Supplementary information

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]*

[a]State Key Laboratory of Bio-organic Chemistry and Natural Products, Shanghai Institute of Organic Chemistry, Chinese Academy of Sciences, Shanghai, 200032, China. Email: wliu@mail.sioc.ac.cn, Tel +86-21-54925111, Fax: +86-21-64166128

[b]Biotica Technology Ltd, Chesterford Research Park, Cambridge CB10 1XL, UK

Materials and Procedures

Verification genotypes of mutant strains

All of the mutant strains' genotypes except *TL3001* were confirmed by PCR analysis. Genomic DNA of mutants was isolated following standard protocol [1]. With the wild type DNA as control, mutants were verified by PCR using primers listed in Table S2. For *TL3001*, five individual isolates of mutants genomic DNA together with wide type strains' were digested by *Xho*I-*Hin*dIII, and further analyzed by southern blotting with using 1.8 kb *Kpn*I fragment of ermE from pAGe-1 as the probe. Genotype analysis results were showed in Figure S7.

Table S1. Bacterial strains and plasmids used in this study

Strain/Plasmid	Characteristic(s)	Source/Reference
E. coli		
DH5α	Host for general cloning	Invitrogen
XL1-Blue MRF	Host for genomic library construction	Stratagene
S17-1	Donor strain for conjugation between *E.coli* and *Streptomyces*	[1]
S. flaveolus		
DSM 9954	Wild type strain, SFA producing	DSMZ
Streptomyces		
DSM 9954	Wild type strain, SFA producing	DSM
TL3001	Δ(*sfaK~sfaH*) gene replacement mutant, SFA non-producing	This study
TL3002	Δ*orf3* in frame deletion mutant, SFA producing	This study
TL3003	Δ*orf4* in frame deletion mutant, SFA producing	This study
TL3005	Δ*sfaR* gene replacement mutant, SFA producing with a reduced yield	This study
TL3006	Δ*sfaA* gene replacement mutant, SFA non-producing	This study
TL3007	Δ*sfaK* in frame deletion mutant, SFA non-producing	This study
Plasmids		
pGEM-T easy	*E. coli* subcloning vector	Promega
pGEM 5zf	*E. coli* subcloning vector	Promega
pANT841	*E. coli* subcloning vector	AF438749
pSP72	*E. coli* subcloning vector	Promega
SuperCos 1	Cosmid vector for genomic library construction	Stratagene

pAGe-1	Containing erythromycin resistant gene *ermE*	From J.A. Salas
pKC1139	*E.coli-Streptomyces* shuttle vector for gene inactivation, temperature sensitive replication in *Streptomyces* with apramycin resistance	[2]
pKC5201	Derivative of pKC1139 with aparamycin gene replaced by neomycin resistance gene from SuperCos1	This study
pTL3101	SuperCos1-based, *S. flaveolus* genomic library cosmid	This study
pTL3102	SuperCos1-based, *S. flaveolus* genomic library cosmid	This study
pTL3103	SuperCos1-based, *S. flaveolus* genomic library cosmid	This study
pTL3104	SuperCos1-based, *S. flaveolus* genomic library cosmid	This study
pTL3105	SuperCos1-based, *S. flaveolus* genomic library cosmid	This study
pTL3106	SuperCos1-based, *S. flaveolus* genomic library cosmid	This study
pTL3107	SuperCos1-based, *S. flaveolus* genomic library cosmid	This study
pTL3111	pKC1139 derivative for gene replacement of *pks*	This study
pTL3113	pKC5201 derivative for gene replacement of *sfaA*	This study
pTL3114	pKC1139 derivative for gene replacement of *orf3*	This study
pTL3115	pKC1139 derivative for gene replacement of *sfaK*	This study
pTL3116	pKC1139 derivative for gene replacement of *sfaR*	This study
pTL3117	pKC1139 derivative for gene replacement of *orf4*	This study
pTL3201	pSP72 derivative containing *ermE* resistant gene	This study

	from pAGe-1	
pTL3202	pTL3104 derivative for gene replacement	This study
pTL3203	pTL3202 derivative contains *ermE* resistant gene from pTL3201	This study
pTL3205	pTL3102 derivative with the *sfaA* gene replaced by *acc(3)IV-oriT*	This study
pTL3206	pANT841 derivative containing left arm for *orf3* in frame deletion	This study
pTL3207	pGEM 5zf derivative containing right arm for *orf3* in frame deletion	This study
pTL3208	pGEM 5zf derivative containing left arm for *sfaK* in frame deletion	This study
pTL3209	pGEM 5zf derivative containing right arm for *sfaK* in frame deletion	This study
pTL3210	pSP72 derivative containing 4.7 kb *Bgl*II-*Kpn*I fragment of TL3102	This study
pTL3211	pSP72 derivative containing left and right arm for *sfa*R in gene replacement	This study
pTL3212	pTL3211 derivative containing left and right arm for *orf4* gene replacement	This study
pTL3213	pET28a derivative containing *Nde*I-*Hin*dIII fragment of sfaR for protein expression	This study

Table S2. Primers used in this study for verification of the genotype

orf3	orf3-gt-for: 5'-GCGCTGGATCTGCTGAAGGTCC-3'
	orf3-gt-rev: 5'- GATGTAGTCGCGCGAGATGCCG-3'
sfaK	sfaK-gt-for: 5'-ACTGATCAAGGCCGTACTCG-3'
	SfaK-gt-rev: 5'-GACCTTCGTGTCGTACACCTC-3'
sfaR	sfaR-gt-for: 5'-GTGAACGAGATACTCAGCGCC-3'
	sfaR-gt-rev: 5'-TCAGACCGGCTCTGTCTGCGG-3'
orf4	orf4-gt-for: 5'-GCAACGTGCACCACGGCAAGG-3'
	orf4-gt-rev: 5'-CCAAGGCCGAACGCGATCTCC-3'
sfaA	sfaA-gt-for: 5'-TTTCCATGGAAATCGGCTCGGGCGC
	sfaA-gt-rev: 5'-TTTAAGCTTGTATATCAACCGCCATTAG
	TTTTTCAATGGATG

