



**HAL**  
open science

# Genetic engineering of modular PKSs: from combinatorial biosynthesis to synthetic biology

Kira J. Weissman

► **To cite this version:**

Kira J. Weissman. Genetic engineering of modular PKSs: from combinatorial biosynthesis to synthetic biology. *Natural Product Reports*, Royal Society of Chemistry, 2016, 33 (2), pp.203 - 230. 10.1039/c5np00109a . hal-01474079

**HAL Id: hal-01474079**

**<https://hal.univ-lorraine.fr/hal-01474079>**

Submitted on 1 Feb 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Genetic engineering of modular PKSs: from combinatorial biosynthesis to synthetic biology

Kira J. Weissman

UMR 7365, Ingénierie Moléculaire et Physiopathologie Articulaire (IMoPA), CNRS-Université de Lorraine, Biopôle de l'Université de Lorraine, Campus Biologie Santé, 9 Avenue de la Forêt de Haye, CS 50184, 54505 Vandœuvre-lès-Nancy Cedex, France

e-mail: kira.weissman@univ-lorraine.fr

Multienzyme polyketide synthases (PKSs) are molecular-scale assembly lines which construct complex natural products in bacteria. The underlying modular architecture of these gigantic catalysts inspired, from the moment of their discovery, attempts to modify them by genetic engineering to produce analogues of predictable structure. These efforts have resulted in hundreds of metabolites new to nature, as detailed in this review. However, in the face of many failures, the heady days of imagining the possibilities for a truly 'combinatorial biosynthesis' of polyketides have faded. It is now more appropriate to talk about 'PKS synthetic biology' with its more modest goals of delivering specific derivatives of known structure in combination with and as a complement to synthetic chemistry approaches. The reasons for these failures will be discussed in terms of our growing understanding of the three-dimensional architectures and mechanisms of these systems. Finally, some thoughts on the future of the field will be presented.

1. Introduction to polyketides and modular PKSs
2. Theoretical engineering possibilities
3. Practical aspects of PKS genetic engineering
  - 3.1 Genetically engineering native polyketide producers
  - 3.2 Heterologous expression
4. A brief tour of PKS multienzyme genetic engineering
  - 4.1 Domain modification
  - 4.2 Exchanging modules
  - 4.3 Re-engineering subunits
5. Assessment of genetic engineering to date
6. Outlook for the future
7. References

## 1. Introduction to polyketides and modular PKSs

The complex polyketides count among them multiple blockbuster drugs (the antibiotics Biaxin (clarithromycin) and Zithromax (azithromycin) and the immunosuppressants Rapamune (rapamycin, sirolimus) and Prograf/Protopic (FK506, tacrolimus)) (**Fig. 1**), with peak sales in the billions of dollars.<sup>1</sup>

Furthermore, since 2008, five of the 25 natural product derived entities launched into the clinic were reduced polyketides, including fidaxomicin, the first in a new class of narrow-spectrum antibacterials.<sup>2</sup> However, the majority of polyketide natural products are not drugs – that is, they did not evolve to work on targets within the human body. There is therefore a need not only to discover new polyketide structures, but to modify them to increase their suitability as medicines. Indeed, both Biaxin and Zithromax are semi-synthetic derivatives of their parent compound erythromycin A – the introduced chemical modifications served to increase bioavailability and microbiological activity against selected organisms.<sup>3</sup>

The polyketide discovery ‘pipeline’ has been fed to date by two approaches: uncovering novel structures in Nature (an endeavor which has been reinvigorated recently by genome mining<sup>4-6</sup>), and chemical synthesis (either total or semi-synthesis). However, a third complementary strategy was suggested in the early 1990s with the discovery of the first gene cluster in a bacterium encoding the biosynthesis of a complex polyketide. Analysis of the erythromycin A (DEBS) genes in *Saccharopolyspora erythraea*<sup>7,8</sup> revealed that the metabolite is assembled by a gigantic multienzyme megacomplex called a modular polyketide synthase (PKS), in which each task in constructing the molecule is assigned to a discrete catalytic activity (**Fig. 2**). This division-of-labor organization immediately explained how Nature manages to generate large numbers of structurally and stereochemically complex polyketides from a relatively small pool of simple precursors: by dividing the biosynthetic work among many individual enzymes, it becomes possible not only to incorporate a range of different constituents, but to program a specific series of modifications to the building blocks at the point at which they are integrated into the metabolites, include their degree of reduction and the resulting stereochemistry.

Each enzymatic activity within the PKS multienzymes is carried by an independently-folded domain. The domains are grouped into working units called modules, where each module introduces one building block into the growing chain, chemically tailors its functionality, and then hands the intermediate on to the next module in line. Each chain extension module minimally contains an acyl transferase (AT) domain which recruits the next monomer to the synthase (due to the presence of the AT domains within the multienzymes, these systems are known as *cis*-AT PKS, to distinguish them from the *trans*-AT PKS to be discussed below), a ketosynthase (KS) which elongates the chain using Claisen-like chemistry, and an acyl carrier protein (ACP) to which the building blocks and growing intermediates are covalently bound. Polyketide structural diversity largely arises from the inclusion in the modules of optional **processing** activities, including ketoreductase (KR), dehydratase (DH) and enoyl reductase (ER) domains. The KR reduce either the *re*- or *si*- faces of the nascent  $\beta$ -keto functions to give hydroxyl groups, and in many cases also control the stereochemistry at the adjacent  $\alpha$ -methyl centers via epimerization.<sup>9,10</sup> The DH domains can then eliminate water to generate olefinic

moieties which may exhibit either *cis*-<sup>11</sup> or *trans*-<sup>12</sup> configurations, while the ER domains further reduce the double bonds to saturated intermediates, where again two stereochemical outcomes are possible at the adjacent  $\alpha$ -methyl centers.<sup>13</sup> At the end of the synthesis, the finished chain is freed from the multienzyme by a thioesterase (TE) domain by hydrolysis or more commonly macrolactonization using an internal hydroxyl nucleophile.<sup>14</sup> The compounds released from the PKSs are typically not the final products of the pathways, however, as they are often decorated by a series of discrete ‘post-PKS enzymes’ which carry out modifications such as oxidation, glycosylation and methylation, conferring bioactivity on the metabolites.<sup>15</sup> In erythromycin biosynthesis, for example, the first enzyme-free intermediate 6-deoxyerythronolide B is hydroxylated twice and glycosylated with two different sugars (mycarose and desosamine), and the desosamine subsequently modified to cladinose by methylation (**Fig. 2**).<sup>16–20</sup>

It turns out that two distinct types of modular enzymes exist in Nature. The founding member of the second so-called *trans*-AT PKS class was discovered in 1993,<sup>21</sup> but because the gene cluster could not be correlated with a product, it was unclear whether the observed pathway was a non-functional relic of a *cis*-AT PKS.<sup>22</sup> With the identification of several additional ‘aberrant’ systems in the late 1990s/early 2000s,<sup>23–25</sup> it became evident that these so-called *trans*-AT PKSs represented a distinct family of multimodular systems (to-date, some 40 *trans*-AT PKSs have been characterized in detail, while aggregate genomic data suggest that they represent 25% of all assembly line PKSs<sup>26</sup>). The notable feature which distinguishes the *trans*-AT PKSs from the *cis*-AT is that the AT domain is not present within the PKS multienzymes, but is an isolated protein which acts iteratively in *trans* to deliver extender unit to all of the modules (**Fig. 3**). Other distinguishing aspects include the presence of duplicated domains (e.g. doublets or triplets of ACPs), inactive domains and modules, modules for which the component domains are located on at least two proteins, and the intervention in *trans* of large numbers of additional *trans*-acting enzymes (e.g. cassettes which introduce  $\beta$ -methyl groups into the polyketide chains (**Fig. 3**)<sup>27</sup>). These systems also characteristically include modules of nonribosomal peptide synthetase (NRPS) permitting the activation and incorporation of amino acids into the chain-extension intermediates (ca. 50% of *trans*-AT PKS<sup>26</sup>) (**Fig. 3**), a property shared with certain *cis*-AT PKS systems. These fundamental differences between the *cis*- and *trans*-AT PKSs can apparently be explained by their distinct modes of evolution, with *cis*-AT systems arising from repeated rounds of gene duplication coupled with domain diversification via homologous recombination,<sup>28</sup> and the *trans*-AT synthases from the cobbling together of various parts acquired largely by horizontal transfer.<sup>29</sup> On the other hand, the most recent structural data suggest more intimate links between the two types of pathway, with the possibility that at least some *trans*-AT modules evolved through the gradual loss of a parental *cis*-AT.<sup>30</sup>

In any case, for the *cis*-AT PKSs, the direct correspondence between the complement of domains in the multienzymes/the set of post-PKS enzymes and the structures of the final polyketides, makes them (at least in theory) particularly amenable to genetic engineering: all that is required to effect a particular structural change is to identify the domain or post-PKS enzyme responsible, and modify it accordingly. This idea inaugurated the field of PKS 'combinatorial biosynthesis' in the mid-1990s – an endeavor inspired by combinatorial chemistry in which researchers imagined creating vast libraries of novel polyketides by genetic engineering, which could then be subjected to biological testing.<sup>31,32</sup> For the *trans*-AT PKSs, on the other hand, the link between the domain composition of the polypeptides and the products was initially less clear, and so they have not been significantly targeted for manipulation. However, their apparent mode of evolution would seem to render them perhaps even more intrinsically suitable to genetic engineering approaches (a point reviewed in detail in ref. 33).

## 2. Theoretical engineering possibilities

What sorts of changes to PKS multienzymes are possible in theory (**Fig. 4**)? The most obvious starting point is manipulating individual PKS domains. For example, by specific modification of active site residues, the domains could be inactivated or their substrate or stereochemical specificity adjusted. It is also in principle possible to switch one domain for another from the same or a different PKS in order to alter these same properties. For example, a malonate-specific AT could be exchanged for an AT with a preference for methylmalonate thereby introducing a methyl branch into the polyketide backbone, or one KR could be swapped for another in order to change the direction of ketoreduction and/or introduce a 2-methyl epimerization. Processing domains could also be added (gain-of-function) in order to change the extent of reductive tailoring of an intermediate during a specific chain extension cycle. Finally, relocation of the TE domain could cause premature termination of polyketide biosynthesis, thereby acting as a means to control chain length, or one type of TE could be exchanged for another in order to vary the mode of release (for example, from macrocyclization to hydrolysis).

Moving up a structural level, it is also theoretically possible to alter whole modules or indeed entire subunits: deleting them to reduce the size of the chain, exchanging them to modify at once multiple structural features of the intermediate, or adding them to increase overall product length. In view of their critical contribution to bioactivity, the post-PKS enzymes are also attractive targets for genetic engineering; possible modifications include the removal or addition of specific tailoring steps, or changes in the regioselectivity or stereospecificity of the reactions. Evidently, all of the above

changes could in principle be combined to simultaneously introduce several types of structural modifications into the metabolite.

A final approach, referred to as mutasynthesis, is to use genetic engineering to enable PKS pathways to intersect with organic synthesis.<sup>34</sup> This strategy is based on disabling the biosynthesis either by inactivation of an enzyme involved in generating a precursor or by targeting a catalytic domain or domains which operate at an early stage in the pathway (i.e. the loading module, the first KS domain, etc.) (mutasynthesis is thus a genetically-enabled variant of precursor-directed biosynthesis, in which alternative synthetic precursors are simply fed to wild-type strains<sup>35</sup>). Biosynthesis is then jump-started by furnishing analogues of the missing building blocks or more advanced intermediates. If these are loaded onto the PKS and processed by the downstream modules (and post-PKS enzymes), then a novel, bioactive polyketide can result. Mutasynthesis can also be exploited to introduce 'functional handles' into polyketide backbones, which can then be targeted for further diversification by standard chemical methods.

### 3. Practical aspects of PKS genetic engineering

Before summarizing PKS engineering efforts to date, it is worth considering the practical aspects of modifying the systems. Globally, two approaches have been adopted to alter polyketide structures – either manipulating the native producing organism, or transferring the full gene cluster to a so-called 'heterologous host' (Fig. 5). Both of these strategies rely quite heavily on recombination systems borrowed from phage.<sup>36</sup> The choice of engineering approach has largely been dictated by the characteristics of the native producer. These are highly variable, as polyketide clusters are common in at least four distinct bacterial phyla (*Proteobacteria* (e.g. the myxobacteria and pseudomonads), *Actinobacteria* (e.g. the *Streptomyces*), *Firmicutes* (e.g. the Bacilli) and *Cyanobacteria*).

#### 3.1 Genetically engineering native polyketide producers

One significant advantage of targeting the native producer is that the entire biosynthetic pathway is already present, notably including regulatory elements. The host strain is also intrinsically capable of expressing and folding the giant PKS polypeptides, as well as modifying their component ACPs post-translationally with phosphopantetheine prosthetic group to generate the active *holo* forms (a reaction catalyzed by dedicated phosphopantetheinyl transferase (PPTase) enzymes<sup>37</sup>). The native bacterium also contains all of the basic building blocks necessary to supply the biosynthesis, and where appropriate, specific resistance mechanisms against the final product(s).

