Supplementary Information

Maximizing Heterologous Expression of Engineered Type I Polyketide Synthases: Investigating Codon Optimization Strategies

Matthias Schmidt[1,2,3,4], Namil Lee[1,2,4#], Chunjun Zhan[1,2,#], Jacob B. Roberts[1,2,5], Alberto A. Nava[1,2,8], Leah S. Keiser[1,2,8], Aaron A. Vilchez[1,2,8], Yan Chen[1,2], Christopher J. Petzold[1,2], Robert W. Haushalter[1,2], Lars M. Blank[3], Jay D. Keasling[1,2,5,6,7,8,9,10,*]

[1]Joint BioEnergy Institute, 5885 Hollis Street, Emeryville, California 94608, United States.
[2]Biological Systems & Engineering Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, United States.
[3]Institute of Applied Microbiology (iAMB), Aachen Biology and Biotechnology (ABBt), RWTH Aachen University, 52062 Aachen, Germany.
[4]California Institute for Quantitative Biosciences (QB3), University of California, Berkeley, California 94720, United States.
[5]Joint Program in Bioengineering, University of California, Berkeley, California 94720, United States.
[6]Department of Chemistry, University of California, Berkeley, California 94720, United States.
[7]Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, United States.
[8]Department of Chemical and Biomolecular Engineering, University of California, Berkeley, California 94720, United States.
[9]The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark.
[10]Center for Synthetic Biochemistry, Institute for Synthetic Biology, Shenzhen Institutes for Advanced Technologies, Shenzhen 518071, China.

#Authors contributed equally.
*Corresponding author: Jay D. Keasling, jdkeasling@lbl.gov

Table S1: List of all applied constraints during codon optimization.

Codon optimization constraints
UniquifyAllKmers(9, include_reverse_complement=True)
AvoidHairpins(stem_size=10,hairpin_window=100)
AvoidPattern("9xA")
AvoidPattern("9xT")
AvoidPattern("6xC")
AvoidPattern("6xG")
AvoidPattern("NdeI_site")
AvoidPattern("XhoI_site")
AvoidPattern("SpeI_site")
AvoidPattern("BamHI_site")
AvoidPattern("BsaI_site")
EnforceGCContent(mini=0.3, maxi=0.75, window=50)
EnforceTranslation()

Table S2: Primers used in this study.

Primer	Sequence	Description
Confirmation primers for serine recombinase-assisted genome engineering		
cPCR-AG5577_BxB1_F	AGCAAATTCGGCAACACGC	Confirmation of BxB1 integrations into AG5577 BxB1 attB site
cPCR-AG5577_BxB1_R	CTAGGCAGAATTTTGGGAGTGGC	Confirmation of BxB1 integrations into AG5577 BxB1 attB site

cPCR-AG5577_MR11_F	GTGATTTGAAAGAGTTGTCAGTTAGCTCG	Confirmation of MR11 integrations into AG5577 MR11 attB site
cPCR-AG5577_MR11_R	AGGACTCACCTCTAGAACACGC	Confirmation of MR11 integrations into AG5577 MR11 attB site
cPCR-AG6212_BxB1_F	CGAATTCTTTCATTTAAGACCCTAATA	Confirmation of BxB1 integrations into AG6212 BxB1 attB site
cPCR-AG6212_BxB1_R	CTAGGCAGAATTTTGGGAGTGG	Confirmation of BxB1 integrations into AG6212 BxB1 attB site
cPCR-LipPKS-Cg_ubc_F	CGCCATGGAACTGCGTAA	Anneals to 3’ end of codon optimized LipPKS gene. Confirmation of integration into the host genome
cPCR-LipPKS-Cg_ubc_R	GGTATGGTGCAAAGGCATCT	Anneals to 5’ end of codon optimized LipPKS gene. Confirmation of integration into the host genome
cPCR-LipPKS-Cg_mcu_F	CGCCATGGAATTGCGAAAC	Anneals to 3’ end of codon optimized LipPKS gene. Confirmation of integration into the host genome
cPCR-LipPKS-Cg_mcu_R	AAAGACTGTGGCCCAAGTAG	Anneals to 5’ end of codon optimized LipPKS gene. Confirmation of integration into the host genome
cPCR-LipPKS-Cg_hrca_F	AGAGCTGGACTCTGGAACT	Anneals to 3’ end of codon optimized LipPKS gene. Confirmation of integration into the host genome

