UC Berkeley

UC Berkeley Previously Published Works

Title

Heterologous Gene Expression of N-Terminally Truncated Variants of LipPks1 Suggests a Functionally Critical Structural Motif in the N-terminus of Modular Polyketide Synthase.

Permalink

<https://escholarship.org/uc/item/71r9t3td>

Journal

ACS chemical biology, 12(11)

ISSN

1554-8929

Authors

Yuzawa, Satoshi
Bailey, Constance B
Fujii, Tatsuya
et al.

Publication Date

2017-11-01

DOI

10.1021/acscchembio.7b00714

Peer reviewed

Heterologous Gene Expression of *N*-terminally truncated variants of LipPks1 Suggests a Functionally Critical Structural Motif in the *N*-terminus of Modular Polyketide Synthase

Satoshi Yuzawa^{1,2,8*}, Constance B. Bailey¹, Tatsuya Fujii⁷, Renee Jocić^{1,8}, Jesus F. Barajas⁸, Veronica T. Benites^{1,2,8}, Edward E.K. Baidoo^{1,2,8}, Yan Chen^{1,2}, Christopher J. Petzold^{1,2,8}, Leonard Katz³, and Jay D. Keasling^{1-6,*}

¹Biological Systems and Engineering Division, Lawrence Berkeley National Laboratory, Berkeley, California, 94720, United States

²Joint BioEnergy Institute, Emeryville, California, 94608, United States

³QB3 Institute, University of California, Berkeley, California, 94720, United States

⁴Department of Bioengineering, University of California, Berkeley, California, 94720, United States

⁵Department of Chemical and Biomolecular Engineering, University of California, Berkeley, California, 94720, United States

⁶Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kogle, Allé, DK2970-Hørsholm, Denmark

⁷Research Institute for Sustainable Chemistry, National Institute of Advanced Industrial Science and Technology, Higashi-hiroshima, Hiroshima, 739-0046, Japan

⁸Agile BioFoundry, Emeryville, California, 94608, United States

Supporting Information

ABSTRACT: *Streptomyces* genomes have a high G + C content and typically use an ATG or GTG codon to initiate protein synthesis. Although gene-finding tools perform well in low GC genomes, it is known that the accuracy in predicting a translational start site (TSS) is much less for high GC genomes. LipPks1 is a *Streptomyces*-derived, well-characterized modular polyketide synthase (PKS). Using this enzyme as a model, we experimentally investigated the effects of alternative TSSs using a heterologous host, *Streptomyces venezuelae*. One of the TSSs employed boosted the protein level by 59-fold and the product yield by 23-fold compared to the originally annotated start codon. Interestingly, a structural model of the PKS indicated the presence of a structural motif in the *N*-terminus, which may explain the observed different protein levels together with a proline and arginine-rich sequence that may inhibit translational initiation. This structure was also found in 6 other modular PKSs that utilize non-carboxylated starter substrates, which may guide the selection of optimal TSSs in conjunction with start-codon prediction software.

Natural products, such as polyketides and nonribosomal peptides, from the genus *Streptomyces* have been the source of and inspiration for many clinically useful antibacterial, anti-tumor, anti-parasitic, and immunosuppressive drugs as well as herbicidal and insecticidal agents. In fact, over one-third of known polyketides and nonribosomal peptides originate from this genus.¹ Polyketides and nonribosomal peptides are produced from very large enzymes that resemble modular assembly lines, and which are intrinsically capable of creating unique and diverse molecular structures.

Genome mining is a powerful tool to discover novel compounds from sequenced genomes.^{2, 3} Two related strategies have proven successful to identify and characterize target biosynthetic gene clusters (BGCs): gene knock outs in the native hosts and heterologous gene expression in genetically tractable hosts. Heterologous expression of modular polyketide synthase (PKS) and nonribosomal peptide synthetase (NRPS) genes has been demonstrated in various hosts including several *Streptomyces* species (e.g. *S. coelicolor*, *S. albus*, *S. lividans*, and *S. avermitilis*).^{4, 5} Although these studies have significantly contributed to our understanding of many BGCs, this strategy is not always successful.

Streptomyces genomes have a base composition of 70% or more G + C and usually use an ATG or GTG codon to initiate protein synthesis.⁶ Although gene-finding tools perform well in low GC content genomes, it is known that the accuracy is much less for high GC genomes.⁷ One possible explanation of the above failure is the use of improper start codons in heterologous hosts. One may design an expression vector based on a mis-annotated start codon, which may cause translational inhibition and/or produce misfolded, non-functional proteins. Translational machinery in a heterologous host may select for translational start sites (TSSs) different from those used by the native host. Native hosts may also have enzymes to post-translationally truncate the *N*-terminus of proteins to functionalize them.⁸

Streptomyces venezuelae ATCC 10712 (hereafter referred to as *S. venezuelae*) is an emerging promising chassis for heterologous expression of genes derived from *Streptomyces* and closely related genera because of its genetic tractability, and relatively rapid doubling time, as well as having a complete genome sequence, genome engineering/editing tools, several well-characterized natural/synthetic promoters and ribosomal binding sites (RBSs).^{6, 9-11} The host also contains an exceptionally high number of phosphopantetheinyl transferases including eight Sfp types,¹² which are required to convert PKS and NRPS *apo* proteins to their corresponding *holo* forms. Using *S. venezuelae* as a heterologous host, we investigated the effects of alternative TSSs for a *Streptomyces*-derived, well-characterized modular PKS, LipPks1, as a test case.

Figure 1. A model PKS system used in this study. (A) Proposed model for β-lipomycin biosynthesis by the lipomycin synthase. (B) Proposed model for biosynthesis of **1** by LipPks1+TE.

LipPks1 comprises the load and module 1 of the lipomycin synthase from *Streptomyces aureofaciens* Tü117

(**Figure 1A**).¹³ It is composed of a loading didomain consisting of an acyltransferase (AT) domain and an acyl carrier protein (ACP) domain, and an extension module composed of a ketosynthase (KS) domain, an AT domain, a ketoreductase (KR) domain, and an ACP domain. As annotated in GenBank, translational initiation of the *lipPks1* gene starts from the proline and arginine-rich sequence and produces an approximately 170-residue *N*-terminal tail upstream of the catalytically active loading AT domain, as shown in **Figure 2** (<https://www.ncbi.nlm.nih.gov/nuccore/77864456>). In a previous study, we constructed an expression vector that encodes *N*-terminal hexahistidine (6xHis)-tagged LipPks1 with the thioesterase (TE) domain from the erythromycin PKS appended to the *C*-terminus (LipPks1+TE). LipPks1+TE produces 3-hydroxy-2,4-dimethylpentanoic acid (**1**) from isobutyryl-CoA and methylmalonyl-CoA in the presence of NADPH (**Figure 1B**). We also constructed a vector encoding LipPks1+TE that lacks the 170-residue *N*-terminal tail and a vector that encode LipPks1+TE without the *N*-terminal His tag. Heterologous expression analysis of these three genes in *Escherichia coli* indicated that the linker contributed to stable protein folding, at least, in this host¹⁴ and that the protein cannot be produced in the absence of the *N*-terminal His tag (data not shown).

M¹AGPPFPFRRRGPSGRRRCGGRATPGSV²RDRTGRRPAAV³PSRAVCA
ADLCEENDGSKNV⁴SEHRGSAGGSVLFRTGTV⁵LPWVLTGPGAAAV⁶
RARSEALRTHLRSTEWSPAGVGQALLAGTGAGADTHRAVLVLAGDRAQ
TLNALAALSAGADHPAV⁷FTSTRADASPAGPVFVFPGQGSQWTGMAREL
LDSAPVFAKRLHDCADAFAPYLGHSLLDSVTGAAGGPEPVGADVQPA
LFAVMVALTDLWNAAGVAPGALLGHSGLGELAAAHVAGVLSLDSARVV
ARWSQAQATLAGRDMVSVLLPADELADLLDRRWPGRLLVAVENGPGS
AVASGDLDAEAELVAHLTAEGIHARRVDVGLAAHSPHIDAILPRIRAD
IAPIRAHTPSIPVYSALHGGALDGTMPDAAYWCRNLRSTVRFADATRA
ALEAGHTTFVEVSPHPVLTAMEVSATRAAHAATVLGLTLRRGEG

Figure 2. Amino acid sequence of the loading AT of LipPks1 including the 170-residue *N*-terminal tail, which starts from the proline and arginine-rich sequence. Red number indicates potential TSSs in the *N*-terminal tail region. The predicted loading AT domain sequence is shaded in gray.

