Polyketide biosynthesis: understanding and exploiting modularity

By Kira J. Weissman
Department of Biochemistry, University of Cambridge,
80 Tennis Court Road, Cambridge CB2 1GA, UK (kjw21@cus.cam.ac.uk)

Published online 24 September 2004

Polyketide-based pharmaceuticals are some of our most important medicines. They are constructed in micro-organisms (typically bacteria and fungi) by gigantic enzyme catalysts called polyketide synthases (PKSs). The organization of PKSs into molecular assembly lines makes them particularly appealing targets for genetic engineering because, in principle, an alteration in the enzyme organization might translate into a predictable change in polyketide structure. Excitingly, this has been shown repeatedly to work in practice, but the efficiency of the engineered PKSs is frequently too low to be useful for large-scale drug synthesis. To reach this goal, researchers need a deeper understanding of the structure and function of these proteins, which are among the most complex in nature. This review highlights some recent experiments which are providing key information about the molecular organization, mechanism and orchestration of these magnificent catalysts, and opening up fresh prospects of truly combinatorial biosynthesis of novel polyketides as leads in drug discovery.

Keywords: drug discovery; polyketide synthase; multi-enzyme; genetic engineering; combinatorial biosynthesis; erythromycin A

1. Introduction

Although you may not be on first-name terms with them, polyketides are good for you. They include antibiotics (e.g. erythromycin A), immunosuppressants (e.g. rapamycin), anti-parasitics (e.g. avermectin), cholesterol-lowering agents (e.g. lovastatin), anti-fungals (e.g. amphotericin B) and anti-cancer drugs (e.g. doxorubicin) (figure 1) (Staunton & Weissman 2001). In fact, polyketides show at least some activity against a far wider range of both acute and degenerative diseases. Combined sales from the more than 40 polyketide pharmaceuticals already generate more than £10 billion a year. Given their importance and potential, there is enormous worldwide interest in finding new versions of these molecules with improved therapeutic properties.

The majority of polyketides are made by micro-organisms (bacteria and fungi), so until now a very successful approach has been to search in the soil and in the oceans for undescribed and uncultured microbes, and then to isolate, identify and test the molecules housed inside them. As only a small percentage of soil bacteria have ever been grown in the laboratory and most marine micro-organisms

One contribution of 17 to a Triennial Issue ‘Chemistry and life science’.
remain wholly uncharacterized, this strategy should continue to reveal novel bioactive compounds for many years to come (Rouhi 2003). Another successful approach has been to alter the structures of known polyketides using standard synthetic chemistry. This approach has resulted, for example, in the highly successful semi-synthetic erythromycin-based antibiotics azithromycin (Zithromax) and clarithromycin (Biaxin). In general, though, the functional complexity of polyketide structures makes them difficult to modify in a precise way, and in particular, most chemical methods are entirely unsuited to changing their carbon skeletons.

In the late 1980s, inspired by the pioneering work of Professor Sir David Hopwood on the genetics of the biosynthesis of actinorhodin (figure 1) (Malpartida & Hopwood 1984), Peter Leadlay at the University of Cambridge and Leonard Katz, then at Abbott Laboratories in Chicago, began searching for the genes encoding for the assembly of erythromycin A in the soil bacterium *Saccharopolyspora erythraea*. In 1990, both groups published the sequences of the erythromycin genes (Cortés et al. 1990; Tuan et al. 1990; Donadio et al. 1991), an advance that has already begun to transform polyketide drug development. Having the gene sequences meant that it became, in principle, possible to modify any portion of erythromycin A by altering the corresponding step in its assembly.

Since their remarkable discovery, researchers have uncovered the genetic instructions for making over 50 additional polyketides. These sequences have come primarily from *Streptomyces* and related *Saccharopolyspora* species, but also from myxobacteria (e.g. Silakowski et al. 2001; Gerth et al. 2003), mycobacteria (e.g. Stinear et al. 2004), a marine cyanobacterium (Chang et al. 2002; Edwards et al. 2004), pseudomonads (El-Sayed et al. 2003), and from as yet uncultured bacterial symbionts of...
Understanding and exploiting PKS modularity

Figure 2. Correlation between the steps in making erythromycin and the complement of enzymes in its assembly line. The structure requires seven units of propionate (shown in red), so there are seven AT domains. Six KS domains catalyse the critical carbon–carbon bond-forming reactions (the new bonds are shown in orange). The pathway also includes all of the reductive activities (KR, DH and ER) required to modify the oxygen-bearing centres (shown in yellow). Finally, there is a TE to make the ring (new bond shown in green), and seven acyl carrier proteins to hold the chain as it grows. The modifications shown in black are added after the PKS is finished (post-PKS).

