Supplementary Information

Divergent biosynthesis yields a cytotoxic aminomalonate-containing precolibactin

Zhong-Rui Li[1,7], Jie Li[2,7], Jin-Ping Gu[3], Jennifer Y. H. Lai[1], Brendan M. Duggan[4], Wei-Peng Zhang[1], Zhi-Long Li[5], Yong-Xin Li[1], Rong-Biao Tong[5], Ying Xu[6], Dong-Hai Lin[3], Bradley S. Moore[2,4]* & Pei-Yuan Qian[1]*

[1]Division of Life Science and Environmental Science Programs, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, P. R. China. [2]Center for Marine Biotechnology and Biomedicine, Scripps Institution of Oceanography, University of California at San Diego, La Jolla, CA 92093, United States. [3]High-field NMR Research Center, College of Chemistry and Chemical Engineering, Xiamen University, Xiamen 361005, P. R. China. [4]Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California at San Diego, La Jolla, CA 92093, United States. [5]Department of Chemistry, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, P. R. China. [6]Shenzhen Key Laboratory of Marine Bioresource & Ecoenvironmental Science, Shenzhen Engineering Laboratory for Marine Algal Biotechnology, College of Life Sciences and Oceanography, Shenzhen University, Shenzhen 518060, P. R. China. [7]These authors contributed equally to this work. *e-mail: bsmoore@ucsd.edu or boqianpy@ust.hk

Supplementary Results

I. Supplementary Tables

Supplementary Table 1. Plasmids and strains used in this study.

	Description	Source
Strains		
E. coli CFT073	Wild type *E. coli* strain harboring the colibactin (*clb*) gene cluster	1,2
Bacillus subtilis 1779	Wild type *B. subtilis* strain harboring the amicoumacin (*ami*) gene cluster	3
E. coli Top10	Host strain for routine cloning	Invitrogen
E. coli BL21 (DE3)/pLysE	Host strain for protein expression	Novagen
E. coli DH10B	Host for heterologous expression: Δ(*araABC-leu*)*7697*, *araD139*, *deoR*, *endA1*, *galK*, *galU*, Δ(*lac*)*X74mcrA*, Δ(*mcrCB-hsdSMR-mrr*), *nupG*, *recA1*, *rpsL*(Str^r), (φ80 *lacZ*ΔM15)	Invitrogen
E. coli BW25113	K12 derivative, Δ*araBAD*, Δ*rhaBAD*	4
E. coli BW25113/pCAP01-*clb*/pIJ790	*E. coli* BW25113 harboring pCAP01-*clb* and pIJ790, for the further genetic manipulation of pCAP01-*clb*	2
Plasmids		
pET-28a(+)	Protein expression vector, production of N-terminally His-tagged fusion proteins, pBR322 ori, T7 promoter, kan^R	Novagen
pETDuet-1	Protein expression vector, pBR322 ori, T7 promoter, amp^R	Novagen
pIJ790	λ -Red (*gam*, *bet*, *exo*), *cat*, *araC*, *rep101*ts	5
pIJ773	Source of $apra^R$	5
pDR111	Source of amp^R	6
Protein expression plasmids constructed		
pETDuet-1-*amiD*	pETDuet-1, the thioesterase AmiD for *in vivo* complementation	This study
pETDuet-1-*clbQ*	pETDuet-1, the thioesterase ClbQ for *in vivo* complementation	This study
pET-28a-*amiD*	pET-28a(+), recombinant *N*-His_6-AmiD for *in vitro* assay	This study
pET-28a-*clbQ*	pET-28a(+), recombinant *N*-His_6-ClbQ for *in vitro* assay	This study
Plasmids for heterologous expression and genetic manipulation of *clb* pathway		
pCAP01-*clb*	pCAP01 derivative that carries a 70-kb genomic region containing the entire *clb* gene cluster	2
pCAP01-*clb* (Δ*clbP*)	pCAP01-*clb* derivative (Δ*clbP*): $apra^R$	2
pCAP01-*clb* (Δ*clbP*/Δ*clbQ*)	pCAP01-*clb* derivative (Δ*clbP* & Δ*clbQ*): $apra^R$	This study
pCAP01-*clb*	pCAP01-*clb* derivative (Δ*clbP* & Δ*clbQ*): $apra^R$; (Δ*clbB*):	This study

(Δ*clbP*/Δ*clbQ*/Δ*clbB*)	*amp*^{R}	
pCAP01-*clb* (Δ*clbP*/Δ*clbQ*/Δ*clbC*)	pCAP01-*clb* derivative (Δ*clbP* & Δ*clbQ*): *apra*^{R}; (Δ*clbC*): *amp*^{R}	This study
pCAP01-*clb* (Δ*clbP*/Δ*clbQ*/Δ*clbD EF*)	pCAP01-*clb* derivative (Δ*clbP* & Δ*clbQ*): *apra*^{R}; (Δ*clbDEF*): *amp*^{R}	This study
pCAP01-*clb* (Δ*clbP*/Δ*clbQ*/Δ*clbG*)	pCAP01-*clb* derivative (Δ*clbP* & Δ*clbQ*): *apra*^{R}; (Δ*clbG*): *amp*^{R}	This study
pCAP01-*clb* (Δ*clbP*/Δ*clbQ*/Δ*clbH*)	pCAP01-*clb* derivative (Δ*clbP* & Δ*clbQ*): *apra*^{R}; (Δ*clbH*): *amp*^{R}	This study
pCAP01-*clb* (Δ*clbP*/Δ*clbQ*/Δ*clbI*)	pCAP01-*clb* derivative (Δ*clbP* & Δ*clbQ*): *apra*^{R}; (Δ*clbI*): *amp*^{R}	This study
pCAP01-*clb* (Δ*clbP*/Δ*clbQ*/Δ*clbJ*)	pCAP01-*clb* derivative (Δ*clbP* & Δ*clbQ*): *apra*^{R}; (Δ*clbJ*): *amp*^{R}	This study
pCAP01-*clb* (Δ*clbP*/Δ*clbQ*/Δ*clbK*)	pCAP01-*clb* derivative (Δ*clbP* & Δ*clbQ*): *apra*^{R}; (Δ*clbK*): *amp*^{R}	This study
pCAP01-*clb* (Δ*clbP*/Δ*clbQ*/Δ*clbO*)	pCAP01-*clb* derivative (Δ*clbP* & Δ*clbQ*): *apra*^{R}; (Δ*clbO*): *amp*^{R}	This study
***E. coli* CFT073 mutants**		
E. coli CFT073 Δ*clbP*	Δ*clbP*: *apra*^{R}, deletion of *clbP* gene, the *clbP* gene was replaced by an apramycin resistance gene (*apra*^{R}) by λ Red-mediated recombination in *E. coli* CFT073	This study
E. coli CFT073 Δ*clbP*/Δ*clbQ*	Δ*clbP* & Δ*clbQ*: *apra*^{R}, deletion of *clbP* & Δ*clbQ* genes, the *clbP* & Δ*clbQ* genes were replaced by an apramycin resistance gene (*apra*^{R}) by λ Red-mediated recombination in *E. coli* CFT073	This study
***clb*$^+$ heterologous expression strains**		
E. coli DH10B/pCAP01-*clb*	*E. coli* DH10B harboring pCAP01-*clb*	2
E. coli DH10B/pCAP01-*clb* (Δ*clbP*)	*E. coli* DH10B harboring pCAP01-*clb* (Δ*clbP*)	2
E. coli DH10B/pCAP01-*clb* (Δ*clbP*/Δ*clbQ*)	*E. coli* DH10B harboring pCAP01-*clb* (Δ*clbP*/Δ*clbQ*)	This study
E. coli DH10B/pCAP01-*clb* (Δ*clbP*/Δ*clbQ*/Δ*clbB*)	*E. coli* DH10B harboring pCAP01-*clb* (Δ*clbP*/Δ*clbQ*/Δ*clbB*)	This study
E. coli DH10B/pCAP01-*clb* (Δ*clbP*/Δ*clbQ*/Δ*clbC*)	*E. coli* DH10B harboring pCAP01-*clb* (Δ*clbP*/Δ*clbQ*/Δ*clbC*)	This study
E. coli DH10B/pCAP01-*clb* (Δ*clbP*/Δ*clbQ*/Δ*clbD EF*)	*E. coli* DH10B harboring pCAP01-*clb* (Δ*clbP*/Δ*clbQ*/Δ*clbDEF*)	This study
E. coli	*E. coli* DH10B harboring pCAP01-*clb*	This study

DH10B/pCAP01-*clb* (Δ*clbP*/Δ*clbQ*/Δ*clbG*)	(Δ*clbP*/Δ*clbQ*/Δ*clbG*)	
E. coli DH10B/pCAP01-*clb* (Δ*clbP*/Δ*clbQ*/Δ*clbH*)	*E. coli* DH10B harboring pCAP01-*clb* (Δ*clbP*/Δ*clbQ*/Δ*clbH*)	This study
E. coli DH10B/pCAP01-*clb* (Δ*clbP*/Δ*clbQ*/Δ*clbI*)	*E. coli* DH10B harboring pCAP01-*clb* (Δ*clbP*/Δ*clbQ*/Δ*clbI*)	This study
E. coli DH10B/pCAP01-*clb* (Δ*clbP*/Δ*clbQ*/Δ*clbJ*)	*E. coli* DH10B harboring pCAP01-*clb* (Δ*clbP*/Δ*clbQ*/Δ*clbJ*)	This study
E. coli DH10B/pCAP01-*clb* (Δ*clbP*/Δ*clbQ*/Δ*clbK*)	*E. coli* DH10B harboring pCAP01-*clb* (Δ*clbP*/Δ*clbQ*/Δ*clbK*)	This study
E. coli DH10B/pCAP01-*clb* (Δ*clbP*/Δ*clbQ*/Δ*clbO*)	*E. coli* DH10B harboring pCAP01-*clb* (Δ*clbP*/Δ*clbQ*/Δ*clbO*)	This study
Protein expression strains for Δ*clbQ* complementation assay		
E. coli DH10B/pCAP01-*clb* (Δ*clbP*/Δ*clbQ*)*::*pET Duet-1-*amiD*	Protein expression strain for *in vivo* Δ*clbQ* mutant complementation with thioesterase AmiD	This study
E. coli DH10B/pCAP01-*clb* (Δ*clbP*/Δ*clbQ*)*::*pET Duet-1-*clbQ*	Protein expression strain for *in vivo* Δ*clbQ* mutant complementation with thioesterase ClbQ	This study
Protein expression strains for production of recombinant proteins		
E. coli BL21 (DE3)/pLysE/pET-28 a-*amiD*	Protein expression strain for production of *N*-His_6-AmiD	This study
E. coli BL21 (DE3)/pLysE/pET-28 a-*clbQ*	Protein expression strain for production of *N*-His_6-ClbQ	This study

Supplementary Table 2. Oligonucleotides used in this work. Restriction sites are marked in bold and homologous arms for recombination are underlined.

