Molecular BioSystems

Cite this: *Mol. BioSyst.*, 2011, **7**, 1549–1563

www.rsc.org/molecularbiosystems

PAPER

Insights into the complex biosynthesis of the leupyrrins in *Sorangium cellulosum* So ce690†

Maren Kopp,[ab] Herbert Irschik,[b] Katja Gemperlein,[a] Kathrin Buntin,[a] Peter Meiser,[ac] Kira J. Weissman,‡[a] Helge B. Bode§[a] and Rolf Müller*[ab]

Received 20th October 2010, Accepted 6th January 2011
DOI: 10.1039/c0mb00240b

The anti-fungal leupyrrins are secondary metabolites produced by several strains of the myxobacterium *Sorangium cellulosum*. These intriguing compounds incorporate an atypically substituted γ-butyrolactone ring, as well as pyrrole and oxazolinone functionalities, which are located within an unusual asymmetrical macrodiolide. Previous feeding studies revealed that this novel structure arises from the homologation of four distinct structural units, nonribosomally-derived peptide, polyketide, isoprenoid and a dicarboxylic acid, coupled with modification of the various building blocks. Here we have attempted to reconcile the biosynthetic pathway proposed on the basis of the feeding studies with the underlying enzymatic machinery in the *S. cellulosum* strain So ce690. Gene products can be assigned to many of the suggested steps, but inspection of the gene set provokes the reconsideration of several key transformations. We support our analysis by the reconstitution *in vitro* of the biosynthesis of the pyrrole carboxylic starter unit along with gene inactivation. In addition, this study reveals that a significant proportion of the genes for leupyrrin biosynthesis are located outside the core cluster, a 'split' organization which is increasingly characteristic of the myxobacteria. Finally, we report the generation of four novel deshydroxy leupyrrin analogues by genetic engineering of the pathway.

Introduction

The myxobacteria, a diverse group of soil-dwelling Gram-negative bacteria, are increasingly recognized as producers of secondary metabolites with potential in clinical therapy.[1,2] Many of these natural products are polyketides (PK), non-ribosomal polypeptides (NRP), or more commonly, hybrids of the two structural types.[3] These classes of metabolites are biosynthesized by gigantic enzymatic 'assembly lines' called polyketide synthases (PKSs) and nonribosomal polypeptide synthetases (NRPSs), respectively.[4–6] Both PKS and NRPS are organized into functional units denoted as modules, in which each module is responsible for incorporating one building block (a carboxylic acid or amino acid, respectively) into the growing chain. PKS modules minimally comprise the keto-synthase (KS), acyl transferase (AT) and acyl carrier protein (ACP) domains required for building block selection and carbon–carbon bond formation, but may additionally include a set of activities (ketoreductase (KR), dehydratase (DH) and enoyl reductase (ER)), which together accomplish the reductive modification of the intermediates; less commonly, a *C*- or an *O*-methyl transferase (MT) may also be present.[7] Within NRPSs, the analogous set of core enzymes includes condensation (C) (or dual condensation/heterocyclization (HC)), adenylation (A) and peptidyl carrier protein (PCP) domains, which may be augmented by processing enzymes such as methyl transferases and epimerization (E) activities. In both cases, chain extension is typically terminated by a thioesterase domain, which effects release of the mature intermediate using a nucleophilic group on the chain, or alternatively by hydrolysis.[8]

Despite the presence of these biosynthetic machineries in multiple bacterial orders, myxobacterial compounds often exhibit rare structural elements relative to other metabolites.[3,9] Chemical functionalities of note include the β-methoxyacrylates of myxothiazol[10] and melithiazol,[11,12] the cyclopropane of ambruticin,[13,14] and the isochromanone ring of the adjuazols.[15] Arguably, however, the structurally most remarkable family of myxobacterial natural products is the leupyrrins.[16] Extensive

[a] *Helmholtz Institute for Pharmaceutical Research, Helmholtz Center for Infection Research and Department of Pharmaceutical Biotechnology, Saarland University, P.O. Box 151150, 66041 Saarbrücken, Germany*
[b] *Microbial Drugs, Helmholtz Center for Infection Research, Inhoffenstrasse 7, 38124 Braunschweig, Germany*
[c] *URSAPHARM Arzneimittel GmbH & Co. KG, Industriestraße 35, 66129 Saarbrücken, Germany*
† Electronic supplementary information (ESI) available. See DOI: 10.1039/c0mb00240b
‡ Present address: AREMS, UMR CNRS-UHP 7214, Université Henri Poincaré, Nancy 1, Faculté des Sciences et Technologies, Bld. des Aiguillettes, BP 70239, 54506 Vandoeuvre-lès-Nancy, France.
§ Present address: Molekulare Biotechnologie, Institut für Molekulare Biowissenschaften, Biozentrum/Campus Riedberg, Goethe Universität Frankfurt, Max-von-Laue-Str. 9, 60438 Frankfurt am Main, Germany.

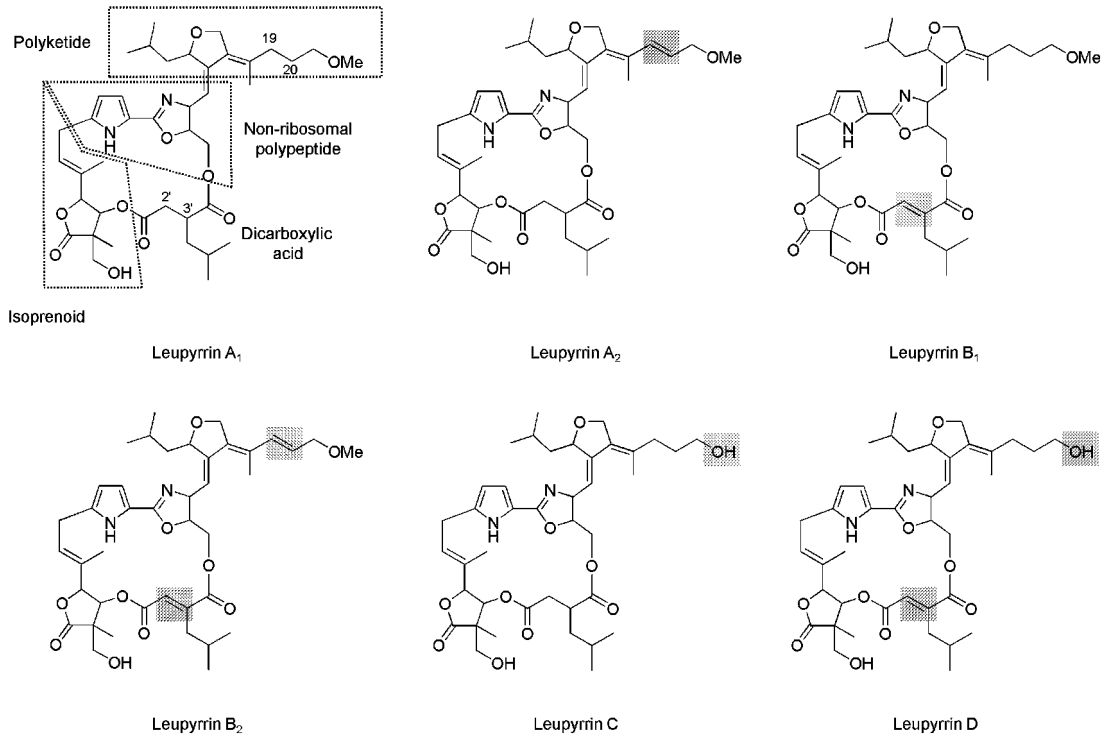

Fig. 1 Structures of the leupyrrins identified from *Sorangium cellulosum* strains So ce690 and So ce705. The shaded regions show the structural differences with leupyrrin A_1.

feeding studies have revealed that these metabolites are comprised of four distinct biosynthetic units: in addition to PK and NRP building blocks, the molecules include an isoprenoid-derived element and a dicarboxylic acid moiety.[17] Six of these unusual, 'super-hybrid' molecules have been isolated to date from *Sorangium cellulosum* strains So ce690 and So ce705, which differ in the presence or absence of C19/C20 and C2′/3′ double bonds, and methylation at the C21 hydroxyl position (Fig. 1).[16]

On the basis of the feeding studies, we previously developed a detailed proposal for the pathway to the leupyrrins, which incorporated several novel transformations.[17] To provide further experimental support for this biosynthetic scheme, as well as to pave the way for rational modification of the structures, we aimed to locate the gene cluster responsible for leupyrrin biosynthesis in *S. cellulosum* So ce690. Here, we show that examination of the identified gene set supports many of the earlier propositions, but also prompts the revision of several central hypotheses. In addition, the analysis strongly suggests that many genes associated with leupyrrin biosynthesis lie outside of the core cluster, a so-called 'split' organization which currently hinders full elucidation of the biosynthetic pathway. We also report the reconstitution *in vitro* of the biosynthesis of the leupyrrin starter unit, pyrrole carboxylic acid, as well as the generation by genetic engineering of four C22-deshydroxyl leupyrrin analogues.

Results and discussion

Identification of the leupyrrin gene cluster and determination of putative cluster boundaries

The NRP-derived portion of the leupyrrins incorporates an oxazoline ring, which we previously hypothesized would derive from threonine by cyclization.[7,18] Therefore, to locate the gene cluster, we aimed to amplify the coding sequence for the HC domain responsible for this transformation from the genome of So ce690. Degenerate HC domain primers[19] were used in a PCR reaction with So ce690 genomic DNA. The product was cloned into pCR2.1TOPO and sequenced, revealing homology to HC domains. The fragment was subsequently reamplified, and cloned into the conjugation vector pSUPHyg.[20] The resulting vector was then used to selectively inactivate the HC-encoding gene by insertional mutagenesis, generating 200 strains uniformly deficient in leupyrrin production as judged by bioassay with the yeast *Debaryomyces hansenii*[21] and HPLC-MS of two culture extracts. The identity of the two mutants was confirmed by Southern Blot (see Materials and methods). A probe based on this sequence was then used for screening a 2300-clone cosmid library of *S. cellulosum* So ce690. Based on the typical genome size of the myxobacteria (9–13 Mbp),[22,23] this cosmid library should represent at least a 10-fold coverage of the genome. This process yielded 16 putatively NRPS-positive hits, which were then restriction digested and re-screened with the HC probe alongside digested genomic DNA from wild type So ce690. Among the two cosmids which yielded the same sized band as the wild type DNA, a single cosmid (C9) was selected for sequencing. The T3 end of this cosmid showed homology to NRPS genes, while the genes on the T5 end could not be reasonably correlated to leupyrrin biosynthesis. 'Cosmid walking' was then used to obtain an additional 30 999 bp of sequence on cosmid C9.6. The nucleotide sequence of the leupyrrin gene cluster has been deposited in GenBank under accession number HM639990.

