NPR

Uncovering the structures of modular polyketide synthases

Journal:	*Natural Product Reports*
Manuscript ID:	NP-REV-07-2014-000098.R1
Article Type:	Review Article
Date Submitted by the Author:	11-Sep-2014
Complete List of Authors:	Weissman, Kira; Lorraine University, Molecular and Structural Enzymology Group UMR 7365 CNRS-UL : IMoPA

Journal Name

RSCPublishing

ARTICLE

Uncovering the structures of modular polyketide synthases

Cite this: DOI: 10.1039/x0xx00000x

Kira J. Weissman[a]

Received 00th January 2012,
Accepted 00th January 2012

DOI: 10.1039/x0xx00000x

www.rsc.org/

Covering: Up to 2014

The modular polyketide synthases (PKSs) are multienzyme proteins responsible for the assembly of diverse secondary metabolites of high economic and therapeutic importance. These molecular 'assembly lines' consist of repeated functional units called 'modules' organized into gigantic polypeptides. For several decades, concerted efforts have been made to understand in detail the structure and function of PKSs in order to facilitate genetic engineering of the systems towards the production of polyketide analogues for evaluation as drug leads. Despite this intense activity, it has not yet been possible to solve the crystal structure of a single module, let alone a multimodular subunit. Nonetheless, on the basis of analysis of the structures of modular fragments and the study of the related multienzyme of animal fatty acid synthase (FAS), several models of modular PKS architecture have been proposed. This year, however, the situation has changed – three modular structures have been characterized, not by X-ray crystallography, but by the complementary methods of single-particle cryo-electron microscopy and small-angle X-ray scattering. This review aims to compare the cryo-EM structures and SAXS-derived structural models, and to interpret them in the context of previously obtained data and existing architectural proposals. The consequences for genetic engineering of the systems will also be discussed, as well as unresolved questions and future directions.

Introduction

The reduced polyketide secondary metabolites, which exhibit a stunning array of functionality, stereochemistry and bioactivity (Fig. 1),[1] continue to inspire synthetic chemists to attempt their total syntheses. In the laboratory, this is often accomplished by a convergent route in which subsections are assembled and then linked together. Nature, on the other hand, uses a highly linear approach to constructing these molecules, coupling the coenzyme A (CoA)-derived building blocks head-to-tail via repeated cycles of chain extension. This process, which may be accompanied at each stage by specific modification of the newly-incorporated extension unit, is carried out by gigantic multienzyme complexes called polyketide synthases (PKSs).[2]

Chain building within PKSs is performed by a conserved trio of functional domains (Fig. 2a):[3] an acyltransferase (AT) which selects the appropriate monomer from the cellular pool, a ketosynthase (KS) domain which joins the chosen building block to the growing polyketide chain using Claisen-type chemistry, and an acyl carrier protein (ACP) whose phosphopantetheine (Ppant) prosthetic group serves as a covalent attachment point for the intermediate during the assembly process (analogous to solid-phase synthesis). The β-keto group of the resulting intermediate may or may not undergo further reductive tailoring prior to the next round of chain extension. These reactions are carried out by ketoreductase (KR), dehydratase (DH) and enoyl reductase (ER) domains, which yield respectively a β-hydroxyl, an α,β double bond, or fully-reduced methylene. The variable deployment of these reductive domains contributes significantly to the structural variety generated by PKSs relative to animal fatty acid synthases (FASs), to which they share striking functional similarities. The final, chemically tricky macrocyclization reaction is accomplished by a dedicated thioesterase (TE) domain, located at the C-terminal end of the PKS. In reflection of this division-of-labor organization, PKS are often referred to as 'molecular-scale assembly lines'. Within the PKS, the catalytic and carrier protein domains are joined together by linker regions (defined here as sequences outside of the conserved, functional domains, whether or not they are structured) and organized into operational units called modules, such that each module typically carries out one round of chain extension and reductive processing. In many cases, a single PKS polypeptide (called a subunit) will incorporate two or more such modules, leading to average molecular weights on the order of 300 kDa.

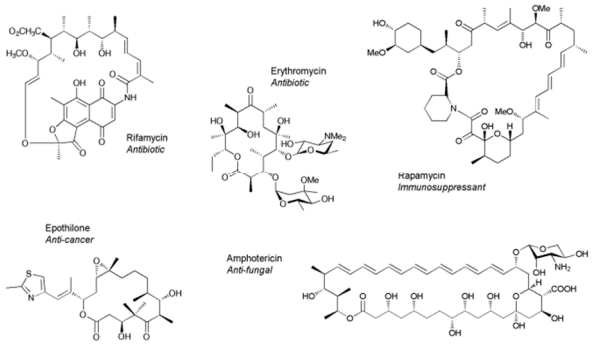

Fig. 1 Structures and biological activities of reduced polyketides.

Apparently reflecting the efficiency of this mode of biosynthesis for generating complex molecular scaffolds, Nature has convergently evolved two distinct classes of modular PKS, referred to as *cis*-AT and *trans*-AT.[4] Within *cis*-AT PKS, such as the prototypical 6-deoxyerythronolide B synthase (DEBS) responsible for assembling the core of erythromycin (Fig. 2a),[5] there is a striking relationship between the complement of domains in each module and the structure of the resulting chain extension intermediate. In contrast, the *trans*-AT PKS exhibit architectures which correlate significantly less well with the products of the biosynthesis (Fig. 2b). In addition to the presence of one or more free-standing ATs which acts iteratively to furnish extender unit to all of the modules,[6] such PKS routinely incorporate duplicated and inactive domains, additional enzymatic functions (e.g. *C*-methyl transferases (MT), GCN5-related *N*-acetyl transferases (GNAT),[7] pyran synthase (PS),[8] and β-branching (B) activities[9]), unusual domain orderings, as well as split modules, where the component domains are distributed between two polypeptides.[4] Further characteristic features include interaction in *trans* with a set of enzymes that introduces a β-methyl group using isoprenoid-like chemistry,[10] as well as the presence of nonribosomal peptide synthetase (NRPS) modules which recruit amino acids into the growing chain (Fig. 2b). Although several of these elements are also found in *cis*-AT PKS, particularly those of the marine cyanobacteria,[11] they are significantly more common in the *trans*-AT systems,[4] a difference which can be accounted for by the apparently divergent evolutionary origins of the two systems: while the *cis*-AT PKS are thought to have originated by repeated rounds of gene duplication followed by diversification of domains sets by homologous recombination,[12] the *trans*-AT PKS were apparently cobbled together from multiple gene segments, many of which were acquired by horizontal gene transfer.[13] It is nonetheless possible to predict the product of a given *trans*-AT PKS module based on the presumed substrate specificity of the downstream KS domain.[13]

On viewing the modular genetic architecture of the *cis*-AT PKS in the early 1990s − that is, the direct link between nucleotide sequence and chemical structure − it was immediately hoped that valuable analogues of predictable structure might be obtained by genetic engineering.[14] This idea led to empirical efforts, which now span several decades, to modify polyketide biosynthesis by approaches such as swapping of individual catalytic domains, modules and subunits, both within and between different PKS. These experiments, while leading in a number of cases to the anticipated products, have been nonetheless plagued by dramatic reductions in yield relative to the wild type systems and/or the presence of unexpected side products.[15] More discouragingly, many have simply failed outright. To date, no concerted efforts have been made to reprogram biosynthesis by *trans*-AT PKS, although their mosaic architectures would appear to favor such an approach.[16]

What is clear for both classes of PKS is that if genetic engineering is ever to become routine, we will need to significantly bolster our fundamental understanding of how these multienzymes operate. One of the key areas for which information has been scant until only recently, is the three dimensional architecture of PKS modules.[17] Despite intensive efforts, it has not yet been possible to solve the high-resolution structure of a PKS module by X-ray crystallography, much less that of a multi-modular subunit. This failure is likely explained by the high inherent flexibility of the multi-domain PKS proteins. This year, however, three independent groups have structurally characterized whole *cis*-AT and *trans*-AT PKS modules, using the complementary methods of single-particle cryo-electron microscopy (cryo-EM)[18,19] and small-angle X-ray scattering (SAXS).[20,21] One of the most striking common findings of these studies is that none of the observed structures agrees in detail with existing models of PKS architecture.[22–25]

A brief primer on structure elucidation by single-particle cryo-EM and SAXS

X-ray crystallography is the gold standard method in structure elucidation, because a structure solved at high resolution can reveal atomic-level details of the protein. However, the most significant hurdle with this approach remains its requirement for a good supply of diffraction-quality crystals.[26] Such crystals can be difficult if not impossible to obtain with certain multienzyme polypeptides or multi-protein complexes which exhibit high intrinsic mobility, and thus substantial structural heterogeneity. Additionally, a crystal structure only provides a static snapshot of a system, and so key features of dynamic proteins may not be captured.[26]

Both cryo-EM and SAXS are inherently lower-resolution approaches (maximum 5−10 Å for cryo-EM[27] and ~20 Å for SAXS[28]), but critically, neither method requires proteins in the crystalline state. Both techniques are also suitable for probing protein conformational dynamics. In single-particle cryo-EM (Fig. 3a),[29] radiation-sensitive specimens are analyzed in a transmission electron microscope under cryogenic conditions, generating a large number of 2D projection images featuring identical copies of the protein in different orientations. Related images are aligned and classified, and then combined computationally to generate average 3D reconstructions (or maps) of the particle at a much higher signal-to-noise ratio than the individual images. Although the obtained resolution is generally insufficient for *de novo* structure determination,[27] when atomic structures derived from X-ray or NMR are available for some or all of the sub-components of the complex, they may be fitted into the calculated density map to provide a pseudo-atomic model, considerably extending the information provided by this technique.

In a SAXS experiment (Fig. 3b), a sample in solution is illuminated by X-rays, and the radiation scattered at small angles is registered by a detector.[30] This analysis can be done in 'batch' in which the entirety of the sample is visualized, or directly following gel filtration chromatography.[31] This additional step, which allows selection of only those particles exhibiting a particular radius of gyration (*R*g) (a parameter which loosely defines as the average protein dimension), greatly increases the chances of obtaining an aggregate-free, monodisperse sample which permits confident interpretation of the SAXS data. The resulting scattering curve contains precise information about the size, shape and oligomeric state of the molecules responsible for the scattering.[32] Although the experiment yields only 1D data, it is nonetheless possible to use this information to computationally construct *ab initio* 3D models of the structure. On the other hand, as a 1D curve cannot uniquely define the shape of a three-dimensional object, what is obtained is a series of 3D forms which are all consistent with the data. These are then analyzed to identify common features, and combined to generate an

