Natural Product Reports

ROYAL SOCIETY OF CHEMISTRY

REVIEW

View Article Online
View Journal

Cite this: DOI: 10.1039/d6np00010j

Decision making by modular polyketide synthases and implications for genetic engineering

Kira J. Weissman

Covering: up to 2025

Modular polyketide synthases (PKSs) are often described as molecular-scale assembly lines – a moniker that implies a high-degree of coordinated and regulated activity. Belying this characterization, all of these systems must navigate a number of critical biosynthetic 'choices' in order to faithfully produce one or a limited number of products. Such alternative options occur at all stages of the assembly process, including building block selection, chain extension, intersubunit transfer, modification by in-*trans* acting enzymes, and chain termination. Here we discuss research over many decades that has shed substantial light on the detailed and often stunningly complex molecular mechanisms that underpin PKS biosynthetic fidelity. These insights have inspired increasingly effective efforts to intervene at these decision points by genetic engineering to redirect the pathways towards alternative outcomes and their associated novel products.

Received 30th January 2026

DOI: 10.1039/d6np00010j

rsc.li/npr

Université de Lorraine, CNRS, IMoPA, F-54000 Nancy, France. E-mail: kira.weissman@univ-lorraine.fr

1. Introduction

The reduced or complex polyketide natural products frequently inspire synthetic chemists to attempt their total synthesis, often resulting in multi-step protocols that require dozens of chemical reagents (Fig. 1).[1] Yet bacteria manage to build these highly complex molecules from a limited pool of simple precursors with robust catalytic, energy and atom efficiency.[2] Unlike laboratory-based approaches in which elaborate sub-structures are synthesized and then homologated together, nature constructs polyketides in linear fashion, block by block (Fig. 1).[3,4] While the basic algorithm is simple, structural diversity of the resulting chains arises from the meticulous tailoring of each monomer at the point of entry into the process, among other factors.

In this context, it's worth thinking about the enzymology that Nature might have used to program this series of modifications. One possibility would have been to co-opt the enzymes of fatty acid biosynthesis in bacteria to build polyketides, as the underlying chemistry is the same (Fig. 2).[5] To assemble fatty acids, an acetyl-CoA starter unit is condensed with a malonate extender unit attached to a non-catalytic acyl carrier protein

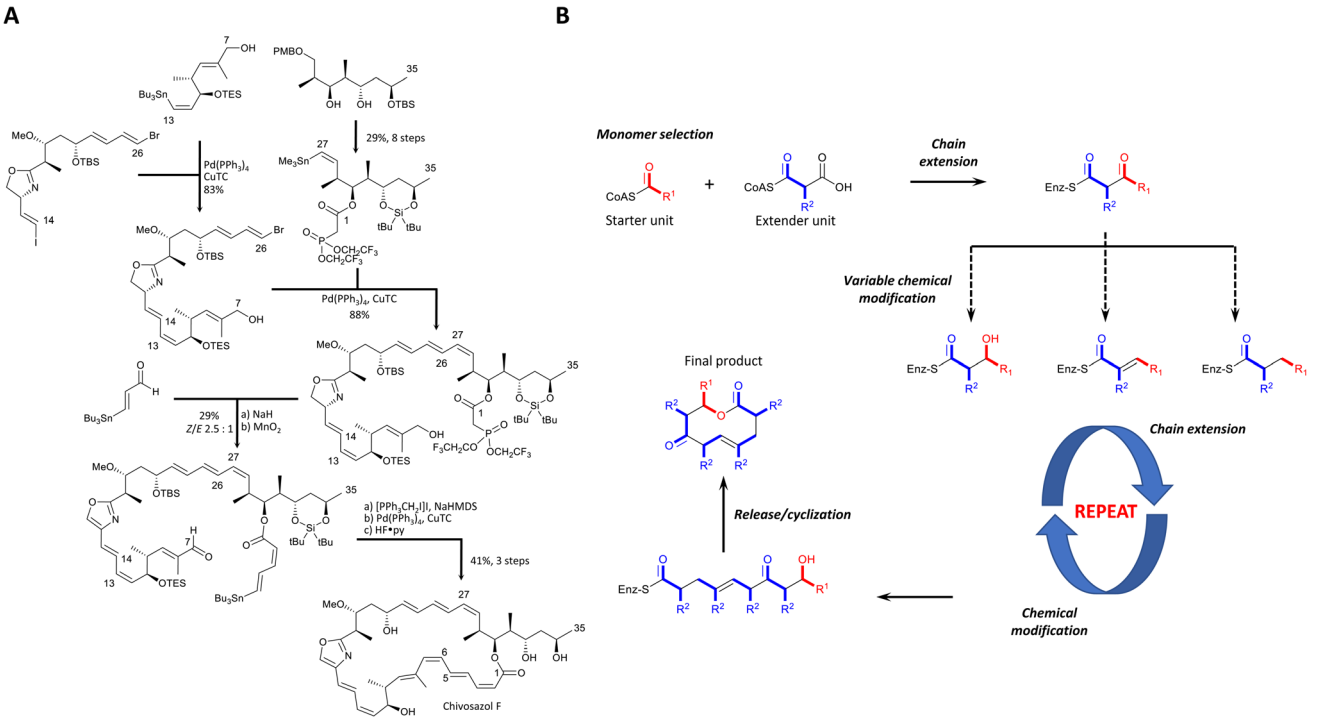

Fig. 1 Comparison between (A) the chemical synthesis of complex polyketides, and (B) their biosynthesis in bacteria by polyketide synthases (PKSs). (A) Multi-step, multi-reagent synthetic route to the myxobacterial polyketide chivosazol F.[1] This synthesis involved assembly of several complex fragments which were then combined. (B) Complex polyketides are constructed in linear fashion, by the addition of simple building blocks to a growing chain. Each of the monomers may undergo chemical modification before the cycle repeats a defined number of times. The process terminates with release of the polyketide from the PKS, which often results in macrocyclization.

(ACP), most typically by a ketosynthase (KAS III) homologue (FabH).[6] The resulting β-keto chain is then acted upon successively by ketoreductase (KR; FabG), dehydratase (DH; *e.g.* FabZ) and enoyl reductase (ER; *e.g.* FabI) enzymes, yielding respectively β-hydroxyl, olefin and fully reduced intermediates (Fig. 2). This cycle is then iterated (with KSs FabB and FabF taking over for FabH) until a chain of the appropriate length is reached (typically C_{16}). Although it would have been in principle possible to redirect this pathway to making polyketides, this would have required programmed short-cutting of the processing reactions to produce the diverse functionality characteristic of these molecules (Fig. 2), among other modifications. These observations may explain why Nature did not elect to use this mode of biosynthesis to make complex polyketides, but a collection of discrete enzymes is deployed to generate so-called type II polyketides[7,8] which classically exhibit polyaromatic or linear polyene[9] structures (*i.e.* the underlying programming of chain building is much simpler).

Instead, Nature uses a division-of-labor approach to build complex polyketides, in which each and every step in the pathways is carried out by a dedicated enzymatic domain (Fig. 3).[3] Taking assembly of the erythromycin A **1** core, 6-deoxyerythronolide B **2** (6-dEB), as an example, it is constructed from one propionyl-CoA starter unit and six of its carboxylated equivalents, (2*S*)-methylmalonyl-CoA, as extender units.[10] The system thus includes 7 acyl transferase (AT) domains capable of selecting these monomers from the cellular pool. Joining them together requires six carbon–carbon bonds – a task carried out by six KS domains (note: the KS and AT domains together are referred to as the 'condensing wing' portion of the systems). The pathway also incorporates precisely the set of enzymes required to set the pendant functional groups on the macrocycle: six KRs

Kira J. Weissman

Kira J. Weissman holds a BS from Stanford University, and MPhil and PhD degrees in chemistry from the University of Cambridge, UK. Her PhD and post-doc work was supported by a number of competitive fellowships, including the Churchill Scholarship, and a Royal Society Dorothy Hodgkin Fellowship. She is currently Professor of Enzymology (PR Ex1) at the Université de Lorraine (UL), Nancy, France, where she carries out research on the mechanistic enzymology, structural biology and synthetic biology of modular polyketide synthases (PKSs) and allied mega-enzyme systems. She is the author of >100 primary publications, reviews, commentaries, book chapters and popular science articles, Associate Editor of JACS Au, *and Editor-in-Chief of the RSC's Chemical Biology Book Series. She is also co-head of the Biomolecules for the Bioeconomy (B4B) program of the UL (currently 27 UL laboratories and 27 partner companies).*

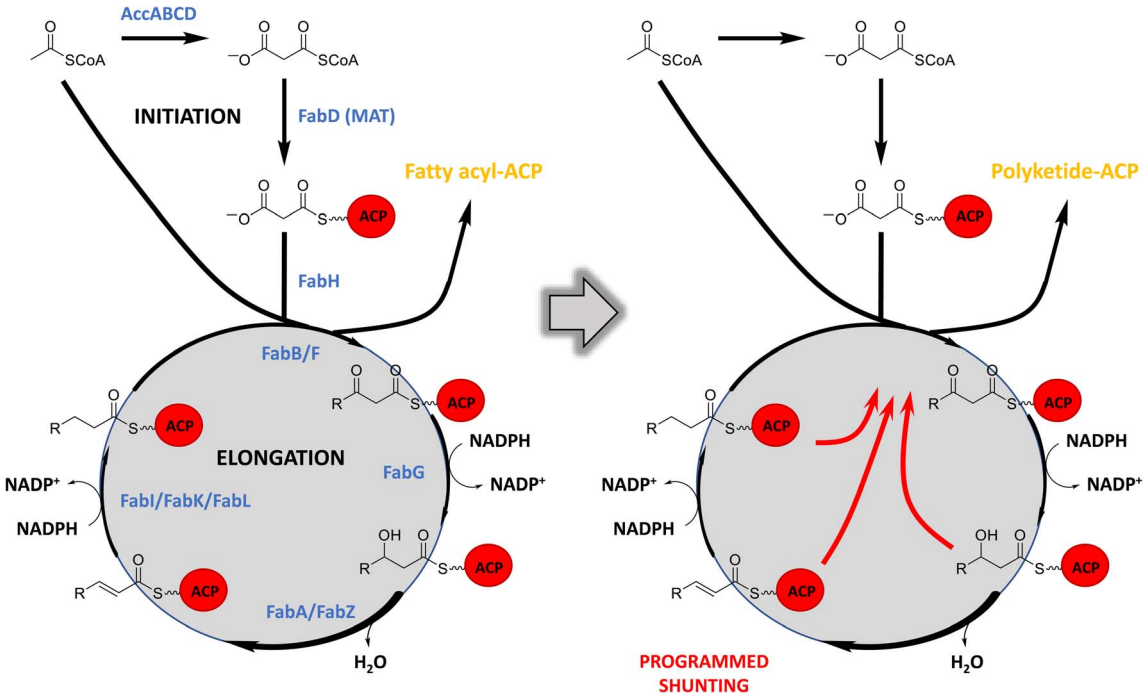

Fig. 2 Scenario for the repurposing of type II fatty acid biosynthesis in bacteria to generate complex polyketides. Fatty acids are assembled by a series of discrete enzymes from acetyl-CoA and malonyl-CoA building blocks.[5,318] The first round of chain extension is typically carried out by the ketosynthase FabH,[6] and subsequent rounds by its homologues FabB and FabF. Each cycle of growth is accompanied by full processing of the β-keto intermediate by ketoreductase FabG, dehydratases FabA and FabZ, and enoyl reductases FabI/FabK/FabL, to yield the saturated chain. During the process, the chains remain tethered to a non-catalytic acyl carrier protein (ACP), with the initial transfer of malonate onto the ACP catalyzed by a malonyl-CoA/acetyl-CoA transacylase (MAT), FabD. In principle, this cycle could have been co-opted to give rise to the diverse functionality observed in polyketides, but this would have required programmed shunting of the process at a specific stage in every cycle, among other adaptations. Key: Acc, acetyl-CoA carboxylase. The squiggly line linking the ACP to the pendant intermediates is the phosphopantetheine (Ppant) cofactor.

(note, one of the KRs (KR^0) does not catalyze ketoreduction, but only epimerization at the α-methyl position[11]), one DH and one ER. Finally, the critical release of the chain as a macrolactone is effected by a thioesterase (TE) domain.[12]

As in fatty acid biosynthesis (Fig. 2), both the substrates and intermediates are attached to ACPs throughout the process *via* their phosphopantetheine (Ppant) prosthetic groups.[13] This tethering allows the chains to be shepherded to all of the catalytic partners, a feature that has obvious kinetic benefits given the obligate, extended reaction series. Sets of domains are organized into functional modules – one to initiate the biosynthesis (typically AT–ACP, but see Section 2.1) and six for chain extension (minimally incorporating KS, AT and ACP domains, and comprising a variable complement of processing activities). The modules are distributed among multiple gigantic polypeptides called subunits (*e.g.* DEBS 1, 2 and 3 (Fig. 3)[14]) that are arranged linearly in the order in which they act, giving rise to a veritable molecular-scale assembly line for constructing a polyketide small molecule, called a modular polyketide synthase (PKS). Another important aspect of these enzymes is their obligate homodimeric character,[15] with the KS, DH[16] and TE[17] domains among other elements contributing to dimerization.[18,19] Two large families of modular PKS have been identified to date – the *cis*-AT class in which the ATs are present within the multienzymes, and the *trans*-AT class, in which this function is supplied by a discrete enzyme that acts iteratively to service all of the PKS modules.[20]

The assembly line biosynthetic logic elegantly solves certain problems inherent to polyketide biosynthesis – which enzyme does what and when – but it creates others. Indeed, if we make the reasonable assumption that the systems evolved to synthetize one or at most a few related structures, then each PKS must negotiate multiple decision points where alternative outcomes are possible. In this review, we discuss how modular PKSs handle these biosynthetic 'choices', and the opportunities that they present for modifying the structures by genetic engineering towards the generation of novel derivatives.

2. The choice of building blocks (starter/extender units)

2.1 Substrate/ACP partner selection by *cis*-AT PKS ATs

The first decision point in polyketide biosynthesis is the choice of monomers from which to build the chains (Fig. 4). It is likely that a given PKS system will encounter a wide range of acyl-CoAs in the cellular environment, including those of primary metabolism and any generated by cluster-specific enzymes.[21,22] Common starter units include acetyl-CoA and propionyl-CoA,

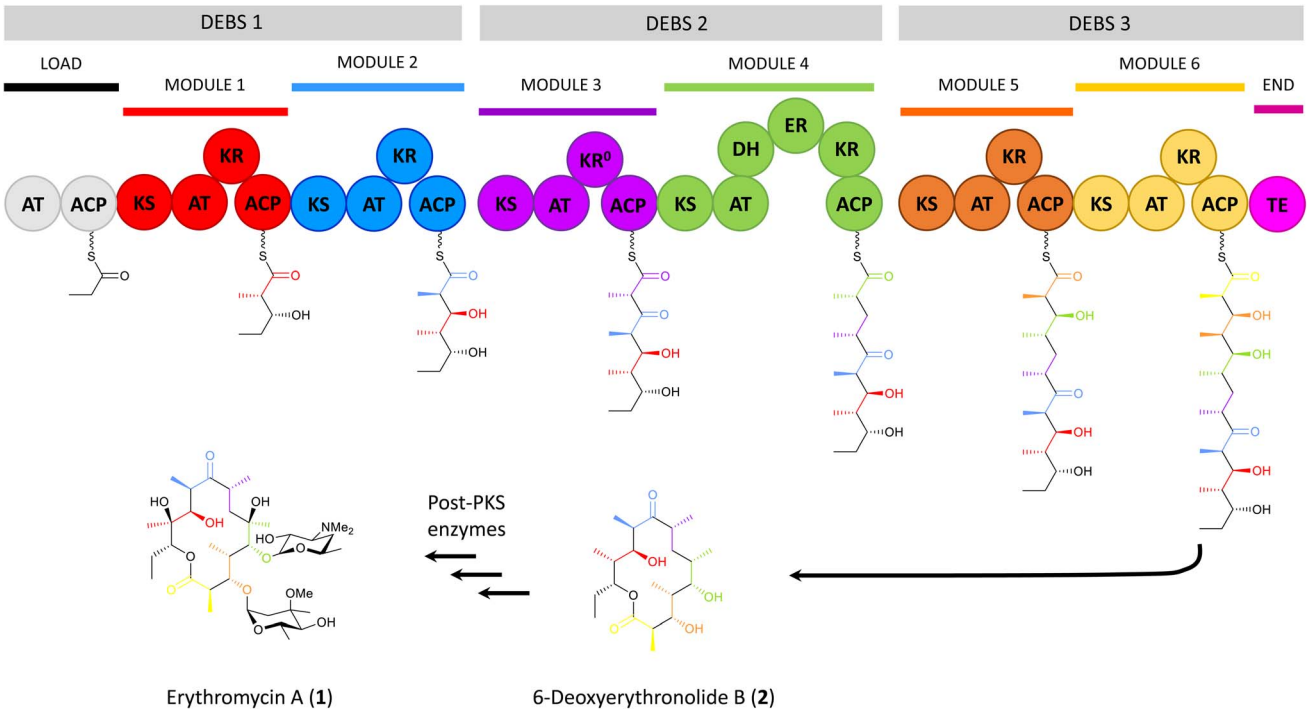

Fig. 3 The erythromycin A (**1**) polyketide synthase (6-deoxyerythronolide B synthase, DEBS). The PKS is composed of three subunits, DEBS 1, 2 and 3. Each subunit comprises multiple modules, of which two are involved in initiating and terminating the assembly process, and the remaining six in chain extension. Each of the chain extension modules minimally includes ketosynthase (KS), acyl transferase (AT) and acyl carrier protein (ACP) domains, and a variable complement of processing domains (ketoreductase (KR), dehydratase (DH) and enoyl reductase (ER)). The assembly line ends in a thioesterase (TE) that liberates the first enzyme-free intermediate 6-deoxyerythronolide (6-dEB) (**2**) as a macrolide. 6-dEB undergoes subsequent transformation by a set of post-PKS enzymes, yielding the final bioactive erythromycin A. Key: KR^0, type C2 ketoreductase that is inactive for ketoreduction, but which catalyzes epimerization at the α-position of the intermediate.[319]

while the carboxylated equivalents, malonyl-CoA and (2*S*)-methylmalonyl-CoA, are the most common extender units.[23] The pool of employed acyl-CoA starter units also includes more exotic species such as the amino acid-derived isobutyryl-CoA and 2-methylbutyryl-CoA (as in the avermectins), as well as guanidinobutyryl-CoA (azalomycin F), for example (for full reviews see: ref. 21, 22 and 24). Additional atypical precursors are assembled on ACP supports by cluster-encoded sets of enzymes. These include:[22] hydroxymalonyl-ACP and aminomalonyl-ACP (zwittermicin A),[25,26] 5-hexynoyl-ACP (jamaicamide)[27] and L-lactyl-ACP (FR901464).[28] Although less diverse than starter moieties,[22] alternative extender units include ethylmalonyl-CoA (*e.g.* monensin), allylmalonyl-CoA (*e.g.* FK506), butylmalonyl-CoA (*e.g.* reveromycin), 3-oxobutylmalonyl-CoA (*e.g.* sanglifherin A), isobutylmalonyl-CoA (*e.g.* ansalactam), hexylmalonyl-CoA/isoheptyl-malonyl-CoA (*e.g.* the stambomycins), and benzyl-malonyl-CoA (*e.g.* splenocin).[23]

As mentioned earlier, the choice of monomer is most typically made by AT domains. In this case, the selection reaction takes place in two steps (*i.e.* it is a ping-pong mechanism):[29–31] the AT first self-acylates its active site serine with the substrate, and then transacylates the building block to the Ppant arm of an ACP acceptor (Fig. 5). The larger, catalytic subdomain of the ATs exhibits the αβ hydrolase fold characteristic of proteases,[32] but which has evolved to transfer the acyl-enzyme intermediates to a downstream ACP domain with which it forms a complex.[26] According to molecular dynamics (MD) simulations, the smaller, appended ferredoxin-like subdomain contributes to determining the overall volume and mobility of the active site,[33] and also helps to orient the ACP for binding to the catalytic portion.[34] The AT catalytic apparatus includes the aforementioned Ser which forms a catalytic dyad with a conserved His residue (a main chain NH may serve as the third member of a triad), as well as an oxyanion hole comprising backbone amide residues.[35] In the context of loading modules, starter-unit ATs can partner with variant KS domains – so-called KS_Qs in which the active site Cys is mutated to Gln – that catalyze decarboxylation of an ACP-tethered extender unit (usually malonate) selected by the AT, to yield a typical starter unit (*e.g.* acetate) (Fig. 5).[36] The same functions can be accomplished by a Gcn5-related *N*-acetyltransferase (GNAT)/ACP pair. In this case, the ACP is acylated by some as yet unidentified enzyme, and then the GNAT catalyzes its decarboxylation to yield acetyl-ACP (Fig. 5).[37]

The majority of ATs that recruit extender units exhibit strict specificity for their carboxyacyl-CoA substrates, even discriminating among monomers that are closely similar in structure such as malonate, methylmalonate and ethylmalonate.[29] This tight fidelity is likely necessary because KS domains (at least those assayed directly) are tolerant towards variation in extender unit structure.[38–40] Nonetheless, promiscuous examples of ATs have been identified,[39,41–43] particularly those that

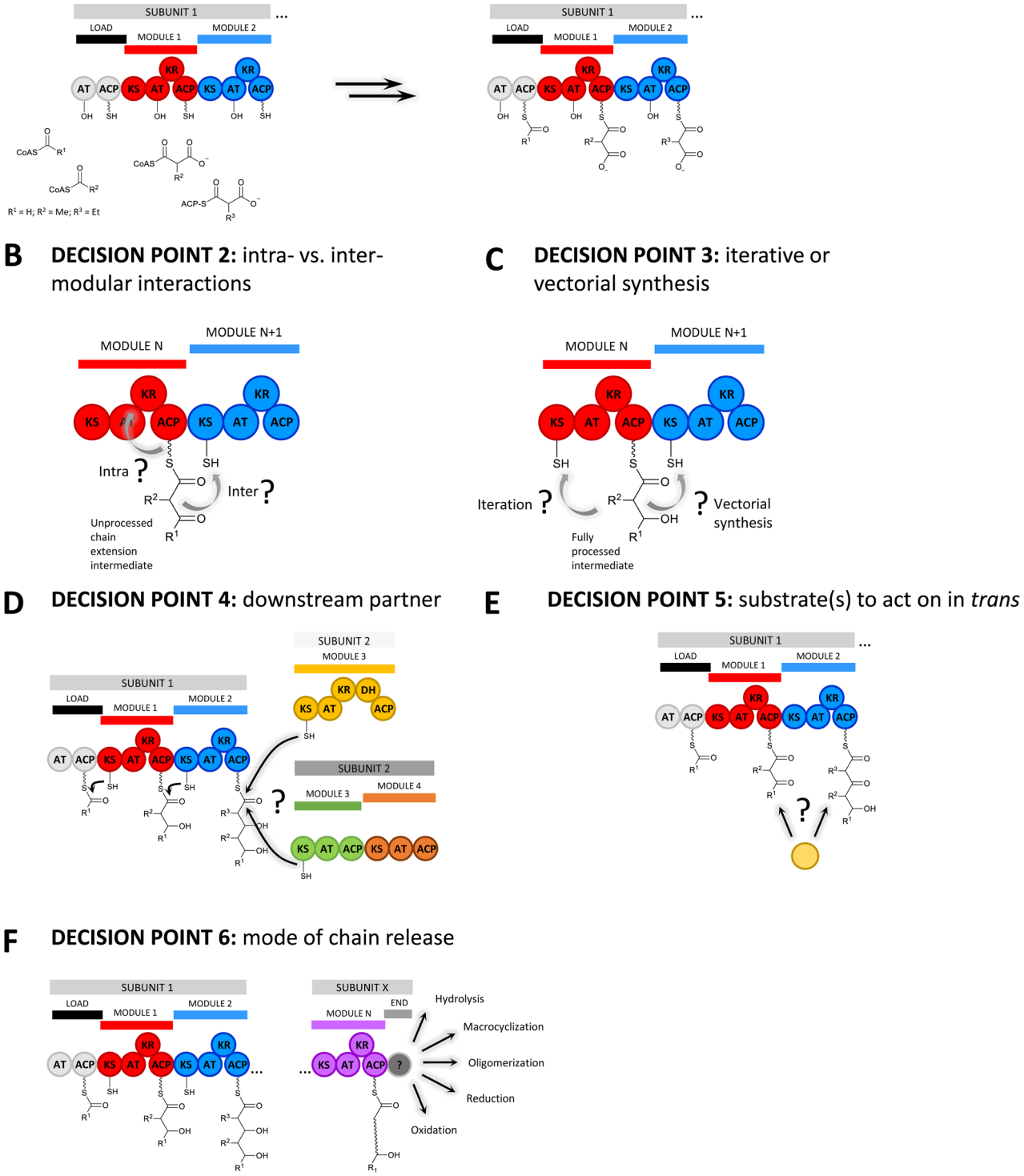

Fig. 4 Key decision points in modular polyketide biosynthesis. (A) The first concerns selection of building blocks by the acyl transferase (AT) domains. In the case of the loading module, typical monomers include acetyl-CoA and malonyl-CoA, while extension modules often choose the corresponding carboxylated building blocks, malonyl-CoA and methylmalonyl-CoA. The overall reaction involves self-acylation by the AT followed by transfer of the monomer to the downstream ACP domain (see Fig. 5). (B) The second decision is whether, following chain extension to generate a β-keto intermediate, the chain is ferried to one or more processing domains (intramodular interactions), or handed off to the N-terminal domain of the downstream module (*e.g.* a KS; intermodular interaction). (C) Having completed a full round of chain extension/processing, any given module must decide whether to act iteratively by back transfer of the fully processed intermediate to the KS domain, or vectorially, by shuttling the chain to the downstream module. In both cases, the KS active site must be empty in order to accept the incoming substrate. (D) Having decided to engage in chain transfer, the system must choose among multiple potential partners (subunit 2s, in this case). (E)

recognize extender units generated by cluster-encoded crotonyl-CoA/reductase (CCR) homologues.[29]

From the perspective of biosynthetic alternatives, the key question concerns the amino acid basis for substrate selection, as it underpins attempts to modify choice by site-directed mutagenesis. Efforts over decades to understand AT substrate specificity have resulted in the identification of multiple sequence motifs correlating with monomer preference (summarized in ref. 31). These include a four-residue motif spanning the catalytic His that is located 100 residues C-terminal of the active site serine in a loop connecting the large and small sub-domains, and which interacts directly with the bound substrate (Fig. 5). This motif is HAFH in malonyl-CoA-specific ATs and YASH in methylmalonyl-CoA-specific ATs, a difference rationalized by the role of the Phe in sterically excluding methylmalonyl-CoA and other α-branched extender units.[44] In addition, the catalytic His of YASH has been proposed to favor the (2*S*)-isomer of methylmalonyl-CoA over the (2*R*) configuration, again *via* physical blockage.[32] More recently, it was shown for ATs with atypical substrate specificity, that there is a rough correlation between the size of the extender unit side chains and the AlphaFold2-modeled active site volumes.[43]

In terms of recognition of discrete acyl-ACP substrates, a selection mechanism was initially proposed based on the solved crystal structure of an AT domain from the zwittermicin A PKS (ZmaA AT).[26] This analysis revealed a conserved, solvent-exposed hydrophobic patch on the surface of the AT domain (formed by three residues just upstream of the catalytic His) that was suggested to serve as a docking site for the carrier protein. Furthermore, partner binding was proposed to lead to induced structuration of the AT, opening up the active site for substrate entry, and thus serving as an additional specificity determinant in substrate selection.[26]

More detailed information has been obtained for an AT (VinK) that acts in *trans* to transfer a dipeptide between a stand-alone ACP VinL, and an ACP (VinP1ACP$_L$) that constitutes the starting module of the vicenistatin *cis*-AT PKS (Fig. 6).[45,46] This AT thus makes two choices in terms of its potential ACP partners. To understand the basis for its specificity, the contact surfaces between VinK and the two ACPs were probed by X-ray crystallography, with critical stabilization of the binary complexes provided by chemically cross-linking Ser-to-Cys mutants of VinK to the two ACPs. In the case of the ACP VinL,[45] recognition takes place between three residues of VinK situated at the inter-subdomain interface (Arg and Met contributed by the ferredoxin-like domain, and a second positively charged Arg of the catalytic domain) and the negatively-charged helix αII region of the ACP that encompasses the Ser attachment point of the Ppant cofactor (Fig. 6) (note: prototypical ACPs comprise four α-helices, αI–αIV, with the N-terminus of αII bearing the modified Ser residue; a short α-helix (αII′) is often present in the long loop αI–αII between the first two α-helices.[13]) The overall contact area is small, consistent with the necessarily transient nature of the interaction. Concerning the VinK/VinP1ACP$_L$ interface,[46] it again involves helix αII that contacts the two AT sub-domains, with both charged and hydrophobic residues forming the interface. However, the precise residues involved and the orientation of the ACP relative to VinK differ (Fig. 6). Overall, while these data provide important insights into how VinK recognizes its two ACP partners, they do not clarify the basis for anti-selection towards all of the other ACPs within the vicenistatin system.

2.2 Substrate/ACP partner selection by *trans*-AT PKS ATs

Concerning the ATs of *trans*-AT PKSs, the majority of these enzymes select malonyl-CoA as substrate,[47] but a few ATs specific for alternative building blocks have been characterized (*e.g.* ethylmalonyl-CoA (KirCII),[48] succinyl-CoA (aurantinins (**3**), gladiolin (**4**), etnangien (**5**)[49] *etc.*)), and which also exhibit varying degrees of substrate tolerance.[50] The ATs may be present in these systems as discrete proteins, or as domains within multi-domain proteins (*e.g.* tandem ATs, AT–ER, AT–AT–ER, *etc.*[51]). Interestingly, the majority of the discrete ATs clade separately from the ATs of *cis*-AT PKSs and more closely resemble the enzymes of bacterial fatty acid synthesis (which are also malonyl-CoA specific), with the exception of the ethylmalonyl-CoA-specific KirCII which is more similar to *cis*-AT PKS ATs.[51] Like certain *cis*-AT PKS-derived ATs, KirCII also exhibits broad specificity towards alternative extender units including allylmalonyl-CoA and propargylmalonyl-CoA.[52]

In the case of malonate-specific ATs, they typically charge all of the ACPs present in the systems with the same extender unit. Thus, there is no need for the enzymes to discriminate among potential substrates. In terms of the basis for ACP recognition, only a single structure is available – that of the disorazol (Dis) PKS *trans*-AT in complex with the ACP of the first chain extension module.[53] Capturing this interaction was again achieved by the use of a cross-linking agent in combination with introduction of a Cys residue into the AT. In this structure, the ACP contacts both the AT ferredoxin and αβ-hydrolase sub-domains, but *via* structural elements distinct from those deployed in *cis*-AT PKSs (*i.e.* the C-terminus of loop I–αII and helix αIII) (Fig. 7). Correspondingly, relative to the orientation of the ACP in the VinK/VinL complex, the Dis ACP is rotated by 180° (Fig. 7). Three residues present in these ACP regions are conserved among the six other ACPs of the disorazol system, making it likely that they serve as common recognition determinants for the AT.

In contrast to typical *trans*-ATs, the small number of acyltransferases with specificity other than malonate must precisely recognize a particular ACP within the target PKS system. To date, only the ethylmalonyl-CoA-selective KirCII has been characterized in this regard.[54] In this system, KirCII interacts

A fifth decision point involves choices made by *trans*-acting enzymes, because at least in theory, the catalyzed chemistry is compatible with multiple intermediates (here, the β-keto group resulting from chain extension). (F) The last decision PKS decision point concerns the mode of chain release, a choice typically made by the terminal domain of the subunit. Shown are several of the chemical options characterized to date.

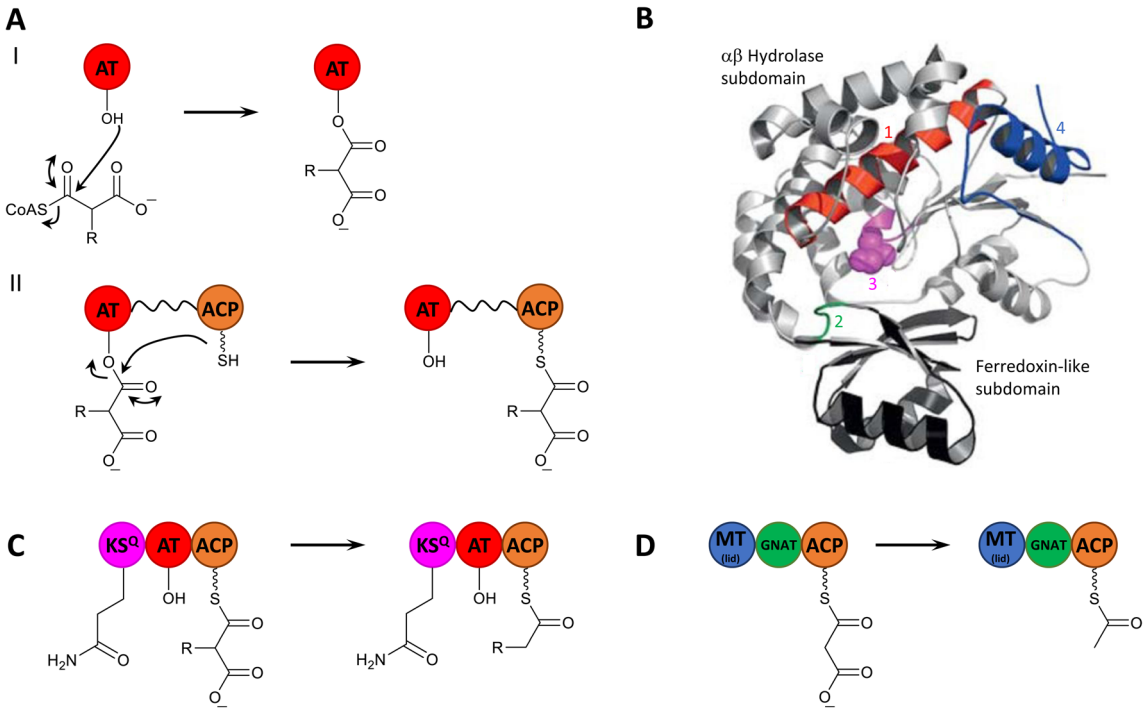

Fig. 5 Choices of building blocks. (A) Overall reaction catalyzed by the acyl transferase (AT) domains, as illustrated for selection of an extender unit. The reaction occurs in two steps: (I) AT self-acylation; and, (II) AT-catalyzed transfer to the downstream ACP domain. (B) Crystal structure of the AT_5 domain from DEBS[31] (PDB ID: 2HG4).[32] The domain consists of a catalytic αβ hydrolase subdomain and a ferredoxin-like subdomain. The active site is located at the inter-subdomain interface (the catalytic Ser is shown in pink). Identified specificity determinants (1–4) are highlighted, including 2 that indicates the location of the YASH motif. Reproduced from Smith and Tsai, The type I fatty acid and polyketide synthases: a tale of two megasynthases, *Nat. Prod. Rep.*, 2007, **24**, 1041–1072 (ref. 31). (C) Certain loading modules incorporate a KS^Q domain in addition to an AT and an ACP. This domain, in which the catalytic Cys is replaced by Gln, catalyzes the decarboxylation of ACP-tethered malonate (or its derivatives) to yield acyl-ACP starter units. (D) Many *trans*-AT PKSs and certain *cis*-AT systems initiate the biosynthesis *via* a Gcn5-related *N*-acetyltransferase (GNAT) domain, which catalyzes the decarboxylation of ACP-linked malonate. In the case of the curacin system depicted here, this domain is preceded by the inactive remnants of a methyltransferase (MT) domain.[37]

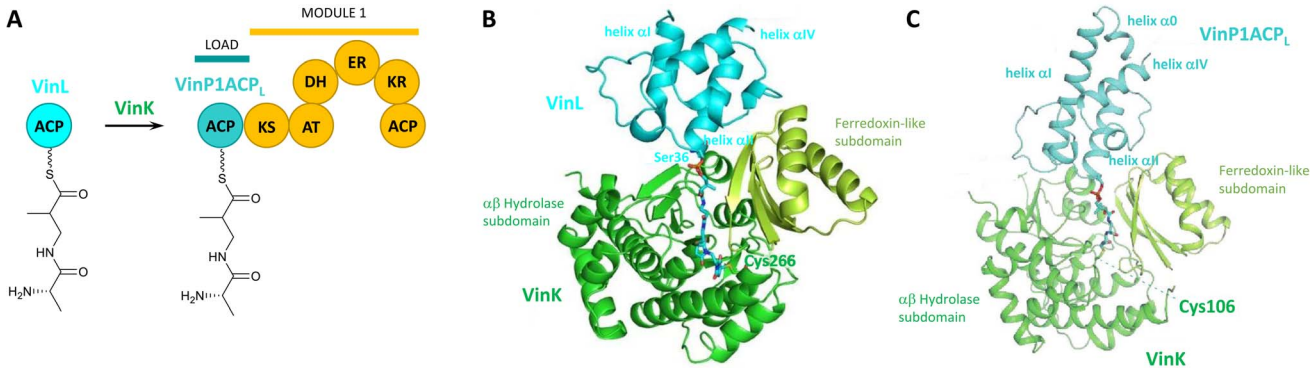

Fig. 6 Choices made by the *trans*-acting AT VinK during vicenistatin biosynthesis. (A) ACPs partners of VinK. The enzyme has to recognize both the substrate attached to the discrete ACP VinL, and transacylate it specifically onto the PKS-integrated $VinP1ACP_L$. (B) Solved crystal structure of VinL ACP (blue) in complex with VinK (shades of green) (PDB IDs: 5CZC, 5CZD).[45] The complex was stabilized by replacing VinK Ser266 by Cys, which allowed cross-linking with the ACP phosphopantetheine (Ppant) cofactor *via* a bifunctional maleimide reagent. The protein–protein interface is formed principally between three residues of VinK that contact helix αII of the ACP (note, the nomenclature of the α-helices has been modified from the original paper. The stereochemistry at the α center of the intermediate is not known[320]). Reproduced and adapted from Miyanaga, *et al.*, Structure-based analysis of the molecular interactions between acyltransferase and acyl carrier protein in vicenistatin biosynthesis, *Proc. Natl. Acad. Sci. U. S. A.*, 2016, **113**, 1802–1807 (ref. 45). (C) Crystallographic complex of $VinP1ACP_L$ covalently bound to VinK (PDB ID: 8H6S).[46] The linkage in this case was generated by engineering a Cys mutant of the catalytic Ser106 of VinK, and reacting it with $VinP1ACP_L$ bearing a bromoacetamide-functionalized Ppant cofactor.[321,322] The interface again involves helix αII that contacts both the catalytic and ferredoxin-like subdomains. The observed docking pose is similar to that in (B), but the involved residues and precise ACP orientation differ. Reproduced and adapted from Miyanaga, *et al.*, Structural basis of transient interactions of acyltransferase VinK with the loading acyl carrier protein of the vicenistatin modular polyketide synthase, *Biochemistry*, 2023, **62**, 17–21, copyright 2023, with permission from the American Chemical Society (ref. 46).

