Commensal bacteria produce GPCR ligands that mimic human signaling molecules

Authors: Louis J. Cohen[1,2], Daria Esterhazy[3], Seong-Hwan Kim[1], Christophe Lemetre[1], Rhiannon R. Aguilar[1], Emma A. Gordon[1], Amanda J. Pickard[6], Justin R. Cross[6], Ana B. Emiliano[4], Sun M. Han[1], John Chu[1], Xavier Vila-Farres[1], Jeremy Kaplitt[1], Aneta Rogoz[3], Paula Y. Calle[1], Craig Hunter[5], J. Kipchirchir Bitok[1], Sean F. Brady[1]

Author Affiliation:
[1] Laboratory of Genetically Encoded Small Molecules, The Rockefeller University
[2] Division of Gastroenterology, Department of Medicine, Icahn School of Medicine at Mount Sinai
[3] Laboratory of Mucosal Immunology, The Rockefeller University
[4] Laboratory of Molecular Genetics, The Rockefeller University
[5] Comparative Biosciences Center, The Rockefeller University
[6] Donald B. and Catherine C. Marron Cancer Metabolism Center, Memorial Sloan Kettering Cancer Center

Author Contributions:
L.J.C. and S.F.B. designed research; L.J.C. assisted with all experiments; S-H.K. assisted with molecule characterization; E.A.G., P.Y.C., J.K.B., and R.R.A. assisted with gene cloning; D.E., A.B.E., S.M.H., C.H. and A.R. assisted with mouse experiments; J.C., X.V-F., J.K. assisted with molecule synthesis; A.J.P. and J.R.C. assisted with metabolite analysis in human/mouse samples; L.J.C. and C.L. analyzed data; L.J.C. and S.F.B. wrote the paper

Author Information:
Authors declare no financial competing interests. Correspondence and requests for materials should be addressed to S.F.B. (sbrady@rockefeller.edu).

Contact: Laboratory of Genetically Encoded Small Molecules, The Rockefeller University, 1230 York Avenue, New York, New York, 10065

Keywords: GPCR, microbiome, metagenome, signaling, *N*-acyl amide

Supplementary NMR Discussion, Figures and Tables

Compound isolation and NMR structure determination overview

For each compound family, dried ethyl acetate extracts were partitioned by reversed phase flash column chromatography (Teledyne Isco, C_{18} RediSep RF Gold^{Tm} 15 g) using the following mobile phase conditions: solvent A:B (water:acetonitrile with 0.1% formic acid) 10% B isocratic for 5 min, gradient to 100% B over 20 minutes (30 ml/min). Fractions containing clone specific metabolites as identified by LCMS were pooled and semi-preparative reversed phase HPLC was used to separate individual *N*-acyl amide molecules (Waters XBridge^{Tm} C_{18}, 10 mm x 250 mm: 4.8 ml/min: solvent A:B, water:acetonitrile with 0.1% TFA). Chromatographic details for each metabolite analyzed by NMR.

Molecule **2**: retention time 8.5 min, gradient 85% B to 100% B over 20 min
Molecule **3:** retention time 13 min, isocratic 70% B
Molecule **4a:** retention time 17 min, isocratic 40% B for 5 min, gradient from 40% to 72% B over 15 min
Molecule **4b:** retention time 16 min, isocratic 40% B for 5 min, gradient from 40% to 72% B over 15 min
Molecule **5:** retention time 9.5 min, isocratic 60% B for 5 min, gradient from 60% to 100% B over 10 min
Molecule **6:** retention time 17 min, isocratic 50% B for 5 min, gradient from 50% to 100% B over 15 min

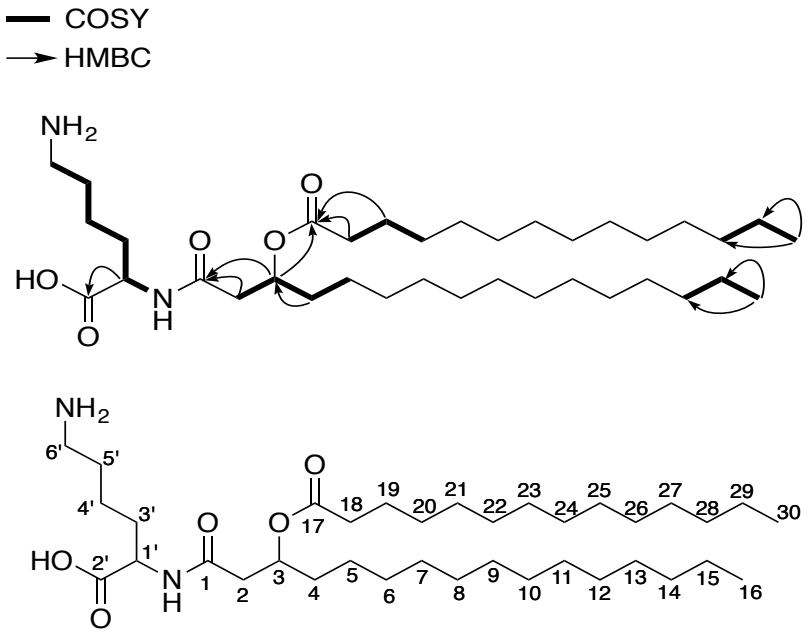
Each *E. coli* strain transformed with a single hm-NAS gene produced a family of related *N*-acyl amides. With the exception of compound family 4 (compounds **4a** and **4b**), ^{1}H NMR and MS analysis indicated that cultures produced metabolites with the same amine head group but different acyl substituents. Based on MS data acyl substituents were predicted to be fully saturated or mono-unsaturated and differ only slightly in length. The most common acyl chains incorporated by hm-NASs are from 14-18 carbons in length. These can be modified by β-hydroxylation or a single unsaturation. In the case of family 4 ^{1}H NMR data suggested two different amine head groups. The major *N*-acyl amide produced by weight for each family was selected for in depth structural analysis. In the case of family 4**,** the major *N*-acyl amide by weight for each head group was structurally characterized (compounds **4a** and **4b**). The following pages include in depth structure determination analyses for each purified compound.

Family 2 – compound 2

For family 2 we were not able to completely separate individual *N*-acyl amides from each other. The material we structurally characterized contained a mixture of two *N*-acyl amides (one major and one minor metabolite) that differed by 26 units (*m/z*: $[M+H]^+$ 611, 637). The HRMS predicted molecular formula of the dominant compound in the mixture was $C_{36}H_{70}N_2O_5$ (*m/z*: $[M+H]^+$ Calcd $C_{36}H_{71}N_2O_5$ 611.5363; found 611.5385). The COSY spectrum defined 5 spin systems. The 5-carbon-NH COSY spin system together with an HMBC correlation between H-1′(δ_H 3.03) and C-2′ (δ_C 176.8) supported the presence of an N-acylated lysine substructure. The 4-carbon COSY spin system together with an HMBC correlation from H-2 to the C-1 carbonyl indicated the presence of a C-3 oxidized fatty acid. The presence of two separate acyl chains was suggested by the presence of two terminal methyl triplets in the ^{1}H NMR spectrum. The appearance of the mass for *N*-hydroxypalmitoyl lysine in the HR MS/MS fragmentation analysis allowed us to define the structure of one acyl substituent as [C16:3-OH]. The length of the second acyl group was predicted based on the predicted molecular formula [C14]. Based on this analysis the final structure of compound **2** was determined to be 3-(myristoyloxy)palmitoyl lysine. Based on the small olefinic proton signal (δ_H 5.32) in the ^{1}H NMR spectrum and MS/MS fragmentation analysis the minor compound in the mixture (*m/z*: $[M+H]^+$ 637) was predicted to contain a double bond and longer acyl substituents.

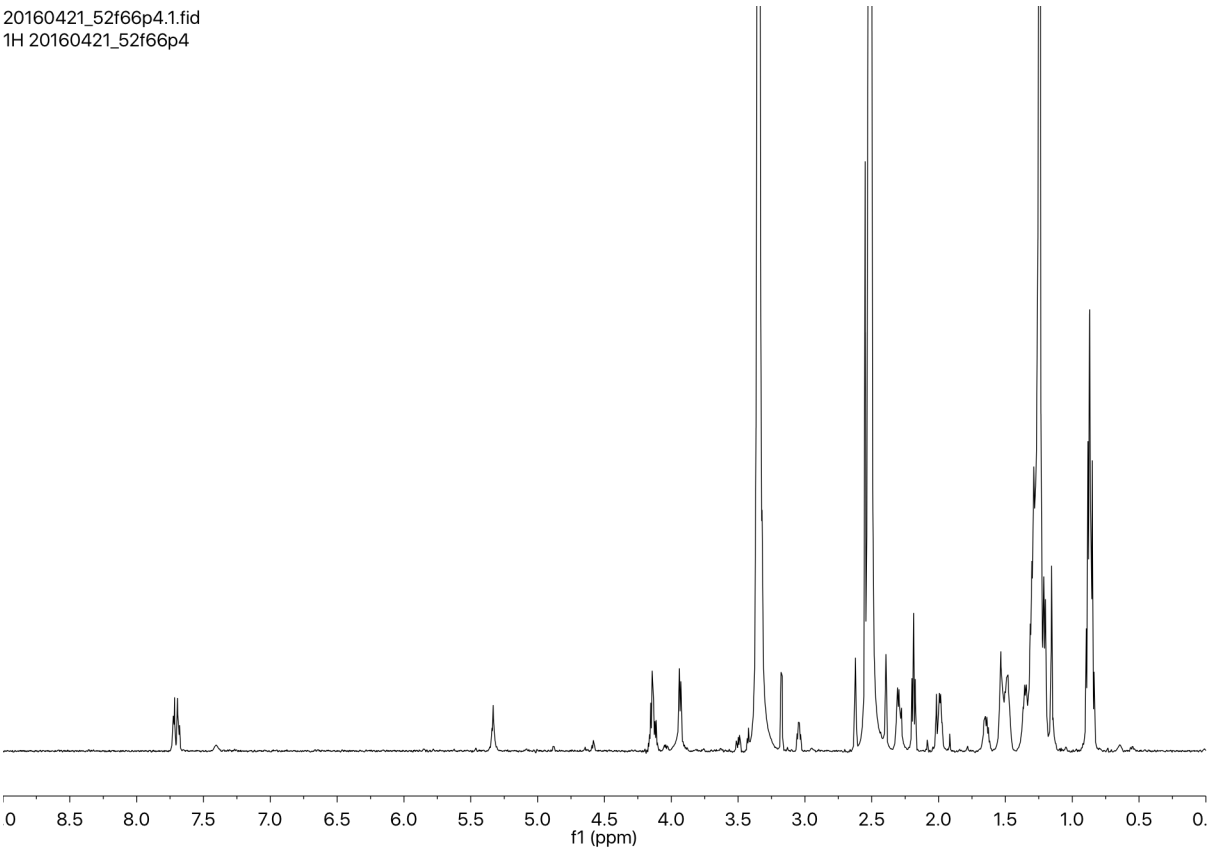
NMR Table Compound 2

Position	Compound 2 [a c]	
	δ_C, type	δ_H, mult (*J* in Hz)
1	172.8, C	
2	33.3, CH_2	2.30, dd (6.5, 6.5)
3	65.3, CH	3.92, m
4	31.0, CH_2	1.21, m [b]
5	28.7, CH_2	1.23, m [b]
6	28.8, CH_2	1.22, m [b]
7-13 [b]	28.8, CH_2	1.22, m [b]
14	31.0, CH_2	1.23, m [b]
15	22.8, CH_2	1.29, m [b]
16	13.6, CH_3	0.86, t (7.0)
17	174.8, C	
18	33.3, CH_2	2.18, t (7.5)
19	24.5, CH_2	1.47, m
20-27 [b]	28.8, CH_2	1.22, m [b]
28	29.4, CH_2	1.23, m [b]
29	23.0, CH_2	1.29, m [b]
30	10.5, CH_3	0.86, t (7.0)
NH		7.69, d (7.0)
1′	41.0, CH	3.03, dd (6.0, 6.0)
2′	176.8, C	
3′	24.1, CH_2	1.47, m
4′	22.0, CH_2	1.22, m
5′	23.8, CH_2	1.50, m
6′	36.0, CH_2	2.26, m

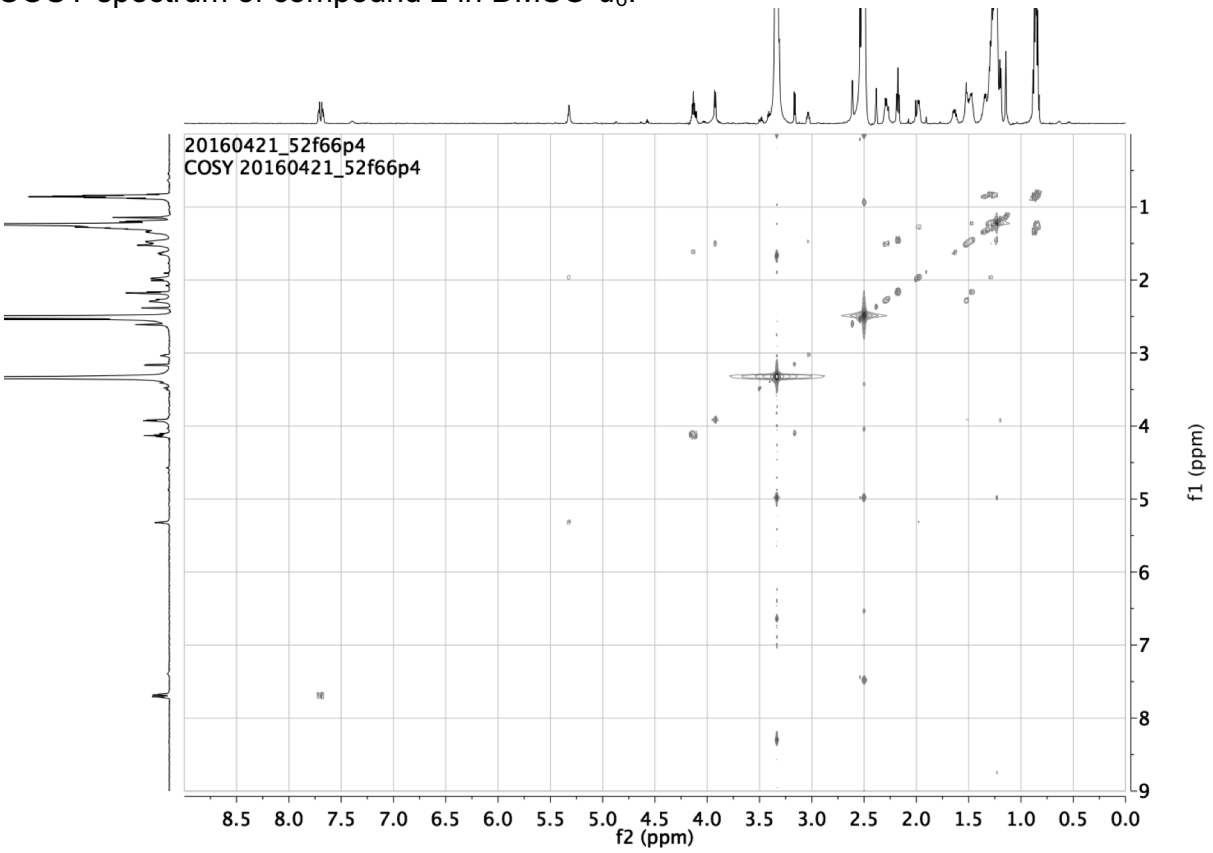


[a] 600MHz for ^{1}H, [b] overlapped signal, [c] Carbon chemical shift from HSQC and HMBC correlations

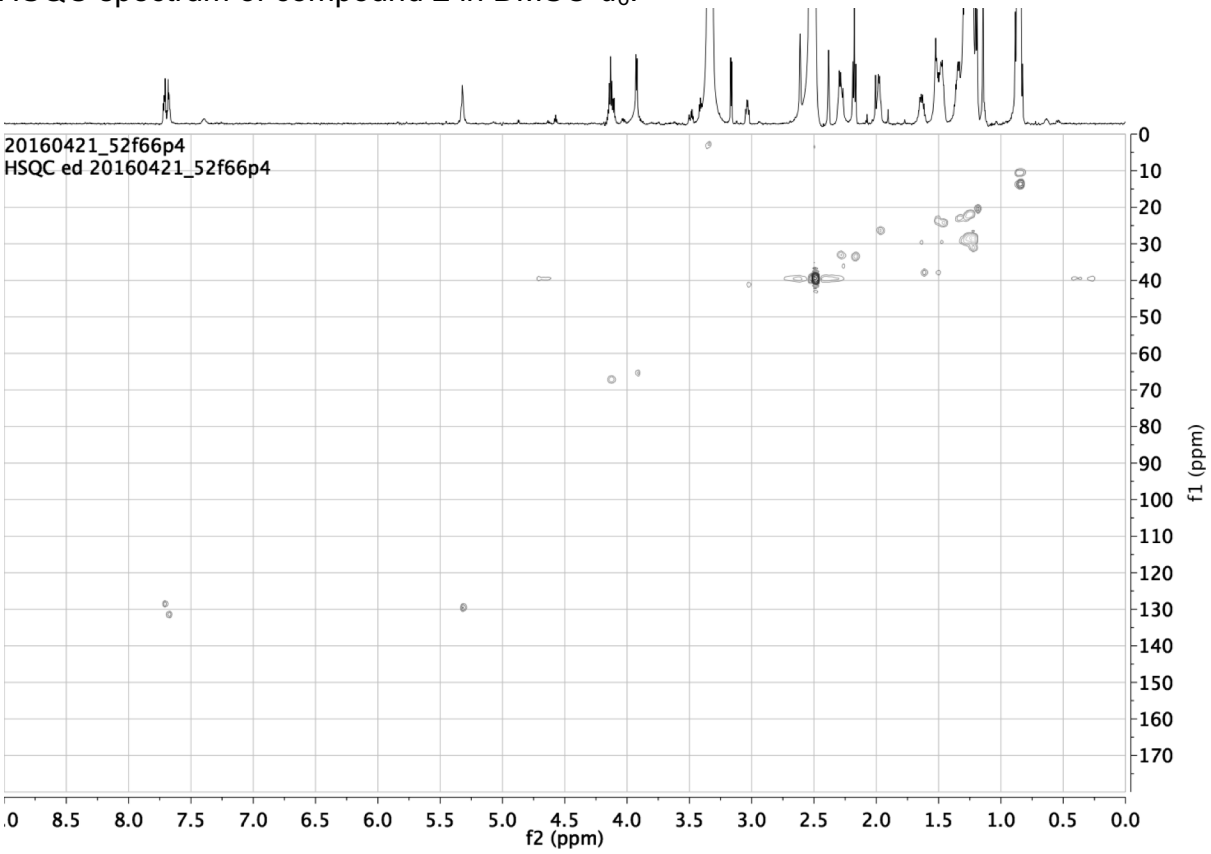
^{1}H NMR spectrum of compound **2** in DMSO-d_6