Primer	Sequence	Description
Primers for gene deletions		
clbB-knock out-F	GTGGACGCTGGTTGTAGGAGGGGAATCGGGATTAACACTGTTCAGTGATGGACACTCCTTATTTGATTTT	Deletion of gene *clbB* from pCAP01-*clb* (*clbP*/*clbQ*).
clbB-knock out-R	TAAGCGCACGTAACTCAATAGGATCACCCAGCACCGTACCGGTGCCGTGCTTGGTCTGACAGTTACCAAT	
clbC-knock out-F	AACGGCATGGAAATCGCCATTATTGGTATGGCGGTCCGTTTCCCGCAGTCGACACTCCTTATTTGATTTT	Deletion of gene *clbC* from pCAP01-*clb* (*clbP*/*clbQ*).
clbC-knock out-R	ACTCCTGCGGCTGTATCGGGATATAAAACGGTGAATGCGCCAGATCCAGCTTGGTCTGACAGTTACCAAT	
clbDEF-knock out-F	TGCAGGAGTAATGGGAACTGGCGTCGCTCATAACATGGCGCAATACGGCAGACACTCCTTATTTGATTTT	Deletion of genes *clbD–F* from pCAP01-*clb* (*clbP*/*clbQ*).
clbDEF-knock out-R	TTTCTTTTTCCCGTTCAGCCGCCAACCGTCGCCATCCTGCTGTAATTCTGTTGGTCTGACAGTTACCAAT	
clbG-knock out-F	GTGGACGCTGGTTGTAGGAGGGGAATCGGGATTAACACTGTTCAGTGATGGACACTCCTTATTTGATTTT	Deletion of gene *clbG* from pCAP01-*clb* (*clbP*/*clbQ*).
clbG-knock out-R	TAAGCGCACGTAACTCAATAGGATCACCCAGCACCGTACCGGTGCCGTGCTTGGTCTGACAGTTACCAAT	
clbH-knock out-F	ACGCGGAGAATCTGTCGCACTGCAACTGCCTTTTTGTTTCGAATTGATTAGACACTCCTTATTTGATTTT	Deletion of gene *clbH* from pCAP01-*clb* (*clbP*/*clbQ*).
clbH-knock out-R	TTGATGGTAGTGAAGCGCAGCAGGTCAACCAACGCCACGTGCTGACCGCATTGGTCTGACAGTTACCAAT	
clbI-knock out-F	ATAGCTATCATTGGGATGGCGGGGCGTTTCCCTCAAGCCGATACGGTACAGACACTCCTTATTTGATTTT	Deletion of gene *clbI* from pCAP01-*clb* (*clbP*/*clbQ*).
clbI-knock out-R	GCTGTTATCGGAAAACGCCCGACAGTGGCCATCGGCGGCGGTGATCCCACTTGGTCTGACAGTTACCAAT	
clbJ-knock out-F	GATCATGTGGCCCGCGCCCTGTTAGCGCTGGGCGTGCAGCATGGCGACCGGACACTCCTTATTTGATTTT	Deletion of gene *clbJ* from pCAP01-*clb* (*clbP*/*clbQ*).
clbJ-knock out-R	AAATAGCTCAGCAATAGGTACCGTAACCTTAAAAATCTCCTCAATACGGCTTGGTCTG	

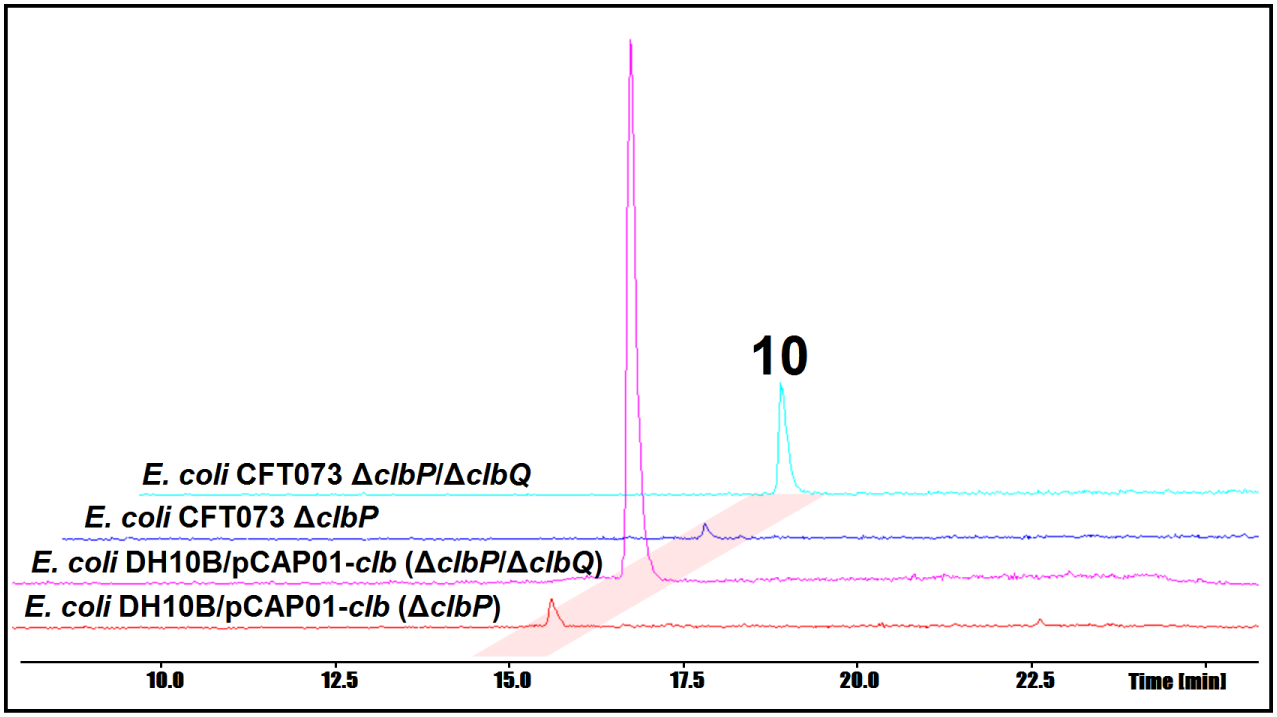
clbK-knock out-F	ACAGTTACCAAT GTACACGGCATTTTACGACTGGGTGCGGT CTATCTGCCAGTGGATCCGGTGACACTCC TTATTTGATTTT	Deletion of gene *clbK* from pCAP01-*clb* (*clbP*/*clbQ*).
clbK-knock out-R	CTGCCCGATAATCGCCTCAAGTGCCTGCT GAATACGCACCAATTCTATAGTTGGTCTG ACAGTTACCAAT	
clbO-knock out-F	TGGCTCACTGGATATTGCCATTATTGGCAT GAGCGGGCGTTTTTCCGGTGGACACTCCT TATTTGATTTT	Deletion of gene *clbO* from pCAP01-*clb* (*clbP*/*clbQ*).
clbO-knock out-R	TGTGCGGCGCACATGCCGGTGCGAAAACC CGGTGCAGAGCTTCAAGCTCATTGGTCTG ACAGTTACCAAT	
clbP-knock out-F	ACACGTTAGCATTAAAACATTATATCATC TCCTGTGCTGTATGCTGCTCTCTCACGTTA AGGGATTTTGG	Deletion of genes *clbP* & *clbQ* in the wild type colibactin producer *E. coli* CFT073. Deletion of genes *clbP* & *clbQ* from *E. coli* DH10B harboring pCAP01-*clb*.
clbP-knock out-R	TCGTTTAATTTGATGATTTAATGTCAGAAC GAAAGCTAACAGGATAATTCGCTCATGAG CTCAGCCAATC	
clbPQ-knock out-F	ACACGTTAGCATTAAAACATTATATCATC TCCTGTGCTGTATGCTGCTCTCTCACGTTA AGGGATTTTGG	Deletion of genes *clbP* & *clbQ* in the wild type colibactin producer *E. coli* CFT073. Deletion of genes *clbP* & *clbQ* from *E. coli* DH10B harboring pCAP01-*clb*.
clbPQ-knock out-R	TCTACCCTACTATTTCGAGTGATTCAATCGT CTGGTTCACATAACCTACCGCTCATGAGCT CAGCCAATC	
Colony PCRs for correct insert check		
clbB-knock out check-F	GCAACGCCGTGTCCACCACGA	Colony PCR for correct insert check (*clbB* knockout).
clbB-knock out check-R	TGCGGCGACCGAGTTGCTCTT	
clbC-knock out check-F	TGGCGCGTCACTATCCGCAAGTG	Colony PCR for correct insert check (*clbC* knockout).
clbC-knock out check-R	TGCGGCGACCGAGTTGCTCTT	
clbDEF-knock out check-F	ACGCACCACCCTTATCAGGCACG	Colony PCR for correct insert check (*clbD–F* knockout).
clbDEF-knock out check-R	TGCGGCGACCGAGTTGCTCTT	
clbG-knock out check-F	GCAACGCCGTGTCCACCACGA	Colony PCR for correct insert check (*clbG* knockout).
clbG-knock out check-R	TGCGGCGACCGAGTTGCTCTT	
clbH-knock out check-F	GCACCTGGTGGCGCAGTGGA	Colony PCR for correct insert check (*clbH* knockout).

clbH-knock out check-R	TGCGGCGACCGAGTTGCTCTT	
clbI-knock out check-F	GCAGCAATACATCGGGCAGCAGTG	Colony PCR for correct insert check (*clbI* knockout).
clbI-knock out check-R	TGCGGCGACCGAGTTGCTCTT	
clbJ-knock out check-F	GATCGAGTTGGCTGGGGAGTTGCA	Colony PCR for correct insert check (*clbJ* knockout).
clbJ-knock out check-R	TGCGGCGACCGAGTTGCTCTT	
clbK-knock out check-F	CTCCTGCACGCCCTAGCCCAG	Colony PCR for correct insert check (*clbK* knockout).
clbK-knock out check-R	TGCGGCGACCGAGTTGCTCTT	
clbO-knock out check-F	TGCGGCATGCACCGGAAGACT	Colony PCR for correct insert check (*clbO* knockout).
clbO-knock out check-R	TGCGGCGACCGAGTTGCTCTT	
clbP-knock out check-F	CGCTGTTGGGCACTCTTTGGCAA	Colony PCR for correct insert check (*clbP* knockout).
clbP-knock out check-R	CGAGTGAGGTGGCAGGGGCAAT	
clbPQ-knock out check-F	CGCTGTTGGGCACTCTTTGGCAA	Colony PCR for correct insert check (*clbPQ* knockout).
clbPQ-knock out check-R	CGAGTGAGGTGGCAGGGGCAAT	
Protein expression for *in vitro* enzymatic assay		
*clbQ*1-BamHI-F	agtgagt**GGATCC**ATGAGTAATATCAGTTTGTATTG	Amplification of gene *clbQ* for protein expression, inserted into the expression vector pET28a, His tag.
*clbQ*1-XhoI-R	agtgagt**CTCGAG**CTACCCTACTATTTCGAGTG	
*amiD*1-HindIII-F	agtgagt**AAGCTT**GCATGATCAAATTATTCTGTCTGCC	Amplification of gene *amiD* for protein expression, inserted into the expression vector pET28a, His tag.
*amiD*1-XhoI-R	agtgagt**CTCGAG**TCATACCACTCCTGTCTGATTC	
Protein expression for *in vivo* complementation assay		
*amiD*2-HindIII-F	agtgagt**AAGCTT**ATGATCAAATTATTCTGTCTGCC	Amplification of gene *amiD* for protein expression, inserted into the expression vector pETDuet-1.
*amiD*2-NotI-R	agtgagt**GCGGCCGC**TCATACCACTCCTGTCTGATTC	
*clbQ*2-BamHI-F	agtgagt**GGATCC**GATGAGTAATATCAGTTTGTATTG	Amplification of gene *clbQ* for protein expression, inserted into the expression vector pETDuet-1.
*clbQ*2-HindIII-R	agtgagt**AAGCTT**CTACCCTACTATTTCGAGTG	

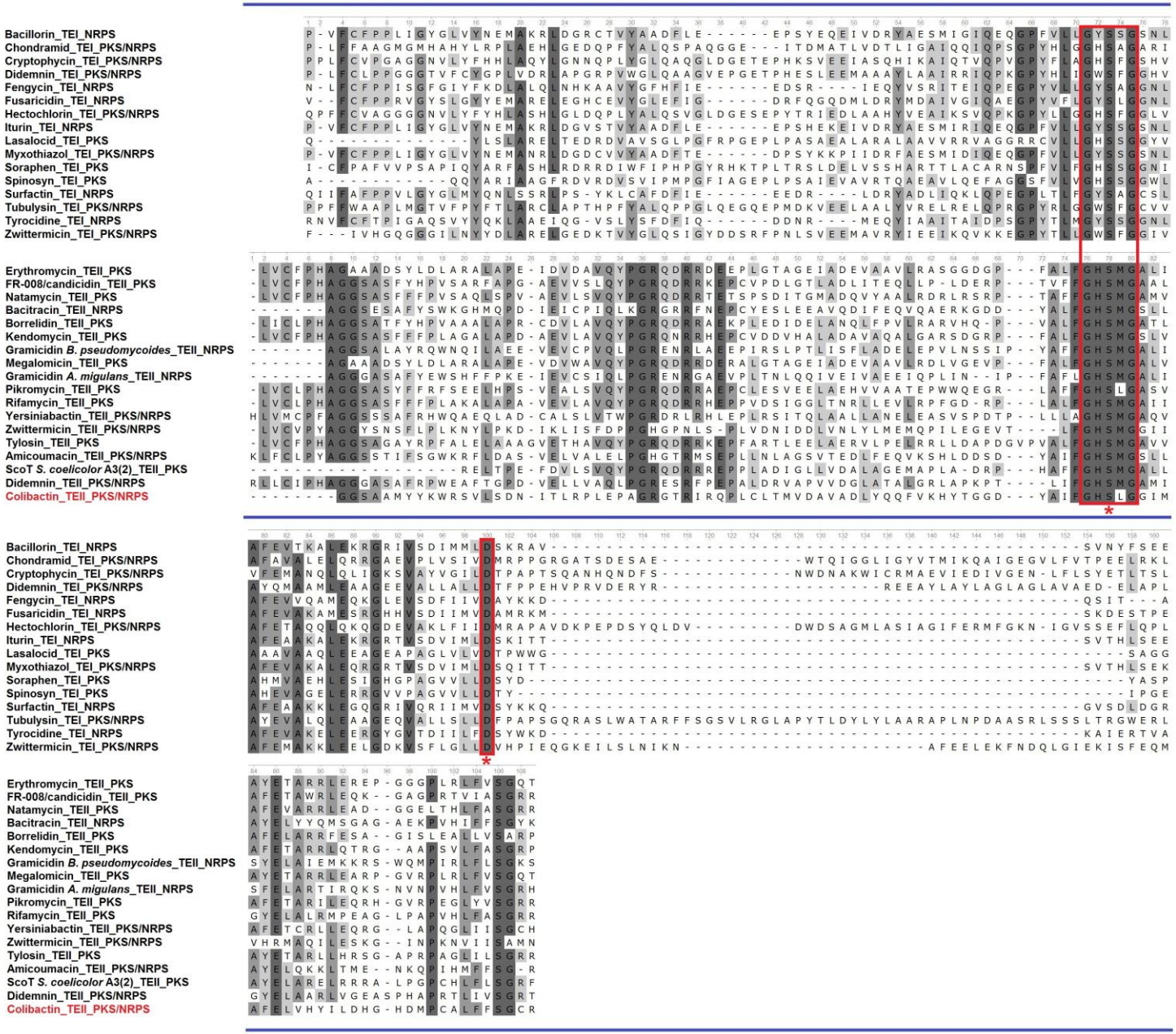
Supplementary Table 3. Accession numbers and descriptions of TEs used for alignment analysis.