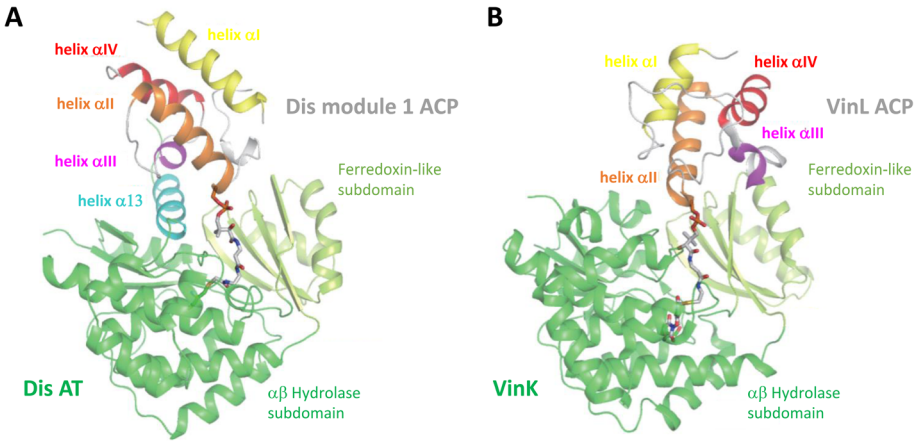

Fig. 7 Recognition of ACP domains by *trans*-acting ATs. (A) Crystal structure of the cross-linked complex between the discrete Dis AT and its module 1 ACP partner (PDB ID: 5ZK4).[53] The contact interface on the ACP includes the C-terminus of loop αI–αII and helix III which interact with both the ferredoxin-like and αβ-hydrolase subdomains of Dis AT. Relative to the orientation of the ACP in the VinK/VinL complex[45] (B), the Dis ACP is rotated by 180°. Reproduced and adapted from Miyanaga, *et al.*, Structural basis of protein–protein interactions between a *trans*-acting acyltransferase and acyl carrier protein in disorazole biosynthesis, *J. Am. Chem. Soc.*, 2018, **140**, 7970–7978, copyright 2018, with permission from the American Chemical Society (ref. 53).

preferentially with the ACP of module 5 (ACP_5). The remaining 13 ACPs of the system are charged with malonate by the second *trans*-acting AT_2 situated within the tridomain protein KirCI (AT_1–AT_2–ER), which conversely does not act on ACP_5.[54,55] Interestingly, when KirCI AT_2 is excised from the multi-domain protein, it can malonate ACP_5, showing that its enzymatic context and/or the flanking domains contribute to its specificity. Concerning KirCII, while its preferred substrate is ACP_5, it can interact productively *in vitro* with at least three other ACPs from the same system, raising the question of whether these contacts occur *in vivo* and lead to as yet undiscovered kirromycin derivatives bearing additional branching ethyl groups.[54]

In terms of the basis for recognition of partner ACP by KirCII, the binding motif on ACP_5 was mapped using alanine-scanning mutagenesis, while the complex between the two proteins was computationally modeled (Fig. 8).[54] Both of these approaches converged on the same residues of the two partners and which are uniformly charged, and indeed several overall surface regions in the modeled proteins exhibit complementary electrostatic features. In ACP_5, these amino acids are located on the C-terminus of helix α1 and the N-terminus of helix αII, as well as the intervening long loop αI–αII, which encompasses helix αII′ (V57–D60). The contact motifs in KirCII are positioned on both the small and large sub-domains, and an upstream region that resembles the folded linker region present in *cis*-AT PKSs between the KS and AT domains (the so-called linker domain (LD)),[32] which together form an extended binding cleft. The precise basis for AT docking is thus distinct from all of the other complexes described earlier.

Validation for the interaction model was provided by introducing charge-reversal mutations into the ACP positions that reduced or eliminated KirCII transacylase activity, as well as by creating a chimeric ACP_5/ACP_{10} protein incorporating the HαI–LαI–αII–HαII region of ACP_5 that was recognized by KirCII.[54] The overall picture that emerged from this work is that partner choice by KirCII is a consequence of a unique set of both

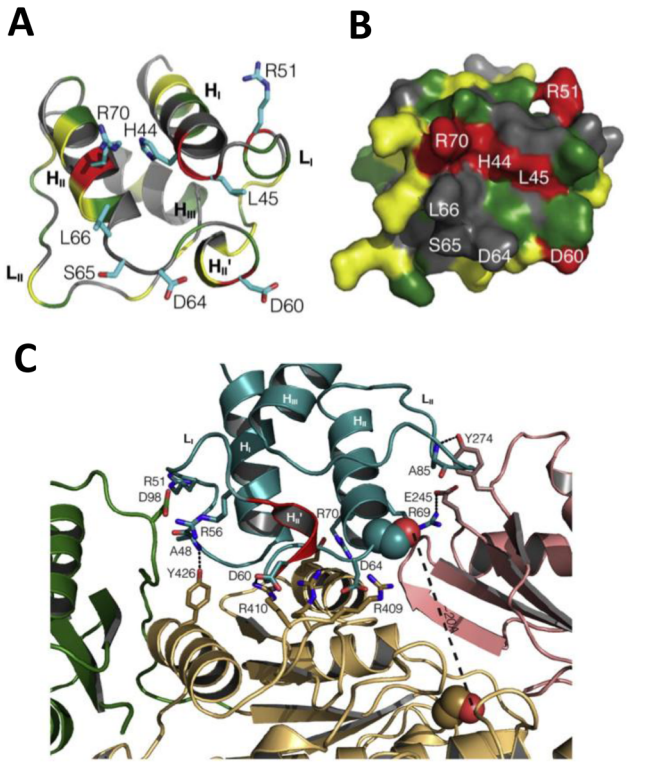

Fig. 8 Basis for recognition of Kir ACP_5 by KirCII.[54] Residues on ACP_5 involved in binding, as revealed by alanine-scanning mutagenesis coupled with activity assay, shown in (A) a homology model ribbon diagram, and (B) the corresponding computed surface. Red, <20% activity; green 20–80% activity; yellow, >80% activity (grey, not determined). (C) Computational model for the KirCII/ACP_5 complex obtained by docking of the respective homology models. The KirCII catalytic subdomain is shown in light orange, the ferredoxin-like subdomain in light pink, the linker domain (LD) in green, and ACP_5 in teal. The active Ser of both ACP_5 and KirCII are represented as spheres. Reproduced from Ye, *et al.*, Reprogramming acyl carrier protein interactions of an acyl-CoA promiscuous *trans*-acyltransferase, *Chem. Biol.*, 2014, **21**, 636–646, copyright 2014, with permission from Elsevier (ref. 54).

charged and hydrophobic residues on the ACP_5 surface relative to the other ACPs of the same systems, including the critical residues Arg51, Arg70 (Fig. 8), Leu58 and Val61. Notably, introduction of Leu in place of Phe at the corresponding position of an ACP from the erythromycin system was sufficient to boost *trans*-acylation activity by 14-fold relative to the wild type protein *in vitro*. However, it remains to be demonstrated whether such ACP engineering can be used to efficiently redirect substrate choice by *trans*-ATs *in vivo* to generate novel analogues (see Section 2.3).

2.3 Engineering monomer choice

Note: all recommendations for PKS engineering based on decision point analysis are summarized in Table 1. Having elucidated sequence elements of ATs involved in mediating their choice of substrates (both the small molecule to be transferred and its carrier (CoA or ACP), and the target PKS ACP domain), it becomes possible, at least in principle, to manipulate these specificities by targeted engineering. The interest of this approach is that it minimizes the risks of perturbing both the AT structure and its interaction interfaces with other domains in *cis* or in *trans*.[29] In this context, much effort has been expended to alter extender unit choice in *cis*-AT PKSs by site-directed mutation. Although such experiments have resulted in modified specificity,[56–59] they have not yet yielded a complete shift in substrate selection. Moreover, they are systematically accompanied by a loss in catalytic efficiency.[30] To our knowledge, only one attempt has been made to influence substrate selection by *trans*-AT PKS ATs (*vide infra*), but the intrinsic substrate tolerance of KirCII has been exploited to direct the biosynthesis of several kirromycin analogues, albeit as minor components in a mixture with the native metabolite.[50,52]

Globally, work in this area demonstrates that residues beyond those targeted contribute to AT small-molecule substrate selection, potentially *via* effects on the interactions between the catalytic and ferredoxin sub-domains,[33] as well as overall conformational modulation of the enzyme structure.[60] Indeed, amino acids observed to impact specificity are present in both sub-domains, and many do not interact directly with the substrate.[33] While improved results have been obtained by exchanging more extensive structural motifs,[33] the distributed origin of AT substrate specificity argues for alternative approaches for modifying building block selection in PKS systems.

In this context, two major strategies have been pursued (Fig. 9): (i) exchange of entire AT domains of alternative specificity; and, (ii) inactivation of AT domains coupled with in-*trans* complementation. Swapping of AT domains has been a go-to engineering strategy for decades,[48,51,53–61] as, at least theoretically, structural alteration of the domain itself should be minimal. However, choices must always be made concerning the sites used for the swapping experiments. Traditionally, the upstream fusion points were located within the conserved C-terminal boundary of the upstream KS domain (thus the AT was cloned with its upstream KS–AT linker) or the conserved N-terminal boundary of the AT domain (thus preserving the upstream KS–AT linker from the host PKS), while the downstream junction was typically engineered at several sites within a relatively conserved region within the linker region following

Table 1 Roadmap for PKS engineering based on decision point analysis

Targeted modification	Suggested engineering strategies	Select relevant references
Specificity of building block selection	(i) Exchange of AT domains using several established fusion points	61 and 62
	(ii) AT inactivation/*trans*-AT complementation, but currently limited in substrate scope	65–69
Swapping of PKS domains (aside from ATs)	Generally low-yielding, reflecting the high interdependence of PKS domains	64 and 76
Toggling between vectorial and iterative behavior of individual modules	Difficult due to lack of understanding of the factors underpinning the intrinsically iterative behavior of certain modules, and KS proof-reading	129 and 132
Use of modules and subunits as building blocks (including subunit splitting)	(i) Evaluation of various domain sets (classical modules, XUs and XUAs), coupled in future with strategies to broaden KS substrate tolerance	162, 185 and 323
	(ii) Deployment of intrinsically orthogonal sets of docking domains (DDs)	193
Harnessing *trans*-acting enzymes to catalyze novel, regiospecific transformations and/or modifying their substrate specificity	Complicated by the variability in control mechanisms, especially in *trans*-AT PKSs. No successful experiments of this type reported to date	N.A.
Altering the timing of chain liberation	Relocation of a broad-specificity thioesterase or other release domain to the new termination point	80 and 313
Changing TE release chemistry and substrate specificity	Modifications achieved by site-directed mutagenesis, but the generalizability of these results is not yet clear	254 and 269

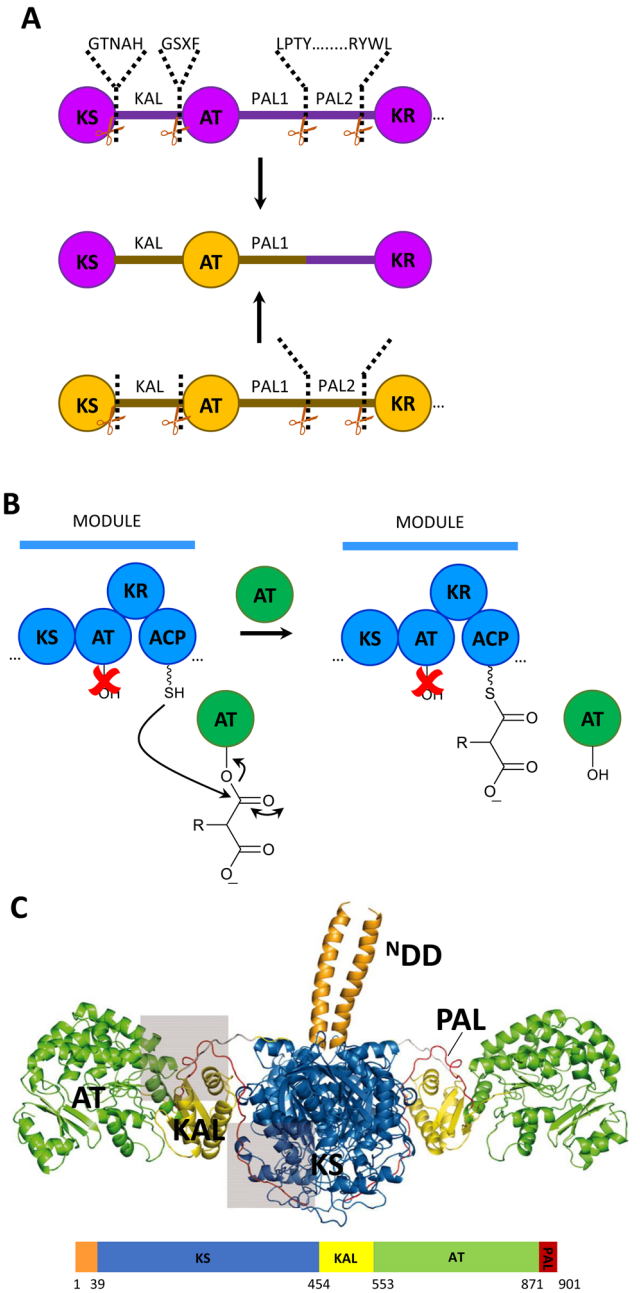

Fig. 9 Genetic engineering of AT domains. (A) Sites within the KAL (KS–AT linker), PAL1 and PAL2 (PAL, post-AT linker) regions typically used for AT swaps. The hybrid generated by cutting at the GTNAH of the KAL region and the LPTY motif at the end of PAL1, is shown in the middle of the panel. (B) AT complementation strategy. In this approach, inactivation of the *cis*-AT of the target module by sitedirected mutagenesis is coupled with expression of a *trans*-acting AT of alternative specificity. (C) Crystal structure of the DEBS KS_5–AT_5 didomain homodimer (PDB ID: 2HG4).[32] This structure revealed interactions between the KAL and PAL regions, and that they also contact the KS and AT domains (grey boxes), explaining why junctions within these linker motifs were found to disrupt the function of engineered modules.[40] Panel (C) was reproduced and adapted from Tang, *et al.*, The 2.7 Å crystal structure of a 194-kDa homodimeric fragment of the 6-deoxyerythronolide B synthase, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 11124–11129, copyright 2006, with permission from the National Academy of Sciences (ref. 32). Legend: ^{N}DD, N-terminal docking domain.

the AT domain (Fig. 9). However, certain of these commonly-used junctions were shown in model engineered systems to disrupt the ability of the KS and ACP domains to interact effectively to catalyze chain extension, while leaving the AT activity intact, as predicted – an effect likely attributable to perturbation of the overall module architecture (in this case, the fusion sites were positioned just upstream of the AT domain and within the downstream AT–KR linker).[40] This observation can be rationalized by the fact the KS–AT linker (referred to as 'KAL') and the post-AT linker ('PAL', and in particular, 'PAL2', the residues in the last half of this region) interact with each other and contact both the KS and AT domains (Fig. 9).[32]

Thus, more recently, efforts were made to identify 'optimal' sites for these modifications. Two approaches were used: comprehensive study *in vitro* of combinations of the previously exploited fusion points,[61] and engineering of a fluorescence-based biosensor that reported on *in vivo* protein solubility as a proxy for catalytic activity (Fig. 10).[62] The first strategy revealed that KAL-AT-PAL1 exchange fragments yielded the most stable and active enzymes (note: in this case, the heterologous AT is introduced alongside its complete upstream linker and the first ('PAL1') portion of the downstream linker) (Fig. 9). Nevertheless, the authors tempered their result following the solubility screening, as it showed that several different combinations of sites up and downstream of the AT produced functional hybrids, and the recommended junctions did not include the conserved end of the KS in combination with PAL1, as suggested by the earlier work.[61] Nonetheless, in subsequent experiments, the same authors successfully used the KAL-AT-PAL1 boundaries to assess the ability of twelve ATs of atypical specificity to function in a common modular context,[43] and to engineer the PKS-catalyzed biosynthesis of multiple derivatives of δ-valerolactam.[63]

In the second strategy (Fig. 9), a particular AT domain of a *cis*-AT PKS is inactivated, and a *trans*-AT of different specificity is provided to replace the missing activity *in vivo*. The success of such cross-complementation evidently depends on the capacity of the *trans*-acting AT to recognize the non-native, *cis*-AT PKS ACP target, and less explicitly, on the ability of the other *cis*-acting ATs to outcompete the *trans*-AT for charging of their cognate ACPs. The fact that this strategy works is consistent with the accumulated data obtained to date showing that the ACP/AT interaction in modular PKS systems is fundamentally plastic.[45,46,54,64] Multiple specific solutions have been found by evolution in terms of secondary structures/amino acids employed by both partners to form active complexes that position the 19 Å-long Ppant cofactor within reach of the AT active site.

To date, this approach has been shown to function with exogenously added stand-alone malonyl-CoA:ACP transacylase (MAT) that participates in the type II biosynthesis of fatty acids and aromatic polyketides in *Streptomyces coelicolor*,[65,66] and the *trans*-ATs from the disorazol, kirromycin and bryostatin *trans*-AT PKS systems.[67–69] Direct comparison between the KirCII and Dis *trans*-ATs showed the Dis AT to be substantially more effective at complementing *cis*-AT PKS ATs, potentially because the three conserved residues of its native ACP targets are highly

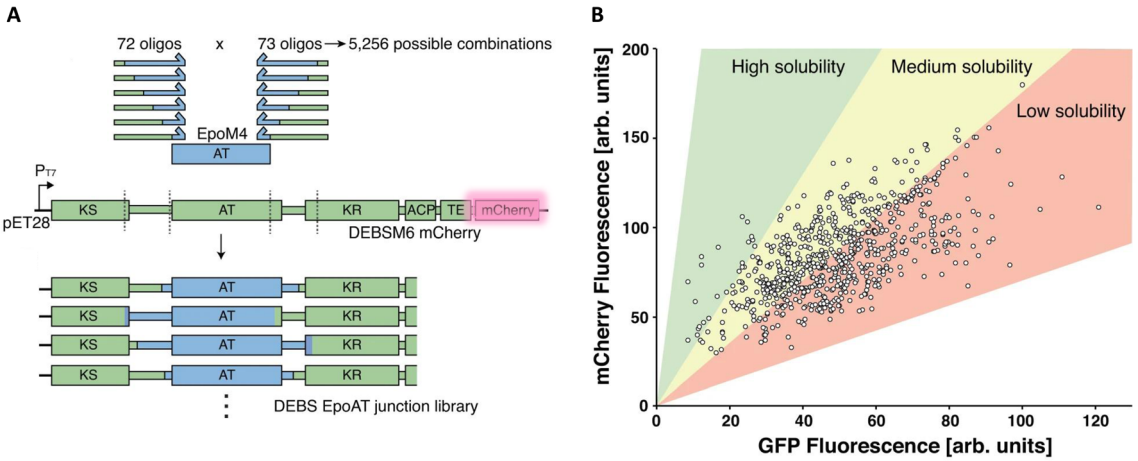

Fig. 10 Use of a fluorescence-based biosensor to identify 'optimal' junctions for AT exchanges.[62] (A) A library of constructs was generated to test thousands of different fusion points between the parental DEBS module 6 and the heterologous epothilone (Epo) module 4 AT, and the resulting proteins expressed from vector pET28. Junctions were located with the KAL and PAL regions, but additionally within the catalytic KS, AT and KR domains. (B) Screening for the solubility of the hybrid PKSs as a proxy for activity was carried out by comparing the mCherry fluorescence to that of a GFP-based biosensor for misfolded proteins. Reproduced from Englund, *et al.*, Biosensor guided polyketide synthases engineering for optimization of domain exchange boundaries, *Nat. Commun.*, 2023, **14**, 4871 (ref. 62).

conserved in the ACP domains of *cis*-AT PKSs.[51] A particularly notable use of the Dis *trans*-AT was to complement AT-null mutants of an engineered mini-PKS, resulting in site-selective incorporation of fluoromalonyl-CoA (**6**) (Fig. 11).[69] The ability of the AT to utilize fluoromalonyl-CoA in preference to malonyl-CoA for *in vivo* applications was subsequently improved by structure-guided engineering.[70] For this, saturation mutagenesis was carried out of a conserved residue correlated with recognition of malonate (F190), and coupled with two mutations (L87V/A) designed to compromise positioning of the extender unit relative to the AT oxyanion hole – modifications hypothesized to be more deleterious in the presence of malonate than activated fluoromalonate. One mutant exhibiting improved selectivity towards fluoromalonyl-CoA (**6**) as well as reduced hydrolysis was used to generate multiple fluorinated versions of the erythromycin macrolide *in vitro* (Fig. 11).

Despite these evident successes, the complementation approach is presently limited by the relatively restricted panel of

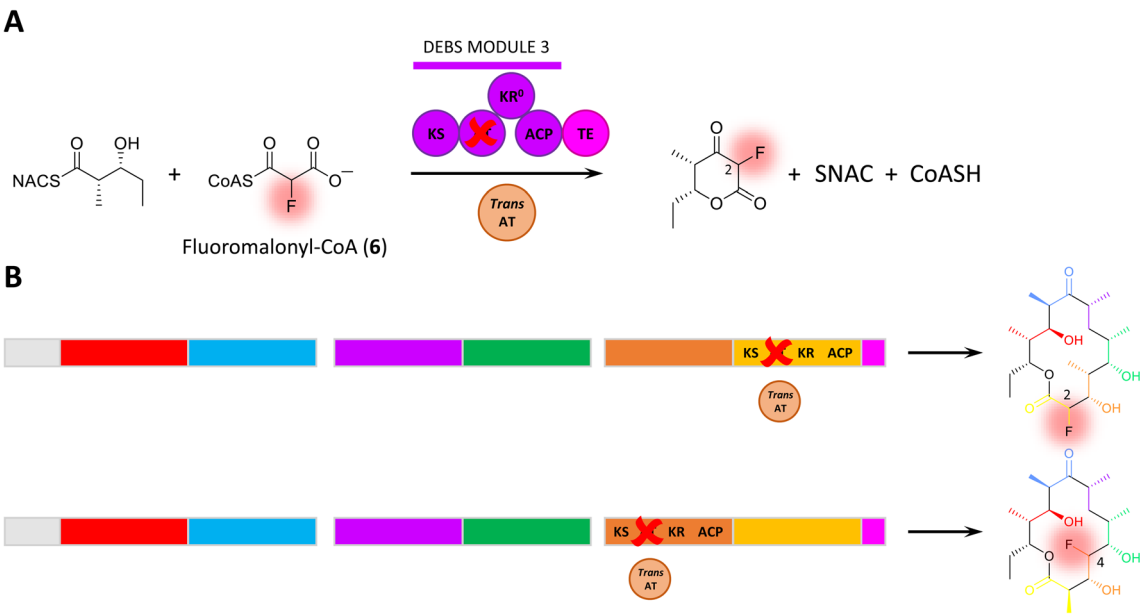

Fig. 11 Incorporation of fluoromalonate into triketide keto lactone. (A) In this experiment, the Dis *trans*-AT was used to complement an engineered version of DEBS module 3-TE in which the AT domain was inactivated. Running the reaction *in vitro* in the presence of fluoromalonyl-CoA (**6**) and native DEBS diketide derivatized as its *N*-acetylcysteamine (SNAC) thioester, resulted in the target 2-fluoro-2-desmethyl-triketide lactone. (B) Following structure-guided engineering, the Dis *trans*-AT was used to complement AT mutants of modules 5 and 6 within the context of a complete DEBS assembly line, giving rise to several fluorinated analogues of 6-dEB. Key: KS, ketosynthase; AT, acyl transferase; KR^0, KR inactive for ketoreduction; ACP, acyl carrier protein.

substrates recognized by *trans*-AT PKS AT domains. In addition, the extent of off-target binding of *trans*-ATs with other ACPs from the engineered systems has not been directly investigated. Such events could lead to stalled PKSs and/or the production of unwanted derivatives. Indeed, as alluded to earlier, the factors underlying ACP anti-selection by *trans*-AT PKS ATs remain largely mysterious, even in the context of their native systems, and would be an interesting subject of future work.

3. Navigating intra/intermodular options in *cis*-AT and *trans*-AT PKSs

Once a starter unit or the growing polyketide is loaded onto the KS domain of a particular module, KS-catalyzed Claisen-type condensation can take place between the substrate and an ACP-tethered extender unit. This homologation reaction results in a β-keto intermediate attached to the ACP. At this stage, every PKS module must make the critical decision whether to traffic the chain to the full suite of processing enzymes within the module or to transfer the intermediate to a downstream domain, typically a KS, or at the end of the PKS, a TE domain (Fig. 4). Or in other words, a choice must be made between intra- and intermodular interactions. As judged by the fact that PKS systems give rise to one or a limited number of products,[71] this process is highly faithful. Perhaps counterintuitively, no choice must be made concerning which of the available modification enzymes to visit at any given step, as there are no chemical consequences of soliciting inappropriate partners (*e.g.* if the β-keto intermediate is presented to a DH or ER domain, no transformation is possible). It is also relevant that data obtained *in vitro* have shown that a new cycle of chain extension by any given *cis*-AT PKS module cannot be initiated until the intermediate is handed off to a downstream module, and thus the function of each module has been likened to a turnstile.[72]

Several theoretical models can be envisioned for how this intra- *vs.* intermodular control could be achieved, which differ in the extent of ACP mobility (Fig. 12).[4,73,74] In the first, ACP dynamics are entirely unconstrained, and therefore the domain searches out its reaction partners using a random walk.[75] In this mode, non-productive interactions can occur that will slow down the overall process, but the only possibilities in terms of the available chemistry concern processing reactions *vs.* chain transfer. Thus, in order that product fidelity be maintained, the downstream acceptor domain would need to have strict specificity for only those intermediates that had been fully modified by the upstream domains – effectively sending the chains back for additional processing if they arrive prematurely.

In the second possibility, the trajectory of the ACP domain is restrained by the overall conformational state of the module – *i.e.* the positioning of the other domains relative to each other and to the ACP physically encourages communication with a subset of the domains while disfavoring or even sterically prohibiting others. The advantage of this model is that untimely transfer of the chain could be avoided by confining the acyl-ACP to its own module until the complete set of transformations occurs. In the third option, the architectural state of the module

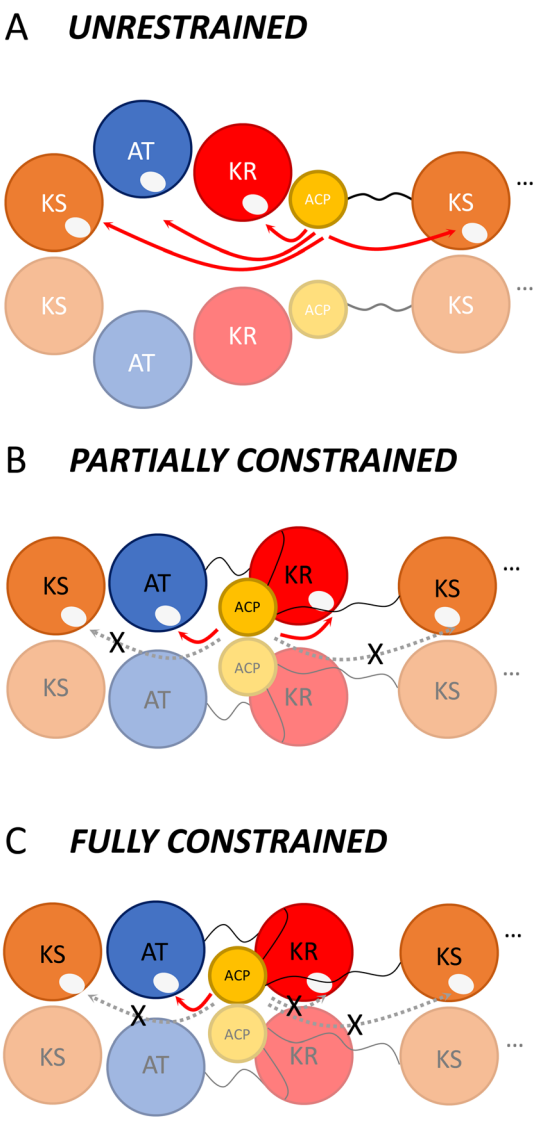

Fig. 12 Models for the control of intra- *vs.* intermodular interactions in modular PKSs. (A) Model in which ACP movement is completely unconstrained. In this case, the downstream KS acts as the key gatekeeper, ensuring that all modifications have occurred in the upstream module before accepting the intermediate for extension. (B) Partially constrained model, in which access of the ACP to certain active sites (in this case, the up- and downstream KSs, shown by dotted arrows) is blocked by the overall conformational state of the module. (C) In this model, the conformational status of the module forces the ACP into a single interaction (here, with the AT domain), so no choice is required. In (A–C), the active sites of the individual domains are indicated with white ellipses. Key: KS, ketosynthase; AT, acyltransferase; KR, ketoreductase; ACP, acyl carrier protein.

at each stage of the cycle is such that the ACP is directed only to the appropriate next domain by excluding all incorrect associations (here, there is in fact no choice).

For the *cis*-AT PKSs, data in support of the second model have been obtained by the recent elucidation at high-resolution of the structures of two intact modules of similar domain composition, *via* a combination of X-ray crystallography and cryo-electron microscopy[64,76] (note: the cryo-EM structures of a PKS module reported earlier are not discussed here as they

diverge substantially from the more recent results, and indeed, their physiological relevance has been called into question[77]). One module was sourced from the prototypical PKS system responsible for biosynthesis of erythromycin A (DEBS module 1),[14] and the second from the lasalocid A assembly line (Lsd14).[78,79] DEBS M1 was modified genetically to include both N- and C-terminal dimerization/protein–protein interaction elements (a docking domain (DD)[19] (see Section 5.1) and the DEBS TE,[80] respectively), while Lsd14 is a mono-modular subunit that natively incorporates both N- and C-terminal homodimeric DDs, as well as an additional dimerization element (DE) between its AT and KR domains (Fig. 13).[18] The presence in particular of C-terminal dimerization motifs is important, as they confine the two ACPs to a single reaction chamber.[64] In any case, the key enabling approach was to reduce the high conformational heterogeneity of the modules by using an antigen-binding fragment called Fab 1B2 (ref. 77) to stabilize the homodimeric KS–AT didomain portion of the structures (in the case of Lsd14, DEBS N-terminal DDs were introduced in place of the natural pair to enable this interaction[64]).

The principal findings of these efforts have been nicely summarized in several recent reviews.[81,82] Thus, here, we will focus on the consequences for choices at this point in the biosynthesis. One of the most important observations is that the modules adopt multiple distinct conformational states. In one of the DEBS M1 conformations, a single ACP was docked against the KS domain of the partner monomer, a positioning consistent with the chain extension state. Indeed, only one of the two ACPs was ever visualized in the three DEBS M1 structures, presumably because the second ACP sharing the same reaction center remained too mobile. This finding has led to the proposal that chain elongation and modification occur asynchronously in one reaction center, followed by the equivalent process in the second center[64,83] (although there is no reason in principle that the two centers should alternate[82]). In a second state (termed 'turnstile-closed'), the two AT domains adjacent to the KS homodimer were 'flexed' – *i.e.* they adopted a bent conformation that fully occluded the KS active site. Although no ACP domains were visible in this structure,[76] this configuration presumably promotes the interaction between the acyl-ACP substrate and the processing domains. Finally, the authors observed a conformation in which only one of the two AT domains was flexed, which is suggested to represent a half-open intermediate between the turnstile-closed and chain extension states.[76,82]

Overall, the obtained data on DEBS are consistent with the intermediate control model discussed above (Fig. 12), in which the global configuration of the module favors a subset of ACP/partner contacts. Interestingly, they could also be extrapolated to explain the absence of premature chain transfer between modules. If, as has been hypothesized previously,[84] the consecutive modules of modular PKSs function simultaneously, then the presence of two flexed ATs in the module downstream of any given module would prevent the next KS from accepting the growing chain, and thus any productive intermodular interactions. In this scheme, starting from the last module and progressing along the assembly line, the sequential transfer of intermediates to a downstream acceptor – also a necessary prerequisite to turnstile opening[72] – would progressively deblock the KS active sites, allowing for the next round of KS charging and subsequent chain extension (Fig. 14).[72] Thus, in this 'vending machine' model, it is the conformation of the

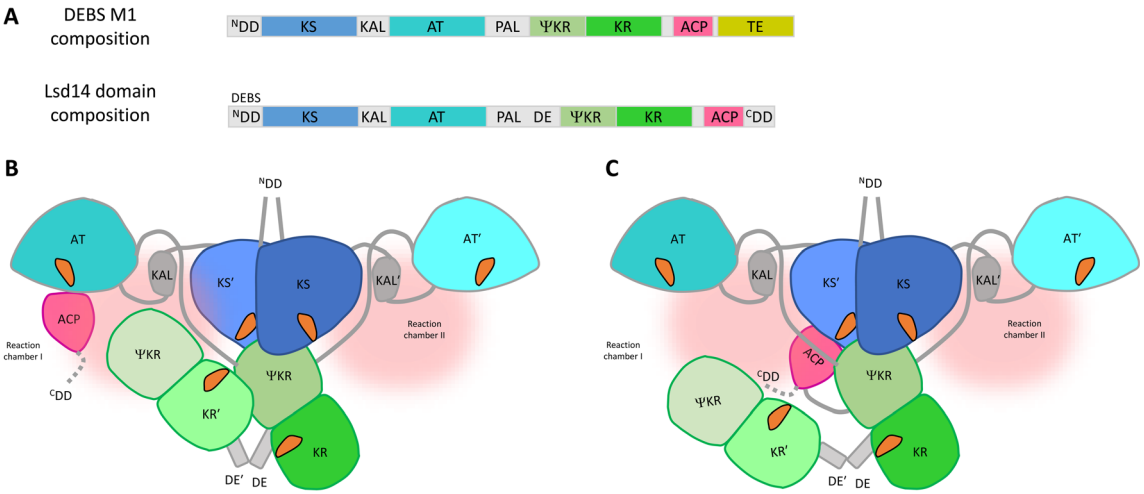

Fig. 13 Structural analysis of whole modules in distinct conformational states. (A) The domain composition of the DEBS module 1 and lasalocid (Lsd) module 14 model systems.[64,76] Both constructs include the same functional domains KS, AT, KR and ACP, as well as artificially grafted N-terminal docking domains (^{N}DD) from DEBS. The C-termini of the proteins differ, as DEBS M1 terminates in an engineered thioesterase (TE), while Lsd14 ends in its natural C-terminal DD (^{C}DD). Lsd14 additionally incorporates a dimerization element (DE) N-terminal to its KR domain. (B) Schematic of the three-dimensional organization of Lsd14 in which the ACP is docked against the AT domain.[64] The domains belonging to the second monomer are indicated with a prime. The overall configuration of the homodimeric module gives rise to two reaction chambers, only one of which is occupied by an ACP, due to the presence of the downstream, homodimeric ^{C}DDs. In this state, access by the ACP to the KS active site is sterically blocked by the KR. (C) Schematic of the conformation of Lsd14 in which the KR has moved relative to its position in (B), allowing the ACP to engage the KS domain. Key: KS, ketosynthase; KAL, KS–AT linker; AT, acyltransferase; PAL, post-AT linker; ΨKR, structural portion of the ketoreductase; KR, catalytic portion of the ketoreductase; ACP, acyl carrier protein.

downstream module that acts as a critical control element for the intra/intermolecular decision point.

