## Supporting Information

# Proteomics Guided Discovery of Flavopeptins: Anti-Proliferative Aldehydes Synthesized by a Reductase Domain-Containing Nonribosomal Peptide Synthetase

Yunqiu Chen‡, Ryan A. McClure‡, Yupeng Zheng, Regan J. Thomson*, Neil L. Kelleher*

Department of Chemistry, Northwestern University, 2145 Sheridan Road, Evanston, IL, 60208, USA

**Corresponding Author**

* n-kelleher@northwestern.edu

* r-thomson@northwestern.edu

‡These authors contributed equally.

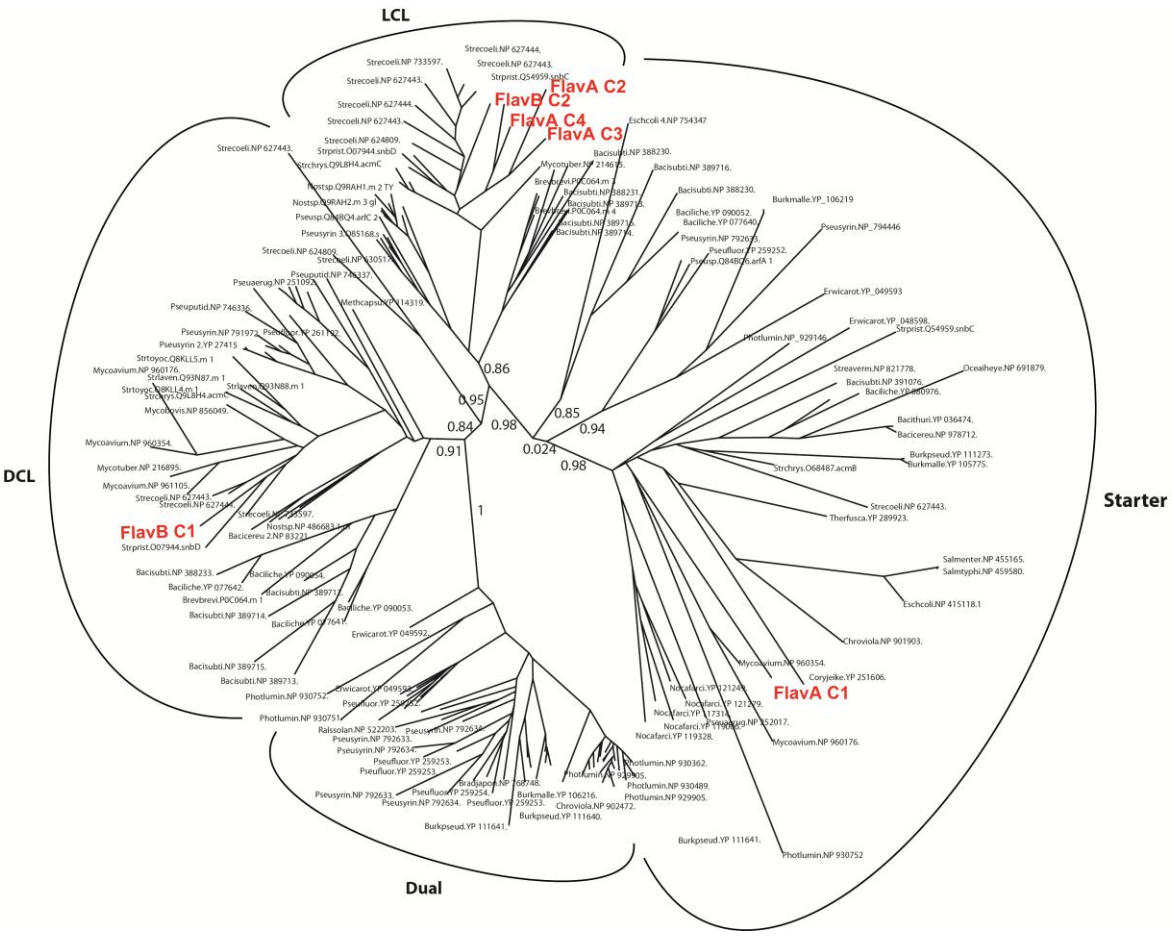



**Figure S1.** Phylogenetic tree built using 159 NRPS condensation domains and the six condensation domains from the flavopeptin biosynthetic gene cluster (shown in red). The 159 condensation domains with known subtypes (LCL, DCL, dual and starter) were extracted from reference 20 (Rausch, C.; Hoof, I.; Weber, T.; Wohlleben, W.; Huson, D. H. *BMC Evol. Biol.* **2007**, *7*, 78.). Sequence alignment was done by MUSCLE and the phylogeny was reconstructed using phyml, employing the JTT model of amino acid substitution. FlavA C1 falls into the category of starter C domain, FlavB C1 belongs to DCL subtype, and other C domains are all LCL types.

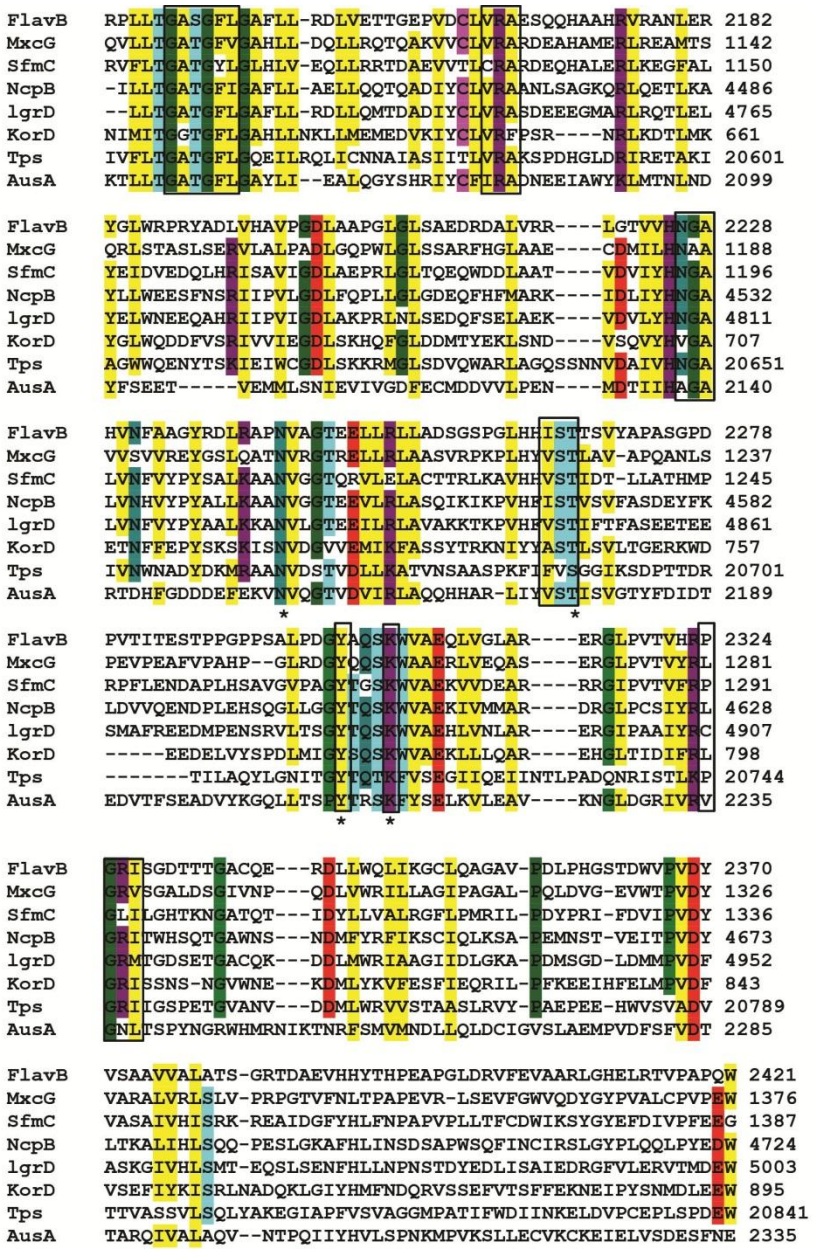



**Figure S2.** Sequence alignment of the reductase domain from FlavB with other NRPS terminal reductases: mxcG (*Stigmatella aurantiaca* Sg a15), sfmC (*Streptomyces Lavendulae*), ncpB (*Nostoc sp.* ATCC 53789), lgrD (*Brevibacillus brevis NBRC 100599*), KorD (*Bacillus sp.* NK2003), TPS (*Hypocrea virens*) and AusA (*Staphylococcus aureus* subsp. aureus str. JKD6008). Conserved residues are colored as follows: hydrophobic in yellow, negatively charged in red, positively charged in violet, serine and threonine in

turquoise, proline and glycine in green, cysteine in pink and asparagine/glutamine in teal. Active site residues are indicated by a star '*'. Residues that interact with NAD(P)H are enclosed in black boxes. The accession numbers for the synthetases indicated are: myxochelin (AAG31130), saframycin (ABI22133), nostocyclopeptide (AAO23334), gramicidin (226312537), koranimine (AEC14349), peptaibol (AAM78457) and aureusimine (384860829).