Fig. S1 Alignment of the partial amino acid sequences of crotonyl-CoA reductases. Conserved residues shaded in yellow are used for degenerate primer design. S.avermt: CCR from *S. avermitilis* MA-4680 (Accession no. NP_823087); S.cinnam: CCR from *S. cinnamonensis* (Accession no. AAD53915); S.coelic: CCR from *S. coelicolor A3(2)* (Accession no. NP_630556); S.collin: CCR from *S. collinus* (Accession no. AAA92890); S.diasta: CCR from *S. diastaticus* (Accession no. AAR16523); S. fradia: CCR from *S. fradiae* (Accession no. CAA57474); S.hygros: CCR from *S. hygroscopicus* (Accession no. AAR32675); S.kanamy: CCR from *S. kanamyceticus* (Accession no. CAI96521); S.spHK80: CCR from *S. sp. HK803* (Accession no. AAQ84149).

```
                70                                  352
S.avermt    LVAVMASSVNYNSVW............RYLWMSLKRIIGSH
S.cinnam    LVAVMASSVNYNSVW............RYLWMSLKRIIGSH
S.coelic    LVAVMASSVNYNSVW............RYLWMSLKRIIGSH
S.collin    LVAVMASSVNYNSVW............RYLWMSLKRIIGSH
S.diasta    LVAVMASSVNYNSVW............RYLWMSLKRIVGSH
S.fradia    LVAVMASSVNYNTVW............RYLWMSLKKIVGSH
S.hygros    LVAVMASSINYNTVW............RYLWMKLKKIVGSH
S.kanamy    LVAVMASSVNYNSVW............RYLWMSLKRIIGSH
S.spHK80    LIAVMASSINYNTVW............RYLWMSLKRIVGSH
Consensus       MASSvNYN                    WMsLKrIi
```

Fig. S2 Restriction map of the 150 kb DNA region from *Streptomyces flaveolus* DSM9954. Three overlapping cosmids pTL3102, pTL3014 and pTL3106 were selected for further DNA sequencing. Solid rectangles indicate the probe loci. B, *Bam*HI.

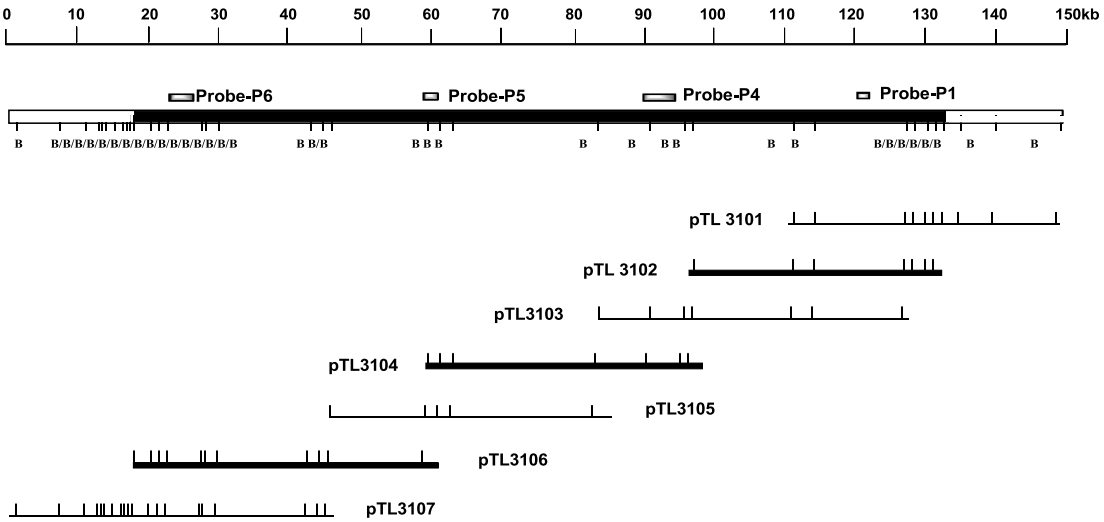


Fig. S3 Alignment of the partial amino acid sequences of the AT domains from SFA cluster. Sequence in blocks A, B, C correlate with malonyl-CoA, (2*S*)-2-methylmalonyl-CoA, and putative (2*S*)-2-(2-oxobutyl)malonyl-CoA specificity. Residues shaded in bright yellow are essential for catalytic activity. The residue G marked in red is assumed to enlarge the binding cavity of SfaI-M13-AT to accept the more bulky substrate. Position of residues are corresponding to the prototypical AT, *E. coli* FabD [3].

```
                 A              B                     C
Malonyl-CoA     63             94           121         199
SfaF-M4-AT      RTVYAQ.......GHSVGE.......R........HAFH
SfaG-M6-AT      RTAYAQ.......GHSLGE.......R........HAFH
SfaH-M7-AT      RTLYAQ.......GHSVGE.......R........HAFH
SfaH-M9-AT      RTVYAQ.......GHSVGE.......R........HAFH
SfaH-M10-AT     RTVYAQ.......GHSVGE.......R........HAFH
SfaH-M11-AT     DTGYGQ.......GHSVGE.......R........HAFH
SfaK-AT         DTRLAQ.......GHSYGE.......R........AAFH

(2S)-2-methylmalonyl-CoA
SfaE-M1-AT      RVDVVQ.......GHSQGE.......R........YASH
SfaF-M2-AT      RVDVVQ.......GHSQGE.......R........YASH
SfaF-M3-AT      RVDVVQ.......GHSQGE.......R........YASH
SfaG-M5-AT      RVDVVQ.......GHSQGE.......R........YASH
SfaH-M8-AT      RVDVVQ.......GHSQGE.......R........YASH
SfaI-M12-AT     RVDVVQ.......GHSQGE.......R........YASH

(2S)-2-(2-oxobutyl)malonyl-CoA
SfaI-M13-AT     RVDVVQ.......GHSQGE.......R........GAGH
```