Figure 3. Organization of the erythromycin assembly line. (a) The 28 individual domains within the assembly line are grouped into (b) modules. The loading module initiates erythromycin manufacture by selecting the first building block, while the remaining modules (1–6) catalyse the six rounds of chain extension and reductive processing. (c) The modules themselves are organized into multi-modular subunits, called DEBS 1, 2 and 3.

other organisms (e.g. the Paederus beetle (Piel 2002)). All genes resemble those for erythromycin, although not in exact detail.

Six years after the discovery of the erythromycin cluster, the Cambridge group reported that a hybrid gene with a piece of the rapamycin polyketide synthase (PKS) patched into the erythromycin PKS successfully directed the synthesis of a new compound (Oliynyk et al. 1996). This result marked the start of a veritable explosion of activity in this area.

2. Polyketide synthases: nature’s assembly lines

The sequencing of the erythromycin genes revealed that nature has adopted an assembly-line process for polyketide biosynthesis, in that the number of chemical steps required to make the molecule matches the complement of enzymatic domains in the pathway (figure 2). Erythromycin contains seven three-carbon building blocks (propionate), and so there are seven acyltransferase (AT) domains that can select this material from the intracellular pool. Linking the building blocks together requires six new carbon–carbon bonds, and so there are six separate ketosynthase (KS) domains to carry out this key reaction (in chemical terms, a Claisen condensation). The multi-enzyme also contains the appropriate ketoreductase (KR), dehydratase (DH), and enoylreductase (ER) activities to carry out all specific reductive processing at oxygen-bearing centres. Finally, there is a thioesterase (TE) domain, at the end of the assembly line, that terminates chain construction and releases the product. Throughout the manufacturing process, the incomplete polyketide chains are tethered to small domains called acyl carrier proteins (ACPs) through long ‘swinging arms’ (Weissman et al. 2004), which ensures that they are correctly shepherded among the various active centres rather than wandering off into the cell.

The domains carrying these active centres are joined together into large polypeptides, in approximately the order they are used, to form what is essentially a Ford production line for polyketide biosynthesis, called a PKS (figure 3). In the case of erythromycin biosynthesis, these gigantic proteins are named DEBS 1, 2 and 3, after the polyketide intermediate that they produce, 6-deoxyerythronolide B (6-dEB) (Donaire et al. 1991; Bevitt et al. 1992). Within these enormous protein subunits, sets of activities are grouped into modules, such that each module adds one building block to the growing chain, and carries out any essential processing, before handing the intermediate on to the next module in the line. (In passing, it should be mentioned that simpler, less highly processed polyketides such as actinorhodin (figure 1) (Fernandez-Moreno et al. 1992) are constructed by iterative synthases which re-use a single set of enzymes several times, as do most fungal PKSs.)

The product of the PKS is very often only an intermediate on the way to becoming a bioactive compound. Once it is released from the PKS, it is set upon by a host of other ‘post-PKS’ enzymes which may add functional groups (methyls, hydroxyls and carbonyls, among others) or which decorate the structure with sugars (Gaisser & Leadlay 1998; Borisova et al. 1999; Tang & McDaniel 2001; Blanco et al. 2001). In the case of a group of polyketides called polyethers, such activities carry out a series of oxidations and ring-forming reactions, which totally alter the molecular shape (Oliynyk et al. 2003; Hughes-Thomas et al. 2003).

3. The promise of modularity

The strikingly repetitive organization of the modular PKSs (even at the level of the gene sequences) suggests that they have evolved by duplication and shuffling of modular units, whether as domains, modules or multiple modules (Bedford et al. 1996); once relocated, a particular piece of DNA would undergo random mutation and further evolution. These processes, coupled with the generous exchange of DNA between different bacteria, could explain how the genes for making polyketides ended up in so many different species of bacteria and fungi, and yet still resemble each.
**Understanding and exploiting PKS modularity**

Table 1. List, for each type of domain, of characteristics that would be ideal for combinatorial biosynthesis.

