Angewandte Chemie
Eine Zeitschrift der Gesellschaft Deutscher Chemiker

Supporting Information

Alternative Benzoxazole Assembly Discovered in Anaerobic Bacteria Provides Access to Privileged Heterocyclic Scaffold

*T. Horch, E. M. Molloy, F. Bredy, V. G. Haensch, K. Scherlach, K. L. Dunbar, J. Franke, C. Hertweck**

SUPPORTING INFORMATION

Experimental Procedures

Bacterial strains and culturing conditions

Escherichia coli strains were grown at 37 °C in Luria-Bertani medium (10 g BD Bacto™ Tryptone, 5 g BD Bacto™ yeast extract, 10 g NaCl, 1000 mL water; deionized water was used for all media and components described) with agitation or on LB agar plates. The following antibiotics were supplemented as appropriate: 100 µg mL^{-1} kanamycin, ampicillin 50 µg mL^{-1}, 100 µg mL^{-1} ampicillin, 10 µg mL^{-1} gentamycin. Plasmid construction and storage was performed with *E. coli* Top10 or XL-1 Blue, while *E. coli* BL21 (DE3) was used for heterologous expression experiments. All generated *E. coli* strains were stored long-term in 10 % (*v/v*) glycerol (final concentration) at –80 °C. *Clostridium cavendishii* DSM 21758 was obtained from the Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures GmbH and maintained under an anaerobic atmosphere (N_2:H_2:CO_2, 85:5:10 vol:vol:vol) in a Whitley A35 anaerobic work station (Don Whitley Scientific) operating at 37 °C or 30 °C for 68–72 hours. Routine cultivation of the *C. cavendishii* wild-type and Rif1 strains was performed in VM Medium (modified from original recipe[1]: 1 g KH_2PO_4, 4.5 g $K_2HPO_4 \cdot 3H_2O$, 2.1 g urea, 10 g MOPS, 2 g yeast extract, $MgCl_2 \cdot 6H_2O$ (5 mM), $CaCl_2 \cdot 2H_2O$ (1 mM), 1.25 mg mL^{-1} $FeSO_4 \cdot 7H_2O$, 10 mL vitamin solution, 1 mL trace element solution, 1 g L^{-1} cysteine-HCl, 5.1 g cellobiose. Vitamin Solution: 250 mg p-aminobenzoic acid, 250 mg thiamine, 100 mg D-biotin, 150 mg nicotinic acid, 250 mg riboflavin, 100 mg cyanocobalamine, 270 mg calcium pantothenate, 1000 mL water. Trace element solution: 50 mL 10 M HCl, 5 g $FeSO_4$ ($7 \cdot H_2O$), 1.44 g $ZnSO_4$ ($7 \cdot H_2O$), 1.12 g $MnSO_4 \cdot 7H_2O$, 30 mg H_3BO_3, 20 mg $CoCl_2 \cdot 6H_2O$, 40 mg $NiCl_2 \cdot 6H_2O$, 20 mg Na_2SeO_4, 200 mg $Na_2B_4O_7$ ($10 \cdot H_2O$), 1 g $(NH_4)_6Mo_7O_{24}$ ($4 \cdot H_2O$), 1000 mL water. Transformation procedure was performed in modified GS-2 medium.[2] To allow for gas exchange, all cultures were grown in unsealed test-tubes or glass bottles with a loosened lid.

Experimental procedures with *C. cavendishii* strains

Variation of culture conditions

In order to generate metabolite profiles of *C. cavendishii* under various conditions, the strain was cultivated in rich medium (TGY medium: tryptone 5.0 g, yeast extract 3.0 g, D-glucose 1.0 g, 1000 mL water, pH 7.0), minimal medium (P2 medium: 20 g D-glucose, 0.5 g H_2HPO_4, 0.5 g KH_2PO_4, 2.2g CH_3COONH_4, 2.0 g $MgSO_4 \cdot 7H_20$, 0.1 g $MnSO_4 \cdot H_2O$, 0.1g NaCl, 0.1 g $FeSO_4 \cdot 7H_20$, 100 mg paraaminobenzoic acid, 100 mg thiamine, 1 mg biotin, 1000 mL water, pH 7.0) and semi-defined medium (VM and CM3[1, 3]). Best closoxazole production was obtained when cultivated in VM medium, which was used for all further experiments.

General culture extractions for metabolic profiling

For metabolic profiling of *C. cavendishii*, cultures were extracted once with 1:1 (*v/v*) ethyl acetate, then the extract was dried over Na_2SO_4 and filtered through filter paper (Rotilabo Typ 600P). Upon concentration under reduced pressure, the extracts were dissolved in methanol (500x concentration) and subjected to analytical LC-MS.

Selection of the rifampicin resistant mutant C. cavendishii Rif1

In an attempt to enhance closoxazole production by *C. cavendishii*, we took a pleiotropic approach that involves selecting for mutations in genes encoding RNA polymerase (RNAP), a strategy with previous success in upregulating the translation of biosynthetic gene clusters.[4] *C. cavendishii* was cultivated overnight in VM broth, then 100 µL volumes were spread on VM agar plates containing 2-fold serial dilutions of rifampicin (concentration range from 4 µg mL^{-1} to 0.125 µg mL^{-1}), an antibiotic that targets the RNAP β-subunit. The minimum inhibitory concentration (MIC) of rifampicin against *C. cavendishii* under these conditions was 0.5 µg mL^{-1}. Based on this result, *C. cavendishii* was again cultivated overnight in VM broth, then 100 µL volumes were spread on VM agar plates containing 2x, 5x, and 10x MIC of rifampicin (1 µg mL^{-1}, 2.5 µg mL^{-1}, and 5 µg mL^{-1}, respectively) to select for spontaneous resistant mutants. Eleven colonies grew at a concentration of 1 µg mL^{-1} rifampicin, while none emerged at 2.5 µg mL^{-1} and 5 µg mL^{-1}. A T-streak of each resistant colony was performed on VM agar containing 1 µg mL^{-1} rifampicin, then an isolated colony of each strain was transferred to VM broth containing 1 µg mL^{-1} rifampicin for cultivation. The strains were stocked and arbitrarily named Rif1 – Rif11. *C. cavendishii* Rif1 – Rif11 were subjected to metabolic profiling under the established closoxazole production conditions. Compared to the wild type, *C. cavendishii* Rif1 displayed a modest increase in yield. Thus, subsequent scale-up and knock-out experiments used *C. cavendishii* Rif1. The vast majority of mutations induced by rifampicin have been found in the so-called rifampicin resistance-determining region (RRDR) of the *rpoB* gene.[5] We therefore performed a colony PCR using primers Ccav_rpoB_F and Ccav_rpoB_R (Table S3) to amplify the equivalent region of the *C. cavendishii* Rif1 genome, then sequenced the PCR product. The wild-type sequence was maintained in the *C. cavendishii* Rif1 RRDR, meaning that the reason for the increased rifampicin resistance and enhanced closoxazole production is unknown.

Plasmid construction for the generation of C. cavendishii Rif1 ΔclxA

For the creation of the in-frame nonsense mutation of *clxA* in the genome of *C. cavendishii* Rif1, we generated a CRISPR-based knockout plasmid as previously described with minor changes.[6] In short, we used the webtool CRISPy-web[7] to determine a protospacer flanked by a PAM sequence (5'-NGG-3') suited for CRISPR-nCas9-based genome editing. The PAM site was chosen so that it was present only once in the genome to minimize off-target effects. Gene synthesis (GENEWIZ, Germany) was used to obtain the DNA fragment *clxA-ko* containing the sgRNA under the regulation of the mini-P4 promoter as well as the homologous arms for

homology-directed repair (Table S4). The obtained vector was *Bsa*I-digested, the desired fragment extracted using the Monarch Gel Extraction Kit (New England Biolabs) and ligated into *Bsa*I-linearized pCasC using T4 Ligase (Thermo Scientific) to result in pCas-clxA-ko (Table S2). For verification, the region of pCas-clxA-ko containing the fragment was sequenced (GENEWIZ, Germany) using primers HB1-seq-rev and pUC-ori-seq-rev (Table S3).

Generation of C. cavendishii Rif1 ΔclxA

The single knock-out mutant *C. cavendishii* Rif1Δ*clxA* was generated by a procedure based on the genetic modification of *R. cellulolyticum*.[6] *C. cavendishii* Rif1 was transformed with the plasmid pCas-clxA-ko. For this, *C. cavendishii* Rif1 was revived from stock in 5 mL VM medium. Overnight cultures were subcultured once (2.5% *v*/*v* inoculum) in 5 mL VM medium. 2.5 mL of the subculture was transferred to 50 mL GS-2 medium and grown until mid-log phase was reached (optical density at 600 nm (OD_{600}) of 0.3–0.45). Glycine was added to a final concentration of 10 mg mL^{-1} and cells were grown for 1 hour at 37 °C. Cells were harvested at 6000 x *g* for 10 minutes at 4 °C and washed twice with 15 mL of ice-cold electroporation buffer (270 mM sucrose, 5 mM K_2HPO_4, pH 6.5). The pellet was resuspended in 2.5 mL electroporation buffer and 0.5 mL of this suspension added to 1 µg of plasmid DNA. Cells were transferred to an ice cold 0.4 cm electroporation cuvette and incubated for 10 minutes on ice. Cells were electroporated (2000 V, 25 µF, 1000 Ω) and immediately added to 2 mL pre-warmed GS-2 medium. Cells were recovered for 2 hours and then centrifuged at 4000 x *g* for 10 minutes. Pelleted cells were resuspended in 200 µL GS-2 medium and spread on GS-2 agar plates containing 10 µg mL^{-1} erythromycin. Colony-forming units (CFU) that appeared after 1–2 days were transferred to fresh GS-2 plates containing 15 µg mL^{-1} erythromycin.

Successful gene editing of the target gene resulted in the introduction of an in-frame STOP codon and, to facilitate screening, an *Eco*RV restriction site (5'-TAAGATATC-3'). The presence or absence of the restriction site (corresponding to edited or unedited *clx*A) was identified by colony PCR using primers ko_clxA_fwd and ko_clxA_rev and subsequent digest of the PCR fragment with *Eco*RV for 1 hour at 37 °C. PCR amplicons were analysed on a 1% agarose gel stained with ethidium bromide and visualized with UV light. The edited target site would result in two fragments (552 and 460 bp), whereas the unedited target site would result in a product of 1012 bp. Absence of the full-length PCR product and presence of both expected fragments were assigned as fully edited mutants, i.e. *C. cavendishii* Rif1Δ*clxA*. The PCR fragment obtained from the wild-type control remained undigested and the negative control (water instead of template) did not show any band. *C. cavendishii* Rif1Δ*clxA* was cultivated in presence of 20 µg mL^{-1} erythromycin to maintain the editing vector when checking for closoxazole production.

General analytical methods

NMR: All 1D (^{1}H, ^{13}C, DEPT) and 2D NMR (^{1}H-^{1}H COSY, HSQC, HMBC) were recorded in deuterated solvents using a Bruker AVANCE II 300, AVANCE III 500 or 600 MHz instrument equipped with Bruker Cryo Platform. The chemical shifts are reported in ppm relative to the solvent residual signal (^{1}H: δ (CH_2Cl_2) = 5.32 ppm, δ (DMSO) = 2.50 ppm, ^{13}C: δ (CD_2Cl_2) = 53.84 ppm, δ (DMSO-d_6) = 39.52 ppm.[8] The following abbreviations are used for multiplicities of resonance signals: s = singlet, d = doublet, t = triplet, q = quartet, qt = quintet, br = broad.

LC-MS: LC-MS measurements were performed using Q Exactive Orbitrap High Performance Benchtop LC-MS or Exactive Hybrid-Quadrupole-Orbitrap with electrospray ion source and Accela HPLC system (Thermo Fisher Scientific, Bremen) or Ultimate3000 UHPLC system (Thermo Fisher Scientific, Bremen), or LTQ Velos Ion Trap Benchtop LC-MS with electrospray ion source and Surveyor HPLC system (Thermo Fisher Scientific, Bremen).

HPLC conditions using Exactive: A Betasil C18 column; (2.1 × 15 0mm, 5 µm, Thermo Fisher) was used with gradient elution as follows: MeCN (0.1% (*v*/*v*) HCOOH)/H_2O (0.1% (*v*/*v*) HCOOH). method: initially at 5:95, reaching 2:98 over 6 min then maintaining 2:98 for 3 min. The flow rate was 0.2 mL · min^{-1} and injection volume was 3 µL.

HPLC conditions using Q Exactive: An Accucore C18 column (2.1 × 100 mm, 2.6 µm, Thermo Fisher) was used with gradient elution as follows: MeCN (0.1% (*v*/*v*) HCOOH)/H_2O (0.1% (*v*/*v*) HCOOH) initially at 5:95, reaching 2:98 over 10 min, then maintaining 2:98 for 4 min. The flow rate was 0.2 mL · min^{-1} and injection volume was 3 µL.

UHPLC conditions using Q Exactive: An Accucore C18 column (2.1 × 100 mm, 2.6 µm, Thermo Fisher) was used with gradient elution as follows: MeCN (0.1% (*v*/*v*) HCOOH)/H_2O (0.1% (*v*/*v*) HCOOH) initially at 5:95, reaching 2:98 over 10 min, then maintaining 2:98 for 4 min. The flow rate was 0.2 mL · min^{-1} and injection volume was 3 µL.

UHPLC conditions using Q Exactive: An Accucore C18 column (2.1 × 100 mm, 2.6 µm, Thermo Fisher) was used with gradient elution as follows: MeCN (0.1% (*v*/*v*) HCOOH)/H_2O (0.1% (*v*/*v*) HCOOH) initially at 5:95, reaching 2:98 over 7 min, then maintaining 2:98 for 3 min. The flow rate was 0.2 mL · min^{-1} and injection volume was 3 µL.

HPLC conditions using LTQ: A C18 column (Phenomenex Kinetex XB-C18, 2.6 µm, 100 × 3 mm) was used with gradient elution as follows: MeCN (0.1% (*v*/*v*) HCOOH)/H_2O (0.1% (*v*/*v*) HCOOH) initially at 10:90 for 1 min, reaching 100:0 over 8 min, maintaining 100:0 for 4 min. The flow rate was 0.6 mL · min^{-1} and injection volume was 5 µL.

HPLC:

Analytical HPLC: HPLC measurements were performed on an analytical RP-HPLC system A C18 Waters Symmetry column, 5 µm or EC250/4.6 Neucleodur C18 HTEC 5 µm was used with gradient elution as follows: MeCN/H_2O (0.1% (*v*/*v*) TFA) initially at 20:80 for 5 min, reaching 99:1 over 35 min, maintaining 99:1 for 5 min and a flow rate of 1 mL min^{-1} and injection volume was 10 µL.

Preparative HPLC: Fractions containing the desired compound were further purified using the preparative HPLC System LC-8A/CBM-20A (Shimazu) using a C18 column Phenomenex Synergi 4 µm Fusion RP, 80 A; 250 x 21.2 mm with gradient elution as follows:

Closoxazole A and B: MeCN (83% in H_2O)/H_2O (0.01% (*v*/*v*) TFA initially at 10:90 for 5 minutes, reaching 100:0 in 40 minutes, then maintaining 100:0 for 10 minutes and a flow rate of 10 mL min^{-1}. Tafamidis analogue (**25**): MeCN/H_2O (0.01% (*v*/*v*) TFA initially at 30:70 for 5 minutes, reaching 100:0 in 35 minutes, then maintaining 100:0 for 10 minutes and a flow rate of 12 mL min^{-1}.

For purification of synthetic compounds: using a C18 column Phenomenex Luna 10 u C18(2) 100 A AXIA (250 × 21.2 mm, 10 µm) with gradient elution as follows: MeCN (100%)/H_2O (0.01% (*v*/*v*) TFA initially at 25:75 for 4 minutes, reaching 100:0 in 24 minutes, then maintaining 100:0 for 10 minutes and a flow rate of 18 mL · min^{-1}, or using a C18 column Nucleodur C18 HTec VP250/10 (250 × 10 mm, 10 µm) with gradient elution as follows: MeCN (100%)/H_2O (0.01% (*v*/*v*) TFA initially at 25:75 for 4 minutes, reaching 100:0 in 24 minutes, then maintaining 100:0 for 10 minutes and a flow rate of 5 mL min^{-1}.

Experimental procedures with *E. coli* strains

Plasmid construction for E. coli expression vectors

The target *clx* genes were amplified by PCR from genomic DNA isolated from *C. cavendishii* Rif1. Primers were designed to contain 15 bp overhang specific for the target vector (Table S3). PCRs were performed with Phusion® High-Fidelity PCR Master Mix and amplicons purified using Monarch Gel Extraction Kit (New England Biolabs). Purified PCR fragments were ligated into linearized vector (Table S2) using NEBuilder according to the manufacturers protocol at 50 °C for 15 min. Ligation reactions were added to electrocompetent *E. coli* Top10 or *E. coli* XL-1 strains and incubated for 15 min on ice, followed by electroporation (2250 V). The cells were recovered for 1–2 h at 37 °C in LB medium with agitation, then spread on LB agar plates supplemented with the appropriate antibiotic. Colony PCR was used to ensure that transformants possessed vectors with inserted PCR fragments. For verification, plasmids were isolated and subjected to restriction analysis to verify insert size.

Heterologous expression of clx biosynthetic gene cluster

For heterologous expression of the *clx* genes, *E. coli* BL21 (DE3) were transformed as described above with pET28a and/or pETDuet derived expression vectors. A single colony was used to create a glycerol stock for long-term storage. The relevant strain was revived from stock 2 mL LB (0.5% *v*/*v* inoculum) and grown overnight with the appropriate antibiotic. For production and isolation of benzoxazoles, 50 mL of autoinduction medium (6 g Na_2HPO_4, 3 g KH_2PO_4, 20 g BD Bacto™ Tryptone, 5 g BD Bacto™ yeast extract, 5 g NaCl, 7.56 g glycerol, 0.5 g D-glucose, 2 g lactose, 1000 mL water; optional: 100 µg mL^{-1} kanamycin, 50 µg mL^{-1} ampicillin) was inoculated with 1% *v*/*v* of freshly grown overnight culture. Cultures were grown for 12 h at 37 °C with agitation in baffled cultivation flasks. For chemical complementation experiments, synthetic 3,4-AHBA was added to a final concentration of 200 µM to the relevant cultures during the exponential phase (OD_{600} ca. 0.5–0.8).

Supplementation of closoxazole-producing E. coli with varying aromatic acids

For precursor directed biosynthesis experiments, the respective *E. coli* strains were handled as described above. Chemical supplements were sterile filtered through Rotilabo-syringe filters (CME, pore size 0.22 µM) and added to a final concentration of 200 µM to *E. coli* in the exponential growth stage (OD_{600} ca. 0.5–0.8). When supplementing two precursor substrates, both compounds were added in final concentrations of 200 µM each.

General culture extractions for metabolic profiling

For metabolic profiling of *E. coli*, cultures were extracted once with 1:1 (*v*/*v*) ethyl acetate, then the extract was dried over Na_2SO_4 and filtered through filter paper (Rotilabo Typ 600P). Upon concentration under reduced pressure, the extracts were dissolved in methanol (500x concentration) and subjected to analytical LC-MS.

Large-scale extraction and isolation of closoxazoles

Closoxazoles were purified from *C. cavendishii* Rif1 culture broth and *E. coli* culture broth. Briefly, *C. cavendishii* Rif 1 was cultured in 500 mL Schott bottles in an anerobic chamber at 30 °C for 3 days. For extraction with XAD-2 beads, 20 g L^{-1} beads were equilibrated with water, then added to the culture broth and incubated aerobically overnight with mixing. The XAD2 beads were separated from culture broth and washed twice with water, followed by extraction in methanol for 1–3 hours. For extraction with ethyl acetate, the culture broth was extracted twice with ethyl acetate, dried under reduced pressure, and dissolved in methanol. After concentration under reduced pressure, the extract was pre-purified using an open size exclusion column (Sephadex L20 resin, MeOH as eluent) or silica column (Kieselgel 60, pore size 20–50 µm, Merck KGaA, elution gradient ethyl acetate:methanol). Fractions were checked for the presence of closoxazole A and B by analytical HPLC. Fractions containing the desired compound were further purified by preparative HPLC.

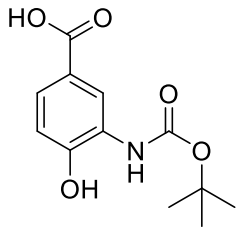
Closoxazole A
^{1}H NMR: (600 MHz, DMSO-d_6, 27 ºC): δ = 11.91–13.07 (m, 1H), 10.68 (s, 1H), 10.01 (br s, 1H), 9.66 (s, 1H), 8.31 (d, *J* = 1.2 Hz, 1H), 8.29 (d, *J* = 2.0 Hz, 1H), 7.97 (dd, *J* = 8.5, 1.6 Hz, 1H), 7.83 (d, *J* = 8.5 Hz, 1H), 7.67 (dd, *J* = 8.4, 2.1 Hz, 1H), 7.50 (d, *J* = 2.1 Hz, 1H), 7.35 (dd, *J* = 8.1, 2.1 Hz, 1H), 6.99 (d, *J* = 8.5 Hz, 1H), 6.84 (d, *J* = 8.2 Hz, 1H), 4.56–5.36 ppm (m, 2H).
^{13}C NMR (151 MHz, DMSO-d_6, 27 ºC): δ = 167.1, 165.1, 164.8, 154.1, 152.1, 148.3, 142.0, 137.4, 131.0, 127.7, 126.3, 125.6, 124.6, 121.4, 118.6, 117.1, 117.1, 115.6, 114.5, 112.7, 110.4 ppm.
HRMS (ESI$^+$) calcd. for $C_{21}H_{16}N_3O_6^+$: 406.1034; found: 406.1031.

Closoxazole B
^{1}H NMR (600 MHz, DMSO-d_6, 27 ºC): δ = 12.35–12.68 (m, 1H), 10.69 (s, 1H), 9.63–9.67 (m, 1H), 8.32 (d, *J* = 2.0 Hz, 1H), 8.28 (d, *J* = 1.3 Hz, 1H), 7.94–7.97 (m, 1H), 7.89–7.92 (m, 2H), 7.79–7.82 (m, 1H), 7.66–7.71 (m, 1H), 6.99–7.03 (m, 1H), 6.70–6.74 (m, 2H), 6.08 ppm (br s, 1H).
^{13}C NMR (151 MHz, DMSO-d_6, 27 ºC): δ = 167.5, 165.6, 165.4, 154.4, 153.4, 152.5, 142.7, 131.4, 129.7, 128.1, 126.7, 126.1, 124.7, 121.9, 118.6, 116.0, 114.0, 112.6, 110.6 ppm.
HRMS (ESI$^+$) calcd. for $C_{21}H_{16}N_3O_5^+$: 390.1084; found: 390.1076.

Extraction and isolation of **25**
Compound **25** was purified from *E. coli* culture broth. Briefly, the culture broth was extracted twice with ethyl acetate, dried under reduced pressure and dissolved in methanol. The obtained extract was pre-purified using an open silica column (Kieselgel 60, pore size 20–50 µm, Merck KGaA, elution gradient ethyl acetate:methanol). Fractions were checked for the presence of the compound by LC-MS. The fraction containing the desired compound **25** was dried under reduced pressure and redissolved in methanol/water/hexane (45:5:50 v/v) for 5 h with occasional shaking. The two layers formed were separated and checked for the presence of the compound by LC-MS and both subjected for further purification by preparative HPLC.
^{1}H NMR (600 MHz, DMSO-d_6, 27 ºC): δ = 13.49 (s, 1H), 9.88–9.92 (m, 1H), 8.34 (br d, *J* = 1.1 Hz, 1H), 8.18 (dd, *J* = 1.9 Hz, 2H), 8.08–8.09 (m, 1H), 8.10 (dd, *J* = 8.6, 1.6 Hz, 1H), 7.94–7.95 (m, 1H), 7.96 (br t, *J* = 1.9 Hz, 2H), 7.90 ppm (d, *J* = 8.6 Hz, 1H).
^{13}C NMR (151 MHz, DMSO-d_6, 27 ºC): δ = 166.9, 161.0, 152.8, 141.2, 135.2, 131.5, 129.3, 127.7, 125.9, 121.3, 111.1 ppm.
HRMS (ESI$^+$) calcd. for $C_{14}H_8Cl_2NO_3^+$: 307.9876; found: 307.9873.

Synthesis procedures

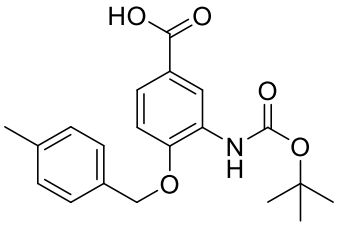
All reagents were obtained from commercial suppliers (Sigma Aldrich, TCI, Alfa Aesar, etc.) and used without further purification, unless otherwise explained. Some reactions were performed under inert gas (argon) by using the Schlenk technique and dried solvents. Dichloromethane (DCM), acetonitrile (MeCN), methanol (MeOH) and chloroform were used from a solvent purification system (Innovative Technologies). Open column chromatographic separations were performed on silica gel (Kieselgel 60, 15–40 µm, Merck KGaA). Reaction progresses were monitored by thin layer chromatography (TLC) (silica gel on aluminum sheets 20 × 20 cm with fluorescent indicator 254 nm, Merck KGaA) or HPLC-(HR)MS.
The synthesis of **14** was performed first, then the route was later modified for the synthesis of **11**. It was not possible to simultaneously use Boc and TBS as protecting groups for 3,4-AHBA, probably because of steric hindrance. 3,4-AHBA protected with Boc and acetyl led to an undesirable major product after microwave-assisted benzoxazole formation. Ultimately, Boc and 4-Me-Bn used as protecting groups for 3,4-AHBA enabled microwave-assisted benzoxazole formation, although the harsh reaction conditions led to the partial deprotection of the Boc group. Thus, after the microwave-assisted reaction, benzoxazole **20** was obtained in addition to unreacted starting material (**19**) and Boc-protected **20**.



15

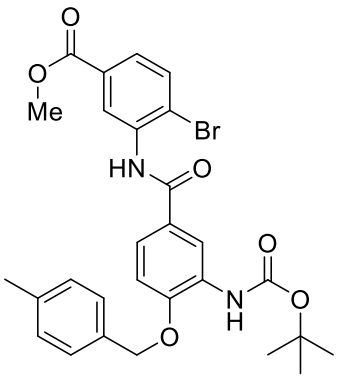
Synthesis of 3-((*tert*-butoxycarbonyl)amino)-4-hydroxybenzoic acid.
3-Amino-4-hydroxybenzoic acid (**13**, 851 mg, 5.56 mmol, 1 eq.) was suspended in DMF (7 mL), then Boc_2O (1.27 g, 5.82 mmol, 1.05 eq.) was added. The mixture was stirred at 70 °C for 3 h, cooled to 20 °C and poured into water (30 mL). The white precipitate was filtered, washed with water (2 × 30 mL) and dried under reduced pressure to give the title product (1.12 g, 4.43 mmol, 80%) as a white powder.

^{1}H NMR (300 MHz, DMSO-d_6, 21 ºC): δ = 12.49 (br s, 1H), 10.65 (br s, 1H), 8.27 (d, *J* = 1.8 Hz, 1H), 7.90 (s, 1H), 7.51 (dd, *J* = 8.5, 2.1 Hz, 1H), 6.88 (d, *J* = 8.5 Hz, 1H), 1.46 ppm (s, 9H).
^{13}C NMR (75 MHz, DMSO-d_6, 22 ºC): δ = 167.2, 152.7, 151.5, 126.1, 125.6, 122.1, 121.4, 114.5, 79.4, 28.0 ppm.
HRMS (ESI$^+$) calcd. for $C_{12}H_{16}NO_5^+$: 254.1023; found: 254.1020.



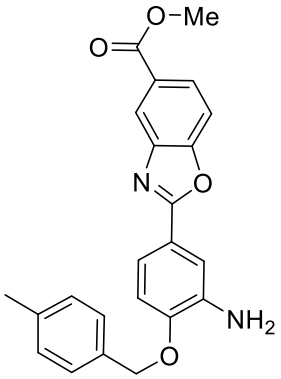
16

Synthesis of 3-((*tert*-butoxycarbonyl)amino)-4-((4-methylbenzyl)oxy)benzoic acid.
3-Amino-4-hydroxybenzoic acid (**13**, 1.00 g, 6.53 mmol, 1 eq.) was dissolved in DMF (10 mL), then Boc_2O (1.50 g, 6.85 mmol, 1.05 eq.) was added at 0 °C. The mixture was stirred at 70 °C for 3 h, then 4-MeBnBr (3.63 g, 19.59 mmol, 3 eq.) and K_2CO_3 (2.71 g, 19.59 mmol, 3 eq.) were added and stirred at 70 °C for 1 h. Water (30 mL) was added under strong stirring and the solution was cooled to 4 °C. A highly viscous layer formed, the water phase was decanted, and the organic phase was washed with water (2 × 30 mL). MeOH (30 mL), water (20 mL) and NaOH (1.28 g, 32 mmol, 4.9 eq.) were added and the mixture was stirred at 70 °C for 4 h until everything was dissolved. Water (30 mL) was added, the solution was cooled to 20 °C and acetic acid (1.87 mL, 1.96 g, 32.64 mmol, 5 eq.) was added slowly under strong stirring. A white precipitate formed and was filtered, washed with water (2 × 30 mL), and dried under reduced pressure to give the title product (1.98 g, 5.54 mmol, 85%) as a white powder.
^{1}H NMR (500 MHz, DCM-d_2, 18 ºC): δ = 8.80 (br s, 1H), 7.78 (dd, *J* = 8.6, 2.2 Hz, 1H), 7.35–7.38 (m, *J* = 8.0 Hz, 2H), 7.26–7.29 (m, *J* = 7.8 Hz, 2H), 7.14 (s, 1H), 7.04 (d, *J* = 8.6 Hz, 1H), 5.20 (s, 2H), 2.41 (s, 3H), 1.55 ppm (s, 9H).
^{13}C NMR (126 MHz, DCM-d_2, 19 ºC): δ = 170.9, 153.0, 151.5, 139.1, 133.2, 129.9, 129.0, 128.4, 125.7, 122.3, 120.0, 111.4, 81.3, 71.4, 28.5, 21.5 ppm.
HRMS (ESI$^+$) calcd. for $C_{20}H_{24}NO_5^+$: 358.1649; found: 358.1646.



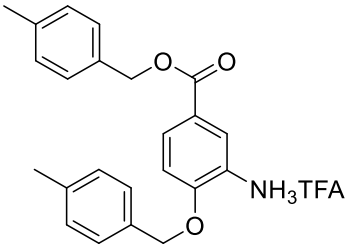
19

Synthesis of methyl 4-bromo-3-(3-((*tert*-butoxycarbonyl)amino)-4-((4-methylbenzyl)oxy)benzamido)benzoate.
3-((*Tert*-butoxycarbonyl)amino)-4-((4-methylbenzyl)oxy)benzoic acid (**16**, 127.4 mg, 356 µmol, 2.05 eq.) was suspended in DCM (6 mL), then DIPEA (134.8 mg, 181.7 µL, 1.04 mmol, 6eq.) was added to give a clear solution. $SOCl_2$ (82.7 mg, 51.7 µL, 695 mmol, 4 eq.) was added and stirred at 20 °C for 10 min. All volatiles were removed under reduced pressure and DCM (6 mL), DIPEA (134.8 mg, 181.7 µL, 1.04 mmol, 6 eq.), a catalytic amount of DMAP (~1 mg) and methyl 3-amino-4-bromobenzoate (**18**, 40 mg, 173 µmol, 1 eq.) were added and stirred overnight at 20 °C. Water (15 mL) was added and the organic phase was washed with NaOH (5 mL, 1 M), HCl (5 mL, 1 M), water (5 mL), dried over Na_2SO_4 and the solvent was removed under reduced pressure. The crude product was purified by preparative HPLC to give the title product (44 mg, 77 µmol, 44%) as a white solid.
^{1}H NMR (500 MHz, DCM-d_2, 27 ºC): δ = 9.09 (d, *J* = 1.8 Hz, 1H), 8.79 (s, 1H), 8.47 (s, 1H), 7.69–7.75 (m, 1H), 7.74 (s, 1H), 7.62 (dd, *J* = 8.5, 2.3 Hz, 1H), 7.36–7.39 (m, *J* = 8.0 Hz, 2H), 7.27–7.29 (m, *J* = 7.9 Hz, 2H), 7.19 (s, 1H), 7.09 (d, *J* = 8.5 Hz, 1H), 5.21 (s, 2H), 3.96 (s, 3H), 2.42 (s, 3H), 1.56 ppm (s, 9H).
^{13}C NMR (126 MHz, DCM-d_2, 27 ºC): δ = 166.7, 165.3, 153.1, 150.3, 139.1, 137.0, 133.4, 133.0, 131.2, 130.0, 129.6, 128.4, 127.5, 126.3, 123.1, 122.5, 119.5, 117.1, 112.0, 81.3, 71.5, 52.9, 28.6, 21.5 ppm.
HRMS (ESI$^+$) calcd. for $C_{28}H_{30}BrN_2O_6^+$: 569.1282; found: 569.1286.



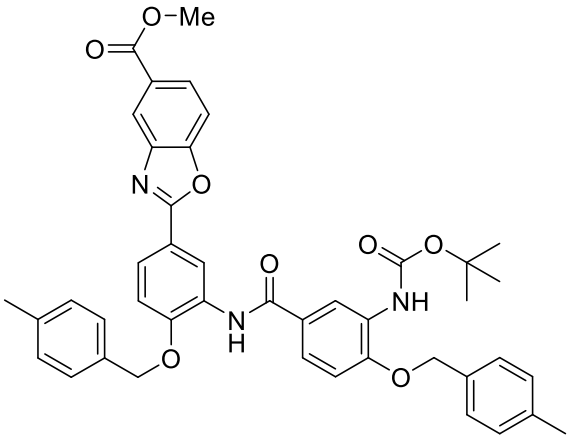
20

Synthesis of methyl 2-(3-amino-4-((4-methylbenzyl)oxy)phenyl)benzo[*d*]oxazole-5-carboxylate.
A solution of methyl 4-bromo-3-(3-((*tert*-butoxycarbonyl)amino)-4-((4-methylbenzyl)oxy)benzamido)benzoate (**19**, 20 mg, 35 µmol, 1 eq.), CuI (50 mg, 263 µmol, 7.5 eq.), Cs_2CO_3 (100 mg, 307 µmol, 8.7 eq.) and 9,10-phenantroline (50 mg, 280 µmol, 8 eq.) in MeCN (10 mL) was prepared. After sonication, the mixture was heated to 210 °C in a microwave (2 min 500 W, then 90 min 150 W) for 1.5 h. The solvent was evaporated under reduced pressure and the crude product was purified by preparative HPLC to give the title product (2.5 mg, 6.4 µmol, 18%) as a white solid.
^{1}H NMR (500 MHz, DCM-d_2, 27 ºC): δ = 8.39 (d, *J* = 1.4 Hz, 1H), 8.08 (dd, *J* = 8.5, 1.6 Hz, 1H), 7.64–7.67 (m, 2H), 7.64–7.67 (m, 1H), 7.63 (d, *J* = 8.5 Hz, 1H), 7.38–7.41 (m, *J* = 7.9 Hz, 2H), 7.25–7.28 (m, *J* = 7.8 Hz, 2H), 7.04 (d, *J* = 9.0 Hz, 1H), 5.18 (s, 2H), 3.97 (s, 3H), 2.41 ppm (s, 4H).
^{13}C NMR (126 MHz, DCM-d_2, 27 ºC): δ = 166.6, 164.8, 153.7, 149.5, 142.6, 138.2, 137.2, 133.5, 129.3, 127.8, 126.9, 126.3, 121.2, 119.5, 118.5, 113.3, 111.7, 110.1, 70.4, 52.1, 20.9 ppm.
HRMS (ESI$^+$) calcd. for $C_{23}H_{21}N_2O_4^+$: 389.1496; found: 389.1495.



17

Synthesis of 2-((4-methylbenzyl)oxy)-5-(((4-methylbenzyl)oxy)carbonyl)benzenaminium trifluoroacetate.
3-((*Tert*-butoxycarbonyl)amino)-4-hydroxybenzoic acid (**15,** 523 mg, 2.07 mmol, 1 eq.) was suspended in DMF (5 mL), then 4-MeBnBr (1.15 g, 6.21 mmol, 3 eq.) and K_2CO_3 (857 mg, 6.20 mmol, 3 eq.) were added. The mixture was heated to 70 °C for 1 h, poured into water (50 mL) and stirred vigorously for 5 min until a highly viscous second layer formed. The water phase was decanted, the viscous liquid was washed with water (50 mL) and dried under reduced pressure. Everything was dissolved in DCM (10 mL), then TFA (1 mL) was added. The mixture was stirred at 40 °C for 12 h, then toluene (20 mL) was added and all solvents were evaporated under reduced pressure. The resulting white solid was suspended in DCM (5 mL), filtered, washed with DCM (2 mL) and dried under reduced pressure to give the TFA salt of the title product (708 mg, 1.49 mmol, 72%) as a white solid.
^{1}H NMR (300 MHz, DMSO-d_6, 21 ºC): δ = 7.53 (d, *J* = 2.1 Hz, 1H), 7.43 (br d, *J* = 2.1 Hz, 1H), 7.39 (d, *J* = 8.0 Hz, 2H), 7.30–7.36 (m, 2H), 7.20 (d, *J* = 8.2 Hz, 4H), 7.09 (d, *J* = 8.7 Hz, 1H), 5.84 (br s, 3H), 5.24 (s, 2H), 5.19 (s, 2H), 2.31 ppm (s, 6H).
^{13}C NMR (75 MHz, DMSO-d_6, 22 ºC): δ = 165.8, 151.4, 137.9, 137.7, 133.9, 133.8, 133.2, 129.5, 129.4, 128.6, 128.1, 122.7, 122.6, 117.9, 112.4, 70.1, 66.2, 21.2 ppm.
HRMS (ESI$^+$) calcd. for $C_{23}H_{24}NO_3^+$: 362.1751; found: 362.1745.



26

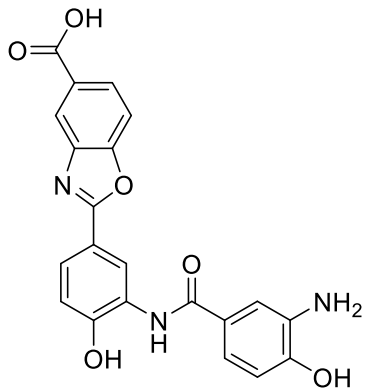
Synthesis of methyl 2-(3-(3-((*tert*-butoxycarbonyl)amino)-4-((4-methylbenzyl)oxy)benzamido)-4-((4-methylbenzyl)oxy)phenyl)benzo[*d*]oxazole-5-carboxylate.

Crude *tert*-butyl (5-(chlorocarbonyl)-2-((4-methylbenzyl)oxy)phenyl)carbamate was prepared by mixing 3-((tert-butoxycarbonyl)amino)-4-((4-methylbenzyl)oxy)benzoic acid (**16**, 3 mg, 8.4 µmol, 1 eq.) with DIPEA (3.25 mg, 4.4 µL, 25.2 µmol, 3 eq.) and $SOCl_2$ (2 mg, 1.25 µL, 16.8 µmol, 2 eq.) in DCM (50 µL) for 10 min, followed by removal of all volatile compounds. A solution of methyl 2-(3-amino-4-((4-methylbenzyl)oxy)phenyl)benzo[*d*]oxazole-5-carboxylate (**20**, 1.67 mg, 4.3 µL, 1 eq.), crude *tert*-butyl (5-(chlorocarbonyl)-2-((4-methylbenzyl)oxy)phenyl)carbamate (2 eq.) and DIPEA (1.7 mg, 2.3 µL, 12.9 µmol, 3 eq.) in DCM (600 µL) was prepared and stirred at 20 °C for 20 h. The solvent was removed under reduced pressure and the mixture was purified by preparative HPLC to give the title compound (1.5 mg, 2.1 µmol, 48%) as a white solid.

^{1}H NMR (600 MHz, DCM-d_2, 27 ºC): δ = 9.40 (d, *J* = 2.1 Hz, 1H), 8.75 (br s, 1H), 8.68 (s, 3H), 8.42–8.45 (m, 1H), 8.11 (dd, *J* = 8.5, 1.7 Hz, 2H), 7.99–8.08 (m, 2H), 7.69 (d, *J* = 8.5 Hz, 2H), 7.55 (dd, *J* = 8.5, 2.3 Hz, 2H), 7.45 (d, *J* = 8.0 Hz, 3H), 7.34–7.41 (m, 2H), 7.29 (d, *J* = 7.9 Hz, 2H), 7.26 (d, *J* = 7.9 Hz, 2H), 7.22 (d, *J* = 8.6 Hz, 1H), 7.19 (s, 1H), 7.06 (d, *J* = 8.5 Hz, 2H), 5.32 (s, 3H), 5.21 (s, 2H), 3.98 (s, 4H), 2.42 (s, 3H), 2.39 (s, 3H), 1.57 ppm (s, 9H).

^{13}C NMR (151 MHz, DCM-d_2, 27 ºC): δ = 166.7, 164.9, 164.4, 153.8, 152.5, 150.2, 149.5, 142.6, 138.5, 138.4, 132.9, 129.4, 129.4, 128.8, 127.8, 127.7, 127.6, 127.0, 126.5, 123.6, 121.8, 121.4, 119.7, 118.7, 116.5, 112.1, 111.3, 110.3, 80.7, 71.1, 70.9, 52.1, 28.0, 20.9 ppm.