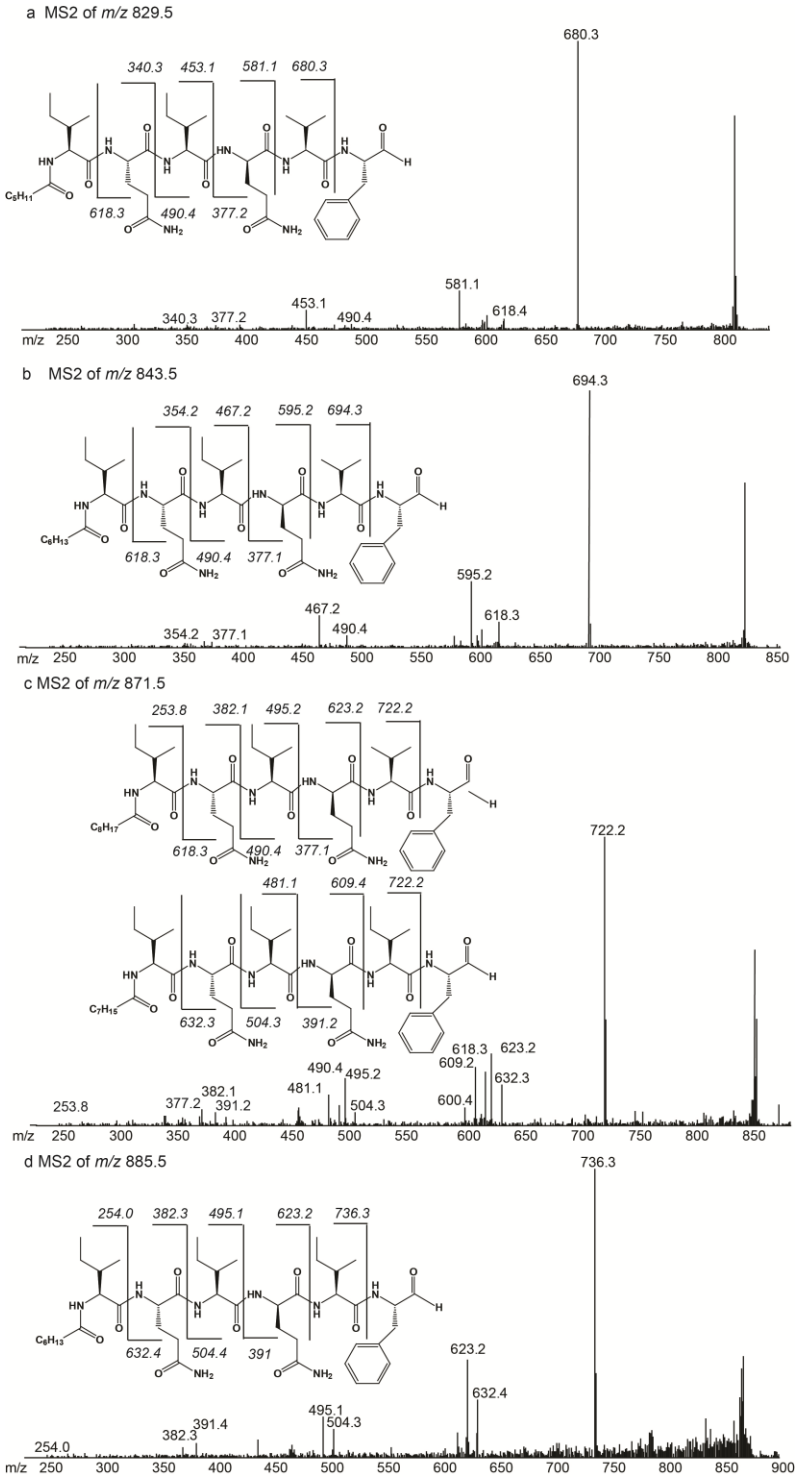



**Figure S3.** $MS^2$ fragmentation spectra for different flavopeptin species with *m/z* 829.5, 843.5, 871.5 and 885.5, respectively.

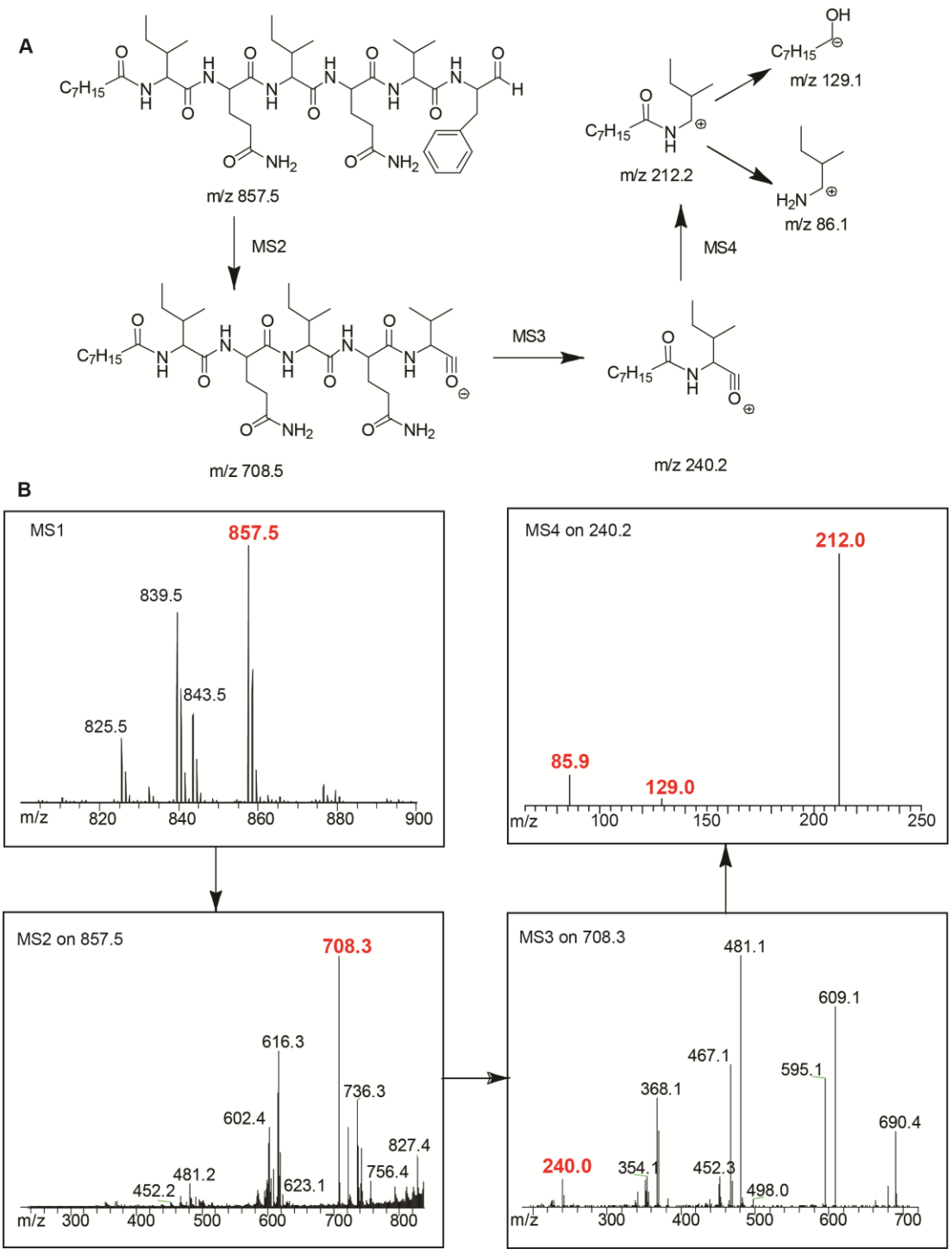
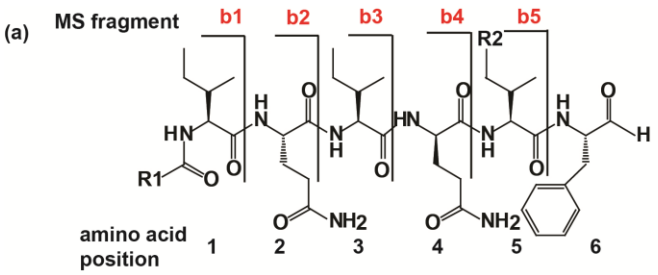



**Figure S4.** Tandem MS for the localization of the first isoleucine residue. (a) Tandem MS reaction mechanism for flavopeptin. (b) The actual spectrum at each stage of MS event. $MS^4$ spectrum on *m/z* 240.2 agrees with the predicted fragmentation pattern for the fatty acyl-isoleucine substructure.




| 13C6 Ile | $R_1=C_7H_{15}$, $R_2=H$ | | $R_1=C_6H_{13}$, $R_2=CH_3$ | | |
|---|---|---|---|---|---|
| | 1 | 3 | 1 | 3 | 5 |
| **b1** | 246.2 | 240.2 | 232.2 | 226.2 | 226.2 |
| **b2** | 374.3 | 368.3 | 360.2 | 354.2 | 354.2 |
| **b3** | 487.3 | 487.3 | 473.3 | 473.3 | 467.3 |
| **b4** | 615.4 | 615.4 | 595.4 | 595.4 | 601.4 |
| **b5** | 714.5 | 714.5 | 714.5 | 714.5 | 714.5 |

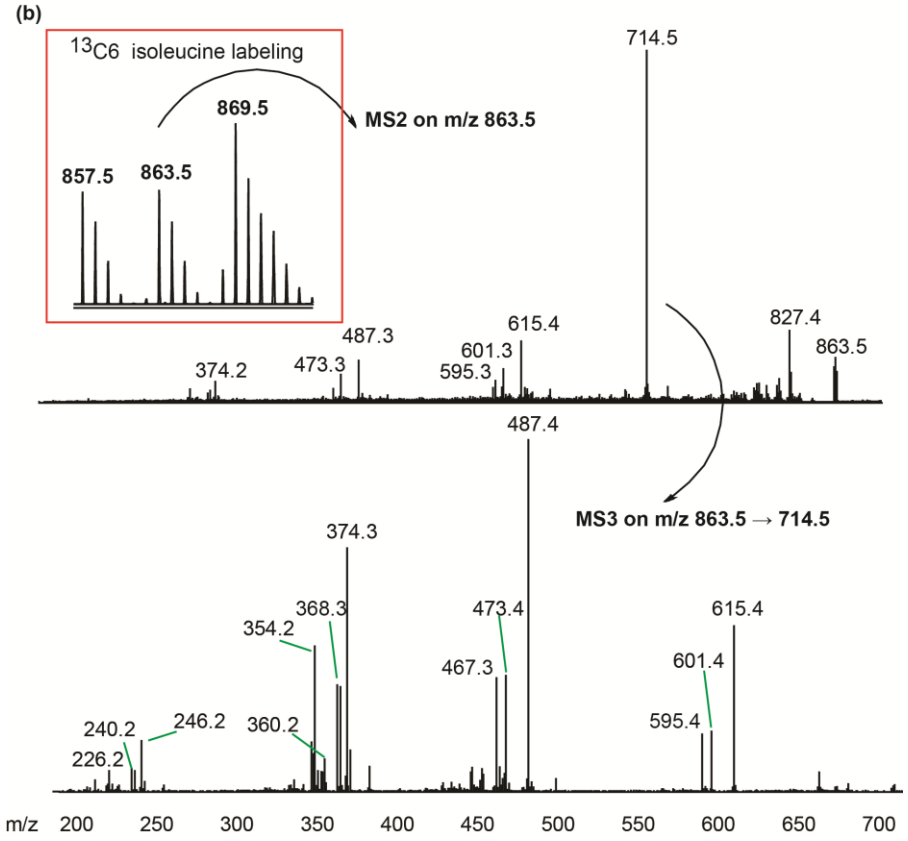



**Figure S5.** Tandem MS showing isoleucine can be incorporated into amino acid positions 1, 3 and 5. (a) Chemical structure and the fragment ion masses for $^{13}C_6$-isoleucine labeled flavopeptin (with *m/z* 863.5) at all possible sites. (b) $MS^2$ and $MS^3$ fragmentation of *m/z* 863.5 confirmed that isoleucine can be incorporated into amino acid position 1, 3 and 5.

