Supporting Information: Nathan A. Magarvey, Monika Ehling-Schulz, Christopher T. Walsh

Characterization of the cereulide NRPS α-hydroxy acid specifying modules: Activation of α-keto acids and chiral reduction on the assembly line

Figure S1. Genetic Architecture of the vlm Biosynthetic Gene Cluster from *Streptomyces levoris* A9. 1.
Genes in blue (*vlm*1 and *vlm*2) are the Vlm NRPSs. Genes shown white are putative *vlm* biosynthetic genes: *vlm*C, putative *vlm* resistance determinant (Major Super Facilitator transporter family); *vlm*D, putative deaminase (generation of α-keto acids: pyruvate and α-KIV); *vlm*E, putative NAD(P)H-dependent FMN reductase (no predicted function) and genes shown in black orf1-3 are those predicted to flank the downstream portion of the *vlm* biosynthesis gene cluster.

Figure S2. Multiple Block Alignment of Vlm1, CesA proteins. Regions of the *S. levoris* A9 and *S. tsusimaensis* Vlm1 NRPS 2 and the *B. cereus* F4810/72 CesA NRPS 3 were compared with the AlignX Blocks program a component of the Vector NTI Advance 10.1.1 software. Blocks show in color demark regions of similarity between the three enzymes (CesA, Vlm1 *S. levoris* and Vlm1 *S. tsusimaensis*) shown with an SP-score (Sum of the Pairs Score) of greater than 200 (statistically significant). A spacer region was identified between the A and KR domains of CesA and the Vlm1 amino acid sequences. The spacer in *S. levoris* A9 and *S. tsusimaensis* Vlm1 protein is larger than CesA but does not contain a region with homology to an active site of an aminotransferase.
Figure S3. Ribbon diagram of GrsA PheA A domain X-ray Structure. Adenylation domain motifs are highlighted in color and insertion point for the $\alpha$-KR domains within the Ces and Vlm NRPSs is shown with a red arrow.
Figure S4.

Predicted Specificity Code

<table>
<thead>
<tr>
<th></th>
<th>235</th>
<th>236</th>
<th>239</th>
<th>278</th>
<th>299</th>
<th>301</th>
<th>322</th>
<th>330</th>
<th>331</th>
<th>517</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ces NRPS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CesA (A1)</td>
<td>V</td>
<td>G</td>
<td>V</td>
<td>W</td>
<td>V</td>
<td>G</td>
<td>T</td>
<td>S</td>
<td>G</td>
<td>K</td>
<td>α-Kic</td>
</tr>
<tr>
<td>CesB (A1)</td>
<td>V</td>
<td>G</td>
<td>F</td>
<td>W</td>
<td>V</td>
<td>A</td>
<td>V</td>
<td>S</td>
<td>D</td>
<td>K</td>
<td>α-Kiv</td>
</tr>
<tr>
<td>CesA (A2)</td>
<td>D</td>
<td>V</td>
<td>A</td>
<td>N</td>
<td>F</td>
<td>A</td>
<td>I</td>
<td>I</td>
<td>Y</td>
<td>K</td>
<td>Ala</td>
</tr>
<tr>
<td>CesB (A2)</td>
<td>D</td>
<td>A</td>
<td>F</td>
<td>W</td>
<td>I</td>
<td>G</td>
<td>G</td>
<td>T</td>
<td>F</td>
<td>K</td>
<td>Val</td>
</tr>
<tr>
<td>Vlm NRPS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vlm1 A1</td>
<td>A</td>
<td>A</td>
<td>L</td>
<td>W</td>
<td>I</td>
<td>A</td>
<td>V</td>
<td>S</td>
<td>G</td>
<td>K</td>
<td>α-Kiv</td>
</tr>
<tr>
<td>Vlm2 A1</td>
<td>V</td>
<td>V</td>
<td>I</td>
<td>W</td>
<td>I</td>
<td>A</td>
<td>E</td>
<td>N</td>
<td>M</td>
<td>K</td>
<td>Pyruvate</td>
</tr>
<tr>
<td>Vlm1 A2</td>
<td>D</td>
<td>A</td>
<td>F</td>
<td>W</td>
<td>V</td>
<td>G</td>
<td>G</td>
<td>T</td>
<td>?</td>
<td>K</td>
<td>Val</td>
</tr>
<tr>
<td>Vlm2 A2</td>
<td>D</td>
<td>A</td>
<td>L</td>
<td>W</td>
<td>L</td>
<td>G</td>
<td>G</td>
<td>T</td>
<td>?</td>
<td>K</td>
<td>Val</td>
</tr>
<tr>
<td>Enniatin NRPS5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ENSYN A1</td>
<td>G</td>
<td>A</td>
<td>L</td>
<td>H</td>
<td>V</td>
<td>V</td>
<td>G</td>
<td>I</td>
<td>C</td>
<td>K</td>
<td>D-Hiv</td>
</tr>
</tbody>
</table>