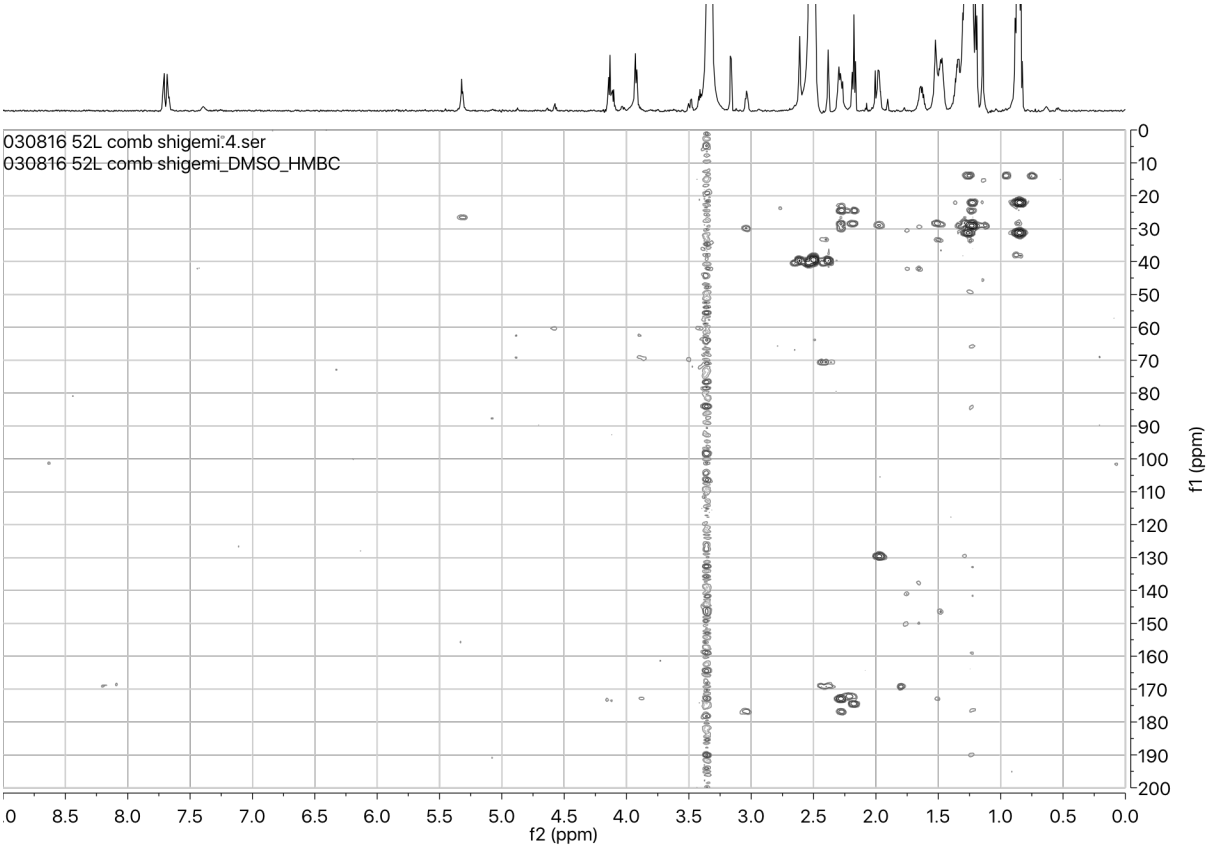
COSY spectrum of compound **2** in DMSO-d_6.



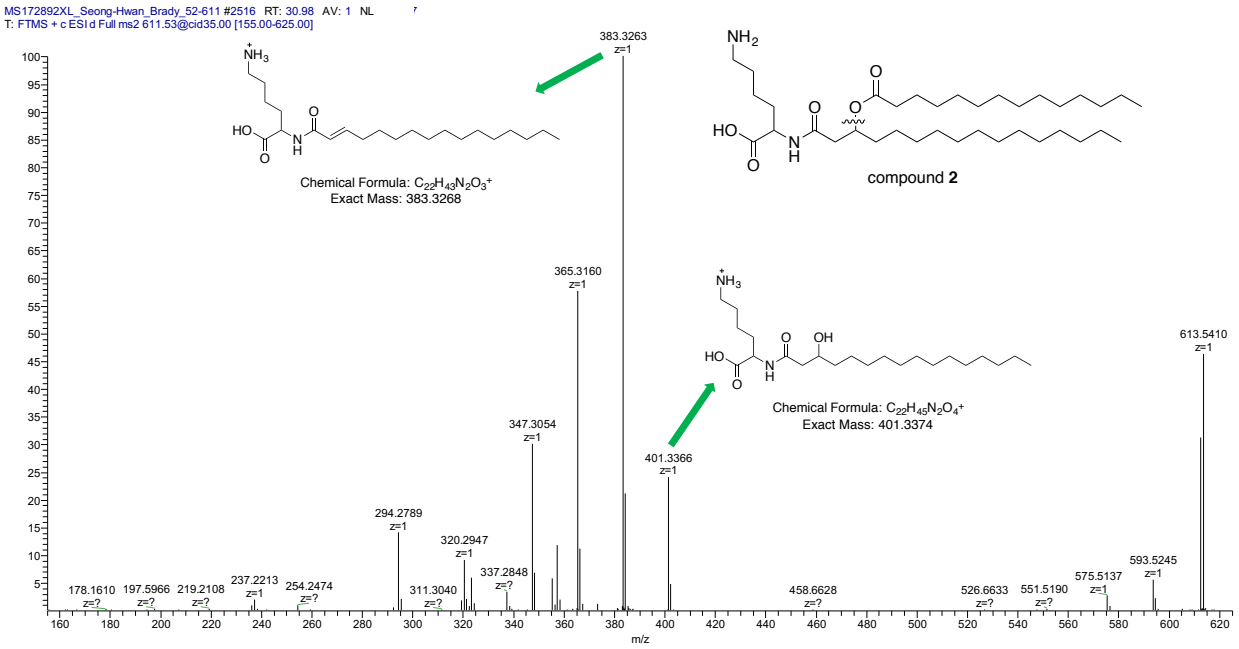
HSQC spectrum of compound **2** in DMSO-d_6.



HMBC spectrum of compound **2** in DMSO-d_6.



HRESI-MS/MS fragmentation for compound **2**

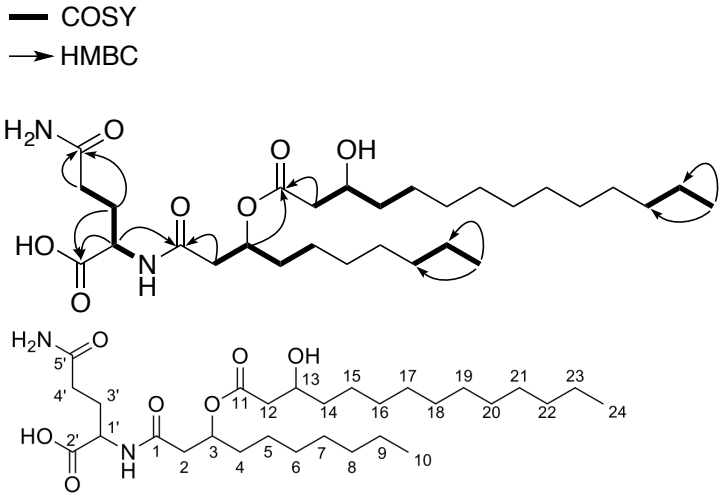


Family 3 – compound 3

The molecular formula predicted by HRMS for compound **3** was $C_{29}H_{54}N_2O_7$ (*m/z*: $[M+H]^+$ Calcd $C_{29}H_{55}N_2O_7$ 543.4009, found 543.4009). The ^1H NMR of compound **3** exhibited two oxygenated methines, a group of highly overlapped aliphatic methylene proton signals (δ_H 1.24~1.21) and two terminal methyl triplets. The ^{13}C NMR of compound **3** exhibited four carbonyl carbons, two oxygen bearing carbons, an aliphatic methine carbon, two methyl carbons, ten distinguished aliphatic methylene carbons and additional overlapping aliphatic methylene carbons (δ_C 28.8-28.6). COSY correlations defined five spin systems. Starting from the 3-carbon-NH COSY spin system, HMBC correlations from H-1′ (δ_H 4.12) and H_2-3′ (δ_H 1.91 and 1.73) to C-2′ (δ_C 173.4) and H-4′ (δ_H 2.10) and H_2-3′ (δ_H 1.91 and 1.73) to C-5′ (δ_C 173.4) defined the structure of glutamine. An HMBC correlation from H-3 (δ_H 5.07) to C-11 (δ_C 170.7) was used to connect the glutamine through an amide bond to a C-3 oxidized fatty acid. HMBC correlations from H-1′ (δ_H 4.12) and H_2-12 to C-11 allowed us to connect a second fatty acid to this substructure through an ester bond. The exact nature of each fatty acid was defined by HRESI-MS/MS fragmentation analysis. Based on this analysis the *N*-acyl fatty acid chain was predicted to contain 10 carbons and the acyl fatty chain was predicted to contain 14 carbons. Thus, the structure of **3** was determined to be 3-[(3-OH-myristoyl)oxy]decanoyl glutamine.

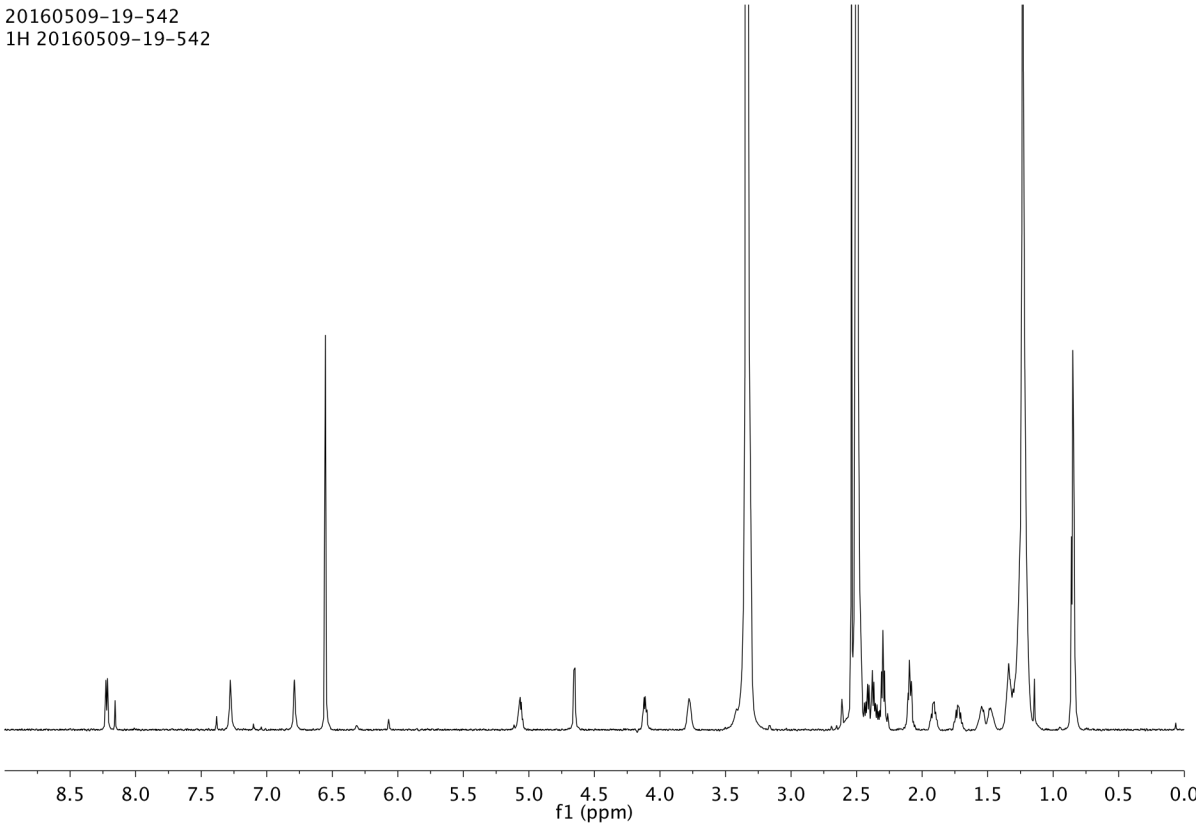
NMR Table Compound 3

Position	Compound 3 [a]	
	δ_C, type	δ_H, mult (*J* in Hz)
1	169.1, C	
2	39.8, CH_2	2.41 2.36, dd (7.5, 7.5)
3	70.7, CH	5.07, m
4	33.2, CH_2	1.54, m[b] 1.48, m[b]
5	28.6, CH_2	1.23, m[b]
6	28.8, CH_2	1.22, m[b]
7	28.9, CH_2	1.22, m[b]
8	31.3, CH_2	1.21, m[b]
9	22.1, CH_2	1.23, m[b]
10	14.0, CH_3	0.85, t (7.0)
11	170.7, C	
12	42.8, CH_2	2.30
13	67.2, CH	3.78, m[b]
14	36.7, CH_2	1.34, t (7.0) 1.30
15	28.6, CH_2	1.24, m
16	28.6, CH_2	1.24, m
17	28.6, CH_2	1.24, m
18	28.6, CH_2	1.24, m
19	28.6, CH_2	1.24, m
20	28.6, CH_2	1.24, m
21	28.6, CH_2	1.24, m
22	31.3, CH_2	1.21, m
23	22.1, CH_2	1.23, m
24	14.0, CH_3	0.85, t (7.0)
NH		8.03, d (7.0)
1′	51.6, CH	4.12, dq (7.5, 7.0)
2′	173.4, C	
3′	26.8, CH_2	1.91, d (7.5) 1.73,
4′	31.3, CH_2	2.10, m
5′	173.4, C	

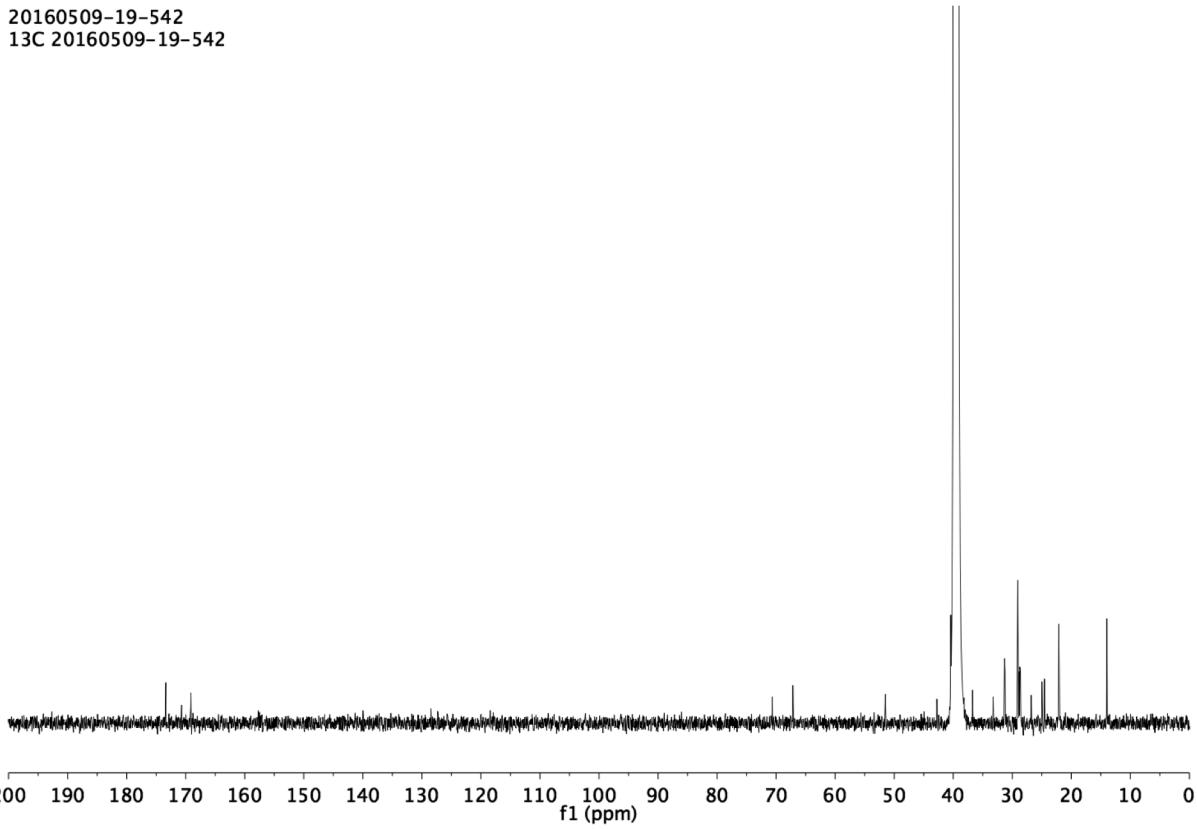


[a]600 MHz for ^1H and 150 MHz for ^{13}C, [b]overlapped signal

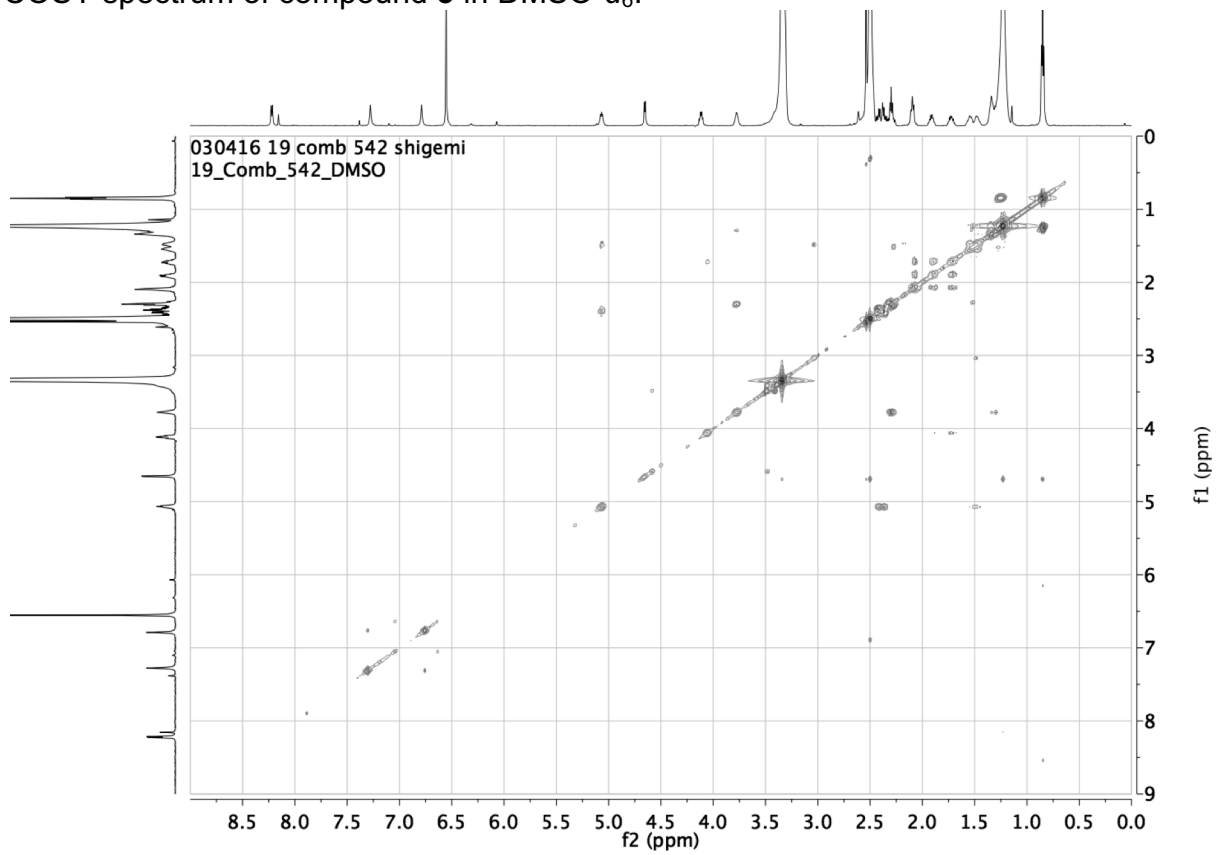
^{1}H NMR spectrum of compound **3** in DMSO-d_6.