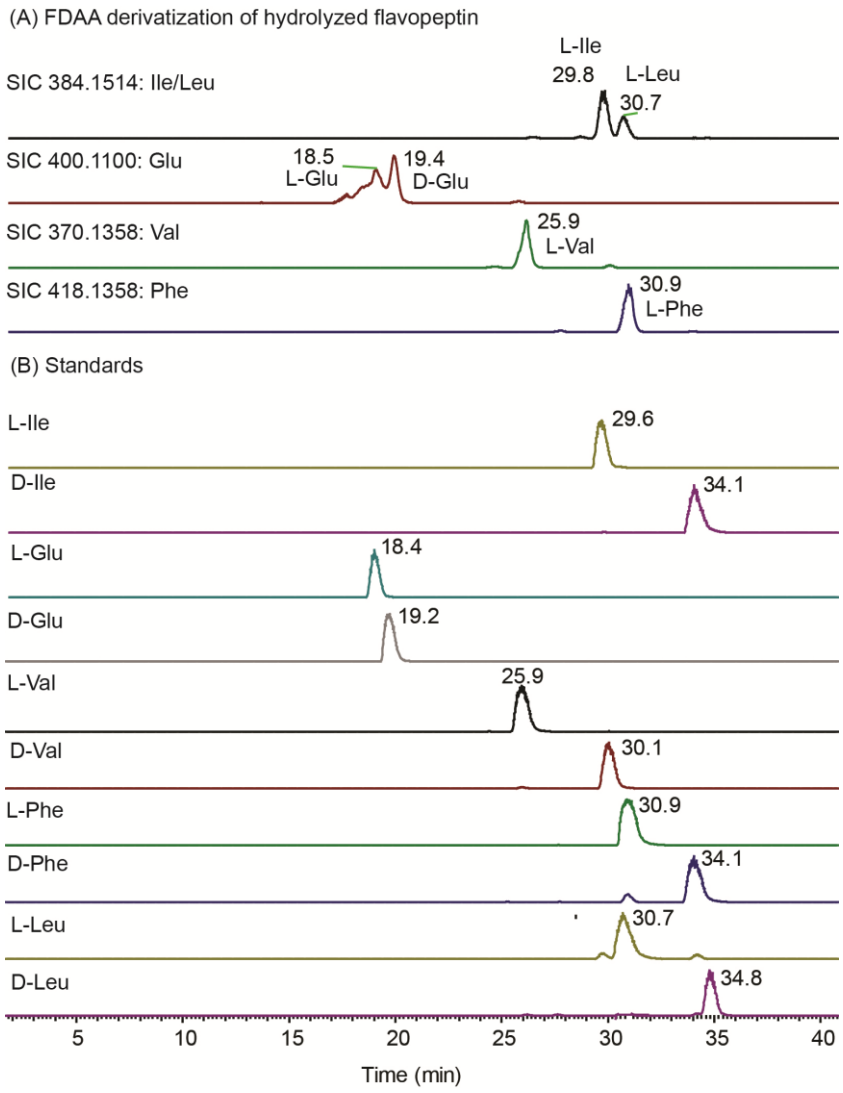
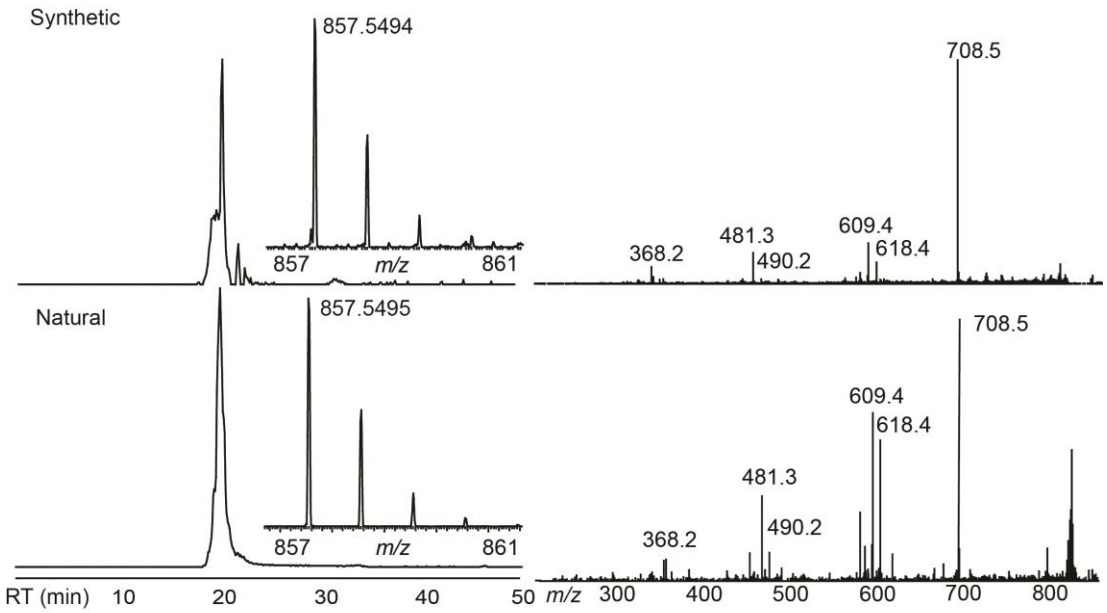
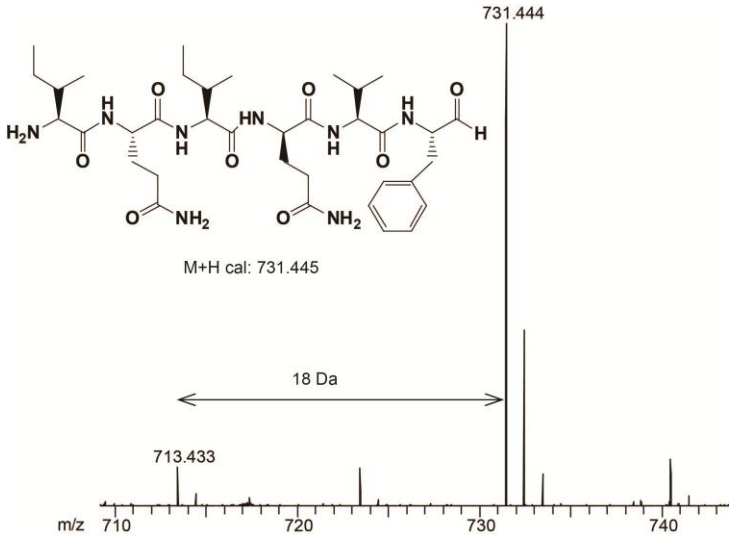



**Figure S6** Determination of the stereochemistry of amino acid building blocks of flavopeptins by Marfey's method. Flavopeptins were hydrolyzed in 6 N HCl at 110°C for 24 h before derivatization by the Marfey's reagent fluorodinitrophenyl-5-L-alanine amide (FDAA) followed by reverse-phase LC-MS analysis. (A) Selected ion chromatograms (SICs) for the FDAA derivatives of amino acids from hydrolyzed flavopeptins. Note that glutamine was converted to glutamic acid during strong acid hydrolysis. The stereochemistry of each amino acid was determined by comparing the retention time with the corresponding standard amino acid derivatives shown in panel (B). The mass tolerance for all SICs was set to be 10 ppm.




(a)




(b)

**Figure S7.** Synthetic flavopeptins. (a) LC–MS comparison between the synthetic *N*-octanoyl flavopeptin (**1**, top) and natural flavopeptins isolated from F-6652 (bottom). Shown are the selected ion chromatograms of *m/z* 857.5, with the mass spectra shown as an insert and the $MS^2$ fragmentation on right. (c) Mass spectrum for the *N*-free amine flavopeptin analog (**3**) showing a dominant peak at 731.4, with very minor (< 10%) dehydrated form.

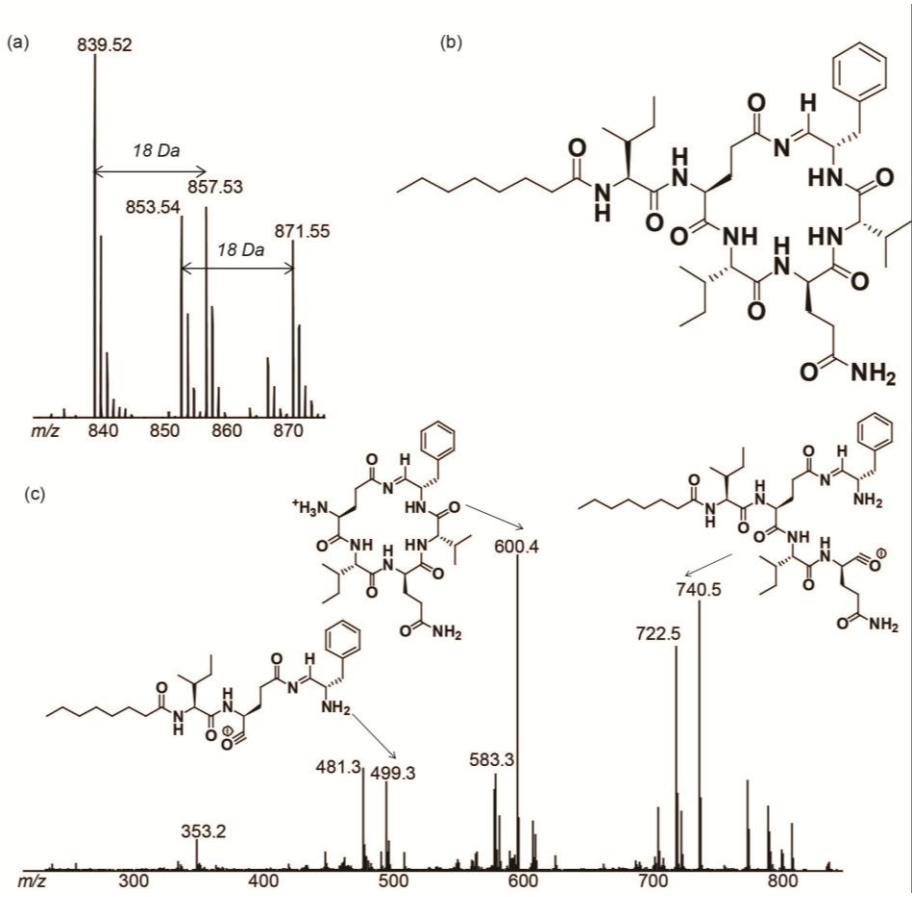



**Figure S8.** Dehydrated form of flavopeptins. (a) Mass spectrum of flavopeptins after organic solvent extraction and drying, showing the dehydrated products at *m/z* 839.5 and 853.5. (b) Proposed structure for the dehydrated form at *m/z* 839.5. (c) $MS^2$ spectrum of *m/z* 839.5, with the structures of the major fragment masses shown.