Table S1 NRPS codes for Ces, Vlm NRPS A domains and the proposed Enniatin NRPS α-hydroxy acid activating A domain.
NRPS codes were generated according to Stachelhaus et al., 1999.6
Supporting Information: Nathan A. Magarvey, Monika Ehling-Schulz, Christopher T. Walsh

**Materials.** Chemically competent TOP10 and BL21(DE3) *E. coli* and the pTRC-TOPO-TA vector were obtained from Invitrogen; pET28b was purchased from Novagen. All restriction endonucleases and T4 DNA ligase were obtained from New England Biolabs. Synthetic DNA oligonucleotides were purchased from Integrated DNA Technologies. PCR amplifications were run with Phusion DNA Polymerase from Finnzymes on a BioRad MyCycler thermal cycler. Pre-cast SDS-PAGE gels and protein molecular weight markers were obtained from BioRad. Ni-NTA chromatography resin was purchased from Qiagen. L-[U-14C]Valine and L-[U-14C]Leucine were obtained from American Radiolabeled Chemicals. BAS-IIIs phosphorimager plates were obtained from Fuji. DNA sequencing and quantitative amino acid analysis were performed at the Biopolymers Facility at the Harvard Medical School. d-α-hydroxy isocaproic acid was obtained from Bachem. All other materials were purchased from Sigma-Aldrich.

**Sequencing the Valinomycin Biosynthetic gene cluster from Streptomyces levoris A9.** The cosmid pVB32 generated previously was used to transform *E. coli* Top10 competent cells (Invitrogen) to ampicillin resistance. The pVB32 DNA was sheared, blunt-ended, DNA fragments were shot-gun cloned into a sequencing vector, sequenced and assembled into a single contig (SeqWright Corp.).

**Protein Overexpression and Purification of CesA, CesB fragments and CesP.** The *cesB* NRPS gene fragment encoding A-KR-PCP domains was amplified from *B. cereus* F4810/72 genomic DNA with the following primers: 5'-CCA TGG TTA AAG TGA AAG ATA TTT ATC CAC -3' (cesBNMW2_forward) and 5'-CTC GAG TAT TAA TTC AAG TAT TTT AAT CGT TG -3' (cesBNMW5_reverse). *Nco*I and *Xho*I sites are underlined. The resulting fragment was cloned into the pSmart -HCAmp vector (Lucigen) via T/A cloning. The *cesA* amplicon was generated from *B. cereus* F4810/72 genomic DNA with the following primers: 5'- CGA GTG GAA GAA CAT GAT CAC -3' (cesA_forward) and 5'-GTC CAT TTC ATG CGT ACG AGT-3' (cesA_reverse) and ligated directly into pTRC-TOPO-His. The *cesP* amplicon was amplified with the following primers: 5'-CCA TGG GAA AAC ACA TAG AGC ATA AAA ATT CC -3' (cesP_forward) and 5'-CTC GAG GCT CTC ATC CAT AAG ATT GTC GGT C-3' (cesP_reverse) cloned directing into the pCR SMART (Lucigen) where it was liberated with *Nco*I and *Xho*I, purified by agarose gel electrophoresis, and ligated into similarly digested pET28b to generate the expression construct, pET28b-CesP (C-His), encoding CesP fused to a C-terminal hexa-histidine tag. DNA sequencing of each gene or gene fragment within overexpression vectors (i.e. pET28b and pTRC-TOPO TA) was used to determine that each was free of PCR generated error. The *cesA*, *cesB* and *cesP* expression plasmids were each used to transform a *E. coli* strain BL21(DE3) strain and the resulting recombinant strains grown in LB media supplemented with 50 µg/mL kanamycin. Each liter of culture was inoculated with 10 mL of overnight starter culture. Cultures were grown at 15°C to an OD600 of 0.4-0.6, and induced with 2.5 µM IPTG (CesA), 10 uM IPTG (CesB), and 100uM (CesP) then incubated at 15°C for an additional 24 h. Cells were harvested by centrifugation (20 min at 6000g), re-suspended in lysis buffer (300 mM NaCl, 25 mM Tris pH 8.0, 10% glycerol), lysed (2 passes at 10 000-15 000 psi, Avestin EmulsiFlex-C5 high-pressure homogenizer), and the cell debris was removed by centrifugation (35 000 rpm, for 30 min, at 4°C, Beckman L7 Ultracentrifuge, 70Ti rotor).

