

## Supplementary Information

### Discovery and biosynthesis of bosamycin from *Streptomyces* sp. 120454

Zi Fei Xu,<sup>†</sup> Sheng Tao Bo,<sup>†</sup> Mei Jing Wang, Jing Shi, Yang Sun, Rui Hua Jiao, Qiang Xu, Ren Xiang Tan,\* Hui Ming Ge<sup>\*a</sup>

## Table of Contents

### Experimental procedures:

Heterologous expression of cosmid pHG06015 in *Streptomyces albus* J1074.

Co-expression of *bsmF*, *bsmG* and *bsmH* in *Streptomyces albus* J1074 /pHG06015 strain.

Isolation and purification of bosamycins from *S. sp.* 120454 wild-type and recombinant strains.

Physical data for bosamycins.

Preparation and analysis of Marfey's derivatives.

Chemical synthesis of 5-OMe Tyr.

Chemical synthesis of 5-OH Tyr.

Chemical complementation of 5-OMe Tyr or 5-OH Tyr into  $\Delta bsmF$ ,  $\Delta bsmG$  and  $\Delta bsmH$  mutant.

Chemical complementation of L-erythro- $\beta$ -OH-Asp into  $\Delta bsmC$  mutant.

Protein expression and purification.

Adenylation activities of A domain.

*In vitro* assay of BsmH.

Biological activity assay of SHP2.

### Supplementary tables:

**Table S1.** Bacterial plasmids and strains.

**Table S2.** Oligonucleotide primers used in this study.

**Table S3.** Substrate specificity predictions for the adenylation domains of the NRPSs encoded in the *bsm* gene cluster.

**Table S4.**  $^1\text{H}$  (600 MHz) and  $^{13}\text{C}$  (150 MHz) NMR data of compound **1** in DMSO- $d_6$ .

**Table S5.**  $^1\text{H}$  (600 MHz) and  $^{13}\text{C}$  (150 MHz) NMR data of compound **2** in DMSO- $d_6$ .

**Table S6.**  $^1\text{H}$  (600 MHz) and  $^{13}\text{C}$  (150 MHz) NMR data of compound **3** in DMSO- $d_6$ .

**Table S7.**  $^1\text{H}$  (600 MHz) and  $^{13}\text{C}$  (150 MHz) NMR data of compound **4** in DMSO- $d_6$ .

**Table S8.**  $^1\text{H}$  (600 MHz) and  $^{13}\text{C}$  (150 MHz) NMR data of compound **5** in DMSO- $d_6$ .

**Table S9.**  $^1\text{H}$  (600 MHz) and  $^{13}\text{C}$  (150 MHz) NMR data of compound **6** in DMSO- $d_6$ .

**Table S10.**  $^1\text{H}$  (600 MHz) and  $^{13}\text{C}$  (150 MHz) NMR data of compound **13** in DMSO- $d_6$ .

**Table S11.**  $^1\text{H}$  (600 MHz) and  $^{13}\text{C}$  (150 MHz) NMR data of compound **14** in DMSO- $d_6$ .

**Table S12.**  $^1\text{H}$  NMR (400 MHz) data of compound **A-E** in DMSO- $d_6$ .

**Table S13.**  $^1\text{H}$  NMR (400 MHz) data of **G**, **I** and **J** in DMSO- $d_6$ .

**Table S14.** Conserved sequence regions in the alignment comparisons of BsmF and other known P450s.

### Supplementary figures:

**Figure S1.** Structures of bosamycins.

**Figure S2.** LC-MS analysis of L-FDAA and D-FDAA derivatives of the amino acid residues in **1** and **3**.

**Figure S3.** Identification of compound **5** as a novel inhibitor of SHP2.

**Figure S4.** Generation of the *S. sp.* 120454 mutant strains.

**Figure S5.** MS/MS analysis of metabolite extracts from wild type and mutant strains.

**Figure S6.** Chemical complementation of 5-OMe Tyr into mutants.

**Figure S7.** Chemical complementation of 5-OH Tyr into mutants.

**Figure S8.** HPLC analysis of metabolite extracts from mutant strains and gene complementation strains.

**Figure S9.** Chemical complementation fed L-erythro- $\beta$ -OH-Asp into  $\Delta bsmC$  mutant.

**Figure S10.** Verification of cosmid pHG06015.

**Figure S11.** SDS-PAGE analysis of proteins.

**Figure S12.** Sequence alignment of BsmF-P450 with other P450 proteins.

**Figure S13.** Domain analyze of BsmF.

**Figure S14.** Sequence alignment of BsmC with other  $\beta$ -hydroxylases.

**Figure S15.** Sequence alignment of BsmB-T<sub>2</sub> with other T domains.

**Figure S16.** Sequence alignment of BsmA-C<sub>1</sub> with other C domains.

**Figure S17-22.** MS/MS analysis of **1-6**.

**Figure S23-25.** MS/MS analysis of **7-9** in  $\Delta bsmF$  mutant strain.

**Figure S26-28.** MS/MS analysis of **7-9** in  $\Delta bsmG$  mutant strain.

**Figure S29-31.** MS/MS analysis of **7-9** in  $\Delta bsmH$  mutant strain.

**Figure S32-34.** MS/MS analysis of **10-12** in  $\Delta bsmC$  mutant strain.

**Figure S35-36.** MS/MS analysis of **13-14**.

**Figure S37.** HRESIMS spectrum of **1**.

**Figure S38-44.** 1D and 2D NMR spectrum of compound **1** in DMSO-*d*<sub>6</sub>.

**Figure S45.** HRESIMS spectrum of **2**.

**Figure S46-52.** 1D and 2D NMR spectrum of compound **2** in DMSO-*d*<sub>6</sub>.

**Figure S53.** HRESIMS spectrum of **3**.

**Figure S54-62.** 1D and 2D NMR spectrum of compound **3** in DMSO-*d*<sub>6</sub>.

**Figure S63.** HRESIMS spectrum of **4**.

**Figure S64-70.** 1D and 2D NMR spectrum of compound **4** in DMSO-*d*<sub>6</sub>.

**Figure S71.** HRESIMS spectrum of **5**.

**Figure S72-78.** 1D and 2D NMR spectrum of compound **5** in DMSO-*d*<sub>6</sub>.

**Figure S79.** HRESIMS spectrum of **6**.

**Figure S80-86.** 1D and 2D NMR spectrum of compound **6** in DMSO-*d*<sub>6</sub>.

**Figure S87.** HRESIMS spectrum of **13**.

**Figure S88-94.** 1D and 2D NMR spectrum of compound **13** in DMSO-*d*<sub>6</sub>.

**Figure S95.** HRESIMS spectrum of **14**.

**Figure S96-102.** 1D and 2D NMR spectrum of compound **14** in DMSO-*d*<sub>6</sub>.

**Figure S103.** HRESIMS spectrum of **A**.

**Figure S104.** <sup>1</sup>H NMR (400 MHz) spectrum of compound **A** in DMSO-*d*<sub>6</sub>.

**Figure S105.** HRESIMS spectrum of **B**.

**Figure S106.** <sup>1</sup>H NMR (400 MHz) spectrum of compound **B** in DMSO-*d*<sub>6</sub>.

**Figure S107.** HRESIMS spectrum of **C**.

**Figure S108.** <sup>1</sup>H NMR (400 MHz) spectrum of compound **C** in DMSO-*d*<sub>6</sub>.

**Figure S109.** HRESIMS spectrum of **D**.

**Figure S110.** <sup>1</sup>H NMR (400 MHz) spectrum of compound **D** in DMSO-*d*<sub>6</sub>.

**Figure S111.** HRESIMS spectrum of **E**.

**Figure S112.** <sup>1</sup>H NMR (400 MHz) spectrum of compound **E** in DMSO-*d*<sub>6</sub>.

**Figure S113.** HRESIMS spectrum of **G**.

**Figure S114-115.** <sup>1</sup>H NMR (400 MHz) and <sup>13</sup>C (100 MHz) spectrum of compound **G** in DMSO-*d*<sub>6</sub>.

**Figure S116.** HRESIMS spectrum of **I**.

**Figure S117-118.** <sup>1</sup>H NMR (400 MHz) and <sup>13</sup>C (100 MHz) spectrum of compound **I** in DMSO-*d*<sub>6</sub>.

**Figure S119.** HRESIMS spectrum of **J**.

**Figure S120-121.**  $^1\text{H}$  NMR (400 MHz) and  $^{13}\text{C}$  (100 MHz) spectrum of compound **J** in  $\text{DMSO-}d_6$ .

**References**

## Experimental Procedures

### Heterologous expression of cosmid pHG06015 in *Streptomyces albus* J1074.

Cosmid library for strain 120454 was constructed according to the protocol.<sup>2</sup> The cosmid pHG06015 that covers partial *bsm* gene cluster was obtained through PCR screening from 2000 clones using primers listed in Table 3. The cosmid pHG06015 was transferred into the *E. coli*/ET12567 (pUZ8002) strain, and then introduced into *Streptomyces albus* J1074 by intergrative conjugation. The recombinant strain was grown on MS agar medium supplemented with 50 µg/mL of apramycin for sporulation. The seed culture was prepared by inoculating fresh spores into 250-mL baffled flasks containing 50 mL of TSB medium (17.0 g tryptone, 3.0 g soytone, 2.5 g glucose, 5.0 g sodium chloride, 2.5 g Na<sub>2</sub>HPO<sub>4</sub> in 1 L water, pH 7.0) for 1 days at 30 °C and 250 rpm. Subsequently, 15 mL seed cultures were inoculated into 2 L baffled flasks containing 300 mL of the fermentation medium (dextrin 40 g, tomato paste 7.5 g, NZ Amine 2.5 g, primary yeast 5 g in 1 L distilled water, pH 7.0), and incubated for 7 days at 160 rpm and 30 °C. Finally, the fermentation broth was filtered and absorbed with XAD-16 resin. The resin was washed with water and eluted with acetone.

### Co-expression of *bsmF*, *bsmG* and *bsmH* in *Streptomyces albus* J1074 /pHG06015 strain.

A DNA fragment containing three genes of *bsmF*, *bsmG* and *bsmH* was amplified from genomic DNA and subcloned into *E. coli*-*Streptomyces* expression shuttle vector pUWL201PWT plasmid to afford pHG06016 plasmid. The plasmid pHG06016 was transformed into *E. coli* ET12567/pUZ8002 and then introduced into *S. albus*/pHG06015 strain by conjugation. Clones harboring pHG06016 plasmid were selected by thiostrepton resistance and verified by diagnostic PCR. The resulting recombinant strain was then fermented for 7 days at 30 °C. The crude extract was analyzed by LC-MS and HPLC.

### Isolation and purification of bosamycins from *S. sp.* 120454 wild-type and recombinant strains.

For isolation of compounds, a large scale fermentation (20 L) for wild-type strain was carried out using the same medium as mention above. The resin was harvested after seven days' cultivation, and extracted with methanol for three times. The combined methanol phases were evaporated to dryness, and the resulting extract was subjected to silica-gel column, and eluted with the mixture of methylene dichloride and methanol (100:1 to 1:1). Fractions were combined according to HPLC analysis, and further separated by Sephadex LH-20 chromatography. Fractions containing the target compounds were finally purified by semi-preparative HPLC. Compounds **1-6** were purified from *S. sp.* 120454 wild-type strain. Compound **13** were purified from HG06012 strain. Compound **14** were purified from HG06013 strain.

### Physical data for bosamycins.

Compound **1**: white amorphous solid;  $[\alpha]_D^{25}$  -26.0 (c 0.10, MeOH); UV (MeOH)  $\lambda_{max}$  (log  $\epsilon$ ) 218 (3.84), 228 (3.83), 278 (3.31); NMR data see Table S4; HRESIMS  $m/z$  876.4166 [M+ H]<sup>+</sup> (calcd for C<sub>40</sub>H<sub>58</sub>N<sub>7</sub>O<sub>15</sub>, 876.3913).

Compound **2**: white amorphous solid;  $[\alpha]_D^{25}$  +30.8 (c 0.01, MeOH); UV (MeOH)  $\lambda_{max}$  (log  $\epsilon$ ) 212 (3.61), 226 (3.57), 279 (3.08); NMR data see Table S5; HRESIMS  $m/z$  1082.4697 [M+ H]<sup>+</sup> (calcd for C<sub>50</sub>H<sub>68</sub>N<sub>9</sub>O<sub>18</sub>, 1082.4677).

Compound **3**: white amorphous solid;  $[\alpha]_D^{25}$  -20.0 (c 0.05, MeOH); UV (MeOH)  $\lambda_{max}$  (log  $\epsilon$ ) 218 (4.23), 228 (4.27), 281 (3.58); NMR data see Table S6; HRESIMS  $m/z$  1126.4572 [M+ H]<sup>+</sup> (calcd for C<sub>51</sub>H<sub>68</sub>N<sub>9</sub>O<sub>20</sub>, 1126.4575).

Compound **4**: white amorphous solid;  $[\alpha]_D^{25}$  +8.7 (c 0.01, MeOH); UV (MeOH)  $\lambda_{max}$  (log  $\epsilon$ ) 218 (4.01), 229 (3.12), 280 (3.49); NMR data see Table S7; HRESIMS  $m/z$  1081.4725 [M+ H]<sup>+</sup> (calcd for C<sub>51</sub>H<sub>69</sub>N<sub>8</sub>O<sub>18</sub>, 1081.4724).

Compound **5**: white amorphous solid;  $[\alpha]_D^{25}$  -6.7 (c 0.14, MeOH); UV (MeOH)  $\lambda_{max}$  (log  $\epsilon$ ) 218 (3.88), 229 (3.98), 278 (3.48); NMR data see Table S8; HRESIMS  $m/z$  1140.4743 [M+ H]<sup>+</sup> (calcd for C<sub>52</sub>H<sub>70</sub>N<sub>9</sub>O<sub>20</sub>, 1140.4732).

Compound **6**: white amorphous solid;  $[\alpha]_D^{25} +8.7$  (c 0.02, MeOH); UV (MeOH)  $\lambda_{\max}$  (log  $\epsilon$ ) 218 (4.07), 227 (4.08), 280 (3.64); NMR data see Table S9; HRESIMS  $m/z$  1095.4879  $[M+H]^+$  (calcd for  $C_{52}H_{71}N_8O_{18}$ , 1095.4881).

Compound **13**: white amorphous solid;  $[\alpha]_D^{25} +8.0$  (c 0.03, MeOH); UV (MeOH)  $\lambda_{\max}$  (log  $\epsilon$ ) 218 (3.86), 229 (3.97), 281 (3.40); NMR data see Table S10; HRESIMS  $m/z$  1009.4564  $[M+H]^+$  (calcd for  $C_{48}H_{65}N_8O_{16}$ , 1009.4513).

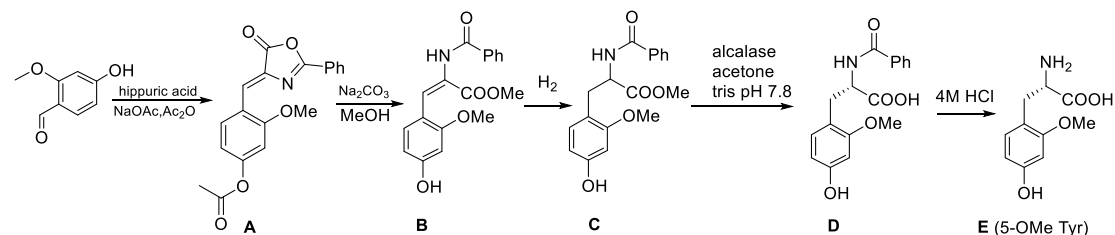
Compound **14**: white amorphous solid;  $[\alpha]_D^{25} +5.3$  (c 0.04, MeOH); UV (MeOH)  $\lambda_{\max}$  (log  $\epsilon$ ) 218 (3.65), 227 (3.67), 278 (3.26); NMR data see Table S11; HRESIMS  $m/z$  1039.4631  $[M+H]^+$  (calcd for  $C_{49}H_{67}N_8O_{17}$ , 1039.4619).

#### Preparation and analysis of Marfey's derivatives.

The absolute configuration of the bosamycins was determined by advanced Marfey's method.<sup>3</sup> Briefly, Compound **3** (0.5 mg) was dissolved in 6 N HCl (1 mL) and hydrolyzed at 110 °C for 12 h. After cooling, the solution was evaporated to dryness and dissolved in H<sub>2</sub>O (100  $\mu$ L). To this mixture was added a 1% (w/v) solution (200  $\mu$ L) of 1-fluoro-2,4-dinitrophenyl-5-L-alanine amide (L-FDAA) and 1-fluoro-2,4-dinitrophenyl-5-D-alanine amide (D-FDAA) in acetone respectively. After adding NaHCO<sub>3</sub> solution (1 M, 50  $\mu$ L), the reaction mixture was heated at 45 °C for 1 h and then acidified with 2 N HCl (25  $\mu$ L). The standards L-Leu, L-OMe Tyr, L-Ser, *erythro* L-OHAsp, *threo* L-OHAsp and L-Tyr were derivatized in a similar manner. Derivatized hydrolysate (20  $\mu$ L) and standard amino acids (20  $\mu$ L) were subjected to LC-MS analysis.

Because compound **3** has two tyrosines that have different configurations, whereas compound **1** only have one tyrosine that can be assigned unambiguously by Marfey's method. Thus, **1** was hydrolyzed and derivatized in a similar manner. Results can be found in Figure S2.

#### Chemical synthesis of 5-OMe Tyr.<sup>4-7</sup>



Compound **A**: 4-Hydroxy-2-methoxybenzaldehyd (1 g, 6.6 mmol), hippuric acid (1.2 g, 6.7 mmol), sodium acetate (0.56 g, 7 mmol) and acetic anhydride (1.5 ml, 1.6 g, 16 mmol) were added to a 100-ml round-bottom flask, heated in an oil bath at 100 °C for 2 h. The resulting solid mixture was cooled to room temperature before H<sub>2</sub>O (10 ml) was added. The mixture was filtrated, washed with aqueous Na<sub>2</sub>CO<sub>3</sub> and dried under vacuum. The desired compound **A** was then obtained after recrystallization from acetone/H<sub>2</sub>O (2:1) as a yellow solid; yield: 1.2 g (53%). <sup>1</sup>H NMR data see Table S13. HRESIMS  $m/z$ : 338.0992  $[M+H]^+$  (calcd. for  $[C_{19}H_{15}NO_5H]^+$  338.0878).

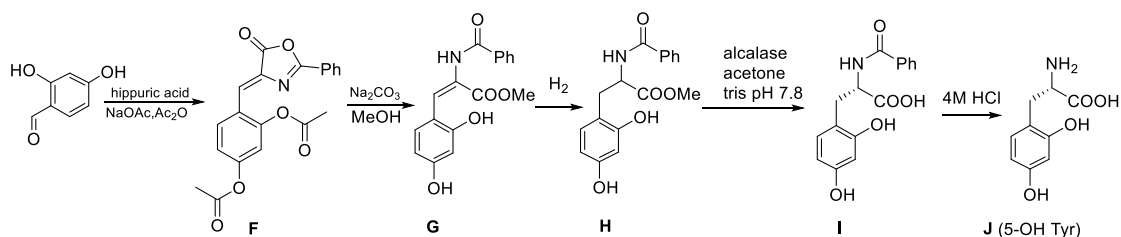
Compound **C**: To a 100-ml round bottom flask was added compound **A** (1 g, 3.38 mmol), 1:1 mixture of CH<sub>2</sub>Cl<sub>2</sub> and MeOH (12 ml) and Na<sub>2</sub>CO<sub>3</sub> (283 mg, 1.41 mmol). The mixture was stirred at room temperature for 12 hours, then filtered and concentrated to afford compound **B** (850mg, 85%). <sup>1</sup>H NMR data see Table S13. HRESIMS  $m/z$ : 350.1016  $[M+Na]^+$  (calcd. for  $[C_{18}H_{17}NO_5Na]^+$  350.1107). Compound **B** was dissolved in a solution of Pd/C (70 mg) in MeOH (15 mL) and acetic acid (2.5 mL). The reaction suspension was then hydrogenated (50 bar H<sub>2</sub>) for 24 h. The mixture was filtered and concentrated to give compound **C** as a white crystalline solid (723 mg, 72%). <sup>1</sup>H NMR data see Table S13. HRESIMS  $m/z$ : 330.0549  $[M+H]^+$  (calcd. for  $[C_{18}H_{19}NO_5H]^+$  330.1297).

Compound **D**: Compound **C** (500 mg, 1.5 mmol) was dissolved in DMSO (1.5 mL) and diluted with acetone (12 mL) then diluted with Tris buffer (80 mM) pH 7.8. The mixture was warmed to 37 °C, and Alcalase 1ml (3 mL, > 2.4 U / mL, Sigma) from *Bacillus licheniformis* was added to the reaction system. The reaction was periodically adjusted to pH 7.8 by the addition of 1 M NaOH until conversion was stopped by HPLC (2 days). The volatiles were removed

in vacuo. The organic phase was acidified with HCl to pH = 2, and then extracted with EtOAc. The organic extract was concentrated, and the resulting solid was recrystallized by diluting in MeOH and EtOAc (~1:1) to give compound **D** as a white needle (323 mg, 67%). <sup>1</sup>H NMR data see Table S13. HRESIMS *m/z*: 338.1012 [M+Na]<sup>+</sup> (calcd. for [C<sub>17</sub>H<sub>17</sub>NO<sub>5</sub>Na]<sup>+</sup> 338.1140).

Compound **E** (5-OMe Tyr): Compound **D** (200 mg, 0.63 mmol) was hydrolyzed with 4N HCl. The mixture was cooled, washed with EtOAc, and concentrated to give a crude brown compound **E** (75 mg, 56%). <sup>1</sup>H NMR data see Table S13. HRESIMS *m/z*: 212.0940 [M+H]<sup>+</sup> (calcd. for [C<sub>10</sub>H<sub>12</sub>NO<sub>4</sub>H]<sup>+</sup> 212.0878).

#### Chemical synthesis of 5-OH Tyr.<sup>4-7</sup>



5-OH Tyr (30 mg) was prepared using the method as mentioned in 5-OMe Tyr. NMR data see Table S13, Figures S120 and S121. HRESIMS *m/z*: 198.0873 [M+H]<sup>+</sup> (calcd. for [C<sub>19</sub>H<sub>15</sub>NO<sub>5</sub>H]<sup>+</sup> 198.0688).

#### Chemical complementation of 5-OMe Tyr or 5-OH Tyr to $\Delta bsmF$ , $\Delta bsmG$ and $\Delta bsmH$ mutant.

$\Delta bsmF$ ,  $\Delta bsmG$  and  $\Delta bsmH$  mutant was individually cultured in a 50 ml flasks containing 20 mL B medium. After 24 hours cultivation, 5-OMe Tyr (1.3 mg) or 5-OH Tyr (1.3 mg) dissolved in DMSO was supplemented into fermentation broth and cultured for another 6 days. After extraction by XAD-16, the eluted organic solution was then analyzed by LC-MS. The LC-MS analysis was performed using a 18 min solvent gradient from 10 % to 90 % methanol in water supplied with 0.1 TFA at a flow rate of 0.4 mL/min.

#### Chemical complementation of L-erythro- $\beta$ -OH-Asp into $\Delta bsmC$ mutant.

$\Delta bsmC$  mutant was individually cultured in a 50 ml flasks containing 20 mL B medium. After 24 hours cultivation, L-erythro- $\beta$ -OH-Asp (1.2 mg) dissolved in DMSO was supplemented into fermentation broth and cultured for another 6 days. After extraction by XAD-16, the eluted organic solution was then analyzed by LC-MS. The LC-MS analysis was performed using a 18 min solvent gradient from 10 % to 90 % methanol in water supplied with 0.1 TFA at a flow rate of 0.4 mL/min.

#### Protein expression and purification.

DNA fragments containing target genes including BsmA (A<sub>1</sub>), BsmB (A<sub>4</sub>-T<sub>4</sub>), BsmD (C<sub>6</sub>-A<sub>6</sub>-T<sub>6</sub>), BsmF (A<sub>0</sub>-T<sub>0</sub>) and BsmH were individually amplified from genomic DNA of *S. sp.* 120454 with primers listed in Table S3. The purified PCR product of BsmA (A<sub>1</sub>), BsmB (A<sub>4</sub>-T<sub>4</sub>) and BsmF (A<sub>0</sub>-T<sub>0</sub>) were ligated with linearized pET28a (linearized by NdeI and HindIII) to afford pHG06017, pHG06018 and pHG06020, DNA fragments containing BsmD (C<sub>6</sub>-A<sub>6</sub>-T<sub>6</sub>) and BsmH were ligated with linearized pET22b (treated with NdeI and HindIII) to afford pHG06019 and pHG06021. The obtained plasmids were transformed into *E. coli* BL21(DE3), respectively. A single colony was picked to inoculate a 4 mL LB starter culture grown overnight at 37 °C, 200 rpm. The following day, 0.4 L LB media supplemented with kanamycin or ampicillin was inoculated with the starter culture and incubated at 37 °C, 200 rpm until the OD<sub>600</sub> reached 0.6. The culture was cooled to 4 °C and induced with 0.125 mM IPTG. Cultures were incubated at 16 °C, 200 rpm for 18 h. Cell pellets were resuspended in lysis buffer (100 mM Tris, pH 8.0, 15 mM imidazole, 300 mM

NaCl, 10 % glycerol) and sonicated on ice. After centrifugation at 15000 rpm for 30 min, the supernatant was filtered and loaded onto a 5 mL Histrap HP column (GE lifesciences). Fractions containing the proteins were pooled and desalted by a PD10 column (GE Healthcare) with 100 mM Tris-HCl buffer (pH 7.5) and 10% glycerol and stored at -80°C.

#### **Adenylation activities of A domain.**

The A domain specificity assays were conducted in a 50 µL reaction volume containing 100 mM Tris-HCl, pH 7.5, 20 µM NRPS protein, 12.5 mM MgCl<sub>2</sub>, 2.0 mM TCEP, 2 mM amino acid, 4 mM ATP. After reaction at room temperature for 1 hour, an equal volume of the Master Reaction Mix (Sigma-Aldrich Pyrophosphate Assay Kit MAK 168) were added to each of the sample, and incubated for another 30 minutes. Then, the fluorescence intensity ( $\lambda_{ex}=316$  /  $\lambda_{em}=456$  nm) was measured by a microplate reader (TECAN infinite M200PRO).

#### ***In vitro* assay of BsmH.**

Enzymatic reaction was performed in 100 mM phosphate buffer (pH 7.2) containing 10 µM BsmH, 0.5 mM 5-OH Tyr, 1 mM SAM. After incubation at 30 °C for 2 h, 50 µL acetonitrile were added to quench the reaction. Then the mixture was centrifuged at 14,000 g for 10 min and the supernatant was analyzed by analytic HPLC using a 20 min solvent gradient from 5% to 20% acetonitrile in water supplied with 0.1 TFA at a flow rate of 0.5 mL/min.

#### **Biological activity assay of SHP2.**

The catalytic activity of SHP2 was monitored using the surrogate substrate DiFMUP in a prompt fluorescence assay format.<sup>8</sup> The phosphatase reactions were performed at room temperature in 96-well black polystyrene plate, flat bottom, low flange, nonbinding surface (Corning, cat. no. 3575) using a final reaction volume of 100 µl and the following assay buffer conditions: 60 mM HEPES, pH 7.2, 75 mM NaCl, 75 mM KCl, 1 mM EDTA, 0.05% P-20, 5 mM DTT. 1 nM of SHP2<sup>WT</sup>(residues 1-525) was co-incubated with of 1 µM of bisphosphorylated IRS1 peptide (sequence:H2N-LN(pY)IDLDLV(dPEG8)LST(pY)ASINFQK-amide) and 30 µM of tested compounds. Under the same buffer conditions, the phosphatase assays of SHP2<sup>E76K</sup> (0.5 nM) or SHP2<sup>PTP</sup> (1 nM) was incubated with Compound **5** at various concentrations. After 30-60 min incubation at 25 °C, the surrogate substrate DiFMUP (Invitrogen, cat. no. D6567, 100 µM) was added to the reaction and incubated at 25 °C for 30 min. The reaction was then quenched by the addition of 20 µl of a 160 µM solution of bpV (Phen) (Enzo Life Sciences cat. no. ALX-270-204). The fluorescence signal was monitored using a microplate reader (TECAN, M200PRO) using excitation and emission wavelengths of 340 and 450 nm, respectively.

Compounds **1-6**, **13-14** were screened at 30 µM against Src homology 2-containing protein tyrosine phosphatase 2 (SHP2), which is a major phosphatase involved in growth factor and cytokine-mediated signaling. Studies have shown that SHP2 allosteric inhibitors have shown remarkable anti-tumor benefits.<sup>8-11</sup> In our initial single-concentration assays, only compound **5** had the inhibitory effect on the SHP2 (Figure S3, A). To further evaluate the acting mechanisms of compound **5** we used three SHP2 proteins to test sensitivities of **5**: (1) wild-type (WT) SHP2 (residues 1–525); (2) SHP2<sup>E76K</sup> mutant with a partially open conformation in SHP2; (3) SHP2 PTP domain with a completely open conformation. Compound **5** was shown to inhibit SHP2 enzyme activity in a dose-dependent manner, and the IC<sub>50</sub> value of SHP2<sup>WT</sup>, SHP2<sup>E76K</sup>, SHP2<sup>PTP</sup> were 24.25, 45.56, and 89.98 µM, respectively (Figure S3, B and C), suggesting **5** could be a novel allosteric inhibitor of SHP2.



**Table S1.** Bacterial plasmids and strains.

Plasmid/Strain	Relevant characteristics	Source
<b>Plasmid</b>		
pKC1139	<i>E.coli-Streptomyces</i> shuttle plasmid used for gene disruption, temperature sensitive	12
PJTU2554	Cosmid vector for genomic library construction	13
pSET152-kasOp*	pSET152 derived plasmid containing the promoter <i>kasOp</i> *	14
pUWL201PWT	<i>E. coli-Streptomyces</i> expression shuttle vector harboring <i>oriT</i> (cloned into the <i>PstI</i> site)	15
pET28a	Protein expression vector used in <i>E.coli</i> , encoding N-terminal His-tag, kanamycin resistance	Novagen
pET22b	Protein expression vector used in <i>E.coli</i> , encoding C-terminal His-tag, ampicillin resistance	Novagen
pHG06001	pKC1139 derived plasmid for disruption of <i>bsmA-C<sub>1</sub></i>	This study
pHG06002	pKC1139 derived plasmid for disruption of <i>bsmC</i>	This study
pHG06003	pKC1139 derived plasmid for disruption of <i>bsmD</i>	This study
pHG06004	pKC1139 derived plasmid for disruption of <i>bsmF</i>	This study
pHG06005	pKC1139 derived plasmid for disruption of <i>bsmG</i>	This study
pHG06006	pKC1139 derived plasmid for disruption of <i>bsmH</i>	This study
pHG06007	pKC1139 derived plasmid for disruption of <i>orf(-1)</i>	This study
pHG06008	pKC1139 derived plasmid for disruption of <i>bsmI</i>	This study
pHG06009	pSET152-kasOp* derived plasmid for complementation of <i>bsmC</i>	This study
pHG06010	pSET152-kasOp* derived plasmid for complementation of <i>bsmF</i>	This study
pHG06011	pSET152-kasOp* derived plasmid for complementation of <i>bsmG</i>	This study
pHG06012	pSET152-kasOp* derived plasmid for complementation of <i>bsmF-T281A</i>	This study
pHG06013	pSET152-kasOp* derived plasmid for complementation of <i>bsmF-F380A</i>	This study
pHG06014	pSET152-kasOp* derived plasmid for complementation of <i>bsmF-C387A</i>	This study
pHG06015	Cosmid which contains <i>bsm</i> biosynthetic gene cluster	This study
pHG06016	pUWL201PWT derived plasmid harboring <i>bsmF</i> , <i>bsmG</i> , <i>bsmH</i>	This study
pHG06017	pET28a derived plasmid for expressing N-terminal His-tag BsmA (A <sub>1</sub> )	This study
pHG06018	pET28a derived plasmid for expressing N-terminal His-tag BsmB (A <sub>4</sub> -T <sub>4</sub> )	This study
pHG06019	pET22b derived plasmid for expressing C-terminal His-tag BsmD (C <sub>6</sub> -A <sub>6</sub> -T <sub>6</sub> )	This study
pHG06020	pET28a derived plasmid for expressing N-terminal His-tag BsmF (A <sub>0</sub> -T <sub>0</sub> )	This study
pHG06021	pET22b derived plasmid for expressing C-terminal His-tag BsmH	This study
<b><i>E. coli</i> strains</b>		
DH5α	General cloning host	16
BL21 (DE3)	Heterologous host for protein expression	NEB

ET12567 (pUZ8002)	Methylation-deficient host used for <i>E. coli-Streptomyces</i> intergeneric conjugation	3
<b>Strains</b>		
<i>S. albus</i> J1074	Model actinomycete used for gene heterologous expression	17
120454	Wild type strain for bosamycins production	This study
HG06001	$\Delta bsmA-C_1$ , in-frame deletion mutant strain in WT, bosamycin D producing	This study
HG06002	$\Delta bsmC$ , in-frame deletion mutant strain in WT, bosamycins non-producing	This study
HG06003	$\Delta bsmD$ , in-frame deletion mutant strain in WT, bosamycins non-producing	This study
HG06004	$\Delta bsmF$ , in-frame deletion mutant strain in WT, bosamycins non-producing	This study
HG06005	$\Delta bsmG$ , in-frame deletion mutant strain in WT, bosamycins non-producing	This study
HG06006	$\Delta bsmH$ , in-frame deletion mutant strain in WT, bosamycins non-producing	This study
HG06007	$\Delta orf(-1)$ , in-frame deletion mutant strain in WT, bosamycins producing	This study
HG06008	$\Delta bsmI$ , in-frame deletion mutant strain in WT, bosamycin D producing	This study
HG06009	complementation of $\Delta bsmC$ mutant by <i>bsmC</i> , bosamycins producing	This study
HG06010	complementation of $\Delta bsmF$ mutant by <i>bsmF</i> , bosamycins producing	This study
HG06011	complementation of $\Delta bsmG$ mutant by <i>bsmG</i> , bosamycins producing	This study
HG06012	<i>Streptomyces albus</i> J1074 integrated with plasmid pJTU2554	This study
HG06013	<i>Streptomyces albus</i> J1074 integrated with plasmid pHG06015 which contains <i>bsm</i> biosynthetic gene cluster	This study
HG06014	HG06013 containing plasmid pHG06016	This study

**Table S2.** Oligonucleotide primers used in this study.

Oligonucleotide	Sequence <sup>a</sup>	Enzyme sites
<b>a. for amplification of homologous arms from genomic DNA for gene disruption (5'-3')</b>		
<i>orf(-1)</i> -up-F	<b>AACGACGGCCAGTGCCAAGCTT</b> GCTTCAAGAGCACGTCGGATAC	<i>Hind</i> III
<i>orf(-1)</i> -up-R	ATCCCGGCACTACCAACGACCGCACAGCGTCGAAG	
<i>orf(-1)</i> -down-F	CTTCGACGCTGTGCGGTCGTTGGTGAGTGCCGGGAT	
<i>orf(-1)</i> -down-R	<b>AGCTATGACATGATTACGAATTC</b> CTGCTCACCATCCACCATCT	<i>Eco</i> RI
<i>bsmI</i> -up-F	<b>AACGACGGCCAGTGCCAAGCTT</b> GTCGTCCACTCCAGCAATAG	<i>Hind</i> III
<i>bsmI</i> -up-R	GGCATGCAGCTGATCACCCACTGGCCGACGAGATC	
<i>bsmI</i> -down -F	GATCTCGTCGGCCAGTGGGGTGATCAGCTGCATGCC	
<i>bsmI</i> -down-R	<b>AGCTATGACATGATTACGAATTC</b> GCACAAGACACCCAGACAAC	<i>Eco</i> RI
<i>bsmA-C<sub>1</sub></i> -up-F	<b>AACGACGGCCAGTGCCAAGCTT</b> GGACCATCAGGCACGACATAAC	<i>Hind</i> III
<i>bsmA-C<sub>1</sub></i> -up-R	CAATGGCCTTTGACCGCACCGCTGCACATCACGGTG	
<i>bsmA-C<sub>1</sub></i> -down-F	CACCGTGATGTGCAGCGGTGCGGTCAAAGGCCATTG	
<i>bsmA-C<sub>1</sub></i> -down -R	<b>AGCTATGACATGATTACGAATTC</b> GAACCAGACGATCAGCAAGAAC	<i>Eco</i> RI
<i>bsmC</i> -up-F	<b>AACGACGGCCAGTGCCAAGCTT</b> GTTGGCGGTGCTCAAG	<i>Hind</i> III
<i>bsmC</i> -up-R	TTGTGATCAGCAGCAGATGTACTCGGTGAGGTGTAGAT	
<i>bsmC</i> -down-F	ATCTACACCTCGACCGAGTACATCTGCTGCTGATCGACAA	
<i>bsmC</i> -down -R	<b>AGCTATGACATGATTACGAATTC</b> ATGTGCCGTCTGTTC	<i>Eco</i> RI
<i>bsmD</i> -up-F	<b>AACGACGGCCAGTGCCAAGCTT</b> AAGACCCTACCACCCTCTT	
<i>bsmD</i> -up-R	<b>TTGCTCGGCCTGGACACGCAGTTCGGGTAGCGCGC</b>	
<i>bsmD</i> -down-F	<b>GCGCGTACCCGGAAGTGC</b> GTGTCAGGCCGAGCAA	
<i>bsmD</i> -down -R	<b>AGCTATGACATGATTACGAATTC</b> CGATACGGAAACCACGCAACT	
<i>bsmF</i> -up-F	<b>AACGACGGCCAGTGCCAAGCTT</b> GTATCGAACTGGGAGAAGTG	<i>Hind</i> III
<i>bsmF</i> -up-R	CATGGCCAGCGAGTACCTTTCTCCTTGGGTGTCAT	
<i>bsmF</i> -down -F	ATGACACCCAAGGAGAAAGGTGACTCGCTGGCCATG	
<i>bsmF</i> -down -R	<b>AGCTATGACATGATTACGAATTC</b> CCCTTCCTCGTACTTCTTGTC	<i>Eco</i> RI
<i>bsmG</i> -up-F	<b>AACGACGGCCAGTGCCAAGCTT</b> GCTGGCAGTTGGTCTTC	<i>Hind</i> III
<i>bsmG</i> -up-R	ATCTCTGTACCAATTGGGCCTGTATCGCTGTGTCGTTCA	
<i>bsmG</i> -down -F	TGAACGACACAGCGATACAGGCCCAATTGGTACGAGAGAT	
<i>bsmG</i> -down -R	<b>AGCTATGACATGATTACGAATTC</b> CCACTCCATACGATGACCTTAC	<i>Eco</i> RI
<i>bsmH</i> -up-F	<b>AACGACGGCCAGTGCCAAGCTT</b> GATCAGTGCCGCGATCAT	<i>Hind</i> III
<i>bsmH</i> -up-R	CTACTTCTGTGCCCGATGAGGTGGAATGAGTCCGTAGAC	
<i>bsmH</i> -down -F	GTCTACGGACTCATTTCCACCTCATCGGGCACAAGAAGTAG	
<i>bsmH</i> -down -R	<b>AGCTATGACATGATTACGAATTC</b> CCGTGGAAGTAGCCGAAG	<i>Eco</i> RI
<b>b. for screening of the correct mutants (5'-3')</b>		
<i>orf(-1)</i> - Diag-F	CAGCGGATCATGCTATCT	
<i>orf(-1)</i> - Diag-R	ACAGGTGGCGTTCAACTATT	
<i>bsmI</i> - Diag-F	AGTCTGTGTCGGCCATTCTA	
<i>bsmI</i> - Diag-R	CACAGGTGGCGTTCAACTATT	
<i>bsmA-C<sub>1</sub></i> - Diag-F	GACCTGCTTTGCGAGTTTAC	
<i>bsmA-C<sub>1</sub></i> - Diag-R	CGGAGAACCGATTACCTCTATG	

<i>bsmC</i> - Diag-F	CACAGGGAACAGAGGAGAATG	
<i>bsmC</i> - Diag-R	GCAAGGTCGAACGAGTACC	
<i>bsmD</i> - Diag-F	TCCCGAACCCACTGATGA	
<i>bsmD</i> - Diag-R	CAACTCACCGTACGAGAACAC	
<i>bsmF</i> - Diag-F	TGCGCTACATCGAAGAGAAC	
<i>bsmF</i> - Diag-R	CGAAGAGTTCGACGAGCAG	
<i>bsmG</i> - Diag-F	AGCTGGAGACCGACTTCTT	
<i>bsmG</i> - Diag-R	GTGTCCGGCTTGCCTTC	
<i>bsmH</i> - Diag-F	CTTCGACCCGATGACGTTT	
<i>bsmH</i> - Diag-R	GATGCCACATCCGGACAG	
<b>c. for genes complementation (5'-3')</b>		
152- <i>bsmC</i> -F	<b>TGCTGCATGCATACGTACTAGT</b> CTCAAGCGCCCGGAAAGG	<i>SpeI</i>
152- <i>bsmC</i> -R	<b>CTATGACATGATTACGAATTC</b> ACCCCGGACATGGCGAC	<i>EcoRI</i>
152- <i>bsmF</i> -F	<b>TGCTGCATGCATACGTACTAGT</b> CCGTTTCGAGGCGTACCGC	<i>SpeI</i>
152- <i>bsmF</i> -R	<b>CTATGACATGATTACGAATTC</b> ACTGTGCCTCTCGCCC	<i>EcoRI</i>
152- <i>bsmG</i> -F	<b>TGCTGCATGCATACGTACTAGT</b> GGTGACTCGCTGGCCATG	<i>SpeI</i>
152- <i>bsmG</i> -R	<b>CTATGACATGATTACGAATTC</b> ATGGACGTTGCTCCTC	<i>EcoRI</i>
152- <i>bsmF</i> -T281A-F	CTGACGCACCTCGTCTCCACC	
152- <i>bsmF</i> -T281A-R	GCGCTGGTAACCCCTGCGACGAG	
152- <i>bsmF</i> -F380A-F	GGTTCGGTCCGCACTACTGC	
152- <i>bsmF</i> -F380A-R	GCGCCGAGGTGGCGGCCGTC	
152- <i>bsmF</i> -C387A-F	CTGGGCGCCCGTTTCGCC	
152- <i>bsmF</i> - C387A-R	GCGGTAGTGCGGACCGGAACCGAA	
201- <i>bsmFGH</i> -F	<b>AAAGAGGAGAAATTACATATG</b> ATGACACCCAAGGAGAAA	<i>NdeI</i>
201- <i>bsmFGH</i> -R	<b>CAGGAATTCGATATCAAGCTT</b> AACTCAGCGCGTGATA	<i>HindIII</i>
<b>d. for protein expression</b>		
BsmA (A <sub>1</sub> )-28a-F	<b>GTGCCGCGCGGCAGCCATATG</b> CTGTTTCGAGGCGCGGGTT	<i>NdeI</i>
BsmA (A <sub>1</sub> )-28a-R	<b>CTCGAGTGCAGCCGCAAGCTT</b> TTCAGCGTAGTTTACCTGGTC	<i>HindIII</i>
BsmB (A <sub>4</sub> -T <sub>4</sub> )-28a-F	<b>GTGCCGCGCGGCAGCCATATG</b> CTGGCGAGCCTCCTCGAT	<i>NdeI</i>
BsmB (A <sub>4</sub> -T <sub>4</sub> )-28a-R	<b>CTCGAGTGCAGCCGCAAGCTT</b> TTCACGACAGCGGCATCCGCTC	<i>HindIII</i>
BsmD (C <sub>6</sub> -A <sub>6</sub> -T <sub>6</sub> )-22b-F	<b>AAGAAGGAGATATACATATG</b> TGTCGCTTGTGCTTT	<i>NdeI</i>
BsmD(C <sub>6</sub> -A <sub>6</sub> -T <sub>6</sub> )-22b-R	<b>CTCGAGTGCAGCCGCAAGCTT</b> GGGCACCTCGCCACTCC	<i>HindIII</i>
BsmF (A <sub>0</sub> -T <sub>0</sub> )-28a-F	<b>GTGCCGCGCGGCAGCCATATG</b> GCCGAGTGAACGACACC	<i>NdeI</i>
BsmF (A <sub>0</sub> -T <sub>0</sub> )-28a-R	<b>CTCGAGTGCAGCCGCAAGCTT</b> TTCAGTGTGCCTCTCGCCC	<i>HindIII</i>
BsmH-22b-F	<b>AAGAAGGAGATATACATATG</b> ATGTCCCGCCTTGTGAG	<i>NdeI</i>
BsmH-22b-R	<b>CTCGAGTGCAGCCGCAAGCTT</b> CTTCTTGTGCCGATGAC	<i>HindIII</i>

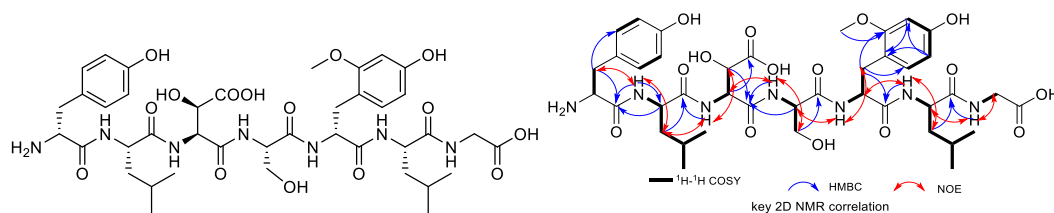
<sup>a</sup> Letters highlighted in bold are sequences used for ligation independent cloning and the enzyme sites are indicated by underline

**Table S3.** Substrate specificity predictions for the adenylation domains of the NRPSs encoded in the *bsm* gene cluster.

Module	Substrate recognition sequence	Corresponding amino acid	Predicted amino acid
M1	DASTIAAVCK	Tyr	Tyr
M2	DASTIAAVCK	Tyr	Tyr
M3	DAWMVGAVCK	Leu	Phe
M4	DLTKLGVVVK	Asp	Asn
M5	DVWHFSLVDK	Ser	Ser
M6	DASTIGAVCK	OMe-Tyr	Tyr
M7	DAWMVGAVCK	Leu	Phe
M8	DILQLGVIWK	Gly	Gly
M0	DGSIAALVWK	Tyr	Tyr

The prediction of the substrate specificity was based on NRPS Predictor2.<sup>18</sup>

**Table S4.**  $^1\text{H}$  (600 MHz) and  $^{13}\text{C}$  (150 MHz) NMR data of compound **1** in  $\text{DMSO-}d_6$ .

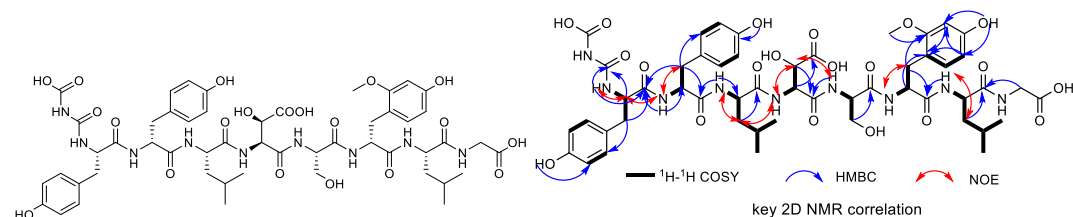


No	$\delta_{\text{C}}$ (type)	$\delta_{\text{H}}$ , multi. (J in Hz)	No	$\delta_{\text{C}}$ (type)	$\delta_{\text{H}}$ , multi. (J in Hz)
D-Tyr	169.44, C		D-OCH <sub>3</sub> -Tyr	171.49, C	
	54.67, CH	3.92, d (7.4)		54.28, CH	4.30, dd (14.6, 7.1)
	36.85, CH <sub>2</sub>	2.86, d (7.2)		31.57, CH <sub>2</sub>	2.92, dd (13.5, 6.5)
	125.39, C				2.72, dd (13.3, 8.5)
	130.69, CH	7.00, d (8.1)		115.96, C	
	115.65, CH	6.68, d (8.0)		157.74, C	
	156.96, C			55.55, CH <sub>3</sub>	3.70, s
	115.65, CH	6.68, d (8.0)		99.06, CH	6.31, s
	130.69, CH	7.00, d (8.1)		158.49, C	
L-Leu	171.47, C		L-Leu	131.51, CH	6.83, d (8.1)
	51.27, CH	4.28, m		106.87, CH	6.18, d (8.0)
	39.41, CH <sub>2</sub>	1.39, 1.26, m		NH	8.12, d (6.5)
	23.99, CH	1.02, m			
	23.63, CH <sub>3</sub>	0.75, d (6.5)			
	23.7, CH <sub>3</sub>	0.77, d (6.5)			
	NH	8.67, d (7.4)			
L-erythro-OHAsp	169.72, C				
	57.63, CH	4.29, m			
	72.86, CH	3.94, d (4.1)			
	174.47, C				
	NH	8.04, d (5.7)			
L-Ser	170.33, C		Gly	171.86, C	
	55.91, CH	4.12, m		41.68, CH <sub>2</sub>	3.70, d (5.4)
	61.92, CH <sub>2</sub>	3.60, m		NH	8.21, t (5.4)
	NH	7.78, d (7.2)			

**Table S5.**  $^1\text{H}$  (600 MHz) and  $^{13}\text{C}$  (150 MHz) NMR data of compound **2** in  $\text{DMSO-}d_6$ .

No	$\delta_{\text{C}}$ (type)	$\delta_{\text{H}}$ , multi. ( $J$ in Hz)	No	$\delta_{\text{C}}$ (type)	$\delta_{\text{H}}$ , multi. ( $J$ in Hz)
L-Tyr	172.41, C		L-Ser	170.29, C	
	54.82, CH	4.25, m		55.76, CH	4.18, q (6.0)
	38.04, CH <sub>2</sub>	2.57, dd (13.8, 4.0), 2.35, dd (13.9, 8.6)		62.06, CH <sub>2</sub>	3.56, dd (10.6, 5.3), 3.48, dd (10.1, 3.6)
	128.32, C		NH	7.79, d (7.4)	
	130.62, CH	6.70, d (8.1)	D-OCH <sub>3</sub> -Tyr	171.31, C	
	115.16, CH	6.54, d (8.4)		54.2, CH	4.33, dd (14.7, 7.4)
	156.1, C			31.87, CH <sub>2</sub>	2.82, dd (13.6, 4.3) 2.71, dd (13.3, 7.4)
	115.16, CH	6.54, d (8.4)		115.53, C	
	130.62, CH	6.70, d (8.1)	157.84, C		
	D-Tyr	171.62, C		55.53, CH <sub>3</sub>	3.70, s
54.92, CH		4.46, dd (14.2, 8.7)	99.07, CH	6.30, d (1.9)	
37.67, CH <sub>2</sub>		2.86, 2.59, m	158.52, C		
128.17, C			131.44, CH	6.79, d (8.1)	
130.69, CH		7.06, d (8.3)	106.88, CH	6.17, dd (8.1, 2.2)	
115.23, CH		6.62, d (8.4)	NH	7.93, br s	
156.31, C			L-Leu	172.88, C	
115.23, CH		6.62, d (8.4)		51.31, CH	4.11, dd (15.4, 8.1)
130.69, CH	7.06, d (8.3)	40.73, CH <sub>2</sub>	1.35, m		
NH	8.24, d (8.6)	24.08, CH	1.17, m		
L-Leu	172.55, C		21.69, CH <sub>3</sub>	0.68, d (6.5)	
	51.27, CH	4.42, m	21.74, CH <sub>3</sub>	0.77, d (6.5)	
	41.35, CH <sub>2</sub>	1.46, 1.41, m	NH	7.88, d (8.2)	
	24.45, CH	1.42, m	Gly	171.53, C	
	23.62, CH <sub>3</sub>	0.76, d (5.7)		41.13, CH <sub>2</sub>	3.71, s
	23.71, CH <sub>3</sub>	0.80, d (5.7)		NH	8.11, br s
	NH	8.19, d (9.0)			
L-erythro-OHAsp	169.09, C				
	55.53, CH	4.70, t (5.4)			
	71.46, CH	4.08, d (5.3)			
	173.27, C				
NH	8.29, br s				

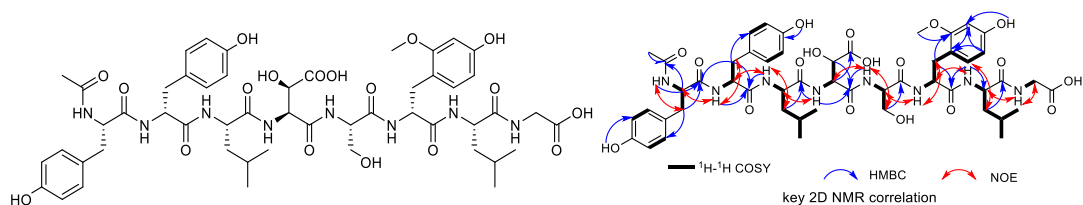
**Table S6.**  $^1\text{H}$  (600 MHz) and  $^{13}\text{C}$  (150 MHz) NMR data of compound **3** in  $\text{DMSO-}d_6$ .