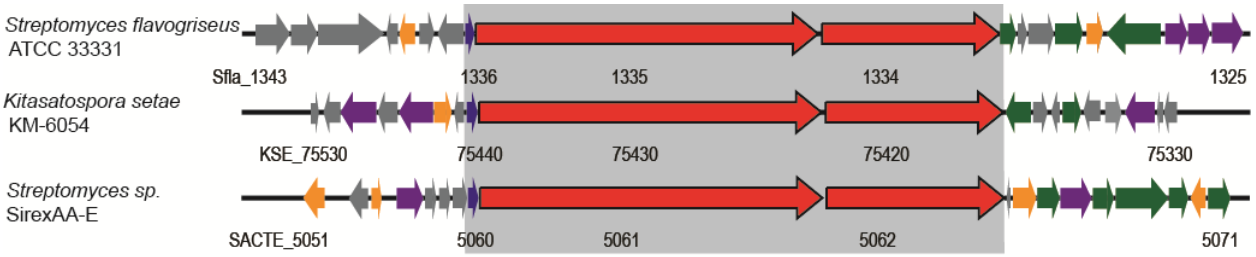
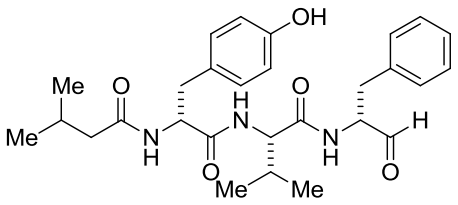


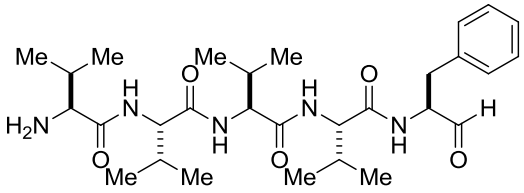

**Figure S9** Compare the flavopeptin biosynthetic gene cluster among *Streptomyces flavogriseus* ATCC 33331 (top), *Kitasatospora setae* KM-6054 (middle) and *Streptomyces* sp. *Sirex*AA-E (bottom). The three genes (*mbtH, flavA and flavB*) in shadow show high sequence similarity among the three strains while other neighboring genes are not highly conserved. Color coding for the genes: red for NRPS, dark blue for MbtH, orange for transcriptional regulators, purple for transporters, green for genes with other annotated functions, grey for hypothetical proteins. The descriptions of genes are shown in the table below.

| ***Streptomyces flavogriseus* ATCC 33331** | | ***Kitasatospora setae* KM-6054** | | ***Streptomyces* sp. *Sirex*AA-E** | |
|---|---|---|---|---|---|
| Sfla_1343 | von Willebrand factor A | KSE_75510 | hypothetical protein | SACTE_5051 | LysR family transcriptional regulator |
| Sfla_1342 | ATPase AAA | KSE_75500 | hypothetical protein | SACTE_5053 | NmrA family protein |
| Sfla_1341 | hypothetical protein | KSE_75490 | putative drug resistance protein | SACTE_5054 | HxlR transcriptional regulator |
| Sfla_1340 | hypothetical protein | KSE_75480 | hypothetical protein | SACTE_5056 | extracellular ligand-binding receptor |
| Sfla_1339 | transcriptional regulator | KSE_75470 | putative major facilitator superfamily transporter | SACTE_5057 | hypothetical protein |
| Sfla_1338 | hypothetical protein | KSE_75460 | putative MarR family transcriptional regulator | SACTE_5058 | hypothetical protein |
| Sfla_1337 | hypothetical protein | KSE_75450 | hypothetical protein | SACTE_5059 | hypothetical protein |
| Sfla_1336 | MbtH | KSE_75440 | MbtH (74/81%) | SACTE_5060 | MbtH (79/90%) |
| Sfla_1335 | NRPS | KSE_75430 | NRPS (69/75%) | SACTE_5061 | NRPS (80/85%) |
| Sfla_1334 | NRPS | KSE_75420 | NRPS (69/75%) | SACTE_5062 | NRPS (80/85%) |
| Sfla_1333 | N-acetyltransferase GCN5 | KSE_75410 | putative polyprenyl diphosphate synthase | SACTE_5063 | hypothetical protein |

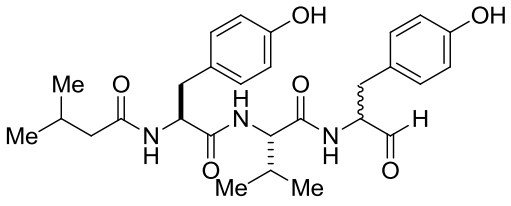
| | | | | | |
|---|---|---|---|---|---|
| Sfla_1332 | hypothetical protein | KSE_75400 | hypothetical protein | SACTE_5064 | AsnC transcriptional regulator |
| Sfla_1331 | hypothetical protein | KSE_75390 | hypothetical protein | SACTE_5065 | polysaccharide deacetylase |
| Sfla_1330 | acetyl xylan esterase | KSE_75380 | methyltransferase | SACTE_5066 | general substrate transporter |
| Sfla_1329 | IclR family transcriptional regulator | KSE_75370 | hypothetical protein | SACTE_5067 | short-chain dehydrogenase |
| Sfla_1328 | fumarate reductase | KSE_75365 | hypothetical protein | SACTE_5068 | 2-dehydropantoate 2-reductase |
| Sfla_1327 | binding-protein-dependent transport system inner membrane protein | KSE_75360 | major facilitator superfamily transporter | SACTE_5069 | short-chain dehydrogenase |
| Sfla_1326 | binding-protein-dependent transport system inner membrane protein | KSE_75350 | hypothetical protein | SACTE_5070 | TetR family transcriptional regulator |
| Sfla_1325 | ABC transporter | KSE_75330 | hypothetical protein | SACTE_5071 | aldo/keto reductase |



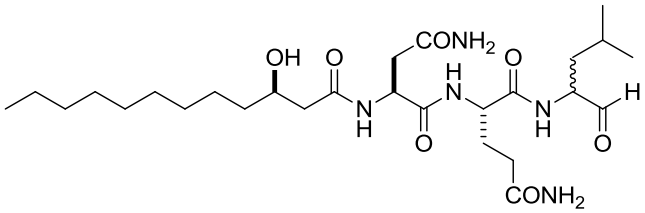
Nerfilin 1
cysteine protease cathepsin and papain inhibitor



