Supplementary Information

Macrocyclic colibactin induces DNA double-strand breaks via copper-mediated oxidative cleavage

Authors: Zhong-Rui Li[1,2], Jie Li[4,6], Wenlong Cai[2], Jennifer Y. H. Lai[1], Shaun M. K. McKinnie[4], Wei-Peng Zhang[1], Bradley S. Moore[4,5], Wenjun Zhang[2,3]* & Pei-Yuan Qian[1]*

Affiliations:

[1]Department of Ocean Science and Division of Life Science, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, China.

[2]Department of Chemical and Biomolecular Engineering, University of California, Berkeley, California 94720, USA.

[3]Chan Zuckerberg Biohub, San Francisco, California 94158, USA.

[4]Center for Marine Biotechnology and Biomedicine, Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California 92093, USA.

[5]Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego, La Jolla, California 92093, USA.

[6]Present address: Department of Chemistry and Biochemistry, University of South Carolina, Columbia, South Carolina, USA

*e-mail: boqianpy@ust.hk; wjzhang@berkeley.edu

1. Materials and Methods

Bacterial strains, plasmids and general materials. All bacterial strains and plasmids used in the present study are listed in Supplementary Table 4. All oligonucleotide primers (ordered from GenScript) used are listed in Supplementary Table 5 and the sequences of the homology arms used for recombination are underlined. All polymerase chain reactions (PCRs) were performed using Phusion polymerase (Thermo Scientific), 200 μM dNTPs (Thermo Scientific), and 0.5 μM of each oligonucleotide primer. Luria–Bertani (LB) medium (Affymetrix) for the cultivation of *E. coli* strains was purchased from Thermo Scientific. All antibiotics were purchased from Sigma-Aldrich, Inc., and were used for plasmid maintenance at the following concentrations: ampicillin (100 μg/mL), apramycin (50 μg/mL), chloramphenicol (25 μg/mL), and kanamycin (50 μg/mL). L-[1-^{13}C]serine and L-[^{15}N]serine were purchased from Cambridge Isotope Laboratories, Inc. for the feeding experiments. The plasmid pBR322 (Invitrogen) at a concentration of 1000 μg/mL was purchased from New England BioLabs for the DNA double-strand break assays. Catalase and neocuproine were purchased from Sigma-Aldrich, Inc., and superoxidase dismutase was purchased from MP Biomedicals, LLC..

Inactivation of *clbS* in the *E. coli* heterologous expression system and genetic complementation of the Δ*clbS* mutant with ClbS. Using λ Red–mediated recombination[1], the inactivation of the peptidase-encoding gene *clbP*, the thioesterase-encoding gene *clbQ*, and the resistance protein-encoding gene *clbS* was carried out to generate the *clb* Δ*clbP*Δ*clbQ*Δ*clbS* mutant (the three genes were adjacent to each other in the bacterial artificial chromosome (BAC) pCAP01-*clb*[2] (Supplementary Fig. 1a)). The PCR amplification product of the *aac(3)-IV* apramycin-resistance gene (*apra*^{R}) flanked with homologous arm was obtained by PCR using a pair of primers *clbPQS*-knockout-F and *clbPQS*-knockout-R from the plasmid pIJ773[3]. Through electroporation, the purified PCR product was transformed into the L-arabinose-induced *E. coli* BW25113 cells that carried both the BAC pCAP01-*clb* and the λ Red recombinase expression plasmid pIJ790. The homologous recombination between the PCR product and pCAP01-*clb* resulted in an in-frame deletion of *clbPQS* in pCAP01-*clb*. The recombinants were selected from an LB agar plate containing apramycin and kanamycin; the temperature-sensitive plasmid pIJ790 was eliminated by incubation at 37 °C. The mutant colonies were selected by colony PCR using the primers *clbPQS*-knockout check-F and *clbPQS*-knockout check-R. The construct was then examined by restriction analysis with KpnI and XhoI digestion (Supplementary Fig. 2a, b). The newly created BAC pCAP01-*clb* Δ*clbP*Δ*clbQ*Δ*clbS* was isolated from the *E. coli* BW25113 and introduced into the *E. coli* DH10B cells through electroporation to obtain the heterologous expression strain *E. coli* DH10B/pCAP01-*clb* Δ*clbP*Δ*clbQ*Δ*clbS*.

The genomic DNA of *E. coli* CFT073 was isolated as the template for the PCR amplification of *clbS*[4]. The PCR product was cloned into the expression vector pETDuet-1 to generate pETDuet-1-ClbS. The construct was identified and confirmed by PCR screening, restriction analysis, and sequencing, and was finally transformed into the competent *E. coli* DH10B cells harboring the pCAP01-*clb* Δ*clbP*Δ*clbQ*Δ*clbS* through electroporation. The transformed cell was grown on an LB agar plate containing ampicillin, apramycin and kanamycin at 37 °C overnight.

All strains, including *E. coli* DH10B/pCAP01-*clb* Δ*clbP*Δ*clbQ*[5], *E. coli* DH10B/pCAP01-*clb* Δ*clbP*Δ*clbQ*Δ*clbS*, and *E. coli* DH10B/pCAP01-*clb* Δ*clbP*Δ*clbQ*Δ*clbS::*pETDuet-1-ClbS, were inoculated into 10 mL of LB broth containing kanamycin (for Δ*clbP*Δ*clbQ* and Δ*clbP*Δ*clbQ*Δ*clbS* mutants) or ampicillin and kanamycin (for Δ*clbP*Δ*clbQ*Δ*clbS::*pETDuet-1-ClbS) and incubated for 1 day at 37 °C and 250 r.p.m. Using the UPLC–MS analysis method described below, the chemical profiles of these strains were recorded.

UPLC–MS-based analysis of *clb* biosynthetic pathway metabolites. For UPLC–MS-based analysis, a single colony of the strain tested was picked for culture. After incubation, the culture broth was extracted with an equal volume of ethyl acetate (EtOAc) twice. The organic phase extract was separated by centrifugation (10

min, 4,200 r.p.m.), evaporated, and finally resuspended in 100 μL of methanol (MeOH). The prepared extract was filtered through a 0.2-μm regenerated cellulose filter (Agilent Technologies, Inc.) before it was subjected to analysis by ultra-high-performance liquid chromatography–mass spectrometry (UPLC–MS).

For the UPLC–MS analysis, a 5 μL aliquot was injected onto a Waters BEH C18 reversed-phase UPLC column (1.7 μm, 150 mm × 2.1 mm inner diameter) and analyzed with a Bruker microTOF-Q II mass spectrometer (Bruker Daltonics GmbH) coupled with a Waters ACQUITY UPLC system (Waters ACQUITY) by a gradient elution (A, CH_3CN with 0.1% formic acid (FA); B, H_2O with 0.1% FA: 5% A over 2 min, 5–100% A from 2 to 22 min, 100% A from 22 to 25 min, and 100–5% A from 25 to 27 min; flow rate, 250 μL/min). The TOF-MS settings were as follows: acquisition, mass range *m*/*z* 200–2,000 Da, MS scan rate 1 s^{-1}; source, gas temperature 200 °C, gas flow 8 L/min; nebulizer 4 Bar, ion polarity positive; scan source parameters, capillary exit 140 V, skimmer 50 V. HyStar chromatography software (Bruker Daltonik) was used to control the system. Bruker Compass DataAnalysis 4.0 software (Bruker Daltonik) was used for data analysis.

Data analysis. The chemical profiles of the *E. coli* DH10B heterologous expression host harboring pCAP01-*clb* Δ*clbP*Δ*clbQ*, pCAP01-*clb* Δ*clbP*Δ*clbQ*Δ*clbS* or pCAP01-*clb* Δ*clbP*Δ*clbQ*Δ*clbS*::pETDuet-1-ClbS were recorded in quintuplicate ($N = 5$). The relative abundances of all the previously or newly identified precolibactin metabolites were compared, calculated and expressed as means ± standard deviation. All data analyses were carried out with version 17.0 of SPSS for Windows (SPSS Inc.).

Cloning and overexpression of the transcriptional regulator ClbR. The transcriptional regulator gene *clbR* was PCR amplified from *E. coli* CFT073 genomic DNA using the primers shown in Supplementary Table 5[6]. The PCR product was cloned into the expression vector pETDuet-1 to generate pETDuet-1-ClbR. The construct was identified and confirmed by PCR screening, restriction analysis, and sequencing of the purified plasmid DNA. Through electroporation, the correct construct was transformed into the *E. coli* DH10B competent cells harboring the pCAP01-*clb* Δ*clbP*Δ*clbQ*Δ*clbS* for coexpression. The transformed cells were grown on an LB agar plate containing ampicillin, apramycin, and kanamycin at 37 °C overnight. Then a single colony of the *E. coli* DH10B heterologous expression strain harboring both pCAP01-*clb* Δ*clbP*Δ*clbQ*Δ*clbS* and pETDuet-1-ClbR was picked from the LB agar plate and inoculated into 10 mL of LB broth containing ampicillin and kanamycin. The culture was incubated at 37 ºC on a rotary shaker (250 r.p.m.) for 1 day. Using the UPLC–MS analysis method described above, the chemical profile of the ClbR overexpression strain was recorded.

Inactivation of *clb* biosynthetic pathway genes in the *E. coli* heterologous expression system. Individual PKS-NRPS building block genes (including *clbB*, *clbC*, *clbH*, *clbI*, *clbJ*, *clbK*, and *clbO*)[7] and 2-aminomalonyl building block-related genes (including biosynthetic gene cassette *clbDEF* and *clbG*)[8,9] were disrupted in the newly created BAC pCAP01-*clb* Δ*clbP*Δ*clbQ*Δ*clbS*. Using λ Red–mediated recombination in the L-arabinose-induced *E. coli* BW25113 cells that harbored both pCAP01-*clb* Δ*clbP*Δ*clbQ*Δ*clbS* and plasmid pIJ790, each of these genes (or gene cassettes) was replaced by an ampicillin-resistance marker to generate the following nine derivative BACs: pCAP01-*clb* Δ*clbP*Δ*clbQ*Δ*clbS*Δ*clbB*, pCAP01-*clb* Δ*clbP*Δ*clbQ*Δ*clbS*Δ*clbC*, pCAP01-*clb* Δ*clbP*Δ*clbQ*Δ*clbS*Δ*clbH*, pCAP01-*clb* Δ*clbP*Δ*clbQ*Δ*clbS*Δ*clbI*, pCAP01-*clb* Δ*clbP*Δ*clbQ*Δ*clbS*Δ*clbJ*, pCAP01-*clb* Δ*clbP*Δ*clbQ*Δ*clbS*Δ*clbK*, pCAP01-*clb* Δ*clbP*Δ*clbQ*Δ*clbS*Δ*clbO*, pCAP01-*clb* Δ*clbP*Δ*clbQ*Δ*clbS*Δ*clbDEF*, and pCAP01-*clb* Δ*clbP*Δ*clbQ*Δ*clbS*Δ*clbG*. All of these BACs were isolated from the *E. coli* BW25113 cells and individually introduced into the *E. coli* DH10B cells through electroporation for heterologous expression. The chemical profiles of these *clb* pathway-related heterologous expression strains were analyzed with UPLC–MS as mentioned above.

Screening of potential novel *clb* biosynthetic pathway metabolites. *Clb*$^+$ heterologous expression strain *E. coli* DH10B/pCAP01-*clb* Δ*clbP*Δ*clbQ*Δ*clbS*Δ*clbO* was selected for a further screening of potential novel *clb* pathway metabolites. The culture of Δ*clbP*Δ*clbQ*Δ*clbS*Δ*clbO* mutant was prepared and then extracted as described above. The prepared extract was filtered through a 0.2-μm regenerated cellulose filter before it was subjected to analysis by UPLC–MS.

For the further screening, a new gradient elution was set using MeOH/H_2O as the elution mixture to increase the resolution: A, MeOH with 0.1% formic acid (FA); B, H_2O with 0.1% FA: 5% A for 2 min, 5–100% A from 2 to 52 min, 100% A from 52 to 57 min, and 100–5% A from 57 to 60 min; flow rate, 250 μL/min.

Fermentation and isolation of precolibactin-795a (8), precolibactin-795b (9), and precolibactin-969 (11). The *clb*$^+$ heterologous expression strain *E. coli* DH10B/pCAP01-*clb* Δ*clbP*Δ*clbQ*Δ*clbS*Δ*clbO* showed a higher yield of **8** and thus, was selected for production of this metabolite (Fig. 2c). A single colony of the Δ*clbP*Δ*clbQ*Δ*clbS*Δ*clbO* mutant was picked from a freshly streaked LB agar plate containing ampicillin, apramycin, and kanamycin and inoculated into 10 mL of LB medium containing the same three antibiotics in a sterile 50-mL Falcon tube. After an overnight culture at 37 °C and 250 r.p.m., 2 mL of the preculture was inoculated into 200 mL of LB medium containing ampicillin, apramycin, and kanamycin for overnight incubation a second time at 37 °C and 250 r.p.m.. Ten mL of overnight culture was added to 1 L of LB medium in a large Fernbach flask (2.8 L) containing only kanamycin. In total, 500 L of LB medium were cultured at 32 °C and 160 r.p.m. for 40 hours and were then extracted with EtOAc (2 × 500 L) to obtain a dried crude extract (60 g). The crude extract was fractionated on a flash chromatography system (2,500 g of Sepra C18 (50 μm) sorbent) and eluted with an MeOH–H_2O gradient (10:90, 30:70, 50:50, 80:20, and 100:0) to yield five fractions (F01–F05). UPLC–MS analyses showed that the fraction F04 (15 g) contained **8**; F04 was then further chromatographed over the same Sepra C18 column with isocratic elution (MeOH–H_2O 70:30) to obtain 10 subfractions (F0401–F0410). F0406 (4 g), which contained **8**, was subjected to three rounds of HPLC purification, all using the same semi-preparative C18 Phenomenex Luna column (5 μm, 250 mm × 10 mm inner diameter). In the first round of HPLC separation, F0406 was eluted with an CH_3CN–H_2O gradient mixture (A, CH_3CN with 0.1% FA; B, H_2O with 0.1% FA; 40% A over 5 min, and 40–80% A from 5 to 100 min) at a flow rate of 4.0 mL/min to obtain subfraction F040603 (t_R = 38–42 min; 500 mg), which was found to contain **8** based on UPLC–MS analysis. This subfraction was purified through HPLC again with an MeOH–H_2O gradient elution system (A, MeOH with 0.1% FA; B, H_2O with 0.1% FA: 60% A over 5 min, and 60–100% A from 5 to 120 min; flow rate, 2.5 mL/min) to obtain a further subfraction F04060305 (t_R = 58–59 min; 6.4 mg), which contained **8**. This subfraction was subjected to a final round of HPLC purification using an CH_3CN–H_2O elution system (A, CH_3CN with 0.1% FA; B, H_2O with 0.1% FA: 50% A over 5 min, and 50–90% A from 5 to 100 min; flow rate, 4.0 mL/min) to yield pure **8** (t_R = 19.7 min; 1.1 mg).

The *clb* Δ*clbP*Δ*clbQ*Δ*clbS*Δ*clbG* mutant was selected to produce **9**. The preparation and purification of **9** were conducted following a similar method to that for the production of **8**, and 4.5 mg of **9** was obtained from 50 L of the *clb* Δ*clbP*Δ*clbQ*Δ*clbS*Δ*clbG* mutant fermentation broth. After one round of purification using a Sepra C18 (50 μm) flash chromatography system, three rounds of HPLC separation were performed to obtain pure **9** for an NMR test. In the first round, the parameters were set as follows: CH_3CN–H_2O gradient mixture: A, CH_3CN with 0.1% FA; B, H_2O with 0.1% FA; 40% A over 5 min, and 40–80% A from 5 to 100 min; flow rate, 4.0 mL/min; F040603 (35 mg, t_R = 38–42 min) was collected. In the second round, the parameters were set as follows: MeOH–H_2O gradient mixture: A, MeOH with 0.1% FA; B, H_2O with 0.1% FA: 60% A over 5 min, and 60–100% A from 5 to 120 min; flow rate, 2.5 mL/min; F04060307 (6.3 mg, t_R = 62–63 min) was collected. In the third round, the parameters were set as follows: CH_3CN–H_2O gradient mixture: A, CH_3CN with 0.1% FA; B, H_2O with 0.1% FA; 50% A over 5 min, and 50–90% A from 5 to 100 min; flow rate, 4.0 mL/min; **9** (4.5 mg, t_R = 19.5 min) was collected.

All of the *clb* PKS/NRPS building block genes and the 2-aminomalonyl building block-related genes were indispensable to the biosynthesis of **11** (Fig. 2a). Therefore, the heterologous expression strain *E. coli* DH10B that harbored pCAP01-*clb* Δ*clbP*Δ*clbQ*Δ*clbS* was selected for the production of **11**. A large-scale fermentation was performed, following a similar method to that as mentioned above. From 2,000 L of the fermentation broth of the Δ*clbP*Δ*clbQ*Δ*clbS* mutant, 230 g of crude extract was collected. All the crude extract was subjected to a flash chromatography system (2,500 g of Sepra C18 (50 μm) sorbent) with two rounds of elution using an MeOH–H_2O gradient elution (eluted with MeOH–H_2O, 10:90, 30:70, 50:50, 80:20 and 100:0) and an MeOH–H_2O isocratic elution (eluted with MeOH–H_2O, 65:35), to obtain a subfraction F0408 (24 g) containing **11**. The subfraction F0408 was purified using semi-preparative HPLC with five rounds of purification. In the first round, the parameters were set as follows: CH_3CN–H_2O gradient mixture: A, CH_3CN with 0.1% FA; B, H_2O with 0.1% FA; 40% A over 5 min, and 40–80% A from 5 to 100 min; flow rate, 4.0 mL/min; F040804 (2.5 g, t_R = 34–38 min) was collected. In the second round, the parameters were set as follows: MeOH–H_2O gradient mixture: A, MeOH with 0.1% FA; B, H_2O with 0.1% FA: 50% A over 5 min, and 50–100% A from 5 to 120 min; flow rate, 3.0 mL/min; F04080405 (450 mg, t_R = 60–63 min) was collected. In the third round, the parameters were set as follows: CH_3CN–MeOH–H_2O gradient mixture: A, CH_3CN/MeOH (1:1, vol/vol) with 0.1% FA; B, H_2O with 0.1% FA; 40% A over 5 min, and 40–90% A from 5 to 120 min; flow rate, 4.0 mL/min; F0408040507 (25 mg, t_R = 48–52 mim) was collected. In the fourth round, the parameters were set as follows: MeOH–H_2O gradient mixture: A, MeOH with 0.1% FA; B, H_2O with 0.1% FA; 50% A over 5 min, and 50–85% A from 5 to 140 min; flow rate, 3.0 mL/min; F040804050706 (867 μg, t_R = 96–101 min) was collected. In the fifth round, the parameters were set as follows: CH_3CN–H_2O gradient mixture: A, CH_3CN with 0.1% FA; B, H_2O with 0.1% FA; 35% A over 5 min, and 35–60% A from 5 to 180 min; flow rate, 4.0 mL/min; **11** was collected (50 μg, t_R = 151–155 min). Similar to precolibactin-886 (**10**)[5], **11** present as an approximately equal mixture of two isomers which were collected together.

High-resolution ESI-MS, MS/MS, and NMR characterization of 8, 9, and 11. All of the *clb* pathway metabolite samples were dried and resuspended in MeOH. Then, 20 μL of the suspension were injected onto an Agilent Technologies 6520 Accurate-Mass Q-TOF LC–MS instrument fitted with an Agilent Eclipse Plus C18 column (4.6 × 100 mm) for LC–HRMS analysis. A linear gradient elution was performed using CH_3CN–H_2O (A, CH_3CN with 0.1% FA; B, H_2O with 0.1% FA: 40% A over 2 min, and 40–100% A from 2 to 18 min, at a flow rate of 0.5 mL/min). LC–HRMS/MS analyses were performed using the same program described above with the target mass *m*/*z* 796.37 $[M + H]^+$ and collision energy of 30 V or 45 V around the retention time 12.6 ± 0.5 min for **8**; with the target mass *m*/*z* 796.35 $[M + H]^+$ and collision energy of 30 V or 45 V around the retention time 12.5 ± 0.5 min for **9**; with the target mass *m*/*z* 887.38 $[M + H]^+$ and collision energy of 15 V or 30 V around the retention time 12.2 ± 0.5 min for **10**; and with the target mass *m*/*z* 970.38 $[M + H]^+$ and collision energy of 15 V or 30 V around the retention time 12.1 ± 0.5 min for **11**. All MS data were analyzed using Agilent MassHunter Qualitative Analysis software.

^1H, ^{13}C, ^1H–^1H COSY, ^1H–^{13}C HSQC, and ^1H–^{13}C HMBC NMR spectra for **8**, ^1H, ^{13}C, ^1H–^1H COSY, ^1H–^1H NOESY, ^1H–^{13}C HSQC, ^1H–^{13}C HMBC, and ^1H–^{15}N HSQC NMR spectra for **9**, and ^1H, ^1H–^1H COSY, ^1H–^1H TOCSY, ^1H–^1H NOESY, ^1H–^{13}C HSQC, and ^1H–^{13}C HMBC NMR spectra for **11** were acquired on a Bruker Ascend 850 NMR spectrometer (850 MHz for ^1H and 212.5 MHz for ^{13}C) equipped with a cryoprobe. For all of the NMR tests, the samples were dissolved in DMSO-d_6 (DMSO-d_6; Cambridge Isotope Laboratories) and loaded into 2.5 mm NMR tubes. Data were collected and reported as follows: chemical shift, integration multiplicity (s, singlet; d, doublet; t, triplet; m, multiplet), and coupling constant. Chemical shifts were reported using the DMSO-d_6 resonance as the internal standard for ^1H-NMR DMSO-d_6: δ = 2.50 ppm and ^{13}C-NMR DMSO-d_6: δ = 39.6 ppm.

Feeding experiments for biosynthetic pathway study. The newly generated heterologous expression strain *E.*

coli DH10B/pCAP01-*clb* Δ*clbP*Δ*clbQ*Δ*clbS* was selected to inoculate 5 mL of LB medium containing apramycin and kanamycin, and incubated at 37 °C on a shaker (250 r.p.m.) overnight. The next day, 100 μL of seed culture were used to inoculate 10 mL of LB medium supplemented with kanamycin and 0.25 mg/mL isotope-labeled L-[1-^{13}C]serine or L-[^{15}N]serine in a 50-mL sterile Falcon tube, and incubated at 37 °C with shaking (250 r.p.m.) for 1 day. After incubation, each culture supernatant was extracted with an equal volume of EtOAc twice. The organic layer was separated by centrifugation at 4,200 r.p.m. for 10 min, collected, and evaporated to dryness. All of the prepared extract samples were redissolved in 20 μL of MeOH for UPLC–MS analysis following the method described above.

Amino acid homology analysis of ClbK$_{PKS}$ and ClbO. The amino acid sequences of ketosynthase (KS) domains of ClbK$_{PKS}$, ClbO, and 78 other polyketide synthases (PKSs) were collected (both *cis*-acyltransferase PKS and *trans*-acyltransferase PKS were involved; the PKS domains of gene clusters of some classical thiazole-containing secondary metabolites were selected). Their accession numbers and amino acid sequences used for homology analysis are listed in Supplementary Tables 6 and 7, respectively.

Phylogenetic analysis to visualize relationships among the KS domains of ClbK$_{PKS}$, ClbO, and closely related PKSs was performed with version 7.0 of Molecular Evolutionary Genetics Analysis (MEGA) using the maximum-likelihood method based on the JTT matrix-based model[10]. The analysis of 80 amino acid sequences was conducted with a total of 390 aligned amino acid positions and 500 bootstrap replicates.

Construction and expression of peptidase ClbP. The peptidase gene *clbP* was amplified by PCR from the genomic DNA of *E. coli* CFT073 and then cloned into pET28a(+) expression vector by using appropriate restriction sites (Supplementary Table 5). The construct was identified and confirmed by PCR screening, restriction analysis, and sequencing and was then transformed into the competent *E. coli* BL21 (DE3) for recombinant protein expression.

Cleavage of precolibactin-969 (11) and precolibactin-886 (10). To examine if peptidase ClbP cleaves **11** or **10** into the corresponding mature colibactin-645 (**13**) or **15**, respectively, whole cell assays were performed. Ten mL culture of the *E. coli* BL21 (DE3) harboring pET-28a-ClbP was inoculated from a single colony and grown in LB medium supplemented kanamycin at 37 °C. At an OD_{600} of 0.5, 100 μL of this culture were transferred into a 500-μL tube. One μL of **11** dissolved in DMSO was then added into the culture to a final concentration of 1 μM. After incubation for 2 hours at 37 °C with shaking at 250 r.p.m., the cleavage reaction of **11** by ClbP was quenched by the addition of 100 μL of MeOH. After centrifugation, the supernatant was collected and lyophilized. The lyophilized powder was extracted with 100 μL MeOH by vortexing the mixture. The sample was then centrifuged and the supernatant was analyzed by UPLC–MS with the method mentioned above.

Meanwhile, the same reaction procedure was followed with the treatment of metal chelators. To remove the trace metals, 1) all of LB medium used in the assay was pre-treated with Chelex-100; 2) the *E. coli* BL21 cells harboring pET-28a-ClbP were washed once prior to the addition of compound; 3) ethylenediaminetetraacetic acid (EDTA) at a final concentration of 500 μM was added into the culture during the *in vivo* reaction. Following the same protocol for cleaving **11**, the ability of peptidase ClbP to cleave **10** was examined.

Isolation and identification of colibactin-645 (13) and 15. The **13** and **15** as the respective cleaved products of **11** and **10** by peptidase ClbP (as mentioned above; with the treatment of metal chelators) were collected. To purify **13**, the lyophilized powder was loaded into a flash chromatography system with Sepra C18 sorbent and eluted with an MeOH–H_2O gradient to afford fractions. F02 (from 25% MeOH elution) was subsequently subjected to semi-preparative HPLC with two rounds of purification. In the first round, the parameters were set

as follows: $CH_3CN–H_2O$ gradient mixture: A, CH_3CN with 0.1% FA; B, H_2O with 0.1% FA; 5% A over 5 min, and 5–30% A from 5 to 120 min; flow rate, 4.0 mL/min; F0226 (1.2 mg, t_R = 55–59 min) was collected. In the second round, the parameters were set as follows: MeOH–H_2O gradient mixture: A, MeOH with 0.1% FA; B, H_2O with 0.1% FA: 5% A over 5 min, and 5–35% A from 5 to 180 min; flow rate, 3.0 mL/min; **13** (20 μg, t_R = 105–108 min) was collected. The purification of **15** was conducted, following a similar method to that for purifying **13**. Briefly, F02 (from 25% MeOH elution) was collected from Sepra C18 sorbent separation; F0224 (1.8 mg, t_R = 56–60 min) was collected from the first round of HPLC separation; and **15** (80 μg, t_R = 109–112 min) was collected from the second round of HPLC separation.

Both **13** and **15** were identified by analyzing their HR-MS/MS fragmentation patterns. Twenty μL of each sample were injected onto an Agilent Technologies 6520 Accurate-Mass Q-TOF LC–MS instrument fitted with an Agilent Eclipse Plus C18 column. A linear gradient elution was performed using $CH_3CN–H_2O$ (A, CH_3CN with 0.1% FA; B, H_2O with 0.1% FA: 2% A over 2 min, and 2–35% A from 2 to 25 min, at a flow rate of 0.5 mL/min). LC–HRMS/MS analyses were performed with the target mass *m*/*z* 646.14 $[M + H]^+$ and collision energy of 15 V around the retention time 13.6 ± 0.5 min for **13**, and with the target mass *m*/*z* 563.14 $[M + H]^+$ and collision energy of 15 V around the retention time 13.8 ± 0.5 min for **15**.

Investigation of metabolites from native *clb*$^+$ *E. coli* strain. Glycerol stock containing the native *clb*$^+$ strain *E. coli* CFT073 was streaked on an LB agar plate and grown for 1 day at 37 °C. A resulting colony was picked up and inoculated into 20 mL of LB medium for overnight incubation at 37 °C and 250 r.p.m.. Ten mL of overnight culture were added to 1 L of LB medium in a large Fernbach flask. In total, 2 L of LB medium were cultured at 37 °C and 160 r.p.m.. Following the incubation for 2 days, the cells were harvested by centrifugation at 4 °C (20 min, 4,800 r.p.m.). The pellets were washed twice with PBS and then extracted twice with MeOH by ultrasonic treatment. The dried extract was saved as the crude cellular extract. The supernatant from the centrifugation step was filtered aseptically using sterile cellulose pyrogen free disposable filters of 0.2 μm pore size. The cell-free broth was then transferred to a new sterile flask with a polypropylene mesh bag containing 100 g of washed Amberlite XAD16N resin (Sigma-Aldrich Co.) and incubated on a rotary shaker (220 r.p.m.) for 8 hours at room temperature. Residual broth was pressed out of the resin bag and the resin was extracted with 40% of MeOH. After evaporation, the dried extract was saved as the crude cell-free extract.

Both the cellular extract and cell-free extract were individually chromatographed over Sepra C18 sorbent and then purified by semi-preparative HPLC with the same separation conditions used in the separation of **13** (as mentioned above). Briefly, the crude extract was loaded into the Sepra C18 sorbent column, the fraction F02 eluted by 25% of MeOH was collected; F02 was purified by HPLC with an $CH_3CN–H_2O$ gradient elution system, the subfraction F0202 with retention time of 55–59 min was collected; then F0202 was purified by HPLC again with an MeOH–H_2O gradient elution system, the subfraction F020202 with retention time of 105–108 min was collected. F020202CE and F020202CFE were obtained from the crude cellular extract and cell-free extract, respectively, and analyzed by UPLC–MS as mentioned above. Additionally, both the cellular and cell-free extracts from the *E. coli* CFT073 *clb*$^-$ mutant were prepared and analyzed using a similar method as mentioned above.

Effect of metal chelators on the yields of precolibactins. A single colony of *E. coli* DH10B/pCAP01-*clb* Δ*clbP*Δ*clbQ*Δ*clbS* was picked up and inoculated into 10 mL of LB medium containing apramycin and kanamycin. Following an overnight incubation at 37 °C and 250 r.p.m., 100 μL of the preculture were inoculated into 10 mL of LB medium containing kanamycin with or without metal chelators (2 mM EDTA or 2 mM D-DDC, or 1 mg Chelex-100). After incubation for 2 days at 37 °C, each culture broth was individually extracted with an equal volume of EtOAc twice. All of extracts were analyzed by UPLC–MS as mentioned above. To examine the yields of various precolibactins, the ion chromatogram traces corresponding to **2**, **5**, **7**, and **11** were extracted and compared. These four metabolites represent linear, aza-spirocyclopropane,

thiazole-containing, and aminomalonate-containing precolibactins, respectively.

Plasmid DNA cleavage assays. The ability of colibactin-645 (**13**) to cause DNA double-strand breaks (DSBs) *in vitro* was assessed using the plasmid pBR322 DNA by following the conversion of supercoiled DNA (Form I) to the open circular (Form II) and linear DNA (Form III) and using agarose gel electrophoresis to separate the cleavage products[11,12]. In each 15-μL of reaction mixture, 0.15 μL of **13** dissolved in dimethyl sulfoxide (DMSO) were added to a solution of 0.5 μL pBR322 DNA (250 ng) in 10.0 mM Tris-HCl buffer (pH 7.4) with or without Cu(II) to achieve the final concentration of 15 μM of **13**. Cupric chloride was added to the mixture of pBR322 DNA and **13** to reach the final concentration of Cu(II) ions of 3 μM or 30 μM. The negative control was prepared by adding 0.15 μL of DMSO instead of **13** to the sample solution. All of these reactions were incubated for 12 hours at 37 °C in a darkroom. After incubation, all of the reaction samples were individually mixed with 6X DNA loading dye (3 μL) and separated over a 0.8% agarose gel stained with SYBR Safe DNA Gel Stain. The electrophoresis was carried out at room temperature in 1X-TAE buffer at 60 V for 120 min. After electrophoresis, the gel was viewed with a CCD camera gel documentation system (ChemiDoc, Bio-Rad) and photographed (Fig. 4a). The ability of precolibactin-969 (**11**) to induce DSBs *in vitro* with or without metal ions (Cu(II), Fe(III) or Fe(II)) was examined by following a similar method (described in the legend to Supplementary Figure 15).

The effect of the Cu(II)-mediated DNA cleavage of pBR322 DNA by **11** was further evaluated in the presence of a specific Cu(I) chelator neocuproine (1 mM)[13] or different reductants, including β-mercaptoethanol (β-ME) (5 mM) or dithiothreitol (DTT) (5 mM). The Cu(I) chelator neocuproine or reductants were added alternatively to the reaction mixture of **11** (15 μM), pBR322 DNA (250 ng), and Cu(II) (30 μM). After incubation at 37 °C for 4 hours in a darkroom, the effectiveness of different reagents at preventing or promoting DNA breakage was observed using electrophoresis as described above (Fig. 4c).

To explore the mechanism underlying the Cu(II)-mediated DNA DSBs, DNA cleavage assays in the presence of various reactive oxygen species (ROS) scavengers were conducted. The ROS scavengers selected in the current study were as follows: potassium iodide (KI) which is a hydrogen peroxide (H_2O_2) scavenger [at a final concentration of 1 mM][14]; catalase which mediates the decomposition of H_2O_2 [at a final concentration of 0.1 mg/mL][12]; superoxide dismutase (SOD) which catalyzes the conversion of the superoxide radical into H_2O_2 [at a final concentration of 10 units][12]; and mannitol [at a final concentration of 50 mM] and DMSO [at a final concentration of 10%, v/v], both of which are hydroxyl radical scavengers[12,15]. These ROS scavengers were added alternatively to the reaction mixture of **11** (15 μM), pBR322 DNA (250 ng), and Cu(II) (30 μM). All of reactions were incubated for 12 hours at 37 °C in a darkroom and then analyzed using agarose gel electrophoresis as described above (Fig. 4d).

Freifelder–Trumbo analysis. Using the plasmid DNA cleavage assay method described above, the ratio of the number of single-strand breaks (SSBs) (n_1) to the number of double-strand breaks (DSBs) (n_2) per DNA molecule was calculated using the Freifelder–Trumbo relationship[12,16]. A time-course assay was conducted using the samples of pBR322 plasmid DNA (250 ng) that were treated with 15 μM **11** in the presence of 30 μM Cu(II) in 10.0 mM Tris-HCl buffer (pH 7.4) for various durations (1 h, 2 h, 4 h, 6 h, 8 h, 10 h, and 12 h) (Fig. 4b). Following incubation at 37 °C in a darkroom, Form I (supercoiled), Form II (nicked), and Form III (linearized) plasmid DNA in the reaction mixture were separated by agarose gel (0.8%, w/v) electrophoresis following the method described above; and their amounts were quantitated using Quantity One 1-D analysis software (Supplementary Figure 16b). The said ratio was finally calculated according to the equations in Supplementary Figure 16b.

Detection of Cu(II) reduction. The concentration of free Cu(I) was determined spectrophotometrically using a specific Cu(I) chelator neocuproine to evaluate the ability of precolibactin-886 (**10**) or precolibactin-969 (**11**)

to reduce Cu(II), with or without plasmid pBR322 DNA (500 ng). Briefly, 150 μM **11** or **10**, 300 μM cupric chloride, and 1 mM neocuproine in 10.0 mM Tris-HCl (pH 7.4) were used in combination experiments. Following a 16-hours incubation at 37 °C in a darkroom, the amount of Cu(I) was determined by measuring the visible absorbance at 450 nm of the cuprous-neocuproine complex using a NanoDrop (*ND*-8000, 8-Sample Spectrophotometer)[17,18]. The final Cu(I) concentrations were determined through comparison to a standard curve; and β-mercaptoethanol was selected as a bioreductant for establishing the standard curve. Additionally, in order to characterise the ability of **10** to reduce Cu(II), the concentration of Cu(I) was recorded at different time points (0.5 h, 1 h, 2 h, 4 h, 6 h, 8 h, 12 h, and 16 h). Meanwhile, the same admixture without neocuproine was directly injected into the mass spectrometer to record the disappearance of **10** and to detect the copper-bound colibactin species.

Determination of the binding constant. To calculate the binding constant of Cu(II) with macrocyclic colibactin, a modified titration experiment was used[19]. Upon gradual addition of Cu(II), the intensity of precolibactin-886 (10) was measured by using the mass spectrometer method mentioned above. The binding constant K_a was determined from the Benesi–Hildebrand plot of $I_0/(I_0 - I)$ against $[Cu(II)]^{-1}$, where I_0 and I are the intensities before and after the addition of the Cu(II). The binding constant was calculated from the ratio of intercept/slope[20].

Cell culture. HeLa cells, human normal colon fibroblast CCD-112 CoN cells, and colorectal cancer HCT-116 cells were provided by the Cell Culture Facility at the University of California, Berkeley. Human normal colon epithelial FHC cells were obtained from the ATCC (Manassas, VA). Cells were grown in the Cell Culture Facility at the University of California, Berkeley, with expert technical assistance. HeLa cells, HCT-116 cells, and CCD-112 CoN cells were cultured in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 10% fetal bovine serum. FHC cells were cultured in DMEM:F12 Medium with 10% fetal bovine serum. All of cells were maintained at 37 °C and 5% CO_2.

Immunofluorescence. HeLa cells and FHC cells were seeded on glass coverslips prior to the treatment of compounds. Cells were treated with 50 nM of **13** or 500 nM of **15** for 4 hours at 37 °C in a darkroom. The mouse monoclonal anti-phospho-histone H2A.X (Ser139) (EMD Millipore Corporation) and rabbit polyclonal anti-53BP1 (Novus Biologicals) were employed as the primary antibodies; the Goat anti-Mouse IgG (Alexa Fluor 488, Life Technologies) and Goat anti-Rabbit IgG (Alexa Fluor 647, Life Technologies) were employed as the secondary antibodies. Prior to immunostaining, cells were fixed in solution containing 4% paraformaldehyde for 15 min at room temperature, permeabilized with 0.25% Triton X-100 solution for 20 min at 4 °C, and blocked with 3% bovine serum albumin in PBS for 10 min at room temperature. Subsequently, cells were incubated with the primary antibodies (1:100) for 1 hour at 37 °C in dark, washed with PBS, and then incubated with the secondary antibodies (1:100) for 1 hour at 37 °C in dark. Coverslips were washed four times with PBS, then mounted onto glass slides with Vectashield mounting medium containing DAPI (Vector Laboratories) to counterstain nuclear DNA and visualized with a Zeiss Axio Imager M1 fluorescence microscope. Filters for DAPI, GFP, and Cy5 were used to acquire images.

Neutral comet assays. The DNA damage was evaluated by using the Reagent Kit for Single Cell Gel Electrophoresis Assay/CometAssay (Trevigen), according to the manufacturer's instructions. Briefly, HeLa cells were treated with **13** at different concentrations (20 nM, 50 nM, or 200 nM) for 4 hours at 37 °C in dark. Following treatment, the cells were harvested in cold PBS at a density of 1×10^5 cells/mL. An aliquot of this cell suspension was mixed with pre-warmed LM Agarose (37 °C) at a 1:10 ratio (v/v) and immediately pipetted onto a comet slide. Comet slides were immersed in cold lysing solution for 1 hour, followed by an incubation in cold 1X neutral electrophoresis buffer for 30 min. Electrophoresis was then carried out for 30

min at 15V (~1V/cm) at 4 °C. DNA was stained in a 1:10,000 dilution of SYBR Gold (Molecular Probes). Cells were photographed using a Zeiss Axio Imager M1 fluorescence microscope.

To investigate the prevention of DNA damage in cells by metal chelators, HeLa cells were either treated with **13** (50 nM) or infected by the *clb*$^{+}$ *E. coli* CFT073 (multiplicity of infection (MOI) = 100), then either ethylenediaminetetraacetic acid (EDTA) or bathocuproinedisulfonic acid (BCS)[21] was added into the cell culture at a final concentration of 2.5 mM EDTA or 2 mM BCS. Following the incubation (4 hours at 37 °C), infected cells with or without the treatment of metal chelators were washed three times with PBS. To quantify the DNA damage levels in these cells, all of cells were collected with trypsinization, then the neutral comet assay was preformed. On each slide, at least 50 cells were chosen randomly for examine. Tail moment which is positively correlated with the extent of DNA breakage in a cell was used as an index of DNA damage. [Tail moment combines a measure of length of the comet tail and the proportion of DNA to migrate into the tail]. In the controls, cell cultures with the treatment of EDTA or BCS alone in the same condition did not show measurable difference comparing to the control of DMSO.

2. Supplementary Text

Structure elucidation of precolibactin-795b (9)

The molecular formula of **9** was determined as $C_{39}H_{53}N_7O_7S_2$ by high-resolution mass spectrometry (HRMS) on the basis of its protonated molecular ion peak at *m*/*z* 796.3520 (calculated, 796.3521) (Supplementary Fig. 9b). In the current study, the molecular formula of **9** was consistent with that of the previously reported synthetic precolibactin C[22]. We initially detected **9** from the ethyl acetate extracts of *clb*$^+$ heterologous expression strain *E. coli* DH10B/pCAP01-*clb* Δ*clbP*. By comparing its MS/MS fragmentation with that of precolibactin-712 (**7**) which had been identified by nuclear magnetic resonance (NMR) spectroscopy, we proposed its bithiazole-containing structure[2] (Supplementary Fig. 1). Then Balskus and co-workers provided the first ^{1}H NMR spectrum of this bithiazole-containing metabolite, which showed two proton resonances at δ_H 8.16 and 8.28 ppm, and implied the presence of a bithiazole moiety compared with one thiazole proton resonance in the ^{1}H NMR spectrum of **7**[9]. However, they did not provide any two-dimensional NMR spectra to confirm the exact structure of this natural bithiazole-containing *clb* pathway metabolite. Subsequently, Herzon and co-workers synthesized this bithiazole-containing compound (named precolibactin C), based on the proposed structure[22]. Considering that two similar *clb* pathway-related metabolites **8** and **9** were detected, we performed thorough analysis to first confirm the chemical structure of **9**.

Using an ethyl acetate extract prepared from 50 L of the fermentation broth of *E. coli* DH10B/pCAP01-*clb* Δ*clbP*Δ*clbQ*Δ*clbS*Δ*clbG*, we obtained 4.5 mg of metabolite **9** which appeared as a white and amorphous powder. A comparison of the 1D and 2D NMR spectra of **9** and **7** indicated that **9** contained the same *N*-myristoyl-D-asparagine residue and 1*H*-pyrrolo[3,4-*c*]pyridine-3,6(2*H*,5*H*)-dione unit as did **7** (Fig. 1, Supplementary Fig. 10, Supplementary Table 2). However, unlike the ^{1}H–^{13}C HSQC spectrum of **7**, the ^{1}H–^{13}C HSQC spectrum of **9** showed two thiazole proton resonances at δ_H 8.33/δ_C 127.8 (H-34/C-34) and δ_H 8.28/δ_C 119.0 (H-37/C-37) (Supplementary Fig. 10f). The ^{1}H–^{13}C HMBC correlations from H-34 to C-33 (δ_C 161.7) and C-35 (δ_C 150.2), from H-37 to C-36 (δ_C 166.7) and C-38 (δ_C 147.2) indicated the assigned bithiazole moiety (Supplementary Fig. 10g). Herein we confirmed the chemical structure of the natural **9**.

Structure elucidation of precolibactin-795a (8)

The molecular formula of **8** was determined as $C_{39}H_{53}N_7O_9S_1$ by HRMS on the basis of its protonated molecular ion peak at *m*/*z* 796.3697 (calculated, 796.3698) (Supplementary Fig. 9a). Interestingly, the only difference between the molecular formulas of **8** and **9** was the substitution of a sulfur atom in **9** with two oxygen atoms in **8**. This similarity led to an initial co-elution of metabolites **8** and **9** during the metabolite screening and prevented us from realizing that they were in fact two distinct metabolites. After optimizing the isolation conditions, we successfully separated **8** and **9**. Next, based on the observation that the deletion of PKS gene *clbO* increased the yield of **8** (Fig. 2c), we cultivated the *E. coli* DH10B/pCAP01-*clb* Δ*clbP*Δ*clbQ*Δ*clbS*Δ*clbO* strain in LB medium and obtained 1.1 mg of **8** from a 500 L culture for its structural identification.

Metabolite **8** was a white and amorphous powder. The 1D and 2D NMR spectra of **8** suggested it contained the *N*-myristoyl-D-asparagine residue and the 1*H*-pyrrolo[3,4-*c*]pyridine-3,6(2*H*,5*H*)-dione unit (Fig. 1, Supplementary Fig. 11, Supplementary Table 3). Being different from **9**, only one thiazole proton (H-34, δ_H 8.76) resonance was observed in the ^{1}H–^{13}C HSQC spectrum of **8**, and the ^{1}H–^{13}C HMBC correlations from H-34 to C-33 (δ_C 164.1) and C-35 (δ_C 142.9) were consistent with the assignment of a thiazole moiety. An HMBC correlation from H-32 (δ_{Ha} 5.51, δ_{Hb} 5.59) to C-33 suggested an attachment between the thiazole and the 1*H*-pyrrolo[3,4-*c*]pyridine-3,6(2*H*,5*H*)-dione unit via C-32 (δ_C 44.4). In high-resolution tandem mass spectrometry (HRMS^n) fragmentation analysis, **8** and **9** showed similar fragmentation patterns, including a group of decarboxylated fragmentation species, which indicated the existence of the terminal carboxyl group in **8** (Supplementary Fig. 12). From the ^{1}H NMR spectrum of **8**, no extra proton signal was observed, but what

we did observe were four extra ^{13}C NMR signals of chemical shifts at δ_C 123.9, 133.0, 160.9, and 164.0. The ^{13}C NMR signal of shift at δ_C 160.9 was assigned as the terminal carboxyl carbon (C-39). The remaining structural assignment of **8** required a C_3HNO_2 moiety and four degrees of unsaturation. On the basis of the remaining molecular formula and HMBC correlations, we proposed that **8** contained a newly formed 5-hydroxy oxazole moiety. Three quaternary carbons were assigned as C-36 (δ_C 164.0), C-37 (δ_C 133.0) and C-38 (δ_C 123.9); and a key HMBC correlation from H-34 to C-36 revealed the linkage between the thiazole moiety and the 5-hydroxy oxazole moiety.

Then the HRMS^n data of both **8** and **9** were further compared. The majority of the MS/MS fragmentation species of **8** were assigned (Supplementary Fig. 12a, d) and their counterparts in the HRMS^n fragmentation spectrum of **9** were identified (Supplementary Fig. 12c, e), including three identical fragmentation species (which could be observed in the HRMS^n fragmentation spectra of both **8** and **9**) and 14 pairs of corresponding fragmentation species (which displayed the same difference in molecular weight due to 5-hydroxy oxazole moiety [C_3HNO_2] *vs* thiazole moiety [C_3HNS]) (Supplementary Fig. 12f). In addition, two more distinctive MS/MS fragmentation species were observed in the fragmentation spectra of **8** but not in that of **9** (*fragmentation species **18** and **19*** in Supplementary Fig. 12). These two carbonium-based ions were produced by the hydroxy group-induced elimination of one H_2O molecule from their decarboxylated precursor ions, suggesting the presence of a hydroxy group in the distinctive moiety of C_3HNO_2.

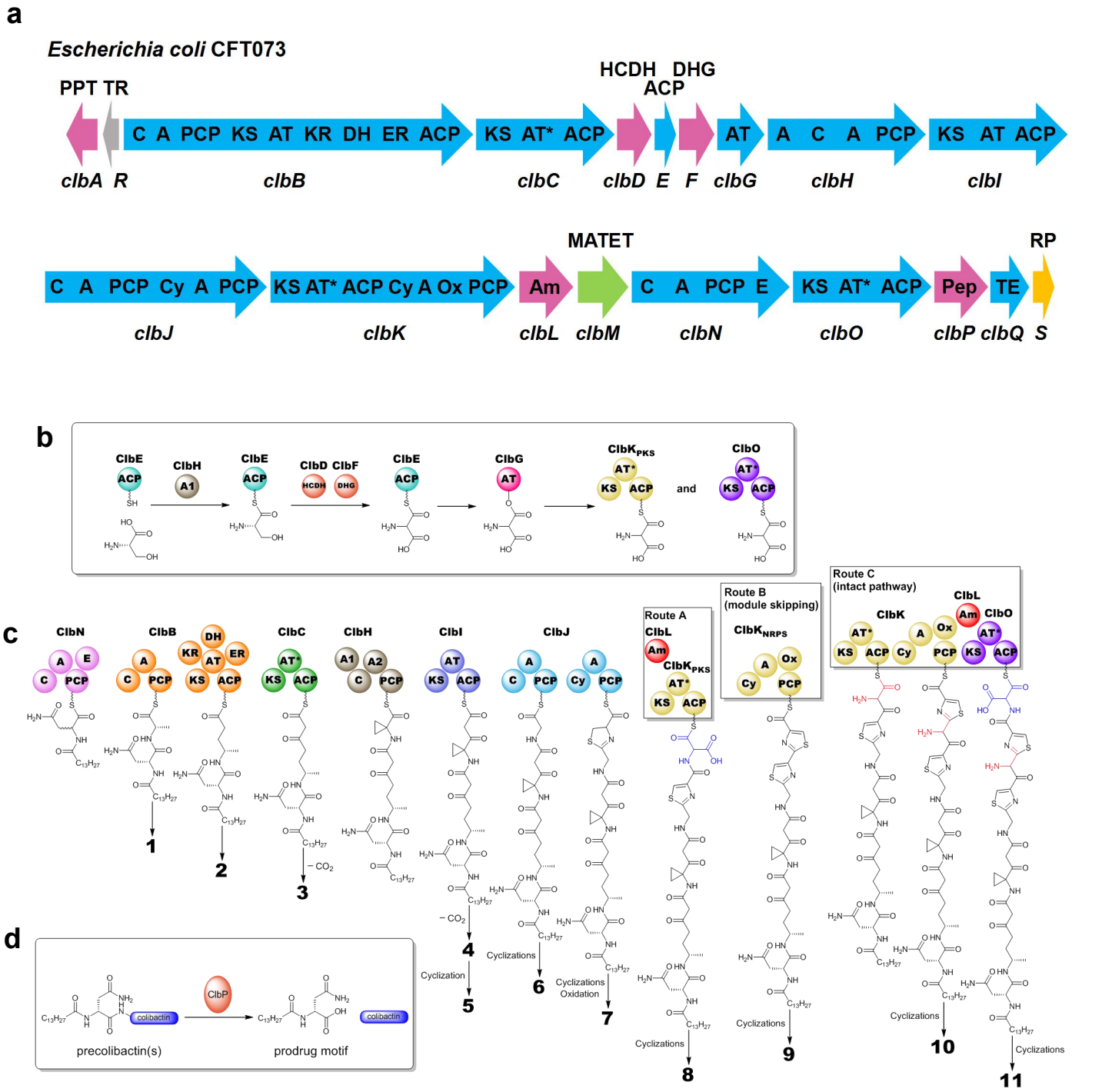
Structure elucidation of precolibactin-969 (11)

The molecular formula of **11** was determined as $C_{44}H_{59}N_9O_{12}S_2$ by HRMS on the basis of its protonated molecular ion peak at *m*/*z* 970.3799 (calculated, 970.3797) (Supplementary Fig. 3). From 2,000 L of fermentation broth of *E. coli* DH10B/pCAP01-*clb* Δ*clbP*Δ*clbQ*Δ*clbS*, we obtained 50 μg of **11** as a white and amorphous powder for its structure elucidation. Similar to precolibactin-886 (**10**), **11** was isolated as an approximately equal mixture of two diastereomers (Supplementary Fig. 5). By comparing the HRMS^n fragmentation data of **10** (molecular formula: $C_{41}H_{58}N_8O_{10}S_2$) and **11** (molecular formula: $C_{44}H_{59}N_9O_{12}S_2$), we were able to assign the majority of the MS/MS fragmentation species of **10** (Supplementary Fig. 7a, c) and identify their counterparts in the HRMS^n fragmentation spectrum of **11** (Supplementary Fig. 7b, d), including eight identical fragmentation species (which were N-terminal ions and could be observed in the HRMS^n fragmentation spectra of both **10** and **11**) and 10 pairs of corresponding fragmentation species (which were C-terminal ions and displayed a difference in molecular weight of 83.00 daltons) (Supplementary Fig. 7e). All of these HRMS^n fragmentation data indicated that an additional C_3HNO_2 fragment in **11** is likely attached to the carboxyl terminal of **10**.