Table 1 Proteins encoded in the sequenced region of the *Sorangium cellulosum* So ce690 genome and their putative functions

PKS/NRPS portion of the gene cluster

Protein	aa	Proposed function (protein domains with their position in the sequence)[a]
Leu9	599	C (31–477), PCP (500–572)
LeuA	3001	HC (74–519), A (521–1043), PCP (1049–1118), KS (1146–1571), AT (1719–2033), DH (2036–2336), KR (2366–2855), ACP (2878–2958)
LeuB	2260	KS (33–460), AT (612–926), DH (927–1211), *C*-MT (1401–1610), KR (1611–2117), ACP (2133–2213)
LeuC	1866	KS (34–465), AT (607–923), DH (925–1208), KR (1238–1725), ACP (1739–1819)
LeuD	2230	KS (31–457), AT (600–919), DH (921–1199), *C*-MT (1391–1600), KR (1601–2094), ACP (2122–2202)
LeuE	1417	C (77–510), A (511–1036), PCP (1053–1122), TE (1136–1367)

Non-PKS/NRPS portion of the gene cluster

Protein	aa	Proposed function of the homologous protein[b]	Source of the homologous protein	Identity/similarity	GenBank accession no.
Orf1	702	Hypothetical protein Sce_8270	*Sorangium cellulosum*	89%/93%	CAN98440
Orf2	278	Phospholipid/glycerol acyl transferase	*Desulfatibacillum alkenivorans*	42%/62%	ACL04784
Orf3	173	Hypothetical protein PPSIR1_03833	*Plesiocystis pacifica*	35%/52%	EDM79237
Orf4	473	Protein kinase-like protein	*Saccharophagus degradans*	34%/54%	ABD83141
Leu1	259	GrsT (type II thioesterase)	*Brevibacillus brevis*	40%/57%	AAA58717
Leu2	766	Formate dehydrogenase	*Plesiocystis pacifica*	40%/57%	EDM77908
Leu3	434	Predicted protein	*Physcomitrella patens*	25%/45%	EDQ80279
		Hydroxycinnamoyl-CoA shikimate/quinate hydroxycinnamoyltransferase-like	*Populus trichocarpa*	25%/41%	EEF08790
Leu4	478	WS/DGAT/MGAT acyltransferase	*Conexibacter woesei*	41%/59%	ADB50260
Leu5	522	Proline adenylation protein	*Oscillatoria* sp.	50%/65%	ACR33075
Leu6	379	Prolyl-ACP dehydrogenase	*Oscillatoria* sp.	54%/74%	ACR33074
Leu7	92	Acyl carrier protein	*Geobacter bemidjiensis*	36%/61%	ACH37576
Leu8	253	Dehydrogenase	*Rhodococcus erythropolis*	42%/61%	EEN85838
Leu10	399	Probable carboxylase/reductase TgaD	*Sorangium cellulosum*	48%/67%	GQ981380
Leu11	376	Phospholipase, patatin family	*Stigmatella aurantiaca*	42%/63%	EAU66327
Leu12	400	FAD-binding domain protein	*Myxococcus xanthus*	36%/50%	ABF89221
		Geranylgeranyl reductase	*Frankia symbiont*	26%/35%	EFD29699
Leu13	474	Cytochrome P450 family protein	*Plesiocystis pacifica*	39%/57%	EDM74971
Leu14	304	Methylase type II	*Cellulomonas flavigena*	45%/60%	ADG73421
		O-Methyltransferase (SpiB)	*Polyangium cellulosum*	46%/61%	CAL58679
Leu15	337	Rieske (2Fe–2S) iron–sulfur domain protein	*Haliangium ochraceum*	35%/53%	ACY14415
Leu16	246	TetR family transcriptional regulator	*Saccharophagus degradans*	25%/50%	ABD83140
Leu17	601	Predicted very-long-chain acyl-CoA synthetase	*Uncultured bacterium MedeBAC49C08*	39%/60%	AAY82643
Leu18	482	Hypothetical protein PPSIR1-15490	*Plesiocystis pacifica*	43%/56%	EDM73770
		Carotenoid oxygenase	*Haliangium ochraceum*	42%/55%	ACY17942
Leu19	373	HpcH/HpaI aldolase	*Kangiella koreensis*	70%/83%	ACV26976
Leu20	298	HpcH/HpaI aldolase	*Salinispora tropica*	49%/67%	ABP54183
Leu21	175	Dehydratase	*Micromonospora sp.*	68%/78%	EEP74957
Leu22	482	Cytochrome P450 hydroxylase (AjuJ)	*Chondromyces crocatus*	40%/60%	CAQ1883
Orf5	715	Hypothetical protein Sce_8271	*Sorangium cellulosum*	92%/96%	CAN98441
Orf6	410	Hypothetical protein Sce_8272	*Sorangium cellulosum*	89%/92%	CAN98442
Orf7	308	Putative methanol dehydrogenase regulatory protein Sce_8273	*Sorangium cellulosum*	98%/99%	CAN98443
Orf8	309	Hypothetical protein Sce_8274	*Sorangium cellulosum*	93%/97%	CAN98444

[a] Domain boundaries were assigned, when possible, based on solved structures of homologous domains. [b] Typically the closest homolog is shown, except in cases where functional assignment of a more distant homolog was particularly informative.

To define the boundaries of the gene cluster, we identified putative orfs using Frameplot,[24] and assigned gene boundaries and functions on the basis of BLAST analysis[25] (Table 1 and Fig. 2). In the case of the PKS and NRPS genes, determination of the start codons was aided by sequence alignment of the translated N-terminal docking domain regions (Fig. S1, ESI†); these short sequences, located at the extreme termini of PKS and NRPS proteins, have been shown to mediate specific recognition between the multienzymes.[26,27] This analysis revealed that the protein products of genes *orf1* and orfs 5–8 bear high homology (>89% identity) to proteins encoded within the recently-sequenced genome of *S. cellulosum* So ce56.[23] Moreover, the homologous genes in So ce56 (*sce8270–8274*, respectively) are located sequentially within the genome, suggesting that a leupyrrin biosynthetic 'island' may have been inserted between genes *orf1* and *orf5* in So ce690. A similar observation was recently made for the sorangicin producer, So ce12.[28] Based on this analysis, the leupyrrin gene cluster spans approximately 65 kbp, and exhibits an average G + C content of 71%.

Model for the biosynthesis of the leupyrrins

With the gene cluster in hand, we aimed to evaluate our previous proposal for leupyrrin biosynthesis[17] in light of the encoded enzymes. The earlier feeding studies showed that the pyrrole moiety within the NRP portion of the structure, the presumed starter unit for the biosynthesis, derives from

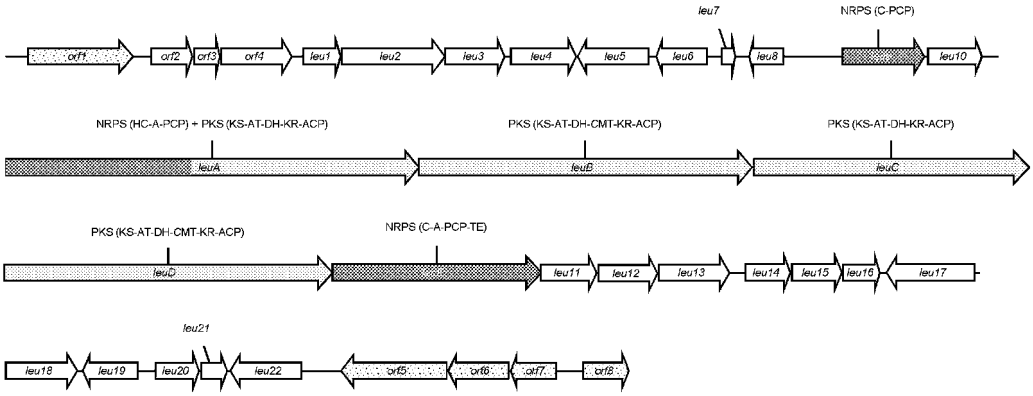

Fig. 2 Genetic organization of the leupyrrin gene cluster in *S. cellulosum* So ce690. The genes are shown approximately to scale. Key: dark grey, NRPS; light grey, PKS; dotted, genes located outside the leupyrrin gene cluster.

proline (Fig. 3). Transformation of L-proline to pyrrole has been demonstrated in a number of other natural product pathways (*e.g.* pyoluteorin,[29] undecylprodigiosin,[29] clorobiocin,[30] coumermycin A_1[30] and anatoxin-a)[31] to be accomplished by a three-protein system, comprising an L-proline-specific adenylation domain, a carrier protein (CP) domain, and a flavoprotein dehydrogenase. (Although the carrier proteins involved in these pathways have been referred to as both acyl carrier proteins (ACPs) and peptidyl carrier proteins (PCPs),[32,33] multiple sequence alignment (Fig. S2, ESI†) shows that they belong to a separate, sequence-diverse class which is not clearly ACP or PCP, hence we prefer the designation CP).[34] As shown by assays *in vitro*, L-proline **7** is activated as its L-prolyl-AMP by the A domain, loaded onto the CP to yield the tethered L-prolyl-*S*-CP protein **8**, and then oxidized to the pyrrole **9** by two consecutive C–N desaturation steps carried out by the dehydrogenase.[33] Consistent with this precedent, inspection of the *leu* gene cluster readily revealed three candidate genes for this series of reactions: *leu5* (A domain), *leu6* (dehydrogenase) and *leu7* (CP). Indeed, the closest homologs to the encoded Leu5 and Leu6 (Table 1) are, respectively, the proline-specific A domain and the dehydrogenase from the recently-characterized pathway to the anatoxin-a pyrrole moiety.[31] Furthermore, identification of residues in Leu5 (DLLYLALT) correlated with A domain substrate specificity (ref. 35 and 36 and http://www-ab.informatik.uni-tuebingen.de/toolbox) reveals near identity with the proline-conferring 'code' (DLLYLALV) distilled from several other discrete proline-specific A domains.[37] Inactivation of *leu5* by insertional mutagenesis resulted in the abolition of leupyrrin production (Fig. S3, ESI†), consistent with an essential function of Leu5 in the pathway.

In order to confirm the role of Leu5–Leu7 in assembly of the pyrrole, we aimed to reconstitute formation of pyrrolyl-*S*-Leu7 *in vitro*. For this, each of Leu5, Leu6 and Leu7 were obtained as a homogeneous protein (Fig. 4) following expression in *E. coli* and purification by affinity chromatography. The identities of Leu5 and Leu6 were confirmed by MALDI-ToF MS/MS, while that of Leu7 was demonstrated by ESI-FT-Orbitrap-MS/MS (*vida infra*). Purified Leu6 showed the characteristic yellow color of a flavoprotein, and the presence of bound FAD was confirmed by release of the co-factor from the protein by denaturation, followed by HPLC-MS analysis. In order to be functional, the carrier protein Leu7 must be modified post-translationally by the addition of a phosphopantetheine prosthetic group, catalyzed by a phosphopantetheinyl transferase (PPTase).[38] As it is often the case that the native PPTases present in *E. coli* are unable to efficiently modify heterologous CPs,[38,39] we expressed Leu7 in the presence of the known broad-specificity PPTase MtaA from *Stigmatella aurantiaca* DW4/3-1.[40] High-resolution nanospray ESI-FT-Orbitrap-MS analysis indicated that Leu7 was expressed in its *apo* (unmodified) form in *E. coli* in the absence of MtaA, but that co-expression with MtaA resulted in essentially complete transformation to its active, *holo* form (+340 Da) (Table 2). We further confirmed the presence of the phosphopantetheine arm using the recently-developed 'phosphopantetheine ejection' assay[41,42] (Table 2).

We next characterized the relative adenylation specificity of Leu5 towards a panel of amino acids comprising L-proline, D-proline, L- and D-pipecolic acid, L-glutamate and glycine, using an end-point ATP-[^{32}P]PP$_i$ exchange assay.[43] In the case of L-proline, we confirmed formation of intermediate L-prolyl-AMP by mass spectrometry (Table 2), and demonstrated that the exchange rate was linear over at least 8 min (data not shown). As shown in Fig. 4, only L-proline was activated to a significant extent following incubation for 1.5 min. Full kinetic characterization of Leu5 with L-proline yielded the following parameters: $k_{cat} = 2.9 \pm 0.1\ s^{-1}$; $K_M = 17 \pm 9\ \mu M$; $k_{cat}/K_M = 1.7 \times 10^5\ M^{-1}\ s^{-1}$. A comparison of the catalytic efficiency of Leu5 to other stand-alone L-proline activating A domains (*e.g.* CloN4 ($4.2 \times 10^2\ M^{-1}\ s^{-1}$)[30] from clorobiocin biosynthesis, CouN4 ($35\ M^{-1}\ s^{-1}$)[30] from the coumermycin A_1 pathway, RedM ($1.8 \times 10^3\ M^{-1}\ s^{-1}$)[29] from undecylprodigiosin biosynthesis and PltF ($1.1 \times 10^4\ M^{-1}\ s^{-1}$)[29] from the pyoluteorin pathway) shows that Leu5 exhibits a significantly higher catalytic efficiency than the other A domains towards this substrate. This effect is largely due to the low K_M value measured for Leu5, as the calculated k_{cat} is within the same range as that for RedM ($2.8\ s^{-1}$) and PltF ($5.5\ s^{-1}$).[29] Using high-resolution analysis by ESI-FT-Orbitrap-MS coupled with the Ppant ejection assay, we next showed that Leu5 readily converted *holo* Leu7 into prolyl-*S*-Leu7, in the presence of L-proline and ATP. Modification of 20 μM Leu7 was *ca.* 80% complete within 90 min in the presence of 0.2 μM Leu5. Finally, incubation of

