Molecular BioSystems

Cite this: *Mol. BioSyst.*, 2011, **7**, 852–861

www.rsc.org/molecularbiosystems

PAPER

Cloning, sequencing and characterization of the biosynthetic gene cluster of sanglifehrin A, a potent cyclophilin inhibitor†

Xudong Qu,[a] Nan Jiang,[a] Fei Xu,[a] Lei Shao,[a] Gongli Tang,[a] Barrie Wilkinson[b] and Wen Liu*[a]

Received 15th October 2010, Accepted 16th November 2010
DOI: 10.1039/c0mb00234h

Sanglifehrin A (SFA), a potent cyclophilin inhibitor produced by *Streptomyces flaveolus* DSM 9954, bears a unique [5.5] spirolactam moiety conjugated with a 22-membered, highly functionalized macrolide through a linear carbon chain. SFA displays a diverse range of biological activities and offers significant therapeutic potential. However, the structural complexity of SFA poses a tremendous challenge for new analogue development *via* chemical synthesis. Based on a rational prediction of its biosynthetic origin, herein we report the cloning, sequencing and characterization of the gene cluster responsible for SFA biosynthesis. Analysis of the 92 776 bp contiguous DNA region reveals a mixed polyketide synthase (PKS)/non-ribosomal peptide synthetase (NRPS) pathway which includes a variety of unique features for unusual PKS and NRPS building block formation. Our findings suggest that SFA biosynthesis requires a crotonyl-CoA reductase/carboxylase (CCR) for generation of the putative unusual PKS starter unit (2*R*)-2-ethylmalonamyl-CoA, an iterative type I PKS for the putative atypical extender unit (2*S*)-2-(2-oxo-butyl)malonyl-CoA and a phenylalanine hydroxylase for the NRPS extender unit (2*S*)-*m*-tyrosine. A spontaneous ketalization of significant note, may trigger spirolactam formation in a stereo-selective manner. This study provides a framework for the application of combinatorial biosynthesis methods in order to expand the structural diversity of SFA.

Introduction

Sanglifehrin A (SFA), along with 20+ biosynthetic congeners, was isolated from *Streptomyces flaveolus* DSM 9954 (previously known as *Streptomyces* sp. A92-308110) as a result of screening microbial broth extracts for new cyclophilin A (CypA) ligands.[1,2] The highly complex structure of SFA is unprecedented (Fig. 1), featuring a [5.5] spirolactam moiety conjugated with a 22-membered macrolide by a linear carbon chain. The spirolactam system, with a quaternary carbon at its center, possesses seven stereogenic centers (SFA has seventeen in total), and the macrolide moiety displays rich structural diversity reflected in a 2-oxobutyl side-chain and a tripeptide moiety containing the unusual non-proteinogenic amino acids (2*S*)-*m*-tyrosine (*m*-Tyr, **1**) and (3*S*)-3-carboxypiperazine (Pip, **3**). Distinct from other known piperazic-acid-containing natural products isolated thus far, which exclusively have an amide bond linkage on the α-nitrogen, SFA utilizes the β-nitrogen atom of Pip for amide formation.

SFA exhibits potent activity against lymphocyte (*e.g.* T and B cells) proliferation[3] and viral infection (*e.g.* HCV and HIV),[2,4] and may be useful for promoting recovery after myocardial infarction.[5] The mechanism for immunosuppression remains unclear, however it is distinct from other immunophilin-binding drugs such as cyclosporin A (CsA), FK506 and rapamycin.[6] CsA binds to CypA to form an initial binary complex, whereas rapamycin and FK506 bind to a different receptor known as FKBP12. The CsA–CypA and FK506–FKBP complexes interact with the same target protein, calcineurin, to inhibit its Ser/Thr phosphatase activity thereby arresting T-cell proliferation in the G_0–G_1 stage, whilst the rapamycin–FKBP complex acts on mTOR.[7] Although SFA shows strong affinity for CypA binding (60-fold higher than that of CsA), the ultimate target of the SFA–CypA complex remains elusive.[6] SFA has no effect on the growth of human dendritic cells but inhibits the production of IL-12,[8] which plays a key role in regulating Th1 and NK cell proliferation, and in linking innate immunity with adaptive immunity, raising the possibility that its bioactivity is independent of SFA–CypA complex formation.

[a] *State Key Laboratory of Bio-organic and Natural Products Chemistry, Shanghai Institute of Organic Chemistry, Chinese Academy of Sciences, Shanghai, 200032, China. E-mail: wliu@mail.sioc.ac.cn; Fax: +86-21-64166128; Tel: +86-21-54925111*
[b] *Biotica Technology Ltd, Chesterford Research Park, Cambridge CB10 1XL, UK*

† Electronic supplementary information (ESI) available: bacterial strains and plasmids (Table S1); primers (Table S2); sequence aligment (Fig. S1–S6 and S9); chromosome walking (Fig. S2); genotype verification for the mutant strains (Fig. S7) and SDS-PAGE analysis of the purified protein (Fig. S8). See DOI: 10.1039/c0mb00234h

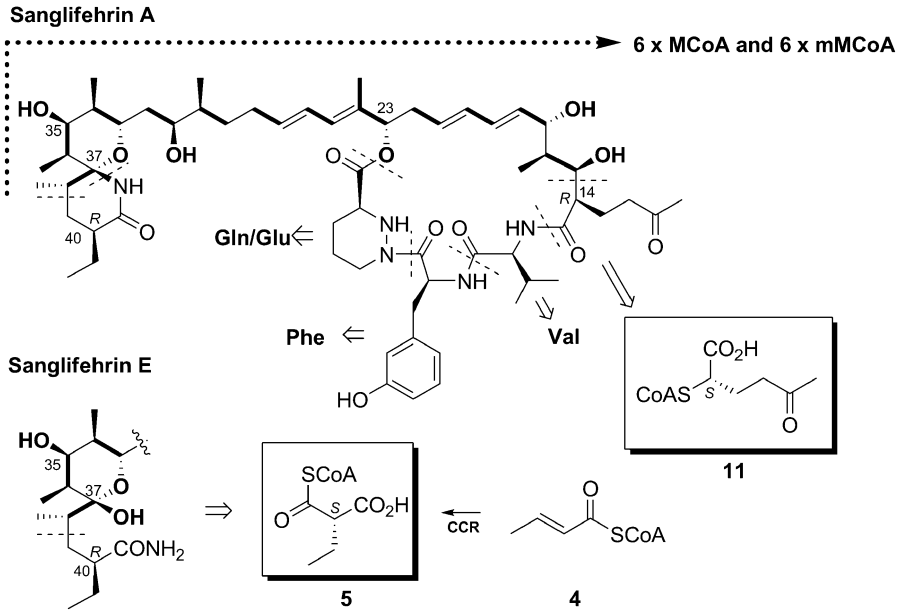

Fig. 1 Retro-biosynthetic analysis of SFA. MCoA: malonyl-CoA; mMCoA: (2*S*)-2-methylmalonyl-CoA; Gln/Glu: glutamine/glutamic acid; Phe: phenylalanine and Val: valine. Key building blocks are highlighted within boxes.

Therapeutic treatment with the immunosuppressant agents currently in clinical use can lead to severe side effects such as renal and central nervous system toxicity,[9] thus their use in certain immune dysfunction diseases is hindered. For example, calcineurin inhibition is the underlying cause of both the immunosuppressive and toxic effects of CsA and FK506. SFA, with a distinct mechanism of action, provides an alternative lead structure for the generation of new potent immunosuppressant drugs with lower toxicity. Since the complex architecture of SFA poses a tremendous challenge for chemical synthesis, combinatorial biosynthesis offers a promising way to access structural diversity. However, no *in vivo* or *in vitro* studies of SFA biosynthesis have been reported so far, a prerequisite for this approach.