cPCR-LipPKS-Cg_hrca_R	GCGTTCCAAAGGTCGGTTA	Anneals to 5’ end of codon optimized LipPKS gene. Confirmation of integration into the host genome
cPCR-LipPKS-Pp_ubc_F	CGCCATGGAACTGCGTAA	Anneals to 3’ end of codon optimized LipPKS gene. Confirmation of integration into the host genome
cPCR-LipPKS-Pp_ubc_R	GTCACGCTATCCAGCAGAC	Anneals to 5’ end of codon optimized LipPKS gene. Confirmation of integration into the host genome
cPCR-LipPKS-Pp_mcu_F	GCCATGGAACTGCGAAAC	Anneals to 3’ end of codon optimized LipPKS gene. Confirmation of integration into the host genome
cPCR-LipPKS-Pp_mcu_R	GTCACGGAATCCAGCAGAC	Anneals to 5’ end of codon optimized LipPKS gene. Confirmation of integration into the host genome
cPCR-LipPKS-Pp_hrca_F	CTTGACCGCCATGGAACT	Anneals to 3’ end of codon optimized LipPKS gene. Confirmation of integration into the host genome
cPCR-LipPKS-Pp_hrca_R	TCACGCTATCCAGCAGAGA	Anneals to 5’ end of codon optimized LipPKS gene. Confirmation of integration into the host genome
cPCR-LipPKS-Ec_ubc_F	GCGACGTTGGCTTTGATAG	Anneals to 3’ end of codon optimized LipPKS gene. Confirmation of integration into the host genome

cPCR-LipPKS-Ec_ubc_R	CAAATAACGCTGGCTGCAC	Anneals to 5’ end of codon optimized LipPKS gene. Confirmation of integration into the host genome
cPCR-LipPKS-Ec_mcu_F	CGCCATGGAACTGCGTAA	Anneals to 3’ end of codon optimized LipPKS gene. Confirmation of integration into the host genome
cPCR-LipPKS-Ec_mcu_R	CACGGAATCCAGCAGACT	Anneals to 5’ end of codon optimized LipPKS gene. Confirmation of integration into the host genome
cPCR-LipPKS-Ec_hrca_F	TGGGCTTCGACAGCTTAAC	Anneals to 3’ end of codon optimized LipPKS gene. Confirmation of integration into the host genome
cPCR-LipPKS-Ec_hrca_R	TCACGCTATCCAGCAGACT	Anneals to 5’ end of codon optimized LipPKS gene. Confirmation of integration into the host genome
cPCR-LipPKS-WT_R	CGCAGTCGTGCAGCTTA	Anneals to 5’ end of codon optimized LipPKS gene. Confirmation of integration into the host genome
cPCR-LipPKS-WT_F	TCTTCGACCACCCGACA	Anneals to 3’ end of codon optimized LipPKS gene. Confirmation of integration into the host genome
cPCR-Sc_mmCoA_F	GTCGGCAAGAGCACGTT	Anneals to 3’ end of S. cellulosum So ce56 mmCoA operon. Confirmation of

		integration into the host genome
cPCR-Sc_mmCoA_R	ATCGTGAAGCGCAGCTC	Anneals to 5' end of S. cellulosum So ce56 mmCoA operon. Confirmation of integration into the host genome
Confirmation primers for gene deletions and replacements in AG6212		
cPCR-prpDBC2_F	GTGTTACCGATCCACTGGGCGTCAACG	Confirmation of *prpDBC2* deletion
cPCR-prpDBC2_F	CATTGCGCATTCCGATCATGCGCGTCTGCG	Confirmation of *prpDBC2* deletion
cPCR-kivd-CCL4_F	GACATCATTCACCAGCAGGTCGGTGGACTTCGTGC	Confirmation of *kivd-CCL4* replacing Cgl0605
cPCR-kivd-CCL4_R	GCGGTGTGCGGGATGACTTCGGAGTAGATG	Confirmation of *kivd-CCL4* replacing Cgl0605
cPCR-sfp_F	GTCAATGTTGACGCGGCCTGGACGGGTGGAACCGGT	Confirmation of *sfp* replacing Cgl1016
cPCR-sfp_F	ACCATCCGCCACATCGAGTCTGTCCACCAGCT	Confirmation of *sfp* replacing Cgl1016
Primers for RT-qPCR analysis		
qPCR-LipPKS-Cg_hrca_F	ATGTCCGAACACAGAGGAAGT	Amplification of LipPKS cDNA for qPCR
qPCR-LipPKS-Cg_mcu_F	CATATGTCCGAGCACAGGG	Amplification of LipPKS cDNA for qPCR
qPCR-LipPKS-Cg_ubc_F	ATGTCCGAACACAGGGGTAG	Amplification of LipPKS cDNA for qPCR
qPCR-LipPKS-Ec_hrca_F	ATGAGCGAACACAGAGGATCA	Amplification of LipPKS cDNA for qPCR
qPCR-LipPKS-Ec_mcu_F	ATGAGCGAACATCGTGGTAGT	Amplification of LipPKS cDNA for qPCR