According to the nucleotide sequence data in GenBank, the native *lipPks1* gene starts from a GTG (= GTG₁) codon. To investigate the effect of alternate start codons in the *N*-terminal linker region, we inspected the native gene and found 6 GTG codons (= GTG₂-GTG₇) that may be used as TSSs but there was no ATG codon (**Figure 2**, see also **Table S2**). In *Streptomyces*, several different tools including FramePlot, Prodigal, and Glimmer have been used to predict open reading frames.^{7, 15, 16} FramePlot predicted GTG₁ as a start codon, which is the same start codon annotated in GenBank. However, GTG₂ was predicted as a start codon in both Prodigal and Glimmer. To investigate the impact of different TSSs in a heterologous host, *S. venezuelae*, we constructed chromosome integration vectors encoding six different *N*-terminal-truncated LipPks1+TE variants that are designed to initiate translation at one of the six GTG codons, GTG₂-GTG₇, which can be transferred from *E. coli* 12567 into *S.*

venezuelae. The introduced vector is then site-specifically integrated into the *S. venezuelae* chromosome at the *Streptomyces* phage Φ C31 attachment site mediated by the Φ C31 integrase and cognate attP site encoded in each integration vector. We also constructed integration vectors encoding *N*-terminal 6xHis-tagged, GTG₁-initiated LipPks1+TE, which was previously characterized *in vitro*,¹⁴ GTG₁-initiated LipPks1+TE as a positive control, and mCherry as a negative control. These genes were placed under the well-characterized strong kasO* promoter.^{10, 11, 17} Because the vectors encode the apramycin resistant gene, apramycin resistant transconjugants were isolated and examined. Correct integrations were verified by the polymerase chain reaction (PCR) using primers listed in **Table S1** (see also **Figure S1**). The sequence of each PCR fragment that contained the promoter and each *N*-terminal tail was also confirmed, indicating the correct integration of the plasmid into the chromosome in each *S. venezuelae* strain. Each strain was named by the start codon used (see **Figure 3**).

Figure 3. *In vivo* analysis of engineered *S. venezuelae* strains. (A) Production of **1** from each strain cultured in 042 medium was measured at 5 days by LC-TOF-MS with an authentic standard (n = 4). (B) Relative quantity of LipPks1 at day 5 in each strain was determined by MRM-LCMS method targeting two LipPks1-derived unique peptides (n = 4).

Each of the nine recombinant strains was grown in medium 042 for 5 days. Medium 042 has previously been

developed to produce secondary metabolites from *Streptomyces* sp. 17944¹⁸ and gave the best titers in our experiments among several different media tested (data not shown). The culture broth was mixed with an equal volume of methanol and filtered. Crude extracts were then analyzed by liquid chromatography time-of-flight mass spectrometry (LC-TOF-MS). As shown in **Figure 3A**, all the strains that harbored one of the LipPks1+TE variants in the genomes produced the expected **1** as a major product but the titers varied from strain to strain. The absolute titers were determined using the authentic standard.¹⁹ These strains also produced 3-hydroxy-2,4-dimethylhexanoic acid, originating from the use of 2-methylbutyryl-CoA as the starter substrate in place of isobutyryl-CoA (**Figure S2**). Their relative production levels were very similar to those of **1** (the absolute titers were not quantified). Strain GTG1 produced 0.6 ± 0.2 mg/L of **1**. Interestingly, strain GTG4, whose start codon was not predicted by any of gene-finding tools used in this study, produced the highest amount of **1** (13.8 ± 0.5 mg/L). Although there was no predicted start site with various gene-finding, it appears that there is a reasonable RBS (GGAG) 7 bp prior to the GTG4 start codon in the native gene (**Table S2**). Time course analysis of the strain showed almost linear increase in the titer over 5 days (**Figure S3**).

To explain the difference in product titers, we used the following three approaches: proteomics analysis to determine protein levels *in vivo*, *in vitro* kinetic analysis, and structural analysis through homology modeling. To compare protein levels of the PKSs in the different *S. venezuelae* strains, we took cell samples at 5 days, extracted proteins, digested the proteins using trypsin, and analyzed the tryptic digests using liquid chromatography triple-quadrupole-time-of-flight mass spectrometer (LC-QqQ) and targeted two unique PKS-derived peptides. The sum peak area of the two peptides was used to represent relative quantity of the protein in each strain. The data indicate that the protein levels reflected the amounts of **1** (**Figure 3B**), which indicates that protein levels are major determinants of the product titers. Comparison between strains GTG1 and GTG4 revealed a 59-fold difference in the protein levels at 5 days (161 ± 60 vs. 9447 ± 2126).

We previously determined the k_{cat} (0.053 min^{-1}) and K_M ($2.9 \mu\text{M}$) for production of **1** from the *N*-terminal 6xHis-tagged, GTG₁-initiated LipPks1+TE *in vitro*.¹⁴ To determine the kinetic parameters for the LipPks1+TE variant where GTG₄ is the TSS, we expressed the *C*-terminally His-tagged protein in *E. coli* and purified the protein using the same protocol. The k_{cat} and K_M values for production of **1** were $0.026 \pm 0.001 \text{ min}^{-1}$ and $1.9 \pm 0.1 \mu\text{M}$, respectively (**Figure S4**). These data suggest that the strain that encodes *N*-terminal 6xHis-tagged, GTG₁-initiated LipPks1+TE (N-His-GTG1 strain) produce twice the amount of **1** compared to the GTG4 strain if the protein levels were the same and substrate concentrations were saturated. From the proteomics studies (**Figure 3B**), we determined that 13-fold difference in the protein levels between N-His-GTG1 and GTG4 strains. Assuming that

the *in vitro* determined kinetics are reflective of those *in vivo*, one would predict 6.5-fold less production of **1** by the N-His-GTG1 strain compared to GTG4 strain, or approximately 2 mg/L. The actual titer from N-His-GTG1 strain was 1.5 ± 0.2 mg/L, which demonstrates a nice correlation between the *in vitro* and *in vivo* data.

To account for the huge difference in protein levels in the engineered *S. venezuelae* strains tested, we analyzed the modeled structure of LipPks1 loading AT including the 170-residue *N*-terminal tail (**Figure 4**, **Figure S6**). Interestingly, the model indicates the presence of a structure that resembles a subdomain between KS and AT domains in four existing KS-AT didomain crystal structures (here referred to as KS-AT linker)²⁰⁻²³. The KS-AT linker is usually ~100 residues and forms an overall β - α - α - β - α fold. This subdomain was previously proposed to partially serve as a docking site for an ACP domain in the polyketide chain elongation reaction²⁴. It may also stabilize the neighboring AT structure to strengthen substrate binding (decreased K_M) as previously demonstrated in an AT domain swapping study.²⁵ This finding may explain the unstable folding without the *N*-terminal tail in the aforementioned *E. coli* study¹⁴ and significantly decreased protein and the product levels in GTG5-7 strains. The location of GTG4 codon is approximately 20 residues before this structural motif. The construct that initiated protein synthesis from the GTG3 codon resulted in much less protein level compared to those that initiated with the GTG2 or GTG4 codons. One of the possible explanation is that the nascent polypeptide chain may form an inhibitory β -sheet structure in the ribosomal tunnel but the reason is still unclear²⁶. As shown in **Figure 2**, GTG1 codon give rise to a proline and arginine-rich sequence, which is also known to inhibit translational initiation²⁷.

Figure 4. Structural model of the loading AT of LipPks1 including the 170-residue *N*-terminal tail. Methionine and valine residues corresponding to GTG1-7 are shown in red.

In summary, we have investigated a causal relationship between protein production, catalytic activity, and TSS in a heterologous host, *S. venezuelae*, using *lipPks1* from the lipomycin BGC as a model. Importantly, our analysis indicates that the current start codon prediction algorithm for *Streptomyces* was not

adequate at least for the PKS tested. One important finding in our structural analysis was that the KS-AT linker-like structure likely exists in the *N*-terminal tail in LipPks1, which has not been experimentally characterized in any PKS that contains a loading AT. Interestingly, we also found this structural motif in 6 other PKSs analyzed including the erythromycin PKS (**Figure S5**, see also **Table S3**). An erythromycin PKS mutant that lacks the first 107 residues, which corresponds to the entire *N*-terminal tail region including the KS-AT linker-like structure, showed approximately 20-fold decreased production of erythromycin, which is consistent with our data.²⁸ It appears that the *N*-terminal tail of the borrelidin PKS also contains a part of a KS structure (**Figure S6**), which may indicate that the loading AT-ACP didomain was evolved from a KS-AT-ACP tridomain. To date, the only loading AT structure solved is the loading AT from the avermectin PKS.²⁹ Despite it possessing a high degree of sequence similarity (~50%) to that of the lipomycin PKS, the avermectin loading AT appears to lack this subdomain. From a sequence alignment of all loading ATs from PKSs in the SBSPKS database (<http://202.54.249.142/~pkssdb/sbspks/master.html>), it appears that avermectin is anomalous (**Figure S7**). The avermectin loading AT is very well-known for the substrate promiscuity³⁰. This may indicate that the avermectin PKS evolved the unusual substrate binding ability by losing the KS-AT linker-like structure.