To date, the majority of modifications in the producing strains have been made by homologous recombination. Most often, introduction of a suicide vector (i.e. one which cannot replicate in the host) carrying a selectable marker and a region of homology to the intended integration site has been used to insertionaly **inactivate** a specific gene (to provide evidence, for example, for pathway identity), with screening for correct integrants based on the acquisition of antibiotic resistance. Subsequent excision of the resistance element has been achieved via flanking recombinase target sites (e.g. *loxP* or *FRT* which are recognized respectively by the Cre and Flp recombinases<sup>36</sup>), allowing for marker recycling as well as improving the stability of the introduced mutation (though a 'scar' sequence remains). Site-specific insertion has also been carried out at bacterial genomic attachment sites (*attB*), by equipping the introduced vector with a phage integrase gene (*int*) and a corresponding phage attachment site (*attP*). This method has been applied almost exclusively in the Actinomycetes, for which elements derived from the *Streptomyces* phages  $\phi$ C31 and VWB<sup>38</sup> have been heavily exploited.

'Clean' genomic modifications can be generated by double homologous recombination, as all vector-derived sequences including the resistance marker are excised in the second step. Screening for double recombinants, which was traditionally a laborious procedure, has been greatly improved by utilization of counter-selectable markers (for example, the *Bacillus subtilis*-derived SacB which is lethal to Gram-negative bacteria in the presence of sucrose<sup>39</sup>), which provide a positive read-out for successful loss of the vector. **A recent development in this area is the use of the protein encoded by *gusA*,  $\beta$ -glucuronidase (GUS), as a method for gene targeting in streptomycetes.<sup>40</sup> Advantages of the system include the fact that most streptomycetes do not carry an endogenous GUS activity, and that its activity is cofactor-independent and can be assayed with a wide range of commercially-available substrates (including, for example, the chromogenic substrate X-Gluc, which forms a blue precipitate, 5,5-dibromo-4,4-dichloro-indigo, after hydrolysis by GUS). In this way, it was possible to develop a facile visual screen for double-cross over events, as clones with double crossovers (those having lost a vector backbone containing a *gusA* gene) when overlaid with X-Glu solution are white, while those containing only a single cross over or the free plasmid, are blue.**

The **recombination** approach is particularly powerful when coupled with 'recombineering' in *Escherichia coli*, as cosmid- or BAC-sized regions of a gene cluster can be modified, and then introduced with high efficiency into the target genome via double-homologous recombination within the resulting long regions of sequence identity (these may exceed 20 kbp). (Recombineering derives from 'recombination-mediated genetic engineering', but the approach is also referred to in the literature as 'PCR-targeting'<sup>41</sup> or the 'REDIRECT system'<sup>42</sup> when applied specifically to *Streptomyces*). Recombineering is, in fact, another phage-derived recombination system comprising one of two equivalent sets of proteins, either RecE/T from  $\lambda$  prophage or Red $\alpha\beta\gamma$  from coliphage  $\lambda$ . The

approach does not depend on suitably placed restriction sites meaning it is highly flexible, it can be used to manipulate large DNA fragments (inserts of up to 80 kbp have been cloned), and only short regions of homology (40–50 bp) are required between the targeting DNA and the destination site, allowing the homology ‘arms’ to be introduced in synthetic oligonucleotides.<sup>43</sup>

It was of course only a matter of time before the transformative CRISPR-Cas9 was applied to polyketide-producing organisms. The discovery and decryption of this bacterial adaptive immune system<sup>44</sup> has led to a veritable explosion of applications in genome engineering,<sup>45</sup> with the technology adapted to date to a range of eukaryotic cells and even whole animals,<sup>46–49</sup> as well as some bacteria including *Escherichia coli*,<sup>50,51</sup> *Streptococcus pneumoniae*<sup>51</sup> and *Lactobacillus reuteri*.<sup>52</sup> In the first three *Streptomyces* examples which were published in 2015, CRISPR-Cas9-based editing plasmids were constructed and used to delete specific PKS genes as well as several complete gene clusters in the native producers, with the overall experiments requiring significantly less time than conventional approaches.<sup>53–55</sup> It is almost certain that this system will be used in the near future to make more subtle changes (e.g. exchange or gain-of-function) to clusters in other Actinomycetes and polyketide-producers, and thus represents a formidable addition to the PKS synthetic biology toolbox.

### 3.2 Heterologous expression

Although it is clearly possible to target native producers, many polyketide-synthesizing bacteria exhibit characteristics unfavorable to routine use in synthetic biology (e.g. poor laboratory growth, inefficient transformation and homologous recombination, etc.) while effective genetic tools (inducible promoters, activators, antibiotic and counter-selection markers, etc.) are only available for a select group of species. It is therefore desirable in many cases to transfer entire clusters to strains which are intrinsically more amenable to genetic manipulation, a process known as heterologous expression (**Fig. 5**). Indeed, many strains cannot be cultured in the laboratory, and in these cases, heterologous expression is the only means to access their encoded chemical diversity.<sup>56</sup> The technically challenging steps of cluster reconstitution can be carried out in a work-horse bacterial strain such as *E. coli* or alternatively by transformation-associated recombination (TAR) in yeast,<sup>56,57</sup> and then the vectors mobilized into a heterologous host or hosts of choice. Once clusters are reassembled, making targeted modifications to PKS genes and other types of genetic changes is relatively straightforward.

On the other hand, it is important to verify on a case-by-case basis that the chosen alternative host is capable of producing the PKS proteins in active form, and contains the necessary precursors to support the biosynthesis. If an appropriately broad-specificity PPTase or specific building blocks

are lacking, then these can be supplied by inclusion of additional genes (e.g. the complete pathway to a specific precursor) in the heterologous expression constructs. To date, no universal 'superhost' capable of supporting any PKS pathway has been identified, although efforts are underway to engineer improved versions of currently-used strains.<sup>58,59</sup> Thus, heterologous expression experiments have often relied on strains phylogenetically close to the original producing organism (e.g. *Streptomyces coelicolor* or *S. lividans* for Actinomycete clusters<sup>60</sup> and *Myxococcus xanthus* for myxobacterial pathways) as this increases the likelihood that native transcriptional elements function properly, translation is efficient, the resulting proteins fold correctly, etc. *E. coli* has also been used successfully to produce complex polyketides, notably including erythromycin A,<sup>61-64</sup> although this required substantial metabolic engineering of the strain.<sup>65</sup>

The practical aspects of heterologous expression have been covered in detail in several excellent, recent reviews,<sup>60,66-68</sup> and so they will only be summarized here. In the earliest versions of these experiments, modular PKS pathways were typically reconstituted on multiple, compatible vectors which were then introduced simultaneously into a target host.<sup>69</sup> The plasmids were then either maintained episomally via compatible replicons, and/or introduced into the host genome via homologous recombination, transposition, or phage-mediated integration (for examples of each of these strategies, see ref. 68). With advances in recombineering, it also became possible to reassemble ('stitch together') entire gene clusters located on several cosmids into single expression vectors.<sup>68</sup> It was shown more recently that large biosynthetic gene clusters (up to 52 kbp) liberated from genomic DNA by restriction digest can be cloned directly into linearized *E. coli* expression vectors. This process, which is mediated by the recombineering RecE/T couple, is referred to as 'linear plus linear homologous recombination (LLHR)', and has been used to transfer 10 PKS and NRPS pathways from *Phototrhhabdus luminescens* into *E. coli*.<sup>70</sup> Similarly, several *Bacillus* gene clusters were recently captured directly from genomic DNA using TAR cloning in yeast, and expressed in *B. subtilis*.<sup>57</sup> A final advance worth mentioning in this context is the use of bifunctional phage P1-derived artificial chromosomes (PAC vectors) in order to create DNA libraries of polyketide-producing organisms in *E. coli*, which then allow transplantation of whole clusters into model actinomycetes; as these vectors can hold as much as 200 kbp of DNA, their use avoids reassembly of even large gene clusters ( $\geq 70$  kbp). In the first example of this approach,<sup>71</sup> the entire 83.5 kbp gene cluster for FK506 (tacrolimus) was transferred from *Streptomyces tsukubaensis* NRRL 18488 via a 130 kbp PAC clone into four different derivatives of *S. coelicolor*. The initially relatively low yields of FK506 (1.2 mg/L) were improved 5-fold by overexpression of an FK506 regulator gene.

In light of the increasing affordability of DNA synthesis and the numerous methods now available for efficient and seamless gene assembly (e.g. sequence- and ligation-independent cloning (SLIC),<sup>72</sup> Gibson isothermal assembly,<sup>73</sup> and Golden Gate assembly<sup>74</sup>), in future, entire gene clusters may be

routinely constructed *de novo* from small DNA fragments, a feat first demonstrated for the erythromycin cluster.<sup>75</sup> The major advantage of this approach would be to enable facile generation of cluster variants as specific regions could be altered at the DNA synthesis stage of the assembly process, or of novel, hybrid assembly lines, by coupling fragments encoding domains, modules or subunits from several PKS systems.<sup>76,77</sup> It would likewise be possible at this point to optimize other elements of the gene cluster for heterologous expression, including codon usage, promoters and terminators, etc.

## 4. A tour of PKS multienzyme genetic engineering

The aim of this section is to summarize the notable successes and failures of PKS genetic manipulation, with a particular emphasis on recent developments (2012–present). Examples will be limited to engineering of the multienzymes themselves, as strategies for modifying post-PKS enzymes<sup>15</sup> and for mutasynthesis<sup>78,79</sup> have been treated extensively elsewhere. No attempt has been made to be comprehensive, but apologies are made if any key experiments have been omitted inadvertently.

### 4.1 Domain modification

Altering individual domains is to date the most frequent type of modification made, and all theoretical changes mentioned earlier have been achieved. In fact, the first successful PKS engineering experiment involved inactivation of specific reductive domains: by disrupting cofactor binding in KR and ER domains of DEBS, several analogues were produced bearing the predicted modifications.<sup>8,80</sup> The success of these experiments presumably derives from the ‘minimally invasive’ nature of inactivation mutagenesis, which is unlikely to introduce major structural modifications into the PKS multienzymes. On the other hand, a contemporaneous attempt to generate a DEBS DH mutant failed to produce full-length derivative, even though only a single active site His was altered.<sup>81</sup> More recently, disabling of KR, DH and ER domains has been used to produce a library of 22 oxidized derivatives of premonensin **(the highest yields obtained were ca. 10-fold reduced from the wild type metabolites)** (Fig. 6), a shunt metabolite of the monensin pathway which structurally resembles the anti-cancer polyketide discodermolide<sup>82</sup>, as well as to introduce modifications within the polyene class of antifungal polyketides, leading to improved solubility and reduced hemolytic activity.<sup>83</sup>

Another notable inactivation strategy is to target a specific AT domain within a full-length PKS and to complement the missing activity in *trans* with an AT of alternative specificity. By using a malonate-

specific AT to rescue an inactivated methylmalonate-specific AT of the DEBS PKS expressed heterologously in *E. coli*, exclusively the expected desmethyl derivative of 6-deoxyerythronolide B was obtained at yields comparable to the wild type.<sup>84</sup> More recently, several discrete AT domains from *trans*-AT PKS systems were used to complement an inactive AT *in vitro* in a reconstituted DEBS PKS.<sup>85</sup> The *trans*-acting AT domains were again able to deliver extender units to the PKS resulting in successful incorporation of both malonate and ethylmalonate in place of the native methylmalonate, albeit with loss of efficiency in the case of ethylmalonate. AT complementation was also employed to achieve directed incorporation of fluorine atoms into polyketides in *E. coli*.<sup>86</sup> In this case, the AT of module 6 of DEBS was disabled via an active site Ser to Ala mutation. The resulting module was able to use fluoromalonyl-CoA as extender unit (efficiency was in fact higher than with malonyl-CoA), presumably via direct loading of the relatively more active building block onto the ACP of module 6. However, complementation with the broad-specificity *trans*-acting AT from the disorazol PKS improved the yield of fluoroketide. This strategy was subsequently extended to the AT domains of DEBS modules 2 and 3 within the context of a mini-PKS (M2-M3-TE), giving rise to both fluorinated regioisomers of the expected tetraketide lactone.

Changes in substrate choice have also been achieved by targeting putative specificity-determining residues within AT domains. For example, comparative sequence analysis of malonate- and methylmalonate-specific ATs identified several sequences correlating with substrate preference (among them, an active site motif<sup>87,88</sup> which is YASH for methylmalonate-specific ATs and HAFH for malonate-specific ATs).<sup>89,90</sup> However, exchange of these key motifs *in vivo* in order to convert methylmalonate-specific ATs into malonate-specific domains produced only an incomplete shift in extender unit choice, indicating that further elements of the AT active site contribute to specificity as well as a 10–20% drop in yield. To address this question, computational modeling of AT<sub>5</sub> of the erythromycin PKS was recently used to help identify additional amino acids which might participate in substrate selection.<sup>91</sup> This analysis revealed a Val residue which was not highlighted by sequence analysis, but which when mutated allowed utilization of 2-propargylmalonate as extender unit (Fig. 7). Thus a single point-mutation sufficed to expand AT specificity to encompass a building block suitable for subsequent chemical modification, although at least for the moment, the yields remain outside of the realm of synthetic utility.