The ^1H NMR spectroscopic data of **11** were almost identical to those of **10**, with the exception of some shifts of corresponding proton signals (Supplementary Fig. 6b). The ^1H–^1H COSY, ^1H–^1H TOCSY, ^1H–^{13}C HMQC and ^1H–^{13}C HMBC data indicated that **11** and **10** shared the same carbon skeleton from C-1 to C-40, including a heterocycle-fused macrocycle and bisected bithiazole (Supplementary Fig. 6, Supplementary Table 1). A key HMBC correlation from H-22 (δ_H 34.6) to C-23 (δ_C 107.9) established the hemiacetal functionality, which was a unique connectivity for the macrocycle formation; a linear connection from C-24 to C-33 was established by comprehensive and clear HMBC correlations along with a characteristic ketone resonance at δ_C 205.4 (C-29) as indicated by the HMBC correlations from H-30 (δ_{Ha} 3.08 and δ_{Hb} 3.73) to C-29; two thiazole protons at δ_H 7.97/δ_C 119.8 (H-34/C-34) and δ_H 9.07/δ_C 120.5 (H-39/C-39) were observed from both ^1H NMR and ^1H–^{13}C HSQC spectra, presenting two thiazole moieties.

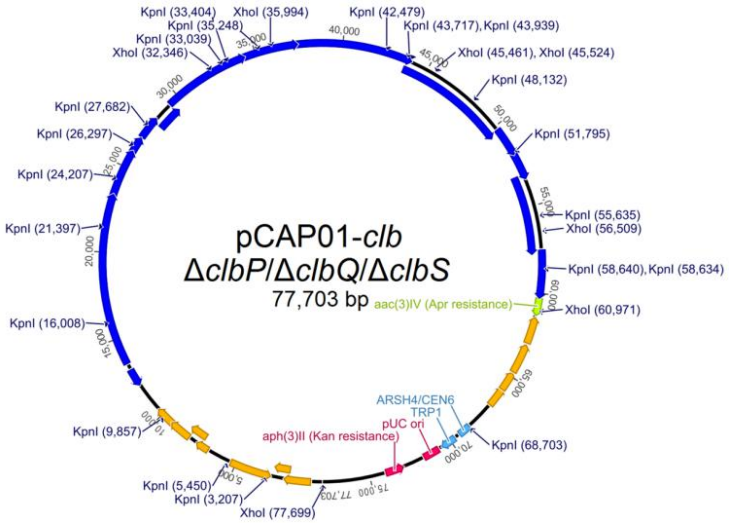
We did not obtain adequate HMBC correlations to the additional molecular formula C_3HNO_2, because of its proton-deficient structure. Being enlightened by the chemical structure of **8** and the results from the feeding experiment which showed that **11** was labeled by L-[1-^{13}C]serine or L-[^{15}N]serine with a relative [*M* + H + 1] signal intensity approximately twice as high as those of **8** and **10** (Supplementary Fig. 4), we proposed that the remaining C_3HNO_2 moiety in **11** forms a 5-hydroxy oxazole ring, the same as in **8** (Fig. 1).

3. Supplementary Figures

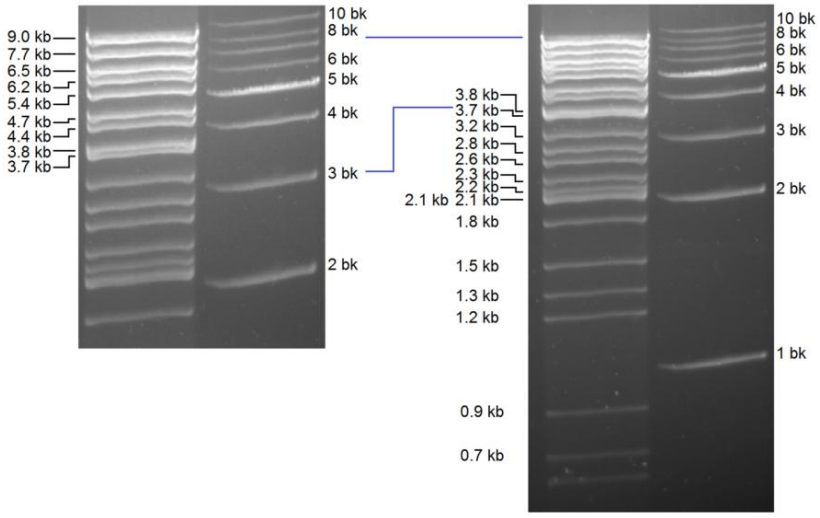


Supplementary Figure 1. A summary of findings from previous and current study on the colibactin biosynthesis. a, Organization of the colibactin (*clb*) biosynthetic gene cluster (not drawn to scale). Genes encoding nonribosomal peptide synthetase (NRPS), polyketide synthase (PKS), and NRPS-PKS hybrid enzyme are shown in blue; genes encoding tailoring enzyme or accessory enzyme are shown in pink; the *clb* pathway transcriptional regulator (TR) gene *clbR* is shown in grey; the multidrug and toxic compound extrusion transporter (MATET) gene *clbM* is shown in green; and the *clb* pathway resistance protein (RP) gene *clbS* is shown in yellow. The predicted functions are shown using the following abbreviations: A, adenylation; ACP, acyl carrier protein; Am, amidase; AT, acyltransferase; C, condensation; Cy, cyclization; DH, dehydratase; DHG, dehydrogenase; E, epimerase; ER, enoyl reductase; HCDH, hydroxylacyl-CoA dehydrogenase; KR, ketoreductase; KS, ketosynthase; Ox, oxidase; PCP, peptidyl carrier protein; Pep, peptidase (a periplasmic D-amino peptidase); PPT, phosphopantetheinyl-transferase; TE, thioesterase. AT* domains are predicted based on structural topology as ancestral inactive relics. **b**, Formation and transportation of an unusual PKS extender unit aminomalonyl-ACP, which is involved in the biosynthesis of (pre)colibactins. **c**, Proposed biosynthesis of precolibactins. **d**, A prodrug mechanism is involved in the *clb* biosynthetic pathway. Precolibactin(s) can be recognized and hydrolyzed by the peptidase ClbP, resulting in the release of mature colibactin(s) and a prodrug motif (an *N*-myristoyl-D-asparagine). **e**, Chemical structures of precolibactins.

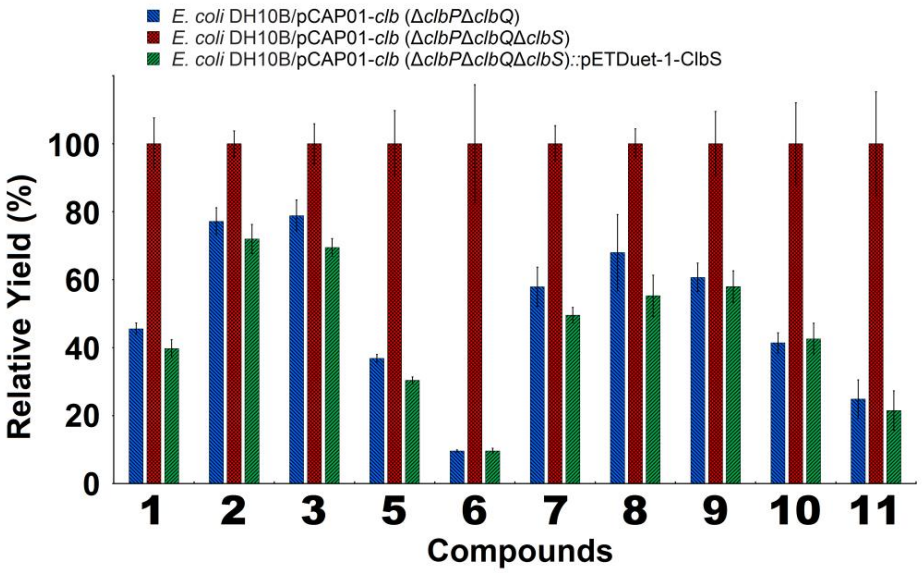
a


pCAP01-clb
ΔclbP/ΔclbQ/ΔclbS
77,703 bp
aac(3)IV (Apr resistance)
ARSH4/CEN6
TRP1
pUC ori
aph(3)II (Kan resistance)
KpnI (33,404)
XhoI (35,994)
KpnI (42,479)
KpnI (35,248)
KpnI (43,717),KpnI (43,939)
KpnI (33,039)
XhoI (45,461), XhoI (45,524)
XhoI (32,346)
KpnI (48,132)
KpnI (27,682)
KpnI (51,795)
KpnI (26,297)
KpnI (24,207)
KpnI (55,635)
XhoI (56,509)
KpnI (21,397)
KpnI (58,640),KpnI (58,634)
XhoI (60,971)
KpnI (16,008)
KpnI (9,857)
KpnI (68,703)
KpnI (5,450)
KpnI (3,207)
XhoI (77,699)

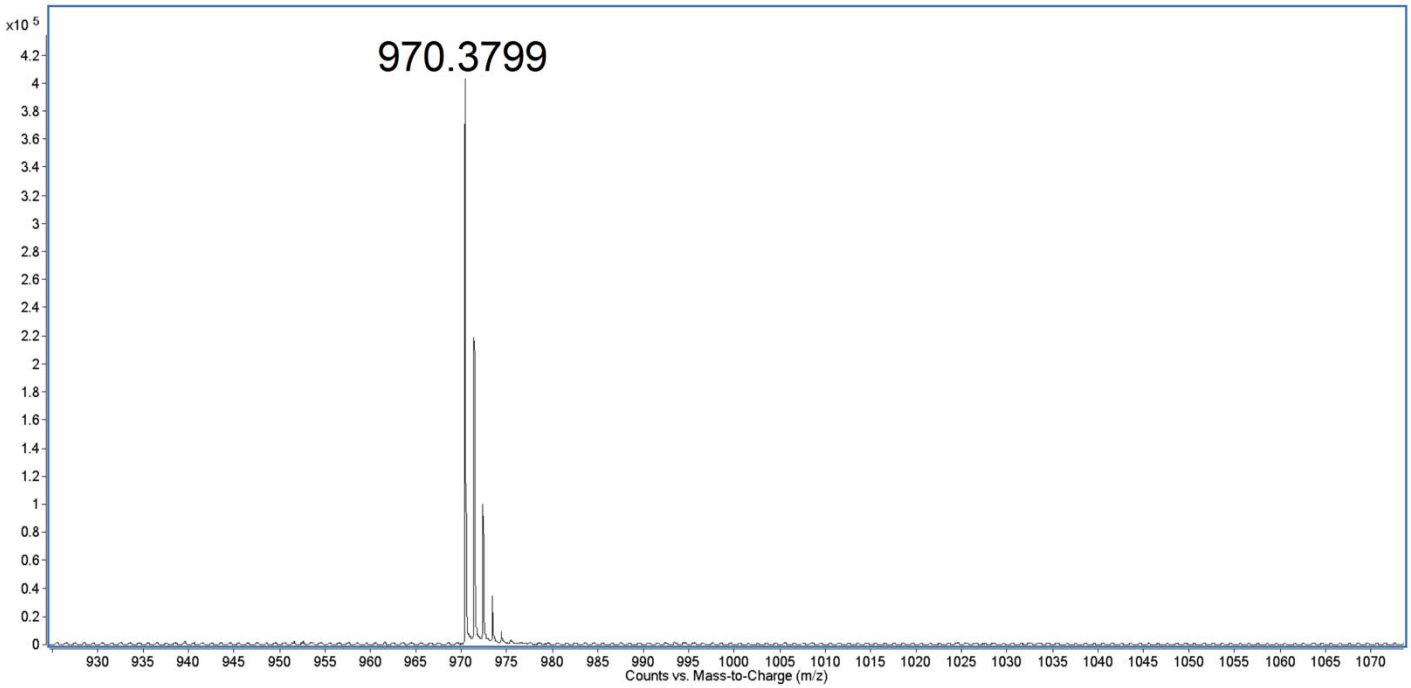
b


9.0 kb
7.7 kb
6.5 kb
6.2 kb
5.4 kb
4.7 kb
4.4 kb
3.8 kb
3.7 kb
10 bk
8 bk
6 bk
5 bk
4 bk
3 bk
2 bk
3.8 kb
3.7 kb
3.2 kb
2.8 kb
2.6 kb
2.3 kb
2.2 kb
2.1 kb
2.1 kb
1.8 kb
1.5 kb
1.3 kb
1.2 kb
0.9 kb
0.7 kb
10 bk
8 bk
6 bk
5 bk
4 bk
3 bk
2 bk
1 bk

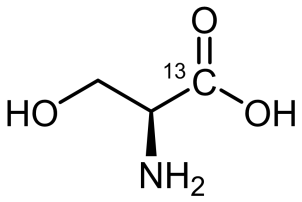
c


E. coli DH10B/pCAP01-clb (ΔclbPΔclbQ)
E. coli DH10B/pCAP01-clb (ΔclbPΔclbQΔclbS)
E. coli DH10B/pCAP01-clb (ΔclbPΔclbQΔclbS)::pETDuet-1-ClbS
Relative Yield (%)
0
20
40
60
80
100
1
2
3
5
6
7
8
9
10
11
Compounds

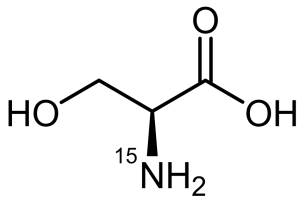
Supplementary Figure 2. Inactivation of resistance protein gene *clbS* and the effect of ClbS on the production of *clb* pathway-related precolibactins. a, **b**, Physical map of bacterial artificial chromosome pCAP01-*clb* Δ*clbP*Δ*clbQ*Δ*clbS* with predicted KpnI + XhoI cleavage sites and restriction fragment sizes (**a**) and the experimentally determined restriction map with KpnI + XhoI (**b**). **c**, Effect of resistance protein ClbS on the production of various *clb* pathway-related precolibactins. A comparison of the relative abundances of individual precolibactin derivatives from extracts of *ClbS*-containing (pCAP01-*clb* Δ*clbP*Δ*clbQ*), *ClbS*-knockout (pCAP01-*clb* Δ*clbP*Δ*clbQ*Δ*clbS*), and *ClbS*-complemented (pCAP01-*clb* Δ*clbP*Δ*clbQ*Δ*clbS*::pETDuet-1-C*lbS*) *E. coli* strains, based on the extracted ion chromatograms. Data are shown as mean ± s.d. (n = 5).



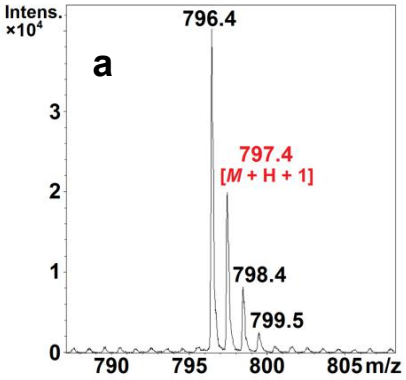
Supplementary Figure 3. The high-resolution ESI-MS of precolibactin-969 (**11**); its molecular formula was determined as $C_{44}H_{59}N_9O_{12}S_2$ on the basis of its protonated molecular ion peak at *m*/*z* 970.3799 (calculated, 970.3797).

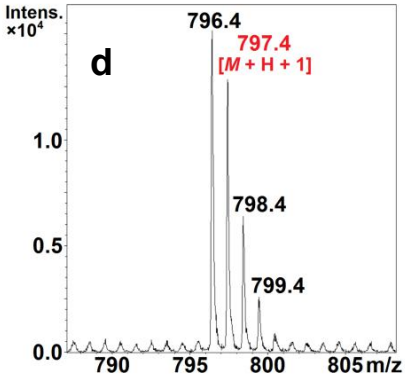


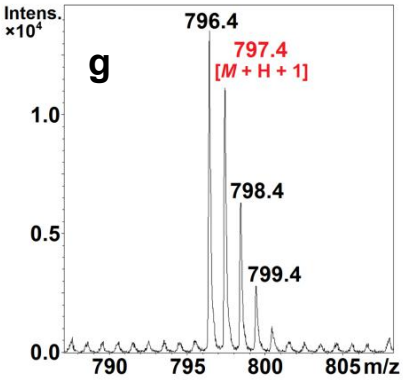
L-[1-^{13}C]serine

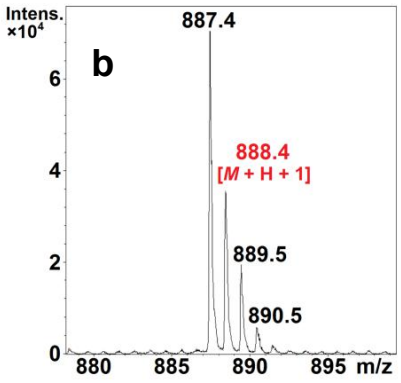


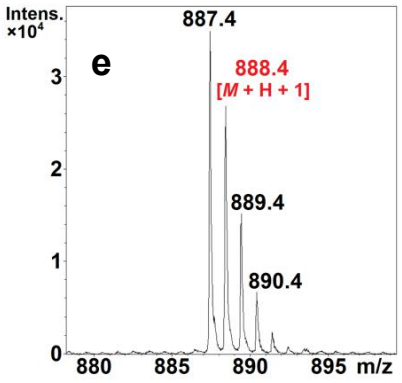
L-[^{15}N]serine

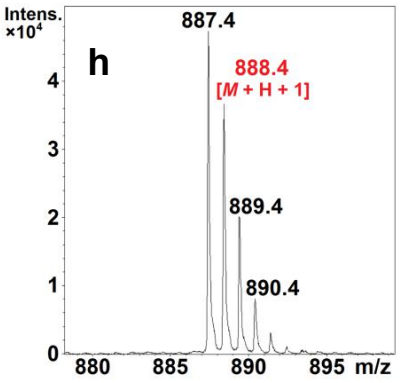


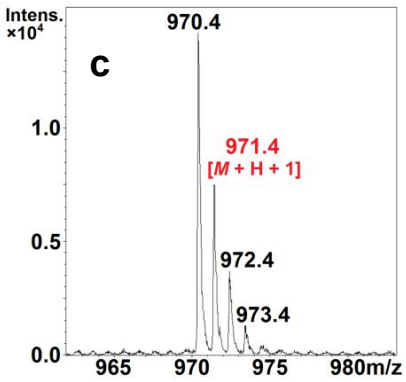


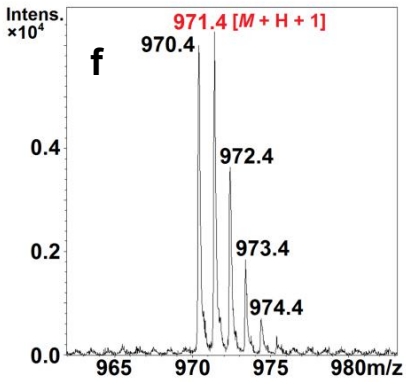


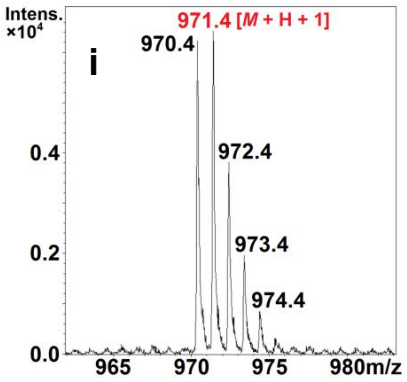




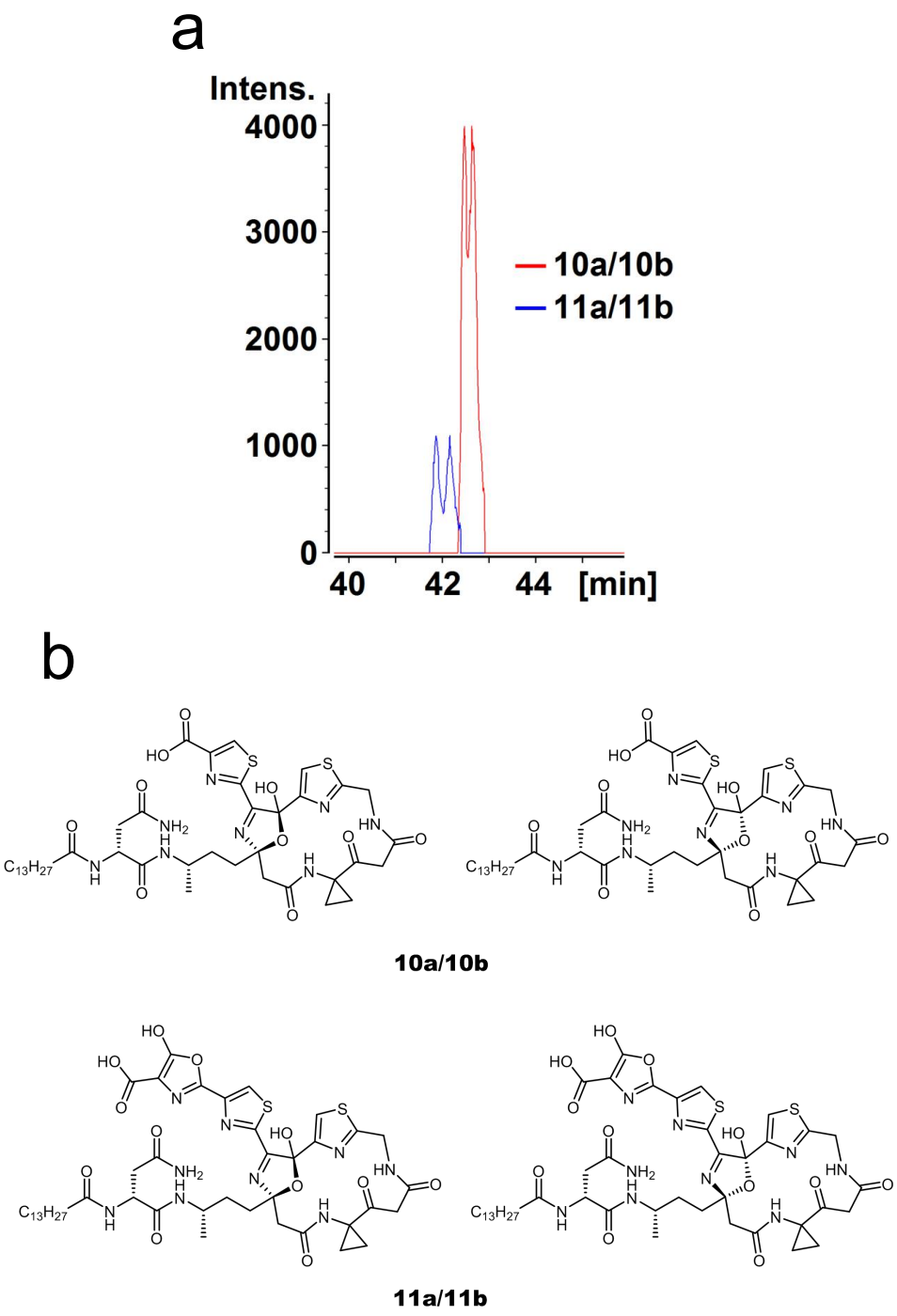






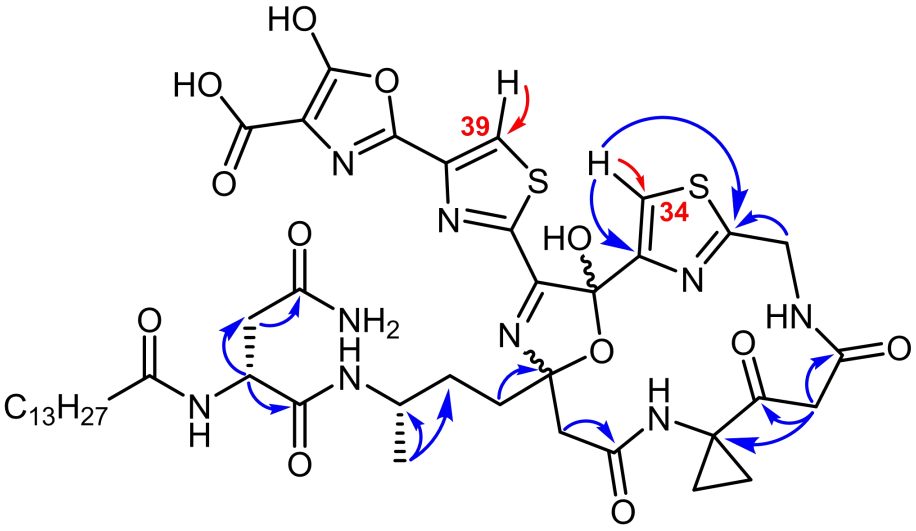


Supplementary Figure 4. UPLC–MS analysis of the feeding experiment. a–i, Considering that the genes related to the 2-aminomalonyl building block, including biosynthetic gene cassette *clbDEF* and *clbG*, have been proven to form and transfer the L-serine-derived rare PKS extender unit[8,9], in order to interrogate the incorporation of the aminomalonyl extender unit into precolibactin-795a (**8**), precolibactin-886 (**10**), or precolibactin-969 (**11**), both L-[1-^{13}C]serine and L-[^{15}N]serine were selected for the feeding experiments. L-[1-^{13}C]serine or L-[^{15}N]serine at a concentration of 0.25 mg/mL was added to the LB medium for culturing the heterologous expression strain *E. coli* DH10B/pCAP01-*clb* Δ*clbP*Δ*clbQ*Δ*clbS*. Compared to the blank tests without isotope feeding (**a** for **8**, **b** for **10**, and **c** for **11**), the isotope-labeled metabolites were observed, displaying mass shifts of [*M* + H + 1] with different abundances. (**d**) **8** was labeled by L-[1-^{13}C]serine; (**e**) **10** was labeled by L-[1-^{13}C]serine; (**f**) **11** was labeled by L-[1-^{13}C]serine; (**g**) **8** was labeled by L-[^{15}N]serine; (**h**) **10** was labeled by L-[^{15}N]serine; (**i**) **11** was labeled by L-[^{15}N]serine. The results show that all of these metabolites could be labeled by L-[1-^{13}C]serine or L-[^{15}N]serine; however, **11** was labeled by L-[1-^{13}C]serine or L-[^{15}N]serine with relative [*M* + H + 1] signal intensity approximately twice as higher as those of **8** and **10**, indicating that **11** contains two L-serine derived moieties.



Supplementary Figure 5. a, The detection of precolibactin-886 (**10**) diastereomer pairs and precolibactin-969 (**11**) diastereomer pairs. The metabolic profile of *clb*$^{+}$ heterologous expression strain *E. coli* DH10B/pCAP01-*clb* Δ*clbP*Δ*clbQ*Δ*clbS* was investigated using UPLC–MS-based analysis, displaying that both of the **10** and **11** were produced as approximately equal mixtures of diastereomer pairs which were separated by chromatography. **b**, Proposed structures of **10a**/**10b** and **11a**/**11b**.

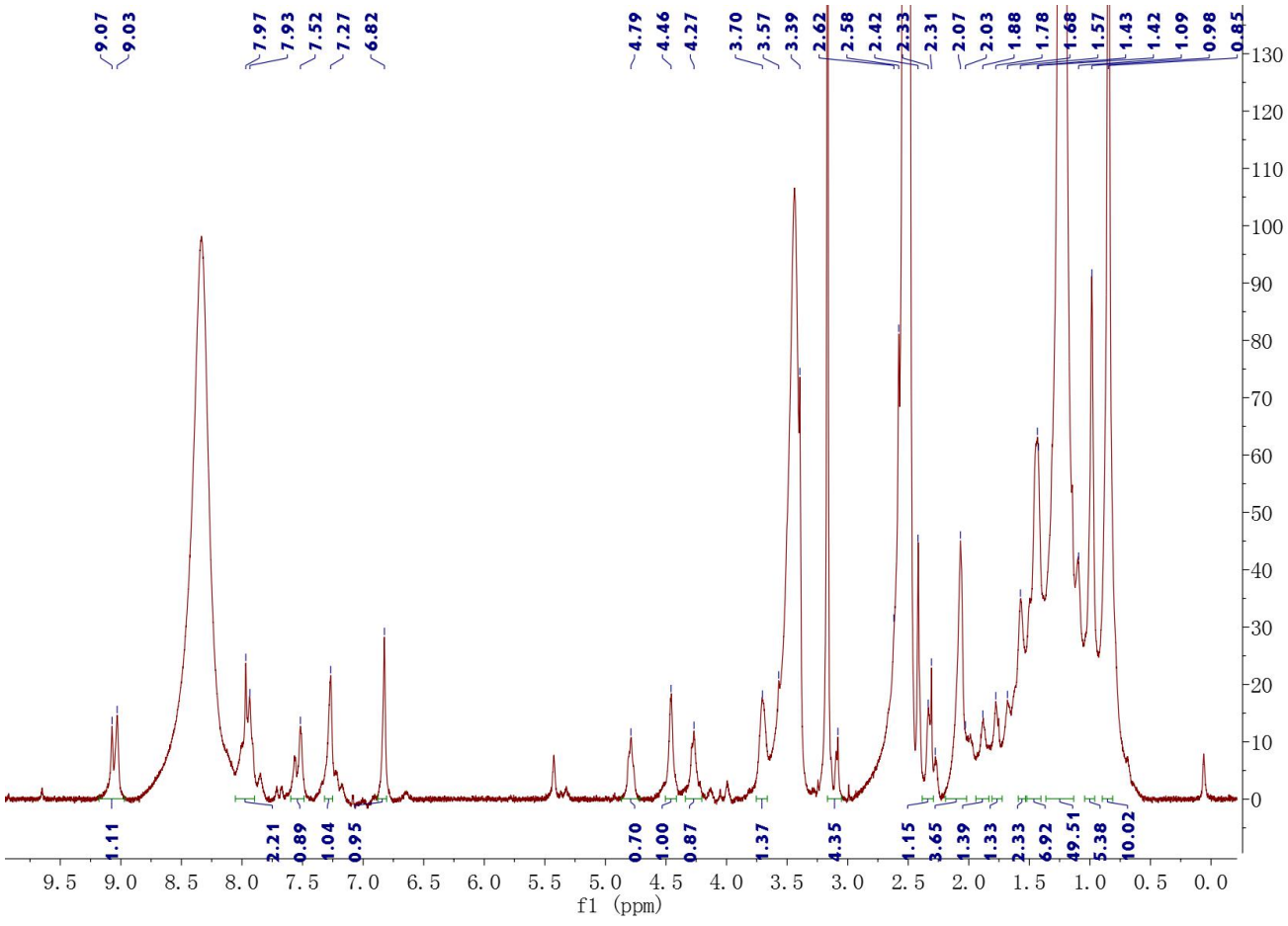
a



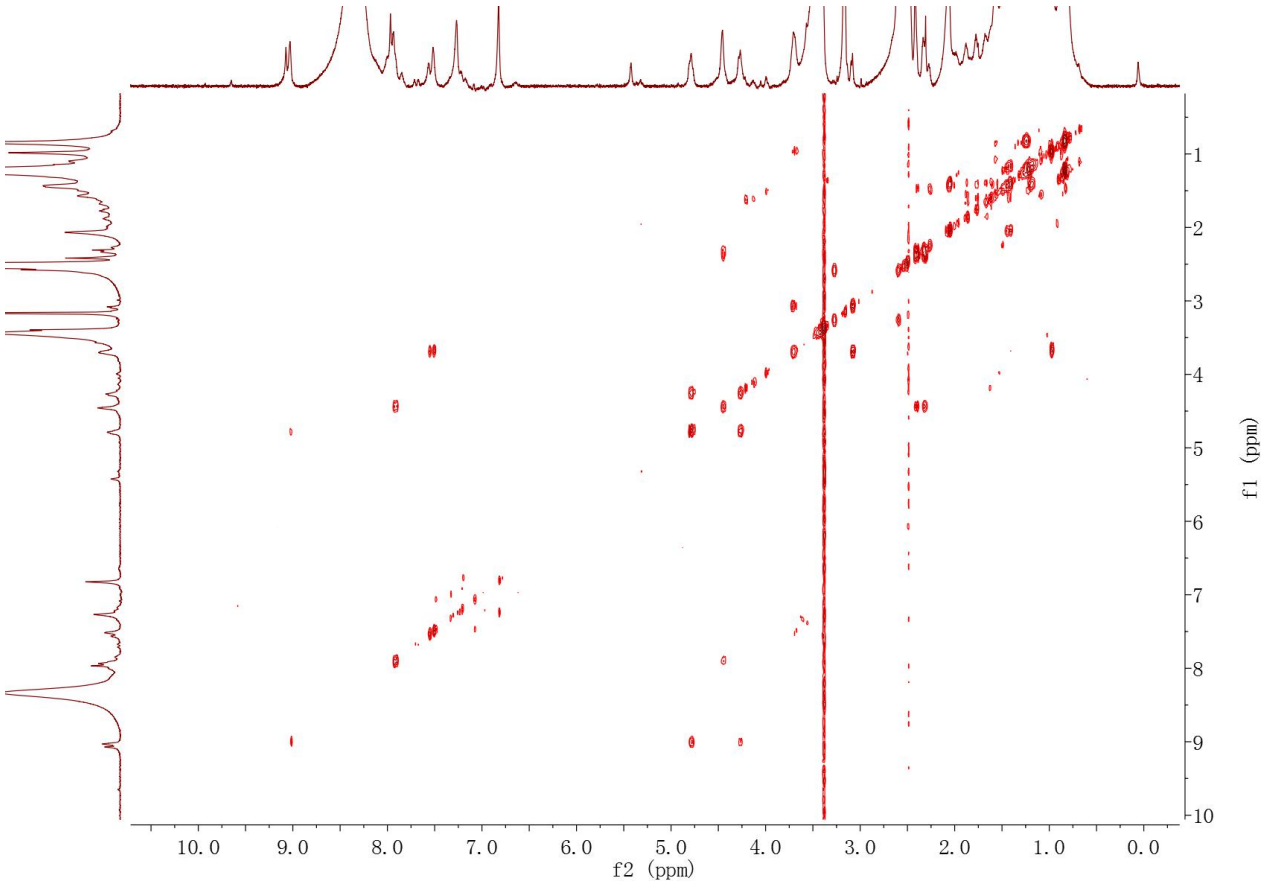
11

Key HMBC **Key HSQC**

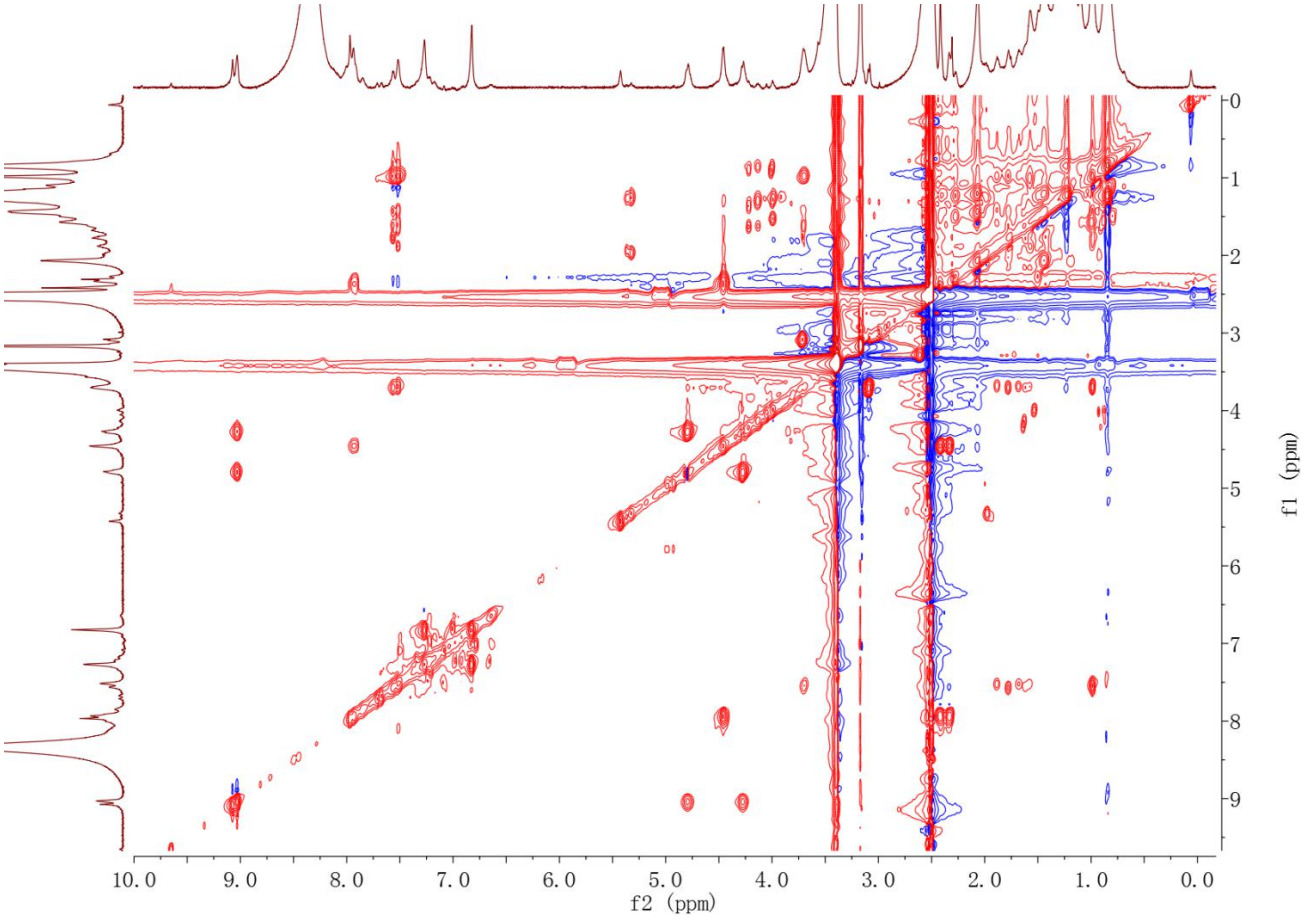
b



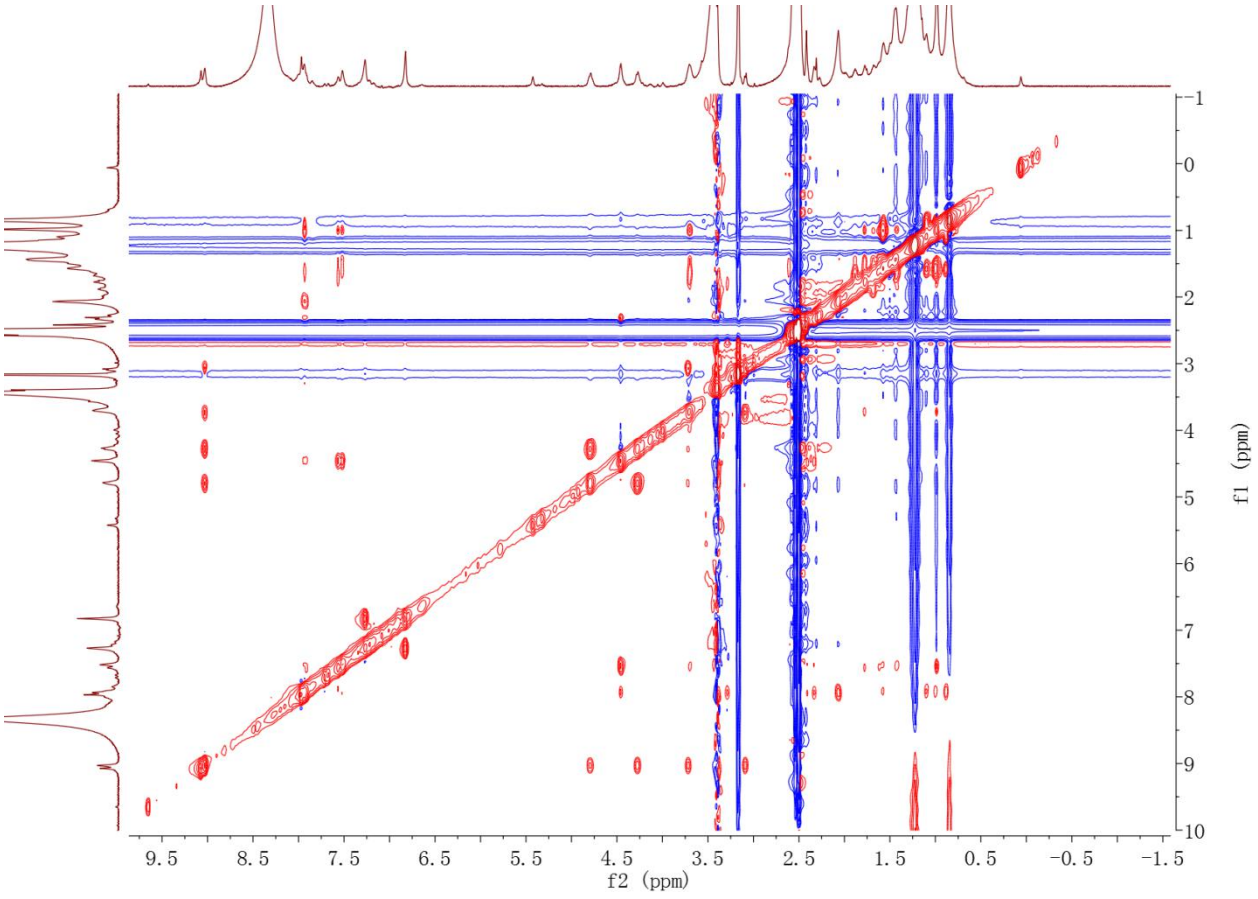
c


10.0
9.0
8.0
7.0
6.0
5.0
4.0
3.0
2.0
1.0
0.0
f2 (ppm)
1
2
3
4
5
6
7
8
9
10
f1 (ppm)

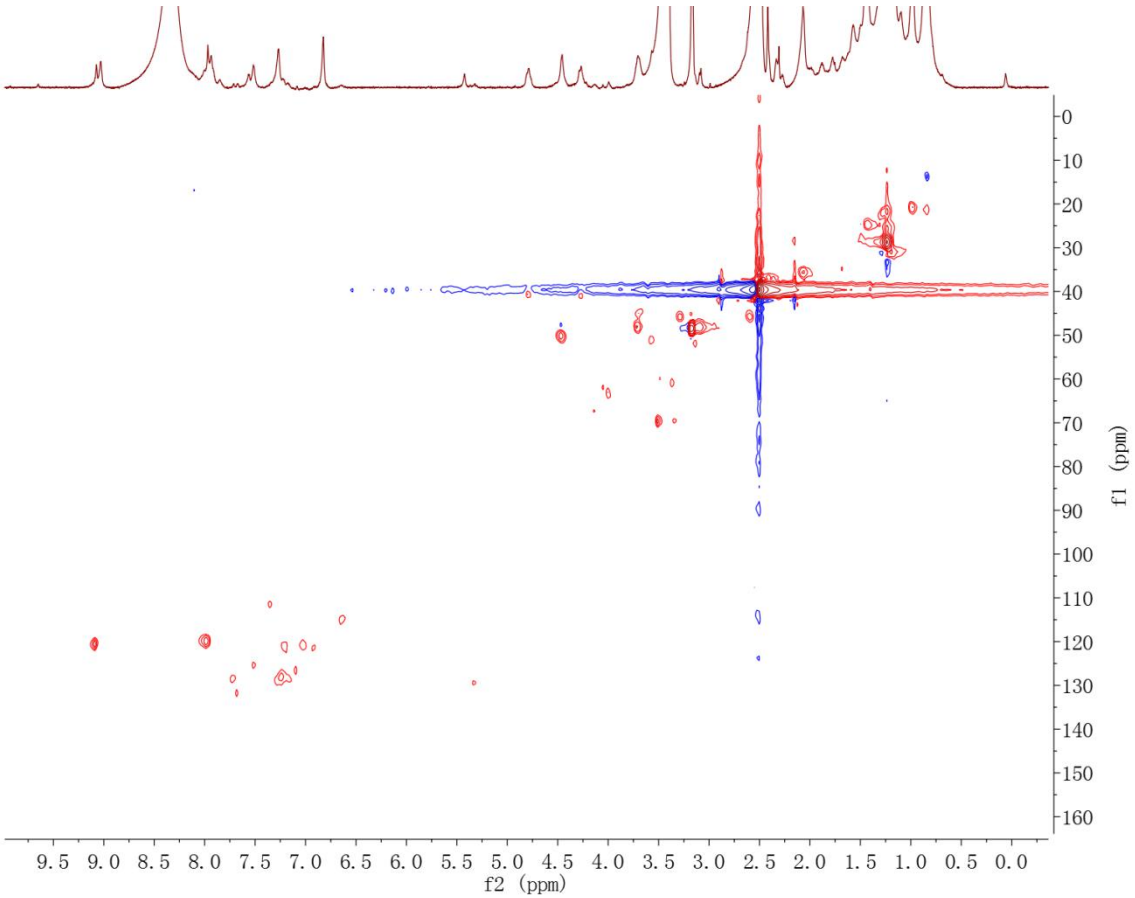
d


10.0
9.0
8.0
7.0
6.0
5.0
4.0
3.0
2.0
1.0
0.0
f2 (ppm)
0
1
2
3
4
5
6
7
8
9
f1 (ppm)