In the case of Lsd14, two conformations were observed in which a single ACP was docked against either the AT or KS domains, corresponding respectively to the extender unit-charging and condensation-competent states (Fig. 13). These data again confirm that while both ACPs occupy the same reaction center due to the presence of up- and downstream dimerization motifs, only one ACP at a time engages with the catalytic partners. Curiously, in no case were bent ATs observed, even though a third configuration analogous to the turnstile-closed state was identified in which the two ACPs were undocked. Together, these data have inspired a related, but distinct proposal for modular control, wherein specific ACP/partner interactions at each stage are kinetically favored based on the physical proximity of the two domains coupled with blocking of other active sites.[64,82]

More specifically, at the beginning of the catalytic cycle, the sole active ACP is directed to the AT domain of its reaction center due to steric obstruction of the KS active site by a KR domain. Movement of the KR then deblocks the KS allowing the ACP that now bears the extender unit to dock, and chain extension occurs. A larger scale 'pendulum' movement then directs the ACP away from the KS–AT condensing region and towards the KR domains, again resulting in an increased probability of interaction.[82] Finally, transfer to the downstream module is promoted by the closer proximity of the acceptor KS relative to that within the ACP's own module.[82] The next round of chain extension is suggested to take place in the opposite reaction center that has in the meantime become transacylated with the polyketide intermediate, due to combined swinging and rotation of the processing portion of the module (DE–KR–ACP)$_2$ around a hinge created by the DE (*i.e.* the two reaction chambers are equivalent relative to the DE pivot point). Overall, according to this model, the key choice between intra- and intermodular interactions arises naturally from the progressive distancing of the ACP from its potential reaction partners within its own module, with a concomitant kinetic penalty for the corresponding reactions. Nonetheless, it remains to establish the factors underpinning the module's 'decision' to adopt this particular sequence of conformations.

At present, structural data are completely lacking for full *cis*-AT PKS modules incorporating additional processing domains (typically DH or DH + ER). Nonetheless, recent analysis of a tri-domain DH–ER–KR portion of the juvenimicin *cis*-AT PKS has provided initial insight (Fig. 15).[85] This work revealed that the region homodimerizes *via* the DH domain, as expected based on studies of discrete DHs,[16,86–91] and that the ER is monomeric, as observed previously for excised ER[92] and ER–KR fragments.[64] In principle, it would seem that the global architecture adopted by Lsd14 (ref. 64 and 76) could accommodate the presence of a homodimeric DH, as it would take the place of the similarly homodimeric DE region just downstream of the AT domains (Fig. 13).[64] Furthermore, the outward projection of the succeeding monomeric ER and KR domains from both sides of the central DH would lead naturally to two reaction chambers harboring the complete set of enzymes. Measurements of the

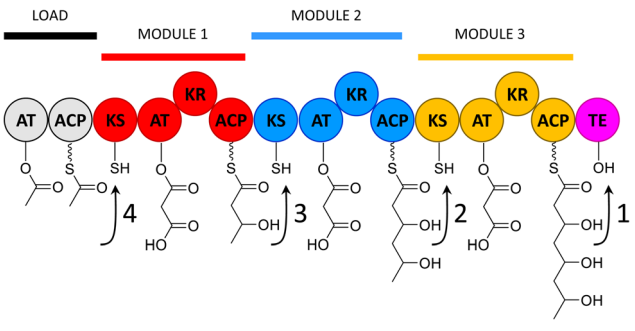

Fig. 14 The 'vending machine' model for sequential chain transfers in modular polyketide biosynthesis.[84] Deblocking of the last module by transfer of intermediate to the downstream TE results in turnstile opening,[72] allowing the KS of the same module to be transacylated by the acyl chain tethered to the upstream module. The upstream module is thus deblocked, and the cycle repeats until the KSs of all modules bear substrates in preparation for the next rounds of chain extension. Key: AT, acyl transferase; ACP, acyl carrier protein; KS, ketosynthase; KR, ketoreductase; TE, thioesterase.

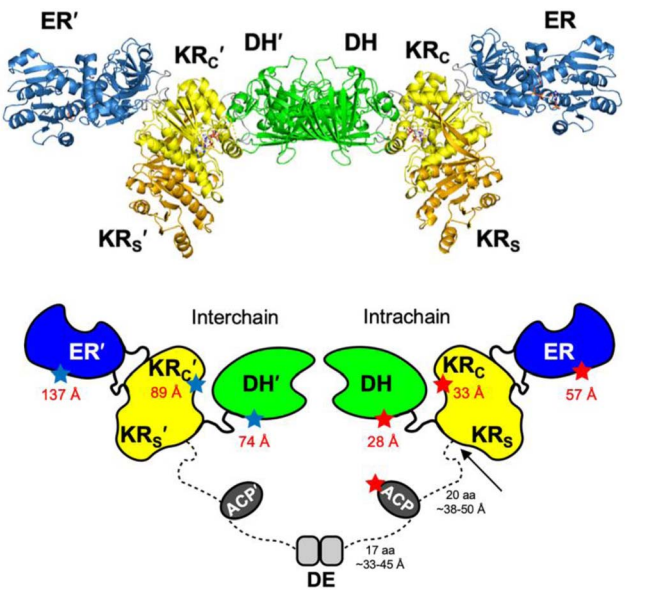

Fig. 15 Crystal structure and corresponding schematic of the homodimeric DH–ER–KR tridomain of the juvenimicin *cis*-AT PKS, showing the relative disposition of the domains and their intimate association with the intervening linkers (PDB ID: 8G7W).[85] Consistent with earlier work on homologous excised domains, the DH is homodimeric and so contributes to the overall homodimerization of the region/module, whereas the ER is monomeric. Shown are the distances measured from the Ppant attachment point of the ACP to the partner domain active sites. Intramonomer interactions are indicated with red stars, and intermonomer contacts with blue stars. In view of the measured distances between the active sites, it has been hypothesized that the ACP can solicit the three domains within its own monomer without architectural rearrangement of the tridomain region. Key: ER, enoyl reductase; KR$_S$, structural domain of the ketoreductase; KR$_C$, catalytic domain of the ketoreductase; DH, dehydratase; ACP, acyl carrier protein; DE, dimerization element. Reproduced from McCullough, *et al.*, Structure of a modular polyketide synthase reducing region, *Structure*, 2023, **31**, 1109–1120, copyright 2023, with permission from Elsevier (ref. 85).

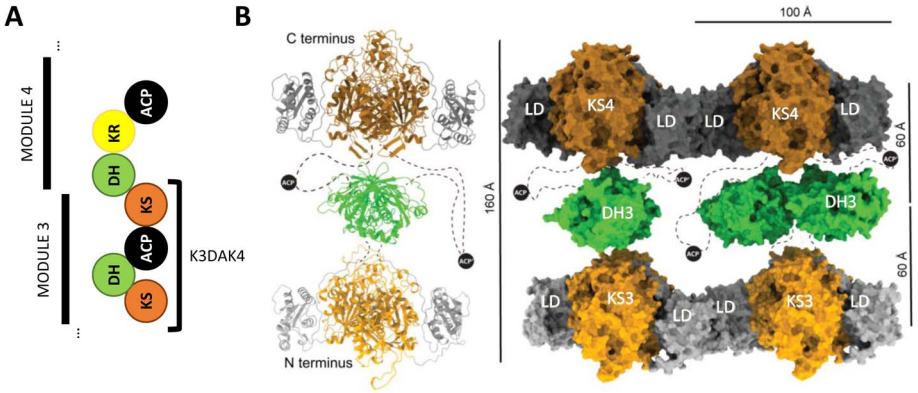

Fig. 16 Analysis by cryo-EM of a 'bimodule core' sourced from a *trans*-AT PKS of *Bacillus brevis*.[93] (A) Schematic of the PKS illustrating the analyzed tetradomain portion (K3DAK4) that spans an intermodular junction. (B) Plausible positions for the three catalytic domains based on a pseudo-atomic model of the core region. The ACP domains were not resolved, consistent with their high intrinsic mobility, and so are indicated schematically. (C) Surface representation of the bimodule core, incorporating two models of the DH to reflect its rotational heterogeneity. The distances between the center of mass of the catalytic domains are indicated, as well as the overall dimensions of the region. The image also shows the lateral associations between the linker domains (LDs) of equivalent KS domains mediated by their LINKS motifs,[94] that on the scale of many module cores, give rise to polymeric filaments. Reproduced and adapted from Tittes, *et al.*, The structure of a polyketide synthase bimodule core, *Sci. Adv.*, 2022, **8**, eabo6918 (ref. 93).

distances between the three active sites in one monomer and the point of attachment of the ACP to the KR (Fig. 13), suggest that all processing steps can occur without conformational rearrangement of the tridomain region.[64] Indeed, the tight association between the three tailoring domains and their intervening linkers is argued to intrinsically limit the flexibility of this part of the module, and in particular, its ability to undergo larger scale pendulum-like movements as proposed for Lsd14. Nonetheless, in the absence of high-resolution structural information in the presence of the ACPs and at various stages of the processing sequence, minor conformational adjustments to favor a particular modification partner cannot be excluded. Structures of complete modules will also be critical for deepening our understanding of how the choice between intra- and intermodular options is made when this more extensive series of tailoring reactions is required.

Concerning the *trans*-AT PKSs, structural data are relatively scant. The only information relevant to this key decision point has been obtained at high-resolution by cryo-EM for a 'bimodule core' sourced from the orphan BGC11 PKS from *Bacillus brevis*.[93] The investigated construct (K3DAK4) encompassed the KS, DH and ACP domains of the third module of the system, as well as the downstream KS domain of the next module, and an intervening linker (KS_3–DH_3–ACP_3-linker–KS_4) (Fig. 16). While the three catalytic domains were visualized, the evidently highly mobile ACPs were not resolved in the electron density. The two KSs and DH domain adopt an overall linear arrangement along the KS homodimeric axis, in which the rotational orientation of the sandwiched DH varies. Furthermore, the two KSs associate laterally with the equivalent KSs of another monomer *via* previously identified laterally interacting ketosynthase sequences (LINKS),[94] giving rise to long polymeric filaments (average of 20 bimodule cores). Although not directly observed, the ACPs are proposed to occupy the reaction centers created by the other domains that encompass the intermodular junctions, with the two ACPs in a given center originating from distinct homodimers. In terms of choices, given the absence of a KR domain in module 3, no β-keto processing may occur, and thus there would be no reaction competing with transfer to module 4. However, as the product of the PKS has not been characterized, it is not possible to determine which enzymes actually act in the third module. Indeed, it cannot be excluded that a KR from a different module functions in *trans*[95–97] prior to DH_3-catalyzed dehydration (inspection of the DH_3 sequence shows both conserved catalytic residues to be present (unpublished data)), or that the DH acts on a β-hydroxy intermediate installed by the previous module.[44,98] In these cases, given the high accessibility of domains from both modules 3 and 4 to both of the ACPs, it is entirely unclear how the system would ensure that chain modification occurs prior to intermodular transacylation.

The overall view that emerges from these studies is that individual modular PKS modules are highly integrated units. Sets of domains within the modules are deeply interdependent, and their interactions are orchestrated to a substantial extent, such that the overall state of the module is primed for the consecutive steps in the chain extension cycle. The conformational transitions depend intimately on sets of evolving interdomain interfaces – likely explaining the low efficacy of engineering experiments based on domain swapping that perturbed these contacts.[99] Although as mentioned earlier, AT domain exchange has been relatively effective, these data cumulatively argue for an approach based on whole modules as units (or biobricks) in future engineering experiments.[100]

4. Single-use *vs.* iterative modules

4.1 The molecular basis for iteration

Unlike the presumed fatty acid synthase ancestor of modular PKSs that acts iteratively,[101] most PKS modules carry out only one round of chain extension/processing on any given substrate before sending it downstream. This mode of growth is referred

to as 'directional' or 'vectorial' synthesis,[102] and again depends on a key choice made following the complete round of tailoring reactions: handing off the polyketide to the downstream module rather than sending it upstream to the KS for further chain extension (Fig. 4) (here, we do not consider back transfer to the KS prior to the full processing sequences, as to our knowledge, this outcome does not occur naturally). As mentioned earlier, the module structures solved to date identify two possible barriers to internal KS re-acylation – the flexed state of the ATs following chain elongation that physically excludes back transfer, or the fact that transacylation to a downstream module kinetically outcompetes translocation to the upstream KS.

However, neither of these two models can account for the finding that multiple, naturally iterative modules exist within the context of otherwise modular PKSs. Examples of such intriguing modular/iterative systems among *cis*-AT PKSs include: ambruticin,[103] stigmatellin,[104] aureothin (**7**)/neoaureothin (**8**),[105,106] the DKxanthenes (**9**),[107] borrelidin,[108] crocacin,[109] thiolactomycin[110,111] and azalomycin/kanchanamycin,[96,112–114] while the 9-methylstreptimidone,[115] patellazole,[116] nocardiosis-associated polyketide (NOCAP),[117,118] gladiolin (**4**)/etnangien (**5**),[109,110] and lankacidin[121]/chejeunolide[122] *trans*-AT PKSs also exhibit this behavior.[111,112] Some of these repeating modules iterate in error yielding trace side products,[120,123,124] while others act a defined or even a variable number of times in programmed fashion. For example, the iterating modules of the stigmatellin and aureothin systems perform twice, that of the borrelidin PKS repeats three times, and one or more modules of the DKxanthene assembly line act two or three times (Fig. 17). Perhaps importantly, the majority of repeat-acting *cis*-AT PKS modules incorporate the same complement of domains (KS–AT–DH–KR–ACP), although two additionally include an ER domain. The iterating module of the etnangien (**5**) *trans*-AT PKS exhibits the same domain composition minus the AT (KS–DH–KR–ACP), while the 9-methylstreptimodone[115] and patellazole[116] systems incorporate an additional MT domain (KS–DH–KR–MT–ACP). The iterating modules of the ambruticin, NOCAP and lankacidin/chejeunolide assembly lines contain KR and DH domains, with those of ambruticin and NOCAP split over two subunits, among other differences (*vide infra*). Together, these observations hint that these particular domain sets are linked to iterative operation, and indeed, all four modification domains are present in bacterial monomodular iterative systems, which are the proposed ancestors of the modular PKSs (KS–AT–DH–MT–ER–KR–ACP–TE).[125,126]

In all of these cases, some mechanism has to act to kinetically impede premature transfer of the intermediate to the downstream acceptor, as well as to ensure that the active site of the iterating KS remains open (Fig. 4) (note, back transfer in the aureothin system has been shown to occur between the ACP of one monomer and the KS of the other monomer[127] – *i.e.* the same partners that are involved in chain extension[64,76]). It is intriguing that with the exception of the 9-methylstreptimidone and NOCAP systems,[115] the iterating modules uniformly operate at intersubunit junctions, suggesting that the location of the acceptor module on a distinct polypeptide also contributes to this mechanism (*vide infra*).

The factors underpining iteration have been directly investigated for a number of the *cis*-AT PKS systems, including aureothin,[98,99] DKxanthene,[100] and azalomycin/kanchanamycin.[92,106–108] Together, these studies have identified at least three elements that favor iterative behavior. The first is the naturally repetitive

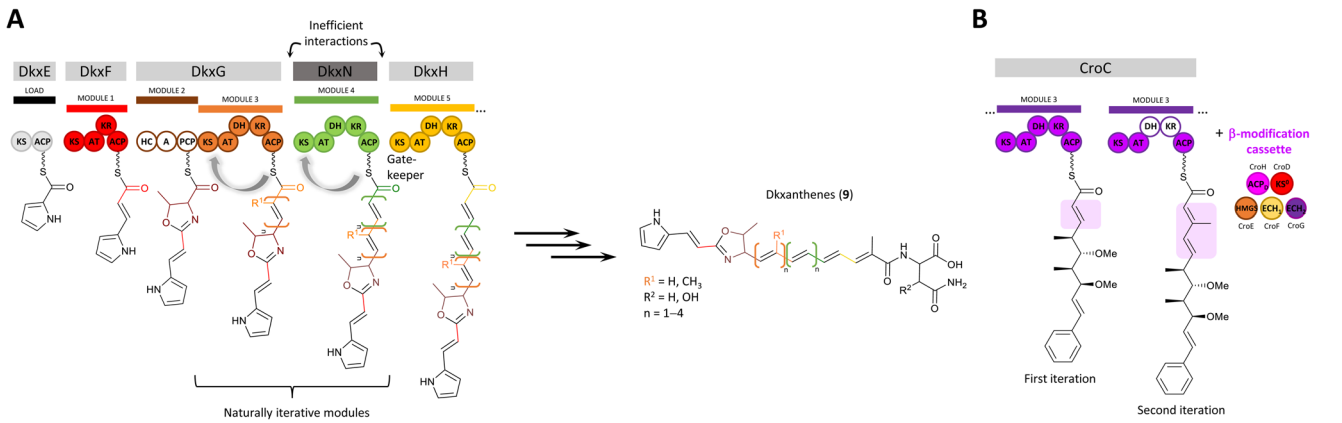

Fig. 17 Iterative behavior by modular PKS modules. (A) The family of DKxanthene polyenes (**9**) is likely to arise by variable iteration of both module 3 of DkxG and module 4 housed on DkxN. Several factors are likely to contribute to this repeat behavior, including inefficient docking interactions between DkxG/DkxN and DkxN/DkxH – a situation presumed to arise by the insertion over the course of evolution of DkxN between the two other subunits – and the substrate specificity of the acceptor KS of DkxH. The polyketide synthase (PKS) modules of the system are shown using solid spheres, and the nonribosomal peptide synthetase (NRPS) module *via* open spheres. (B) Module 3 of the crocacin modular PKS also exhibits iterative behavior, but with the outcome varying between the two rounds. In the first, both the KR and DH domains act to yield a double bond, while in the second, the in *cis* action by the KR domain (and subsequently the DH) is successfully outcompeted by a β-methylation cassette, resulting in installation of a β-methyl group. The mechanisms controlling these competing reactions are unknown. Key: KS, ketosynthase; ACP, acyl carrier protein; KR, ketoreductase; HC, heterocyclization; A, adenylation; PCP, peptidyl carrier protein; DH, dehydratase; KS^0, KS that is inactive for condensation but that catalyzes decarboxylation; HMGS, 3-hydroxy-3-methylglutaryl-synthase homologue; ECH, enoyl-CoA hydratase homologue.

character of certain modules. In other words, these modules are intrinsically inclined to catalyze back transfer of extended intermediates and re-catalysis of chain extension. As mentioned earlier, this mode-of-action evidently requires that the KS active site Cys remains vacant – *i.e.* that it is not re-acylated by the intermediate produced by the upstream module. Instead, the KS receives as substrate the newly-elongated chain attached to the ACP of its own module. In the case of both aureothin[128] and azalomycin[114] biosynthesis in which the first module iterates, the repeatedly-acting KS domains have been shown to control their own priming with starter unit, allowing them to remain open to back transfer. Presumably, the same mechanism operates for iterating KSs that instead extend polyketide intermediates.

In order for the reverse translocation to the KS to occur, the ACP must dock against the acyl transfer site of the module, a protein–protein interaction motif that is distinct from the chain extension site.[83,129] As discussed previously, the turnstile model posits that following chain extension, access to both the translocation and chain extension sites on the KS is blocked by AT flexing until the downstream ACP domain is cleared of intermediate.[76] Clearly, this mechanism is incompatible with back transfer. Thus, the AT domains in naturally iterative modules may not flex, leaving the KS domains open. In addition, the protein–protein recognition motifs of the ACPs of iterating modules may, unlike their single-use counterparts, mediate interactions with both the chain transfer and elongation sites on the KS. In principle, another contributing factor could be the intrinsic substrate specificity of the iterating KS domain. Indeed, if the KS has a strong preference for the penultimate intermediate in the iterative cycle, this would ensure that acylation of the KS with the upstream intermediate is relatively slow, and that the KS carries out repetitive condensation efficiently until the correct chain length is reached. To date, these hypotheses remain untested, and thus the molecular basis for the iterative propensity of certain modules remains unknown.

Characterization of iteration in the DKxanthene PKS identified another feature promoting iteration (Fig. 17).[107] As noted above, the majority of repeating modules operate at inter-subunit junctions, and thus transfer of intermediate to a downstream module depends on specific protein–protein interactions with the next polypeptide (see Section 5.1). Conversely, if subunit association is sub-optimal, then ACP-to-KS back transfer can kinetically outcompete chain translocation to the next module. In the DKxanthene case, sequence analysis suggests that the gene encoding a likely iterative module/subunit, DknN, was acquired by horizontal transfer. Insertion of this protein in the pathway between two other subunits may thus have resulted in inefficient interactions with both the upstream and downstream partners. This would have the consequence of slowing intermediate passage to DkxN from the preceding subunit DkxG, as well as translocation downstream to DkxH, enabling back transfer and therefore iteration by both DkxG and DkxN – behavior that correlates perfectly with the observed DKxanthene product spectrum in several hosts.[107]

As demonstrated experimentally for the aureothin[128] PKS, the final control element in module iteration is the specificity of the KS downstream of the repeatedly acting module (Fig. 17). This molecular recognition ensures that only the intermediates exhibiting the correct length/functionality can be translocated. Conversely, immature polyketide chains stall on the ACP domain of the iterating module, encouraging their return to the KS for further chain extension. Thus, overall, the control of iteration is multi-faceted: it depends on the natively iterative nature of certain modules, but is enabled by additional features such as acceptor KS gatekeeping, and relatively inefficient docking that serves to functionally isolate the modules.

Another aspect of choice in certain *cis*-AT PKS systems is the fact that different cycles of iteration are associated with use of a distinct set of processing domains. For example, module 3 of the crocacin system acts twice, but in the first round the KR and DH domains are deployed, while in the second, these in *cis* reactions are somehow out-competed by in *trans* β-methylation (see Section 6.4) (Fig. 17).[109] Similarly, in thiolactomycin assembly, the KR–DH duo functions in the first two repetitions, but is inactive during the third iteration, resulting in a β-keto group,[110] while in azalomycin biosynthesis,[114] the ER domain of the iterating module only acts in the second round. The mechanistic origin of these choices remains unresolved, but they are hypothesized to derive from the intrinsic substrate preferences of the involved variably acting domains, coupled with specificity-based gate-keeping by the downstream KS.

The characterized *trans*-AT PKSs further expand the complexity of iterative function. For example, the iterating module of the NOCAP PKS is split between two subunits and additionally exhibits an unusual domain ordering: KS/DH–ACP–KR.[119] Thus, in this case, back transfer to the KS from the ACP occurs in *trans* across an inter-polypeptide junction, and yet somehow kinetically competes with in *cis* translocation to the downstream module (perhaps due to the unusual localization of the ACP within the module).

Recent work revealed the origin of the divergent behavior by the gladiolin (**4**) and etnangien (**5**) systems (Fig. 18).[120] The responsible *trans*-AT PKSs are nearly identical, and yet etnangien contains a triene at a position corresponding to a single, fully saturated unit in gladiolin. The basis for these alternative outcomes was elucidated *via* a comprehensive series of *in vitro* experiments combined with gene deletion/complementation. Overall, it was shown that interaction between a discrete ER (GbnE) and module 5 of the gladiolin system suppresses the intrinsically iterative behavior of the module, an effect reinforced by the substrate specificity of the module 6 acceptor KS domain that favors the α,β-saturated intermediate. For an as yet unknown reason, the homologous ER of the etnangien PkS (EtnL) does not associate with its corresponding module 5, and thus the module is free to express its iterative tendency, resulting in the characteristic polyene.

The lankacidin family of *trans*-AT PKSs is arguably the most complex example of iteration identified to date. As recently determined for the chejeunolide system,[122,130,131] a single copy of the second subunit (CheC) iterates a total of five times, the first together with the upstream subunit CheA, and the last in partnership with the downstream subunit, CheF. These collaborative associations are enabled by the atypical domain

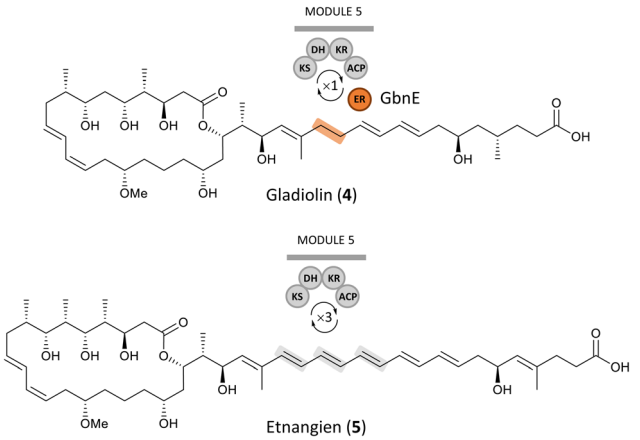

Fig. 18 Basis for the divergent synthesis of gladiolin (**4**) and etnangien (**5**) by nearly identical *trans*-AT PKS systems.[120] In the case of gladiolin, the presence of a discrete enoyl reductase (ER) suppresses the inherently iterative behavior of module 5 of the PKS, resulting in a α,β saturated intermediate during this extension cycle. By contrast, in the absence of the ER, the equivalent module of the etnangien PKS iterates three times, yielding a triene.

ordering of CheC (KR–MT–ACP–ACP–KS–DH) that allows it to form a functional model with the upstream KS of CheA and the downstream KR–ACP of CheF, and the proposed circular architecture adopted by the subunit. As with certain *cis*-AT PKSs mentioned earlier, CheC must selectively deploy its modification domain in specific cycles: the DH halts in the second chain extension cycle but otherwise functions, while the MT only acts during the third. Again, how these choices are made is entirely unclear, as well as what triggers CheC to stop repeating in order to engage with CheF in the fifth round of chain extension.

4.2 Engineering iterative systems

In terms of genetic engineering of iterative systems, to date, efforts have been relatively limited. In one notable example, module iteration was probed using assays *in vitro*, with the explicit aim of converting a single-use module (DEBS module 3) into a repeating system.[129] The specific strategy was based on the assumption that chain translocation and extension require the two involved ACPs (ACP_2 and ACP_3) to dock against distinct charged sites on the condensing wing, a hypothesis confirmed by the recent high-resolution structural analysis[83] (see Section 3). This implies, as alluded to earlier, that back transfer from the ACP that participates in condensation cannot occur due to failure to engage the chain translocation site. Thus, to transform a standard module into one capable of iteration, the authors replaced the first 22 residues of ACP_3 that encompass the N-terminus of helix αI, with that of ACP_2 that naturally interacts with KS_3 to effect intermodular transfer. Chain extension from a chemically synthesized diketide precursor (**10**) was then reconstituted *in vitro* using the mutated ACP_3 in combination with KS_3–AT_3 didomain, KR_2 (because KR_3 is naturally non-reducing) and NADPH, resulting in a tetraketide product corresponding to two rounds of chain elongation (*i.e.* classical extension and one round of iteration) (Fig. 19). The same reaction in the presence of wild-type ACP_3 only yielded triketide, demonstrating the critical contribution of the exchanged region to the additional extension cycle. The authors surmised that the system stopped at two rounds of extension due to an intrinsic size limit in the KS_3 active site that prohibited a second back transfer and/or the third extension cycle. This observation, coupled with the failure of KR_2 to act during the second round of elongation, illustrates the complexities inherent in trying to introduce defined iterative behavior into these systems, and which are only likely to be amplified in an *in vivo* context.

The most extensive work *in vivo* was carried out on the neoaureothin (**8**) *cis*-AT PKS and was geared to exploring the plasticity of the iterating module (Fig. 20).[132] The neoaureothin system is the presumed ancestor of the aureothin (**7**) PKS, as it incorporates an additional bimodular subunit, NorA′. Based on the finding that the first module (NorA) could catalyze up to four repeated condensations *in vitro*[128] (relative to the two observed *in vivo*), the authors evaluated its ability to compensate for the deletion of one or two modules present in the subsequent subunit, NorA′. While the target product aureothin (**7**) was not obtained when the entirety of bimodular NorA′ was removed (Aur* PKS), deletion of only one module (Har PKS) resulted in double and triple catalysis by NorA, giving rise, respectively, to a novel analogue, homoaureothin (**11**), and to neoaureothin (**8**) (note, for the engineered PKS to function, it was necessary to preserve the native NorA ACP/NorA′ KS interface as well as the intervening docking domains, a point that will be returned to in Section 5.1). This result demonstrated a certain size tolerance in the acceptor KS domain of NorA′, but the failure to obtain a tetraene product is again consistent with a gatekeeper role.

In future, to extend these types of approaches, it will evidently be interesting to deepen our understanding of the multiple factors known to influence repetitive action, including the molecular features that distinguish single-use from iterative modules, and the elements that modulate this functional choice: suboptimal intermodular interactions, gatekeeping by the downstream KS, and the intervention of *trans*-acting elements (*i.e.* the gladiolin ER that is recruited to and suppresses iterative function[120]).

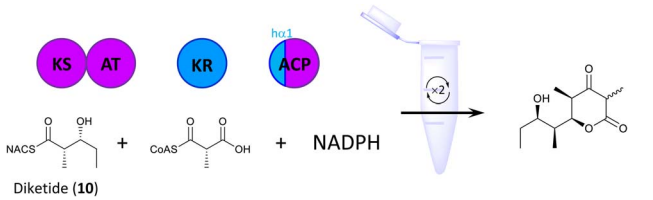

Fig. 19 Conversion of a single-use module into an iterative system *in vitro*.[129] This experiment was based on the hypothesis that the ACP mediating intermodular chain transfer to the KS and that which participates in chain extension dock to distinct sites on the KS. Thus, to engineer an iterative module, an ACP was modified to incorporate both the chain transfer site of the upstream ACP (helix α1 shown in blue), and the chain extension site (remaining purple region), allowing it to interact twice with its partner KS to yield a tetraketide product starting from a diketide chain (**10**). Key: KS, ketosynthase; AT, acyl transferase; KR, ketoreductase; ACP, acyl carrier protein.

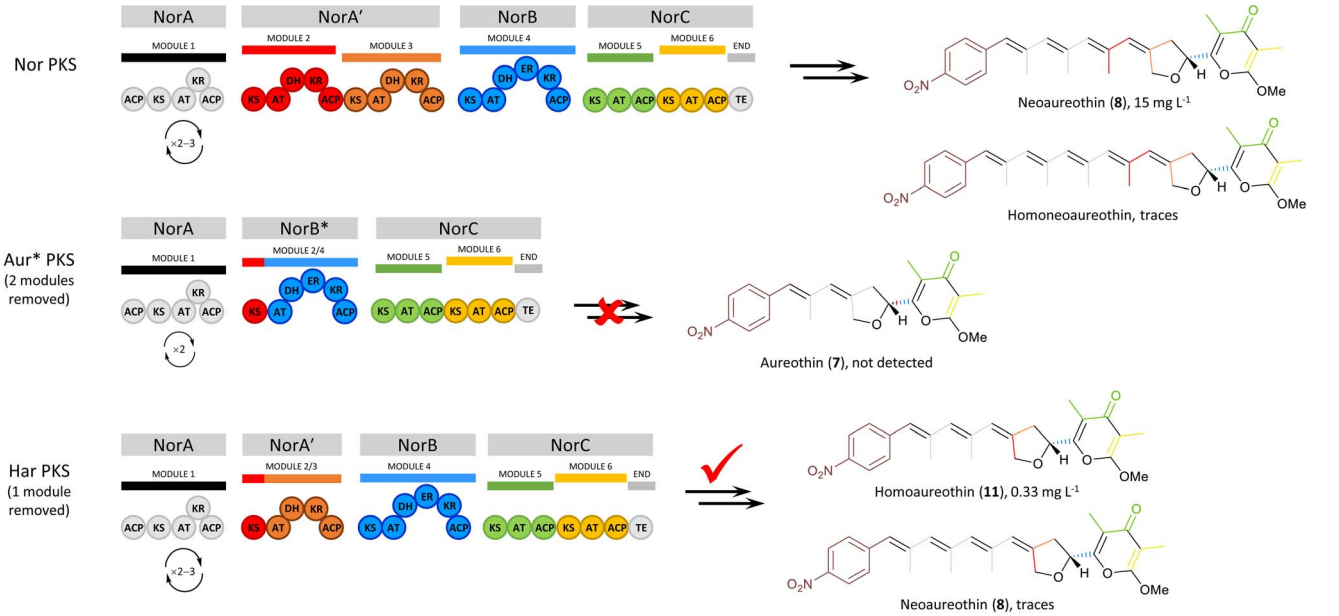

Fig. 20 Engineering of the neoaureothin (Nor) *cis*-AT PKS to test the tolerance of iterative systems.[132] In the native system, the first module acts twice (and occasionally a third time) to generate a major product neoaureothin (**8**) incorporating a tetraene, and traces of a pentaene, homoneoaureothin. An attempt was made to transform the Nor PKS into an aureothin (Aur*) PKS, by fusing the KS of NorA′ to the AT of NorB (thus removing the intervening two complete modules). However, the target aureothin (**7**) was not detected. When only one of the two modules of NorA′ was eliminated (Har PKS), the system functioned, giving rise to homoaureothin (**11**) (2× action by NorA) and traces of neoaureothin (**8**) (3× action by NorA). Note: the products are color-coded to identify the source modules. Key: ACP, acyl carrier protein; KS, ketosynthase; KR, ketoreductase; DH, dehydratase; ER, enoyl reductase; TE, thioesterase.

5. Timing and selection of downstream modular/subunit partners

5.1 Factors acting at intermodular interfaces

Having completed a full chain extension cycle and the requisite tailoring reactions, the chain extension intermediate must either be transferred to a downstream module so that additional rounds of growth can occur, or to a terminal TE/chain-releasing domain (see Section 7).[133] According to the turnstile hypothesis,[72] translocation to an acceptor module can only take place when the acceptor module has itself already handed off its product. However, this model does not actually address how the next module is chosen. Partner selection is simplified when the two modules participating in the hand-off are covalently joined within a single PKS subunit by a short linker region (*i.e.* the transfer occurs intramolecularly). Here, the choice is facilitated by the physical proximity of the two modules, and thus of the directly involved ACP_n and KS_{n+1} domains. Less obvious is what occurs when the next module resides on a distinct polypeptide, as happens at every inter-subunit junction (Fig. 4).

Work in a number of different laboratories has revealed that three distinct elements operate to ensure the fidelity of chain translocation at such junctions: so-called 'docking domains' (DDs),[19,134] compatible domain/domain interfaces (in *cis*-AT PKSs, at least),[135,136] and acceptor KS[137] gatekeeping. The fact that partner choice at intersubunit interfaces requires three levels of control implies that any one of these features in insufficient to ensure faithful chain transfer, and highlight the importance of this decision point in these systems.

Concerning DDs, they are protein–protein recognition motifs between 10–80 residues in length that are situated at the extreme C- and N-termini of both *cis*-AT and *trans*-AT PKS subunits (note: the members of matched pairs of DDs are referred to as ^{C}DD and ^{N}DD, respectively). Research in several labs has revealed the architectures of multiple DD complexes operating in both *cis*-AT and *trans*-AT systems (Fig. 21).[19,138–150] These structures will not be discussed in detail here as an in-depth review has been published previously,[134] but we will instead emphasize some of the overall lessons that have emerged in terms of understanding how DDs allow subunits to interact productively, while achieving exquisite partner selectivity.