AUTHOR INFORMATION

Corresponding Author

*Email: syuzawa@lbl.gov

*Email: jdkeasling@lbl.gov

Author Contributions

S.Y., T.F., R.J., and J.F.B. constructed plasmids, purified proteins, and determined kinetic parameters. V.T.B. and E.E.B. conducted metabolite analysis by LC-TOF. Y.C. and C.J.P. conducted protein analysis by LC-QqQ. S.Y. and C.B.B. performed structural analysis. S.Y., L.K., and J.D.K. were responsible for experimental design. All authors contributed to the preparation of the manuscript.

Note

J.D.K. has a financial interest in Amyris, Lygos, Demetrix, and Constructive Biology. L.K. has a financial interest in Lygos.

ACKNOWLEDGMENT

This work was funded by the Joint BioEnergy Institute, which is funded by the U.S. Department of Energy (DOE), Office of Science, Office of Biological and Environmental Research, and the Co-Optimization of Fuels & Engines project and Agile BioFoundry sponsored by the U.S. DOE Office of Energy Efficiency and Renewable Energy, Bioenergy Technologies and Vehicle Technologies Offices, under Contract DEAC02-05CH11231 between DOE and

Lawrence Berkeley National Laboratory. The U.S. Government retains and the publisher, by accepting the article for publication, acknowledges that the U.S. Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes. The authors declare no competing financial interest.

ASSOCIATED CONTENT

Supporting Information.

Supporting information is available, listing the contents of the material supplied as supporting information. The material is available free of charge via the internet at <http://pubs.acs.org>.

REFERENCES

- [1] Medema, M. H., and Kottmann, R., and Yilmaz, P., and Cummings, M., and Biggins, J. B., and Blin, K., and de Bruijn, I., and Chooi, Y. H., and Claesen, J., and Coates, R. C., and Cruz-Morales, P., and Duddela, S., and Dusterhus, S., and Edwards, D. J., and Fewer, D. P., and Garg, N., and Geiger, C., and Gomez-Escribano, J. P., and Greule, A., and Hadjithomas, M., and Haines, A. S., and Helfrich, E. J., and Hillwig, M. L., and Ishida, K., and Jones, A. C., and Jones, C. S., and Jungmann, K., and Kegler, C., and Kim, H. U., and Kotter, P., and Krug, D., and Masschelein, J., and Melnik, A. V., and Mantovani, S. M., and Monroe, E. A., and Moore, M., and Moss, N., and Nuttmann, H. W., and Pan, G., and Pati, A., and Petras, D., and Reen, F. J., and Rosconi, F., and Rui, Z., and Tian, Z., and Tobias, N. J., and Tsunematsu, Y., and Wiemann, P., and Wyckoff, E., and Yan, X., and Yim, G., and Yu, F., and Xie, Y., and Aigle, B., and Apel, A. K., and Balibar, C. J., and Balskus, E. P., and Barona-Gomez, F., and Bechthold, A., and Bode, H. B., and Borris, R., and Brady, S. F., and Brakhage, A. A., and Caffrey, P., and Cheng, Y. Q., and Clardy, J., and Cox, R. J., and De Mot, R., and Donadio, S., and Donia, M. S., and van der Donk, W. A., and Dorrestein, P. C., and Doyle, S., and Driessen, A. J., and Ehling-Schulz, M., and Entian, K. D., and Fischbach, M. A., and Gerwick, L., and Gerwick, W. H., and Gross, H., and Gust, B., and Hertweck, C., and Hofte, M., and Jensen, S. E., and Ju, J., and Katz, L., and Kaysser, L., and Klassen, J. L., and Keller, N. P., and Kormanec, J., and Kuipers, O. P., and Kuzuyama, T., and Kypides, N. C., and Kwon, H. J., and Lautru, S., and Lavigne, R., and Lee, C. Y., and Linquan, B., and Liu, X., and Liu, W., and Luzhetskyy, A., and Mahmud, T., and Mast, Y., and Mendez, C., and Metsa-Ketela, M., and Micklefield, J., and Mitchell, D. A., and Moore, B. S., and Moreira, L. M., and Muller, R., and Neilan, B. A., and Nett, M., and Nielsen, J., and O'Gara, F., and Oikawa, H., and Osbourn, A., and Osburne, M. S., and Ostash, B., and Payne, S. M., and Pernodet, J. L., and Petricek, M., and Piel, J., and Ploux, O., and Raaijmakers, J. M., and Salas, J. A., and Schmitt, E. K., and Scott, B., and Seipke, R. F., and Shen, B., and Sherman, D. H., and Sivonen, K., and Smanski, M. J., and Sosio, M., and Stegmann, E., and Sussmuth, R. D., and Tahlan, K., and Thomas, C. M., and Tang, Y., and Truman, A. W., and Viaud, M., and Walton, J. D., and Walsh, C. T., and Weber, T., and van Wezel, G. P., and Wilkinson, B., and Willey, J. M., and Wohlleben, W., and Wright, G. D., and Ziemert, N., and Zhang, C., and Zotchev, S. B., and Breitling, R., and Takano, E., and Glockner, F. O. (2015) Minimum Information about a Biosynthetic Gene cluster, *Nat Chem Biol* 11, 625-631.
- [2] Doroghazi, J. R., Albright, J. C., Goering, A. W., Ju, K. S., Haines, R. R., Tchaluikov, K. A., Labeda, D. P., Kelleher, N. L., and Metcalf, W. W. (2014) A roadmap for natural product discovery based on large-scale genomics and metabolomics, *Nat Chem Biol* 10, 963-968.
- [3] Dejong, C. A., Chen, G. M., Li, H., Johnston, C. W., Edwards, M. R., Rees, P. N., Skinnider, M. A., Webster, A. L., and Magarvey, N. A. (2016) Polyketide and nonribosomal peptide retro-biosynthesis and global gene cluster matching, *Nat Chem Biol* 12, 1007-1014.
- [4] Baltz, R. H. (2010) *Streptomyces* and *Saccharopolyspora* hosts for heterologous expression of secondary metabolite gene clusters, *J Ind Microbiol Biotechnol* 37, 759-772.
- [5] Komatsu, M., Komatsu, K., Koiwai, H., Yamada, Y., Kozono, I., Izumikawa, M., Hashimoto, J., Takagi, M., Omura, S., Shin-ya, K., Cane, D. E., and Ikeda, H. (2013) Engineered *Streptomyces avermitilis* host for heterologous expression of biosynthetic gene cluster for secondary metabolites, *ACS Synth Biol* 2, 384-396.
- [6] Harrison, J., and Studholme, D. J. (2014) Recently published *Streptomyces* genome sequences, *Microb Biotechnol* 7, 373-380.
- [7] Hyatt, D., Chen, G. L., Locascio, P. F., Land, M. L., Larimer, F. W., and Hauser, L. J. (2010) Prodigal: prokaryotic gene recognition and translation initiation site identification, *BMC Bioinformatics* 11, 119.
- [8] Hartmann, E. M., and Armengaud, J. (2014) N-terminomics and proteogenomics, getting off to a good start, *Proteomics* 14, 2637-2646.
- [9] Zhang, M. M., Wong, F. T., Wang, Y., Luo, S., Lim, Y. H., Heng, E., Yeo, W. L., Cobb, R. E., Enghiad, B., Ang, E. L., and Zhao, H. (2017) CRISPR-Cas9 strategy for activation of silent *Streptomyces* biosynthetic gene clusters, *Nat Chem Biol*.
- [10] Bai, C., Zhang, Y., Zhao, X., Hu, Y., Xiang, S., Miao, J., Lou, C., and Zhang, L. (2015) Exploiting a precise design of universal synthetic modular regulatory elements to unlock the microbial natural products in *Streptomyces*, *Proc Natl Acad Sci U S A* 112, 12181-12186.
- [11] Phelan, R. M., Sachs, D., Petkiewicz, S. J., Barajas, J. F., Blake-Hedges, J. M., Thompson, M. G., Reider, A. P., Rasor, B. J., Katz, L., and Keasling, J. D. (2017) Development of Next Generation Synthetic Biology Tools for Use in *Streptomyces venezuelae*, *ACS Synth Biol* 6, 159-166.
- [12] Baltz, R. H. (2016) Genetic manipulation of secondary metabolite biosynthesis for improved production in *Streptomyces* and other actinomycetes, *J Ind Microbiol Biotechnol* 43, 343-370.
- [13] Bihlmaier, C., Welle, E., Hofmann, C., Welzel, K., Vente, A., Breitling, E., Muller, M., Glaser, S., and Bechthold, A. (2006) Biosynthetic gene cluster for the polyenoyltetramic acid alpha-lipomycin, *Antimicrob Agents Chemother* 50, 2113-2121.
- [14] Yuzawa, S., Eng, C. H., Katz, L., and Keasling, J. D. (2013) Broad substrate specificity of the loading didomain of the lipomycin polyketide synthase, *Biochemistry* 52, 3791-3793.
- [15] Ishikawa, J., and Hotta, K. (1999) FramePlot: a new implementation of the frame analysis for predicting protein-coding regions in bacterial DNA with a high G + C content, *FEMS Microbiol Lett* 174, 251-253.
- [16] Delcher, A. L., Bratke, K. A., Powers, E. C., and Salzberg, S. L. (2007) Identifying bacterial genes and endosymbiont DNA with Glimmer, *Bioinformatics* 23, 673-679.
- [17] Wang, W., Li, X., Wang, J., Xiang, S., Feng, X., and Yang, K. (2013) An engineered strong promoter for streptomycetes, *Appl Environ Microbiol* 79, 4484-4492.
- [18] Rateb, M. E., Yu, Z., Yan, Y., Yang, D., Huang, T., Vodanovic-Jankovic, S., Kron, M. A., and Shen, B. (2014) Medium optimization

