an enoyl-reductase (ER), which reduces the double bond to a fully saturated carbon-carbon bond [16,38]. It is important to keep in mind that the modifying enzymes are optional, and even if present in a PKS, don't necessarily act at each step. The resulting product therefore contains differentially modified acyl-moieties. This variation in modifications allows for a larger chemical space that is accessible to PKSs. Interestingly, despite the accessible chemical space, individual PKSs are highly specific and typically only synthesize a single product. Their versatility is therefore a consequence of the large family of PKSs producing a wide variety of secondary metabolites.

On top of these commonly found enzymatic functions, a number of less common modifications can occur. Some require distinct enzymes such as methyl-transferases (MTs) [39], whereas others replace *i.e.* the DH with an isomerase[40]. In the case of the DH there are a number of other enzymatic functions known for proteins of the same

fold[41]. Another way of diversifying the product is the incorporation of different building blocks[42], such as methyl-malonyl-CoA[43]. Typically, these specialized modifications are specific to a certain class of PKSs.

The last common enzyme involved in polyketide synthesis is the thioesterase (TE). It performs the last step in the synthesis, releasing the final product from the ACP by cleaving the thioester bond to the phosphopantetheine arm[16]. Its enzymatic actions may be limited to simple cleavage or involve a cyclization step[44].

It is important to note, that while the core PKS enzymes discussed above are responsible for chain extension and optional modification, the final product after which many PKSs are named often requires additional tailoring enzymes acting on the PKSs product[45-48]. Tailoring steps may include methylations or the addition of sugar moieties[16].

Type I PKS

The multienzymatic type I class of PKSs form homodimers[23]. Compared to type II or type III PKSs their individual proteins are much larger as a consequence of their multienzyme nature. Thus type I PKSs contain all required enzymes for one step of polyketide chain extension and modification on a single protein, as opposed to the individual enzymatic domains of type II and type III PKSs[49]. A PKS protein containing all the condensing and modifying domains required for or involved in one step of chain extension and modification is termed a module. A single type I PKS protein may contain multiple modules[50]. This, together with their homodimeric nature, makes the individual type I PKS proteins very large (>200 kDA), especially compared to the mean size of proteins in their bacterial hosts [51]. The whole type I PKS may even extend over several multimodule proteins, which interact via non-covalent linker domains[52,53].

Type I PKSs are sub classified around their use of modules throughout polyketide synthesis. There are two main classes of type I PKSs:

(1) Iterative PKSs (iPKSs) consist of a single module, which is utilized repeatedly to obtain the product. Despite only having a single module the modification pattern is not necessarily consistent through each iteration of condensation and modification, meaning not every modifying domain will act in each step[20]. Additionally, final chain length depends on both the initial substrate as well as the number of chain extension, neither of which is directly apparent from just the module architecture. It is therefore nontrivial to predict iPKSs products.

(2) Modular PKSs (modPKSs) on the other hand use each module once during product synthesis[18,54]. This, in combination with the fact that each module makeup typically correlates with its synthetic function and is encoded on the genome in the same order, makes the final product more estimable based on the makeup of the modPKSs modules from a genome study of a given organism.

The linear correlation between operon, modules and final product in modPKSs is termed the rule of collinearity[49]. It was first discovered for the best studied of modPKSs, 6-deoxyerythronolide B synthase (DEBS). DEBS produces the precursor for the antibiotic erythromycin[55] and was the first to be shown to have a clear link between its operon and the protein, in that it has three proteins with two modules each which act sequentially to achieve product synthesis[56]. The rule of collinearity

has since been shown to have exceptions[57,58] but still holds true for many modPKSs[59].

PKS modules differ in makeup depending on their product

Modules from different type I PKS can look quite differently, but generally follow the same logic[60]. The first part of the protein contains the KS followed by the AT, after which the modifying domains, if present, appear typically in the order DH-ER-KR. Lastly the modules ACP is located in the sequence after the modifying domains. If the module is the last in its assembly line, it might contain the TE after the ACP. This overall order and optionality of the modifying domains allows for individual modules to have a specific purpose in their enzymatic activity to achieve the final PKS product. While the order of the individual enzymatic domains is set, their presence is not, making PKS modules highly versatile (Fig. 1.3).

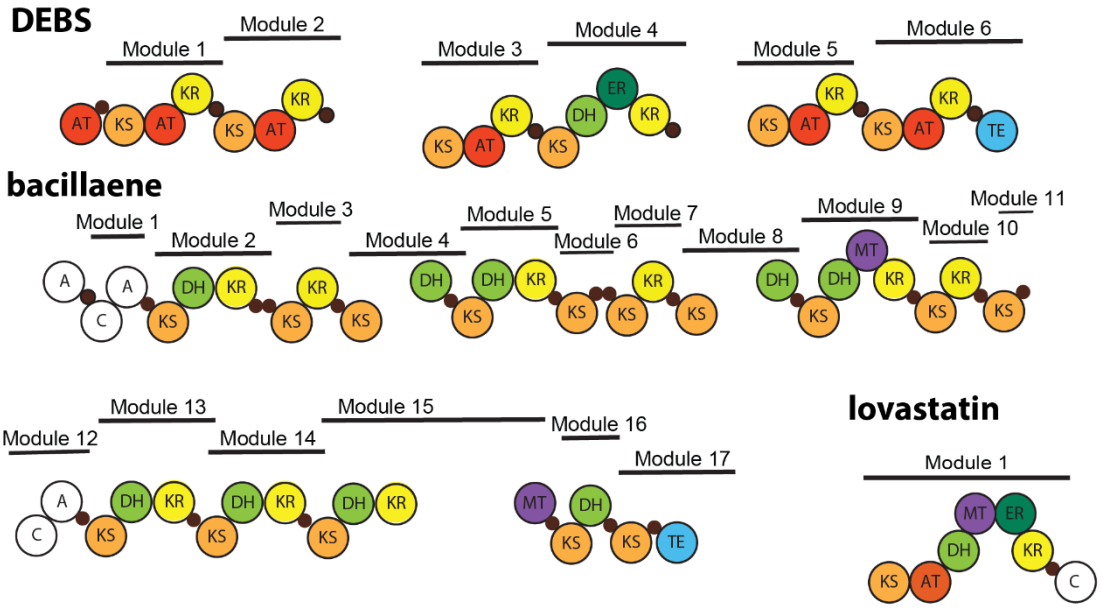


Figure 1.3: Modules from different classes of type I PKSs. *Despite large differences between individual modules, the key logic of KS-X-ACP holds true for each and every module. All PKS domains are color coded, ACPs are depicted as small black circles. Domains with white background belong to NRPSs. DEBS and the bacillaene are modPKS forming assembly lines, whereas the lovastatin PKS is an iPKS using only a single module.*

modPKS: Cis-AT PKSs

ModPKSs following the classical module organization are termed cis-AT PKSs[53]. The most well studied PKS, DEBS, is part of his class of PKS. It consists of 6 modules, divided into 3 proteins with two modules each [61,62]. Both information on the architecture of individual modules as well as the individual enzymatic action are well understood[50,62]. Another example of a well-known cis-AT PKS is the rapamycin producing PKS [63]. While following the same pattern it is much larger with a total of 14 modules involved in product synthesis. The higher order interactions, meaning how multiple modules interact, especially when factoring in inter-protein interactions, are however not understood. Many of the hypotheses of how modPKSs organize are based on DEBS and range from helical arrangements[64] to a serpentine stacking of the three proteins[50,62]. It is known that substrate hand over from one protein to the next requires a N-or C-terminal docking domain in terminal modules, allowing the two modules on different proteins to interface and closing the distance for the ACP to

hand over the substrate to the downstream KS[65]. This is especially crucial as many cis-AT PKS, such as the rapamycin PKS are distinctively larger than DEBS and can contain 12 modules or more, with typically two to four modules encoded in one protein. As the rapamycin PKS proves, more modules (six in the case of rapB) on a single protein are still possible. The need for modules to interact across proteins is therefore crucial for efficient and precise polyketide synthesis. As most cis-AT PKSs follow the rule of collinearity, and contain all the enzymes for each step on the respective module they are a prime target for product predictions.

modPKS: trans-AT PKSs

The second class of modPKSs are called trans-AT PKS[66]. The name refers to them breaking the classical module architecture in that both the AT and the ER are not part of each module's protein in this class of PKS. If both are present, given the optional nature of the ER, they are expressed in trans, often together as one protein. Despite this specialty trans-AT PKS do still loosely follow the rule of collinearity, however prediction of product is more difficult especially as ER modifications are hard to foresee. Typically, trans-AT PKS have one to two ATs for the whole cluster, meaning each module is loaded with the same building blocks by the same AT, as opposed to each module containing its distinct AT in cis-AT PKSs[66]. The outsourcing of domains from both the condensing and modifying regions makes trans-AT modules significantly smaller compared to those of cis-AT PKSs. This is exemplified by the trans-AT PKS module RhiE[41]. It contains both a KS and a DH. Although small, it shows that despite the lacking domains trans-AT PKSs follow the same general layout for the module architecture as other type I PKSs. Just like cis-AT PKS they rely on docking domains to hand over substrate across protein boundaries. Although the concept is the same the architecture of cis-and trans-AT docking domains is markedly different[67]. Well known PKSs of the trans-AT class include the bacillaene PKS[68] and the rhizoxin PKS[69]. Both also exemplify another specialty found more commonly in trans-AT PKS, in that they are hybrid biosynthetic gene cluster (BGC) including an NRPS part each. Phylogenetic studies suggest that trans-AT PKSs evolved distinctly from cis-AT PKSs. Specifically, cis-AT PKSs appear to rely more on module duplication events followed by divergence, whereas trans-AT PKS genes more often show signs of (multi-)module acquisition via horizontal gene transfer[60].

iPKS: HR-iPKSs

iPKSs that incorporate all modifying enzymes are termed highly-reducing iPKSs (HR-iPKSs) [70]. As their name suggests, they typically reduce much of their substrates making them akin to FAS in both using highly alike enzymes and incorporating similar building blocks. Contrary to FAS however, they do not fully reduce their substrate in each step, instead relying on a programmed pattern of modification in each step. The exact mechanisms of this programming are not well understood to date, precluding both prediction and engineering efforts on HR-iPKSs. Examples of HR-iPKSs include the lovastatin [39] and equisetin [71] PKSs. Both have a minimal NRPS part attached. They also exemplify a module deviation typical to HR-iPKSs, in that their modules include an MT, another feature similar to FAS. Interestingly, they do not commonly contain a TE, making their substrate release condition a point of interest [20]. Another peculiarity in their module makeup is the fact that, while they typically do contain an ER, they may have an additional ER present in trans. Earlier experiments showed, that despite the presence of the ER in the main

module it may only fulfill a scaffolding role, as it was found incompetent to perform the required ER steps in lovastatin biosynthesis [39].

iPKS: NR-iPKSs

The other class of iPKSs are the non-reducing iPKSs (NR-iPKSs) [20]. They do not contain any of the canonical modifying enzymes, instead often using an altered DH which functions as a cyclization domain. This domain is termed the product template domain (PT) [44]. In addition to the typical module architecture NR-iPKSs often have an additional AT, termed the starter acyl transferase (SAT) [72]. It is responsible for loading the first building block onto the ACP, which can deviate largely from typical building blocks, such as incorporating hexanoyl-CoA in the aflatoxin PKS [73]. Another example of an NR-iPKS is the cercosporin [74] PKS. While featuring a highly similar module architecture a difference is the presence of two ACPs in the cercosporin producing PKS as opposed to the aflatoxin PKS. Given their non-reducing nature and the high reactivity of the poly-beta-ketones they produce almost all NR-iPKS products are cyclized, often producing aromatic compounds. In this context the incorporation of the NR-iPKSs special domain, the PT is crucial, in that it ensures defined cyclization patterns [73]. Additionally, NR-iPKSs typically do contain a TE, which releases the product. Instead of a simple cleavage reaction, it is often involved in a final cyclization step [73].

iPKS: PR-iPKSs

There is an additional sub-class between HR- and NR- iPKSs, termed the partially reducing iPKSs (PR-iPKSs) [75]. PR-iPKSs follow a similar structure to HR-iPKSs, with no TE and no SAT domain. Typically, they do not contain an ER or DH, instead having a thioester hydrolase (TH) as a specialized chain release enzyme. Their overall module architecture commonly is KS-AT-TH-KR-ACP. The best studied examples of PR-iPKSs include 6-methylsylicylic acid synthase (6-MSAS) and the 2-hydroxyl-5-methylnaphthoic acid PKS [75]. Despite 6-MSAS being among the first fungal PKS to be purified in 1970 [76], PR-iPKSs are the least well understood class of iPKSs overall, as relatively few examples are known. Many PR-iPKSs produce aromatic compounds typically containing two rings as exemplified by the mellein biosynthesis[77].

NRPS

NRPSs are a completely distinct group of secondary metabolite producers from PKSs. As with PKSs, NRPSs are typically not constitutively active and their products are also of interest in medicine, for much the same reasons, including antibacterial (bacitracin) and antitumor (actinomycin D) agents [17]. They produce polypeptide chains instead of polyketide chains and use a completely different set of enzymes. Despite these differences there is a certain overlap. As mentioned before, PKS and NRPS can exist in hybrid clusters. The reason goes back to similarities between the two system. NRPSs follow a step-wise chain elongation and modification pattern, homologous to that of PKSs. Their growing polypeptide chain is linked via a thioester to a peptidyl-carrier-protein (PCP). Like the ACP the PCP is loaded with a phosphopantetheine arm unto which the growing substrate is tethered. The adenylation domain (A), is the analogue of the AT and loads amino acids onto the PCP. The condensation domain (C) performs a similar role to the KS in that it extends the growing polypeptide chain. Its mechanism is however markedly different,

and less well understood than that of the KS. As in PKSs, the substrate can be optionally modified by *i.e.* epimerization (E), formylation (F), methylation (M), heterocyclization (Cy), reduction (R) and oxidation (Ox) domains [17]. Like type I PKSs, NRPSs are divided into modules, with each module involved in one step of chain extension and modification [78]. This similarity to modPKSs means the rule of collinearity does apply to NRPSs. Unlike ribosomes NRPSs can incorporate non proteinogenic amino acids, giving them access to a vastly expanded chemical space [18]. Like PKSs, NRPSs are therefore a very versatile family of secondary metabolite producers. Well known examples of NRPS products include mycobactin, various cyanobactins as well as penicillin [79]. The use of a thioester tether in both PKS and NRPS allows for the interconnectivity of the two systems as is evident in a number of hybrid clusters, such as the bacillaene BGC [68]. This means a NRPS product can be handed over to and extended upon in a PKS and vice versa [18].

The structural biology of type I PKSs

All the key enzymatic domains of PKSs are individually well understood with a multitude of structures from a variety of systems existing for each of them [16,27]. While minor differences appear between systems, generally a KS from a cis-AT PKS will still have the same fold as that of *i.e.* a NR-iPKS. Open questions in the field include the arrangement of domains within the modules of type I PKS [49], substrate specificity of key domains, the mechanism of specialized enzymatic domains, the mechanism of substrate shuttling by the ACP within the context of a module or the inherent flexibility within PKS systems. In recent years several structures of complete type I PKS modules have emerged, shedding a light on their quaternary structure.

Specialized domains for programmed synthesis

The PT is a specialized enzymatic domain exclusive to NR-iPKSs, belonging to the DH fold family, occurring in both fungal and the bacterial NR-iPKSs. It guides the cyclization of the NR-iPKSs poly-beta-ketone products. It structurally resembles a DH with two additional alpha helices along the edge of the double hot-dog fold (Fig. 1.4) [44] [73]. While also dimeric, its dimerization interface differs drastically from that of any other DH in that the PT dimerizes via the end opposite to the N-terminus as opposed to the N-terminal interactions known for many cis- and trans-AT DHs. The active site, while positioned in the same place as that of more common DHs is not fully understood. Several PTs are known to guide a two-step cyclization reaction resulting in cyclization patters at C7-C12, C9-C14, C4-C9, C2-C7 or C6-C11 followed by a second cyclization resulting in a double aromatic ring formation [44]. Despite this biochemical knowledge the mechanistic basis for the divergence of cyclization patterns across different PTs is not understood.

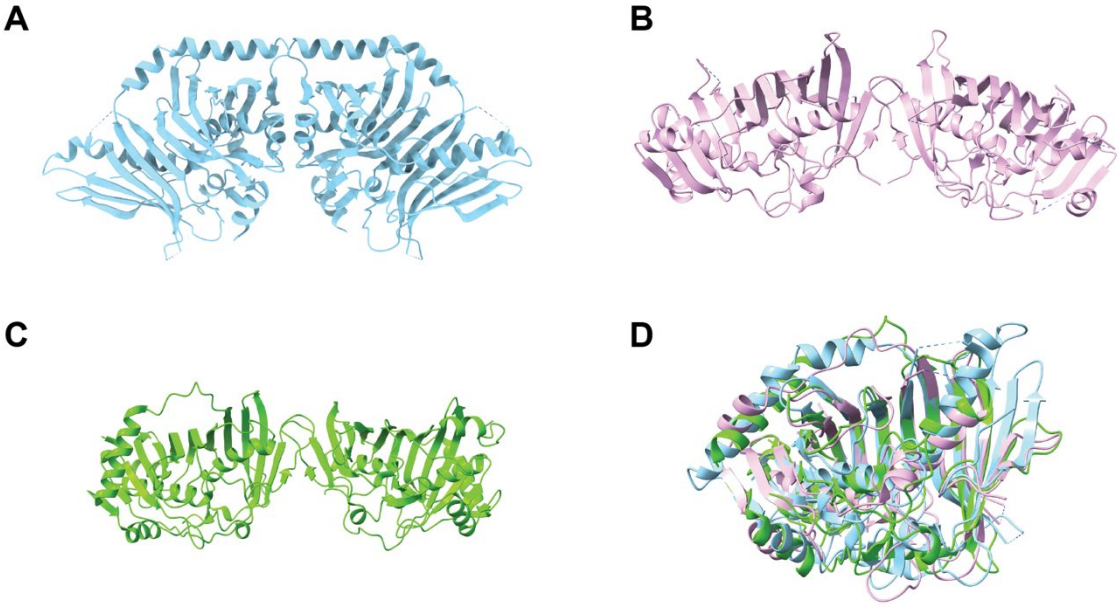


Figure 1.4: Comparison of a PT (blue) and two DHs (green, pink) dimers. (A) *PT dimer from PksA (PDB ID 3HRR).* ***(B)*** *DH dimer from an orphan PKS in Brevibacillus brevis (PDB ID 5HU7).* ***(C)*** *DH dimer from RhiE (PDB ID 4KC5)* ***(D)*** *Overlay of the protomers of all three. The RSMD (C-alpha) between the protomers of the three is below 1.3 Å, despite their different enzymatic activity, highlighting the conservation of the DH fold.*

The architecture of individual PKS modules

Due to their large degree of flexibility, entire PKS modules have been elusive targets for structural studies. Earlier studies focused on individual regions, as they typically constitute a more rigid and compact complex, as exemplified by the NR-iPKS CTB1s condensing region [72] or that of the cis-AT modPKS DEBS module 5 [80]. First steps towards modular organization were made using small angle X-ray scattering (SAXS) to obtain low resolution, but in solution insights into type I PKS modules. Efforts to use both crystallography and cryo-electron microscopy (cryo-EM) were successful in pushing the limits of resolution allowing for detailed insights into the architecture of type I PKS modules.

Low resolution solution studies of cis- and trans-AT modPKSs

A glimpse at the module architecture of cis-AT PKSs was afforded in 2014 by the Khosla Lab [50]. Specifically, they resolved DEBS module 3 using SAXS, followed by modelling high resolution X-ray structures of the contained domains into the SAXS maps. DEBS module 3 consists of KS-AT-KR-ACP. In addition, they engineered a modified DEBS module 3 with the TE added at its C-terminus. Interestingly, the modified DEBS module 3 was shown to be capable of iterative action in vitro [81]. The model displays an extended conformation for module 3 with the TE fused, in which the KS forms a dimer, with each protomer being flanked by its respective AT. Atop each of the KSs sit one of the KRs, without direct contact between the KR protomers. The KR in turn interfaces with the TE which homodimerizes on top of the KS following an overall P2 symmetry. Evaluating the models obtained by the study, a large discrepancy in relative positioning of the KR is observable. In the majority it is placed directly adjacent to the linker domain (LD) of the KS, at an angle of 50°

relative to the KS-AT dimer, but when taking the lesser populated models into account the entire modifying region rotates by up to 90 ° relative to the condensing region. It is important to remember that, while useful, SAXS does not afford high resolution data. The indication of rotational freedom does however fit expected behavior, mainly derived from the porcine FAS structure (Figure 1.5A) [82]. Porcine FAS is highly similar to type I PKSs in domain architecture and biochemistry [83]. It produces a similar product via the same step-wise extension process making the comparison relevant.

Also in 2014, a minimalistic trans-AT PKS module was investigated using SAXS. VirA module 5 consists of the domain architecture KS-ACP-ACP-dimerization motif and is derived from the virginiamycin NRPS-PKS hybrid cluster from *Streptomyces virginiae* [84]. The core KS of the module was modeled via homology to the KS of DEBS module 5 with a sequence identity of 40%. Overall, the model for the full module displays the KS dimerizing at the center with both of the ACPs reaching outward and upward towards the next module. The authors hypothesized that this clamping motion may be required for interaction with the downstream module. The presence of two ACPs and their relative positioning is accredited to their potential interactions with trans-acting enzymes. Additionally, the LD is slightly rotated by 18-21° as compared to the previously known KS-AT from DEBS or RhiE. As the module does not contain a modifying region and stems from a trans-AT PKS, the only region that does not display inherent flexibility is the KS domain. While the rotation of the LD is an interesting note, at the given resolution and with no interactions observed it might just rotate in solution as the LD is connected to the KS via only two linkers. The authors report density for both ACPs, placing the first of the two directly next to the KS, potentially interfacing with the active site on each protomer and the second one free floating. It is however important to note that long linkers between both the KS-ACP and ACP-ACP allow for many more positions, as the ACP needs to be able to reach all catalytic domains, in this specific case the modules own KS and the succeeding modules KS, as well as at least the trans-acting AT.

The organization of trans-AT PKS modules

In 2013 Hertweck *et al* determined the crystal structure of a trans-AT PKS module core with the architecture KS-DH, from the rhizoxin PKS [41], specifically the last module of RhiE. While minimalistic, the modules X-ray crystallographic structure of 2.1 Å displayed both domains as individually dimerizing, with the DH dimer placed directly on top of the KS dimer (Figure 1.5B). The DH domain of RhiE is a specialized domain which is involved in vinylogous chain branching, and another example of the diverse functionality within the DH fold family of domains. This is underlined by the fact that the canonical dehydratase function would not be able to act without a KR on the keto-acyl-moiety added by the KS. In the crystal the DH dimer is placed at a 90° angle relative to the KS. When comparing the model to the previously known porcine FAS, and KS-AT subregion from DEBS, the overall architecture is consistent when factoring in the systematic differences, such as the lack of an AT in RhiE. Despite its minimalistic nature, RhiE contains both a condensing and a modifying region. The lack of an ACP does not allow for any comments on substrate shuffling, while the method of crystallography locks the module core into a single conformation, where questions regarding degrees of freedom between the two regions remain inherently unanswered.

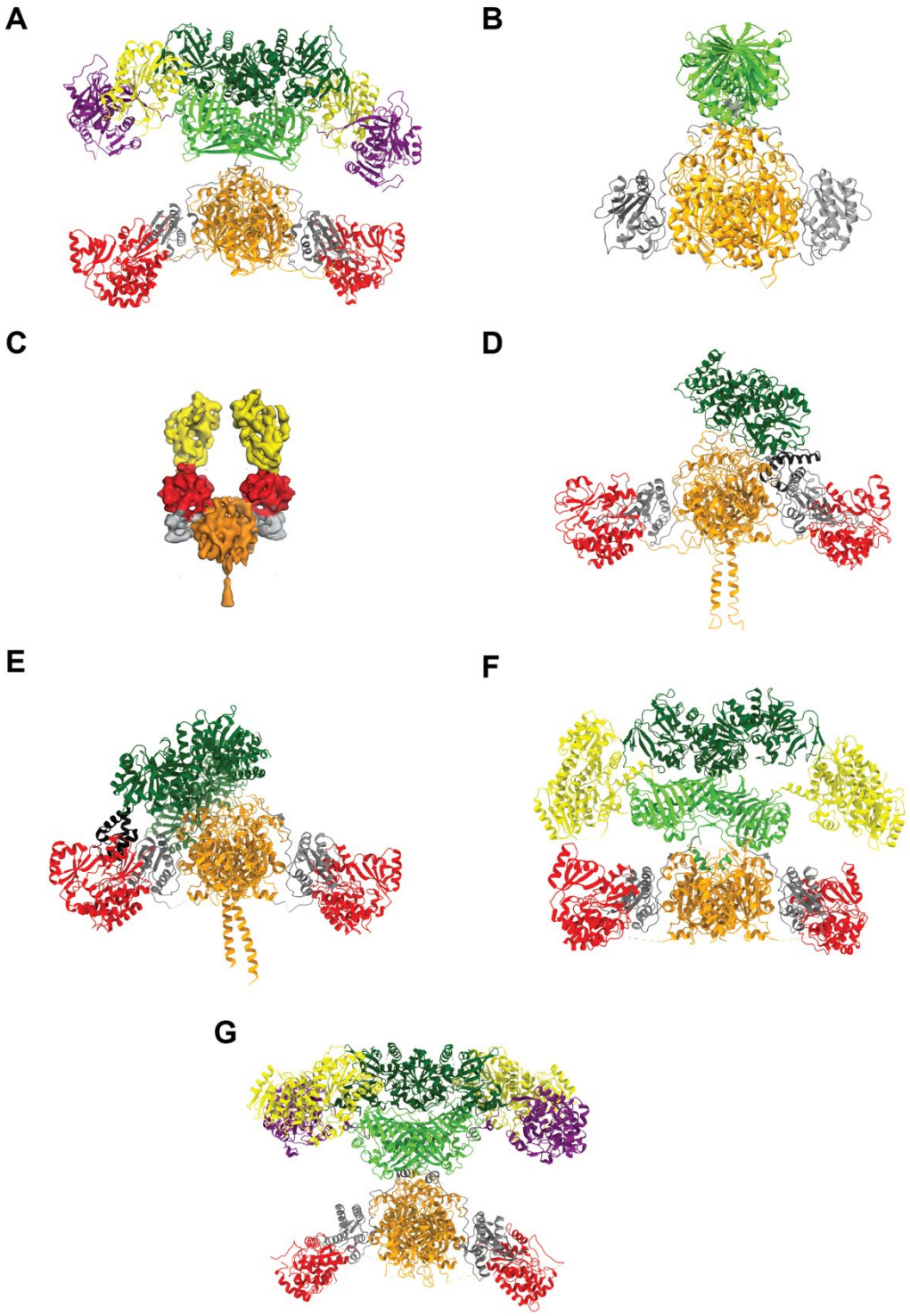


Figure 1.5: Known type I PKS modules and porcine FAS. *Enzymatic domains are color coded with the KS in orange, AT in red, DH in light green, ER in dark green, KR in yellow and the MT in purple. The ACP is depicted in black with the LD being grey.* ***(A)*** *Porcine FAS following KS-AT-DH-MT-ER-KR domain architecture solved by X-ray crystallography. PDB ID: 2VZ8* ***(B)*** *RhiE module core consisting of KS-DH solved by crystallography.* ***(C)*** *PikAIII module following KS-AT-KR-ACP architecture analysed by cryo-EM. EM-map colored by domain fitting as suggested by the authors.* ***(D)*** *DEBS module 1 with KS-AT-KR-ACP, analysed by cryo-EM. PDB ID: 7M7F* ***(E)*** *Lsd14 module 7 with KS-AT-KR-ACP analysed by cryo-EM. PDB ID: 7S6B* ***(F)*** *MAS module with KS-AT-DH-ER-KR architecture. Module consists of the condensing and modifying region crystallized individually and are placed using porcine FAS as a template. PDB IDs: 5BP1, 5BP4* ***(G)*** *lovastatin B module consisting of KS-AT-DH-MT-ER-KR, analysed by cryo-EM. PDB ID: 7CPX*

The architecture of cis-AT PKS modules

Interestingly all known cis-AT PKS modules follow the same domain architecture, highlighting differences between the three and allowing for the evaluation of differences within this class of PKSs, while at the same time complementing each other to arrive at a better overall understanding.

The first cis-AT PKS module structure was published in 2014. Contrary to prior belief the PikAIII module with a KS-AT-KR-ACP architecture involved in pikromycin and methymycin biosynthesis from a cis-AT PKS[85], adopts an arched conformation (Fig. 1.5C) as opposed to the conventionally expected extended conformation which was based on porcine FAS[82], observed in 2008, and a KS-AT subregion from the Khosla Lab[80]. Like all type I PKSs it forms a homodimer, with the KS domain being the only dimerizing domain of the module. All available conformations of the module follow a C2 symmetry and were analyzed at moderate resolutions of 7.3- 9.2 Å by cryo-EM. The arc itself is formed by the AT, which is a in downward bend as compared to other known models, with each KR protomer interfacing with their respective AT. This led to a reaction chamber hypothesis in which the KS would be loaded by its upstream ACP from the conventional ACP binding site and unloaded from a newly identified bottom entrance to the active site cavity via the module's own ACP. Questions regarding, amongst others, how this might affect the Claisen-condensation in the KS as the substrate itself would block the bottom entrance once loaded, have since relegated this model as a base architecture for all type I PKS modules[86]. Additional studies confirmed biochemically the activity of a DEBS module trapped in its extended state[87,88], as well as showing via SAXS and cross-linking mass spectrometry (XL-MS) that DEBS module 2 forms the extended structure in at least the dominant fraction of soluble molecules[89].

DEBS module 1 is part of the most well studied cis-AT PKS DEBS. It was excised from the first protein of the DEBS PKS, which consists of module 1 and 2. DEBS module 1 follows a KS-AT-KR-ACP architecture and was analyzed in 2021 by cryo-EM at 3.2 Å resolution[90] (Fig. 1.5D). Interestingly, this means it has the same module architecture as PikAIII. In previous studies this module was shown to be a stable dimer as a standalone module as well as being functionally active[91,92]. In order to make the module more suitable for structural studies it was fused N-terminally with the docking domains of DEBS module 3 as a Fab-fragment binding site and C-terminally with the TE from DEBS module 6 to increase the constructs solubility. The module follows the extended conformation with an interesting twist to the AT, which has both an extended conformation and a flexed conformation where its two subunits are apart and close together respectively. The cryo-EM study described three states for this construct, two of which include an ACP stalled at the KS. Both the KRs displayed a large degree of mainly rotational freedom, similar to what was described for porcine FAS. This freedom is enhanced in the case of DEBS module 1 as the KRs do not dimerize and are therefore two independent appendages to the modules condensing region.

At the same time as DEBS module 1 another cis-AT PKS module following the KS-AT-KR-ACP was solved at 2.4 Å using a mix of cryo-EM and X-ray crystallography[93] (Fig 1.5E). Termed Lsd14, it is the lone standing module 7 of the 11- module lasalocid A PKS assembly line. The module differs from DEBS module 1 in that it contains a dimerizing element placed between the AT and KR domains, therefore

between the condensing and the modifying region. It too follows the extended conformation and has three individual states resolved. In the apo from the ACP is stalled at the AT, in one of the cryo-EM models the ACP is not bound to any domain and in the last state the ACP is stalled at the KS. As with DEBS module 1 the modifying region is seen to have a large degree of freedom in both rotational and swiveling motions[93]. In both Lsd14 and DEBS module 1 the linkers between AT and KR could not be resolved, meaning the link between the condensing and modifying region, and its potential fold remains an open question. Interestingly, some of the KR-KS interactions observed for the module block the KS active site entrance, while simultaneously directing the ACP in close proximity to the AT.

When comparing these three structural models it is important to keep the difference in resolution in mind. While PikAIIIs cryo-EM map is an intermediate resolution map with a consequentially approximate placement of domains, both DEBS module 1 and Lsd14 are high resolution models, in all three states for both of them. Additionally, once again comparing either of them to porcine FAS as a blueprint both Lsd14 and DEBS do conform, following the same overall architecture whereas PikAIII appears to be an outlier. Interestingly the low-resolution structure of DEBS module 3 also conforms with both porcine FAS and both DEBS module 1 and Lsd14 in terms of the extended fold. Focusing on Lsd14 and DEBS module 1, it is relevant to note that the both had to be stabilized via an antibody (Fab 1B2) in order to constitute a stable dimer in solution. While DEBS module 1 may find additional dimerization stability when in its native protein with module 2, which also typically dimerizes, the standalone Lsd14 module might indeed not be a constitutive dimer in solution, or rely on docking domain interactions with both preceding and succeeding modules. Both raise interesting follow up questions as the homodimeric nature of cis-AT PKSs is commonly expected to help in substrate shuffling, or be straight required for certain enzymatic action, especially for some KSs.