qPCR-LipPKS-Ec_ubc_F	ATGAGCGAACATCGTGGC	Amplification of LipPKS cDNA for qPCR
qPCR-LipPKS-Pp_hrca_F	CATATGAGCGAGCACAGGG	Amplification of LipPKS cDNA for qPCR
qPCR-LipPKS-Pp_mcu_F	CATATGTCGGAGCACAGGG	Amplification of LipPKS cDNA for qPCR
qPCR-LipPKS-Pp_ubc_F	CATATGAGCGAGCACCGG	Amplification of LipPKS cDNA for qPCR
qPCR-LipPKS-WT_GTG_F	CATGTGTCCGAACACCGTG	Amplification of LipPKS cDNA for qPCR
qPCR-LipPKS-WT_ATG_F	CATATGTCCGAACACCGTGG	Amplification of LipPKS cDNA for qPCR
qPCR-LipPKS-Cg_hrca_R	TCTAAGATGAGTCCTCAGAGCCT	Amplification of LipPKS cDNA for qPCR
qPCR-LipPKS-Cg_mcu_R	CTAAGATGCGTTCGCAGAGC	Amplification of LipPKS cDNA for qPCR
qPCR-LipPKS-Cg_ubc_R	GAGATGGGTTCGCAGAGCTT	Amplification of LipPKS cDNA for qPCR
qPCR-LipPKS-Ec_hrca_R	AGATGAGTTCTCAGAGCCTCG	Amplification of LipPKS cDNA for qPCR
qPCR-LipPKS-Ec_mcu_R	GTTCGCAGAGCCTCGCTA	Amplification of LipPKS cDNA for qPCR
qPCR-LipPKS-Ec_ubc_R	GTAAATGGGTTCGCAGAGCTTC	Amplification of LipPKS cDNA for qPCR
qPCR-LipPKS-Pp_hrca_R	GTTCGCAGAGCCTCGGAT	Amplification of LipPKS cDNA for qPCR
qPCR-LipPKS-Pp_mcu_R	GTTCGCAGAGCCTCGCTA	Amplification of LipPKS cDNA for qPCR
qPCR-LipPKS-Pp_ubc_R	GTTCGCAGAGCCTCGCTA	Amplification of LipPKS cDNA for qPCR
qPCR-LipPKS-WT_GTG_R	AGATGCGTGCGCAGAG	Amplification of LipPKS cDNA for qPCR

qPCR-LipPKS-WT_ATG_R	AGATGCGTGCGCAGAG	Amplification of LipPKS cDNA for qPCR
qPCR-Pp-RpoD_F	CAGGGCTACCTGACTTACGC	*P. putida* housekeeping gene used as internal control for qPCR
qPCR-Pp-RpoD_R	ACGTTGATCCCCATGTCGTT	*P. putida* housekeeping gene used as internal control for qPCR
qPCR-Ec-RpoD_F	GGAGCAAAACCCGCAGTC	*E. coli* housekeeping gene used as internal control for qPCR
qPCR-Ec-RpoD_R	CGACGATATCTTCCGGCAG	*E. coli* housekeeping gene used as internal control for qPCR
qPCR-Cg-RpoC_F	GTGCTCGACGTAAACGTCTTC	*C. glutamicum* housekeeping gene used as internal control for qPCR
qPCR-Cg-RpoC_R	GAGGGTTCGGTAGTTGATGGT	*C. glutamicum* housekeeping gene used as internal control for qPCR
Gibson primers for assembly of pAN001 LipPKS-WT_GTG		
gPCR-WT-LipM1-Frag1_F	CGAATTCAAAAGATCTTTTAAGAAGGAGATATACATGTGTCCGAACACCGTGGCAGTGC	Amplification of WT sequence of LipM1 fragment 1
gPCR-WT-LipM1-Frag1_R	CCTCGTCGCGAGGGCGAACGCCACGTCGGCG	Amplification of WT sequence of LipM1 fragment 1
gPCR-WT-LipM1-Frag2_F	ACGTGGCGTTCGCCCTCGCGACGAGGCGTACTGC	Amplification of WT sequence of LipM1 fragment 2
gPCR-WT-LipM1-Frag2_R	TCCCGCTGTCCAGCTCGGCCTTGAGGTACGCG	Amplification of WT sequence of LipM1 fragment 2

