Electronic Supplementary Material (ESI) for Chemical Science.

Supporting Information

Thioester reduction and aldehyde transamination are universal steps in actinobacterial polyketide alkaloid biosynthesis

U. R. Awodi, J. L. Ronan, J. Masschelein, E. L. C. de los Santos and G. L. Challis

Department of Chemistry, University of Warwick, Gibbet Hill, Coventry, CV4 7AL, United Kingdom.

PKS chain release mechanisms

A wide variety of chain release mechanisms are employed by modular PKSs (Figure S1).

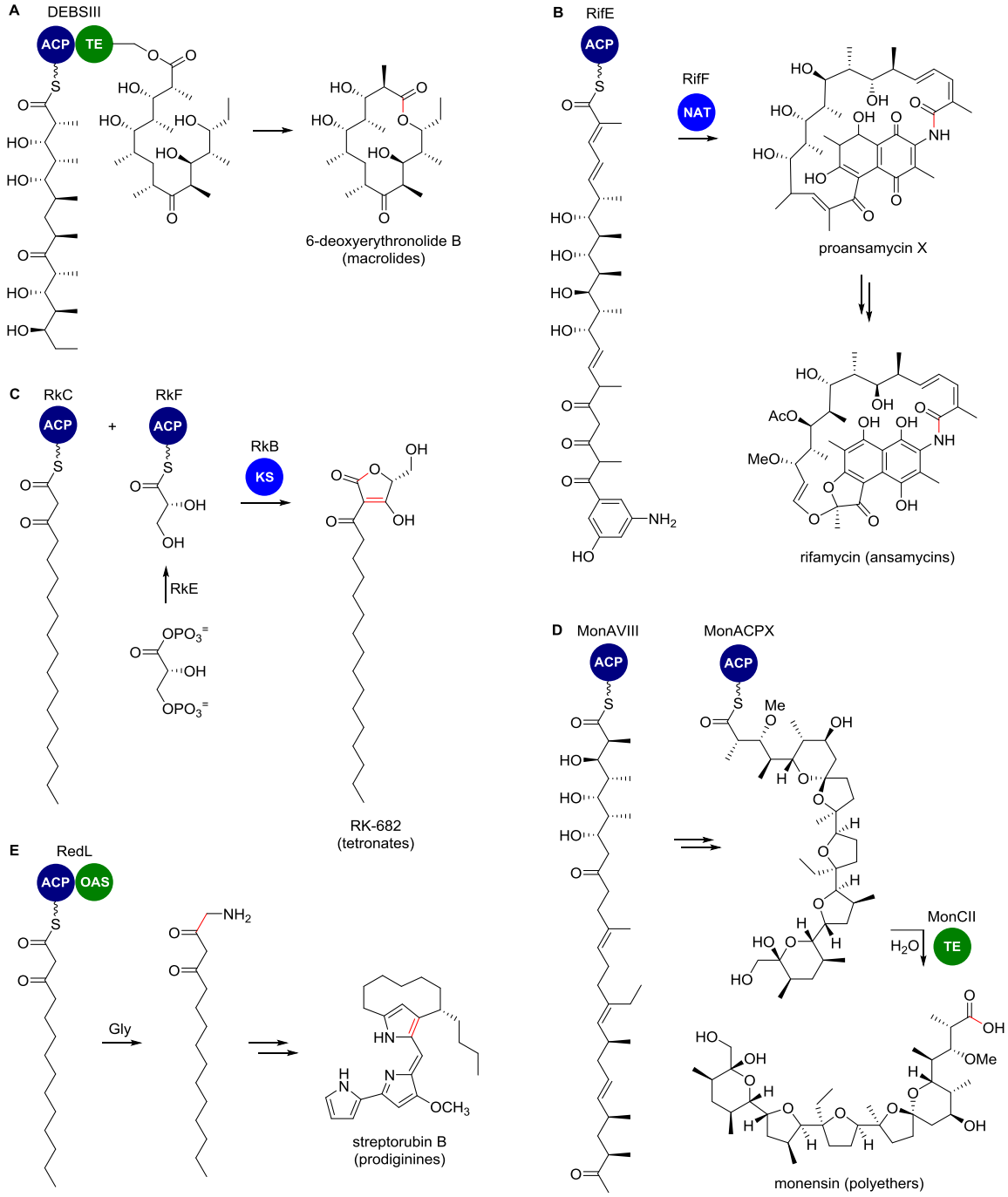

Figure S1 Chain release mechanisms employed by type I modular PKSs for the assembly of various metabolite classes (note modular PKS domains other than the final acyl carrier protein (ACP) domain and the domain responsible for chain release, where relevant, have been omitted for clarity). (A) Macrolide antibiotics are typically assembled via thioesterase (TE) domain-mediated macrolactonisation, as exemplified by 6-deoxyerythronolide B biosynthesis. (B) The assembly of ansamycins such as rifamycin involves macrolactamization catalyzed by a standalone N-acyl transferase (NAT). (C) Tetronates, such as RK-682, are assembled via condensation of the polyketide chain with a glyceryl-ACP thioester catalyzed by a standalone ketosynthase (KS)-like domain. (D) The polyketide chain in the biosynthesis of monensin and other polyethers is transferred to a dedicated ACP where a series of tailoring reactions occur. A standalone TE then releases the mature antibiotic via hydrolysis. (E) An α-oxamine synthase (OAS) domain is responsible for chain release via decarboxylative condensation with glycine in the biosynthesis of streptorubin B and other prodiginine alkaloids.

Cloning and overexpression of *cpkG* and the region of *cpkC* encoding the TR domain in *E. coli*

cpkG (1561 bp) and the region of *cpkC* encoding the TR domain (1158 bp) were amplified by PCR using the primers listed in Table S1. A CACC sequence was introduced at the 5'-end of the forward primers to allow for directional cloning of the blunt-ended PCR products into pET151/D-TOPO®. The PCR reaction was performed with Readymix *Taq* PCR mix (Sigma Aldrich) according to the manufacturer's instructions. Reaction conditions consisted of an initial denaturation step at 94 °C for 2 min, followed by 30 cycles of 94 °C for 45 s, 60 °C for 45 s and 72 °C for 90 s. PCR products were separated on a 1 % agarose gel containing 0.05 µL/ml GelRed Nucleic Acid Stain (Cambridge Bioscience UK). Bands of the appropriate size were excised from the gel and purified using the GeneJET Gel Extraction kit (Thermo Fisher Scientific). The purified PCR products were ligated with the linearized expression vector using the Champion pET151 Directional TOPO Expression kit (Invitrogen) according to the manufacturer's guidelines. The ligation mixtures were used to transform One Shot TOP10 chemically competent *E. coli* cells. Transformants were selected on LB agar plates supplemented with ampicillin (100 µg/mL). Plasmids were purified from ampicillin-resistant colonies using the GeneJET Plasmid Miniprep kit (Thermo Fisher Scientific) and the sequences of the cloned genes were confirmed by Sanger sequencing (GATC Biotech). One correct clone was used to transform BL21 Star(DE3) chemically competent *E. coli* cells for expression of *cpkG* and the TR domain of *cpkC* as *N*-terminal His_6-tagged fusion proteins. For overexpression of each gene, *E. coli* BL21 Star(DE3)/pET151-*cpkG* and *E. coli* BL21 Star(DE3)/pET151-*cpkC-TR* cells were cultured in LB medium (1 L) supplemented with ampicillin (100 µg/ml) at 37 °C with shaking at 180 rpm. Incubation was continued until the optical density of the cultures at 600 nm reached 0.5-0.6, at which time isopropyl-β-*D*-thiogalactopyranoside was added to a final concentration of 0.5 mM to induce expression. The cell cultures were then incubated overnight at 15 °C with shaking at 180 rpm.

