

Supplementary Information

Divergent biosynthesis yields a cytotoxic aminomalonate-containing precolibactin

Zhong-Rui Li^{1,7}, Jie Li^{2,7}, Jin-Ping Gu³, Jennifer Y. H. Lai¹, Brendan M. Duggan⁴, Wei-Peng Zhang¹, Zhi-Long Li⁵, Yong-Xin Li¹, Rong-Biao Tong⁵, Ying Xu⁶, Dong-Hai Lin³, Bradley S. Moore^{2,4*} & Pei-Yuan Qian^{1*}

¹Division of Life Science and Environmental Science Programs, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, P. R. China. ²Center for Marine Biotechnology and Biomedicine, Scripps Institution of Oceanography, University of California at San Diego, La Jolla, CA 92093, United States. ³High-field NMR Research Center, College of Chemistry and Chemical Engineering, Xiamen University, Xiamen 361005, P. R. China. ⁴Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California at San Diego, La Jolla, CA 92093, United States. ⁵Department of Chemistry, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, P. R. China. ⁶Shenzhen 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. ⁷These 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		
<i>E. coli</i> CFT073	Wild type <i>E. coli</i> strain harboring the colibactin (<i>clb</i>) gene cluster	1,2
<i>Bacillus subtilis</i> 1779	Wild type <i>B. subtilis</i> strain harboring the amicoumacin (<i>ami</i>) gene cluster	3
<i>E. coli</i> Top10	Host strain for routine cloning	Invitrogen
<i>E. coli</i> BL21 (DE3)/pLysE	Host strain for protein expression	Novagen
<i>E. coli</i> DH10B	Host for heterologous expression: $\Delta(\textit{araABC-leu})7697$, $\textit{araD139}$, \textit{deoR} , $\textit{endA1}$, \textit{galK} , \textit{galU} , $\Delta(\textit{lac})X74mcrA$, $\Delta(\textit{mcrCB-hsdSMR-mrr})$, \textit{nupG} , $\textit{recA1}$, $\textit{rpsL}(\text{Str}^r)$, ($\phi 80 \textit{lacZ}\Delta M15$)	Invitrogen
<i>E. coli</i> BW25113	K12 derivative, $\Delta\textit{araBAD}$, $\Delta\textit{rhaBAD}$	4
<i>E. coli</i> BW25113/pCAP01- <i>clb</i> /pIJ790	<i>E. coli</i> BW25113 harboring pCAP01- <i>clb</i> and pIJ790, for the further genetic manipulation of pCAP01- <i>clb</i>	2
Plasmids		
pET-28a(+)	Protein expression vector, production of N-terminally His-tagged fusion proteins, pBR322 ori, T7 promoter, \textit{kan}^R	Novagen
pETDuet-1	Protein expression vector, pBR322 ori, T7 promoter, \textit{amp}^R	Novagen
pIJ790	λ -Red (\textit{gam} , \textit{bet} , \textit{exo}), \textit{cat} , \textit{araC} , $\textit{rep101ts}$	5
pIJ773	Source of \textit{apra}^R	5
pDR111	Source of \textit{amp}^R	6
Protein expression plasmids constructed		
pETDuet-1- <i>amiD</i>	pETDuet-1, the thioesterase AmiD for <i>in vivo</i> complementation	This study
pETDuet-1- <i>clbQ</i>	pETDuet-1, the thioesterase ClbQ for <i>in vivo</i> complementation	This study
pET-28a- <i>amiD</i>	pET-28a(+), recombinant N-His ₆ -AmiD for <i>in vitro</i> assay	This study
pET-28a- <i>clbQ</i>	pET-28a(+), recombinant N-His ₆ -ClbQ for <i>in vitro</i> assay	This study
Plasmids for heterologous expression and genetic manipulation of <i>clb</i> pathway		
pCAP01- <i>clb</i>	pCAP01 derivative that carries a 70-kb genomic region containing the entire <i>clb</i> gene cluster	2
pCAP01- <i>clb</i> ($\Delta\textit{clbP}$)	pCAP01- <i>clb</i> derivative ($\Delta\textit{clbP}$): \textit{apra}^R	2
pCAP01- <i>clb</i> ($\Delta\textit{clbP}/\Delta\textit{clbQ}$)	pCAP01- <i>clb</i> derivative ($\Delta\textit{clbP}$ & $\Delta\textit{clbQ}$): \textit{apra}^R	This study
pCAP01- <i>clb</i>	pCAP01- <i>clb</i> derivative ($\Delta\textit{clbP}$ & $\Delta\textit{clbQ}$): \textit{apra}^R ; ($\Delta\textit{clbB}$):	This study

($\Delta clbP/\Delta clbQ/\Delta clbB$)	amp^R				
pCAP01- <i>clb</i>	pCAP01- <i>clb</i> derivative ($\Delta clbP$ & $\Delta clbQ$):	$apra^R$; ($\Delta clbC$):			This study
($\Delta clbP/\Delta clbQ/\Delta clbC$)	amp^R				
pCAP01- <i>clb</i>	pCAP01- <i>clb</i> derivative ($\Delta clbP$ & $\Delta clbQ$):	$apra^R$;			This study
($\Delta clbP/\Delta clbQ/\Delta clbD$ <i>EF</i>)	($\Delta clbDEF$):	amp^R			
pCAP01- <i>clb</i>	pCAP01- <i>clb</i> derivative ($\Delta clbP$ & $\Delta clbQ$):	$apra^R$; ($\Delta clbG$):			This study
($\Delta clbP/\Delta clbQ/\Delta clbG$)	amp^R				
pCAP01- <i>clb</i>	pCAP01- <i>clb</i> derivative ($\Delta clbP$ & $\Delta clbQ$):	$apra^R$; ($\Delta clbH$):			This study
($\Delta clbP/\Delta clbQ/\Delta clbH$)	amp^R				
pCAP01- <i>clb</i>	pCAP01- <i>clb</i> derivative ($\Delta clbP$ & $\Delta clbQ$):	$apra^R$; ($\Delta clbI$):			This study
($\Delta clbP/\Delta clbQ/\Delta clbI$)	amp^R				
pCAP01- <i>clb</i>	pCAP01- <i>clb</i> derivative ($\Delta clbP$ & $\Delta clbQ$):	$apra^R$; ($\Delta clbJ$):			This study
($\Delta clbP/\Delta clbQ/\Delta clbJ$)	amp^R				
pCAP01- <i>clb</i>	pCAP01- <i>clb</i> derivative ($\Delta clbP$ & $\Delta clbQ$):	$apra^R$; ($\Delta clbK$):			This study
($\Delta clbP/\Delta clbQ/\Delta clbK$)	amp^R				
pCAP01- <i>clb</i>	pCAP01- <i>clb</i> derivative ($\Delta clbP$ & $\Delta clbQ$):	$apra^R$; ($\Delta clbO$):			This study
($\Delta clbP/\Delta clbQ/\Delta clbO$)	amp^R				

***E. coli* CFT073 mutants**

<i>E. coli</i> CFT073 $\Delta clbP$	$\Delta clbP$: $apra^R$, deletion of <i>clbP</i> gene, the <i>clbP</i> gene was replaced by an apramycin resistance gene ($apra^R$) by λ Red-mediated recombination in <i>E. coli</i> CFT073				This study
<i>E. coli</i> CFT073 $\Delta clbP/\Delta clbQ$	$\Delta clbP$ & $\Delta clbQ$: $apra^R$, deletion of <i>clbP</i> & $\Delta clbQ$ genes, the <i>clbP</i> & $\Delta clbQ$ genes were replaced by an apramycin resistance gene ($apra^R$) by λ Red-mediated recombination in <i>E. coli</i> CFT073				This study

***clb*⁺ heterologous expression strains**

<i>E. coli</i> DH10B harboring pCAP01- <i>clb</i>					2
<i>E. coli</i> DH10B harboring pCAP01- <i>clb</i> ($\Delta clbP$)					2
<i>E. coli</i> DH10B harboring pCAP01- <i>clb</i> ($\Delta clbP/\Delta clbQ$)					This study
<i>E. coli</i> DH10B harboring pCAP01- <i>clb</i> ($\Delta clbP/\Delta clbQ/\Delta clbB$)					This study
<i>E. coli</i> DH10B harboring pCAP01- <i>clb</i> ($\Delta clbP/\Delta clbQ/\Delta clbC$)					This study
<i>E. coli</i> DH10B harboring pCAP01- <i>clb</i> ($\Delta clbP/\Delta clbQ/\Delta clbDEF$)					This study
<i>E. coli</i> DH10B harboring pCAP01- <i>clb</i>					This study

DH10B/pCAP01- <i>clb</i> ($\Delta clbP/\Delta clbQ/\Delta clbG$)	<i>E. coli</i>	DH10B	harboring	pCAP01- <i>clb</i>	This study
DH10B/pCAP01- <i>clb</i> ($\Delta clbP/\Delta clbQ/\Delta clbH$)	<i>E. coli</i>	DH10B	harboring	pCAP01- <i>clb</i>	This study
DH10B/pCAP01- <i>clb</i> ($\Delta clbP/\Delta clbQ/\Delta clbI$)	<i>E. coli</i>	DH10B	harboring	pCAP01- <i>clb</i>	This study
DH10B/pCAP01- <i>clb</i> ($\Delta clbP/\Delta clbQ/\Delta clbJ$)	<i>E. coli</i>	DH10B	harboring	pCAP01- <i>clb</i>	This study
DH10B/pCAP01- <i>clb</i> ($\Delta clbP/\Delta clbQ/\Delta clbK$)	<i>E. coli</i>	DH10B	harboring	pCAP01- <i>clb</i>	This study
DH10B/pCAP01- <i>clb</i> ($\Delta clbP/\Delta clbQ/\Delta clbO$)	<i>E. coli</i>	DH10B	harboring	pCAP01- <i>clb</i>	This study
Protein expression strains for $\Delta clbQ$ complementation assay					
DH10B/pCAP01- <i>clb</i> ($\Delta clbP/\Delta clbQ$)::pET Duet-1- <i>amiD</i>	<i>E. coli</i>	Protein expression strain for <i>in vivo</i> $\Delta clbQ$ mutant complementation with thioesterase AmiD			This study
DH10B/pCAP01- <i>clb</i> ($\Delta clbP/\Delta clbQ$)::pET Duet-1- <i>clbQ</i>	<i>E. coli</i>	Protein expression strain for <i>in vivo</i> $\Delta clbQ$ mutant complementation with thioesterase ClbQ			This study
Protein expression strains for production of recombinant proteins					
(DE3)/pLysE/pET-28 a- <i>amiD</i>	<i>E. coli</i>	BL21	Protein expression strain for production of N-His ₆ -AmiD		This study
(DE3)/pLysE/pET-28 a- <i>clbQ</i>	<i>E. coli</i>	BL21	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		
<i>clbB</i> -knock out-F	<u>GTGGACGCTGGTTGTAGGAGGGGAATCGG</u> <u>GATTAACACTGTTTCAGTGATGGACTCC</u> TTATTTGATTTT	Deletion of gene <i>clbB</i> from pCAP01- <i>clb</i> (<i>clbP/clbQ</i>).
<i>clbB</i> -knock out-R	TAAGCGCACGTA ^{ACTCAATAGGATCACCC} <u>AGCACCGTACCGGTGCCGTGCTTGGTCTG</u> ACAGTTACCAAT	
<i>clbC</i> -knock out-F	<u>AACGGCATGGAAATCGCCATTATTGGTAT</u> <u>GGCGGTCCGTTTCCCAGTCCGACTCC</u> TTATTTGATTTT	Deletion of gene <i>clbC</i> from pCAP01- <i>clb</i> (<i>clbP/clbQ</i>).
<i>clbC</i> -knock out-R	<u>ACTCCTGCGGCTGTATCGGGATATAAAAC</u> <u>GGTGAATGCGCCAGATCCAGCTTGGTCTG</u> ACAGTTACCAAT	
<i>clbDEF</i> -knock out-F	<u>TGCAGGAGTAATGGGAACTGGCGTCGCTC</u> <u>ATAACATGGCGCAATACGGCAGACTCC</u> TTATTTGATTTT	Deletion of genes <i>clbD–F</i> from pCAP01- <i>clb</i> (<i>clbP/clbQ</i>).
<i>clbDEF</i> -knock out-R	<u>TTTCTTTTTCCCGTTCAGCCGCAACCGTC</u> <u>GCCATCCTGCTGTAATTCTGTTGGTCTGAC</u> AGTTACCAAT	
<i>clbG</i> -knock out-F	<u>GTGGACGCTGGTTGTAGGAGGGGAATCGG</u> <u>GATTAACACTGTTTCAGTGATGGACTCC</u> TTATTTGATTTT	Deletion of gene <i>clbG</i> from pCAP01- <i>clb</i> (<i>clbP/clbQ</i>).
<i>clbG</i> -knock out-R	TAAGCGCACGTA ^{ACTCAATAGGATCACCC} <u>AGCACCGTACCGGTGCCGTGCTTGGTCTG</u> ACAGTTACCAAT	
<i>clbH</i> -knock out-F	<u>ACGCGGAGAATCTGTGCGCACTGCAACTGC</u> <u>CTTTTTGTTTCGAATTGATTAGACTCCT</u> TATTTGATTTT	Deletion of gene <i>clbH</i> from pCAP01- <i>clb</i> (<i>clbP/clbQ</i>).
<i>clbH</i> -knock out-R	<u>TTGATGGTAGTGAAGCGCAGCAGGTCAAC</u> <u>CAACGCCACGTGCTGACCGCATTGGTCTG</u> ACAGTTACCAAT	
<i>clbI</i> -knock out-F	<u>ATAGCTATCATTGGGATGGCGGGGCGTTT</u> <u>CCCTCAAGCCGATACGGTACAGACTCC</u> TTATTTGATTTT	Deletion of gene <i>clbI</i> from pCAP01- <i>clb</i> (<i>clbP/clbQ</i>).
<i>clbI</i> -knock out-R	<u>GCTGTTATCGGAAAACGCCCGACAGTGGC</u> <u>CATCGGCGGCGGTGATCCCACCTTGGTCTG</u> ACAGTTACCAAT	
<i>clbJ</i> -knock out-F	<u>GATCATGTGGCCC GCGCCCTGTTAGCGCT</u> <u>GGGCGTGCAGCATGGCGACCGGACTCC</u> TTATTTGATTTT	Deletion of gene <i>clbJ</i> from pCAP01- <i>clb</i> (<i>clbP/clbQ</i>).
<i>clbJ</i> -knock out-R	<u>AAATAGCTCAGCAATAGGTACCGTAACCT</u> <u>TAAAAATCTCCTCAATACGGCTTGGTCTG</u>	