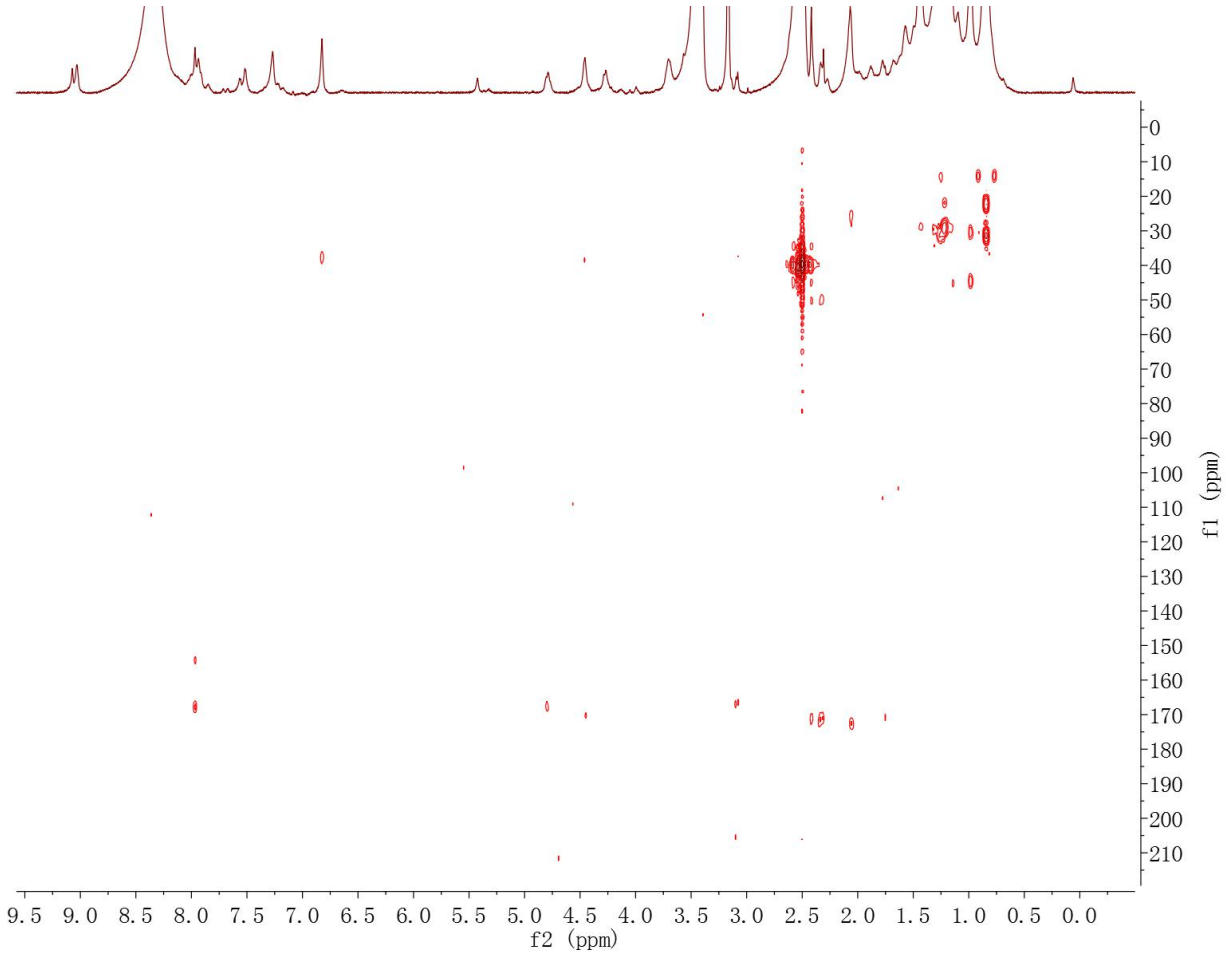
e



f

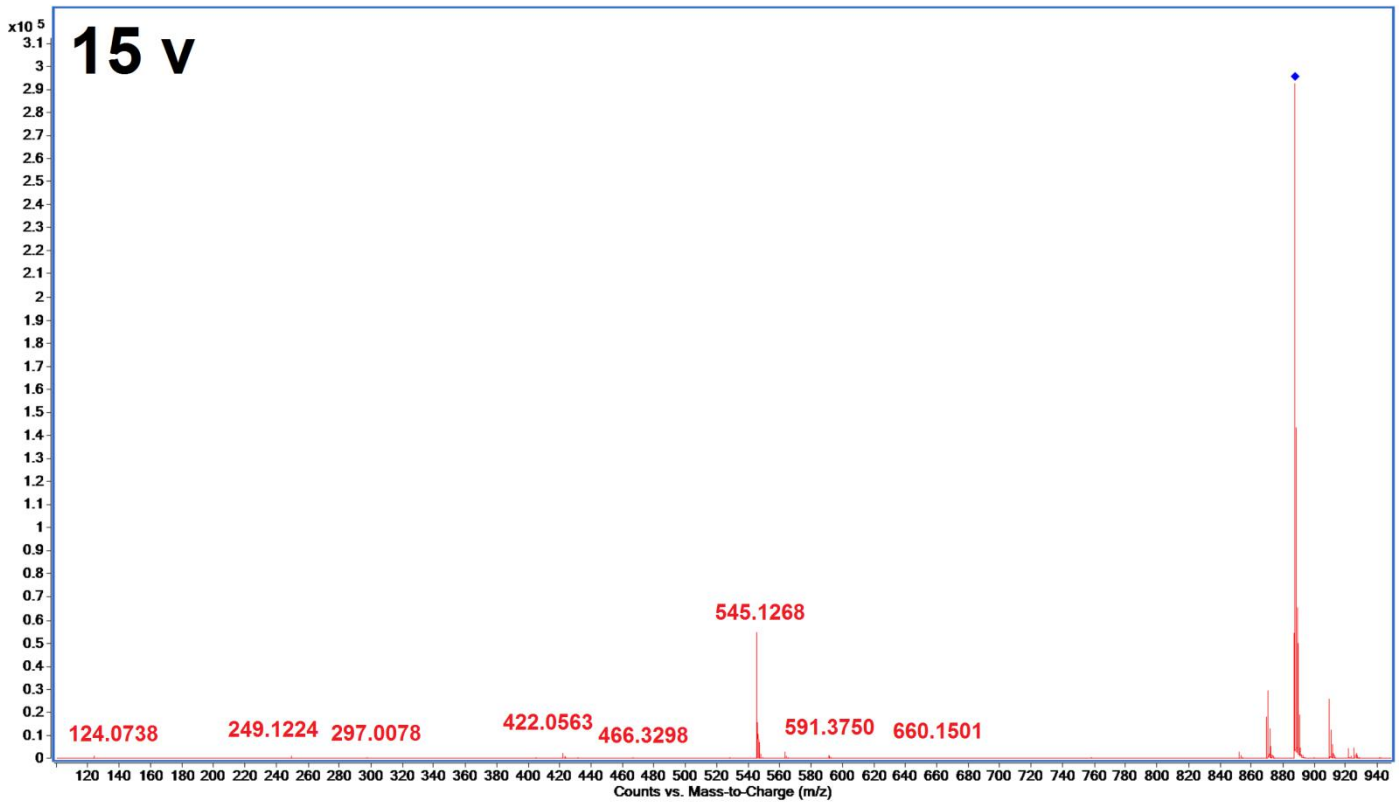


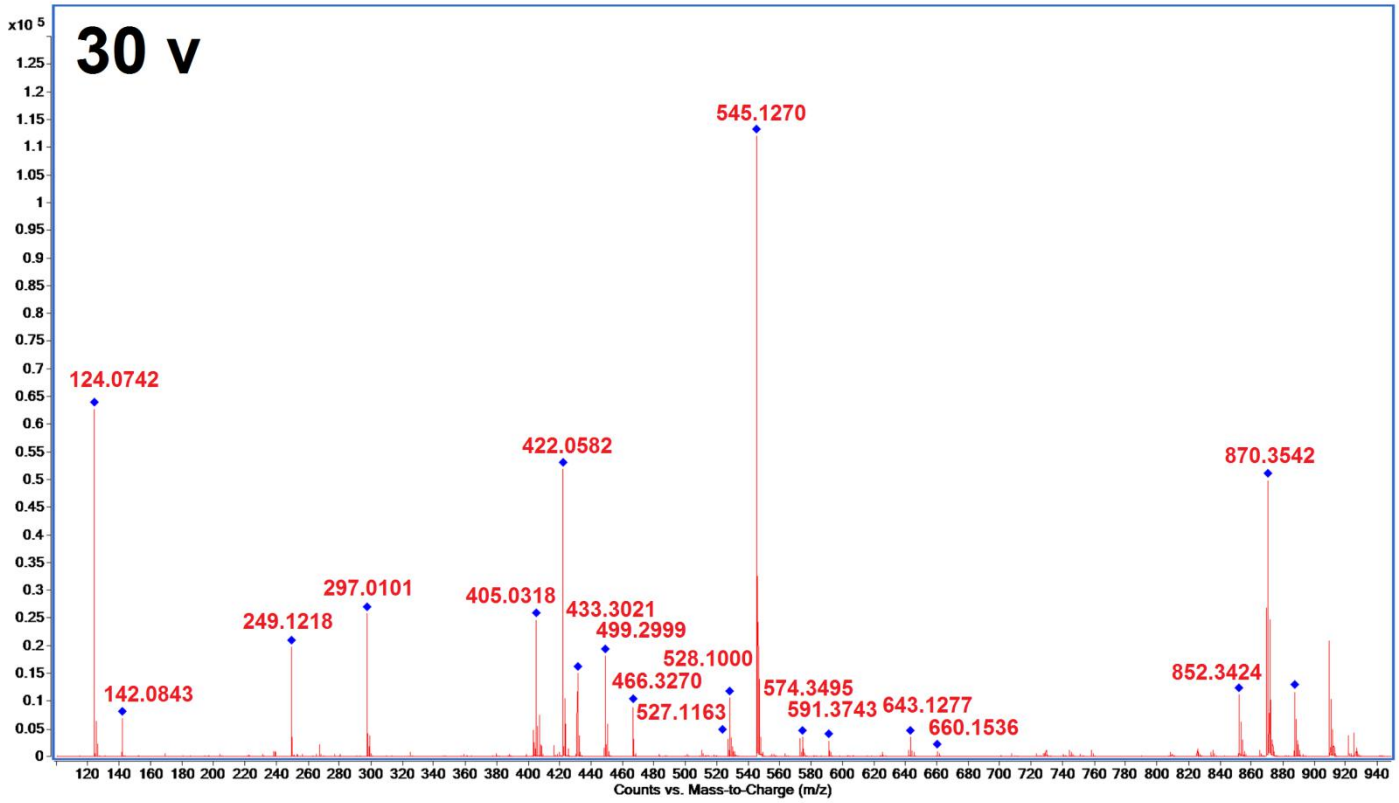
g



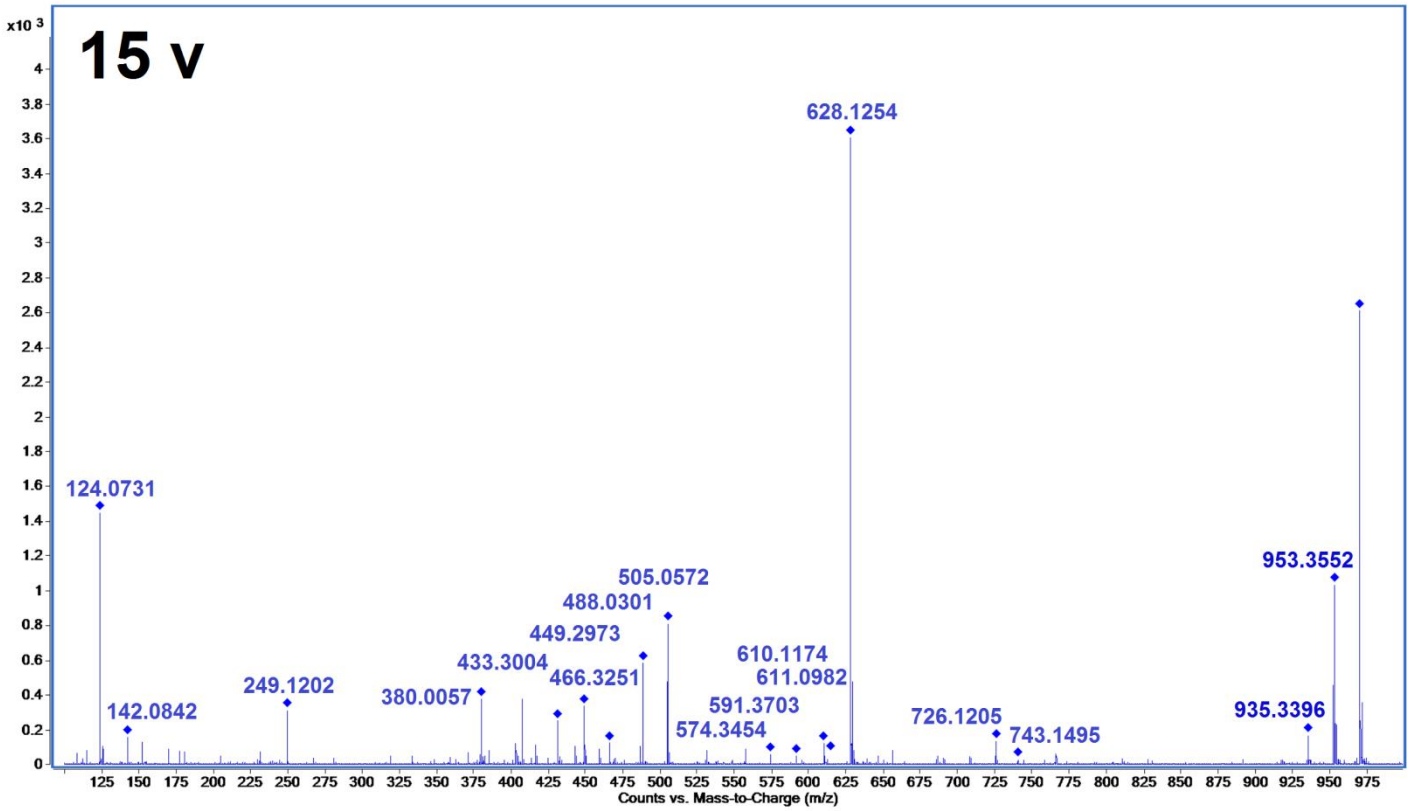
Supplementary Figure 6. 1D and 2D NMR spectra of precolibactin-969 (**11**). **a**, NMR-based key correlations for the structural assignment of **11**. **b**, ^{1}H NMR spectrum, **c**, ^{1}H–^{1}H COSY spectrum, **d**, ^{1}H–^{1}H TOCSY spectrum, **e**, ^{1}H–^{1}H NOESY spectrum, **f**, ^{1}H–^{13}C HSQC spectrum, and **g**, ^{1}H–^{13}C HMBC spectrum. Recorded with 850 MHz for ^{1}H and 212.5 MHz for ^{13}C, cold probe, in DMSO-d_6.

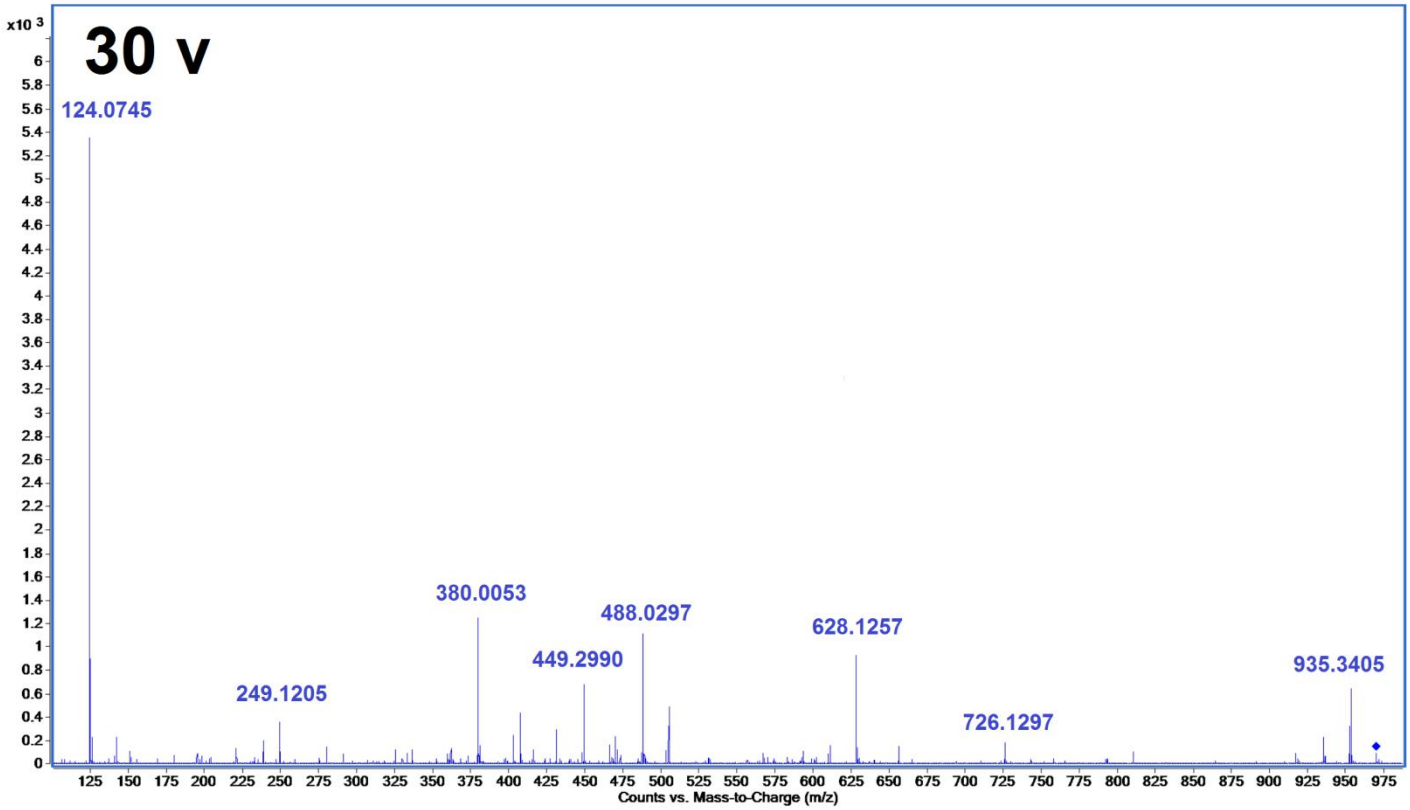
a


15 v
x10 5
545.1268
124.0738
249.1224
297.0078
422.0563
466.3298
591.3750
660.1501
Counts vs. Mass-to-Charge (m/z)

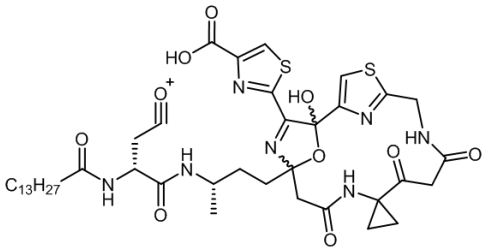

30 v
x10 5
545.1270
124.0742
422.0582
870.3542
297.0101
405.0318
433.3021
249.1218
499.2999
528.1000
466.3270
852.3424
142.0843
574.3495
591.3743
643.1277
527.1163
660.1536
Counts vs. Mass-to-Charge (m/z)

b

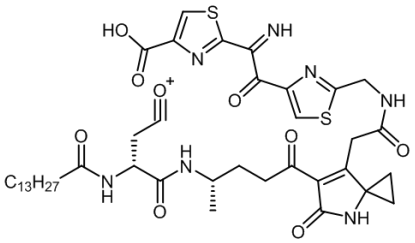

15 v
x10 3
628.1254
124.0731
953.3552
505.0572
488.0301
449.2973
433.3004
610.1174
611.0982
466.3251
249.1202
380.0057
591.3703
142.0842
574.3454
726.1205
935.3396
743.1495
Counts vs. Mass-to-Charge (m/z)


30 v
x10 3
124.0745
380.0053
488.0297
628.1257
449.2990
935.3405
249.1205
726.1297
Counts vs. Mass-to-Charge (m/z)

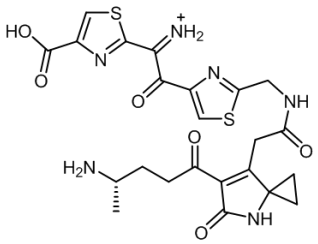
C



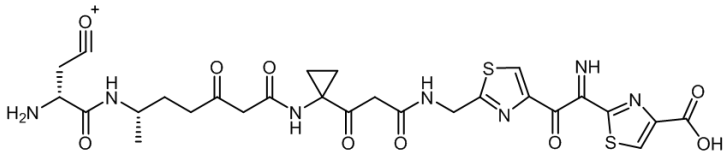
Frag. 1 (870.3525)



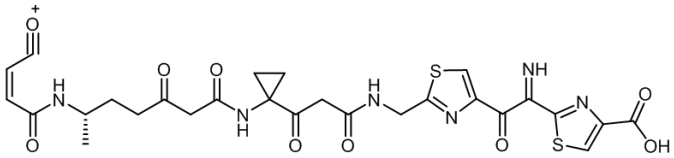
Frag. 2 (852.3419)



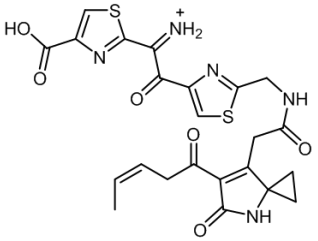
Frag. 7 (545.1272)



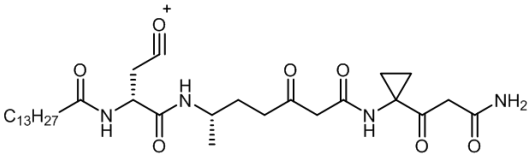
Frag. 3 (660.1541)



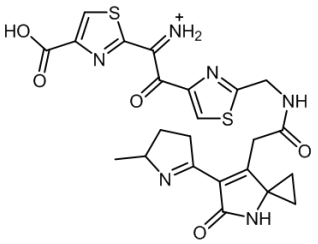
Frag. 4 (643.1275)



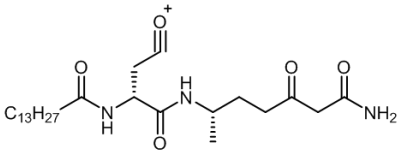
Frag. 8 (528.1006)



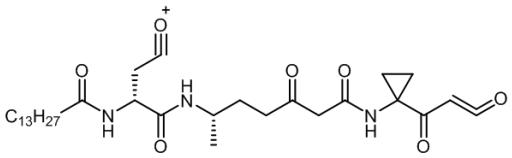
Frag. 5 (591.3752)



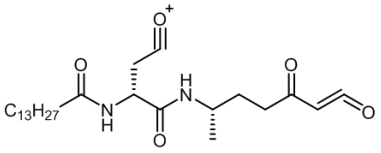
Frag. 9 (527.1166)



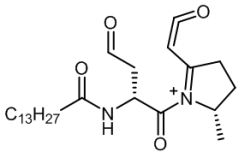
Frag. 10 (466.3275)



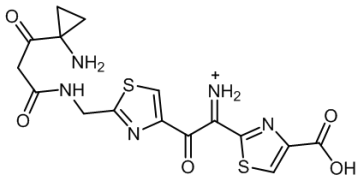
Frag. 6 (574.3487)



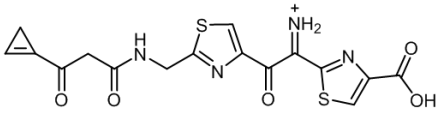
Frag. 11 (449.3010)



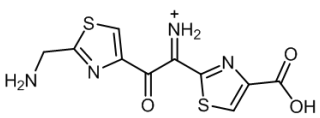
Frag. 12 (433.3061)



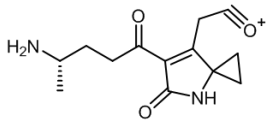
Frag. 13 (422.0587)



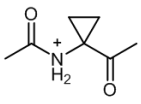
Frag. 14 (405.0322)



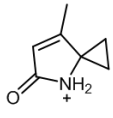
Frag. 15 (297.0111)



Frag. 16 (249.1234)

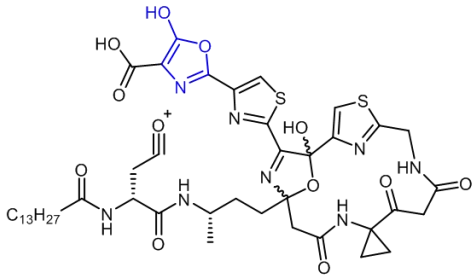


Frag. 17 (142.0863)

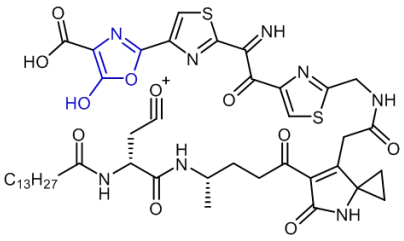


Frag. 18 (124.0757)

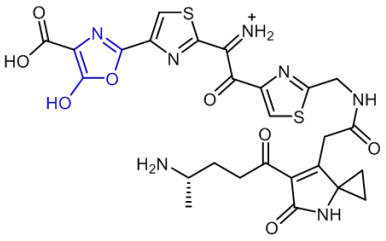
d



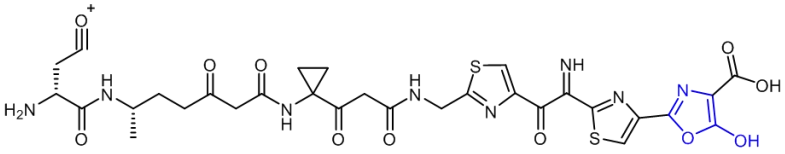
Frag. 1 (953.3532)



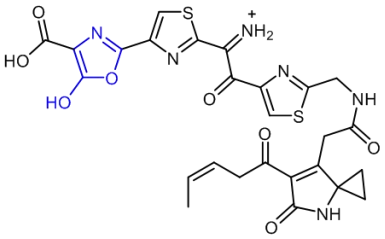
Frag. 2 (935.3426)



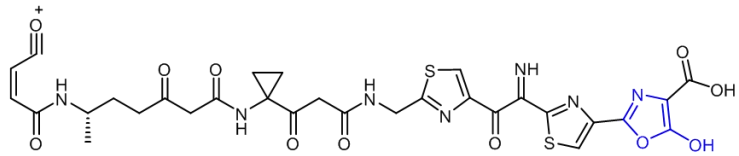
Frag. 7 (628.1279)



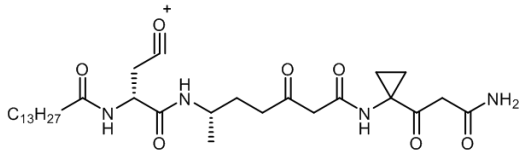
Frag. 3 (743.1548)



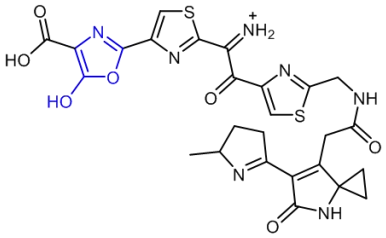
Frag. 8 (611.1013)



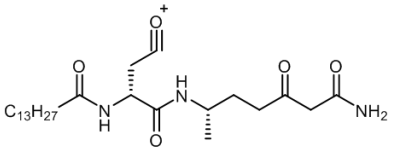
Frag. 4 (726.1283)



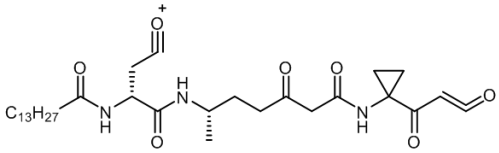
Frag. 5 (591.3752)



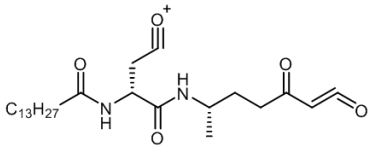
Frag. 9 (610.1173)



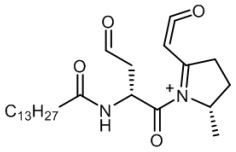
Frag. 10 (466.3275)



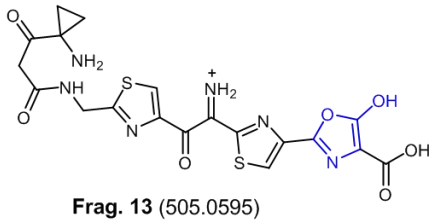
Frag. 6 (574.3487)



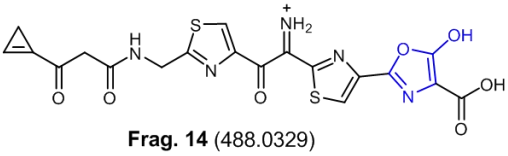
Frag. 11 (449.3010)



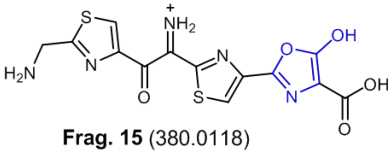
Frag. 12 (433.3061)



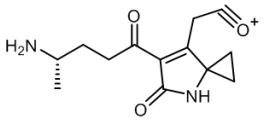
Frag. 13 (505.0595)



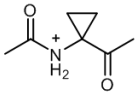
Frag. 14 (488.0329)



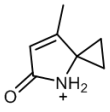
Frag. 15 (380.0118)



Frag. 16 (249.1234)



Frag. 17 (142.0863)

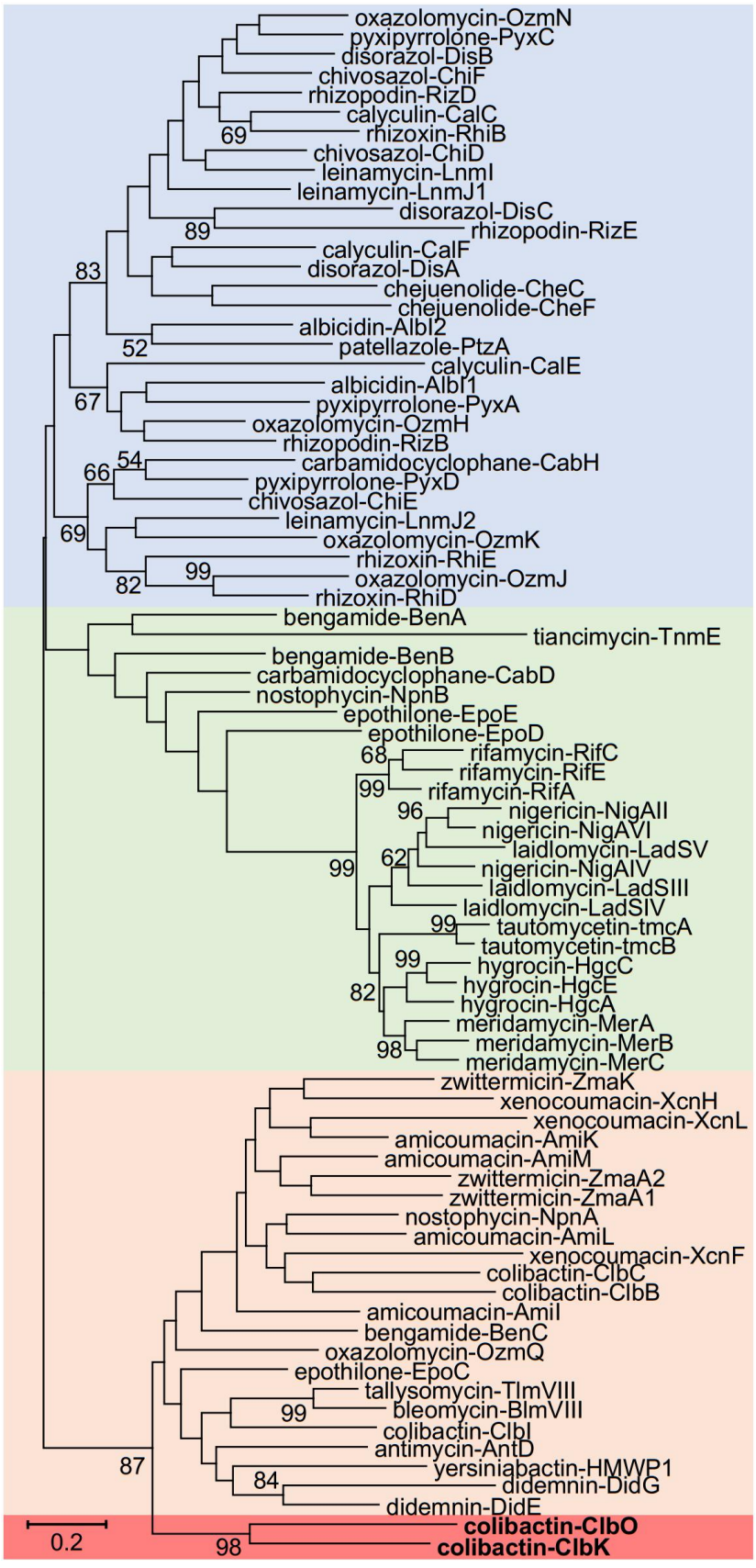


Frag. 18 (124.0757)

e

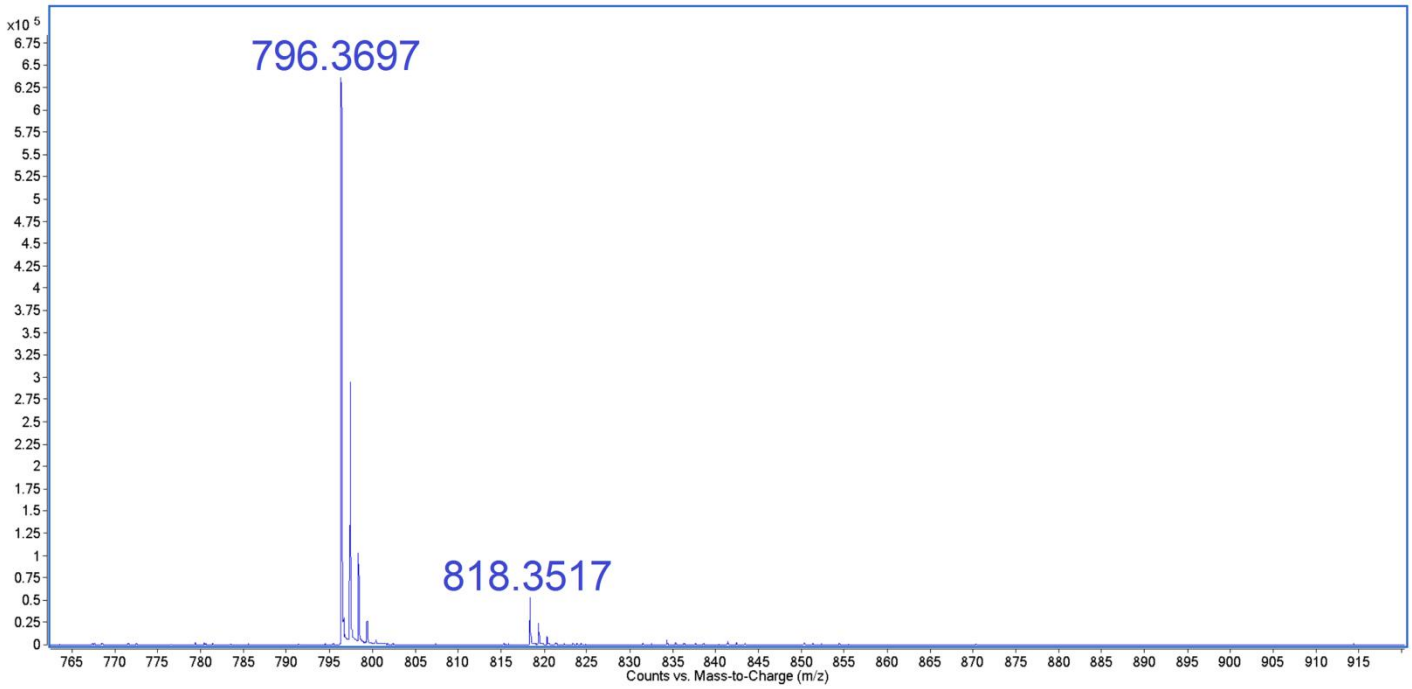
Fragmentation	Obs. mass	Calc. mass	Error [Da]	Obs. mass	Calc. mass	Error [Da]	Note
	precolibactin-886 (**10**)			precolibactin-969 (**11**)			
1	870.3542	870.3525	0.0017	953.3552	953.3532	0.0020	+ C_3HNO_2
2	852.3424	852.3419	0.0005	935.3396	935.3426	-0.0030	+ C_3HNO_2
3	660.1536	660.1541	-0.0005	743.1495	743.1548	-0.0053	+ C_3HNO_2
4	643.1277	643.1275	0.0002	726.1205	726.1283	-0.0078	+ C_3HNO_2
5	591.3743	591.3752	-0.0009	591.3703	591.3752	-0.0049	Identical
6	574.3495	574.3487	0.0008	574.3454	574.3487	-0.0033	Identical
7	545.1270	545.1272	-0.0002	628.1254	628.1279	-0.0025	+ C_3HNO_2
8	528.1000	528.1006	-0.0006	611.0982	611.1013	-0.0031	+ C_3HNO_2
9	527.1163	527.1166	-0.0003	610.1174	610.1173	0.0001	+ C_3HNO_2
10	466.3270	466.3275	-0.0005	466.3251	466.3275	-0.0024	Identical
11	449.2999	449.3010	-0.0011	449.2973	449.3010	-0.0037	Identical
12	433.3021	433.3061	-0.0040	433.3004	433.3061	-0.0057	Identical
13	422.0582	422.0587	-0.0005	505.0572	505.0595	-0.0023	+ C_3HNO_2
14	405.0318	405.0322	-0.0004	488.0301	488.0329	-0.0028	+ C_3HNO_2
15	297.0101	297.0111	-0.0010	380.0057	380.0118	-0.0061	+ C_3HNO_2
16	249.1218	249.1234	-0.0016	249.1202	249.1234	-0.0032	Identical
17	142.0843	142.0863	-0.0020	142.0842	142.0863	-0.0021	Identical
18	124.0742	124.0757	-0.0015	124.0731	124.0757	-0.0026	Identical

Supplementary Figure 7. a, HR-MS/MS fragmentation pattern of precolibactin-886 (**10**). Fragmentation was acquired with collision energies of 15 and 30 V. **b**, HR-MS/MS fragmentation pattern of precolibactin-969 (**11**). Fragmentation was acquired with collision energies of 15 and 30 V. **c**, The major fragmentation species from HR-MS/MS measurement of **10**. **d**, The major fragmentation species from HR-MS/MS measurement of **11**. **e**, A comparison between the fragmentation species of **10** and **11**. Obs. = observed; Calc. = calculated.

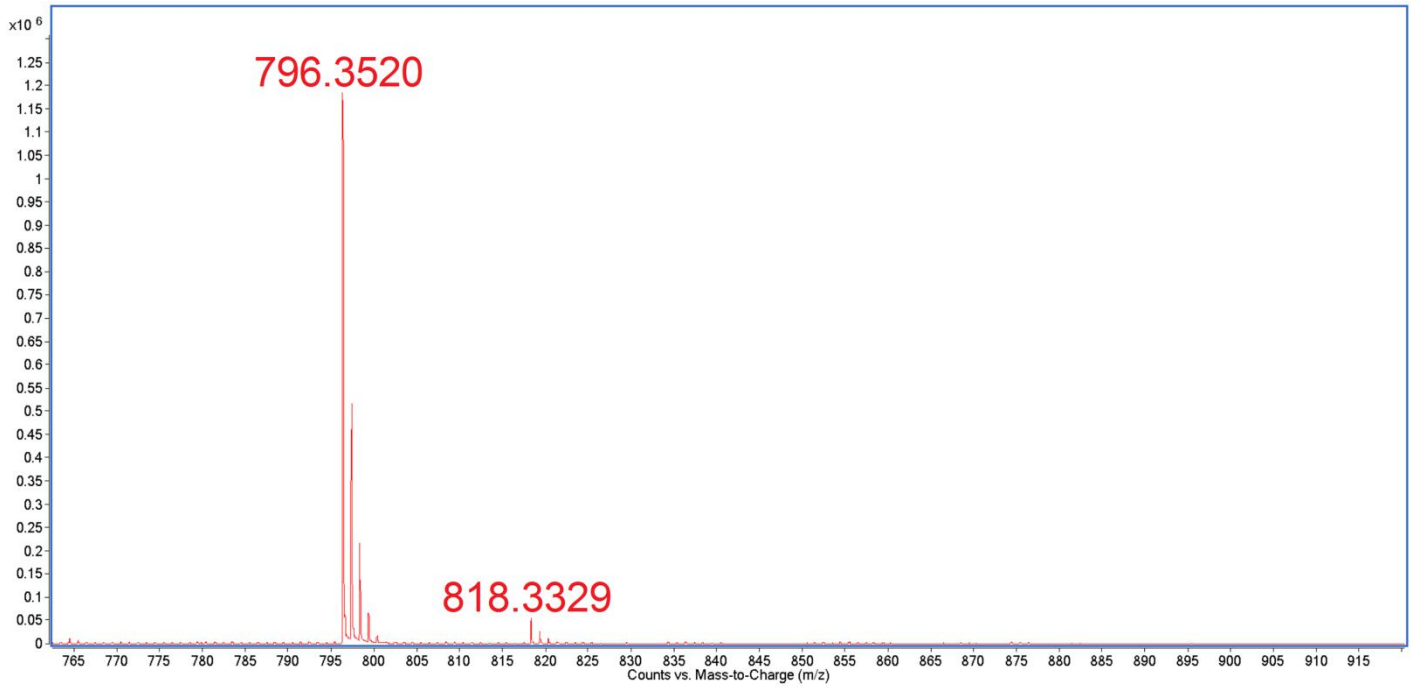


Supplementary Figure 8. Maximum-likelihood phylogenetic analysis of ClbK$_{PKS}$ and ClbO. Molecular phylogenetic analysis by the maximum likelihood method to indicates the relationships among the ketosynthase (KS) of ClbK$_{PKS}$, ClbO, and 78 other representative polyketide synthase (PKS) domains. The tree with the highest log likelihood (-43888.4672) is shown. The tree is drawn to scale with branch lengths measured in the number of substitutions per site.

a

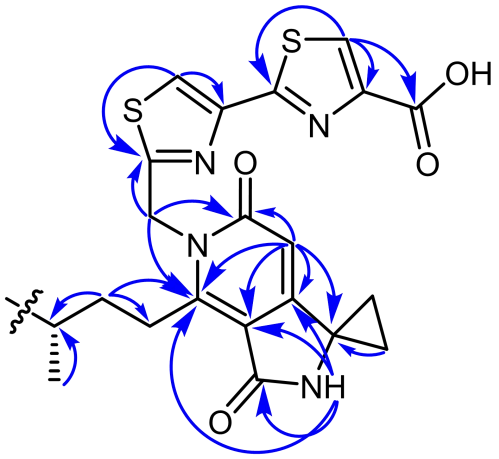


b



Supplementary Figure 9. The high-resolution ESI-MS of precolibactin-795a (**8**) (**a**) and precolibactin-795b (**9**) (**b**). Their molecular formulas were determined as $C_{39}H_{53}N_7O_9S_1$ and $C_{39}H_{53}N_7O_7S_2$ on the basis of their protonated molecular ion peaks at *m*/*z* 796.3697 (calculated, 796.3698) and 796.3520 (calculated, 796.3521), respectively.

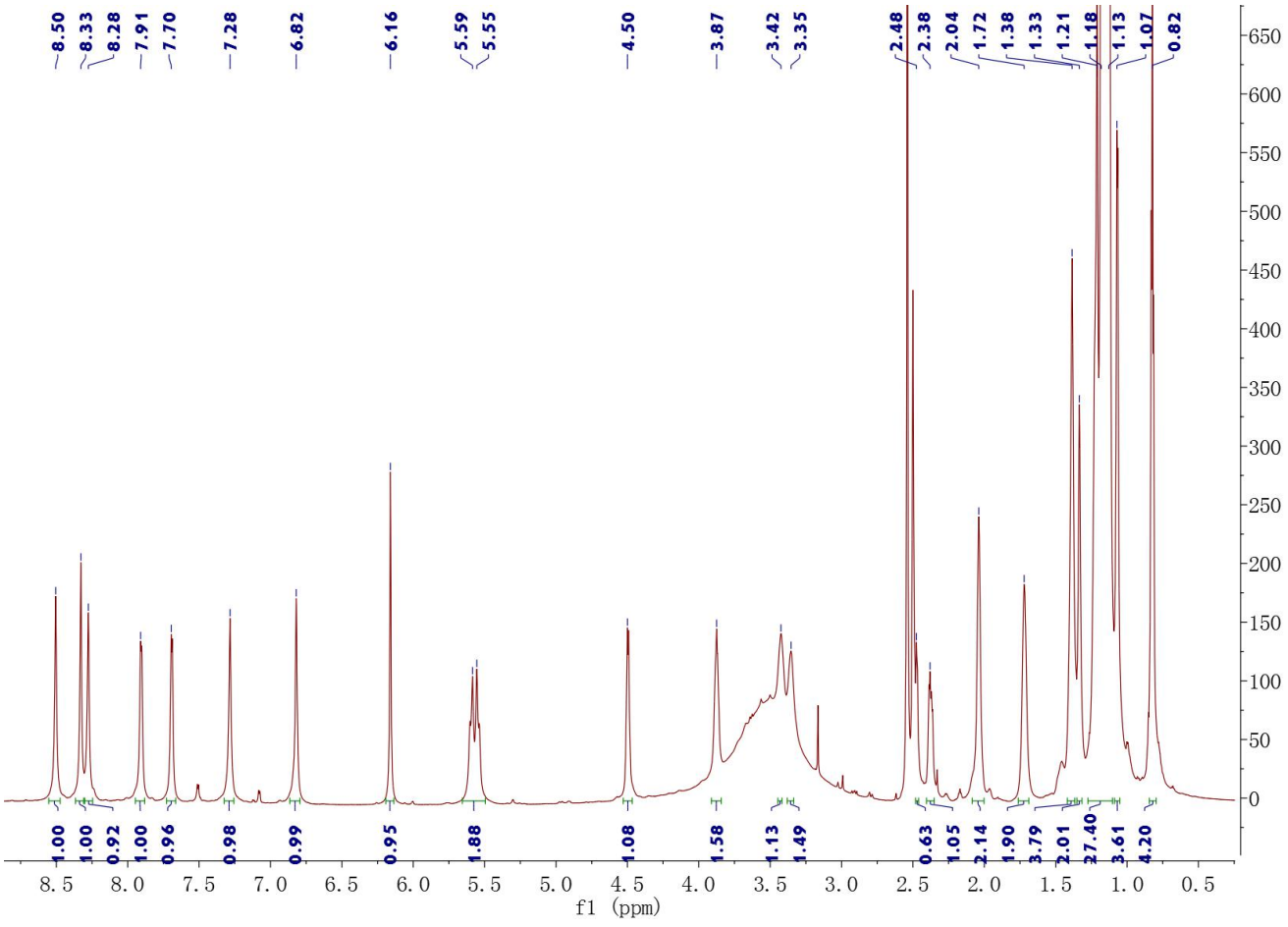
a


S
OH
S
N
O
N
O
NH
O

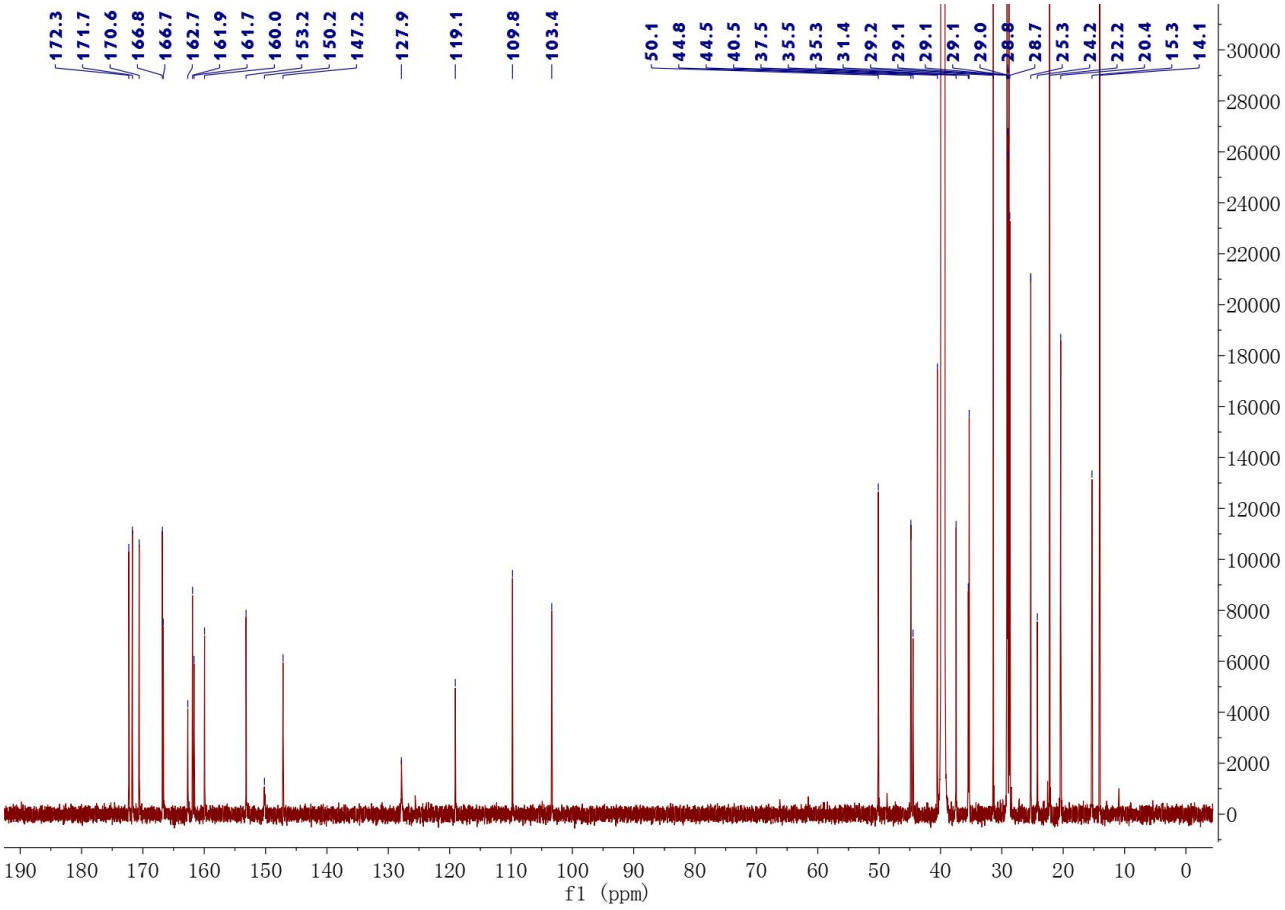
9

Key HMBC

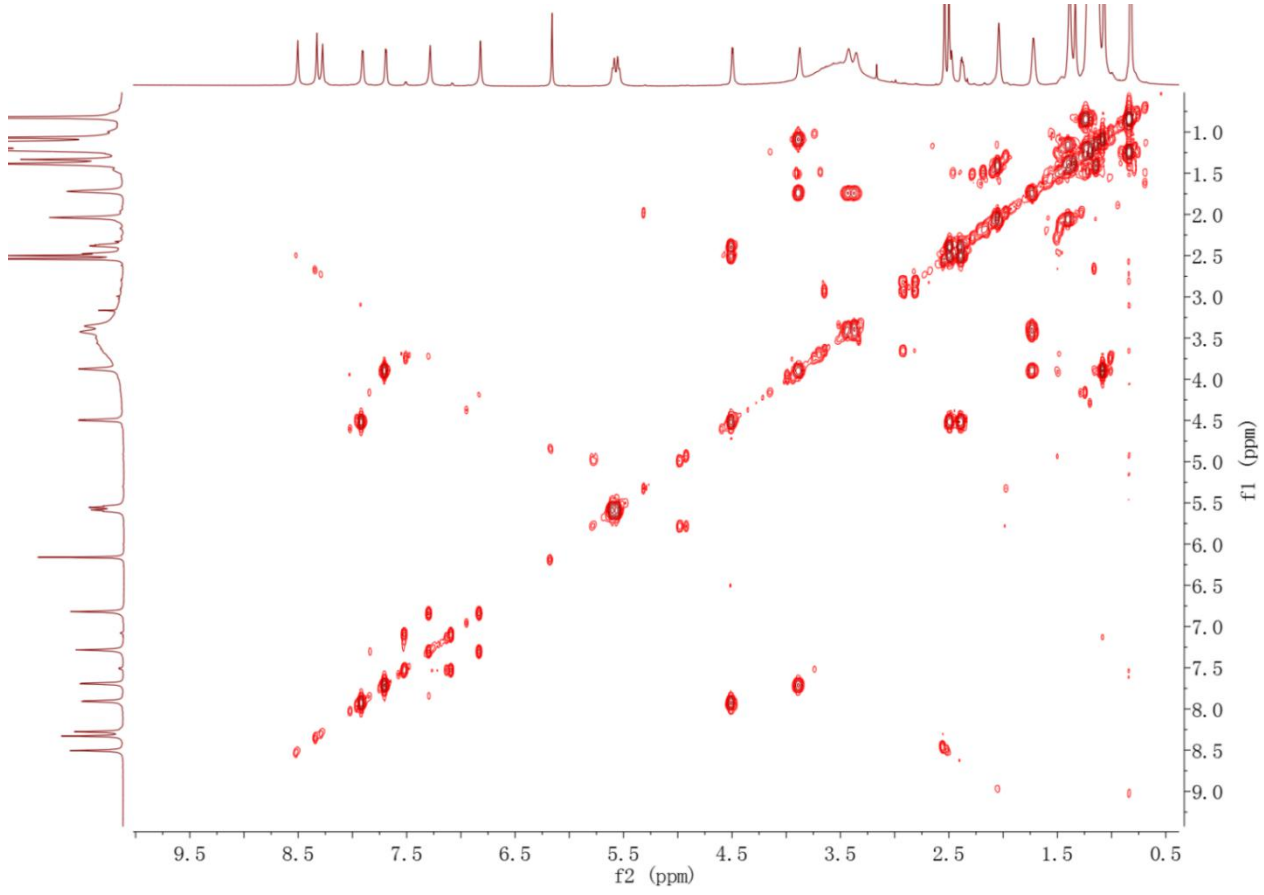
b


8.50
8.33
8.28
7.91
7.70
7.28
6.82
6.16
5.59
5.55
4.50
3.87
3.42
3.35
2.48
2.38
2.04
1.72
1.38
1.33
1.21
1.18
1.13
1.07
0.82
1.00
1.00
0.92
1.00
0.96
0.98
0.99
0.95
1.88
1.08
1.58
1.13
1.49
0.63
1.05
2.14
1.90
3.79
2.01
27.40
3.61
4.20
f1 (ppm)

