

# Parallelized gene cluster editing illuminates mechanisms of epoxyketone proteasome inhibitor biosynthesis

Chuan Huang<sup>1,2,4,5,\*</sup>, Daniel Zabala<sup>1</sup>, Emmanuel L.C. de los Santos<sup>1,2,3</sup>, Lijiang Song<sup>1</sup>, Christophe Corre<sup>1,2,3</sup>, Lona M. Alkhalaf<sup>1</sup> and Gregory L. Challis<sup>1,2,4,5,\*</sup>

<sup>1</sup>Department of Chemistry, University of Warwick, Coventry CV4 7AL, UK, <sup>2</sup>Warwick Integrative Synthetic Biology Centre, University of Warwick, Coventry CV4 7AL, UK, <sup>3</sup>School of Life Sciences, University of Warwick, Coventry CV4 7AL, UK, <sup>4</sup>Biomedicine Discovery Institute, Department of Biochemistry and Molecular Biology, Monash University, Clayton, Victoria 3800, Australia and <sup>5</sup>ARC Centre of Excellence for Innovations in Peptide and Protein Science, Monash University, Clayton, Victoria 3800, Australia

Received August 08, 2022; Revised December 22, 2022; Editorial Decision December 23, 2022; Accepted January 06, 2023

## ABSTRACT

Advances in DNA sequencing technology and bioinformatics have revealed the enormous potential of microbes to produce structurally complex specialized metabolites with diverse uses in medicine and agriculture. However, these molecules typically require structural modification to optimize them for application, which can be difficult using synthetic chemistry. Bioengineering offers a complementary approach to structural modification but is often hampered by genetic intractability and requires a thorough understanding of biosynthetic gene function. Expression of specialized metabolite biosynthetic gene clusters (BGCs) in heterologous hosts can surmount these problems. However, current approaches to BGC cloning and manipulation are inefficient, lack fidelity, and can be prohibitively expensive. Here, we report a yeast-based platform that exploits transformation-associated recombination (TAR) for high efficiency capture and parallelized manipulation of BGCs. As a proof of concept, we clone, heterologously express and genetically analyze BGCs for the structurally related nonribosomal peptides epone mycin and TMC-86A, clarifying remaining ambiguities in the biosynthesis of these important proteasome inhibitors. Our results show that the epone mycin BGC also directs the production of TMC-86A and reveal contrasting mechanisms for initiating the assembly of these two metabolites. Moreover, our

data shed light on the mechanisms for biosynthesis and incorporation of 4,5-dehydro-L-leucine (dhL), an unusual nonproteinogenic amino acid incorporated into both TMC-86A and eponemycin.

## INTRODUCTION

Rapid progress in DNA sequencing and bioinformatics has revealed opportunities as well as challenges for studying the biosynthesis of structurally complex specialized metabolites in microorganisms (1,2). To overcome some of the challenges, such as genetic intractability, complicated systems for regulation of gene expression, and complex culture conditions that are difficult to reproduce, expression of biosynthetic gene clusters (BGCs) in well-characterized heterologous hosts is often employed (3,4). This requires effective methods for cloning and genetic manipulation of BGCs.

Several approaches for cloning whole BGCs have been developed, including direct capture from genomic DNA (e.g. using transformation-associated recombination (TAR), RecET recombineering, Cas9-mediated targeting of chromosome segments and *in vitro*  $\lambda$  packaging and ligation) (5–8) and BGC assembly (e.g. DNA assembler and Direct Pathway Cloning) (9,10). Among these methods, TAR cloning is widely used for direct capture of chromosomal fragments from diverse species into a circular artificial chromosome that can propagate, segregate, and be selected for in yeast (11). To facilitate the heterologous expression of directly cloned bacterial BGCs in *Streptomyces* hosts, the Moore lab developed a yeast-*Escherichia coli*-*Streptomyces* shuttle vector, pCAP01 which can be employed to capture a specific BGC via homologous recombination with two

\*To whom correspondence should be addressed. Tel: +44 02476 574024; Email: g.l.challis@warwick.ac.uk  
Correspondence may also be addressed to Chuan Huang. Tel: +61 03 9905 1750; Email: chuan.huang@monash.edu  
Present addresses:

Daniel Zabala, Faes Farma, Bilbao, Spain.

Emmanuel L.C. de los Santos, UCB Celltech, Slough, Berkshire SL1 3WE, UK.

~1 kb flanking sequences (5). One disadvantage of this method is the high background level of plasmid recircularization via non-homologous end joining. To solve this problem and simplify vector construction, pCAP03 was developed. This utilizes two ~50 bp BGC flanking sequences inserted between a counter-selectable marker (*URA3*) and its promoter (*pADHI*) (12). However, a potential drawback of this method is that the short flanking sequences may lower the homologous recombination efficiency for some BGCs.

In-frame gene deletion is a powerful approach for elucidating the biosynthetic functions of proteins encoded by BGCs (13–18). However, most methods for constructing such deletions involve several time-consuming and labor-intensive steps, necessitating the serial construction and analysis of numerous mutants to investigate the function of all proteins encoded by a typical BGC. The development of parallelizable methods for construction and analysis of gene deletions would significantly accelerate our understanding of specialized metabolite biosynthesis.

Early approaches to the construction of in-frame gene deletions involved two sequential homologous recombination events in the native metabolite producer using marker-assisted selection and screening (19), which is a slow and laborious process. More recently, tools for efficient recombination in *E. coli* (e.g.  $\lambda$ -Red-mediated recombination) have streamlined the process of creating such mutants. Using a PCR-generated resistance cassette, genes in cosmids containing partial BGCs can be replaced in a single step (20). The engineered cosmids can be used to replace the corresponding genes on the chromosomes of native metabolite producers via double homologous recombination and the resistance gene can be excised from the resulting mutants to create in-frame deletions containing a ‘scar’ sequence (20). In addition to being difficult to parallelize, these methods are only applicable to genetically tractable native producers, or cloned BGCs that have already been incorporated into a heterologous host.

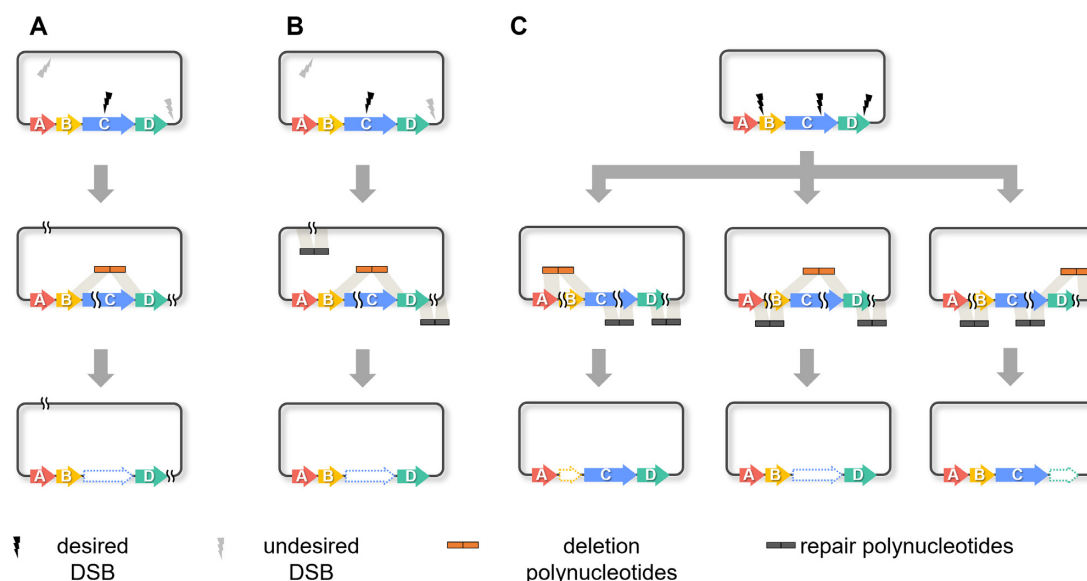
The ability to routinely clone complete BGCs enables the creation of in-frame deletions prior to expression in a heterologous host, which is easier to parallelize. TAR in yeast has been combined with selectable markers to swap targeted regions (5,21). Moreover, double-strand break (DSB)-stimulated recombination, initiated using naturally occurring or artificially introduced unique restriction sites, enables various modifications of cloned BGCs, including promoter engineering, gene deletion and targeted mutation (22,23). Site-specific DSBs can also be created *in vivo* or *in vitro* using the programmable CRISPR/Cas9 endonuclease editing system. This enables multiplexed creation of unmarked gene deletions (CREasPy-Cloning), promoter engineering (mCRISTAR, miCASTAR) and site-directed mutagenesis in clones containing whole BGCs (24–28). However, CRISPR/Cas9-based methods are often hampered by off-target effects that yield unpredictable DSBs and/or require tedious processes for multiplexing, such as the manufacture and introduction of customized gRNA expression plasmids for each *in vivo* modification, or optimization of multiple parameters for efficient digestion at each site *in vitro*. Direct assembly of overlapping DNA fragments in yeast offers an alternative approach for multiplexed BGC manipulation. However, this method becomes prohibitively

expensive for large BGCs and unintended mutations resulting from errors in DNA synthesis must be excluded prior to assembly (29). Thus, despite considerable promise, yeast has yet to be developed into a widely applicable, operationally simple, and cost-effective platform for cloning and parallelized editing of complete BGCs.

Here, we describe the development of a yeast-based platform for efficient capture and genetic manipulation of complete BGCs. Key features of this platform include: (i) the use of commercially available restriction endonucleases to create a predictable set of DSBs in vectors containing captured BGCs; (ii) efficient TAR-mediated reassembly of digested vectors into derivatives containing in-frame deletions in target genes using different combinations of repair and deletion polynucleotides and (iii) modest amounts of PCR screening (typically <20 colonies) to identify clones with the desired genotype (Figure 1A and B). The multiplexability of the restriction digestion-reassembly cascade (Figure 1C), coupled with the operational simplicity of the overall process, greatly facilitates parallelized construction of BGC variants.