	ACAGTTACCAAT	
<i>clbK</i> -knock out-F	<u>GTACACGGCATT</u> <u>TTTACGACTGGGTGCGGT</u> <u>CTATCTGCCAGTGGATCCGGTGACTCC</u> TTATTTGATTTT	Deletion of gene <i>clbK</i> from pCAP01- <i>clb</i> (<i>clbP/clbQ</i>).
<i>clbK</i> -knock out-R	CTGCCCCGATAATCGCCTCAAGTGCCTGCT <u>GAATACGCACCAATTCTATAGTTGGTCTG</u> ACAGTTACCAAT	
<i>clbO</i> -knock out-F	<u>TGGCTCACTGGATATTGCCATTATTGGCAT</u> <u>GAGCGGGCGTTTTTCCGGTGGACTCCT</u> TATTTGATTTT	Deletion of gene <i>clbO</i> from pCAP01- <i>clb</i> (<i>clbP/clbQ</i>).
<i>clbO</i> -knock out-R	TGTGCGGCGCACATGCCGGTGCGAAAACC <u>CGGTGCAGAGCTTCAAGCTCATTGGTCTG</u> ACAGTTACCAAT	
<i>clbP</i> -knock out-F	<u>ACACGTTAGCATTAAAACATTATATCATC</u> <u>TCCTGTGCTGTATGCTGCTCTCTCACGTTA</u> AGGGATTTTGG	Deletion of genes <i>clbP</i> & <i>clbQ</i> in the wild type colibactin producer <i>E. coli</i> CFT073. Deletion of genes <i>clbP</i> & <i>clbQ</i> from <i>E. coli</i> DH10B harboring pCAP01- <i>clb</i> .
<i>clbP</i> -knock out-R	TCGTTTAATTTGATGATTTAATGTCAGAAC <u>GAAAGCTAACAGGATAATTCGCTCATGAG</u> CTCAGCCAATC	
<i>clbPQ</i> -knock out-F	<u>ACACGTTAGCATTAAAACATTATATCATC</u> <u>TCCTGTGCTGTATGCTGCTCTCTCACGTTA</u> AGGGATTTTGG	Deletion of genes <i>clbP</i> & <i>clbQ</i> in the wild type colibactin producer <i>E. coli</i> CFT073. Deletion of genes <i>clbP</i> & <i>clbQ</i> from <i>E. coli</i> DH10B harboring pCAP01- <i>clb</i> .
<i>clbPQ</i> -knock out-R	TCTACCCTACTATTTTCGAGTGATTCAATCGT <u>CTGGTTACATAACCTACCGCTCATGAGCT</u> CAGCCAATC	

Colony PCRs for correct insert check

<i>clbB</i> -knock out check-F	GCAACGCCGTGTCCACCACGA	Colony PCR for correct insert check (<i>clbB</i> knockout).
<i>clbB</i> -knock out check-R	TGCGGCGACCGAGTTGCTCTT	
<i>clbC</i> -knock out check-F	TGGCGCGTCACTATCCGCAAGTG	Colony PCR for correct insert check (<i>clbC</i> knockout).
<i>clbC</i> -knock out check-R	TGCGGCGACCGAGTTGCTCTT	
<i>clbDEF</i> -knock out check-F	ACGCACCACCCTTATCAGGCACG	Colony PCR for correct insert check (<i>clbD-F</i> knockout).
<i>clbDEF</i> -knock out check-R	TGCGGCGACCGAGTTGCTCTT	
<i>clbG</i> -knock out check-F	GCAACGCCGTGTCCACCACGA	Colony PCR for correct insert check (<i>clbG</i> knockout).
<i>clbG</i> -knock out check-R	TGCGGCGACCGAGTTGCTCTT	
<i>clbH</i> -knock out check-F	GCACCTGGTGGCGCAGTGGA	Colony PCR for correct insert check (<i>clbH</i> knockout).

<i>clbH</i> -knock out check-R	TGCGGCGACCGAGTTGCTCTT	
<i>clbI</i> -knock out check-F	GCAGCAATACATCGGGCAGCAGTG	Colony PCR for correct insert check (<i>clbI</i> knockout).
<i>clbI</i> -knock out check-R	TGCGGCGACCGAGTTGCTCTT	
<i>clbJ</i> -knock out check-F	GATCGAGTTGGCTGGGGAGTTGCA	Colony PCR for correct insert check (<i>clbJ</i> knockout).
<i>clbJ</i> -knock out check-R	TGCGGCGACCGAGTTGCTCTT	
<i>clbK</i> -knock out check-F	CTCCTGCACGCCCTAGCCCAG	Colony PCR for correct insert check (<i>clbK</i> knockout).
<i>clbK</i> -knock out check-R	TGCGGCGACCGAGTTGCTCTT	
<i>clbO</i> -knock out check-F	TGCGGCATGCACCGGAAGACT	Colony PCR for correct insert check (<i>clbO</i> knockout).
<i>clbO</i> -knock out check-R	TGCGGCGACCGAGTTGCTCTT	
<i>clbP</i> -knock out check-F	CGCTGTTGGGCACTCTTTGGCAA	Colony PCR for correct insert check (<i>clbP</i> knockout).
<i>clbP</i> -knock out check-R	CGAGTGAGGTGGCAGGGGCAAT	
<i>clbPQ</i> -knock out check-F	CGCTGTTGGGCACTCTTTGGCAA	Colony PCR for correct insert check (<i>clbPQ</i> knockout).
<i>clbPQ</i> -knock out check-R	CGAGTGAGGTGGCAGGGGCAAT	

Protein expression for *in vitro* enzymatic assay

<i>clbQ1</i> -BamHI- F	agtgagtGGATCCATGAGTAATATCAGTTTGT ATTG	Amplification of gene <i>clbQ</i> for protein expression, inserted into the expression vector pET28a, His tag.
<i>clbQ1</i> -XhoI-R	agtgagtCTCGAGCTACCCTACTATTTTCGAGT G	
<i>amiD1</i> -HindIII -F	agtgagtAAGCTTGCATGATCAAATTATTCTGT CTGCC	Amplification of gene <i>amiD</i> for protein expression, inserted into the expression vector pET28a, His tag.
<i>amiD1</i> -XhoI-R	agtgagtCTCGAGTCATACCACTCCTGTCTGA TTC	

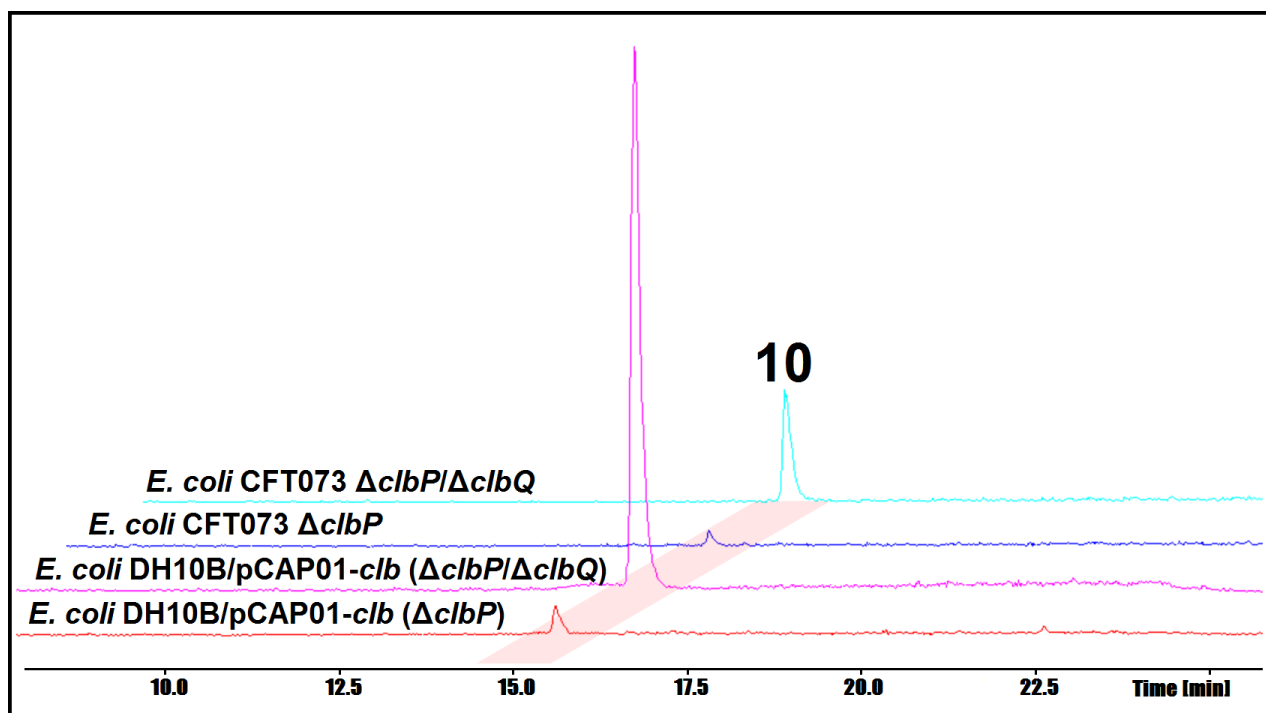
Protein expression for *in vivo* complementation assay

<i>amiD2</i> -HindIII -F	agtgagtAAGCTTATGATCAAATTATTCTGTCT GCC	Amplification of gene <i>amiD</i> for protein expression, inserted into the expression vector pETDuet-1.
<i>amiD2</i> -NotI-R	agtgagtGCGGCCGCTCATACCACTCCTGTCT GATTC	
<i>clbQ2</i> -BamHI- F	agtgagtGGATCCGATGAGTAATATCAGTTT GTATTG	Amplification of gene <i>clbQ</i> for protein expression, inserted into the expression vector pETDuet-1.
<i>clbQ2</i> -HindIII- R	agtgagtAAGCTTCTACCCTACTATTTTCGAGT G	