No	$\delta_{\text{C}}$ $\delta_{\text{N}}$ (type)	$\delta_{\text{H}}$ , multi. ( $J$ in Hz)	No	$\delta_{\text{C}}$ $\delta_{\text{N}}$ (type)	$\delta_{\text{H}}$ , multi. ( $J$ in Hz)	
L-Tyr	169.84, C		L- <i>erythro</i> -OHAsp	168.65, C		
	53.85, CH	4.47, m		55.34, CH	4.71, dd (7.6, 5.8)	
	36.59, CH <sub>2</sub>	2.69, 2.58, m		71.85, CH	4.06, d (3.7)	
	127.11, C			171.14, C		
	130.11, CH	6.69, d (8.3)		NH	8.353, br s	
	114.82, CH	6.53, d (8.4)		L-Ser	169.74, C	
	155.92, C				55.24, CH	4.16, m
	114.82, CH	6.53, d (8.4)			61.56, CH <sub>2</sub>	3.56, dd (10.4, 5.0), 3.49, dd (10.4, 4.0)
	130.11, CH	6.69, d (8.3)			115.34, NH	7.76, d (6.6)
	114.93, NH	8.20, d (8.6)		D-OCH <sub>3</sub> -Tyr	170.89, C	
158.84, C		54.20, CH	4.33, dd (14.9, 7.5)			
158.07, C		31.37, CH <sub>2</sub>	2.71, dd (13.0, 7.7), 2.89, dd (13.4, 4.5)			
D-Tyr	170.85, C		115.23, C			
	54.92, CH	4.54, m	156.31, C			
	37.68, CH <sub>2</sub>	2.88, dd (13.4, 4.5), 2.63, m	55.09, CH <sub>3</sub>		3.70, s	
	128.05, C		98.62, CH	6.30, d (2.0)		
	130.23, CH	7.07, d (8.4)	157.35, C			
	114.79, CH	6.63, d (8.4)	131.04, CH	6.80, d (8.1)		
	155.79, C		106.43, CH	6.18, dd (8.1, 2.0)		
	114.79, CH	6.63, d (8.4)	120.75, NH	7.96, br s		
130.23, CH	7.07, d (8.4)	L-Leu	172.43, C			
117.76, NH	8.40, d (8.3)		50.97, CH	4.12, dd (15.7, 7.9)		
L-Leu	172.07, C			40.27, CH <sub>2</sub>	1.36, m	
	50.67, CH		4.44, m	23.68, CH	1.23, m	
	41.16, CH <sub>2</sub>	1.43, m	21.29, CH <sub>3</sub>	0.69, d (6.4)		
	24.09, CH	1.43, m	21.22, CH <sub>3</sub>	0.78, d (6.4)		
	23.17, CH <sub>3</sub>	0.80, d (6.6)	120.37, NH	7.89, d (8.3)		
	23.27, CH <sub>3</sub>	0.76, d (6.6)	Gly	172.84, C		
	119.15, NH	8.25, d (8.4)		40.85, CH <sub>2</sub>	3.71, s	

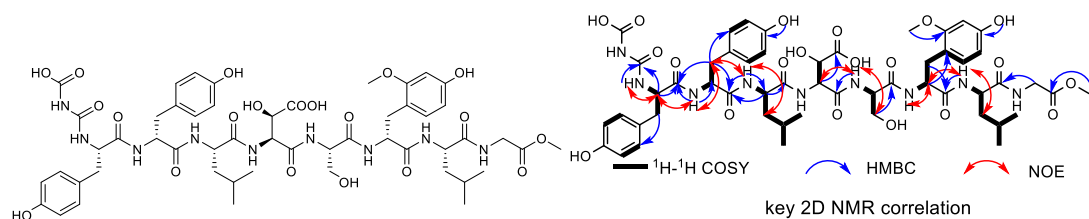


**Table S7.**  $^1\text{H}$  (600 MHz) and  $^{13}\text{C}$  (150 MHz) NMR data of compound **4** in  $\text{DMSO-}d_6$ .



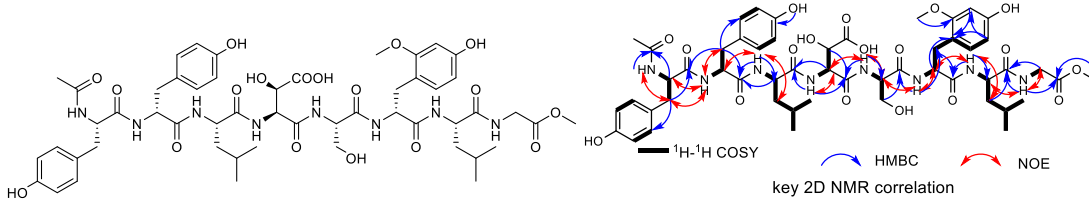
No	$\delta_{\text{C}}$ (type)	$\delta_{\text{H}}$ , multi. ( $J$ in Hz)	No	$\delta_{\text{C}}$ (type)	$\delta_{\text{H}}$ , multi. ( $J$ in Hz)	
L-Tyr	170.89, C		L- erythro –OHAsp	168.66, C		
	54.73, CH	4.47, m		55.62, CH	4.72, m	
	37.27, $\text{CH}_2$	2.37, 2.58, m		71.40, CH	4.00, m	
	128.45, C			171.34, C		
	130.5, CH	6.88, d (7.8)		NH	8.322, d (7.6)	
	115.34, CH	6.61, d (8.1)		L-Ser	169.82, C	
	155.8, C				55.69, CH	4.19, m
	115.34, CH	6.61, d (7.8)			62.03, $\text{CH}_2$	3.50, 3.56, m
	130.5, CH	6.88, d (8.1)		D-OCH <sub>3</sub> -Tyr	NH	7.80, d (7.1)
	NH	8.27, d (8.2)			171.11, C	
169.26, C		54.21, CH	4.31, m			
D-Tyr	22.83, $\text{CH}_3$	1.721, s	31.93, $\text{CH}_2$	2.71, 2.82, m		
	D-Tyr	169.73, C		115.57, C		
		54.68, CH	4.54, dd (13.4, 8.3)	157.42, C		
		37.95, $\text{CH}_2$	2.85, 2.60, m	55.51, $\text{CH}_3$	3.71, s	
		128.02, C		99.13, CH	6.31, s	
		130.67, CH	7.01, d (7.9)	158.07, C		
		115.33, CH	6.59, d (7.8)	131.42, CH	6.79, d (8.0)	
		156.28		106.88, CH	6.18, d (8.1)	
		115.33, CH	6.59, d (7.8)	NH	7.93, br s	
		130.67, CH	7.01, d (7.9)	L-Leu	172.4, C	
NH		8.40, d (7.9)	51.33, CH		4.11, dd (13.4,6.6)	
L-Leu	172.15, C		40.78, $\text{CH}_2$		1.35, m	
	51.11, CH	4.41, m	24.12, CH	1.17, m		
	41.63, $\text{CH}_2$	1.43, m	21.74, $\text{CH}_3$	0.69, d (6.1)		
	24.58, CH	1.42, m	21.57, $\text{CH}_3$	0.76, d (6.1)		
	23.65, $\text{CH}_3$	0.77, d (6.4)	NH	7.89, d (8.0)		
	23.74, $\text{CH}_3$	0.80, d (6.4)	Gly	172.73, C		
	NH	8.19, d (7.8)		41.15, $\text{CH}_2$	3.71, s	
				NH	8.12, br s	

**Table S8.**  $^1\text{H}$  (600 MHz) and  $^{13}\text{C}$  (150 MHz) NMR data of compound **5** in  $\text{DMSO-}d_6$ .



No	$\delta_{\text{C}}$ (type)	$\delta_{\text{H}}$ , multi. ( $J$ in Hz)	No	$\delta_{\text{C}}$ (type)	$\delta_{\text{H}}$ , multi. ( $J$ in Hz)	
L-Tyr	170.24, C		L- erythro -OHAsp	169.24, C		
	54.43, CH	4.48, m		56.49, CH	4.63, m	
	37.67, CH <sub>2</sub>	2.72, 2.62, m		72.98, CH	3.98, d (3.8)	
	127.52, C			170.04, C		
	130.56, CH	6.71, d (8.3)		L-Ser	170.39, C	
	115.26, CH	6.53, d (8.4)			55.66, CH	4.11, m
	156.37, C				62.01, CH <sub>2</sub>	3.60, 3.52, m
	115.26, CH	6.53, d (8.4)		NH	7.67, d (6.9)	
	130.56, CH	6.71, d (8.3)		D-OCH <sub>3</sub> -Tyr	54.94, CH	4.52, m
	NH	8.25, d (9.1)			171.51, C	
159.31, C		54.18, CH	4.34, dd (14.9, 7.8)			
D-Tyr	171.30, C		31.64, CH <sub>2</sub>	2.70, m		
	54.94, CH	4.52, m		2.98, dd (13.3, 4.7)		
	37.97, CH <sub>2</sub>	2.89, 2.66, m	115.23, C			
	128.18, C		156.39, C			
	130.56, CH	7.06, d (8.4)	55.54, CH <sub>3</sub>	3.70, s		
	115.25, CH	6.63, d (8.4)	99.05, CH	6.30, d (1.9)		
	156.27, C		158.49, C			
	115.25, CH	6.63, d (8.4)	131.59, CH	6.82, d (8.2)		
	130.56, CH	7.06, d (8.4)	106.88, CH	6.18, dd (8.1, 1.9)		
	NH	8.40, d (8.3)	NH	8.15, br s		
L-Leu	172.35, C		L-Leu	173.25, C		
	51.25, CH	4.43, m		51.68, CH	4.16, dd (14.6, 9.1)	
	40.68, CH <sub>2</sub>	1.44, m		41.11, CH <sub>2</sub>	1.44, m	
	24.46, CH	1.44, m		24.27, CH	1.314 m	
	23.75, CH <sub>3</sub>	0.82, d (5.4)		21.79, CH <sub>3</sub>	0.73, d (6.4)	
	23.52, CH <sub>3</sub>	0.80, d (5.4)		21.77, CH <sub>3</sub>	0.79, d (6.4)	
	NH	8.22, d (9.1)		NH	7.96, d (8.3)	
				Gly	52.06, CH <sub>3</sub>	3.58, s
		170.58, C				
			41.36, CH <sub>2</sub>	3.81, dd (7.6, 6.6)		

**Table S9.**  $^1\text{H}$  (600 MHz) and  $^{13}\text{C}$  (150 MHz) NMR data of compound **6** in  $\text{DMSO-}d_6$ .

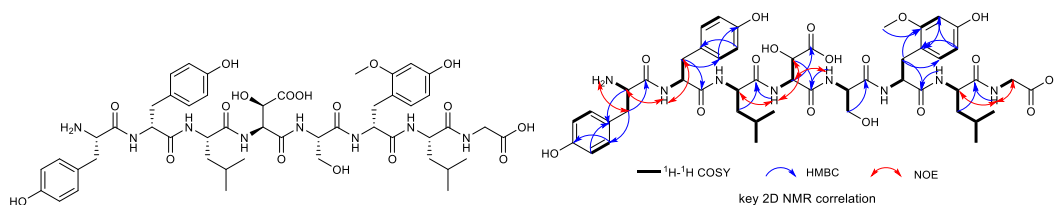


No	$\delta_c$ (type)	$\delta_H$ , multi. ( $J$ in Hz)	No	$\delta_c$ (type)	$\delta_H$ , multi. ( $J$ in Hz)	
L-Tyr	171.82, C		L- <i>erythro</i> -OHAsp	169.04, C		
	54.8, CH	4.38, m		55.45, CH	4.73, dd (8.77, 5.9)	
	37.35, CH <sub>2</sub>	2.38, dd (13.7, 9.6)		71.38, CH	4.10, d (5.8)	
		2.55, dd (13.2, 9.4)		173.08, C		
	128.45, C			NH	8.32, d (6.5)	
	130.53, CH	6.88, d (8.4)		L-Ser	170.37, C	
	115.32, CH	6.61, d (8.3)			55.69, CH	4.20, dd (12.4, 5.0)
	156.15, C				62.06, CH <sub>2</sub>	3.50, 3.55, dd (11.0, 5.3)
	115.34, CH	6.61, d (8.3)			NH	7.77, d (7.5)
	130.51, CH	6.88, d (8.3)		D-OCH <sub>3</sub> -Tyr	171.4, C	
NH	7.91, d (7.9)	54.24, CH	4.30, dd (14.7, 7.4)			
169.79, C		31.82, CH <sub>2</sub>	2.71, dd (13.3, 7.1), 2.82, dd (13.6, 5.1)			
22.86, CH <sub>3</sub>	1.73, s					
D-Tyr	171.65, C			115.57, C		
	54.91, CH	4.46, d (14.3, 8.5)		157.86, C		
	37.71, CH <sub>2</sub>	2.83, 2.61, m		55.54, CH <sub>3</sub>	3.70, s	
	128.02, C			99.03, CH	6.31, d (6.3)	
	130.67, CH	7.01, d (8.5)		158.52, C		
	115.33, CH	6.59, d (8.3)		131.41, CH	6.77, d (8.2)	
	156.28, C			106.88, CH	6.18, dd (8.1, 2.0)	
	115.33, CH	6.59, d (8.5)		NH	7.96, d (6.7)	
	130.67, CH	7.01, d (8.5)	L-Leu	173.11, C		
	NH	8.29, d (7.8)		51.23, CH	4.11, d (5.8)	
		40.57, CH <sub>2</sub>		1.34, m		
L-Leu	172.64, C			24.04, CH	1.15, m	
	51.29, CH	4.37, m		21.74, CH <sub>3</sub>	0.68, d (6.5)	
	41.17, CH <sub>2</sub>	1.46, m		21.58, CH <sub>3</sub>	0.75, d (6.5)	
	24.41, CH	1.42, m		NH	7.93, d (5.5)	
	23.57, CH <sub>3</sub>	0.77, d (5.6)	Gly	52.26, CH <sub>3</sub>	3.60, s	
	23.68, CH <sub>3</sub>	0.79, d (5.6)		170.57, C		
	NH	8.19, d (8.2)		41.05, CH <sub>2</sub>	3.80, d (5.9)	
				NH	8.16, d (6.0)	

**Table S10.**  $^1\text{H}$  (600 MHz) and  $^{13}\text{C}$  (150 MHz) NMR data of compound **13** in  $\text{DMSO-}d_6$ .

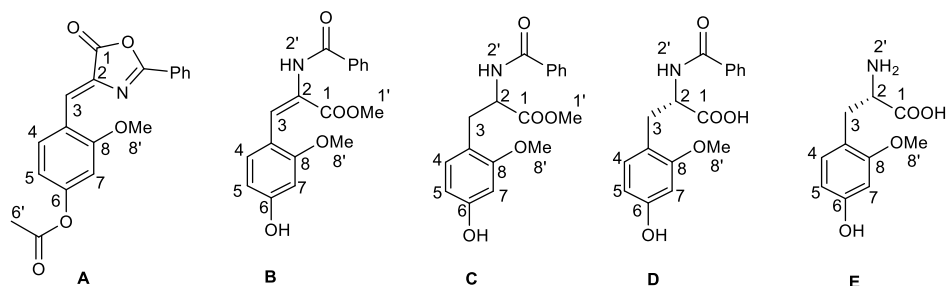
No	$\delta_{\text{C}}$ (type)	$\delta_{\text{H}}$ , multi. (J in Hz)	No	$\delta_{\text{C}}$ (type)	$\delta_{\text{H}}$ , multi. (J in Hz)
L-Tyr	170.64, C		L-Ser	169.40, C	
	54.12, CH	4.70, s		55.10, CH	4.22, m
	36.19, CH <sub>2</sub>	2.57, m		61.70, CH <sub>2</sub>	3.50, dd (10.3, 5.3), 3.54, dd (10.3, 5.3)
		2.79, m		NH	7.74, d (7.6)
	127.15, C		D-Tyr	170.54, C	
	130.56, CH	6.79, d (8.7)		54.80, CH	4.41, m
	114.81, CH	6.61, d (8.7)		37.95, CH <sub>2</sub>	2.69, m
	156.55, C				2.79, m
	114.81, CH	6.61, d (8.7)		127.26, C	
	130.56, CH	6.79, d (8.7)		130.28, CH	7.03, d (8.4)
NH	8.78, d (8.6)	114.81, CH		6.61, d (8.4)	
D-Tyr	167.73, C		156.06, C-OH	9.18, s	
	54.80, CH	4.40, m	114.81, CH	6.61, d (8.4)	
	37.05, CH <sub>2</sub>	2.69, 2.78, m	130.28, CH	7.03, d (8.4)	
	124.60, C		NH	7.89, br s	
	130.16, CH	6.96, d (8.4)	L-Leu	172.39, C	
	114.81, CH	6.61, d (8.4)		50.75, CH	4.21, dd (15.4, 8.1)
	155.84, C-OH	9.16, s		40.63, CH <sub>2</sub>	1.35, m
	114.81, CH	6.61, d (8.4)		23.77, CH	1.27, m
	130.16, CH	6.96, d (8.4)	21.27, CH <sub>3</sub>	0.72, d (6.3)	
	NH	8.47, d (8.2)	21.11, CH <sub>3</sub>	0.75, d (6.3)	
L-Leu	172.34, C		NH	8.13, d (8.3)	
	50.79, CH	4.38, m	Gly	171.08, C	
	40.67, CH <sub>2</sub>	1.44, m		40.62, CH <sub>2</sub>	3.71, t (5.9)
	24.09, CH	1.44, m		NH	8.19, t (5.9)
	23.17, CH <sub>3</sub>	0.82, d (6.2)			
	23.28, CH <sub>3</sub>	0.79, d (6.2)			
	NH	7.94, d (7.5)			
L-erythro-OHAsp	168.60, C				
	54.10, CH	4.68, m			
	71.00, CH	4.07, d (5.7)			
	172.66, C				
	NH	8.33, d (7.4)			

**Table S11.**  $^1\text{H}$  (600 MHz) and  $^{13}\text{C}$  (150 MHz) NMR data of compound **14** in  $\text{DMSO-}d_6$ .



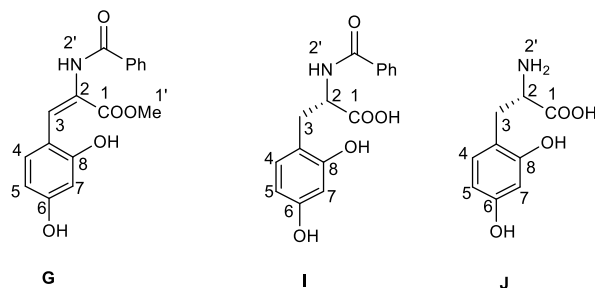
No	$\delta_{\text{C}}$ (type)	$\delta_{\text{H}}$ , multi. ( $J$ in Hz)	No	$\delta_{\text{C}}$ (type)	$\delta_{\text{H}}$ , multi. ( $J$ in Hz)	
L-Tyr	171.26, C		L-Ser	170.08, C		
	55.00, CH	4.71, m		55.50, CH	4.21, m	
	38.46, CH <sub>2</sub>	2.58, m		61.90, CH <sub>2</sub>	3.60, dd (10.9, 5.0), 3.64, dd (11.0, 5.4)	
		2.81, m			7.78, d (7.7)	
	127.60, C			D-OCH <sub>3</sub> -Tyr	171.04, C	
	130.74, CH	7.03, d (8.4)			54.09, CH	4.35, dd (14.7, 7.4)
	115.44, CH	6.62, d (8.4)			32.05, CH <sub>2</sub>	2.79, m
	156.50, C					2.71, m
	115.44, CH	6.62, d (8.4)			115.44, C	
	130.74, CH	7.03, d (8.4)		158.53, C		
	8.78, d (8.7)					
D-Tyr	168.16, C		55.55, CH <sub>3</sub>	3.70, s		
	54.03, CH	4.40, m	99.08, CH	6.31, d (1.3)		
	36.64, CH <sub>2</sub>	2.77, 2.80, m	157.87, C-OH	9.20, s		
	125.08, C		131.41, CH	6.79, dd (8.1, 3.9)		
	131.03, CH	6.79, d (8.1)	106.88, CH	6.18, d (8.0)		
	115.44, CH	6.62, d (8.3)		7.88, br s		
	157.87, C-OH	9.17, s	L-Leu	172.77, C		
	115.44, CH	6.62, d (8.3)		51.26, CH	4.13, m	
	131.03, CH	6.79, d (8.3)		40.81, CH <sub>2</sub>	1.34, m	
		8.47, d (8.2)		24.05, CH	1.17, m	
L-Leu	172.64, C		21.77, CH <sub>3</sub>	0.69, d (6.5)		
	51.18, CH	4.40, m	21.54, CH <sub>3</sub>	0.79, d (6.5)		
	41.40, CH <sub>2</sub>	1.44, m		7.90, d (7.5)		
	24.55, CH	1.44, m	Gly	173.07, C		
	23.62, CH <sub>3</sub>	0.76, d (5.7)		41.08, CH <sub>2</sub>	3.71, s	
	23.72, CH <sub>3</sub>	0.83, d (5.7)			8.10, d (5.8)	
		7.90, d (9.0)				
	L-erythro -OHAsp	169.06, C				
55.00, CH		4.71, m				
71.44, CH		4.07, d (5.8)				
172.89, C						
		8.37, d (8.7)				

**Table S12.**  $^1\text{H}$  NMR (400 MHz) data of **A-E** in  $\text{DMSO-}d_6$ .



No	A	B	C	D	E
	$\delta_{\text{H}}$ (multi, <i>J</i> , Hz)	$\delta_{\text{H}}$ (multi, <i>J</i> , Hz)	$\delta_{\text{H}}$ (multi, <i>J</i> , Hz)	$\delta_{\text{H}}$ (multi, <i>J</i> , Hz)	$\delta_{\text{H}}$ (multi, <i>J</i> , Hz)
1'		3.67 (s)	3.60 (s)		
2			4.58 (dd, 13.6, 9.0)	4.55 (m)	3.96 (t, 5.5)
2'-NH			8.66 (d, 7.6)	8.49 (d, 8.0)	8.12 (s)
3	7.51 (s)	7.70 (s)	2.86 (dd, 13.5, 9.7)	2.82 (dd, 13.5, 10.5)	2.86 (dd, 14.0, 7.1)
			3.10 (dd, 13.5, 5.5)	3.16 (dd, 13.5, 4.4)	3.30 (dd, 14.0, 6.4)
4	8.81 (d, 8.6)	7.48 (d, 7.7)	6.97 (d, 8.1)	7.00 (d, 8.0)	6.91 (d, 8.1)
5	6.94 (dd, 8.6, 1.8)	6.20 (d, 8.6)	6.97 (d, 8.1)	6.23 (dd, 8.1, 1.5)	6.29 (dd, 8.1, 2.2)
6-OH		9.75 (brs)	9.35 (brs)	9.24 (brs)	9.47 (brs)
6'	2.31 (s)				
7	7.00 (d, 1.8)	6.32 (s)	6.37 (d, 1.5)	6.36 (d, 1.5)	6.40 (d, 2.1)
8'	3.91 (s)	3.75 (s)	3.74 (s)	3.73 (s)	3.71 (s)
Ph	8.12 (d, 7.4)	7.95 (d, 7.5)	7.79 (d, 7.4)	7.77 (d, 7.5)	
	8.12 (d, 7.4)	7.95 (d, 7.5)	7.79 (d, 7.4)	7.77 (d, 7.5)	
	7.73 (m)	7.56 (t, 7.1)	7.53 (t, 7.2)	7.53 (t, 7.2)	
	7.64 (t, 7.6)	7.48 (t, 7.6)	7.46 (t, 7.1)	7.45 (t, 7.5)	
	7.64 (t, 7.6)	7.48 (t, 7.6)	7.46 (t, 7.1)	7.45 (t, 7.5)	

**Table S13.** <sup>1</sup>H NMR (400 MHz) data of **G**, **I** and **J** in DMSO-*d*<sub>6</sub>.



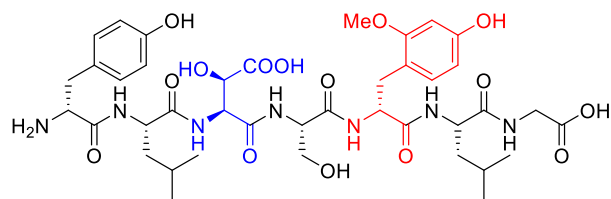
No	<b>G</b>	<b>I</b>	<b>J</b>
	$\delta_{\text{H}}$ (multi, <i>J</i> , Hz)	$\delta_{\text{H}}$ (multi, <i>J</i> , Hz)	$\delta_{\text{H}}$ (multi, <i>J</i> , Hz)
1'	3.69 (s)		
2		4.55 (m)	3.95 (t, 5.5 )
2'-NH		8.49 (brd, 7.7)	8.28 ((brd, 7.7)
3	7.70 (s)	2.80 (dd,13.6, 10.3)	2.85 (dd, 13.8, 7.4)
		3.08 (dd, 13.6, 4.3)	2.98 (dd, 13.8, 6.6)
4	7.52 (d, 8.6)	6.92 (d, 8.2)	6.83 (d, 8.2)
5	6.19 (dd, 8.6, 2.1)	6.09 (dd, 8.1, 2.5)	6.16 (dd, 8.1, 2.4 )
6-OH	9.75 (s)	9.03 (brs)	
7	6.37 (d, 2.1)	6.27 (d, 2.3)	6.39 (d, 2.3 )
8-OH	10.10 (brs)	9.47 (brs)	9.66 (brs)
Ph	7.96 (d, 7.4)	7.77 (d, 7.5)	
	7.96 (d, 7.4)	7.77 (d, 7.5)	
	7.58 (t, 7.4)	7.52 (t, 7.5)	
	7.48 (t, 7.4)	7.45 (t, 7.5)	
	7.48 (t, 7.4)	7.45 (t, 7.5)	

**Table S14.** Conserved sequence regions in the alignment comparisons of BsmF and other known P450s (numbering indicated for BsmF).<sup>19,20</sup>

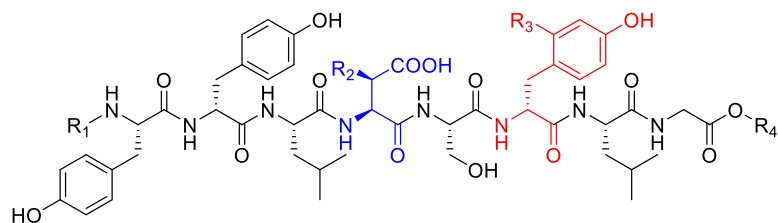
P450	B-B2 loop N-term	B-B2 loop C-term	F-helix	G-helix	I-helix	$\beta$ -1 sheet
	87 90	107 108 109 111 114 115	194 195 196 197	216 217 218 219 222	270 271 277	324 325 326
OxyD	G I	S G G M V S	H A F G	A H T E V	N C G	A M H
CloI	G L	A S G M V T	H A W S	A K N E L	N C G	S L H
NovI	G L	A S G M V T	H A W S	A K N E L	N C G	S L H
SimD1	G L	A S R M L T	H A L S	A K N E L	N C G	S L H
Sky32	G L	A A G M V T	S A L S	A R N E L	N C G	A M H
Consensus*	G L	A <sup>(1)</sup> G M V T	H A <sup>(2)</sup> S	A <sup>(3)</sup> N E <sup>(4)</sup>	<b><u>N</u></b> C <b><u>G</u></b>	<sup>(5)</sup> LM H
BsmF-P450	A G	M G S Q F N	S Y E R	L L D K A	N A G	S Q Y

\* Identity residues shown in bold and underlined, mismatching residues or similar residues indicated in normal font. Exceptions are: <sup>(1)</sup> Small residue (S, G, A), <sup>(2)</sup> large hydrophobic (W, F, L), <sup>(3)</sup> positively charged residue (K, H, R), <sup>(4)</sup> majority hydrophobic (L, V; also S and G), <sup>(5)</sup> majority small (S, A; also V). Protein accession number: OxyD (3MGX\_A); CloI (AAN65225); NovI (Q9L9F9); SimD1 (AAK06805); Sky32 (4L0F\_A).



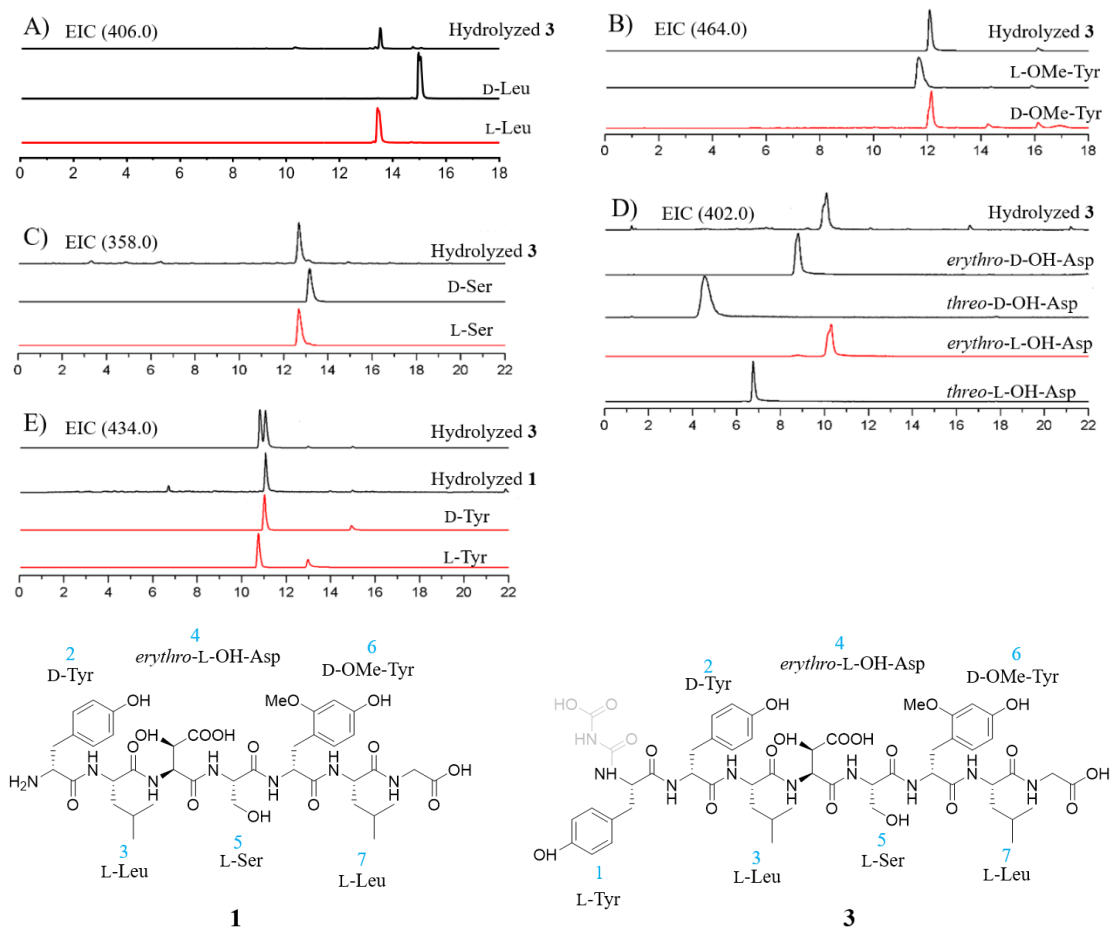


1 bosamycin A

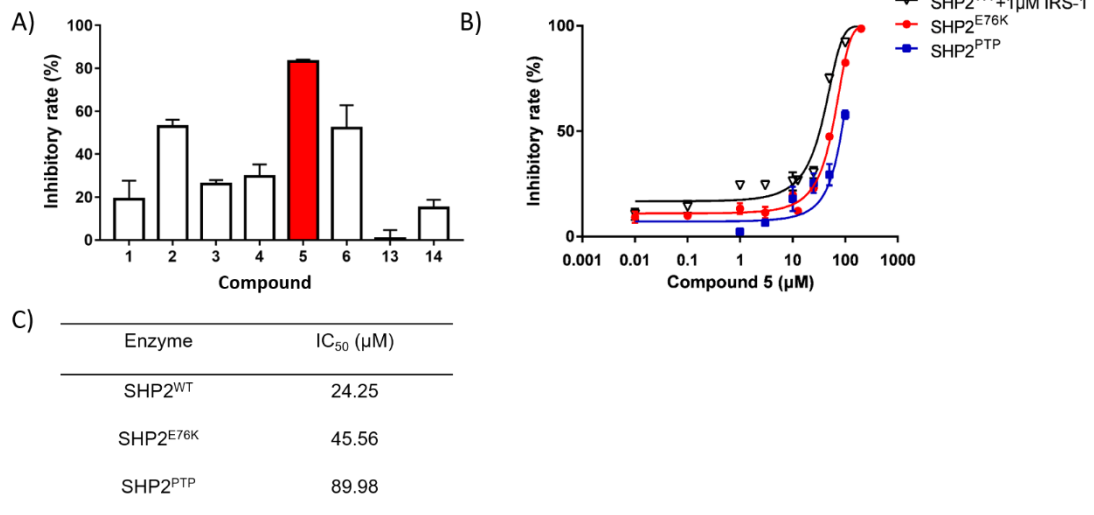


		R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	R <sub>4</sub>
2	bosamycin B		-OH	-OMe	H
3	bosamycin C		-OH	-OMe	H
4	bosamycin D		-OH	-OMe	H
5	bosamycin E		-OH	-OMe	-Me
6	bosamycin F		-OH	-OMe	-Me
7	bosamycin G		-OH	H	H
8	bosamycin H		-OH	H	H
9	bosamycin I		-OH	H	H
10	bosamycin J		H	-OMe	H
11	bosamycin K		H	-OMe	H
12	bosamycin L		H	-OMe	-Me
13	bosamycin M	H	-OH	H	H
14	bosamycin N	H	-OH	-OMe	H

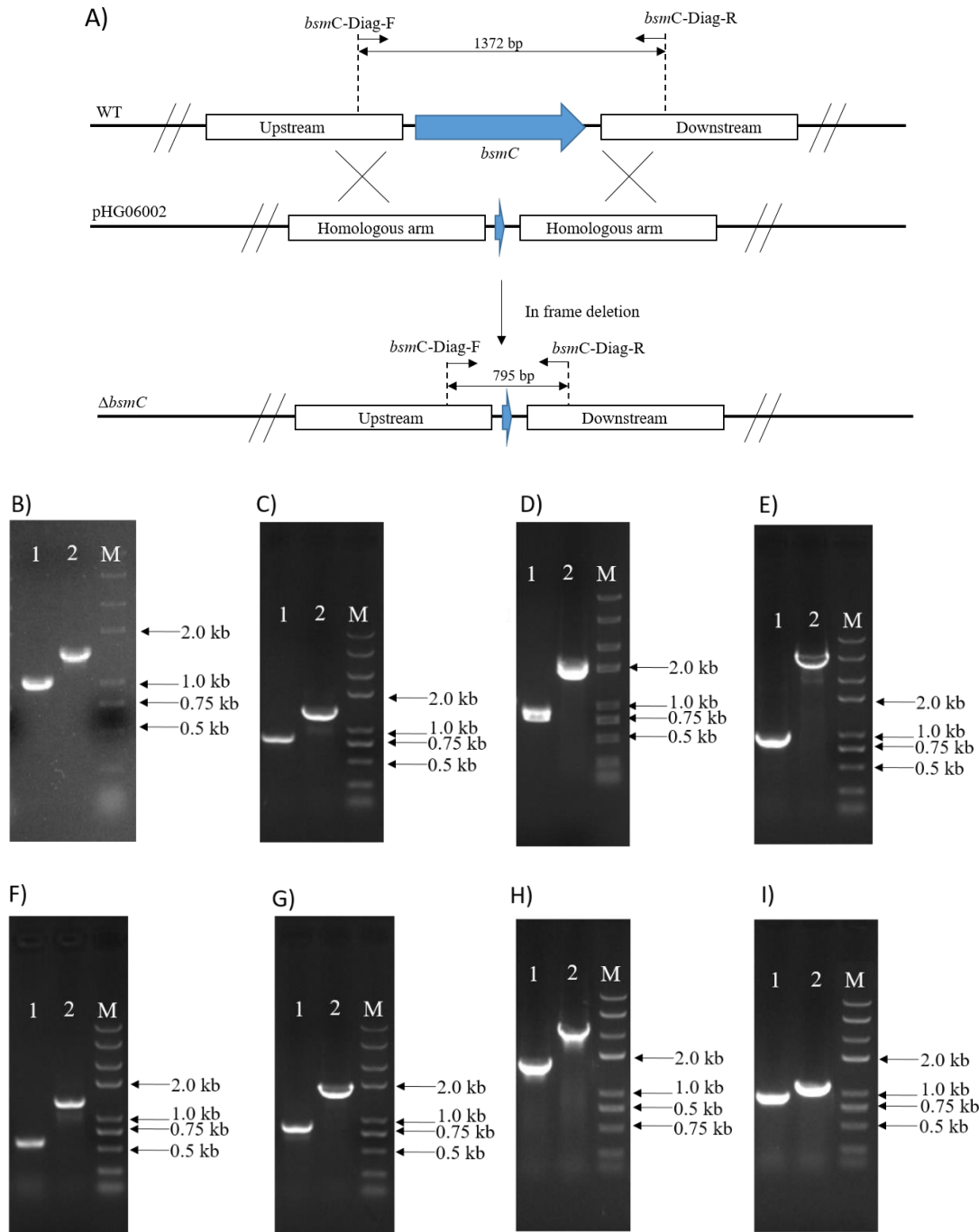
Figure S1. Structures of bosamycins.



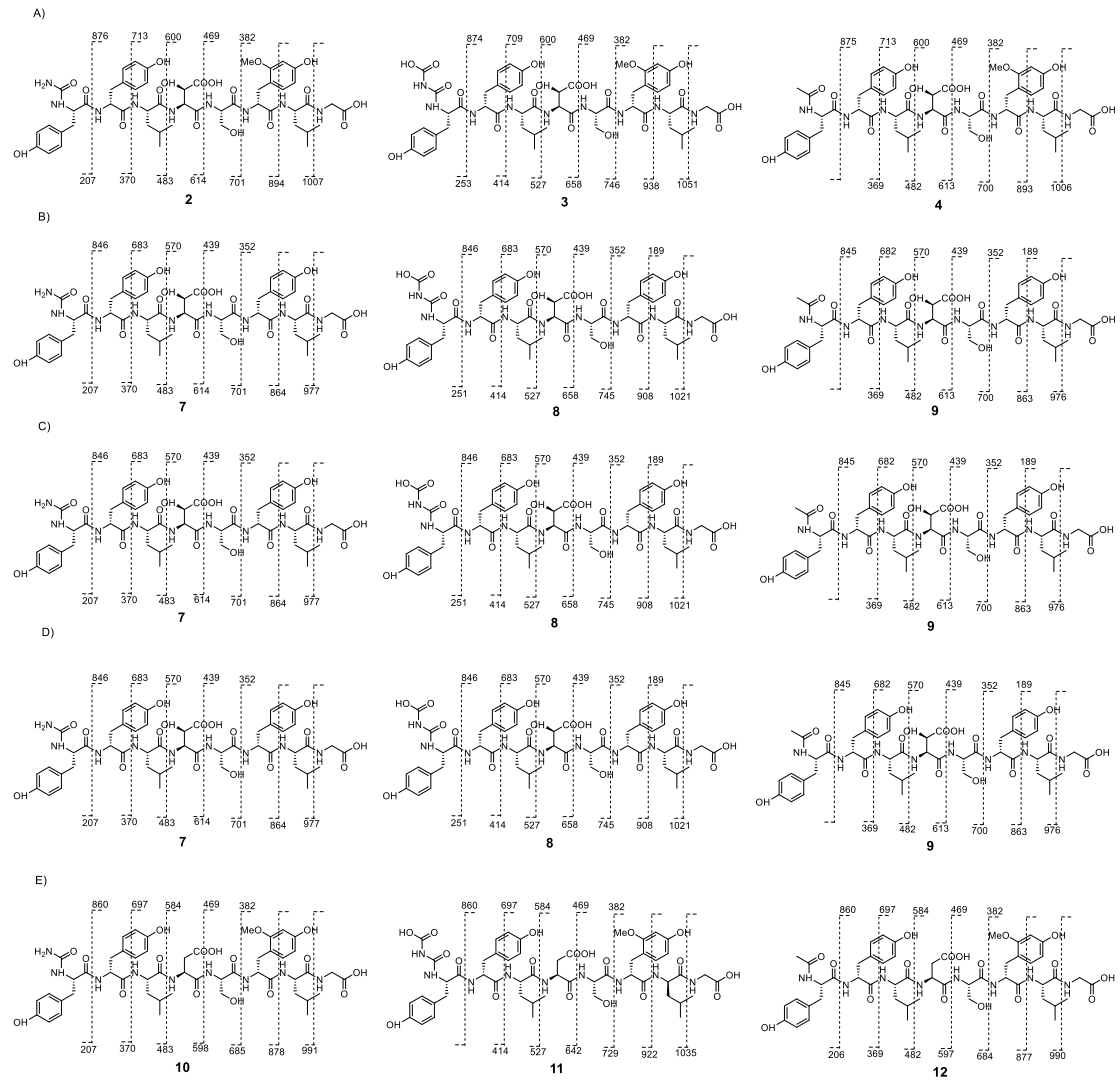
**Figure S2.** LC-MS analysis of L-FDAA and D-FDAA derivatives of the amino acid residues in **1** and **3**. Panel A indicates 3<sup>rd</sup> and 7<sup>th</sup> Leu in **3** is L-type; Panel B indicates 6<sup>th</sup> OMe-Tyr in **3** is D-type, Panel C indicates 5<sup>th</sup> Ser in **3** is L-type; Panel D indicates 4<sup>th</sup> OH-Asp in **3** is *erythro*-L-OH-Asp; Panel E indicates 2<sup>nd</sup> Tyr in **1** and **3** is D-type, 1<sup>st</sup> Tyr in **3** is L-type. The deduced D-type configurations in 2<sup>nd</sup>, and 6<sup>th</sup> amino acid residues are consistent with the presence of E domains in their corresponding modules (Scheme 1).



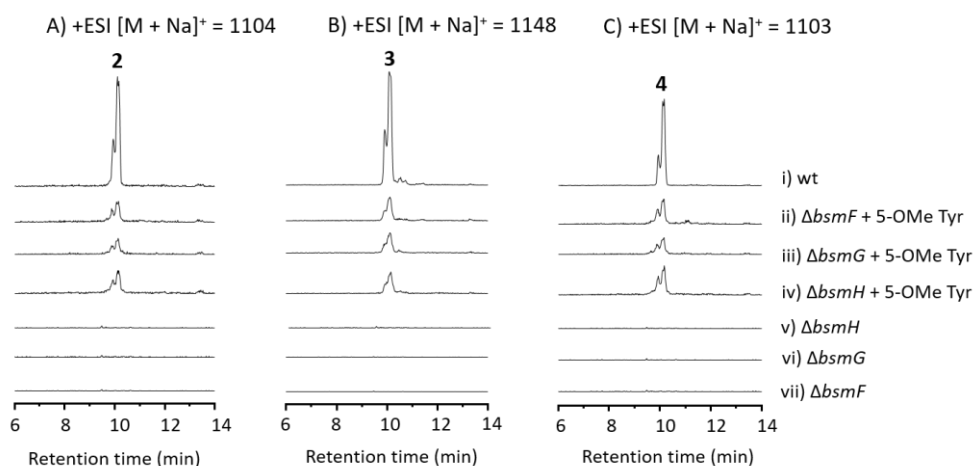
**Figure S3.** Identification of compound **5** as a novel inhibitor of SHP2. A) Primary screen the biological activity of compounds on SHP2 enzyme activity were examined. SHP2 was screened in the presence of 1  $\mu$ M 2P-IRS-1 and 30  $\mu$ M of each compound. B) Phosphatase activities of SHP2<sup>WT</sup>, SHP2<sup>E76K</sup>, SHP2<sup>PTP</sup> were assessed in the presence of compound **5** at various concentrations. C) The IC<sub>50</sub> value of **5** against SHP2<sup>WT</sup>, SHP2<sup>E76K</sup>, SHP2<sup>PTP</sup>.



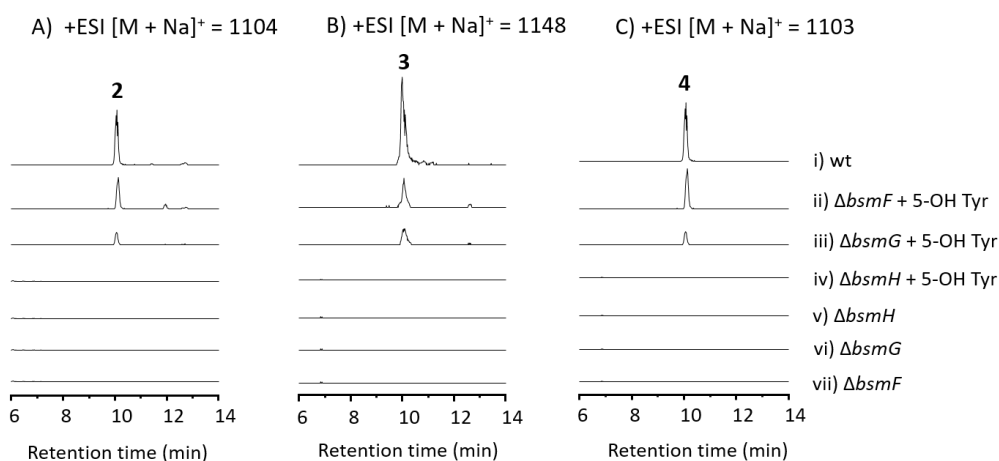
**Figure S4.** Generation of the *S. sp.* 120454 mutant strains. A) Gene disruption with homologous recombination strategies. B) The *S. sp.* 120454 HG06001 mutant ( $\Delta bsmA-C_1$ ). Lane 1, mutant strain; Lane 2, *S. sp.* 120454 WT; C) The *S. sp.* 120454 HG06002 mutant ( $\Delta bsmC$ ). Lane 1, mutant strain; Lane 2, *S. sp.* 120454 WT; D) The *S. sp.* 120454 HG06003 mutant ( $\Delta bsmD$ ). Lane 1, mutant strain; Lane 2, *S. sp.* 120454 WT; E) The *S. sp.* 120454 HG06004 mutant ( $\Delta bsmF$ ). Lane 1, mutant strain; Lane 2, *S. sp.* 120454 WT; F) The *S. sp.* 120454 HG06005 mutant ( $\Delta bsmG$ ). Lane 1, mutant strain; Lane 2, *S. sp.* 120454 WT; G) The *S. sp.* 120454 HG06006 mutant ( $\Delta bsmH$ ). Lane 1, mutant strain; Lane 2, *S. sp.* 120454 WT; H) The *S. sp.* 120454 HG06007 mutant ( $\Delta orf(-1)$ ). Lane 1, mutant strain; Lane 2, *S. sp.* 120454 WT; I) The *S. sp.* 120454 HG06008 mutant ( $\Delta bsmI$ ). Lane 1, mutant strain; Lane 2, *S. sp.* 120454 WT; Lane M, *Trans2K*<sup>®</sup> Plus II DNA marker.



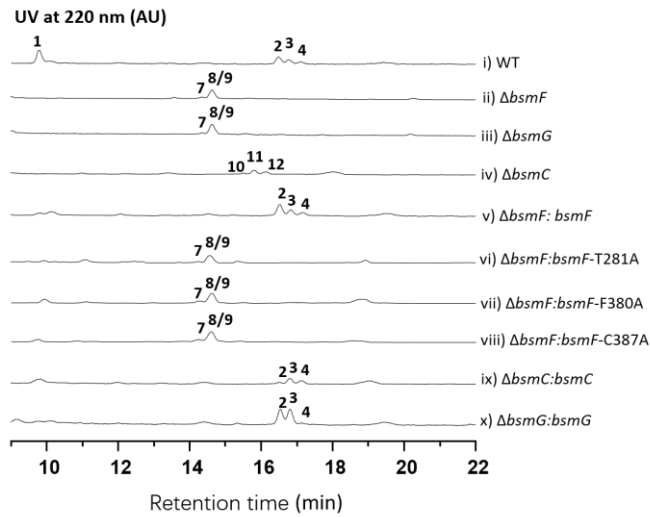
**Figure S5.** MS/MS analysis of metabolite extracts from wild type and mutant strains. A) *S. sp.* 120454 WT. B) The *S. sp.* 120454 HG06004 mutant ( $\Delta bsmF$ ). C) The *S. sp.* 120454 HG06005 mutant ( $\Delta bsmG$ ). D) The *S. sp.* 120454 HG06005 mutant ( $\Delta bsmH$ ). E) The *S. sp.* 120454 HG06002 mutant ( $\Delta bsmC$ ).



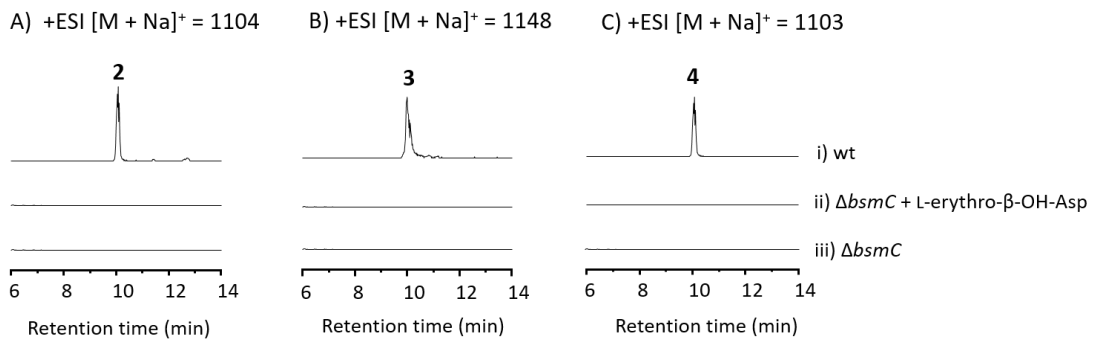
**Figure S6.** Chemical complementation of 5-OMe Tyr into mutants. LC-MS analyses of A) compound **2**, B) compound **3** and C) compound **4** for different mutant strains. i) wt; ii)  $\Delta bsmF$  mutant fed with 5-OMe Tyr; iii)  $\Delta bsmG$  mutant fed with 5-OMe Tyr; iv)  $\Delta bsmH$  mutant fed with 5-OMe Tyr; v)  $\Delta bsmH$  mutant; vi)  $\Delta bsmG$  mutant; vii)  $\Delta bsmF$  mutant.



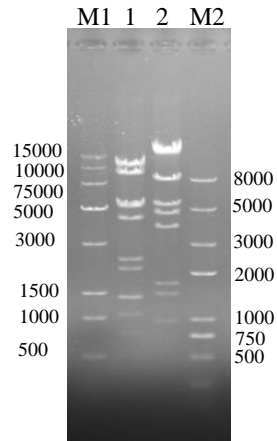
**Figure S7.** Chemical complementation of 5-OH Tyr into mutants. LC-MS analyses of A) compound **2**, B) compound **3** and C) compound **4** for different mutant strains. i) wt; ii)  $\Delta bsmF$  mutant fed with 5-OH Tyr; iii)  $\Delta bsmG$  mutant fed with 5-OH Tyr; iv)  $\Delta bsmH$  mutant fed with 5-OH Tyr; v)  $\Delta bsmH$  mutant; vi)  $\Delta bsmG$  mutant; vii)  $\Delta bsmF$  mutant.