gPCR-WT-EryM6-TE_F	CCTCAAGGCCGAGCTGGACAGCGGGACTCCCGCCC	Amplification of WT sequence of EryM6-TE
gPCR-WT-EryM6-TE_R	GTTTTATTTGATGCCTGGAGATCCTTACTCGATCAGTGGTGGTGGTGGTGGTGC	Amplification of WT sequence of EryM6-TE
Gibson primers for assembly of pAN001 LipPKS-WT_ATG		
gPCR-WT-LipM1-ATG-Frag1_F	CGAATTCAAAAGATCTTTTAAGAAGGAGATATACAT**A**TGTCCGAACACCGTGGCAGTGC	Amplification of WT sequence of LipM1 fragment 1 with ATG start codon mutation
Gibson primers for assembly of pBH026 RFP		
gPCR-pBH026_F	GAAGGTCGTCACTCCACCGGTGCTTAAGGATCCAAACTCGAGTAAGGATCTCCAGG	Amplification of pBH026 plasmid backbone
gPCR-pBH026_R	ACGTCTTCGCTACTCGCCATATGTATATCTCCTTCTTAAAAGATCTTTTGAATTCG	Amplification of pBH026 plasmid backbone
gPCR-RFP_F	AGAAGGAGATATACATATGGCGAGTAGCGAAGACG	Amplification of *rfp* gene
gPCR-RFP_R	TTAAGCACCGGTGGAGTGACGACCT	Amplification of *rfp* gene
Gibson primers for assembly of pK18 ΔCgl0605::kivd-CCL4		
gPCR-pK18-Cgl0605_F	TCGGGTGGGCCTTTCTGCGTTTATAGCCCTTGATTATTGCCAAAGAAACCTTTAAGGACT	Amplification of pK18 ΔCgl0605 plasmid backbone
gPCR-pK18-Cgl0605_R	CGAGCCGCAGCCGAATGTGACTAGTTAACGTGCAGGCTTACCTTTTGGAAGC	Amplification of pK18 ΔCgl0605 plasmid backbone
gPCR-kivd-CCL4_F	GCTTCCAAAAGGTAAGCCTGCACGTTAACTAGTCACATTCGGCTGCGGCTCG	Amplification of *kivd* and *CCL4* expression cassette

gPCR-kivd-CCL4_R	AGTCCTTAAAGGTTTCTTTGGCAATAATCAAGGGCTATAAACGCAGAAAGGCCCACCCGA	Amplification of *kivd* and *CCL4* expression cassette
Gibson primers for assembly of pJH209 Sc_mmCoA		
gPCR-pJH209_F	GGCCAGGAACCGTAAAAAAGTCAAAAGCCTCCGGTCGGAGG	Amplification of pJH209 plasmid backbone
gPCR-pJH209_R	CGACTTCGTGACAACGATCCCCCAACTGAGAGAACTCAAAGG	Amplification of pJH209 plasmid backbone
gPCR-Ptac-Sc_mmCoA_F	AGTTCTCTCAGTTGGGGGATCGTTGTCACGAAGTCGACTACG	Amplification of *S. cellulosum* So56 mmCoA operon under the control of P_{tac}
gPCR-Ptac-Sc_mmCoA_F	GACCGGAGGCTTTTGACTTTTTTACGGTTCCTGGCCTTTTGC	Amplification of *S. cellulosum* So56 mmCoA operon under the control of P_{tac}
Gibson primers for assembly of pGingerBG-NahR LipPKS-Pp_mcu		
gPCR-pGingerBG-NahR_F	GCGGCAATAGCTGAGGATCCAAACTCGAGTAAGGATCTCC	Amplification of pGingerBG-NahR plasmid backbone
gPCR-pGingerBG-NahR_R	ACCCCTGTGCTCCGACATATGTATATCTCCTTCTTAAATGATGGCTTTATTG	Amplification of pGingerBG-NahR plasmid backbone
gPCR-LipPKS-Pp_mcu-Frag1_F	GAAGGAGATATACATATGTCGGAGCACAGGGGTAGTG	Amplification of LipPKS-Pp_mcu fragment 1
gPCR-LipPKS-Pp_mcu-Frag1_R	CGCAAGCCTGCCATAGTGCGGCCAGACTCACC	Amplification of LipPKS-Pp_mcu fragment 1
gPCR-LipPKS-Pp_mcu-Frag2_F	TCTGGCCGCACTATGGCAGGCTTGCGGGGTG	Amplification of LipPKS-Pp_mcu fragment 2