HRMS (ESI$^+$) calcd. for $C_{43}H_{42}N_3O_8^+$: 728.2966; found: 728.2972.

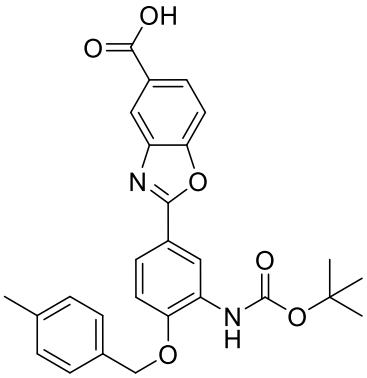


14

Synthesis of 2-(3-(3-amino-4-hydroxybenzamido)-4-hydroxyphenyl)benzo[*d*]oxazole-5-carboxylic acid.

Methyl 2-(3-(3-((*tert*-butoxycarbonyl)amino)-4-((4-methylbenzyl)oxy)benzamido)-4-((4-methylbenzyl)oxy)phenyl)benzo[*d*]oxazole-5-carboxylate (**26**, 1.23 mg, 1.69 µmol, 1 eq.) was dissolved in DCM (100 µL) and TFA (200 µL). The solution was stirred for 3.5 h at 65 °C and the solvents were removed under a stream of nitrogen. The crude product was redissolved in MeOH (50 µL) and NaOH (200 µL, 1 M), then stirred at 20 °C for 30 min. HCl (200 µL, 1 M) was added and the crude product was purified by preparative HPLC to yield the TFA salt of the title product (0.7 mg, 1.35 µmol, 80%) as a white solid. Storage under reduced pressure (<1 mbar) over three days gave the free amine.

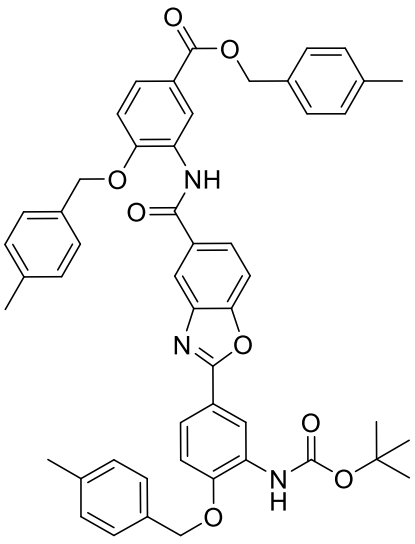
^{1}H NMR (600 MHz, DMSO-d_6, 27 ºC): δ = 12.72–13.43 (m, 1H), 11.06 (br s, 1H), 10.27–10.27 (m, 1H), 9.93 (br s, 1H), 9.17 (br s, 1H), 8.85 (d, *J* = 1.6 Hz, 1H), 8.22–8.33 (m, 1H), 8.02 (dd, *J* = 8.5, 1.5 Hz, 1H), 7.90 (dd, *J* = 8.4, 2.0 Hz, 1H), 7.88 (d, *J* = 8.5 Hz, 1H), 7.34 (br s, 1H), 7.20–7.27 (m, 1H), 7.14 (d, *J* = 8.4 Hz, 1H), 6.82 (br d, *J* = 8.1 Hz, 1H), 4.76–5.76 ppm (m, 1H).
^{13}C NMR (151 MHz, DMSO-d_6, 27 ºC): δ = 167.5, 165.8, 164.5, 153.5, 152.4, 148.8, 142.4, 130.1, 128.1, 127.8, 127.0, 125.7, 125.1, 121.8, 121.0, 118.2, 117.2, 116.5, 115.1, 114.4, 111.2 ppm.
HRMS (ESI$^+$) calcd. for $C_{21}H_{16}N_3O_6^+$: 406.1034; found: 406.1031.



27

Synthesis of 2-(3-((*tert*-butoxycarbonyl)amino)-4-((4-methylbenzyl)oxy)phenyl)benzo[*d*]oxazole-5-carboxylic acid.

Methyl 2-(3-amino-4-((4-methylbenzyl)oxy)phenyl)benzo[*d*]oxazole-5-carboxylate (**20**, 2.55 mg, 6.6 µmol, 1 eq.) was dissolved in DMF (500 µL) and Boc_2O (100 mg, 458 µmol, 70 eq.) was added. The solution was stirred at 70 °C for 8 h, then water (1 mL) was added and stirred at 20 °C overnight. The precipitate was filtered and washed with water (2 × 1.5 mL). The white solid was suspended in MeOH (1.5 mL) and NaOH (200 µL, 1 M) and stirred at 70 °C for 1 h. HCl (100 µL, 2 M) was slowly added and the precipitate was filtered and washed with water (1 mL) and dried under reduced pressure to give the title product (2.3 mg, 4.9 µmol, 74%) as a white solid.
^{1}H NMR (600 MHz, DCM-d_2, 27 ºC): δ = 9.04 (br s, 1H), 8.50 (d, *J* = 1.5 Hz, 1H), 8.17 (dd, *J* = 8.6, 1.7 Hz, 1H), 7.94 (dd, *J* = 8.4, 2.2 Hz, 1H), 7.71 (d, *J* = 8.8 Hz, 1H), 7.39 (d, *J* = 8.1 Hz, 2H), 7.29 (d, *J* = 7.7 Hz, 2H), 7.22 (s, 1H), 7.14 (d, *J* = 8.8 Hz, 1H), 5.23 (s, 2H), 2.42 (s, 3H), 1.58 ppm (s, 9H).
^{13}C NMR (151 MHz, DCM-d_2, 27 ºC): δ = 168.9, 164.7, 154.3, 152.5, 149.7, 142.7, 138.6, 132.8, 129.4, 129.2, 127.9, 127.1, 125.7, 122.4, 122.1, 119.4, 117.0, 111.7, 110.5, 80.8, 71.0, 28.0, 20.9 ppm.
HRMS (ESI$^+$) calcd. for $C_{27}H_{27}N_2O_6^+$: 475.1864; found: 475.1861.



21

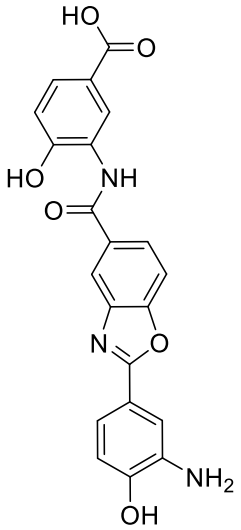
Synthesis of 4-methylbenzyl 3-(2-(3-((*tert*-butoxycarbonyl)amino)-4-((4-methylbenzyl)oxy)phenyl)benzo[d]oxazole-5-carboxamido)-4-((4-methylbenzyl)oxy)benzoate.

A solution of 2-(3-((*tert*-butoxycarbonyl)amino)-4-((4-methylbenzyl)oxy)phenyl)benzo[*d*]oxazole-5-carboxylic acid (**27**, 2.3 mg, 4.9 µmol, 1 eq.) was dissolved in DCM (1 mL) and DIPEA (1.88 mg, 2.5 µL, 14.5 µmol, 3 eq.), followed by the addition of $SOCl_2$ (1.15 mg, 720.8 µL, 9.7 µmol, 2 eq.). The mixture was stirred at 20 °C for 10 min and the volatile solvents were removed under reduced pressure. DCM (1 mL), DIPEA (5.64 mg, 7.5 µL, 43.5 µmol, 9 eq.) and 2-((4-methylbenzyl)oxy)-5-(((4-methylbenzyl)oxy)carbonyl)benzenaminium trifluoroacetate (**17**, 2.9 mg, 6.1 µmol, 1.25 eq.) were added and the mixture was stirred at 20 °C and reaction process was monitored by HPLC-MS. After 1.5 h the solvent was removed under reduced pressure and the crude product was purified by preparative HPLC to give the title product (2.49 mg, 3.1 µmol, 63%) as a white solid.

^{1}H NMR (600 MHz, DCM-d_2, 27 ºC): δ = 9.17 (d, *J* = 1.8 Hz, 1H), 9.05 (br s, 1H), 8.70 (s, 1H), 8.22 (d, *J* = 1.5 Hz, 1H), 7.94 (dd, *J* = 8.6, 2.0 Hz, 1H), 7.90 (dd, *J* = 8.4, 1.8 Hz, 1H), 7.84 (dd, *J* = 8.8, 2.2 Hz, 1H), 7.70 (d, *J* = 8.4 Hz, 1H), 7.41 (d, *J* = 1.0 Hz, 2H), 7.39 (d, *J* = 7.7 Hz, 4H), 7.29 (br d, *J* = 7.3 Hz, 2H), 7.28 (br d, *J* = 7.7 Hz, 6H), 7.25 (d, *J* = 8.1 Hz, 2H), 7.23 (s, 1H), 7.12–7.16 (m, 1H), 7.09 (d, *J* = 8.8 Hz, 1H), 5.35 (br s, 2H), 5.28–5.31 (m, 2H), 5.23 (s, 2H), 2.42 (s, 3H), 2.41 (s, 3H), 2.40 (s, 3H), 1.58 ppm (s, 9H).

^{13}C NMR (151 MHz, DCM-d_2, 27 ºC): δ = 165.9, 164.7, 153.0, 152.5, 151.1, 149.7, 142.8, 138.6, 138.1, 133.4, 132.9, 132.8, 131.7, 129.5, 129.4, 129.2, 128.2, 128.2, 127.9, 127.5, 126.0, 124.2, 123.3, 122.4, 120.8, 119.5, 118.5, 117.0, 111.7, 111.4, 110.7, 80.8, 71.2, 71.0, 66.5, 28.0, 21.0, 20.9 ppm.

HRMS (ESI$^+$) calcd. for $C_{50}H_{48}N_3O_8^+$: 818.3436; found: 818.3438.



11

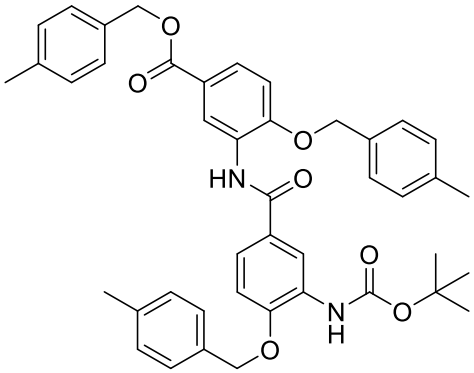
Synthesis of 3-(2-(3-amino-4-hydroxyphenyl)benzo[*d*]oxazole-5-carboxamido)-4-hydroxybenzoic acid.

A solution of 4-methylbenzyl 3-(2-(3-((*tert*-butoxycarbonyl)amino)-4-((4-methylbenzyl)oxy)phenyl)benzo[*d*]oxazole-5-carboxamido)-4-((4-methylbenzyl)oxy)benzoate (**21**, 2.0 mg, 2.45 µmol) in DCM (400 µL) and TFA (500 µL) was stirred in a sealed vial at 65 °C for 70 min. The reaction was monitored by HPLC-MS. All volatiles were removed, and the crude product was purified by preparative HPLC and dried under reduced pressure to give the title compound as the TFA salt (1.0 mg, 1.93 µmol, 79%). Storage under reduced pressure (<1 mbar) over three days gave the free amine (0.79 mg, 1.95 µmol, 80%) as a white powder.

^{1}H NMR (600 MHz, DMSO-d_6, 27 ºC): δ = 10.73 (br s, 1H), 10.16 (br s, 1H), 9.68 (s, 3H), 8.32 (d, *J* = 1.2 Hz, 1H), 8.31 (d, *J* = 2.0 Hz, 1H), 7.94–8.04 (m, 1H), 7.84 (d, *J* = 8.4 Hz, 1H), 7.69 (dd, *J* = 8.4, 2.1 Hz, 1H), 7.53 (d, *J* = 2.1 Hz, 1H), 7.39 (dd, *J* = 8.2, 2.1 Hz, 1H), 6.97–7.06 (m, 1H), 6.86 ppm (d, *J* = 8.2 Hz, 1H).

^{13}C NMR (151 MHz, DMSO-d_6, 27 ºC): δ = 167.2, 165.2, 164.9, 154.2, 152.2, 148.5, 142.0, 137.0, 131.1, 127.9, 126.4, 125.6, 124.7, 121.5, 118.6, 117.5, 117.1, 115.6, 114.6, 113.1, 110.5 ppm.

HRMS (ESI$^+$) calcd. for $C_{21}H_{16}N_3O_6^+$: 406.1034; found: 406.1033.



28

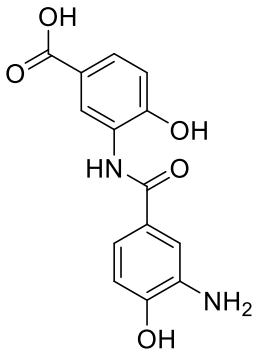
Synthesis of 4-methylbenzyl 3-(3-((*tert*-butoxycarbonyl)amino)-4-((4-methylbenzyl)oxy)benzamido)-4-((4-methylbenzyl)oxy)benzoate.

3-((*Tert*-butoxycarbonyl)amino)-4-((4-methylbenzyl)oxy)benzoic acid (**16**, 100 mg, 0.28 mmol, 1 eq.) was suspended in DCM (5 mL), then DIPEA (108.5 mg, 146.2 µL, 0.84 mmol, 3 eq.) was added to give a clear solution. $SOCl_2$ (66.6 mg, 39.6 µL, 0.56 mmol, 2 eq.) was added and stirred at 20 °C for 10 min. All volatiles were removed under reduced pressure, then DCM (5 mL), DIPEA (108.5 mg, 146.2 µL, 0.84 mmol, 3 eq.) and 2-((4-methylbenzyl)oxy)-5-(((4-methylbenzyl)oxy)carbonyl)benzenaminium trifluoroacetate (**17**, 133.6 mg, 0.28 mmol, 1 eq.) were added and stirred 1 h at 20 °C. The solvent was removed under reduced pressure and the crude mixture was purified by preparative HPLC to give the title product (148.2 mg, 0.21 mmol, 76%).

^{1}H NMR (500 MHz, DCM-d_2, 27 ºC): δ = 9.16 (d, *J* = 2.1 Hz, 1H), 8.75 (s, 1H), 8.61 (s, 1H), 7.85 (dd, *J* = 8.6, 2.2 Hz, 1H), 7.53 (dd, *J* = 8.5, 2.3 Hz, 1H), 7.44 (d, *J* = 4.1 Hz, 2H), 7.42 (d, *J* = 4.1 Hz, 2H), 7.39 (d, *J* = 8.0 Hz, 2H), 7.29 (d, *J* = 7.9 Hz, 2H), 7.27 (d, *J* = 4.4 Hz, 2H), 7.26 (d, *J* = 4.4 Hz, 2H), 7.22 (s, 1H), 7.10 (d, *J* = 8.6 Hz, 1H), 7.04 (d, *J* = 8.6 Hz, 1H), 5.37 (s, 2H), 5.27 (s, 2H), 5.19 (s, 2H), 2.43 (s, 4H), 2.42 (s, 3H), 2.40 (s, 3H), 1.59 ppm (s, 9H).

^{13}C NMR (126 MHz, DCM-d_2, 27 ºC): δ = 166.0, 164.8, 152.5, 151.1, 149.5, 138.5, 138.3, 138.0, 133.5, 132.9, 129.4, 129.2, 128.8, 128.3, 128.2, 127.9, 127.7, 127.6, 125.8, 123.3, 121.8, 120.9, 116.6, 111.3, 80.7, 71.0, 70.9, 66.4, 28.1, 21.0 ppm.

HRMS (ESI$^+$) calcd. for $C_{43}H_{45}N_2O_7^+$: 701.3221; found: 701.3220.



23

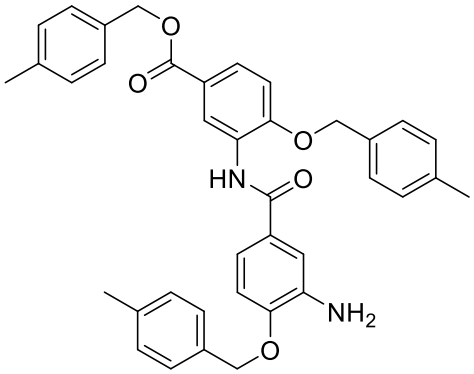
Synthesis of 3-(3-amino-4-hydroxybenzamido)-4-hydroxybenzoic acid.

A solution of 4-methylbenzyl 3-(3-((*tert*-butoxycarbonyl)amino)-4-((4-methylbenzyl)oxy)benzamido)-4-((4-methylbenzyl)oxy)benzoate (**28**, 50 mg, 0.07 mmol, 1 eq.) in DCM (2 ml) was prepared, then TFA (2 mL) was added and the solution was stirred in a vial at 65 °C for 60 min. The solvent was removed under reduced pressure and the crude mixture was purified by preparative HPLC to give the title product as the TFA salt (24.5 mg, 0.06 mmol, 86%). Storage under reduced pressure (<1 mbar) over three days gave the free amine.

^{1}H NMR (500 MHz, DMSO-d_6, 27 ºC): δ = 12.52 (br s, 1H), 10.77 (br s, 1H), 9.26 (s, 1H), 8.40 (d, *J* = 2.1 Hz, 1H), 7.64 (dd, *J* = 8.4, 2.1 Hz, 1H), 7.60 (br s, 1H), 7.52–7.57 (m, 1H), 6.98 (d, *J* = 8.4 Hz, 1H), 6.95 (dd, *J* = 8.3, 1.4 Hz, 1H), 3.81 ppm (br s, 2H).

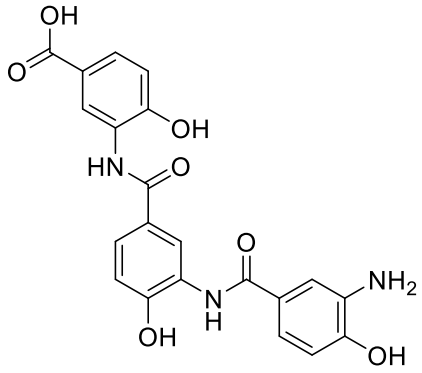
^{13}C NMR (126 MHz, DMSO-d_6, 27 ºC): δ = 167.5, 165.1, 153.4, 151.4, 127.5, 127.3, 126.4, 125.8, 125.2, 123.4, 121.9, 119.5, 115.7, 115.2 ppm.

HRMS (ESI$^+$) calcd. for $C_{14}H_{13}N_2O_5^+$: 289.0819; found: 289.0817.



29

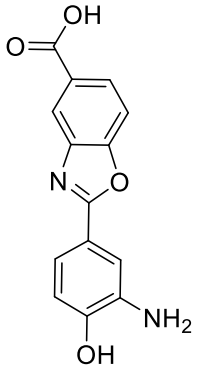
Synthesis of 4-methylbenzyl 3-(3-amino-4-((4-methylbenzyl)oxy)benzamido)-4-((4-methylbenzyl)oxy)benzoate.
A solution of 4-methylbenzyl 3-(3-((*tert*-butoxycarbonyl)amino)-4-((4-methylbenzyl)oxy)benzamido)-4-((4-methylbenzyl)oxy)benzoate (**28**, 50 mg, 0.071 mmol) in DCM (2 mL) was prepared, then TFA (100 µL) was added and the solution was stirred at room temperature and reaction progress was monitored by HPLC-MS. After 4 h the solvent was evaporated and the crude mixture was purified by preparative HPLC to give the title product as the TFA salt (45.4 mg, 0.064 mmol, 89%). Storage under reduced pressure (<1 mbar) over three days gave the free amine.
^{1}H NMR (600 MHz, DMSO-d_6, 27 ºC): δ = 9.18 (s, 1H), 8.47 (d, *J* = 2.2 Hz, 1H), 7.77 (dd, *J* = 8.7, 2.2 Hz, 1H), 7.39–7.41 (m, 2H), 7.39 (d, *J* = 4.3 Hz, 2H), 7.35 (d, *J* = 8.0 Hz, 2H), 7.27 (d, *J* = 2.2 Hz, 1H), 7.24 (d, *J* = 8.8 Hz, 1H), 7.21 (d, *J* = 7.4 Hz, 4H), 7.18 (d, *J* = 7.8 Hz, 2H), 7.13 (dd, *J* = 8.3, 2.1 Hz, 1H), 6.96 (d, *J* = 8.4 Hz, 1H), 5.28 (s, 2H), 5.26 (s, 2H), 5.16 (s, 2H), 5.11 (br s, 1H), 2.31 (s, 7H), 2.30 ppm (s, 3H).
^{13}C NMR (151 MHz, DMSO-d_6, 27 ºC): δ = 165.7, 165.6, 154.2, 148.6, 138.1, 137.9, 137.7, 137.5, 134.4, 133.8, 133.8, 129.5, 129.5, 129.4, 128.6, 128.2, 128.1, 127.9, 127.4, 127.3, 124.5, 122.2, 116.3, 113.8, 113.1, 111.8, 70.5, 69.8, 66.4, 21.3 ppm.
HRMS (ESI$^+$) calcd. for $C_{38}H_{37}N_2O_5^+$: 601.2697; found: 601.2696.



24

Synthesis of 3-(3-(3-amino-4-hydroxybenzamido)-4-hydroxybenzamido)-4-hydroxybenzoic acid.
3-((*Tert*-butoxycarbonyl)amino)-4-((4-methylbenzyl)oxy)benzoic acid (**16**, 8.92 mg, 0.025 mmol, 1.5 eq.) was suspended in DCM (100 µL), then DIPEA (6.45 mg, 8.7 µL, 0.05 mmol, 3 eq.) was added to give a clear solution. $SOCl_2$ (3.96 mg, 2.4 µL, 0.033 mmol, 2 eq.) was added and stirred at 20 °C for 10 min. All volatiles were removed under reduced pressure and DCM (100 µL), DIPEA (6.45 mg, 8.7 µL, 0.05 mmol, 3 eq.) and 4-methylbenzyl 3-(3-amino-4-((4-methylbenzyl)oxy)benzamido)-4-((4-methylbenzyl)oxy)benzoate (**29**, 10 mg, 0.016 mmol, 1 eq.) were added. The mixture was stirred at 20 °C for 1 h and the reaction progress was monitored by HPLC-MS. All volatiles were removed under reduced pressure. The crude 4-methylbenzyl 3-(3-(3-((*tert*-butoxycarbonyl)amino)-4-((4-methylbenzyl)oxy)benzamido)-4-((4-methylbenzyl)oxy)benzamido)-4-((4-methylbenzyl)oxy)benzoate was dissolved in DCM (100 µL) and TFA (100 µL), then heated to 60 °C in an sealed vial. The reaction was monitored by HPLC-MS and stopped after 165 min. The crude mixture was purified by preparative HPLC to give the title product as the TFA salt (~5 mg, ^{13}C-NMR is shown). Small impurities were detected by HPLC-HRMS and the product was purified a second time by preparative HPLC to give the title product as a white solid (0.6 mg, 0.001 mmol, 9%). ^{1}H-NMR and HPLC-HRMS showed no more impurities, but due to the small amount not all signals could be found in ^{13}C-NMR, therefore the ^{13}C-NMR after the first preparative purification is shown.
^{1}H NMR (600 MHz, DMSO-d_6, 27 ºC): δ = 12.56 (br s, 1H), 10.76 (s, 1H), 10.67 (s, 1H), 10.15–10.47 (br s, 1H), 9.33 (s, 1H), 9.30 (s, 1H), 8.42 (d, *J* = 2.3 Hz, 1H), 8.40 (d, *J* = 2.1 Hz, 1H), 7.70 (dd, *J* = 8.5, 2.3 Hz, 1H), 7.65 (dd, *J* = 8.4, 2.2 Hz, 1H), 7.45 (br s, 1H), 7.33–7.42 (br s, 1H), 7.03 (d, *J* = 8.4 Hz, 1H), 6.99 (d, *J* = 8.4 Hz, 1H), 6.85–6.88 (br d, *J* = 8.4 Hz, 1H), 6.54 ppm (br s, 3H). ^{13}C

NMR (126 MHz, DMSO-d_6, 27 ºC): δ = 167.6, 165.3, 165.3, 153.5, 152.9, 151.8, 127.5, 127.4, 126.6, 126.5, 125.6, 125.4, 125.1, 123.7, 123.6, 122.2, 121.9, 119.8, 115.8, 115.8, 115.4 ppm. HRMS (ESI$^+$) calcd. for $C_{21}H_{18}N_3O_7^+$: 424.1139; found: 424.1136.



22

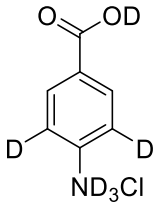
Synthesis of 2-(3-amino-4-hydroxyphenyl)benzo[*d*]oxazole-5-carboxylic acid.

A solution of 3-((*tert*-butoxycarbonyl)amino)-4-((4-methylbenzyl)oxy)benzoic acid (**16**, 176 mg, 0.49 mmol, 1 eq.) was suspended in DCM (2 mL) and DIPEA (191 mg, 257 µL, 1.48 mmol, 3 eq.) was added. Then $SOCl_2$ (88 mg, 0.74 mmol, 54 µL, 1.5 eq.) was added and the mixture was stirred at room temperature for 10 min. Then all volatile compounds were removed under reduced pressure. The solid residue was redissolved in DCM (2 mL) and DIPEA (191 mg, 257 µL, 1.48 mmol, 3 eq.) and 3-((*tert*-butoxycarbonyl)amino)-4-hydroxybenzoic acid (**15**, 100 mg, 0.39 mmol, 0.8 eq.) were added. The mixture was stirred at room temperature for 6 h and all volatile compounds were removed under reduced pressure. The solid residue was redissolved in DCM (2 mL) and TFA (2 mL) was added and the mixture was refluxed at 65 °C for 21 h. The solvents were removed under a stream of nitrogen and the crude mixture was purified by preparative HPLC to give the title product (19.1 mg, 0.071 mmol, 14%).

^{1}H NMR (600 MHz, DMSO-d_6, 26 ºC): δ = 12.12–13.86 (m, 1H), 10.55–10.92 (m, 1H), 8.24 (d, *J* = 1.1 Hz, 1H), 8.00 (dd, *J* = 8.5, 1.4 Hz, 1H), 7.82 (d, *J* = 8.5 Hz, 1H), 7.76 (br s, 1H), 7.61–7.67 (m, 1H), 6.99 ppm (br d, *J* = 8.3 Hz, 1H).

^{13}C NMR (151 MHz, DMSO-d_6, 29 ºC): δ = 167.5, 164.6, 153.4, 151.1, 142.4, 131.5, 128.1, 126.9, 122.0, 120.9, 117.5, 117.3, 115.9, 111.1 ppm.

HRMS (ESI$^+$) calcd. for $C_{14}H_{11}N_2O_4^+$: 271.0713; found: 271.0709.



30

4-(Carboxy-*d*)benzen-2,6-*d*$_2$-aminium-*d*$_3$

4-Aminobenzoic acid (250 mg, 1.82 mmol) was suspended in D_2O (7 mL) and DCl in D_2O (3 mL, 38%). The solution was degassed with argon, sealed in a microwave vial, and was heated to 110°C in a microwave for 15 h. The mixture was cooled to room temperature and the solvent was removed under reduced pressure to yield the title product as a grey powder (303.7 mg, 1.69 mmol, 93%).

HRMS (ESI$^+$) calcd. for $C_7H_6D_2NO_2^+$: 140.0675; found: 140.0674.

Bioinformatic analysis

Identification of closoxazole biosynthetic gene cluster and sequence alignments

Searches for genes encoding canonical benzoxazole biosynthetic enzymes in the *C. cavendishii* genome were done using BLAST.[9] To identify the candidate cluster *clx*, the putative functions of enzymes encoded in the *clx* locus were assigned based on BLAST and HHPred searches.[9-10] Sequence alignments were created using ClustalOmega.[11]

Genome Neighborhood Diagram

In order to provide support for our proposal that *clxA–E* (GenBank locus tags: BUB01_RS09140 – BUB01_RS09155) are the minimal set of genes required for closoxazole formation, we created Genome Neighborhood Diagrams using the Enzyme Function Initiative - Genome Neighborhood Network Tool (EFI-GNT) provided by the Enzyme Function Initiative Enzyme Similarity Tool (EFI-EST; https://efi.igb.illinois.edu/efi-gnt/).[12] For this, the corresponding amino acid sequences were obtained from UniProt and submitted to EFI-GNT using the default parameters. The local genomic region of the *clx* locus, consisting of 10 open reading frames upstream and downstream, was manually checked for genes putatively encoding enzymes associated with secondary metabolism.

Sequence Similarity Network

ClxD (NCBI accession: SHJ39914.1) and NatAM (NCBI accession: CEK42821) were used to generate the sequence similarity network (SSN) using EFI-EST (https://efi.igb.illinois.edu/efi-est/).[12] The primary networks were created at an expectation-value (e-value) of 10^{-50} for the SSN based on ClxD and 10^{-120} for that based on NatAM, the secondary and final network at e-value of 10^{-60}. The alignment score was determined empirically. The primary and secondary SSNs were analysed using the representative node network where proteins with ≥90% sequence identity were collapsed to a single node. The final network visualized sequences with 100% identity as a single node and consisted of 243 nodes. All networks were visualized using Cytoscape (v. 3.8.0) with the organic layout.[13] The accession numbers of the included sequences can be found in **Table S7**. Associated Genome Neighborhood Diagrams were generated using the default parameters and analyzed manually to search the local genomic regions adjacent to the genes encoding the NatAM and ClxD homologs.

Table S1. Strains used in this study.

Strain	Purpose, genotype	Abbreviated name	Reference
Clostridium cavendishii	wild type (DSM 21758)	-	
C. cavendishii Rif1	DSM 21758-derived, rifampicin-resistant strain	-	this study
C. cavendishii Rif1	negative control for gene inactivation, Ery^R	-	this study
C. cavendishii Rif1*ΔclxA*	*C. cavendishii* Rif1 with CRISPR/Cas9-inactivated *clxA*, Ery^R	-	this study
Escherichia coli TOP10	*general cloning strain*	-	laboratory strain
E. coli XL1-Blue	*general cloning strain*	-	laboratory strain
E. coli BL21 (DE3)	*heterologous expression strain*	-	laboratory strain
E. coli pET-28a (+)[a,c]	*E. coli* BL21 (DE3) containing pET-28a	-	this study
E. coli pETDuet-1 [b,c]	*E. coli* BL21 (DE3) containing pETDuet-1	-	this study
E. coli pET-28a- pET-Duet1	*E. coli* BL21 (DE3) containing pET28 and pETDuet	-	this study
E. coli pET28a-*clxABCDE*	*E. coli* BL21 (DE3) containing pET28a-*clxABCDE*	*E. coli* pET28a-*clxA–E*	this study
E. coli pET28-*clxBCDE*	*E. coli* BL21 (DE3) containing pET28-*clxBCDE*	*E. coli* pET-*clxBCDE*	this study
E. coli pETDuet-*clxB*	*E. coli* BL21 (DE3) containing pETDuet-*clxB*	*E. coli* pET-*clxB*	this study
E. coli pETDuet-*clxE*	*E. coli* BL21 (DE3) containing pETDuet -*clxE*	*E. coli* pET-*clxE*	this study
E. coli pETDuet-*clxBE*	*E. coli* BL21 (DE3) containing pETDuet-*clxBE*	*E. coli* pET-*clxBE*	this study
E. coli pET28-*clxACD*	*E. coli* BL21 (DE3) containing pET28-*clxACD*	*E. coli* pET-*clxACD*	this study
E. coli pET28-*clxAD* pETDuet-*clxBE*	*E. coli* BL21 (DE3) containing pET28-*clxAD and pETDuet-clxBE*	*E. coli* pET-*clxABDE*	this study
E. coli pET28-*clxAC*- pETDuet -*clxBE*	*E. coli* BL21 (DE3) containing pET28-*clxAC and pETDuet-clxBE*	*E. coli* pET-*clxABCE*	this study
E. coli pET28-*clxA*- pETDuet -*clxBE*	*E. coli* BL21 (DE3) containing pET28-*clxA and pETDuet-clxBE*	*E. coli* pET-*clxABE*	this study
E. coli pET28*clxC*- pETDuet *t*-*clxBE*	*E. coli* BL21 (DE3) containing pET28-*clxC and* pETDuet-*clxBE*	*E. coli* pET-*clxCBE*	this study
E. coli pET28-*clxAD*	*E. coli* BL21 (DE3) containing pET28-*clxAD*	*E. coli* pET-*clxAD*	this study
E. coli pET28-*clxD*	*E. coli* BL21 (DE3) containing pET28-*clxD*	*E. coli* pET-*clxD*	this study
E. coli pET28-*clxA*	*E. coli* BL21 (DE3) containing pET28-*clxA*	*E. coli* pET-*clxA*	this study
E. coli pET28-*clxC*	*E. coli* BL21 (DE3) containing pET28-*clxC*	*E. coli* pET-*clxC*	this study

a: from here on called pET28, b: from here on called pETDuet; c: for simplification of the naming of the two-plasmid system, all *E. coli* strains used for heterologous expression studies are referred to as *E. coli* pET, regardless if they carry a pET28a- or pETDuet-derived expression vectors or both.

Table S2. Plasmids used in this study.

Name	Description	Reference
pCasC	pSOS-Cas-GmR with D10A Cas9 mutation; Clostridia-adapted vector for CRISPR/Cas editing in *R. cellulolyticum*	[6]
pCas-clxA-ko	1-step CRISPR/nCas9 knockout vector for *ccav_clxA*	this study
pET28a	heterologous expression of synthetic genes, KanR, IPTG/Lactose inducible	Sigma-Aldrich
pETDuet-1	heterologous expression of synthetic genes, AmpR, IPTG/Lactose inducible	Sigma-Aldrich
pET28a-*clxA–E*	*clxABCDE* downstream of T7 polymerase promotor and RBS in pET28 (*Nco*I/*Nhe*I)	this study
pET28-*clxBCDE*	*clxBCDE* downstream of T7 polymerase promotor and RBS in pET28 (*Nco*I/*Nhe*I)	this study
pETDuet-*clxB*	*clxB* downstream of T7 polymerase promotor and RBS in pETDuet (*Nde*I/*Pac*I)	this study
pETDuet-*clxE*	*clxE* downstream of T7 polymerase promotor and RBS in pETDuet (*Nco*I/*Bam*HI)	this study
pETDuet-*clxBE*	*clxB* downstream of T7 polymerase promotor and RBS in pETDuet-*clxE* (*Nde*I/*Pac*I)	this study
pET28-*clxA*	*clxA* downstream of T7 polymerase promotor and RBS in pET28 (*Nco*I/*Nde*I)	this study
pET28-*clxAC*	*clxC* downstream of T7 polymerase promotor and RBS in pET28-*clxA* (*Nhe*I/*Bam*HI)	this study
pET28-*clxACD*	*clxCD* downstream of T7 polymerase promotor and RBS in pET28-*clxA* (*Nhe*I/*Bam*HI)	this study
pET28-*clxAD*	*clxD* downstream of T7 polymerase promotor and RBS in pET28-*clxA* (*Nhe*I/*Bam*HI)	this study
pET28-*clxD*	*clxD* downstream of T7 polymerase promotor and RBS in pET28 (*Nco*I/*Nhe*I)	this study
pET28-*clxC*	*clxC* downstream of T7 polymerase promotor and RBS in pET28 (*Nco*I/*Nde*I)	this study

AmpR – ampicillin resistance, GmR – gentamicin resistance, EryR – erythromycin resistance, KanR – kanamycin resistance, RBS – ribosome binding site

Table S3. Primers used in this study.

Name	Sequence (5' – 3')	Purpose
ccav_rpoB_F	GGATATGTAGATGATATAGATCACTTAGG	forward primer for sequencing of RRDR of *rpoB* from *C. cavendishii* Rif1
ccav_rpoB_R	CTTCTTCTTGTGCTCTTACTGTAAC	reverse primer for sequencing of RRDR of *rpoB* from *C. cavendishii* Rif1
ko_clxA_fwd	CTTATAGAAATTAAACCAATTGCTAAACC	colony-PCR Δ*clxA*
ko_clxA_rev	CAGCATCATCTTGAGCTTTGAC	colony-PCR Δ*clxA*
HB1-seq-rev	AACAAGCCATGAAAACCG	forward sequencing pCas-*clxA*-ko
pUC-ori-seq-rev	CTACGGGGTCTGACGCTCAG	reverse sequencing pCas-*clxA*-ko
pET28a-*clxABCDE* fwd	TTTAACTTTAAGAAGGAGATATACCATGTATAAATATACTAAAGATGA	cloning of pET28a-*clxABCDE*
pET28a-*clxABCDE* rev	TTTGCTGTCCACCAGTCATGCTAGCTTATACAACTGAAAACTTGT	cloning of pET28a-*clxABCDE*
pET28a-*clxBCDE* fwd	TTTAACTTTAAGAAGGAGATATACCATGATTAACAAAAAAGATTTATG	cloning of pET-*clxBCDE*
pET28a-*clxBCDE* rev	TTTGCTGTCCACCAGTCATGCTAGCTTATACAACTGAAAACTTGT	cloning of pET-*clxBCDE*
pETDuet-*clxB* fwd1	TTAAGTATAAGAAGGAGATATACATATGATTAACAAAAAAGATTT	cloning of pET-*clxB*
pETDuet-*clxB* rev1	GGTGGCAGCAGCCTAGGTTAATTAATTATTTTTCTATTATATTTTC	cloning of pET-*clxB*
pETDuet-*clxE* fwd	AACTTTAAGAAGGAGATATACCATGAGAGGTAAATACAGTAGATT	cloning of pET-*clxE*
pETDuet-*clxE* rev	GCGCGCCGAGCTCGAATTCGGATCCTTATACAACTGAAAACTTGT	cloning of pET-*clxE*
pETDuet-*clxB* fwd2	TTAAGTATAAGAAGGAGATATACATATGATTAACAAAAAAGATTT	cloning of pET-*clxBE*
pETDuet-*clxB* rev2	GGTGGCAGCAGCCTAGGTTAATTAATTATTTTTCTATTATATTTTC	cloning of pET-*clxBE*
pET28-*clxA* fwd	TTTAACTTTAAGAAGGAGATATACCATGTATAAATATACTAAAGATGA	cloning of pET-*clxA*
pET28-*clxA* rev	GTCCACCAGTCATGCTAGCCATATGTTAATCTACATATTCTACAT	cloning of pET-*clxA*
pET28-*clxC* fwd1	ATATGTAGATTAACATATGGCTAGGTGACTATTTTAGAATTATGT	cloning of pET-*clxAC*
pET28-*clxC* rev1	GTCGACGGAGCTCGAATTCGGATCCCTAGTCTTCGTAAAGTACTC	cloning of pET-*clxAC*
pET28-*clxCD* fwd	ATATGTAGATTAACATATGGCTAGGTGACTATTTTAGAATTATGT	cloning of pET-*clxACD*
pET28-*clxCD* rev	GTCGACGGAGCTCGAATTCGGATCCCTAGTCTTCGTAAAGTACTC	cloning of pET-*clxACD*
pET28-*clxC* fwd2	ACTTTAAGAAGGAGATATACCATGTTGTTAATAATCAATAATAA	cloning of pET-*clxC*
pET28-*clxC* rev2	GTCCACCAGTCATGCTAGCCATATGCTAGTCTTCGTAAAGTACTC	cloning of pET-*clxC*
pET28-*clxD* fwd2	TTTAACTTTAAGAAGGAGATATACCATGATTATAGATGCACATGCTC	cloning of pET-*clxD*
pET28-*clxD* rev2	TTTGCTGTCCACCAGTCATGCTAGCTTAACCTCTTTTCTGTAATAAATT T	cloning of pET-*clxD*

Table S4. Sequence of the cassette for CRISPR-Cas9n-based gene knockout. Italic sequence represents restriction sites; underlined sequences represent miniP4, protospacer and sgRNA; homologous arms are marked in grey; insert is marked in green.