Site-directed mutagenesis has also been used to alter the stereospecificity of both keto- and enoyl reduction. In the case of the KR domains of *cis*-AT PKSs, efforts have focused on amino acids which by sequence analysis show a strong correlation with the direction of reduction.<sup>92,93</sup> KRs catalyzing *re*-type reduction to generate a L-3-hydroxyacyl intermediate are classified as ‘A-type’, while those reducing from the opposite *si*-face to give a D-3-hydroxyacyl chain are ‘B-type’. The strongest predictor of B-type ketoreduction is a Leu-Asp-Asp (‘LDD motif’) situated in a loop near the active

site, while A-type KRs lack this motif, and instead typically exhibit a conserved Trp ('W motif') which is located on the opposite side of the active site. These residues have been suggested to interact with the  $\beta$ -keto intermediate in order to direct it into one or the other side of the active site, positioning it appropriately relative to the NADPH cofactor.<sup>9</sup> Recently, however, an alternative mechanism was proposed on the basis of detailed structural comparison of all available A- and B-type KRs:<sup>94</sup> the active sites of A-type KRs are structured in the absence of substrate, with the lid loop and lid helix which lie adjacent to the NADPH allowing access to only one side. In this scheme, the conserved W might serve to orient the phosphopantetheine of the incoming substrate via hydrogen bonding. In contrast, it is suggested that B-type KR active sites only become structured when the chain enters from the opposite side, with the LDD motif contributing to this substrate-assisted assembly.

Certain KRs also catalyze epimerization at the adjacent 2-methyl centers, even when they are inactive as ketoreductases.<sup>10,95</sup> This has led to the classification of KR domains into six types: those catalyzing ketoreduction but not epimerization (A1 and B1), ketoreduction and epimerization (A2 and B2), epimerization but not reduction (C2) and fully inactive KRs (C1). However, no corresponding active site motifs can reliably be correlated as yet to the epimerization activity. Alteration of KRs from the DEBS system whether by exchanging A- and B-type motifs<sup>96</sup> or by saturation mutagenesis at these same positions,<sup>97</sup> resulted in altered stereocontrol when isolated KR domains were assayed *in vitro* with surrogate substrates, albeit with significant loss of activity (60–80%<sup>96</sup>). However, the equivalent mutations introduced into a model DEBS-derived PKS *in vivo* did not result in the expected changes in stereochemistry.<sup>98</sup> Thus, additional factors likely contribute to KR stereocontrol within intact multienzymes.

Active-site surgery on ER domains which set the stereochemistry of certain methyl branches (Fig. 8) has also yielded mixed results. Comparative sequence analysis revealed a strong relationship between a single active site residue and the direction of enoyl reduction – a highly-conserved active site Tyr<sup>99</sup> is present in ERs generating a (2S)-methyl configuration, while the corresponding position in ERs giving rise to (2R)-methyl groups is most often a Val (Fig. 8).<sup>13</sup> Notably, a Tyr-to-Val mutation in a model bimodular PKS based on DEBS produced a complete switch in the methyl branch configuration from (2S) to (2R), with no appreciable drop in yield. However, a Val-to-Tyr substitution at the equivalent position in a (2R)-specific ER failed to yield the (2S)-methyl product. Thus, as with PKS KR domains, the full set of stereo-determinants in ER domains remains to be identified.

Complete changes in the direction of ketoreduction have been achieved by a more dramatic domain-level modification, which is to exchange one domain for another. Several proof-of-principle KR swap experiments were published,<sup>100,101</sup> before this strategy was significantly enabled by a 'reductive loop swapping' (RLS) approach, in which a specific A1-type KR domain within a DEBS-derived model PKS was replaced with an oligonucleotide ('polylinker') containing multiple restriction

sites (**Fig. 9**).<sup>102</sup> In this way, it was possible to graft in a range of KR domains from the same and different PKS systems, as well as to vary the fusion points with the parent synthase within the linkers separating the KR domains from the upstream AT and downstream ACP domains. Introduction of both A1- and B1-type KR domains gave good to essentially wild type yields of the expected products (depending on the KR used and the specific fusion sites), consistent with the idea that the direction of ketoreduction is intrinsic to the domains. On the other hand, insertion of one example each of A2- and B2-type KR domains resulted in only traces of the epimerized products, while the bulk of polyketide produced derived from the absence of reduction. This contrasted with the finding *in vitro* with reconstituted modules that placing an epimerizing KR into an otherwise non-epimerizing context could result in a significant proportion of 2-methyl epimerization.<sup>103</sup> More recently, examination *in vivo* of a much wider panel of A2- and B2-type KR domains within the RLS model, revealed several A2-type KR domains which gave the epimerized products with significantly higher efficiency.<sup>104</sup> However, mixtures of products were always observed, either due to lack of epimerization or absence of ketoreduction in some cycles. Furthermore, the B2-type KR domains were largely recalcitrant to relocalization, with only two of the transplanted six domains giving traces of the epimerized product. Why B1 KR domains should be functional in an A1 context and not B2 KR domains remains to be established. Globally, however, the results suggest that to obtain both efficient ketoreduction and a change in 2-methyl group stereochemistry will require swapping complete modules instead of just the responsible KR domains.

Specific domain exchange has also been extensively employed with ATs in order to alter substrate choice (see ref. 105 for a summary of such experiments). In fact, the first hybrid PKS incorporating domains from two distinct assembly lines was obtained by swapping a DEBS AT for an AT from the rapamycin (RAPS) system.<sup>106</sup> Indeed, in this way it has been possible to modify all six of the extender units incorporated into the erythromycin A backbone,<sup>107-110</sup> to generate ethylmalonate extender, it was also necessary to metabolically engineer precursor supply by introducing a crotonyl-CoA reductase-carboxylase (*ccr*)<sup>111</sup> capable of transforming crotonyl-CoA into (2S)-ethylmalonyl-CoA.<sup>112</sup> However, the price of success in these experiments has been minor to substantial reductions in overall yield. Individual KR domains have also been exchanged for sets of reductive domains (i.e. DH-KR and DH-ER-KR). This 'gain-of-function' type of experiment was first used to convert a hydroxyl group in a tetraketide to a double bond by exchanging a KR for a DH-KR didomain within a DEBS-derived trimodular PKS.<sup>113</sup> The authors then went on to generate a tetraketide capable of cyclizing to form an 8-membered ring lactone:<sup>114</sup> introduction of a DH-ER-KR segment from the RAPS PKS in place of a KR in the second module of the mini-PKS eliminated a nucleophile competing with the terminal hydroxyl, thus allowing exclusive formation of an 8-membered instead of a 6-membered ring at yields of ca. 20 mg/L (**Fig. 10**) (by comparison, insertion of DH-ER-KR from DEBS module 4 resulted in the corresponding unsaturated compound (<10 mg/L), due to failure of the ER to act). A more

extensive set of experiments was carried out with the RLS model PKS, resulting in multiple products bearing the expected double bond or fully-reduced methylene, respectively, at fair to good yield **relative to the parental system**.<sup>102</sup> However, incompletely-reduced and non-reduced polyketides were also obtained, evidence that, as observed previously, the introduced domains were skipped in some cycles.

AT and KR exchanges within modules 2, 5 and 6 of the DEBS system were combined with gain-of-function experiments (both substitution of DH-KR and DH-ER-KR for KR domains), and deletion mutagenesis (replacement of KR domains with a synthetic 18 aa fragment) in order to generate a library of 6-deoxyerythronolide B (6-dEB) derivatives in several heterologous *Streptomyces* hosts.<sup>109</sup> The sites used were located at the boundaries of the AT and KR domains, and were shown not to effect the overall yields of 6-dEB. The resulting >50 6-dEB derivatives purportedly incorporated modifications at branching centers, to the degree of reduction at one or multiple locations in the macrocycle, and of stereochemistry (**Fig. 11**). On the other hand, structural verification was obtained largely by correlating mass spectrometry data with the predicted structures, and so neither the location of the introduced modifications nor the stereochemistry of the products was conclusively demonstrated. In addition, there was a strong correlation between the number of introduced changes and the yields, with those arising from double and triple mutants (<0.1 mg/L) dramatically below that of the wild type system (20 mg/L). A similar strategy was subsequently applied but with the three DEBS genes located on separate, compatible plasmids, yielding many of the same metabolites produced by the earlier swap experiments, but 15 additional derivatives, as judged by mass spectrometry.<sup>69</sup> However, as in the original experiments, yields dropped as the number of mutations increased, and a number of side products were observed, typically corresponding to lack of action by an inserted reductive domain.

Relocation of the DEBS TE domain to the C-terminus of various ACPs has also proved a productive strategy, with the consequence of causing premature release of the polyketide chain. This approach was applied initially by moving the TE from its location at the end of the third subunit, DEBS 3, to the C-terminal extremity of the first (DEBS 1), where it was fused to the ACP of the second chain extension module (**Fig. 9**).<sup>115,116</sup> In its new context, the TE accepted the triketide generated by module 2 (instead of its native heptaketide), cyclizing it to a triketide lactone (**yield of 10–15 mg/L, only slightly less than the amount of erythromycin produced by the wild-type strain**). The resulting DEBS 1-TE mini-PKS served subsequently as a powerful model system for efforts to understand a number of mechanistic issues in polyketide biosynthesis.<sup>102,104,117–120</sup> The same TE was moved to three other positions with DEBS – to the end of module 1 to yield the expected diketide (**at 30 mg/L**),<sup>121</sup> as well as to the termini of modules 3 and 5, giving respectively tetraketide and hexaketide products.<sup>122</sup> Fusion to the DEBS TE has also been employed to off-load intermediates in at least three other PKS

pathways, spinosyn (at 1 mg/L) (Fig. 12),<sup>123</sup> monensin (44 mg/L),<sup>124</sup> and lipomycin,<sup>125</sup> in order to decipher specific mechanistic questions. In the spinosyn case, for example, it was shown that a hydroxyl group introduced by ketoreduction during chain extension is likely reoxidized to the ketone at the post-PKS stage in order to increase the acidity of the adjacent proton, allowing for a series of reactions culminating in a Diels-Alder cyclization. What is particularly notable about the TE relocation experiments relative to other types of domain-level modification is the relatively high yields of the expected products.

## 4.2 Exchanging modules

Moving up a structural level, a number of experiments have successfully exchanged whole modules, with the replacement modules sourced from other PKS systems. For example, the loading module of DEBS has been swapped with that from the avermectin (Ave) PKS of *Streptomyces avermitilis* (derivative yields ca. 2 mg/L) (Fig. 13),<sup>126</sup> those of both DEBS and Ave used to replace the loading module of the spinosyn assembly line (15–25 mg/L),<sup>127</sup> and the spiramycin loading module exchanged with that from the tylosin PKS.<sup>128</sup> In the DEBS case, this modification resulted in several novel erythromycin derivatives from the native producing strain *Sac. erythraea*, derived from incorporation of the branched-chain starter units isobutyrate and 2-methylbutyrate characteristic of avermectin in place of the native propionate and acetate. The intrinsically broad specificity of the Ave loading module<sup>129</sup> was further exploited by supplying the recombinant *Sac. erythraea* strain with a library of short chain fatty acids, resulting in 12 novel erythromycin derivatives which retained modest antibacterial activity.<sup>130</sup>

By disabling the primary metabolic pathways leading to the native starter units, the inherent promiscuity of the Ave loading module was also used to generate a number of avermectin derivatives in *S. avermitilis*. The most effective antiparasitic among them,<sup>131</sup> doramectin (Dectomax), incorporates a cyclohexane carboxylic acid (CHC) as starter unit. It is, in fact, a showcase for polyketide pathway engineering, as it is the first drug to be derived from this approach, and is now one of the most successful veterinary anthelmintics.<sup>132</sup> Recently, an attempt was made to engineer an alternative doramectin producer (Fig. 14) by exchanging the Ave loading module with that from the phoslactomycin PKS<sup>133</sup> whose native substrate is CHC-CoA, coupled with introduction of a CHC-CoA biosynthetic pathway.<sup>131</sup> The resulting strain did yield doramectin in amounts higher than that derived from supplementing the wild type strain with CHC (53 vs. 9 µg/mL), but which were overall significantly reduced relative to avermectin production from the parent (500 µg/mL) – and native avermectin was also produced. Thus this system will require further improvements in order to be suitable for industrial-scale production of doramectin.