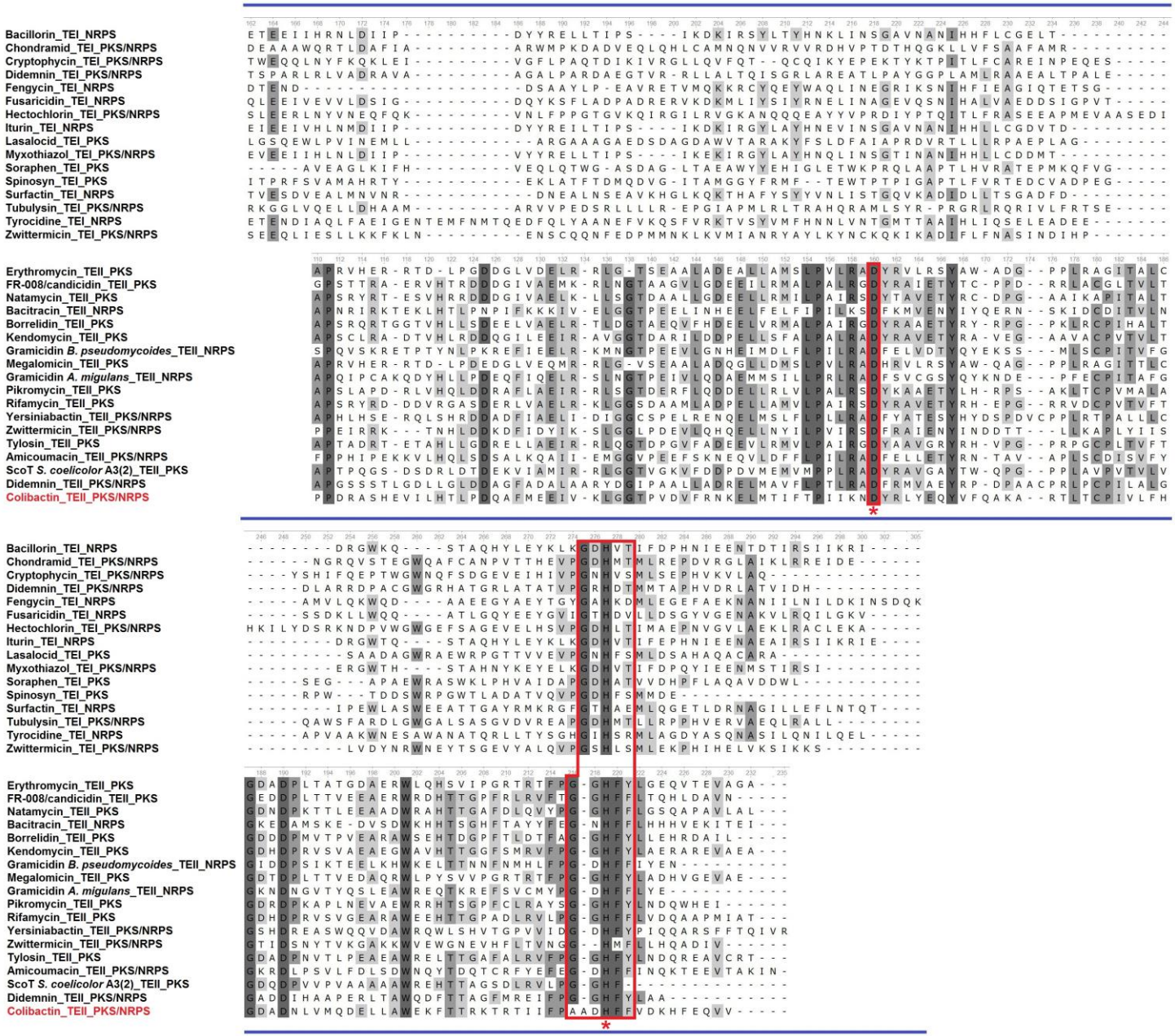
Thioesterase	**Accession number**
Bacillorin_TEI_NRPS	WP_007410142
Chondramid_TEI_PKS/NRPS	Q0VZ70
Cryptophycin_TEI_PKS/NRPS	ABM21572
Didemnin_TEI_PKS/NRPS	WP_014748210
Fengycin_TEI_NRPS	AAB00093
Fusaricidin_TEI_NRPS	ABQ96384
Hectochlorin_TEI_PKS/NRPS	AAY42398
Iturin_TEI_NRPS	ABY89500
Lasalocid_TEI_PKS	BAG85032
Myxothiazol_TEI_PKS/NRPS	AAF19815
Soraphen_TEI_PKS	AAA79984
Spinosyn_TEI_PKS	AAG23262
Surfactin_TEI_NRPS	1JMK_C
Tubulysin_TEI_PKS/NRPS	CAF05651
Tyrocidine_TEI_NRPS	O30409
Zwittermicin_TEI_PKS/NRPS	ACM79812
Erythromycin_TEII_PKS	AAA21345
FR-008/candicidin_TEII_PKS	AAQ82559
Natamycin_TEII_PKS	ADX66462
Bacitracin_TEII_NRPS	WP_020452080
Borrelidin_TEII_PKS	WP_019330222
Kendomycin_TEII_PKS	CAQ52621
Gramicidin *B. pseudomycoides*_TEII_NRPS	WP_006096422
Megalomicin_TEII_PKS	AAG13923
Gramicidin *A. migulans*_TEII_NRPS	P14686
Pikromycin_TEII_PKS	AAC69333
Rifamycin_TEII_PKS	AAG52991
Yersiniabactin_TEII_PKS/NRPS	WP_001551291
Zwittermicin_TEII_PKS/NRPS	ACM79811
Tylosin_TEII_PKS	KDS84464
Amicoumacin_TEII_PKS/NRPS	WP_019257684
ScoT *S. coelicolor* A3(2)_TEII_PKS	AAF43096
Didemnin_TEII_PKS/NRPS	WP_014748194
Colibactin_TEII_PKS/NRPS	AE014075

II. Supplementary Figures

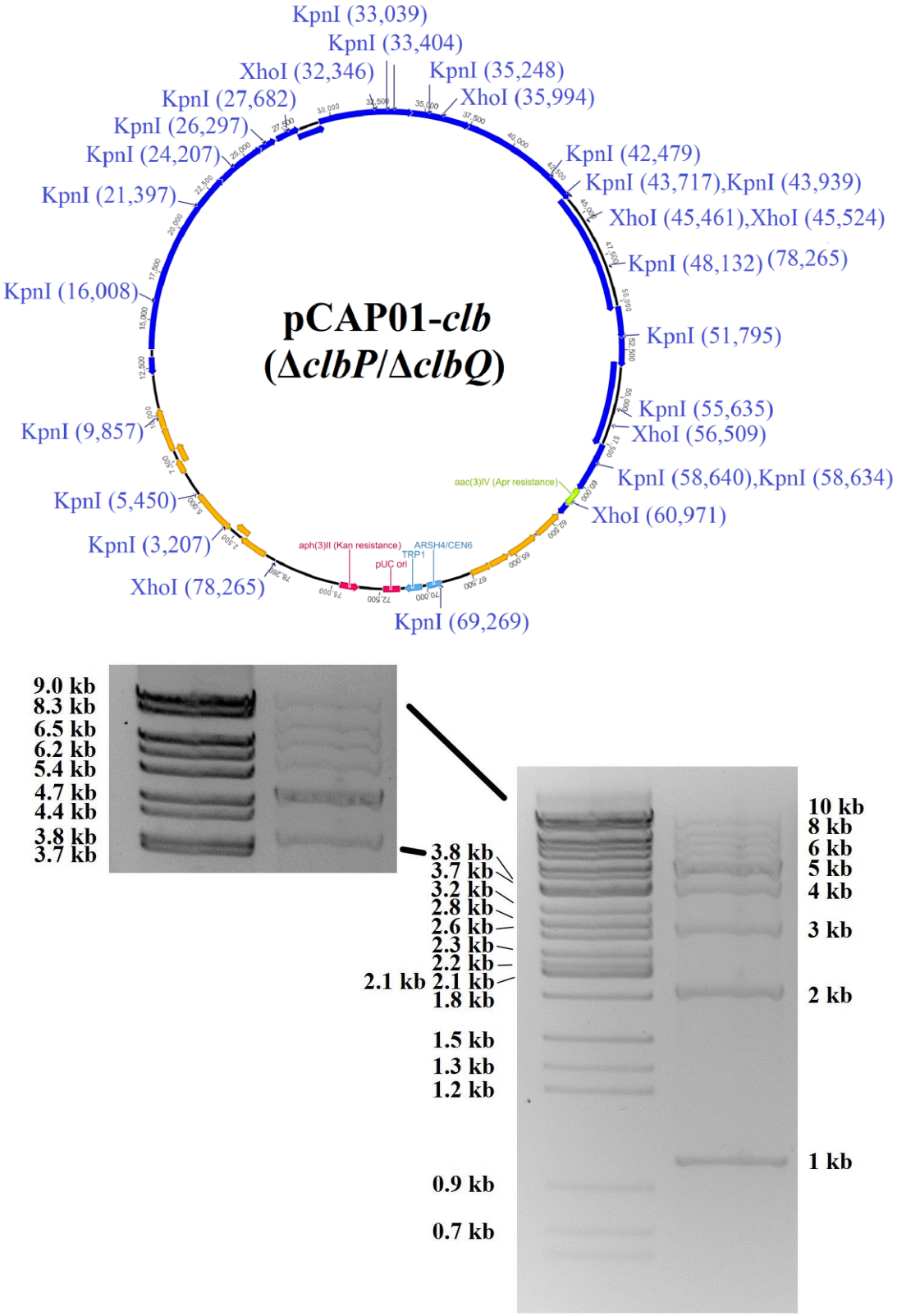


Supplementary Fig. 1. The comparison of UPLC-MS extracted ion chromatogram traces of EtOAc extracts obtained from *E. coli* CFT073 Δ*clbP* and *E. coli* CFT073 Δ*clbP*/Δ*clbQ*; *E. coli* DH10B/pCAP01-*clb* (Δ*clbP*) and *E. coli* DH10B/pCAP01-*clb* (Δ*clbP*/Δ*clbQ*), respectively. EIC+ = 887.38 ± 0.01, corresponding to compound **10** (designated as precolibactin-886 in the present study).

