SUPPLEMENTARY INFORMATION

Engineered biosynthesis of hybrid macrolide polyketides containing D-angolosamine and D-mycaminose moieties

Ursula Schell[a,¥], Stephen F. Haydock[b,¥], Andrew L. Kaja[a], Isabelle Carletti[a], Rachel E. Lill[a], Eliot Read[a], Lesley S. Sheehan[a], Lindsey Low[a], Maria-Jose Fernandez[b], Friederike Grolle[b], Hamish A. I. McArthur[c], Rose M. Sheridan[a], Peter F. Leadlay[b], Barrie Wilkinson[a*] and Sabine Gaisser[a,d*]

[a]Biotica Technology Ltd, Chesterford Research Park, Little Chesterford, Saffron Walden, Essex CB10 1XL, UK

[b]Department of Biochemistry, University of Cambridge, 80 Tennis Court Road, Cambridge CB2 1GA, UK

[c]Bioprocess Research, Central Research Division, Pfizer Inc., Groton, CT 06340, USA

[d]Hochschule Biberach, Karlstraße 11, 88400 Biberach/Riß, Germany (current address)

*Authors to whom correspondence should be addressed: gaisser@fh-biberach.de (biology); barrie.wilkinson@biotica.com (chemistry)

[¥] These authors made equal contributions

Construction of biosynthetic gene cassettes

The strategy represents a further development of the *Xba*I based method described previously.[1] Expression vector pSG142[1,2] was adapted to allow the assembly of gene cassettes directly and expression vector pSG144 was created. This expression vector allows the assembly of gene cassettes directly to replace *eryBV* and then build up the cassette without prior assembly in pUC-derived vectors as was the method described previously.[1] To achieve this pSG142 was digested with *Xba*I and a fill-in reaction was performed using standard protocols. The DNA was ligated and used to transform *E. coli* DH10B. Construct pSG143 was isolated and the removal of the *Xba*I site was confirmed by sequence analysis. Plasmid pSGLit1 (see below) was digested with *Nde*I and *Bgl*II and an approximately 1.3 kb insert was isolated. Plasmid pSG143 was digested with *Nde*I and *Bgl*II, the vector band was isolated and ligated with the approximately 1.3 kb band from pSGLit1 followed by transformation of *E. coli* DH10B. Plasmid pSG144 (Figure S1) was isolated and the construct was verified by sequence analysis.

Isolation of conversion vector pSGLit1

Conversion vector pSGLit1 was used to include the glycosyltransferase genes at the end of pre-assembled gene cassettes. To do this, gene *eryBV* was amplified by PCR using the primers BIOSG1 5'-GGGTCTAGATCCGGACGAACGCATCGATTAATTAAGGAGGACACATA-3' and 7966 5'-GGGGAATTCAGATCTGGTCTAGAGGTCAGCCGGCGTGGCGGCGCGTGAGTTCCTCCAGTCGCGGGACGATCT-3', to introduce a *Xba*I site sensitive to Dam

