Polyketides, proteins and genes in fungi: programmed nano-machines begin to reveal their secrets

Russell J. Cox*

Received 22nd March 2007
First published as an Advance Article on the web 10th May 2007
DOI: 10.1039/b704420h

Fungi produce a wide variety of biologically active compounds. Among these, the polyketides form a large and structurally diverse group. These compounds are synthesised by highly programmed and very large iterative multifunctional proteins, the polyketide synthases, with nm dimensions. This review outlines the current state of knowledge regarding the links between gene sequence, protein architecture and biosynthetic programming for fungal polyketide synthases.

1.0 Introduction

Fungi inhabit almost all known environments on earth—from the antarctic, to the hyper saline waters of the dead sea,[1] from temperate soils and forests to deserts and from oceans to tropical rain forests. They also inhabit surprising environments such as the intestinal tracts of insects,[2] the phloem of plants[3] and even the epidermis of humans. Less surprising habitats include crops, foodstuffs and domestic animals. Conservative estimates suggest that there are around 1.5 million different fungal species on earth[4]—this compares with estimates of around 300 000 plant species. Around 75 000 (*ca.* 5%) of fungi have been collected and described but only a small proportion of these chemically investigated. Many fungi are pathogens—particularly of plants and insects, but also of mammals, and play a role in numerous human diseases.[5] The ability of fungi to opportunistically colonise new environments has even been linked to the extinction of the dinosaurs![6]

From a chemist's perspective, the fungi offer a fascinating example of molecular evolution. Fungi have existed on earth for at least one thousand million years,[7] and for much of that time they have exploited and evolved secondary metabolism for the production of bioactive compounds. These compounds include toxins such as α-amanitin **1** the toxic principal of *Amanita phalloides* (death cap), vomitoxin (deoxynivalenol) **2** produced by the family of plant pathogens the *Fusaria* and aflatoxin B_1 **3** produced by *Aspergillus* species. Fungi also produce potent psychoactive compounds such as muscimol **4**, psilocybin **5** and xenovulene A **6**, pharmaceuticals such as the β-lactams and the statin lovastatin **7a**. These compounds illustrate the utility and diversity of chemical structures produced by fungi, and the diverse biosynthetic potential of these organisms.

The biosynthetic pathways towards non-ribosomal peptides (precursors of **1**, for example), and towards terpenoids such as **2**, have been studied in considerable depth and are well understood. By contrast, detailed understanding of the early chemical steps leading to the polyketides such as lovastatin **7a**, aflatoxin B_1 **3**, and xenovulene A **6**, is only just yielding to chemical investigations.

1.1 Polyketide synthases: sources, genes and proteins

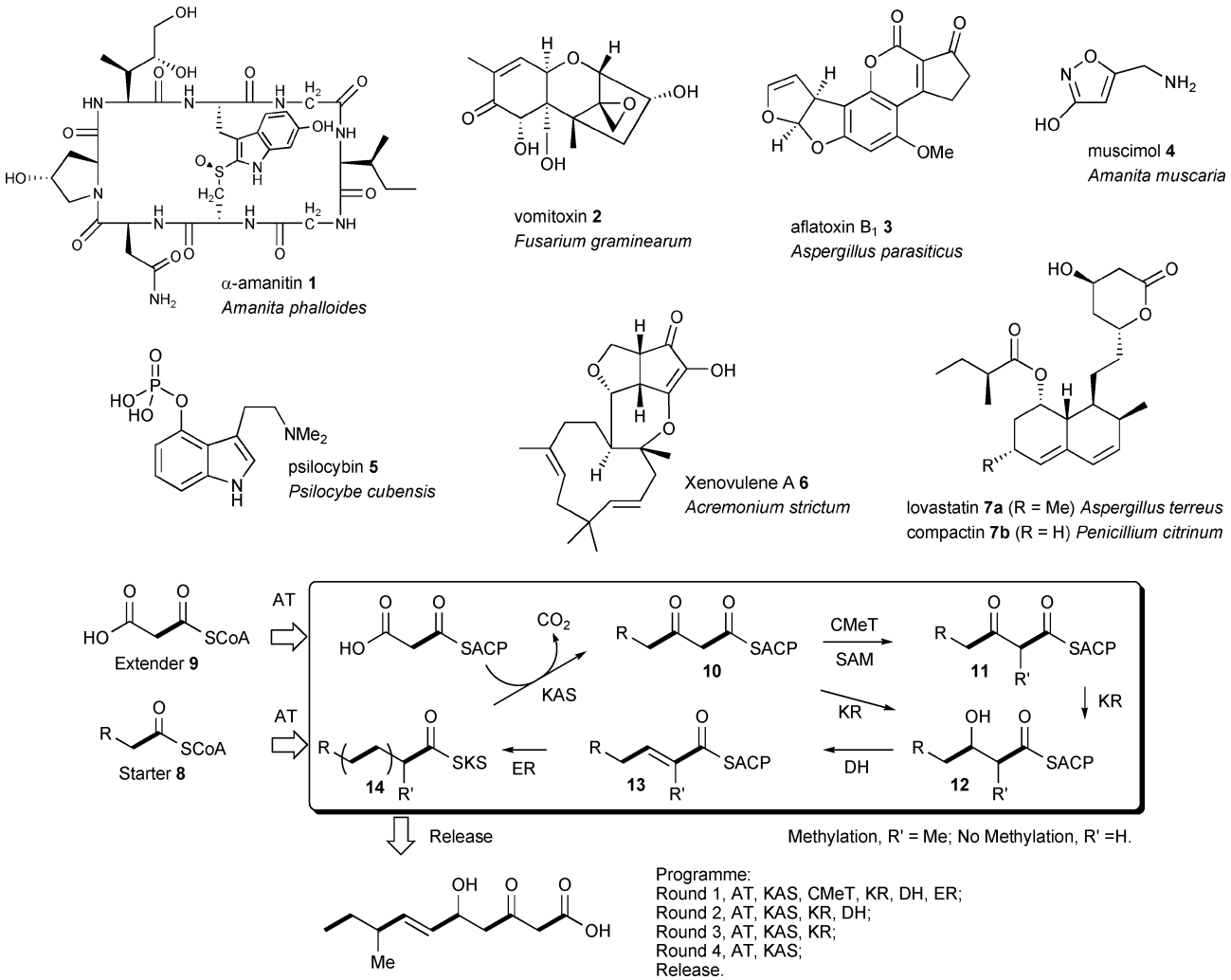
Polyketides have long been recognised as one of the most important classes of secondary metabolites. They occur in plants, bacteria and marine organisms as well as in fungi. Though diverse in structure, the class is defined by the common biosynthetic origin of the carbon atoms: these are derived from small carboxylic acids such as acetate, propionate and, rarely, butyrate. As long ago as 1953, Birch realised that polyketide biosynthesis is related to fatty acid biosynthesis.[8]

Numerous feeding experiments using isotopically labelled precursors showed that the key carbon–carbon bond forming step in polyketide and fatty acid biosynthesis is a decarboxylative Claisen reaction between an acyl thiolester **8** and a malonyl thiolester **9** (Scheme 1).[9] This reaction is catalysed by a β-ketoacyl synthase (KAS) enzyme and all fatty acid synthases (FAS) and

School of Chemistry, University of Bristol, Cantock's Close, Bristol, UK BS8 1TS. E-mail: r.j.cox@bris.ac.uk; Fax: +44 (0) 117 929 8611

Dr Russell Cox is Reader in Organic and Biological Chemistry in the School of Chemistry at the University of Bristol, UK. He was born in the New Forest in 1967. He studied for his first Degree in Chemistry at the University of Durham, where he stayed to study for a Ph.D. in bioorganic chemistry with Professor David O'Hagan. Post-doctoral positions with Prof. John Vederas at the University of Alberta, Canada, and Prof. David Hopwood at the John Innes Centre, UK, were followed by his first academic appointment at the University of Bristol in 1996.

Russell J. Cox



Scheme 1 Generic chemical reactions catalysed by iterative fungal PKS.

polyketide synthases (PKS) must possess KAS activity (Scheme 1). All FAS use an acyl carrier protein (ACP) to carry the malonyl extender units and transiently hold the growing acyl chain—this feature is preserved in almost all PKS where ACP is also required. The *apo*-ACP requires post-translational modification through the addition of phosphopantetheine (PP) derived from CoA to a conserved serine. This reaction is carried out by a PP transferase (also known as *holo*-ACP synthase, ACPS).[10–12] Most FAS and PKS proteins also require an acyl transferase (AT) enzyme to transfer acyl groups from CoA onto the KS and ACP components.

During fatty acid biosynthesis, the newly formed β-ketothiolester **10** is subjected to further chemical processing while attached to the terminal thiol of the ACP PP: first it is reduced by a β-ketoacyl reductase (KR) to a secondary alcohol **12**; this then undergoes a dehydratase (DH) catalysed dehydration to form an unsaturated thiolester **13**; and final enoyl reduction (ER) yields a fully saturated thiolester **14**. Fungal PKS deploy all these reactions, but additionally the chain can be methylated, using a methyl group from *S*-adenosylmethionine (SAM). This probably occurs after KAS, giving an α-methyl-β-ketothiolester **11**. During the biosynthesis of palmitate (C_{16}) there are seven cycles of these reactions. The final reaction of FAS is hydrolysis of the thiolester by a dedicated thiolesterase (TE).

The understanding of the relationship between FAS and PKS proteins, the application of molecular genetics, and more latterly genomics, has greatly facilitated the discovery and understanding of polyketide synthases from diverse sources. The homology in catalytic function between FAS and PKS enzymes is preserved in their respective gene sequences.

It is now clear that there are at least two different protein architectures for FAS proteins. Type I FAS proteins are large multifunctional proteins in which single (or sometimes two) peptides contain the sequences for KAS, ACP, AT, KR, DH, ER and TE activities—these catalytic functions are carried out by particular functional domains. The genes for type I FAS proteins are correspondingly large, single open reading frames. Type II FAS consist of complexes of single proteins, encoded by cognate single genes. Type I FAS are found in animals and fungi while type II FAS are found in plants and bacteria. Both type I and type II FAS are *iterative*—that is, they contain single copies of each catalytic function and these must act again and again to produce the mature fatty acid. Recent crystallographic investigations have revealed the type I mammalian FAS to be a nanoscale object, measuring approximately 18 × 21 × 9 nm (Fig. 1).[13]

PKS proteins are characterised in the same way: type I PKS consist of very large multifunctional proteins with individual

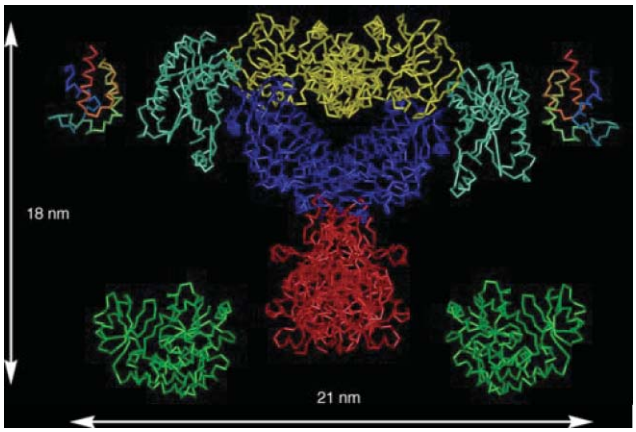


Fig. 1 Architecture of the mammalian FAS (PDB: 2CF2): green, AT; red, KAS; dark blue, DH; yellow, ER; light blue, KR. ACP (multicolour) shown in proposed position.

functional domains, type II PKS consist of individual proteins. Type III PKS (chalcone and stilbene synthases) do not correspond with any known FAS—these are very simple KAS proteins which do not require AT, ACP, KR, DH, ER or TE functionalities. Type I PKS are found in bacteria and fungi, type II PKS are restricted to bacteria while type III PKS are found in plants, bacteria and fungi.

PKS use the same array of chemical reactions as FAS—but the key difference is that of *programming*: both FAS and PKS have to control chain length (*i.e.* the number of extensions), but PKS are able to additionally control starter and extender unit selection and the extent of reduction during each condensation cycle. Fungal PKS are also able to programme the extent of chain methylation. The issue of programming is key to understanding and exploiting PKS. In the case of the bacterial modular polyketide synthases, each condensation cycle is catalysed by a discreet module containing all the catalytic domains required. In this case the programme is explicit in the order and composition of the modules.[9] However for the iterative type I fungal polyketide synthases, the programme is cryptic—encoded in the PKS itself.

Fungal PKS proteins closely resemble mammalian FAS proteins—in fact, much more closely than they resemble fungal FAS proteins. Thus the 3D model of the mammalian FAS (Fig. 1) is probably a good approximation of the fungal PKS structure. However, the fact that fungal PKS are programmed means that there is much to learn about the molecular programming of these nanoscale molecular machines.