Based on a rational prediction of its biosynthetic origin, we now report the specific access to the SFA biosynthetic machinery by cloning of the crotonyl-CoA reductase/carboxylase (CCR) gene from *S. flaveolus* DSM 9954. Characterization of the SFA gene cluster unveils a hybrid polyketide synthase (PKS)/nonribosomal peptide synthetase (NRPS) system for assembly of the SFA skeleton, and a number of new genes believed to be responsible for the synthesis of unusual building blocks for assembling the polyketide peptide hybrid backbone.

Results and discussion

Prediction of the biosynthetic origin

The chemical structure of SFA implies that its biosynthesis consists of two phases (Fig. 1). The first phase yields a carbon chain of variable saturation and alkyl substitution, and is typical of polyketide biosynthesis. In the second phase the presumptive polyketide intermediate can be further elongated by the addition of three amino acid residues. Such a process is characteristic of the polyketide/polypeptide backbones assembled by a range of hybrid modular PKS/NRPS systems.[10] Our initial attempts, utilizing degenerate primers to amplify PKS and NRPS encoding fragments as described previously,[11,12] yielded a group of distinct PCR products including 7 for PKS and 11 for NRPS. Unfortunately, gene disruption experiments using each of these failed to eliminate SFA production in *S. flaveolus* DSM 9954 (data not shown), suggesting that this SFA-producing strain, like many actinobacteria, possesses the genetic machinery for production of multiple PKS and NRPS based metabolites.

Based on the α-alkylation, β-hydroxylation and double bond geometry pattern of the polyketide chain, in conjunction with the orientation of amide bond formation, we propose that SFA assembly begins from the spirolactam moiety (Fig. 1). This analysis further suggests that polyketide biosynthesis may end with the incorporation of an unusual 6-carbon unit, (2*S*)-2-(2-oxobutyl)malonyl-CoA (**11**). Dissection of the spirolactam system to a linear chain allows prediction of a unique starter unit (2*R*)-2-ethylmalonamyl-CoA (**7**) for PKS priming. Structurally, both of these atypical building blocks are reminiscent of (2*S*)-2-ethylmalonyl-CoA (**5**), a somewhat uncommon PKS building block whose biogenesis typically arises in crotonyl-CoA (**4**), and requires α-carboxylation of **4** by CCR in a NADPH dependent manner.[13,14] The *ccr* gene is generally clustered with the genes required for producing polyketides with an α-ethyl branch such as salinosporamide[14] and monensin.[15] These combined observations indicate that screening for a *ccr* gene may provide selective access to the SFA biosynthetic machinery.

Cloning of the SFA biosynthetic gene cluster

We therefore designed a pair of degenerate PCR primers according to the conserved motifs of known CCRs in α-ethyl branched polyketide biosynthesis (Fig. S1†). Subsequent DNA amplification from the genomic DNA of *S. flaveolus* gave a

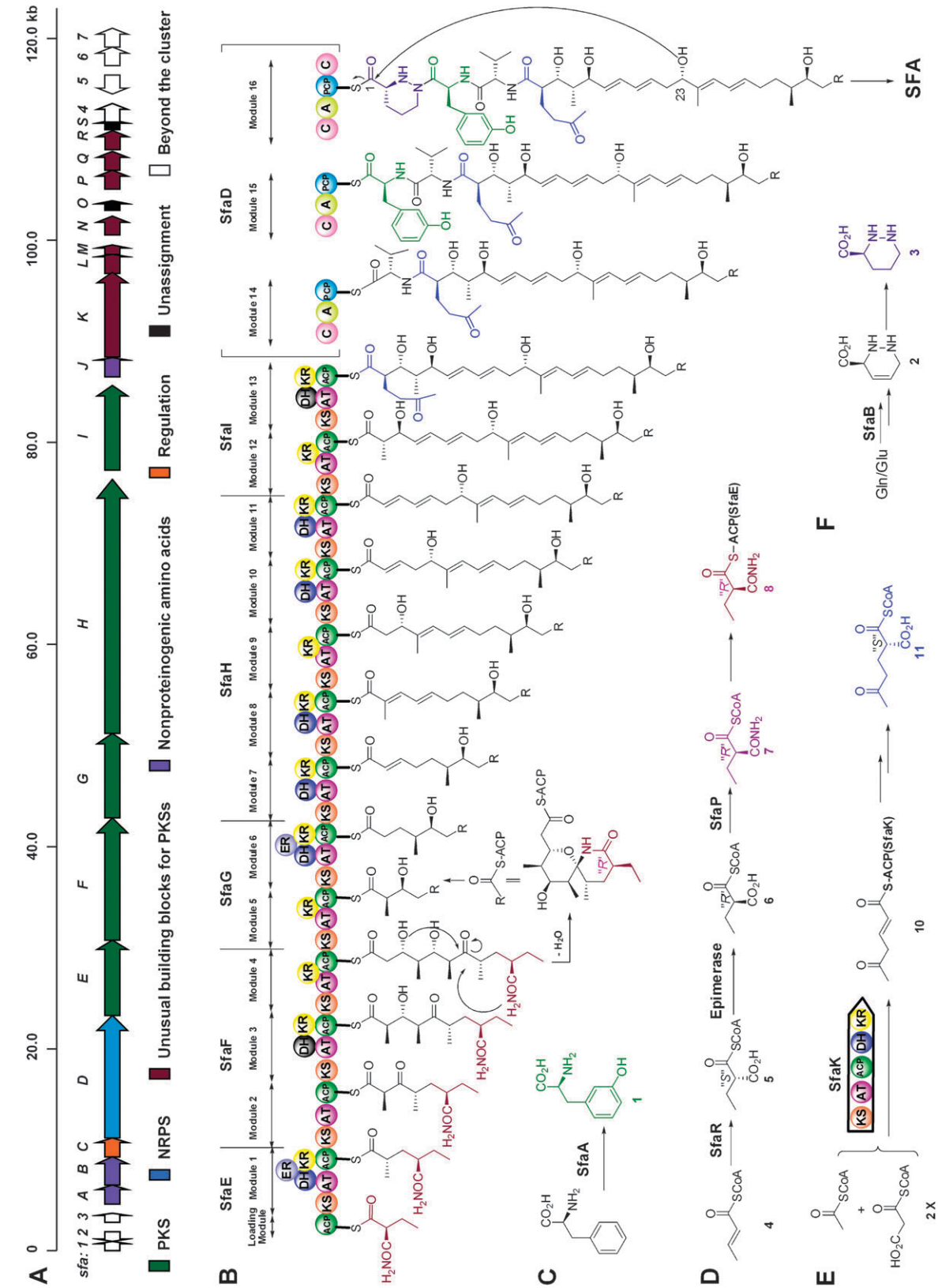

Fig. 2 Gene cluster and proposed biosynthetic pathway. A, Organization of the SFA biosynthetic genes, the deduced functions of which are labeled in color and summarized in Table 1. B, PKS/NRPS-directed skeleton assembly. C–F, Pathways for building block biosynthesis: C, (2*S*)-*m*-tyrosine (green); D, (2*R*)-2-ethylmalonamyl-CoA (cherry-red); E, (2*S*)-2-(2-oxobutyl)malonyl-CoA (blue); and F, (3*S*)-3-carboxypiperazine (purple).

single product band on an agarose gel with the expected size of 900 bp. The amplified DNA was cloned, and sequencing of multiple clones resulted in three distinct *ccr* gene fragments which were used as probes to screen a genomic library of *S. flaveolus* by colony hybridization. This led to the finding of a single genetic locus harboring both a *ccr* and type I modular PKS genes. Chromosome walking from this locus led to the identification of a 150 kb contiguous DNA region on the chromosome (Fig. S2†). To validate that this locus is responsible for SFA biosynthesis, an internal DNA fragment (subsequently shown to encompass parts of *sfaF* and *sfaH*, and all of *sfaG*) was replaced by the erythromycin antibiotic resistance gene *ermE* *via* double-crossover homologous recombination. The resulting mutant *TL3001* completely lost the ability to produce SFA (Fig. 3, trace VII), confirming the essential nature of this locus for SFA production. Cosmids pTL3102, pTL3104 and pTL3106 were then selected for sequencing, yielding 118 372 bp of contiguous sequence with an average GC content of 74.3%.