Table S1 Primers used for PCR amplification of *cpkG* and the region of *cpkC* encoding the TR domain

Primer name	Sequence (5'-3')
*cpkC*_TR_fw	CACCTTCGCCGCCGAGGTC
*cpkC*_TR_rev	TCACCCGGCCGCCGGGAAG
*cpkG*_fw	CACCATGACCGACCAGCC
*cpkG*_rev	TCAGCCGATCGGAGCCA

Purification and analysis of His_6-CpkG and His_6-CpkC-TR

For the purification of His_6-CpkG and His_6-CpkC-TR, the appropriate *E. coli* cells were harvested by centrifugation (1509 g, 4 °C, 15 min) and cell pellets were re-suspended in 10 mL washing buffer (50

mM TRIS-HCl (pH 8.0), 200 mM NaCl, 10 mM imidazole, 10 % glycerol). During His_6-CpkG purification, all buffers were supplemented with 20 µM pyridoxal 5'-phosphate (PLP). Phenylmethanesulfonyl fluoride (PMSF) (100 µL) was added and the cells were lysed with a high pressure cell disruptor (Constant Systems Ltd TS-Series Cabinet) at 20,000 psi. After removal of the cellular debris by centrifugation (7800 g, 4 °C, 30 min), the cell-free lysate was applied to a 1 mL HiTrap HP affinity column (GE Healthcare) and purified using an ÄKTA Purifier (GE Healthcare). Unbound and weakly-bound proteins were removed by washing the column with 20 mL washing buffer. The recombinant fusion proteins were eluted with 3 mL elution buffer (50 mM TRIS-HCl (pH 8.0), 200 mM NaCl, 200 mM imidazole, 10 % glycerol) at a flow rate of 0.45 mL/min. The absorbance was monitored at 280 nm and fractions were analyzed by SDS-PAGE (Figure S2). Those containing the His_6-tagged proteins were pooled, concentrated and exchanged into storage buffer (CpkC-TR: 50 mM TRIS-HCl (pH 8.0), 200 mM NaCl, 10 % glycerol; CpkG: 50 mM HEPES (pH 7.8), 100 mM NaCl, 20 µM PLP, 10 % glycerol) using Amicon Ultra centrifugal filters with a 30 kDa molecular weight cut-off membrane (Millipore). The proteins were aliquoted, flash frozen in liquid N_2 and stored at -80 °C. The native oligomerization state of His_6-CpkG was determined by gel filtration analysis using an ÄKTA Purifier fitted with a HiLoad 75 Superdex 12 prep grade column, which had previously been calibrated with molecular weight markers between 12,000-20,000 Da. The elution volume (*ca*. 55 mL) indicated that the protein was dimeric (Figure S3). Protein concentrations were determined by measuring A_{280} on a NanoDrop spectrophotometer (Thermo Scientific), using calculated extinction coefficients of 54,890 L mol^{-1} cm^{-1} (CpkC-TR) and 20,860 L mol^{-1} cm^{-1} (CpkG). UHPLC-ESI-Q-TOF-MS using a Dionex UltiMate 3000 RS UHPLC connected to an ACE 3 C4-300 reverse phase column (Advanced Chromatography Technologies, Aberdeen, UK; 100 × 2.1 mm) coupled to a Bruker MaXis IMPACT mass spectrometer was used to determine the molecular weights of the purified proteins (Figure S4). The elution profile is given in Table S2. Sodium formate (10 mM) was used for internal calibration. The mass spectrometer was operated in positive ion mode with a scan range of 50–3000 *m/z*.

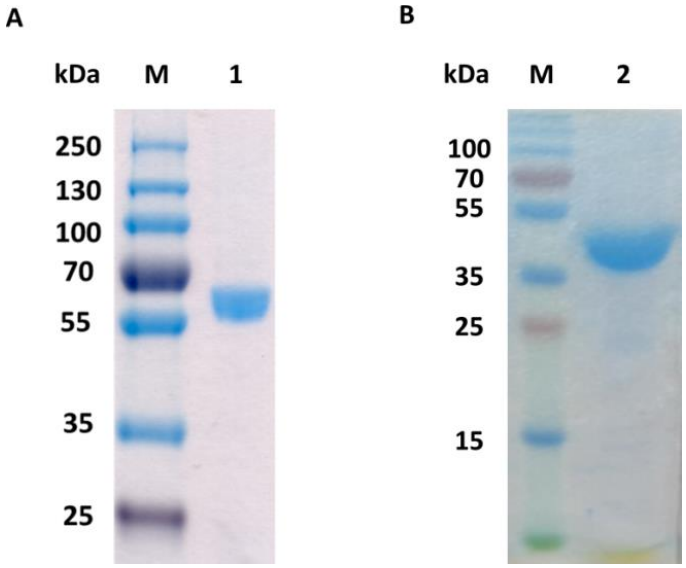

Figure S2 SDS-PAGE analysis of purified His_6-CpkG (A) and His_6-CpkC-TR (B). His_6-CpkG (lane 1) and His_6-CpkC-TR (lane 2) were analysed on a 12 % polyacrylamide gel. Lane M contains proteins of known molecular weight (PageRuler Plus Prestained Protein Ladder, Thermo Scientific).