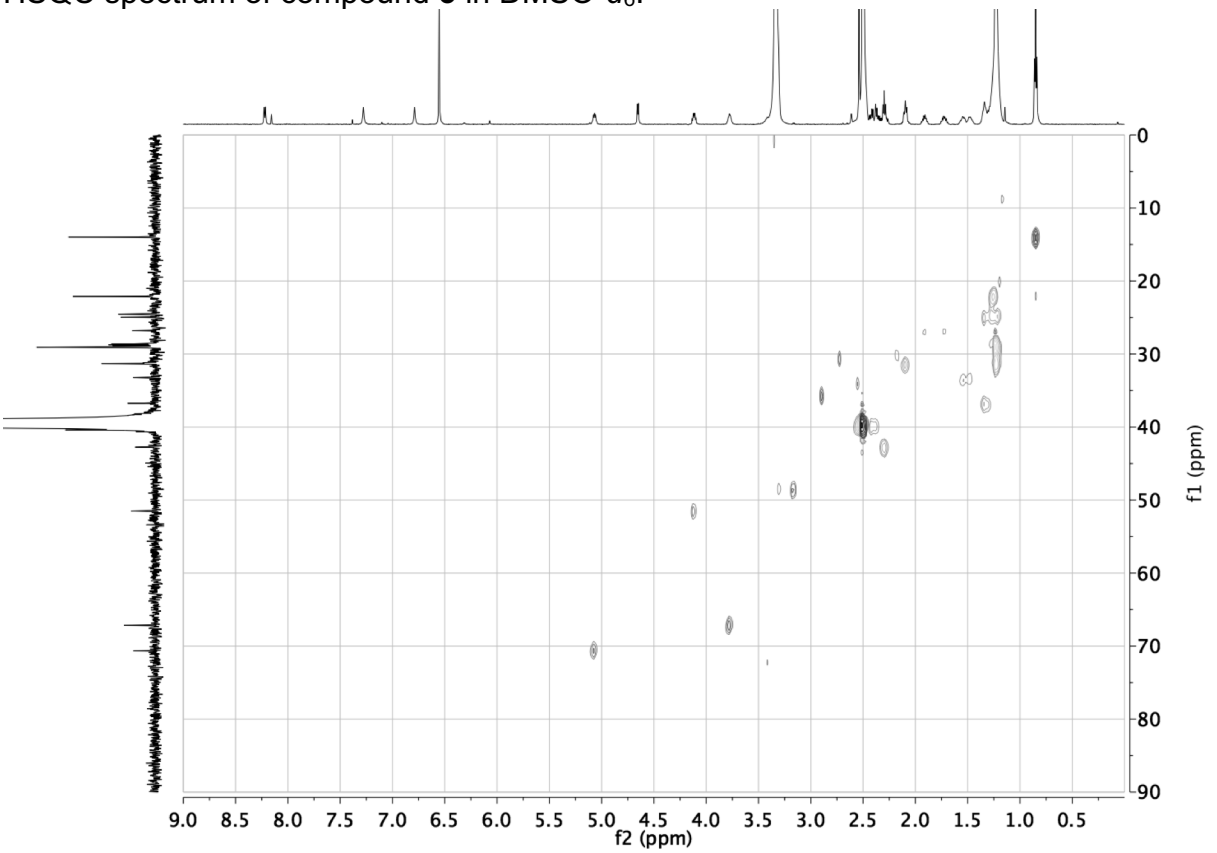
^{13}C NMR spectrum of compound **3** in DMSO-d_6.



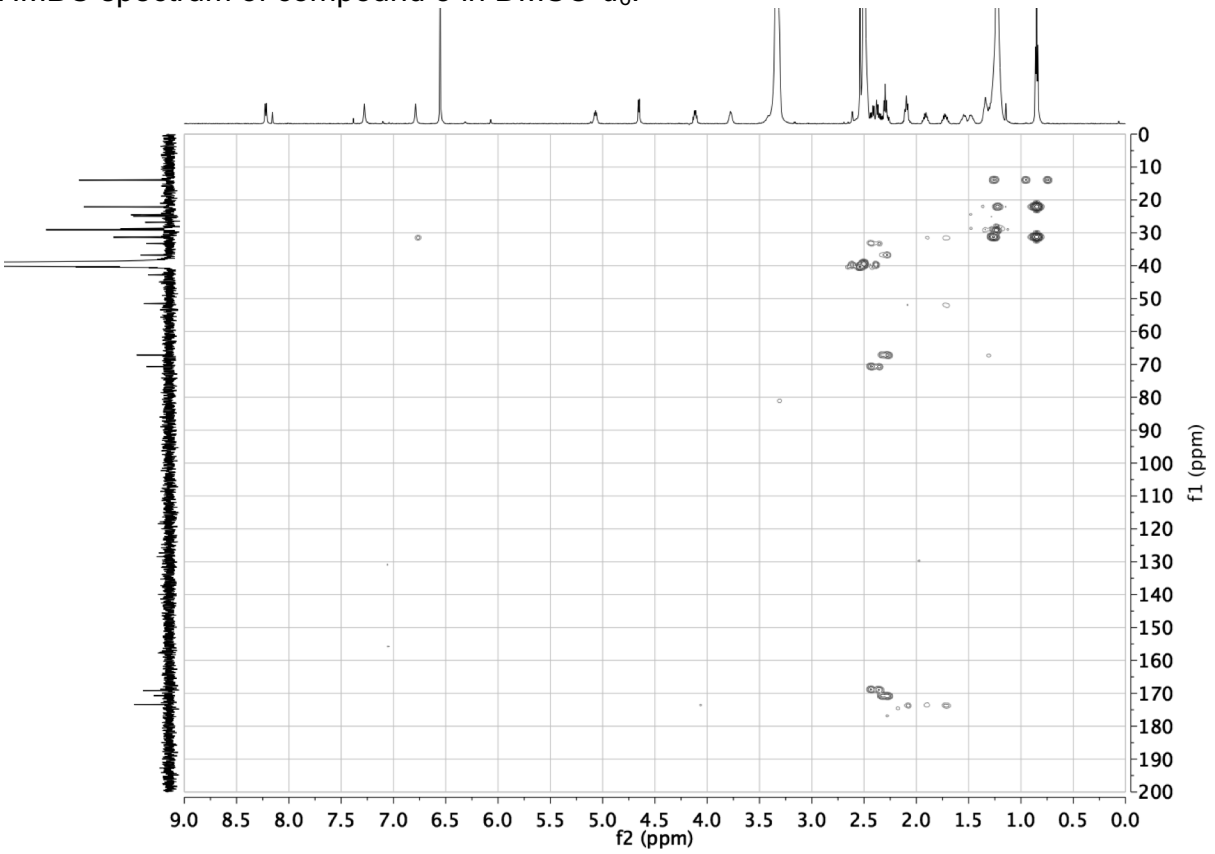
COSY spectrum of compound **3** in DMSO-d_6.



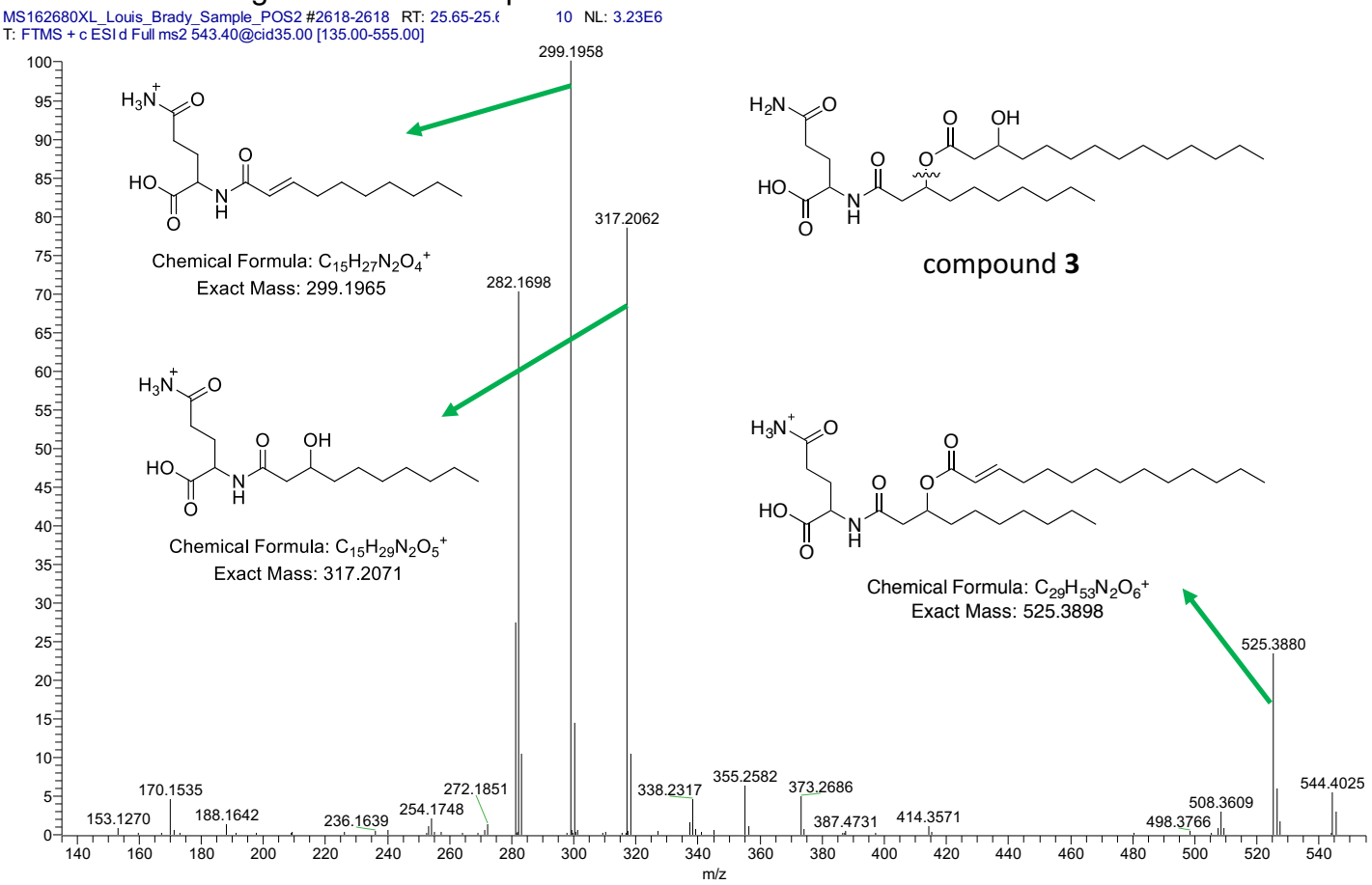
HSQC spectrum of compound **3** in DMSO-d_6.



HMBC spectrum of compound **3** in DMSO-d_6.



HRESI-MS/MS fragmentation of compound **3**

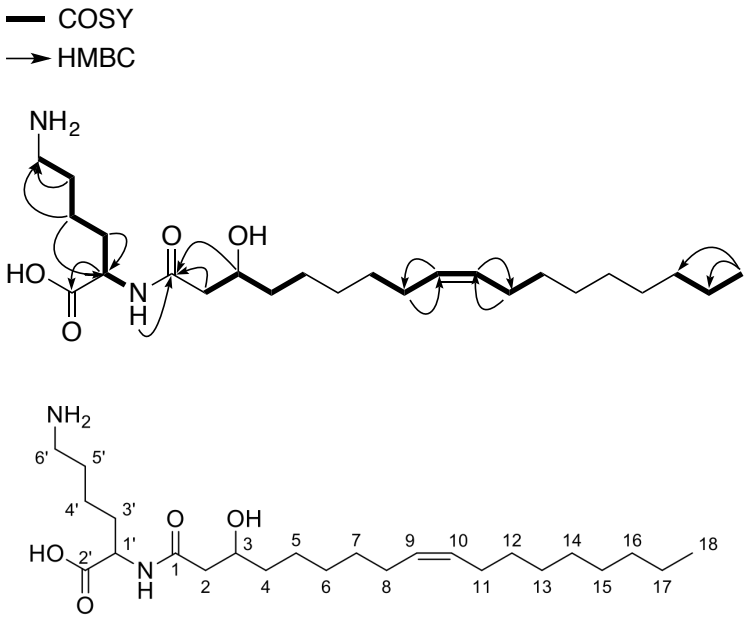


Family 4 – Compound 4a

The major metabolite in family 4 (**4a**) was predicted by HRMS to have the following molecular formula: $C_{24}H_{46}N_2O_4$ (*m/z*: $[M+H]^+$ Calcd $C_{24}H_{47}N_2O_4$ 427.3536, found 427.3531). Through analysis of ^1H and edited HSQC spectra two olefinic protons, an oxymethine proton, a deshielded aliphathic methine proton, a terminal methyl triplet proton and a group of overlapping aliphatic methylene (δ_H 1.24~1.22) were revealed. From the COSY spectrum four spin systems were established. The 5 carbon spin system with COSY correlations from H-1′ to H_2-6′ along with empirical carbon and proton chemical shift data and HMBC correlations from H-1′ (δ_H 3.80) to C-2′ (δ_C 173.6) led to the construction of lysine. HMBC correlations from H_2-2 to C-1 and from NH to C-1 of HMBC indicated that the lysine as connected to a C-3 hydroxylated fatty acid moiety through an amide bond. Based on the predicted molecular formula for compound **4a** the fatty acid side chain was determined to be [C18:1]. The position of the double bond is predicted based on the position that is most frequently seen in *E. coli* lipids and has been seen in other *N*-acyl amino acids heterologously produced in *E. coli*.[1] Thus, the structure of **4a** was determined to be *N*-3-OH-oleoyl lysine.

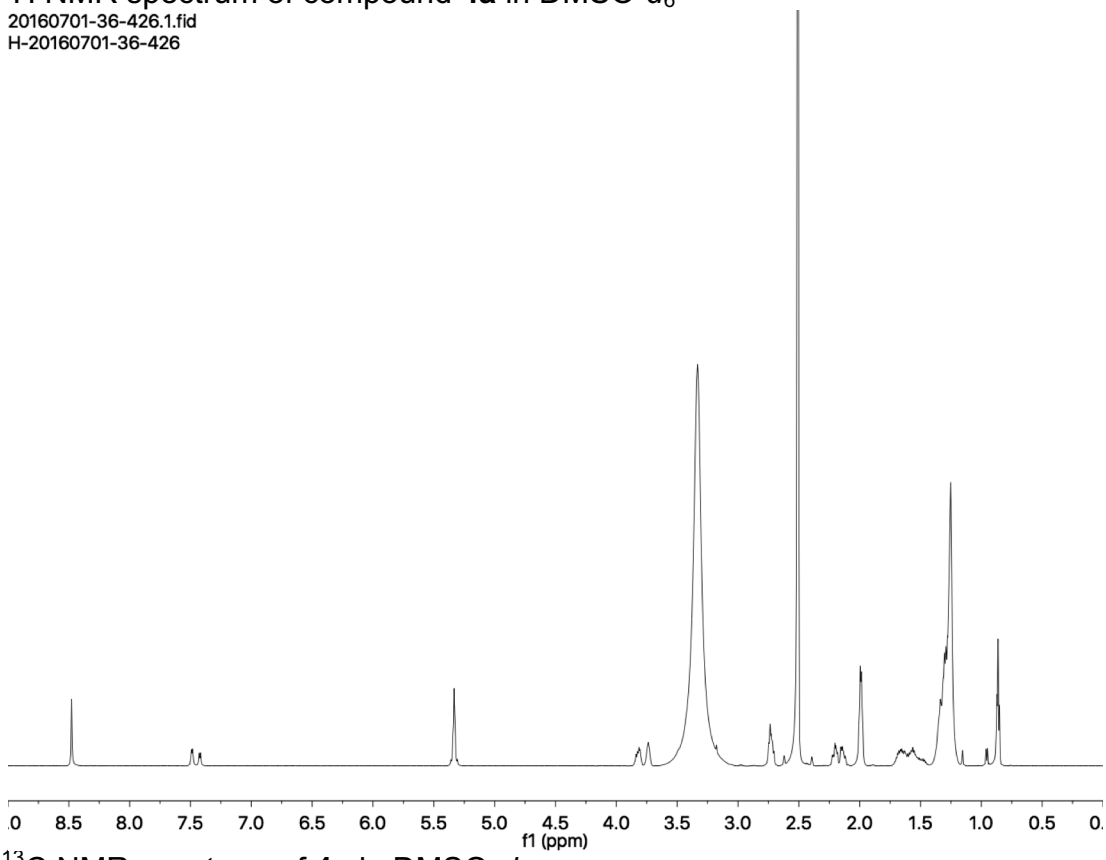
NMR Table Compound **4a**

Position	Compound 4a	
	δ_C,type	δ_H, mult (*J* in Hz)
1	170.0, C	
2	43.8, CH_2	2.16, m
3	67.6, CH	3.73, m
4	37.0, CH_2	1.31, m
5	28.7, CH_2	1.22, m[b]
6	28.8, CH_2	1.22, m[b]
7	29.0, CH_2	1.22, m[b]
8	26.6, CH_2	1.98, m
9	129.6, CH_2	5.33, m
10	129.6, CH_2	5.33, m
11	26.6, CH_2	1.98, m
12	29.0, CH_2	1.22, m[b]
13	29.0, CH_2	1.22, m[b]
14	29.0, CH_2	1.22, m[b]
15	29.0, CH_2	1.22, m[b]
16	31.1, CH_2	1.24, m[b]
17	22.0, CH_2	1.25, m
18	13.9, CH_3	0.85, t (7.0)
NH		7.45, d (7.0)
1′	53.3, CH	3.80, m
2′	173.6, C	
3′	31.4, CH_2	1.62, 1.54, m
4′	21.8, CH_2	1.34, 1.25, m
5′	27.5, CH_2	1.65, 1.56, m
6′	38.8, CH_2	2.73, 2.71, m