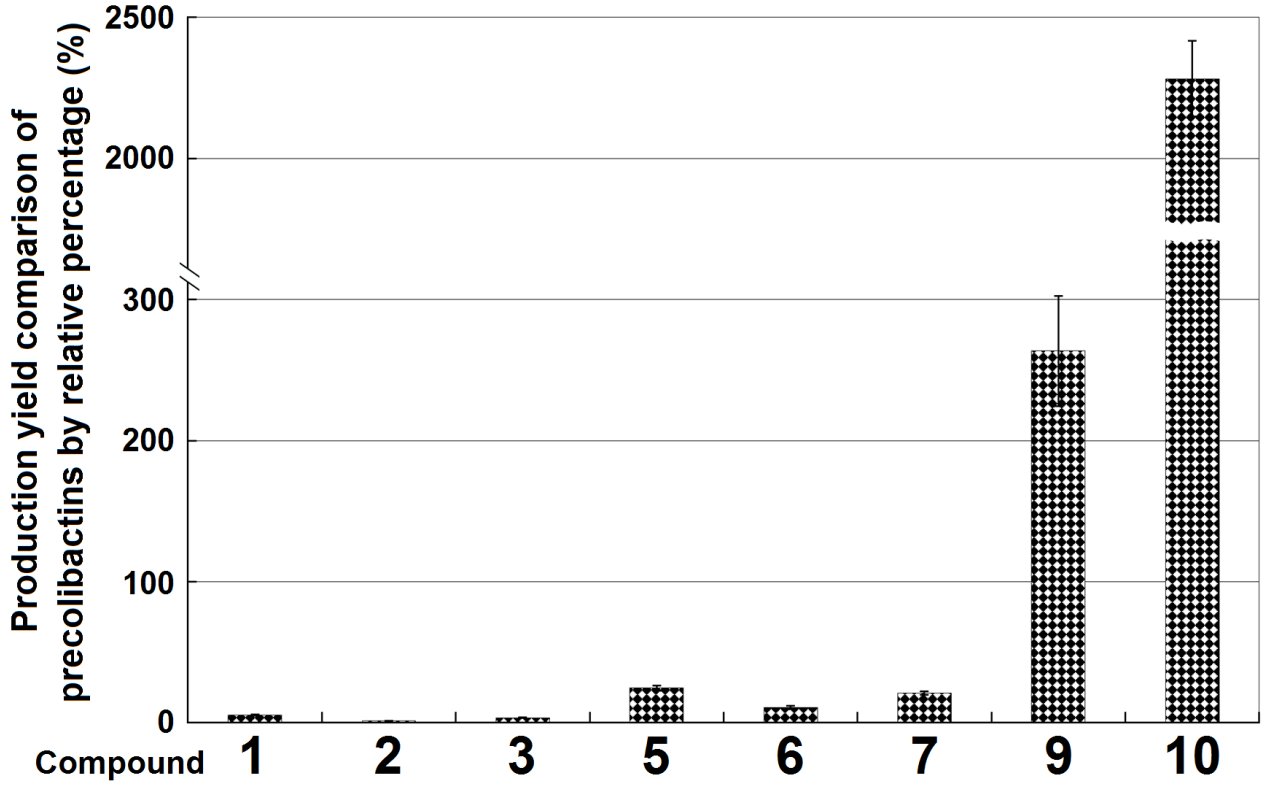

Bacillorin_TEI_NRPS
Chondramid_TEI_PKS/NRPS
Cryptophycin_TEI_PKS/NRPS
Didemnin_TEI_PKS/NRPS
Fengycin_TEI_NRPS
Fusaricidin_TEI_NRPS
Hectochlorin_TEI_PKS/NRPS
Iturin_TEI_NRPS
Lasalocid_TEI_PKS
Myxothiazol_TEI_PKS/NRPS
Soraphen_TEI_PKS
Spinosyn_TEI_PKS
Surfactin_TEI_NRPS
Tubulysin_TEI_PKS/NRPS
Tyrocidine_TEI_NRPS
Zwittermicin_TEI_PKS/NRPS
Erythromycin_TEII_PKS
FR-008/candicidin_TEII_PKS
Natamycin_TEII_PKS
Bacitracin_TEII_NRPS
Borrelidin_TEII_PKS
Kendomycin_TEII_PKS
Gramicidin B. pseudomycoides_TEII_NRPS
Megalomicin_TEII_PKS
Gramicidin A. migulans_TEII_NRPS
Pikromycin_TEII_PKS
Rifamycin_TEII_PKS
Yersiniabactin_TEII_PKS/NRPS
Zwittermicin_TEII_PKS/NRPS
Tylosin_TEII_PKS
Amicoumacin_TEII_PKS/NRPS
ScoT S. coelicolor A3(2)_TEII_PKS
Didemnin_TEII_PKS/NRPS
Colibactin_TEII_PKS/NRPS


Bacillorin_TEI_NRPS
Chondramid_TEI_PKS/NRPS
Cryptophycin_TEI_PKS/NRPS
Didemnin_TEI_PKS/NRPS
Fengycin_TEI_NRPS
Fusaricidin_TEI_NRPS
Hectochlorin_TEI_PKS/NRPS
Iturin_TEI_NRPS
Lasalocid_TEI_PKS
Myxothiazol_TEI_PKS/NRPS
Soraphen_TEI_PKS
Spinosyn_TEI_PKS
Surfactin_TEI_NRPS
Tubulysin_TEI_PKS/NRPS
Tyrocidine_TEI_NRPS
Zwittermicin_TEI_PKS/NRPS
Erythromycin_TEII_PKS
FR-008/candicidin_TEII_PKS
Natamycin_TEII_PKS
Bacitracin_TEII_NRPS
Borrelidin_TEII_PKS
Kendomycin_TEII_PKS
Gramicidin B. pseudomycoides_TEII_NRPS
Megalomicin_TEII_PKS
Gramicidin A. migulans_TEII_NRPS
Pikromycin_TEII_PKS
Rifamycin_TEII_PKS
Yersiniabactin_TEII_PKS/NRPS
Zwittermicin_TEII_PKS/NRPS
Tylosin_TEII_PKS
Amicoumacin_TEII_PKS/NRPS
ScoT S. coelicolor A3(2)_TEII_PKS
Didemnin_TEII_PKS/NRPS
Colibactin_TEII_PKS/NRPS

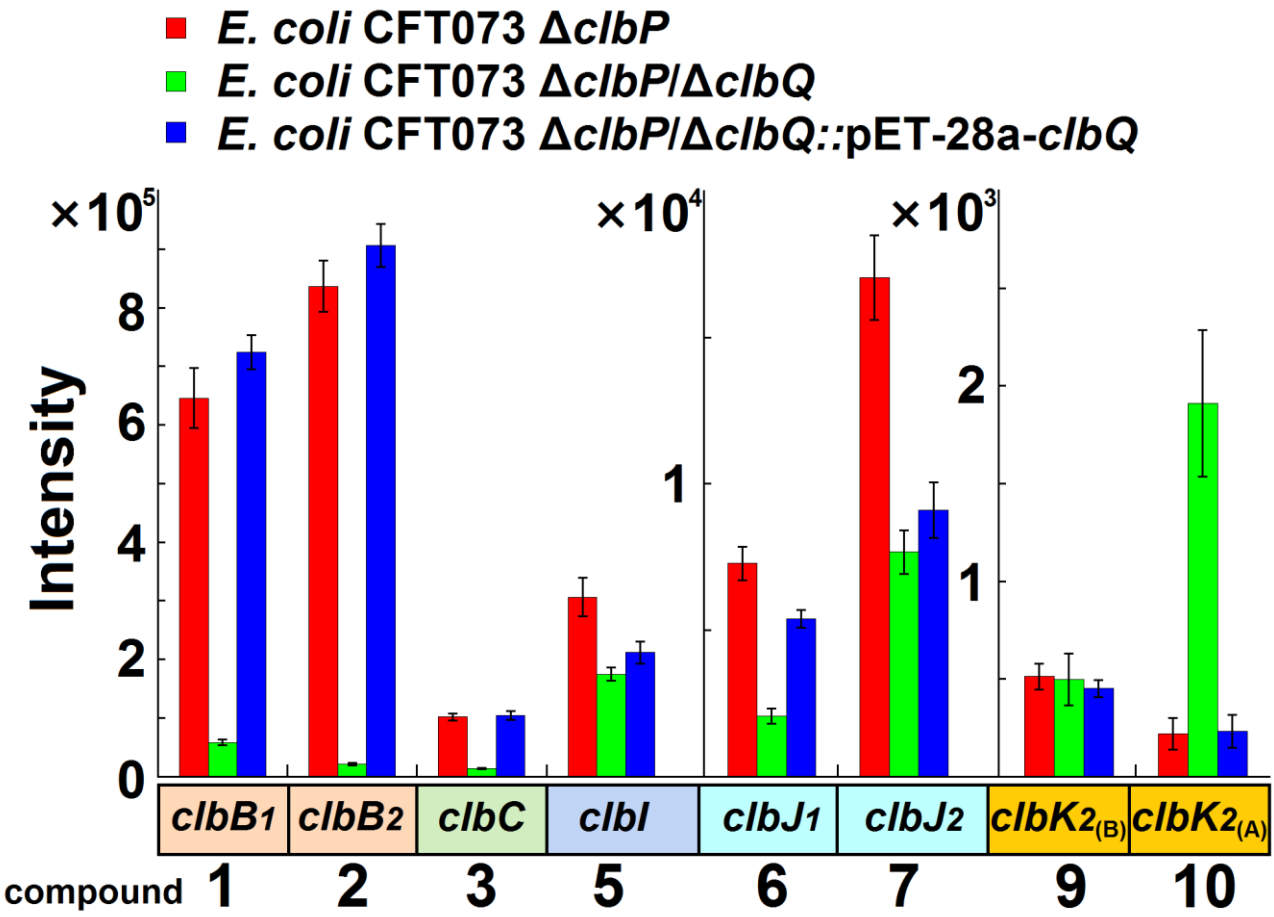
Supplementary Fig. 2. Protein sequence alignment of ClbQ with other representative type I and type II thioesterases, showing that the Ser-His-Asp catalytic triad of type II but not type I TEs is conserved in ClbQ. However, variation of a few amino acid residues adjacent to the catalytic sites were observed and might portend an unusual catalytic function. The signature GXSXG motif containing the strictly conserved active site Ser and the GXHXX motif containing the catalytic His residue are in red frames. The characteristic Ser-His-Asp catalytic triad of thioesterase or thioesterase domain is highlighted by asterisks.



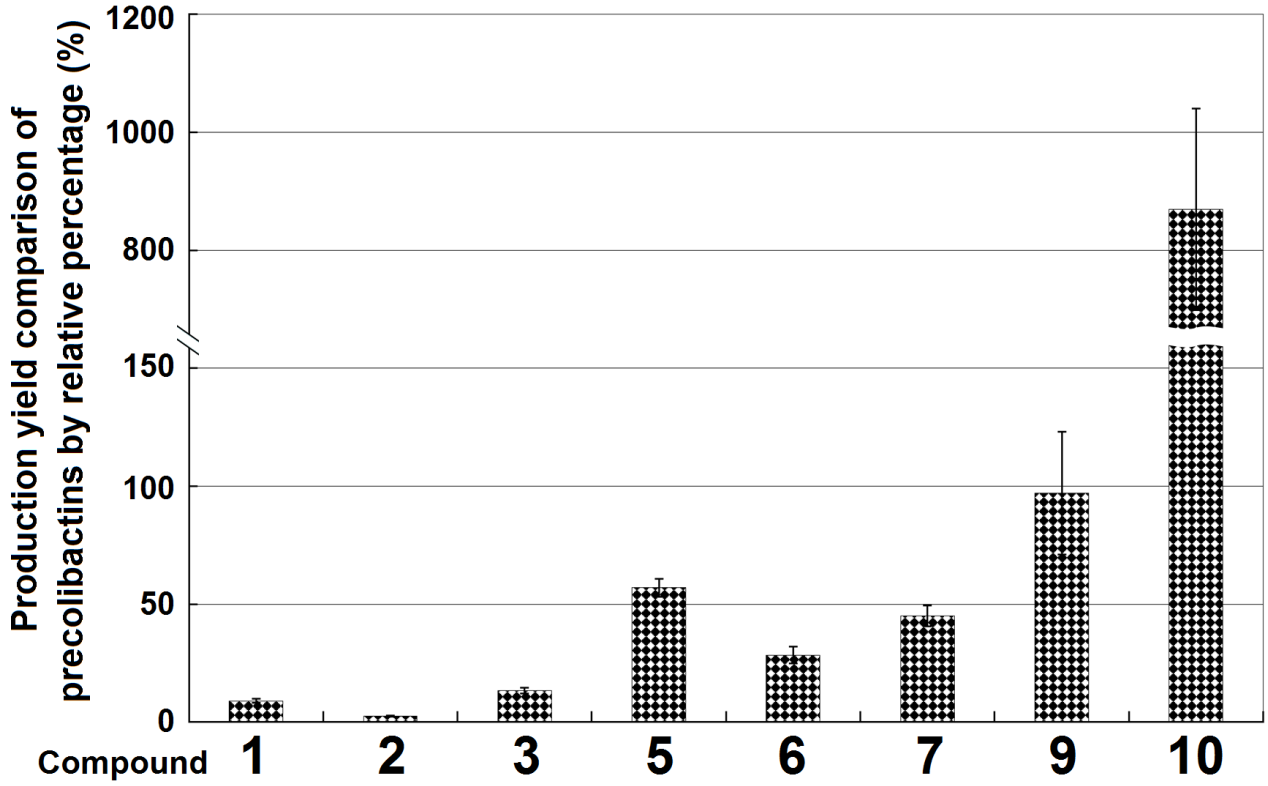
Supplementary Fig. 3. Physical map of bacterial artificial chromosome pCAP01-*clb* (Δ*clbP*/Δ*clbQ*) with predicted KpnI + XhoI cleavage sites and restriction fragment sizes (upper) and the experimentally determined restriction map with KpnI + XhoI (lower).