of *Streptomyces* sp. 17944 for tirandamycin B production and isolation and structural elucidation of tirandamycins H, I and J, *J Antibiot* (Tokyo) 67, 127-132.

[19] Eng, C. H., Yuzawa, S., Wang, G., Baidoo, E. E., Katz, L., and Keasling, J. D. (2016) Alteration of Polyketide Stereochemistry from anti to syn by a Ketoreductase Domain Exchange in a Type I Modular Polyketide Synthase Subunit, *Biochemistry* 55, 1677-1680.

[20] Tang, Y., Kim, C. Y., Mathews, I. I., Cane, D. E., and Khosla, C. (2006) The 2.7-Angstrom crystal structure of a 194-kDa homodimeric fragment of the 6-deoxyerythronolide B synthase, *Proc Natl Acad Sci U S A* 103, 11124-11129.

[21] Tang, Y., Chen, A. Y., Kim, C. Y., Cane, D. E., and Khosla, C. (2007) Structural and mechanistic analysis of protein interactions in module 3 of the 6-deoxyerythronolide B synthase, *Chem Biol* 14, 931-943.

[22] Whicher, J. R., Smaga, S. S., Hansen, D. A., Brown, W. C., Gerwick, W. H., Sherman, D. H., and Smith, J. L. (2013) Cyanobacterial polyketide synthase docking domains: a tool for engineering natural product biosynthesis, *Chem Biol* 20, 1340-1351.

[23] Herbst, D. A., Jakob, R. P., Zahringer, F., and Maier, T. (2016) Mycocerosic acid synthase exemplifies the architecture of reducing polyketide synthases, *Nature* 531, 533-537.

[24] Yuzawa, S., Kapur, S., Cane, D. E., and Khosla, C. (2012) Role of a conserved arginine residue in linkers between the ketosynthase and acyltransferase domains of multimodular polyketide synthases, *Biochemistry* 51, 3708-3710.

[25] Yuzawa, S., Deng, K., Wang, G., Baidoo, E. E., Northen, T. R., Adams, P. D., Katz, L., and Keasling, J. D. (2017) Comprehensive in Vitro Analysis of Acyltransferase Domain Exchanges in Modular Polyketide Synthases and Its Application for Short-Chain Ketone Production, *ACS Synth Biol* 6, 139-147.

[26] Wilson, D. N., and Beckmann, R. (2011) The ribosomal tunnel as a functional environment for nascent polypeptide folding and translational stalling, *Curr Opin Struct Biol* 21, 274-282.

[27] Ito, K., Chiba, S., and Pogliano, K. (2010) Divergent stalling sequences sense and control cellular physiology, *Biochem Biophys Res Commun* 393, 1-5.

[28] Pereda, A., Summers, R. G., Stassi, D. L., Ruan, X., and Katz, L. (1998) The loading domain of the erythromycin polyketide synthase is not essential for erythromycin biosynthesis in *Saccharopolyspora erythraea*, *Microbiology* 144 (Pt 2), 543-553.

[29] Wang, F., Wang, Y., Ji, J., Zhou, Z., Yu, J., Zhu, H., Su, Z., Zhang, L., and Zheng, J. (2015) Structural and functional analysis of the loading acyltransferase from avermectin modular polyketide synthase, *ACS Chem Biol* 10, 1017-1025.

[30] Moore, B. S., and Hertweck, C. (2002) Biosynthesis and attachment of novel bacterial polyketide synthase starter units, *Natural Product Reports* 19, 70-99.

Supplemental Figures

Figure S1. Genomic PCR analysis. (A) Expected PCR fragments from each strain. (B) Comparison between PCR fragments from each strain that contain a part of Φ C31 integrase, kasO* promoter, and different sizes of *N*-terminal tails. Lane 1 = N-His-GTG1; Lane 2 = GTG1; Lane 3 = GTG2; Lane 4 = GTG3; Lane 5 = GTG4; Lane 6 = GTG5; Lane 7 = GTG6; Lane 8 = GTG7. (C) Genomic PCR data for LipPks1+TE region. Lane 1 = the first PCR fragment (= 1711 or 1183 bp); Lane 2 = the second PCR fragment (= 2381 bp); Lane 3 = the third PCR fragment (= 2477 bp); Lane 4 = the fourth PCR fragment (= 2709 bp)

Figure S2. Production of 3-hydroxy-2,4-dimethylhexanoic acid in genomically engineered *S. venezuelae* strains. (A) Proposed model for biosynthesis of 3-hydroxy-2,4-dimethylhexanoic acid by LipPks1+TE. (B) Production of 3-hydroxy-2,4-dimethylhexanoic acid from each strain cultured in 042 medium was measured at 5 days by LC-TOF (n = 4). Peak areas were used to compare the relative production levels.

(B)

1 2 3 4 5 6 7 8

Figure S3. Time course analysis of (A) 3-hydroxy-2,4-dimethylpentanoic acid production (n = 3), (B) protein production (n = 3), and (C) cell density (estimated by Bradford assay of the cell pellets) for N-His-GTG1 and GTG4 strains (n = 3).

(A)

(B)

(C)

Figure S4. Initial velocity analysis for GTG4-initiated LipPks1+TE for different starter acyl-CoA substrates. Prism (<https://www.graphpad.com/>) was used to calculate kinetic parameters and errors.

	Initial velocity (min ⁻¹)	Std. Error	Initial velocity (min ⁻¹)
kcat	0.05196	kcat	0.0007801
Km	1.074	Km	0.2574

	Initial velocity (min ⁻¹)	Std. Error	Initial velocity (min ⁻¹)
kcat	0.0289	kcat	0.0004605
Km	5.628	Km	0.3958

	Initial velocity (min ⁻¹)	Std. Error	Initial velocity (min ⁻¹)
kcat	0.0258	kcat	0.0001498
Km	1.898	Km	0.0994

	Initial velocity (min ⁻¹)	Std. Error	Initial velocity (min ⁻¹)
kcat	0.08457	kcat	0.0009411
Km	1.351	Km	0.1774

	Initial velocity (min ⁻¹)	Std. Error	Initial velocity (min ⁻¹)
kcat	0.08294	kcat	0.003016
Km	47.53	Km	4.309

Figure S5. Structural models of the loading AT domain of eryAI (from the erythromycin PKS), aveAI (from the avermectin PKS), borAI (from the borrelidin PKS), cfa6 (from the coronafacic acid PKS), lkmAI (from the lankamycin PKS), megAI (from the megalomicin PKS), and plm1 (from the phoslactomycin PKS).

Erythromycin PKS

There is no valine residue (GTG codon) before its KS-AT linker-like structure.

Avermectin PKS

KS-AT linker-like structure was not observed.

Borrelidin PKS

There is a valine residue (GTG codon) before a structural element in the KS domain. Potential TSSs are the following.

M₁TGSAV₂SAPFLQPPEPVS...