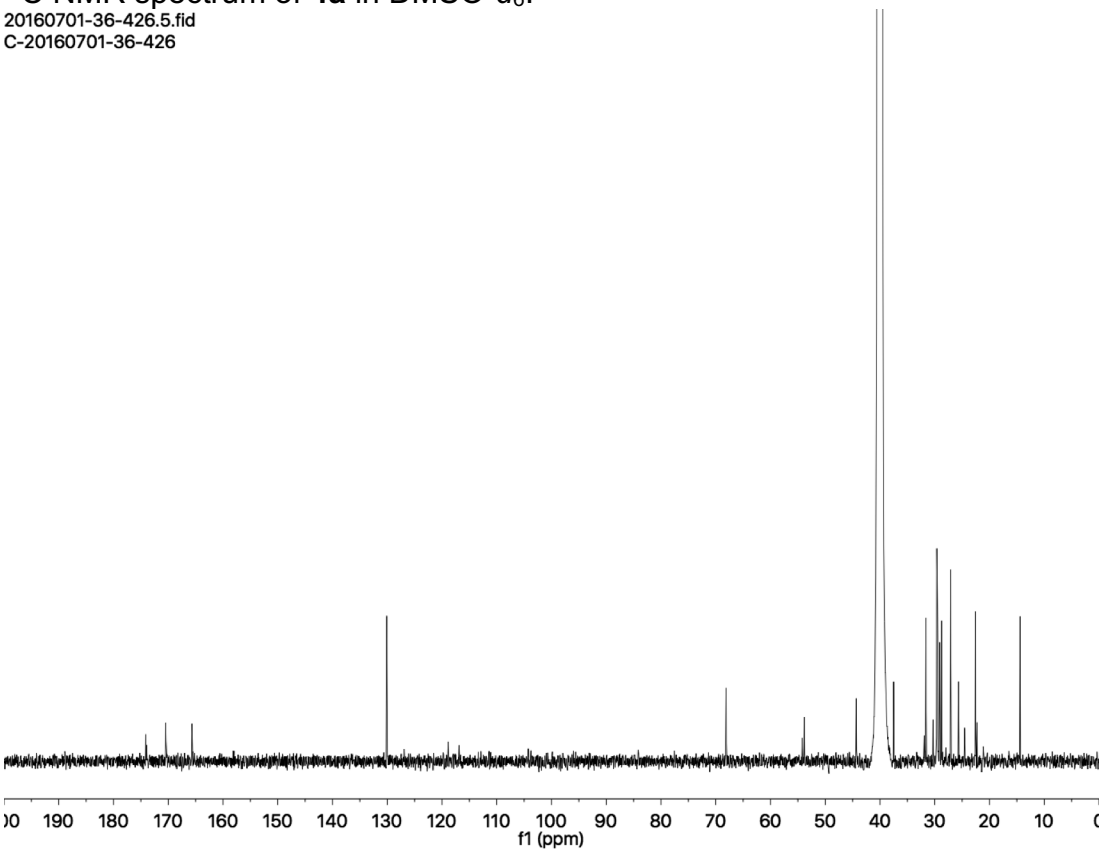


[a]600 MHz for ^1H and 150 MHz for ^{13}C, [b]overlapped signal

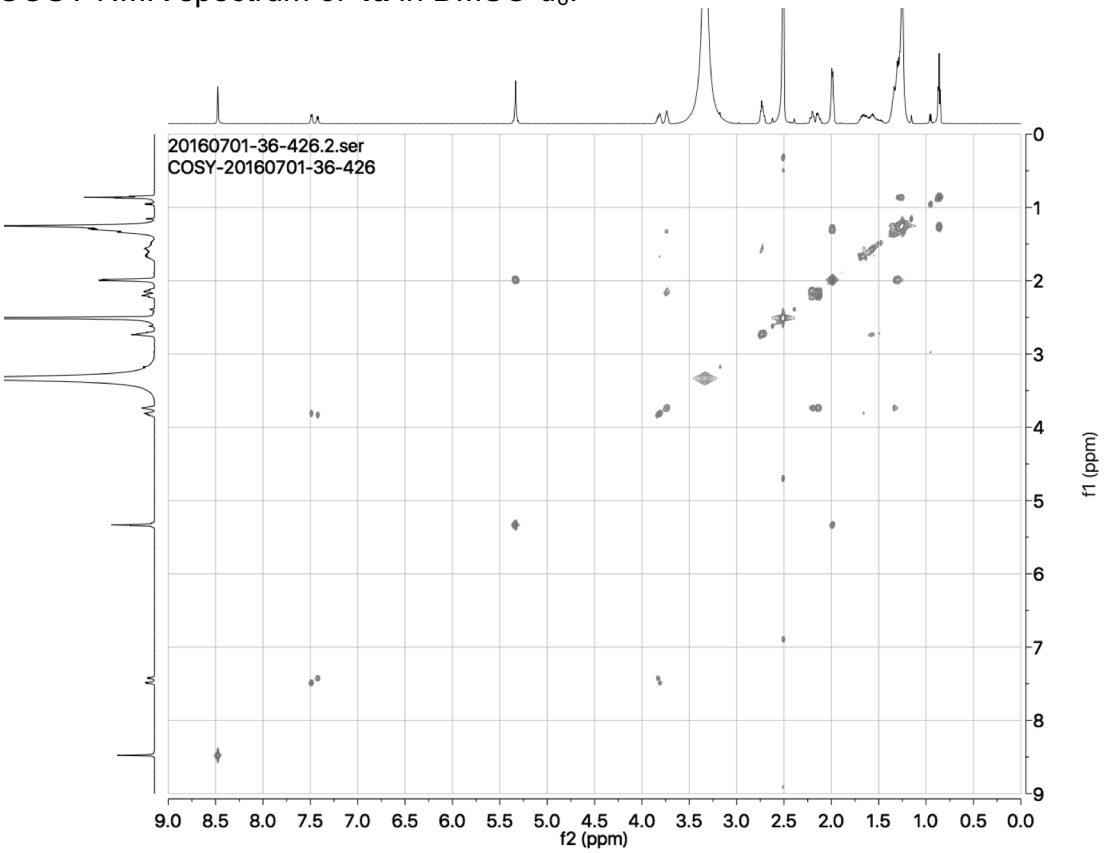
^{1}H NMR spectrum of compound **4a** in DMSO-d_6



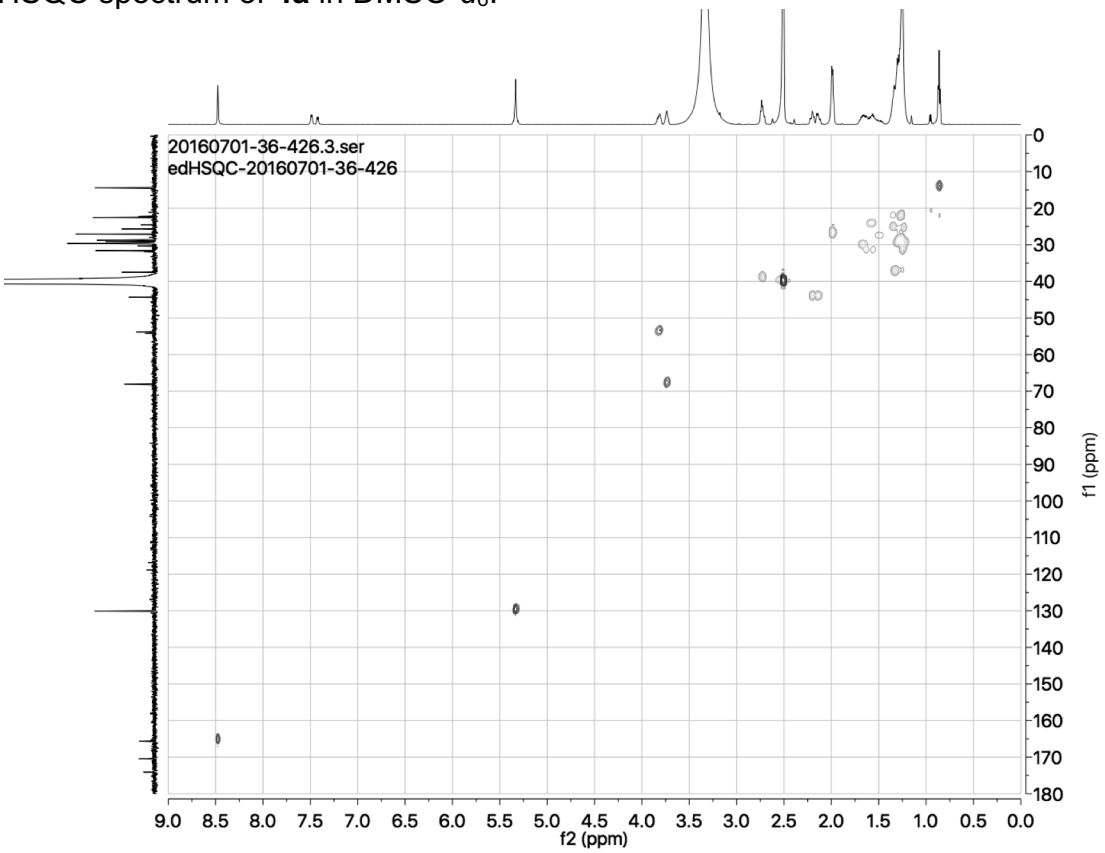
^{13}C NMR spectrum of **4a** in DMSO-d_6.



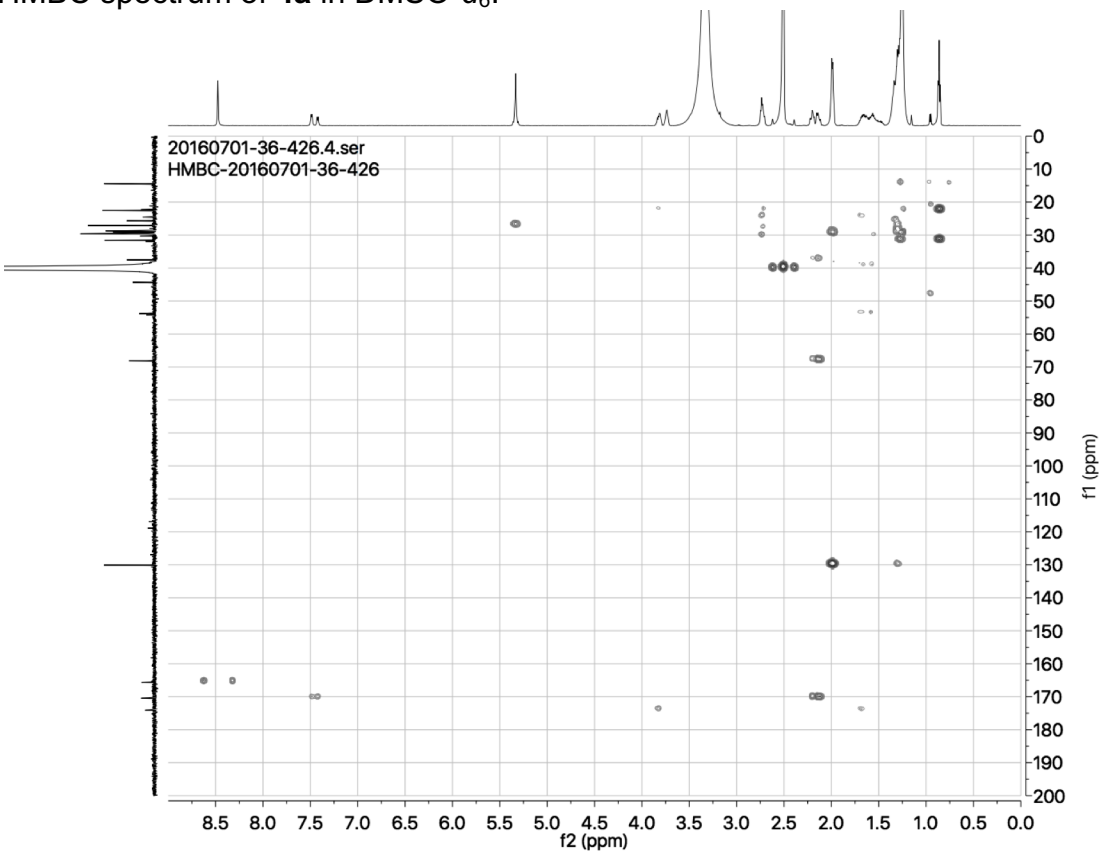
COSY NMR spectrum of **4a** in DMSO-d_6.



HSQC spectrum of **4a** in DMSO-d_6.



1 HMBC spectrum of **4a** in DMSO-d_6.



2

3

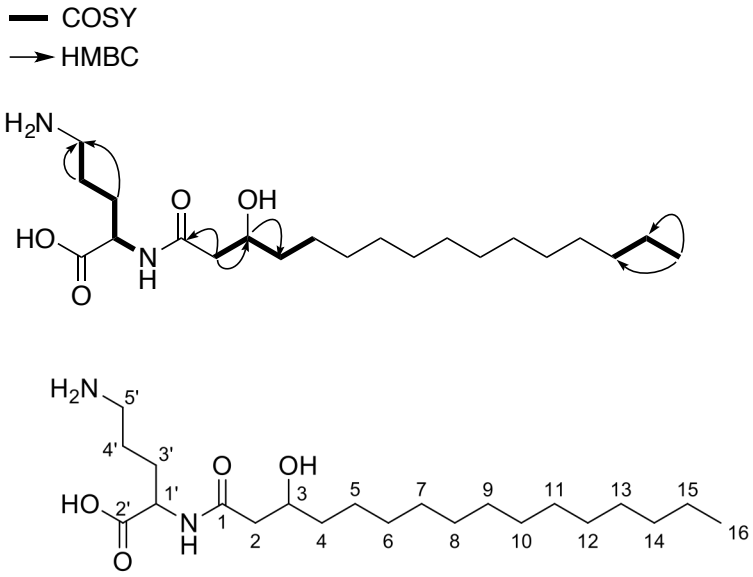
4

Family 4 – compound 4b

The second major *N*-acyl amide from family 4 by weight had an HRMS predicted molecular formula of $C_{21}H_{42}N_2O_4$ (compound **4b**: *m/z*: $[M+H]^+$ Calcd $C_{21}H_{43}N_2O_4$ 387.3223, found 387.3226). NMR spectra for **4b** were nearly identical to those collected for **4a** with the exception of 1) the disappearance of the olefinic protons in the ^1H NMR and 2) the replacement of the 5-carbon-NH COSY spin system with a 4-carbon-NH COSY spin system. Based on these differences and the HRMS predicted molecular formula for **4b** the structure of **4b** was determined to be *N*-3-OH-palmitoyl ornithine.

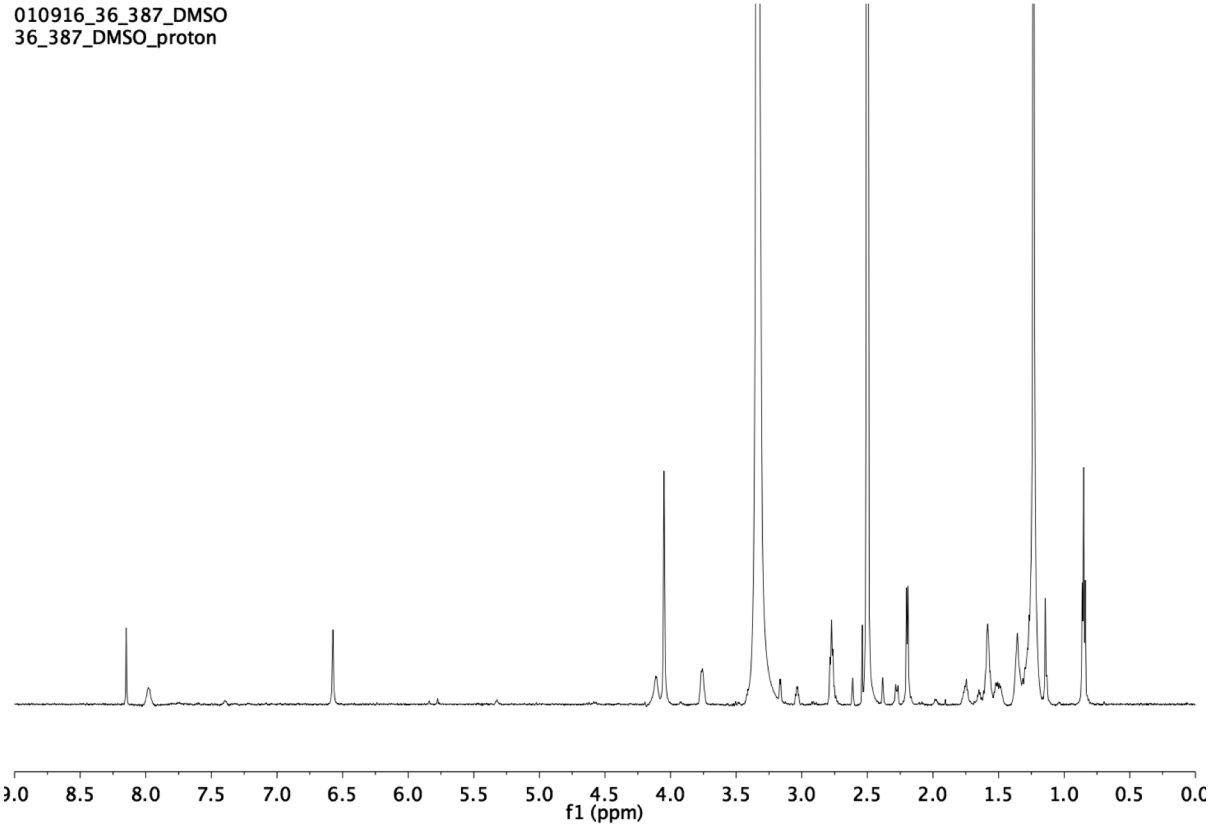
NMR Table Compound 4b

Position	Compound 4b [a,c]	
	δ_C, type	δ_H, mult (*J* in Hz)
1	170.8, C	
2	43.1, CH_2	2.19, dd (7.5, 7.5)
3	67.4, CH	3.76, m
4	36.5, CH_2	1.36, m[b] 1.29, m[b]
5	28.7, CH_2	1.22, m[b]
6	28.8, CH_2	1.22, m[b]
7	28.9, CH_2	1.22, m[b]
8	29.0, CH_2	1.22, m[b]
9	29.0, CH_2	1.22, m[b]
10	29.0, CH_2	1.22, m[b]
11	29.1, CH_2	1.22, m[b]
12	29.0, CH_2	1.21, m[b]
13	29.1, CH_2	1.24, m[b]
14	31.2, CH_2	1.22, t (7.0)
15	22.1, CH_2	1.26, m
16	14.1, CH_3	0.85, t (7.0)
NH		7.98, d (7.0)
1′	51.6, CH	4.10, dq (7.5, 7.0)
2′	174.3, C	
3′	28.5, CH_2	1.74, d (7.5) 1.59,
4′	23.8, CH_2	1.58, m
5′	38.1, CH_2	2.77, m
6′		