- S4 -
Ni-NTA resin (1 mL per liter of culture) and 2 mM imidazole were added to the clarified lysate and allowed to batch bind for 1 h at 4°C. The resin was washed with 5 column volumes of 2 mM imidazole in lysis buffer. Protein was eluted with a step gradient of increasing imidazole concentrations in lysis buffer (5, 25, 50, 100, 200 mM). Most of the CesA and CesB protein fragments eluted in the 100 and 200 mM fractions, which were pooled and dialyzed at 4°C in two steps: 1 h in 100 mM NaCl, 50 mM Tris pH 8.0, 1 mM EDTA, 10% glycerol; and an additional overnight dialysis in the same buffer. Concentrated protein solutions were stored at -80°C. Protein concentrations were determined by the Bradford reagent assay with BSA as a standard.

ATP-[\(^{32}\)P]PP\(_\text{i}\) Exchange Assays for Aminoacyl-AMP Formation. To determine substrate specificity of CesA and CesB A domains, ATP-[\(^{32}\)P]PP\(_\text{i}\) reactions (100 µL) containing Tris-HCl (pH 7.5) (75 mM), MgCl\(_2\) (10 mM), 1,4-Dithiothreitol (DTT) (5 mM), ATP (5 mM), 100 µg/mL BSA, \(\alpha\)-amino acid, \(\alpha\)-keto acids, \(\alpha\)-hydroxy acid substrates (5 mM), and 1 mM [\(^{32}\)P]PP\(_\text{i}\) (2 Ci/mM), 100 nM CesA and CesB were performed at room temperature. Reactions were initiated by addition of protein. At 5 and 10 min, 50 µL of the reactions was quenched into 500 µL of 1.6% (w/v) activated charcoal, 4.5% (w/v) tetrasodium pyrophosphate, and 3.5% (v/v) perchloric acid. The charcoal was collected by centrifugation, washed twice with 500 µL of 4.5% (w/v) tetrasodium pyrophosphate and 3.5% (v/v) perchloric acid, then resuspended in 500 µL of wash solution, added to scintillation fluid, and counted by a Beckman Coulter LS 6500 liquid scintillation counter.

Preparation of radio-labeled substrates

Generation of \(^{14}\)C labeled \(\alpha\)-keto isovaleric acid from L-[U-\(^{14}\)C] valine, as well as, \(^{14}\)C \(\alpha\)-keto isocaproic acid from L-[U-\(^{14}\)C]leucine was achieved by enzymatic catalysis with the *Crotalus adamanteus* L-amino acid oxidase using a procedure adapted from Rudiger et al.\(^7\) For \(^{14}\)C labeled \(\alpha\)-keto isovaleric acid synthesis 500 µL L-[U-\(^{14}\)C]valine (0.05 mCi), 500 µL ddH\(_2\)O, 50 µL catalase (1 unit /µL), 200 µL *Crotalus adamanteus* L-amino acid oxidase (0.065 units/µL) and 500 µL Tris-HCL 1M pH 7.5 was incubated 30°C for 3h. For preparation of \(^{14}\)C labeled \(\alpha\)-keto isocaproic acid a reaction containing 500 µL L-[U-\(^{14}\)C]leucine (0.05 mCi), 500 µL H2O 50 ul catalase (1 unit /µL) 200 µL amino acid oxidase (0.065 units/µL) 500 µL Tris-HCL 1M pH 7.8 was incubated at 37°C for 3h. Purification of the labeled \(\alpha\)-keto acids (\(^{14}\)C labeled \(\alpha\)-keto isovaleric acid and \(\alpha\)-keto isocaproic acid) from the reaction mixtures was achieved by passage of reaction contents over a column containing a 5 ml of a slurry of activated Dowex-50Wx8-200 ion exchange resin and elution with 10 ml of ddH\(_2\)O. Fractions were collected at 500 µL intervals with the majority of the radioactive \(\alpha\)-keto acids collected in the seventh and eight fractions.