Coronafacic acid PKS

There is a valine residue (GTG codon) before its KS-AT linker-like structure. Potential TSSs are the following (shown in red in the structure).

M₁CRAWTP TSMNERYGNNLPGSCCALS NRIDWTSPSRSTPSGKKA AVNRFQYFCRS
LPV₂TTFT...

Lankamycin PKS

There is no valine residue (GTG codon) before its KS-AT linker-like structure.

Megalomicin PKS

There is a valine residue (GTG codon) before its KS-AT linker-like structure. Potential TSSs are the following (shown in red in the structure).

M₁VDV₂PDLLGTRTPHPGPLP...

There are valine residues (GTG codon) before its KS-AT linker-like structure. Potential TSSs are the following (shown in red in the structure).

M₁VPSARSRV₂V₃LLLARTDEEDQPDWDDVPGV₄SV₅CAPRDSGSIFATAETCLRAD
 RPALVVVHDETGGV₆AV₇ALALETAGAEHYGPLCDDAEERAAQLRALRPRDRTSAL
 DATT...

Figure S6. A structural model of (A) the loading AT domain of LipPks1 (from the lipomycin PKS, green) superimposed on the KS-AT didomain from module 3 of the erythromycin PKS (PDB = 2QO3, yellow) and (B) the loading AT domain of borAI (from the borrelidin PKS, blue) superimposed on the KS-AT didomain from module 3 of the erythromycin PKS (PDB = 2QO3, yellow).

(A)

Lipomycin PKS loadingAT + Erythromycin PKS module 3 KS-AT

(B)

Borrelidin PKS loadingAT + Erythromycin PKS module 3 KS-AT

Figure S7. A sequence alignment analysis of loading AT domains including *N*-terminal tails. Abbreviations: Lip, lipomycin; Ery, erythromycin; Ave, avermectin; Bor, borrelidin; Cor, coronafacic acid; Lan, lanlamycin; Meg, megalomicin; Pho, phoslactomycin. Loading AT domain were shown in bold.

```
Cor -----
Pho -----
Lan -----
Ery -----
Meg -----
Bor MTGSAVSAPFLQPPEPVSGHSEKSDPVLVVGAGRRARMADAVRAAGAQAQAGIDPAVLRRT
Lip -----
Ave -----
```

```
Cor -----
Pho -----M
Lan -----
Ery -----
Meg -----
Bor RATLITAGSAGAAGRLAAALRLTGATISLDTRETPTLLALHLAAQALRAGDTSYAVVGAE
Lip -----
Ave -----
```

```
Cor -----M
Pho VPSARSRVVL-----LLARTDEEDQPDWDDVPGVSVCAPRDSGSIFATAET
Lan -----
Ery -----
Meg -----
Bor LPDGNCALILARQSAATAEGAVPQAIVRTTTADRTTTAD----HAPAPDDHGSPA--REA
Lip -----MAG----PPPFERRRGPSG--RRR
Ave -----
```

```
Cor C---RAWTP TSMNERYGN-----NLPGCCALSNRIDWTSPSRSTPSGKKA AV
Pho CLRADRPALVVVHDETG GVAVALALETAGAEHYGPLCDDA---EER-AAQLRALRPRD--
Lan -----MNA
Ery -----MAD
Meg -----MVD
Bor PHATRTLSPGITQAPAE GFPG-----LLATLHDDT---PLR-PTAVTEHGSDATT
Lip C--GGRATPGSVRDRTGRRPA AVP---SRAVCAADLCEEN---DDG-SKNVSEHRGSAGG
Ave -----
```

```
Cor NRFQYFCRSLPVTTFTVPLILSDASPAALAATAQRLRAEHETR SFAEWQVFC-----RQQ
Pho ---RTSALDATTAGTVPLLLSGRTEAGLRAHAGRLATHLQER--TDISVADAAF--TQA
Lan SPLQGENISAATRGAA RPWVLSAADDRALRVWARRLLDHLRAH--PESGADGVGAALAHQ
Ery LSKLSDSRTAQPGRIVRPWPLSGCNESALRARARQLRAHLDRF--PDAGVEGVGAALAHQ
Meg VPDLLGTRTPHPGPLPFPWPLCGHNEPELRARARQLHAYLEGI--SEDDVVAVGAALARE
Bor VLVLLDQPQDAAPAAPLPWVVSAPHTRALRATAATLAVHLDTT--P-AAPADVAHTLL--
Lip SV-L-----FPRTGTVLPWVLTGPGAAAVRARSEALRTHLRAS--TEWSPAGVGQALLAG
Ave -----
```

```
Cor LTRDHAEYRAAILATDQASLLKGLDMLARGRASRHLVLGRA-----DQLRQPVLVVFPQGQ
Pho TARTLWPHRAMVAVADRDAVVSALRAMAEGRAGDA AVRAGS-----GGPQSAVFVFPQGQ
Lan DR--RLPRRAVIVGSSRAEILDGLAAVADDRPHPAVARGTA-----AAPAPVTFVFPQGQ
Ery EQADAGPHRAVVVASTSELLDGLAAVADGRPHASVVRGVA-----RPSAPVVFVFPQGQ
Meg TRAQDGPHRAVVVASTVELTAALAAALQGRPHPSVVRGVA-----RPTAPVVFVLPQGQ
Bor -TARPDHRHAAVVGADRATLTDLRLALATGGDAPHLVHGTA-----TGSRPRVFVFPQGQ
```

Lip TGAGADTHRAVVLAGDRAQTLNALAALSAGADHPA---VFTSTRADASPAGPVFVFPQGQ
Ave -----MQRMDGG-EEPRPAAGEVLGVADADGGVVFVFPQGQ
: : . . : * : * * *

Cor PLWPRMTTRLSATLPSYQQALHRHGDLEARIGWNPALAE---GERIERLLRIQPLQF
Pho AQWVGMARELAATTPVFRDKLAECARELRPFVDFELDDVLSG---ALPLERVEVIQPALF
Lan SQWPGMATELLGESAVFADAIERCAQVFAAESGWSLTEALETAAAPGTVPRVEIVQPALY
Ery AQWAGMAGELLGESRVFAAAMDACARAFEPVTDWTLAQVLDS---PEQSRRVEVVQPALF
Meg AQWPGMATRLLAESPVFAAAMRACERAFDEVTDWSLTEVLDS---PEHLRRVEVVQPALF
Bor SQWPGMAAELLETSPEFHDVSHACADALAEFVDWSVLDVLRQAPDAPPLRRVDVLQPTLW
Lip SQWTGMARELLDSAPVFARKLHDCADAFAPYLGHSLDSVTGAAGGPEPVGADVQPALF
Ave PQWPGMGRELLDASDVRESVRACEAFAFYVDWSVEQVLRDSPDAPGLDRVDVVQPTLF
* * . * : : : . : * * :

Cor ALACSLAETWQAFGIEPAAVIGHSVGEAAAAHVGGYLDEASGAALTVLWGETTLTAEQGQ
Pho AVMVSLAELWRSNGVTPAAVVGHSFGIEAAVTAAGALTVTDGARLVAAVSKALARLQNGG
Lan AVSAALETWRSFGVEPDVVGHSMDGLAAHAAGAVTIEDGARAVARWSKELRPLVGKG
Ery AVQTSLAALWRSFGVTPDAVVGHSIGELAAAHVCGAAGAADAARAAALWSREMIPLVNGG
Meg AVQTSLAALWRSFGVTPDAVVGHSIGELAAAEVCGAVDVEAAARAAALWSREMIPLVGRG
Bor ATMVSLAEVWRSYGVEPAAVVGHCCEIAAAQVAGALDMRDAARLLAHRRAWLRLVGKG
Lip AVMVALTDLWNAAGVAPGALLGHSLGELAAAHVAGVLSLDD SARVVARWSQAQATLAGRG
Ave AVMISLAALWRSQGVPECAVLGHSLGEIAAAHVSGGLSLADAARVVTLWSQAQTTLAGTG
* : * : * : * : * : * : * : * : * : * : * : * : *