Journal Name RSCPublishing

ARTICLE

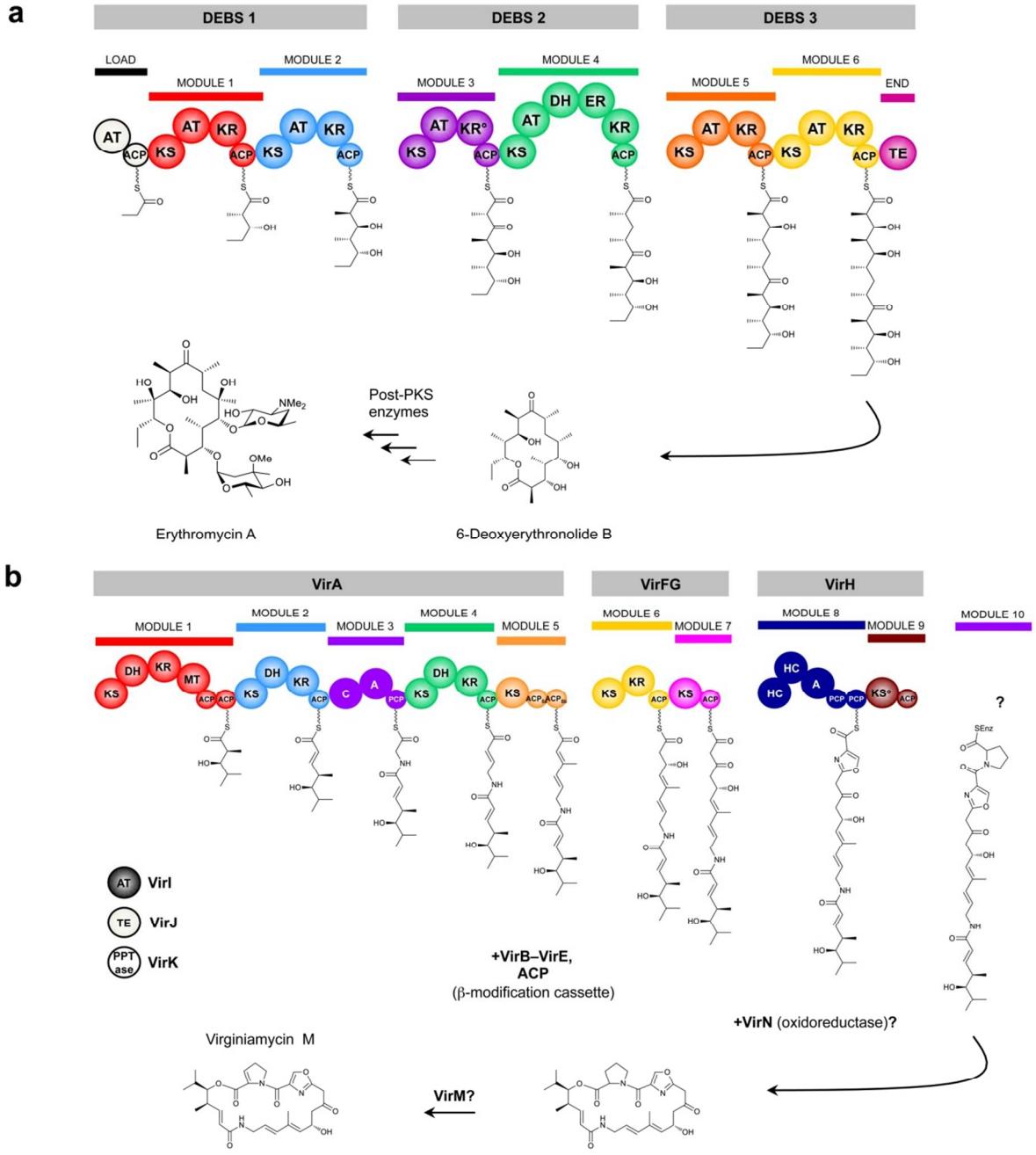

Fig. 2 Schematic of the biosynthesis by representative *cis*-AT and *trans*-AT PKS systems. a) The *cis*-AT PKS which assembles erythromycin A consists of three gigantic multienzymes (DEBS 1, DEBS 2 and DEBS 3) comprising loading, chain extension and termination modules. For this system, there is co-linearity between the order and domain composition of the modules and the structures of the resulting chain-extension intermediates. b) The *trans*-AT PKS responsible for virginiamycin biosynthesis includes at least nine chain extension modules (seven polyketide synthase and two non-ribosomal peptide synthetase); the presumed tenth module for incorporation of proline was not identified in the sequenced region.[84] This system exhibits several typical characteristics of the *trans*-AT PKS, including a discrete, iteratively acting AT (VirI), duplicated domains (ACPs of modules 1 and 5, PCPs of module 8), a condensation-inactive module (module 9), etc. Key: AT, acyltransferase; ACP, acyl carrier protein; KS, ketosynthase; KR, ketoreductase; DH, dehydratase; ER, enoyl reductase; TE, thioesterase; MT, methyltransferase; C, condensation; A, adenylation; PCP, peptidyl carrier protein; HC, condensation/heterocyclization; °, inactive domain.

Journal Name RSCPublishing

ARTICLE

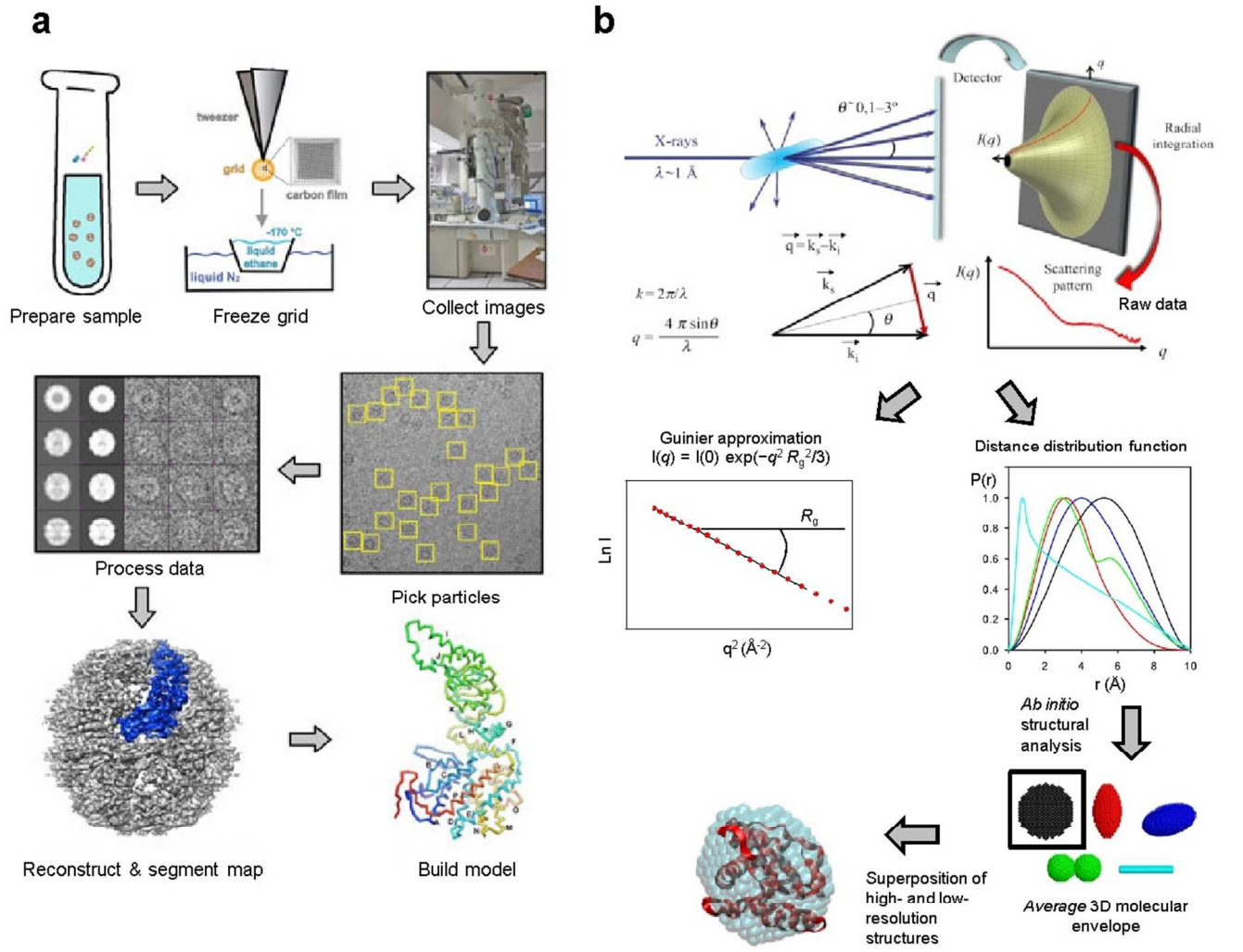

Fig. 3 Methods complementary to X-ray crystallography for studying modular PKS architecture. a) Schematic of a typical single-particle cryo-EM experiment. (Reproduced and adapted by permission of J. Zhang, http://cryoem.tamu.edu). b) Schematic of a typical SAXS experiment. Reprinted and adapted from H. D. T. Mertens and D. I. Svergun, *J. Struct. Biol.*, 2010, **172**, 128−141, and M. Czjzek, H.-P. Fierobe, and V. Receveur-Bréchot, *Methods Enzymol*. 2012, **510**, 183−210, with permission from Elsevier.

average molecular envelope. As with cryo-EM, one of the strengths of the SAXS approach is that it is then possible to place high resolution X-ray or NMR structures into the obtained average molecular form. Alternatively, architectural models of multidomain proteins can be generated directly from the NMR or crystal structures of individual domains or multidomain constructs by treating them as rigid bodies, and comparing theoretical scattering curves calculated from different configurations of the components within the complex against the actual SAXS data.[30]

PKS structural biology – the story until 2014

Prior to 2014, no structural data existed for an intact PKS module. However, information gleaned on mammalian FAS[23] fueled significant reflection on modular PKS architecture. The domain composition of animal FAS (Fig. 4a) corresponds to that of a PKS module in which the full complement of reductive domains, KR, DH and ER, is present, as well as a C-terminal TE (note: the term MAT is used instead of AT to denote the dual function of this domain as a malonyl-/acetyltransferase). Thus, the FAS can be viewed as a single PKS module which is specialized for iterative fatty acid synthesis. Furthermore, the catalytic and ACP domains of FAS and PKS exhibit convincing sequence homology,[12,33] although the corresponding linker regions are considerably less similar.[18]

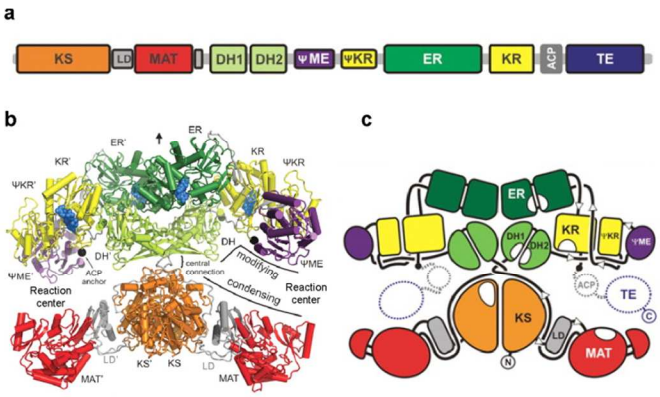

Fig. 4 Structure and organization of the animal fatty acid synthase (FAS). a) Linear sequence of domains in the FAS, drawn approximately to scale. DH1 and DH2 identify the two halves of the pseudo-dimeric 'double-hotdog' fold, while the gray boxes designate linker regions that together form a folded structural domain. b) Solved structure of the animal FAS, colored according to the domains shown in a). Structured and unstructured linker regions are indicated in gray. Bound $NADP^+$ cofactors and the attachment sites for the C-terminal ACP–TE didomain are shown as blue and black spheres, respectively. The pseudo-twofold rotational axis of the dimer is indicated by an arrow. Domains within the second polypeptide of the dimer are designated with a prime. The modifying and condensing regions of the FAS are joined together through a short linker, as indicated. c) Schematic representation of the structure shown in b). The hypothesized positions of the ACP and TE domains are indicated. From T. Maier, M. Leibundgut and N. Ban, *Science*, 2008, **321**, 1315–1322. Reprinted with permission from AAAS.

These features explain why, until recently, structural information obtained on the FAS was considered to be highly relevant to modular PKS. In this context, the most important development in the FAS field was the publication in 2006 of the 4.5 Å crystal structure of porcine FAS,[34] which was refined to 3.2 Å in 2008.[23] At the higher resolution, all of the functional domains and intervening linkers were visible, with the exception of the ACP and the downstream TE (Fig. 4b).

The structure revealed that the homodimeric FAS adopts an overall X-shaped structure in which the two polypeptides are aligned head-to-head and tail-to-tail. The dimer interface is maintained largely through contacts between the two copies of the KS and ER domains, with additional contributions from the dimeric DH. The lower 'condensing' portion of the structure comprises the KS and MAT domains, while the upper 'modifying' portion includes the three reductive domains (KR, DH and ER), but additionally a catalytically-inactive methyltransferase (ΨMT), which had not previously been identified by sequence analysis. The domains are interconnected by flexible linkers, which trace a circuitous path through the structure. The one exception is the linker lying between the KS and MAT which adopts, along with a portion of the MAT-DH linker, an αβ fold, and which serves to maintain the respective positions of the KS and MAT domains.

The disposition of the domains around the central two-fold axis of the complex gives rise to two, independent reaction centers, each housing an ACP and the downstream TE (Fig. 4c). Catalysis in the two reaction chambers appears to be asynchronous: elongation occurs on one side of the complex within a relatively closed chamber, and acyl chain reduction in the second, consequently more open reaction cleft.[35] Based on the measured distances between the active sites, the ACP was predicted to be highly mobile, shuttling back and forth across its reaction chamber to access the various domains. In fact, as revealed by subsequent analysis by single-particle negative-staining EM,[35] the entire FAS structure shows extraordinary flexibility, adopting a near continuum of conformations in which domains in the modifying portion of the structure move with respect to each other, and the lower, chain-extension region swings and swivels as a unit. Furthermore, the distribution of conformations could be influenced by the presence of substrates or by inactivation of certain catalytic domains, suggesting a direct correlation between the assumed configuration and the stage of the biosynthesis.

In the case of modular PKS, initial studies of intact DEBS subunits by limited proteolysis, chemical cross-linking and analytical ultracentrifugation, inspired the first model of modular PKS architecture,[22] in which the two polypeptides are also aligned head-to-head and tail-to-tail, but twisted about each other to form a double-helical structure – a fundamentally different topology to that of the FAS. Subsequently, it proved possible to solve the X-ray or NMR structures of all domain types present in *cis*-AT PKS (Fig. 5).[36] These studies confirmed the dimeric nature of the PKS, and showed that, at least in isolation, the structures of many of the PKS domains or didomains closely resemble their counterparts in animal FAS.[37–40] For example, the overall topology of the KS-AT didomain is identical, with the positions of the KS and AT domains fixed relative to each other by the intervening, well-folded KS-AT linker. Together, these data inspired a series of models for PKS modular structure which were clearly extrapolated from that of the FAS (Fig. 6).[23–25] Nonetheless, the structural studies also revealed several notable differences between certain PKS and FAS domains: while the ER domain of FAS is homodimeric and the TE monomeric,[23,41] PKS ERs are monomeric,[24] and the TEs homodimeric[42] (Fig. 4c, 5). And while both the PKS and FAS DH domains are homodimeric, the interfaces are formed by different structural elements.[23,43] These initial architectural discrepancies, as well as considerations of PKS mechanism,[44] raised the question of whether it was in fact valid to model the structure of modular *cis*-AT PKS on that of animal FAS.