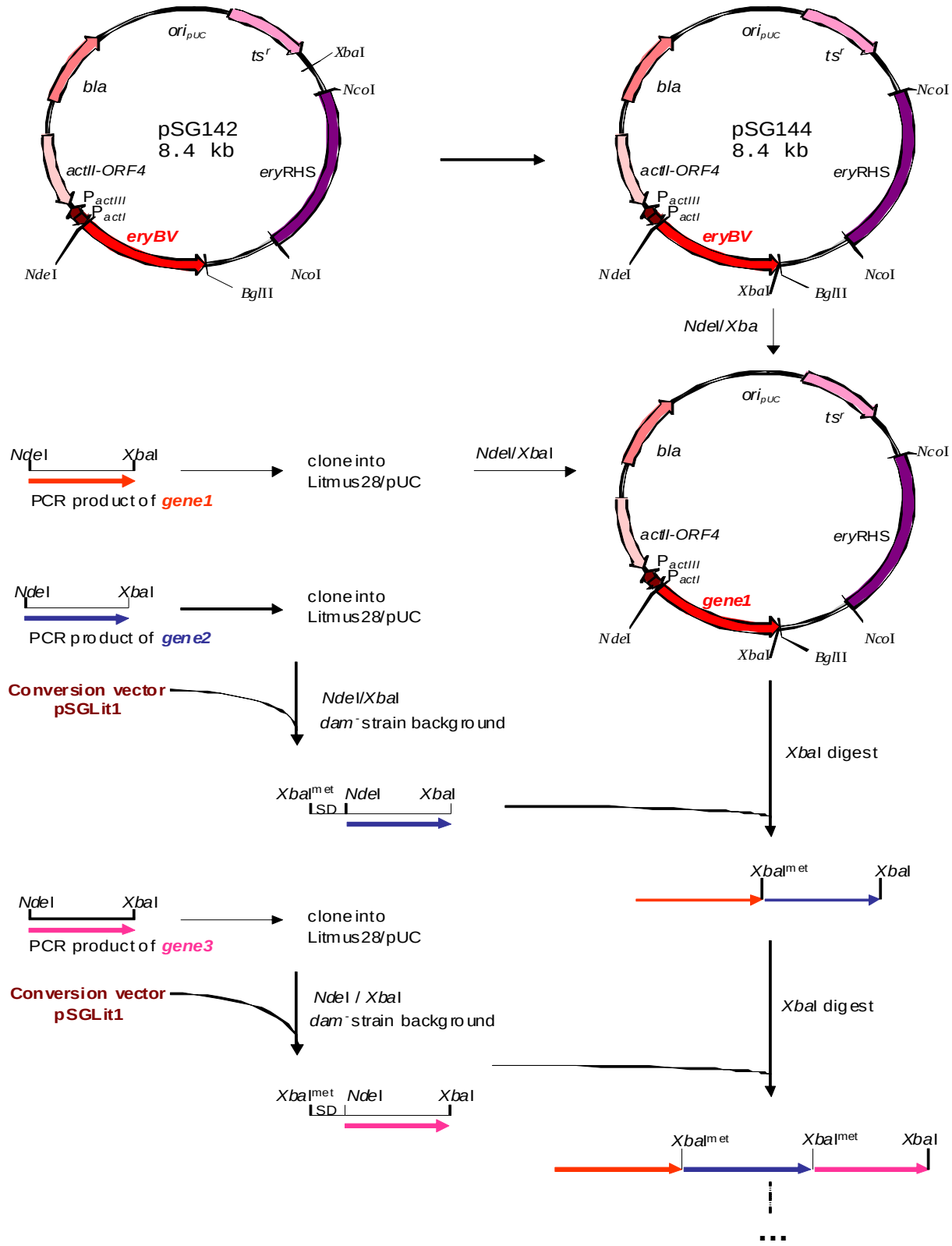

Figure S1. Iterative assembly of biosynthetic gene cassettes; *Xba*I^{met}, site sensitive to the Dam methylase in *E. coli*; *EryRHS* denotes a DNA fragment from the *ermE* distal flank of the erythromycin biosynthetic gene cluster; SD denotes a Shine Dalgarno sequence.

