

## Supporting Information

# Proteomics Guided Discovery of Flavo-peptins: 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.



```

FlavB RPLLTGASFLSAFLL-RDLVETTGEFVDCLVRAESQQAHAHVVRANLER 2182
MxcG QVLLTGATGFVSAHLL-DQLLRQTQAKVVCLVRADEAHAMELIREAMTS 1142
SfmC RVFLTGATGYLGLHLV-EQLLRRTDAEVVTLCRARDEQHALELKEGFAL 1150
NcpB -ILLTGATGFISAFLL-AELLQQTQADIYCLVRAANLSAGKQLQETLKA 4486
lgrD --LLTGATGFLSAFLL-RDLLQMTDADIYCLVRADEEEGMAHLRQTLEL 4765
KorD NIMITGATGFLSAHLNKLLEMEMEDVKIYCLVRFPSR----NKLKDTLMK 661
Tps IVFLTGATGFLSQELRLQLICNNAIASIITLVRAKSPDHGLDIRETAKI 20601
AusA KTLITGATGFLSAYLI--EALQGYSHRIYCFIRADNEEIAWYILMTNLND 2099

FlavB YGLWRPRYADLVHAVPSDLAAPGLLSAEDRDALVRR---LGTVVVNGA 2228
MxcG QRLSTASLSEIVLALPADLGGPWLGLSSARFHGLAAE---CMILNNA 1188
SfmC YEIDVEDQLHHSIAVIGDLAEPRGLTQEQWDDLAA---VVIYVNGA 1196
NcpB YLLWEEFNSRIIPVIGDLFQPLLGLGDEQFHFMARK---IDLIYNGA 4532
lgrD YELWNEEQAHRIIPVIGDLAKPRINLSEDFSELAEK---VIVLYNGA 4811
KorD YGLWQDDFVSRIVVIEGDLKHKQFLDDMTYEKLSND---VSQVYVGA 707
Tps AGWQENYTSRIEIVCSGLSKKRMGLSDVQWARLAGQSSNNVDAIVNGA 20651
AusA YFSEET----VEMMLSNIEVIVGDFECMDDVVLLEN---MVTIIHAA 2140

FlavB HVNFAAGYRDLAPNVAETEEELRLADSGSPGLHHSITSVYAPASGP 2278
MxcG VVSVVREYGSLOATNVRTRELLRLAASVRPKPLHYVSTLAV-APQANLS 1237
SfmC LVNFVYPYSALKAANVGTQRVLELACTTRLKAVHHVSTIDT-LLATHMP 1245
NcpB LVNHVYPYALLKAANVGTTEEVLRLASQIKIKPVHHSITSVVFASDEYFK 4582
lgrD LVNFVYPYAALKKANVLTTEEILRLAVAKTKPVHHSITFTFASEETEE 4861
KorD ETNFFEPYSKSKISNVDVVEMTKFASSYTRKNIYYASTLSVLTGERKWD 757
Tps IVNWNADYDKMKAANVSTVDLLKATVNSAASPKFIFVSGGKSDPTTDR 20701
AusA RTDHFQDDDEFKVNVDITVDVIRLAQQHHR-LIYVSTISVGTDFDIDT 2189
* *

FlavB PVTITESTPPGPPSALPDVYASQKWAERLVGLAR---ERGLPVTVHFP 2324
MxcG PEVPEAFVPAHP--GLRDCYQSKWAAERLVEQAS---ERGLPVTVYRL 1281
SfmC RPFLENDAPLHSAVGVPAQYTGSKWAEKVVDEAR---RRGLPVTVHFP 1291
NcpB LDVVQENDPLEHSQGLLGGYTSQKWAERKIVMMAR---DRGLPCSIYRL 4628
lgrD SMAFREEDMPENSRVLTSEYTSQKWAERHLVNLAR---ERGLPAAIYRC 4907
KorD -----EDELVYSPDLMIQYSQKWAERKLLQAR---EHGLTIDIFRL 798
Tps -----TILAQYLGNTSYTQTKFVSEGI IQEIIINTLPADQNRISTLEP 20744
AusA EDVTFSEADVYKQLLTSYTRSKFYSELKVLAV---KNGLDGRIVRV 2235
* *

FlavB GRISGDTTTCACQE---RDLWQLIKGCLQAGAV--DLPHGSTDWVVDY 2370
MxcG GRVSGALDSIVNP---QDLVWRILLAGIPAGAL--QLDVG-EVWTFVDY 1326
SfmC GLILGHTKNQATQT---IDYLLVALRGFLPMRIL--DYPRI-FDVIVDY 1336
NcpB GRITWHSQTAWNS---NDMFYRFIKSCIQLKSA--HEMNST-VEITVDY 4673
lgrD GRMTGDSETACQK---DMLMWRIAAGIIDLGKA--DMSGD-LDMMVDF 4952
KorD GRISNSN-NEVWNE---KMLYKVFESFIEQRIL--FKKEIHFELMVDV 843
Tps GRIGSPETIVANV---DMLWRVVSTAASLRVY--FAEPPEE-HWVSVADV 20789
AusA GNLTSPYNGRWHMRNIKTRNSMVMNDLLQLDCIGVSLAEMPVDFSFVDT 2285

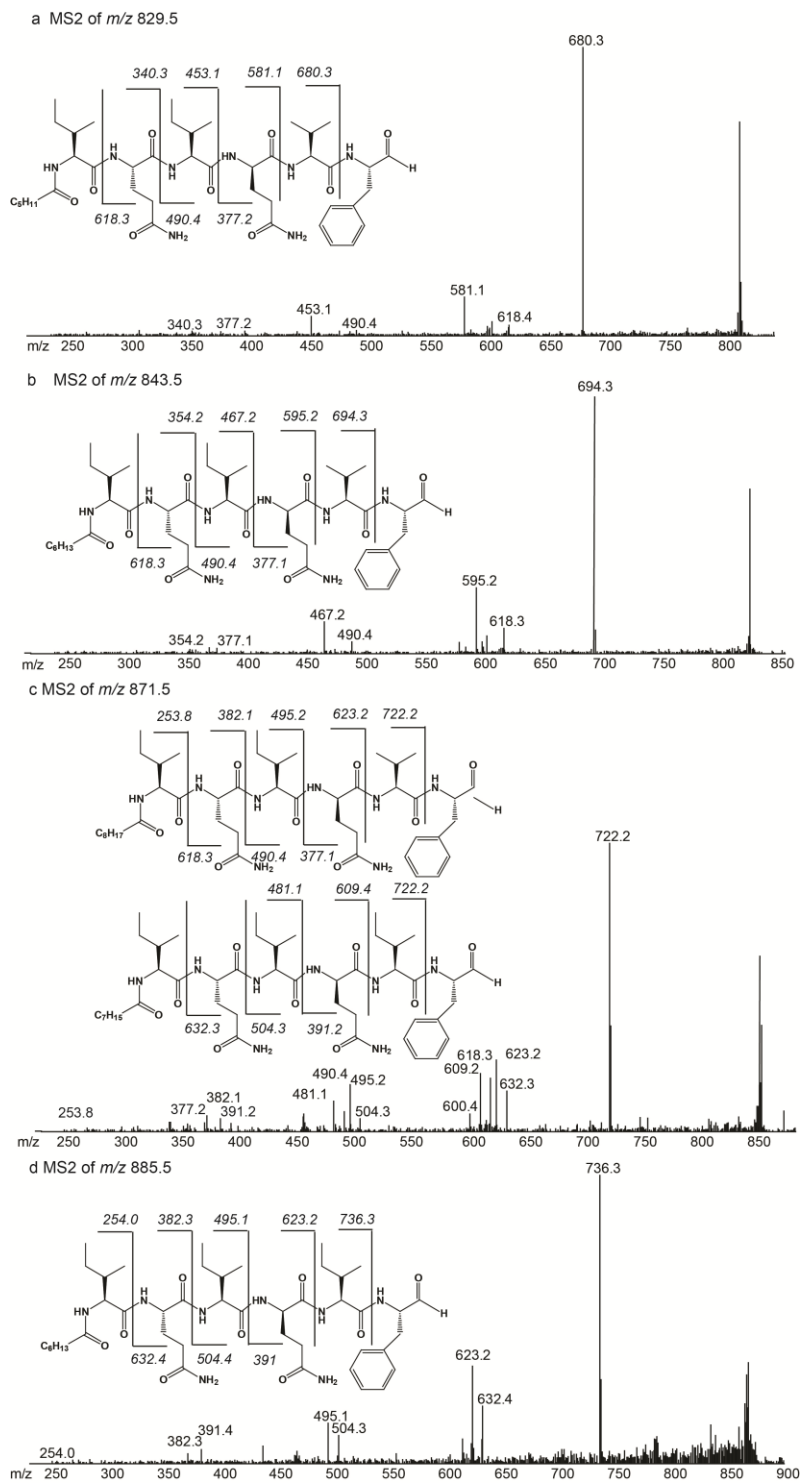
FlavB VSAAVVALATS-GRDAEVHHYTHPEAPGLDRVFEVAARLGHELRTVPAPQW 2421
MxcG VARALVRLSLV-PRPGTVFNLTPEPEVR-LSEVFGWVQDYGYPVALCPVPEW 1376
SfmC VASAIVHISRK-REAIIDGFYHLFNPAVPVPLLTFCDWIKSYGYEFDIVPFEEG 1387
NcpB LTKALIHLSQQ-PESLGKAFHLINSDSAPWSQFINCIRSLGYPLQLPYEDW 4724
lgrD ASKGI VHLSMT-EQSLSENFHLLNPNSTDYEDLISAIEDRGFVLERVMTDEW 5003
KorD VSEFIYKISRLNADQKLGIIYHMFNDQVSSEFVTSFFEKNEIPIYSNMDLEW 895
Tps TTVASSVLSQLYAKEGIAFFVSVAGGMPATIFWDIINKELDVPEPLSPDEW 20841
AusA TARQIVALAQV--NTPQIIYHVLSPNKMPVKSLLCEVKCKEIELVSDSEFNE 2335

