COMBINATORIAL BIOSYNTHESIS OF REDUCED POLYKETIDES

Kira J. Weissman and Peter F. Leadlay

Abstract | The bacterial multienzyme polyketide synthases (PKSs) produce a diverse array of products that have been developed into medicines, including antibiotics and anticancer agents. The modular genetic architecture of these PKSs suggests that it might be possible to engineer the enzymes to produce novel drug candidates, a strategy known as 'combinatorial biosynthesis'. So far, directed engineering of modular PKSs has resulted in the production of more than 200 new polyketides, but key challenges remain before the potential of combinatorial biosynthesis can be fully realized.

SCAFFOLD
The backbone atoms of a molecule, on which functionality is displayed.

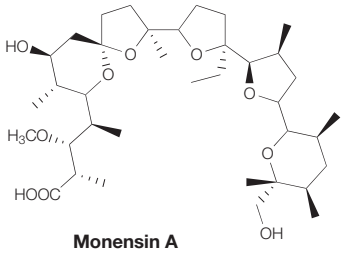
Department of Biochemistry, University of Cambridge, Cambridge, CB2 1GA, UK.
e-mails: kjw21@cus.cam.ac.uk; leadlay@mole.bio.cam.ac.uk
doi:10.1038/nrmicro1287

Polyketide natural products are among the most important microbial metabolites in human medicine, targeting both acute and degenerative diseases. They are in clinical use as antibiotics (erythromycin A, rifamycin S), anticancer drugs (doxorubicin, epothilone), cholesterol-lowering agents (lovastatin), antiparasitics (avermectin), antifungals (amphotericin B), insecticides (spinosyn A) and immunosuppressants (rapamycin) (FIG. 1). Polyketide-derived pharmaceuticals comprise 20% of the top-selling drugs, with combined worldwide revenues of over UK£10 billion per year.

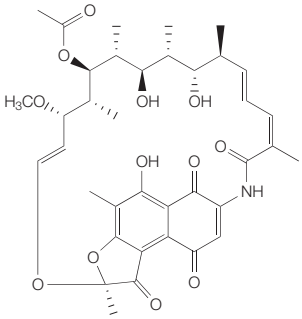
There is intense interest in discovering analogues of polyketide natural products with improved or novel pharmacological properties — the need for new antibiotics is particularly well recognized[1]. Polyketide-drug discovery efforts have traditionally taken two forms: isolation and identification of new compounds from the biosphere, or the semi-synthetic modification of existing drug SCAFFOLDS using medicinal chemistry. Both approaches are likely to yield new structures in the foreseeable future. For the past few years, however, an alternative strategy has also been used to produce new polyketides — genetically engineering polyketide synthase (PKS) biosynthetic pathways. This advance was made possible by the initial cloning of several polyketide gene clusters in the 1980s, including the clusters for synthesis of the aromatic antibiotic actinorhodin[2], and the reduced polyketide antibiotic erythromycin A[3,4]. By mixing and matching some of the actinorhodin PKS genes with their counterparts from other aromatic-antibiotic gene clusters, hybrid molecules were often produced[5,6], giving rise to the concept now known as 'combinatorial biosynthesis'[7]. Given the modular organization within the giant PKS proteins that catalyse erythromycin biosynthesis, simple mixing and matching of these intact proteins would be of limited interest. However, once methods were devised to 'cut and paste' portions of these giant genes to produce chimaeric genes, in a way that retained the activity of the multiple enzyme activities that they encode, it became apparent that the structure of erythromycin (and therefore other related polyketides) might be rationally modified by creating patchwork PKS genes from two or more natural PKS genes spliced together[8–10], an idea which now dominates combinatorial biosynthesis[11–13].

Although modular PKS genetic engineering has now repeatedly been demonstrated[14–16], the low efficiency of these experiments limits their routine use in industrial drug discovery, and this type of combinatorial engineering remains too labour intensive to be described as high-throughput. Therefore, research has also focused on developing improved tools for genetic engineering and heterologous expression, as well as understanding the structure and function of modular PKS proteins and post-PKS tailoring enzymes. These efforts have recently resulted in a biological source of

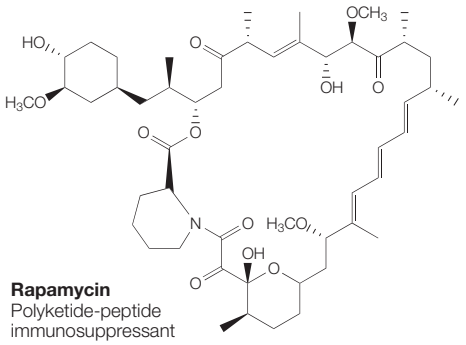
a Reduced polyketides



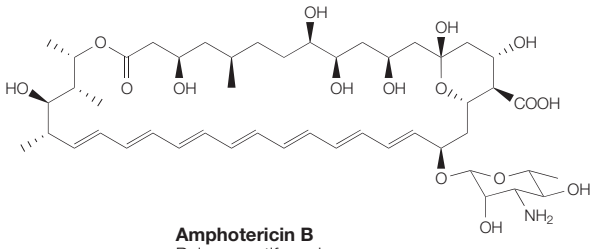
Monensin A
Polyether antibiotic



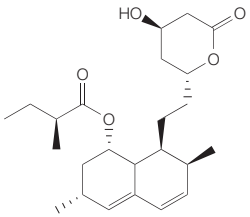
Rifamycin S
Ansamycin antibiotic



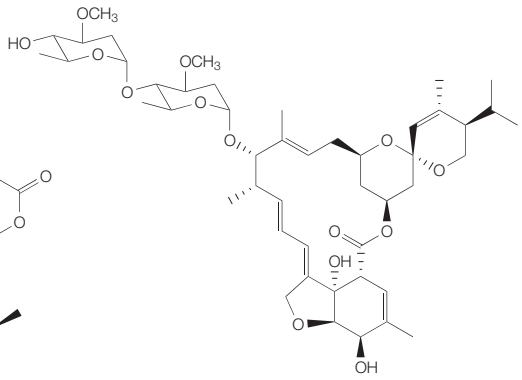
Rapamycin
Polyketide-peptide immunosuppressant



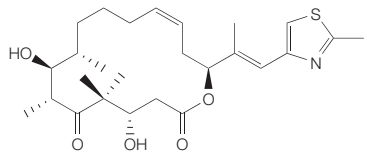
Amphotericin B
Polyene antifungal



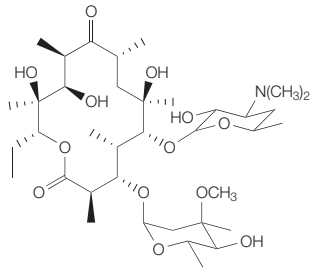
Lovastatin
Cholesterol-lowering agent



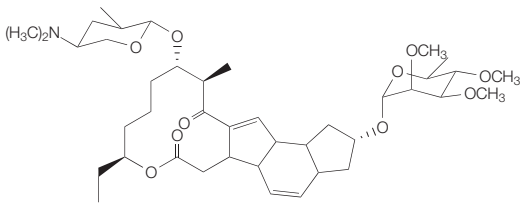
Avermectin B1b
Macrolide antiparasitic



Epothilone
Polyketide-polypeptide anticancer agent

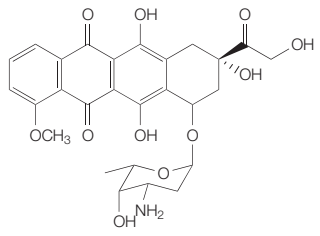


Erythromycin A
Macrolide antibiotic

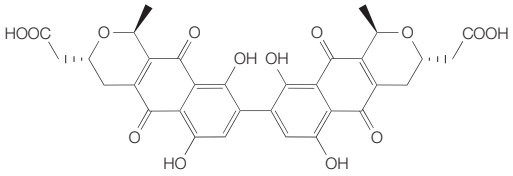


Spinosyn A
Macrolide insecticide

b Aromatic polyketides

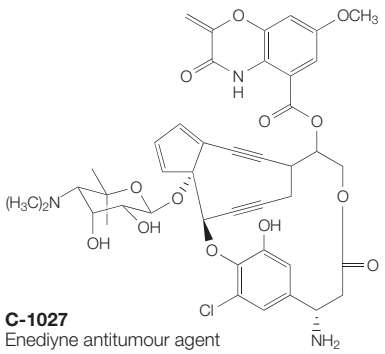


Doxorubicin
Aromatic antitumour agent



Actinorhodin
Aromatic antibiotic

c Unclassified



C-1027
Enediyne antitumour agent

Figure 1 | **Classes of polyketide metabolites and representative structures.** This conventional classification distinguishes those metabolites (**a**) in which polyketide-chain synthesis occurs with concomitant reduction from (largely) aromatic metabolites, such as in (**b**). However, many polyketides (**c**) resist easy classification.

the semi-synthetic drug ivermectin[14], as well as several promising chemotherapeutic-drug leads derived from geldanamycin[15] and epothilone[16].

Polyketides: structure and function

Polyketides make good drugs. Of the 7,000 known polyketide structures, more than 20 (0.3%) have been commercialized, which is considerably better than the typical <0.001% 'hit rate' from standard pharmaceutical screens. Despite their relatively low molecular weights (on average 800 Da), polyketides bristle with diverse functional groups, and many are rich in stereochemistry.

Polyketides have been classified both on the basis of their structures and on their mode of biosynthesis. For simplicity, we distinguish between polyketides that are derived from unreduced polyketone chains and are largely aromatic (for example, doxorubicin (FIG. 1b)) and those in which many of the carbonyl functionalities have been reduced during biosynthesis (for example, erythromycin A (FIG. 1a)). In the case of the reduced polyketides, the initial linear chain is often constrained by MACROCYCLIZATION into a biologically active conformation, which promotes precise interaction with protein targets[17]. In this review, we will focus only on the reduced polyketides produced by modular PKSs, as their biosynthesis has to date proved most amenable to rational genetic engineering.