Name	Sequence (5' – 3')	Purpose
clxA-ko	5'-*TACCGCGAGACC*<u>TTGACAAATTTATTTTTTAAAGTTAAAATTAAGTTGTCAGTAAGCCAAGTTGAAACGTTTTAG AGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTT TGAATTCTCTAGAGTCGA</u>ATTTAAAGTTAATAATTTCGAATAAATACTTACAAAATATTATTAATTTAATATTTTAAG GGTCATTTGTTAAATTATTGAATTTAATATATTGTATAAATATACTAAAGATGAAATTAAAACAAATTTGCAAATTTT GTTATTTAATAGTATAATACATTCATATTTAGCAAAAAAATACACTTTTTCAAACTTAAAGGAGGTAATTATGGACA GTAATTCCCAATTAATAACAAAATTGAATTCAGCATTACAAATTGCTACAAAAGCCAACTTTTACAAGGACAGGCT AGGTAATATTGAAATTAAATCTTTAGATGATTTTTCAAAATTACCTTTAACTACAAAAGAAGACCTTCGAAAGCTAA AACCAATGGAAGCTTTAACTGTAGATATAGAAGACCTTTTCCAATATCACGAATCTTTTGGATAAGATATCACTTG GCTTACTGAAAAAGATTTTAATGCTTATGGAGATCAACTAAATGAATTTGGAGTAAATTTCAAAAGTACAGATATT GTATTAAATAGATTCCCATATGCTATTTCAGTTCCAGCTCATATCTTTACCAATGCAATTCATAAGAAAGGAGCTT GTGTTATACCGGTTAGTAAGGCCTCAGCAATTTCACCATTAAAAAGAGTTGCTAATCTAATTTATAAACTACGCCC AAGCATTCTTACTGGAATTCCAGATGAATTAATAAAACTTAATAAAGTTGCTAAGTTTATGGACATATCCCTTAAAG ATCTTGGTTGCATTAGAGCAATATGTACTGCTGGAGAAATGCTTTCTGAAGGAAGGAAAGCTAAACTAGAAAGCA TCTTTGGAGCTAAAGTCTATAATTATTATGGATGTACAGAGTGTG*TGCGTGGAGACC*-3'	cloning pCas-*clxA*-ko

Results and Discussion

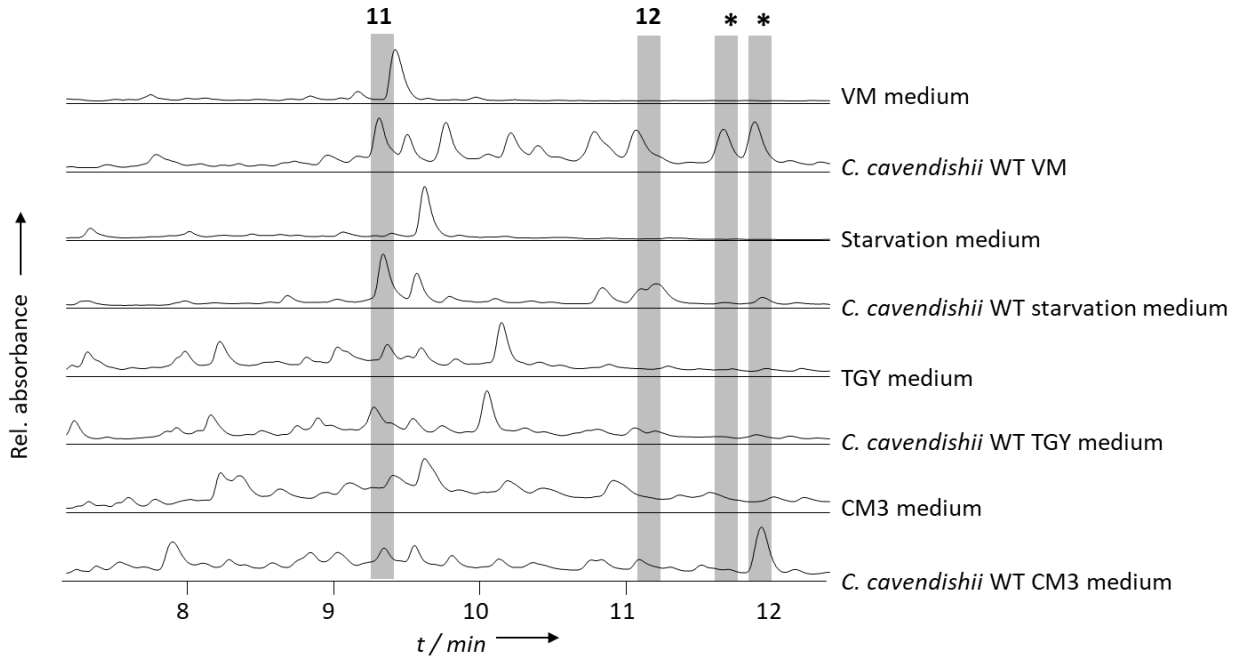


Figure S1. Metabolite profiles of extracts from *C. cavendishii* cultivated under varying conditions. Displayed are HPLC profiles when cultured in full (TGY medium), semi-defined (CM3, VM) and starvation medium (VM medium with reduced carbon source: 0.05 g L^{-1} cellobiose). No growth was observed in minimal medium (P2 medium). Asterisks (*) indicate signals corresponding to putative congeners (the earlier eluting compound is referred to as congener 1 and the later eluting compound as congener 2 in Figure S2 and S3). Based on the results of this media screen, cultivation of *C. cavendishii* Rif1 was scaled up in VM (Figure 2A)

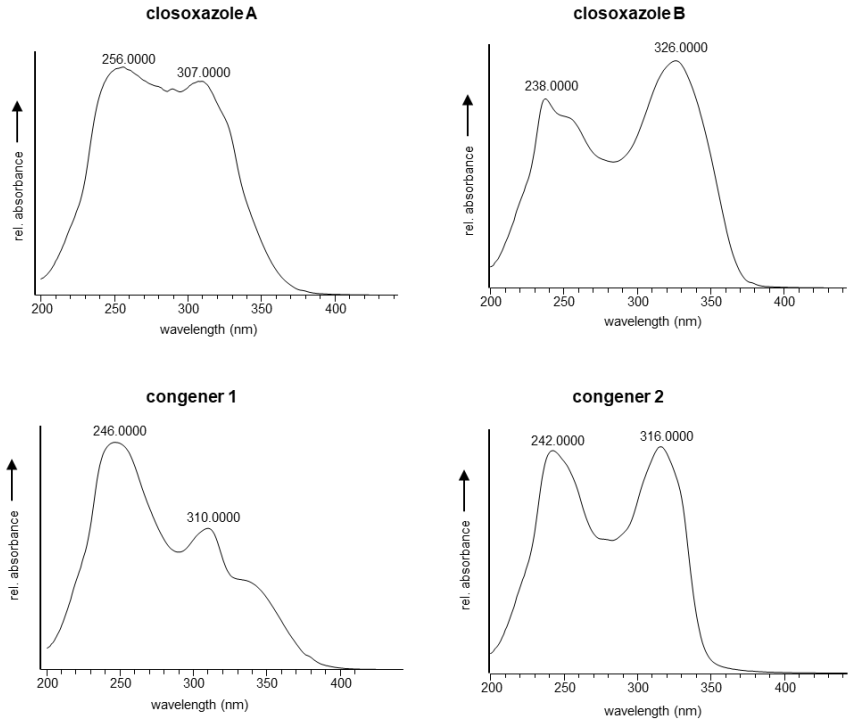


Figure S2. UV/Vis spectra of closoxazoles produced by *C. cavendishii*. UV/Vis spectra detected in ethyl acetate extract of *C. cavendishii* corresponding to closoxazole A and B and the two putative congeners.

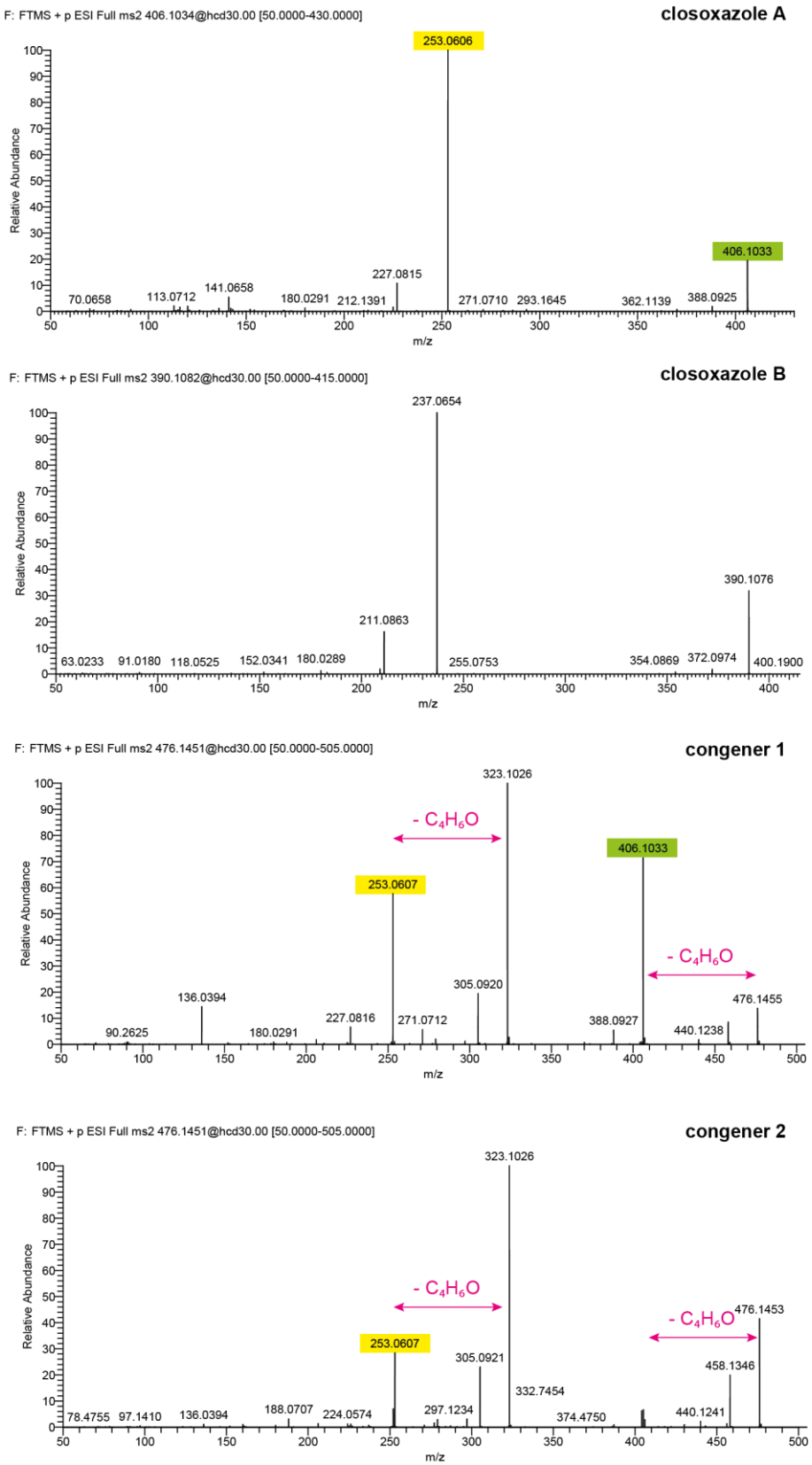


Figure S3. MS/MS fragmentation patterns of closoxazoles from *C. cavendishii*. MS/MS fragmentation patterns of *m/z* 406.1034 $[M+H]^+$ (**11**), *m/z* 390.1082$[M+H]^+$ (**12**) and *m/z* 476.1451 $[M+H]^+$ from ethyl acetate extracts of *C. cavendishii* Rif1 cultures. Fragments that indicate a shared core structure of the respective compound are marked in green and yellow. The fragmentation patterns from the respective $[M+H]^+$ ionic species of the later eluting pair of metabolites indicate a loss of *m/z* 70.0419 (calculated sum formula C_4H_6O, marked in pink), suggesting them to be *N*- or *O*-acylated derivatives of closoxazole A.

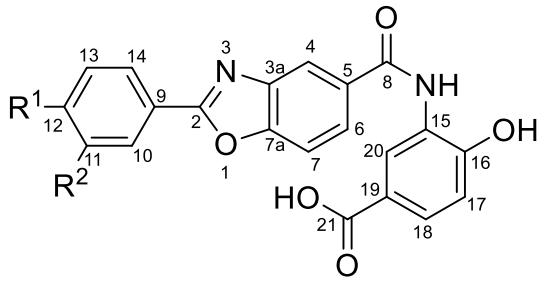


Figure S4. Alternative C-atom numbering of **11** and **12.** The carbon atoms of the closoxazoles are numbered according to previously reported benzoxazole-based natural products.

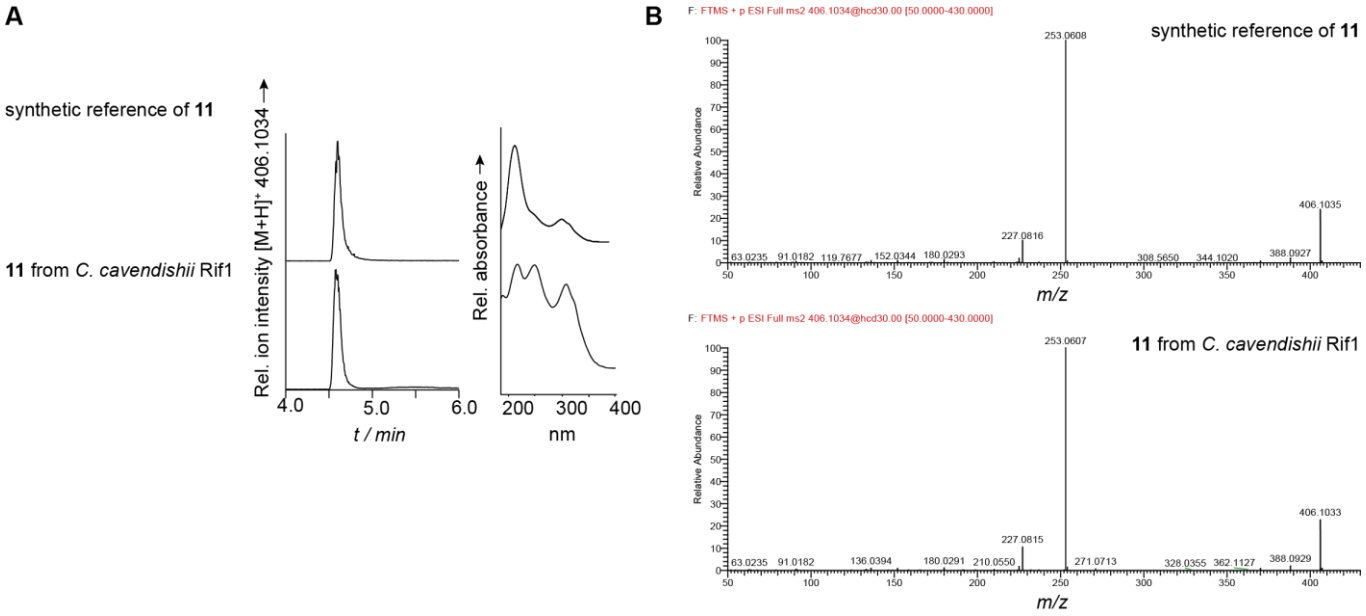


Figure S5. Comparison of the synthetic reference compound of **11** and **11** isolated from *C. cavendishii* Rif1**.** A) LC-HRMS with extracted ion chromatograms (EICs) corresponding to the $[M+H]^+$ ionic species of closoxazole A and the corresponding UV/Vis spectra. B) MS/MS fragmentation patterns of the $[M+H]^+$ ionic species of **11**.

Table S5. NMR shifts of closoxazole A (**11**) isolated from the native producer and its synthetic reference in DMSO-d_6

11 from *C. cavendishii* Rif1	**Synthetic Reference of 11**	**Difference**
δC [ppm]	δC [ppm]	
110.41	110.51	-0.10
112.74	113.09	-0.36
114.49	114.62	-0.14
115.56	115.62	-0.05
117.06	117.15	-0.09
117.08	117.48	-0.40
118.56	118.62	-0.06
121.43	121.49	-0.06
124.64	124.74	-0.10
125.59	125.63	-0.05
126.31	126.40	-0.09
127.75	127.85	-0.10
131.04	131.11	-0.07
137.40	137.02	0.38
142.01	142.04	-0.03
148.26	148.49	-0.23
152.13	152.20	-0.06
154.07	154.15	-0.09
164.84	164.86	-0.02
165.11	165.22	-0.10
167.08	167.16	-0.08

Determination of *clx* gene cluster boundaries

Genome Neighborhood Diagrams showed that the ten open reading frames upstream and downstream were not conserved outside *C. cavendishii*, suggesting that *clxA*–*clxE* are sufficient for the biosynthesis of closoxazoles.

Table S6. Annotated functions of open reading frames of the *clx* biosynthetic gene cluster and the surrounding genomic region. ORFs of *clx* are shaded in grey.

Name	Locus tag	Size	Annotated function	Protein homolog and origin [a]	Percent identity/similarity (%)
	BUB01_RS09125	598 aa	Thiamine pyrophosphate-binding protein		
	BUB01_RS09130	217 aa	HAD hydrolase-like protein		
clxA	**BUB01_RS09135**	**440 aa**	**PaaK-like**	NatL2 *Streptomyces* sp. Tü 6176	27.1/44.0
clxB	**BUB01_RS09140**	**375 aa**	**3-Amino-4-hydroxybenzoic acid synthase**	GriH *Streptomyces griseus*	37.9/54.9
clxC	**BUB01_RS09145**	**442 aa**	**PaaK-like**	NatL2 *Streptomyces* sp. Tü 6176	22.8/37.4
clxD	**BUB01_RS09150**	**255 aa**	**Amidohydrolase**	NatAM *Streptomyces* sp. Tü 6176	10.1/16.3
clxE	**BUB01_RS09155**	**274 aa**	**2-Amino-3,7-dideoxy-D-threo-hept-6-ulosonate synthase**	GriI *Streptomyces griseus*	31.7/51.8
	BUB01_RS09160	189 aa	Hypothetical protein		
	BUB01_RS09170	328 aa	Gfo/Idh/MocA family oxidoreductase		

[a] Selected homologs reported to be involved in benzoxazole formation or 3,4-AHBA biosynthesis.

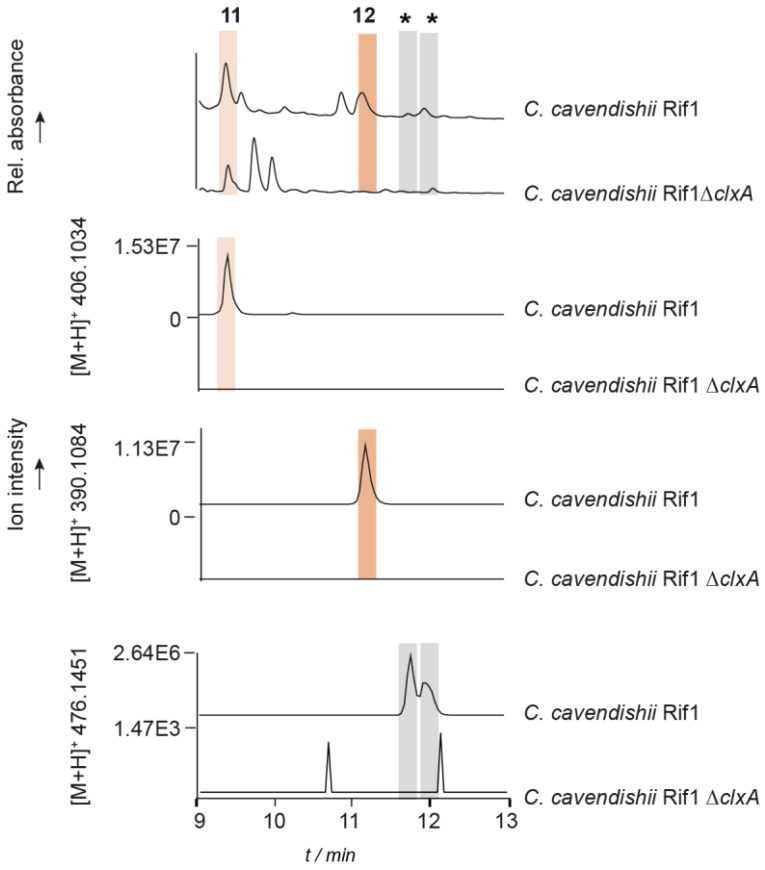


Figure S6. Detection of closoxazoles in *C. cavendishii* Rif1 and *C. cavendishii* Rif1Δ*clxA* extracts. Shown are HPLC chromatograms and EICs corresponding to the $[M+H]^+$ ionic species of closoxazoles A, B (**11, 12**) and putative congeners (marked with asterisks) from crude extracts of *C. cavendishii* Rif1 and from *C. cavendishii* Rif1Δ*clxA*. EICs of the calculated exact mass of the compounds are shown with *m/z* ± 5 ppm.

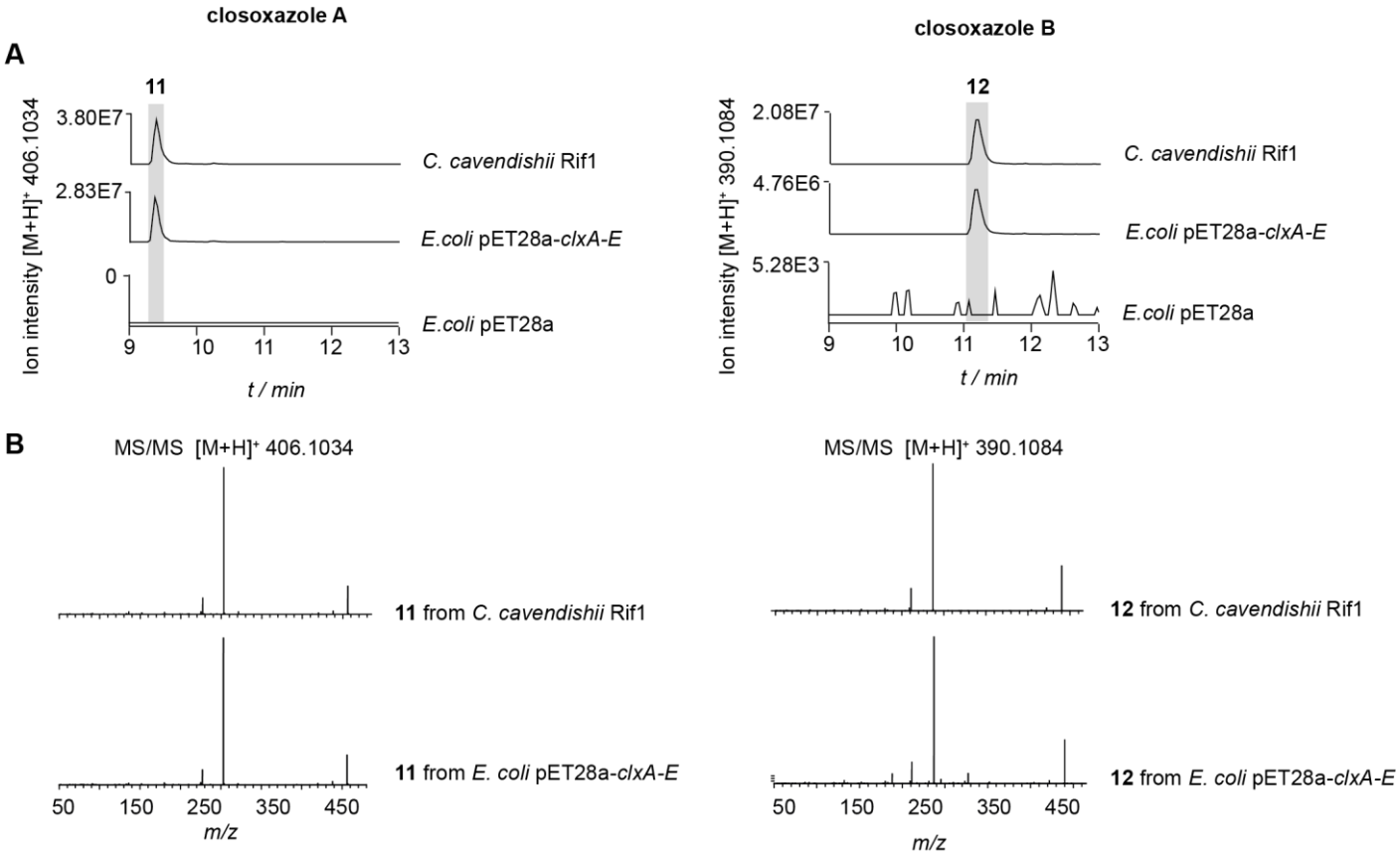


Figure S7. Production of closoxazole A and B by *C. cavendishii* Rif1 and *E. coli* pET28a-*clxA–E*. A) Extracted ion chromatograms (EICs) corresponding to the $[M+H]^+$ ionic species of closoxazoles A (**11**) and B (**12**) from crude extracts of the native producer *C. cavendishii* Rif1 and from *E. coli* pET28a-*clxA–E* (heterologous producer). EICs of the calculated exact mass of the compounds are shown with m/z ± 5 ppm. B) MS/MS fragmentation patterns of **11** and **12** isolated from *C. cavendishii* Rif1 and *E. coli* pET28a-*clxA–E*.

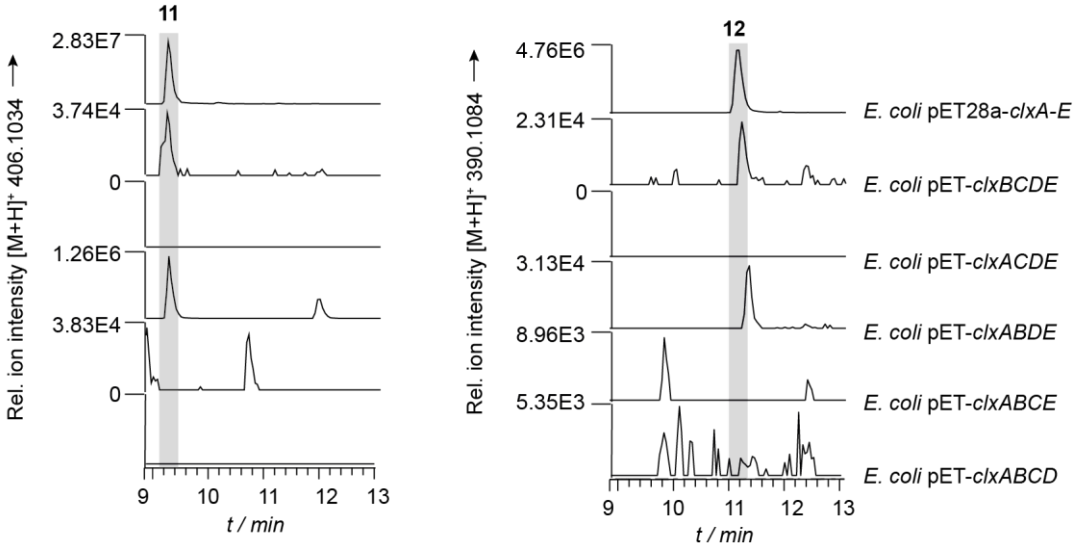


Figure S8. Production of closoxazole A (**11**) and B (**12**) by *E. coli* encoding the listed combinations of *clx* biosynthetic genes. Shown are EICs corresponding to the $[M+H]^+$ ionic species of **11** and **12** from crude extracts of *E. coli* expressing the full *clx* gene cluster and the *clx* gene cluster missing each of the genes in turn. EICs of the calculated exact mass of the compounds are shown with *m/z* ± 5 ppm. Trace amounts of closoxazole A and B can be detected in the metabolic profile of *E. coli* pET-*clxBCDE* (lacking ClxA) and reduced production in the metabolic profile of *E. coli* pET-*clxABDE* (lacking ClxC). This is probably due to promiscuous activity of either ClxC or ClxA, respectively, or other ANL superfamily members encoded in the *E. coli* genome.

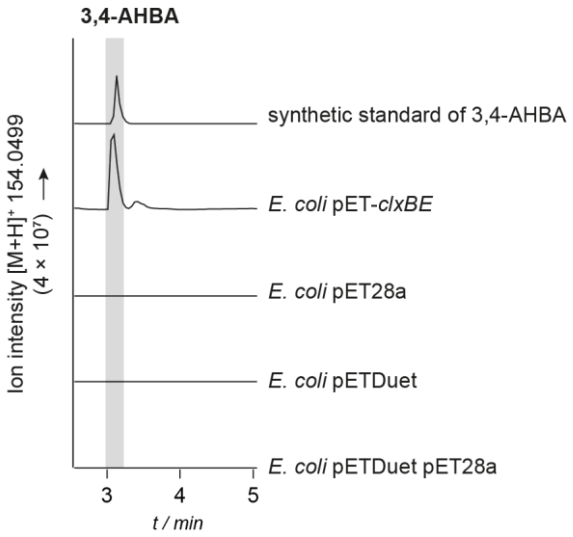


Figure S9. Production of 3,4-AHBA by *E. coli* pET*clxBE*. Shown are EICs corresponding to the $[M+H]^+$ ionic species of 3,4-AHBA: 3,4-AHBA synthetic standard, and 3,4-AHBA from crude extracts of *E. coli* pET-*clxBE* and *E. coli* carrying one or both of the relevant empty vectors used in this study.

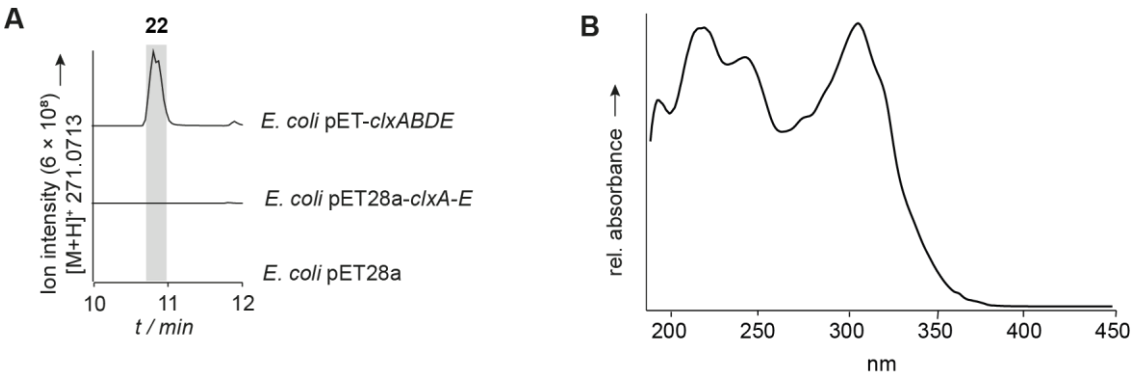


Figure S10. Detection of benzoxazole dimer **22** in *E. coli* pET-*clxABDE* extracts. Extracted ion chromatograms (EIC) of $[M+H]^+$ 271.0713, with the corresponding sum formula ($C_{14}H_{10}N_2O_4$) matching that of a benzoxazole derived from two 3,4-AHBA units (**22**), are shown for organic extracts of *E. coli* pET-*clxABDE* (lacking ClxC) and *E. coli* pET28a-*clxA–E* (complete *clx* gene cluster). EICs of the calculated exact mass of the compounds are shown with *m/z* ± 5 ppm. B) Absorption spectrum of **22** detected from organic culture extracts of *E. coli* pET-*clxABDE*.

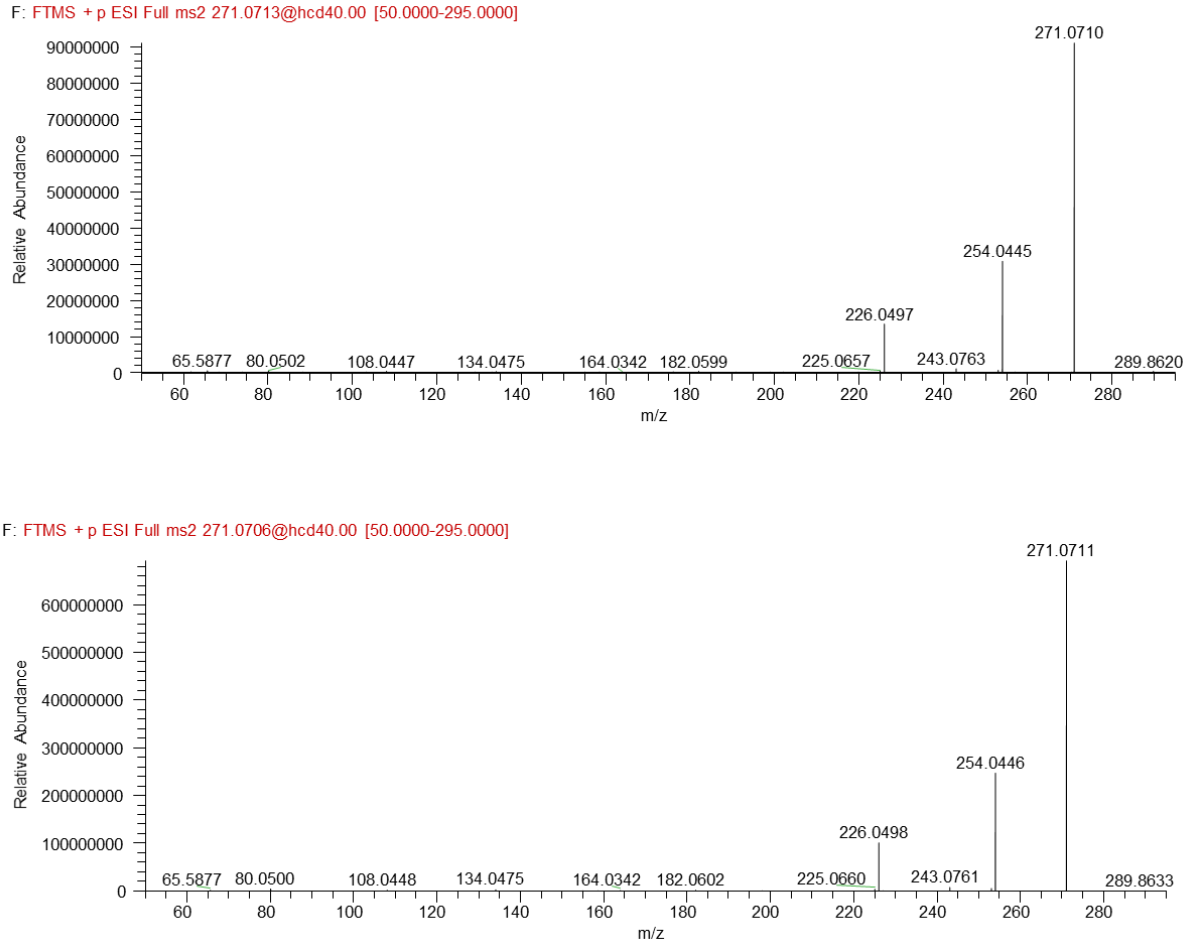


Figure S11. MS/MS fragmentation patterns of synthetic and *E. coli*-derived benzoxazole dimers **22**. MS/MS fragmentation patterns were generated from the respective $[M+H]^+$ ionic species of **22** detected in the HRMS profile of the synthetic reference (upper panel) and crude extract from *E.coli* pET-*clxABDE* (lower panel).

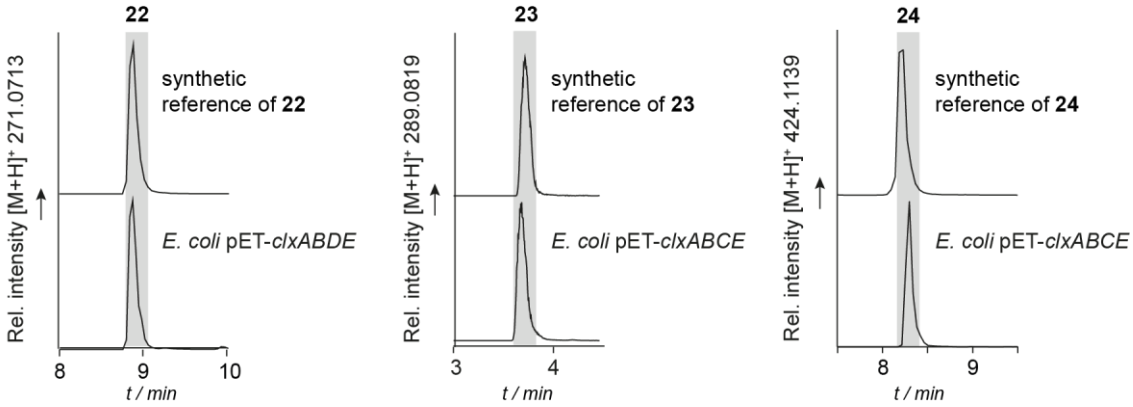


Figure S12. HR-LCMS analysis of synthetic reference of **22**–**24** and heterologously produced **22**–**24.** Shown are EICs corresponding to the $[M+H]^+$ ionic species of benzoxazole dimer **22**, the amide dimer **23** and trimer **24** of the synthetic standards, as well as from crude extracts of *E. coli* pET-*clxABCE* (lacking *clxD*) and *E. coli* pET-*clxABDE* (lacking *clxC*).

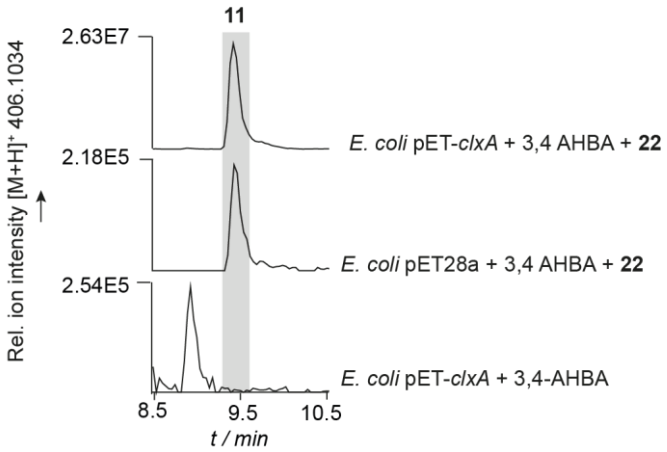


Figure S13. Supplementation of *E. coli* pET-*clxA* with the benzoxazole dimer 22. Shown are EICs corresponding to the $[M+H]^+$ ionic species of closoxazole A (**11**) from crude extracts of *E. coli* pET-*clxA* and *E. coli* pET28a (empty vector control) both supplemented with **22** and 3,4-AHBA, and *E. coli* pET-*clxA* supplemented with 3,4-AHBA alone.

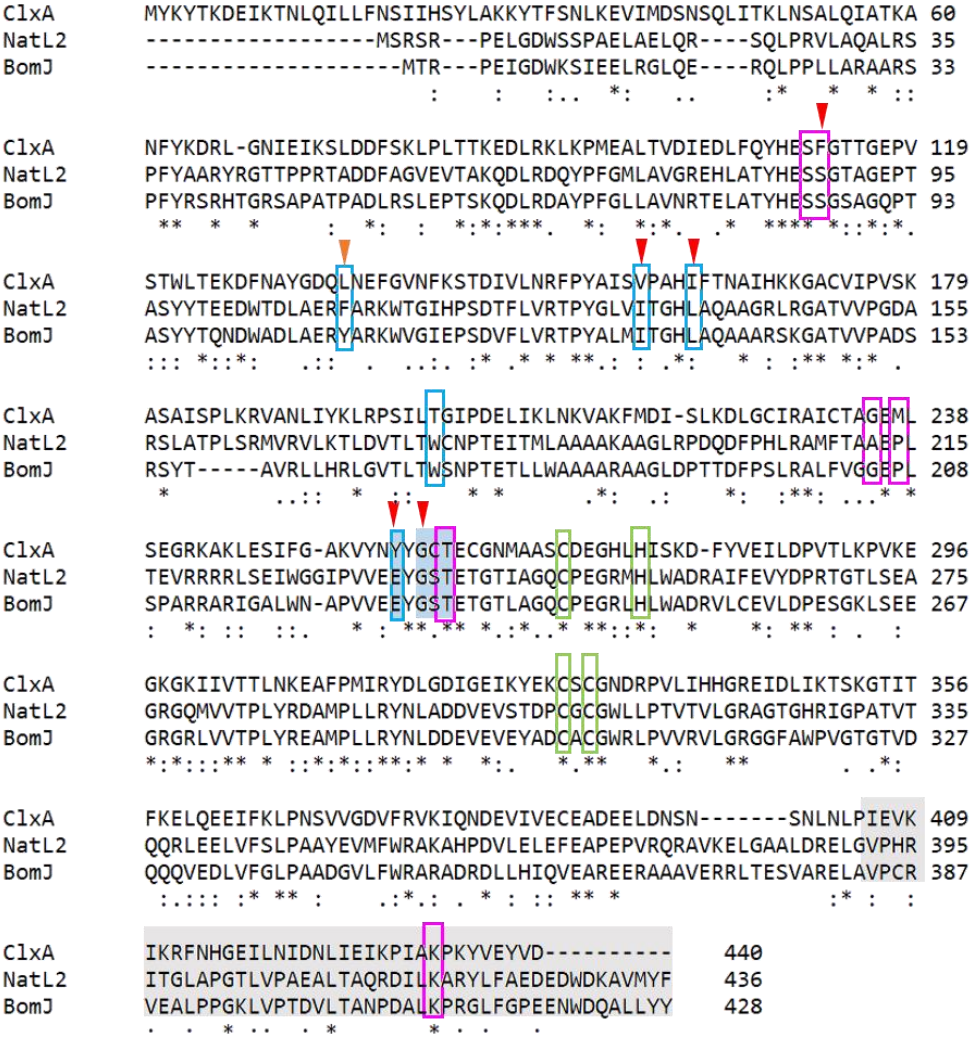


Figure S14. Sequence alignment of ClxA (*C. cavendishii*), NatL2 (*Streptomyces* sp. Tü6176) and BomJ (*S.* sp. NRRL12068). The functions of the residues are derived from their role in structurally characterized NatL2[14]. The sequence alignment shows that the overall sequence of ClxA is similar to that of 3-HAA-using ligases. When it comes to specific features, ClxA possesses the C-terminal extension (marked in grey) similar to that observed in NatL2[14] (compared to other PaaK ligases), albeit it is ten amino acids shorter. The structurally relevant Zn^{2+} has been described as a unique feature of NatL2; the residues that coordinate zinc in NatL2 are also conserved in ClxA. Furthermore, Lys418, a crucial residue of the nucleotide binding site in NatL2, is conserved in ClxA. Lastly, residues Thr239 and Gly237, which are involved in 3-HAA binding by NatL2, are conserved in ClxA. However, there are some obvious differences between ClxA and NatL2. For example, Pro214 and Ser88 of the NatL2 nucleotide binding pocket are not conserved in ClxA, with a Met and a Phe residue, respectively, being found at the comparable positions. ClxA also differs from NatL2 in some key residues involved in 3-HAA binding, substitutions which could conceivably influence substrate specificity. When it comes to Glu235 of NatL2, which interacts with the hydroxyl group of 3-HAA, ClxA possesses a Tyr residue at the comparable position. Furthermore, Ser238 of NatL2, which is involved in stacking interaction with the 3-HAA benzene ring, is not conserved in ClxA, which instead has a Cys residue at the comparable position.