Chain extension modules of several PKS have also been modified. One of the first demonstrations of this approach was to generate several bimodular mini-PKS, in which the second module of DEBS 1 was substituted by downstream modules from DEBS (3 and 6), as well as module 5 from the rifamycin (Rif) PKS (**Fig. 15**).<sup>134</sup> The critical parameter in each of these experiments was to maintain the native linker region joining modules 1 and 2 within DEBS 1. The approach was subsequently extended by showing that Rif module 5 could communicate successfully with the next subunit, DEBS 2, when engineered to incorporate the docking domain<sup>135</sup> normally present at the C-terminus of DEBS 1 (*vide infra*), resulting in synthesis of 6-dEB in the heterologous host *S. coelicolor* (at a yield of 15 mg/L). Shortly thereafter, work on the DEBS 1-TE system showed that replacing both modules 1 and 2 with counterparts from the RAPS PKS could yield simple analogues of the blockbuster anti-cholesterol statins (yields in the range of 1–2.5 mg/L).<sup>136</sup> Here, several points of fusion between the parental DEBS PKS and the RAPS modules were tested, revealing not only the importance of an intact intermodular linker, but preservation of the native ACP/KS interface between the two modules (i.e. the modular unit can be considered to run from  $AT_n$ - $KS_{n+1}$ ) – a result obtained independently by others.<sup>137</sup> This knowledge was successfully employed to create a hybrid tetraketide synthase.<sup>136</sup> A functional tetraketide PKS was also engineered by fusing covalently the third module of DEBS to DEBS 1, giving product at yields comparable to when the natural mode of communication between modules 2 and 3 is preserved.<sup>137</sup> In contemporaneous work, module 6 of DEBS was replaced with module 6 from the pikromycin (Pik) system within the context of the full-length assembly line, and again maintenance of a native Pik  $ACP_5/KS_6$  interface proved critical to obtaining product (yields of ca. 3 mg/L).<sup>138</sup>

An attempt was also made to insert a module within a DEBS subunit, with the aim of increasing the length of the final polyketide chain by two carbons.<sup>139</sup> As initial proof-of-principle (**Fig. 16**), RAPS extension modules (2 and 5) were covalently spliced between the first two chain extension modules of DEBS 1-TE. The resulting expanded PKSs did give rise to the predicted tetraketide product (1–2 mg/L), but the major product in both cases (ca. 20-fold more) was triketide resulting from ‘skipping’ of the interpolated modules. The yields also dropped relative to the native PKS, consistent with the fact that both the RAPS modules and DEBS module 2 were confronted with non-native substrate. Attempts to improve communication between the DEBS and RAPS modules by preserving the native ACP/KS interfaces did not improve the obtained amount of tetraketide. The equivalent modification was then made to the entire DEBS assembly line, leading to production of the expected 16-membered ring-expanded macrolide (ca. 3 mg/L), but again the major product (14-membered macrolactone, 10 mg/L) corresponded to efficient bypassing of the introduced module. The mechanistic basis for these observation was subsequently elucidated using the DEBS 1-TE model system, showing that instead of being transferred from  $ACP_n$ (DEBS) to  $KS_{n+1}$ (RAPS) of the interpolated

module, the polyketide chain is handed off to ACP<sub>n+1</sub>(RAPS), thus skipping the chain extension reaction (**Fig. 16**).<sup>140</sup> In the native situation, transacylation from ACP<sub>n</sub>(DEBS) to KS<sub>n+1</sub>(DEBS) occurs preferentially, but may be slowed in the hybrid PKS due to poor substrate recognition by KS<sub>n+1</sub>(RAPS), allowing direct ACP<sub>n</sub>-to-ACP<sub>n+1</sub> transfer to become competitive.

In recent work, whole module engineering has been used to elucidate aspects of biosynthesis by a related family of PKSs producing the nitrosubstituted polypropionate aureothin (aur), neoareothin (nor, also known as spectinibilin) and luteoreticulín.<sup>141,142</sup> In the aur<sup>143</sup> and nor cases, the first module of the assembly lines is capable of catalyzing multiple rounds of chain extension (typically two, but as many as four<sup>144</sup>), a programmed iterative behavior noted for several other PKS systems (e.g. borrelidin,<sup>145</sup> DKXanthene,<sup>146</sup> stigmatellin,<sup>147</sup> etnangien,<sup>148</sup> and crocacin<sup>149</sup>). In an initial experiment, several of the nor PKS modules were deleted in a heterologous host in order to create an aur synthase, recapitulating the presumed evolutionary event that gave rise to the aur PKS (**Fig. 17**).<sup>150</sup> However, the recombinant PKS was found to be inactive even though the native ACP/KS interfaces had been preserved, presumably due to incompatible substrate specificity of downstream modules. A new construct was therefore created in which only the third extension module was lacking, again maintaining the native KS/ACP contacts. This gave rise to a novel product called homoaureothin (at **0.33 mg/L**), the missing triene member of the aureothin family. Traces of neoareothin were also generated due to three cycles of chain extension by the first module, showing that its repetitive action can compensate for the lack of one downstream module. Overall, access to four members of the polypropionate family allowed analysis of the relationship between backbone length and activity, showing that the longest congener exhibited the strongest cytotoxic effects, while short chain length improved antifungal properties. In a related study, net module deletion coupled with AT exchange was used to evolve an aur PKS into a luteoreticulín synthase (yield of luteoreticulín = **0.12 mg/L**), the first time that a PKS has been reprogrammed to generate a polyketide from a different bacterium.<sup>142</sup>

### 4.3 Re-engineering subunits

Another successful engineering approach has been to generate novel combinations of subunits comprising single or multiple modules. In many cases, these experiments have been underpinned by our increasing understanding of the mechanistic basis for recognition between native PKS polypeptides. These interactions are mediated, at least in part, by short, independently-folding regions at the extreme C- and N-termini of the subunits, which are referred to as docking domains (DDs).<sup>135</sup> The initial important studies in this area were to show that these regions are modular and so can be substituted by other such partners without impairing intersubunit communication<sup>151,152</sup> (a result confirmed by several independent studies<sup>153,154</sup>) and that they can mediate contacts between

modules which do not normally interact (as in the Rif module 5 case mentioned earlier).<sup>134,155</sup> Subsequently, bioinformatic analysis identified three distinct sequence classes of docking domains in PKS systems (although two of these have subsequently been combined into class 1),<sup>156</sup> while a fourth class was discovered to operate at junctions between NRPS proteins and their partners (both downstream NRPS and PKS subunits) in mixed *cis*-AT PKS-NRPS systems.<sup>157</sup> High-resolution structural information is now available for all of these docking domains (**Fig. 18**),<sup>135,154,157,158</sup> showing the DD classes to adopt distinct structures, definitively identifying their boundaries, and highlighting residues which might be targeted in attempts to alter interaction specificity. Nonetheless, in the case of PKS/PKS interfaces, it appears likely that contacts between the flanking ACP and KS domains which engage in acyl transfer also contribute to specificity and the affinity of the overall docking interactions.<sup>159,160</sup>

The intrinsic capacity of evolutionarily-related docking domains to communicate with each other was demonstrated by several subunit exchange experiments. For example, hybrid synthases were constructed in the heterologous host *S. lividans* by combining the first two subunits from the Pik PKS with the third subunits of the DEBS and oleandomycin (Ole) assembly lines (**Fig. 19**).<sup>138</sup> The resulting chimeric PKSs produced the expected 14-membered macrolides at yields comparable to the wild type system in *S. lividans*, demonstrating the inherent ability of the subunits to interact productively. The approach was subsequently extended to DEBS subunits incorporating various modifications (domain deletions, and substitutions of AT and KR domains with activities from the RAPS PKS), with some success in obtaining the anticipated analogues. Similarly, combining the DEBS 1 subunit with subunits PikAII–PikAIV of the Pik PKS produced the anticipated hybrid metabolites albeit at low yields ( $\mu\text{g/L}$ ), but additionally significant quantities ( $>11 \text{ mg/L}$ ) of triketide lactones arising from an unanticipated communication between subunits DEBS 1 and PikAIV.<sup>161</sup> Replacing the C-terminal docking domain of DEBS 1 with that of PikAI dramatically reduced lactone production, demonstrating the critical role of this region in mediating the interaction. Comparative sequence analysis of the C-terminal DDs of DEBS 1, PikAI and PikAIII<sup>161</sup> shows that the DEBS 1 DD is overall more similar to the PikAIII DD than to the PikAI DD (particularly across the critical docking  $\alpha$ -helix<sup>135</sup>), potentially explaining this result. Finally, reliance on the inherent similarity between the PKSs biosynthesizing avermectin and milbemycin (Mil) allowed for replacement of the first avermectin subunit AveA1 with MilA1, within an avermectin industrial strain.<sup>162</sup> This resulted in high yields ( $3400 \text{ mg/L}$ ) of two novel metabolites called the tenvermectins, which show enhanced activity relative to avermectin against certain insects (**Fig. 20**).

An analogous set of experiments was carried out with synthases responsible for biosynthesis of the structurally related 16-membered macrolides tylosin (Tyl), spiramycin (Srm) and chalomycin (Chm), in a heterologous host derived from a mutant of *Streptomyces fradiae*. In this strain, the

tylosin PKS-encoding genes were inactivated or absent, but those for the post-PKS enzymes were present.<sup>163</sup> Hybrid systems were created by combining the loading and first three chain extension modules of the Chm pathway with the final four extension modules of either the Tyl or Srm PKSs. To encourage efficient communication between the Chm and Spr/Tyl polypeptides, the docking domain at the C-terminal end of Chm module 3 (subunit ChmGII) was replaced with the corresponding elements from either the Spr (from subunit SrmGII) or Tyl (from subunit TylGII) systems. In the case of the Chm/Spr chimera, a novel compound was obtained at yields on par with those of tylosin from the parental strain. However, as the control was not carried out (i.e. to determine if unmodified ChmGII was also capable of communicating effectively with SrmGIII using its native docking domain), the effect of the docking domain engineering is unclear. Another notable result of these experiments is that the majority of post-PKS tailoring enzymes of tylosin biosynthesis acted efficiently on the hybrid metabolites, demonstrating the useful tolerance of these catalysts to structural variation.

In extensions of this work, docking domain engineering was used to enable communication between less closely-related proteins. For example, a two-plasmid system was created allowing for co-expression of DEBS 1 and a single downstream module fused to the DEBS TE to enable off-loading.<sup>164</sup> To facilitate chain transfer to the acceptor module from DEBS 1, the N-terminal docking domain of DEBS 2 was incorporated at its N-terminus. Using this system, successful communication was observed between DEBS 1 and modules from the Pik and rifamycin (Rif) PKSs. Interactions were also established between DEBS 2 (module 4) and two non-native modular partners (DEBS module 6 and Rif module 5) using the same strategy (i.e. introduction of the N-terminal DD of DEBS 3 in front of the partner modules), resulting in production of the expected hexaketide macrocycles. In the same vein, it was shown that by suitable transplantation of docking domains, subunits EpoA and EpoB from the epothilone mixed PKS-NRPS could be replaced *in vitro* with ACPs from the rapamycin and enterobactin biosynthetic pathways (**Fig. 21**).<sup>165</sup>

A variant strategy which is also worth mentioning is not to induce cross-talk between existing multienzymes, but to use docking domain engineering to create two smaller polypeptides from a larger subunit. This idea has been demonstrated using the Pik system, in which a matched pair of docking domains from the phoslactomycin PKS was introduced in place of two intermodular linkers within the subunit PikAI (**Fig. 22**).<sup>166</sup> The resulting yields of pikromycin were reduced by only 50% relative to when PikAI was intact. This is encouraging for the prospect of engineering PKS assembly lines which comprise very large proteins (for example, RAPS<sup>167</sup>), as it shows that they might be broken down into more manageable functional units by insertion of docking domains at specific locations.

The power of the subunit engineering approach was dramatically increased with the development of a system allowing the facile interchange of modules sourced from many different PKS (the so-

called 'legoization'<sup>168</sup> of polyketide biosynthesis). Here, gene synthesis was employed to create two compatible plasmids, the first encoding the DEBS loading module and a variable 'donor' module connected to it by a linker region, and the second an 'acceptor' module fused covalently to the broad-specificity DEBS TE domain (**Fig. 23**).<sup>169</sup> Communication between the two modules was assured by inclusion at the respective termini of compatible docking domains, also derived from the DEBS PKS. Evaluation of 154 bimodular combinations created by combining 14 modules from eight PKS clusters in an *E. coli* strain adapted to polyketide synthesis,<sup>170</sup> revealed that nearly half of the bimodular combinations gave the expected triketide lactone product (yields 0.02–23 mg/L). This experiment thus showed the intrinsic capacity of the investigated chain extension modules to function out of context and on analogues of their native substrates, in the presence of suitable intermodular elements (linkers and docking domains). In follow up work, it was shown that non-productive modular combinations could be rescued by preserving intact the original ACP<sub>n</sub>/KS<sub>n+1</sub> interfaces,<sup>171</sup> as predicted by the earlier studies.<sup>136,137</sup> Using the design rules identified from the study of the bimodular PKS (i.e. if module A is capable of communicating with B and B with C in bimodular contexts, then it should be possible to create a functional trimodular system comprising A, B and C, in that order), the system was subsequently extended to synthesize tetraketides of predicted structure, with an impressive 95% success rate (as judged by product synthesis at any yield).<sup>172</sup> However, metabolite amounts decreased as a function of the number of unnatural intermodular junctions that were introduced, which would seem to limit the possibility to extend the synthases to generate full-length polyketides.

## 5. Assessment of genetic engineering to date

As is clear from the previous sections, PKS engineering can produce polyketide analogues, many of which are new to Nature. However, we are far from being able to realize the optimistic goals of 'combinatorial biosynthesis',<sup>32</sup> which aimed to generate large libraries of novel compounds in amounts sufficient for biological testing. Indeed, the vast majority of modification experiments reduce the yields of polyketide relative to the parental systems, often drastically, and many simply fail (although the full extent of this is difficult to assess from the available literature). It now seems more reasonable to speak of PKS synthetic biology with the objective of generating specific derivatives of known polyketides at acceptable yields. Desirable structures could be identified, for example, by *in silico* screening of a biological target against a virtual library of analogues. Alternatively, a new derivative might be produced by classical semi-synthesis, but it might then be

possible to develop a synthetic biology route to the same compound which is more efficient and higher yielding and has the added advantage of being environmentally friendly.

For future engineering efforts, what lessons emerge from the complete set of PKS engineering efforts to date? Although modifying individual domains has been a central strategy, the various approaches have met with only mixed success. Inactivation – the least invasive method – has yielded the best results, but its potential for introducing large structural changes is limited. Active-site surgery remains attractive as a means to alter both substrate specificity and stereochemistry as it too introduces modest structural perturbation to the PKS, but this approach is currently hampered by our incomplete understanding of the molecular basis for these key aspects of the biosynthetic pathways. Indeed, a recent study has suggested that mutated AT domains recognize alternative extender units only because their substrate specificity and overall activity is compromised.<sup>173</sup> Further high-resolution data on the catalytic domains, particularly in the presence of substrates or their analogues which to have to date been largely lacking, would be highly informative. On the other hand, poor overall yields of an engineered polyketide can also derive from the low activity of domains downstream from that which was modified, perhaps also necessitating optimization of their substrate specificity towards the non-native intermediates, replacement by broader specificity domains/modules, etc.