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	AAV42398
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 <i>B. pseudomycooides</i> _TEII_NRPS	WP_006096422
Megalomicin_TEII_PKS	AAG13923
Gramicidin <i>A. migulans</i> _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 <i>S. coelicolor</i> 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 $\Delta clbP$ and *E. coli* CFT073 $\Delta clbP/\Delta clbQ$; *E. coli* DH10B/pCAP01-*clb* ($\Delta clbP$) and *E. coli* DH10B/pCAP01-*clb* ($\Delta clbP/\Delta clbQ$), 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

```

152 154 156 158 160 162 164 166 168 170 172 174 176 178 180 182 184 186 188 190 192 194 196 198 200 202 204 206 208 210 212 214 216 218 220 222 224 226 228 230 232 234 236 238 240 242 244
E T E E I I H R N L D I I P - - - - - D Y Y R E L L T I P S - - - - - I K D K I R S Y L T Y H N K L I N S G A V N A N I H H F L C G E L T - - - - -
D E A A A W Q R T L D A F I A - - - - - A R W M P K D A D V E Q L Q H L C A M N Q V V R V R D H V P T D T H Q G K L L V F S A A F A M R - - - - -
T S P A R L R L V A D R A V A - - - - - A G A L P A R D A E G T V R - R L L A L T Q I S G R L A R E A T L P A Y G G P L A M L R A A E A L T P A L E - - - - -
D T E N D - - - - - D S A A Y L P - E A V R E T V M Q K K R C Y Q E Y W A Q L I N E G R I K S N I H F I E A G I Q T E T S G - - - - -
Q L E E I V E V V L D S I G - - - - - D Q Y K S F L A D P A D R R E R V K D K M L I Y S I Y R N E L I N A G E V Q S N I H A L V A E D D S I G P V T - - - - -
S L E E R L N Y V N E Q F K - - - - - V N L F P P G T G V K I R G I L R V G K A N Q Q E A Y Y V P R D I N Y P T Q I T L F R A S E E A P M E V A A S E D I
E I E I V H L N M D I I P - - - - - D Y R E I L T I P S - - - - - I K D K I R G Y L A Y H N E V I N S G A V N A N I H H L C G D V D T - - - - -
L G S Q E W L P V I N E M L L - - - - - A R G A A A G A E D S D A G D A V W T A R A K Y F S L D F A I A P R D V R T L L L R P A E P L A G - - - - -
E V E E I I H L N L D I I P - - - - - V Y R E L L T I P S - - - - - I K E K I R G Y L A Y H N Q L I N S G T I N A N I H H L L C D D M T - - - - -
- - - - - A V E A G L K I F H - - - - - V E Q L Q T W G - A S D A G - L T A E A W Y Y E H I G L E T W K P R Q L A A P T L H V R A T E P M K Q F V G - - - - -
I T P R F S V A M A H R T Y - - - - - E K L A T F T D M Q D V G - I T A M G G Y F R M F - - T E W T P T I G A P T L F V R T E D C V A D P E G - - - - -
T V E S D V E A L M N V N R - - - - - D N E A L N S E A V K H G L K Q K T H A F Y S Y Y V N L I S T G V Q K A D I D L L T S G A D F D - - - - -
R K G G L V Q E L L D H A A M - - - - - A R V V P E D S R L L L L R - E P G I A P M L R L T R A H Q R A M L S Y R - P R G L R Q R I V L F R T S E - - - - -
E T E N D I A Q L F A E I G E N T E M F N M T Q E D F Q Y A A N E F V K Q S F V R K T V S V M F H N L V N T G M T T A I I H L I Q S E L A D E A D E - - - - -
S E E Q L I E S L L K K F K L N - - - - - E N S C Q Q N F E D P M M N K L V M I A N R Y A Y L K Y N C K Q I K A D I F L F N A S I N D I H P - - - - -
  
```

Erythromycin_TEI_PKS
 FR-008/candicidin_TEI_PKS
 Natamycin_TEI_PKS
 Bacitracin_TEI_NRPS
 Borrelidin_TEI_PKS
 Kendomycin_TEI_PKS
 Gramicidin B_pseudomycoides_TEI_NRPS
 Megalomicin_TEI_PKS
 Gramicidin A_migulans_TEI_NRPS
 Pikromycin_TEI_PKS
 Rifamycin_TEI_PKS
 Yersiniabactin_TEI_PKS/NRPS
 Zwittermicin_TEI_PKS/NRPS
 Tylosin_TEI_PKS
 Amicoumacin_TEI_PKS/NRPS
 ScoT S. coelicolor A3(2)_TEI_PKS
 Didemnin_TEI_PKS/NRPS
 Colibactin_TEI_PKS/NRPS

```

110 112 114 116 118 120 122 124 126 128 130 132 134 136 138 140 142 144 146 148 150 152 154 156 158 160 162 164 166 168 170 172 174 176 178 180 182 184 186
A P R V H E R - R T D - L P G D D G L V D E L R - R L G - T S E A A L A D E A L L A M S L P V L R A R Y R V L R S Y A W - A D G - - P P L R A G I T A L C
G P S T T R A - E R V H T R D D G I V A E M K - R N G T A A G V L G D E E I L R M A L P A L R G D Y R A I E T Y T C - P P D - - R R L A C G L T V L T
A P S R Y R T - E S V H R R D D G I V A E L K - L L S G T D A A L L G D E E L L R M I L P A I R S D Y T A V E T Y R C - D P G - - A A I K A P I T A L T
A P N R I R K T E K L H T L P N P I F K K K I V - E L G G T P E E L I N H E E L F E L F I P I L K S D F K M V E N Y I Y Q E R N - S K I D C D I T V L N
A P S R Q R T G G T V H L L S D E E L V A E I R - T L D G T A E Q V F H D E E L V R M A L P A I R G D Y R A A E T Y R Y - R P G - - P K L R C P I H A L T
A P S C L R A - D T V H L R D D G I V A E I R - A V G G T D A R I D D P E L S F A L P A R A R A V E T Y R A - V E G - - A A V A C P V T V L T
S P Q V S K R E T P T Y N L P K R E F I E L L R - K M N G T P E E V I G N H E I M D L F L P I L R A D F E L V D T Y Q E K S S - - M L S C P I T V F G
A P R V H E R - R T D - L P G D D G L V D E Q M R - R L G - V S E A A L A D Q E L D M S L P V L R A D H R V L R S Y A W - Q A G - - P P L R A G I T A L C
A P Q I P C A K Q D Y H L L P D E Q F I Q E L R - S I N G T P E I V I Q D A E M M S I L L R L R A D F S V C G S Y Q Y K N D E - - P F E C P I T A F G
A P S L A P D - R L V H Q L D D R A F L A E I R - R I S G T D R E F Q D D E L R L V L P A L R S D Y K A E T Y L H - R P S - - A K I T C P V M A I A
A P S R Y R D - D D V R G A S D R F V A E L R - K L G G S D A A M A D P E L L A M V L P A I R S D Y R A V E T Y R H - E P G - - R R V D C P V T V T F
A P H L H S E - R Q L S H R D D A D F I A E L I - D I G G C S P E L R E N Q E M S L F L L R A D F Y A T E S H Y D S P D V C P P L R T P A L L L C
P P E I R R K - - - T N H L D D K D F I D Y I K - S L G L P D E V L Q H E L N Y I L P V I R S D F R A I E N Y I N D D T T - - L L K A P L Y I S
A P T A D R T - E T A H L L G D R E L L A E I R - R Q G T D P G V F A D E E V L R M V L P A I R G D Y A A V G R Y R H - V P G - - P R P G C P L T V T F
P P P H I P E K K V L Q L S D S A L K Q A I I - E M G G V P E E F S K N E Q V L D F F L P I L R A D F E L L E T Y R N - T A V - - A P L S C D I S V F Y
A P T P Q G S - D S R L D T D E K V I A M I R - R E G G T V G K V F D D P V M E M V M P L R A D Y R A V A G Y T W - Q P G - - P P L A V P V T V L V
A P G S S T L G D L L G D D A G F A D A L A R Y D G I P A A L L A D R E I M A V F L P L R A D F R M V A E Y R P - D P A A C P R L C P I L A L G
P P D R A S H E V I L H T L P D A F M E E I V - K L G G T P V D V F R N K E I M T I F T I I K N D Y R L Y E Q V F Q A K A - - R T L T C P I V L F H
  