methylation at the 5' end and a *Xba*I site as well as a *Bgl*II site at the 3' end of *eryBV*. Plasmid pUC18eryBVcas was used as a template. Using standard techniques the PCR product was ligated with *Sma*I-cut pUC18 and used to transform *E. coli* DH10B. The construct was then digested using *Bam*HI and *Bgl*II and an about 1.3 kb DNA band was isolated followed by the ligation with *Bam*HI and *Bgl*II digested Litmus 28 vector DNA using standard procedures. The vector pSGLit1 was isolated and the DNA sequence of the insert was verified by sequence analysis. (Plasmid pUC18eryBVcas was constructed as follows: The gene *eryBV* was amplified by PCR using the primers casOleG21 casoleG21 5'-GGGGAAGCTTGCCGACGATGACGACGACCACCGGACGAACGCATCGATTAATTAAG-3' and 7966 5'-GGGGAATTCAGATCTGGTCTAGAGGTCAGCCGGCGTGGCGGCGCGTGAGTTCCTCCAGTCGCGGGACGATCT-3' and pSG142[2] as template. The PCR fragment was cloned using standard procedures and plasmid pUC18eryBVcas was isolated with an *Nde*I site overlapping the start codon of *eryBV* and *Xba*I and *Bgl*II sites following the stop codon). Plasmid pSGlit1 was used to replace the glycosyltransferase gene *eryBV* with *eryCIII*, *tylMII* and *angMII* using *Nde*I and *Bgl*II digest and standard cloning procedures. To create pSGLit1eryCIII and pSGLit1tylMII plasmids pSGCIII and pSGTYLM2[2] were used to isolate the *Nde*I and *Bgl*II insert band. The primers employed to amplify *angMII* were BIOSG63 5'-GGGCATATGCGTATCCTGCTGACGTCGTTCGCGCACAACAC-3' and BIOSG80 5'-GGAGATCTGGCGCGGCGGTGCGCGGCGGTGAGGCGTTCG-3'. The cosmid5B2 containing a fragment of the angolamycin biosynthetic pathway was used as template. The plasmids were used to

transform *E. coli* ET12567 (*dam*$^-$) and the plasmids isolated to include the glycosyltransferases at the end of pre-assembled gene cassettes (Figure S1) using standard procedures. This introduced a his6-tag at the C-terminal end of EryCIII, TylMII and AngMII.

Isolation of conversion vector pSGLit2

This plasmid was used to introduce a 5'-region containing an *Xba*I site sensitive to Dam methylation and a Shine-Dalgarno region, thus converting genes which were originally cloned with an *Nde*I site overlapping the start codon and an *Xba*I site adjacent to the stop codon for the assembly of gene cassettes. Plasmid Litmus 28 was digested with *Spe*I and *Xba*I and the vector fragment was isolated. Plasmid pSGLit1 (*dam*$^-$) was digested with *Xba*I and the insert band was isolated and ligated with the *Spe*I and *Xba*I digested vector fragment of Litmus 28 followed by the transformation of *E. coli* DH10B using standard techniques. Plasmid pSGLit2 was isolated and the construct was verified by restriction digest and sequence analysis. This conversion included the transformation of the ligations into *E. coli* ET12567 followed by the isolation of *dam*$^-$ DNA and *Xba*I digests. Examples for this strategy are outlined below.

Isolation of conversion vector pSGLit3relig1

For the multiple use of promoter sequences in *act*-controlled gene cassettes a 240 bp fragment was amplified by PCR using the primers BIOSG78 5'-GGGCATATGTGTCCTCCTTAATTAATCGAT GCGTTCGTCC-3' and BIOSG79 5'-GGAGATCTGGTCTAGATCGTGTTCCCCTCCCTGCCTCGTGGTCCCTCACGC-3'

and plasmid pSG142[2] as template. The 0.2 kb PCR fragment was cloned employing standard procedures using *Eco*RV digested plasmid Litmus28. Plasmid conv no1 was isolated. The construct was verified by sequence analysis. Plasmid conv no1 was digested with *Nde*I and *Bgl*II and the about 0.2 kb fragment was isolated and ligated with the *Bam*HI and *Nde*I digested vector fragment of pSGLit2. The ligation was used to transform *E. coli* DH10B and plasmid pSGLit3relig1 was isolated using standard procedures. This construct was verified using restriction digests and sequence analysis.