Fig. S4 Alignment of the partial amino acid sequences of the KR domains from the SFA cluster. Boxed residues are important for NADPH binding and characteristic residues for each KR-type are marked in blue or red. The yellow shading is used to highlight catalytic residues for ketoreductase activity. Based on the model [4], the KR of M5 (with residue W141 and without H146) belongs to the A1 type to form *"syn"* L-products; M12 (with residues W141 and H146) belongs to the A2 type to form *"anti"* L-products; and the remaining are B1 type KRs (with the characteristic LDD motif and without P151) which form *"anti"* D-products. The SfaK-DH/KR didomain shows typical characteristic residues of B1-type KR (the NADPH binding site locates at -151 compare to others). Positions of residues are corresponding to SfaF-M4-AT.

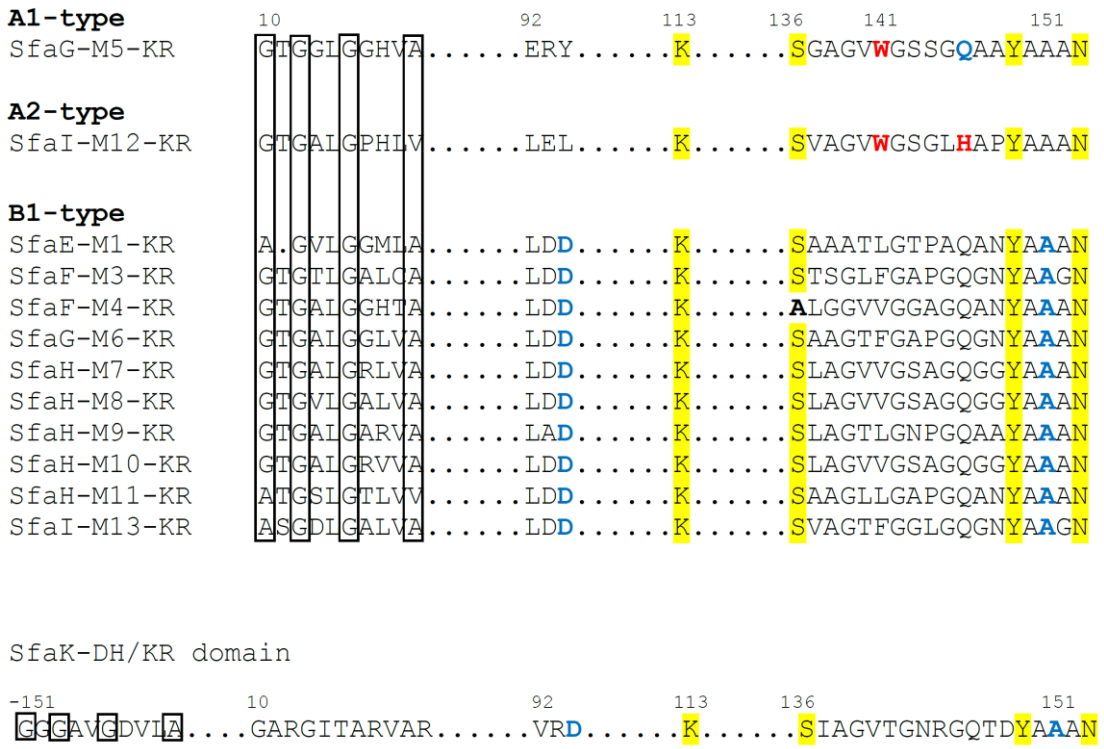


Fig. S5 Alignment of the partial amino acid sequences of the DH, ER and ACP domains from SFA cluster. Bright yellow is used to highlight catalytic residues for dehydratation **a)**, enoyl reduction **b)** and acyl carrier activity **c)**. SfaH-M13-DH and SfaF-M3-DH are anticipated to be inactive due to the lack of histidine 68. Position of residues are referenced relative to SfaE-M1-DH, SfaE-M1-ER and SfaE-LM-ACP, respectively.

a

```
                    36                  68   72     77
SfaE-M1-DH          HPLLA...........HTVLGTALLP
SfaG-M6-DH          HPLLS...........HTVRGSALLP
SfaH-M7-DH          HPLLR...........HSVLGTPVLP
SfaH-M8-DH          HPLLG...........HVVLGTTLLP
SfaH-M10-DH         HPLLG...........HAILGPALLP
SfaH-M11-DH         HPLLG...........HTVLDTVLLP
SfaH-M13-DH         HPLLG...........E.......AP
SfaF-M3-DH          HPLLG...................LP
```

b

```
                   114                 130
SfaE-M1-ER      RVLVHAAAGGVGMAAVRVAR
SfaG-M6-ER      RVLVHAAAGGVGMAAVRVAR
```

c

```
                       34
SfaE-LM-ACP         GLRSL
SfaE-M1-ACP         GFDSL
SfaF-M2-ACP         GFDSL
SfaF-M3-ACP         GFDSL
SfaF-M4-ACP         GFDSL
SfaG-M5-ACP         GFDSL
SfaG-M6-ACP         GFDSL
SfaH-M7-ACP         GFDSL
SfaH-M8-ACP         GFDSL
SfaH-M9-ACP         GFDSM
SfaH-M10-ACP        GFDSL
SfaH-M11-ACP        GFDSL
SfaI-M12-ACP        GFDSI
SfaI-M13-ACP        GFDSL
SfaK-ACP            GVDSL
```