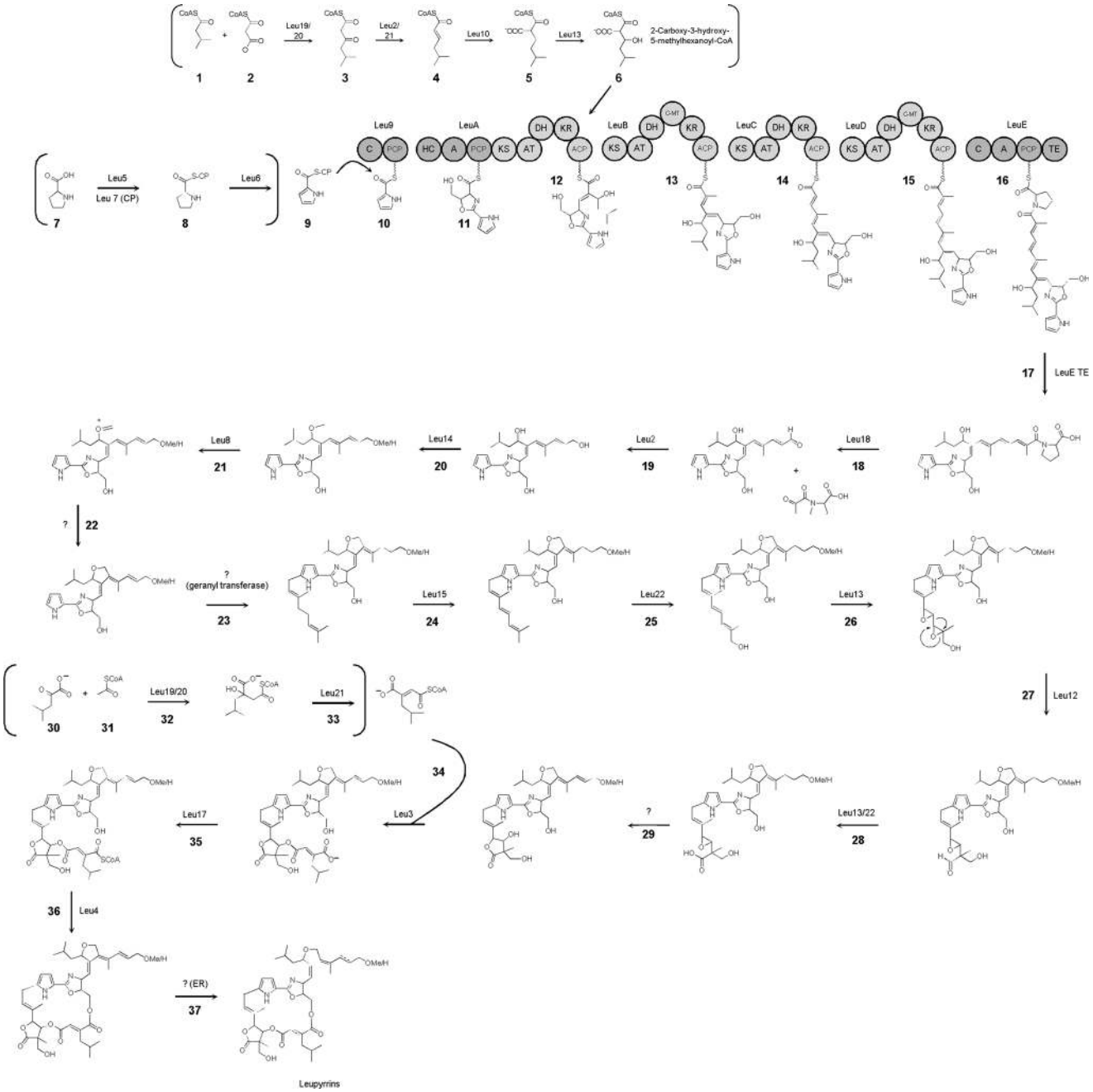

Fig. 3 Summary of the proposed biosynthesis of the leupyrrins, based on a synthesis of the previous feeding studies[17] and analysis of the gene cluster. The biosyntheses of specific building blocks are shown in parentheses. Because not all of the proteins encoded in the leupyrrin biosynthetic gene cluster can be clearly assigned to roles in the pathway, it may be that the functions are instead provided by other enzymes encoded outside of the gene cluster.

Leu5, Leu6 and Leu7 with ATP, L-proline, and FAD yielded the four-electron oxidation product pyrrolyl-*S*-Leu7, as judged by high-resolution MS analysis and Ppant ejection (Fig. 5 and Table 2). Thus, Leu6 is confirmed as a prolyl-*S*-CP dehydrogenase. Consistent with the transient nature of the partially-oxidized dehydro-L-prolyl-*S*-CP,[30] we did not detect any of this intermediate by MS. Taken together, these data firmly establish that together Leu5–Leu7 represent a functional unit for generating the pyrrole carboxylic acid starter unit of leupyrrin biosynthesis.

The next step in the pathway was proposed to be a condensation between the carrier-bound pyrrole carboxylic acid and either hydroxythreonine or threonine (which is later oxidized to hydroxythreonine), followed by a heterocyclization reaction to form the observed oxazoline **11** (Fig. 3).[18] Inspection of the annotated NRPS-PKS portion of the gene cluster shows that only the NRPS module of LeuA contains the HC domain required in this reaction sequence (Fig. 2). To try to determine whether threonine or hydroxythreonine is used in the condensation, we carried out substrate specificity prediction analysis of the

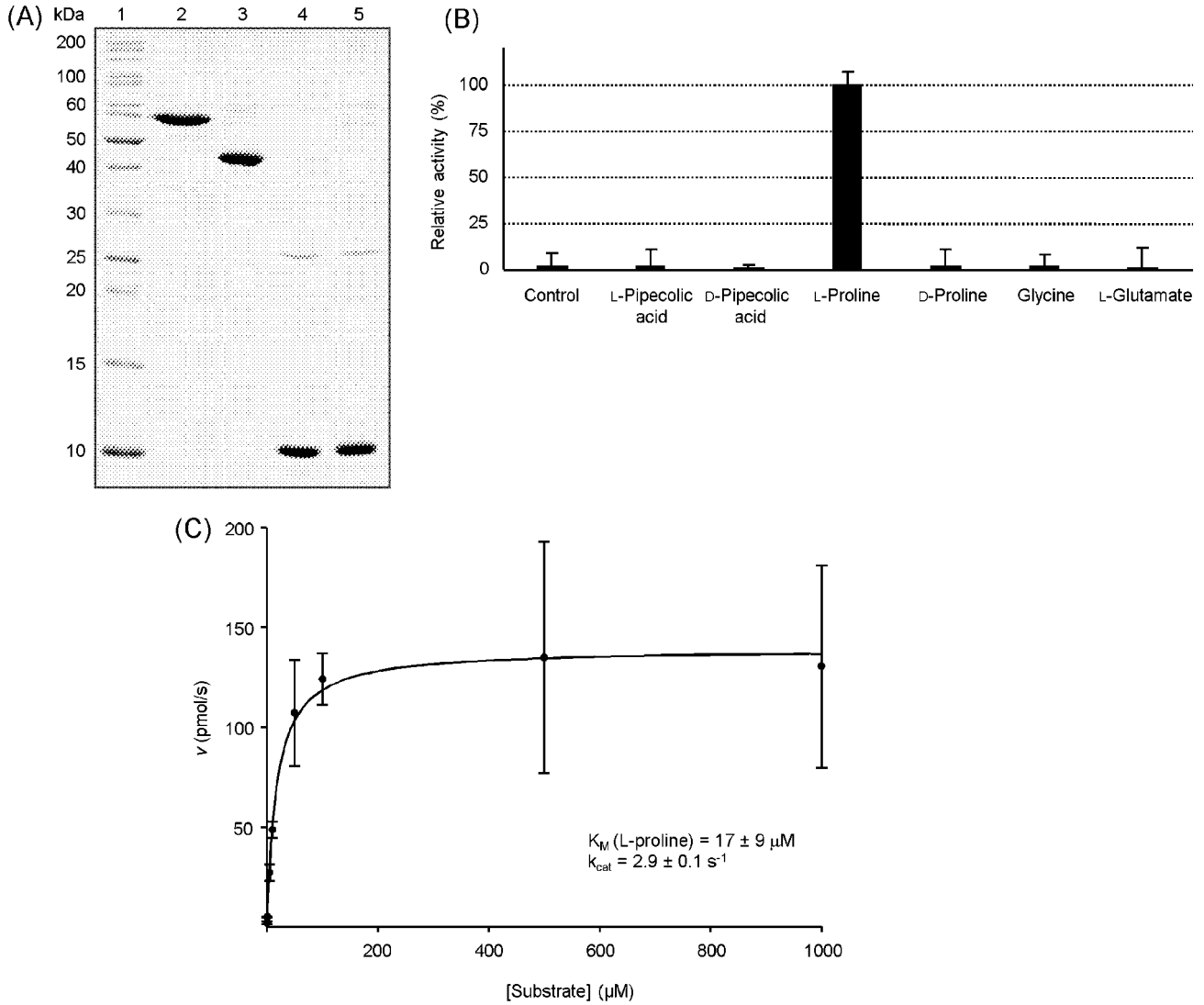

Fig. 4 Expression of the pyrrole biosynthetic cassette and kinetic characterization of Leu5. (A) SDS-PAGE analysis of proteins Leu5, Leu6, and Leu7 expressed in this study. The calculated molecular weights are as follows: Leu5 (58.7 kDa; lane 2), Leu6 (43.4 kDa; lane 3), Leu7 expressed in the absence of MtaA (10.4 kDa; lane 4), and Leu7 expressed in the presence of MtaA (10.7 kDa; lane 5). Lane 1 contains a molecular weight marker. (B) Results of the end-point ATP–[^{32}P]PP_i exchange assay to determine the substrate specificity of the A domain Leu5. Adenylation activities are expressed relative to the activation of L-proline (100%). (C) Determination of the kinetic parameters for activation of L-proline by Leu5.

Table 2 Molecular mass data (Da) obtained by high resolution nanospray ESI-FT-Orbitrap-MS/MS

Identity of protein/4′-phosphopantetheinyl fragment	Calcd[a]	Found
apo Leu7	10385.47	10385.47[b]
holo Leu7	10725.55	10725.57[b]
Pantetheine	260.12	260.12
Phosphopantetheine	358.10	358.10
L-Prolyl-*S*-Leu7	10822.61	10822.61[b]
L-Prolyl-*S*-phosphopantetheine	455.15	455.15
L-Prolyl-*S*-pantetheine	357.17	357.17
Pyrrolyl-2-carboxyl-*S*-Leu7	10818.58	10818.58[b]
Pyrrolyl-2-carboxyl-*S*-phosphopantetheine	451.12	451.12
Pyrrolyl-2-carboxyl-*S*-pantetheine	353.14	353.14

[a] Monoisotopic mass. [b] Detected mass after deconvolution.

LeuA A domain (ref. 35 and 36 and http://www-ab.informatik.uni-tuebingen.de/toolbox). However, the strongest hit was to cysteine-activating A domains, leaving the question of the amino acid activated by LeuA unresolved. The nature of the threonine hydroxylase is likewise unclear, although candidates include the two cytochrome P450s (Leu13 and Leu22) encoded by the cluster. An additional issue is the identity of the protein which acts as the pyrrole-carboxylic acid donor for the condensation reaction. In principle, the condensation could occur between LeuA-bound (hydroxy)threonine and pyrrolyl-*S*-Leu7. However, the cluster encodes another NRPS module Leu9, which comprises a C-PCP didomain. Interestingly, Leu9 exhibits a convincing C-terminal docking domain, which could interact with the 'β-sheet-type' N-terminal docking domain of LeuA[26,44] (Fig. S1, ESI†), while no such recognition

