COMMUNICATION

ChemComm
www.rsc.org/chemcomm

Rapid cloning and expression of a fungal polyketide synthase gene involved in squalestatin biosynthesis†

Russell J. Cox,*[a] Frank Glod,[a] Deirdre Hurley,[a] Colin M. Lazarus,[b] Thomas. P. Nicholson,[a] Brian A. M. Rudd,[c]‡ Thomas J. Simpson,[a] Barrie Wilkinson[c]‡ and Ying Zhang[a]

[a] *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 9298611*

[b] *School of Biological Sciences, University of Bristol, Woodland Road, Bristol BS8 1UG*

[c] *GlaxoSmithKline, Gunnels Wood Road, Stevenage, Herts, SG1 2NY*

Received (in Cambridge, UK) 4th August 2004, Accepted 3rd September 2004
First published as an Advance Article on the web 24th September 2004

PCR primers designed to selectively amplify the unique *C*-methyltransferase domain of fungal polyketide synthases were used to selectively clone a polyketide synthase gene involved in the biosynthesis of the squalene synthase inhibitor squalestatin S1 1, heterologous expression of which led to the biosynthesis of the squalestatin side-chain.

Polyketides biosynthesised by fungi display a remarkably rich diversity of structural motifs and accompanying biological activities. As part of our ongoing efforts to exploit the biosynthetic potential of fungal polyketide synthases (PKS) for the potential directed biosynthesis of new compounds we have devised a strategy which allows the selective and rapid cloning of specific fungal PKS genes.[1] Squalestatin S1 **1** (also known as zaragozic acid A) is a potent and selective inhibitor of squalene synthase, identified by scientists at Glaxo and Merck as a metabolite of *Phoma sp.* and other filamentous fungi[2] which displays significant potential for the treatment of diseases associated with elevated serum cholesterol concentrations.

The biosynthesis of **1** has been investigated using stable isotope feeding experiments (Fig. 1).[3] These experiments indicate that **1** is composed of two polyketide chains. The main chain (chain A) is a hexaketide, composed of a benzoate derived starter unit and five acetate-derived extender units. The side chain (chain B) is a tetraketide, with acetate as the starter unit. Four more carbons are derived from succinate, forming the distinctive 5-hydroxy-[1,3]dioxane-4,5,6-tricarboxylic acid core of **1**.

A feature of fungal polyketide synthases is the mechanism of attachment of pendant methyl groups. In Type I bacterial PKS, propionate is activated to methylmalonate and used as an extender unit. However, in fungi, a distinct *C*-methyltransferase (*C*-MeT) domain of the Type I PKS transfers a methyl group from *S*-adenosyl methionine (SAM). Squalestatin possesses four carbons derived in this way, two on each of the polyketide chains.

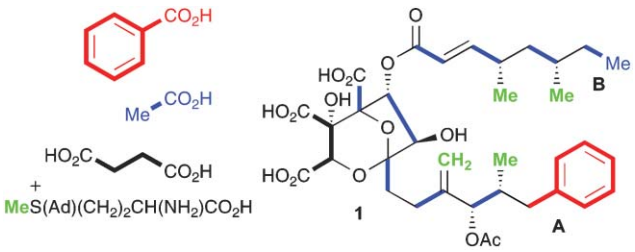


Fig. 1 The biosynthetic origin of squalestatin S1 **1**.

We reasoned that because both of the PKS proteins responsible for the biosynthesis of the component chains of **1** should bear *C*-MeT domains, a gene-probing approach using a *C*-MeT specific probe should give rapid access to squalestatin biosynthetic genes in the host fungi. We planned to obtain the gene-probe by polymerase chain reaction (PCR) from the genome of the target organism using short degenerate oligonucleotides based upon known conserved sequences of fungal PKS *C*-MeT domains.[1] Importantly, the sequence of these PCR primers was biased away from the ubiquitous *O*- and *N*-methyl transferases involved in primary metabolism and polynucleotide processing.

We began by plotting a time-course of **1** production (Fig. 2) in *Phoma sp.* C2932. Results from these experiments showed that **1** begins production on day 4 and is rapidly produced between days 4 to 6. This suggested that gene expression of the squalestatin PKS occurred from day 4 onwards and thus RNA isolated on these days would contain transcripts of interest. This was confirmed when cDNA prepared from RNA collected on day 4 was successfully used as a template in PCR reactions using the degenerate *C*-MeT primers. These PCR reactions gave products of the expected 320 base pairs (bp) in length, and sequence analysis confirmed them to be highly homologous to known *C*-MeT encoding nucleotide sequences.

We next constructed a λ-phage cDNA library from RNA isolated on day 5. Approximately 3×10^5 λ-plaques were transferred to a nylon membrane and probed with the ^{32}P-labelled 320 bp PCR products. Sequence analysis of hybridised plaques revealed the presence of a 4.8 Kb fragment of a Type I PKS, containing the polyadenylated 3′ end of the sequence. This clone contained (5′→3′) a truncated fragment of a dehydratase domain (DH) and complete *C*-MeT, enoyl reductase (ER), keto-reductase (KR) and acyl carrier protein (ACP) domains.