The first important observation concerning *cis*-AT PKSs is that the majority of characterized DDs incorporate 1–3 α-helices (Fig. 21).[19,138,139,144] It may be an evolutionary happenstance that α-helices were selected for this function, because indeed β-strands/sheets also serve as protein–protein recognition motifs,[151] but evidently, the possibility of varying their lengths and surface features allows for facile modulation of their interactions. Indeed, it has recently been shown that structurally equivalent DDs (*e.g.* in terms of number and lengths of α-helices and their respective orientations) can be deployed in several different ways to form DD complexes of divergent topology.[146] These distinct architectural modes of interaction translate into multiple 'types' of DD complexes (referred to as 1a,[19] 1b,[139] 2,[138] 1-like[144] and 2-like[144]) that are intrinsically orthogonal to each other (*i.e.* one type of DD cannot productively engage another type). Thus, DD type is one layer of

organization that contributes to the choice of potential partner subunits in PKS systems. Intriguingly a single PKS may utilize the same type of DD at multiple junctions,[152] and therefore a particular ^{C}DD may be confronted with multiple ^{N}DDs with which it is in principle compatible, and *vice versa*. In such cases, selectivity appears to be achieved *via* the strategic deployment of charged interface residues that initially drive the appropriate interactions, coupled with shape complementarity in the resulting complexes.

In the *trans*-AT PKS case, although the individual DDs are again often α-helical, the structures of the DD complexes are distinct from those seen in *cis*-AT PKSs.[140,145] Characterized DD types include 4 α-helix bundle (4HB),[145] 3 α-helix bundle (3HBb)[146] and DH-docking (DHD) (Fig. 21).[140] It is also notable that selectivity is often achieved by induced folding.[145,146] That is, one or both of the DD partners exhibits strong characteristics of an intrinsically disordered region (IDR), a type of protein sequence that only adopts a stable structure in the presence of its correct partner(s).[153,154] Notably, the use of IDRs is an elegant way to ensure that docking interactions take place with high target selectivity, and in addition, with μM affinity. Indeed, all of the DDs investigated to date, whether IDRs or not, exhibit K_Ds in the low to medium μM range. Thus, docking can be categorized as a weak, transient interaction,[155] which may reflect the need for the ACP domain to interact temporarily with its downstream partner for chain transfer, while allowing it to readily dissociate in order to re-engage the domains of its own module in a subsequent catalytic cycle.

The second contributor to partner selectivity concerns the formation of productive interactions between the donor ACP domain of the upstream subunit and domains in the acceptor subunit, including the KS that is directly transacylated. Recently, a model docking interface was visualized for the first time by chemically cross-linking native DEBS ACP_2-^{C}DD to ^{N}DD-KS_3-AT_3 didomain (Fig. 22).[83] This structure revealed that ACP_2 sits within a cleft created by the KS_3-LD-AT_3 of the PKS monomer opposite to that which houses its partner KS (in the context of an intact PKS system, this cleft would therefore arise from the same subunit harboring the ACP domain). In this configuration, helix αI of ACP_2 (the same region discussed earlier in the context of iteration (Section 4.2)) forms a suite of electrostatic interactions with the AT_3 domain, as well as contacting positively charged residues on its partner KS_3 *via* the conserved Asp adjacent to its appended Ppant cofactor (Fig. 22). The interface is augmented by interactions between the ^{C}DD and the ^{N}DD, giving rise to a total buried surface area of ≈3600 Å^2. For

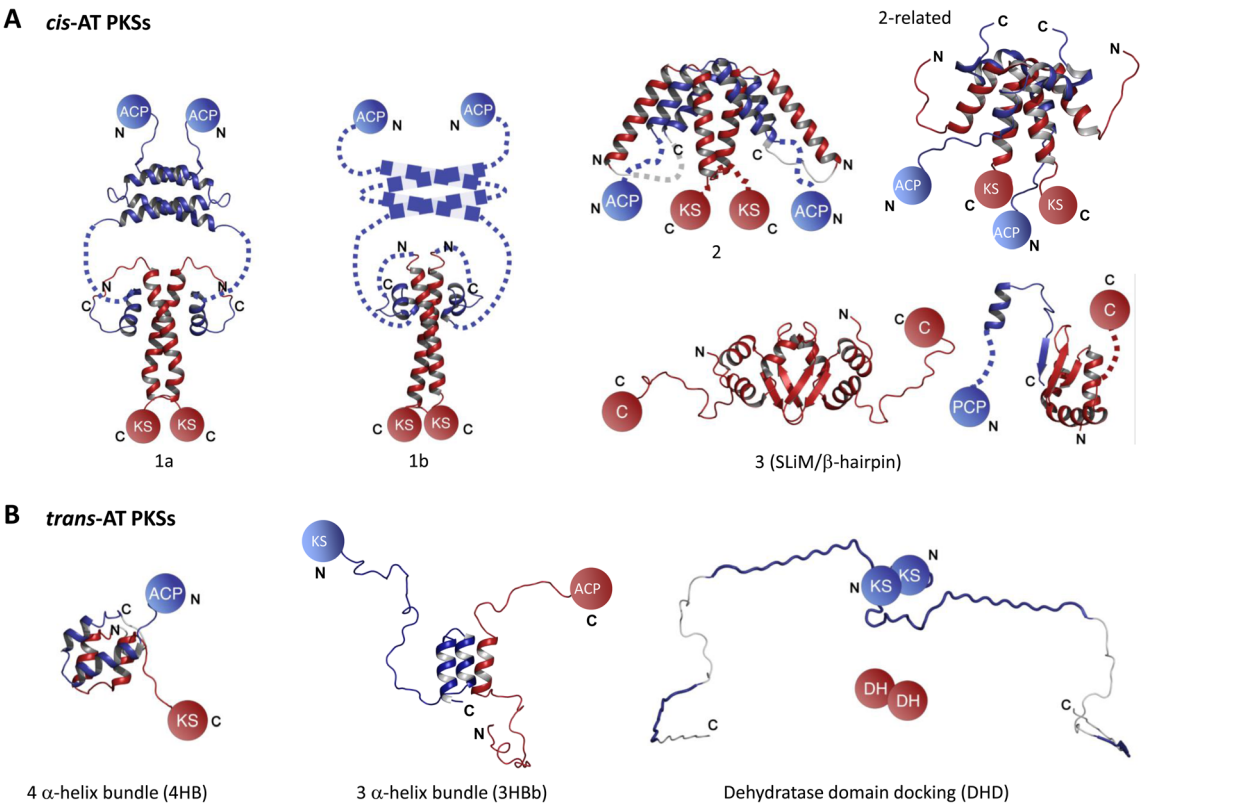

Fig. 21 Structures of docking domain complexes from *cis*-AT and *trans*-AT PKSs. (A) Docking domain types characterized from *cis*-AT PKSs.[19,138,139,142–144] Complexes identified to date are designated type 1 (that includes 1a (PDB IDs: 1PZQ, 1PZR) and 1b (PDB ID: 3F5H)), 2 (PDB IDs: 4MYY, 4MYZ), 2-related (PDB ID: 6TDN) and 3 (also designated SLiM/β-hairpin; PDB IDs: 2JUG, 6EWV). With the exception of type 3, all DD interfaces involve contacts between α-helices. (B) Docking domains identified in *trans*-AT PKSs (PDB IDs: 2N5D, 8RAJ).[140,145–147] The structures of these complexes notably differ relative to those in (A). In the case of the dehydratase docking domain (DHD), the intrinsically disordered ^{C}DD interacts directly with the downstream DH domain. In both panels, the C-terminal DD (^{C}DD) and the directly upstream domain are shown in blue, while the partner ^{N}DD and the downstream domain are represented in red.

reasons that are presently unclear, only a single ACP_2-^{C}DD was observed bound to the ^{N}DD-KS_3-AT_3 (Fig. 22), although all DD complexes solved to date are symmetrical.[19,138–150] In any case, this extensive interface explains, at least in part, the repeated finding that non-native pairings of acceptor/donor modules often result in inefficient chain transfer,[137,156] as they presumably result in sub-optimal electrostatic contacts. Nonetheless, given the varied docking modes observed for ACP domains and their AT partners (see Section 2.1), it will be important in future to solve the structures of additional chain transfer complexes.

Even less is currently known about intersubunit interactions in *trans*-AT PKSs, as no high-resolution structure of an interface has been solved to date. Such a junction from the virginiamycin system was characterized at lower resolution by SAXS,[145] but as the ACP_n and KS_{n+1} domains were too far apart (80 Å) to contact each other, even taking into account the distance spanned by the Ppant cofactor (19 Å), the visualized architecture is not compatible with chain transfer. Thus, further work should aim to visualize *trans*-AT PKS chain transfer complexes at high-resolution in order to determine whether both *cis*-AT and *trans*-AT systems have independently arrived at the same structural solution for guiding partner choice.

The final control element at intersubunit junctions is the substrate specificity of the acceptor KS domain.[156] Although the extent to which an N-terminal KS natively encounters incorrect substrates due to subunit mis-docking is unknown, in engineered systems at least, the low substrate tolerance of these domains has repeatedly been demonstrated to act as a barrier to efficient chain transfer between modified or heterologous modules.[129,135,157–164] KS domains even act as blocks to chain extension in modules downstream of engineered interfaces,[159,162] arguing that they play an important role in substrate progression even under native biosynthetic conditions.

In terms of the molecular basis for KS substrate choice, for *trans*-AT PKSs, a strong correlation exists between the phylogeny of the KS sequences and the structures of the incoming intermediate at the α and β positions – *i.e.* KSs co-evolve with the suite of modifying domains located in the upstream module that together set these two functionalities.[165–168] In the case of *cis*-AT PKSs, the relationship between KS sequence and substrate structure is less evident. Nonetheless, in the specific case of the polyol family of gigantic *cis*-AT polyketides, a convincing relationship between the KSs and the upstream processing domains was discerned, with the KSs found to form substrate-specific clades.[169] A KS sequence motif correlating with substrate structure was also identified that is located near the active site and the dimer interface,[169] and indeed, mutation of a single amino acid in this region had been shown previously to expand substrate specificity.[163]

This type of study was subsequently extended to over 120 actinomycete-derived *cis*-AT PKS systems.[170] In total, 739 KSs were categorized into 115 groups based on their native substrates, with the analysis covering the α, β and γ positions. This structure-based exercise revealed specific amino acids and more extended motifs among 32 substrate binding tunnel residues that broadly correlate with substrate choice. However, given the strong overlap in specificity determinants and therefore the limited predictability of these motifs, success at manipulating or broadening substrate selection in future will depend on identifying additional amino acids underlying specificity that evidently lie beyond the KS binding pocket.[171]

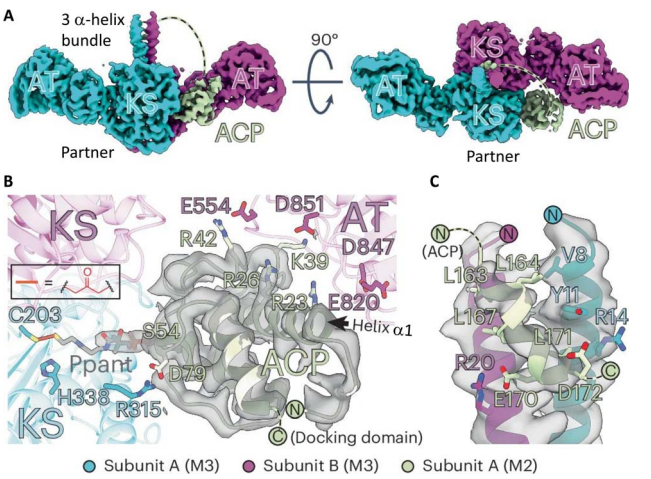

Fig. 22 Cryo-EM structure of an intermodular chain transfer complex.[83] (A) Structure of the homodimeric KS–AT fragment sourced from DEBS module 3 cross-linked to the upstream ACP_2. The ^{C}DD C-terminal to ACP_2 docks against the ^{N}DD upstream of the KS to form a 3 α-helical bundle, while ACP_2 sits in a cleft formed by the KS–LD–ACP of one monomer (pink), facilitating its interaction with its KS partner (blue). (B) Zoom on the ACP_2 docking interface showing the interface residues (color-coded according to the domain). The structure of the cross-link (red) is inset. (C) Zoom on the interface between the ^{C}DD (beige) and the coiled-coil formed by the ^{N}DD (pink and blue). Reproduced and adapted from Cogan, *et al.*, Structural basis for intermodular communication in assembly-line polyketide biosynthesis, *Nat. Chem. Biol.*, 2024, **21**, 876–882, copyright 2024, with permission from SNCSC (ref. 83).

5.2 Engineering intersubunit interactions

The intrinsic complexity and high domain interdependence of PKS modules strongly argues for employing whole modules as biosynthetic units in genetic engineering,[99,100,172,173] as a complement to domain swapping. However, this approach necessitates specifying exactly what is meant by a PKS 'module'.

For decades since their discovery,[14,174] PKS modules have been defined as the minimum set of domains necessary for all of the steps in extending the chain – selection of a building block, transfer of the chosen monomer to the ACP domain, and chain extension – and include optional processing domains. A traditional PKS module thus comprises KS–AT–modifying domains–ACP (Fig. 23). However, the finding that certain KSs (most notably those of *trans*-AT PKSs) co-migrate evolutionarily with the upstream modifying domains, prompted arguments for 'redefining' PKS modules as starting with the AT domains and terminating after the downstream KSs (*i.e.* AT–modifying domains–ACP–KS) (Fig. 23).[175,176]

While such 'alternative' modules have been shown in multiple cases to produce more active hybrid systems than the use of classical modules,[63,113,159,162,164,177–180] this set of domains

has no functional significance, as it cannot accomplish chain elongation. Indeed, classical modules are also present in the nonribosomal peptide synthetase (NRPS) systems[181,182] that evolved independently and that form productive hybrids with traditional PKS modules,[20,183] attesting to the fundamental nature of this organization (more precisely, the NRPS condensation (C), adenylation (A) and peptidyl carrier protein (PCP) domains correspond respectively to the KS, AT and ACP domains, and are present in the same order). This observation motivates the use of the term eXchange unit (XU) for different groups of PKS domains (a term borrowed from the NRPS field[184]), reflecting their particular utility in genetic engineering applications. Indeed, it has recently been shown that a third domain set that exploits a site downstream of the AT previously used for AT domain swapping (*i.e.* a module defined as: processing domains–ACP–KS–AT, 'XUA' unit (Fig. 23)), gave higher product yields than the 'alternative' module in a number of engineered PKSs.[185]

Regardless of the strategy employed (classical modules *vs.* XUs/XUAs), module-based engineering always results in one or more non-native elements in the resulting systems. In the case of traditional modules whose interactions are mediated by DDs, both the ACP/acceptor module chain-transfer interface and KS substrate specificity may be mismatched (Fig. 23). The use of alternative module XUs preserves the native KS chain preference, but the KS–AT portion of the intermodular interface is altered (see Section 5.1), and this choice of boundaries also introduces a non-natural KS/ACP chain extension interface into the resulting hybrid (Fig. 23). Finally, deploying processing domains–ACP–KS–AT XUAs conserves both KS substrate specificity and the full KS–AT docking site for the ACP, but again results in mismatching between the two chain elongation partners in the hybrid module (Fig. 23). This analysis likely explains why, overall, which set of domains works best appears to be highly system-dependent.[63,173,186]

Even when several engineering strategies are tried, proofreading by KS domains/modules[137] downstream of the modified interface continues to be a substantial block to obtaining wild type titers of polyketide derivatives. For example, in recent work, XUs (AT–modifying domains–ACP–KS) were employed to generate 5 triketide, 25 tetraketide, and 125 pentaketide synthases derived from the pikromycin (Pik) PKS.[162] For each XU added, the obtained yields dropped precipitously (6.4% of pentatketide synthases were functional *vs.* 32% of tetraketide and 60% of triketides synthases), an effect attributed largely to KS gatekeeping. In another case, a new interface between the fourth and ninth PKS subunits of the stambomycin PKS was created by docking domain engineering.[159] Although the structure of the intermediate generated by the fourth subunit corresponded to that produced by the eighth subunit at both the α and β positions and thus to the substrate normally seen by the ninth subunit, no productive chain transfer occurred. An

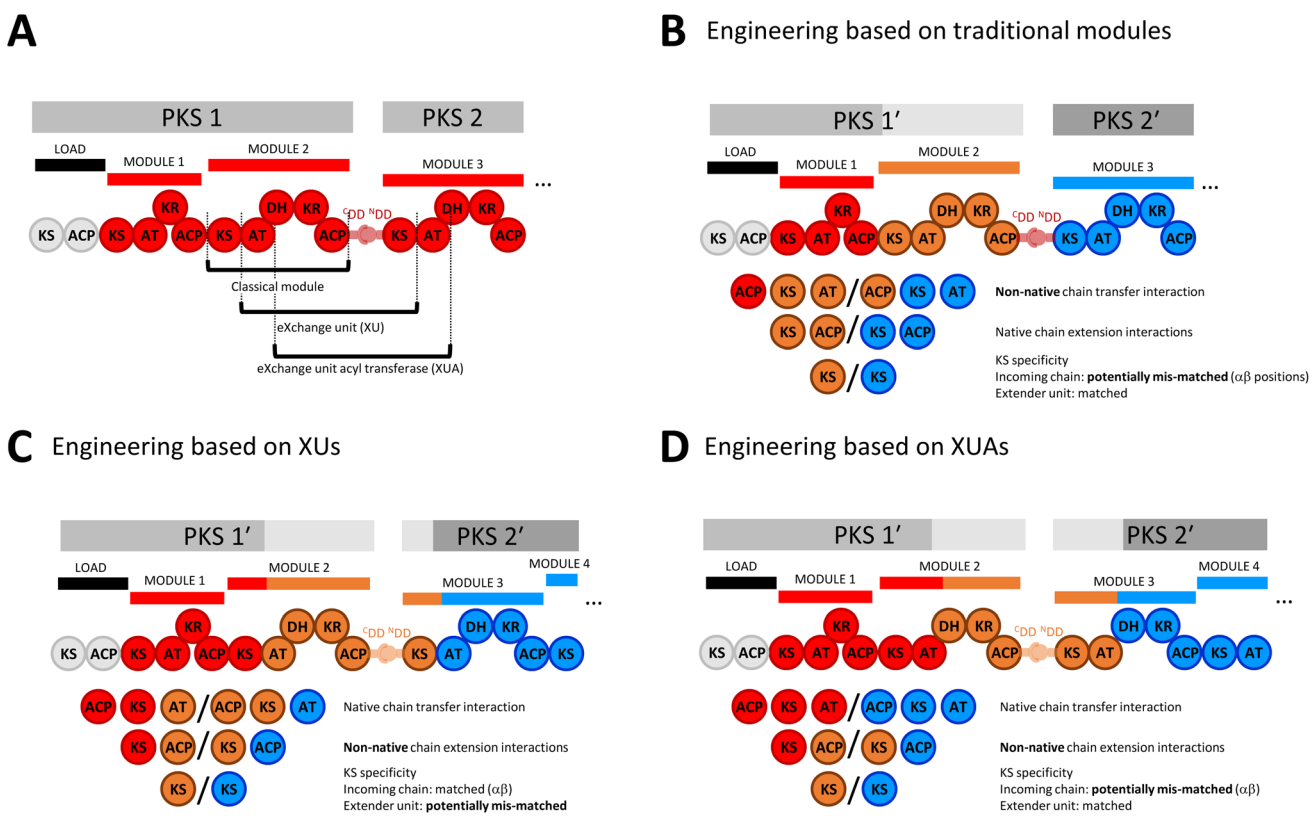

Fig. 23 Domain sets used for engineering. (A) Domain sets comprising classical modules, exchange units (XUs) and exchange units acyl transferase (XUAs). Both the XUs and the XUAs encompass domains that span the junction between classical modules (which in this case involves matched docking domains). Native and non-native interdomain interactions arising from exchanges based on (B) classical modules, (C) XUs, and (D) XUAs. Non-native contacts, which are engendered by all three strategies, are highlighted in bold.

attempt to improve the now heterologous interface between modules 13 and 21 by site-directed mutagenesis of ACP_{13} (ref. 129) resulted in the target product, but in low amounts at best. Moving the fusion point both into and beyond the downstream KS domain (*i.e.* the use of XUs) resulted in the desired metabolites, but the yields were still 8-fold reduced relative to the parental titers.

These results, when taken together, are sobering as they mean that typical KS domains are sensitive not only to the functionality at the acyl terminus of their substrates, but that located farther down the chain (at the γ and δ positions, and maybe beyond). Thus, the interdependence among modular PKS components discussed earlier extends beyond the domains of individual modules to the relationships between the modules themselves. Surmounting this barrier will depend either on identifying modules of varying domain composition that incorporate intrinsically promiscuous KSs (although this could theoretically result in a mixture of products due to recognition of incompletely processed intermediates generated by the upstream module (see Section 3)), or increasing our ability to redirect/broaden KS specificity by site-directed mutagenesis. In this context, it has previously been noted that the rapamycin[163,187,188] and mycolactone[189] systems are particularly attractive sources for such modules, as many KSs from each PKS exhibit high mutual sequence homology and yet operate on intermediates of diverse functionality and chain lengths. In the meantime, the highest success rates for module-based engineering have been achieved with mini-PKS systems (1–3 modules) giving rise to diverse small molecule products.[63,172,185,186]

Another relevant approach is to create new intersubunit interactions by insertion of docking domains between modules within existing PKS polypeptides. This experiment was first reported for the Pik PKS, and produced a functional synthase, but yields 2-fold lower than wild type.[190] More recently, the same strategy was applied to three different systems (butenylspinosyn (Bus) (Fig. 24), avermectin (AveA) and epothilone (Epo)), resulting in 7–13-fold increases in product titers.[191] This effect was traced to improved translation of modules encoded farther along the large PKS genes, arguing for the interest of keeping the genes small. As a substantial number of identified PKS subunits[192] incorporate two or more modules and thus can in theory be split in this way, this approach represents a potentially facile way to boost polyketide production.

Nonetheless, such experiments, as well as attempts to productively engineer combinations of classical modules, necessitate that the correct choice of DDs to deploy is made.[193] More specifically, the introduced pairs of DDs must be orthogonal to those that are native to the system, thereby ensuring that the natural order of subunits is maintained. As noted recently,[193] consideration of the available DD literature offers clear guidelines for how such elements can be chosen. In cases where it's possible on the basis of sequence analysis to classify the native DDs in the target system into types (see Section 5.1), a different type can be used in engineering to enforce orthogonality. Alternatively, and in particular where sequence-based classification is unreliable,[144] intersubunit communication can instead be mediated by wholly non-natural and therefore intrinsically orthogonal elements. These include synthetic protein–protein interaction motifs called SYNZIPs[194,195] and semi-synthetic SpyTag/SpyCatcher complexes, and potentially also split inteins.[172] Engineered PKS systems incorporating SYNZIPs and SpyTag/SpyCatcher pairs have been shown to function,[157,196] even though this strategy replaces non-covalent DD interactions with covalent (or effectively covalent in the case of SYNZIPs) associations, but care must be taken in deploying these elements, as they introduce structural constraints into the hybrid proteins.[196] A final approach is to transplant a complete set of compatible DDs from another PKS into the modified system[193] (for an in-depth discussion of module-based synthetic biology strategies, the reader is referred to ref. 172).

One last engineering application worth mentioning that goes beyond modular polyketide biosynthesis is the utilization of PKS-derived DDs to increase the effective concentration of enzymes acting sequentially in metabolic pathways.[148,197] Such an approach depends on the intrinsic modularity of DDs – *i.e.* their ability to retain their protein–protein recognition functions when separated from their natural PKS contexts.[147,198] This strategy has been applied, for example, to the astaxanthin terpenoid biosynthetic pathway in *E. coli*. Use of DD pairs derived from the DEBS and rapamycin (RAPS) PKSs whose compatibility was initially demonstrated *in vitro* using Bio-Layer Interferometry, resulting in an overall 2.4-fold increase in product yield.[197] A similar effect was observed by fusing 4HB-type DDs from the macrolactin *trans*-AT PKS to two consecutive enzymes involved in flavonoid biosynthesis (4-coumarate-CoA ligase and chalcone synthase), resulting in an approximately 2-fold increase in naringenin production.[148] While the application of DDs to metabolic engineering is interesting and could potentially be applied to more extended series of enzymes, the observed impacts have to date been modest. This result potentially reflects the fact that boosted proximity cannot overcome the strongly rate-limiting effects of one or more pathway enzymes.

6. Choices by *trans*-acting enzymes

In *cis*-AT PKS systems, but much more commonly during biosynthesis by *trans*-AT PKSs, certain enzymes act in *trans* to modify the chain extension intermediates attached to the ACP domains. In the next section we discuss the choices presented to these catalysts (Fig. 4), what is presently understood about how they are negotiated, and the consequent prospects for engineering these types of modifications. At the outset we would note that *trans*-acting enzymes would appear to be at a kinetic disadvantage relative to *cis*-located enzymes competing for the same substrates. Therefore, the *trans vs. cis* context could act as a universal control element in PKS systems.

6.1 A *cis*-located ER that acts in cross-modular fashion in *trans*

During azalomycin (Azl) assembly by a *cis*-AT PKS, an ER domain housed within the first iterating module (termed

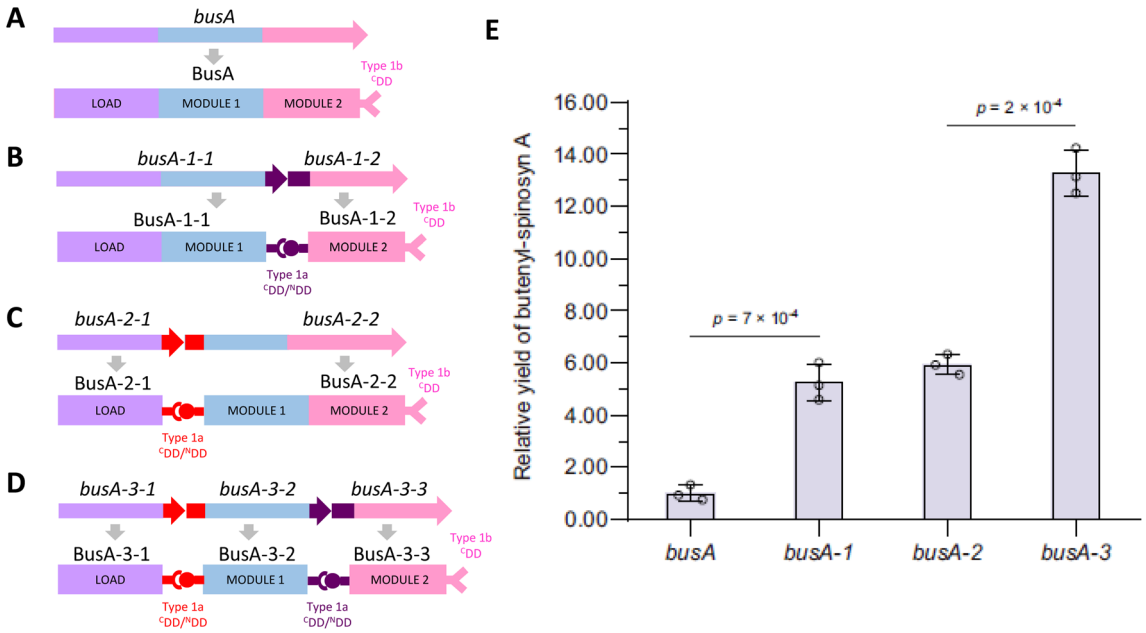

Fig. 24 Subunit-splitting strategy applied to the butenyl-spinosyn (Bus) system.[191] (A) Encoding gene and protein organization of the native BusA subunit. The multienzyme comprises a loading and two chain extension modules, and a type 1b ^{C}DD.[193] BusA was split into two ((B) and (C)) fragments by insertion of matched pairs of type 1a DDs, and (D) into three pieces by simultaneous incorporation of the two pairs. (E) Relative yields of butenyl-spinosyn produced by the wild-type system *vs.* full assembly lines incorporating the various split versions of BusA. An overall maximum increase of *ca.* 13-fold was obtained *via* this approach. Reproduced and adapted with permission from Liu, *et al.*, Improving polyketide biosynthesis by rescuing the translation of truncated mRNAs into functional polyketide synthase subunits, *Nat. Commun.*, 2025, **16**, 774 (ref. 191).

module 1/2, see Section 4.1) acts in cross-modular fashion on the intermediate tethered to the ACP of module 3 (ACP_3) (Fig. 25).[96,113] Such in *trans* action is necessary, as module 3 generates an α,β-saturated intermediate yet lacks an ER domain. Although in principle this reaction could have occurred *via* back-transfer of the chain from module 3 to module 1/2, it was shown unexpectedly to involve a direct intermodular interaction between the $ER_{1/2}$ domain and the substrate attached to ACP_3. Inspection of the full azalomycin PKS shows that potential α,β-unsaturated substrates for $ER_{1/2}$ are also generated by the 5th, 6th, 19th and 20th modules in the system, meaning that $ER_{1/2}$ somehow chooses the correct chain to modify. This tight selection is particularly intriguing, given the close evolutionary relationship between modules 3 and 6. Evidently, the fact that AzlA and AzlB act successively in the biosynthesis could help limit the interaction between $ER_{1/2}$ and ACPs in other subunits, but it was demonstrated directly that the docking interaction between AzlA and AzlB did not fully explain the cross-module activity.[113]

This enigma was at least partially resolved by exchanging all of the module 6 domains (KS, AT, KR, DH, ACP) individually and in combination with their counterparts from module 3, and testing the ability of the hybrid modules to function downstream of module 1/2 (for this, the intervening subunit AzlB was deleted, and the ^{N}DD of AzlB module 3 was grafted to the N-terminus of module 6).[113] Collectively, these experiments resulted in the anticipated C_6-truncated derivatives of azalomycin, and showed that the presence of KS_3 is necessary but not entirely sufficient for in *trans* recruitment of $ER_{1/2}$, as a small proportion of metabolite remained unreduced. The absence of in *trans* enoylreduction in the 6th, 19th and 20th modules presumably reflects the lack of an analogous $ER_{1/2}$ docking site on these KS domains. Heterologous ERs could only substitute poorly for $ER_{1/2}$, although a close homologue of $ER_{1/2}$ from the polaramycin PKS could function in its place, demonstrating that the particular characteristics of this domain also play a role in substrate selection. The correct sequence of events is reinforced by proof-reading by KS_4 of the downstream module that is selective for the fully-reduced tetraketide. It remains to be seen in future if cross-module enoyl reduction can be introduced into additional positions in azalomycin or alternative PKSs *via* an analogous engineering approach.

6.2 The *trans*-acting ER GbnE that modulates modular/iterative behavior

As discussed earlier (see Section 4.1), recruitment of a *trans*-acting ER GbnE to module 5 of the gladiolin PKS suppresses the intrinsically iterative propensity of the module (Fig. 18). While GbnE also acts during biosynthesis by split-module 10 (KS/DH–KR–ACP), it ignores the α,β-unsaturated intermediates generated by modules 3, 4 and 16. The available data demonstrate that the ACPs play a role in recruiting the ER domain, but the full molecular basis for the ER's choice of where to act remains to be elucidated – information which will be necessary to leverage this type of chemistry for future engineering.

6.3 *Trans*-acting α-hydroxylases in *trans*-AT PKS systems

On-line α-hydroxylation of polyketide intermediates has been investigated to date for both the oocydin[199] and mupirocin[200]

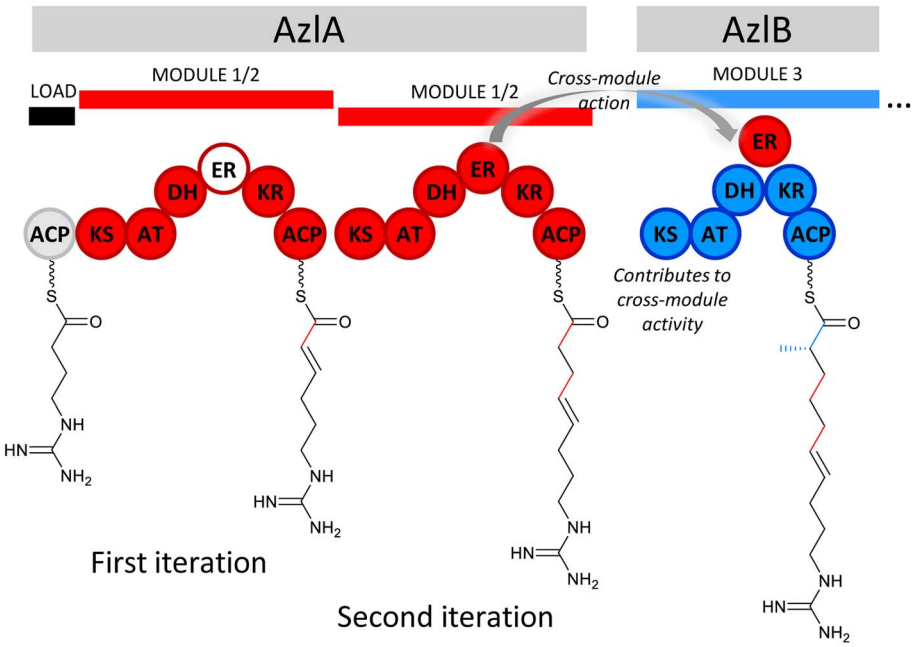

Fig. 25 Cross-module action by an ER during biosynthesis of azalomycin (Azl) by a *cis*-AT PKS.[96,113] The ER domain of the iterative module 1/2 of AzlA acts in *trans* during chain extension by module 3, an activity which depends on the KS of module 3. Neither the molecular basis for the choice of cross-modular partner by the ER, nor for its failure to act during the first round of chain extension by module 1/2, is fully understood at present. The portions of the intermediates added by each module are color-coded. Key: ACP, acyl carrier protein; KS, ketosynthase; AT, acyl transferase; DH, dehydratase; ER, enoyl reductase (the blank circle indicates that this domain is inactive); KR, ketoreductase.

trans-AT PKSs. In the oocydin case, this reaction is catalyzed by a flavin-dependent monooxygenase homologue OocM and takes place at the intersubunit junction between module 4 (OocL KS_4–KR_4–ACP_4) and module 5 (OocN KS^0_5–ACP_5; KS^0 designates a non-elongating KS domain that catalyzes intermodular chain translocation) (Fig. 26).[199] Using assays *in vitro* with recombinant proteins, it was established that OocM acts on a β-keto substrate tethered to ACP_4, and thus α-hydroxylation precedes KR_4-catalyzed ketoreduction at the β-position of the chain. Intriguingly, the reaction also requires KS^0_5 that is proposed to act as a protein–protein adaptor between OocM and ACP_4. Such a specificity mediator would appear to be necessary, as each and every cycle of PKS-catalyzed chain extension results in a potential β-keto substrate for OocM (but *vide infra*). Nonetheless, ACP_4 is also likely to contribute to choice, as ACPs that serve as supports for α-hydroxylation in *trans*-AT PKSs form a clade distinct from non-targeted ACPs in phylogenetic trees.[199]

As in the oocydin system, α-hydroxylation during mupirocin biosynthesis occurs at the interface between two subunits, MmpD and MmpA, and the two flanking modules have the same domain composition (MmpD KS_4–KR_4–ACP_4; MmpA KS^0–ACP_1).[200] The associated monooxygenase MupA exhibits high sequence homology to OocM, and its activity is ACP-dependent. Surprisingly, however, assay *in vitro* containing only MmpA and purified MmpD ACP_4 modified with native β-keto substrate, showed conversion to the α-hydroxylated product. Thus, unlike the oocydin case, the downstream KS^0 is not necessary for the reaction, although it was observed to interact weakly with ACP_4. Together, these results suggest that the enzymatic components required for regiospecific α-hydroxylation may vary among *trans*-AT PKS systems.

Interestingly, MupA was also observed to bind MmpA ACP_1 by NMR, but this interaction did not result in α-hydroxylation of the same tethered substrate. Given that both the ACP_1 and the polyketide chain (when attached to MmpD ACP_4) are recognized by MupA, it is difficult to understand why the reaction also did not take place on ACP_1 *in vitro*. Resolving this question may not be critical, however, as if α-hydroxylation happens systematically *in vivo* on ACP_4 prior to chain transfer to ACP_1, the situation will never arise naturally. Indeed, ACP_1 has been proposed to play an alternative role by contributing to MupA docking against the MmpD/MmpA bimodular complex. According to this model, the resulting network of protein–protein interactions would strongly enforce the choice of α-hydroxylation substrate, potentially explaining the observed lack of proofreading by the module 5 KS^0 against non-α-hydroxylated chains.[200] An intriguing remaining question is how the KRs in both the oocydin and mupirocin systems decide when to act – *i.e.* after the α-hydroxylation – but the timing could in principle be controlled by an important contribution of the α-hydroxyl group itself to substrate recognition.

Concerning prospects for introducing α-hydroxylation rationally, the most straightforward approach would be to transplant existing monooxygenase/biomodule pairs into alternative contexts, coupled with matching of KS substrate specificity in the downstream module.[165,168] Screening of the sequences of the other ACPs in the PKS for their phylogenetic relatedness to the target ACP could provide reassurance that no side-reactions would occur.