Fig. S6 a) Alignment of substrate specificity-conferring residues for the SFA NRPS SfaD. Bright yellow is used to highlight key residues of adenylation domains for substrate selectivity. GrsA from *Brevibacillus brevis* (Accession no. P0C061); SfaD-M14-A, SfaD-M15-A, SfaD-M16-A are from this study. **b)** Adenylation domains with most identical substrate specificity-conferring motifs are shown. Conserved residues are highlighted in red. Positions of residues are referenced relative to GrsA.

a

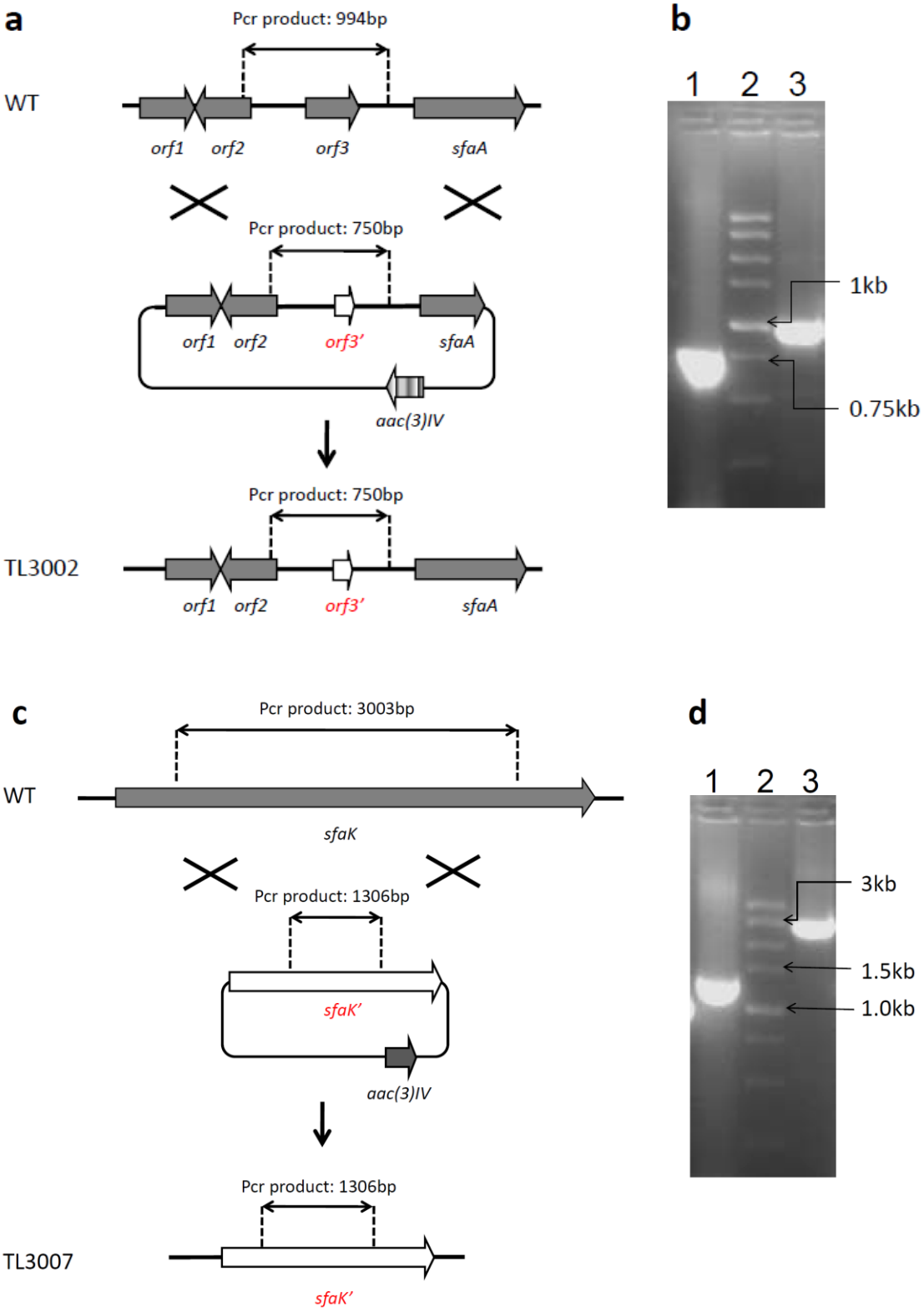
```
                235   239      278      299        322      330       517
GrsA            DASVW......T......ITA......A......TC......K
SfaD-M14-A      DASTY......W......VVG......G......TF......K
SfaD-M15-A      DISTY......C......LTG......A......AF......K
SfaD-M16-A      DVHVQ......F......SQA......H......VV......K
```