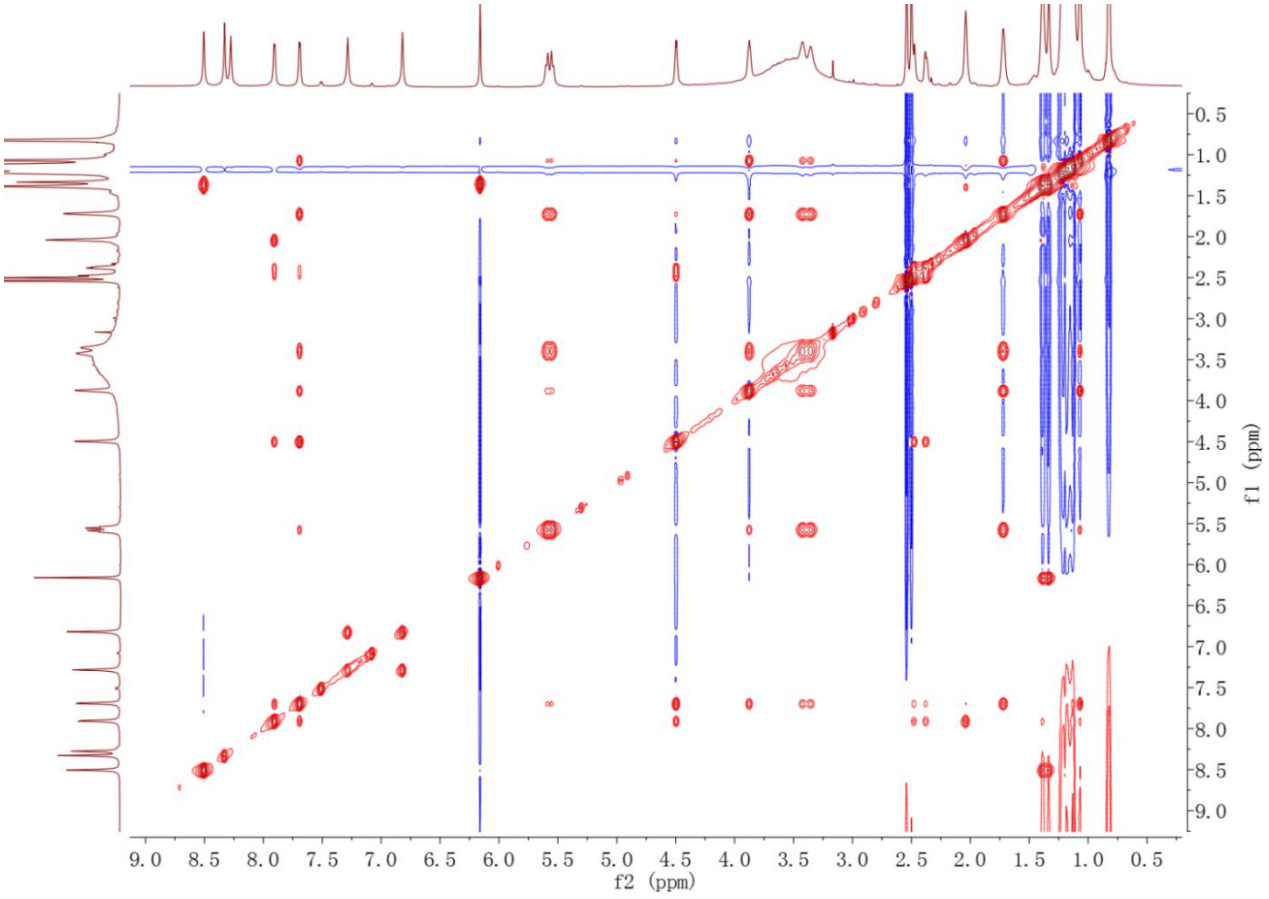
c


172.3
171.7
170.6
166.8
166.7
162.7
161.9
161.7
160.0
153.2
150.2
147.2
127.9
119.1
109.8
103.4
50.1
44.8
44.5
40.5
37.5
35.5
35.3
31.4
29.2
29.1
29.1
29.1
29.0
28.8
28.7
25.3
24.2
22.2
20.4
15.3
14.1
f1 (ppm)

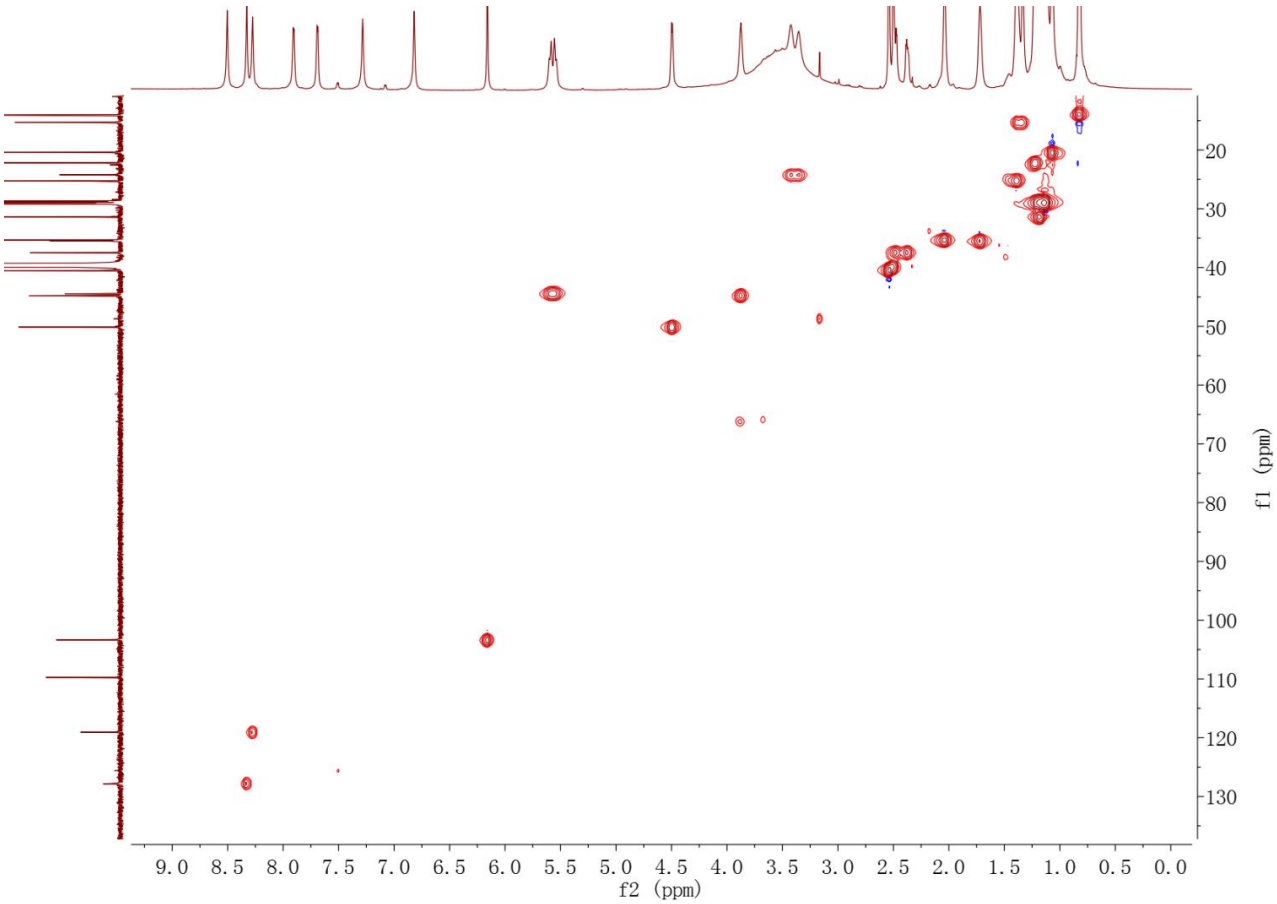
d


f2 (ppm)
f1 (ppm)

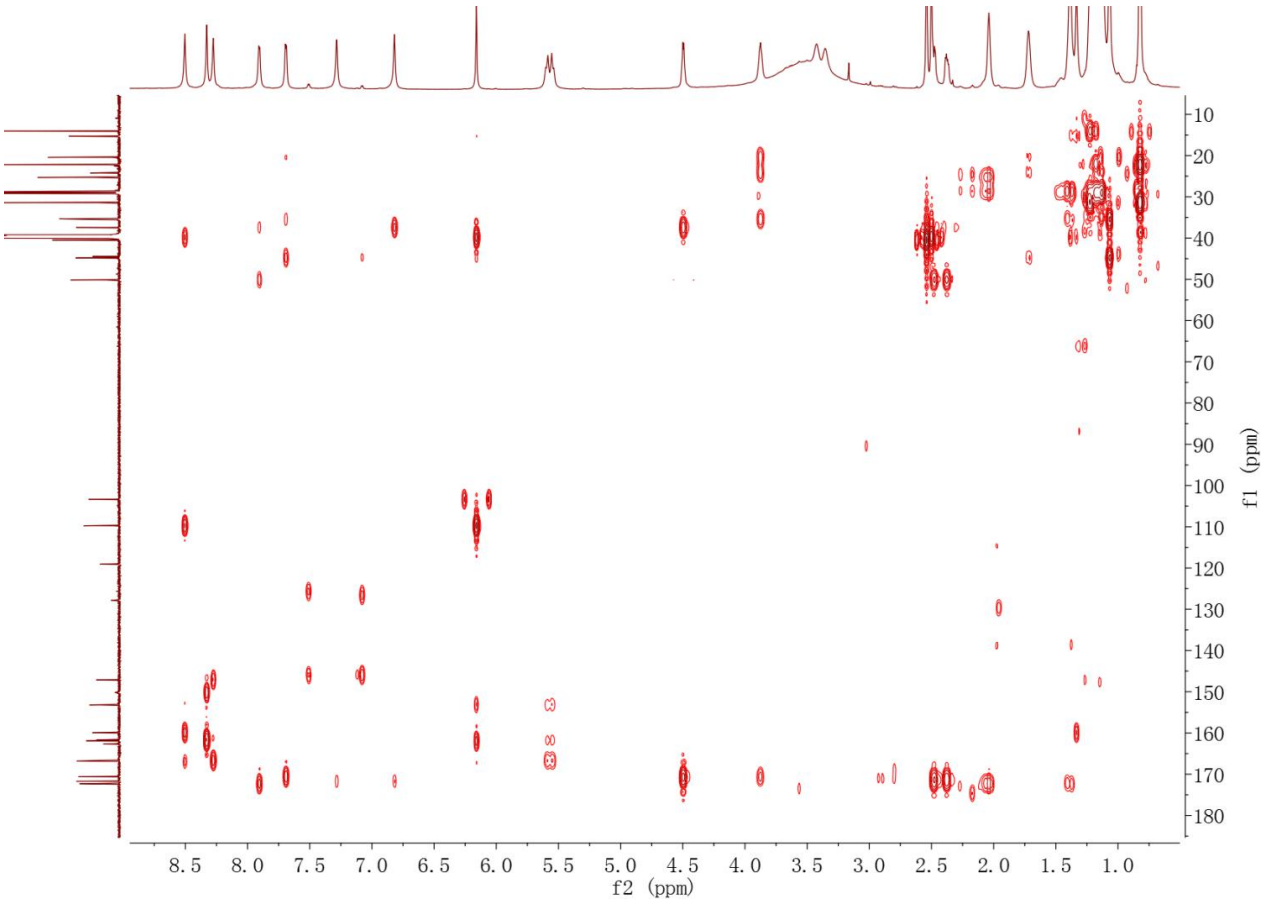
e



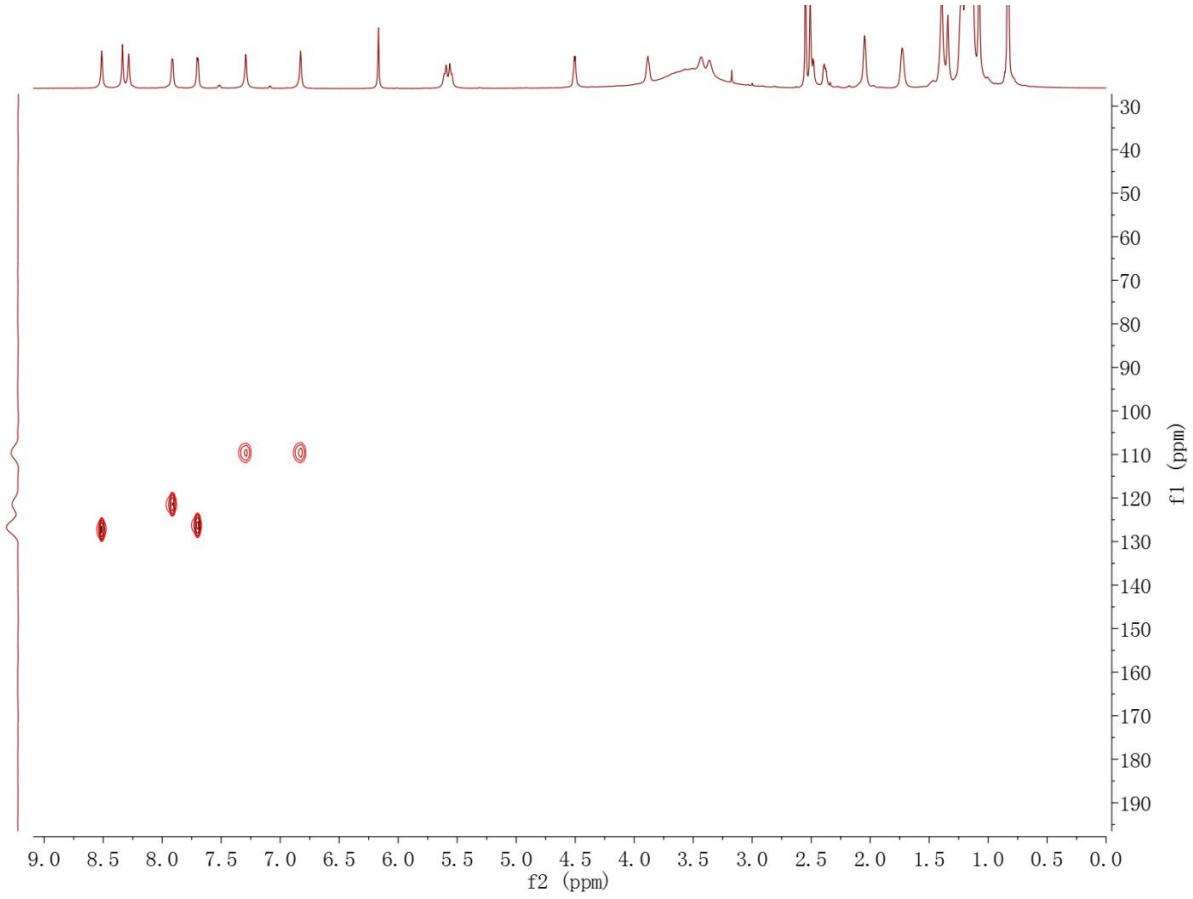
f



g

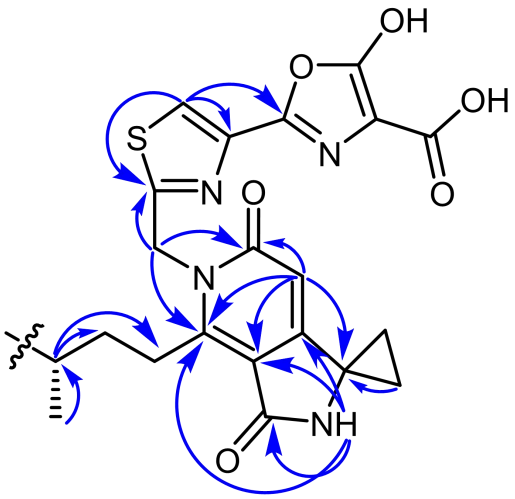


h



Supplementary Figure 10. 1D and 2D NMR spectra of precolibactin-795b (**9**). **a**, NMR-based key correlations for the structural assignment of **9**. **b**, ^{1}H NMR spectrum, **c**, ^{13}C NMR spectrum, **d**, ^{1}H–^{1}H COSY spectrum, **e**, ^{1}H–^{1}H NOESY spectrum, **f**, ^{1}H–^{13}C HSQC spectrum, **g**, ^{1}H–^{13}C HMBC spectrum, and **h**, ^{1}H–^{15}N HSQC spectrum. Recorded with 850 MHz for ^{1}H and 212.5 MHz for ^{13}C, cold probe, in DMSO-d_6.

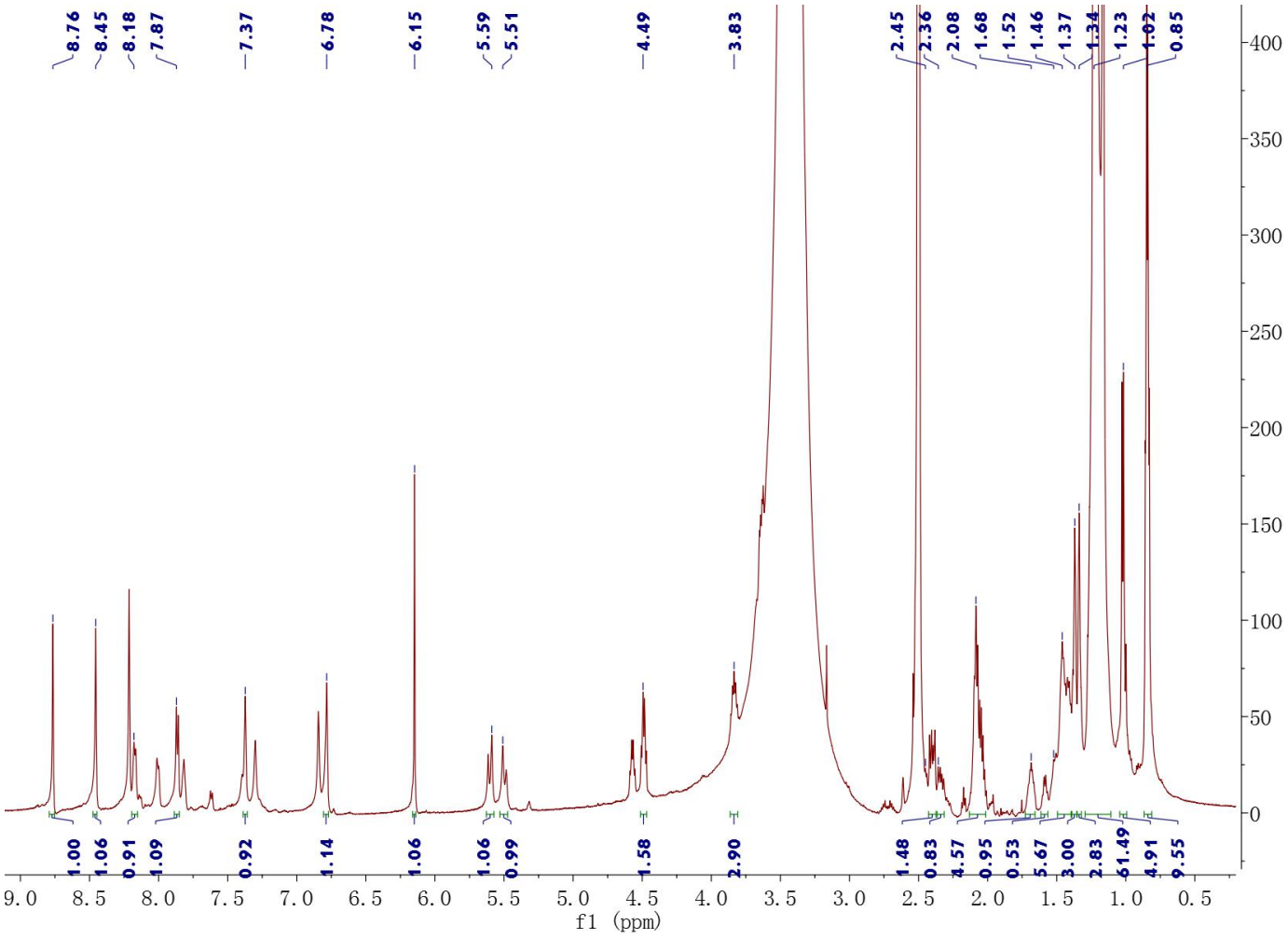
a


OH
O
OH
S
N
N
O
O
N
NH
O

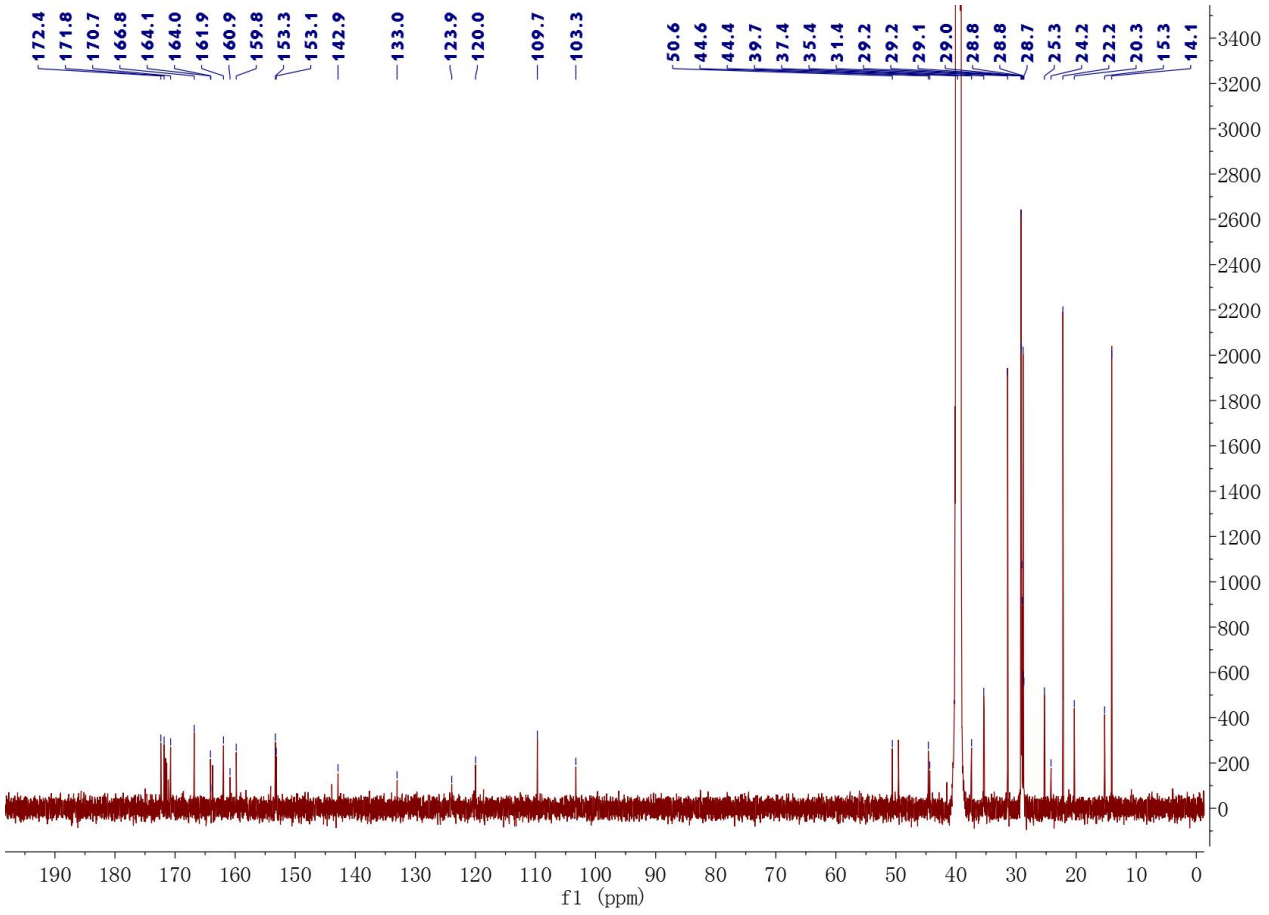
8

Key HMBC

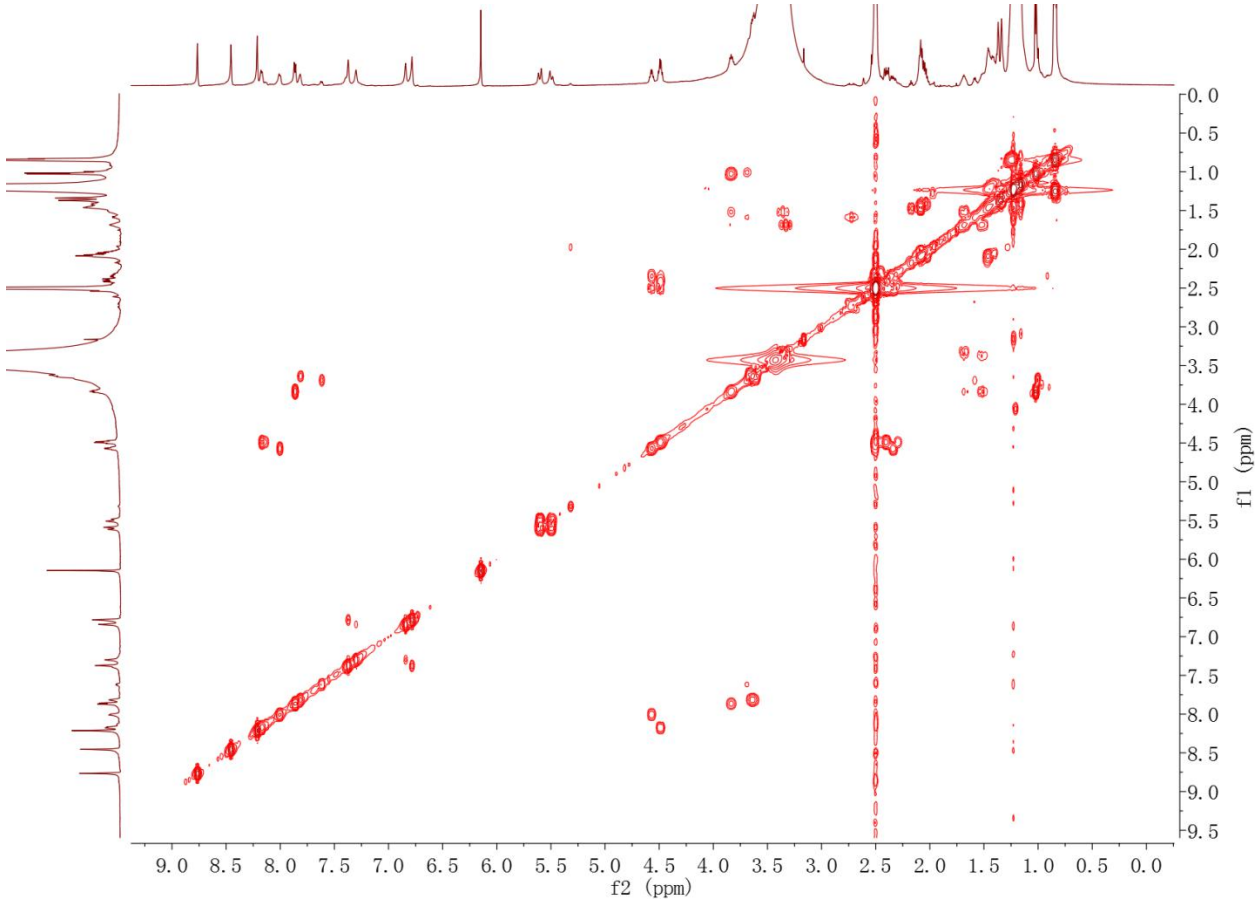
b


8.76
8.45
8.18
7.87
7.37
6.78
6.15
5.59
5.51
4.49
3.83
2.45
2.36
2.08
1.68
1.52
1.46
1.37
1.34
1.23
1.02
0.85
1.00
1.06
0.91
1.09
0.92
1.14
1.06
1.06
0.99
1.58
2.90
1.48
0.83
4.57
0.95
0.53
5.67
3.00
2.83
61.49
4.91
9.55
f1 (ppm)

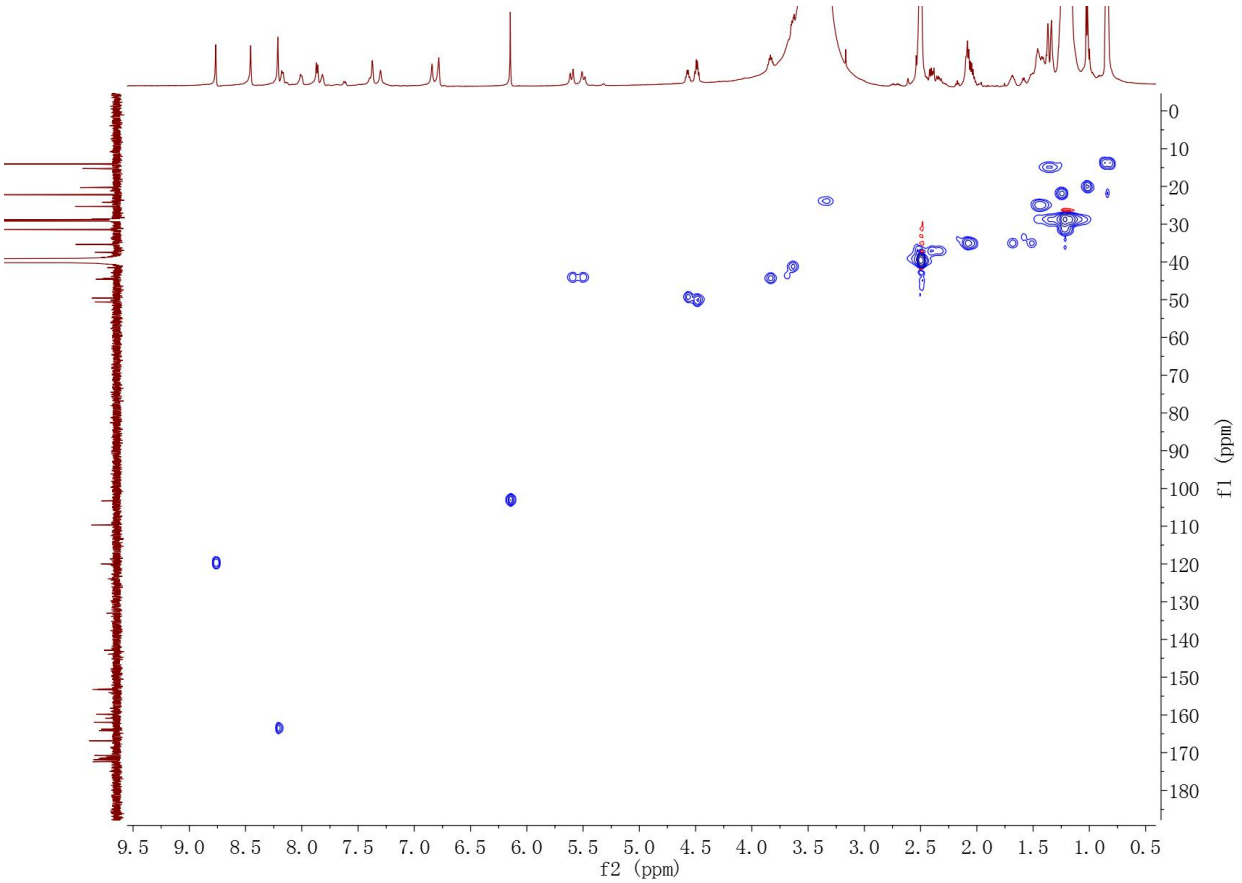
c


172.4
171.8
170.7
166.8
164.1
164.0
161.9
160.9
159.8
153.3
153.1
142.9
133.0
123.9
120.0
109.7
103.3
50.6
44.6
44.4
39.7
37.4
35.4
31.4
29.2
29.2
29.1
29.0
28.8
28.8
28.7
25.3
24.2
22.2
20.3
15.3
14.1
f1 (ppm)

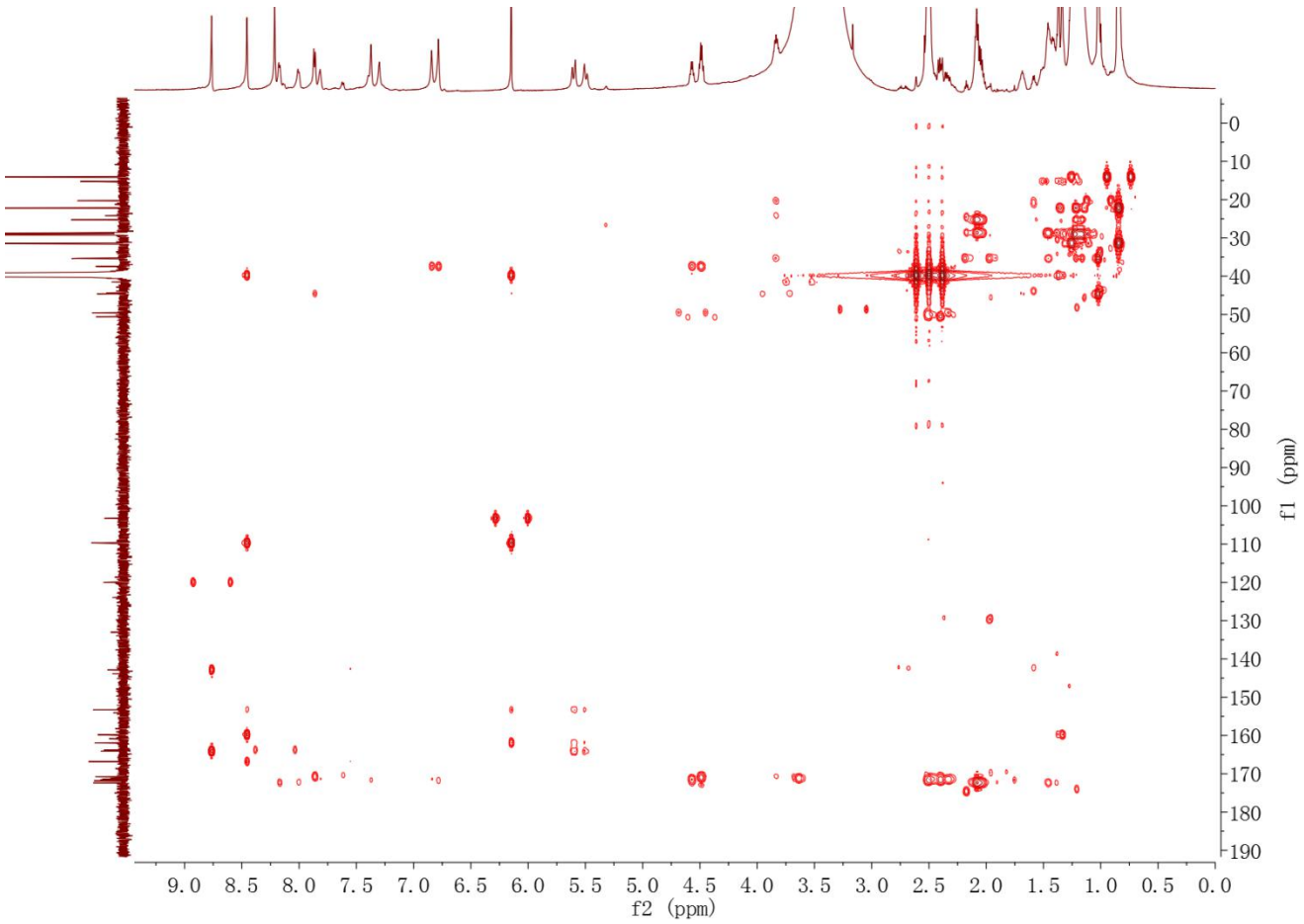
d


f2 (ppm)
f1 (ppm)

e

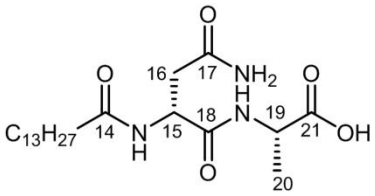


f



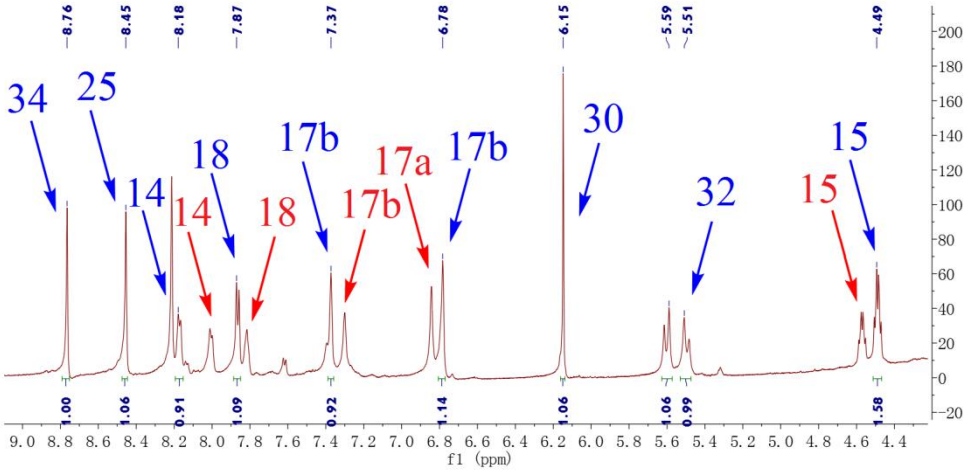
g

1

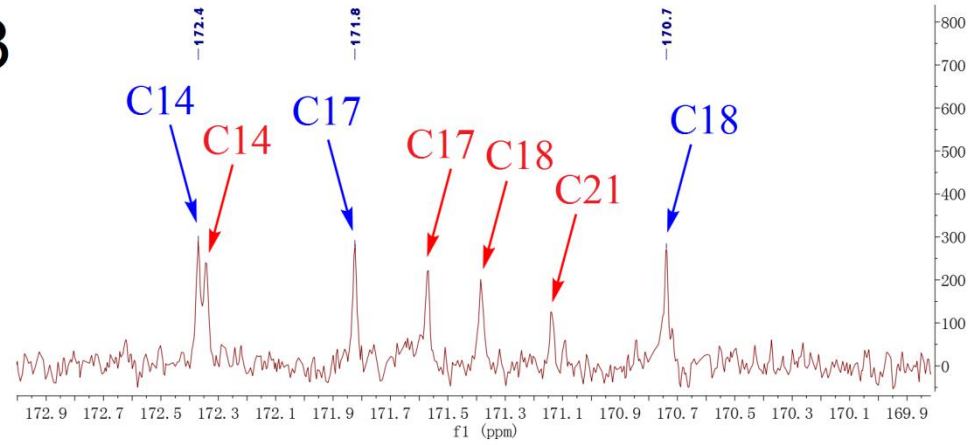


1 precolibactin-413

2

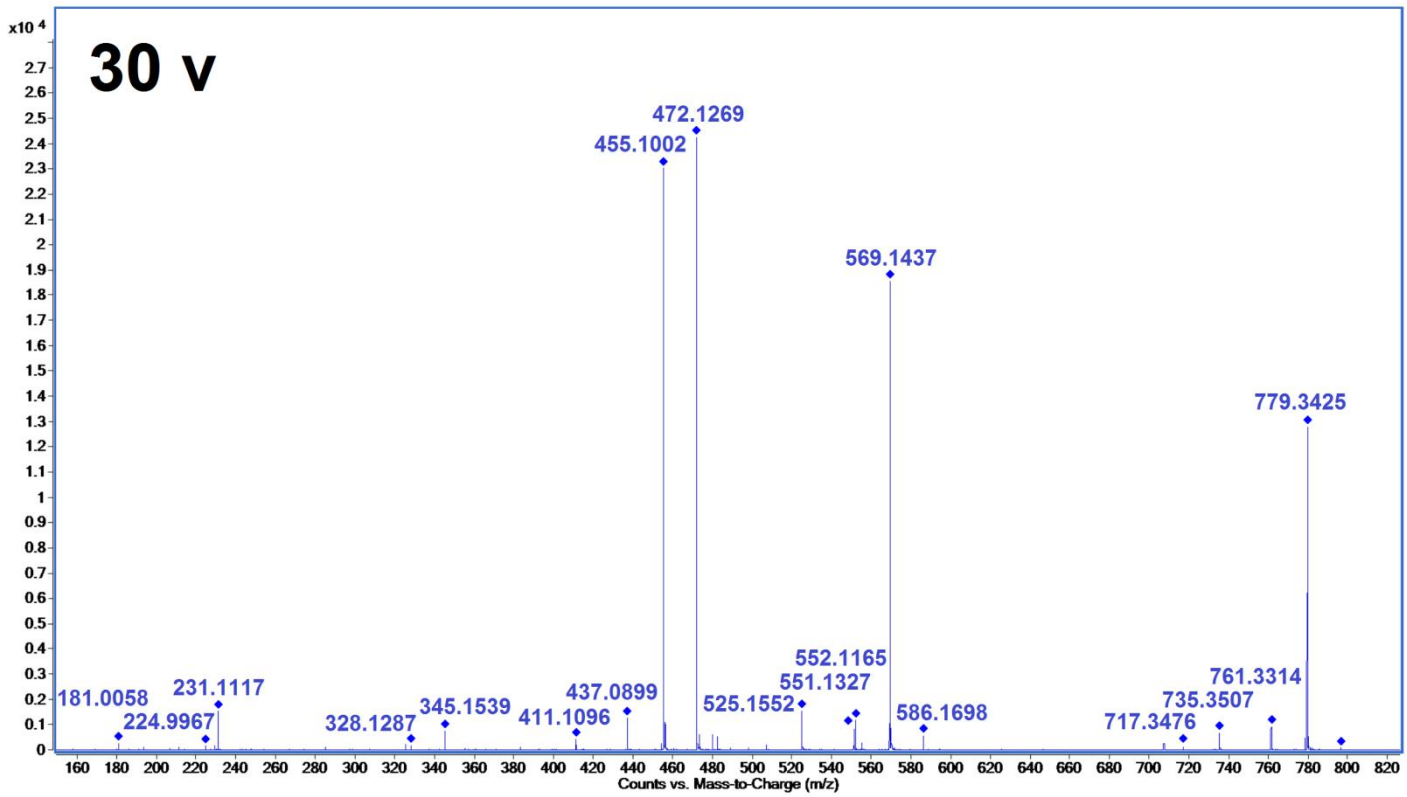


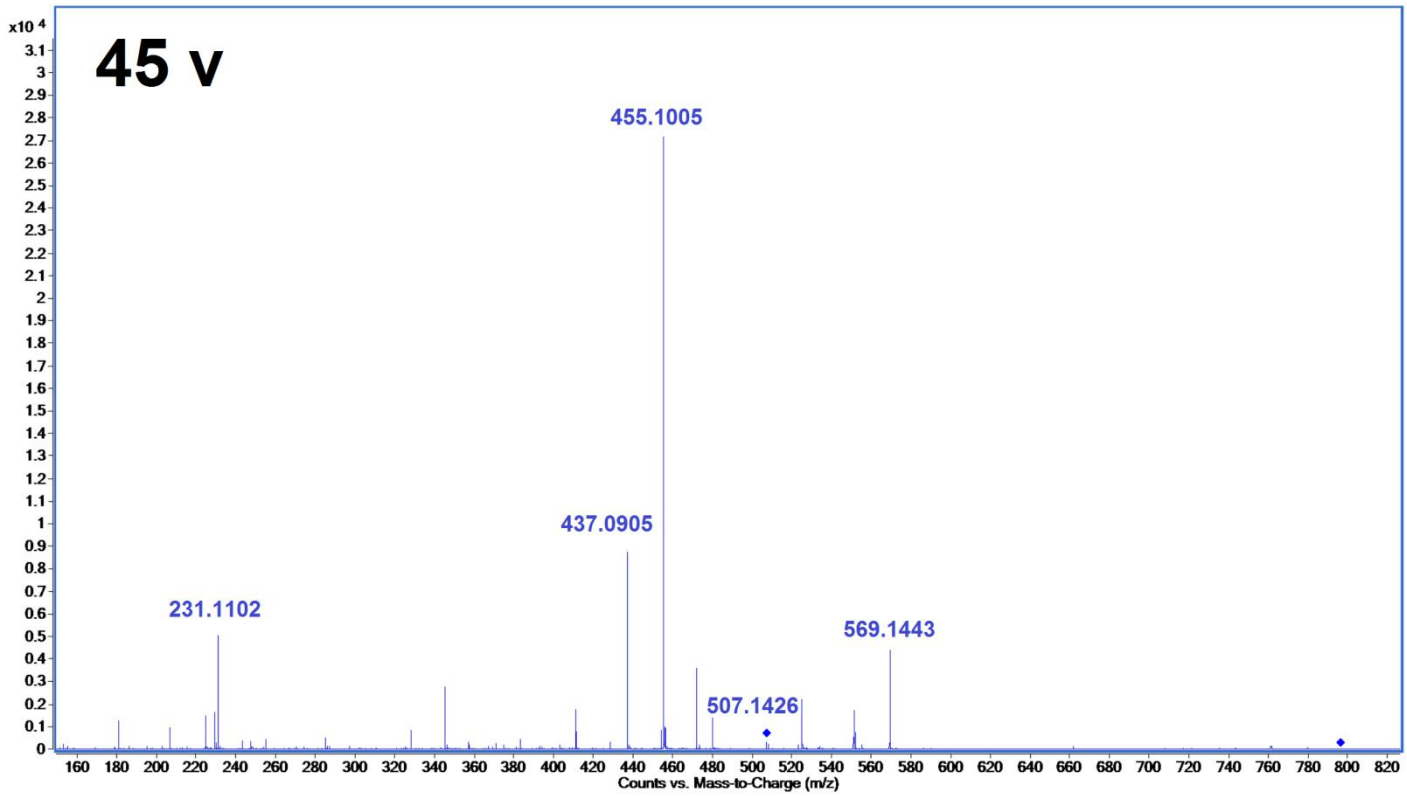
3



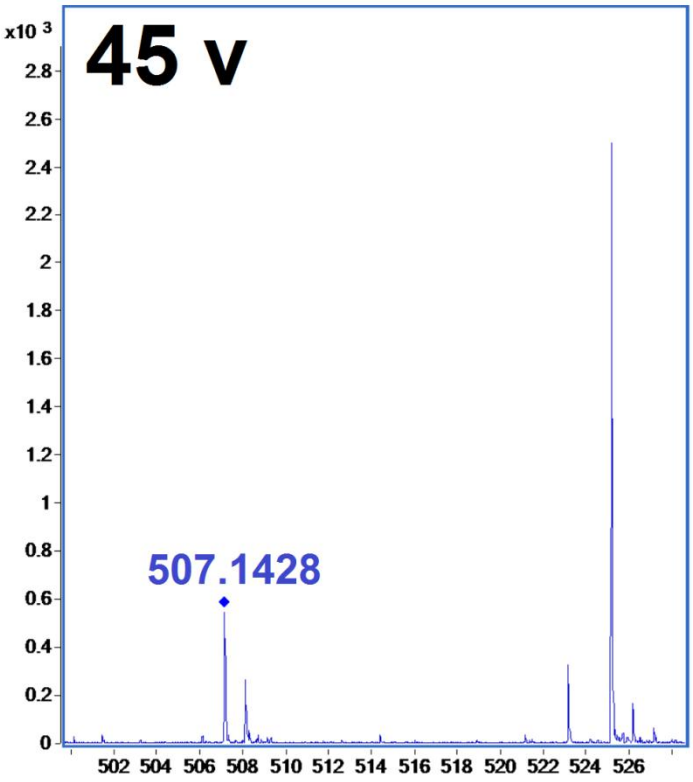
Supplementary Figure 11. 1D and 2D NMR spectra of precolibactin-795a (**8**). **a**, NMR-based key correlations for the structural assignment of **8**. **b**, ^{1}H NMR spectrum, **c**, ^{13}C NMR spectrum, **d**, ^{1}H–^{1}H COSY spectrum, **e**, ^{1}H–^{13}C HSQC spectrum, and **f**, ^{1}H–^{13}C HMBC spectrum. Recorded with 850 MHz for ^{1}H and 212.5 MHz for ^{13}C, cold probe, in DMSO-*d*$_6$. In the NMR test, **8** was mixed a small quantity of precolibactin-413 (**1**) (**g1**). We labeled the overlapping ^{1}H NMR signals (**g2**) and ^{13}C NMR signals (**g3**) of **8** and **1** with blue (for **8**) and red (for **1**) arrows.