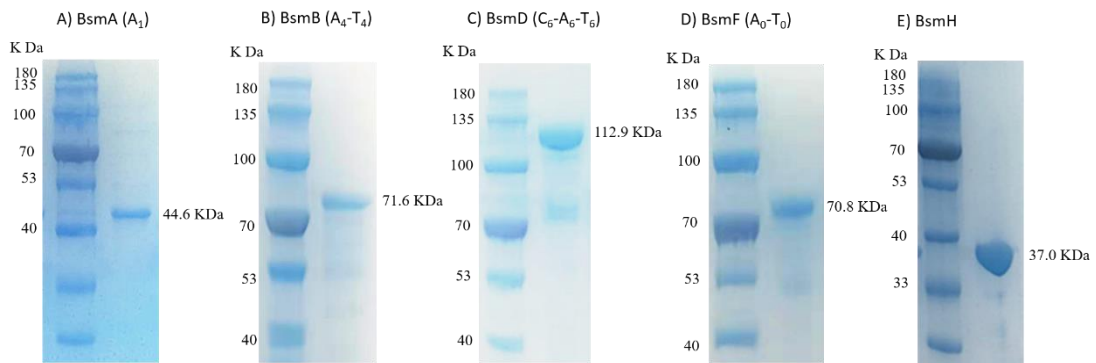
**Figure S8.** HPLC analysis of metabolite extracts from mutant strains and gene complementation strains. i) wild type; ii)  $\Delta bsmF$  mutant strain; iii)  $\Delta bsmG$  mutant strain; iv)  $\Delta bsmC$  mutant strain; v) complementation of  $\Delta bsmF$  mutant by *bsmF*; vi) complementation of  $\Delta bsmF-T281A$  mutant by *bsmF*; vii) complementation of  $\Delta bsmF-F281A$  mutant by *bsmF*. viii) complementation of  $\Delta bsmF-C281A$  mutant by *bsmF*; ix) complementation of  $\Delta bsmC$  mutant by *bsmC*; x) complementation of  $\Delta bsmG$  mutant by *bsmG*. 8 and 9 have identical retention time.



**Figure S9.** Chemical complementation fed L-erythro- $\beta$ -OH-Asp into  $\Delta bsmC$  mutant. LC-MS analyses of A) compound **2**, B) compound **3** and C) compound **4** for different strains. i) wt; ii)  $\Delta bsmC$  mutant fed L-erythro- $\beta$ -OH-Asp; iii)  $\Delta bsmC$  mutant.

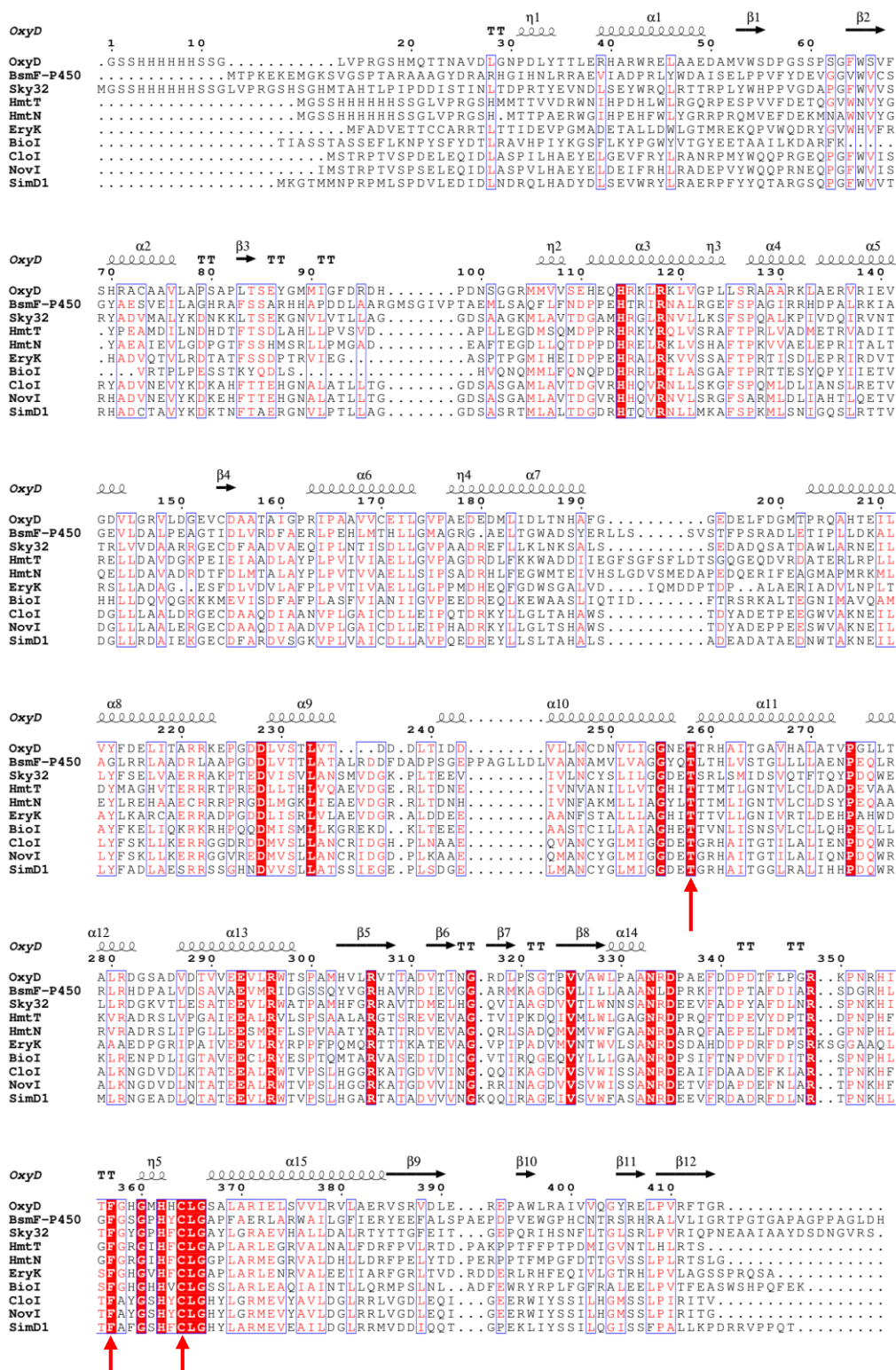


**Figure S10.** Verification of cosmid pHG06015. A) Physical map of pJTU2554 harboring the *bsm* gene cluster. B) agarose gel electrophoresis of pHG06015 cosmid. Legend: M1, 15K marker; 1, pHG06015 digested by BamHI; 2, pHG06011 digested by KpnI; M2, 8K marker



**Figure S11.** SDS-PAGE analysis of proteins. A) BsmA ( $A_1$ ) (calculated molecule weight: 44.6 KDa); B) BsmB ( $A_4-T_4$ ) (calculated molecule weight: 71.6 KDa); C) BsmD ( $C_6-A_6-T_6$ ) (calculated molecule weight: 112.9 KDa); D) BsmF ( $A_0-T_0$ ) (calculated molecule weight: 70.8 KDa). E) BsmH (calculated molecule weight: 37.0 KDa).





**Figure S12.** Sequence alignment of BsmF-P450 with other P450 proteins. The Glu/Thr residues are important in interactions with and proton transfer to iron-oxo intermediates in the P450 catalytic cycle. The heme-binding motif conserved Cys residues that act as the proximal ligand to the heme iron, the conserved phenylalanines as a regulator of heme iron potential.<sup>21,22</sup> Protein accession number: OxyD (3MGX\_A); Sky32 (4LOF\_A); HmtT (CBZ42154); HmtN (5XW2\_A); EryK (P48635.3); CloI (AAN65225); NovI (Q9L9F9); SimD1 (AAK06805); BioI (AGG62423.1). The multiple alignment was generated by ClustalW servers and rendered with ESPrnt 3.0.<sup>23</sup>

P450 domain  
1 10 20 30 40 50 60  
BsmF . . . . . MTPKEKEMGKSVGSP TARAAAGYDRAHG IHNLRRAEVIADPRLY WDAISELPPVFVYDEVGGVWCS  
OxyD . . . . . GSSHSHHSSSG . . . . . LVPRGSHMOTTNAV D LNPDLTYTLERHARWR ELAAEDAMVWSDPGSSPSGFWVSVF  
Sky32 . . . . . MGSSHHHHHSSGLVPRGSHSHTAHLPIPIDDISTIN LTPDRTYEVDLSEYWRQLRTTRPLWHPVPGDAPGFWSVVF  
HmtT . . . . . MGSSHHHHHSSGLVPRGSHSMMTTVVDRWNIHPDHLWLRGQRPESPVVFDVETQGVWVYGF  
HmtN . . . . . MGSSHHHHHSSGLVPRGSHSHTTPAERWGIHPEHFWLYGRRPRQMFVEDEKMNWVYGF  
EryK . . . . . MFADVETTCARRTITIDEVPGMADETTALLD LVTGMREKQPVVQDQRYGVWVFR  
EryF . . . . . ATVPDLESDFHVDWYSTYAE IRETAPVTVPVFLGQDAWLVTYGDEAKAALSDDLRLSSDDPKK  
BioI . . . . . TIASSTASSEFLKNPYSFYDT LRAVHPYIKGSFLLKYPGWYVTGYEETAAILKDFK . . . . .  
CloI . . . . . MSTRPTVSPDELEQIDL LASPILHAEYELGEVFRYL RANRPMYWQQPRGEPGFWVVIS  
NovI . . . . . IMSTRPTVSPSELEQIDL LASPVLHAEYELDEIFRLLRADEPVPYQQPRNEQPGFWVVIS  
SimD1 . . . . . MKGTMNNRPMLSPDVLDEDI LNDRQLHADYDLSEVWRYLRAERFFYQTAGRSQFQVWVT

P450 domain  
70 80 90 100 110 120 130 140  
BsmF GYAESVETLGLGHRAFSSARHAFDDLAARGMSGIVPTAEMLSAQFLFNDPPETRTIRNLRGELFSPAGTRHHPFALRRIA  
OxyD SHRACAAVLAAPSAPLTSYGMMLGFDRD . . . . . PDNSGGRMVSEHEQERKLRKLVGLLRAAAARKLAERVRIE  
Sky32 RYADVMAAARRGECDFEADVAEQIPLNTIISDILGVPAADEFFLLKLNKSA L . . . . . SEDADQSDAIDLARNEIL  
HmtT . YPEAMDLNLDHDTFTSDLAHLLPVSD . . . . . AP LLEGMSONMPDRRKYRQLVSRAFTPLRVADMETRVADIT  
HmtN . YAEAIEVLGDPGTFSSHMSTRLLPMGAD . . . . . EAFTEGLLQTDPPDRRELRKLVSHAFTPKVVVAELEPRITALT  
EryK . HADVQTVLRDTATFSSDPTRVTEG . . . . . ASPTPGMIHEIDPPETRRRKLVSQSAFTPTRTISDLEPRTRDVT  
EryF . HELLDEVEFFPAGLGFPEVDR . . . . . NYFATNMGTSDPPETTRRKRKLVQSAFTVRRVEAMRFRVEQIT  
BioI . . . . . VRTFLPESSSTKYQDLS . . . . . HVQNMMLFQNQPDRRRLTLAGSAGFTPTRTTESYQPYIIEV  
CloI RYADVNEVYKDKAHFTTEHGNALATLLTG . . . . . GDSASGMLAVTDGVRHQVRLNLSKGFSPQMLDLIANSLRETV  
NovI RHADVNEVYKDKAHFTTEHGNALATLLTG . . . . . GDSASGMLAVTDGVRHQVRLNLSRGSFARMDDLIAHTLQETV  
SimD1 RHADVNEVYKDKAHFTTEHGNALATLLTG . . . . . GDSASRTMLAVTDGDRHQVRLNLSKAFSEKMLSNIGQSLRTTV

P450 domain  
150 160 170 180 190 200 210 220  
BsmF GEVLIDAEPEAGTIDLVRDFARLEFLHMLTHLLGMAGRGAEELTGWADSV ERLLS . . . . . SVS TFFPSRABLETIFLDLKA  
OxyD GDVLGRLVLDGEVCDAAITAIGFRIPAAVVCEIILGVPAADEEPLIDLNTNHA FG . . . . . GEDELDFDGMTPRQAHTEIL  
Sky32 TRLVVDAARRGECDFEADVAEQIPLNTIISDILGVPAADEFFLLKLNKSA L . . . . . SEDADQSDAIDLARNEIL  
HmtT RELLDVAVDGKPEIEIAADLAPLPVTVVAELILGVPAGDRDLFKKWAADDIEGFSGSFLDTS GGQEQQVDRDATELRLL  
HmtN QELLDVAVDKPTIEMTALAYLPVTVVAELILGVPAGDRHLFEWGMMTELVHSLGDVSMEDAPEDEQRIEFAGMAPMKRML  
EryK RSLLDVADAG . . . E SFLVLDVDFAPPVTVVAELILGVPAGDRHQFGDWSGALVD . . . . . IQMDDPTDP . . . . . ALAERADVLNPLT  
EryF AEELLDVEVGDSSGVVDVDRFAHPLEIKVICELILGVPAADEEARFAGFRWSS . EILVMD . . . . . PERAE . . . . . QRQAAREVV  
BioI HHLLDQVQGGKKMEVVIDFAFPVLSFLVIANIIGVPEE DRQLKEWAASLIQTID . . . . . FTRSRKALLTEGNIIMAVQAM  
CloI DGLLLAALDRGECDAADQIAANVPLGAI CDLLEIPOTDRKYLGLTAHAWSS . . . . . TQYADETPEEGWVAKNEIL  
NovI DGLLAALDRGECDAADQIAANVPLGAI CDLLEIPOTDRKYLGLTAHAWSS . . . . . TQYADEPPEESVWAKNEIL  
SimD1 DGLLRDAIEKGECDFAIRDVSGKVFVLVVICDLLEAVQEDREYLLSLTAH . . . . . ALS . . . . . ADEADATAEDNWTAKNEIL

P450 domain  
230 240 250 260 270 280 290 300  
BsmF AGLRRLAADRTAAFGDDLVTTLATAALRDDFDAPSGEPFAGLLDLVAANAMVLVAGCYQTLTLVSTGLLQI AENPEQLR  
OxyD VFDELITARRRKEPGDDLVSTLVT . . . . . DD . . . . . DLTIDD . . . . . VLNCDNVLIIGNE TTRHAITGAHALATVPGLLT  
Sky32 LYFSELVAEERRAKPTEDVISVLANSMVSGK . PLTEEV . . . . . I VLNCSYL LIGGDETSRLSLIDSVQITQYDQWE  
HmtT DVMAGHVTERRRTRREDLTLLVQAEVGE . RLTDNE . . . . . I VNVANILLVVTGHITTTMTLNGNTLCLDADPEVAA  
HmtN EYLRHMAECRRRPGDMLGKLEAEVDGR . ALDDNE . . . . . I VNFAKMLL LAGCYLTTMTLNGNTVLCDSYEPHQA  
EryK AYLKAQCAERRADPGDDLSRLVLAEVDRGR . ALDDNE . . . . . AANFSTALLLAGGHTTTVLLGNTVRTLDEHFAWD  
EryF NFI L DLVERRRTEPGDDLSALYSVQDDDRGLSADE . . . . . L TSIALVLL LAGAEASVSLIGITGYLLQLHPEQLA  
BioI AYF K E L L K E R R R G G D R P Q D M S M L K G R E K D . . . . . K L T E E . . . . . A A S T C I L L A I A G H E T V N L S N S V L L Q H E P G E L L  
CloI LYFSKLLKERKGGDRPQDDMSMLKGREK . . . . . KLTEE . . . . . AASTCILLAIAGHETVNLNSVLLQLHPEQGLL  
NovI LYF K L L K E R R G G V R E D M V S L A N C R I D G D . P L K A A E . . . . . Q V A N C Y G L M I G G E T G R H A I T G T L A L I E N P D Q W R  
SimD1 LYFADLAE SRRSSGHNDVVS L L A T S S I T G E . P L S D G E . . . . . L M A N C Y G L M I G G E T G R H A I T G T L A L I E N P D Q W R

P450 domain  
310 320 330 340 350 360 370  
BsmF RLRLHDPALVDVSAVAVEMRLDGSQYVGRHVARDIEVGC ARMKAGDGVLLILLAANLDPKRFITDPTAFDILR . . . SDGRHL  
OxyD ALRDGGSADVDTVVEVLRNTS PAMHVI RVTTADVTIN . . . . . RD LP SGT P V V A W L P A A N R D P A E F D P D T F L P G R . . . . . K P N R H I  
Sky32 LLR DG K V T L E S A T E V L R W A T P A M H F G R A V T D M E L H G . Q V I A A G D V L L W N N S A N R D E E V F A D P Y A F D L N R . . . . . S P N K H I  
HmtT KVRADRSLVPGAIEBALRVLSPSAALARGTSREVEVAG . TVIPKQIVMLWLGAGNRDPROFTDPEVYDFPTR . . . . . D P W P H F  
HmtN RVRADRSLLIPGLLESMRRLSPVAATYRATTDRDVEVAG . QRLSADQMVMVWFAGANR D A R O F A E P E L F D M T R K . . . . . G P W P H L  
EryK AA A E D P G R I P A I V E V L R V R P P F P Q M O R T T T K A T E V A G . V P I P A D V M V N T V W L S A N R D S D A H D P D R F D P S R K S G G A A Q L V  
EryF LVRADP S A L P N A V E C L R Y E S P Q M T A R V A S E D I D C G . V A I P O Y S T V L V A N G A A N R D P S O F P D P H R F V D V T R . . . . . D T R G H L  
BioI KLRENPDLIGTAVBECLRYE SPTQMTARVASEDIDICG . VTIRQEQVYLLLGAANR D P S I F T N P D V F D I T R . . . . . T P N K H F  
CloI ALKNGVDL LNTATBALRMTVPSLHGGKATGDVVING . QRIRKAGDVVS V W I S S A N R D E A I F D A A D E F K L A R . . . . . T P N K H F  
NovI ALKNGVDL LNTATBALRMTVPSLHGGKATGDVVING . GRIRINAGDVVS V W I S S A N R D E T V F D A P D E F N L A R . . . . . T P N K H F  
SimD1 M L R N G E A D L O T A T E V L R N T V P S L H G A R T A T A D V V I N G K Q Q I R A G E I V S V W F A S A N R D E E V F R A D A R F D I L N R . . . . . T P N K H L

P450 domain linker region  
380 390 400 410 420 430 440 450  
BsmF GFSGSPHHCAGFAERLARVA L G F I E R Y E E F A L S P A E P D P V E W G P H C N T R S R H R A L V L I G R T P G T A P A G P P A G L D H  
OxyD TFGHGMHCAGSALRIELS VVLRVLAERVSRV D L E . . . . . R E P A W L R A I V V Q Y R E L P V R F T G R . . . . . G A P P A G L D H  
Sky32 TFGYGFHFCGAYLGRAEVA L L D A L R T Y T T G F E I T . . . . . G E P Q R I H S N F L T G L S R L P V R I Q P N E A A I A Y D S D N G V R S .  
HmtT TFGHGMHCAGLARLEGRVALNALDFRFFVLRD . PAKP PTFPTPD M I G V N T L H L R T S . . . . . P A P P A G L D H  
HmtN GFGRGHFCGGLPARMEGRVALHLLDRFFEL Y T D . P E R P P T F M P G F D T T G V S S L P L R T S L G . . . . . P A P P A G L D H  
EryK SPFGHGMHCAGLARLEGRVALEEIARFGRLTV D . R D D E R L R H F E I Q V L T R H L P V L A G S S P R Q S A . . . . . P A P P A G L D H  
EryF SPFGHGMHCAGSRLAKLEGEVALRALFGRFPAL S L G I D A D D V V W R R S L L L R G I D H L P V R L D G . . . . . P A P P A G L D H  
BioI SPFGHGMHCAGSRLARLEAQIAINTLQRMPSLNL . . . . . A D F E W R Y R P L F G F R A L E E L P V T F E A S W S H P Q E K . . . . . P A P P A G L D H  
CloI TFAYGSHYCGHYLGRMEVYAVLDGLRRLVGDLEQI . . . . . G E R W I Y S S I L H G M S S L P I R I T V . . . . . P A P P A G L D H  
NovI TFAYGSHYCGHYLGRMEVYAVLDGLRRLVGDLEQI . . . . . G E R W I Y S S I L H G M S S L P I R I T V . . . . . P A P P A G L D H  
SimD1 TFAFGSHYCGHYLGRMEVYAVLDGLRRLVGDLEQI . . . . . G E K L Y S S I L O G I S S F A L L K P D R R V P P Q T . . . . . P A P P A G L D H

linker regions

460 470 480 490 500 510 520 530  
 BsmF GAEVGVAETCAAEVTTAAEADVPEAKALSDLERHQVVAEWNDDGTAGTAAQICWQLVFEERARLAPDSTAVEDEGVTYS  
 BsmA-A1 .....S  
 BsmA-A2 .....T  
 BsmB-A1 .....S  
 BsmB-A2 .....T  
 BsmB-A3 .....S  
 BsmC-A1 .....S  
 BsmC-A2 .....S  
 BsmC-A3 .....T

A domain

540 550 560 570 580 590 600 610  
 BsmF YAEIDAMANGTAHRLRDIQVGEETVVAISNERSVRFIVAILAVNKAAGAFLLAAADAPSERLKTMLEQASARLVITADNAT  
 BsmA-A1 YGELNARANRLARLLVDRGVGFEQVVALALPDRSPELVVAMLAWLKRGAANTBIDITYPADRIRFMIEDARFTLVLTHTTA  
 BsmA-A2 YEDLNARANRLARLLTERGVGFEHIVALALPDRSPELITLAWLKRGAANTBIDITYPADRIRFMIEDARFTLVLTHTTID  
 BsmB-A1 YGELNARANRLARLLVDRGVGFEQVVALGLPDRSADVIVAVLAWWKRGAANTBIDITSYVDRIRFMVEDARFALVLTHTTA  
 BsmB-A2 YEDLNARANRLARLLTERGVGFEHIVALALPDRSPELITLAWLKRGAANTBIDITYPADRIRFMIEDARFTLVLTHTTA  
 BsmB-A3 YGELDARARRLARLLAERGVGAEHIVALALPDRSPELVVALLAWLKRGAANTBIDITGYPVERIRFMVGDARFTLVLTHTGT  
 BsmC-A1 YAEINARANRLARLLVDRGVGAEQVVALALPDRSADVIIVAVLAWWKRGAANTBIDITYPADRIRFMVEDARFALVLAHSSA  
 BsmC-A2 YGELNARANRLARLLVDRGVGFEQVVALGLPDRSADVIVAVLAWWKRGAANTBIDITSYVDRIRFMVEDARFALVLTHTAT  
 BsmC-A3 YEDLNARANRLARLLTERGVGFEHIVALALPDRSPELVVALLAWLKRGAANTBIDITGYPVERIRFMVEDARFTLVLTDTAA

A domain

620 630 640 650 660 670 680  
 BsmF AARTI...EPFAFPAEVVRDLTARAPAPP...ITGVLPNTAYVVFVTSGTGRPKGIANSHEALTNLHIAQRKVF  
 BsmA-A1 AGLW...QDSTPTLLDDDTATQRELASFDTADPTDADRSPLEDPHPVYVIYTSGSTGVFKGVVTHHTGLINLALAQSDQW  
 BsmA-A2 AGLW...QDSTPTLLDDDTATQRELASFDTADLATAP...DPHPVYVIYTSGSTGVFKGVVTHHTGLINLALAQSDQW  
 BsmB-A1 AGLW...EDGTATVFLDDDAVHAEALAGFEAADLTDADRSPLEDSACPAYVIYTSGSTGVFKGVVTHHTGLINLALAQSDQW  
 BsmB-A2 AGLW...QDSTPTLLDDDTATQRELASFDTADPTDADRSPLEDPHPVYVIYTSGSTGVFKGVVTHHTGLINLALAQSDQW  
 BsmB-A3 AGLW...TEGTPVHLLDDDAVGRLLAGFEESVDP...VAVPDPHPVYVIYTSGSTGVFKGVVTHHTGLINLALAQSDQW  
 BsmC-A1 GDLVPEQDGTSTVVLDDPAVQESLGLDASDLTDADRVTALDLAHPAYVIYTSGSTGVFKGIAPPVHGLANFALAQSKQW  
 BsmC-A2 AGLW...VDGAATVFLDDDAVQAEALAGFGADDLA...TVLDSACPAYVIYTSGSTGVFKGVVTHHTGLINLALAQSDQW  
 BsmC-A3 SASW...ADGIPTLCPDDPAVQEARLRHEATDP...VAPFDLADHPAYVIYTSGSTGVFKGVVTHHTGLINLALAQSDQW

A domain

690 700 710 720 730 740 750 760  
 BsmF RIRPRDRVLCFLSPNFDGCSISVVLALLCGATLVVARTAQITPGLPPLRLRERRISAAIMTSVWVSLPGDP.LPDLFPD  
 BsmA-A1 HIGPGSRVLCFAFSPSFDAAASEVFTLLIGGTTVATTDLTPGDALHTLTDTAITHTCLPSSALSVLDTT...IPP  
 BsmA-A2 HIGPGSRVLCFAFSPSFDAAASEVFTLLIGGTTVATTDLTPGEALHTLTDTAITHTCLPSSALSVLDTT...IPP  
 BsmB-A1 VGRGGRVLCFAFSPSFDAAATWWSLALLSGAALVVAAGAEELAPGAALMGVLDGAGVYTCMVPVSVLPLLDVGR...VPA  
 BsmB-A2 .TGAOGVALLHTFSPSFDASITALLFTPLVSGRVLVAELR...DGLGDEETRSLGCDFLKATPSHLALLLEGLP.EEVSP  
 BsmB-A3 GLGADDVLCVKTFAFSDVSVVWFFWLLQGASVVLAKPDGHDKARYLAEELVSEGVTTAHFVPSMLDAFLGEPAGRCAS  
 BsmC-A1 GIGAGSRVLCFLSPSFDAAASEFFTALLIGGABVLADADRLTPGPDLELLKNEGITHTCLVSVLSEEG...LPS  
 BsmC-A2 VGRGGRVLCFAFSPSFDAAATWWSLALLSGAALVVAAGAEELAPGAALMGVLDGAGVYTCMVPVSVLPLLDVGR...VPA  
 BsmC-A3 LREGRERLPSVTLIAFDLAGEIYLYLPCGAGVLPVPGATVANDFLAMAGLADTGVTVVQATPSLWRELAASGAQGLG

A domain

770 780 790 800 810 820 830  
 BsmF LRIAFAFGERLPGLDVRFWTAFGR...RLNLYGPAEAAVWAWHFCAG...EDAPP.TGRPVA GKRYVMDDEHRLLP  
 BsmA-A1 AMTLIVA GSTPDTIQRWSTGR...TMINAYGPEETTVCAAMSEPLS...GEVVPV.TGRPNDNVRTYV.DQNLSFVPP  
 BsmA-A2 AMTLIVA GSTPDTIQRWSTGR...TMINAYGPEETTVCAAMSEPLS...GAAAPP.TGRPNDNVRTYV.DQNLSFVPP  
 BsmB-A1 SVTVVVGGAACGPDVAGRWSAGR...MMFNAYGPEETTVCAAMSEPLS...GEVVPV.TGRPNDNVRTYV.DQNLSFVPP  
 BsmB-A2 RKELLVGGALLGALRTWRA GHPDVIVLNVYGPTEATVNCSEFRIEPDELDPGFPV.TGRPNDNVRTYV.DAGLRFPVPP  
 BsmB-A3 LRRVFCSSGVLPALVTRFHSVLPVAVLHNLVYGPTEATVDFVHFPCADGDPRIAPP.TGRPNDNVRTYV.DAGLRFPVPP  
 BsmC-A1 TLTLLVGNCCPEIVERWSAGR...RMFNAYGPEETTVCAAMSEPLS...GAVVPP.TGRPNDNVRTYV.DGGLCFVPP  
 BsmC-A2 SVTVVVGGAACGPDVAGRWSAGR...MMFNAYGPEETTVCAAMSEPLS...GEVVPV.TGRPNDNVRTYV.DQNLSFVPP  
 BsmC-A3 LRRVLVGGAVSAALETLRGLGRSVTLNVYGPTEETLWSTAADLDGAGDG.AAPS.TGRPNDNVRTYV.DAGLRFPVPP

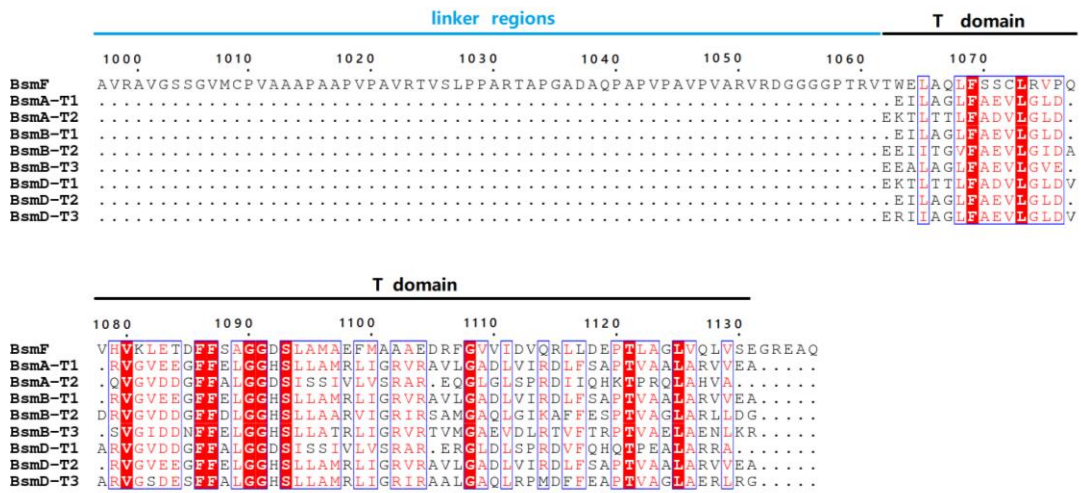
A domain

840 850 860 870 880 890 900 910  
 BsmF GSPGELYIAGTGVGR.VLGRPLMAGSFVDFFAERPCRLMYRTGDLVWREDCSLEIYVGRRROVKTRGQVVELDELER  
 BsmA-A1 GVPGELYVAGAGVARGYLNRRPGLTSERFVADDPYGP.AGSRMYRTGDLARWNSDGLHFIIGRABDQVKTRGRIELGEVEA  
 BsmA-A2 RVPGELYVAGSGVARGYLNRRPGLTSERFVADDPYGP.PGSRMYRTGDLARWNSDGLHFIIGRABDQVKTRGRIELGEVEA  
 BsmB-A1 GVPGELYVSGTGVARGYLNRRPGLTSERFVADDPYGP.AGSRMYRTGDLARWNSDGLHFIIGRABDQVKTRGRIELGEVEA  
 BsmB-A2 GVPGELYVAGVCLARGYLNRRPFTGQRFVADDPYGP.AGSRMYRTGDLARWNSDGLLVFIIGRABDQVKTRGRIELGEIEA  
 BsmB-A3 GVPGELYLAGAGLARGYLNRRPGLTAEERFADDPYGP.PGARMYRTGDLASWNSDGLRYIIGRABDQVKTRGRIELGEIEA  
 BsmC-A1 GVPGELYVSGAGVARGYLNRRPRLTAEERFVADDPYGP.AGSRMYRTGDLVWNSDGLQFIIGRABDQVKTRGRIELGEIDA  
 BsmC-A2 GVPGELYVAGAGVARGYLNRRPGLTSERFVADDPYGP.AGSRMYRTGDLARWNSDGLHFIIGRABDQVKTRGRIELGEVEA  
 BsmC-A3 GVS GELYIAGAGLARGYLNRRPGLTAEERFADDPYGP.AGSRMYRTGDLARWNSDGLRIFIGRABDQVKTRGRIELGEVEA

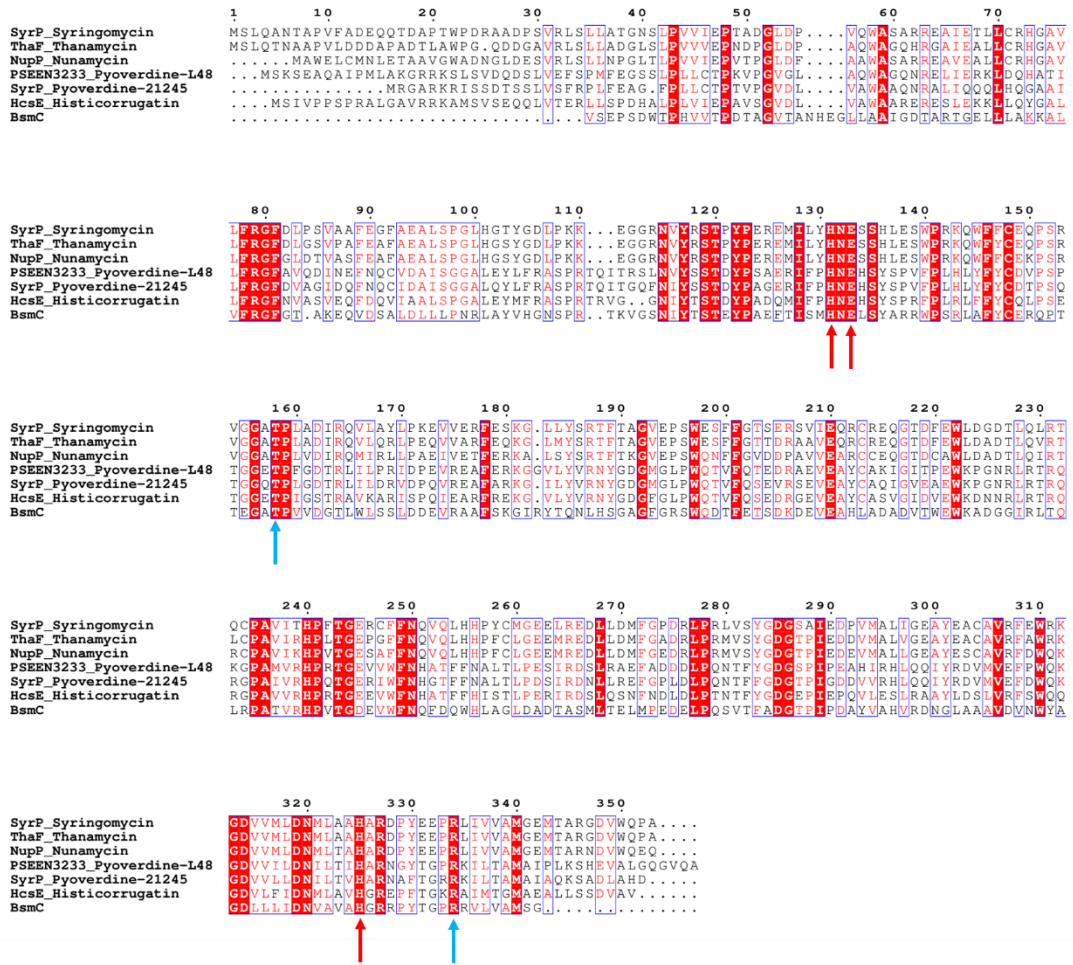
A domain

linker regions

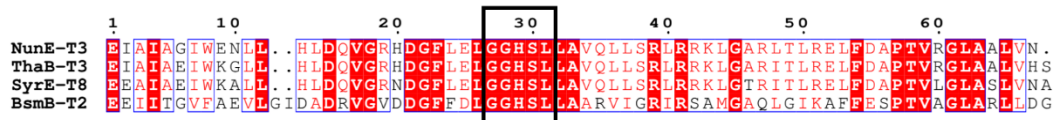
920 930 940 950 960 970 980 990  
 BsmF VLEKAPGVASCHAVERNRNLHALVVRDPGSDWEPEVRRHLAERLHGGMIPATFTVMDQLPLTNSGKVGQEHGTAERNGG  
 BsmA-A1 VLAACPGVAAAV.....  
 BsmA-A2 ALAACPGVAAAA.....  
 BsmB-A1 VLAACPGVAAAV.....  
 BsmB-A2 ALTGCAQVRAAA.....  
 BsmB-A3 ALAACPGVTAAGV.....  
 BsmC-A1 VLAACPGVASAA.....  
 BsmC-A2 VLAACPGVAAAV.....  
 BsmC-A3 ALAACPGVESAAV.....



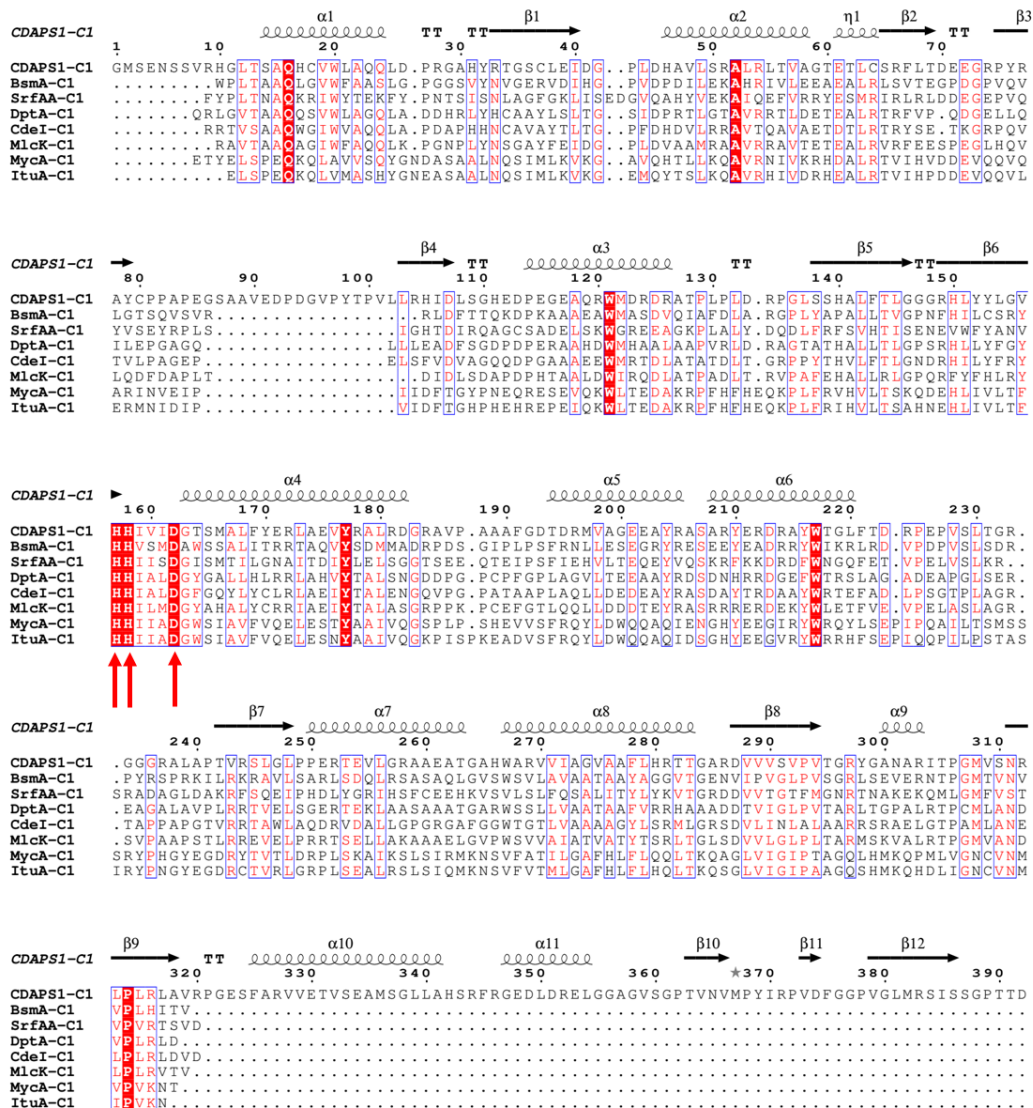
**Figure S13.** Domain analyze of BsmF. P450 domain (27-436 aa); A domain (537-930 aa); T domain (1062-1131 aa); linker regions (437-536 aa and 931-1061 aa).



**Figure S14.** Sequence alignment of BsmC with other  $\beta$ -hydroxylases. Fe(II) and  $\alpha$ -ketoglutarate binding residues are marked with red arrows and blue arrows.<sup>24</sup> Protein accession number: SyrP\_Syringomycin (AKF46133.1); ThaF\_Thanamycin (ALG65284.1); NupP\_Nunamycin (KPN90375.1); PSEEN3233\_Pyoverdine-L48 (WP\_011534378.1); SyrP\_Pyoverdine-21245 (AJW67533.1); HcsE\_Histicorrugatin (WP\_053122094.1). The multiple alignment was generated by ClustalW servers and rendered with ESPrift 3.0.<sup>23</sup>



**Figure S15.** Sequence alignment of BsmB-T<sub>2</sub> with other T domains. Strictly conserved GGHSL motif in the thiolation domain are marked with black box.<sup>24</sup> Protein accession number: NupE-T<sub>3</sub> (KPN90369.1); SyrE-T<sub>8</sub> (AAY37647.1); ThaB-T<sub>3</sub> (AED90003.1). The multiple alignment was generated by ClustalW servers and rendered with ESPrnt 3.0.<sup>23</sup>



**Figure S16.** Sequence alignment of BsmA-C1 with other C domains. The conserved histidine and aspartate residues are marked by red arrows.<sup>25</sup> Protein accession number: CDAPS1-C1 (CAB38517); SrfAA-C1 (CAE02630); DptA-C1 (AHX36919); CdeI-C1 (QBC75021); MlcK-C1 (ARU08073); MycA-C1 (Q9R9J1); ItuA-C1 (BAB69698). The multiple alignment was generated by ClustalW servers and rendered with ESPrnt 3.0.<sup>23</sup>

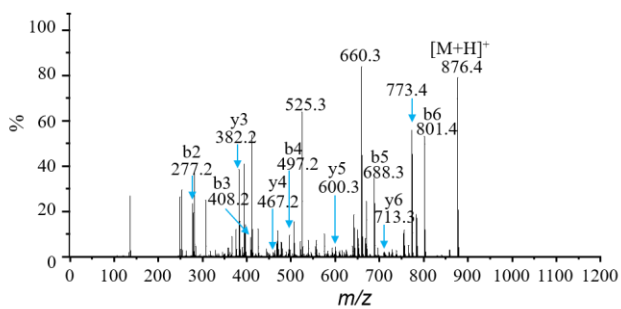
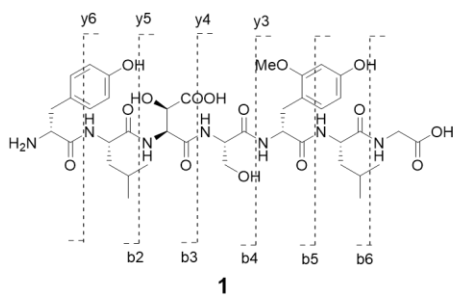


Figure S17. MS/MS analysis of 1.

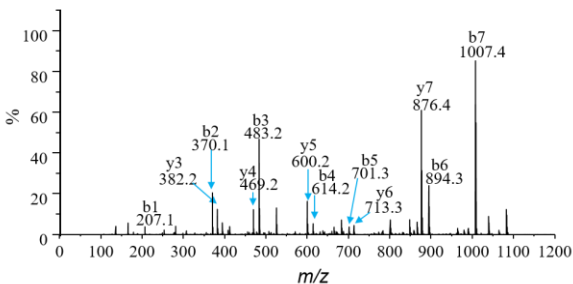
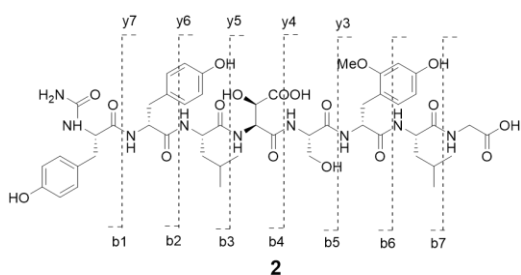


Figure S18. MS/MS analysis of 2.

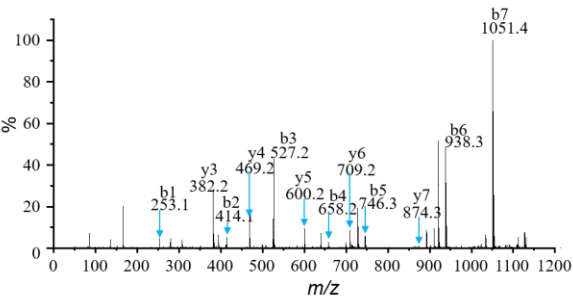
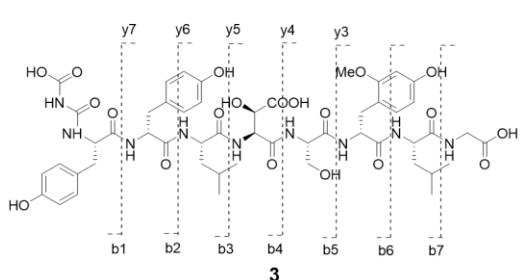


Figure S19. MS/MS analysis of 3.

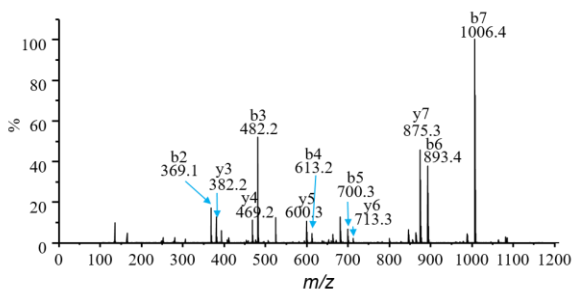
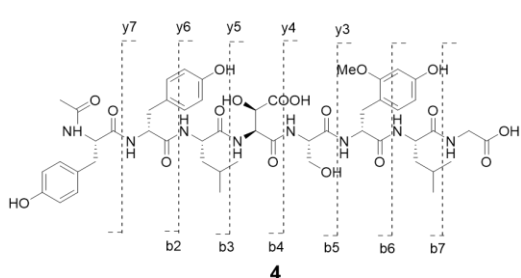


Figure S20. MS/MS analysis of 4.

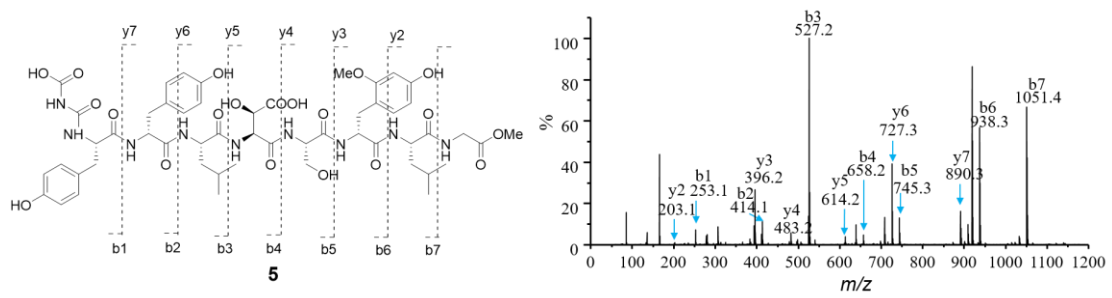


Figure S21. MS/MS analysis of 5

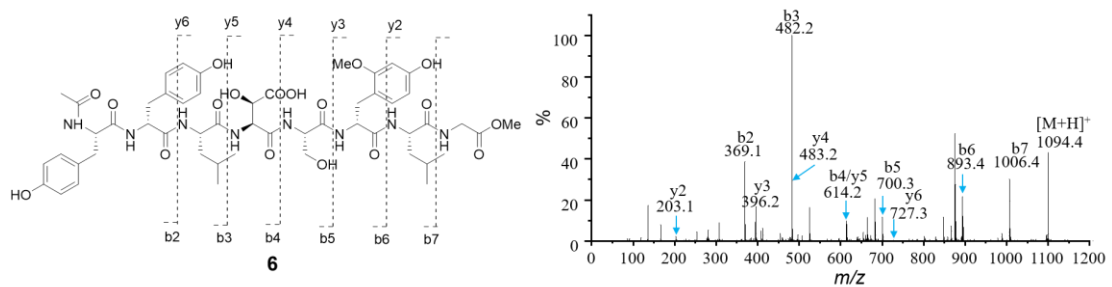


Figure S22. MS/MS analysis of 6.

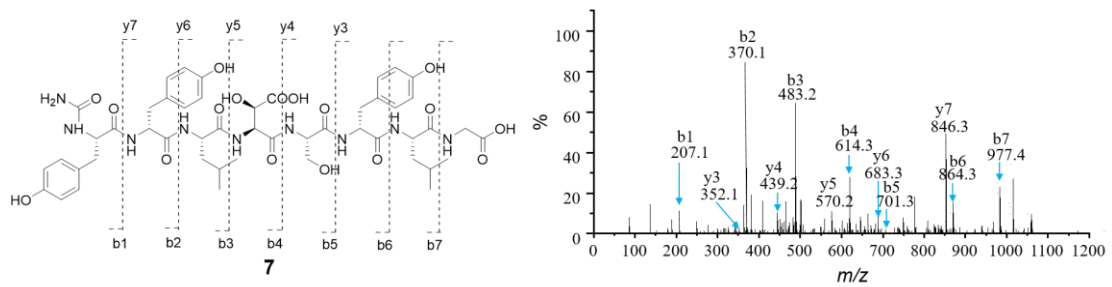


Figure S23. MS/MS analysis of 7 in *ΔbsmF* mutant strain.

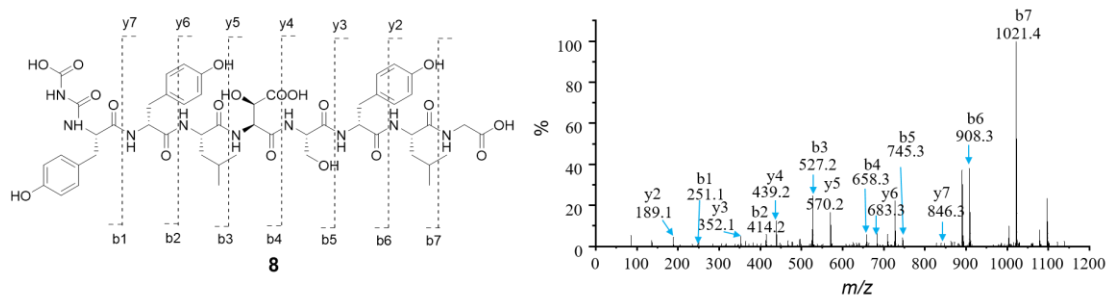


Figure S24. MS/MS analysis of 8 in *ΔbsmF* mutant strain.



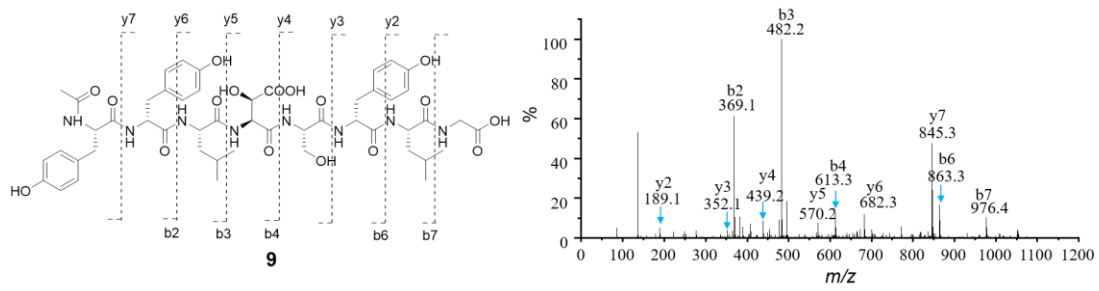


Figure S25. MS/MS analysis of **9** in  $\Delta bsmF$  mutant strain.

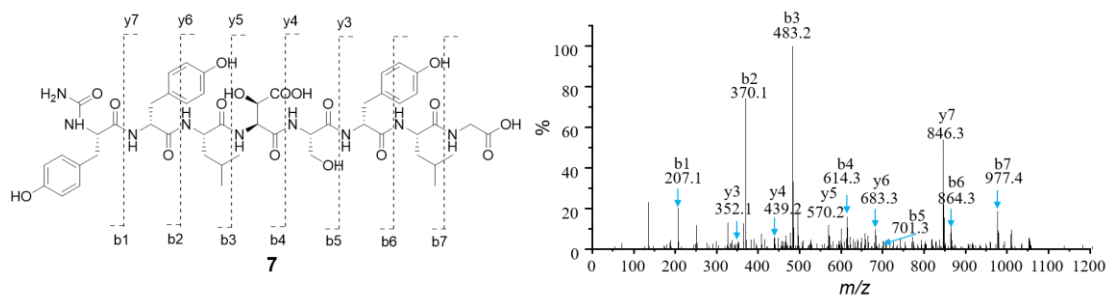


Figure S26. MS/MS analysis of **7** in  $\Delta bsmG$  mutant strain.