Bioinformatic analysis revealed 19 open reading frames (ORFs) (Fig. 2A and Table 1), which are proposed to constitute the *sfa* gene cluster based upon functional assignment of the deduced gene products. These ORFs have the same transcriptional orientation and may be translationally coupled as judged by the overlapping of many start and stop codons. Consistent with the structure of SFA, 5 modular type I PKS genes (*sfaEFGHI*) and 1 NRPS gene (*sfaD*) for assembly of SFA are present. Two genes (*sfaAB*) are suggested to be involved in biosynthesis of the non-proteinogenic amino acids **1** and **3**, and five genes (*sfaMKPQR*) may be required for supplying the unusual starter and extender unit building blocks **7** and **11** for polyketide biosynthesis. There is one putative regulatory gene (*sfaC*), and five additional genes (*sfaJ*, *L*, *O*, *sfaN* and *sfaS*) whose function in SFA biosynthesis could not be assigned based on sequence analysis alone. To identify the boundaries of the SFA gene cluster, we inactivated *orf3* and *orf4*. The resulting mutant strains *TL3002* and *TL3003* produced SFA at a titer comparable to that of the wild strain (Fig. 3, traces III and IV), indicating that the *sfa* gene cluster, spanning 92 776 bp, encompasses the 19 ORFs from *sfaA* to *sfaS*.

Polyketide assembly

The five genes *sfaE-I* encode a type I PKS system comprising of a mono-domain loading module and thirteen typical extension modules, which are organized co-linearly with their activities in the biosynthetic assembly process as shown in Fig. 2B. Each module consists of ketoreductase (KS), acyltransferase (AT), and acyl carrier protein (ACP) domains for carbon chain extension, and optionally the dehydratase (DH), enoylreductase (ER) and ketoreductase (KR) domains for reductive modifications to process the β-oxo functionality. This is in agreement with our biosynthetic prediction, which suggests the presence of unusual starter and extender units.

Analysis of the amino acid motifs conferring substrate selectivity[16–19] shows that the AT domains of modules 4, 6, 7, 9, 10 and 11 are specific for malonyl-CoA, while the other ATs (except for module 13) are specific for (2*S*)-2-methylmalonyl-CoA (Fig. S3†). This is consistent with the structure of SFA. In the motif C (GAGH) of the module 13 AT domain the usual Ser residue within YASH for (2*S*)-2-methylmalonyl-CoA specificity is replaced by Gly. This variation may broaden the substrate binding cavity,[17]

Table 1 Deduced functions of ORFs in the SFA biosynthetic gene cluster

Gene	Size[a]	Protein homologue and origin[b]	% Identity/ similarity	Proposed function
orf1	88	*SACE_5995* (YP_001108100); from *Saccharopolyspora erythraea* NRRL 2338	80/62	Transglycosylase-associated protein
orf2	146	*orf7* (AAA88561); from *S. fradiae*	72/69	Unknown protein
orf3	155	*SSEG_08631* (YP_002208581); from *S. sviceus* ATCC 29083	55/41	Conserved hypothetical protein
sfaA	276	*PAH* (NP_000268); from *Homo sapiens*	53/37	Phenylalanine hydroxylase
sfaB	446	*KtzI* (ABV56589); from *Kutzneria* sp. 744	63/47	Lysine/ornithine-*N*-monooxygenase
sfaC	231	*orf4* (ABV56600); from *Kutzneria* sp. 744	50/41	Negative transcriptional regulator
sfaD	3609	*OciB* (ABI26078); from *Planktothrix agardhii* NIVA-CYA 116	55/38	NRPS
sfaE	2217	*MerB* (ABJ97438); from *S. violaceusniger*	64/52	PKS
sfaF	4206	*ObsA* (AAS00419); from *Saccharopolyspora spinosa*	56/45	PKS
sfaG	3628	*NlmA1* (AAS46341); from *S. nanchangensis*	63/53	PKS
sfaH	8301	*FscC* (AAQ82564); from *S.* sp. *FR-008*	65/54	PKS
sfaI	3405	*NanA8* (AAP42874); from *S. nanchangensis*	60/49	PKS
sfaJ	388	*MupE* (AAM12917); from *Pseudomonas fluorescens*	38/27	Enoyl reductase
sfaK	1515	*MxaC* (AAK57187); from *Stigmatella aurantiaca*	47/34	Iterative PKS
sfaL	425	*PksE* (AAO25884); from *Micromonospora* sp. 046/Eco11	40/32	Acyl transferase
sfaM	239	*AAur_3586* (YP_949278); from *Arthrobacter aurescens* TC1	57/40	Short chain dehydrogenase
sfaN	340	*EhpO* (AAN40904); from *Pantoea agglomerans*	49/27	3-Oxoacyl-(ACP) synthase III (KASIII)
sfaO	84	*Snoa3* (CAA12019); from *S. nogalater*	60/40	ACP
sfaP	616	*OxyD* (AAZ78328); from *S. rimosus*	76/64	Asparagine synthetase homologue
sfaQ	259	*RifR* (AAG52991); from *Amycolatopsis mediterranei*	68/55	Thioesterase
sfaR	453	*RimJ* (AAR16523); from *S. diastaticus*	91/86	Crotonyl-CoA reductase/carboxylase
sfaS	71	*KtzJ* (ABV56590); from *Kutzneria* sp. 744	72/60	MbtH like protein
orf4	151	*orf7* (CAD18995); from *S. cattleya*	81/65	Hypothetical protein
orf5	152	*SSDG_02397* (YP_002199237); from *S. pristinaespiralis* ATCC 25486	63/53	Conserved hypothetical protein
orf6	144	*Strop_2904* (YP_001159721); from *Salinispora tropica* CNB-440	57/45	Hypothetical protein Strop_2904
orf7	177	*Strop_3504* (YP_001160313); from *Salinispora tropica* CNB-440	63/52	Putative transposase IS4

[a] Numbers are in amino acids. [b] NCBI accession numbers are given in parentheses.

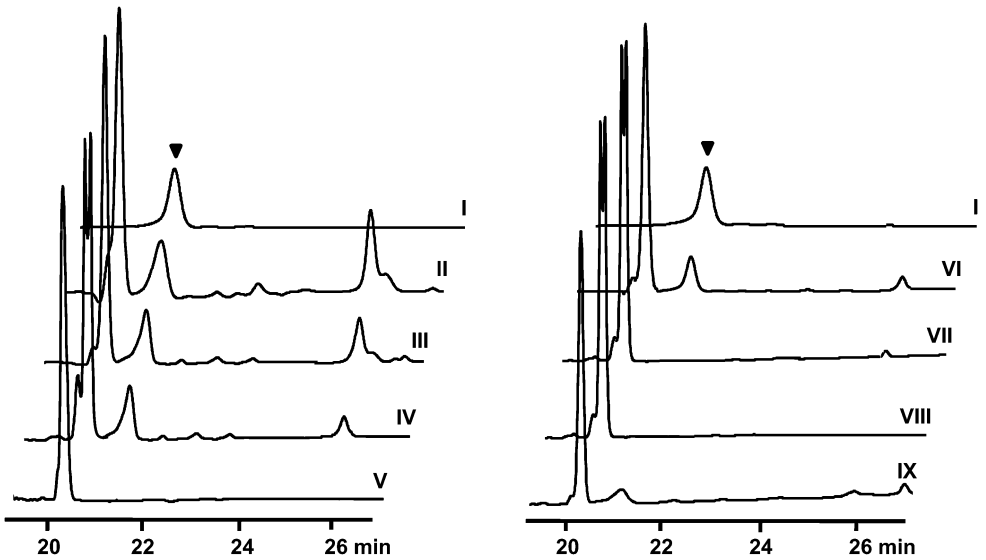

Fig. 3 HPLC analysis of SFA production. Authentic SFA (I); wild type strain *S. flaveolus* (II); mutant strain *TL3002* (Δ*orf3*, III); mutant strain *TL3003* (Δ*orf4*, IV); mutant strain *TL3007* (Δ*sfaK*, V); mutant strain *TL3005* (Δ*sfaR*, VI); mutant strain *TL3001* (Δ*sfaF*∼Δ*sfaH*, VII); mutant strain *TL3006* (Δ*sfaA*, VIII) and mutant strain *TL3006* fed with 2 mM DL-*m*-tyrosine (IX). Solid triangle indicates SFA.