Modular architecture of iPKSs

In terms of iPKSs thus far two structures of entire modules exist, both derived from HR-iPKSs. Despite their homology some key differences are observable. Mycocerosic acid-like synthase (MAS) is a HR-iPKS from *Mycobacterium smegmatis* involved in the production of C_{20}-C_{28} branch chained fatty acids, key virulence factors of its bacterial host. Its structure was solved by X-ray crystallography in two steps, with the condensing region consisting of KS-AT and its modifying region DH-ER-KR being crystallized individually at 2.2 and 3.8 Å resolution [94] respectively (Fig. 1.5F). Interestingly the KS-AT part on its own was monomeric in crystal whereas DH-ER-KR crystallized as a dimer. KS-AT was still found to be in a monomer-dimer equilibrium at a K_d of 0.4 mM[94]. Taken together the module follows the extended conformation although their direct interfacing could not be resolved. Stabilized by crystal contacts, the formation of an alpha helix in the linker between the two regions was observed. Despite the placement being based on prior knowledge from porcine FAS, which shares a striking resemblance, the linker would not allow for much further spacing of the two regions regardless. While the overall organization of the modifying region is similar between porcine FAS and MAS, excluding the absence of a MT from MAS, differences such as the completely changed DH dimerization angles at 222° instead of the FAS's DHs interprotomer angle of 96° exist. Naturally this comes with a different dimerization interface despite the highly conserved DH protomer fold between the two systems, showing that regardless of their similarities

large differences exist between PKSs and FASs. Given the chimeric nature of the model, questions regarding relative movement between the two regions could not be answered.

Like MAS, the lovastatin PKS is a HR-iPKS; it produces the statin precursor lovastatin. The lovastatin PKS employs LovB, the main module with the domain architecture KS-AT-DH-MT-KR-ER-ACP-C in concert with the ER protein LovC expressed in trans [39]. Its structure was solved in 2020 at 2.9 Å resolution by cryo-EM for LovB and at 3.6 Å for the LovBC complex (Fig. 1.5G). The structure of LovB revealed several important aspects of structural organization. For one, like MAS, LovB was visualized in an extended conformation. However, unlike MAS, its DH is highly similar in terms of dimerization interface to that of porcine FAS, making LovB a “FAS-like PKS”. Surprisingly, and in stark contrast to porcine FAS, LovB exhibit a rather rigid conformation, without substantial relative motions of its modifying and condensing regions, presumably due to a stabilizing small (approx. 304.7 $Å^2$) direct interface between the KS and DH domains [39]. Previously, the large-scale rotation between the two subregions previously observed in FAS has often been assumed to carry over to all type I PKSs. Instead of large-scale conformational changes, LovB apparently relies entirely on the mobility of the ACP that is tethered to its C-terminus via a long linker, to shuttle the substrate between active sites, without large overall conformational changes. The complex of LovB with the LovC ER is crucial for product synthesis despite the existence of an ER domain in LovB.
From a structural point of view, the LovBC structure provides the first example of a trans-acting enzyme interfacing with a multidomain PKS module. LovC interfaces with the AT of LovB, notable in an asymmetric manner, with only one protomer of the dimeric LovB module interacting with LovC, without large-scale conformational changes in the LovB module itself [39]. The observed binding pose of LovC to LovB would allow for the ACP to reach it, based on ACP-linker lengths analysis. LovC binding to only one protomer of LovB conforms with data suggesting that type I PKSs might only use one of their two enzymatic domains at a time[72].

MAS and LovB belong to the same class of HR-iPKSs, yet differentiate significantly. The more FAS-like LovB is the first type I PKS module to display any observable rigidity between the condensing and modifying region. The MAS structures chimeric nature makes a direct discussion of degrees of freedom difficult. However, when examining the linker between modifying and condensing region, it appears more plausible that it is more similar to porcine FAS than LovB in terms of degrees of freedom between the two domains despite its DH differing more from FAS than that of LovB. Additionally, the use of a trans-acting ER in the lovastatin synthesis regardless of the presence of the same domain type on its module highlights that a close look at any given modules BGC is essential to capture the full picture, even for iPKSs, which are typically expected to exclusive rely their one module for polyketide synthesis. Unfortunately, to date no complete module structures for either NR- or PR-iPKSs exist. NR-iPKSs are interesting here as their condensing region is markedly different to that of all other type I PKSs and porcine FAS with the inclusion of a second specialized AT.

Modular architecture of type I PKSs

Based on the observed structures of modules as outlined above, we can elaborate a general trend, which most type I PKS modules follow: Type I PKS modules are not only entities in the sense of enabling a step of polyketide chain extension and modification but are also structurally distinct entities (Fig. 1.5). Interestingly, the split between condensation and modification occurs on a structural level as well, with a KS dimer being at the center of the modules condensing region. Each of its protomers links to an individual AT protomer via the only part of scaffolding protein found in type I PKS, the LD. The modifying region sits on top of the condensing region, connected to the C-terminus of the condensing region via a short linker. This linker ensures distance constraints between the two regions, with the modifying region typically located ~10-20 Å above the condensing region. The modifying region itself consists of up to a KR, a DH and an ER. If all three enzymes are present, the DH homodimerizes in the center above the KS. Each of the DH protomers links to a KR protomer attached to its outer flanks. On top of the DH, the ER also homodimerizes making contacts with both the DH and the KR. If not all of the modifying regions are present, they follow the same architecture individually. The DH always homodimerizes, as does the ER, whereas the KR is always a monomer and can only interact with the other two to form a quaternary structure element. Both the condensing region and the modifying region therefore form stable complexes, that are attached to one another via a short linker. Lastly the ACP and the TE are attached to the multienzyme via a long linker protruding from the modifying regions C-terminus with another linker spacing the ACP and TE from one another. This allows for the ACP to reach all enzymatic domains in a restricted diffusion-based manner.

Type I PKS modules are, with the exception of LovB, highly flexible, with a distinct point being the degrees of both rotational and translational freedom between the modifying and condensing regions of a given module. Additionally, the ACP and its movement are thus far near impossible to trace, with the only available data being when it is bound to a given enzymatic domain.

The multimodular organization of modPKSs

While hypothesized on from the point onward where the homodimeric nature of modPKSs became apparent, with ideas ranging from helical or stacking [64] to a serpentine [62] arrangement, to date only two low resolution models of multimodular organization of modPKSs exists, both obtained using SAXS.

A glimpse at the multimodule architecture of cis-AT PKSs was afforded in 2014 by the Khosla Lab [50]. Specifically, they analyzed DEBS3 consisting of DEBS modules 5,6 and the TE, using SAXS, followed by modelling high resolution X-ray structures of the contained domains into the SAXS maps. DEBS3 has a KS-AT-KR-ACP-KS-AT-KR-ACP-TE architecture (Fig. 1.3), where module 5 consists of the first KS-AT-KR-ACP and module 6 of the second KS-AT-KR-ACP-TE. The multimodule follows a similar architecture to that of individual modules described above, with both KS-ATs being in an extended state and the KSs dimerizing. Once again, each KS protomer has one KR protomer interfacing with it, without the KR protomers contacting one another in either module. As in the SAXS model of DEBS module 3 the TE dimerizes and is placed atop module 6s KR, centrally above the KSs, without any direct interactions, however at a defined distance. As the model obtained is arguably in line

with knowledge on individual modules, the essential point to be derived here is a linear stacking of the two modules along a vertical axis aligning with the KSs dimer axis.

In 2016 an investigation of docking domains between VirA module 5 and its succeeding module VirFG module 6, led to a composite model for their interaction and a first insight into a trans-AT bimodular structure [95]. This is especially interesting as it depicts an inter-protein bimodule interaction, hence the need for docking domains. The docking domains observed differ from those found in cis-AT PKSs but fulfill the same purpose of mediating interactions between two modules on different proteins. The authors generated a new SAXS model for the KS of VirFG module 6, with its docking domain and generated a composite model, based on the docking domains need to interact and their relative positioning across VirA module 5 and VirFG module 6. The available domain configuration across the two proteins is therefore KS-ACP-ACP-docking domain-docking domain-KS, meaning an excised VirFG module 6 excluding its KR and ACP was used. The overall model places the KS dimers in a linear fashion, similar to that of DEBS3, and argues that the docking itself takes place at the outer edges of the VirFG module 6s LD, unlike DEBS3 where the C-terminal end of module 5 is covalently tethered to the N-terminus of module 6 at the center of module 6s KS dimer. The model therefore reinforces the clamping model hypothesized for trans-AT PKS module interactions by the authors previous work on VirA. As this conformation does not allow for interaction between the second ACP of VirA module 5 and the KS of VirFG module 6 a certain degree of conformational freedom is expected. While the overall stacking of modules is similar to that observed for DEBS3 it is important to note that neither of the modules includes a modifying region (module 5) or had its modifying region included (module 6) in the experiment, representing a maximally minimalistic system. Additionally, the authors did not generate a map including both the VirA module 5 and the excised part of VirFGs module 6, relying entirely on inference drawn from the relative positioning of the docking domains across two individual SAXS maps.

In terms of multimodular architecture of type I PKSs the data available to date is as described above unfortunately of low resolution owing mainly to the incredible flexibility of the individual regions and modules connected via linkers. In order to fully comprehend the action of PKSs it is crucial to investigate both their higher order architecture and its role in mechanisms such as substrate shuttling and substrate specificity. The resolution revolution in cryo-EM[96] and its suitedness to the observation of more flexible proteins and protein complexes have paved the way to take the next steps in elucidating the function of PKSs.

Aims of the thesis

With their repetitive chain extension and modification as well as their incredible versatility as a protein family PKSs have long been a prime target for engineering efforts to access complex organic compounds. Overall extensive knowledge on individual type I PKS key enzymatic domains exists from both a biochemical and a structural point of view. Recent studies described consistently the architecture of both cis-AT PKS and iPKS module organization. However, the key question of the higher order architecture characteristic of modPKS assembly lines, as well as the role of specialized domains in determining product characteristics in iPKS remained enigmatic. This work contributes to addressing these open questions and specifically addresses two aims:

1. *To elucidate the higher order architecture of a PKS assembly line (chapter 2).* For this task the core of a trans-AT PKS bimodule from *Brevibacillus Brevis* with the domain architecture KS-DH-ACP-KS was selected for structural analysis, as a minimal bimodule variant with two different module architectures. To investigate structures from the level of isolated domains to the entire core of the bimodule, cryo-EM single-particle reconstruction was combined with X-ray crystallography.
2. *To reveal the mechanistic basis of cyclization specificity in PTs (chapter 3).* PTs are known to cyclize a poly-beta-ketone substrate in NR-iPKSs. While several cyclization patterns are known biochemically, the mechanism for their specificity is unclear. In order to elucidate this mechanism a substrate analogue mimicking a poly-beta-ketone after the first of the two cyclization steps was crosslinked to the PksA PT, followed by visualization using X-ray crystallography.

Chapter 2

The structure of a polyketide synthase bimodule core

Yves U. Tittes[1], Dominik A. Herbst[1,2], Solène F.X. Martin[1], Hugo Munoz-Hernandez[1], Roman P. Jakob[1], Timm Maier[1]*

[1] Biozentrum, University of Basel, Spitalstrasse 41, 4056 Basel, Switzerland
[2] current affiliations: California Institute for Quantitative Biology (QB3), University of California, Berkeley, CA 94720. and Molecular Biophysics and Integrative Bio-Imaging Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA.
*Corresponding author: Timm.maier@unibas.ch

Statement of own contribution

I have personally performed all steps from cloning, protein purification, cryo-EM and X-ray sample preparation, cryo-EM and X-ray data collection, cryo-EM and X-ray data analysis, computational analysis of genomics data, the preparation of the manuscript and all figures within, with additional contributions on individual steps from others as described in the author contributions. The exception is Suppl. Fig 6, which was adapted from a figure generated by Solène F.X. Martin.

Abstract

Polyketide synthases (PKSs) are predominantly microbial biosynthetic enzymes. They assemble highly potent bioactive natural products from simple carboxylic acid precursors. The most versatile families of PKSs are organized as assembly lines of functional modules. Each module performs one round of precursor extension and optional modification, followed by directed transfer of the intermediate to the next module. While enzymatic domains and even modules of PKSs are well understood, the higher-order modular architecture of PKS assembly lines remains elusive. Here, we visualize a PKS bimodule core using cryo electron microscopy and resolve a two-dimensional meshwork of the bimodule core formed by homotypic interactions between modules. The sheet-like organization provides the framework for efficient substrate transfer and for sequestration of *trans*-acting enzymes required for polyketide production.

Teaser

A microbial polyketide assembly line organizes into a sheet-like supramolecular structure of aligned and mobile domains.

Introduction

Polyketide synthases (PKSs) are biosynthetic factories, primarily occurring in microbes, for the production of bioactive secondary metabolites, which have applications in medicine as antibiotics, immunosuppressants, and anti-cancer drugs[15,97]. PKSs assemble polyketides by stepwise condensation and modification of acyl building blocks[16].The most versatile class of PKSs, Type I PKSs, are organized into dimeric functional modules. Each module contains a condensing region comprising the condensing ketosynthase (KS) and a substrate-loading acyl transferase (AT) domain. Modules may further comprise an optional modifying region that often contains a double-hot dog-fold domain (DH). DH domains can have different enzymatic functions to act as dehydratases[16], isomerases[40], branching enzymes[41], or epimerases[98]. Throughout stepwise polyketide synthesis, the intermediates are tethered to an acyl-carrier protein (ACP) domain. ACPs are integral parts of the multidomain PKS polypeptides and are flanked by flexible peptide linkers. Modular Type I PKS (modPKSs) consist of multiple modules acting as an assembly line. Each module performs one step of chain extension and optional modification of the added moiety and passes on the new intermediate to the KS of the downstream module for the next extension step. Individual modPKS proteins typically consist of two to four modules and they carry docking domains at their termini, mediating interactions to upstream or downstream modules in the assembly line (Fig. 2.1A). A complete assembly line commonly consists of more than ten modules[65,99].

Two distinct types of modPKSs have evolved that either incorporate AT domains in each module (*cis*-AT PKSs) [53] or utilize *trans*-acting AT domains (*trans*-AT PKSs) [66]. In many modPKSs the organization of modules is colinear from gene order to protein interaction and biosynthetic reaction step[49], but *trans*-AT PKSs are known for exceptions to this rule of collinearity[57,58]. In *cis*-AT PKSs, the AT is connected to the KS by a ~130 AA linker domain (LD), the only common non-enzymatic scaffolding domain in modPKSs. Despite the absence of AT domains, *trans*-AT PKSs still contain LDs. Based on homotypic interactions of LDs in crystal structures of isolated KS domains, Keatinge-Clay and colleagues proposed LDs of *trans*-AT PKSs to form lateral interactions between equivalent KS domains, termed Laterally INteracting

Ketosynthase Sequences (LINKS) [100]. Structural data is available on individual PKS domains, two- to three-domain-fragments[49] and individual modules of *trans*-AT[41] and iterative- or *cis*-AT[39,85,90,93,94] PKSs, but for the higher-order architecture of PKS multimodules only conceptual models of serpentine[62], helical[64], sheet[100] and linear[50] arrangements have been proposed. Clusters of *trans*-AT PKSs of up to 0.2 micrometer size were identified by immunostaining electron microscopy (EM) in *Bacillus subtilis* [68]. ModPKSs are important targets for natural product engineering, but efforts towards re-engineering PKS assembly lines thus far are only partially successful [101]. They are hampered by a lack of structural insights into the interplay of domains and modules and the architectural requirements for productive multi-modular biosynthesis[49]. Here, we resolve the organization of a PKS bimodule core by combining crystallographic and cryoEM analysis of a fragment from a bacterial *trans*-AT PKS. Our analysis includes structures of isolated domains and of domains in the context of a bimodular assembly line as well as the higher-order organization of the bimodule core into a LINKS-bridged array.

Results

Structural analysis of a trans-AT PKS bimodule core

We studied the KS3-DH3-ACP-KS4 (K3DAK4) fragment of a 438 kDA four-module protein (GenBank:BBR47_39880) from the orphan biosynthetic gene cluster 11 (BGC11) of *Brevibacillus brevis* strain NBRC 100599 (Fig. 2.1A,B). This fragment contains two subsequent KS domains, compared to a minimal PKS bimodule, it only lacks a flexibly-tethered terminal ACP. We refer to this fragment as bimodule core. We combined X-ray crystallographic analysis of recombinantly expressed isolated enzymatic domains at 1.6 Å (KS3), 2.2 Å (DH3) and 3.2 Å (KS4) (Fig. 2.2, 2.3, Table S1), with cryoEM visualization of domains in the context of the bimodule core at 2.9 Å resolution (KS3, KS4 Fig. 2.1C, Fig. S1, S2, S3, Table S2) as well as the entire bimodule core polymerized into filaments at 6.8 Å resolution with maps focused on subregions at 4.0 Å and 3.9 Å resolution, respectively. (Fig. 2.1D, Fig. S1, S2, S3). The mobile ACP domains with their linkers remain unresolved.

Individual K3DAK4 bimodule cores adopt a single type of compact structure based on cryoEM 2D class analysis (Fig. S1B, S3B), a linear arrangement of enzymatic domains (Fig. 2.1C) with substantial variability of domain positioning, in particular for DH3. We represent the individual bimodule core by a hybrid model assembled from high-resolution cryoEM structures of KS3 and KS4 in the context of the bimodule core and a crystallographic model of DH3. The presence of alternative non-compact arrangements in solution can't be ruled out. Folds and dimerization interfaces of individual domains (Fig. S2A-D, Fig. 2.3) are preserved upon integration into the dimeric K3DAK4 bimodule core. The overall organization of domains in the third module with the DH3 domain positioned along the central dimer axis of KS3 (Fig. 2.1C) above the KS active sites agrees with previous models for iterative PKSs (LovB[39] , MAS-like PKS5 [94]) or *trans*-AT PKS didomains (RhiE[41]), but differs from a model proposed for the *cis*-AT PKS PikAIII[85].

The K3DAK4 bimodule core forms filaments

Filaments of the K3DAK4 bimodule core form spontaneously and co-occur on cryoEM grids with individual bimodule cores in varying populations. The average visualized filament (Fig. 2.2A,B) contains around 20 dimeric bimodule cores with dimensions of 100 Å length (in filament direction) and 160 Å height (Fig. 2.1D), equivalent to a total

filament length of 200 nm. Filaments can bend by up to approx. 12° per module. Our structural analysis does not identify any specific mechanism of filament length determination. Filament formation occurs via homotypic LINKS interactions (Fig. 2.1D, Fig. S4), providing lateral connections of each KS domain with two identical domains of the two neighboring bimodule cores. The linear organization of stacked KS3, DH3 and KS4 domains resembles that observed for individual bimodule cores, but LINKS interactions reduce the overall conformational variability. In the filament, the two KS domains of all bimodule cores are locked in the same orientation, while the DH3 domains maintain a higher degree of rotational freedom (Fig. 2.2A,B).
The distances between the centers of mass of KS3-DH3 and DH3-KS4 in the filaments are highly similar at around 60 Å. Given the discrepancy between the KS3-DH3 linker length of only nine amino acids (AAs) and the linkers between DH3 and KS4 comprising eighty-nine AAs plus the ACP domain, the distance similarity between the DH and both KS domains is intriguing. No fully ordered rigid contact site between DH3 and KS4 or local binding of linker regions enforcing proximity is observed, suggesting that transient dynamic interactions are responsible for maintaining this organization. An obvious candidate is the transient binding of ACP to individual enzymatic domains distributed across the filaments. Additionally, a loop region in KS4 (residues 1994-2006), termed here the clamp loop (Fig. S5, S6), protrudes towards DH3 and is posed to mediate further transient interactions. This loop is more prominent in *trans*-AT KSs (Fig. S5, S6) although also present in a small subset of less than 1% of *cis*-AT KSs (Fig. S5D). Given its preferential occurrence in *trans*-AT PKS it may also serve as a docking point for *trans*-acting enzymes. We hypothesize that the avidity of multiple transient ACP and clamp loop interactions across extended filaments results in the observed fixed intermodular distance, which determines the spatial constraints for substrate transfer.

The lateral interactions of KS domains

The two KS domains of the bimodule core adopt a canonical thiolase fold with complete active sites based on the structural conservation of the Cys-His-His motif (Cys694-His829-His869 for KS3, Cys1766-His1901-His1941 for KS4) (Fig. 2.2C-F). Both KS3 and KS4 clade with active KSs in transATor, a computational prediction tool for *trans*-AT PKS BGC products [102]. The crystal structures of the separately expressed KS domains and the cryoEM structures of the KS domains in the context of the isolated bimodule core closely resemble each other (Fig. S2A-G). The largest difference is observed for a region around the active site of KS4, which contains an extended ~20 AA loop absent from KS3. This region remains disordered in the crystal structure of isolated KS4, but is ordered in the cryoEM analysis of the bimodule core (Fig. S2E-G, S3A). KS3 and KS4 feature an LD domain with a lateral trigonal three-α-helix LINKS motif [100]. Homotypic LINKS interactions of these motifs combined with the dimeric structure of KS domains are the sole basis for filament formation; LINKS-based contacts are also involved in crystal lattice formation in our crystal structures of both isolated KS domains (Fig. S4A-C). The LINKS interaction in KS3 covers the sequence region of Asn1061 to Lys1096 with 16 AAs directly contributing to the interface. In KS4 the interaction region is between residues Asn2132 to Gln2167 with eighteen AAs involved in contacts (Fig. S4A-C). At 38.9 %, the sequence identity between KS3 and KS4 is lower in this region than across the entire KS domains. Distinct AA differences in the LINKS interfaces, such as Phe2139 of KS4 occupying the position equivalent to Leu1068 in KS3, explain the homospecificity of LINKS interactions (Fig. S4A-C). The integration of functional domains into the context of a multimodular PKS provides strong avidity of dimeric interactions across modules. The multimodular context thus

reduces the evolutionary pressure on maintaining stable dimerization interfaces of individual domains (Fig. S7). Consequently, KS domains derived from modular PKS are more likely to dissociate into monomers in solution then monofunctional dimeric KS enzymes, and such behaviour is observed here for KS3 (Fig. S7). Intriguingly, in cryoEM analysis, we also observed the isolated KS3 domain to dimerize via its LINKS interface without formation of the canonical thiolase fold KS dimerization interface (Fig. S1A).

The dimeric organization of the K3DAK4 DH3 domain

The double-hot dog fold DH3 is the central domain in the bimodule core. It defines the connecting region between KS3 and KS4 of the bimodule core, may engage in transient interactions across module borders and provides the N-terminal anchoring point for the subtrate-shuttling ACP domain. Still, DH3 retains a large degree of conformational variability within the bimodule core, which we interpret as rotational flexibility modeled by positioning the well-ordered crystal structure of the dimeric DH3 domain in two extreme orientations in the complete filament composite model (Fig. 2.1D, 2.2B). Based on crystallographic analysis, the DH3 domain forms a canonical double-hot dog fold observed in PKS dehydratases, isomerases and product template domains, and contains active site residues (Asp1349-His1172) characteristic of enzymatically active dehydratase domains[98] (Fig. 2.3A). The overall shape of the DH3 dimer is the result of a distinct mode of dimerization, different from those observed in *cis*-AT PKS or the iteratively acting PKS LovB (Fig. 2.3B, Fig. S8). *Trans*- and *cis*-AT PKS use 15-20 residues within the first 100 N-terminal AAs for forming the DH dimer interface. DHs of *cis*-AT PKS dimerize via N-terminal loops interacting with the solvent facing side of strands β_1- β_2 and β_5- β_6 of the double-hot dog fold. In DH3 however, the same loop between β_1- β_2 is involved in homotypic dimerization with its counterpart of the second protomer. The inter-protomer angle in *trans*- and *cis*-AT PKS DHs is similar, in the range of ~140°-190° (Fig. S8E), but as a consequence of the different interaction topology, the overall shape of the DH3 dimer is clearly different from those of *cis*-AT PKS DH domains (Fig. S8C,D). It also strongly differs from the DH dimer of the iterative LovB PKS, which is bent in opposite direction (Fig. S8A,B). To analyse the conservation of DH dimer organization within BGC11 and more generally the family of *trans*-AT PKS, we investigated crystal structures of DH4 (PDB 5HQW) from BGC11 and of the MlnD DH (PDB 5IL5) from the Macrolactin *trans*-AT PKS *of Bacillus amyloquefaciens,* which we have determined previously. Indeed, both MlnD and DH4 use the same mode of dimer formation as DH3 (Fig S8F), supporting the existence of a third fundamental type of dimeric DH architecture specific to *trans*-AT PKS, reflecting the distinct evolutionary path of *trans*-AT PKSs, iterative fatty acid synthase-like PKSs, such as LovB, and *cis*-AT PKSs as previously inferred from phylogenetic and structural analysis[39,60,103].

Discussion

The organization of a single complete PKS assembly as a linear, but flexible stack of modules, is obtained by propagating the module-to-module stacking observed here along the central dimer axis. The conformational variability of DH domains in such models increases the space the ACP can access. *Trans*-AT PKS modules mediating different steps of modifications of intermediates differ in their domain composition and may contain a ketoreductase (KR) [16] or DH-KR segment instead of the single DH domain. Due to the absence of defined rigid DH interactions, the linear assembly line architecture proposed here could accommodate all variations of the modifying

domains, and would even be compatible with the modular structure of *cis*-AT PKSs, which lack LINKS elements for lateral filament formation.
Horizontal filament formation between assembly lines via the LD[100] domains at each KS complements the vertical stacking of modules (Fig. 2.4). Association of extended assembly lines is favored by the avidity of multiple docking domains and LINKS-mediated interactions, even if individual domains in an assembly line would exhibit weak or no homotypic association, e.g. for individual *trans*-AT KS domains without LINKS motifs[100], or if the affinity of inter-polypeptide docking elements are low[99]. We hypothesize that multivalent transient interactions in two dimensions will result in module alignment and distance restraints between modules beyond the direct effect of peptide linker length to provide the framework for ACP-mediated substrate shuttling. Homospecific LINKS interactions ensure that assembly lines are always zipped up into filaments in register with product extension. Together with the moderate distance between neighboring assembly lines, this architecture could support ACP-mediated hopping of intermediates between neighboring assembly lines (see ACP positions in Fig. 2.4) as a strategy to overcome road blocks and to increase efficiency of precursor extension.
Extending the architecture of K3DAK4s filaments observed in cryoEM with an average length of around 20 bimodules onto the entire BGC11 with 17 modules yields an extended two-dimensional sheet structure, with dimensions that would be consistent with *in vivo* size estimates of 0.1-0.2 micrometer obtained for a complex of many copies of the bacillaene *trans*-AT PKS[68,100]. Local condensation of entire assembly lines into a two-dimensional mesh would support sequestration of *trans*-acting enzymes, in particular the essential *trans*-acting AT-AT-ER domains by avidity of a multitude of weak interactions.
Notably, *trans*-AT PKS, and in particular also the BGC11 studied here, form hybrid assembly lines with monomeric non-ribosomal peptide synthetase (NRPSs) modules to link amino acid and carboxylic acid building blocks. NRPSs require large-scale conformational flexibility for their multistep reactions[104-107]. A two-dimensional meshwork organization of *trans*-AT PKS provides different options for linking to or integrating NRPS modules (Fig. S9), but also imposes steric restraints. Understanding the interplay between these two types of biosynthetic modules will be a crucial step towards engineering hybrid systems to massively expand the chemical product space.
Our analysis of a *trans*-AT PKS bimodule core covers all levels of organization from individual domains to filaments of multimodular PKS systems (Fig. 2.4). The visualization of a sheet like organization of a *trans*-AT PKS assembly line provides a comprehensive basis for probing the role of *trans*-AT PKS assembly formation *in vivo* as well as for engineering bio-inspired enzymatic systems.