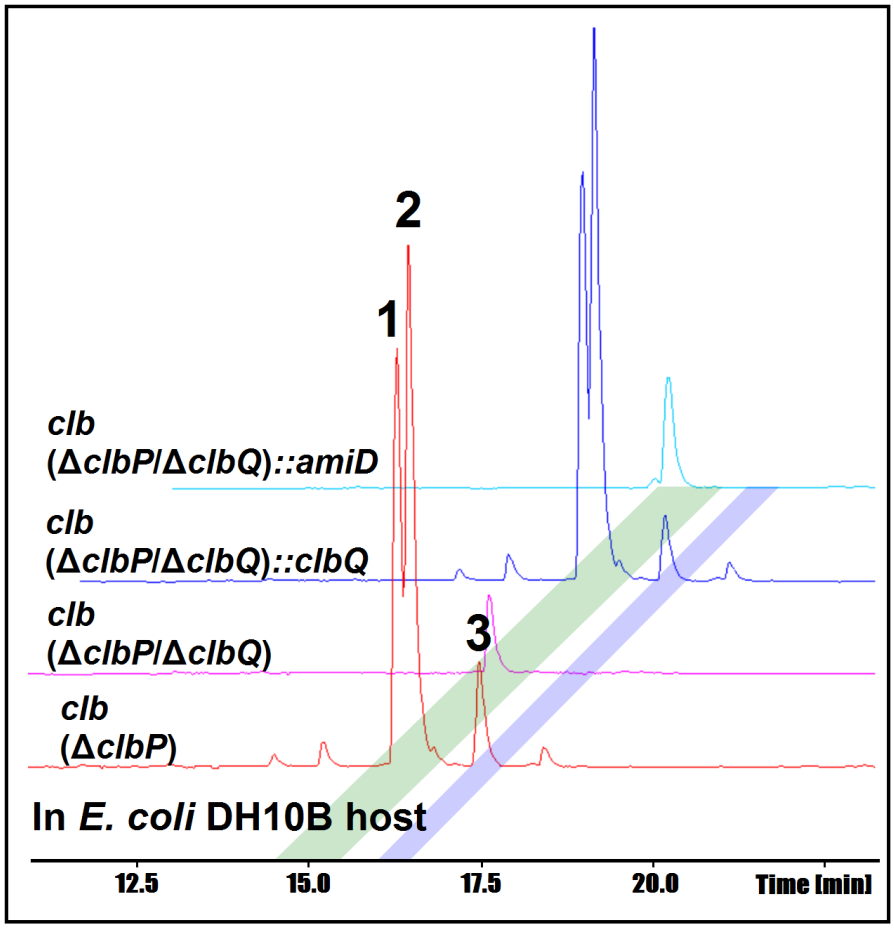
Supplementary Fig. 4. The abundance of individual *clb* pathway-related compound in *E. coli* DH10B/pCAP01-*clb* (Δ*clbP*/Δ*clbQ*) relative to *E. coli* DH10B/pCAP01-*clb* (Δ*clbP*) (in percentage, %). All values are mean ± SD; $N = 5$.



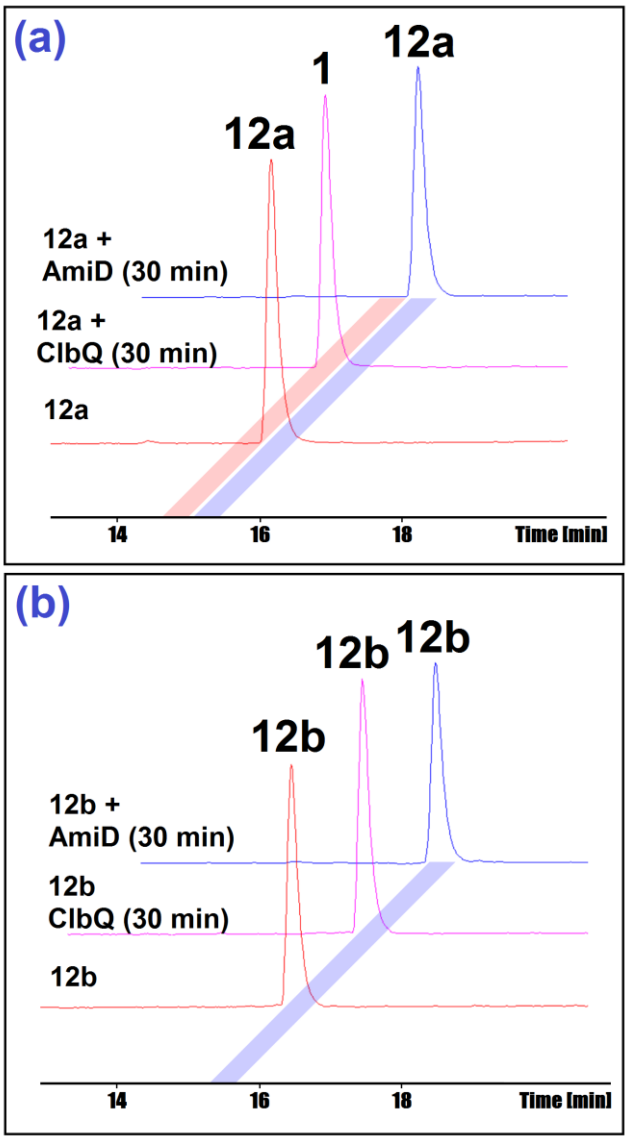
Supplementary Fig. 5. The comparison of the abundance of individual *clb* pathway-related compound from the extracts of colibactin native producer mutants *E. coli* CFT073 Δ*clbP*, *E. coli* CFT073 Δ*clbP*/Δ*clbQ* and *E. coli* CFT073 Δ*clbP*/Δ*clbQ::*pET-28a-*clbQ*, respectively. The extracted ion chromatograms corresponding to the same compound from different mutants were compared. All values are mean ± SD; N = 5. This result showed the same trend as observed in the corresponding *clb*$^+$ heterologous expression host mutants (see **Figure 2** in the main text).



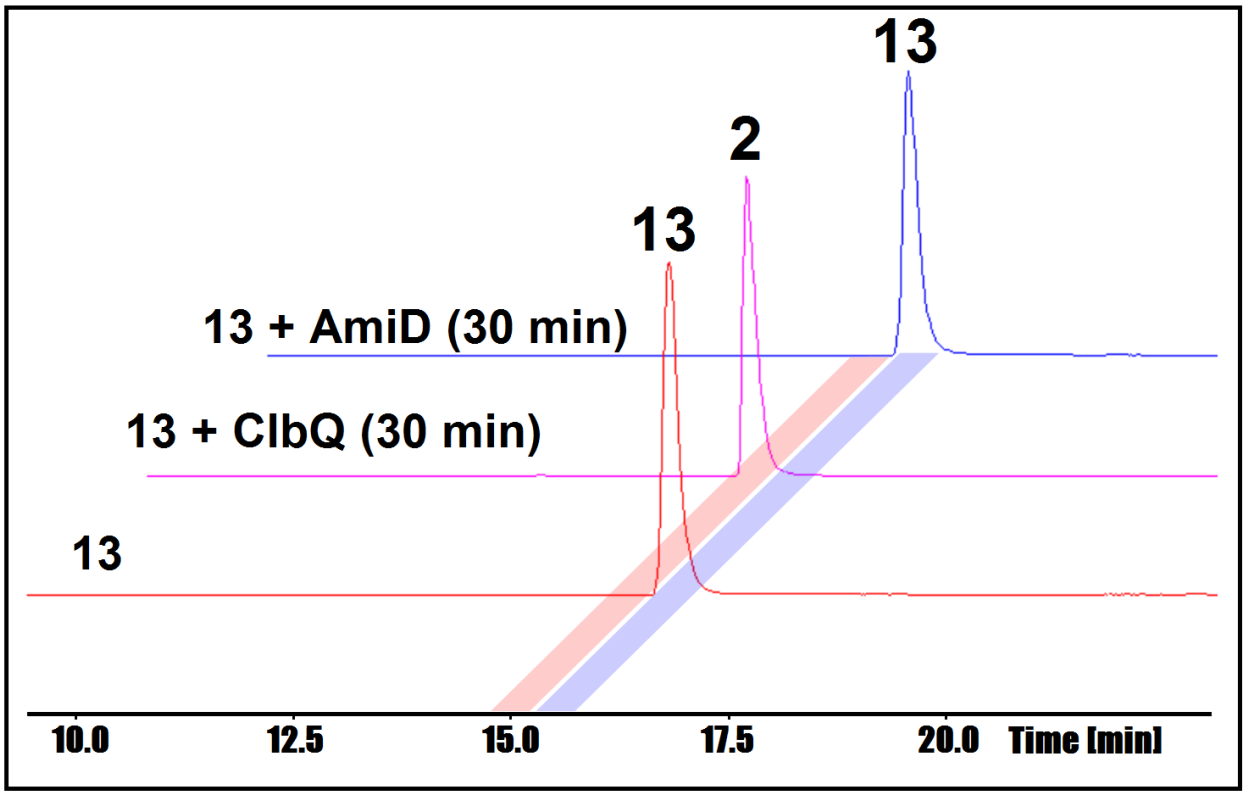
Supplementary Fig. 6. The abundance of individual *clb* pathway-related compound in *E. coli* CFT073 Δ*clbP*/Δ*clbQ* relative to *E. coli* CFT073 Δ*clbP* (in percentage, %). All values are mean ± SD; $N = 5$.



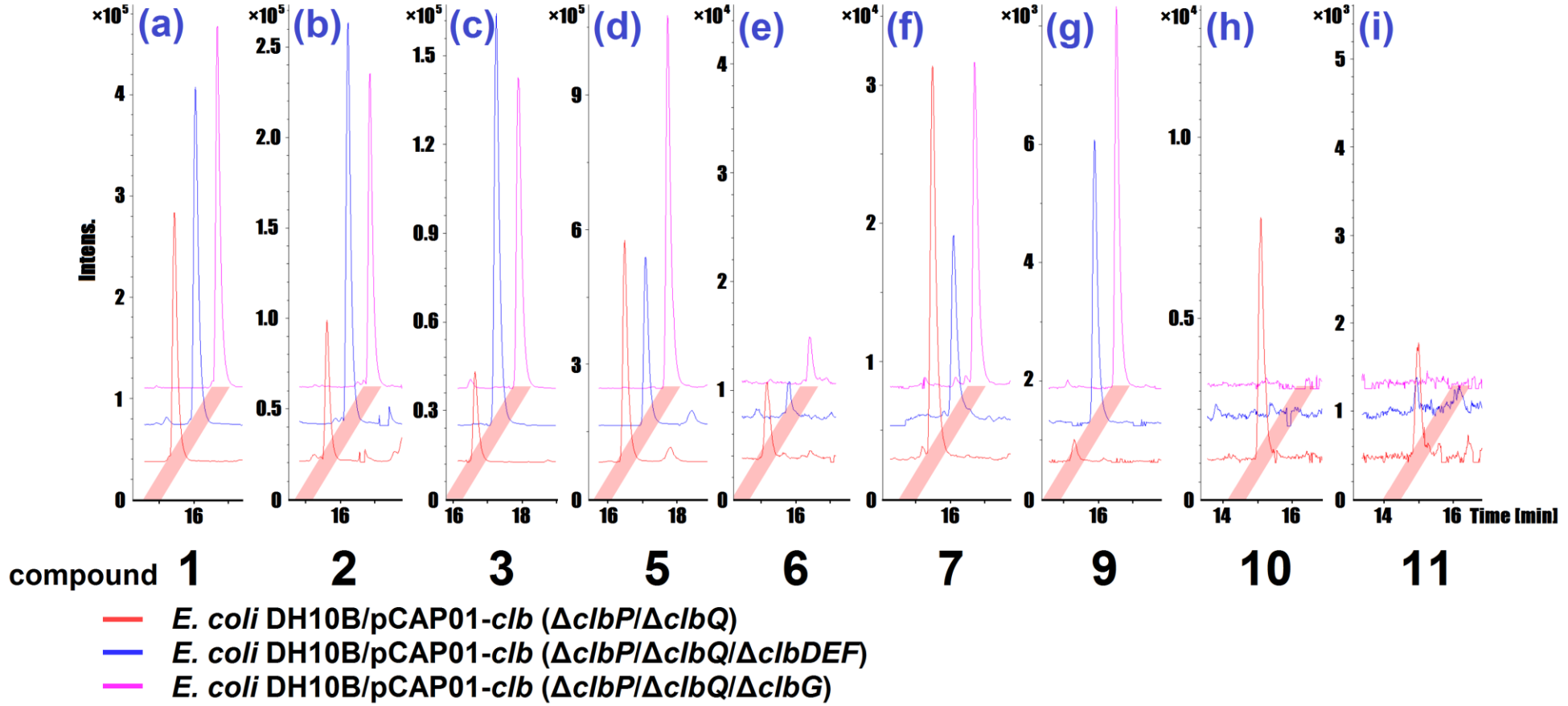
Supplementary Fig. 7. UPLC-MS extracted ion chromatogram traces displaying the production of the *clb* pathway-related compounds in the *clb*$^+$ *E. coli* DH10B Δ*clbP*/Δ*clbQ* mutant and the complementation strains with different thioesterase genes. Extracted ion chromatograms (EIC+ = 414.30 ± 0.01, 442.33 ± 0.01 and 440.35 ± 0.01, corresponding to compounds **1**, **2** and **3**, respectively) of the extracts obtained from the *E. coli* DH10B carrying either pCAP01-*clb* (Δ*clbP*) or pCAP01-*clb* (Δ*clbP*/Δ*clbQ*) were first compared. Then, a protein expression plasmid pETDuet-1 carrying ClbQ or AmiD was transformed into *E. coli* DH10B/pCAP01-*clb* (Δ*clbP*/Δ*clbQ*) by electroporation. Their chemical profiles were compared, which showed that the unique activity of ClbQ in mediating intermediates off-loading could not be complemented by AmiD, the amicoumacin thioesterase close to ClbQ in phylogenetic analysis.