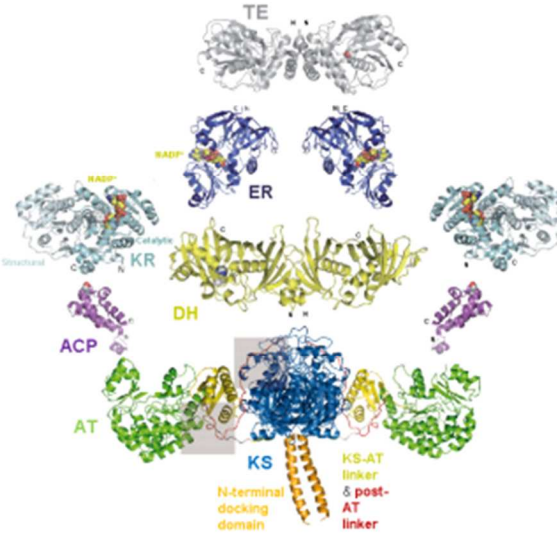

Fig. 5 High-resolution structures of individual PKS domains and didomains. Shown are the: crystal structure of the DEBS KS_5–AT_5 didomain (orange, N-terminal docking domain; blue, ketosynthase (KS); green, acyl transferase (AT); yellow, KS-to-AT linker; red, post-AT linker). Reprinted with permission from Y. Tang, C.-Y. Kim, I. I. Mathews, D. E. Cane and C. Khosla, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 11124–11129. Copyright (2006), National Academy of Sciences; NMR solution structure of DEBS ACP_2, with the position of the Ser targeted for phosphopantetheinylation shown in spheres; crystal structure of the homodimeric DEBS DH_4, with the catalytic His shown in spheres; crystal structure of KR_2 from the amphotericin PKS (A1-type KR) indicating the structural and catalytic subdomains, with the catalytic Tyr and $NADP^+$ cofactor shown in spheres; crystal structure of the monomeric ER from the spinosyn PKS, with the position of the $NADP^+$ cofactor indicated in spheres; crystal structure of the homodimeric DEBS TE. Reproduced from Ref. 38, with permission from the Royal Society of Chemistry.

Journal Name RSCPublishing

ARTICLE

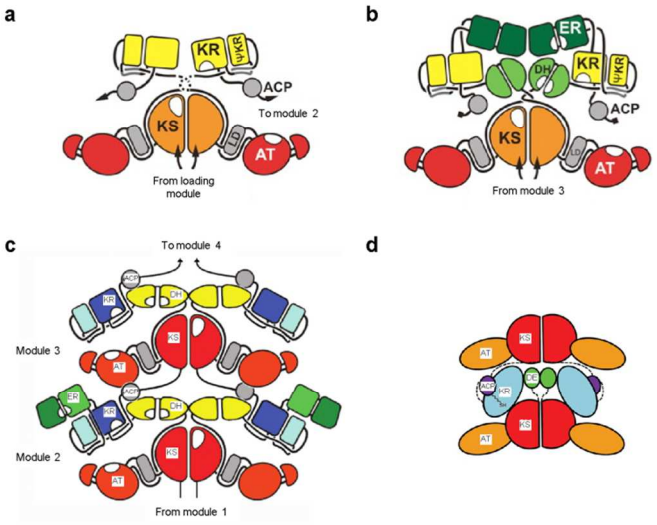

Fig. 6 Proposed models for modular PKS architecture. a) and b) Models proposed for DEBS modules incorporating just a KR domain (module 1), or the full complement of reductive domains (KR, DH and ER, module 4) following elucidation of the porcine FAS crystal structure. From T. Maier, M. Leibundgut and N. Ban, *Science*, 2008, **321**, 1315–1322. Reprinted with permission from AAAS. c) Model proposed for modules of the mycolactone PKS[100] including KR and DH, or all reductive domains, following determination of the crystal structures of DH and ER domains which revealed fundamental differences to their FAS equivalents.[24,43] Reprinted by permission from Macmillan Publishers Ltd: *Nat. Chem. Biol.*, 2012, **8**, 615−621. d) Model of spinosyn PKS module 6 inspired by the discovery and structural characterization of the post-AT dimerization domain.[25] Reprinted with permission from *ACS Chem. Biol.*, 2013, **8**, 1263−1270. Copyright 2013 American Chemical Society. In all of the illustrated models, the topology of the twin PKS polypeptides gives rises to two independent reaction chambers which each houses an ACP domain.

In the case of the *trans*-AT PKS, high-resolution structural information remains relatively scarce, as only the crystal structures of a *trans*-AT[45] and two KS domains (one as a monodomain,[46] and the other as a didomain with a branching (B) activity[9]), and the NMR structure of an ACP-ACP didomain[47] have been published to date. This paucity of insights has not permitted development of corresponding structural models for these systems, a situation complicated by their high compositional heterogeneity.

The first structures of intact PKS modules – what do they reveal?

The structure of PikAIII by single-particle cryo-EM

In June this year, Skiniotis, Smith, Sherman and colleagues achieved a breakthrough in PKS structural biology, by obtaining cryo-EM snapshots of a modular PKS modified to mimic all stages of its catalytic cycle (back-to-back papers published in *Nature* by Dutta, *et al.*[18] and Whicher, *et al.*[19]). The PKS module selected for these studies (PikAIII), performs the fifth chain extension cycle in the biosynthesis of the antibiotic pikromycin (Fig. 7a). The stand-alone nature of the module (that is, the PikAIII polypeptide contains only module 5) means that it engages in chain transfer in *trans* with both the upstream module 4 carried on PikAII and the downstream module 6 located on PikAIV. PikAIII has a relatively simple composition, as it includes just the three core domains (KS_5, AT_5 and ACP_5) and a KR (KR_5), a post-ACP dimerization element comprising two consecutive α-helices, and a C-terminal docking α-helix.[48,49] This module was presumably chosen, at least in part, because it is discrete, which obviated the need to define artificial N- and C-termini in order to express it heterologously. Reconstructions of PikAIII in multiple states were obtained at calculated resolutions of 7~11 Å. At this resolution, it was possible to fit the crystal and NMR structures of homologous domains from the DEBS PKS[37,38,40] into the EM maps by treating them as rigid bodies, leading to pseudo-atomic resolution structures.

Initial analysis of the *holo* form of PikAIII (that is, with the ACP modified with phosphopantetheine), yielded the first surprise: although the subunits were aligned head-to-head and tail-to-tail as expected,[22] together they form a symmetrical arch. Thus, unlike the two isolated reaction clefts predicted by existing models (Fig. 6),[23–25] this topology results in a *single* reaction chamber, into which all of the domain active sites face (Fig. 7b, c). The base of the arch comprises the homodimeric KS, while the AT and KR, which are positioned consecutively below the KS, form its sides. The configuration of the KS and AT domains resembles that in the double-helical model,[22] but is dramatically different from that seen in crystallographic analysis of the isolated DEBS KS-AT domains in which the two domains form an extended, slightly angular structure (Fig. 7d).[38,39] To arrive at the configuration seen in PikAIII, the KS-AT linker and the AT must rotate a full 120° relative to their previously observed positions, and certain structural elements must also rearrange to relieve steric clashes with the KS. In addition, the newly-observed AT-KR conformation means that a region of the linker following the AT (the so-called 'post-AT linker') which in the KS-AT didomain structures[38,39] contacts the side of the KS domain, must be situated elsewhere. This result serves as a warning for the potential pitfalls of extrapolating the structure of an entire module from those of smaller fragments.

The ACPs, which could be clearly visualized in the EM maps, are located in two distinct positions – either near AT_5 (43% of particle projections) or KR_5 (57%) (State 1, Fig. 8). The two ACPs occupy equivalent positions, suggesting that they move together as a unit due to the steric constraints imposed by the downstream dimerization domain (Fig. 7a).[48] Although the ACP must be charged with extender unit by the AT to initiate a catalytic cycle, its Ppant prosthetic group is not inserted into the AT_5 active site, as the Ser attachment point is directed away from the domain; likewise, the interface with KR_5 would be unproductive. Nonetheless, the fact that the ACP is confined within the reaction center must facilitate its interaction with AT_5 once the appropriate substrates are present. In this context, one obvious question is where the ACPs are located in the *apo* form of the module (whose structure was not reported), in order to permit their modification by a phosphopantetheinyl transferase (PPTase).[50] Is placement adjacent to the KR compatible with this interaction, or would the ACPs instead have to be situated outside the reaction chamber, entering only once the prosthetic group is attached? In the latter case, new contacts between the Ppant and

Journal Name RSCPublishing

ARTICLE

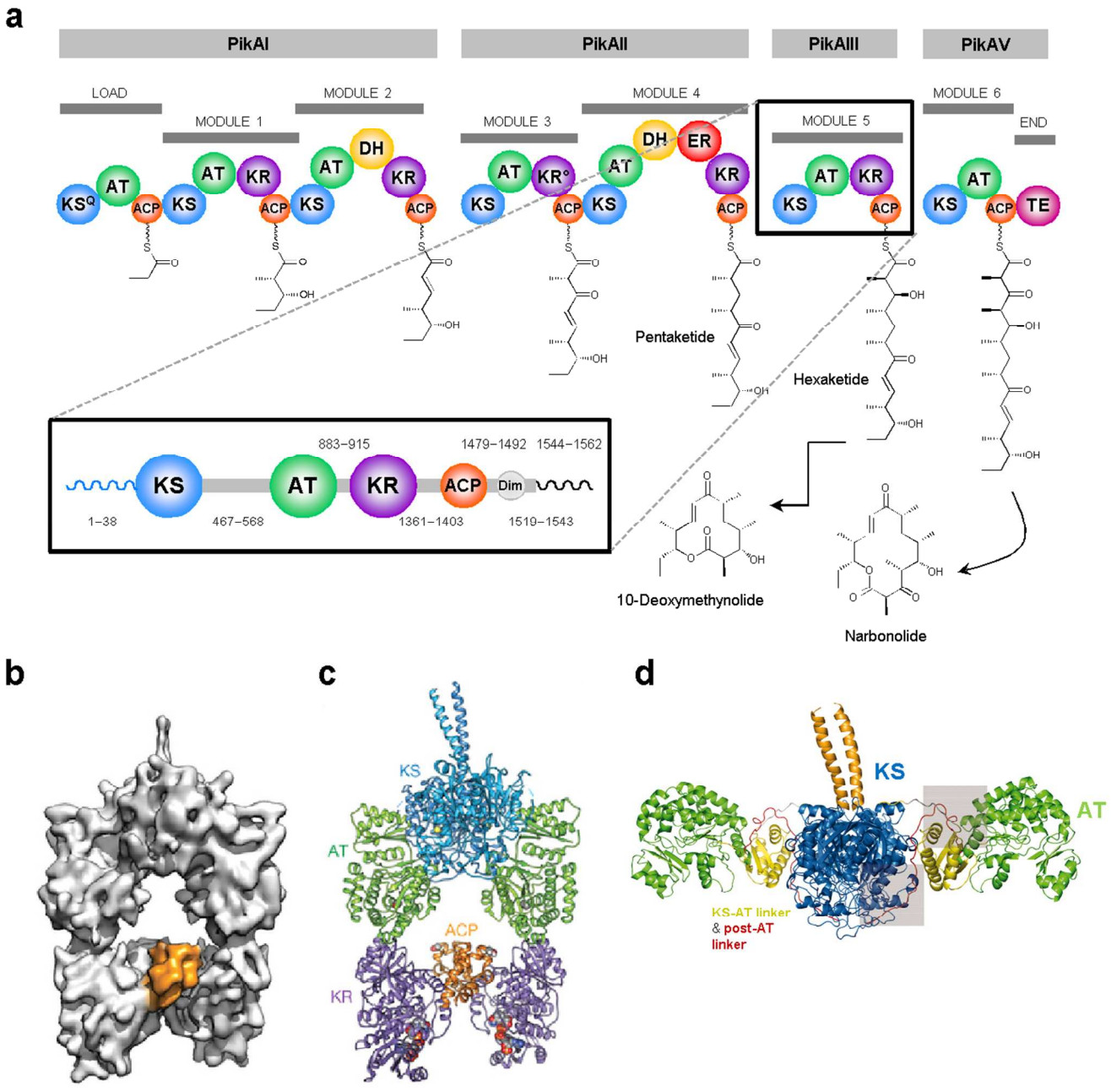

Fig. 7 Analysis of PikAIII by single-particle cryo-EM.[18,19] a) Schematic of biosynthesis by the pikromycin (Pik) PKS. The PKS comprises a loading and six chain extension modules distributed among four polypeptides (PikAI–IV). A polyketide product, either 10-deoxymethynolide or narbonolide, is off-loaded by the thioesterase (TE) domain. Inset is a schematic of PikAIII, the module selected for study by single-particle cryo-EM. The residue ranges of the docking domains (squiggles) and linker regions are shown. Key: KS^Q, decarboxylase; Dim, dimerization α-helices. Adapted from Ref. 18. b) Solid rendering of the cryo-EM map of one of two conformers of *holo* PikAIII, in which ACP_5 (orange) sits near KR_5. c) Structural model for PikAIII derived by fitting the structures of DEBS KS, AT, KR and ACP domains as rigid bodies into the cryo-EM map shown in b). Active site residues and/or cofactors are shown. Panels b and c reprinted by permission from Macmillan Publishers Ltd: *Nature*, 2014, **510**, 512−517. d) Crystal structure of discrete DEBS KS_5-AT_5 didomain, showing the divergent positioning of the KS-AT linker and AT domain relative to that in c). Reproduced with permission from Y. Tang, C.-Y. Kim, I. I. Mathews, D. E. Cane and C. Khosla, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 11124–11129; copyright (2006), National Academy of Sciences.