The 5′ portion of the PKS was then obtained using a process involving the rapid amplification of cDNA ends (RACE) to produce an overlapping clone, again of 4.8 Kb containing β-ketoacyl synthase (KS), acyl transferase (AT), dehydratase (DH) and *C*-MeT domains. A contiguous, full-length clone was then constructed by digest and religation of the two cDNA fragments.

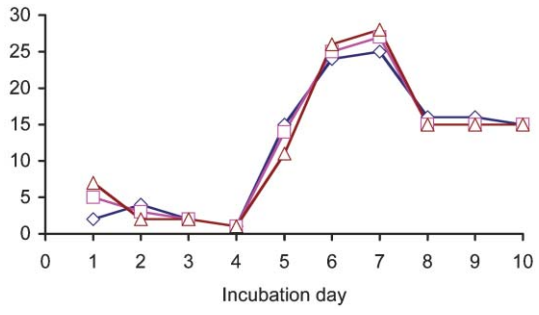


Fig. 2 Time course of squalestatin production for three replicates.

† Electronic supplementary information (ESI) available: chemical characterisation of **2** and details of its isolation and purification. See http://www.rsc.org/suppdata/cc/b4/b411973h/

‡ Current Addresses: BAMR, Novacta Biosystems, UH Innovations Centre, College Lane, Hatfield, Herts, AL10 9AB; BW, Biotica Technology Ltd, Chesterford Research Park, Little Chesterford Essex CB10 1XL.

DOI: 10.1039/b411973h

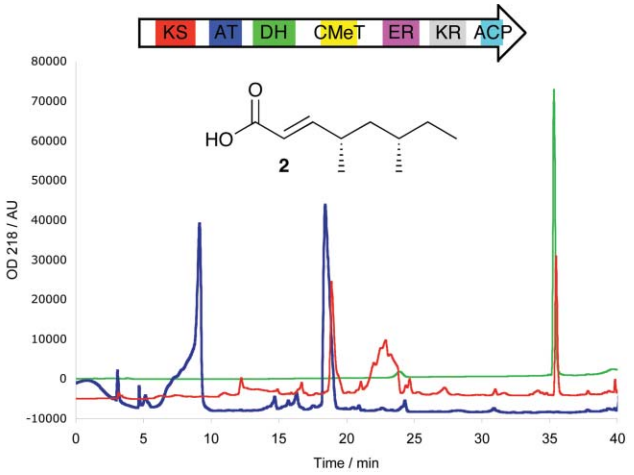


Fig. 3 Approximate domain locations in SQTKS. HPLC traces show WT *A. oryzae* extract (blue); *A. oryzae* expression clone grown in presence of starch (red) and purified **2** (green).[6]†

The reconstructed 8142 bp PKS (Phoma PKS1) contained KS, AT, DH, *C*-MeT, ER, KR and ACP domain encoding regions (Fig. 3). The translated gene, encoding a polypeptide of 2714 amino acids, shows high end-to-end amino acid homology with other known fungal PKS involved in the biosynthesis of highly reduced polyketides such as the lovastatin diketide synthase (LDKS, 59% identity),[4] and the compactin diketide synthase (60% identity).[5]

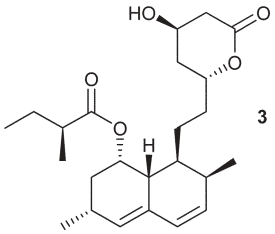
Close inspection of individual domains suggested they were catalytically competent. For example in the ER domain, the conserved NADPH binding motif (LxHx(G/A)xGGVG) was present (LIHAASGGVG). This contrasts with the situation in other fungal PKS genes such as the lovastatin nonaketide synthase (LNKS) where the ER domain is probably inactive due to deleterious mutations in the NADPH binding region.[4]

We then exploited the *Aspergillus oryzae* fungal expression system which has been used by Ebizuka and Fujii for the successful expression of fungal unreduced PKS.[7] Thus Phoma PKS1 was inserted into the expression vector pTAex3 to form a 15747 bp expression construct in *Escherichia coli*. *A. oryzae* strain M-2-3 (*arg*B$^-$) protoplasts were transformed with the expression vector and selected on arginine deficient plates. True transformants were selected by repeated sub-cloning on arginine deficient media, and further confirmed by colony PCR using Phoma PKS1 specific primers and by Southern blotting.

The pTAex3 expression system utilises the *amy*B promoter which is repressed by glucose and induced by starch. We thus grew the *A. oryzae* transformants in the presence of starch, and grew wild-type (WT) *A. oryzae* in parallel as a control. Organic extracts of the medium from the WT and transformant strains were then analysed by RP-HPLC. Of five transformants shown to have integrated the Phoma PKS1, one strain showed the presence of a new compound in the HPLC trace (Fig. 3). This compound was isolated and purified by repeated RP-HPLC.