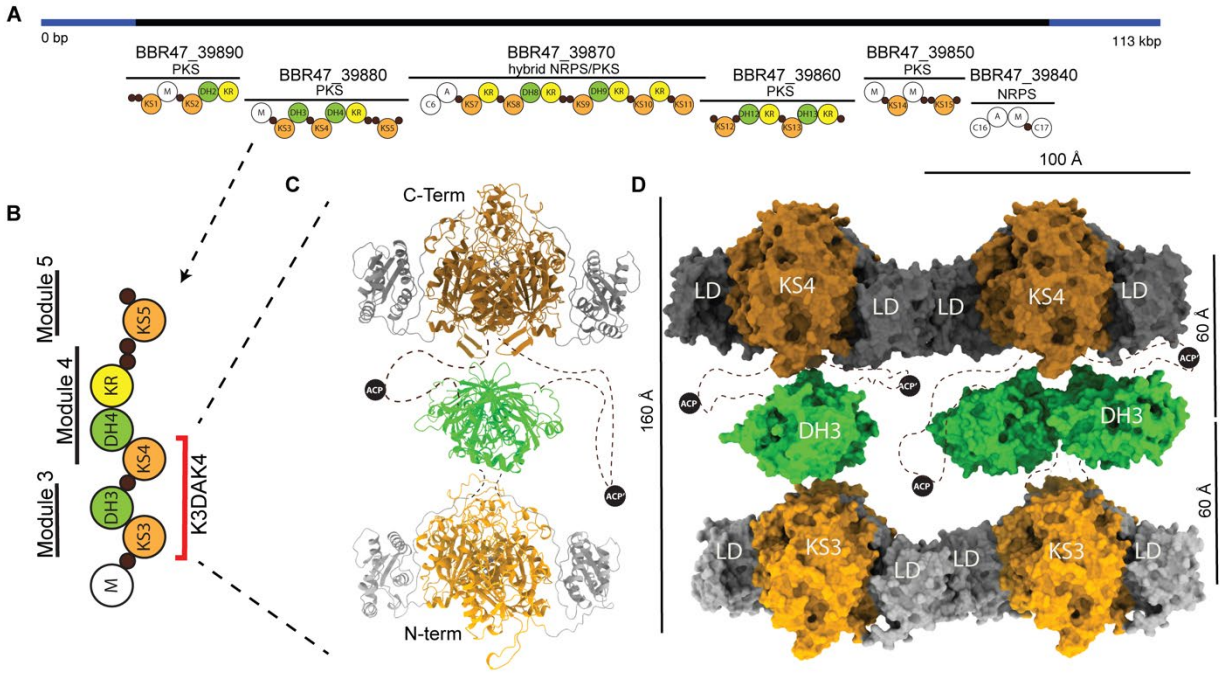


Figure 2.1. The architecture of the K3DAK4 bimodule core of the Brevibacillus brevis BGC11 PKS. (A) *Functional annotation of BGC11 as provided by Antismash[108]. Regions coding for PKS and NRPS core enzymes (black) or auxiliary enzymes(blue) are indicated. The domain organization of encoded PKS and NRPS proteins is shown schematically, PKS domains are labelled and shown in color, NRPS condensation (C), adenylation (A), and methylation (M) domains are shown as white circles, small black circles represent carrier proteins.* ***(B)*** *Domain and modular organization of the BBR47_39880 protein, the minimal K3DAK4 bimodular fragment containing two subsequent KS domains is indicated in red.* ***(C)*** *Pseudo-atomic model of an isolated dimeric K3DAK4 bimodule core assembled from individual high-resolution structures based on cryo EM 2D classes (Fig. S1B, S2C-F, S3B). The ACP is not resolved, plausible positions are depicted schematically.* ***(D)*** *Surface representation of the K3DAK4 bimodule core (vertical) filament (horizontal) (PDB 7ZSK). Individual KS and DH dimer models were rigid-body fitted into the overall intermediate resolution cryoEM map (map (1) in Fig. S1), with DH3 initial models positioned in two different orientations for visualization of rotational disorder, the mobile ACP is indicated schematically. Distances between the center of mass of dimeric domains (right) and overall dimensions are indicated (left, top).*

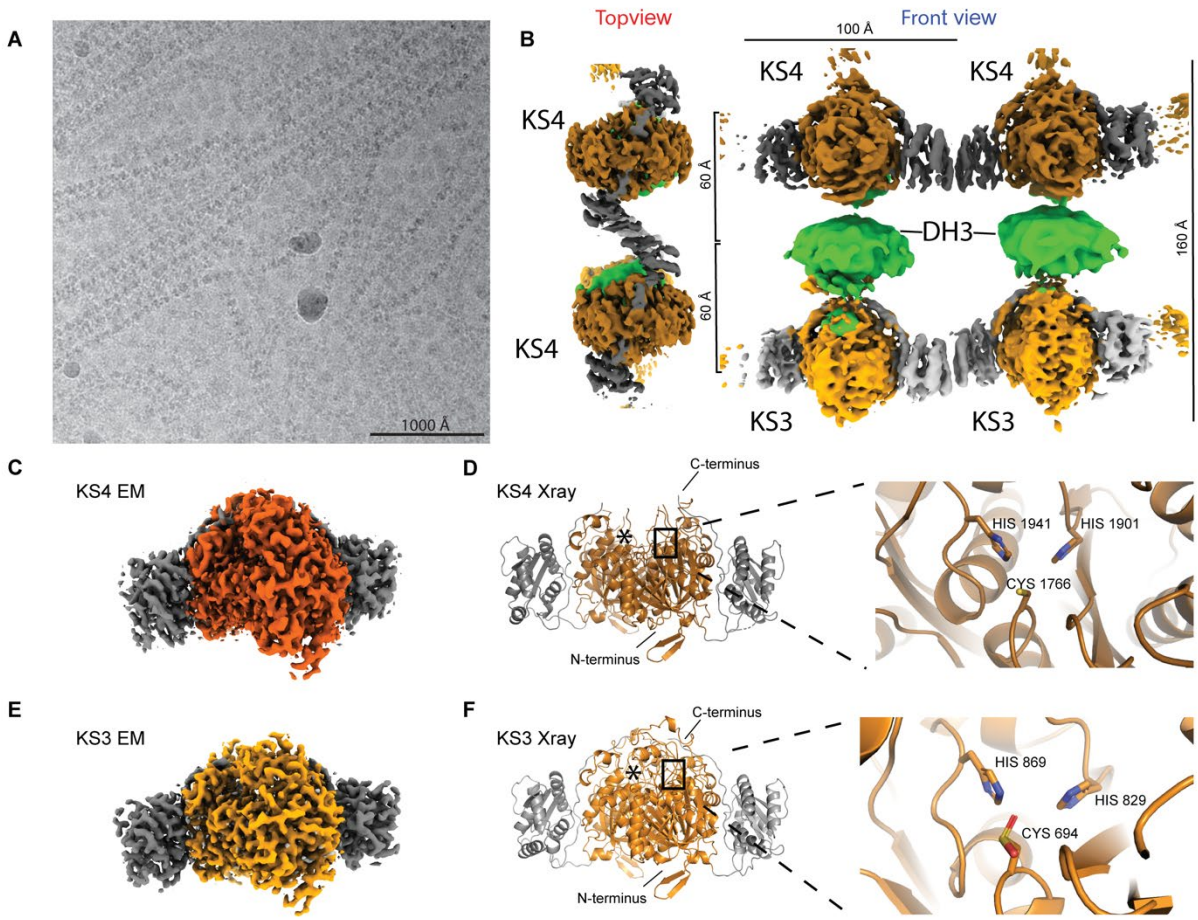


Figure 2.2. CryoEM analysis of K3DAK4 filaments and individual bimodule cores. (A) *CryoEM micrograph from the filament data set.* ***(B)****, Hybrid representation of two adjacent bimodule cores within the filament assembled from two focused refinement maps for laterally linked pairs of KS3 (light grey/orange) and KS4 (dark grey/orange) dimers (maps (2) and (3) in Fig. S1) and an excised region covering a pair of DH3 dimers (green) from an overall map of the bimodule core filament (map (1) in Fig. S1). KS3 and KS4 form homotypic LINKS interactions. In the top view, the KS3 and DH domains are obscured by the KS4 domain (cf. also Movie S1).* ***(C-F),*** *Structural analysis of dimeric KS3 and KS4 domains in isolation or in the context of K3DAK4 bimodule cores.* ***(C)*** *cryoEM map derived from focused refinement of KS4 in individual bimodule cores at 2.9 Å resolution.* ***(D)****, X-ray crystallographic model of the isolated KS4 domain at 3.2 Å resolution. Notably, the area around the C-terminus of KS4 is better ordered in the cryoEM map (also see Fig. S2, S3A).* ***(E)*** *cryoEM map derived from focused refinement of KS3 in individual bimodule cores at 2.9 Å resolution.* ***(F)*** *X-ray crystallographic model of the KS3 domain at 1.7 Å resolution. The active site cysteine is oxidized in the crystal to a sulfinic acid. Asterisks in (D, F) depict position of the second protomers active site.*

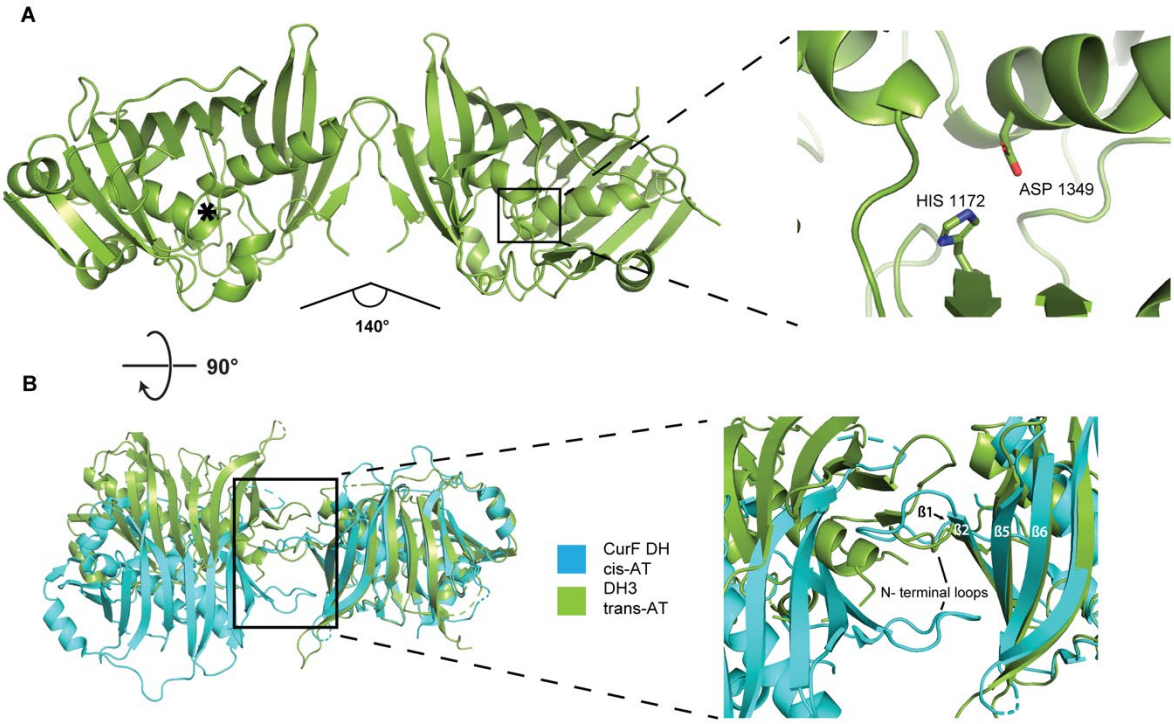


Figure 2.3. The dimeric structure of the isolated DH3 domain of K3DAK4. ***(A)*** *Overview of the 2.2 Å resolution crystal structure of DH3 (left) and a close-up view of one active site (right). The active site of the second protomer is indicated by an asterisk. The angle between the two protomers is indicated. (**B**) Comparison of the dimer interface of DH3 (green) with the dehydratase domain of the cis-AT PKS CurF (turquoise). Models are aligned on the right protomer. The inset highlights the different topology of ß-strand and N-terminal loop interactions in the dimerization interfaces of DH3 and CurF DH.*

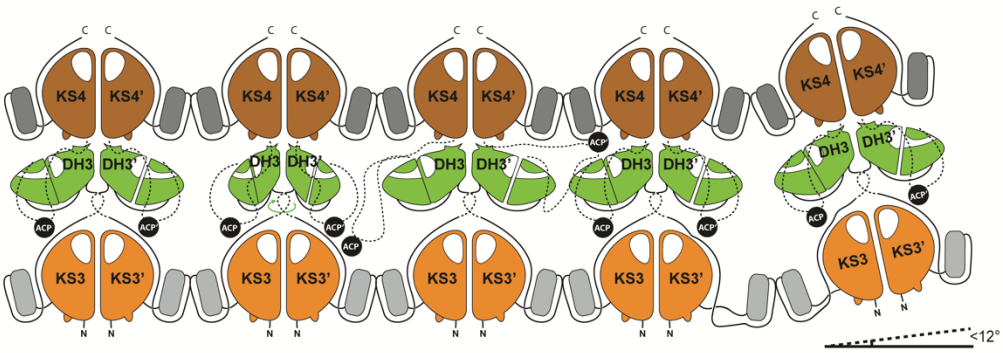


Figure 2.4. Schematic depiction of a K3DAK4 filament. *Each bimodule core forms LINKS interactions with both its neighboring bimodule cores. The DH dimers in the middle portray a large degree of rotational freedom that together with flexible linkers (schematically indicated by dashed lines) may enable the ACP to reach also laterally to neighboring bimodule cores in the filament. The filament bends by up to 12 ° in direction of filament extension, twisting in other directions occurs with smaller magnitude.*

Materials and Methods

Cloning and Protein Expression

Brevibacillus Brevis (strain 47 / JCM 6285 / NBRC 100599 obtained from NBRC) was cultured at 30 °C for 2 days in LB. Cells were pelleted and washed using Lysis buffer (10 mM Tris-HCl pH 8.0, 25 mM EDTA, 100 mM NaCl). Cells were resuspended in Lysis buffer and incubated for 1 day at 37 °C after addition of RNase and lysozyme, followed by incubation for 1 day at 50 °C supplemented with 0.028 % of SDS and 0.22 mg/ml Proteinase K. DNA was purified using phenol-chloroform extraction[109] and dissolved in TE buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA). The BbC0ZGQ6 (*Brevibacillus brevis* strain 47 / JCM 6285 / NBRC 100599, AA 532-2220, construct KS3-DH3-ACP-KS4) bimodule core construct was cloned into the Gateway entry vector pDONR221 and transferred via Gateway-cloning into a modified pACEBac1 coding for an N-terminal His10-myc-Flag-tag and pETG10A-N-H6. The Baculovirus expression system was used for protein production [110]. Bacmids were transfected into and virus generation was carried out in Sf21 insect cells (Expression Systems, IPLB-Sf-21-AE), grown in HyClone insect cell medium (GE Life Sciences). Baculovirus infected Sf21 cells (BICs) were harvested at generation 2 after transfection and stored at -80 °C. The individual domains KS3 (AA 532-1135), DH3 (AA 1135-1434) and KS4 (AA 1587-2220) were each cloned out of the pETG10A-N-H6 vector into p7XNH3 vectors using FX cloning[111].
For expression of the full length bimodule core K3DAK4, BICs were used to infect Sf21 cells and cells were harvested after 3 days by centrifugation. Pellets were stored at -80 °C. The isolated domains KS3, DH3, KS4 were expressed in *Escherichia Coli* BL21 (DE3) with induction using 0.1 mM IPTG at an OD_{600} of 0.8 followed by overnight expression at 30 °C. Cells were harvested using centrifugation and stored at -80 °C.

Protein Purification

Sf21 cells were resuspended in lysis buffer (50 mM $NaPO_4$ pH 7.5, 75 mM NH_4SO_4, 50 mM imidazole, 10 % glycerol, 4 µM bestatin, 2 µM pepstatinA, 20 µM phenantroline, 2 µM phosphoramidon, 5 mM beta-mercaptoethanol, 10 mM $MgCl_2$, DNAse, RNAse) and lysed by sonication. The lysate was cleared by centrifugation (158 420 g, 1 h, 4

°C), filtered (0.45 µm) and loaded onto two serially connected 5 ml high affinity Ni-charged resin column (Genscript) pre-equilibrated with lysis buffer. Unbound protein was eluted with 10 column volumes (CVs) lysis buffer 15 CV NiA (50 mM $NaPO_4$ pH 7.5, 75 mM NH_4SO_4, 100 mM Imidazole, 10 % v/v glycerol, 4 µM bestatin, 2 µM pepstatinA, 20 µM phenantroline, 2 µM phosphoramidon, 5 mM beta-mercaptoethanol). The sample was eluted with 4 CV of NiB (20 mM Tris-HCl pH 8.0, 50 mM NH_4Cl, 350 mM imidazole, 10 % glycerol, 4 µM bestatin, 2 µM pepstatinA, 20 µM phenantroline, 2 µM phosphoramidon, 5 mM beta-mercaptoethanol) and diluted 1:4 with AIC-A (20 mM Tris-HCl pH 8.0, 20 mM NH_4Cl, 10 % glycerol, 4 µM bestatin, 2 µM pepstatinA, 20 µM phenantroline, 2 µM phosphoramidon, 5 mM beta-mercaptoethanol). The sample was loaded onto a 6.5 ml anion exchange column (PL-SAX 4000 Å 10 µm, Agilent) followed by a wash step of 8 CV using AIC-A. Elution was performed with a gradient from 30 % AIC-B (20 mM Tris-HCl pH 8.0, 500 mM NH_4Cl, 10 % glycerol, 4 µM bestatin, 2 µM pepstatinA, 20 µM phenantroline, 2 µM phosphoramidon, 5 mM beta-mercaptoethanol) to 100 % AIC-B in 9 CV including a hold of 2.5 CV at 49 %. Pure fractions were pooled and concentrated (50 kDa MWCO Sartorius vivaspin 20, 3220 g, 5 min cycles) followed by size exclusion chromatography (Superdex 200 prep grade 16/600, GE Healthcare) using GF buffer (50 mM $NaPO_4$ pH 7.5, 75 mM NaCl, 5% v/v glycerol, 1 mM TCEP). Elution fractions were analyzed by SDS-PAGE, pure fractions were combined, concentrated to 1.1 mg/ml (50 kDa MWCO Satorius vivaspin 20, 3220 g, 5 min intervals) and stored at -80 °C.

E. coli cells were resuspended in lysis buffer 2 (50 mM Hepes-KOH pH 8.0, 500 mM NaCl, 20 mM imidazole, 10 % glycerol, 2 µM pepstatinA, 20 µM phenantroline, 4 µM E-64, 2 µM phosphoramidon, 5 mM beta-mercaptoethanol) then lysed on ice using sonication. The lysate was cleared by ultracentrifugation (158 420 g, 30 min, 4 °C), filtered (0.8 µm), and loaded onto a with lysis buffer pre-equilibrated 5 ml high affinity Ni-charged resin column (Genscript). Two wash steps were performed with 9 CV of lysis buffer then 9 CV of NiA2 (50 mM Hepes-KOH pH 8.0, 50 mM NaCl, 20 mM imidazole, 10 % v/v glycerol, 1 µM pepstatinA, 10 µM phenantroline, 2 µM E-64, 1 µM phosphoramidon, 5 mM β-mercaptoethanol) followed by elution using 8 CV of NiB2 (50 mM Hepes-KOH pH 8.0, 50 mM NaCl, 400 mM imidazole, 10 % v/v glycerol, 1 µM pepstatinA, 10 µM phenantroline, 2 µM E-64, 1 µM phosphoramidon, 5 mM β-mercaptoethanol). The sample was dialyzed over night into 1.6 L dialysis buffer (12.5 mM Hepes-KOH pH 8.0, 31.25 mM NaCl, 11.875 % v/v glycerol, 5 mM beta-mercaptoethanol) under addition of 200 µg of C3 protease (homemade) to cleave the N-terminal His10 tag followed by removal of cleaved tag and uncleaved protein using an orthogonal 5 ml column packed with High affinity Ni-charged resin (Genscript). Subsequently, the sample was concentrated (10 kDa MWCO Satorius vivaspin 20, 3220 g, 4 °C, 10 min intervals) and subjected to size exclusion chromatography (Superdex 200 prep grade 16/600, GE Healthcare) in GF buffer 2(20 mM Hepes-KOH pH 8.0, 250 mM NaCl, 5 % v/v glycerol, 5 mM DTT). All isolated domains (KS3, DH3, KS4) were purified following the same protocol but using a Superdex 75 prep grade 16/600 GE Healthcare column for the DH3 domain. All purification steps were carried out at 4 °C.

Crystallographic analysis of isolated domains

Crystallization experiments were performed using the sitting drop vapor diffusion technique. KS3 crystals were grown at 4 °C with a mix of 0.2 µl drops of 12 mg ml^{-1} protein in GF buffer2 with 0.2 µl of reservoir solution (1 % w/v PEG MME 2k, 1 M Na

succinate). DH3 crystals were grown at 20 °C with a mix of 0.2 µl drops of 3.9 mg ml^{-1} protein in GF buffer2 with 0.2 µl of reservoir solution (20 % w/v PEG 500 MME; 10% PEG 20k, 0.1 M Tris;bicine pH 8.5, 0.06 M $MgCl_2$ hexahydrate, 0.06 M $CaCl_2$ dihydrate). KS4 crystals were grown at 20 °C with a mix of 0.2 µl drops of 4 mg ml^{-1} protein in GF buffer2 supplemented by 10 mM $MgSO_4$ with 0.15 µl of reservoir solution (0.2 M $(NH_4)_2SO_4$, 0.1 M Sodiumcitrate pH 5.22, 8 % w/v PEG 3350). All crystals were cryoprotected in mother liquor with ethylene glycol added to 25% (v/v) before flash freezing in liquid nitrogen. All data were collected at the Swiss Light Source (Paul-Scherrer Institute, Villigen, Switzerland) at 100 K. KS3 data sets were collected at PXIII, DH3 and KS4 datasets were collected at PXI, with wavelength of 1.00, 1.00 and 0.98 Å respectively. Data reduction was performed using XDS[112] and XSCALE[112] and phases were determined by molecular replacement in PHASER[113,114]. The final models were built using model building in COOT[115] and refinement in phenix[116] (Table S1).

KS3 (7ZM9) crystallized in space group C2 with a unit cell of a=126.78 Å, b= 92.69 Å, c= 99.27 Å and β=92.54°. Our earlier structure of the isolated KS2 domain from BGC11 was used for replacement (PDB 4Z37). The final model was refined to R_{Work}/R_{Free} of 0.15/0.17 at 1.62 Å resolution. Ramachandran values are 97.46%/2.54%/0.00% (favored/allowed/outliers).

DH3 (7ZMF) crystallized in space group of $P2_12_12$ with a unit cell of a=72.38 Å, b=197.36 Å, c= 39.17 Å. Our earlier structure of an isolated DH domain of another module (DH4) of the BGC11 served as a replacement model (PDB 5HQW). The final model was refined to a R_{Work}/R_{Free} of 0.22/0.25 at 2.21 Å resolution. Ramachandran values are 97.58%/2.42%/0.00% (favored/allowed/outliers).

KS4 (7ZMC) crystallized in space group $C222_1$ with a unit cell of a=73.56 Å, b= 191.79 Å, c= 293.81 Å. Two data sets from a single crystal were merged using XSCALE. A search model was generated in Swiss Model[117] using KS3 as a template, refined into the EM density of KS4 using phenix [116] and then used as a replacement model for molecular replacement in Phaser[114]. The final model was refined to an R_{Work}/R_{Free} of 0.26/0.29 at 3.1 Å resolution. Ramachandran values are 93.14%/6.67%/0.19% (favored/allowed/outliers).

CryoEM sample preparation, data collection and analysis

Samples were thawed, centrifuged (30 min, 4 °C, 21k g), dialyzed overnight (o/n) at 4 °C into 37.5 mM $NaPO_4$ with 5 mM TCEP, and centrifuged again (5 min, 20 °C, 21k g). Sample were diluted to 0.15 mg/ml (~1 µM) and incubated at 23 °C for 2-6 h before grid preparation. 300 mesh Cu lacey grids (Agar Scientific) were glow discharged (45 s) in air, 4 µl sample was applied using a clipped pipette tip, adsorbed for 60 s, and pluge frozen in liquid ethane using a Leica EM GP (3 sec blot, 90 % humidity, 20 °C). Three data sets were collected on a Titan Krios G1 (Thermo Fisher Scientific) electron microscope equipped with a Gatan LS Quantum energy filter, and a K2 Summit detector (Gatan) in super-resolution mode. Movies were collected with 50 frames over 10 s of exposure at a dose rate of ~6 e^- sec^{-1} $Å^{-1}$ and a pixel size of 1.058 Å (0.529 Å super resolution pixel size) (Table S2) Movies were motion corrected using patch motion correction followed by patch ctf correction in cryoSPARC (Structura Biotechnology Inc.). Particles were picked using the blob picker (filaments) and template picker (individual bimodule cores) followed by multiple rounds of 2D classification and sorting in cryoSPARC (Fig. S1). For the filaments no symmetry was applied, whereas the individual bimodule core set had two fold symmetry applied in the final steps.

Ab initio classes for both the bimodule core data set as well as the filament data set were generated in cryoSPARC. After picking with an initial 2d class derived template, a set of 99 566 particles of individual bimodule cores was selected from 2D classes. A large degree of flexibility between the individual dimeric domain prevented high-resolution reconstruction of the entire bimodule core. From the same data set, using a template representing a single KS domain, a total of 353,464 particles after 2D classification were selected for further processing of individual KSs. An ab initio model was calculated in cryoSPARC, and non-uniform homogenous refinement, iterated with local and global CTF refinement yielded a map at 2.9 Å resolution, which represents a combination of KS3 and KS4 domains as evident form the densities for non-identical sidechains and few differences of mainchain regions. 3D variability analysis (3DVA) on the whole set followed by clustering into two clusters provided two starting models for heterogeneous refinement starting at a high initial resolution of 5Å, leading to a split of 182,315 in the first and 171,149 particles in the second class, respectively. Map analysis confirmed successful sorting into KS3 and KS4 based on side-chain densities, few main-chain differences and relative position of residual density representing further domains of the bimodule core. Further non-uniform refinements for each subset applying C2 symmetry followed by local and global CTF refinements and a final step of local refinement using a global mask and C2 symmetry, lead to final maps at 2.90 Å (KS4) and 2.93 Å (KS3) resolution. The maps are of excellent quality and clearly distinguish KS3 and KS4 at the sequence level (Fig. S3). In both KSs, parts of the LINKS helices, that aren't engaged in LINKS interactions, are less well resolved. Both KS3 and KS4 X-ray crystallographic models (PDB 7ZM9 and 7ZMC) were real-space refined into their respective maps using phenix, followed by manual model building in coot and a final round of real-space refinement in phenix.

To determine an approximate average length of filaments 7 random micrographs were picked and all continuous filaments identified by having more than 4 repeat units were manually traced and the repeat units were counted. Crossovers of filaments and the presence of filaments reaching beyond the field of view prevent an exact quantification of filament length.

For the filament data set, all processing was performed in cryoSPARC in a single particle workflow rather than by helical analysis due to the large size of the repeat unit of the flat filament and the variable connections between filament units. Four ab-initio models were generated of which one representing a total of four bimodule core dimers was finally selected as initial model. After two rounds of non-uniform refinement a map region representing two adjacent bimodule core dimers was excised using chimera, and expanded into a mask and used as model for all further reconstructions. Refinement of the whole filament using the mask around two bimodule core yielded a resolution of 6.8 Å in cryoSPARC. Overall resolution was limited by flexibility originating from the variety of filament conformations originating from the bend filament as well as an overrepresentation of top views. 3DVA in cryoSPARC revealed only minor shifts with regards to the positioning of KS domains, that are presumably linked to different bent-angles of filaments. Local maps covering LINKS-bridged pairs of KS dimers were obtained from 260,601 particles of two bimodule core dimers in masked and local refinements of the corresponding regions for KS3 and KS4 (maps (2) and (3) in Fig. S1, included as additional maps in EMDB-14945) from a set of local motion-corrected and reextracted particles. No improvements in resolution or map quality were observed using attempts of 3D classification in relion3.1[118]. Pronounced positional variability, predominantly rotation of individual DH dimers, precluded

improvements of map quality in masked and local refinements of a lateral pair of DH3 dimers.
For obtaining an overall model of two laterally linked bimodule cores (PDB 7ZSK), we assembled start models for KS3, KS4, and DH based on the individual crystal and high-resolution cryoEM structures (PDB 7ZMD, 7ZMA, 7ZM9, 7ZMC, 7ZMF). Models for KS3 and KS4 were optimized by chain-, domain- and segment-wise rigid-body fitting and overall coordinate and ADP refinement into the respective maps of pairs of KS dimers from local refinement (map (2),(3) in Fig. S1C, EMDB-14945 additional maps). A model of two laterally linked bimodule core dimers was then assembled from the refined KS3 and KS4 dimer pairs and DH3 dimers. Positional disorder prevents unambiguous identification of a single DH3 dimer orientation. To model this disorder, we manually defined two different starting orientations for rigid body refinement, representing divergent states of rotation of the DH3 dimer. Finally, rigid-body and ADP refinement with one rigid group per KS3, KS4 or DH3 dimer against the overall intermediate resolution map (EMDB-14945, map (1) in Fig. S1C) was carried out. The final results of rigid body refinement of DH3 are influenced by the choice of start model orientation. Refinement at intermediate resolution is not suited to optimize local sidechain interactions at the LINKS interface, which is best visualized in the respective higher resolution X-ray models of KS3 and KS4 dimers.

Sequence and structure analysis

All sequence comparisons were performed using BLAST[119] and visualized in Geneious Prime (version 2019-03-05). Clading analysis of the KSs was performed using TransATor [102]. For module assignment, we employ traditional module boundaries, but acknowledge the relevance of alternate definitions based on evolutionary patterns[120,121]. Structural comparisons were performed using PDBeFold[122], interfaces were analysed using PISA at the European Bioinformatics Institute[123] or QtPISA from CCP4[113]. For the comparison of DH dimer shape, one DH protomer was superposed based on secondary structure elements. Angles between DH protomers were analyzed using phenix.angle based on ordered secondary structure elements only. BGC analysis was performed using Antismash[108]. Sequence-based comparison of the clamp loop lengths was performed with a custom python script (DOI: 10.5281/zenodo.6550895) to count the clamp loop lengths in amino acids for all entries in the antismash DB[124] that match both terms, “KS” and “type I PKS”. All sequences were aligned against KS3 and loop length was defined as distance between residues matching KS3 clamp loop anchor points 920 and 936. Analysis of clamp loop lengths in KSs with structural models in the pdb was performed based on structural alignments using superpose[122] (secondary structure matching) from CCP4 instead of sequence alignments and manual counting of residues using the same anchor points.

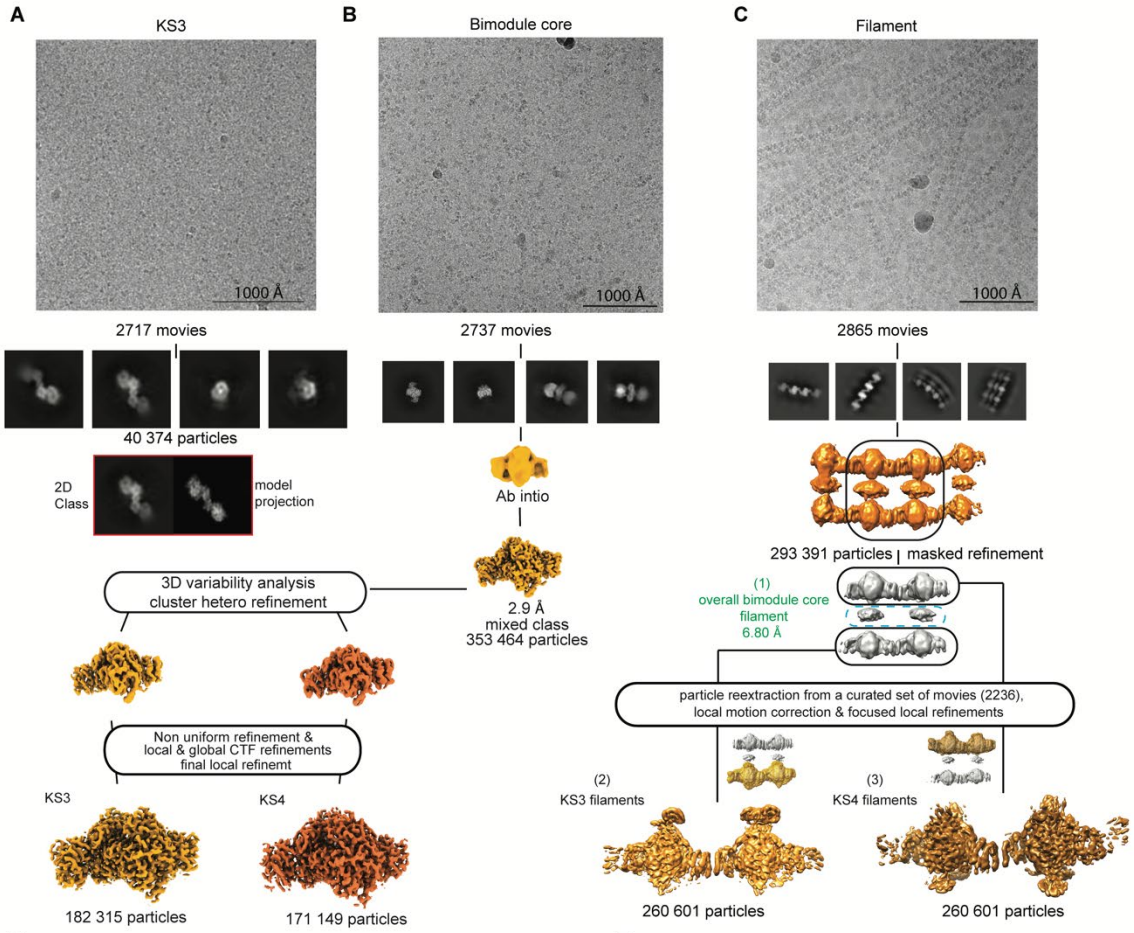
Primers

K3DAK4 gateway fwd 532
GGGGACAAGTTTGTACAAAAAAGCAGGCTTAGAGAATCTGTATTTCCAGGGGTC
GCCTTCATCGGTGGTGCG
K3DAK4 gateway rev 2220
GGGGACCACTTTGTACAAGAAAGCTGGGTACTACAAGGGATGCAGGTAAGCCG
G

KS3 FX cloning fwd
TATATGCTCTTCGAGTTCGCCTTCATCGGTGGTGCG
KS3 FX cloning rev
TATATGCTCTTCATGCGCTGCTAACAGACCTGGCGTG

DH3 FX cloning fwd
TATATGCTCTTCGAGTAGCACAGGTGTGGGTGTGATTC
DH3 FX cloning rev
TATATGCTCTTCATGCTGATCTAACAGATACATTTGCAATCTTGC

KS4 FX cloning fwd
TATATGCTCTTCGAGTCCAAAAGGGGCGTTGGACATTG
KS4 FX cloning rev
TATATGCTCTTCATGCGGGATGCAGGTAAGCCGGC


A
KS3
1000 Å
2717 movies
40 374 particles
2D Class
model projection
3D variability analysis
cluster hetero refinement
Non uniform refinement &
local & global CTF refinements
final local refinemt
KS3
KS4
182 315 particles
171 149 particles
B
Bimodule core
1000 Å
2737 movies
Ab intio
2.9 Å
mixed class
353 464 particles
C
Filament
1000 Å
2865 movies
293 391 particles
masked refinement
(1)
overall bimodule core
filament
6.80 Å
particle reextraction from a curated set of movies (2236),
local motion correction & focused local refinements
(2)
KS3 filaments
(3)
KS4 filaments
260 601 particles
260 601 particles