1.2 Fungal polyketide synthases

Fungi make some of the simplest and some of the most complex polyketides known.[14] It is useful to consider a hierarchy of complexity when considering the structures of fungal polyketides, because the complexity in chemical structure is generated by enzymes and ultimately genes.[15]

The simplest structures are those such as orsellinic acid **15** (and its derivatives), a tetraketide, generated by many fungi in the genera *Aspergillus* and *Penicillium*. Addition of an extra acetate gives a pentaketide such as tetrahydroxynaphthalene **16**, a compound widely distributed in fungi and which is involved in melanization—a key component of apressorium formation and invasion of plant cells by plant pathogens such as *Magnaporthe grisea*, *Colletotrichum lagenarium* and others.[16] Similar aromatic compounds such as the anthraquinone norsolorinic acid **18** are the precursors of aflatoxin B_1 **3**[17] and the related sterigmatocystins.

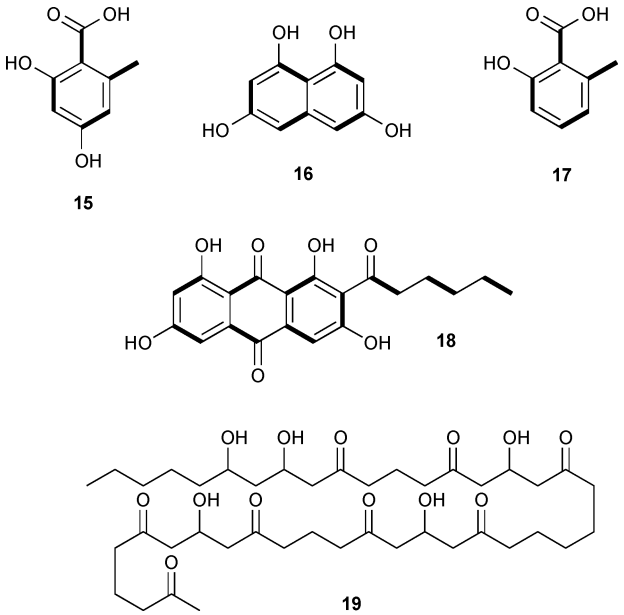
Further complexity is represented by compounds such as 6-methyl salicylic acid **17**, where a programmed reduction reaction occurs during biosynthesis. Additional complexity is then observed in compounds such as T-toxin **19** produced by *Cochliobolus heterostrophus*[18] and lovastatin **7a** produced by *A. terreus*. In these compounds, many more carbon atoms are used and many more reduction and dehydration reactions occur. Additionally, in many cases, pendant methyl groups have been added from the *S*-methyl of methionine, catalysed by a *C*-methyl transferase domain offering a further dimension for programming.

1.3 Linking PKS genes and compounds in fungi

The hierarchy of fungal polyketide chemical structures is due to differences in programming of their PKS proteins—apparent increases in product structural complexity are due to increasing use and control of reductive, dehydrative and methylating steps by the PKS. This must be due to differences in PKS protein sequence and structure. This fact has been exploited in the development of rapid methods for the cloning of fungal PKS genes associated with the biosynthesis of particular fungal polyketide types.

Simpson, Lazarus and Bingle realised that these subtle protein sequence differences should be reflected in DNA sequence and that polymerase chain reaction (PCR) primers could be designed to selectively amplify fragments of fungal PKS genes from fungal genomic DNA (or cDNA).[14] In early work in this area, Simpson, Lazarus and Bingle hypothesised that fungal polyketides could be grouped into two classes: non-reduced (NR) compounds such as orsellinic acid **15**, norsolorinic acid **18** and tetrahydroxy naphthalene **16**, and partially reduced (PR) compounds such as 6-MSA **17**. At the time, very few fungal PKS genes were known, and based on very limited sets of sequences, they designed degenerate PCR primers which were complementary to conserved DNA sequences in the KAS domains in fungal PKS responsible for the biosynthesis of NR and PR compounds.[14] Later, the same analysis was extended to the KAS domains of highly reduced (HR) compounds such as lovastatin **7a** when DNA sequence data became available for the lovastatin nonaketide and diketide synthases (LNKS and LDKS respectively).[15] The availability of these sequences also allowed the development of selective PCR primers for *C*MeT domains.

Application of these selective primers in PCR reactions with diverse filamentous fungi yielded DNA products which were fragments of new PKS genes. Detailed sequence analysis showed that the newly generated sequences clustered with other PKS gene sequences from the expected classes. For example, PCR primers designed based on NR PKS sequences amplified fragments of other NR PKS, primers based on 6MSA-type sequences amplified fragments of further examples of PR PKS genes. Likewise primers exploiting similarities between LNKS and LDKS gene sequences amplified fragments of new HR PKS genes.



This sequence analysis has been significantly extended as genomic approaches have been applied to fungi in the recent past.[19] Full genome sequences have now been obtained for more than a dozen fungi. In each organism, many PKS genes have been discovered. For example *Aspergillus niger* contains 34 PKS genes,[20] so there are now several hundred fungal PKS genes known. Sequence comparison of all these new PKS genes, however, shows that the three classes of fungal PKS genes predicted by Simpson, Lazarus and Bingle are the same three classes observed in the most recent sequence comparisons.[19]

Despite the fact that so many fungal PKS genes have been discovered, however, relatively few genes have been definitively linked to the biosynthesis of specific compounds (Table 1). This review seeks to clarify these links. Since the NR, PR and HR nomenclature is useful for describing both the chemical products and their cognate genes, it is sensible to survey the state of knowledge of fungal PKS in this way.

2.0 Fungal NR PKS

One of the first discovered fungal PKS, orsellinic acid synthase (OSAS) was isolated from *Penicillium madriti* and reported in 1968.[21] The tetraketide orsellinic acid **15** is the simplest tetraketide, requiring no reductions during its biosynthesis. Despite the early work with the protein, however, the OSAS encoding gene has not yet been discovered and nothing is known of the catalytic domains or their organisation. However, genes involved in the biosynthesis of a number of other non-reduced polyketides are now known and a general pattern of domain organisation has emerged. In all cases known to date, these genes encode type I iterative PKS proteins. At the N-terminus, a domain is present which appears to mediate the loading of a *starter unit* (Fig. 2). It appears that the starter unit can derive from either a dedicated FAS, another PKS or an acyl CoA. The starter unit loading domain is followed by typical KAS and AT domains responsible for chain extension and malonate loading. Beyond the AT is a conserved domain with as-yet unproven function. Sequence analysis of this domain, however, suggests that it may be involved in control of chain-length

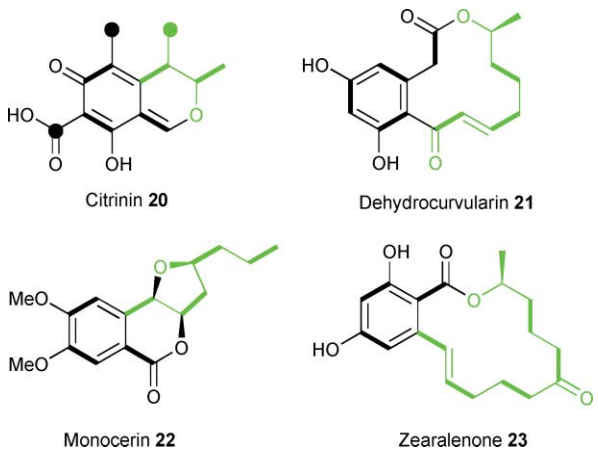
Fig. 2 General architecture of NR PKS genes in fungi. SAT, starter unit ACP transacylase; KAS, β-ketoacylsynthase; AT, acyl transferase; PT, product template; ACP, acyl carrier protein; TE, thiolesterase; CLC, Claisen cyclase.

(*vide infra*). This is followed, in-turn, by an ACP. Some NR PKS appear to terminate after the ACP, but many feature a diverse range of different domains including cyclases, methyl transferases, and reductases. It thus appears that these synthases are arranged with an N-terminal *loading component*, a central *chain extension component* consisting of KS, AT and ACP domains with possible control over the number of extensions, and a C-terminal *processing component*.

These three components correspond to the three important elements of programming—starter unit selection, chain-length determination and post-PKS processing. These features are exemplified in the examples outlined below.

2.1 NR PKS loading component

Feeding experiments with isotopically,[17] and in some cases ^{19}F,[22] labelled precursors have shown that many NR fungal polyketides are formed by the use of 'advanced' starter units. For example in the classic case of norsolorinic acid **18** biosynthesis, it has long been known that hexanoate forms the starter unit. Differential specific incorporations of acetate into the early and late positions in compounds such as citrinin **20** have been used to argue that these compounds may have been formed by more than one PKS so that one PKS makes an advanced starter unit which is passed to a second PKS for further extension.[23]



Dehydrocurvularin[24] **21**, monocerin[25] **22** and zearalenone[26] **23** are all good potential examples of this: here the structures of the compounds indicate high levels of reduction early during biosynthesis and no reductions during later steps. In the case of dehydrocurvularin **21**, advanced labelled tetraketide precursors were incorporated during biosynthesis, possibly indicating that a tetraketide formed by a HR PKS forms the starter unit for a second NR-PKS.[27]

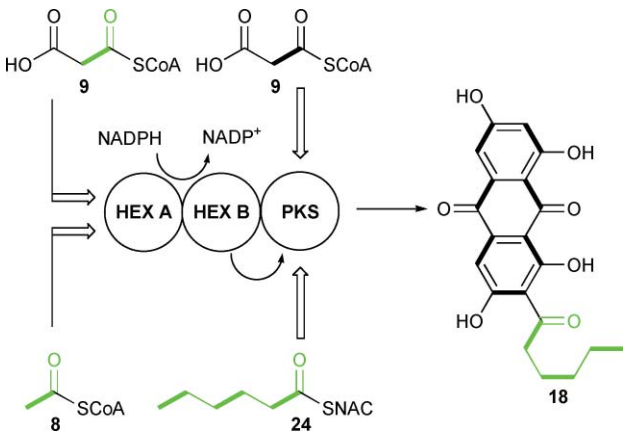
Very recently Townsend has defined the molecular basis for the ability of NR-PKS to utilise starter units derived from other

Table 1 Proven relationships between PKS genes and compounds in fungi for non-, partially- and highly-reducing PKS

Organism	Gene	Protein	Final product	Ref.
Aspergillus parasiticus	*pksA*	NSAS	Aflatoxin B_1 **3**	17
Aspergillus nidulans	*pksST*	NSAS	Sterigmatocystin	17
Dothistroma septosporum	*pksA*	NSAS	Dothistromin	55
Aspergillus nidulans	*wA*	WAS	YWA1 27	46
Aspergillus fumigatus	*alb1*	alb1p	YWA1 27	53
Colletotrichum lagenarium	*PKS1*	THNS	Tetrahydroxy naphthalene **16**	48
Wangiella dermatitidis	*WdPKS1*	THNS	Tetrahydroxy naphthalene **16**	38
Gibberella zeae	*PKS13*	ZS–B	Zearalenone **23**	32
Monascus purpureus	*pksCT*	CitS	Citrinin **20**	56
Penicillium patulum	*MSAS*	MSAS	6-MSA **17**	59
Aspergillus terreus	*atX*	MSAS	6-MSA **17**	61
Glarea lozoyensis	*pks2*	MSAS	6-MSA **17**	62
Aspergillus terreus	*lovB*	LNKS	Lovastatin **7a**	72
Aspergillus terreus	*lovF*	LDKS	Lovastatin **7a**	72
Penicillium citrinum	*mlcA*	CNKS	Compactin **7b**	73
Penicillium citrinum	*mlcB*	CDKS	Compactin **7b**	73
Giberella zeae	*PKS4*	ZS–A	Zearalenone **23**	32
Phoma sp.	*PhPKS1*	SQTKS	Squalestatin tetraketide **48**	77
Cochliobolus heterostrophus	*pks1*	TTS1	T-toxin **19**	78
Cochliobolus heterostrophus	*pks2*	TTS2	T-toxin **19**	79
Giberella fujikuroi	*fum1*	FUMS	Fumonisin B1 **40**	80
Fusarium moniliforme	*ORF3*	FUSS	Fusarin C **56**	87
Fusarium heterosporum	*eqiS*	EQS	Equisetin **57**	89
Beauveria bassiana	*ORF4*	TENS	Tenellin **58**	91

FAS or PKS systems. Townsend realised that two genes in the aflatoxin biosynthetic gene cluster of *A. parasiticus* (*stcJ* and *stcK*) encoded the α and β components of a typical fungal FAS (HexA and HexB).[28] Clustering of these FAS genes with the NSAS PKS suggested that HexA and HexB probably produced hexanoate for use as the norsolorinic acid starter unit.

The protein complex formed between NSAS, HexA and HexB, known as NorS, was then isolated and characterised.[29] This 1.4 MDa protein complex synthesises norsolorinic acid **18** from acetyl CoA **8**, malonyl CoA **9** and NADPH. Townsend showed that hexanoyl CoA is not a free intermediate produced by NorS, suggesting that hexanoate produced as an ACP derivative by the HexA–HexB FAS must be passed directly to NSAS. In the absence of NADPH, hexanoate cannot be formed by the FAS components and thus no norsolorinic acid is formed. Norsolorinic acid biosynthesis is restored, however, when exogenous hexanoyl SNAC **24** is used. However, the extreme difficulty of working with a protein complex of 1.4 MDa consisting of three very large proteins at WT concentrations precluded further studies into the transfer mechanism (Scheme 2).