As a proof of concept, we have applied this platform to BGCs directing the biosynthesis of eponemycin (1) and TMC-86A (2), two members of a group of epoxyketone proteasome inhibitors that inspired the development of the multiple myeloma drug Carfilzomib and the orally bioavailable derivative Oprozomib, in addition to selective inhibitors of the immunoproteasome for the treatment of autoimmune diseases (Figure 2A) (30,31). Epoxyketones have been isolated from various microbes, including Actinobacteria, Cyanobacteria, Myxobacteria and environmental microbiomes (Figure 2A) (32–37). The epoxyketone moiety specifically targets the catalytic threonine residue in the  $\beta$ -subunit of the proteasome (32,33,38,39). Very similar BGCs for eponemycin (1) and TMC-86A (2), which bear a close structural relationship, have been identified in *Streptomyces hygroscopicus* ATCC 53709 and *Streptomyces chromofuscus* ATCC 49982 respectively (Figure 2A and B) (40,41). Previous *in vitro* reconstitution experiments showed that the epoxyketone warhead of TMC-86A and eponemycin is assembled from an  $\alpha$ -dimethyl- $\beta$ -keto acid biosynthetic intermediate by the unique flavin-dependent decarboxylase-desaturase-monooxygenase TmcF/EpnF and undergoes subsequent hydroxylation by the cytochrome P450 (CYP) TmcI/EpnI (40).

The biosynthesis of TMC-86A is proposed to commence with the condensation of a butanoyl thioester with L-Ser and L-Leu, catalyzed by the nonribosomal peptide synthetase (NRPS) TmcG (40). The resulting *N*-acyl-dipeptidyl thioester is then elongated with a malonyl extender unit and dimethylated by the polyketide synthase (PKS) TmcH (Figure 2C), prior to thioesterase domain-mediated hydrolysis from the PKS to produce the  $\alpha$ -dimethyl- $\beta$ -keto acid intermediate (40). Eponemycin is proposed to be biosynthesized via an analogous pathway (Figure 2C), except that a 6-methylheptanoyl thioester is used as the starter unit (42,43). However, the mechanisms responsible for incorporation of different starter units and the common dhL residue into TMC-86A and eponemycin remain unclear, notwithstanding previous efforts to genetically characterize the eponemycin BGC (42,43). Our results demon-



**Figure 1.** Scheme highlighting the benefits of the approach described here over other methods for yeast-mediated editing of cloned BGCs. (A) Linearization of the vector using a unique restriction site in the target gene, or CRISPR-Cas9, enables an in-frame deletion to be created by co-transformation with a deletion polynucleotide (23). However, in practice most genes in a BGC do not contain a unique restriction site and CRISPR-Cas9 can be hampered by off-target effects leading to additional undesired DSBs at unpredictable locations. These problems limit the general applicability of this approach. (B) For genes lacking a unique restriction site, a DSB is created in the target gene along with a handful of undesired DSBs at other sites; however, their location is predictable. Thus, co-transformation of yeast with the digested vector, a deletion polynucleotide, and several repair polynucleotides enables an in-frame deletion to be introduced into every gene in the BGC. (C) The same restriction site is often present in several different genes, enabling several deletion mutants to be created from a vector digested with a single endonuclease by multiplexed reassembly.

strate that both BGCs direct the biosynthesis of TMC-86A, using the ketoacyl-ACP synthase III (KASIII) homologue TmcD/EpnD and the acyl carrier protein (ACP) TmcE/EpnE to generate the butanoyl thioester starter unit. Eponemycin and related congeners are shunt metabolites that appear to result from hijacking of alternative starter units from the primary metabolic fatty acid synthase (FAS). Moreover, they show that desaturation of the L-Leu side chain precedes eponemycin assembly and involves the CYP EpnK and the NRPS EpnJ.

## MATERIALS AND METHODS

### Bacterial strains, plasmids and culture conditions

The strains and plasmids used in this study are listed in Supplementary Table S1. *E. coli* Top10 (Invitrogen), used for routine cloning, was grown in Luria-Bertani medium (tryptone 1%, yeast extract 0.5%, NaCl 1%) at 37°C with appropriate antibiotic selection when necessary (ampicillin 100 µg/ml, apramycin 50 µg/ml, kanamycin 25 µg/ml or chloramphenicol 25 µg/ml).

*S. hygrosopicus* ATCC53709, *S. chromofuscus* ATCC49982, and *Streptomyces albus* J1074 were grown on either TSB medium (Becton Dickinson) or SFM medium (19) at 30°C for routine cultivation. For epoxyketone production, 20 µl spore suspension was inoculated on 25 ml R5 agar medium (19) and incubated at 30°C for 7 days. For the dhL feeding experiment, substrate was dissolved in 1 ml water and overlaid on the agar plates after 48 h from culture initiation (10 mM final concentration).

*Saccharomyces cerevisiae* VL6 – 48N was grown on YPD medium (glucose 2%, yeast extract 1%, peptone 2%)

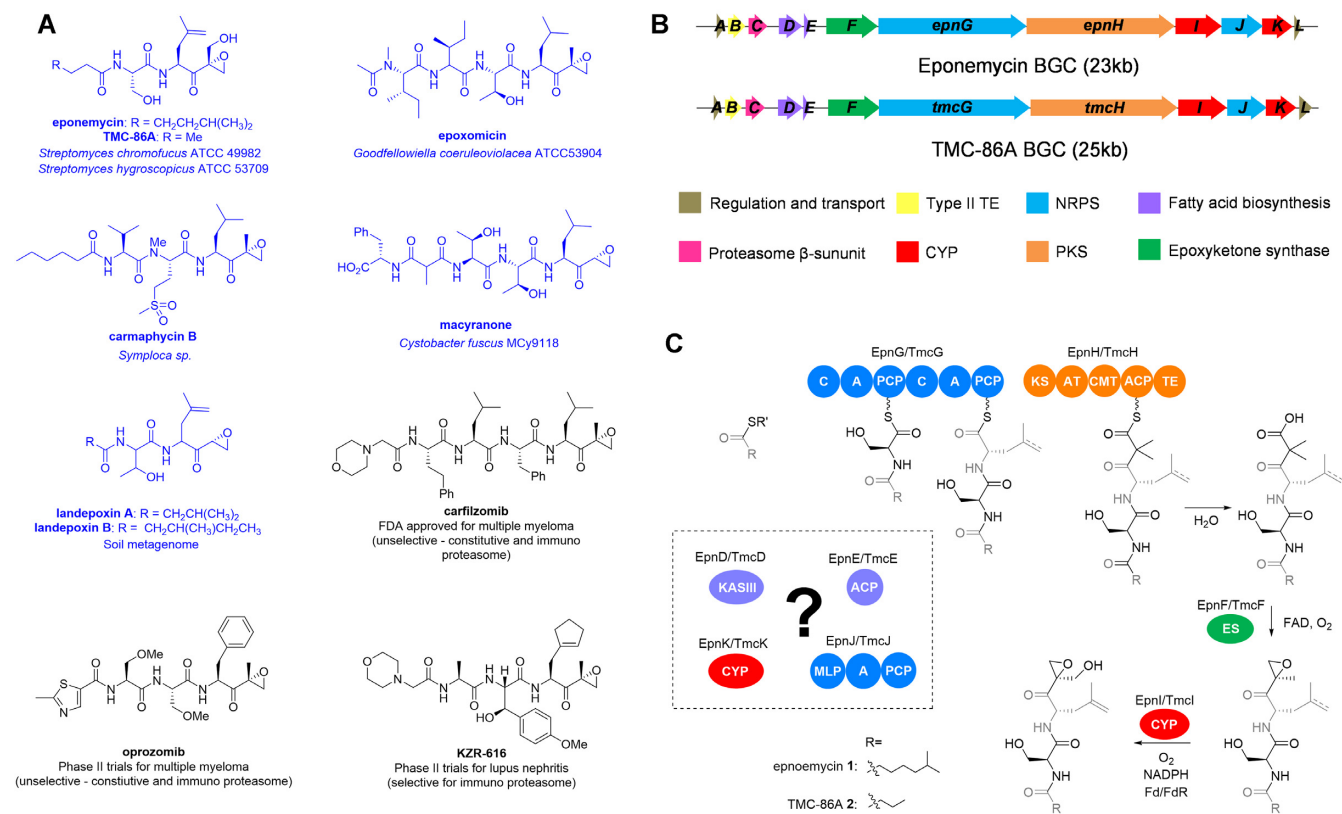
supplemented with 100 mg/l adenine at 30°C. For TAR cloning, tryptophan-deficient medium (sorbitol 18.2%, glucose 2.2%, yeast nitrogen base without amino acid and ammonium sulfate 0.17%, yeast synthetic drop-out medium supplements without tryptophan 0.19%, ammonium sulfate 0.5% and agar 2%) was used with 5-Fluoroorotic acid (1 mg/ml) when necessary (12).

### Construction of plasmids

DNA isolation and manipulation was carried out according to standard methods for *E. coli* and *Streptomyces* (19,44). To generate the *pADH1-URA3* cassette for pCAP1000, *pADH1* and *URA3* genes were amplified from pCAP03 separately and joined using overlap PCR (see primer list in Supplementary Table S2). For pCAP1000 linearization before yeast transformation, BamHI, PmeI and NdeI restriction sites were introduced between *pADH1* and *URA3*. The *pADH1-URA3* cassette was sub-cloned into the pCR™2.1 vector for verification. After verification by DNA sequencing, the *pADH1-URA3* cassette was amplified (see Primers List in Supplementary Table S2) and cloned into SpeI and XhoI-digested pCAP01 using Gibson Assembly to generate pCAP1000.