Alignment was prepared with ClustalOmega. Residues involved in 3-HAA binding are shaded in light blue, residues surrounding the 3-HAA binding pocket are marked with a blue box. Key residues involved in Zn^{2+} coordination are marked with a green box, residues involved in nucleotide binding are marked with a pink box. Key residues different in ClxA compared to NatL2 are marked with a red triangle, orange triangle if the residue is not conserved in any of the sequences. C-terminal extension is marked in grey. Consensus symbols as follows: asterisk (*) indicates fully conserved residues, (:) indicates conservation between groups of strongly similar properties and (.) indicates conservation between groups of weakly similar properties.

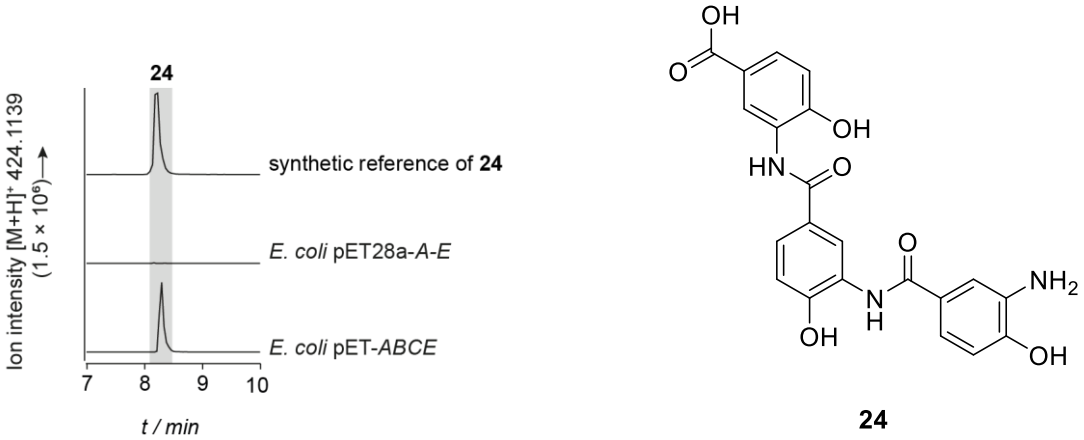


Figure S15. Detection of the trimer **24** produced by *E. coli* pET-*clxABCE*. Shown are EICs corresponding to the $[M+H]^+$ ionic species of synthetic trimer **24** and **24** detected in crude extracts of *E. coli* pET28a-*clxA–E* (expressing the full *clx* biosynthetic gene cluster) and *E. coli* pET-*clxABCE* (lacking *clxD*).

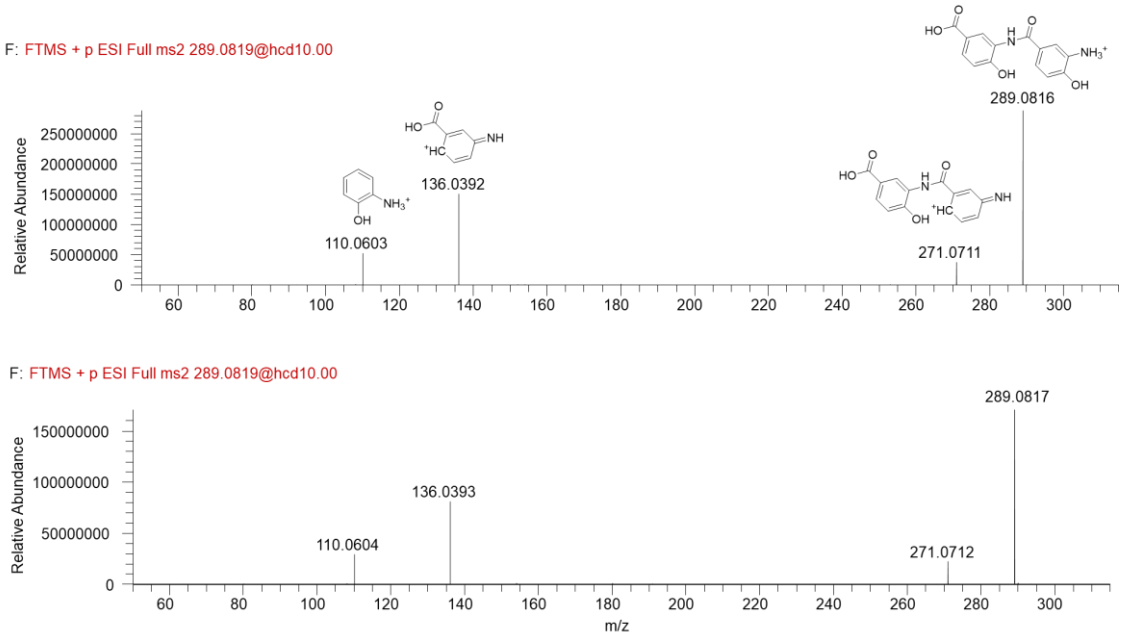


Figure S16. MS/MS fragmentation patterns of synthetic and heterologously produced amide dimer **23**. MS/MS fragmentation patterns were generated from the respective $[M+H]^+$ ionic species of **23** detected in the HR-MS profile of the synthetic reference (upper panel) and crude extract of *E.coli* pET-*clxABCE* (lower panel).

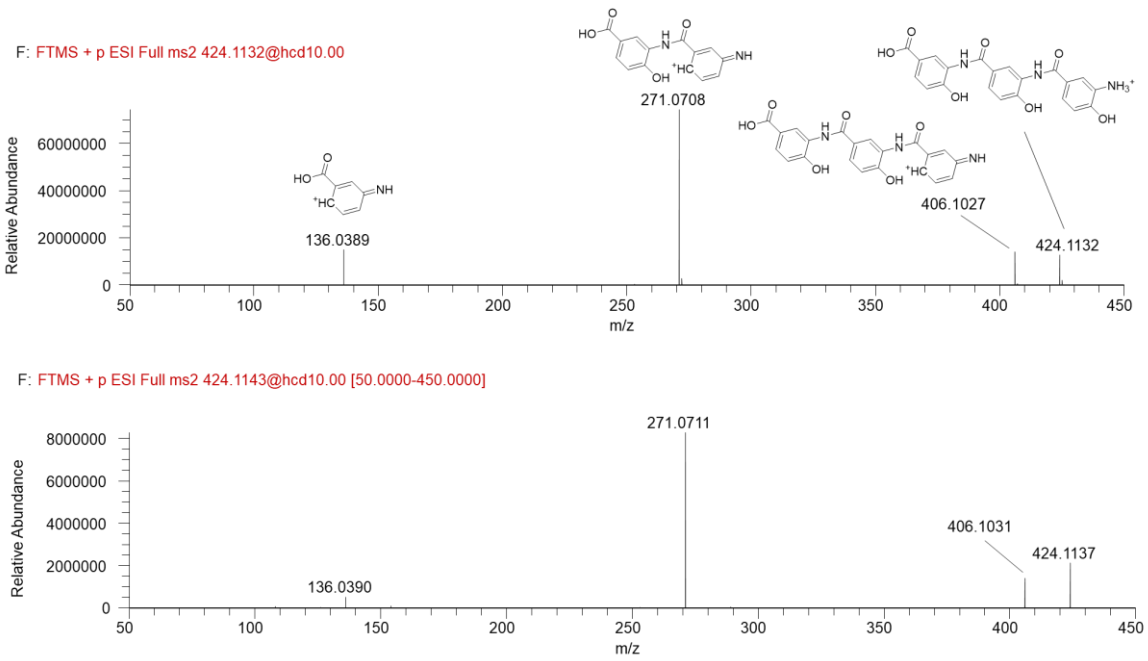


Figure S17. MS/MS fragmentation patterns of synthetic and heterologously produced trimer **24.** MS/MS fragmentation patterns were generated from the respective $[M+H]^+$ ionic species of **24** detected in the HR-MS profile of the synthetic reference (upper panel) and crude extract of *E.coli* pET-*clxABCE* (lower panel).

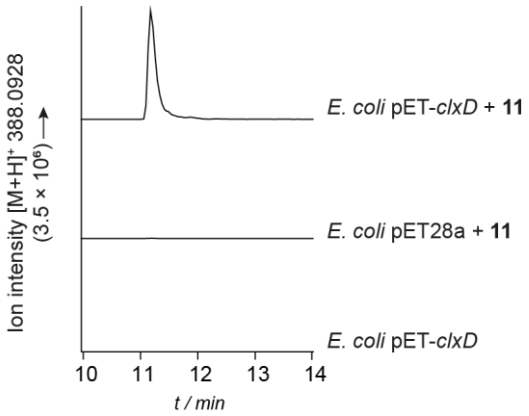


Figure S18. Supplementation of *E. coli* pET-*clxD* with the closoxazole A (**11**). Shown are EICs corresponding to the $[M+H]^+$ ionic species of which the calculated sum formula corresponds to a compound derived from **11**, which contains two instead of one benzoxazole ring (calcd. for $C_{21}H_{14}N_3O_6$: 388.0928, found: 388.0924). Analysis of crude extracts of *E. coli* pET-*clxD* and *E. coli* pET-28a (empty vector control) supplemented with **11** and unsupplemented *E. coli* pET-*clxD*. The compound elutes later than **11**, indicating a less polar molecule compared to **11**, as expected if formation of a second heterocycle occurs. Minor levels of the putative double benzoxazole (1.46×10^4) are detectable in *E. coli* pET-28a supplemented with **11**, probably due to promiscuous/off-target activity of amidohydrolase superfamily members encoded in the host genome.

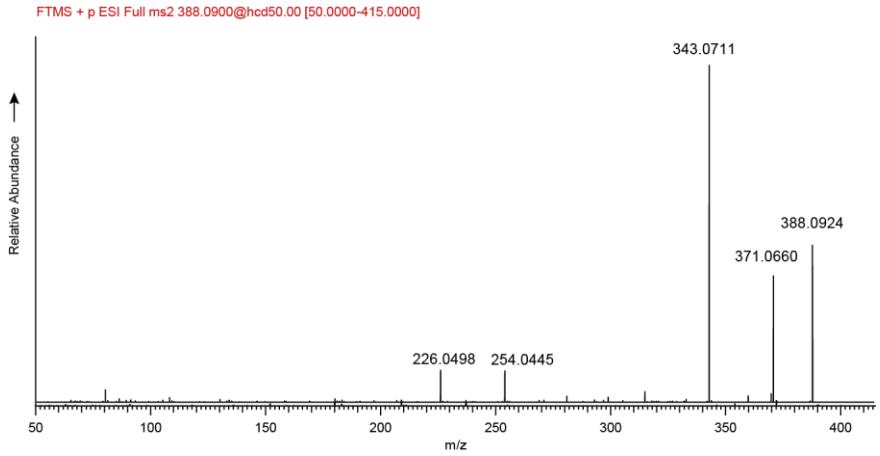


Figure S19. MS/MS fragmentation pattern of the putative benzoxazole carrying two heterocycles. Displayed are the $[M+H]^+$ ionic species of which the calculated sum formula corresponds to a closoxazole-derived compound harboring two benzoxazole heterocycles. The compound was detected in the crude extract of a culture of *E. coli* pET-*clxD* supplemented with **11**. Fragmentation occurs only under high collision energy as observed for benzoxazole **22**. The fragment pattern observed shows similar mass losses (loss of 45.0219 and 17.0265) as observed for **22** (see Figure S11).

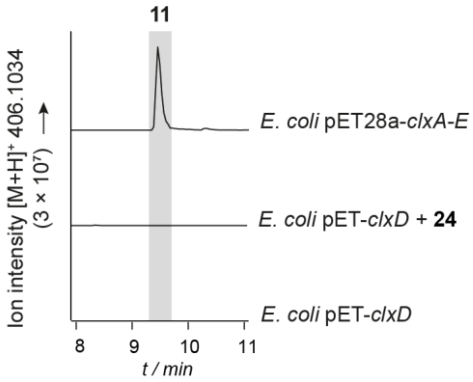


Figure S20. Supplementation of *E. coli* pET-*clxD* with the trimer **24**. Shown are EICs corresponding to the $[M+H]^+$ ionic species of closoxazole A (**11**) from crude extracts of *E. coli* pET28a-*clxA–E* (expressing the full closoxazole biosynthetic gene cluster), *E. coli* pET-*clxD* supplemented with the trimer **24** and *E. coli* pET-*clxD* (control).

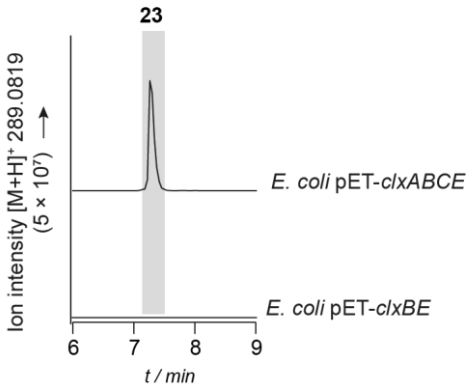


Figure S21. Detection of amide dimer **23** in crude extracts of *E. coli* pET-*clxBE*. Shown are EICs corresponding to the $[M+H]^+$ ionic species of dimer **23** from crude extracts of *E. coli* pET-*clxABCE* (lacking the heterocyclase ClxD) and *E. coli* pET-*clxBE* (only 3,4-AHBA biosynthetic genes). No production of dimer **23** was detected for *E. coli* pET-*clxBE*.

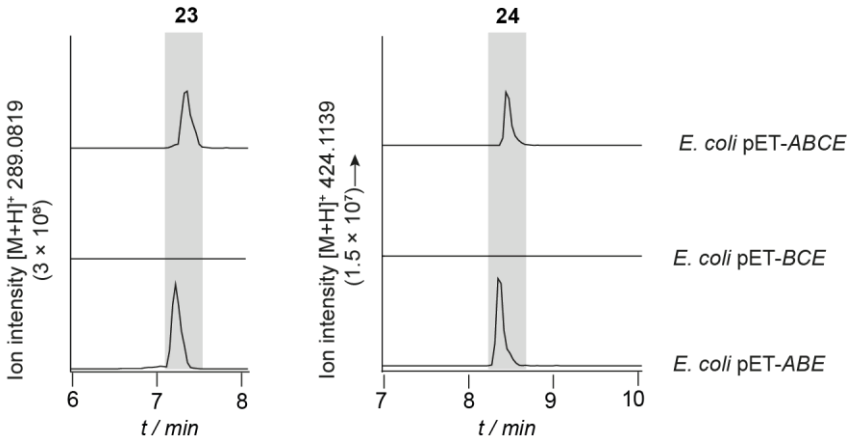


Figure S22. Detection of amide conjugates composed of two and three 3,4-AHBA units. The strain *E. coli* pET-*clxABCE* (lacking ClxD) produces the amide dimer **23** and trimer **24**. *E. coli* pET-*clxBCE* (lacking the first ligase ClxA and ClxD) does not produce amide dimer **23** and trimer **24**. *E. coli* pET-*clxABE* (lacking the second ligase ClxC and ClxD) produces both compounds. Shown are EICs corresponding to the [M+H]+ ionic species of the dimer (**23**) and trimer (**24**).

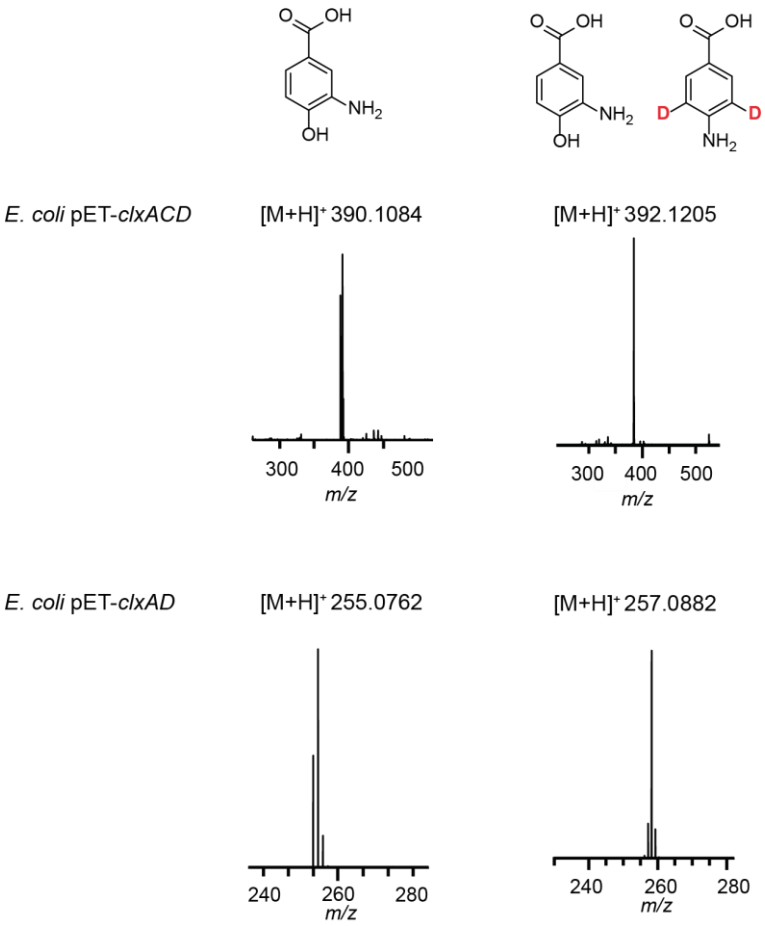


Figure S23. Detection of deuterated PABA incorporation into closoxazole B and its respective dimer benzoxazole. *E. coli* expressing either ClxD and the ligase ClxA (*E. coli* pET-*clxAD*), or ClxD and both ligases ClxA and ClxC (*E. coli* pET-*clxACD*), was supplemented with either synthetic 3,4-AHBA or synthetic 3,4-AHBA and deuterated PABA (PABA-d_2). Incorporation of PABA-d_2 is demonstrated by a mass shift of 2 amu. Upper panel: ESI-MS spectra (positive ion mode) from organic culture extracts of **12** ($[M+H]^+$ 390.1084) and **12** with the predicted 2 amu mass shift ($[M+H]^+$ 392.1205), being derived from two units 3,4-AHBA and one unit PABA or PABA-d_2. Lower panel: ESI-HRMS spectra (positive ion mode) from organic culture extracts of the dimer benzoxazole ($[M+H]^+$ 255.0762) derived from one unit PABA and one unit 3,4-AHBA and the dimer benzoxazole with the predicted 2 amu mass shift ($[M+H]^+$ 257.0882), derived from one unit 3,4-AHBA and one unit PABA-d_2. PABA: *p*-aminobenzoic acid.

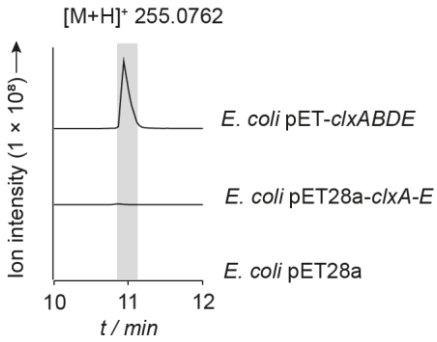


Figure S24. Detection of a putative benzoxazole dimer derived from PABA and 3,4-AHBA. Extracted ion chromatograms (EIC) of $[M+H]^+$ 255.0762, which has a deduced sum formula ($C_{14}H_{10}N_2O_3$) matching that of a benzoxazole derived from one PABA unit and one 3,4-AHBA unit, are shown for organic extracts of *E. coli* pET-*clxABDE* (lacking *clxC*) and E. coli pET-*clxABDE* (complete *clx* gene cluster). EICs of the calculated exact mass of the compounds are shown with *m/z* ± 5 ppm.

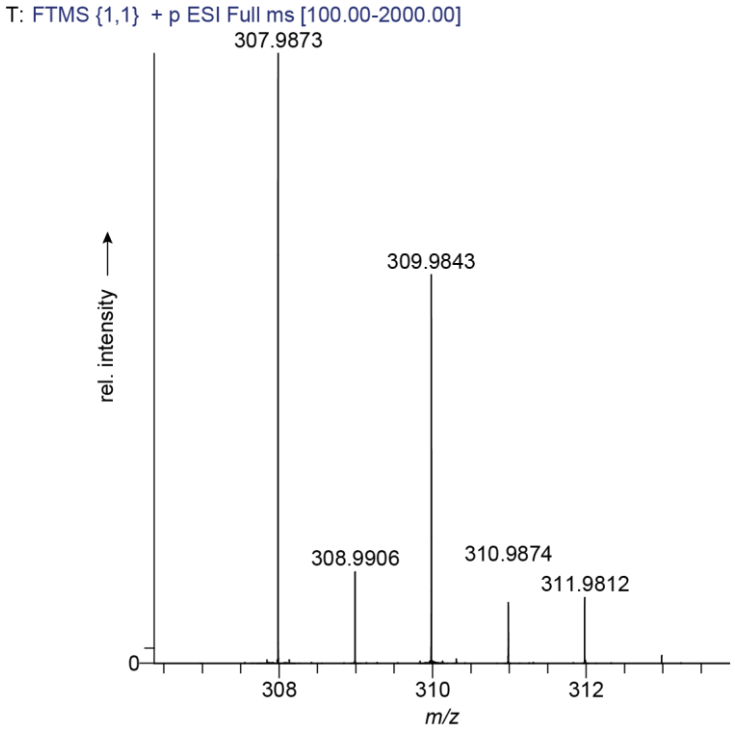


Figure S25. ESI-MS (positive ion mode) of isolated tafamidis analogue **25**.

NMR Spectra

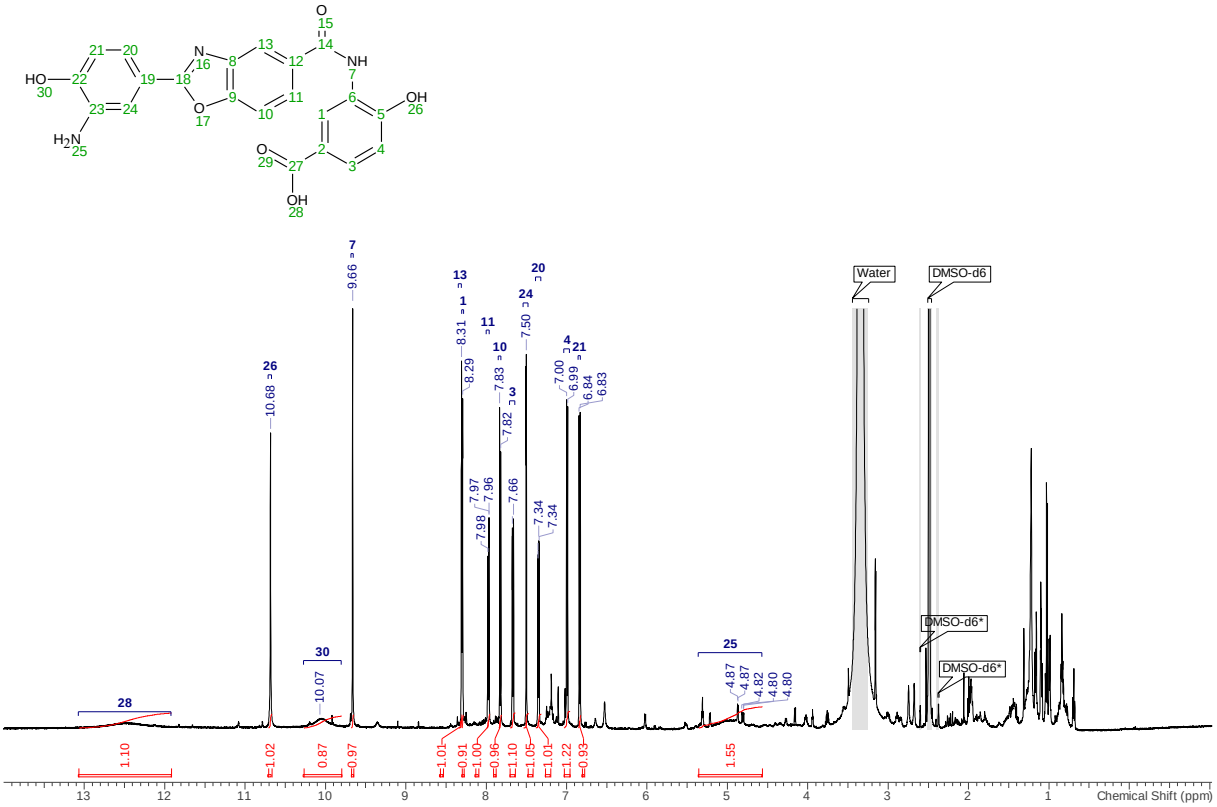


Figure S26. ^{1}H-NMR spectrum of **11** isolated from *C. cavendishii* Rif1. Itsmall impurities which lead to signals in the range of δ = 0.8–3.1 ppm.

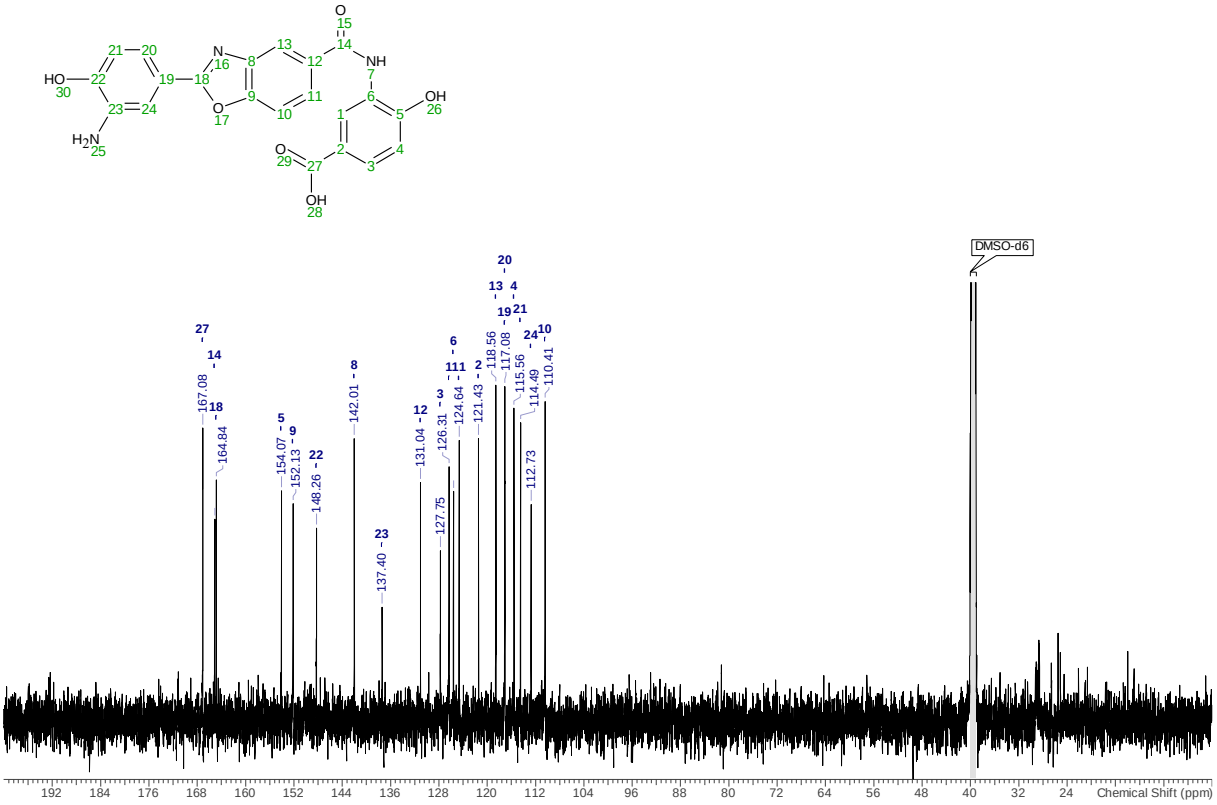


Figure S27. ^{13}C-NMR spectrum of **11**.

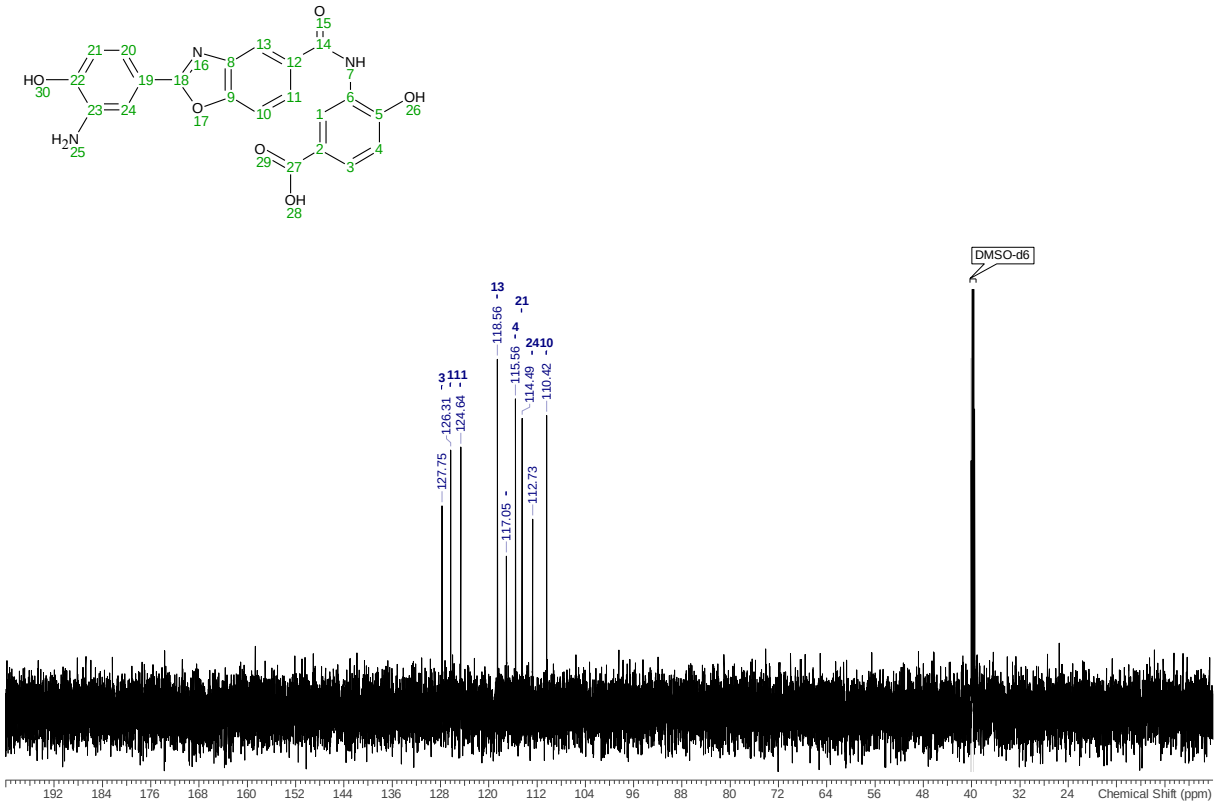


Figure S28. DEPT-135 NMR spectrum of **11.**

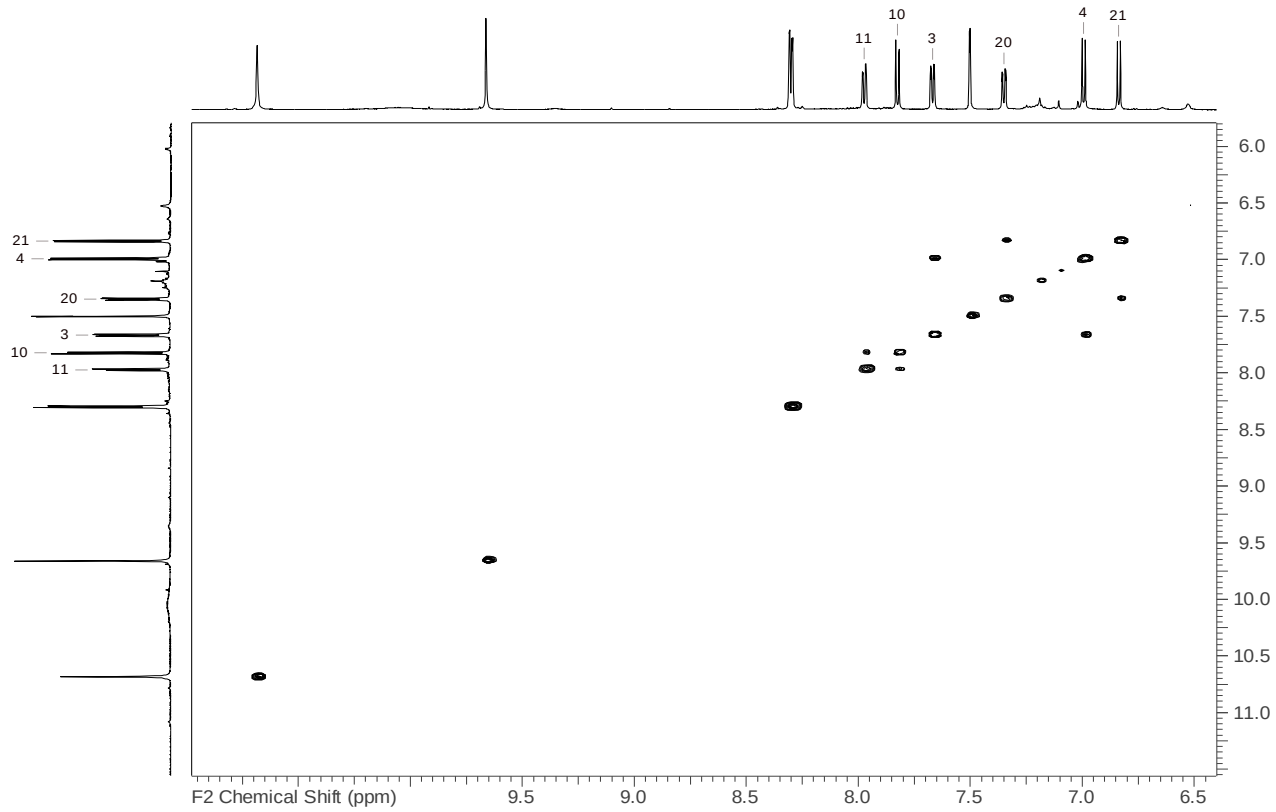


Figure S29. ^{1}H, ^{1}H COSY NMR spectrum of **11**.

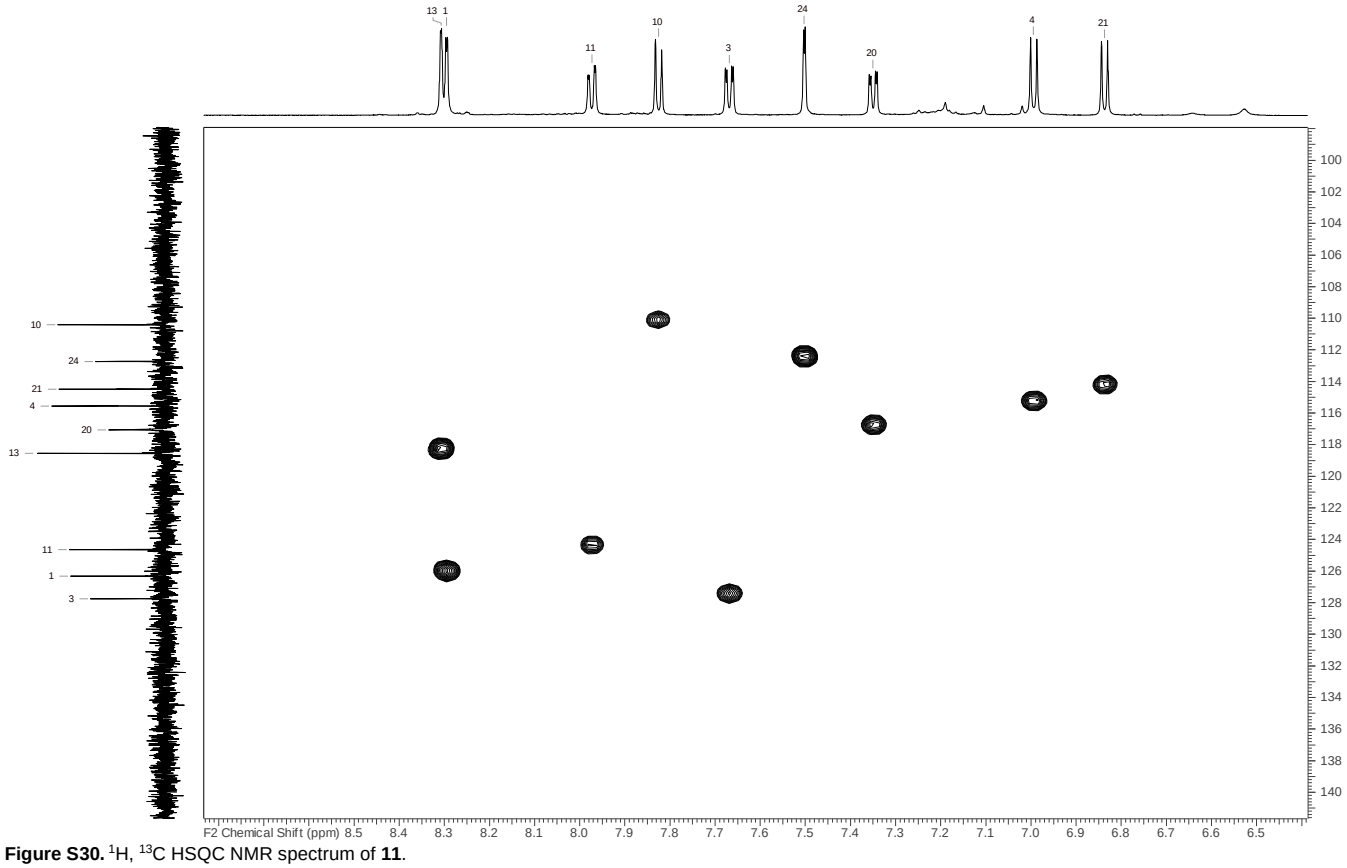


Figure S30. ^{1}H, ^{13}C HSQC NMR spectrum of **11**.

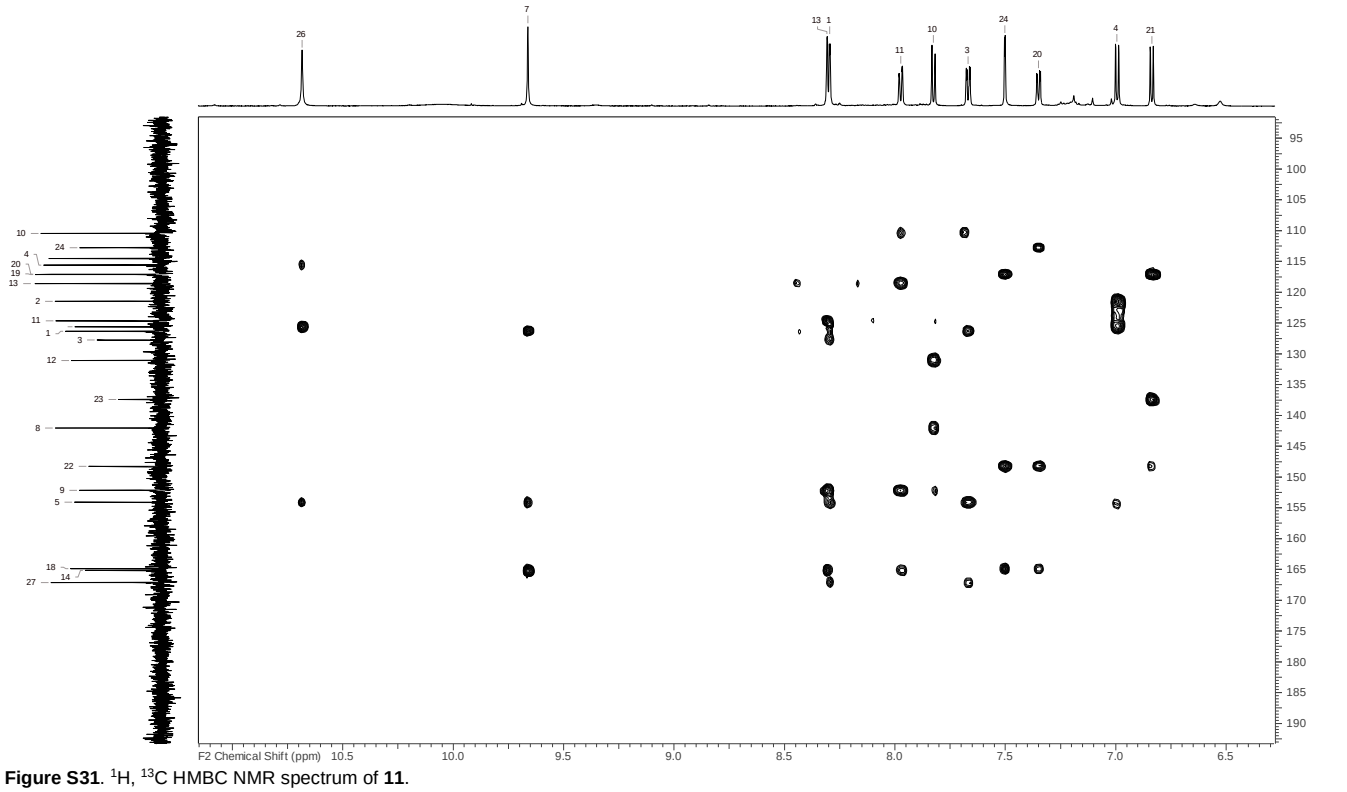


Figure S31. ^{1}H, ^{13}C HMBC NMR spectrum of **11**.