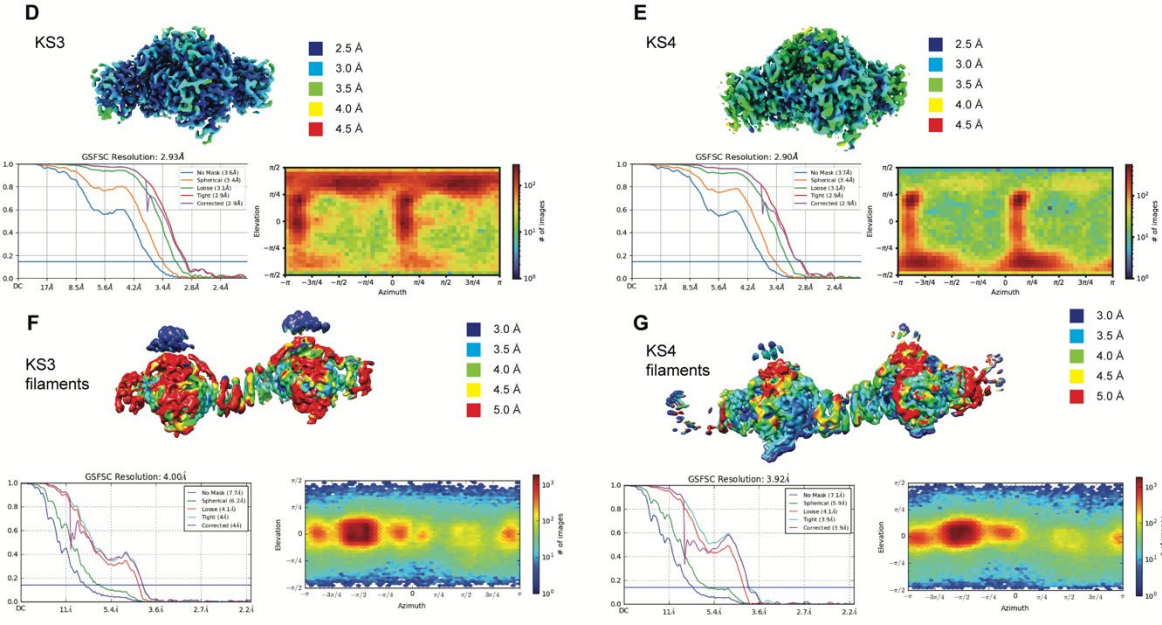

D
KS3
2.5 Å
3.0 Å
3.5 Å
4.0 Å
4.5 Å
GSFSC Resolution: 2.93Å
E
KS4
2.5 Å
3.0 Å
3.5 Å
4.0 Å
4.5 Å
GSFSC Resolution: 2.90Å
F
KS3
filaments
3.0 Å
3.5 Å
4.0 Å
4.5 Å
5.0 Å
GSFSC Resolution: 4.00Å
G
KS4
filaments
3.0 Å
3.5 Å
4.0 Å
4.5 Å
5.0 Å
GSFSC Resolution: 3.92Å

Fig. S1.
Schematic representation of cryoEM data processing. *A total of three datasets were collected, one of the recombinantly expressed isolated KS3 (A), one of individual bimodule core dimers (B) and one of bimodule core filaments (C).* ***(A)*** *Data set for individual KS3 domains. 2D class averages reveal dimerization via the LINKS interface of the LD instead of the canonical KS dimer interface, as shown by comparison of a 2D class to a projection based on a model of LINKS-bridged KS3 monomers. Particle number, size and variability precluded high-resolution 3D reconstruction.* ***(B)*** *Data set for individual K3DAK4 bimodule cores. The overall bimodule core is highly flexible, but 3D variability analysis-based sorting into particle sets representing either KS3 (EMDB-14795) or KS4 (EMDB-14793) provided individual reconstructions for each KS to 2.9 Å resolution. Validation of sorting based on map analysis is provided in Figure S3.* ***(C)*** *Analysis of the K3DAK4 filament data set. To prevent misalignment due to pseudosymmetry a box-size large enough to fit four adjacent bimodule cores was chosen in the initial particle picking and reconstruction steps. The intermediate resolution overall bimodule core filament map (EMDB-14945) (1) was used for the rigid body refinement of the overall model. Focused refinements on pairs of KS dimers connected via LINKS interaction lead to improved local maps for pairs of KS3 (2) and KS4 (3) dimers at resolutions of 4.0 Å and 3.9 Å, respectively. Masks used in local refinement of pairs of KS3 or KS4 dimers are shown and the region excised from the overall map (1) for representing DH3 domains in the hybrid model in Fig. 2 is schematically indicated by a blue dashed line.* ***(D-G)*** *Local resolution maps, FSC plots and viewing direction distribution plots for individual reconstructions as indicated.*

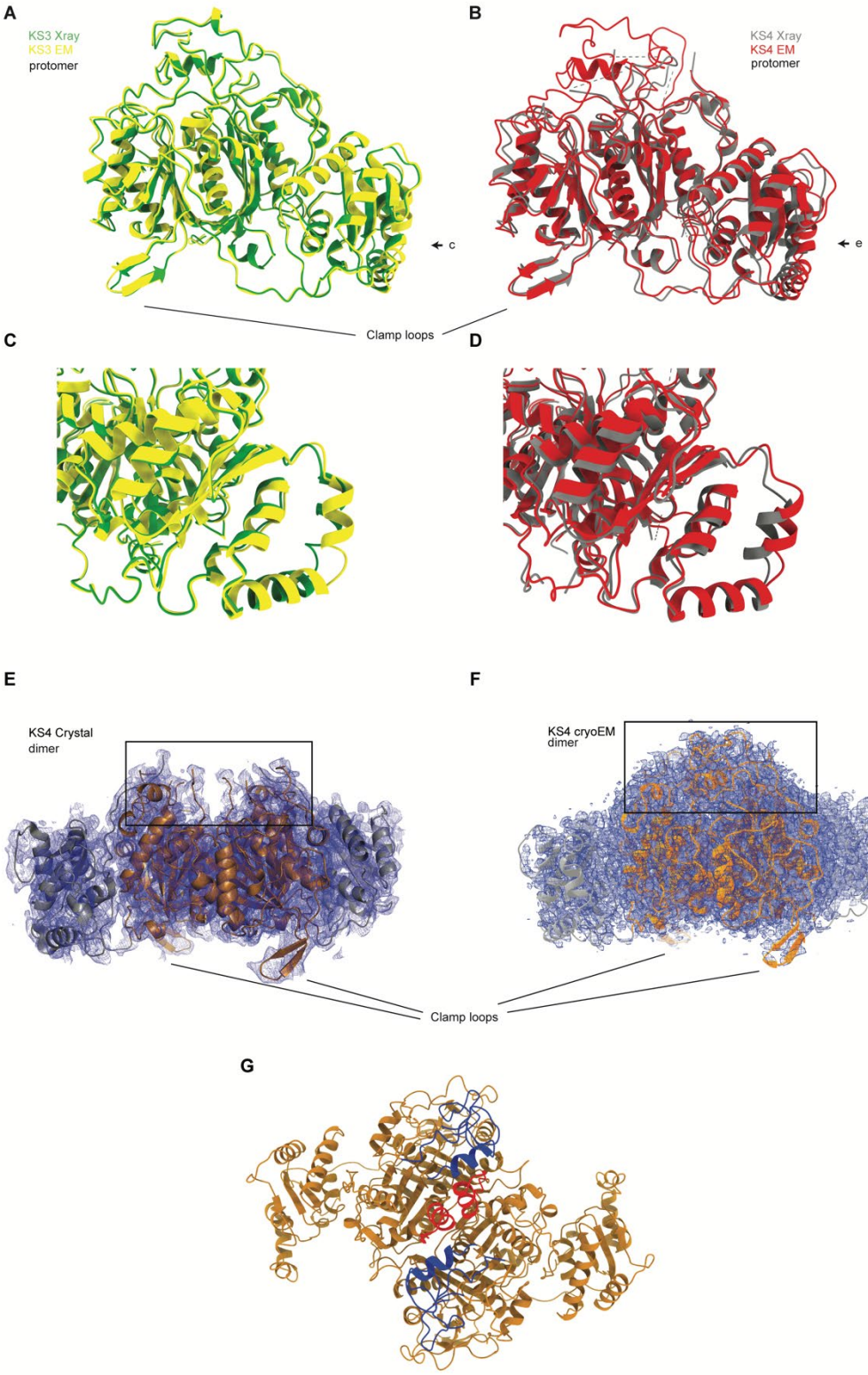

A
KS3 Xray
KS3 EM
protomer
c
B
KS4 Xray
KS4 EM
protomer
e
Clamp loops
C
D
E
KS4 Crystal
dimer
F
KS4 cryoEM
dimer
Clamp loops
G

Fig. S2.
Analysis of KS3 and KS4 domain structures. (A-F) *Comparison of models of KSs derived from X-ray crystallographic analysis of the isolated domains and cryoEM analysis of KS domains in the context of individual bimodule cores for KS3 (A,C; yellow: cryoEM, green: crystallographic) and KS4 (B,D; red: cryoEM, grey: crystallographic)* ***(E,F,G)****, A central region around the active site (indicated by rectangles in (E,F) and red or blue color in (G)) is disordered in crystallographic analysis and not visualized in the crystallographic map at a contour level of 1.7 sigma (E), but is visualized in the cryoEM-derived map at contour level 0.26 in arbitrary units (F). (G) Cartoon style overview representation of KS4 based on cryoEM analysis. Regions shown in blue and red are only visualized in the cryoEM map (F), but not in the crystallographic map (E). The helix colored in red is a 20 AA extension of KS4 relative to KS3.*

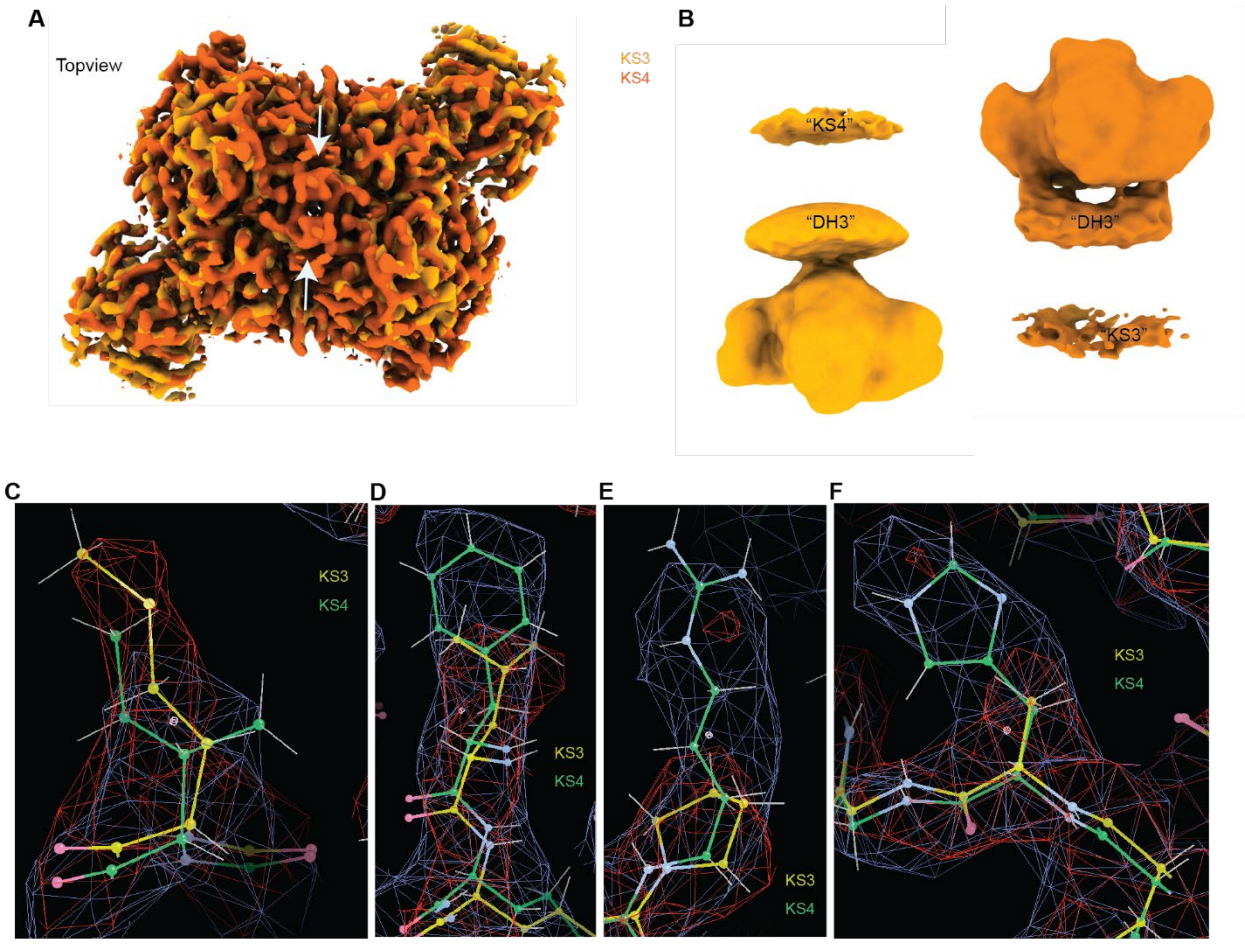


Fig. S3.
Comparison of KS3 and KS4 cryoEM maps and models. (A) *Overlay of the individual maps for the closely-related KS3 and KS4 domains shows an overall highly similar structure. Differences are apparent around the center region close to the C-terminus above the active sites, where a loop (indicated by an arrow) in KS4 is around 20 AA longer than in KS3. In* ***(B)*** *B-factor softened (600 Å^2) maps for KS3 and KS4 reconstructions reveal residual low-resolution densities presumably representing the additional domains of the bimodule core in agreement with their relative positioning to each KS.* ***(C-F)*** *Individual examples of differences in sequence between KS3 and KS4 reflected in the respective cryoEM maps, confirming successful sorting into particles representing either KS3 or KS4. Model and map for KS3 in yellow and red; for KS4 in green and blue respectively.*

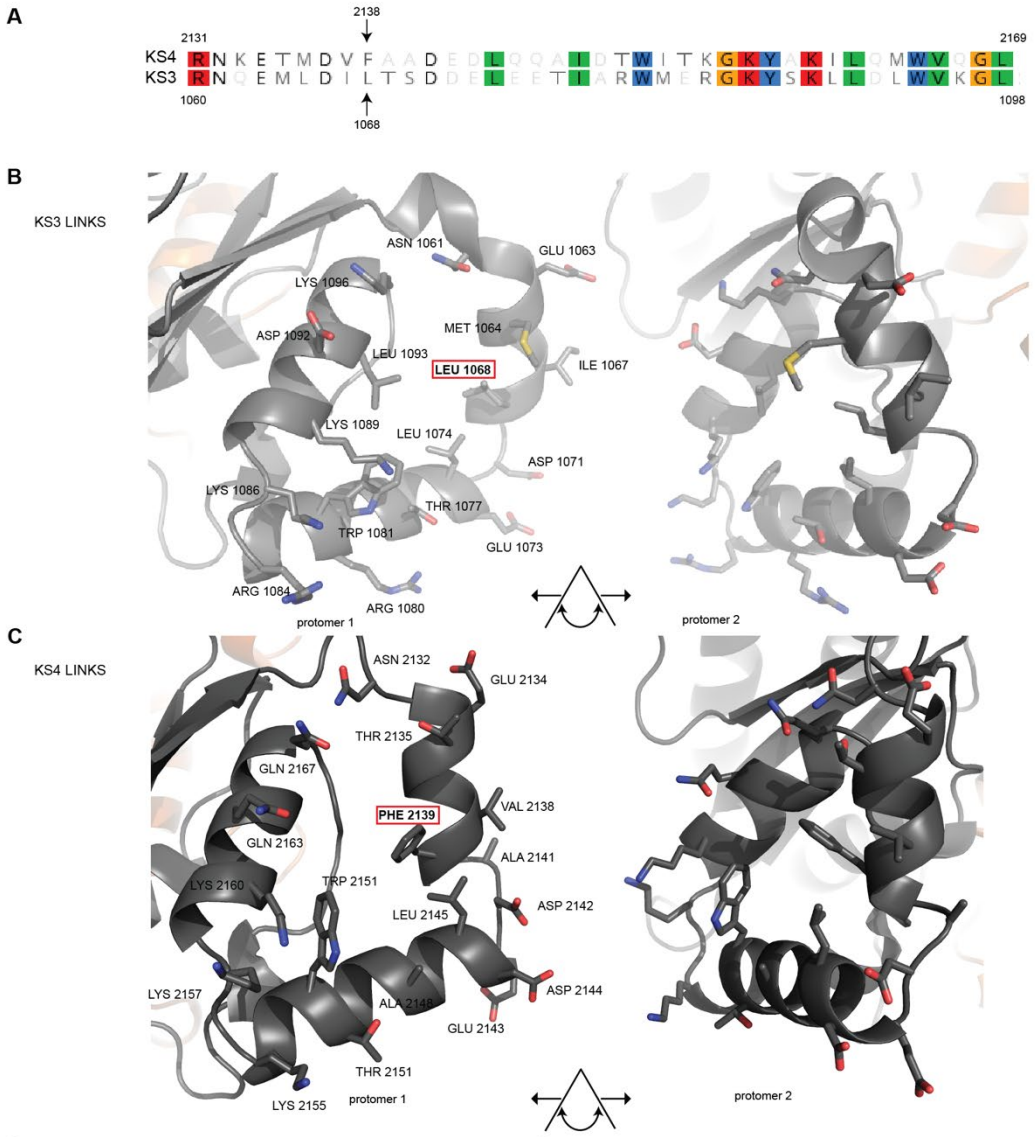

A
2138
2131
2169
KS4
KS3
1060
1068
1098
B
KS3 LINKS
ASN 1061
GLU 1063
LYS 1096
MET 1064
ASP 1092
LEU 1093
LEU 1068
ILE 1067
LYS 1089
LEU 1074
ASP 1071
LYS 1086
THR 1077
TRP 1081
GLU 1073
ARG 1084
ARG 1080
protomer 1
protomer 2
C
KS4 LINKS
ASN 2132
GLU 2134
THR 2135
GLN 2167
PHE 2139
VAL 2138
GLN 2163
ALA 2141
LYS 2160
TRP 2151
LEU 2145
ASP 2142
LYS 2157
ASP 2144
GLU 2143
THR 2151
LYS 2155
protomer 1
protomer 2
D
E

Fig. S4.
Structural analysis of KS3 and KS4 LINKS interfaces. (A) *Sequence alignment of the three-alpha helical LINKS motifs in KS3 and KS4.* ***(B,C)*** *LINKS interfaces of KS3 (B) and KS4 (C) as observed in crystal structures. For visualization, the interfaces are opened up by rotation and translation. The overall three-alpha helical fold is similar for KS3 and KS4, but variations of individual residues such as LEU1068 in KS3 matching PHE2139 in KS4 (indicated also in (A)) ensure that steric clashes prevent mispairing between the KS domains of module 3 and 4.* ***(D,E)*** *Structural superimposition of the KS domains KS3 (light orange) and KS4 (dark orange) (reported here), isolated KS2 from the same BGC11 (green, PDB 4Z37), KS9 of the bacillaene cluster in Bacillus amyloliquifaciens (cyan, PDB 6MHK) and KS6 of the bacillaene cluster from Bacillus subtilis (purple, PDB 5ERF), the overall RMSDs (Cα) are below 1.1 Å. (D) Side view and (E) view onto the LINKS interface of superimposed KS domains.*

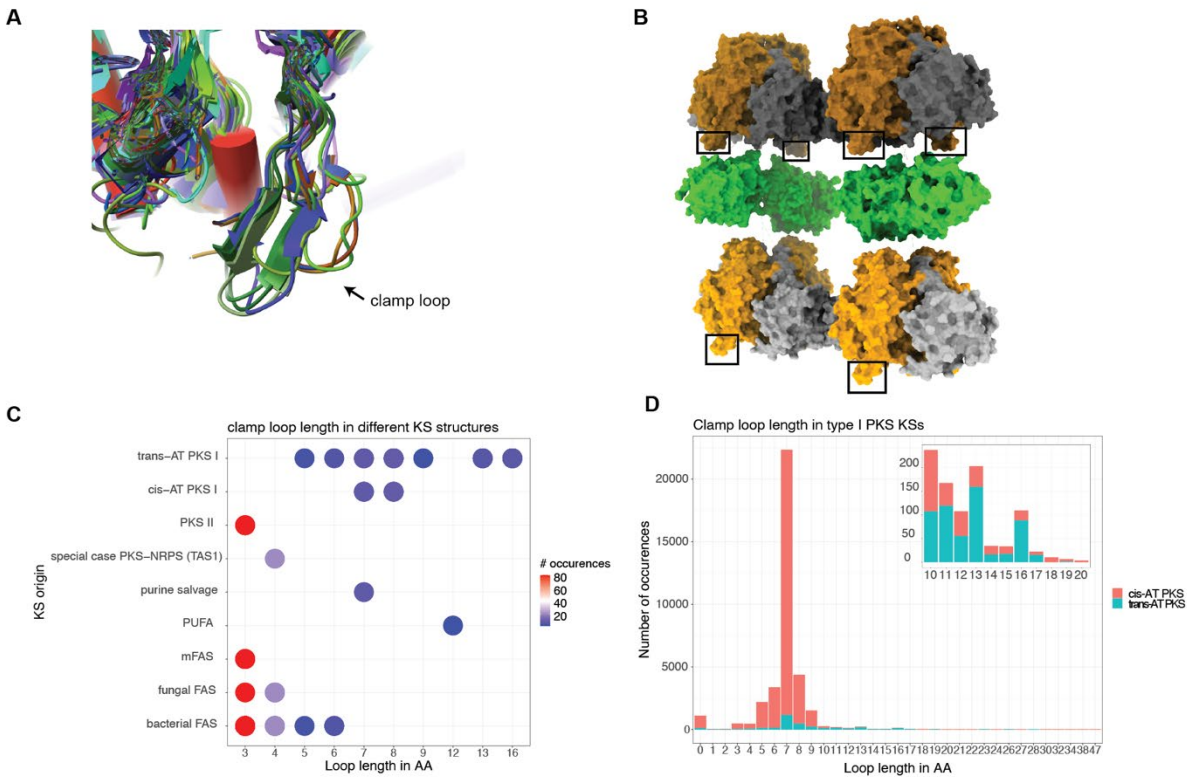


Fig. S5.
Analysis of clamp loops in KS domains. (A) *Zoom-in view onto the clamp loop region of structurally characterized KSs (listed in Data S1) superimposed on KS3. The clamp loop is positioned between residues 920 and 936 in KS3, locating it just before the core β-strand 13.* ***(B)*** *Localization of the KS clamp loops in the assembled K3DAK4 filament (two clamp loops for KS3 are obscured from vision by other parts of the molecule).* ***(C)*** *Visualization of clamp loop lengths for different classes of KS domains derived from the structural comparison shown in (A); only KS domains of trans-AT PKSs exhibit long clamp loops (13/16 residues), but short clamp loops also occur in trans-AT PKS. In type II PKS, mammalian and fungal fatty acid synthases, the clamp loop is completely absent. Most cis-AT PKS KSs feature a shorter clamp loop, similar to that of short clamp loops in trans-AT PKS KSs.* ***(D)*** *Analysis of clamp loops across a set of ~36'500 type I PKS KS sequences aligned using ClustalOmega [125]. There is a distinct preference towards loops of 7 AAs length in both cis-and trans-AT KSs. Longer loops are less common in both systems; however, they are much more common in trans-AT PKSs than in cis-AT PKSs. Sequences and their classification as cis- or trans-AT are derived from "Antismash" [124] and have not been checked individually for misclassifications. The insert shows a zoom of the 10-20 AA loop-length region.*

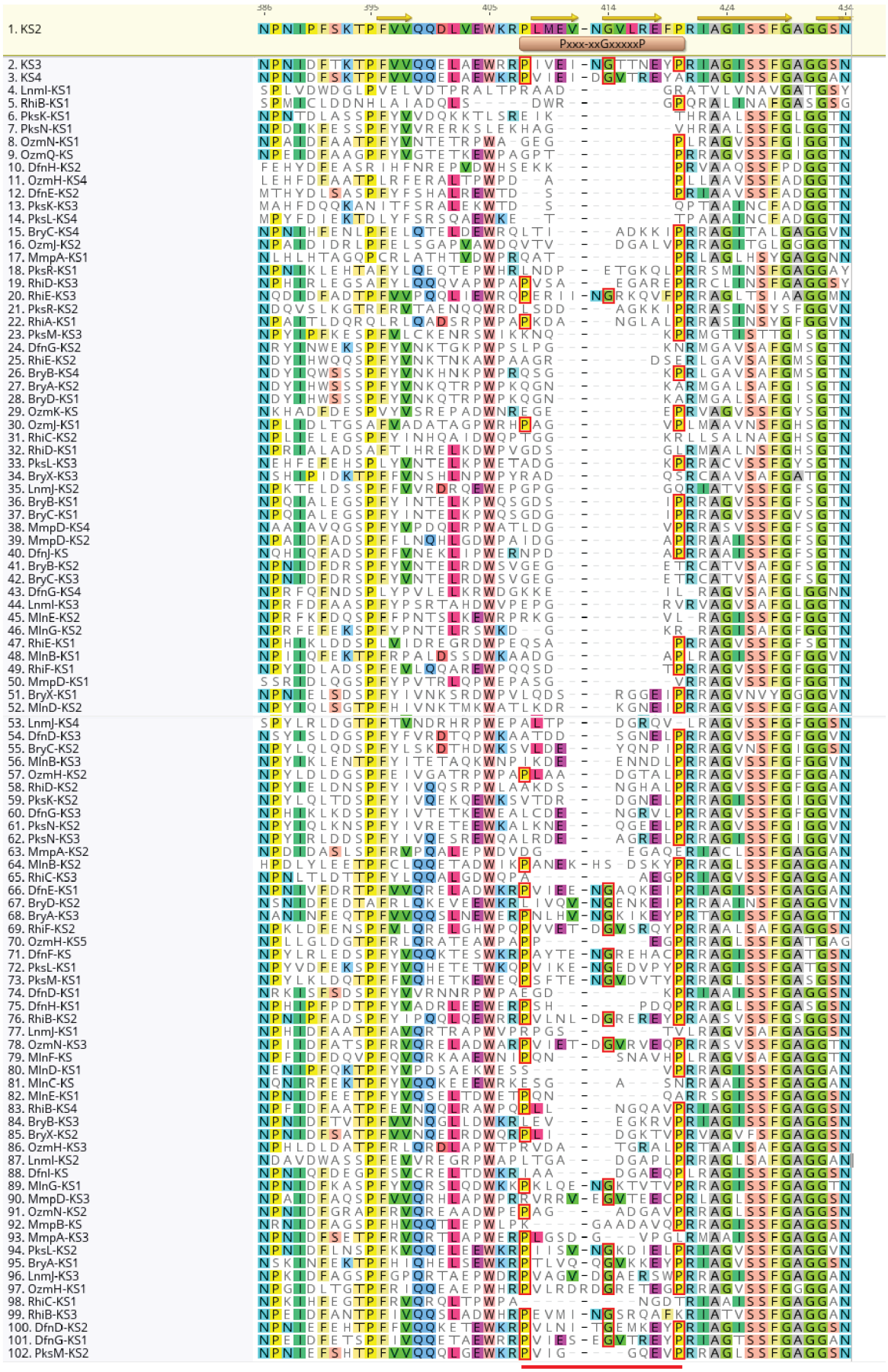

386
395
405
414
424
434
1. KS2
Pxxx-xxGxxxxxP
2. KS3
3. KS4
4. LnmI-KS1
5. RhiB-KS1
6. PksK-KS1
7. PksN-KS1
8. OzmN-KS1
9. OzmQ-KS
10. DfnH-KS2
11. OzmH-KS4
12. DfnE-KS2
13. PksK-KS3
14. PksL-KS4
15. BryC-KS4
16. OzmJ-KS2
17. MmpA-KS1
18. PksR-KS1
19. RhiD-KS3
20. RhiE-KS3
21. PksR-KS2
22. RhiA-KS1
23. PksM-KS3
24. DfnG-KS2
25. RhiE-KS2
26. BryB-KS4
27. BryA-KS2
28. BryD-KS1
29. OzmK-KS
30. OzmJ-KS1
31. RhiC-KS2
32. RhiD-KS1
33. PksL-KS3
34. BryX-KS3
35. LnmJ-KS2
36. BryB-KS1
37. BryC-KS1
38. MmpD-KS4
39. MmpD-KS2
40. DfnJ-KS
41. BryB-KS2
42. BryC-KS3
43. DfnG-KS4
44. LnmI-KS3
45. MlnE-KS2
46. MlnG-KS2
47. RhiE-KS1
48. MlnB-KS1
49. RhiF-KS1
50. MmpD-KS1
51. BryX-KS1
52. MlnD-KS2
53. LnmJ-KS4
54. DfnD-KS3
55. BryC-KS2
56. MlnB-KS3
57. OzmH-KS2
58. RhiD-KS2
59. PksK-KS2
60. DfnG-KS3
61. PksN-KS2
62. PksN-KS3
63. MmpA-KS2
64. MlnB-KS2
65. RhiC-KS3
66. DfnE-KS1
67. BryD-KS2
68. BryA-KS3
69. RhiF-KS2
70. OzmH-KS5
71. DfnF-KS
72. PksL-KS1
73. PksM-KS1
74. DfnD-KS1
75. DfnH-KS1
76. RhiB-KS2
77. LnmJ-KS1
78. OzmN-KS3
79. MlnF-KS
80. MlnD-KS1
81. MlnC-KS
82. MlnE-KS1
83. RhiB-KS4
84. BryB-KS3
85. BryX-KS2
86. OzmH-KS3
87. LnmI-KS2
88. DfnI-KS
89. MlnG-KS1
90. MmpD-KS3
91. OzmN-KS2
92. MmpB-KS
93. MmpA-KS3
94. PksL-KS2
95. BryA-KS1
96. LnmJ-KS3
97. OzmH-KS1
98. RhiC-KS1
99. RhiB-KS3
100. DfnD-KS2
101. DfnG-KS1
102. PksM-KS2

Fig. S6.
Sequence comparison of the clamp loop region across a subset of KS domains. *KS domains with long clamp loops (13-16AA) from trans-AT PKSs show a recurring PxxxxxGxxxxP motif, presumably linked to ß-hairpin formation.*

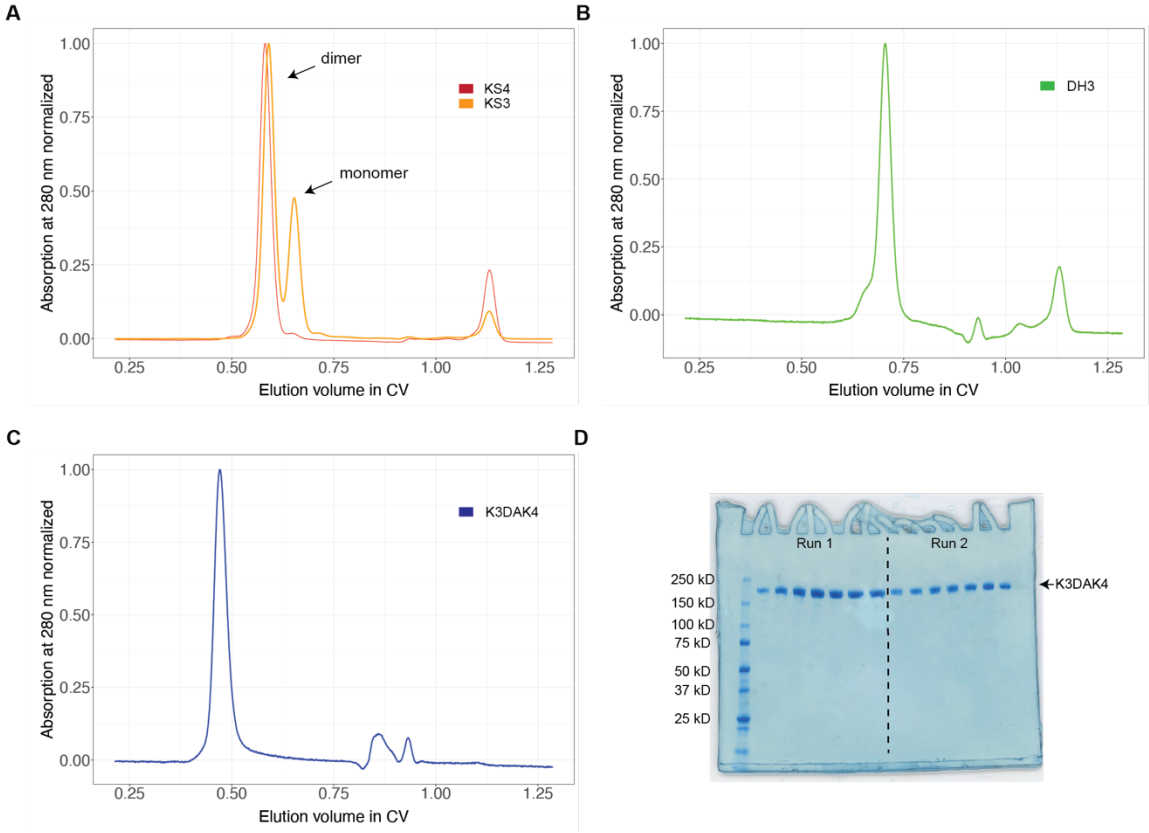


Fig. S7.
Size-exclusion chromatography of KS3, KS4, DH and the intact K3DAK4 bimodule core. *Chromatography was performed on an Superdex 200 increase column at 23°C.* ***(A)*** *Overlay of KS3 and KS4 elution. KS4 elutes as a stable dimer, whereas KS3 elutes in a monomer-dimer equilibrium.* ***(B)*** *Monodisperse elution of DH3. (****C****) The K3DAK4 bimodule elutes as a monodisperse dimer.* ***(D)*** *Uncropped SDS-PAGE analysis of two individual size-exclusion chromatography runs on an Superdex 200 column of a K3DAK4 purification.*