For pCAP03-based vectors, homologous arms were synthesized and cloned into pCAP03 by GENEWIZ. For pCAP1000 and pCAP01-based vectors, homologous arms for each clusters were amplified from genomic DNA of each wild type strain using Phusion® High-Fidelity DNA polymerase (New England Biolabs), with ~15 bp primer-introduced overhangs for Gibson Assembly/GeneArt™ Seamless Cloning and Assem-





**Figure 2.** Structures, origins, and therapeutic applications of epoxyketone proteasome inhibitors, and comparison of the biosynthetic gene clusters and proposed biosynthetic pathways for eponemycin and TMC-86A. (A) Structures and sources of natural epoxyketones (in blue), and structures and applications of synthetic epoxyketones (in black) inspired by the natural products. (B) Comparison of the *S. hygroscopicus* eponemycin and *S. chromofuscus* TMC-86A BGCs. (C) Previously proposed biosynthetic pathways for eponemycin **1** and TMC-86A **2** in *S. hygroscopicus* and *S. chromofuscus* respectively. The variable *N*-acyl group and conserved dhL residue in eponemycin and TMC-86A, which are assembled via uncharacterized mechanisms, are shown in grey. The dashed 4,5-double bond in the dhL residue indicates that the timing of its introduction is unknown. The dashed box contains putative biosynthetic enzymes with experimentally unvalidated functions encoded by the eponemycin and TMC-86A BGCs. R' = ACP or CoA.

bly. The counter-selectable cassettes were amplified from pCR-ADHURA3 or pCAP1000 using Phusion<sup>®</sup> High-Fidelity DNA polymerase, with ~15 bp primer-introduced overhangs for Gibson Assembly/GenArt<sup>™</sup> Seamless Cloning and Assembly. All paired homologous arms were cloned into SpeI/XhoI-digested pCAP01 (without counter-selectable cassettes) and SpeI/XhoI-digested pCAP1000 (together with counter-selectable cassettes) to create the following plasmids: pCAP01epn and pCAP1000epn for epnBGC capture; pCAP01tmc and pCAP1000tmc for tmcBGC capture. The integrity of all plasmids was confirmed by restriction digestion and sequencing.

For genetic complementation experiments, target genes were amplified from genomic DNA using Phusion DNA polymerase and cloned into pOSV556t via HindIII/StuI restriction sites to generate pOSV-epnD, pOSV-epnE, pOSV-epnDE, pOSV-tmcD, pOSV-epnI and pOSV-epnK. All constructs were verified by restriction digestion and sequencing.

#### Yeast transformation and *Streptomyces/E. coli* conjugation

To create appropriate genomic fragments for TAR cloning of BGCs, genomic DNA isolated from the eponemycin and

TMC-86A producing organisms was digested with suitable restriction enzymes (PstI for the eponemycin BGC, and NdeI and XhoI for the TMC86A BGC), and purified by ethanol precipitation. 2–4  $\mu$ g of digested genomic DNA were used for yeast spheroplast transformation, together with 0.1–0.2  $\mu$ g of the appropriate PmeI-linearized capture vector. For gene editing, plasmids were digested with appropriate restriction enzymes that cut within or adjacent to target genes and purified by ethanol precipitation. 0.1–0.2  $\mu$ g of digested plasmid and 0.05–0.1  $\mu$ g of the requisite repair and deletion polynucleotides were used to co-transform yeast spheroplasts. Yeast spheroplast preparation and transformation were carried out using a previously reported protocol (11). After spheroplast transformation, yeast colonies were picked and screened by PCR using a previously reported protocol with primers designed to amplify fragments from the middle and each end of the target BGC (11). Plasmids were isolated from PCR-positive yeast colonies and used to transform *E. coli* Top10. After verification by restriction digestion, plasmids harboring target BGCs or their derivatives were introduced into *Salbus* J1074 by triparental mating from *E. coli* ET12567, via the helper strain *E. coli* ET12567/pUB307, using a standard protocol (45). Kanamycin resistance was used for ex-conjugant screening.

Genetic complementation plasmids were introduced into *S. albus*/epnBGC- $\Delta$ *epnD* (pOSV-*epnD*), *S. albus*/epnBGC- $\Delta$ *epnE* (pOSV-*epnE*), *S. albus*/epnBGC- $\Delta$ *epnDE* (pOSV-*epnDE*), *S. albus*/tmcBGC- $\Delta$ *tmcD* (pOSV-*tmcD* and pOSV-*epnD*), *S. albus*/epnBGC- $\Delta$ *epnI* (pOSV-*epnI*), *S. albus*/epnBGC- $\Delta$ *epnK* (pOSV-*epnK*) and *S. albus*/epnBGC- $\Delta$ *epnJ* (pOSV-*epnJ*) via conjugal transfer from *E. coli* ET12567/pUZ8002. Hygromycin was used for transconjugant selection.

### Growth of *Streptomyces* strains and analysis of epoxyketone production

After 7 days incubation, the agar was cut into small chunks and 15 ml of ethyl acetate was added to each plate. After 2 h, the organic extracts were decanted and concentrated by rotary evaporation. The residues were separately dissolved in 1 ml of methanol, and 2  $\mu$ l of each sample were analyzed by UHPLC-ESI-Q-TOF-MS/MS on a Bruker MaXis II mass spectrometer (or Bruker MaXis Impact mass spectrometer) coupled to a Dionex UltiMate 3000 UHPLC fitted with an Agilent Zorbax Eclipse Plus C18 column (100  $\times$  2.1 mm, 1.8  $\mu$ m). Using a flow rate of 0.2 ml/min, the column was eluted with a combination of water and acetonitrile as follows: 5% (v/v) acetonitrile for 5 min, 5–100% (v/v) acetonitrile over 21 min, 100% (v/v) acetonitrile for 3 min, 100–5% (v/v) acetonitrile over 2 min and 5% (v/v) acetonitrile for 3 min.

## RESULTS

### Cloning and heterologous expression of the eponemycin and TMC-86A BGCs

To capture the eponemycin and TMC-86A BGCs, we initially constructed pCAP03-based vectors containing 50 bp flanking sequences (pCAP03epnBGC and pCAP03tmcBGC, respectively). However, no clones were obtained for either cluster using these vectors, despite several attempts (Supplementary Table S3). Reasoning that low homologous recombination efficiency, due to the short flanking sequences, was the cause of this failure, we next attempted to capture the two clusters using pCAP01 containing 1 kb flanking sequences. While we succeeded in cloning the eponemycin BGC using this approach, a high level of plasmid recircularization was observed, and we still failed to clone the TMC-86A BGC (Supplementary Table S3).

To surmount these problems, we sought to use 1 kb flanking sequences in conjunction with the *URA3* counter selectable marker. However, this was not possible in pCAP03 because the flanking sequences are inserted between the *URA3* gene and its promoter (*pADHI*) (12,21,46,47), which limits their combined length to 130 bp. We therefore cloned a modified *pADHI-URA3* cassette into pCAP01 to create pCAP1000 (Figure 3A). The 1kb flanking sequences initially inserted in pCAP01 were then cloned into pCAP1000 either side of the *pADHI-URA3* cassette, generating pCAP1000epn and pCAP1000tmc, respectively. Using these vectors, we succeeded in cloning both clusters and a low background of plasmid recircularization was observed (four out of six clones correct using pCAP1000epn

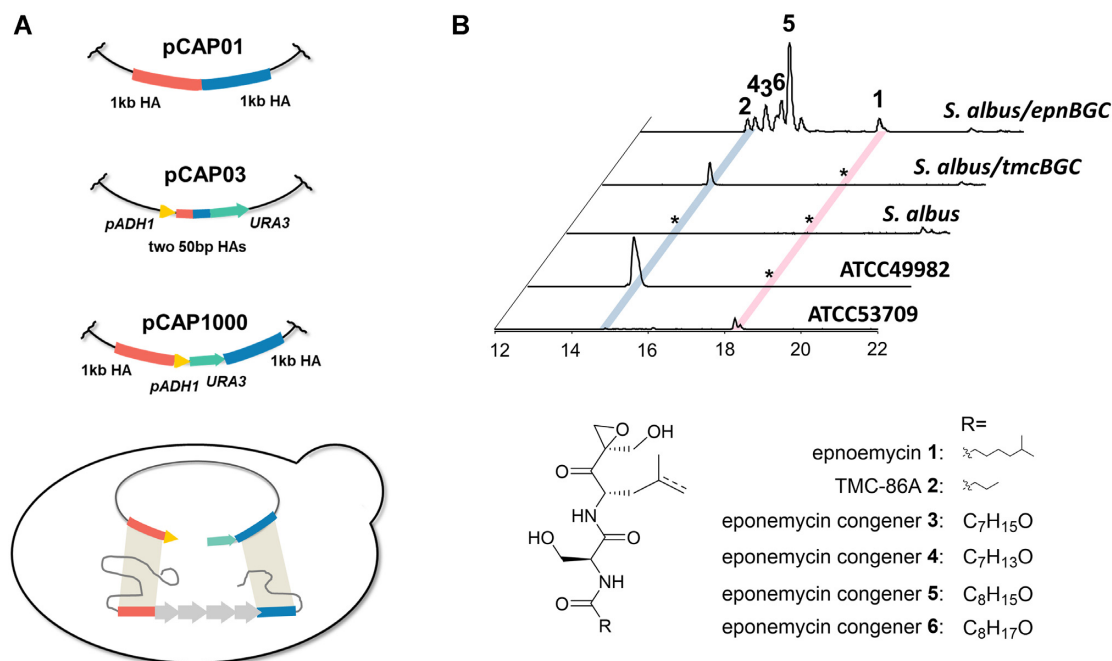
and all four clones correct using pCAP1000tmc; see Supplementary Table S3). The inserts of the resulting plasmids, pCAP1000epnBGC and pCAP1000tmcBGC-initial, were validated using restriction enzyme digests (Supplementary Figure S1).