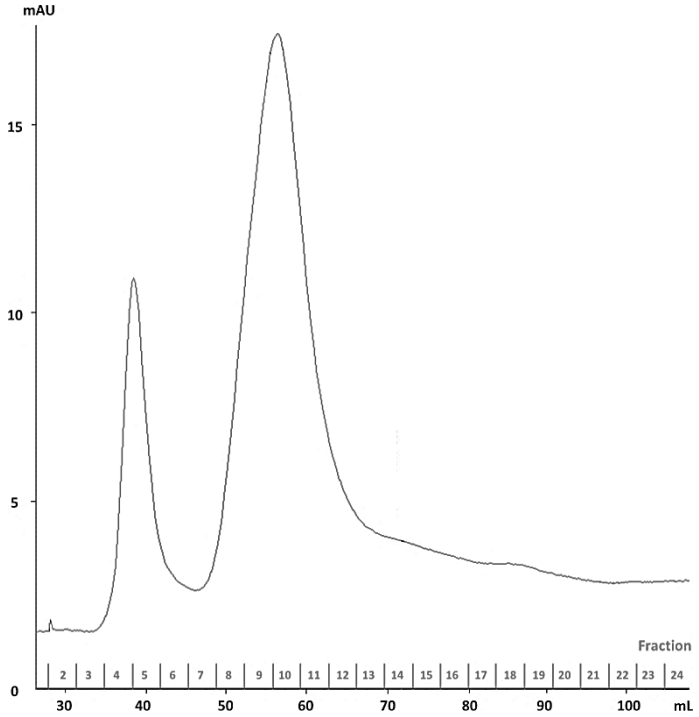

Figure S3 Chromatogram from gel filtration analysis of purified His_6-CpkG. UV absorbance at 280 nm was monitored.

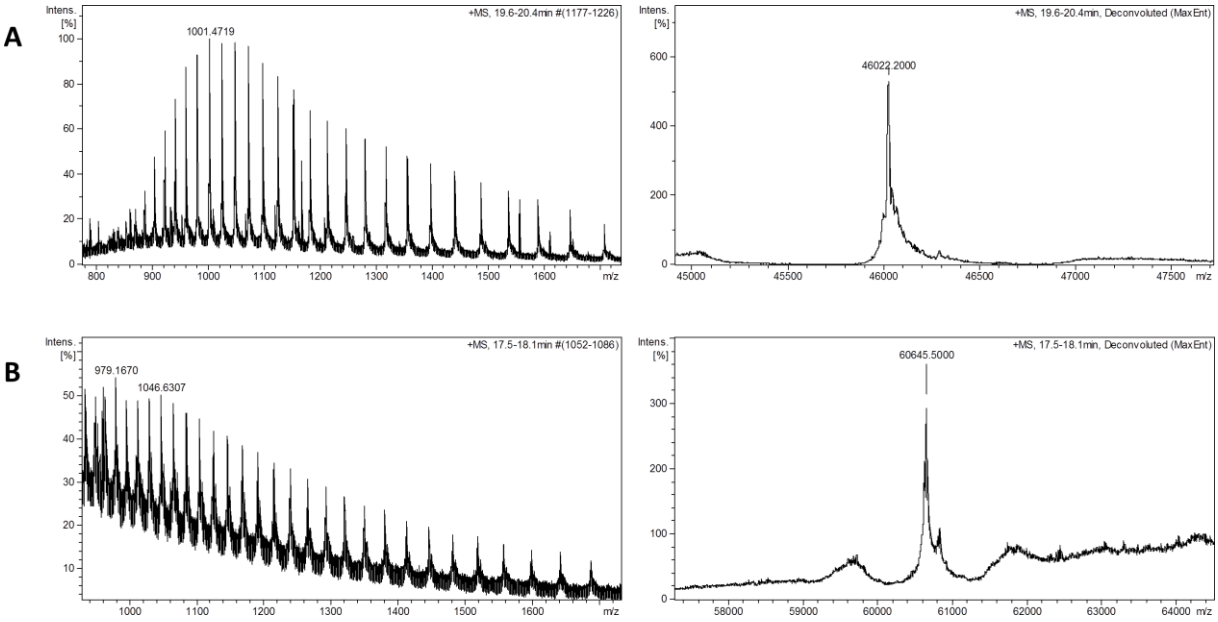

Figure S4 Measured (left) and deconvoluted (right) mass spectra of His_6-CpkC-TR (A) and His_6-CpkG (B). The calculated molecular weights of His_6-CpkC-TR and His_6-CpkG are 46022 Da and 60646 Da, respectively.

Table S2 Elution conditions used for LC-MS analyses of purified recombinant proteins

Time (min)	Water / 0.1 % formic acid (%)	Methanol / 0.1 % formic acid (%)	Flow rate (mL/min)
0.0	90	10	0.1
8.0	90	10	0.1
47.5	0	100	0.1
52.5	0	100	0.1
57.5	90	10	0.1

Analysis of NAD(P)H consumption by CpkC-TR

The consumption of NAD(P)H by His_6-CpkC-TR was examined as described by Chhabra *et al.*[1] Octanoyl-CoA (200 µM) was incubated with NAD(P)H (200 µM) and buffer (10 mM TRIS-HCl (pH 8.0), 50 mM NaCl, 5 % glycerol) in a total volume of 214 µL. Reactions were initiated by addition of 36 µL of purified His_6-CpkC-TR (425 µM) and the absorbance at 340 nm was continuously monitored for 10 min at 25 °C using a Varian Cary® 50 Bio UV-Visible Spectrophotometer. Control reactions were performed from which the enzyme was omitted. The data shown in Figure 3 are from a single experiment but the same results were obtained in three independent experiments.

Analysis of CpkC-TR-catalyzed reduction of octanoyl-CoA and octanal to octanol

The conversion of octanal to octanol (Figure S5) was examined by incubating purified His_6-CpkC-TR (100 µM) with NADH (20 mM) and octanal (20 mM) in NH_4HCO_3 buffer (25 mM), in a total volume of 500 µL. The conversion of octanoyl-CoA to octanol was investigated similarly except that His_6-CpkC-TR (125 µM) and NADH (40 mM) were used. Following incubation at room temperature for 24 h, equal volumes of $CHCl_3$ were added to the reaction mixtures to precipitate the protein and extract the reaction products. The $CHCl_3$ extracts were subsequently dried with $MgSO_4$ and filtered through cotton wool. One drop of *N,O*-Bis-(trimethylsilyl)acetamide (BSA) was added and the samples were analyzed by GC-MS (Varian 400). 5 µL of each sample was injected using a CP8400 sample injector in splitless mode and passed through a VF-5ms column (30 m length, 0.25 mm internal diameter) using helium as the carrier gas. The temperature was held at 50 °C for 1 min before being ramped up to 300 °C at a rate of 25 °C/min. The temperature was held at 300 °C for 9 min before being ramped back down to 50 °C. The sample was ionized using electron impact with 70 eV electrons and detected with an ion trap mass analyzer. Two independent assays were performed and a reaction containing boiled His_6-CpkC-TR was used as a negative control.