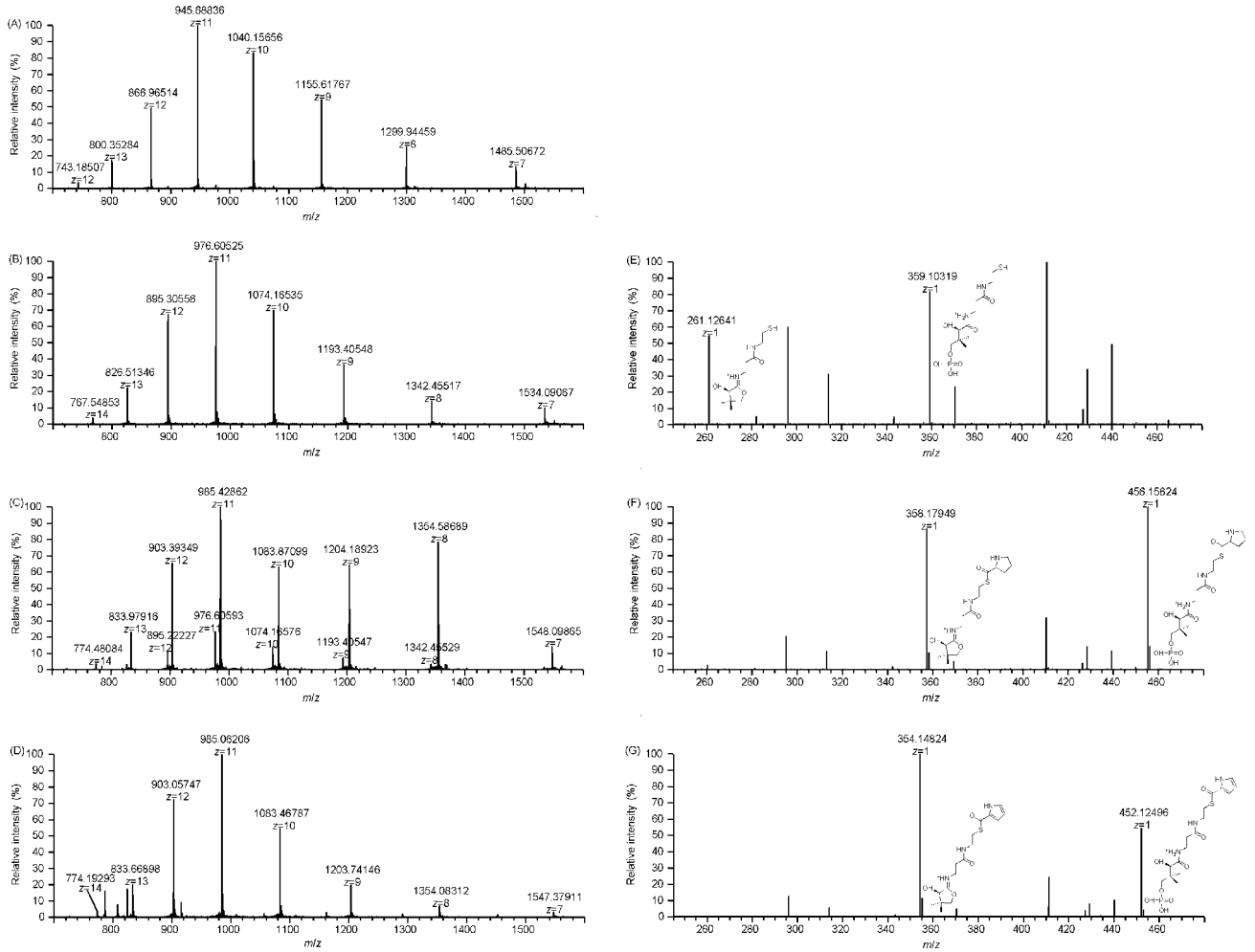

Fig. 5 Reconstitution *in vitro* of pyrrolyl-*S*-Leu7 biosynthesis as demonstrated by ESI-FT-Orbitrap-MS coupled with the Ppant ejection assay. (A) *apo* Leu7. (B) *holo* Leu7. (C) Analysis of Leu7 following incubation with L-proline, ATP, and Leu5 (90 min). (D) Analysis of Leu7 following incubation with L-proline, ATP, FAD, Leu5 and Leu6 (120 min). Ppant ejection analysis of: (E) *holo* Leu7, (F) L-prolyl-*S*-Leu7 and (G) pyrrolyl-*S*-Leu7. The structures of the ejected fragments are shown.

elements are present on Leu7. Thus, another possibility (Fig. 3) is that the pyrrole carboxylic acid moiety is passed from Leu7 **9** to the PCP domain of Leu9 (possibly facilitated by the C domain), and then condensation takes place between pyrrolyl-*S*-Leu9 **10** and the amino acid bound to the LeuA PCP. In either case, the HC domain would then act on the resulting dipeptide to form the oxazoline ring **11** (Fig. 3).

According to the feeding studies, biosynthesis continues with condensation of the threonine carboxylate with an unusual extender unit, 2-carboxy-3-hydroxy-5-methylhexanoyl-CoA **6** (Fig. 3), derived from isovaleryl-CoA. In principle, condensation between isovaleryl-CoA and a suitable building block (*e.g.* malonyl-CoA) could be carried out by a PKS module, which could then install a C2 methyl *via* a *C*-MT domain and the required hydroxy functionality by KR-catalyzed reduction of the 3-ketone. Subsequent oxidation of the C2 methyl to the carboxylic acid (possibly by one of the two P450 enzymes Leu13 or Leu22) would furnish the appropriate extender unit. However, inspection of the PKS modules encoded by the cluster (within LeuA–LeuD) reveals that none incorporates exclusively the set of domains (KS, AT, *C*-MT, KR, ACP) required for this sequence of transformations. In addition, the cluster lacks a PKS loading module (frequently an AT-ACP didomain) which would typically furnish the isovaleryl-CoA to the KS domain. An alternative origin for the extender unit is suggested by the presence in the cluster of two genes, *leu19* and *leu20*, encoding HpcH/HpaI aldolases. Members of this enzyme family catalyze reversible, stereospecific C–C bond formation.[45] Thus, one possible source of the extender units which is consistent with the previous feeding studies[17] is Leu19- or Leu20-catalyzed condensation between malonyl-CoA **2** and isovaleryl-CoA **1** to yield 3-keto-5-methylhexanoyl-CoA **3**, which could then be reduced and dehydrated **4**, and finally reductively carboxylated at C2 to yield the dicarboxylic acid **5** (Fig. 3). Leu10 shows significant homology to a rapidly growing class of enzymes that reductively carboxylate activated unsaturated fatty acids.[46,47] Oxidation at C3 may occur before or after PKS extension to yield **6**. Consistent with this proposal, disruption of *leu19* by insertional mutagenesis resulted in almost complete abolition of leupyrrin biosynthesis (data not shown). Nonetheless, further experiments will be required to conclusively demonstrate the origin of the extender unit.

Table 3 Acyl transferase (AT) active site (underlined) and additional conserved residues that correlate with domain specificity

Domain	11	63	90	91	92	93	94	117	200	201	231	250	255	15	58	59	60	61	62	70	72	197	198	199
LeuA	F	Q	G	H	S	L	G	R	A	H	N	Q	V	S	D	L	S	Y	A	E	C	R	R	A
LeuB	Q	Q	G	H	S	L	G	R	F	H	T	H	V	Q	D	A	V	T	A	Q	A	D	H	A
LeuC	Q	Q	G	H	S	I	G	R	F	H	N	H	V	H	D	T	A	Y	A	E	A	S	H	A
LeuD	Q	A	G	H	S	L	G	R	F	H	N	H	V	Y	T	G	I	Y	S	E	A	P	H	A
Mal consensus	Q	Q	G	H	S	LVIFAM	G	R	FP	H	ANTGEDS	NHQ	V	R	RQED	T	GRLE	YWFH	TAS	EQ	S	S	G	A
MeMal consensus	Q	Q	G	H	S	QM	G	R	S	H	T	N	V	W	REDQ	VIDA	D	V	VLI	MAEQ	SAG	D	YV	A

(Mal = malonate; MeMal = methylmalonate)

As with A domains, particular amino acids within AT domains have been correlated with their substrate specificity.[48–50] Inspection of these residues in the LeuA AT (Table 3) reveals a divergence from the canonical 'malonate' code at several positions, which may reflect its ability to select the atypical building block, 2-carboxy-3-hydroxy-5-methylhexanoyl-CoA. In any case, following the condensation between the NRP portion and the extender unit, the initially-generated C3 ketone group would be processed to the double bond by the sequential action of the KR and DH domains present in the LeuA PKS module **12** (Fig. 3).

We previously proposed that the remainder of the polyketide skeleton of leupyrrin would be generated by two consecutive cycles of chain elongation with malonate. Chain extension would be accompanied in the first cycle by C2-methylation and in both cycles by reduction at C3 to the double bond (**13**, **14**). Consistent with this hypothesis, the domain complements in the downstream PKSs LeuB and LeuC are exactly those required to generate the observed functionalities, and both of the AT domains are predicted to exhibit malonate specificity (Fig. 2 and Table 3). We were thus surprised to discover that the pathway contains an additional PKS module (encoded by LeuD) as well as a subsequent NRPS module (LeuE) (Fig. 2), neither of which seems to be required. However, a possible explanation for these apparently surperfluous modules derives from consideration of the intermediate that would be generated by further chain extension and processing by LeuD—the polyene portion of the chain **15** bears striking resemblance to a carotenoid, and so could be a substrate for the carotenoid oxygenase homolog (Leu18)[51] encoded by the cluster. Indeed, targeted disruption of *leu18* completely abolished leupyrrin production (Fig. S3, ESI†). Subsequent oxidative cleavage of the double bond to yield an aldehyde could occur at a number of stages: following chain extension by LeuD **15**, after further chain extension by LeuE **16**, or following release of the fully-extended intermediate as a free acid by the LeuE TE domain **17** (the latter possibility is depicted in Fig. 3) the predicted specificity of LeuE for proline is again based on analysis of the specificity code (ref. 35 and 36 and http://www-ab.informatik.uni-tuebingen.de/toolbox). The first two mechanisms would result in acyl groups remaining attached to the respective ACP or PCP domains, but these could be released hydrolytically by the type II 'proof-reading' TE[52–54] encoded in the cluster (Leu1). Consistent with the idea that chain extension continues beyond LeuC, analysis of the docking domain regions of LeuC–LeuE identifies appropriate sequence elements at each potential interface (Fig. S1, ESI†), while inspection of all active site motifs within LeuD and LeuE does not reveal any obvious, disabling mutations (data not shown). Following the cleavage reaction **18**, the aldehyde terminus could be reduced to the required alcohol functionality by a dehydrogenase such as Leu2 **19** (Fig. 3).

Completion of the polyketide portion of the molecule requires installation of a five-membered cyclic ether (Fig. 1). On the basis of the feeding studies, the ring was proposed to arise by a mechanism analogous to methylene bridge formation during berberine biosynthesis of plants:[55,56] methylation of the C26 hydroxyl **20**, oxidation of the methyl group **21**, cyclization and finally, deprotonation **22** (Fig. 3). The cluster

contains a gene encoding a candidate *O*-methyltransferase (Leu14) to accomplish the modification of the C26 OH, and the promiscuous activity of this enzyme towards the C21 hydroxyl **20** could also account for the leupyrrins which are methylated at this position (Fig. 1). However, while the dehydrogenase Leu8 may reasonably be proposed to catalyze the desaturation, there are no clear candidates in the cluster to accomplish the subsequent ring formation.

Similarly, it is difficult to account for all of the steps required to introduce and modify the third structural unit of the leupyrrins, the isoprenoid portion (C1–C7/C22–C24). The series of transformations (Fig. 3), which was proposed on the basis of feeding experiments with doubly-^{13}C-labeled acetate,[17] comprises prenylation of the pyrrole moiety with a geranyl-pyrophosphate moiety **23**, followed by desaturation **24** and hydroxylation **25**. Epoxidation of the two terminal double bonds **26** followed by a semipinacol-like rearrangement **27**, aldehyde oxidation **28** and opening of the remaining epoxide **29**, would then generate the observed γ-butyrolactone ring. Most obviously, the cluster lacks a prenyltransferase homolog, and so this enzyme must be encoded elsewhere in the genome. (The prenyltransferase encoded downstream of the assigned cluster boundaries was insertionally inactivated to exclude the possibility that the enzyme may have a dual function, and operate in the leupyrrin pathway. The resulting mutant was found to be unaffected in leupyrrin biosynthesis; data not shown.) The next step, desaturation, may be catalyzed by the gene product of *leu15* which shows homology to Rieske (2Fe–2S) iron–sulfur domain proteins (Table 1). We reasoned that the subsequent hydroxylation and epoxidation reactions were likely to be performed by the two P450 enzymes (Leu13 and Leu22). To probe directly the possible roles of Leu13 and Leu22 in the biosynthesis, *leu13* and *leu22* were inactivated by insertional mutagenesis. *leu22* lies at the cluster boundary and is transcribed in the opposite direction to *leu21*, so we did not anticipate any polar effects from this gene inactivation (Fig. 2). On the other hand, analysis of the intergenic region separating *leu13* and *leu14* did not reveal any obvious σ54 or σ70 promoter binding sequences (data not shown), suggesting that transcription of *leu13* may be coupled to that of the downstream genes *leu14–leu16*. Analysis by HPLC-MS of the *leu22*$^-$ mutant showed that it no longer produced the leupyrrins, but instead a series of novel metabolites with molecular weights corresponding to deshydroxy derivatives of the leupyrrins (deshydroxy leupyrrin A_1 (rt = 19.2 min; [M + H] = 723.42055 (calculated: 723.42151 Δ = −1.3 ppm)); deshydroxy leupyrrin A_2 (rt = 19.5 min; [M + H] = 721.40502 (calculated: 721.40586 Δ = −1.2 ppm)); deshydroxy leupyrrin B_1 (rt = 21.0 min; [M + H] = 721.40485 (calculated: 721.40586 Δ = −1.4 ppm)); and deshydroxy leupyrrin B_2 (rt = 21.2 min; [M + H] = 719.38924 (calculated: 719.39021 Δ = −1.3 ppm))) (Fig. 6).

