Supporting Information

Evaluating translational efficiency of noncanonical amino acids to inform the design of druglike peptide libraries

Alix I Chan†, Manali S. Sawant†₴, Daniel J. Burdick‡, Jeffrey Tom†, Aimin Song†, and Christian N. Cunningham†*

† Department of Peptide Therapeutics, Genentech, South San Francisco, California 94080, United States
₴ Current Address: Department of Antibody and Protein Engineering, 23andMe, Inc., South San Francisco, California 94080, United States
‡ Department of Discovery Chemistry, Genentech, South San Francisco, California 94080, United States
* Email: cunningham.christian@gene.com

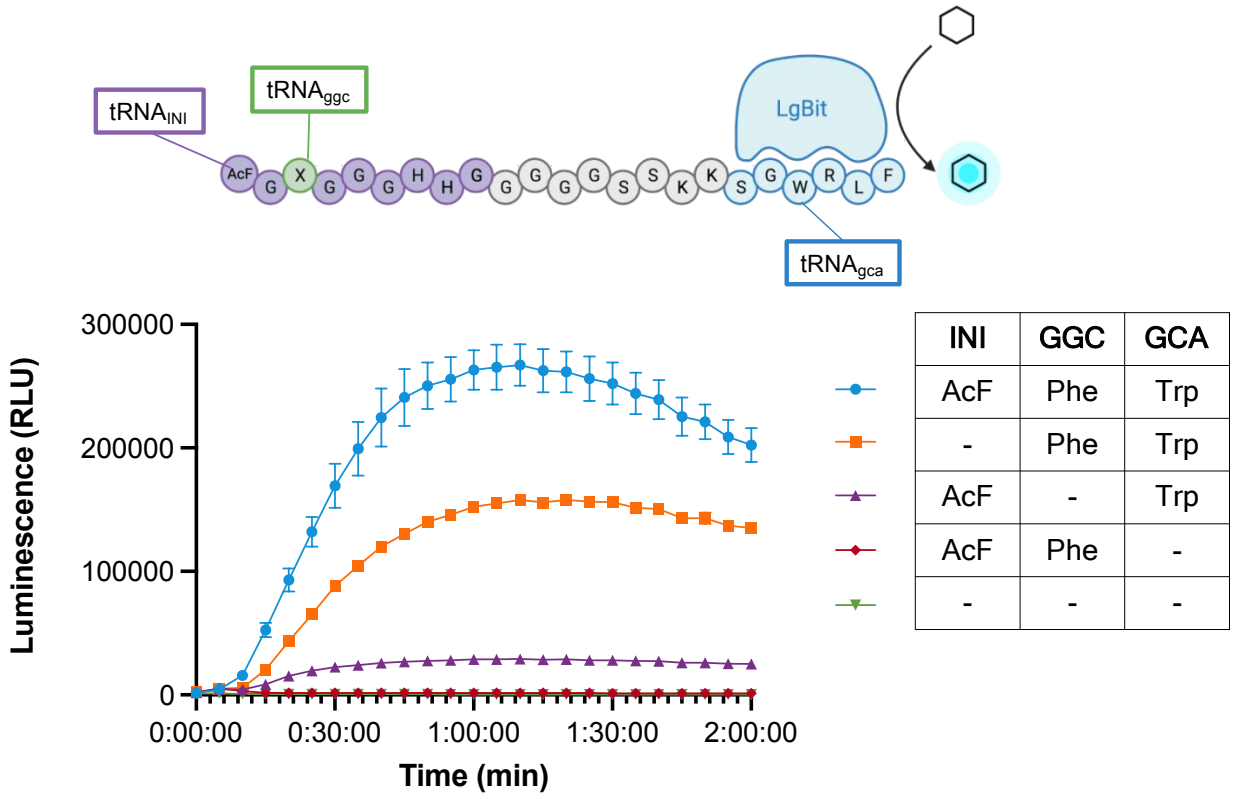
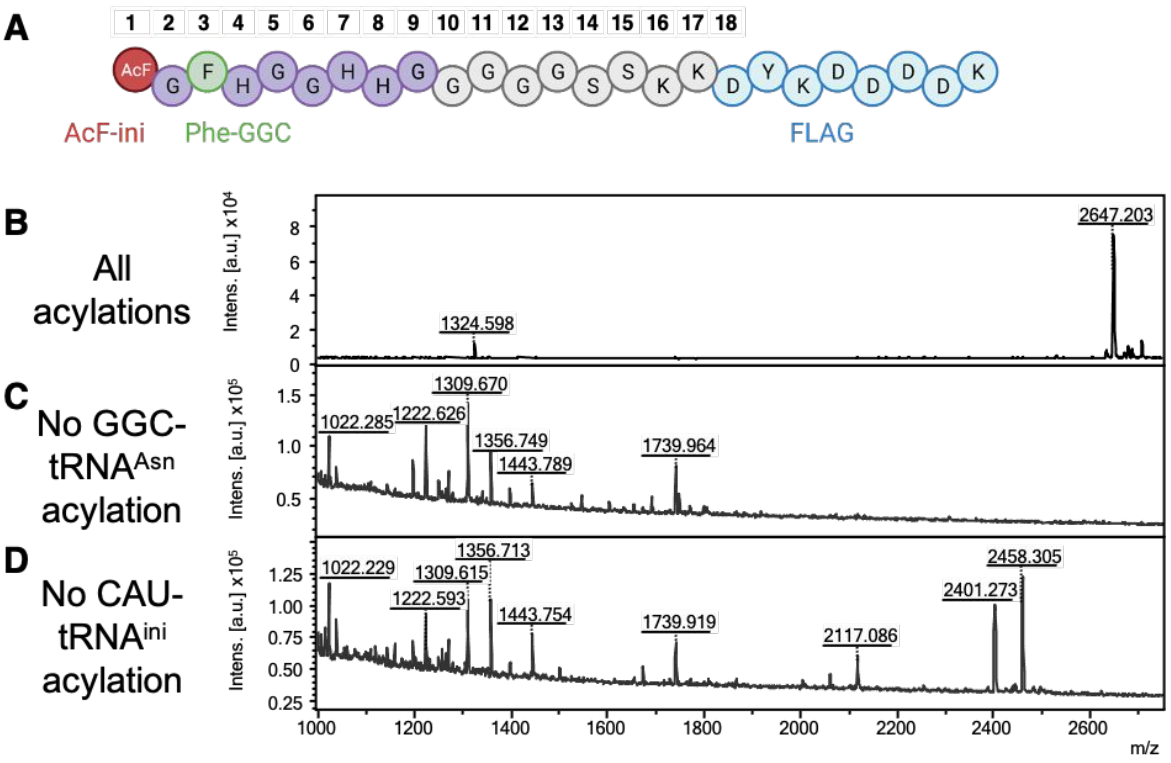


Figure S1. Luminescence signal in the FIT reaction of P3 is dependent on the aminoacylations of AcF-CAU-tRNA^{ini}, Trp-GCA-tRNA^{Asn}, and Phe-GCA-tRNA^{Asn}. Luminescence production is negligible when Trp-GCA-tRNA^{Asn} is omitted, and reduced in the absence of AcF-acylated CAU-tRNA^{ini} or Phe-acylated GGC-tRNA^{Asn}.



E

Initiation site	Calculated Mass	Observed Mass		
		B. All acylations	C. No GGC-tRNA^{Asn} acylation	D. No CAU-tRNA^{ini} acylation
1	2647.2	2647.2		
2	2458.1			2358.3
3	2401.1			2401.3
4	2254			
5	2116.9			2117.0
6	2059.9			2060.0
7	2002.9			
8	1865.8			
9	1728.8			
10	1671.8			1671.9
11	1614.7			
12	1557.7			
13	1500.7			1500.8
14	1443.7		1443.8	1443.8
15	1356.6		1356.7	1356.7
16	1269.6		1269.7	
17	1141.5			
18	1013.4			

Figure S2. MALDI-TOF characterization of translation readthrough or reinitiation products. (A) FLAG-containing peptide sequence translated, with the sites of flexizyme-aminoacylated tRNAs noted. (B-D) MALDI-TOF spectra of peptides from IVT reactions under various acylation conditions. (E) Table of expected masses for peptide truncation products and observed masses in the spectra show in (B-D).