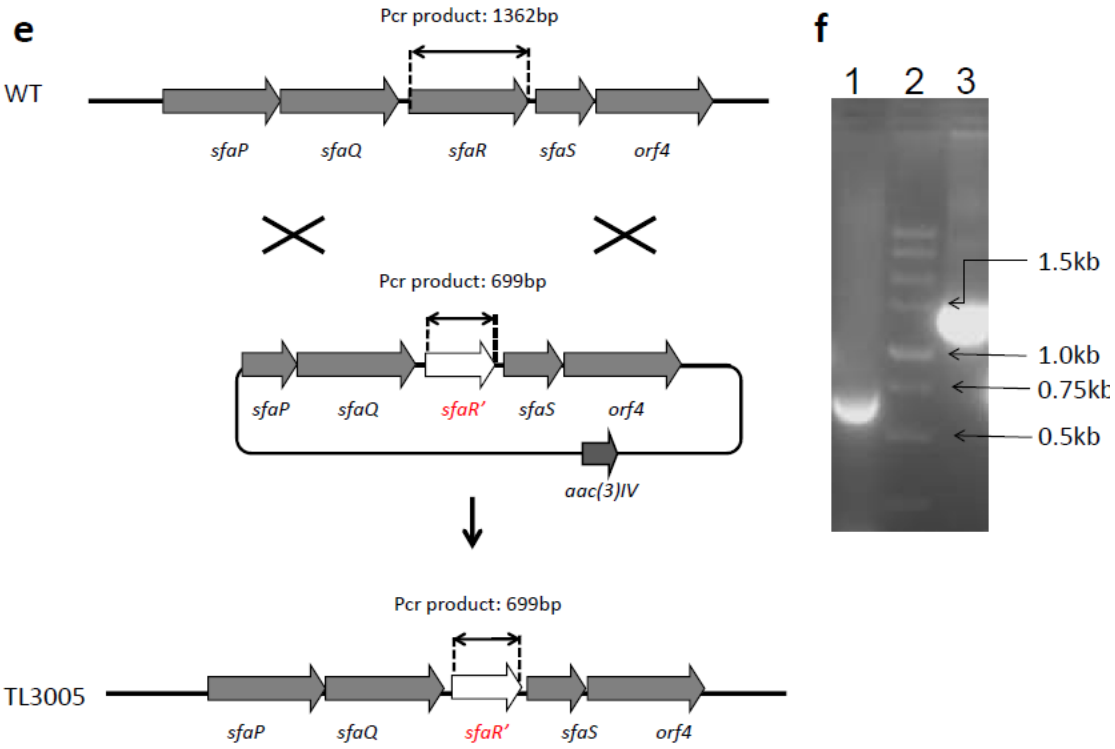
b

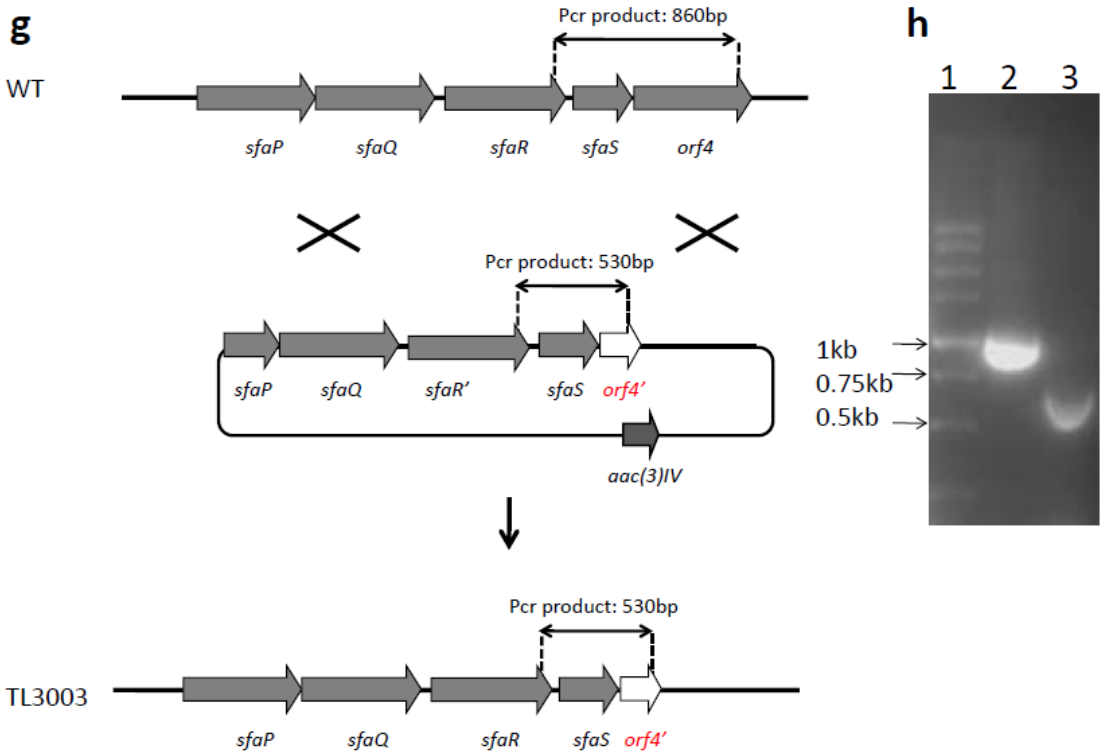
```
SfaD-M14-A: DAYWVGGTFK
Val         DAFWIGGTFK

SfaD-M15-A: DIYCLGAAFK
No hits

SfaD-M16-A: DVQFSAHVVK
Pro         DVQFAAHVVK
```