Reduced polyketides can be divided into two main classes, the macrocycles and the polyethers. Macrocycles, such as erythromycin A and rapamycin (FIG. 1a), contain a large lactone or lactam ring[18] and have a wide variety of biological activities. Antibiotic and antifungal polyenes[19], such as amphotericin B (FIG. 1a) and nystatin, form a subset of the macrolides that incorporate three to seven conjugated double bonds. The polyene macrolides are among the giants of the polyketide world, with rings that contain more than 44 carbons. A second subset of the macrolides is the ansamycin antibiotics, which includes rifamycin S (FIG. 1a) and geldanamycin. These molecules incorporate the non-proteinogenic amino acid 3-amino-5-hydroxybenzoic acid (AHBA) as the first building block and, in common with rapamycin and epothilone, are hybrids of polyketides and peptides[20].

The polyether ionophore antibiotics, such as monensin A[21] (FIG. 1a) and salinomycin, have a carboxylate group and two to five ether oxygen atoms (usually as tetrahydrofuran or tetrahydropyran rings) that chelate alkali-metal cations. Other polyketides, for example the potent antitumour enediyne molecules (C-1027, FIG. 1c), contain both reduced and aromatic groups[22,23], and cannot easily be classified.

Polyketide producers

Reduced polyketides are mainly synthesized in bacteria but are also found in terrestrial and marine invertebrates, including insects, molluscs and sponges, although there is evidence that the true sources of these metabolites in these organisms are symbiotic bacteria[24]. The actinomycetes, a diverse order of Gram-positive bacteria, synthesize most of all known polyketides. Among them, the soil-dwelling, filamentous *Streptomyces* and *Saccharopolyspora* species are the most prolific polyketide producers, although myxobacteria[25] and pseudomonads[26] are also important sources of these compounds.

The functions of polyketides, which are non-essential for the growth of producer organisms at least *in vitro*, are unknown[27]. It is often assumed that they provide a competitive advantage in the environment through uses in intra- and interspecies communication or in self-defence. Self-defence as an explanation is particularly persuasive for non-motile *Streptomyces* spp.[27] Others have argued that some secondary metabolites have no value at present to their producer organisms, but that the ability to produce these molecules provides a selective advantage[28].

PKSs are modular

For the past 15 years, the biosynthesis of the polyketide core of erythromycin A, 6-deoxyerythronolide B (6-dEB) (FIG. 2), has provided the paradigm for understanding the structure and function of the PKSs that are responsible for assembling complex polyketides. 6-dEB is constructed on a gigantic molecular 'assembly line', with a dedicated enzymatic domain for each biosynthetic step[29] (FIG. 2), a feature known as the 'collinearity rule.'

The boundaries and active sites of the PKS domains in the gene sequence were identified by their homologies to the corresponding enzymes in animal fatty-acid biosynthesis[30]. Sequence analysis revealed that the domains are covalently linked and grouped into functional units called 'modules', so that each module carries out a single chain-extension step and any required processing reactions before the intermediate is passed on to the downstream module (FIG. 2); therefore, these multienzyme complexes have been named the modular PKSs. The modules are further organized into multimodular subunits — in erythromycin biosynthesis these subunits are 6-dEB synthases (DEBS) 1, 2 and 3. A typical subunit contains two or three modules, although a multienzyme (MLSA1) in the PKS that synthesizes the polyketide toxin mycolactone incorporates nine modules[31]. MLSA1 contains over 50 active sites and weighs more than 2 MDa, making it one of the largest and most complex catalysts yet discovered in nature.

The organization of the modular PKSs allows a wide range of complex polyketide products to be assembled from simple precursors. Also, the structure (and important elements of the stereochemistry[32]) of a polyketide product can often be predicted by examining the complement of PKS genes. Furthermore, with only one or two exceptions[33], bacterial and fungal PKS and post-PKS genes are clustered together in the genome with transcriptional regulators and genes for self-resistance[34]. This genetic organization simplifies the identification and sequencing of complete novel PKS clusters.

MACROCYCLIZATION
Formation of a large macrolactone or macrolactam ring.

DIFFUSIVE LOADING
Binding of a substrate at an enzyme active site by diffusion through the cytoplasm.

Combinatorial biosynthesis

The 'one enzyme, one function' organization of modular PKS genes indicates an obvious strategy for manipulating the pathways. If this modularity extends to the proteins — that is, if the component enzymes are portable and can carry out their natural functions in alternative contexts and with new substrates — then it should be possible to rearrange the domains to create novel PKSs that produce novel products. Encouragingly, many domains receive their substrates from PKS enzymes that catalyse prior steps in the biosynthesis and not by DIFFUSIVE LOADING, which suggests that they

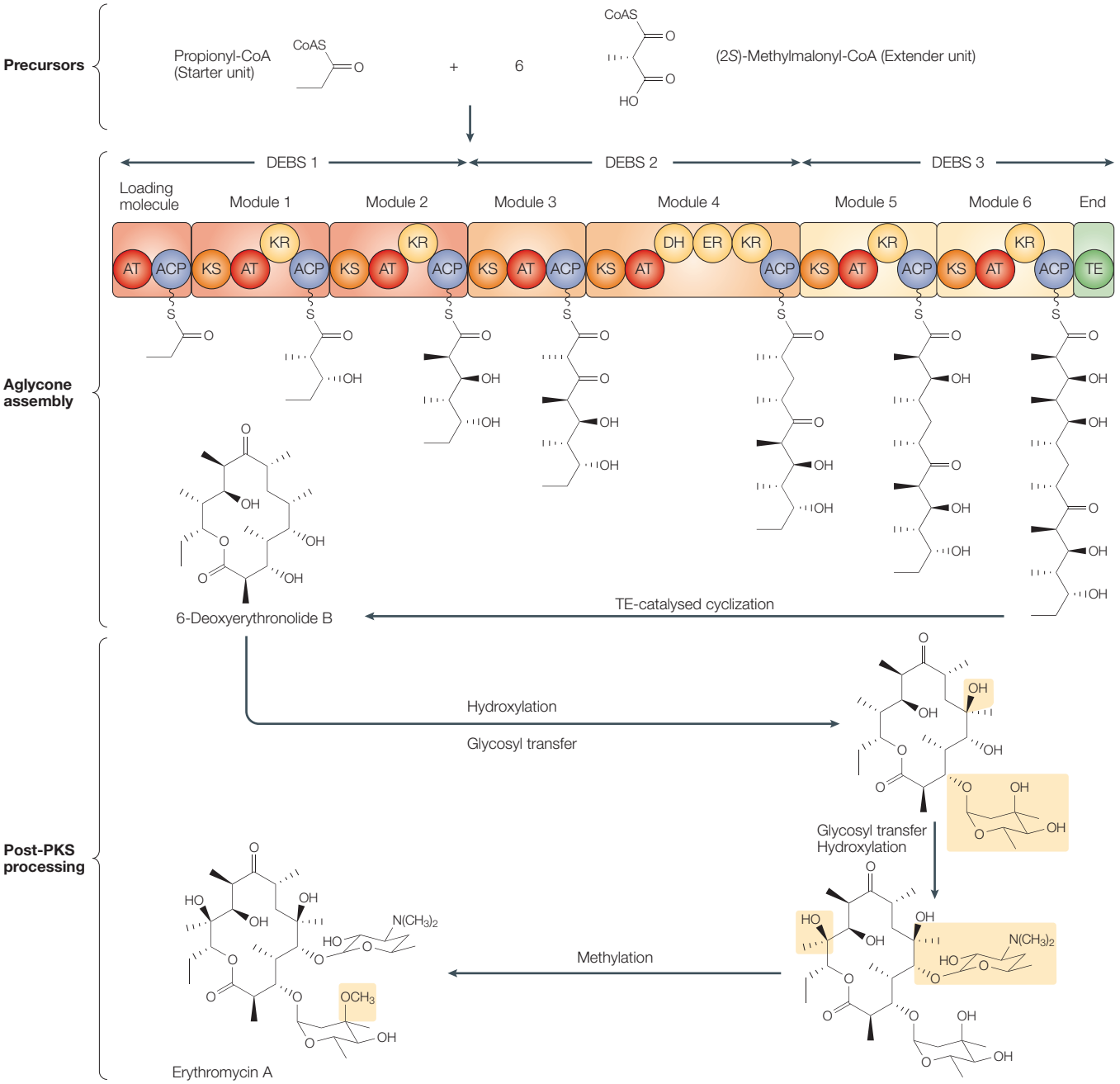


Figure 2 | **The modular polyketide synthase (PKS) paradigm.** In this figure, erythromycin biosynthesis is used to illustrate modular PKS function. The precursors of polyketides are small carboxylic acids, such as acetate, propionate and malonate, and are activated as their coenzyme A (CoA) thioesters. One starter unit of propionyl-CoA and six extender units of (2*S*)-methylmalonyl-CoA (carboxylated propionyl-CoA) are required for the biosynthesis of erythromycin A. The erythromycin PKS (6-deoxyerythronolide B synthase, DEBS) consists of a loading module, six chain-extension modules and a chain-terminating thioesterase (TE), distributed across three gigantic proteins which are named DEBS 1, 2 and 3. Each chain-elongation module contains an acyl transferase (AT), a ketosynthase (KS) and an acyl carrier protein (ACP) domain, and variable numbers of reductive elements, as required. Following TE-catalysed release from the PKS, the aglycone 6-deoxyerythronolide B is hydroxylated twice, glycosylated twice and methylated on a sugar residue to form the fully active antibiotic erythromycin A. DH, dehydratase; ER, enoyl reductase; KR, ketoreductase.

might not have been under evolutionary pressure to develop strict substrate specificities, and therefore that they might be inherently amenable to rearrangement. The combinatorial strategy is also helped by our increasing ability to distinguish domains and their boundaries. It might be feasible to have a truly 'combinatorial biosynthesis' of polyketides, in which all the possible combinations of domains, modules or subunits are constructed, resulting in vast libraries of new compounds[12], and because of the PKS collinearity rule, such changes could in principle generate products with predictable structures.

Although domain modularity is important, successful combinatorial biosynthesis also depends on several other factors, including suitable host cells for library production (such as plant and insect cells[35]), phosphopantetheinyl transferase (PPTase) enzymes to activate PKS subunits by post-translational modification[36], a complete set of small-molecule precursors[37], post-PKS enzymes with the required flexibility to modify unnatural substrates[38], and so on (for a more extended discussion, see REF. 39).