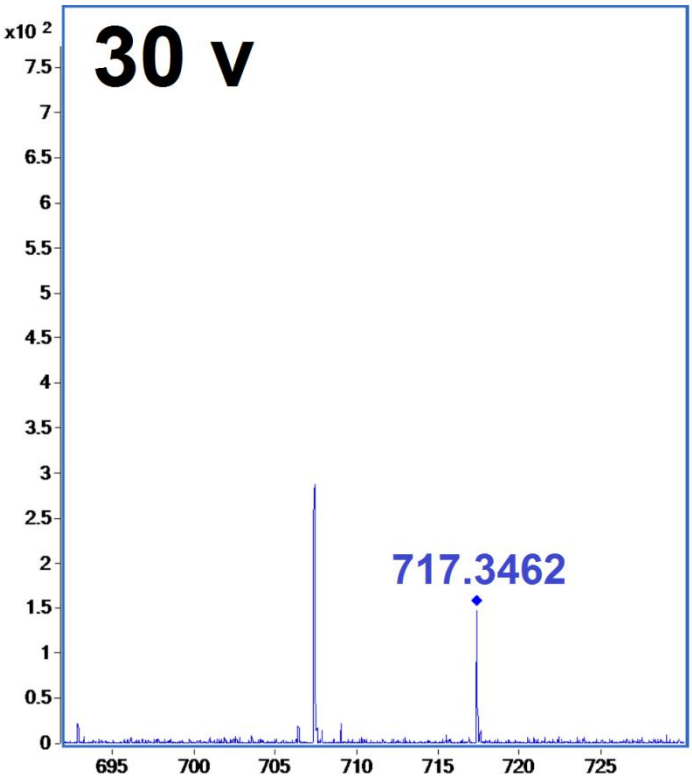
a


30 v
x10 4
472.1269
455.1002
569.1437
779.3425
552.1165
551.1327
761.3314
735.3507
181.0058
231.1117
224.9967
345.1539
328.1287
437.0899
411.1096
525.1552
586.1698
717.3476
Counts vs. Mass-to-Charge (m/z)

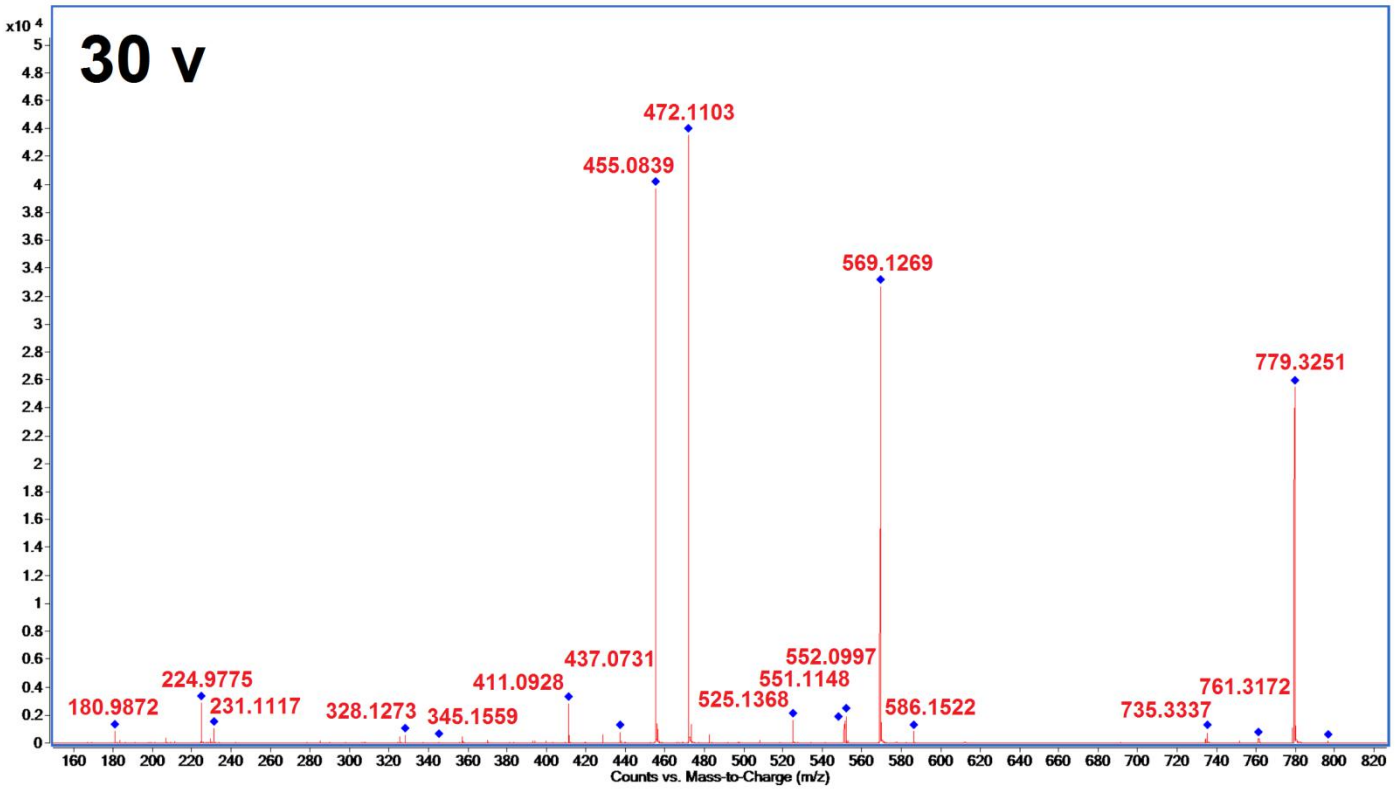

45 v
x10 4
455.1005
437.0905
231.1102
569.1443
507.1426
Counts vs. Mass-to-Charge (m/z)

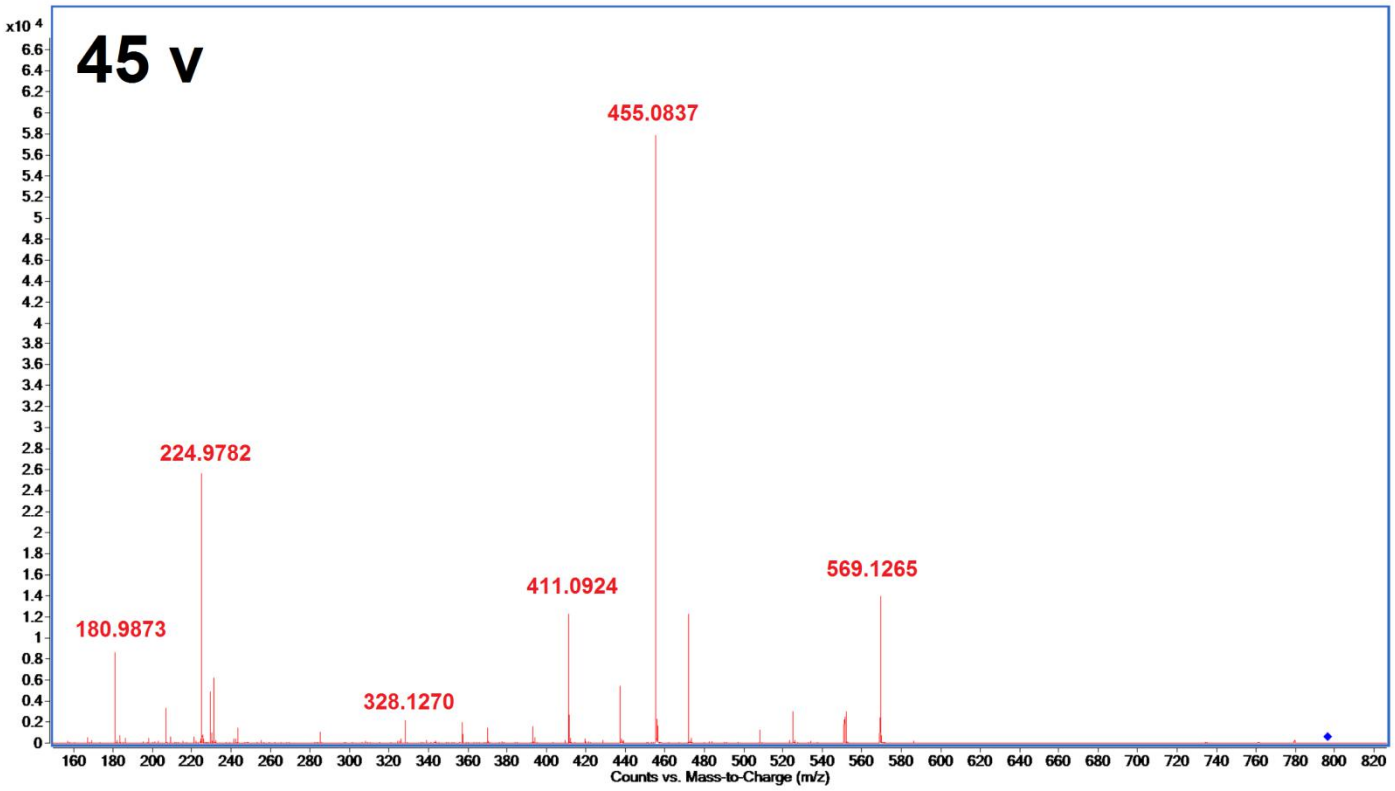
b


45 v
x10 3
2.8
2.6
2.4
2.2
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
507.1428
502 504 506 508 510 512 514 516 518 520 522 524 526

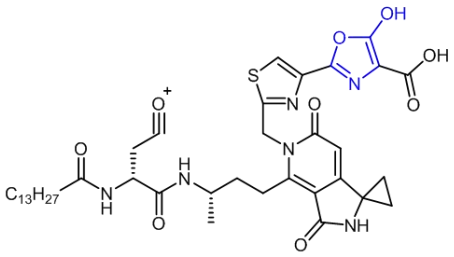

30 v
x10 2
7.5
7
6.5
6
5.5
5
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
717.3462
695 700 705 710 715 720 725

C

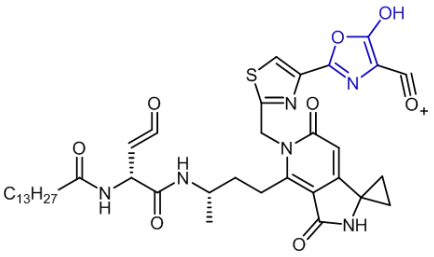

30 v
x10 4
472.1103
455.0839
569.1269
779.3251
224.9775
180.9872
231.1117
328.1273
345.1559
411.0928
437.0731
552.0997
551.1148
525.1368
586.1522
761.3172
735.3337
Counts vs. Mass-to-Charge (m/z)


45 v
x10 4
455.0837
224.9782
569.1265
411.0924
180.9873
328.1270
Counts vs. Mass-to-Charge (m/z)

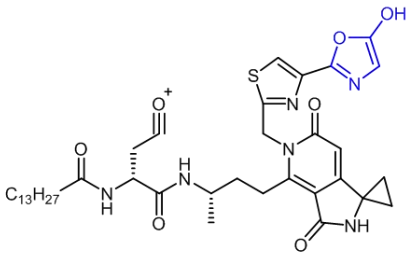
d



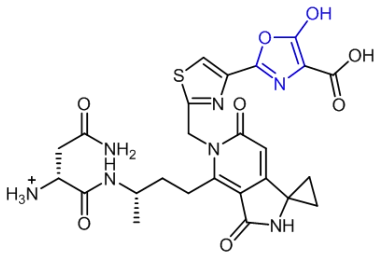
Frag. 1 (779.3433)



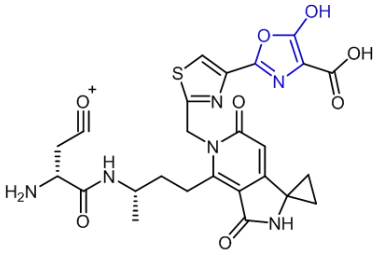
Frag. 2 (761.3327)



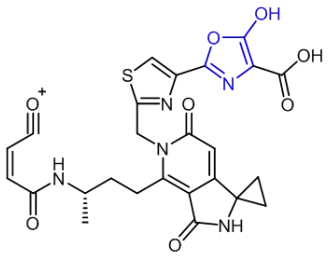
Frag. 3 (735.3534)



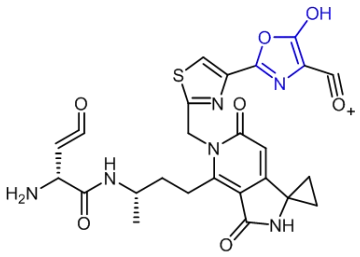
Frag. 4 (586.1715)



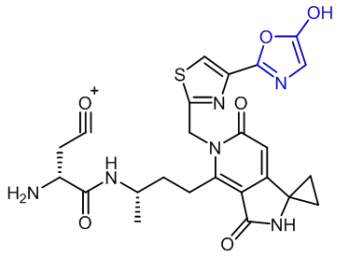
Frag. 5 (569.1449)



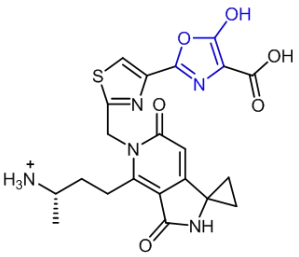
Frag. 6 (552.1184)



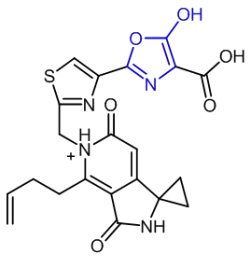
Frag. 7 (551.1343)



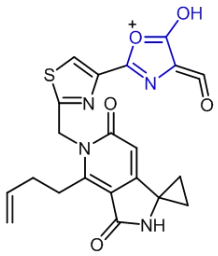
Frag. 8 (525.1551)



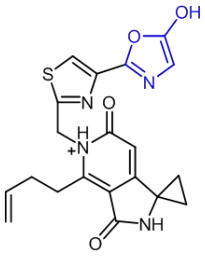
Frag. 9 (472.1285)



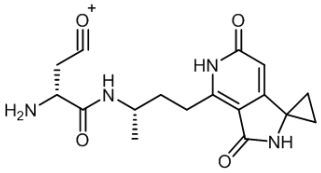
Frag. 10 (455.1020)



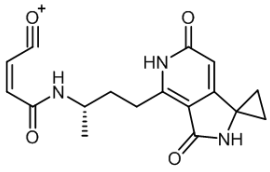
Frag. 11 (437.0914)



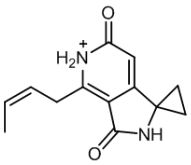
Frag. 12 (411.1122)



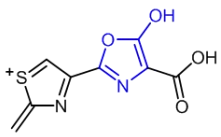
Frag. 13 (345.1557)



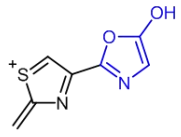
Frag. 14 (328.1292)



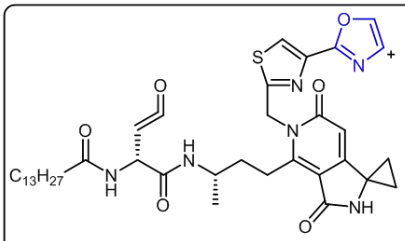
Frag. 15(231.1128)



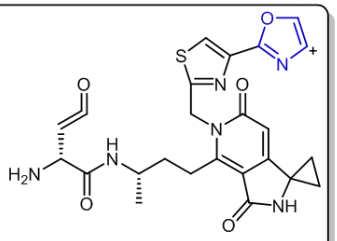
Frag. 16 (224.9965)



Frag. 17 (181.0066)

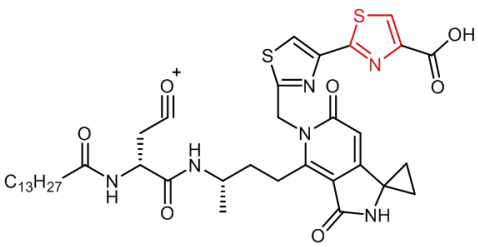


Frag. 18 (717.3429)

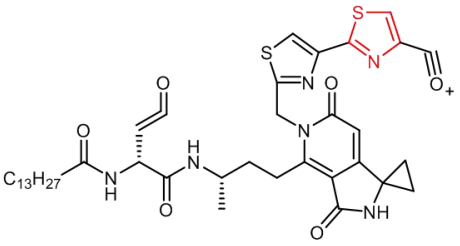


Frag. 19 (507.1445)

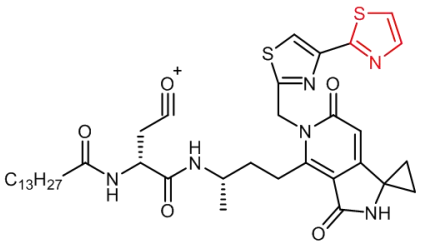
e



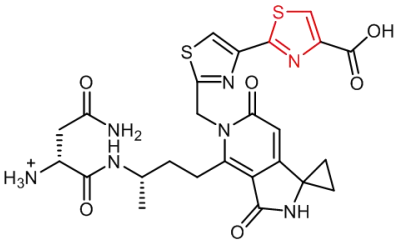
Frag. 1 (779.3255)



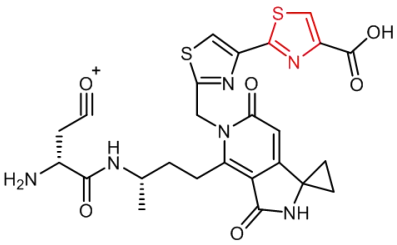
Frag. 2 (761.3150)



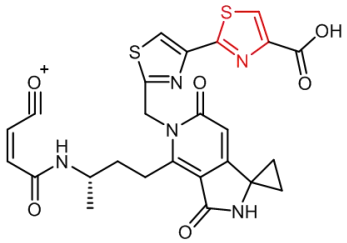
Frag. 3 (735.3357)



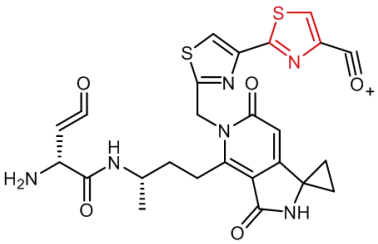
Frag. 4 (586.1537)



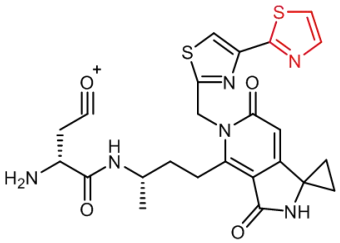
Frag. 5 (569.1272)



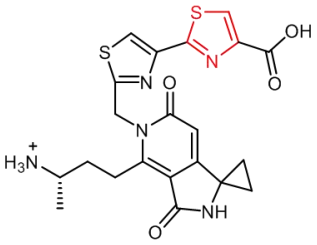
Frag. 6 (552.1006)



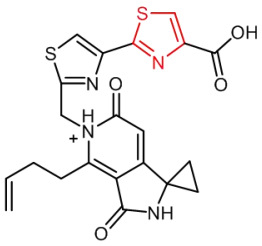
Frag. 7 (551.1166)



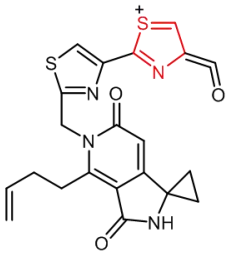
Frag. 8 (525.1373)



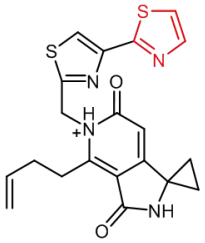
Frag. 9 (472.1108)



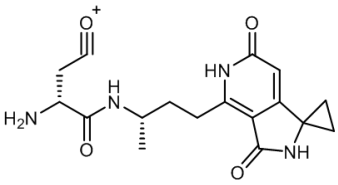
Frag. 10 (455.0842)



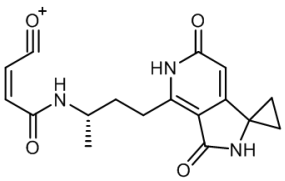
Frag. 11 (437.0737)



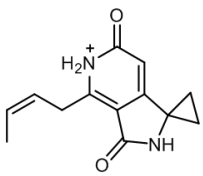
Frag. 12 (411.0944)



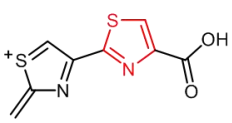
Frag. 13 (345.1557)



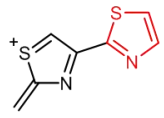
Frag. 14 (328.1292)



Frag. 15 (231.1128)



Frag. 16 (224.9787)



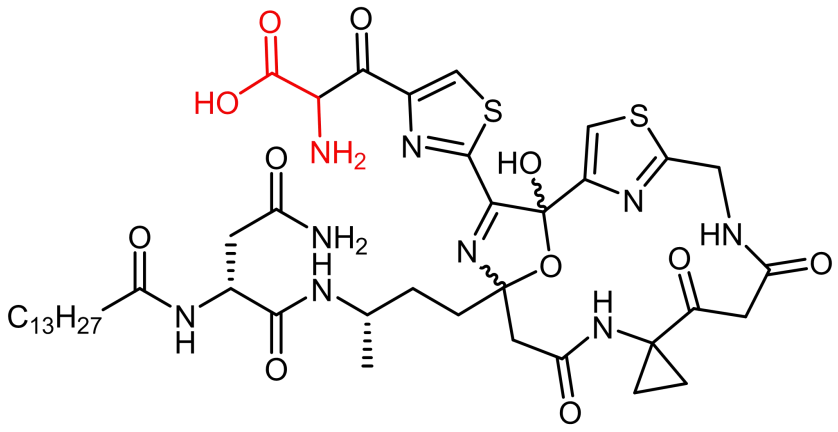
Frag. 17 (180.9889)

f

Fragmentation	Obs. mass	Calc. mass	Error [Da]	Obs. mass	Calc. mass	Error [Da]	Note
	precolibactin-795a (**8**)			precolibactin-795b (**9**)			
1	779.3425	779.3433	-0.0008	779.3251	779.3255	-0.0004	C_3HNO_2 *vs* C_3HNS
2	761.3314	761.3327	-0.0013	761.3172	761.3150	0.0022	C_3HNO_2 *vs* C_3HNS
3	735.3507	735.3534	-0.0027	735.3337	735.3357	-0.0020	C_3HNO_2 *vs* C_3HNS
4	586.1698	586.1715	-0.0017	586.1522	586.1537	-0.0015	C_3HNO_2 *vs* C_3HNS
5	569.1437	569.1449	-0.0012	569.1269	569.1272	-0.0003	C_3HNO_2 *vs* C_3HNS
6	552.1165	552.1184	-0.0019	552.0997	552.1006	-0.0009	C_3HNO_2 *vs* C_3HNS
7	551.1327	551.1343	-0.0016	551.1148	551.1166	-0.0018	C_3HNO_2 *vs* C_3HNS
8	525.1552	525.1551	0.0001	525.1368	525.1373	-0.0005	C_3HNO_2 *vs* C_3HNS
9	472.1269	472.1285	-0.0016	472.1103	472.1108	-0.0005	C_3HNO_2 *vs* C_3HNS
10	455.1002	455.1020	-0.0018	455.0839	455.0842	-0.0003	C_3HNO_2 *vs* C_3HNS
11	437.0899	437.0914	-0.0015	437.0731	437.0737	-0.0006	C_3HNO_2 *vs* C_3HNS
12	411.1096	411.1122	-0.0026	411.0928	411.0944	-0.0016	C_3HNO_2 *vs* C_3HNS
13	345.1539	345.1557	-0.0018	345.1559	345.1557	0.0002	Identical
14	328.1287	328.1292	-0.0005	328.1273	328.1292	-0.0019	Identical
15	231.1117	231.1128	-0.0011	231.1117	231.1128	-0.0011	Identical
16	224.9967	224.9965	0.0002	224.9775	224.9787	-0.0012	C_3HNO_2 *vs* C_3HNS
17	181.0058	181.0066	-0.0008	180.9872	180.9889	-0.0017	C_3HNO_2 *vs* C_3HNS
18	717.3476	717.3429	0.0047				Discrepant fragment
19	507.1426	507.1445	-0.0019				Discrepant fragment

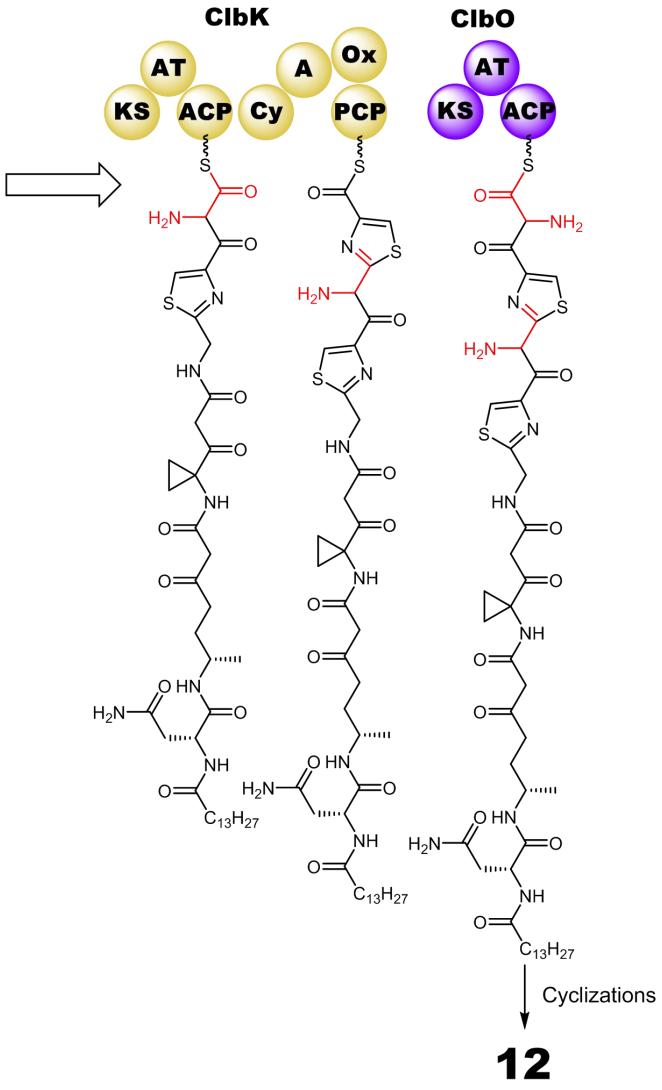
Supplementary Figure 12. a, HR-MS/MS fragmentation pattern of precolibactin-795a (**8**). Fragmentation was acquired with collision energies of 30 and 45 V. **b**, Tow highlighted HR-MS/MS fragments displaying the difference in HR-MS/MS fragmentation pattern between **8** and precolibactin-795b (**9**). **c**, HR-MS/MS fragmentation pattern of **9**. Fragmentation was acquired with collision energies of 30 and 45 V. **d**, The major fragmentation species from HR-MS/MS measurement of **8**. Tow discrepant fragmentation species of **8** from **9** are highlighted. **e**, The major fragmentation species from HR-MS/MS measurement of **9**. **f**, A comparison between the fragmentation species of **8** and **9**. Obs. = observed; Calc. = calculated.

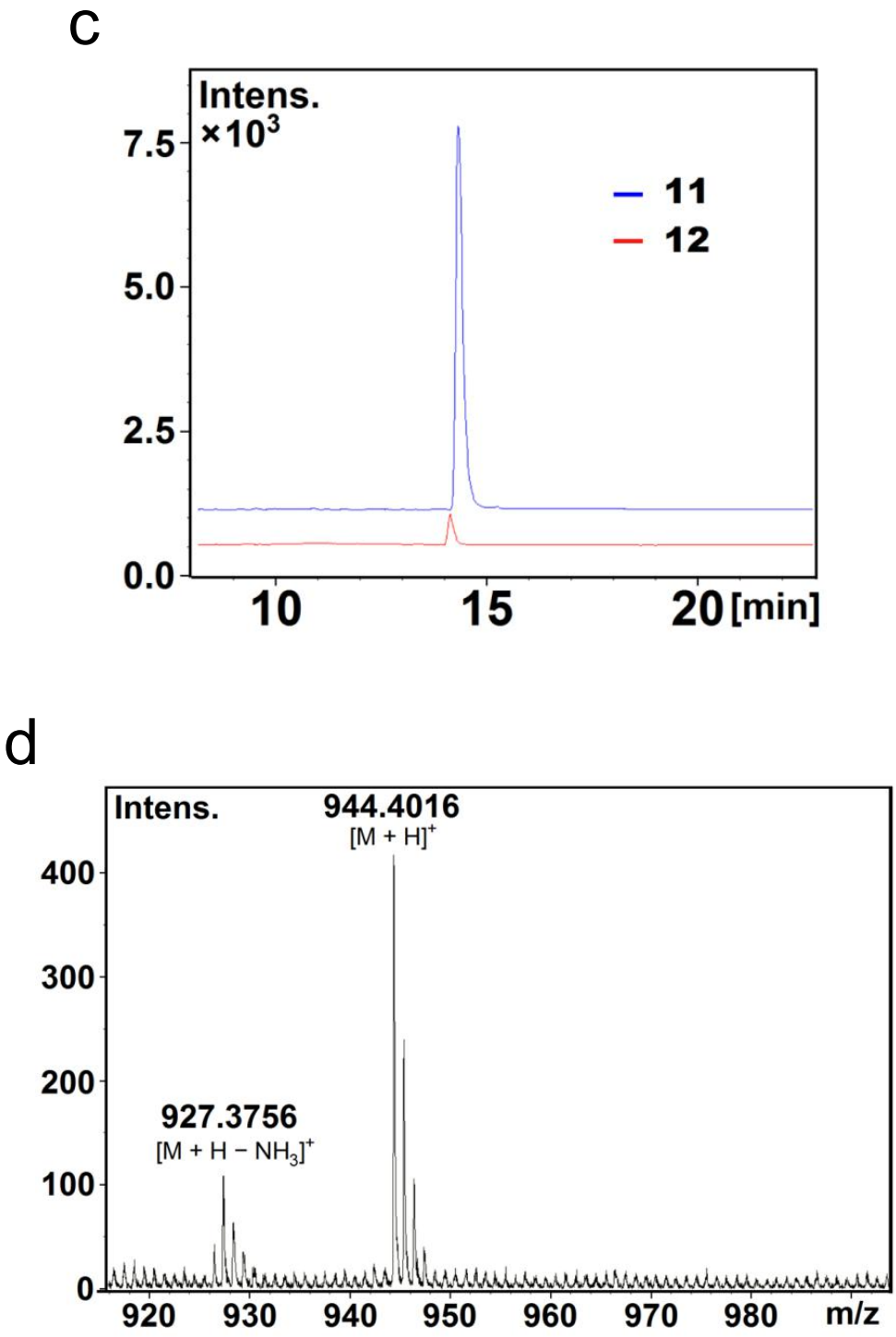
a



12 precolibactin-943

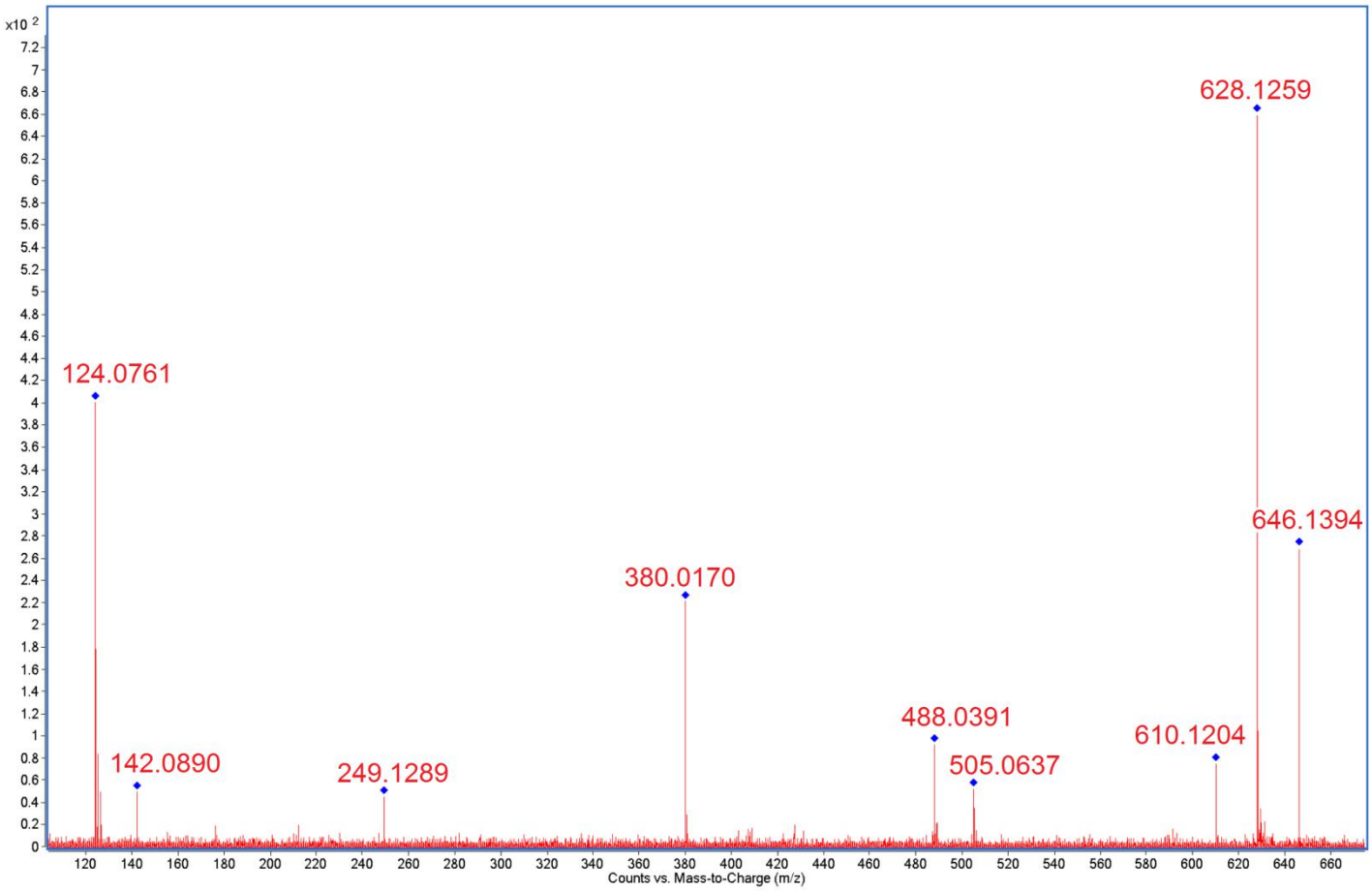
b



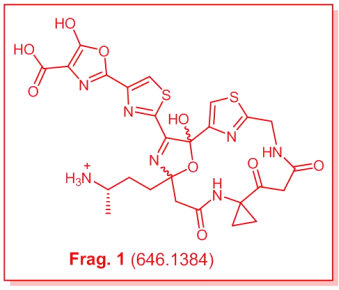


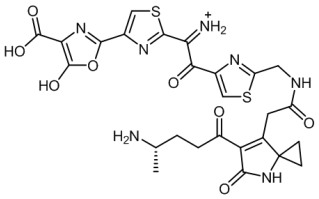
Supplementary Figure 13. The detection of precolibactin-943 (12). a, Proposed structure of **12**. **b**, Proposed biosynthesis of **12**. Compared to the biosynthesis of precolibactin-969 (**11**) (Fig. 1), the assembly of **12** shows a different biosynthetic logic for the utilization of the aminomalonate extender unit. ClbO is shown to incorporate the aminomalonate through a decarboxylative Claisen condensation in forming **12**. **c**, The metabolic profile of *clb*$^+$ heterologous expression strain *E. coli* DH10B/pCAP01-*clb* Δ*clbP*Δ*clbQ*Δ*clbS* was investigated using UPLC–MS-based analysis, revealing the different retention times of **11** and **12**. EIC+ = 944.40 ± 0.01 and 970.38 ± 0.01, corresponding to metabolites **12** and **11**, respectively. **d**, The high-resolution ESI-MS of **12**. Compared to **11**, the $[M + H - NH_3]^+$ MS fragmentation of **12** is observed.

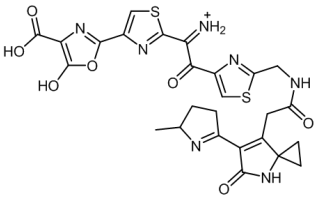
a

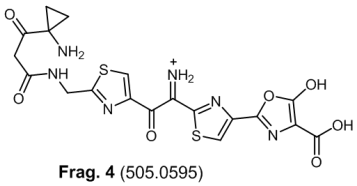

x10 2
628.1259
124.0761
646.1394
380.0170
488.0391
610.1204
142.0890
249.1289
505.0637
Counts vs. Mass-to-Charge (m/z)

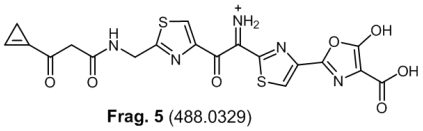
b

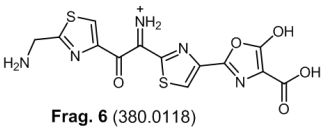

Frag. 1 (646.1384)

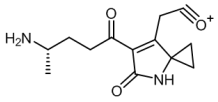

Frag. 2 (628.1279)

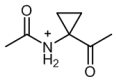

Frag. 3 (610.1173)


Frag. 4 (505.0595)


Frag. 5 (488.0329)


Frag. 6 (380.0118)

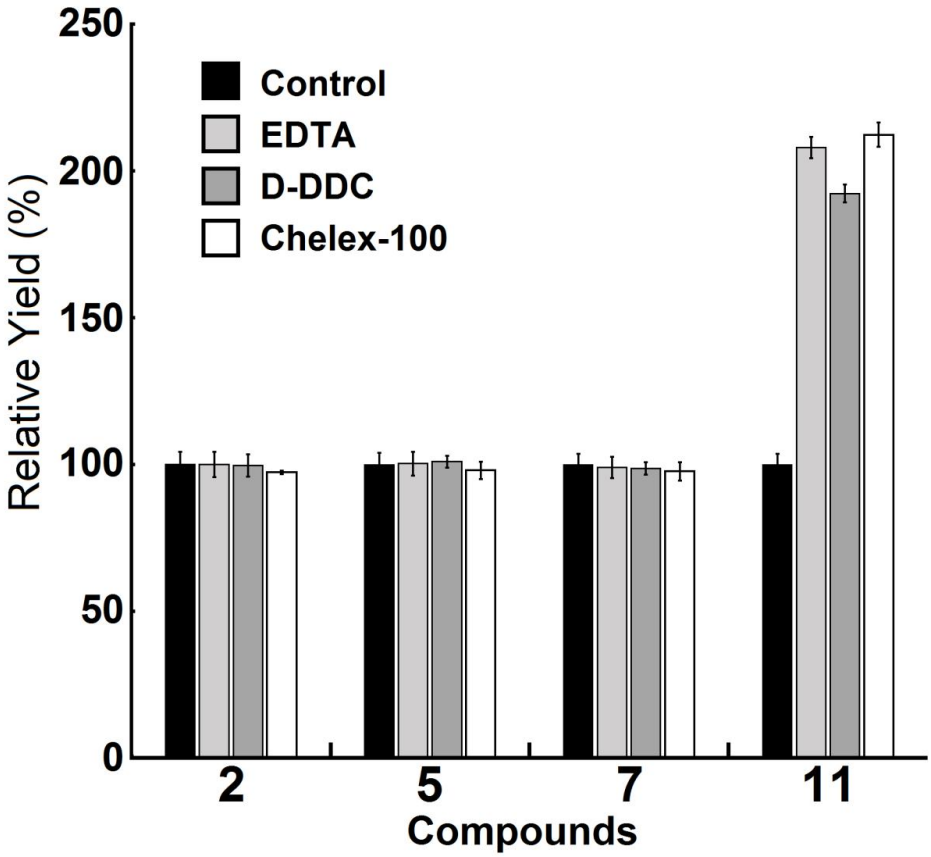

Frag. 7 (249.1234)


Frag. 8 (142.0863)

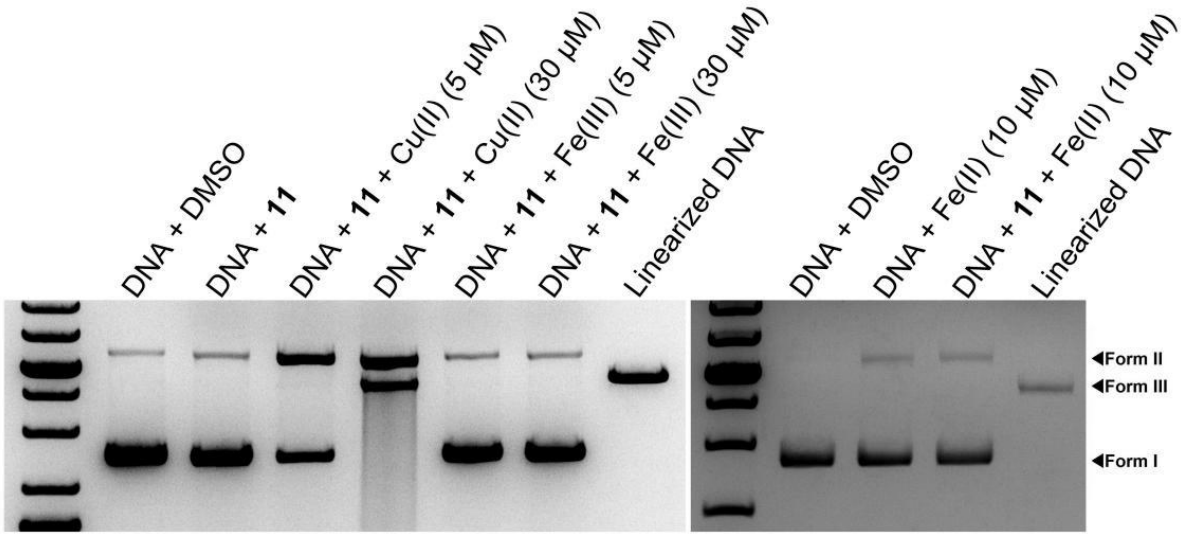

Frag. 9 (124.0757)

Fragmentation	Obs. mass	Calc. mass	Error [Da]
	colibactin-645 (**13**)		
1	646.1394	646.1384	0.0010
2	628.1259	628.1279	-0.0020
3	610.1204	610.1173	0.0031
4	505.0637	505.0595	0.0042
5	488.0391	488.0329	0.0062
6	380.0170	380.0188	-0.0018
7	249.1289	249.1234	0.0055
8	142.0890	142.0863	0.0027
9	124.0761	124.0757	0.0004

Supplementary Figure 14. Structure analysis of colibactin-645 (**13**) by HR-MS/MS. **a**, HR-MS/MS fragmentation pattern of **13**. Fragmentation was acquired with collision energy of 15 V. **b**, The major fragmentation species from HR-MS/MS measurement of **13**. Obs. = observed; Calc. = calculated.



Supplementary Figure 15. Effect of various metal chelators, including ethylenediaminetetraacetic acid (EDTA), diethylammonium diethyldithiocarbamate (D-DDC), and Chelex-100, on the production of different precolibactins with diverse carbon skeletons from *clb*$^{+}$ Δ*clbP* mutant. Precolibactin-441 (**2**), precolibactin-546 (**5**), precolibactin-712 (**7**), and precolibactin-969 (**11**) were selected and evaluated, which represent linear, aza-spirocyclopropane, thiazole-containing, and aminomalonate-containing structural derivatives, respectively. Data are shown as mean ± s.d. (n = 3).



Supplementary Figure 16. Analysis of the effect of precolibactin-969 (**11**) on plasmid pBR322 DNA in the presence of Cu(II), Fe(III) or Fe(II). Reactions were performed at 15 μM **11** in the presence of Cu(II) (5 μM or 30 μM), Fe(III) (5 μM or 30 μM) or Fe(II) (10 μM), or in the absence of metal ion for 16 hours at 37 °C. DNA cleavage by **11** was observed only in the co-incubation of Cu(II) and **11**, but not Fe(III) or Fe(II), in which nicked and linearized DNA appeared while supercoiled DNA completely disappeared. The control (reactions without **11**) of Cu(II) or Fe(III) at 30 μM showed DNA cleavage similar to the negative control presented in the figure (the lane with DNA and DMSO only). Top band, nicked DNA (Form II); middle band, linearized DNA (Form III); bottom band, supercoiled DNA (Form I). *Eco*RI-linearized pBR322 DNA is shown as the standard.

a

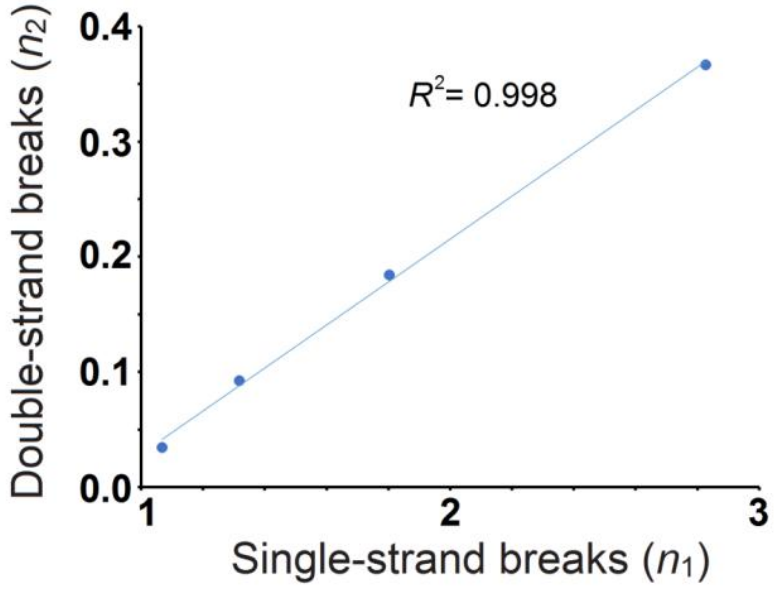
	Form I	Form II	Form III
Band 3 (2 h)	33.2 %	63.5 %	3.3 %
Band 4 (4 h)	24.4 %	67.2 %	8.4 %
Band 5 (6 h)	13.7 %	71.0 %	15.3 %
Band 6 (8 h)	4.1 %	70.5 %	25.4 %

	n_1	n_2
Band 3 (2 h)	1.0685	0.0341
Band 4 (4 h)	1.3185	0.0921
Band 5 (6 h)	1.8039	0.1839
Band 6 (8 h)	2.8278	0.3664

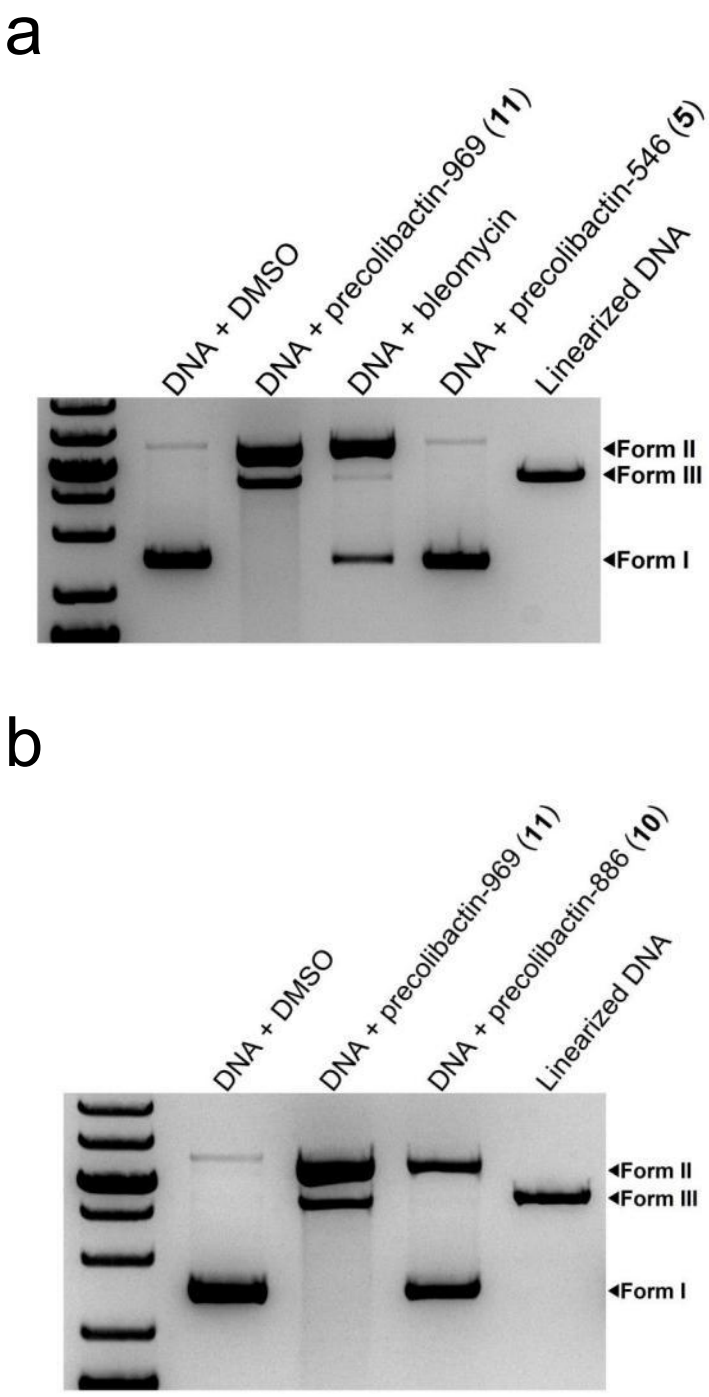
$$f_{III}=n_2e^{-n_2}$$

$$f_I=e^{-(n_1+n_2)}$$

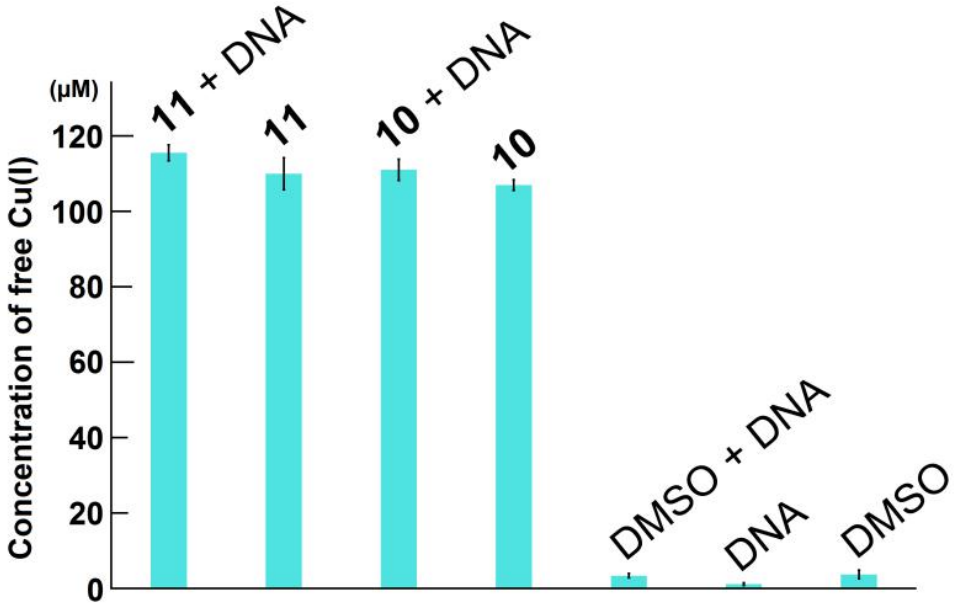
b



Supplementary Figure 17. Freifelder–Trumbo analysis of precolibactin-969 (11)-induced DNA linearization. a, Quantification of Form I, Form II, and Form III plasmid DNA; and the equations for calculating the number of single-strand breaks (SSBs) (n_1) and the number of double-strand breaks (DSBs) (n_2) per plasmid molecule. f_I and f_{III} represent the fractions of Form I and Form III DNA, respectively, after treatment. **b**, Ratio of DNA SSBs (n_1) to DSBs (n_2) per DNA molecule in a plasmid cleavage assay using **11**. Data points (left to right) correspond to reactions with incubation times of 2, 4, 6 and 8 hours. The Freifelder–Trumbo relationship postulates that if the ratio of n_1 to n_2 is approximately 120, the DSBs arise from the accumulation of SSBs randomly distributed over the plasmid molecule[23]. The value was constant for **11** (5.35:1), and was much lower than 120, indicating the **11**-induced DNA linearization was formed through DSBs rather than SSBs.