The plasmids were integrated into the chromosome of *S. albus* J1074 via triparental conjugation (48), and the resulting strains were cultivated for 7 days at 30°C on R5 agar medium. UHPLC-ESI-Q-TOF-MS analysis of ethyl acetate culture extracts identified six metabolites in the strain containing pCAP1000epnBGC that were absent from wild type *S. albus*. Two of these were identified as eponemycin **1** and TMC-86A **2**, based on molecular formulae deduced from high resolution mass spectra and retention time comparisons with authentic standards produced by *S. hygroscopicus* ATCC 53709 and *S. chromofuscus* ATCC 49982, respectively (Figure 3B and Supplementary Figure S3). The remaining four metabolites yielded  $[M+H]^+$  ions with  $m/z = 413.2286$  (calculated for  $C_{20}H_{33}N_2O_7^+$ : 413.2282),  $m/z = 415.2444$  (calculated for  $C_{20}H_{35}N_2O_7^+$ : 415.2439),  $m/z = 427.2440$  (calculated for  $C_{21}H_{35}N_2O_7^+$ : 427.2439) and  $m/z = 429.2599$  (calculated for  $C_{21}H_{37}N_2O_7^+$ : 429.2595) and were proposed to be eponemycin congeners **3**, **4**, **5** and **6** with alterations to the N-acyl group (Supplementary Figure S2). MS/MS analyses were consistent with these structural assignments (Supplementary Figure S3).

In contrast, no new metabolites were produced by *S. albus* containing pCAP1000tmcBGC-initial. Close inspection of the insert captured in this plasmid indicated that the promoter region upstream of *tmcA* (which encodes a putative LuxR-like transcriptional activator) is truncated, likely affecting the expression of this gene (40,43). Reasoning that TmcA may function as a transcriptional activator of the downstream biosynthetic genes, we amplified 995 bp from the intergenic region upstream of *tmcA* flanked by 50 bp arms with homology to the 5' end of the insert and the adjacent vector backbone. A DSB upstream of the *tmcA* start codon was generated using a unique XhoI restriction site and yeast was co-transformed with the linearized plasmid and the PCR product, leading to incorporation of the 995 bp intergenic region into pCAP1000tmcBGC-initial. The resulting construct (named pCAP1000tmcBGC) was integrated into the chromosome of *S. albus* and, after 7 days cultivation at 30°C on R5 agar medium, UHPLC-ESI-Q-TOF-MS analysis of ethyl acetate extracts showed that this strain produces TMC-86A **2** (Figure 3B). Interestingly, neither eponemycin **1**, nor its congeners **3**, **4**, **5** and **6** were detected in the extracts (Figure 3B).

### Parallelized construction of in-frame deletions in *epnD*, *epnE*, *epnI*, *epnJ* and *epnK*

Previous attempts to elucidate the roles in eponemycin biosynthesis played by *epnD*, *epnE*, *epnI*, *epnJ* and *epnK* via construction and analysis of the corresponding in-frame deletion mutants failed to provide useful insights (43). During resequencing of the eponemycin BGC cloned into pCAP1000, we identified several errors in the originally reported sequence (41). In particular, cytosine and guanine residues are missing from the originally reported sequences



**Figure 3.** Cloning and heterologous expression of the eponemycin and TMC-86A BGCs using pCAP1000. (A) Comparison of homologous arms (HAs) and counter-selectable markers incorporated into pCAP01, pCAP03 and pCAP1000. The two HAs are shown in red and blue, respectively. The pADH1 promoter is shown in yellow and URA3 is shown in green. The combination of 1kb HAs and a counter-selectable marker in pCAP1000 improves the efficiency of BGC capture via double homologous recombination, which is illustrated schematically below the vector comparison. (B) Extracted ion chromatograms at  $m/z = 421.2309 \pm 0.002$ ,  $365.1683 \pm 0.002$ ,  $437.2258 \pm 0.002$ ,  $435.2102 \pm 0.002$ ,  $451.2415 \pm 0.002$  and  $449.2258 \pm 0.002$ , corresponding to  $[M + Na]^+$  for eponemycin 1, TMC-86A 2, and eponemycin congeners 3, 4, 5 and 6, respectively, from UHPLC-ESI-Q-TOF-MS analysis of culture extracts of *S. albus* with and without pCAP1000epnBGC and pCAP1000tmcBGC integrated into its chromosome, the eponemycin producer *S. hygrosopicus* ATCC53709, and the TMC-86A producer *S. chromofuscus* ATCC49982.

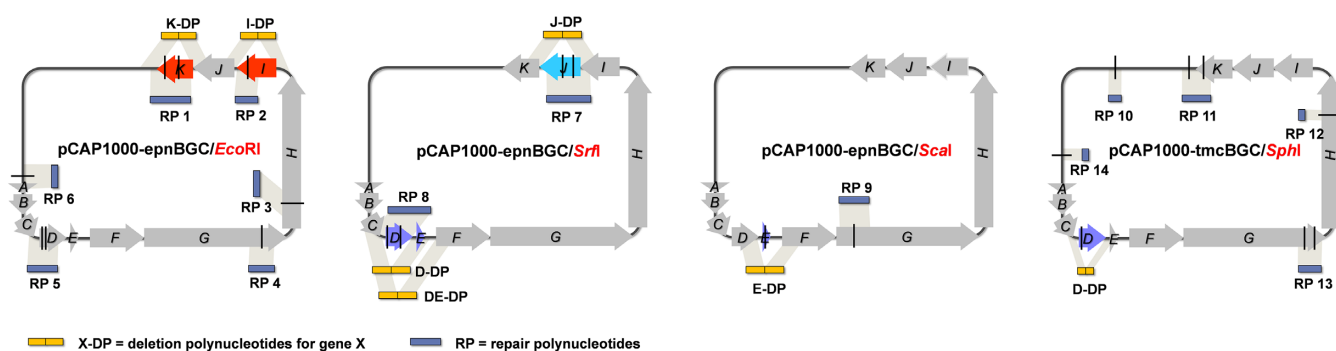
of *epnD* and *epnK*, respectively. These errors resulted in frame shifts in the originally reported *epnD* and *epnK* sequences that eliminated a stop codon 27 nucleotides downstream of the originally assigned start codon and introduced a stop codon 11 nucleotides upstream of the true stop codon, respectively (Supplementary Figure S4). Thus, *epnD* is 27 bp shorter and *epnK* is 15 bp longer than originally assigned. These errors likely contributed to the failure of the previous genetic studies to establish clear roles for several of the eponemycin biosynthetic genes.

To reevaluate the roles of *epnD*, *epnE*, *epnI*, *epnJ* and *epnK* in eponemycin biosynthesis we designed a yeast-based method for parallelized construction of in-frame deletions in these five genes. Three different restriction enzymes were employed to create one or two DSBs in each gene: EcoRI for *epnI* and *epnK*, SrfI for *epnJ* and *epnD*, and ScaI for *epnE* (Figure 4). These enzymes introduced DSBs at up to six other locations in pCAP1000epnBGC, resulting in a total of eight, four and two fragments from the EcoRI, SrfI and ScaI digests, respectively. 1 kb repair polynucleotides that symmetrically span the DSB (or pair of DSBs if they are within 1 kb of each other) were amplified from pCAP1000epnBGC for each off target cut site and 1 kb deletion polynucleotides for each target gene were assembled by overlap PCR using two 500 bp amplimers from the 5' and 3' flanking regions. The fragments from each restriction digest were mixed with the appropriate combination of repair and deletion polynucleotides and used

to transform yeast. 60–100% of transformants resulting from the mixtures containing the SrfI and ScaI restriction fragments harboured reassembled plasmids with the desired genotype (Supplementary Table S4). Reassembly efficiency was reduced when the EcoRI restriction fragments were used (13–15% positives—see Supplementary Table S4), but only a modest amount of screening was needed to identify a correct clone. The resulting plasmids, named pCAP1000epnBGC- $\Delta$ *epnX* (where X is the identifying letter for each gene), contained 1188, 984, 186, 1080 and 1509 bp in-frame deletions in *epnI*, *epnD*, *epnE*, *epnK* and *epnJ*, respectively. The genotypes of these plasmids were confirmed by restriction digestion and sequencing, and no adventitious mutations were detected (Supplementary Figure S1). Each plasmid was separately transferred into *S. albus* via conjugation from *E. coli*.