Stacopin P1
cysteine protease inhibitor papain and calpain



Tyropeptin
inhibit proteasome, cystein proteases like cathepsin L and calpain



Fellutamide
proteasome and cysteine protease cathepsin B inhibitor

**Figure S10**. Other peptide aldehyde natural products that inhibit cysteine proteases and the proteasome.

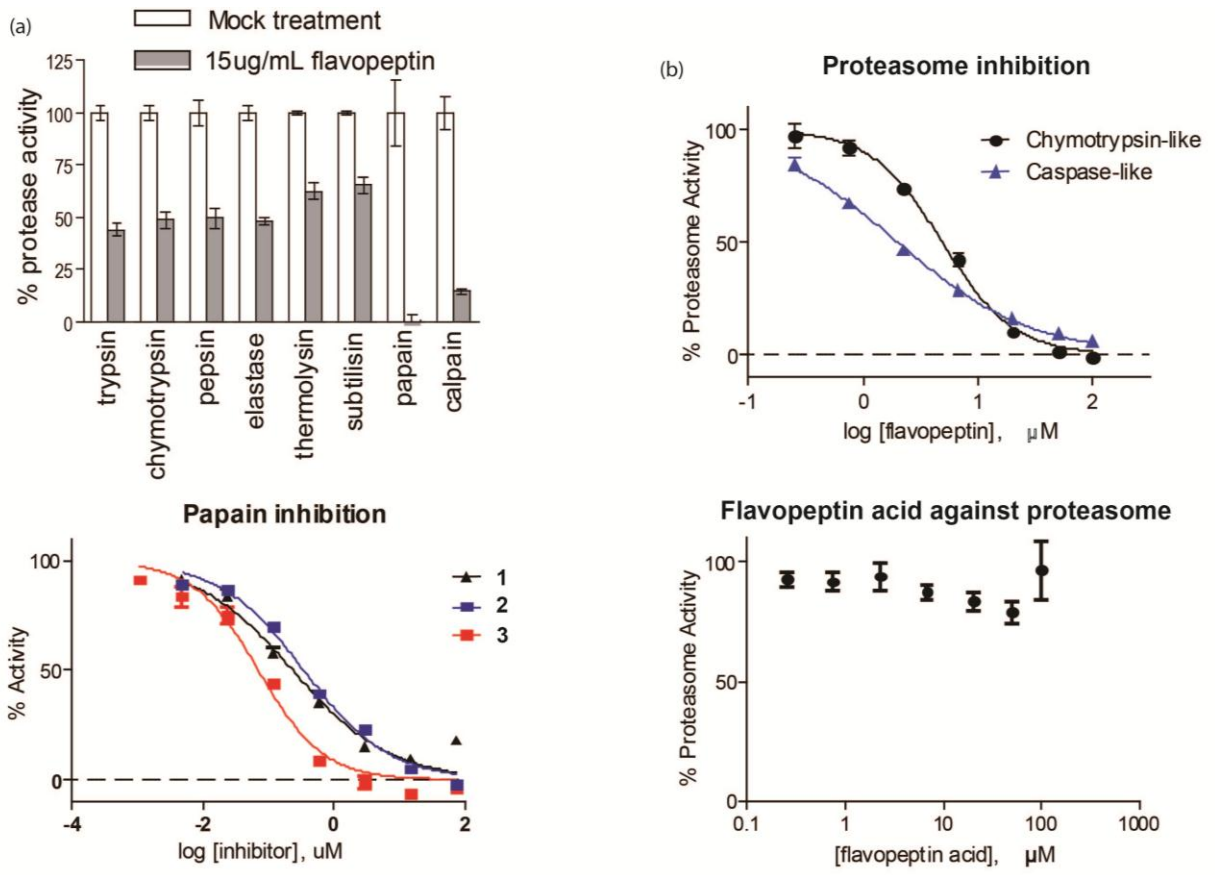



**Figure S11**. Bioactivities of flavopeptins. (a) Top: Flavopeptins specifically inhibited cysteine proteases such as papain and calpain, while proteases of other types, such as serine (trypsin, chymotrypsin, elastase, subtilisin), aspartic (pepsin), and metallo-protease (thermolysin) were less inhibited. Bottom: The $IC_{50}$ plots against papain using the three synthetic flavopeptins (**1**, octanoyl-I-Q-I-Q-V-F-CHO, **2**, heptanoyl-I-Q-I-Q-I-F-CHO and **3**, $NH_2$-I-Q-I-Q-V-F-CHO). (b) Flavopeptins showed micromolar range inhibition against the chymotrypsin and caspase-like activities of the human 20S proteasome. Top, the $IC_{50}$ plot for Compound **1**. Bottom, the flavopeptin acid derivative (**4**) did not inhibit the chymotrypsin-like activity of proteasome up to 100 µM. All data points were in triplicates.