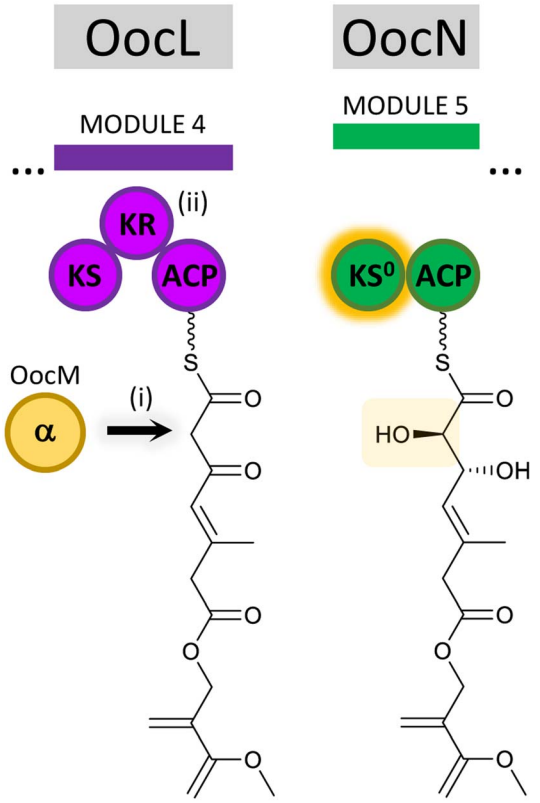

Fig. 26 α-Hydroxylation during biosynthesis of oocydin by a *trans*-AT PKS.[199] (i) The hydroxylase OocM is capable of selecting the β-keto intermediate tethered to the acyl carrier protein (ACP) of module 4. This transformation requires the presence of the KS^0 of the downstream module, for reasons that remain to be elucidated. (ii) Reduction by the ketoreductase (KR) domain at the β-position follows the α-hydroxylation. Key: KS, ketosynthase; KS^0, KS that is inactive for condensation but which catalyzes intermodular chain transfer.

6.4 Choices by β-methylation cassettes

β-Modification is a common on-line transformation in *trans*-AT PKSs, but less common in *cis*-AT systems.[201–203] In its simplest form, the initial β-keto group arising from chain extension is converted into a β-methyl (of either *endo* or *exo* stereochemistry), but many variations on this theme have been identified.[201] Despite the chemical simplicity of a β-methyl, its installation requires a 'cassette' of five different proteins: a discrete ACP (called the ACP donor, ACP_D), a condensation-inactive but decarboxylating KS homologue (KS^0), a 3-hydroxy-3-methylglutaryl-CoA synthase (HMGS) homologue, and two enoyl-CoA hydratase homologues (ECH1 and ECH2). In the first step, malonate attached to ACP_D is decarboxylated by KS^0 (Fig. 27). The resulting acetate is then employed as a nucleophile by the HMGS for an aldol addition to the β-keto group of a specific intermediate tethered to an ACP of the PKS (referred to as ACP acceptor, ACP_A). Hydrolytic release of ACP_D yields an HMG-like species linked to ACP_A. This intermediate is then successively dehydrated and decarboxylated by the two ECH enzymes to give the final β-methyl group (when the ECH2 acts in *trans*, the resulting β-methyl is *endo*, but when the ECH2 is present in *cis* within the targeted module, the *exo* stereochemistry is typically generated (*vide infra*)). As with α-hydroxylation, what is particularly intriguing about this type of modification is that a potential β-keto substrate for the cassette enzymes is produced by every round of chain extension. Aberrant β-methylation/modification is likely to be efficiently outcompeted, however, in modules comprising *cis*-acting tailoring domains, but would appear to be an issue for minimal modules where such domains are absent (*i.e.* KS–AT–ACP in *cis*-AT PKSs and KS–ACP in *trans*-AT PKSs).

Before discussing the current state of knowledge, it is interesting to consider how β-methylation/modification could be controlled, at least in theory (Fig. 28). Among the three enzymes (HMGS, ECH1 and ECH2) that directly interact with the polyketide linked to ACP_A, only the HMGS has to be specific – if it doesn't act, no substrate is available for the two ECHs. Concerning reaction specificity, in principle, the enzymes could recognize exclusively the particular acyl-ACP they need to target, and thus essentially disregard all of the other ACP-bound chains. Alternatively, the enzymes might instead only bind certain polyketide chains while ignoring the ACP, with potentially a focus on the acyl terminus that distinguishes each intermediate from the preceding one. Or, the enzymes might leverage both of these specificity determinants to make their choice. As mentioned earlier, β-modification has to occur competitively with alternative chemistries, minimally interaction with a downstream module (for chain extension) or domain (*e.g.* a TE). A corollary to this situation is that if the competing reactions occur fast enough, they could effectively suppress the in-*trans* reactions, even though the substrate could be recognized by the cassette enzymes.

To date, these issues have been investigated directly for the virginiamycin (Vir) M[204,205] (Fig. 29) and weishanmycin (Wsm)[206] (Fig. 30) hybrid *trans*-AT PKS-NRPS systems, and the bongkrekic acid *trans*-AT PKS.[207] In the case of virginiamycin M, historically, β-methylation was thought to occur exclusively during chain extension by module 5.[208–210] In common with a large number of modules targeted for this modification, Vir module 5 incorporates a curious tandem of ACP domains (ACP_{5a} and ACP_{5b}). Prior structural analysis of module 5 by small-angle X-ray scattering (SAXS)[211] had revealed that the two ACPs are divergently positioned relative to the central KS homodimer, with ACP_{5a} tucked up against the KS, and ACP_{5b} located at the end of a relatively extended linker region. ACP_{5a} was thus hypothesized to participate in chain extension, leaving ACP_{5b} to serve as the site for the subsequent β-methylation reaction series.[211] Indeed, when binary interactions between the β-methylation cassette members VirC (HMGS), VirD (ECH1) and VirE (ECH2) and the two ACPs were studied by Trp fluorescence quenching, only ACP_{5b} was recognized by the enzymes.[204] This finding is consistent with the division-of-labor model for the biosynthesis suggested by the module 5 SAXS structure. The holo form of ACP_{5b} was preferred relative to the apo protein, demonstrating a role for the Ppant group in recognition, but the addition of acyl groups mimicking the natural substrates of the three enzymes did not result in improved affinities. Interestingly, while a control ACP from module 6 was not recognized, as expected, by any of the β-methylation cassette, 'control' ACP_7 was

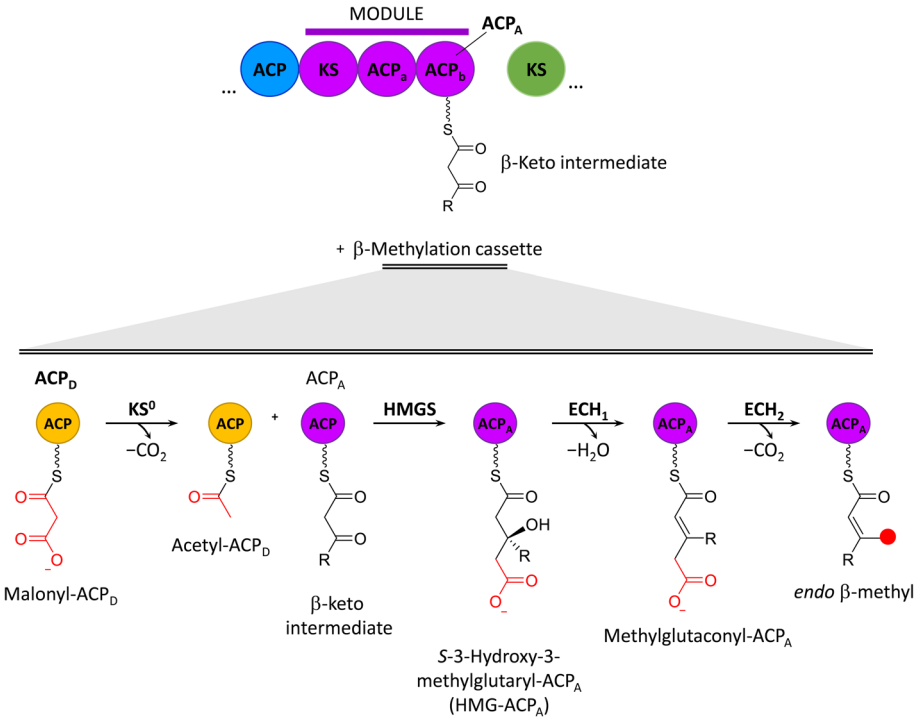

Fig. 27 Enzymology of β-methylation in polyketide biosynthesis. The reaction series starts with KS^0-catalyzed decarboxylation of malonate linked to an ACP donor (ACP_D). The resulting acetate is then used as a nucleophile by a 3-hydroxy-3-methylglutaryl-synthase (HMGS) homologue to attack a β-keto intermediate attached to an ACP acceptor (ACP_A). Curiously, many modules targeted for β-modification incorporate repeated ACP domains (typically two, but sometimes three). Following hydrolysis of ACP_D, the resulting HMG-like intermediate is successively dehydrated and decarboxylated by enoyl-CoA hydratase homologues (ECH1 and ECH2, respectively) to yield the final β-methyl. Throughout the figure, the structure arising from the acetate nucleophile is shown in red. Importantly from the perspective of choices, only the HMGS and ECH enzymes interact directly with chains tethered to the PKS, and so participate in substrate selection. Key: KS^0, KS that is inactive for condensation but that catalyzes decarboxylation.

bound with essentially wild type affinity – a point that we will return to later.

In an effort to understand how ACP_{5a} and ACP_{5b} are distinguished, the NMR structures of all of the ACPs were solved in both their apo and holo forms,[204] and compared to the structures of a tandem of ACPs from the mupirocin system that also participate in β-methylation.[210] This analysis revealed no evident architectural differences among the domains that could

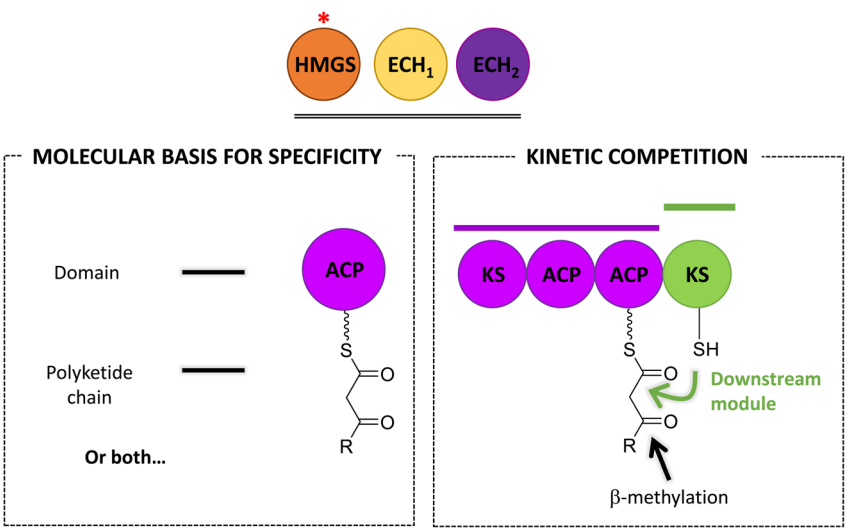

Fig. 28 Models for control of β-modification. In principle, substrate selection by the HMGS, the first enzyme to act in the series, would be sufficient, as in the absence of its catalysis, no substrate is present for the ECHs. Selection could derive from recognition of the ACP domain which carries the substrate, the structural features of the substrate itself, or both of these elements. Another potential contributor to specificity is the ability of alternative reactions (*e.g.* chain transfer to a downstream module) to kinetically outcompete β-modification.

underpin their selective recognition. Nonetheless, the domains were observed to exhibit distinct electrostatic surface features,[212] with that of recognized ACP_7 globally more similar to ACP_{5b} than those of unrecognized ACP_{5a} and ACP_6 (Fig. 29). Thus, one contribution to the choice of partner is likely to be the compatibility between these surfaces and those of their partners that helps to drive correct interactions while disfavoring others.

Two further elements were revealed by the crystallographic analysis of VirC and VirD.[204] In the case of VirD, the structure of the isolated enzyme was solved, as well as its complex with ACP_{5b} (Fig. 29). Comparison of the two showed that the interaction with ACP_{5b} provokes increased structuration of VirD, and the effect is even more dramatic for VirC (Fig. 29).[205] As this architectural remodeling is likely fundamental to the catalysis, it serves as a further layer of specificity in these interactions (*i.e.* interactions with non-native acyl-ACP partners will not yield active enzymes[205]). Finally, inspection of the ACP_{5b}/VirD complex revealed that the precise positioning of the Ppant arm at the interdomain interface also constitutes a specificity determinant (Fig. 29). Unexpectedly, the ACP_{5b} residues at the VirD interface are, with one exception, strictly conserved with their ACP_{5a} equivalents, and therefore do not contribute to substrate selection.

Concerning the surprising finding that ACP_7 is recognized, it allowed assessment of the extent to which competition with chain extension by downstream NRPS module 8 suppresses β-methylation – in other words, did the interaction observed *in vitro* between ACP_7 and the β-methylation cassette translate into catalysis *in vivo*, or was it effectively inhibited by interaction with module 8. For this, extracts of a virginiamycin M producer were scrutinized for metabolites corresponding to a second β-methylation event in module 7. This analysis indeed revealed a previously undetected doubly-methylated product, at a yield approximately 0.5% that of virginiamycin M. Thus, despite the multiple levels of regulation at work, the choice of β-methylation substrate is imperfect. This apparent biosynthetic error likely reflects the limited number of specificity determinants available on the small ACP domains to enable their discrimination, as well as the fact that all polyketide intermediates in any given system share substructures.

In the case of weishanmycin (Wsm) A1 (**12**),[213] a member of the leinamycin family,[206,214–216] β-modification occurs in both

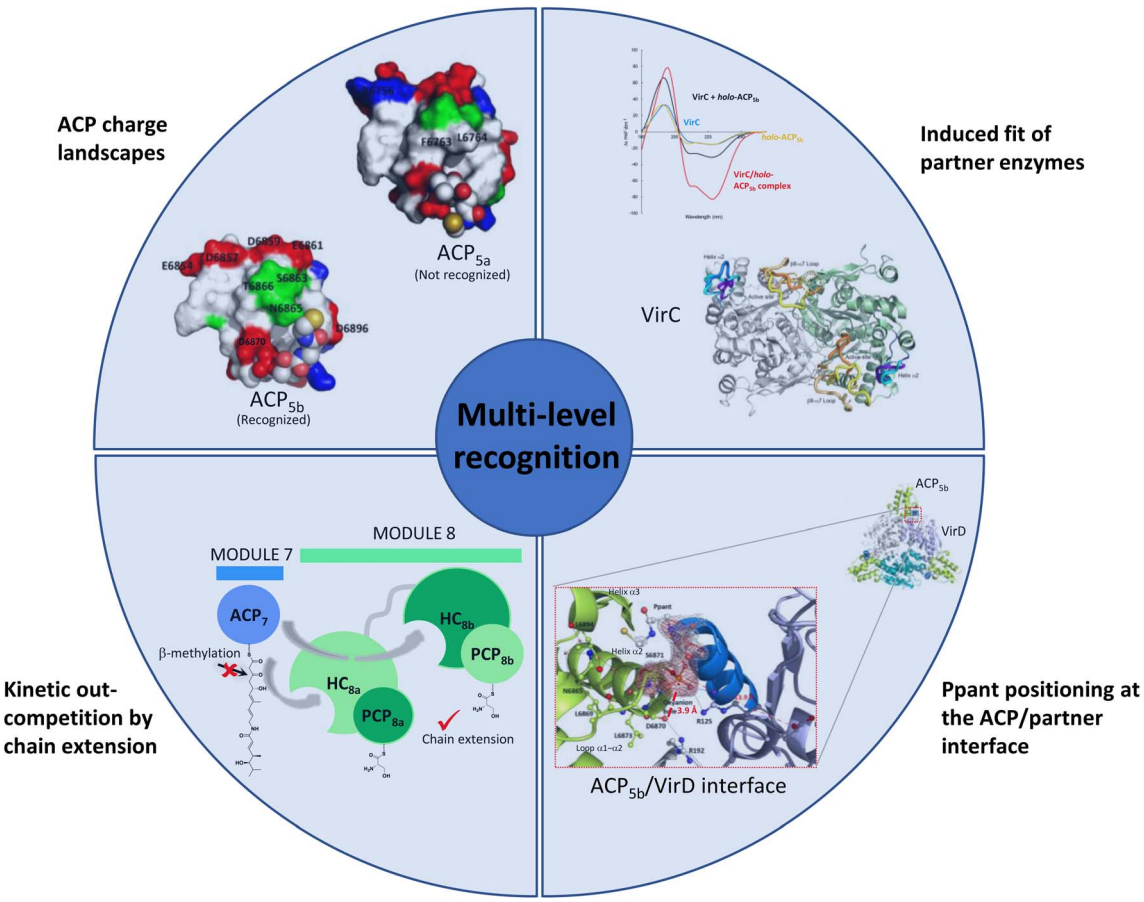

Fig. 29 Complex control of β-methylation during biosynthesis of virginiamycin M.[204,205] Substrate choice derives from multiple features of the acyl-ACP target and the partner enzymes. These include the electrostatic surface landscape of the ACP that drives the interactions with the cassette members, the positioning of the Ppant cofactor at the resulting interprotein interface, and induced structuration of the partner enzymes that occurs upon complex formation. A final element of control is provided in module 7, at least, by kinetic out-competition with chain extension catalyzed by the nonribosomal peptide synthetase (NRPS) module 8. Key: ACP, acyl carrier protein; HC, heterocyclization; PCP, peptidyl carrier protein.

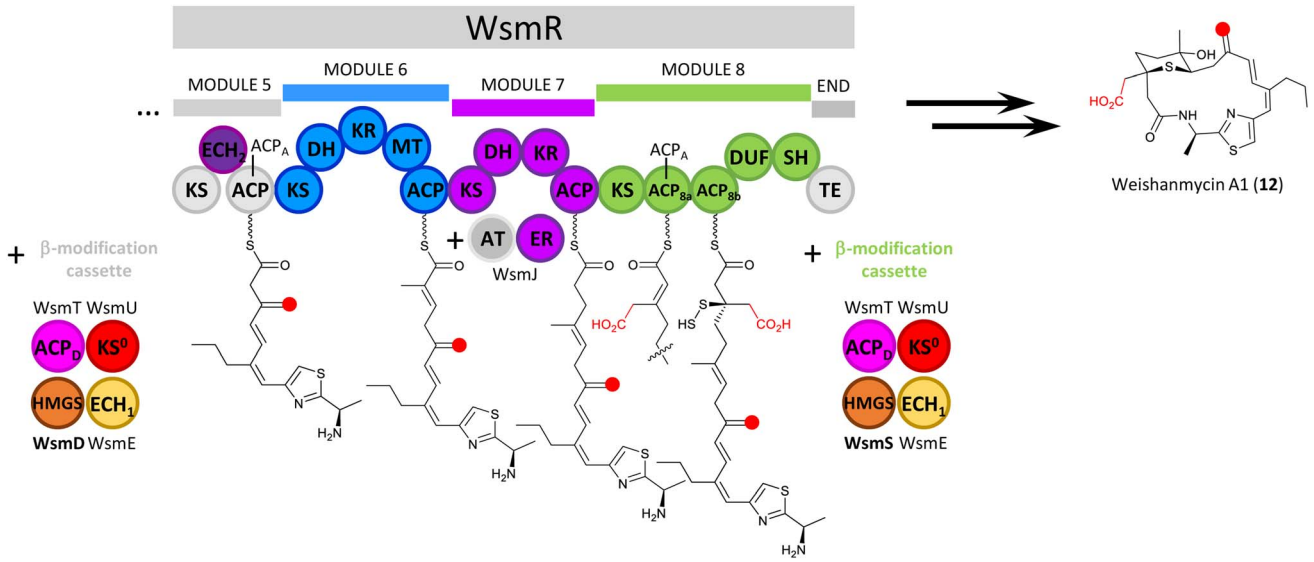

Fig. 30 Control of β-modification during weishanmycin A1 (**12**) assembly.[213,217] β-Modification reaction series occur during chain assembly by both modules 5 and 8. In the case of module 5 that incorporates a single ACP_A domain, the cassette of four proteins WmsD, WsmE, WsmU and WsmT is complemented by an ECH2 domain located within the module. As is classical for this in *cis* localization of the ECH2, the resulting β-methyl group exhibits an *exo* configuration. The structures arising from the acetate nucleophile are shown in red. WmsE, WsmT and WsmU also act following chain extension by module 8, but alongside a second HMGS homologue, WmsS, specific for ACP_{8a} (ACP_A). This reaction series stops at the carboxyl-branching stage, consistent with the absence of either an in *trans*- or in *cis*-acting ECH2. Instead, once the intermediate is transferred to ACP_{8b}, it undergoes sulfur insertion carried out by the thiocysteine lyase (SH) domain. The basis for ACP selection in this system appears to derive from unique characteristics of the ACP interface residues. Key: KS, ketosynthase; ECH2, enoyl-CoA hydratase homologue 2; ACP_A, acceptor acyl carrier protein; ACP_D, donor acyl carrier protein; KS^0, KS that is inactive for condensation but which catalyzes decarboxylation; HMGS, 3-hydroxy-3-methylglutaryl synthase homologue; ECH1, enoyl-CoA hydratase homologue 1; DH, dehydratase; KR, ketoreductase; MT, C-methyl transferase; AT, acyl transferase; ER, enoyl reductase; DUF, domain of unknown function; TE, thioesterase.

modules 5 and 8 (Fig. 30). The β-modification cassette includes five members: WsmT (ACP_D), WsmU (KS^0), WsmD and WsmS (HMGSs), and WsmE (ECH1). Module 8 exhibits an unusual domain composition (KS_8–ACP_{8a}–ACP_{8b}–DUF_8–SH_8–TE; DUF, domain of unknown function; SH, thiocysteine lyase[217]) and results in installation of both a carboxylated β-branch and a hydropersulfide moiety, while module 5 yields an *exo*-β-methylene (Fig. 30). The reaction series stops at the carboxylate in module 8 because the cassette lacks an ECH2 homologue. Correspondingly, module 5 incorporates a *cis*-acting ECH2, as expected for the observed *exo* β-methyl stereochemistry, and, unusually, a single ACP domain.[201]

Via in vitro studies using recombinant components of module 8 coupled with intact protein mass spectrometry, the two ACP domains were demonstrated to play distinct roles in the pathway, consistent with their low sequence identity and divergent phylogenetic clading. Specifically, chain extension and all steps required to introduce the carboxylated β-branch occur on ACP_{8a} (*i.e.* it is the sole ACP_A). As clarified by subsequent work with the guangnanmycin system,[217] ACP_{8b} serves as the site for the sulfur insertion, and also interacts with the downstream TE leading to chain release. The same functional inequivalence was also demonstrated *in vitro* for the tandem ACPs of the leinamycin system. Each of the Wsm HMGSs interacts with the ACP of only one module (WsmD with module 5 and WsmS with module 8) – a situation previously observed with the myxovirescin[218] and largimycin *trans*-AT PKS[215] – but partner with the same ACP_D (WsmT). The sole ECH1 (WsmE) functions with both modules, and as with VirD,[204] substrate recognition involves the Ppant cofactor.[213]

In terms of what allows the cassette enzymes to distinguish between the weishanmycin tandem ACPs, conserved differences in amino acids were noted between the presumed ACP_As and non-ACP_As within the whole leinamycin family.[165,167] The Wsm ACP residues lie in regions of the domains (*e.g.* the N-termini of helices αII and αIII) previously implicated in forming productive interfaces with both HMGS and ECH1.[204,212] Supporting their contribution to recognition, mutation of Wsm ACP_{8a} at five of these amino acids completely abolished its ability to function as an ACP_A with HMGS WsmS.[213] This situation contrasts with the Vir PKS, in which, as mentioned earlier, the residues deployed by ACP_{5b} at the interface with ECH1 VirD are almost uniformly conserved with those of ACP_{5a}, and therefore do not contribute to specificity. Thus, the basis for partner choice differs between the Vir and Wsm systems, consistent with the idea that *trans*-AT PKSs evolved independently.[167] Future work on the Wsm system could address the multiple aspects that remain uncertain, including how the HMGSs select between the two available ACP_As, whether the ECH1 is also ACP_A-specific (as shown for VirD) and the amino acid basis for this choice, and the features that ensure that the β-carboxylate and hydropersulfide are installed prior to chain termination.

Concerning bongkrekic acid,[219] β-modification again takes place during chain extension by two modules (1 and 11), but

halts in module 11 after catalysis by the first ECH1 BonH, leaving a C3 carboxymethyl in place.[207] Both modules incorporate a tandem of ACP domains, with those of module 1 exhibiting high (98%) mutual sequence identity. Consistent with the results obtained on the Vir system, the second ACP of module 11 (equivalent to Vir ACP_{5b}) was shown to be recognized by the HMGS homologue BonG, but whether the first ACP is also a substrate could not be ascertained due to protein insolubility. BonG was also active in catalysis with one of the two essentially identical ACPs of module 1, but not towards two ACP_As sourced from other modules, demonstrating that ACP recognition contributes to reaction control. Surprisingly, however, both the ECHs BonH and BonI were able to efficiently and consecutively process HMG tethered to the ACP_As of modules 1 and 11 to yield *endo*-β-methyl groups. Therefore, the lack of BonI activity in module 11 does not arise from anti-selection against the component ACP domains. Indeed, both BonH and BonI were even able to transform substrates tethered to non-ACP_As. Thus, in this system, substrate choice is simpler than in the Vir case, as the HMGS BonG acts as a gatekeeper for the entire β-modification series (in agreement with the minimal control mechanism proposed earlier, Fig. 28). As for how BonI-catalyzed decarboxylation is suppressed in module 11, multiple mechanisms have been proposed, including kinetic out-competition by transfer in *cis* to an unusual downstream off-loading KS domain, inefficient recognition by BonI of the longer and more structurally complex module 11 substrate, or steric exclusion by the overall modular architecture (although why action by BonG and BonH wouldn't similarly be impeded is unclear).[207]

An extended form of β-modification also occurs during biosynthesis of curacin A (Cur) (**13**) by a *cis*-AT system incorporating 10 PKS modules and a single NRPS module (Fig. 31).[220] Unusually, the targeted module in subunit CurA includes a triplet of consecutive ACP domains with high mutual sequence identity (*ca.* 90%), while the ECH2 of the β-modification cassette is fused to PKS subunit CurF, but gives rise to *endo* stereochemistry. The tandem ACPs do not interact with each other, and unlike in the Vir and Wsm systems, they are functionally equivalent, with all three serving as sites for the classical reactions of β-methylation.[221] Further distinguishing the pathways, the intermediates are processed in *cis* by both a halogenase[222] (Cur Hal) that acts prior to ECHs 1 and 2, and in *trans* by an ER located downstream of the ECH2 that catalyzes the final step in the tailoring series (Fig. 31).[223] Although nothing is presently known about how the β-modification cassette members recognize the three ACP_As, anti-selection against the remaining ACPs of the system is presumably aided by kinetic out-competition by the *cis*-acting modification domains. Concerning the order of the β-modification reactions in CurA, while ECH1 and ECH2 were shown to recognize the non-chlorinated intermediate, the chlorinated form reacted *ca.* 4-fold faster, at least *in vitro*.[223] These data argue that the sequence of reactions (Cur Hal → ECH1/ECH2 (ref. 224)) is kinetically controlled, while catalysis by the ER is constrained to the last step by the reactivity requirements of its substrate.

Overall, the available data reveal substantial intersystem variability in how β-modification is regulated. In both the weishanmycin A1 (**12**) and bongkrekic acid cases, particular sequence features underlie selective acyl-ACP_A recognition by the cassette enzymes, whereas in the virginiamycin case, it is rather overall electrostatics and Ppant positioning that are decisive. In the Vir case, all three cassette members act as gatekeepers, with their activity dependent on interaction with the correct partners *via* induced structuration, while only BonG of the bongkrekic acid PKS exhibits strong selectivity. Finally, as demonstrated for the virginiamycin, bongkrekic acid and curacin (**13**) pathways, the relative kinetics of competing reactions also serve as key determinants of the overall chemical outcome.

Together, these results have clear implications for any attempts to engineer β-modification. Given the inconsistency and complexity of the control mechanisms, it will be difficult to redirect the cassette enzymes to alternative acyl-ACP intermediates by site-directed mutagenesis alone. A more promising strategy will instead be to manipulate whole modules targeted for this chemistry along with their accompanying cassettes, using the module engineering approaches discussed earlier (Section 5). Strategic omission of the ECH2 or both enzymes from classical cassettes would result in an abbreviated series of reactions, giving rise, respectively, to β-carboxyl or HMG-like intermediates. Such a strategy would of course need to be coupled with identification of KS domains/modules of suitably broad substrate specificity to act downstream. In parallel, it has been shown at least for the Vir *trans*-AT PKS, that intrinsic weaknesses in β-modification control points can lead to natural derivatives of known compounds.[204] It will be interesting in future to intervene actively in these decision points to boost the yields of such minor products, but also for the potential generation of new-to-nature structures.

6.5 Reprogramming of a PKS *via* an in-*trans* ECH^Q-TE didomain

The aurantinins (ARTs) (**3**) of *Bacillus subtilis* arise from the coupling of long and short polyketide chains[225] synthesized, respectively, by 12 extension module (Art10–Art15) and bimodular (Art16) *trans*-AT PKSs (Fig. 32).[226] In addition to the classical ECH1 (Art19) and ECH2 (Art20) involved in β-branching during the third and fourth chain extension cycles, the pathway includes Art21, a protein encompassing a second ECH2 homologue fused to a proof-reading thioesterase[227] (TEII) (see Section 7.1). In the Art21 ECH2, a Gln residue replaces the catalytic His necessary for typical decarboxylation activity, suggestive of an alternative role (it is therefore referred to as ECH^Q). Remarkably, when Art21 ECH^Q is inactivated by removal of *ca.* 190 internal residues, the PKS system shifts from assembling tetracyclic aurantinins (**3**) to making bicyclic auritriacids (ATAs (**14**), Fig. 32), but expression of the ECH^Q in *trans* restores ART production.

The ART metabolites are proposed to arise from homologation of the intermediate generated by Art10–Art15 with a succinyl-derived octaketide assembled by Art16 (Fig. 32), catalyzed by the terminal KS_{13} domain of Art15.[228] In the presence of inactive Art21 ECH^Q, subunits Art10–Art16 instead

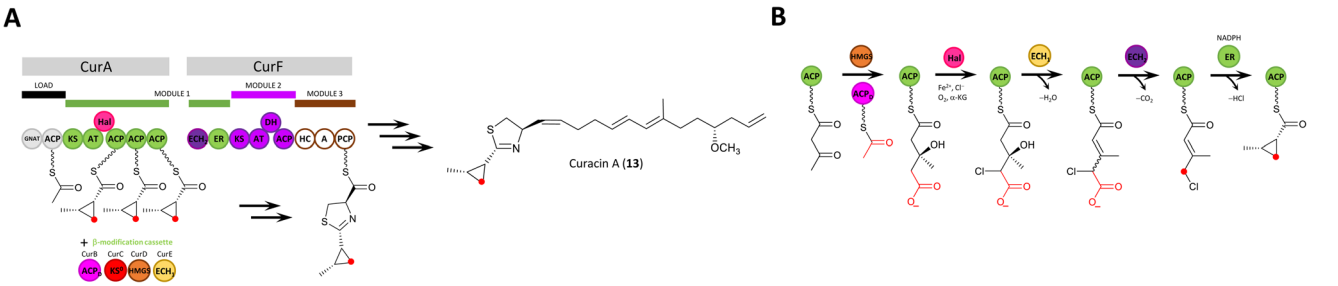

Fig. 31 β-Modification reactions during curacin A (Cur) (**13**) biosynthesis.[220] (A) The targeted module of CurA incorporates a triplet of ACPs, all of which are thought to act as ACP acceptors (ACP_As). The β-modification cassette includes in addition to KS^0, ACP_D, HMGS and ECH1 (CurB–CurE), a halogenase (Hal) domain, and ECH2 and ER domains located in the next subunit CurF. The structures arising from the acetate nucleophile are shown in red. (B) Sequences of transformations leading to installation of the curacin cyclopropane. Following the HMGS-catalyzed reaction, the intermediate is chlorinated at the γ position, and then undergoes ECH1-catalyzed dehydration to yield the vinyl chloride. The final steps involve ECH2-mediated decarboxylation, followed by an unusual ER-catalyzed cyclopropanation reaction. Nothing is known to date concerning substrate choice in this system. Key: GNAT, Gcn5-related *N*-acetyl transferase; ACP, acyl carrier protein; KS, ketosynthase, AT, acyl transferase; Hal, halogenase; ACP_D, donor acyl carrier protein; KS^0, KS that is inactive for condensation but which catalyzes decarboxylation; HMGS, 3-hydroxy-3-methylglutaryl synthase homologue; ECH1, enoyl-CoA hydratase homologue 1; ECH2, enoyl-CoA hydratase homologue 2; ER, enoyl reductase; DH, dehydratase; HC, heterocyclization; A, adenylation; PCP, peptidyl carrier protein.

function in co-linear fashion, an organization presumably enabled by the splitting of the thirteenth module between Art15 and Art16 (Art15···KS-DH*/MT-ACP···Art16; * indicates an inactive and presumed structural domain). This change in function has multiple consequences. Specifically, instead of Art16 deploying succinate as a substrate to initiate the biosynthesis, the hybrid Art15/Art16 module extends the polyketide chain assembled by Art10–Art15 with malonate, and in addition, the terminal module of Art16 now functions as a site of HCS-catalyzed modification of the intermediate (Fig. 32). Interestingly, when a Gln → His mutant of Art21 ECH^Q was expressed in *trans*, it failed to rescue the ART phenotype, while

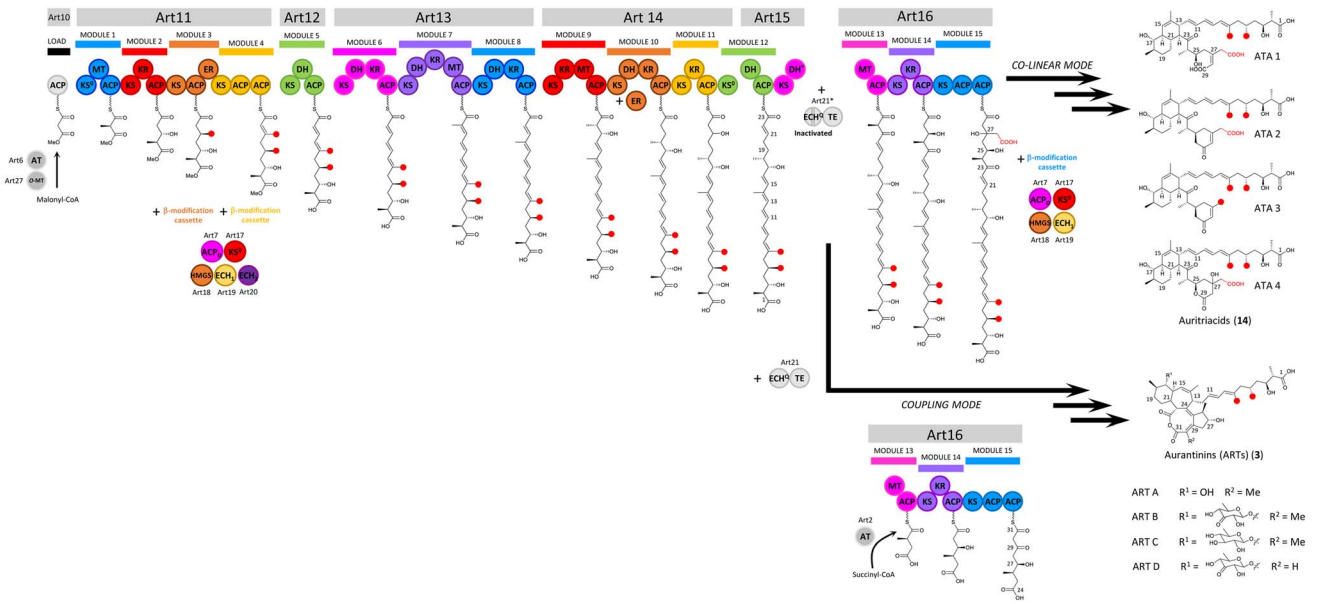

Fig. 32 ECH^Q-provoked pathway divergence.[226,228] The aurantinins (**3**) (ARTs) are biosynthesized *via* coupling of a chain generated by Art10–Art15 with that produced by Art16. Two modules of Art11 are targeted for β-methylation by a common cassette of enzymes (Art7, Art17–Art20). The structures arising from the acetate nucleophile are shown in red. The homologation involves a non-decarboxylating ECH2 homologue (designated ECH^Q), as internal truncation of ECH^Q leads to a linear mode of biosynthesis in which Art15 forms a functional module with the N-terminal didomain of Art16. In this alternative case, the terminal module 15 is also targeted by the β-methylation cassette, a reaction which stops at the HMG-like stage. Downstream modifications then yield diverse auritriacids (ATAs) 1–4 (**14**). Key: *O*-MT, *O*-methyltransferase; AT, acyl transferase; ACP, acyl carrier protein; KS^0, KS that is inactive for condensation but which catalyzes chain transfer or decarboxylation; MT, C-methyltransferase; KS, ketosynthase; KR, ketoreductase; ER, enoyl reductase; ACP_D, donor acyl carrier protein; HMGS, 3-hydroxy-3-methyl-glutaryl synthase homologue; ECH1, enoyl-CoA hydratase homologue 1; ECH2, enoyl-CoA hydratase homologue 2; DH, dehydratase; DH^0, inactive DH which may play a structural role in the module; ECH^Q, ECH2 homologue in which the catalytic His has been replaced by Gln; TE, thioesterase.

at the same time disabling ATA synthesis. This result suggested a critical role for the Gln itself in productive pathway rerouting, and more specifically the amide side chain, as substitution of Gln with Asn also restored ART biosynthesis, albeit partially.