To further confirm the fidelity of plasmid reassembly, we initially investigated the production of eponemycin-related metabolites by the *epnI* mutant. The *epnI* gene encodes a protein with 89% identity to TmI from *S. chromofuscus*, a CYP that catalyzes the final hydroxylation step in TMC-86A biosynthesis (Figure 2B) (40). UHPLC-ESI-Q-TOF-MS analysis of the *epnI* mutant showed that the production of both TMC-86A and eponemycin was abolished and two new products, with molecular formulae corresponding to deshydroxy-eponemycin 11 ( $[M+H]^+m/z$  383.2525, calculated for  $C_{20}H_{35}N_2O_5^+$ : 383.2540) and deshydroxy-TMC-86A 12 ( $[M+H]^+m/z$  327.1914, calculated for



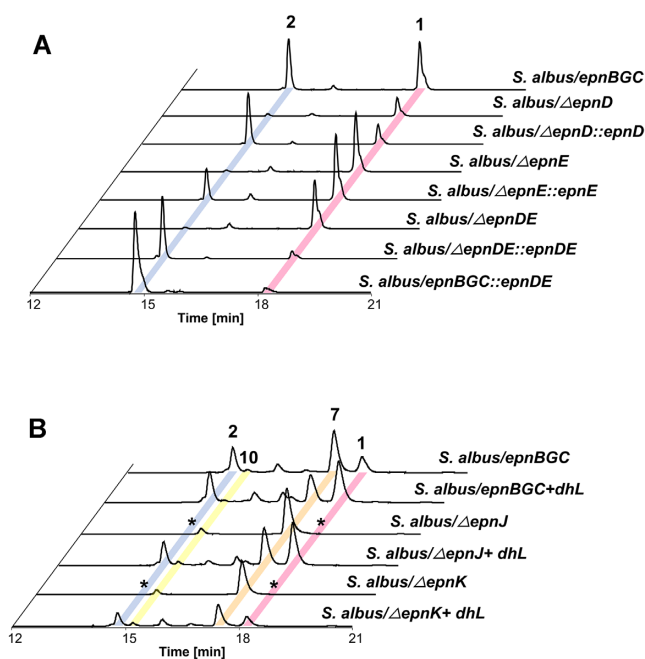


**Figure 4.** Scheme illustrating parallelized construction of in-frame deletions in the eponemycin and TMC-86A BGCs using homology-directed DSB repair in yeast. DSBs are generated by restriction enzyme digestion *in vitro*. Using different combinations of repair polynucleotides (RP) and deletion polynucleotides (DP), the fragmented plasmids can be reassembled in yeast to create a series of in-frame deletion mutants.

$C_{16}H_{27}N_2O_5^+$ : 327.1914) were produced (Supplementary Figure S3). Moreover, production of the eponemycin congeners **3**, **4**, **5** and **6** was also abolished and new compounds giving rise to  $[M+H]^+$  ions with  $m/z = 397.2325$  (calculated for  $C_{20}H_{33}N_2O_6^+$ : 397.2333), 399.2484 (calculated for  $C_{20}H_{35}N_2O_6^+$ : 399.2490), 411.2483 (calculated for  $C_{21}H_{35}N_2O_6^+$ : 411.2490) and 413.2639 (calculated for  $C_{21}H_{37}N_2O_6^+$ : 413.2646), consistent with molecular formulae for the deshydroxy derivatives of the congeners **13**, **14**, **15** and **16**, were observed (Supplementary Figure S2A). Comparison of the MS/MS spectra for **11**, **12**, **13**, **14**, **15** and **16**, with those for **1**, **2**, **3**, **4**, **5** and **6**, respectively, provided further evidence that the hydroxymethyl groups appended to the epoxyketone warhead in the latter are replaced by methyl groups in the former (Supplementary Figure S3). Introduction of pOSV-*epnI* (an integrative vector containing *epnI* under the control of the constitutive *ermE*\* promoter) into the *epnI* mutant restored the production of eponemycin **1**, eponemycin congeners **3–6** and TMC-86A **2** (Supplementary Figure S2A). These data show that the function of all the eponemycin biosynthetic genes, except *epnI*, are unaffected in the pCAP1000epnBGC- $\Delta epnI$  construct, demonstrating that yeast-mediated plasmid reassembly proceeds with a very high degree of fidelity.

### EpnD and EpnE are required for assembly of the TMC-86A butanoyl group

To investigate the biosynthetic roles of the proteins encoded by *epnD* and *epnE*, which are KASIII and ACP homologs, respectively, the profile of metabolites produced by *S. albus* containing pCAP1000epnBGC, pCAP1000epnBGC- $\Delta epnD$  and pCAP1000epnBGC- $\Delta epnE$  were compared using UHPLC-ESI-Q-TOF-MS. While there was only a modest decrease in the production of eponemycin **1** and congeners **3**, **4**, **5** and **6** in the *epnD* mutant (~70% of wild type), TMC-86A **2** production was strongly affected (Figure 5A and Supplementary Figure S2B). Introduction of pOSV-*epnD* into the  $\Delta epnD$  mutant restored wild type levels of TMC-86A **2**, without affecting the production of eponemycin **1**, or congeners **3**, **4**, **5** and **6** (Figure 5A, Supplementary Figure S2B). In the *epnE* mutant, TMC-86A **2** production was also strongly affected,



**Figure 5.** Extracted ion chromatograms (EICs) from UHPLC-ESI-Q-TOF-MS analyses of culture extracts from *S. albus* containing pCAP1000epnBGC and various engineered derivatives. (A) Comparison of EICs at  $m/z = 421.2309 \pm 0.002$  and  $365.1683 \pm 0.002$ , corresponding to the  $[M+Na]^+$  ions for eponemycin **1** and TMC-86A **2**, respectively, for *S. albus* containing pCAP1000epnBGC;  $\Delta epnD$ ,  $\Delta epnE$  and  $\Delta epnDE$  derivatives;  $\Delta epnD$ ,  $\Delta epnE$  and  $\Delta epnDE$  derivatives plus *epnD*, *epnE* and *epnDE* under the control of the constitutive *ermE*\* promoter; and pCAP1000epnBGC plus *epnDE* under the control of the *ermE*\* promoter. The corresponding EICs for eponemycin congeners **3–6** are in Supplementary Figure S2B. (B) EICs at  $m/z = 421.2309 \pm 0.002$ ,  $365.1683 \pm 0.002$ ,  $407.2516 \pm 0.002$ , and  $367.1840 \pm 0.002$ , corresponding to the  $[M+Na]^+$  ions for eponemycin **1**, TMC-86A **2**, deshydroxy-dihydro-eponemycin congener **7**, and dihydro-TMC-86A congener **10**, respectively, for *S. albus* containing pCAP1000epnBGC with and without added dhL, and  $\Delta epnK$  and  $\Delta epnJ$  derivatives with and without added dhL. The asterisks highlight the absence of peaks for **1** and **2** in the  $\Delta epnK$  and  $\Delta epnJ$  mutants. The corresponding EICs for eponemycin congeners **3–6**, dihydro-eponemycin **8**, and a dihydro-eponemycin congener **9** are in Supplementary Figures S2D and S3.

whereas eponemycin **1** and congeners **3**, **4**, **5** and **6** were produced at wild type levels (Figure 5A, Supplementary Figure S2B). Introduction of pOSV-*epnE* into the mutant restored

TMC-86A **2** production to 50% of the wild-type level, without affecting the production of eponemycin **1**, or congeners **3**, **4**, **5** and **6** (Figure 5, Supplementary Figure S2B). These data indicate that EpnD and EpnE both play an important role in the biosynthesis of TMC-86A **2** but neither is required for the assembly of eponemycin **1** or its congeners.

To further confirm the key role played by EpnD and EpnE in TMC-86A **2** biosynthesis, we used our genetic manipulation platform to create a double mutant by deleting 1242 bp in pCAP1000epnBGC between the start codon of *epnD* and the stop codon of *epnE* (including the 18bp intergenic region; Figure 4 and Supplementary Table S4). The integrity of the resulting plasmid (pCAP1000epnBGC- $\Delta$ *epnDE*) was confirmed by restriction enzyme digestion and sequencing (Supplementary Figure S1), and it was introduced into *S. albus*. UHPLC-ESI-Q-TOF-MS analysis of ethyl acetate extracts demonstrated that TMC-86A **2** production is strongly suppressed in the resulting strain, whereas the levels of eponemycin **1** and congeners **3**, **4**, **5** and **6** are not significantly altered (Figure 5A, Supplementary Figure S2B). Introduction of pOSV-*epnDE* into the strain dramatically increased the production of TMC-86A **2** relative to eponemycin **1** and congeners **3**, **4**, **5** and **6** (Figure 5A, Supplementary Figure S2B). The more marked increase in titers of TMC-86A **2** relative to eponemycin **1** and congeners when *epnD* and *epnE* are co-expressed in the *epnDE* double mutant and epnBGC heterologous expression mutant compared to the complemented *epnD* and *epnE* individual mutants indicates that EpnD and EpnE cooperate closely in the assembly of the butanoyl starter unit for TMC-86A biosynthesis.

We also created a 984 bp in-frame deletion in *tmcD* (which encodes an EpnD orthologue) to investigate the function of this gene in the *S. chromofuscus* TMC-86A BGC. *SphI* digestion of pCAP1000tmcBGC produced eight fragments, which were used with a *tmcD* deletion polynucleotide and four repair polynucleotides to co-transform yeast (Figure 4 and Supplementary Table S4). 1 out of 14 colonies screened contained the correctly reassembled plasmid (pCAP1000tmcBGC- $\Delta$ *tmcD*). The integrity of this plasmid was confirmed by restriction digestion and sequencing (Supplementary Figure S1), and it was transferred into *S. albus* by conjugation from *E. coli*. UHPLC-ESI-Q-TOF-MS analysis of culture extracts showed that TMC-86A **2** production is strongly suppressed in this strain and eponemycin **1** plus congeners **3**, **4**, **5** and **6** are produced (Supplementary Figure S2C). Introduction of either pOSV-*tmcD* or pOSV-*epnD* into this strain greatly increased the levels of TMC-86A **2** relative to eponemycin **1** and its congeners (Supplementary Figure S2C).

Collectively, these data show that the principal difference between the epoxyketone biosynthetic pathways in *S. hygroscopicus* and *S. chromofuscus* is the respective ability of EpnD/EpnE and TmcD/TmcE to supply the butanoyl starter unit to the NRPSs EpnG and TmcG. In the former, this process is relatively inefficient, allowing alternative starter units, presumably assembled by the primary metabolic FAS, to be utilized. This results in the production of eponemycin **1** and congeners **3**, **4**, **5** and **6**, in addition to TMC-86A **2** via the *S. hygroscopicus* pathway.