<table>
<thead>
<tr>
<th>Domain Type</th>
<th>Requirements for Combinatorial Use</th>
</tr>
</thead>
<tbody>
<tr>
<td>KS</td>
<td>partner any ACP; accept any incoming chain; accept any building block in chain extension</td>
</tr>
<tr>
<td>AT</td>
<td>use exclusively one loading/extender unit; transfer building block to any ACP domain</td>
</tr>
<tr>
<td>KR</td>
<td>carry out exclusively ‘A-type’ or ‘B-type’ ketoreduction, irrespective of substrate structure</td>
</tr>
<tr>
<td>DH, ER</td>
<td>perform reactions in a substrate-independent manner</td>
</tr>
<tr>
<td>ACP</td>
<td>be recognizable to all other domains within the module</td>
</tr>
<tr>
<td>TE</td>
<td>perform regiospecific macrolactonization (or alternatively hydrolysis)</td>
</tr>
</tbody>
</table>

other. Importantly, this model for PKS evolution suggests that we might be able to replicate this process of drug discovery in the laboratory, by exchanging pieces of PKS genes among existing synthases. If the changes we make give a predictable outcome, then we can imagine in future carrying out truly ‘combinatorial’ biosynthesis in which millions of novel PKSs, of varying lengths and composition, are created by combining different domains and/or modules, and each oversees the production of a correspondingly novel polyketide.

### 4. Combinatorial biosynthesis: the challenges

For this combinatorial strategy to work, there are a number of engineering challenges: these include host cells that would survive making these new molecules (which is certainly not guaranteed, given that many polyketides are antibiotics) and which contain all of the necessary chemical building blocks, as well as enzymes capable of activating the PKS proteins by adding swinging arms to the ACPs. However, the most important components are the PKS domains themselves, each of which must satisfy particular requirements in order to be useful in combinatorial biosynthesis. For example, KS domains must be able to partner with any ACP domain they encounter, accept any incoming polyketide change regardless of its structure, and accept any building block in chain extension. The requirements for the other domains are shown in table 1.

Clearly, substrate preference is a central issue, as once a domain is relocated it must be able to recognize and process any unnatural polyketide that it encounters. This applies in particular to the KS, reductive and TE domains, whose substrates are determined by the action of previous modules (in contrast, the AT domains need only select extender units from the cellular pool and affix them to ACPs). Therefore, to construct functional hybrids we will have to explore the substrate tolerance of individual domains, in the hope of identifying some that are particularly ‘promiscuous’ (e.g. Dutton et al. 1991; Jacobsen et al. 1997; Weissman et al. 1998; Cane et al. 2002; Watanabe et al. 2003). In parallel, we will need to understand the molecular basis for this substrate choice so that if suitable domains cannot be found, we might be able to engineer activities to exhibit a greater level of flexibility.

Figure 4. Set of reactions required to catalyse one round of polyketide chain extension: (a) the chain is transferred from the ACP domain upstream to the KS; (b) the AT selects an extender unit and loads it onto the ACP; (c) the KS catalyses carbon–carbon bond formation; the product remains attached to the swinging arm of the ACP; (d) the chain is passed to the reductive activities (KR, DH, ER) for processing; (e) the chain is transferred to the KS in the downstream module.

Assembling a polyketide can be a complicated affair, involving, in some cases, over a hundred individual enzymes. At the heart of each module is the ACP domain that holds onto the polyketide chain as it is lengthened and modified. For chain extension to happen, the ACP in turn interacts with all of the other domains within its own module, as well as with the KS in the adjacent module (figure 4). Although the details of these processes are largely unknown, it is clear that some mechanism must prevent premature handing-on of the chain to the next module before the processing steps are complete.

Given the subtlety of their interactions, one of the most important challenges in combinatorial engineering is how to attach domains, modules or subunits together. The resulting ‘hybrid’ proteins must fold properly so that all the domains are appropriately arranged around the ACP, and also interact precisely with each other. One strategy would be to use individual domains as the basic combinatorial unit. However, to do so, we will have to learn how to fit them together to form ‘designer’
Figure 5. The ‘mini-synthase’ DEBS 1-TE. This bimodular protein was created by relocating the chain-terminating thioesterase from the end of DEBS 3 to the end of DEBS 1, using genetic engineering. The new PKS produced the expected triketide lactone.