Progress and challenges in PKS engineering

By directed manipulation of domains, it has been possible to decrease[40] and increase chain lengths[41], to alter the selection of starter units[42] and building blocks[43,44], and to modify the level of β-keto reduction[14] (FIG. 3). Swapping whole modules or subunits of PKSs in wild-type and mutated forms has yielded products that are hybrids of several polyketides[45]. The largest reported number of new structures from a single experiment is more than 50 derivatives of 6-dEB, including analogues that are altered at one, two or three carbon centres[46]. In total, more than 200 new polyketides have been produced by PKS genetic engineering[47], which represents a substantial fraction of the number of natural polyketides identified to date.

Despite these encouraging results, attempts to engineer PKSs have shown that the 'modular' synthases are not completely modular; that is, individual domains function with a preferred set of partner activities[48], and all domains have some degree of substrate specificity[49]. Rational control of stereochemistry is another challenging problem, as initial mechanistic

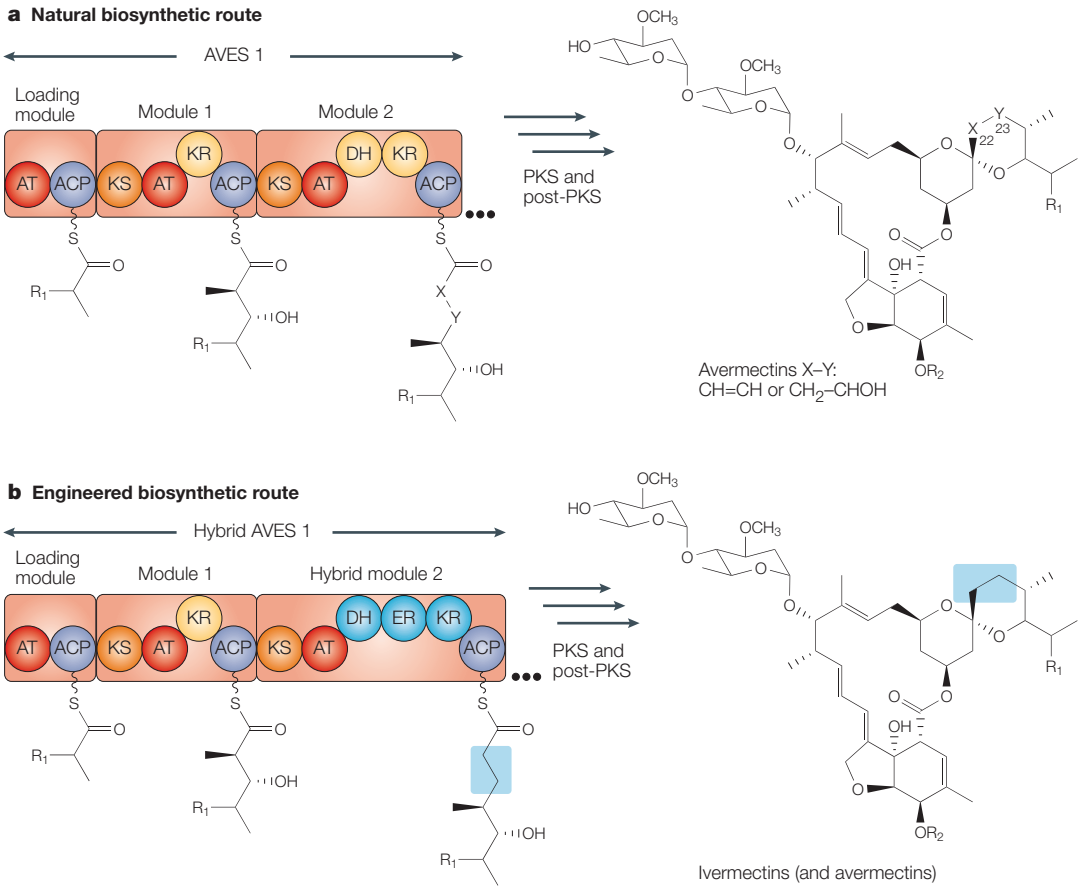


Figure 3 | **Direct production of ivermectin-like drugs by genetic engineering of the avermectin polyketide synthase (PKS). a** | The natural biosynthetic pathway results in a mixture of avermectins incorporating either a double bond at C22–C23 or a hydroxyl at C23. **b** | By genetic engineering, the dehydratase (DH) and ketoreductase (KR) domains of the avermectin PKS were replaced with a complete set of reductive elements, DH, enoyl reductase (ER) and KR, from the rapamycin PKS. The resulting hybrid synthase also produced avermectins because of incomplete processing of some chains, but yielded additional structures that are fully reduced at C22 and C23. This result indicates that it might be possible to obtain ivermectins directly by fermentation instead of by semi-synthesis. AVES, avermectin PKS.

studies have revealed that the stereochemical outcome at both alkyl and hydroxyl centres results from a complex interplay between the ketosynthase (KS) and ketoreductase (KR) domains and is, therefore, difficult to predict[50,51].

Furthermore, we do not understand the spatial relationships among domains within the modules that allow them to cooperate in a sequential fashion. These relationships must also be maintained in any engineered enzyme. These interactions might be mediated, in part, by the amino-acid LINKERS between the conserved domains[52]. In experiments in which domains have been exchanged, fusion points are often selected within these linker regions based on sequence homologies (so that the hybrid sequence is most similar to the parents) and/or the ease of different cloning strategies, without considering possible structural or functional effects. Consequently, many of these experiments have resulted in low yields of modified polyketides, or have failed. As a result, it is often necessary to try to use different fusion sites within a given linker to increase the chances of success[44].

The context-dependent behaviour of domains lowers the potential for truly combinatorial re-ordering of PKS domains, although large libraries are still an important goal. Therefore, in the short term, attention is being focused on three areas[53]: directed engineering of analogues of valuable compounds to improve their activity or pharmacodynamics, installing functionalities into existing scaffolds that are more amenable to specific modification by standard semi-synthesis methodologies, and generating advanced intermediates for use in the chemical synthesis of natural products. Even to achieve these more modest goals, we will need to develop increasingly sophisticated genetic tools for the efficient manipulation and expression of PKS gene clusters, while substantially augmenting our knowledge of both PKS architecture and mechanisms.

New PKS clusters

Sequencing in whole or in part of approximately 50 PKS clusters has revealed important variations on the modular PKS 'paradigm' of DEBS, both in terms of the order and composition of domains within PKS modules and some aspects of their operation[54,55]. These variations suggest new strategies for genetic engineering, enhancing the prospects for truly combinatorial biosynthesis. New clusters, particularly those that synthesize metabolites with unusual structural features, also provide additional domains for inclusion in a combinatorial 'tool kit'. As 99% of bacterial strains cannot be cultured using standard fermentation conditions[56], the polyketide 'diversome', particularly of marine microorganisms[57], remains largely unexplored.

Finding new clusters. Microbial genome sequencing, as already shown for *Streptomyces coelicolor*[58] and *Streptomyces avermitilis*[59], allows us to identify all of the predicted PKSs present in an organism, as well as (in principle) the associated regulatory networks. Another approach to finding new clusters is to screen cosmid or bacterial artificial chromosome (BAC) libraries of DNA from a particular bacterium for modular PKS genes; such screens can isolate the genes that code for production of specific metabolites, or cryptic clusters that code for production of as-yet-unidentified polyketide products.

There has also been progress in obtaining DNA from unculturable bacteria and obligate microbial symbionts. DNA extracted from soil samples has been cloned into bacterial hosts such as *Streptomyces lividans* without prior cultivation of the bacteria[60,61]. These METAGENOMIC libraries can represent up to 5% of the microbial diversity that is present in the soil bacterial community, a 24-fold improvement on traditional culturing methods[61]. Metagenomic libraries from aquatic and terrestrial invertebrates such as sponges and beetles — which also contain the DNA of their bacterial symbionts[62] — have been constructed, and screening of these libraries could uncover new PKS genes.

PKS genes in DNA libraries are usually detected using screening by PCR with primers designed for the most conserved regions of KS domains[63]. Locating a single PKS gene can enable identification of a whole cluster if the insert size of the library is large enough[64]. Finding a cluster that synthesizes a specific metabolite is more difficult, because many organisms produce more than one polyketide. For example, *S. avermitilis* contains eight modular PKS clusters[59]. Unfortunately, PCR-based methods are biased and can only identify a proportion of the KS domains that are present, with the consequence that the metabolic potential within a given sample is likely to be underestimated[65].

Nonetheless, screening approaches have shown that modular PKS genes are widely distributed among all the major actinomycete families and genera[66], as well as in marine cyanobacteria[67], dinoflagellates[68], bacterial symbionts of marine invertebrates[62] and many other species. Such diversity will undoubtedly inspire many new approaches to manipulating PKS gene clusters.

Improvements in engineering technologies

Engineering PKS systems is laborious, owing to the size of polyketide gene clusters (10–100 kb), and also, several versions of the same experiment are often required. In general, two strategies have been used to reconfigure PKS gene clusters. In the first strategy, homologous recombination in the natural host has been used to delete, alter or replace segments of genomic DNA. The advantage of this approach is that the host bacterium can produce the original metabolite, but one important drawback is that homologous recombination is often difficult to achieve, particularly with slow-growing organisms such as the *Streptomyces* spp. Furthermore, although the heterologous cloning of PKSs in actinomycetes has been established[69], the genetic tools to manipulate many other classes of polyketide-producing microorganisms are poorly developed[39]. The alternative strategy is to transfer (portions of) the PKS cluster to heterologous hosts that are more tractable to genetic engineering. Where the chosen expression hosts are

LINKERS
Sequences of amino acids that adopt no fixed structure and that covalently join certain enzymatic domains within PKS modules.

METAGENOMIC
Genetic material obtained from an environmental sample.

non-actinomycetes, it is often necessary to modify the hosts so that they can provide an adequate supply of small-molecule precursors as well as a PPTase enzyme for efficient post-translational activation of the PKS proteins[36].