Journal Name RSCPublishing

ARTICLE

the catalytic domains could favor the sequestration of the ACPs inside the reaction cleft.

During polyketide biosynthesis, each KS catalyzes two distinct reactions: transfer of the growing intermediate from the upstream ACP (ACP_{n-1}) onto its active site Cys, and chain elongation using extender unit tethered to the downstream ACP, ACP_n (Fig. 9a).[51] This dual functionality raises questions about how the KS discriminates between its two acyl-ACP partners, thus avoiding mis-transfer of the extended chain back onto its catalytic Cys which would lead to iteration by the module. Previous analyses by computational docking in combination with NMR and site-directed mutagenesis,[52–54] have suggested that both the upstream and the downstream partner ACPs of a given KS_n dock into a deep cleft formed by the KS, the AT and the intervening KS-AT adaptor region (Fig. 9b). The precise docking interfaces (and thus, modes of substrate presentation to the KS active site) were suggested to be distinct, explaining how the KS distinguishes between the two ACPs. Although mutagenesis based on this idea was used to induce ACP_n to KS_n back transfer and thus iterative behavior from a model DEBS module,[53] it does not agree with the new data obtained by Dutta, *et al.*

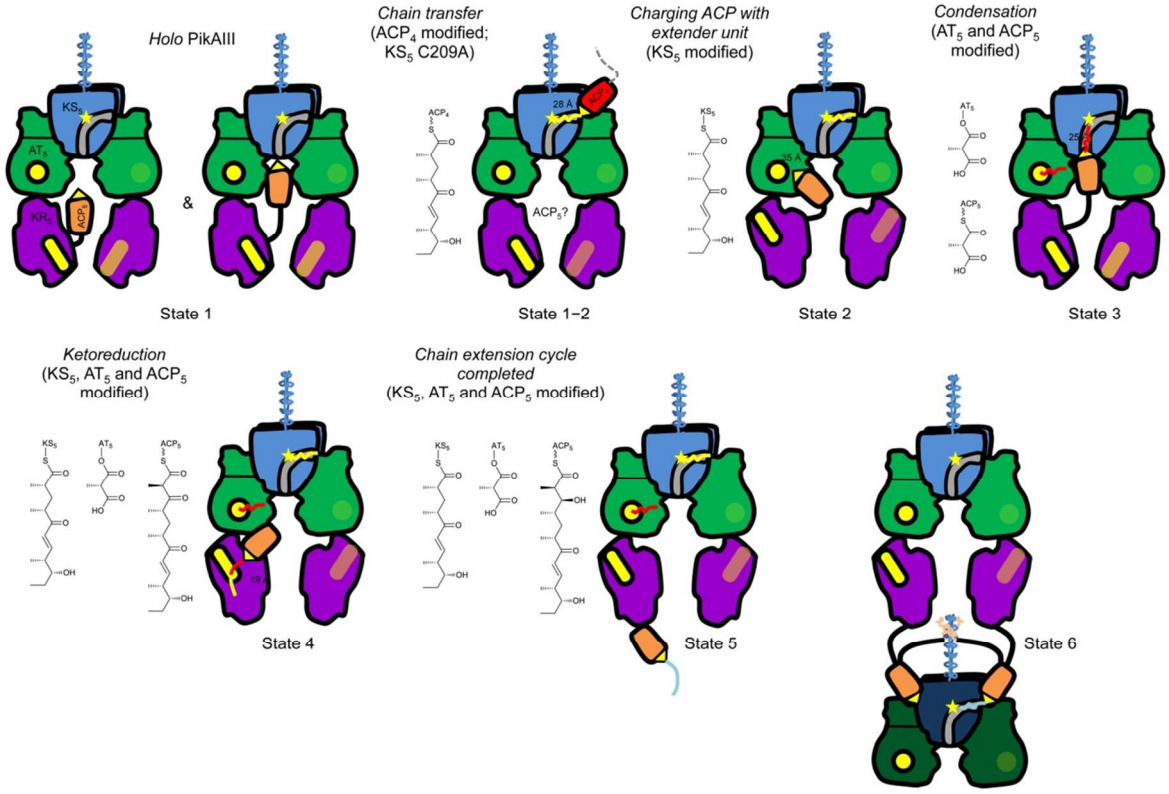

Fig. 8 Conformational states identified for PikAIII as a function of covalent modification of one or more component domains.[18,19] The domains are color-coded as in Fig. 7, with the locations of the active sites/residues indicated in yellow. Where appropriate, the structures of the substrates attached to the various domains are shown. The numbers indicate the distance estimated between the catalytic residues of specific domains and the Ser phosphopantetheine attachment point of the ACP. State 1 represents the *holo* protein (ACP modified with phosphopantetheine cofactor). In this state, the ACPs (only one is shown for clarity) occupy one of two positions, either adjacent to the KR (57%) or to the AT (43%). State 1−2 represents the conformation which permits transfer of the chain extension intermediate from the upstream module to PikAIII. To allow unequivocal identification of the ACP_4 domain, ACP_5 was removed from the construct – thus its location at this stage of the biosynthesis is unclear. In State 2, the KS is now modified with pentaketide intermediate (yellow line), which prompts the ACP to localize near the AT active site. Notably, the KR has flipped end-to-end with respect to its position in States 1 and 1–2. In State 3, the ACP is now acylated with methylmalonate (as is the AT; red lines), readying it to participate in chain extension with the KS (although in this case, the KS is unmodified). Correspondingly, the ACP now occupies a position below the KS, where it can access its bottom active site entrance. Curiously, the KR at this stage is unflipped. State 4 represents the system following chain extension (β-ketohexaketide-ACP), although both the KS and the AT are also modified. At this stage, the ACP is positioned for the next step, ketoreduction, at the KR domain (which resumes its flipped configuration); from this location, the intermediate (red−yellow line) can be inserted into either side of the active site. Following ketoreduction to generate the β-hydroxyhexaketide-ACP (blue line), the ACP is ejected out of the reaction chamber (State 5), freeing it to engage in chain transfer with the downstream module located on PikAIV. Shown is one of three conformers observed for the module, in which only the position of the ACPs varies relative to the catalytic domains. State 6 (not directly characterized) is the conformation proposed to allow transfer of the intermediate to PikAIV. The positioning of the ACPs on either side of the homodimeric KS would seem to necessitate unraveling of the post-ACP dimerization domain. Adapted from Ref. 19.

Journal Name RSCPublishing

ARTICLE

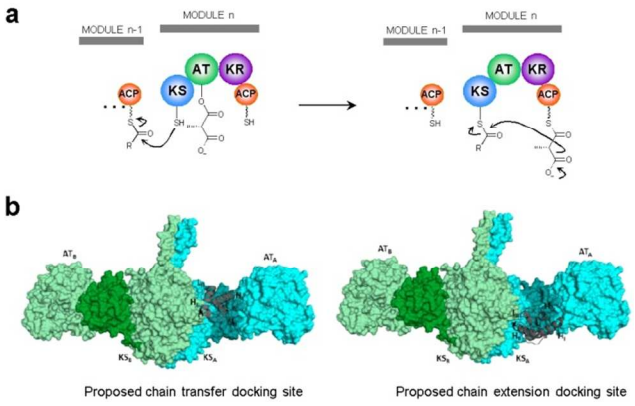

Fig. 9 Mechanistic aspects of the KS-catalyzed reaction. a) During polyketide biosynthesis, the KS catalyzes two distinct reactions. In the first, the KS transfers the chain extension intermediate tethered to the upstream ACP (ACP_{n-1}) onto its catalytic Cys. It then cooperates with the ACP of its own module (ACP_n) in chain extension. To avoid inappropriate back-transfer of the extended chain to the KS, it must have a means to distinguish between the two ACPs. b) Model for ACP discrimination suggested by computational docking in combination with NMR and site-directed mutagenesis.[52–54] According to this model, the KS is able to distinguish between ACP_{n-1} and ACP_n as they dock against two distinct motifs formed by the KS in combination with the AT and the intervening KS-AT adaptor region, changing their mode of substrate presentation to the KS active site. Reprinted from S. Kapur, B. Lowry, S. Yuzawa, S. Kenthirapalan, A. Y. Chen, D. E. Cane and C. Khosla, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**, 4110–4115.

To investigate the issue of how chain transfer occurs *in trans* between ACP_4 (PikAII) and KS_5 (PikAIII) during pikromycin biosynthesis, Dutta, *et al.* generated a version of PikAIII lacking ACP_5 but including the downstream sequences (PikAIII(ΔACP_5)), as a construct in which the dimerization helices were absent was unexpectedly found to be monomeric. Interestingly, the cryo-EM structure of PikAIII(ΔACP_5) revealed that the KR domains had rotated by 165° about each wall of the arch, showing that the position of the KRs depends on the presence of the following ACP domain. To trap the intersubunit interaction, they fused the N-terminal docking domain of PikAIII(ΔACP_5) to the docking domain C-terminal to ACP_4 via a 10 amino-acid linker.[55] Using the broad specificity PPTase, Svp,[56] they then attached the native pentaketide generated by module 4 to ACP_4, and coupled this with mutation of the KS_5 catalytic Cys to Ala, to prevent ACP_4-to-KS_5 chain transfer. In the resulting 8.6 Å resolution cryo-EM structure of pentaketide-ACP_4-PikAIII(C209A/ΔACP_5) (State 1−2, Fig. 8), ACP_4 sits at the side entrance of KS_5 at an estimated distance of 28 Å from the KS_5 active site. In this configuration, the domain lies outside of the module 5 reaction chamber, establishing a clear, structural basis for discrimination between itself and ACP_5. And while the ACP_4 interface residues agree with the earlier analyses,[53] the overall mode of interaction is entirely different, as the recognition surface on PikAIII does not include the AT or the KS-AT linker. Interestingly, in the absence of tethered pentaketide, *holo* ACP_4 does not contact KS_5, showing that protein-protein interactions between the matched docking domains and ACP_4/KS_5[49] are insufficient to define a stable interface. This result is consistent with the weak affinity measured for intersubunit interactions in *cis*-AT PKS (2−24 μM),[49,55,57] which places them in the class of transient protein-protein interactions,[58,59] and suggests that the substrate makes a decisive contribution to docking.

As ACP_5 resides within the reaction chamber, it was not obvious how it would access the identified KS_5 side entrance. To address this question, Dutta, *et al.* solved the structure of PikAIII in which ACP_5 was loaded with methylmalonate (MM) using the natural acyl transfer activity of AT_5 (State 3, Fig. 8). In the normal catalytic cycle, this modification would prime the ACP to engage with the KS in chain extension. In this case, however, the KS was unmodified, raising the question of whether the state of the module accurately reflects that prior to condensation, as KS acylation provokes large conformational adjustments of the PKS (*vide infra*). In any case, in the resulting 7.3 Å resolution cryo-EM structure in which both ACP_5 and AT_5 carry bound extender unit, the ACP resides at the bottom of the KS, ~25 Å from its active site. It has thus moved ~20 Å from its location near AT_5 in the *holo* enzyme. As this position is far from the active site entrance used by the chain transfer ACP_4, this structure identified a second, independent route to the KS_5 catalytic machinery via the central chamber. (On the other hand, as this alternative entrance is occluded by loops in homologous KS from non-modular PKS and FAS pathways,[23,60–63] correct extender unit positioning is apparently possible via both active site channels.) In light of the identified interaction surfaces, both ACPs could in principle dock onto the KS simultaneously – potentially explaining the direct ACP-to-ACP transfer observed in some engineered PKS systems which results in aberrant 'skipping' of chain extension by a specific module.[64]

In the companion paper, Whicher, *et al.*[19] probed the state of PikAIII derivatized to mimic all of the remaining stages in the reaction cycle. In the first instance, they used a thiophenol-activated form of the native pentaketide to acylate KS_5, while leaving ACP_5 in its *holo* form. (What is slightly curious about this and several subsequent experiments, however, is that the same substrate was reported only to 'prefer' KS_5 relative to ACP_5 in the context of intact PikAIII,[65] and indeed, thiophenol esters have been shown to efficiently load the Ppant arms of peptidyl carrier protein (PCP) domains from NRPS.[66–69] From the published data, it is not clear whether any PikAIII molecules carried modifications on *both* the KS and ACP domains, or at any other locations.) The resulting 7.9 Å resolution structure (State 2, Fig. 8) suggested that acylation of the KS Cys induces several conformational changes in the KS domain which are then transmitted allosterically to the adjacent AT. As a result, the AT moves by ~2 Å to partially-occlude the KS_5 side-entrance, potentially protecting the labile KS thioester from hydrolysis. Remarkably, modification of KS_5 also leads ACP_5 to relocate by 10 or 30 Å, respectively, from its positions in the *holo* structure, to form an interface with the AT_5 domain. The reactive Ser of the ACP is 'directed' towards the AT active site in this configuration, although an estimated 35 Å separates the two. While these results suggest that transfer of the growing polyketide to the KS leads to a domain arrangement favorable to charging of the ACP with extender unit by the AT, the available literature data[70–72] strongly argue that PKS modules *in vivo* are heavily charged with extender units on both the AT and ACP domains even in the absence

of acylation of the upstream KS. Thus, the observed conformation may be an emergency response of the module which promotes rapid loading with extender unit.