```

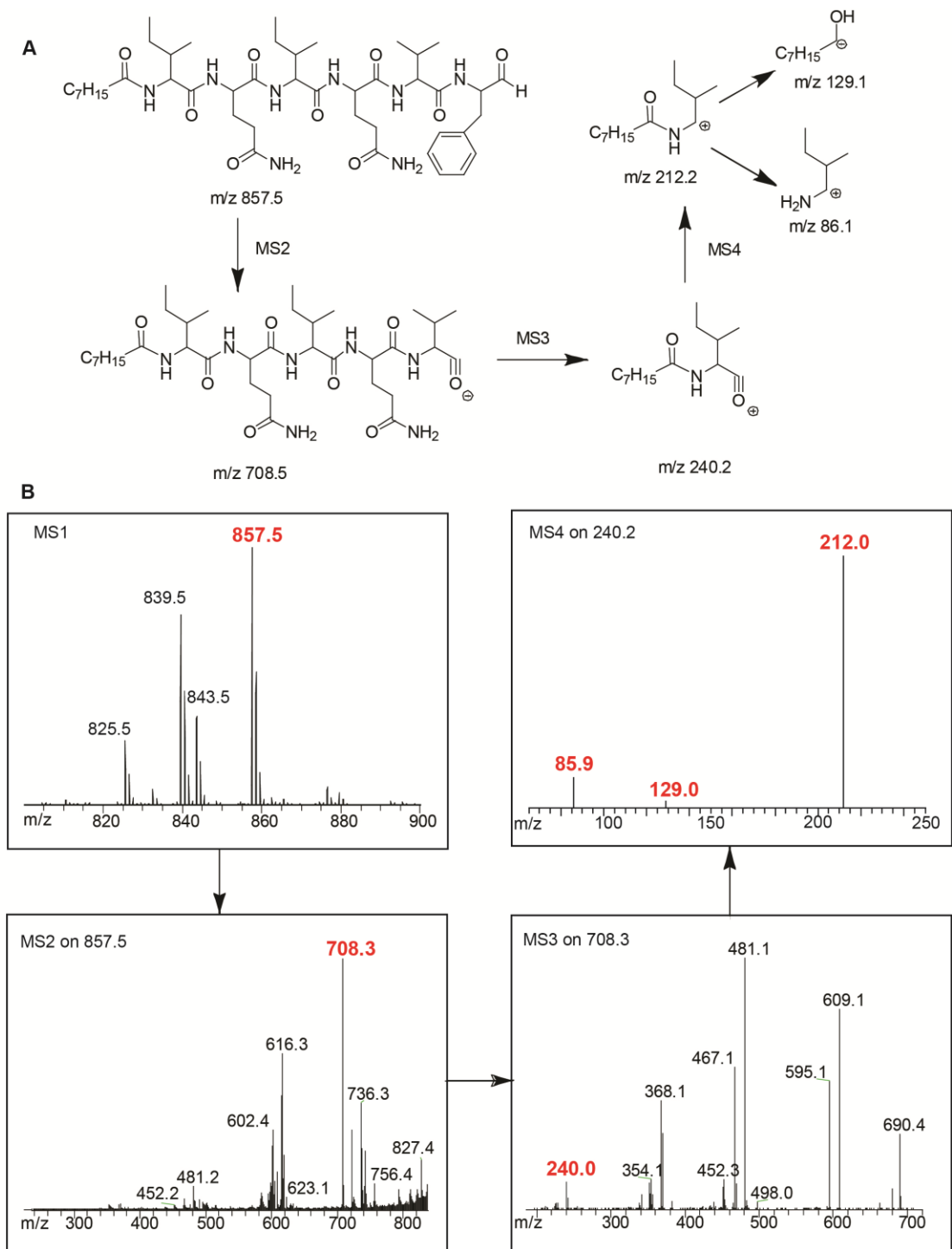
**Figure S2.** Sequence alignment of the reductase domain from FlavB with other NRPS terminal reductases:

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

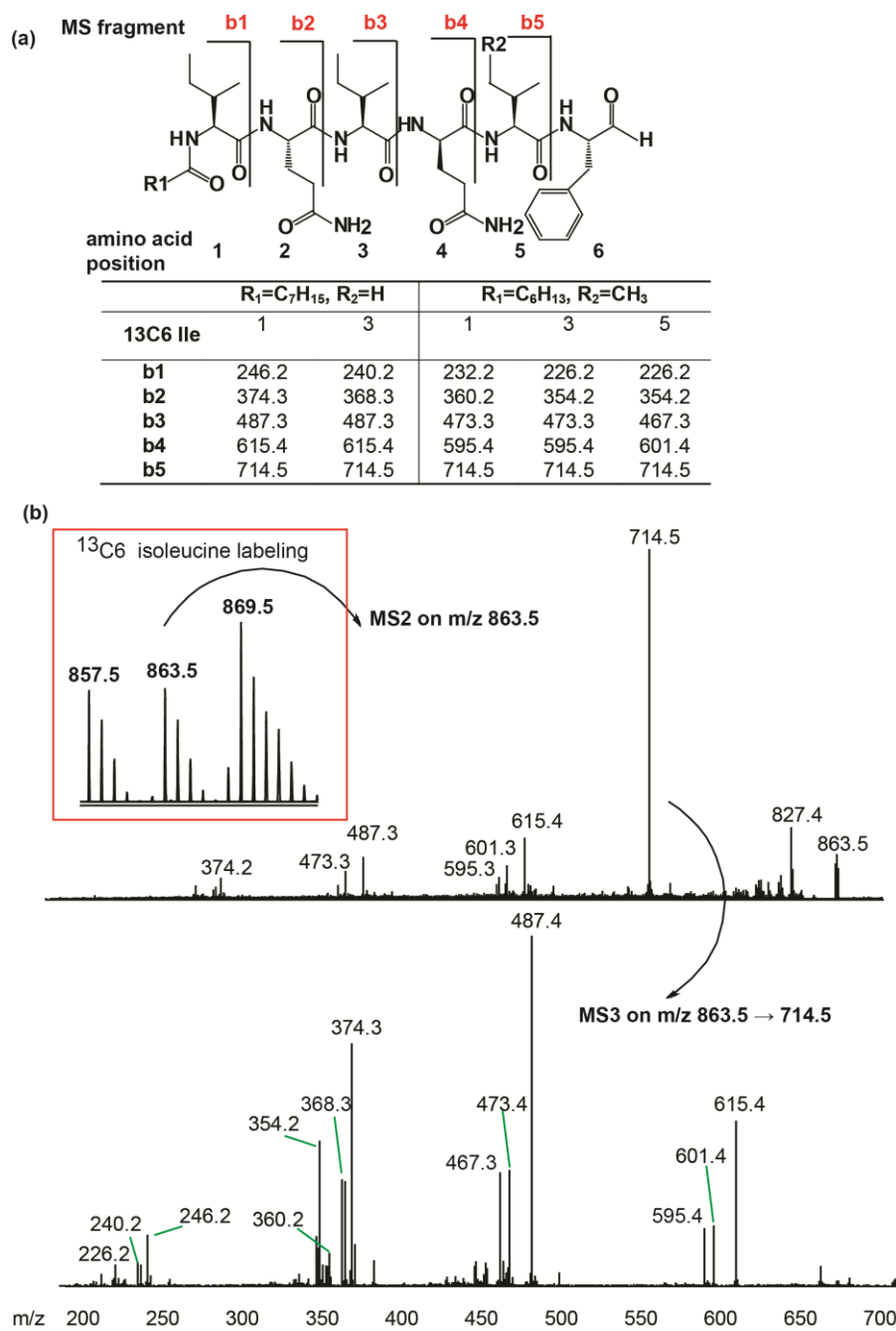
turquoise, proline and glycine in green, cysteine in pink and asparagine/glutamine in teal. Active site residues are indicated by a star \*\*. Residues that interact with NAD(P)H are enclosed in black boxes. The accession numbers for the synthetases indicated are: myxochelin (AAG31130), saframycin (ABI22133),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