Isolation of cassettes containing genes derived from the tylosin biosynthetic gene cluster

The primers used for gene amplification are listed in Table S1. PCR products were cloned into pUC-derived cloning vectors and the constructs were verified by sequence analysis. These plasmids were then digested with *Nde*I and *Xba*I and the insert bands were isolated and ligated into plasmid pSGLit2 digested with *Nde*I and *Xba*I. The pSGLit2 derived vectors were then used to transform *E. coli* ET12567 (*dam*$^-$) and DNA was isolated. These were digested with *Xba*I and the inserts isolated (see Figure S1). For iterative gene cassette assembly plasmid pSG144 and pUC19*tylAI* were digested with *Nde*I and *Xba*I and the insert band of pUC19*tylAI* and the vector band of pSG144 were isolated, ligated together and used to transform *E. coli* DH10B. Plasmid pSG144*tylAI* was thus isolated. As a second step, plasmid pSGLit2*tylAII* (*dam*$^-$) was digested with *Xba*I and ligated with *Xba*I digested plasmid pSG144*tylAI*. The ligation was used to transform *E. coli* DH10B and plasmid pSG144*tylAItylAII* was isolated and verified using standard protocols. As a third step plasmid pSGLit2*tylMIII* (*dam*$^-$) was digested with *Xba*I and the insert band was ligated with *Xba*I digested plasmid pSG144*tylAItylAII*. The

ligation was used to transform *E. coli* DH10B and plasmid pSG144*tylAItylAIItylMIII* was isolated using standard protocols. The construct was verified using restriction digests and sequence analysis. Using this strategy, the gene cassettes were built up step by step inserting the desired gene as appropriate. To complete the gene cassette plasmid pUC*tylMI* was *Pac*I digested and the insert was ligated with the *Pac*I digested vector fragment of pSGLit1*eryCIII* using standard procedures. Plasmid pSGLit1*tylMIeryCIII* was isolated and the orientation was confirmed by restriction digests and sequence analysis. Plasmid pSGLit1*tylMIeryCIII* was digested with *Xba*I and *Bgl*II and the insert band was isolated and ligated with the *Xba*I and *Bgl*II digested vector fragment of plasmid pSG144*tylAItylAIItylMIIItylBtylIa*. Plasmid pSG144*tylAItylAIItylMIIItylBtyl1atylMIeryCIII* was isolated using standard procedures and the construct was confirmed using restriction digests and sequence analysis. Using the same approach, plasmid pSGLit1*tylMItylMII* was isolated and the gene cassette pSG144*tylAItylAIItylMIIItylBtylIatylMItylMII* was created. Plasmid preparations were used to transform *S. erythraea* mutant strains with standard procedures.

Isolation of biosynthetic gene cassettes containing TDP-D-angolosamine biosynthetic genes

Genes of interest were amplified using standard protocols and primers listed in Table S1. These were cloned into pUC-vectors or Litmus 28. The resulting constructs were verified by sequence analysis. The resulting genes were then cloned into pSGlit2 as *Nde*I and *Xba*I digested fragments using the assembly protocol described above. Transformations were then performed using *E. coli* ET12567 (*dam*$^-$) and DNA isolated. Subsequently *Xba*I

restriction digests were performed, the gene fragments isolated and used for the construction of gene cassettes after the first gene of the cassette, *angAI,* had been cloned using *Nde*I and *Xba*I digest and ligated with *Nde*I and *Xba*I digested DNA of pSG144. Using the gene cassette approach outlined in Figure S1 the biosynthetic gene cassettes were assembled. To insert the glycosyltransferase gene at the end of the cassette plasmid for example pSGLit1*eryCIII* (isolated from *E. coli* ET12567) was digested with *Xba*I and *Bgl*II and the about 1.2 kb fragment was isolated and ligated with the *Xba*I digested and partially *Bgl*II digested vector fragment of the cassette construct. The *Bgl*II partial digest was necessary due to the presence of a *Bgl*II site in *angB*. Using the same strategy, the glycosyltransferase genes *angMII* and *tylMII* were cloned at the end of the gene cassettes. During the construction of further gene cassettes the presence of the *Bgl*II site in *angB* was used such that the DNA fragments carrying *angMI* and the glycosyltransferase genes were lifted from already existing gene cassettes (as described above) using *Bgl*II digests and incorporated. Gene cassette constructs such as pSG144*angAIangAIIorf14angMIIIangBangMIeryCIII* and pSG144*angAIangAIIorf14angMIIIangBangMIangMII* were isolated.