Advances in heterologous expression. Using heterologous hosts for both cloning and polyketide overproduction is often the method of choice. Strains that have already been optimized by classical improvement strategies, such as *Saccharopolyspora erythraea* and *Streptomyces fradiae*, are obviously useful, and the genetically well characterized strains *S. coelicolor*, *S. lividans* and *Streptomyces venezualae* have also been exploited. These species can coordinately regulate all the genes that are required for polyketide assembly, can express and fold the gigantic PKS proteins, and can synthesize a wide range of precursors. By constructing deletion mutants of these strains to reduce native PKS production, 'cleaner' backgrounds for heterologous expression have been produced[6,70]. PKS gene sets from less tractable bacteria, and engineered hybrid systems, can be assembled in *Escherichia coli* and then transferred by conjugation into these optimized strains[71]. The development of *E. coli–Streptomyces* artificial chromosome (ESAC) vectors allows PKS cluster-sized fragments (>90 kDa) to be transferred in this manner[72]. Once introduced into the *Streptomyces* strains, the genes are then integrated into the chromosomes by homologous recombination, or at naturally occurring or engineered attachment sites.

Progress has also been made in identifying activator/promoter systems that function well in these hosts, for example the *actII-ORF4/PactI* activator–promoter pair from the erythromycin cluster, although no single system functions in every species[71,73,74]. Recent research has also revealed a superfamily of PPTases that catalyse the essential post-translational modification of PKS proteins[75]. This work has resulted in a collection of PPTases that have different substrate preferences (including the extraordinarily substrate-tolerant enzyme Sfp from *Bacillus subtilis*[76]), which ensures that a range of PKS proteins can be expressed in activated form. Sequencing of new PKS clusters has also revealed gene sets for precursor metabolites required in building-block assembly. Mobilizing these genes onto plasmids allows for precursor production in heterologous hosts, for example methoxymalonyl acyl carrier protein (ACP) extender units in *S. fradiae*[77]. Several complex polyketides have been produced heterologously, including 6-dEB[70], narbonolide and 10-deoxymethynolide (precursors to pikromycin and methymycin, respectively)[78], 8,8a-deoxyoleandolide (precursor to oleandomycin)[79], megalomicin[80], epothilone[81] and aureothin[82].

One useful host for PKS expression is *E. coli*, although extensive engineering has been required for this to be feasible. *E. coli* produces acetyl-coenzyme A (acetyl-CoA) and malonyl-CoA, but lacks sufficient quantities of other precursors needed for polyketide assembly, such as the starter unit propionyl-CoA and the extender unit (2*S*)-methylmalonyl-CoA[83]. Furthermore, the *E. coli* PPTase enzyme cannot interact with modular PKS proteins[84]. By making a series of metabolic modifications, researchers at Kosan Biosciences (Hayward, California) have engineered *E. coli* to produce the polyketide core of erythromycin A, 6-dEB[83], and an analogue of 6-dEB incorporating butyrate as a starter unit[85]. The modified strain of *E. coli* has also been used to synthesize yersiniabactin, a hybrid natural product incorporating both polyketide and peptide elements[86], and the AHBA precursor to the ansamycin antibiotics such as rifamycin[87]. In an important advance, researchers have recently succeeded in introducing two post-PKS glycosylation operons into *E. coli*, and showed their function by converting inactive 6-dEB to a biologically active glycosylated derivative, erythromycin C[88]. If this result can be generalized to other pathways, *E. coli* could be a useful host for the production of novel, therapeutically relevant polyketides.

Engineering strategies. Many attempts to engineer PKSs have taken the domain as the combinatorial unit. In these experiments, individual activities are deleted, introduced or exchanged using sites within the linker regions that separate the domains, but many of these experiments fail. One way to circumvent this problem is to carry out less invasive domain 'surgery' by site-directed mutagenesis. Our limited knowledge of the molecular basis of specificity and mechanisms means that this approach has so far only been reported for the acyl transferase (AT) domains[43,89]. A potentially powerful alternative technique is to disable a function by specific mutagenesis at the active site, and then to provide in *trans* a replacement function from a different domain. In one example, the AT domain from the sixth module of DEBS was inactivated, and substituted with the discrete malonyl-CoA:ACP transacylase (MAT) from *S. coelicolor*, which selects acetate as a building block instead of propionate[90]; the expected des-methyl, 6-dEB product was synthesized. In the future, it should be possible to alter the extender-unit specificity of the MAT domain to incorporate a broader range of building blocks at selected positions.

Another tactic is to swap complete PKS modules and entire subunits; the domain interactions within these regions have already been optimized, so this type of strategy is less likely to perturb the structure and function of the PKS. The most successful strategy for joining two non-cognate modules together is to use an existing intermodular linker sequence and, in particular, the linker associated with the upstream module[48,91]. To carry out engineering on the subunit level, whole multienzymes have been expressed from several compatible vectors with different resistance markers[46,92]. Mutant versions of each subunit were then combined through co-transformation of different plasmid combinations, rather than by making new plasmids for each of the several mutants. Therefore, this strategy marks an important step towards truly combinatorial polyketide engineering[47]. However, this

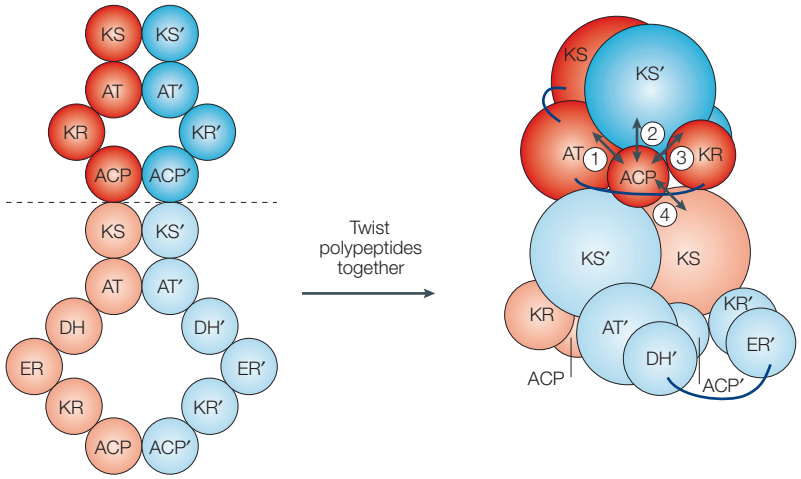


Figure 4 | **The 'double helical' model for modular polyketide synthases (PKSs).** Each PKS subunit is a homodimer, and the individual proteins are twisted around each other to form a double helical architecture. The ketosynthase (KS), acyl transferase (AT) and acyl carrier protein (ACP) domains are located at the core of the helix, whereas the reductive activities occupy the periphery of the structure. This arrangement places the ACP at the spatial centre of the module, where it can interact with all of the other domains (as shown), in particular the KS domain on the opposite subunit. In this representation, the linkers in between the domains are shown as lines. Interdomain interactions during chain extension and processing are shown in the right hand panel. (1) ACP and AT interact during loading of the extender unit. (2) ACP and KS′ cooperate during chain extension. (3) ACP and KR cooperate during ketoreduction and (4) ACP_n and KS_{n+1} interact during intermodular chain transfer. DH, dehydratase; ER, enoyl reductase; KR, ketoreductase.

approach only works when the constituent subunits are derived from homologous sets of PKS genes, so that they retain an inherent ability to interact and function with each other[45].

Advances in molecular biology have also impacted on PKS engineering. For example, engineered homologous recombination[93] has proved useful for manipulating large sections of PKS genes in *E. coli*, such as regions encoding AT domains[15]; if required, the resulting plasmids can then be transferred by conjugation into an alternative host for polyketide production[94,95]. This technique has recently been coupled with the directed evolution of PKS domains, which has the considerable advantage that many experiments (for example, choice of linker splice sites) can be carried out simultaneously. 4,000 variants of the LOADING DIDOMAIN AT–ACP of the pikromycin PKS were created by DNA family shuffling, and the resulting mutant synthases were screened for biological activity in *S. lividans*; only three altered PKSs led to production of bioactive compounds[96]. Advances in the total synthesis of long DNA sequences could enable assembly of entire hybrid PKSs, or even gene clusters, from scratch, introducing convenient restriction sites to aid in genetic engineering, and tailoring codon usage to the particular expression organism[97].

Emerging structural information

Determining the architecture of the modular PKSs is crucial to the success of engineering, but the size and complexity of PKSs make their structures difficult to analyse. Any structural models must account not only for the ability of all of the domains within each module to interact with the ACP in a coordinated fashion, but also for communication between successive modules (ACP_n–KS_{n+1}) during chain transfer (FIG. 4). Also, as each PKS contains several protein subunits (up to 8, REF. 21), these must be arranged correctly to synthesize only the desired end product.

A structural model for modular PKSs was proposed in 1996, on the basis of biophysical studies of purified DEBS proteins[98]. These experiments revealed that each subunit is a homodimer, and that individual domains (particularly the KS–AT didomains and thioesterase (TE) activities) remain dimeric even when released from the PKS as discrete proteins; by contrast, the reductive activities were always monomeric. These studies also showed that a particular KS preferentially interacts with the ACP domain on the opposite subunit during chain extension. These findings were incorporated into a 'double-helical' model for modular PKSs[98] (FIG. 4). In this structure, individual proteins in each dimeric subunit are wrapped around each other. This topology places the KS (and possibly AT) domains at the centre of the helix, consistent with their intimate homodimeric behaviour; by contrast, the reductive activities occupy the periphery of the structure. Nonetheless, all of the domains in each module seem to be able to interact with an ACP domain (FIG. 4), and the closest contact between the KS and ACP domains within the same module occurs across the subunit interface, as required by the crosslinking studies.

Despite efforts to determine the structures of a module or subunit, the only structure solved to date is of a single domain, the TE/cyclase from the erythromycin and pikromycin systems[99,100]. The crystallographic structures revealed that the TEs are homodimers, as expected from the earlier proteolysis work[98], and that they belong to the α/β hydrolase family, but with unusually open substrate channels that span the proteins. The shape of the channels agrees with the known substrate preferences of each enzyme.

Other relevant structures are those of the DOCKING DOMAINS from the erythromycin PKS. This designation refers to the sequences of amino acids at the C and N termini of PKS subunits, which are thought to participate in the correct ordering of PKS multienzymes[101]. By fusing together two docking domains, it was possible to solve the structure of a docked complex by NMR[102]. This structure, which is representative of a large group of such docking elements, revealed the shared basis for the association of subunits, as well as an unexpected role in their dimerization. It also indicated how the DEBS proteins discriminate against inappropriate partners.