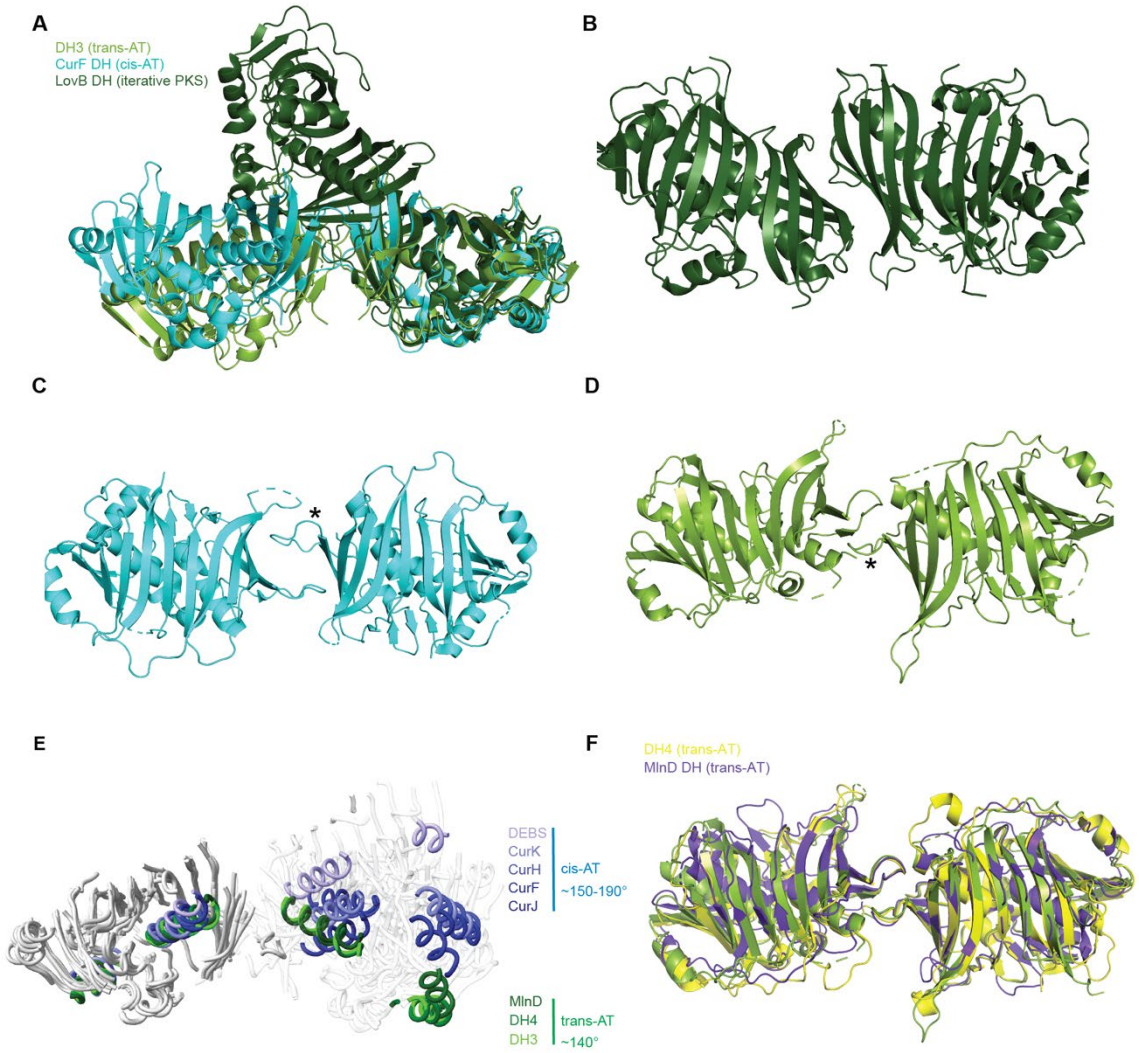


Fig. S8.
DH dimerization interfaces in different types of PKS. *The trans-AT PKS DH3 from BGC11 (reported here) is shown in green, the DH from the cis-AT PKS CurF (PDB 3KG6) in cyan and the DH of the iterative PKS LovB (PDB 7CPX) in dark green.* ***(A)*** *Comparison of all three DH dimers based on superposition of the right protomer. The two cis-and trans-AT PKS DH dimers exhibit similar interdomain angles, clearly distinct from that of the iterative PKS LovB.* ***(B-D)*** *Depiction of the individual dimerization interfaces. While cis-and trans-AT PKS rely on the N-terminal loop to dimerize, the iterative LovB PKS dimerizes via interactions of the beta-strands of the double-hotdog fold. Asterisks in (C) and (D) depict the N-terminal loop of the aligned protomer in DH3 and CurF DH. cis-AT DHs dimerize via their N-terminal loops docking to reach β1-2 and β5-6 from one side, whereas the loops of trans-AT PKS DHs approach from the other side.* ***(E)*** *Shape comparison of DH dimer from cis- and trans-AT PKS. DH dimers are superimposed based on the left protomer. Superpositioning, visualization and calculation of interdomain angles is based only on ordered secondary structure elements. Two conserved helices in each protomer are shown in color to highlight differences in protomer orientation, these helices are colored in shades of blue for cis-AT PKS and in shades of green for trans-AT PKS DH domains as indicated. (F) Comparison of the dimerization mode of three trans-AT PKS DH domains, DH3 (green), DH4 (yellow, PDB 5HQW) of BGC11 studied here, and MlnD (purple, PDB 5IL5). DH dimers are superimposed based on the left protomer only. The mode dimerization is highly similar for all three DH domains.*

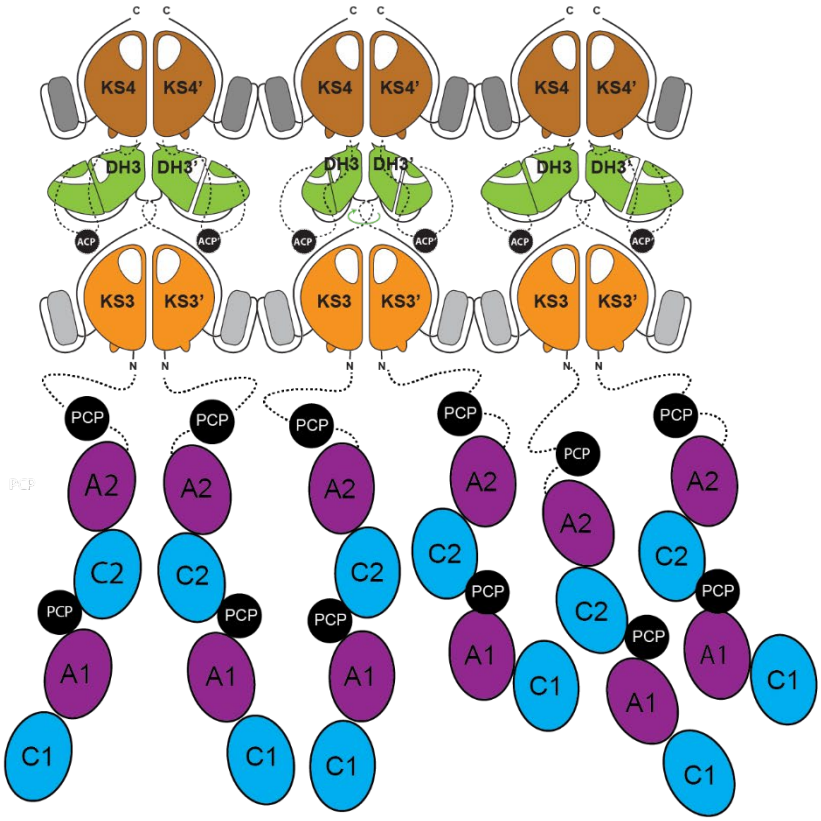


Fig. S9.
Schematic model of a hybrid NRPS-PKS assembly line. *C represents the NRPS condensation domain, whereas A refers to the adenylation domain of NRPS and PCP to the peptidyl carrier protein of NRPS. While large conformational rearrangements are common in NRPS and necessary for catalytic activity, the PKS meshwork restrains conformational space.*

Table S1.
X-ray crystallographic data collection and refinement statistics

	KS3 PDB 7ZM9	DH3 PDB 7ZMF	KS4 PDB 7ZMC
Data collection			
Space group	C 1 2 1	P 21 21 2	C 2 2 21
Cell dimensions			
a, *b*, *c* (Å)	126.78 92.69 99.27	72.38 197.36 39.17	73.56 191.79 293.81
α, β, γ (°)	90 92.54 90	90 90 90	90 90 90
Resolution (Å)	60.48 - 1.62 (1.66 - 1.62)[1]	48.68 - 2.21 (2.26 - 2.21)[1]	48.97 - 3.10 (3.20 - 3.10)[1]
R_{sym} or R_{merge}	4.8 (171.2)	14.68 (205.2)	28.3 (460.4)
$I / \sigma I$	16.65 (1.01)	14.37 (1.33)	8.44 (0.53)
Completeness (%)	99.70 (99.40)	99.0 (98.30)	99.55 (98.11)
Redundancy	6.80 (7.00)	13.39 (13.69)	13.61 (13.66)
CC1/2	100.0 (62.5)	99.9 (66.7)	99.9 (41.9)
Refinement			
Resolution (Å)	60.48 - 1.62 (1.66 - 1.62)	48.68 - 2.21 (2.29 - 2.21)	48.97 - 3.10 (3.21 - 3.10)
No. reflections	145,063	28,951	38,108
R_{work} / R_{free}	0.15/ 0.17	0.22/ 0.25	0.26/ 0.29
No. atoms			
Protein	9,260	8,510	16,679
Ligand/ion	110	16	0
Water	518	78	0
B-factors (Å^2)			
Protein	53.8	58.88	125.72
Ligand/ion	76.28	68.17	-
Water	54.87	55.85	-
R.m.s. deviations			
Bond lengths (Å)	0.012	0.005	0.004
Bond angles (°)	1.54	0.60	0.67
Clashscore	2.62	2.10	5.6

1 Dataset is derived from a single crystal.
*Values in parentheses are for highest-resolution shell.

Table S2.
Cryo-EM data collection, refinement and validation statistics

	KS3 EMDB-14795 PDB 7ZMD	KS4 EMDB-14793 PDB 7ZMA	K3DAK4 EMDB-14945 PDB 7ZSK
Data collection and processing			
Magnification	130,000	130,000	130,000
Voltage (kV)	300	300	300
Electron exposure (e–/Å^2)	67	67	67
Defocus range (μm)	1.5-2.5	1.5-2.5	0.8-1.8
Pixel size (Å)	1.058	1.058	2.116
Symmetry imposed	C2	C2	C1
Initial particle images (no.)	3,332,737	3,332,737	1,288,160
Final particle images (no.)	182,315	171,149	293,391
Map resolution (Å)	2.93	2.90	6.80
FSC threshold	0.143	0.143	0.143
Map resolution range (Å)	4.5-2.5	4.5-2.5	6.0-14.5
Refinement type	Coordinate	Coordinate	Rigid-body
Initial model used (PDB code)	Xray KS3	Xray KS4	Combined
Model resolution (Å)	3.10	3.20	10.4
FSC threshold	0.5	0.5	0.5
Map sharpening *B* factor (Å^2)	90.8	94.7	387.4
Model composition			
Non-hydrogen atoms	9,294	9,484	46,128
Protein residues	1,191	1,214	5,916
Ligands	0	0	0
B factors (Å^2)			
Protein	32.28	40.90	735.96
Ligand	-	-	-
R.m.s. deviations			
Bond lengths (Å)	0.002	0.003	
Bond angles (°)	0.522	0.894	
Validation			
MolProbity score	1.27	1.61	
Clashscore	3.75	4.52	
Poor rotamers (%)	0.00	0.00	
Ramachandran plot			
Favored (%)	97.47	94.46	
Allowed (%)	2.53	5.54	
Disallowed (%)	0.00	0.00	

Movie S1.
Overview of the bimodule core filament maps and model. Blend over between experimental map and model-based cartoon and surface representations in different views with transition to hypothesized 2D-mesh formation, as described in the main text.

Data S1.
Analysis of clamp loop length based on experimental KS structures. Includes pdb id, type of the BGC, organism, clamp loop length and boundary residues. Structures are shown in Fig S5 A and clamp loop length distribution is visualized in Fig S5 C. For selection of clamp loop boundaries KS3 residues 920 and 936 were used as reference.

Acknowledgements
We thank Janine Brunner and Raimund Dutzler for the donation of the FX-cloning plasmids used in this study (p7XNH3, addgene #47064), and EMBL Heidelberg for providing the pETG10A vector, M. Chami, K. Goldie and C. Alampi from the University of Basel BioEM facility as well as I. Mohammed for their help in operating the electron microscopes, A. Mazur for help in coding the sequence comparisons and the National Institute of Technology and Evaluation Biological Resource Center for providing the Brevibacillus Brevis NBRC 100599 strain. Calculations were performed at sciCORE (http://scicore.unibas.ch/) scientific computing center at University of Basel.

Funding
Swiss National Science Foundation project funding 179323 and 159696 to T.M.
Swiss National Science Foundation R'EQUIP grant 177084 to T.M.
AL4BIOCH project of the H2020-MSCA-IF-2018 to H.M-H.

Author contributions
Conceptualization: DAH, TM
Methodology: YUT, DAH, RPJ, TM
Software: YUT
Investigation: YUT, HMH, SFXM, RPJ
Visualization: YUT
Project administration: TM
Supervision: RPJ, TM
Writing – original draft: YUT, DAH, TM
Writing – review and editing: YUT, TM

Competing interests
The authors declare no competing interests.

Code availability
All custom code is available under the Zenodo DOI provided in the methods section.

Data availability
All data used in this work has been deposited in the pdb under 7ZM9 (KS3, X-ray crystallography), 7ZMA (KS4, cryoEM), 7ZMC (KS4, X-ray crystallography), 7ZMD (KS3, cryoEM),7ZMF (DH3, X-ray crystallography) and 7ZSK (K3DAK4, cryoEM). CryoEM maps have been deposited in the emdb under EMD-14795 (KS3), EMD-14793 (KS4) and EMD-14945 (K3DAK4).

Supplementary Materials
Supplementary Materials containing Figs. S1 to S9 and Tables S1 to S2

Supplementary Data S1

Supplementary Movie S1

Chapter 3

Statement of own contribution

All work regarding the crystallization, X-ray data collection and interpretation of crystallographic data was performed by me, whereas protein expression, chemical synthesis of the ligand and crosslinking thereof were performed by the collaborating group of Craig Townsend at Johns Hopkins University. Text and figures relating to protein expression, chemical synthesis of the ligand and crosslinking thereof have been provided by Jamie Alley.

Introduction

Polyketide synthases (PKSs) produce a vast variety of bioactive secondary metabolites in both bacteria and fungi to afford the host organisms a competitive advantage in their environmental niche. Some of these secondary metabolites and derivatives thereof are in clinical use as antibiotics, immunosuppressants and anti-cancer drugs [15]. Type I PKSs are homodimeric multienzymes, comprised of two classes, modular type I PKSs (modPKSs) and iterative type I PKSs (iPKSs)[16]. ModPKSs use each enzymatic domain once in their product synthesis, allowing for collinearity between their operon and the final product [16]. In contrast iPKSs use the same enzymatic domain iteratively to produce a final product. Specialization comes from specific starter units as well as programmed use of each enzyme during the iterative cycles [16]. In non-reducing iPKS (NR-iPKS) specialized enzymes, such as the product template domain (PT), perform specific cyclizations to reach the final product [73].

One of the best studied NR-iPKS is PksA, which produces the aflatoxin B_1 precursor norsolorinic acid in *Aspergillus parasiticus* and *Aspergillus flavus* [44]. Both *strains are* concerning contaminants in food production due to their prevalence combined with the highly carcinogenic nature of aflatoxin B_1[126]. The PksA protein consists of several functional domains: an acyl carrier protein (ACP), the only non-enzymatic domain, a starter unit:ACP transacylase (SAT), a malonyl-CoA:ACP transacylase (MAT), a ketosynthase (KS) domain, a PT and a thioesterase (TE) domain[73] . The KS extends the growing ketide chain by catalyzing a decarboxylative Claisen condensation. It uses two types of building blocks: The starter substrate is a hexanoyl unit produced by the dedicated fungal fatty acid synthase system HexA/B, and the elongation substrate is malonyl-CoA; both are delivered as thioesters attached to the ACPs covalent cofactor phosphopantetheine. The ACP is loaded with the substrates by the SAT, for the starting substrate, and by the MAT for the seven elongation building blocks, respectively [73] (Fig 3.1).

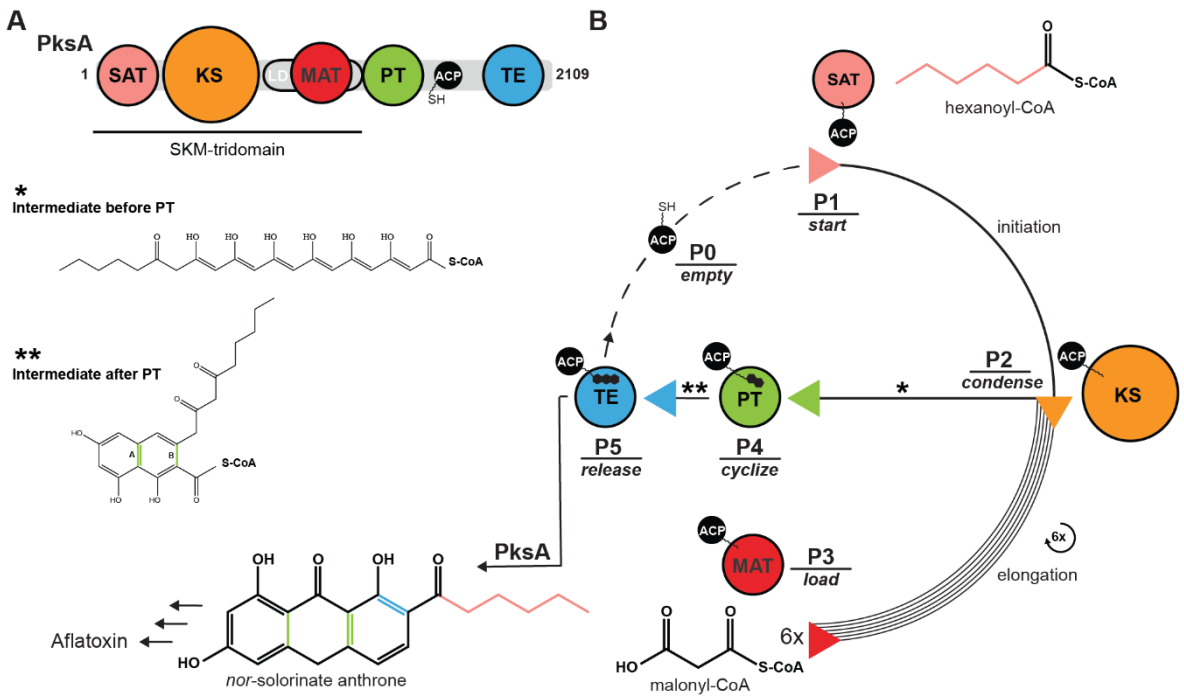


Figure 3.1: Overview of PksA's protein layout and enzymatic function. *(**A**) As an iPKS, PksA uses it's enzymatic domains iteratively to achieve product synthesis. (**B**) Schematic depiction of PksA's reaction cycle. First the SAT loads the ACP with hexanoyl, which is transferred onto the KS, followed by six rounds of chain extension by malonyl, which is loaded onto the ACP by the MAT. Having reached the programmed length, the PT cyclizes the substrate in two steps, indicated as A,B in green on both the ** intermediate and the final product. Finally, the TE cleaves the product from the ACP, leading to the final ring closure (indicated in blue on nor-solorinate anthrone).*

Upon reaching the programmed 20-carbon extension the highly reactive poly-beta-ketone substrate is stereo specifically and selectively cyclized twice by the PT followed by cleavage of the substrate from the ACP by the TE resulting in norsolorinic acid anthrone[73] (Fig 3.1). PksA PT follows the so-called F-pattern cyclization[73], yielding in two steps first a C4-C9 and then a C2-C11 cyclized product. The PT has a double hot-dog (DH) fold similar to that of dehydratase domains in other PKSs[73]. Despite the conserved fold, DHs are known for a variety of distinct enzymatic functions[16,40,41,98]. Additionally, DH domains are known to have divergent modes of dimerization [127].

Given the consistent fold and varying function a specific positioning of the substrate is paramount to achieving consistent and specific substrate cyclization. Three models of PksA PT cocrystalized with different ligands have been determined (PDB ID: 3HRR, 3HRQ and 5KBZ) [44,73], however no ligand-free structure is known. The models mimic the initial substrate **3** bound state (PDB: 5KBZ) and the final substrate analogue **2** (PDB: 3HRR), while one structure contains a palmitate ligand **1** (PDB: 3HRQ). Chemically, one of the key factors precluding easy observation is the highly reactive nature of its native substrate, a poly-beta-ketone **4**, making it unsuitable for structural studies and too instable for chemical ones. Both the initial substrate mimic bound, and the final product bound state afford a glimpse at the PTs enzymatic function, yet the mechanism of the two consecutive cyclization steps is still not fully understood. Specifically, the N, O and S heteroatoms of the previous bis-isoxazole substrate analog **3** (PDB ID: 5KBZ) were intended to mimic the alternating placement of carbonyl oxygens of the ACP-bound polyketide intermediate **4** of PksA biosynthesis (Fig 3.2). In its x-ray structure bound in the PT active site ionic interaction of the terminal *O*-

phosphate with the surface residues Arg1623 and Lys1396 anchor one end of the substrate mimic to array the two heterocycles deeper into the active site. The C–S–C bond angles approximate 90° and the rotational constraints of two 5-membered rings bring into question how faithful a structural mimic the bis-isoxazole is of the true poly-b-ketone.

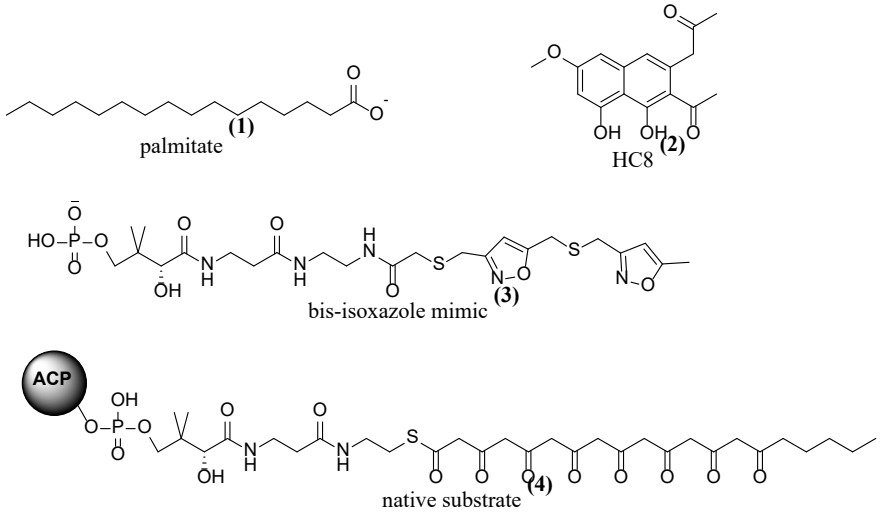


Figure 3.2 Native substrate and previous mimics. *The three previous mimics cocrystalized into the PT monodomain compared to the native linear substrate.*

To apply the isoxazole as a simpler, less constrained diketone surrogate and to gain insight into the intramolecular aldol reaction catalysed by the PT domain, the whole conception of a polyketide intermediate mimic was reimagined. Rather than trying to imitate the poly-β-ketone motif, it was realized that the first-ring intermediate **5** would have to be favourably bound to enable the second cyclization to occur. The decision to incorporate this initial benzene ring would confer greater chemical stability and more ready synthetic access to a more accurate and potentially useful probe of its function.

Here we design and synthesize an isoxazole ligand with a benzene ring mimicking the product of the first cyclization step, characterize its crosslinking biochemically and investigate a 2.7 Å X-ray crystallographic model of it crosslinked to PksA PT to obtain a snapshot at an intermediate step in the cyclization to elucidate the PTs enzymatic function.

Results & Discussion

To trap the first cyclization of PksA-PT we designed a new isoxazole ligand, following and improving upon the bis-isoxazole analogue used previously (PDB 5KBZ). Beyond their hoped-for ability to mimic structurally a di- or poly-β-ketone, it has been long known that isoxazoles can be reductively cleaved to β-iminoketones, which undergo hydrolysis in aqueous acid to diketones. A methylene bridge would, therefore, place the isoxazole, or diketone, in the correct register to with reference to the phenyl ring (Fig 3.4). Then to trap the new substrate mimic in a covalent complex with the PT, the classic Bloch alkyne thioester mechanism-based inactivator of a DH domain from fatty acid synthesis was further adapted from a modification reported by Burkart[128]. The pKa's of hydrogens adjacent to a thioester or ketone are roughly equivalent (~20).[2] As had Burkart, we replaced the native pantetheine thioester with a hydrolytically stable methylene (–CH_2–). In principle then, when the keto-alkyne **7** binds to the PksA-PT, the catalytic His1346 would deprotonate between the ketone and the alkyne as illustrated in Figure 3.3 and immediately be captured by the resulting keto-allene to form the precedented *E*-alkene adduct **5**[128]. In a final adjustment to the evolving synthetic plan, the two phenol groups were removed to simplify the synthesis and to avoid steric clashes in the reaction chamber visualized in preliminary modelling studies (Fig 3.4).

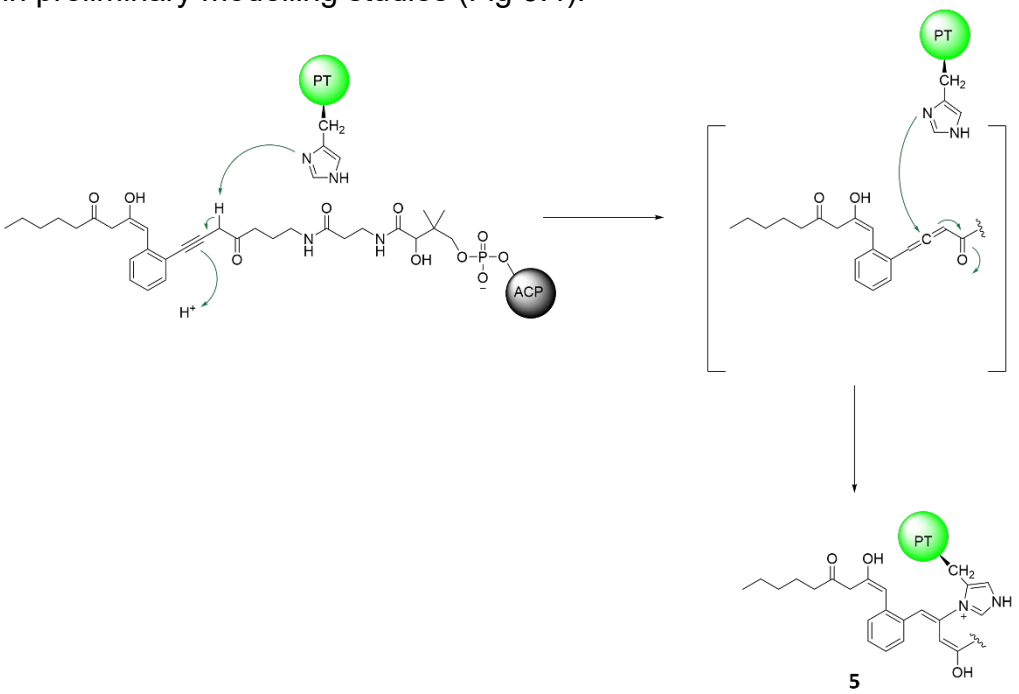


Figure 3.3 DH crosslinker mechanism *FAS DH=ACP covalent crosslinking resulted from histidine deprotonation and subsequent nucleophilic attack in the DH active site.*

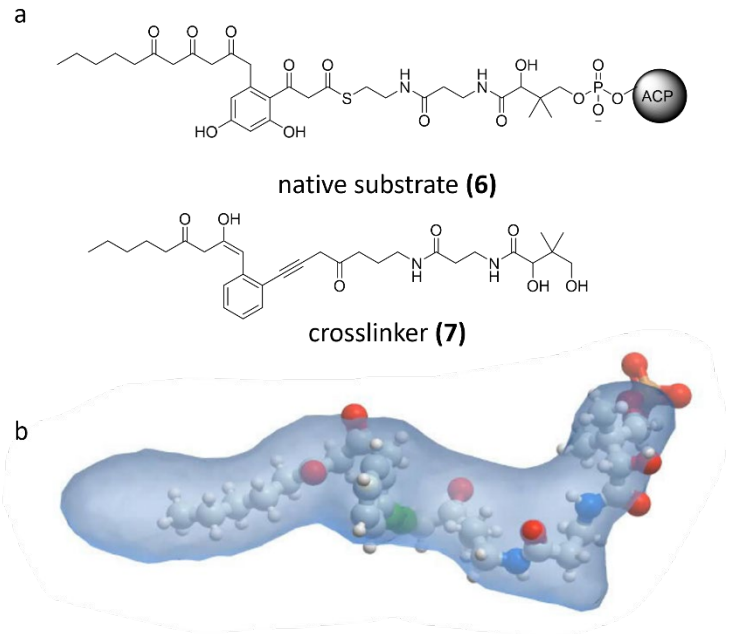


Figure 3.4 Comparison and modeling of target PksA-PT crosslinker. A*, Comparison of crosslinker **7** to the native substrate for the second cyclization. **B**, Modeling of the PT crosslinker **7** in active site pocket. The phenyl ring sits in the reaction chamber and the two arms sit in the appropriate channels.*

The final synthesis of the keto-alkyne inactivator **7** was carried out as summarized in Figure 4.7. 2-Iodo-benzylbromide was reacted with trimethylsilyl (TMS)-acetylide anion generated with ethylmagnesium bromide in the presence of copper(I) bromide. The TMS group was removed with fluoride ion to yield **8**. This terminal alkyne was combined with hexyloxime chloride and base to effect elimination and induce a regiospecific nitrile oxide [3+2]-cycloaddition to afford isoxazole **9** bearing a five-carbon alkyl chain. This length was chosen because earlier experiments had shown that a longer side chain demonstrated poor water solubility and soap-like properties. To complete the keto-alkyne inactivator **7**, the pantetheine-like arm was built up from 5-bromopentene in a quantitative classical Gabriel synthesis and epoxidation of the terminal alkene (Fig 3.5). Again using TMS-acetylide, the epoxide could be opened in the presence of boron trifluoride etherate to **11**. At this point, a modified Sonagashira reaction substituting silver ion for the traditional copper(I) co-catalyst[129] was used to couple the aryliodide **9** and TMS-alkyne **11** to produce **12** in a good yield.