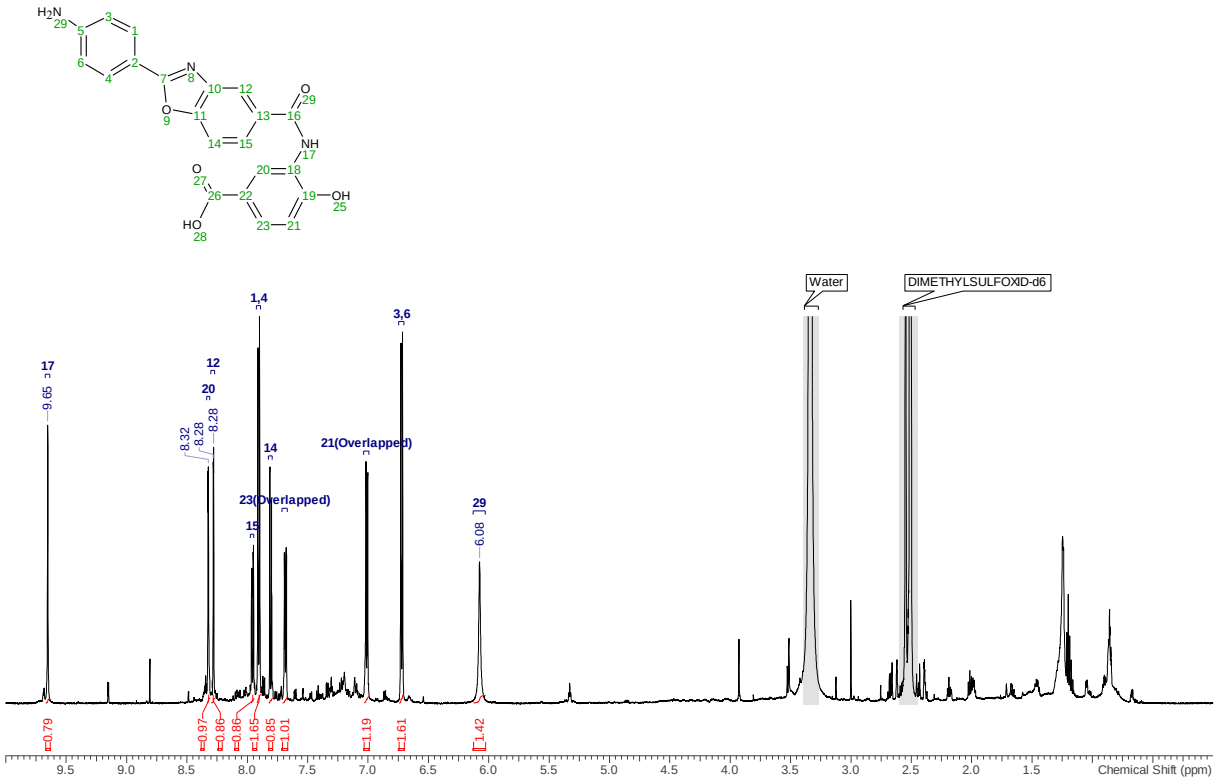


Figure S32. ^{1}H spectrum of **12** isolated from *C. cavendishii* Rif1. Itsmall impurities which lead to signals in the range of δ = 0.8–4.0 ppm.

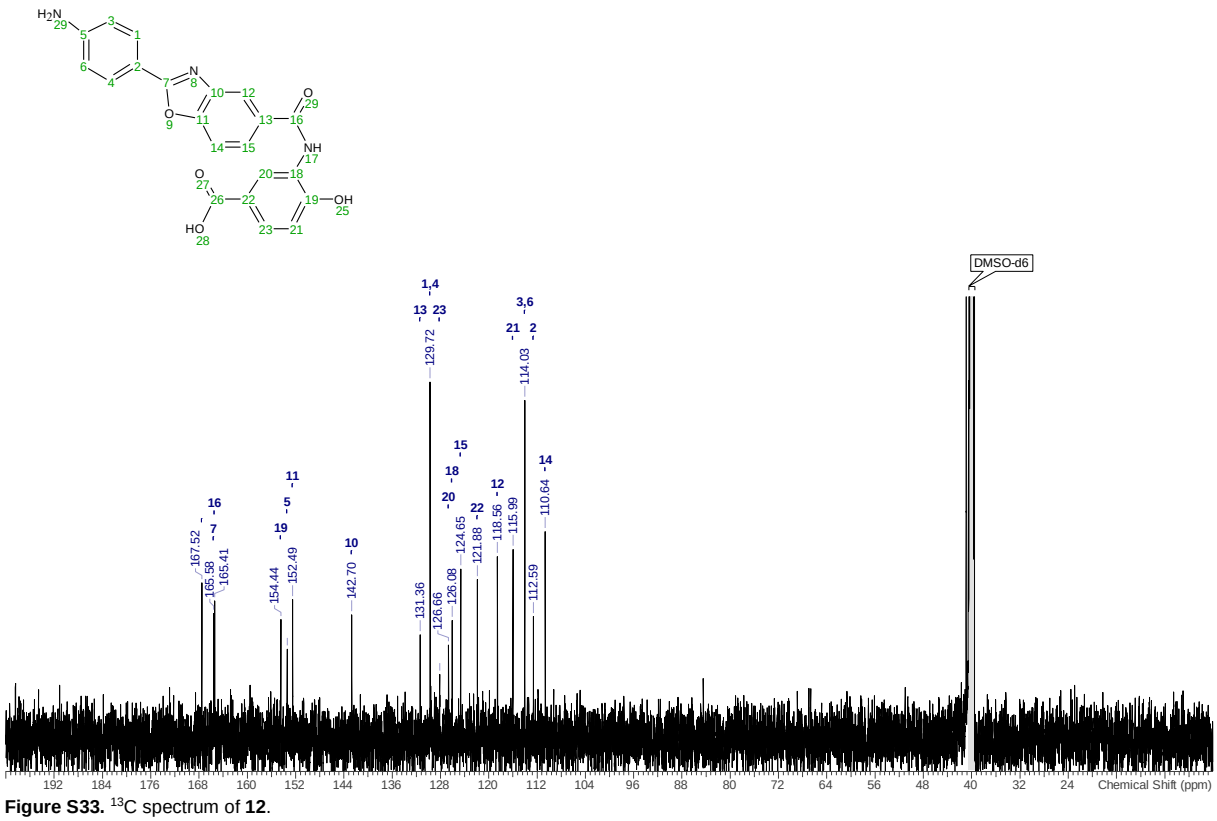


Figure S33. ^{13}C spectrum of **12**.

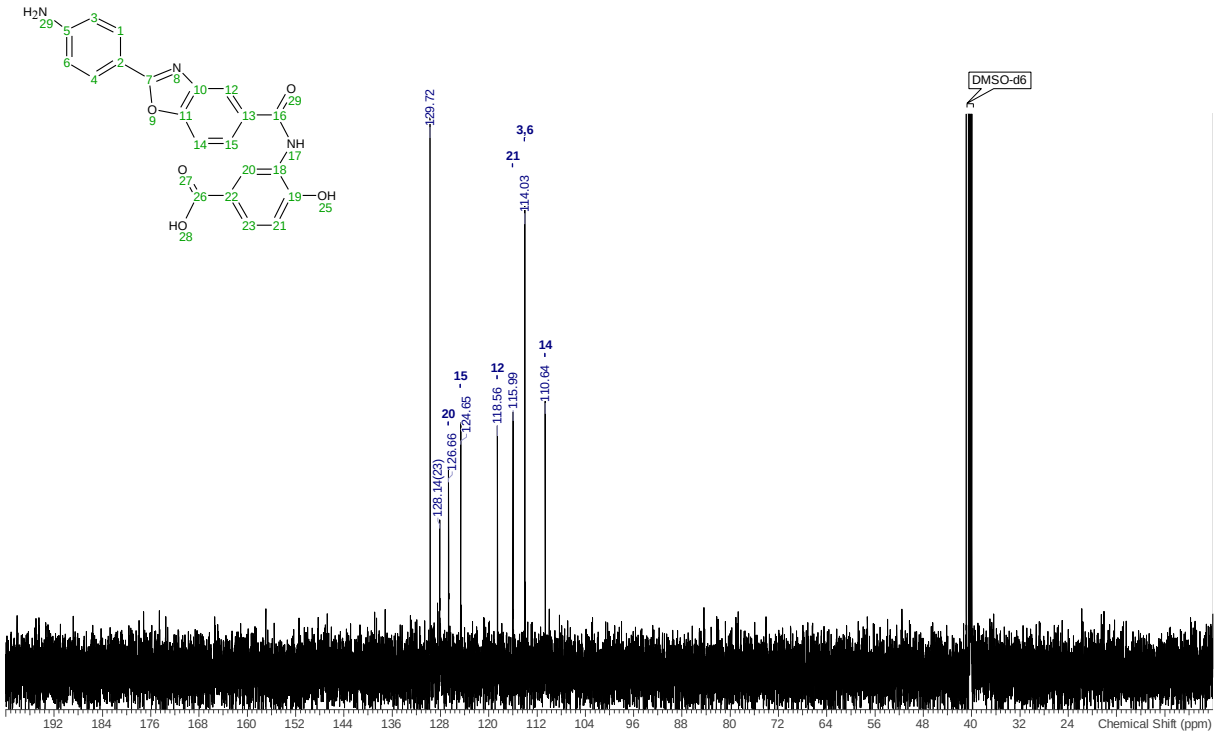


Figure S34. DEPT-135 NMR spectrum of **12**.

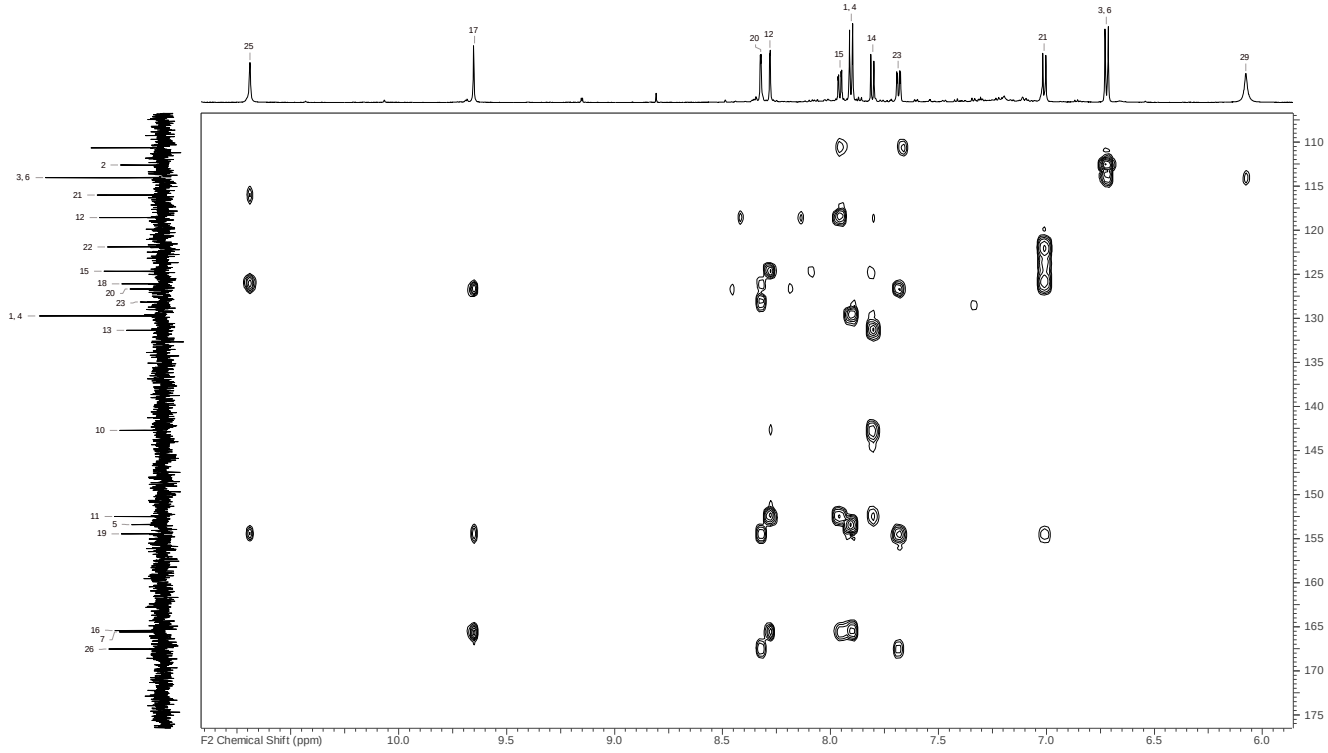


Figure S35. ^{1}H-^{13}C HMBC NMR spectrum of **12**.

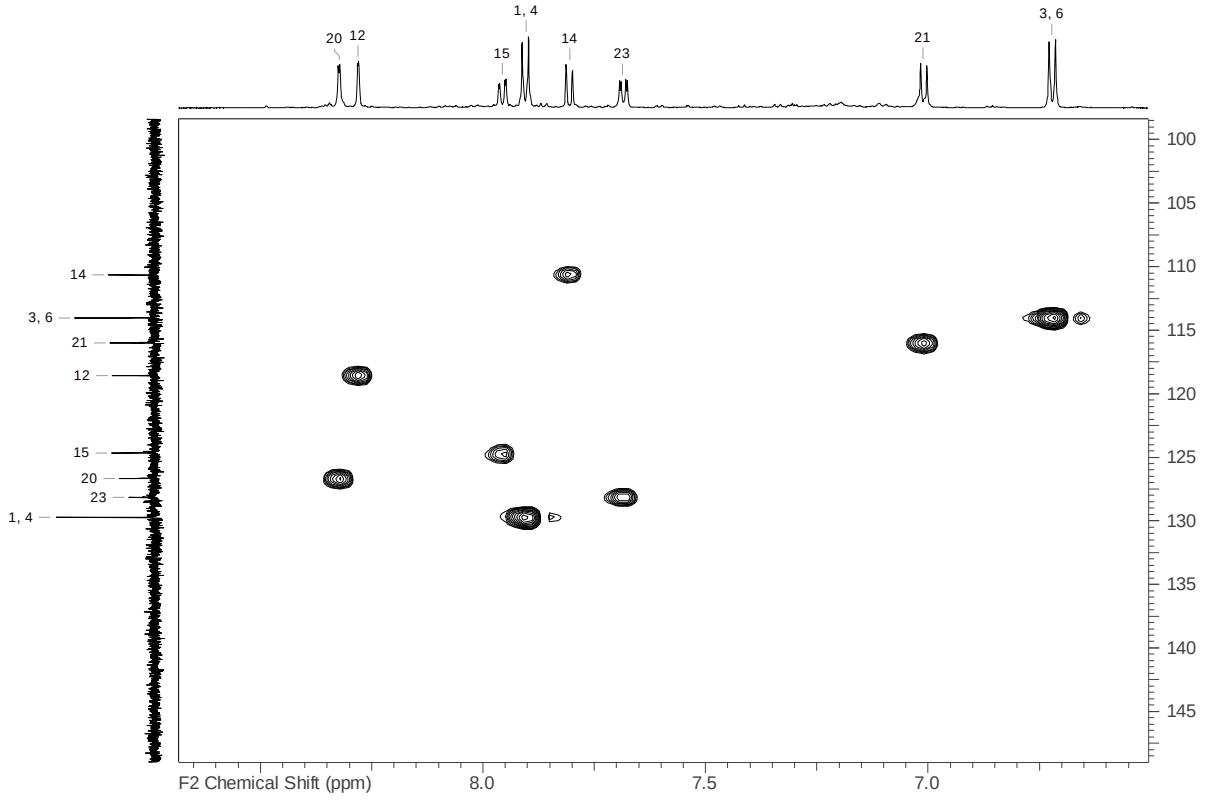


Figure S36. ^{1}H-^{13}C HSQC NMR spectrum of **12**.

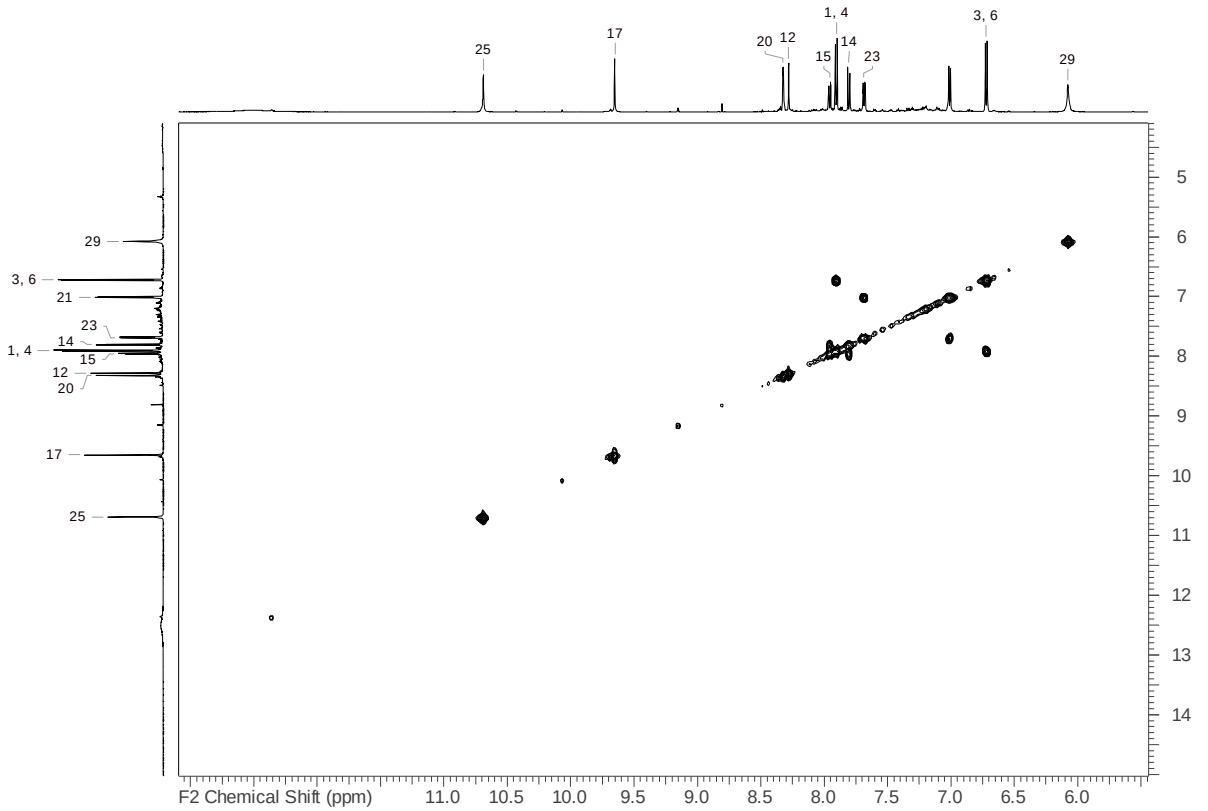


Figure S37. ^{1}H, ^{1}H COSY NMR spectrum of **12**.

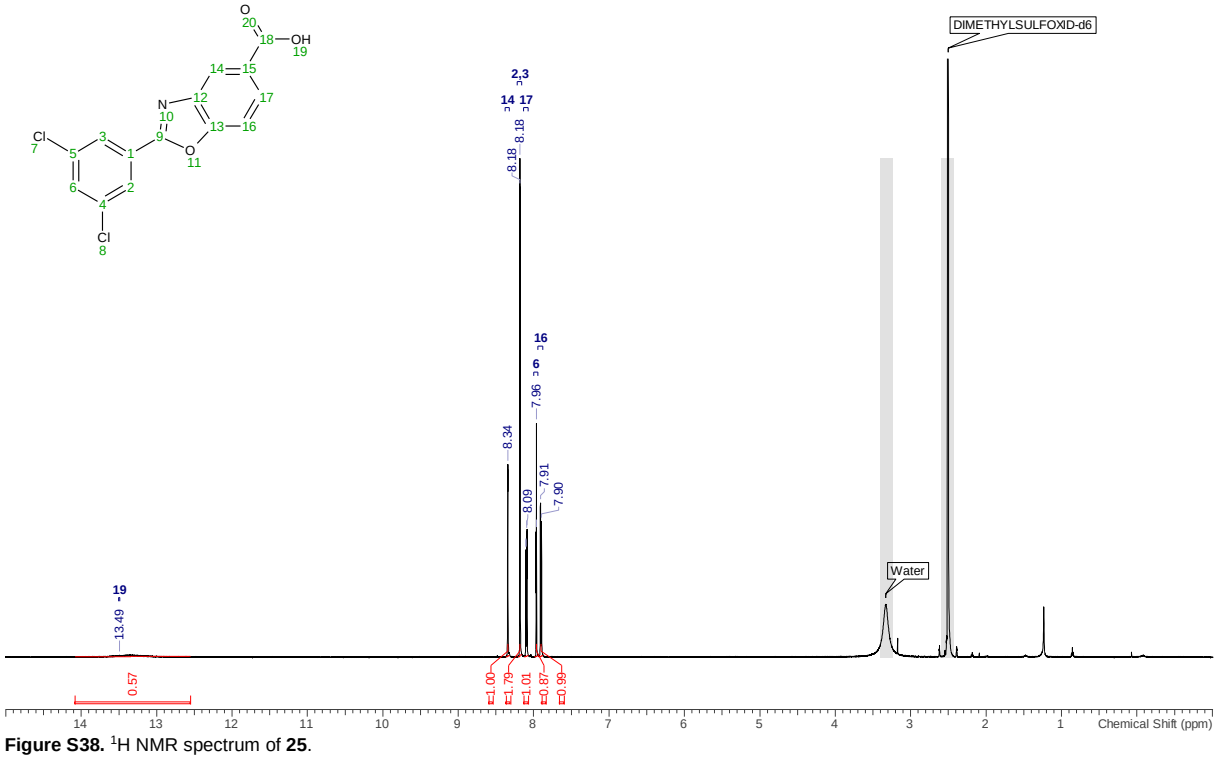


Figure S38. ^{1}H NMR spectrum of **25**.

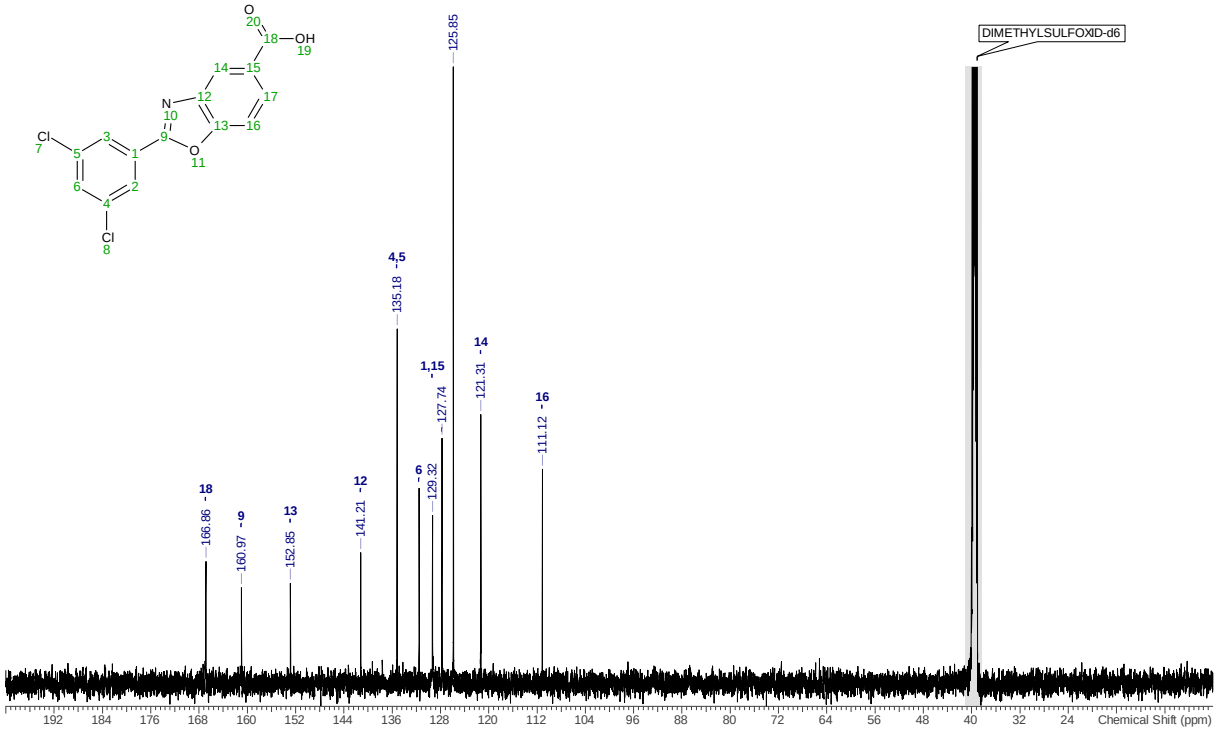


Figure S39. ^{13}C NMR spectrum of **25**.

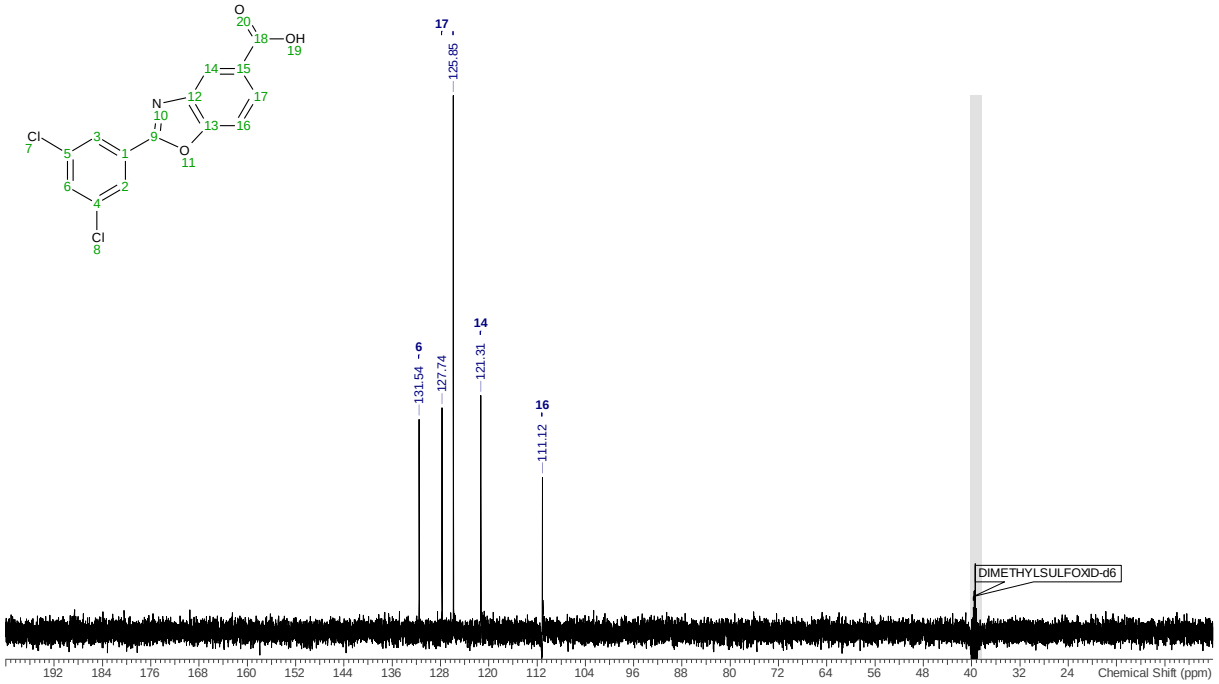


Figure S40. DEPT-135 NMR spectrum of **25**.

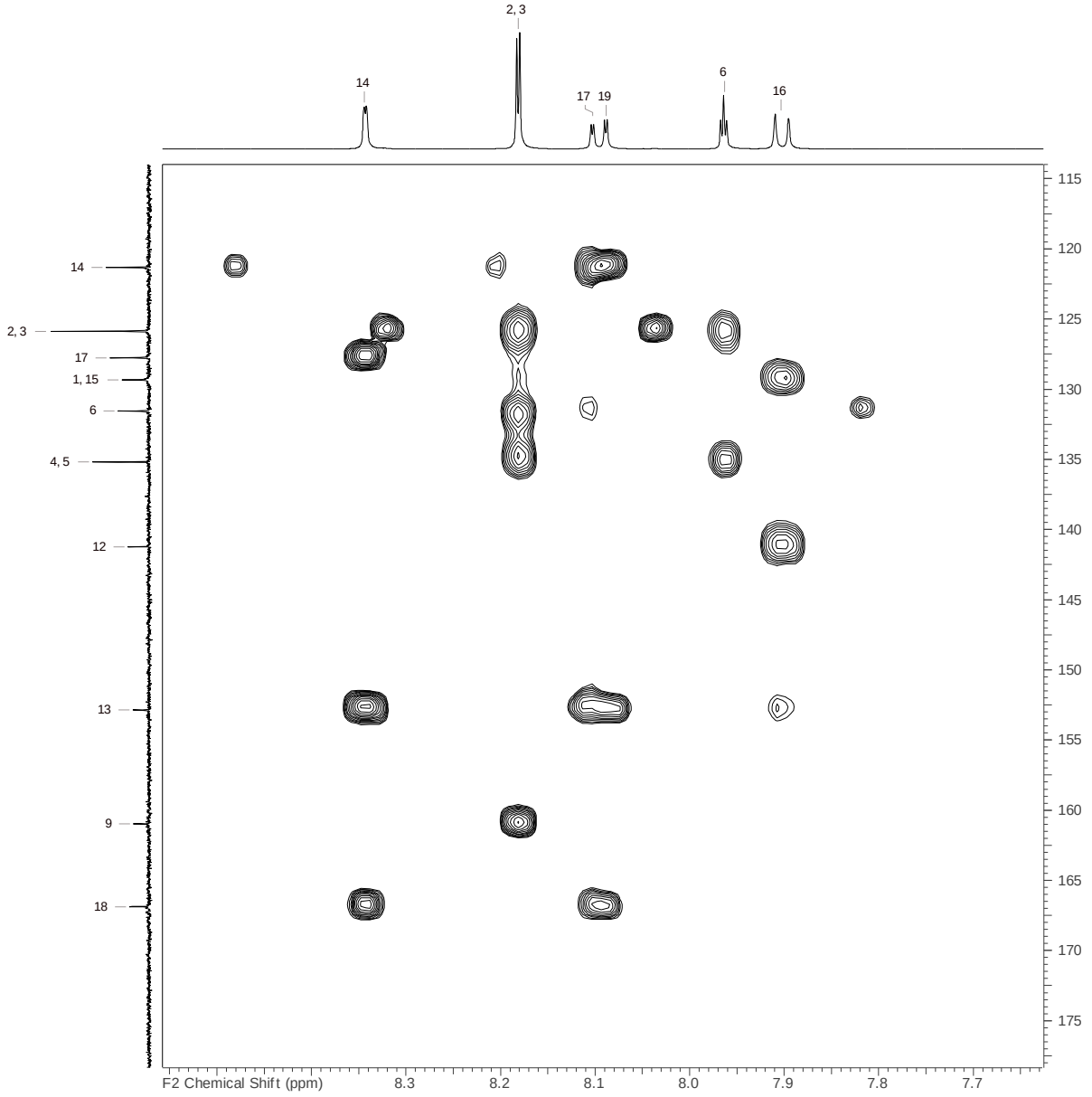


Figure S41. ^{1}H-^{13}C HMBC NMR spectrum of **25**.

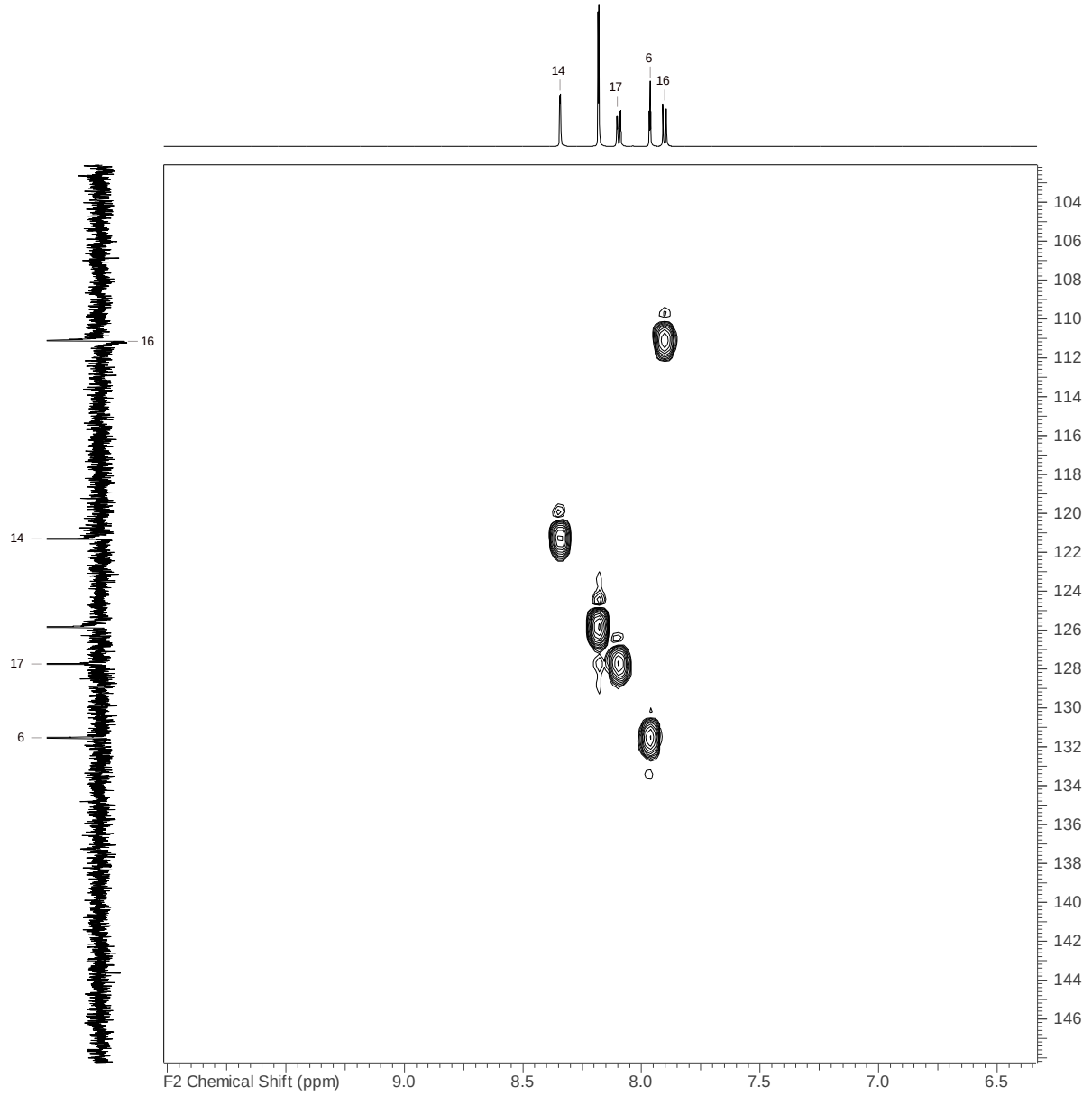


Figure S42. ^{1}H-^{13}C HSQC NMR spectrum of **25**.

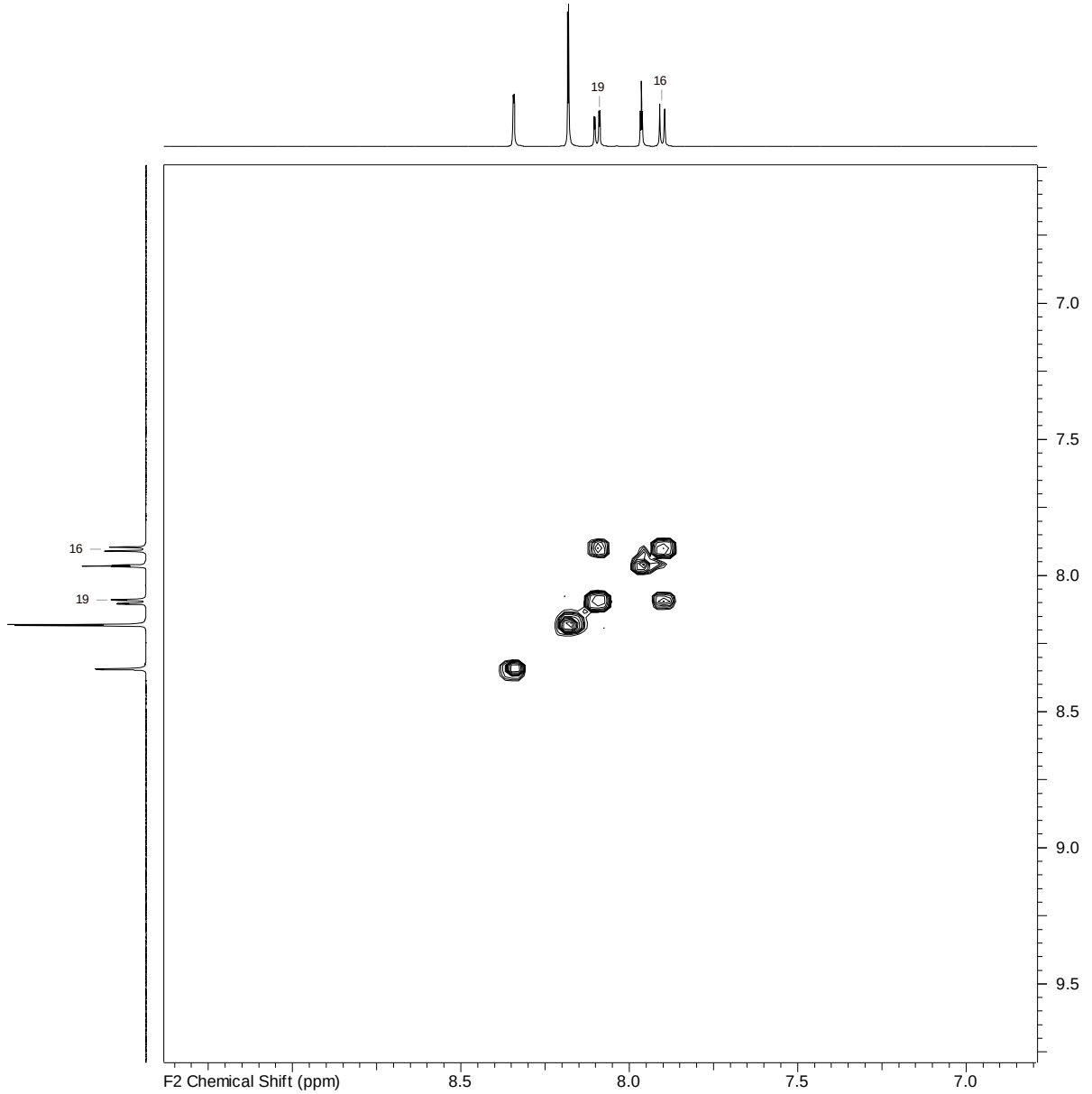


Figure S43. ^{1}H, ^{1}H COSY NMR spectrum of **25**.

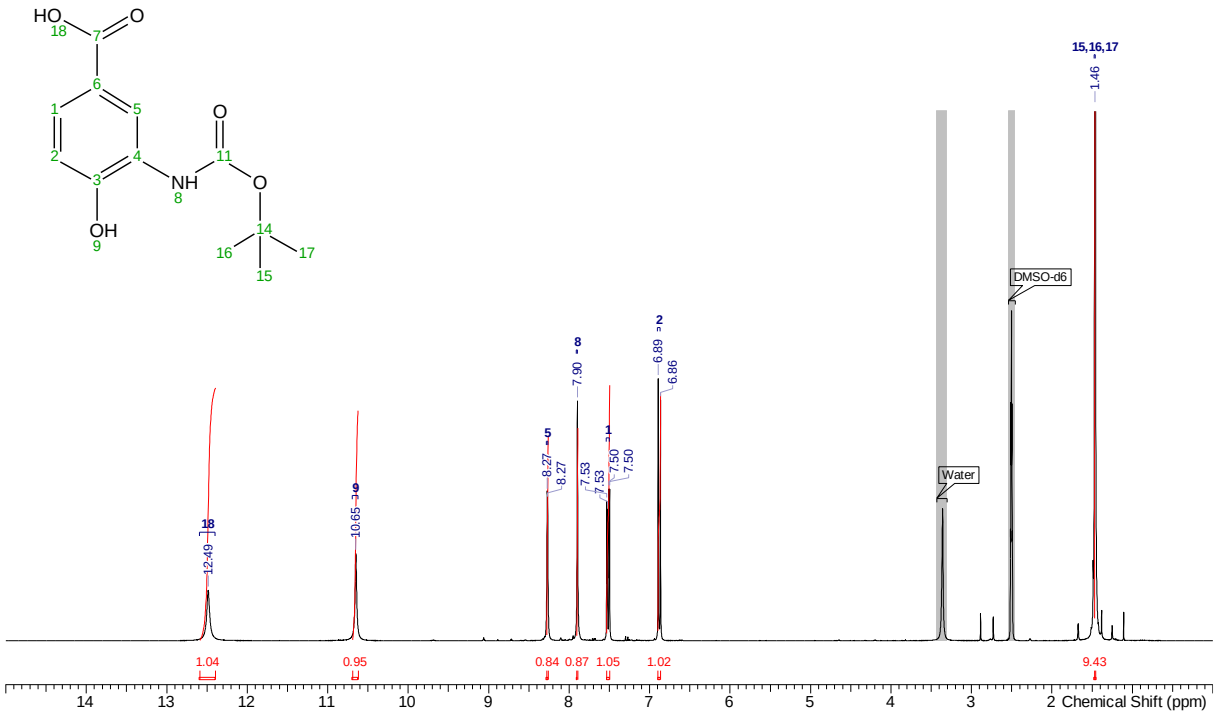


Figure S44. ^{1}H NMR spectrum of **15**.