```

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

```

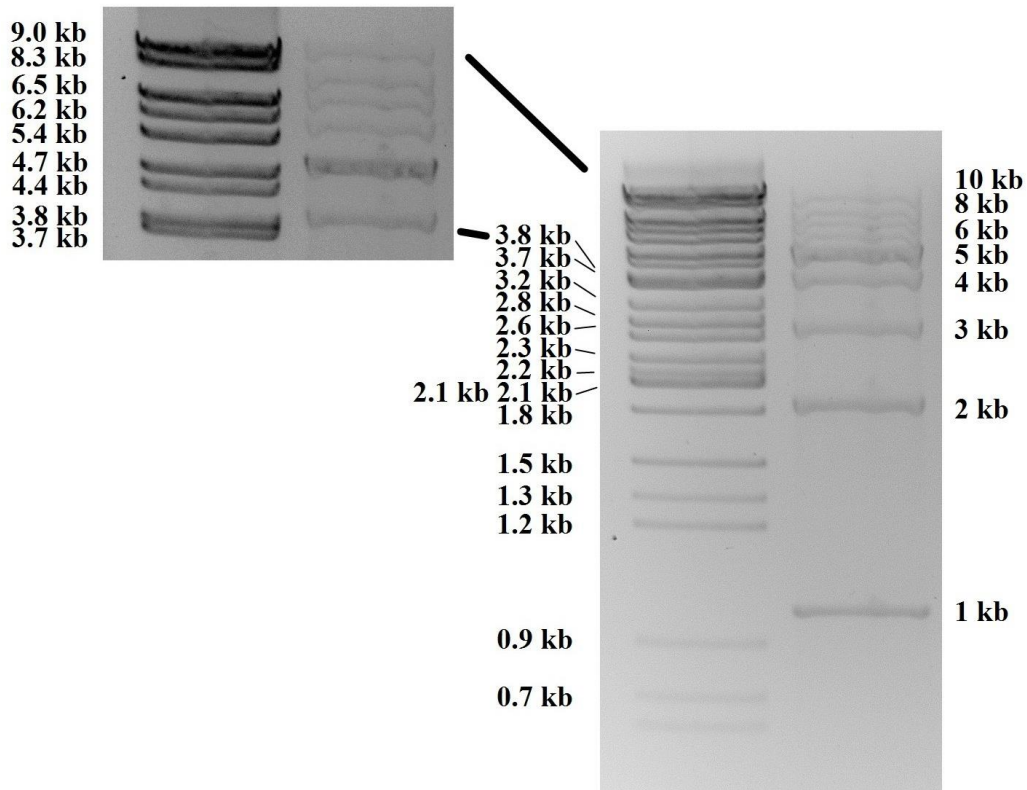
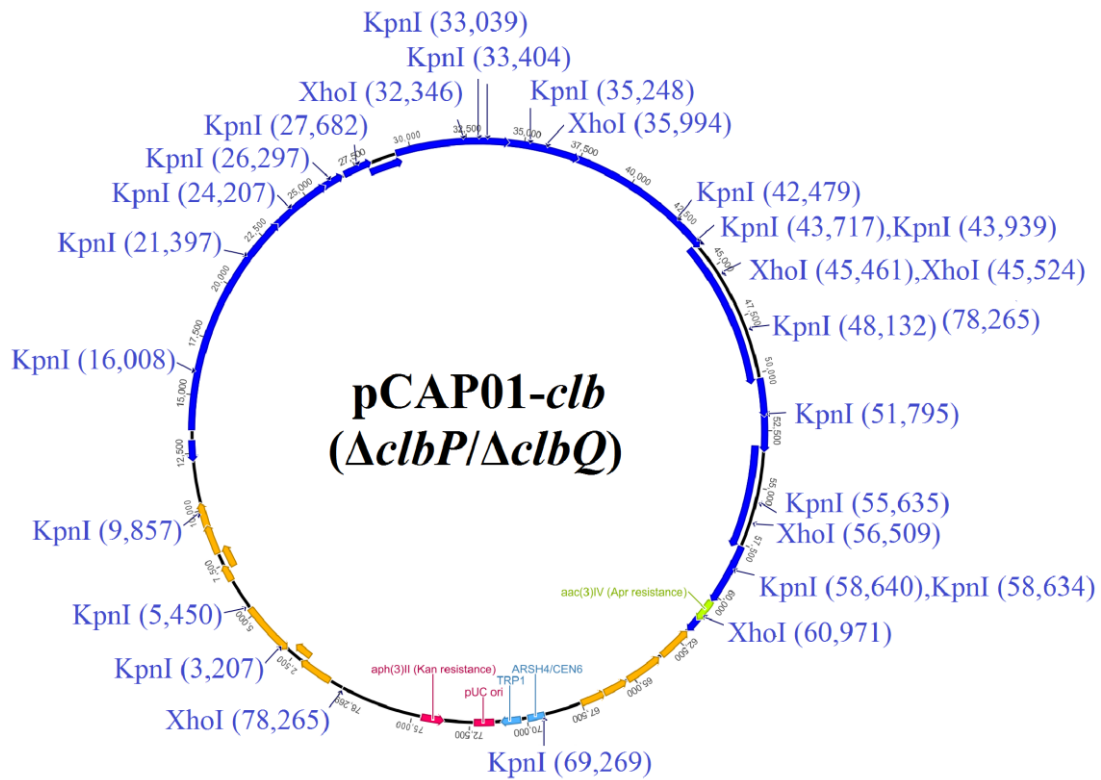
246 248 250 252 254 256 258 260 262 264 266 268 270 272 274 276 278 280 282 284 286 288 290 292 294 296 298 300 302 304 306
- - - - - D R G W K Q - - - - - S T A Q H Y L E Y K L K G D H V T I F D P H N I E E N T D T I R S I I K R I - - - - -
- - - - - N G R Q V S T E G W Q A F C A N P V T T H E V P G D H M T M L R E P D V R G L A I K L R R E I D E - - - - -
- - - - - Y S H I F Q E P T W G N Q F S D G E V E I H I V P G N H V S M L S E P H V K V L A Q - - - - -
- - - - - D L A R R D P A C G W G R H A T G R L A T A T V P G R H D T M M T A P H V D R L A T V I D H - - - - -
- - - - - A M V L K W Q D - - - - - A A E E G Y A E Y T G Y G A H K D M L E G E F A E K N A N I L N I L D K I N S D Q K
- - - - - S S D K L L W Q Q - - - - - A T L G Q Y E E Y G V I G T H D V L L D S G Y V E N A K V L R Q I L G K V - - - - -
H K I L Y D S R K N D P V W G E F S A G E V E L H S V P G D H L T I M A E P N V G V L A E K L R A C L E K A - - - - -
- - - - - D R G W T Q - - - - - S T A Q H Y L E Y K L K G D H V T I F E P H N I E E N A E A I R S I I K R I E - - - - -
- - - - - S A A D A G W R A E W R P G T T V V E V P G N H S M L D S A H A Q A C A R A - - - - -
- - - - - E R G W T H - - - - - S T A H N Y K E Y E L K G D H V T I F D P Q Y I E E N M S T I R S I - - - - -
- - - - - S E G - - - - - A P A E W R A S W K L P H V A I D A P G D H A T V V D H F L A Q A V D D W L - - - - -
- - - - - R P W - - - - - T D S W R P G W T L A D A T V Q V P G D H F S M M D E - - - - -
- - - - - I P E W L A S W E E A T T G A Y R M K R G F G T H A E M L Q G E T L D R N A G I L L E F L N T Q T - - - - -
- - - - - Q A W S F A R D L G W G A L S A S G V D V R E A P G D H M T L L R P P H V E R V A E Q L R A L L - - - - -
- - - - - A P V A A K W N E S A W A N A T Q R L L T Y S G H G I H S R M L A G D Y A S Q N A S I L Q N I L Q E L - - - - -
- - - - - L V D Y N R W N E Y T S G E V Y A L Q V P G S H L S M L E K P H I H E L V K S I K K S - - - - -
  
```

Erythromycin_TEI_PKS
 FR-008/candicidin_TEI_PKS
 Natamycin_TEI_PKS
 Bacitracin_TEI_NRPS
 Borrelidin_TEI_PKS
 Kendomycin_TEI_PKS
 Gramicidin B_pseudomycoides_TEI_NRPS
 Megalomicin_TEI_PKS
 Gramicidin A_migulans_TEI_NRPS
 Pikromycin_TEI_PKS
 Rifamycin_TEI_PKS
 Yersiniabactin_TEI_PKS/NRPS
 Zwittermicin_TEI_PKS/NRPS
 Tylosin_TEI_PKS
 Amicoumacin_TEI_PKS/NRPS
 ScoT S. coelicolor A3(2)_TEI_PKS
 Didemnin_TEI_PKS/NRPS
 Colibactin_TEI_PKS/NRPS

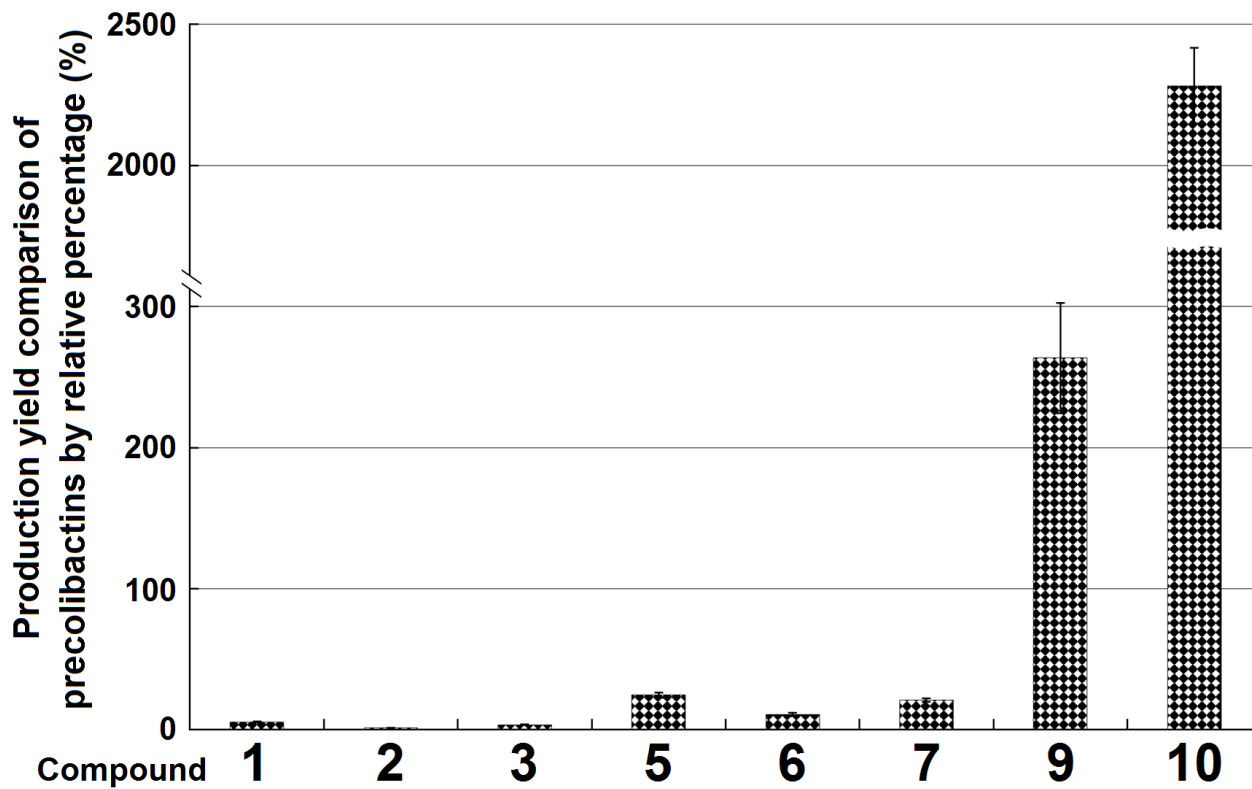
```

188 190 192 194 196 198 200 202 204 206 208 210 212 214 216 218 220 222 224 226 228 230 232 234
G D A D P L T A T G D A E R W L Q H S V I P G R T R T F P G - G H F Y L G E Q V T E V A G A - - -
G E D D P L T T V E E A E R W R D H T T G P F R L R V F T G - G H F L T Q H L D A V N - - -
G D N D P K T T L E E A A D W R A H T T G A F D L Q V Y P G - G H F L L G S Q A P A V L A L - - -
G K E D A M S K E - D V S D W K H H T S G H F T A Y Y F E G - N H F L H H H V E K I T E I - - -
G D D D M V T P V E A R A W S E H T D G P F T L D T F A C - G H F Y L L E H R D A I L - - -
G D H D P R V S V A E A E G W A V H T T G G F S M R V F P G - G H F Y L A E R A R E V A E A - - -
G I D D S I K T E E L K H W K E L T I N N F N M H L F P G - D H F F I Y E N - - -
G D T D P L T T V E D A Q R W L P Y S V V P G R T R T F P G - G H F Y L A D H V G E V A E - - -
G K N D N G V T Q S L E A W R E Q T K R E F S V C M Y P G - D H F F L Y E - - -
G D R D P K A P L N E V A E W R R H T S G P F C L R A Y S G - G H F Y L N D Q W H E I - - -
G D H D P R V S V G E A R A W E E H T T G P A D L R V L P G - G H F L V D Q A A P M I A T - - -
G S H D R E A S W Q V D A W R Q W L S H V T G P V I D G - D H E Y P I Q A R S F F T Q I V R
G T I D S N Y T V K G A K K W V E W G N E V H F L T V N G G - - H M F L L H Q A D I V - - -
G D A D N V T L P E A E A W R E L T T G A F A L R V F P G - G H F Y L N D Q R E A V C R T - - -
G K R D L P S V L F D L S D N N Q Y D Q T C R F Y E F E G - D H F F I N Q K T E E V T A K I N -
G Q D D V P V P A A A A A W R E H T T A G S D R L R V L P G - G H F - - -
G A D D I H A A P E R L T A W Q D F T T A G F M R E I F P G - G H F Y L A A - - -
G D A D N L V M Q D E L L A W E K F T T R K T R T I I F A A D H F F V D K H F E Q V V - - -
  