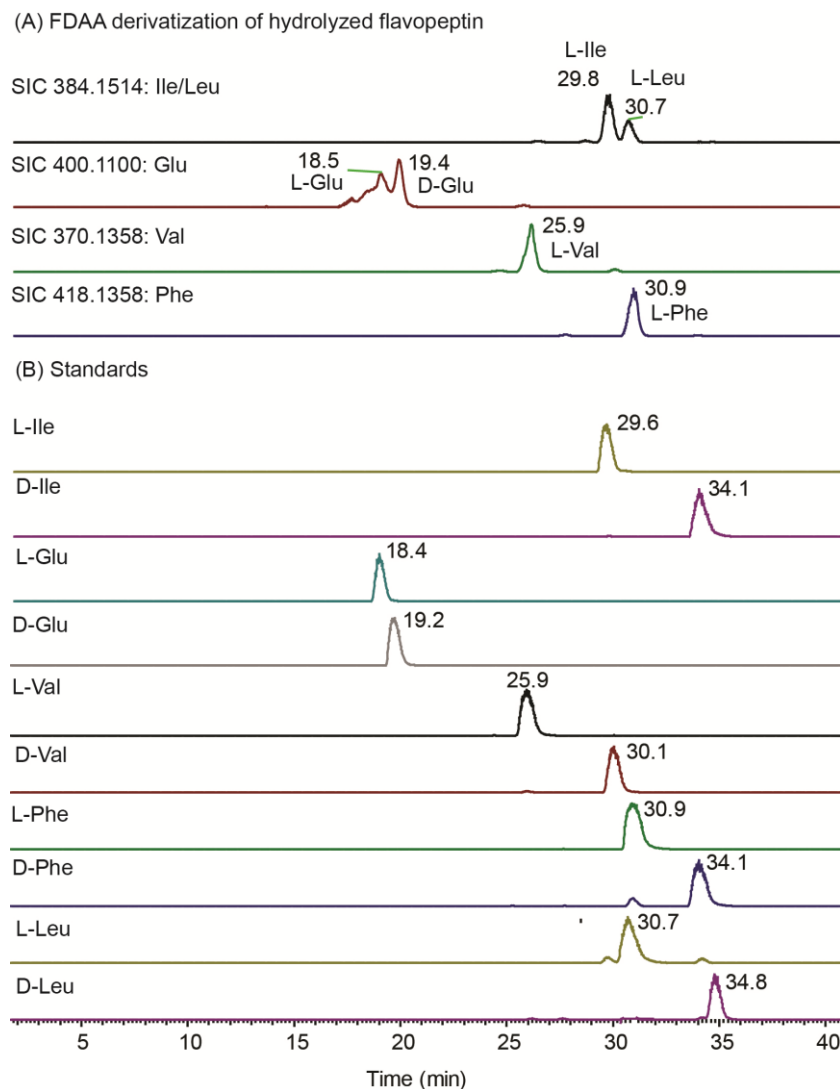
**Figure S3.** MS<sup>2</sup> fragmentation spectra for different flavopeptin species with *m/z* 829.5, 843.5, 871.5 and 885.5, respectively.



**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<sup>4</sup> spectrum on *m/z* 240.2 agrees with the predicted fragmentation pattern for the fatty acyl-isoleucine substructure.

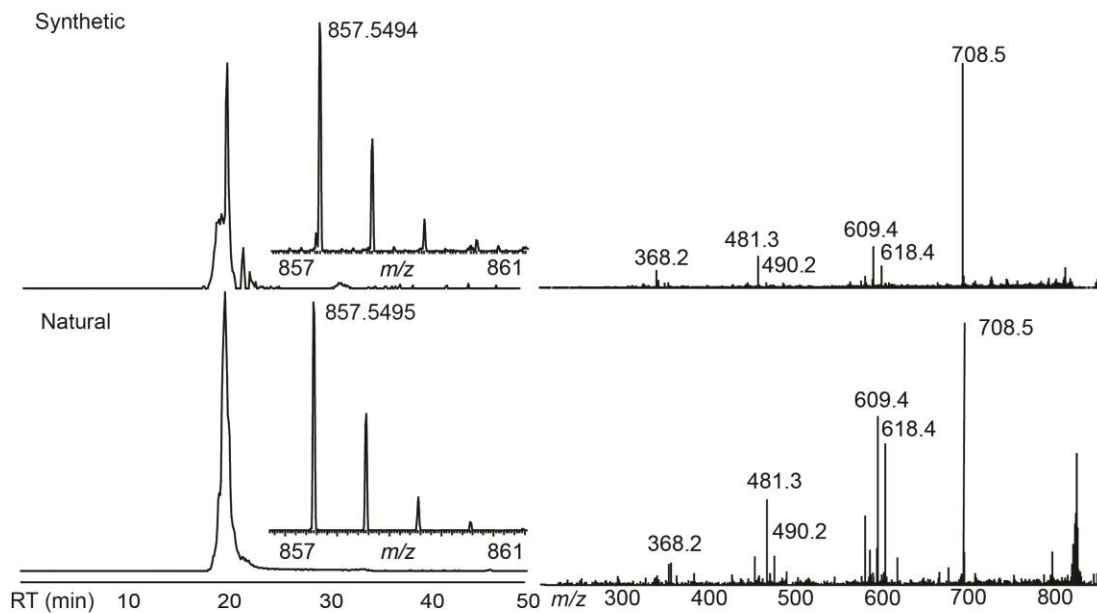


**Figure S5.** Tandem MS showing isoleucine can be incorporated into amino acid positions 1, 3 and 5. (a) Chemical structure and the fragment ion masses for  $^{13}C_6$ -isoleucine labeled flavopeptin (with  $m/z$  863.5) at all possible sites. (b) MS<sup>2</sup> and MS<sup>3</sup> fragmentation of 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

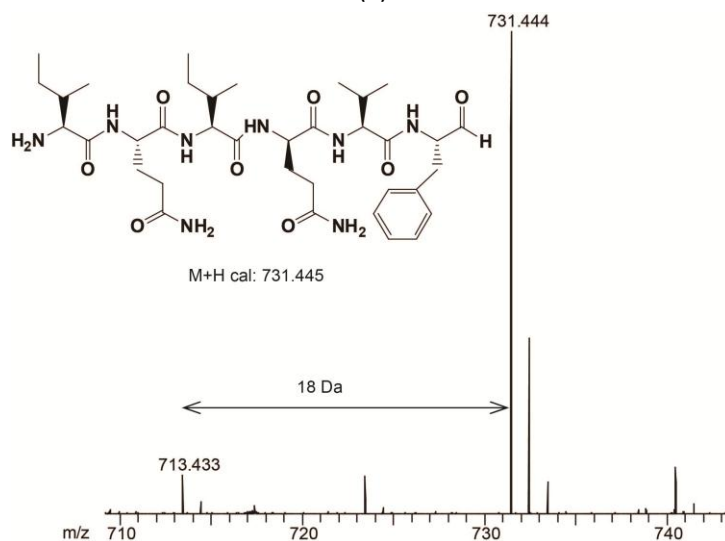


**Figure S6** Determination of the stereochemistry of amino acid building blocks of flavopeptins by Marfey's method. Flavopeptins were hydrolyzed in 6 N HCl at 110°C for 24 h before derivatization by the Marfey's reagent fluorodinitrophenyl-5-L-alanine amide (FDAA) followed by reverse-phase LC-MS analysis. (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.



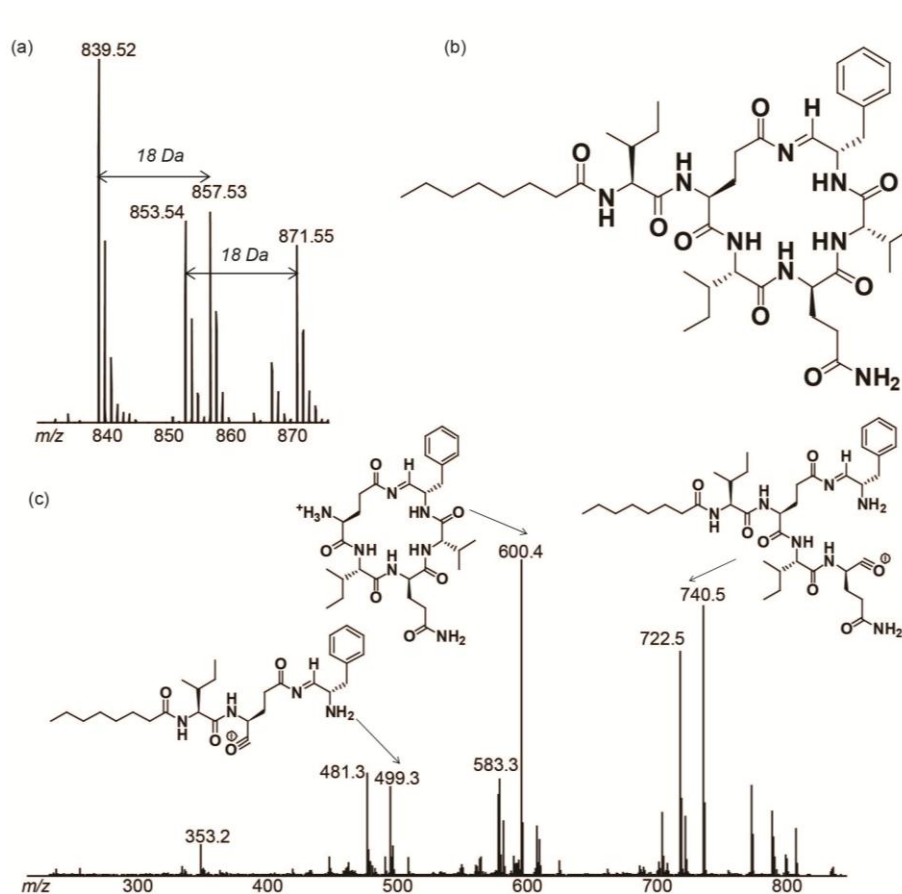


(a)