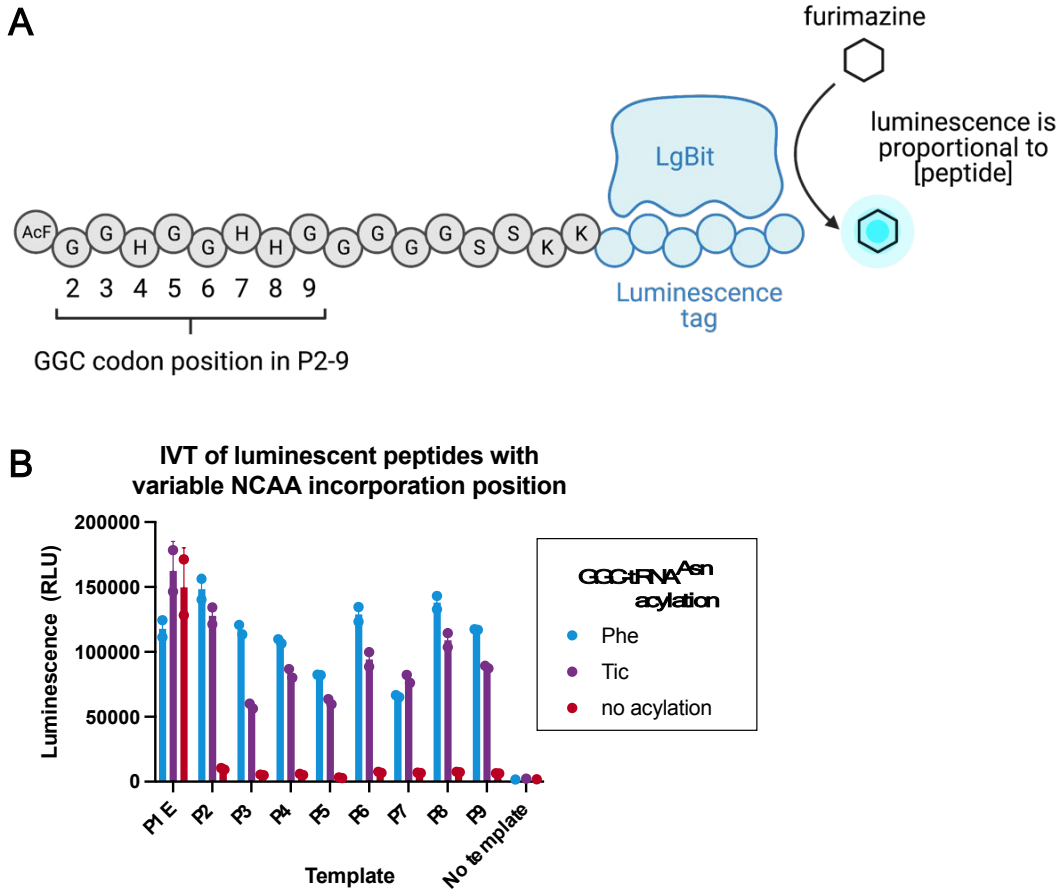


Figure S3 Selection and validation of test sequences for luminescence-based translational efficiency assay. (A) Assay schematic for luminescence-based detection of translated peptides. (B) Positional scan of amino acid incorporation and the effect on luminescence signal when incorporating either Phe, Tic or non-acylated GGC-tRNA^{Asn}. Template numberings correspond to positions annotated in Figure S3A.

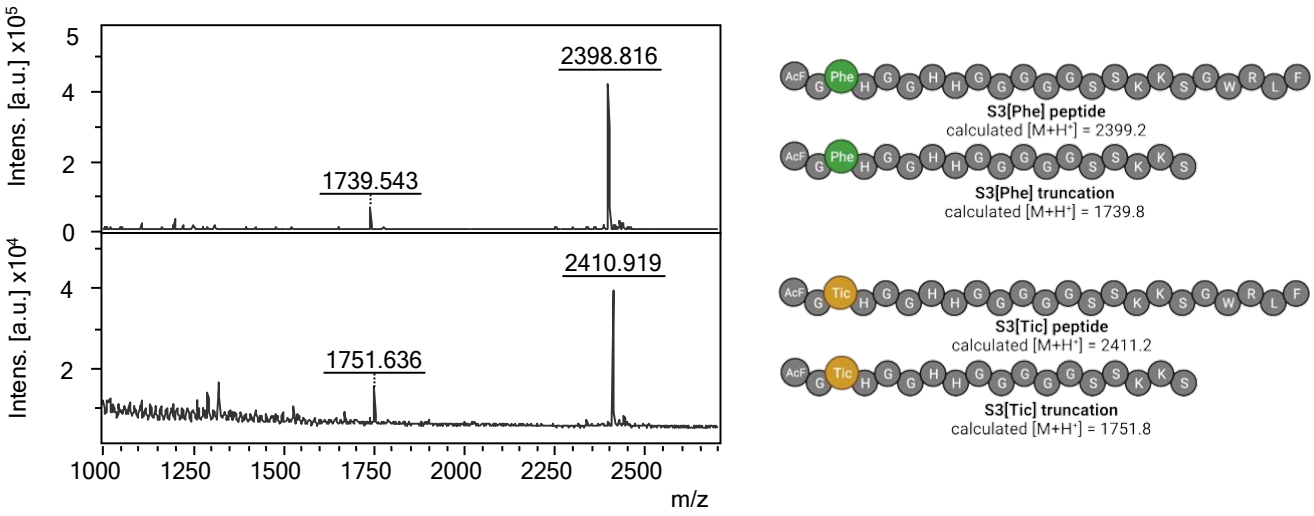


Figure S4. MALDI spectra from IVT reactions of **P3**[Phe] and **P3**[Tic]. Major products correspond to full-length expected product, with the only other observed product corresponding to truncation products after Ser18.

Table S1. Elongation amino acids tested in the study and their aminoacylation conditions. C = cyanomethyl ether, DNB = 3,5-dinitrobenzyl ester, ABT = amino-derivatized benzyl thioester H = HEPES pH 7.5; B = Bicine pH 9.0.

General name	**Abbreviation**	**Activation Group**	**Flexizyme**	**Buffer**	**(hr)**
L-Phenylalanine	Phe	CME	eFx	H	2
L-Alanine	Ala	DNB	dFx	H	**5**
L-Proline	Pro	DNB	dFx	H	2
Glycine	Gly	DNB	dFx	H	2
L-Tryptophan	Trp	CME	eFx	H	2
L-Serine	Ser	DNB	dFx	H	**5**
L-Threonine	Thr	DNB	dFx	H	**5**
L-Tyrosine	Tyr	CME	eFx	H	2
L-Cysteine	Cys	DNB	dFx	B	2
D-Phenylalanine	D-Phe	CME	eFx	H	5
D-Alanine	D-Ala	DNB	dFx	H	5
D-Proline	D-Pro	DNB	dFx	H	5
D-Tyrosine	D-Tyr	CME	eFx	H	5
D-Cysteine	D-Cys	DNB	dFx	H	5
L-Norleucine	Nle	DNB	dFx	H	5
(S)-2-Aminoheptanoic acid	Ahp	DNB	dFx	H	5
β-Cyclohexyl-L-alanine	Cha	DNB	dFx	H	5
α-Aminoisobutyric acid	Aib	DNB	dFx	H	5
1-Aminocyclopropane-1-carboxylic acid	Acp	DNB	dFx	H	5
(S)-1,2,3,4-Tetrahydroisoquinoline-3-carboxylic acid	Tic	ABT	aFx	H	2
L-4-methoxy-Phe	F4m	CME	eFx	B	2
L-Diphenylalanine	DiF	CME	eFx	H	2
4-Phenyl-L-phenylalanine, Biphenylalanine	Bip	CME	eFx	H	5
Ser(OMe)	SOM	DNB	dFx	B	2
Thr(OMe)	TOM	DNB	dFx	B	2
L-4-thiazolyl-Ala	4TzA	DNB	dFx	B	2
δ-pyrrolidyl-L-Asp	PyD	DNB	dFx	H	5
β-2-(S)-Homo-Phenylalanine	β^2Phe	CME	eFx	B	2
β-2-(S)-Homo-Leucine	β^2Leu	DNB	dFx	B	5
β-2-(S)-Homo-Valine	β^2Val	DNB	dFx	B	18
γ-Glycine	γGly	DNB	dFx	B	2

N-Methyl-L-phenylalanine	MeF	CME	eFx	H	5
N-Methyl-L-alanine	MeA	DNB	dFx	H	5
N-Methyl-L-norleucine	MeNle	DNB	dFx	H	5
N-Methyl-L-glycine (Sarcosine)	Sar	DNB	dFx	H	2
N-(2-Phenylethyl)-glycine	PEtG	ABT	aFx	H	2
Cyclopropyl-methyl-glycine	CpmG	DNB	dFx	B	2
N-n-Propyl-glycine	PrG	DNB	dFx	B	2
3-[(1-pyrenylacetyl)amino]-L-alanine	Dap(pyr)	CME	eFx	H	16
3-[[(7-methoxy-2-oxo-2H-1-benzopyran-3-yl)carbonyl]amino-L-alanine	Dap(Cou)	CME	eFx	H	2
3-[(7-Nitro-2,1,3-benzoxadiazol-4-yl)amino]-L-alanine	Dap(NBD)	DNB	aFx	H	16
L-Propargyl-Glycine	PropG	DNB	dFx	H	2
L-homopropargylglycine	hPropG	DNB	dFx	H	2
L-bishomopropargylglycine	hhPropG	DNB	dFx	H	2
4-Ethynyl-L-phenylalanine	Epa	CME	eFx	B	2
L-Azido-Alanine	AN_3	DNB	dFx	B	2
L-Azido-Homo-Alanine	hAN_3	DNB	dFx	H	2
L-Azido-Ornithine	$OrnN_3$	DNB	dFx	B	2
L-Azido-Lysine	$LysN_3$	DNB	dFx	B	2
Biocytin	KBio	DNB	dFx	B	2
4-Benzoyl-L-phenylalanine	Bpa	CME	eFx	H	5

Table S2. Initiator amino acids tested in the study and their aminoacylation conditions. H = HEPES pH 7.5; B = Bicine pH 9.0.

