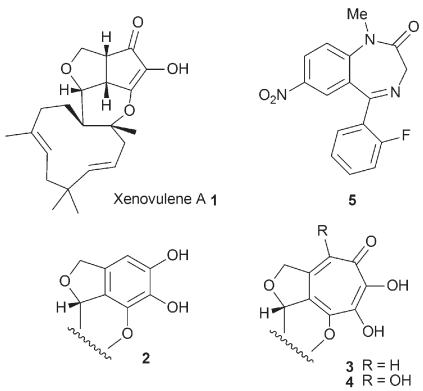
Characterisation of 3-methylorcinaldehyde synthase (MOS) in *Acremonium strictum*: first observation of a reductive release mechanism during polyketide biosynthesis†

Andrew M. Bailey,[b] Russell J. Cox,*[a] Kate Harley,[a] Colin M. Lazarus,[b] Thomas J. Simpson*[a] and Elizabeth Skellam[a]

Received (in Cambridge, UK) 6th June 2007, Accepted 28th June 2007
First published as an Advance Article on the web 25th July 2007
DOI: 10.1039/b708614h

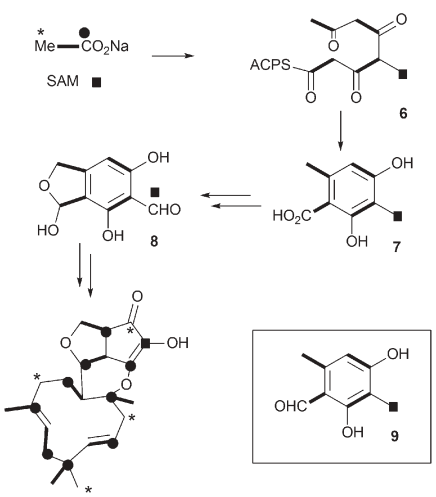
Isolation and sequencing of a PKS gene isolated from xenovulene-producing cultures of *Acremonium strictum* indicated the presence of NT-, KS-, AT-, PT-, *C*-MeT- and R-domains; heterologous expression in *Aspergillus oryzae* resulted in the production of 3-methylorcinaldehyde, demonstrating the role of the terminal reductase domain in product release.

The fungal plant pathogen *Acremonium strictum* produces xenovulene A **1**, together with lesser quantities of the related compounds **2**, **3** and **4**.[1] Xenovulene A **1** is able to displace high affinity ligands from the γ-aminobutyric acid $(GABA)_A$ receptor.[2] For example, **1** inhibits flunitrazepam **5** (Rohypnol) binding to rat forebrain $GABA_A$ receptors with a K_i of 7 nM. The biosynthesis of such pharmacologically useful compounds is of high interest because genetic manipulation of the corresponding biosynthetic genes can lead to the generation of analogous compounds with varying biological properties.[3]



On the basis of incorporation studies with ^{13}C-labelled acetate and methionine it was proposed that the xenovulenes are biosynthesised *via* 3-methylorsellinic acid **7**, itself derived from a *C*-methylated tetraketide intermediate **6** (Scheme 1).[4] Subsequent modifications would produce the lactol **8**, which could be linked, possibly *via* dehydration to an *o*-quinonoid intermediate and subsequent hetero-Diels Alder cycloaddition, to humulene.[5] This would be followed by an oxidative ring expansion and two ring contraction steps to the cyclopentenone moiety of **1** *via* the tropolone (**3/4**) and phenolic **2** co-metabolites. Closely related tropolone–humulene fused metabolites pycnidione and eupenifeldin have been isolated from cultures of a *Phoma* sp.[6] and *Eupenicillium brefeldianum*[7] respectively. In this paper we describe molecular genetic studies leading to the isolation and heterologous expression of a polyketide synthase (PKS) gene, which suggest that the true initial polyketide-derived intermediate is 3-methylorcinaldehyde (2,4-dihydroxy-3,5-dimethylbenzaldehyde) **9**.