Figure 6. The structure of modular PKSs. (a) In the ‘Cambridge’ double-helix model, each subunit consists of two identical polypeptides that are twisted around each other into a helix. The core of the helix contains the KS, AT and ACP activities, while the reductive domains form loops at the periphery of the molecule. Cross-linking experiments have shown that the KS of one subunit cooperates with the ACP on the opposite subunit, and in this model they are correctly placed to interact (see arrows). (b) The crystal structure of the erythromycin thioesterase. As predicted from the model and limited proteolysis studies, the TE is a dimer, with a twofold axis of symmetry (indicated). In this figure, a homology model of the adjacent ACP domain (ACP₆) has been docked against the TE.

modules. In natural modules, domains are joined by sequences of amino acids called ‘linkers’ that vary in size from 20 to over 250 amino acids, giving the appearance of beads on a string. We might opt to join domains together in these linker regions or instead, at the domain boundaries (thereby keeping the linkers intact). In order
to evaluate these options, we need to understand the role that linkers play in organizing domains within each module. Not only that, we will need to understand how the activities recognize each other, so that the appropriate interfaces can be created between unnatural partners.

In principle, it would seem simpler to treat whole modules as combinatorial units, because the interactions between their domains have already been optimized. Again, however, the issue is how to join modules together most effectively; specifically, whether to preserve the linkers between the modules or not (Ranganathan et al. 1999; Gokhale et al. 1999a). Experiments have shown that failing to connect modules correctly can result in the inserted module simply being ‘skipped’ over (Thomas et al. 2002), or can completely disable the whole synthase (Gokhale et al. 1999a). Regardless of the strategy, it would be very useful to be able to develop a set of universal ‘interdomain’ and ‘intermodular’ linkers that could be used to join any pieces of PKS together, regardless of their sequences, to make a functional assembly line.

Typically, the modules of a modular PKS are distributed over several different proteins, which means that the polyketide chains have to be able to pass between them. But not only do the individual subunits have to identify their correct partners, they have to resist forming incorrect associations that might result in the wrong products. These interactions are handled, at least in part, by folded structures called ‘docking domains’, located at the ends of PKS proteins (Broadhurst et al. 2003). Therefore, to allow us to construct synthases containing multiple proteins, we have to understand how these docking domains enable both association and discrimination between the multi-enzymes. Such insights could then be used to develop a set of mutually exclusive (orthogonal) pairs of docking domain that can simply be attached to the ends of engineered PKS proteins to ensure that the assembly line comes together in the correct order.

5. How combinatorial are PKSs?

One obvious way to assess the combinatorial potential of PKSs is simply to try to engineer them and see what happens. Such experiments might yield an empirical set of design rules for putting PKS pieces together, as well as revealing particular domains, modules or subunits that meet the criteria for combinatorial use. An important early example of this strategy involved truncating the erythromycin PKS by relocating the chain-terminating thioesterase activity to the end of the first subunit, to give a mini-synthase called DEBS 1-TE (figure 5) (Cortés et al. 1995; Kao et al. 1995). This protein assembled the predicted small-molecule triketide lactone, demonstrating that the TE domain could cyclize a significantly smaller substrate than normal. This type of experiment has also yielded valuable information about the likely structure of many PKS-bound intermediates (for a recent example on the pathway that produces the polyketide insecticide spinosyn, see Martin et al. (2003)). Other recent engineering experiments include altering domains (additions, deletions and swaps) (e.g. Gaisser et al. 2003; Petkovic et al. 2003; Starks et al. 2003), and exchanging whole modules (e.g. Gokhale et al. 1999a) or multi-modular subunits (e.g. Kim et al. 2002). It is clear that the outcome is improved if several alternative ‘donor’ pieces of PKS are tried in addition to several different positions for cutting and pasting the donor DNA into its new context (Petkovic et al. 2003). However, many of these experiments fail to produce the expected polyketides (and even if
they are made, the yields are depressingly low), and it is usually difficult to trace
the source of the problem.
In this situation, we clearly need to know more about fundamental aspects of
polyketide construction: details of how selective individual domains are for their
substrates; the three-dimensional structure of domains, modules and subunits; how
domains and modules recognize each other, and how they link or dock together.
Eventually, this approach might bring us closer to achieving truly combinatorial
biosynthesis, but even in the short term it should allow more efficient directed alter-
ations to synthases with the aim of producing particularly desirable new structures.