General name	Abbreviation	Activation Group	Flexizyme	Buffer	(hr)
N-Acetyl-L-phenylalanine	AcF	CME	eFx	H	2
N-Acetyl-L-alanine	AcA	DNB	dFx	B	5
N-Chloroacetyl-L-phenylalanine	ClAcF	CME	eFx	H	2
N-Chloroacetyl-L-alanine	ClAcA	DNB	dFx	B	2
N-Chloroacetyl-L-tyrosine	ClAcY	CME	eFx	H	2
N-Chloroacetyl-D-phenylalanine	ClAcf	CME	eFx	H	2
N-Chloroacetyl-D-alanine	ClAca	DNB	dFx	B	5
N-Chloroacetyl-D-tyrosine	ClAcy	CME	eFx	H	2
N-Biotinyl-L-phenylalanine	BtnPhe	CME	dFx	H	3

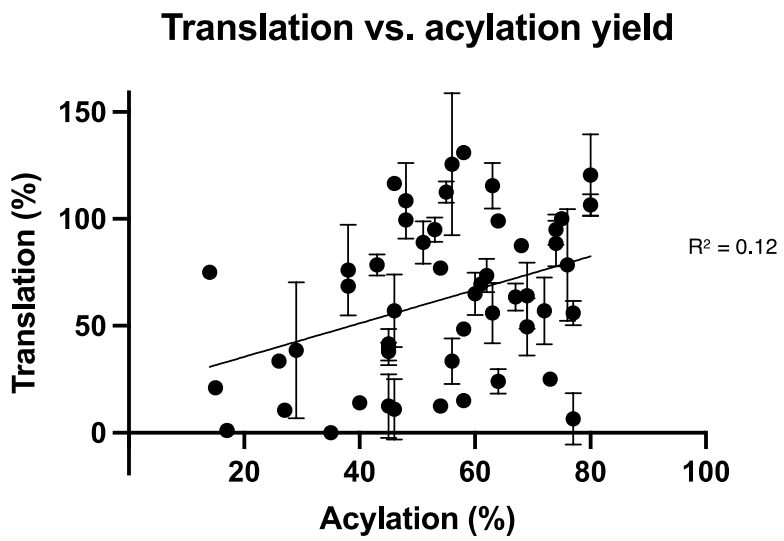


Figure S5. Translation efficiency vs acylation efficiency for all elongation amino acids tested in this study. No correlation is observed between acylation and translation efficiency (R^2 = 0.12).

Table S3. Theoretical diversities and calculated fractions of competent templates of the L8 and L14 library pools when translated with limited amino acid pools (See Figure 5B).

	L8-HiBit217		L14-HiBiT-217	
#aa	**Diversity**	**% competent**	**Diversity**	**% competent**
15	1.70×10^{7}	100%	2.80×10^{14}	100%
14	1.10×10^{7}	68%	1.30×10^{14}	46%
13	7.50×10^{6}	45%	5.70×10^{13}	20%
12	4.80×10^{6}	29%	2.30×10^{13}	8%
11	3.00×10^{6}	18%	8.90×10^{12}	3%
10	1.80×10^{6}	11%	3.10×10^{12}	1%
9	1.00×10^{6}	6.00%	1.00×10^{12}	0.40%
8	5.30×10^{5}	3.20%	2.80×10^{11}	0.10%
7	2.60×10^{5}	1.60%	6.90×10^{10}	0.02%
6	1.20×10^{5}	0.70%	1.40×10^{10}	0.01%

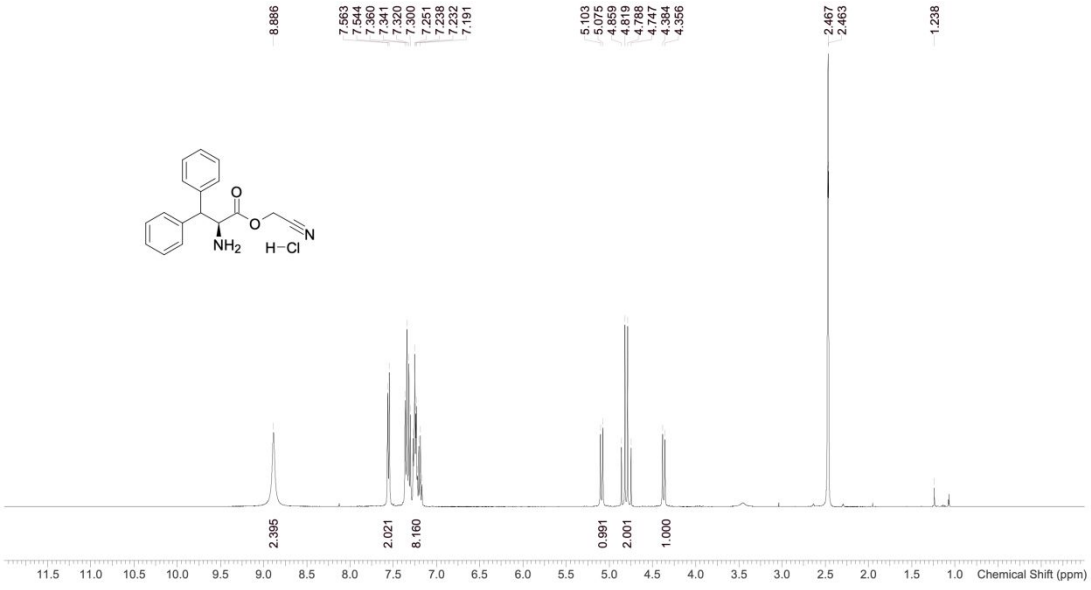
Table S4. Quantitation of library concentrations after IVT, as measured against a HiBiT217 standard.

Scaffold	Codon table	Final Concentration (µM)
L8-HiBiT217	Natural	17.1
L8-HiBiT217	N-methyl	16.3
L8-HiBiT217	Low IVT efficiency	10.4
L14-HiBiT217	Natural	9.0
L14-HiBiT217	N-methyl	9.0
L14-HiBiT217	Low IVT efficiency	7.2