Post-translational modification and substrate loading of CesA and CesB PCP domains

CesA and CesB fragments were combined with CesP (3 µM) and coenzyme A (CoA, 150 µM) in the stock reaction buffer [50 mM Tris (pH 7.5), 5 mM MgCl\(_2\), and 5 mM TCEP]
and incubated at 25°C for 1 h to allow for complete phosphopantetheinylatation. Radiolabeled α-keto acids (α-KIC and α-KIV) (5 mM) and ATP (5 mM) were then added, and to the CesA and CesB reaction mixtures, respectively, and each reaction was incubated at 25°C for an additional 1 h. In parallel, CesA and CesB fragments (10 µM) were combined with CesP (3 µM) and [1-14C]acetyl-CoA (180 µM, 56 Ci/mol) in the stock reaction buffer, and each reaction incubated at 25°C for 1 h. The reactions containing [1-14C]acetyl-CoA were subjected to liquid scintillation counting to ensure that CesP was sufficient phosphopantetheinylation had occurred. Likewise the CesA and CesB reactions containing the radiolabeled α-keto acids (α-KIC, CesA and α-KIV, CesB) were analyzed by liquid scintillation counting to ensure that substrates had been loaded onto the CesA and CesB PCP domains.

**Ketoreduction assay and α-hydroxy acid detection**

Substrate loaded forms of CesA (i.e. 14C α-keto isocaproyl-S-PCP) and CesB (i.e. 14C α-keto isovaleryl-S-PCP) were assayed for ketoreductase activity in 25mM HEPES buffer pH 7.0 supplemented with 10mM NADPH and incubated at 25°C for 1h. Substrates were liberated from CesA and CesB by addition of 25 µM TycF 9 and further incubation at 25°C for 30 min. Reactions were extraction with an equal volume of ethyl acetate and the organic extract collected and spotted onto a 12.5 cm TLC plate, Merck Silica gel 60 F254 (Merck KGaA), which were then developed in a water saturated butanol, 0.5% formic acid solvent system. Authentic cold standards of α-keto isocaproic acid, α-keto isovaleric acid, D,L α-hydroxy isocaproic acid, D,L α-hydroxy isovaleric acid, L-leucine and L-valine were included on TLC plates. Plates were stained with bromocresol green and ninhydrin and the spots identifying the cold standards spotted with a solution of L-[U-14C]valine. Developed TLC plates were then exposed to BAS-IIIIs phosphorimager plates for 12h and the phosphorimager plates scanned with an Amersham Typhoon 8600 Variable Mode Imager and the image analyzed with the ImageQuant 5.2 software (Molecular Dynamics). The retention factors for α-hydroxy acids, α-keto acids and α-amino acids were calculated: α-hydroxy isovaleric acid $R_F=0.8$, α-keto isovaleric acid $R_F=0.65$, α-hydroxy isocaproic acid $R_F=0.85$, α-keto isocaproic acid $R_F=0.7$, L-valine $R_F=0.24$ and $R_F=0.30$ for L-isoleucine.

**Chiral Chromatography of α-hydroxy acids and radio-HPLC**

The α-hydroxy acids generated by the CesA and B KR domains were released from their respective PCPs using TycF9, extracted with an equal volume of ethyl acetate and concentrated in vacuo. Chiral separation of underivatized enantiomerically pure L-HIV, D-HIV was performed on a 4.5 mm X 250mm Chirex 3126 D-pencillamine chiral column (Phenomenex) using an isocratic method with 2mM Copper (II) sulfate in 15 % isopropanol as a mobile phase (1 ml/min flow rate) and UV detection (254nm). Separation of underivatized L- HIC and D-HIC was achieved by chiral chromatography with a 50 x 4.6 mm ChiralPak MA(+) (Daicel Chemical Industries) chiral column using an isocratic method with 2mM Copper (II) sulfate in 10 % acetonitrile as a mobile phase (1 ml/min flow rate) and UV detection (254nm). Chirality of 14C-labeled α-HIC and α-HIV generated by the CesA and B α-KR domains was determined using the chiral
Supporting Information: Nathan A. Magarvey, Monika Ehling-Schulz, Christopher T. Walsh

chromatography described above with the aid of a IN/US Systems β-RAM Model 2B Radioactivity HPLC Detector.