The four major compounds were purified and their structures elucidated in full by combined 1D and 2D NMR spectroscopy (Table 4). This analysis revealed the analogues to be the C22-deshydroxyl derivatives of leupyrrins A_1, A_2, B_1 and B_2, demonstrating the identity of Leu22 as a C22 hydroxylase. This finding implicates Leu13 as the required epoxidase, and indeed multiple sequence alignment shows that it groups with

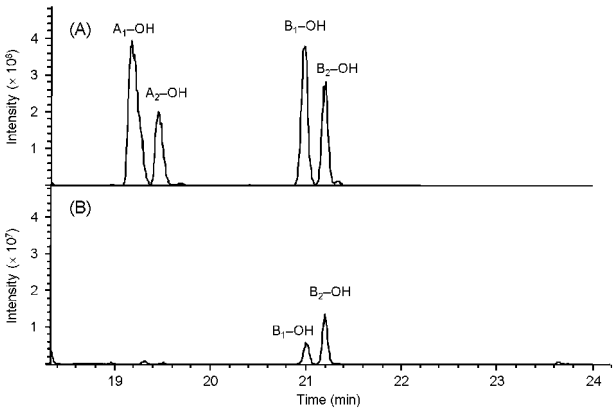

Fig. 6 Base peak chromatogram (m/z = 718–725) of the (A) *leu22*$^-$ and (B) *S. cellulosum* So ce690 wild type. In the mutant, peaks corresponding to following new leupyrrin derivatives were identified: deshydroxy leupyrrin A_1 (rt = 19.2 min; [M + H] = 723.42055 (calculated: 723.42151 Δ = −1.3 ppm)), deshydroxy leupyrrin A_2 (rt = 19.5 min; [M + H] = 721.40502 (calculated: 721.40586 Δ = −1.2 ppm)), deshydroxy leupyrrin B_1 (rt = 21.0 min; [M + H] = 721.40485 (calculated: 721.40586 Δ = −1.4 ppm)), deshydroxy leupyrrin B_2 (rt = 21.2 min; [M + H] = 719.38924 (calculated: 719.39021 Δ = −1.3 ppm)). The *S. cellulosum* So ce690 wild type also produces deshydroxy leupyrrin B_1 and deshydroxy leupyrrin B_2 in relatively lower amounts (B).

other P450 epoxidases from various secondary metabolic pathways (*e.g.* EpoK,[57] PimD,[58] MycG[59] and OleP[60]) (Fig. S4, ESI†). Consistent with an essential role in the biosynthesis, HPLC-MS analysis of the *leu13*$^-$ mutant revealed the complete absence of leupyrrins (Fig. S3, ESI†). However, no novel derivatives were detected. Thus, either Leu13 plays additional roles at earlier stages in the pathway as suggested previously (*e.g.* as a tyrosine hydroxylase or during extender unit biosynthesis), or disruption of *leu13* resulted in polar effects on genes *leu14–leu16*, obscuring Leu13's function. The identity of the enzyme responsible for the subsequent semipinacol rearrangement is not immediately apparent, but a possible candidate is Leu12, as it shows weak homology to a geranylgeranyl reductase (Table 1), while lacking the typical FAD-binding motif of these enzymes. The only strong candidates in the cluster for the oxidation of the aldehyde to the carboxylic acid are again Leu13 and Leu22. Ring opening of the epoxide to form the γ-butyrolactone may occur spontaneously, or be performed by yet another enzyme absent from the identified gene set.

In the next stage of the biosynthesis, the NRP-PK-isoprenoid intermediate is coupled to both ends of a dicarboxylic acid 'linker' to form the overall macrolactone ring. By means of the feeding experiments, the dicarboxylic acid was determined to derive from condensation of leucine-derived 2-ketoisocaproate **30** with acetyl-CoA **31**.[17] The mechanistic similarity between this reaction and that required to generate the extender unit suggests that either of the aldolases Leu19 or Leu20 may catalyze this reaction. The condensation **32** would then be followed by dehydration, possibly accomplished by the dehydratase Leu21 **33**. The first link to the NRP-PK-isoprenoid intermediate would logically be formed by attack of the C3 hydroxyl on the CoA-activated end of the dicarboxylic acid **34**.

Table 4 ^{1}H NMR spectroscopic assignments of the deshydroxy leupyrrins A_1, A_2, B_1 and B_2

Position (H)	Deshydroxy leupyrrin A_1 δ_H (ppm) (m, *J* [Hz])	Deshydroxy leupyrrin A_2 δ_H (ppm) (m, *J* [Hz])	Deshydroxy leupyrrin B_1 δ_H (ppm) (m, *J* [Hz])	Deshydroxy leupyrrin B_2 δ_H (ppm) (m, *J* [Hz])
3	5.52 (d, 10.0)	5.6 (d, 10.5)	5.50 (d, 9.0)	5.61 (d, 9.5)
4	4.73 (d, 10.0)	4.95 (d, 10.5)	4.7 (d, 9.0)	4.75 (d, 9.5)
6	5.48 (m)	5.40 (m)	5.38 (m)	5.41 (m)
7	3.31 (m)	3.32 (m)	3.33 (m)	3.33 (m)
	3.74 (m)	3.75 (m)	3.74 (m)	3.74 (dd, 11.5, 14.5)
9	6.04 (m)	6.20 (m)	6.25 (m)	6.25 (d, 4.0)
10	7.04 (m)	7.10 (m)	7.16 (m)	7.15 (d, 4.0)
13	5.10 (m)	5.04 (m)	5.00 (m)	5.01 (m)
14	4.50 (m)	4.77 (m)	4.75 (m)	4.75 (dd, 9.5, 4.0)
15	5.47 (d, 10.0)	5.60 (d, 10.0)	5.50 (d, 9.5)	5.60 (d, 9.5)
19	2.03 (m)	6.18 (m)	2.03 (m)	6.20 (m)
20	1.67 (m)	5.90 (ddd, 15.0, 5.0, 5.0)	1.70 (m)	5.92 (ddd, 15.0, 5.0, 5.0)
21	3.3 (m)	4.02 (dd, 6.0, 1.0)	3.30 (m)	4.02 (dd, 6.0, 1.0)
22	1.53 (s)	1.46 (s)	1.53 (s)	1.53 (s)
23	1.12 (s)	1.15 (s)	1.12 (s)	1.12 (s)
24	1.91 (s)	1.80 (s)	1.95 (s)	1.94 (s)
25	4.47 (dd, 12.0, 2.5)	4.47 (m)	4.45 (dd 12.0, 2.5)	4.45 (m)
	4.62 (m)	4.63 (m)	4.63 (m)	4.74 (m)
26	4.88 (m)	4.90 (m)	4.80 (m)	4.80 (m)
27	1.42 (m)	1.40 (m)	1.40 (m)	1.40 (m)
	1.59 (m)	1.60 (m)	1.60 (m)	1.58 (ddd, 14.5, 10.0, 3.0)
28	1.95 (m)	1.85 (m)	1.83 (m)	1.81 (m)
29	0.88 (d, 6.5)	1.01 (d, 6.5)	1.00 (d, 6.5)	1.00 (d, 6.5)
30	0.96 (d, 6.5)	0.95 (d, 6.5)	0.98 (d, 6.5)	0.98 (d, 6.5)
31	4.67 (m)	4.67 (m)	4.69 (m)	4.67 (m)
	4.63 (m)	4.63 (m)	4.64 (m)	4.63 (m)
32	2.03 (br s)	2.04 (br s)	2.10 (s)	2.00 (br s)
2′	2.38 (m)	2.44 (m)	6.60 (s)	6.61 (s)
	2.50 (m)	2.64 (m)		
3′	2.72 (m)	2.72 (m)	—	—
5′	1.60 (m)	1.6 (m)	2.39 (dd, 12.5, 7.0)	2.40 (dd, 12.5, 7.0)
	1.23 (m)	1.24 (m)	2.49 (dd, 12.5, 7.0)	2.56 (dd, 12.5, 7.0)
6′	1.54 (m)	1.53 (m)	1.70 (m)	1.66 (ddd, 13.5, 7.0, 7.0)
7′	0.86 (d, 6.0)	0.86 (d, 6.0)	0.87 (m)	0.85 (d, 6.5)
8′	0.84 (d, 6.0)	0.84 (d, 6.0)	0.77 (d, 6.5)	0.76 (d, 6.5)
OCH3	3.31 (s)	3.35 (s)	3.31 (s)	3.35 (s)
NH	8.90 (br s)	8.90 (br s)	—	8.80 (br,s)

This reaction may be accomplished by the putative acyl transferase Leu3. Leu3 shows weak but convincing homology (Table 1) to hydroxycinnamoyl CoA shikimate/quinate hydroxycinnamoyltransferase-like proteins, whose shikimate/quinate substrates bear some resemblance to the leupyrrin γ-butyrolactone. Activation of the second carboxylic acid may be accomplished by Leu17 **35**, which exhibits homology to very-long chain acyl-CoA synthetases, followed by ring closure catalyzed by Leu4 **36**, a predicted member of the WS/DGAT/MGAT family which acts as acyl-coenzyme A:glycerol acyl transferases.

The final modifications to the structure appear to be the partial reduction of the C19/C20 and C2/3 double bonds **37** (Fig. 3). However, the cluster does not contain any gene obviously encoding for the enoyl reductase domain(s) responsible for these transformations, suggesting that these genes are also located outside of the cluster boundaries.

Analysis of the remaining genes

The protein product of *leu16* shows convincing homology to TetR-family transcriptional regulators (Table 1), helix-turn-helix (HTH) DNA-binding proteins which are widely distributed among bacteria.[61] This type of regulator is known to govern the transcriptional control of secondary metabolic pathways, and so is presumably involved in regulating leupyrrin biosynthesis. The proteins encoded by the remaining genes (*leu11* and orfs 2–4) (Table 1) do not have an obvious role to play in the pathway. Unfortunately, it was not possible to evaluate their possible functions by gene inactivation due to probable polar effects on adjacent genes (Fig. 2). In any case, double cross-over experiments to generate markerless deletions have not, to our knowledge, been reported for the genus *Sorangium*, despite concerted efforts by various groups.

Conclusions

Identification of the core gene cluster encoding for leupyrrin biosynthesis has allowed us to assess our previous model for the pathway which was developed on the basis of feeding studies.[17] Although gene products can be assigned to a number of postulated steps, it is clear that the encoded complement of enzymes is insufficient to account for all of the required transformations. This lack is particularly clear in the case of oxidative functions, and the required prenyl transferase. Thus, it appears that a significant fraction of the pathway genes is located elsewhere in the genome, a 'split' cluster organization which is increasingly common in the myxobacteria.[9,62]

Fischbach, *et al*. recently coined the term 'super cluster' to describe a gene set which has evolved by the merging of several biosynthetic sub-clusters into a single locus.[63] Evidently in the case of leupyrrin, these 'sub-cluster joining events' are as yet incomplete.

Examination of the cluster also inspired the revision of several key hypotheses, including most notably, biosynthesis of the PK-NRP portion of the molecule. Our analysis suggests that the intermediate is extended by a polyketide unit and possibly by an amino acid, before undergoing an oxidative cleavage reaction catalyzed by a carotenoid-oxygenase-like domain. Cleavage of an extended precursor has a precedent in the pathways to the myxobacterial metabolite myxothiazol[64] and the insect toxin pederin;[65,66] in the case of myxothiazol, the intermediate is elongated with glycine prior to oxidative release, while the pederin chain may be extended by as many as three acetate units and one arginine, before the oxygenase acts. Nonetheless, the proposed mechanism of oxidative release in the leupyrrin case would be wholly novel. While gene inactivation in *Sorangium* species is notoriously difficult,[67] we were able to provide direct support for the role of several genes in the pathway using this method. In the majority of cases (*leu5*, *leu13*, and *leu17–leu19*), these modifications lead to the complete loss of leupyrrin production, but disabling of *leu22* resulted in the first four leupyrrin analogues produced by genetic engineering, C22-deshydroxy leupyrrin A_1, A_2, B_1 and B_2. Additional proof for our hypothesis has been provided by the complete reconstitution *in vitro* of the biosynthesis of the pyrrole carboxylic acid starter unit.