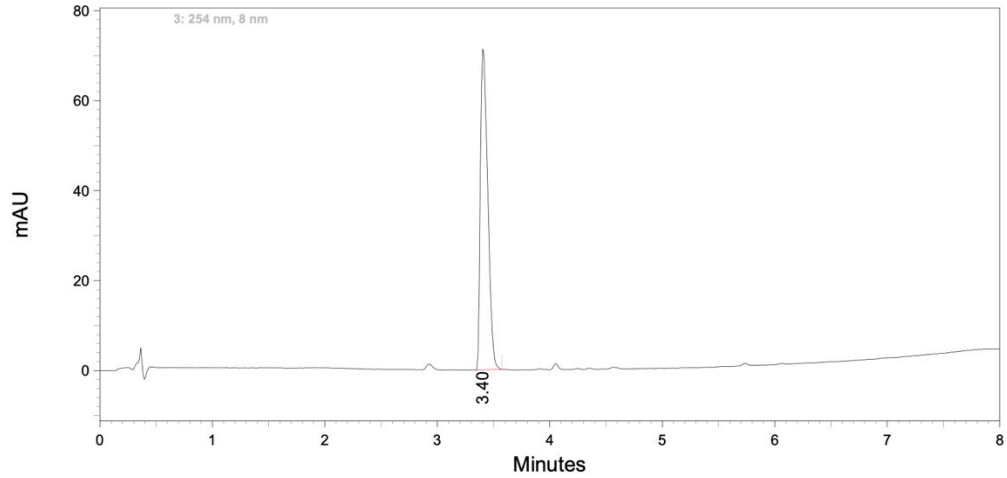
Characterization data of new compounds

^{1}H NMR spectrum of **DiF-CME**

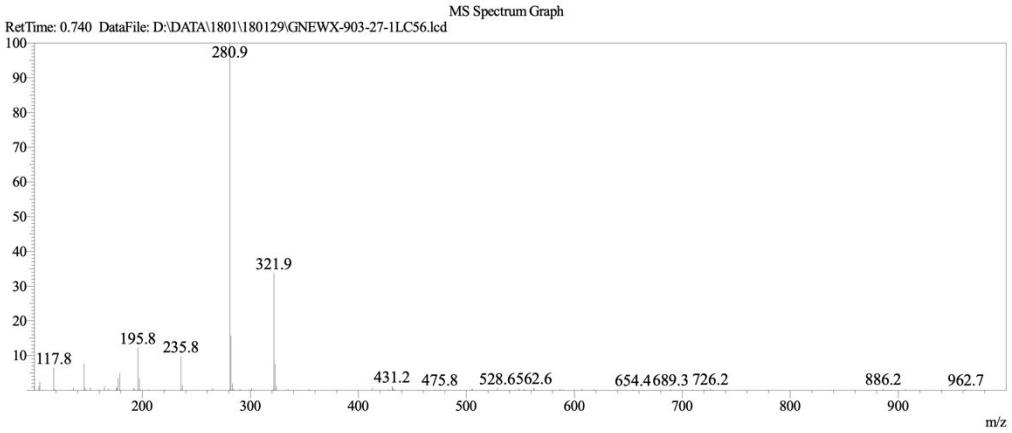
Cyanomethyl (S)-2-amino-3,3-diphenylpropanoate hydrochloride (**DiF-CME**). White solid. Yield 72%. ^{1}H NMR (400 MHz, DMSO-d_6) δ (ppm) 8.89 (s, br, 2H), 7.56 - 7.54 (m, 2H), 7.36 - 7.19 (m, 8H), 5.08 (d, *J* = 11.2 Hz, 1H), 4.80 (q, *J* = 12.0 Hz, 2H), 4.37 (d, *J* = 11.2 Hz, 1H). ESI-MS $[M+H]^+$ calcd. 281.1. found 280.9.



Analytical HPLC spectrum of **DiF-CME**

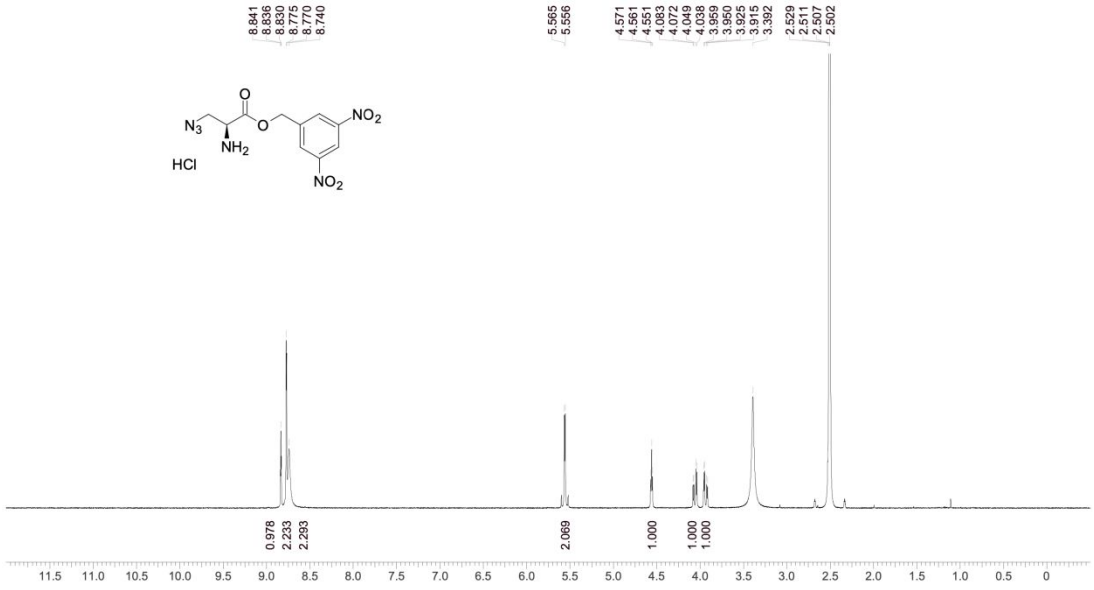


ESI-MS spectrum of **DiF-CME**

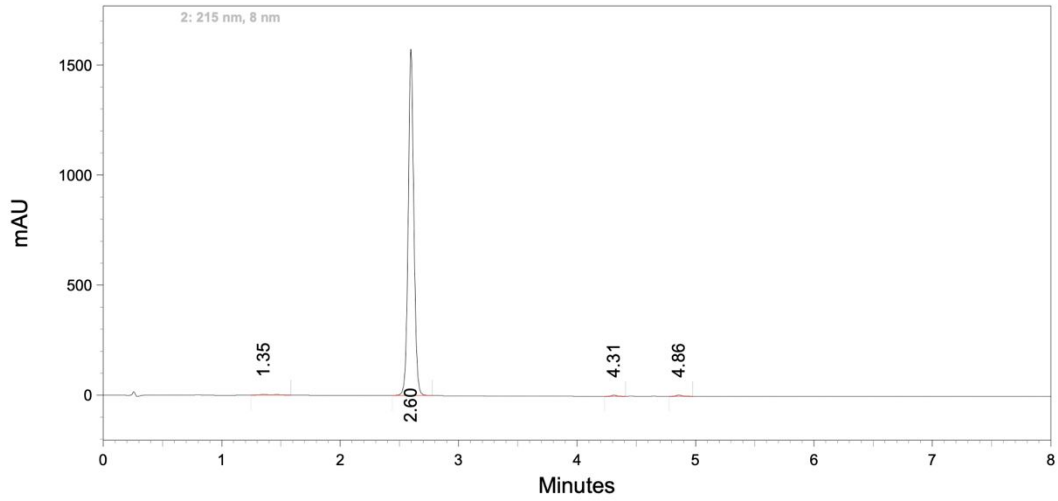


^{1}H NMR spectrum of **AN_3-DNB**

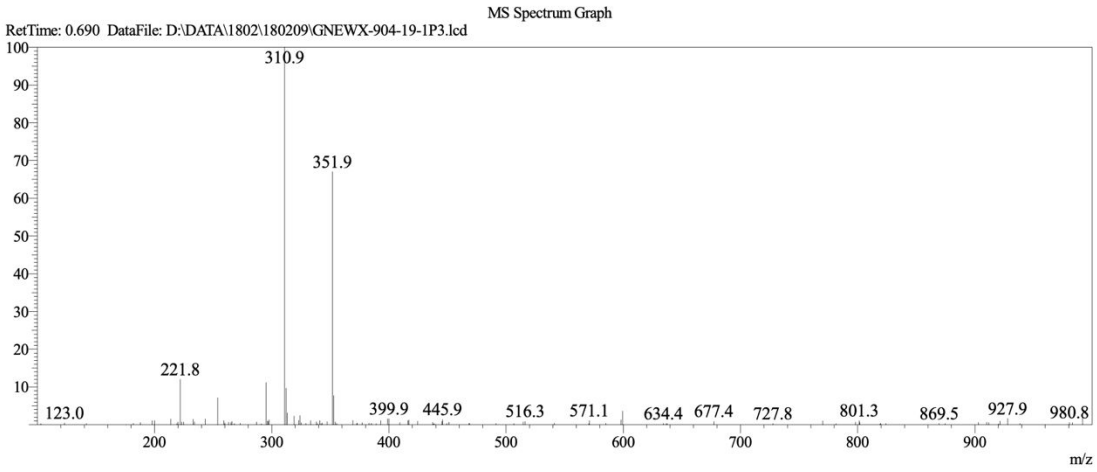
3,5-Dinitrobenzyl (S)-2-amino-3-azidopropanoate hydrochloride (**AN_3-DNB**). White solid. Yield 93%. ^{1}H NMR (400 MHz, DMSO-d_6) δ (ppm) 8.84 (t, *J* = 2.0 Hz, 1H), 8.77 (d, *J* = 2.0 Hz, 2H), 8.74 (s, 2H), 5.56 (d, *J* = 3.6 Hz, 2H), 4.56 (t, *J* = 4.0 Hz, 1H), 4.06 (dd, *J* = 13.6, 4.0 Hz, 1H), 3.94 (dd, *J* = 13.6, 4.0 Hz, 1H). ESI-MS $[M+H]^+$ calcd. 311.1. found 310.9.