We have previously described the design and utility of degenerate oligonucleotides with biased sequences that can be used for the amplification from fungal genomes of defined short DNA sequences corresponding to biosynthetic genes.[8] In particular, we have described the design of a set of oligonucleotide polymerase chain reaction (PCR) primers that are able to selectively amplify fragments of fungal PKS genes that are involved in the biosynthesis of separate chemical types of polyketide.[9] We defined these chemical types to be non-reduced (NR), partially reduced (PR) and highly reduced (HR). Since we first showed the relationship between DNA sequence and chemical structure for fungal polyketide synthases, based on limited available sequences, others have extended our analysis to include



Scheme 1 Incorporation of C_1 and C_2 units into **1**.

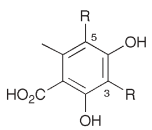
[a]*School of Chemistry, University of Bristol, Cantock's Close, Bristol, UK BS8 1TS. E-mail: tom.simpson@bristol.ac.uk; r.j.cox@bristol.ac.uk; Fax: +44 117 9298611; Tel: +44 117 9287656*
[b]*School of Biological Sciences, University of Bristol, Woodland Road, Bristol, UK BS8 1UG*

† Electronic supplementary information (ESI) available: Details of cloning and expression studies, DNA sequence of *ASpks1*, peptide sequence of MOS, isolation and purification of 3-methylorcinaldehyde and X-ray crystallographic details. See DOI: 10.1039/b708614h

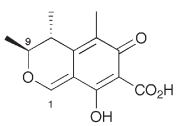
numerous PKS genes discovered by genomics programmes.[10] These new analyses have included a limited number of examples where both gene sequence and corresponding chemical structure are known, but also very many examples where PKS genes have not yet been linked to chemical structures. However, despite these limitations, the relationship between gene sequence and chemical structure is remarkably well conserved for the fungal PKS genes.

Thus, although no fungal PKS genes are known that biosynthesise orsellinic acid **10**, or its *C*-methylated derivatives (*e.g.* **7**, **11**), we speculated that we could predict related sequences from fungal genomic datasets. We expected a fungal PKS gene involved in methylorsellinic acid biosynthesis to utilise keto-synthase (KS), acyl transferase (AT), acyl carrier protein (ACP) and *C*-MeT-domains, but no reductive activities such as β-ketoacyl reductase (KR) or enoyl reductase (ER)-domains. Kroken and co-workers have identified a class of fungal PKS genes that encode domains corresponding to this analysis—the so-called non-reducing Clade III,[10] although only one of the sequences in this clade has been linked to a chemical structure, *vide infra*. We nevertheless designed degenerate PCR primers (KHKS2 and KHKS3c) based on conserved regions of KS-domains from the non-reducing Clade III sequences—the primer positions were chosen such that their position was unique to this clade and their sequence unlikely to amplify members of other clades.†

Genomic DNA was extracted from *A. strictum* and used as the template for PCR reactions with the KHKS2 and KHKS3c primers. The genomic DNA was also used to prepare a library. The PCR reaction afforded a 245 bp product, which was sequenced and shown to be highly homologous to other DNA sequences of the non-reducing Clade III PKS genes. This was labelled with ^{32}P and used to probe the *A. strictum* genomic DNA library, leading to the isolation of a 22 226 bp contig containing nine putative open reading frames (ORF). The largest of these is a 9.6 kb ORF encoding a Type I iterative PKS (*ASpks1*).