Experimental

Strains and culture conditions

S. cellulosum So ce690 was grown at 30 °C and 160 rpm in liquid HS medium and on solid PM plates as described previously.[16] Conjugation was performed as described previously,[67] and mutants were cultivated in HS containing 150 μg ml^{-1} hygromycin B. For secondary metabolite analysis, the strain was grown in SG medium. *E. coli* DH10B, *E. coli* ET12567/pUZ8002 and *E. coli* SURE were grown in liquid LB medium at 37 °C. The indicator strain *Debaryomyces hansenii* was grown in MYC medium and on EBS agar. When used, additional antibiotics were present at the following concentrations: kanamycin sulfate (Kan) (50 μg ml^{-1}), chloramphenicol (Cm) (20 μg ml^{-1}), and ampicillin (Amp) (20 μg ml^{-1}).

General molecular biology methods

S. cellulosum So ce690 chromosomal DNA was prepared as described previously.[68] Plasmid DNA was isolated using the GeneJET™ Plasmid Miniprep Kit (Fermentas). DNA fragments were purified from agarose gels with the NucleoSpin Extract gel extraction kit (Macherey-Nagel). Polymerase chain reactions (PCR) were carried out using Taq DNA polymerase (MBI Fermentas) or Phusion Polymerase (Invitrogen). DMSO was added to the reaction mixtures to a final concentration of 5%. Conditions for amplification with an Eppendorf Mastercycler Gradient Thermal Cycler were as follows: denaturation, 15–30 s at 95/98 °C; annealing, 20 s at 50–62 °C; extension, 15–60 s at 72 °C (30 cycles); final extension at 72 °C for 10 min. All other DNA manipulations were performed according to standard protocols.[69] *In silico* DNA and amino acid sequence analyses were performed with the VectorNTI software package and ClustalW,[70] and BLAST[25] was used for comparison with GenBank data. Annotation of the leupyrrin gene cluster was performed using Frameplot.[24]

Identification and annotation of the leupyrrin gene cluster

A 1.1 kbp fragment was amplified from *S. cellulosum* So ce690 genomic DNA using degenerate HC domain primers.[19] This fragment was subcloned into pCR2.1TOPO (Invitrogen) and sequenced. The obtained HC domain sequence was used for generating an appropriate probe for screening the cosmid library, as previously described.[47] The initial hybridization revealed 2 cosmids, of which C9 was T7 and T3 end sequenced. Cosmid C9 showed homology to genes unrelated to leupyrrin biosynthesis at its T7 end and homology to NRPS genes at its T3 end. A portion of the NRPS region was amplified from cosmid C9 and used to re-screen the cosmid library for overlapping sequence. This analysis yielded 5 cosmids, of which cosmid C9.6 showed the lowest extent of overlap with cosmid C9 based on PCR analysis. Sequencing of the two cosmids was performed using a shotgun approach, as described previously.[65]

The amplified HC domain was also cloned into the conjugation plasmid pSUPHyg, which was then used for conjugation into *S. cellulosum*, generating 200 mutant clones. Leupyrrin-negative mutants were then identified by bioassay. *S. cellulosum* So ce690 HC mutants were grown on PM plates without hygromycin to visible colonies. The PM plates were then overlaid with the indicator strain *Debaryomyces hansenii* mixed with liquid EBS soft agar (1 : 500). The plates were then incubated overnight at 30 °C. Leupyrrin-negative mutants were identified by their inability to inhibit the growth of *D. hansenii*. Additional analysis of two leupyrrin negative mutants by HPLC-MS confirmed that leupyrrin production had been abolished. Genetic verification of the mutants was carried out by Southern Blot. For this, the genomic DNA of the mutants was digested with *Mlu*I and hybridized with a DIG-labeled internal fragment of the HC domain. One distinct band (*ca.* 10 kbp) was detected in the case of the wild type strain, whereas the mutants yielded a much larger band (*ca.* 19 kbp) resulting from the integrated pSUPHygHC.

Gene inactivation experiments

Inactivation of the genes *leu5*, *leu13*, *leu17*, *leu18*, *leu19* and *leu22* in *S. cellulosum* So ce690 was achieved by amplifying homologous inactivation fragments from cosmid C9 using the following primers: leu_orf5_fwd (5′-CTC**AAGCTT**GAGCGGCCCGTCGCTTC-3′), leu_orf5_rev (5′-CTGGCC**GGATCC**AGTACCCCTTCGTCAG-3′), leu_orf13fwd (5′-TCCT**AGATCT**TGAAGGCAGGAGTACG-3′), leu_orf13_rev (5′-CTCGGG**AAGCTT**CTCCCAGTGATCA-3′), leu_orf17fwd (5′-CTCATC**AAGCTT**CGCCACGCTGGTAG-3′), leu_orf17rev (5′-GAGGCG**GGATCC**CCACGAACGAGACC-3′), leu_orf18fwd (5′-GACC**AAGCTT**GAGAACGGGAGCATG-3′), leu_orf18rev (5′-GGAAACT**GGATCC**CACCCCACAGG-3′),

leu_orf19fwdI (5′-GTGTGA**AAGCTT**GTGGTCCAGCAG-ATCAC-3′), leu_orf19revI (5′-CCTTCTGCCGTCAGCCGTC-CTCC-3′), leu_orf19fwdII (5′-GGAGGACGGCTGACGGCA-GAAGG-3′), leu_orf19revII (5′-GTGCTT**AGATCT**GGTGGA-TGACCGGCTTG-3′), leu_orf22_fwd (5′-TGAAGC**AGATCT**-TGATTTGCAGGATCCAC-3′), and leu_orf22_rev (5′-CGGGT-T**AAGCTT**GCTTCTTGTCGAG-3′) (introduced restriction sites *BamH*I, *Hind*III and *Bgl*II are shown in bold). The PCR products were cloned into pCR2.1TOPO (Invitrogen). The fragments were then excised using *BamH*I, *Hind*III or *Bgl*II restrictions sites, and subcloned into pSUPHyg to generate the conjugation plasmids pSUPHygleu5, pSUPHygleu13, pSUPHygleu17, pSUPHygleu18, pSUPHygleu19 and pSUPHygleu22. After conjugation in *S. cellulosum* So ce690, correct insertion of the conjugation plasmids into the *S. cellulosum* genome was confirmed by Southern Blot (data not shown).

Analysis of secondary metabolite production

S. cellulosum So ce690 and mutants were grown in SG medium containing 1% XAD adsorber resin (Rohmer and Haas) until the starch and glucose were consumed completely. The XAD beads were then separated from the culture and extracted stepwise with methanol, a mixture of methanol/ethanol/2-propanol (400/75/25) and finally with acetone. The combined extracts were evaporated and the residue was dissolved in methanol resulting in a 100-fold concentration of the original culture volume. The extracts were analyzed using a HPLC system 1090 series II equipped with a diode array detector (Hewlett-Packard). Separation was achieved using a Nucleosil column (Macherey and Nagel; 125 × 3 mm, 120-5-C18, flow 0.5 ml min^{-1}), with a solvent system consisting of water and methanol. The following gradient was applied: 45–100% methanol over 30 min. Compounds were detected at 300 nm with a bandwidth of 20 nm.

Cloning of the expression constructs for Leu5–Leu7

Amplification of the genes *leu5* and *leu6* by PCR was achieved using cosmid C9 containing the appropriate part of the leupyrrin gene cluster. The forward primers for *leu5* and *leu6* (leu5fwd (5′-CTGCT**CATATG**ACGTACCTGTTGCATCA-GC-3′), leu6fwd (5′-TTCCC**CATATG**TCATCCTACAC-CAGG-3′)) contained an artificial *Nde*I site, whereas the reverse primers (leu5rev (5′-GACAC**GCGGCCGC**CTAC-TCGCCCCGTTCAAGC-3′), leu6rev (5′-TACGA**GCGGCCGC**-TCAAAGGCCCATGAGCTG-3′)) incorporated an artificial *Not*I site (restriction sites shown in bold). The PCR products were cloned into pJET1.2 (Fermentas), digested with *Nde*I and *Not*I, and subsequently ligated into the expression vector pET-28b+ (Novagen) previously digested with *Nde*I and *Not*I, for expression as N-terminally His_6-tagged proteins. The obtained expression constructs were designated as pET28b-leu5 and pET28b-leu6, and were verified by sequencing. Amplification of the gene *leu7* by PCR was also achieved using cosmid C9. The forward primer for *leu7* (leu7fwd (5′-ACAGC**GGATCC**-ATGGAGGACATCAAAGCTCCGCTCC-3′)) included an artificial *Bam*HI site, whereas the reverse primer (leu7rev (5′-GGGGA**GTCGAC**CTAGACGCTGCGGGCC-3′)) contained an artificial *Sal*I site (restriction sites shown in bold). The PCR product was cloned into pJET1.2 (Fermentas), digested with *Bam*HI and *Sal*I, and subsequently ligated into the expression vector pGEX-6P-1 (GE Healthcare) previously digested with *Bam*HI and *Sal*I, for expression as a C-terminal fusion with glutathione-*S*-transferase (GST). The obtained expression construct was designated as pGEX-6P-leu7, and was verified by sequencing.

Heterologous expression and purification of Leu5–Leu7

The expression construct pET28b-leu5 was transformed into *E. coli* strain Rosetta 2 (DE3) pLysS (Novagen). Expression was carried out in LB medium (200 ml containing kanamycin sulfate (40 μg ml^{-1}) and chloramphenicol (20 μg ml^{-1}) at 37 °C. Protein expression was induced at OD_{600} = 0.8 by addition of isopropyl-β-D-thiogalactopyranoside (IPTG) to a final concentration of 0.2 mM. After induction, the cells were cultivated at 16 °C overnight and then harvested by centrifugation (15 344*g*, 5 min, 4 °C). The expression construct pET28b-leu6 was co-transformed with the plasmid pGro7 (Takara Bio Inc.)[71,72] encoding the *E. coli* chaperones GroEL and GroES into *E. coli* strain BL21 (DE3) (Novagen). Expression was carried out in LB medium (200 ml) containing kanamycin sulfate (50 μg ml^{-1}), chloramphenicol (20 μg ml^{-1}), and L-arabinose (2 mg ml^{-1}) at 37 °C. Expression of Leu6 was induced at OD_{600} = 0.9 by addition of IPTG to a final concentration of 1 mM. After induction, the cells were cultivated at 30 °C for 3 h and then harvested by centrifugation (15 344*g*, 5 min, 4 °C). The expression construct pGEX-6P-leu7 was transformed into *E. coli* strain BL21 (DE3) (Novagen). Expression was carried out in LB medium (50 ml) containing ampicillin (100 μg ml^{-1}) at 37 °C. Protein expression was induced at OD_{600} = 0.9 by addition of IPTG to a final concentration of 0.2 mM. After induction, the cells were cultivated at 16 °C overnight and then harvested by centrifugation (15 344*g*, 5 min, 4 °C). In a second approach, the protein Leu7 was coexpressed with the broad specificity 4′-phospho-pantetheinyl transferase (PPTase) MtaA from *Stigmatella aurantiaca* DW4/3-1, encoded by the plasmid pSUMtaA.[40] During expression, the conditions were as described above for the expression of Leu7 alone, but the medium additionally contained chloramphenicol (20 μg ml^{-1}).