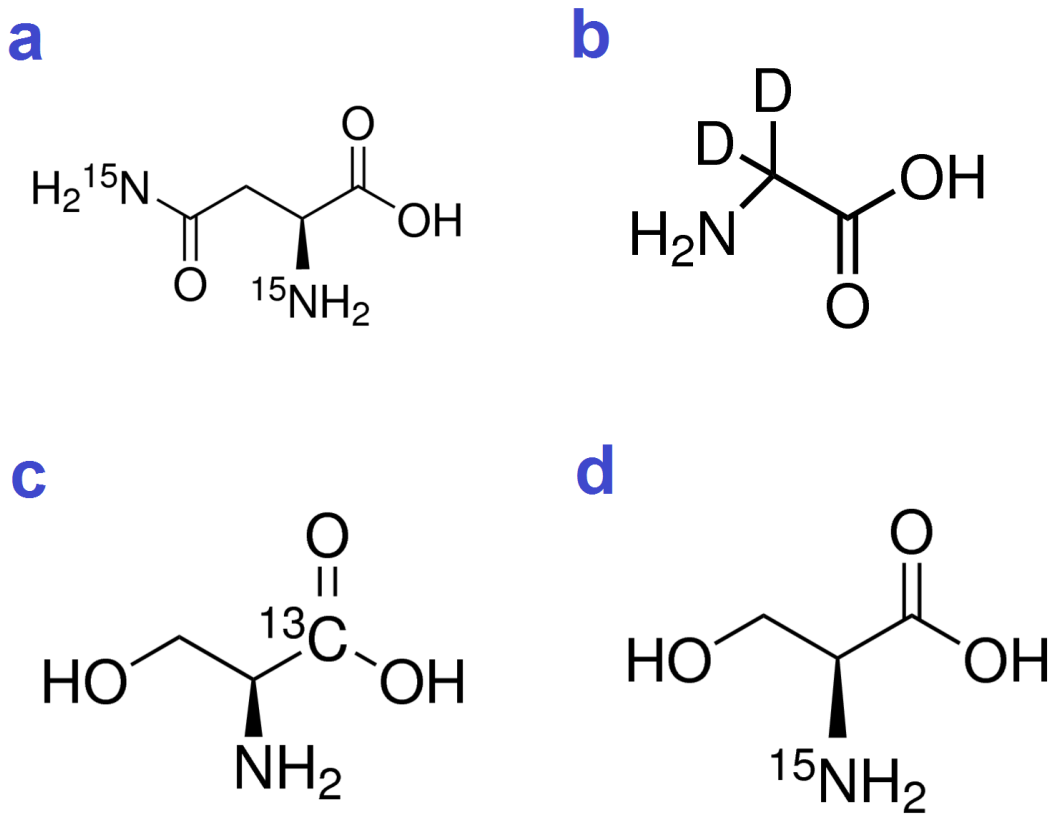
Supplementary Fig. 8. UPLC-MS extracted ion chromatogram traces displaying a complete hydrolysis of **12a** into **1** catalyzed by recombinant ClbQ, but not by recombinant AmiD, after an incubation at 30 °C for 30 min (a). Neither ClbQ nor AmiD could hydrolyze **12b** (b). These results showed that ClbQ is a highly selective thioesterase towards the SNAC derivatives of natural precolibactins. EIC+ = 414.30 ± 0.01 and 515.33 ± 0.01, corresponding to precolibactin-413 (**1**) and its two SNAC thioester epimers (**12a** and **12b**).



Supplementary Fig. 9. UPLC-MS extracted ion chromatogram traces showing a complete hydrolysis of **13** into **2** by recombinant ClbQ, but not by recombinant AmiD, after an incubation at 30 °C for 30 min. EIC+ = 442.33 ± 0.01 and 543.36 ± 0.01, corresponding to the precolibactin-441 (**2**) and its SNAC thioester (**13**).

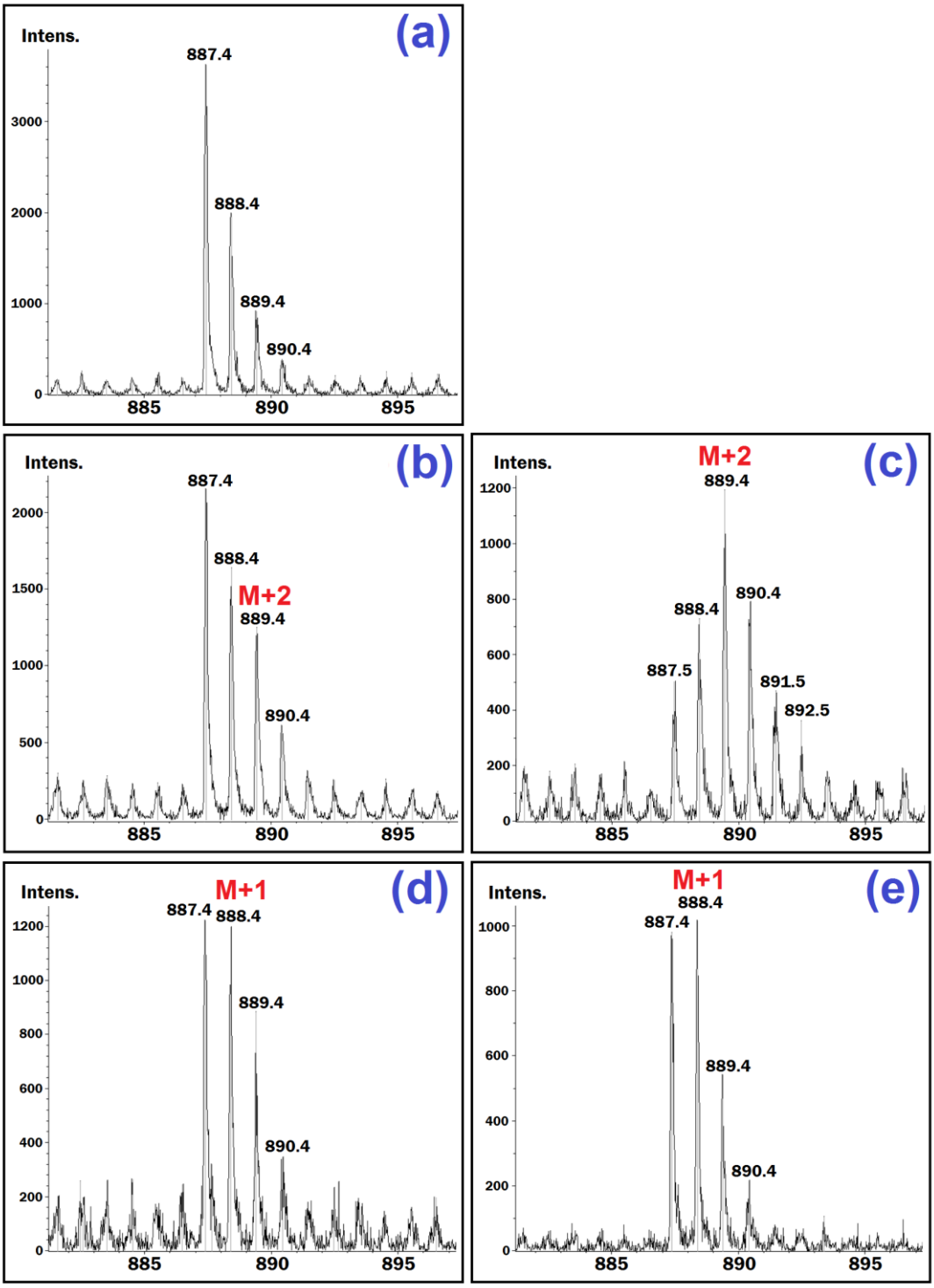


Supplementary Fig. 10. The comparison of UPLC-MS extracted ion chromatogram traces of the EtOAc extracts of *E. coli* DH10B/pCAP01-*clb* and its Δ*clbDEF* and Δ*clbG* mutants. EIC+ = 414.30 ± 0.01, 442.33 ± 0.01, 440.35 ± 0.01, 547.39 ± 0.01, 630.39 ± 0.01, 713.37 ± 0.01, 796.35 ± 0.01, 887.38 ± 0.01 and 970.37 ± 0.01, corresponding to compounds **1**, **2**, **3**, **5**, **6**, **7**, **9**, **10** and **11**. This systematic gene disruption result showed that the aminomalonyl-ACP biosynthetic gene cassette *clbDEF* and the *trans*-AT encoding gene *clbG* were involved in the biosynthesis of **10** and **11**.

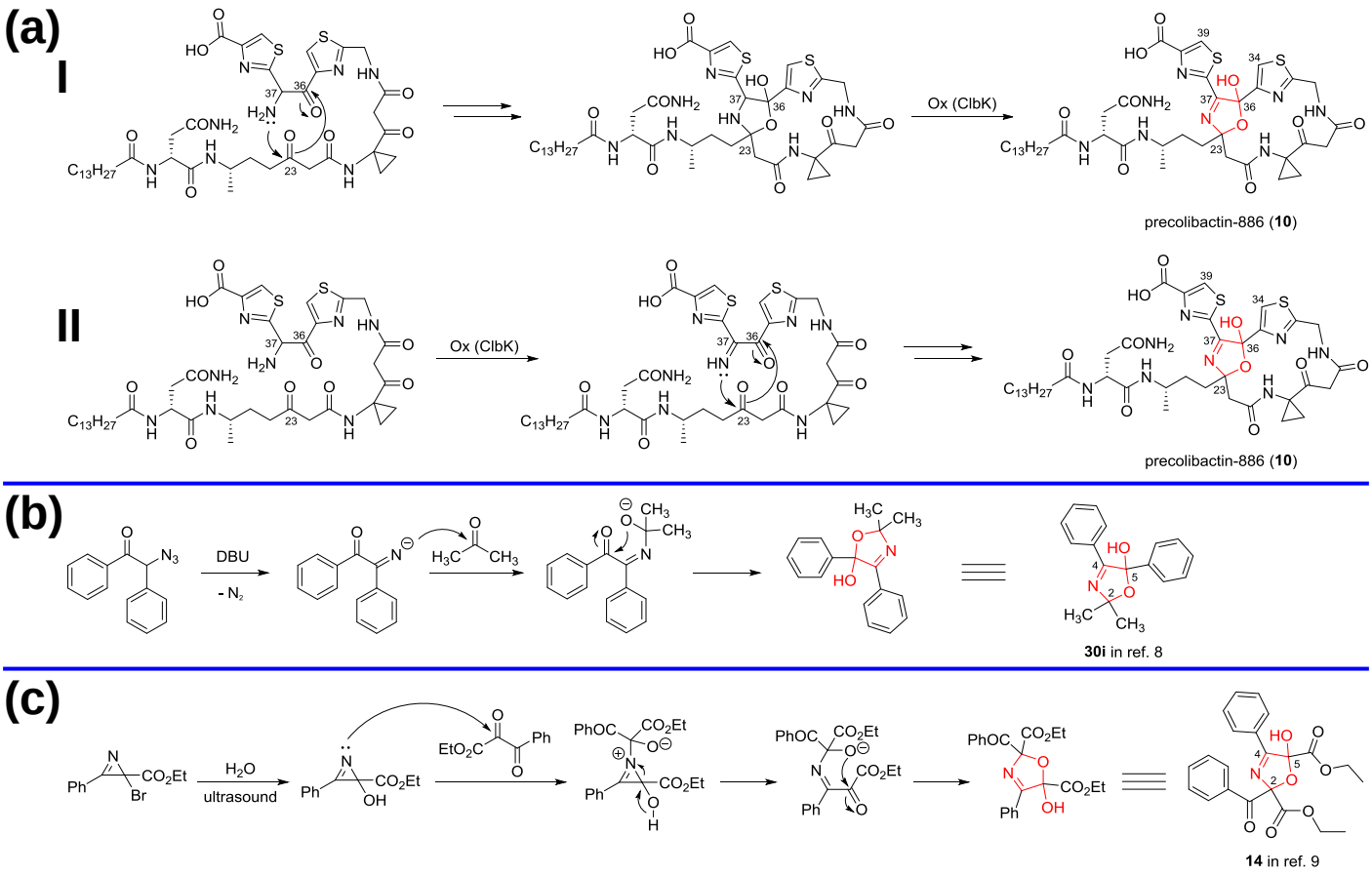


Supplementary Fig. 11. The structures of isotope-labelled amino acids used in the present study. (a) L-[$^{15}N_2$]asparagine; (b) [2,2-D_2]glycine; (c) L-[1-^{13}C]serine; (d) L-[^{15}N]serine. Structures were referred to the Official Website of Sigma-Aldrich Co. LLC.

http://www.sigmaaldrich.com



Supplementary Fig. 12. UPLC-MS analysis of the feeding experiment. L-[$^{15}N_2$]asparagine (b), [2,2-D_2]glycine (c), L-[1-^{13}C]serine (d) or L-[^{15}N]serine (e) at a concentration of 0.5 mg/mL was added into the Luria-Bertani media for culturing the heterologous expression host *E. coli* DH10B harboring pCAP01-*clb* (Δ*clbP*/Δ*clbQ*). Compared to the control without isotope feeding (a), the isotope-labeled versions of compound **10** were observed, displaying mass shifts of [M + 2] (b), [M + 2] (c), [M + 1] (d) and [M + 1] (e), respectively.