Rather unexpectedly, Whicher and colleagues also observed that the conformational effects provoked by KS modification are transmitted to the AT_5-KR_5 linker, inducing KR_5 to flip over end-to-end relative to its position in the *holo* and methylmalonate-modified structures (compare State 2 to States 1 and 1−2, Fig. 8). These domain acrobatics reverse the positions of the KR structural and catalytic sub-domains,[37] bringing its active site in proximity to that of the AT near the center of the reaction chamber. Overall, these results in fact imply that the KR flips back and forth during the first three stages of the reaction (compare States 1−3, Fig. 8). On the other hand, as noted previously, State 3 may not fully represent the chain extension configuration, as AT_5 and ACP_5 are modified, but KS_5 is not, leaving open the possibility that KR_5 would also adopt the flipped configuration in a more native, triply-charged protein. If this were the case, only the *holo* protein (and possibly the *apo* form) would incorporate an un-flipped KR.

To investigate the next stage in the biosynthesis, Whicher, *et al.* allowed chain extension to proceed by incubating *holo* PikAIII with thiophenol-pentaketide and methylmalonyl-CoA, leading to β-ketohexaketide tethered to ACP_5. As an excess of both substrates was used, the final protein was also modified with pentaketide on KS_5 and methylmalonate on AT_5 (State 4). Although this method avoided the synthesis of an additional advanced precursor, the resulting triply-modified state may not be strictly relevant to the global catalytic cycle. Recharging of KS_5 following condensation would rely on rapid chain transfer from ACP_4, a reaction that appears unlikely given that AT_5 in the β-ketohexaketide-PikAIII structure has shifted towards KS_5 by 8 Å, completely occluding its side entrance (in this context, it must be assumed that the observed modification of KS_5 with pentaketide occurred via its lower active site entrance). Indeed, the authors argue on the basis of this structure that movement of AT_5 is designed to prevent premature handover of intermediate from the upstream PikAII. Given the significant conformational consequences of KS acylation, it will be of interest to study the structure of PikAIII modified only with methylmalonate on AT_5 and β-ketohexaketide on ACP_5 (generated by, for example, transfer from the corresponding CoA using Svp).

In the β-ketohexaketide-PikAIII structure (State 4, Fig. 8), ACP_5 docks against the 'lid helix' and 'lid loop' of the catalytic subdomain of KR_5 (Fig. 5), positioning its active Ser at a distance of 19 Å from the active site. KR_5 is classified as an A2-type KR, as it catalyzes NADPH-dependent reduction from the *re* face of the β-keto substrate, without prior epimerization at the adjacent α-methyl center.[73,74] KRs are thought to control the stereochemistry of ketoreduction by guiding the substrates into one or the other sides of their active sites, thus situating them appropriately with respect to the NADPH cofactor. From its observed position, ACP_5 could in principle deliver its substrate into the KR_5 active site from either direction. Thus, the directionality of entry does not appear to be determined by the ACP docking mode, but instead by intrinsic features of the KR active sites. Indeed, it has been shown that the active sites of A-type KRs are well formed even in the absence of substrate, while in B-type KRs, which catalyze reduction from the opposite *si* sense, the lid helix and lid loop only become ordered upon substrate binding.[75] It is thus somewhat surprising that Whicher, *et al.* found that the electron density corresponding to PikAIII KR_5 was better modeled with the structure of the *holo* B1-type KR of DEBS module 1[37] than by several A-type KRs. In any case, the detailed mechanism by which the KRs control substrate entry awaits higher-resolution structural information on KR domains in the presence of their substrates.

To generate the state corresponding to the completed chain extension cycle, PikAIII was incubated with thiophenol-pentaketide and methylmalonyl-CoA, but additionally with NADPH, permitting the KR_5-catalyzed reaction to occur. In fact, mass spectrometry analysis revealed that reduction of the β-ketohexaketide-ACP to β-hydroxyhexaketide-ACP only went to ~50%, and thus it was presumably necessary to separate projections representing State 4 from the newly-formed State 5 (Fig. 8), in order to reconstruct the cryo-EM maps. Again, under the experimental conditions, the final protein would also have been acylated on the KS_5 and AT_5 active sites with pentaketide and methylmalonate, respectively, which somewhat complicates interpretation of the data. In the resulting structure (State 5), ACP_5 resides in three distinct positions below the KR_5 structural subdomain. Thus, with the catalytic cycle now finished, the ACP-tethered substrate has been ejected from the reaction chamber and is oriented away from the module, facilitating its transfer to the downstream PikAIV. This result provides a physical basis for the 'retardation control' mechanism proposed by Staunton and Hill, in which each PKS module ensures that all keto group modifications go effectively to completion prior to handing off of the intermediate to the next module.[76] On the other hand, the catalytic domains of PikAIII occupy essentially the same positions as in State 4, with AT_5 continuing to fully block the side entrance to the KS_5 active site. Thus, even though PikAIII could in principle initiate another round of chain extension, acylation of the KS_5 Cys by the upstream ACP_4 appears to be precluded. Presumably, an additional conformational change, perhaps associated with docking of PikAIII against PikAIV (State 6, Fig. 8), must occur in order to free up the KS side channel.

A final, major contribution of these papers is to help refine a model for ACP-based interactions in modular PKS.[77] Consistent with computational studies on fungal FAS,[78] the ACP may search out potential partners by diffusion within the reaction chamber. What the study by Whicher, *et al.* indicates is that this 'random walk' leads the ACP to sound out potential docking sites on the various partner domains via complementary charge:charge interactions. A specific partner would then be selected when the catalytic domain recognizes its correct substrate attached to the ACP – providing additional contacts which stabilize the otherwise highly transient complex. In this way, the essentially random movements of the ACP could be channeled into the sequence of ACP/partner interactions which precisely corresponds to the chemistry required for chain extension and processing. On the other hand, not all catalytic steps require formation of a specific protein-protein complex, as at least in the DEBS system, the TE domain has been shown to interact only with the substrate attached to the upstream ACP, not the domain itself.[79] Given the broad substrate specificity of the TE domains,[80–83] such a mechanism would seem to risk premature off-loading of the polyketide chains. The finding by Whicher and colleagues that the ACP segregates into its reaction chamber until reductive processing is complete nicely resolves this programming issue.

Structural characterization of DEBS module 3-TE and DEBS 3 by SAXS

Just prior to publication of the PikAIII structures, the Khosla laboratory reported analysis by SAXS of several PKS modules (Edwards, *et al.*[20]). Two multienzymes were analyzed: DEBS module 3 fused to the off-loading thioesterase of the same system (module 3-TE), and the intact bimodular subunit, DEBS 3, comprising modules 5 and 6 (Fig. 2a); module 3-TE was obtained in its *holo* form, but the modification state of DEBS 3 was not

Journal Name RSCPublishing

ARTICLE

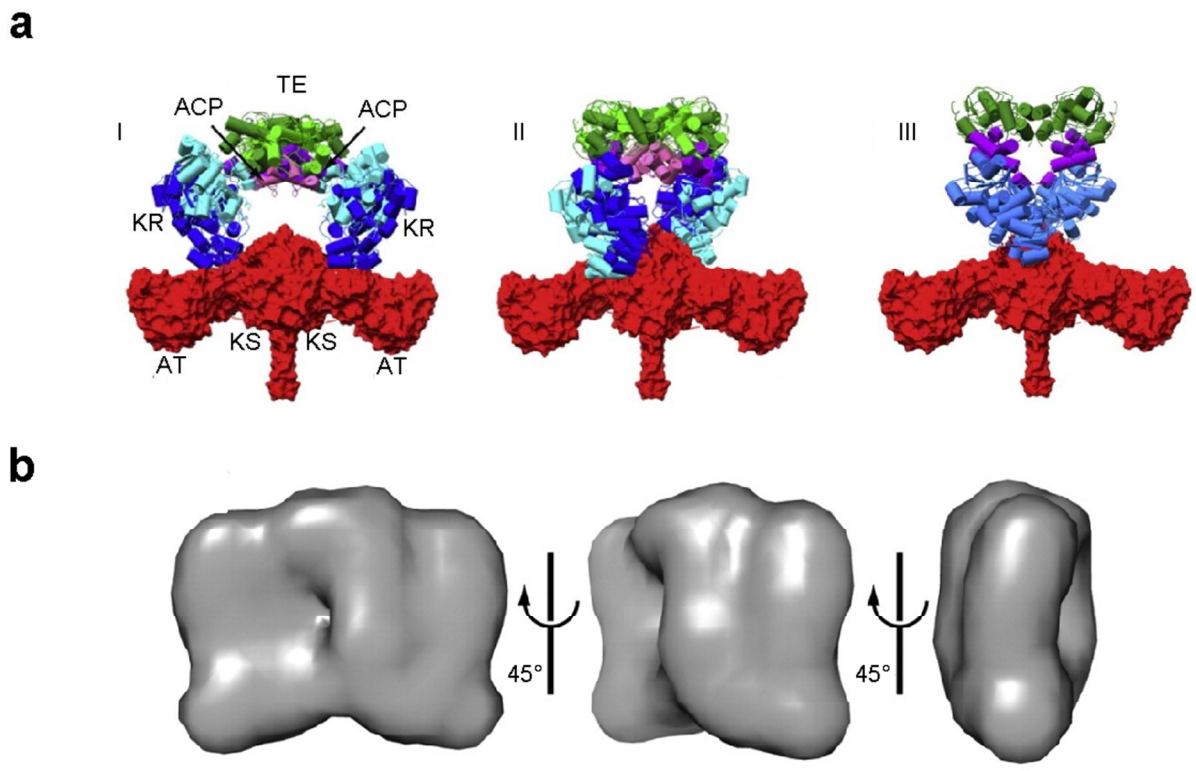

Fig. 10 Analysis by SAXS of DEBS module 3-TE. a) Three architectural models derived by rigid-body modeling of the component domains against the obtained SAXS data. In all cases, the AT domain sits outside the single reaction chamber, while the ACPs are positioned adjacent to the C-terminal TE. b) Three orientations of the molecular envelope of module 3-TE calculated with imposed second-order symmetry. Reprinted from A.L. Edwards, T. Matsui, T. M. Weiss and C. Khosla, *J. Mol. Biol.*, 2014, **426**, 2229−2245, with permission from Elsevier.

explicitly stated. All of the modules analyzed have the same domain composition as PikAIII (KS-AT-KR-ACP), but differ in that in each case, the ACP is followed by a homodimeric domain (KS (in the case of module 5) or TE (modules 3 and 6)).

SAXS analysis was initially performed on mono- and didomain fragments of module 3-TE (KS-AT, KR, ACP and TE), yielding calculated average molecular envelopes in good agreement with the atomic-resolution structures revealed by crystallography or NMR. Based on this finding, Edwards and colleagues opted to use the high-resolution structures as the basis for rigid body modeling of the overall multi-domain complex with imposition of second-order symmetry, refining their obtained models (Fig. 10a) against the SAXS scattering data. The SAXS data were additionally used to calculate an average *ab initio* molecular form, with approximate dimensions of 180 Å × 170 Å × 80 Å (Fig. 10b).

Three rigid body refinement models consistent with the data were obtained (Fig. 10a), suggesting the existence of multiple conformational states. However, none of these agrees in detail with the PikAIII structures. The most obvious difference is the shape of the KS-AT portion of the module, which because it was treated as a unit during rigid body modeling, retains its extended form. This configuration serves to exclude the AT from what would otherwise be a central reaction chamber containing all of the other domains. While the KR domains are located below the KS homodimer as in PikAIII, the ACPs in the three models are not docked near the KR and AT active sites, but situated adjacent to the TE homodimer. This divergent positioning does not appear to be a consequence of the downstream TE, as rigid body modeling of module 3 lacking the TE produced the same ACP configuration.

One possible origin of the discrepancies between the SAXS- and cryo-EM derived models is that useful SAXS data were obtained for only the lowest scattering angles ($q < 0.15$; $q = 4\pi\sin\theta/\lambda$ where 2θ is the scattering angle and λ is the wavelength of the incident beam), limiting resolution to an estimated 40 Å. Nonetheless, the overall shape of the favored rigid-body model (I, Fig. 10a) fits well within the flat, disc-shaped molecular envelope (Fig. 10b), raising the possibility that DEBS module 3 actually adopts the observed configuration (in addition to those described for PikAIII) in solution. On the other hand, one wonders whether Edwards and colleagues would have obtained a domain arrangement closer to that in PikAIII, by attempting to place the KS and AT domains individually into the calculated molecular form.