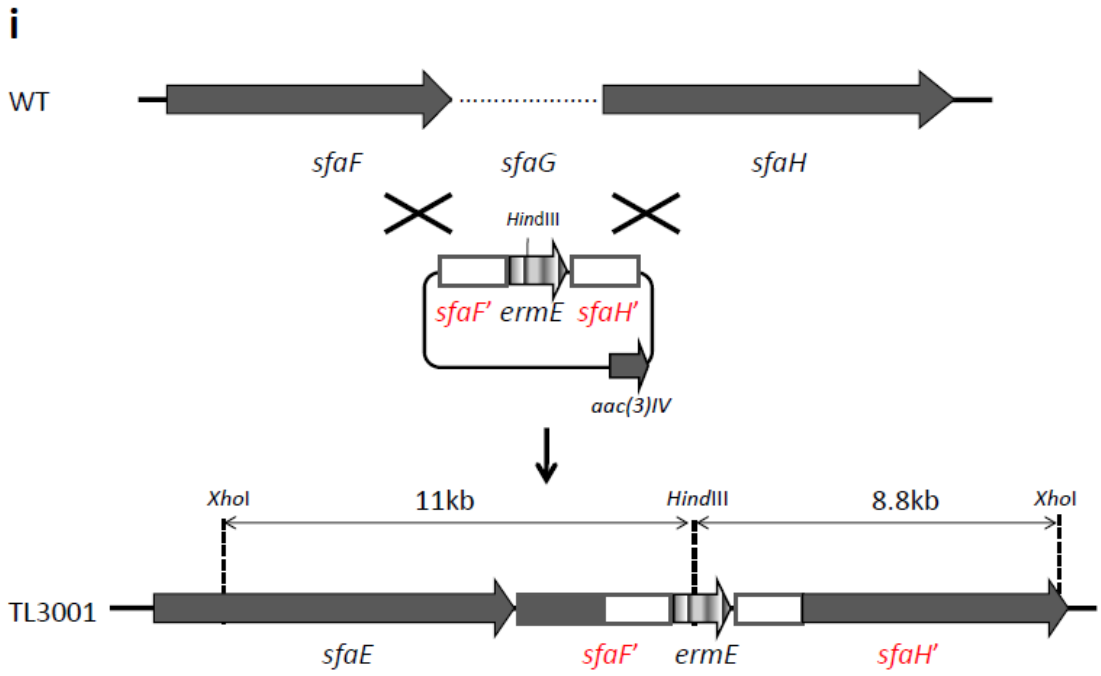
Fig. S7 Verification of *S. flaveolus* mutant genotypes

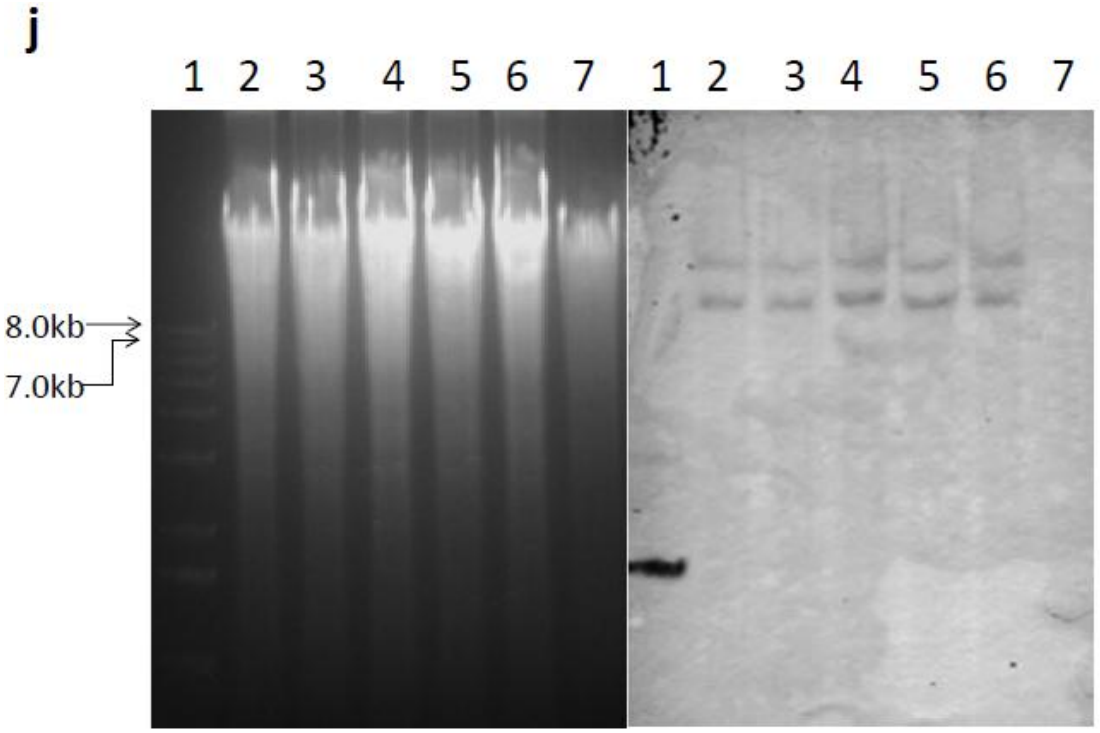
a) Construction of the *orf3* inframe deletion mutant and the size of the PCR fragments from wild-type strain and TL3002.
b) PCR analysis of TL3002 (lane 1) and wild-type strain (lane 3). Lane 2, molecular weight marker.
c) Construction of the *sfaK* inframe deletion mutant and the size of the PCR fragments from wild-type strain and TL3007.
d) PCR analysis of TL3007 (lane 1) and wild-type strain (lane 3). Lane 2, molecular weight marker.
e) Construction of the *sfaR* gene deletion mutant and the size of the PCR fragments from wild-type strain and TL3005.
f) PCR analysis of TL3005 (lane 1) and wild-type strain (lane 3). Lane 2, molecular weight marker.
g) Construction of the *orf4* gene deletion mutant and the size of the PCR fragments from wild-type strain and TL3003.
h) PCR analysis of TL3003 (lane 3) and wild-type strain (lane 2). Lane 1, molecular weight marker.
i) Construction of the pks deletion mutant (partial *sfaF, sfaH* and intact *sfaG*) and the restriction map of TL3001 showing the fragment sizes upon *Xho*I-*Hin*dIII digestion.
j) Gel (left) and Southern (right) analysis of TL3001 genomic DNA (Lanes 2~6 are five individual isolates). A 1.8 kb *Kpn*I fragment from pAGe-1 containing *erm*E was used as the probe. *Xho*I-*Hin*dIII digested TL3001 and wild-type strain genomic DNA (Lane 7) were tested by Southern analysis. Lane 1, molecular weight marker.
k) Construction of the *sfaA* gene replacement mutant and the size of the PCR fragments from wild-type strain and TL3006.
l) PCR analysis of TL3006 (lane 2) and wild-type strain (lane 3). Lane 1, molecular weight marker.