```

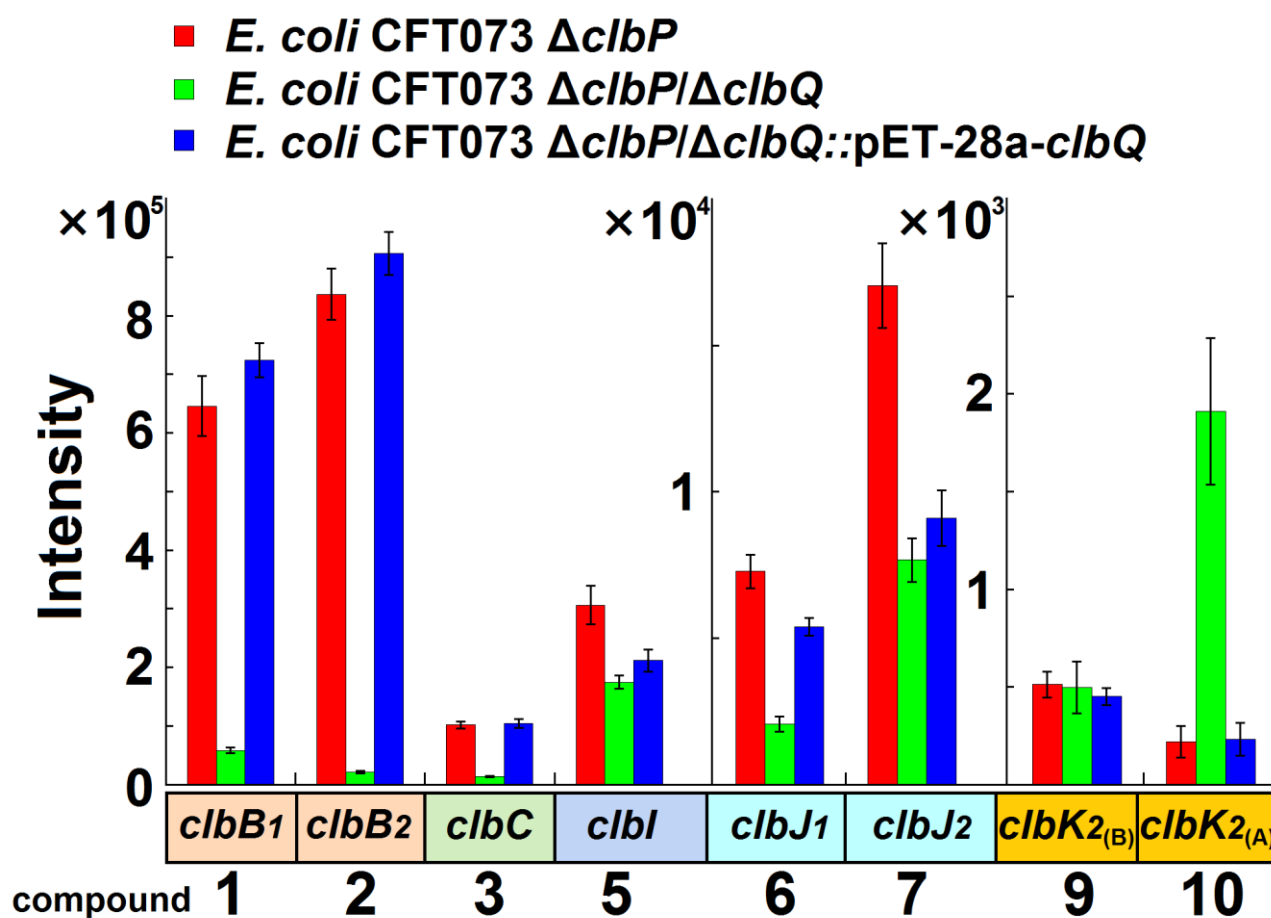
Supplementary Fig. 2. Protein sequence alignment of ClbQ with other representative type I and type II thioesterases, showing that the Ser-His-Asp catalytic triad of type II but not type I TEs is conserved in ClbQ. However, variation of a few amino acid residues adjacent to the catalytic sites were observed and might portend an unusual catalytic function. The signature GX SXG 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.



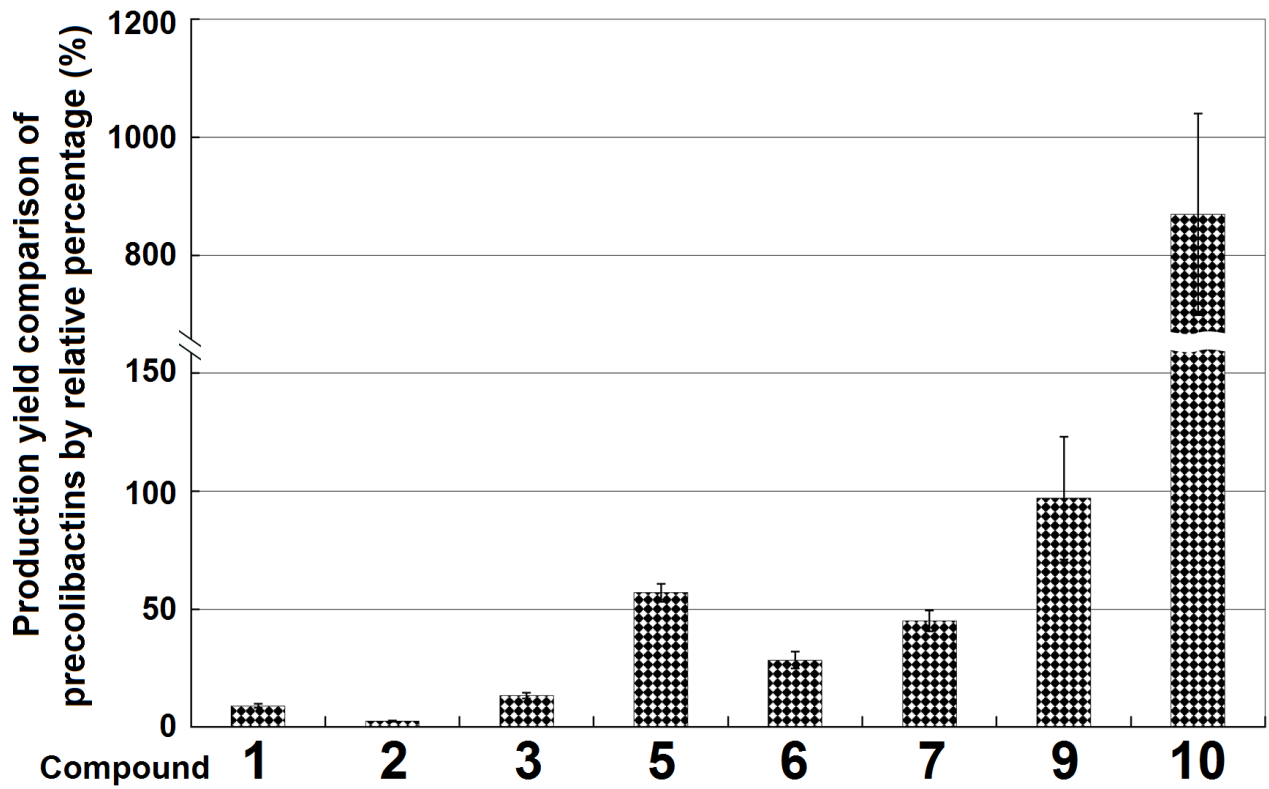
Supplementary Fig. 3. Physical map of bacterial artificial chromosome pCAP01-clb ($\Delta clbP/\Delta clbQ$) with predicted KpnI + XhoI cleavage sites and restriction fragment sizes (upper) and the experimentally determined restriction map with KpnI + XhoI (lower).



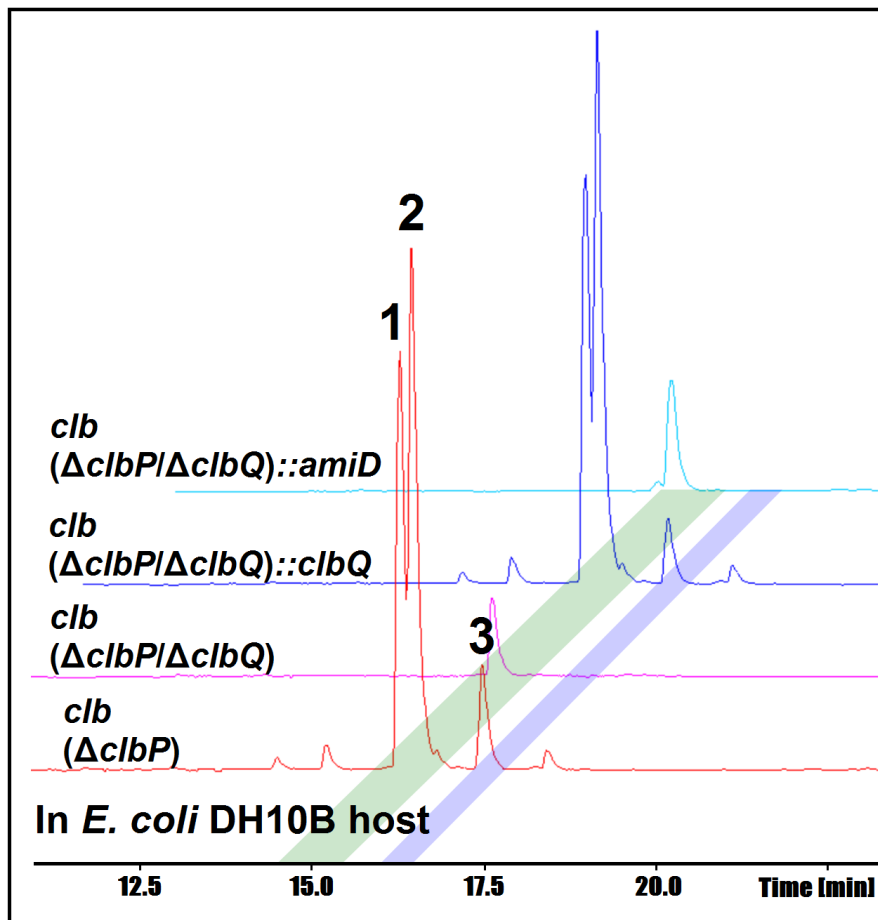
Supplementary Fig. 4. The abundance of individual *clb* pathway-related compound in *E. coli* DH10B/pCAP01-*clb* ($\Delta clbP/\Delta clbQ$) relative to *E. coli* DH10B/pCAP01-*clb* ($\Delta clbP$) (in percentage, %). All values are mean \pm SD; $N = 5$.



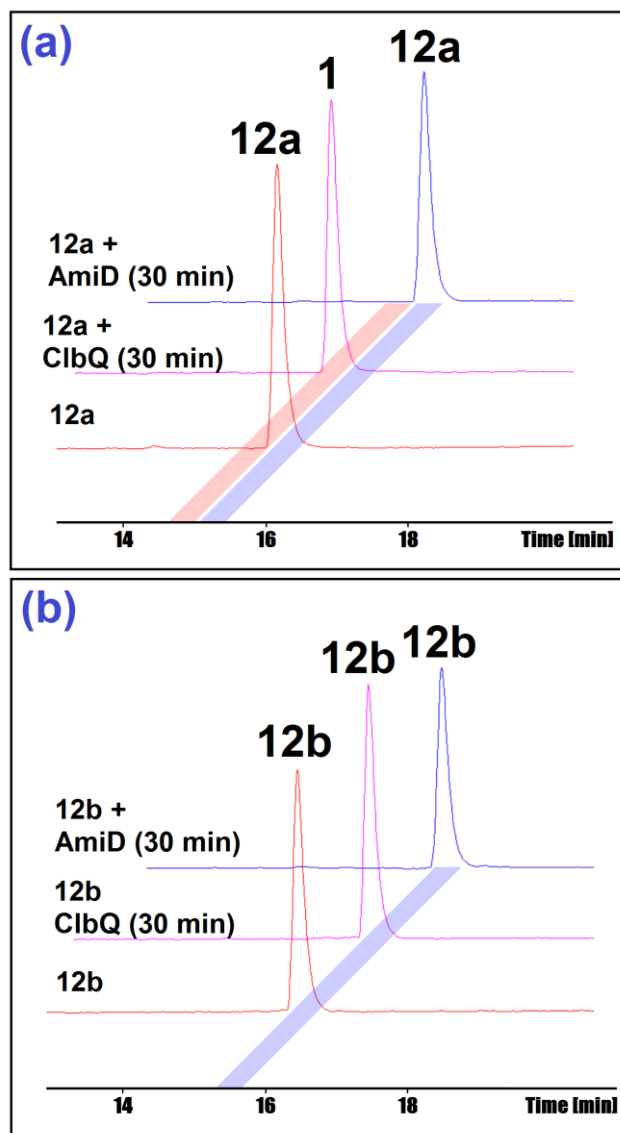
Supplementary Fig. 5. The comparison of the abundance of individual *clb* pathway-related compound from the extracts of colibactin native producer mutants *E. coli* CFT073 $\Delta clbP$, *E. coli* CFT073 $\Delta clbP/\Delta clbQ$ and *E. coli* CFT073 $\Delta clbP/\Delta clbQ::pET-28a-clbQ$, respectively. The extracted ion chromatograms corresponding to the same compound from different mutants were compared. All values are mean \pm 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).