Analysis of DEBS 3 at ca. 50 Å resolution revealed that the particle is dimeric and ellipsoidal, with dimensions of approximately 300 Å × 190 Å × 100 Å. Consideration of how subunits interact via their docking domains[48,55] predicts that consecutive modules will

Journal Name RSCPublishing

ARTICLE

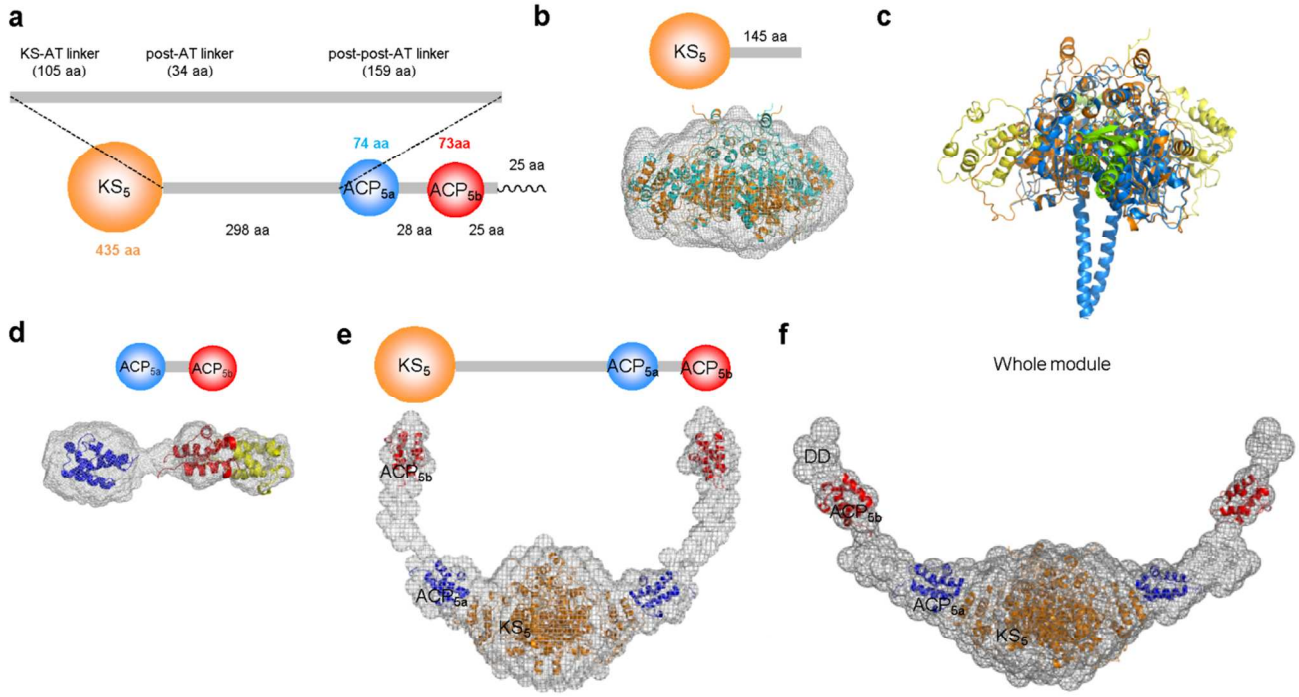

Fig. 11 Analysis by SAXS of VirA module 5. a) Schematic of VirA module 5. The residue ranges of the docking domain (squiggle) and linker regions are shown, as are the domain sizes. The first two portions of the 298 amino acid linker between the KS_5 and ACP_{5a} domains show homology consecutively to the KS-AT and post-AT linkers identified in structures of discrete KS-AT didomains from *cis*-AT PKS. b) Superposition of KS_5 homology models on the calculated molecular envelope for the KS_5-linker construct as determined by SAXS. The model in orange was generated using DEBS KS_5 as template, while that in cyan was obtained following asymmetric rotation of the two KS-AT linkers. c) Superposition of the VirA KS_5-linker homology model on the KS_5-linker region of PikAIII (PDB file kindly provided by G. Skiniotis). The VirA KS_5 is shown in orange and the linker in yellow, while the KS_5 of PikAIII is shown in blue and the linker in green. d) Superposition of the NMR structures of ACP_{5a} and ACP_{5b} on the molecular envelope calculated for the ACP_{5a}-ACP_{5b} didomain. This analysis shows that one of the two domains undergoes a spring-like movement along the inter-ACP axis, but that it does not contact the other ACP domain. e) The molecular envelope obtained by SAXS analysis of KS_5-ACP_{5a}-ACP_{5b}, with fitting of the individual domains. The module adopts an arch-like shape, with the KS at the base and the ACPs and intervening linker forming the walls. The distinct positioning of ACP_{5a} and ACP_{5b} suggests that they play non-equivalent roles in the biosynthesis. f) Molecular envelope derived from SAXS for intact module 5. The presence of the presumed C-terminal dimerization sequences did not induce closure of the module, leading to their reassignment as putative docking domains (DD).

adopt a co-linear configuration – a supposition entirely consistent with the measured length of the DEBS 3 protein, and the form of the obtained average molecular envelopes. By rigid body modeling, Edwards, *et al.* identified domain positions analogous to those in module 3-TE, but with modules 5 and 6 rotated with respect to each other by as much as 70° relative to the *xy*-plane of module 5 (a plane running the length of the module, and passing through the KS-AT didomain and the ACP) (Fig. 10). Another significant prediction from this analysis is that even when modules are covalently linked, interactions between them are minimal, a feature proposed to facilitate PKS evolution by whole module duplication.

Structural characterization of VirA module 5 by SAXS

Davison and colleagues likewise used SAXS to study the structure of an intact module derived from the *trans*-AT PKS responsible for synthesizing the antibiotic virginiamycin[84] (Fig. 2b). The chosen module, module 5, sits at the C-terminal end of subunit VirA, and comprises a KS and a tandem of ACP domains (ACP_{5a} and ACP_{5b}), separated by linkers of variable length (Fig. 11a). Although architecturally simple, module 5 interacts with a large number of partners, providing a test case for understanding protein-protein interactions. These include module 4 (*in cis*), module 6 located on subunit VirFG, the discrete AT VirI, the PPTase VirK, a proof-reading (type II) thioesterase[85] VirJ, and a β-methylation cassette (VirC−VirE) (*in trans*).[10] Indeed, duplicated ACP domains are characteristic of modules which introduce β-methyl groups, where they are thought to function either in parallel or in series to increase the overall throughput of the biosynthesis.[10,86]

To investigate the architecture of *apo* module 5 by SAXS, Davison, *et al.*[21] took a 'dissect and build approach', studying several modular fragments in order to reconstitute the structure and dynamic behavior of the whole module.[87] One construct analyzed included KS_5 and 139-residues of the downstream 298 amino acid linker (Fig. 11a). Successive portions of this linker show homology to the KS-AT and post-AT linkers indentified in the structures of

excised KS-AT didomains from *cis*-AT PKS, and homologous sequences are present in ~90% of *trans*-AT PKS modules.[46] In the two crystal structures of KSs from *trans*-AT PKS (PksKS2[46] and a KS-B didomain,[9] the corresponding linker regions together form the same αβ fold observed in the KS-AT didomains, despite the absence of the AT. Similarly, a homology model of the homodimeric VirA KS_5-linker region which incorporated this structural subdomain, fit well into the SAXS-derived molecular envelope (Fig. 11b). Nonetheless, improved agreement to the data was obtained by rotating the linker regions with respect to the KS by 18 and 21°, respectively (Fig. 11b).[9,46] Even in this alternative configuration, however, the positioning of the subdomain relative to the KS differs substantially from that seen in the PikAIII structure (Fig. 11c). This observation raises the question of whether the location of the KS-AT linker in this fragment of the module reflects its native orientation, given the difference reported by Dutta, *et al.* between its positioning in excised KS-AT didomain constructs and the intact module. To explain this discrepancy, they noted that all PKS modules incorporate C-terminal dimerization domains which exert an 'outsize' influence on modular structure, and which were absent from the truncated proteins.[18] In the case of *trans*-AT PKS, however, sequence analysis (KJ Weissman, unpublished observations) strongly suggests that many *trans*-AT PKS subunits, including VirA (*vide infra*) and RhiE (origin of the KS-B crystallographic fragment), lack such C-terminal dimerization elements. In light of this, the position of the structural subdomain observed in KS-containing fragments of *trans*-AT PKS modules may accurately reflect the native architecture.

A second important result came from examination of a portion of module 5 containing just the two ACP domains and the intervening 28-residue linker. In the obtained dumbbell-shaped molecular envelope, atomic distribution corresponding to both the domains and the linker was clearly visible (Fig. 11d). Furthermore, fitting of the NMR structures of the ACPs into the form showed that one of the two domains undergoes a spring-like motion along the inter-ACP axis due to the flexibility of the tethering region, but that it does not directly contact the second ACP. The absence of inter-ACP interaction was subsequently confirmed by independent NMR analysis of the ACP_{5a}-ACP_{5b} didomain, and is consistent with previous literature data on other multi-ACP constructs.[47,88] This result underlies the utility of SAXS for revealing conformational flexibility in solution.

Davison and colleagues next investigated a construct incorporating the three functional domains, KS_5, ACP_{5a} and ACP_{5b}, using the data obtained on the two sub-fragments to place the domains into the resulting molecular envelope (estimated resolution 20 Å). This experiment revealed that, like PikAIII, module 5 adopts an overall arch-like shape with the KS at its base (Fig. 11e). On the other hand, in the absence of additional catalytic domains (AT, KR, etc.), the walls of the arch are formed by the two ACP domains themselves and the intervening linker region. As a consequence of this organization, ACP_{5a} is tucked up against KS_5, while ACP_{5b} is situated at the extremity of the module, and is highly accessible – a differential localization which could account for the in-series mode of operation observed for certain tandem ACPs of *trans*-AT PKS.[86] According to this model, ACP_{5a} would participate in chain extension with KS_5, while ACP_{5b} would serve as the site for modification at the β-center, before handing off the intermediate to module 6. On the other hand, while the measured inter-ACP distances are compatible with the direct ACP-to-ACP transfer required by this model, the estimated distance between the ACP_{5a} Ser and the KS active site (50 Å) is too large to permit an interaction. As with PikAIII, it may be necessary to study substrate-modified versions of module 5 in order to visualize the conformational change(s) which permits the condensation reaction to occur.

The other notable result from the form calculation on KS_5-ACP_{5a}-ACP_{5b} is that the 159-residue linker C-terminal to the KS–AT and post-AT linkers (Fig. 11a) must also adopt a compact structure, as no region of the atomic distribution could be correlated with a long, unstructured peptide. Interestingly, this sequence has no homologues in the protein database, and no mutual sequence homology is discernible between comparable KS-ACP linkers in 35 other *trans*-AT PKS modules. However, all of these regions exhibit an amino acid bias characteristic of intrinsically unstructured polypeptides, which often adopt a defined fold only in the presence of other proteins and are implicated in molecular recognition processes with multiple partners.[89] This observation suggests that these regions may play structural roles in the modules and/or be involved in facilitating interactions with enzymes acting *in trans* − but these hypotheses remain to be validated.

Module 5 terminates in ~25 residues (DD, Fig. 11a) which show homology to a proposed dimerization element from the CurA subunit of the curacin *cis*-AT PKS.[90] Surprisingly, SAXS analysis of the whole module, again at an estimated 20 Å resolution, revealed no interaction between the putative dimerization domains on the two polypeptides, and consequently, that the module retained its open form (Fig. 11f). Indeed, closer sequence analysis showed that the DD region has low, but convincing homology to 'class 2' C-terminal docking domains present in the *cis*-AT PKS of cyanobacteria and myxobacteria.[55,91] Correspondingly, inspection of the N-terminus of the downstream subunit VirFG (Fig. 2b) revealed its similarity to the partner class 2 N-terminal docking domains. Resolution of the crystal structure of a covalent complex of such domains[55] has shown that, unlike the C-terminal docking domains found in the DEBS and Pik PKSs (class 1), this type of docking element does not contain a dimerization motif. Thus, the open form of VirA module 5 likely reflects its native state, and is proposed to allow one or both of the carrier protein domains to interact with their numerous *in trans* catalytic partners. Nonetheless, in view of the compact nature of the class 2 docking domain complex,[55] the module may need to clamp down transiently to permit chain transfer to the homodimeric KS of module 6. This consideration coupled with the results obtained on the KS-AT linker and the ACP_{5a}-ACP_{5b} didomain, strongly suggest that *trans*-AT PKS, like their *cis*-AT counterparts, are highly dynamic systems.

Unresolved questions

Although the studies discussed here significantly advance our understanding of PKS architecture, they raise a large number of questions for future work. Based on the localization of the ACP domains in the PikAIII structure, Skiniotis and colleagues proposed that the ACP/partner interactions are largely driven by the chemical nature of the intermediate attached to the carrier protein domain.[18,19] On the other hand, in view of the 19 Å spanned by the Ppant arm, the measured distances between the ACP Ser to which the prosthetic group is attached and the various catalytic domains (with the sole exception of the KR), are inconsistent with insertion of the attached substrates into the active sites (i.e. distance pentaketide-ACP_4/KS_5: 28 Å; ACP_5/AT_5 in the presence of pentaketide-KS_5: 35 Å; methymalonyl-ACP_5/KS_5: 25 Å;) (Fig. 8). Thus, it will be important to confirm with higher resolution data that the observed placement of the ACP does in fact result from substrate binding into the various active sites. A related, though technically-challenging task, is to identify the minimum structural features required to achieve appropriate positioning of the polyketide-ACPs, as this will clearly have consequences for the ability of the systems to process substrate

analogues. In addition, our picture of ACP-based substrate delivery will be incomplete without high-resolution information on the linker regions, allowing us to understand their contribution to directed domain movements as well as modular reorganization. Correspondingly, it will be critical to study the architectures of *trans*-AT PKS modules in the presence of native polyketide chains and analogues to determine whether such substrate-driven effects are also operational.