As for domain swapping approaches, these are inherently more disruptive, not only because they result in hybrid linker regions flanking the exchanged domains, but because they also potentially modify interdomain interfaces. This is important, as the recent cryo-electron microscopy structures of Pik module 5 modified to mimic various stages of the chain extension cycle,<sup>160,174</sup> show that interdomain contacts are critical for establishing the functional state of the module, and that transitions between such states rely on evolving interfaces between the catalytic domains. At least a subset of such contacts must be disrupted when domains are replaced with homologues from other modules or PKSs. In support of this, a detailed study into why an AT-swapped module was inactive showed that while the AT introduced using classical linker sites remained functional, communication no longer occurred between the neighboring KS and ACP domains, suggestive of a higher-order change to the overall structure of the module.<sup>175</sup> In future, comparable studies aimed at elucidating the structures of modified (i.e. domain-swapped) modules whether functional or non-functional, could help to reveal the architectural perturbations which arise in these experiments, allowing more rational development of alternative strategies. Not only that, despite lacking strong sequence conservation, the linker regions appear to play critical roles in transmitting conformational information between the domains,<sup>160,174</sup> and these sequences have generally been altered in swapping experiments. Thus, sequence/function relationships are urgently needed for these portions of PKS multienzymes.

Collectively, these observations argue that exchanging intact catalytic units, whether these are modules or subunits, may ultimately be a more productive approach. However, for module swaps within subunits the same set of problems arises, as it has been shown repeatedly that higher yields are obtained by maintaining ACP<sub>n</sub>/KS<sub>n+1</sub> interfaces intact, meaning that fusion points must therefore be chosen between KS and AT domains on either side of the inserted module. The Pik module 5 cryo-EM structure has provided new insights into the KS/AT interface which were not evident from earlier crystallographic studies of isolated KS-AT didomains.<sup>87,88</sup> Specifically, the data suggest that some amino acids that form part of the AT domain in the structures of excised KS-AT may in fact belong to the KS-AT linker in the context of the full module. Incorporating this information into engineering strategies may allow us to improve the future efficacy of module swaps based on AT<sub>n</sub>-KS<sub>n+1</sub> units. Subunit swapping has been carried out successfully by strategic use of multiple classes of docking domains. In theory, it is now possible to engineer PKS systems comprising multiple heterologous subunits, in which the four known structural classes of DD could be deployed at the various inter-peptide interfaces to ensure their orthogonality. Another attractive approach is re-engineering interaction specificity within a single structural class of DD by site-directed mutagenesis of key residues. While candidates have been identified for the various classes,<sup>135,156,158</sup> it has not yet been shown systematically that targeted alteration of these amino acids can lead to a predictable change in partner preference. It thus remains to conclusively elucidate all of the molecular determinants for subunit discrimination in modular PKS systems.

## 6. Outlook for the future

PKS synthetic biology holds significant promise for obtaining designer polyketides. However, realizing the true potential of this approach depends on further deepening our understanding of the molecular genetics, structure and function of both *cis*- and *trans*-AT PKSs, and, critically, converting these insights into more productive strategies for modifying these systems. In this context, certain 'sequence homogenized' domains of *cis*-AT PKS systems – those domains which show significant sequence homology due to substantial recombination within the PKS cluster, but which nevertheless operate on diverse substrates – have been flagged for their potential in domain-swapping approaches, as they are likely to have high intrinsic functional interoperability.<sup>176</sup> Similarly, the naturally mosaic character of *trans*-AT PKS gene clusters due to their highly recombinatorial mode of evolution, strongly argues for their inherent amenability to rearrangement.<sup>33,177</sup>

Another interesting question is whether we will be able to take advantage of the many polyketide diversification chemistries that have been unveiled in recent years –  $\beta$ -methylation (and variants,

leading for example to vinyl chloride and cyclopropane groups<sup>178</sup>), vinylogous chain branching,<sup>179</sup> cyclic ether formation,<sup>180</sup> and decarboxylative chain release,<sup>181</sup> just to name a few – by introducing these types of structural modifications at will into other polyketide metabolites. Similarly, it will be interesting to try to exploit the growing family of *ccr* enzymes<sup>182</sup> and broad-specificity acid-CoA ligases<sup>183–185</sup> to expand the range of extender units that can be incorporated into polyketides, particularly those that might be subject to downstream semi-synthetic chemistry. Such experiments could rely on exchanging a given AT for one of inherently broader specificity, such as exist in several PKS systems.<sup>186</sup> Equally, the scope of the AT inactivation/complementation strategy discussed earlier could be broadened by exploiting ACP-promiscuous<sup>187</sup> ATs of alternative building block specificity, particularly those sourced from the growing number of *trans*-AT PKS systems.<sup>22</sup>

**In future, it will also be worth pursuing approaches** which allow for the rational construction of novel hybrid molecules – not only those formed from portions of multiple native polyketides, but incorporating additional types of building block such as amino acids and fatty acids at defined positions. Polyketide/peptide/fatty acid hybrids exist in nature,<sup>188</sup> but our knowledge of how these distinct biosynthetic systems communicate productively remains scant at present, hampering efforts to rationally engineer such chimaeric assembly lines in the laboratory. Recently, a way forward has been suggested for efficiently generating such ‘conglomerate’ molecules – merging ‘genetic sub-clusters’ encoding the key building blocks in the ways that Nature does.<sup>176</sup>

Another promising area is to further exploit directed evolution as a means to establish sequence/function relationships for PKSs that can be used to guide genetic engineering within the multienzymes. However, high throughput mutagenesis/screening experiments have yet to become mainstream in modular PKS engineering, with only two examples reported at present.<sup>189,190</sup> In the first of these from 2004,<sup>189</sup> DNA family shuffling between the DEBS and Pks loading modules was used to generate a sequence library, with the aim of changing the specificity of building block selection from methylmalonate to propionate (but with no net change in structure of the resulting metabolite). The resulting hybrid loading modules were used to complement a strain in which the corresponding gene had been inactivated, with screening for restoration of antibiotic activity by bioassay. Analysis of 4000 transformants revealed three functional chimeric synthases. However, they contained little DEBS DNA, and so it is likely that biosynthesis in these mutant systems was initiated by direct loading of propionate onto the KS of the first chain extension module. In the more recent work published in 2015,<sup>190</sup> subunits from the same two systems were hybridized by homologous recombination in the yeast *Saccharomyces cerevisiae*, exploiting the fact that the DEBS and Pks genes contain numerous short sections of identical sequence. Recombination, for example, of the mono-modular PkAIII with the bimodular DEBS 3 yielded both active hybrid mono- and bimodular proteins, with the fusion points occurring largely within the conserved catalytic domains

of the first module of DEBS 3 (that which corresponds evolutionarily to PikAIII) (**Fig. 24**). A number of the resulting constructs were biosynthetically competent *in vitro*, yielding the predicted chimeric macrolides when supplied with suitable substrate. One of these was then used to complement a *Streptomyces venezuelae* PikAIII/PikAIV knockout; action by two of the native post-PKS enzymes resulted in dihydropikromycin (**Fig. 24**).

It remains a goal in the long-term to use PKS synthetic biology to produce compounds of interest for which no natural analogues exist.<sup>191</sup> For this, we will need a set of suitable PKS components (domains, linkers, docking domains, modules, subunits, etc.) with appropriately-broad substrate and partner specificity and which are capable of catalyzing unique chemistries, and a set of fundamental design rules for assembling them (covalently or non-covalently) into productive mini-PKSs. On the other hand, the PKS components required for such experiments will likely derive from a variety of organisms, complicating the choice of heterologous host. It would therefore be prudent to assemble the biosynthetic systems in broad host-range vectors, allowing testing of a number of existing, polyketide-friendly production organisms. In parallel, additional heterologous hosts may be created via on-going efforts to minimize genome size by deletion of non-essential regions such as endogenous secondary metabolite clusters; this would in principle direct more cellular resources towards the production of the introduced metabolite(s), and simplify detection and eventual purification. Regulation of biosynthesis is another issue to be tackled in such engineered systems. Here, a 'refactoring' approach is avocated, in which endogenous regulatory elements are replaced with designer regulatory circuits, so that the pathways are under control of the end user.<sup>77</sup>

In summary, while the idea of truly combinatorial biochemistry looks to have gone the way of combinatorial chemistry as reliable means to fill the drug discovery pipeline, PKS synthetic biology aimed at small, focused libraries appears to have a bright future – particularly when allied with traditional chemical approaches, as in the recent generation of 356 new derivatives of the hybrid polyketide-nonribosomal peptide antimycin.<sup>192</sup> The approach will no doubt continue to be bolstered by deepening knowledge of modular PKS structure and function, the exponentially increasing information on PKS gene clusters which permit instructive comparison between pathways as well as providing clues as to the natural modes of evolution of the systems, and to continued advancements in the many technologies underlying genetic engineering.

## 6. Acknowledgements

Work in the author's laboratory is supported by the Agence Nationale de la Recherche, the Centre National de la Recherche Scientifique (CNRS), and the University of Lorraine/the Lorraine Region.

## 7. References

1. Natural product pharmaceuticals - the third generation; Winter 08. *Drug Discovery World* at <<http://www.ddw-online.com/p-92813>>
2. M. S. Butler, A. A. B. Robertson and M. A. Cooper, *M. A. Nat. Prod. Rep.*, 2014, **31**, 1612–1661.
3. K. Peltier, *Conn. Med.* 1992, **56**, 371–374.
4. B. O. Bachmann, S. G. Van Lanen and R. H. Baltz, *J. Ind. Microbiol. Biotechnol.*, 2014, **41**, 175–184.
5. P. Monciardini, M. Iorio, S. Maffioli, M. Sosio and S. Donadio, *Microb. Biotechnol.*, 2014, **7**, 209–220.
6. P. J. Rutledge and G. L. Challis, *Nat. Rev. Microbiol.*, 2015, **13**, 509–523.
7. J. Cortés, S. F. Haydock, G. A. Roberts, D. J. Bevitt and P. F. Leadlay, *Nature*, 1990, **348**, 176–178.
8. S. Donadio, M. J. Staver, J. B. McAlpine, S. J. Swanson and L. Katz, *Science*, 1991, **252**, 675–679.
9. A. T. Keatinge-Clay, *Chem. Biol.*, 2007, **14**, 898–908.
10. A. Garg, C. Khosla and D. E. Cane, *J. Am. Chem. Soc.*, 2013, **135**, 16324–16327.
11. M. M. Alhamadsheh, N. Palaniappan, S. DasChouduri and K. A. Reynolds, *J. Am. Chem. Soc.* 2007, **129**, 1910–1911.
12. J. Wu, T. J. Zaleski, C. Valenzano, C. Khosla and D. E. Cane, *J. Am. Chem. Soc.*, 2005, **127**, 17393–17404.
13. D. H. Kwan, Y. Sun, F. Schulz, H. Hong, B. Popovic, J. C. Sim-Stark, S. F. Haydock and P. F. Leadlay, *Chem. Biol.*, 2008, **15**, 1231–1240.
14. M. E. Horsman, T. P. A. Hari and C. N. Boddy, *Nat. Prod. Rep.*, 2015. doi:10.1039/c4np00148f.
15. C. Olano, C. Méndez and J. Salas, *Nat. Prod. Rep.*, 2010, **27**, 571–616.
16. J. F. Andersen and C. R. Hutchinson, *J. Bacteriol.*, 1992, **174**, 725–735.
17. J. M. Weber, J. O. Leung, G. T. Maine, R. H. Potenz, T. J. Paulus and J. P. DeWitt, *J. Bacteriol.*, 1990, **172**, 2372–2383.
18. S. F. Haydock, J. A. Dowson, N. Dhillon, G. A. Roberts, J. Cortés and P. F. Leadlay, *Mol. Gen. Genet.*, 1991, **230**, 120–128.
19. D. Stassi, S. Donadio, M. J. Staver and L. Katz, *J. Bacteriol.*, 1993, **175**, 182–189.
20. S. Gaisser, G. A. Böhm, J. Cortés and P. F. Leadlay, *Mol. Gen. Genet.*, 1997, **256**, 239–251.
21. C. Scotti, M. Piatti, A. Cuzzonji, P. Perani, A. Tognoni, G. Grandi, A. Galizzi and A. M. Albertini, *Gene*, 1993, **130**, 65–71.
22. J. Piel, *Nat. Prod. Rep.*, 2010, **27**, 996–1047.
23. Y. Paitan, G. Alon, E. Orr, E. Z. Ron and E. Rosenberg, *J. Mol. Biol.*, 1999, **286**, 465–474.
24. G. Huang, L. Zhang and R. G. Birch, *Microbiology*, 2001, **147**, 631–642.
25. J. Piel, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 14002–14007.
26. R. V. O'Brien, R. W. Davis, C. Khosla and M. E. Hillenmeyer, *J. Antibiot.*, 2014, **67**, 89–97.
27. C. T. Calderone, *Nat. Prod. Rep.*, 2008, **25**, 845–853.
28. H. Jenke-Kodama, A. Sandmann, R. Müller and E. Dittmann, *Mol. Biol. Evol.*, 2005, **22**, 2027–2039.
29. 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.
30. J. R. Lohman, M. Ma, J. Osipiuk, B. Nocek, Y. Kim, C. Chang, M. Cuff, J. Mack, L. Bigelow, H. Li, M. Endres, G. Babnigg, A. Joachimiak, G. N. Phillips Jr and B. Shen, *Proc. Natl. Acad. Sci. U.S.A.*, **112**, 12693–12698.
31. P. F. Leadlay, *Curr. Opin. Chem. Biol.*, 1997, **1**, 162–168.
32. K. J. Weissman and P. F. Leadlay, *Nat. Rev. Microbiol.*, 2005, **3**, 925–936.
33. M. Till and P. R. Race, *Biotechnol. Lett.*, 2014, **36**, 877–888.
34. K. J. Weissman, *Trends Biotechnol.*, 2007, **25**, 139–142.