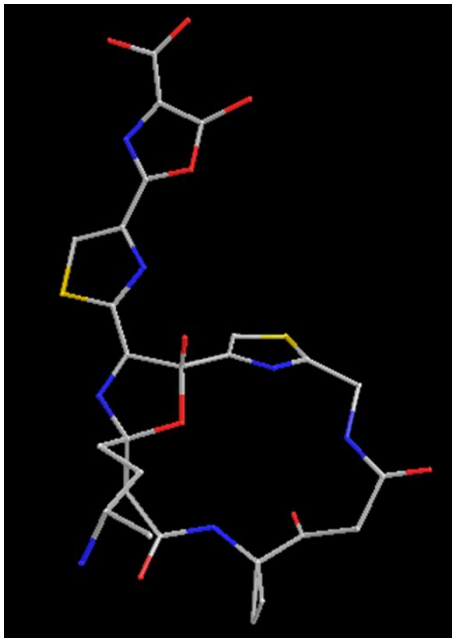


Supplementary Figure 18. a, The DNA double-strand breaks (DSBs) activities of precolibactin-969 (**11**), bleomycin and precolibactin-546 (**5**) were evaluated. Reactions were performed with either **11** (15 μM), bleomycin (15 μM) or **5** (5 mM) in the presence of both Cu(II) (30 μM) and β-mercaptoethanol (5 mM) for 12 hours at 37 °C. [We tested the DNA DSBs ability of **5** under a variety of assay conditions, including higher Cu(II) concentrations and higher reductant concentrations. **5** did not display any DNA DSBs ability *in vitro* in our assays using plasmid pBR322 DNA (data not shown).] **b**, The DNA DSBs activities of **11** and precolibactin-886 (**10**) were compared. Reactions were performed at the same concentration of **11** (15 μM) or **10** (15 μM) in the presence of Cu(II) (30 μM) for 12 hours at 37 °C. **a**, **b**, All of the controls (reactions without compound) of each reagent or metal ion show no DNA cleavage similar to the negative control presented in the figure (the lane with DNA and DMSO only). Top band, nicked DNA (Form II); middle band, linearized DNA (Form III); bottom band, supercoiled DNA (Form I). *Eco*RI-linearized pBR322 DNA is shown as the standard.

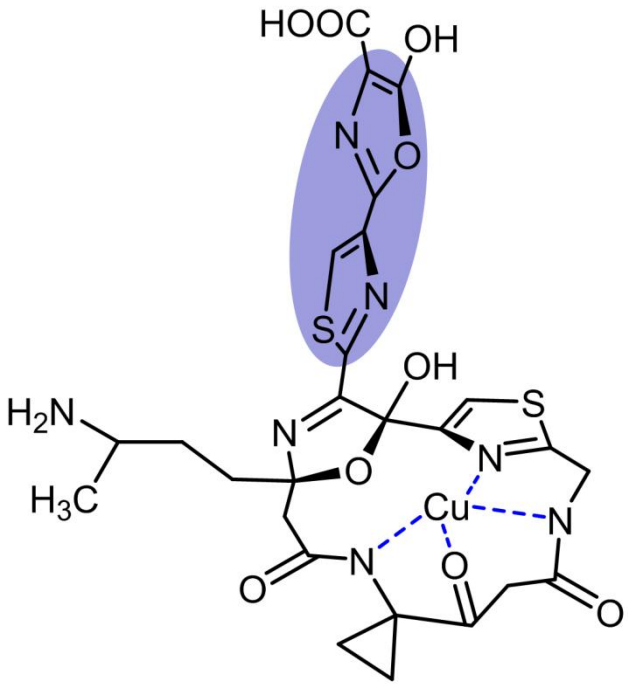


Supplementary Figure 19. Reduction of Cu(II) by macrocyclic colibactins. The formation of Cu(I) indicates that both precolibactin-886 (**10**) and precolibactin-969 (**11**) have the ability to reduce Cu(II) to Cu(I), with and without plasmid pBR322 DNA. Reactions were performed at 150 μM **11** or **10** in the presence of 300 μM cupric chloride and 1 mM neocuproine, with or without adding plasmid pBR322 DNA (500 ng). In control, only DMSO solvent was added. Following incubation for 16 hours at 37 °C in a darkroom, the amount of Cu(I) was determined by measuring the visible absorbance at 450 nm of the cuprous-neocuproine complex using a NanoDrop. Data are shown as mean ± s.d. (n = 3).

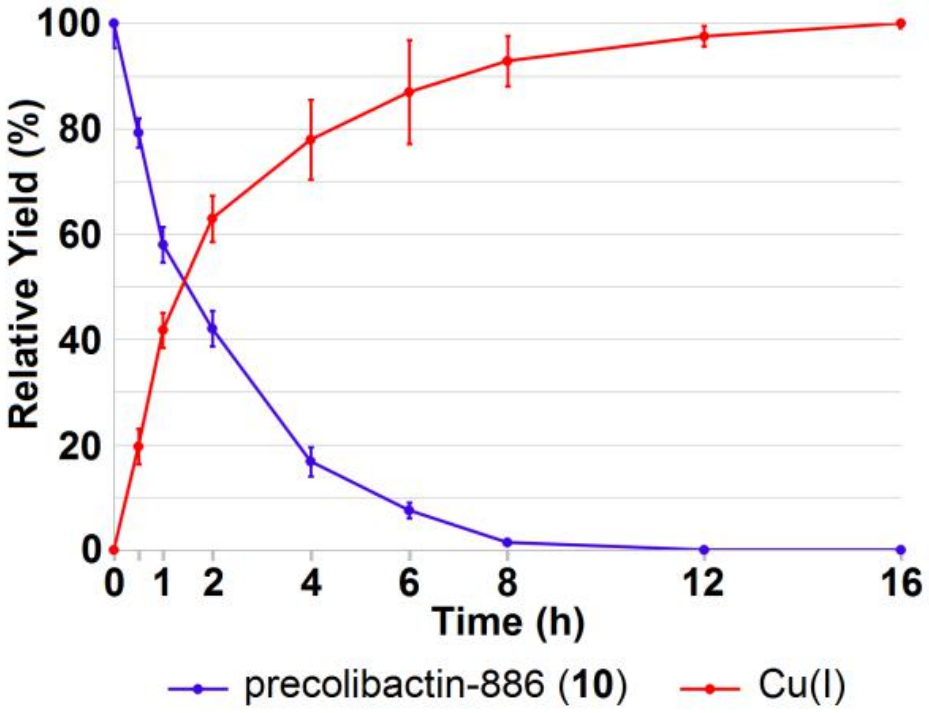
a



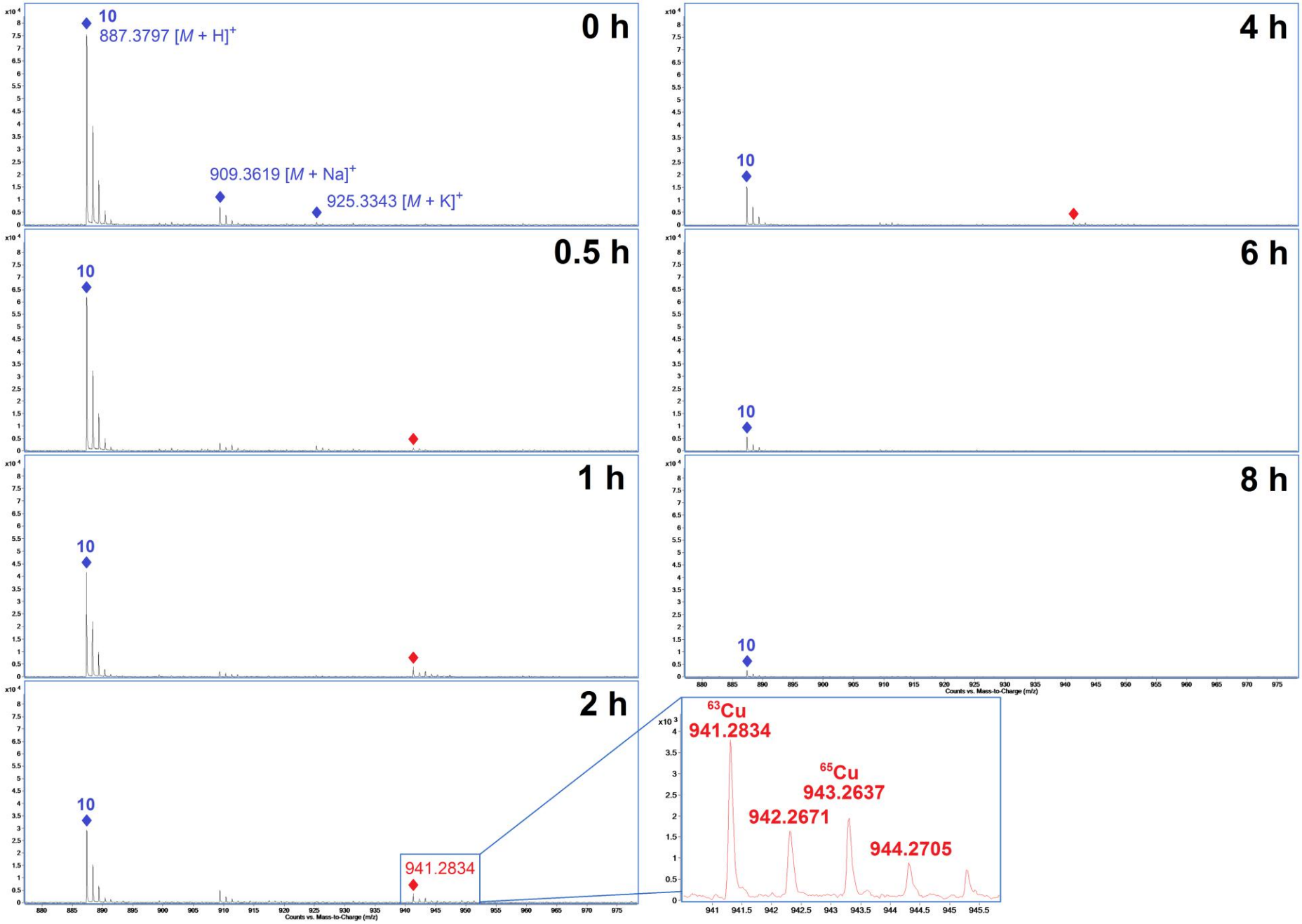
b



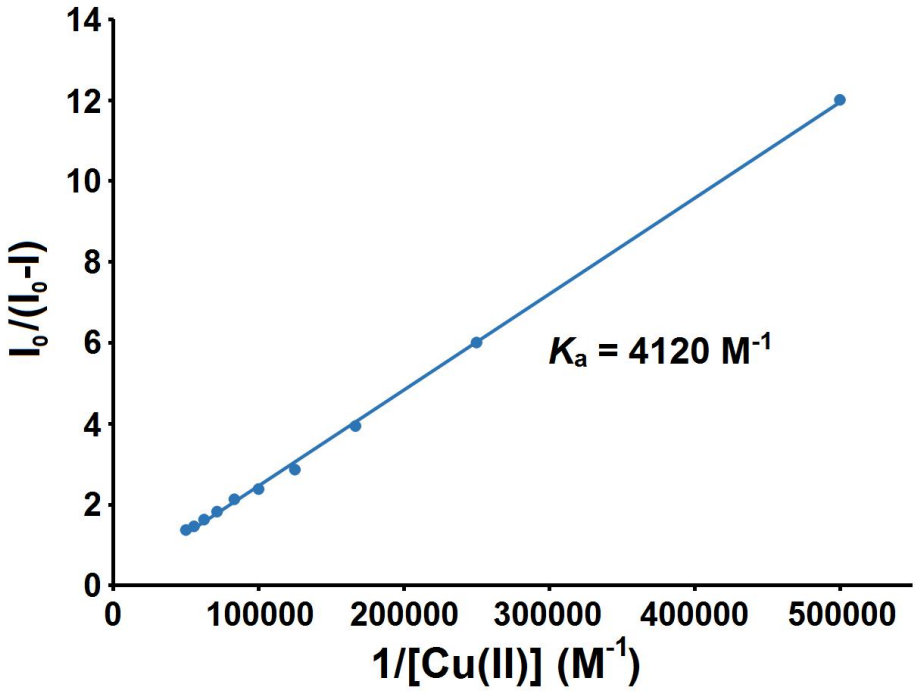
Supplementary Figure 20. a, Backbone structure of the lowest energy conformer of colibatin-645 (**13**) calculated by molecular modeling. **b**, Proposed structure of colibatin·Cu(II). The potential DNA-binding moiety in **13**, the thiazole/5-hydroxy oxazole moiety, is highlighted.



Supplementary Figure 21. Degradation of precolibactin-886 (**10**) along with the formation of Cu(I) after mixing **10** and Cu(II) over time. The degradation of **10** was monitored by the loss of its mass signal in HRMS analysis. The formation of Cu(I) was measured by adding neocuproine as mentioned in **Supplementary Figure 19**. Data are shown as mean ± s.d. (n = 3).

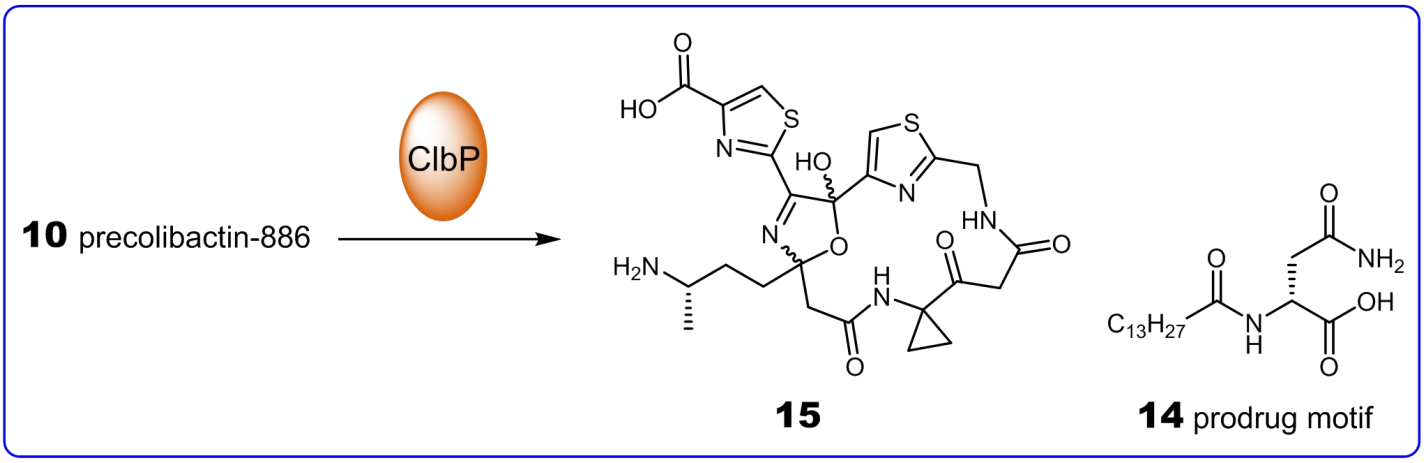

0 h
10
887.3797 [M + H]+
909.3619 [M + Na]+
925.3343 [M + K]+
0.5 h
1 h
2 h
941.2834
4 h
6 h
8 h
Counts vs. Mass-to-Charge (m/z)
63Cu
941.2834
65Cu
943.2637
942.2671
944.2705

Supplementary Figure 22. HR-ESI-MS analyses of the colibactin·copper complex. In this assay, the precolibactin-886 (**10**) which contains the putative active center for Cu(II) binding and reduction was selected as a representative for the mature colibactin, because this precursor was more readily available. After incubation with Cu(II), the mass signal of **10** disappeared over time, and a putative copper-bound colibactin species was tracked with mass signal consistent with the isotopic distribution of a copper-bound complex. The mass signal of the copper-bound colibactin species was weak and transient, suggesting the instability of this species.

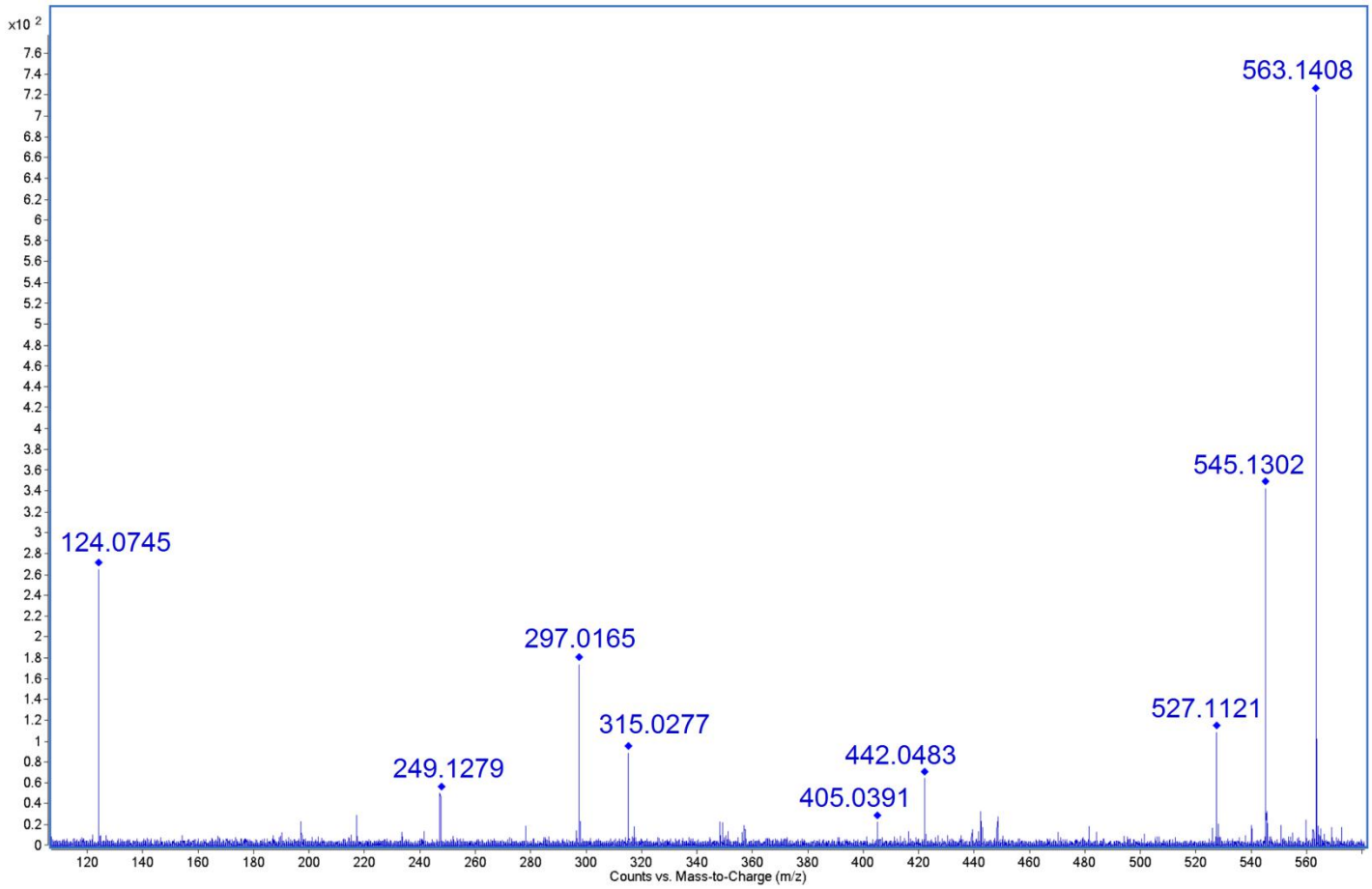


Supplementary Figure 23. Double-reciprocal (Benesi–Hildebrand) plot derived from a titration experiment of precolibactin-886 (**10**) with Cu(II).

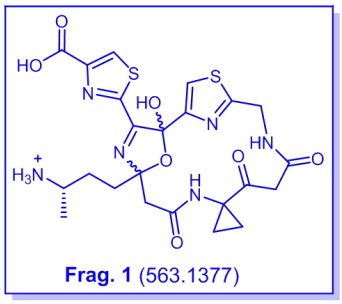
a


ClbP
10 precolibactin-886
15
14 prodrug motif

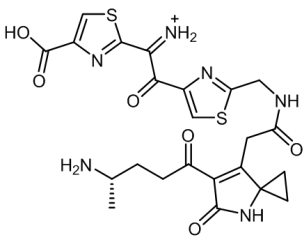
b


x10 2
563.1408
545.1302
124.0745
297.0165
315.0277
527.1121
249.1279
442.0483
405.0391
Counts vs. Mass-to-Charge (m/z)

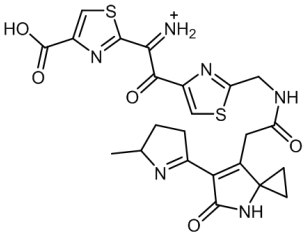
C



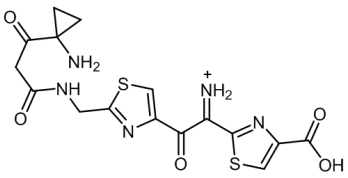
Frag. 1 (563.1377)



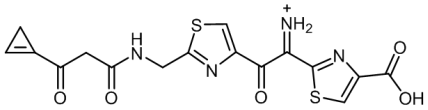
Frag. 2 (545.1272)



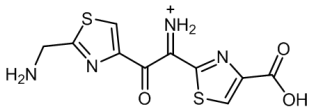
Frag. 3 (527.1166)



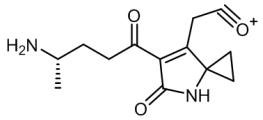
Frag. 4 (422.0587)



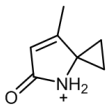
Frag. 5 (405.0322)



Frag. 6 (297.0111)



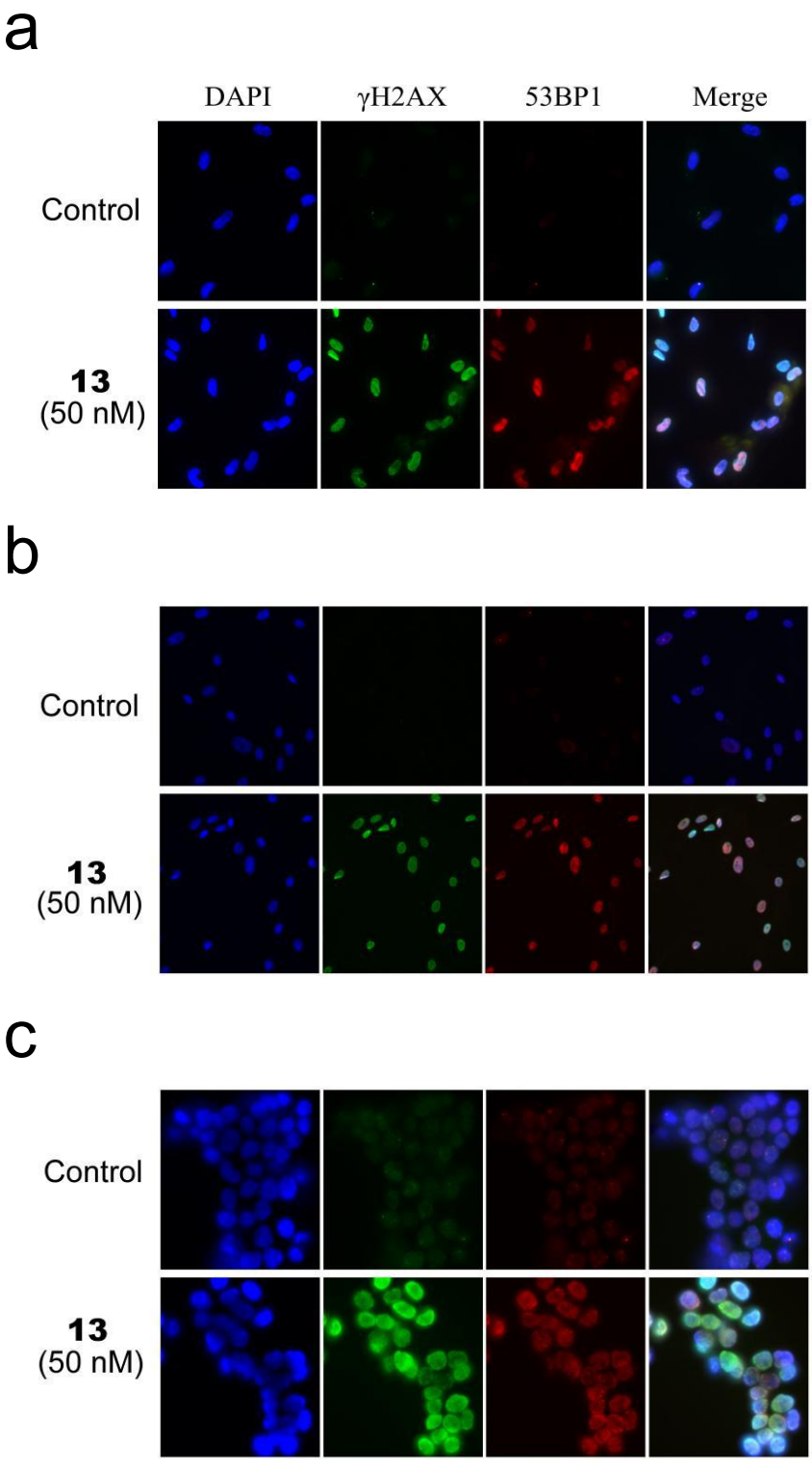
Frag. 7 (249.1234)



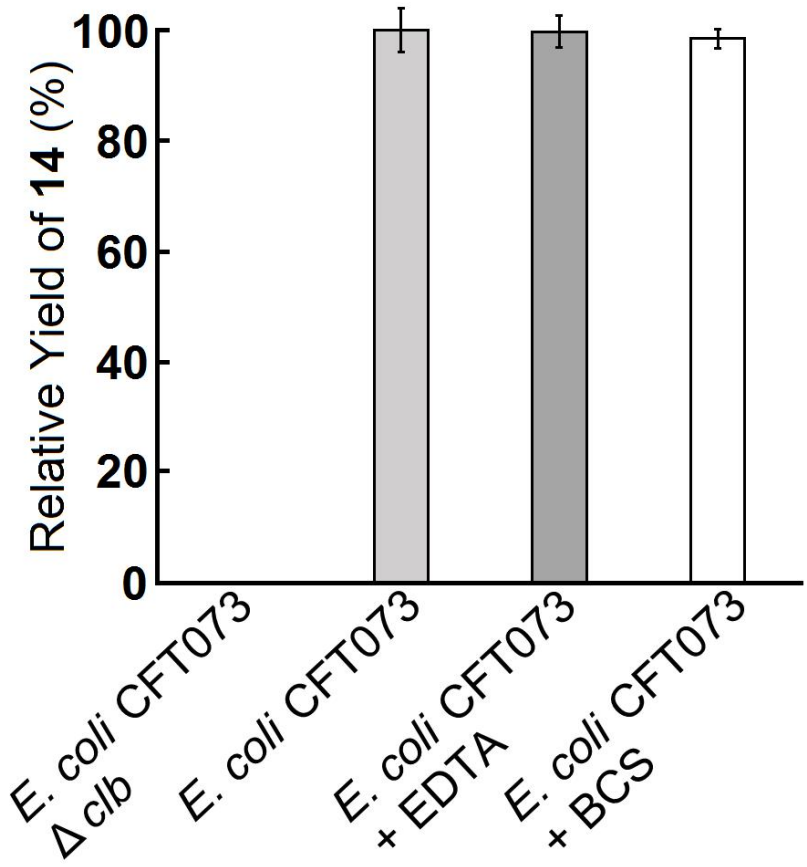
Frag. 8 (124.0757)

Fragmentation	Obs. mass	Calc. mass	Error [Da]
	15		
1	563.1408	563.1377	0.0031
2	545.1302	545.1272	0.0030
3	527.1121	527.1166	-0.0045
4	422.0483	422.0587	-0.0104
5	405.0391	405.0322	0.0069
6	297.0165	297.0111	0.0054
7	249.1279	249.1234	0.0045
8	124.0745	124.0750	-0.0005

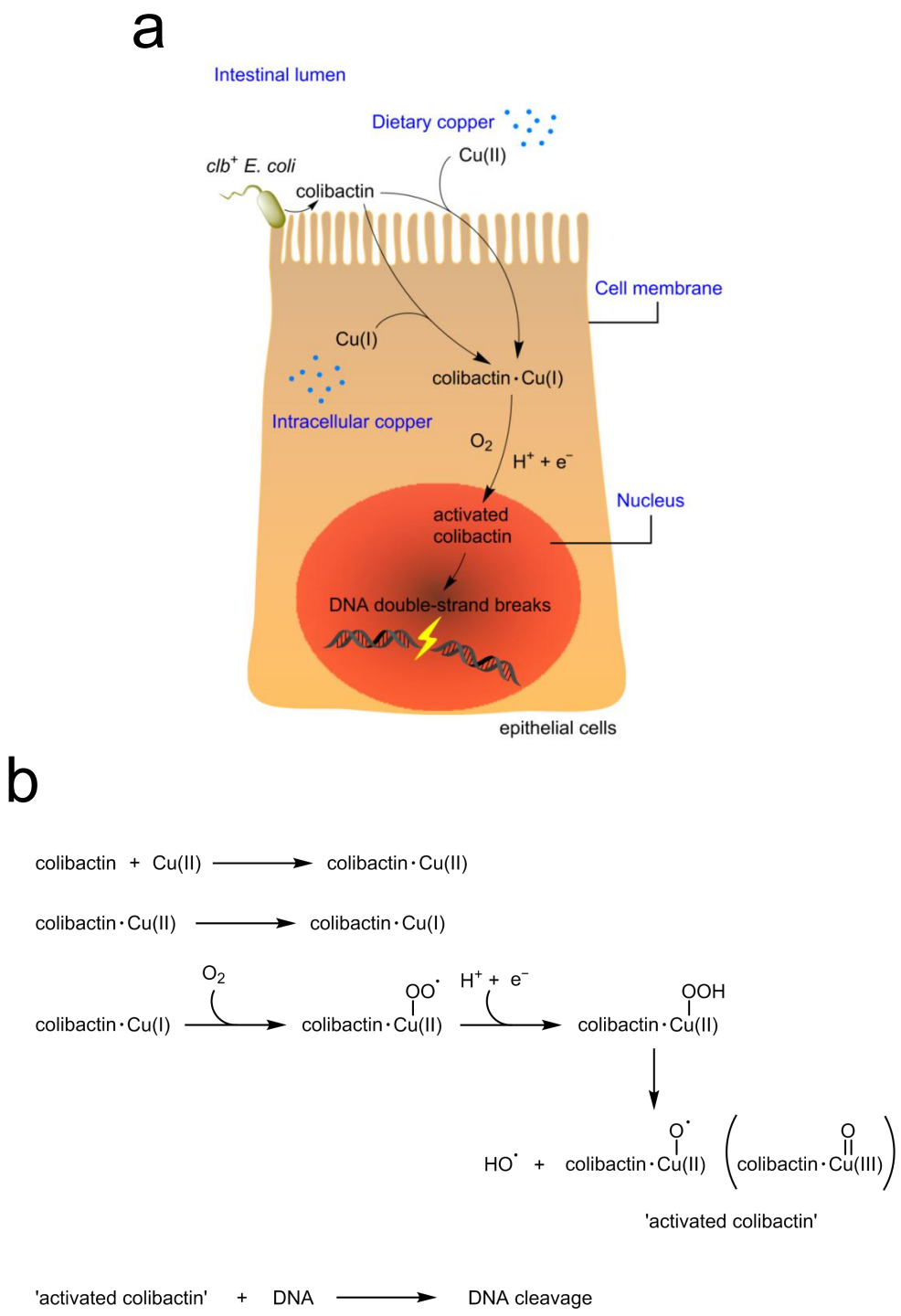
Supplementary Figure 24. Structure analysis of **15** by HR-MS/MS. **a**, Scheme of maturation of **15** from the hydrolysis of its precursor precolibactin-886 (**10**) by expressed peptidase ClbP, with the release of a prodrug motif *N*-myristoyl-D-asparagine (**14**) simultaneously. **b**, HR-MS/MS fragmentation pattern of **15**. Fragmentation was acquired with collision energy of 15 V. **c**, The major fragmentation species from HR-MS/MS measurement of **15**. Obs. = observed; Calc. = calculated.



Supplementary Figure 25. Analysis of colibactin-induced DNA damage in cell cultures. Immunofluorescence imaging of γH2AX and 53BP1 foci in human normal colon epithelial FHC cells (**a**), human normal colon fibroblast CCD-112 CoN cells (**b**), and colorectal cancer HCT-116 cells (**c**) that are treated with **13** (50 nM), showing these foci are induced and colocalize in these cells. In control, only DMSO solvent was added. Columns from left to right, nucleus (blue), γH2AX (green), 53BP1 (red), and merge.



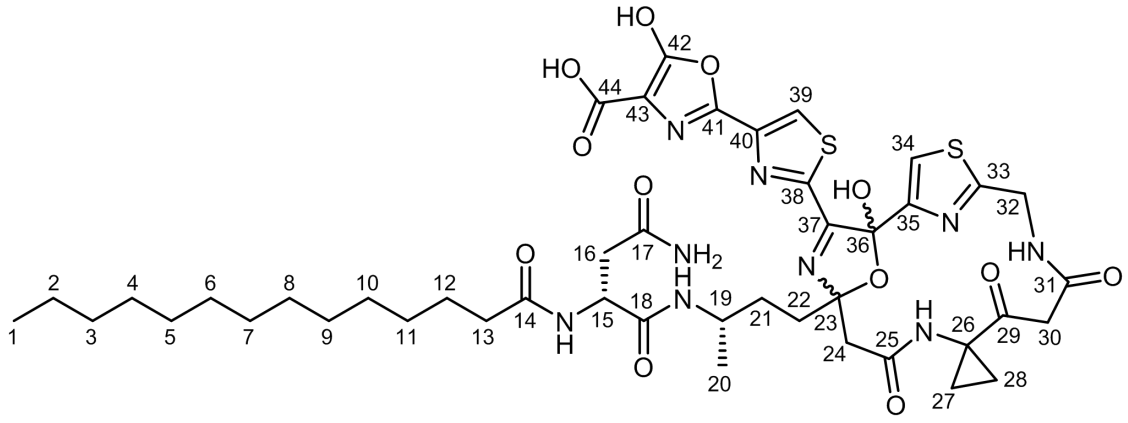
Supplementary Figure 26. Production of the prodrug motif **14** of *clb* pathway was selected as a proxy for flux through the biosynthetic pathway to evaluate the effects of different metal chelators on metabolite production during a transient infection process. The results showed that both ethylenediaminetetraacetic acid (EDTA, 2.5 mM) and bathocuproinedisulfonic acid (BCS, 2 mM) could not suppress the production of *clb* pathway-related metabolites in the transient infection process.



Supplementary Figure 27. Proposed mechanism for the generation of 'activated colibactin' *in vivo*. a, After being secreted from a producing bacterium that localizes close to or in contact with the intestinal brush border[24], colibactin (colibactin-645, **13**) binds to exchangeable copper in the intestinal lumen, likely coming from diet[25], to form a colibactin·Cu(II) complex. This complex is quickly transported into the epithelial cell while reduced to a colibactin·Cu(I) complex, and the coordination of O_2 to this cuprous complex in cells generates 'activated colibactin' that attacks DNA and initiates DNA cleavage. Additionally, this mechanism dose not exclude the possibility that colibactin quickly enters the epithelial cell and then binds the intracellular copper to exert its activity. **b**, Proposed reactions in colibactin activation. Cu(II)─O$^{\bullet}$ (or Cu(III)═O) is proposed to be the active species in the 'activated colibactin' complex susceptible of DNA carbon–hydrogen bond activation[26], which is consistent with the observed inhibitory effects of H_2O_2 scavengers on the DNA cleavage reaction *in vitro* as colibactin·Cu(II)─OOH is a key intermediate to colibactin·Cu(II)─O$^{\bullet}$.

4. Supplementary Tables

Supplementary Table 1. Structure characterization data of precolibactin-969 (**11**).



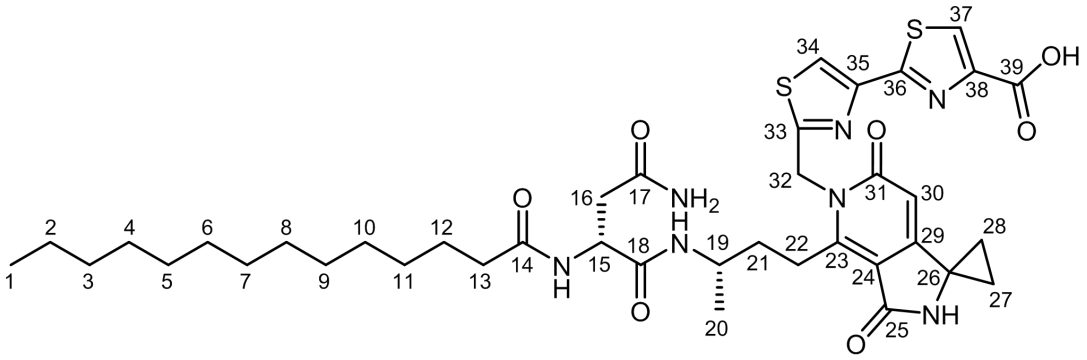
NMR spectroscopic data for **11** in DMSO-*d*6

C	δ_C*	δ_H	COSY	TOCSY	NOESY	HMBC (^{1}H to ^{13}C)
1	13.7 (CH_3)	0.85	2	2		2, 3
2	22.3 (CH_2)	1.23	1	1		1, 3
3	31.4 (CH_2)	1.19-1.23				
4-10	28.7 (CH_2), 28.8 (CH_2)	1.19-1.23				
11	28.8 (CH_2)	1.19-1.23				
12	24.9 (CH_2)	1.43	13			11
13	35.6 (CH_2)	2.08	12		14NH	14
14	172.5 (C)	-				
		NH, 7.93	15	15, 16	13, 15, 20	
15	50.3 (CH)	4.46	14NH, 16	14NH, 16	14NH, 16, 18NH	16, 18
16	38.4 (CH_2)	a 2.30	15, 16b	14NH	15	15, 17
		b 2.41	15, 16a	14NH	15	15, 17
17	171.2 (C)	-				
		a NH, 6.82	17bNH	17bNH	17bNH	16
		b NH, 7.27	17aNH	17aNH	17aNH	
18	170.4 (C)	-				
		NH, 7.52	19	19, 20, 21	15, 19, 20, 21	
19	48.1 (CH)	3.70	18NH, 20	18NH	18NH	
20	20.7 (CH_3)	0.98	19	18NH	14NH, 18NH	19, 21
21	30.4 (CH_2)	1.42	22	18NH	18NH	
22	34.6 (CH_2)	1.78	21			23
23	107.9 (C)	-				
24	45.7 (CH_2)	a 2.60	24b		24b	25
		b 3.29	24a		24a, 25NH	
25	170.6 (C)	-				
		NH, 7.94			24b	
26	37.3 (C)	-				
27	21.7 (CH_2)	1.23	28			28
28	21.7 (CH_2)	1.23	27			27
29	205.5 (C)	-				

30	48.1 (CH_2)	a 3.08	30b	30b	30b, 31	26, 29, 31
		b 3.73	30a	30a	30a, 31	
31	167.0 (C)	-				
		NH, 9.03	32b	32a, 32b	30a, 30b, 32a, 32b	
32	40.8 (CH_2)	a 4.27	32b	31, 32b	31, 32b	
		b 4.79	31NH, 32a	31, 32a	31, 32a	33
33	167.8 (C)	-				
34	119.8 (CH)	7.97				33, 35
35	154.3 (C)	-				
36	– (C)	-				
37	– (C)	-				
38	– (C)	-				
39	120.5 (CH)	9.07				
40	– (C)	-				
41	– (C)	-				
42	– (C)	-				
43	– (C)	-				
44	– (C)	-				

* Determined from $^{1}H–^{13}C$ HSQC and $^{1}H–^{13}C$ HMBC NMR data

Supplementary Table 2. Structure characterization data of precolibactin-795b (**9**).

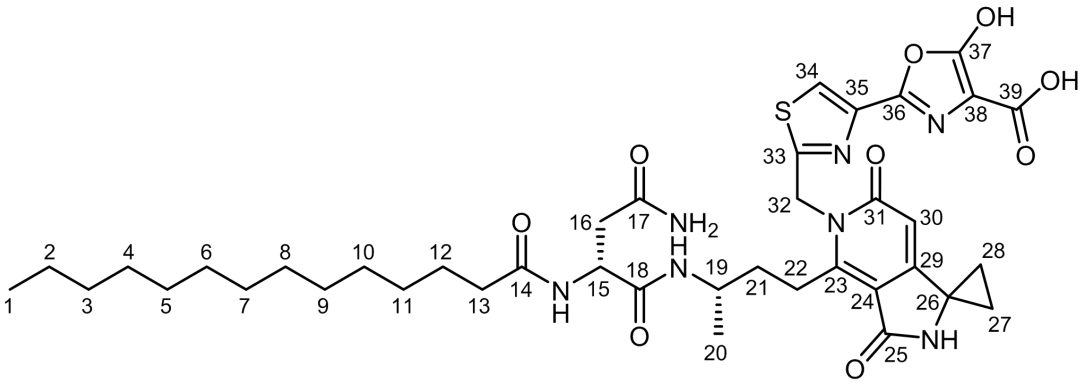


NMR spectroscopic data for **9** in DMSO-*d*$_6$

C	δ_C	δ_H (*J* in Hz)	COSY	HMBC (^{1}H to ^{13}C)
1	14.1	0.82 t (6.8)	2	2, 3
2	22.2	1.21 m	1	1, 3
3	31.4	1.19 m		
4-10	28.7, 29.0, 29.1, 29.1, 29.1, 29.1, 29.2	1.13-1.19 m		
11	28.8	1.13-1.19 m		
12	25.3	1.39 m	13	
13	35.3	2.04 m	12	11, 12, 14
14	172.3	-		
		NH, 7.91 d (6.0)	15	14, 15, 16
15	50.1	4.50 m	14NH, 16a, 16b	16, 18
16	37.5	a 2.38 m	15, 16b	15, 17
		b 2.48 m	15, 16a	
17	171.7	-		
		a NH, 6.82 s	17bNH	17
		b NH, 7.28 s	17aNH	17
18	170.6	-		
		NH, 7.70 d (6.8)	19	18, 19, 20, 21
19	44.8	3.87 m	18NH, 20, 21	18, 20, 21, 22
20	20.4	1.07 d (5.1)	19	19, 21
21	35.5	1.72 m	19, 22a, 22b	19, 20, 22, 23
22	24.2	a 3.35 m	21, 22b	
		b 3.42 m	21, 22a	
23	153.2	-		
24	109.8	-		
25	166.8	-		
		NH, 8.50 s		23, 24, 25, 26, 29
26	40.5	-		
27	15.3	a 1.35 m	27b, 28a, 28b	26, 28, 29
		b 1.39 m	27a, 28a, 28b	26, 28, 29
28	15.3	a 1.35 m	27a, 27b, 28b	26, 27, 29
		b 1.39 m	27a, 27b, 28a	26, 27, 29
29	160.0	-		
30	103.4	6.16 s		23, 24, 26, 29, 31
31	161.9	-		

32	44.5	a 5.55 d (6.0)	32b	23, 31, 33
		b 5.59 d (6.0)	32a	23, 31, 33
33	166.7	-		
34	119.1	8.23 s		33, 35
35	147.2	-		
36	161.7	-		
37	127.9	8.33 s		36, 38, 39
38	150.2	-		
39	162.7	-		

Supplementary Table 3. Structure characterization data of precolibactin-795a (**8**).



NMR spectroscopic data for **8** in DMSO-*d6*

C	δ_C	δ_H (*J* in Hz)	COSY	HMBC (^{1}H to ^{13}C)
1	14.1	0.85 t (6.9)	2	2, 3
2	22.2	1.24 m	1	1, 3
3	31.4	1.19-1.25 m		
4-10	28.8, 28.8, 29.0, 29.1, 29.2, 29.2, 29.2	1.19-1.25 m		
11	28.7	1.19-1.25 m		
12	25.3	1.46 m	13	11, 13, 14
13	35.4	2.08 m	12	11, 12, 14
14	172.4	-		
		NH, 8.18 d (7.6)	15	14, 15
15	50.6	4.49 m	14NH, 16, 18NH	16, 18
16	37.4	2.33 m	15	15, 17
		2.53 m	15	15, 17
17	171.6	-		
		a NH, 6.78 s	17bNH	16, 17
		b NH, 7.37 s	17aNH	17
18	170.7	-		
		NH, 7.87 d (7.6)	15, 19	18, 19
19	44.6	3.83 m	18NH, 20, 21a, 21b	18, 20, 21, 22
20	20.3	1.02 d (6.0)	19	19, 21
21	35.4	a 1.51 m	19, 21b, 22	19, 22
		b 1.68 m	19, 21a, 22	19, 22
22	24.2	3.35 m	21a, 21b	23
23	153.1	-		
24	109.7	-		
25	166.8	-		
		NH, 8.45 s		23, 24, 25, 26, 29
26	39.7	-		
27	15.3	a 1.34 m	27b, 28a, 28b	26, 28, 29
		b 1.37 m	27a, 28a, 28b	26, 28, 29
28	15.3	a 1.34 m	27a, 27b, 28b	26, 27, 29
		b 1.37 m	27a, 27b, 28a	26, 27, 29
29	159.8	-		
30	103.3	6.15 s		23, 24, 26, 31

31	161.9	-		
32	44.4	a 5.51 d (5.1)	32b	23, 31, 33
		b 5.59 d (5.1)	32a	23, 31, 33
33	164.1	-		
34	120.0	8.76 s		33, 35, 36
35	142.9	-		
36	164.0	-		
37	133.0	-		
38	123.9	-		
39	160.9	-		

Supplementary Table 4. Plasmids and strains used in this study.