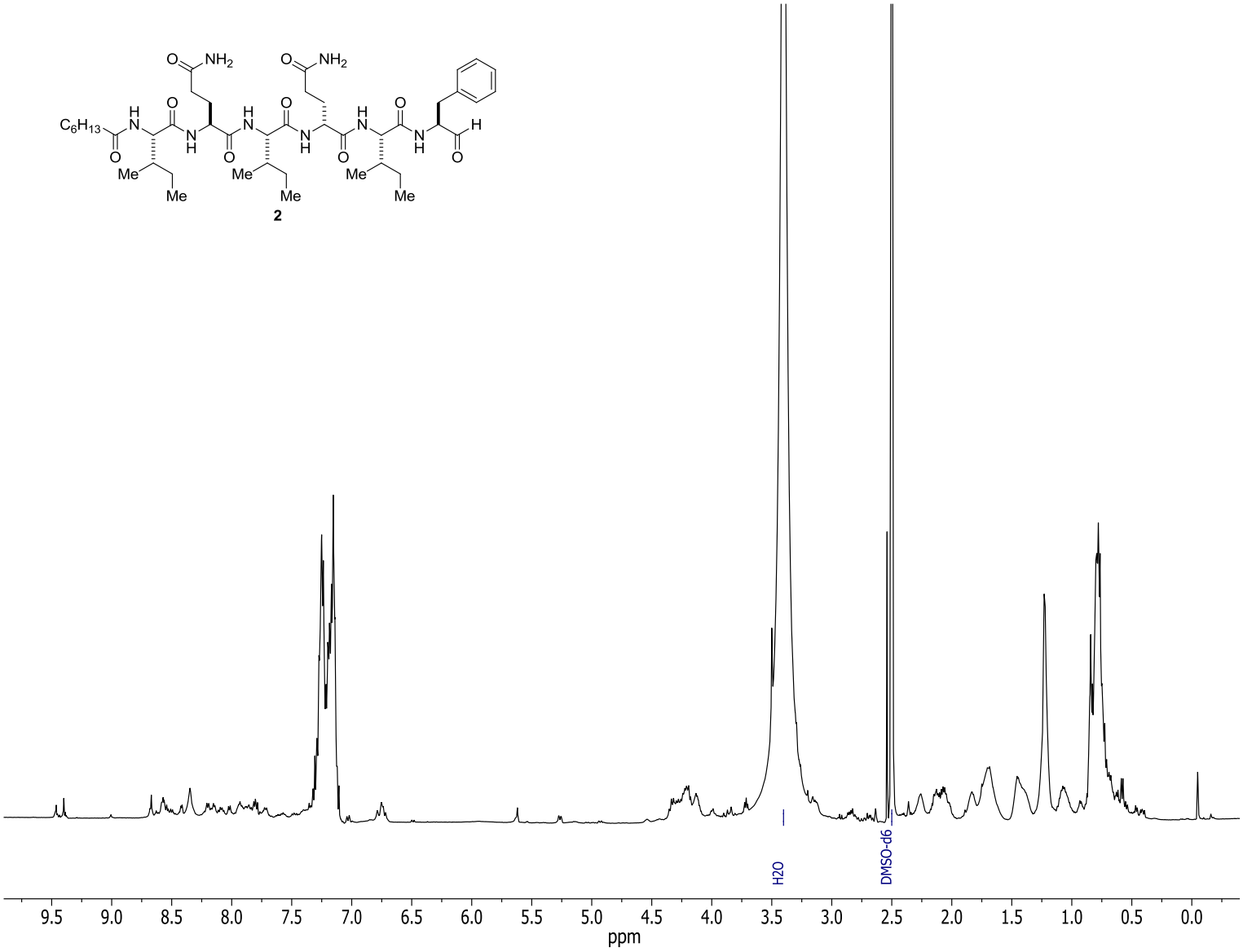



**Figure S12**. $^{1}$H NMR spectrum for **2** in $d_6$-DMSO.

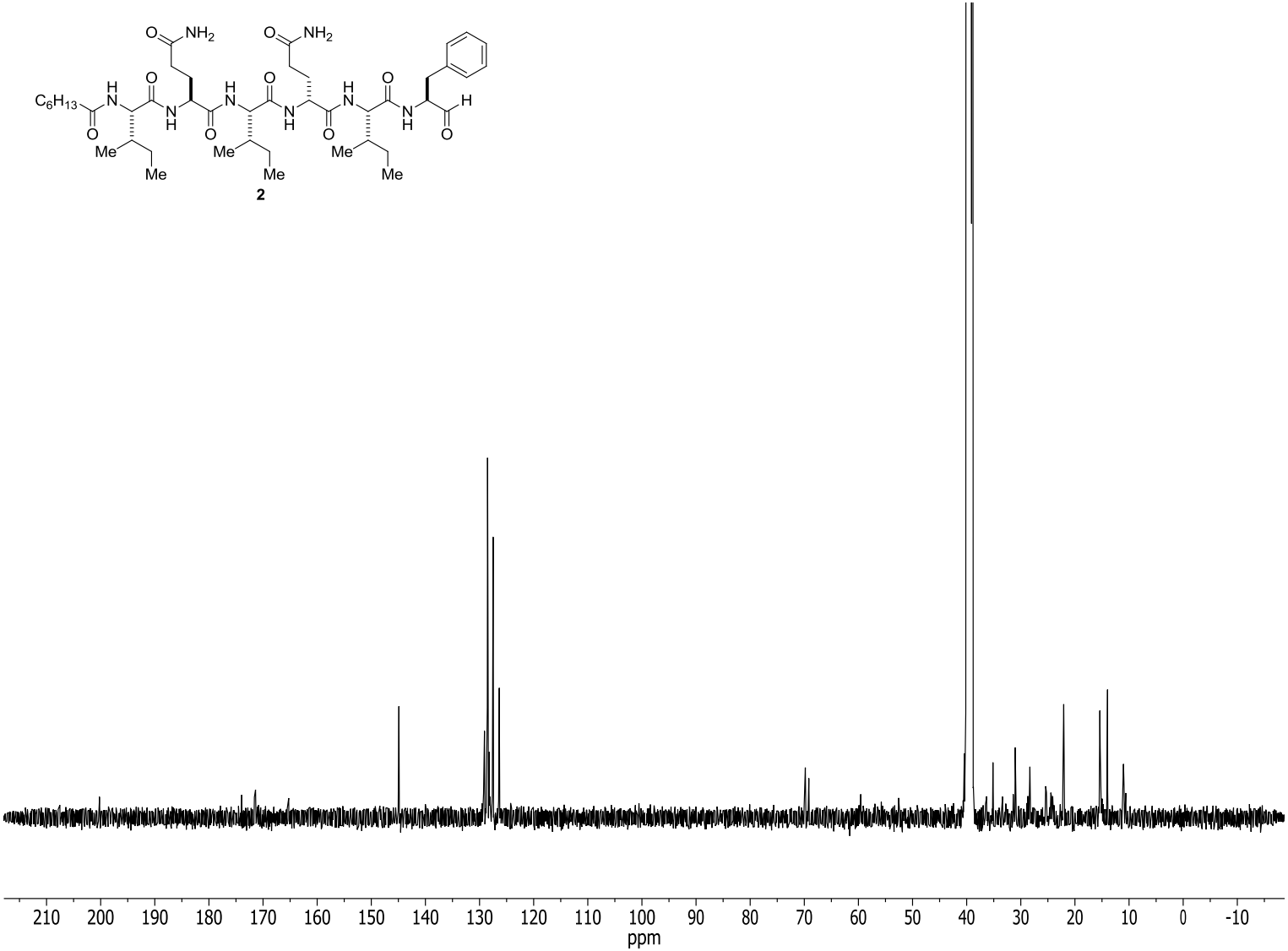



**Figure S13**. $^{13}$C NMR spectrum for **2** in $d_6$-DMSO.

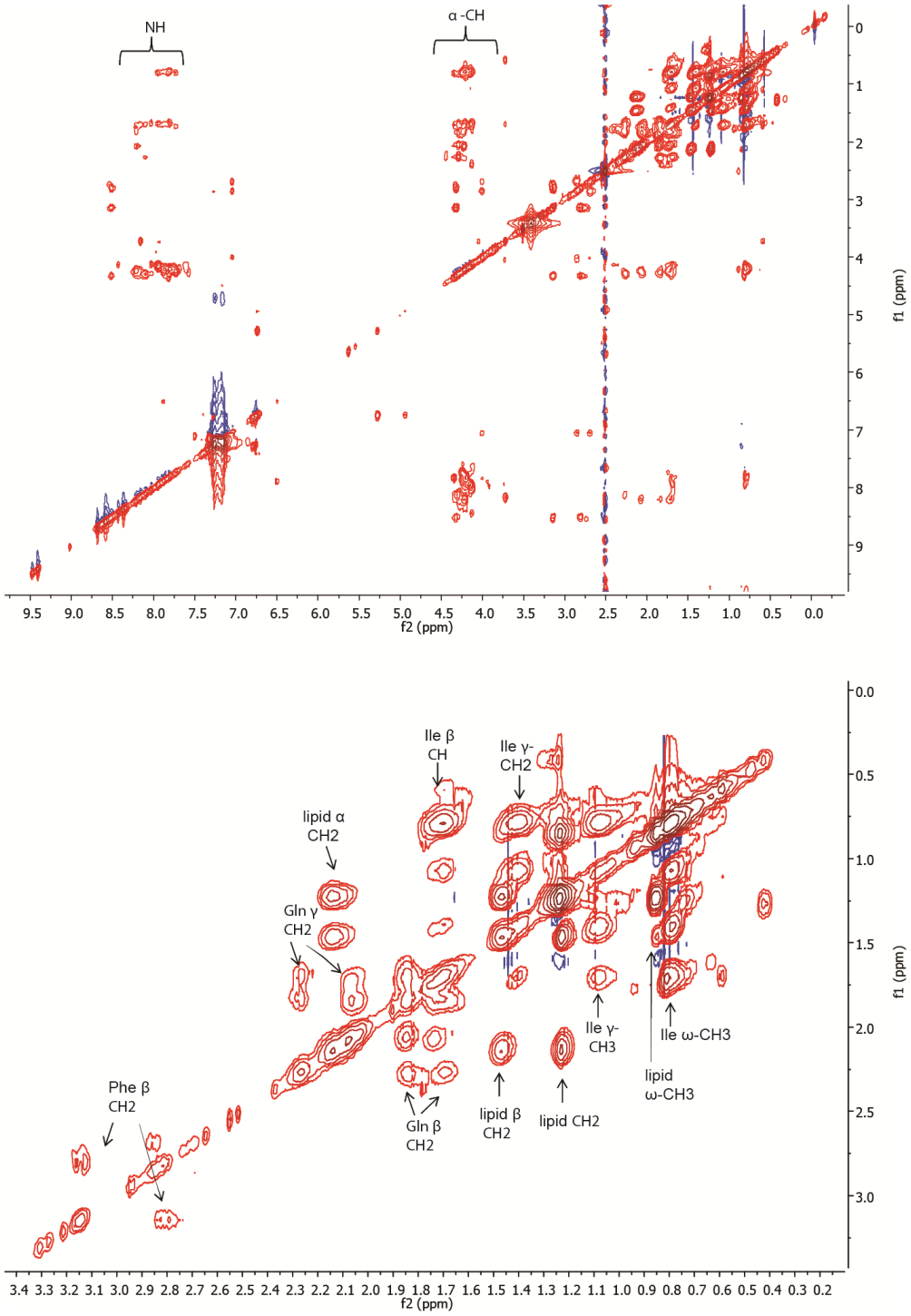



**Figure S14**. Two-dimensional homonuclear ($^1$H) TOCSY spectrum for **2** in $d_6$-DMSO. Top: full TOCSY spectrum with amide and α proton regions highlighted. Bottom: side chain signals and associated spin systems are highlighted.

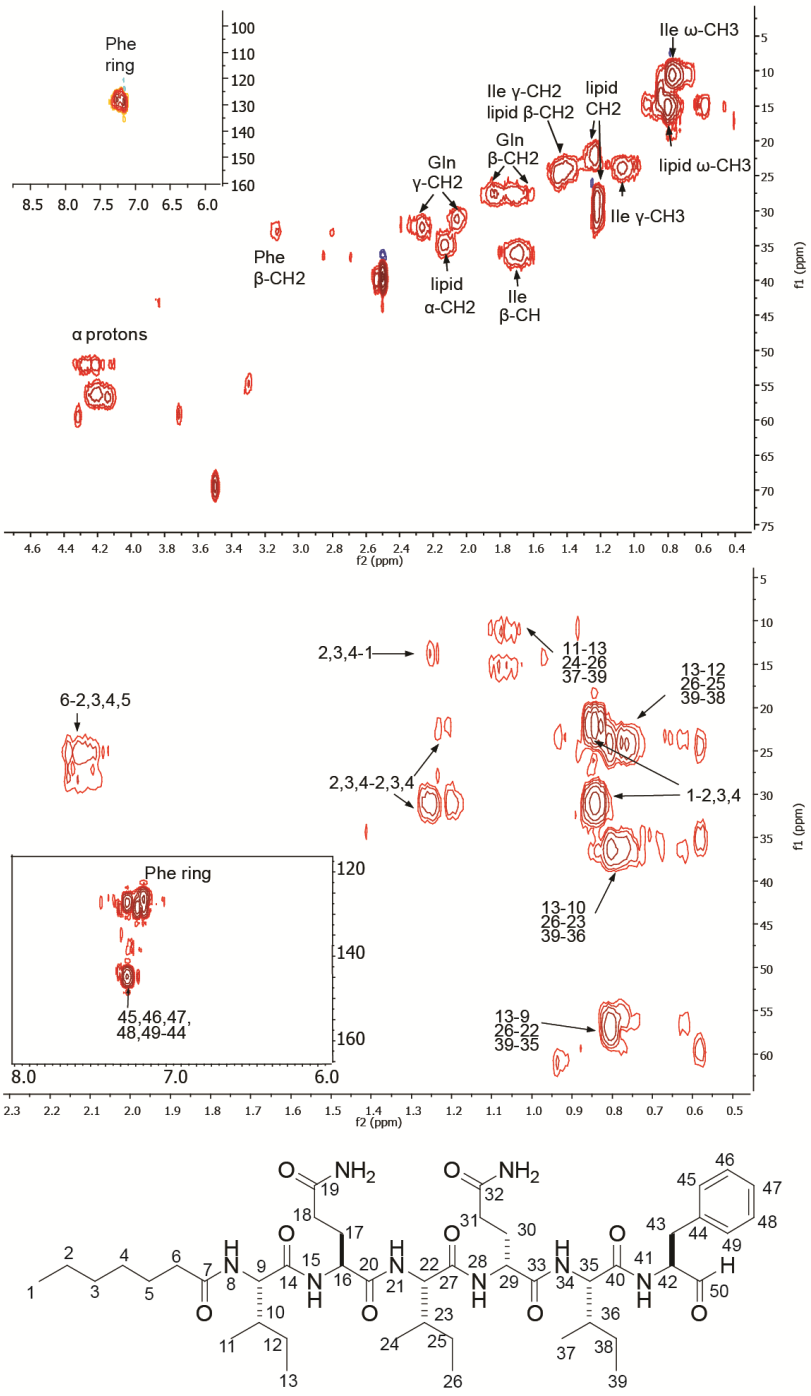



**Figure S15**. Heteronuclear ($^{1}$H-$^{13}$C) NMR spectra for **2** in $d_6$-DMSO with assignments. Top: HSQC spectrum. Bottom: HMBC spectrum.