Both the cryo-EM and SAXS studies investigated the structures of *cis*-AT PKS modules with the same complement of catalytic domains (KS, AT and KR). Thus, in future, it will be vital to extend these approaches to modules comprising additional functional activities (e.g DH and ER) (Fig. 2a). The effects on the open reaction chamber of a dimeric DH[43,92] or the recently-identified post-AT 'dimerization element'[25] are particularly anticipated. Likewise, *trans*-AT PKS modules are notable for their compositional heterogeneity, with unusual domain orderings, duplicated domains, and non-classical catalytic functions (Fig. 2b). It will thus be interesting to determine how this varying domain content is accommodated within the modular architecture. In addition, modules in *trans*-AT systems depend on interactions *in trans* with multiple discrete enzymes (minimally a PPTase and *trans*-AT) to complete their catalytic cycles. Structural studies could reveal whether this suite of modifications occurs within the context of multi-protein complexes which form on the modules, or by successive, transient visits by the *trans*-acting partners.[21]

Further outstanding questions concern communication between successive modules in the PKSs. Dutta and colleagues were able to visualize pentaketide-ACP_4 docked against KS_5,[18] capturing the interaction between modules of *cis*-AT PKS which communicate via class 1[48,49] docking domains (State 1−2, Fig. 8). Nonetheless, several details remain to be established, including the fate of the dimerization element downstream of the ACPs. To affect chain transfer, the ACPs dock against the side entrances to the KS active sites, on opposite sides of the KS homodimer (State 6, Fig. 8). Given the wide separation between the ACPs, this positioning would seem to necessitate unraveling of the C-terminal dimerization motif. Indeed, the observed distance between the ACP domains in two of the conformers of β-hydroxyhexaketide-PikAIII (State 5, Fig. 8), is also consistent with the absence of downstream dimerization. In these cases, the homodimeric KS of the next module might serve as a replacement dimerization element. Confirmation of these hypotheses, however, awaits structure elucidation of a docked complex at a high enough resolution to reveal the positions of the dimerization α-helices. In addition, to fully comprehend intersubunit interactions in *cis*-AT PKS, we will also have to study interfaces mediated by class 2 docking domains.[55,91]

In the case of *trans*-AT PKS, the results of Davison, *et al.*[21] suggest that communication between at least some subunits occurs via class 2-like docking domains. Nonetheless, even a superficial analysis of interfaces in *trans*-AT PKS reveals that many lack obvious sequences at the extreme C- and N-termini of the participating subunits that could function as docking domains (KJ Weissman, unpublished observations). This is particularly obvious for modules in which the component domains are in fact split between two different polypeptides, as occurs in the majority of *trans*-AT systems characterized to date.[4] How functional modules are reconstituted in these cases is entirely unclear.

Similarly, the situation for modules which interact *in cis* within the context of a single PKS polypeptide remains largely unresolved. In *cis*-AT PKS, the linkers joining ACP and KS domains of successive modules are relatively short, typically 20 residues[21] (68 Å[93] maximum length). While the post-ACP dimerization α-helices are mobile elements and hence do not appear to constrain the ACP trajectories, direct tethering to a subsequent module via a relatively short peptide would seem to represent a much more significant impediment to free movement. Clearly, this issue can only be resolved by higher resolution analysis of multi-modular proteins. The same question arises for *trans*-AT PKS systems, although in this case, the intermodular linkers are significantly longer (*ca.* 46 residues[21]). In fact, the linkers joining the ACPs of β-modification modules to the downstream KS domains are on average substantially longer still (*ca.* 63 amino acids[21])), which argues that the open architecture observed for VirA module 5 (Fig. 11f) could be accommodated within the context of a PKS multienzyme.

For both *cis*-AT and *trans*-AT systems, it would also be desirable to understand how PKS modules engage in chain transfer with NRPS modules, which may form part of the same polypeptide (Fig. 2b), or be located on an adjacent subunit. A related question is the oligomeric state of such NRPS modules, as although docking domains upstream of an NRPS subunit in a mixed system were found to be homodimeric,[94] the only solved structure of an NRPS module shows it to be monomeric.[95] A final issue that might be settled by detailed structural studies is the basis for iterative operation of certain modules within the context of an otherwise modular assembly line, as observed in the biosynthesis of borrelidin, aureothin, and DKxanthene, among other PKS.[96–98] All such modules are stand-alone, but it remains unclear whether iteration is due to inefficient docking with the up- and/or downstream subunits, some inherent property of the modules, or a combination of both factors.

Modular PKS architecture: lessons for genetic engineering

It is hoped that insights obtained into PKS modular architecture will significantly boost efforts by synthetic biologists to redesign these systems. What emerges from these three initial studies is a complex picture of PKS operation, which nonetheless suggests several ways forward. Most importantly, the cryo-EM analysis of PikAIII has revealed that the module adopts at least seven different functional states in response to the structure of the chain tethered to the ACP, modification of the KS active site, and a set of distinct, evolving interfaces between the functional domains which allow this chemical information to be transmitted throughout the modular structure, configuring the system for the next step in the biosynthesis (Fig. 8). Thus, to optimally engineer a PKS module by, for example, exchange of a catalytic domain, it will be necessary to have detailed information on both its substrate specificity and the interfaces it will form with neighboring domains at each stage in the chain extension cycle, as well as with the ACP during the catalytic act.

In the PikAIII system, the non-domain regions also play critical roles, relaying conformational messages between the domains (KS-AT, AT-KR linkers), permitting the relatively free movement of the ACP between the various active sites (KR-ACP linker), or contributing to subunit stability through homodimerization of the module C-terminus (post-ACP α-helices). Likewise, linkers in VirA module 5 serve essential functions as either structural/scaffolding domains, or highly flexible tethers. In neither case do we understand the precise characteristics of these regions that permit these varied roles, a situation which is complicated by a lack of strict sequence conservation. Clearly, detailed structure-function relationships must be established for linkers sourced from a panel of PKS. Similarly, we are starting to understand more clearly how the non-catalytic docking domains of both *cis*- and *trans*-AT PKS contribute to intersubunit communication,[48,49,55] in combination with specific contacts between the interacting ACP and KS domains.[18] On the

other hand, at least in *cis*-AT PKS, productive intermodular chain transfer clearly depends on the substrate,[18] and so KS specificity issues will also need to be considered in attempts to manipulate interpolypeptide interfaces.

Perhaps the most promising way forward would be to carry out detailed comparative studies by cryo-EM and/or SAXS on a set of closely-similar PKS modules from a single PKS or, in the case of the *cis*-AT PKS, several evolutionarily-related systems (for example, the 12-, 14-and 16-membered macrolide synthases). In this way, it might be possible to discern how evolution has accommodated specific structural differences (domain exchanges to alter substrate specificity/stereochemistry, module and subunit insertions, etc.), so that such changes might be efficiently recapitulated in the laboratory. In parallel to this rational approach, a library strategy involving high-throughput generation and screening of chimaeric synthases could also yield important PKS sequence/activity information.[15]

Conclusions

Since the discovery of modular PKS multienzymes in the 1990s, a major goal has been to uncover the three-dimensional structures of these extraordinarily complex catalysts. Although high-resolution information on whole modules is still lacking, it may yet prove possible to solve the crystal structure of a whole module, perhaps by trapping it in a specific conformational state using a domain cross-linking approach.[99] Even in this case, however, it is clear that a single crystallographic snapshot will not complete the mechanistic picture. In the meantime, the three studies highlighted in this review demonstrate the strong complementarity of single-particle cryo-EM and SAXS for studying these systems, and in particular, for directly revealing the dynamical changes that occur during the chain extension cycles. The next chapter in the PKS structural biology story is eagerly anticipated.

Acknowledgements

Work in the author's laboratory is supported by the Agence Nationale de la Recherche (ANR JCJC 2011 PKS-PPIs), the Centre National de la Recherche Scientifique (CNRS), the University of Lorraine, and the Lorraine Region (BQR grants). A Gruez and B Chagot are thanked for helpful discussions and comments on the manuscript.

Notes and references

[a]Molecular and Structural Enzymology Group, Université de Lorraine, IMoPA, UMR 7365, Vandœuvre-Les-Nancy, F-54500, France. CNRS, IMoPA, UMR 7365, Vandœuvre-Les-Nancy, F-54500, France. E-mail: kira.weissman@univ-lorraine.fr.

1 G. T. Carter, *Nat. Prod. Rep.*, 2011, **28**, 1783–1789.
2 C. Hertweck, *Angew. Chem. Int. Ed Engl.*, 2009, **48**, 4688–4716.
3 J. Staunton and K. J. Weissman, *Nat. Prod. Rep.*, 2001, **18**, 380–416.
4 J. Piel, *Nat. Prod. Rep.*, 2010, **27**, 996–1047.
5 J. Cortés, S. F. Haydock, G. A. Roberts, D. J. Bevitt and P. F. Leadlay, *Nature*, 1990, **348**, 176–178.
6 Y.-Q. Cheng, G.-L. Tang and B. Shen, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 3149–3154.
7 L. Gu, T. W. Geders, B. Wang, W. H. Gerwick, K. Håkansson, J. L. Smith and D. H. Sherman, *Science*, 2007, **318**, 970–974.
8 P. Pöplau, S. Frank, B. I. Morinaka and J. Piel, *Angew. Chem. Int. Ed Engl.*, 2013, **52,** 13215–13218.
9 T. Bretschneider, J. B. Heim, D. Heine, R. Winkler, B. Busch, B. Kusebauch, T. Stehle, G. Zocher and C. Hertweck, *Nature*, 2013, **502**, 124–128.
10 C. T. Calderone, *Nat. Prod. Rep.*, 2008, **25**, 845–853.
11 A. C. Jones, E. A. Monroe, E. B. Eisman, L. Gerwick, D. H. Sherman and W. H. Gerwick, *Nat. Prod. Rep.*, 2010, **27**, 1048–1065.
12 H. Jenke-Kodama, A. Sandmann, R. Müller and E. Dittmann, *Mol. Biol. Evol.*, 2005, **22**, 2027–2039.
13 T. Nguyen, K. Ishida, H. Jenke-Kodama, E. Dittmann, C. Gurgui, T. Hochmuth, S. Taudien, M. Platzer, C. Hertweck and J. Piel. *Nat. Biotechnol.*, 2008, **26**, 225–233.
14 K. J. Weissman and P. F. Leadlay, *Nat. Rev. Microbiol.*, 2005, **3**, 925–936.
15 S. Poust, A. Hagen, L. Katz and J. D. Keasling, *Curr. Opin. Biotechnol.* 2014, **30**, 32–39.
16 M. Till and P. R. Race, *Biotechnol. Lett.*, 2014, **36**, 877–888.
17 W. Xu, K. Qiao and Y. Tang, *Crit. Rev. Biochem. Mol. Biol.*, 2013, **48**, 98–122.
18 S. Dutta, J. R. Whicher, D. A. Hansen, W. A. Hale, J. A. Chemler, G. R. Congdon, A. R. H. Narayan, K. Håkansson, D. H. Sherman, J. L. Smith and G. Skiniotis, *Nature*, 2014, **510**, 512–517.
19 J. R. Whicher, S. Dutta, D. A. Hansen, W. A. Hale, J. A. Chemler, A. M. Dosey, A. R. H. Narayan, K. Håkansson, D. H. Sherman, J. L. Smith and G. Skiniotis, *Nature*, 2014, **510**, 560–564 (2014).
20 A. L. Edwards, T. Matsui, T. M. Weiss and C. Khosla, *J. Mol. Biol.*, 2014, **426**, 2229–2245.
21 J. Davison, J. Dorival, H. Rabeharindranto, H. Mazon, B. Chagot, A. Gruez and K. J. Weissman, *Chem. Sci.*, 2014, **5**, 3081–3095.
22 J. Staunton, P. Caffrey, J. F. Aparicio, G. A. Roberts, S. S. Bethell and P. F. Leadlay, *Nat. Struct. Biol.*, 1996, **3**, 188–192.
23 T. Maier, M. Leibundgut and N. Ban, *Science*, 2008, **321**, 1315–1322.
24 J. Zheng, D. C. Gay, B. Demeler, M. A. White and A. T. Keatinge-Clay, *Nat. Chem. Biol.*, 2012, **8**, 615–621.
25 J. Zheng, C. D. Fage, B. Demeler, D. W. Hoffman and A. T. Keatinge-Clay, *ACS Chem. Biol.*, 2013, **8**, 1263–1270.
26 E. F. Garman, *Science*, 2014, **343**, 1102–1108.
27 X. Li, P. Mooney, S. Zheng, C. R. Booth, M. B. Braunfeld, S. Gubbens, D. A. Agard and Y. Cheng, *Nat. Methods*, 2013, **10**, 584–590.
28 C. E. Blanchet and D. I. Svergun, *Ann. Rev. Phys. Chem.*, 2013, **64**, 37–54.
29 J. L. S. Milne, M. J. Borgnia, A. Bartesaghi, E. E. Tran, L. A. Early, D. M. Schauder, J. Lengyel, J. Pierson, A. Patwardhan and S. Subramaniam, *FEBS J.*, 2013, **280**, 28–45.
30 H. D. T. Mertens and D. I. Svergun, *J. Struct. Biol.*, 2010, **172**, 128–141.
31 G. David and J. Pérez, *J. Appl. Crystallogr.*, 2009, **42**, 892–900.
32 D. A. Jacques and J. Trewhella, *Protein Sci.*, 2010, **19**, 642–657.
33 S. Smith and S. C. Tsai, *Nat. Prod. Rep.*, 2007, **24**, 1041–1072.
34 T. Maier, S. Jenni and N. Ban, Science, 2006, **311**, 1258–1262.
35 E. J. Brignole, S. Smith and F. J. Asturias, *Nat. Struct. Mol. Biol.*, 2009, **16**, 190–197.
36 A. T. Keatinge-Clay, *Nat. Prod. Rep.*, 2012, **29**, 1050–1073.
37 A. T. Keatinge-Clay and R. M. Stroud, *Structure*, 2006, **14,** 737–748.
38 Y. Tang, C.-Y. Kim, I. I. Mathews, D. E. Cane and C. Khosla, *Proc. Natl. Acad. Sci. U. S. A.* (2006) **103**, 11124–11129.
39 Y. Tang, A. Y. Chen, C.-Y. Kim, D. E. Cane and C. Khosla, *Chem. Biol.*, 2007, **14**, 931–943.