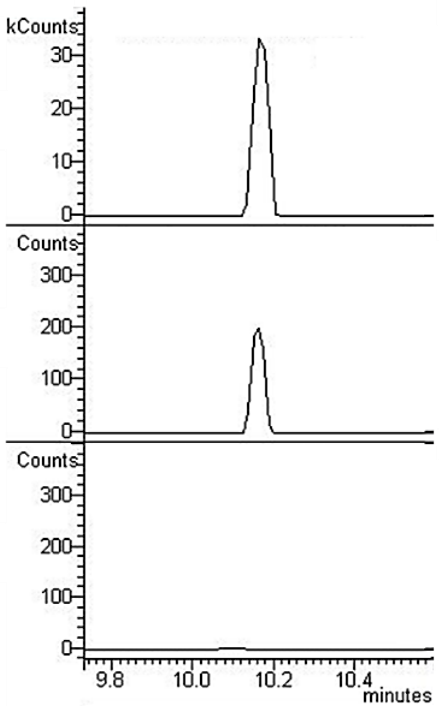

Figure S5 Extracted ion chromatograms at *m/z* = 187.2 (corresponding to the fragment ion of the BSA derivative of octanol resulting from neutral loss of a methyl group) from GC-MS analyses of the CpkC-TR-catalyzed reduction of or octanal with NADH. The top chromatogram is for the authentic standard of BSA-derivatized octanol. The middle chromatogram corresponds to the BSA-derivatized product of the enzymatic reaction, while the bottom chromatogram is from BSA-derivatization of the negative control reaction containing heat-denatured enzyme.

UV-Vis spectroscopic analysis of PLP binding to CpkG

UV/Vis absorbance spectra were obtained for His_6-CpkG (60 µM) and PLP (60 µM) independently, using a Perkin Elmer Lambda 35 UV/Vis Spectrometer (Figure S6). Prior to measurements, His_6-CpkG was exchanged from HEPES storage buffer (50 mM HEPES (pH 7.8), 100 mM NaCl, 20 µM PLP, 10 % glycerol) into TRIS buffer (20 mM TRIS-HCl, (pH 8.0), 100 mM NaCl) to remove unbound PLP.

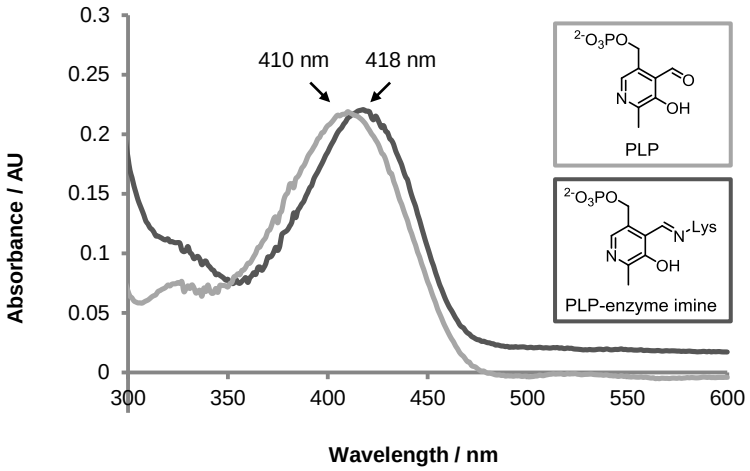

Figure S6 UV-Vis spectra of PLP (light grey) and the aldimine adduct of PLP with CpkG (dark grey). The λ_{max} values for the absorbances observed in each case are indicated. The shift in λ_{max} from 410 nm to 418 nm indicates that PLP has formed an internal aldimine adduct with His_6-CpkG.[2] The near-identical absorbance intensities at λ_{max} are consistent with a 1:1 molar binding ratio between CpkG and PLP.

Spectrophotometric analysis of the substrate specificity of CpkG

The enzymatic activity and substrate specificity of CpkG were characterized by employing a spectrophotometric assay that relies on the detection of a blue α-amino acid-copper (II) complex.[3] His_6-CpkG (20 µM) was combined in separate reactions with 10 mM amino donor (octylamine, butylamine, isobutylamine, cyclohexylamine or benzylamine) and 10 mM amino acceptor (pyruvate or α-ketoglutarate) in 100 mM potassium phosphate buffer (pH 7) in a total volume of 160 µL. Reactions were carried out in 96-well plates and incubated overnight at room temperature. Following overnight incubation, 40 µL of a $CuSO_4$/methanol solution was added to each reaction and the resulting mixtures were filtered by centrifugation (15871 g, 1 min) using PVDF micro spin filters (Grace Davison Discovery Sciences). The $CuSO_4$/methanol solution was prepared as described by Hwang and Kim.[2] Supernatants were subsequently analyzed by measuring the absorbance at 595 nm using a Tecan GENios Microplate Reader. Reactions were performed in triplicate. Reaction mixtures incubated without enzyme served as negative controls. Blank samples consisted of 100 mM potassium phosphate buffer (pH 7.0) with and without 20 µM His_6-CpkG. The results were expressed

as relative specificities calculated by first subtracting the mean absorbance value of the blank samples from each measured absorbance value, to obtain background-corrected values. Subsequently, the mean absorbance values of the test samples were divided by the mean absorbance value of the control samples. Relative specificities were then determined by comparison to the specificity of CpkG for octylamine and pyruvate as amino donor and acceptor (Figure S7).

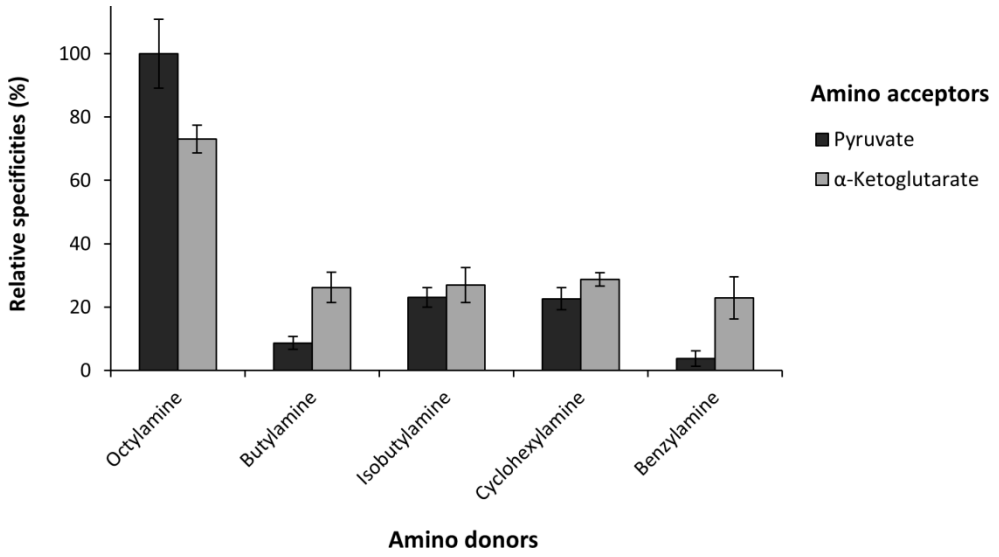

Figure S7 Amino donor and acceptor specificities of CpkG as determined by the $CuSO_4$/MeOH staining method. The specificities for the different substrate combinations are shown relative to the specificity of CpkG for octylamine and pyruvate as amino donor and acceptor. Data are represented as means ± 1 standard deviation from triplicate experiments.