It might soon be possible to assemble an overall structure for a PKS module from the X-ray and NMR structures of its constituent domains, as has been accomplished for other large multienzyme complexes, such as the 2-oxoacid dehydrogenases[103]. These efforts will almost certainly be aided by sophisticated homology modelling based on solved structures[104]. Recent progress in the expression of individual domains and didomains[105] is encouraging.

LOADING DIDOMAIN
Consists of domains, minimally an acyl-transferase–acyl-carrier protein didomain, which initiate biosynthesis by selection of a starter unit.

DOCKING DOMAINS
Sequences of amino acids at the end of polyketide-synthase subunits which adopt distinct three-dimensional structures and are putatively involved in mediating protein–protein recognition between the multienzymes.

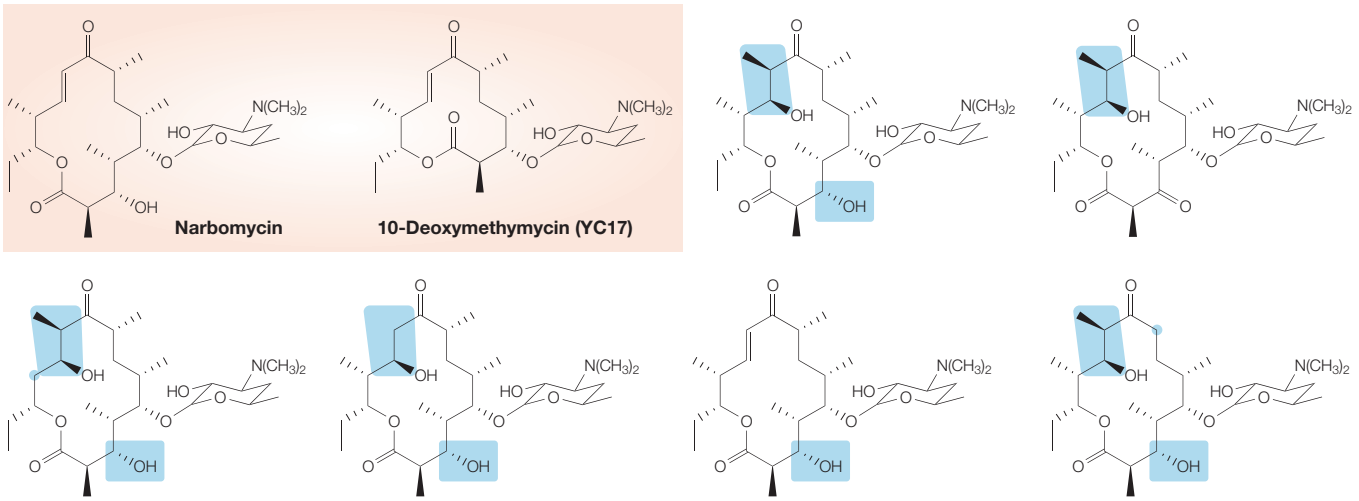


Figure 5 | **Post-polyketide-synthase elaboration of a polyketide library.** Structures of putative bioactive compounds produced by co-expressing the gene for the desosamine-specific glycosyl transferase DesVII and a plasmid library encoding different macrolactones in *Streptomyces lividans*. In each case, liquid chromatography–mass spectrometry analysis yielded a mass consistent with the anticipated structure of the desosaminylated polyketide. The pink box indicates the natural products of DesVII — 14-membered narbomycin and the 12-membered 10-deoxymethymycin (YC17) — whereas the blue boxes indicate where the alternative macrolides differ in structure from narbomycin.

Domain modularity

Combinatorial engineering depends on domain portability, and the substrate preferences of several activities have been studied to identify domains of each type that are suitable for combinatorial use. The ATs have strict substrate and stereochemical specificities, typically accepting a single extender unit[106]. Some KS[107,108] and TE[109] domains are significantly more tolerant to alternative structures, although they are not entirely promiscuous. Therefore, until it is possible to manipulate substrate specificities by, for example, using solved crystallographic structures to identify specific active-site residues involved in substrate selection, these preferences will limit the potential of combinatorial methods. Encouragingly, two groups have recently shown that sequence motifs shared among AT domains[110] are located at the active sites and have a direct role in substrate choice[43,89]; by altering these specific residues, the specificity of an AT was relaxed from methylmalonate only to both malonate and methylmalonate.

The KR domains also tolerate different substrates[111], and these activities have been successfully swapped among different PKSs[112]. The stereochemical outcome of the natural KR-catalysed reactions can be reliably predicted by aligning conserved amino-acid motifs within the domains[32,113], but the direction of reduction in engineered PKSs can be affected in unpredictable ways by both the substrates and their context within the hybrid synthases[50,51]. So whereas KR domains show the substrate flexibility required for combinatorial applications, we need to improve our ability to predict or to control the stereochemistry at the resulting hydroxyl centres.

The mycolactone PKS[31] might provide the most useful set of domains for combinatorial applications. Sequencing has revealed that the amino-acid identity among many comparable domains (for example, KSs and ACPs) in all 16 modules of this PKS is very high. Remarkably, these virtually identical activities operate on the polyketide chain throughout the chain-extension process, meaning that they can tolerate large variations in substrate size and functionality. This observation indicates that the mycolactone domains might be the least discriminating PKS enzymes identified so far; if this hypothesis is proved correct, these domains might prove invaluable in future PKS engineering.

Post-PKS enzymes

The post-PKS decoration of polyketide structures, for example by oxidation, methylation, alkylation or glycosylation, dramatically alters their molecular binding properties and is crucial to their biological activity[38]. A current challenge, therefore, is to develop combinatorial approaches that lead to novel polyketide skeletons and also to complete structures that can serve as drug candidates.

The function of post-PKS enzymes is typically tested using gene knockouts, followed by characterization of the resulting metabolites; in this way, it is also often possible to draw conclusions about the sequence in which modification events occur[38,114]. An alternative approach is to express a particular post-PKS gene in a heterologous host in the presence of a polyketide scaffold[79]. Once the role of a particular enzyme has been determined, it can then be evaluated for its use in combinatorial biosynthesis. Glycosyltransferases (GTs), for example, have been investigated *in vivo* in various ways, for both their tolerance to alternative sugars and to novel aglycones. These experiments are usually carried out in heterologous hosts, using combinations of native and exogenous GTs, sugars or sugar gene sets and PKS pathways[115–117]. In one example, the pathway

to the sugar desosamine and the associated GT gene *desVII* from the pikromycin pathway were co-expressed with a library of plasmids encoding different macrolactones[118] (FIG. 5); the host in this case was a mutant strain of *S. lividans* in which the polyketide production had been disabled[119]. Although conversion of the polyketides was generally low, bioassays revealed that the novel macrolides had antibiotic activity. Analogous experiments have been used to probe the specificity of polyketide P450 mono-oxygenases[45,79,117,120,121] and methyltransferases[122].

These types of experiments have revealed that, whereas some post-PKS enzymes have specific substrate preferences, others are flexible enough to allow the efficient production of novel compounds. For example, in its native role, the P450 enzyme PikC from the pikromycin pathway catalyses the hydroxylation of both 12- and 14-membered macrolactones and recognizes two alternative positions within the smaller ring[123,124]; it has also been shown to act on hybrid polyketides that incorporate portions of pikromycin, tylosin and erythromycin[125]. The desosamine GT from the same pathway, DesVII, shows similarly broad tolerance for hybrid macrolides[118,125]. Other promising enzymes include the GTs from the oleandomycin[116,126] and tylosin[115] pathways, the spinosyn methyltransferases[122] and the post-PKS enzymes of ansamitocin biosynthesis[127]. However, as existing post-PKS activities are not likely to modify all possible new polyketide backbones, in the future, it might be necessary to relax or tailor these specificities further using site-directed mutagenesis or directed evolution approaches.

Perspectives

So far, genetic engineering of PKSs has resulted in more than 200 new compounds[47], some of which show promise as drug candidates. From an engineering perspective, it now remains to develop even more efficient and rapid methods for manipulating gene clusters in various species, and for facile transfer of engineered gene sets to heterologous hosts which have been optimized for all aspects of polyketide production. No single 'superhost' strain is likely to be the answer for all cases. Better knowledge of the three-dimensional architecture of PKSs, the details of how domains operate and interact, and the molecular basis for substrate specificity and stereocontrol will pay dividends in the design of hybrid synthases that are likely to be functional. In particular, if amino-acid units could be reliably engineered into polyketides using modules from non-ribosomal peptide synthetases[128], the diversity of the molecular libraries produced would be greatly magnified. One way forward is hinted at by the successful chemical assembly of peptide–polyketide materials on solid supports and their ensuing combinatorial PKS enzyme-catalysed cyclization into soluble libraries[129]. Also, the ideas and methods of forced enzyme evolution[130,131], already highly successful at improving the bioprocessing fitness of conventional enzymes, might yet be brought fruitfully to bear on these assembly-line multienzymes. Routine and effective combinatorial engineering of polyketide biosynthesis as an integral part of modern drug discovery remains both an alluring and a tangible goal.

Note added in proof

In recent work, Menzella *et al.*[132] synthesized codon-optimized genes encoding 14 chain-extension modules derived from eight PKS clusters. The modules were designed to incorporate unique restriction sites not only at the module boundaries, but also flanking each domain within the modules, the intermodular linker sequences and the docking domains. Using these gene cassettes, the authors engineered 154 distinct bimodular combinations, fronted by the loading module from DEBS and terminating in the DEBS TE. On the basis of analysis by liquid chromatography/tandem mass spectrometry, nearly half of the chimaeric systems yielded the predicted triketide lactone products in an engineered strain of *E. coli*[83]. The authors' use of synthetic 'building blocks' is important as a logical way of speeding up the search for productive combinations of all types of PKS units — modules, domains and linkers. Although in this case each experiment was individually engineered, this research represents an important step towards combinatorial production of polyketides.