The means by which Art21 ECH^{Q} directs the biosynthetic choices made in the thirteenth module has not yet been conclusively defined, although a role in influencing which extender unit is selected was ruled out by studies *in vitro*.[226] In the meantime, homologues of Art21 ECH^{Q}-TE were identified in a number of other *trans*-AT PKS systems. When expressed in the Art21 ECH^{Q} mutant strain, certain of these proteins were able to rescue ART production to the exclusion of ATA, while others resulted in a mixture of the two metabolites. The latter result is consistent with the simultaneous existence of both routes within the cell, and thus incomplete pathway control. Overall, these data raise the intriguing possibility that manipulating the equivalent decision points by engineering in the other identified systems could lead to additional, novel polyketide analogues, particularly as some of the systems also incorporate NRPS modules.[226]

7. Terminating PKS-based assembly

Chain liberation is the final step in biosynthesis by modular PKSs. This reaction occurs in two contexts – release of both incorrect building blocks and polyketide intermediates by type II (stand-alone) thioesterases (TEIIs), or the final assembly line products by a diversity of enzymes.[223] Choices made in this critical step can influence both the stage at which the products are disengaged from the PKS, as well as the chemistry employed, thus profoundly shaping the final structures (Fig. 4).

7.1 Enzymology of chain liberation by type II enzymes

The need for on-line proof-reading is hypothesized to arise, for example, from rare KS-catalyzed decarboxylation of ACP-tethered extender units in the absence of chain extension intermediate (Fig. 33). This aberrant activity would result, following protonation of the enolate, in a starter-like unit attached to the ACP domain – *e.g.* acetate derived from malonate, propionate from methylmalonate, or butyrate from ethylmalonate. Additional potential sources of error include the use of an acyl-CoA instead of CoA during PPTase-catalyzed phosphopantetheinylation resulting in acylated ACP,[229] building block mis-selection by loading modules (*e.g.* choosing acetate instead of propionate),[230–232] and mis-processing of chain extension intermediates (Fig. 33).[227] Overall, these situations would provoke stalling of the chains on the ACP domains due to gatekeeping by the downstream acceptor domains, and thus wasteful blockage of the assembly process.

Studies on a number of *cis*-AT PKS systems[227,230,233–235] have identified dedicated, most typically stand-alone TEIIs whose role it is to hydrolytically release these abnormal substrates. Indeed, when these enzymes are selectively inactivated, product yields usually drop,[232,236,237] but the activity can be restored by complementation.[231,236,238] Catalysis occurs in two steps, with initial self-acylation of the TE active Ser (*i.e.* *trans*-acylation), followed by water-mediated release.[239] Concerning their mode of substrate recognition, all of the assayed TEIIs exhibit substrate promiscuity *in vitro*, although general trends in specificity are nonetheless apparent.[227,235] Based on structural studies of RifR from the rifamycin pathway,[234] tolerance towards the acyl chains is thought to arise from the intrinsic flexibility of several elements that define the size and shape of the substrate-binding chamber (*i.e.* lid α-helices and lid loop), as well as its generally hydrophilic surfaces features (Fig. 34). These observations coupled with the low rate of TEII hydrolytic activity relative to PKS-catalyzed chain extension,[234] lead to a 'hedge-trimmer' model[227] of proof-reading, in which long-enough lived and physically accessible substrates are systematically removed from any ACPs in a given system. Whether or not this chemistry requires formation of specific ACP/TEII complexes is not presently known. Anti-selection against comparable acyl-CoAs that would be energetically wasteful is proposed to be controlled by their inability to provoke the opening of the TEII lid that is necessary for substrate entry.[234]

In *trans*-AT PKS systems, this corrective release activity is furnished by acyl hydrolase (AH) enzymes, potentially because type II TEs fail to recognize the constituent ACPs.[240] AHs are AT homologues that act like TEIIs, but with a clear preference for short-chain acyl-ACPs.[240,241] They are present in *trans*-AT pathways as either mono-functional proteins, or in the context of didomain or tridomain multienzymes (*e.g.* AH–AT and AH–AT–ER). Based on analysis of the bacillaene AH (PksD), the preference for shorter chains is proposed to derive from the partially hydrophilic nature of the binding channel, which disfavors interaction with long, hydrophobic substrates (Fig. 35).[241] The corresponding acyl-CoAs are repelled by a high proportion of negatively-charged residues in the channel leading to the active site. Concerning anti-selection against correct ACP-appended extender units, an Arg residue that normally forms a salt bridge with the carboxylate moiety in AT domains is replaced in AHs by smaller, uncharged residues, thereby removing a critical binding interaction. Recognition of at least a subset of ACPs in the model bacillaene system is mediated by a series of residues proximal to the Ppant attachment site that contact amino acids at the entrance to the substrate binding pocket of the AH,[240] with recognition largely deriving from shape complementarity. Globally, the interaction is similar to that observed in the *trans*-AT/ACP binary complexes (see Section 2.2), consistent with the idea that AHs are ATs repurposed by evolution to carry out hydrolysis. Nonetheless, AlphaFold multimer predicts at least four distinct binding modes, leaving opening the possibility that recognition of ACP targets is plastic. Such a feature would obviously favor broad-specificity chain clearing.

7.2 Chain release by type I TEs and other enzymes

Once PKS-catalyzed chain assembly is completed, the products require release from the multienzymes prior to further processing by so-called post-PKS enzymes.[242] This chemistry is typically, although not always,[243] carried out in *cis* by the C-terminal domain of the final PKS subunit. Type I TEs are the most commonly observed enzymes, but a number of other

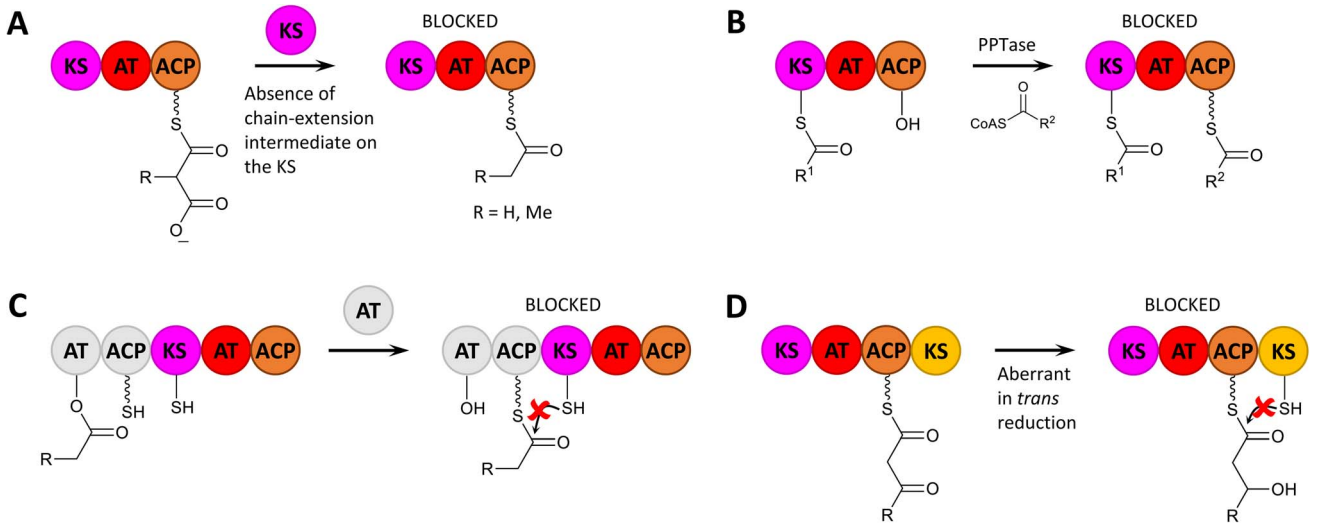

Fig. 33 Biosynthetic events that result in blocking of PKS modules, and thus the need for proof-reading activities. (A) KS-catalyzed decarboxylation of an extender unit in the absence of chain extension intermediate on the KS active site. (B) Transfer by a phosphopantetheinyl transferase (PPTase) of acyl-phosphopantetheine onto the ACP instead of phosphopantetheine. (C) Selection by the AT of the loading module of the incorrect substrate, resulting in its non-acceptance by the gatekeeping ketosynthase domain of module 1. (D) Aberrant activity of a *trans*-acting enzyme (in this case, a ketoreductase), resulting in a substrate that is not accepted by the downstream, gatekeeping KS domain. Key: KS, ketosynthase; AT, acyl transferase; ACP, acyl carrier protein.

possibilities have been reported, including but not limited to reductive release by thioester reductase (TR) domains and or conversely by oxidative mechanisms,[239] as well as hydrolysis catalyzed by type II TEs (Fig. 4)[239,244–246] (note: a comprehensive catalogue of release modes was provided in ref. 239). Even the chemistry effected by type I TEs is variable: macrolactonization and hydrolysis are the most common outcomes, but several families of TEs have been reported to accomplish coupled hydrolysis, decarboxylation and β-hydroxyl elimination to generate terminal dienes.[247,248] In all of these cases, the choice of chemistry is typically determined by the type of enzyme involved, but non-native substrate features can provoke alternative outcomes (*i.e.* hydrolysis by TE domains instead of macrocyclization).[249–251]

As TEs are the prototypical off-loading domains, the rest of this discussion will focus on these catalysts. A number of TEs sourced from *cis*-AT PKSs have been characterized at high-resolution by X-ray crystallography in order to understand the basis both for initial substrate choice and the mode of chain release (Fig. 36).[17,247,249,252–256] Collectively, this work has revealed that the homodimeric TEs exhibit an overall canonical αβ-hydrolase tertiary fold, but atypically, contain a substrate channel that traverses the entire protein and whose structural integrity depends on dimer formation. The classical catalytic triad (Ser, His, Asp) is located at the mid-point of this tunnel, consistent with a mechanism in which the ACP-tethered substrate enters from one side of the domain and the liberated cyclic or linear product exits from the other. The mobile 'lid' seen in type II PKSs that corresponds to helices α6 and α7 in dimeric type I TEs, is maintained in a closed position by the dimeric interface. Indeed, no major conformational change has been detected upon substrate binding, and thus both substrate

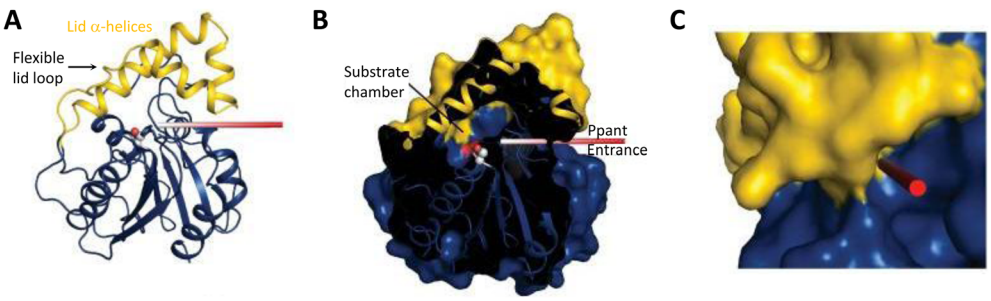

Fig. 34 Structural analysis of the type II TE RifR from the rifamycin pathway.[234] (A) Ribbon diagram of the TE, with the active site Ser shown in spheres (PDB ID: 3FLA, 3FLB). The lid region in its closed conformation is shown in yellow, and the linker between the lid and β-strand 5 is indicated. The flexibility of these elements underpins the structural plasticity of the reaction chamber, allowing it to accommodate a range of acyl chains. (B) Surface diagram of RifR TEII, showing that access to the active site (rod) is blocked by the lid α-helices. (C) Close-up view of the closed entrance tunnel to the RifR active site. Lid opening is proposed to be facilitated by interaction with 'correct' ACP partners. Reproduced from Claxton, *et al.*, *J. Biol. Chem.*, 2009, **284**, 5021–5029 (ref. 234).

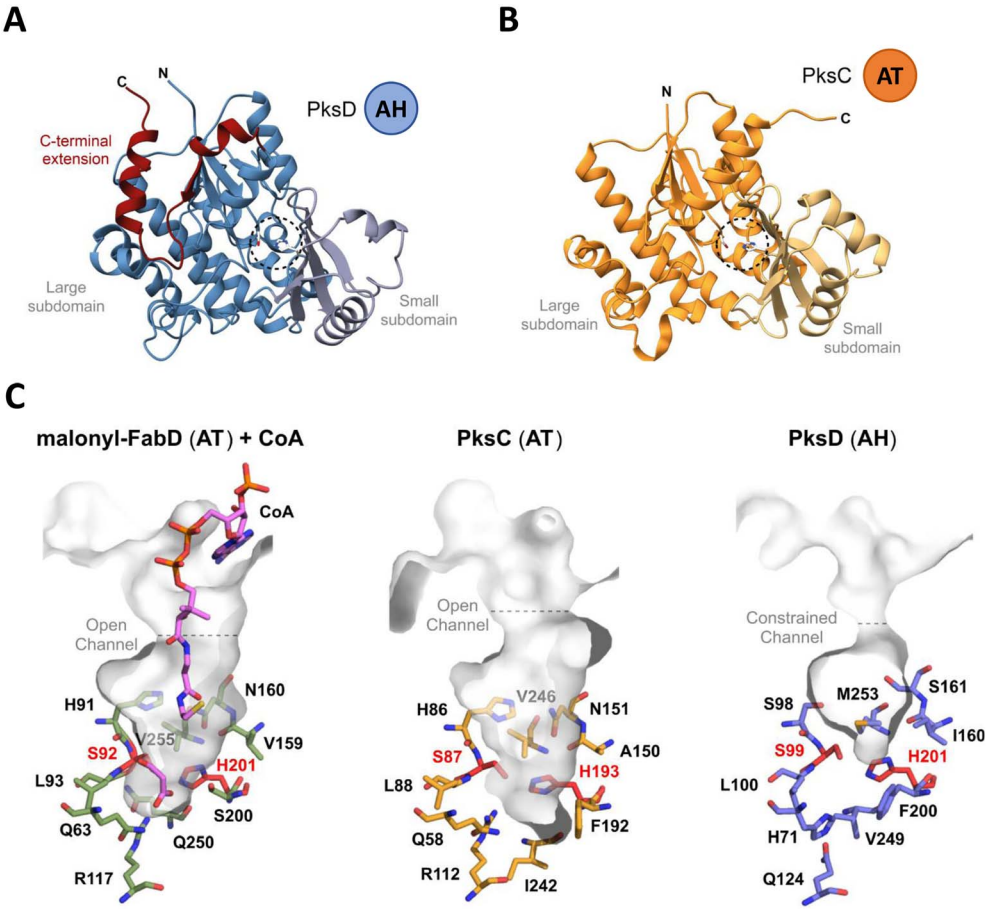

Fig. 35 Structural analysis of the bacillaene acyl hydrolase (AH) PksD (PDB ID: 8AVZ).[240] (A and B) Structural comparison between the AH PksD and the AT PksC (PDB ID: 5DZ6) from the same system, with the catalytic residues highlighted in stick form. The two enzymes exhibit the same global fold consistent with a common evolutionary origin, but PksD contains a C-terminal extension (shown in red). (C) Comparison of the substrate binding pocket of PksD relative to the AT PksC, and the malonyl-CoA-acyl carrier protein transacylase FabD bound to CoA (PDB ID: 2G2Z). Both the entrance channel and the substrate binding pocket are more constrained in PksD than the other enzymes, and the binding channel is partially hydrophilic, consistent with its preference for shorter chains. Reproduced and adapted with permission from Fage, *et al.*, Molecular basis for short-chain thioester hydrolysis by acyl hydrolases in *trans*-acyltransferase polyketide synthases, *JACS Au*, 2025, **5**, 144–157 (ref. 240).

selection and catalytic outcome are proposed to depend on the intrinsic features of the TE active sites (*i.e.* the shape, length and chemical environment of the acyl cavity at the exit end of the domain (Fig. 36)) and the substrates themselves (*i.e.* in-built structural constraints such as double bonds).[252]

In terms of substrate selection, the linear mode of biosynthesis implies that the TEs will only ever be confronted with fully-extended intermediates. However, to avoid premature chain transfer to the TE active site, the domains must discriminate between various levels of intermediate processing when modification domains are present in the upstream module. Given the covalent linkage of the TE to the upstream ACP (*i.e.* this situation is analogous to ACP → KS intermodular transfers within subunits, see Section 3), this decision point would seem to be particularly critical. Surprisingly, assays *in vitro* have demonstrated that TEs exhibit generally broad specificity towards their substrates in the acylation step.[12,250,257–262] Aberrant TE self-acylation with carboxylated extender unit (malonate, methylmalonate, *etc.*) is, on the other hand, likely discouraged by the negatively-charged character of the entrance channel.[249] According to the structural analysis, the broad substrate tolerance of the TEs arises from the overall large size and generally hydrophobic character of the acyl cavities, as well as a relative paucity of specific hydrogen bonding interactions, allowing for accommodation of a range of differently functionalized chains of varying length.[17,249,252,253]

In the case of the erythromycin (DEBS) PKS, at least, no specific protein–protein complex is formed with the upstream ACP domain in the absence of substrate,[263,264] explaining the ability of this TE to partner with diverse, non-native ACP domains in engineered contexts[80,265–267] (see Section 7.3). It is thus not at all clear how premature TE self-acylation is avoided. As noted previously,[264] one possible control mechanism is relative kinetics – *i.e.* if chain release by the TE is slow compared to catalysis by the upstream module, all modification reactions will have time to occur before the TE active site is vacated. Little relevant kinetic information is available from *in vitro* studies, however, as TEs have routinely been evaluated with simple

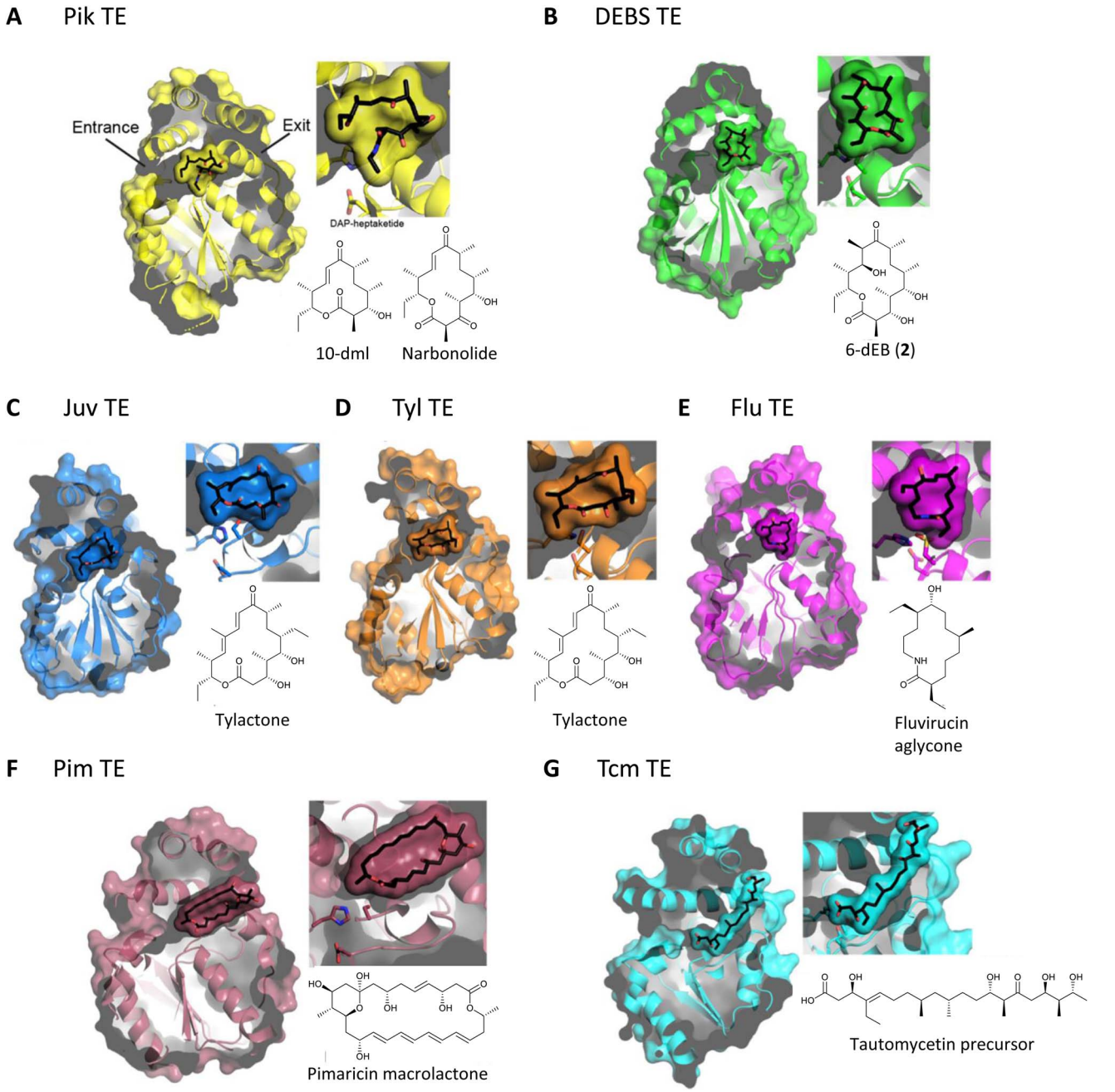

Fig. 36 Comparison of the reaction chambers in solved crystal structures of TE domains.[252] (A) Pik TE incorporating the non-natural amino acid 2,3-diaminopropionic acid (DAP) in place of the catalytic Ser, in the presence of bound heptaketide (PDB ID: 9CGL).[252] (B) DEBS TE modeled in the presence of 6-dEB (PDB ID: 5D3K). (C) Juvemicin TE modeled with bound tylactone (PDB ID: 9CEL). (D) Tylosin TE modeled with tylactone (PDB ID: 9CGO). (E) Fluvirucin TE modeled with fluvirucin aglycone (PDB ID: 9CFJ). (F) Pimaricin TE solved as a complex with an analogue of the native macrolactone product (PDB ID: 7VO5). (G) Tautomycetin TE in the presence of modeled product (PDB ID: 3LCR). In each case, the overall shape of the active site closely matches that of the products, consistent with its critical role in determining the reaction outcome. Reproduced and adapted from McCullough, *et al.*, Substrate trapping in polyketide synthase thioesterase domains: structural basis for macrolactone formation, *ACS Catal.*, **14**, 12551–12563, copyright 2024, with permission from the American Chemical Society (ref. 252).

model compounds[255,257,259,262,263] instead of close substrate mimic. Nonetheless, evidence has been obtained both *in vitro*[268,269] and *in vivo*[133] that the TE can influence the kinetics of competing reactions. Specifically, the presence of a downstream TE has been shown to provoke skipping of ketoreduction by the upstream module, presumably because the TE prefers the β-keto substrate. An alternative or perhaps simultaneous control mechanism is that access to the TE by the ACP only becomes physically possible following the full suite of processing reactions due to the conformational state of the upstream module (see Section 3).

A choice must also be made concerning the release mode employed – attack by an external nucleophile (usually water), or more typically, (macro)cyclization using a particular

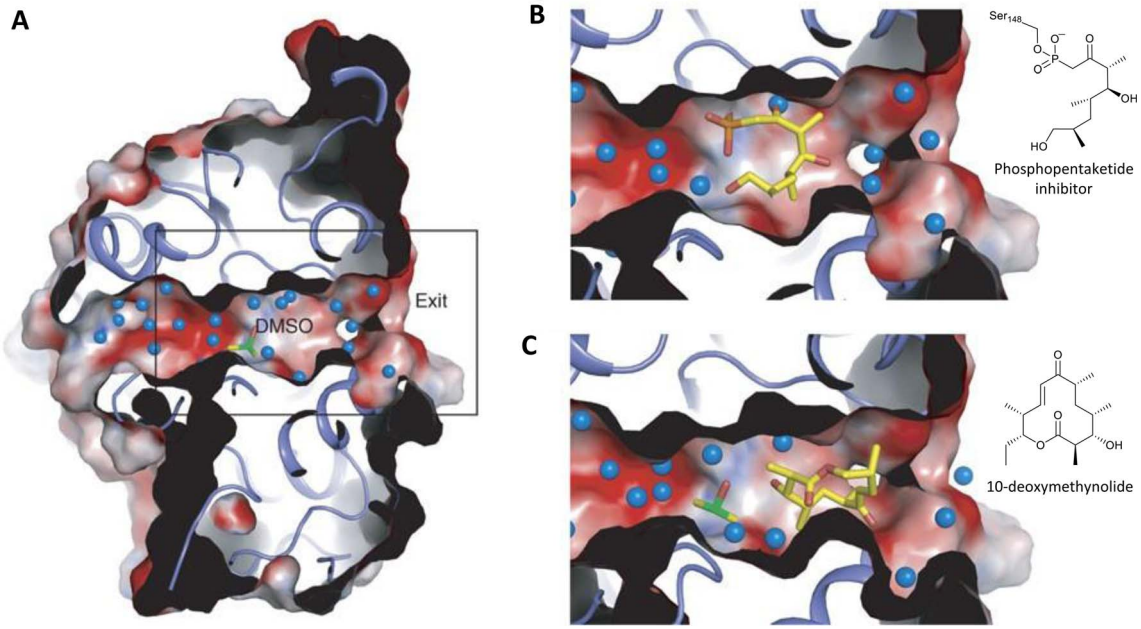

Fig. 37 Structural basis for macrocyclization by the PikTE. (A) Cut-away view of the active site of Pik TE in the absence of ligand. The electrostatic surface of the voluminous substrate channel is shown, alongside the water network (blue spheres), and a molecule of DMSO trapped in the oxyanion hole of the active site. The side from which the macrocyclic product exits the channel is indicated. (B) Zoom (region indicated by the rectangle in (A)) on phophopentaketide inhibitor (structure shown to the right) bound to the active site Ser. The presence of a hydrophilic zone at the distal end of the channel provokes the substrate to curl into a conformation favoring macrocyclization. (C) Zoom onto bound 10-deoxymethynolide in the substrate channel, with DMSO again occupying the oxyanion hole. The macrocyclic product engages in few direct contacts with the TE, with hydrogen bonds mediated only by water and the DMSO. Reproduced and adapted with permission from Akey, *et al.*, Structural and mechanistic insights into polyketide macrolactonization from polyketide-based affinity labels, *Nat. Chem. Biol.*, 2006, **2**, 537–542, copyright 2006, with permission from SNCSC (ref. 249).

nucleophile internal to the chain.[12] To our knowledge, only one case has been reported of a TE natively using multiple release mechanisms for the same intermediate: the alternative formation of diolide elaiophylin or linear, spiroketal pteridic acids.[270] In the case of macrocyclization, the enzyme must select which nucleophile to use in the process; characterized TEs employ hydroxyls and both amine and amide nitrogens as the attacking groups.[239] Rare TEs catalyze the release of two different intermediates in the pathway (*e.g.* 12- and 14-membered macrolides by the pikromycin TE[271]), or oligomerization of two full-length chains to yield dimers (macrodiolides).[272–274]

The available structures suggest a convincing model for the factors governing macrocyclization *vs.* hydrolysis. Globally, the acyl cavities of all the TEs exhibit shape and surface features that are complementary to their products (Fig. 36). For macrocyclizing TEs in particular, the cavity interior is highly hydrophobic near the catalytic triad, but more polar near the exit region. The presence of this distal 'hydrophilic barrier' directs the flexible chains to fold back upon themselves within the cavity, bringing the internal nucleophiles within range of the acyl terminus (Fig. 37).[249] In this context, the TEs are selective both for the type of nucleophile to be used (*e.g.* hydroxyl or amine) and in the case of DEBS and Pik TEs, at least, its stereochemistry,[251,275,276] but show more flexibility towards chain length (*i.e.* as long as the nucleophile can be brought into proximity to the acyl terminus, the TE will recognize it).[12] Although there is little direct contact with the substrates, those intermediates that do not exhibit a minimal set of structural determinants are hydrolyzed instead of macrocyclized.[249–251]

Concerning the model hydrolyzing tautomycetin (Tcm) TE, the acyl pocket is too narrow to house a curled substrate, and thus water is employed as the nucleophile. A special case concerns the CurM TE that must selectively hydrolyze an (*R*)-β-sulfated intermediate while ignoring its direct non-sulfated precursors.[247] Although there is little interaction between the active site and the substrate, a particular Arg residue interacts directly with the sulfate, resulting in the correct positioning of the chain relative to the catalytic machinery, while non-sulfated chains are not similarly oriented.

Turning to more unusual TEs, progress has been made towards understanding their modes of action. For example, in the case of oligomerizing TEs, the catalytic mechanism has been shown to involve TE-catalyzed attack of a specific hydroxyl of the upstream ACP-tethered substrate on the acyl group of its attached chain, resulting in a dimeric intermediate attached to the ACP (Fig. 38).[272,273] The linear dimer is then returned to the active site of the TE domain, and undergoes classical chain release as a macrodiolide using the same internal nucleophile. Even more remarkably, in the case of conglobatin (**15**) biosynthesis, the TE can also employ a hydroxyl nucleophile of another polyketide, benwamycin I, to offload both conglobatin monomer and dimer to form the corresponding esters.[277] Decrypting the molecular basis for this chemistry awaits structure elucidation of a model cyclodimerizing TE.

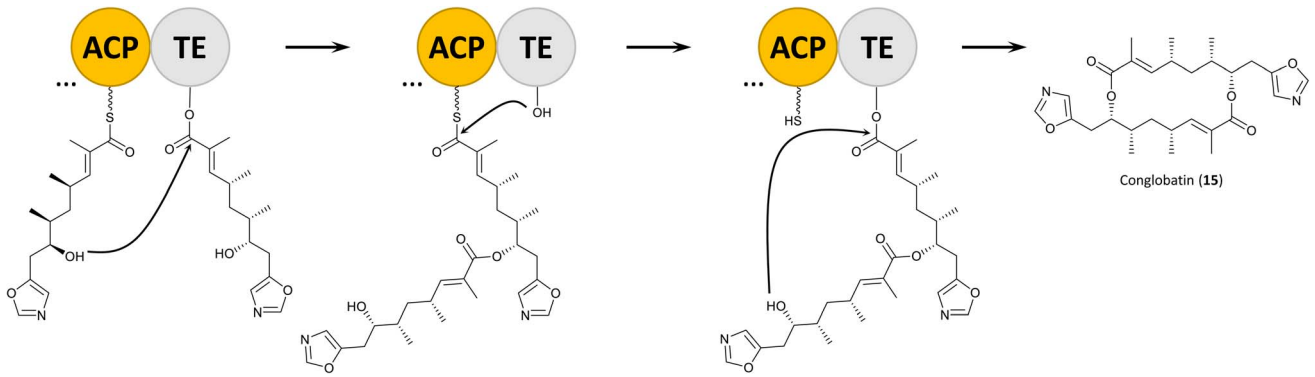

Fig. 38 Mechanism of polyketide cyclodimerization,[272,273] as illustrated for conglobatin (**15**) biosynthesis. First, an internal nucleophile of the fully elaborated intermediate attached to the ACP domain is used by the TE domain to attack the acyl terminus of an equivalent monomer tethered to the TE active site. The resulting linear dimer is then transferred onto the TE Ser. Finally, the TE catalyzes a second attack using the same nucleophile on its acyl terminus, resulting in concomitant release and macrocyclization of the product.

Finally, in terms of the pikromycin decision point that results in both twelve- and fourteen-membered macrolactones, release of the hexaketide from subunit PikAIII depends on an interaction between the last subunit PikAIV that houses the TE, and the upstream subunit PikAIII.[271] For this, PikAIV remains docked to PikAIII, but is proposed to adopt a configuration distinct from the normal co-linear subunit ordering that favors its interaction with the terminal ACP of PikAIII. Although action at a distance also explains release of premature intermediates from several engineered PKSs,[159,259] the factors underlying this configurational heterogeneity remain unknown.

7.3 Engineering release mechanisms

Many published examples exist in which TE domains have been challenged with unnatural substrates by genetic engineering, either by relocating them to alternative positions within the same systems,[80,159,266,278–280] altering their associated PKSs (*e.g.* by

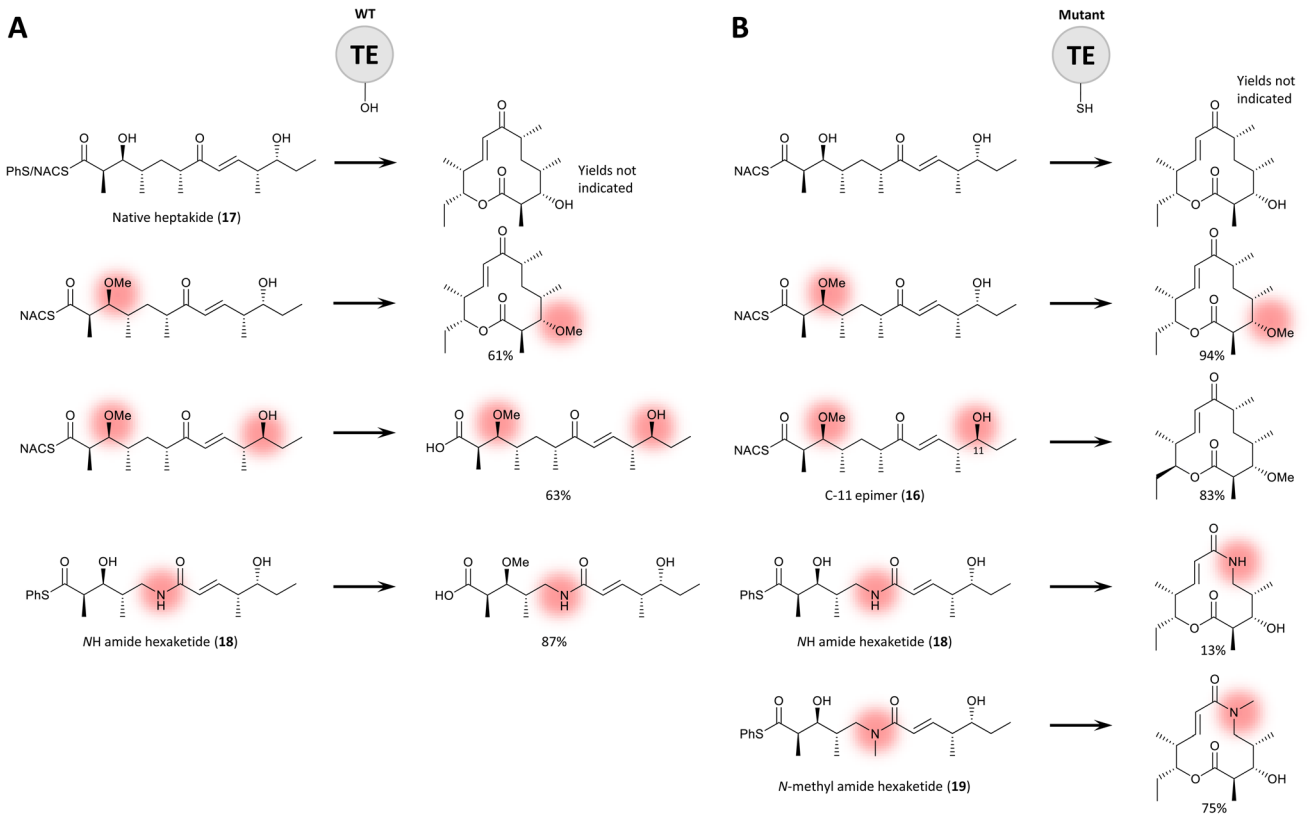

Fig. 39 Effect of an active site Ser → Cys mutation on catalysis by the Pik TE.[269] (A) Substrates macrocyclized by the wild type (Ser) Pik TE and their yields. The structural variations relative to the native hexaketide (**17**) are indicated in red. (B) Reactivity trends with the Cys active site mutant of the Pik TE. The mutant TE is able to macrocyclize multiple chains that are subject to hydrolytic release by the wild type enzyme, including several substrates incorporating amide functionality (**18**, **19**).