Determination of alanine absolute stereochemistry using Marfey's method

The absolute stereochemistry of alanine, formed by the CpkG-catalyzed transamination of pyruvate with octylamine, was determined by derivatization with Marfey's reagent,[4] followed by UHPLC-ESI-Q-TOF-MS analysis using a Dionex UltiMate 3000 RS UHPLC connected to a Zorbax Eclipse Plus column (C18, 100 × 2.1 mm, 1.8 µm) coupled to a Bruker MaXis IMPACT mass spectrometer. His_6-CpkG (20 µM) was combined with octylamine (10 mM), pyruvate (10 mM) and phosphate buffer (100 mM, pH 7.0) in a final volume of 80 µL. After incubation for 2 h at room temperature, proteins were removed by centrifugation (13523 g, 15 min) using a Vivaspin 500 centrifugal filter with a 5 kDa MWCO (Sartonius Stedim Biotech). Acetone (65 µL), $NaHCO_3$ (1 M, 7.5 µL) and 1 % Marfey's reagent in acetone (17 µL) were then added to 40 µL of the filtrate and the resulting mixture was incubated at 37 °C for 1.5 h. Authentic standards of the Marfey's derivatives of L- and D-alanine were prepared similarly by adding acetone (65 µL), 1 M $NaHCO_3$ (7.5 µL) and 1 % Marfey's reagent in acetone (17 µL) to solutions of L- and D-alanine (10 mM) in phosphate buffer (100 mM pH 7.0) in a final volume of 40 µL. These mixtures were incubated as described above. Reactions were quenched with 1 M HCl (7.5 µL) and the remaining acetone was removed by evaporation. Deionised water (200 µL) and

acetonitrile (200 µL) were then added to each sample. Precipitates were removed by centrifugation (13523 g, 1 min) using PVDF micro spin filters (Grace Davison Discovery Sciences). The samples were subsequently analysed by LC-MS using to the elution profile in Table S3 (Figure S8). Independent experiments were performed in duplicate.

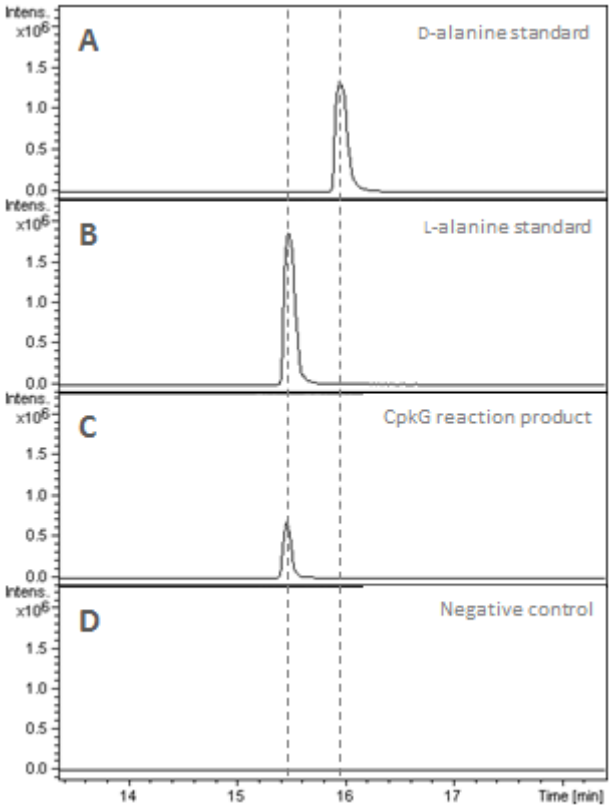

Figure S8 Extracted ion chromatograms at *m/z* = 342.1044 ± 0.005 (corresponding to the $[M+H]^+$ for the Marfey's derivative of alanine) from LC-MS analyses of (A) Marfey's derivative of D-alanine; (B) Marfey's derivative of L-alanine; (C) Marfey's derivative of the products of the CpkG-catalyzed transamination of pyruvate with octylamine; and (D) Marfey's derivatization of the negative control reaction without enzyme.

Table S3 Elution conditions used for LC-MS analyses of small molecules

Time (min)	Water / 0.1 % formic acid (%)	Acetonitrile / 0.1 % formic acid (%)	Flow rate (mL/min)
0.0	95	5	0.2
5.3	95	5	0.2
18.3	0	100	0.2
22.3	0	100	0.2
25.3	95	5	0.2
34.0	95	5	0.2

Detection of octanal

Octanal formation in the CpkG-catalyzed reaction with pyruvate and octylamine was detected by derivatization with D-cysteine and subsequent high-resolution LC-MS analysis.[5] His_6-CpkG (90 µM) was combined with octylamine (200 µM), pyruvate (200 µM) and phosphate buffer (25 mM, pH 7.0) in a final volume of 125 µL. After incubation for 15 min at 37 °C, 400 µM D-cysteine was added and the resulting mixture was incubated for 10 min at 50 °C. Methanol (125 µL) was then added to precipitate the enzyme. Precipitates were removed by centrifugation (13523 g, 1 min) using PVDF micro spin filters (Grace Davison Discovery Sciences) and the supernatants were analyzed using a Dionex UltiMate 3000 RS UHPLC connected to a Zorbax Eclipse Plus column (C18, 100 × 2.1 mm, 1.8 µm) coupled to a Bruker MaXis IMPACT mass spectrometer and the elution profile in Table S3. An authentic octanal standard (200 µM) was derivatized and analyzed similarly. Independent assays were performed in duplicate. The derivatized assay products displayed the same *m/z* and retention time as the derivatized octanal standard, (4*S*)-2-((*Z*)-4-hydroxyhept-2-en-1-yl)thiazolidine-4-carboxylic acid (Figure S9).

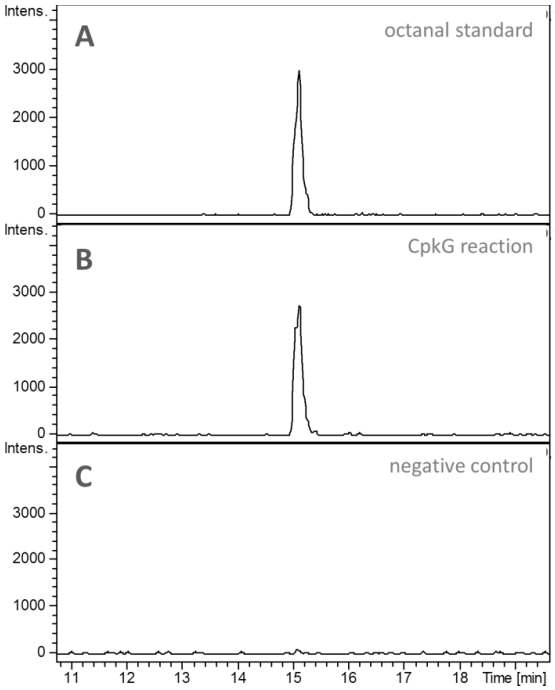

Figure S9 Extracted ion chromatograms at *m/z* = 232.1365 ± 0.005 (corresponding to the $[M+H]^+$ ion of the D-cysteine derivative of octanal) from LC-ESI-MS analyses of the His_6-CpkG-catalyzed reaction with octylamine and pyruvate. (A) Authentic standard of D-cysteine-derivatized octanal; (B) product of the enzymatic reaction after derivatization with D-cysteine; and (C) the negative control reaction from which the enzyme was omitted after derivatization with D-cysteine.