6. What shape are PKSs?
We do not yet know the three-dimensional structure of a PKS multi-enzyme, so our
working model is based on one developed by Jim Staunton and Peter Leadlay in
1996 (Staunton et al. 1996), which aimed to show how the entire range of natural
PKS modules might be accommodated in a common structure. The ‘Cambridge
model’ was deduced from experiments in which purified DEBS multi-enzymes were
progressively chopped into their constituent modules and domains by proteolytic
enzymes, followed by careful characterization of the fragments (Aparicio et al. 1994;
Staunton et al. 1996). These experiments revealed that each of the three different
DEBS multi-enzymes is present in duplicate, and that the two identical copies bind
tightly to each other (in technical terms, they are ‘dimeric’). Individual domains or
groups of domains (e.g. KS–AT didomains, the TE) remained as dimers, even when
released from the rest of the PKS. In contrast, the reductive activities (KR, DH and
ER) appeared to be singletons.
Staunton and Leadlay proposed that the twin polypeptides of each subunit twist
around each other to form a double helix (figure 6a). In this model, the KS, AT,
ACP and TE domains sit at the core of the structure, directly against their counter-
parts on the opposite protein, while the reductive activities (KR, DH, ER) occupy
positions in loops on the periphery of the molecule. The interaction between one
pair of identical modules and those on either side appears rather limited, restricted
to contacts between ACP and KS domains and their associated linkers or docking
domains. This topology nicely accounts for several features of the assembly process
that had been observed experimentally. For example, the KS domain of one subunit
is correctly positioned to interact with the ACP of the other (figure 6a) (Staunton
et al. 1996).
Such a model is challenged by the increasing number of modular PKSs that have
been found to have unusual domain organizations, for example, independent ATs
(Piel 2002; Cheng et al. 2003), or the presence of multiple ‘core’ activities (for example,
Additionally, we need much more high-resolution structural information to reveal
the structure and location of the linker regions between domains and modules, as
well as how subunits dock together.
Very recently, Robert Stroud and his colleagues at the University of California,
San Francisco, have solved the first X-ray crystal structures of an individual PKS
domain: the TE of DEBS (figure 6b) (Tsai et al. 2001) and of the pikromycin PKS
(Tsai et al. 2002), respectively. As expected from earlier work (Aparicio et al. 1994;
Staunton et al. 1996), the TEs are dimers with twofold rotational symmetry. The

Figure 7. Schematic of experiments in which docking domains have been swapped. (a) The circled regions indicate the portions of the PKS called ‘docking domains’ which we think play a role in protein–protein recognition. (b) The matched docking domain pair which normally allows modules 4 and 5 to communicate was put in place of the docking domains which normally sit between modules 2 and 3; chain transfer between the modules was largely unaffected. In contrast, when the docking domains between modules 2 and 3 were intentionally mismatched, the modules lost much of their ability to cooperate in chain assembly.

The characteristic shape of the substrate channels accords with the known substrate preferences of each enzyme. As with other multi-enzymes (Milne et al. 2002), the piecemeal, domain-by-domain approach to obtaining detailed structure can obviously be extended to all the PKS domains. Together with sophisticated modelling based on known enzyme structures, these should provide an ever-clearer view of PKS domains and especially of their active centres.

7. Investigating linkers

Several groups have developed computational methods for determining the locations of linkers in modular PKSs (Udwary et al. 2002; Yadav et al. 2003a, b; Ansari et al. 2004). From their sequences alone, it is obvious that linkers are not simply ‘molecular strings’ (Perham 2000). Their amino acid content confers a degree of stiffness to them that would appear to keep domains apart yet permit them to move about relative to each other. Chaitan Khosla (Stanford University), David Cane (Brown University) and their colleagues have even proposed that the exact sequence of the linkers between modules (‘intermodular linkers’) is critical and that these linkers actively participate in productive polyketide chain transfer (Gokhale et al. 1999a). They suggest that, in order to join modules together, the linker associated with the upstream module must be preserved in its entirety.