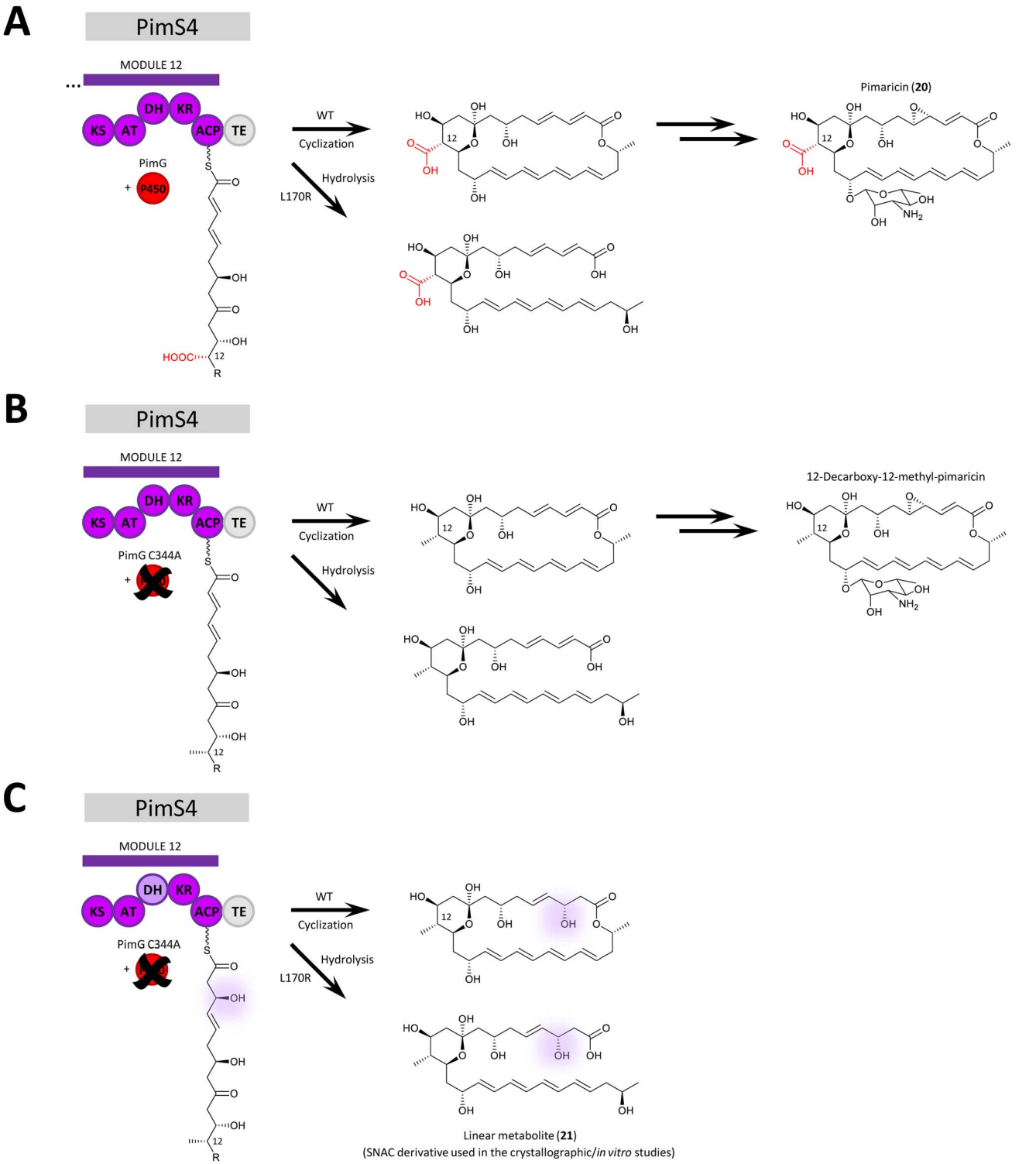

Fig. 40 Genetic engineering of the pimaricin (**20**) PKS TE.[254] (A) Introduction of an L170R mutation into the TE shifts the release mode from macrocyclization to hydrolysis. (B) When the *trans*-acting cytochrome P450 PimG is inactivated, the PKS generates C-12-Me chains instead of C-12-COOH. (C) Combining the mutations from (A) and (B) resulted principally in linear product (**21**), whose *N*-acetylcysteamine (SNAC) derivative was used to solve the structure of the TE/product complex. Note: the DH of module 12 is only partially active, resulting in a C-3 hydroxyl group (purple shading).

inactivation,[66,281–284] surgical modification[285,286] and exchange[29,267,287–296] of domains, or by removal of certain modules,[159,278] *etc.*), or transplanting the domains into heterologous assembly lines.[265,297–300] The high success rate of many of these experiments correlates with the generally broad specificity of the acylation half-reaction, and in the case of the

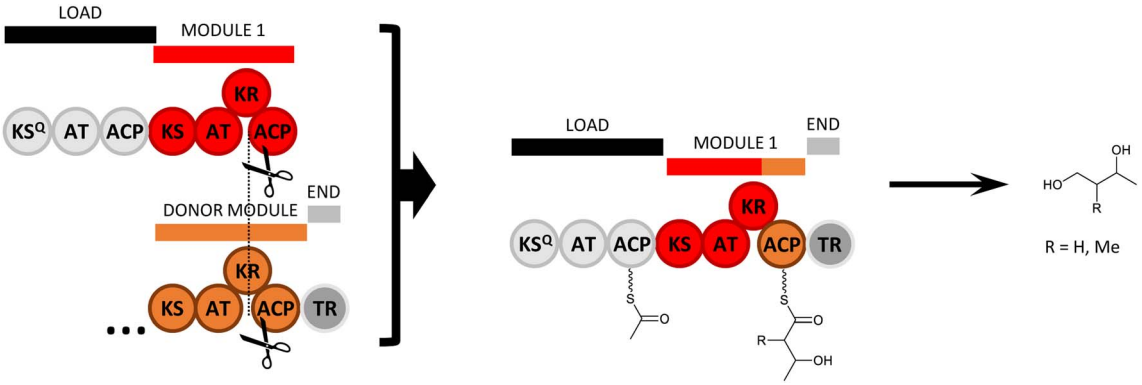

Fig. 41 Genetic engineering of chain release.[313] A series of mini-PKSs giving rise to diastereoisomers of 1,3-butanediols and 2-methyl-1,3-butanediols were engineered by attachment of C-terminal thioester reductase (TR) domains. The most efficient systems were obtained by introducing the TRs along with their native, upstream ACP partners, using a site N-terminal to the ACP.

relocation experiments, with the lack of a defined ACP/TE interface.[263,264] Nonetheless, close analysis of a handful of TEs both *in vitro* and *in vivo* (*e.g.* DEBS[250,257–259,275,301–303] and Pik[251,261,268,304–309]) has revealed limits to their tolerance, and therefore that these domains can become critical bottlenecks in engineered systems.[254,276,278] Non-respect for TE structural requirements in their substrates can result in hydrolysis as opposed to the desired macrolactonization, and/or low turnover to product, due to the greater intrinsic stringency of the chain release step.

Promisingly, it was recently demonstrated that an active site Ser to Cys mutant (S148C) of the Pik TE substantially boosted its efficiency with a C-11 epimer (**16**) of its natural hexaketide substrate (**17**) that is not accepted by the wild type enzyme (Fig. 39).[269] The lack of activity with the wild type was traced using quantum mechanical modeling (QMM) to a high rate-limiting activation barrier, which is correspondingly reduced in the mutant TE. Subsequently, the equivalent mutation introduced into a panel of TEs sourced from diverse systems almost systematically improved the absolute and relative (compared to hydrolytic) yields of the desired macrocyclic products starting from non-native hexa- and heptaketide substrates.[304] Furthermore, random mutagenesis (site-saturation coupled with error-prone PCR) of Pik TE S148C coupled with screening for macrocyclic product by HPLC-MS, allowed identification of mutant TEs that exhibited improved efficiency towards two hexaketides (**18**, **19**) incorporating non-natural amide functionality (Fig. 39).[310]

Specificity engineering was also carried out on the pimaricin (**20**) TE, a representative member of a family of TEs that has evolved to lactonize polyhydroxylated polyenes (Fig. 40). Notably, introduction of a single mutation into this TE within the context of a pimaricin pathway engineered to generated C-12 decarboxylated chains (*i.e.* bearing C-12-Me instead of C-12-COOH *via* inactivation of the cytochrome P450 PimG),[311,312] resulted in a pronounced shift from macrolactone formation towards hydrolysis (Fig. 40).[254] Cleverly, the *N*-acetylcysteamine (SNAC) derivative of this linear metabolite (**21**) was then leveraged both to solve the crystal structure of the TE/product complex (*i.e.* the substrate was macrocyclized during the crystallization), and in *in vitro* assays to test structure-guided mutations for their ability to increase turnover to the decarboxylated macrolactone. Overall, yields *in vitro* of the target macrocycle from mutant Pim TE were boosted by 39%, a result that was then recapitulated *in vivo* (37%) using the full PKS. Application of these various TE engineering strategies to additional intact PKS systems is eagerly anticipated.

An attractive alternative to modifying TE choices is to alter the fundamental chemistry of chain release.[313] This strategy was recently employed to generate small-molecule alcohols (diastereoisomers of 1,3-butanediols and 2-methyl-1,3-butanediols) using engineered bimodular PKSs, by tethering them to heterologous thioester reductase (TR) domains (Fig. 41).[314,315] TR domains are NADPH-dependent enzymes that catalyze 2- and 4-electron reduction reactions of acyl-thioesters to yield the corresponding aldehydes or alcohols, respectively.[316,317] Among the various fusion strategies tested, the highest titers were achieved when the TR was introduced along with its native, upstream ACP partner, using a site N-terminal to the ACP domain.[313] As suggested by the authors, this approach may preserve the key ACP/TR recognition interface involving ACP helix αII, but it nonetheless introduces non-native interactions with other domains within the module (*e.g.* the KS), and so may require further optimization.

8. Concluding remarks/outlook

As is clear from this review, a multitude of complicated, multi-layer mechanisms underlie the ability of modular PKS systems to navigate, albeit imperfectly, the many choices made during polyketide assembly. Such control processes often operate through specific structural features of the individual ACPs and their domain/enzyme partners, but additionally *via* their interactions, as well as at the multi-domain, whole module and multi-modular levels. While our understanding of these elements has advanced substantially, much remains to be discovered. Unresolved issues of note include, but are not limited to: (i) the basis for anti-selection of ACPs by partner

enzymes acting in *trans* that contributes to their selectivity; (ii) whether existing models for control at intra- *vs.* inter-modular decision junctions apply to *cis*-AT PKS modules of augmented domain composition, as well as to *trans*-AT PKS modules; (iii) the molecular features underlying the intrinsically iterative propensity of certain modules, as well as their capacity to deploy processing domains only in select cycles; (iv) the generality of the observed ACP/KS docking interaction at intermodular interfaces; (v) the amino acid basis for KS substrate specificity; and, (vi) the extent to which TE substrate choice and release mode can be influenced by site-directed mutagenesis and directed evolution-based approaches. It is anticipated that future research on these questions will boost our ability to rationally manipulate these key decision points, further enabling ongoing efforts to redirect PKS pathways towards the synthesis of novel analogues.

9. Conflicts of interest

The author declares no conflict of interest.

10. Data availability

No primary research results, software or code have been included and no new data were generated or analysed as part of this review.

11. Acknowledgements

Research in the author's lab is supported by the France 2030 program Lorraine Initiative of Excellence (ANR-15-IDEX-04-LUE), ANR grants BiosynKADH (ANR-24-CE92-0032-02) and PKSEng (ANR-24-CE93-0010-01), the Université de Lorraine and the CNRS.

12. References

1 I. Paterson and N. Y. S. Lam, *J. Antibiot.*, 2018, **71**, 215–233.
2 D.-S. Tian, X. Zhang and R. J. Cox, *Nat. Prod. Rep.*, 2025, **42**, 720–738.
3 J. Staunton and K. J. Weissman, *Nat. Prod. Rep.*, 2001, **18**, 380–416.
4 K. J. Weissman, in *Comprehensive Natural Products III*, ed. H.-W. (Ben) Liu and T. P. Begley, Elsevier, Oxford, 2020, pp. 4–46.
5 J. E. Cronan and J. Thomas, *Methods Enzymol.*, 2009, **459**, 395–433.
6 Q. Guo, C. Zhong, H. Dong, J. E. Cronan and H. Wang, *J. Biol. Chem.*, 2024, **300**, 105600.
7 A. Chen, R. N. Re and M. D. Burkart, *Nat. Prod. Rep.*, 2018, **35**, 1029–1045.
8 J. Wang, R. Zhang, X. Chen, X. Sun, Y. Yan, X. Shen and Q. Yuan, *Microb. Cell Fact.*, 2020, **19**, 110.
9 G. L. C. Grammbitter, M. Schmalhofer, K. Karimi, Y.-M. Shi, T. A. Schöner, N. J. Tobias, N. Morgner, M. Groll and H. B. Bode, *J. Am. Chem. Soc.*, 2019, **141**, 16615–16623.
10 A. F. Marsden, P. Caffrey, J. F. Aparicio, M. S. Loughran, J. Staunton and P. F. Leadlay, *Science*, 1994, **263**, 378–380.
11 A. Garg, C. Khosla and D. E. Cane, *J. Am. Chem. Soc.*, 2013, **135**, 16324–16327.
12 M. E. Horsman, T. P. A. Hari and C. N. Boddy, *Nat. Prod. Rep.*, 2016, **33**, 183–202.
13 J. Crosby and M. P. Crump, *Nat. Prod. Rep.*, 2012, **29**, 1111–1137.
14 J. Cortes, S. F. Haydock, G. A. Roberts, D. J. Bevitt and P. F. Leadlay, *Nature*, 1990, **348**, 176–178.
15 J. Staunton, P. Caffrey, J. F. Aparicio, G. A. Roberts, S. S. Bethell and P. F. Leadlay, *Nat. Struct. Biol.*, 1996, **3**, 188–192.
16 D. L. Akey, J. R. Razelun, J. Tehranisa, D. H. Sherman, W. H. Gerwick and J. L. Smith, *Structure*, 2010, **18**, 94–105.
17 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.
18 J. Zheng, C. D. Fage, B. Demeler, D. W. Hoffman and A. T. Keatinge-Clay, *ACS Chem. Biol.*, 2013, **8**, 1263–1270.
19 R. W. Broadhurst, D. Nietlispach, M. P. Wheatcroft, P. F. Leadlay and K. J. Weissman, *Chem. Biol.*, 2003, **10**, 723–731.
20 E. J. N. Helfrich and J. Piel, *Nat. Prod. Rep.*, 2016, **33**, 231–316.
21 M. C. Wilson and B. S. Moore, *Nat. Prod. Rep.*, 2012, **29**, 72–86.
22 L. Ray and B. S. Moore, *Nat. Prod. Rep.*, 2016, **33**, 150–161.
23 Y. A. Chan, A. M. Podevels, B. M. Kevany and M. G. Thomas, *Nat. Prod. Rep.*, 2009, **26**, 90–114.
24 B. S. Moore and C. Hertweck, *Nat. Prod. Rep.*, 2002, **19**, 70–99.
25 Y. A. Chan, M. T. Boyne, A. M. Podevels, A. K. Klimowicz, J. Handelsman, N. L. Kelleher and M. G. Thomas, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 14349–14354.
26 H. Park, B. M. Kevany, D. H. Dyer, M. G. Thomas and K. T. Forest, *PLoS One*, 2014, **9**, e110965.
27 D. J. Edwards, B. L. Marquez, L. M. Nogle, K. McPhail, D. E. Goeger, M. A. Roberts and W. H. Gerwick, *Chem. Biol.*, 2004, **11**, 817–833.
28 H.-Y. He, H. Yuan, M.-C. Tang and G.-L. Tang, *Angew Chem. Int. Ed. Engl.*, 2014, **53**, 11315–11319.
29 B. J. Dunn and C. Khosla, *J. R. Soc. Interface*, 2013, **10**, 20130297.
30 B. J. Dunn, D. E. Cane and C. Khosla, *Biochemistry*, 2013, **52**, 1839–1841.
31 S. Smith and S.-C. Tsai, *Nat. Prod. Rep.*, 2007, **24**, 1041–1072.
32 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.
33 E. Kalkreuter, K. S. Bingham, A. M. Keeler, A. N. Lowell, J. J. Schmidt, D. H. Sherman and G. J. Williams, *Nat. Commun.*, 2021, **12**, 2193.
34 Y. Feng, F. Zhang, S. Huang, Z. Deng, L. Bai and J. Zheng, *Acta Crystallogr., Sect. D: Struct. Biol.*, 2022, **78**, 779–791.
35 C. Oefner, H. Schulz, A. D'Arcy and G. E. Dale, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2006, **62**, 613–618.

36 C. Bisang, P. F. Long, J. Cortés, J. Westcott, J. Crosby, A. L. Matharu, R. J. Cox, T. J. Simpson, J. Staunton and P. F. Leadlay, *Nature*, 1999, **401**, 502–505.
37 M. A. Skiba, C. L. Tran, Q. Dan, A. P. Sikkema, Z. Klaver, W. H. Gerwick, D. H. Sherman and J. L. Smith, *Structure*, 2020, **28**, 63–74.
38 A. Y. Chen, N. A. Schnarr, C.-Y. Kim, D. E. Cane and C. Khosla, *J. Am. Chem. Soc.*, 2006, **128**, 3067–3074.
39 I. Koryakina, J. B. McArthur, M. M. Draelos and G. J. Williams, *Org. Biomol. Chem.*, 2013, **11**, 4449–4458.
40 M. Hans, A. Hornung, A. Dziarnowski, D. E. Cane and C. Khosla, *J. Am. Chem. Soc.*, 2003, **125**, 5366–5374.
41 L. Laureti, L. Song, S. Huang, C. Corre, P. Leblond, G. L. Challis and B. Aigle, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**, 6258–6263.
42 L. Su, Y. Souaibou, L. Hôtel, C. Jacob, P. Grün, Y.-N. Shi, A. Chateau, S. Pinel, H. B. Bode, B. Aigle and K. J. Weissman, *Chem. Sci.*, 2025, **16**, 5076–5088.
43 E. Englund, M. Schmidt, A. A. Nava, A. Lechner, K. Deng, R. Jocic, Y. Lin, J. Roberts, V. T. Benites, R. Kakumanu, J. W. Gin, Y. Chen, Y. Liu, C. J. Petzold, E. E. K. Baidoo, T. R. Northen, P. D. Adams, L. Katz, S. Yuzawa and J. D. Keasling, *J. Am. Chem. Soc.*, 2023, **145**, 8822–8832.
44 A. T. Keatinge-Clay, *Chem. Rev.*, 2017, **117**, 5334–5366.
45 A. Miyanaga, S. Iwasawa, Y. Shinohara, F. Kudo and T. Eguchi, *Proc. Natl. Acad. Sci. U. S. A.*, 2016, **113**, 1802–1807.
46 A. Miyanaga, K. Kawada, T. Chisuga, F. Kudo and T. Eguchi, *Biochemistry*, 2023, **62**, 17–21.
47 C. Zhao, J. M. Coughlin, J. Ju, D. Zhu, E. Wendt-Pienkowski, X. Zhou, Z. Wang, B. Shen and Z. Deng, *J. Biol. Chem.*, 2010, **285**, 20097–20108.
48 E. M. Musiol, T. Härtner, A. Kulik, J. Moldenhauer, J. Piel, W. Wohlleben and T. Weber, *Chem. Biol.*, 2011, **18**, 438–444.
49 L. Liu, W. Wang, M. Chen, Y. Zhang, H. Mao, D. Wang, Y. Chen and P. Li, *Appl. Microbiol. Biotechnol.*, 2023, **107**, 2403–2412.
50 E. M. Musiol-Kroll, F. Zubeil, T. Schafhauser, T. Härtner, A. Kulik, J. McArthur, I. Koryakina, W. Wohlleben, S. Grond, G. J. Williams, S. Y. Lee and T. Weber, *ACS Synth. Biol.*, 2017, **6**, 421–427.
51 E. M. Musiol and T. Weber, *Med. Chem. Commun.*, 2012, **3**, 871–886.
52 I. Koryakina, J. McArthur, S. Randall, M. M. Draelos, E. M. Musiol, D. C. Muddiman, T. Weber and G. J. Williams, *ACS Chem. Biol.*, 2013, **8**, 200–208.
53 A. Miyanaga, R. Ouchi, F. Ishikawa, E. Goto, G. Tanabe, F. Kudo and T. Eguchi, *J. Am. Chem. Soc.*, 2018, **140**, 7970–7978.
54 Z. Ye, E. M. Musiol, T. Weber and G. J. Williams, *Chem. Biol.*, 2014, **21**, 636–646.
55 E. M. Musiol, A. Greule, T. Härtner, A. Kulik, W. Wohlleben and T. Weber, *ChemBioChem*, 2013, **14**, 1343–1352.
56 C. D. Reeves, S. Murli, G. W. Ashley, M. Piagentini, C. R. Hutchinson and R. McDaniel, *Biochemistry*, 2001, **40**, 15464–15470.
57 F. Del Vecchio, H. Petkovic, S. G. Kendrew, L. Low, B. Wilkinson, R. Lill, J. Cortés, B. A. M. Rudd, J. Staunton and P. F. Leadlay, *J. Ind. Microbiol. Biotechnol.*, 2003, **30**, 489–494.
58 U. Sundermann, K. Bravo-Rodriguez, S. Klopries, S. Kushnir, H. Gomez, E. Sanchez-Garcia and F. Schulz, *ACS Chem. Biol.*, 2013, **8**, 443–450.
59 H. Petković, A. Sandmann, I. R. Challis, H.-J. Hecht, B. Silakowski, L. Low, N. Beeston, E. Kuscer, J. Garcia-Bernardo, P. F. Leadlay, S. G. Kendrew, B. Wilkinson and R. Müller, *Org. Biomol. Chem.*, 2008, **6**, 500–506.
60 S. Bhattacharya, E. G. Margheritis, K. Takahashi, A. Kulesha, A. D'Souza, I. Kim, J. H. Yoon, J. R. H. Tame, A. N. Volkov, O. V. Makhlynets and I. V. Korendovych, *Nature*, 2022, **610**, 389–393.
61 S. Yuzawa, K. Deng, G. Wang, E. E. K. Baidoo, T. R. Northen, P. D. Adams, L. Katz and J. D. Keasling, *ACS Synth. Biol.*, 2017, **6**, 139–147.
62 E. Englund, M. Schmidt, A. A. Nava, S. Klass, L. Keiser, Q. Dan, L. Katz, S. Yuzawa and J. D. Keasling, *Nat. Commun.*, 2023, **14**, 4871.
63 N. Lee, M. Schmidt, C. Li, C. J. Filbin, S. Klass, A. Yaguchi, A. L. Fear, T. W. H. Backman, B. A. Abel, Y.-M. Kim, W. E. Kim, C. W. Johnson, Y. Chen, J. W. Gin, C. J. Petzold, E. E. K. Baidoo, Y. Gao, M. C. Burnet, G. Babnigg, P. D. Laible, C. A. Singer, W. E. Michener, D. Salvachúa, H. Garcia Martin, R. W. Haushalter and J. D. Keasling, *Nat. Catal.*, 2025, **8**, 389–402.
64 S. R. Bagde, I. I. Mathews, J. C. Fromme and C.-Y. Kim, *Science*, 2021, **374**, 723–729.
65 Y. Tang, T. S. Lee, S. Kobayashi and C. Khosla, *Biochemistry*, 2003, **42**, 6588–6595.
66 P. Kumar, A. T. Koppisch, D. E. Cane and C. Khosla, *J. Am. Chem. Soc.*, 2003, **125**, 14307–14312.
67 B. J. Dunn, K. R. Watts, T. Robbins, D. E. Cane and C. Khosla, *Biochemistry*, 2014, **53**, 3796–3806.
68 N. B. Lopanik, J. A. Shields, T. J. Buchholz, C. M. Rath, J. Hothersall, M. G. Haygood, K. Håkansson, C. M. Thomas and D. H. Sherman, *Chem. Biol.*, 2008, **15**, 1175–1186.
69 M. C. Walker, B. W. Thuronyi, L. K. Charkoudian, B. Lowry, C. Khosla and M. C. Y. Chang, *Science*, 2013, **341**, 1089–1094.
70 S. Sirirungruang, O. Ad, T. M. Privalsky, S. Ramesh, J. L. Sax, H. Dong, E. E. K. Baidoo, B. Amer, C. Khosla and M. C. Y. Chang, *Nat. Chem. Biol.*, 2022, **18**, 886–893.
71 J. D. Kittendorf and D. H. Sherman, *Bioorg. Med. Chem.*, 2009, **17**, 2137–2146.
72 B. Lowry, X. Li, T. Robbins, D. E. Cane and C. Khosla, *ACS Cent. Sci.*, 2016, **2**, 14–20.
73 K. J. Weissman, *Nat. Chem. Biol.*, 2015, **11**, 660–670.
74 C. Anselmi, M. Grininger, P. Gipson and J. D. Faraldo-Gómez, *J. Am. Chem. Soc.*, 2010, **132**, 12357–12364.
75 C. H. Mak, P. Pham, S. A. Afif and M. F. Goodman, *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.*, 2015, **92**, 032717.
76 D. P. Cogan, K. Zhang, X. Li, S. Li, G. D. Pintilie, S.-H. Roh, C. S. Craik, W. Chiu and C. Khosla, *Science*, 2021, **374**, 729–734.

77 X. Li, N. Sevillano, F. La Greca, L. Deis, Y.-C. Liu, M. C. Deller, I. I. Mathews, T. Matsui, D. E. Cane, C. S. Craik and C. Khosla, *J. Am. Chem. Soc.*, 2018, **140**, 6518–6521.
78 L. Smith, H. Hong, J. B. Spencer and P. F. Leadlay, *ChemBioChem*, 2008, **9**, 2967–2975.
79 A. Migita, M. Watanabe, Y. Hirose, K. Watanabe, T. Tokiwano, H. Kinashi and H. Oikawa, *Biosci., Biotechnol., Biochem.*, 2009, **73**, 169–176.
80 J. Cortes, K. E. Wiesmann, G. A. Roberts, M. J. Brown, J. Staunton and P. F. Leadlay, *Science*, 1995, **268**, 1487–1489.
81 A. M. Soohoo, D. P. Cogan, K. L. Brodsky and C. Khosla, *Annu. Rev. Biochem.*, 2024, **93**, 471–498.
82 M. Grininger, *Nat. Chem. Biol.*, 2023, 1–15.
83 D. P. Cogan, A. M. Soohoo, M. Chen, Y. Liu, K. L. Brodsky and C. Khosla, *Nat. Chem. Biol.*, 2025, **21**, 876–882.
84 A. M. Hill and J. Staunton, in *Comprehensive Natural Products II*, ed. H.-W. (Ben) Liu and L. Mander, Elsevier, Oxford, 2010, pp. 385–452.
85 T. M. McCullough, A. Dhar, D. L. Akey, J. R. Konwerski, D. H. Sherman and J. L. Smith, *Structure*, 2023, **31**, 1109–1120.
86 A. Keatinge-Clay, *J. Mol. Biol.*, 2008, **384**, 941–953.
87 J. F. Barajas, R. P. McAndrew, M. G. Thompson, T. W. H. Backman, B. Pang, T. de Rond, J. H. Pereira, V. T. Benites, H. G. Martín, E. E. K. Baidoo, N. J. Hillson, P. D. Adams and J. D. Keasling, *J. Ind. Microbiol. Biotechnol.*, 2019, **46**, 1225–1235.
88 D. Gay, Y.-O. You, A. Keatinge-Clay and D. E. Cane, *Biochemistry*, 2013, **52**, 8916–8928.
89 D. Kawasaki, A. Miyanaga, T. Chisuga, F. Kudo and T. Eguchi, *Biochemistry*, 2019, **58**, 4799–4803.
90 K. H. Sung, G. Berkhan, T. Hollmann, L. Wagner, W. Blankenfeldt and F. Hahn, *Angew Chem. Int. Ed. Engl.*, 2018, **57**, 343–347.
91 G. J. Dodge, D. Ronnow, R. E. Taylor and J. L. Smith, *ACS Chem. Biol.*, 2018, **13**, 2699–2707.
92 D. Khare, W. A. Hale, A. Tripathi, L. Gu, D. H. Sherman, W. H. Gerwick, K. Håkansson and J. L. Smith, *Structure*, 2015, **23**, 2213–2223.
93 Y. U. Tittes, D. A. Herbst, S. F. X. Martin, H. Munoz-Hernandez, R. P. Jakob and T. Maier, *Sci. Adv.*, 2022, **8**, eabo6918.
94 D. C. Gay, D. T. Wagner, J. L. Meinke, C. E. Zogzas, G. R. Gay and A. T. Keatinge-Clay, *J. Struct. Biol.*, 2016, **193**, 196–205.
95 J. Fang, Y. Zhang, L. Huang, X. Jia, Q. Zhang, X. Zhang, G. Tang and W. Liu, *J. Bacteriol.*, 2008, **190**, 6014–6025.
96 G. Zhai, W. Wang, W. Xu, G. Sun, C. Hu, X. Wu, Z. Cong, L. Deng, Y. Shi, P. F. Leadlay, H. Song, K. Hong, Z. Deng and Y. Sun, *Angew Chem. Int. Ed. Engl.*, 2020, **59**, 22738–22742.
97 B. Zhang, Z. Xu, Q. Teng, G. Pan, M. Ma and B. Shen, *Angew Chem. Int. Ed. Engl.*, 2017, **56**, 7247–7251.
98 O. A. Trivedi, P. Arora, A. Vats, M. Z. Ansari, R. Tickoo, V. Sridharan, D. Mohanty and R. S. Gokhale, *Mol. Cell*, 2005, **17**, 631–643.
99 K. J. Weissman, *Nat. Prod. Rep.*, 2016, **33**, 203–230.
100 A. A. Nava, J. Roberts, R. W. Haushalter, Z. Wang and J. D. Keasling, *ACS Synth. Biol.*, 2023, **12**, 3148–3155.
101 C. Khosla, D. Herschlag, D. E. Cane and C. T. Walsh, *Biochemistry*, 2014, **53**, 2875–2883.
102 H. Chen and L. Du, *Appl. Microbiol. Biotechnol.*, 2016, **100**, 541–557.
103 X. Zhong, S. Liu, B. Ma, K. Gao, D. Jiang, Y. Hou, H. Chen, J. Lv, J. I. Bowen, M. P. Crump, C. L. Willis and L. Wang, *Nat. Commun.*, 2025, **16**, 7996.
104 N. Gaitatzis, B. Silakowski, B. Kunze, G. Nordsiek, H. Blöcker, G. Höfle and R. Müller, *J. Biol. Chem.*, 2002, **277**, 13082–13090.
105 J. He and C. Hertweck, *Chem. Biol.*, 2003, **10**, 1225–1232.
106 N. Traitcheva, H. Jenke-Kodama, J. He, E. Dittmann and C. Hertweck, *ChemBioChem*, 2007, **8**, 1841–1849.
107 P. Meiser, K. J. Weissman, H. B. Bode, D. Krug, J. S. Dickschat, A. Sandmann and R. Müller, *Chem. Biol.*, 2008, **15**, 771–781.
108 C. Olano, B. Wilkinson, S. J. Moss, A. F. Braña, C. Méndez, P. F. Leadlay and J. A. Salas, *Chem. Commun.*, 2003, 2780–2782.
109 S. Müller, S. Rachid, T. Hoffmann, F. Surup, C. Volz, N. Zaburannyi and R. Müller, *Chem. Biol.*, 2014, **21**, 855–865.
110 M. E. Yurkovich, R. Jenkins, Y. Sun, M. Tosin and P. F. Leadlay, *Chem. Commun.*, 2017, **53**, 2182–2185.
111 W. Tao, M. E. Yurkovich, S. Wen, K. E. Lebe, M. Samborskyy, Y. Liu, A. Yang, Y. Liu, Y. Ju, Z. Deng, M. Tosin, Y. Sun and P. F. Leadlay, *Chem. Sci.*, 2016, **7**, 376–385.
112 H. Hong, Y. Sun, Y. Zhou, E. Stephens, M. Samborskyy and P. F. Leadlay, *Beilstein J. Org. Chem.*, 2016, **12**, 2164–2172.
113 G. Zhai, Y. Zhu, G. Sun, F. Zhou, Y. Sun, Z. Hong, C. Dong, P. F. Leadlay, K. Hong, Z. Deng, F. Zhou and Y. Sun, *Nat. Commun.*, 2023, **14**, 612.
114 W. Xu, G. Zhai, Y. Liu, Y. Li, Y. Shi, K. Hong, H. Hong, P. F. Leadlay, Z. Deng and Y. Sun, *Angew Chem. Int. Ed. Engl.*, 2017, **56**, 5503–5506.
115 B. Wang, Y. Song, M. Luo, Q. Chen, J. Ma, H. Huang and J. Ju, *Org. Lett.*, 2013, **15**, 1278–1281.
116 J. C. Kwan, M. S. Donia, A. W. Han, E. Hirose, M. G. Haygood and E. W. Schmidt, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**, 20655–20660.
117 K. P. Yuet, C. W. Liu, S. R. Lynch, J. Kuo, W. Michaels, R. B. Lee, A. E. McShane, B. L. Zhong, C. R. Fischer and C. Khosla, *J. Am. Chem. Soc.*, 2020, **142**, 5952–5957.
118 K. M. Guzman, K. P. Yuet, S. R. Lynch, C. W. Liu and C. Khosla, *J. Org. Chem.*, 2021, **86**, 11100–11106.
119 D. Menche, F. Arikan, O. Perlova, N. Horstmann, W. Ahlbrecht, S. C. Wenzel, R. Jansen, H. Irschik and R. Müller, *J. Am. Chem. Soc.*, 2008, **130**, 14234–14243.
120 X. Jian, F. Pang, C. Hobson, M. Jenner, L. M. Alkhalaf and G. L. Challis, *J. Am. Chem. Soc.*, 2024, **146**, 6114–6124.
121 S. Tatsuno, K. Arakawa and H. Kinashi, *J. Antibiot.*, 2007, **60**, 700–708.