supporting an unusual function such as the ability to utilize the putative extender unit **11** for the last polyketide chain extension. Sequence analysis of the 12 KR domains, according to a published model for predicting the stereochemistry of the secondary hydroxyl-group products,[20–22] allowed their classification into three groups, and is consistent with the stereochemistry of SFA (Fig. S4†). The DH domains in SfaF-M3 and SfaI-M13 are apparently inactive as both lack the essential His and Gly residues in the conserved catalytic motif (Fig. S5†).[23,24] Finally, it is notable that the PKS loading module housed in SfaE contains only an ACP domain, suggesting an "in *trans*" priming process for selection and loading of the starter unit. Taken together, SfaEFGHI appear sufficient for the assembly of a nascent ACP-tethered polyketide intermediate *via* 13 decarboxylative condensations.

Non-proteinogenic amino acid biosynthesis and peptide assembly

The gene products of *sfaA* and *sfaB* are likely to be involved in biosynthesis of the non-proteinogenic amino acids *m*-Tyr (**1**) and Pip (**3**) as shown in Fig. 2(C and F). SfaA belongs to the family of phenylalanine hydroxylases and may act on L-phenylalanine to introduce a hydroxyl group at the *meta*-position to give **1**.[25,26] Inactivation of *sfaA* by gene replacement with *aac*(*3*) *IV* (an apramycin resistant gene) completely abolished SFA production, which could then be restored by feeding of 2 mM DL-*m*-tyrosine into the resulting mutant strain *TL3006*, strongly supporting the above assumption (Fig. 3, VIII and IX). SfaB is highly homologous to various lysine/ornithine *N*-mono-oxygenases and may catalyze the flavin-dependent oxidation of an ornithine-like precursor to initiate ring closure and form the γ,δ-dehydropiperazoate precursor of **2** as has been proposed for KtzI during kutznerides biosynthesis.[27] Enamine to imine tautomerization of **2** may be followed by a reduction, resulting in **3** as the final building block required for incorporation into the SFA backbone.

sfaD encodes the only NRPS identified within the biosynthetic gene cluster, supporting its role in tripeptide biosynthesis as outlined in Fig. 2(A and B). SfaD consists of three typical NRPS modules, each of which comprises an adenylation (A) domain, a peptidyl carrier protein (PCP) domain and a condensation (C) domain. The third module contains an additional C domain at the C-terminus. To predict the substrates of SfaD the substrate specificity-conferring motif, which encompass ten amino acids for each A domain, were identified by sequence alignment with the A domain of GrsA (Fig. S6a and b†).[28–30] For SfaD-M14-A the sequence DAYWVGGTFK is typical for selection of L-valine as is found in SFA. For SfaD-M15-A the sequence DIYCLGAAFK does not correlate with any known A domain, as would be anticipated for incorporating the unusual **1** residue of SFA (consistent with above DL-*m*-tyrosine feeding experiment performed in the *sfaA* mutant strain *TL3006*). Finally, for SfaD-M16-A the sequence DVQFSAHVVK fits most closely with the specificity-conferring codes to A domains for L-proline residue activation, in agreement with the structural similarity between Pro and **3** residues.

These combined data are consistent with an assembly line process in which the PKS SfaEFGHI and NRPS SfaD are together responsible for biosynthesis of SFA. In analogy to (for example) rapamycin biosynthesis, release of the template chain along with intra-molecular lactonization is likely to be catalyzed by the C domain located at the C-terminus of SfaD, rather than by the more commonly found thioesterase (TE) domains.[31]

PKS starter unit biosynthesis

Our proposal for SFA biosynthesis requires (2*R*)-2-ethyl-malonamyl-CoA (**7**) as the starter unit for PKS assembly (Fig. 2D). The most logical biosynthetic scheme for (2*R*)-2-ethylmalonamyl-CoA (**7**) starts with crotonyl-CoA (**4**) in analogy to (2*S*)-2-ethylmalonyl-CoA (**5**) biosynthesis. As discussed above, *sfaR* encodes a putative CCR capable of the

reductive-carboxylation of **4** to give **5** in an NADPH dependent manner.[32] Epimerization of **5** into (2*R*)-2-ethylmalonyl-CoA (**6**) can be catalyzed by either a pathway specific epimerase or a 2-methylmalonyl-CoA epimerase such as that which has been characterized from the acetate assimilation pathway of *Rhodobacter sphaeroides*.[33] As no putative epimerase gene could be identified in the SFA gene cluster, we presume the latter option to be the most probable. Amidation of **6** to form **7** could be catalyzed by SfaP, an asparagine synthetase homologue. This is supported by the observation that the highly homologous protein OxyD performs the conversion of malonyl-CoA into malonamyl-CoA during oxytetracycline biosynthesis.[34] Finally, a "in *trans*" process could be adopted to prime the starter unit **7** onto the C-terminal ACP module of SfaE (Fig. 2B).

In an attempt to validate our hypothesis for **7** biosynthesis we inactivated *sfaR* by gene deletion. The resultant mutant *TL3005* produced SFA at significantly reduced levels (approx. 33%) *versus* the wild-type strain (Fig. 3, trace VI). This fact suggests that the additional *ccr* homologues identified during degenerate PCR amplification may participate in supplying **5** (>60%) for SFA biosynthesis. A similar result has been reported for monensin biosynthesis.[15] To confirm its function as a CCR we then expressed *sfaR* in *E. coli* BL21 (DE_3) and assayed the purified SfaR *in vitro* (Fig. 4B). In the presence of biocarbonate and NADPH, **4** was converted rapidly into the carboxylated product **5** along with a small quantity of the reduced product *n*-butyryl-CoA **9** (approx. 13%). When biocarbonate was removed from the assay system, **4** was totally converted into **9**. These data clearly demonstrated that SfaR is a CCR capable of generating 2-ethylmalonyl-CoA **5** from crotonyl-CoA **4** *via* a reductive carboxylation mechanism

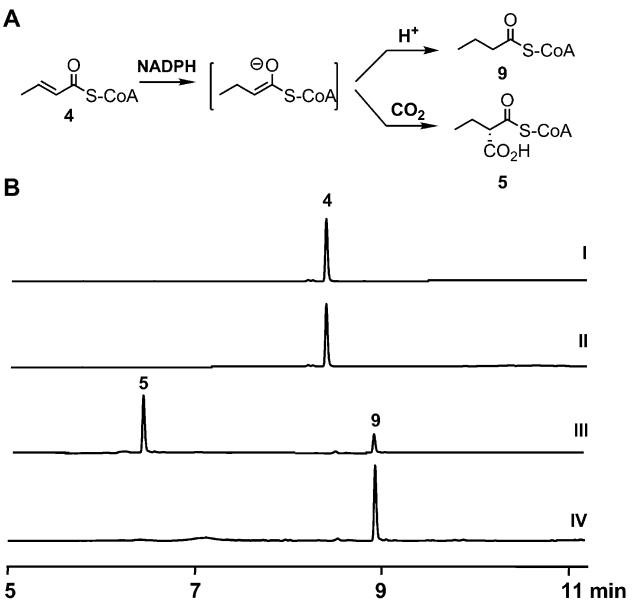

Fig. 4 Characterization of SfaR as a CCR *in vitro*. A, mechanism of SfaR- catalyzed reaction. B, product examination by HPLC. Standard crotonyl-CoA (I); assay system contains crotonyl-CoA, $NaHCO_3$, NADPH, but no SfaR (II); assay system contains crotonyl-CoA, $NaHCO_3$, NADPH and SfaR (III); and assay system contains crotonyl-CoA, NADPH and SfaR (IV).