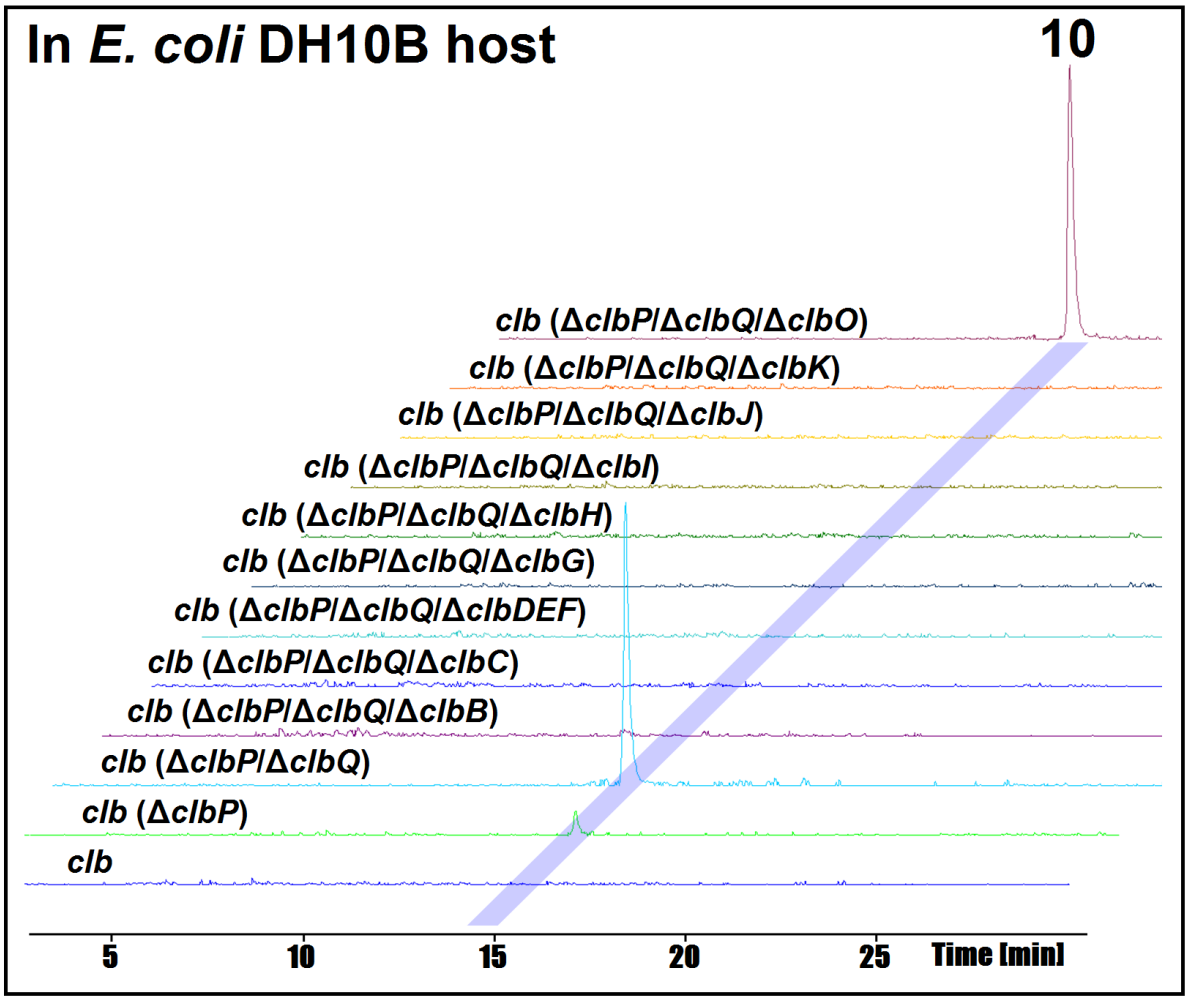


(d)

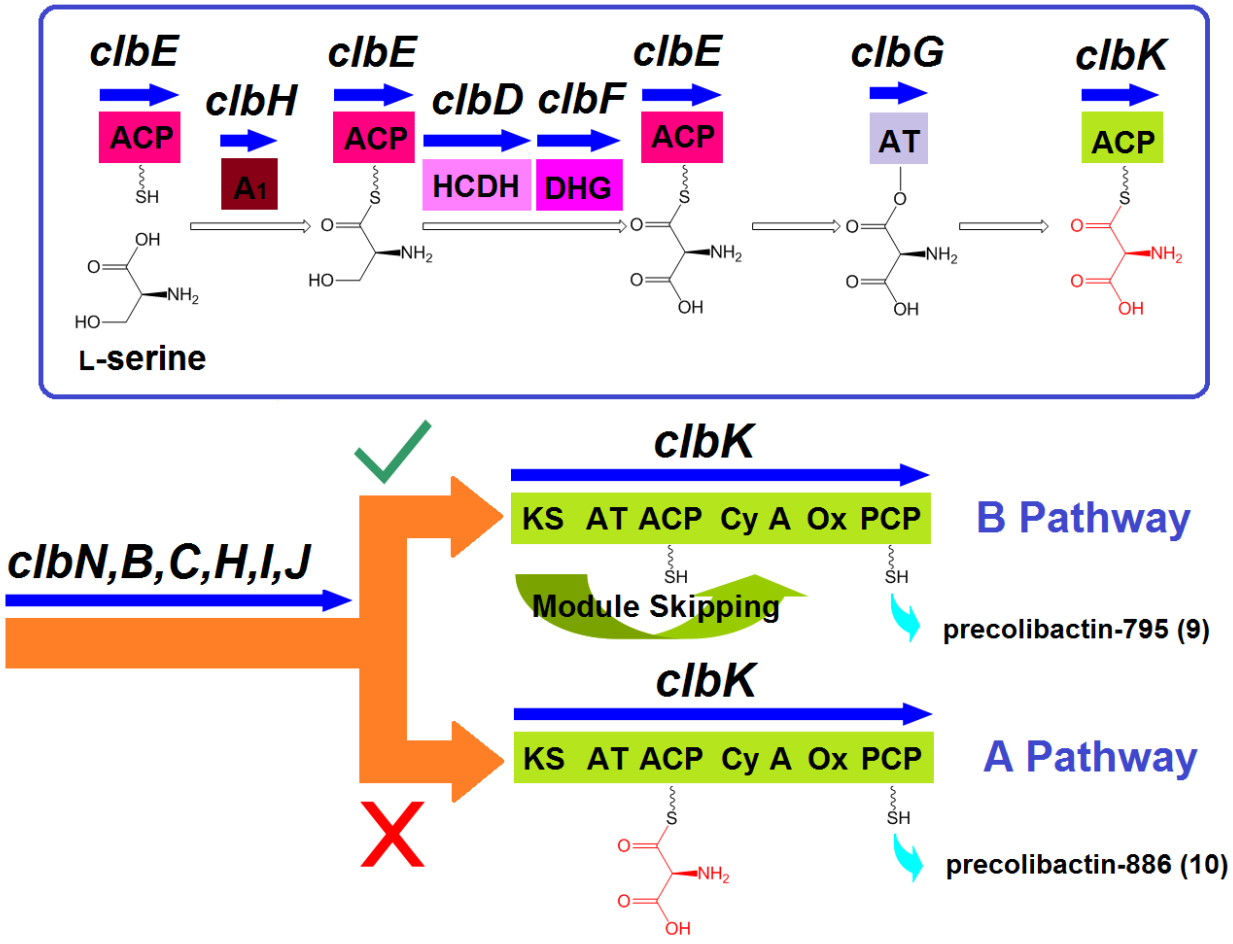
10 in this study	chemical shift (in DMSO-d_6)	30i in ref. 8	chemical shift (in $CDCl_3$)	14 in ref. 9	chemical shift (in $CDCl_3$)
C-23	107.8	C-2	107.3	C-2	105.5
C-36	107.6	C-5	108.5	C-5	109.9
C-37	160.4	C-4	165.4	C-4	165.5

Supplementary Fig. 13. Comparison of the generation mechanism and NMR chemical shifts of the unique 2,5-dihydro-5-hydroxyoxazole ring in precolibactin-886 (**10**) and previously synthesized compounds. (a) Proposed biosynthesis of the 2,5-dihydro-5-hydroxyoxazole ring in precolibactin-886 (**10**). It could be envisioned that the precursor ketone at C-23, a chemically reactive site of the linear assembly product and the point of cyclization in other cyclic precolibactins, could undergo a nucleophilic attack by the aminomalonyl-derived nitrogen atom, and the resultant oxoanion at C-23 could further attack another ketone at C-36 to form a unique 2,5-dihydro-5-hydroxyoxazole ring. The double bond between C-37 and N-37 is speculated to be introduced by the oxidation domain of ClbK that has been previously proposed to install both the Δ^{34} and Δ^{39} double bonds of the two thiazole rings[7]. The timing of the introduction of this unsaturation is currently not clear, could be either prior to (as in II) or after (as in I) the formation of the oxazolidin-5-ol heterocycle. This proposed biosynthesis of the 2,5-dihydro-5-hydroxyoxazole ring in precolibactin-886 (**10**) is consistent with previously reported synthetic strategies for the preparation of