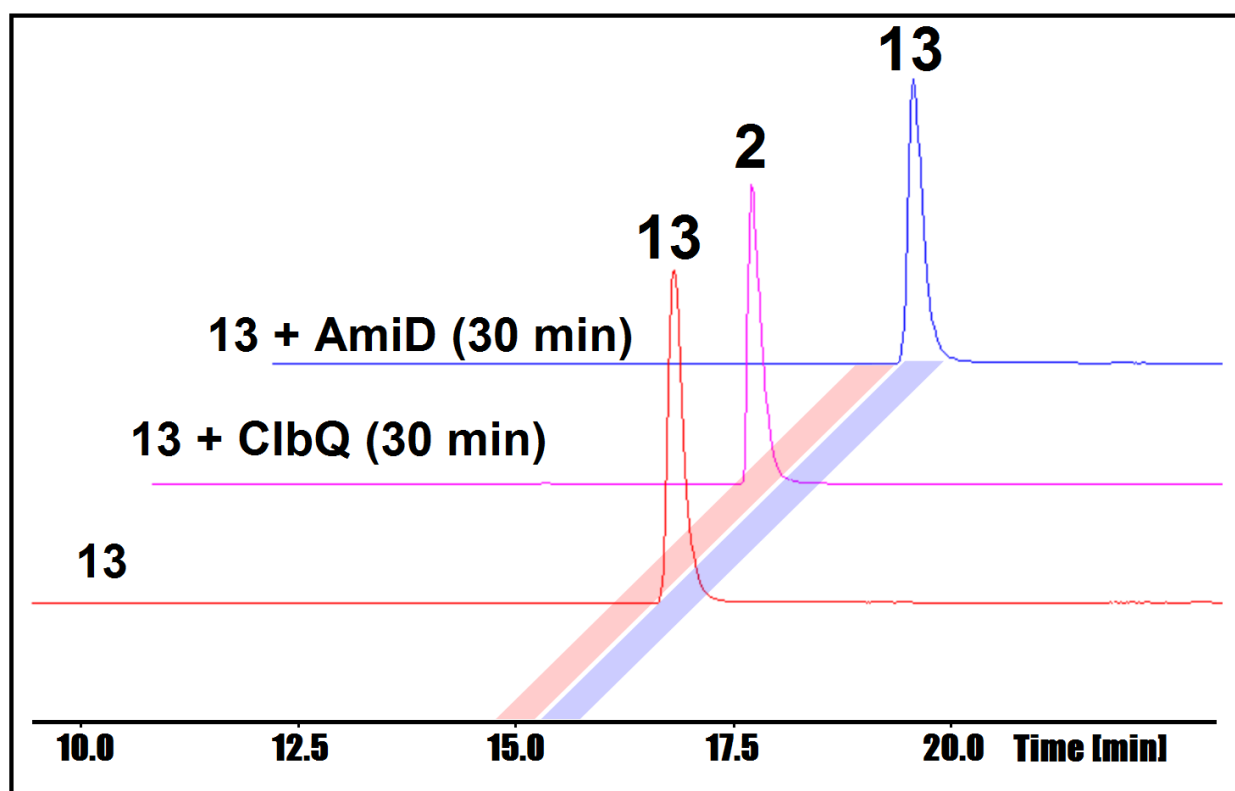
Supplementary Fig. 6. The abundance of individual *clb* pathway-related compound in *E. coli* CFT073 $\Delta clbP/\Delta clbQ$ relative to *E. coli* CFT073 $\Delta clbP$ (in percentage, %). All values are mean \pm SD; $N = 5$.



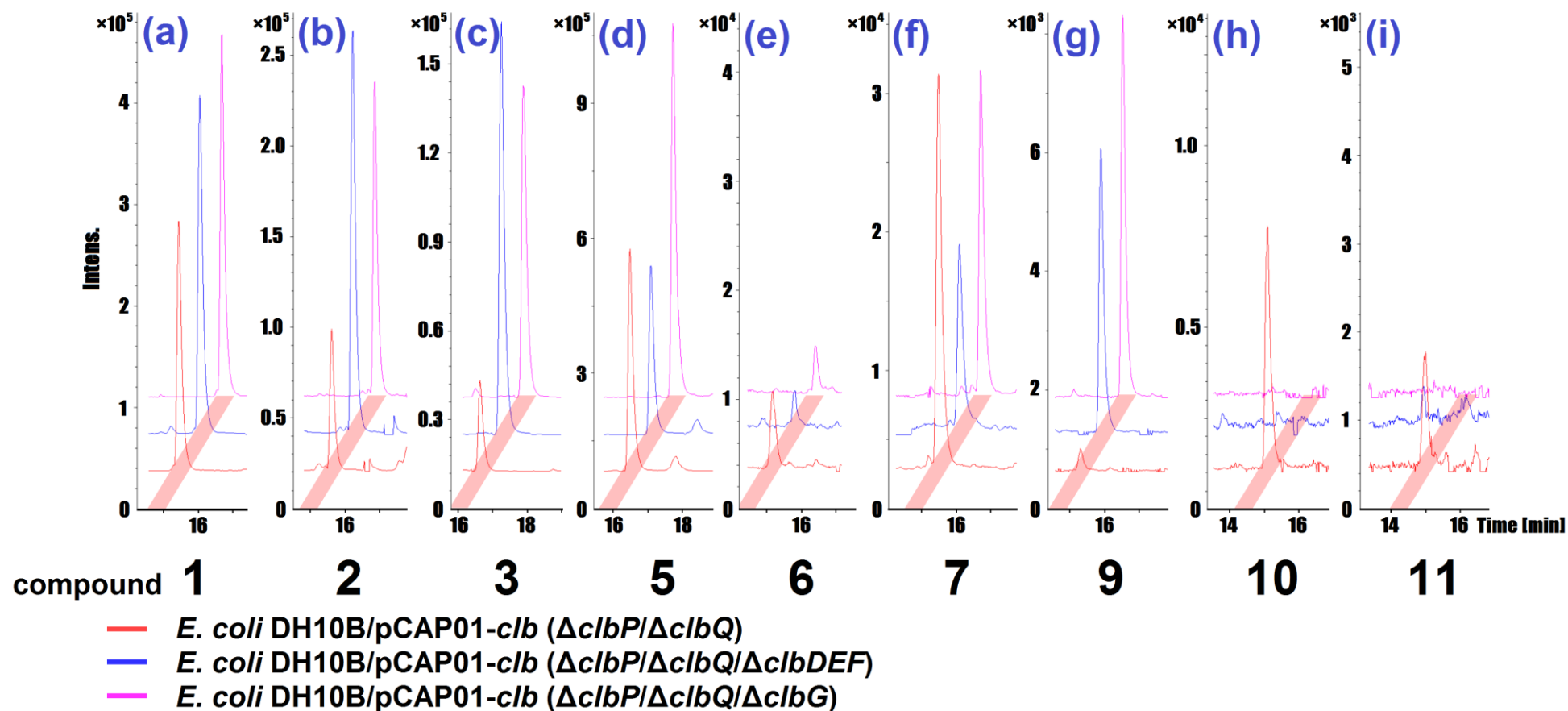
Supplementary Fig. 7. UPLC-MS extracted ion chromatogram traces displaying the production of the *clb* pathway-related compounds in the *clb*⁺ *E. coli* DH10B $\Delta clbP/\Delta clbQ$ mutant and the complementation strains with different thioesterase genes. Extracted ion chromatograms (EIC⁺ = 414.30 ± 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* ($\Delta clbP$) or pCAP01-*clb* ($\Delta clbP/\Delta clbQ$) were first compared. Then, a protein expression plasmid pETDuet-1 carrying ClbQ or AmiD was transformed into *E. coli* DH10B/pCAP01-*clb* ($\Delta clbP/\Delta clbQ$) 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.



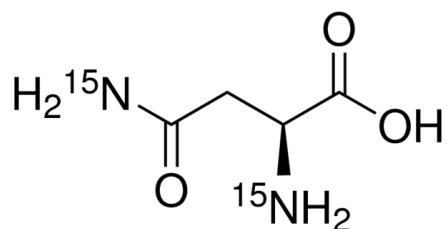
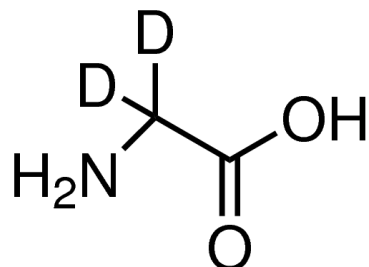
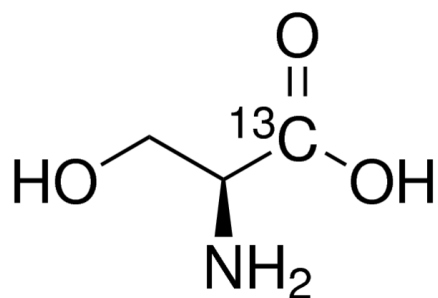
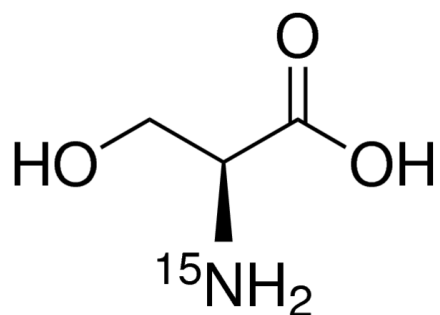
Supplementary Fig. 8. UPLC-MS extracted ion chromatogram traces displaying a complete hydrolysis of **12a** into **1** catalyzed by recombinant ClbQ, but not by recombinant AmiD, after an incubation at 30 °C for 30 min (a). Neither ClbQ nor AmiD could hydrolyze **12b** (b). These results showed that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

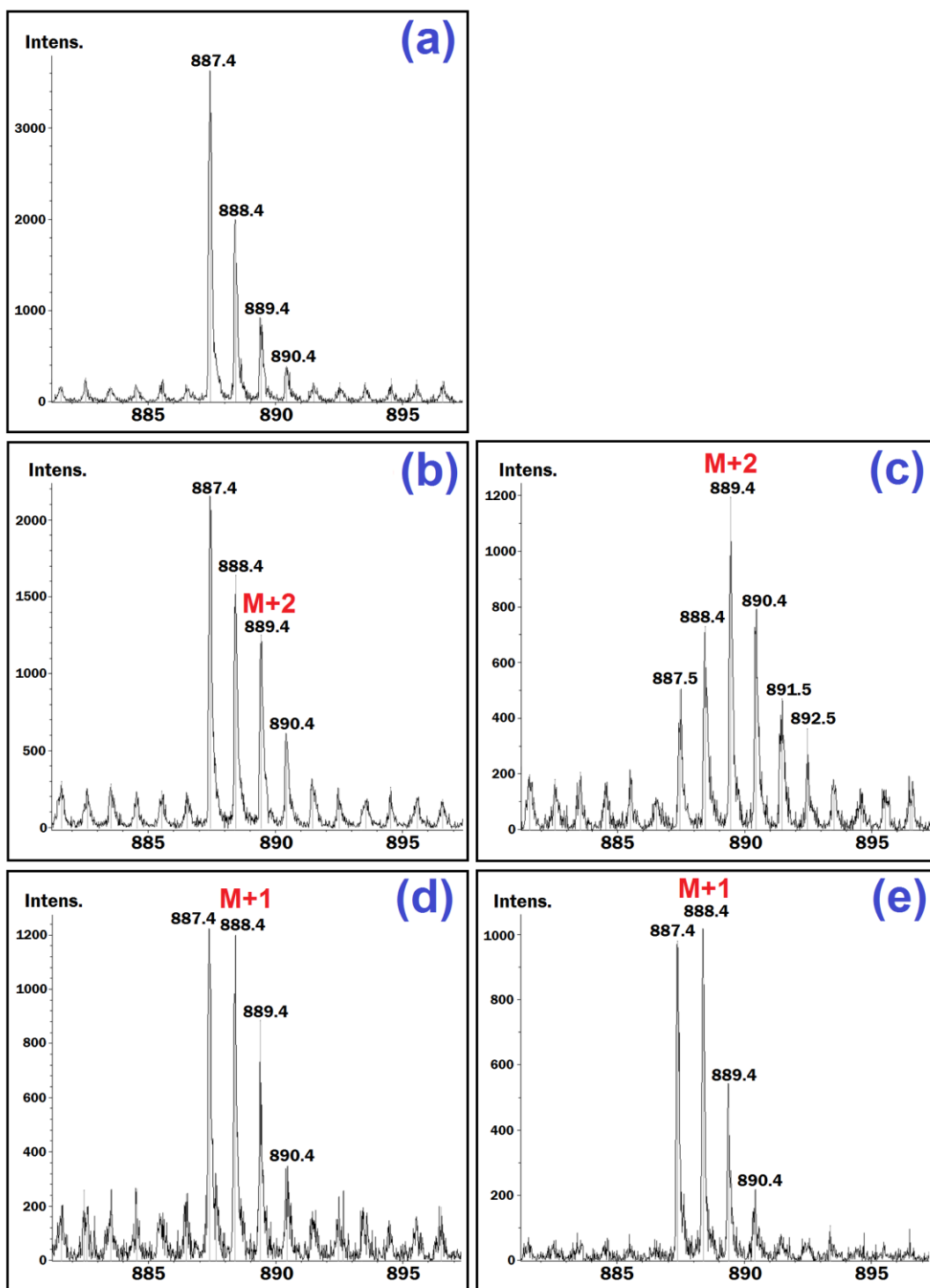


Supplementary Fig. 10. The comparison of UPLC-MS extracted ion chromatogram traces of the EtOAc extracts of *E. coli* DH10B/pCAP01-*clb* and its $\Delta clbDEF$ and $\Delta clbG$ mutants. EIC₊ = 414.30 ± 0.01, 442.33 ± 0.01, 440.35 ± 0.01, 547.39 ± 0.01, 630.39 ± 0.01, 713.37 ± 0.01, 796.35 ± 0.01, 887.38 ± 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**.