Analytical HPLC spectrum of **AN_3-DNB**

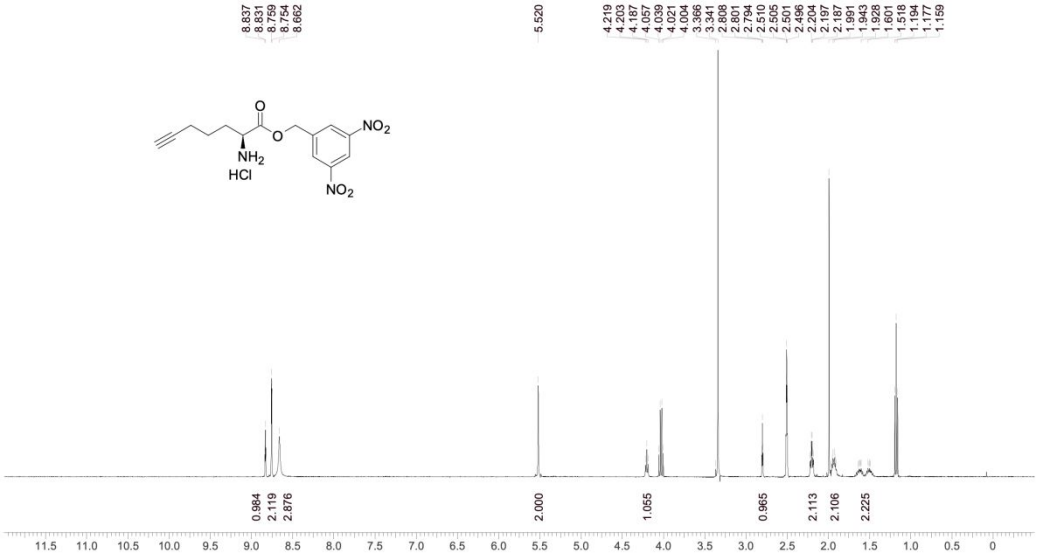


ESI-MS spectrum of **AN_3-DNB**

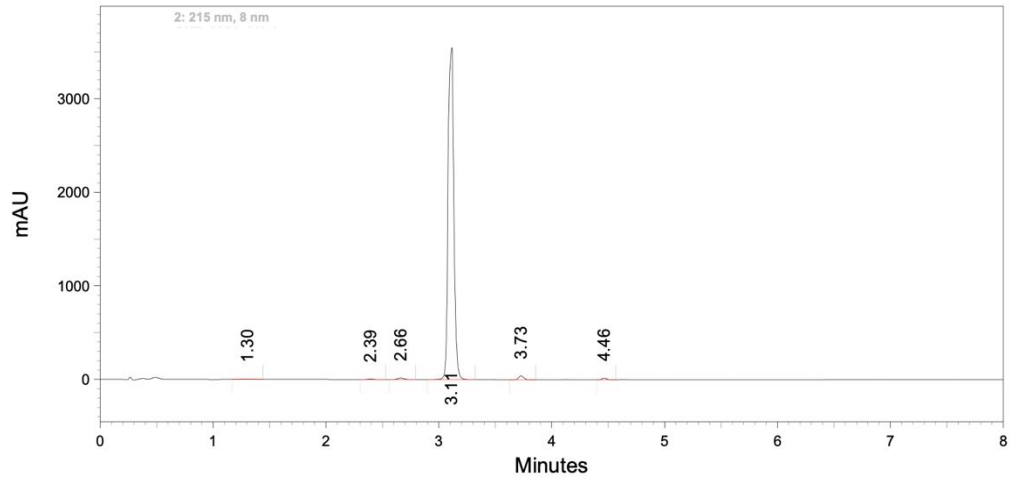


^{1}H NMR spectrum of **hhPropG-DNB**

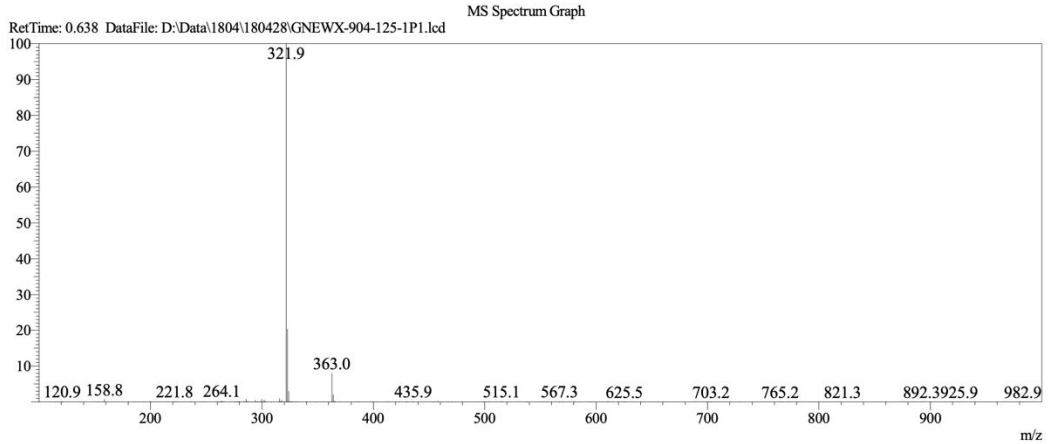
3,5-Dinitrobenzyl (S)-2-aminohept-6-ynoate hydrochloride (**hhPropG-DNB**). Yellow solid. Yield 81%. ^{1}H NMR (400 MHz, DMSO-d_6) δ (ppm) 8.83 (t, *J* = 2.0 Hz, 1H), 8.76 (d, *J* = 2.0 Hz, 2H), 8.66 (s, br 3H), 5.52 (s, 2H), 4.20 (t, *J* = 2.4 Hz, 1H), 2.80 (t, *J* = 2.8 Hz, 1H), 2.22 - 2.18 (m, 2H), 1.97 - 1.90 (m, 2H), 1.64 - 1.49 (m, 2H). ESI-MS $[M+H]^+$ calcd. 322.1. found 321.9.