Detection of pyruvate and octylamine

Pyruvate formation in the CpkG-catalyzed reaction with alanine and octanal was analysed with a lactate dehydrogenase (LDH)-coupled assay. LDH catalyzes the reduction of pyruvate to L-lactate by NADH. The progress of this reaction can be followed spectrophotometrically by measuring the decrease in absorbance at 340 nm (due to the depletion of NADH).[6] Reactions contained 90 µM His_6-CpkG, 200 µM octanal, 200 µM L- or D-alanine, 1 µM lactate dehydrogenase (from *Lactobacillus leichmanii*), 200 µM NADH and 25 mM phosphate buffer (pH 7.0) in a total volume of 250 µL. The absorbance at 340 nm was monitored for 15 min at 25 °C using a Varian Cary 50 Bio UV-Visible Spectrophotometer (Figure S10). After 15 min, methanol (450 µL) was added to the reactions to precipitate the enzymes. Precipitates were removed by centrifugation (13523 g, 1 min) using PVDF micro spin filters (Grace Davison Discovery Sciences) and the supernatants were analysed by UHPLC-ESI-Q-TOF-MS using a Dionex UltiMate 3000 RS UHPLC connected to a Zorbax Eclipse Plus column (C18, 100 × 2.1 mm, 1.8 µm) coupled to a Bruker MaXis IMPACT mass spectrometer and the elution profile in Table S3 (Figure S11).

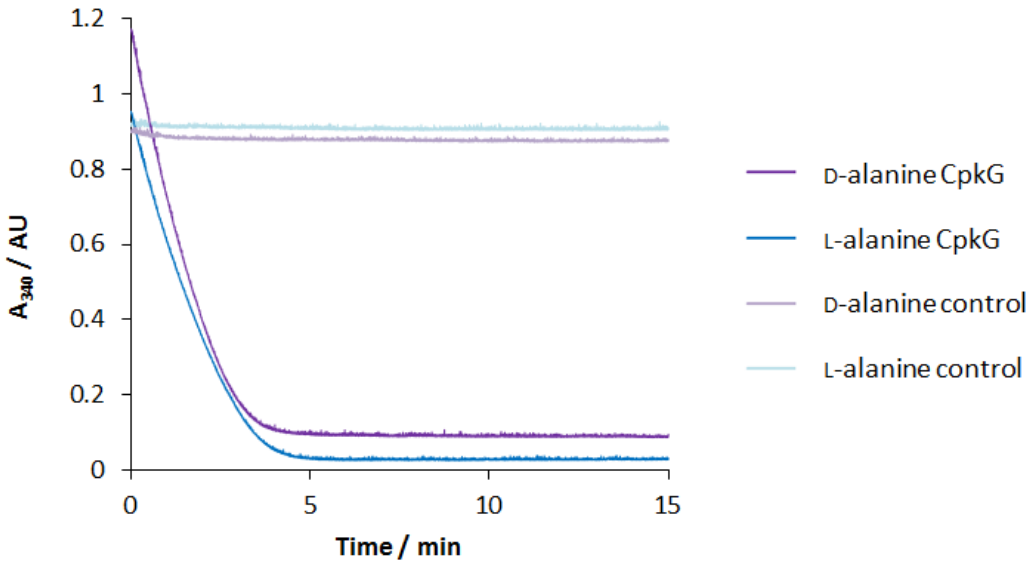

Figure S10 Detection of pyruvate formation by CpkG using a coupled assay with LDH to reduce the pyruvate formed with NADH. The consumption of NADH was determined by monitoring the decrease in absorbance at 340 nm.

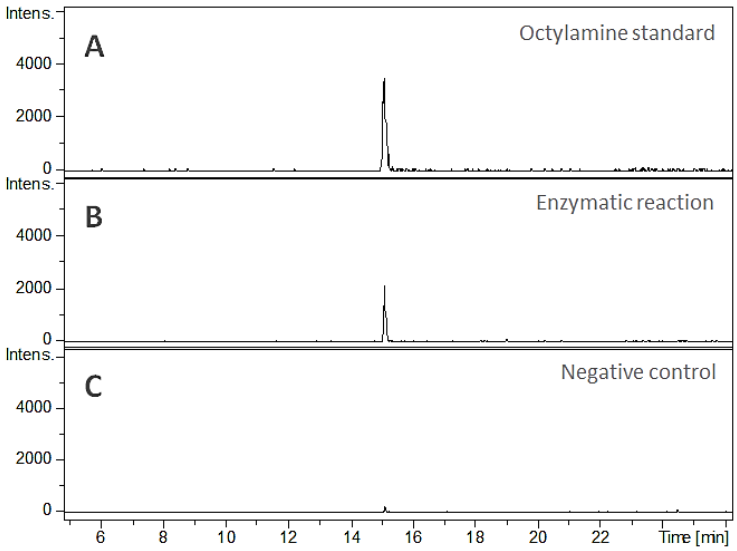

Figure S11 Extracted ion chromatograms at *m/z* = 130.1590 ± 0.005 (corresponding to the $[M+H]^+$ ion for octylamine) from LC-ESI-MS analyses of the His_6-CpkG-catalyzed reaction of octanal with D-alanine. (A) Octylamine authentic standard; (B) enzymatic reaction; and (C) negative control reaction from which the enzyme has been omitted.

Mass spectrometer settings for small molecule analyses

The mass spectrometer was operated in positive ion mode with a scan range of 50–3000 *m/z*. Source conditions were: end plate offset at −500 V; capillary at −4500 V; nebulizer gas (N_2) at 1.6 bar; dry gas (N_2) at 8 L min^{-1}; dry temperature at 180 °C. Ion transfer conditions were: ion funnel RF at 200 Vpp; multiple RF at 200 Vpp; quadrupole low mass at 55 *m/z*; collision energy at 5.0 eV; collision RF at 600 Vpp; ion cooler RF at 50–350 Vpp; transfer time at 121 µs; pre-pulse storage time at 1 µs. Calibration was performed with 10 mM sodium formate through a loop injection of 20 µL at the start of each run.

Bioinformatics analyses

Homology searches were performed using HMMER 3.1b2 (http://hmmer.janelia.org) against a GenBank database compiled with the help of multigeneblast and a database containing bacterial and whole genome shotgun sequences from Refseq.[7] Searches for homologs of *cpkG* that were within 35 kb of coding sequences that contained hits to profile HMMs constructed for ketosynthase (KS) domains and thioester reductase (TR) domains were conducted.[8] The DNA sequences within a 100 kb genomic window centered on the nucleotide midpoint of the three hits (transaminase, KS and TR domain) were extracted and annotated using antiSMASH[9] and BLAST searches (Figure S12, Table 4).