1. Walsh, C. Where will new antibiotics come from? *Nature Rev. Microbiol.* **1**, 65–70 (2003).
2. Malpartida, F. & Hopwood, D. A. Molecular cloning of the whole biosynthetic pathway of a *Streptomyces* antibiotic and its expression in a heterologous host. *Nature* **309**, 462–464 (1984).
3. Cortés, J., Haydock, S. F., Roberts, G. A., Bevitt, D. J. & Leadlay, P. F. An unusually large multifunctional polypeptide in the erythromycin-producing polyketide synthase of *Saccharopolyspora erythraea*. *Nature* **348**, 176–178 (1990).
4. Donadio, S., Staver, M. J., McAlpine, J. B., Swanson, S. J. & Katz, L. Modular organization of genes required for complex polyketide biosynthesis. *Science* **252**, 657–679 (1991).
References 3 and 4 report the first sequencing of a modular PKS gene.
5. Bartel, P. L. *et al.* Biosynthesis of anthraquinones by interspecies cloning of actinorhodin biosynthesis genes in streptomycetes: clarification of actinorhodin gene functions. *J. Bacteriol.* **172**, 4816–4826 (1990).
6. McDaniel, R., Ebert-Khosla, S., Hopwood, D. A. & Khosla, C. Engineered biosynthesis of novel polyketides. *Science* **262**, 1546–1550 (1993).
7. Khosla, C. & Zawada, R. J. Generation of polyketide libraries via combinatorial biosynthesis. *Trends Biotechnol.* **14**, 335–341 (1996).
8. Cortés, J. *et al.* Repositioning of a domain in a modular polyketide synthase to promote specific chain cleavage. *Science* **268**, 1487–1489 (1995).
9. Kao, C. M., Luo, G. L., Katz, L., Cane, D. E. & Khosla, C. Manipulation of macrolide ring size by directed mutagenesis of a modular polyketide synthase. *J. Am. Chem. Soc.* **117**, 9105–9106 (1995).
10. Oliynyk, M., Brown, M. J., Cortés, J., Staunton, J. & Leadlay, P. F. A hybrid modular polyketide synthase obtained by domain swapping. *Chem. Biol.* **3**, 833–839 (1996).
11. Donadio, S. & Sosio, M. Strategies for combinatorial biosynthesis with modular polyketide synthases. *Comb. Chem. High Throughput Screen.* **6**, 489–500 (2003).
12. Weber, T., Welzel, K., Pelzer, S., Vente, A. & Wohlleben, W. Exploiting the genetic potential of polyketide producing streptomycetes. *J. Biotechnol.* **106**, 221–232 (2003).
13. Weissman, K. J. Understanding and exploiting PKS modularity. *Phil. Trans. R. Soc. Lond. A* **362**, 2671–2690 (2004).
14. Gaisser, S. *et al.* Direct production of ivermectin-like drugs after domain exchange in the avermectin polyketide synthase of *Streptomyces avermitilis*. *Org. Biomol. Chem.* **1**, 2840–2847 (2003).
Describes the engineered biosynthesis of ivermectin, a valuable antiparasitic drug.
15. Patel, K. *et al.* Engineered biosynthesis of geldanamycin analogs for Hsp90 inhibition. *Chem. Biol.* **11**, 1625–1633 (2004).
16. Starks, C. M., Zhou, Y., Liu, F. & Licari, P. J. Isolation and characterization of new epothilone analogues from recombinant *Myxococcus xanthus* fermentations. *J. Nat. Prod.* **66**, 1313–1317 (2003).