In this model, the linkers act as switches which allow the ACP domain at the end of a module to change from partnering the KS of its own module to partnering the KS of the module downstream, so that intermodular chain transfer can occur (Gokhale & Khosla 2000). They further suggest that the linkers assume alternative structures in order to perform these functions, and this is why the specific sequence of residues is crucial (Gokhale et al. 1999a). If they are correct, it will be impossible to develop a set of universal linkers for joining modules together, because modules must be kept together with their specific downstream linker sequences. Certainly, we require detailed structural information on linkers to test such ideas directly.
8. How proteins dock

In the Cambridge model for PKS structure, the cooperating multi-enzymes are stacked one on top of another. This model predicts that any docking interactions between successive subunits will occur at their ends between ACP and KS domains, respectively. The ‘interpolypeptide linkers’ highlighted by the Brown and Stanford researchers (Gokhale & Khosla 2000) correspond to the short sequences of amino acids at the termini that obviously do not belong to the nearby domains. By analysing their sequences using structure-prediction programs, we discovered that they are likely to form α helices (three at the end of the proteins and one longer helix at the beginning).

They could directly demonstrate that these regions—which we have since re-named ‘docking domains’ (Broadhurst et al. 2003)—play a role in docking one multi-enzyme with another: when they swapped matched pairs of docking domains, chain transfer between adjacent DEBS modules was not badly disrupted (figure 7) (Tsuji et al. 2001; Wu et al. 2001). In contrast, intentionally mismatching docking domains reduced (but did not abolish) the transfer efficiency (Tsuji et al. 2001; Wu et al. 2002) (see figure 7).

Individual docking domains are small enough for their structures to be solved by nuclear magnetic resonance (NMR) (in full) as well as by X-ray crystallography. However, we wished to solve the structure of the complex, which would be likely to reveal more about how they work together. At first, therefore, we worried about...
how to capture such a short-lived complex: the literature suggested that even correct partners interact only weakly ($K_{diss}$ of the order of $\mu$M) (Gokhale et al. 1999b; Staunton et al. 1996; Caffrey et al. 1992), and such unstable interactions are notoriously difficult to analyse by NMR or X-ray crystallography (Noreen & Thornton 2003). However, we knew that PKS subunits could be joined together end-to-end through their docking domains to make even larger proteins, and that they continued to make polyketides (McDaniel et al. 1997; Squire et al. 2003; Olano et al. 2003). We decided therefore to try to look at the docking domains in isolation, but fused together end-to-end in exactly the same way. These model ‘Dock’ proteins were, we found, much more stable than their unlinked counterparts, and folded into a unique structure. We recently solved the structure of one of these (called ‘Dock 2–3’, which models the interaction between DEBS 2 and 3), at atomic resolution, by using multidimensional NMR (figure 8) (Broadhurst et al. 2003).

The structure of this model protein revealed that docking domains play two critical roles: not only do they help subunits to associate, as we had suspected, but they also appear to assist them in becoming and remaining dimers in the first place. Two features of the docking domain complex stand out: it is a symmetrical dimer (i.e. each polypeptide is present twice), and it contains two non-interacting structural units (labelled domain ‘A’ and domain ‘B’ in figure 8). Our strongest evidence for their role in dimerization is that elements of our structure have been authenticated in other systems as dimerization motifs. The first of these is a very unusual, intertwined bundle of helices in domain A; the only other known protein which shares this shape is the dimerization domain of a DNA-binding protein called HNF-1α (Rose et al. 2000). The second dimerization motif lies within domain B, and is formed by the two copies of the N-terminal docking domain of DEBS 3 that wrap around each other to create a ‘coiled-coil’. Coiled-coil formation is a well-known way for proteins to associate with and stabilize each other (Branden & Tooze 1999). In both cases, the particular hydrophobic residues (the amino acid ‘glue’) which hold the helices together are shared among many different synthases, which suggests that they all may use their docking domains to stabilize the dimer.

The actual docking interaction occurs in domain B: the two copies of the helix at the end of DEBS 2 wrap around the coiled-coil, and in doing so, form a second bundle of four helices with a different topology to that in domain A. Again, these four helices are held together by hydrophobic residues that are shared among many docking domains. (A large subset of docking domains do not share these sequence similarities, and so it remains to be determined whether they have a similar structure.)