(Fig. 4A).[32] By extension to other studied examples we assume that the product of SfaR is (2*S*)-2-ethylmalonyl-CoA.[35]

Spirolactam formation

The spirolactam moiety of SFA is unique among natural products, and the biosynthetic mechanism leading to its formation is of significant interest. Experimental evidence has revealed a hemiketal intermediate during formation of the avermectin spiroketal moiety,[36] and a similar stepwise pathway is likely involved in SFA biosynthesis. During the total synthesis of SFA, formation of the spirolactam moiety was accomplished by spontaneous ketalization triggered by acid deprotection of a linear polyketide chain which is similar to the theoretical PKS bound intermediate of the module 4- ACP domain in SfaF (**12**, Fig. 2B and 5).[37,38]

Sequencing of the SFA gene cluster revealed no candidate genes for functional involvement in formation of the spirolactam moiety, and we speculate that a spontaneous process may occur with a PKS-bound pentaketide intermediate as shown in Fig. 5. First, the keto group at C37 is attacked by the C33 hydroxyl group to form the hemiketal-containing intermediate **13**, which can exist in equilibrium with the oxonium intermediate **14**. Chain elongation of the pentaketide moiety **13** with no further modification would lead to formation of the naturally occurring SFA congener sanglifehrin E (Fig. 1). Intra-molecular cyclization of the C41 amide with the oxonium ion intermediate may then form **15**, and subsequent chain elongation leads to the formation of SFA. Interestingly, this process appears to be stereoselective and computational analysis has revealed an extra hydrogen bond interaction between the amine and the hydroxyl group at C35 in the "pre-*S*" hemiketal.[38] This presumably leads to pre-organization so that nucleophilic attack by the amine is favoured from the *Si* face to form the product with *S*-configuration at C37. Similar results were also achieved by organic synthesis where the ratio of epimers formed is 7 : 1 (*S* : *R*).[38]

Formation of the putative 6-carbon extender unit

Based on analysis of the SFA structure and PKS architecture it appears that the final round of polyketide chain assembly requires an unusual extender unit, which we suggest to be (2*S*)-2-(2-oxobutyl)malonyl-CoA (**11**). Analysis of the SFA gene cluster led to the identification of *sfaK*, encoding an atypical PKS as a probable candidate for biosynthesis of an **11** precursor (Fig. 2, A and E). SfaK consists of four domains of which the N-terminal KS and AT domains are characteristic of known type I PKSs, and the remaining two domains have no significant sequence homologies to any proteins of known function. Based on secondary structural modeling by using the software Phyre[39] and web-based software (Motif Scan, http://hits.isb-sib.ch/cgi-bin/PFSCAN)[40] we predict that the third domain encompassing ~74 amino acids is in fact a unique ACP that contains a catalytic motif GFDSL (the central Ser residue is for phosphopantetheine attachment) (Fig. S5c†), and that the fourth domain of ~220 amino acids at the C-terminus is a KR and DH like di-domain (Fig. 2E, S4 and S9†). This domain architecture is similar to that found for the iterative PKSs involved in enediyne biosynthesis, *e.g.* SgcE which is required for antibiotic

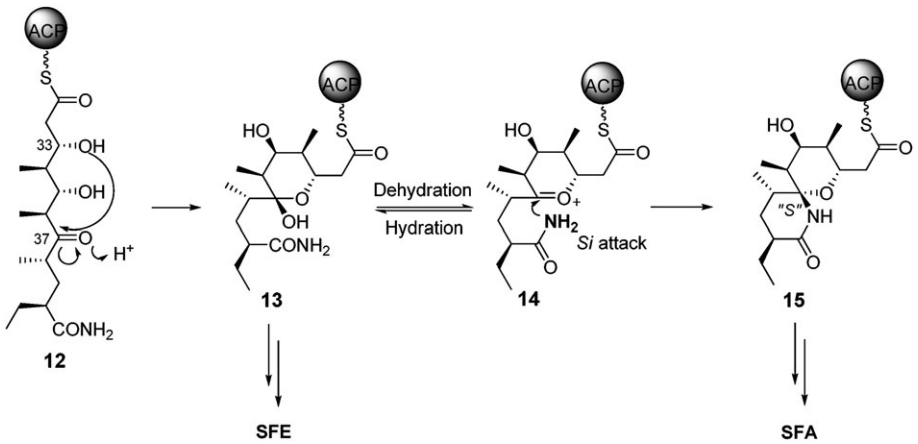

Fig. 5 Proposed cyclization mechanism for spirolactam formation.

C-1027 biosynthesis.[41] In analogy to SgcE, SfaK can be envisaged to catalyze the assembly of a nascent linear unsaturated triketide from acetyl-CoA and two malonyl-CoA units. The conserved motif for malonyl-CoA selection in the SfaK–AT domain supports this prediction (Fig. S3†). How this generates the required α,β-olefin and γ-keto groups is unclear, but could involve some crosstalk with enzymes from primary metabolism in fatty acid biosynthesis.

To convert the putative SfaK product to **11** requires the reductive carboxylation of an α,β-unsaturated crotonyl-CoA like substrate (it is also not clear how this unit might be transferred to CoA or ACP) in a process similar to that catalyzed by SalG during (2*S*)-2-(2-chloroethyl)malonyl-CoA biosynthesis.[14] Giving that the CCR gene *sfaR* within the *sfa* gene cluster is not essential for SFA biosynthesis as shown above, we assume that the reductive carboxylation reaction required to generate **11** can be performed by one of the alternative *ccr* homologues identified in the genome. To determine whether SfaK is essential for SFA biosynthesis, *sfak* was removed by in frame deletion. SFA production was totally abolished in the resultant mutant strain *TL3007* (Fig. 3, trace V), confirming its essential requirement for SFA biosynthesis and consistent with our suggested role in SFA biosynthesis.

Conclusion

We have uncovered the biosynthetic pathway of SFA by cloning, sequencing and characterization of the *sfa* gene cluster in *S. flaveolus* DSM 9954. SFA biosynthesis occurs by a mixed PKS–NRPS system in a linear manner. The discovery of many new enzymes, putatively involved in the synthesis of unusual PKS and NRPS building blocks and formation of structurally unique moieties, presents an excellent opportunity to investigate their new chemistry and enzymology. SFA exhibits a diverse range of biological effects with significant potential of therapeutic values. Structurally, SFA is also a hybrid polyketide/non-ribosomal peptide product of particular novelty and complexity, and due to this the use of SFA as a drug lead for optimization through semi-synthetic methods is technically demanding. The availability of the SFA biosynthetic gene cluster now provides the genetic basis for application of biosynthetic medicinal chemistry methods for its optimization.

Materials and methods

Bacterial strains, plasmids, and reagents

Bacterial strains and plasmids used in this study are summarized in Table S1.† Biochemicals, chemicals, media, restriction enzymes, and other molecular biological reagents were obtained from standard commercial sources.

General biological methods and DNA sequencing

DNA isolation and manipulation in *E. coli* and *Streptomyces* were carried out according to standard methods.[42,43] For Southern analysis, digoxigen labeling of DNA probes, hybridization, and detection were performed according to the protocols provided by the manufacturer (Roche Diagnostics Corp., Indianapolis, IN, USA). T4 polymerase (Dalian TaKaRa Biotechnology Co. Ltd., Dalian, China) was used to generate blunt ends. PCR amplification was carried out on an authorized cycler Eppendorf AG 22331 (Hamburg, Germany) using either Taq DNA polymerase (Dingguo Co. Ltd., China) for routine genotype verification or PrimeSTAR HS DNA polymerase (Dalian TaKaRa Biotechnology Co. Ltd., Dalian, China) for high fidelity amplification. Primer synthesis and DNA sequencing were performed at Shanghai Invitrogen Biotech Co., Ltd. (Shanghai, China) and the Chinese National Human Genome Center (Shanghai, China).