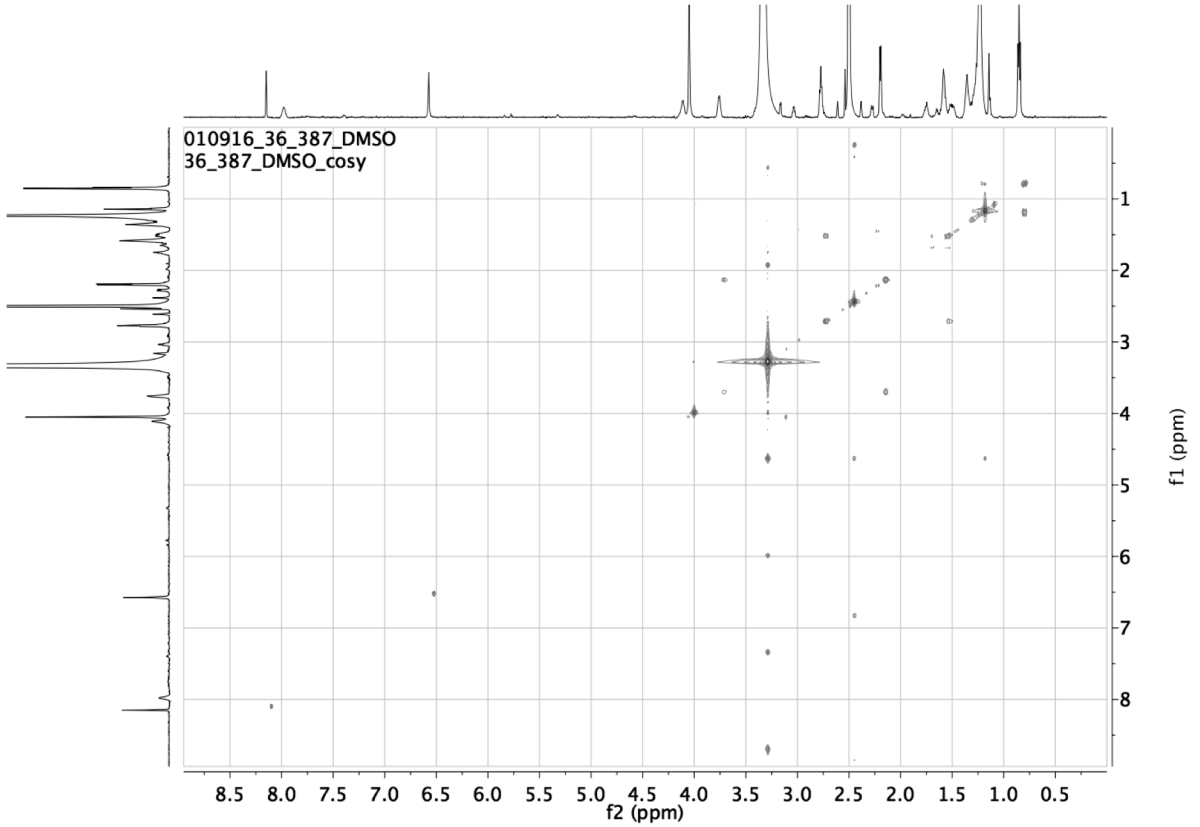


[a]600 MHz for ^1H, [b]overlapped signal, [c]assigned from HMBC correlations

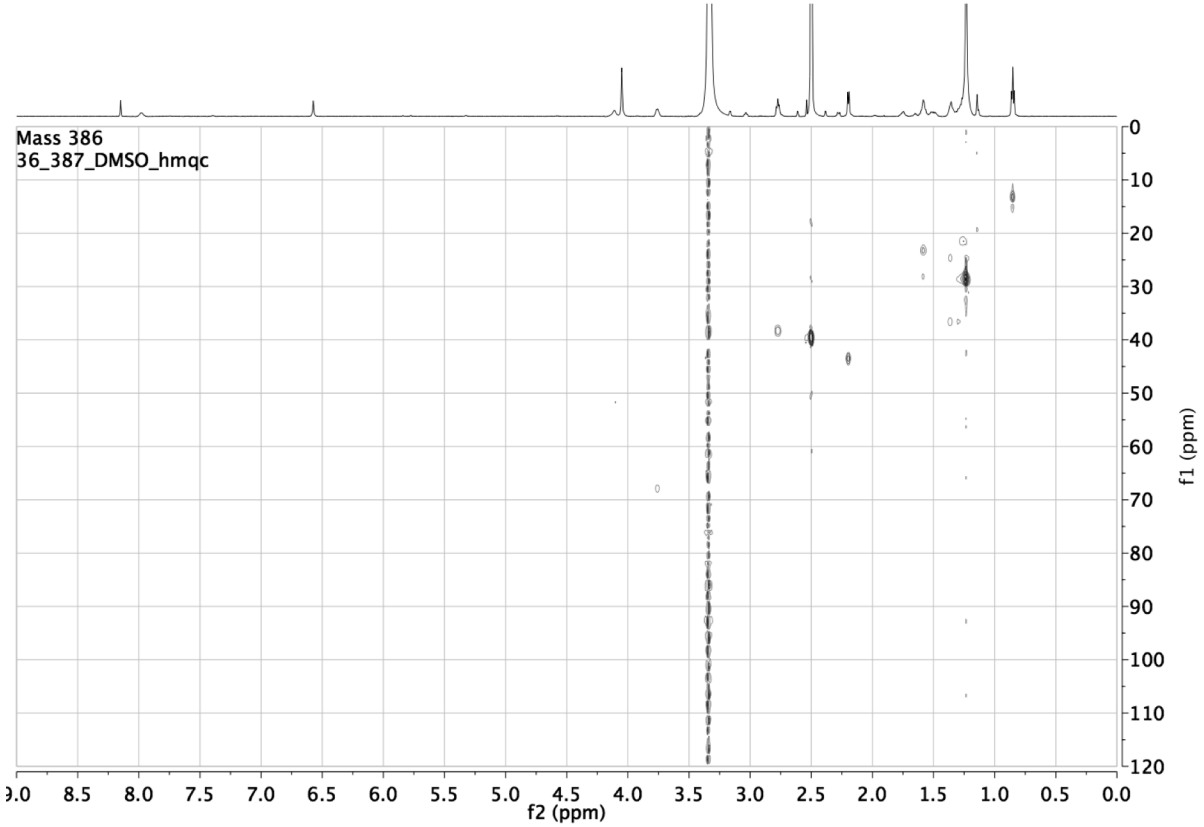
^{1}H NMR spectrum of compound **4b** in DMSO-d_6



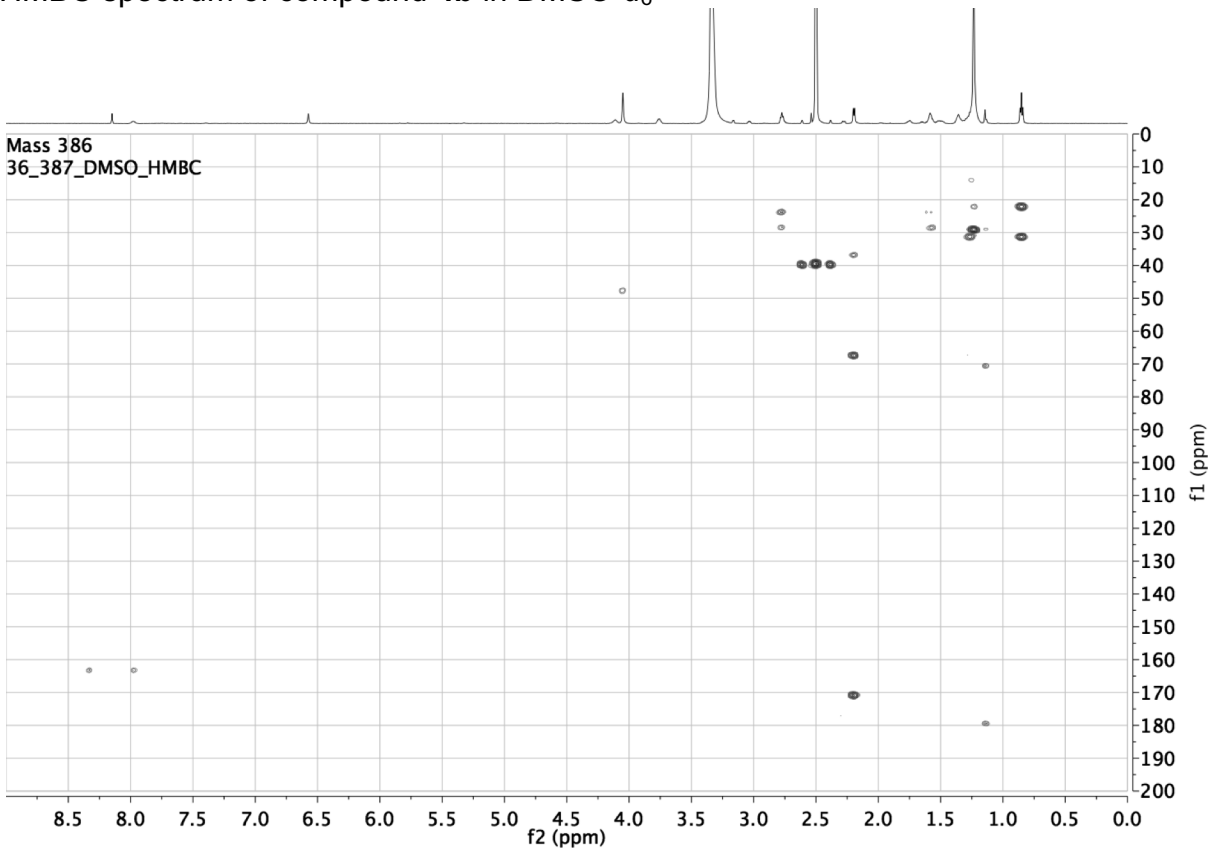
COSY spectrum of compound **4b** in DMSO-d_6.



HMQC spectrum of compound **4b** in DMSO-d_6.



HMBC spectrum of compound **4b** in DMSO-d_6

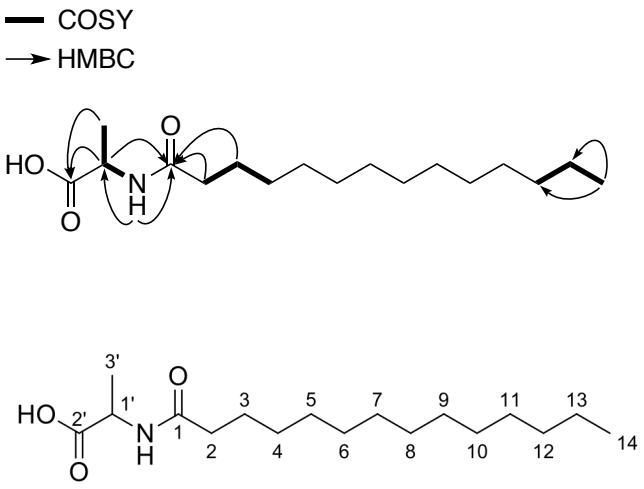


Family 5 – compound 5

The HRMS predicted molecular formula for compound **5** was $C_{17}H_{33}NO_3$ (*m/z*: $[M+Na]^+$ Calcd $C_{17}H_{33}NO_3Na$ 322.2358, found 322.2356). The ^1H NMR spectrum of **5** exhibited chemical shifts characteristic of a saturated fatty acid, a deshielded methine**,** 2 methyls (doublet and triplet) and a deshielded NH. The ^{13}C NMR spectrum of **5** exhibited two clear carbonyl carbons and one predicted N-substituted carbon. Three spin systems were observed in the COSY spectrum. HMBC correlations from the methyl doublet of a 2-carbon COSY spin system to the carbonyl carbon (C-2′) indicated the presence of an alanine in **5.** Empirical ^{13}C NMR chemical shift data and HMBC correlations from H_2-2 to C-1 and the NH proton to C-1 and C-1′ indicated that the fully saturated fatty acid was connected to the alanine through an amide bond. Based on the HRMS predicted molecular formula for **5** the length of a fatty acid was determined to be C14. Thus, the structure of **5** was determined to be *N*-myristoyl alanine.

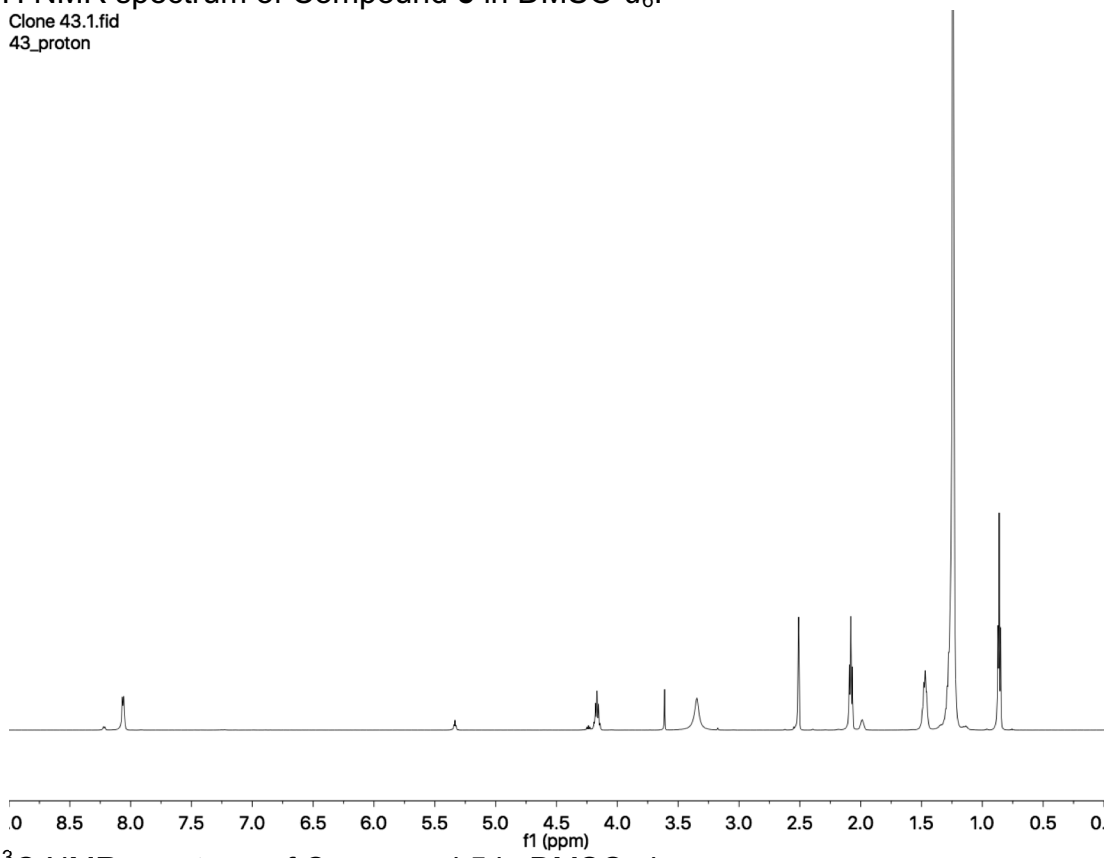
NMR Table for Compound **5**

Position	Compound 5 [a]	
	δ_C, type	δ_H, mult (*J* in Hz)
1	172.0, C	
2	35.0, CH_2	2.07, dd (7.5, 7.5)
3	25.2, CH_2	1.45, m
4	28.6, CH_2	1.22, m[b]
5	28.7, CH_2	1.22, m[b]
6	28.8, CH_2	1.22, m[b]
7	28.9, CH_2	1.22, m[b]
8	29.0, CH_2	1.22, m[b]
9	29.0, CH_2	1.22, m[b]
10	29.0, CH_2	1.22, m[b]
11	29.1, CH_2	1.22, m[b]
12	31.3, CH_2	1.21, m[b]
13	22.1, CH_2	1.24, m[b]
14	14.0, CH_3	0.85, t (7.0)
NH		8.05, d (7.0)
1′	47.3, CH	4.15, dq (7.5, 7.0)
2′	174.3, C	
3′	17.2, CH_3	1.23, d (7.5)