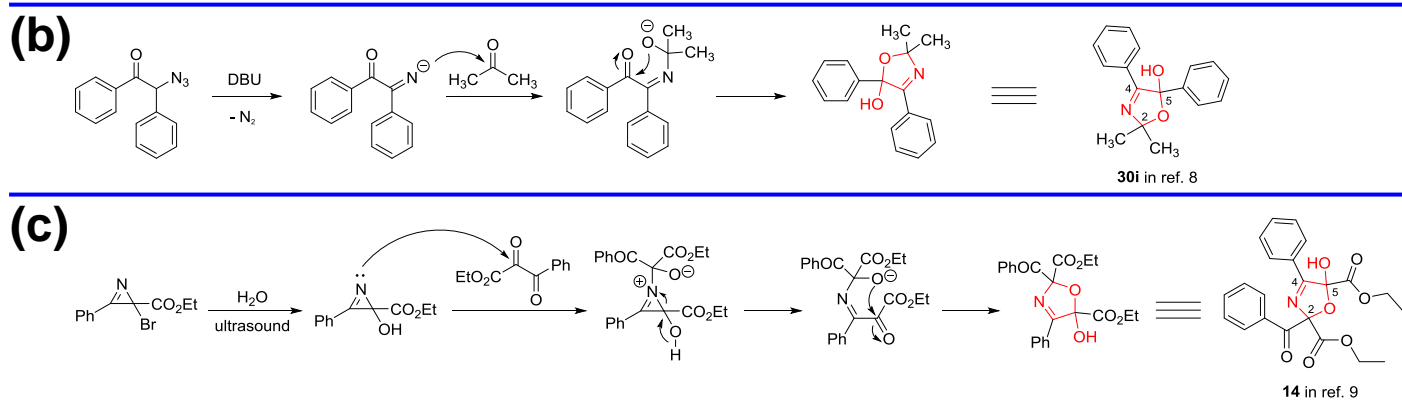
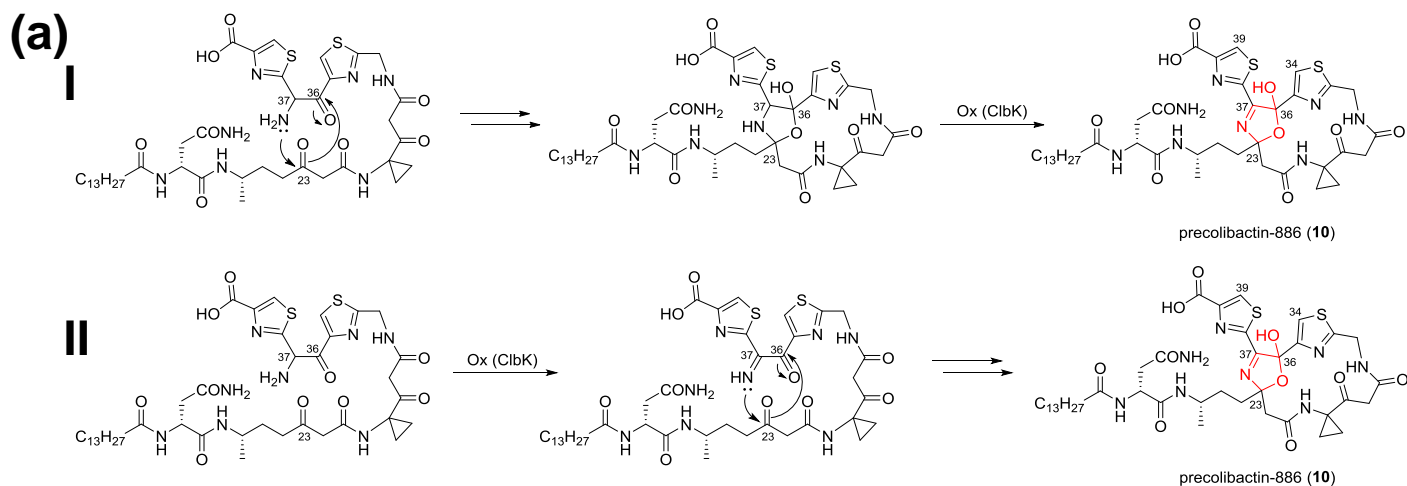
a**b****c****d**

Supplementary Fig. 11. The structures of isotope-labelled amino acids used in the present study. (a) L-[¹⁵N₂]asparagine; (b) [2,2-D₂]glycine; (c) L-[1-¹³C]serine; (d) L-[¹⁵N]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-[$^{15}\text{N}_2$]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* ($\Delta clbP/\Delta clbQ$). 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.

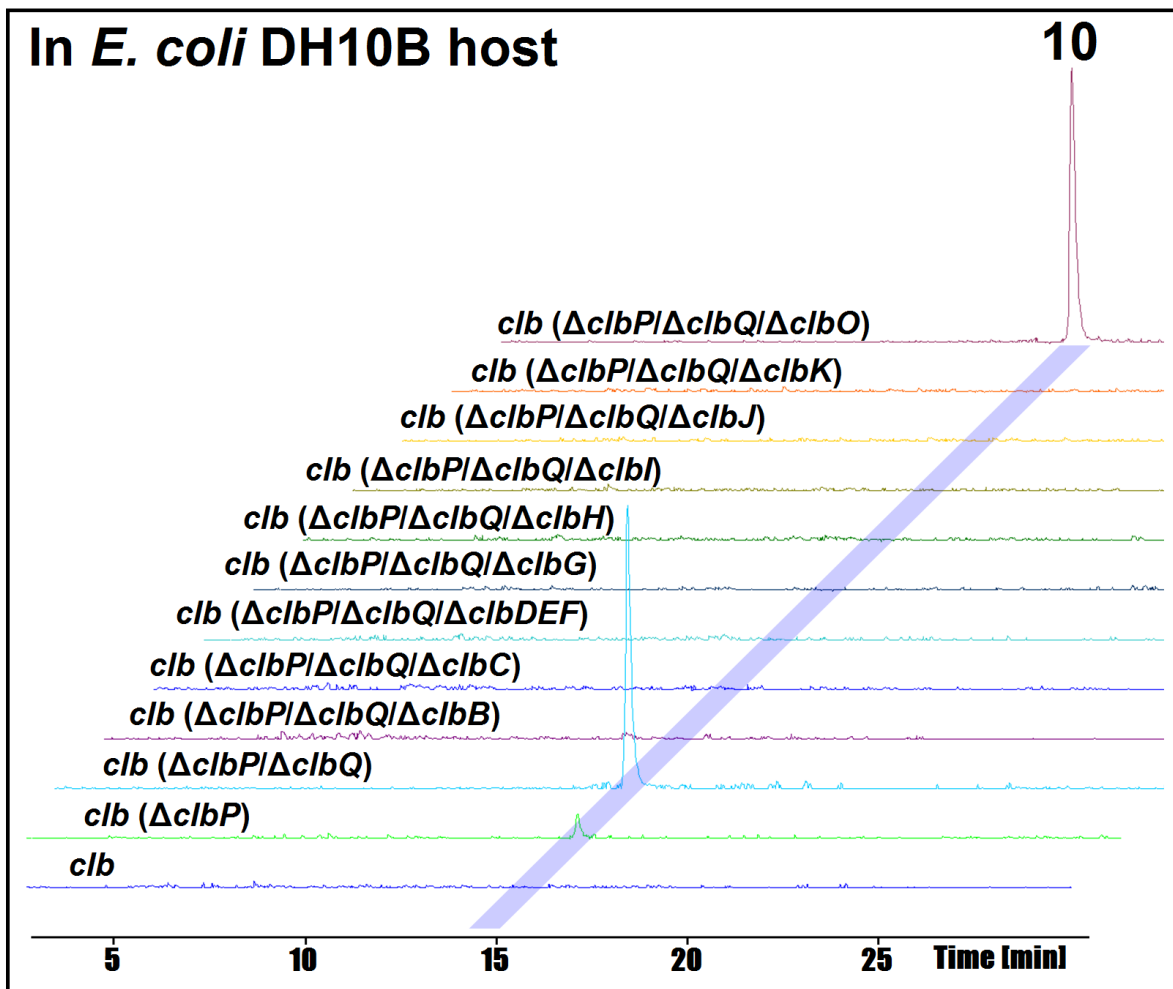


(d)

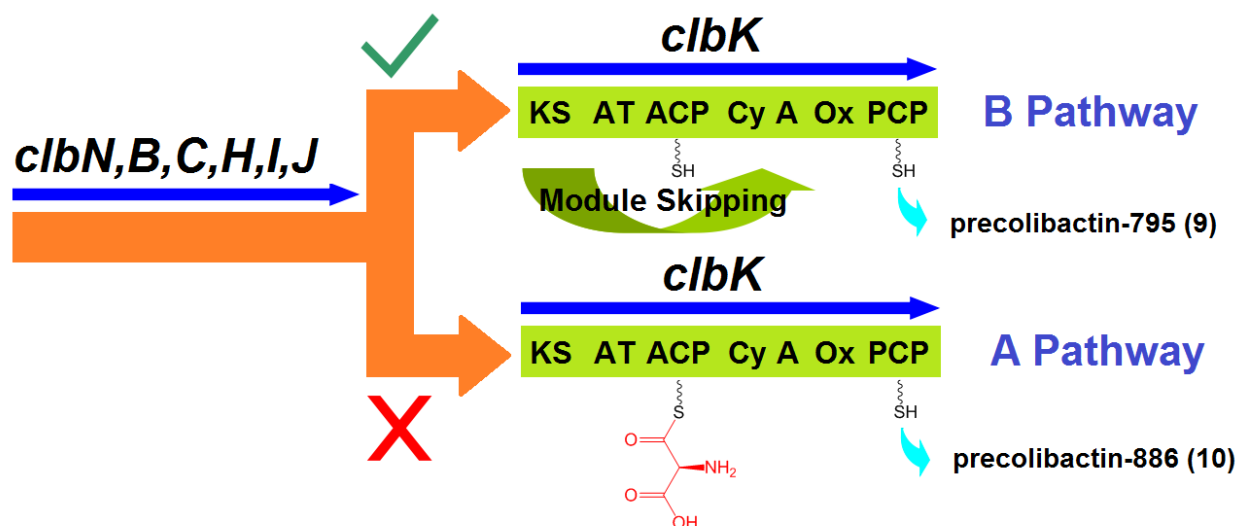
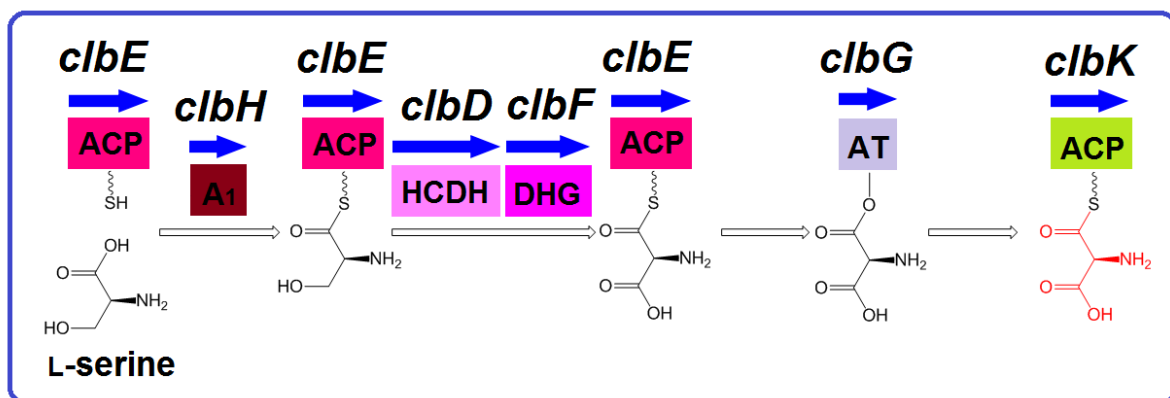
10 in this study	chemical shift (in DMSO-<i>d</i>₆)	30i in ref. 8	chemical shift (in CDCl₃)	14 in ref. 9	chemical shift (in CDCl₃)
C-23	107.8	C-2	107.3	C-2	105.5
C-36	107.6	C-5	108.5	C-5	109.9
C-37	160.4	C-4	165.4	C-4	165.5