gPCR-LipPKS-Pp_mcu-Frag2_R	CGAGTTTGGATCCTCAGCTATTGCCGCCGCCCAG	Amplification of LipPKS-Pp_mcu fragment 2
Gibson primers for assembly of pGingerBG-NahR LipPKS-Ec_mcu		
gPCR-pGingerBG-NahR_F	GGCGGAGGGAATAGTTAAGGATCCAAACTCGAGTAAGGATCTCC	Amplification of pGingerBG-NahR plasmid backbone
gPCR-pGingerBG-NahR_R	CCACGATGTTCGCTCATATGTATATCTCCTTCTTAAATGATGGCTTTATTG	Amplification of pGingerBG-NahR plasmid backbone
gPCR-LipPKS-Ec_mcu-Frag1_F	GAAGGAGATATACATATGAGCGAACATCGTGGTAGTGCAGGT	Amplification of LipPKS-Ec_mcu fragment 1
gPCR-LipPKS-Ec_mcu-Frag1_R	GTTCTTCATCACCGCTACGTAGAACCTCCAGCAGGTTCCAATCCA	Amplification of LipPKS-Ec_mcu fragment 1
gPCR-LipPKS-Ec_mcu-Frag2_F	GGAGGTTCTACGTAGCGGTGATGAAGAACTGAGTAATCGCG	Amplification of LipPKS-Ec_mcu fragment 2
gPCR-LipPKS-Ec_mcu-Frag2_R	CGAGTTTGGATCCTTAACTATTCCCTCCGCCCAGCC	Amplification of LipPKS-Ec_mcu fragment 2

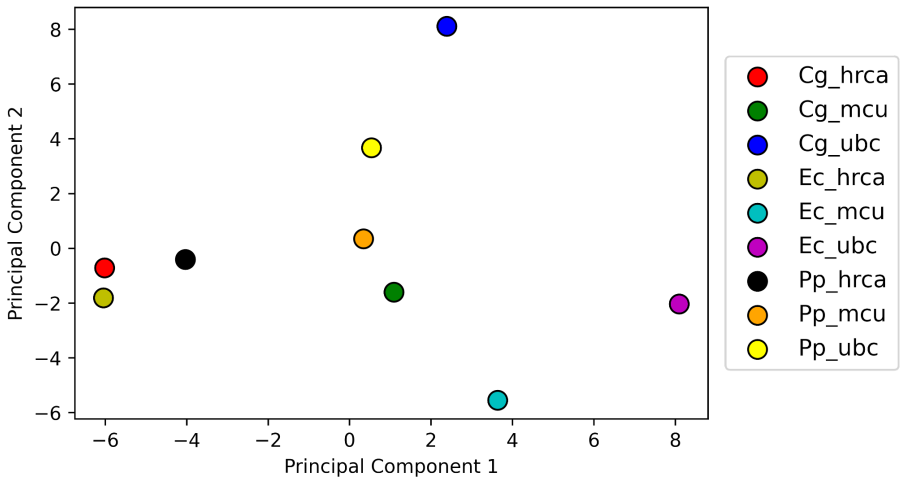

Figure S1: Principal component analysis of optimized LipPKS sequences. A high similarity in codon usage leads to clustering of the respective nucleotide sequences. The hrca optimized genes are in close proximity to each other.

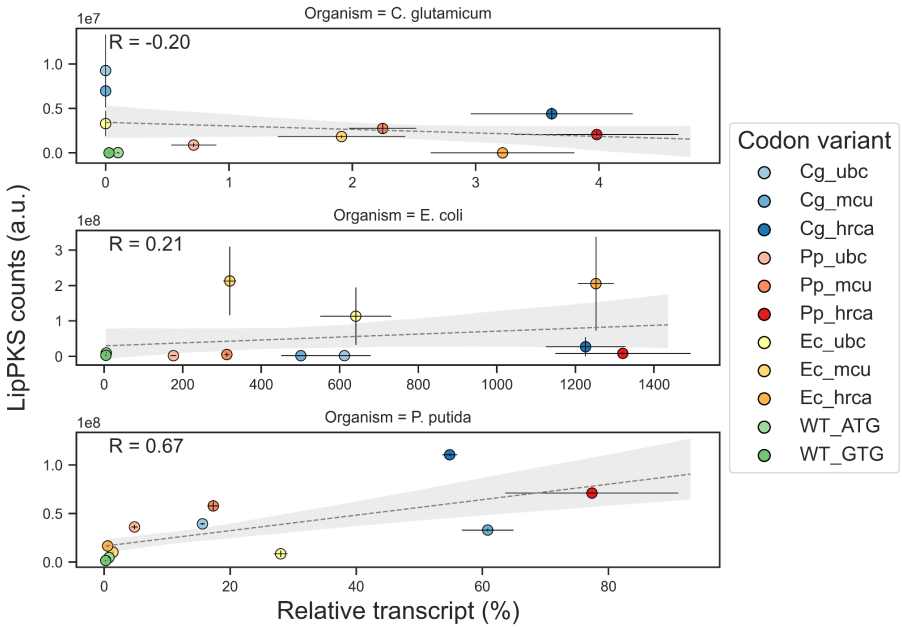

Figure S2: Regression plot of LipPKS counts and relative transcript for *C. glutamicum*, *E. coli*, and *P. putida* (n = 3). Only *P. putida* showed a slight correlation between these two parameters.