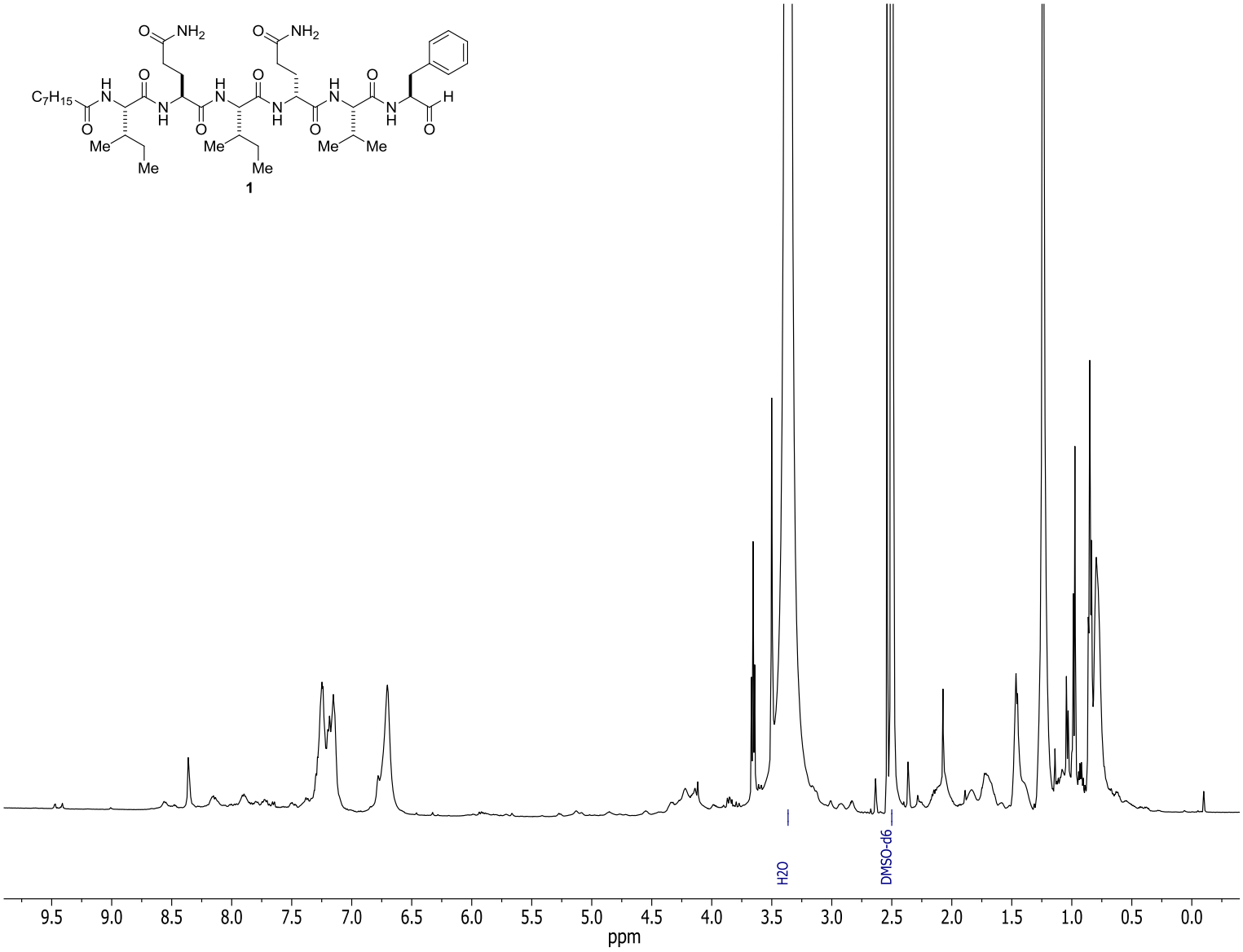



**Figure S16**. $^{1}$H NMR spectrum for **1** in $d_6$-DMSO.

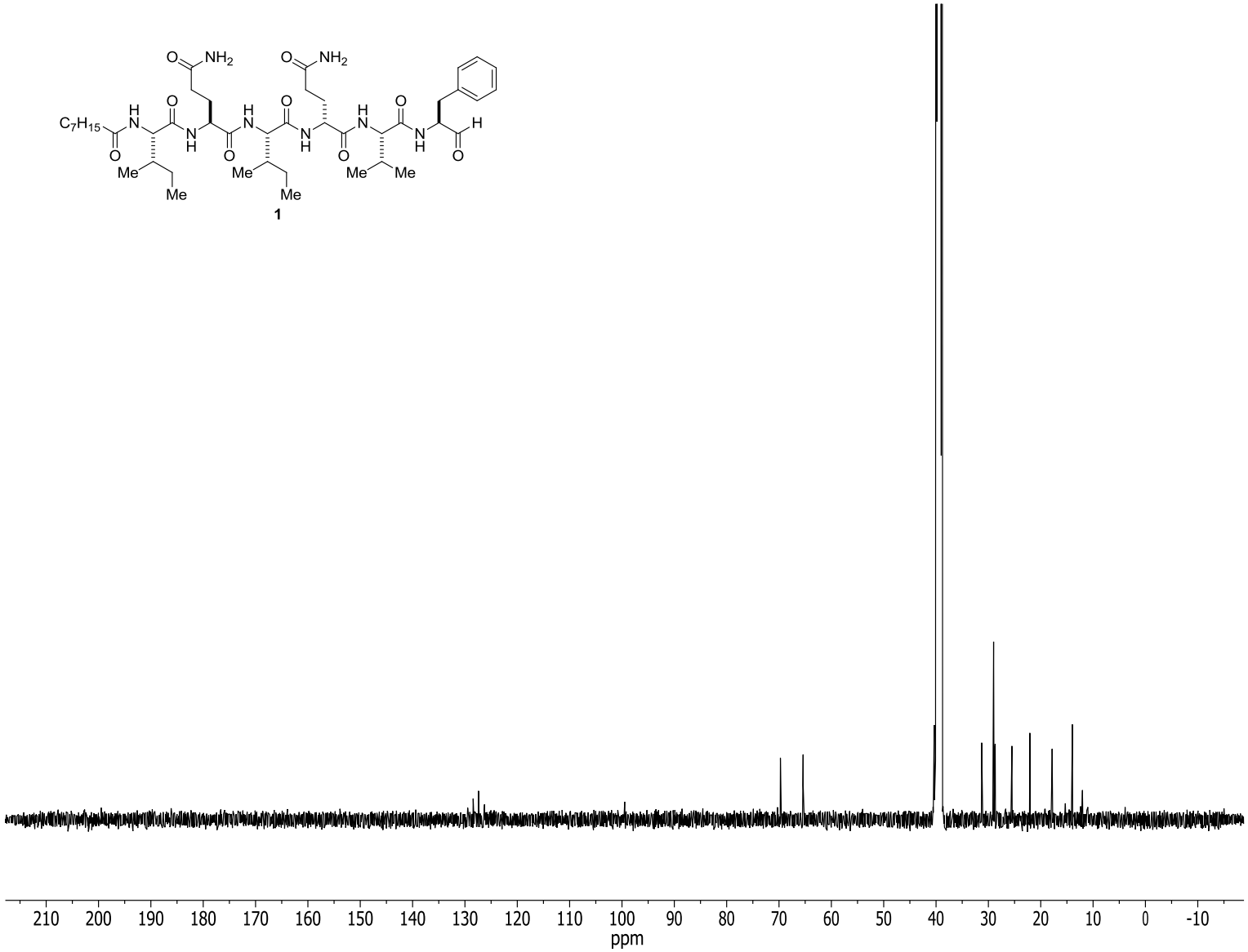



**Figure S17**. $^{13}$C NMR spectrum for **1** in $d_6$-DMSO.

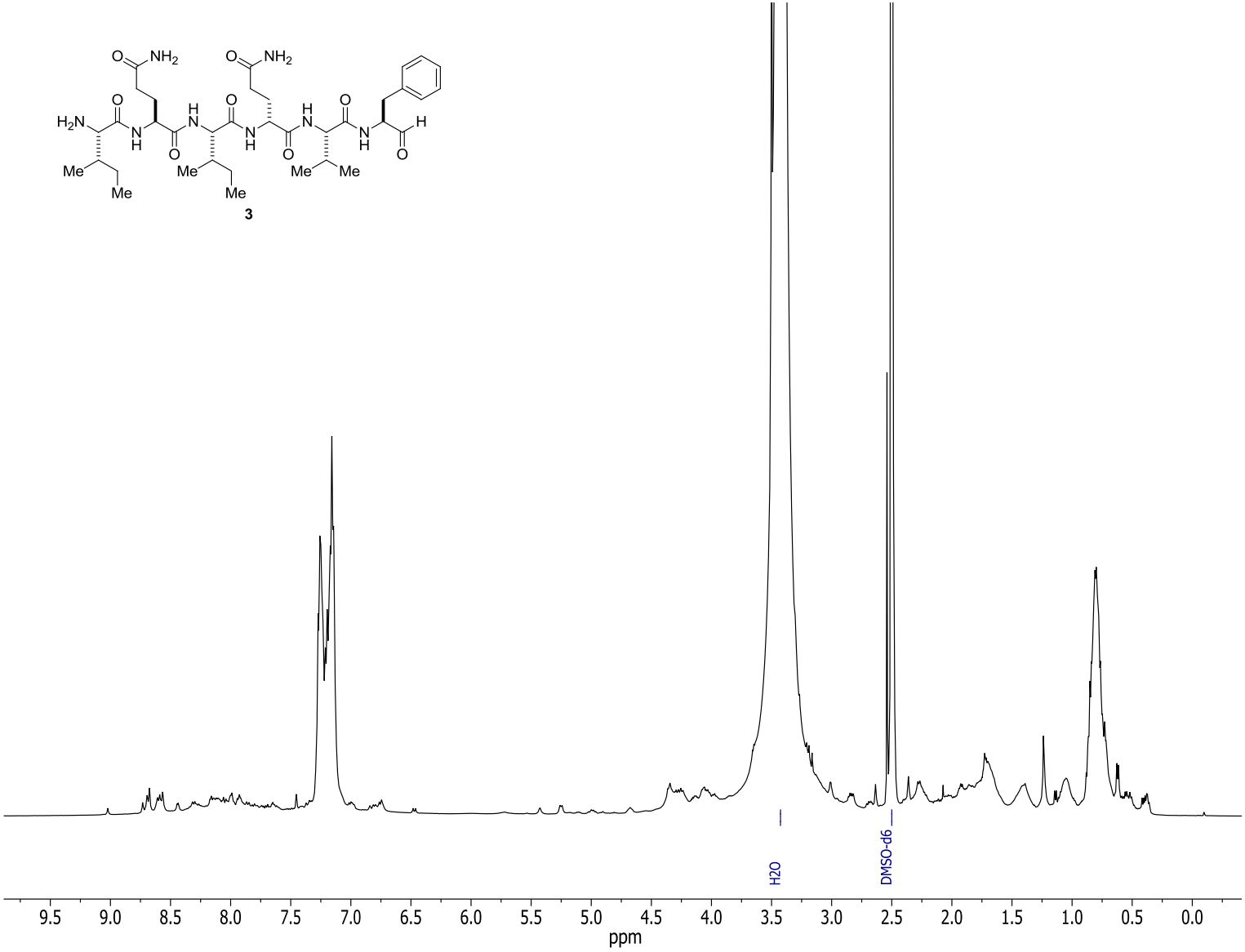



**Figure S18**. $^{1}$H NMR spectrum for **3** in $d_6$-DMSO.

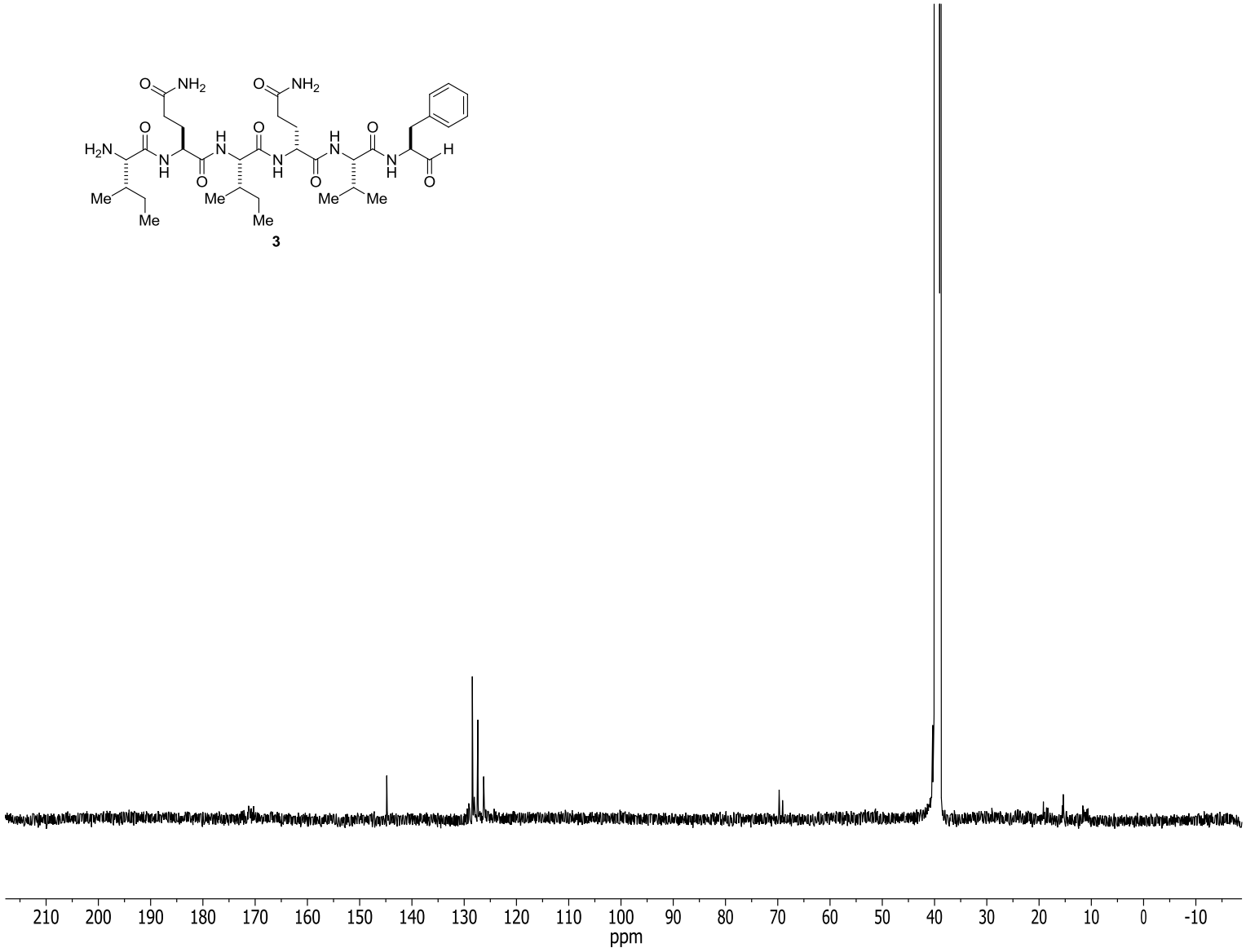