10 R = H
11 R = Me



12

Alignment with consensus sequences revealed frame-shifts and in-frame stop codons indicative of the presence of two possible introns. These were confirmed by reverse transcription (RT)-PCR procedures. The translated intron-free sequence was then analysed for catalytic domains by simple alignment with Type II PKS proteins as we have previously described.[11] This analysis revealed the presence of a putative *N*-terminal (NT)-domain, KS- and AT-domains, a putative product template (PT)-domain, an ACP and, following this, a *C*-MeT-domain (Fig. 1). At the C-terminus of the sequence is a domain corresponding to an NAD(P)H dependent thiolester reductase (R). A number of homologous C-terminal reductases have been found in association with peptidyl carrier proteins in several non-ribosomal peptide synthetases (NRPS) and

Fig. 1 MOS domain structure.

these have been demonstrated to be involved in reductive release processes. For example, in the NRPS responsible for gramicidin biosynthesis, the terminal R-domain of LgrD releases the product as an aldehyde, which is further reduced by a second reductase encoded by *LgrE* located upstream of the NRPS genes.[12] In the case of myxochelin biosynthesis in *Stigmatella aurantiaca*, the C-terminal NRPS module, MxcG, itself contains a C-terminal R-domain, which acts *twice*, releasing an alcohol.[13] We ourselves have shown that the PKS-NRPS systems involved in the biosynthesis of the fungal metabolites fusarin C[14] and tenellin[15] also contain homologous reductase domains at the C-termini of the NRPS, though their function remains to be firmly established. Database analyses have also revealed the presence of similar reductase domains in association with several PKS genes and, while a role in reductive release has been postulated, this has never been established.

In order to determine the role of *ASpks1*, it was cloned into the fungal vector pTAex3 between the *amy*B promoter and terminator. pTAex3 also contains the *arg*B gene, allowing for selection in an arginine autotroph (*arg*B$^-$) of the fungus *Aspergillus oryzae*.[16] Transformation of *A. oryzae arg*B$^-$ with pTAex3-*ASpks1* and subsequent incubation in minimal media resulted in the isolation of 9 transformants. Of these, 5 showed red coloration when grown on either solid or liquid media. TLC and LCMS analysis of the organic extracts of these clones revealed the presence of a new compound. HRMS analysis indicated a molecular formula of $C_9H_{10}O_3$ (M^+, EI, observed 166.0630, calculated 166.0630). Organic extracts were prepared from five clones grown in liquid media (5 × 100 ml). The extracts were combined and purified by flash chromatography to yield 28.4 mg of a colourless solid.† The ^1H NMR spectrum showed the presence of signals that could be assigned to aromatic methyl (2.08 ppm, 3H, s) and (2.50, 3H, s), aldehyde (10.07, 1H, s), and an aromatic hydrogen (6.20, 1H, s), which suggested that the new compound was 3-methylorcinaldehyde **9**. This was confirmed by X-ray crystallography.‡ *ASpks1* thus encodes 3-methylorcinaldehyde synthase, which we have designated MOS.

The C-terminal R-domain must be involved in a reductive release mechanism, reducing a putative thiolester intermediate to the observed aldehyde and concomitantly releasing the free *holo*-ACP thiol. From the structure of the product, the R-domain cannot be involved in any processive reductive chemistry, as is observed in other fungal and bacterial Type I PKS. Indeed, its position *after* the ACP is quite different to that observed in other PKSs, where reductive modifications occur during biosynthesis, and reductive domains are placed *before* the ACP in the peptide sequence. To the best of our knowledge, MOS is the first established example of a PKS using a reductive release mechanism.

The position of the R-domain at the very C-terminus of MOS also points to its role in product release. Known mechanisms for product release from PKS enzymes utilise either thiolesterase (TE) or Claisen cyclase (CLC)[17] activities. Interestingly, the only other Clade III gene to have been chemically characterised is the citrinin PKS gene *pksCT* from *Monascus purpurea*. No domain indicating a release mechanism was reported but re-examination of the reported sequence[18] reveals a homologous C-terminal reductase domain, which is consistent with the structure of citrinin **12**, which requires reduction of the polyketide-derived intermediate to the aldehyde level at C-1 at some stage in the biosynthesis.[19] It would

 This journal is © The Royal Society of Chemistry 2007

thus be reasonable to propose that the *pksCT* encoded R-domain is also implicated in reductive release.