122 Z. P. Mai, B. Zhang, Z. X. Pang, J. Shi, Z. F. Xu, B.-B. Huang, S. Y. Ma, R. H. Jiao, Z.-J. Yao, R. X. Tan and H. M. Ge, *Nat. Synth.*, 2024, **3**, 1255–1265.
123 B. Wilkinson, G. Foster, B. A. Rudd, N. L. Taylor, A. P. Blackaby, P. J. Sidebottom, D. J. Cooper, M. J. Dawson, A. D. Buss, S. Gaisser, I. U. Böhm, C. J. Rowe, J. Cortés, P. F. Leadlay and J. Staunton, *Chem. Biol.*, 2000, **7**, 111–117.
124 I. H. Hardt, H. Steinmetz, K. Gerth, F. Sasse, H. Reichenbach and G. Höfle, *J. Nat. Prod.*, 2001, **64**, 847–856.
125 M. Grininger, *Proc. Natl. Acad. Sci. U. S. A.*, 2020, **117**, 8680–8682.
126 B. Wang, F. Guo, C. Huang and H. Zhao, *Proc. Natl. Acad. Sci. U. S. A.*, 2020, **117**, 8449–8454.
127 B. Busch, N. Ueberschaar, Y. Sugimoto and C. Hertweck, *J. Am. Chem. Soc.*, 2012, **134**, 12382–12385.
128 B. Busch, N. Ueberschaar, S. Behnken, Y. Sugimoto, M. Werneburg, N. Traitcheva, J. He and C. Hertweck, *Angew Chem. Int. Ed. Engl.*, 2013, **52**, 5285–5289.
129 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.
130 B. G. Ng, J. W. Han, D. W. Lee, G. J. Choi and B. S. Kim, *J. Antibiot.*, 2018, **71**, 495–505.
131 K. J. Weissman, *Nat. Synth.*, 2024, **3**, 1194–1196.
132 Y. Sugimoto, K. Ishida, N. Traitcheva, B. Busch, H.-M. Dahse and C. Hertweck, *Chem. Biol.*, 2015, **22**, 229–240.
133 A. A. Koch, J. J. Schmidt, A. N. Lowell, D. A. Hansen, K. M. Coburn, J. A. Chemler and D. H. Sherman, *Angew Chem. Int. Ed. Engl.*, 2020, **59**, 13575–13580.
134 H. G. Smith, M. J. Beech, J. R. Lewandowski, G. L. Challis and M. Jenner, *J. Ind. Microbiol. Biotechnol.*, 2021, **48**, kuab018.
135 N. Wu, S. Y. Tsuji, D. E. Cane and C. Khosla, *J. Am. Chem. Soc.*, 2001, **123**, 6465–6474.
136 S. Y. Tsuji, D. E. Cane and C. Khosla, *Biochemistry*, 2001, **40**, 2326–2331.
137 M. Klaus, L. Buyachuihan and M. Grininger, *ACS Chem. Biol.*, 2020, **15**, 2422–2432.
138 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.
139 T. J. Buchholz, T. W. Geders, F. E. Bartley, K. A. Reynolds, J. L. Smith and D. H. Sherman, *ACS Chem. Biol.*, 2009, **4**, 41–52.
140 M. Jenner, S. Kosol, D. Griffiths, P. Prasongpholchai, L. Manzi, A. S. Barrow, J. E. Moses, N. J. Oldham, J. R. Lewandowski and G. L. Challis, *Nat. Chem. Biol.*, 2018, **14**, 270–275.
141 S. Kosol, A. Gallo, D. Griffiths, T. R. Valentic, J. Masschelein, M. Jenner, E. L. C. de Los Santos, L. Manzi, P. K. Sydor, D. Rea, S. Zhou, V. Fülöp, N. J. Oldham, S. C. Tsai, G. L. Challis and J. R. Lewandowski, *Nat. Chem.*, 2019, **11**, 913–923.
142 C. Hacker, X. Cai, C. Kegler, L. Zhao, A. K. Weickhmann, J. P. Wurm, H. B. Bode and J. Wöhnert, *Nat. Commun.*, 2018, **9**, 4366.
143 C. D. Richter, D. Nietlispach, R. W. Broadhurst and K. J. Weissman, *Nat. Chem. Biol.*, 2008, **4**, 75–81.
144 F. Risser, S. Collin, R. Dos Santos-Morais, A. Gruez, B. Chagot and K. J. Weissman, *J. Struct. Biol.*, 2020, **212**, 107581.
145 J. Dorival, T. Annaval, F. Risser, S. Collin, P. Roblin, C. Jacob, A. Gruez, B. Chagot and K. J. Weissman, *J. Am. Chem. Soc.*, 2016, **138**, 4155–4167.
146 S. Scat, K. J. Weissman and B. Chagot, *RSC Chem. Biol.*, 2024, **5**, 669–683.
147 J. Zeng, D. T. Wagner, Z. Zhang, L. Moretto, J. D. Addison and A. T. Keatinge-Clay, *ACS Chem. Biol.*, 2016, **11**, 2466–2474.
148 S.-Y. Son, D.-W. Bae, E. Kim, B.-G. Jeong, M.-Y. Kim, S.-Y. Youn, S. Yi, G. Kim, J.-S. Hahn, N. K. Lee, Y. J. Yoon and S.-S. Cha, *Structure*, 2024, **32**, 1477–1487.
149 D. P. Dowling, Y. Kung, A. K. Croft, K. Taghizadeh, W. L. Kelly, C. T. Walsh and C. L. Drennan, *Proc. Natl. Acad. Sci. U. S. A.*, 2016, **113**, 12432–12437.
150 J. Watzel, E. Duchardt-Ferner, S. Sarawi, H. B. Bode and J. Wöhnert, *Angew Chem. Int. Ed. Engl.*, 2021, **60**, 14171–14178.
151 A. M. Watkins and P. S. Arora, *ACS Chem. Biol.*, 2014, **9**, 1747–1754.
152 M. Thattai, Y. Burak and B. I. Shraiman, *PLoS Comput. Biol.*, 2007, **3**, e186.
153 L. Mollica, L. M. Bessa, X. Hanoulle, M. R. Jensen, M. Blackledge and R. Schneider, *Front. Mol. Biosci.*, 2016, **3**, 52.
154 R. van der Lee, M. Buljan, B. Lang, R. J. Weatheritt, G. W. Daughdrill, A. K. Dunker, M. Fuxreiter, J. Gough, J. Gsponer, D. T. Jones, P. M. Kim, R. W. Kriwacki, C. J. Oldfield, R. V. Pappu, P. Tompa, V. N. Uversky, P. E. Wright and M. M. Babu, *Chem. Rev.*, 2014, **114**, 6589–6631.
155 I. M. A. Nooren and J. M. Thornton, *J. Mol. Biol.*, 2003, **325**, 991–1018.
156 M. Klaus, M. P. Ostrowski, J. Austerjost, T. Robbins, B. Lowry, D. E. Cane and C. Khosla, *J. Biol. Chem.*, 2016, **291**, 16404–16415.
157 M. Klaus, A. D. D'Souza, A. Nivina, C. Khosla and M. Grininger, *ACS Chem. Biol.*, 2019, **14**, 426–433.
158 N. Wu, F. Kudo, D. E. Cane and C. Khosla, *J. Am. Chem. Soc.*, 2000, **122**, 4847–4852.
159 L. Su, L. Hôtel, C. Paris, C. Chepkirui, A. O. Brachmann, J. Piel, C. Jacob, B. Aigle and K. J. Weissman, *Nat. Commun.*, 2022, **13**, 515.
160 H. G. Menzella, R. Reid, J. R. Carney, S. S. Chandran, S. J. Reisinger, K. G. Patel, D. A. Hopwood and D. V. Santi, *Nat. Biotechnol.*, 2005, **23**, 1171–1176.
161 H. G. Menzella, J. R. Carney and D. V. Santi, *Chem. Biol.*, 2007, **14**, 143–151.

162 K. A. Ray, J. D. Lutgens, R. Bista, J. Zhang, R. R. Desai, M. Hirsch, T. Miyazawa, A. Cordova and A. T. Keatinge-Clay, *Nat. Commun.*, 2024, **15**, 6485.
163 A. C. Murphy, H. Hong, S. Vance, R. W. Broadhurst and P. F. Leadlay, *Chem. Commun.*, 2016, **52**, 8373–8376.
164 H. Peng, K. Ishida, Y. Sugimoto, H. Jenke-Kodama and C. Hertweck, *Nat. Commun.*, 2019, **10**, 3918.
165 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.
166 E. J. N. Helfrich, R. Ueoka, A. Dolev, M. Rust, R. A. Meoded, A. Bhushan, G. Califano, R. Costa, M. Gugger, C. Steinbeck, P. Moreno and J. Piel, *Nat. Chem. Biol.*, 2019, **15**, 813–821.
167 E. J. N. Helfrich, R. Ueoka, M. G. Chevrette, F. Hemmerling, X. Lu, S. Leopold-Messer, H. A. Minas, A. Y. Burch, S. E. Lindow, J. Piel and M. H. Medema, *Nat. Commun.*, 2021, **12**, 1422.
168 M. F. J. Mabesoone, S. Leopold-Messer, H. A. Minas, C. Chepkirui, P. Chawengrum, S. Reiter, R. A. Meoded, S. Wolf, F. Genz, N. Magnus, B. Piechulla, A. S. Walker and J. Piel, *Science*, 2024, **383**, 1312–1327.
169 L. Zhang, T. Hashimoto, B. Qin, J. Hashimoto, I. Kozone, T. Kawahara, M. Okada, T. Awakawa, T. Ito, Y. Asakawa, M. Ueki, S. Takahashi, H. Osada, T. Wakimoto, H. Ikeda, K. Shin-Ya and I. Abe, *Angew Chem. Int. Ed. Engl.*, 2017, **56**, 1740–1745.
170 M. Hirsch, B. J. Fitzgerald and A. T. Keatinge-Clay, *ACS Chem. Biol.*, 2021, **16**, 2515–2526.
171 H. Yamamoto, H. Ueda, M. Aso, M. Lee and S. Yuzawa, *ACS Chem. Biol.*, 2026, **21**, 362–370.
172 G. Kim, D. Lee, J. H. Kim, S. D. Kim, H. Kim, J. H. Kim, S. S. Yim, S.-J. Yeom, J. D. Keasling and B.-K. Cho, *Nat. Prod. Rep.*, 2025, **42**, 1489–1506.
173 Y. Chainani, J. Diaz, M. Guilarte-Silva, V. Blay, Q. Zhang, W. Sprague, K. E. J. Tyo, L. J. Broadbelt, A. Mukhopadhyay, J. D. Keasling, H. G. Martin and T. W. H. Backman, *Nat. Commun.*, 2025, **16**, 5787.
174 J. S. Tuan, J. M. Weber, M. J. Staver, J. O. Leung, S. Donadio and L. Katz, *Gene*, 1990, **90**, 21–29.
175 A. T. Keatinge-Clay, *Angew Chem. Int. Ed. Engl.*, 2017, **56**, 4658–4660.
176 D. A. Vander Wood and A. T. Keatinge-Clay, *Proteins*, 2018, **86**, 664–675.
177 T. Miyazawa, B. J. Fitzgerald and A. T. Keatinge-Clay, *Chem. Commun.*, 2021, **57**, 8762–8765.
178 T. Miyazawa, M. Hirsch, Z. Zhang and A. T. Keatinge-Clay, *Nat. Commun.*, 2020, **11**, 80.
179 K. Kudo, T. Nishimura, I. Kozone, J. Hashimoto, N. Kagaya, H. Suenaga, H. Ikeda and K. Shin-Ya, *Sci. Rep.*, 2021, **11**, 9944.
180 A. Ranganathan, M. Timoney, M. Bycroft, J. Cortés, I. P. Thomas, B. Wilkinson, L. Kellenberger, U. Hanefeld, I. S. Galloway, J. Staunton and P. F. Leadlay, *Chem. Biol.*, 1999, **6**, 731–741.
181 K. D. Patel, M. R. MacDonald, S. F. Ahmed, J. Singh and A. M. Gulick, *Nat. Prod. Rep.*, 2023, **40**, 1550–1582.
182 T. Izoré and M. J. Cryle, *Nat. Prod. Rep.*, 2018, **35**, 1120–1139.
183 A. Miyanaga, F. Kudo and T. Eguchi, *Nat. Prod. Rep.*, 2018, **35**, 1185–1200.
184 K. A. J. Bozhüyük, F. Fleischhacker, A. Linck, F. Wesche, A. Tietze, C.-P. Niesert and H. B. Bode, *Nat. Chem.*, 2018, **10**, 275–281.
185 Z. Huang, S. Xie, R.-Z. Liu, C. Xiang, S. Yao and L. Zhang, *Nat. Chem. Biol.*, 2025, **21**, 1361–1367.
186 J.-M. Massicard, C. Soligot, K. J. Weissman and C. Jacob, *Chem. Commun.*, 2020, **56**, 12749–12752.
187 J. F. Aparicio, I. Molnár, T. Schwecke, A. König, S. F. Haydock, L. E. Khaw, J. Staunton and P. F. Leadlay, *Gene*, 1996, **169**, 9–16.
188 A. Wlodek, S. G. Kendrew, N. J. Coates, A. Hold, J. Pogwizd, S. Rudder, L. S. Sheehan, S. J. Higginbotham, A. E. Stanley-Smith, T. Warneck, M. Nur-E-Alam, M. Radzom, C. J. Martin, L. Overvoorde, M. Samborskyy, S. Alt, D. Heine, G. T. Carter, E. I. Graziani, F. E. Koehn, L. McDonald, A. Alanine, R. M. Rodríguez Sarmiento, S. K. Chao, H. Ratni, L. Steward, I. H. Norville, M. Sarkar-Tyson, S. J. Moss, P. F. Leadlay, B. Wilkinson and M. A. Gregory, *Nat. Commun.*, 2017, **8**, 1206.
189 T. P. Stinear, A. Mve-Obiang, P. L. C. Small, W. Frigui, M. J. Pryor, R. Brosch, G. A. Jenkin, P. D. R. 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.
190 J. Yan, S. Gupta, D. H. Sherman and K. A. Reynolds, *ChemBioChem*, 2009, **10**, 1537–1543.
191 Y. Liu, C. Song, Q. Cui, H. Sun, C. Jiang, R. Guo, R. He, Z. Li, J. Luan and H. Wang, *Nat. Commun.*, 2025, **16**, 774.
192 M. M. Zdouc, *et al.*, *Nucleic Acids Res.*, 2025, **53**, D678–D690.
193 S. Collin and K. J. Weissman, *Nat. Commun.*, 2025, **16**, 6690.
194 K. E. Thompson, C. J. Bashor, W. A. Lim and A. E. Keating, *ACS Synth. Biol.*, 2012, **1**, 118–129.
195 A. W. Reinke, R. A. Grant and A. E. Keating, *J. Am. Chem. Soc.*, 2010, **132**, 6025–6031.
196 L. Buyachuihan, Y. Zhao, C. Schelhas and M. Grininger, *ACS Chem. Biol.*, 2023, **18**, 1500–1509.
197 X. Sun, Y. Yuan, Q. Chen, S. Nie, J. Guo, Z. Ou, M. Huang, Z. Deng, T. Liu and T. Ma, *Nat. Commun.*, 2022, **13**, 5541.
198 J. L. Meinke, A. J. Simon, D. T. Wagner, B. R. Morrow, S. You, A. D. Ellington and A. T. Keatinge-Clay, *ACS Synth. Biol.*, 2019, **8**, 2017–2024.
199 F. Hemmerling, R. A. Meoded, A. E. Fraley, H. A. Minas, C. L. Dieterich, M. Rust, R. Ueoka, K. Jensen, E. J. N. Helfrich, C. Bergande, M. Biedermann, N. Magnus, B. Piechulla and J. Piel, *Angew Chem. Int. Ed. Engl.*, 2022, **61**, e202116614.
200 A. J. Winter, N. Khanizeman, A. M. C. Barker-Mountford, A. J. Devine, L. Wang, Z. Song, J. A. Davies, P. R. Race, C. Williams, T. J. Simpson, C. L. Willis and M. P. Crump, *Angew Chem. Int. Ed. Engl.*, 2023, e202312514.
201 P. D. Walker, A. N. M. Weir, C. L. Willis and M. P. Crump, *Nat. Prod. Rep.*, 2021, **38**, 723–756.

202 C. T. Calderone, W. E. Kowtoniuk, N. L. Kelleher, C. T. Walsh and P. C. Dorrestein, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 8977–8982.
203 C. T. Calderone, *Nat. Prod. Rep.*, 2008, **25**, 845–853.
204 S. Collin, R. J. Cox, C. Paris, C. Jacob, B. Chagot, K. J. Weissman and A. Gruez, *Nat. Commun.*, 2023, **14**, 1327.
205 S. Collin, K. J. Weissman and A. Gruez, *JACS Au*, 2024, **4**, 3833–3847.
206 A. P. Phillips, A. J. Winter, C. M. Hooper, C. Williams, J. Crosby, C. L. Willis and M. P. Crump, *Angew Chem. Int. Ed. Engl.*, 2025, **64**, e202414165.
207 M. E. M. Hiseman, A. P. Phillips, C. Chiriac, L. J. Smith, J. Crosby, C. Williams, C. L. Willis, A. J. Winter and M. P. Crump, *RSC Adv.*, 2025, **15**, 40855–40863.
208 N. Pulsawat, S. Kitani and T. Nihira, *Gene*, 2007, **393**, 31–42.
209 Y. Mast, T. Weber, M. Gölz, R. Ort-Winklbauer, A. Gondran, W. Wohlleben and E. Schinko, *Microb. Biotechnol.*, 2011, **4**, 192–206.
210 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, *Nat. Chem. Biol.*, 2013, **9**, 685–692.
211 J. Davison, J. Dorival, H. Rabeharindranto, H. Mazon, B. Chagot, A. Gruez and K. J. Weissman, *Chem. Sci.*, 2014, **5**, 3081–3095.
212 F. P. Maloney, L. Gerwick, W. H. Gerwick, D. H. Sherman and J. L. Smith, *Proc. Natl. Acad. Sci. U. S. A.*, 2016, **113**, 10316–10321.
213 S. Meng, A. D. Steele, W. Yan, G. Pan, E. Kalkreuter, Y.-C. Liu, Z. Xu and B. Shen, *Nat. Commun.*, 2021, **12**, 5672.
214 Y. Huang, S.-X. Huang, J. Ju, G. Tang, T. Liu and B. Shen, *Org. Lett.*, 2011, **13**, 498–501.
215 A. Becerril, I. Pérez-Victoria, J. M. Martín, F. Reyes, J. A. Salas and C. Méndez, *ACS Chem. Biol.*, 2022, **17**, 2320–2331.
216 G. Pan, Z. Xu, Z. Guo, Hindra, M. Ma, D. Yang, H. Zhou, Y. Gansemans, X. Zhu, Y. Huang, L.-X. Zhao, Y. Jiang, J. Cheng, F. Van Nieuwerburgh, J.-W. Suh, Y. Duan and B. Shen, *Proc. Natl. Acad. Sci. U. S. A.*, 2017, **114**, E11131–E11140.
217 A. D. Steele, S. Meng, G. Li, E. Kalkreuter, C. Chang and B. Shen, *J. Am. Chem. Soc.*, 2024, **146**, 32605–32617.
218 V. Simunovic and R. Müller, *ChemBioChem*, 2007, **8**, 1273–1280.
219 N. Moebius, C. Ross, K. Scherlach, B. Rohm, M. Roth and C. Hertweck, *Chem. Biol.*, 2012, **19**, 1164–1174.
220 Z. Chang, N. Sitachitta, J. V. Rossi, M. A. Roberts, P. M. Flatt, J. Jia, D. H. Sherman and W. H. Gerwick, *J. Nat. Prod.*, 2004, **67**, 1356–1367.
221 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.
222 D. Khare, B. Wang, L. Gu, J. Razelun, D. H. Sherman, W. H. Gerwick, K. Håkansson and J. L. Smith, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 14099–14104.
223 L. Gu, B. Wang, A. Kulkarni, T. W. Geders, R. V. Grindberg, L. Gerwick, K. Håkansson, P. Wipf, J. L. Smith, W. H. Gerwick and D. H. Sherman, *Nature*, 2009, **459**, 731–735.
224 L. Gu, J. Jia, H. Liu, K. Håkansson, W. H. Gerwick and D. H. Sherman, *J. Am. Chem. Soc.*, 2006, **128**, 9014–9015.
225 A. Nakagawa, Y. Konda, A. Hatano, Y. Harigaya, M. Onda and S. Omura, *J. Org. Chem.*, 1988, **53**, 2660–2661.
226 D. Wang, H. Mao, Z. Zhao, L. Liu, Y. Chen and P. Li, *Adv. Sci.*, 2024, **11**, e2401708.
227 M. L. Heathcote, J. Staunton and P. F. Leadlay, *Chem. Biol.*, 2001, **8**, 207–220.
228 P. Li, M. Chen, W. Tang, Z. Guo, Y. Zhang, M. Wang, G. P. Horsman, J. Zhong, Z. Lu and Y. Chen, *Nat. Commun.*, 2021, **12**, 4499.
229 K. Geyer, S. Hartmann, R. R. Singh and T. J. Erb, *Biochemistry*, 2022, **61**, 2662–2671.
230 Y.-Y. Wang, X.-X. Ran, W.-B. Chen, S.-P. Liu, X.-S. Zhang, Y.-Y. Guo, X.-H. Jiang, H. Jiang and Y.-Q. Li, *FEBS Lett.*, 2014, **588**, 3259–3264.
231 M. Kotowska, K. Pawlik, A. Smulczyk-Krawczyszyn, H. Bartosz-Bechowski and K. Kuczek, *Appl. Environ. Microbiol.*, 2009, **75**, 887–896.
232 Z. Hu, B. A. Pfeifer, E. Chao, S. Murli, J. Kealey, J. R. Carney, G. Ashley, C. Khosla and C. R. Hutchinson, *Microbiology*, 2003, **149**, 2213–2225.
233 H. Nakamura, J. X. Wang and E. P. Balskus, *Chem. Sci.*, 2015, **6**, 3816–3822.
234 H. B. Claxton, D. L. Akey, M. K. Silver, S. J. Admiraal and J. L. Smith, *J. Biol. Chem.*, 2009, **284**, 5021–5029.
235 B. S. Kim, T. A. Cropp, B. J. Beck, D. H. Sherman and K. A. Reynolds, *J. Biol. Chem.*, 2002, **277**, 48028–48034.
236 Y. Zhou, Q. Meng, D. You, J. Li, S. Chen, D. Ding, X. Zhou, H. Zhou, L. Bai and Z. Deng, *Appl. Environ. Microbiol.*, 2008, **74**, 7235–7242.
237 M. Kotowska and K. Pawlik, *Appl. Microbiol. Biotechnol.*, 2014, **98**, 7735–7746.
238 A. R. Butler, N. Bate and E. Cundliffe, *Chem. Biol.*, 1999, **6**, 287–292.
239 R. F. Little and C. Hertweck, *Nat. Prod. Rep.*, 2022, **39**, 163–205.
240 C. D. Fage, M. Passmore, B. P. Tatman, H. G. Smith, X. Jian, U. C. Dissanayake, M. E. Foran, G. A. Cisneros, G. L. Challis, J. R. Lewandowski and M. Jenner, *JACS Au*, 2025, **5**, 144–157.
241 M. Jenner, J. P. Afonso, C. Kohlhaas, P. Karbaum, S. Frank, J. Piel and N. J. Oldham, *Chem. Commun.*, 2016, **52**, 5262–5265.
242 C. Olano, C. Méndez and J. A. Salas, *Nat. Prod. Rep.*, 2010, **27**, 571–616.
243 R. Little, F. Trottmann, M. Preissler and C. Hertweck, *RSC Chem. Biol.*, 2022, **3**, 1121–1128.
244 J.-W. Tang, X. Liu, W. Ye, Z.-R. Li and P.-Y. Qian, *Nat. Prod. Rep.*, 2022, **39**, 991–1014.
245 T. Liu, D. You, C. Valenzano, Y. Sun, J. Li, Q. Yu, X. Zhou, D. E. Cane and Z. Deng, *Chem. Biol.*, 2006, **13**, 945–955.

246 B. M. Harvey, H. Hong, M. A. Jones, Z. A. Hughes-Thomas, R. M. Goss, M. L. Heathcote, V. M. Bolanos-Garcia, W. Kroutil, J. Staunton, P. F. Leadlay and J. B. Spencer, *ChemBioChem*, 2006, **7**, 1435–1442.
247 J. J. Gehret, L. Gu, W. H. Gerwick, P. Wipf, D. H. Sherman and J. L. Smith, *J. Biol. Chem.*, 2011, **286**, 14445–14454.
248 D. Gu, R. Zhai, B. Daymo, Y. Xie, C. Luo and W. Zhang, *Biochemistry*, 2024, **63**, 3213–3219.
249 D. L. Akey, J. D. Kittendorf, J. W. Giraldes, R. A. Fecik, D. H. Sherman and J. L. Smith, *Nat. Chem. Biol.*, 2006, **2**, 537–542.
250 R. Aggarwal, P. Caffrey, P. F. Leadlay, C. J. Smith and J. Staunton, *J. Chem. Soc., Chem. Commun.*, 1995, 1519–1520.
251 C. C. Aldrich, L. Venkatraman, D. H. Sherman and R. A. Fecik, *J. Am. Chem. Soc.*, 2005, **127**, 8910–8911.
252 T. M. McCullough, V. Choudhary, D. L. Akey, M. A. Skiba, S. M. Bernard, J. D. Kittendorf, J. J. Schmidt, D. H. Sherman and J. L. Smith, *ACS Catal.*, 2024, **14**, 12551–12563.
253 J. W. Giraldes, D. L. Akey, J. D. Kittendorf, D. H. Sherman, J. L. Smith and R. A. Fecik, *Nat. Chem. Biol.*, 2006, **2**, 531–536.
254 Y. Zhou, W. Tao, Z. Qi, J. Wei, T. Shi, Q. Kang, J. Zheng, Y. Zhao and L. Bai, *ACS Catal.*, 2022, **12**, 762–776.
255 J. B. Scaglione, D. L. Akey, R. Sullivan, J. D. Kittendorf, C. M. Rath, E.-S. Kim, J. L. Smith and D. H. Sherman, *Angew Chem. Int. Ed. Engl.*, 2010, **49**, 5726–5730.
256 S.-C. Tsai, H. Lu, D. E. Cane, C. Khosla and R. M. Stroud, *Biochemistry*, 2002, **41**, 12598–12606.
257 K. J. Weissman, C. J. Smith, U. Hanefeld, R. Aggarwal, M. Bycroft, J. Staunton and P. F. Leadlay, *Angew Chem. Int. Ed. Engl.*, 1998, **37**, 1437–1440.
258 K. K. Sharma and C. N. Boddy, *Bioorg. Med. Chem. Lett.*, 2007, **17**, 3034–3037.
259 R. S. Gokhale, D. Hunziker, D. E. Cane and C. Khosla, *Chem. Biol.*, 1999, **6**, 117–125.
260 M. Wang, P. Opare and C. N. Boddy, *Bioorg. Med. Chem. Lett.*, 2009, **19**, 1413–1415.
261 H. Lu, S.-C. Tsai, C. Khosla and D. E. Cane, *Biochemistry*, 2002, **41**, 12590–12597.
262 Y. Yin, H. Lu, C. Khosla and D. E. Cane, *J. Am. Chem. Soc.*, 2003, **125**, 5671–5676.
263 L. Tran, M. Tosin, J. B. Spencer, P. F. Leadlay and K. J. Weissman, *ChemBioChem*, 2008, **9**, 905–915.
264 L. Tran, R. W. Broadhurst, M. Tosin, A. Cavalli and K. J. Weissman, *Chem. Biol.*, 2010, **17**, 705–716.
265 C. J. Martin, M. C. Timoney, R. M. Sheridan, S. G. Kendrew, B. Wilkinson, J. C. Staunton and P. F. Leadlay, *Org. Biomol. Chem.*, 2003, **1**, 4144–4147.
266 C. M. Kao, G. Luo, L. Katz, D. E. Cane and C. Khosla, *J. Am. Chem. Soc.*, 1995, **117**, 9105–9106.
267 C. M. Kao, M. McPherson, R. N. McDaniel, H. Fu, D. E. Cane and C. Khosla, *J. Am. Chem. Soc.*, 1997, **119**, 11339–11340.
268 J. D. Mortison, J. D. Kittendorf and D. H. Sherman, *J. Am. Chem. Soc.*, 2009, **131**, 15784–15793.
269 A. A. Koch, D. A. Hansen, V. V. Shende, L. R. Furan, K. N. Houk, G. Jiménez-Osés and D. H. Sherman, *J. Am. Chem. Soc.*, 2017, **139**, 13456–13465.
270 Z. Yang, Y. Qiao, N. C. Konakalla, E. Strøbech, P. Harris, G. Peschel, M. Agler-Rosenbaum, T. Weber, E. Andreasson and L. Ding, *Nat. Commun.*, 2023, **14**, 7398.
271 J. D. Kittendorf, B. J. Beck, T. J. Buchholz, W. Seufert and D. H. Sherman, *Chem. Biol.*, 2007, **14**, 944–954.
272 Y. Zhou, A. C. Murphy, M. Samborskyy, P. Prediger, L. C. Dias and P. F. Leadlay, *Chem. Biol.*, 2015, **22**, 745–754.
273 Y. Zhou, P. Prediger, L. C. Dias, A. C. Murphy and P. F. Leadlay, *Angew Chem. Int. Ed. Engl.*, 2015, **54**, 5232–5235.
274 S. Zhang, S. Fan, H. He, J. Zhu, L. Murray, G. Liang, S. Ran, Y. Z. Zhu, M. J. Cryle, H.-Y. He and Y. Zhang, *Chem. Soc. Rev.*, 2025, **54**, 396–464.
275 A. Pinto, M. Wang, M. Horsman and C. N. Boddy, *Org. Lett.*, 2012, **14**, 2278–2281.
276 D. A. Hansen, A. A. Koch and D. H. Sherman, *J. Am. Chem. Soc.*, 2017, **139**, 13450–13455.
277 W. Zhang, M. Yang, W. Li, L. Zhou, Y. Shen, S.-P. Wang, J.-M. Gao, H.-W. Lin, J. Qi and Y. Zhou, *J. Agric. Food Chem.*, 2023, **71**, 7459–7467.
278 Y. Song, M. Hogan, J. Muldoon, P. Evans and P. Caffrey, *Molecules*, 2024, **29**, 1396.
279 C. M. Kao, G. Luo, L. Katz, D. E. Cane and C. Khosla, *J. Am. Chem. Soc.*, 1994, **116**, 11612–11613.
280 I. Böhm, I. E. Holzbaur, U. Hanefeld, J. Cortés, J. Staunton and P. F. Leadlay, *Chem. Biol.*, 1998, **5**, 407–412.
281 S. Donadio, M. J. Staver, J. B. McAlpine, S. J. Swanson and L. Katz, *Science*, 1991, **252**, 675–679.
282 S. Donadio, J. B. McAlpine, P. J. Sheldon, M. Jackson and L. Katz, *Proc. Natl. Acad. Sci. U. S. A.*, 1993, **90**, 7119–7123.
283 S. Kushnir, U. Sundermann, S. Yahiaoui, A. Brockmeyer, P. Janning and F. Schulz, *Angew Chem. Int. Ed. Engl.*, 2012, **51**, 10664–10669.
284 D. Kong, M.-J. Lee, S. Lin and E.-S. Kim, *J. Ind. Microbiol. Biotechnol.*, 2013, **40**, 529–543.
285 K. Bravo-Rodriguez, A. F. Ismail-Ali, S. Klopries, S. Kushnir, S. Ismail, E. K. Fansa, A. Wittinghofer, F. Schulz and E. Sanchez-Garcia, *ChemBioChem*, 2014, **15**, 1991–1997.
286 D. H. Kwan, Y. Sun, F. Schulz, H. Hong, B. Popovic, J. C. C. Sim-Stark, S. F. Haydock and P. F. Leadlay, *Chem. Biol.*, 2008, **15**, 1231–1240.
287 D. Bedford, J. R. Jacobsen, G. Luo, D. E. Cane and C. Khosla, *Chem. Biol.*, 1996, **3**, 827–831.
288 C. M. Kao, M. McPherson, R. N. McDaniel, H. Fu, D. E. Cane and C. Khosla, *J. Am. Chem. Soc.*, 1998, **120**, 2478–2479.
289 L. Kellenberger, I. S. Galloway, G. Sauter, G. Böhm, U. Hanefeld, J. Cortés, J. Staunton and P. F. Leadlay, *ChemBioChem*, 2008, **9**, 2740–2749.
290 T. Annaval, C. Paris, P. F. Leadlay, C. Jacob and K. J. Weissman, *ChemBioChem*, 2015, **16**, 1357–1364.
291 M. Oliynyk, M. J. Brown, J. Cortés, J. Staunton and P. F. Leadlay, *Chem. Biol.*, 1996, **3**, 833–839.

292 X. Ruan, A. Pereda, D. L. Stassi, D. Zeidner, R. G. Summers, M. Jackson, A. Shivakumar, S. Kakavas, M. J. Staver, S. Donadio and L. Katz, *J. Bacteriol.*, 1997, **179**, 6416–6425.
293 L. Liu, A. Thamchaipenet, H. Fu, M. Betlach and G. Ashley, *J. Am. Chem. Soc.*, 1997, **119**, 10553–10554.
294 R. McDaniel, A. Thamchaipenet, C. Gustafsson, H. Fu, M. Betlach and G. Ashley, *Proc. Natl. Acad. Sci. U. S. A.*, 1999, **96**, 1846–1851.
295 H. Petkovic, R. E. Lill, R. M. Sheridan, B. Wilkinson, E. L. McCormick, H. A. I. McArthur, J. Staunton, P. F. Leadlay and S. G. Kendrew, *J. Antibiot.*, 2003, **56**, 543–551.
296 Q. Xue, G. Ashley, C. R. Hutchinson and D. V. Santi, *Proc. Natl. Acad. Sci. U. S. A.*, 1999, **96**, 11740–11745.
297 Z. A. Hughes-Thomas, C. B. W. Stark, I. U. Böhm, J. Staunton and P. F. Leadlay, *Angew Chem. Int. Ed. Engl.*, 2003, **42**, 4475–4478.
298 S. Yuzawa, C. H. Eng, L. Katz and J. D. Keasling, *Biochemistry*, 2013, **52**, 3791–3793.
299 A. Hagen, S. Poust, T. de Rond, J. L. Fortman, L. Katz, C. J. Petzold and J. D. Keasling, *ACS Synth. Biol.*, 2016, **5**, 21–27.
300 A. Tripathi, S.-S. Choi, D. H. Sherman and E.-S. Kim, *J. Ind. Microbiol. Biotechnol.*, 2016, **43**, 1189–1193.
301 P. Argyropoulos, F. Bergeret, C. Pardin, J. M. Reimer, A. Pinto, C. N. Boddy and T. M. Schmeing, *Biochim. Biophys. Acta*, 2016, **1860**, 486–497.
302 T. P. A. Hari, P. Labana, M. Boileau and C. N. Boddy, *ChemBioChem*, 2014, **15**, 2656–2661.
303 M. Wang and C. N. Boddy, *Biochemistry*, 2008, **47**, 11793–11803.
304 Z. C. Litman, Y. Chen and D. H. Sherman, *J. Am. Chem. Soc.*, 2025, **147**, 40305–40318.
305 C. C. Aldrich, B. J. Beck, R. A. Fecik and D. H. Sherman, *J. Am. Chem. Soc.*, 2005, **127**, 8441–8452.
306 J. Wu, W. He, C. Khosla and D. E. Cane, *Angew Chem. Int. Ed. Engl.*, 2005, **44**, 7557–7560.
307 W. He, J. Wu, C. Khosla and D. E. Cane, *Bioorg. Med. Chem. Lett*, 2006, **16**, 391–394.
308 D. A. Hansen, C. M. Rath, E. B. Eisman, A. R. H. Narayan, J. D. Kittendorf, J. D. Mortison, Y. J. Yoon and D. H. Sherman, *J. Am. Chem. Soc.*, 2013, **135**, 11232–11238.
309 D. A. Hansen, A. A. Koch and D. H. Sherman, *J. Am. Chem. Soc.*, 2015, **137**, 3735–3738.
310 M. L. Adrover-Castellano, B. J. Curtis, J. J. Schmidt, H. A. Boesger, C. A. Glasser, D. E. Olukorede, F. Qu and D. H. Sherman, *ACS Catal.*, 2025, **15**, 3405–3417.
311 Z. Qi, Q. Kang, C. Jiang, M. Han and L. Bai, *Appl. Microbiol. Biotechnol.*, 2015, **99**, 6745–6752.
312 Z. Qi, Y. Zhou, Q. Kang, C. Jiang, J. Zheng and L. Bai, *Appl. Microbiol. Biotechnol.*, 2017, **101**, 2427–2436.
313 S. Yao, S. Xie, R.-Z. Liu, Z. Huang and L. Zhang, *Nat. Chem. Biol.*, 2025, **21**, 1368–1375.
314 I. G. U. Pait, S. Kitani, F. W. Roslan, D. Ulanova, M. Arai, H. Ikeda and T. Nihira, *J. Ind. Microbiol. Biotechnol.*, 2018, **45**, 77–87.
315 M. W. Mullowney, R. A. McClure, M. T. Robey, N. L. Kelleher and R. J. Thomson, *Nat. Prod. Rep.*, 2018, **35**, 847–878.
316 Y. Li, K. J. Weissman and R. Müller, *J. Am. Chem. Soc.*, 2008, **130**, 7554–7555.
317 U. R. Awodi, J. L. Ronan, J. Masschelein, E. L. C. de Los Santos and G. L. Challis, *Chem. Sci.*, 2017, **8**, 411–415.
318 L. Zhang, J. Xiao, J. Xu, T. Fu, Z. Cao, L. Zhu, H.-Z. Chen, X. Shen, H. Jiang and L. Zhang, *Cell Res.*, 2016, **26**, 1330–1344.
319 A. T. Keatinge-Clay, *Chem. Biol.*, 2007, **14**, 898–908.
320 F. Kudo, *J. Antibiot.*, 2024, **77**, 486–498.
321 L. E. Misson, J. T. Mindrebo, T. D. Davis, A. Patel, J. A. McCammon, J. P. Noel and M. D. Burkart, *Proc. Natl. Acad. Sci. U. S. A.*, 2020, **117**, 24224–24233.
322 I. Nazi, K. P. Koteva and G. D. Wright, *Anal. Biochem.*, 2004, **324**, 100–105.
323 R. S. Gokhale, S. Y. Tsuji, D. E. Cane and C. Khosla, *Science*, 1999, **284**, 482–485.