Cor GMLSVAALDDIKPLL-----DDQESLSVAAINAPRSVTLSGTLMALDALQARLTDAGFW
Pho EMVAVALPAEQVTALVA---EWDLDLDIAVVGPGSTVVAGTTESATALLERLRARDVR
Lan DMASVELPADLVAPRLA---DWGPDVLVLAGINGPRSVLLAGATDAIERCVAELTAEGVR
Ery DMAAVALSADEIEPRIA---RWDDDVVLVAGVNGPRSVLLTGSPPEVARRVQELSAEGVR
Meg DMAAVALSPAELAAARVE---RWDDDVVPAGVNGPRSVLLTGAPEPIARRVAELAAQGV
Bor TVISVATSGQDITRMA---AWPDSVELAALNGPRSVLAGPPDVLGDIVNDLTDQGIH
Lip DMVSVLLPADELADLL---DRRWPGRLVVAVENGPGSAVASGDLDAAEVLVAHLTAEGIH
Ave ALVSVAAATPEDELLPRIAPWTEDNPARLAVAAVNGPRSTVVSAGAREAVADLVADLTAAQVR
: : * : : : * * * * : * *

Cor AWRVPGGQVAGHGPQVEPLRDAVTSR-APAKVMDGQIPYYSSVSGGRQNGEHLDDVDYWFR
Pho ATLLP-IGIAGHSWRMEPAHEYLVQEAARVPRVTGIPVYTSTTTDPLDTGGLDAEHWFR
Lan ARTIG-VSMAAHS PHMDWVVARVRAALDWFTPGAGAVPLYSSRTGAVLDTRELTAEHWCA
Ery AQVIN-VSMAAHS AQVDDIAEGMRSALAWFAPGGSEVPFYASLTGGAVDTRELVDYWR
Meg AQVVN-VSMAAHS AQVDAVAEGMRSALTWFAPGSDVPYAGLTGGRLDTRELGAHWPR
Bor AKRIPGVDTVGHCSQVEVLRDHLDLVLRPVS PRPAAVPFYSTVDGTERDTTTLDTDYWYL
Lip ARRVD-VGLAAHSPHIDAILPRIRADIAPIRAHTPSIPVYSALHGGALDGTMDAAYWCR
Ave TRMIP-VDVPAAHSPMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRLDTRELDAAWYR
: : . * : : : * : : : : . : *

Cor VLRETVRFQDSLQALIAADGHRFLFIEVSPHPVLTSLIDEALRGAGVVGG-----AVATLDQ
Pho SLREPARFQQVIEELLGRGHRVFMSPHPVLTLPLEETAHHLG-----RDVVVLDTMRR
Lan NFRGRVEFERSVRDCTAERGGVFLELSPHPVLAAAVRQTLDAAG-----SGHPVLATLRR
Ery SFRLPVRFDEAIRSALEVPGTFVEASHPHVLAAALQQTLD AEG-----SSAAVPTLQR
Meg SFRLPVRFDEATRIVLELQPGTFIESSHPHVLAAALQQTLD AEG-----SPAAIVPTLQR
Bor NTRSQVRFHQAVRNLLAAGHRSFVEVSPHPVLLGASIEDTAAEFG-----LDDVAAVGTLRR
Lip NLRSTVRFADATRAALEAGHTTFVEVSPHPVLT TAMEVSA-----TRAHAATVLGTLRR
Ave NMSSTVRFEPARLLLQQGPKTFVEMSPHPVLT MGLQELAPDLGDTTGTADTVIMGTLRR
.. * . * : * * * * : : * : :

Supplemental Tables

Table S1. Primers used in genomic PCR experiments in **Figure S1**.

Primer name	Target	Primer sequence
SY398 (Forward primer)	Amplify 1183-1711 bp fragment	CACTCGACAAGTTCGTTGCGGAA
SY333 (Reverse primer)	Amplify 1183-1711 bp fragment	GTAGCCTGCGCTTGGCTCCA
SY67 (Forward primer)	Amplify 2381 bp fragment	TGGACTCCGCACCGGTTTTTCG
SY513 (Reverse primer)	Amplify 2381 bp fragment	AGAACCCAACCACAACGGGC
SY70 (Forward primer)	Amplify 2477 bp fragment	AGCGATGCGCGTCGTCTGG
SY514 (Reverse primer)	Amplify 2477 bp fragment	GTAACCATCCACAGCGGTGCA
SY73 (Forward primer)	Amplify 2709 bp fragment	GTTACCGCACCGCCTTTAAACG
SY332 (Reverse primer)	Amplify 2709 bp fragment	AGCTGTGCTTCTGTTCGGCT

Table S2. Native gene sequence of the *N*-terminal tail of LipPks1. GTG₄ codon and the potential ribosomal binding site were shown in bold.

PKS name	Gene sequence
Lipomycin PKS	GTGGCGGGGCTCCGCCGTTCCCCCGACGGCGGGGGCCAGCGGCCGTCGGCGGTGCGGCGGAC GGCCACCCCCGGATCCGTGCGTGACCGACCGGGCGGCGCCCGGCCCGTCCCGTACGGGC CGTCTGTGCCGACACCTTTGTGAAGAGAACGACGAC GGGAG CAAGAAC GTG TCCGAACACCGT GGCAGTGCAGGCGGCTCCGTACTGTTCCCCCGACGGGGACCGTGCTGCCCTGGGTCTGACCG GTCCCGGCGCCCGCCGTGCGGGCCCGTCCGAGGCTCTGCGCACGCATCTCCGCGCGAGCAC CGAATGGTGCCTGGCGCGGCGTCCGGACAGGCCCTCCTCGCCGGGACTGGAGCCGGGGCGGACACC CACCGGGCGCTCGTACTCGCCGGGGACCGCGCAGACCCTGAACGCGCTCGCCGCGCTGTCCG CCGGCGCCGACCACCCGGCGGTGTTCAACAGACCCGCGCGGACGCCTCCCCCGGGGGCCG

Table S3. Amino acid sequences used in structural modeling in **Figure 4** and **Figure S5**.

PKS name	Amino acid sequence
Lipomycin PKS	MAGPPFFPRRRGPSRRRCGRATPGSVDRDTGRRPAAVPSRAVCAADLCEENDDGSKNVSEHR GSAGGSVLFPRGTGVLPLWVLTGPGAAAVRARSEALRTHLRASTEWSPAGVGQALLAGTGAGADT HRAVVLGADRAQTNLALAALSAGADHPAVFTSTRADASPAGPVFVFPQGQSQWTGMARELLDSA PVFARKLHDCADAFAPYLGHSLLDSVTGAAGGPEPVGADVQPALFAMVALTDLWNAAGVAPG ALLGHSIGELAAAHVCGAAGAADAARAAALWSREMIPLVNGDMAAVALSADEIEPRIARWDDV RLVAVENGPGSAVASGDLDAAEELVAHLTAEGIHARRVDVGLAAHSPHIDAILPRIRADIAP RAHTPSIPVYSALHGGALDGTMPDAAYWCRNLRSTVRFADATRAALEAGHTTFVEVSPHPVLT AMEVSATRAHAATVLGTLRRGEG
Erythromycin PKS	MADLSKLSDSRTAQPGRIVRPWPLSGCNESALRARARQLRAHLDRFPDAGVEGVGAALAHDEQA DAGPHRAVVASSTSELDDGLAAVADGRPHASVVRGVARPSAPVVFVFPQGAQWAGMAGELLG ESRVFAAAMDACARAFEPVTDWTLAQVLDSPEQSRREVVQPALFAVQTSALALWRSFGVTPDA VVGHSIGELAAAHVCGAAGAADAARAAALWSREMIPLVNGDMAAVALSADEIEPRIARWDDV VLAVNGPRSVLLTGSPPEVARRVQELSAEGVRAQVINVSMAAHSQAQVDDIAEGMRSALAWFAP GGSEVPFYASLTGGAVDTRELVDYWRRSFRPLPVRFDEAIRSALEVPGTFVEASPHPVLAAL QQLDAEGSSAAVPTLQRGQG
Avermectin PKS	MQRMDGGEEPRAAGEVLGVADADGGVVFVFPQGQPQWPGMGRELLDASDVRESVRACEAAFAF VEQVLRDSDPADGLDRVDVQPTLFAVMISLAALWRSQGVEPCAVLGHSIGELAAAHVSGGLSLAD TLWSQAQTTLAGTGALVSAATPDELLPRIAPWTEDNPARLAVAANGPRSTVVSAGAREAVADLVA QVTRMIPVDVPAHSPLMYAIEERVVSGLLPITPRPSRIPFHSSVTGGRDLTRELDAAWYRNMSS PAARLLLQGGPKTFVEMSPHPVLTMLGLQELAPDLGDTTGTADTVIMGTLRRQGQ
Borrelidin PKS	MTGSAVSAPFLQPPEPVSGHSEKSDPVLLVGAGRRARMADAVRAAGAQAQIDPAVLRRTATL ITAGSAGAAGRLAAALRLTGATISLDTRETPTLLALHLAAQALRAGDTSYAVVGAELPDGNCAL