The positioning of the *C*-MeT-domain in MOS after the ACP is unusual in that in all known HR PKSs, *e.g.* SQTKS, LNKS, LDKS, TENS and FUSS,[20] are found before the ACP- between the dehydratase (DH)- and KR-domains and are presumed to introduce the methyl branches in a processive fashion during iterative chain assembly. It is not clear, however, whether the MOS *C*-MeT acts during chain extension, or after chain completion and cyclisation, but before reductive product release. It could conceivably methylate the highly electron rich aromatic intermediate before release rather than after, as there is evidence to suggest that *C*-methylation in *e.g.* 3,5-dimethylorsellinate must occur before product release as both ^{13}C and ^{14}C-labelled 3,5-dimethylorsellinates **11** are efficiently incorporated into the meroterpenoid family of metabolites, *e.g.* andibenin B,[21] in contrast to ^{14}C-labelled orsellinic acid **10**, which showed no evidence for any incorporation.

Orsellinic acid **10** is the simplest tetraketide, requiring only control of chain length during its biosynthesis since no reductive modifications are required. It is thus surprising that no examples of fungal orsellinic acid synthase (OAS) genes have been discovered prior to this study. It might have been expected that putative OAS genes would resemble the well known genes responsible for 6-methylsalicylic acid (6MSA) biosynthesis in terms of domain structure.[22] This is because 6MSA undergoes an identical set of biosynthetic reactions to **10**, except for a single KR reaction. 6MSA synthase (MSAS) thus contains KS-, AT-, DH-, KR- and ACP-domains.

However, while MOS does lack a KR-domain, it also possesses additional NT- and PT-domains as well as the *C*-MeT- and R-domains discussed above. Townsend and co-workers have recently shown[23] that the NT-domain of norsolorinic acid synthase (NSAS) from *Aspergillus parasiticus* is involved in loading the hexanoate starter unit, while we showed[11] that the NSAS AT-domain cannot transfer acetyl or hexanoyl groups. Sequence analysis indicates that the NT-domains of MOS and NSAS share low sequence conservation (15.9% identity; 28.4% similarity). However the key catalytic residues C117 and H245 (NSAS numbering) responsible for acyl transfer are conserved. Conserved NSAS Q13 may correspond with MOS Q23. It appears possible that the MOS NT-domain selects acetate as the starter unit and the low sequence homology between NSAS and MOS NT-domains reflects the different substrate selectivities.

The PT-domain of MOS was also examined. It has been hypothesised that this domain controls chain length[23] Several fungal PKS are known where both PT-domain sequence and chemical structure of the product are known. Phylogenetic analysis of PT-domains from these synthases shows a distinct relationship between chain-length and clade structure of the tree-plot.[20] An alternative explanation for the clustering observed would be the number of chain extensions catalysed by the synthase, ranging from 3 to 7 (Fig. 2). The MOS PT-domain clusters with the PT-domain from zearalenone PKSB, which is also thought to be a tetraketide synthase. Interestingly, this suggests that the citrinin PKS is also a tetraketide synthase, which must therefore use a reduced and methylated C_4 starter unit—this explains the presence of the reduced carbon at C-9 of citrinin **12** despite the lack of a KR-domain in the citrinin PKS. Further clarification of this issue will await matching of further Clade III PKSs to chemical structures.

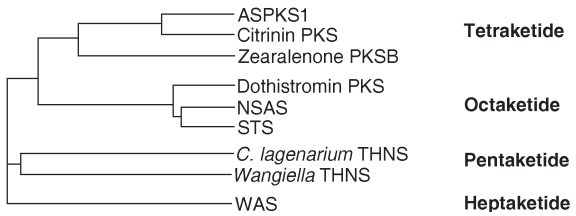


Fig. 2 Phylogenetic analysis of Clade III PKS PT-domains.

We thank the BBSRC for sudentships (KH and ES).