Analytical HPLC spectrum of **hhPropG-DNB**

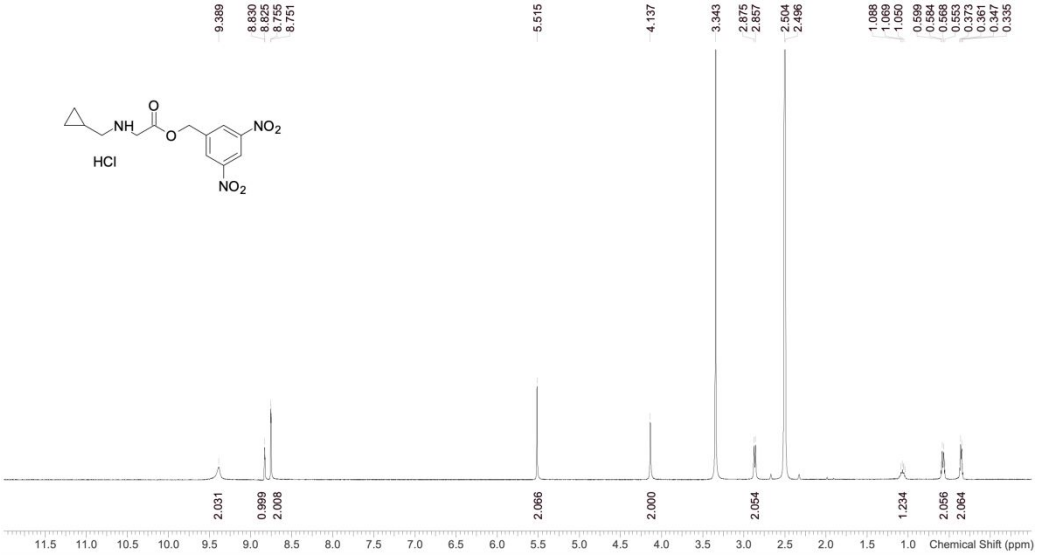


ESI-MS spectrum of **hhPropG-DNB**

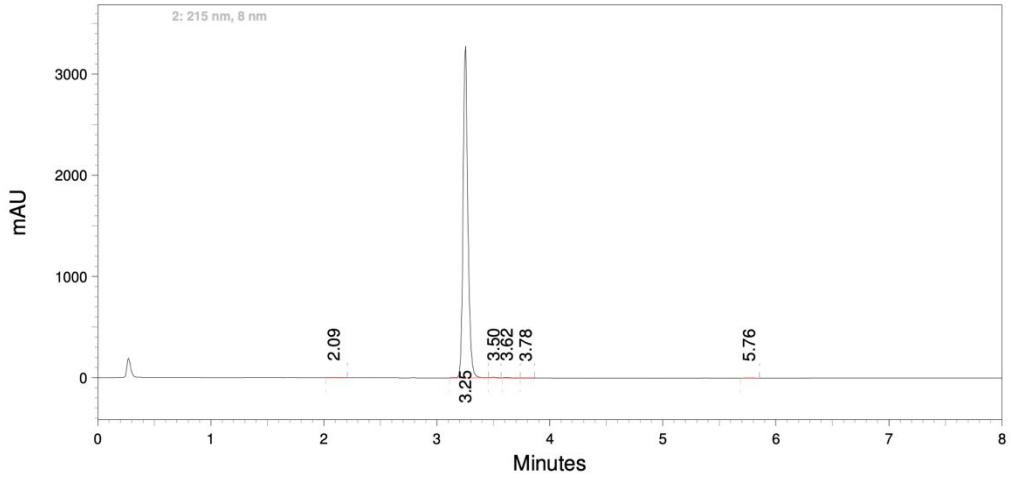


^{1}H NMR spectrum of **CpmG-DNB**

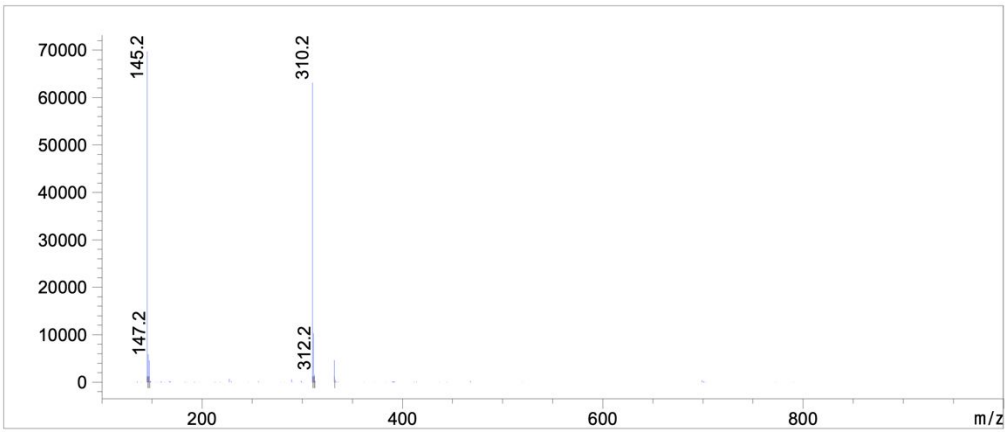
3,5-Dinitrobenzyl (cyclopropylmethyl)glycinate hydrochloride (**CpmG-DNB**). White solid. Yield 79%. ^{1}H NMR (400 MHz, DMSO-d_6) δ (ppm) 9.39 (s, br 2H), 8.85 - 8.81 (m, 1H), 8.77 - 8.73 (m, 2H), 5.52 (s, 2H), 4.14 (s, 2H), 2.87 (d, J = 6.8 Hz, 2H), 1.08 - 1.05 (m, 1H), 0.60 - 0.55 (m, 2H), 0.37 - 0.33 (m, 2H). ESI-MS $[M+H]^+$ calcd. 310.1. found 310.2.



Analytical HPLC spectrum of **CpmG-DNB**



ESI-MS spectrum of **CpmG-DNB**

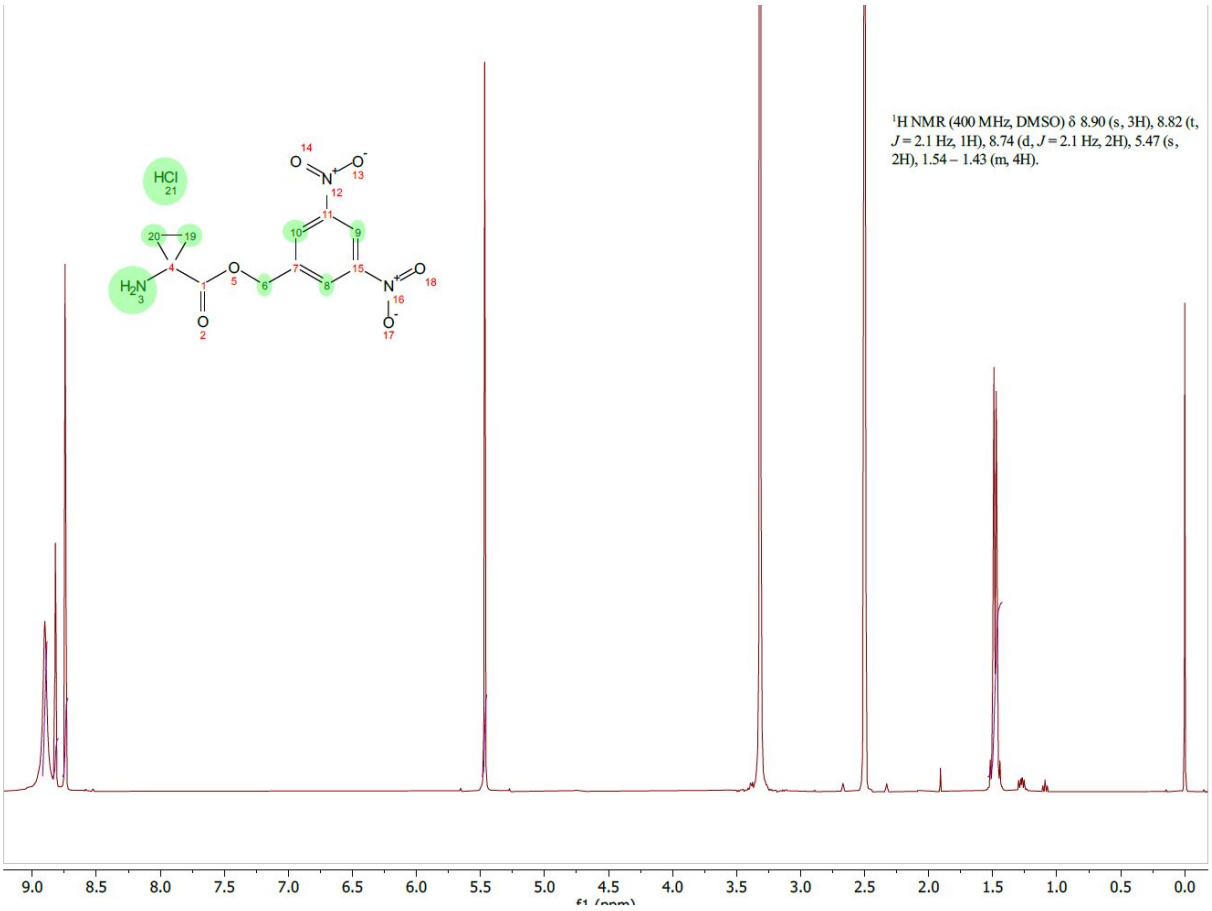


^{1}H NMR spectrum of **Acp-DNB**

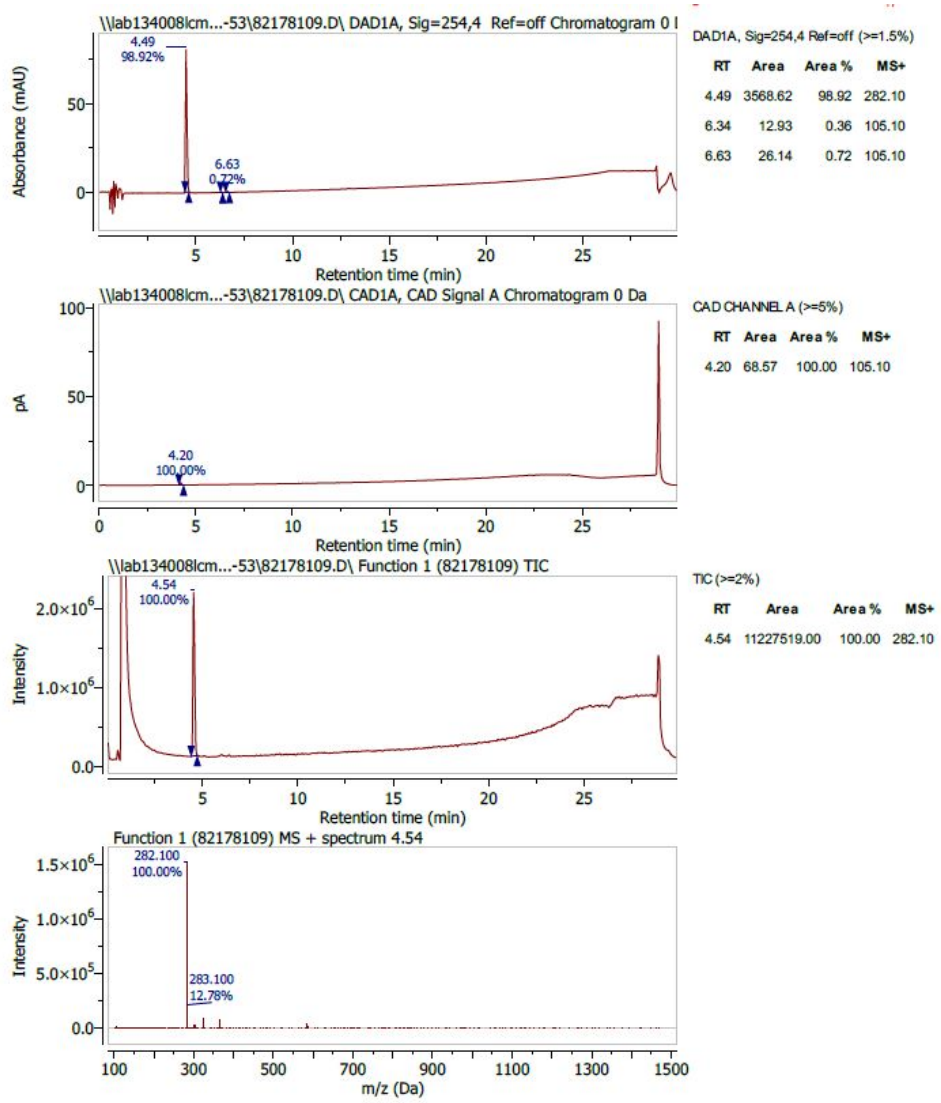
3,5-dinitrobenzyl 1-aminocyclopropane-1-carboxylate hydrochloride (**Acp-DNB**)

1H NMR (400 MHz, DMSO) d 8.90 (s, 3H), 8.82 (t, J = 2.1 Hz, 1H), 8.74 (d, J = 2.1 Hz, 2H), 5.47 (s, 2H), 1.54 – 1.43 (m, 4H).