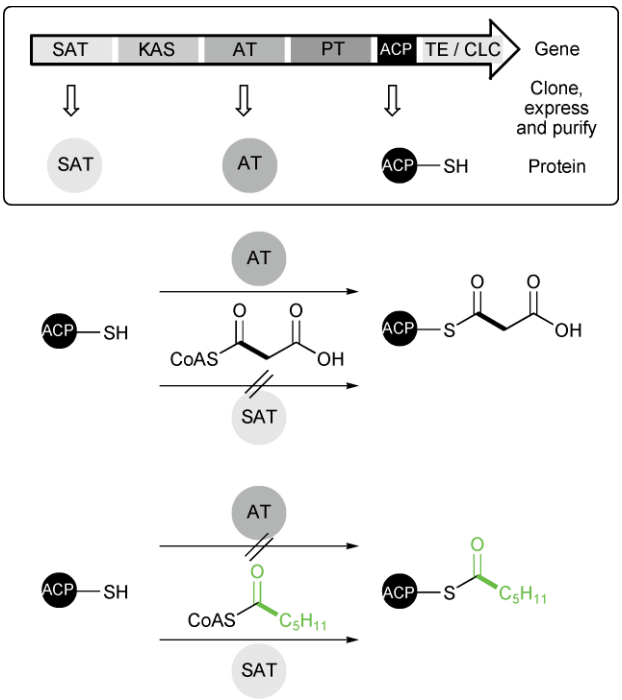


Scheme 2 Substrates of the NorS complex.

Complementary experiments carried out by Cox and coworkers involved the cloning and heterologous expression of the NSAS AT and ACP catalytic domains.[30] These were shown to be properly folded and catalytically active *in vitro.* The AT domain catalysed the transfer of malonyl extender groups from CoA onto ACP, but did not appear to transfer potential acyl starter groups from CoA or FAS ACP ruling out its involvement in starter unit provision (Scheme 3).

Townsend *et al.* realised that the N-terminal domain of NSAS also possesses canonical acyl transferase sequence motifs and that this domain could be a candidate for the required starter unit transferase.[31] The N-terminal domain and ACP were cloned and heterologously expressed. *In vitro*, the N-terminal protein was shown to catalyse the transfer of hexanoate from CoA onto the ACP. Site directed mutagenesis experiments to remove the proposed catalytic cysteine of the transferase resulted in loss of catalytic activity. The N-terminal transferase showed significant selectivity for the transfer of hexanoate over longer or shorter acyl chains. Thus the N-terminal domain of NSAS acts as a starter unit–ACP transacylase (SAT) component (Scheme 3).

Sequence comparison with other known NR PKS suggests that such SAT domains are common. In the few cases where the PKS sequence has been correlated with product structure, the presence of SAT domains now explains prior results from feeding experiments which suggested the use of advanced starter units. For example, it is now known that two PKS genes are involved in the biosynthesis of zearalenone **23**[32]—one of these is a HR-PKS and



Scheme 3 Observed catalytic activities of isolated NSAS domains.

probably provides a highly reduced hexaketide as a starter unit. The second zearalenone PKS is a NR-PKS possessing an N-terminal SAT domain which likely loads the hexaketide ready for three further extensions. Although feeding studies have not been carried out with advanced precursors, strong chemical evidence exists that the biosynthesis of the related dehydrocurvularin **21** does indeed involve the use of a HR-polyketide as the starter unit.[33]

Similar SAT domains are present in the citrinin **20** PKS, making it likely that citrinin is formed by a diketide starter unit, extended three times by a NR-tetraketide synthase—this validates results from isotopic feeding experiments.[23] However, most NR PKS appear to possess potential SAT domains whether they require an acetate starter unit or not, for example the PKS involved in the biosynthesis of YWA1 **27** (WAS) and tetrahydroxynaphthalene **16** (THNS). In the case of THNS from *C. lagenarium*, *in vitro* experiments have shown that the purified protein utilises malonyl CoA as the starter unit.[34] The THNS SAT domain may therefore be involved with loading of malonate to use as a starter unit. Bacterial type II PKS also use malonate as a starter unit, but here the KS_β component decarboxylates it before chain extension.[35]

2.2 NR PKS chain extension component

The extension components of NR polyketide synthases consist of KAS, AT and ACP domains. Located between the AT and ACP domain is another little-understood domain, dubbed the *product template* (PT) domain.[31] The enzymology of the chain extension components is virtually unexplored—certainly in comparison to the KS, AT and ACP components of the bacterial type I modular PKS and the type II iterative PKS. Once again the problem of investigating these systems lies in the large size of the type I proteins and the difficulty of obtaining sufficient protein for enzymology studies.

These problems have begun to be overcome, however, by the heterologous expression of individual catalytic domains. Heterologous expression offers the possibility of obtaining large quantities of protein, but ensuring correct folding of the recombinant proteins can be a hit-and-miss process. It appears that accurate determination of the domain boundaries is important for ensuring correct folding during *Escherichia coli* expression. Two approaches have been taken to determining the likely C-terminal and N-terminal boundaries. Cox and co-workers have used a simplistic model in which the peptide sequences of type II PKS components are aligned with the type I NR PKS and the N and C termini of the type II proteins are assumed to correspond with the type I domain boundaries.[30] Townsend *et al.* have described a more sophisticated model involving a bioinformatic approach which assesses possible secondary structure, hydrophobicity and family homology to determine the boundaries precisely.[36]

The Cox group has successfully expressed and studied the AT and ACP components of NSAS.[30] The AT appears to act as a relatively standard malonyl transferase, transferring malonate from CoA onto the NSAS *holo*-ACP, but not transferring other acyl groups such as acetate or hexanoate. The ACP component reacted with bacterial ACPS enzymes to generate *holo*-ACP. The *holo*-ACP was also able to receive malonyl groups *via* the bacterial *Streptomyces coelicolor* AT (also known as MCAT) as well as *via* the NSAS AT. Remarkably, the NSAS *holo*-ACP is also active in polyketide biosynthesis assays with the KAS components of the bacterial type II actinorhodin PKS. All these facts suggest that the NSAS KS-AT-ACP behave much as the type II iterative PKS systems found in bacteria.

The type II iterative PKS systems found in bacteria, however, differ from the type I systems in that they contain a chain initiation factor (CIF, also known as KS_β or chain length factor) protein which is involved in the decarboxylation of malonate to form acetate which then acts as the starter unit for polyketide biosynthesis.[35] The CIF forms an αβ dimer with the KAS protein and this dimeric species is able to determine chain length (*i.e.* the number of extensions).[37] The type I PKS of fungi lack the CIF component, but it is conceivable that the PT domain is involved in chain length determination. Sequence analysis of PT domains (450–550 residues) from a range of NR PKS genes, where the chemical products are known, suggests that this may well be the case (Fig. 3). For example, comparison of the PT domains from the *Acremonium strictum* PKS1, citrinin PKS, zearalenone PKS-B, NSAS, sterigmatocystin PKS, dothistromin PKS, THNS from *C. lagenarium* and *Wangiella dermatitidis*[38] and WAS suggests that these group into clades which correspond with chain-length. Thus the citrinin, ASPKS1 and zearalenone-B group correspond to tetraketide synthases, the NSAS, sterigmatocystin and dothistromin PKS are all octaketide synthases, the THNS are

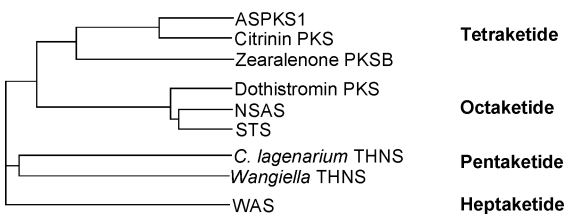


Fig. 3 Phylogenetic analysis of individual PT domains.

pentaketide synthases (but see sections 2.31 and 2.32); and the wA synthase (WAS) forms an outgroup—this is a heptaketide.[39] Very recently, Townsend has validated this preliminary analysis by showing that the expressed PT domain from NSAS is capable of controlling chain length *in vitro*.[40]

2.3 NR PKS processing component

The domains found after the chain extension components at the C-termini of fungal NR-PKS are highly varied. These include putative Claisen-cyclase–thiolesterases (CLC–TE), methyltransferases (MeT), reductases (R), and additional ACP domains. All evidence collected to date suggests that these processing components act *after* chain assembly to modify either a poly-keto or a cyclised intermediate, which is probably still bound to the ACP.

2.31 CLC–TE domains

The first fungal NR PKS gene to be cloned was *A. nidulans wA* (encoding WAS) first reported in 1992.[41,42] At that time, limited domain analysis was carried out to determine the presence of KAS, AT and ACP domians, but N- and C-terminal domains had not been recognised. The *A. nidulans pksST*[43] (encoding the sterigmatocystin NSAS) and the *A. parasiticus pksA*[44] (encoding the aflatoxin NSAS) were then reported in 1995. Leonard and Yu realised that the 3′ sequence of *pksST* encoded a protein domain with homology to known TE domains of FAS and the ACV synthetase involved in β-lactam biosynthesis. Later it was shown that *wA* and *pksA* also possess similar domains and TE domains were recognised as one of the most common processing components of NR PKS.

From the outset it was realised that the TE domain could either operate as a standard thiolesterase, or be involved in a cyclisation-release mechanism. The first experiment which investigated the role of the TE domain occurred when Ebizuka and coworkers expressed *A. nidulans wA* (encoding WAS) in *Aspergillus oryzae*.[45] In an initial experiment the expression strain produced the citreoisocoumarin **25** and the parent compound **26** indicating that WAS is a heptaketide synthase (Fig. 4). The secondary alcohol of **25** is presumably formed by an adventitious reduction *in vivo*—WAS could not catalyse this reaction as it lacks reductive domains.

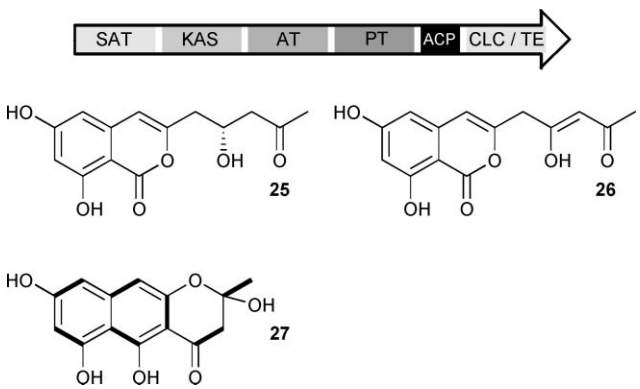
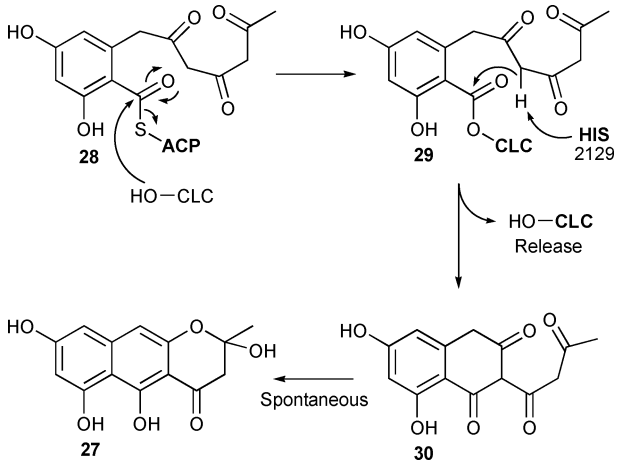


Fig. 4 Gene structure of *wA* and structures of its heptaketide chemical products.

However, it was later realised that the expression construct used in this experiment had a deletion resulting in the expressed WAS missing the final 67 amino acids of the C-terminal TE domain. When the complete *wA* gene was expressed, however, the heptaketide naphthopyrone YWA1 **27** was produced.[46] A series of experiments involving step-wise shortening of the C-terminus of WAS showed that deletion of as few as 32 amino acids resulted in production of the citreoisocoumarin **26**. Site directed mutagenesis of a conserved serine and histidine in the C-terminal domain also resulted in a switch from naphthopyrone production to citreoisocoumarin production.[47]

Isotopic feeding experiments using ^{13}C-labelled acetate showed the folding pattern shown in Fig. 4 for the naphthopyrone **27**.[47] Thus both the citreoisocoumarin **26** and the naphthopyrone **27** must result from the cyclisation of the common intermediate **28** (Scheme 4). This observation suggests that the WAS chain extension component produces a heptaketide and catalyses the cyclisation and aromatisation of the first ring.

The C-terminal domain must therefore catalyse a second (Claisen) cyclisation reaction to form the observed naphthopyrone **27**. The involvement of conserved serine and histidine residues has led to the suggested mechanism shown in Scheme 4. Thus the TE domain has been renamed as CLC (Claisen cyclase). These domains also occur in the known NSAS and THNS proteins where the same chemistry must occur to provide the observed products.



Scheme 4 Proposed mechanism of the WAS CLC domain.