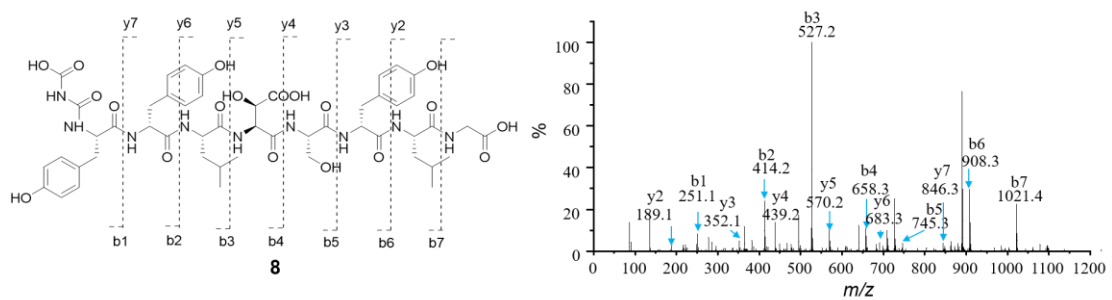


Figure S27. MS/MS analysis of **8** in  $\Delta bsmG$  mutant strain.

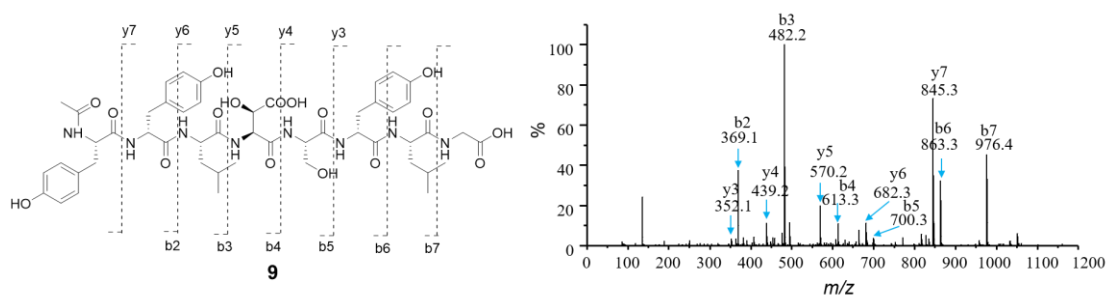


Figure S28. MS/MS analysis of **9** in  $\Delta bsmG$  mutant strain.

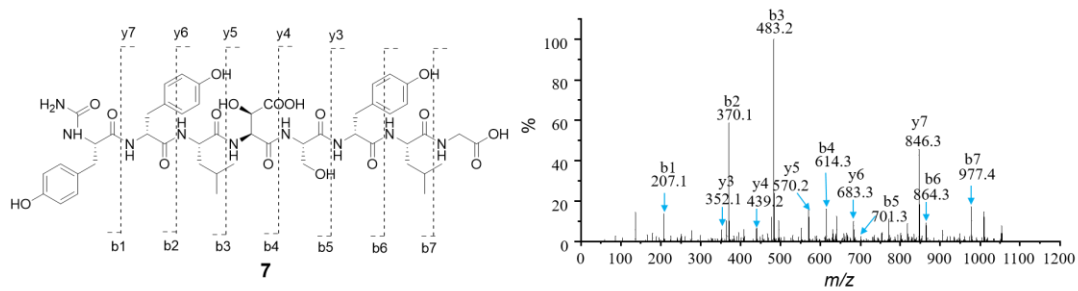


Figure S29. MS/MS analysis of 7 in  $\Delta bsmH$  mutant strain.

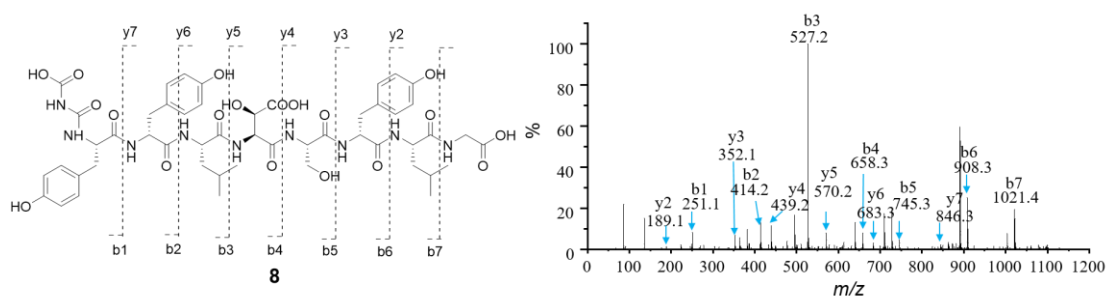


Figure S30. MS/MS analysis of 8 in  $\Delta bsmH$  mutant strain.

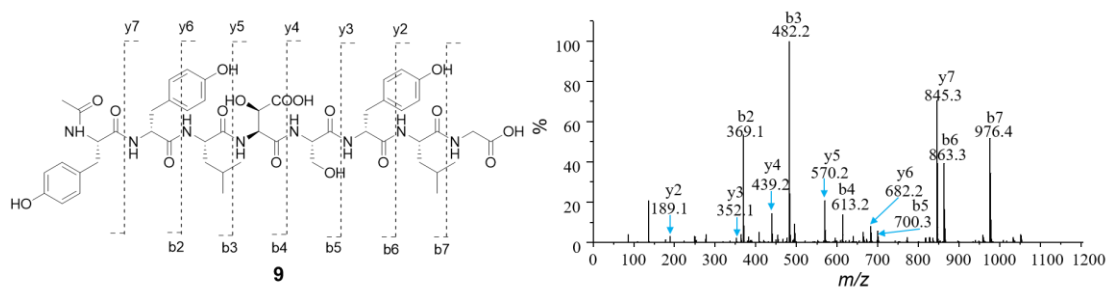


Figure S31. MS/MS analysis of 9 in  $\Delta bsmG$  mutant strain.

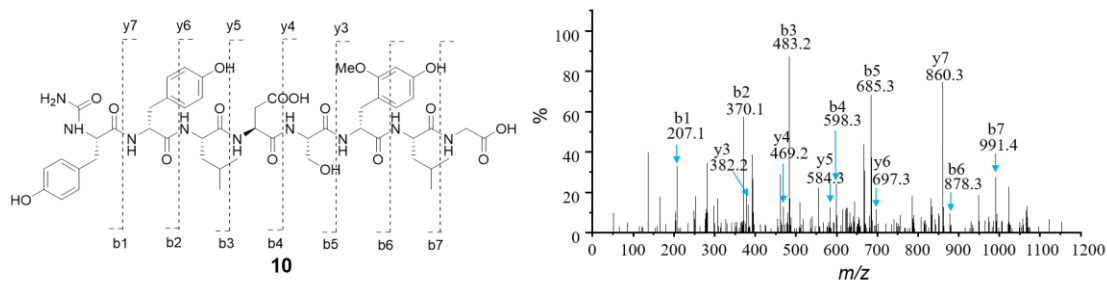


Figure S32. MS/MS analysis of 10 in  $\Delta bsmC$  mutant strain.

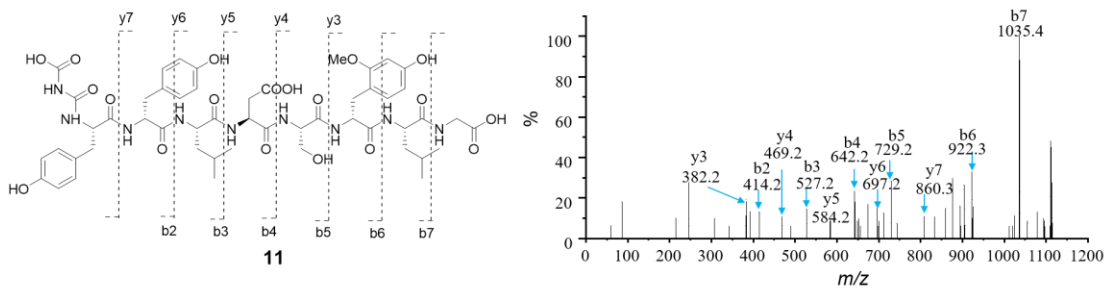


Figure S33. MS/MS analysis of **11** in  $\Delta bsmC$  mutant strain.

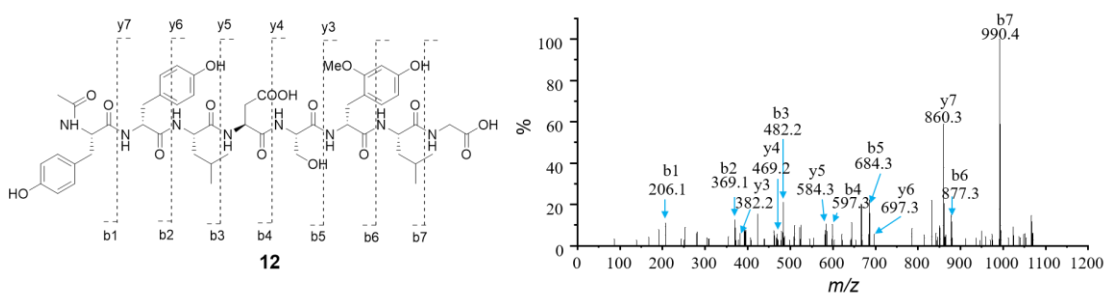


Figure S34. MS/MS analysis of **12** in  $\Delta bsmC$  mutant strain.

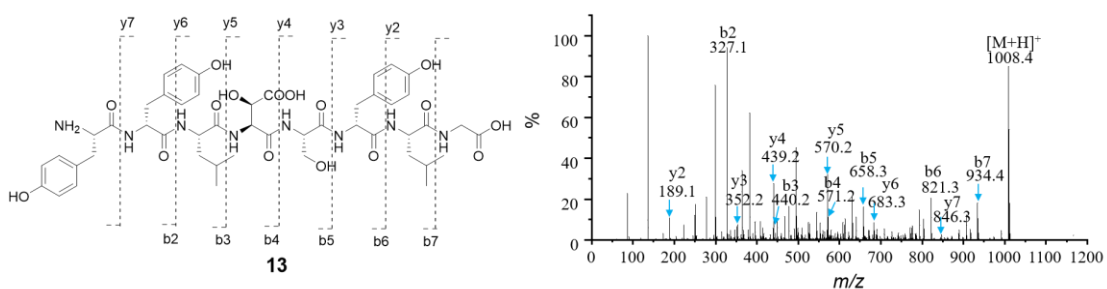


Figure S35. MS/MS analysis of **13**.

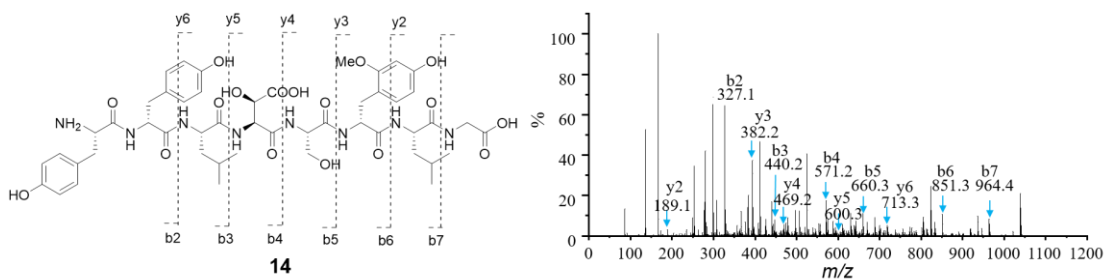


Figure S36. MS/MS analysis of **14**.

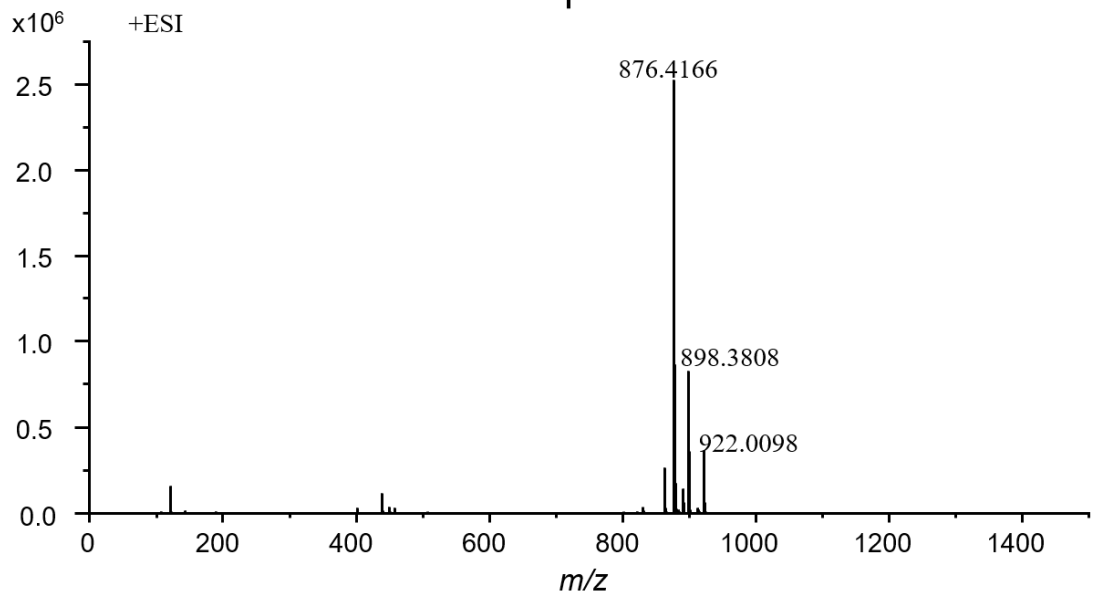
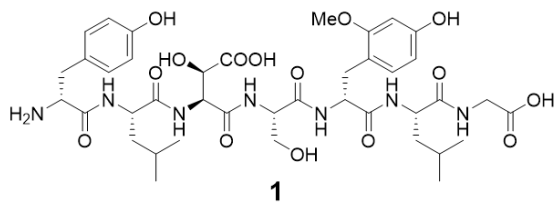


Figure S37. HRESIMS spectrum of **1**.

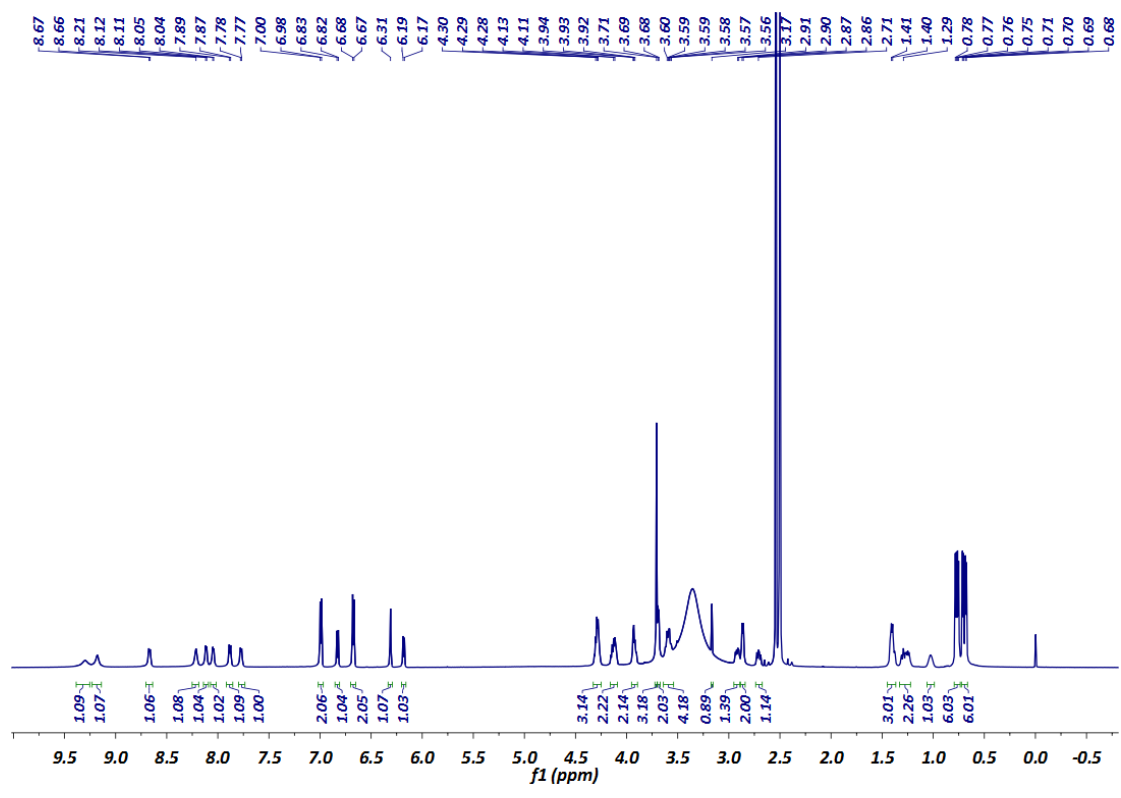


Figure S38. <sup>1</sup>H NMR (600 MHz, DMSO-*d*<sub>6</sub>) spectrum of compound **1**.

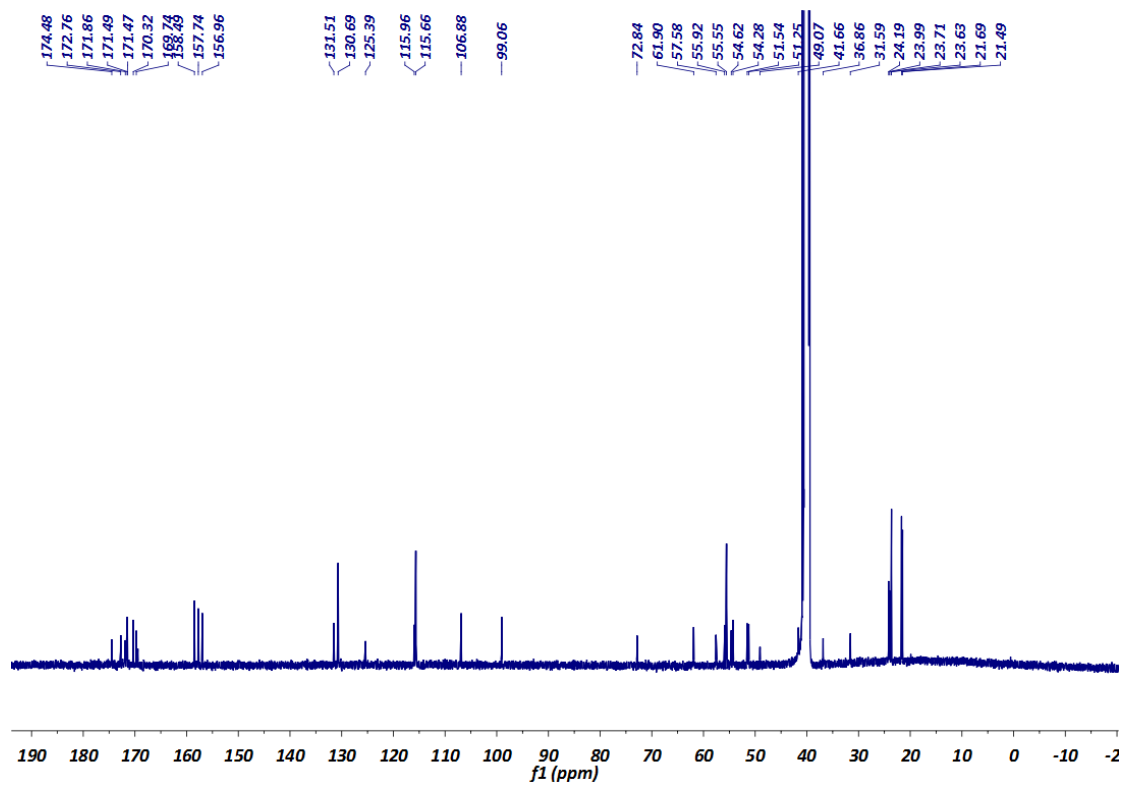


Figure S39.  $^{13}\text{C}$  NMR (150 MHz,  $\text{DMSO}-d_6$ ) spectrum of compound 1.

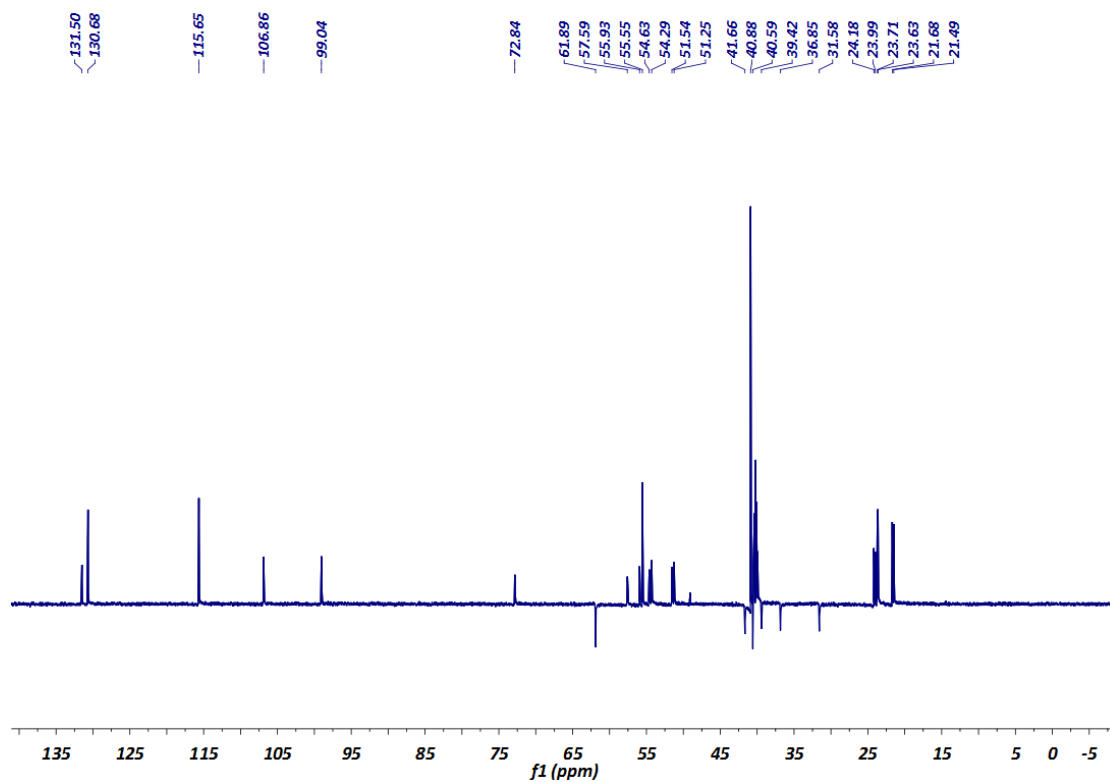


Figure S40. DEPT NMR (150 MHz,  $\text{DMSO}-d_6$ ) spectrum of compound 1.

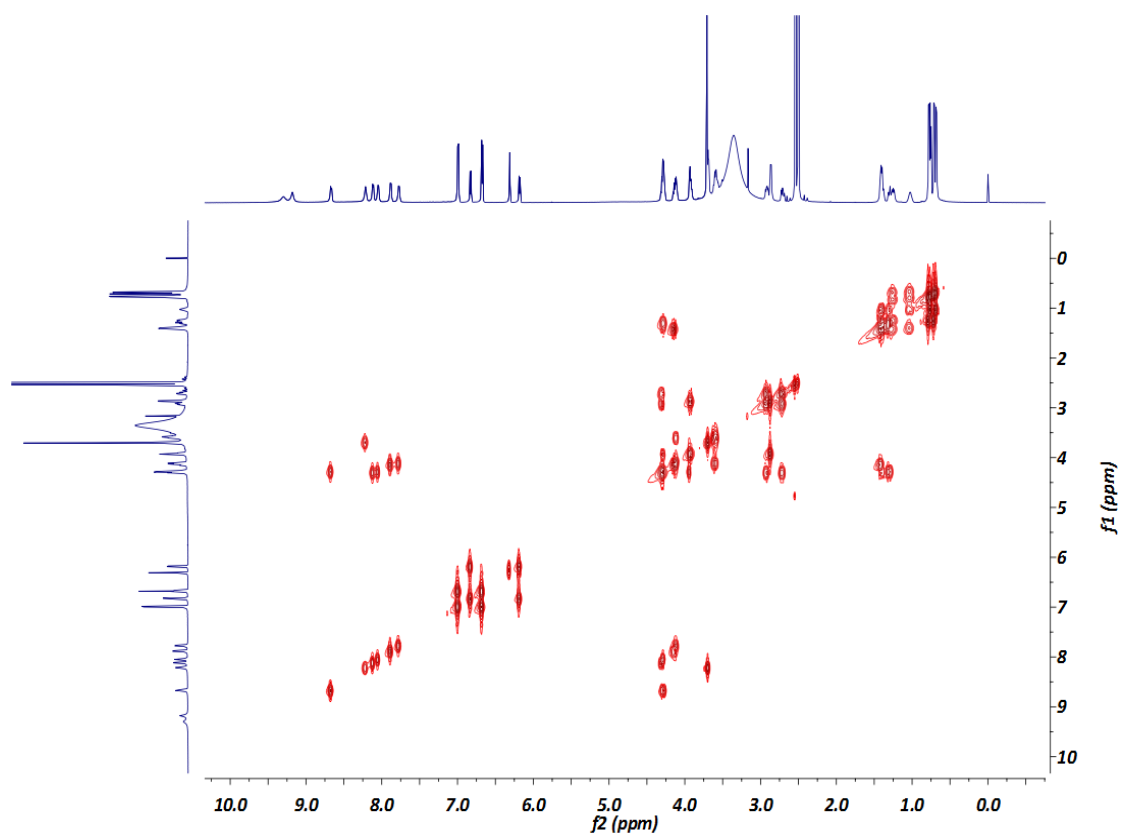


Figure S41.  $^1\text{H}$ - $^1\text{H}$  COSY NMR (600 MHz,  $\text{DMSO-}d_6$ ) spectrum of compound **1**.

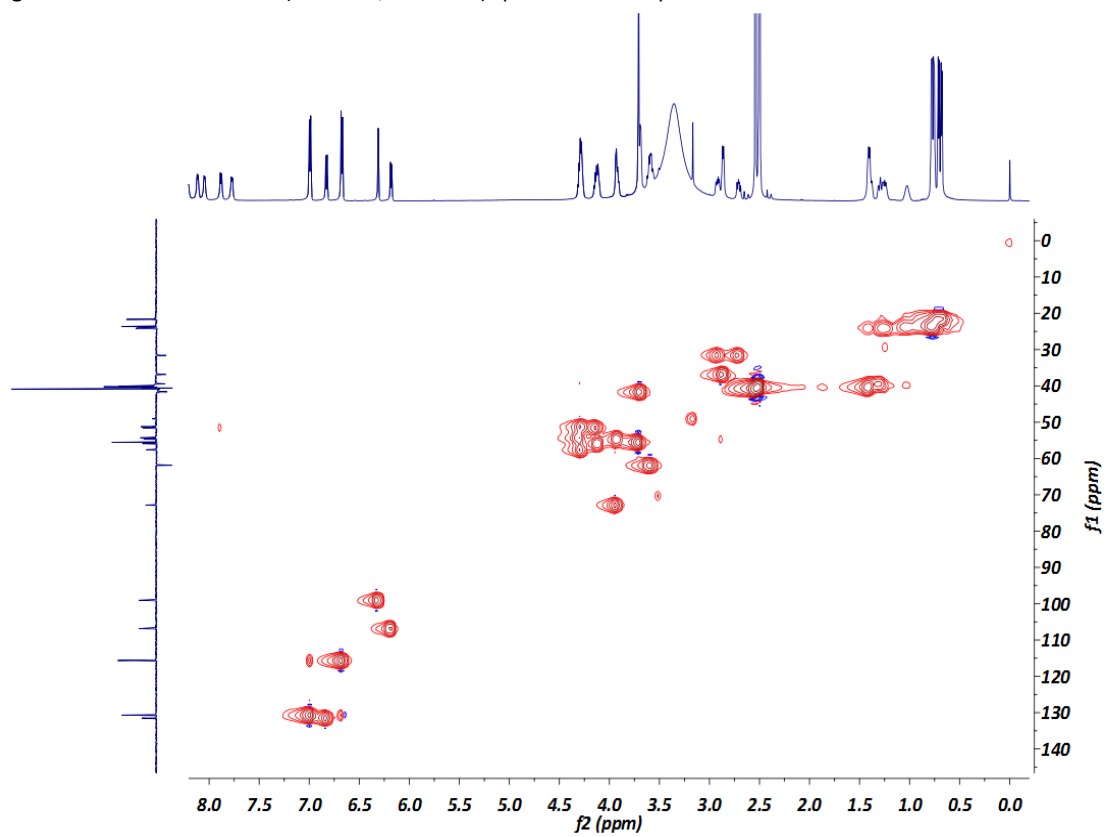


Figure S42. HSQC NMR (600 MHz,  $\text{DMSO-}d_6$ ) spectrum of compound **1**.

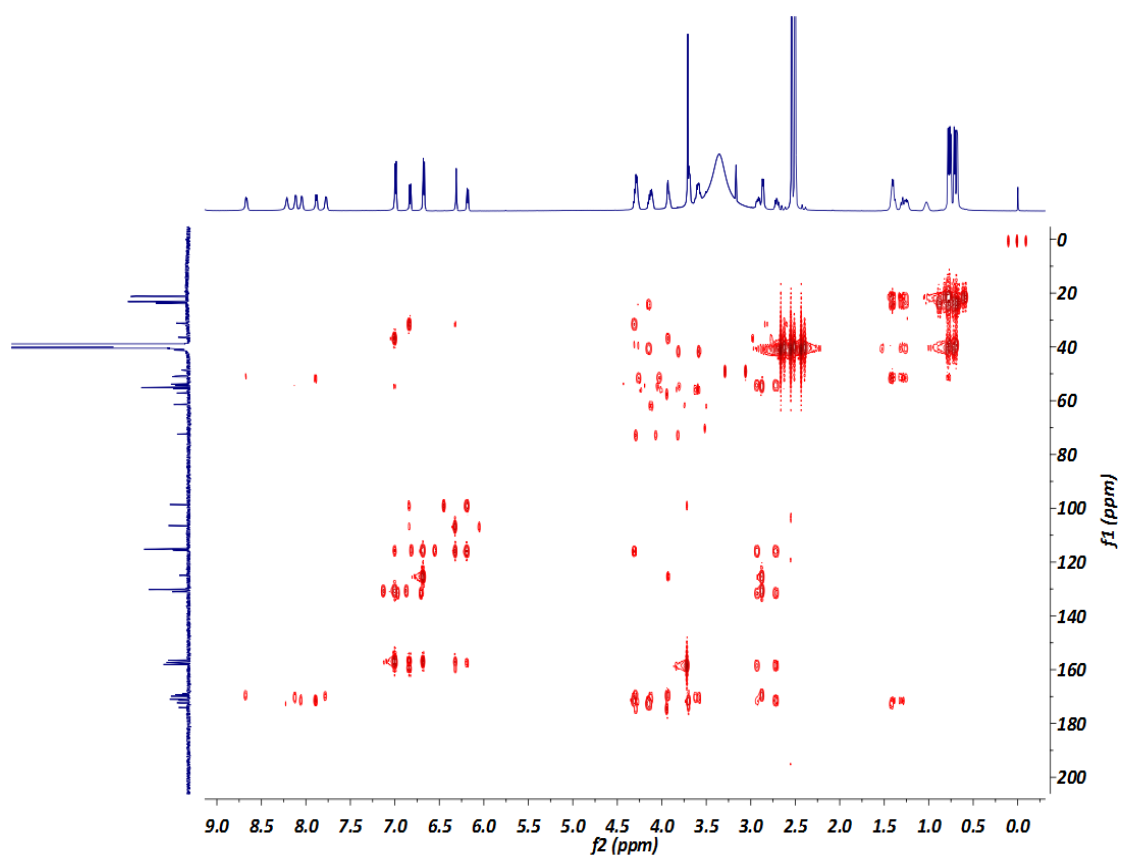


Figure S43. HMBC NMR (600 MHz, DMSO- $d_6$ ) spectrum of compound **1**.

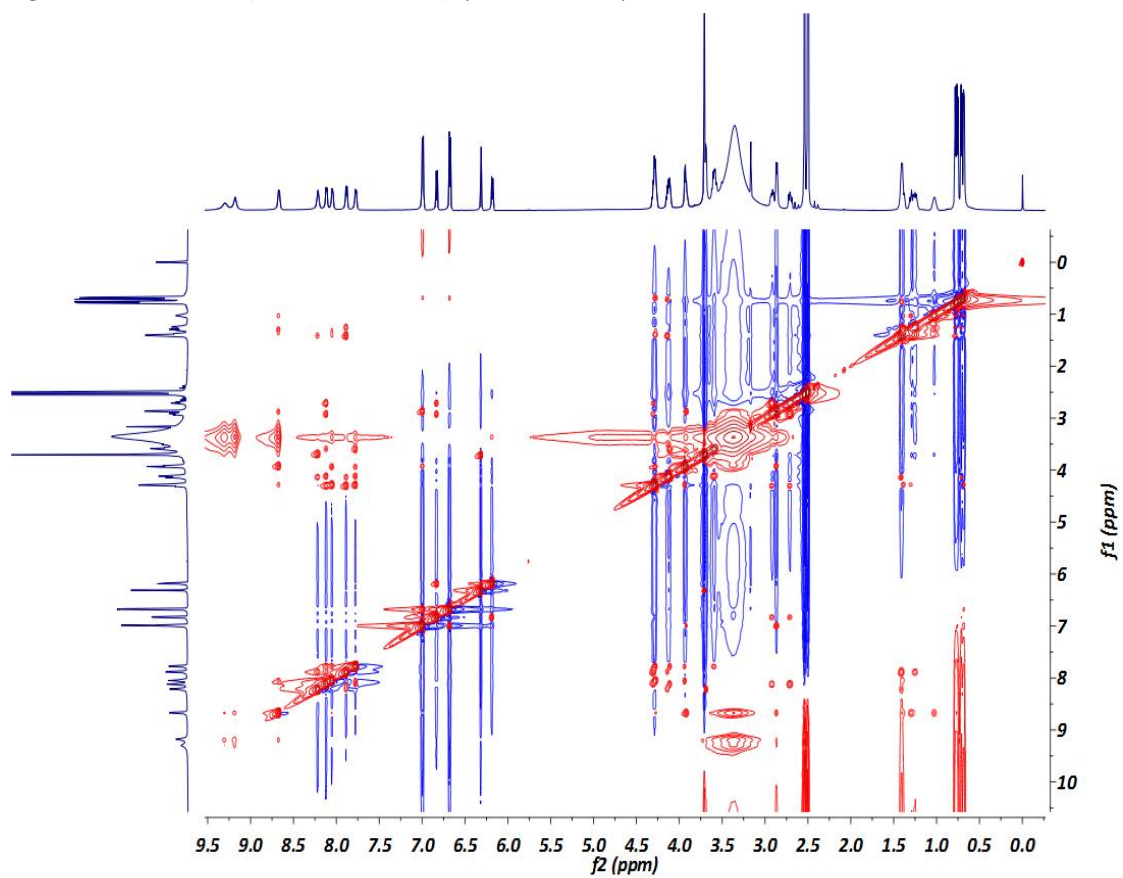


Figure S44. NOESY NMR (600 MHz, DMSO- $d_6$ ) spectrum of compound **1**.

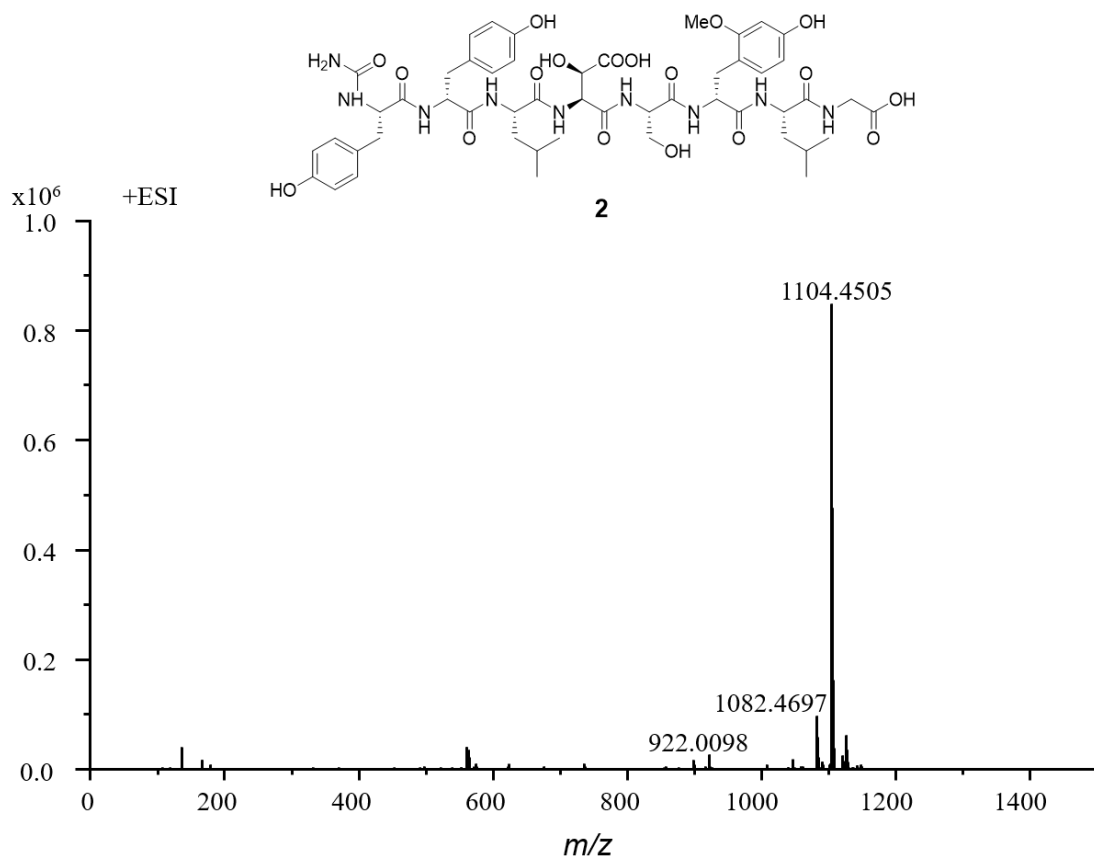


Figure S45. HRESIMS spectrum of 2.

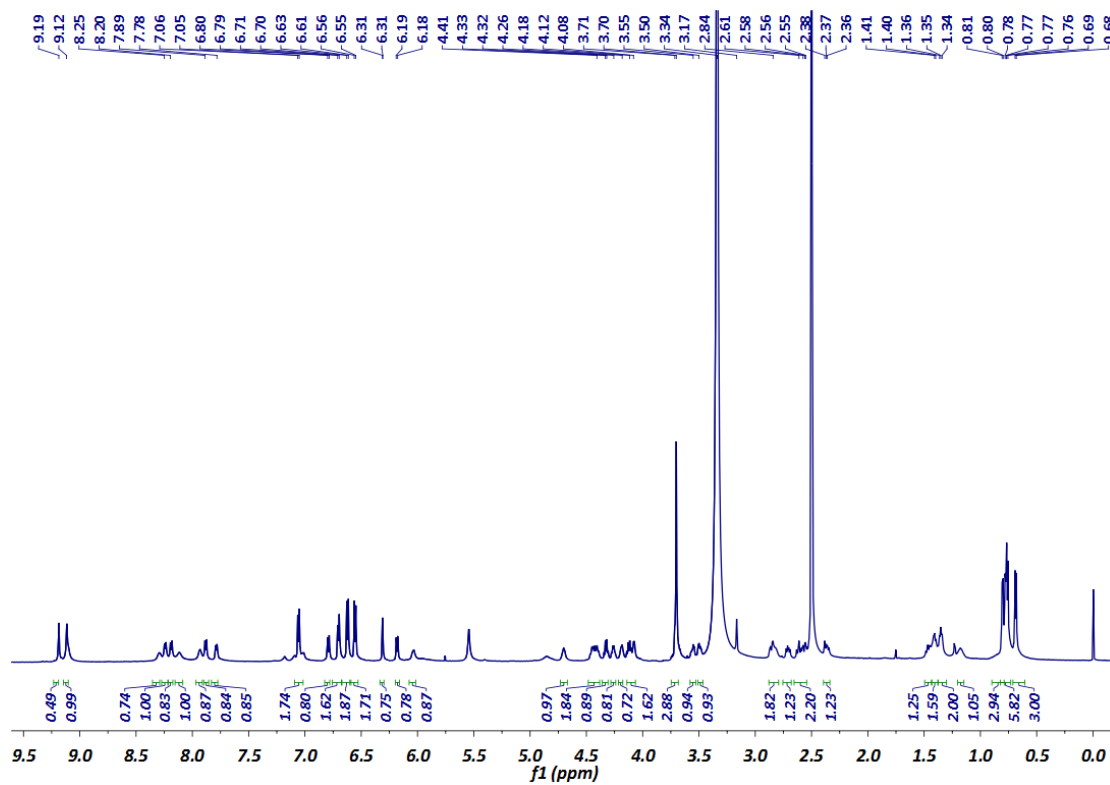


Figure S46. <sup>1</sup>H NMR (600 MHz, DMSO-*d*<sub>6</sub>) spectrum of compound 2.



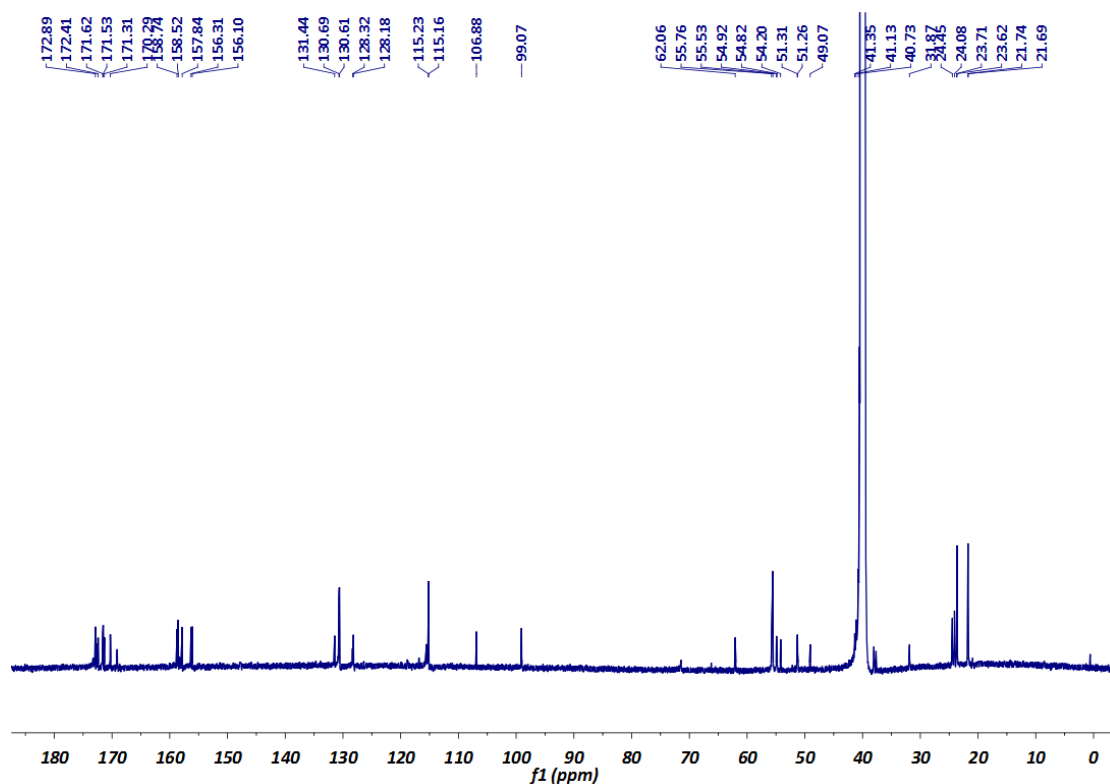


Figure S47.  $^{13}\text{C}$  NMR (150 MHz,  $\text{DMSO}-d_6$ ) spectrum of compound 2.

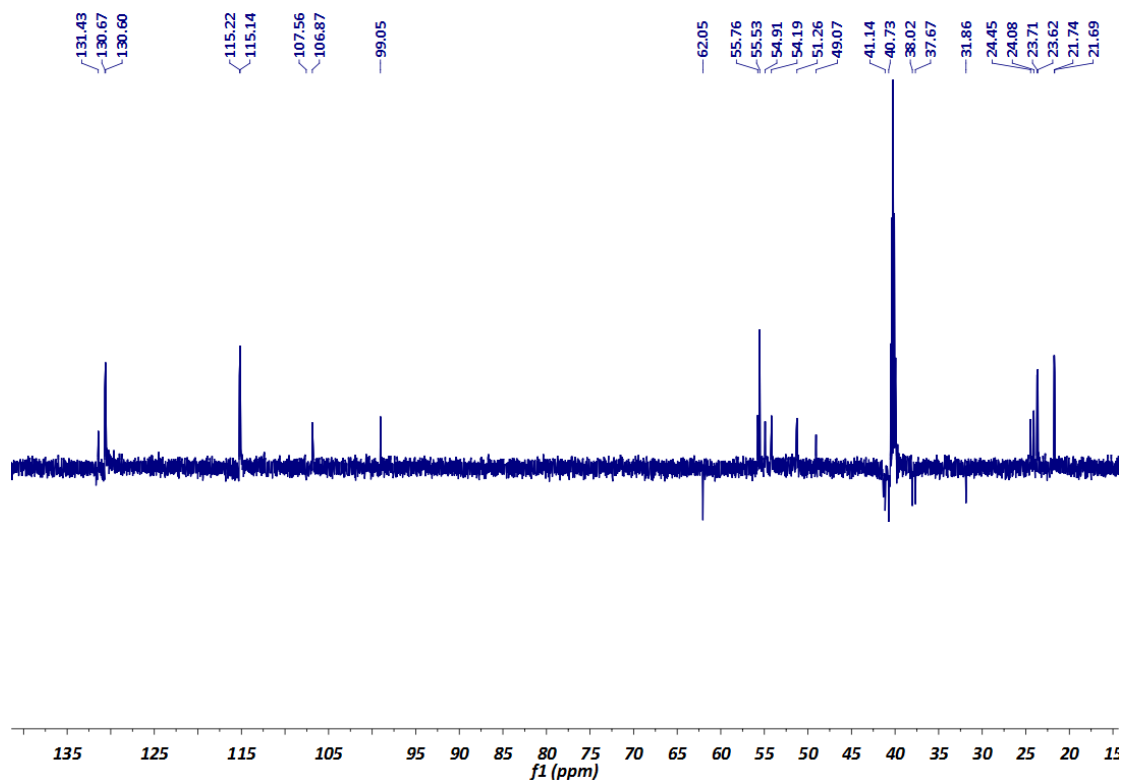


Figure S48. DEPT NMR (150 MHz,  $\text{DMSO}-d_6$ ) spectrum of compound 2.

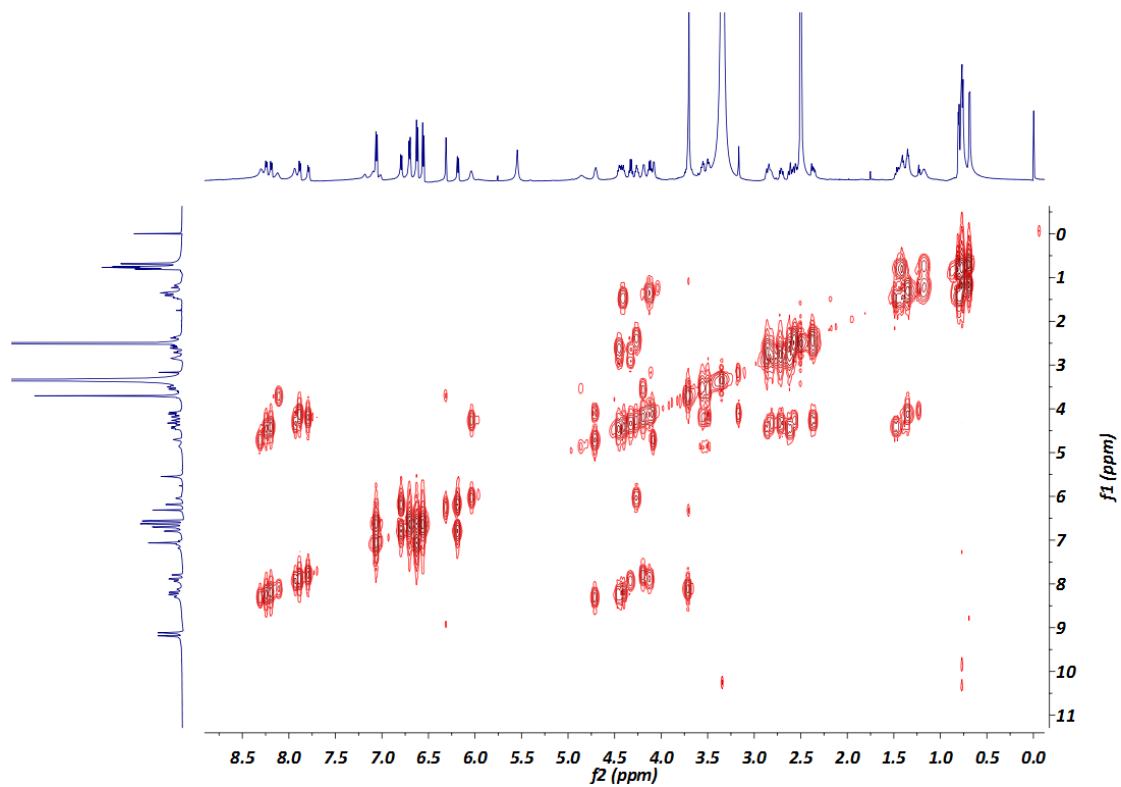


Figure S49.  $^1\text{H}$ - $^1\text{H}$  COSY NMR (600 MHz,  $\text{DMSO}-d_6$ ) spectrum of compound 2.

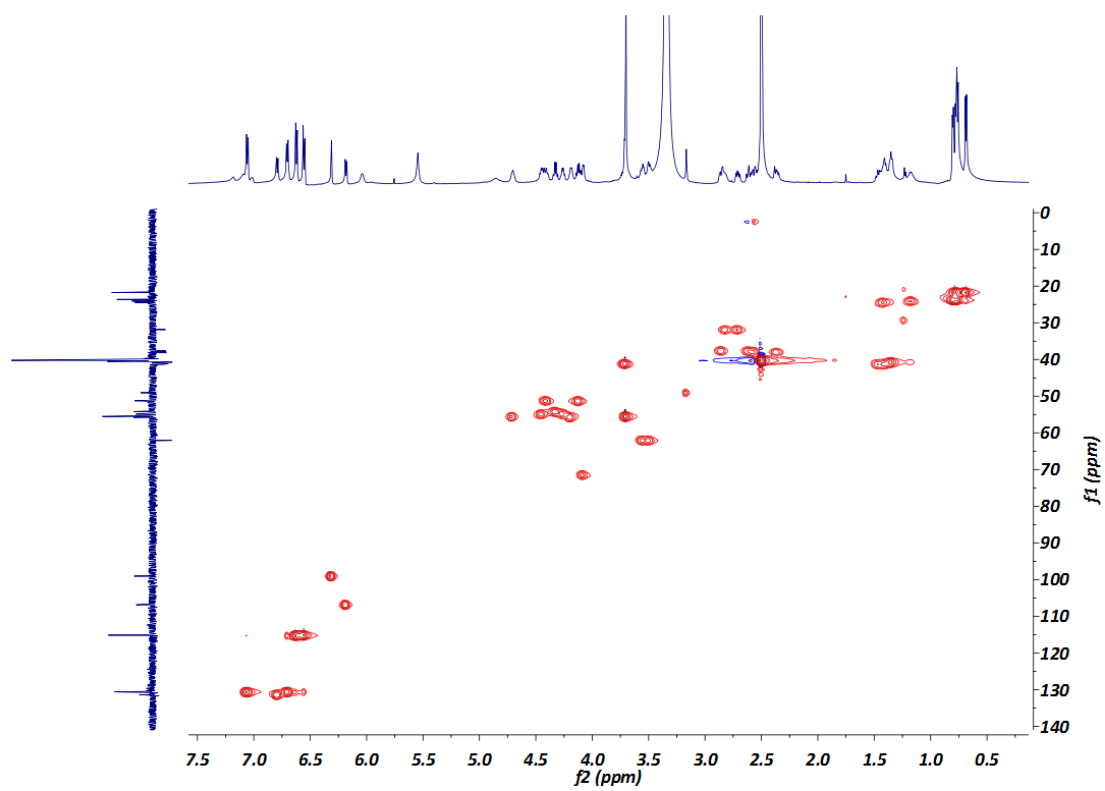


Figure S50. HSQC NMR (600 MHz,  $\text{DMSO}-d_6$ ) spectrum of compound 2.