Purification of Leu5 and Leu6 was carried out at 4 °C using an Äktaprime™ Plus Purification System (GE Healthcare). Cell pellets were resuspended in buffer A (20 mM Tris-HCl (pH 7.8), 200 mM NaCl, 10% glycerol (v/v), 10 mM imidazole; 20 ml). The cells were broken by three passes through a French Press (1000 psi) and the insoluble material was removed from the lysate by centrifugation (15 344*g*, 45 min, 4 °C). The lysate was filtered through a 1.2 μm syringe filter (PALL®), before being applied to a 1 ml HisTrap™ HP column (GE Healthcare). All steps of the purification were carried out at a flow rate of 1 ml min^{-1}. The protein extract (20 ml) was loaded onto the column after an equilibration step with buffer A (20 ml). After loading, the column was washed with buffer A (20 ml), and then the proteins were eluted using a stepwise gradient with buffer B (buffer A + 500 mM imidazole) to give concentrations of 60, 100, 200, 300 and 500 mM imidazole. Elution of the proteins was monitored by

recording the absorbance at 280 nm, and appropriate fractions were analyzed by SDS-PAGE. The fractions containing the recombinant protein Leu5 were pooled, concentrated with an Amicon Ultra-4 concentrator (10 kDa cut-off; Millipore) and desalted using a PD-10 column (GE Healthcare) into buffer C (50 mM Tris-HCl (pH 7.6), 1 mM EDTA, 2 mM DTT, 10% glycerol (v/v)). The fractions containing the recombinant protein Leu6 were pooled, concentrated with an Amicon Ultra-4 concentrator (10 kDa cut-off; Millipore) and desalted using a PD-10 column (GE Healthcare) into a storage buffer (50 mM Tris-HCl (pH 8.0), 100 mM NaCl, 1 mM EDTA, 1 mM DTT, 10% glycerol (v/v)). Purified protein Leu6 was then flash frozen in liquid nitrogen and stored at −80 °C. The identity of the protein was confirmed by MALDI-ToF MS/MS analysis and the protein concentration was determined using the Bradford assay[73] (Bio-Rad). Typically, 8.7 mg of purified protein Leu6 were obtained from 200 ml cell culture.

In the case of Leu5, a second purification step was carried out at 4 °C with the Äktaprime using anion exchange chromatography. For this, a 1 ml HiTrap™ Q HP column (GE Healthcare) was utilized at a flow rate of 1 ml min^{-1}. The pre-purified protein sample (3.5 ml) was loaded onto the column after an equilibration step with buffer C. After loading (15 ml), the column was washed with buffer C (5 ml) followed by elution of the proteins using a linear gradient with buffer D (buffer C + 1 M NaCl; 0–1 M NaCl over 25 ml). Fractions (2 ml) were collected and elution of the proteins was monitored by recording the absorbance at 280 nm. Appropriate fractions were analyzed by SDS-PAGE and fractions containing the recombinant protein Leu5 were pooled, concentrated with an Amicon Ultra-4 concentrator (10 kDa cut-off; Millipore) and desalted using a PD-10 column into storage buffer. Purified protein Leu5 was then flash frozen in liquid nitrogen and stored at −80 °C. The identity of the protein was confirmed by MALDI-ToF MS/MS analysis and the protein concentration was determined using the Bradford assay (Bio-Rad). Typically, 4.5 mg of purified protein Leu5 were obtained from 200 ml cell culture.

Purification of Leu7 (with and without co-expression of MtaA) was carried out at 4 °C using the GST SpinTrap™ Purification Module (GE Healthcare). Cell pellets were resuspended in PBS (140 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, 1 mM DTT (pH 7.3); 1.5 ml). The cells were broken by three passes through a French Press (700 psi) and the insoluble material was removed from the lysate by centrifugation (20 817*g*, 45 min, 4 °C). The lysate was loaded onto a SpinTrap column (GE Healthcare) and the purification was carried out according to the manufacturer's protocol. The GST tag was removed by on-column enzymatic cleavage with 20 U PreScission Protease (GE Healthcare) in 150 µl cleavage buffer (50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 1 mM EDTA, 1 mM DTT) at 4 °C overnight. The collected eluate was analyzed by SDS-PAGE. Purified protein Leu7 was then flash frozen in liquid nitrogen and stored at −80 °C. The protein concentration was determined using the Bradford assay (Bio-Rad). Typically, 53 µg of purified protein Leu7 were obtained from 50 ml cell culture without co-expression of MtaA, and 47 µg of purified protein Leu7 were obtained from 50 ml cell culture after co-expression of MtaA.

Detection of the Leu6 flavin cofactor

150 µg of purified Leu6 protein were boiled for 10 min, and the denatured protein was removed by centrifugation (20 817*g*, 10 min, 4 °C). The flavin present in the yellow supernatant was analyzed using an HCTplus ESI-MS ion trap instrument operating in a negative ionization mode. Separation was achieved by HPLC with a Luna® C_{18}(2)-HST column (Phenomenex; 100 × 2 mm; 2.5 µm particle size, flow rate 0.4 ml min^{-1}) using solvent A (ddH_2O and 0.1% formic acid (v/v)) and solvent B (acetonitrile and 0.1% formic acid (v/v)). The following gradient was applied: 2 min isocratic development at 95% A/5% B; 20 min linear gradient from 95% A/5% B to 5% A/95% B; 3 min isocratic development at 5% A/95% B.

ATP–[^{32}P]PP_i exchange assay for aminoacyl-AMP formation

The substrate specificity of Leu5 was probed using the ATP–[^{32}P]PP_i exchange reaction.[43] Reactions (100 µl, 30 °C) contained 75 mM Tris-HCl (pH 8.0), 10 mM $MgCl_2$, 100 mM NaCl, 2 mM dATP, 10 µM amino acid substrate (L-proline, D-proline, L- and D-pipecolic acid, L-glutamate and glycine) and 0.1 µCi [^{32}P] pyrophosphate. The reactions were initiated by addition of Leu5 to a final concentration of 50 nM. Reactions were incubated for 1 min 30 s and then quenched with charcoal suspension (500 µl of 1.2% (w/v) activated charcoal, 0.1 M tetrasodium pyrophosphate, 0.35 M perchloric acid). The charcoal suspension was pelleted by centrifugation, washed twice with quenching buffer lacking charcoal and then resuspended in 500 µL of water and submitted for liquid scintillation counting. Varying concentrations of L-proline (0.5, 1, 5, 10, 50, 100, 500 µM and 1 mM) were used to measure the kinetic parameters of L-proline activation. The kinetic constants were calculated by fitting the data to the Michaelis–Menten equation by nonlinear regression analysis, using SigmaPlot.

Formation of L-prolyl-AMP was demonstrated directly, as follows. Reactions (50 µl, 30 °C) contained 50 mM Tris-HCl (pH 8.0), 10 mM $MgCl_2$, 5 mM L-proline, 5 mM ATP, 2.5 mM TCEP (pH 8.0), and 2 µM Leu5. The reactions were incubated for 5, 30, and 60 min and then quenched with 1% TCA (w/v). The resulting solution was analyzed using a HCTplus ESI-MS ion trap instrument operating in negative ionization mode. Separation was achieved by HPLC with a Synergi™ Hydro-RP C_{18} column (Phenomenex; 150 × 2 mm; 4 µm particle size, flow rate 0.4 ml min^{-1}), using solvents C (20 mM ammonium formate, and 0.004% formic acid (v/v) in ddH_2O) and D (solvent C + 50% methanol (v/v)). The following gradient was applied: 2 min isocratic development at 95% C/5% D; 20 min linear gradient from 95% C/5% D to 0% C/100% D; 5 min isocratic development at 0% C/100% D.

Reconstitution of the biosynthesis of pyrrole-Leu7 *in vitro*

To assay for transfer of proline to *holo* Leu7 by Leu5, reactions (50 µl, 30 °C) contained 50 mM Tris-HCl (pH 8.0), 10 mM $MgCl_2$, 5 mM L-proline, 5 mM ATP, 2.5 mM TCEP, 0.2 µM Leu5, and 20 µM *holo* Leu7. The reactions were incubated for 30, 60, 90, and 120 min and then analyzed by high resolution nanospray ESI-FT-Orbitrap-MS/MS analysis.

For this, the Leu7 samples (*apo*, *holo*, L-prolyl-, and pyrrolyl-2-carboxyl-*S*-Leu7) were first purified using C_{18} ZipTips (Millipore). To prepare each sample, the ZipTip was washed five times with 10 µl of acetonitrile and five times with 10 µl of H_2O/0.1% TFA (v/v). The sample was loaded onto the ZipTip, the ZipTip was washed eight times with 10 µl of H_2O/0.1% TFA (v/v), and then the sample was eluted into 2 µl of 78% acetontrile/2% acetic acid (v/v). To prepare the sample for injection into the mass spectrometer, 20 µl of nanospray solution (49% methanol, 49% H_2O, and 2% formic acid (v/v/v)) was added to the eluted sample and mixed thoroughly. The samples were analyzed on a high resolution ESI-FT LTQ Orbitrap™ mass spectrometer (Thermo Scientific) operating in positive ionization mode and equipped with a TriVersa™ NanoMate ion source (Advion BioSciences). The samples were introduced using a pressure of 0.4 psi, and a spray voltage of 1.4 kV. The solution flow rate was estimated to be 200 nl min^{-1}. A MS scan was acquired over a mass range of m/z 200–2000. The aminoacylation reaction product was further confirmed *via* the phosphopantetheine (PPant) ejection assay,[41,42] by inducing a collision-induced decomposition with a normalized collision energy of 15%. Reactions (50 µl, 30 °C) to reconstitute the complete pathway from L-proline to pyrrolyl-2-carboxyl-*S*-Leu7 contained 50 mM Tris-HCl (pH 8.0), 10 mM $MgCl_2$, 5 mM L-proline, 5 mM ATP, 100 µM FAD, 2.5 mM TCEP, 0.2 µM Leu5, 10 µM Leu6, and 20 µM *holo* Leu7. The reactions were incubated for 60, 90, and 120 min and then analyzed by high resolution nanospray ESI-FT-Orbitrap-MS/MS analysis, coupled with the Ppant ejection assay.

Isolation, purification and structural elucidation of the C22-deshydroxy leupyrrins A_1, A_2, B_1 and B_2

Mutant *leu22*$^-$ was cultivated with 1% XAD-16 absorber resin as described previously (*2.5*). Isolation and purification of the C22-deshydroxy leupyrrins A_1, A_2, B_1 and B_2 was carried out according to the procedure previously described for the native leupyrrins.[16] Briefly, wet cells and XAD-16 were extracted with MeOH (4 × 200 ml). After removal of the MeOH by evaporation, the resulting aqueous layer was adjusted to pH 10 with aqueous NH_3. Extraction with diethyl ether (4 × 100 ml) and evaporation of the solvent yielded an oily residue, which was dissolved in 3 ml MeOH and extracted with *n*-heptane (4 × 4 ml). Evaporation of the remaining MeOH solution resulted in an orange solid which was subjected to sequential preparative RP-HPLC (yielding 3–8 mg of each purified C22-deshydroxyl leupyrrin). NMR spectra were recorded at 300 K on Bruker AMX400 and ADVANCE DMX 600 NMR spectrometers with CD_3OD as the solvent and internal standard. The structures of the C22-deshydroxy leupyrrins were deduced from detailed 1D (^1H and ^{13}C) and 2D (COSY, HMQC and HMBC) NMR spectral data, and comparison with the data obtained on the respective native, hydroxylated leupyrrins.[16]

Abbreviations

PK	polyketides
NRP	nonribosomal polypeptides
PKS	polyketide synthases
NRPS	nonribosomal polypeptide synthetases
KS	ketosynthase
AT	acyl transferase
ACP	acyl carrier protein
KR	ketoreductase
DH	dehydratase
ER	enoyl reductase
MT	methyl transferase
C	condensation domain
HC	condensation/heterocyclization domain
A	adenylation domain
PCP	peptidyl carrier protein
E	epimerization domain
CP	carrier protein

Acknowledgements

We thank Eva Luxenburger for assistance with the HPLC-MS measurements, and Carsten Kegler for assistance in the analysis of the *leu13*/*leu14* intergenic region. Work in R.M.'s laboratory was funded by the Deutsche Forschungsgemeinschaft (DFG) and the Bundesministerium für Bildung und Forschung (BMBF).