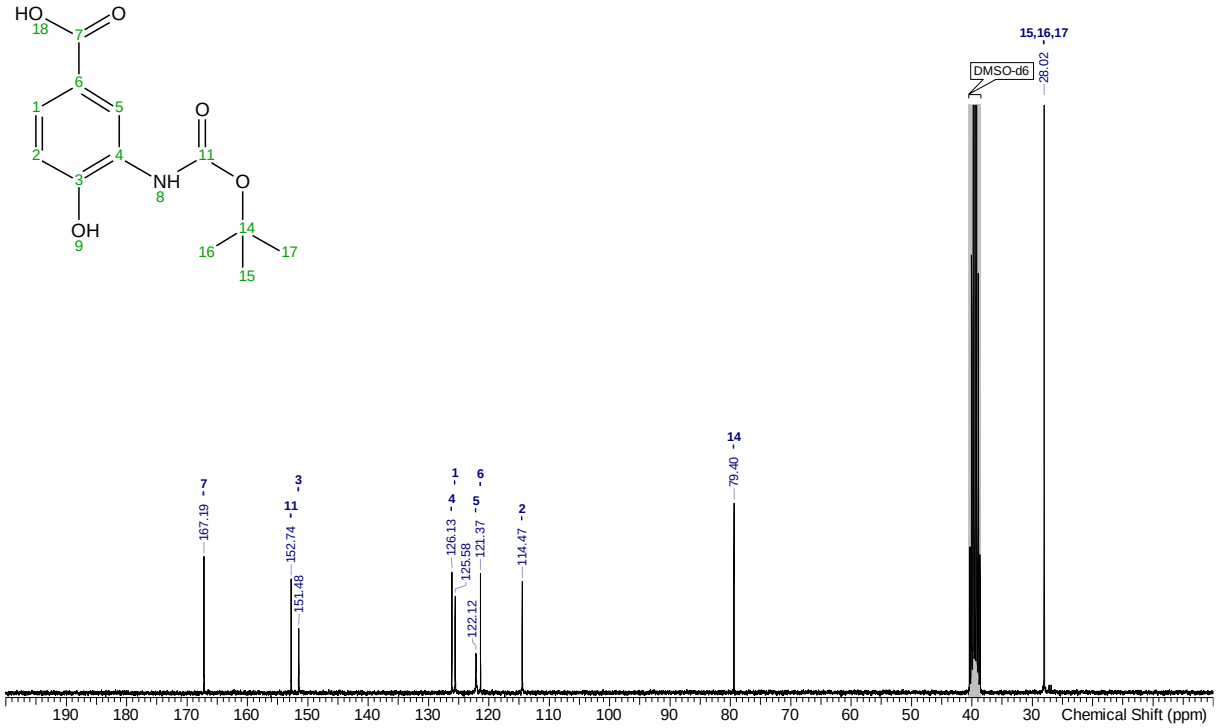


Figure S 45. ^{13}C NMR spectrum of **15.**

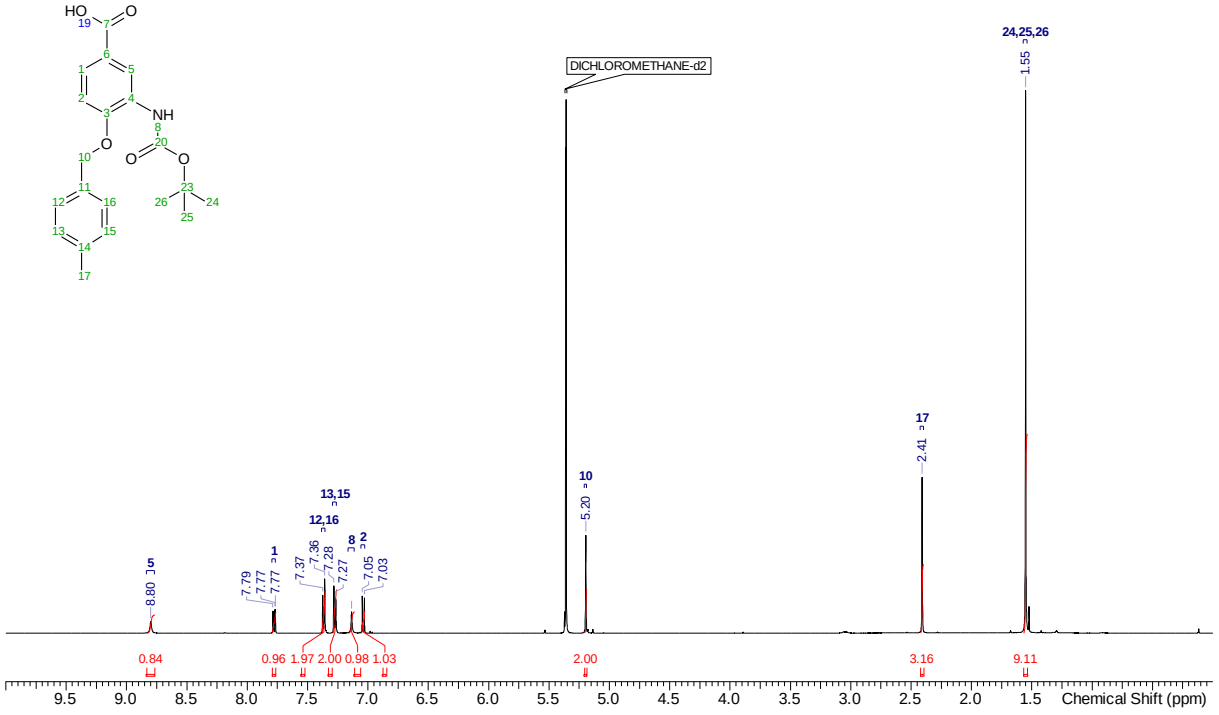


Figure S46. ^{1}H NMR spectrum of **16.**

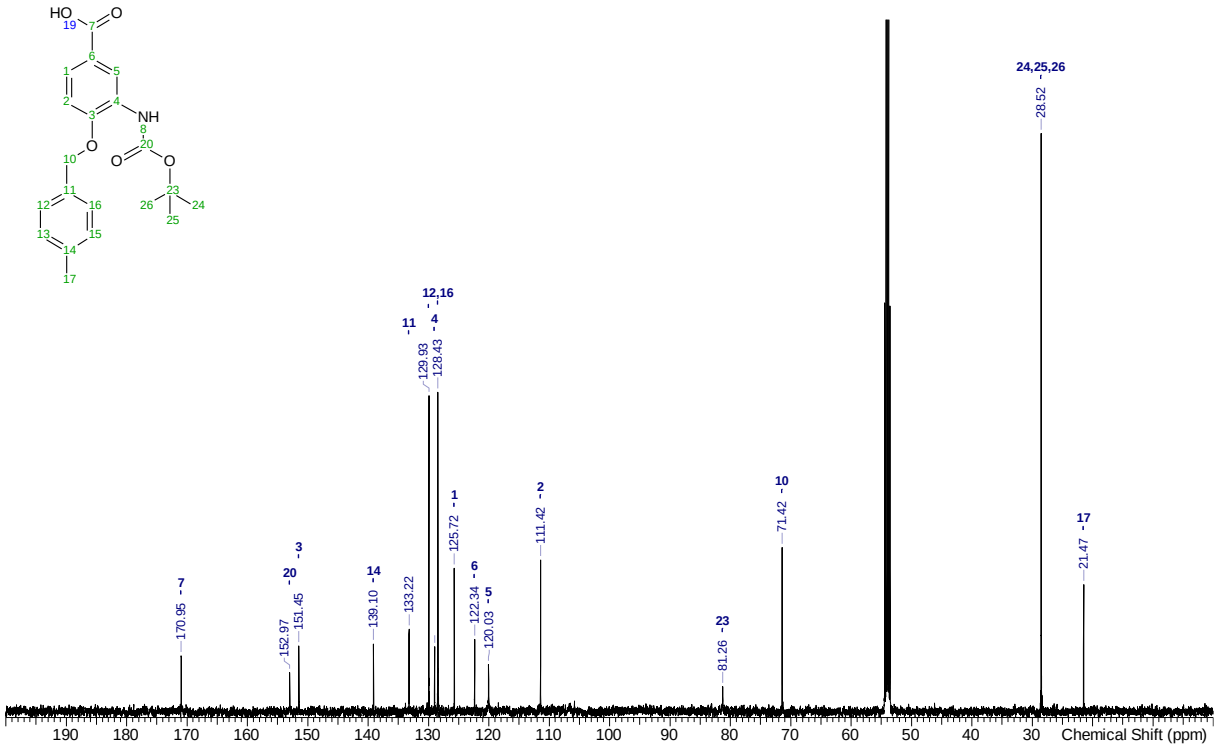


Figure S 47. ^{13}C NMR spectrum of **16.**

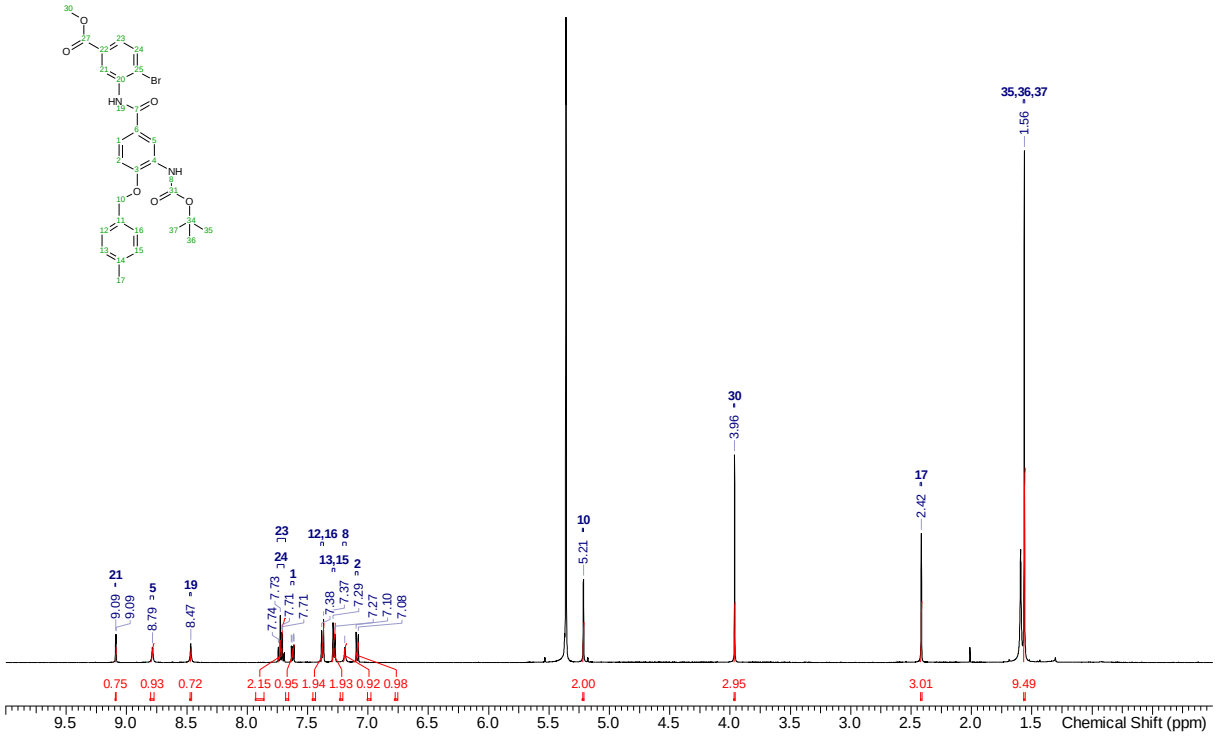


Figure S 48. ^{1}H NMR spectrum of **19.**

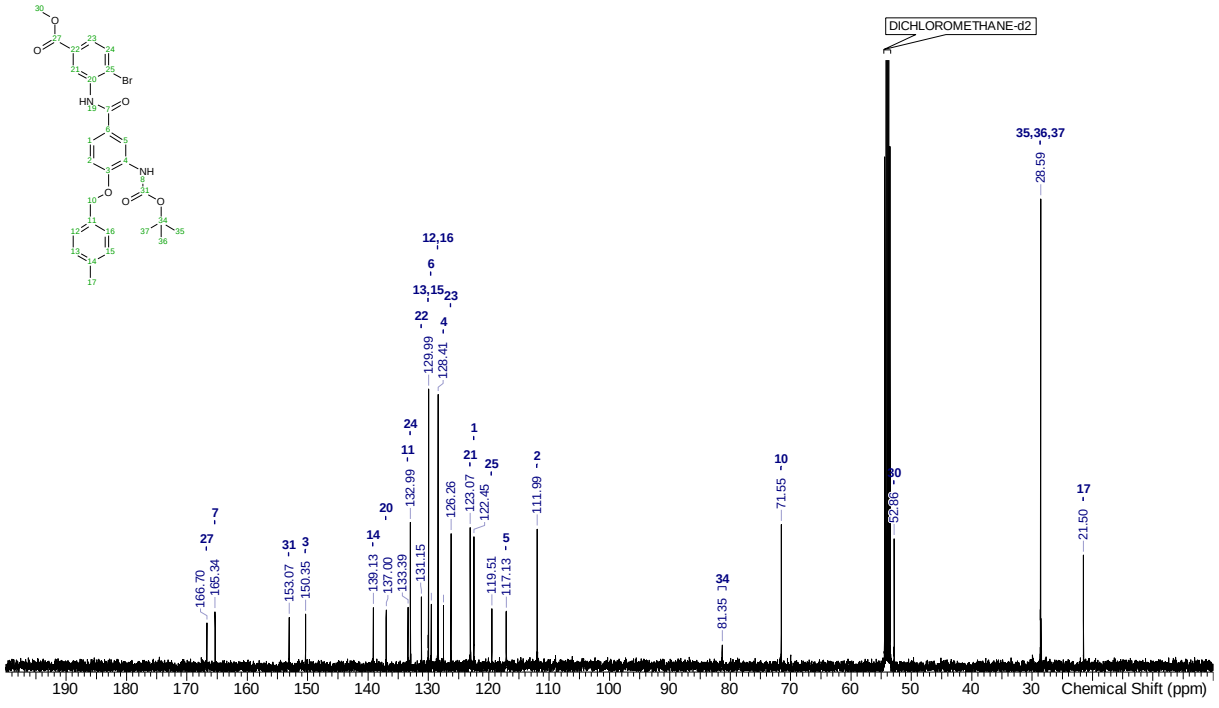


Figure S 49. ^{13}C NMR spectrum of **19.**

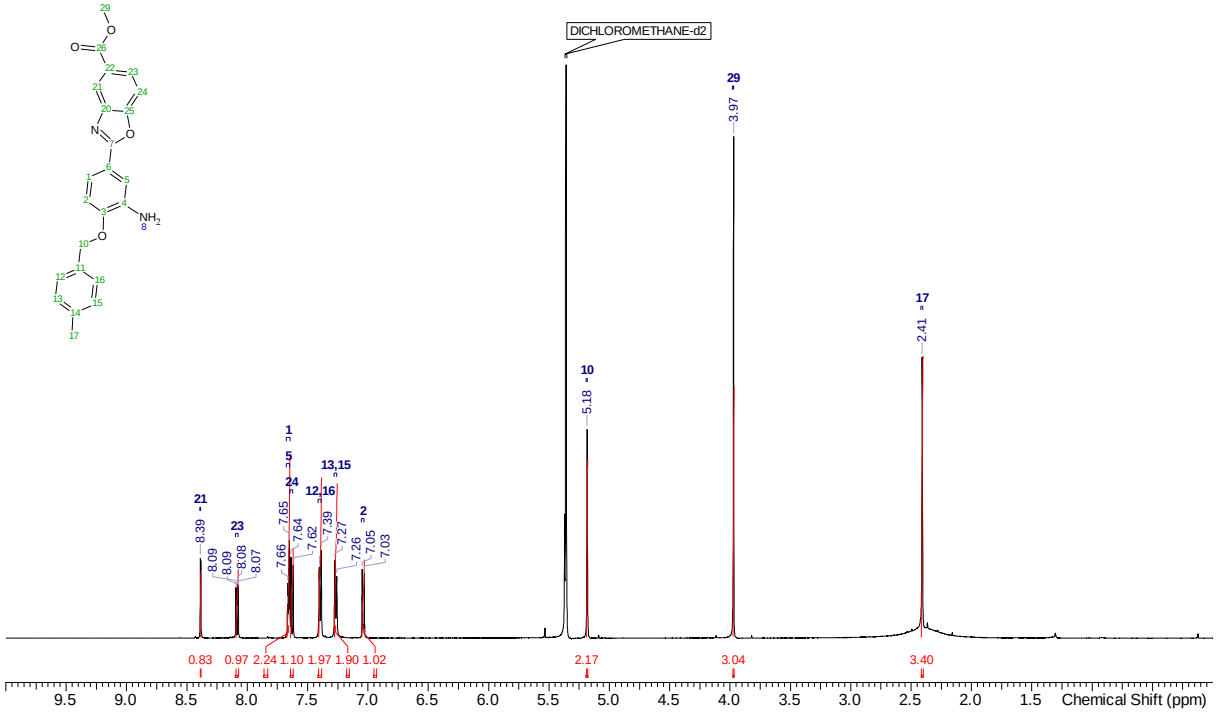


Figure S 50. ^{1}H NMR spectrum of **20.**

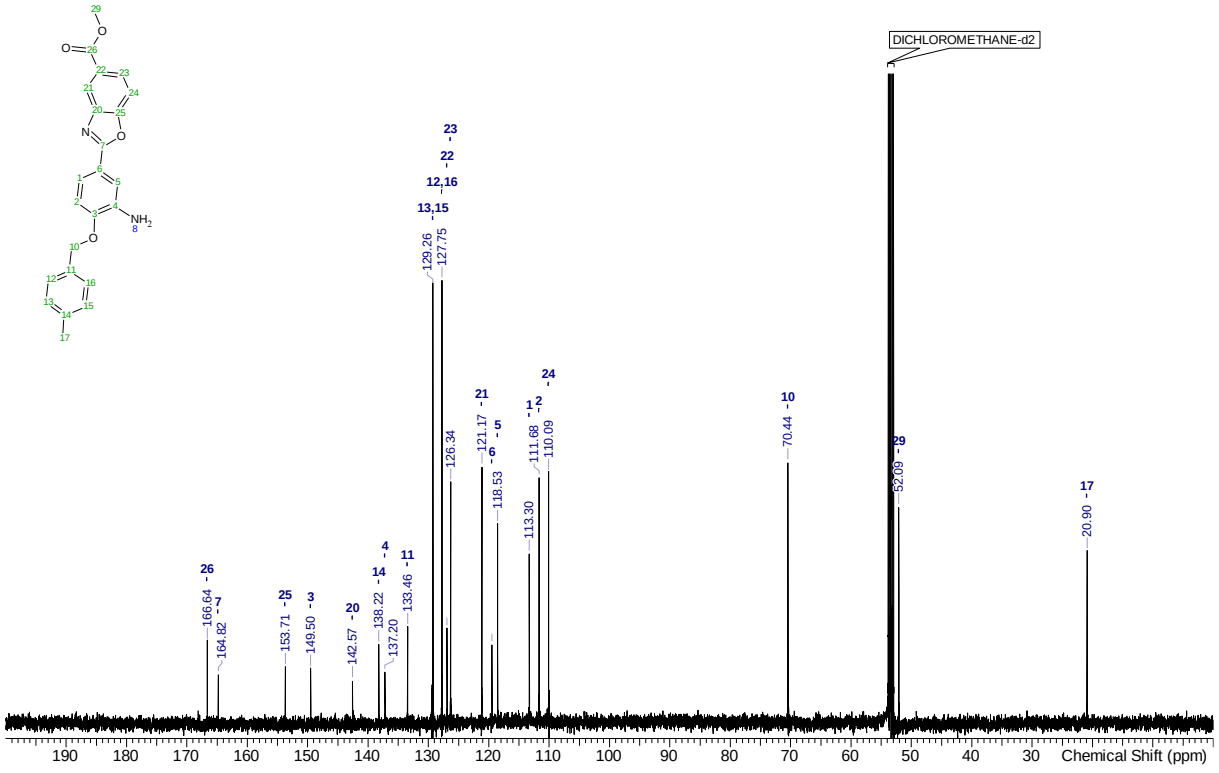


Figure S 51. ^{13}C NMR spectrum of **20.**

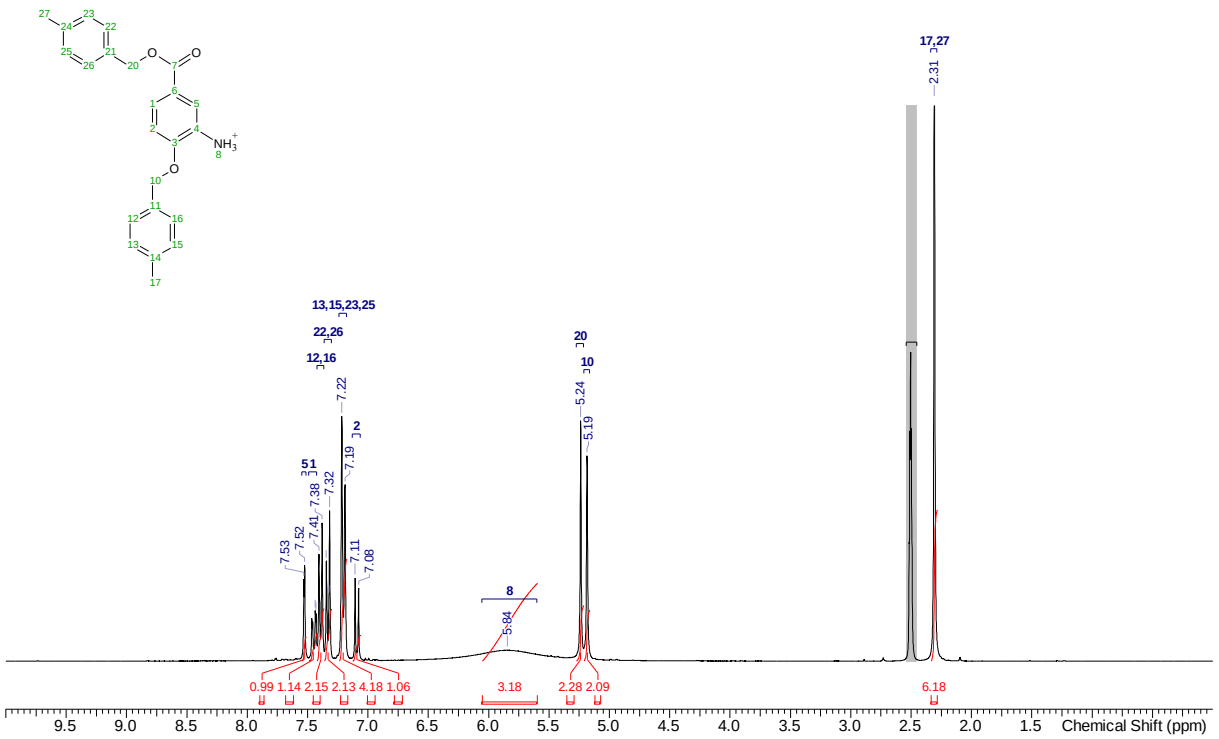


Figure S52. ^{1}H NMR spectrum of **17**.

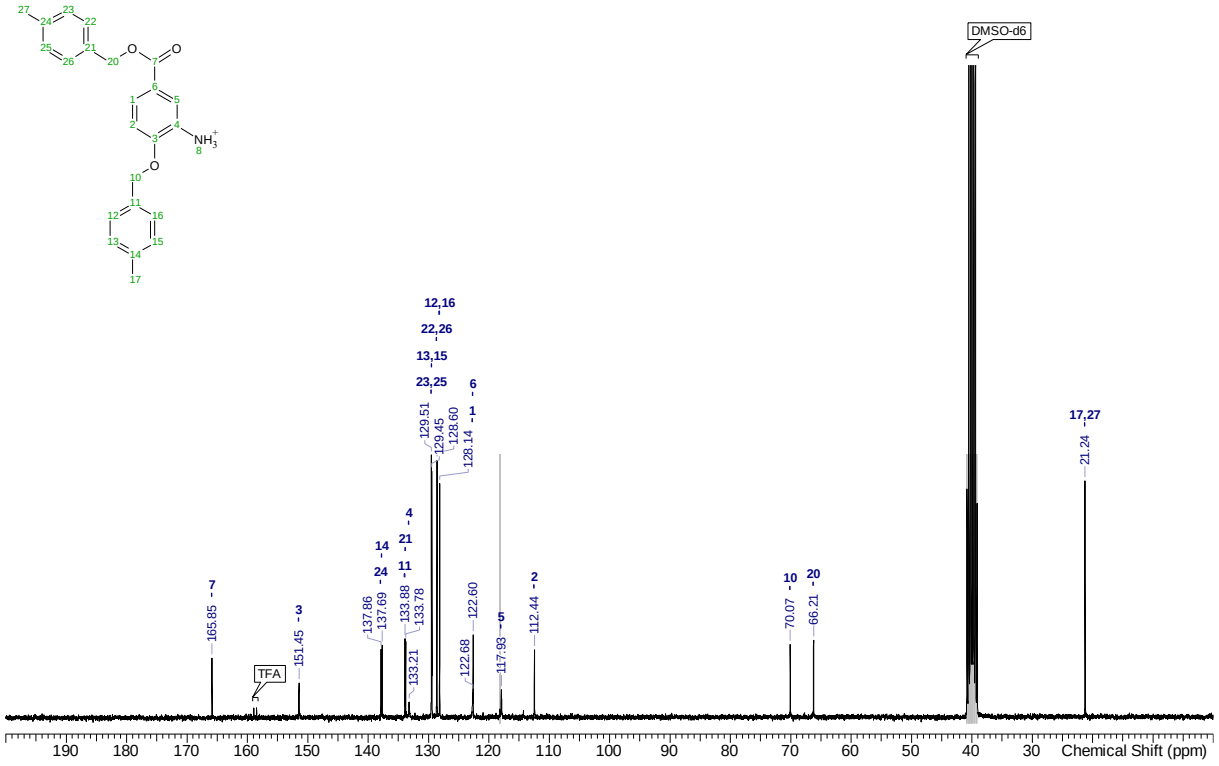


Figure S53. ^{13}C NMR spectrum of **17**.

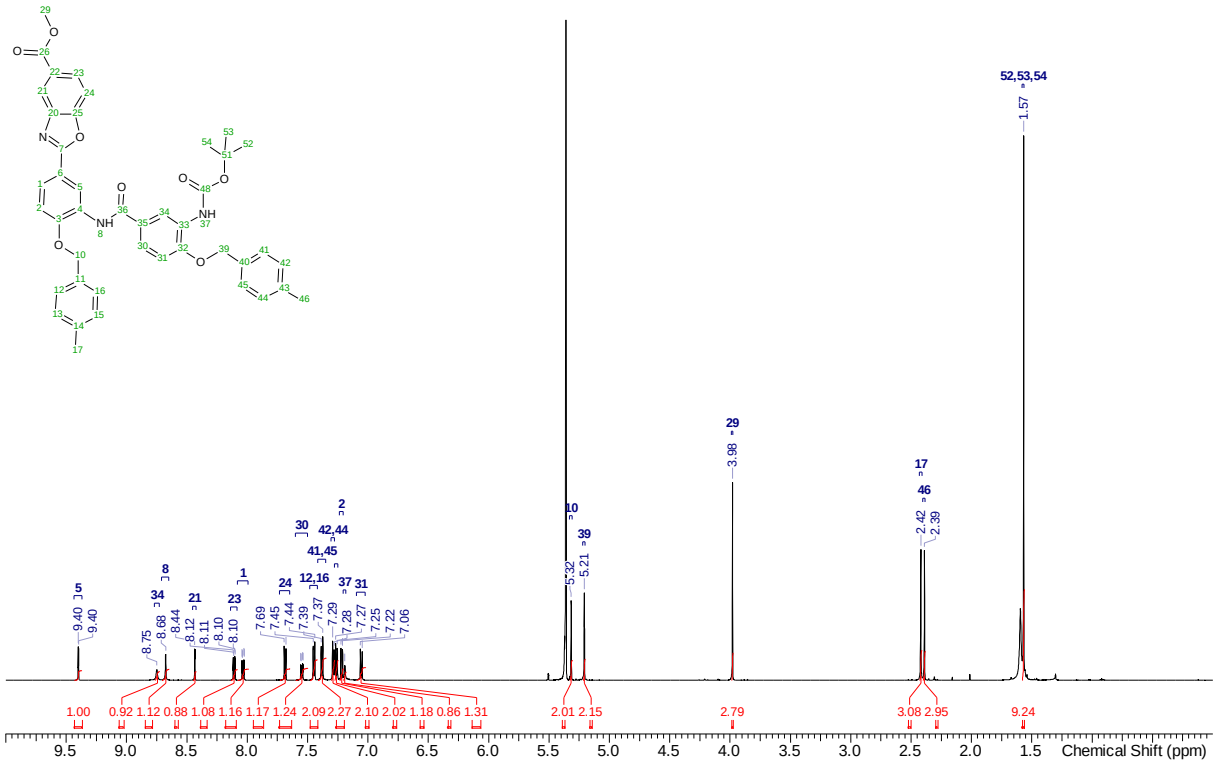


Figure S54. ^{1}H NMR spectrum of **26.**

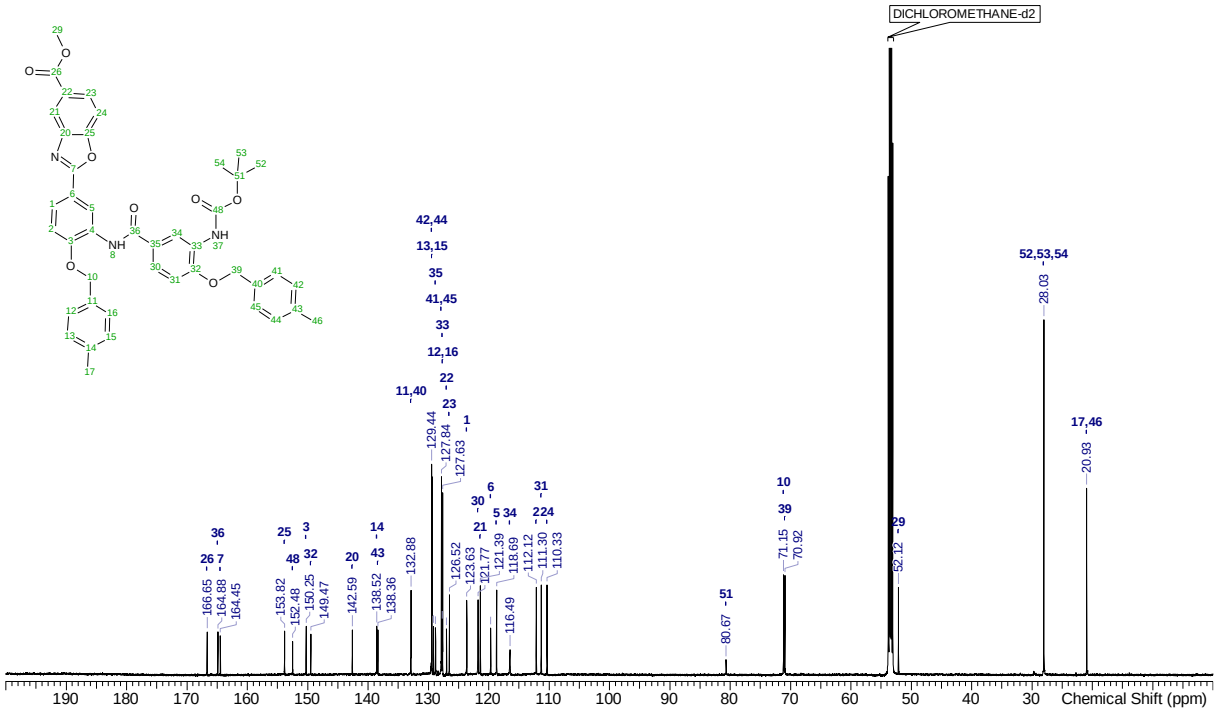


Figure S55. ^{13}C NMR spectrum of **26.**

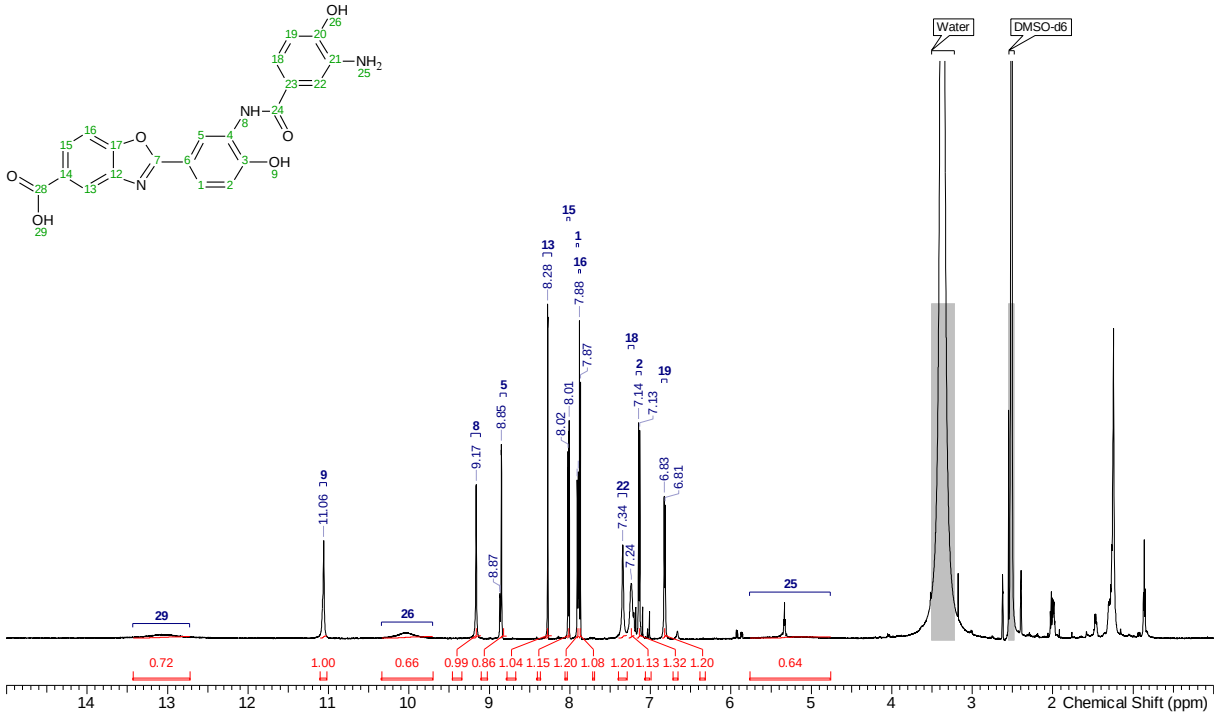


Figure S56. ^{1}H NMR spectrum of **14.**

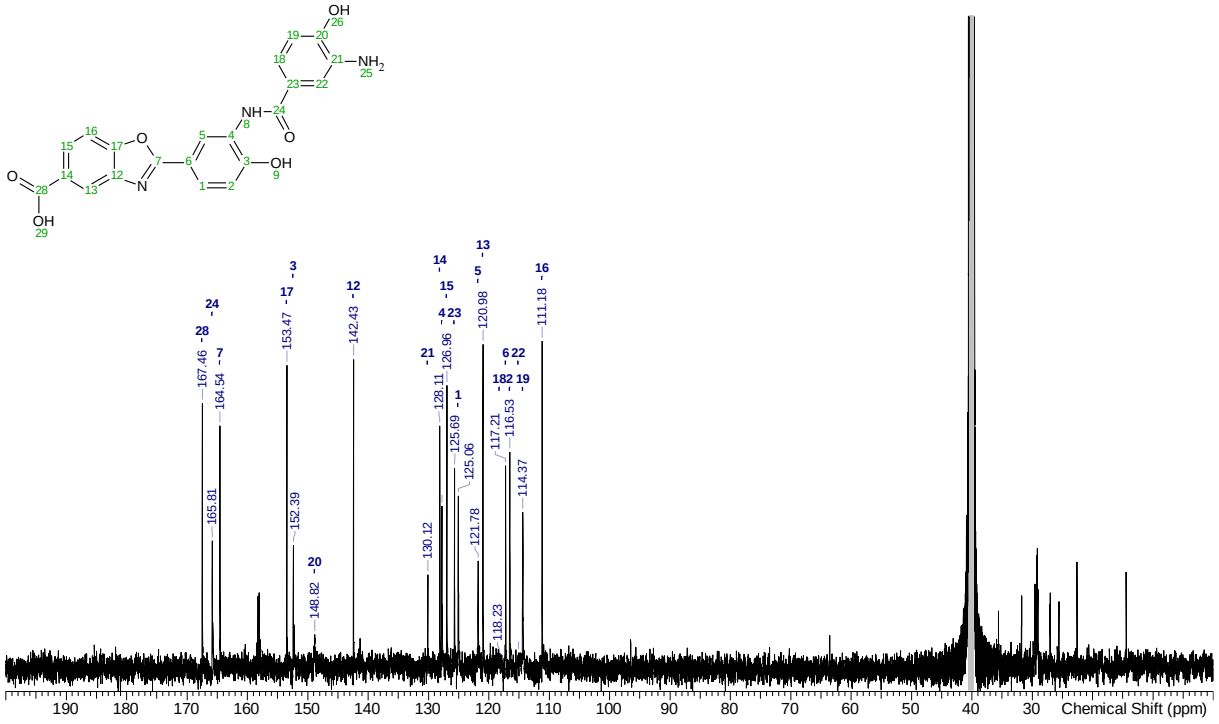


Figure S57. ^{13}C NMR spectrum of **14.**

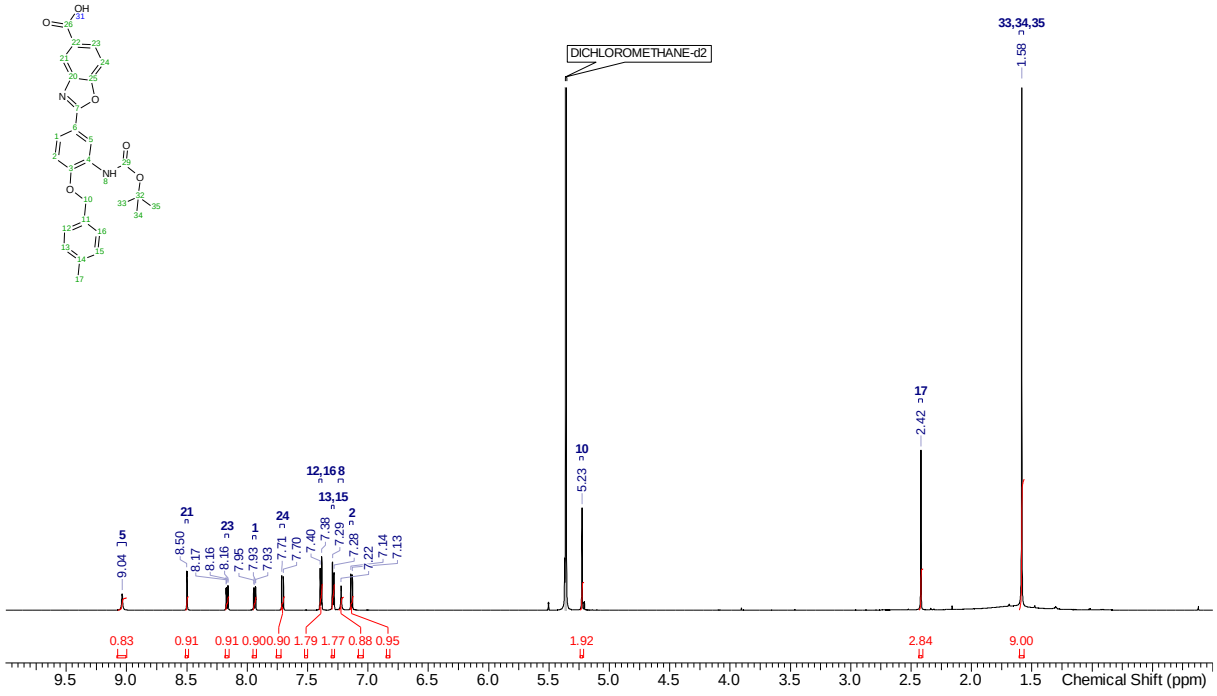


Figure S58. ^{1}H NMR spectrum of **27**.