2.32 Chain shortening reactions

A further role for C-terminal CLC domains has recently been discovered. The *C. lagenarium PKS1* gene is involved in the biosynthesis of the pentaketide tetrahydroxynaphthalene **16**.[48] Ebizuka and Watanabe expressed *C. lagenarium PKS1* in the heterologous fungal host *A. oryzae* and, as expected, observed the production of the pentaketide **16**. However, two additional compounds were produced: the tetraketide orsellinic acid **15** and the pentaketide α-acetyl orsellinic acid **31**.[49] Similar results were obtained *in vitro* using purified THNS and malonyl CoA as the substrate.[34] In this case the system produced a hexaketide **32** (15%), pentaketide THN **16** (50%), pentaketide α-acetyl orsellinic acid **31** (25%) and the tetraketide orsellinic acid **15** (10%) (Fig. 5).[50]

Ebizuka and Watanabe then generated two mutants of the *C. lagenarium PKS1* in which the CLC domain was completely removed or the active site serine of the CLC was mutated to

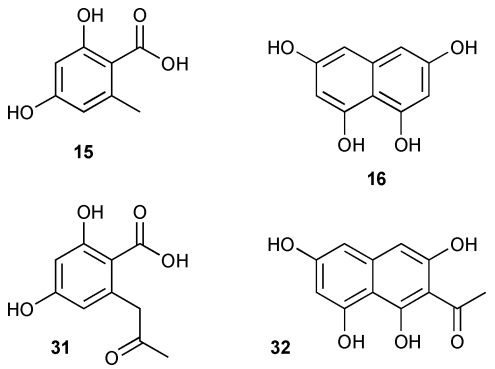
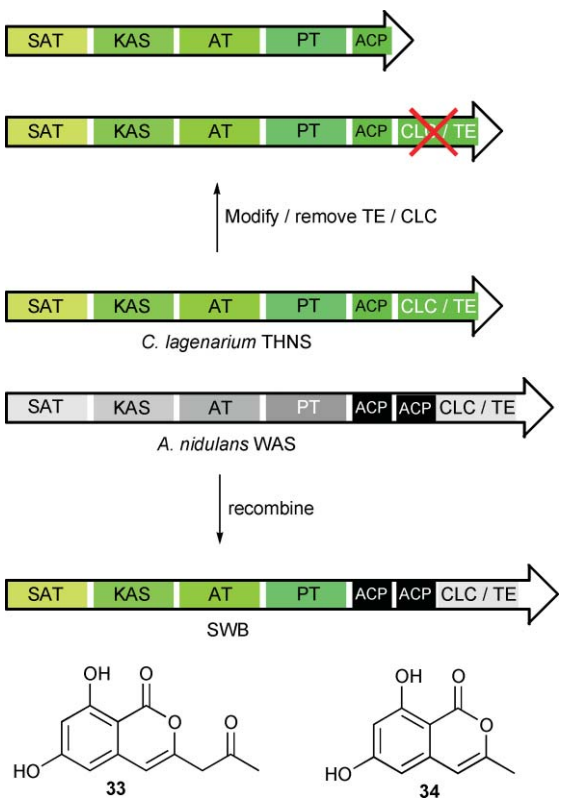


Fig. 5 Tetraketide, pentaketide and hexaketide compounds produced *in vitro* by *C. lagenarium* THNS.

alanine (Scheme 5).[50] In both cases, expression of the modified *C. lagenarium PKS1* genes resulted in the production of >95% hexaketide **33** and *ca.* 5% of pentaketide **34** indicating that the CLC domain must also be involved in chain-length determination—*i.e.* the *prevention of chain extension* to form a hexaketide. This mechanism must involve the active site serine of the CLC but it is not yet clear how the specificity arises.

Watanabe and Ebizuka further investigated the relationship between chain-length and cyclisation by synthesising a fused PKS consisting of the N-terminus of *C. lagenarium* THNS and the C-terminus of *A. nidulans* WAS (Scheme 5). The fusion site was chosen between the AT and ACP domains. The resulting chimeric PKS, known as SWB, produced the hexaketides **32** and **33** as well as the pentaketides **31** and **34**.[51]



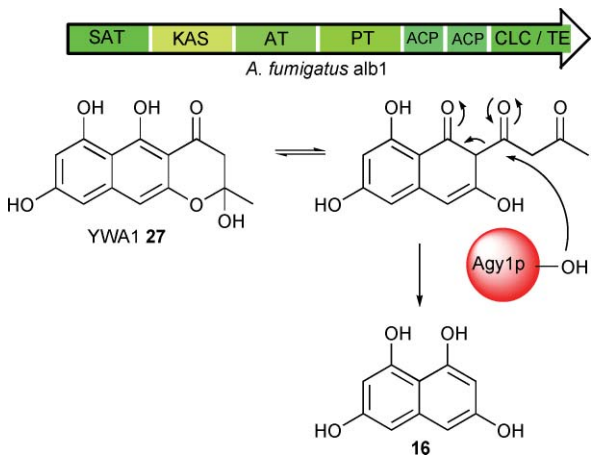
Scheme 5 Modifications to *C. lagenarium* THNS.

This result shows that the fundamental hexaketide selectivity of *C. lagenarium* THNS is preserved and that the WAS CLC domain is capable of cyclising both hexaketides and heptaketides, but the absence of THN as a product shows that the WAS CLC cannot intercept and cyclise a pentaketide intermediate. When the CLC domain was deleted from SWB, only hexaketides were produced showing that the WAS CLC is also able to affect chain length.[50]

The original hypothesis that CLC domains are required for chain release must now be modified. Although it appears that CLC domains can clearly affect the mode and timing of chain release, they are not necessary as indicated by the above experiments where a functional NR PKS can be made lacking a CLC. It is not clear which component of the NR PKS is involved in chain release when the CLC is absent—however, the situation closely mirrors the situation in the bacterial type II PKS where there is no obvious mechanism to release the chain from the ACP when it has reached its final length.[52]

A related process of chain shortening has been observed during the biosynthesis of the pentaketide THN **16** in *Aspergillus fumigatus*. In this organism, the NR PKS gene *alb1* encodes a polyketide synthase (alb1p) which is closely related to the heptaketide synthase WAS. When *alb1* was expressed in *A. oryzae* the heptaketide YWA1 **27** was produced, together with a minor component which proved to be a dehydration product of **27**.[53] This result was initially puzzling because *alb1* had previously been shown to be involved in the biosynthesis of THN **16**.

The conundrum was resolved by analysis of other genes in the *alb1* cluster in *A. fumigatus*. Knockout of the gene *ayg1* (encoding the protein Ayg1p) led to loss of the production of THN **16** and its derivatives, but production of YWA1 **27** was observed. It was therefore concluded that Ayg1p must convert **27** to **16**, presumably by a retro Claisen mechanism (Scheme 6).[54] This was demonstrated by treating YWA1 **27** with partially purified Ayg1p resulting in the formation of **16** *in vitro*.



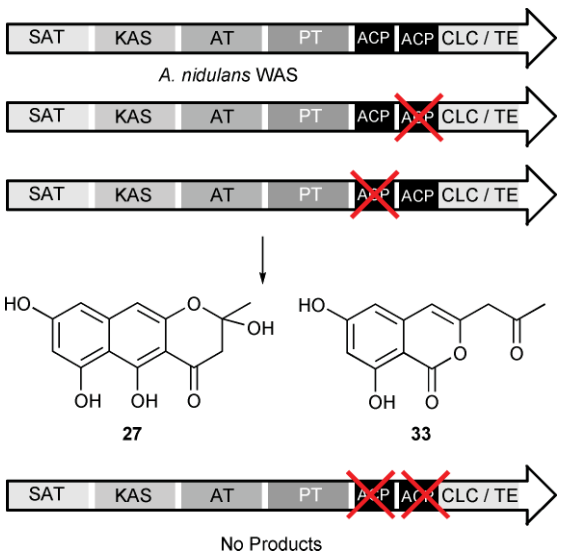
Scheme 6 Conversion of the tetraketide YWA1 **27** to THN **16**.

2.33 ACP domains

Additional ACP domains are a common feature of the processing components of fungal NR PKS. For example the *A. nidulans* WAS and NSAS, and the *C. lagenarium* THNS all possess an additional ACP domain between the ACP of the chain extension component

and the C-terminal CLC–TE domain. In the case of the NSAS involved in dothistromin biosynthesis in *Dothistroma septosporum* there are two additional ACP domains.[55] It is clear that these additional ACP domains are not required because the norsolorinic acid synthase from *A. parasiticus* requires only a single ACP.

In the case of the WAS from *A. nidulans* the role of the two ACP domains has been studied by Ebizuka and coworkers.[47] For an ACP to be active during polyketide biosynthesis it must have been appended with PP at its conserved serine. The *A. nidulans wA* gene encoding WAS was expressed in the heterologous fungal host *Aspergillus oryzae*. WAS was then modified to replace the two ACP serines with alanines (Scheme 7). When either ACP was inactivated singly the synthase was still capable of producing the expected naphthopyrone **27**, but inactivation of both ACP domains resulted in total loss of polyketide production. This experiment indicates that both ACP domains were capable of interacting with the chain extension components and the processing component.



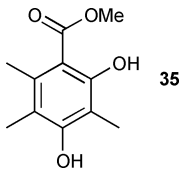
Scheme 7 Analysis of ACP function in WAS.

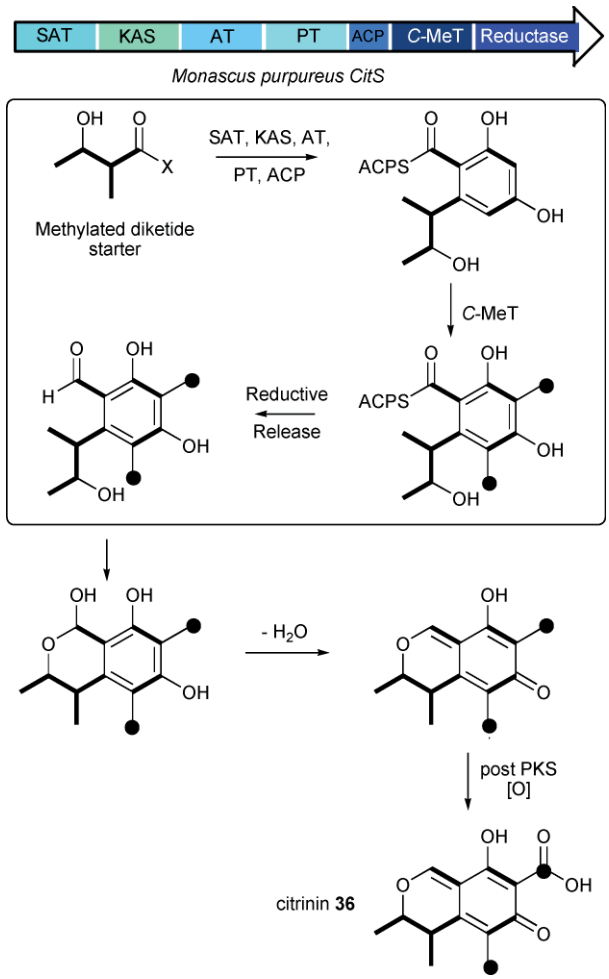
Ebizuka and coworkers have also fused the ACP didomain and CLC from WAS onto the SAT, KAS, AT and PT domains of THNS. This chimeric PKS was active indicating that the ACPs are not 'matched' to the other components—they appear to function independently of any of the other programme elements (Scheme 5).[51]

2.34 *C*-MeT domains

Few NR PKS are known to possess *C*-methylation domains, although a number of known fungal non-reduced polyketides are *C*-methylated such as dimethylorsellinic acid methyl ester **35**. A small group of NR PKS have been identified by Kroken *et al.*[19] which feature a *C*-MeT domain located after the ACP. The only correlation between a gene sequence and a compound comes from the case of citrinin **36** where the PKS involved in citrinin biosynthesis in *Monascus purpureus* has recently been described.[56]

Here the *C*-MeT domain must also be programmed as it acts twice during polyketide biosynthesis when a probable C_5 starter unit is extended.[23] It is not yet clear whether the *C*-MeT domain acts during extension, after chain extension but before aromatisation, or after aromatisation. 1,3-Dihydroxyaromatics are known to tautomerise easily to keto forms[57] and it is conceivable that this could act as the nucleophile for the reaction with SAM (Scheme 8). This would maintain the pattern observed for the other processing component domains where these act after the extension components are finished. Other types of programming remain to be investigated—for example, dimethyl orsellinic acid methyl ester **35** is a known fungal metabolite[58] and this could conceivably be synthesised by a very similar PKS with a differently programmed *C*-MeT domain.





Scheme 8 Possible biosynthesis of citrinin **36**.

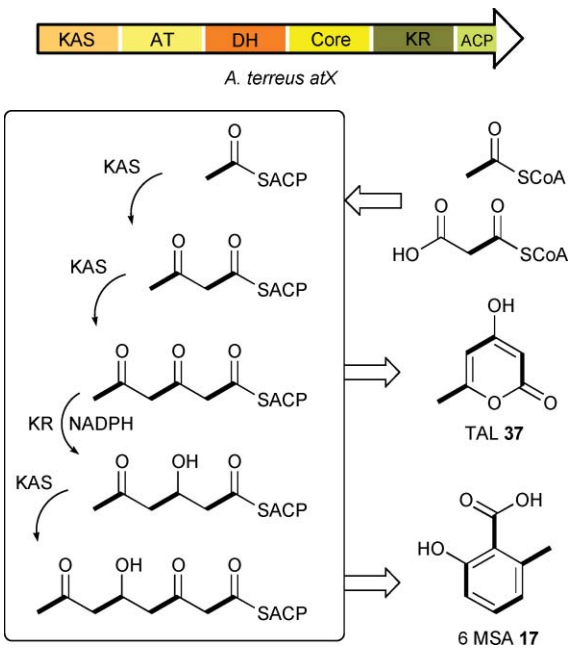
2.35 R domains

Once again, reductases are currently rare as part of the processing component of NR PKS. Although not described in the literature, sequence analysis of the *M. purpureus pksCT* sequence[56] shows that it possesses a C-terminal thiolester reductase domain.[39]

Similar domains are known from NRPS systems where reductase domains are sometimes used as chain release mechanisms, releasing an aldehyde or primary alcohol. In the case of citrinin biosynthesis, the reductive release mechanism makes good sense as this provides the product with C-1 at the correct oxidation state (Scheme 8).