Sequence analysis

ORFs were identified using the FramePlot 4.0b program (http://nocardia.nih.go.jp/fp4/). The corresponding deduced proteins were compared with other known proteins in the databases by using available BLAST methods (http://www.ncbi.nlm.nih.gov/blast/). Amino acid sequence alignments were performed by using the strap (http://www.bioinformatics.org/strap/).

Fermentation, analysis and isolation of SFA

S. flaveolus DSM 9954 (German Collection of Microorganisms and Cell Cultures, Braunschweig Germany) and recombinant streptomyces strains generated in this study were routinely grown at 28 °C on ISP-4 agar medium or TSB liquid medium.[42] ISP-4 agar medium was used to prepare *S. flaveolus* spores. For

fermentation of *S. flaveolus* DSM 9954 and recombinant strains, frozen spore stocks (0.5 mL) were inoculated into 500 mL-baffled flasks containing 50 mL seed medium (4% glycerol, 1% soy peptone, 2.1% malt extract; pH adjusted to 7.0 with 1.0 M NaOH) and shaken at 250 rpm and 27 °C. After 24 h an aliquot of seed culture (3 mL) was used to inoculate a 500 mL-production flask containing 50 mL of production medium (2% toasted soy flour, 4% glycerol, 1.95% MES hydrate, pH adjusted to 6.8 with 10 M NaOH) and shaken at 250 rpm and 24 °C for 5 days. For feeding experiment, DL-*m*-tyrosine (Sigma-aldrich Co.) was added into the 2 days old production broth with a final concentration of 2 mM and fermentation was continued for a further 3 days. Ethyl acetate (50 mL) and fermentation broth (50 mL) were stirred rigorously under sonication for 20 min and then clarified by centrifugation. The upper organic phase was removed and the solvent removed under reduced pressure and the residue dissolved in methanol (1 mL). Analysis by HPLC-ESI-MS was performed on a Thermo Fisher LTQ Fleet ESI mass spectrometer (Thermo Fisher Scientific Inc., USA) coupled to an Agilent 1200 HPLC (Agilent Technologies Inc., USA). Chromatography was achieved over an Aglient Zorbax column (SB-C18, 5 μm, 4.6 × 250 mm) eluting with a flow rate of 1 mL min^{-1} over a 30 min gradient as follows: T = 0, 40% B; T = 5, 40% B; T = 25, 90% B; T = 27, 90% B; T = 29, 40% B, and T = 30, 40% B. A, H_2O + 0.1% formic acid; B, CH_3CN + 0.1% formic acid. UV analysis was made at 242 nm. The retention time of SFA standard was approximately 21.3 ± 0.2 min. For HPLC-UV analysis the same methods and conditions as above were used but the formic acid was replaced with 0.05% TFA.

Cloning and isolation of the SFA biosynthetic gene cluster

A genomic library of *S. flaveolus* DSM 9954 was constructed in SuperCos1 according to a method described previously.[11] The 0.9 kb DNA probes were amplified from the *S. flaveolus* genomic DNA by using the degenerate primers of *ccr*: CCR-for (5′-AG GAA TTC ATG GCC TCC KCS RTS AAC TAC AAY-3′, *Eco*RI site is underlined) and CCR-rev (5′-TC GGA TCC GAT GAT GCG CTT SWS BKD CAT CCA -3′, *Bam*HI site is underlined). The genomic library was screened by colony hybridization with the mixture of three *ccr* fragments as the probe (probe 1). The *ccr* positive cosmids were further screened by PCR with degenerate PKS primers[11] to identify those which contain both *ccr* and *pks* genes. *Bam*HI was then used to digest the resulting cosmids to generate restriction maps. The 6.4 kb BamHI fragment excised from the upstream end of pTL3102 (probe 4), 1.4 kb BamHI fragment derived from the upstream end of pTL3104 (probe 5) and 4.8 kb BamHI fragment from the upstream end of pTL3106 (probe 6), were used as probes for chromosomal walking to isolate overlapping cosmids covering the entire gene cluster (Fig. S2†).

Genetic manipulation of *S. flaveolus*

Genetic manipulation of *S. flaveolus* was carried out using standard literature methods[42] with minor modifications. In brief, approximately 10^9 *S. flaveolus* spores were suspended in TES (500 μL, 0.05 M, pH 8.0) and heat shocked at 50 °C for 10 min. TSB (500 μL) was then added and the whole incubated at 37 °C for 3–5 h. Spores were then recovered by centrifugation and resuspended in LB medium (2 mL) to be used as recipients. *E. coli* S17-1 containing the respective plasmid was grown in LB medium with appropriate antibiotics to an OD_{600} of 0.4–0.6. Cells from 50 mL of culture were recovered by centrifugation, washed twice with 20 mL LB medium, and resuspended in LB medium (1 mL) for use as the donor. Donor (100 μL) and recipients (100 μL) were mixed and distributed onto MS plates supplemented with 10 mM $MgCl_2$. The plates were incubated at 30 °C for 10–16 h then overlaid with 1 mL water of containing nalidixic acid to select against *E. coli* and either apramycin (Am) or erythromycin (Em) or kanamycin (Km) to select for *S. flaveolus* exconjugants (the final concentration of antibiotics is 50 μg mL^{-1}). Incubation was continued at 30 °C until exconjugants appeared.

Generation of the gene replacement vector pKC5201

A 900 bp DNA fragment containing the neomycin resistance cassette was amplified from SuperCos1 using primers neo-for (5′-AA GAG CTC GCA GAA ACG GTG CTG ACC-3′, *Sac*I site is underlined) and neo-rev (5′-AA GAG CTC ATA GAA GGC GGC GGT GGA-3′, *Sac*I site is underlined) and further cloned into the pKC1139 *Sac*I site for replacing the *aac(3)IV* resistance gene to generate pKC5201. The neomycin resistant gene in the resulted plasmid pKC5201 has the same transcription direction to *aac(3)IV* of pKC1139.

Deletion of the *pks* (*sfaK* ~ *sfaH*)

A 1.8 kb *Eco*RI/*Stu*I fragment containing the *ermE* resistance cassette from pAGe-1 was cloned into the *Eco*RI/*Eco*RV site of pSP72 to yield TL3201. Cosmid pTL3104 was digested with *Bam*HI to afford a 9.5 kb DNA fragment containing vector, which further self-ligates to yield pTL3202. The 1.8 kb *Bgl*II fragment containing the *ermE* resistance gene from pTL3201 was cloned into the *Bam*HI site of pTL3202 to generate plasmid pTL3203 which contains the same coding direction for the *ermE* and *pks* genes. Finally, the 4.5 kb *Eco*RI fragment was moved from pTL3203 into the *Eco*RI site of pKC1139 to create pTL3111, which was subsequently used to replace part of the *pks* locus with the *ermE* gene.

Inactivation of *orf3*

Primers orf3L-for (5′-AA GGA TCC GGA TGA CCT GTG ACA GCG AAC CTT G-3′, *Bam*HI site is underlined) and orf3L-rev (5′-AA CAT ATG GCT TGG ACA TGA AGT ATC CCC TCA T-3′, *Nde*I site is underlined) were used to amplify a 1.5 kb fragment from pTL3106 containing *orf3*. The resultant blunt-ended fragment was cloned into the *Stu*I site of pANT841 to generate pTL3206 as the left arm for gene replacement. The right arm for gene replacement was achieved by cloning a 1.5 kb blunt-ended DNA fragment, amplified from pTL3106 with primers orf3R-for (5′-AA CAT ATG GCG GGC GAG ATG CGG GAT GTC GAG G-3′, *Nde*I site is underlined) and orf3R-rev (5′-AA GAA TTC GCG GTA CAG GTT GGC GAA CCG GTC G-3′, *Eco*RI site is underlined), into the *Eco*RV site of pGEM 5zf to yield

pTL3207. Finally, a three part ligation was achieved between the 1.5 kb *Bam*HI-*Nde*I fragment of pTL3206, the 1.5 kb *Nde*I-*Eco*RI fragment from pTL3207, and *Bam*HI-*Eco*RI digested pKC1139 to yield pTL3114. This was the vector used for *orf3* in frame deletion.