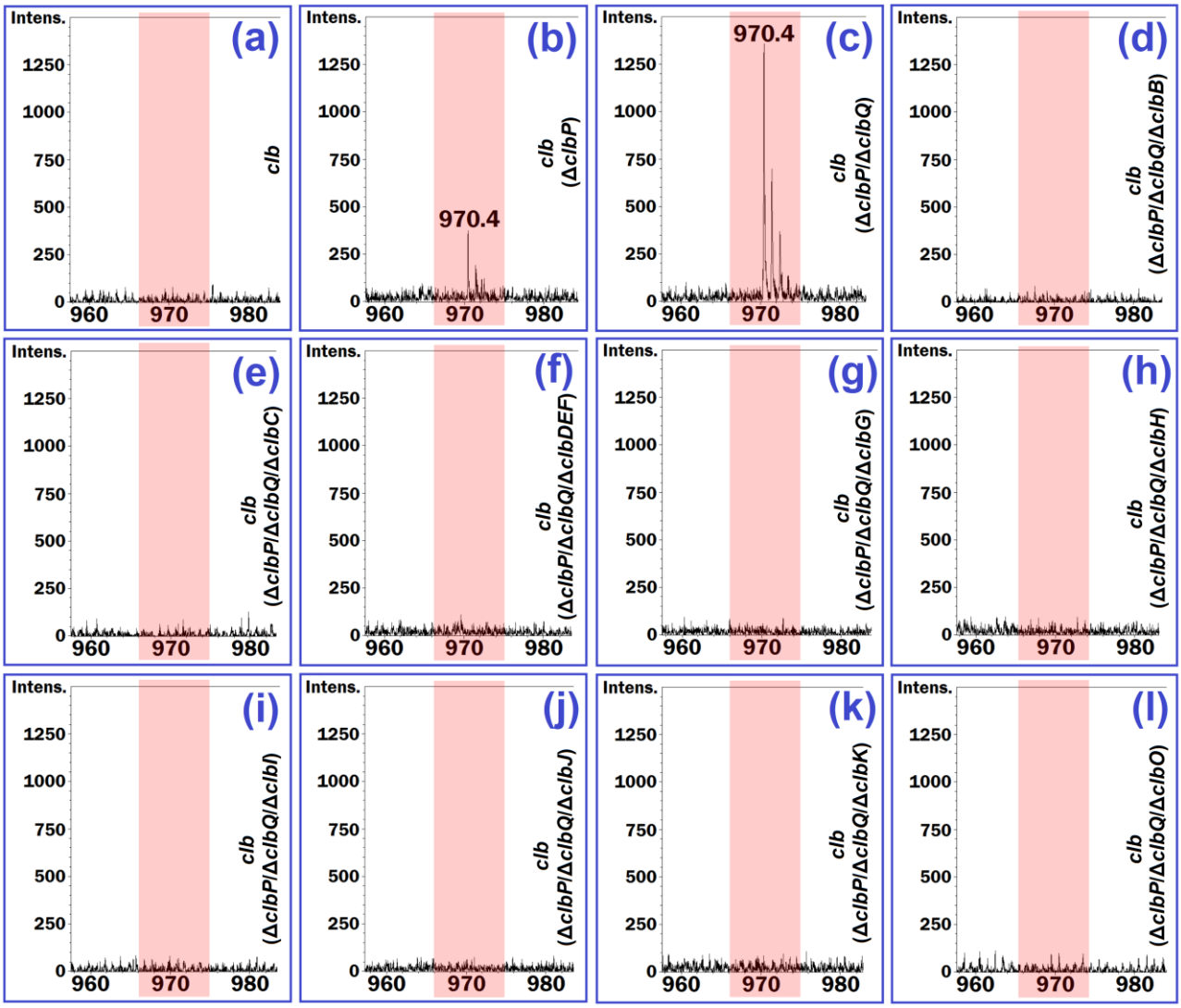
2,5-dihydro-5-hydroxyoxazoles shown in (b) and (c). (b) Nucleophilic addition of an imino anion to a ketone followed by a spontaneous intramolecular cyclization as a synthetic entry to a 2,5-dihydro-5-hydroxyoxazole compound (**30i** in ref. 8)[8]. (c) Preparation of a 2,5-dihydro-5-hydroxyoxazole compound (**14** in ref. 9) from 2-halo-2*H*-azirine features the same nucleophilic addition and cyclization process in the generation of 2,5-dihydro-5-hydroxyoxazole compounds[9]. (d) Comparison of the ^{13}C NMR chemical shifts of the 2,5-dihydro-5-hydroxyoxazole ring in precolibactin-886 (**10**) and synthetic compounds **30i** (in ref. 8) and **14** (in ref. 9).



Supplementary Fig. 14. The comparison of UPLC-MS extracted ion chromatogram traces of the EtOAc extracts of *E. coli* DH10B/pCAP01-*clb* and its eleven mutants. EIC+ = 887.38 ± 0.01, corresponding to compound **10** (precolibactin-886). This systematic gene disruption result showed that only ClbO, the last PKS module of the *clb* pathway, was not involved in the biosynthesis of **10**.

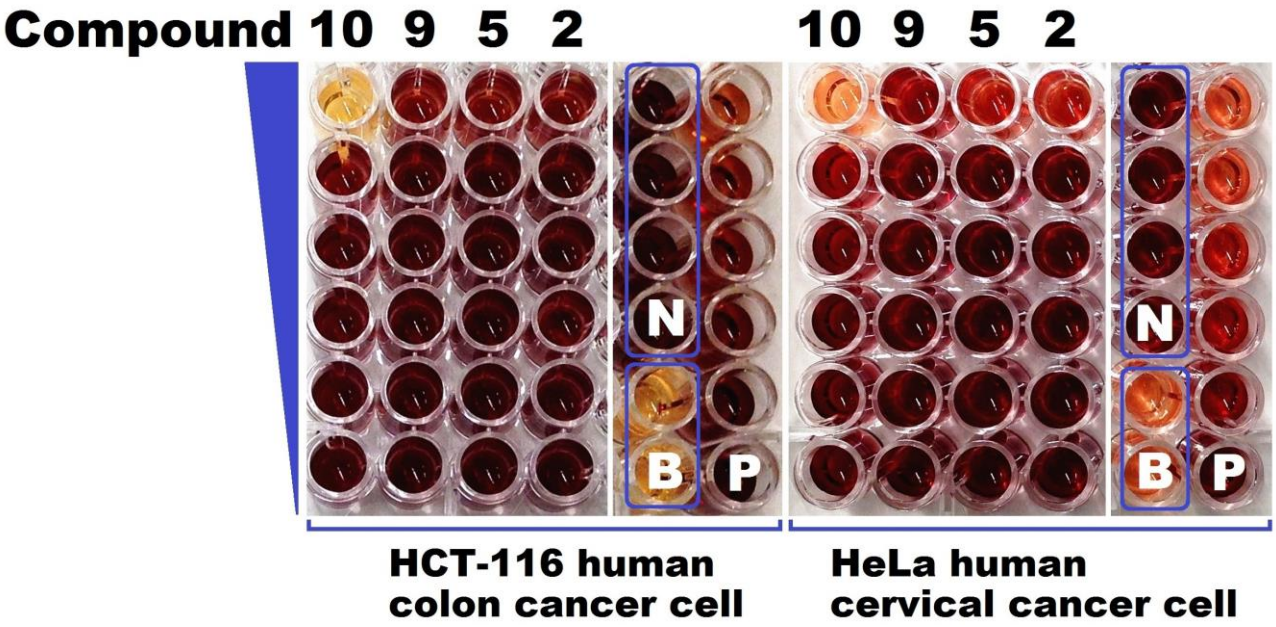


Supplementary Fig. 15. Inactivation of either *clbG* or *clbDEF* would prevent the generation and utilization of the aminomalonyl-ACP extender unit, and thus block path A and significantly redirect the overall pathway flux into path B.



Supplementary Fig. 16. A systematic gene disruption experiment was performed to examine the genes involved in the biosynthesis of precolibacin-969 (**11**). The result showed that any additional knockout of *clbB* (d), *clbC* (e), *clbDEF* (f), *clbG* (g), *clbH* (h), *clbI* (i), *clbJ* (j), *clbK* (k) or *clbO* (l) from the *E. coli* DH10B cells harboring pCAP01-*clb* (Δ*clbP*/Δ*clbQ*) completely abolished the production of **11**. The MS spectra obtained from the extracts of *E. coli* DH10B/pCAP01-*clb* (Δ*clbP*) (b) and *E. coli* DH10B/pCAP01-*clb* (Δ*clbP*/Δ*clbQ*) (c) were used as positive controls, and that of *E. coli* DH10B/pCAP01-*clb* (a) was used as a negative control.

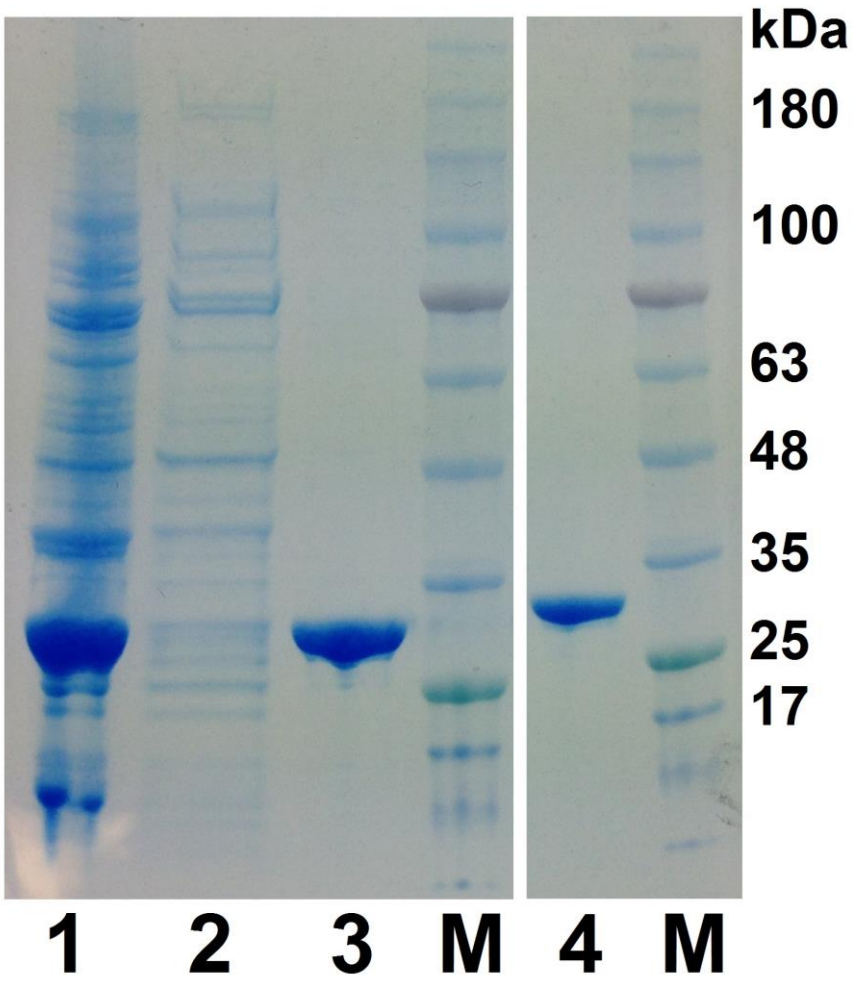
(a)



(b)

	IC_{50} (μM)			
	10	**9**	**5**	**2**
HCT-116 (human colon cancer cell)	**22.3 ± 4.1**	**>100**	**97.9 ± 14.7**	**>100**
HeLa (human cervical cancer cell)	**34.0 ± 5.5**	**>100**	**>100**	**>100**

Supplementary Fig. 17. Cytotoxicity of selected precolibactins. (a) Precolibactin-441 (**2**), -546 (**5**), -795 (**9**), and -886 (**10**) that represent linear, aza-spirocyclopropane, bithiazole, and aminomalonate-containing structural derivatives, respectively, were evaluated in a 96-well plate MTS/PMS cytotoxicity assay against HCT-116 human colon carcinoma and HeLa human cervical carcinoma cell lines, respectively, in concentrations diluted serially from top to bottom rows (78.1, 19.5, 4.88, 1.22, 0.31, and 0.076 μg/mL, respectively). The color indicates cell viability (darker is more viable and thus less cytotoxic). N: negative control (DMSO, final concentration 0.75%); B: blank culture media without cells (McCoy's 5A media for HCT-116 testing and DMEM media for HeLa testing); P: positive control (etoposide, IC_{50} 2.0 ± 0.36 and 1.1 ± 0.19 μM against HCT-116 and HeLa cells, respectively). (b) IC_{50} (the concentration inhibiting cell growth by 50%) values of selected precolibactins **2**, **5**, **9**, and **10**, expressed as mean ± SD, N = 3.



Supplementary Fig. 18. SDS-PAGE analysis of protein expression of recombinant thioesterases from *E. coli* BL21 (DE3)/pLysE/pET28a-*clbQ* and *E. coli* BL21 (DE3)/pLysE/pET28a-*amiD* cells. Columns 1 and 2: insoluble and soluble fractions of ClbQ expression, showing a poor solubility of *N*-His_6-ClbQ expressed. Column 3: purified *N*-His_6-ClbQ. Column 4: purified *N*-His_6-AmiD. Column M: marker.

Supplementary References

1. Brotherton, C.A. & Balskus, E.P. *J. Am. Chem. Soc.* **135**, 3359−3362 (2013).

2. Li, Z.R. *et al. ChemBioChem* **16**, 1715−1719 (2015).

3. Li, Y. *et al. Sci. Rep.* **5**, 9383 (2015).

4. Datsenko, K.A. & Wanner, B.L. *Proc. Natl. Acad. Sci. USA* **97**, 6640−6645 (2000).

5. Gust, B. *et al. Proc. Natl. Acad. Sci. USA* **100**, 1541–1546 (2003).

6. Wagner, J.K., Marquis, K.A. & Rudner, D.Z. *Mol. Microbiol.* **73**, 963–974 (2009).

7. Zha, L., Wilson, M.R., Brotherton, C.A. & Balskus, E.P. *ACS Chem. Biol.* **11**, 1287−1295 (2016).

8. Patonay, T. & Hoffman, R.V. *J. Org. Chem.* **60**, 2368−2377 (1995).

9. Pinho e Melo, T.M.V.D. *et al. J. Org. Chem.* **67**, 66−71 (2002).