	Description	Source
Strains		
E. coli CFT073	Wild type *E. coli* strain harboring the colibactin (*clb*) gene cluster	2, 27
E. coli Top10	Host strain for routine cloning	Invitrogen
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*	1
E. coli BW25113/pCAP01-*clb*/pIJ790	*E. coli* BW25113 harboring pCAP01-*clb* and pIJ790, for the further genetic manipulation of pCAP01-*clb*	2
E. coli BW25113/pCAP01-*clb* Δ*clbP*Δ*clbQ*Δ*clbS*/pIJ790	*E. coli* BW25113 harboring pCAP01-*clb* Δ*clbP*Δ*clbQ*Δ*clbS* and pIJ790, for the further genetic manipulation of pCAP01-*clb* Δ*clbP*Δ*clbQ*Δ*clbS*	This study
Plasmids		
pETDuet-1	Protein expression vector, pBR322 ori, T7 promoter, *amp^R*	Novagen
pIJ790	λ -Red (*gam*, *bet*, *exo*), *cat*, *araC*, *rep101*ts	3
pIJ773	Source of *apra^R*	3
pDR111	Source of *amp^R*	28
Protein expression plasmids constructed		
pETDuet-1-ClbS	pETDuet-1, the *clb* pathway specific resistance protein ClbS for *in vivo* complementation	This study
pETDuet-1-ClbR	pETDuet-1, the *clb* pathway transcriptional regulator ClbR for overexpression	This study
pET-28a-ClbP	pET28a(+), the *clb* pathway peptidase ClbP for expression	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*Δ*clbQ*	pCAP01-*clb* derivative (Δ*clbP* & Δ*clbQ*): *apra^R*	5
pCAP01-*clb* Δ*clbP*Δ*clbQ*Δ*clbS*	pCAP01-*clb* derivative (Δ*clbP* & Δ*clbQ* & Δ*clbS*): *apra^R*	This study
pCAP01-*clb* Δ*clbP*Δ*clbQ*Δ*clbS*Δ*clbB*	pCAP01-*clb* derivative (Δ*clbP* & Δ*clbQ* & Δ*clbS*): *apra^R*; (Δ*clbB*): *amp^R*	This study
pCAP01-*clb* Δ*clbP*Δ*clbQ*Δ*clbS*Δ*clbC*	pCAP01-*clb* derivative (Δ*clbP* & Δ*clbQ* & Δ*clbS*): *apra^R*; (Δ*clbC*): *amp^R*	This study
pCAP01-*clb* Δ*clbP*Δ*clbQ*Δ*clbS*Δ*clbDEF*	pCAP01-*clb* derivative (Δ*clbP* & Δ*clbQ* & Δ*clbS*): *apra^R*; (Δ*clbDEF*): *amp^R*	This study
pCAP01-*clb* Δ*clbP*Δ*clbQ*Δ*clbS*Δ*c*	pCAP01-*clb* derivative (Δ*clbP* & Δ*clbQ* & Δ*clbS*): *apra^R*; (Δ*clbG*): *amp^R*	This study

lbG		
pCAP01-*clb* Δ*clbP*Δ*clbQ*Δ*clbS*Δ*clbH*	pCAP01-*clb* derivative (Δ*clbP* & Δ*clbQ* & Δ*clbS*): *apra*^{R}; (Δ*clbH*): *amp*^{R}	This study
pCAP01-*clb* Δ*clbP*Δ*clbQ*Δ*clbS*Δ*clbI*	pCAP01-*clb* derivative (Δ*clbP* & Δ*clbQ* & Δ*clbS*): *apra*^{R}; (Δ*clbI*): *amp*^{R}	This study
pCAP01-*clb* Δ*clbP*Δ*clbQ*Δ*clbS*Δ*clbJ*	pCAP01-*clb* derivative (Δ*clbP* & Δ*clbQ* & Δ*clbS*): *apra*^{R}; (Δ*clbJ*): *amp*^{R}	This study
pCAP01-*clb* Δ*clbP*Δ*clbQ*Δ*clbS*Δ*clbK*	pCAP01-*clb* derivative (Δ*clbP* & Δ*clbQ* & Δ*clbS*): *apra*^{R}; (Δ*clbK*): *amp*^{R}	This study
pCAP01-*clb* Δ*clbP*Δ*clbQ*Δ*clbS*Δ*clbO*	pCAP01-*clb* derivative (Δ*clbP* & Δ*clbQ* & Δ*clbS*): *apra*^{R}; (Δ*clbO*): *amp*^{R}	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*Δ*clbQ*	*E. coli* DH10B harboring pCAP01-*clb* Δ*clbP*Δ*clbQ*	5
E. coli DH10B/pCAP01-*clb* Δ*clbP*Δ*clbQ*Δ*clbS*	*E. coli* DH10B harboring pCAP01-*clb* Δ*clbP*Δ*clbQ*Δ*clbS*	This study
E. coli DH10B/pCAP01-*clb* Δ*clbP*Δ*clbQ*Δ*clbS*Δ*clbB*	*E. coli* DH10B harboring pCAP01-*clb* Δ*clbP*Δ*clbQ*Δ*clbS*Δ*clbB*	This study
E. coli DH10B/pCAP01-*clb* Δ*clbP*Δ*clbQ*Δ*clbS*Δ*clbC*	*E. coli* DH10B harboring pCAP01-*clb* Δ*clbP*Δ*clbQ*Δ*clbS*Δ*clbC*	This study
E. coli DH10B/pCAP01-*clb* Δ*clbP*Δ*clbQ*Δ*clbS*Δ*clbDEF*	*E. coli* DH10B harboring pCAP01-*clb* Δ*clbP*Δ*clbQ*Δ*clbS*Δ*clbDEF*	This study
E. coli DH10B/pCAP01-*clb* Δ*clbP*Δ*clbQ*Δ*clbS*Δ*clbG*	*E. coli* DH10B harboring pCAP01-*clb* Δ*clbP*Δ*clbQ*Δ*clbS*Δ*clbG*	This study
E. coli DH10B/pCAP01-*clb* Δ*clbP*Δ*clbQ*Δ*clbS*Δ*clbH*	*E. coli* DH10B harboring pCAP01-*clb* Δ*clbP*Δ*clbQ*Δ*clbS*Δ*clbH*	This study
E. coli DH10B/pCAP01-*clb*	*E. coli* DH10B harboring pCAP01-*clb* Δ*clbP*Δ*clbQ*Δ*clbS*Δ*clbI*	This study

Δ*clbP*Δ*clbQ*Δ*clbS*Δ*clbI*		
E. coli DH10B/pCAP01-*clb* Δ*clbP*Δ*clbQ*Δ*clbS*Δ*clbJ*	*E. coli* DH10B harboring pCAP01-*clb* Δ*clbP*Δ*clbQ*Δ*clbS*Δ*clbJ*	This study
E. coli DH10B/pCAP01-*clb* Δ*clbP*Δ*clbQ*Δ*clbS*Δ*clbK*	*E. coli* DH10B harboring pCAP01-*clb* Δ*clbP*Δ*clbQ*Δ*clbS*Δ*clbK*	This study
E. coli DH10B/pCAP01-*clb* Δ*clbP*Δ*clbQ*Δ*clbS*Δ*clbO*	*E. coli* DH10B harboring pCAP01-*clb* Δ*clbP*Δ*clbQ*Δ*clbS*Δ*clbO*	This study
Protein expression strain for Δ*clbS* complementation assay		
E. coli DH10B/pCAP01-*clb* Δ*clbP*Δ*clbQ*Δ*clbS*::pETDuet-1-ClbS	Protein expression strain for *in vivo* Δ*clbS* mutant complementation with the *clb* pathway specific resistance protein ClbS	This study
Protein expression strain for *clbR* overexpression		
E. coli DH10B/pCAP01-*clb* Δ*clbP*Δ*clbQ*Δ*clbS*/pETDuet-1-ClbR or (*E. coli* DH10B/pCAP01-*clb* Δ*clbP*Δ*clbQ*Δ*clbS*::*clbR*)	Protein expression strain for transcriptional regulator ClbR overexpression	This study
Protein expression strain for *clbP* expression		
E. coli BL21 (DE3)/pET-28a-ClbP	Protein expression strain for peptidase ClbP expression	This study

Supplementary Table 5. 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-knockout-F	GTGGACGCTGGTTGTAGGAGGGGAATCG GGATTAACACTGTTCAGTGATGGACACTC CTTATTTGATTTT	Deletion of gene *clbB* from pCAP01-*clb* Δ*clbP*Δ*clbQ*Δ*clbS*
clbB-knockout-R	TAAGCGCACGTAACTCAATAGGATCACCC AGCACCGTACCGGTGCCGTGCTTGGTCTG ACAGTTACCAAT	
clbC-knockout-F	AACGGCATGGAAATCGCCATTATTGGTAT GGCGGTCCGTTTCCCGCAGTCGACACTCC TTATTTGATTTT	Deletion of gene *clbC* from pCAP01-*clb* Δ*clbP*Δ*clbQ*Δ*clbS*
clbC-knockout-R	ACTCCTGCGGCTGTATCGGGATATAAAAC GGTGAATGCGCCAGATCCAGCTTGGTCTG ACAGTTACCAAT	
clbDEF-knockout-F	TGCAGGAGTAATGGGAACTGGCGTCGCT CATAACATGGCGCAATACGGCAGACACT CCTTATTTGATTTT	Deletion of gene *clbDEF* from pCAP01-*clb* Δ*clbP*Δ*clbQ*Δ*clbS*
clbDEF-knockout-R	TTTCTTTTTCCCGTTCAGCCGCCAACCGTC GCCATCCTGCTGTAATTCTGTTGGTCTGA CAGTTACCAAT	
clbG-knockout-F	GTGGACGCTGGTTGTAGGAGGGGAATCG GGATTAACACTGTTCAGTGATGGACACTC CTTATTTGATTTT	Deletion of gene *clbG* from pCAP01-*clb* Δ*clbP*Δ*clbQ*Δ*clbS*
clbG-knockout-R	TAAGCGCACGTAACTCAATAGGATCACCC AGCACCGTACCGGTGCCGTGCTTGGTCTG ACAGTTACCAAT	
clbH-knockout-F	ACGCGGAGAATCTGTCGCACTGCAACTGC CTTTTTGTTTCGAATTGATTAGACACTCCT TATTTGATTTT	Deletion of gene *clbH* from pCAP01-*clb* Δ*clbP*Δ*clbQ*Δ*clbS*
clbH-knockout-R	TTGATGGTAGTGAAGCGCAGCAGGTCAA CCAACGCCACGTGCTGACCGCATTGGTCT GACAGTTACCAAT	
clbI-knockout-F	ATAGCTATCATTGGGATGGCGGGGCGTTT CCCTCAAGCCGATACGGTACAGACACTCC TTATTTGATTTT	Deletion of gene *clbI* from pCAP01-*clb* Δ*clbP*Δ*clbQ*Δ*clbS*
clbI-knockout-R	GCTGTTATCGGAAAACGCCCGACAGTGG CCATCGGCGGCGGTGATCCCACTTGGTCT GACAGTTACCAAT	
clbJ-knockout-F	GATCATGTGGCCCGCGCCCTGTTAGCGCT GGGCGTGCAGCATGGCGACCGGACACTC CTTATTTGATTTT	Deletion of gene *clbJ* from pCAP01-*clb* Δ*clbP*Δ*clbQ*Δ*clbS*
clbJ-knockout-R	AAATAGCTCAGCAATAGGTACCGTAACCT TAAAAATCTCCTCAATACGGCTTGGTCTG ACAGTTACCAAT	

clbK-knockout-F	GTACACGGCATTTTACGACTGGGTGCGGTCTATCTGCCAGTGGATCCGGTGACACTCCTTATTTGATTTT	Deletion of gene *clbK* from pCAP01-*clb* Δ*clbP*Δ*clbQ*Δ*clbS*
clbK-knockout-R	CTGCCCGATAATCGCCTCAAGTGCCTGCTGAATACGCACCAATTCTATAGTTGGTCTGACAGTTACCAAT	
clbO-knockout-F	TGGCTCACTGGATATTGCCATTATTGGCATGAGCGGGCGTTTTTCCGGTGGACACTCCTTATTTGATTTT	Deletion of gene *clbO* from pCAP01-*clb* Δ*clbP*Δ*clbQ*Δ*clbS*
clbO-knockout-R	TGTGCGGCGCACATGCCGGTGCGAAAACCCGGTGCAGAGCTTCAAGCTCATTGGTCTGACAGTTACCAAT	
clbPQS-knockout-F	ACACGTTAGCATTAAAACATTATATCATCTCCTGTGCTGTATGCTGCTCTCTCACGTTAAGGGATTTTGG	Deletion of genes *clbP* & *clbQ* & *clbS* from *E. coli* DH10B harboring pCAP01-*clb*
clbPQS-knockout-R	GGCAAACGGTAAGCACCCCGCTATTCTGCAAGACATTTCTGCAGTTTATTGCTCATGAGCTCAGCCAATC	
Colony PCRs for correct insert check		
clbB-knockout check-F	GCAACGCCGTGTCCACCACGA	Colony PCR for correct insert check (*clbB* knockout)
clbB-knockout check-R	TGCGGCGACCGAGTTGCTCTT	
clbC-knockout check-F	TGGCGCGTCACTATCCGCAAGTG	Colony PCR for correct insert check (*clbC* knockout)
clbC-knockout check-R	TGCGGCGACCGAGTTGCTCTT	
clbDEF-knockout check-F	ACGCACCACCCTTATCAGGCACG	Colony PCR for correct insert check (*clbDEF* knockout)
clbDEF-knockout check-R	TGCGGCGACCGAGTTGCTCTT	
clbG-knockout check-F	GCAACGCCGTGTCCACCACGA	Colony PCR for correct insert check (*clbG* knockout)
clbG-knockout check-R	TGCGGCGACCGAGTTGCTCTT	
clbH-knockout check-F	GCACCTGGTGGCGCAGTGGA	Colony PCR for correct insert check (*clbH* knockout)
clbH-knockout check-R	TGCGGCGACCGAGTTGCTCTT	
clbI-knockout check-F	GCAGCAATACATCGGGCAGCAGTG	Colony PCR for correct insert check (*clbI* knockout)
clbI-knockout check-R	TGCGGCGACCGAGTTGCTCTT	
clbJ-knockout check-F	GATCGAGTTGGCTGGGGAGTTGCA	Colony PCR for correct insert check (*clbJ* knockout)
clbJ-knockout check-R	TGCGGCGACCGAGTTGCTCTT	

clbK-knockout check-F	CTCCTGCACGCCCTAGCCCAG	Colony PCR for correct insert check (*clbK* knockout)
clbK-knockout check-R	TGCGGCGACCGAGTTGCTCTT	
clbO-knockout check-F	TGCGGCATGCACCGGAAGACT	Colony PCR for correct insert check (*clbO* knockout)
clbO-knockout check-R	TGCGGCGACCGAGTTGCTCTT	
clbPQS-knockout check-F	CGCTGTTGGGCACTCTTTGGCAA	Colony PCR for correct insert check (*clbPQS* knockout)
clbPQS-knockout check-R	CGAGTGAGGTGGCAGGGGCAAT	
Protein expression for *in vivo* complementation assay		
clbS-BamHI-F	agtgagt**GGATCC**CCTGTTCCATCATCAAAAGAA	Amplification of gene *clbS* for protein expression, inserted into the expression vector pETDuet-1
clbS-HindIII-R	agtgagt**AAGCTT**TTCTGCAAGACATTTCTGCAG	
Protein overexpression		
clbR-BamHI-F	agtgagt**GGATCC**ATGGGGGGAAACATG	Amplification of gene *clbR* for protein expression, inserted into the expression vector pETDuet-1
clbR-HindIII-R	agtgagt**AAGCTT**GATAATCTCATTCCTGTTAG	
Protein expression		
clbP-BamHI-F	agtgagt**GGATCC**ATGACAATAATGGAACACGTTAGC	Amplification of gene *clbP* for protein expression, inserted into the expression vector pET28a(+)
clbP-XhoI-R	agtgagt**CTCGAG**TTACTCATCGTCCCACTCCTTG	

Supplementary Table 6. Accession numbers and descriptions of ketosynthase domains and polyketide synthases used for alignment analysis.

Compound	Protein	AT Type	Accession Number
albicidin	AlbI1	*trans*-AT	AJ586576
albicidin	AlbI2	*trans*-AT	AJ586576
amicoumacin	AmiI	*cis*-AT	SRS606572
amicoumacin	AmiK	*cis*-AT	SRS606572
amicoumacin	AmiL	*cis*-AT	SRS606572
amicoumacin	AmiM	*cis*-AT	SRS606572
antimycin	AntD	*cis*-AT	AP017424
bengamide	BenA	*cis*-AT	KP143770
bengamide	BenB	*cis*-AT	KP143770
bengamide	BenC	*cis*-AT	KP143770
bleomycin	BlmVIII	*cis*-AT	AF210249
calyculin	CalC	*trans*-AT	AB933566
calyculin	CalE	*trans*-AT	AB933566
calyculin	CalF	*trans*-AT	AB933566
carbamidocyclophane	CabD	*cis*-AT	KT826756
carbamidocyclophane	CabH	*cis*-AT	KT826756
chejuenolide	CheC	*trans*-AT	DI339174
chejuenolide	CheF	*trans*-AT	DI339174
chivosazol	ChiD	*trans*-AT	DQ065771
chivosazol	ChiE	*trans*-AT	DQ065771
chivosazol	ChiF	*trans*-AT	DQ065771
colibactin	ClbB	*cis*-AT	AE014075
colibactin	ClbC	*trans*-AT	AE014075
colibactin	ClbI	*cis*-AT	AE014075
colibactin	ClbK	*trans*-AT	AE014075
colibactin	ClbO	*trans*-AT	AE014075
didemnin	DidE	*trans*-AT	NC_017958
didemnin	DidG	*trans*-AT	NC_017958
disorazol	DisA	*trans*-AT	AJ874112
disorazol	DisB	*trans*-AT	AJ874112
disorazol	DisC	*trans*-AT	AJ874112
epothilone	EpoC	*cis*-AT	AF217189
epothilone	EpoD	*cis*-AT	AF217189
epothilone	EpoE	*cis*-AT	AF217189
hygrocin	HgcA	*cis*-AT	CP023992
hygrocin	HgcC	*cis*-AT	CP023992
hygrocin	HgcE	*cis*-AT	CP023992
laidlomycin	LadSIII	*cis*-AT	JQ793783
laidlomycin	LadSIV	*cis*-AT	JQ793783
laidlomycin	LadSV	*cis*-AT	JQ793783
leinamycin	LnmI	*trans*-AT	AF484556
leinamycin	LnmJ1	*trans*-AT	AF484556
leinamycin	LnmJ2	*trans*-AT	AF484556

meridamycin	MerA	*cis*-AT	DQ351275
meridamycin	MerB	*cis*-AT	DQ351275
meridamycin	MerC	*cis*-AT	DQ351275
nigericin	NigAII	*cis*-AT	CP023992
nigericin	NigAIV	*cis*-AT	CP023992
nigericin	NigAVI	*cis*-AT	CP023992
nostophycin	NpnA	*cis*-AT	JF430079
nostophycin	NpnB	*cis*-AT	JF430079
oxazolomycin	OzmH	*trans*-AT	EF552687
oxazolomycin	OzmJ	*trans*-AT	EF552687
oxazolomycin	OzmK	*trans*-AT	EF552687
oxazolomycin	OzmN	*trans*-AT	EF552687
oxazolomycin	OzmQ	*trans*-AT	EF552687
patellazole	PtzA	*trans*-AT	CP003539
pyxipyrrolone	PyxA	*trans*-AT	KY765914
pyxipyrrolone	PyxC	*trans*-AT	KY765914
pyxipyrrolone	PyxD	*trans*-AT	KY765914
rhizopodin	RizB	*trans*-AT	FR854394
rhizopodin	RizD	*trans*-AT	FR854394
rhizopodin	RizE	*trans*-AT	FR854394
rhizoxin	RhiB	*trans*-AT	AM411073
rhizoxin	RhiD	*trans*-AT	AM411073
rhizoxin	RhiE	*trans*-AT	AM411073
rifamycin	RifA	*cis*-AT	AF040570
rifamycin	RifC	*cis*-AT	AF040570
rifamycin	RifE	*cis*-AT	AF040570
tallysomycin	TlmVIII	*cis*-AT	EF032505
tautomycetin	TmcA	*cis*-AT	DQ983361
tautomycetin	TmcB	*cis*-AT	DQ983361
tiancimycin	TnmE	*cis*-AT	KT716443
xenocoumacin	XcnF	*cis*-AT	CCWM01000217
xenocoumacin	XcnH	*cis*-AT	CCWM01000217
xenocoumacin	XcnL	*cis*-AT	CCWM01000217
yersiniabactin	HMWP1	*cis*-AT	AE014075
zwittermicin	ZmaA1	*trans*-AT	FJ430564
zwittermicin	ZmaA2	*cis*-AT	FJ430564
zwittermicin	ZmaK	*cis*-AT	FJ430564

Supplementary Table 7. Amino acid sequences of ketosynthase domains of 80 polyketide synthases used for homology analysis.

albicidin-AlbI1	IAIVAMHCEVPGAGENTEALWSFLRSDVNAIRPIESTRPDLWAAMRAYPGLAGEQLPRYAGFLDDVDAFDAAFFGISRREAECMDPQQRKVLEMVWKLIEQAGHDPLSWGGQPVGLFVGAHTSDYGELLASQPQLMAQCGAYIDSGSHLTMIPNRASRWFNFTGPSEVINSACSSSLVALHRAVQSLRQGESSVALVLGVNLILAPKVLLASASAGMLSPDGRCKTLDAAADGFVRSEGI
	AGVILKPLAQALADGDRVYGLVRGVAVNHGGRSNSLRAPNVNAQRQLLIRTYQEAGVEPASVGYVELHGTGTSLGDPIEIQALKEAFIALGAQAAPSNCGIGSVKSALGHLEAAAGLTGLIKVLLMLKHGEQAGTRHFSTLNPLIDLRGTSFEVVAQHRAWPSQVGIHGTLLPRRAGISSFGFGGANAHAIVEEH
albicidin-AlbI2	AIIGLAGRFPGADTLEEFWNNLRNGQSSMGEVPGERWDHQHYFDSERQAPGKTYSRWGAFLRDIDGFDAAFFEWPDSVALESDPQARIFLEQAYAGIEDAGYTPGSLSKSQRVGVFVGVMNGYYSGGARFWQIANRVSYQFDFRGPSLAVDTACSASLTAIHLALESLRSGSCEVALAGGVNLLVDPQQYLNLAGAAMLSAGASCRPFGEAADGFVAGEACGVVLLKPLKQARADGDV
	IHAVIRGSMINAGGHTSAFSSPNPAAQAEVVRQALQRAGVAPDSISYIEAHGTGTVLGDAVELGALNKVFDKRAAPCPIGSLKANIGHAESAAGIAGLAKLVLQFRHGELVPSLNAFPLNPYIEFGRFQVQQQPAPWPRRGAQPRRAGLSAFGAGGSNAHLVVEEAP
amicoumacin-AmiI	IAIIGMAGRFPGAKNIDEFWHNLQNGKESISFFTDEELFAEGIDEQTFARTDYVRAKGIIDGPDLFDASFFGYSPGQAEVMDPQIRLLHEYVWKTLEDAGCVSAEYEGKIGLFTGTTSNFQWLQHFADSLDGRMSELFEIGSLNDTYTISTRVAHKLNLKGPAITLQTACSTSLVALHLACKSVLNGESDMALAGGVSILHPVKSGYIYQENMVKSPDGHCRAFDKDAKGTVGGDGVGFVAVKS
	LSRALADGDRIYAVVKGSAVNNDGDQKVGFNAPSVEGQTEVIRDAIADAGIEPETISYVETHGTGTALGDPIEIEALTKAFQTNKTAFCRIGSVKTNIGHLDAAAGAAGLIKTVLSLQHQQLVPSLHFKEANPNIDFQNSPFLVNDKLRVWKRTNTPRRAGVSSFGMGGTNAHVILEEAP
amicoumacin-AmiK	IAVIGMACRFPGAKNIEEFWENIKTGRESITFFSKEELMEAGVPKEQLDHKNYVRAKGKVDLHDHFDAAFFGYSQREAEVMDPQVRMFHEVAWESLENAGYNPETYSQPIGLFGAASANLYWQASSMLLRSSSSSEQFAAVQLTDKDFMNTQISYKLNLKGPSIAVDTACSSSLTAVHLASRALLTGDCKMALAGGVTVTCPHKKGYMYQEGMIMSPDGHCRAFDEEAKGTVGGEGCGV
	VMLKTLKQALKDQDHIYAVIKGSAYNNDGSRKVGYTAPSIEGQADVIRKALNISRVESESISYVEAHGTGTSLGDPVEIEALKQAFQTEKKQFCAIGSVKTNIGHLDSAAGIAGFIKTALSLKARTLPPSLHYKKPNPKIDFENSPFYVSTSAKQWTASDRVFRAGVSSFGIGGTNVHVVMEEP
amicoumacin-AmiL	IAVVGMAGKFPGAANIDEYWSNISKGIESVTQFTDEELKEAGVDETLLRQEAYVKAKPIIDEVSDFDAEFFGYSPREAALMDPQIRLMHECTWEALEHAGCDAERFNGLIGLYAGASSNLSWMGRHMSSLHKNEDAFQIMHLNDSSFASRIAYKLNLKGPSVSVQTACSTSLVAIHMACQGLIGGECDLALAGGVTLSLPHISGYLYQEGMIYSADGHCRPFDEKANGTIFGEGAGVVALKRL
	KDAQADGDVIYAVIKGSAINNDGNNKAGYTAPSAEGQANVIATAMEMAETEPESIGYVEAHGTGTPVGDPIEIEALTRAYNTDKKAYCRIGSVKANIGHLDAASGVAGFIKTVMMLHHKMLPPAFHYQTSNPRIKFENTPFFVNTELTEWESLQGPRRAGVSAFGIGGANAHVVLEEP
amicoumacin-AmiM	IAVIGMAGRFPGAKNVQEFWSNLKNAKETISFFTDEELREEGVNEELLNDPRFVKAKGIVEDVDLFDAGFFDYTPKEAAMMDPQFRIFHECVWTALEDAGYDPFTYKGQIGLYTGAGLNAEWVLRALQGAKSGEDSKTLETAVLNMRDYMATLISYKLNLKGPSMMVQTACSTSMVSIHLAAQALLNGDCHMAVAGGVSIRLPQKSGYLYQEGMIHSPDGHCRVFDDEAAGTVFGDGA
	GAVILKTLEDAEADGDHIYAVIKGTAINNDGSRKVGYTAPSTKGQVTAIRTALHMAEVEPETITYVEAHGTGTTLGDPIEIEALKQAYQTDRRGYCRIGSVKSNIGHLNDASGVAGFMKAVLSLKHKVIPPTLNFEKPNAKIDFEQSPFVVNTKLTPWEENEFPRRAGVSSFGIGGTNAHVILEEAP
antimycin-AntD	VAVIGMAGRFPGADDLDAFWDNLAAGRASVRPVTDEEFLAAGGDPRDLDDPSLIRMASVVEGIDRFDSGFFGYTPAEAAVVDPQQRLLLETAYQALEDGGRLAPGTDNGSFGVYAGSGDSRYYPAHIHPRYAGQPGSIELVHAATANSLGTLATRISYELGLTGPSVSLQTACSTALVAVHTACQDLLDYRCDTALAAAVSLNPSALLGYRHVPGGPFSPDGYCRAFAADAAGTSSGDGVGA
	VVLKRLDDALADGDRIRAVIRGSAVNNDGRRKVGFTAPSSVGQAEVILAAQAHAMVDAGTIGLVEAHGTATRLGDPIEVAALTEAFGQSTERRGYCALGSVKTNIGHLGAAAGIAGLIKAVLALERRQIPPSLHFERPNPLIDFAAGPFRVPTALEDWPDAGHPRRAAISAFGVGGTNAHVVLEEAP
bengamide-BenA	IAIVGLACRLPGAGNPSEFWELLRAGRSAVKPIPAERWNAERWPTPSRWAGFLDRVEPFEPAFFGISPREARLMDPQQWLMMKLAWEALGDAGLPPARLRGSMTGVFLGAVWHDFADVKLRKGEPVAQHSATGRALNMLANRVSYGFGFEGPSLVLDTACSSSLVAVHLAARSLRAGESSLALAGGVNLILSPETLLALSKFGGLSPGPTCWVFDERANGYVRGEGAAIVVLKRYSQALK
	DGNHIYALVRGSAMNNDGASNGLTAPNPLAQQDVLRRACADARVHPSELQYVEAHGTGTPLGDPIEVSALGAVLSAERLPSEPLRIGSVKTNIGHLEGAAGIAGLLKLTLAMHHRELPPTLNFERPNPNIPFAALGVQVQTQRSPWPASSGSPLLGGVSSFGWGGTNCHVILQSA
bengamide-BenB	IAIVGMSCRFPGASSPSAFWQMLRNGTDGISEVPASRWPVDAFYDPDPDRPGKMASRWGGFIDGVDAFDAHFFRCSPAEAAGMDPQQRLFLDVAWEALEDAAQPLDELSGTQTGIFVGAMTNDYAQLQDATQGLKQIDAHYAIGTEFSFIAGRVAHLLGVRGPCMMLSTACSSSLVALHLATRSLRSGESDLSVVGGVNVILSPRMHVASSKLRVMAANGRCKTFDASADGYVRGEG
	CGVVVLKRLSDAVKQGDRIHAVIRGTAVNHDGPSSGLTVPSSDAQAAVIRSALADAGADSATLTLLEAHGSGTPLGDAIEIEGIRRALGAGLEDRRPPLLVGSVKTNVGHLEAAAGMAGLFKVVLALRNGEVPPHLHLKQLNPQLNMSGLQVPVRLGPWPDAPERPLAGVSSFGLSGINAHVVVEAP
bengamide-BenC	IAVIGIAGRFPGARDVQAFWNNLRSGVESIRFFSEEELLGQGVPAELLRNPNYVRAKGILSDVDCFDAGFFGISPAEARLMDPQQRLLIECAWEVLEDAGYAPGQLQERVGMHVGVGSNRYASEHVWPSLKGEHWLEHVEAAIGNEKDTAATRIAYLLDMKGPCFTVQTFCSSSLVAVHQAVSSLLNQECDIALAGGAYVEVPHVVGYLHREGSFTSPDGHCRAFDAGAAGTPFGSAVALV
	ALKRLEDALEDNDAIYGVIKGSAINNDGSQRAGYAAPSADGQAQVISEALGMAGAEPESISYVEAHGTGTSLGDPIEISALTRALSRGGRPHPGCLLGSVKTNIGHTDRASGVVSLIKVLQALAHGEIPPTLHFSKPNPAIDFGPFEVPTRLREWPRRDTPRRAGVSSFGLGGTNAHVIVEEAP
bleomycin-BlmVIII	IAVISLGGRFPGADRVDRLWTNLLDREDAISHFTADERLARGRDPELVRHPRFVGAEGVLGDVSLFDAEFFGCSPREAEVMDPQHRLCLEEAWHVFDTAGYDPAATGTAVGVFLSASLSSYLIRNVLPGGAAQRLLGGFPLLIHNDKDFLATTVSHKLGLTGPSYAVGSACSSSLVAVHLACQSLLTEECDMALAGGVSLQVPQGQGYVHADDGIYSPDGRCAPFDAGAAGTVGGSGVGLVLL
	KRLADAVRDGDRVHAVILGSAVNNDGADKVGYTAPGVTGQSAVVAEALAVAGISAATVGVLEAHGTGTRLGDPVEVAALTRAFRAHTDRSGFCALGSVKANVGHLDAAAGVTGLIKAVLAVREGVIPGTPHYRSPNPAIDFATTPFYVTADTLAWPEADHPRRAGVSSFGIGGTNAHVILEQAP
calyculin-CalC	IAVIGLSGRYPGAPSPEVFWENLKNGVDSISEIPRERWDYRTLPDGAIARWGGFIDDVDCFDPLFFNISPREAERMDPQERLFLETVWHAVEDAGYSREALEAAFAGRVGVFTGAMYGEYQFFSAPLSESGFAISTSYGSIANRVSYIFNLHGPSLAIDSLCSSSLTALHLAVESIRRGECLAAIAGGVNLSLHPNKYLLLAEMQMASADGRCRSFGAGGTGMVPGEGVGAVLLKPLAQAERDG
	DRIYGMIRASAINHDGRTNGYTVPNPNAQAEMIESALARAGTCPEWISYVEAHGTGTPLGDPIEIAGLTRAFATDKTGYCALGSVKSNIGHPEAAAGIAGLTKILLQMQHCQLVPSLHAEETNPSIDFARTPFTLQRECCEWKRPTIDGREIPRLACISSFGAGGANAHAVIEEY
calyculin-CalE	IAIVGVGGFFPGAENFHDFWGKILAGQDLISEMPEPRRSSVYTVYEQEISDLNGLYGGFIKNADRFDADFFQYRHEEVMAMDLQQRLFLEATWATLEDAGYHPLSLSEQQVGVYVGAIGNDYRAFLEDCGYPIDMFYEGTGASLAGIANRTSYFYNFNGPSQIVDTACCSSLYAVDMAVQAIRQSLCETAIAGGVSSICTATGFHMYAAMDYLSKDWRCKTFSDEGDGWSKGEMVAAVYLK
	PYDKALRDRDHIYGLIRSSGTNHGGRAFFYTQPNNSKHVALIKDVYRKAGVDPRDVAYIEAHGSGTKMGDALEYNGISRALKELAREADVGLDNGRCGIGSIKSNIGHTEAAAGIAGLLKALMILHDGQIPASLHIDRENENIAKASPLYLTRERQPLPPARIAGRPRHAGVHSFNFSGATVHALLAE
calyculin-CalF	IAIIGLAGTYPHAPNLETFWQNLLEGRDCIDEIPAERWSLDFFDPDPKAAAQNGRSYGKWGGFIENLYHFDPLFFGISPREAVIMNPKEPLVMQCAWHVLEDAGYSPDGLSDKTVGVFIGVTRAGTDPYPGTFASMTNRISYAFDFQGPSFPIDTMCSSSLVAIHEACRHLRDGECSVALAGGVNAYVDPSHYAVLAAGQFLSPDGRCRSFGADANGMVPGEGVGVLLLKPLSRAEQDGDH
	IYAVIRGSATNHGGRTNGFTVPNPKAHRDLIRLALDRAGVDARQVSYVEAHGTGTTLGDPIEIRGLAQAFAYDTEDTGFCSIGSVKSNIGHLEAAAGISGITKILLQMKHGSLVPSLHSDSLNRQIDFPSTPFFVQQDVEPWNPTDENGDPKARIACVSSFGAGGSNAHVVLEAY
Carbamidocyclophane-CabD	IAIIGIGVRFPGAQNPESFWHLLREGVDAITEIPPERWNVDEFYDPKPATPGKMITRCGGFLDDVGSFDAGFFGISPREAECIDPQQRLVLEVAWEALENAGIAPHTLNSSQTGVFVGIGNHDYGKLQSWQFSRLNAYDGTGRTLGVAANRLSYVLNLQGPSLVVETSCSSSLVAVHYACQSLRTKESNLCLVGGVSVMLSPESHIIFSQAKMMAPDGRCKTFDASADGYVRGEGCGVVVLKR
	LSDALRDGDNILATIRGSAVNQDGRSNGLTAPNGPAQQAVIRKALDNAGVTPDQISYVEAHGTGTALGDPIEVNSLKTVLMKGRQPNQTCYIGSVKTNIGHLEPAAGIAGLIKVVLALHHKQIPPHLHFQQLNPYIKINNTPLTIPTKLQEWDFHEGSRLAGVSAFGFGGTNAHVILEEAP

carbamidocyclophane-CabH	IAIIGISCRYPGAKNWREFWENLKNGVDSVTEVPPGRWQEKEWYHPDPQHPGTSYSKCAGFLDEIDKFDPLFFQISPVEAQFIEPQQRIFLEEAYHAIEDAGYAADSLKGKQCGVFVGAGTNGDYNRLLSLAGLDTNRLALTGNLLSMIPARIAYFLDLKGPVVALDTACSSSLVAVHQACESIQRGESELAIAGGIAIMSTADFQVLSSQFQMLSAAGRCKTFDESASGTVWSEGCGVVLLKSYEQAIRDNDHIYGIIKGTGVNYDGNTNGISAPSGQSQTRLEQGVYEKFGINPETISYVEAHGTATPLGDPIEVEALTEAFSKWTSKKQFCAIGSVKTNIGHAATSAGISGLIKTVLCLKHQKLVPSLHYDQANPHIDFENSPFYVNTEFKDWQVSDRHPRRATVSSFGFSGTNAHLVVEEAP
chejuenolide-CheC	IAIVGMAARLPKAQDVARFWENLRNGRDCTDVIPAERWKQEGFLSEAPLNGRGSYTNRGAFIDDVDAFDHVFFNLTPNEAARMDPQERIMLEQTYRSMLDAGYTRKQWAGSDTAVFVGVMNGDYAWHTPAQTTTAPATSLFWSMANRASYFFDWRGPSMAVDTACSASLTALHLACQALKNGDCEQAVVGGVNLITHPRHYELLCGLHMLSRSEQCKPFGANADGFVDGEGVVCLVVKRAEDAIRDKDRVYAFIKGSAINAGGRSNGYTAPNPEAQAALIGKALRAAGVSPQEVGYVEAHGTGTELGDPIELRALSKSYGDAAPQSIRLGSVKSNLGHLESAAGLCGVLKAVLQMHHGEWVPSLHAEQLNPHLDFTQTPFLLNRERRVWDATAPRVSAVSSFGAGGGNAHVVIQ
chejuenolide-CheF	IAIIGMAGEFPNGATPEAFWRSLAAGEDAVRVIPQSRWDWRRDYSASTAQGEATYGRHGGFMEGVAEFDPAFFNIAPVDAALLDPQERRFLQVSYHALEDAGYFVSPSKDVGVFAAAMFGHYQDLDASRVVSSSFAAIANRVSYAFDLQGPSLTVDTMCSGSLTALHMACNSLRLGECRMALSGGVNIMAHPGKYRLLSQGKFLSASGHCHAFGVEADGYVPGEGAAVVVLKSVADALRDQDVIYAIVRATAINSGGKTSSFTVPSARAQQRVIQDALRKSGVNPAQVNYIEAHGTGTGLGDPIELQALQSAYGANLEGADTPCYLGSVKSNIGHLESAAAMAGLFKVVQQFAHEQLAPTLHCAIENPYLNIEQTRFQLVREKQPWPLAGAATRFAGLSSFGAGGANGHVILQQY
chivosazol-ChiD	IAIIGVAGRYPGADDLREFWQNLREGKDCITEIPADRWDHAAWYDPDRSKLGRIHNKWGGFLRGIDLFDPMFFRISPREAEFMDPQERLFLEVAWSTFEDAGYTRELLQRRHGGCVGVFAGVMYSEYPFYGVEATLRGRPVAVGLGYGSIANRVSFVMDLNGPSLSVDTLCSSSLTCLHLAVESLRRGECAMALAGGVNLSLHPNKYLQHSLFKMTAPDGRCRSFGEGGDGFTPGEGVGCVLLKPLSHALRDGDPIHGVIRGVAVNHGGRSSGYTVPSAKAQAALVKAAVARAGIPARSIGYIEAHGTGTPLGDPIELEGLNQAFQGAAEPGFSCPIGSLKSNMGHLEAAAGVAGLTKVLLQMKHGELAPSLHAEALNPNIDWTATPFFVQRRPAPWERATVEVDGRSQVLPRRAGISSFGAGGANAHVIVEEAP
chivosazol-ChiE	IAVIGLSGRFPGSPDLDAFWDNLAAGRDCITEVPKDRWDPDRIGGSAAQFRWGGFLDDIDKFDPLFFKISGREAELSDPQQRLFLEEAWKALEGAGYTPEALDRARCGVYAGALDGNYQAFMSPKDLDEPQTLWANDTSVIAARIAYFLNLKGPAITINGACSSSLIAIHMACQALWSGEVDLMLAGGVCLMTSHHFHEVAGKAGMLSPSGRCRTFDDGADGFVPGEGVGVVLLKRLDDALRDGDPIRAVIAGSGVNQDGRTNGITAPSALSQAELEARIYRDFDVHPESITLVEAHGTGTRLGDPIEVNALTDAFRTLTSKTGYCAIGSVKTNIGHTALASGVAGFIKAVLAMERGQIPPSLHFRGANEHIDFANSPFFVNTELRPWRPSGPRRAAISSFGMSGTNAHVVLEEP
chivosazol-ChiF	IAVIGVSGRYPMADDLDAFWENLAAGRDCITEVPADRFPVDAYYDPEVGKLGKTYTRWGGFLSDVDRFDPLVFNIAPREAELMDPQERLFLETAWATLENAGYARGALAKRRVGVFAGVMWGQYQLLGAERAGRSTVLPASIFASIANRVSYCFDWRGPSLAVDTMCSSSLTAIHLACESIRRGECEMALAGGVNVTVHPMKYLYLSQGRMVSTDGRCRGFGQGGDGYVPGEGVGAVLLKPLSRALEDGDHIHAVIKGTAINHGGRSSGYTVPNPEAHGALIRDALSAAGFEPESVSYVEAHGTGTSLGDPIEMTGLAAAFPGLPARSRALGTVKSNIGHLEAAAGIAALTKVLLQLEHRQLAPSLHADPPNPNIDFDASPFHVQRELAAWQPPEVPGQQGRARYPRRAGISSFGAGGNNVHVLLEEH
colibactin-ClbB	VAIVGMAGRFPGAANIAALWTLVVGGESGLTLFSDEELRAHGVTPDTLKQANYIKTKGIVDDHEWFDADFFGYTPNEAECMDPQIRLLHQCCWQTLEHAGCDPATFTGAIGIYAGLLTSPHWLNAVMQDTTDSTALYKASILNIHSVTALIAHALNLTGPAVTLDTACSTSAVAIHQACIALRNRDCDAALAGGVSIEMPAYRGYEYHEGMINARDGVCRPFDSQASGTVTGDGLGMLLLKRLDDALADRDCIYGVIKGSAVNNDGNNKIGYTAPSVIGQSTVIRTSLRRAGFDSDSIGLVEAHGTGTVLGDPIELRALNEVFGPTPVPFCVVSALKSNIGHLNSAAGVAGVIKTTLALHHQVLPPTAHFRQLNPAIDLSRSALYVVNQQVQPWPSTRPRRALVSSFGIGGTNASIALEA
colibactin-ClbC	IAIIGMAVRFPQSRTLHEFWHNIVQGKECVTFFSEEELLAEGVEQSTLDNPAYVRAKPYIEGICDFDAAFFGYSHKEAQTLDPKSRVLHEVAYHALEDAGYAQRTSDLITGVFVGASEDVDWLRRSLSQIGGDALNRFESGIYGHKDLLAHLIAYSLNLNGPVYSLYTSCSTSLSATHIACRSLLFGECDLALAGGITIDLPQKSGYFCQQGMIHSTDGHCRPFDSQASGTLFGDGAGVVVLRRLEDALAAGDRIYAVIRGSAVNNDGKQKIGFVAPGHEGQKAVICAACHLAEVSPESIGYVETHGTGTRIGDPIEFAALTEAFDTSHRQYCALGAVKANIGHTHAAAGVAGLIKTALVLHHRTIPPLANYQMPNSKLDLAHSPFYIPIQPQEWPASRMPPRAGVSSFGIGGTNVHMILEG
colibactin-ClbI	IAIIGMAGRFPQADTVQAFWENLLASRECISFYSDEELLAMGISPEFVQHPDYVKAKGEVADIDKFDAAFFGIAPREAELMDPQHRVLLETAWAAFEDAGYVAADYPGDVGIFAGKSMDSYLMLNLMPHFKRVFSSGSLQAAIGNDKDSITTTIAYHLNLRGPAITVQTSSSTSLVAVCVACQSLLTWQCDMAIAGGVTLGPPAKTGYLSQEGGITAADGHCRAFSDNSSGFVPGTGAGLVVLKRVDEALRDGDNIYAVIKGFAVNNDGSEKISYTAPSVDAQARAIAQAQRLAGLTPQDITYVEAHGTGTRLGDPVEFSALSQAFAGASQKQYCALGSVKTNIGHLDTAAGVAGLIKTALAVQQGIIPATLHFERPNAQIDLTNSPFYINTTCQPWQPESGIRRAGVTSLGMGGTNAHVVLEQAP
colibactin-ClbK	IAIVGMNCRYPGVHSVAAFETVLRTGCNILDPKVTPSNGHNHITLNNVYEHMAEFDANFFGYSRAEAEIMDPQQRVFLTCAWEMFEQSGYNPKQHDARVGLYAGVSTSFYLLTHLMNNPDKLAQLGGLQIMVGNDKDHLTSQLAYRLNITGPCVTVQASCATSLVAVHLACEGLLSGQCDMALAGGVTFRMEEQRSYESHGDGLQAEDGLIHTFDAQASGTVYSSGLGMVLLKRATDAQVQGDNILAVIKGSAINNDGGARSGYTVPGVDGQEAVMIEAHSLAEVTPQQIQYLELHGSGTPLGDAIEFAAIKRVFGTPAPNATPWRLGAVKPNVGHVEMASGITSLIKTVLSLTNRVFYPTLNFQRANPQLGLEDSPFEVVSRLTPWPEGTTPRTAGVSAFGLGGTNAHLVVQA
colibactin-ClbO	IAIIGMSGRFSGAESVPEWWDKLLAGEEFTQPTCTEDDNGNPWIRLRNIITAPYDFDAAFFNIPPGEALLMDPQQRIFLECCYNALEHAGYIPTQLKRVGVYGATYANNYFIDRVYPYLKMSGDHHYLQAQIGNEKDYLCAQVAYKLGFTGPAVSVQTACSSSLVAAYLACEGLLTFQADVALAGGVTLGFLQAHGYSPQGDKLVSQDGHCAPFSAEATGTVYSSGAGVVVLKRLEDALRDQDRVYAVIKGGAVNNDGGRRLGFVAPSVEGQVEAINTALAAAAEVVPTDIALIETHGTGTPLGDEIELEALHRVFAPACAPHSIQLGAVKANLGHLGVASGIVSLMKTALTLYTGLVPPQINLVNKHKKLLQPASPFYLSDVVTSVPQTKRIHATVSSFGLGGTNAHLVLQN
didemnin-DidE	IAIIGMACRVPGADGPEALARLLAEGREGIADLPADDGIDDPRWVRRKGLLADPFGFDAGLFGHAPREAALIDPQARLLLELGWEALERAGQPPRATEGQGRTAVYAGAGISTYLLTHLLNNPAAADLATPFEAVLANSGDSLVTRIAYELDLRGPAVDVQTACSTSLVAVHLAVQALIAGECRMALAGGASIDLPAPAGYWYAPGMILSPDGHCRAFAADAAGTVPASGGGMVVLKRLVDALEDGDHIHAVIRGSATNNDGAARIGFTAPGVEGQAEVIAEALTLAGIASEQIGMIEAHGTGTPIGDPIEVQALTRAFGPEVPRGQCLLGSIKSNLGHLNAAAGVVGLIKAVMAVEAGMVPPTLHADQPNPALDLAATPFRLATALSPWPAIPGPRRAGVSSFGMGGSNAHVVIEQAP
didemnin-DidG	IAVTGLACRAIGVDGPDALRRVLLEGRDVLARHDAADAAALGVPAALARDPRYVGTAAMLPGQESFDHALFGLSLREAELLDPQHRLFLELCREALEDAAWLAPAAGRRVGVFGGCGSPIYLMTRLAGAGAALEAAGATALSFAAKPDFLSLRVAYRLDLRGPAMTVQTGCSTGMSVVQAAVAALAAGECEAALAGAACLTRLDPAGYLHHPGAIESPTGRCRPFDAAADGTVGGSGGAVVVLKRLADAERDGDHVHAVIRAVAITNDGGAKIGFTAPGEAAQAEAVARAFRASGIHPDTIGCVEAHGTATALGDPIEVAALTRGFRAAGGTRLAHAALGSVKGNLGHADAAAGMLSIAKAVLSVAHGEIYPTANTTRANPELRLAETPFRLPTAPEAWPVADHPRRMAVNCYGMGGTNLHAILEQAP
disorazol-DisA	IAIIGMSARFPKSPDLDAFWQNLLSGRDCVDEIPAERWDHRRYFAEAAQPHKTYGRWGGFIEDVDRFDPMFFNISPREAEQMDPQQRLFLECAWATMEHAGYGDPRAYGDRAVGLFVGVMWNEYSRIGSQLTLQTARYAGPGSLYWAIANRVSYWMNLTGPSLAIDTACSSSLVAVHQACMSIRNGECDMAMAGGINLSIHPDKYLYLAQSKFLSLDGRCRSFGQGGTGYVPGEGVGAVLLKPLEQALRDGDHVYGIVRGSAINHGGRATGFTVPDPEAQARLVFDALRRARVSPDQLSYIECHGTGTALGDPVEIAGLSKAFRMAGATRTSIPIGSVKSNLGHLEAAAGIAALIKVLLCMQHQAIPKSLHSDVKNPNIRFEEVPFEVVNETRSWQGDGGAPRFAGVSSFGAGGSNAHVILESY
disorazol-DisB	IAIVGLSGRYPMAPDLDAFWENLAAGRDCVVEIPADRWDHGRYFDPNPGAAGKSYSKWGGFLDDVDRFDPLFFNIAPREAEAMDPQERVFLEVAWHALEDAGYARSPLANRATGVFVGVMYGHYQLFGAEALALDRPVSAGSSFASIANRVSYFFDFRGPSVALDTMCSSSLTAIHLACAALQRGEIEMALAGGVNLSLHPQKYILLSRGKFMATDGRCRSFGEGGDGYVPGEGAGAVVLKRLDRAIADGDRIHGVVKASALNHGGKTSGYTVPNPSAQADVVAAALAQSGVDPRTITYVEAHGTGTSLGDPIEIAGLTRAFEASPKERPTCAIGSVKSNVGHLESAAGVAGLTKVLLQMAHEQLVPSIHADPPNPNINFAESPFRVQRELGPWRAPVDERGQRLPLRAGLSSFGAGGANAHLVLEAY