## EpnJ and EpnK play key roles in 4,5-dehydroleucine biosynthesis

Previous genetic experiments have failed to define unambiguous biosynthetic roles for *epnJ* and *epnK* (43). These genes encode an NRPS containing MbtH-like protein (MLP), adenylation (A) and peptidyl carrier protein (PCP) domains and a putative CYP, respectively. We have proposed that the homologous proteins encoded by *tmcJ* and *tmcK* could be involved in the biosynthesis of the dhL residue incorporated into TMC-86A **2** and eponemycin **1** (40).

UHPLC-ESI-Q-TOF-MS analyses showed eponemycin **1**, congeners **3**, **4**, **5** and **6**, and TMC-86A **2** are all absent from culture extracts of *S. albus* containing pCAP1000epnBGC- $\Delta$ *epnJ* and pCAP1000epnBGC- $\Delta$ *epnK* (Figure 5B, Supplementary Figure S2D). Instead, eponemycin congener **7** ( $[M+H]^+m/z$  385.2704, calculated for  $C_{20}H_{37}N_2O_5^+$ : 385.2697), in which the hydroxymethyl group appended to the warhead and the dhL residue have been replaced by a methyl group and a L-Leu residue, respectively, is the major epoxyketone produced. This compound can also be detected in *S. albus* containing pCAP1000epnBGC. Metabolites with molecular formulae and MS/MS fragmentation patterns corresponding to dihydro-eponemycin **8** (calculated for  $C_{20}H_{37}N_2O_6^+$ : 401.2644, measured: 401.2644), a dihydro-eponemycin congener **9** (calculated for  $C_{21}H_{35}N_2O_6^+$ : 429.2595, measured: 429.2597), and dihydro-TMC-86A **10** (calculated for  $C_{16}H_{29}N_2O_6^+$ : 345.2020, measured: 345.2018) were also observed (Figure 5B, Supplementary Figures S2D and S3). Introduction of pOSV556t-*epnJ* and pOSV556t-*epnK* into the corresponding mutants restored the production of eponemycin **1**, congeners **3**, **4**, **5** and **6**, and TMC-86A **2** (Supplementary Figure S2D).

These data confirm that both EpnJ and EpnK play critical roles in the biosynthesis of the dhL residue incorporated into metabolites **1–6**. Moreover, they are consistent with our previous proposal that the A domain of EpnJ loads L-Leu onto the downstream PCP domain and the resulting aminoacyl thioester is desaturated across C4/C5 by EpnK (40). Hydrolytic release (likely catalysed by the type II thioesterase EpnB) would afford dhL, which is incorporated into **1–6** by EpnG. Feeding of dhL to the *epnJ* and *epnK* mutants restored the production of **1–6**, consistent with this nonproteinogenic amino acid being a free intermediate in their biosynthesis (Figure 5B, Supplementary Figure S2D).

## DISCUSSION

In this work, we have developed a yeast-based platform for efficient capture and parallelized manipulation of specialized metabolite BGCs utilizing DSBs created by commercially available restriction enzymes. By combining beneficial features of the previously reported pCAP01 and pCAP03 vectors (5,12), we have created pCAP1000 containing long flanking sequences to increase the efficiency of homologous recombination, and a counterselectable marker to reduce the number of false positives resulting from non-homologous end joining. The enhanced utility of pCAP1000 compared with pCAP01



and pCAP03 was demonstrated by comparing capture frequencies for the eponemycin and TMC-86A BGCs. No clones were obtained for either BGC using pCAP03. Using pCAP01, 23% and 0% of clones contained the eponemycin and TMC-86A BGCs, respectively. In contrast, 67% and 100% of clones contained the eponemycin and TMC-86A BGCs when pCAP1000 was used. Thus, pCAP1000 represents a useful addition to the toolkit for TAR-mediated BGC capture, particularly for cases that fail using pCAP03 due to low homologous recombination efficiency.

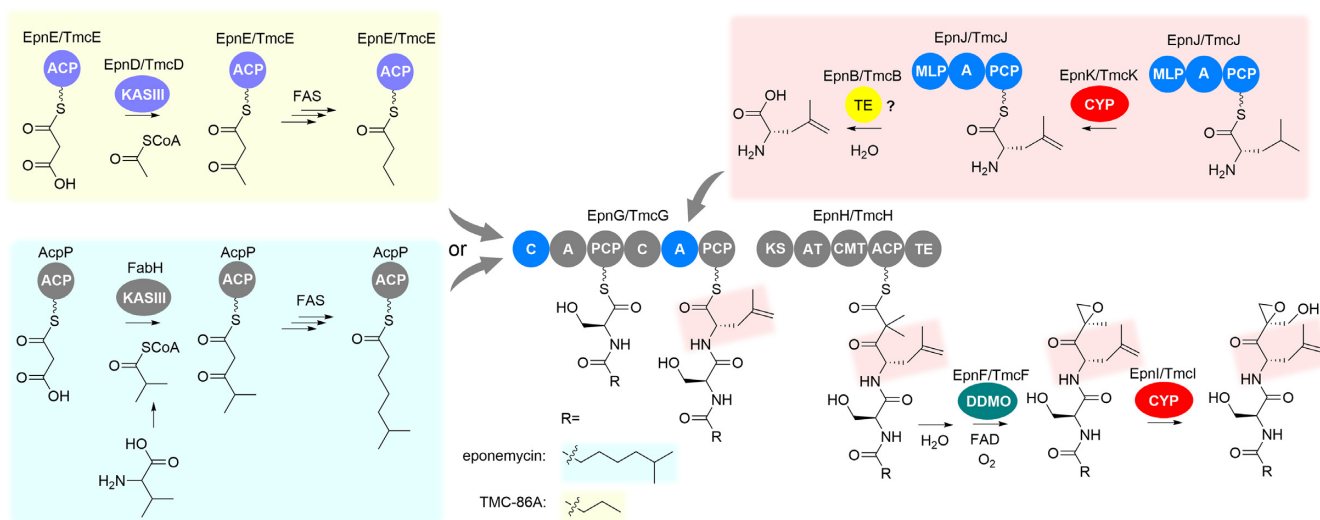
Yeast is an efficient tool for plasmid construction and modification due to its robust homology-directed DNA repair machinery (11,23,49,50). However, current methodology is reliant on the creation of a unique DSB at sites targeted for modification, using either a naturally occurring unique restriction site or a site-specific guide RNA in conjunction with CRISPR-Cas9 (23,50). While such methods are efficient, as illustrated by engineering of the initially captured TMC-86A BGC to incorporate the promoter region upstream of *tmcA*, creation of DSBs at specific locations can be challenging due to the random distribution of unique restriction sites and CRISPR-Cas9 digestion at off-target sites in GC-rich DNA. Our methodology circumvents these problems by exploiting the high efficiency and fidelity of homologous recombination in yeast to reassemble plasmids that are cut at multiple sites by a single restriction endonuclease. Using different combinations of repair and deletion polynucleotides enables combinatorial reassembly of plasmids containing in-frame deletions in various genes. As a proof of concept, we created six in-frame deletions in the eponemycin BGC and a single deletion in the TMC-86A BGC using various restriction enzymes producing as many as eight fragments. In all cases, the fragments were reassembled to produce BGCs containing the desired deletions. Subsequent heterologous expression, genetic and chemical complementation, and UHPLC-ESI-Q-TOF-MS analyses confirmed that all the engineered BGCs were functional and enabled biosynthetic roles to be assigned to the target genes. Thus, our methodology represents a valuable addition to the toolkit for manipulating specialized metabolite BGCs, which can be readily expanded to BGCs cloned into other vectors (such as those isolated from cosmid/fosmid libraries) by incorporating a yeast origin of replication. Moreover, using synthetic (rather than PCR-amplified) repair and deletion polynucleotides would enable this engineering platform to be automated, further facilitating parallelization. One potential limitation of our method is that reassembly efficiency could decrease with larger BGCs, because it may prove necessary to use a greater number of restriction fragments to create in-frame deletions in certain genes. This might necessitate screening of larger numbers of colonies to identify a plasmid with the intended genotype. However, in our hands, deletions can readily be constructed in captured BGCs up to 50 kb in size, and no more than 20 transformants typically needs to be screened to identify a positive clone.

The previously overlooked production of TMC-86A **2** alongside eponemycin **1** and congeners **3–6** when the eponemycin BGC is expressed in *S. albus*, coupled with the exclusive production of **2** when the TMC-86A BGC is in-

troduced into this host, prompted us to investigate the biosynthetic origins of the various N-terminal fatty acyl residues in these metabolites. Our data clearly implicate the KASIII and ACP homologs encoded by *epnD/tmcD* and *epnE/tmcE*, respectively, in the assembly of the butanoyl thioester that initiates TMC-86A **2** biosynthesis (Figure 6). EpnD/TmcD appear to be specific for acetyl-CoA and are proposed to catalyze decarboxylative condensation of an active site-bound acetyl thioester with malonyl-EpnE/TmcE (Figure 6), mirroring the role played by KASIII homologs encoded by BGCs for other specialized metabolites (14). The resulting ACP-bound  $\beta$ -keto thioester is likely reduced to the corresponding butanoyl thioester by the ketoreductase, dehydratase and enoylreductase enzymes of the primary metabolic FAS. Our data also indicate that EpnD/TmcD and EpnE/TmcE play no role in assembly of the 6-methylheptanoyl thioester that initiates eponemycin **1** biosynthesis (Figure 6). This is likely assembled entirely by the primary metabolic FAS. Fatty acid biosynthesis in *Streptomyces* species is initiated primarily by FabH-catalyzed condensation of isobutyl, 2-methylbutyl and 3-methylbutyl-CoA starter units derived from L-Val, L-Ile and L-Leu, respectively, with malonyl-AcpP (51,52). Incorporation of [U-<sup>13</sup>C<sub>5</sub>]-L-valine into the N-terminal fatty acyl residue of eponemycin **1** is consistent with it being assembled by the primary metabolic FAS (43,51,52).