	ILARQSAATAEGAVPQAIIVRTTADRTTADHAPAPDDHGSPAREAPHATRTLSPGITQAPAE FPGLLATLHDDTPLRPTAVTEHGSDATTVLVLLDQPDAAAPAPLPWVVSAPHTRALRATAATL AVHLDTTAPAPADVHTLLTARPDHRHAUVGADRATLTDLRALATGGDAPHLVHGTATGSPR PVFVFPQGGSQWPGMAAELLETFEPFHDSVHACADALAEFVDWSVLDVLRQAPDAPPLRRVDVL QPTLWATMVSLAEVWRSYGVPEAAVVGHCCEIAAAQVAGALDMRDAARLLAHRSRRAWLRVVGK GTVISVATSGQDITRRMAAWPDSVELAALNGPRSVLAGPPDVLGDIVNDLTDQGIHAKRIPGV DTVGHCSQVEVLRDHLDDVLRPVSPPAAVFPYSTVDGTERDTTLDTDYWLNTRSQVRFHQ VRNLLAAGHRSFVEVSPHPLLGASIEDTAAEFGLDDVAAVGTLRRGQG
Coronafacic acid PKS	MCRAWTPSTSMNERYGNLPGSCCALSNRIDWTSRSTPSGKKAANRFQYFCRSLPVTTFTVPLI PAALAATAQRLRAEHETRSFAEWQVFCRQQLTRDHAERYRAAILATDQASLLKGLDMLARGRASRL DQLRQPVLVFPQGGLWPRMTTSLSATLPSYQQALHRHGDLEIARIGWNPARALAEGERIERLLRI ALACSLAETWQAFGIEPAAVIGHSVGEAAAAHVGGYLDEASGAALTVLWGETLTAIEGQGGMLSV IKPLDDQESLSVAANAPRSVTLSGTLMALDALQARLTDAGFWAWRVPGGQVAGHGPPQVEPLRDA PAKVMGQIPYSSVSGGRQNGEHLDDVDYWFRVLRETFRFQDSLQALIADGHRFLFIEVSPHPVLTS LRGAGVVGGAVALTDQRKDD
Lankamycin PKS	MNASPLQGENISAATRGAAWPVLSAADDRALRVWARRLLDHLRAHPESGADGVGAALAHQDRR LPRRAVIVGSSRAEILDGLAAVADDRPHPAVARGTAAAPAPVTFVFPQGGSQWPGMATELLGES AVFADAIERCAQVFAAESGWSLTEALETAAAPGTVPVEIVQFALYAVSAALTELWRSFGVEPD AVVGHSMGDLAAHAAGAVTIEDGARAVARWSKELRPLVKGDMASVELPADLVAPRLADWGPD LVLAGINGPRSVLLAGATDAIERCVAELTAEGVRARTIGVSMAAHSPHMDWVVARVRAALDWFT PGAGAVPLYSSRTGAVLDTRELTAEHWCANFRGRVEFERSVRDCTAERGGVFLELSHPVLAAA VRQTLDAAGSGHPVLATLRRGQG
Megalomicin PKS	MVDVPDLGTRTPHPGPLPFPWPLCGHNEPELRARARQLHAYLEGISEDVAVGAALARETRAQD VVVASSVTELTAAALAQAQRPHPSVVRGVARPTAPVVFVLPQGGAQWPGMATRLLAESPVFAAAM AFDEVTDWSLTEVLDSPEHLRRVEVVQPALFAVQTSALALWRSFGVRPDAVLGHSIGELAAAEVCG AAARAAALWSREMPVLVGRGDMAAVALSPAELAAARVERWDDDVVPAGVNGPRSVLLTGAPEPIARR AQGVRAQVVNVSMAAHSAQVDAVAEGMRSALTWFAPGDSVPYAGLTGGRDLTRELGDHWPERSF FDEATRAVLELQPGTFIESSPHPVLAASLQQTLDDEVGSPAIVPTLQRDQG
Phoslactomycin PKS	MVPSARSRVLLARTDEEDQPDWDDVPGVSVCAPRDSGSIFATAETCLRADRPALVVVHDETGGV LETAGAEHYGPLCDDAEERAAQLRALRPRDRTSALDATTAGTVPLLLSGRTEAGLRAHAGRLATHI ISVADAAFTQATARTLWPHRAMVAVADRDAVVSALRAMAEGRAGDAAVRAGSGGPQSAVFVFPQG MARELAATTPVFRDKLAECARELRPFVDFELDDVLSGALPLERVEVIQPALFAVMVSLAELWRSNG VVGHSFGEIAAVTAAGALTVDGARLVAAVSKALARLQNGEMVAVALPAEQVTALVAEWDLDLDI PGSTVVAGTTESATALLERLRARDVRATLLPIGIAGHSWRMEPAHEYLVEQEAARVPRVTGIPVYT PLDTGGLDAEHWFRLREPARFQQVIEELLGRGHRVFEVMSHPVLTLPTEETAHLGRDVVVLDT

Table S4. Plasmids used in this study

Plasmid	Summary	Source or reference
pSY169 (JPUB_TBD)	kasOp*-GTG1-started LipPks1+TE	This study
pSY170 (JPUB_TBD)	kasOp*-GTG2-started LipPks1+TE	This study
pSY171 (JPUB_TBD)	kasOp*-GTG3-started LipPks1+TE	This study
pSY172 (JPUB_TBD)	kasOp*-GTG4-started LipPks1+TE	This study
pSY173 (JPUB_TBD)	kasOp*-GTG5-started LipPks1+TE	This study
pSY174 (JPUB_TBD)	kasOp*-GTG6-started LipPks1+TE	This study
pSY175 (JPUB_TBD)	kasOp*-GTG7-started LipPks1+TE	This study
pSY196 (JPUB_009552)	T7p-GTG4-started LipPks1+TE-His	This study

Plasmid 22 (JPUB_009550)	kasOp*-mCherry	This study
Plasmid 34 (JPUB_TBD)	kasOp*-His-GTG ₁ -started LipPks1+TE	This study

Supplemental Methods

Chemicals. All chemicals were purchased from Sigma-Aldrich unless otherwise described.

Plasmids. Gene fragments that encode LipPks1+TE or the variants were PCR-amplified from pSY074₁ and subcloned into a slightly modified Streptomyces Integration Vector that encodes ΦC31 integrase and kasO* promoter₂ using NdeI and EcoRI to generate pSY169-pSY175 and plasmid 34. mCherry gene was also subcloned into the same vector (Plasmid 22). To express GTG₄-started LipPks1+TE in *E. coli*, the corresponding gene was subcloned into pET30 vector (pSY196). Plasmids used in this study are listed in **Table S4**. The plasmids along with their associated information have been deposited in the public version of JBEI registry (<http://public-registry.jbei.org>) and are physically available from the authors upon request.

Conjugal transfer of vectors to *S. venezuelae*. *E. coli* ET12567/pUZ8002 was transformed with an appropriate vector and selected for on LB agar containing apramycin (50 µg/mL). A single colony was used to inoculate a 5 mL of LB containing kanamycin (25 µg/mL), chloramphenicol (15 µg/mL), and apramycin (25 µg/mL) at 37°C. The overnight culture was used to seed 10 mL of LB containing the same antibiotics, which was grown at 37°C to an OD₆₀₀ of 0.4-0.6. The *E. coli* cells were pelleted by centrifugation, washed twice with LB, and resuspended in 500 µL of LB. Fresh *S. venezuelae* spores were collected from a glycerol-arginine plate with 5 mL of 2xYT and incubated at 50°C for 10 min. The spores (500 µL) and the *E. coli* cells (500 µL) were mixed, spread onto mannitol soy agar, and incubated at 23°C for 16-20 h. After addition of nalidixic acid (20 µg/mL) and apramycin (40 µg/mL) were added, the plate was further incubated for 3-4 days. A single colony was used to inoculate into TSB containing nalidixic acid (25 µg/mL) and apramycin (25 µg/mL). Following overnight incubation, the culture was spread onto a glycerol-arginine plate. The plate was incubated for 2-3 days. The spores were collected from the plate with 3-4 mL of water and mixed with glycerol to prepare 25% glycerol stock. The glycerol stock was stored at -80°C.