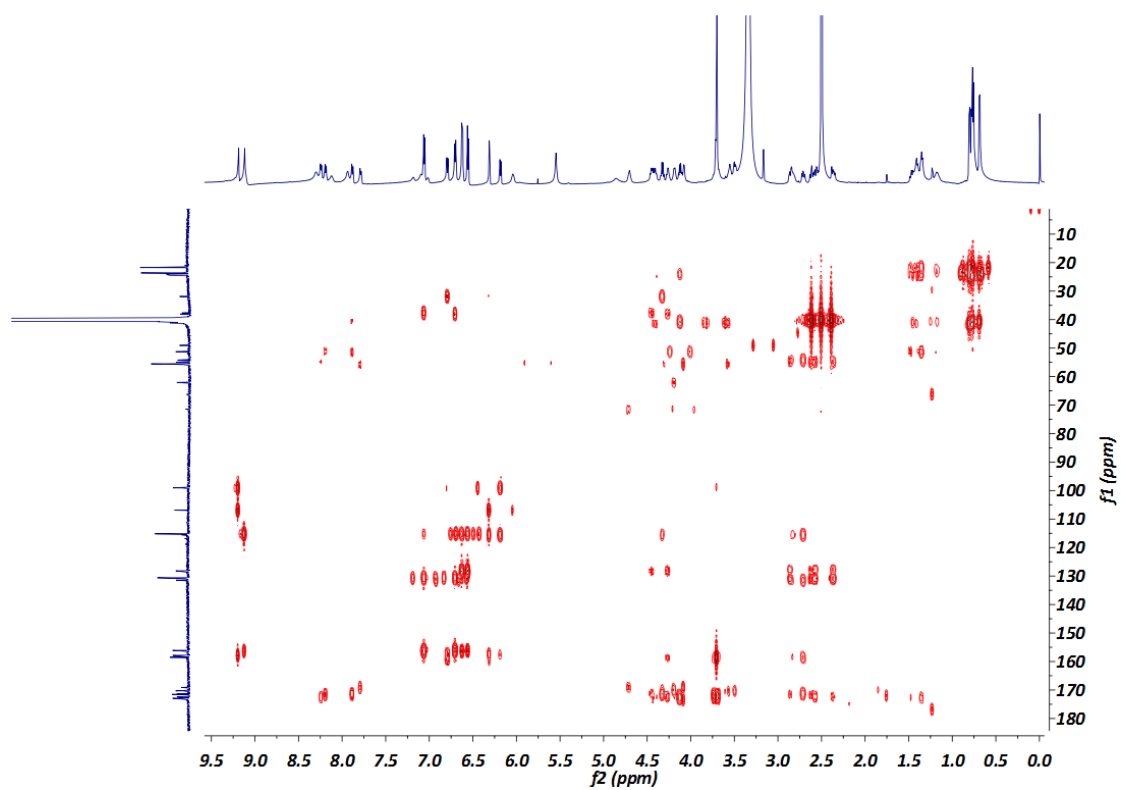


Figure S51. HMBC NMR (600 MHz, DMSO- $d_6$ ) spectrum of compound 2.

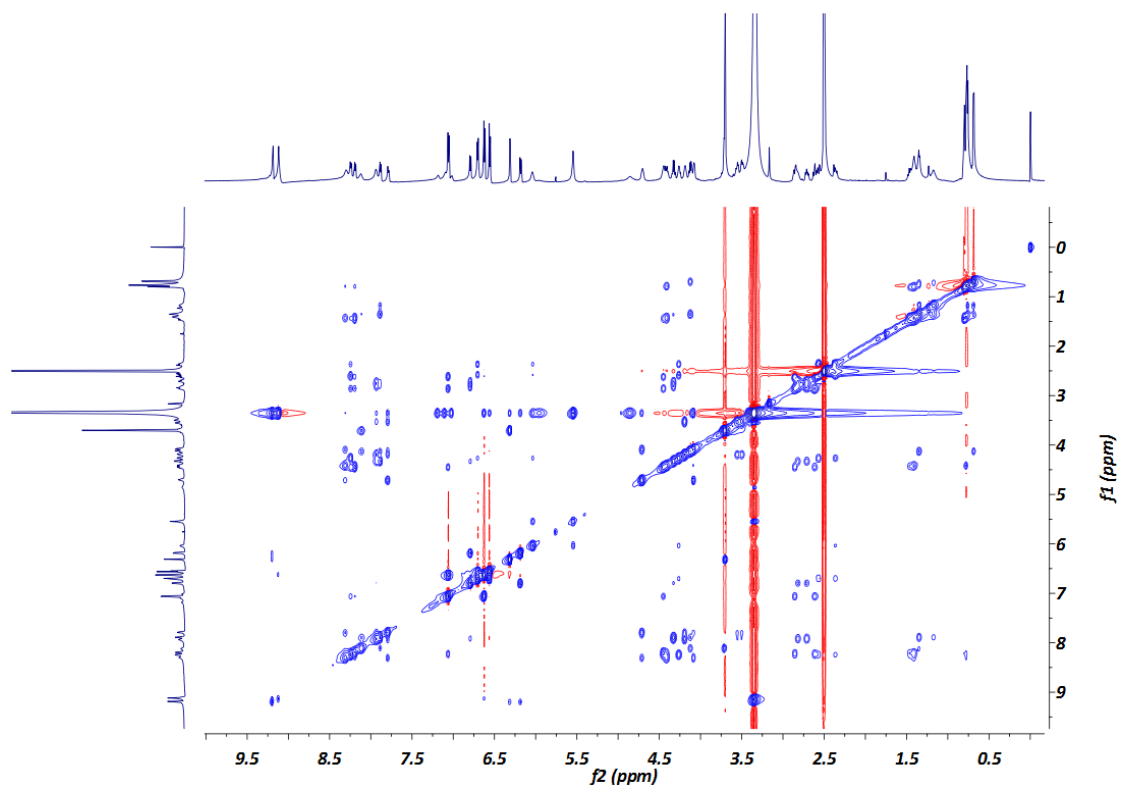


Figure S52. NOESY NMR (600 MHz, DMSO- $d_6$ ) spectrum of compound 2.

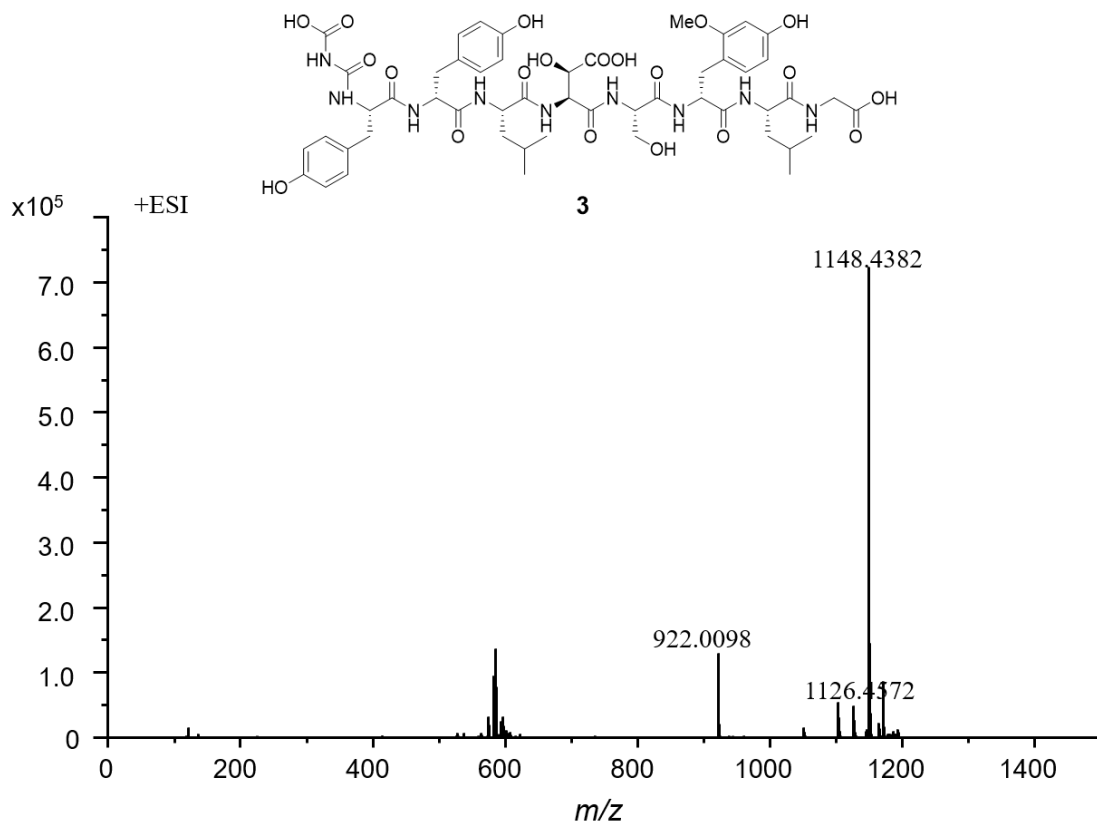


Figure S53. HRESIMS spectrum of **3**.

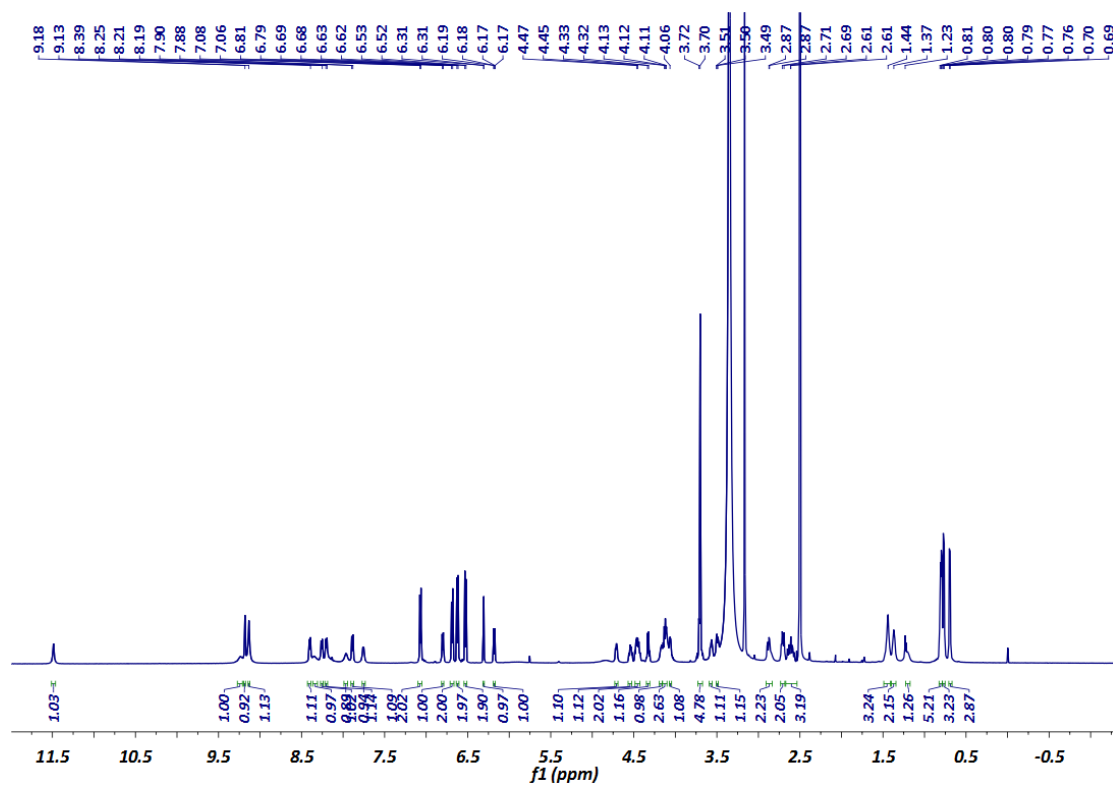


Figure S54. <sup>1</sup>H NMR (600 MHz, DMSO-*d*<sub>6</sub>) spectrum of compound **3**.

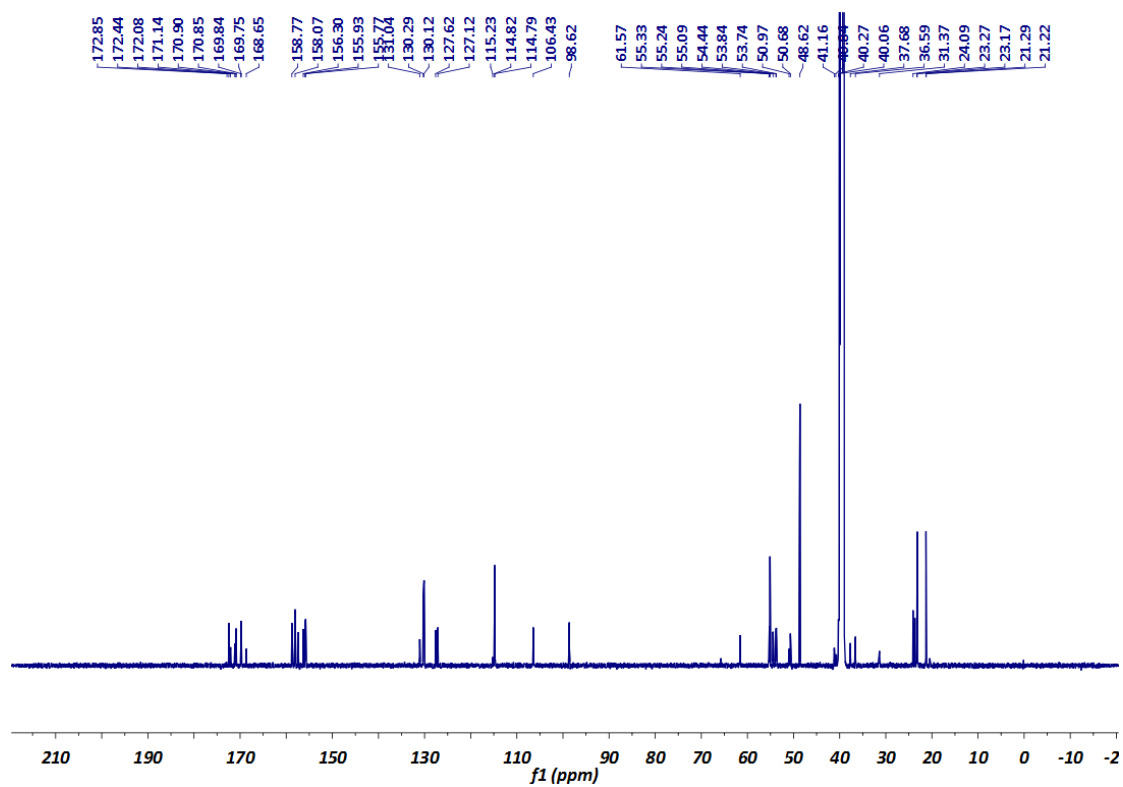


Figure S55.  $^{13}\text{C}$  NMR (150 MHz,  $\text{DMSO-}d_6$ ) spectrum of compound **3**.

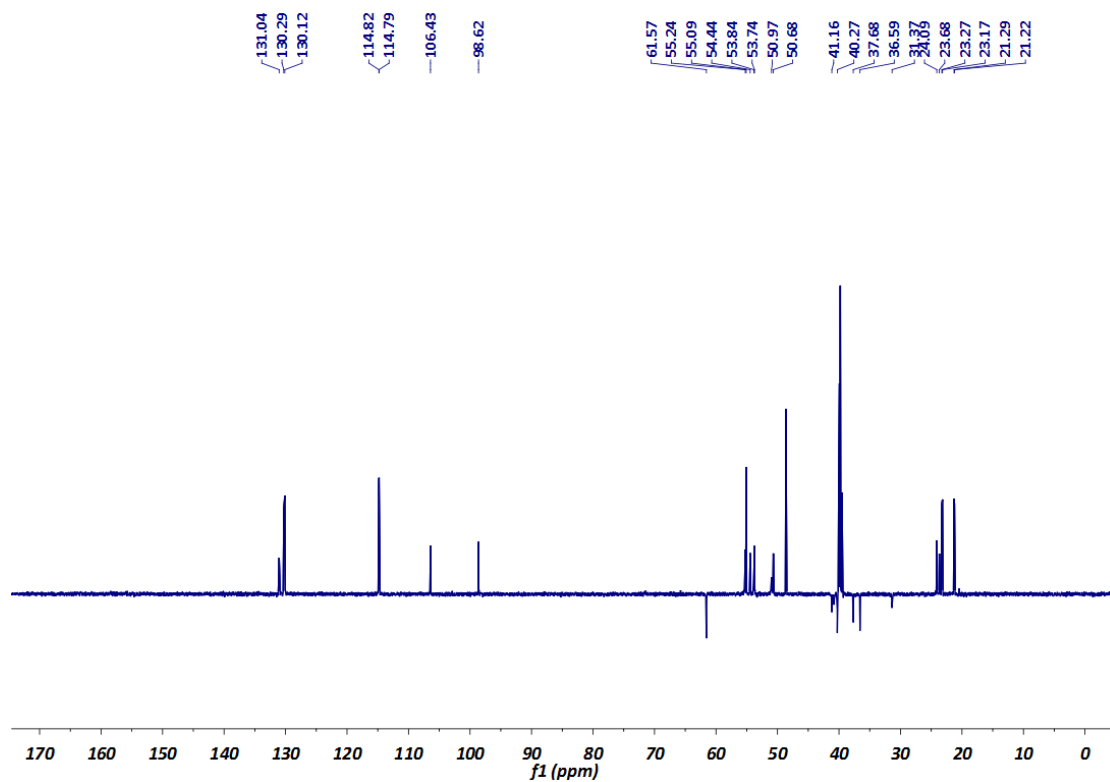


Figure S56. DEPT NMR (150 MHz,  $\text{DMSO-}d_6$ ) spectrum of compound **3**.

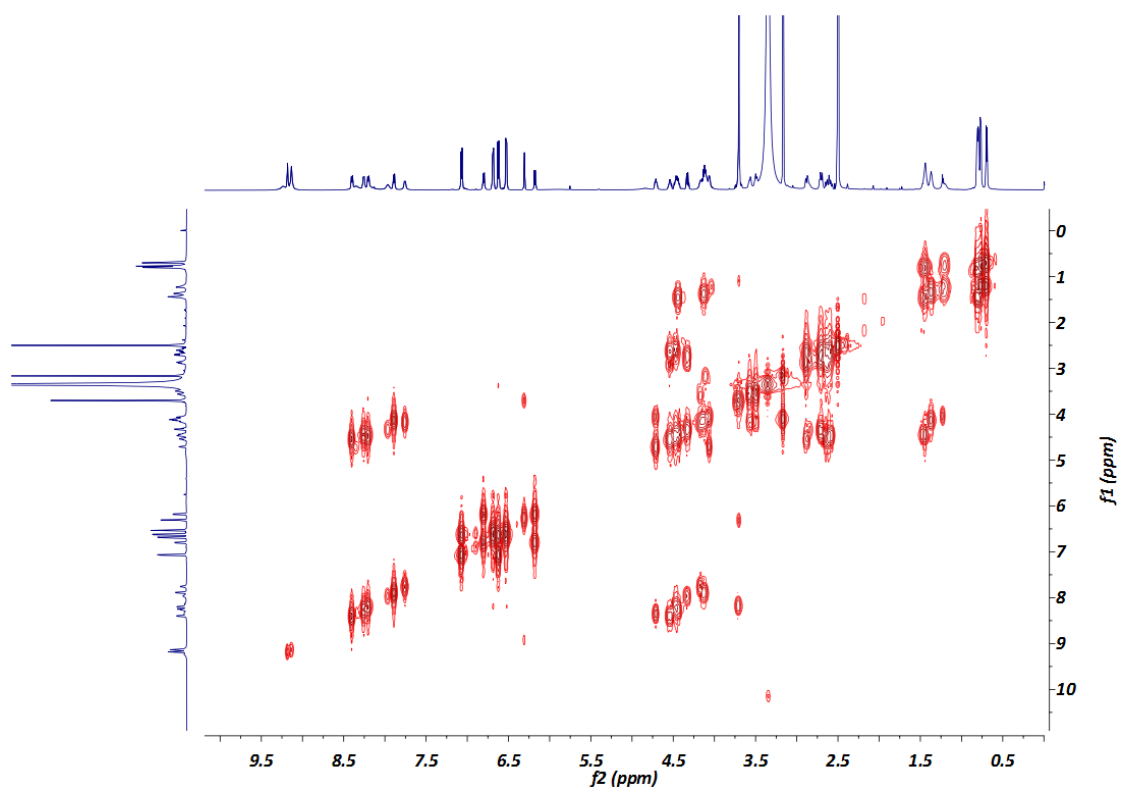


Figure S57.  $^1\text{H}$ - $^1\text{H}$  COSY NMR (600 MHz,  $\text{DMSO-}d_6$ ) spectrum of compound **3**.

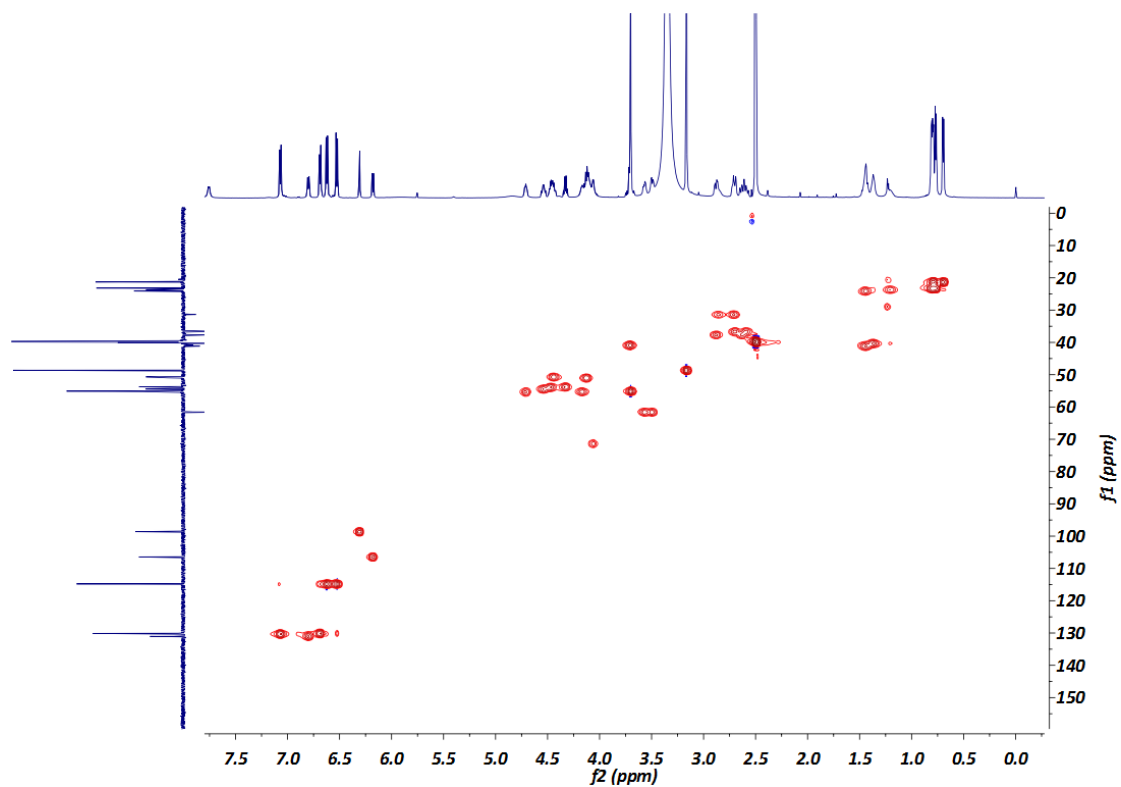


Figure S58. HSQC NMR (600 MHz,  $\text{DMSO-}d_6$ ) spectrum of compound **3**.

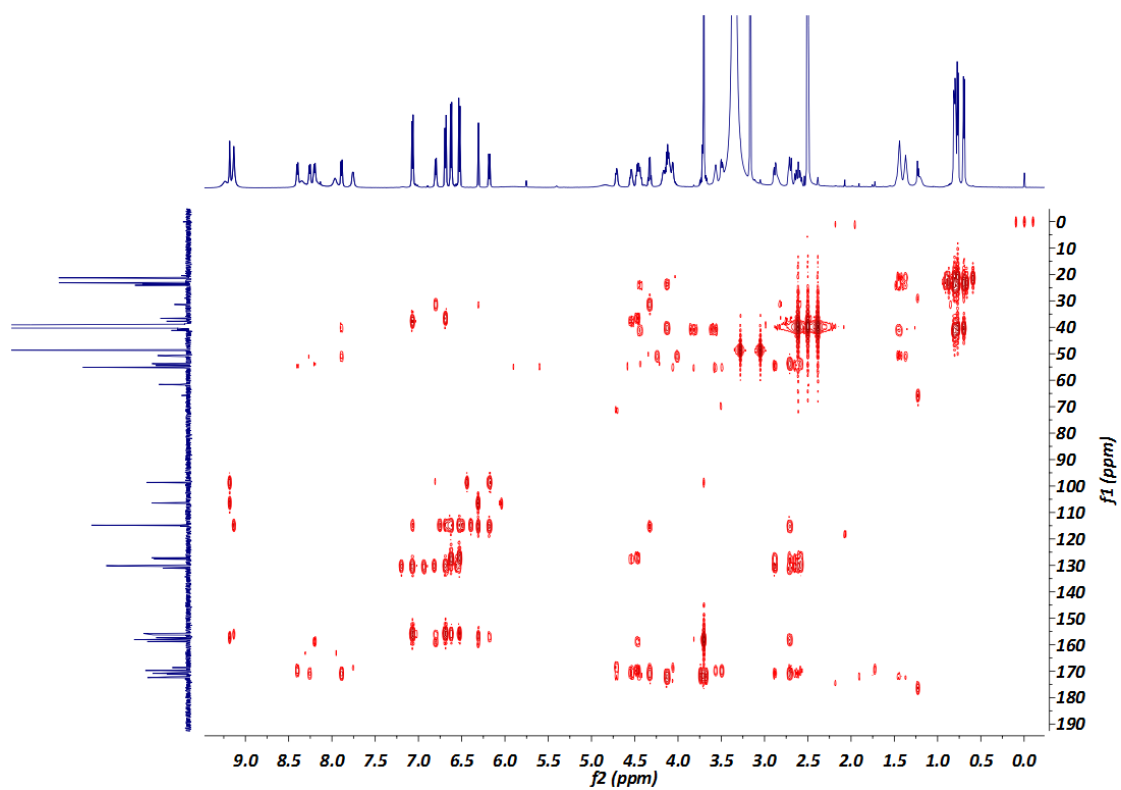


Figure S59. HMBC NMR (600 MHz, DMSO- $d_6$ ) spectrum of compound **3**.

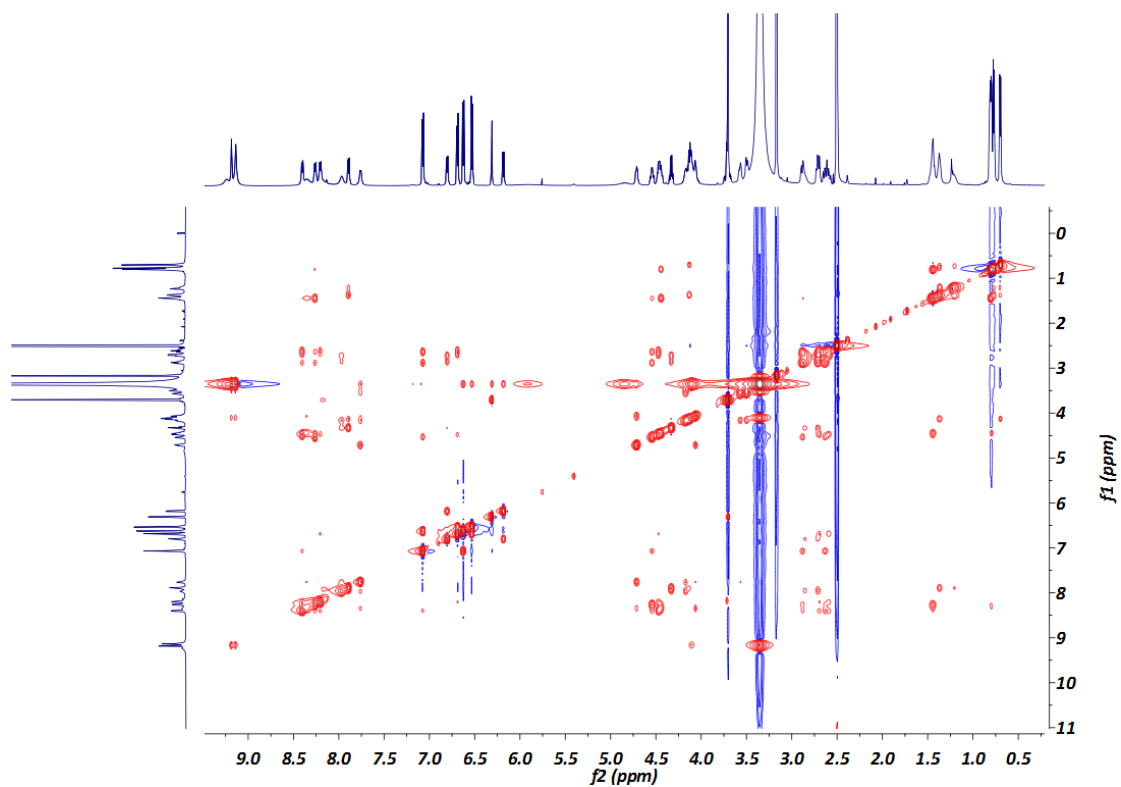


Figure S60. NOESY NMR (600 MHz, DMSO- $d_6$ ) spectrum of compound **3**.

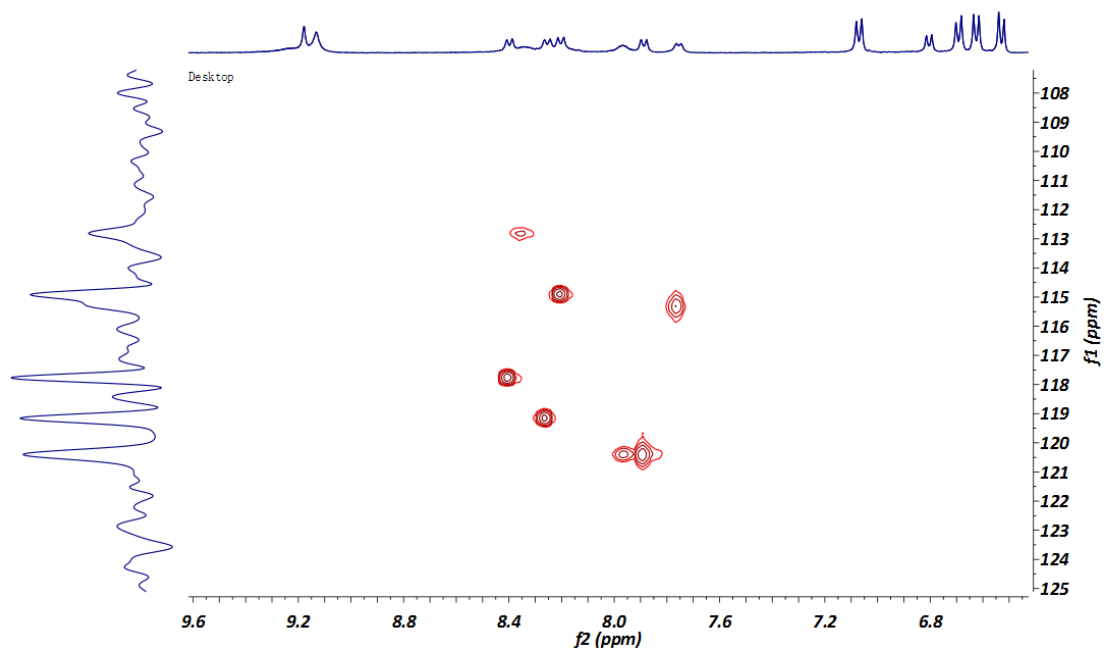


Figure S61. N-H HSQC NMR (600 MHz, DMSO-*d*<sub>6</sub>) spectrum of compound **3**.

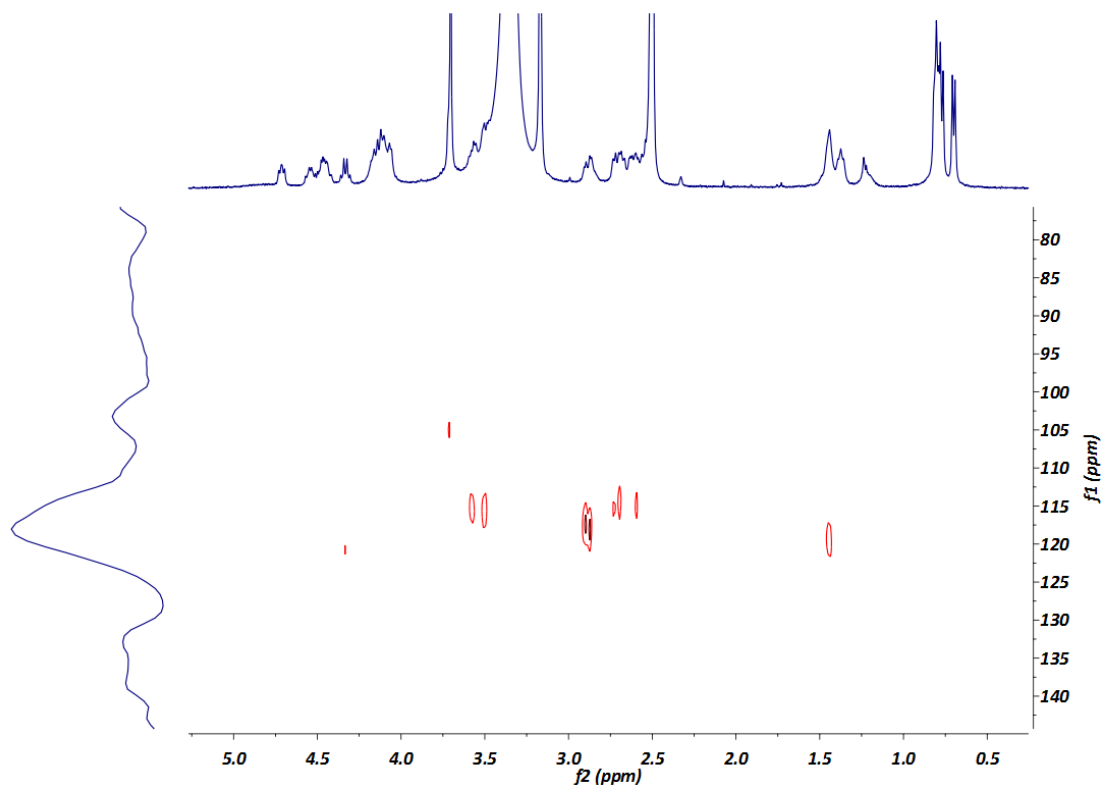


Figure S62. N-H HMBC NMR (600 MHz, DMSO-*d*<sub>6</sub>) spectrum of compound **3**.



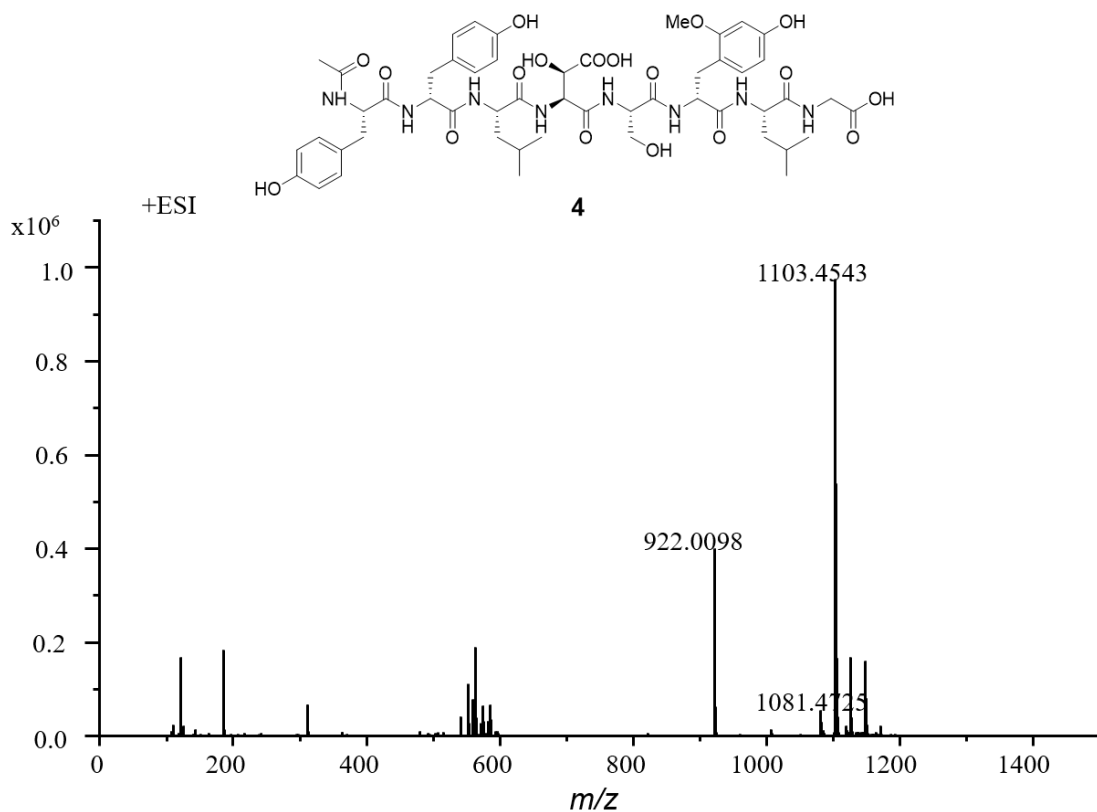


Figure S63. HRESIMS spectrum of **4**.

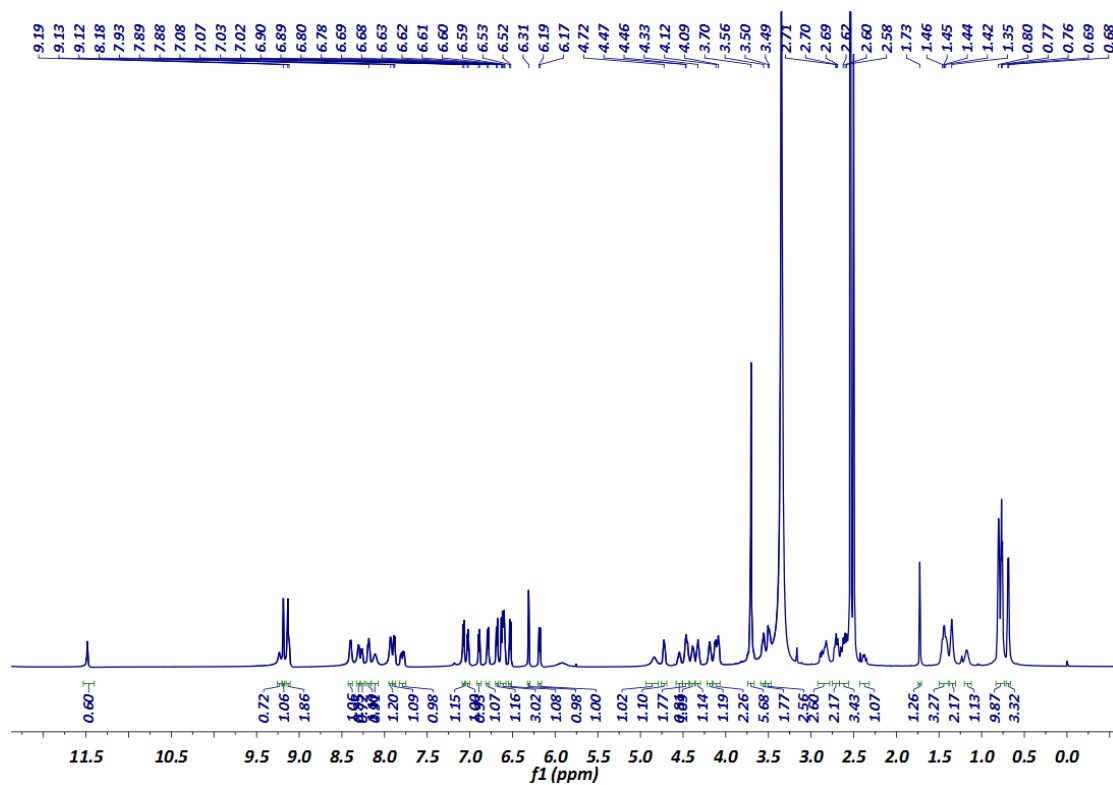


Figure S64.  $^1\text{H}$  NMR (600 MHz,  $\text{DMSO}-d_6$ ) spectrum of compound **4**.

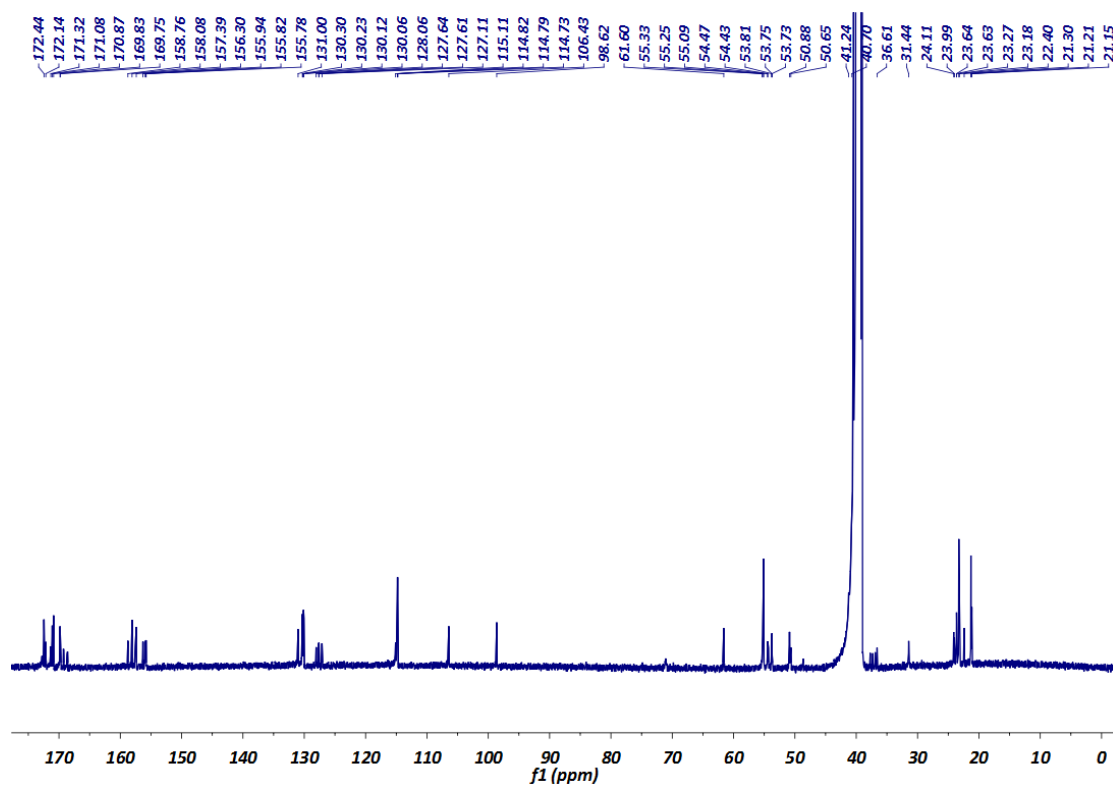


Figure S65.  $^{13}\text{C}$  NMR (150 MHz,  $\text{DMSO}-d_6$ ) spectrum of compound 4.

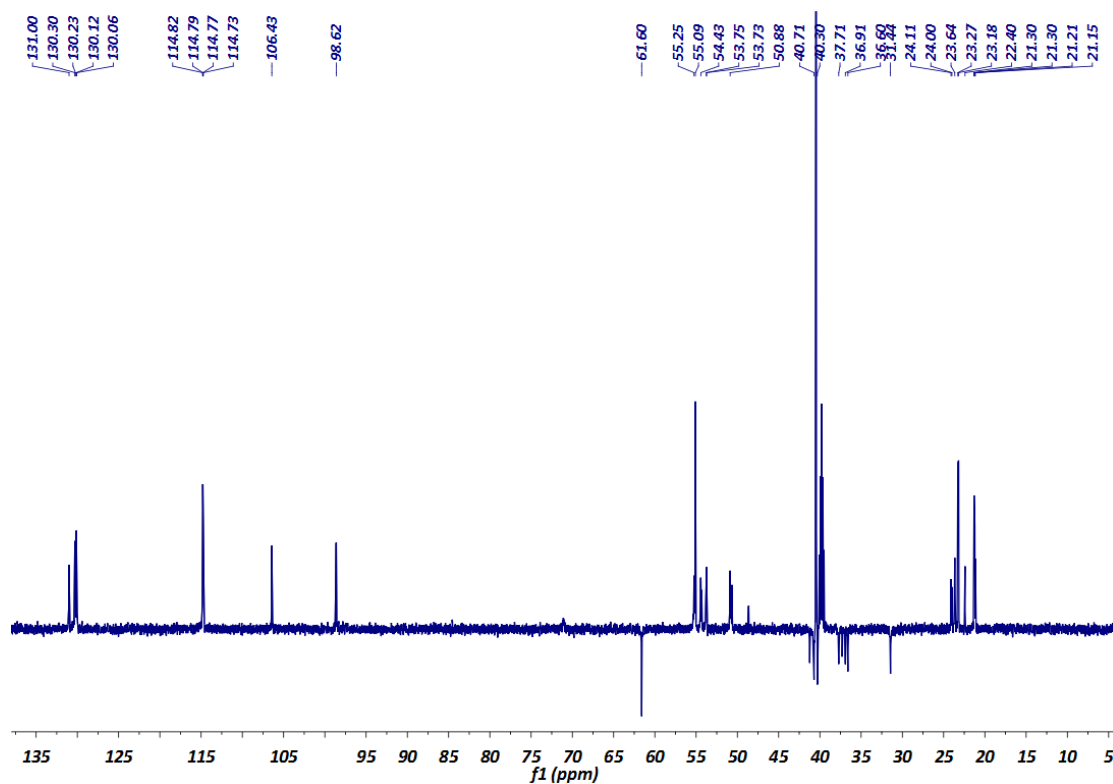


Figure S66. DEPT NMR (150 MHz,  $\text{DMSO}-d_6$ ) spectrum of compound 4.

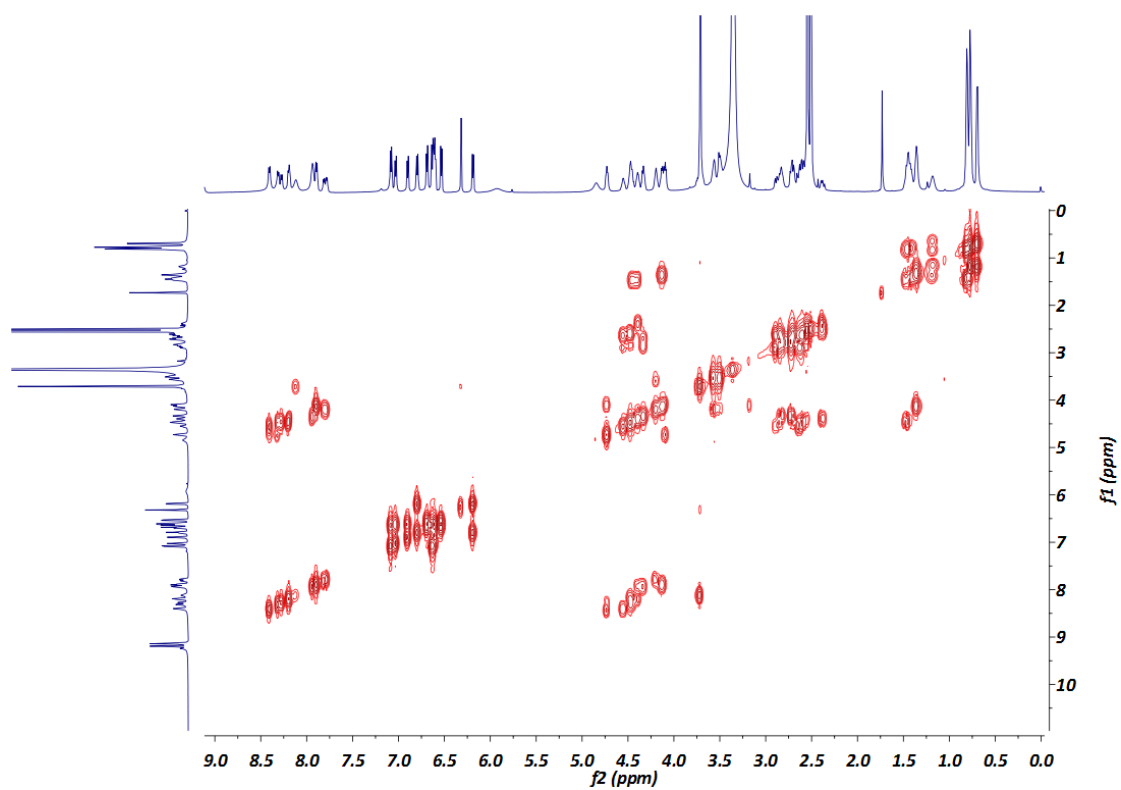


Figure S67.  $^1\text{H}$ - $^1\text{H}$  COSY NMR (600 MHz,  $\text{DMSO-}d_6$ ) spectrum of compound 4.

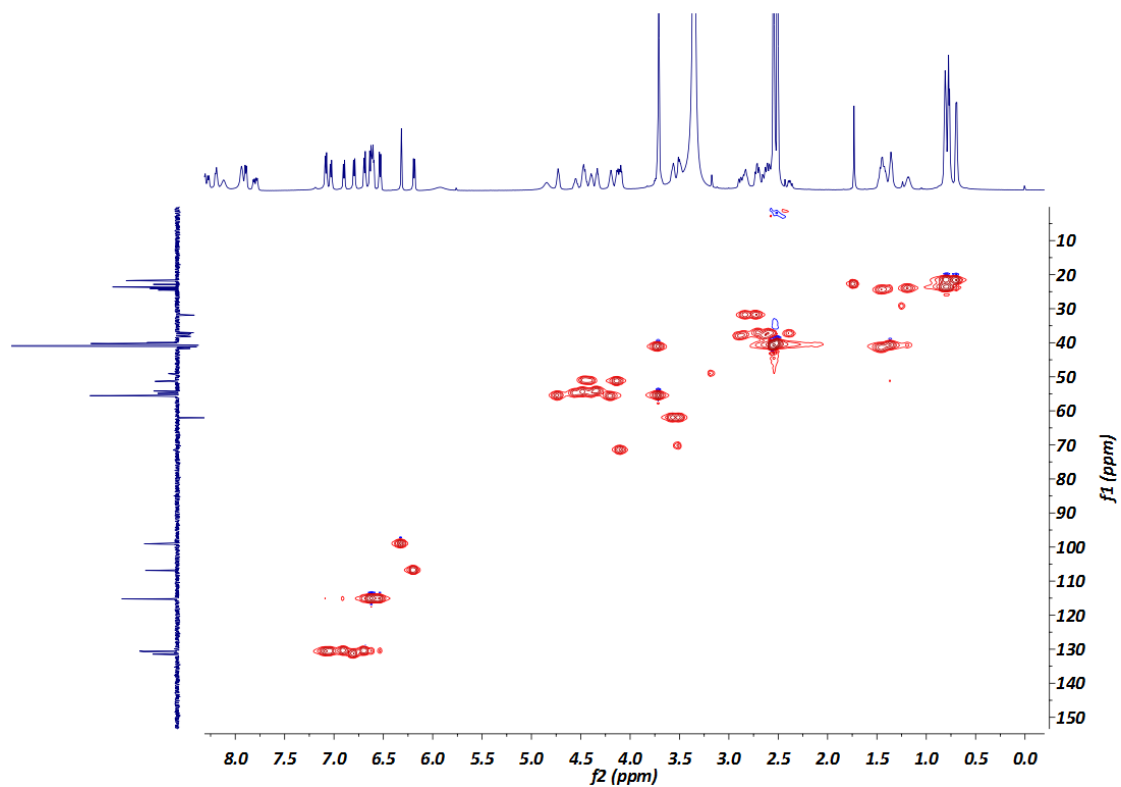


Figure S68. HSQC NMR (600 MHz,  $\text{DMSO-}d_6$ ) spectrum of compound 4.