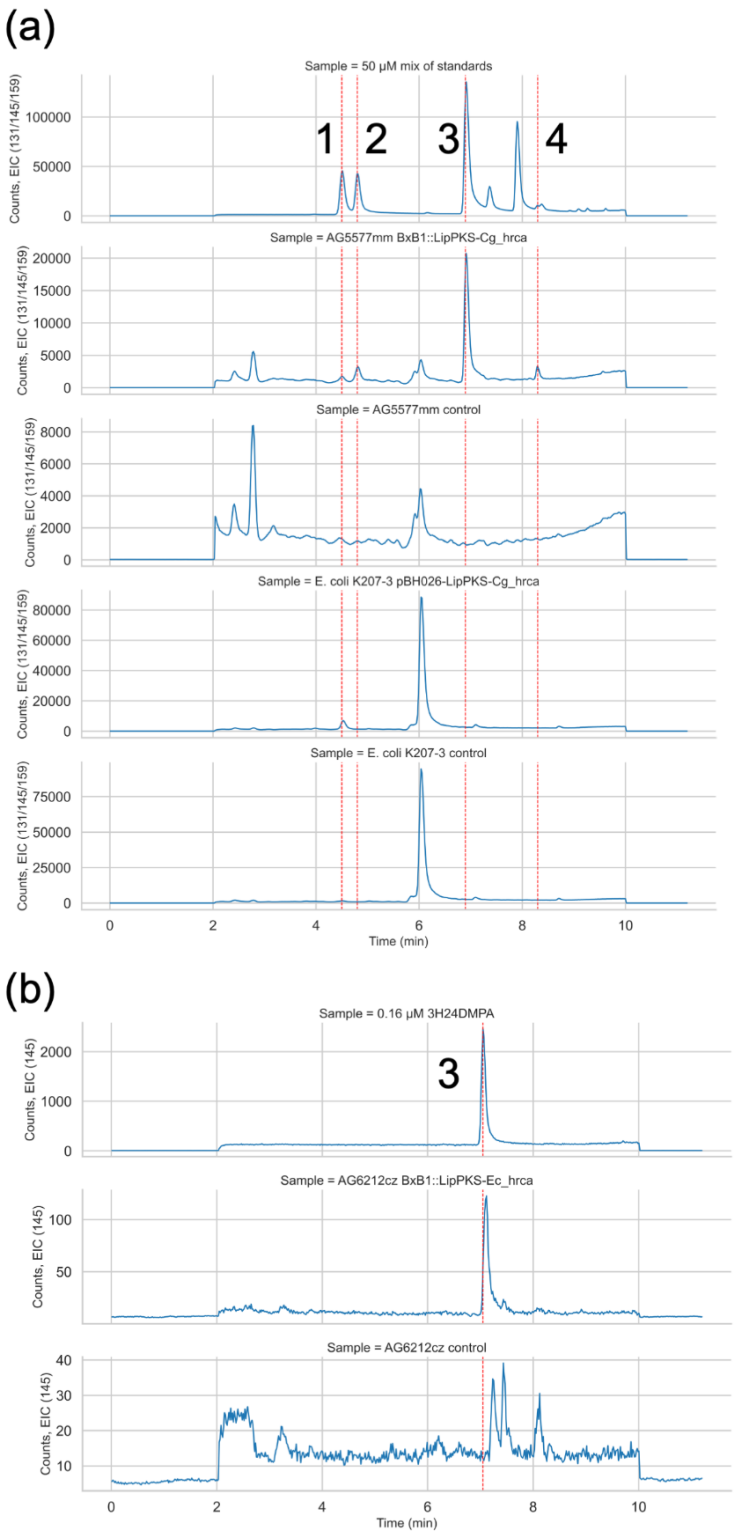
(a)
Sample = 50 µM mix of standards
1
2
3
4
Counts, EIC (131/145/159)
Sample = AG5577mm BxB1::LipPKS-Cg_hrca
Sample = AG5577mm control
Sample = E. coli K207-3 pBH026-LipPKS-Cg_hrca
Sample = E. coli K207-3 control
Time (min)
(b)
Sample = 0.16 µM 3H24DMPA
3
Counts, EIC (145)
Sample = AG6212cz BxB1::LipPKS-Ec_hrca
Sample = AG6212cz control
Time (min)