17. Kohli, R. M. & Walsh, C. T. Enzymology of acyl chain macrocyclization in natural product biosynthesis. *Chem. Commun. (Camb.)* **3**, 297–307 (2003).
18. Ogasawara, Y. *et al.* Cloning, sequencing, and functional analysis of the biosynthetic gene cluster of macrolactam antibiotic vicenistatin in *Streptomyces halstedii*. *Chem. Biol.* **11**, 79–86 (2004).
19. Aparicio, J. F., Caffrey, P., Gil, J. A. & Zotchev, S. B. Polyene antibiotic biosynthesis gene clusters. *Appl. Microbiol. Biotechnol.* **61**, 179–188 (2003).
20. Kim, C. G., Yu, T. W., Fryhle, C. B., Handa, S. & Floss, G. 3-Amino-5-hydroxybenzoic acid synthase, the terminal enzyme in the formation of the precursor of mC7N units in rifamycin and related antibiotics. *J. Biol. Chem.* **273**, 6030–6040 (1998).
21. Oliynyk, M. *et al.* Analysis of the biosynthetic gene cluster for the polyether antibiotic monensin in *Streptomyces cinnamonensis* and evidence for the role of *monB* and *monC* genes in oxidative cyclization. *Mol. Microbiol.* **49**, 1179–1190 (2003).
22. Ahlert, J. *et al.* The calicheamicin gene cluster and its iterative type I enediyne PKS. *Science* **297**, 1173–1176 (2002).
23. Liu, W., Christenson, S. D., Standage, S. & Shen, B. Biosynthesis of the enediyne antitumor antibiotic C-1027. *Science* **297**, 1170–1173 (2002).
24. Piel, J. *et al.* Antitumor polyketide biosynthesis by an uncultivated bacterial symbiont of the marine sponge *Theonella swinhoei*. *Proc. Natl Acad. Sci. USA* **101**, 16222–16227 (2004).
25. Gerth, K., Pradella, S., Perlova, O., Beyer, S. & Müller, R. Myxobacteria: proficient producers of novel natural products with various biological activities — past and future biotechnological aspects with the focus on the genus *Sorangium*. *J. Biotechnol.* **106**, 233–253 (2003).
26. El-Sayed, A. K. *et al.* Characterization of the mupirocin biosynthesis gene cluster from *Pseudomonas fluorescens* NCIMB 10586. *Chem. Biol.* **10**, 419–430 (2003).
27. Challis, G. L. & Hopwood, D. A. Synergy and contingency as driving forces for the evolution of multiple secondary metabolite production by *Streptomyces* species. *Proc. Natl Acad. Sci. USA* **100**, 14555–14561 (2003).
28. Firn, R. D. & Jones, C. G. Natural products — a simple model to explain chemical diversity. *Nat. Prod. Rep.* **20**, 382–391 (2003).
29. Staunton, J. & Weissman, K. J. Polyketide biosynthesis: a millennium review. *Nat. Prod. Rep.* **18**, 380–416 (2001).
30. Smith, S., Witkowski, A. & Joshi, A. K. Structural and functional organization of the animal fatty acid synthase. *Prog. Lipid Res.* **42**, 289–317 (2003).
References 23 and 30 report the gene sequence of a PKS responsible for assembling an enediyne-type polyketide.
31. Stinear, T. P. *et al.* Giant plasmid-encoded polyketide synthases produce the macrolide toxin of *Mycobacterium ulcerans*. *Proc. Natl Acad. Sci. USA* **101**, 1345–1349 (2004).
Reports the sequencing and analysis of the mycolactone PKS.
32. Caffrey, P. Conserved amino acid residues correlating with ketoreductase stereospecificity in modular polyketide synthases. *Chembiochem* **4**, 654–657 (2003).
Identifies residues in KR domains that can be used to predict the direction of ketoreduction.
33. Yu, T. W. *et al.* The biosynthetic gene cluster of the maytansinoid antitumor agent ansamitocin from *Actinosynnema pretiosum*. *Proc. Natl Acad. Sci. USA* **99**, 7968–7963 (2002).
34. Stratigopoulos, G., Bate, N. & Cundliffe, E. Positive control of tylosin biosynthesis: pivotal role of TylR. *Mol. Microbiol.* **54**, 1326–1334 (2004).
35. Pfeifer, B. A. & Khosla, C. Biosynthesis of polyketides in heterologous hosts. *Microbiol. Mol. Biol. Rev.* **65**, 106–118 (2001).
36. Weissman, K. J., Hong, H., Oliynyk, M., Siskos, A. P. & Leadlay, P. F. Identification of a phosphopantetheinyl transferase for erythromycin biosynthesis in *Saccharopolyspora erythraea*. *Chembiochem* **5**, 116–125 (2004).
37. Pfeifer, B., Hu, Z., Licari, P. & Khosla, C. Process and metabolic strategies for improved production of *Escherichia coli*-derived 6-deoxyerythronolide B. *Appl. Environ. Microbiol.* **68**, 3287–3292 (2002).
38. Rix, U., Fischer, C., Remsing, L. L. & Rohr, J. Modification of post-PKS tailoring steps through combinatorial biosynthesis. *Nat. Prod. Rep.* **19**, 542–580 (2002).
39. Cane, D. E., Walsh, C. T. & Khosla, C. Harnessing the biosynthetic code: combinations, permutations and mutations. *Science* **282**, 63–68 (1998).
40. Martin, C. J. *et al.* Heterologous expression in *Saccharopolyspora erythraea* of a pentaketide synthase derived from the spinosyn polyketide synthase. *Org. Biomol. Chem.* **1**, 4144–4147 (2003).
41. Rowe, C. J. *et al.* Engineering a polyketide with a longer chain by insertion of an extra module into the erythromycin-producing polyketide synthase. *Chem. Biol.* **8**, 475–485 (2001).
42. Long, P. F. *et al.* Engineering specificity of starter unit selection by the erythromycin-producing polyketide synthase. *Mol. Microbiol.* **43**, 1215–1225 (2002).
43. Del Vecchio, F. *et al.* Active-site residue, domain and module swaps in modular polyketide synthases. *J. Ind. Microbiol. Biotechnol.* **30**, 489–494 (2003).
44. Petkovič, H. *et al.* A novel erythromycin, 6-desmethyl erythromycin D, made by substituting an acyltransferase domain of the erythromycin polyketide synthase. *J. Antibiot.* **56**, 543–551 (2003).
References 38 and 44 are comprehensive reviews on advances in the understanding and manipulation of post-PKS enzymes.
45. Reeves, C. D. *et al.* Production of hybrid 16-membered macrolides by expressing combinations of polyketide synthase genes in engineered *Streptomyces fradiae* hosts. *Chem. Biol.* **11**, 1465–1472 (2004).
46. Xue, Q., Ashley, G., Hutchinson, C. R. & Santi, D. V. A multiplasmid approach to preparing large libraries of polyketides. *Proc. Natl Acad. Sci. USA* **96**, 11740–11745 (1999).
Reports an important step towards combinatorial PKS engineering.
47. Rodriguez, E. & McDaniel, R. Combinatorial biosynthesis of antimicrobials and other natural products. *Curr. Opin. Chem. Biol.* **4**, 526–534 (2001).
48. Ranganathan, A. *et al.* Knowledge-based design of bimodular and trimodular polyketide synthases based on domain and module swaps: a route to simple statin analogues. *Chem. Biol.* **6**, 731–741 (1999).
49. Watanabe, K., Wang, C. C., Boddy, C. N., Cane, D. E. & Khosla, C. Understanding substrate specificity of polyketide synthase modules by generating hybrid multimodular synthases. *J. Biol. Chem.* **278**, 42020–42026 (2003).
50. Holzbaur, I. E. *et al.* Molecular basis of Celmer's rules: role of the ketosynthase domain in epimerisation and demonstration that ketoreductase domains can have altered product specificity with unnatural substrates. *Chem. Biol.* **8**, 329–340 (2001).
51. Weissman, K. J. *et al.* The molecular basis of Celmer's rules: the stereochemistry of the condensation step in chain extension on the erythromycin polyketide synthase. *Biochemistry* **36**, 13849–13855 (1997).
52. Tsuji, S. Y., Wu, N. & Khosla, C. Intermodular communication in polyketide synthases: comparing the role of protein–protein interactions to those in other multidomain proteins. *Biochemistry* **40**, 2317–2325 (2001).
53. McDaniel, R., Welch, M. & Hutchinson, C. R. Genetic approaches to polyketide antibiotics. *Chem. Rev.* **105**, 543–558 (2005).
54. Müller, R. Don't classify polyketide synthases. *Chem. Biol.* **11**, 4–6 (2004).
55. Shen, B. Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms. *Curr. Opin. Chem. Biol.* **7**, 285–295 (2003).
56. Amann, R. I., Ludwig, W. & Schleifer, K. H. Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. *Microbiol. Rev.* **59**, 143–169 (1995).
57. Salomon, C. E., Magarvey, N. A. & Sherman, D. H. Merging the potential of microbial genetics with biological and chemical diversity: an even brighter future for marine natural product drug discovery. *Nat. Prod. Rep.* **21**, 105–121 (2004).
58. Bentley, S. D. *et al.* Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature* **417**, 141–147 (2002).
59. Ōmura, S. *et al.* Genome sequence of an industrial microorganism *Streptomyces avermitilis*: deducing the ability of producing secondary metabolites. *Proc. Natl Acad. Sci. USA* **298**, 12215–12220 (2001).
60. Courtois, S. *et al.* Recombinant environmental libraries provide access to microbial diversity for drug discovery from natural products. *Applied Environ. Microbiol.* **69**, 49–55 (2003).
Describes a method for accessing microbial diversity from soil samples without cultivation.
61. Ginolhac, A. *et al.* Phylogenetic analysis of polyketide synthase I domains from soil metagenomic libraries allows selection of promising clones. *Appl. Environ. Microbiol.* **70**, 5522–5527 (2004).
62. Piel, J. A polyketide synthase-peptide synthetase gene cluster from an uncultured bacterial symbiont of *Paederus* beetles. *Proc. Natl Acad. Sci. USA* **92**, 14002–14007 (2002).
63. Rascher, A. *et al.* Cloning and characterisation of a gene for geldanamycin production in *Streptomyces hygroscopicus* NRRL 3602. *FEMS Microbiol. Lett.* **218**, 223–230 (2003).
64. Zazopoulous, E. *et al.* A genomics-guided approach for discovering and expressing cryptic metabolic pathways. *Nature Biotechnol.* **21**, 187–190 (2003).
65. Ayuso, A. *et al.* A novel actinomycete strain de-replication approach based on the diversity of polyketide synthase and nonribosomal peptide synthetase biosynthetic pathways. *Appl. Microbiol. Biotechnol.* **67**, 795–806 (2005).
66. Ayuso-Sacido, A. & Genilloud, O. New PCR primers for the screening of NRPS and PKS-I systems in actinomycetes: detection and distribution of these biosynthetic gene sequences in major taxonomic groups. *Microb. Ecol.* **49**, 10–24 (2004).
67. Edwards, D. J. *et al.* Structure and biosynthesis of the jamaicamides, new mixed polyketide-peptide neurotoxins from the marine cyanobacterium *Lyngbya majuscula*. *Chem. Biol.* **11**, 817–833 (2004).
68. Snyder, R. V. *et al.* Polyketide synthase genes from marine dinoflagellates. *Mar. Biotechnol.* **5**, 1–12 (2003).
69. Strohl, W. R., Bartel, P. L., Li, Y., Connors, N. C. & Woodman, R. H. Expression of polyketide biosynthesis and regulatory genes in heterologous streptomycetes. *J. Indust. Microbiol.* **7**, 163–174 (1991).
70. Kao, C. M., Katz, L. & Khosla, C. Engineered biosynthesis of a complete macrolactone in a heterologous host. *Science* **265**, 509–512 (1994).
Reports the first biosynthesis of a polyketide macrolactone in the heterologous host *Streptomyces coelicolor*.
71. Rodriguez, E. *et al.* Rapid engineering of polyketide overproduction by gene transfer to industrially optimised strains. *J. Ind. Microbiol. Biotechnol.* **30**, 480–488 (2003).
72. Sosio, M. *et al.* Artificial chromosomes for antibiotic-producing actinomycetes. *Nature Biotechnol.* **18**, 343–345 (2000).
73. Tao, M. F., Zhou, X. F., Kieser, T. & Deng, Z. X. Construction of a temperature inducible shuttle expression vector and its application in *Streptomyces*. *Sheng Wu Gong Cheng Xue Bao* **18**, 420–423 (2002).
74. Wilkinson, C. J. *et al.* Increasing the efficiency of heterologous promoters in actinomycetes. *J. Mol. Microbiol.* **4**, 417–426 (2002).
75. Lambalot, R. H. *et al.* A new enzyme superfamily — the phosphopantetheinyl transferases. *Chem. Biol.* **3**, 923–926 (1996).
Describes a large family of proteins responsible for post-translational transfer of 4′-phosphopantetheine.
76. Quadri, L. E. *et al.* Characterization of Sfp, a *Bacillus subtilis* phosphopantetheinyl transferase for peptidyl carrier protein domains in peptide synthetases. *Biochemistry* **37**, 1585–1595 (1998).
77. Rodriguez, E. *et al.* Engineered biosynthesis of 16-membered macrolides that require methoxymalonyl-ACP precursors in *Streptomyces fradiae*. *Appl. Microbiol. Biotechnol.* **66**, 85–91 (2004).
78. Tang, L., Fu, H., Betlach, M. & McDaniel, R. Elucidating the mechanism of chain termination switching in the picromycin/methymycin polyketide synthase. *Chem. Biol.* **6**, 553–558 (1999).
79. Shah, S. *et al.* Cloning, characterization, and heterologous expression of a polyketide synthase and P-450 oxidase involved in the biosynthesis of the antibiotic oleandomycin. *J. Antibiot.* **53**, 502–508 (2000).
80. Volchegursky, Y., Hu, Z., Katz, L. & McDaniel, R. Biosynthesis of the anti-parasitic agent megalomicin: transformation of erythromycin to megalomicin in *Saccharopolyspora erythraea*. *Mol. Microbiol.* **37**, 752–762 (2000).
81. Tang, L. *et al.* Cloning and heterologous expression of the epothilone gene cluster. *Science* **287**, 640–642 (2000).
Describes the production of epothilone in the heterologous host *E. coli*.
82. He, J. & Hertweck, C. Iteration as programmed event during polyketide assembly; molecular analysis of the aureothin biosynthesis gene cluster. *Chem. Biol.* **10**, 1225–1232 (2003).
83. Pfeifer, B. A., Admiraal, S. J., Gramajo, H., Cane, D. E. & Khosla, C. Biosynthesis of complex polyketides in a metabolically engineered strain of *E. coli*. *Science* **291**, 1790–1792 (2001).
Describes the production of 6-dEB in an adapted strain of *E. coli*.