References

1 S. C. Wenzel and R. Müller, *Curr. Opin. Drug Discovery Dev.*, 2009, **12**, 220–230.
2 S. C. Wenzel and R. Müller, *Mol. BioSyst.*, 2009, **5**, 567–574.
3 K. J. Weissman and R. Müller, *Bioorg. Med. Chem.*, 2009, **17**, 2121–2136.
4 M. A. Fischbach and C. T. Walsh, *Chem. Rev.*, 2006, **106**, 3468–3496.
5 C. Hertweck, *Angew. Chem., Int. Ed.*, 2009, **48**, 4688–4716.
6 M. A. Marahiel and L. O. Essen, *Methods Enzymol.*, 2009, **458**, 337–351.
7 C. T. Walsh, H. W. Chen, T. A. Keating, B. K. Hubbard, H. C. Losey, L. S. Luo, C. G. Marshall, D. A. Miller and H. M. Patel, *Curr. Opin. Chem. Biol.*, 2001, **5**, 525–534.
8 L. Du and L. Lou, *Nat. Prod. Rep.*, 2010, **27**, 255–278.
9 S. C. Wenzel and R. Müller, *Nat. Prod. Rep.*, 2007, **24**, 1211–1224.
10 W. Trowitzsch, G. Reifenstahl, V. Wray and G. Höfle, *J. Antibiot.*, 1980, **33**, 1480–1490.
11 B. Böhlendorf, M. Herrmann, H. J. Hecht, F. Sasse, E. Forche, B. Kunze, H. Reichenbach and G. Höfle, *Eur. J. Org. Chem.*, 1999, 2601–2608.
12 F. Sasse, B. Böhlendorf, M. Herrmann, B. Kunze, E. Forche, H. Steinmetz, G. Höfle and H. Reichenbach, *J. Antibiot.*, 1999, **52**, 721–729.
13 G. Höfle, H. Steinmetz, K. Gerth and H. Reichenbach, *Liebigs Ann. Chem.*, 1991, 941–945.
14 S. M. Ringel, R. C. Greenough, S. Romm, D. Conor, A. L. Gutt, B. Blair, G. Kanter and M. von Strandtmann, *J. Antibiot.*, 1977, **30**, 371–375.
15 R. Jansen, B. Kunze, H. Reichenbach and G. Höfle, *Eur. J. Org. Chem.*, 2002, 917–921.
16 H. B. Bode, H. Irschik, S. C. Wenzel, H. Reichenbach, R. Müller and G. Höfle, *J. Nat. Prod.*, 2003, **66**, 1203–1206.
17 H. B. Bode, S. C. Wenzel, H. Irschik, G. Höfle and R. Müller, *Angew. Chem., Int. Ed.*, 2004, **43**, 4163–4167.
18 R. S. Roy, A. M. Gehring, J. C. Milne, P. J. Belshaw and C. T. Walsh, *Nat. Prod. Rep.*, 1999, **16**, 249–263.
19 Y. Q. Cheng, G. L. Tang and B. Shen, *J. Bacteriol.*, 2002, **184**, 7013–7024.
20 S. Rachid, D. Krug, I. Kochems, B. Kunze, M. Scharfe, H. Blöcker, M. Zabriski and R. Müller, *Chem. Biol.*, 2006, **14**, 667–681.

21 A. Heredia-Tapia, B. O. Arredondo-Vega, E. J. Nunez-Vazquez, T. Yasumoto, M. Yasuda and J. L. Ochoa, *Toxicon*, 2002, **40**, 1121–1127.
22 B. S. Goldman, W. C. Nierman, D. Kaiser, S. C. Slater, A. S. Durkin, J. Eisen, C. M. Ronning, W. B. Barbazuk, M. Blanchard, C. Field, C. Halling, G. Hinkle, O. Iartchuk, H. S. Kim, C. Mackenzie, R. Madupu, N. Miller, A. Shvartsbeyn, S. A. Sullivan, M. Vaudin, R. Wiegand and H. B. Kaplan, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 15200–15205.
23 S. Schneiker, O. Perlova, O. Kaiser, K. Gerth, A. Alici, M. O. Altmeyer, D. Bartels, T. Bekel, S. Beyer, E. Bode, H. B. Bode, C. J. Bolten, J. V. Choudhuri, S. Doss, Y. A. Elnakady, B. Frank, L. Gaigalat, A. Goesmann, C. Groeger, F. Gross, L. Jelsbak, J. Kalinowski, C. Kegler, T. Knauber, S. Konietzny, M. Kopp, L. Krause, D. Krug, B. Linke, T. Mahmud, R. Martinez-Arias, A. C. McHardy, M. Merai, F. Meyer, S. Mormann, J. Munoz-Dorado, J. Perez, S. Pradella, S. Rachid, G. Raddatz, F. Rosenau, C. Rückert, F. Sasse, M. Scharfe, S. C. Schuster, G. Suen, A. Treuner-Lange, G. J. Velicer, F. J. Vorhölter, K. J. Weissman, R. D. Welch, S. C. Wenzel, D. E. Whitworth, S. Wilhelm, C. Wittmann, H. Blöcker, A. Pühler and R. Müller, *Nat. Biotechnol.*, 2007, **25**, 1281–1289.
24 J. Ishikawa and K. Hotta, *FEMS Microbiol. Lett.*, 1999, **174**, 251–253.
25 S. F. Altschul, T. L. Madden, A. A. Schaffer, J. H. Zhang, Z. Zhang, W. Miller and D. J. Lipman, *Nucleic Acids Res.*, 1997, **25**, 3389–3402.
26 C. D. Richter, D. Nietlispach, R. W. Broadhurst and K. J. Weissman, *Nat. Chem. Biol.*, 2007, **4**, 75–81.
27 R. W. Broadhurst, D. Nietlispach, M. P. Wheatcroft, P. F. Leadlay and K. J. Weissman, *Chem. Biol.*, 2003, **10**, 723–731.
28 H. Irschik, M. Kopp, K. J. Weissman, K. Buntin, J. Piel and R. Müller, *ChemBioChem*, 2010, **11**, 1840–1849.
29 M. G. Thomas, M. D. Burkart and C. T. Walsh, *Chem. Biol.*, 2002, **9**, 171–184.
30 S. Garneau-Tsodikova, A. Stapon, D. Kahne and C. T. Walsh, *Biochemistry*, 2006, **45**, 8568–8578.
31 A. Mejean, S. Mann, G. Vassiliadis, B. Lombard, D. Loew and O. Ploux, *Biochemistry*, 2010, **49**, 103–113.
32 A. Mejean, S. Mann, T. Maldiney, G. Vassiliadis, O. Lequin and O. Ploux, *J. Am. Chem. Soc.*, 2009, **131**, 7512–7513.
33 C. T. Walsh, S. Garneau-Tsodikova and A. R. Howard-Jones, *Nat. Prod. Rep.*, 2006, **23**, 517–531.
34 K. E. Roege and W. L. Kelly, *Org. Lett.*, 2009, **11**, 297–300.
35 T. Stachelhaus, H. D. Mootz and M. A. Marahiel, *Chem. Biol.*, 1999, **6**, 493–505.
36 G. L. Challis, J. Ravel and C. A. Townsend, *Chem. Biol.*, 2000, **7**, 211–224.
37 C. Li, K. E. Roege and W. L. Kelly, *ChemBioChem*, 2009, **10**, 1064–1072.
38 R. H. Lambalot, A. M. Gehring, R. S. Flugel, P. Zuber, M. LaCelle, M. A. Marahiel, R. Reid, C. Khosla and C. T. Walsh, *Chem. Biol.*, 1996, **3**, 923–936.
39 M. R. Mofid, R. Finking and M. A. Marahiel, *J. Biol. Chem.*, 2002, **277**, 17023–17031.
40 N. Gaitatzis, A. Hans, R. Müller and S. Beyer, *J. Biochem. (Tokyo)*, 2001, **129**, 119–124.
41 D. Meluzzi, W. H. Zheng, M. Hensler, V. Nizet and P. C. Dorrestein, *Bioorg. Med. Chem. Lett.*, 2007, **18**, 3107–3111.
42 P. C. Dorrestein, S. B. Bumpus, C. T. Calderone, S. Garneau-Tsodikova, Z. D. Aron, P. D. Straight, R. Kolter, C. T. Walsh and N. L. Kelleher, *Biochemistry*, 2006, **45**, 12756–12766.
43 T. Stachelhaus and M. A. Marahiel, *J. Biol. Chem.*, 1995, **270**, 6163–6169.
44 Y. Li, K. J. Weissman and R. Müller, *ChemBioChem*, 2010, **11**, 1069–1075.
45 S. Takayama, G. J. McGarvey and C. H. Wong, *Annu. Rev. Microbiol.*, 1997, **51**, 285–310.
46 T. J. Erb, V. Brecht, G. Fuchs, M. Müller and B. E. Alber, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 8871–8876.
47 K. Buntin, H. Irschik, K. J. Weissman, E. Luxenburger, H. Blöcker and R. Müller, *Chem. Biol.*, 2010, **17**, 342–356.
48 S. F. Haydock, J. F. Aparicio, I. Molnar, T. Schwecke, A. König, A. F. A. Marsden, I. S. Galloway, J. Staunton and P. F. Leadley, *FEBS Lett.*, 1995, **374**, 246–248.
49 S. F. Haydock, A. N. Appleyard, T. Mironenko, J. Lester, N. Scott and P. F. Leadley, *Microbiology (Reading, U. K.)*, 2005, **151**, 3161–3169.
50 G. Yadav, R. S. Gokhale and D. Mohanty, *J. Mol. Biol.*, 2003, **328**, 335–363.
51 D. P. Kloer and G. E. Schulz, *Cell. Mol. Life Sci.*, 2006, **63**, 2291–2303.
52 M. L. Heathcote, J. Staunton and P. F. Leadley, *Chem. Biol.*, 2001, **8**, 207–220.
53 D. Schwarzer, H. D. Mootz, U. Linne and M. A. Marahiel, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 14083–14088.
54 E. Yeh, R. M. Kohli, S. D. Bruner and C. T. Walsh, *ChemBioChem*, 2004, **5**, 1290–1293.
55 T. M. Kutchan and H. Dittrich, *J. Biol. Chem.*, 1995, **270**, 24475–24481.
56 J. A. Bjorklund, T. Frenzel, M. Rueffer, M. Kobyashi, U. Mocek, C. Fox, J. M. Beale, S. Gröger, M. H. Zenk and H. G. Floss, *J. Am. Chem. Soc.*, 1995, **117**, 1533–1545.
57 H. Ogura, C. Nishida, U. Hoch, R. Perera, J. H. Dawson and P. R. Ortiz de Montellano, *Biochemistry*, 2004, **43**, 14712–14721.
58 M. V. Mendes, N. Anton, J. F. Martin and J. F. Aparicio, *Biochem. J.*, 2005, **386**, 57–62.
59 M. Inouye, Y. Takada, N. Muto, T. Beppu and S. Horinouchi, *Mol. Gen. Genet.*, 1994, **245**, 456–464.
60 A. M. Rodriguez, C. Olano, C. Mendez, C. R. Hutchinson and J. A. Salas, *FEMS Microbiol. Lett.*, 1995, **127**, 117–120.
61 J. L. Ramos, M. Martinez-Bueno, A. J. Molina-Henares, W. Teran, K. Watanabe, X. Zhang, M. T. Gallegos, R. Brennan and R. Tobes, *Microbiol. Mol. Biol. Rev.*, 2005, **69**, 326–356.
62 S. C. Wenzel and R. Müller, *Curr. Opin. Chem. Biol.*, 2005, **9**, 447–458.
63 M. A. Fischbach, C. T. Walsh and J. Clardy, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 4601–4608.
64 B. Silakowski, H. U. Schairer, H. Ehret, B. Kunze, S. Weinig, G. Nordsiek, P. Brandt, H. Blöcker, G. Höfle, S. Beyer and R. Müller, *J. Biol. Chem.*, 1999, **274**, 37391–37399.
65 J. Piel, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 14002–14007.
66 J. Piel, D. Butzke, N. Fusetani, D. Hui, M. Platzer, G. Wen and S. Matsunaga, *J. Nat. Prod.*, 2005, **68**, 472–479.
67 M. Kopp, H. Irschik, F. Gross, O. Perlova, A. Sandmann, K. Gerth and R. Müller, *J. Biotechnol.*, 2004, **107**, 29–40.
68 B. Neumann, A. Pospiech and H. U. Schairer, *Trends Genet.*, 1992, **8**, 332–333.
69 J. Sambrook, E. F. Fritsch and T. Maniatis, *Molecular cloning: A laboratory manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.
70 J. D. Thompson, D. G. Higgins and T. J. Gibson, *Nucleic Acids Res.*, 1994, **22**, 4673–4680.
71 A. de Marco, E. Deuerling, A. Mogk, T. Tomoyasu and B. Bukau, *BMC Biotechnol.*, 2007, **7**, 32–40.
72 K. Nishihara, M. Kanemori, M. Kitagawa, H. Yanagi and T. Yura, *Appl. Environ. Microbiol.*, 1998, **64**, 1694–1699.
73 M. M. Bradford, *Anal. Biochem.*, 1976, **72**, 248–254.