3.0 Fungal PR PKS

Much less is known about the enzymology of the PR PKS. The domain structure is much closer to mammalian FAS, with an N-terminal KAS followed by AT, and DH domains (Scheme 9). A so-called 'core' domain follows the DH, and this is followed by a KR and the PKS terminates with an ACP domain. The domain structure differs considerably from the NR PKS—there is no SAT domain or PT domain, and the PKS terminates after the ACP and appears not to require a CLC–TE domain.



Scheme 9 Domain architecture of MSAS encoded by *A. terreus atX*, *G. lozoyensis pks2* and *P. patulum MSAS* and chemical steps during 6MSA biosynthesis.

Once again there is no obvious catalytic machinery for off-loading the product. The DNA sequence of the KAS domain is distinguishable from NR-PKS and FAS KAS domains using selective PCR primers and DNA probes.[14,15] Overall, however, catalytic domains closely match those of the mamalian FAS, although the 'core' domain is different.[59] Remarkably, the fungal MSAS genes are closely related to recently discovered *bacterial* genes for the synthesis of the non-reduced tetraketide orsellinic acid **15**.[60] For example, the iterative type I PKS from *Micromonospora echinospora*, calO5, appears to be homologous to the fungal PR PKS proteins apart from lacking a *ca.* 450 amino acid region encompassing the KR domain.

Although a number of PR PKS genes are known from genome sequencing projects, only three genes have been matched to chemical products—in all cases, the tetraketide 6-methylsalicylic acid **17**. The first 6-methylsalicylic acid synthase (MSAS) to be discovered was from *Penicillium patulum*, encoded by *MSAS*.[59] The Ebizuka group have worked with the *atX* gene from *Aspergillus terreus*,[61] and most recently Tkacz and coworkers have described an MSAS gene (*pks2*) isolated from *Glarea lozoyensis*.[62]

MSAS from *P. patulum* is one of the smallest type I PKS at 191 KDa. It proved relatively easy to isolate, and many of the earliest *in vitro* studies of the enzymology of any PKS were carried out with this enzyme.[63,64] MSAS requires acetate as a starter unit—unlike THNS and the type II actinorhodin PKS, malonate cannot be used as a starter. The AT domain loads malonate from CoA to use as an extender. The PKS is evidently programmed—acetate must be extended three times and the KR must only act once, after the second extension. When NADPH is unavailable, the KR reaction cannot occur. Under these circumstances, a tetraketide is not produced and triketide lactone (TAL) **37** is produced instead (Scheme 9). This reveals that chain extension and reduction are linked, and indicates that the KR must act during chain extension. This is different to the C-terminal R domain of citrinin synthase which must act after extension and cyclisation. Mammalian FAS also produces TAL **37** under the same circumstances. The similarities between MSAS and FAS were further exemplified by Spencer and Jordan who showed that the KS and ACP components of different peptide chains in dimeric MSAS species could be cross-linked using thiol specific reagents.[65]

Chain length determination in MSAS appears to use a 'counting' mechanism. When acetyl CoA is unavailable as a starter unit, other acyl CoAs can substitute and be extended. In the case of the type II act PKS, experiments of this type show that the length of the starter unit is important so that C_{16} chains are preferred products despite the length of the starter unit.[52] In the case of MSAS denied NADPH and acetyl CoA, but incubated with malonyl CoA extender units and various acyl CoA starter units, two extensions always occurred to produce substituted TALs.[66]

The expression of MSAS in heterologous hosts has allowed more detailed experiments to be performed. Hopwood, Khosla and coworkers first described the expression of *P. patulum* MSAS in the Gram positive bacterium *Streptomyces coelicolor*.[67] Ebizuka, Fujii and coworkers expressed *A. terreus atX* in *Aspergillus nidulans*.[61] However, the most useful expression system to date has proven to be the yeast *Saccharomyces cerevisiae*.[68,69]

Fujii and coworkers developed a *S. cerevisiae* expression system in which two independent copies of *atX* can be produced simultaneously. This allowed them to mutate each copy of *atX* in complementary ways (Fig. 6). Initial experiments determined how much of the N and C termini of MSAS could be deleted without affecting 6-MSA production. It was found that at the N-terminus, a deletion mutant lacking 44 amino acids (*atXN1*) was still active, but that a deletion of 81 amino acids (*atXN2*) prevented 6-MSA formation by removal of key KAS residues. At the C terminus, deletion of as few as 9 amino acids (*atXC1*) rendered the MSAS inactive by removal of key ACP residues.

Both *P. patulum* and *A. terreus* MSAS form homotetramers. Fujii and coworkers exploited this fact in the reconstitution of functional MSAS complexes formed from combinations of the deletion mutants. For example, co-expression of *atXN2* with *atXC1*, both of which are inactive in isolation, led to the reconstitution of an active MSAS complex. This shows that the KAS of one peptide chain must interact with the ACP of another peptide chain. The number of deletion constructs was

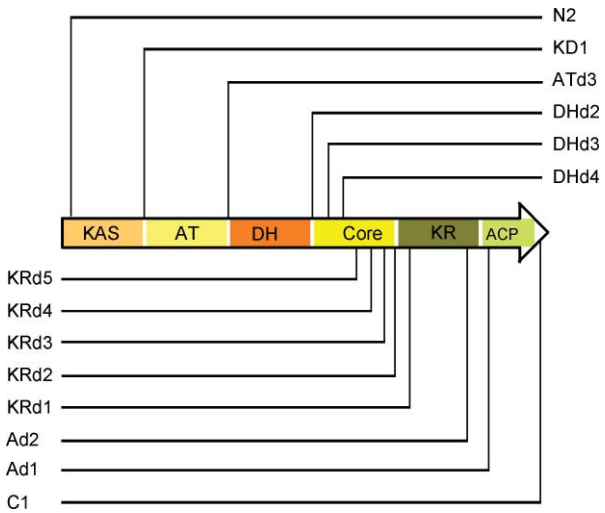
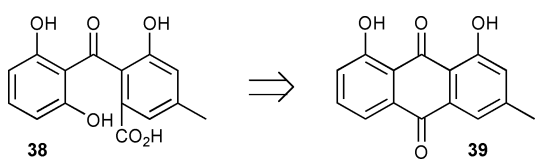


Fig. 6 N-terminal and C-terminal deletions of *atX* expressed in *S. cerevisiae*.

then increased to examine a total of 6 different length deletions from the N-terminus and 8 different length deletions from the C terminus (Fig. 6). Remarkably, almost all combinations of the truncated MSAS proteins proved functional, showing that very severe truncations can be complemented if the activity is present on another peptide chain. However, a short region of the core domain was identified, the presence of which proved essential for successful complementation. It is hypothesised that this region of 122 amino acids probably forms a motif required for subunit–subunit interaction. This core sequence is present in other fungal PR PKS, and in the bacterial PKS such as CalO5. Mammalian FAS also has a central core region which has been shown to mediate assembly of the synthase[70]—however, there are no sequence similarities between the FAS and MSAS core regions.

Tkacz and coworkers expressed the *G. lozoyensis pks2* gene in *A. nidulans*,[62] replicating the earlier work of Fujii and coworkers with *atX*.[61] Tkacz showed that the expression of *pks2* led to the formation of 6-MSA **17**, but a minor metabolite, the octaketide benzophenone **39**, was also produced (Scheme 10). It is not yet clear how this compound arises—Tkacz and coworkers suggested several possibilities such as induction of an as-yet unidentified *A. nidulans* octaketide synthase gene by 6-MSA (although this didn't happen in Fujii's experiments with *atX*), or possible positional effects of the insertion of the *trpC* promoter. However, a very intriguing possibility is that an N-terminal intron in *pks2* can be processed in two different ways (Tkacz showed that 42% of *pks2* transcripts have an alternative mode of intron removal) leading to two different MSAS N-terminal sequences. It is conceivable that one of these could be programmed differently to produce an octaketide such as **39**. Intriguingly, **39** is also partially reduced.



Scheme 10 Production of benzophenone.

4.0 Fungal HR PKS

The final class of fungal PKS are those that produce complex, highly reduced compounds such as lovastatin **7a**, T-toxin **19**, fumonisin B1 **40** and squalestatin **41**.

In all cases found to date, these fungal compounds are produced by iterative type I PKS. These PKS have an N-terminal KAS domain, followed by AT and DH domains. In many cases, the DH is followed by a *C*-MeT domain. Some HR PKS possess an ER domain, but those that don't have a roughly equivalent length of sequence with no known function. This is followed by a KR domain and the PKS often terminates with an ACP. There appears to be no domain similar to the PT domain of the NR PKS or the core domain of the PR PKS, and there is no N-terminal SAT domain as found in the NR PKS (Fig. 7).

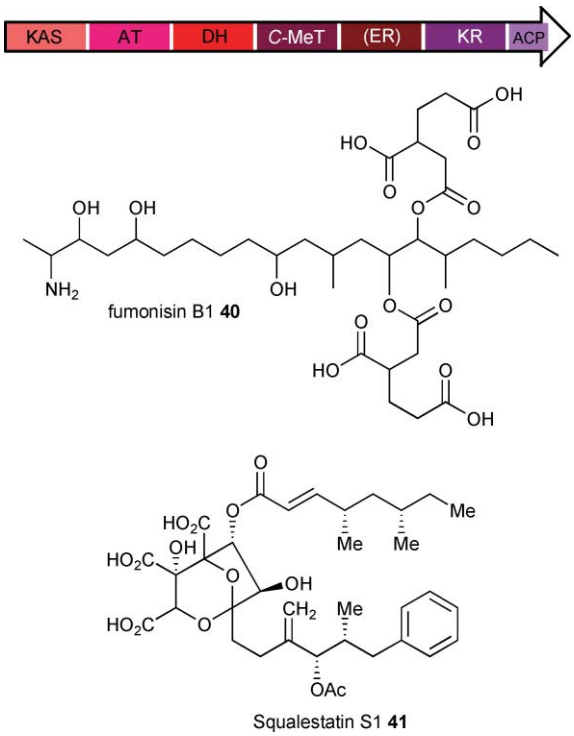


Fig. 7 General domain architecture of HR PKS and compounds produced by HR PKS.

As with the NR and PR PKS, many HR PKS genes are known from the numerous fungal genome sequences, but few gene sequences have been linked to the production of known compounds. However, of the few cases where both gene and chemical product are known, some progress has been made in understanding function and programming.

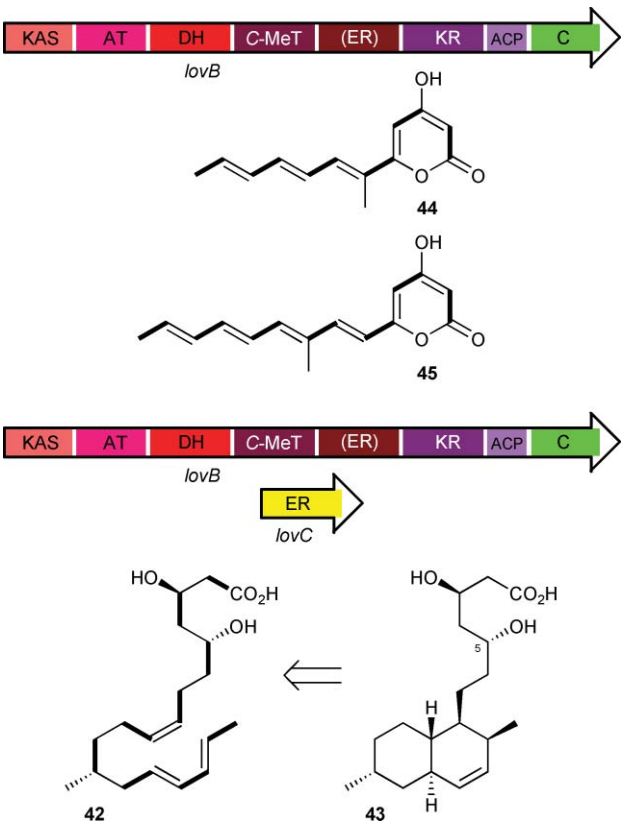
4.1 The lovastatin polyketide synthases

Lovastatin **7a** (also known as mevinolin) is produced by *Aspergillus terreus*. The related compound compactin (mevastatin) **7b**, produced by *Penicillium citrinum* is identical to lovastatin apart from the C-12 methyl group which is absent in compactin. Both compounds are potent inhibitors of mammalian hydroxymethylglutaryl (HMG) CoA reductase and hence find a valuable role as pharmaceuticals used to lower serum cholesterol. The biosynthesis

of these compounds has been studied intensely—isotopic feeding experiments have shown that two polyketide chains are required: a nonaketide and a methylated diketide. The requirement for two polyketide synthases is evident in the gene clusters associated with **7a** and **7b** biosynthesis where two PKS genes are found.