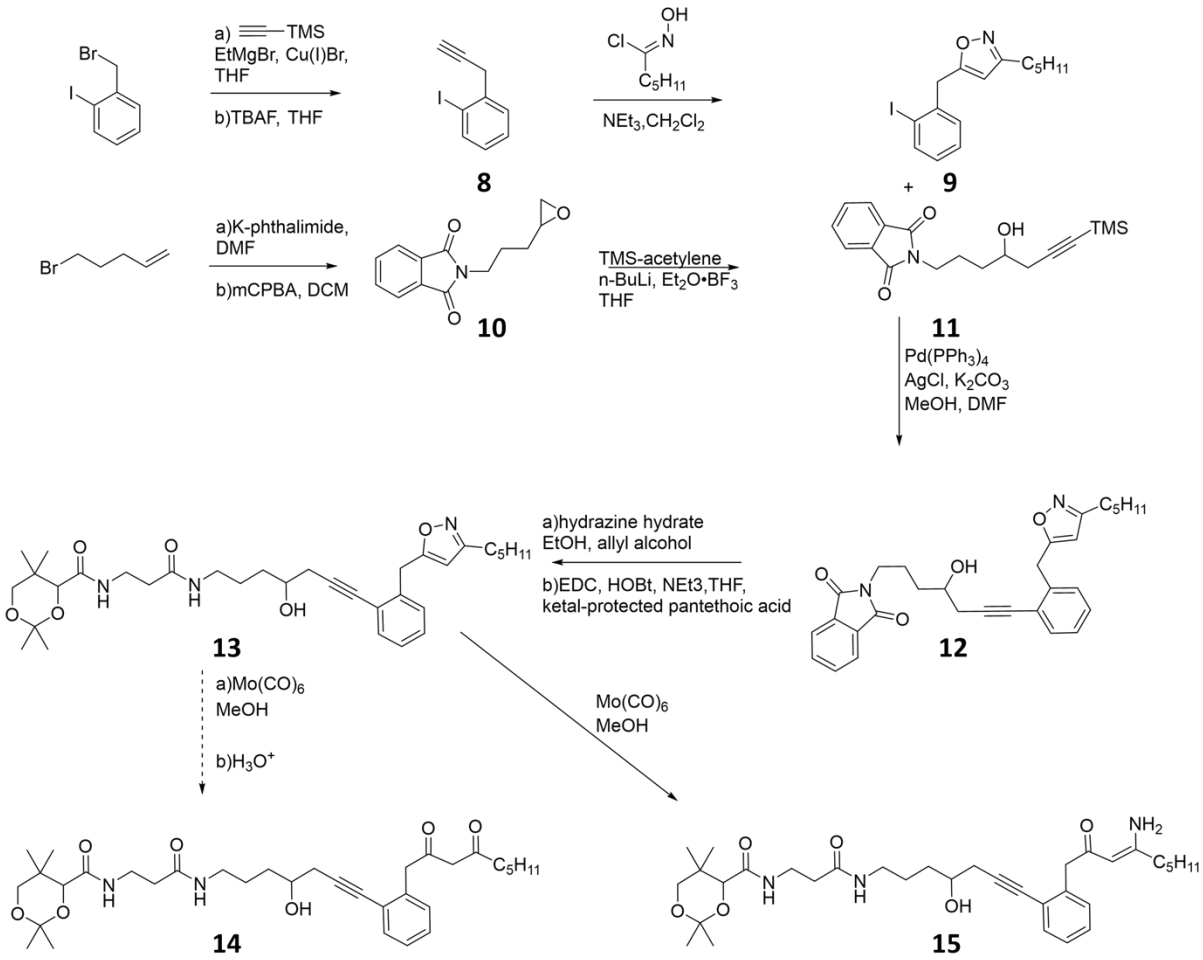


Figure 3.5 PksA-PT crosslinker synthesis.

Finally, the phthalimide amine protecting group was removed with hydrazine hydrate and the free amine was coupled using conventional means with ketal-protected pantethoic acid to give **13**. The isoxazole moiety could be reduced by molybdenum hexacarbonyl[130] to the highly stabilized (intramolecular H-bond) β-keto-enamine **15**, which proved resistant to hydrolysis to the β-diketone **14**. More acidic conditions were attempted but resulted in pantetheine arm degradation and no β-diketone product **14** was observed. Unable to fully unmask the isoxazole ring, Dess-Martin periodinane was used to oxidize the secondary alcohol to the desired keto-alkyne and mild acid ketal deprotection of the pantetheine diol gave **16** with the isoxazole intact (Fig 3.6). Further attempts toward the synthesis of the diketone crosslinker **7** are underway; however, this synthesis to an improved isoxazole mimic **16** allowed biochemical analysis in the PksA-PT domain to proceed immediately.

Figure 3.6 Oxidation and deprotection of isoxazole containing PksA-PT mimic.

PT doping with substrate mimic

To evaluate the necessity of phosphorylation to induce substrate mimic binding to the PksA-PT, as thought to be important from earlier bis-isoxazole experiments,[5] we optionally incubated crosslinker **16** with PanK pantetheine kinase from *Escherichia coli*. Crosslinker **16** had limited solubility in water, producing a cloudy mixture; however, the phosphorylation reaction clarified over time and the crosslinker was fully phosphorylated in six hours at room temperature, which was verified by electrospray ionization mass spectrometry (ESI-MS) in negative mode (M^-= 632.274). Previous work by our lab and others demonstrated[44,73] the necessity of phosphate for binding and co-crystallization. However, we were interested to see if our more accurate substrate mimic would have enough innate binding affinity to overcome the lack of this important ionic interaction. Therefore, we sought to dope both **16** and phospho-**16** into PksA-PT. Each variant of the crosslinker was added in a two-fold excess to the freshly purified PksA-PT. Stoichiometric covalent attachment of **16** to PksA-PT was achieved in two hours at room temperature and confirmed by intact-protein ESI-MS (Fig 3.7). In contrast to the previous mimetics[44], **16** bound and reacted without resin-based pre-treatment of PksA-PT to strip endogenous palmitate from the active site. Even more surprisingly, phospho-**16** did not covalently attach to PksA-PT. We continued to encourage phospho-**16** reaction with higher compound concentrations, DMSO doping and varied pH all to no avail. The protein was completely modified with **16** alone. To confirm the reaction specificity, we exposed the reacted protein to trypsin proteolysis treatment and subsequent MS analysis of the fragments confirmed the adduct was covalently attached to the active site His1346 (Fig 3.7, 89% coverage) as envisioned at the outset. No other modified fragments were detected.

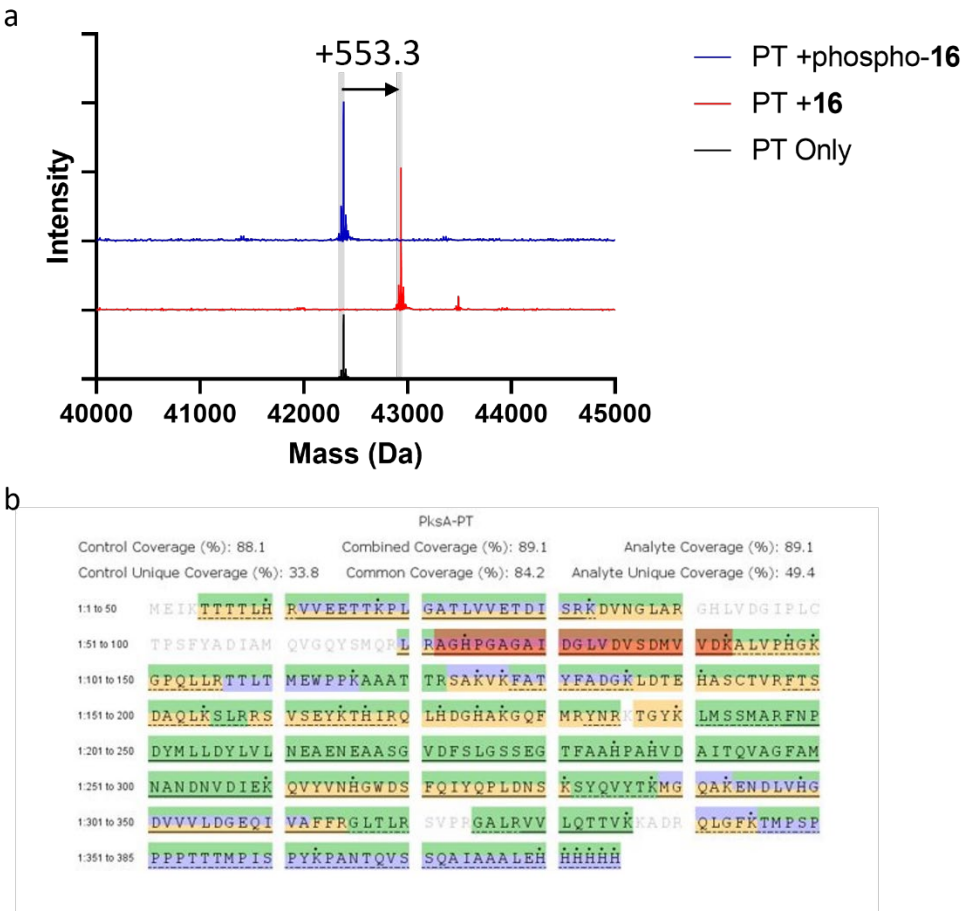


Figure 3.7 PksA-PT mass spectra analysis. A*, The covalent reaction of PksA-PT with* ***16*** *(red), but not phospho-****16*** *(blue) as demonstrated by the intact protein mass calculation of each reaction compared to unreacted control (black).* ***B****, Coverage map of trypsin digestion showed adduct mass added to only red highlighted fragment containing active site His74 in this construct corresponding to His1345 in full-length PksA. Green highlighting indicates fragments observed both in unreacted and reacted PT samples while purple highlighting indicates fragments only detected in unreacted PT and yellow for only reacted PT samples. Gray fragments were not detected in either sample. Dots above the residues indicate possible basic reaction sites calculated for the search.*

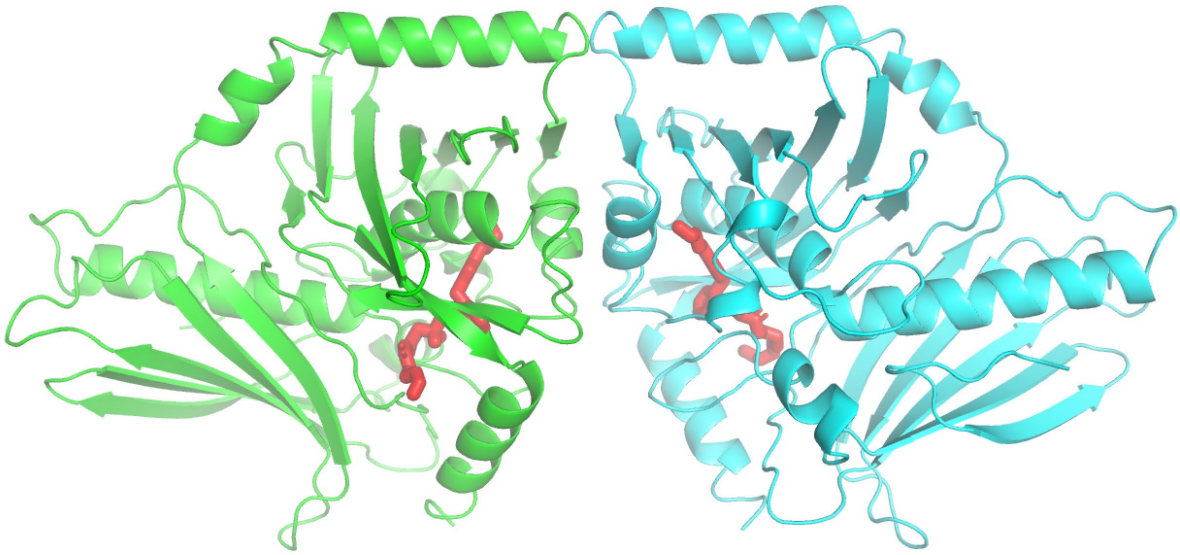


Figure 3.8: Overall model of the 16 crosslinked PksA PT dimer*. Protomers are colored in green and cyan, with the* ***16*** *ligands shown in red stick representation.*

Having crosslinked PksA-PT to the new isoxazole substrate analogue **16** we proceeded to elucidate its structure using X-ray crystallography at 2.7 Å resolution (Fig 3.8). To solve the phase problem, we used the model of PksA-PT with its final substrate analogue (PDB 3HRR) for molecular replacement. The crystal had a C222 pseudosymmetry, with its true symmetry being C2. In total there are four protomers in the asymmetric unit. The PT dimerizes via its canonical interface, the same as in all three other known models. Interestingly, RSMD(C-alpha) between individual models (Ours, PDB: 5KBZ, 3HRR, 3HRQ) is below 1.0 Å, therefore indicating no overall changes regardless of the ligand involved (Fig 3.9). Differences between the models are minor. In the **16-**ligand bound model we added residues forming another helical twist on the C-terminal helix. No residue from the extended regions makes direct contacts to the Phosphopantetheine-arm in the PT active site tunnel. The last three solvent facing carbons of the **16-**ligand were not built as a high degree of flexibility precluded defined placement.

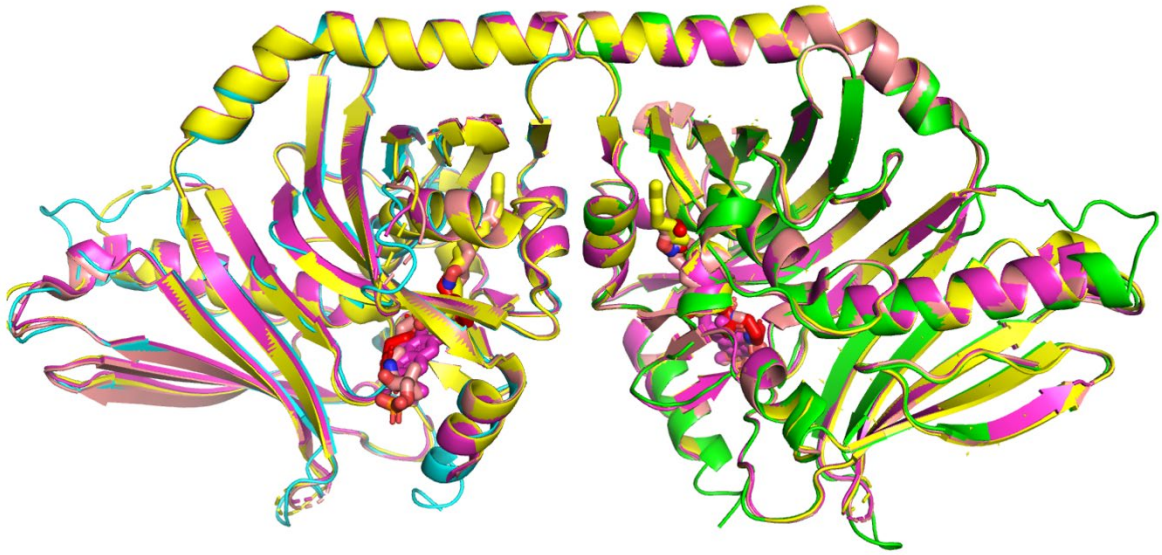


Figure 3.9: Structural comparison of experimental PksA PT models. *All models are aligned on chain A, (left) RMSD of Calpha atoms between individual models is below 1 Å. The* ***16****-crosslinked Pksa PT dimer is shown in green/cyan, In yellow the palmitate* ***1*** *bound state (PDB 3HRQ), in magenta the final substrate analogue* ***2*** *bound state (PDB 3HRR) and in orange bisisoxzaloe* ***3*** *bound state (PDB 5KBZ).*

The crosslinked ligand has well defined density within the active site cavity. Density around the active site His1345 confirms the crosslink (Fig 3.10). In two out of the four chains of our model the ligand has reacted, potentially with surrounding water, forming a gem-diol instead of the originally present carbonyl (Fig 3.10 A-B). When comparing the two instances of the isoxazole ligand present in our model, differences in coordination with their surrounding are minimal, except for the outward most built O16, facing opposite directions which in the case of our carbonyl version means that N11 interacts with Asp1395. Interestingly, when comparing the different ligands within the active site cavity, the bulk of interacting residues appears to interact with the phosphopantetheine arm, as opposed to the actual substrate (Fig 3.11-3.12). The fact that the phosphopantetheine end of the **16-**ligand does not make interactions with the protein and is facing the solvent allows for a large degree of conformational flexibility. Overall, the differences between the protomers in positioning of the phosphopantetheine end of the **16**-ligand, canonically facing towards and being covalently bound to the ACP, can mainly be attributed to weak interactions with surrounding residues and solvent accessibility.

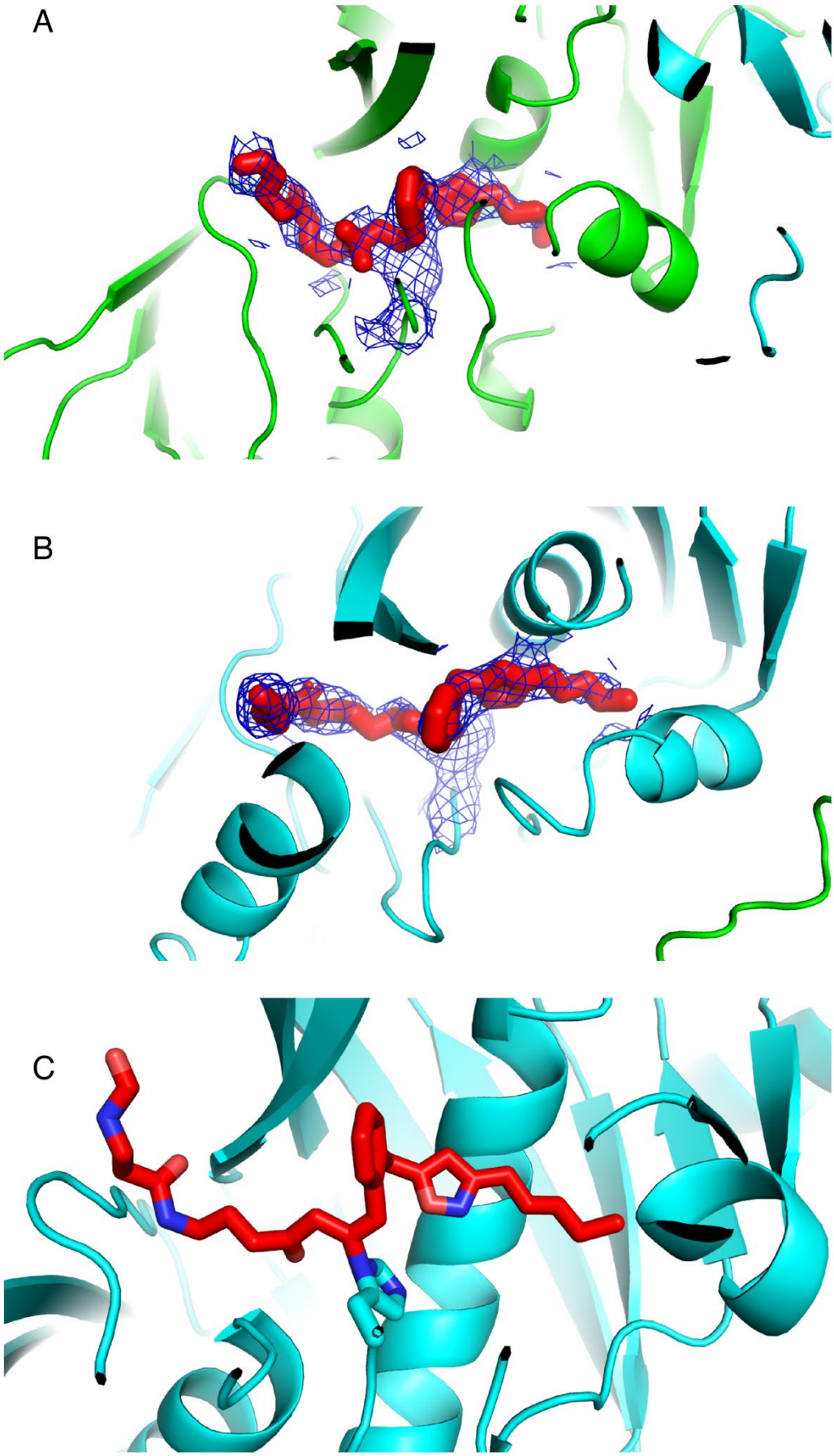

A
B
C

Figure 3.10: Comparison of the two different states of the 16-ligand crosslinked to PksA PT. *Display of the isoxazole **16**-ligand with the surrounding 2(Fo-Fc) electron density at I/sigma(I) = 0.8. (A) displays the gem-diol state of the ligand with clear density linking to the histidine of the crosslink. In (B) density and ligand model for the carbonyl state of the ligand. Similar to (A), strong density for the crosslink is observed. (C) showcases the model of the carbonyl variant of the **16**-ligand crosslinked to the histidine.*

Within the active site cavity, the substrate is surrounded by mostly hydrophobic residues without direct interactions. In the higher resolution structure with the bis-isoxazole bound (PDB: 5KBZ) we only observed indirect interactions via a coordinated water network. We cannot find any similar water networks, granted this might be partly due to the lower resolution of our model. The isoxazole in our model is rotating between three positions as demonstrated in the different chains. Interestingly, the analogous first isoxazole of the bis-isoxazole ligand designed to mimic the initial substrate (PDB ID: 5KBZ), is coordinated with the aforementioned water network in its model and only allows for a single position.

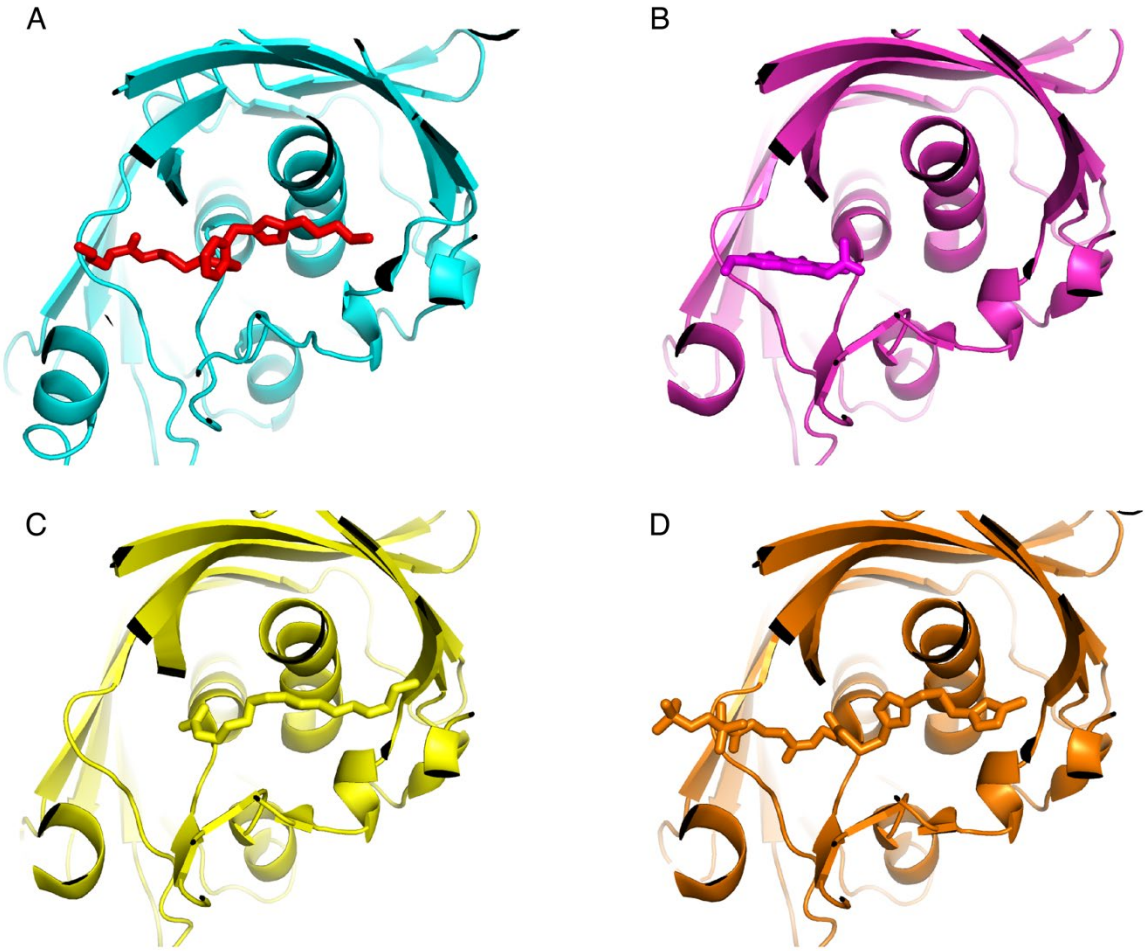


Figure 3.11: Comparison of ligand poses in PksA PT structural models. *The view is consistent allowing for comparison of relative ligand positioning. (A) Isoxazole ligand (this work). (B) PksA PT with final substrate analogue **2** (PDB: 3HRR). (C) PksA PT with palmitate **1** (PDB: 3HRQ). (D) PksA PT with a bisisoxazole mimetic **3** (PDB: 5KBZ).*

Residues Lys1396 and Arg1623 implicated in interacting with the phosphate of the phosphopantetheine arm in 5KBZ, are not directly interacting in our model. As the **16**-ligand does not contain a phosphate this is to be expected. Interestingly however, the ligand is moved outward by ~2-6 Å, as compared to the bis-isoxazole ligand (PDB ID: 5KBZ), the outward movement is related to the placement of the **16**-ligand crosslink, moving the phosphopantetheine arm outward relative to the bis-isoxazole

ligand by 1-2 carbons. While the distance of the outward movement clearly correlates with the position of our crosslink compared to the non-crosslinked ligand in 5KBZ, it also correlates with the unsuccessful crosslinking efforts for the analogue including the phosphate (phospho-**16**) as the outward movement would make the phosphate unreachable for both Lys1396 and Arg1623.

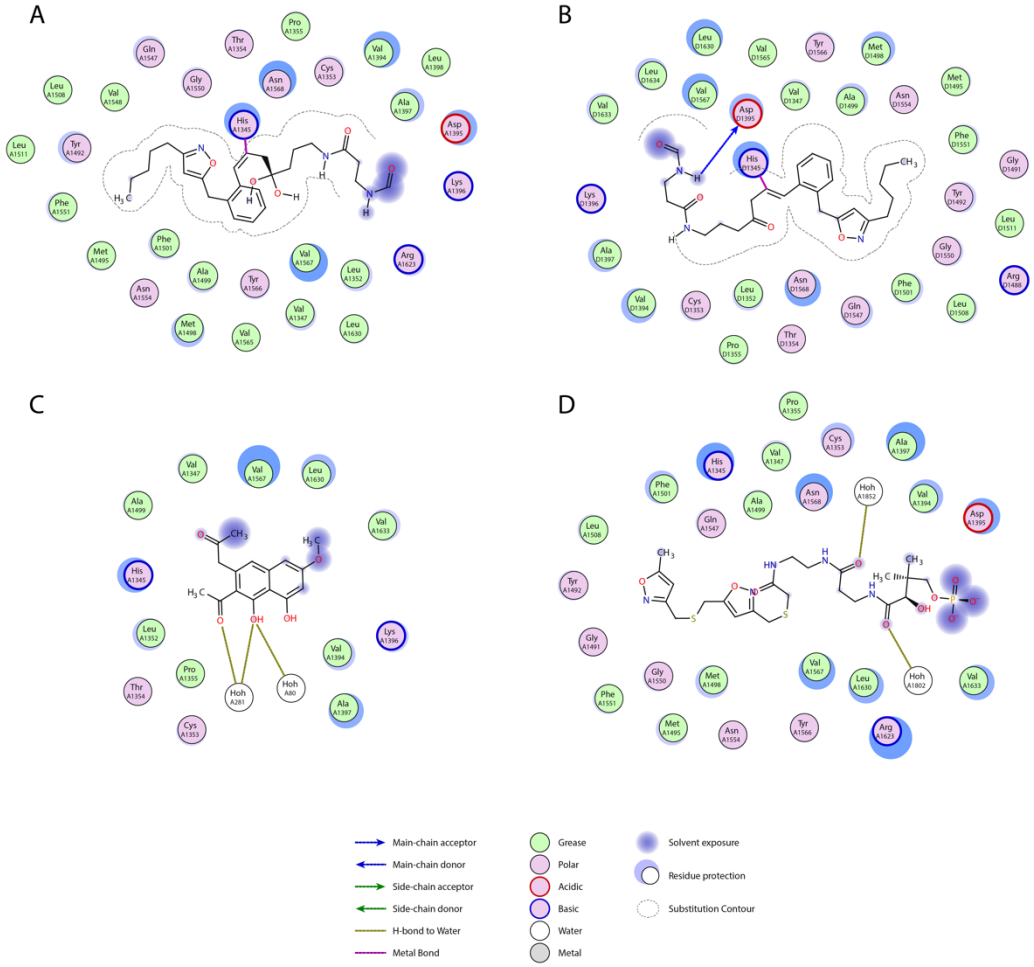


Figure 3.12: Visualization of ligand interactions in PksA PT substrate complexes. *(A The **16**-bound PT with the gemdiol state ligand. (B) The **16**-bound PT with the carbonyl state ligand. (C) 3HRR model with the final substrate analogue **2**. It is positioned in place of the phosphopantetheine arm in the other models. (D) 5KBZ model with its bisisoxazole mimetic. Overall positioning is similar to our model, with minor changes to accommodate the different substrate.*

In total we present the model of a distinct step in the cyclization of substrate in the norsolorinic acid anthrone biosynthesis by PksA. In combination with previously known structures, we conclude that no conformational changes are required for the enzymatic activity of PksA. Movement is however part of the cyclization, as the outward movement of the phosphopantetheine arm relative to the initial substrate bound model (PDB: 5KBZ) displays. This is substantiated by the fact that crosslinking of phospho-**16** was unsuccessful, whereas **16** crosslinked well (Fig. 3.7). Within the active site tunnel we could not observe any meaningful interactions between the ligand and nearby residues, both on the phosphopantetheine side and the substrate side (Fig. 4.12). The relatively consistent placement of the ligand might suggest a potential water network interacting with the ligand, analogous to that observed in the initial substrate analogue bound state (PDB: 5KBZ).

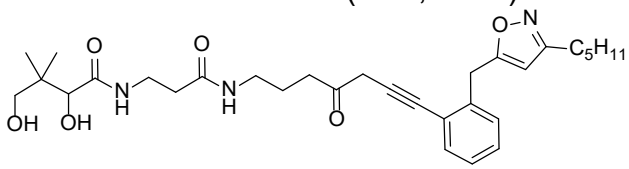
The weak interactions between ligand and protein allow for relative freedom of movement (Fig. 4.12), albeit potentially inhibited by a canonically bound ACP. The outward movement of the substrate after the first round of cyclization might initiate the undocking of the ACP, which is then completed upon the second cyclization reaction. Given the substrates' propensity to self-cyclize for the last step this might even act as a quality control.

From our results we can conclude the PT does not undergo conformational changes, meaning future efforts to adapt PTs will rely on specific substrate length and characteristics to achieve targeted cyclization. Further studies exploring the cyclization patterns of different PTs, will enable design of a substrate and an active site which are compatible and allow for reproducible targeted cyclization. Overall, this study will help guide rational programming of cyclizing enzymes in order to expand the chemical space accessible via bioengineered PKSs.

Materials & Methods

Synthesis of PT substrate inhibitor

For full synthetic details and characterizations, the reader is directed to the dissertation of Dr. Jacob M. Kravetz (JHU, 2021).



2,4-dihydroxy-3,3-dimethyl-N-(3-oxo-3-((4-oxo-7-(2-((3-pentylisoxazol-5-yl)methyl)phenyl)hept-6-yn-1-yl)amino)propyl)butanamide (16): The title compound was synthesized from 2-iodobenzyl bromide and 5-bromopentene over nine steps. **1H NMR** (300 MHz, CDCl3): δ 7.4-7.5 (m, 1H), 7.4-7.4 (m, 1H), 7.3-7.3 (m, 3H), 6.4-6.6 (m, 1H), 5.70 (s, 1H), 4.21 (s, 2H), 3.98 (s, 2H), 3.5-3.6 (m, 3H), 3.4-3.5 (m, 4H), 3.2-3.4 (m, 3H), 2.67 (t, 2H, J=7.1 Hz), 2.5-2.6 (t, J = 7.8 Hz, 2H), 2.41 (dd, 2H, J=4.9, 6.8 Hz), 1.7-1.9 (m, 5H), 1.6-1.6(m, 1H), 1.2-1.4 (m, 5H), 1.02 (s, 4H), 0.92 (s, 3H) 0.89 (t, J= 7.8 Hz, 3H), **$^{13}C\{^{1}H\}$ NMR** (101 MHz, CDCl 3): δ 204.22, 171.76, 171.50, 137.91, 132.60, 129.37, 128.82, 127.30, 122.91, 101.63, 86.61, 82.75, 77.71, 77.20, 70.90, 39.32, 38.86, 38.83, 35.71, 35.26, 34.89, 32.03, 31.35, 27.88, 26.00, 23.26, 22.27, 21.65, 20.41, 13.90 **UPLC-HRMS** (ESI) calc'd for $C_{31}H_{44}N_3O_6^+$ 554.3225; found 554.3227 $[M+H]^+$.

Protein expression and purification

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 r.p.m. until OD_{600} reached 0.7. Cultures were cold-shocked in ice water for 1 hr before induction with 0.5 mM isopropyl-β-d-thiogalactopyranoside (IPTG; GoldBio). Expression was carried out at 18 °C with shaking at 250 r.p.m. for approximately 16 h. Cells were harvested by centrifugation at 4000 × *g* for 15 min and flash frozen in liquid nitrogen for storage at −80 °C.