a
Pcr product: 994bp
WT
orf1
orf2
orf3
sfaA
Pcr product: 750bp
orf1
orf2
orf3'
sfaA
aac(3)IV
Pcr product: 750bp
TL3002
orf1
orf2
orf3'
sfaA
b
1 2 3
1kb
0.75kb
c
Pcr product: 3003bp
WT
sfaK
Pcr product: 1306bp
sfaK'
aac(3)IV
Pcr product: 1306bp
TL3007
sfaK'
d
1 2 3
3kb
1.5kb
1.0kb


e
Pcr product: 1362bp
WT
sfaP
sfaQ
sfaR
sfaS
orf4
Pcr product: 699bp
sfaP
sfaQ
sfaR'
sfaS
orf4
aac(3)IV
Pcr product: 699bp
TL3005
sfaP
sfaQ
sfaR'
sfaS
orf4
f
1 2 3
1.5kb
1.0kb
0.75kb
0.5kb


g
Pcr product: 860bp
WT
sfaP
sfaQ
sfaR
sfaS
orf4
Pcr product: 530bp
sfaP
sfaQ
sfaR'
sfaS
orf4'
aac(3)IV
Pcr product: 530bp
TL3003
sfaP
sfaQ
sfaR
sfaS
orf4'
h
1 2 3
1kb
0.75kb
0.5kb


i
WT
sfaF
sfaG
sfaH
HindIII
sfaF' ermE sfaH'
aac(3)IV
XhoI
11kb
HindIII
8.8kb
XhoI
TL3001
sfaE
sfaF'
ermE
sfaH'


j
1 2 3 4 5 6 7 1 2 3 4 5 6 7
8.0kb
7.0kb

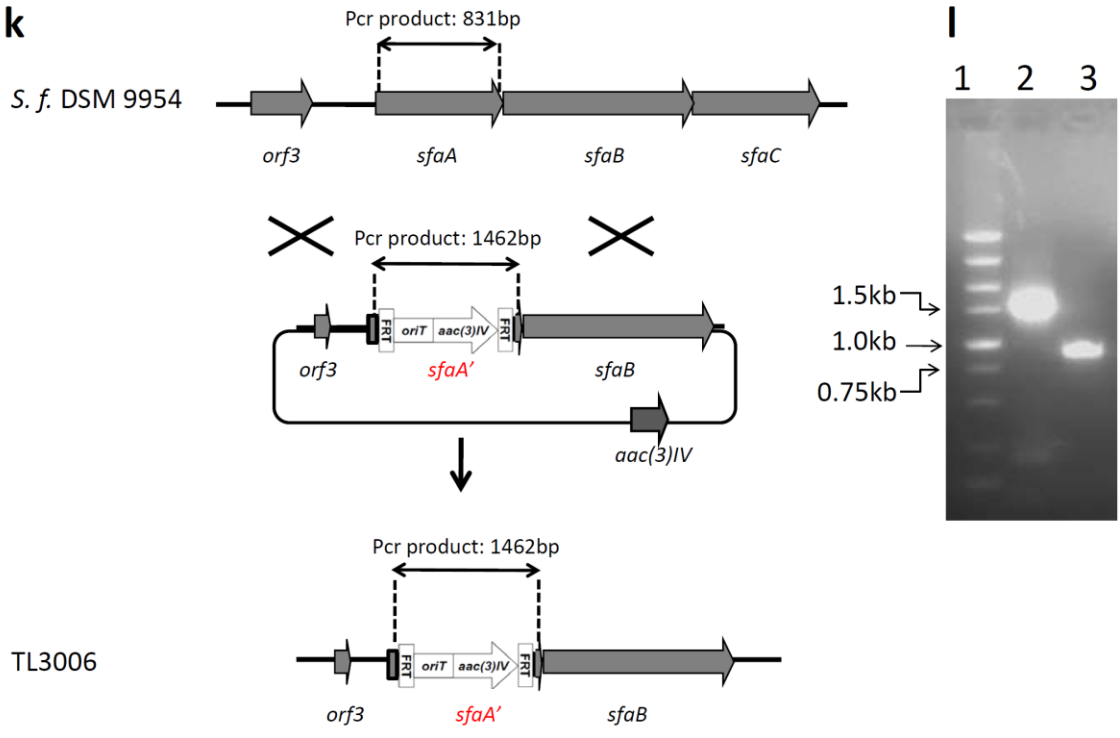

k
Pcr product: 831bp
S. f. DSM 9954
orf3
sfaA
sfaB
sfaC
Pcr product: 1462bp
FRT
oriT
aac(3)IV
FRT
orf3
sfaA'
sfaB
aac(3)IV
Pcr product: 1462bp
TL3006
orf3
sfaA'
sfaB
l
1 2 3
1.5kb
1.0kb
0.75kb

Fig. S8 10% SDS-PAGE of purified SfaR.