Inactivation of *sfaA*

Primers 201-2L (5′-GTG GAA ATC GGC TCG GGC GCG CCC GAA TTA ACC GCG TCG ATT CCG GGG ATC CGT CGA CC-3′) and 201-2R (5′-AAT GGA TGT ATC GTC GCA GGA CGC CCA GAA TTC ACC TGC TGT AGG CTG GAG CTG CTT C-3′) were used to amplify the 1.3 kb *aac(3)IV-oriT* cassette from pIJ773. pTL3205 was constructed *via* λ_{RED}-mediated PCR targeting[44] in order to replace the *sfaA* gene of pTL3106 with the *aac(3)-oriT* cassette. The 8 kb *Bgl*II fragment was then transferred from pTL3205 into the *Bam*HI site of pKC5201 to afford plasmid pTL3113, which was used for *sfaA* gene replacement.

Inactivation of *sfaK*

Similar to in frame deletion of *orf3*, the left arm is amplified from pTL3102 by primers sfaKL-for (5′-AT GAA TTC GTG GCG GTC GTC GGC ATG GCC TG-3′,with *Eco*RI underlined) and sfaKL-rev (5′-AT GGA TCC CCT CGG TGC CGC GGA ACA GGA GC-3, with *Bam*HI site underlined) and cloned into the *Eco*RV site of pGEM 5zf to achieve pTL3208. Right arm amplified from pTL3102 by primers sfaKR-for (5′-AT GGA TCC GGC GAT GAC CGA CTG GCT CGC CAC-3′, with *Bam*HI underlined) and sfaKR-rev (5′-AT AAG CTT GGT CCC CGA CGG TGA GCA GCA CC-3′, with *Hind*III underlined) was cloned into the *Eco*RV site of pGEM 5zf to generate pTL3209. The *Eco*RI–*Bam*HI fragment from pTL3208 and the *Bam*HI–*Hin*dIII fragment from pTL3209 were recovered and then co-ligated into the *Eco*RI–*Hin*dIII site of pKC1139, yielding pTL3113 for in frame deletion of *sfaK*.

Inactivation of *sfaR*

The 4.7 kb *Bgl*II-*Kpn*I fragment from pTL3102 was cloned into the same sites of pSP72 to create pTL3210. The internal 674 bp fragment of pTL3210 was then removed by *Eco*R72I digestion and the rest part further self-ligated to yield pTL3211. The *Bgl*II-*Hin*dIII fragment of pTL3211 was further cloned into *Bam*HI-*Hin*dIII sites of pKC1139 to afford plasmid pTL3116, which is used for *sfaR* gene replacement.

Inactivation of *orf4*

An internal 300 bp DNA fragment of *orf4* was removed from pTL3210 by *Cla*I-*Eco*RI double digestion. The remaining major fragment of pTL3210 was treated with T4 DNA polymerase to generate blunt ends and allowed to self-ligate to generate pTL3212. The *Bgl*II-*Hin*dIII fragment from pTL3212 was moved into *Bam*HI-*Hin*dIII digested pKC1139 to yield pTL3117. This was used for *orf4* gene deletion.

Generation of mutant strains

Each exconjugant colony was streaked onto ISP-4 agar media containing 50 μg mL^{-1} Am and incubated at 37 °C for several days to induce the single crossover. A single colony was then inoculated into 3 mL TSB media without antibiotics and incubated in a shaker at 37 °C for 4 days. This was used to innoculate further TSB media (3 μL, 0.1% v/v) which was incubated under same condition for 4 days. This step was repeated 5 times to induce double crossover. 50 μL of the resulting broths were then streaked on ISP-4 ager media containing 50 μg mL^{-1} Em for pTL3111, or 50 μg mL^{-1} Am for pTL3113, or without any antibiotics for pTL3114–pTL3117 and incubated at 37 °C until colonies were seen. Verification of antibiotic resistance phenotype was performed by replica streaking of colonies onto plates containing different antibiotics and checked resistant phenotype. For pTL3111, the double crossover mutant should be Em^R and Am^S. For pTL3113 it should be Am^R and Km^S. For pTL3114–pTL3117 the Am^S phenotype indicated loss of the vector backbone and could be either wild type strain or inframe deletion mutants. Genotype verification by PCR was performed to verify the mutants (see ESI†).

Synthesis of crotonyl-CoA

Crotonic acid (3 mg, 35 μmol, 3.5 eq., Aldrich Inc.), PyBOP (8.4 mg, 16 μmol, 1.6 eq., GL Biochem Shanghai Ltd.) and K_2CO_3 (5.6 mg, 40 μmol, 4 eq., China National Medicines Co. Ltd.) were dissolved in anhydrous THF (3 mL, China National Medicines Co. Ltd.) using vigorous stirring under argon at room temperature for at least 10 min. A solution of CoASH (7.67 mg, 10 μmol, 1 eq., Sigma Inc.) was then added slowly and the reaction was stirred at room temperature for 1 h. The reaction mixture was purified by Jasco LC-NetII HPLC (Jasco, Inc. Japan) and an Aglient Zorbax column (SB-C18, 5 μm, 4.6 × 250 mm) monitoring at a wavelength of 260 nm with condition of 5–40% B (CH_3CN) and A (Water, 10 mM NH_4OAc) elution in 30 min and structurally determined by ESI-MS (Thermo Fisher LTQ Fleet ESI and Agilent 1200 HPLC) for crotonyl CoA 835.90 ($M + H^+$).

Overexpression, purification and biological assay of SfaR

The *sfaR* gene was amplified by PCR from cosmid TL3102 using primers (SfaR-for: 5′-TC GAA TTC CAT ATG AAC GAG ATA CTC AGC G-3′ and SfaR-rev: 5′-TCA AAG CTT ATC AGA CCG GCT CTG TCT GC-3′, restriction sites *Eco*RI, *Nde*I and *Hin*dIII are underlined). This was then cloned into the *Nde*I and *Hin*dIII sites of His_6-tagged vector pET28a and the resulting plasmid pTL3213 was transferred into *E. coli* BL21 (DE_3) for overexpression. SfaR with a molecular weight of ~50 kD was expressed at 25 °C 250 rpm for 10 h under 100 μM IPTG induction (IPTG was added when OD_{600} reached 0.6) and purified to homogeneity by Ni-NTA affinity chromatography (Fig. S9†). The elutent containing SfaR was exchanged into stocking buffer (50 mM Tris-HCl, 100 mM NaCl, and 10% (v/v) glycerol, pH = 8.0) by PD-10 (GE healthcare), concentrated to 10 mg mL^{-1} by Vivaspin (GE healthcare) and stored at −80 °C. Protein concentration was determined by standard Bradford method.[45] The standard assay system consists of 1 mM NADPH, 60 mM $NaHCO_3$, 20 μM SfaR and 500 μM substrate in Tris-Cl buffer (75 mM, pH = 8.0) with

incubated at 30 °C. In order to further study the reduction step, the same assay was used except $NaHCO_3$ was omitted. The assay system are analyzed by HPLC-ESIMS with a flow rate of 1 mL min^{-1} over a 30 min gradient as follows T = 0, 5% B; T = 30, 40% B (A: 10 mM NH_4OAc in water; B: acetonitrile). (2*S*)-2-ethylmalonyl-CoA ($M + H^+$: 881.90) and butyryl-CoA ($M + H^+$: 837.86) were eluted at 6.6 min and 9.0 min, respectively.

Nucleotide sequence accession number

The nucleotide sequence reported in this study is available in the GenBank database under accession no. FJ809786.