disorazol-DisC	IAVIGLAGRYPGADTPRQLWRALRSGQSAVTRPPAGRFGASAPQGDEPRGGGASPGWGGYLERLDRFDSLFFGISPAEAKLMDPQERLFIEVAWECLEDAGYTPEELRRAAPRVGVFVGAMWSDYQSVGLEAWQRDRRAKAVAFHSSIANRISYLFDLHGPSVAIDTSCSSGLTALHLASRSLRLGECDVALVGGVNLLGHPFHPDLLEGLNLTSRDDKTRAFGAGGSGWVPGEGVGAVLLRRLPEAEERGEHIRCVLKGTALAHAGKAPRYGMPSTRAQAGSIRDALADGGVAASEIDYVECAATGSGIADASEVDALKQAFEGRSPDGPPCLLGSVKPNIGHLESASALSQLTKVILQLEHGEIAPTLHTEPRNPLIQLDGTPFRINRALSPWPRAAGADAPPRRALINAFGATGSSAHAVVEEY
epothilone-EpoC	IAVIGMSGRFPGARDLDEFWRNLRDGTEAVQRFSEQELAASGVDPALVLDPNYVRAGSVLEDVDRFDAAFFGISPREAELMDPQHRIFMECAWEALENAGYDPTAYEGSIGVYAGANMSSYLTSNLHEHPAMMRWPGWFQTLIGNDKDYLATHVSYRLNLRGPSISVQTACSTSLVAVHLACMSLLDRECDMALAGGITVRIPHRAGYVYAEGGIFSPDGHCRAFDAKANGTIMGNGCGVVLLKPLDRALSDGDPVRAVILGSATNNDGARKIGFTAPSEVGQAQAIMEALALAGVEARSIQYIETHGTGTLLGDAIETAALRRVFDRDASTRRSCAIGSVKTGIGHLESAAGIAGLIKTVLALEHRQLPPSLNFESPNPSIDFASSPFYVNTSLKDWNTGSTPRRAGVSSFGIGGTNAHVVLEEAP
epothilone-EpoD	IAIVGAACRFPGGVEDLESYWQLLTEGVVVSTEVPADRWNGADGRGPGSGEAPRQTYVPRGGFLREVETFDAAFFHISPREAMSLDPQQRLLLEVSWEAIERAGQDPSALRESPTGVFVGAGPNEYAERVQDLADEAAGLYSGTGNMLSVAAGRLSFFLGLHGPTLAVDTACSSSLVALHLGCQSLRRGECDQALVGGVNMLLSPKTFALLSRMHALSPGGRCKTFSADADGYARAEGCAVVVLKRLSDAQRDRDPILAVIRGTAINHDGPSSGLTVPSGPAQEALLRQALAHAGVVPADVDFVECHGTGTALGDPIEVRALSDVYGQARPADRPLILGAAKANLGHMEPAAGLAGLLKAVLALGQEQIPAQPELGELNPLLPWEALPVAVARAAVPWPRTDRPRFAGVSSFGMSGTNAHVVLEEAP
epothilone-EpoE	IAIVGIGCRFPGGAGTPEAFWELLDDGRDAIRPLEERWALVGVDPGDDVPRWAGLLTEAIDGFDAAFFGIAPREARSLDPQHRLLLEVAWEGFEDAGIPPRSLVGSRTGVFVGVCATEYLHAAVAHQPREERDAYSTTGNMLSIAAGRLSYTLGLQGPCLTVDTACSSSLVAIHLACRSLRARESDLALAGGVNMLLSPDTMRALARTQALSPNGRCQTFDASANGFVRGEGCGLIVLKRLSDARRDGDRIWALIRGSAINQDGRSTGLTAPNVLAQGALLREALRNAGVEAEAIGYIETHGAATSLGDPIEIEALRTVVGPARADGARCVLGAVKTNLGHLEGAAGVAGLIKATLSLHHERIPRNLNFRTLNPRIRIEGTALALATEPVPWPRTGRTRFAGVSSFGMSGTNAHVVLEEAP
hygrocin-HgcA	IAIVGMACRLPGGVGSPEELWELVLAGGEGIGEFPADRGWDLEKLFDPDPDHAGTSYARRGGFLHDAGEFDADFFGISPREALAMDPQQRLLLETSWEALERAGIDPVSLRGQDVGVFAGMMHQNYGVGAGEGTDAPVGLEGHLMTGTSASVVSGRVSYVLGFEGPAVTVDTACSSSLVALHLAAQALRAGECSMALAGGVTVMAGPDSFVEFSRQRGLAVDGRCKSFSASADGTGWAEGVGVVVLERLSVAERCGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRRALAGAGLVAADVDAVEAHGTGTPLGDPIEAQAVLATYGQDRDRPLLMGSLKSNVGHAQAAAGVAGVIKMVLALRYGVLPRTLHVDEPTPQVDWSSGAVELLTEERVWPEVGRPRRAGVSGFGVSGTNAHVILEQAP
hygrocin-HgcC	IAIVAMACRLPGGVSSPEQLWDLVEHGRDGITGFPVNRGWDLANLFHPDPDHRGTSYVREGGFLHDAGEFDAAFFGISPREAVTMDPQQRIMLELSWEAFERAGIDPTAWRGKDVGVFSGFAGSDYGSGLPELPEALESYLMTSDAGSVLSGRISYTLGFEGPAVTVDTACSASLVALHLAARALRAGDCSMALAGGVTVLGTPRAFIGFSRQRGLAADARCKAFAAAADGTGFSEGAGVLLLERLSDAERNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANAHLTPSEVDAVEAHGTGTALGDPIEAQALLATYGQGRPEDRPLWLGSLKSNIGHTQGAAGVAGVIKMVLALRYGVLPRTLHVDEPSPQVDWSSGAVELLTEQRVWPEVDRPRRAGVSGFGVSGTNAHVILEQAP
hygrocin-HgcE	IAIVGMACRLPGGVTTPEGLWDLVASGGDAISGFPEDRGWDLENLFDPDPDHPGTSYVREGGFLRDAGEFDAGFFGISPREALAMDPQQRIMLELSWEAFERAGIDPTALRGKDIGVFSGVTYHSYGSGARVPEELEAVMGTGTSASVLSGRVSYALGFEGPSVAVDTACSSSLVALHLAVQALRLGECSMALAGGVTVMANPGIFVGFSRQRGMSKDGRCRAFAAAADGTGFSEGAGVLLVERLSDAERNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALASAGLVASDVDAVEAHGTGTALGDPIEAQALLDTYGQDRPEGRPLWLGSLKSNIGHAQAAAGVAGVIKMVLALRCGVLPRTLHVDEPSRQVDWSAGAVELLTEERVWPEVGRPRRAGVSGFGVSGTNAHVILEQAP
laidlomycin-LadSIII	VAVVQMACRYPGGVRTPEDLWELVAAGGHALGAFPDNRGWDLARLFHPDPDHPGTTYASEGGFLHDADLFDPEFFGISPREAQALDPQQRLLLECAWEALERAGIDPQSLQGSRTGVYAGAALPGFGTPHIDLEAEGHLVTGNAPSVLSGRLAYTFGLEGPAVTIDTACSSSLVAVHLAAHALRRRECDLALAGGVTVMTTPYVFTEFSRQRGLAADGRCKPFAAAADGTAFSEGAGLLVLERLSDARRNGHEVLAVIRGSAVNQDGASNGLTAPNGPSQQRVIRAALAGARLSPAEVDAVEAHGTGTKLGDPIEAQALLATYGQERHEDRPLWLGSVKSNIGHTQGAAGAAGLIKMILALRHGMLPPTVNVDRPTPHADWESGAVRLLTEPVTWPRGERVRRAGISSFGISGTNAHLILEEP
laidlomycin-LadSIV	IAIVSMACRFPGGVSSADELWDLVESGQDAMGAFPTDRGWDLDRLFHPDPDHPGTSYSDQGGFLHDAGDFDASFFGISPREALAMDPQQRLLLEASWEVLERAGIDPTSLKGTLTGTYVGVMYHDYGKAFPEADAQLEGYAYLASSGSVVSGRVAYTLGLEGPAVTVDTACSSSLVSIHLAAQALRQGECDLALAGGVTVMADPDVFAGFSRQRGLSPDGRCKPYAAAADGVGFSEGVGLILLERLSDARRRGHRVLGVVRGSAVNQDGASNGLTAPNGPSQERVIRQALASGRLVASDVDVVEGHGTGTTLGDPIEAQALLATYGQGRPDGRPLWLGSVKSNIGHTQAAAGVAGVIKMVLAMRRRGVVPASLHVDEPSPHVDWEAGAVRLVSEAVPWPESDRPRRAGVSSFGASGTNAHLIIEHV
laidlomycin-LadSV	VAVVSMACRFPGGVESPEGLWELVAQGGDAIEGFPADRGWDLDGLYHPDPDHPGTCYVREGGFLSGADRFDSGFFGFSPREALASSPQLRLLLETSWEVLERAGIDPTTLKGSPTGVYVGAATTGNLTQGDPAGKATEGYAGSAPSVLSGRVSFTLGLEGPAVTIETACSSSLVAMHLAAQALRQGECDLALAGGVTVMSTPEVFTGFSRQRGLAPNGRCKPFAAAADGTGWGEGVGLILLERLSDARRNGHKVLALLRGSAVNQDGASNGFTAPNGPSQRRVIRQALAGANLSTSEIDVVEAHGTGTKLGDPIEAEALIATYGKEREDDRPLWLGSVKSNIGHTQAAAGVAGVIKMVMALQRELLPATLHVDEPTPHVQWEGGGVRLLTEPVPWSRSERVRRAGVSSFGISGTNAHVILEEP
leinamycin-LnmI	IAVIGIAGRYPGAGDLETFWSNLAEGVDSVGPLPAERARDGWPTEQMWGGFLDGVDRFDALFFGIAPRDAQLMDPQERQFLQVVWETLEDAGCTRARIREQLGSDVGVFVGTMYNEYPFFGVERSLAGESADTGSAVAGIANRVSYFLDLHGPSLAVDTMCSSSLTALHLAVESLRRGECAAAVAGGVNLSLHPHKFRQQTRLKMSSSDHRCRSFGAGGDGFVPAEGVGAVLLKPLSAAEADGDRIHAVIRGTAVNHGGKTNGYMVPNPVAQGDLVRAALRRAGADPATIGYVEAHGTGTQLGDPVEINGLNRAFAGASVAPASRAIGSVKANIGHAEAAAGIAGLTKVVLQLRHRHLVPSLHTEELNDAVDWASSPFEVVREGRPWAPLTGADGAPLPRRAGLSAFGAGGANAHVVVEEY
leinamycin-LnmJ1	IAIIGVAGRYPEAEDLEAFWRNLAEGRDCVGEVPADRWDHAAYYDPERGKEGRTYGRRGGFLDGVDRFDAASFGISRREAELMDPQERLFLTVGRQAVENAGYRPEELARTRVGVFAGVMWNHYQLCTDGSAEPVAPTALHCSVANRLSYCLDLSGPSMAVDTACSSSLTSLHLAVESIRRGECALAVAGGVNVAAHPQKYLQLAQGRFLSSDGRCRAFGADGDGYVPGEGVGAVLLKPLADALADGDHVHAVIKGSFLNHSGRTSGFTVPSPAAQATLIADALDRSGVAADSVGYIEAHGTGTALGDPIEIEGLRQAFADAGLAPGSCAIGSVKSGIGHLESAAGIAAVTKVLLQMRHRELVPSLHSEQPNPHIDFAATPFAVQRTRAPWVPRPGSTVLRAGVSAFGAGGSNAHVLLESAP
leinamycin-LnmJ2	IAVIGMSGRFPGAEDLDAFWENIAAGRDSFTEVPAQRWDVGPVFDADRLVPDRTYSKWAAMLPEVGRFDAAFFNHSPLEAEVMDPQQRLFLEQSWAALEHAGYAVGADDRTSCGVFVGCAPGDYSTLLTEAGRADTGHAFLGTTSSLLPARIGYFLNLDGPTMAVDTACSSSLVAVHLAADSIRRGECAMALAGGVALMVTPQLHVRASKVGMLSPRGTCVPFDASADGTVLGEGVGAVVLKRLDRAVADGDHIHGVIKATGVNGDGRTNGITAPSALSQAALIADVHRRAGVGADDIGYVEAHGTGTALGDPIEVRALTEVFRRSTDRSGYCGIGTVKANIGHTTMAAGIAGLLKTLLALRHSELPPAPAFDTPNPKTELDSSPFFVVRDRQEWEPGPGGQRIATVSSFGFSGTNAHAVLAQAP
meridamycin-MerA	VVIVGMGCRFPGRAHSPEDLWRIVADGEDAISGFPSDRGWDLAGLYHPDPDHPGTSYARDGGFLYDAAEFDAGFFGISPREAEAMDPQQRLLLETSWEALERAGIPAEHIKGSSTGVFIGASSVGYAADAGEAAEGYQLTGTAASVASGRVSYTLGLEGPAVTVDTACSSSLVALHLAVQSLRAGECSLALAGGVTVMATPAMFVEFSRQRGLAMDGRCKAFAAAADGTGWAEGVGVLVVERLSDAERNGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALASAGLVASDVDAVEAHGTGTTLGDPIEAQALLATVFRRSTDRSGYCGIGTVKANIGHTTMAAGIAGLLKTLLALRHSELPPAPAFDTPNPKTELDSSPFFVVRDRQEWEPGPGGQRIATVSSFGFSGTNAHAVLAQAP
meridamycin-MerB	IVIVGMSCRYPGGITSPEALWDLVRSDGDAISVLPADRGWDLDGLYDPDPDRTGTSYARSGGFVYDAAEFDAAFFGISPREAAAMDPQQRLLLETSWEAFERAGIPATSVKGERIGVFTGVMHHDYLTRLSTTPDAVEGYLGTGAAAGVASGRVAYTFGLEGPAVTVDTACSSSLVALHLAVQALRLGECSLALAGGVTVMSTPTVFVEFSRQRGLAPDGRCKAFAGAADGTGFAEGIGMLLVERLSDARRNGHPVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALASAGLSTVDVDAVEAHGTGTTLGDPIEAQALLATYGQGRDSDRPLLLGSIKSNIGHTQAAAGVAGVIKMVMAMRHGVLPQSLHIDEPTPHVDWSTGAVELLSEQTAWPEAGRPRRAGVSSFGISGTNAHLILEQAP
meridamycin-	IVIVGMSCRFPGDVESPEDLWRLIDSDGDAITAFPTDRGWDLTGLFDTAVGESGTSYARVGGFVHDAGEFDPAFFGISPREATAMDPQQRLLLHAAWEAFERAGIPAASVRGSRTGVFVGASPQGYGAAEASEGYFLTGSSGSVISGRVSYTLGLEGPAVTVDTACSSSLVALHLAVQALRSGECSLALAGGVTVMATPTAFVEFSRQRGLAADGRCKSFAAGADGTGWSEGVGLLLVERLSD

MerC	AERLGHRVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANARLSAVDVDAVEAHGTGTALGDPIEAQALLATYGQGRDVGRPLWLGSVKSNIGHTQAAAGVAGVIKMVMALRHGVLPRTLHVDEPSPHVDWSSGAVELLSERAAWPEMGRPRRAGVSSFGVSGTNAHVVLEQAP
nigericin-NigA II	VAVVGMACRFPGGVTSPEELWELLVSGVDAIGDFPTDRGWDLENLYHPDPEHYGTSCVRQGGFVEADRFDAAFFGISPREALAMDPQQRLVLELAWESLERAGIDPVSLKGTRTGVYAGVSSQDYLSRAPRIPEGFEGYATTGGLTSVISGRVAYTFGLEGPAVTVDTACSASLVAMHLAVQALRQGECALALAGGVTSLATPIMFTEFSRQRALAPDGRCKSFAADADGTGFSEGVGLVVL
	ERLSEARRNGHRVLAVIRGSAINQDGASNGLTAPNDVAQERVIGQALANAQLAPGDVDAVEAHGTGTKLGDPIEAEALIATYGQNRPADRPLLLGSLKSNIGHTHAAAGVAGVIKMVMALRHPQLPANLHLDEPTPHVDWDGSGLRLLTEPVEWPRLERPRRAAVSSFGISGTNAHLILEQAP
nigericin-NigA IV	VAIVSMACRYPGGITSPEELWELVASRGDAIGEFPTDRGWQLDGLFHPDPDHFGTSYVRHGGFLDRADGFDATFFGISPREALAADPQQRQLLEVAWELIERAGIDPTSLKGTATGVYAGAGILGFGTPQIEKSTEGFLLTGNTLSVVSGRVAFTLGLEGPAVTVDTACSSSLVAMHLACQALRQGECTLALAGGVTVMTTPNTFVEFSRQRGLAPDGRCKPFAAAADGTGFSEGVGLLLLERL
	SDAQRNGHQVLAVIRGSAINQDGASNGLTAPNGPSQQRVIRQALINAQLSSAEVDAVEAHGTGTTLGDPIEAEALLAAYGQDRTAEQPLWLGSLKSNIGHAQGAAGVAGVIKMVMALRHELLPATLYVDEPTPHADWSSGAVRLLTDSVEWPRNERPRRAGVSAFGISGTNAHLIVEQAP
nigericin-NigA VI	VAIVSMACRYPGGVGCPEELWELVASGGDGITAFPTDRGWDLDGLYHPDPDHPGTSYVRHGGFLGGTDRTDGTDRFDAEFFGISPREALAMDPQQRLLLEVSWELFERAGVDPVTTKGSRTGVYAGVSSQDYLSRMPRVPEGFEGYATTGSLTSVVSGRVAYTFGLEGPAVTVDTACSSSLVAMHLAAQALRQGECDLALAGGVTALTTPTAFAEFSRQRGLAPDGRCKSFAAGADGTGF
	SEGVGLVLLERLSDARRNGHRVLALIRGSAINQDGASNGLTAPNDVSQERVIRQALAGARLAPDQVDAVEAHGTGTSLGDPIEAHALLATYGQDRPGERPLWLGSLKSNIGHAQAAAGVAGVIKMVMALRHETLPVTLHIDEPTPHVEWEGGGVRLLTEPVPWPTGERPRRAGVSSFGISGTNAHLILEQAP
nostophycin-NpnA	IAVIGMSCRFPGANNIDEFWQNLANGVESISFFTEAEIIAAGVDPTLVKNPNYVKAKPILSDVESFDADFFGYSSREAELMDPQQRLLLECAWESLENAGYNPLTYNGAIGIYAGAVMNTYLLNNVYPNRHQLDVNDNLQVATMDSMGGLQLMVANDKDYLTTRISYKLNLTGPSVNVQTACSTSLVAIHMACANLLSGESDMVLAGGVSINAPQKVGHLYQEGMIVTPDGHCRAFDAQ
	AQGTIFGSGVGLVVLKRLQDAIADQDHIYAVVKGSATNNDGGTKVGYMAPNGDGQTAVVTEAMTMAGVDAETIGYVEAHGTGTPMGDPIEIGGLTQAFRASTQSKNFCGIGSVKTNVGHLQIASGVVGFIKTVLSLYHKQIPPSLHFEQPNPQLDLPNTPFYVNTTLKDWHTQDYPRRAGVNSLGIGGANAHVILEEAP
nostophycin-NpnB	IAIVGMGCRFPGGANNPEAFWQLLRDGVDAVTEVPANRWDIDALYDSNPTTPGKIYTRYGGFVSQLEEFDAQFFGISPREAVSLDPQQRLLLEVSWEALENASINPSQLAGTQSGVFVGISGNDYLQRLWASEATEIDAYQGTGNAHSVAAGRLSYILGLTGPSLAVDTACSSSLVAVHLACQSLRNQESDLAIAGGVNLLLSSEISINLSKARMLAPDGRCKTFDATADGYVRGEGCGVIILKR
	LSDAIKDKNQILALIRGSAVNQDGRSSGLTAPNGPSQQAVIRQALANSGVEPGEVSYLEAHGTGTSLGDPIEVGAMTAVFGKNRLKDEPLTIGSVKTNIGHLEAAAGIAGIIKVVLQMQHQEIAPHLHFQDPSPHINWENLPLIVPTQKSDWQTAEKPLLAGVSSFSFSGTNAHLVLAQP
oxazolomycin-OzmH	IAVVGVAGRFPGSADLAEFWDHLEQGRDLVTEIPGDRWDWRARTGTSRSRWGGFVPGVDRFDAAFFGISPREAELMDPQQRLLLEVVWTAVEDAGYRASDLAGRRVGVFIGTTNSDYAEVQRAGGRPAEAHTLTGAALSVIPNRISYLLDLRGPSVAVDTACSSSLTAVHQAVGALRDGTCDLAIAGGVSLILDPRLYDALSQNEMLSEDGRCKAFDASANGYVRGEGVGVVVLKDHAA
	ARADGDRVAAVIRAAAVNHGGRTTSLTSPNPDAQAELLVEAYRTAGVDPRTVGYIEAHGTGTALGDPIEITGLTEAFQRLGGDGGPGDGGGSGGGGAAPGAGRASCGIGSVKTNIGHLEAAAGIAGLLKVVLALRHRTIPASLHFRERNPYLDLDGSPFEIVGATRPWPAPLAADGTALPRRAGVSSFGFGGANAHVVIEEAP
oxazolomycin-OzmJ	IAVVGMAGQFPGAADVRAFWRNLTEGRVGYGELPAHYTAPDDPAGYRWGGALTERDHFDAEFFGIRPHEADLMSHNQRLLLQESWHALEDAAIDPTSLAGSRTGLFIGAEPTGYPHESFTGASDALVASRLSYFLDLRGAALVVNTACSSSLAAVHLACQSLSSGESTLALAGGVNAGLDARGLDLLVESGAMSPSGECRTFDADANGTVYSEAVAVLVLKRLSDAVADGDHVHGVIRATG
	MNQDGASNGITAPNGAAQEELVLDVYRRFGIPAERVGYVEAHGTGTALGDPVEANALVRAFRRLTGERGFAALGSAKAHIGHTGAPAGVVGMIKVLLSMRYGRLVGMPTLRDLNPLIDLTGSAFVADATAGPWRHPAGVPLMAAVNSFGHSGTNAHVVVEQY
oxazolomycin-OzmK	VAVIGMHCRTAGARDPEEFWQVIASGEDRTTEVDDPVALSLLKEHFPDVAAPRCGRMADSDRFDPAFFRIAPREAAAMDAAQRVLLESCYHALEDAAQDASALRGRAVATIVGSTGLAPQAEYSAHALMGADTSIMAARLAYQLDLSGPAMTVDTACSSSLVAVDIARRLLLDGESDLALAAGVYVANHPGTFVTMEALGTVSPGRACRPFDADADGMLVGEGVGVLVLKRLSDAVRD
	NDRVLGVIRGSATNQDGRTSGITSPSSAAQSALLRDVYRRSGVDMGRLRYVEAHGTGTSLGDPIEVHGLTEAFAEFTDRRRFCAVGSVKANIGHTMGAAGVLGAIKVLLCLRHGQLPPAANFRTENKHADFDESPVYVSREPADWNREGEEPRVAGVSSFGYSGTNAHVIIEEY
oxazolomycin-OzmN	VAVIGLSGRYPMADDVDEFWANLAAGRDCVTEIPADRWDHDRYFDADPSAPGRASTRWGGFLRDVDRFDPLFFGISPREAELMDPQERLFLQNAWHVFEDAGYRRDELSGRPVGVFVGVMYGEYQLYGAADAERGGIRVTGSSFASIANRVSYTLGLNGPSIALDTMCSSSLTAIHLACESLRTGESEAAVAGGVNLSLHPYKYVFLSQGRFVSADGRCRAFGAGGTGYVPGEGVGAVL
	LKPLAAAERDGDHIYGVIRGHAVNHGGRTNGYTVPSPHAQADAVRRALRGAGLEPADIGYVEAHGTGTSLGDPIEIAGLAKAFGTPKEDGPRWPIGSVKSNIGHLESAAGIAGLTKVLLQFRHRRLVPSLHADEPNPNIDFGRAPFRVQREAADWPEPAGADGAVPRRAGLSSFGAGGSNAHLVVEEY
oxazolomycin-OzmQ	IAVIGLACRFPGAATPDTFWKVLSEGRETLTHFSDEELRAAGVAEPLLADDRYVKAGQVLVDADKFDAGLFGITRDEAELIDPQQRQFLECAYEALERAGYDPQRGEQRIGVYAGVGLNTYLLHNLGERYRTASSVDRYRMMITNDKDFVATRTAYKLNLCGPSVSTNTACSTSLVAVHLACLSLLSGDCTMALAGAAHIQADQGEGYLHHEGMIFSPDGHCRAFDAKAQGTVIGNGVGAVV
	LKRLSDALADGDTVHAVIKGTAVNNDGSDKTGYTAPSVQGQAAVVAEAQEIADVGPETVSYVEAHGTATPLGDPIEVAALNQAFNREGAALAPGSCALGSVKTNVGHLDTAAGMAGLIKTILMLRHRTLVPSLCFEAPNPEIDFAAGPFYVGTETKEWPAGPTPRRAGVSSFGIGGTNAHVIVEEP
patellazole-PtzA	IAIIGMAGRYPDAPNLEQYWKNLRQGVCSLHDVPPERWRVEDFTDPSRPDRTTVRKGGFITDVDRFDSMFFGITPKEAKLMDPQHRLFLEVAWQAIEDAGYTPAELKRSAGNGAVGDVGVIVGAMNQPYRLVGLPATASGKVVQNGHWSIANRLSYHLDLTGVSMAVDTACSASLTAVHLACAALTQGDCGVVVVGGVNLILHPLQSLELARIGMLSPSNLCRPFAEGADGTIQGEGIG
	AMVLKPLDRALADGDTIRSIILGSAIDAGGRTSGYTVPNPNAQRRVIQAAMAQARIDPASIAHVECHGTGTDLGDPIEVNALADALKDRHEKILIGSVKGNIGHLESAAGIAGLSKLIMEIEAGEVVPSLGAENLSKRIDFTGQKLEVATTLRSWPCRFDVAGKIQPRRGGISSFGAGGANAHLVIEQAP
pyxipyrrolone-PyxA	VAIIGMAGALPGARDLDELWRRLRAGEDLISEAPLERWDWRAWGEDDGAFLKWGGFAPGFDCFDAPFFNVAPREAELMDPQQRMLLQTTWKALEDAGHPPATLSGRNVGVFMGVWAQDYHHRLSHHGLVHAQVETGNALTMLGNRISYVFNLRGPSEVFNAACASSLVALHRAAQALRSGECELAIVGGVNALVLPEVTRNLAQVGILSRNGRCRTFDASANGYVRGEGAVVVV
	LKRLEQAEADADFIHAVVRGSRQNHGGRASSLTAPNPKAQAELIFGALDEAGFEPETVTYIEAHGTGTELGDPVEVDGLKSAFEQLARKRQRALEGTQFCGLGSIKSNVGHLESAAGLAGLVKVVLAMRHGELPASLHVQKLNPLIRLEGSPFYVVRRTQPWARVEDAEGRPVPRRAGVSSFGFGGSNAHVALEEY
pyxipyrrolone-PyxC	IAIIGVSGRYPMADDLWDFWDNLKAGRNCITEIPVGRWEHHPYFDRDKDAEGRSYSKWGGFLSDVDRFDALFFNISTREAEGMDPQERLFLQTVWSLFEDAGYTRQSLAPAAARTGVFVGVMYGQYEWLGVEAALRGHVSRASSGRWSIANRISYFFNFQGPSLAVDTACSSSLTAIHLACQALRAGECEQAVAGGVNVSIHPHKYTLLSQGRFASSDGLCRSFGAGGDGYVPGEGIGA
	VLLKPLRSALADGDRVYAVVKGSALNHGGRTNGYTVPSPHAQGEVISEALRRARVEAKELGYIEAHGTGTSLGDPIEVTGLSKAFGASGPAAGRCAIGSVKSNIGHLEAAAGIAGVTKVLLQFQHGQLVPSLHAEPPNPNIRFADTPFVVQKELAAWPRPVAGEPRRAAVSSFGAGGANAHVVLEEF
pyxipyrrolone-PyxD	VAIIGMAGRFPGAQNLDEYWSNLARGVDSVTEVPPERWSVAKHFDPDPKKKGTSYSKWGGFLTDIDKFDPLFFSVSPAEARLMDPQQRLFLEQAWKAFEDAGYSPEQLDQARCGVFVGLMSNDYARMILAAEPDRSPSLQMMGNSSSILAARIAYLLNLKGPALCLDTACSSSLVATHLAVQSLLAGETDLMLAGGVTLYLDEESYIQMSKAGMLSPEGRCKTFDDRADGFVPGEGVGA
	VVLKRLSDALRDGDHVYGIIRGSGLNQDGKTNGISAPSSEAQMRLELEVYQRSGISPDSLTYVEAHGTGTRLGDPIEIEALSEAFARYTDRKQFCAIGSVKSNLGHTSAAAGMAGLMKVLLSFRHRQLPPSLHFHRQNSHIDFGSTPFFVNTALRDWVVPGHGPRRAAISGFSFSGTNAHLVLEEP
rhizopodin-RizB	IAIVGAAAVLPGSETLGAFWRNIEEGKDLIREIPAERWDWRTLRGDTSGDVDAESLRWGGFIDGVDQFDGRFFGISRREGELMDPQQRIFLQTVWRALEDAGWDPASLSGTRTGLFVGVCTSDYAELLARSGGQVEAHTSTGNARSILPNRISHLLGLRGPSEPIDTACSSSLIAVQHAIHALRSGAADVAIAGGVNLVLSPFFHLSFSHAGMLSPGGRCRTFDHRADGYVRGEGSAAVVLKP
	LSRAEADGDTIHGLIVGGAVGHAGHANSLTAPNGASQAELLVEAYSNAGIDVRTVGYIEAHGTATRLGDPIEIHALQRAFRELSRDASTVAPWCGLGSVKSNFGHLEAAAGVAGLLKVLLAMRAQTLPATLHFEQLNPHIDLRDSPFFIVERNTPWRVHGDAPRRAGVSSFGFGGVNAHLVLEEY

rhizopodin-RizD	VAIVGVAGRYPMAEDLDQLWENLAAGRDCIIEIPAQRWDHSRYFDSRRGQPGKTYCKWGGFIEDVDCFDPLFFGITPRDAEWIDPQERLFLETAWSTVEDAGYTREALKRSTGDRVGVFVGVMYGEYQLFGDAAGSTPVGLAYGSIANRVSYCLDVSGPSLAVDTLCSSSLTAIHLAVESLRRGECAAAIAGGVNLSLHPSKYLLHAQAQMSASDGRCRSFGAGGDGFAPGEGVGAVLLKSLTQAEADGDHIYAIIRSTAVNHGGRTSGYTVPNPQAQAQLIQQALTVGRIDPASVSYIEAHGTGTSLGDPIEAAGLIQALSSNGRDRCAIGSIKSNIGHLESAAGIAALTKVLLQMRHGRIAPSLHSRVVNPNIDFSGTPFYIPQELTAWETGLDGAGRPLPRRAGISSFGAGGANAHLILEEY
rhizopodin-RizE	VAVIGMSGRYPGARDVRGFYANLKAGRRALSELPGERGSWSRWFDPQPERAAEGRIYRRTGGFISGVELFDPLAFQMTPAEAKGIHPEERLLLQAAWEAIEDAGYAHTALEGKRHGVFIGVNGLSYPLLGLERWAQRGDVLLDQSYFGLPNRISHFFDLRGPSVPFDTGCSSALVALDAARRAIAQGECDGALVGGVNLYLHVSRFSTLCQARLASTHEAPLLFERGGDGFVPGEGVGLVLLKPLDDALRDGDHVYGVIRATTVSHKGRSSSYLLPSPPAQRALIERALTEARVSPSDIGYVELQAVGAEMTDLSEWRSLSAVFGGVEAGARCALGSVKPLIGHLEAASGMAQLTRVLLQLDHGELFPTPVAVNMHEGISVDTAPFYVPAKSGPWSHPRRARGEGWTPRRALLGSAAAGGTQVFTVLEEC
rhizoxin-RhiB	IAIIGLAGRFPGAADIDSFWNNLKNGVDSISEIPAERWEHSKYYSRDRSVKGKIHTKWGGFIDDIDRFDPLFFNISPRDAERMDPQERIFLETAWRTLEDAGYDRAGLRRHYQGEMGVFVGVMHGEYLLYTRSSAPDCTEDAVDSSFGSIANRVSYVFDCNGPSMAVDTICSSSLTALHLAVESLRRGECQLALAGGVNLSLHPNKYFIQSQLTMSSSDGRCRSFGDGGDGFVPGEGVGAVLLKPLSRAEADGDNIYAVIRATSINHNGKTHGYTVPSPNAQGAMIADSLDKAGMDPRELSYIEAHGTGTALGDPIEITGLQLGFTRRSATLGSAMASAQQCAVGSVKSNIGHLESAAGIAGIAKVLLQLKHRQLAPSLHSKRLNPNIPFADSPFYIPQQLQEWRRPVLNLDGREREYPRAASVSSFGSGGSNGHVIIAEY
rhizoxin-RhiD	IAVIGMAGQFPAADNVERFWRNLSTGVDGIVELGERYLDARRSYSSQRQPGKTYCKWGGVLEQRDDFDPLFFNISPRDAESMNPHQRLILQESWKALEDAGYNPKSFRNQAVGIFVGAEPTGYFHETFIGASDAIIASRLSYFLDFKGPALVVNTGCSSSAVAIHLACENLRNRETDLALAGGVFATLGEPSLISLSQMDMLSPTGRCHTFDSRSDGIVLSEAVGMVTLKRLPEAIADGDPIYGVICASGMNQDGASNGITAPSGQAQERLISELYRRFGIDPRQISYVEAHGTGTPLGDPVEVNALARAFRQFTADTGYCTLGSAKAHIGHTSAAAGVVSLIKILLSMKYRQLPGLLNFEQLNPRIALADSAFTIHRELKDWPVGDSGLRMAALNSFGHSGTNVHIVVREA
rhizoxin-RhiE	IAIIGISGRFPGARNADELWRVLAEGRDMVGEIPADRYDWRDYYGDHREDPSKTRCIWMGSVPGVSEFDPMFFEISPREAKTLDPRQRLLLQESWNALEDAGVGPRHLETQRVGMFVGVEEGDFQRLASEESLTSGHNGILASRLAYFLNLTGPTMAINTACSSALVALHQAALSLRNNECEIAVVAGVNLIFSPDAFIGMTQAGMLSPDGKCFAFDRRANGMVPGEAVCCVVLKPLSRAEADGDPIYGVLRGSGINYDGKTNGITAPSGASQQRLLEEVYQRAGMNTEAVEYIVTHGTGTQLGDPVEINALNDAFRGNKPGHCALTSTKTNLGHTFAASGLVSVISLLQAFRHETIPASLHCEQENDYIHWQQSPFYVNKTNKAWPAAGRDSERLGAVSAFGMSGTNAHVVLESY
rifamycin-RifA	IAIVGMACRLPGGVASPEDLWRLVAERVDAVSEFPGDRGWDLDSLIDPDRERAGTSYVGQGGFLHDAGEFDAGFFGISPREAVAMDPQQRLLLETSWEALENAGVDPIALKGTDTGVFSGLMGQGYGSGAVAPELEGFVTTGVASSVASGRVSYVLGLEGPAVTVDTACSSSLVAMHLAAQALRQGECSMALAGGVTVMATPGSFVEFSRQRALAPDGRCKAFAAAADGTGWSEGVGVVVLERLSVARERGHRILAVLRGSAVNQDGASNGLTAPNGLSQQRVIRRALAAAGLAPSDVDVVEAHGTGTTLGDPIEAQALIATYGQDRETPLWLGSLKSNIGHTQAAAGVASVIKVVQALRHGVMPPTLHVDEPSSQVDWSEGAVELLTGSRDWPRGDRPRRAGVSSFGVSGTNVHLIIEEAP
rifamycin-RifC	IAIVGMACRLPGGVTGPGDLWRLVAEGGDAVSGFPTDRCWDLDTLFDPDPDHAGTSYTDQGGFLHDAALFDPGFFGISPREALAMDPQQRLLLEASWEALEGVGLDPASLQGTDVGVFTGAGGSGYGGGLTGPEMQSFAGTGLASSVASGRVSYVFGFEGPAVTIDTACSSSLVAMHLAAQALRQGDCSMALAGGAMVMSGPDSFVVFSRQRGLATDGRCKAFASGADGMVLAEGISVVVLERLSVARERGHRVLAVLRGSAVNQDGASNGLTAPNGPSQQRVIRAALANAGIGPSDVDLVEAHGTGTSLGDPIEAQALLATYGQDRETPLWLGSLKSNIGHTQAAAGVASVIKVVQALRHGVMPPTLHVDEPSSQVDWSEGAVELLTGSRDWPRGDRPRRAGVSSFGVSGTNVHLIIEEAP
rifamycin-RifE	AVVGMACRFPGGVSSPEDLWQLVAGGVDALSDFPDDRGWELDGLFDPDPDHPGTSYTSQGGFLRGAGLFDAGLFDAGLFGISPREALVMDPQQRVLLETSWEALEDAGVDPLSLKGSDVGVFSGVFTQGYGAGAITPDLEAFAGIGAASSVASGRVSYVFGLEGPAVTIDTACSSSLVAIHLAAQALRAGECSMALAGGATVMPTPGTFVAFSRQRVLAADGRSKAFSSTADGTGWAEGAGVLVLERLSVAQERGHRILAVLRGSAVNQDGASNGLTAPNGPSQQRVIRKALAGAGLVASDVDVVEAHGTGTALGDPIEAQALLATYGQGRERPLWLGSVKSNFGHTQAAAGVAGVIKMVQALRHGAMPPTLHVAEPTPEVDWSAGAVELLTEPREWPAGDRPRRAGVSAFGISGTNAHLILEEAP
tallysomycin-TlmVIII	IAIVSLAGRFPGADDVPTFWENLRAGEESIRRYTREELLALGRDPALVDHPRFVGAEGVLGDIAGFDADFFGYSPREAEIMDPQHRLCLETAWAAFDSAGYDPAALTEPVGVFLATSLSSYLVRNLLSDPGLARGIGGFPLLIHNDKDFAATTVSHKLGLTGPSYAVGSACSSSLVAVHLACQSLQSHECDMALAGGVSLQVPQGQGYVYAEDGIYSPDGRCAAFDASAGGTVGGSGVGLVLLKRLSDAVRDRDQVHALVLGTAVNNDGANKVGYTAPSVDGQSAVIVEAHAVAGISADTVGYVEAHGTGTSLGDPIEVTALTQAFRASTDRTGFCALGSVKTNIGHLDAAAGIAGLVKAALAVRDGIIPASLNFREPNPRINFAATPFHVADTTRPWRAEGAPRRAGVSSFGIGGTNVHVVLEQP
tautomycetin-tmcA	IVLVGMACRFPGGVSDPDGLWRLVAEEADATGPFPTDRGWDLDRLQEVSATSRGGFVAGVDGFDAAFFRISPREALATDPQQRLLLEVSWEALEQAGIDPGTLAGTPTGVFAGAYGSGYGDLVSREQLQGHLLTGGAGSVISGRVAYALGLEGPAVSVDTACSSSLVAMHLAAQALRTGECNLALAGGVTVMATPEMFLEFTAQNGLAEDGRCRSFADSADGTGWSEGVGVVVVMERLSDAVRNGHEVLAVMRSSAVNQDGASNGLTAPNGPSQQRVIRQALAAAGLSSADVDAVEAHGTGTRLGDPIEAQALLATYGQDRPGDRPLWLGSLKSNIGHAQAAAGVGGVIKMVMALRHGVLPRTLHVDAPSTQVDWTQGDVRLLTDAVPWPETGRPRRAGVSSFGVSGTNAHVILEA
tautomycetin-tmcB	IVLVGMACRFPGGVADPDDLWQLVAGGIDGVTAFPADRGWDLDALLGADGSAASATAEGGFLAGAGEFDAAFFRISPREALATDPQQRLLLEVSWEALERAGIDPGTLAGSPTGVFAGAFGSGYGELVARSGEQLQGHLITGTAGSAISGRVAYALGLEGPAVSVDTACSSSLVAMHLAAQALRAGECSLALAGGVTVMATADAFVGFTAQGGLSPDGRCRSFADSADGTGWSEGVGVVVLERLSDAVRNGHEVLAVMRSSAVNQDGASNGLTAPNGPSQQRVIRQALAAAGLSSADVDAVEAHGTGTRLGDPIEAQALLSTYGQDRPGDRPLWLGSLKSNIGHAQAAAGVGGVIKMVMALRHGVLPRTLHVDEPSTQVDWTQGDVRLLTDAVPWPETGRPRRAGVSSFGVSGTNAHVILEA
tiancimycin-TnmE	IAVVGMACRYPDADDPHQLWQTVLAGRRAFRRIPPERLRVADYQEGPGGPDAPDGIHARQAAVLENWTFDRSAFRVPGPAYRATDLTHWLALDVAAAALRDAGLPDGDGLDRDRAGVVIGNSLTGEFSRAGLVRLRWPYVRRTLDQVLRDRGTAPAERRELLQALEEAYKAPFPAPGDETLAGGLANTISGRVCNHFDFHGAGYTVDGACASSLISVATACAALAQGDLDFALAGGVDLSLDPFELVGFSRLGALARGTMRVYDRNPTGFLPGEGCGMVALAREDDAVRMGLRVHARIEGWGLASDGHGGLTRPERTGQALALRRAYARAGFGPERTGLVEGHGTGTAVGDDVELAVLDAELRAAGAGGVRPPALGSIKANIGHTKAAAGVAGLIKAVLAVLRDGVLPPTTGCPDPHPLLGADGAALRTLAEPEDWPDRERHASVNSLGFGGINAHVVVG
xenocoumacin-XcnF	IAVVGISGRFPGSHNVREYWKNILSGKEGISFFTDDELIDAGIPEYIFKNKHYVRAKGIVENALDFDKTAFGYTMREAQYMDPQFRLLHEITWEVLDDAGYGNPKYRPVTGLFASSGSNYEWLDRVKEEVKDTTEQFSIGLLNDREFLTTRIAYKLNLTGPAVTVQSACSSSALAITLACQSLRVNQCDLALAGGVSITVPVKSGYMYVNGMVNSADGHCKPFSEDASGTVFSDGIGMVALKRLEDAIENNDHIYGVIAGYGVNNDGYEKAGFTAPSANGQAACIKQSLAMAGISANDVGYLEAHGTATLVGDDIELEGVKSALSVADDNPSICHIGSVKSNLGHTDSAAGVAGFIKTLLTVKHKKIPPTCHTEKNDHGLLQGTRFRLINAATDWDNNNDIRTAGLSSFGIGGTNVHITIQS
xenocoumacin-XcnH	IAIVGMACNFPDATTPQMFWQNLISGKESIRKFTHEELIKNGYKEEEINDPNFVPAKAILKDYKSFDNNVFGYLKDEVDKMDPQCRLLHECCWHALEDAAYVPGETRDRVGLFAGTSFNLPWISNQLKQKLTPAEIFDVVNWSLPESVITRIAYRLGLNGPVMSVHTACSTSLVAIHTACQSILSGDCDLALAGGVSISLPQESGYTYQKGMIRSADGHCRPFDADSGGTVGGNGVGILLLKLLDRALEDNDHIHAIIKGTSVNNDGQRKVGFTAPSIEGQIEVISNAQMLAGVEPEDISYIECHGTATEIGDPIEVTALARVFQHSQHTCWLGAVKSNIGHLDEAAGIAGIIKTVMSLKNKQFAPTLHYKKPNDKLYLDETPFKVVDRPMEWAVPEGKKRMAGVSSFGIGGTNAHIVLAEA
xenocoumacin-XcnL	IAIVGVSCQFPNSPTVESFWQNCLEGNDCITRFSDETLAENGVPENEYKSQNYVNAKGVIENPLKFDAEFFQYNKREAELLEPQIRHLHQCSFDALGSAGIAPSVYKGKIGCYLAASAQLQWQMAAYEESKQDAVGLFSAFSLIEKDFAATRLAYKLGLTGPAVALQTACSSSLTAVHMAARALLLGECQMALVAAAALTLPNQQGYLYLDGMIESPDGMCRPFDEHANGTVGGEGIGAIVLQPLRQALEEKRPIIALIKGSAINNDGARKVGYTAPSIEGQTDVIRSAITVSKVDPKSISYIEAHGTATKLGDPIEMAALKNIFDDCPAGSIAIGSAKSILGHLDSAAGMAGLIKTALMLKNKMLCANKYFNAPNPKLKLESSPFYINTQSQEWHSQSIRRAGVSSFGIGGTNAHVILEEM

yersiniabactin-HMWP1	VAVIGYACHFPESPDGETFWQNLLEGRECSRRFTREELLAVGLDAAIIDDPHYVNIGTVLDNADCFDATLFGYSRQEAESMDPQQRLFLQAVWHALEHAGYAPGAVPHKTGVFASSRMSTYPGREALNVTEVAQVKGLQSLMGNDKDYIATRAAYKLNLHGPALSVQTACSSSLVAVHLACESLRAGESDMAVAGGVALSFPQQAGYRYQPGMIFSPDGHCRPFDASAEGTWAGNGLGCVVLRRLRDALLSGDPIISVILSSAVNNDGNRKVGYTAPSVAGQQAVIEEALMLAAIDDRQVGYIETHGTGTPLGDAIEIEALRNVYAPRPQDQRCALGSVKSNMGHLDTAAGIAGLLKTVLAVSRGQIPPLLNFHTPNPALKLEESPFTIPVSAQAWQDEMRYAGVSSFGIGGTNCHMIVASL
zwittermicin-ZmaA1	IAIIGMAGKFPGANEINQFWDNLKNGVDSISTFTEDELIKEGVNLDVLQHPNFVKAKGYLEEVEYFDESFFAYTPREAKIMDPQIRMLQETTWEALEMAGYNPFDYEGLIGLYVGASTNFNWMKHTPLLNSDSVVEFSEAGTLSYKDAISTLTSYKLGLKGPSFTLYTACSTSLLSIHLACRSLLTGECSIAAAGGVSITYPKKNGYKYHEGMTSSPDGKVRTFDAEAQGAVFSDGVGMVILKRLEDAIADGDTIYGVIKGSAANNDGGRKVGYTAPSVEGQAEVIQAAHSFAEVDPSTISYVETHGTATPLGDSIEVEALKRAFQDVDKKSFCAIGSVKSNVGHLDTAAGVTGLIKTVLSMRHKQLPPTLNVKRPNSKIDFIDSPFYINTDLVNWQSDDNDLLRAGVSAFGYGGTNVHIVLEEAP
zwittermicin-ZmaA2	IAIIGIDGRFPGAQNVNEFWNNIKSGTESIQFFTDEELIESGVNPMEVKSPNYVKAKGYLEGTDNFDAPFFDYTPQDASLMDPQLRVFHECAWSALEHAGYNIETYPGLIGVYSGASPNLYWQVLSTLSEANEPAGQFLISLLNDKDSLSTQISYKFNLKGPSMNIFTGCSTSLVAIHNACQALLQGHCDIAIAGGITLTQPEKAGYTYQEGMLFSSDGHCRPFDENANGMLFGDGVGIVVLKPLQEAINDGDTIHAVIKGTAINNDGNRKIGYTAPSVEGQVEVIKMAQHEANVEPESISYIETHGTATKLGDTIEIKALSEVFNSNEKQSVPIGSVKANVGHLNAASGVAGLIKTVFAMKDQVLPPSVNFTKPNTQIGFEKTPFYVNQQLNEWKEDNKPLRAGVSSFGIGGTNAHIILEEAP
zwittermicin-ZmaK	VAVVGMAVRMPGASNLKQFWSNLEQGIESIRTFTDEELSEMGIPKEVIQKPNYIKSKGYLEDVDAFDNEFFDYSLKEAEMMDPQLRILHECAWEALEQAGEVNSSHMNTGVYVGGSPNFHWLRSIASESSNTLEDFQAMLLNEKDFFATRLAYKLNLKGPAITVQTACSTSLVAIQNAWADLIDGRCDIALAGGVSITYPTKTGYLYEDGMIFSPDGHCRAFDKDAQGTVGGNGAGIVVLKRLEDALRDGNLIHGVIKGAAVNNDGSNKAGFTAPGVDGQAAVIKSAHEIAGVKPEEISYIETHGTGTQLGDPIEVEALKLAFGKQEPGRVLIGSVKTNIGHLDAAAGVGGFIKTVLALKNKKIPPSLHYNNPNPSIDFGKNPFRVNTELIPWKEEVRRAGVSSFGMGGTNAHIVLESV

5. Supplementary References

1. Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci. USA* **97**, 6640–6645 (2000).

2. Li, Z.-R. et al. Critical intermediates reveal new biosynthetic events in the enigmatic colibactin pathway. *ChemBioChem* **16**, 1715–1719 (2015).

3. Gust, B., Challis, G. L., Fowler, K., Kieser, T. & Chater, K. F. PCR-targeted *Streptomyces* gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin. *Proc. Natl. Acad. Sci. USA* **100**, 1541–1546 (2003).

4. Welch, R. A. et al. Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **99**, 17020–17024 (2002).

5. Li, Z.-R. et al. Divergent biosynthesis yields a cytotoxic aminomalonate-containing precolibactin. *Nat. Chem. Biol.* **12**, 773–775 (2016).

6. Brotherton, C. A., Wilson, M., Byrd, G. & Balskus, E. P. Isolation of a metabolite from the *pks* island provides insights into colibactin biosynthesis and activity. *Org. Lett.* **17**, 1545–1548 (2015).

7. Nougayrède, J. P. et al. *Escherichia coli* induces DNA double-strand breaks in eukaryotic cells. *Science* **313**, 848–851 (2006).

8. Brachmann, A. O. et al. Colibactin biosynthesis and biological activity depends on the rare aminomalonyl polyketide precursor. *Chem. Commun.* **51**, 13138–13141 (2015).

9. Zha, L., Wilson, M. R., Brotherton, C. A. & Balskus, E. P. Characterization of polyketide synthase machinery from the *pks* island facilitates isolation of a candidate precolibactin. *ACS Chem. Biol.* **11**, 1287–1295 (2016).

10. Kumar, S., Stecher, G. & Tamura, K. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **33**, 1870–1874 (2016).

11. Balbás, P. et al. Plasmid vector pBR322 and its special-purpose derivatives — a review. *Gene* **50**, 3–40 (1986).

12. Colis, L. C. et al. The cytotoxicity of (–)-lomaiviticin A arises from induction of double-strand breaks in DNA. *Nat. Chem.* **6**, 504–510 (2014).

13. Rahman, A. et al. Strand scission in DNA induced by dietary flavonoids: role of Cu(I) and oxygen free radicals and biological consequences of scission. *Mol. Cell. Biochem.* **111**, 3–9 (1992).

14. Junglee, S., Urban, L., Sallanon, H. & Lopez-Lauri, F. Optimized assay for hydrogen peroxide determination in plant tissue using potassium iodide. *Am. J. Anal. Chem.* **5**, 730–736 (2014).

15. Repine, J. E., Pfenninger, O. W., Talmage, D. W., Berger, E. M. & Pettijohn, D. E. Dimethyl sulfoxide prevents DNA nicking mediated by ionizing radiation or iron/hydrogen peroxide-generated hydroxyl radical. *Proc. Natl. Acad. Sci. USA* **78**, 1001–1003 (1981).

16. Freifelder, D. & Trumbo, B. Matching of single-strand breaks to form double-strand breaks in DNA. *Biopolymers* **7**, 681–693 (1969).

17. Chaturvedi, K. S., Hung, C. S., Crowley, J. R., Stapleton, A. E. & Henderson, J. P. The siderophore yersiniabactin binds copper to protect pathogens during infection. *Nat. Chem. Biol.* **8**, 731–736 (2012).

18. Azam, S., Hadi, N., Khan, N. U. & Hadi, S. M. Antioxidant and prooxidant properties of caffeine, theobromine and xanthine. *Med. Sci. Monit.* **9**, 325–330 (1969).

19. Chu, T., Hu, Y., Wu, J., Zeng, C., Yang, Y. & Ng, S. W. A new luminescent lanthanide supramolecular network possessing free Lewis base sites for highly selective and sensitive Cu^{2+} sensing. *Photochem. Photobiol. Sci.* **15**, 744–751 (2016).

20. Benesi, H. & Hildebrand, J. A spectrophotometric investigation of the interaction of iodine with aromatic hydrocarbons. *J. Am. Chem. Soc.* **71**, 2703–2707 (1949).

21. Zaengle-Barone, J. M. et al. Copper influences the antibacterial outcomes of a *β*-lactamase-activated prochelator against drug-resistant bacteria. *ACS Infect. Dis.* **4**, 1019–1029 (2018).

22. Healy, A. R., Vizcaino, M. I., Crawford, J. M. & Herzon, S. B. Convergent and modular synthesis of candidate precolibactins. structural revision of precolibactin A. *J. Am. Chem. Soc.* **138**, 5426–5432 (2016).

23. Melvin, M. S. et al. Double-strand DNA cleavage by copper·prodigiosin. *J. Am. Chem. Soc.* **122**, 6333–6334 (2000).

24. Cuevas-Ramos, G. et al. *Escherichia coli* induces DNA damage in vivo and triggers genomic instability in mammalian cells. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 11537–11542 (2010).

25. Stern, B. R. et al. Copper and human health: biochemistry, genetics, and strategies for modeling dose-response relationships. *J. Toxicol. Env. Heal. B* **10**, 157–222 (2007).

26. Pitié, M. & Pratviel, G. Activation of DNA carbon-hydrogen bonds by metal complexes. *Chem. Rev.* **110**, 1018–1059 (2010).

27. Brotherton, C. A. & Balskus, E. P. A prodrug resistance mechanism is involved in colibactin biosynthesis and cytotoxicity. *J. Am. Chem. Soc.* **135**, 3359–3362 (2013).

28. Wagner, J. K., Marquis, K. A. & Rudner, D. Z. SirA enforces diploidy by inhibiting the replication initiator DnaA during spore formation in *Bacillus subtilis*. *Mol. Microbiol.* **73**, 963–974 (2009).