Generally, cell pellets were thawed in lysis buffer A (50 mM Tris-HCl, pH 8, 300 mM NaCl, 10% (v/v) glycerol) and lysed by sonication with a Vibra Cell High-Intensity Ultrasonic Liquid Processor equipped with a microtip probe at 40% amplitude for 10 cycles of 10 sec on and 10 sec off while on ice. The lysate was then cleared by

centrifugation (27,000 × *g* for 25 min). The protein was isolated via Cobalt affinity gravity-flow chromatography. After application, the resin was washed with increasing concentrations of imidazole in lysis buffer A and the protein of interest was eluted with 250 mM imidazole in lysis buffer A. Fractions containing the protein were pooled and dialyzed against the appropriate buffer (50 mM Tris pH 7.5, 5% (v/v) glycerol) at 4 °C. Proteins were optionally 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 were calculated using Expasy ProtParam.[7]

Phosphorylation of PT substrate mimic

2,4-Dihydroxy-3,3-dimethyl-N-(3-oxo-3-((4-oxo-7-(2-((3-pentylisoxazol-5-yl)methyl)phenyl)hept-6-yn-1-yl)amino)propyl)butanamide (**16**) was phosphorylated enzymatically under the following conditions: 0.3 mg/mL PanK (pantothenate kinase), 2 mM substrate mimic, 5 mM ATP, 20 mM KCl, 10 mM $MgCl_2$, 50 mM Tris pH 7.5 at room temperature for up to 7 h. Complete phosphorylation was confirmed by UPLC-ESI-MS (Waters Acquity/Xevo-G2, negative ion mode) equipped with an Acquity UPLC BEH C18 column (2.1mm x 50mm, 1.7µm particle size). All protein was removed from the reaction with a 3 kDa MWCO Amicon Ultra centrifuge filter (Millipore) and the resulting reaction product was used without further purification.

PT doping with substrate mimics

PksA-PT was combined with a 2-fold excess of substrate mimic in 50 mM Tris pH 7.5, 10% glycerol (v/v) at room temperature until fully converted as monitored by UPLC-ESI-MS (Waters Acquity Xevo-G2 UPLC-Q-Tof) equipped with an Acquity UPLC Protein BEH C4 column (2.1mm x 50mm, 300Å pore size, 1.7µm particle size). Doped PksA-PT was purified by size-exclusion chromatography using a Bio-Rad NGC chromatography system equipped with an Enrich SEC70 column (10 x 300 mm, Bio-Rad). Fractions containing PksA-PT were visualized by denaturing SDS-PAGE then pooled and concentrated in an Amicon Ultra centrifuge filter (10K MWCO, Millipore).

Trypsin digests and mass spectrometry of PksA-PT

PksA-PT samples were diluted into 8 M urea in 50 mM ammonium bicarbonate, pH 8 and incubated at 37 ºC for 35 min to allow unfolding. Iodoacetamide was then added to 15 mM final concentration and the samples were incubated in the dark for 35 min. The solution was then diluted with 50 mM ammonium bicarbonate buffer to 2 M final urea concentration and 1 µl of 1 mg/ml Trypsin Gold (Mass Spectrometry Grade, Promega, Madison, WI) was added to each reaction to digest at room temperature for 29 hr. Samples of 20 µl were diluted 2-fold with 20% acetonitrile in water. Samples were analyzed by UPLC-ESI-MS (Waters Acquity Xevo-G2 UPLC-Q-Tof, Acquity UPLC HSS T3 C18 column – 2.1mm x 100mm, 1.8µm particle size, 0.3 ml/min, 5-40% acetonitrile in water with 0.1% formic acid over 36 min followed by a 95% acetonitrile in water with 0.1% formic acid wash and re-equilibration to starting conditions). Fragments masses were then compared to calculated masses using BioPharmalynx (Waters).

Crystallization of crosslinked PksA-PT

Crosslinked protein was concentrated using 3k Da cutoff Amicon Ultra filters (lot No. R1SB25160) at 21'000 g/ 6 °C/ 3 min intervals to a final concentration of 10.7 mg/ml. Crystals were grown at 20 °C in a Rock Maker (Formulatrix) using a MRC 3 well plate (producer) sitting drop vapor diffusion with 50 µl of crystallization solution and of 0.2 µl protein and 0.2 µl crystallization solution per drop. Crystallization solution

consisted of 0.2 M sodium potassium phosphate and 22 % polyethyleneglycol (PEG) 3350 with addition of between 0.02 and 0.1 M sodium fluoride or 2-8 % v/v tert-butanol. All crystals were cryoproteced in mother liquor with 25% (v/v) ethylene glycol before flash freezing in liquid nitrogen. Data was collected at PXI (X06SA) of the swiss light source with a wavelength of 1.0 Å. Data reduction was performed using XDS and XSCALE, while phases were determined using molecular replacement in Phaser. Final models were built using model building in Coot and refinement in PHENIX. Ligands were parametrized in eLBOW.

PksA PT crystallized in space group C2 with unit cell dimensions of a = 52.1 Å, b = 98.6 Å and c = 326.7 Å and a beta angle of 90°. The molecular replacement model was PksA PT with a different ligand bound (PDB: 3HRR). The final model was refined to R_{work}/R_{Free} 29.6/31.6 at 2.64 Å resolution. Ramachandran values are 91.07 %/8.18 %/0.74 %(favored/allowed/outliers).

The crystal displayed a C222 pseudosymmetry. Notable worsening of refinement characteristics and density, especially around the active site led us to believe that the primary contributor are the two different states of the ligand, its carbonyl, and its gem-diol state. We therefore chose to refine the crystal under its C2 symmetry.

Varying ligand conformations and crosslink chemistry were explored including the use of occupancy refinements for alternate states. All occupancy refinements on individual protomers converged to a single solution each, resulting in the final model containing one state per protomer, a gem-diol for chains A and B and a carbonyl for chains C and D.

Table 1: X-ray crystallographic data collection and processing.

	PT
Wavelength	1
Resolution range	49.3 - 2.64 (2.734 - 2.64)
Space group	C 1 2 1
Unit cell	52.1 98.6 326.68 90 90.002 90
Total reflections	340878 (34878)
Unique reflections	48166 (4769)
Multiplicity	7.1 (7.3)
Completeness (%)	96.16 (80.14)
Mean I/sigma(I)	8.06 (0.32)
Wilson B-factor	82.75
R-merge	0.2324 (6.133)
R-meas	0.2511 (6.605)
R-pim	0.09405 (2.432)
CC1/2	0.997 (0.275)
CC*	0.999 (0.657)
Reflections used in refinement	46807 (3862)
Reflections used for R-free	2344 (197)
R-work	0.2960 (0.5706)
R-free	0.3160 (0.6197)
CC(work)	0.935 (0.518)
CC(free)	0.909 (0.403)
Number of non-hydrogen atoms	10602
macromolecules	10468
ligands	270
solvent	0
Protein residues	1352
Nucleic acid bases	
RMS(bonds)	0.004
RMS(angles)	0.66
Ramachandran favored (%)	91.07
Ramachandran allowed (%)	8.18
Ramachandran outliers (%)	0.74
Rotamer outliers (%)	4.89
Clashscore	7.43
Average B-factor	106.66
macromolecules	106.81
ligands	94.83

Discussion

In this thesis, we explored the higher order architecture of a trans-AT PKS bimodule core, at an intermediate resolution, which answers the long-standing question on how modPKS assembly lines are structurally arranged. Additionally, we investigated a PT stalled after the first step of cyclization, whose reaction mechanism was so far elusive. After a summary of the results, we will cover concurrent advances in the field of PKS structure and function connectivity and consider both open questions and next steps.

Summary of results

To tackle aim 1, I used a combination of X-ray crystallography and cryo-EM in chapter 2, deriving the composite model of K3DAK4, a trans-AT PKS bimodule core with architecture KS-DH-ACP-KS at an overall resolution of 7 Å, with the LINKS interactions for both KSs within the bimodule being resolved at 4 Å resolution. Local features for each individual domain were resolved at < 3 Å resolution. The model establishes a linear stacking of the modules roughly along the KSs dimerization axis. Additionally, the horizontal filamentation of K3DAK4 via each of its KSs LDs forming LINKS interactions to a neighbouring module was observed.

While K3DAK4 is a trans-AT PKS bimodule core, and a relative minimalistic one, the introduction of further modifying domains would not disrupt the overall architecture. The same holds true when applying the same higher order architecture to cis-AT PKSs, where the addition of the AT does not interfere with a linear stacking of modules within the assembly line. K3DAK4 also displays a few other key aspects: First and foremost, we see the horizontal stacking of modules via the LINKS interactions, a feature exclusive to trans-AT PKSs, as the LD responsible for the LINKS interaction [100] is a scaffold towards the AT in cis-AT PKSs [80]. A first potential glimpse at the LINKS interaction being an *in vivo* feature were given already, when *Straight et al* depict a region of higher density in *Bacillus subtilis* cells, which co-locates with labelled PKS domains from the bacillaene cluster[68]. The higher order architecture also emphasizes that the interaction strength of individual domains, *i.e.* to dimerize can be below the threshold to do so when isolated, meaning it is not always possible to investigate individual enzymatic domains on their own. This is exemplified by the monomer-dimer equilibrium of KS3 in solution. Multimodular PKSs, at least K3DAK4, therefore rely on a multitude of small interactions to keep the overall quaternary structure assembled. The interaction of multiple modPKS proteins across docking domains may follow the same rules, as prior investigations on docking domains [65,99,131] needed extra stabilizing scaffold when isolated, indicating interactions are either brief or rely on avidity.

In regards to aim 2, I employed X-ray crystallography on PksA PT crosslinked with a substrate analogue mimicking a stall between the two cyclization steps in chapter 3. The structure was solved at a resolution of 2.7 Å and allowed for a detailed look at the interaction of the substrate and its phosphopantetheine tether within the active site, revealing an outward movement of the substrate as compared to a substrate analogue pre-first cyclization step.

The PT structurally belongs to the DH class of PKS domains, yet it has unique additions to the fold and, more importantly, an active site which, despite being highly similar in nature to that of *i.e.* dehydratases, performs a very different task. The precise positioning of the ligand together with the unexpected outward movement of the phosphopantetheine arm during synthesis gives a clear indication of the directionality of the substrate's movement, conforming with prior knowledge on a product-analogue bound PksA PT [73]. Taken together, the three known models of PksA PT [44,73] with different stages of substrate bound indicate a cyclization process relying heavily on meticulous positioning of the substrate.

Progress in structural studies of the higher order architecture of type I PKS

At the outset of this thesis, the information on modular PKS structure available was based on the low resolution SAXS information on both a trans-AT PKS minimalist module [84], and the cis-AT DEBS module 3 [50], the intermediate resolution (7.3-9.2 Å) cryo-EM analysis of a cis-AT PKS from PikAIII [85], and two crystallographic models, the first describing MAS, but relying on homology to porcine FAS to determine relative positioning of the individually crystallized modifying (3.8 Å) and condensing (2.2 Å) region [94], and the second describing RhiE, a small trans-AT PKS module core at 2.1 Å resolution [41]. With the exception of the latter two all early modules were of too low resolution to afford detailed insights into higher order architecture. For multimodular architecture only low-resolution information derived from SAXS on the cis-AT DEBS3 bimodule [50], and also low resolution SAXS data on a trans-AT PKS module and its attachment to the succeeding module based on the inference of where the docking domains are placed in either model [95], were available.

However, the independent elucidation of a new HR-iPKS module in the form of LovB in 2020 by cryo-EM at 2.9 Å resolution [39] as well as two highly similar cis-AT PKS modules following the same KS-AT-KR-ACP architecture in 2021, DEBS module 1 at 3.2 Å resolution by cryo-EM [90] and Lsd14 at 2.4 Å resolution using a mix of both X-ray crystallography and cryo-EM [93], greatly advanced the field. Especially Lsd14 and DEBS module 1, which each were analyzed in three different states allowed for first insights into the mechanism of a PKS module in action. When including my work of K3DAK4, high resolution models for the modules of HR-iPKSs, cis- and trans-AT PKSs now exist.

While LovB is representing a complete HR-iPKS due to its iterative nature, the question of how DEBS module 1 integrates with DEBS module 2 to form the first of the three DEBS proteins remains unsolved. Extrapolating from K3DAK4 is the most likely way, as in the DEBS modules stack on top of one another linearly, an idea compounded by earlier SAXS work on the DEBS3 bimodule [50], which also concluded in a linear stacking of modules. However, it is unclear if they are also always kept in close proximity, a phenomenon that we could also not entirely explain for K3DAK4. In the case of Lsd14 this question is extended as its interaction with neighbouring modules includes docking domains as Lsd14 itself is a standalone module [93]. Overall, all 4 module architectures align when considering differences relating to their class of type I PKS and follow the extended architecture similar to both porcine FAS and MAS.

Advances in mechanistic understanding of type I PKSs modules

From a mechanistic point of view the new module structures are of great importance as well. While highly similar in higher order architecture to other modules, LovBs rigid structure with the tight interface between condensing- and modifying region, specifically the interface between KS and DH, is a unique feature. The ACP is therefore shuttling the substrate around inside a predefined reaction chamber, albeit an open one, where it can only reach one of the two sets of enzymatic domains. This is a clear departure from discussions for *i.e.* porcine FAS and DEBS where the rotational freedom of the modifying region relative to the condensing region was typically expected to allow for each ACP to reach most if not all active sites on either protomer. The model of LovB suggests a mechanism, not a structural homology, more similar to that of fungal FAS[132], where strong scaffolding keeps all domains in place and the ACP is the main moving part . The rigidity and two reaction chamber model open up the possibility that LovB might use both of its protomers independently throughout product synthesis. Given the requirement of an interaction with LovC which was exclusively observed on one protomer at a time, it will still likely not use both reaction chambers simultaneously.

Interestingly, all cis-AT modules structurally investigated to date, even at low resolution, follow the same KS-AT-KR-ACP architecture. Given these inherent similarities, and the high-resolution models of Lsd14 and DEBS module 1, we can finally extract mechanistic insights, especially in regards to the multiple states both Lsd14 and DEBS module 1 were found in. Starting with DEBS module 1, in 2016 a turnstile model for vectorial synthesis in modPKSs was described[133]. The new structure of DEBS module 1 elaborates on it and might explain how it works. The core idea is that the KS needs to ensure that the ACP cannot backload its substrate onto the same KS twice, in order to keep the synthesis directional in an assembly line fashion. To do this it is hypothesized to employ two states, one where it is accessible to the upstream ACP and one where it is closed in a turnstile manner. The three models of DEBS module 1 observed displayed one with the ACP bound to the KS, corresponding to the open state, one, where both ATs were found in a flexed state which sterically hinders the ACP binding site to the KS and simultaneously is proposed to render the AT inactive, corresponding to the closed state, and a third intermediate one, where only one of the two ATs is in the flexed position. Interestingly the KRs are interacting with the KS in both the open and the intermediate state to interact with the KS, but distinctively different. In the open state the KR binds to the opposing protomers KS whereas it binds to its own protomers KS in the intermediate state. This indicates again a high degree of rotational freedom in the modifying region and might suggest that this freedom is required for product synthesis as it may be a prerequisite for allosteric changes enabling the turnstile mechanism. Open questions regarding this hypothesis are mainly surrounding how the flexing of the AT is controlled.

The Lsd14 structures paint a similar picture to that of DEBS module 1, but offer a different hypothesis as to what drives vectorial synthesis in modPKSs. Its first state is an apo-Lsd14 crystal structure where the ACP is stalled at the AT. In the second state the ACP is not bound to any of the enzymatic domains, and in the third state the ACP is stalled at the KS. No flexing of the AT could be observed for Lsd14. Instead, the authors suggest a system based on a swivelling motion of the modifying relative to the condensing region, which inhibits double loading simply via proximity

probability. In the apo state the KR blocks access for the ACP to the KS, while simultaneously bringing the ACP in close proximity of the AT by virtue of being the anchor for the ACP linker. In the ACP-KS bound state one of the KR domains has detached from the KS but stays in close proximity, which again increases chances of an ACP-KS interaction. Lastly, in the third state both KR domains are fully detached from the KS, increasing probability of the ACP interacting with the KR itself and simultaneously increasing the probability of substrate handoff to the downstream KS. This proximity based approach is supported by the known substrate specificity of KS domains [33], which on its own might challenge backloading, and by extension the specificity of ACP-substrate-KS interactions.

While both models are intriguing and plausible the kinetic likelihood model described for Lsd14 is the more generalist of the two, with no restrictions to either larger modifying regions or other classes of type I PKSs, as both iterative and trans-AT type I PKSs may follow this rule. It is however important to note that the static nature of LovB is directly contradicting this model and suggests that at least a certain class of HR-iPKSs may not follow it. Conversely, the turnstile model proposed is excluding both iPKSs and trans-AT PKSs, the prior as they actively need to backload their substrate and the latter for the lack of an AT on the module itself. While the Lsd14 and DEBS systems are different it appears unlikely that these highly similar, and to a degree evolutionary related [60,134], systems would realise vastly different mechanisms to tackle the same problem. Further research will be required to definitively describe the mechanism of substrate translocation through PKS assembly lines.

Implications of type I PKSs higher order architecture for trans-acting enzymes

The recent advances in our understanding of modules and potential to visualize them finally enables us to do the same for the interaction of modules with their trans-acting enzymes [66]. As a trans-AT PKS, K3DAK4 heavily relies on its AT-AT-ER multienzyme. It will be of special interest to investigate how this trans-acting protein interfaces with the trans-AT PKS assembly line mesh, as well as determining which of the two ATs acts on which step. To the best of my knowledge the only structural model of a trans-acting enzyme and its docking to the main PKS module currently available is the LovBC complex [39]. Given its iterative nature and the presence of an AT the way LovC binds to said AT cannot be a blueprint for trans-AT PKSs, it is however crucial in understanding that despite the existence of a given enzymatic domain within a module this function may still be outsourced. When engineering PKSs the activity as well as the docking of trans-acting enzymes will prove to be of utmost importance as they are a demonstrated key part in certain PKSs [135] beyond the previously expected trans-AT PKS systems. Additionally, they may offer to diversify a given PKSs product without the need to insert a domain into the PKSs module itself.

Product programming outside of core PKS enzymes

Individual highly specialized domains are a key feature of certain PKS subclasses with the prevalent example in this thesis being the PT. Despite not being completely understood mechanistically, on a biological level several PTs are known to promote different forms of cyclization patterns [136,137]. A precise understanding of such cyclization patterns, as well as a detailed differentiation between PTs with varying

cyclization patterns will enable the targeted choice or even modification of individual PTs to suit a given purpose. With the movement of the substrate through the different stages of enzymatic activity observed in our study, considerations of ACP docking and affinity will be essential.

Another highly specialized domain not found in all PKSs is the MT. While its enzymatic activity is very different to that of a PT, it also is outside of the core PKS enzymes and much less well understood. Interestingly, it is one domain whose function can be completely outsourced outside the actual PKS and reintroduced by ATs, either via the direct loading of methylated acyl-moieties, such as DEBS loading methyl-malonyl-CoA [43], or the use of a different specialized enzyme [138]. Additionally, the MT, if present on the PKS module, comes in two different classes inherently, the C-Methyltransferases (C-MT) [139] and the O-methyltransferases (O-MT), with a distinct fold each and domain positioning in PKSs they are part of[140]. As our understanding of the functions of core PKS enzymes and even modules grow, it is time to focus on these less ubiquitous enzymes to gain access to the full potential of the highly versatile family of secondary metabolite producers that are PKSs.

Sub cellular localization of type I PKSs and export of their products

A currently understudied aspect of PKS biosynthesis is the cellular localization of PKSs as well as the export of their products. As many of their products are essentially weapons designed to afford an advantage in their biological niche by killing competition [10,141-144] or sometimes symbiosis [145], the secretion of their secondary metabolites becomes important. It is expected that some of them are secreted via general transporters, however with the attachment of the bacillaene PKS cluster to the cell wall, as well as the existence of an annotated ABC transporter in the K3DAK4 BGC [108], some PKSs might have specific exporters. Taking the bacillaene PKSs reported positioning [68], the PKS itself could interact with an exporter in order to be located at said position, a hypothesis compounded by the finding of PKS proteins at the cell membrane in *Myxococcus xanthus* [146]. This, together with investigation of tailoring enzymes [147] will help to further understanding of PKSs as parts of the secondary metabolite synthesis and distribution.

Current problems in structural studies of PKSs

Given the recent results and context discussed above, I think there are two key areas to focus on in the near future for type I PKSs. The first concerns specialized domains, such as the PT, and how to both elucidate their exact mechanism as well as thereby enabling targeted manipulation and the option to incorporate them into PKSs they are not native to. The second relates to the modular and multimodular mechanisms of especially substrate shuffling and control thereof in both iterative and modular PKSs. Going together with the substrate movement throughout product synthesis a close look at trans-acting enzymes will be essential, as well as the interface of PKS and NRPS.

Mechanistic insights into the cyclization patterns of PTs

Regarding the PT I would aim on one hand to specifically target the variation in cyclization patterns. Given the biochemical data, it is clear that at least two classes of cyclization patterns exist between fungal and bacterial systems, each with two to three variations within. Elucidation of PT structures from a set of systems with

differing cyclization patterns [136], potentially in multiple substrates bound states analogous to what exists now for PksA PT, should allow for the investigation of minute differences to establish a clear logic for how cyclization specificity is achieved, therefore allowing for the targeted change thereof, as already elucidated for a single PT [137]. The similarity in overall fold will allow to focus on the crucial details. Additionally, it is essential to include the ACP in follow up experiments. With the delta in phosphopantetheine arm positioning displayed in our **16**-bound PT to the bisisoxale bound state, the exact position of the ACP throughout the cyclization process might be changing, therefore also making its binding site not just static but a point to consider in terms of affinity of different interfacing patches.

How to investigate type I PKS cellular localization and interaction partners

In terms of following up on the achievements in structurally elucidating modules and a multimodule, I'd argue for a split between classes of type I PKSs. While the K3DAK4 model can act as a blueprint for cis-AT PKS assembly lines higher order architecture when regarding only its vertical stacking of modules, conformation is still required. Additionally, it might be of interest to investigate whether cis-AT PKSs have a way homologous to the LINKS interaction in order to keep multiples of the same assembly line in close proximity. In regards to trans-AT PKSs a key factor will be the localisation in their natural hosts. While an initial glimpse has been shown for the bacillaene cluster, with the advancement in technology of gene modification, protein labelling and light microscopy as well as potentially cryo-electron tomography (cryo-ET), a much clearer and more detailed picture of their cellular localisation should be attainable, a main concern being the non-constitutive expression of almost all PKSs [7,8,148,149]. Advances in genome mining give access to tens of thousands of thus far mostly unexplored BGCs [124], one of which is sure to be a suitable target for such a study. Given the sheer size of a trans-AT PKS mesh, it should be relatively easy to find within the cell, it might even be possible to also capture positioning of both trans-acting enzymes and tailoring enzymes. Especially cryo-ET will prove crucial here in allowing to observe native interactions for the thus far elusive trans-acting enzymes. Additionally, if the association with cell walls is a more general feature of trans-AT PKS meshes, it may be possible to investigate the anchors within the membrane, potentially also serving an exporting function, to obtain a complete mechanistic picture from acyl-building block to secreted product of a PKS BGC.

While *in vivo* studies would be highly advantageous, a focus on trans-acting enzymes and their interaction with their PKSs modules *in vitro* is attainable now. With K3DKA4 we have a model to which we could directly add its trans-acting enzyme AT-AT-ER *in vitro* and try to observe it in either cryo-EM or cryo-ET. Recent progress in single particle cryo-EM, as highlighted by the LovBC complex, makes it reasonable to sort for modules interfacing with the trans-acting enzymes even if only 10-20 % of all modules or multimodules were to interact with them. Given current limitations to cryo-ET this approach is also more likely to yield higher resolution answers quicker.

Step-wise analysis of the mechanism of a type I PKS module

Concerning mechanistic insights into the modules of type I PKSs, I would focus on cis-AT PKSs for now as they typically contain all required domains on their modules,

reducing the complexity as compared to the inclusion of trans-acting enzymes for trans-AT PKS, or the multiple interactions with each enzyme in iPKSs. Similarly, to the approach for the PT, stalling the ACP at each domain via crosslinking each individual step of a biochemically well-defined system such as DEBS and analysing them via cryo-EM or X-ray crystallography will give snapshots of all the steps, which can then be combined to arrive at an overall synthesis movie. This will also provide insights into potential conformational rearrangements of individual domains or repositioning of domains when active and inactive, as already initially described for both DEBS module 1 and Lsd14. Unfortunately, direct tracing of the ACP throughout product synthesis is not currently possible, making this focus on the key moments the next best thing.

The interplay of NRPSs and PKSs in hybrid biosynthetic gene clusters

Lastly, the interaction of NRPS and PKS modules is a feature found across all classes of type I PKSs. Knowledge on NRPS domains and modules is at a similar stage to that of PKSs, with a di-module having been structurally analysed [104]. It is known that docking domains bridging pure NRPS and PKS modules or multimodules exist [131,150]. What is however unclear is how the monomeric NRPS would act when attached to the always homodimeric type I PKS module, especially considering the requirement for big conformational changes throughout product synthesis of NRPS [107]. Rather than starting with modules requiring docking domains, I would advocate to use one of the hybrid multimodules, containing both a NRPS and a PKS module to observe this interaction. A key point of concern here is to investigate whether the individual PKS module is dimerizing, especially when excising it from say a NRPS-PKS-PKS module architecture. With recent results we know that it is possible an individual PKS module does not dimerize with a high affinity and would therefore need the stability of dimerization granted by multiple sequentially dimerizing domains. Alternatively, it is possible to graft an NRPS and a PKS module together onto a single protein that would typically only interact via their docking domains, although folding might become an issue. In either case once a stable system is found investigation via cryo-EM has the potential to establish a structural view on their interactions, opening the possibility of follow up studies on substrate transfer across these two highly different yet compatible systems.

Approaches for engineering PKS biosynthetic modules and pathways

With their versatility, their step-wise functionality and how well understood the core enzymatic domains of PKSs are it comes as no surprise that the intention to engineer PKSs to produce desired chemicals is old. Type I PKSs, especially those following the rule of collinearity have been a prime target, due to their predictability [151]. Without, or with limited structural knowledge, the exchange of modules or domains in order to change the product of a given PKS have been tried [152], with the method being so popular it was termed combinatorial biosynthesis. A key issue often arising is substrate specificity of domains in chimeric PKSs, as well as unpredictable behaviour of modules when excised from their native systems, as exemplified by the iterative nature of DEBS module 3 when acting as a standalone module [50,153]. In recent years progress was made mainly by chimerizing well known PKSs, such as DEBS, by exchanging individual domains[36,154,155] or even whole modules[36,92,156].

Efforts to change the enzymatic activity of individual domains were also made, and at times, successful[31,137,157-159]. The last common method to acquire targeted PKS engineering was to incorporate specialized building blocks via the AT[36,155]. In order to give bioengineers the required tools to perform more precise and elegant adjustments to a given PKS system information on the interplay of multiple modules and even across multimodular proteins will be essential. Detailed information on ACP interactions as well as the gatekeeping function of individual domains, all of which are substrate specific to varying degrees[152], are crucial to build a functioning system. When looking at the recent progress described for the Lsd14 and DEBS module 1 systems, it appears especially important to determine the exact mechanism of vectorial synthesis in modPKSs to enable a consistent and targeted approach to engineering of PKSs. The delicate balance of affinities which also rely on the substrate itself for each step will be a key problem until we have an exact understanding of the modular mechanics of modPKS. While vectorial synthesis pertains to modPKSs only, the same inherent problem holds true for iPKSs, maybe even more so as in each step all enzymatic domains are theoretically available and guidance of the substrate is expected to be largely specificity and ACP docking based.

With my work I contributed to the understanding of multimodular systems, adding a new dimension to the variables that need to be considered, especially regarding avidity of the complex quaternary structure of PKS multimodules in at least trans-AT PKS systems. It is crucial for function that, while necessarily occurring, not all interactions are very stable, *i.e.* an ACP that binds a given domain with too high an affinity is inherently a stop for any PKS. Additionally, the existence of trans-acting enzymes adds an additional barrier in that the insertion, excision or replacement of a domain may inhibit or reduce the affinity of the interaction with an essential trans-acting enzyme. As we do not know the trans-enzymes interaction sites excisions are especially problematic.

In an effort to make more specific chemical features available to bioengineers, the inclusion of specialized domains such as the MT or PT will be crucial. High efficiency throughput of an organic compound with a defined cyclization pattern is of interest beyond the medical use of natural secondary metabolites and their derivatives we are focusing on today and might, together with other systems, lend itself to becoming a standard in the chemical industry as well.

Lastly an in-depth look at the natural interaction of PKS and NRPS will both further enhance our understanding of either system, and, when properly studied, vastly expand the possibilities of bioengineered PKS-NRPS hybrids to produce highly complex organic compounds. Initial suggestions regarding NRPS-PKS combinatorial engineering revolves around the adaptation of docking domains for the targeted transfer of substrates between systems[150], however substrate specificity is an expected problem.

A future with selective structure-based PKS design

In the long term I see two big advantages to the investigation of PKSs. On the one hand we are currently on the verge of facing an antibiotics crisis, with strains such as methicillin-resistant *Staphylococcus aureus* being on the rise especially in hospital settings. The overuse of antibiotics in meat production being another key potential

source of more highly resistant bacterial strains. A relevant portion of clinically available antibiotics are of PKS origin or derivatives thereof [15]. With an advancing understanding of how PKSs operate we will be able to use the already available, and surely expanding, vast number of genome mined BGCs[124] and be more successful in both predicting their products[160] as well as being able to artificially use them in order to produce said compounds, potentially even alter and improve upon them within the BGC itself. I am certain, that among them we will find many more compounds with high medical value, not only as antibiotics but for many other types of diseases as well. The second key advance regards the use of PKSs as vehicles to produce basically any polyketide based organic compound. With our increasing understanding of the system, engineering approaches are already becoming more and more successful. In the future I could envision these to become very customizable, potentially finding introduction into the chemical industry. The large versatility existing in the PKS family gives a good indication of the chemical space they could cover, which expands even further when factoring in NRPS. Both from an ecological and a material point of view this approach has much potential, as some chemical reactions that today rely on rare minerals as catalysts, consume huge amounts of energy and produce potentially highly dangerous side products could be exchanged with fermenters of a given bacterium or fungus containing an engineered PKS which just secret the compound of choice. As neither antibiotic resistance nor climate change are issues that will just resolve themselves, I firmly believe that the advancement of our knowledge on BGCs in general and PKSs in particular will be a crucial contribution to sustaining ourselves and to a certain degree our standard of living in the future.

Acknowledgements

First and foremost, I would like to thank Timm for giving me the opportunity to work on this project, your support and mentorship, as well as discussions. I want to thank Henning for being my secondary supervisor and contributing valuable input through the PAC meetings. I also want to thank Sebastian for stepping in and taking over as secondary supervisor for the final steps. Thank you, Marcus, for agreeing to be my external expert and being both cheerful and helpful during the PAC meetings.

The core facilities have always been a source of great help, I would like to especially thank Tim from the biophysics facility, Mohamed, Kenny and Carola from the EM facility, Katarzyna from the Mass spectrometry facility, Adam from Research IT as well as Guillermo, Pablo and Martin at Scicore.

Then I would like to thank everyone who has been in the lab over the years, so thank you Moritz, Anna, Freddy, Domi, Eddi, Yana, Leo, Fra, Stefan, Janine, Flo, Flu, Kai, Shubi, Hugo, Malik, Niko, Roman, LMC, Karo, Elsa, Mirko, Matze and Matze II, for amazing apèros, great discussions and the good atmosphere.

A big thank you also to the folks from the floor; Elia, Morris, Majed and David always available for either discussion or a break if needed.

To my friends who have kept me grounded and gave me the opportunity to catch a break every now and then: Thank you!

And lastly, thank you to my family for always supporting me and encouraging scientific thinking from a young age.