Construction of strain Saccharopolyspora erythraea *SGQ2 (BIOT-2175)*

Plasmid pGG1 was used to introduce a 900 bp deletion in the *eryCIV* gene[3] of the strain *S. erythraea* SGT2[2] in order to create the quadruple mutant SGQ2 using procedures described previously[1,2,4]. Plasmid pGG1 was isolated as follows: Plasmid pNCO62[5] was isolated from the *dam*$^-$ strain *E. coli* 12567 and digested with *Bal*I and *Bcl*I. To introduce a 0.9 kb deletion into *eryCIV* as previously described.[3] These were end filled-in and then

ligated together using standard protocols. The ligation mix was used to transform *E. coli* DH10B. Plasmid pGG17 was isolated and confirmed by sequence analysis and restriction digestion. A 1.1 kb fragment containing the thiostrepton resistance gene was isolated from plasmid pIB060 by digestion with *Hin*dIII/*Sph*I This fragment was ligated with the about 5.3 kb *Sph*I/*Hin*dIII fragment of pGG17 and plasmid pGG1 was thus isolated using standard protocols.

Construction of strain Saccharopolyspora erythraea *Q42/1 (BIOT-2166)*

A chromosomal mutation was introduced into *S. erythraea* SGQ2 to prevent the biosynthesis of TDP-L-mycarose. Plasmid pSGKC1 was isolated by cloning the approximately 700 bp fragment of the *eryBVI* gene by using PCR amplification with cosmid2 a template and the primers 646 (5'-CATCGTCAAGGAGTTCGACGGT- 3') and 874 (5'-GCCAGCTCGGCGACGTCC ATC- 3') using standard protocols. Cosmid 2 containing the right hand site of the *ery*-cluster was isolated from an existing cosmid library[5] by screening with *eryBV* as a probe using standard protocols. The amplified DNA fragment was isolated and cloned into *Eco*RV digested pKC1132[6] using standard protocols. The ligation mixture was used to transform *E. coli* DH10B and plasmid pSGKC1 was isolated. The construct was verified by DNA sequence analysis. Plasmid pSGKC1 was used to transform *S. erythraea* SGQ2 using standard techniques followed by selection with apramycin. The thiostrepton/apramycin resistant transformant *S. erythraea* Q42/1 was thus isolated.

Construction of strain Saccharopolyspora erythraea *LB1 (BIOT-2634)*

To introduce a deletion comprising the PKS and majority of post PKS genes in *S. erythraea* a region of the left hand side of the *ery* cluster (LHS) containing a portion of *eryCI*, the complete *ermE* gene and a fragment of the *eryBI* gene were cloned together with a region of the right hand side of the *ery* cluster (RHS) containing a portion of the *eryBVII* gene, the complete *eryK* gene and a fragment of DNA adjacent to *eryK*. This construct should theoretically enable homologous recombination into the genome in either of the LHS or RHS regions, which after a second recombination event, would result in the isolation of a strain containing a deletion between these two regions of DNA. The LHS fragment (2201 bp) was PCR amplified using *S. erythraea* chromosomal DNA as template and primers BIdelNde (5'-CCCATATGACCGGAGTTCGAGGTACGCGGCTTG-3') and BIdelSpe (5'-GATACTAGTCCGCCGACCGCACGTCGCTGAGCC-3'). Primer BIdelNde contains an *Nde*I restriction site (underlined) and primer BIdelSpe contains a *Spe*I restriction site (underlined) which were used for later cloning steps. The PCR product was cloned into the *Sma*I restriction site of pUC19 and plasmid pLSB177 was isolated using standard procedures. The construct was confirmed by sequence analysis. Similarly, RHS (2158 bp) was amplified by PCR using *S. erythraea* chromosomal DNA as template and primers BVIIdelSpe (5'-TGCACTAGTGGCCGGGCGCTCGACGT CATCGTCGACAT-3') and BVIIdelEco (5'-TCGATATCGTGTCCTGCGGTTTCACC TGCAACGCTG-3'). Primer BVIIdelSpe contains a *Spe*I restriction site (underlined) and primer BVIIdelEco contains an *Eco*RV restriction site (underline). The PCR product of these was cloned into the *Sma*I restriction site of pUC19 in the orientation with *Spe*I positioned adjacent to *Kpn*I and *Eco*RV positioned adjacent to *Xba*I. The plasmid pLSB178 was isolated and confirmed