Having discovered that many PKS subunits assemble via a very similar set of residues, we were very keen to understand why PKS associations are not wholly indiscriminate. Indeed, the model structure also suggests a possible way in which the DEBS proteins might manage to ignore incorrect partners: two charged residues that are found at critical positions in the docking interface (figure 8). When DEBS 2 docks against DEBS 3 (and similarly DEBS 1 against DEBS 2), the charges at these positions are matched, while, if DEBS 2 tries to dock against itself or DEBS 1 with DEBS 3, the resulting charge–charge interactions are instead repulsive. In support of this idea, studies of the $\alpha$-helical proteins Fos and Jun have shown that their preference for forming a coiled-coil with each other instead of with another copy of themselves is based on appropriately placed charged residues (Glover & Harrison
1995). Although we will need to confirm their role by experiment, these charged residues are our best candidates for some sort of docking ‘code’ in DEBS.

Unfortunately, this code fails to explain what happens in other PKS systems. For example, to make the polyether antibiotic monensin, eight different polypeptides must organize themselves into the correct order; perplexingly, multiple sets of docking domain partners turn out to be identical at every critical position we have identified (Oliynyk et al. 2003). This finding strongly suggests that the docking domains are not the whole story, and that the ACP and KS domains play essential roles in recognizing correct PKS partners (Ranganathan et al. 1999; Wu et al. 2002). From a combinatorial perspective, this result is discouraging: engineering efficient chain transfer seems to require not only matched docking domains but also an optimized interface between the ACP and KS domains.

Possible insight into the nature of the ACP-KS interface in PKSs has recently been provided by a computational approach in which the NMR structure of a discrete ACP domain from a bacterial fatty acid synthase (FAS) system was docked against the crystal structure of an FAS KS (Zhang et al. 2001) (bacterial FASs have the same complement of domains as a PKS module, and so features of fatty acid biosynthesis have relevance for understanding polyketide assembly). This analysis revealed that the ACP docks against a region on the KS domain that is adjacent to the KS active site, and also highlighted the specific amino acids that are involved in this interdomain recognition. This docking site would allow the swinging arm of the ACP to reach into the KS domain as required during the condensation reaction. Whether such an arrangement occurs in PKSs remains to be determined.

9. How combinatorial are domains?

It has already proved very useful to study the structures and specificities of domains more directly. The goal here has been to identify domains of each type which are suited to combinatorial use, but also to gain the understanding of their shapes and functions that will allow us to alter their specificity by directed mutation within active centres, leaving the overall architecture of the domain undisturbed and minimizing the introduction of ‘bad’ contacts against adjacent domains. In this respect, AT domains have been particularly well studied.

AT domains make the critical choice of which building blocks to use in making a specific polyketide. The domains come in two types, ‘loading’ ATs, which in most PKSs initiate the assembly process by choosing a starter unit, and ‘extender’ ATs, which select the structural units used during chain extension. These building blocks are typically provided in activated form as acyl-CoAs, such as acetyl- or propionyl-CoA (for the chemistry of chain extension to work, the extender units must also contain \(\alpha\)-carboxylate groups; therefore, these building blocks are more formally malonyl-CoA and methylmalonyl-CoA).

Extender ATs may be particularly selective about which substrates they use (Khosla et al. 1999; Liou et al. 2002). For example, the DEBS ATs will only accept methylmalonate, and of its two possible stereoisomers, only the (2S)-isomer (Marsden et al. 1994). ATs in other synthases are sometimes less fussy, but they normally exhibit a preference for malonyl-CoA, ethylmalonyl-CoA, propylmalonyl-CoA or methoxymalonyl-CoA, respectively. As first shown for ATs that recognize malonate or methylmalonate (Haydock et al. 1995), specific sequences (motifs) reliably
predict which extender unit a particular AT will select. Therefore, by analysing the differences in their structures, we may be able to manipulate the substrate preferences of AT domains. As a step in this direction, two groups have recently demonstrated that motifs shared among AT domains are present at active centres and play a direct role in substrate choice (Reeves et al. 2001; Del Vecchio et al. 2003). The specificity of an AT was relaxed from methylmalonate only to malonate and methylmalonate by making tiny changes in specific residues. Expression and study of individual ATs as purified proteins has allowed quantification of these effects (Lau et al. 2000; Liou et al. 2002) and may eventually lead to a crystal structure.