We previously suggested that the dhL residue of TMC-86A could be assembled by the NRPS encoded by *tmcJ* and the CYP encoded by *tmcK* (40), prior to being transferred onto the second PCP domain of TmcG (Figure 6). The data presented here are fully consistent with this proposal—TMC-86A **2** and eponemycin **1** production is abolished in *epnK* and *epnJ* mutants, but analogues containing L-Leu in place of the dhL residue are produced and feeding of dhL to the mutants restores production. Our earlier observation that both dhL and L-Leu can be adenylated by the second A domain of TmcG (40) indicates that dhL is hydrolytically released after EpnK/TmcK-catalyzed desaturation of the L-Leu residue bound to the PCP domain of EpnG/TmcG. The type II thioesterase (TE) encoded by *epnB/tmcB* is likely to catalyze dhL release (Figure 6). However, this hypothesis is difficult to verify genetically, because deletion of genes encoding type II TEs is known to dramatically impact metabolite production levels in several NRPS and PKS systems (53,54). Intriguingly, the synthetic epoxyketone KZR-616, which selectively targets the immunoproteasome and is currently in phase II trials for lupus nephritis (55), contains a L-(1-cyclopentenyl)alanine (CPA) residue in the position corresponding to the dhL residue in eponemycin **1** and TMC-86A **2**. It will therefore be interesting to investigate whether the biosynthetic machinery encoded by the eponemycin and TMC-86A BGCs is able to incorporate CPA instead of dhL into novel epoxyketone derivatives.

In conclusion, we have developed a robust and versatile platform for efficient cloning and parallelized manipulation of BGCs that is expected to find wide-spread application in elucidating mechanisms of specialized metabolite biosynthesis. Application of this platform to the eponemycin and TMC-86A BGCs has illuminated key steps in the assem-



**Figure 6.** Revised biosynthetic pathway for eponemycin **1** and TMC-86A **2**. The newly identified pathways for assembly of the butanoyl thioester starter unit for TMC-86A **2** by EpnD/TmcD and EpnE/TmcE and the non-proteinogenic amino acid dHL, proposed to involve EpnJ/TmcJ-catalyzed desaturation of L-Leu following activation and covalent tethering by EpnJ/TmcJ are highlighted by the yellow and red boxes, respectively. The 6-methylheptanoyl thioester starter unit for eponemycin **1** biosynthesis is hypothesized to be assembled by the primary metabolic fatty acid synthase (FAS) as shown in the blue box.

bly of these metabolites and lays the foundation for future biosynthetic engineering efforts directed toward the production of clinically useful epoxyketones.

## DATA AVAILABILITY

All data needed to interpret, verify, and extend the research are provided in the main manuscript and supplementary data. The raw LC-MS data, which were processed via standard extracted ion chromatogram procedures, are available upon written request to the corresponding author.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## FUNDING

Biotechnology and Biological Science Research Council [BB/K002341/1 to G.L.C., BB/M017982/1 to G.L.C., C.C.]; Australian Research Council [CE200100012 to G.L.C.]; Wolfson Research Merit Award from the Royal Society [WM130033 to G.L.C.]. Funding for open access charge: UKRI institutional allocation.

**Conflict of interest statement.** G.L.C. is a non-executive director, shareholder and paid consultant of Erebagen Ltd; the other authors declare no competing interests.

## REFERENCES

- Rutledge, P.J. and Challis, G.L. (2015) Discovery of microbial natural products by activation of silent biosynthetic gene clusters. *Nat. Rev. Microbiol.*, **13**, 509–523.
- Gomez-Escribano, J., Alt, S. and Bibb, M. (2016) Next generation sequencing of actinobacteria for the discovery of novel natural products. *Mar. Drugs*, **14**, 78.
- Gomez-Escribano, J.P. and Bibb, M.J. (2014) Heterologous expression of natural product biosynthetic gene clusters in *Streptomyces coelicolor*: from genome mining to manipulation of biosynthetic pathways. *J. Ind. Microbiol. Biotechnol.*, **41**, 425–431.
- Baltz, R.H. (2010) *Streptomyces* and *saccharopolyspora* hosts for heterologous expression of secondary metabolite gene clusters. *J. Ind. Microbiol. Biotechnol.*, **37**, 759–772.
- Yamanaka, K., Reynolds, K.A., Kersten, R.D., Ryan, K.S., Gonzalez, D.J., Nizet, V., Dorrestein, P.C. and Moore, B.S. (2014) Direct cloning and refactoring of a silent lipopeptide biosynthetic gene cluster yields the antibiotic taromycin A. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, 1957–1962.
- Fu, J., Bian, X., Hu, S., Wang, H., Huang, F., Seibert, P.M., Plaza, A., Xia, L., Müller, R., Stewart, A.F. et al. (2012) Full-length RecE enhances linear-linear homologous recombination and facilitates direct cloning for bioprospecting. *Nat. Biotechnol.*, **30**, 440–446.
- Jiang, W., Zhao, X., Gabrieli, T., Lou, C., Ebenstein, Y. and Zhu, T.F. (2015) Cas9-assisted targeting of chromosome segments CATCH enables one-step targeted cloning of large gene clusters. *Nat. Commun.*, **6**, 8101.
- Tao, W., Chen, L., Zhao, C., Wu, J., Yan, D., Deng, Z. and Sun, Y. (2019) *In vitro* packaging mediated one-step targeted cloning of natural product pathway. *ACS Synth. Biol.*, **8**, 1991–1997.
- Shao, Z., Zhao, H. and Zhao, H. (2009) DNA assembler, an *in vivo* genetic method for rapid construction of biochemical pathways. *Nucleic Acids Res.*, **37**, e16.
- Greunke, C., Duell, E.R., D'Agostino, P.M., Glöckle, A., Lamm, K. and Gulder, T.A.M. (2018) Direct pathway cloning (DiPaC) to unlock natural product biosynthetic potential. *Metab. Eng.*, **47**, 334–345.
- Kouprina, N. and Larionov, V. (2008) Selective isolation of genomic loci from complex genomes by transformation-associated recombination cloning in the yeast *Saccharomyces cerevisiae*. *Nat. Protoc.*, **3**, 371–377.
- Tang, X., Li, J., Millán-Aguinaga, N., Zhang, J.J., O'Neill, E.C., Ugalde, J.A., Jensen, P.R., Mantovani, S.M. and Moore, B.S. (2015) Identification of thiotetronic acid antibiotic biosynthetic pathways by target-directed genome mining. *ACS Chem. Biol.*, **10**, 2841–2849.
- Stanley, A.E., Walton, L.J., Kourdi Zerikly, M., Corre, C. and Challis, G.L. (2006) Elucidation of the *Streptomyces coelicolor* pathway to 4-methoxy-2,2'-bipyrrrole-5-carboxaldehyde, an intermediate in prodiginine biosynthesis. *Chem. Commun. (Camb)*, **38**, 3981–3983.
- Mo, S.J., Sydor, P.K., Corre, C., Alhamadsheh, M.M., Stanley, A.E., Haynes, S.W.W., Song, L., Reynolds, K.A. and Challis, G.L. (2008)