References

1 Shen, G. Campbell biology (edited by Lisa Urry, Michael Cain, Steven Wasserman, Peter Minorsky and Jane Reece). *J Biol Res (Thessalon)* **27**, 19 (2020). https://doi.org:10.1186/s40709-020-00127-0

2 O'Sullivan, B. P. & Freedman, S. D. Cystic fibrosis. *Lancet* **373**, 1891-1904 (2009). https://doi.org:10.1016/S0140-6736(09)60327-5

3 Robinson, P. K. Enzymes: principles and biotechnological applications. *Essays Biochem* **59**, 1-41 (2015). https://doi.org:10.1042/bse0590001

4 Moldave, K. Nucleic Acids and Protein Biosynthesis. *Annu Rev Biochem* **34**, 419-448 (1965). https://doi.org:10.1146/annurev.bi.34.070165.002223

5 Wakil, S. J., Stoops, J. K. & Joshi, V. C. Fatty acid synthesis and its regulation. *Annu Rev Biochem* **52**, 537-579 (1983). https://doi.org:10.1146/annurev.bi.52.070183.002541

6 Maplestone, R. A., Stone, M. J. & Williams, D. H. The evolutionary role of secondary metabolites--a review. *Gene* **115**, 151-157 (1992). https://doi.org:10.1016/0378-1119(92)90553-2

7 Shwab, E. K. & Keller, N. P. Regulation of secondary metabolite production in filamentous ascomycetes. *Mycol Res* **112**, 225-230 (2008). https://doi.org:10.1016/j.mycres.2007.08.021

8 Macheleidt, J. *et al.* Regulation and Role of Fungal Secondary Metabolites. *Annu Rev Genet* **50**, 371-392 (2016). https://doi.org:10.1146/annurev-genet-120215-035203

9 Patra, B., Schluttenhofer, C., Wu, Y., Pattanaik, S. & Yuan, L. Transcriptional regulation of secondary metabolite biosynthesis in plants. *Biochim Biophys Acta* **1829**, 1236-1247 (2013). https://doi.org:10.1016/j.bbagrm.2013.09.006

10 Hashimi, S. M., Wall, M. K., Smith, A. B., Maxwell, A. & Birch, R. G. The phytotoxin albicidin is a novel inhibitor of DNA gyrase. *Antimicrob Agents Chemother* **51**, 181-187 (2007). https://doi.org:10.1128/AAC.00918-06

11 Hara, M., Saitoh, Y. & Nakano, H. DNA strand scission by the novel antitumor antibiotic leinamycin. *Biochemistry* **29**, 5676-5681 (1990). https://doi.org:10.1021/bi00476a005

12 Staunton, J. & Wilkinson, B. Biosynthesis of Erythromycin and Rapamycin. *Chem Rev* **97**, 2611-2630 (1997). https://doi.org:10.1021/cr9600316

13 Craney, A., Ahmed, S. & Nodwell, J. Towards a new science of secondary metabolism. *J Antibiot (Tokyo)* **66**, 387-400 (2013). https://doi.org:10.1038/ja.2013.25

14 Tsukamoto, S., Matsunaga, S., Fusetani, N. & Toh-E, A. Theopederins F-J: Five new antifungal and cytotoxic metabolites from the marine sponge, Theonella swinhoei. *Tetrahedron* **55**, 13697-13702 (1999). https://doi.org:Doi 10.1016/S0040-4020(99)00841-8

15 Newman, D. J. & Cragg, G. M. Natural Products as Sources of New Drugs over the Nearly Four Decades from 01/1981 to 09/2019. *J Nat Prod* **83**, 770-803 (2020). https://doi.org:10.1021/acs.jnatprod.9b01285

16 Hertweck, C. The biosynthetic logic of polyketide diversity. *Angew Chem Int Ed Engl* **48**, 4688-4716 (2009). https://doi.org:10.1002/anie.200806121

17 Sussmuth, R. D. & Mainz, A. Nonribosomal Peptide Synthesis-Principles and Prospects. *Angew Chem Int Ed Engl* **56**, 3770-3821 (2017). https://doi.org:10.1002/anie.201609079

18 Hwang, S., Lee, N., Cho, S., Palsson, B. & Cho, B. K. Repurposing Modular Polyketide Synthases and Non-ribosomal Peptide Synthetases for Novel Chemical Biosynthesis. *Front Mol Biosci* **7**, 87 (2020). https://doi.org:10.3389/fmolb.2020.00087
19 Weissman, K. J. & Muller, R. Protein-protein interactions in multienzyme megasynthetases. *Chembiochem* **9**, 826-848 (2008). https://doi.org:10.1002/cbic.200700751
20 Cox, R. J. & Simpson, T. J. Fungal type I polyketide synthases. *Methods Enzymol* **459**, 49-78 (2009). https://doi.org:10.1016/S0076-6879(09)04603-5
21 Shimizu, Y., Ogata, H. & Goto, S. Type III Polyketide Synthases: Functional Classification and Phylogenomics. *Chembiochem* **18**, 50-65 (2017). https://doi.org:10.1002/cbic.201600522
22 Hertweck, C., Luzhetskyy, A., Rebets, Y. & Bechthold, A. Type II polyketide synthases: gaining a deeper insight into enzymatic teamwork. *Nat Prod Rep* **24**, 162-190 (2007). https://doi.org:10.1039/b507395m
23 Smith, S. & Tsai, S. C. The type I fatty acid and polyketide synthases: a tale of two megasynthases. *Nat Prod Rep* **24**, 1041-1072 (2007). https://doi.org:10.1039/b603600g
24 Herbst, D. A., Townsend, C. A. & Maier, T. The architectures of iterative type I PKS and FAS. *Nat Prod Rep* **35**, 1046-1069 (2018). https://doi.org:10.1039/c8np00039e
25 Morita, H., Wong, C. P. & Abe, I. How structural subtleties lead to molecular diversity for the type III polyketide synthases. *J Biol Chem* **294**, 15121-15136 (2019). https://doi.org:10.1074/jbc.REV119.006129
26 Gay, D. C. *et al.* A close look at a ketosynthase from a trans-acyltransferase modular polyketide synthase. *Structure* **22**, 444-451 (2014). https://doi.org:10.1016/j.str.2013.12.016
27 Tsai, S. C. & Ames, B. D. Structural enzymology of polyketide synthases. *Methods Enzymol* **459**, 17-47 (2009). https://doi.org:10.1016/S0076-6879(09)04602-3
28 Keatinge-Clay, A. Crystal structure of the erythromycin polyketide synthase dehydratase. *J Mol Biol* **384**, 941-953 (2008). https://doi.org:10.1016/j.jmb.2008.09.084
29 Reid, R. *et al.* A model of structure and catalysis for ketoreductase domains in modular polyketide synthases. *Biochemistry* **42**, 72-79 (2003). https://doi.org:10.1021/bi0268706
30 Dunn, B. J., Cane, D. E. & Khosla, C. Mechanism and specificity of an acyltransferase domain from a modular polyketide synthase. *Biochemistry* **52**, 1839-1841 (2013). https://doi.org:10.1021/bi400185v
31 Zhang, L. *et al.* Stereospecificity of Enoylreductase Domains from Modular Polyketide Synthases. *ACS Chem Biol* **13**, 871-875 (2018). https://doi.org:10.1021/acschembio.7b00982
32 Chen, A., Jiang, Z. & Burkart, M. D. Enzymology of standalone elongating ketosynthases. *Chem Sci* **13**, 4225-4238 (2022). https://doi.org:10.1039/d1sc07256k
33 Klaus, M., Buyachuihan, L. & Grininger, M. Ketosynthase Domain Constrains the Design of Polyketide Synthases. *ACS Chem Biol* **15**, 2422-2432 (2020). https://doi.org:10.1021/acschembio.0c00405

34 Cantu, D. C., Forrester, M. J., Charov, K. & Reilly, P. J. Acyl carrier protein structural classification and normal mode analysis. *Protein Sci* **21**, 655-666 (2012). https://doi.org:10.1002/pro.2050
35 Moretto, L., Heylen, R., Holroyd, N., Vance, S. & Broadhurst, R. W. Modular type I polyketide synthase acyl carrier protein domains share a common N-terminally extended fold. *Sci Rep* **9**, 2325 (2019). https://doi.org:10.1038/s41598-019-38747-9
36 Yuzawa, S. *et al.* Comprehensive in Vitro Analysis of Acyltransferase Domain Exchanges in Modular Polyketide Synthases and Its Application for Short-Chain Ketone Production. *ACS Synth Biol* **6**, 139-147 (2017). https://doi.org:10.1021/acssynbio.6b00176
37 Lohr, N. A. *et al.* The Ketosynthase Domain Controls Chain Length in Mushroom Oligocyclic Polyketide Synthases. *Chembiochem*, e202200649 (2022). https://doi.org:10.1002/cbic.202200649
38 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 (2017). https://doi.org:10.1039/c6sc03496a
39 Wang, J. *et al.* Structural basis for the biosynthesis of lovastatin. *Nat Commun* **12**, 867 (2021). https://doi.org:10.1038/s41467-021-21174-8
40 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 (2014). https://doi.org:10.1021/cb500459b
41 Bretschneider, T. *et al.* Vinylogous chain branching catalysed by a dedicated polyketide synthase module. *Nature* **502**, 124-128 (2013). https://doi.org:10.1038/nature12588
42 Demachi, A. *et al.* An Unusual Extender Unit Is Incorporated into the Modular Polyketide Synthase of Scopranones Biosynthesis. *Biochemistry* **58**, 5066-5073 (2019). https://doi.org:10.1021/acs.biochem.9b00908
43 Chan, Y. A., Podevels, A. M., Kevany, B. M. & Thomas, M. G. Biosynthesis of polyketide synthase extender units. *Nat Prod Rep* **26**, 90-114 (2009). https://doi.org:10.1039/b801658p
44 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* **114**, E4142-E4148 (2017). https://doi.org:10.1073/pnas.1609001114
45 Hothersall, J. *et al.* Mutational analysis reveals that all tailoring region genes are required for production of polyketide antibiotic mupirocin by pseudomonas fluorescens - Pseudomonic acid B biosynthesis precedes pseudomonic acid A. *Journal of Biological Chemistry* **282**, 15451-15461 (2007). https://doi.org:10.1074/jbc.M701490200
46 Aron, Z. D., Dorrestein, P. C., Blackhall, J. R., Kelleher, N. L. & Walsh, C. T. Characterization of a new tailoring domain in polyketide biogenesis: the amine transferase domain of MycA in the mycosubtilin gene cluster. *J Am Chem Soc* **127**, 14986-14987 (2005). https://doi.org:10.1021/ja055247g
47 Paitan, Y., Orr, E., Ron, E. Z. & Rosenberg, E. Cloning and characterization of a Myxococcus xanthus cytochrome P-450 hydroxylase required for biosynthesis of the polyketide antibiotic TA. *Gene* **228**, 147-153 (1999). https://doi.org:10.1016/s0378-1119(98)00609-x

48 Palaniappan, N., Alhamadsheh, M. M. & Reynolds, K. A. cis-Delta(2,3)-double bond of phoslactomycins is generated by a post-PKS tailoring enzyme. *J Am Chem Soc* **130**, 12236-12237 (2008). https://doi.org:10.1021/ja8044162
49 Weissman, K. J. The structural biology of biosynthetic megaenzymes. *Nat Chem Biol* **11**, 660-670 (2015). https://doi.org:10.1038/nchembio.1883
50 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 (2014). https://doi.org:10.1016/j.jmb.2014.03.015
51 Brocchieri, L. & Karlin, S. Protein length in eukaryotic and prokaryotic proteomes. *Nucleic Acids Res* **33**, 3390-3400 (2005). https://doi.org:10.1093/nar/gki615
52 Keatinge-Clay, A. T. The Structural Relationship between Iterative and Modular PKSs. *Cell Chem Biol* **23**, 540-542 (2016). https://doi.org:10.1016/j.chembiol.2016.05.005
53 Dodge, G. J., Maloney, F. P. & Smith, J. L. Protein-protein interactions in "cis-AT" polyketide synthases. *Nat Prod Rep* **35**, 1082-1096 (2018). https://doi.org:10.1039/c8np00058a
54 Staunton, J. & Weissman, K. J. Polyketide biosynthesis: a millennium review. *Natural Product Reports* **18**, 380-416 (2001). https://doi.org:10.1039/a909079g
55 Khosla, C., Herschlag, D., Cane, D. E. & Walsh, C. T. Assembly line polyketide synthases: mechanistic insights and unsolved problems. *Biochemistry* **53**, 2875-2883 (2014). https://doi.org:10.1021/bi500290t
56 Boakes, S. *et al.* A new modular polyketide synthase in the erythromycin producer Saccharopolyspora erythraea. *J Mol Microbiol Biotechnol* **8**, 73-80 (2004). https://doi.org:10.1159/000084562
57 Piel, J. Biosynthesis of polyketides by trans-AT polyketide synthases. *Nat Prod Rep* **27**, 996-1047 (2010). https://doi.org:10.1039/b816430b
58 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 (2008). https://doi.org:10.1038/nbt1379
59 Katz, L. The DEBS paradigm for type I modular polyketide synthases and beyond. *Methods Enzymol* **459**, 113-142 (2009). https://doi.org:10.1016/S0076-6879(09)04606-0
60 Nivina, A., Yuet, K. P., Hsu, J. & Khosla, C. Evolution and Diversity of Assembly-Line Polyketide Synthases. *Chem Rev* **119**, 12524-12547 (2019). https://doi.org:10.1021/acs.chemrev.9b00525
61 Caffrey, P., Bevitt, D. J., Staunton, J. & Leadlay, P. F. Identification of DEBS 1, DEBS 2 and DEBS 3, the multienzyme polypeptides of the erythromycin-producing polyketide synthase from Saccharopolyspora erythraea. *FEBS Lett* **304**, 225-228 (1992). https://doi.org:10.1016/0014-5793(92)80624-p
62 Khosla, C., Tang, Y., Chen, A. Y., Schnarr, N. A. & Cane, D. E. Structure and mechanism of the 6-deoxyerythronolide B synthase. *Annu Rev Biochem* **76**, 195-221 (2007). https://doi.org:10.1146/annurev.biochem.76.053105.093515
63 Schwecke, T. *et al.* The biosynthetic gene cluster for the polyketide immunosuppressant rapamycin. *Proc Natl Acad Sci U S A* **92**, 7839-7843 (1995). https://doi.org:10.1073/pnas.92.17.7839
64 Staunton, J. *et al.* Evidence for a double-helical structure for modular polyketide synthases. *Nat Struct Biol* **3**, 188-192 (1996). https://doi.org:10.1038/nsb0296-188

65 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). https://doi.org:10.1016/s1074-5521(03)00156-x
66 Helfrich, E. J. & Piel, J. Biosynthesis of polyketides by trans-AT polyketide synthases. *Nat Prod Rep* **33**, 231-316 (2016). https://doi.org:10.1039/c5np00125k
67 Jenner, M. *et al.* Mechanism of intersubunit ketosynthase-dehydratase interaction in polyketide synthases. *Nat Chem Biol* **14**, 270-275 (2018). https://doi.org:10.1038/nchembio.2549
68 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 (2007). https://doi.org:10.1073/pnas.0609073103
69 Partida-Martinez, L. P. & Hertweck, C. A gene cluster encoding rhizoxin biosynthesis in "Burkholderia rhizoxina", the bacterial endosymbiont of the fungus Rhizopus microsporus. *Chembiochem* **8**, 41-45 (2007). https://doi.org:10.1002/cbic.200600393
70 Liu, L., Tang, M. C. & Tang, Y. Fungal Highly Reducing Polyketide Synthases Biosynthesize Salicylaldehydes That Are Precursors to Epoxycyclohexenol Natural Products. *J Am Chem Soc* **141**, 19538-19541 (2019). https://doi.org:10.1021/jacs.9b09669
71 Sims, J. W., Fillmore, J. P., Warner, D. D. & Schmidt, E. W. Equisetin biosynthesis in Fusarium heterosporum. *Chem Commun (Camb)*, 186-188 (2005). https://doi.org:10.1039/b413523g
72 Herbst, D. A. *et al.* The structural organization of substrate loading in iterative polyketide synthases. *Nat Chem Biol* **14**, 474-479 (2018). https://doi.org:10.1038/s41589-018-0026-3
73 Crawford, J. M. *et al.* Structural basis for biosynthetic programming of fungal aromatic polyketide cyclization. *Nature* **461**, 1139-1143 (2009). https://doi.org:10.1038/nature08475
74 de Jonge, R. *et al.* Gene cluster conservation provides insight into cercosporin biosynthesis and extends production to the genus Colletotrichum. *Proc Natl Acad Sci U S A* **115**, E5459-E5466 (2018). https://doi.org:10.1073/pnas.1712798115
75 Soehano, I. *et al.* Insights into the programmed ketoreduction of partially reducing polyketide synthases: stereo- and substrate-specificity of the ketoreductase domain. *Organic & Biomolecular Chemistry* **12**, 8542-8549 (2014). https://doi.org:10.1039/c4ob01777c
76 Dimroth, P., Walter, H. & Lynen, F. [Biosynthesis of 6-methylsalicylic acid]. *Eur J Biochem* **13**, 98-110 (1970). https://doi.org:10.1111/j.1432-1033.1970.tb00904.x
77 Sun, H. *et al.* Synthesis of (R)-mellein by a partially reducing iterative polyketide synthase. *J Am Chem Soc* **134**, 11924-11927 (2012). https://doi.org:10.1021/ja304905e
78 Marahiel, M. A. A structural model for multimodular NRPS assembly lines. *Nat Prod Rep* **33**, 136-140 (2016). https://doi.org:10.1039/c5np00082c
79 Gaudelli, N. M., Long, D. H. & Townsend, C. A. beta-Lactam formation by a non-ribosomal peptide synthetase during antibiotic biosynthesis. *Nature* **520**, 383-387 (2015). https://doi.org:10.1038/nature14100

80 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 (2006). https://doi.org:10.1073/pnas.0601924103
81 Gokhale, R. S., Tsuji, S. Y., Cane, D. E. & Khosla, C. Dissecting and exploiting intermodular communication in polyketide synthases. *Science* **284**, 482-485 (1999). https://doi.org:10.1126/science.284.5413.482
82 Maier, T., Leibundgut, M. & Ban, N. The crystal structure of a mammalian fatty acid synthase. *Science* **321**, 1315-1322 (2008). https://doi.org:10.1126/science.1161269
83 Kohli, G. S., John, U., Van Dolah, F. M. & Murray, S. A. Evolutionary distinctiveness of fatty acid and polyketide synthesis in eukaryotes. *ISME J* **10**, 1877-1890 (2016). https://doi.org:10.1038/ismej.2015.263
84 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 (2014). https://doi.org:10.1039/c3sc53511h
85 Dutta, S. *et al.* Structure of a modular polyketide synthase. *Nature* **510**, 512-517 (2014). https://doi.org:10.1038/nature13423
86 Rittner, A. & Grininger, M. Modular polyketide synthases (PKSs): a new model fits all? *Chembiochem* **15**, 2489-2493 (2014). https://doi.org:10.1002/cbic.201402432
87 Li, X. *et al.* Structure-Function Analysis of the Extended Conformation of a Polyketide Synthase Module. *J Am Chem Soc* **140**, 6518-6521 (2018). https://doi.org:10.1021/jacs.8b02100
88 Cogan, D. P. *et al.* Antibody Probes of Module 1 of the 6-Deoxyerythronolide B Synthase Reveal an Extended Conformation During Ketoreduction. *J Am Chem Soc* **142**, 14933-14939 (2020). https://doi.org:10.1021/jacs.0c05133
89 Klaus, M. *et al.* Solution Structure and Conformational Flexibility of a Polyketide Synthase Module. *JACS Au* **1**, 2162-2171 (2021). https://doi.org:10.1021/jacsau.1c00043
90 Cogan, D. P. *et al.* Mapping the catalytic conformations of an assembly-line polyketide synthase module. *Science* **374**, 729-734 (2021). https://doi.org:10.1126/science.abi8358
91 Lowry, B. *et al.* In vitro reconstitution and analysis of the 6-deoxyerythronolide B synthase. *J Am Chem Soc* **135**, 16809-16812 (2013). https://doi.org:10.1021/ja409048k
92 Klaus, M., D'Souza, A. D., Nivina, A., Khosla, C. & Grininger, M. Engineering of Chimeric Polyketide Synthases Using SYNZIP Docking Domains. *ACS Chem Biol* **14**, 426-433 (2019). https://doi.org:10.1021/acschembio.8b01060
93 Bagde, S. R., Mathews, II, Fromme, J. C. & Kim, C. Y. Modular polyketide synthase contains two reaction chambers that operate asynchronously. *Science* **374**, 723-729 (2021). https://doi.org:10.1126/science.abi8532
94 Herbst, D. A., Jakob, R. P., Zahringer, F. & Maier, T. Mycocerosic acid synthase exemplifies the architecture of reducing polyketide synthases. *Nature* **531**, 533-537 (2016). https://doi.org:10.1038/nature16993
95 Dorival, J. *et al.* Characterization of Intersubunit Communication in the Virginiamycin trans-Acyl Transferase Polyketide Synthase. *J Am Chem Soc* **138**, 4155-4167 (2016). https://doi.org:10.1021/jacs.5b13372
96 Kuhlbrandt, W. Biochemistry. The resolution revolution. *Science* **343**, 1443-1444 (2014). https://doi.org:10.1126/science.1251652

97 Piel, J. Metabolites from symbiotic bacteria. *Nat Prod Rep* **26**, 338-362 (2009). https://doi.org:10.1039/b703499g
98 Labonte, J. W. & Townsend, C. A. Active site comparisons and catalytic mechanisms of the hot dog superfamily. *Chem Rev* **113**, 2182-2204 (2013). https://doi.org:10.1021/cr300169a
99 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 (2016). https://doi.org:10.1021/acschembio.6b00345
100 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 (2016). https://doi.org:10.1016/j.jsb.2015.12.011
101 Klaus, M. & Grininger, M. Engineering strategies for rational polyketide synthase design. *Nat Prod Rep* **35**, 1070-1081 (2018). https://doi.org:10.1039/c8np00030a
102 Helfrich, E. J. N. *et al.* Automated structure prediction of trans-acyltransferase polyketide synthase products. *Nat Chem Biol* **15**, 813-821 (2019). https://doi.org:10.1038/s41589-019-0313-7
103 Scott, T. A. & Piel, J. The hidden enzymology of bacterial natural product biosynthesis. *Nat Rev Chem* **3**, 404-425 (2019). https://doi.org:10.1038/s41570-019-0107-1
104 Reimer, J. M. *et al.* Structures of a dimodular nonribosomal peptide synthetase reveal conformational flexibility. *Science* **366** (2019). https://doi.org:10.1126/science.aaw4388
105 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 (2016). https://doi.org:10.1038/nature16503
106 Drake, E. J. *et al.* Structures of two distinct conformations of holo-non-ribosomal peptide synthetases. *Nature* **529**, 235-238 (2016). https://doi.org:10.1038/nature16163
107 Bonhomme, S., Dessen, A. & Macheboeuf, P. The inherent flexibility of type I non-ribosomal peptide synthetase multienzymes drives their catalytic activities. *Open Biol* **11**, 200386 (2021). https://doi.org:10.1098/rsob.200386
108 Blin, K. *et al.* antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res* **47**, W81-W87 (2019). https://doi.org:10.1093/nar/gkz310
109 Sambrook, J. & Russell, D. W. Purification of nucleic acids by extraction with phenol:chloroform. *CSH Protoc* **2006** (2006). https://doi.org:10.1101/pdb.prot4455
110 Fitzgerald, D. J. *et al.* Protein complex expression by using multigene baculoviral vectors. *Nat Methods* **3**, 1021-1032 (2006). https://doi.org:10.1038/nmeth983
111 Geertsma, E. R. & Dutzler, R. A versatile and efficient high-throughput cloning tool for structural biology. *Biochemistry* **50**, 3272-3278 (2011). https://doi.org:10.1021/bi200178z
112 Kabsch, W. Xds. *Acta Crystallogr D Biol Crystallogr* **66**, 125-132 (2010). https://doi.org:10.1107/S0907444909047337
113 Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* **67**, 235-242 (2011). https://doi.org:10.1107/S0907444910045749

114 McCoy, A. J. *et al.* Phaser crystallographic software. *J Appl Crystallogr* **40**, 658-674 (2007). https://doi.org:10.1107/S0021889807021206
115 Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-2132 (2004). https://doi.org:10.1107/S0907444904019158
116 Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-221 (2010). https://doi.org:10.1107/S0907444909052925
117 Waterhouse, A. *et al.* SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res* **46**, W296-W303 (2018). https://doi.org:10.1093/nar/gky427
118 Scheres, S. H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J Struct Biol* **180**, 519-530 (2012). https://doi.org:10.1016/j.jsb.2012.09.006
119 Boratyn, G. M. *et al.* BLAST: a more efficient report with usability improvements. *Nucleic Acids Res* **41**, W29-33 (2013). https://doi.org:10.1093/nar/gkt282
120 Vander Wood, D. A. & Keatinge-Clay, A. T. The modules of trans-acyltransferase assembly lines redefined with a central acyl carrier protein. *Proteins* **86**, 664-675 (2018). https://doi.org:10.1002/prot.25493
121 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 (2017). https://doi.org:10.1002/anie.201611371
122 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 (2004). https://doi.org:10.1107/S0907444904026460
123 Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J Mol Biol* **372**, 774-797 (2007). https://doi.org:10.1016/j.jmb.2007.05.022
124 Blin, K., Shaw, S., Kautsar, S. A., Medema, M. H. & Weber, T. The antiSMASH database version 3: increased taxonomic coverage and new query features for modular enzymes. *Nucleic Acids Res* **49**, D639-D643 (2021). https://doi.org:10.1093/nar/gkaa978
125 Madeira, F. *et al.* Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic Acids Res* (2022). https://doi.org:10.1093/nar/gkac240
126 Gourama, H. & Bullerman, L. B. Aspergillus flavus and Aspergillus parasiticus : Aflatoxigenic Fungi of Concern in Foods and Feeds (dagger) : A Review. *J Food Prot* **58**, 1395-1404 (1995). https://doi.org:10.4315/0362-028X-58.12.1395
127 Tittes, Y. U. *et al.* The structure of a polyketide synthase bimodule core. *Sci Adv* **8**, eabo6918 (2022). https://doi.org:10.1126/sciadv.abo6918
128 Konno, S., La Clair, J. J. & Burkart, M. D. Trapping the Complex Molecular Machinery of Polyketide and Fatty Acid Synthases with Tunable Silylcyanohydrin Crosslinkers. *Angew Chem Int Ed Engl* **57**, 17009-17013 (2018). https://doi.org:10.1002/anie.201806865
129 Halbes, U. & Pale, P. A new mild procedure for the direct coupling of 1-trimethylsilyl acetylenes with vinyl triflates or aryl iodide. *Tetrahedron Lett* **43**, 2039-2042 (2002). https://doi.org:Pii S0040-4039(02)00209-5

Doi 10.1016/S0040-4039(02)00209-5
130 Nitta, M. & Kobayashi, T. Reductive Ring-Opening of Isoxazoles with Mo(Co)6 and Water. *J Chem Soc Chem Comm*, 877-878 (1982). https://doi.org:DOI 10.1039/c39820000877
131 Richter, C. D., Nietlispach, D., Broadhurst, R. W. & Weissman, K. J. Multienzyme docking in hybrid megasynthetases. *Nat Chem Biol* **4**, 75-81 (2008). https://doi.org:10.1038/nchembio.2007.61
132 Maier, T., Leibundgut, M., Boehringer, D. & Ban, N. Structure and function of eukaryotic fatty acid synthases. *Q Rev Biophys* **43**, 373-422 (2010). https://doi.org:10.1017/S0033583510000156
133 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 (2016). https://doi.org:10.1021/acscentsci.5b00321
134 Ridley, C. P., Lee, H. Y. & Khosla, C. Evolution of polyketide synthases in bacteria. *Proc Natl Acad Sci U S A* **105**, 4595-4600 (2008). https://doi.org:10.1073/pnas.0710107105
135 Kennedy, J. *et al.* Modulation of polyketide synthase activity by accessory proteins during lovastatin biosynthesis. *Science* **284**, 1368-1372 (1999). https://doi.org:10.1126/science.284.5418.1368
136 Li, Y. *et al.* Classification, prediction, and verification of the regioselectivity of fungal polyketide synthase product template domains. *J Biol Chem* **285**, 22764-22773 (2010). https://doi.org:10.1074/jbc.M110.128504
137 Xu, Y. *et al.* Rational reprogramming of fungal polyketide first-ring cyclization. *Proc Natl Acad Sci U S A* **110**, 5398-5403 (2013). https://doi.org:10.1073/pnas.1301201110
138 Simunovic, V. & Muller, R. 3-hydroxy-3-methylglutaryl-CoA-like synthases direct the formation of methyl and ethyl side groups in the biosynthesis of the antibiotic myxovirescin A. *Chembiochem* **8**, 497-500 (2007). https://doi.org:10.1002/cbic.200700017
139 Storm, P. A., Pal, P., Huitt-Roehl, C. R. & Townsend, C. A. Exploring Fungal Polyketide C-Methylation through Combinatorial Domain Swaps. *ACS Chem Biol* **13**, 3043-3048 (2018). https://doi.org:10.1021/acschembio.8b00429
140 Skiba, M. A. *et al.* Structural Basis of Polyketide Synthase O-Methylation. *ACS Chem Biol* **13**, 3221-3228 (2018). https://doi.org:10.1021/acschembio.8b00687
141 Zhang, L. & Birch, R. G. Mechanisms of biocontrol by Pantoea dispersa of sugar cane leaf scald disease caused by Xanthomonas albilineans. *J Appl Microbiol* **82**, 448-454 (1997). https://doi.org:10.1046/j.1365-2672.1997.00135.x
142 Nishimura, S. *et al.* 13-Deoxytedanolide, a marine sponge-derived antitumor macrolide, binds to the 60S large ribosomal subunit. *Bioorg Med Chem* **13**, 449-454 (2005). https://doi.org:10.1016/j.bmc.2004.10.012
143 Lee, K. H. *et al.* Inhibition of protein synthesis and activation of stress-activated protein kinases by onnamide A and theopederin B, antitumor marine natural products. *Cancer Sci* **96**, 357-364 (2005). https://doi.org:10.1111/j.1349-7006.2005.00055.x
144 Piel, J. *et al.* Exploring the chemistry of uncultivated bacterial symbionts: antitumor polyketides of the pederin family. *J Nat Prod* **68**, 472-479 (2005). https://doi.org:10.1021/np049612d

145 Schmidt, E. W. Trading molecules and tracking targets in symbiotic interactions. *Nat Chem Biol* **4**, 466-473 (2008). https://doi.org:10.1038/nchembio.101
146 Simunovic, V., Gherardini, F. C. & Shimkets, L. J. Membrane localization of motility, signaling, and polyketide synthetase proteins in Myxococcus xanthus. *J Bacteriol* **185**, 5066-5075 (2003). https://doi.org:10.1128/JB.185.17.5066-5075.2003
147 Cummings, M., Breitling, R. & Takano, E. Steps towards the synthetic biology of polyketide biosynthesis. *FEMS Microbiol Lett* **351**, 116-125 (2014). https://doi.org:10.1111/1574-6968.12365
148 Loper, J. E., Kobayashi, D. Y. & Paulsen, I. T. The Genomic Sequence of Pseudomonas fluorescens Pf-5: Insights Into Biological Control. *Phytopathology* **97**, 233-238 (2007). https://doi.org:10.1094/PHYTO-97-2-0233
149 Brakhage, A. A. Regulation of fungal secondary metabolism. *Nat Rev Microbiol* **11**, 21-32 (2013). https://doi.org:10.1038/nrmicro2916
150 Miyanaga, A., Kudo, F. & Eguchi, T. Protein-protein interactions in polyketide synthase-nonribosomal peptide synthetase hybrid assembly lines. *Nat Prod Rep* **35**, 1185-1209 (2018). https://doi.org:10.1039/c8np00022k
151 Reeves, C. D. The enzymology of combinatorial biosynthesis. *Crit Rev Biotechnol* **23**, 95-147 (2003). https://doi.org:10.1080/713609311
152 Weissman, K. J. & Leadlay, P. F. Combinatorial biosynthesis of reduced polyketides. *Nat Rev Microbiol* **3**, 925-936 (2005). https://doi.org:10.1038/nrmicro1287
153 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 (2012). https://doi.org:10.1073/pnas.1118734109
154 Eng, C. H. *et al.* Alteration of Polyketide Stereochemistry from anti to syn by a Ketoreductase Domain Exchange in a Type I Modular Polyketide Synthase Subunit. *Biochemistry* **55**, 1677-1680 (2016). https://doi.org:10.1021/acs.biochem.6b00129
155 Musiol-Kroll, E. M. & Wohlleben, W. Acyltransferases as Tools for Polyketide Synthase Engineering. *Antibiotics (Basel)* **7** (2018). https://doi.org:10.3390/antibiotics7030062
156 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 (2016). https://doi.org:10.1074/jbc.M116.730531
157 Murphy, A. C., Hong, H., Vance, S., Broadhurst, R. W. & Leadlay, P. F. Broadening substrate specificity of a chain-extending ketosynthase through a single active-site mutation. *Chem Commun (Camb)* **52**, 8373-8376 (2016). https://doi.org:10.1039/c6cc03501a
158 Wang, F. *et al.* Structural and functional analysis of the loading acyltransferase from avermectin modular polyketide synthase. *ACS Chem Biol* **10**, 1017-1025 (2015). https://doi.org:10.1021/cb500873k
159 Kalkreuter, E., CroweTipton, J. M., Lowell, A. N., Sherman, D. H. & Williams, G. J. Engineering the Substrate Specificity of a Modular Polyketide Synthase for Installation of Consecutive Non-Natural Extender Units. *J Am Chem Soc* **141**, 1961-1969 (2019). https://doi.org:10.1021/jacs.8b10521

160 Gross, H. Strategies to unravel the function of orphan biosynthesis pathways: recent examples and future prospects. *Appl Microbiol Biotechnol* **75**, 267-277 (2007). https://doi.org:10.1007/s00253-007-0900-5