Sequence similarity between putative tailoring enzymes encoded within the cryptic gene clusters and those associated with the streptazone E biosynthetic cluster was investigated using phmmer (http://hmmer.janelia.org/search/phmmer).[10] This allowed us to tentatively link four cryptic gene clusters from our bioinformatics search to the biosynthesis of streptazone E or closely related metabolites.

The structures of polyketide products were predicted by analyzing the domain and module organization of the PKSs. Searches for conserved domains were performed using SBSPKS and the NCBI conserved domain search.[11,12] The substrate specificity of acyltransferase (AT) domains was determined by multiple sequence alignments with AT domains of known specificity.[13] The stereospecificity of ketoreductase (KR) domains was predicted by sequence comparisons with KR domains of known stereospecificity.[14] Dehydratase (DH) domain activity was evaluated by *in silico* analysis of conserved active site residues.[15] The stereospecificity of enoyl reductase (ER) domains was predicted by sequence comparisons to ER domains of known stereospecificity.[16]

The gene clusters were compared to known biosynthetic gene clusters in the MiBIG database.[17] This was done by performing a BLAST[18] query of the protein sequences in the clusters with the proteins in MiBIG. The "distance" between two proteins was calculated by subtracting one from the ratio of the bitscore of the highest scoring alignment to a protein and the highest possible bitscore for the protein, thus the distance between a protein and itself would be 0 and the distance between non hits would be 1. Cluster distances were computed by calculating pairwise distances between coding sequences in the cluster and pairing proteins up to minimize the total distance between the clusters. In the cases where the predicted structures of the polyketide backbone corresponded to products with known biosynthetic gene clusters in the MiBIG database, the known cluster appeared as one of the lowest scoring clusters in the comparison (Figure S13). The bioinformatics analyses were performed with the help of clusterTools (https://github.com/emzodls/clusterTools), a bespoke Python toolkit for analyzing biosynthetic gene clusters, the details of which will be published elsewhere.

Figure S12 Overview of gene clusters harboring PKS genes as well as gene encoding homologs of both CpkC-TR and CpkG. The predicted metabolic product(s) of each of these clusters are listed on the right.

Table S4 Overview of actinobacterial strains harboring gene clusters that contain genes encoding homologs of both *cpkC*-TR and *cpkG*. The GenBank accession numbers of the homologous genes are indicated, as well as the sequence identity of the encoded proteins to CpkC-TR and CpkG. Clusters are ranked according to average sequence identity to CpkC-TR/CpkG.

GenBank Acces. No.	Species	Transaminase Gene Acces. No.	ID to CpkG (%)	TR Gene Acces. No.	ID to CpkC-TR (%)	Cluster Start	Cluster End	Average ID (%)
AL939127	*Streptomyces coelicolor* A3(2)	CAC37880	100	CAC22144	100	7868	72337	100
CP009124	*Streptomyces lividans* TK24	AIJ12365	99.807	AIJ12371	99.583	1504474	1583683	99.695
CP005080	*Streptomyces fulvissimus* DSM 40593	AGK79562	61.437	AGK79554	80.833	5106569	5171035	71.135
CP010519	*Streptomyces albus* DSM 41398	AJE82594	58.679	AJE82585	80.417	2726967	2790890	69.548
NZ_LATD01000059	*Streptomyces* sp. NRRL B-24891	WP_046496905	58.979	WP_046496888	79.583	66	47710	69.281
CP003987	*Streptomyces* sp. 769	AJC52685	57.634	AJC52690	72.5	38073	172009	65.067
GG657757	*Streptomyces viridochromogenes* DSM 40736	EFL33839	58.144	EFL33831	71.488	4852429	4915328	64.816
NZ_BAGC01000109	*Nocardia niigatensis* NBRC 100131	WP_040863151	56.395	WP_040863148	72.5	66635	126073	64.4475
LN881739	*Streptomyces venezuelae* ATCC 15439	CUM36546	57.834	CUM36537	70	974021	1042192	63.917
GG657757	*Streptomyces viridochromogenes* DSM 40736	EFL30368	54.528	EFL30361	69.038	926850	1004427	61.783
LGED01000198	*Saccharothrix* sp. NRRL B-16348	KOX23500	53.62	KOX23546	67.917	72586	139469	60.7685
HE804045	*Saccharothrix espanaensis* DSM 44229	CCH32004	53.686	CCH32002	67.083	5177164	5281786	60.3845
NZ_JOEM01000006	*Streptomyces cyaneofuscatus* NRRL B-2570	WP_032771882	53.947	WP_032771870	66.529	205904	270608	60.238
NZ_KB891309	*Streptomyces* sp. ScaeMP-e10	WP_026241859	53.759	WP_018514182	66.529	713786	776692	60.144
CP002638	*Verrucosispora maris* AB-18-032	AEB44520	55.299	AEB44524	64.167	2712082	2777065	59.733
NZ_JOHZ01000031	*Streptomyces* sp. NRRL F-5193	WP_043226386	53.019	WP_043226388	65.145	10001	77785	59.082
CP001630	*Actinosynnema mirum* DSM 43827	ACU38291	55.212	ACU38286	62.5	5214948	5289106	58.856
NZ_JNYO01000050	*Saccharothrix syringae* NRRL B-16468	WP_033434782	47.451	WP_051766815	69.583	1084	51786	58.517
LN831790	*Streptomyces leeuwenhoekii*	CQR65737	49.049	CQR65742	67.5	7356643	7438425	58.2745
CP001630	*Actinosynnema mirum* DSM 43827	ACU38525	49.808	ACU38528	65	5568526	5626090	57.404
GL877172	*Streptomyces griseus* XylebKG-1	EGE46015	53.131	EGE46010	59.167	7817499	7882757	56.149