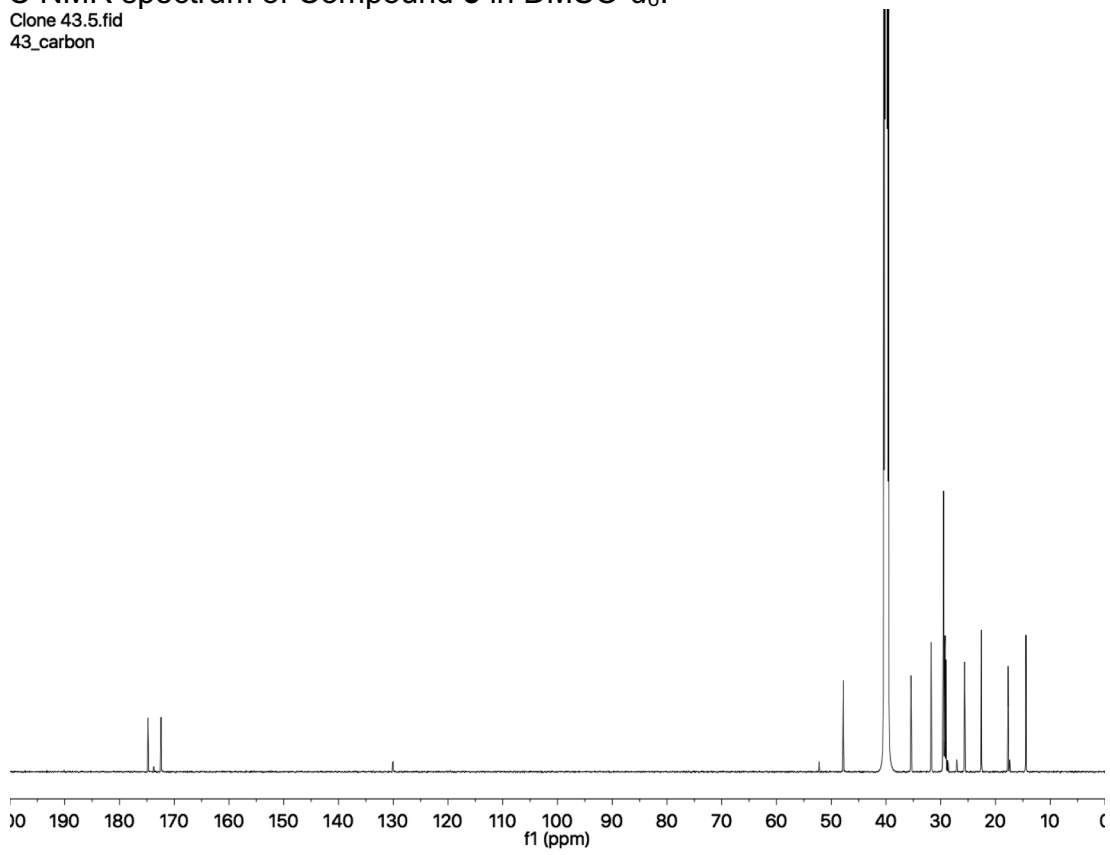


[a]600 MHz for ^1H and 150 MHz for ^{13}C, [b]overlapped signal

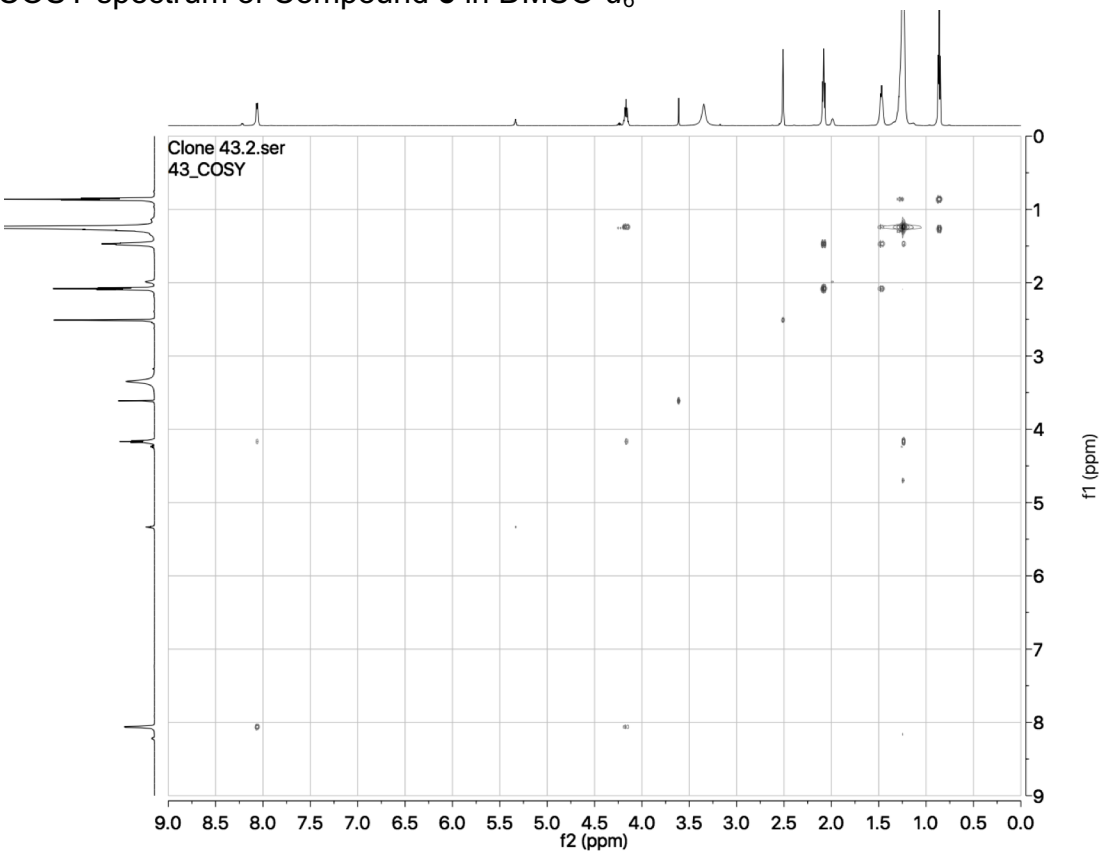
^{1}H NMR spectrum of Compound **5** in DMSO-d_6.



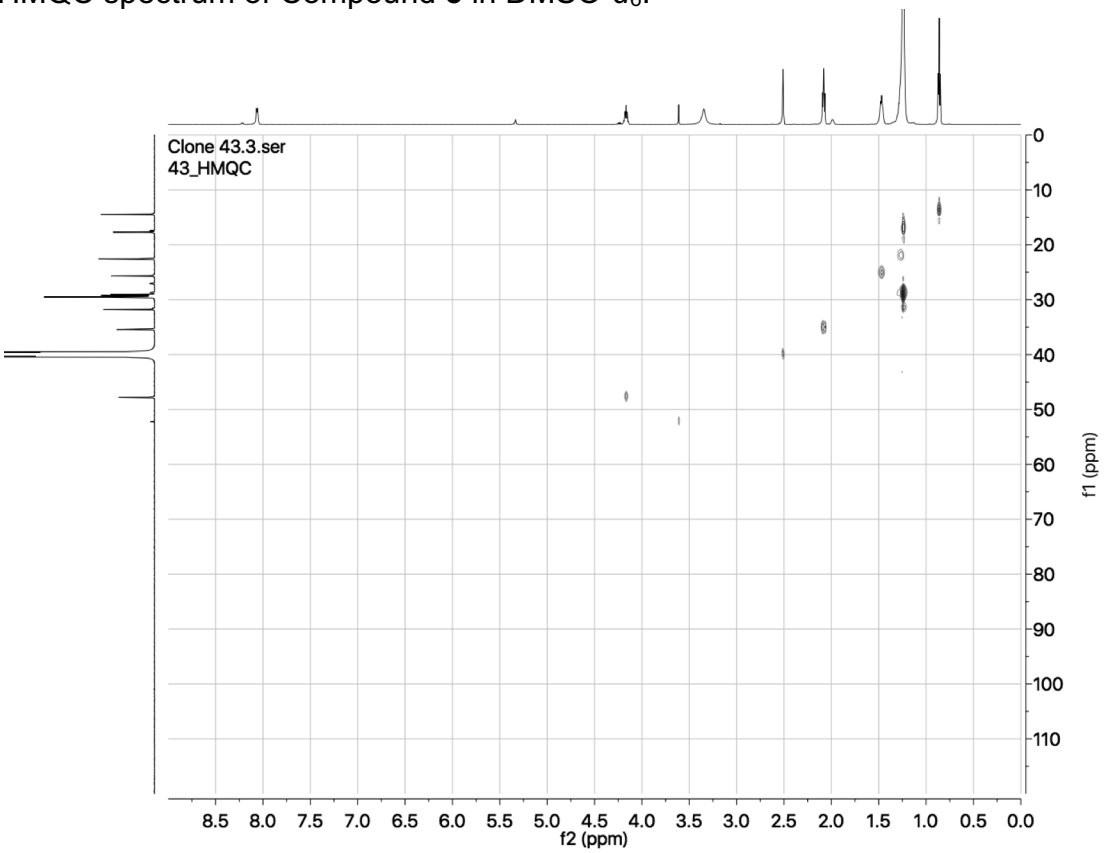
^{13}C NMR spectrum of Compound **5** in DMSO-d_6.



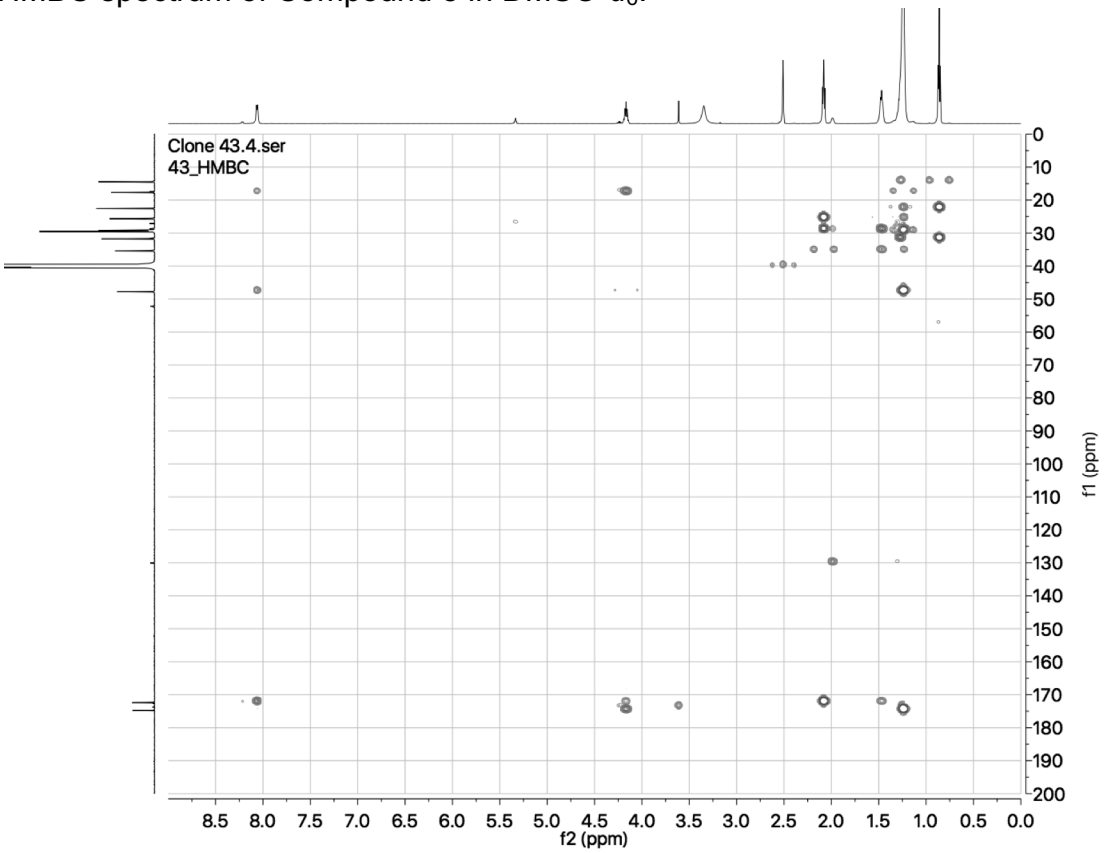
COSY spectrum of Compound **5** in DMSO-d_6



HMQC spectrum of Compound **5** in DMSO-d_6.



HMBC spectrum of Compound **5** in DMSO-d_6.

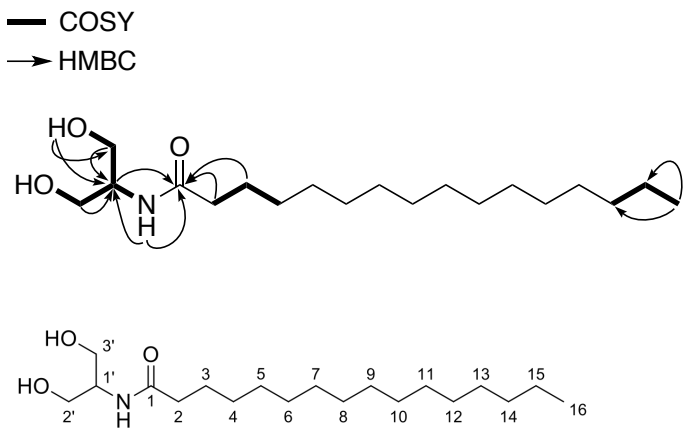


Family 6 – compound 6

The HRMS predicted molecular formula for compound **6** was $C_{19}H_{39}NO_3$ (*m/z*: $[M+H]^+$ Calcd $C_{19}H_{40}NO_3$ 330.3008, found 330.3014). Analysis of the ^1H NMR spectrum indicated the presence of a saturated fatty acid moiety [e.g., overlapping aliphatic methylene proton signals (δ_H 1.23~1.21) and a terminal methyl triplet proton triplet (δ_H 0.85)]. In the ^{13}C and HMQC NMR spectra of **6** we observed one carbonyl carbon, two equivalent oxymethylene carbons (C-2′ and C-3′) and a deshielded aliphatic methine carbon. Analysis of the COSY spectrum identified three spin systems. Extensive COSY and HHMC (see figure below) correlations established a serinol substructure. Based on HMBC correlations from NH to C-1 and C-1′ and weak long-range HMBC correlations from H-1′ to C-1 the serinol was connected to the fatty acid through an amide bond. The length of a fatty acid was determined as C16 based on the predicted molecular formula. Thus, the structure of **6** was determined to be *N*-palmitoyl serinol.

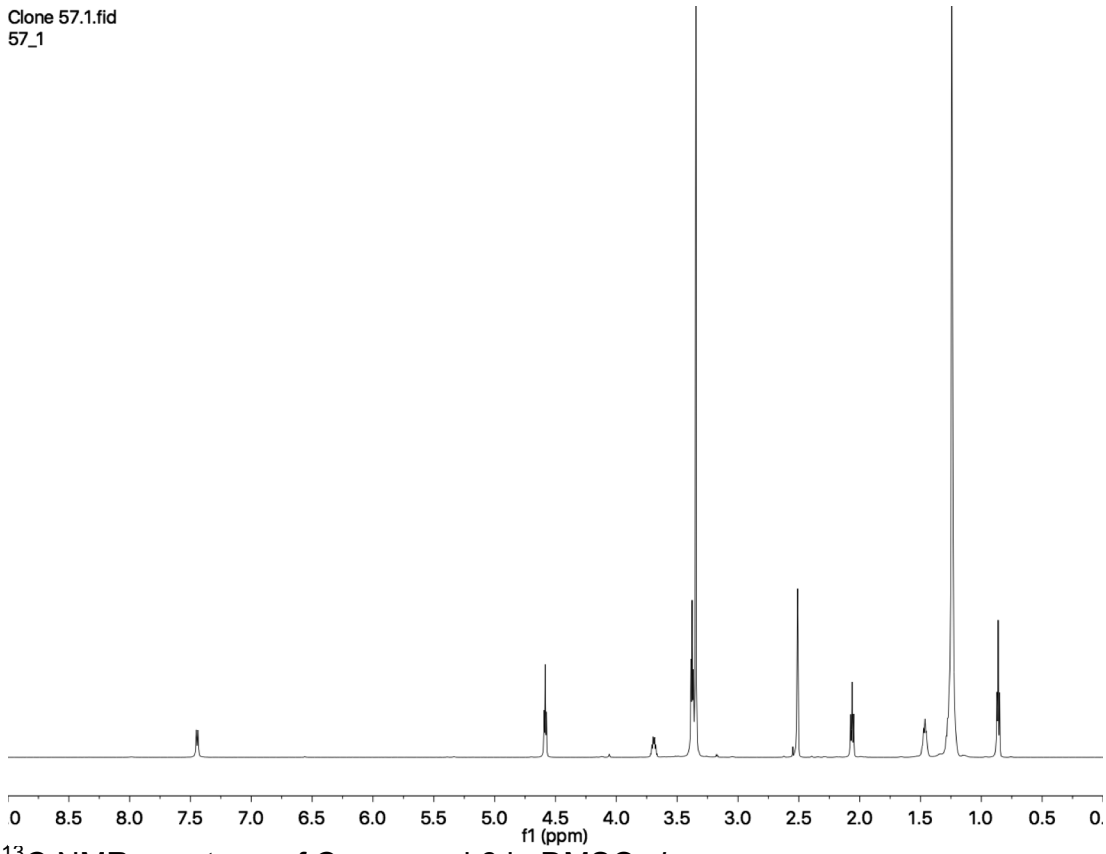
NMR Table Compound **6**

Position	Compound 6 [a]	
	δ_C, type	δ_H, mult (*J* in Hz)
1	172.1, C	
2	35.4, CH_2	2.05, dd (7.5, 7.5)
3	25.3, CH_2	1.45, m
4	28.7, CH_2	1.21, m
5	28.7, CH_2	1.22, m[b]
6	28.8, CH_2	1.22, m[b]
7	29.0, CH_2	1.22, m[b]
8	29.0, CH_2	1.22, m[b]
9	29.0, CH_2	1.22, m[b]
10	29.0, CH_2	1.22, m[b]
11	29.0, CH_2	1.22, m[b]
12	29.0, CH_2	1.22, m[b]
13	29.0, CH_2	1.22, m[b]
14	31.3, CH_2	1.22, m[b]
15	22.1, CH_2	1.27, m[b]
16	14.0, CH_3	0.85, t (7.0)
NH		7.44, d (8.0)
1′	52.7, CH	3.67, ddd (8.0, 5.5, 5.5)
2′	60.2, CH_2	3.37, t (5.5)
2′-OH		4.57, t (5.5)
3′	60.2, CH_2	3.37, t (5.5)
3′-OH		4.57, t (5.5)