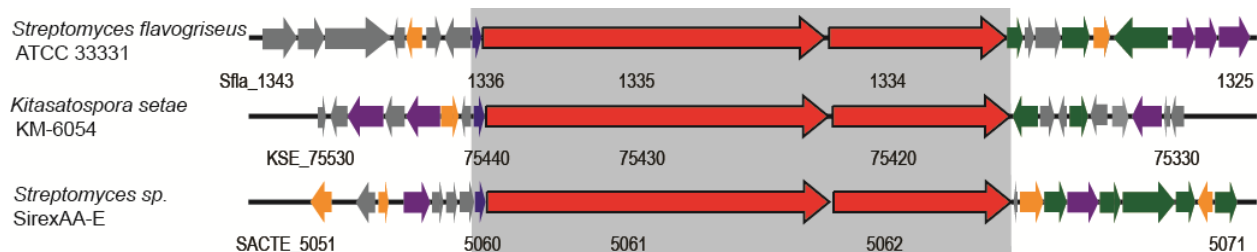


(b)

**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<sup>2</sup> 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.



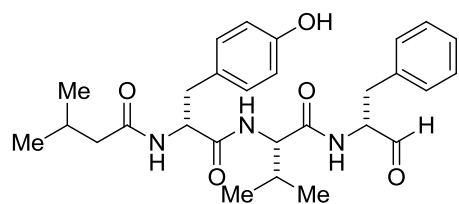
**Figure S8.** Dehydrated form of flavopeptins. (a) Mass spectrum of flavopeptins after organic solvent extraction and drying, showing the dehydrated products at  $m/z$  839.5 and 853.5. (b) Proposed structure for the dehydrated form at  $m/z$  839.5. (c) MS<sup>2</sup> spectrum of  $m/z$  839.5, with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



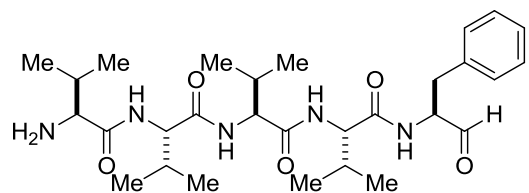
**Figure S9** Compare the flavopectin biosynthetic gene cluster among *Streptomyces flavogriseus* ATCC 33331 (top), *Kitasatospora setae* KM-6054 (middle) and *Streptomyces* sp. *SirexAA-E* (bottom). The three genes (*mbtH*, *flavA* and *flavB*) in shadow show high sequence similarity among the three 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.

<i>Streptomyces flavogriseus</i> ATCC 33331		<i>Kitasatospora setae</i> KM-6054		<i>Streptomyces</i> sp. <i>SirexAA-E</i>	
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

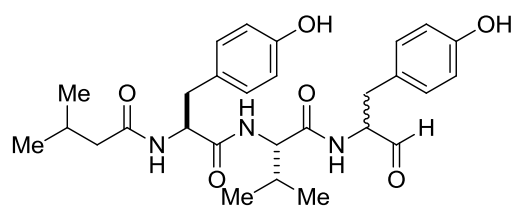
Sfla_1332	hypothetical protein	KSE_75400	hypothetical protein	SACTE_5064	AsnC transcriptional regulator
Sfla_1331	hypothetical protein	KSE_75390	hypothetical protein	SACTE_5065	polysaccharide deacetylase
Sfla_1330	acetyl xylan esterase	KSE_75380	methyltransferase	SACTE_5066	general substrate transporter
Sfla_1329	lclR 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



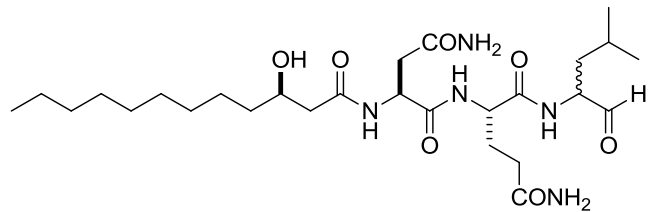
Nerfilin 1  
cysteine protease cathepsin and papain inhibitor



Stacopin P1  
cysteine protease inhibitor papain and calpain

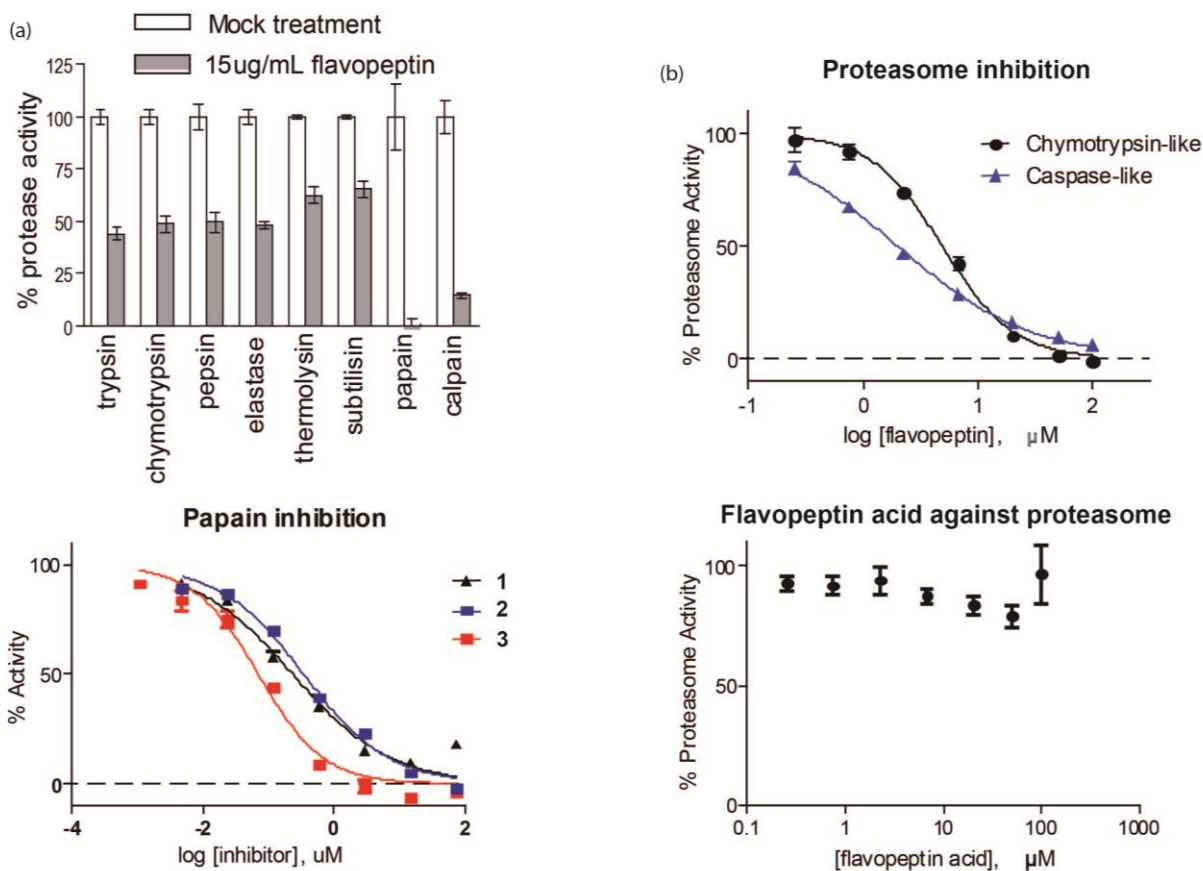


Tyropeptin  
inhibit proteasome, cystein proteases like  
cathepsin L and calpain