35. R. Thiericke and J. Rohr, *Nat. Prod. Rep.*, 1993, **10**, 265–289.
36. N. Nafissi and R. Slavcev, *Appl. Microbiol. Biotechnol.*, 2014, **98**, 2841–2851.
37. R. H. Lambalot, A. M. Gehring, R. S. Flugel, P. Zuber, M. LaCelle, M. A. Marahiel, R. Reid, C. Khosla and C. T. Walsh, *Chem. Biol.*, 1996, **3**, 923–936.
38. M. Bierman, R. Logan, K. O'Brien, E. T. Seno, R. N. Rao and B. E. Schoner, *Gene*, 1992, **116**, 43–49 (1992).
39. J.-M. Reyrat, V. Pelicic, B. Gicquel and R. Rappuoli, *Infect. Immun.*, 1998, **66**, 4011–4017.
40. M. Myronovskyi, E. Welle, V. Fedorenko and A. Luzhetskyy, *Appl. Environ. Microbiol.*, 2011, **77**, 5370–5383.
41. B. Gust, G. L. Challis, K. Fowler, T. Kieser and K. F. Chater, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 1541–1546.
42. 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.
43. Y. Zhang, F. Buchholz, J. P. Muyrers and A. F. Stewart, *Nat. Genet.*, 1998, **20**, 123–128.
44. M. Jinek, K. Chylinski, I. Fonfara, M. Hauer, J. A. Doudna and E. Charpentier, *Science*, 2012, **337**, 816–821.
45. S. H. Sternberg and J. A. Doudna, *Mol. Cell*, 2015, **58**, 568–574.
46. S. J. Gratz, A. M. Cummings, J. N. Nguyen, D. C. Hamm, L. K. Donohue, M. M. Harrison, J. Wildonger and K. M. O'Connor-Giles, *Genetics*, 2013, **194**, 1029–1035.
47. H. Wang, H. Yang, C. S. Shivalila, M. M. Dawlaty, A. W. Cheng, F. Zhang and R. Jaenisch, *Cell*, 2013, **153**, 910–918 (2013).
48. Z. Hou, Y. Zhang, N. E. Propson, S. E. Howden, L. F. Chu, E. J. Sontheimer and J. A. Thomas, *Proc. Natl. Acad. Sci. U. S. A.*, 2013, **110**, 15644–15649.
49. S. W. Cho, S. Kim, J. M. Kim and J.-S. Kim, *Nat. Biotechnol.*, 2013, **31**, 230–232.
50. Y. Jiang, B. Chen, C. Duan, B. Sun, J. Yang and S. Yang, *Appl. Environ. Microbiol.*, 2015, **81**, 2506–2514.
51. W. Jiang, D. Bikard, D. Cox, F. Zhang and L. A. Marraffini, *Nat. Biotechnol.*, 2013, **31**, 233–239.
52. J.-H. Oh and J.-P. van Pijkeren, *Nucleic Acids Res.*, 2014, **42**, e131.
53. R. E. Cobb, Y. Wang and H. Zhao, *ACS Synth. Biol.*, 2015, **4**, 723–728.
54. H. Huang, G. Zheng, W. Jiang, H. Hu and Y. Lu, Y. One-step high-efficiency CRISPR/Cas9-mediated genome editing in *Streptomyces*. *Acta Biochim. Biophys. Sin.*, 2015, **47**, 231–243.
55. Y. Tong, P. Charusanti, L. Zhang, T. Weber and S. Y. Lee, *ACS Synth. Biol.* (2015). doi:10.1021/acssynbio.5b00038.
56. J. H. Kim, Z. Feng, J. D. Bauer, D. Kallifidas, P. Y. Calle and S. F. Brady, *Biopolymers*, 2010, **93**, 833–844.
57. Y. Li, Z. Li, K. Yamanaka, Y. Xu, W. Zhang, H. Vlamakis, R. Kolter, B. S. Moore and P. Y. Qian, *Sci. Rep.* 2015, **5**, 9383.
58. M. Komatsu, T. Uchiyama, S. Ōmura, S., Cane, D. E. & Ikeda, H. *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 2646–2651.
59. M. Komatsu, K. Komatsu, H. Koiwai, Y. Yamada, I. Kozono, M. Izumikawa, J. Hashimoto, M. Takagi, S. Ōmura, K. Shin-ya, D. E. Cane and H. Ikeda, *ACS Synth. Biol.*, 2013, **2**, 384–396.
60. H. Zhang, B. A. Boghigian, J. Armando and B. A. Pfeifer, *Nat. Prod. Rep.*, 2011, **28**, 125–151.
61. B. A. Pfeifer, S. J. Admiraal, H. Gramajo, D. E. Cane and C. Khosla, *Science*, 2001, **291**, 1790–1792.
62. S. Peirú, H. G. Menzella, E. Rodríguez, J. Carney and H. Gramajo, *Appl. Environ. Microbiol.*, 2005, **71**, 2539–2547.
63. H. Zhang, Y. Wang, J. Wu, K. Skalina and B. A. Pfeifer, *Chem. Biol.*, 2010, **17**, 1232–1240.
64. M. Jiang, H. Zhang and B. A. Pfeifer, *B. A. J. Vis. Exp.*, 2013, e4346.
65. X. Gao, P. Wang and Y. Tang, *Appl. Microbiol. Biotechnol.*, 2010, **88**, 1233–1242.
66. R. H. Baltz, *J. Ind. Microbiol. Biotechnol.*, 2010, **37**, 759–772.
67. J. Piel, *Annu. Rev. Microbiol.*, 2011, **65**, 431–453.
68. S. E. Ongley, X. Bian, B. A. Neilan and R. Müller, *Nat. Prod. Rep.*, 2013, **30**, 1121–1138.

69. Q. Xue, G. Ashley, C. R. Hutchinson and D. V. Santi, *Proc. Natl. Acad. Sci. U. S. A.*, 1999, **96**, 11740–11745.
70. J. Fu, X. Bian, S. Hu, H. Wang, F. Huang, P. M. Seibert, A. Plaza, L. Xia, R. Müller, A. F. Stewart, Y. Zhang, *Nat. Biotechnol.*, 2012, **30**, 440–446.
71. A. C. Jones, B. Gust, A. Kulik, L. Heide, M. J. Buttner and M. J. Bibb, *PLoS One*, 2013, **8**, e69319.
72. M. Z. Li and S. J. Elledge, *Nat. Methods*, 2007, **4**, 251–256.
73. D. G. Gibson, L. Young, R. Y. Chuang, J. C. Venter, C. A. Hutchinson 34d and H. O. Smith, *Nat. Methods*, 2009, **6**, 343–345.
74. C. Engler, R. Kandzia and S. A. Marillonnet, *PloS One*, 2008, **3**, e3647.
75. S. J. Kodumal, K. G. Patel, R. Reid, H. G. Menzella, M. Welch and D. V. Santi, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 15573–15578.
76. C. Hertweck, *Trends Biochem. Sci.*, 2015, **40**, 189–199.
77. M. Cummings, R. Breitling and E. Takano, *FEMS Microbiol. Lett.*, 2014, **351**, 116–125.
78. A. Kirschning, F. Taft and t. Knobloch, *Org. Biomol. Chem.*, 2007, **5**, 3245–3259.
79. A. Kirschning and F. Hahn, *Angew. Chem. Int. Ed Engl.*, 2012, **51**, 4012–4022.
80. S. Donadio, J. B. McAlpine, P. J. Sheldon, M. Jackson and L. Katz, *Proc. Natl. Acad. Sci. U. S. A.*, 1993, **90**, 7119–7123.
81. D. J. Bevitt, J. Staunton and P. F. Leadlay, *Biochem. Soc. Trans.*, 1993, **21**, 30S.
82. S. Kushnir, U. Sundermann, S. Yahiaoui, A. Brockmeyer, P. Janning and F. Schulz, *Angew. Chem. Int. Ed Engl.*, 2012, **51**, 10664–10669.
83. D. Kong, M.-J. Lee, S. Lin and E.-S. Kim, *J. Ind. Microbiol. Biotechnol.*, 2013, **40**, 529–543.
84. P. Kumar, A. T. Koppisch, D. E. Cane and C. Khosla, *J. Am. Chem. Soc.*, 2003, **125**, 14307–14312.
85. B. J. Dunn, K. R. Watts, T. Robbins, D. E. Cane and C. Khosla, *Biochemistry*, 2014, **53**, 3796–3806.
86. M. C. Walker, B. W. Thuronyi, L. K. Charkoudian, B. Lowry, C. Khosla and M. C. Chang, *Science*, 2013, **341**, 1089–1094.
87. 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.
88. Y. Tang, A. Y. Chen, C.-Y. Kim, D. E. Cane and C. Khosla, *Chem. Biol.*, 2007, **14**, 931–943.
89. F. Del Vecchio, H. Petkovic, S. G. Kendrew, L. Low, B. Wilkinson, R. Lill, J. Cortés, B. A. Rudd, J. Staunton and P. F. Leadlay, *J. Ind. Microbiol. Biotechnol.*, 2003, **30**, 489–494.
90. C. D. Reeves, S. Murli, G. W. Ashley, M. Piagentini, C. R. Hutchinson and R. McDaniel, *Biochemistry*, 2001, **40**, 15464–15470.
91. 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.
92. P. Caffrey, *ChemBioChem*, 2003, **4**, 654–657.
93. R. Reid, M. Piagentini, E. Rodriguez, G. Ashley, N. Viswanathan, J. Carney, D. V. Santi, C. R. Hutchinson and R. McDaniel, *Biochemistry*, 2003, **42**, 72–79 (2003).
94. S. A. Bonnett, J. R. Whicher, K. Papireddy, G. Florova, J. L. Smith and K. A. Reynolds, *Chem. Biol.*, 2013, **20**, 772–783.
95. A. Garg, X. Xie, A. Keatinge-Clay, C. Khosla and D. E. Cane, *J. Am. Chem. Soc.*, 2014, **136**, 10190–10193.
96. A. Baerga-Ortiz, B. Popovic, A. P. Siskos, H. M. O’Hare, D. Spiteller, M. G. Williams, N. Campillo, J. B. Spencer and P. F. Leadlay, *Chem. Biol.*, 2006, **13**, 277–285.
97. H. M. O’Hare, A. Baerga-Ortiz, B. Popovic, J. B. Spencer and P. F. Leadlay, *Chem. Biol.*, 2006, **13**, 287–296.
98. D. H. Kwan, M. Tosin, N. Schläger, F. Schulz and P. F. Leadlay, *Org. Biomol. Chem.*, 2011, **9**, 2053–2056.
99. J. Zheng, D. C. Gay, B. Demeler, M. A. White and A. T. Keatinge-Clay, *Nat. Chem. Biol.*, 2012, **8**, 615–621.
100. D. Bedford, J. R. Jacobsen, G. Luo, D. E. Cane and C. Khosla, *Chem. Biol.*, 1996, **3**, 827–831.