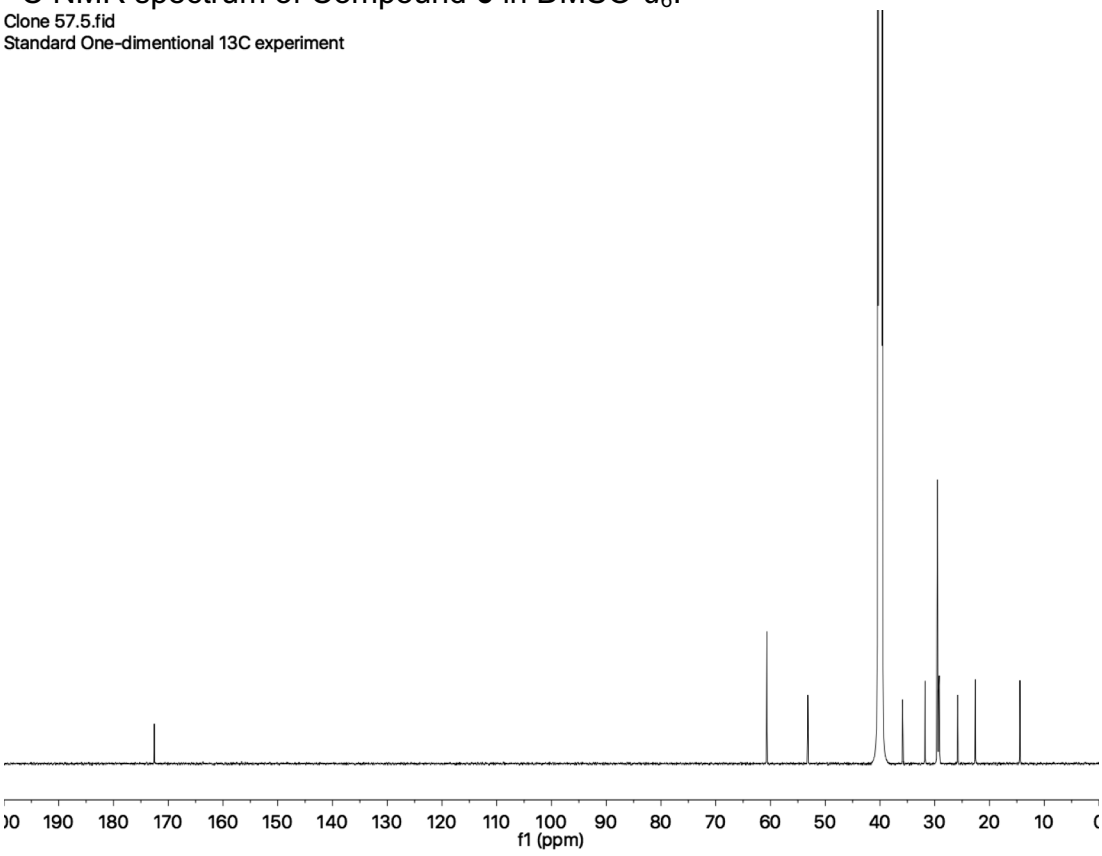


[a]600 MHz for ^1H and 150 MHz for ^{13}C, [b]overlapped signal

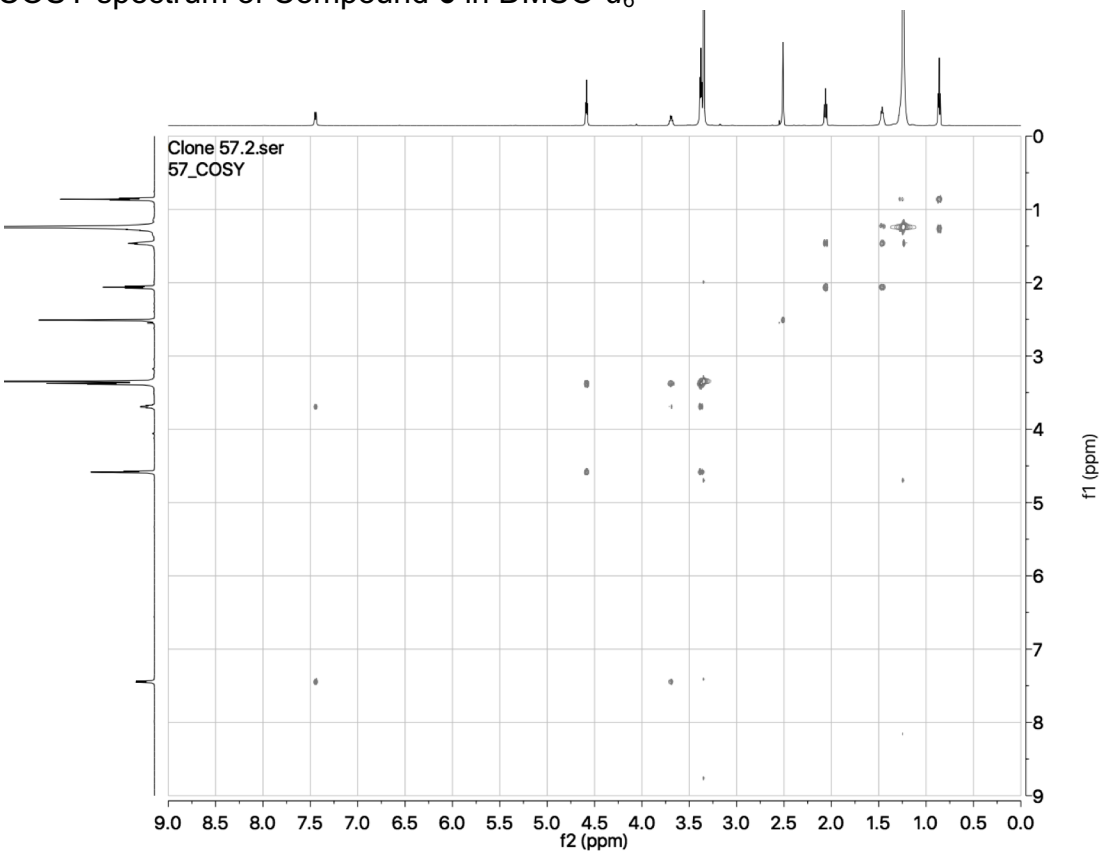
^{1}H NMR spectrum of Compound **6** in DMSO-d_6



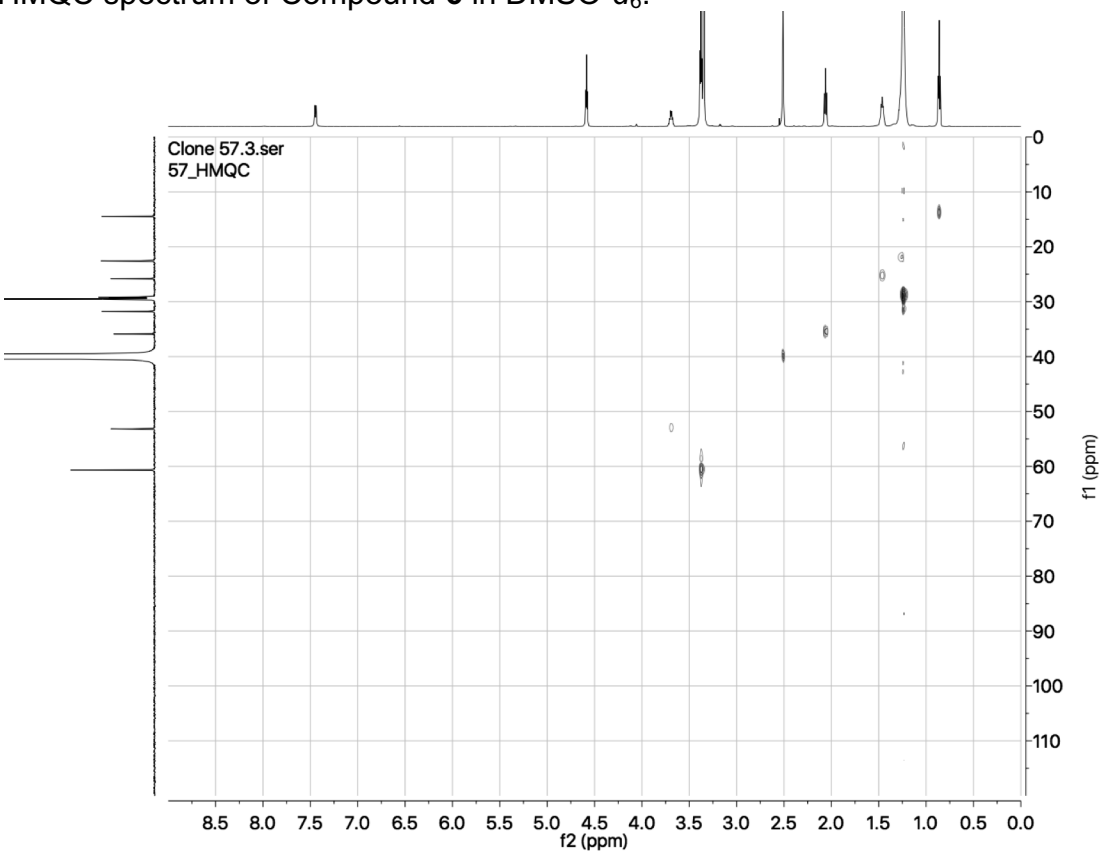
^{13}C NMR spectrum of Compound **6** in DMSO-d_6.



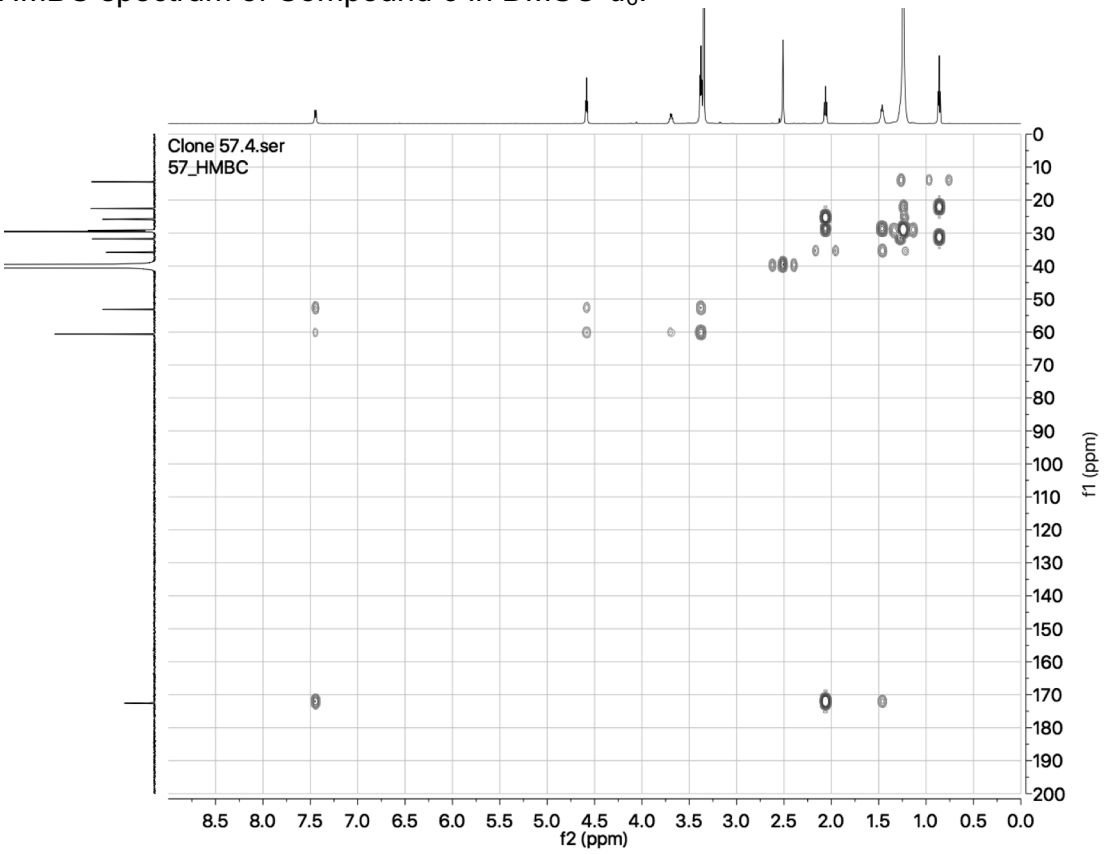
COSY spectrum of Compound **6** in DMSO-d_6



HMQC spectrum of Compound **6** in DMSO-d_6.



1 HMBC spectrum of Compound **6** in DMSO-d_6.



2
3
4

Supplementary Info Tables

Supplementary Info Table 1. hm-NAS Genes Selected for Heterologous Expression. This set included all hm-NAS genes from clades sparsely populated with hm-NAS sequences and representative examples from clades heavily populated with hm-NAS sequences.

Clone Number	EBI Gene	Organism	Gene Size (bp)	Molecule Family
1	EFI7261	Bacteroides oral 274 F0058	1191	No production
2	EHB91285	Alistipes indistinctus YUT 12060	921	1
3	EEK17761	Porphyromonas uenonis	960	No production
5	EEY82825	Bacteroides sp 2_1_33B	987	1
6	EHP49568	Odoribacter laneus YIT 12061	969	No production
7	EHG23013	Alloprevotella rava F0323	1008	1
8	EFA42931	Prevotella bergensis DSM 17361	999	1
9	EFL47029	Prevotella disiens FB035	1005	1
10	EHO75052	Prevotella micans F0438	1005	1
11	ADK95845	Prevotella melaninogenica ATCC 25845	1011	1
12	EFV04460	Prevotella salivae DSM 15606	1017	1
13	EHH01788	Paraprevotella clara YIT 11850	945	1
14	EDY97076	Bacteroides plebius DSM 17135	1002	1
15	CBW20928	Bacteroides fragilis 638R	1026	1
16	EDS14876	Bacteroides stercoris ATCC 43183	1035	1
17	EDO52243	Bacteroides uniformis ATCC 8492	990	1
18	CBK67812	Bacteroides xylanisolvens XB1A	1029	1
19	ACI09609	Klebsiella pneumonia 342	1713	3
21	ABV66681	Acrobacter butzleri RM4018	1716	2
24	EHT12133	Klebsiella oxytoca 10-5246	1731	2
26	EFE54303	Providencia rettgeri DSM 1131	1743	2
27	EFE94777	Serratia odorifera DSM 4582	1734	2
29	EER56350	Neisseria flavescens SK114	768	No production
30	EET45812	Neisseria sicca ATCC 29256	783	4
31	ACS62992	Ralstonia pickettii 12D	846	4
33	BAH33083	Rhodococcus erythropolis PR4	849	No production
35	EFG73978	Mycobacterium parascrofulaceum ATCC BAA 614	870	No production
36	CAW29482	Pseudomonas aeruginosa LESB58	768	4
37	EFH13337	Roseomonas cervicalis ATCC 49957	813	4
38	EGP09383	Bradyrhizobiaceae bacterium SG-6C	1041	No production

39	EEV22085	Enhydrobacter aerosaccus SK60	1011	No production
40	EEY94333	Acinetobacter junii SH205	789	No production
41	EFF83269	Acinetobacter haemolyticus ATCC 19194	789	No production
42	CAP01857	Acinetobacter baumannii SDF	816	4
43	EGP10046	Bradyrhizobiaceae bacterium SG-6C	804	5
50	EFK33376	Chryseobacterium gleum ATCC 35910	1854	No production
51	EEK14630	Capnocytophaga gingivalis ATCC 33624	1815	No production
52	EFS97491	Capnocytophaga ochracea F0287	1848	2
53	CBK85930	Enterobacter cloacae NCTC 9394	1713	2
54	EHM48796	Yokenella regensburgei ATCC 43003	1713	2
55	EEK89350	Bacilus cereus m1550	1596	No production
56	EHL05550	Desulfitobacterium hafniense DP7	1638	6
57	EFV76279	Bacillus sp 2_A_57_CT2	1623	6
58	GL883582	Gemella Haemolysans M341	1576	6

1

Supplementary Info Table 2. Reference Genome Analysis. The phyla Bacteroidetes contained the largest number of reference genomes with hm-NAS genes (86/111 hm-NAS genes, families 1 and 2), while Proteobacterial species contained hm-NAS genes that encode for the greatest number of different *N*-acyl molecule families (families 2, 3, 4, 5). *N*-acyloxyacyl lysine hm-NAS genes (family 2) align to both Proteobacterial and Bacteroidetes species but hm-NAS genes from each phylum are present in distinct branches of the same clade (Fig 1a). Bacteroidetes are one of the most common phyla in the human microbiome, so it is not surprising that hm-NAS genes that map to Bacteroidetes species are the most abundant in patient samples (Fig. 1a). In comparison, hm-NAS genes that encode for *N*-acyl lysine/ornithine (Proteobacteria) and *N*-acyl serinol (Firmicutes) are comparatively rare in reference genomes. These rare genes are however, frequently found in patient samples because their carrier bacterial species are highly prevelant.[44]

PFAM13444 Gene	N-acyl Amide Molecule Family	Body site*	Phyla	Score	E-Value	% Identity	Length	HMP reference genome**
R6A3N1_9BACT/51-156	1	Oral	Bacteroidetes	110	2.00E-22	76.88	199	>ADDV01000044 Prevotella oris C735
R6EH40_9BACT/51-155	1	Oral	Bacteroidetes	281	3.00E-74	72.53	892	>ADDV01000044 Prevotella oris C735
R7PBT6_9BACT/52-156	1	Oral	Bacteroidetes	58.4	6.00E-07	100	31	>ADCT01000041 Prevotella sp. C561
R7NN97_9BACE/51-155	1	Gastrointestinal_tract	Bacteroidetes	1790	0	99.59	981	>AQHY01000032 Bacteroides massiliensis B84634
A0A0C3RD59_9PORP/51-157	1	Gastrointestinal_tract	Bacteroidetes	82.4	4.00E-13	82.8	93	>GG705232 Bacteroides sp. 3_1_33FAA
A6L081_BACV8/51-155	1	Gastrointestinal_tract	Bacteroidetes	1807	0	99.9	981	>ADKO01000098 Bacteroides vulgatus PC510
A6LEV2_PARD8/51-155	1	Gastrointestinal_tract	Bacteroidetes	1762	0	99.28	975	>ACPW01000045 Parabacteroides sp. D13
D4IM11_9BACT/57-158	1	Gastrointestinal_tract	Bacteroidetes	1868	0	100	1011	>ADKO01000098 Bacteroides vulgatus PC510
D5EVS3_PRER2/52-157	1	Gastrointestinal_tract	Bacteroidetes	459	2.00E-126	75.3	996	>DS995534 Bacteroides dorei DSM 17855
D6D060_9BACE/51-155	1	Gastrointestinal_tract	Bacteroidetes	1879	0	100	1017	>GG705232 Bacteroides sp. 3_1_33FAA
E6SVI0_BACT6/51-155	1	Gastrointestinal_tract	Bacteroidetes	907	0	84.02	945	>FP929032 Alistipes shahii WAL 8301
CBK67812_CBK67812.1_Bacteroides_xylanisolvens_XB1A_hypothetical_protein	1	Gastrointestinal_tract	Bacteroidetes	1879	0	100	1017	>GG703854 Prevotella copri DSM 18205
ENA_CBW20928_CBW20928.1_Bacteroides_fragilis_638R_putative_hemolysin_A	1	Gastrointestinal_tract	Bacteroidetes	1873	0	100	1014	>FP929033 Bacteroides xylanisolvens XB1A
ENA_EDO52243_EDO52243.1_Bacteroides_uniformis_ATCC_8492_hemolysin	1	Gastrointestinal_tract	Bacteroidetes	1807	0	100	978	>GL882689 Bacteroides fluxus YIT 12057
ENA_EDS14876_EDS14876.1_Bacteroides_stercoris_ATCC_43183_hemolysin_	1	Gastrointestinal_tract	Bacteroidetes	1890	0	100	1023	>FP929033 Bacteroides xylanisolvens XB1A
ENA_EDY97076_EDY97076.1_Bacteroides_plebeius_DSM_17135_hemolysin_	1	Gastrointestinal_tract	Bacteroidetes	1829	0	100	990	>JH636044 Bacteroides sp. 3_2_5