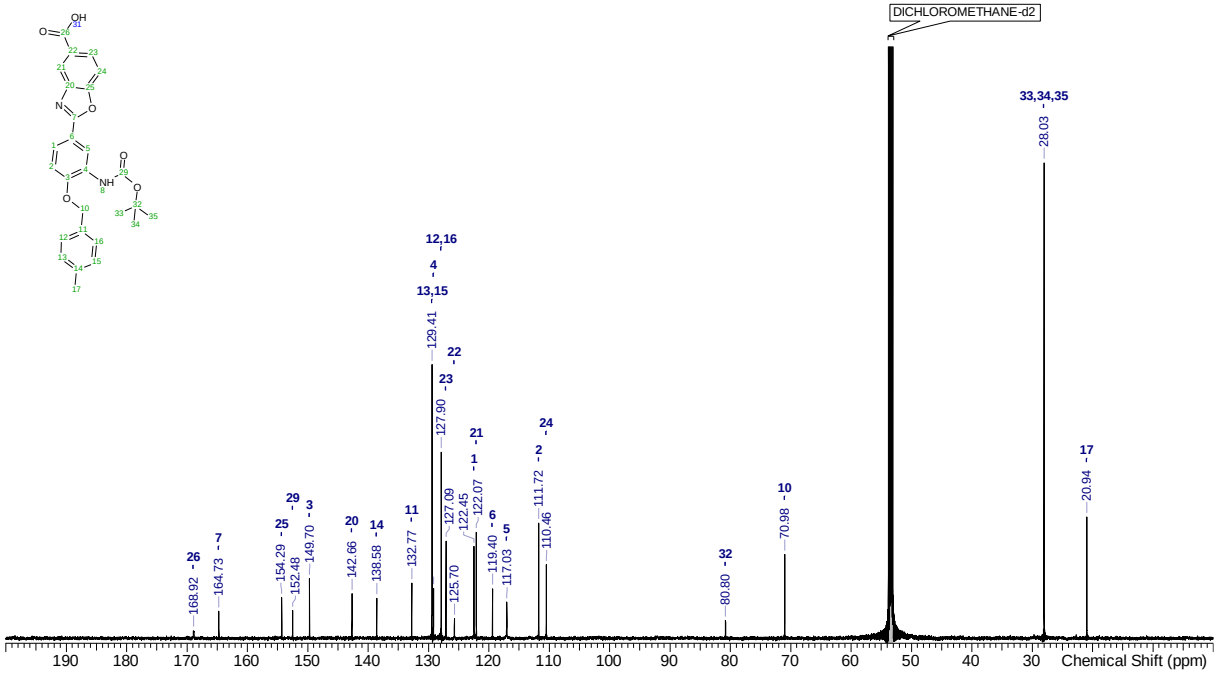


Figure S59. ^{13}C NMR spectrum of **27**.

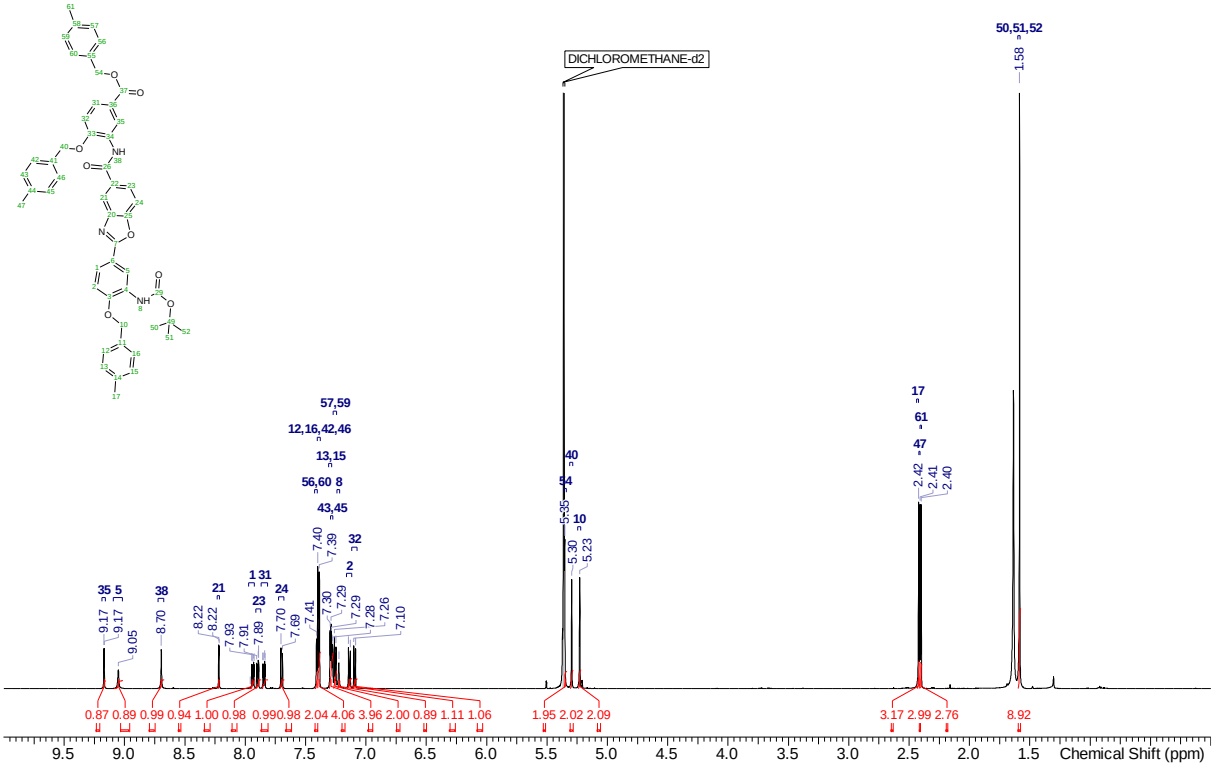


Figure S60. ^{1}H NMR spectrum of **21.**

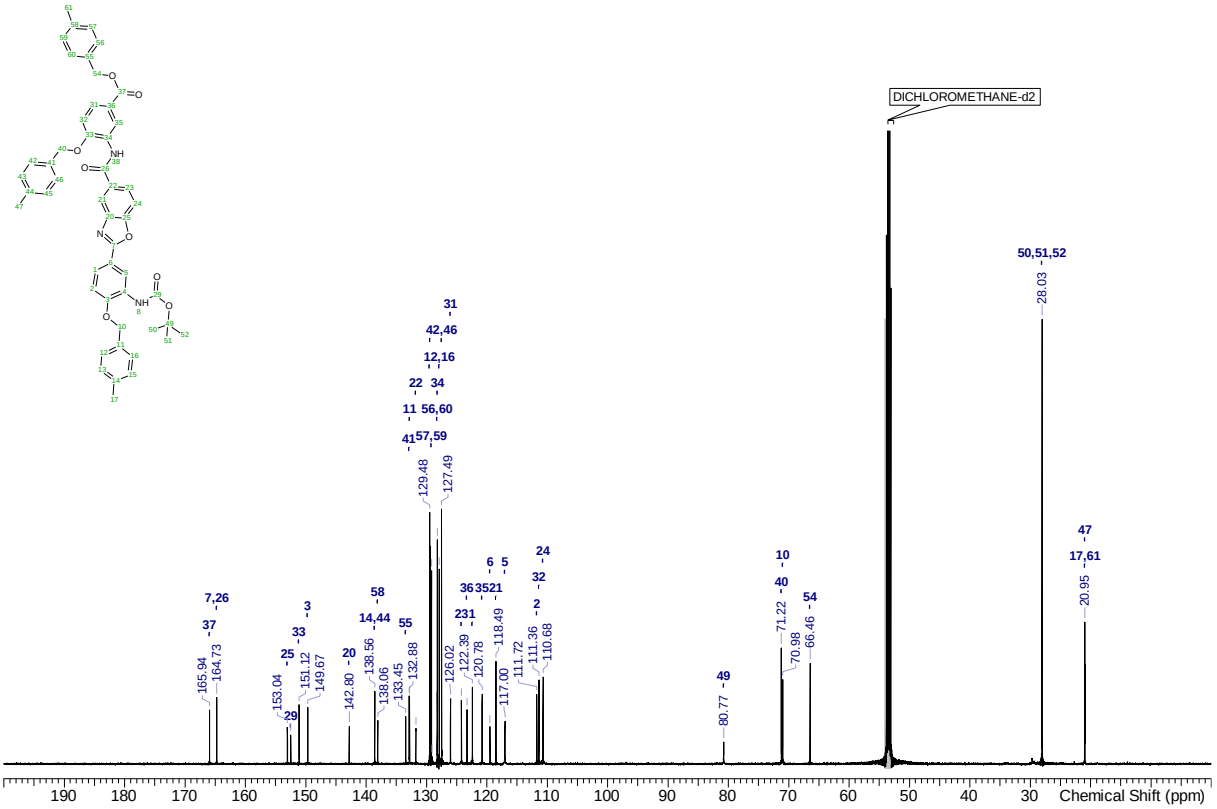


Figure S61. ^{13}C NMR spectrum of **21.**

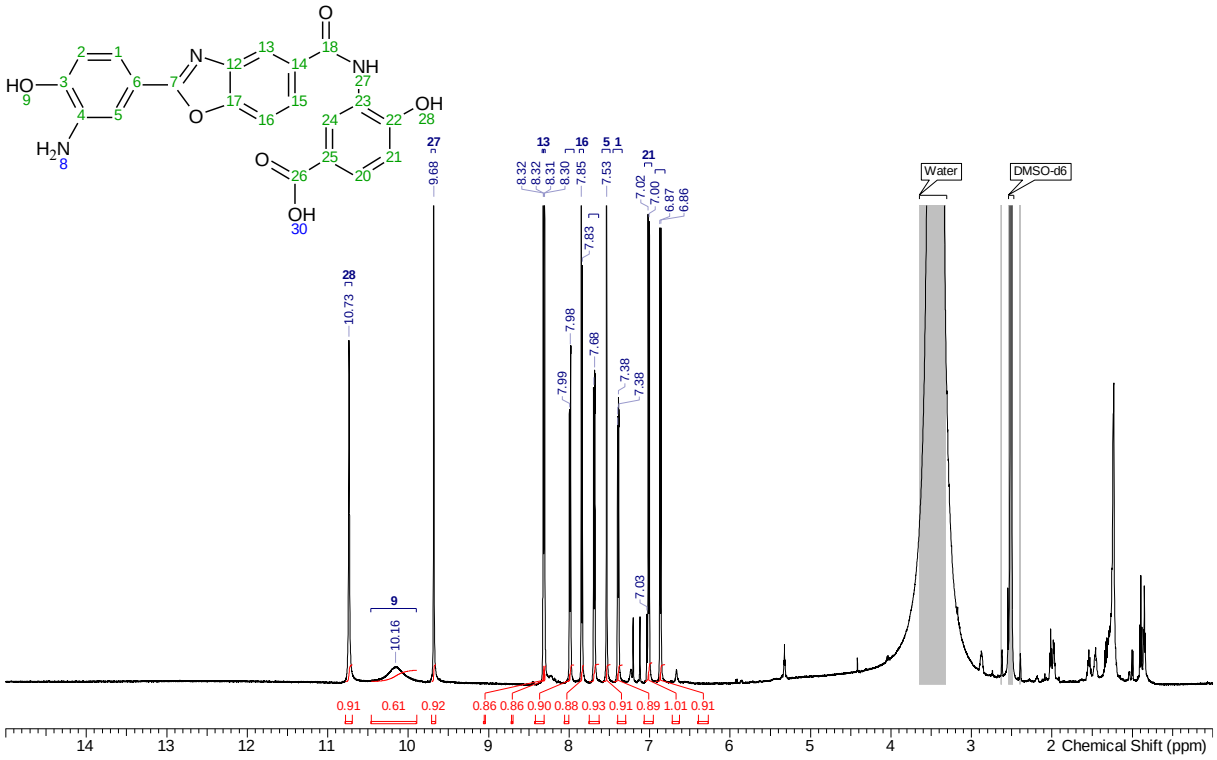


Figure S62. ^{1}H NMR spectrum of synthetic **11.**

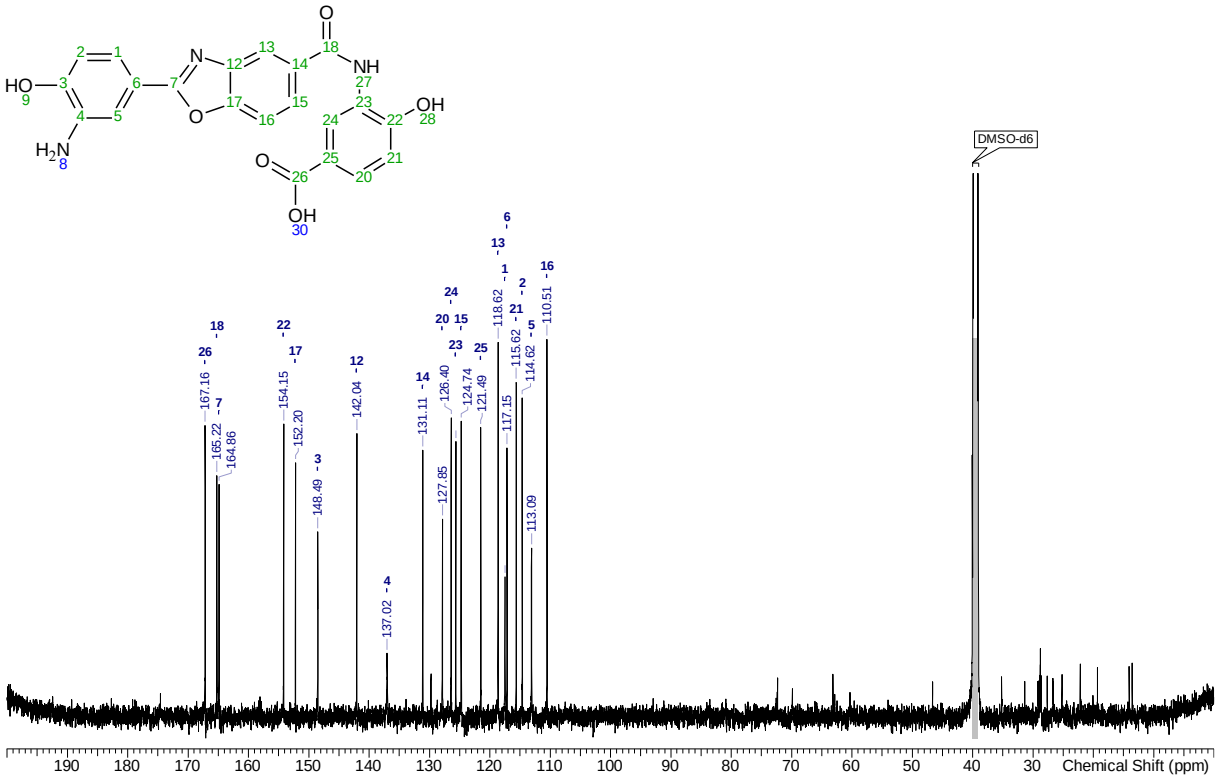


Figure S63. ^{13}C NMR spectrum of synthetic **11.**

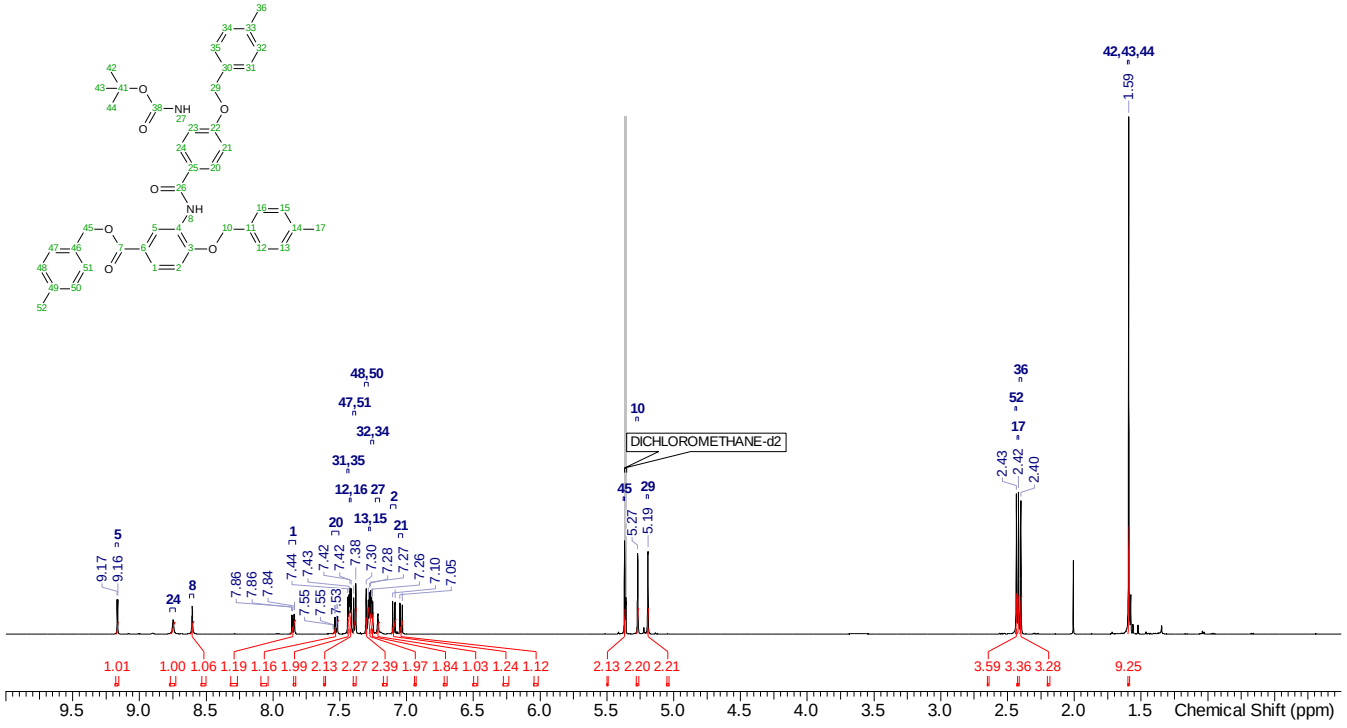


Figure S64. ^{1}H NMR spectrum of **28.**

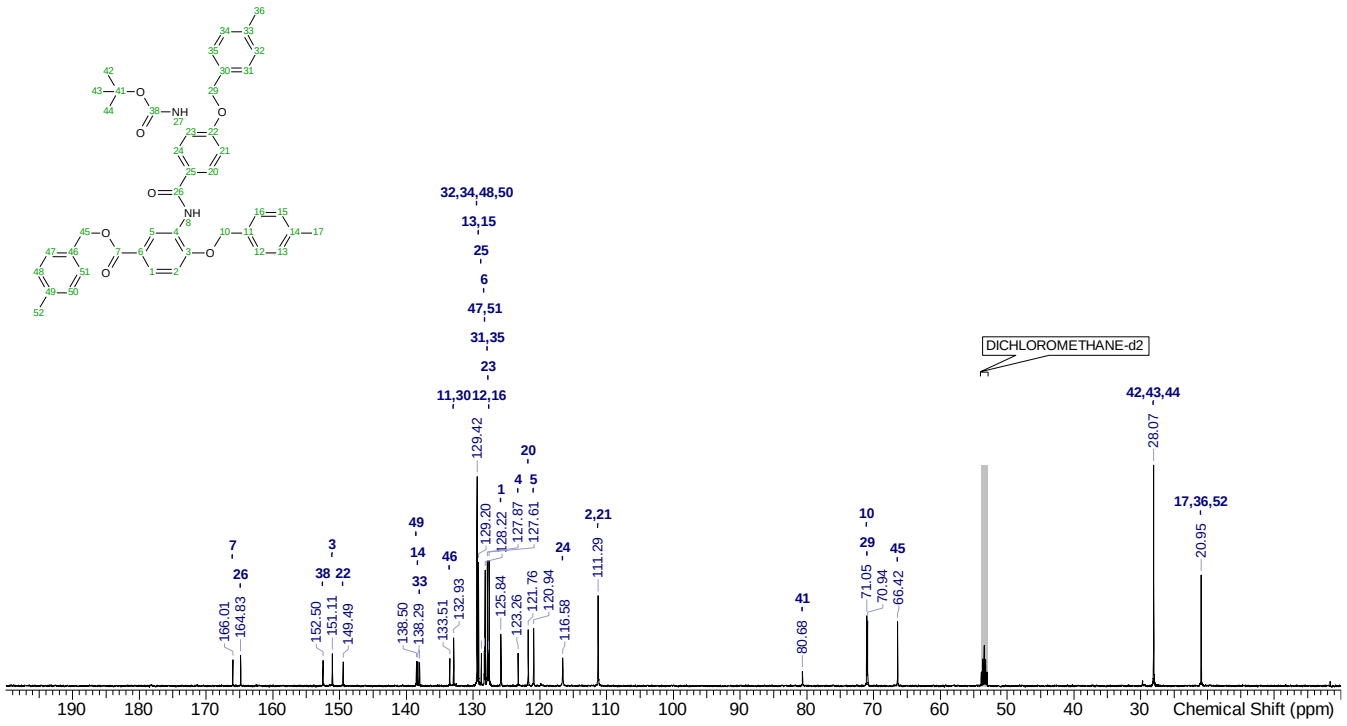


Figure S65. ^{13}C NMR spectrum of **28.**

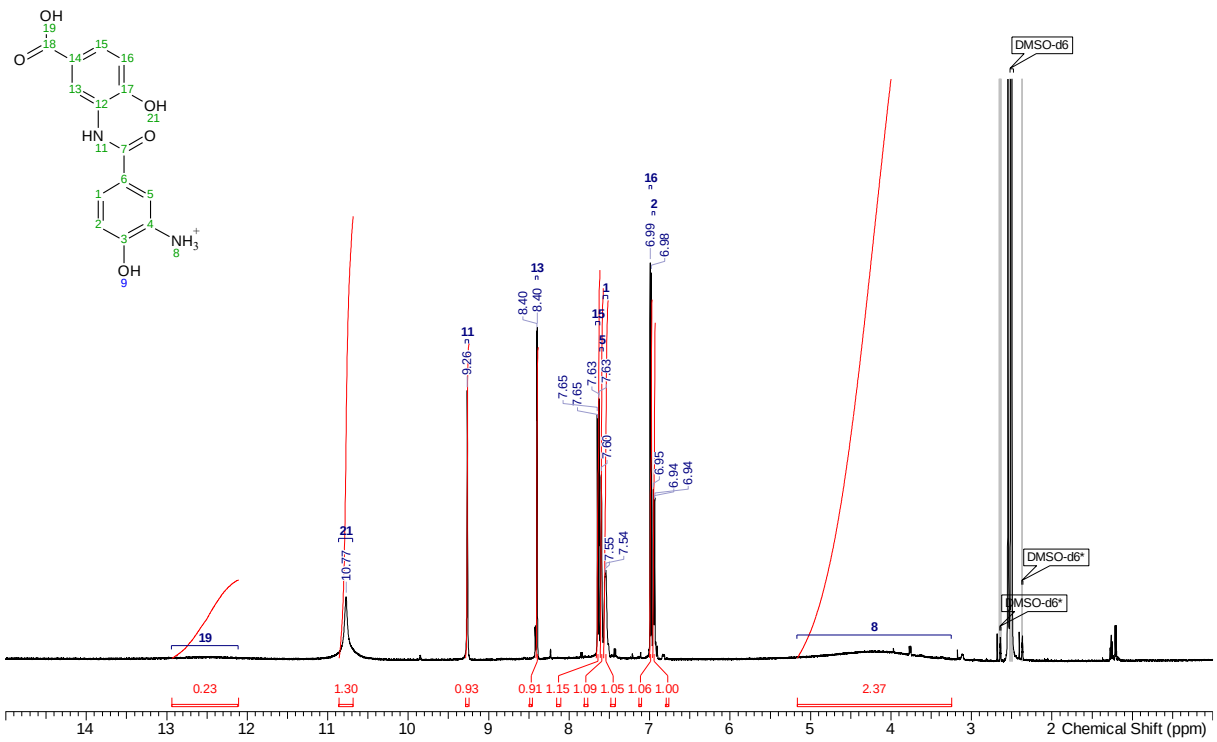


Figure S66. ^{1}H NMR spectrum of **23.**

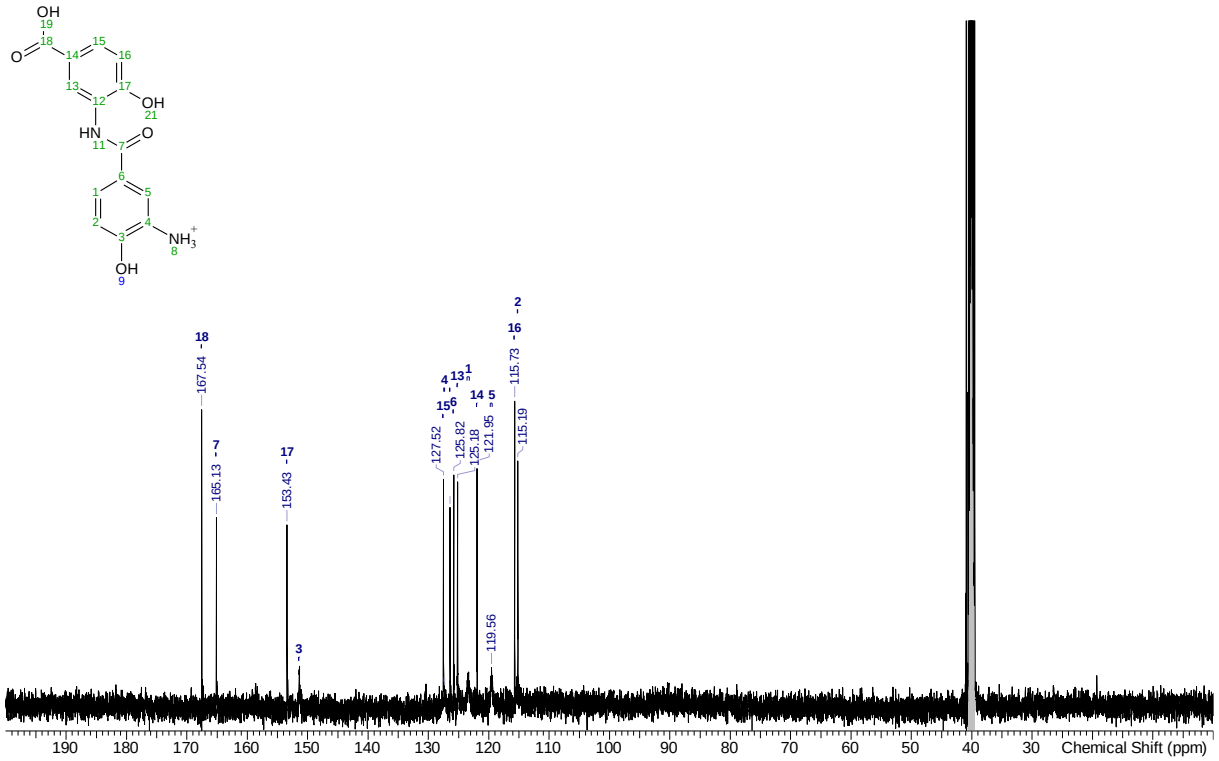


Figure S67. ^{13}C NMR spectrum of **23.**

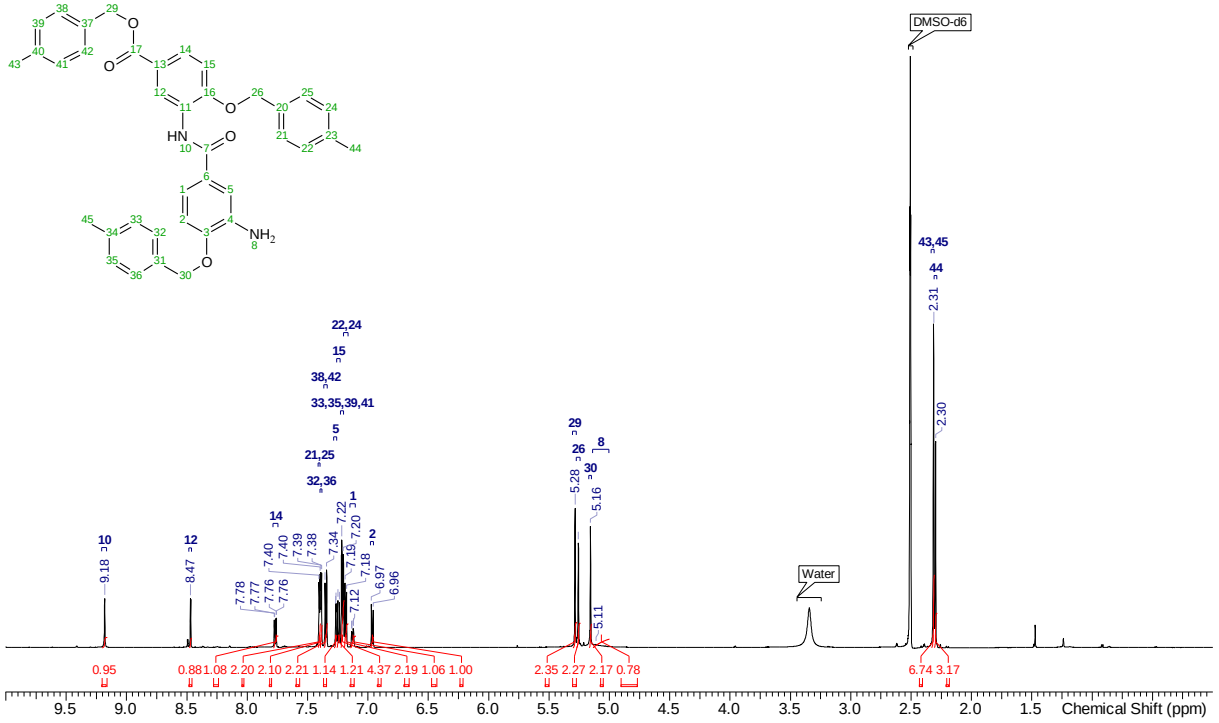


Figure S68. ^{1}H NMR spectrum of **29.**

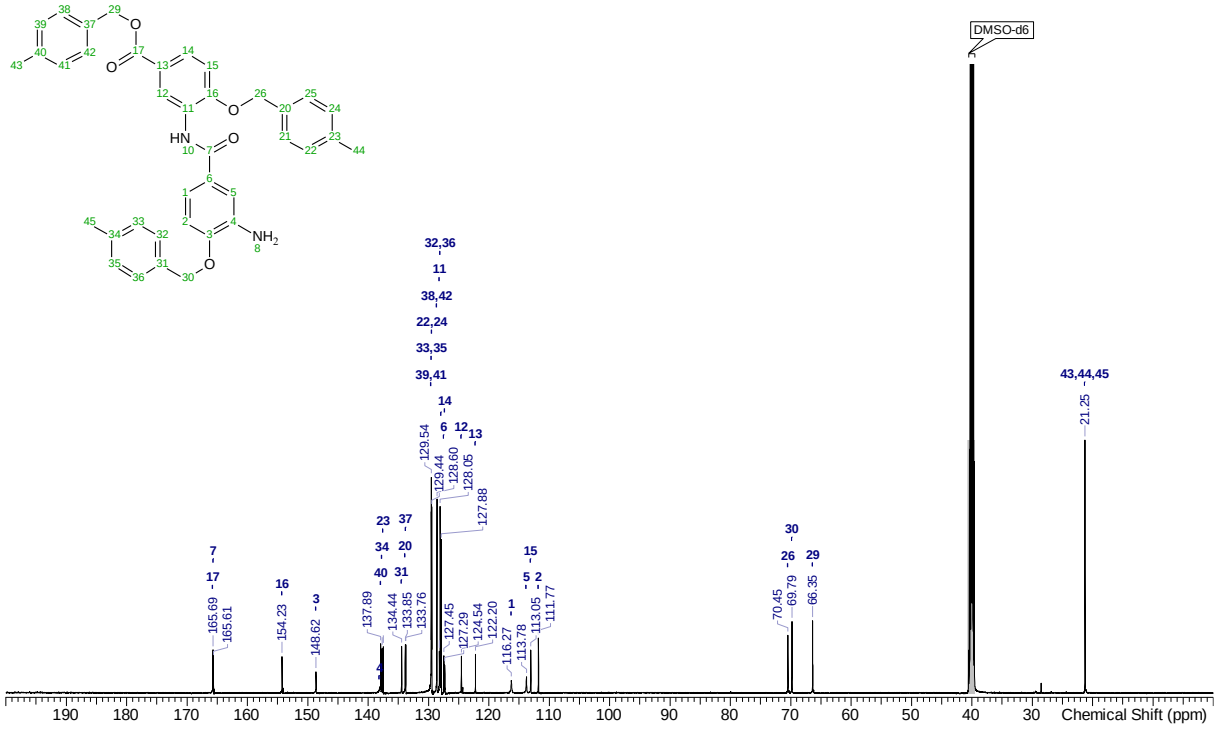


Figure S69. ^{13}C NMR spectrum of **29.**

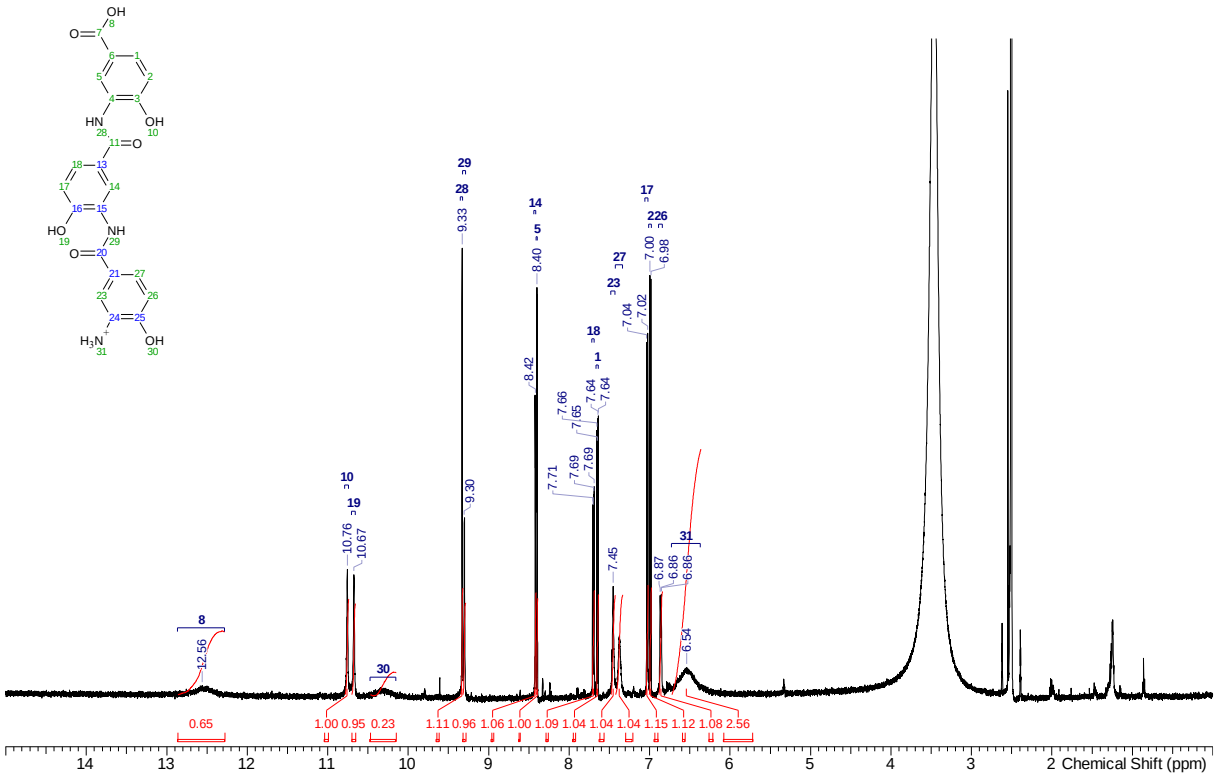


Figure S70. ^{1}H NMR spectrum of **24** after the second preparative HPLC purification.

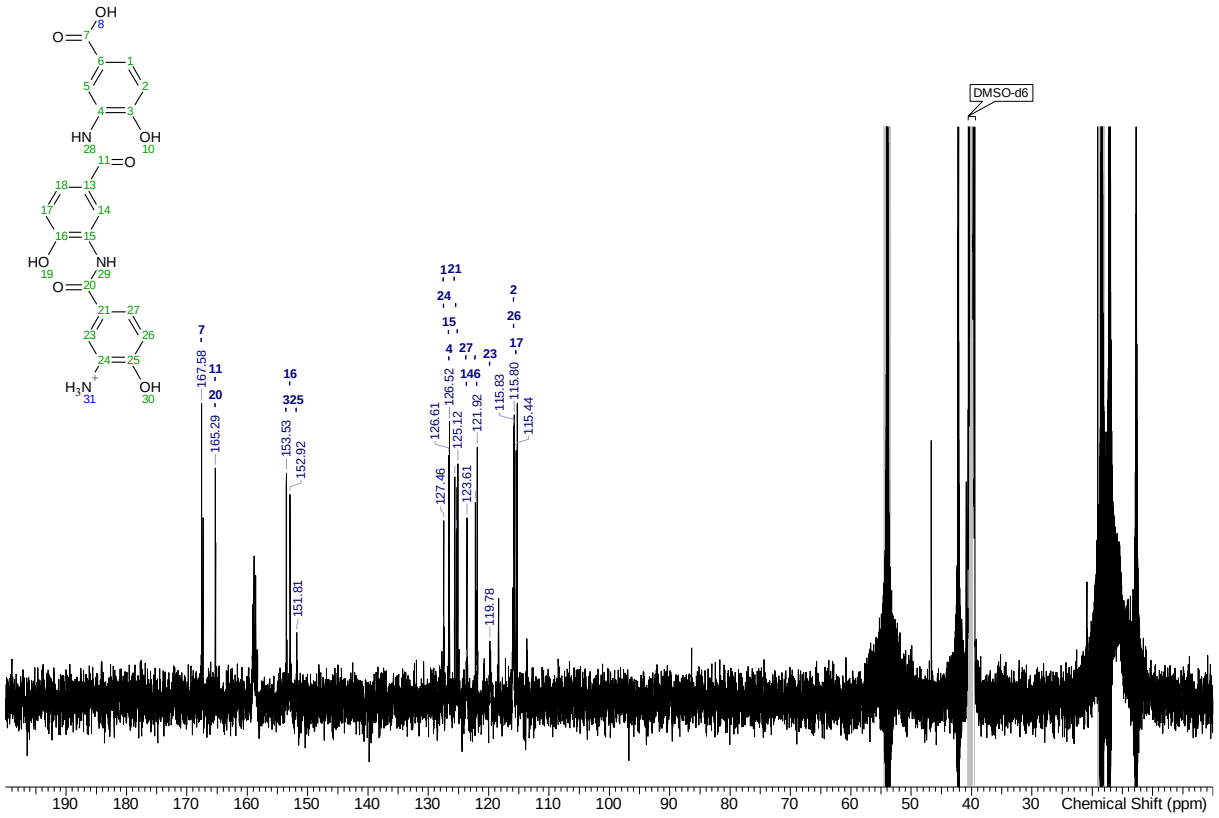


Figure S71. ^{13}C NMR spectrum of **24** after the first preparative HPLC purification. Contains impurities in the range δ = 12.7–54 ppm.

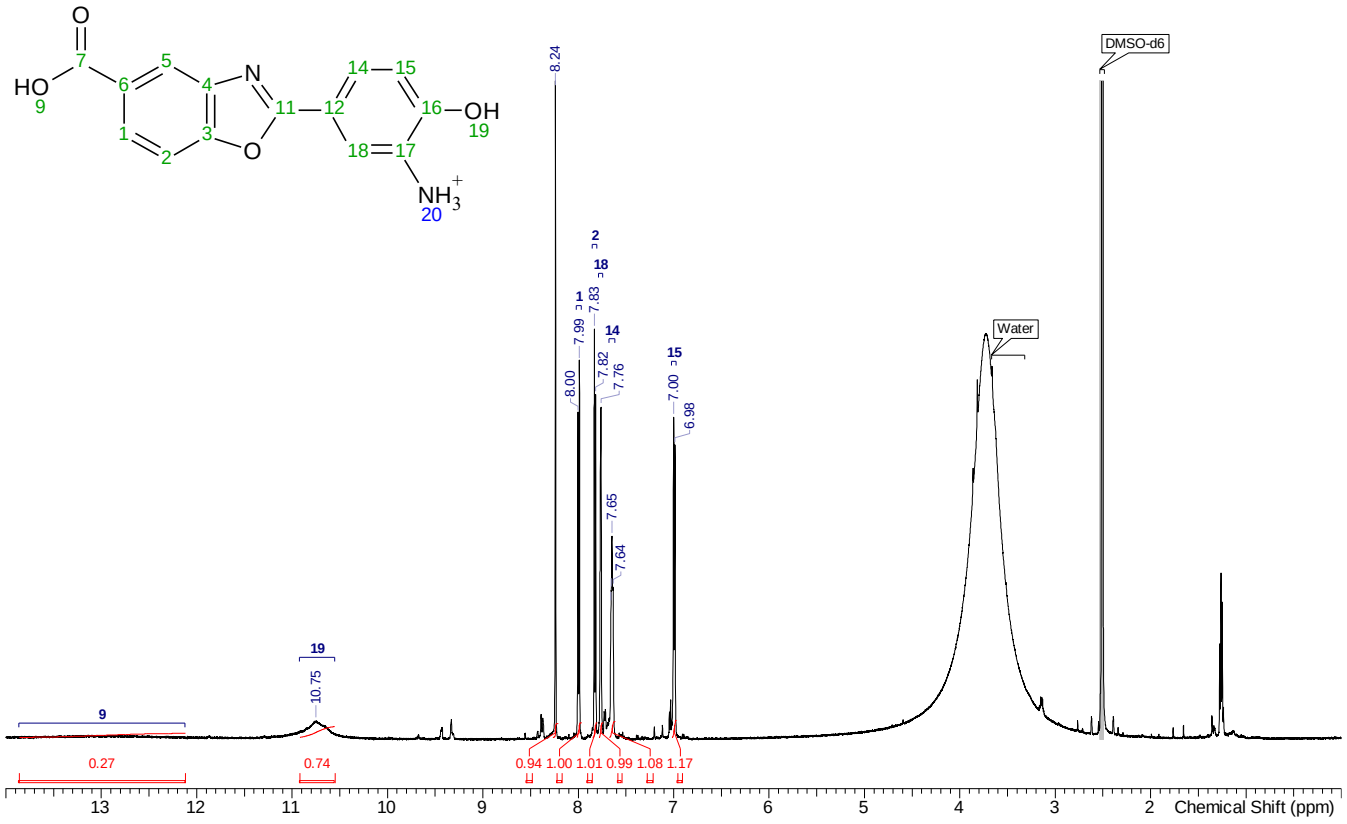


Figure S72. ^{1}H NMR spectrum of **22**.