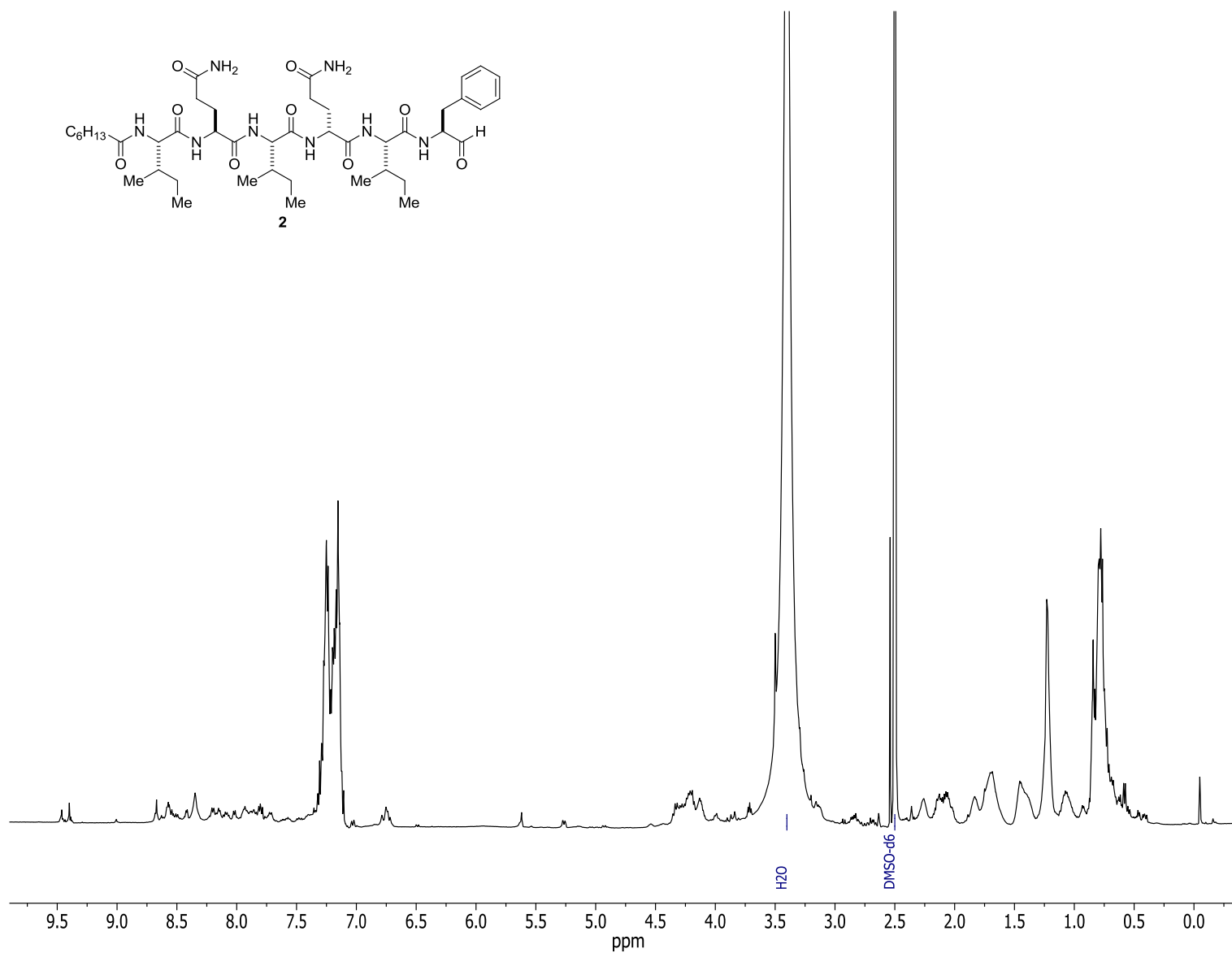


Fellutamide  
proteasome and cysteine protease cathepsin B  
inhibitor

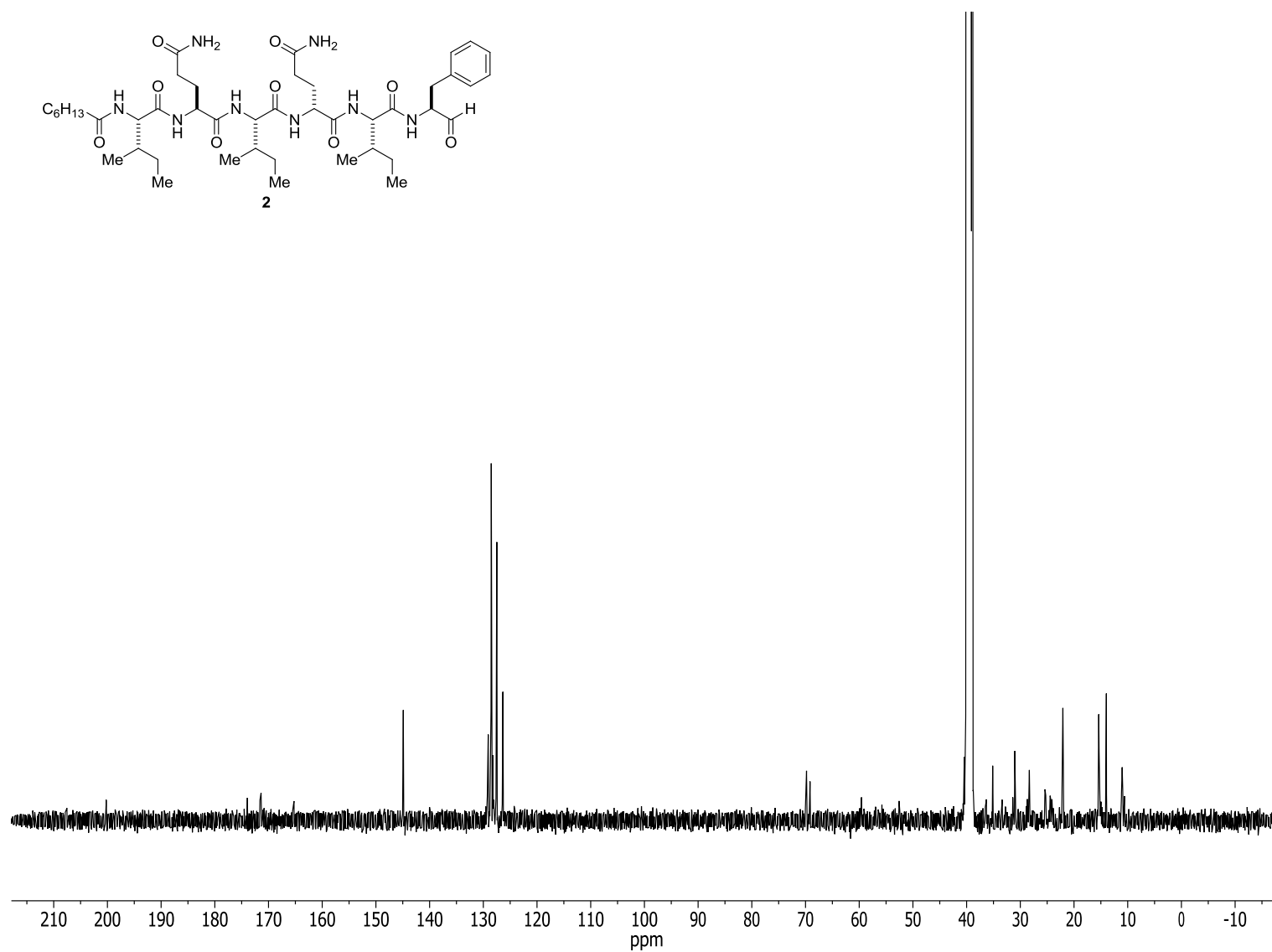
**Figure S10.** Other peptide aldehyde natural products that inhibit cysteine proteases and the proteasome.



**Figure S11.** Bioactivities of flavopeptides. (a) Top: Flavopeptides 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  $\text{IC}_{50}$  plots against papain using the three synthetic flavopeptides (**1**, octanoyl-I-Q-I-Q-V-F-CHO, **2**, heptanoyl-I-Q-I-Q-I-F-CHO and **3**,  $\text{NH}_2$ -I-Q-I-Q-V-F-CHO). (b) Flavopeptides showed micromolar range inhibition against the chymotrypsin and caspase-like activities of the human 20S proteasome. Top, the  $\text{IC}_{50}$  plot for Compound **1**. Bottom, the flavopeptin acid derivative (**4**) did not inhibit the chymotrypsin-like activity of proteasome up to 100  $\mu\text{M}$ . All data points were in triplicates.