10. Mycolactone PKS: nature’s combinatorial set?

Although much has been learned about how to engineer the biosynthesis of valuable polyketides by altering the genes for the PKSs, the development of truly combinatorial biosynthesis based on sets of domains and modules remains a daunting challenge. The key problem is that we cannot yet design ‘universal’ intermodular interfaces, nor broaden the specificity of individual enzymes without badly affecting catalytic efficiency.

In this context, it is therefore intriguing that the recently discovered PKS genes for the polyketide toxin mycolactone (Stinear et al. 2004) show evidence of very recent evolution, such that the natural sequence identity of comparable KS and ACP domains in all 16 modules is astonishingly high. Stewart Cole (Institut Pasteur, Paris) and his colleagues point out that interchanging such domains is less likely to provoke unfavourable interactions than in other PKSs, and (although it remains to be tested), that the KS and other enzyme domains may not yet have evolved a particular specificity. If this is proved correct, it will open up a new and promising vista for engineering.

11. The future

Polyketide synthases construct extraordinarily complex molecules using what may be regarded as a molecular assembly line. This organization has led to attempts to reconfigure the enzymes by swapping modular units such as domains, modules or multi-modular subunits. Done in a combinatorial fashion, this might provide a straightforward route to thousands of new drug leads. To date, genetic engineering of PKSs has resulted in over 100 new compounds, some of which show promise as drug candidates. For example, last year, a research team at Biotica (Cambridge, UK) created, as an alternative to costly semi-synthesis, a direct biological route to the valuable veterinary antiparasitic drug Ivermectin by appropriate engineering of the polyketide synthase for its parent compound avermectin (Gaisser et al. 2003). Meanwhile, at Kosan Biosciences in California, researchers have produced nine new versions of the anti-tumour polyketide epothilone (Starks et al. 2003).

Undoubtedly, the flood of information from genome sequencing projects will only increase our knowledge of PKS systems (see, for example, Zazapoulos et al. 2003; Koumoutsi et al. 2004; Weissman et al. 2004). These efforts will continue to provide alternative or novel domains for inclusion in a ‘combinatorial toolbox’. In addition, progress in understanding and manipulating the enzymes that perform the post-PKS elaboration of polyketide structures, particularly those that add sugars to the
Understanding and exploiting PKS modularity

structures (Borisova et al. 1999; Rodriguez et al. 2000; Gaisser et al. 2001; Tang & McDaniel 2001; Wohlert et al. 2001), should significantly increase the diversity of engineered polyketide libraries.

However, PKS systems are significantly more complicated than they superficially appear, so despite some solid successes, many such engineering experiments continue to fail. Therefore, in order to realize the immense potential of PKSs, we must gain a significantly deeper understanding, at the molecular level, of the way they operate.

From structural studies, a picture is already emerging of the topology of whole multi-enzymes, but also of domains and of the docking regions that allow the multiple subunits within each synthase to associate with each other. In parallel, mechanistic investigations of activities both in their natural contexts and, more recently, as individual domains, have begun to shed light on their innate substrate preferences and functions. Although much more needs to be done, it remains a tangible goal to engineer PKSs to make large collections of novel polyketides, rapidly, efficiently and in good yield. In the meantime, the synthases have much to teach us about the mechanism, structure and integration of complex, multi-enzyme systems.

The author thanks Professor Peter Leadlay for helpful comments on this manuscript, and the Royal Society Dorothy Hodgkin Fellowship for support.

References


Understanding and exploiting PKS modularity


AUTHOR PROFILE

Kira J. Weissman

Kira Weissman was born in 1973 in Stanford, California. She obtained her BS in Chemistry from Stanford in 1995, graduating with honours and distinction. She received a Churchill Fellowship for study at the University of Cambridge in 1995, and carried out her MPhil and PhD research in chemistry with Professor Jim Staunton. After obtaining her PhD in 1999, she was awarded a Junior Research Fellowship from Newnham College, Cambridge, to carry out postdoctoral work with Professor Peter Leadlay in the Department of Biochemistry. Since 2002, she has held a Dorothy Hodgkin Research Fellowship in the Biochemistry Department. Her research interests concern the structure and mechanism of polyketide synthase multi-enzymes, a class of enzymes which assemble some of our most important pharmaceuticals. In her spare time, she enjoys writing popular science articles, principally for *Chemistry World*. 