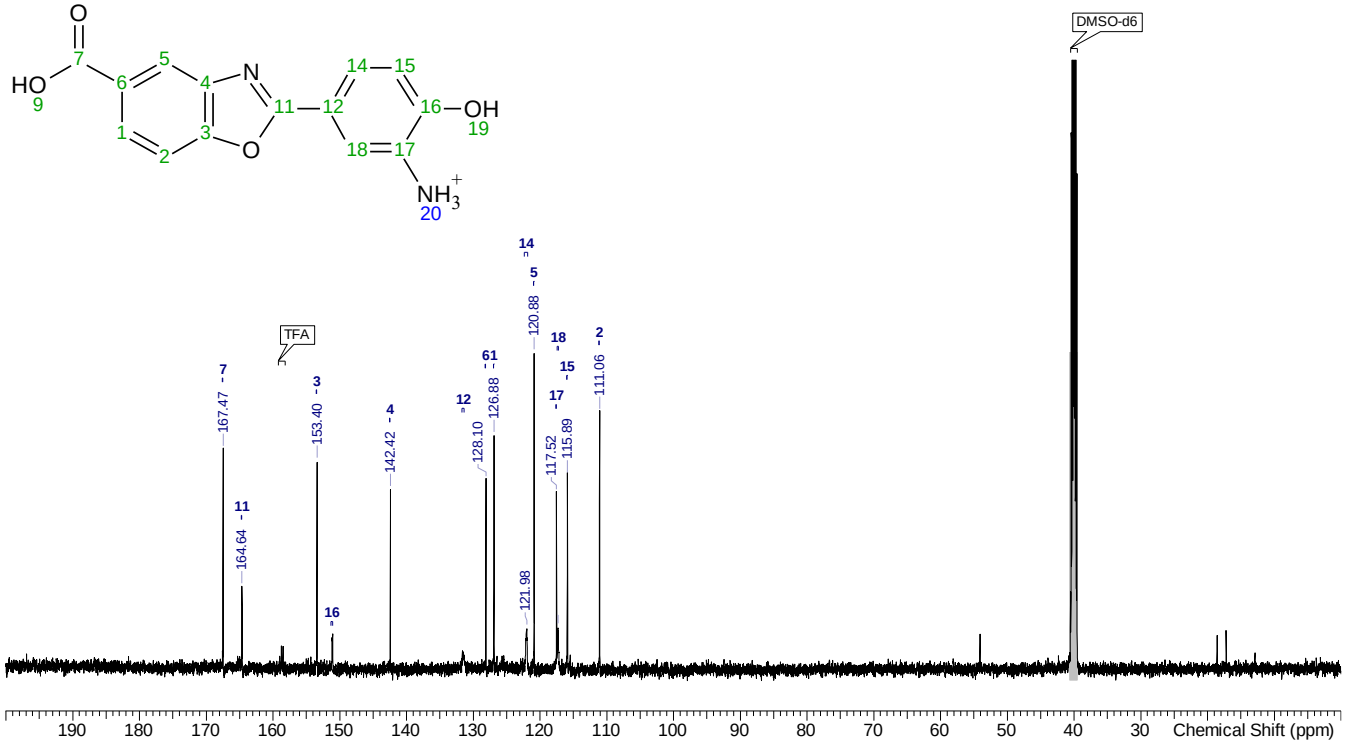


Figure S73. ^{13}C NMR spectrum of **22**.

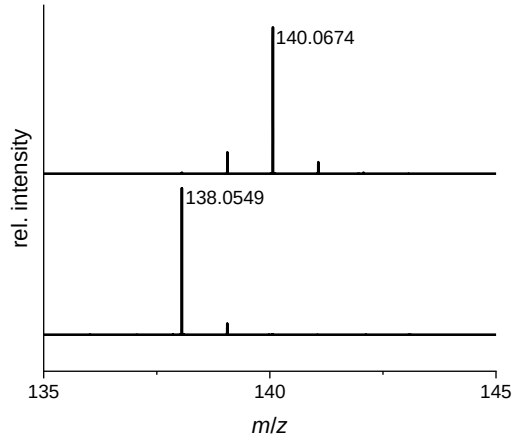


Figure S74. ESI-MS (positive mode) of 4-aminobenzoic acid and 4-aminobenzoic-3,5-d_2 acid (**30**).

NatAM/ClxD hybrid sequence similarity network

To generate a sequence similarity network (SSN) that unites homologs of ClxD and NatAM, we performed several steps of bioinformatic analysis. Both ClxD and NatAM are assigned members of the amidohydrolase superfamily, which consists of over 167,000 sequences with a large sequence diversity coupled with high functional diversity. As a result, our initial SSN that was solely based on ClxD contained a large number of false positives (mostly associated with genes from primary metabolism, e.g. tRNA synthesis and sugar metabolism). We therefore constrained our search to ClxD homologs that are encoded in the vicinity of genes encoding enzymes putatively involved in amide bond formation/esterification, since benzoxazole ring formation requires the initial activation of the precursor molecule.[14] Furthermore, we recognized that our initial ClxD-based SSNs did not contain sequences of characterized heterocyclases involved in benzoxazole formation, like NatAM.[14] This prompted us to create a hybrid SSN derived from two SSNs that were based on ClxD and NatAM, respectively.

We created two independent primary SSNs based on homologs of ClxD and NatAM. Observing the respective Genome Neighborhood Diagrams from the resulting two networks, we manually selected homologs that were encoded in the vicinity of a gene encoding a putative carboxylate-activating enzyme (i.e. belonging to the enzyme family of e.g. CoA-ligases, ATP-grasp or non-ribosomal peptide synthetase), as activation of the precursor molecule is required prior the formation of the benzoxazole ring.[14] Specifically, if a group contained at least three biosynthetic gene clusters with this feature, all sequences of that group were selected to create a combined secondary network of both primary networks. This secondary network consisted of 1101 nodes, with a large amount of false positive sequences, meaning that the surrounding local genomic region indicated that the encoded proteins were putatively not involved in heterocycle formation. Therefore, we performed another round of refinement from this secondary network, to create a third-tier network. For this, we included only sequences of amidohydrolase-like enzymes that were found in synteny to genes encoding for enzymes putatively involved in precursor activation. All nodes corresponding to proteins not fulfilling this criterion were removed from the final cluster. We further excluded sequences we assumed to be involved in primary metabolism (e.g. fatty acid synthesis). The curated set of sequences was submitted to the EFI-EST tool to create the final network (Figure 4A, Table S7). Each node represents a ClxD/NatAM homolog located within a locus putatively involved in biosynthesis of benzoxazole-containing natural products.

Table S7. List of ClxD and NatAM homologs used to construct the sequence similarity network.

UniProt ID	Phylum	Organism	Cluster Number	Name	Phylum	Organism	Cluster Number
A0A022MQ12	Actinobacteria	*Streptomyces* sp. Tu 6176.	1	A0A1H6BQ33	Firmicutes	*Paenibacillus* sp. UNC499MF.	3
A0A069JYK6	Actinobacteria	*Streptomyces* sp. NTK 937.	1	A0A1H7Q462	Proteobacteria	Stigmatella aurantiaca.	3
A0A0F4NQZ7	Proteobacteria	*Pseudoalteromonas piscicida*	1	A0A1H7VL29	Proteobacteria	*Stigmatella aurantiaca*.	3
A0A0M3TGE0	Actinobacteria	*Streptomyces* sp. NRRL 12068.	1	A0A1I0CG30	Proteobacteria	*Stigmatella erecta*.	3
A0A0S2JZW8	Proteobacteria	*Pseudoalteromonas phenolica*	1	A0A1I0K391	Proteobacteria	*Stigmatella erecta.*	3
A0A1A9CVR2	Actinobacteria	*Streptomyces* sp. OspMP-M45.	1	A0A1L3MR92	Firmicutes	*Bacillus weihaiensis.*	3
A0A1A9J422	Actinobacteria	*Streptomyces* sp. SAT1.	1	A0A1L8I6X9	Firmicutes	*Bacillaceae bacterium* G1.	3
A0A1B9ESB4	Actinobacteria	*Streptomyces* sp. PTY087I2.	1	A0A1L8I7L1	Firmicutes	*Bacillaceae bacterium* G1.	3
A0A1C3EJY0	Proteobacteria	*Enterovibrio pacificus.*	1	A0A1L9AUS5	Proteobacteria	*Cystobacter ferrugineus.*	3
A0A1I1EPQ5	Proteobacteria	*Pseudoalteromonas denitrificans* DSM 6059.	1	A0A1L9BFB9	Proteobacteria	*Cystobacter ferrugineus.*	3
A0A1S1Q9W8	Actinobacteria	*Frankia soli.*	1	A0A1L9BJS0	Proteobacteria	*Cystobacter ferrugineus.*	3
A0A1S2NZ06	Actinobacteria	*Streptomyces* sp. MUSC 93.	1	A0A1M6IZP1	Firmicutes	*Clostridium cavendishii* DSM 21758.	3
A0A1T3NS67	Actinobacteria	*Embleya scabrispora*.	1	A0A250J4T2	Proteobacteria	*Cystobacter fuscus.*	3
A0A1V2PKF7	Actinobacteria	*Actinosynnema* sp. ALI-1.44.	1	A0A250JDR1	Proteobacteria	*Cystobacter fuscus.*	3
A0A1V2QX27	Actinobacteria	*Saccharothrix* sp. ALI-22-I.	1	A0A2C1JZD7	Firmicutes	*Bacillus* sp. AFS040349.	3
A0A291RWE1	Actinobacteria	*Nocardia terpenica*.	1	A0A2L0EPE2	Proteobacteria	*Sorangium cellulosum*	3
A0A2G2IY40	Proteobacteria	*Robiginitomaculum* sp.	1	A0A2L0F948	Proteobacteria	*Sorangium cellulosum*	3
A0A2N3XUE2	Actinobacteria	*Saccharopolyspora spinosa*.	1	A0A2N2AFP4	Firmicutes	*Firmicutes* bacterium HGW-Firmicutes-8.	3
A0A2R3ZQ04	Actinobacteria	*Streptomyces* sp. FXJ1.235.	1	A0A2N2DCE7	Firmicutes	*Firmicutes* bacterium HGW-Firmicutes-14.	3
A0A2R3ZQ18	Actinobacteria	*Streptomyces* olivaceus.	1	A0A2N6RMA0	Firmicutes	*Bacillus* sp. UMB0899.	3
A0A2S2GIZ8	Actinobacteria	*Streptomyces* sp. SM18.	1	A0A327S4I1	Firmicutes	*Bacillus* sp. YR335.	3
A0A2X3LF49	Actinobacteria	*Frankia* sp. Ea1.12.	1	A0A353GUG8	Firmicutes	*Firmicutes* bacterium.	3
A0A327SUQ6	Actinobacteria	*Streptomyces* sp. DpondAA-E10.	1	A0A358R170	Firmicutes	*Desulfotomaculum* sp.	3
A0A345HY52	Actinobacteria	*Streptomyces* sp. GSSD-12.	1	A0A369BHC4	Firmicutes	*Anaerobacterium chartisolvens*.	3
A0A353K7B1	Actinobacteria	*Streptomyces* sp.	1	A0A3A6N8Y5	Firmicutes	*Ammonifex* sp.	3
A0A384IVX2	Actinobacteria	*Streptomyces* sp. AC1-42W.	1	A0A3N5C9S5	Firmicutes	*Aquisalibacillus elongatus.*	3
A0A3B6XY02	Proteobacteria	*Pseudoalteromonas piscicida*.	1	A0A410WQE7	Firmicutes	*Paenibacillus chitinolyticus.*	3
A0A3G8G7K0	Actinobacteria	*Streptomyces* olivaceus.	1	A0A415S683	Firmicutes	*Ruminococcus gnavus.*	3

A0A3N1KNU6	Actinobacteria	Streptomyces sp. 844.5.	1	A0A4P2PXC6	Proteobacteria	Sorangium cellulosum	3
A0A3R9WZL5	Actinobacteria	*Streptomyces* sp. WAC05292.	1	A0A4P2QL77	Proteobacteria	*Sorangium cellulosum*	3
A0A3T0VUV6	Proteobacteria	*Hahella* sp. KA22.	1	A0A4R1SC97	Firmicutes	*Hydrogenispora ethanolica.*	3
A0A401Z649	Actinobacteria	*Embleya hyalina*.	1	A0A4R2B250	Firmicutes	*Bacillus foraminis.*	3
A0A494UYY7	Actinobacteria	*Streptomyces fungicidicus*.	1	A0A4R2RZ62	Firmicutes	*Heliophilum fasciatum.*	3
A0A495W5P6	Actinobacteria	*Saccharothrix australiensis.*	1	A0A4Y8IMR3	Firmicutes	*Filobacillus milosensis.*	3
A0A495XLR0	Actinobacteria	*Saccharothrix variisporea.*	1	A0A4Z0QP04	Firmicutes	*Desulfosporosinus* sp. Sb-LF.	3
A0A4Q6Z1F2	Proteobacteria	*Pseudoalteromonas* sp. CO342X.	1	A0A5C6W596	Firmicutes	*Bacillus litoralis.*	3
A0A4Q7IIK6	Proteobacteria	*Pseudoalteromonas phenolica*.	1	B0TEU6	Firmicutes	*Heliobacterium modesticaldum* (strain ATCC 51547 / Ice1)	3
A0A4R1DAU6	Actinobacteria	*Frankia* sp. BMG5.11.	1	E3FE84	Proteobacteria	*Stigmatella aurantiaca* (strain DW4/3-1)	3
A0A4R4WL11	Actinobacteria	*Nonomuraea* sp. KC712.	1	F5LT05	Firmicutes	*Paenibacillus* sp. HGF7.	3
A0A4R7BQY5	Actinobacteria	*Streptomyces* sp. BK561.	1	L0FBW9	Firmicutes	*Desulfitobacterium dichloroeliminans* (strain LMG P-21439 / DCA1)	3
A0A4U3MHQ0	Actinobacteria	*Herbidospora galbida*.	1	Q08U29	Proteobacteria	*Stigmatella aurantiaca* (strain DW4/3-1)	3
A0A514K0I8	Actinobacteria	*Streptomyces asterosporus*.	1	S9PHF5	Proteobacteria	*Cystobacter fuscus* DSM 2262.	3
A0A542DPU5	Actinobacteria	*Amycolatopsis cihanbeyliensis*.	1	S9PI54	Proteobacteria	*Cystobacter fuscus* DSM 2262.	3
A0A542Q2C3	Actinobacteria	*Streptomyces* sp. SLBN-118.	1	S9QJL5	Proteobacteria	*Cystobacter fus*cus DSM 2262.	3
A0A542W7M6	Actinobacteria	*Streptomyces* sp. SLBN-134.	1	A0A0S7YXA1	Proteobacteria	*Gammaproteobacteria* bacterium SG8_30.	4
A0A560XIX7	Actinobacteria	*Streptomyces* sp. T12.	1	A0A127F0Q4	Proteobacteria	*Rhodoplanes* sp. Z2-YC6860.	4
A0A5P1YK20	Actinobacteria	*Streptomyces* tendae.	1	A0A1E3ZKZ0	Proteobacteria	*Bordetella* sp. SCN 67-23.	4
A0A5S3RZX5	Proteobacteria	*Pseudoalteromonas flavipulchra.*	1	A0A1I1A2G5	Proteobacteria	*Rhizobium* sp. NFR07.	4
A0A5S3YRH6	Proteobacteria	*Pseudoalteromonas phenolica.*	1	A0A1I4A2B0	Proteobacteria	*Roseomonas stagni* DSM 19981.	4
A0A652K8Z4	Actinobacteria	*Streptomyces* sp. ms191.	1	A0A1I4DT21	Proteobacteria	*Roseomonas stagni* DSM 19981.	4
A0A6B1LQW8	Actinobacteria	*Streptomyces* sp. SID8360.	1	A0A1Q3Y2P9	Proteobacteria	*Burkholderiales* bacterium 67-32.	4
A8L0C1	Actinobacteria	*Frankia* sp. (strain EAN1pec)	1	A0A1W7MCR7	Proteobacteria	*Novosphingobium* sp. MD-1.	4
D3FAQ8	Actinobacteria	Conexibacter woesei (strain DSM 14684 / JCM 11494 / NBRC 100937 /ID131577)	1	A0A1Y5TT08	Proteobacteria	*Oceanibacterium hippocampi*.	4
D9XKS5	Actinobacteria	*Streptomyces* griseoflavus Tu4000.	1	A0A1Z9DWP9	Proteobacteria	*Cellvibrionales* bacterium TMED122.	4
G2NKI2	Actinobacteria	*Streptomyces* sp. (strain SirexAA-E / ActE)	1	A0A1Z9JI97	Proteobacteria	*Rhodospirillaceae* bacterium TMED167.	4
Q2S818	Proteobacteria	*Hahella chejuensis* (strain KCTC 2396)	1	A0A208XTQ8	Proteobacteria	*Pigmentiphaga* sp. NML030171.	4
A0A081XYL9	Actinobacteria	*Streptomyces* toyocaensis.	2	A0A208XXT8	Proteobacteria	*Pigmentiphaga* sp. NML080357.	4
A0A0K3BLZ2	Actinobacteria	*Kibdelosporangium* sp. MJ126-NF4.	2	A0A226WPR1	Proteobacteria	*Caballeronia sordidicola*.	4

A0A0W7X8R0	Actinobacteria	*Streptomyces* silvensis.	2	A0A226WQT7	Proteobacteria	*Caballeronia sordidicola.*	4
A0A132MUD4	Actinobacteria	*Streptomyces* thermoautotrophicus.	2	A0A2D9HTF5	Other	*Dehalococcoidia bacterium.*	4
A0A132N6J5	Actinobacteria	*Streptomyces* thermoautotrophicus.	2	A0A2E2P9F9	Proteobacteria	*Pseudooceanicola* sp.	4
A0A1B1MGF2	Actinobacteria	*Streptomyces* lincolnensis.	2	A0A3D3DQY0	Proteobacteria	*Gammaproteobacteria* bacterium.	4
A0A1C6RY45	Actinobacteria	*Micromonospora inyonensis.*	2	A0A3G8GL07	Proteobacteria	*Pigmentiphaga* sp. H8.	4
A0A1C6TJD2	Actinobacteria	*Micromonospora pallida.*	2	A0A3R9YDS3	Proteobacteria	*Mesorhizobium carbonis*.	4
A0A1I5AS71	Actinobacteria	*Streptomyces* sp. cf124.	2	A0A4D7QKP7	Proteobacteria	*Phreatobacter* sp. NMCR1094.	4
A0A1M7KXY3	Actinobacteria	*Cryptosporangium aurantiacum*.	2	A0A4V2F345	Proteobacteria	*Pigmentiphaga kullae.*	4
A0A1M7Q7X0	Actinobacteria	*Streptomyces* paucisporeus.	2	A0A520NMM0	Proteobacteria	*Rhodospirillaceae bacterium.*	4
A0A1Q5HKL7	Actinobacteria	*Micromonospora* sp. TSRI0369.	2	A0A523FZ10	Proteobacteria	*Alphaproteobacteria bacterium.*	4
A0A1Q7BS35	Actinobacteria	Actinobacteria bacterium 13_2_20CM_2_71_6.	2	A0A561R7J2	Proteobacteria	*Neorhizobium alkalisoli.*	4
A0A1Q7WC88	Actinobacteria	*Actinobacteria* bacterium 13_1_20CM_3_71_11.	2	A0A5B0X485	Proteobacteria	*Halioglobus* sp. NY5.	4
A0A1Q8C5E0	Actinobacteria	*Actinophytocola xanthii*.	2	A0A5B8LX18	Proteobacteria	*Devosia ginsengisoli*.	4
A0A1R3UQE1	Actinobacteria	*Nocardiopsis* sp. JB363.	2	A0A5N7RPN6	Proteobacteria	Alcaligenaceae bacterium SAGV5.	4
A0A1R3UUX6	Actinobacteria	*Nocardiopsis* sp. JB363.	2	Q0FX15	Proteobacteria	Salipiger bermudensis (strain DSM 26914 / JCM 13377 / KCTC 12554 /HTCC2601)	4
A0A1V2QKL5	Actinobacteria	*Actinosynnema* sp. ALI-1.44.	2	A0A4R4X5S5	Actinobacteria	*Nonomuraea* sp. KC712.	5
A0A250JBU6	Proteobacteria	*Cystobacter fuscus*.	2	A0A351CR38	Actinobacteria	*Micrococcaceae* bacterium.	5
A0A2K9F6Q1	Actinobacteria	*Streptomyces* sp. CMB-StM0423.	2	A0A2Z5JSH8	Actinobacteria	*Streptomyces* atratus.	5
A0A2M8LVH4	Actinobacteria	*Streptomyces* carminius.	2	A0A2M9MEH7	Actinobacteria	*Streptomyces* sp. CB02959.	5
A0A2M9AKN2	Actinobacteria	*Streptomyces* sp. CNZ306.	2	A0A1G7Z4A2	Actinobacteria	*Nonomuraea jiangxiensis.*	5
A0A2N0GUU5	Actinobacteria	*Streptomyces* sp. Ag109_G2-1.	2	A0A2N0LDN1	Other	SAR202 cluster bacterium Io17-Chloro-G3.	5
A0A2P2GUI3	Actinobacteria	*Streptomyces* showdoensis.	2	A0A2E3AKC5	Other	*Chloroflexi* bacterium.	5
A0A2T0LTG0	Actinobacteria	*Prauserella shujinwangii.*	2	C5CJW1	Proteobacteria	*Variovorax paradoxus* (strain S110)	5
A0A316EXH0	Actinobacteria	*Actinoplanes xinjiangensis.*	2	A0A561CK72	Proteobacteria	*Variovorax* sp. T529.	5
A0A317RQL4	Actinobacteria	*Actinokineospora mzabensis.*	2	A0A3P3EJH2	Proteobacteria	*Variovorax* sp. T529.	5
A0A372GGC5	Actinobacteria	*Actinomadura* sp. LHW52907.	2	A0A1H4DZB1	Proteobacteria	*Variovorax* sp. YR216.	5
A0A385ATE4	Actinobacteria	*Micromonospora* sp. B006.	2	S9TMZ5	Proteobacteria	*Ralstonia* sp. AU12-08.	6
A0A3D9ZUZ7	Actinobacteria	*Asanoa ferruginea*.	2	A0A401K1J1	Proteobacteria	*Ralstonia* sp. SET104.	6
A0A3N5AKY6	Actinobacteria	*Streptomyces* sp. Ag109_G2-6.	2	A0A2N5L629	Proteobacteria	*Ralstonia mannitolilytica*.	6

A0A3S8X5S5	Actinobacteria	*Streptomyces* sp. WAC 01438.	2	A0A1I5RUY0	Proteobacteria	*Ralstonia* sp. NFACC01.	6
A0A423V2B7	Actinobacteria	*Streptomyces globisporus*.	2	A0A1C0XPQ9	Proteobacteria	*Ralstonia pickettii*	6
A0A428WTV2	Actinobacteria	*Actinoplanes* sp. ATCC 53533.	2	A0A192A6J3	Proteobacteria	*Ralstonia insidiosa.*	6
A0A428YCH1	Actinobacteria	*Kibdelosporangium aridum.*	2	A0A100HUI7	Proteobacteria	*Ralstonia* sp. NT80.	6
A0A429A977	Actinobacteria	*Streptomyces* sp. WAC 01420.	2	A0A0G3EQ02	Proteobacteria	*Pandoraea thiooxydans.*	6
A0A495JN82	Actinobacteria	*Micromonospora* pisi.	2	A0A0F0EA18	Proteobacteria	*Burkholderiaceae* bacterium 26.	6
A0A495QKK8	Actinobacteria	*Actinomadura pelletieri* DSM 43383.	2	A0A1V2KUV2	Actinobacteria	*Frankia* sp. CcI49.	7
A0A4D4JEJ6	Actinobacteria	*Gandjariella thermophila*.	2	A0A0S4QH26	Actinobacteria	*Frankia irregularis*.	7
A0A4Q8BQ11	Actinobacteria	*Streptomyces* sp. CNZ288.	2	A0A0N8HBW1	Actinobacteria	*Frankia* sp. R43.	7
A0A4R3S2P3	Actinobacteria	*Streptomyces* sp. BK215.	2	A0A1F7P2J3	Other	Candidatus Rokubacteria bacterium RIFCSPLOWO2_02_FULL_72_37.	8
A0A4R6JCI3	Actinobacteria	*Actinoplanes brasiliensis*.	2	A0A1F3YND1	Proteobacteria	Betaproteobacteria bacterium RBG_16_64_18.	8
A0A505HIW9	Actinobacteria	*Micromonospora* sp. XM-20-01.	2	A0A140GVS8	Proteobacteria	*Rhodoplanes* sp. Z2-YC6860.	8
A0A538ER75	Actinobacteria	*Actinobacteria* bacterium.	2	A0A378REZ6	Actinobacteria	*Mycolicibacterium aichiense*.	9
A0A538N4C9	Actinobacteria	*Actinobacteria* bacterium.	2	A0A1J0UEI2	Actinobacteria	*Mycobacterium* sp. WY10.	9
A0A542W6J4	Actinobacteria	*Streptomyces* sp. SLBN-134.	2	A0A0Q9JAR7	Actinobacteria	*Mycobacterium* sp. Soil538.	9
A0A559UE18	Actinobacteria	*Streptomyces* sp. CNZ289.	2	A0A3D1QC12	Firmicutes	*Syntrophomonas* sp.	10
A0A562WC22	Actinobacteria	*Micromonospora sagamiensis.*	2	A0A356UE87	Firmicutes	*Syntrophomonas* sp.	10
D9XNL5	Actinobacteria	*Streptomyces griseoflavus* Tu4000.	2	A0A1V5CUD0	Proteobacteria	*Syntrophorhabdus* sp. PtaU1.Bin058.	10
H8G469	Actinobacteria	Saccharomonospora azurea NA-128.	2	A0A1T3P4I3	Actinobacteria	*Embleya scabrispora*.	11
I1WE21	Actinobacteria	*Streptomyces* sp. WT3.	2	A0A0P6XQ18	Other	*Thermanaerothrix daxensis*.	11
R4LPJ6	Actinobacteria	*Actinoplanes* sp. N902-109.	2	A0A1M6QP99	Proteobacteria	*Desulfatibacillum alkenivorans* DSM 16219.	11
S9Q2R2	Proteobacteria	*Cystobacter fuscus* DSM 2262.	2	A0A3A4TTV9	Firmicutes	*Firmicutes* bacterium.	12
W7V9Z2	Actinobacteria	*Micromonospora* sp. M42.	2	A0A3M1R8R6	other	*Nitrospinae* bacterium.	12
A0A017SY51	Proteobacteria	*Chondromyces apiculatus* DSM 436.	3	A0A5J6MM47	Proteobacteria	*Hypericibacter terrae*.	12
A0A0E4GCE9	Firmicutes	*Syntrophomonas zehnderi* OL-4.	3	A0A662VY29	Other	*Archaeoglobales archaeon.*	13
A0A0K1E9U0	Proteobacteria	*Chondromyces crocatus.*	3	A0A661JQ04	Proteobacteria	*Deltaproteobacteria* bacterium.	13
A0A0M1N1E3	Firmicutes	*Paenibacillus solani.*	3	A0A3A4PJF5	Actinobacteria	*Actinobacteria* bacterium.	singleton
A0A0Q0XDI6	Proteobacteria	*Acidovorax* sp. SD340.	3	A0A197SU18	Actinobacteria	*Streptomyces* sp. ERV7.	singleton
A0A0Q6L5W3	Firmicutes	*Bacillus* sp. Leaf406.	3	A0A3R7TJK6	Bacteroidetes	*Flavobacteriales* bacterium TMED96.	singleton
A0A150TEY9	Proteobacteria	*Sorangium cellulosum*	3	A0A1W9RSX8	Other	Candidatus Omnitrophica bacterium 4484_70.2.	singleton

A0A179SSS7	Firmicutes	*Bacillus litoralis.*	3	A0A1V4SP16	Firmicutes	*Ruminiclostridium hungatei*	singleton
A0A179SYB7	Firmicutes	*Bacillus litoralis.*	3	A0A2E4MRF2	Other	*Gemmatimonadetes bacterium.*	singleton
A0A1G3LDE2	Other	*Spirochaetes* bacterium GWB1_27_13.	3	E1YGH5	Proteobacteria	*Desulfobacterium* sp.	singleton
				A0A3G8GUG1	Proteobacteria	*Pigmentiphaga* sp. H8.	singleton
				A0A291IHW0	Proteobacteria	*Methylomonas koyamae.*	singleton
				A0A1M5ZE55	Proteobacteria	*Candidimonas bauzanensis.*	singleton

Identification of a putative nocarbenzoxazole BGC

We used BLAST to find that *N. lucentensis* DSM 44048 genome contains a gene encoding a ClxD homolog (the whole genome sequence of *N. lucentensis* DSM 44048 is not deposited in the Uniprot Database meaning that it is absent from the SSN). Upon closer inspection of the *N. lucentensis* DSM 44048 locus, it became apparent that genes encoding ClxA, ClxB and ClxE homologs are present in the vicinity of the gene encoding the *clxD* homolog (Figure S75). Drawing on our knowledge of the function ClxA–E in closoxazole biosynthesis and taking the structure of nocarbenzoxazole G (**6**) into account, we could envision a putative biosynthetic route to the *meta*-substituted benzoxazole core of nocarbenzoxazole G (**6**). We identified additional genes in the genome neighborhood that could possibly encode the necessary enzymes for further modification of the core structure. We tentatively named this locus *noc*.

Specifically, the *noc* locus harbors genes encoding homologs of ClxB (*nocB*) and ClxE (*nocC*) that we assign as likely involved in the formation of 3,4-AHBA, which has been proposed as the required building block for nocarbenzoxazole G.[15] The putative aspartokinase (*nocA*) is probably involved in early steps to form aspartate semialdehyde, a precursor in 3,4-AHBA formation.[16] Furthermore, we found genes encoding homologs of ClxA/C (*nocN*) and ClxD (*nocM*), which could catalyze fusion and heterocyclization of two 3,4-AHBA moieties resulting in the benzoxazole **22** (Figure 4C). We envision that subsequent tailoring steps leading to the cytotoxic nocarbenzoxazole G (**6**) may be catalyzed by proteins encoded by *nocJ–L* and *nocD,E,G*. NocJ shows similarity to EhpF, a member of the ANL superfamily that is believed to act together with EhpG, a putative NAD(P)H-dependent aldehyde dehydrogenase, in the reduction of a carboxylic acid to the corresponding aldehyde in D-alanylgriseoluteic acid biosynthesis.[17] In analogy, the enzyme encoded by *nocJ* and the putative aldehyde dehydrogenase encoded in *nocK* may co-operate to reduce the carboxylic acid of **22** to form the corresponding aldehyde. The gene *nocL* is annotated to encode an alcohol dehydrogenase which likely catalyzes further reduction of the aldehyde to afford the corresponding alcohol. Consequently, NocJ–L are probably involved in the reduction of the carboxylic acid of the dimer benzoxazole **22** to form the corresponding alcohol. Finally, NocG is a predicted FAD-binding oxidoreductase, and a member of this family has recently been reported to conduct deamination in saframycin A biosynthesis.[18] The gene *nocD* encodes a member of the tyrosinase family and *nocE* its cofactor, which are known to perform oxidation of phenols.[19] Therefore, NocD, E and G possibly act together in deamination of the hydroxyanilin moiety. We suspect that subsequent methylation to yield the final product **6** is probably conducted by an *O*-methyltransferase encoded elsewhere in the genome, as proposed in nataxazole biosynthesis.[20] Our characterization of the *clx* cluster and bioinformatic analysis lead us to identify a strong candidate locus for nocarbenzoxazole biosynthesis, potentially laying the foundation to link an orphan natural product to its cognate BGC.

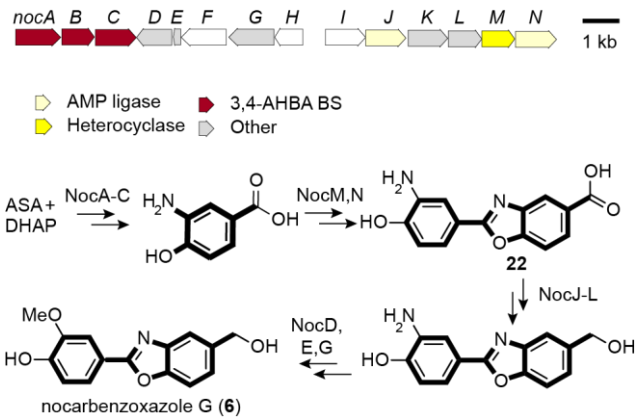


Figure S75. Proposed pathway for biosynthesis of nocarbenzoxazole G by *N. lucentensis* DSM 44048. ASA: L-aspartate-4-semialdehyde. BS: Biosynthesis. DHAP: dihydroxyacetone phosphate.

Table S8. Annotated functions of open reading frames of the putative nocarbenzoxazole biosynthetic gene cluster.

***N. lucentensis* DSM 44048**				***Nocardiopsis* sp. JB363**			**Identity/Similarity**
Gene	**Locus Tag**	**Size [aa]**	**Annotated function**	**Locus Tag**	**Size [aa]**	**Annotated function**	
nocA	D471_RS0121390	435	Aspartate kinase	BQ8420_20030	433	Aspartokinase	53.5/64.2
nocB	D471_RS0121385	273	Hypothetical protein	BQ8420_20035	273	2-amino-3,7-dideoxy-D-threo-hept-6-ulosonate synthase	72.9/83.2
nocC	D471_RS0121380	366	3-dehydroquinate synthase II family protein	BQ8420_20040	366	3,7-dideoxy-D-threo-hepto-2, 6-diulosonate synthase	84.2/91.6
nocD	D471_RS0121375	305	Tyrosinase family protein	BQ8420_20045	304	Putative tyrosinase	80.3/87.9
nocE	D471_RS0121365	108	Hypothetical protein	BQ8420_20050	99	Tyrosinase cofactor	49.5/61.6
nocF	D471_RS0121365	461	MFS transporter	BQ8420_20055	426	Major facilitator family transporter	65.2/76.6
nocG	D471_RS0121360	470	FAD-binding oxidoreductase	BQ8420_20060	470	FAD-dependent oxygenase	63.7/76.6
nocH	D471_RS0121360	222	AfsR/SARP family transcriptional regulator	BQ8420_20065	234	Putative regulatory protein (AfsR-like protein)	61.7/74.8
nocI	D471_RS0121350	381	helix-turn-helix transcriptional regulator	BQ8420_20070	376	HTH luxR-type domain-containing protein	69.0/78.0
nocJ	D471_RS0121345	364	AMP-binding protein	BQ8420_20075	363	EhpF	70.3/80.2
nocK	D471_RS35455-D471_RS0121335	369		BQ8420_20080	454	Aldehyde dehydrogenase	61.9/71.8
nocL	D471_RS0121330	351	NAD(P)-dependent Alcohol dehydrogenase	BQ8420_20085	335	Alcohol dehydrogenase	65.8/79.6
nocM	D471_RS0121325	253	Amidohydrolase family protein	BQ8420_20090	251	Amidohydro-rel domain-containing protein	67.9/75.8
nocN	D471_RS0121320	437	Phenylacetate--CoA ligase family protein	BQ8420_20095	439	Phenylacetate-coenzyme A ligase	65.7/73.0

Author contributions

Research design: C.H.
Conceptualization E.M., T.H.,
Cloning and genetic manipulation of *Clostridium* spp.: T.H., E.M.
Ribosome engineering: E.M.
Optimization of extraction, fermentation: T.H., E.M.
Isolation and structure elucidation: T.H., K.S., K.D.
Synthesis of authentic references and probes: V.H., J.F.
Bioinformatics analyses, gene cluster identification, SSN: T.H., K.D.
Cloning and heterologous expression: F.B.
Precursor directed biosynthesis: F.B., T.H.
Bioassays: T.H., E.M.
Compiling data, writing paper, figures: T.H., E.M., F.B., V.H.
Supervision, writing – Review and Editing: C.H.

References

[1] W. Higashide, Y. Li, Y. Yang, J. C. Liao, *Appl. Environ. Microbiol.* **2011**, *77*, 2727-2733.
[2] G.-z. Cui, W. Hong, J. Zhang, W.-l. Li, Y. Feng, Y.-j. Liu, Q. Cui, *J. Microbiol. Methods* **2012**, *89*, 201-208.
[3] T. Lincke, S. Behnken, K. Ishida, M. Roth, C. Hertweck, *Angew. Chem. Int. Ed.* **2010**, *49*, 2011-2013.
[4] P. J. Rutledge, G. L. Challis, *Nat. Rev. Microbiol.* **2015**, *13*, 509-523.
[5] a) H. Aubry-Damon, C.-J. Soussy, P. Courvalin, *Antimicrob. Agents Chemother.* **1998**, *42*, 2590-2594; b) B. P. Goldstein, *J. Antibiot.* **2014**, *67*, 625-630.
[6] K. L. Dunbar, H. Büttner, E. M. Molloy, M. Dell, J. Kumpfmüller, C. Hertweck, *Angew. Chem. Int. Ed.* **2018**, *57*, 14080-14084.
[7] K. Blin, L. E. Pedersen, T. Weber, S. Y. Lee, *Synth. Syst. Biotechnol.* **2016**, *1*, 118-121.
[8] G. R. Fulmer, A. J. Miller, N. H. Sherden, H. E. Gottlieb, A. Nudelman, B. M. Stoltz, J. E. Bercaw, K. I. Goldberg, *Organometallics* **2010**, *29*, 2176-2179.
[9] G. M. Boratyn, A. A. Schäffer, R. Agarwala, S. F. Altschul, D. J. Lipman, T. L. Madden, *Biol. Direct* **2012**, *7*, 12.
[10] L. Zimmermann, A. Stephens, S.-Z. Nam, D. Rau, J. Kübler, M. Lozajic, F. Gabler, J. Söding, A. N. Lupas, V. Alva, *J. Mol. Biol.* **2018**, *430*, 2237-2243.
[11] F. Madeira, M. Pearce, A. Tivey, P. Basutkar, J. Lee, O. Edbali, N. Madhusoodanan, A. Kolesnikov, R. Lopez, *Nucleic Acids Res.*, DOI: 10.1093/nar/gkac240.
[12] R. Zallot, N. Oberg, J. A. Gerlt, *Biochemistry* **2019**, *58*, 4169-4182.
[13] P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, T. Ideker, *Genome Res.* **2003**, *13*, 2498-2504.
[14] H. Song, C. Rao, Z. Deng, Y. Yu, J. H. Naismith, *Angew. Chem. Int. Ed.* **2020**, *59*, 6054-6061.
[15] M. Sun, X. Zhang, H. Hao, W. Li, C. Lu, *J. Nat. Prod.* **2015**, *78*, 2123-2127.
[16] H. Suzuki, Y. Ohnishi, Y. Furusho, S. Sakuda, S. Horinouchi, *J. Biol. Chem.* **2006**, *281*, 36944-36951.
[17] A. K. Bera, V. Atanasova, S. Gamage, H. Robinson, J. F. Parsons, *Acta Crystallogr. Sect. D: Biol. Crystallogr.* **2010**, *66*, 664-672.
[18] L. Q. Song, Y. Y. Zhang, J. Y. Pu, M. C. Tang, C. Peng, G. L. Tang, *Angew. Chem. Int. Ed.* **2017**, *129*, 9244-9248.
[19] Y. Wang, F. Zhai, Y. Hasebe, H. Jia, Z. Zhang, *Bioelectrochemistry* **2018**, *122*, 174-182.
[20] C. Cano‐Prieto, R. García‐Salcedo, M. Sánchez‐Hidalgo, A. F. Braña, H. P. Fiedler, C. Méndez, J. A. Salas, C. Olano, *ChemBioChem* **2015**, *16*, 1461-1473.