Full structural characterisation revealed the new compound to be the doubly methylated unsaturated acid **2**. Comparison of optical rotation data with synthetic material proved it to be the 4*S*,6*S* enantiomer, and thus chain B of squalestatin. We thus name the protein encoded by Phoma PKS1 as SQTKS (Squalestatin Tetraketide Synthase).

Very few successful expression experiments with highly reduced fungal polyketide synthase genes have been reported. A rare case is that of the lovastatin **3** nonaketide synthase (LNKS) from *A. terreus*.[8] In the case of LNKS, heterologous expression in *A. nidulans* leads to the production of polyunsaturated pyrones which are shorter than the expected nonaketide. The polyunsaturated products indicate the inactivity of the LNKS ER. However coexpression with an accessory protein (*lov*C) ensures correct enoyl reduction and chain extension. In *A. terreus* strains where *lov*C is knocked out, and thus the nonaketide is not made, the methylbutyrate diketide side-chain is not observed as a by-product. The fact that supplementing the *lov*C deficient strain with synthetic nonaketide then gives a high yield of **3** indicates that the diketide sidechain has to be transferred to the nonaketide to off-load it from its synthase, LDKS.[9]



SQTKS shows high homology to LDKS, but expression of SQTKS leads to detectable concentrations of the squalestatin sidechain in the media, indicating a possible dissociative mode of attachment to the squalestatin A chain. This is supported by the observation that **1** co-occurs with numerous cometabolites featuring a wide range of different acyl group side chains.

It is also evident, that unlike LNKS, all the domains of SQTKS are active. The biosynthesis must involve three rounds of chain extension. After the first and second rounds methyl-transfer occurs, and in all rounds of extension the ketoreductase and dehydratase are active. The enoyl reductase and *C*-MeT are not active in the final round of extension. SQTKS thus provides an intriguing system to study programming by fungal PKS. SQTKS is also similar to single modules of bacterial Type I PKS, the key difference being that SQTKS is iterative. The SQTKS expression system will thus form an excellent comparison with simple modular PKS systems such as DEBS1-TE[10] which have proven useful test-beds for *in vivo* and *in vitro* enzymology studies.

We thank the School of Chemistry, University of Bristol for funding and Professor Isao Fujii for the gift of pTAex-3. TPN and FG thank GlaxoSmithKline for CASE awards.

Notes and references

1 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.
2 M. J. Dawson, J. E. Farthing, P. S. Marshall, R. F. Middleton, M. J. O'Neill, A. Shuttleworth, C. Stylli, R. M. Tait, P. M. Taylor, H. G. Wildman, A. D. Buss, D. Langly and M. V. Hayes, *J. Antibiot.*, 1992, **45**, 639–647; J. D. Bergstrom, M. M. Kurz, D. J. Rew, A. M. Amend, J. D. Karkas, R. G. Bostedor, V. S. Bansal, C. Dufresne, F. L. Vanmiddlesworth, O. D. Hensens, J. M. Liesch, D. L. Zink, K. E. Wilson, J. Onishi, J. A. Milligan, G. Bills, L. Kaplan, M. N. Omstead, R. G. Jenkins, L. Huang, M. S. Meinz, L. Quinn, R. W. Burg, Y. L. Kong, S. Mochales, M. Mojena, I. Martin, F. Palaez, M. T. Diez and A. W. Alberts, *Proc. Natl. Acad. Sci. USA*, 1993, **90**, 80–84.
3 C. A. Jones, P. J. Sidebottom, R. J. P. Cannell, D. Noble and B. A. M. Rudd, *J. Antibiot.*, 1992, **45**, 1492–1498.
4 J. Kennedy, K. Auclair, S. G. Kendrew, C. Park, J. C. Vederas and C. R. Hutchinson, *Science*, 1999, **284**, 1368–1372.
5 Y. Abe, T. Suzuki, C. Ono, K. Iwamoto, M. Hosobuchi and H. Yoshikawa, *Mol. Genet. Genom.*, 2002, **267**, 636–646.
6 See ESI for details of HPLC analysis and characterisation of **2**.
7 I. Fujii, Y. Ono, H. Yada, K. Gomi, Y. Ebizuka and U. Sankawa, *Mol. Gen. Genet.*, 1996, **253**, 1–10; I. Shibuya, K. Tsuchiya, G. Tamura, T. Ishikawa and S. Hara, *Biosci. Biotechnol. Biochem.*, 1992, **56**, 1674–1675.
8 K. Auclair, A. Sutherland, J. Kennedy, D. J. Witter, J. P. Van den Heever, C. R. Hutchinson and J. C. Vederas, *J. Am. Chem. Soc.*, 2000, **122**, 11519–11520.
9 K. Auclair, J. Kennedy, C. R. Hutchinson and J. C. Vederas, *Bioorg. Med. Chem. Lett.*, 2001, **11**, 1527–1531.
10 A. Ranganathan, M. Timoney, M. Bycroft, J. Cortes, I. P. Thomas, B. Wilkinson, L. Kellenberger, U. Hanefeld, I. S. Galloway, J. Staunton and P. F. Leadlay, *Chem. Biol.*, 1999, **6**, 731–741.