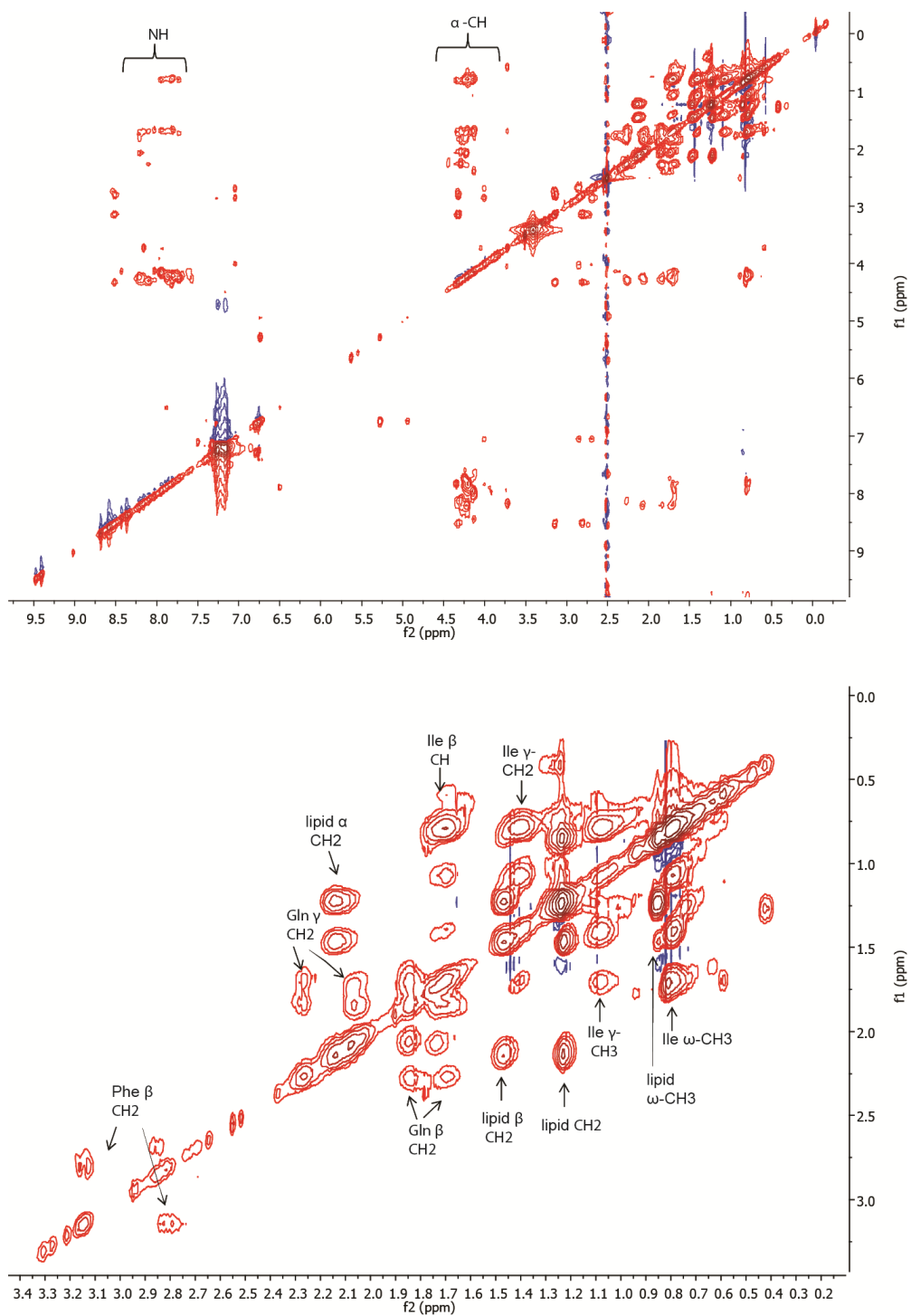


**Figure S12.** <sup>1</sup>H NMR spectrum for **2** in *d*<sub>6</sub>-DMSO.

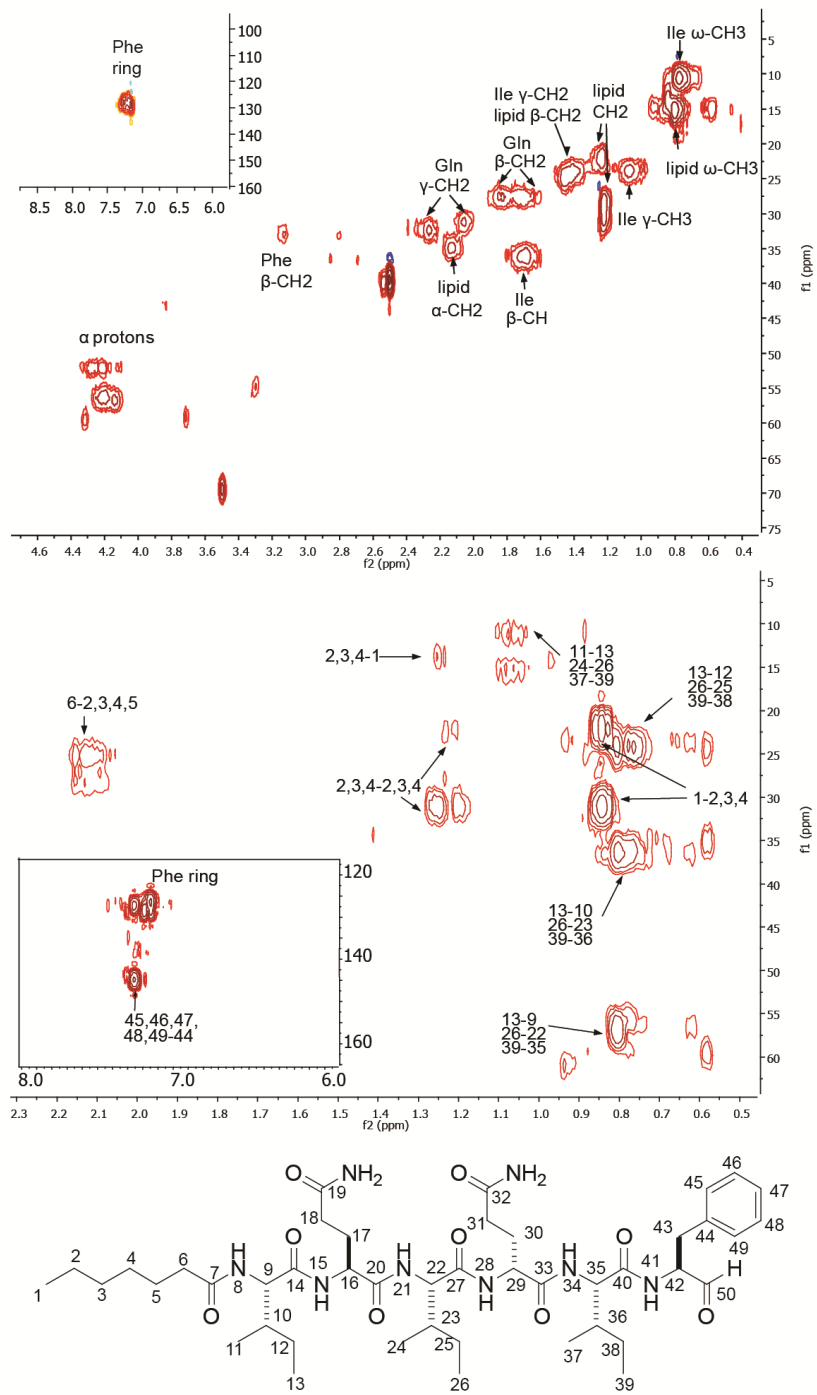


**Figure S13.**  $^{13}\text{C}$  NMR spectrum for **2** in  $d_6$ -DMSO.

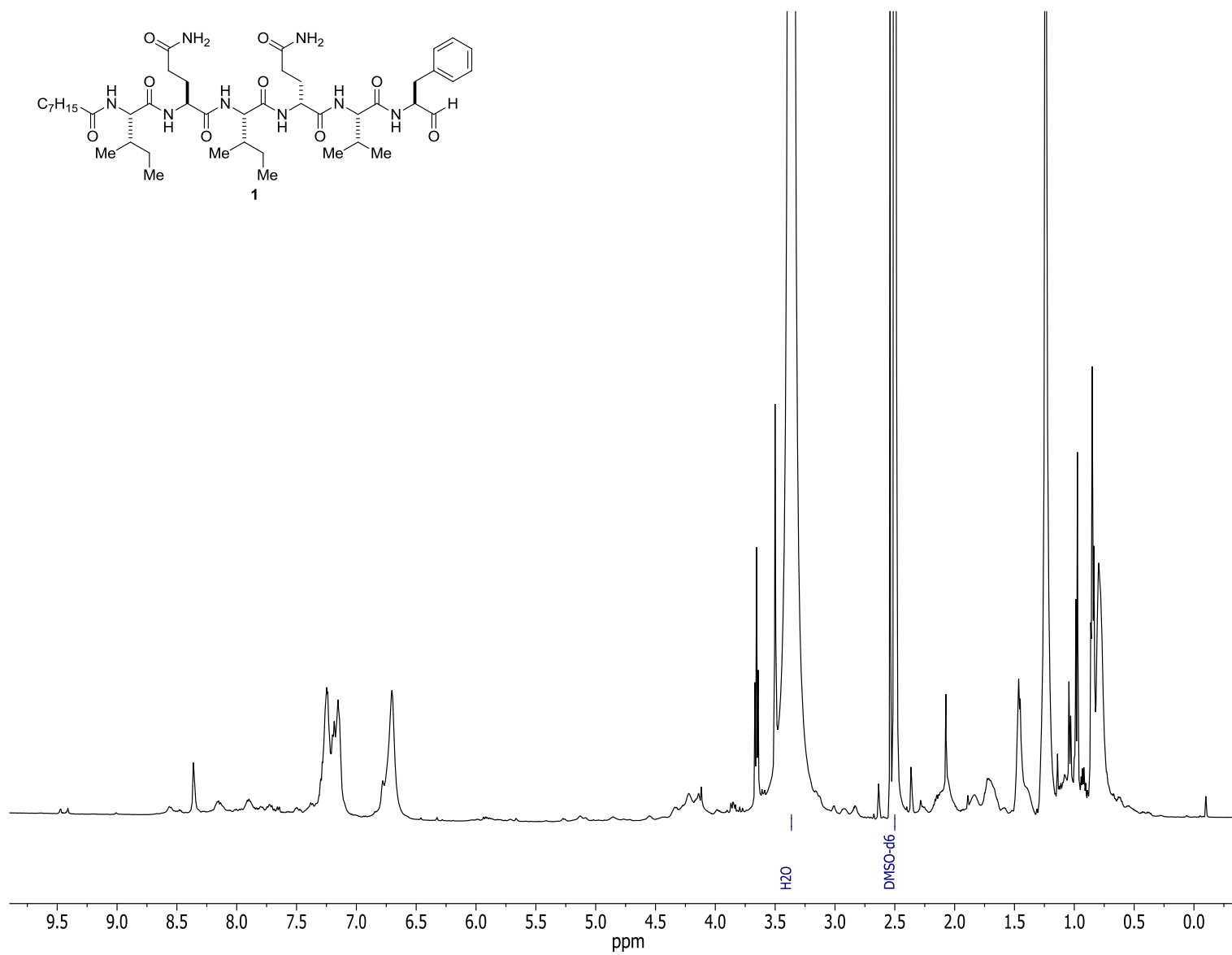




**Figure S14.** Two-dimensional homonuclear ( $^1\text{H}$ ) TOCSY spectrum for **2** in  $d_6$ -DMSO. Top: full TOCSY spectrum with amide and  $\alpha$  proton regions highlighted. Bottom: side chain signals and associated spin systems are highlighted.

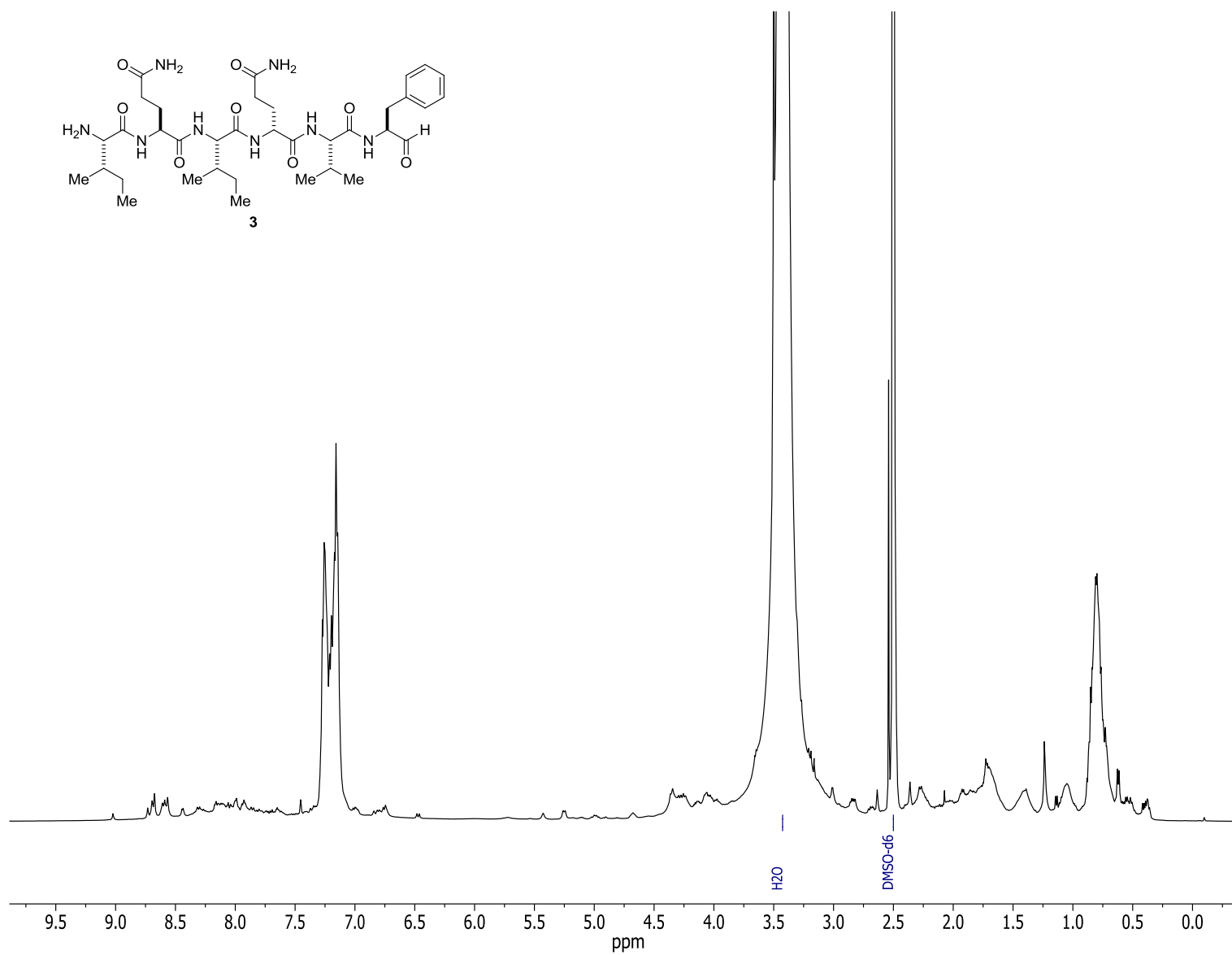


**Figure S15.** Heteronuclear ( $^1\text{H}$ - $^{13}\text{C}$ ) NMR spectra for **2** in  $d_6$ -DMSO with assignments. Top: HSQC spectrum. Bottom: HMBC spectrum.