Figure S3: LC-MS chromatograms showing the production of unnatural polyketides in *C. glutamicum* (a), *E. coli* and *P. putida* (b). Red lines indicate the retention time of the authentic standard. Additional peaks for the authentic standards are most likely caused by racemic mixtures of the 3-hydroxy acids. Standards for (2S,3S)-3-hydroxy-2,4-dimethylpentanoic acid (3H24DMPA) and (2S,3S)-3-hydroxy-2-methylpentanoic acid (3H2MPA) are enantiopure. Standards for 3-hydroxy-4-methylpentanoic acid (3H4MPA) and 3-hydroxy-2,4-dimethylhexanoic acid (3H24DMHA) are racemic mixtures. 1 = 3H2MPA; 2 = 3H4MPA; 3 = 3H24DMPA; 4 = 3H24DMHA

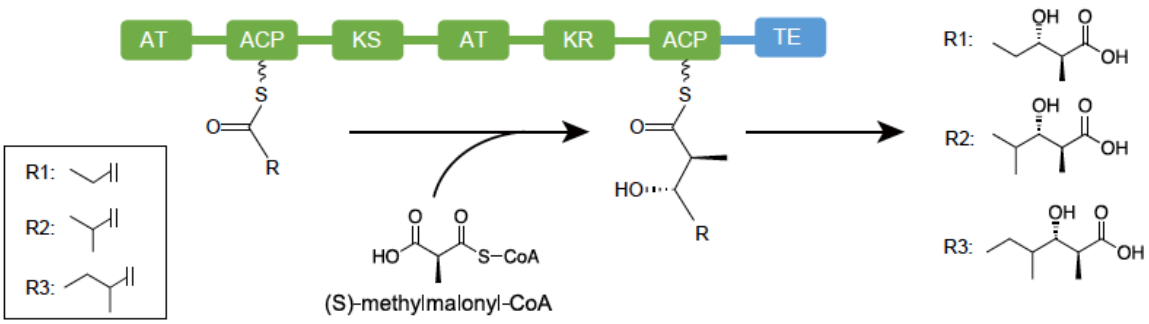

Figure S4: Polyketide synthesis mechanism with various loading substrates and possible products. The final polyketide depends on the availability of the starter unit propionyl-CoA (R1), isobutyryl-CoA (R2) or 2-methylbutyryl-CoA (R3). In the engineered strains *C. glutamicum* AG6212cz and *P. putida* AG5577mm, the most likely product is (2S,3S)-3-hydroxy-2,4-dimethylpentanoic acid (R2). In *E. coli* K207-3, the only possible product is (2S,3S)-3-hydroxy-2-methylpentanoic acid (R1).