- Elucidation of the *Streptomyces coelicolor* pathway to 2-undecylpyrrole, a key intermediate in undecylprodiginine and streptorubin B biosynthesis. *Chem. Biol.*, **15**, 137–148.
15. Song, L., Laureti, L., Corre, C., Leblond, P., Aigle, B. and Challis, G.L. (2013) Cytochrome P450-mediated hydroxylation is required for polyketide macrolactonization in stambomycin biosynthesis. *J. Antibiot. (Tokyo)*, **67**, 71–76.
  16. Sydor, P.K., Barry, S.M., Odulate, O.M., Barona-Gomez, F., Haynes, S.W., Corre, C., Song, L. and Challis, G.L. (2011) Regio- and stereodivergent antibiotic oxidative carbocyclizations catalysed by Rieske oxygenase-like enzymes. *Nat. Chem.*, **3**, 388–392.
  17. Huang, C., Huang, F., Moison, E., Guo, J., Jian, X., Duan, X., Deng, Z., Leadlay, P.F. and Sun, Y. (2015) Delineating the biosynthesis of gentamicin X2, the common precursor of the gentamicin C antibiotic complex. *Chem. Biol.*, **22**, 251–261.
  18. Inahashi, Y., Zhou, S., Bibb, M.J., Song, L., Al-Bassam, M.M., Bibb, M.J. and Challis, G.L. (2017) Watasemycin biosynthesis in *Streptomyces venezuelae*: thiazoline C-methylation by a type B radical-SAM methylase homologue. *Chem. Sci.*, **8**, 2823–2831.
  19. Kieser, T., Bibb, M.J., Buttner, M.J., Chater, K.F. and Hopwood, D.A. (2000) In: *Practical Streptomyces Genetics*. John Innes Foundation. John Innes Centre, Norwich.
  20. Gust, B., Challis, G.L., Fowler, K., Kieser, T. and Chater, K.F. (2003) PCR-targeted *Streptomyces* gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin. *Proc. Natl. Acad. Sci. U.S.A.*, **100**, 1541–1546.
  21. Montiel, D., Kang, H.-S., Chang, F.-Y., Charlop-Powers, Z. and Brady, S.F. (2015) Yeast homologous recombination-based promoter engineering for the activation of silent natural product biosynthetic gene clusters. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, 8953–8958.
  22. Oldenburg, K., Vo, K.T., Michaelis, S. and Paddon, C. (1997) Recombination-mediated PCR-directed plasmid construction in vivo in yeast. *Nucleic Acids Res.*, **25**, 451–452.
  23. Eyles, T.H., Vior, N.M. and Truman, A.W. (2018) Rapid and robust yeast-mediated pathway refactoring generates multiple new bottromycin-related metabolites. *ACS Synth. Biol.*, **7**, 1211–1218.
  24. Kang, H.S., Charlop-Powers, Z. and Brady, S.F. (2016) Multiplexed CRISPR/Cas9- and TAR-mediated promoter engineering of natural product biosynthetic gene clusters in yeast. *ACS Synth. Biol.*, **5**, 1002–1010.
  25. Kim, S.H., Lu, W., Ahmadi, M.K., Montiel, D., Ternei, M.A. and Brady, S.F. (2019) Atolypenes, tricyclic bacterial sesterterpenes discovered using a multiplexed *in vitro* Cas9-TAR gene cluster refactoring approach. *ACS Synth. Biol.*, **8**, 109–118.
  26. Ruiz, E., Talenton, V., Dubrana, M.P., Guesdon, G., Lluch-Senar, M., Salin, F., Sirand-Pugnet, P., Arfi, Y. and Lartigue, C. (2019) CReasPy-cloning: a method for simultaneous cloning and engineering of megabase-sized genomes in yeast using the CRISPR-Cas9 system. *ACS Synth. Biol.*, **8**, 2547–2557.
  27. Zhang, J.J. and Moore, B.S. (2020) Site-directed mutagenesis of large biosynthetic gene clusters via oligonucleotide recombineering and CRISPR/Cas9 targeting. *ACS Synth. Biol.*, **9**, 1917–1922.
  28. Xu, W., Klumbys, E., Ang, E.L. and Zhao, H. (2020) Emerging molecular biology tools and strategies for engineering natural product biosynthesis. *Metab. Eng. Commun.*, **10**, e00108.
  29. Zhang, J.J., Tang, X. and Moore, B.S. (2019) Genetic platforms for heterologous expression of microbial natural products. *Nat. Prod. Rep.*, **36**, 1313–1332.
  30. Kim, K.B. and Crews, C.M. (2013) From epoxomicin to carfilzomib: chemistry, biology, and medical outcomes. *Nat. Prod. Rep.*, **30**, 600–604.
  31. Johnson, H.W.B., Lowe, E., Anderl, J.L., Fan, A., Muchamuel, T., Bowers, S., Moebius, D.C., Kirk, C. and McMinn, D.L. (2018) Required immunoproteasome subunit inhibition profile for anti-inflammatory efficacy and clinical candidate KZR-616 ((2S, 3R)-N-((S)-3-(cyclopent-1-en-1-yl)-1-((R)-2-methylxiran-2-yl)-1-oxopropan-2-yl)-3-hydroxy-3-(4-methoxyphenyl)-2-((S)-2-(2-morpholinoacetamido)propanamido)propanamide). *J. Med. Chem.*, **61**, 11127–11143.
  32. Koguchi, Y., Kohno, J., Suzuki, S., Nishio, M., Takahashi, K., Ohnuki, T. and Komatsubara, S. (1999) TMC-86A, B and TMC-96, new proteasome inhibitors from *streptomyces* sp. TC 1084 and *Saccharothrix* sp. TC 1094. I. Taxonomy, fermentation, isolation, and biological activities. *J. Antibiot. (Tokyo)*, **52**, 1069–1076.
  33. Sugawara, K., Hatori, M., Nishiyama, Y., Tomita, K., Kamei, H., Konishi, M. and Oki, T. (1990) Eponemycin, a new antibiotic active against B16 melanoma. I. Production, isolation, structure and biological activity. *J. Antibiot. (Tokyo)*, **43**, 8–18.
  34. Hanada, M., Sugawara, K., Kaneta, K., Toda, S., Nishiyama, Y., Tomita, K., Yamamoto, H., Konishi, M. and Oki, T. (1992) Epoxomicin, a new antitumor agent of microbial origin. *J. Antibiot. (Tokyo)*, **45**, 1746–1752.
  35. Pereira, A.R., Kale, A.J., Fenley, A.T., Byrum, T., Debonsi, H.M., Gilson, M.K., Valeriote, F.A., Moore, B.S. and Gerwick, W.H. (2012) The carmaphycins: new proteasome inhibitors exhibiting an  $\alpha,\beta$ -epoxyketone warhead from a marine cyanobacterium. *Chem. Bio. Chem.*, **13**, 810–817.
  36. Keller, L., Plaza, A., Dubiella, C., Groll, M., Kaiser, M. and Müller, R. (2015) Macyranoles: structure, biosynthesis, and binding mode of an unprecedented epoxyketone that targets the 20S proteasome. *J. Am Chem. Soc.*, **137**, 8121–8130.
  37. Owen, J.G., Charlop-Powers, Z., Smith, A.G., Ternei, M.A., Calle, P.Y., Reddy, B.V.B., Montiel, D. and Brady, S.F. (2015) Multiplexed metagenome mining using short DNA sequence tags facilitates targeted discovery of epoxyketone proteasome inhibitors. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, 4221–4226.
  38. Groll, M., Kim, K.B., Kairies, N., Huber, R. and Crews, C.M. (2000) Crystal structure of epoxomicin:20S proteasome reveals a molecular basis for selectivity of  $\alpha',\beta'$ -epoxyketone proteasome inhibitors [12]. *J. Am Chem. Soc.*, **122**, 1237–1238.
  39. Schrader, J., Henneberg, F., Mata, R.A., Tittmann, K., Schneider, T.R., Stark, H., Bourenkov, G. and Chari, A. (2016) The inhibition mechanism of human 20 S proteasomes enables next-generation inhibitor design. *Science*, **353**, 594–598.
  40. Zabala, D., Cartwright, J.W., Roberts, D.M., Law, B.J.C., Song, L., Samborsky, M., Leadlay, P.F., Micklefield, J. and Challis, G.L. (2016) A flavin-dependent decarboxylase-dehydrogenase-monooxygenase assembles the warhead of  $\alpha,\beta$ -epoxyketone proteasome inhibitors. *J. Am Chem. Soc.*, **138**, 4342–4345.
  41. Schorn, M., Zettler, J., Noel, J.P., Dorrestein, P.C., Moore, B.S. and Kayser, L. (2014) Genetic basis for the biosynthesis of the pharmaceutically important class of epoxyketone proteasome inhibitors. *ACS Chem. Biol.*, **9**, 301–309.
  42. Liu, J., Zhu, X. and Zhang, W. (2015) Identifying the minimal enzymes required for biosynthesis of epoxyketone oroteasome inhibitors. *Chem. Bio. Chem.*, **16**, 2585–2589.
  43. Zettler, J., Zubeil, F., Kulik, A., Grond, S. and Kayser, L. (2016) Epoxomicin and eponemycin biosynthesis involves gem-dimethylation and an acyl-CoA dehydrogenase-like enzyme. *Chem. bio. chem.*, **17**, 792–798.
  44. Green, Michael.R. and Sambrook, Joseph. (2012) In: *Molecular Cloning: A Laboratory Manual*. 4th edn, Cold Spring Harbor Laboratory Press, NY.
  45. Flett, F., Mersinias, V. and Smith, C.P. (1997) High efficiency intergeneric conjugal transfer of plasmid DNA from *Escherichia coli* to methyl DNA-restricting streptomycetes. *FEMS Microbiol. Lett.*, **155**, 223–229.
  46. Manivasakam, P., Weber, S.C., McElver, J. and Schiestl, R.H. (1995) Micro-homology mediated PCR targeting in *Saccharomyces cerevisiae*. *Nucleic Acids Res.*, **23**, 2799–2800.
  47. Noskov, V.N., Kouprina, N., Leem, S.-H., Ouspenski, I., Barrett, J.C. and Larionov, V. (2003) A general cloning system to selectively isolate any eukaryotic or prokaryotic genomic region in yeast. *BMC Genomics*, **4**, 16.
  48. MacNeil, D.J., Gewain, K.M., Ruby, C.L., Dezeny, G., Gibbons, P.H. and MacNeil, T. (1992) Analysis of *Streptomyces avermitilis* genes required for avermectin biosynthesis utilizing a novel integration vector. *Gene*, **111**, 61–68.
  49. Joska, T.M., Mashruwala, A., Boyd, J.M. and Belden, W.J. (2014) A universal cloning method based on yeast homologous recombination that is simple, efficient, and versatile. *J. Microbiol. Methods*, **100**, 46–51.
  50. Stovicek, V., Holkenbrink, C. and Borodina, I. (2017) CRISPR/cas system for yeast genome engineering: advances and applications. *FEMS Yeast Res.*, **17**, fox030.
  51. Choi, K.H., Heath, R.J. and Rock, C.O. (2000) beta-ketoacyl-acyl carrier protein synthase III (FabH) is a determining factor in branched-chain fatty acid biosynthesis. *J. Bacteriol.*, **182**, 365–370.



52. Mo,S., Kim,B.S. and Reynolds,K.A. (2005) Production of branched-chain alkylprodiginines in *S. Coelicolor* by replacement of the 3-ketoacyl ACP synthase III initiation enzyme, *RedP*. *Chem. Biol.*, **12**, 191–200.
53. Zhou,Y., Meng,Q., You,D., Li,J., Chen,S., Ding,D., Zhou,X., Zhou,H., Bai,L. and Deng,Z. (2008) Selective removal of aberrant extender units by a type II thioesterase for efficient FR-008/candicidin biosynthesis in *Streptomyces* sp. Strain FR-008. *Appl. Environ. Microbiol.*, **74**, 7235–7242.
54. Schneider,A. and Marahiel,M.A. (1998) Genetic evidence for a role of thioesterase domains, integrated in or associated with peptide synthetases, in non-ribosomal peptide biosynthesis in *Bacillus subtilis*. *Arch. Microbiol.*, **169**, 404–410.
55. Kirk,C.J., Muchamuel,T., Wang,J. and Fan,R.A. (2022) Discovery and early clinical development of selective immunoproteasome inhibitors. *Cells*, **11**, 9.