40 V. Y. Alekseyev, C. W. Liu, D. E. Cane, J. D. Puglisi and C. Khosla, *Protein Sci*., 2007, **16**, 2093–2107.
41 B. Chakravarty, Z. Gu, S. S. Chirala, S. J. Wakil and F. Quiocho, *Proc. Natl. Acad. Sci. U. S. A*., 2004, **101**, 15567–15572.
42 S. C. Tsai, L. J. Miercke, J. Krucinski, R. Gokhale, J. C. Chen, P. G. Foster, D. E. Cane, C. Khosla and R. M. Stroud, *Proc. Natl. Acad. Sci. U. S. A*., 2001, **98**, 14808–14813 (2001).
43 A. Keatinge-Clay, *J. Mol. Biol.*, 2008, **384**, 941–953.
44 K. J. Weissman, *ChemBioChem*, 2008, **9**, 2929–2931.
45 F. T. Wong, X. Jin, I. I. Mathews, D. E. Cane and C. Khosla, *Biochemistry*, 2011, **50**, 6539–6548.
46 D. C . Gay, G. Gay, A. J. Axelrod, M. Jenner, C. Kohlhass, A. Kampa, N. J. Oldham, J. Piel and A. T. Keatinge-Clay, *Structure*, 2014, **22**, 444–451 (2014).
47 A. S. Haines, X. Dong, Z. Song, R. Farmer, C. Williams, J. Hothersall, E. Płoskoń, P. Wattana-amorn, E. R. Stephens, E. Yamada, R. Gurney, Y. Takebayashi, J. Masschelein, R. J. Cox, R. Lavigne, C. L. Willis, T. J. Simpson, J. Crosby, P. J. Winn, C. M. Thomas and M. P. Crump. (2013) *Nat. Chem. Biol.* **9**, 685–692.
48 R. W. Broadhurst, D. Nietlispach, M. P. Wheatcroft, P. F. Leadlay and K. J. Weissman, *Chem. Biol.*, 2003, **10**, 723–731.
49 T. J. Buchholz, T. W. Geders, F. E. Bartley 3rd, K. A. Reynolds, J. L. Smith and D. H. Sherman, *ACS Chem. Biol.*, 2009, **4**, 41–52.
50 R. H. Lambalot, A. M. Gehring, R. S. Flugel, P. Zuber, M. LaCelle, M. A. Marahiel, R. Reid, C. Khosla and C. T. Walsh, *Chem. Biol.* 1996, **3**, 923–936.
51 A. Y. Chen, N. A. Schnarr, C. Y. Kim, D. E. Cane and C. Khosla, *J. Am. Chem. Soc.*, 2006, **128**, 3067–3074.
52 L. K. Charkoudian, C. W. Liu, S. Capone, S. Kapur, D. E. Cane, A. Togni, D. Seebach and C. Khosla, *Protein Sci.*, 2011, **20**, 1244–1255.
53 S. Kapur, B. Lowry, S. Yuzawa, S. Kenthirapalan, A. Y. Chen, D. E. Cane and C. Khosla, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**, 4110–4115 (2012).
54 S. Kapur, A. Y. Chen, D. E. Cane and C. Khosla, *Proc. Natl. Acad. Sci. U. S. A*., 2010, **107**, 22066–22071.
55 J. R. Whicher, S. S. Smaga, D. A. Hansen, W. C. Brown, W. H. Gerwick, D. H. Sherman and J. L. Smith, *Chem. Biol.*, 2013, 20, 1340–1351.
56 C. Sánchez, L. Du, D. J. Edwards, M. D. Toney and B. Shen, *Chem. Biol.*, 2001 **8**, 725–738.
57 R. S. Gokhale, D. Hunziker, D. E. Cane and C. Khosla, *Chem. Biol.*, 1999, **6**, 117–125.
58 I. M. Nooren and J. M. Thornton, *J. Mol. Biol.*, 2003, **325**, 991–1018.
59 S. E. A. Ozbabacan, H. B. Engin, A. Gursoy and O. Keskin, *Protein Eng. Des. Sel.*, 2011, **24**, 635–648.
60 W. Huang, J. Jia, P. Edwards, K. Dehesh, G. Schneider and Y. Lindqvist, *EMBO J.*, 1998, **17**, 1183–1191.
61 J. A. Chemler, T. J. Buchholz, T. W. Geders, D. L. Akey, C. M. Rath, G. E. Chlipala, J. L. Smith and D. H. Sherman, *J. Am. Chem. Soc.*, 2012, **134**, 7359–7366.
62 A. T. Keatinge-Clay, D. A. Maltby, K. F. Medzihradszky, C. Khosla and R. M. Stroud, *Nat. Struct. Mol. Biol.*, 2004, **11**, 888–893.
63 J. L. Ferrer, J. M. Jez, M. E. Bowman, R. A. Dixon and J. P. Noel, *Nat. Struct. Biol.*, 1999, **6**, 775–784.
64 I. Thomas, C. J. Martin, C. J. Wilkinson, J. Staunton and P. F. Leadlay, *Chem. Biol.*, 2002, **9**, 781–787.
65 D. A. Hansen, C. M. Rath, E. B. Eisman, A. R. Narayan, J. D. Kittendorf, J. D. Mortison, Y. J. Yoon and D. H. Sherman, *J. Am. Chem. Soc.*, 2013, **135**, 11232–11238.
66 J. Grünewald and M. A. Marahiel, *Microbiol. Mol. Biol. Rev*., 2006, **70**, 121–146.
67 J. Grünewald, S. A. Sieber and M. A. Marahiel, *Biochemistry*, 2004, **43**, 2915–2925.
68 S. A. Sieber, J. Tao, C. T. Walsh and M. A. Marahiel, *Angew. Chem. Int. Ed Engl*., 2004, **43**, 493–498.
69 F. Kopp, C. Mahlert, J. Grünewald and M. A. Marahiel, *J. Am. Chem. Soc.*, 2006, **128**, 16478–16479.
70 B. J. Dunn, D. E. Cane and C. Khosla, *Biochemistry*, 2013, **52**, 1839–1841.
71 H. Hong, P. F. Leadlay and J. Staunton, *FEBS J.*, 2009, **276**, 7057–7069.
72 S. A. Bonnett, C. M. Rath, A.-R. Shareef, J. R. Joels, J. A. Chemler, K. Håkansson, K. Reynolds and D. H. Sherman, *Chem. Biol.*, 2011, **18**, 1075–1081.
73 P. Caffrey, *ChemBioChem*, 2003, **4**, 654–657.
74 A. T. Keatinge-Clay, *Chem. Biol.*, 2007, **14**, 898–908.
75 S. A. Bonnett, J. R. Whicher, K. Papireddy, G. Florova, J. L. Smith and K. A. Reynoldds, *Chem. Biol.*, 2013, **20**, 772–783.
76 A. M. Hill and J. Staunton, in *Comprehensive Natural Products II: Chemistry and Biology*, Elsevier, 2010.
77 K. J. Weissman and R. Müller, *ChemBioChem*, 2008, **9**, 826–848.
78 C. Anselmi, M. Grininger, P. Gipson and J. D. Faraldo-Gómez, *J. Am. Chem. Soc.*, 2010, **132**, 12357–12364.
79 L. Tran, R. W. Broadhurst, M. Tosin, A. Cavalli and K. J. Weissman, *Chem. Biol.*, 2010, **17**, 705–716.
80 R. Aggarwal, P. Caffrey, P. F. Leadlay, C. Smith and J. Staunton, *J. Chem. Soc.Chem. Commun.*, 1995, 1519–1520.
81 J. Cortés, K. E. Wiesmann, G. A. Roberts, M. J. Brown, J. Staunton and P. F. Leadlay, *Science*, 1995, **268**, 1487–1489.
82 C. M. Kao, G. L. Luo, L. Katz, D. E. Cane and C. Khosla, *J. Am. Chem. Soc.*, 1996, **118**, 9184–9185.
83 C. M. Kao, G. Luo, L. Katz, D. E. Cane and C. Khosla, *J. Am. Chem. Soc.*, 1995, **117**, 9105–9106.
84 N. Pulsawat, S. Kitani and T. Nihira, *Gene*, 2007, **393**, 31–42.
85 M. L. Heathcote, J. Staunton and P. F. Leadlay, *Chem. Biol.*, 2001, **8**, 207–220.
86 A. S. Rahman, J. Hothersall, J. Crosby, T. J. Simpson and C. M. Thomas, *J. Biol. Chem.*, 2005, **280**, 6399–6408.
87 M. Czjzek, H.-P. Fierobe and V. Receveur-Bréchot, *Methods Enzymol.*, 2012, **510**, 183–210.
88 U. Trujillo, E. Vázquez-Rosa, D. Oyola-Robles, L. J. Stagg, D. A. Vassallo, I. E. Vega, S. T. Arold and A. Baerga-Ortiz, *PloS One*, 2013, **8**, e57859.
89 P. E. Wright and H. J. Dyson, *Curr. Opin. Struct. Biol.*, 2009, **19**, 31–38.
90 L. Gu, E. B. Eisman, S. Dutta, T. M. Franzmann, S. Walter, W. H. Gerwick, G. Skiniotis and D. H. Sherman, *Angew. Chem. Int. Ed Engl.*, 2011, **50**, 2795–2798 (2011).
91 M. Thattai, Y. Burak and B. I. Shraiman, *PLoS Comput. Biol.*, 2007, **3**, 1827–1835.
92 D. Akey, J. R. Razelun, J. Tehranisa, D. H. Sherman, W. H. Gerwick and J. L. Smith, *Structure*, 2010, **18**, 94–105.
93 G. Yang, C. Cecconi, W. A. Baase, I. R. Vetter, W. A. Breyer, J. A. Haack, B. W. Matthews, F. W. Dahlquist and C. Bustamante, *Proc. Natl. Acad. Sci.*, 2000, **97**, 139–144 (2000).
94 C. D. Richter, D. Nietlispach, R. W. Broadhurst and K. J. Weissman, *Nat. Chem. Biol.*, 2008, **4**, 75–81.
95 A. Tanovic, S. A. Samel, L.-O. Essen and M. A. Marahiel, *Science*, 2008, **321**, 659–663.
96 J. He and C. Hertweck, *ChemBioChem*, 2005, **6**, 908–912.
97 P. Meiser, K. J. Weissman, H. B. Bode, D. Krug, J. S. Dickschat, A. Sandmann and R. Müller, *Chem. Biol.*, 2008, **15**, 771–781.

98 C. Olano, B. Wilkinson, S. J. Moss, A. F. Braña, C. Méndez, P. F. Leadlay and J. A. Salas, *Chem. Commun.*, 2003, 2780–2782.
99 C. Nguyen, R. W. Haushalter, D. J. Lee, P. R. Markwick, J. Bruegger, G. Caldara-Festin, K. Finzel, D. R. Jackson, F. Ishikawa, B. O'Dowd, J. A. McCammon, S. J. Opella, S. C. Tsai and M. D. Burkart, *Nature*, 2014, **505**, 427–431.
100 T. P. Stinear, A. Mve-Obiang, P. L. Small, W. Frigui, M. J. Pryor, R. Brosch, G. A. Jenkin, P. D. Johnson, J. K. Davies, R. E. Lee, S. Adusumilli, T. Garnier, S. F. Haydock, P. F. Leadlay and S. T. Cole, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 1345–1349.