In the case of lovastatin **7a** biosynthesis, the two PKS genes are *lovB* and *lovF*,[71,72] encoding lovastatin nonaketide synthase (LNKS) and lovastatin diketide synthase (LDKS). Two homologous genes are involved during compactin **7b** biosynthesis: *mlcA* (encoding CNKS) and *mlcB* (encoding CDKS).[73]

Isotopic feeding experiments had suggested that LNKS should synthesise a fully elaborated nonaketide such as **42** which could undergo a biological Diels–Alder reaction to form dihydromonacolin L **43**—the observed first PKS-free intermediate. In order to confirm this, *lovB* was expressed in the heterologous fungal host *A. nidulans*.[72] Surprisingly, the polyunsaturated compounds **44** and **45** were isolated. These pyrones are related to the expected nonaketide—for example, the methylation has occurred at the correct position, but it is evident that reductions—specifically enoyl reductions, and later keto reductions—have not occurred correctly and that chain extension has terminated prematurely (Scheme 11).



Scheme 11 Expression of *lovB* and *lovC* in *A. nidulans*.

Close inspection of the sequence of LNKS suggests that the NADPH binding pocket of the ER domain might be disfunctional, explaining the lack of observed ER activity. An enoyl reductase encoding gene, *lovC*, occurs downstream from *lovB* and Vederas and coworkers showed that in coexpression experiments, the lovC protein could complement the inactive ER domain and the expected dihydromonacolin L **43** was produced (Scheme 11). It is evident that the lovC protein must interact with LNKS intimately, drastically affecting programming by ensuring enoyl reduction at the correct positions and allowing complete chain extension. It now appears that lovC type proteins are a common feature of HR PKS systems in fungi, although by no means ubiquitous.

The fact that 'defective' ER domains are not removed from the HR PKS where lovC-type proteins are used to achieve programmed enoyl reductions may be significant. It is evident that the lovC protein must dock with the PKS in order to effect reduction, and the rump ER in the PKS sequence may act as a docking anchor—however, protein experiments have not yet been carried out to prove this hypothesis. Similar interactions between type I PKS and mono-functional proteins also occur in bacterial systems, for example, between multimodular and monofunctional proteins involved in mupirocin biosynthesis in *Pseudomonas fluorescens*.[74]

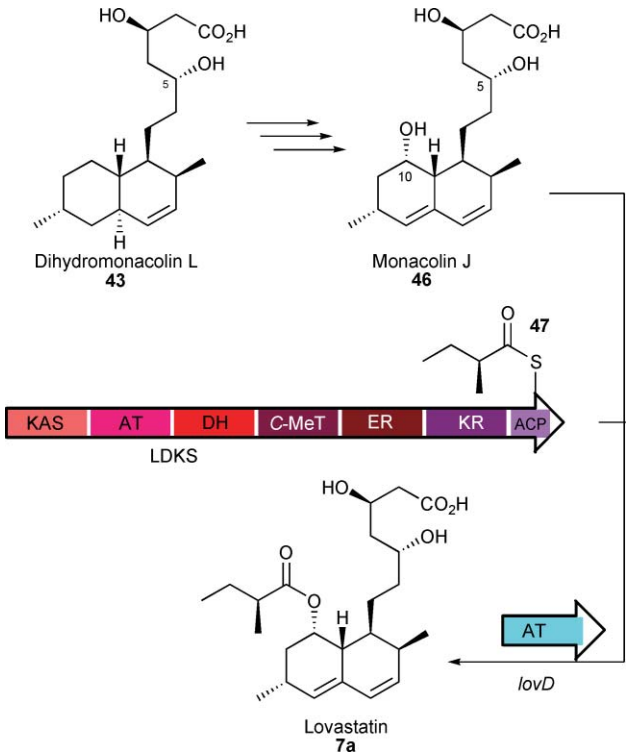
In the bacterial modular PKS systems, there is usually a C-terminal TE domain involved in off-loading the product—either to another PKS or to solution.[9] LNKS appears to possess part of an NRPS condensation (C) domain immediately downstream of the ACP, and this domain has been proposed to be involved in product release, presumably by either activating water as a nucleophile or the C-5 hydoxyl (rather than the nitrogen of an aminothiolester as activated by most NRPS C domains).

LDKS is closely related to LNKS, but its ER domain appears to be intact. LDKS is unusual among fungal PKS because it is not iterative—a single round of extension and processing affords the diketide **47** (Scheme 12). In this respect, LDKS closely resembles a single module of a bacterial modular PKS.[9] LDKS lacks an obvious product release domain, ending immediately after the ACP. A specialised acyl transferase, encoded by *lovD*, transfers **47** from LDKS onto the C-10 hydroxyl of monacolin J **46**.[72] Tang and coworkers expressed *lovD* in *E. coli* and showed that the recombinant protein will also transfer various acyl groups from CoA to the C-10 hydroxyl of **46** (Scheme 12).[75]

4.2 The squalestatin S1 polyketide synthases

Squalestatin S1 **41** is a potent inhibitor of mammalian squalene synthase which also possesses unprecedented curative properties towards prion infected neurons.[76] It is produced by *Phoma* species, and like lovastatin, consists of two polyketide chains: a main chain hexaketide and a sidechain tetraketide. Like lovastatin, both chains are methylated, but unusually for a fungal HR polyketide, the main chain is formed from a non-acetate starter unit—benzoate is incorporated at this position.

Cox and coworkers cloned a HR PKS gene from *Phoma* sp., *PhPKS1*. This was expressed in the heterologous fungal host *Aspergillus oryzae* which produced the tetraketide **48**. *PhPKS1* thus encodes the squalestatin tetraketide synthase (SQTKS).[77] SQTKS is highly homologous to LDKS, lacking the C-terminal NRPS condensation domain of LNKS. Like LDKS, SQTKS possesses a functional ER domain, but SQTKS carries out three extensions. Like LDKS, all modification reactions occur after the first extension. The stereochemistry of the branching methyl group is the same in each case. Two further extensions occur—all modifying reactions occur again after the first of these, but neither ER nor CMeT are used after the final extension (Fig. 8). Sequence comparisons of LDKS and LNKS, however, reveal few clues as



Scheme 12 The *lovD* AT transfers 2-methylbutyrate **47** from LDKS to monacolin J **46**.

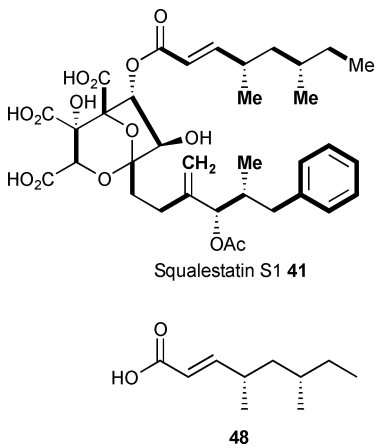


Fig. 8 Squalestatin S1 **41** and the squalestatin tetraketide **48**.

to the structural basis for the difference in programming between the two.

4.3 The fumonisin and T-toxin polyketide synthases

Fumonisin B1 **40** and T-toxin **19** are among the longest fungal polyketides. Fumonisin B1, a nonaketide, is produced by *Gibberella fujikuroi* (anamorph *Fusarium moniliforme*, syn. *Fusarium verticilloides*) and is a mycotoxin responsible for several human and animal diseases. T-toxin, a C_{41} linear polyketide, is produced by race-T of the maize pathogen *Cochliobolus heterostrophus*. Fumonisin B1 **40** is methylated and esterified to two unusual C_7 tricarboxylic acids, while T-toxin **19** is unbranched and its hydroxyls are unmodified.

T-toxin was among the first fungal compounds to be linked to a particular PKS gene.[78] More recently it has been shown that two PKS genes (*pks1* and *pks2* encoding TTS1 and TTS2 respectively) are involved in its biosynthesis—unusually, *pks1* and *pks2* are not clustered on the *C. heterostropus genome*.[79] Domain analysis of the two PKS genes indicates that both are typical HR PKS and that neither possesses a SAT domain—they differ in this respect from the only other fungal PKSs known to use multiple synthases, NSAS and ZS. TTS1 (2528 aa) appears to posses a CMeT domain, despite T-toxin not requiring C-methylation, while TTS2 (2144 aa) does not appear to possess a CMeT domain.

The fumonisin biosynthetic genes of *G. fujikuroi* are clustered. The PKS fumonisin synthase (FUMS) is encoded by *fum1* (previously known as *fum5*) which appears to be a fairly typical PKS of the HR type, with KS, AT, DH, CMeT, ER, KR and ACP domains.[80] It appears that unlike the case of LNKS where the ER is inactive, the ER of FUMS is probably fully functional—consistent with this observation is the lack of a *lovC* homologue in the gene cluster. It appears that FUMS operates a fairly simple programme—chain length is controlled to 18 carbons and, apart from the CMeT, all activities operate in all cycles. Methylation is the only programmed event, occurring during the second and fourth extension cycles only.[81]

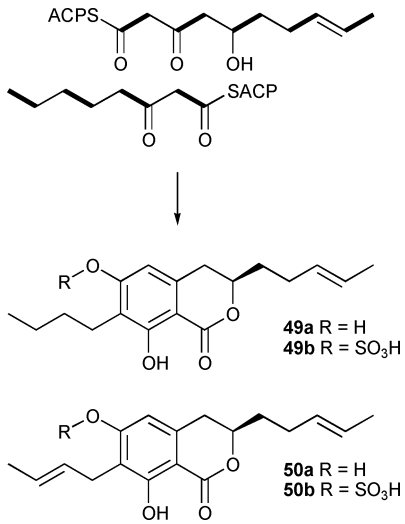
The function of the TTS and FUMS genes has been proven by knockout experiments, but none of the genes has yet been reported as expressed heterologously. However, FUMS has been the target for the first HR-PKS directed domain swap experiments. In initial work, Du and coworkers described the development of a method to exchange regions of *fum1* with homologous regions of *C. heterostropus pks1*. In effect, the KS domain from TTS1 was inserted into FUMS. This domain swap appeared to have no effect on either levels of production of fumonisin B1 **40**, or on programmed events during biosynthesis.[82] Replacement of the whole of FUMS by TTS1 resulted in no production—however, as it is now known that two PKS are required for T-toxin production, it is unsurprising that this experiment produced no new compounds.

Further work, however, in which the FUMS KS was replaced with the LDKS KS did result in significant reprogramming of the PKS. The new chimeric PKS produced two unusual dihydroisocoumarins **49** and **50** (Scheme 13).[83] These compounds appear to be the result of the condensation of two separate polyketide chains—in each case a tetraketide with a pentaketide. It is not clear what underlies these changes in programming, or the unusual dimerisation, but it does suggest that the LDKS and FUMS KS domains are, at least partially, involved in the molecular interactions which help specify chain length and use of reductive domains.

4.4 HR PKS from *Alternaria solani*

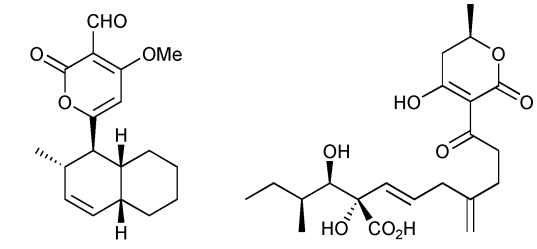
Alternaria solani is a plant pathogen and the causative agent of early blight in solanum species. It produces numerous polyketides such as solanopyrone A **51** and alternaric acid **52**, and is thus an ideal target species for speculative PKS gene-fishing expeditions.

Fujii and Ebizuka have conducted just such investigations, using PCR primers based on conserved PKS sequences as probes with genomic DNA libraries. An early investigation yielded two hits—one, a HR PKS gene named *alt5*, and another, a NR PKS named *pksA*.[84] The *alt5* gene encodes a typical HR PKS, known as

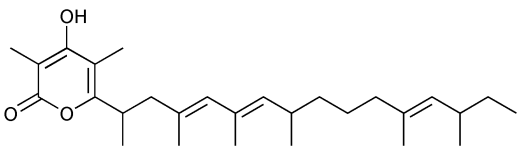


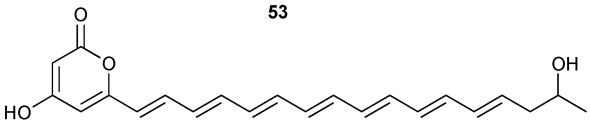
Scheme 13 New compounds produced by *F. verticilloides* after the KAS domain of FUMS was replaced with the LDKS KAS.

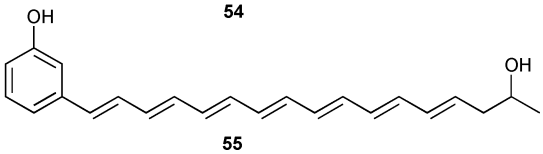
PKSN, with the usual array of catalytic domains. Inspection of the ER sequence suggested that it should be functional like those from LDKS and SQTKS. Expression of *alt5* showed this to be correct—a single compound was synthesised in good yield (*ca.* 15 mg l^{-1}) which proved to be the octamethylated decaketide pyrone **53**, named alternapyrone. This compound is the most complex polyketide yet reported to be produced by an iterative PKS, showing programmed chain length control, keto-reduction, methylation and enoyl reduction.