Lane1: SfaR; Lane2: protein ladder (Fermentas, SM0431)

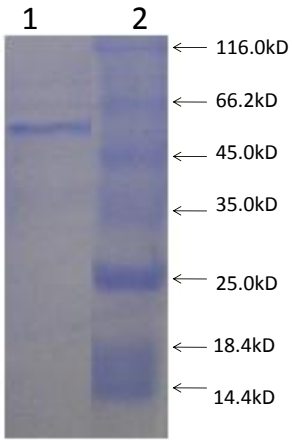
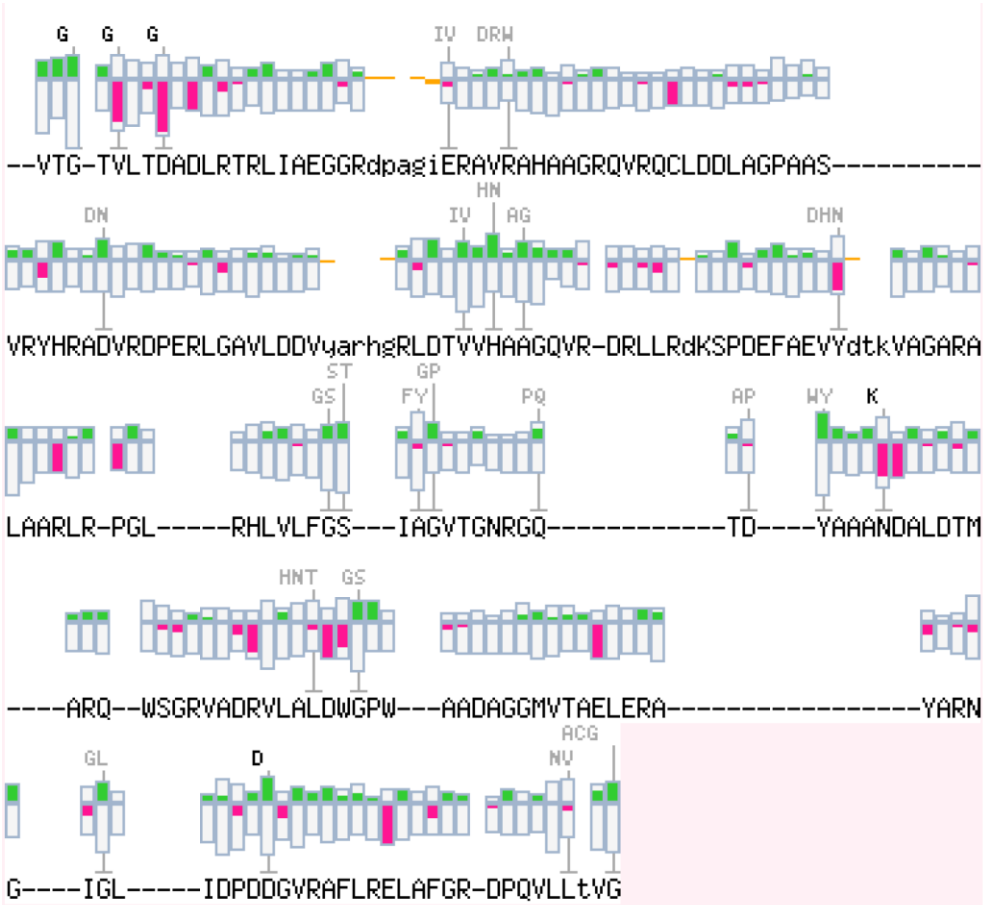


Fig. S9 Sequence analysis of SfaK-DH/KR didomain.

a) The DH/KR didomain sequence; Characteristic residues for NADPH binding are marked in blue; and active site residues for KR are marked in red (also see Figure S4).

ATRTGPDGPGPDGPGRQTAVPEPETMPAGASDTVRHVVEPVPEEPPAGPL
PTLTRAAVSGGGAVGDVLATLLKERGTEVVSDPAGCDALLLLDALDGGDYTL
PGRFTDIRAAVLGGLRTLLLATCHDAGPAGSGVHGLARALSREHPGLAVTAV
DLPAGQPAEEAARTLLAELGGSRPSVTHTDGRRAVWRTRPAPLPAADVTAG
DLGLDRDSVVLLTGGARGITARVARALATLTGCHLELVGRSEPVTGTVLTDAD
LRTRLIAEGGRDPAGIERAVRAHAAGRQVRQCLDDLAGPAASVRYHRADVR
DPERLGAVLDDVYARHGRLDTVVHAAGQVRDRLLRDKSPDEFAEVYDTKVA
GARALAARLRPGLRHLVLFGSIAGVTGNRGQTDYAAANDALDTMARQWSG
RVADRVLALDWGPWAADAGGMVTAELERAYARNGIGLIDPDDGVRAFLREL
AFGRDPQVLLTVGDPAGFGSALD

b) The achieved data from Motif Scan. Characteristic residues for NADPH dependent dehydratase family are shown in the top of each line; below is the sequence of DH/KR didomain (position: 249-474, raw-score = -61.1, N-score = 10.821 and E-value = 3.2 e-4)



References

1. T. Kieser, M. Bibb, M. Butter, K. F. Chater and D. A. Hopwood, *Practical Streptomyces Genetics*, The John Innes Foundation, Norwich., 2001.

2. M. Bierman, R. Logan, K. O'Brien, E. T. Seno, R. N. Rao and B. E. Schoner, *Gene*, 1992,**116**, 43–49.

3. C. D. Reeves, S. Murli, G. W. Ashley, M. Piagentini, C. R. Hutchinson and R. McDaniel, *Biochemistry*, 2001, **40**, 15464–15470.

4. A. T. Keatinge-Clay, *Chem. Biol.*, 2007, **14**, 898–908.