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 rings⁷. 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

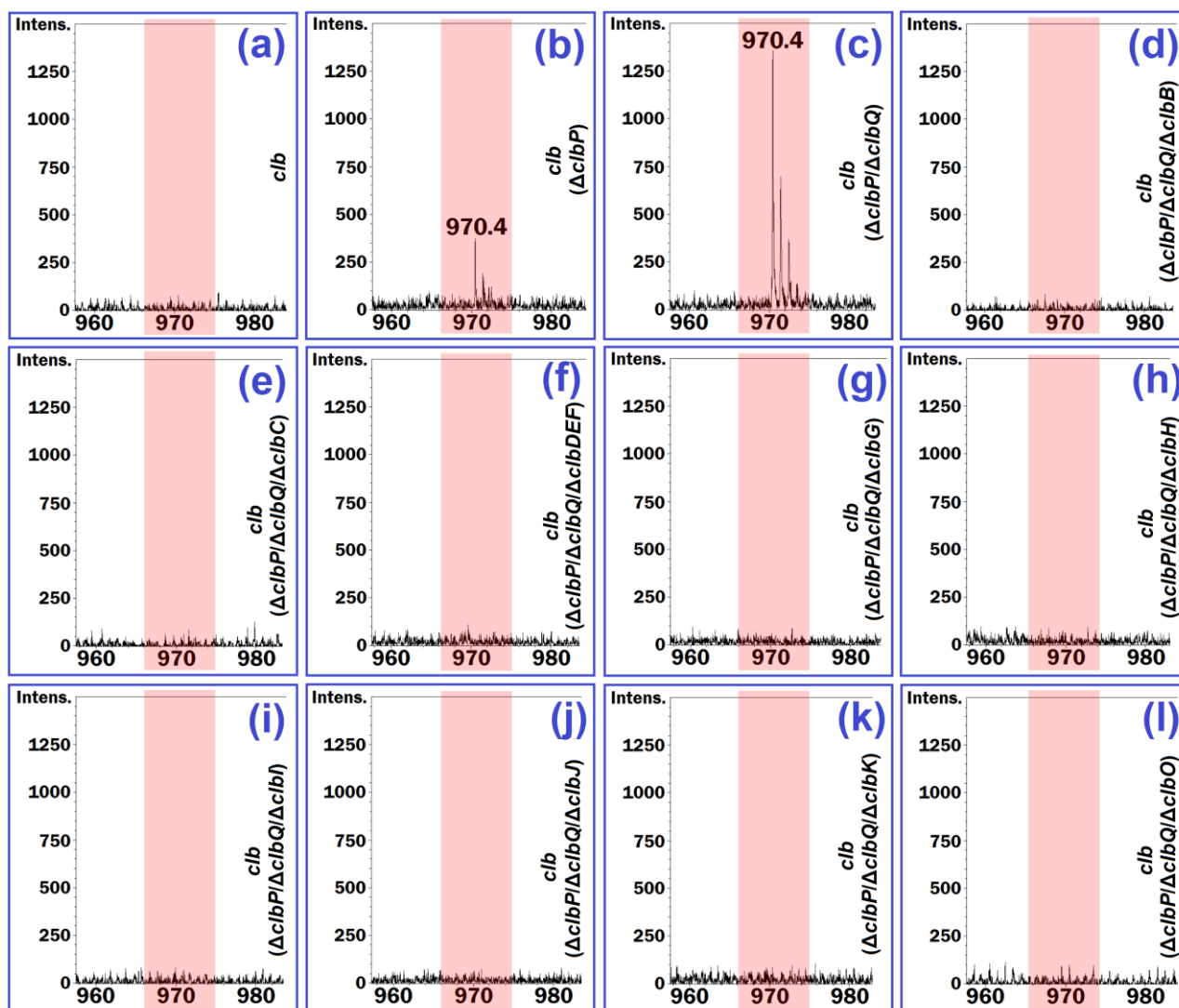
2,5-dihydro-5-hydroxyoxazoles shown in (b) and (c). (b) Nucleophilic addition of an imino anion to a ketone followed by a spontaneous intramolecular cyclization as a synthetic entry to a 2,5-dihydro-5-hydroxyoxazole compound (**30i** in ref. 8)⁸. (c) Preparation of a 2,5-dihydro-5-hydroxyoxazole compound (**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 compounds⁹. (d) Comparison of the ¹³C 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. $EIC^+ = 887.38 \pm 0.01$, 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**.

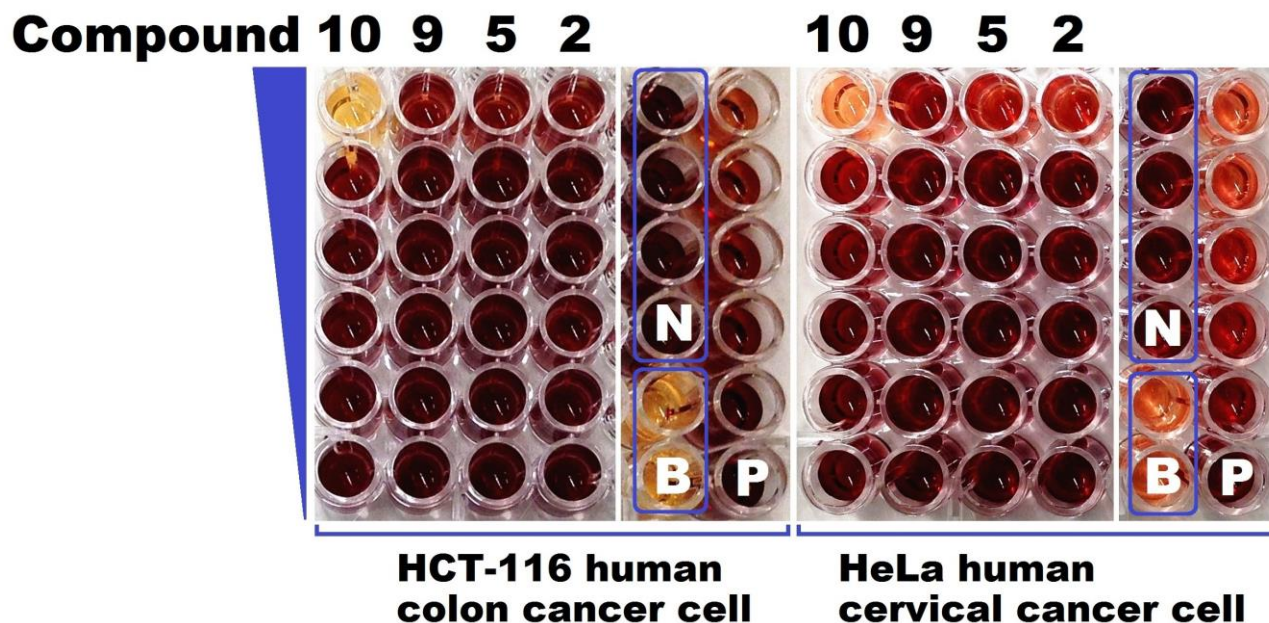


Supplementary Fig. 15. Inactivation of either *clbG* or *clbDEF* would prevent the generation and utilization of the aminomalonyl-ACP extender unit, and thus block path A and significantly redirect the overall pathway flux into path B.



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* ($\Delta clbP/\Delta clbQ$) completely abolished the production of **11**. The MS spectra obtained from the extracts of *E. coli* DH10B/pCAP01-*clb* ($\Delta clbP$) (b) and *E. coli* DH10B/pCAP01-*clb* ($\Delta clbP/\Delta clbQ$) (c) were used as positive controls, and that of *E. coli* DH10B/pCAP01-*clb* (a) was used as a negative control.

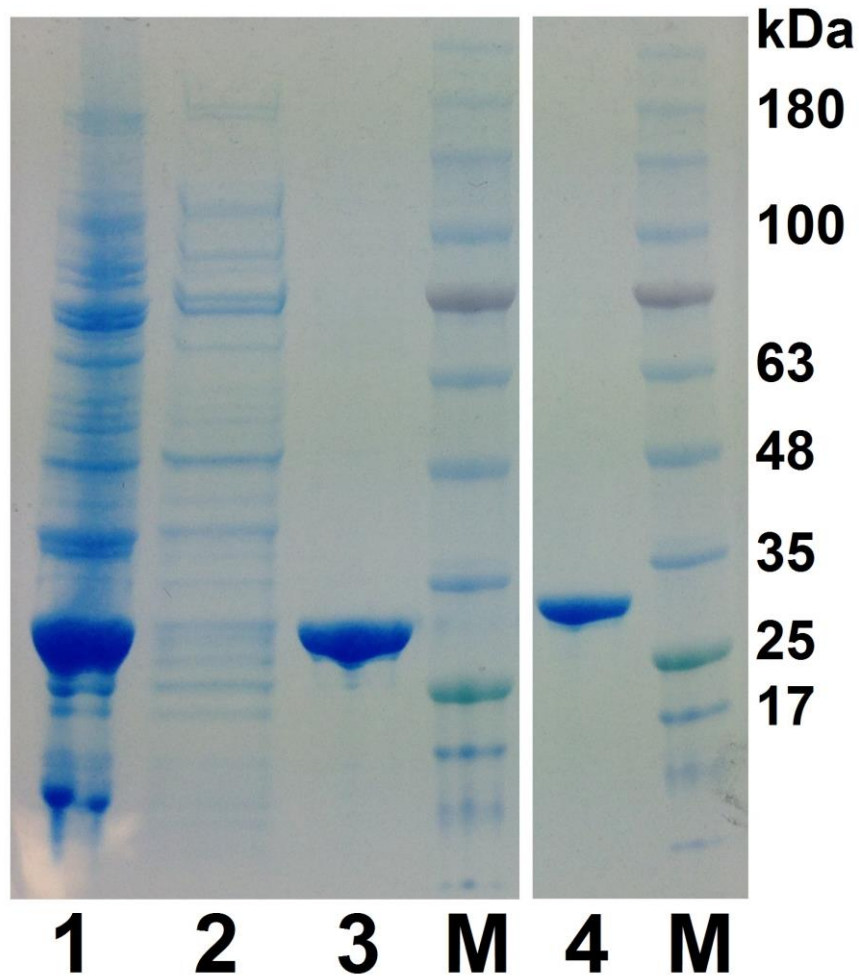
(a)



(b)

	IC₅₀ (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, IC₅₀ 2.0 ± 0.36 and 1.1 ± 0.19 μM against HCT-116 and HeLa cells, respectively). (b) IC₅₀ (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*-His₆-ClbQ expressed. Column 3: purified *N*-His₆-ClbQ. Column 4: purified *N*-His₆-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).