Solanapyrone A **51**

Alternaric acid **52**


53


54

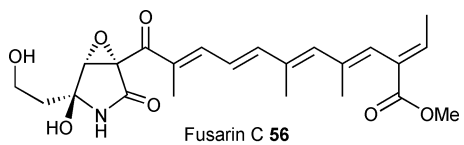

55

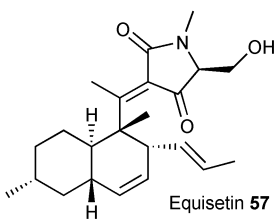
Further gene probing of *A. solani* afforded two more HR PKS, PKSF and PKSK encoded by *pksF* and *pksK* respectively. Unlike PKSN, these synthases did not seem to possess *C*MeT domains, but were otherwise similar to other HR PKS. PKSK proved to be inactive when expressed in *A. oryzae*, but PKSF did produce the new compounds aslanipyrone **54** and aslaniol **55**—albeit in low yields, together with nine other related minor compounds.[85] These polyunsaturated compounds must arise because of the inactivity of the ER domain, and detailed sequence analysis suggested that the NADPH binding motif of the ER domain may be defective in PKSF. It is not yet known whether the PKSF cluster possesses a *lovC* homologue as the lovastatin and tenellin clusters do, but coexpression of PKSF with such a gene could form new polyketides.

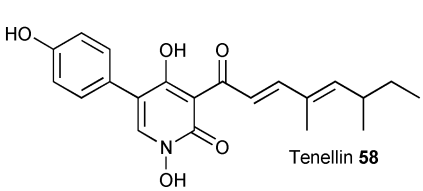
4.5 HR PKS–NRPS

Fungi produce a wide range of bioactive compounds derived from polyketides fused to amino acids. Examples include fusarin C **56**, equisetin **57** and tenellin **58**. Fusarin C **56** consists of a tetramethylated heptaketide fused to homoserine[86] and is produced by strains of the plant pathogens *Fusarium moniliforme* and *Fusarium venenatum*. Genomic DNA libraries from these organisms were used to isolate a gene cluster centred around a 12Kb ORF encoding a HR PKS fused to a non-ribosomal peptide synthetase (NRPS) module.[87] The PKS region is homologous to LNKS: KAS, AT and DH domains are followed by CMeT, a defective ER, KR and ACP domains. Like LNKS, the ACP is upstream of an NRPS condensation (C) domain, but in this case, the NRPS module is complete, featuring downstream adenylation (A), thiolation (T) and C-terminal thiolester reductase (R) domains.

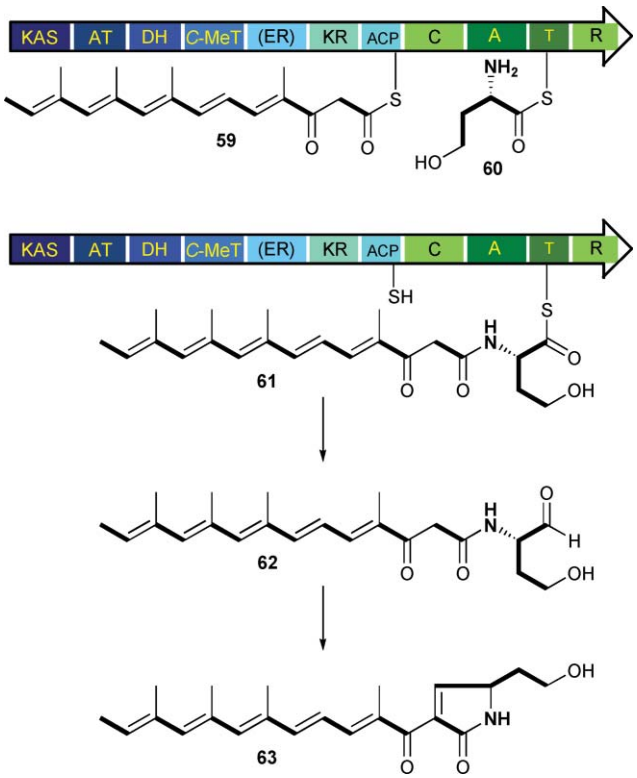
Directed knockout of the PKS-NRPS gene proved it to be involved in the biosynthesis of **56** and it was thus named fusarin synthetase (FUSS). The disfunctional ER domain and the fact that there appears to be no *lovC* homologue in the cluster, is consistent with the polyunsaturated nature of the polyketide moiety.


Fusarin C **56**


Equisetin **57**


Tenellin **58**

It is probable that FUSS assembles a tetramethylated heptaketide **59** attached to the ACP (Scheme 14)—the structure of which is remarkably similar to the heptaketide pyrone **45** produced by LNKS in the absence of the lovC protein. In parallel, the A domain of the NRPS module appears to select, activate and attach homoserine **60** to the thiolation domain.[86] The C domain



Scheme 14 Involvement of FUSS during the biosynthesis of pre-fusarin **63**.

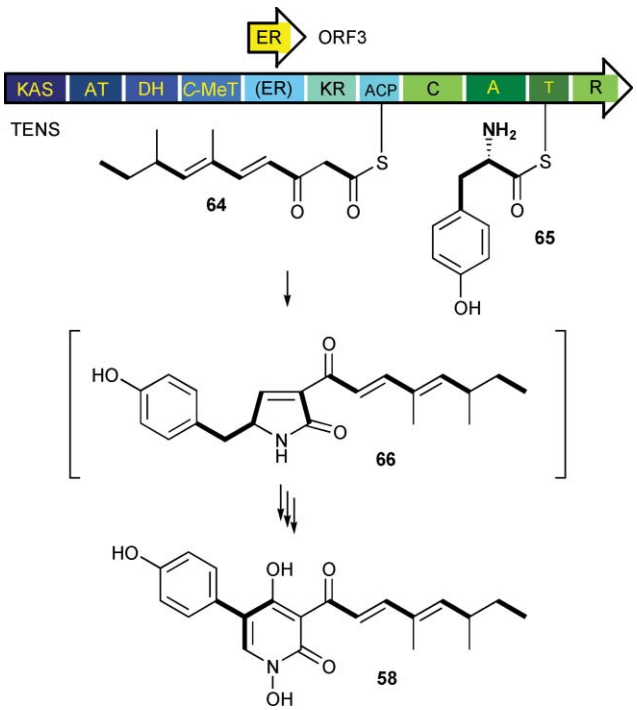
then utilises the amine of homoserine to form an amide with the ACP-bound polyketide, forming a covalently bound intermediate peptide **61**. A similar reaction is probably catalysed by the C domain of LNKS which may use a water nucleophile, or the C-5 hydroxyl of the putative enzyme-bound lovaststin precursor **42** (Scheme 11), to release dihydromonacolin L **43**.

The final reaction catalysed by FUSS may be reductive release of the thiolester, forming peptide aldehyde **62**. A homologous thiolester reductase domain forms the C-terminal feature of fungal α-aminoadipate reductase and this catalyses a single reduction of a thiolester to form an aldehyde.[88] Finally, Knoevenagel cyclisation would give the putative prefusarin **63**. Other genes in the FUSS cluster are presumably responsible for the required further transformation of **63** to fusarin C **56** itself: epoxidation, oxidation of a pendant methyl to a carboxylate and esterification, and hydroxylation α to nitrogen.

A highly homologous PKS-NRPS gene has been shown to be involved in the biosynthesis of equisetin **57** in *Fusarium heterosporum*.[89] EQS posseses the same catalytic domains as FUSS, but examination of the structure of **57** indicates that the pyrollidinone carbon derived from the carboxylate of the amino acid (serine in this case) is not reduced, indicating either a reoxidation mechanism, or the fact that the R domain does not produce an aldehyde intermediate in this case.

PKS-NRPS are also involved in the biosynthesis of as yet unknown compounds in the plant pathogen *Magnaporthe grisea*.[90] The *Ace1* PKS-NRPS gene forms part of a biosynthetic gene cluster which is fleetingly expressed only during apressorium formation during the initial penetration event of pathogenesis. Other genes in the *Ace1* cluster encode cytochrome P450 oxidases as well as lovC-type ER homologues. The compound produced by this cluster can be detected by specific genotypes of rice and, if detected, this produces a signalling cascade resulting in the effective mounting of resistance to further fungal penetration. Thus when *Ace1* and its neighbouring genes are expressed, *M. grisea* is rendered avirulent.[90]

PKS-NRPS proteins are also involved in the biosynthesis of fungal 2-pyridones. For example the yellow pigment tenellin **58** from the insect pathogen *Beauveria bassiana* is produced by a PKS-NRPS, named TENS, homologous to FUSS.[91] Like FUSS, the ER domain of TENS appears to be inactive. However, the TENS gene cluster also contains a *lovC* homologue, ORF3, and this presumably provides the programmed ER event required during the first round of polyketide biosynthesis. It appears that TENS selects tyrosine **65** and fuses it to a doubly methylated pentaketide **64** to form putative pretenellin **66** after reductive release and Knoevenagel cyclisation (Scheme 15).



Scheme 15 Involvement of TENS during the biosynthesis of tenellin **58**.

Two other genes in the TENS cluster encode cytochrome P450 oxidases and it is likely that these catalyse oxidative ring expansion to form the 2-pyridone, and *N*-hydroxylation.[91] *B. bassiana* is an effective insect pathogen and it has been hypothesised that tenellin may be involved in this process. However, directed knockout of TENS did not significantly reduce the pathogenicity of *B. bassiana* towards wax moth larvae.[91]

5.0 Perspective

A combination of molecular genetic and enzymological methods have revealed that fungal polyketide synthases are programmed nano-machines of unprecedented complexity. In particular, linking genes and compounds (Table 1) has greatly assisted the categorisation and understanding of these proteins. The smallest

fungal polyketide synthases are the 6-methylsalicylic acid synthases where the polypeptide chain has five functional domains and is *ca.* 191 kDa, while the largest PKS-NRPS weigh in at over 450 kDa and have up to eleven catalytic units. In all cases, the PKS is programmed. The simplest programmes control only chain length—three extensions for orsellinic acid for example, while the most complex programmes, involved in lovastatin **7a** biosynthesis, for example, control chain length, methylation, reduction and dehydration and additionally influence cyclisation. While similarly complex programmes are evident in the bacterial modular PKS, the fungal PKS are remarkable in achieving this control by a *single* set of enzymes contained within a single polypeptide.

Recognition that the fungal PKS can be categorised into three subsets has allowed more detailed consideration of the programming elements. The NR PKS are arranged into loading, extension and processing components, and to some extent this hypothesis has been verified by expression and study of individual catalytic domains and by the construction of hybrid NR PKS genes. However, the programming elements of the PR and HR PKS remain obscure. The first experiments to probe programming in HR PKS have involved domain swaps, but few conclusions have yet been drawn. Little information is known about the 3D structure of fungal PKS—while sequence and domain organisation similarity with mammalian FAS mean that broad descriptions of the architecture of the catalytic domains can be modelled, detailed hypotheses involving individual domains or peptide motifs cannot yet be linked with programming.