using sequence analysis. Plasmid pLSB177 was digested with *Nde*I and *Spe*I, and the ~2.2kb fragment was isolated. Similarly plasmid pLSB178 was digested with *Nde*I and *Spe*I and the ~4.6 kb fragment was isolated. The two fragments were ligated together and plasmid pLSB188 containing LHS and RHS combined together at a *Spe*I site in pUC19 was thus isolated. pLSB188 was digested with *Nde*I and *Xba*I and a ~4.4 kbp fragment was isolated. This was ligated into *Spe*I and *Nde*I treated pCJR24.[7] The ligation mixture was used to transform *E. coli* DH10B and plasmid pLSB189 was isolated using standard methods. Plasmid pLSB189 was used to transform *S. erythraea* and transformants were selected using thiostrepton. One such resulting mutant named *S. erythraea* Del18 was isolated and inoculated into TSB medium (6 ml) and grown for 2 days using standard procedures. This strains was subcultured three times using the same procedure but growing without thiostrepton (5% inoculum each time). The final culture (100 µl) was plated onto R2T20 agar[8] and incubation at 30^{o}C until sporulation occured. Spores were harvested, filtered, diluted and plated onto R2T20 agar using standard procedures. Colonies were replica plated onto R2T20 plates with and without addition of thiostrepton. Colonies no longer resistant to thiostrepton were selected and further grown in TSB medium. One such colony was isolated and confirmed using PCR and Southern blot analysis. This strain was designated *S. erythraea* LB1 (BIOT-2634). For further analysis, the production of erythromycin was assessed as described previously. [1,2] This strain was cultured using standard methods and the lack of erythromycin production was confirmed by LCMS analysis. As anticipated, in bioconversion assays this strain was unable to glycosylate exogenously added erythronolide B, whereas exogenously added

erythromycin D was efficiently hydroxylated at C12 to yield erythromycin C, indicating functional *eryK*.

Table S1. PCR primers used for cloning genes discussed in above.

tylAI	BIOSG34	5'-GGGCATATGAACGACCGTCCCCGCCGCGCCATGAAGGG-3'
		5'-CCCCTCTAGAGGTCACTGTGCCCGGCTGTCGGCGGCGGCCCCGCGCATGG-3'
tylAII		5'-CCCCTCTAGAGGTCATGCGCGCTCCAGTTCCCTGCCGCCCGGGGACCGCTTG-3'
		5'–GGGTCTAGATCGATTAATTAAGGAGGACATTCATGCGCGTCCTGGTGACCGGAGGT
		GCGGGCTTCATCGGCTCGCACTTCA-3'
tylMIII	TylM31	5'-GGCGGGGAGAGAGGAGAGCATATGAACACGGCAGCCGGCCCGACC-3'
	TylM32	5'-CCCCCTCTAGAGGTCACTCGGGGACATACGGGGCGACGGGCAGCCG-3'
tylB	TylB1	5'-CCCCCTCTAGATCTTAATTAAGGAGGACACCCATGACAGGGCTGCCGCGGCCCGCCGTCCGGGTG-3'
	TylB2	5'-GGGGGTCTAGAGGTCACGGGCCTTCCTCCCAGGAGTCCAGCGCGGCGGA-3'
tyl1a	BIOSG 88	5'-GGGCATATGGCGGCGAGCACTACGACGGAGGGGAATGT-3'
	BIOSG 89	5'-GGGTCTAGAGGTCACGGGTGGCTCCTGCCGGCCCTCAG-3'
tylMI	TylMI1	5'-GGGGGTCTAGATCTTAATTAAGGAGGACAACCATGGCCCATTCATCCGCCACGGCCGGACCGCAGG
		CCGA-3'
	TylMI2	5'-GGGGGTCTAGAGGCATATGTGTCCTCCTTAATTAATCACCGGGTTTCTCCCTTCGCTCCGGGGA
		GCCCGGT-3'
angAI	BIOSG73	5'-GGGCATATGAAGGGC ATCATCCTGGCGGGCGGCAGCGGC-3'
	BIOSG74	5'-GGTCTAGAGGTCAT GCGGCCGGTCCGGACATGAGGGTCTCCGCCAC-3'
angAII	BIOSG71	5'-GGGCATATGCGGCTG CTGGTCACCGGAGGTGCGGGC-3'
	BIOSG72	5'-GGTCTAGAGGTCAGTCG GTGCGCCGGGCCTCCTGCG-3'