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





**Figure S18.** <sup>1</sup>H NMR spectrum for **3** in d<sub>6</sub>-DMSO.

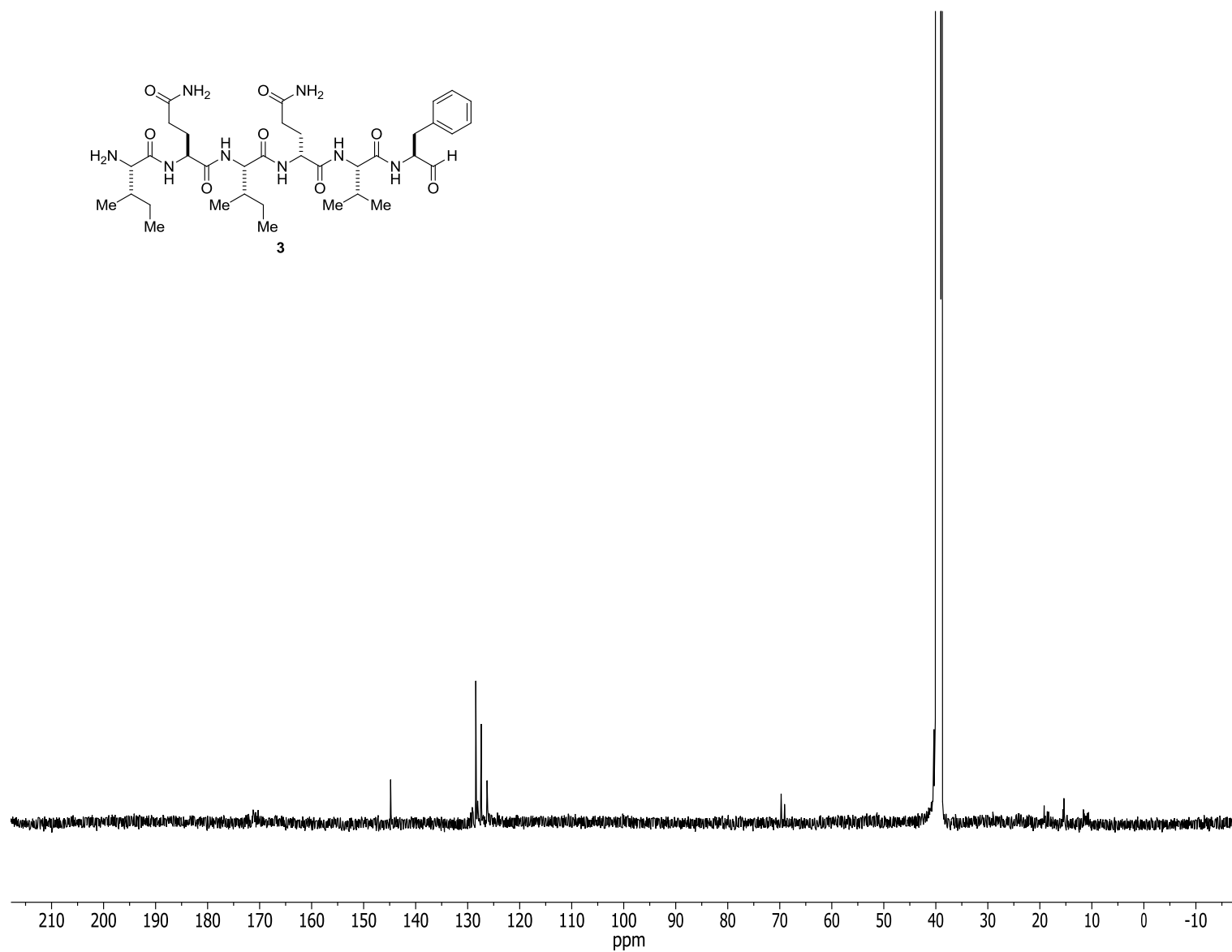
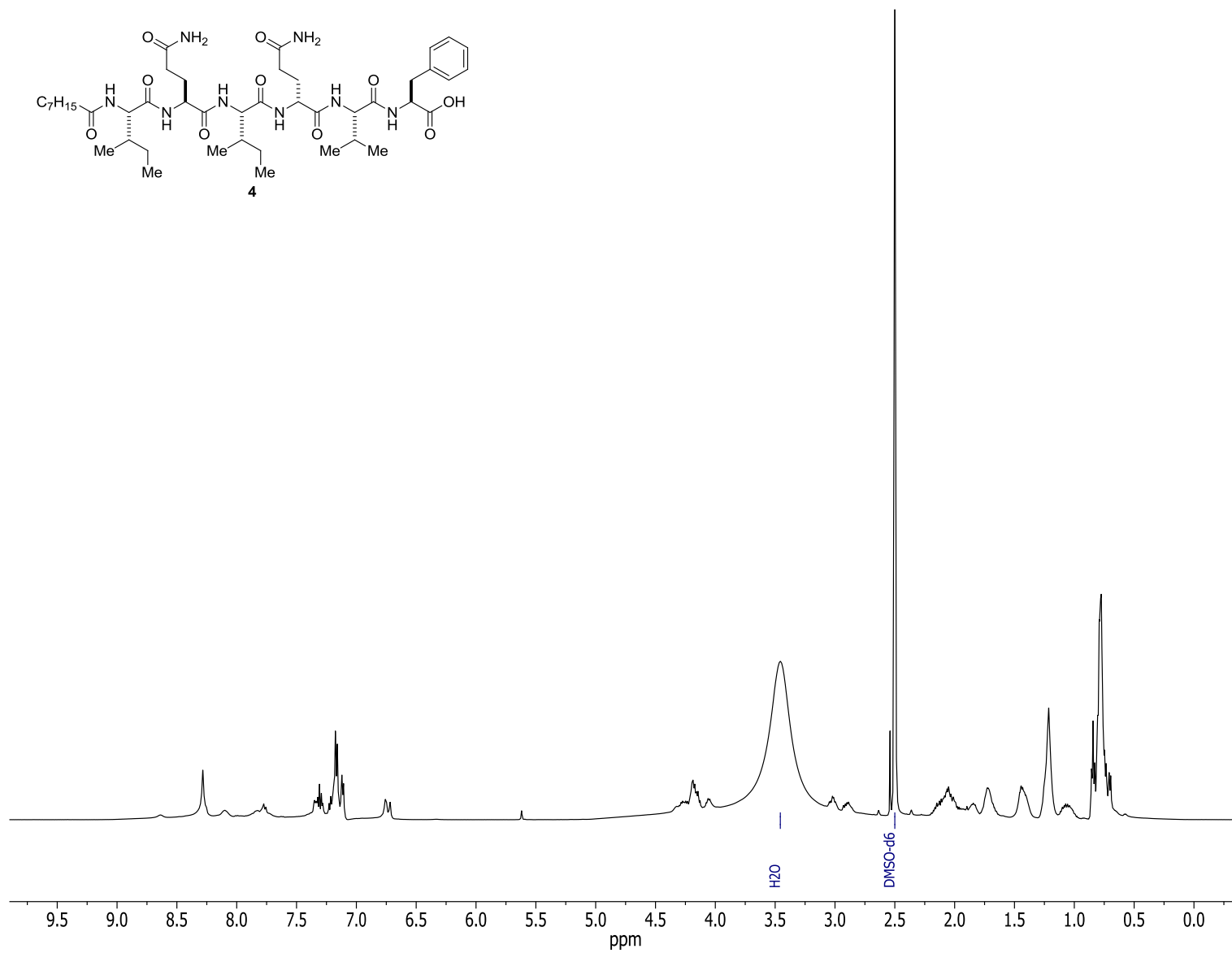


Figure S19.  $^{13}\text{C}$  NMR spectrum for **3** in  $d_6$ -DMSO.



**Figure S20.** <sup>1</sup>H NMR spectrum for **4** in *d*<sub>6</sub>-DMSO.

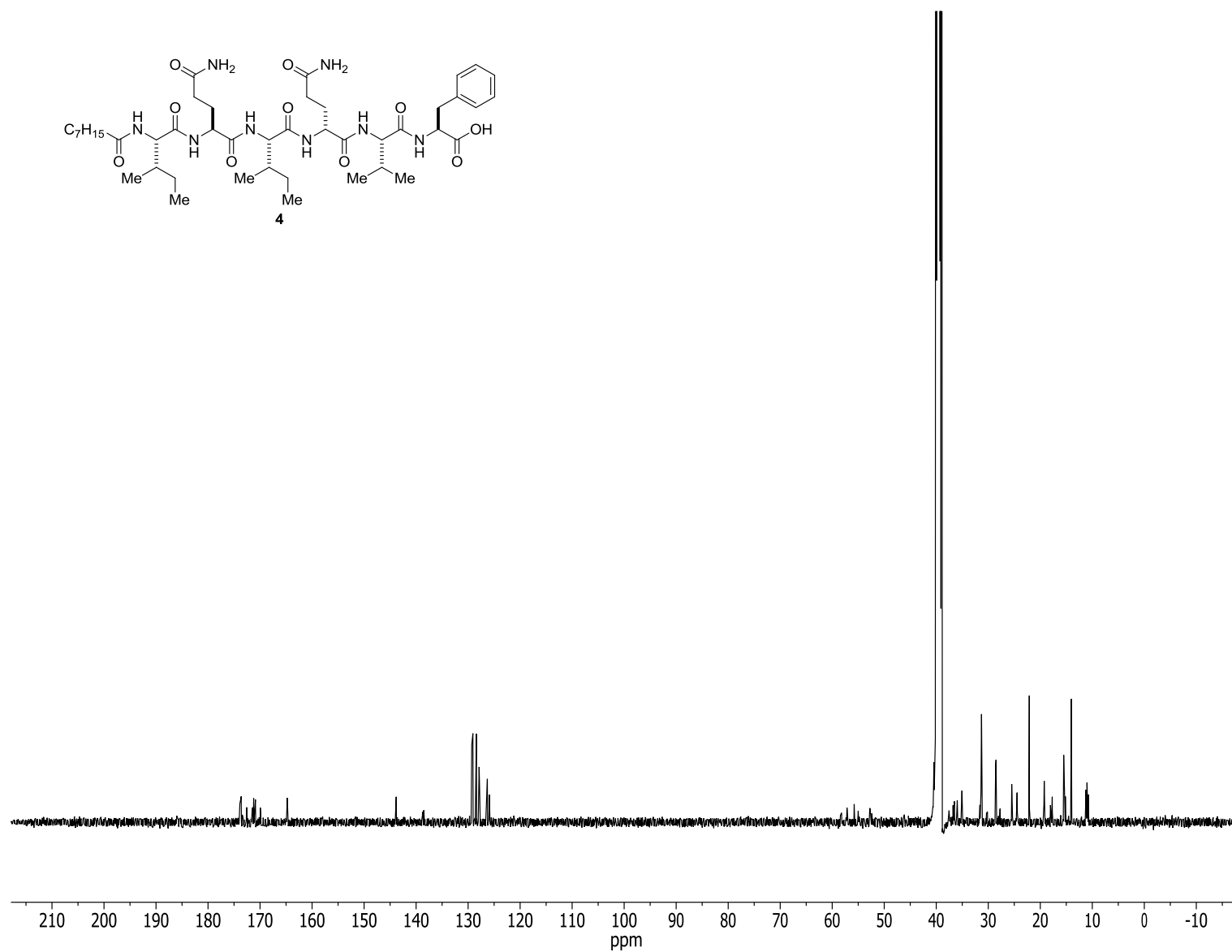


Figure S21.  $^{13}\text{C}$  NMR spectrum for **4** in  $d_6$ -DMSO.