Notes and references

1 T. Kis-Papo, I. Grishkan, A. Oren, S. P. Wasser and E. Nevo, *Mycol. Res.*, 2001, **105**, 749–756.
2 S.-O. Suh, J. V. McHugh, D. D. Pollock and M. Blackwell, *Mycol. Res.*, 2005, **109**, 261–265.
3 C. L. Schardl, *Ann. Rev. Phytopathol.*, 1996, **34**, 109–130.
4 D. L. Hawksworth, *Mycol. Res.*, 2001, **105**, 1422–1432.
5 J. E. Cutler, G. S. Deepe and B. S. Klein, *Nat. Rev. Microbiol.*, 2007, **5**, 13–28.
6 A. Casadevall, *Fungal Genet. Biol.*, 2005, **42**, 98–106.
7 D. S. Heckman, D. M. Geiser, B. R. Eidell, R. L. Stauffer, N. L. Kardos and S. B. Hedges, *Science*, 2001, **293**, 1129–1133.
8 A. J. Birch and F. W. Donovan, *Aust. J. Chem.*, 1953, **6**, 360–368.
9 J. Staunton and K. J. Weissman, *Nat. Prod. Rep.*, 2001, **18**, 380–416.
10 R. J. Cox, T. S. Hitchman, K. J. Byrom, I. S. C. Findlow, J. A. Tanner, J. Crosby and T. J. Simpson, *FEBS Lett.*, 1997, **405**, 267–272.
11 R. J. Cox, J. Crosby, O. Daltrop, F. Glod, M. E. Jarzabek, T. P. Nicholson, M. Reed, T. J. Simpson, L. H. Smith, F. Soulas, A. E. Szafranska and J. Westcott, *J. Chem. Soc., Perkin Trans. 1*, 2002, 1644–1649.
12 D. Keszenman-Pereyra, S. Lawrence, M. E. Twfieg, J. Price and G. Turner, *Curr. Genet.*, 2003, **43**, 186–190.
13 T. Maier, S. Jenni and N. Ban, *Science*, 2006, **311**, 1258–1262.
14 L. E. H. Bingle, T. J. Simpson and C. M. Lazarus, *Fungal Genet. Biol.*, 1999, **26**, 209–223.
15 T. P. Nicholson, B. A. M. Rudd, M. Dawson, C. M. Lazarus, T. J. Simpson and R. J. Cox, *Chem. Biol.*, 2001, **8**, 157–178.
16 R. J. Howard, M. A. Ferrari, D. H. Roach and N. P. Money, *Proc. Natl. Acad. Sci. U. S. A.*, 1991, **88**, 11281–11284.
17 R. E. Minto and C. A. Townsend, *Chem. Rev.*, 1997, **97**, 2537–2555.
18 G. E. Yang, M. S. Rose, B. G. Turgeon and O. C. Yoder, *Plant Cell*, 1996, **8**, 2139–2150.
19 S. Kroken, N. L. Glass, J. W. Taylor, O. C. Yoder and B. G. Turgeon, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 15670–15675.
20 H. J. Pel, J. H. de Winde and D. B. Archer, *et al.*, *Nat. Biotechnol.*, 2007, **25**, 221–231.
21 G. M. Gaucher and M. G. Shepherd, *Biochem. Biophys. Res. Commun.*, 1968, **32**, 664.
22 D. S. J. McKeown, C. McNicholas, T. J. Simpson and N. J. Willett, *Chem. Commun.*, 1996, 301–302.
23 H. Hajjaj, A. Klaebe, M. O. Loret, G. Goma, P. J. Blanc and J. Francois, *Appl. Environ. Microbiol.*, 1999, **65**, 311–314.
24 K. Arai, B. J. Rawlings, Y. Yoshizawa and J. C. Vederas, *J. Am. Chem. Soc.*, 1989, **111**, 3391–3399.
25 F. E. Scott, T. J. Simpson, L. A. Trimble and J. C. Vederas, *J. Chem. Soc., Chem. Commun.*, 1984, 756–758.
26 S. V. Pathre, P. V. Khadikar and C. J. Mirocha, *Appl. Environ. Microbiol.*, 1989, **55**, 1955–1966.
27 Y. Yoshizawa, Z. Li, P. B. Reese and J. C. Vederas, *J. Am. Chem. Soc.*, 1990, **112**, 3212–3213; Z. Li, F. M. Martin and J. C. Vederas, *J. Am. Chem. Soc.*, 1992, **114**, 1531–1533.
28 T. S. Hitchman, E. W. Schmidt, F. Trail, M. D. Rarick, J. E. Linz and C. A. Townsend, *Bioorg. Chem.*, 2001, **29**, 293–307.
29 C. M. H. Watanabe and C. A. Townsend, *Chem. Biol.*, 2002, **9**, 981–988.
30 Y. Ma, L. H. Smith, R. J. Cox, P. Beltran-Alvarez, C. J. Arthur and T. J. Simpson, *ChemBioChem*, 2006, **7**, 1951–1958.
31 J. M. Crawford, B. C. R. Dancy, E. A. Hill, D. Udwary and C. A. Townsend, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 16728–16733.
32 I. Gaffoor and F. Trail, *Appl. Environ. Microbiol.*, 2006, **72**, 1793–1799.
33 Y. Liu, Z. Li and J. C. Vederas, *Tetrahedron*, 1998, **54**, 15937–15958.
34 I. Fujii, Y. Mori, A. Watanabe, Y. Kubo, G. Tsuji and Y. Ebizuka, *Biochemistry*, 2000, **39**, 8853–8858.
35 C. Bisang, P. F. Long, J. Cortes, J. Westcott, J. Crosby, A. L. Matharu, R. J. Cox, T. J. Simpson, J. Staunton and P. F. Leadlay, *Nature*, 1999, **401**, 502–505.
36 D. W. Udwary, M. Merski and C. A. Townsend, *J. Mol. Biol.*, 2002, **323**, 585–598.
37 A. Keatinge-Clay, D. A. Maltby, K. F. Medzihradszky, C. Khosla and R. M. Stroud, *Nat. Struct. Mol. Biol.*, 2004, **11**, 888–893.
38 B. Feng, X. Wang, M. Hauser, S. Kaufmann, S. Jentsch, G. Haase, J. M. Becker and P. Szaniszlo, *Infect. Immun.*, 2001, **69**, 1781–1794.
39 R. J. Cox, *unpublished results*.
40 C. A. Townsend, *personal communication*.
41 M. E. Mayorga and W. E. Timberlake, *Genetics*, 1990, **126**, 73–79.
42 M. E. Mayorga and W. E. Timberlake, *Mol. Gen. Genet.*, 1992, **235**, 205–212.
43 J.-H. Yu and T. J. Leonard, *J. Bacteriol.*, 1995, **177**, 4792–4800.
44 P. K. Chang, J. W. Cary, J. Yu, D. Bhatnagar and T. E. Cleveland, *Mol. Gen. Genet.*, 1995, **248**, 270–277.
45 A. Watanabe, Y. Ono, I. Fujii, U. Sankawa, M. E. Mayorga, W. E. Timberlake and Y. Ebizuka, *Tetrahedron Lett.*, 1998, **39**, 7733–7736.
46 A. Watanabe, I. Fujii, U. Sankawa, M. E. Mayorga, W. E. Timberlake and Y. Ebizuka, *Tetrahedron Lett.*, 1999, **40**, 91–94.
47 I. Fujii, A. Watanabe, U. Sankawa and Y. Ebizuka, *Chem. Biol.*, 2001, **8**, 189–197.
48 Y. Takano, Y. Kubo, K. Shimizu, K. Mise, T. Okuno and I. Furusawa, *Mol. Gen. Genet.*, 1995, **249**, 162–167.
49 I. Fujii, Y. Mori, A. Watanabe, Y. Kubo, G. Tsuji and Y. Ebizuka, *Biosci., Biotechnol., Biochem.*, 1999, **63**, 1445–1452.
50 A. Watanabe and Y. Ebizuka, *Chem. Biol.*, 2004, **11**, 1101–1106.
51 A. Watanabe and Y. Ebizuka, *Tetrahedron Lett.*, 2002, **43**, 843–846.
52 T. P. Nicholson, C. Winfield, J. Westcott, J. Crosby, T. J. Simpson and R. J. Cox, *Chem. Commun.*, 2003, 686–687.
53 A. Watanabe, I. Fujii, H.-F. Tsai, Y. C. Chang, K. J. Kwon-Chung and Y. Ebizuka, *FEMS Microbiol. Lett.*, 2000, **192**, 39–44.
54 H.-F. Tsai, I. Fujii, A. Watanabe, M. H. Wheeler, Y. C. Chang, Y. Yasuoka, Y. Ebizuka and K. J. Kwon-Chung, *J. Biol. Chem.*, 2001, **276**, 29292–29298.
55 R. E. Bradshaw, H. P. Jin, B. S. Morgan, A. Schwelm, O. R. Teddy, C. A. Young and S. G. Zhang, *Mycopathologica*, 2006, **161**, 283–294.
56 T. Shimizu, H. Kinoshita, S. Ishihara, K. Sakai, S. Nagai and T. Nihira, *Appl. Environ. Microbiol.*, 2005, **71**, 3453–3457.
57 T. J. Simpson and M. K. B. Weerasooriya, *J. Chem. Soc., Perkin Trans. 1*, 2000, 2771–2775.
58 A. G. Soman, J. B. Gloer and D. T. Wicklow, *J. Nat. Prod.*, 1999, **62**, 386–388.
59 J. Beck, S. Ripka, A. Siegner, E. Schiltz and E. Schweizer, *Eur. J. Biochem.*, 1990, **192**, 487–498.
60 J. Ahlert, E. Shepard, N. Lomovskaya, E. Zazopoulos, A. Staffa, B. O. Bachmann, K. Huang, L. Fonstein, A. Czisny, R. E. Whitwam, C. M. Farnet and J. S. Thorson, *Science*, 2002, **297**, 1173–1176.

61 I. Fujii, Y. Ono, H. Tada, K. Gomi, Y. Ebizuka and U. Sankawa, *Mol. Gen. Genet.*, 1996, **253**, 1–10.
62 P. Lu, A. Zhang, L. M. Dennis, A. M. Dahl-Roshak, Y.-Q. Xia, B. Arison, Z. An and J. S. Tkacz, *Mol. Genet. Genomics*, 2005, **273**, 207–216.
63 R. J. Light, *J. Biol. Chem.*, 1967, **242**, 1880–1886; P. Dimroth, H. Walter and F. Lynen, *Eur. J. Biochem.*, 1970, **13**, 98–110.
64 J. B. Spencer and P. M. Jordan, *Biochem. J.*, 1992, **288**, 839–846.
65 C. J. Child, J. B. Spencer, P. Bhogal and P. M. Shoolingin-Jordan, *Biochemistry*, 1996, **35**, 12267–12274.
66 I. D. G. Campuzano and P. M. Shoolingin-Jordan, *Biochem. Soc. Trans.*, 1998, **26**, S284–S284.
67 D. J. Bedford, E. Schweizer, D. A. Hopwood and C. Khosla, *J. Bacteriol.*, 1995, **177**, 4544–4548.
68 J. T. Kealey, L. Liu, D. V. Santi, M. C. Betlach and P. J. Barr, *Proc. Natl. Acad. Sci. U. S. A.*, 1998, **95**, 505–509.
69 T. Moriguchi, Y. Ebizuka and I. Fujii, *ChemBioChem*, 2006, **7**, 1869–1874.
70 A. Witkowski, A. K. Joshi and S. Smith, *Biochemistry*, 2004, **43**, 10458–10466.
71 L. Hendrickson, C. R. Davis, C. Roach, D. K. Nguyen, T. Aldrich, P. C. McAda and C. D. Reeves, *Chem. Biol.*, 1999, **6**, 429–439.
72 J. Kennedy, K. Auclair, S. G. Kendrew, C. Park, J. C. Vederas and C. R. Hutchinson, *Science*, 1999, **284**, 1368–1372.
73 Y. Abe, T. Suzuki, C. Ono, K. Iwamoto, M. Hosobuchi and H. Yoshikawa, *Mol. Genet. Genomics*, 2002, **267**, 636–646.
74 A. K. El-Sayed, J. Hothersall, S. M. Cooper, E. Stephens, T. J. Simpson and C. M. Thomas, *Chem. Biol.*, 2003, **10**, 419–430.
75 X. Xie, K. Watanabe, W. A. Wojcicki, C. C. C. Wang and Y. Tang, *Chem. Biol.*, 2006, **13**, 1161–1169.
76 C. Bate, M. Salmona, L. Diomede and A. Williams, *J. Biol. Chem.*, 2004, **279**, 14983–14990.
77 R. J. Cox, F. Glod, D. Hurley, C. M. Lazarus, T. P. Nicholson, B. A. M. Rudd, T. J. Simpson, B. Wilkinson and Y. Zhang, *Chem. Commun.*, 2004, 2260–2261.
78 G. Wang, M. S. Rose, B. G. Turgeon and O. C. Yoder, *Plant Cell*, 1996, **8**, 2139–2150.
79 S. E. Baker, S. Kroken, P. Inderbitzin, T. Asvarak, B.-Y. Li, L. Shi, O. C. Yoder and B. G. Turgeon, *Mol. Plant-Microbe Interact.*, 2006, **19**, 139–149.
80 R. H. Proctor, A. E. Desjardins, R. D. Plattner and T. M. Hohn, *Fungal Genet. Biol.*, 1999, **27**, 100–112.
81 R. A. E. Butchko, R. D. Plattner and R. H. Proctor, *J. Agric. Food Chem.*, 2006, **54**, 9398–9404.
82 X. Zhu, F. Yu, R. S. Bojja, K. Zaleta-Rivera and L. Du, *J. Ind. Microbiol. Biotechnol.*, 2006, **33**, 859–868.
83 X. Zhu, F. Yu, X.-C. Li and L. Du, *J. Am. Chem. Soc.*, 2007, **129**, 36–37.
84 I. Fujii, N. Yoshida, S. Shimomaki, H. Oikawa and Y. Ebizuka, *Chem. Biol.*, 2005, **12**, 1301–1309.
85 K. Kasahara, I. Fujii, H. Oikawa and Y. Ebizuka, *ChemBioChem*, 2006, **7**, 920–924.
86 D. O. Rees, N. Bushby, R. J. Cox, J.R. Harding, T. J. Simpson and C. L. Willis, *ChemBioChem*, 2007, **8**, 46–50.
87 Z. S. Song, R. J. Cox, C. M. Lazarus and T. J. Simpson, *ChemBioChem*, 2004, **5**, 1196–1203.
88 M. J. Hijarrubia, J. F. Aparicio and J. F. Martin, *J. Biol. Chem.*, 2003, **278**, 8250–8256.
89 J. W. Sims, J. P. Fillmore, D. D. Warner and E. W. Schmidt, *Chem. Commun.*, 2005, 186–188.
90 H. U. Bohnert, I. Fudal, W. Dioh, D. Tharreau, J. L. Notteghem and M. H. Lebrun, *Plant Cell*, 2004, **16**, 2499–2513.
91 K. L. Eley, L. M. Halo, Z. Song, H. Powles, R. J. Cox, A. M. Bailey, C. M. Lazarus and T. J. Simpson, *ChemBioChem*, 2007, **8**, 289–297.