101. C. M. Kao, M. McPherson, R. N. McDaniel, H. Fu, D. E. Cane and C. Khosla, *J. Am. Chem. Soc.*, 1998, **120**, 2478–2479.
102. L. Kellenberger, I. S. Galloway, G. Sauter, G. Böhm, U. Hanefeld, J. Cortés, J. Staunton and P. F. Leadlay, *ChemBioChem*, 2008, **9**, 2740–2749.
103. C. R. Valenzano, R. J. Lawson, A. Y. Chen, C. Khosla and D. E. Cane, *J. Am. Chem. Soc.*, 2009, **131**, 18501–18511.
104. T. Annaï, C. Paris, P. F. Leadlay, C. Jacob and K. J. Weissman, *ChemBioChem*, 2015, **16**, 1357–1364.
105. B. J. Dunn and C. Khosla, *J. R. Soc. Interface*, 2013, **10**, 20130297.
106. M. Oliynyk, M. J. B. Brown, J. Cortés, J. Staunton and P. F. Leadlay, *Chem. Biol.*, 1996, **3**, 833–839.
107. 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 (1997).
108. L. Liu, A. Thamchaipenet, H. Fu, M. Betlach and G. Ashley, *J. Am. Chem. Soc.*, 1997, **119**, 10553–10554.
109. R. McDaniel, A. Thamchaipenet, C. Gustafsson, H. Fu, M. Betlach and G. Ashley, *Proc. Natl. Acad. Sci. U. S. A.*, 1999, **96**, 1846–1851.
110. H. Petkovic, R. E. Lill, R. M. Sheridan, B. Wilkinson, E. L. McCormick, H. A. McArthur, J. Staunton, P. F. Leadlay and S. G. Kendrew, *J. Antibiot.*, 2003, **56**, 543–551.
111. T. J. Erb, I. A. Berg, V. Brecht, M. Müller, G. Fuchs and B. E. Alber, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 10631–10636.
112. D. L. Stassi, S. J. Kakavas, K. A. Reynolds, G. Gunawardana, S. Swanson, D. Zeidner, M. Jackson, H. Liu, A. Buko and L. Katz, *Proc. Natl. Acad. Sci. U. S. A.*, 1998, **95**, 7305–7309.
113. R. McDaniel, C. M. Kao, H. Fu, P. Hevezi, C. Gustafsson, M. Betlach, G. Ashley, D. E. Cane and C. Khosla, *J. Am. Chem. Soc.*, 1997, **119**, 4309–4310.
114. C. M. Kao, M. McPherson, R. N. McDaniel, H. Fu, D. E. Cane and C. Khosla, *J. Am. Chem. Soc.*, 1997, **119**, 11339–11340.
115. J. Cortés, K. E. Wiesmann, G. A. Roberts, M. J. Brown, J. Staunton and P. F. Leadlay, *Science*, 1995, **268**, 1487–1489.
116. C. M. Kao, G. Luo, L. Katz, D. E. Cane and C. Khosla, *J. Am. Chem. Soc.*, 1994, **116**, 11612–11613.
117. R. Pieper, R. S. Gokhale, G. Luo, D. E. Cane and C. Khosla, *Biochemistry*, 1997, **36**, 1846–1851.
118. K. J. Weissman, M. Timoney, M. Bycroft, P. Grice, U. Hanefeld, J. Staunton and P. F. Leadlay, *Biochemistry*, 1997, **36**, 13849–13855.
119. K. J. Weissman, M. Bycroft, J. Staunton and P. F. Leadlay, *Biochemistry*, 1998, **37**, 11012–11017.
120. K. J. Weissman, M. Bycroft, A. L. Cutter, U. Hanefeld, E. J. Frost, M. C. Timoney, R. Harris, S. Handa, M. Roddis, J. Staunton and P. F. Leadlay, *Chem. Biol.*, 1998, **5**, 743–754.
121. I. Böhm, I. E. Holzbaier, U. Hanefeld, J. Cortés, J. Staunton and P. F. Leadlay, *Chem. Biol.*, 1998, **5**, 407–412.
122. C. Kao, G. Luo, L. Katz, D. E. Cane and C. Khosla, *J. Am. Chem. Soc.*, 1995 **117**, 9105–9106.
123. C. J. Martin, M. C. Timoney, R. M. Sheridan, S. G. Kendrew, B. Wilkinson, J. Staunton and P. F. Leadlay, *Org. Biomol. Chem.*, 2003, **1**, 4144–4147.
124. 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.
125. S. Yuzawa, C. H. Eng, L. Katz and J. D. Keasling, *Biochemistry*, 2013, **52**, 3791–3793.
126. A. F. Marsden, B. Wilkinson, J. Cortés, N. J. Dunster, J. Staunton and P. F. Leadlay, *Science*, 1998, **279**, 199–202.
127. L. S. Sheehan, R. E. Lill, B. Wilkinson, R. M. Sheridan, W. A. Vousden, A. L. Kaja, G. D. Crouse, J. Gifford, P. R. Graupner, L. Karr, P. Lewer, T. C. Sparks, P. F. Leadlay, C. Waldron and C. J. Martin, *J. Nat. Prod.*, 2006, **69**, 1702–1710.
128. S. Kuhstoss, M. Huber, J. R. Turner, J. W. Paschal and R. N. Rao, *Gene*, 1996, **183**, 231–236.

129. C. J. Dutton, S. P. Gibson, A. C. Goudie, K. S. Holdom, M. S. Pacey, J. C. Ruddock, J. D. Bu'Lock, and M. K. Richards, *J. Antibiot.*, 1991, **44**, 357–365.
130. M. S. Pacey, J. P. Dirlam, R. W. Geldart, P. F. Leadlay, H. A. McArthur, E. L. McCormick, R. A. Monday, T. N. O'Connell, J. Staunton and T. J. Winchester, *J. Antibiot.*, 1998, **51**, 1029–1034.
131. A. C. Goudie, N. A. Evans, K. A. Gration, B. F. Bishop, S. P. Gibson, K. S. Holdom, B. Kaye, S. R. Wicks, D. Lewis and A. J. Weatherley, *Vet. Parasitol.*, 1993, **49**, 5–15.
132. J.-B. Wang, H.-X. Pan and G.-L. Tang, *Bioorg. Med. Chem. Lett.*, 2011, **21**, 3320–3323.
133. N. Palaniappan, B. S. Kim, Y. Sekiyama, H. Osada and K. A. Reynolds, *J. Biol. Chem.*, 2003, **278**, 35552–35557.
134. R. S. Gokhale, S. Y. Tsuji, D. E. Cane and C. Khosla, *Science*, 1999, **284**, 482–485.
135. R. W. Broadhurst, D. Nietlispach, M. P. Wheatcroft, P. F. Leadlay and K. J. Weissman, *Chem. Biol.*, 2003, **10**, 723–731.
136. 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 (1999).
137. R. McDaniel, C. M. Kao, S. J. Hwang and C. Khosla, *Chem. Biol.*, 1997, **4**, 667–674.
138. L. Tang, H. Fu and R. McDaniel, *Chem. Biol.*, 2000, **7**, 77–84.
139. C. J. Rowe, I. U. Böhm, I. P. Thomas, B. Wilkinson, B. A. Rudd, G. Foster, A. P. Blackaby, P. J. Sidebottom, Y. Roddis, A. D. Buss, J. Staunton and P. F. Leadlay, *Chem. Biol.*, 2001, **8**, 475–485.
140. I. Thomas, C. J. Martin, C. J. Wilkinson, J. Staunton and P. F. Leadlay, *Chem. Biol.*, 2002, **9**, 781–787.
141. Y. Sugimoto, K. Ishida, N. Traitcheva, B. Busch, H. M. Dahse and C. Hertweck, *Chem. Biol.*, 2015, **22**, 229–240.
142. Y. Sugimoto, L. Ding, K. Ishida and C. Hertweck, *Angew. Chem. Int. Ed Engl.*, 2014, **53**, 1560–1564.
143. J. He and C. Hertweck, *Chem. Biol.*, 2003, **10**, 1225–1232.
144. 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.
145. C. Olano, B. Wilkinson, S. J. Moss, A. F. Braña, C. Méndez, P. F. Leadlay and J. A. Salas, *Chem. Commun.* 2003, 2780–2782.
146. P. Meiser, K. J. Weissman, H. B. Bode, D. Krug, J. S. Dickschat, A. Sandmann and R. Müller, *Chem. Biol.*, 2008, **15**, 771–781.
147. N. Gaitatzis, B. Silakowski, B. Kunze, G. Norsiek, H. Blöcker, G. Höfle and R. Müller, *J. Biol. Chem.*, 2002, **277**, 13082–13090.
148. 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.
149. S. Müller, S. Rachid, T. Hoffmann, F. Surup, C. Volz, N. Zaburannyi and R. Müller, *Chem. Biol.*, 2014, **21**, 855–865.
150. N. Traitcheva, H. Jenke-Kodama, J. He, E. Dittmann and C. Hertweck, *ChemBioChem*, 2007, **8**, 1841–1849.
151. S. Y. Tsuji, D. E. Cane and C. Khosla, *Biochemistry*, 2001, **40**, 2326–2331.
152. N. Wu, S. Y. Tsuji, D. E. Cane and C. Khosla, *J. Am. Chem. Soc.*, 2001, **123**, 6465–6474.
153. K. J. Weissman, *ChemBioChem*, 2006, **7**, 485–494.
154. 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.
155. N. Wu, D. E. Cane and C. Khosla, *Biochemistry*, 2002, **41**, 5056–5066.
156. M. Thattai, Y. Burak and B. I. Shraiman, *PLoS Comput. Biol.*, 2007, **3**, 1827–1835.
157. C. D. Richter, D. Nietlispach, R. W. Broadhurst and K. J. Weissman, *Nat. Chem. Biol.*, 2008, **4**, 75–81.
158. T. J. Buchholz, T. W. Geders, F. E. Barley 3<sup>rd</sup>, K. A. Reynolds, J. L. Smith and D. H. Sherman, *ACS Chem. Biol.*, 2009, **4**, 41–52.
159. N. Wu, D. E. Cane and C. Khosla, *Biochemistry*, 2002, **41**, 5056–5066.

160. S. Dutta, J. R. Whicher, D. A. Hansen, W. A. Hale, J. A. Chemler, G. R. Congdon, A. R. Narayan, K. Håkansson, D. H. Sherman, J. L. Smith and G. Skiniotis, *Nature*, 2014, **510**, 512–517.
161. B. S. Kim, T. A. Cropp, G. Florova, Y. Lindsay, D. H. Sherman and K. A. Reynolds, *Biochemistry*, 2002, **41**, 10827–10833.
162. J. Huang, A. L. Chen, H. Zhang, Z. Yu, M. H. Li, N. Li, J. T. Lin, H. Bai, J. D. Wang and Y. G. Zheng, *Appl. Environ. Microbiol.*, 2015, **81**, 5326–5334.
163. C. D. Reeves, S. L. Ward, W. P. Reville, H. Suzuki, M. Marcus, O. V. Petrakovsky, S. Marquez, H. Fu, S. D. Dong and L. Katz, *Chem. Biol.*, 2004, **11**, 1465–1472.
164. K. Watanabe, C. C. C. Wang, C. N. Boddy, D. E. Cane and C. Khosla, *J. Biol. Chem.*, 2003, **278**, 42020–42026.
165. S. E. O'Connor, C. T. Walsh and F. Liu, *Angew. Chem. Int. Ed Engl.*, 2003, **42**, 3917–3921.
166. J. Yan, S. Gupta, D. H. Sherman and K. A. Reynolds, *ChemBioChem*, 2009, **10**, 1537–1543.
167. 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.
168. D. H. Sherman, *Nat. Biotechnol.*, 2005, **23**, 1083–1084.
169. 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.
170. S. Murli, J. Kennedy, L. C. Dayem, J. R. Carney and J. T. Kealey, *J. Ind. Microbiol. Biotechnol.*, 2003, **30**, 500–509.
171. S. S. Chandran, H. G. Menzella, J. R. Carney and D. V. Santi, *Chem. Biol.*, 2006, **13**, 469–474.
172. H. G. Menzella, J. R. Carney and D. V. Santi, *Chem. Biol.*, 2007, **14**, 143–151.
173. B. J. Dunn, D. E. Cane and C. Khosla, *Biochemistry*, 2013, **52**, 1839–1841.
174. J. R. Whicher, S. Dutta, D. A. Hansen, W. A. Hale, J. A. Chemler, A. M. Dosey, A. R. Narayan, K. Håkansson, D. H. Sherman, J. L. Smith and G. Skiniotis, *Nature*, 2014, **510**, 560–564.
175. M. Hans, A. Hornung, A. Dziarnowski, D. E. Cane and C. Khosla, *J. Am. Chem. Soc.*, 2003, **125**, 5366–5374.
176. M. H. Medema, P. Cimermancic, A. Sali, E. Takano and M. A. Fischbach, *PLoS Comput. Biol.*, 2014, **10**, e1004016.
177. R. Ueoka, A. R. Uria, S. Reiter, T. Mori, P. Karbaum, E. E. Peters, E. J. Helfrich, B. I. Morinaka, M. Gugger, H. Takeyama, S. Matsunaga and J. Piel, *Nat. Chem. Biol.*, 2015, **11**, 705–712.
178. 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.
179. T. Bretschneider, J. B. Heim, D. Heine, R. Winkler, B. Busch, B. Kusebach, T. Stehle, G. Zocher and C. Hertweck, *Nature*, 2013, **502**, 124–128.
180. P. Pöplau, S. Frank, B. I. Morinaka and J. Piel, *Angew. Chem. Int. Ed Engl.*, 2013, **52**, 13215–13218.
181. J. J. Gehret, L. Gu, W. H. Gerwick, P. Wipf, D. H. Sherman and J. L. Smith, *J. Biol. Chem.*, 2011, **286**, 14445–14454.
182. M. C. Wilson and B. S. Moore, *Nat. Prod. Rep.*, 2012, **29**, 72–86.
183. 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.
184. M. K. Go, J. Y. Chow, V. W. N. Cheung, Y. P. Lim and W. S. Yew, *Biochemistry*, 2012, **51**, 4568–4579.
185. A. J. Hughes and A. Keatinge-Clay, *Chem. Biol.*, 2011, **18**, 165–176.
186. 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.
187. 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.
188. J. Mareš, J. Hájek, P. Urajová, J. Kopecký and P. A. Hrouzek, *PLoS ONE*, 2014, **9**, e111904.
189. B. S. Kim, D. H. Sherman and K. A. Reynolds, *Protein Eng. Des. Sel.*, 2004, **17**, 277–284.

190. J. A. Chemler, A. Tripathi, D. A. Hansen, M. O'Neil-Johnson, R. B. Williams, C. Starks, S. R. Park and D. H. Sherman, *J. Am. Chem. Soc.* (2015). doi:10.1021/jacs.5b04842
191. S. Poust, A. Hagen, L. Katz and J. D. Keasling, *Curr. Opin. Biotechnol.*, 2014, **30C**, 32–39.
192. Y. Yan, J. Chen, L. Zhang, Q. Zheng, Y. Han, H. Zhang, D. Zhang, T. Awakawa, I. Abe and W. Liu, *Angew. Chem. Int. Ed.*, 2013, **52**, 12308–12312.
193. N. Pulsawat, S. Kitani and T. Nihira, *Gene*, 2007, **393**, 31–42.
194. J. Staunton and K. J. Weissman, *Nat. Prod. Rep.*, 2001, **18**, 380–416.