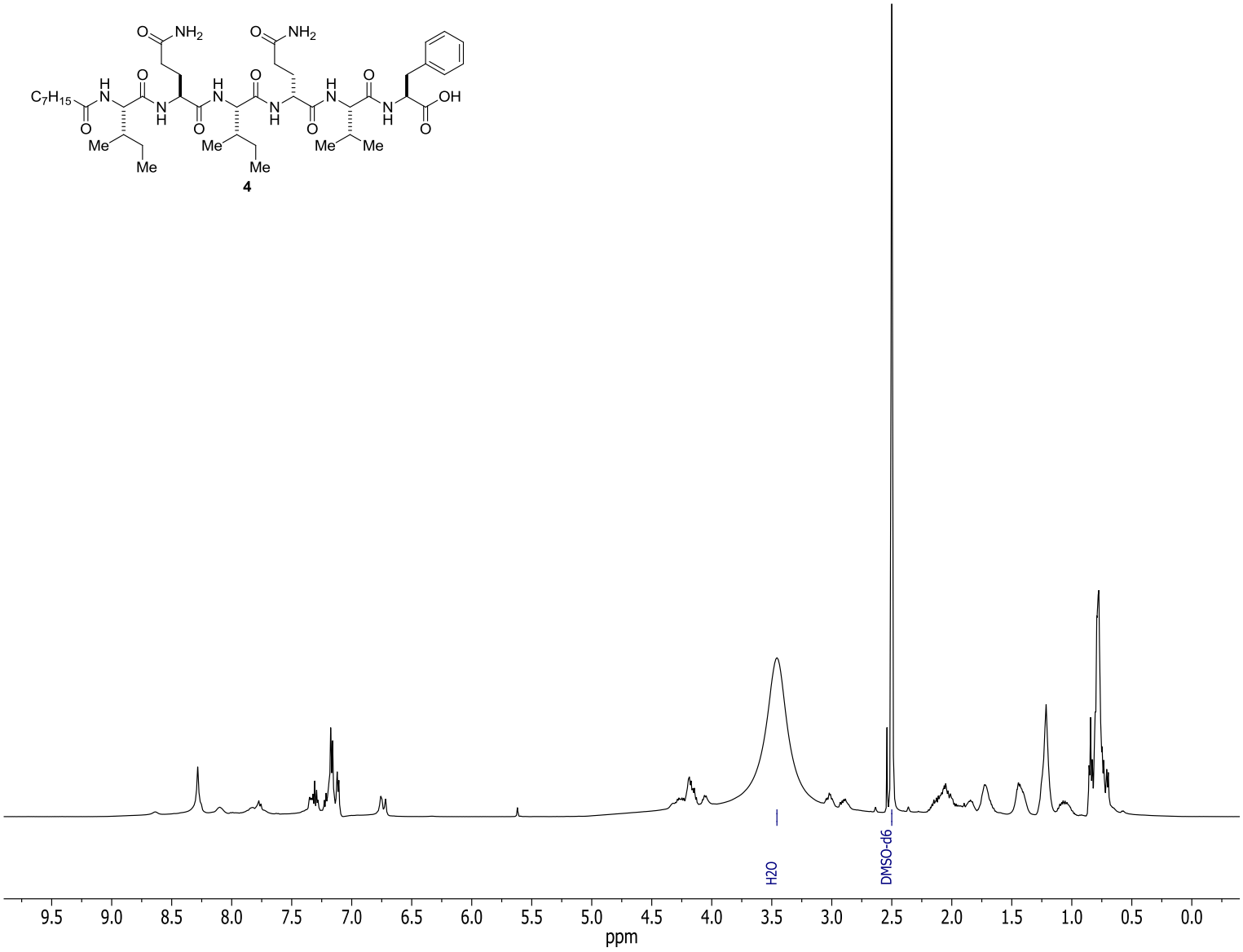



**Figure S19**. $^{13}$C NMR spectrum for **3** in $d_6$-DMSO.




**4**

**Figure S20**. $^{1}$H NMR spectrum for **4** in $d_6$-DMSO.

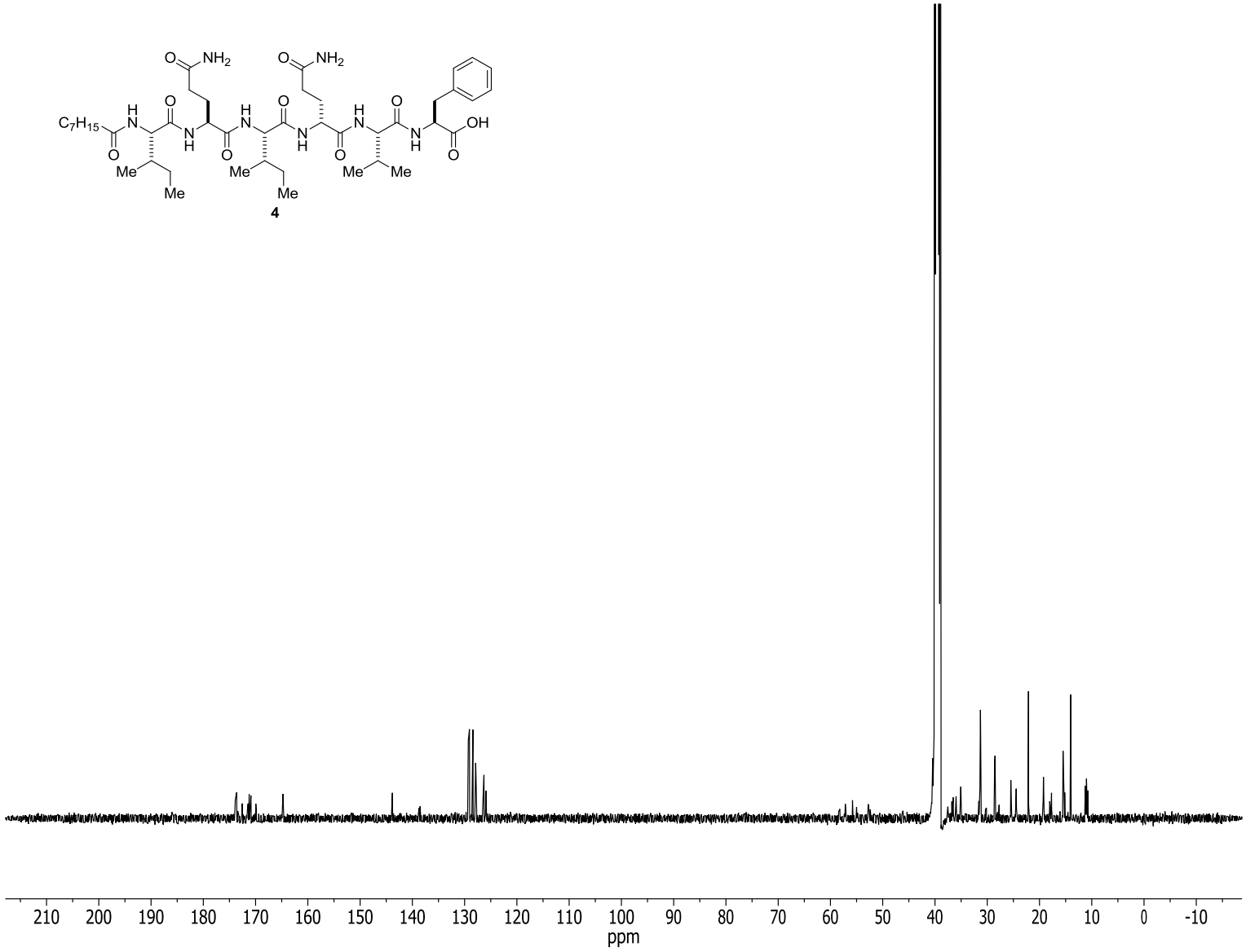



**Figure S21**. $^{13}$C NMR spectrum for **4** in $d_6$-DMSO.