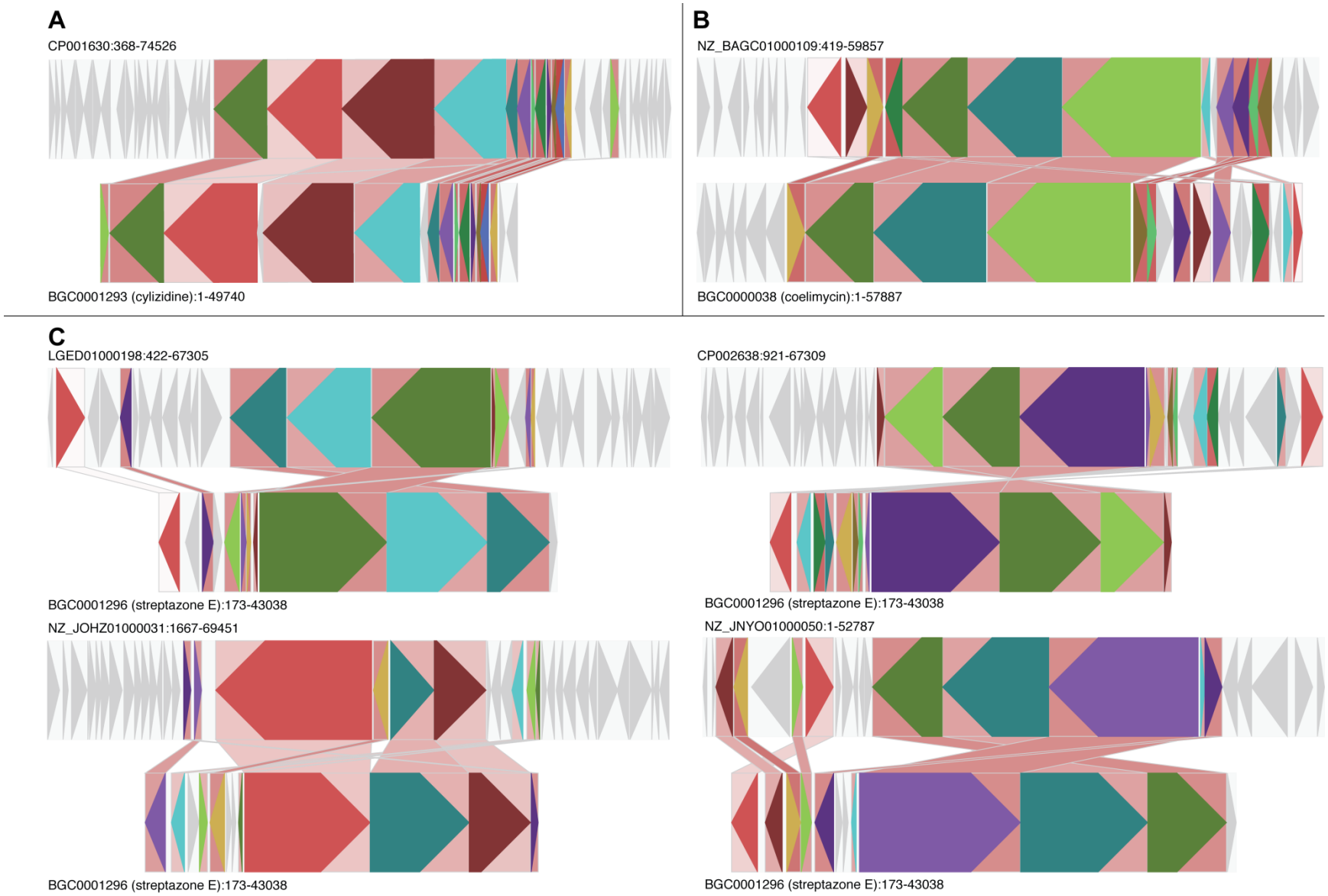

Figure S13 Comparison of cryptic gene clusters for which the metabolic products are predicted to be known compounds with published polyketide alkaloid biosynthetic gene clusters. Genes from the cryptic clusters were compared with the genes in biosynthetic gene clusters in the MiBIG[19] database using BLAST.[18] The best scoring alignments were kept and proteins from the cryptic cluster were paired with their best scoring homologs in the known biosynthetic gene clusters with redder pairings corresponding to greater homology.

References

1 A. Chhabra, A. S. Haque, R. K. Pal, A. Goyal, R. Rai, S. Joshi, S. Panjikar, S. Pasha, R. Sankaranarayanan and R. S. Gokhale, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**, 5681.
2 R. A. Vacca. S. Giannattasio, G. Capitani, E. Marra, P. Christen, *BMC Biochem.*, 2008, **9**:17.
3 B. Hwang and B. Kim, *Enzyme Microb. Technol*., 2004, **34**, 249.
4 R. Bhushan and H. Brückner, *J. Chromatogr. B. Analyt. Technol. Biomed. Life Sci.*, 2011, **879**, 3148.
5 H. J. Kim and H. S. Shin, *Anal. Chim. Acta.*, 2011, **702**, 225.
6 M. Sova,, G. Čadez, S. Turk, V. Majce, S. Polanc, S. Batson, A. J. Lloyd, D. I. Roper, C. W. G. Fishwick and S. Gobec, *Bioorg. Med. Chem. Lett*., 2009, **19**, 1376.
7 M. H. Medema, E. Takano and R. Breitling, *Mol. Biol. Evol.*, 2013, **30**, 1218.
8 B. O. Bachmann and J. Ravel, *Methods Enzymol*., 2009, **458**, 181.
9 T. Weber, K. Blin, S. Duddela, D. Krug, H. U. Kim, R. Bruccoleri, S. Y. Lee, M. A. Fischbach, R. Müller, W. Wohlleben, R. Breitling, E. Takano and M. H. Medema, *Nucleic Acids Res.*, 2015, **43**, W237.
10 S. Ohno, Y. Katsuyama, Y. Tajima, M. Izumikawa, M. Takagi, M. Fujie, N. Satoh, K. Shin-Ya and Y. Ohnishi, *Chembiochem*., 2015, **16**, 2385.
11 S. Anand, M. V. R. Prasad, G. Yadav, N. Kumar, J. Shehara, M. Z. Ansari and D. Mohanty, *Nucleic Acids Res*., 2010, **38**, W487.
12 A. Marchler-Bauer, M. K. Derbyshire, N. R. Gonzales, S. Lu, F. Chitsaz, L. Y. Geer, R. C. Geer, J. He, M. Gwadz, D. I. Hurwitz, C. J. Lanczycki, F. Lu, G. H. Marchler, J. S. Song, N. Thanki, Z. Wang, R. A. Yamashita, D. Zhang, C. Zheng and S. H. Bryant, *Nucleic Acids Res.*, 2015, D222.
13 G. Yadav, R. S. Gokhale and D. Mohanty, *J. Mol. Biol.*, 2003, **328**, 335.
14 T. A. Keatinge-Clay, *Chem. Biol.*, 2007, **14**, 898.
15 T. A. Keatinge-Clay, *J. Mol. Biol.*, 2008, **384**, 941.
16 D. H. Kwan, Y. Sun, F. Schulz, H. Hong, B. Popovic, J. C. C. Sim-Stark, S. F. Haydock and P. F. Leadlay, *Chem*. *Biol.*, 2008, **15**, 1231.
17 M. H. Medema *et al*., *Nat. Chem. Biol.*, 2015, **11**, 625.
18 C. Camacho, G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer and T. L. Madden, *BMC Bioinformatics*, 2008, **10**:421.
19 J. P. Gomez-Escribano, L. Song, D. J. Fox, V. Yeo, M. J. Bibb and G. L. Challis, *Chem. Sci.*, 2012, **3**, 2716.
20 W. Huang, S. J. Kim, J. Liu and W. Zhang, *Org. Lett.*, 2015, **17**, 5344.