Purity 99 %



HPLC and MS spectra of **Acp-DNB**

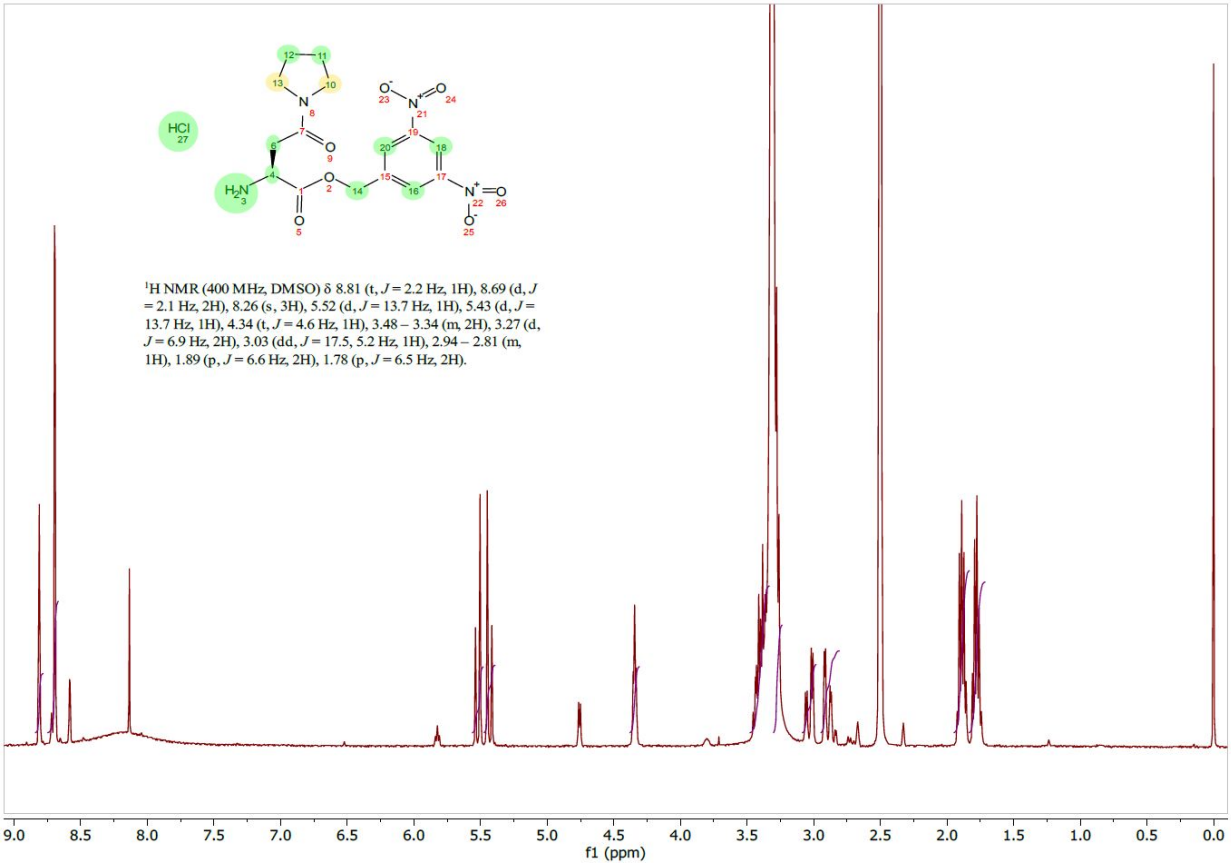


^{1}H NMR spectrum of **pyD-DNB**

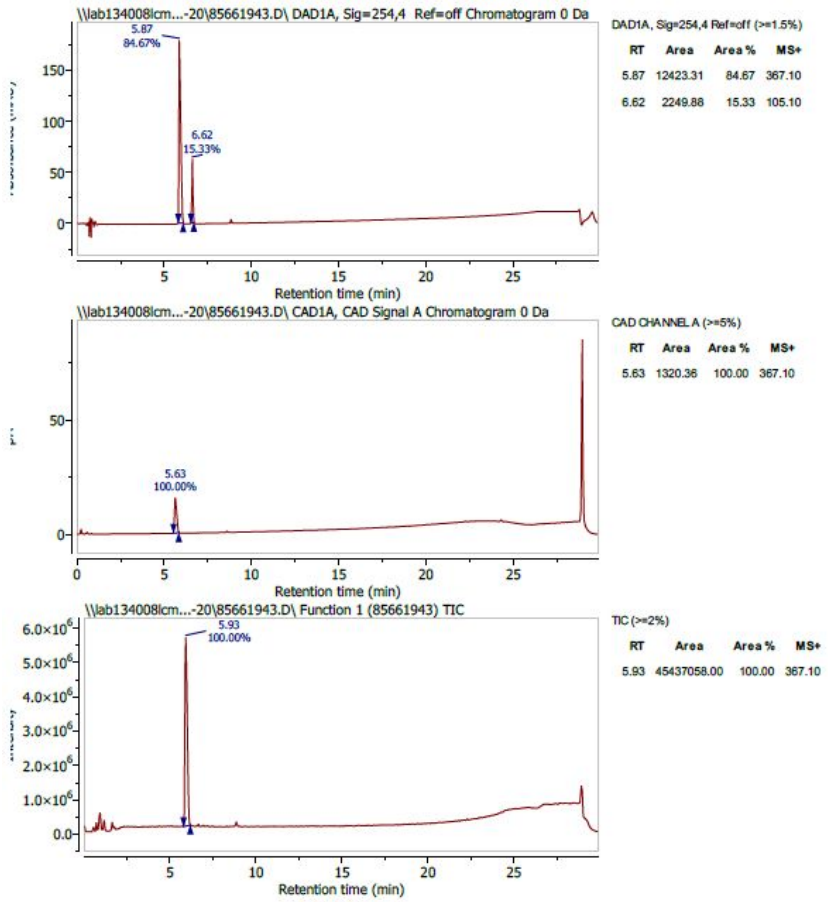
3,5-dinitrobenzyl (S)-2-amino-4-oxo-4-(pyrrolidin-1-yl)butanoate hydrochloride (**pyD-DNB**)

1H NMR (400 MHz, DMSO) d 8.81 (t, J = 2.2 Hz, 1H), 8.69 (d, J = 2.1 Hz, 2H), 8.26 (s, 3H), 5.52 (d, J = 13.7 Hz, 1H), 5.43 (d, J = 13.7 Hz, 1H), 4.34 (t, J = 4.6 Hz, 1H), 3.48 – 3.34 (m, 2H), 3.27 (d, J = 6.9 Hz, 2H), 3.03 (dd, J = 17.5, 5.2 Hz, 1H), 2.94 – 2.81 (m, 1H), 1.89 (p, J = 6.6 Hz, 2H), 1.78 (p, J = 6.5 Hz, 2H).