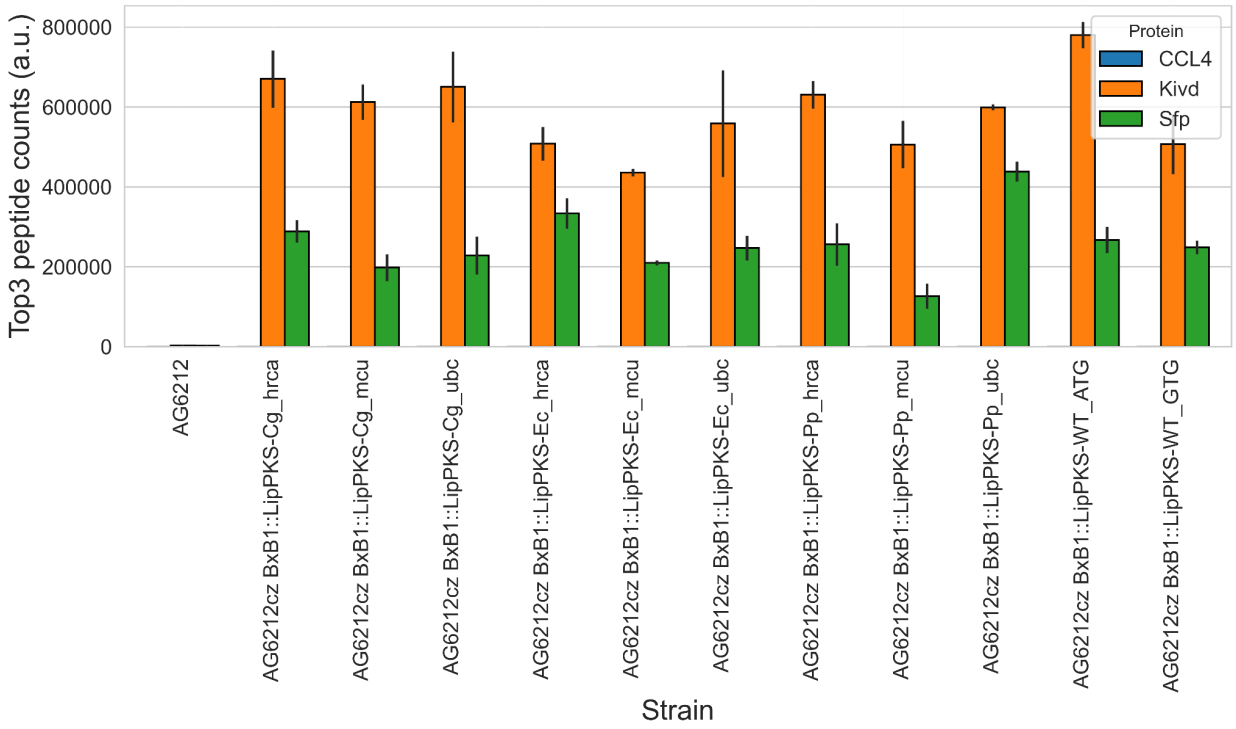

Figure S5: Detection of peptides for supplementary pathways in the *C. glutamicum* ATCC 13032 derivative AG6212cz (n = 3). The corresponding peptides for the phosphopantetheinyl transferase (PPTase) Sfp and ketoisovalerate decarboxylase, Kivd, were detected in all AG6212cz strains. CCL4 peptides could not be detected.

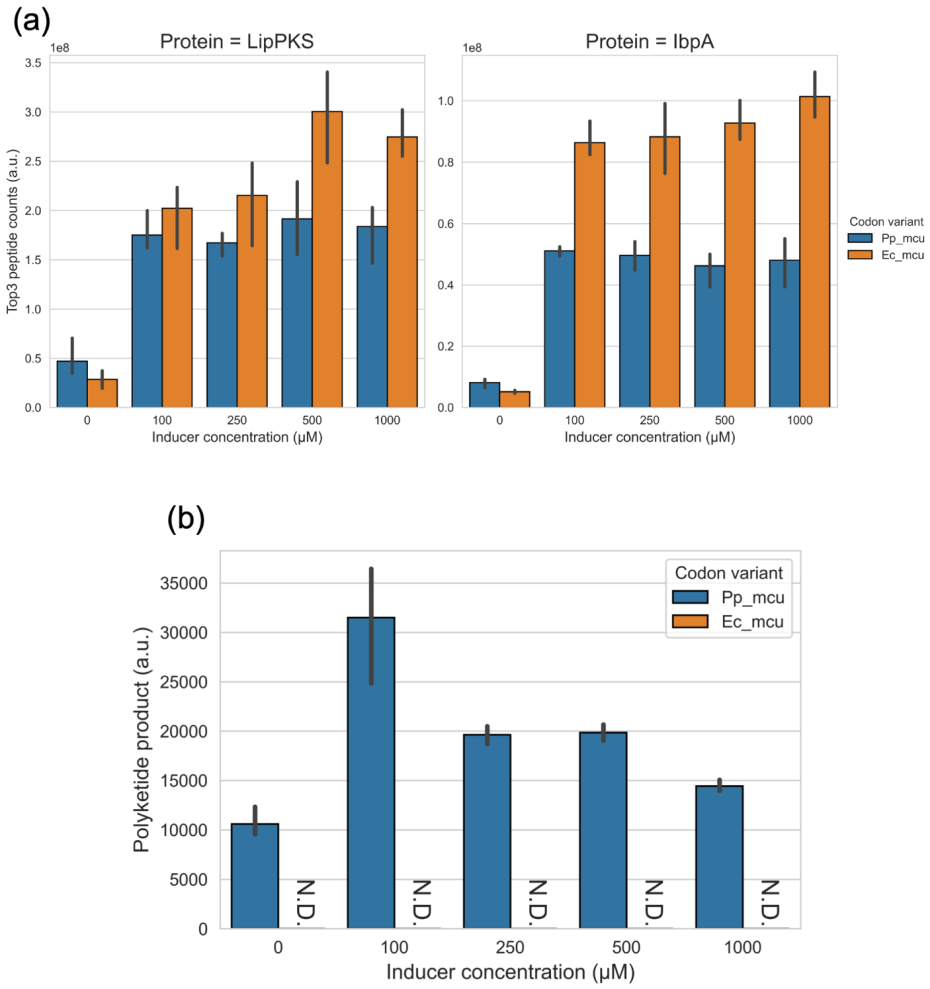

Figure S6: Expression of the LipPKS codon variants Pp_mcu and Ec_mcu from the inducible vector system pGingerBG-NahR in *P. putida* (n = 3). The inducer salicylate was added at the time of inoculation. (a) LipPKS protein levels were comparable, while IbpA levels were significantly higher for Ec_mcu. (b) LC-MS analysis of the same samples (n = 3) revealed no production of the corresponding polyketide for the Ec_mcu variant.