## Figure legends

**Fig. 1** Structures of several polyketide blockbuster medicines. The erythromycin derivatives Biaxin and Zithromax act as antibiotics, and the native polyketides Rapamune and Prograf, as immunosuppressants.

**Fig. 2** The erythromycin (DEBS) polyketide synthase, the prototypical *cis*-AT modular PKS.<sup>7,8</sup> The PKS is organized into three gigantic subunits, DEBS 1, DEBS 2 and DEBS 3, each of which comprises several modules. The modules which sequentially initiate the biosynthesis (load), extend the polyketide chain by one building block and reductively process its functionality (modules 1–6), and terminate the assembly process (end), are composed of multiple independently-folding functional domains separated by linker regions (not shown). Following liberation of the first enzyme-free intermediate, 6-deoxyerythronolide B, it is further processed by a set of post-PKS enzymes to yield the biologically active erythromycin A. Key to domains: AT, acyl transferase; ACP, acyl carrier protein; KS, ketosynthase; KR, ketoreductase; DH, dehydratase; ER, enoyl reductase; TE, thioesterase.

**Fig. 3** The *trans*-AT PKS responsible for synthesis of the antibiotic virginiamycin M (Vir). Like members of its class, the Vir PKS exhibits a number of features which distinguish it from classic *cis*-AT PKSs: numerous *trans*-acting enzymes (notably the AT VirI, and a cassette of enzymes responsible for installation of a  $\beta$ -methyl functionality), tandemly-repeated domains (i.e. the ACPs of module 1), inactive domains (KS<sup>o</sup> of module 9), and the inclusion of modules which incorporate amino acids into the growing chain (modules 3 and 8). Retrobiosynthetic analysis indicates that a module must operate to join proline to the chain, but its encoding gene was not present in the sequenced region.<sup>193</sup> Furthermore, enzymes associated with starter-unit selection are as yet unidentified, while the respective roles of the oxidases VirM and VirN in the biosynthesis remain to be demonstrated directly. Key to domains: MT, C-methyltransferase; ACP, acyl carrier protein; C, condensation; A, adenylation; PCP, peptidyl carrier protein; HC, heterocyclization.

**Fig. 4** Theoretical possibilities for modifying PKS structures. The types of changes made are shaded, and color coded according to the domains or stand-alone (post-PKS) enzymes responsible. All of the indicated alterations could in principle be combined.

**Fig. 5** Workflow for a typical heterologous expression experiment. Reproduced from ref. 68 with permission from the Royal Society of Chemistry.

**Fig. 6** Set of derivatives of premonensin generated by site-directed inactivation of reductive domains, as demonstrated by HRMS. The domain targeted in each case is indicated (by type of domain (e.g. KR) and module of origin (e.g. 2)), and the introduced modification boxed. Adapted from ref. 82.

**Fig. 7** Use of modeling-guided site-directed mutagenesis to expand the substrate specificity of an AT domain.<sup>91</sup> Mutation of Val295 to Ala in AT<sub>6</sub> of DEBS 3 resulted in a multienzyme capable of utilizing 2-propargylmalonate as extender unit, yielding 2-propargylerythromycin which incorporated the same suite of post-PKS modifications as the native erythromycin A.

**Fig. 8** Efforts to understand and control the direction of reduction by ER domains. (a) NADPH-dependent reduction by ER domains can yield either the (2*S*)- or (2*R*)-methyl centers. Specific active site residues, Y and V respectively, are associated with these outcomes. (b) Multiple sequence alignment of ER domains showing the positions of the conserved residues correlating with stereochemistry (black triangles). The location of the NADPH-binding motif is indicated with grey bars. Reproduced from 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 with permission from Elsevier. (c) Schematic of the mini-PKS, DEBS 1-TE, used to test the role of the identified residues in stereocontrol. Initially, the native KR domain of module 2 was replaced with a ‘reductive loop’ from module 4 of the same PKS consisting of DH, ER and KR domains. The parental ER sequence in which the conserved Y is present gave rise to the predicted 2*S* stereochemistry. Mutation of the Y to the alternative V resulted in a quantitative switch in configuration at the 2-methyl center. (d) The reverse experiment in which the native V present in the ER domain from a rapamycin reductive loop was substituted with Y, failed to yield the expected (2*S*)-methyl.

**Fig. 9** The ‘reductive loop swapping’ (RLS) system.<sup>102,104</sup> (a) Schematic of the DEBS 1-TE model PKS and its triketide lactone product. The native KR in module 1 is of the B2 type, and that present in module 2 is A1. (b) Replacement of the module 2 KR with a 122 bp polylinker allowed introduction in its place of a panel of different KR domains (types A1/2 and B1/B2). The presence of multiple restriction sites permitted testing of the effect of the specific fusion sites on both yield and stereochemical outcome. The structures of the four lactone products generated in these experiments are shown, with the three corresponding to changes in stereochemistry boxed.

**Fig. 10** Generation of a tetraketide capable of cyclizing to form an 8-membered ring.<sup>114</sup> (a) A tetraketide synthase was initially created by relocating the TE domain from the end of DEBS 3 to the

terminus of module 3. Preferential use of the internal C5-hydroxyl as nucleophile resulted in a 6-membered ring lactone. (b) The TE was forced to use the terminal hydroxyl to form an 8-membered ring by removal of the C5-hydroxyl (pink box). For this, the KR domain of module 2 was replaced with a DH-ER-KR segment from the RAPS PKS – a net ‘gain of function’ experiment. An unidentified KR is presumed to have carried out ketoreduction at C3.

**Fig. 11** Composition of a 6-deoxyerythronolide B analogue library generated by making multiple changes to the DEBS PKS, including AT and KR exchanges within modules 2, 5 and 6, gain-of-function experiments (both substitution of DH-KR and DH-ER-KR for KR domains), and deletion mutagenesis (replacement of KR domains with a synthetic 18 aa fragment). The six DEBS modules are color-coded to indicate where the modifications were introduced. Evidence for these structures (including stereochemistry) was in the majority of cases only provided by mass spectrometry.

**Fig. 12** Generality of the TE relocation approach. Grafting of the DEBS TE onto module 4 of the spinosyn (Spn) PKS resulted in production of the expected pentaketide lactone product.<sup>123</sup>

**Fig. 13** Genetic engineering of PKS loading.<sup>126</sup> (a) The loading module of the DEBS PKS is specific for propionyl-CoA. (b) Replacement of the DEBS loading module with that from the avermectin (Ave) system, resulted in altered specificity of chain initiation: erythromycin analogues were generated incorporating the isobutyrate and 2-methyl butyrate starter units characteristic of avermectin. As shown, the resulting metabolites underwent substantial post-PKS modification.

**Fig. 14** Genetic engineering of a doramectin producer.<sup>131</sup> (a) Structures of the metabolites phoslactomycin and avermectin combined to generate doramectin, a highly successful antiparasitic used in veterinary medicine. (b) The doramectin PKS was created by replacing the avermectin loading module (in grey) with that from phoslactomycin (blue) which is specific for cyclohexane carboxylic acid.

**Fig. 15** Whole module swapping as an engineering approach.<sup>134</sup> Communication between module 1 of DEBS and two modules from the same system (3 and 6) was achieved by preserving the native intermodular linker between modules 1 and 2. The same strategy resulted in successful cross-talk between module 1 and module 5 from the rifamycin (Rif) PKS. By grafting the docking domain normally present at the end of DEBS 1 onto Rif module 5, it was able to engage in chain transfer with module 3, resulting in production of 6-deoxyerythronolide B.

**Fig. 16** Attempted addition of a module to the DEBS system. (a) RAPS module 2 was inserted at an internal position within DEBS 1, between modules 1 and 2. Although the predicted tetraketide was observed in minor amounts, the major product was triketide lactone, arising from skipping of the interpolated module.<sup>139</sup> (b) A separate study<sup>140</sup> elucidated the molecular basis for the by-passing event, showing that the intermediate undergoes direct ACP-to-ACP transfer involving the inserted module.

**Fig. 17** Genetic engineering of the neo-aureothin PKS to produce a homo-aureothin synthase. (a) Architecture of the neo-aureothin PKS. All modules of the PKS act once except for module 1 (NorA) which can act twice or three times, producing the two products shown (precursors to neo-aureothin and homo-aureothin). (b) An attempt was made to convert the nor PKS into an aureothin (Aur) synthase by removing modules 2 and 3, keeping intact the native ACP<sub>1</sub>/KS<sub>2</sub> junction. However, the resulting truncated PKS was not functional. (c) A more conservative construct in which only module 2 had been removed (again preserving the ACP<sub>1</sub>/KS<sub>2</sub> interface) was found to be active, producing polyketides corresponding to two and three iterations by NorA (precursors to homo-aureothin and neo-aureothin, respectively).

**Fig. 18** Three classes of docking domains from *cis*-AT PKSs. (a) NMR structure of a complex of covalently-fused Class 1a docking domains (PDB: 1PZQ, 1PZR).<sup>135</sup> The C-terminal DDs (red) comprise three  $\alpha$ -helices, the first two of which form a four  $\alpha$ -helical bundle which acts as a dimerization element. The third helix wraps around the coiled-coil formed by the partner N-terminal docking domain (blue). The long, flexible linker joining the second and third  $\alpha$ -helices of the C-terminal DD is represented as a dashed line. C and N indicate C- and N-termini, respectively. (b) X-ray crystal structure of a covalent complex of Class 1b docking domains (PDB: 3F5H).<sup>158</sup> Although by sequence analysis, this class also incorporates a C-terminal dimerization motif, it was not included in the investigated construct. The overall structure resembles that in a, but the C-terminal DD comprises two smaller  $\alpha$ -helices and the precise hydrophobic and polar interactions at the interface differ. (c) X-ray crystal structure of a covalent complex of class 2 docking domains (PDB: 4MYE).<sup>154</sup> The C-terminal DD does not incorporate a dimerization region. Instead, its two  $\alpha$ -helices associate with two  $\alpha$ -helices provided by the N-terminal docking domain to form an overall 8  $\alpha$ -helical bundle. (d) NMR structure of a class 3 N-terminal docking domain (PDB: 2JUG).<sup>157</sup> Each protomer within the DD homodimer comprises three  $\alpha$ -helices and two  $\beta$ -strands, which adopt an overall  $\alpha\beta\beta\alpha\alpha$  topology. For all docking domain complexes, charge:charge interactions are critical determinants of specificity.

**Fig. 19** Whole subunit swaps to alter macrolide structure.<sup>138</sup> In initial experiments, the first two subunits from the pikromycin (Pik) PKS (red) were combined with intact multienzymes from the DEBS (yellow) and oleandomycin (blue) systems, yielding the predicted hybrid 14-membered macrolactones. Subsequently, variants of DEBS 3 were used which contained domains sourced from the RAPS PKS, as well as an inactive KR domain. Again, the expected products were observed. Reproduced from ref. 194 with permission from the Royal Society of Chemistry.

**Fig. 20** Generation of tenvermectins by exchanging subunit AveA1 for MilA1.<sup>162</sup> The structures of the parent avermectin and milbemycin are shown (with colors to highlight their differences), as well as those of the hybrid tenvermectins created by the subunit swap.

**Fig. 21** Use of docking domain engineering to enable communication between unrelated proteins. (a) The epothilone PKS-NRPS was used as a model system. Biosynthesis is initiated by recruitment of acetate by EpoA, which is then condensed with Ser tethered to EpoB, followed by oxidative cyclization to form methylthiazol-5-EpoB. This is then extended with malonate by EpoC. (b) Productive communication with EpoB and EpoC and two unrelated ACP domains was achieved by grafting the native docking partners at the ACP C-termini.

**Fig. 22** Dissection of subunit PikAI by introduction of docking domains. (a) Structure of native PikAI which comprises a loading domain and two chain extension modules. (b) PikAI was divided into two smaller proteins by introducing matched pairs of docking domains from the phoslactomycin PKS in place of the intermodular linkers present between the loading module and module 1 and modules 1 and 2. Yields of pikromycin were only reduced by 50% by these modifications.

**Fig. 23** The 'legoization' of polyketide biosynthesis. (a) Building blocks for the expression plasmids, comprising a loading module (LM), intermodular linkers (LI), N-terminal docking domain (LN), extender module, C-terminal docking domain (LC) and thioesterase. (b) Organization of the two compatible plasmids (vector pAng, with a CloDF13 replication origin and a streptomycin-resistance selection marker, and pBru containing a ColE1 replication origin and a carbenicillin selection marker). (c) Mini-PKS resulting in *E. coli* from co-expression of pAng and pBru, which gives rise to product triketide lactone. Reprinted with permission from MacMillan Publishers Ltd: 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.

**Fig. 24** Generation of hybrid PKS by recombination in yeast. Recombinants were obtained between the genes encoding the mono-modular PikAIII and the bimodular DEBS 3, by exploiting short identical sequences. One of the resulting hybrid PKSs was used to complement a PikAIII/PikAIV knockout mutant of *S. venezuelae*, generating the novel pikromycin derivative dihydropikromycin (the change relative to pikromycin is indicated in grey).

## TOC graphic/text

This reviews covers on-going efforts at engineering the gigantic modular polyketide synthases (PKSs), highlighting both notable successes and failures