Polyketide production. Engineered *S. venezuelae* spores were grown in 5 mL of TSB medium containing nalidixic acid (50 µg/mL) and apramycin (50 µg/mL) for 1-2 days at 30°C. The overnight culture was used to seed 5 mL of 042 containing the same antibiotics, which was grown at 30°C for overnight. 3% of the overnight culture was then used to inoculate 30 mL of 042 medium containing the same antibiotics, which was grown for 2-5 days at 30°C. 300 µL of supernatants were collected from the culture, mixed with 300 µL of methanol, and then incubated at 50°C for overnight (the tube was covered with parafilm). The resulting solution was cooled down at 4°C, centrifuged, and filtered using Amicon Ultra Centrifugal filters, 3 kDa Ultracel, 0.5 mL device (Millipore). The flow-through was analyzed by LC-TOF MS.

Polyketide detection. LC separation of 3-hydroxycarboxylic acids was conducted on a Kinetex XB-C18 column (100-mm length, 3.0-mm internal diameter, and 2.6-µm particle size; Phenomenex, Torrance, CA USA) using an Agilent Technologies 1200 Series Rapid Resolution HPLC system (Agilent Technologies, Santa Clara, CA, USA). A sample injection volume of 3 µL was used. The sample tray and column compartment were set to 6 and 50°C, respectively.

The mobile phase was composed of 0.1% formic acid (Sigma-Aldrich, St. Louis, MO, USA) in water (A) and 0.1% formic acid in methanol (B). 3-hydroxycarboxylic acids were separated using the following gradient: linearly increased from 5% to 72.1% B in 6.5 min, further increased to 95% B in 1.3 min, held at 95 % for 1 min, decreased from 95% to 5% B in 0.4 min, and held at 5% B for 2 min. The flow rate was held at 0.42 mL/min for 8.8 min, linearly increased from 0.42 mL/min to 0.65 mL/min in 0.4 min, and held at 0.65 mL/min for 2 min. The total LC run time was 11.2 min. The HPLC system was coupled to an Agilent Technologies 6210 series time-of-flight mass spectrometer (TOF-MS). Drying and nebulizing gases were set to 11 L/min and 30 lb/in², respectively, and a drying-gas temperature of 330°C was used throughout. MS detection of [M – H][–] ions was conducted in the negative ion mode via electrospray ionization. A capillary voltage of 3,500 V was used throughout. The fragmentor, skimmer, and OCT 1 RF V_{pp} voltages were set to 150, 50, and 170 V, respectively. The acquisition range was from 95-1000 m/z, and the acquisition rate was 0.86 spectra/sec. The TOF-MS was tuned with the Agilent ESI-L Low concentration tuning mix in the range of 50-1700 m/z. Data acquisition and processing were conducted by Agilent MassHunter Workstation and Agilent MassHunter Profinder, respectively.

Proteomics. Engineered *S. venezuelae* strains were grown in 042 medium for 5 days. The cells were harvested by centrifugation. Protein extraction from *E. coli* cell pellets were accomplished using a chloroform/methanol precipitation method as previously described³. Briefly, cell pellets were resuspended in 400 µl of methanol and briefly vortexed, followed by sequential additions of 100 µl of chloroform and 300 µl of water with short intervals of vortexing in between. Phase separation was achieved by centrifugation at 14,000 rpm for 1 minutes. The methanol/water layer was removed and 300 µl of methanol was added afterwards. The protein was pellet after another centrifugation, and the supernatant was decanted. The resulted protein was resuspended in 100 mM ammonium bicarbonate buffer supplemented with 20% methanol and urea at final concentration of 3 M. Protein concentration was determined by the DC assay (BioRad, Hercules, CA). Proteins were then reduced 5 mM tris 2-(carboxyethyl) phosphine for 30 min at room temperature, and alkylated with 10 mM iodoacetamide for 30 min at room temperature in the dark. Proteins were digested overnight at a final concentration of 1 mg/mL with trypsin at a 1:50 trypsin: total protein ratio. The resulted tryptic peptides were subjected to LCMS analysis. All SRM methods development and assays were performed on an Agilent 6460 QQQ mass spectrometer system coupled with an Agilent 1290 UHPLC system (Agilent Technologies). A Sigma–Aldrich Ascentis Peptides ES-C18 column (2.1 mm × 50 mm, 2.0 µm particle size, operated at 60) was used to separate peptides at 0.400 mL/min standard flow rate. The mobile phase was composed of 0.1% formic acid in water (solvent A) and 0.1% formic acid in acetonitrile (solvent B). Peptides were separated using the following gradient: 5% to 35% B for 20 min, followed by a short ramp up to 80% B and held for 1 min, and then ramped down to 5% B and re-equilibrate the column. Peptides were introduced to the mass spectrometer from LC and ionized by using an Agilent Jet Stream source (Agilent Technologies) operating in positive-ion mode with the following parameter settings: Sheath Gas flow=11 l/min, Sheath Gas Temperature=350C, Nozzle Voltage=1000 V, Nebulizing Pressure=30 psi, Chamber Voltage=4500 V. The targeted MRM method and acquired data were public available in panorama website under the following project link: <https://goo.gl/8fHMvX>.

Protein purification. An *E. coli* K207-3 strain⁴ harboring pSY196 was grown in LB medium supplemented with appropriate antibiotics at 37°C until the OD₆₀₀ reached 0.4-0.6. The cultures were then cooled to 18°C and induced with 250 μ M isopropyl- β -D-galactopyranoside for 16-18 h. The cells were harvested by centrifugation and resuspended in lysis/wash buffer (50 mM phosphate, pH 7.6, 300 mM sodium chloride, 10 mM imidazole, 4°C). The cells were lysed by sonication, and cellular debris was removed by centrifugation. Nickel-NTA agarose resin (Qiagen) was added directly to the supernatant (10 mL of resin slurry per L of culture) and mixed for 1 h at 4°C. The resulting mixture was poured into a fritted column, washed with 10 resin volumes of lysis/wash buffer (4°C), and eluted with 2 resin volumes of elution buffer (150 mM phosphate, pH 7.6, 50 mM sodium chloride, 150 mM imidazole, 4°C). The eluted protein was then applied to a HiTRAP Q anion exchange column (GE Healthcare), washed with 5 resin volumes of wash buffer (50 mM phosphate, pH 7.6, 200 mM sodium chloride, 8% glycerol, 4°C), and eluted at approximately 375 mM sodium chloride. The fractions containing target PKSs were concentrated using an Amicon Ultra-15 Centrifugal Filter, 100K device (Millipore). The resulting purified protein was stored at -80°C.

Enzyme kinetics. 3-Hydroxycarboxylic acid production was carried out in 100 mM phosphate buffer, pH 7.2, containing 2.5 mM TCEP at 23°C in the presence of 0.5 μ M GTG₄-started LipPks1+TE, various concentrations of starter acyl-CoA substrates (propionyl-CoA, butyryl-CoA, isobutyryl-CoA, 2-methylbutyryl-CoA, and isovaleryl-CoA), 200 μ M of methylmalonyl-CoA and 500 μ M NADPH. The reactions were quenched at different time points by adding an equal volume of methanol and centrifuged. The supernatants were then analyzed by LC-MS. Initial velocity at each concentration of substrate was determined by analyzing 3 different time points (n = 1).

Structural modeling. Amino acid sequences used in structural modeling are listed in **Table S2**. I-TASSER (<https://zhanglab.ccmb.med.umich.edu/>) was used to generate all tertiary structure models. The internal C-score scoring function in I-TASSER was utilized to select the best homology models for each amino acid sequence used in **Figure 4**, **Figure S5**, and **Figure S6**.

Multiple sequence alignment. Amino acid sequences listed in **Table S2** are analyzed by Clustal Omega (<http://www.ebi.ac.uk/Tools/msa/clustalo/>).

Supplemental References

- [1] Yuzawa, S., Deng, K., Wang, G., Baidoo, E. E., Northen, T. R., Adams, P. D., Katz, L., and Keasling, J. D. (2017) Comprehensive in Vitro Analysis of Acyltransferase Domain Exchanges in Modular Polyketide Synthases and Its Application for Short-Chain Ketone Production, *ACS Synth Biol* 6, 139-147.
- [2] Phelan, R. M., Sachs, D., Petkiewicz, S. J., Barajas, J. F., Blake-Hedges, J. M., Thompson, M. G., Reider Apel, A., Rasor, B. J., Katz, L., and Keasling, J. D. (2017) Development of Next Generation Synthetic Biology Tools for Use in *Streptomyces venezuelae*, *ACS Synth Biol* 6, 159-166.
- [3] Wessel, D., and Flugge, U. I. (1984) A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids, *Anal Biochem* 138, 141-143.
- [4] Murli, S., Kennedy, J., Dayem, L. C., Carney, J. R., and Kealey, J. T. (2003) Metabolic engineering of *Escherichia coli* for improved 6-deoxyerythronolide B production, *J Ind Microbiol Biotechnol* 30, 500-509.