ENA_EEY82825_EEY82825.1_Bacteroides_sp._2_1_33B_hemolysin	1	Gastrointestinal_tract	Bacteroidetes	1801	0	100	975	>ACPT01000029 Bacteroides sp. D20
ENA_EFV04460_EFV04460.1_Prevotella_salivae_DSM_15606_hemolysin_	1	Gastrointestinal_tract	Bacteroidetes	1857	0	100	1005	>ABFZ02000020 Bacteroides stercoris ATCC 43183
ENA_EHB91285_EHB91285.1_Alistipes_indistinctus_YIT_12060_hypothetical_protein_	1	Gastrointestinal_tract	Bacteroidetes	1679	0	100	909	>ABQC02000004 Bacteroides plebeius DSM 17135
ENA_EHH01788_EHH01788.1_Paraprevotella_clara_YIT_11840_hemolysin	1	Gastrointestinal_tract	Bacteroidetes	1724	0	100	933	>GG705151 Bacteroides sp. 2_1_33B
ENA_EHP49568_EHP49568.1_Odoribacter_laneus_YIT_12061_hypothetical_protein	1	Gastrointestinal_tract	Bacteroidetes	1768	0	100	957	>GL629647 Prevotella salivae DSM 15606
I3YLB0_ALIFI/56-157	1	Gastrointestinal_tract	Bacteroidetes	941	0	84.58	953	>JH370372 Alistipes indistinctus YIT 12060
Q5LII1_BACFN/51-155	1	Gastrointestinal_tract	Bacteroidetes	1873	0	100	1014	>JH376579 Paraprevotella clara YIT 11840
Q8A247_BACTN/51-155	1	Gastrointestinal_tract	Bacteroidetes	1873	0	100	1014	>JH594596 Odoribacter laneus YIT 12061
R5C642_9BACE/51-155	1	Gastrointestinal_tract	Bacteroidetes	436	8.00E-120	75.43	924	>FP929032 Alistipes shahii WAL 8301
R5FQF1_9BACT/53-157	1	Gastrointestinal_tract	Bacteroidetes	416	1.00E-113	74.59	972	>ACWI01000002 Bacteroides sp. 2_1_56FAA
R5I942_9PORP/51-156	1	Gastrointestinal_tract	Bacteroidetes	111	5.00E-22	74.23	291	>JH636041 Bacteroides sp. 1_1_6
R5JGR8_9BACE/51-155	1	Gastrointestinal_tract	Bacteroidetes	1823	0	99.6	999	>KB905466 Bacteroides salyersiae WAL 10018
R5KD71_9BACT/52-157	1	Gastrointestinal_tract	Bacteroidetes	606	6.00E-171	78.43	955	>GL629647 Prevotella salivae DSM 15606
R5MMX8_9BACE/51-155	1	Gastrointestinal_tract	Bacteroidetes	1768	0	98.99	987	>ACWH01000030 Bacteroides ovatus 3_8_47FAA
R5NZI1_9BACT/51-155	1	Gastrointestinal_tract	Bacteroidetes	1690	0	99.36	933	>KB905466 Bacteroides salyersiae WAL 10018
R5UEV5_9BACE/51-155	1	Gastrointestinal_tract	Bacteroidetes	1857	0	99.7	1014	>JH379426 Prevotella stercorea DSM 18206
R5UPI5_9PORP/51-157	1	Gastrointestinal_tract	Bacteroidetes	1762	0	99.9	957	>ABJL02000006 Bacteroides intestinalis DSM 17393
R5VW07_9BACE/51-155	1	Gastrointestinal_tract	Bacteroidetes	1546	0	94.85	990	>JH376579 Paraprevotella clara YIT 11840
R6B4U0_9BACT/52-156	1	Gastrointestinal_tract	Bacteroidetes	726	0	80.02	991	>AAVM02000009 Bacteroides caccae ATCC 43185
R6BXV9_9BACT/52-157	1	Gastrointestinal_tract	Bacteroidetes	1707	0	97.87	987	>GG703854 Prevotella copri DSM 18205
R6DH15_9BACE/51-155	1	Gastrointestinal_tract	Bacteroidetes	1120	0	86.61	1016	>GG688329 Bacteroides finegoldii DSM 17565
R6FKP1_9BACE/51-155	1	Gastrointestinal_tract	Bacteroidetes	789	0	81.18	983	>DS499674 Bacteroides stercoris ATCC 43183
R6FUQ8_9BACT/52-158	1	Gastrointestinal_tract	Bacteroidetes	1474	0	93.45	993	>JH379426 Prevotella stercorea DSM 18206
R6KTM3_9BACE/51-155	1	Gastrointestinal_tract	Bacteroidetes	1807	0	99.7	987	>ACCH01000127 Bacteroides cellulosilyticus DSM 14838
R6LNJ9_9BACE/51-154	1	Gastrointestinal_tract	Bacteroidetes	1812	0	99.8	987	>AFBM01000001 Bacteroides clarus YIT 12056
R6MX16_9BACE/51-155	1	Gastrointestinal_tract	Bacteroidetes	817	0	81.75	981	>DS981492 Bacteroides coprocola DSM 17136
R6QE29_9BACT/52-157	1	Gastrointestinal_tract	Bacteroidetes	785	0	81.74	942	>GG703854 Prevotella copri DSM 18205
R6S950_9BACE/51-155	1	Gastrointestinal_tract	Bacteroidetes	1862	0	99.8	1014	>GG688329 Bacteroides finegoldii DSM 17565
R6SC61_9BACE/51-155	1	Gastrointestinal_tract	Bacteroidetes	1807	0	99.8	984	>ACBW01000097 Bacteroides coprophilus DSM 18228
R6VUA1_9BACT/56-157	1	Gastrointestinal_tract	Bacteroidetes	970	0	85.5	931	>FP929032 Alistipes shahii WAL 8301
R6XGV7_9BACT/52-157	1	Gastrointestinal_tract	Bacteroidetes	390	6.00E-106	74.54	923	>GG703854 Prevotella copri DSM 18205
R6YIB5_9BACE/51-155	1	Gastrointestinal_tract	Bacteroidetes	442	2.00E-121	76.02	880	>ACTC01000036 Bacteroides sp. 4_1_36
R7DDR3_9PORP/51-155	1	Gastrointestinal_tract	Bacteroidetes	1657	0	98.31	945	>ACWX01000035 Tannerella sp. 6_1_58FAA_CT1

R7EIP8_9BACE/51-155	1	Gastrointestinal_tract	Bacteroidetes	1768	0	99.28	978	>ACPT01000029 Bacteroides sp. D20
R7F021_9BACT/51-157	1	Gastrointestinal_tract	Bacteroidetes	76.8	2.00E-11	90	60	>AFZZ01000132 Prevotella stercorea DSM 18206
R7HSG0_9BACT/37-143	1	Gastrointestinal_tract	Bacteroidetes	126	2.00E-26	72.5	440	>AFZZ01000132 Prevotella stercorea DSM 18206
R7IYP9_9BACT/59-165	1	Gastrointestinal_tract	Bacteroidetes	233	1.00E-58	72.63	844	>JH379426 Prevotella stercorea DSM 18206
R7JHM4_9BACT/51-152	1	Gastrointestinal_tract	Bacteroidetes	1829	0	99.9	993	>ABFK02000017 Alistipes putredinis DSM 17216
E6K481_9BACT/52-156	1	Oral	Bacteroidetes	1834	0	100	993	>AEPD01000010 Prevotella buccae ATCC 33574
ENA_ADK95845_ADK95845.1_Prevotella_melaninogenica_ATCC_25845_hemolysin_	1	Oral	Bacteroidetes	1845	0	100	999	>CP002122 Prevotella melaninogenica ATCC 25845
ENA_EFI17261_EFI17261.1_Bacteroidetes_oral_taxon_274_str._F0058_hemolysin	1	Oral	Bacteroidetes	2176	0	100	1178	>ADCM01000011 Bacteroidetes oral taxon 274 str. F0058
ENA_EHG23013_EHG23013.1_Alloprevotella_rava_F0323_hypothetical_protein	1	Oral	Bacteroidetes	1840	0	100	996	>JH376829 Prevotella sp. oral taxon 302 str. F0323
ENA_EHO75052_EHO75052.1_Prevotella_micans_F0438_hypothetical_protein	1	Oral	Bacteroidetes	1834	0	100	993	>JH594521 Prevotella micans F0438
F2KX19_PREDF/64-168	1	Oral	Bacteroidetes	1895	0	100	1026	>CP002589 Prevotella denticola F0289
F9D3S1_PREDD/52-156_1	1	Oral	Bacteroidetes	1879	0	100	1017	>GL982488 Prevotella dentalis DSM 3688
I1YUM9_PREI7/53-157	1	Oral	Bacteroidetes	364	1.00E-98	73.71	985	>GG703886 Prevotella oris F0302
Q7MTR9_PORGI/53-158	1	Oral	Bacteroidetes	1801	0	100	975	>AJZS01000078 Porphyromonas gingivalis W50
R5CSR0_9BACT/52-157	1	Oral	Bacteroidetes	420	3.00E-115	75.28	906	>AWEY01000007 Prevotella baroniae F0067
R5GFN8_9BACT/51-155	1	Oral	Bacteroidetes	134	4.00E-29	70.21	866	>ACZS01000081 Prevotella sp. oral taxon 472 str. F0295
R5Q4D6_9BACT/52-157	1	Oral	Bacteroidetes	392	6.00E-107	74.28	972	>AWET01000051 Prevotella pleuritidis F0068
R6W2Q2_9BACT/52-156	1	Oral	Bacteroidetes	569	3.00E-160	77.34	993	>GL872283 Prevotella multiformis DSM 16608
R7CYB8_9BACE/51-155	1	Oral	Bacteroidetes	87.9	3.00E-15	71.47	375	>CP002122 Prevotella melaninogenica ATCC 25845
W0EP20_9PORP/51-155	1	Oral	Bacteroidetes	180	5.00E-43	71.8	773	>AWEY01000007 Prevotella baroniae F0067
C7M608_CAPOD/352-453	2	Oral	Bacteroidetes	3230	0	98.42	1836	>AMEV01000023 Capnocytophaga sp. oral taxon 324 str. F0483
ENA_EEK14630_EEK14630.1_Capnocytophaga_gingivalis_ATCC_33624_Acyltransferase_	2	Oral	Bacteroidetes	3330	0	100	1803	>ACLQ01000018 Capnocytophaga gingivalis ATCC 33624
ENA_EFS97491_EFS97491.1_Capnocytophaga_ochracea_F0287_Acyltransferase	2	Oral	Bacteroidetes	3391	0	100	1836	>AKFV01000035 Capnocytophaga ochracea str. Holt 25
F9YU78_CAPCC/351-452	2	Oral	Bacteroidetes	612	8.00E-173	73.1	1792	>AMEV01000023 Capnocytophaga sp. oral taxon 324 str. F0483
H1Z9S5_MYROD/346-447	2	Oral	Bacteroidetes	172	2.00E-40	72.59	540	>ALNN01000028 Capnocytophaga sp. CM59
ENA_EFA42931_EFA42931.1_Prevotella_bergensis_DSM_17361_hemolysin	1	Oral	Bacteroidetes	1823	0	100	987	>GG704783 Prevotella bergensis DSM 17361
A0A095ZG93_9BACT/52-156	1	Oral	Bacteroidetes	1596	0	95.41	1002	>ADEG01000046 Prevotella buccalis ATCC 35310
E7RNE3_9BACT/52-156	1	Oral	Bacteroidetes	1829	0	100	990	>AEPE02000002 Prevotella oralis ATCC 33269
ENA_EEK17761_EEK17761.1_Porphyromonas_uenonis_60-3_hemolysin_	1	Oral	Bacteroidetes	1751	0	100	948	>ACLR01000009 Porphyromonas uenonis 60-3
ENA_EFL47029_EFL47029.1_Prevotella_disiens_FB035-09AN_hemolysin_	1	Oral	Bacteroidetes	1834	0	100	993	>AEDO01000009 Prevotella disiens FB035-09AN

F4KL89_PORAD/55-160	1	Oral	Bacteroidetes	1735	0	99.48	954	>AENO01000054 Porphyromonas asaccharolytica PR426713P-I
I4Z8L9_9BACT/52-156	1	Oral	Bacteroidetes	1829	0	100	990	>ADFO01000053 Prevotella bivia JCVIHMP010
R6CE12_9BACE/51-155	1	Oral	Bacteroidetes	75	1.00E-11	72.32	289	>AEDO01000009 Prevotella disiens FB035-09AN
R6XAK6_9BACT/52-156	1	Oral	Bacteroidetes	436	1.00E-120	75.76	887	>AEPE02000002 Prevotella oralis ATCC 33269
ENA_EHL05550_EHL05550.1_Desulfitobacterium_hafniense_DP7_aminotransferase_class_V_	6	Gastrointestinal_tract	Firmicutes	3003	0	100	1626	>JH414482 Desulfitobacterium hafniense DP7
ENA_EFV76279_EFV76279.1_Bacillus_sp._2_A_57_CT2_serine-pyruvate_aminotransferase	6	Oral	Firmicutes	2976	0	100	1611	>GL635754 Bacillus sp. 2_A_57_CT2
A6T596_KLEP7/322-423	2	Oral	Proteobacteria	3081	0	99.01	1719	>JH930419 Klebsiella pneumoniae subsp. pneumoniae WGLW2
D8MWX6_ERWBE/367-468	2	Oral	Proteobacteria	525	3.00E-147	73.37	1506	>GG753567 Serratia odorifera DSM 4582
ENA_EFE94777_EFE94777.1_Serratia_odorifera_DSM_4582_Acyltransferase	2	Oral	Proteobacteria	3181	0	100	1722	>GG753567 Serratia odorifera DSM 4582
Q6CZN2_PECAS/322-423	2	Oral	Proteobacteria	399	2.00E-109	71.6	1634	>ADBY01000051 Serratia odorifera DSM 4582
A0A0B5CH45_NEIEG/32-132	4	Oral	Proteobacteria	1386	0	100	750	>ADBF01000232 Neisseria elongata subsp. glycolytica ATCC 29315
E5UJR0_NEIMU/32-132	4	Oral	Proteobacteria	1397	0	100	756	>ACRG01000005 Neisseria mucosa C102
ENA_EET45812_EET45812.1_Neisseria_sicca_ATCC_29256_hypothetical_protein	4	Oral	Proteobacteria	1424	0	100	771	>ACKO02000002 Neisseria sicca ATCC 29256
ENA_ACI09609_ACI09609.1_Klebsiella_pneumoniae_342_conserved_hypothetical_protein	3	Gastrointestinal_tract	Proteobacteria	3059	0	99.12	1701	>ACXA01000063 Klebsiella sp. 1_1_55
A4W746_ENT38/322-423	2	Gastrointestinal_tract	Proteobacteria	1417	0	81.88	1689	>FP929040 Enterobacter cloacae subsp. cloacae NCTC 9394
ENA_CBK85930_CBK85930.1_Enterobacter_cloacae_subsp._cloacae_NCTC_9394_Putative_hemolysin_	2	Gastrointestinal_tract	Proteobacteria	3142	0	100	1701	>FP929040 Enterobacter cloacae subsp. cloacae NCTC 9394
ENA_EFE54303_EFE54303.1_Providencia_rettgeri_DSM_1131_Acyltransferase	2	Gastrointestinal_tract	Proteobacteria	3197	0	100	1731	>ACCI02000039 Providencia rettgeri DSM 1131
ENA_EHM48796_EHM48796.1_Yokenella_regensburgei_ATCC_43003_Acyltransferase	2	Gastrointestinal_tract	Proteobacteria	3142	0	100	1701	>JH417874 Yokenella regensburgei ATCC 43003
F9ZAJ4_ODOSD/341-443	2	Gastrointestinal_tract	Proteobacteria	1013	0	77.5	1738	>JH594597 Odoribacter laneus YIT 12061
G9Z3T1_9ENTR/322-423	2	Gastrointestinal_tract	Proteobacteria	3142	0	100	1701	>JH417874 Yokenella regensburgei ATCC 43003
R5UYM1_9PORP/338-439	2	Gastrointestinal_tract	Proteobacteria	3314	0	99.89	1800	>ADMC01000028 Odoribacter laneus YIT 12061
ENA_ACS62992_ACS62992.1_Ralstonia_pickettii_12D_conserved_hypothetical_protein_	4	Gastrointestinal_tract	Proteobacteria	1541	0	100	834	>GL520222 Ralstonia sp. 5_7_47FAA
ENA_CAW29482_CAW29482.1_Pseudomonas_aeruginosa_LESB58_putative_hemolysin	4	Gastrointestinal_tract	Proteobacteria	1369	0	99.34	756	>ACWU01000206 Pseudomonas sp. 2_1_26
A0A089UDH2_9ENTR/323-424	2	Oral	Proteobacteria	870	0	76.22	1695	>ALNJ01000086 Klebsiella sp. OBRC7
E6WAC8_PANSA/322-423	2	Oral	Proteobacteria	233	7.00E-59	72.78	709	>GL892086 Enterobacter hormaechei ATCC 49162
ENA_EHT12133_EHT12133.1_Raoultella_ornithinolytica_10-5246_hypothetical_protein	2	Oral	Proteobacteria	1829	0	85.9	1723	>ALNJ01000086 Klebsiella sp. OBRC7

G7LV45_9ENTR/322-423	2	Oral	Proteobacteria	387	5.00E-105	75.31	875	>ALNJ01000086 Klebsiella sp. OBRC7
ENA_EER56350_EER56350.1_Neisseria_flavescens_SK114_hypothetical_protein_	4	Oral	Proteobacteria	1397	0	100	756	>ACQV01000022 Neisseria flavescens SK114
A0A077KL19_9FLAO/353-454	2	Oral	Proteobacteria	2289	0	89.09	1842	>GL379781 Chryseobacterium gleum ATCC 35910
A7MLT3_CROS8/322-423	2	Gastrointestinal_tract	Proteobacteria	630	1.00E-178	74.42	1591	>AMLL01000012 Klebsiella pneumoniae subsp. pneumoniae WGLW1
ENA_EFK33376_EFK33376.1_Chryseobacterium_gleum_ATCC_35910_Acyltransferase_	2	Oral	Proteobacteria	3402	0	100	1842	>GL379781 Chryseobacterium gleum ATCC 35910
ENA_CAP01857_CAP01857.2_Acinetobacter_baumannii_SDF_conserved_hypothetical_protein_	4	Pathogen	Proteobacteria	1441	0	99	804	>ACQB01000026 Acinetobacter baumannii ATCC 19606

***** Red text indicates that the original body reference genome location specified airways or blood, which likely are not the origin microbiome site. Analysis of available datasets based on reference genome analysis confirms the location indicated.

Supplementary Info Table 3. Synthetic *N*-acyl amino acid MS data (observed *m/z* in positive ion mode).

Code	Amine moiety	MW	Obs *m/z*
A	Alanine	327.3	328.35
R	Arginine	412.3	413.49
N	Asparagine	370.3	371.40
D	Aspartatic acid	371.3	372.40
C	Cysteine	359.2	360.37
Q	Glutamine	384.3	385.44
E	Glutamic acid	385.3	386.43
G	Glycine	313.3	314.34
H	Histidine	393.3	394.44
I	Isoleucine	369.3	369.23
L	Leucine	369.3	369.24
K	Lysine	384.3	385.47
M	Methionine	387.3	388.42
F	Phenylalanine	403.3	404.45
P	Proline	353.3	354.42
S	Serine	343.3	344.35
T	Threonine	357.3	358.42
W	Tryptophan	442.3	443.48
Y	Tyrosine	419.3	420.48
V	Valine	355.3	356.48

References

1 Brady, S. F., Chao, C. J. & Clardy, J. New natural product families from an environmental DNA (eDNA) gene cluster. *Journal of the American Chemical Society* **124**, 9968-9969 (2002).