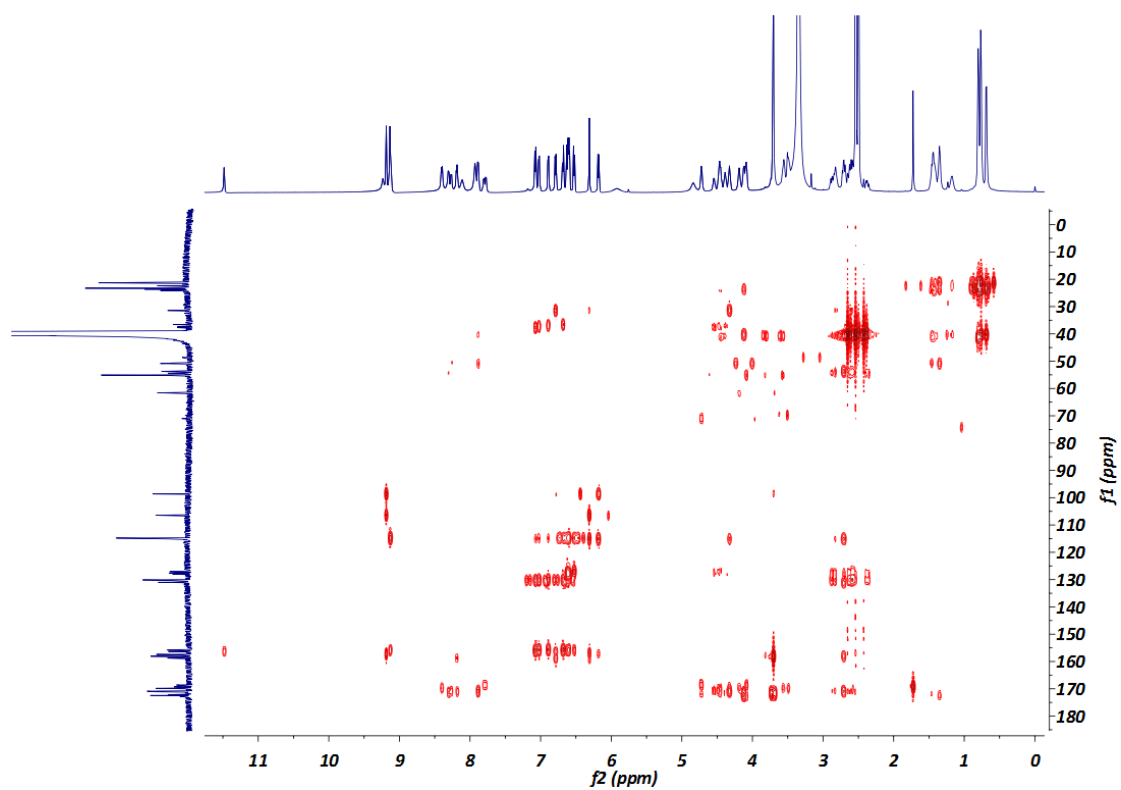


Figure S69. HMBC NMR (600 MHz, DMSO- $d_6$ ) spectrum of compound **4**.

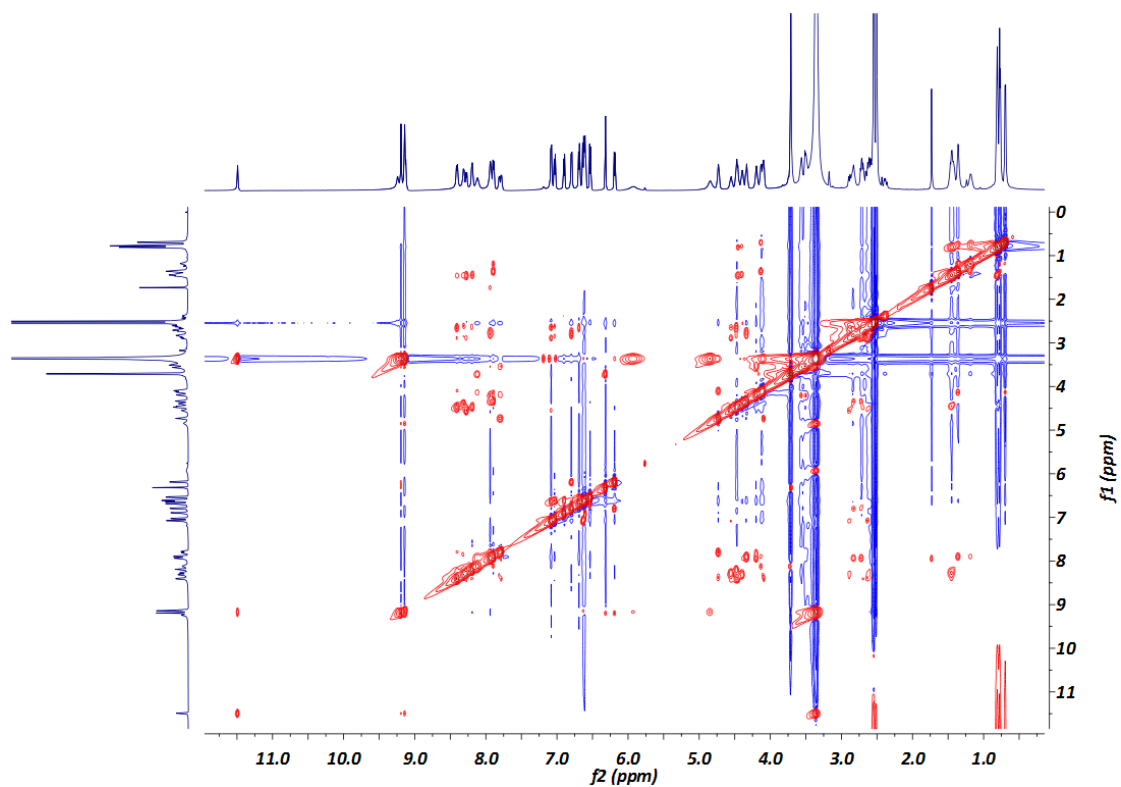


Figure S70. NOESY NMR (600 MHz, DMSO- $d_6$ ) spectrum of compound **4**.

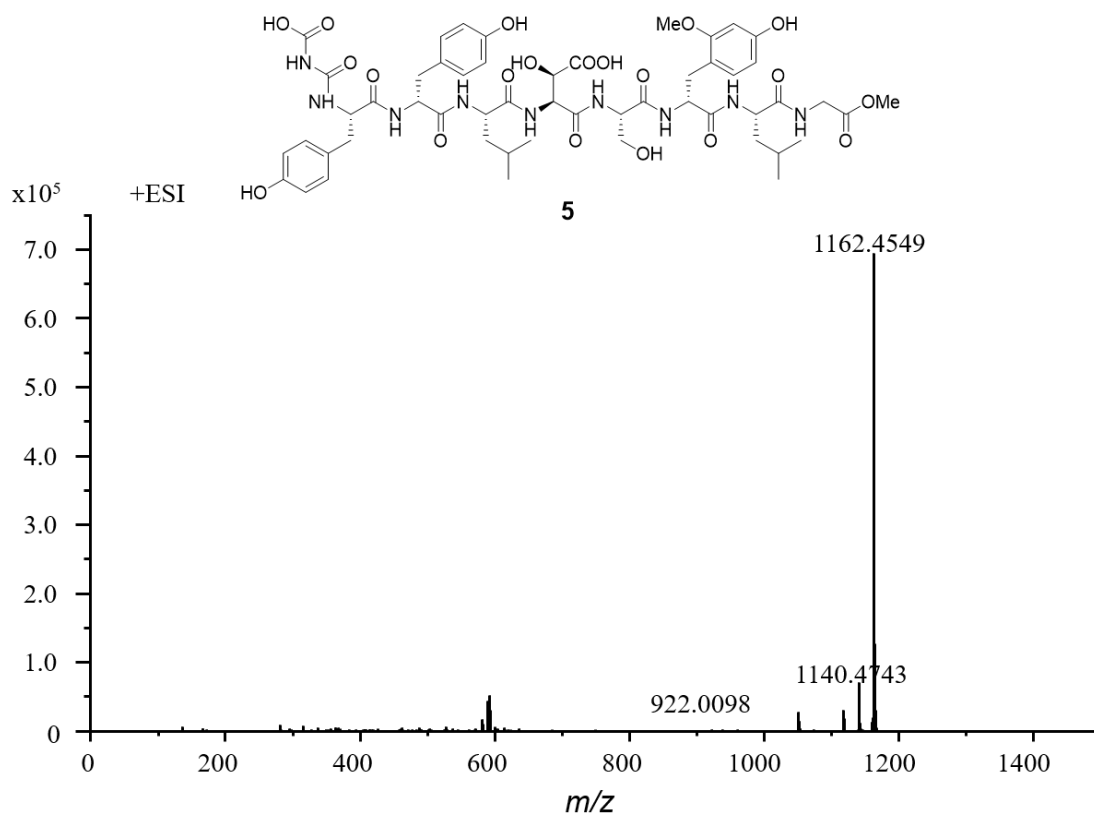


Figure S71. HRESIMS spectrum of 5

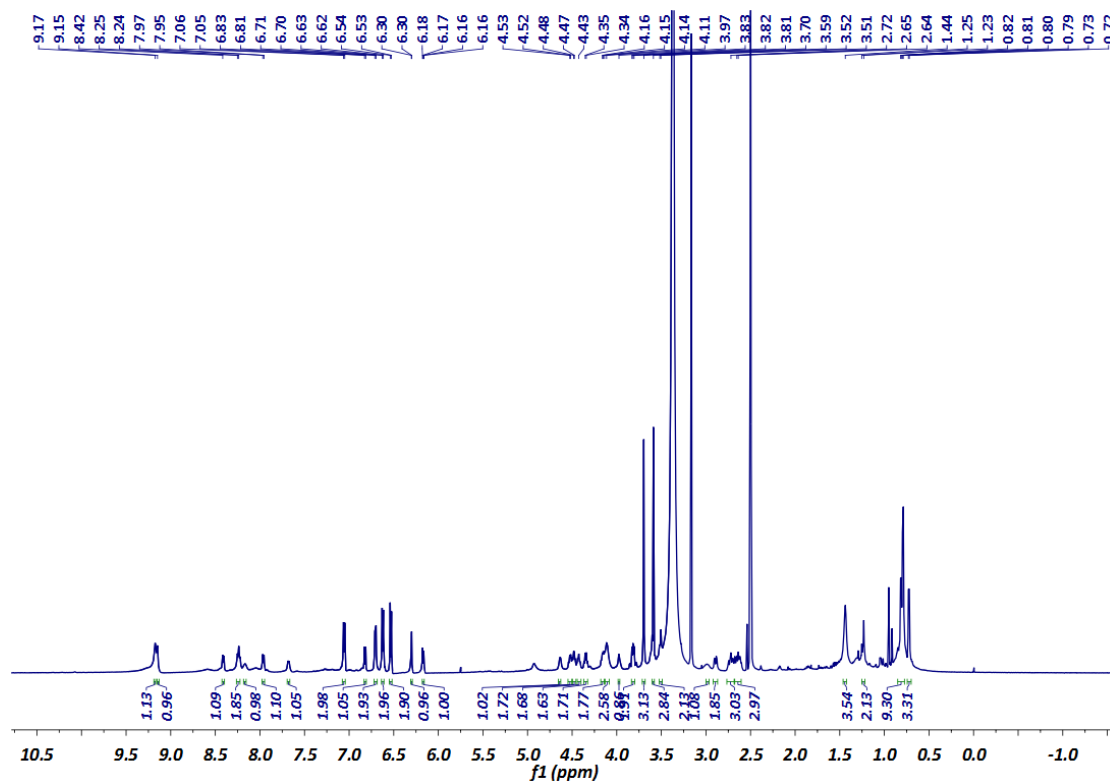


Figure S72. <sup>1</sup>H NMR (600 MHz, DMSO-d<sub>6</sub>) spectrum of compound 5.

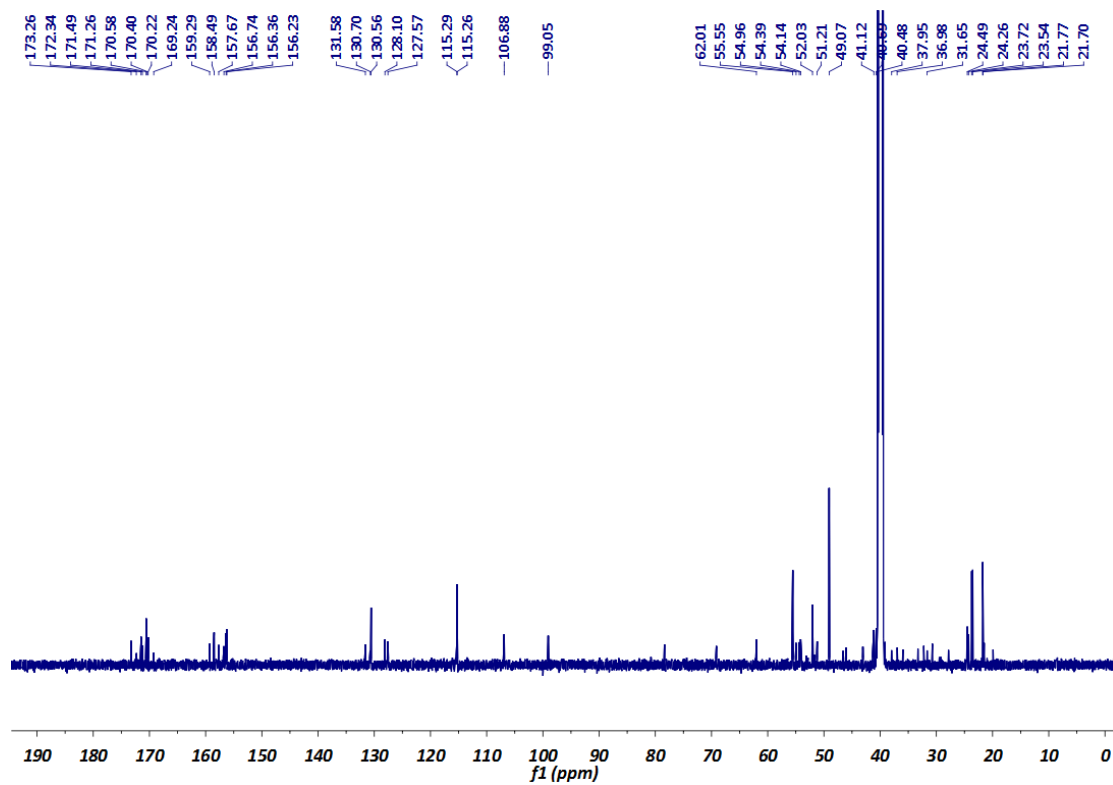


Figure S73.  $^{13}\text{C}$  NMR (150 MHz,  $\text{DMSO}-d_6$ ) spectrum of compound 5.

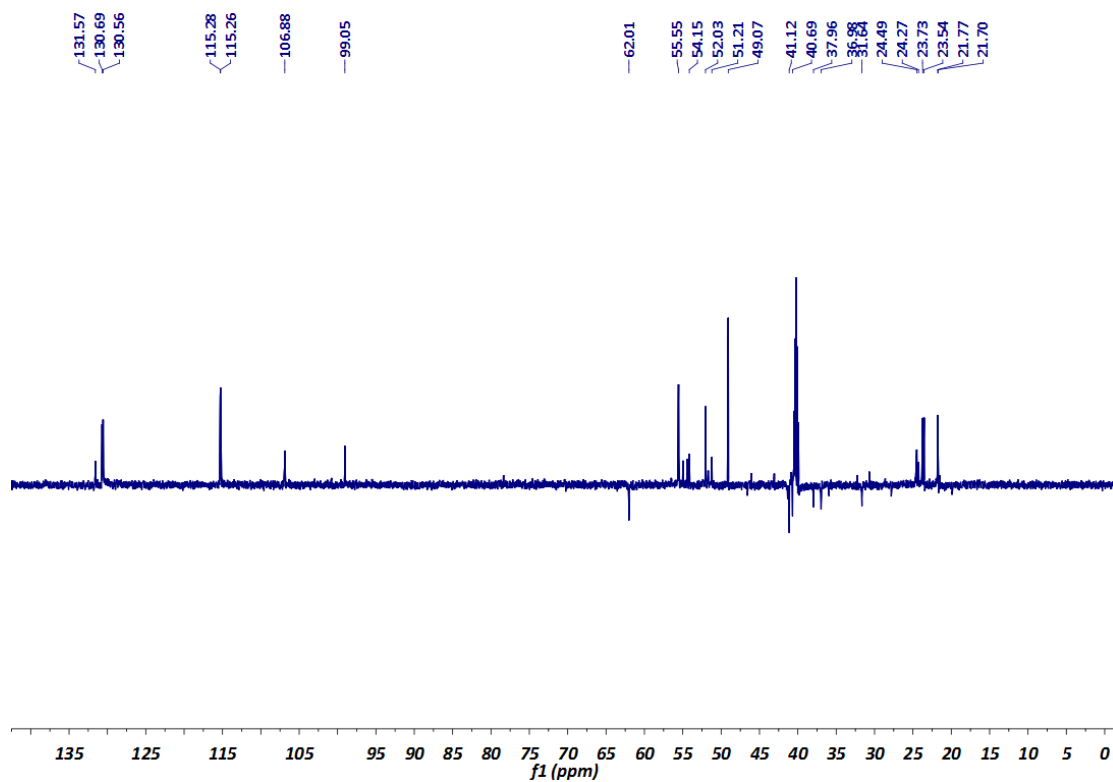


Figure S74. DEPT NMR (150 MHz,  $\text{DMSO}-d_6$ ) spectrum of compound 5.

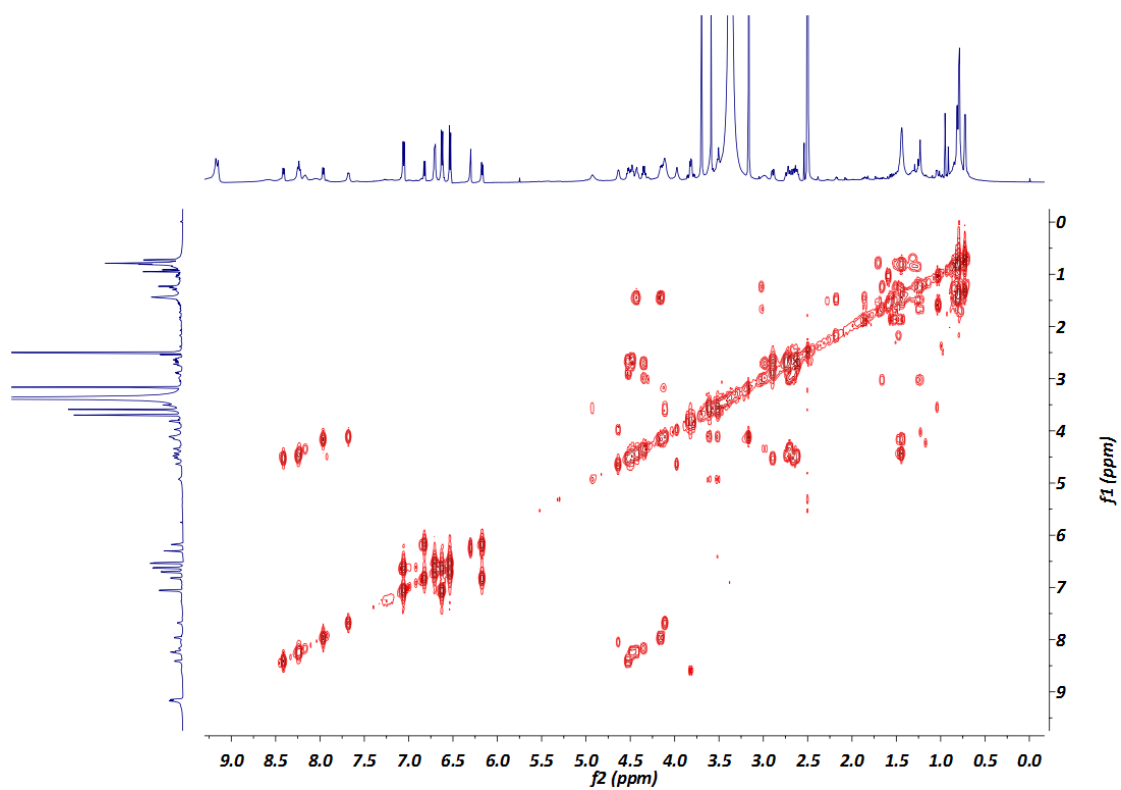


Figure S75.  $^1\text{H}$ - $^1\text{H}$  COSY NMR (600 MHz,  $\text{DMSO-}d_6$ ) spectrum of compound 5.

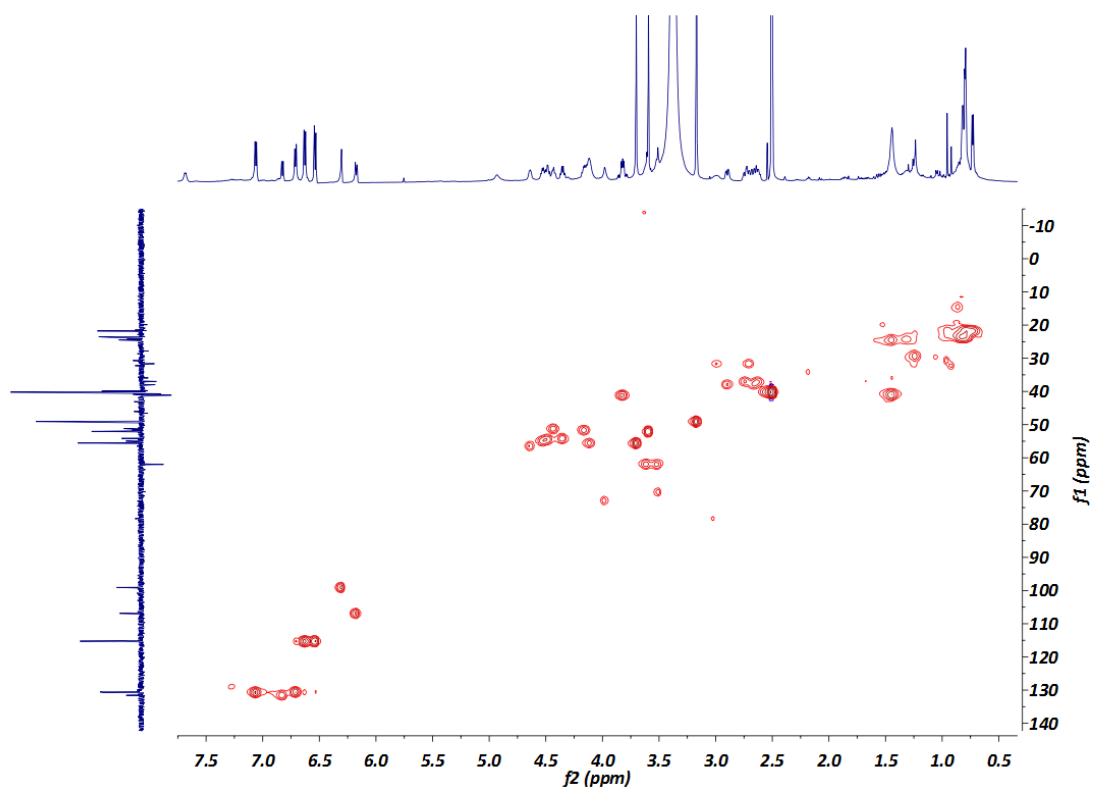


Figure S76. HSQC NMR (600 MHz,  $\text{DMSO-}d_6$ ) spectrum of compound 5.

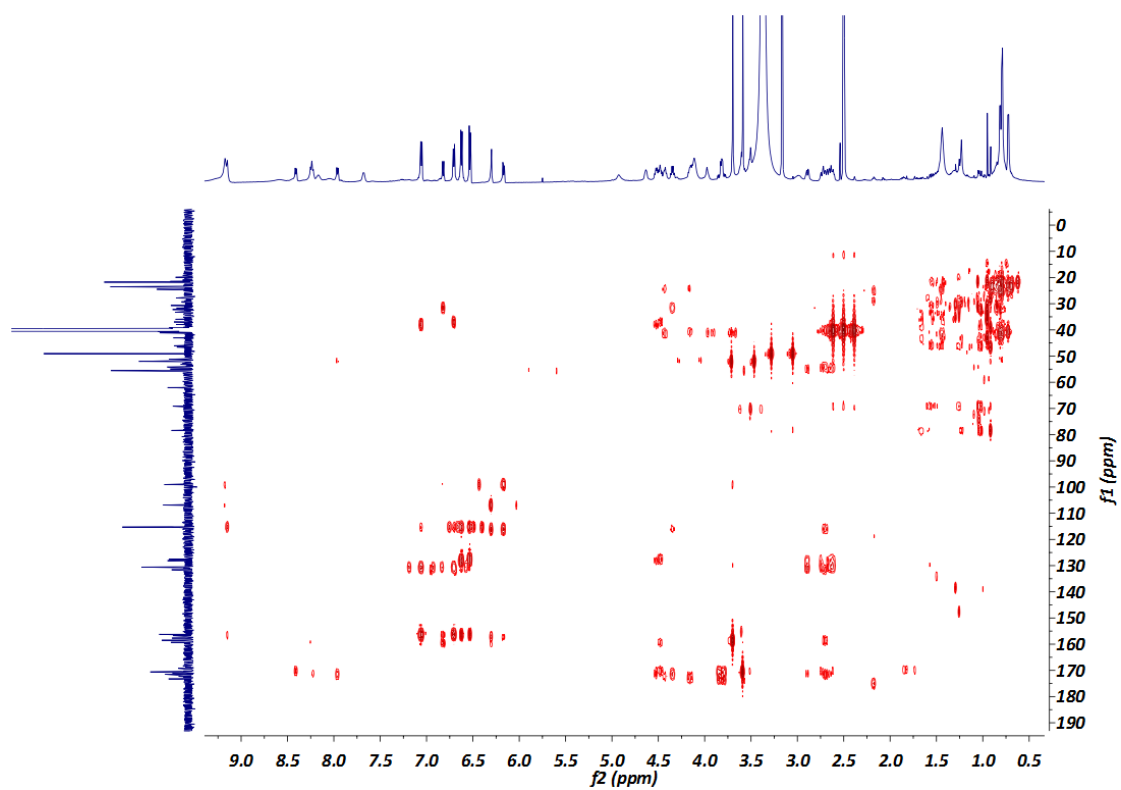


Figure S77. HMBC NMR (600 MHz, DMSO-*d*<sub>6</sub>) spectrum of compound 5.

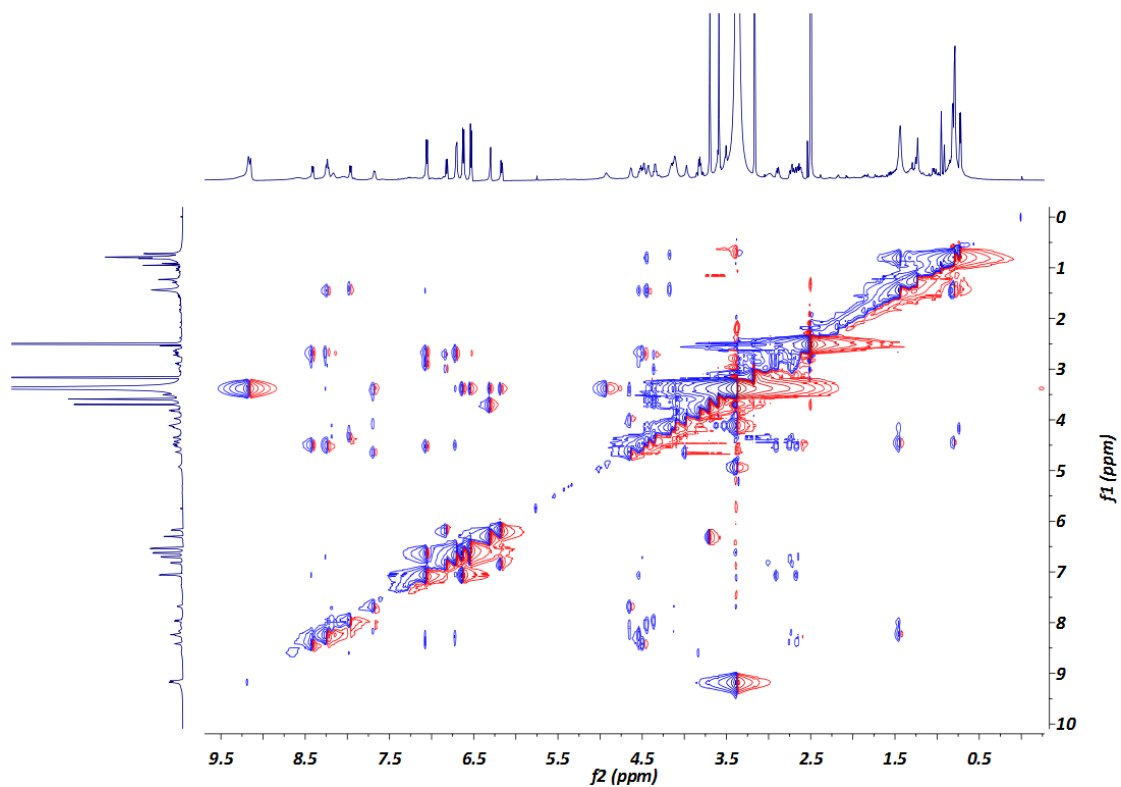


Figure S78. NOESY NMR (600 MHz, DMSO-*d*<sub>6</sub>) spectrum of compound 5.



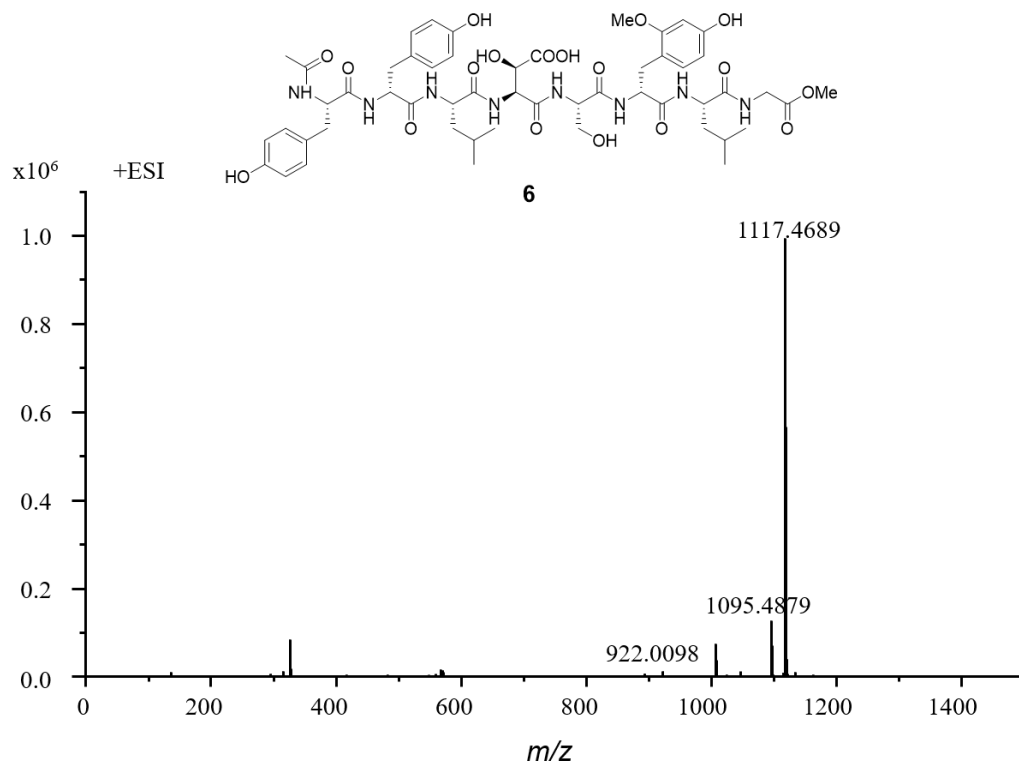


Figure S79. HRESIMS spectrum of **6**.

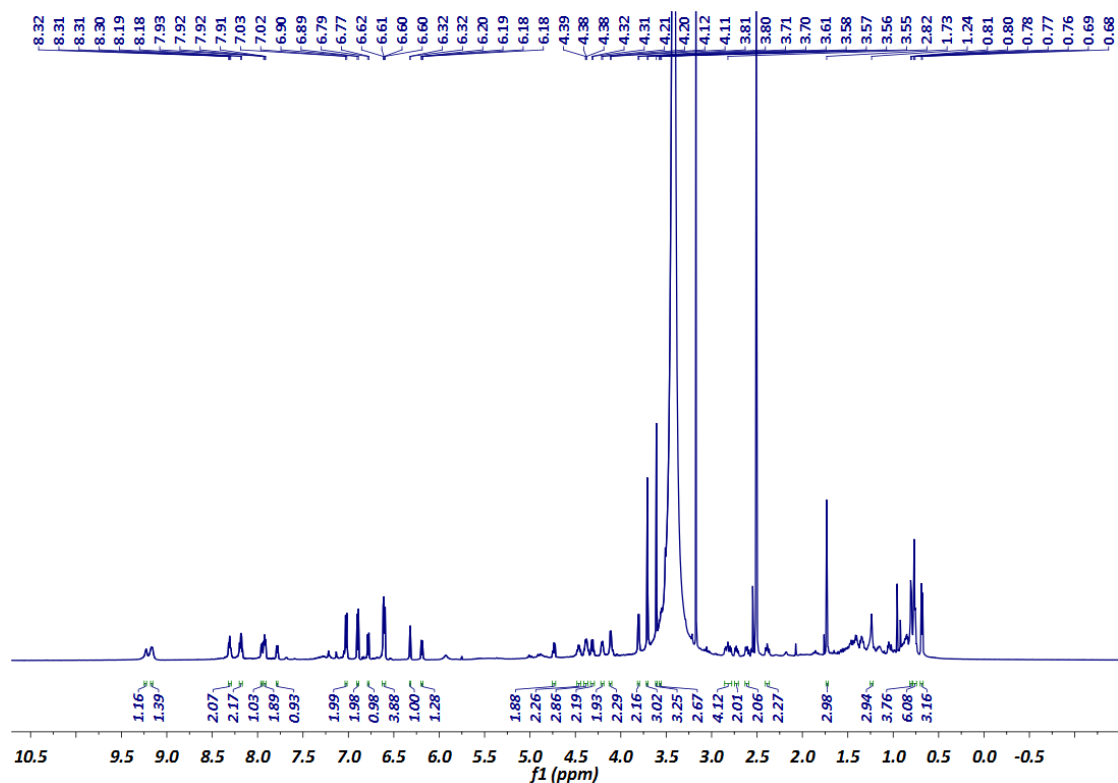


Figure S80. <sup>1</sup>H NMR (600 MHz, DMSO-*d*<sub>6</sub>) spectrum of compound **6**.

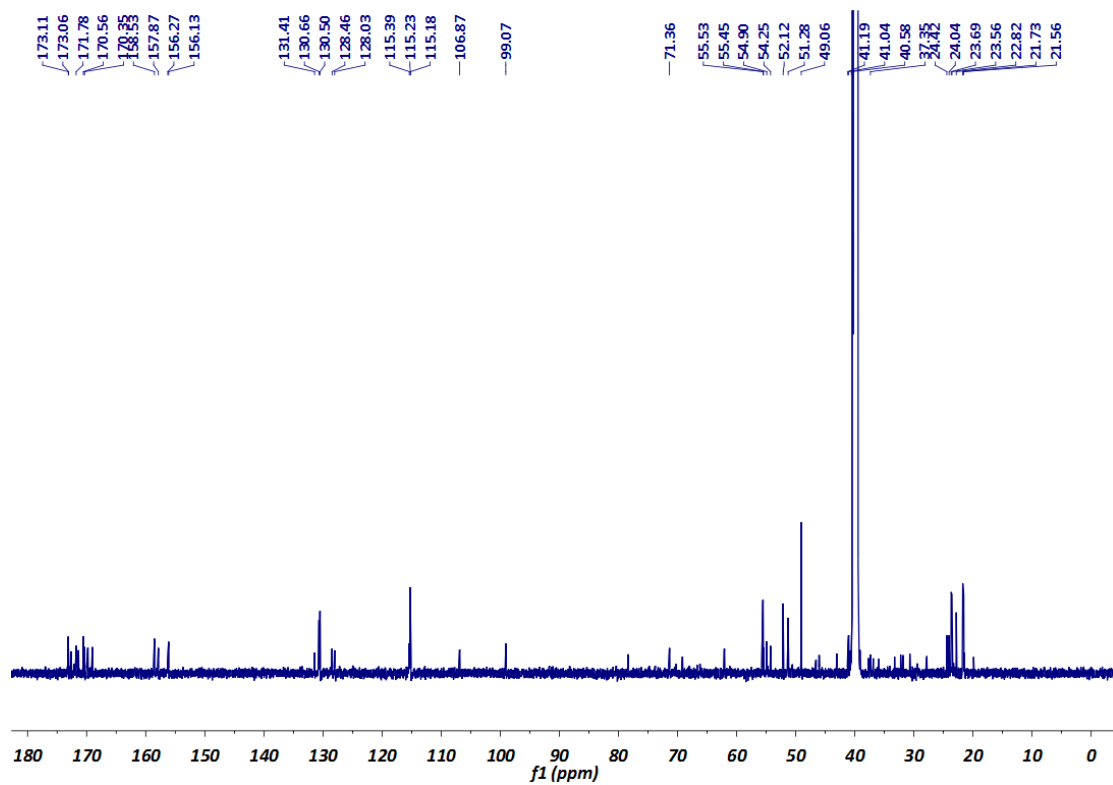


Figure S81.  $^{13}\text{C}$  NMR (150 MHz,  $\text{DMSO}-d_6$ ) spectrum of compound **6**.

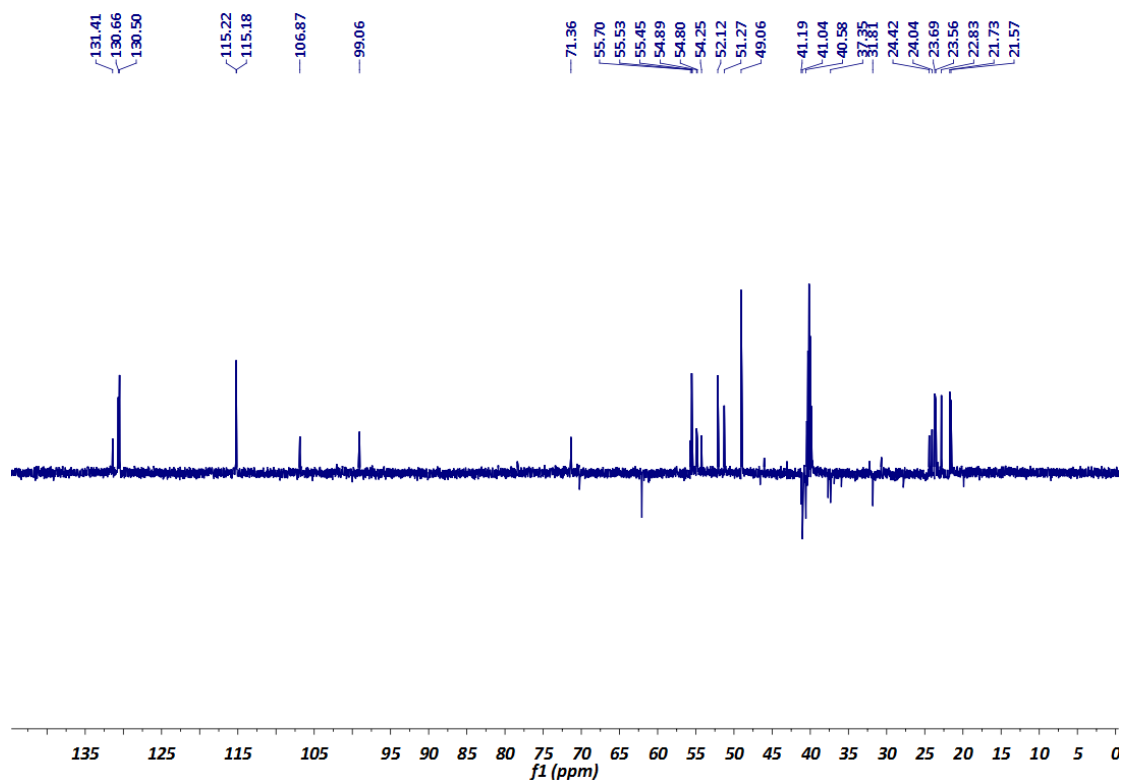


Figure S82. DEPT NMR (150 MHz,  $\text{DMSO}-d_6$ ) spectrum of compound **6**.

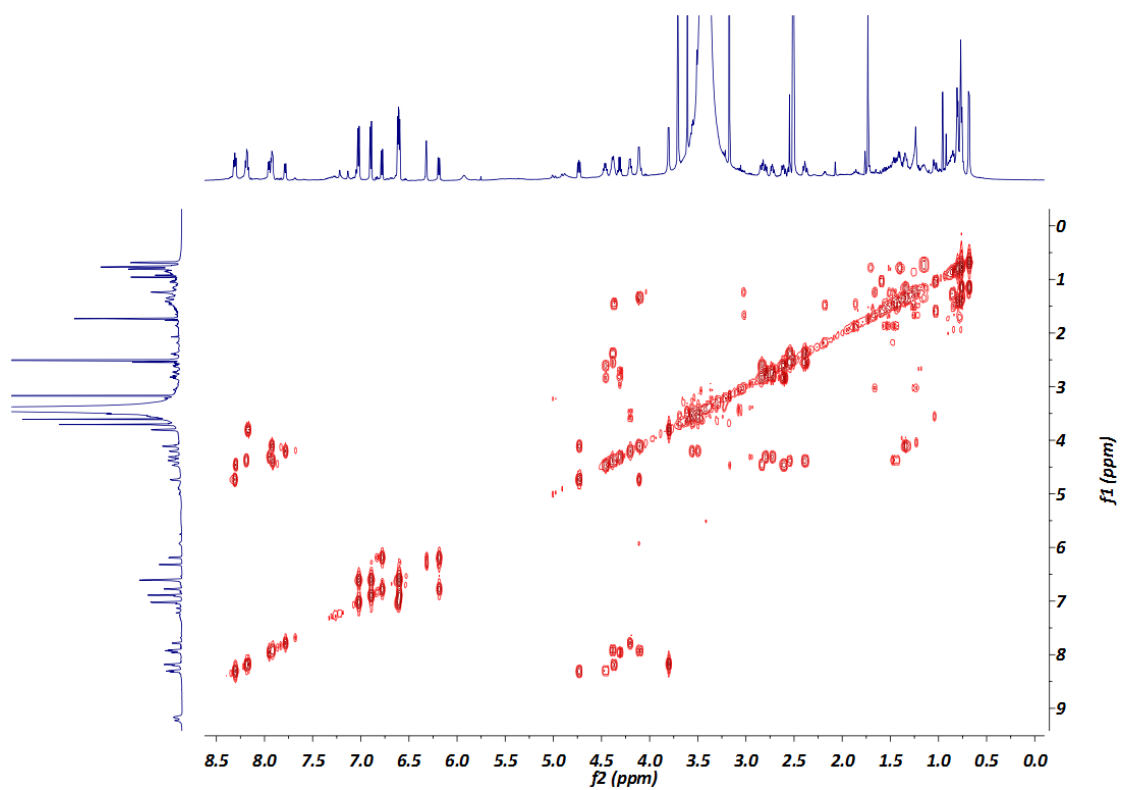


Figure S83.  $^1\text{H}$ - $^1\text{H}$  COSY NMR (600 MHz,  $\text{DMSO-}d_6$ ) spectrum of compound 6.

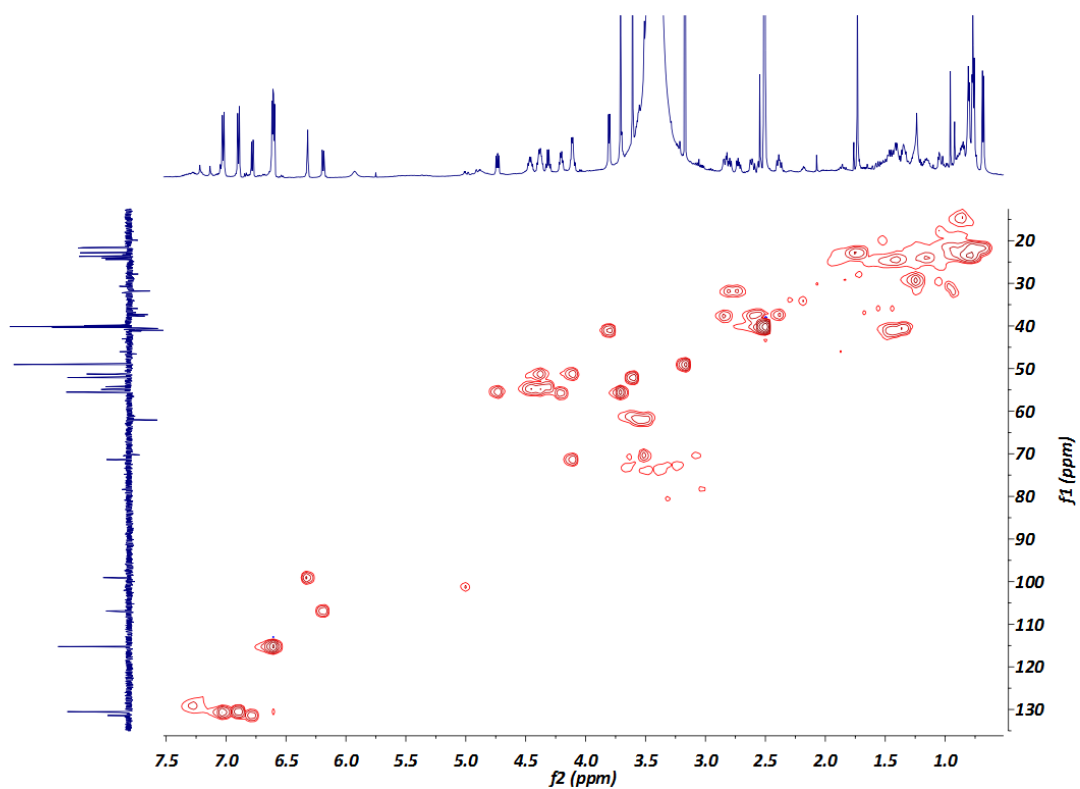


Figure S84. HSQC NMR (600 MHz,  $\text{DMSO-}d_6$ ) spectrum of compound 6.

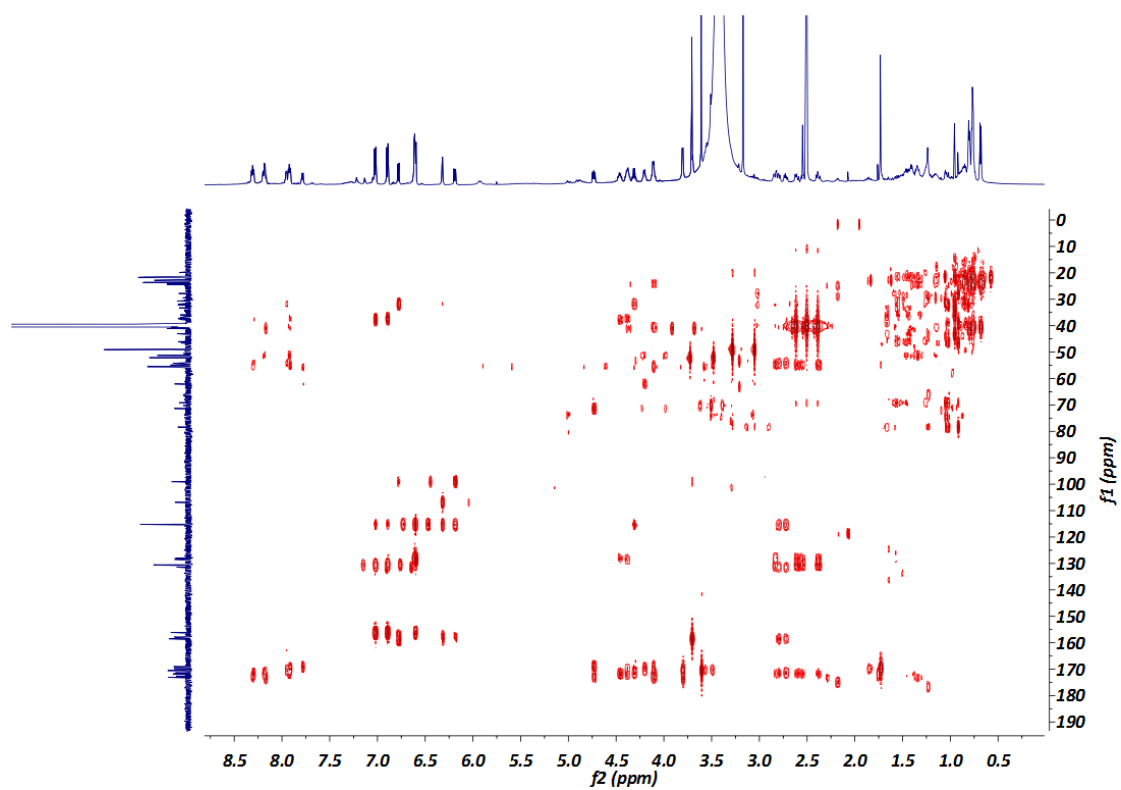


Figure S85. HMBC NMR (600 MHz, DMSO-*d*<sub>6</sub>) spectrum of compound 6.

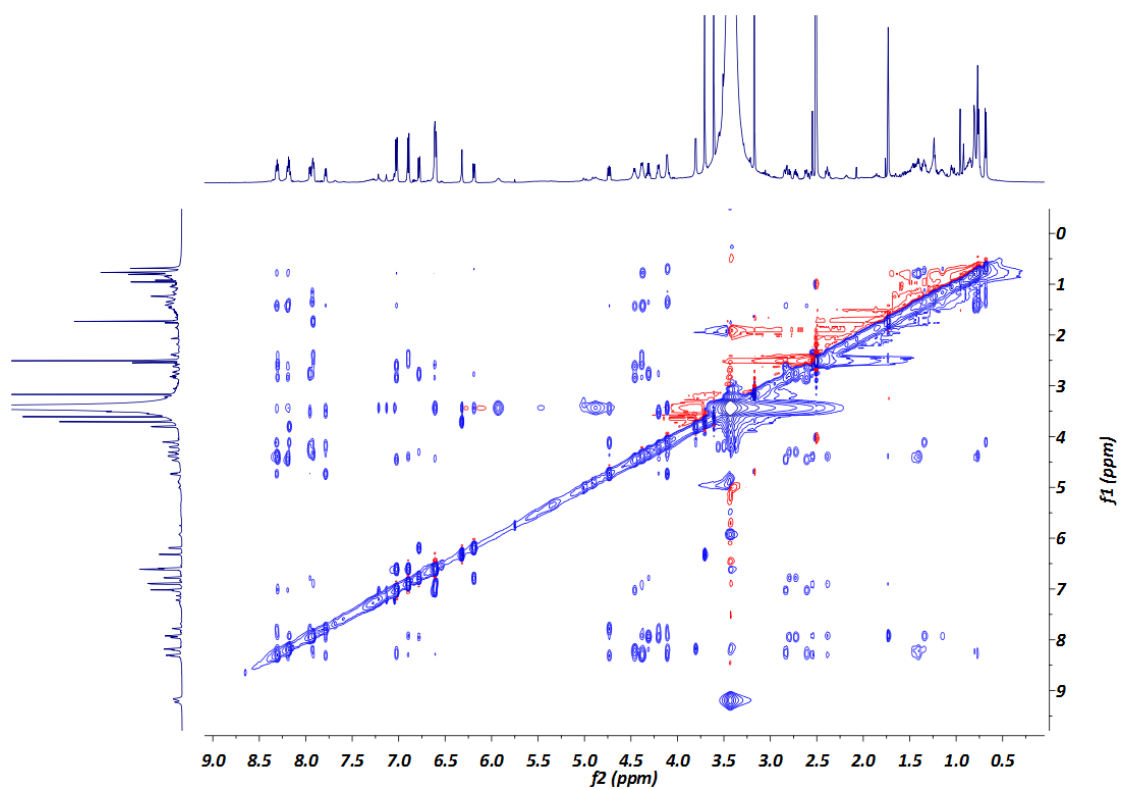


Figure S86. NOESY NMR (600 MHz, DMSO-*d*<sub>6</sub>) spectrum of compound 6.