84. Roberts, G. A., Staunton, J. & Leadlay, P. F. Heterologous expression in *Escherichia coli* of an intact multienzyme component of the erythromycin-producing polyketide synthase. *Eur. J. Biochem.* **214**, 305–311 (1993).
85. Kennedy, J., Murli, S. & Kealey, J. T. 6-Deoxyerythronolide B analogue production in *Escherichia coli* through metabolic pathway engineering. *Biochemistry* **42**, 14342–14348 (2003).
86. Pfeifer, B. A., Wang, C. C., Walsh, C. T. & Khosla, C. Biosynthesis of yersiniabactin, a complex polyketide-nonribosomal peptide, using *Escherichia coli* as a heterologous host. *Appl. Environ. Microbiol.* **69**, 6698–6702 (2003).
87. Watanabe, K., Rude, M. A., Walsh, C. T. & Khosla, C. Engineered biosynthesis of an ansamycin polyketide precursor in *E. coli*. *Proc. Natl Acad. Sci. USA* **100**, 9774–9778 (2003).
88. Peiru, S., Menzella, H. G., Rodriguez, E., Carney, J. & Gramajo, H. Production of the potent antibacterial polyketide erythromycin C in *Escherichia coli*. *Appl. Environ. Microbiol.* **71**, 2539–2547 (2005).
89. Reeves, C. D. *et al.* Alteration of the substrate specificity of a modular polyketide synthase acyltransferase domain through site-specific mutations. *Biochemistry* **40**, 15464–15470 (2001).
90. Kumar, P., Koppisch, A. T., Cane, D. E. & Khosla, C. Enhancing the modularity of the modular polyketide synthases: transacylation in modular polyketide synthases catalyzed by malonyl-CoA:ACP transacylase. *J. Am. Chem. Soc.* **125**, 14307–143112 (2003).
91. Gokhale, R. S., Tsuji, S. Y., Cane, D. E. & Khosla, C. Dissecting and exploiting intermodular communication in polyketide synthases. *Science* **284**, 482–485 (1999).
92. Tang, L., Fu, H. & McDaniel, R. Formation of functional heterologous complexes using subunits from the picromycin, erythromycin and oleandomycin polyketide synthases. *Chem. Biol.* **7**, 77–84 (2000).
93. Zhang, Y., Buchholz, F., Muyrers, J. P. & Stewart, A. F. A new logic for DNA engineering using recombination in *Escherichia coli*. *Nature Genet.* **20**, 123–128 (1998).
94. Gust, B. *et al.* λ Red-mediated genetic manipulation of antibiotic-producing *Streptomyces*. *Adv. Appl. Microbiol.* **54**, 107–128 (2004).
95. Wenzel, S. C., Gross, F., Zhang, Y., Fu., J., Stewart, A. F. & Muller, R. Heterologous expression of a myxobacterial natural products assembly line in pseudomonads via Red/ET recombineering. *Chem. Biol.* **13**, 349–356 (2005).
96. Kim, B. S., Sherman, D. H. & Reynolds, K. A. An efficient method for creation and functional analysis of libraries of hybrid type I polyketide synthases. *Protein Eng. Des. Sel.* **17**, 277–284 (2004).
Describes a directed evolution approach to PKS engineering.
97. Kodumal, S. J. *et al.* Total synthesis of long DNA sequences: synthesis of a 32-kb polyketide synthase gene cluster. *Proc. Natl Acad. Sci. USA* **101**, 15573–15578 (2004).
98. Staunton, J. *et al.* Evidence for a double-helical structure for modular polyketide synthases. *Nature Struct. Biol.* **3**, 188–192 (1996).
Describes a model for the modular polyketide synthases.
99. Tsai, S. C. *et al.* Crystal structure of the macrocycle-forming thioesterase domain of the erythromycin polyketide synthase: versatility from a unique substrate channel. *Proc. Natl Acad. Sci. USA* **98**, 14808–14813 (2001).
Reports the first crystal structure for a PKS domain.
100. Tsai, S. C., Lu, H., Cane, D. E., Khosla, C. & Stroud, R. M. Insights into channel architecture and substrate specificity from crystal structures of two macrocycle-forming thioesterases of modular polyketide synthases. *Biochemistry* **41**, 12598–12606 (2002).
101. Tsuji, S. Y., Cane, D. E. & Khosla, C. Selective protein–protein interactions direct channeling of intermediates between polyketide synthase modules. *Biochemistry* **40**, 2326–2331 (2001).
102. Broadhurst, R. W., Nietlispach, D., Wheatcroft, M. P., Leadlay, P. F. & Weissman, K. J. The structure of docking domains in modular polyketide synthases. *Chem. Biol.* **10**, 723–731 (2003).
Reports the NMR solution structure of a PKS docking-domain complex.
103. Milne, J. L. S. *et al.* Molecular architecture and mechanism of an icosahedral pyruvate dehydrogenase complex: a multifunctional catalytic machine. *EMBO J.* **21**, 5587–5598 (2002).
104. Leadlay, P. & Baerga-Ortiz, A. Mammalian fatty acid synthase: closure on a textbook mechanism? *Chem. Biol.* **10**, 101–103 (2003).
105. Kim, C. Y. *et al.* Reconstituting modular activity from separated domains of 6-deoxyerythronolide B synthase. *Biochemistry* **43**, 13892–13898 (2004).
106. Marsden, A. F. A. *et al.* Stereospecific acyl transfers on the erythromycin-producing polyketide synthase. *Science* **263**, 373–380 (1994).
107. Beck, B. J., Aldrich, C. C., Fecik, R. A., Reynolds, K. A. & Sherman, D. H. Substrate recognition and channeling of monomodules from the pikromycin polyketide synthase. *J. Biol. Chem.* **278**, 42020–42026 (2003).
108. Cane, D. E., Kudo, F., Kinoshita, K. & Khosla, C. Precursor-directed biosynthesis: biochemical basis of the remarkable selectivity of the erythromycin polyketide synthase toward unsaturated triketides. *Chem. Biol.* **9**, 131–142 (2002).
109. Aggarwal, R., Caffrey, P., Leadlay, P. F., Smith, C. J. & Staunton, J. The thioesterase of the erythromycin-producing polyketide synthase: mechanistic studies *in vitro* to investigate its mode of action and substrate specificity. *J. Chem. Soc., Chem. Commun.* 1519–1520 (1995).
110. Haydock, S. F. *et al.* Divergent sequence motifs correlated with the substrate specificity of (methyl)malonyl-CoA:acyl carrier protein transacylase domains in modular polyketide synthases. *FEBS Lett.* **274**, 246–248 (1995).
Describes sequence motifs in AT activities that can be used to predict their substrate specificity.
111. Kao, C. M. *et al.* Alcohol stereochemistry in polyketide backbones is controlled by the β-ketoreductase domains of modular polyketide synthases. *J. Am. Chem. Soc.* **120**, 2478–2479 (1998).
112. Kao, C. M. *et al.* Gain-of-function mutagenesis of the erythromycin polyketide synthase. 2. Engineered biosynthesis of an eight-membered ring tetraketide lactone. *J. Am. Chem. Soc.* **119**, 11339–11340 (1997).
113. Reid, R. *et al.* A model of structure and catalysis for ketoreductase domains in modular polyketide synthases. *Biochemistry* **42**, 72–79 (2003).
114. Hong, Y.-S. *et al.* Inactivation of the carbamoyltransferase gene refines post-polyketide synthase modification steps in the biosynthesis of the antitumor agent geldanamycin. *J. Am. Chem. Soc.* **126**, 11142–11143 (2004).
115. Butler, A. R., Bate, N., Kiehl, E. E., Kirst, H. A. & Cundliffe, E. Genetic engineering of aminodeoxyhexose biosynthesis in *Streptomyces fradiae*. *Nature Biotechnol.* **20**, 713–716 (2002).
116. Gaisser, S. et al. A defined system for hybrid macrolide biosynthesis in *Saccharopolyspora erythraea*. *Mol. Microbiol.* **36**, 391–401 (2000).
117. Gaisser, S. *et al.* Parallel pathways for oxidation of 14-membered polyketide macrolactones in *Saccharopolyspora erythraea*. *Mol. Microbiol.* **44**, 771–781 (2002).
118. Tang, L. & McDaniel, R. Construction of desosamine containing polyketide libraries using a glycosyltransferase with broad specificity. *Chem. Biol.* **8**, 547–555 (2001).
Describes an approach to creating libraries of engineered, bioactive polyketides.
119. Ziermann, R. & Betlach, M. Recombinant polyketide synthesis in *Streptomyces*: engineering of improved host strains. *Biotechniques* **26**, 106–110 (1999).
120. Desai, R. P., Rodriguez, E., Galazzo, J. L. & Licari, P. Improved bioconversion of 15-fluoro-6-deoxyerythronolide B to 15-fluoro-erythromycin A by overexpression of the *eryK* gene in *Saccharopolyspora erythraea*. *Biotechnol. Prog.* **20**, 1660–1665 (2004).
121. Lee, S. K. *et al.* The role of erythromycin C-12 hydroxylase, EryK, as a substitute for PikC hydroxylase in pikromycin biosynthesis. *J. Bioorg. Chem.* **32**, 549–449 (2004).
122. Gaisser, S. *et al.* New erythromycin derivatives from *Saccharopolyspora erythraea* using sugar *O*-methyltransferases from the spinosyn biosynthetic gene cluster. *Mol. Microbiol.* **41**, 1223–1231 (2001).
123. Xue, Y., Wilson, D., Zhao, L., Liu, H. & Sherman, D. H. Hydroxylation of macrolactones YC-17 and narbomycin is mediated by the PikC-encoded cytochrome P450 in *Streptomyces venezualae*. *Chem. Biol.* **5**, 661–667 (1998).
124. Zhang, Q. & Sherman, D. H. Isolation and structure determination of novamethymycin, a new bioactive metabolite of the methymycin biosynthetic pathway in *Streptomyces venezualae*. *J. Nat. Prod.* **64**, 1147–1150 (2001).
125. Yoon, Y. J. *et al.* Generation of multiple bioactive macrolides by hybrid modular polyketide synthases in *Streptomyces venezuelae*. *Chem. Biol.* **9**, 203–214 (2002).
126. Yang, M. *et al.* Probing the breadth of macrolide glycosyltransferases: *in vitro* remodeling of a polyketide antibiotic creates active bacterial uptake and enhances potency. *J. Am. Chem. Soc.* **127**, 9336–9337 (2005).
127. Spiteller, P. *et al.* The post-polyketide synthase modification steps in the biosynthesis of the antitumor agent ansamitocin in *Actinosynnema pretiosum*. *J. Am. Chem. Soc.* **125**, 14236–14237 (2003).
128. Walsh, C. T. Polyketide and nonribosomal peptide antibiotics: modularity and versatility. *Science* **303**, 1805–1810 (2004).
129. Kohli, R. M., Burke, M. D., Tao, J. & Walsh, C. T. Chemoenzymatic route to macrocyclic hybrid peptide/polyketide-like molecules. *J. Am. Chem. Soc.* **125**, 7160–7161 (2003).
130. Bloom, J. D. *et al.* Evolving strategies for enzyme engineering. *Curr. Opin. Struct. Biol.* **15**, 1–6 (2005).
131. Powell, K. A. *et al.* Directed evolution and biocatalysis. *Angew. Chem. Int. Ed. Engl.* **40**, 3948–3959 (2001).
132. Menzella, H. G. *et al.* Combinatorial polyketide biosynthesis by *de novo* design and rearrangement of modular polyketide synthase genes. *Nature Biotechnol.* **23**, 1171–1176 (2005).

Acknowledgements

The authors wish to thank the Royal Society Dorothy Hodgkin Fellowship and the Biotechnology and Biological Sciences Research Council for support.

Competing interests statement

The authors declare no competing financial interests.

Online links

DATABASES

The following terms in this article are linked online to:

Entrez: http://www.ncbi.nlm.nih.gov/Entrez

Bacillus subtilis | *Escherichia coli* | *Streptomyces avermitilis* | *Streptomyces coelicolor*

Swiss-Prot: http://www.expasy.org

Sfp

FURTHER INFORMATION

Kira J. Weissman's homepage:

http://www.bioc.cam.ac.uk/uto/weissman.html

Peter F. Leadlay's homepage:

http://www.bioc.cam.ac.uk/uto/leadlay.html

Access to this interactive links box is free online.