Acknowledgements

This work was supported in part by grants from NNSF (30900021, 20902102, 20832009, 90713012 and 20921091), MST (2009ZX09501-008), "973 program" (2010CB833200), CAS (KJCX2-YW-H08 and KSCX2-YW-G-06), and STCSM (09QH1402700) of China. We thank Nigel Coates, Matthew Gregory, Teresa Foster, Steven Moss, Mahmood Piraee, Tony Warneck and Ming Zhang at Biotica Technology Ltd. for help with fermentation methods and analysis of strains.

References

1 J. J. Sanglier, V. Quesniaux, T. Fehr, H. Hofmann, M. Mahnke, K. Memmert, W. Schuler, G. Zenke, L. Gschwind, C. Maurer and W. Schilling, *J. Antibiot*., 1999, **52**, 466–473.
2 B. Weidmann, K. Zimmermann and K. Lin, *Intern. Patent. Appl*., WO/2006/138507 , 2006.
3 C. Hartel, P. Iblher, A. Puzik, K. Wortmeier, B. Ebel, C. Schultz and M. Muller-Steinhardt, *Scand. J. Immunol*., 2006, **63**, 26–34.
4 E. Sokolskaja, D. M. Sayah and J. Luban, *J. Virol*., 2004, **78**, 12800–12808.
5 S. Javadov, M. Karmazyn and N. Escobales, *J. Pharmacol. Exp. Ther*., 2009, **330**, 670–678.
6 J. Kallen, R. Sedrani, G. Zenke and J. Wagner, *J. Biol. Chem*., 2005, **280**, 21965–21971.
7 M. T. Ivery, *Med. Res. Rev*., 2000, **20**, 452–484.
8 C. Steinschulte, T. Taner, A. W. Thomson, G. Bein and H. Hackstein, *J. Immunol*., 2003, **171**, 542–546.
9 L. W. Miller, *Circulation*, 1996, **94**, 1209–1211.
10 R. McDaniel, M. Welch and C. R. Hutchinson, *Chem. Rev*., 2005, **105**, 543–558.
11 X. Jia, Z. Tian, L. Shao, X. Qu, Q. Zhao, J. Tang, G. Tang and W. Liu, *Chem. Biol*., 2006, **13**, 575–585.
12 L. Li, W. Deng, J. Song, W. Ding, Q. Zhao, C. Peng, W. Song, G. Tang and W. Liu, *J. Bacteriol*., 2008, **190**, 251–263.
13 T. J. Erb, I. A. Berg, V. Brecht, M. Muller, G. Fuchs and B. E. Alber, *Proc. Natl. Acad. Sci. U. S. A*., 2007, **104**, 10631–10636.
14 Y. Liu, C. Hazzard, A. S. Eustaquio, K. A. Reynolds and B. S. Moore, *J. Am. Chem. Soc*., 2009, **131**, 10376–10377.
15 H. Liu and K. A. Reynolds, *J. Bacteriol*., 1999, **181**, 6806–6813.
16 F. H. Stephen, F. A. Jesus, M. Istvan, S. Torsten, K. Lake Ee, K. Ariane, F. A. M. Andrew, S. G. Ian, S. James and F. L. Peter, *FEBS Lett*., 1995, **374**, 246–248.
17 C. D. Reeves, S. Murli, G. W. Ashley, M. Piagentini, C. R. Hutchinson and R. McDaniel, *Biochemistry*, 2001, **40**, 15464–15470.
18 F. Del Vecchio, H. Petkovic, S. G. Kendrew, L. Low, B. Wilkinson, R. Lill, J. Cortés, B. A. M. Rudd, J. Staunton and P. F. Leadlay, *J. Ind. Microbiol. Biotechnol*., 2003, **30**, 489–494.
19 H. Petkovic, R. E. Lill, R. M. Sheridan, B. Wilkinson, E. L. McCormick, H. A. McArthur, J. Staunton, P. F. Leadlay and S. G. Kendrew, *J. Antibiot*., 2003, **56**, 543–551.
20 R. Reid, M. Piagentini, E. Rodriguez, G. Ashley, N. Viswanathan, J. Carney, D. V. Santi, C. R. Hutchinson and R. McDaniel, *Biochemistry*, 2003, **42**, 72–79.
21 P. Caffrey, *ChemBioChem*, 2003, **4**, 654–657.
22 A. T. Keatinge-Clay, *Chem. Biol*., 2007, **14**, 898–908.
23 D. J. Bevitt, J. Cortes, S. F. Haydock and P. F. Leadlay, *Eur. J. Biochem*., 1992, **204**, 39–49.
24 X. Ruan, D. Stassi, S. A. Lax and L. Katz, *Gene*, 1997, **203**, 1–9.
25 P. F. Fitzpatrick, *J. Am. Chem. Soc*., 1994, **116**, 1133–1134.
26 W. Zhang, B. Ostash and C. T. Walsh, *Proc. Natl. Acad. Sci. U. S. A*., 2010, **107**, 16828–16833.
27 D. G. Fujimori, S. Hrvatin, C. S. Neumann, M. Strieker, M. A. Marahiel and C. T. Walsh, *Proc. Natl. Acad. Sci. U. S. A*., 2007, **104**, 16498–16503.
28 T. Stachelhaus, H. D. Mootz and M. A. Marahiel, *Chem. Biol*., 1999, **6**, 493–505.
29 G. L. Challis, J. Ravel and C. A. Townsend, *Chem. Biol*., 2000, **7**, 211–224.
30 C. Rausch, T. Weber, O. Kohlbacher, W. Wohlleben and D. H. Huson, *Nucleic Acids Res*., 2005, **33**, 5799–5808.
31 L. Du and L. Lou, *Nat. Prod. Rep*., 2010, **27**, 255–278.
32 T. J. Erb, V. Brecht, G. Fuchs, M. Muller and B. E. Alber, *Proc. Natl. Acad. Sci. U. S. A*., 2009, **106**, 8871–8876.
33 T. J. Erb, J. Retey, G. Fuchs and B. E. Alber, *J. Biol. Chem*., 2008, **283**, 32283–32293.
34 W. Zhang, B. D. Ames, S. Tsai and Y. Tang, *Appl. Environ. Microbiol*., 2006, **72**, 2573–2580.
35 Y. Chan, A. Podevels, B. Kevany and M. Thomas, *Nat. Prod. Rep*., 2009, **26**, 90–114.
36 H. Ikeda and S. Omura, *Chem. Rev*., 1997, **97**, 2591–2610.
37 K. C. Nicolaou, J. Xu, F. Murphy, S. Barluenga, O. Baudoin, H. Wei, D. L. Gray and T. Ohshima, *Angew. Chem., Int. Ed*., 1999, **38**, 2447–2451.
38 L. A. Paquette, M. Duan, I. Konetzki and C. Kempmann, *J. Am. Chem. Soc*., 2002, **124**, 4257–4270.
39 L. A. Kelley and M. J. E. Sternberg, *Nat. Protoc*., 2009, **4**, 363–371.
40 L. Falquet, M. Pagni, P. Bucher, N. Hulo, C. J. A. Sigrist, K. Hofmann and A. Bairoch, *Nucleic Acids Res*., 2002, **30**, 235–238.
41 W. Liu, S. D. Christenson, S. Standage and B. Shen, *Science*, 2002, **297**, 1170–1173.
42 T. Kieser, M. Bibb, M. Butter, K. F. Chater and D. A. Hopwood, *Practical Streptomyces Genetics*, The John Innes Foundation, Norwich, 2001.
43 J. Sambrook and D. W. Russell, *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, New York, 3rd edn, 2001.
44 B. Gust, T. Kieser and K. F. Chater, *REDIRECT Technology: PCR Targeting System in STREPTOMYCES Coelicolor*, John Innes Centre, Norwich Research Park, Colney, Norwich, NR4 7UH, United Kingdom, 2002.
45 J. E. Coligan, B. M. Dunn, H. L. Ploegh, D. W. Speicher and P. T. Wingfield, *Current Protocols in Protein Science*, John Wiley and Sons, Inc., New York, 1995.