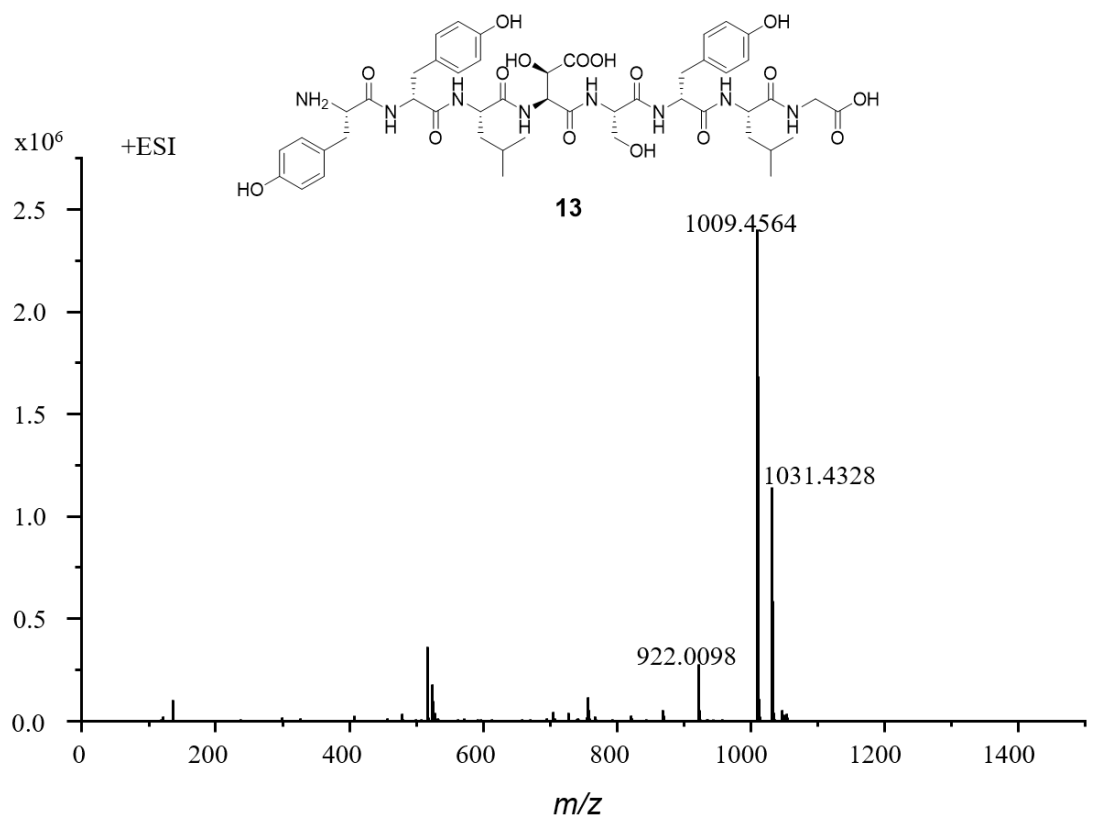


Figure S87. HRESIMS spectrum of **13**.

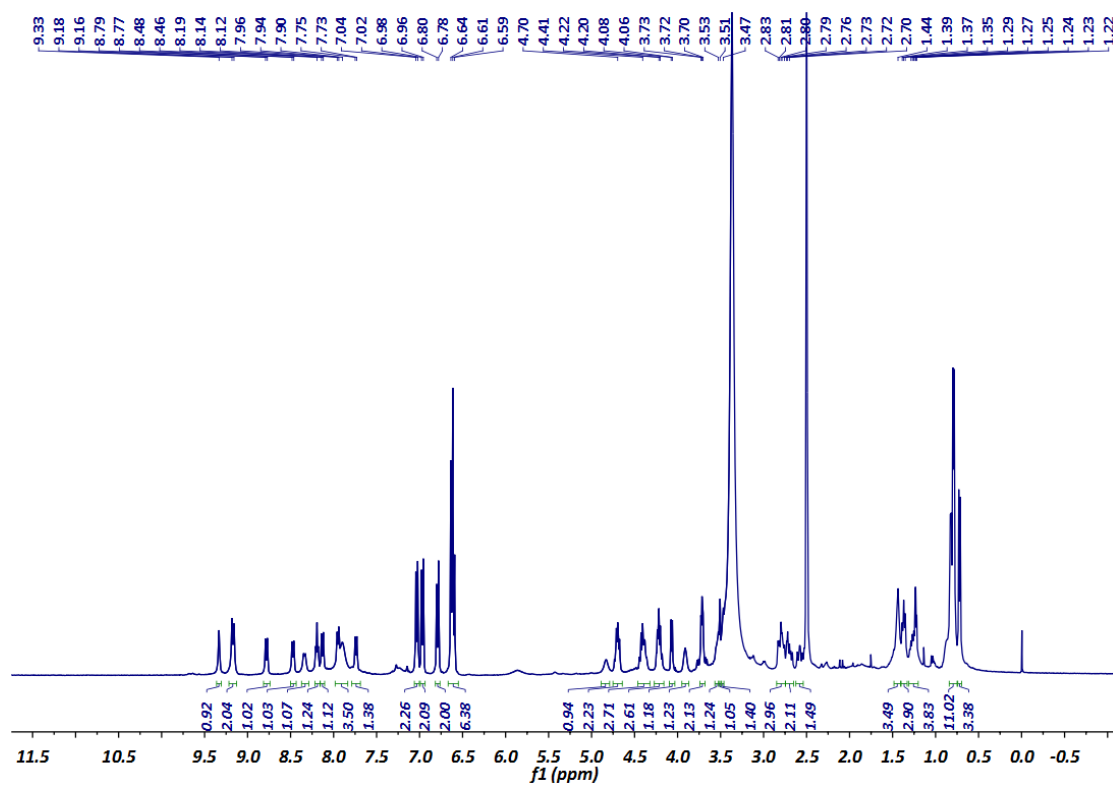


Figure S88.  $^1\text{H}$  NMR (600 MHz,  $\text{DMSO}-d_6$ ) spectrum of compound **13**.

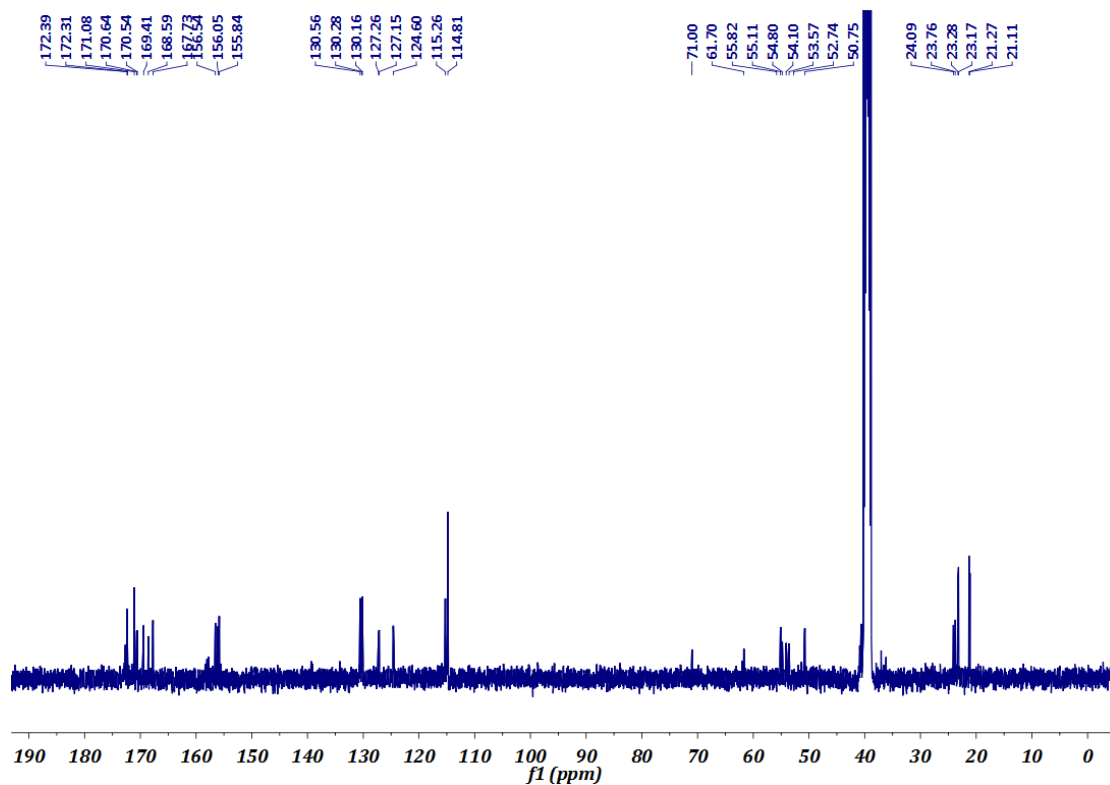


Figure S89.  $^{13}\text{C}$  NMR (150 MHz,  $\text{DMSO}-d_6$ ) spectrum of compound **13**.

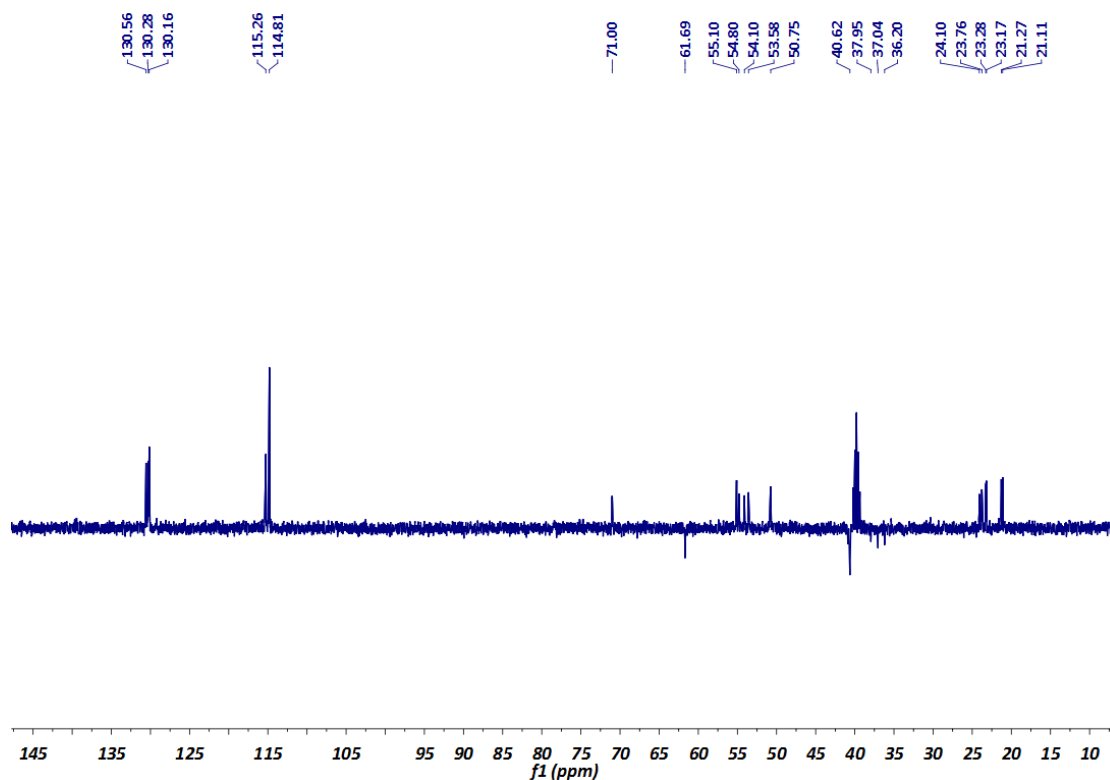


Figure S90. DEPT NMR (150 MHz,  $\text{DMSO}-d_6$ ) spectrum of compound **13**.

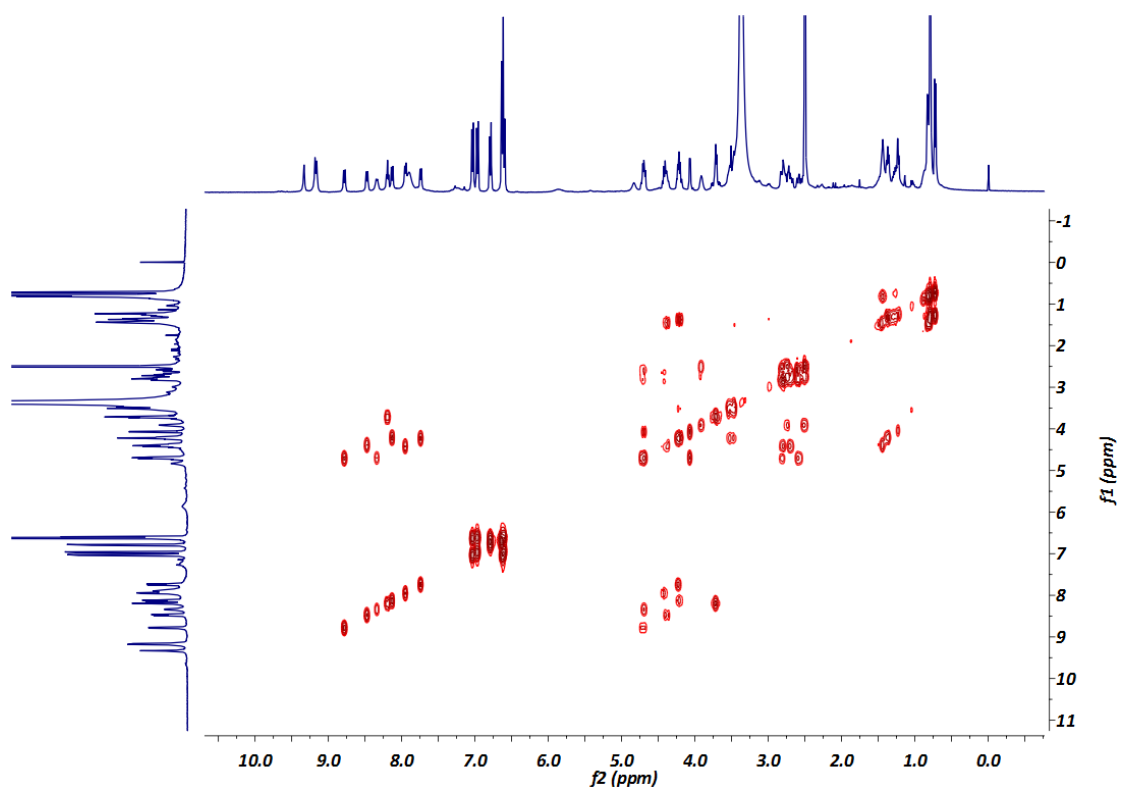


Figure S91.  $^1\text{H}$ - $^1\text{H}$  COSY NMR (600 MHz,  $\text{DMSO-}d_6$ ) spectrum of compound **13**.

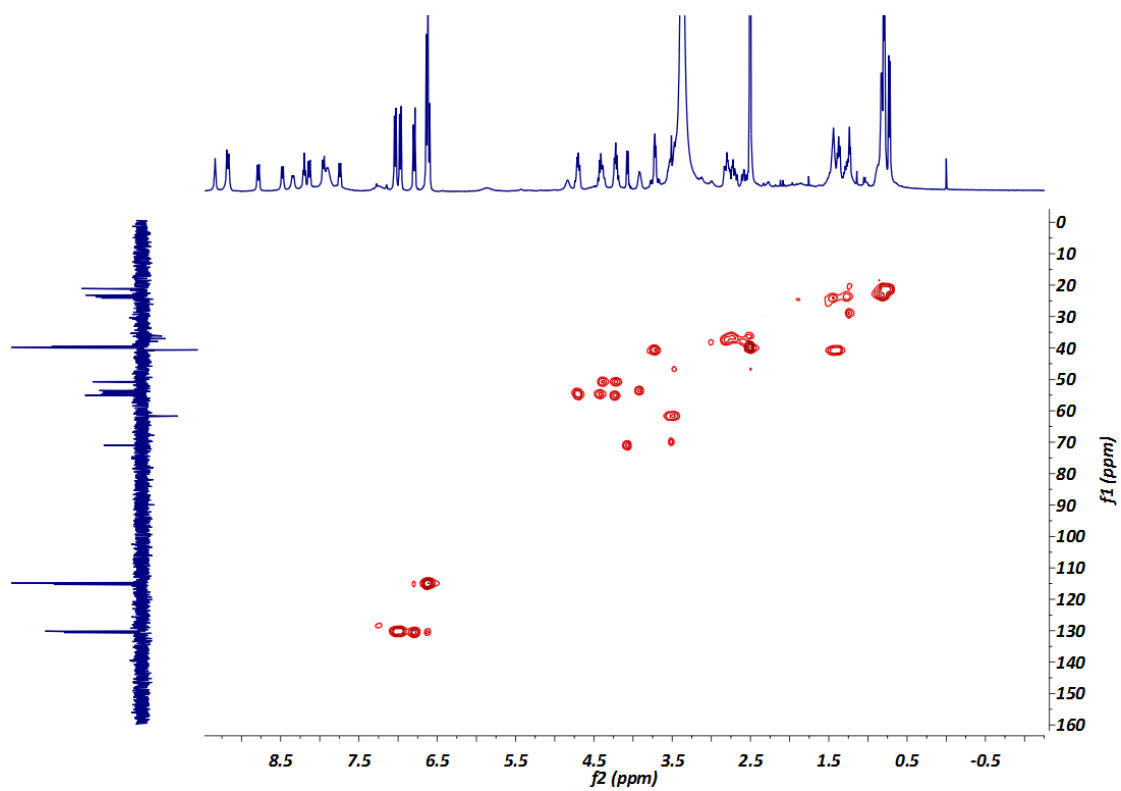


Figure S92. HSQC NMR (600 MHz,  $\text{DMSO-}d_6$ ) spectrum of compound **13**.

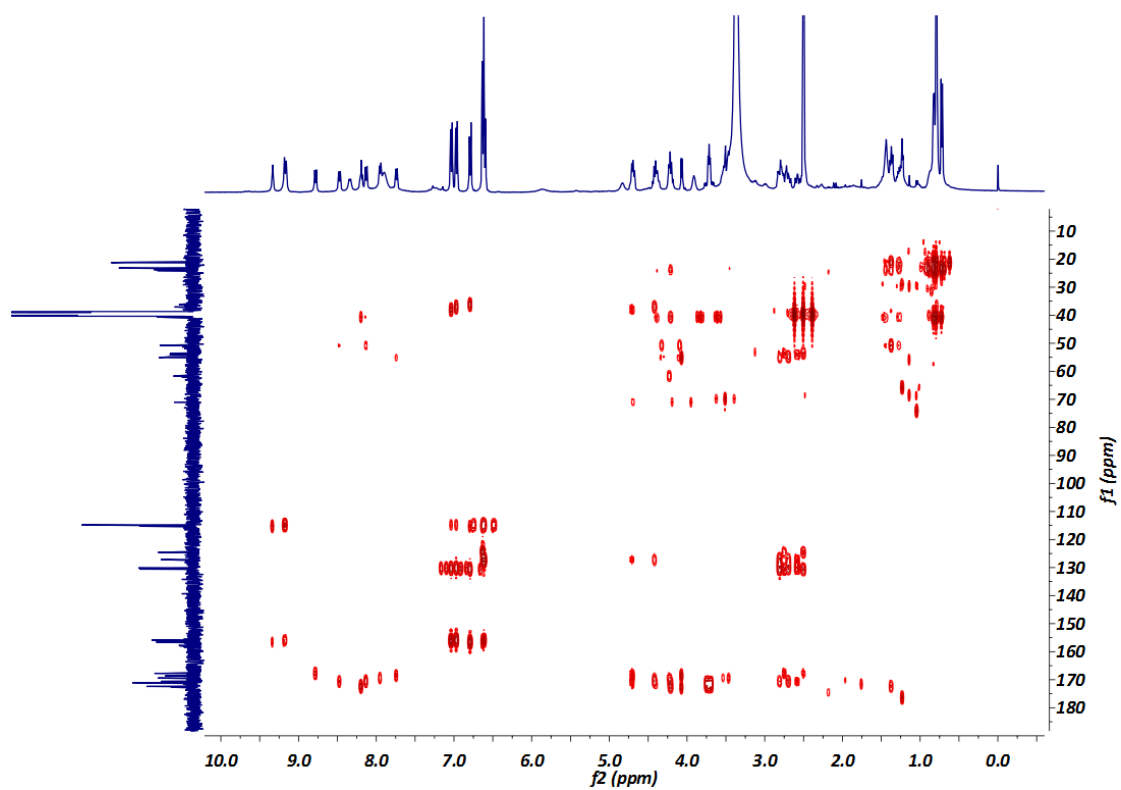


Figure S93. HMBC NMR (600 MHz, DMSO- $d_6$ ) spectrum of compound **13**.

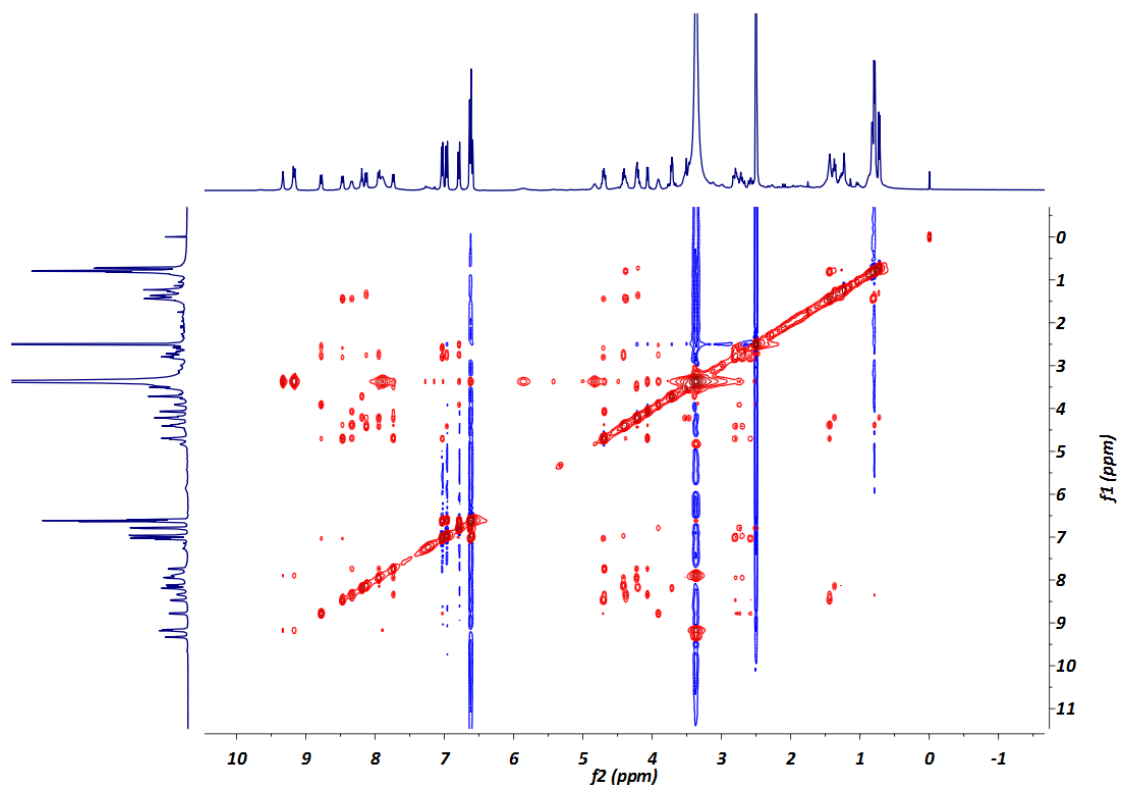


Figure S94. NOESY NMR (600 MHz, DMSO- $d_6$ ) spectrum of compound **13**.



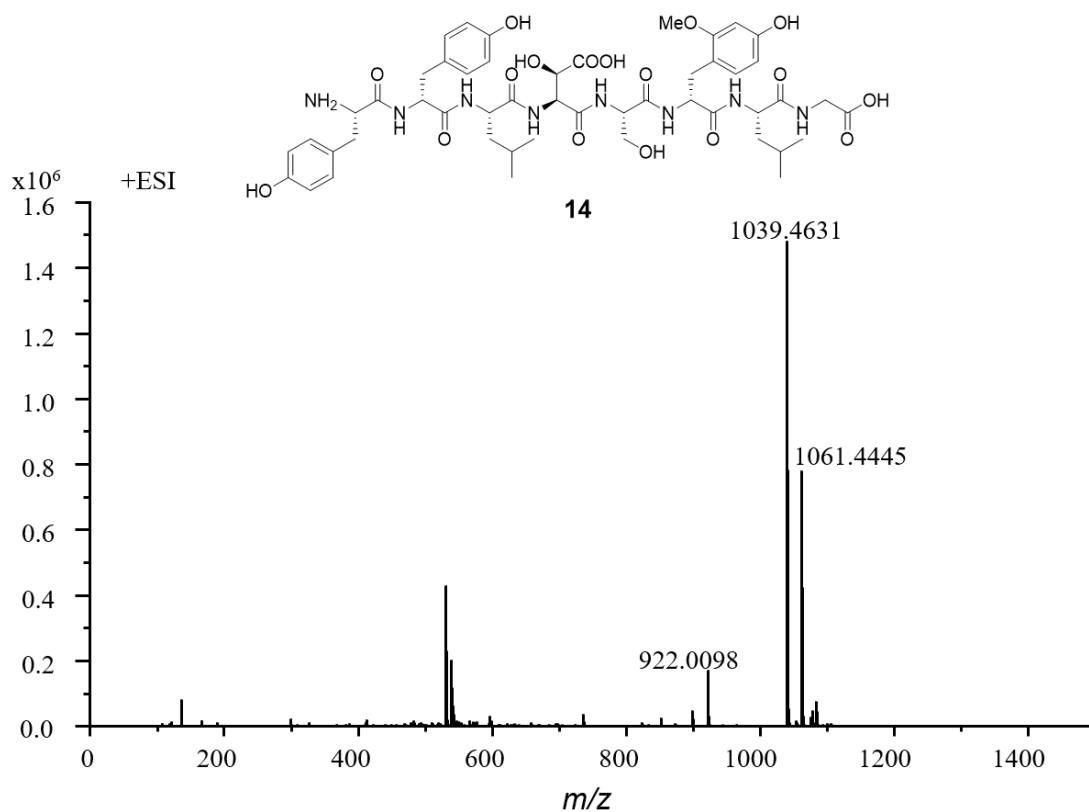


Figure S95. HRESIMS spectrum of **14**.

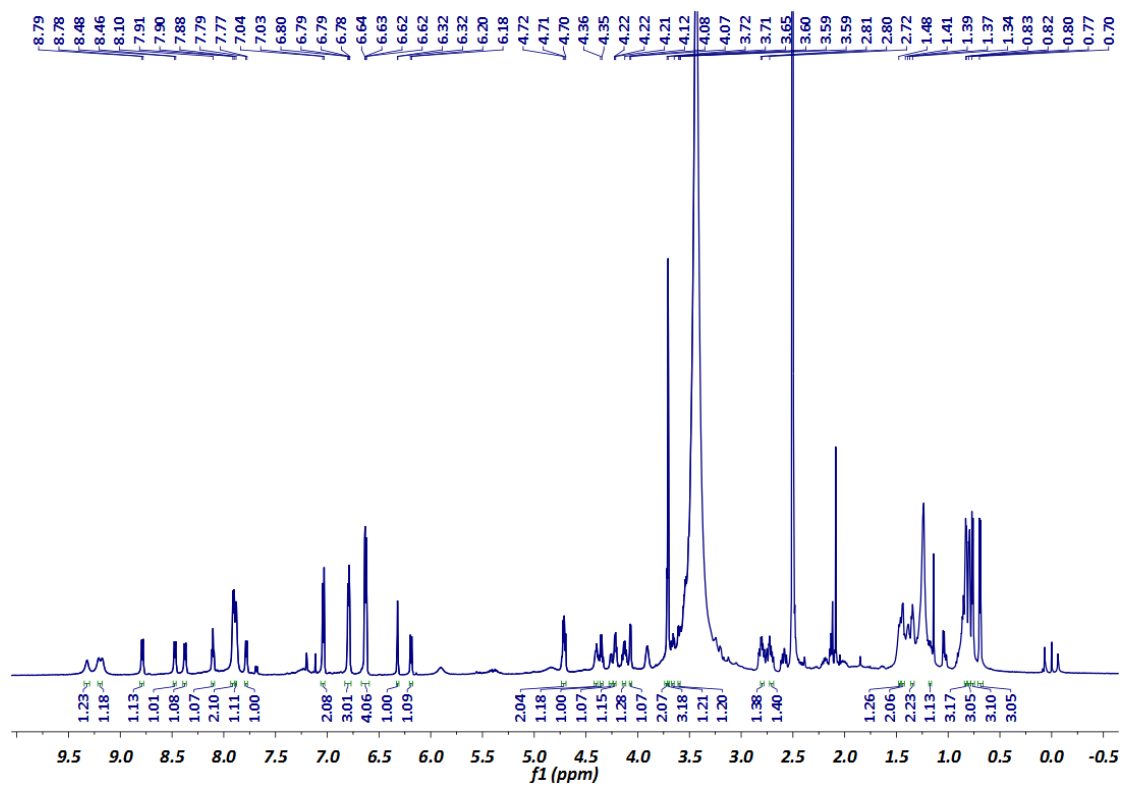


Figure S96.  $^1\text{H}$  NMR (600 MHz,  $\text{DMSO-}d_6$ ) spectrum of compound **14**.

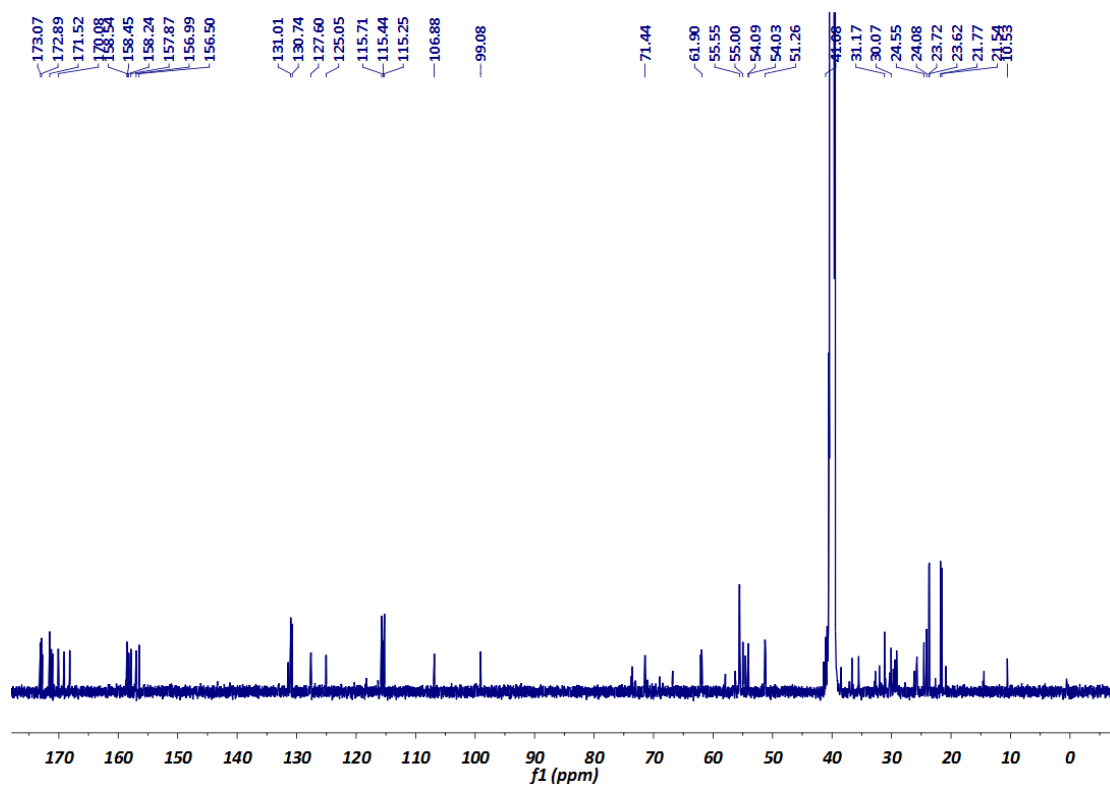


Figure S97.  $^{13}\text{C}$  NMR (150 MHz,  $\text{DMSO}-d_6$ ) spectrum of compound **14**.

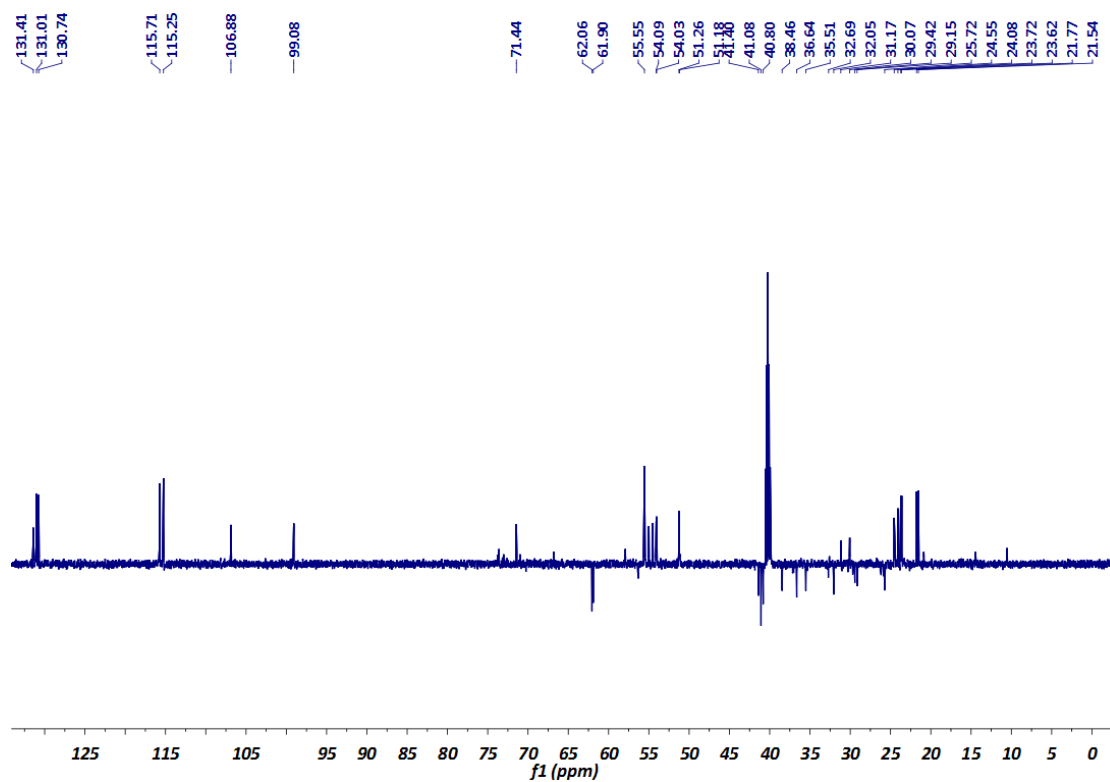


Figure S98. DEPT NMR (150 MHz,  $\text{DMSO}-d_6$ ) spectrum of compound **14**.

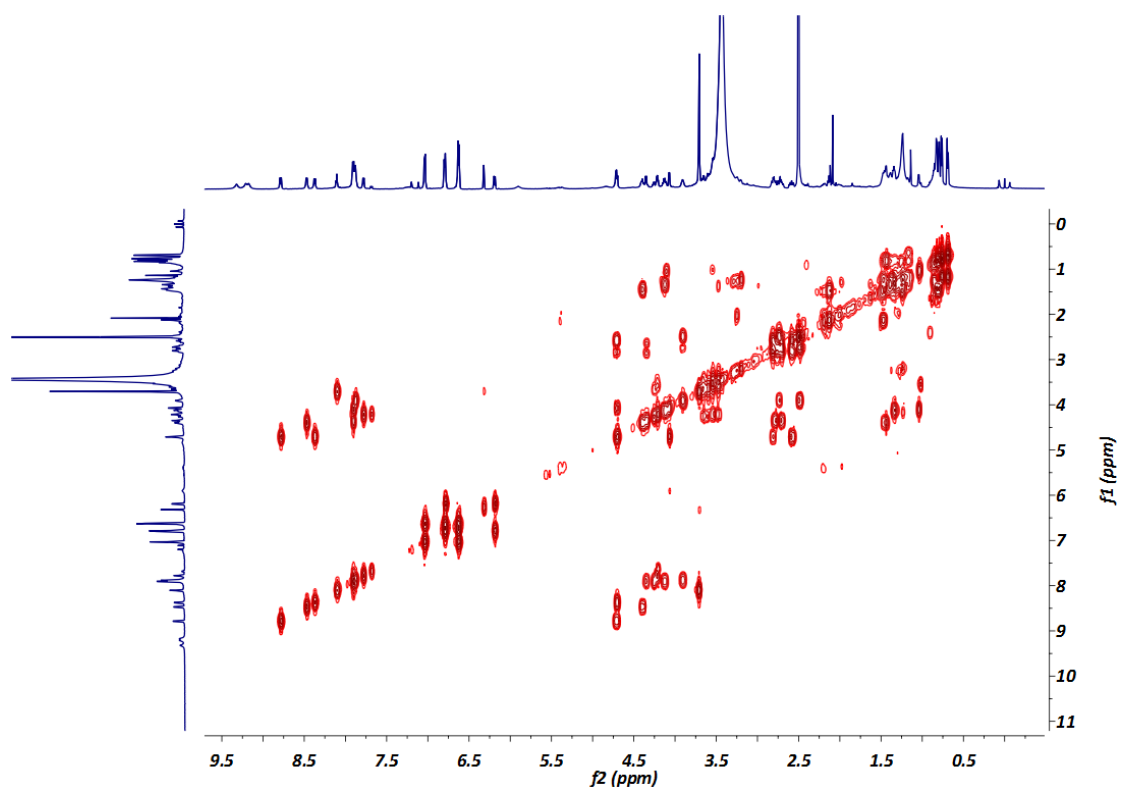


Figure S99.  $^1\text{H}$ - $^1\text{H}$  COSY NMR (600 MHz,  $\text{DMSO}-d_6$ ) spectrum of compound **14**.

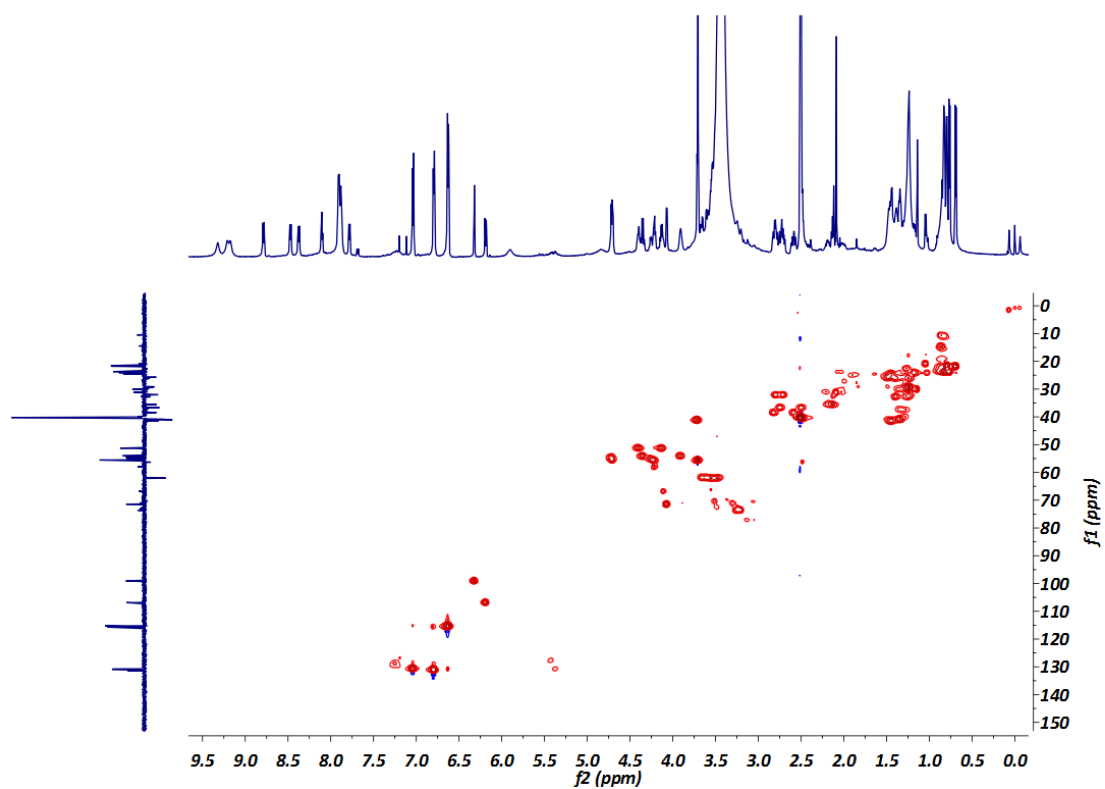


Figure S100. HSQC NMR (600 MHz,  $\text{DMSO}-d_6$ ) spectrum of compound **14**.

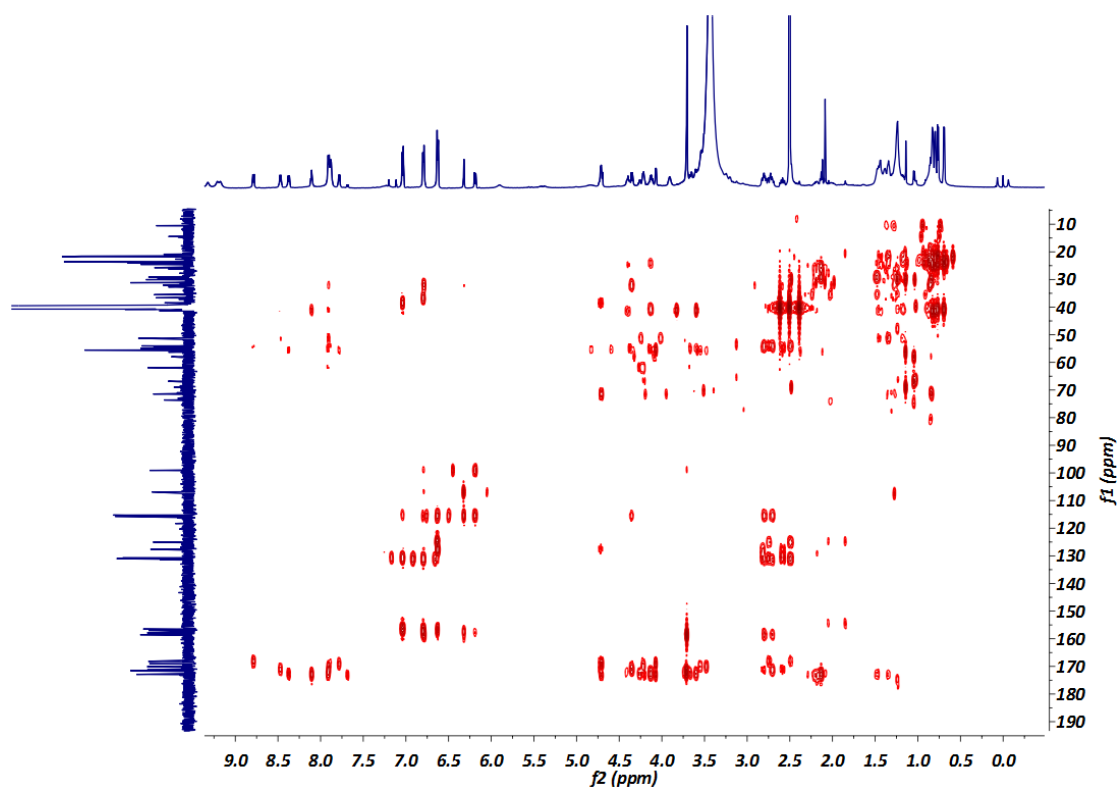


Figure S101. HMBC NMR (600 MHz, DMSO- $d_6$ ) spectrum of compound **14**.

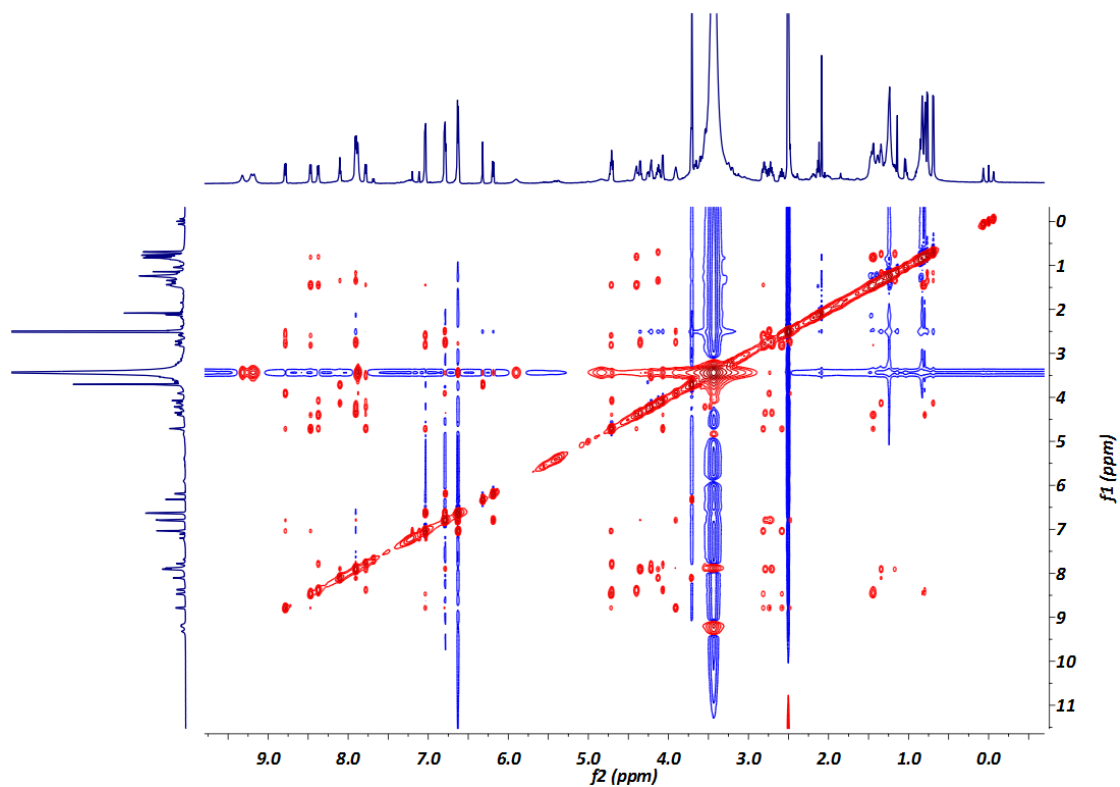


Figure S102. NOESY NMR (600 MHz, DMSO- $d_6$ ) spectrum of compound **14**.

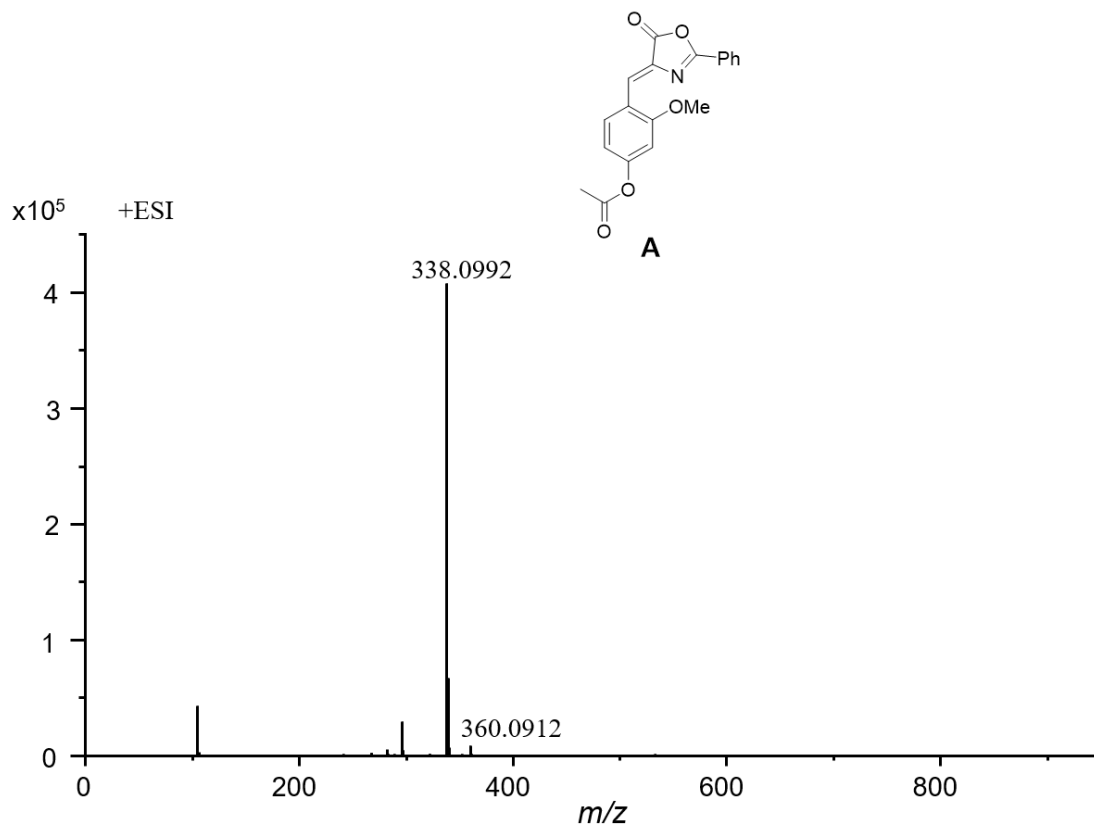


Figure S103. HRESIMS spectrum of A.

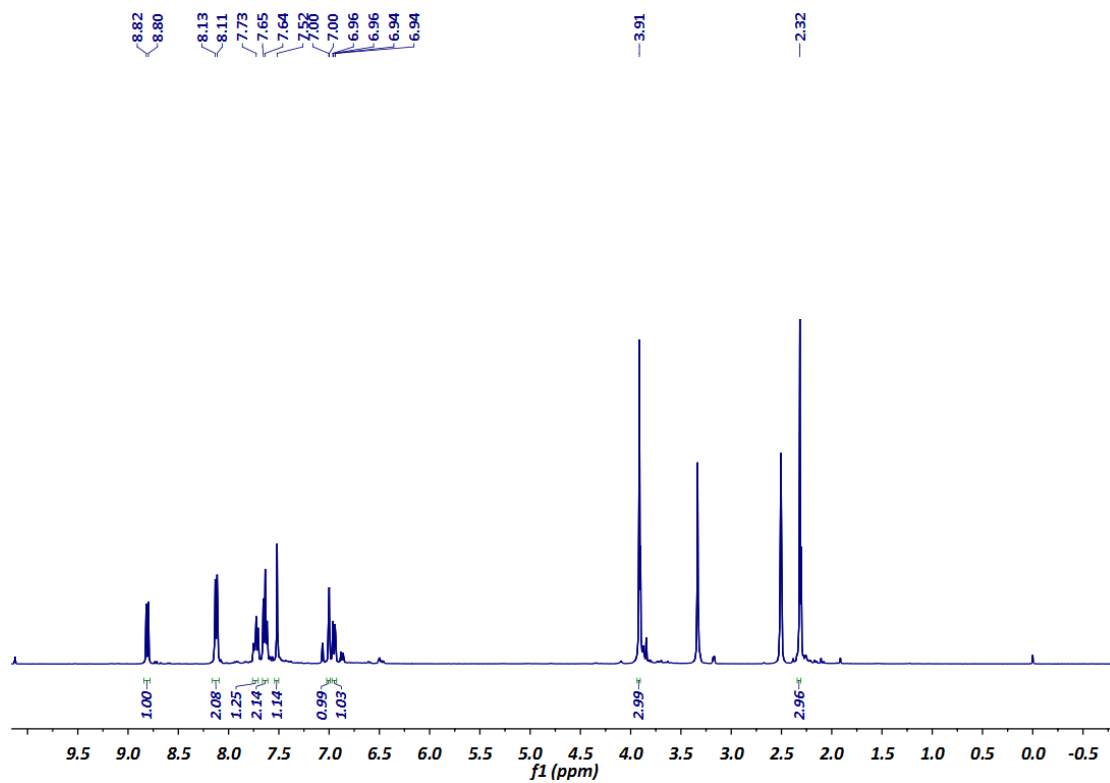


Figure S104. <sup>1</sup>H NMR (400 MHz, DMSO-d<sub>6</sub>) spectrum of compound A.