angMIII	BIOSG61	5'-GGGCATATGAGCCCCGCACCCGCCACCGAGGACCC-3'
	BIOSG62	5'-GGTCTAGAGGTCAGTTCCGCGGTGCGGTGGCGGGCAGGTCAC-3'
spnO	BIOSG41	5'-GGGCATATGAGCAGTTCTGTCGAAGCTGAGGCAAGTG-3'
	BIOSG42	5'-GGTCTAGAGGTCATCGCCCCAACGCCCACAAGCTATGCAGG-3'
angorf14	BIOSG69	5'-GGGCATATGGTGAACGATCCGATGCCGCGCGGCAGTGGCAG-3'
	BIOSG70	5'-GGTCTAGAGGTCAACCTCCAGAGTGTTTCGATGGGGTGGTGGG-3'
angB	BIOSG67	5'-GGGCATATGACTACCTACGTCTGGGACTACCTGGCGG-3'
	BIOSG68	5'-GGTCTAGAGGTCAGAGCGTGGCCAGTACCTCGTGCAGGGC-3'
angMI	BIOSG65	5'-GGGCATATGAACCTCGAATACAGCGGCGACATCGCCCGGTTG-3'
	BIOSG66	5'-GGTCTAGAGGTCAGGCCTGGACGCCGACGAAGAGTCCGCGGTCG-3'
angorf4	BIOSG75	5'-GGGCATATGAGCACCCCTTCCGCACCACCCGTTCCG-3'
	BIOSG76	5'-GGTCTAGAGGTCAGTACAGCGTGTGGGCACACGCCACCAG-3'
eryCVI	BIOSG28	5'-GGGCATATGTACGAGGGCGGGTTCGCCGAGCTTTACGACC-3'
	BIOSG29	5'-GGGGTCTAGAGGTCATCCGCGCACACCGACGAACAACCCG-3'

References

1 S. Gaisser, R. Lill, J. Staunton, C. Méndez, J. Salas and P.F. Leadlay, *Mol. Microbiol.*, 2002, **44**, 771-781.

2 S. Gaisser, J. Reather, G. Wirtz, L. Kellenberger, J. Staunton and P.F. Leadlay, *Mol. Microbiol.*, 2000, **36**, 391-401.

3 K. Salah-Bey, M. Doumith, J. -M. Michel, S. Haydock, J. Cortés, P. F. Leadlay and M.-C. Raynal, 1998, *Mol. Gen. Genet.*, **257**, 542-553.

4 S. Gaisser, G. A. Böhm, M. Doumith, M. -C. Raynal, N. Dhillon, J. Cortés and P. F. Leadlay, *Mol. Gen. Genet.*, 1998, **258**, 78-88.

5 S. Gaisser, G. A. Böhm, J. Cortés and P. F. Leadlay, *Mol. Gen. Genet.*, 1997, **256**, 239-251.

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

7 C. J. Rowe, J. Cortés, S. Gaisser, J. Staunton and P. F. Leadlay, 1998, *Gene* **216**, 215-223.

8 H. Yamamoto, K. H. Maurer, C. R. Hutchinson. *J. Antibiot.*, 1986, **34**, 1304-1313.