Notes and references

‡ CCDC 649826. For crystallographic data in CIF or other electronic format see DOI: 10.1039/b708614h

1 A. M. Ainsworth, M. I. Chicarelli-Robinson, B. R. Copp, U. Fauth, P. J. Hylands, J. A. Holloway, M. Latif, G. B. O'Beirne, N. Porter, D. V. Renno, M. Richards and N. Robinson, *J. Antibiot.*, 1995, **48**, 568.
2 P. Thomas, H. Sundarum, B. J. Krishek, P. Chazot, X. Xie, P. Bevan, S. J. Brocchini, C. J. Latham, P. Charlton, M. Moore, S. J. Lewis, D. M. Thorton, F. A. Stephenson and T. G. Smart, *J. Pharmacol. Exp. Ther.*, 1997, **282**, 513.
3 K. J. Weissman and P. F. Leadlay, *Nat. Rev. Microbiol.*, 2005, **3**, 925.
4 M. E. Raggatt, T. J. Simpson and M. I. Chicarelli-Robinson, *Chem. Commun.*, 1997, 2245.
5 For a biomimetic example see: J. E. Baldwin, A. V. W. Mayweg, K. Neumann and G. J. Pritchard, *Org. Lett.*, 1999, **1**, 1933.
6 G. H. Harris, K. Hoogsteen, K. C. Silverman, S. L. Raghoobar, G. F. Bills, R. B. Lingham, J. L. Smith, H. W. Dougherety, C. Cascales and F. Pelaez, *Tetrahedron*, 1993, **49**, 2139.
7 F. Mayerl, Q. Gao, S. Huang, S. E. Klohr, J. A. Matson, D. R. Gustavson, R. L. Berry, C. Fairchild and W. C. Rose, *J. Antibiot.*, 1993, **46**, 1082.
8 L. E. H. Bingle, T. J. Simpson and C. M. Lazarus, *Fungal Genet. Biol.*, 1999, **26**, 209.
9 T. P. Nicholson, B. A. M. Rudd, M. Dawson, C. M. Lazarus, T. J. Simpson and R. J. Cox, *Chem. Biol.*, 2001, **8**, 157.
10 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.
11 Y. Ma, L. H. Smith, R. J. Cox, P. Beltran-Alvarez, C. J. Arthur and T. J. Simpson, *ChemBioChem*, 2006, **7**, 1951.
12 N. Schrake, U. Linne and C. Mahlert amd M. A. Marahiel, *Biochemistry*, 2005, **44**, 8507.
13 N. Gaitatzis, B. Kunze and R. Mueller, *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **98**, 11136.
14 Z. Song, R. J. Cox, C. M. Lazarus and T. J. Simpson, *ChemBioChem*, 2004, **5**, 1196.
15 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.
16 A. Watanabe, Y. Ono, I. Fujii. U. Sankawa, M. E. Mayorga, W. E. Timberlake and Y. Ebizuka, *Tetrahedron Lett.*, 1998, **39**, 7733.
17 A. Watanabe, I. Fujii, U. Sankawa, M. E. Mayorga, W. E. Timberlake and Y. Ebizuka, *Tetrahedron Lett.*, 1999, **40**, 91.
18 T. Shimizu, H. Kinoshita, S. Ishihara, K. Sakai, S. Nagai and T. Nihira, *Appl. Environ. Microbiol.*, 2005, **71**, 3453.
19 L. Colombo, C. Gennari, D. Potenza, C. Scolastico, F. Aragozzini and C. Merendi, *J. Chem. Soc., Perkin Trans. 1*, 1981, 2594.
20 R. J. Cox, *Org. Biomol. Chem.*, 2007, **5**, 2010.
21 A. J. Bartlett, J. S. E. Holker, E. O'Brien and T. J. Simpson, *J. Chem. Soc., Chem. Commun.*, 1981, 1198.
22 J. Beck, S. Ripka, A. Siegner, E. Schiltz and E. Schweizer, *Eur. J. Biochem.*, 1990, **192**, 487.
23 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.