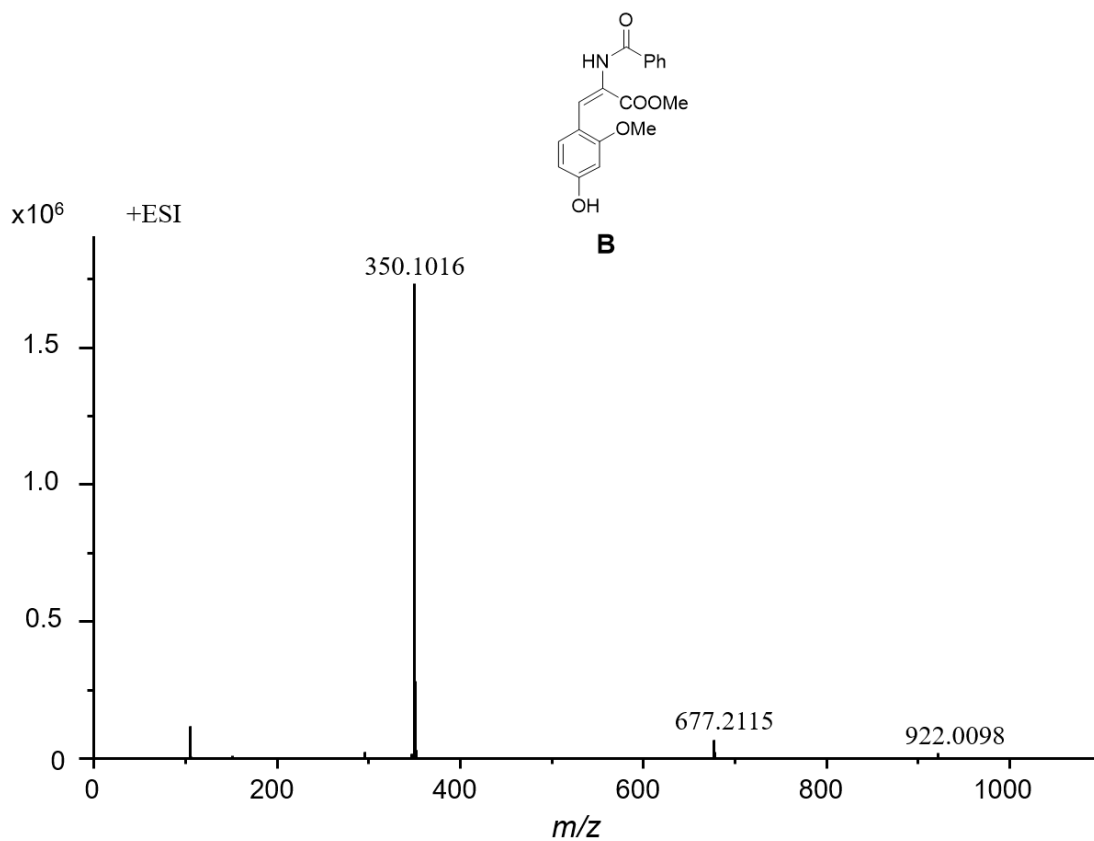


Figure S105. HRESIMS spectrum of **B**.

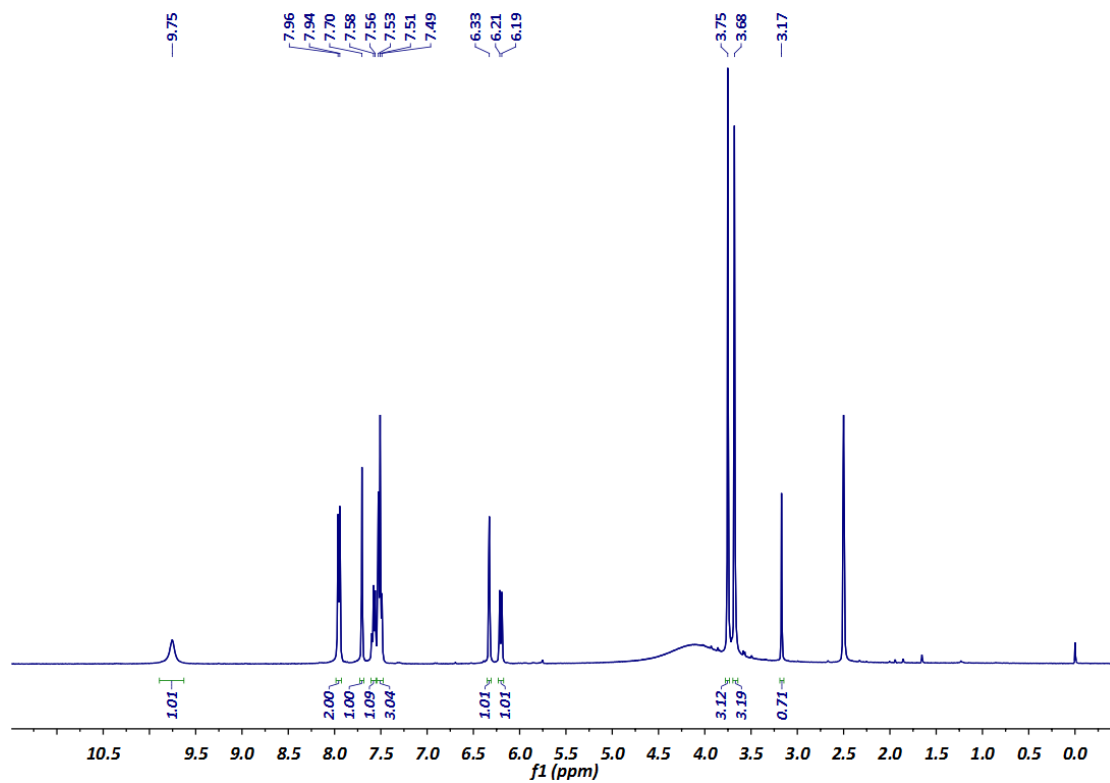


Figure S106.  $^1\text{H}$  NMR (400 MHz,  $\text{DMSO}-d_6$ ) spectrum of compound **B**.

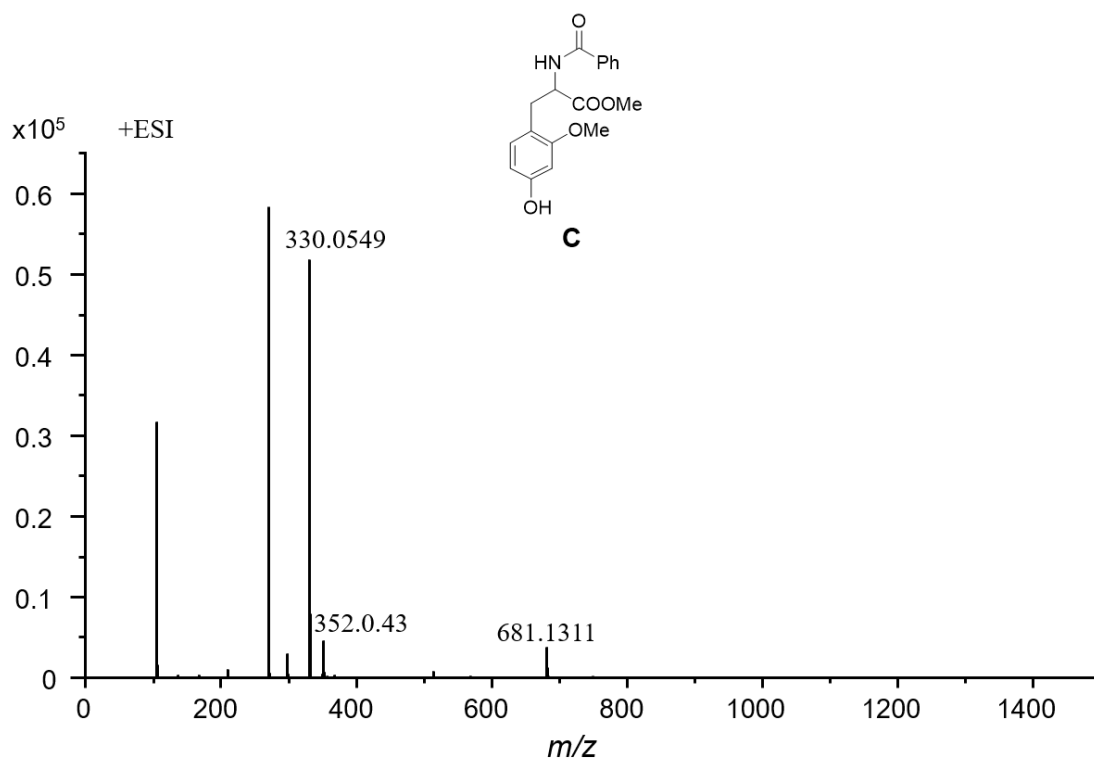


Figure S107. HRESIMS spectrum of **C**.

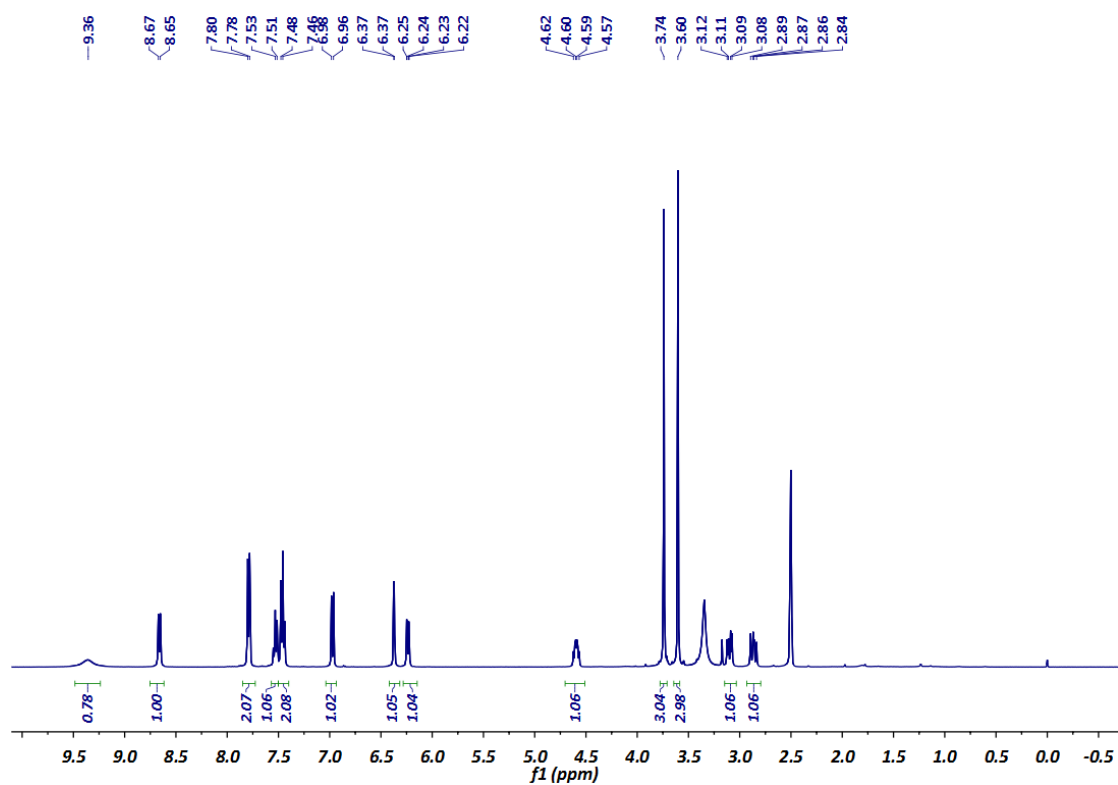


Figure S108.  $^1\text{H}$  NMR (400 MHz,  $\text{DMSO}-d_6$ ) spectrum of compound **C**.

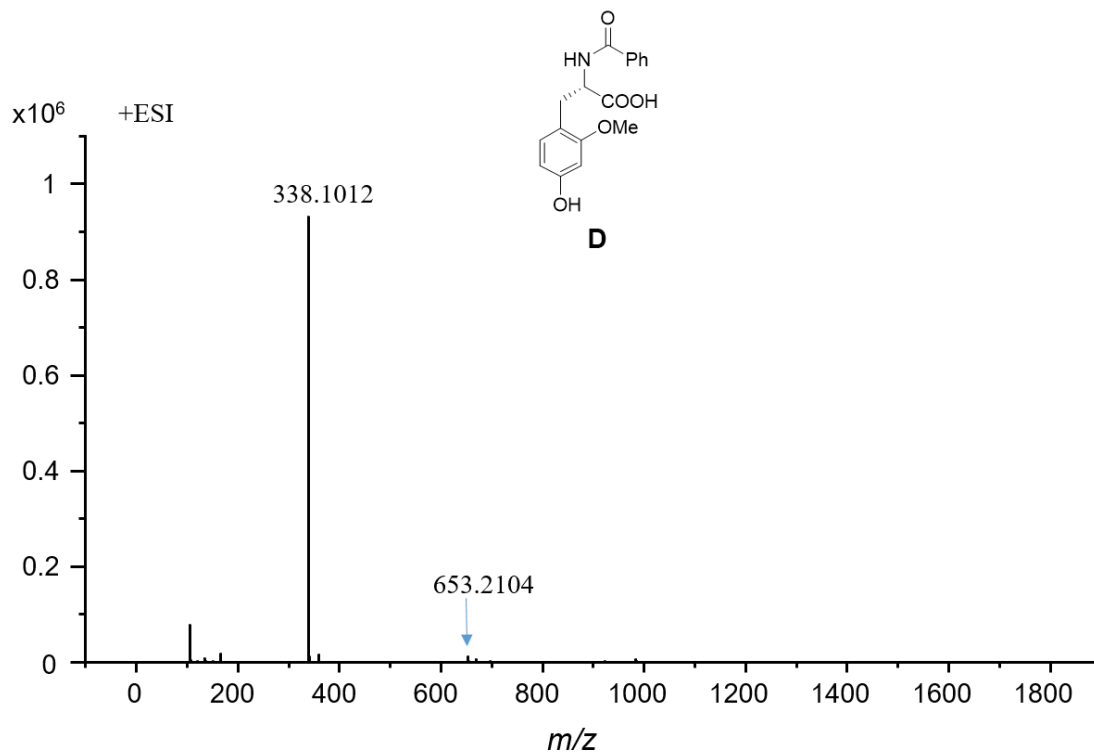


Figure S109. HRESIMS spectrum of **D**.

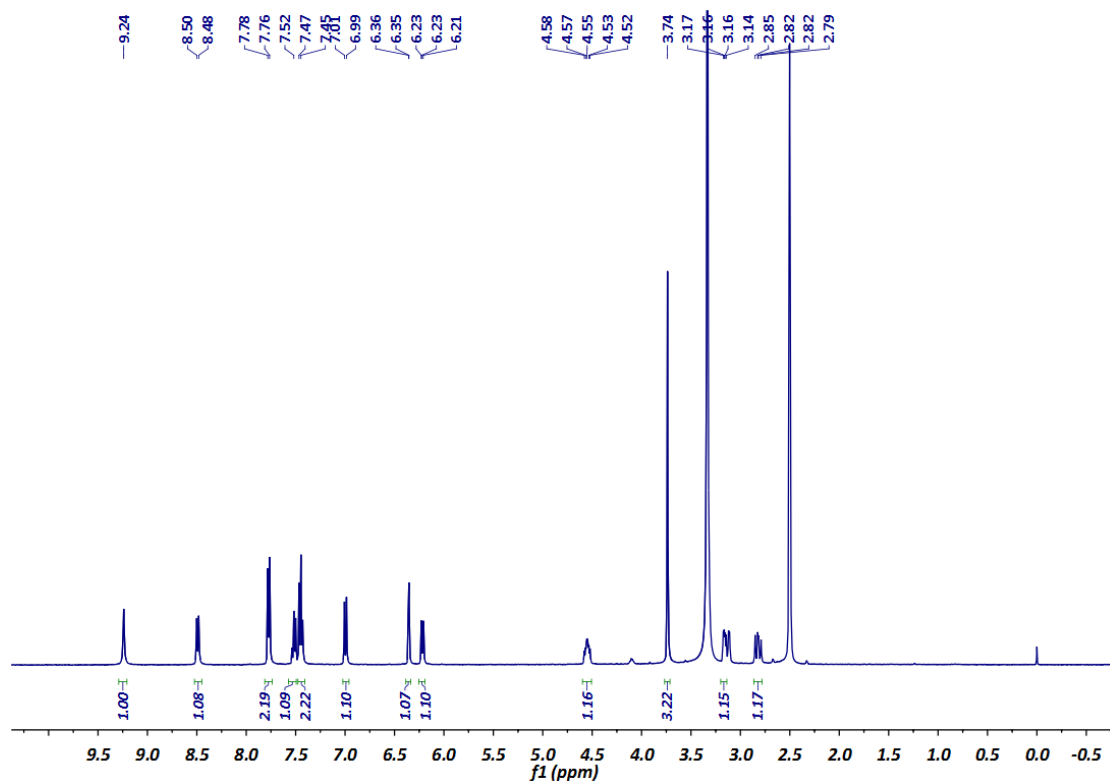


Figure S110.  $^1\text{H}$  NMR (400 MHz,  $\text{DMSO}-d_6$ ) spectrum of compound **D**.



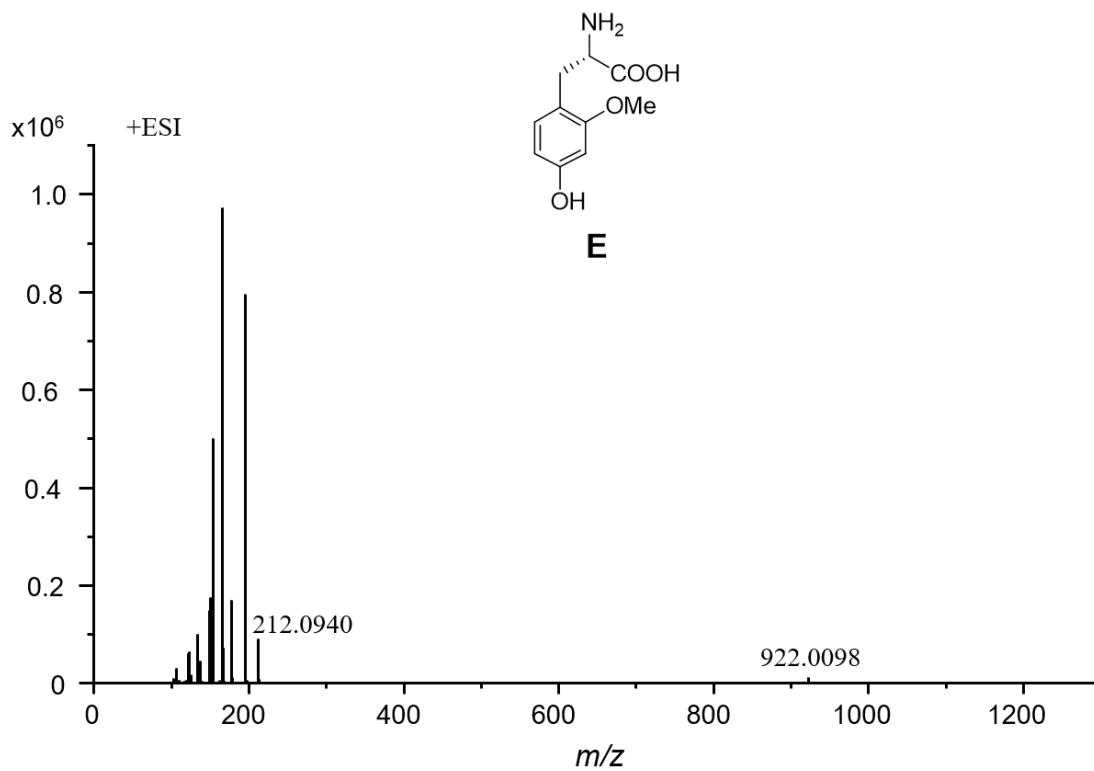


Figure S111. HRESIMS spectrum of E.

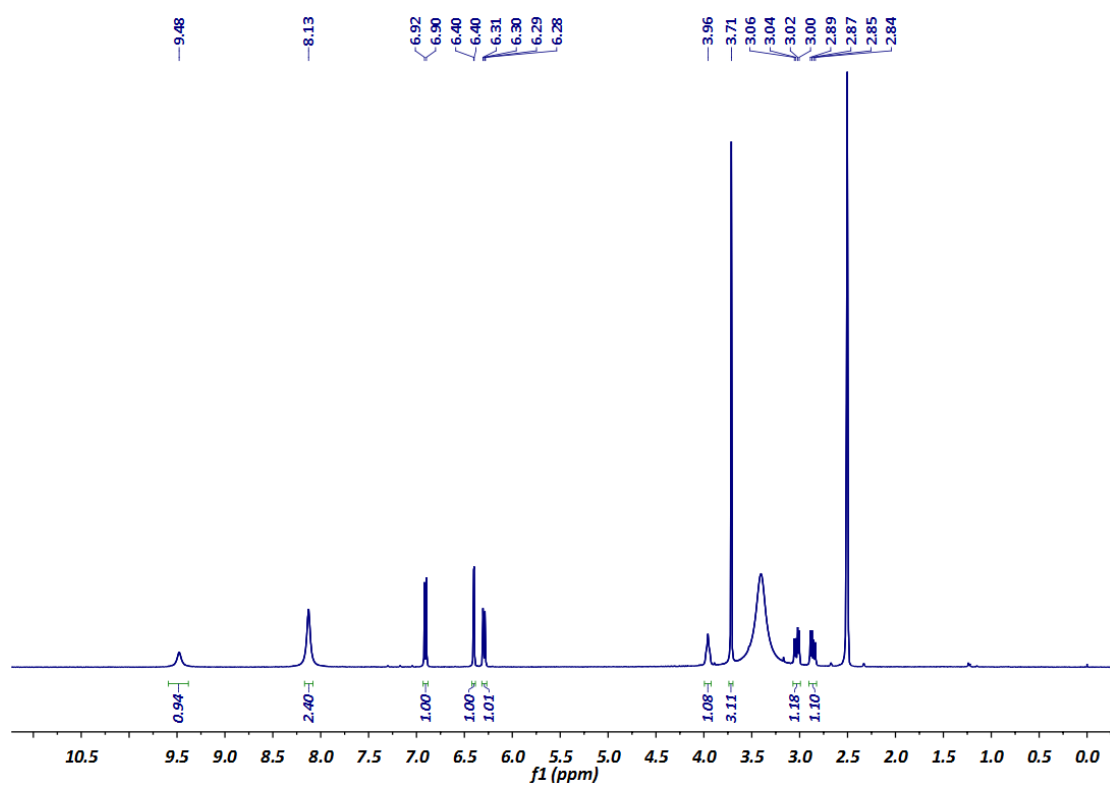


Figure S112. <sup>1</sup>H NMR (400 MHz, DMSO-d<sub>6</sub>) spectrum of compound E.

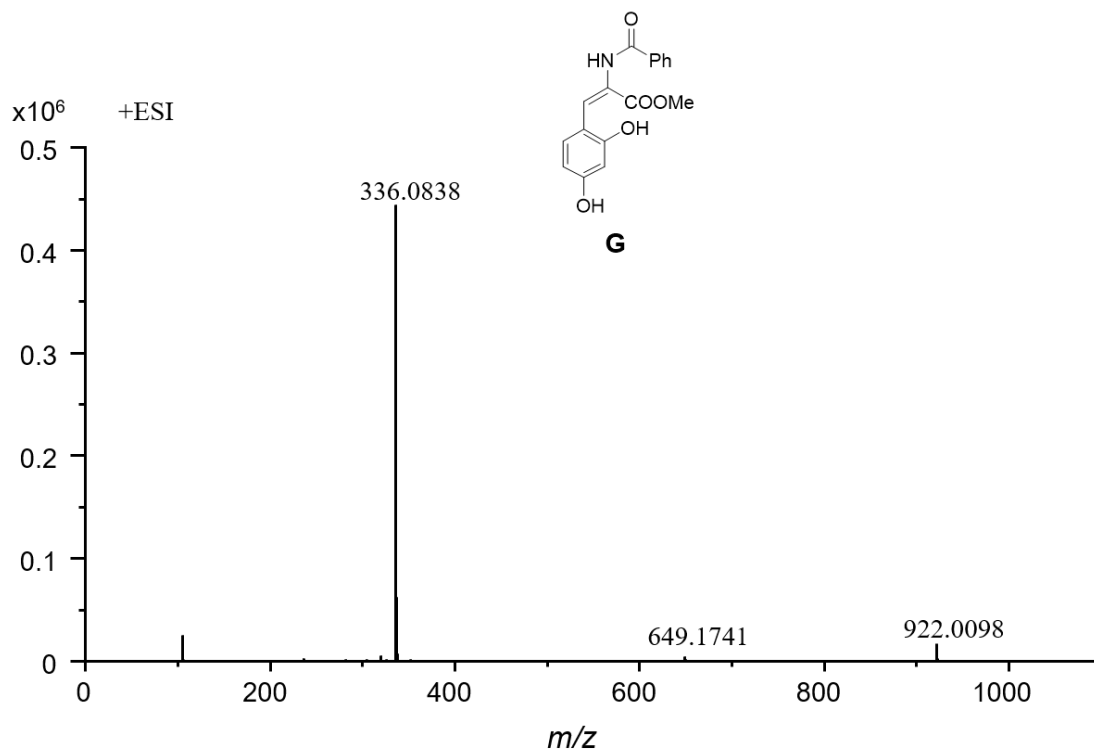


Figure S113. HRESIMS spectrum of **G**.

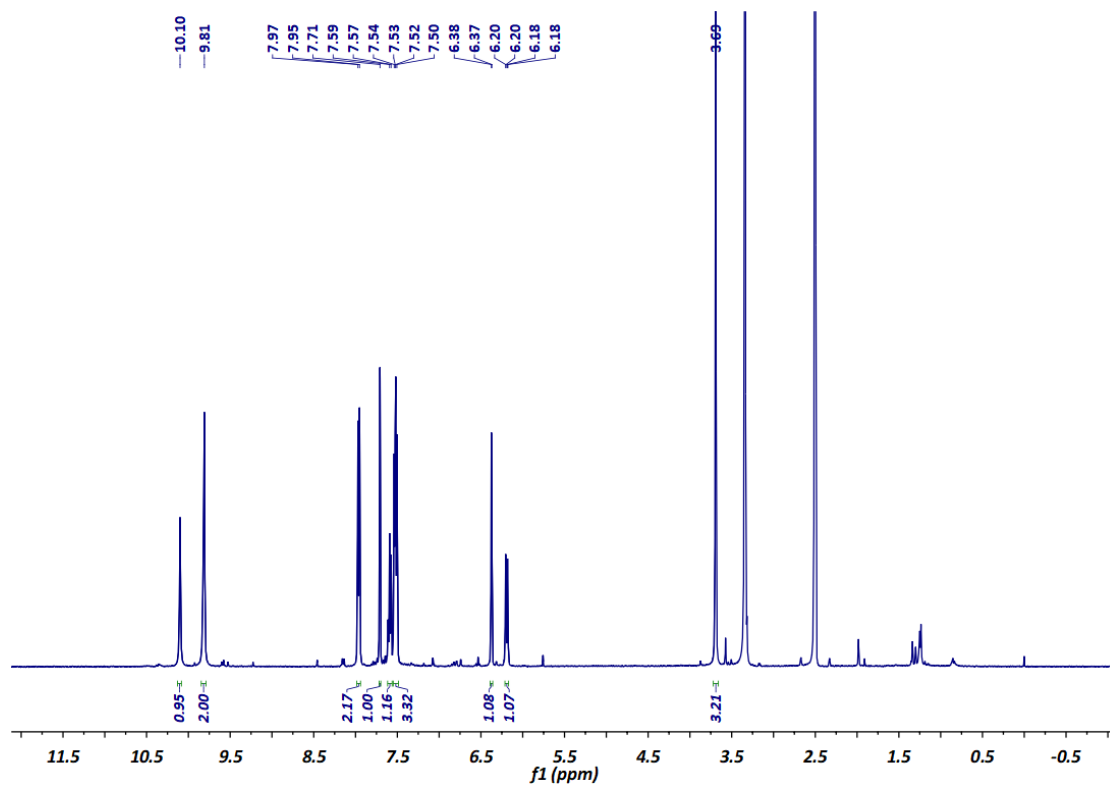


Figure S114.  $^1\text{H}$  NMR (400 MHz,  $\text{DMSO-}d_6$ ) spectrum of compound **G**.

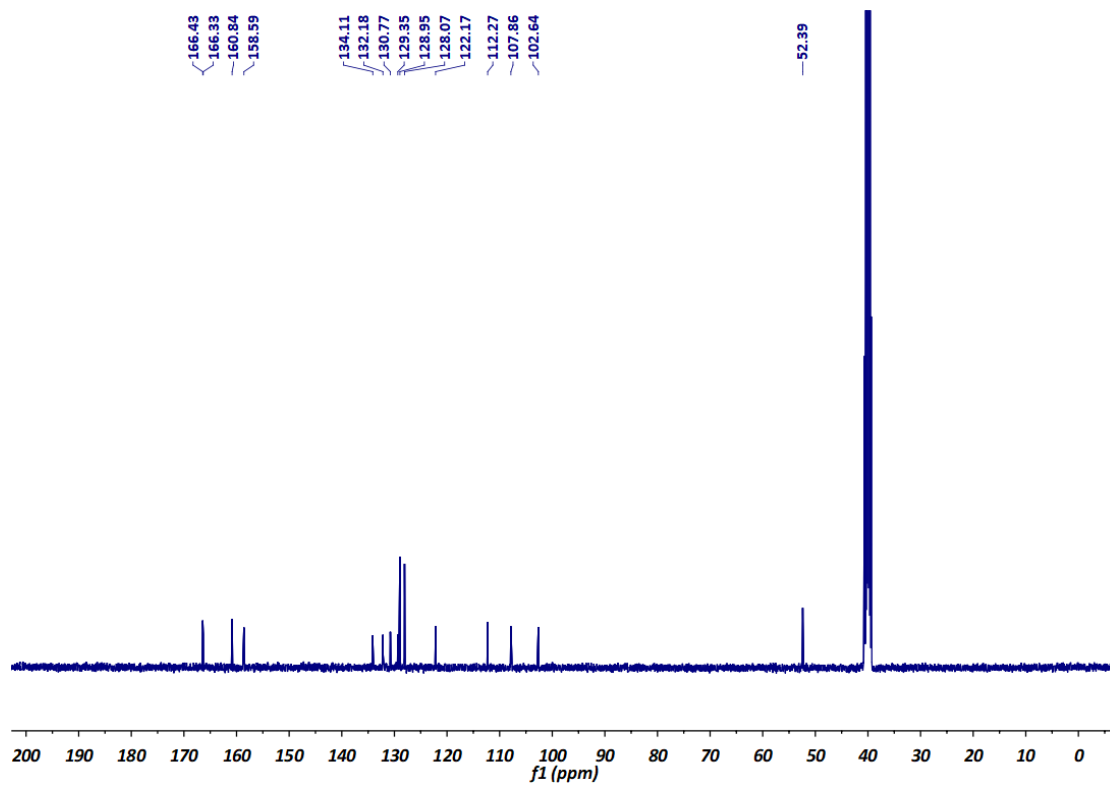


Figure S115.  $^{13}\text{C}$  NMR (100 MHz,  $\text{DMSO-}d_6$ ) spectrum of compound G.

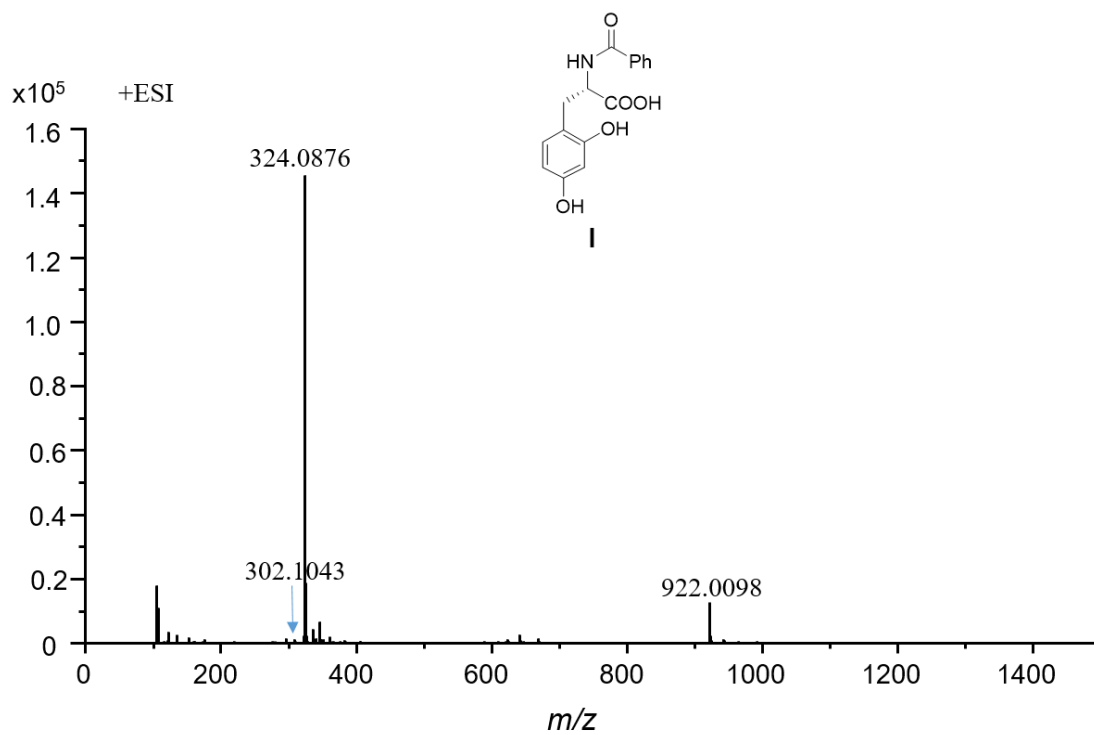


Figure S116. HRESIMS spectrum of I.

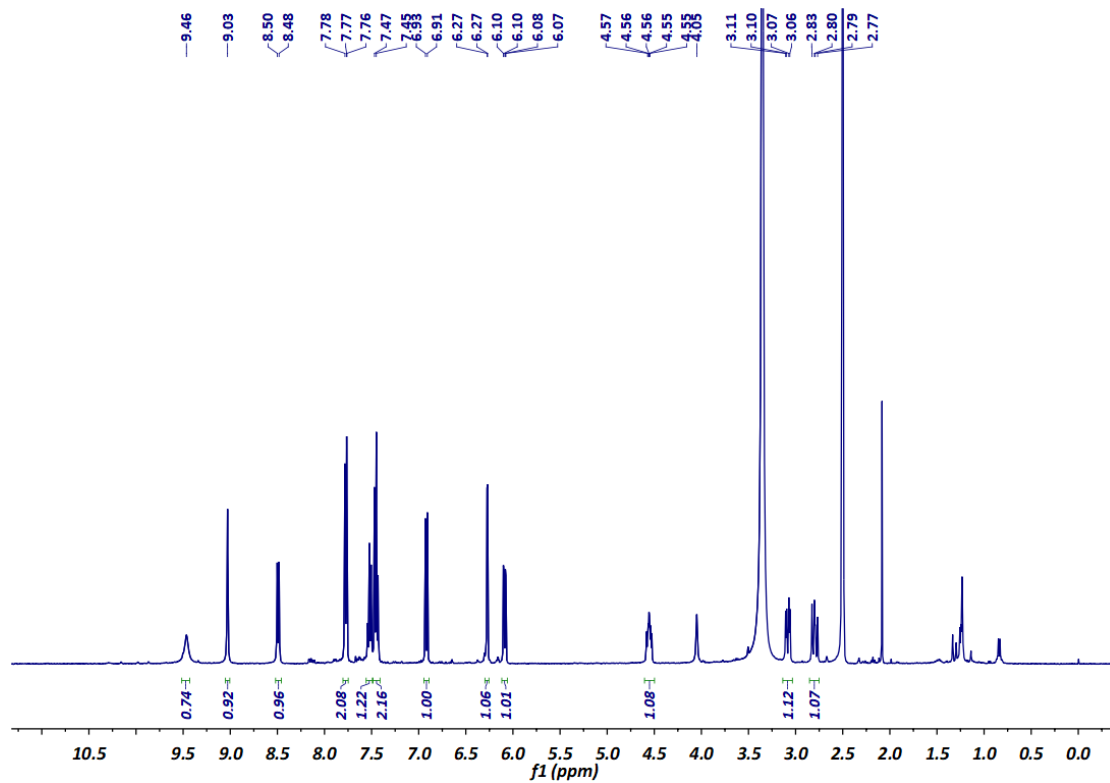


Figure S117.  $^1\text{H}$  NMR (400 MHz,  $\text{DMSO}-d_6$ ) spectrum of compound I.

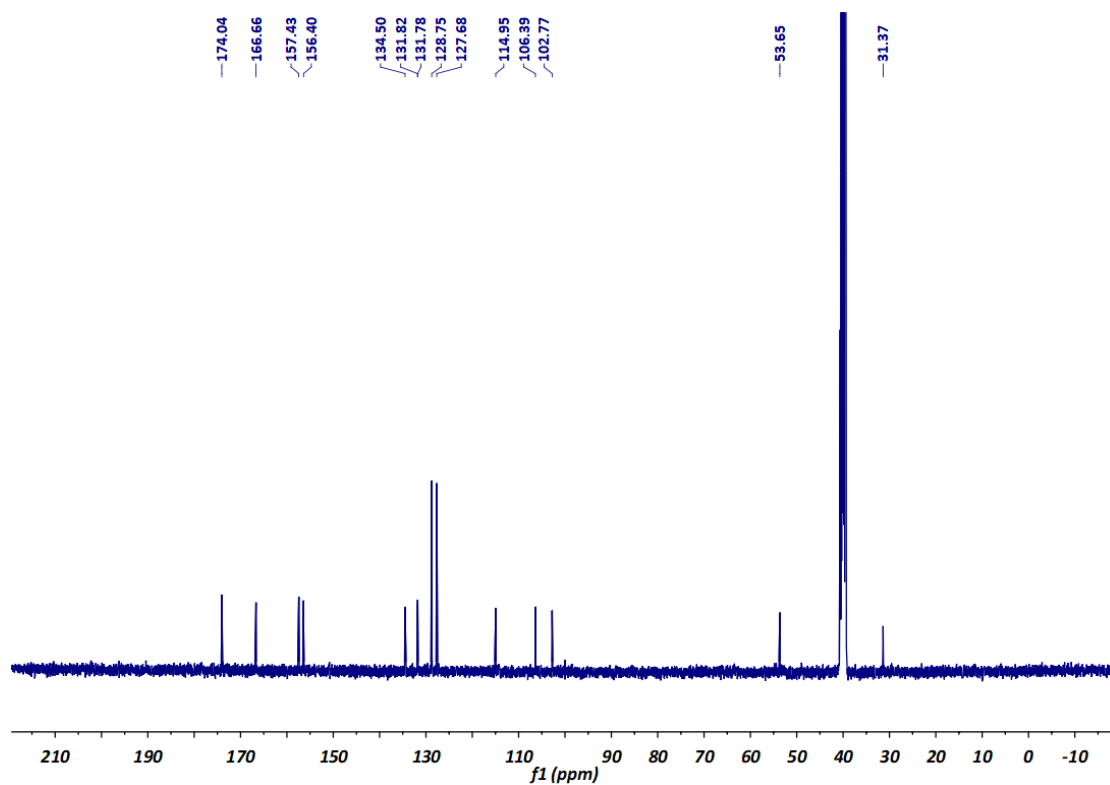


Figure S118.  $^{13}\text{C}$  NMR (100 MHz,  $\text{DMSO}-d_6$ ) spectrum of compound I.

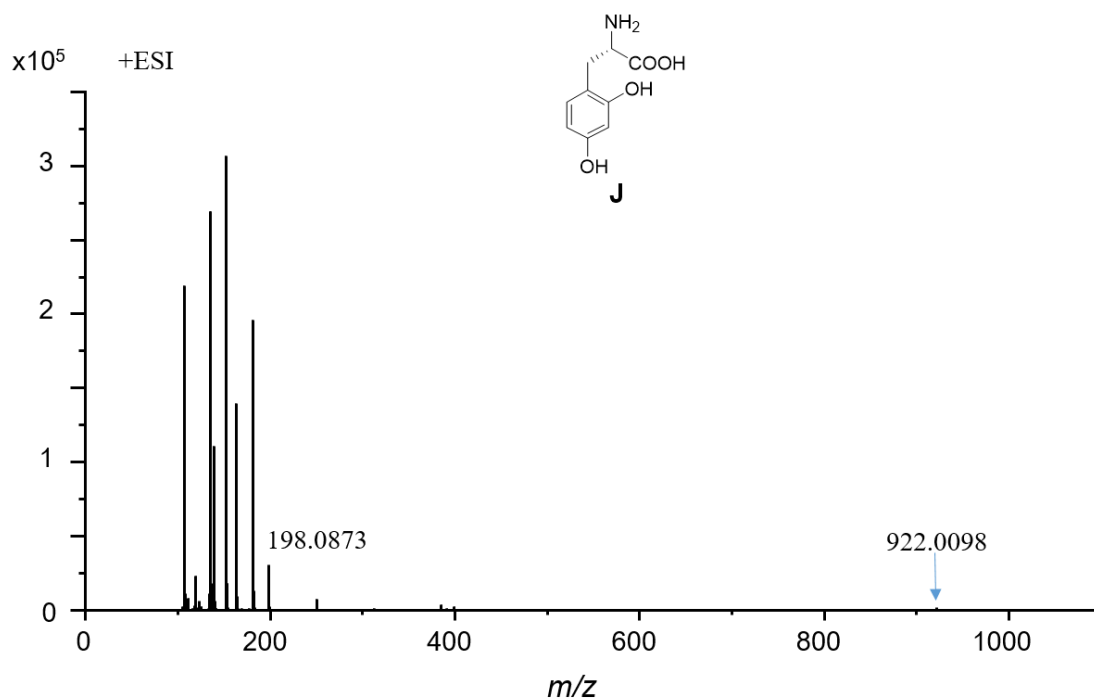


Figure S119. HRESIMS spectrum of J.

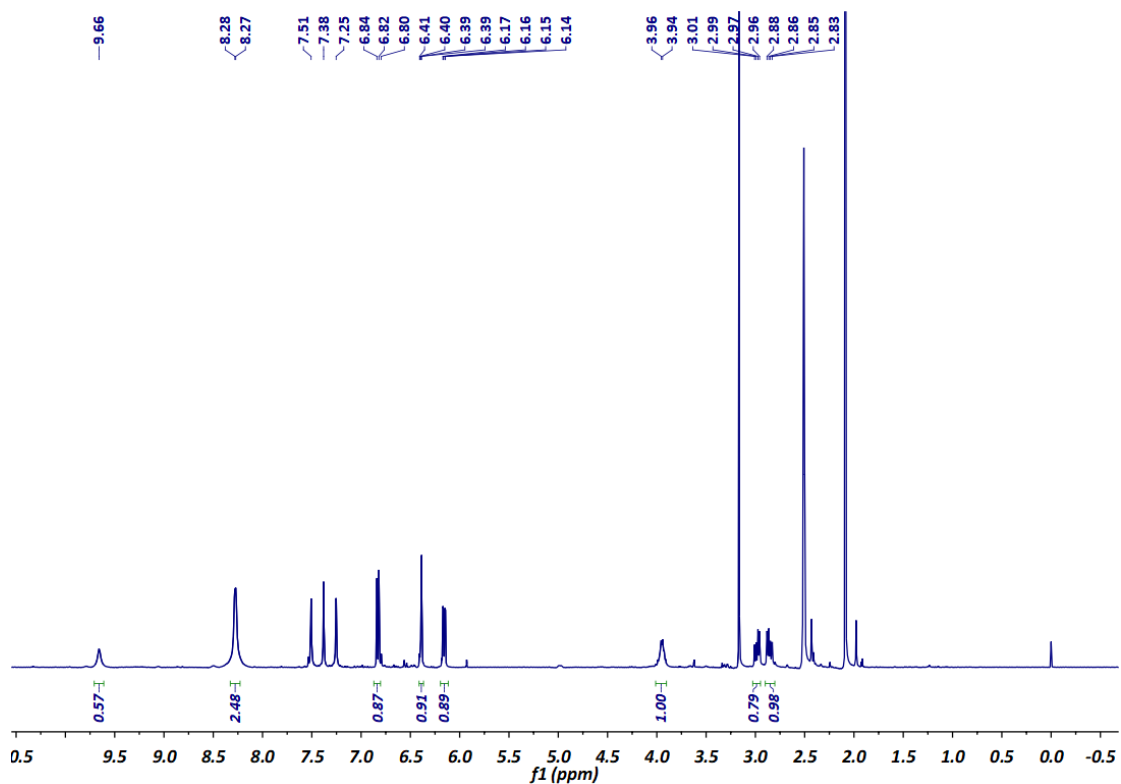


Figure S120. <sup>1</sup>H NMR (400 MHz, DMSO-*d*<sub>6</sub>) spectrum of compound J.

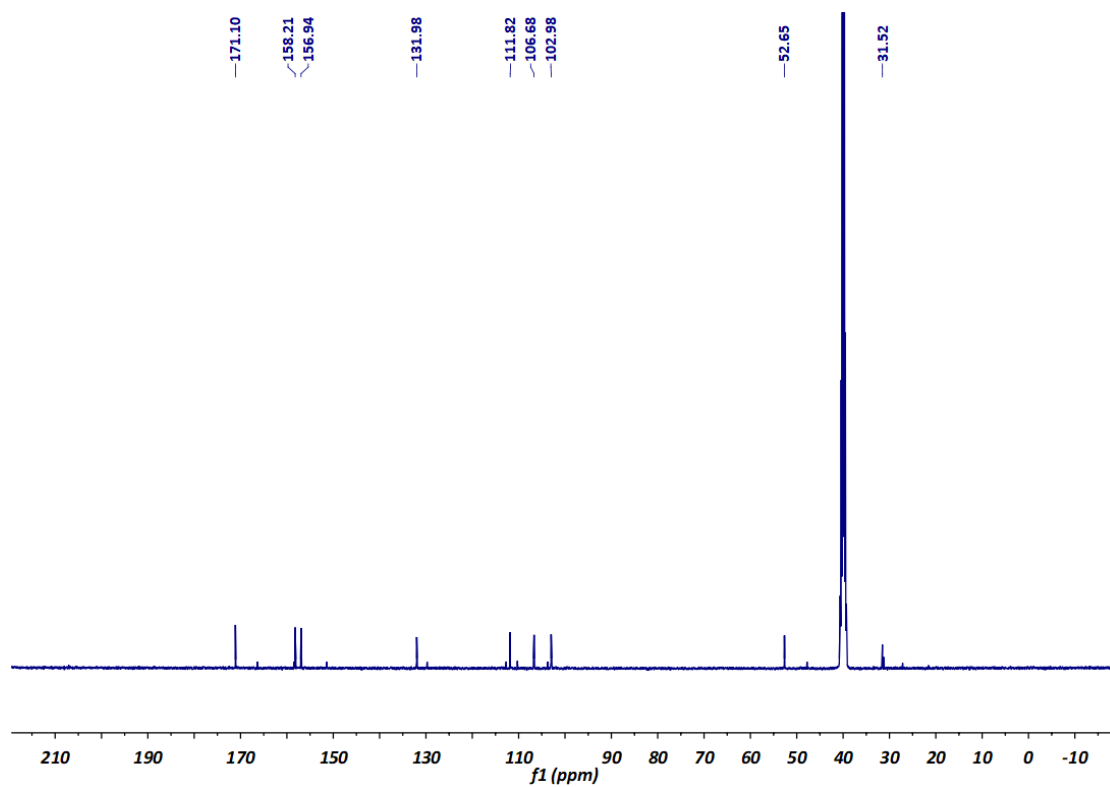


Figure S121.  $^{13}\text{C}$  NMR (100 MHz,  $\text{DMSO-}d_6$ ) spectrum of compound J.

## References

1. Y. N. Song, R. H. Jiao, W. J. Zhang, G. Y. Zhao, H. Dou, R. Jiang, A. H. Zhang, Y. Y. Hou, S. F. Bi, H. M. Ge and R. X. Tan, *Org. Lett.*, 2015, **17**, 556-559.
2. J. Sambrook, E. F. Fritsch and T. Maniatis, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.
3. R. Bhushan and H. Bruckner, *Amino Acids* 2004, **27**, 231-247.
4. K. Kim, K. Parang, O. D. Lau and P. A. Cole, *Bioorg. Med. Chem.*, 2000, **8**, 1263-1268.
5. F. M. Bautista, J. M. Campelo, A. Garcia, D. Luna, J. M. Marinas and A. A. Romero, *J. Chem. Soc. Perkin Trans.*, 2002, **2**, 227-234.
6. U. Nagel, *Angew. Chem., Int. Ed.*, 1984, **96**, 425-426.
7. M. Tagaki and K. Yamamoto, *Tetrahedron* 1991, **47**, 8869-8882.
8. Y. N. Chen, M. J. LaMarche, H. M. Chan, P. Fekkes, J. Garcia-Fortanet, M. G. Acker, B. Antonakos, C. H. Chen, Z. Chen, V. G. Cooke, J. R. Dobson, Z. Deng, F. Fei, B. Firestone, M. Fodor, C. Fridrich, H. Gao, D. Grunenfelder, H. X. Hao, J. Jacob, S. Ho, K. Hsiao, Z. B. Kang, R. Karki, M. Kato, J. Larrow, L. R. La Bonte, F. Lenoir, G. Liu, S. Liu, D. Majumdar, M. J. Meyer, M. Palermo, L. Perez, M. Pu, E. Price, C. Quinn, S. Shakya, M. D. Shultz, J. Slisz, K. Venkatesan, P. Wang, M. Warmuth, S. Williams, G. Yang, J. Yuan, J. H. Zhang, P. Zhu, T. Ramsey, N. J. Keen, W. R. Sellers, T. Stams and P. D. Fortin, *Nature* 2016, **535**, 148-152.
9. Q. Liu, J. Qu, M. Zhao, Q. Xu and Y. Sun, *Pharmacol. Res.*, 2020, **152**, 104595-104609.
10. H. Ran, R. Tsutsumi, T. Araki and B. G. Neel, *Cancer Cell* 2016, **30**, 194-196.
11. A. Flemming, *Nat. Rev. Drug Discov.*, 2016, **15**, 530-531.
12. M. Bierman, R. Logan, O'Brien, K. Seno, E. T. Rao and R. N. Schoner, B. E. *Gene* 1992, **116**, 43-49.
13. L. Li, Z. Xu, J. Wu, Y. Zhang, X. He, T. M. Zabriskie and Z. Deng, *Chembiochem.*, 2008, **9**, 1286-1294.
14. L. Li, G. Pan, X. Zhu, K. Fan, W. Gao, G. Ai, J. Ren, M. Shi, C. Olano, J. A. Salas and K. Yang, *Appl. Microbiol. Biot.*, 2017, **101**, 5291-5300.
15. J. D. Rudolf, L. B. Dong, T. Huang and B. Shen, *Mol. BioSyst.*, 2015, **11**, 2717-2726.
16. M. R. Green, J. Sambrook, *Molecular Cloning: A Laboratory Manual* 4th edn (Cold Spring Harbor Laboratory Press, 2012).
17. F. Xu, Nazari, B. Moon, K. L. B. Bushin and M. R. Seyedsayamdost, *J. Am. Chem. Soc.*, 2017, **139**, 9203-9212.
18. M. Röttig, M. H. Medema, K. Blin, T. Weber, C. Rausch and O. Kohlbacher, *Nucleic Acids Res.*, 2011, **39**, 362-367.
19. M. J. Cryle, A. Meinhart and I. Schlichting, *J. Biol. Chem.*, 2010, **285**, 24562-24574.
20. K. Haslinger, C. Brieke, n S. Uhlman, L. Sieverling, R. D. Süßmuth, M. J. Cryle. *Angew. Chem., Int. Ed.*, 2014, **53**, 8518-8522.
21. H. Yeom, S. G. Sligar, H. Li, T. L. Poulos and A. J. Fulco, *Biochemistry* 1995, **34**, 14733-14740.
22. T. W. Ost, C. S. Miles, A. W. Munro, J. Murdoch, G. A. Reid and S. K. Chapman, *Biochemistry* 2001, **40**, 13421-13429.
23. X. Robert and P. Gouet, *Nucleic Acids Res.*, 2014, **42**, W320-W324.
24. Z. L. Reitz, C. D. Hardy, J. Suk, J. Bouvet and A. Butler, *Proc. Natl. Acad. Sci. U. S. A.*, 2019, **116**, 19805-19814.
25. R. D. Süßmuth and A. Mainz, *Angew. Chem., Int. Ed.*, 2017, **56**, 3770-3821.