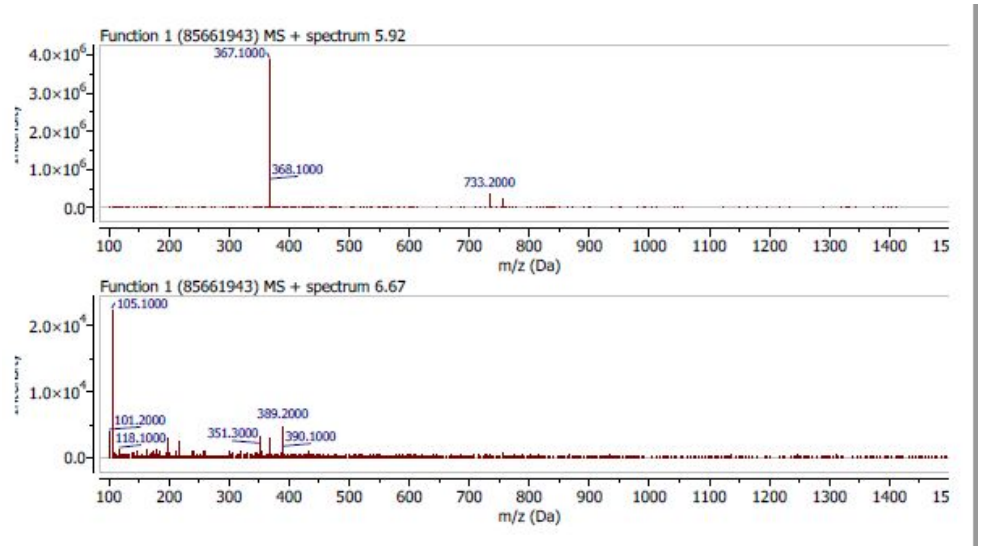
Purity 84.7 %



Analytical HPLC spectrum of **pyD-DNB**



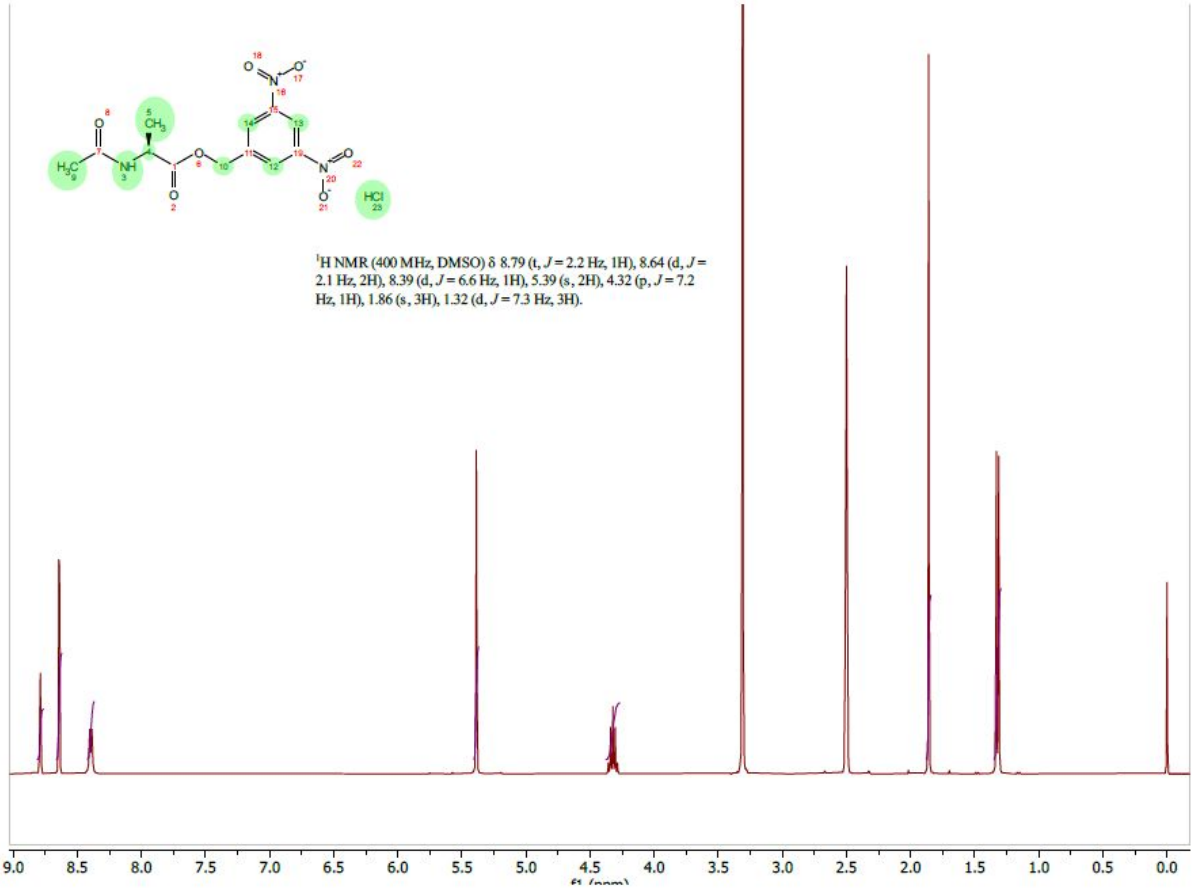
ESI-MS spectrum of **pyD-DNB**



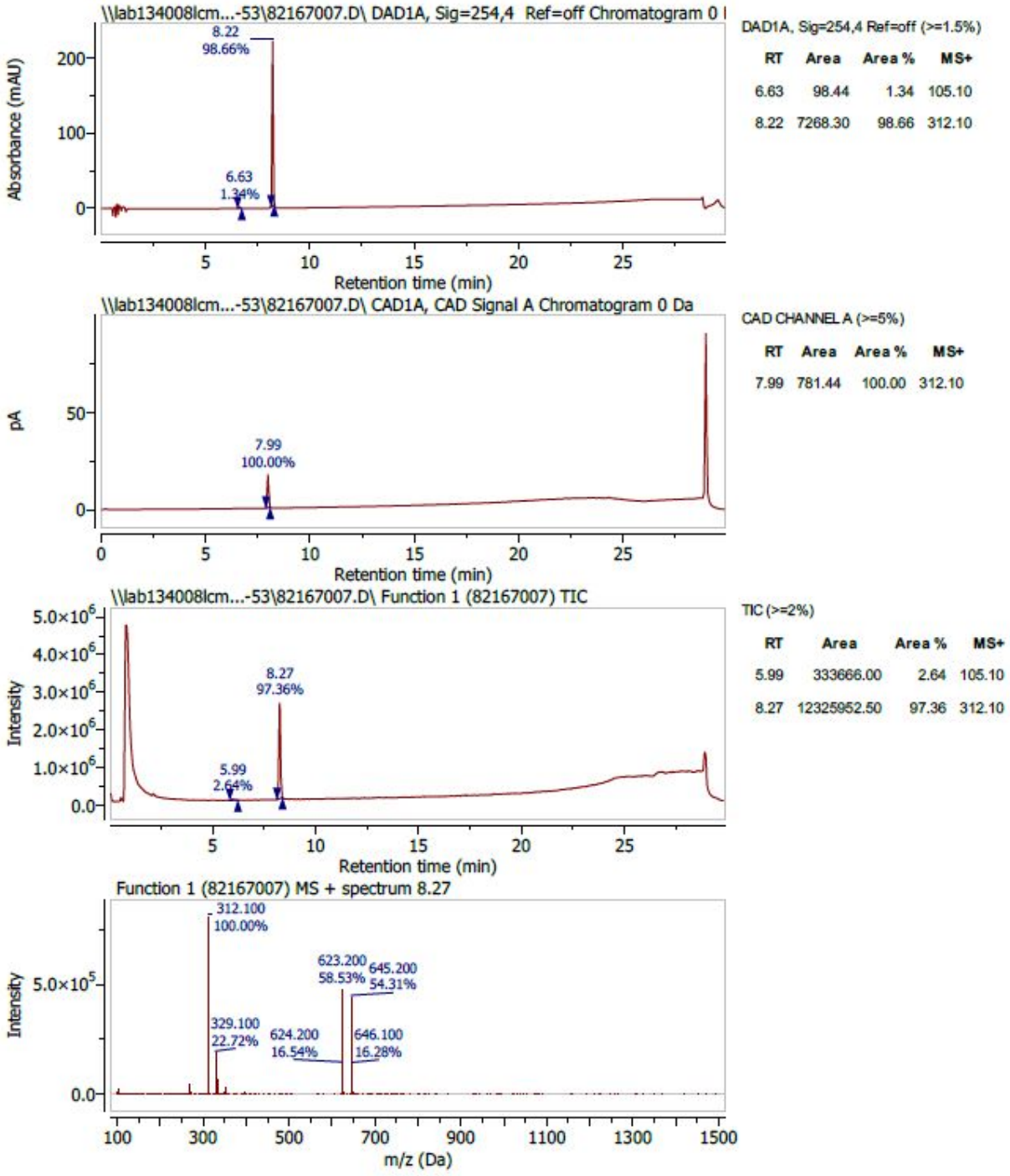
^{1}H NMR spectrum of 3,5-dinitrobenzyl acetyl-L-alaninate (**AcA-DNB**)

1H NMR (400 MHz, DMSO) d 8.79 (t, J = 2.2 Hz, 1H), 8.64 (d, J = 2.1 Hz, 2H), 8.39 (d, J = 6.6 Hz, 1H), 5.39 (s, 2H), 4.32 (p, J = 7.2 Hz, 1H), 1.86 (s, 3H), 1.32 (d, J = 7.3 Hz, 3H).

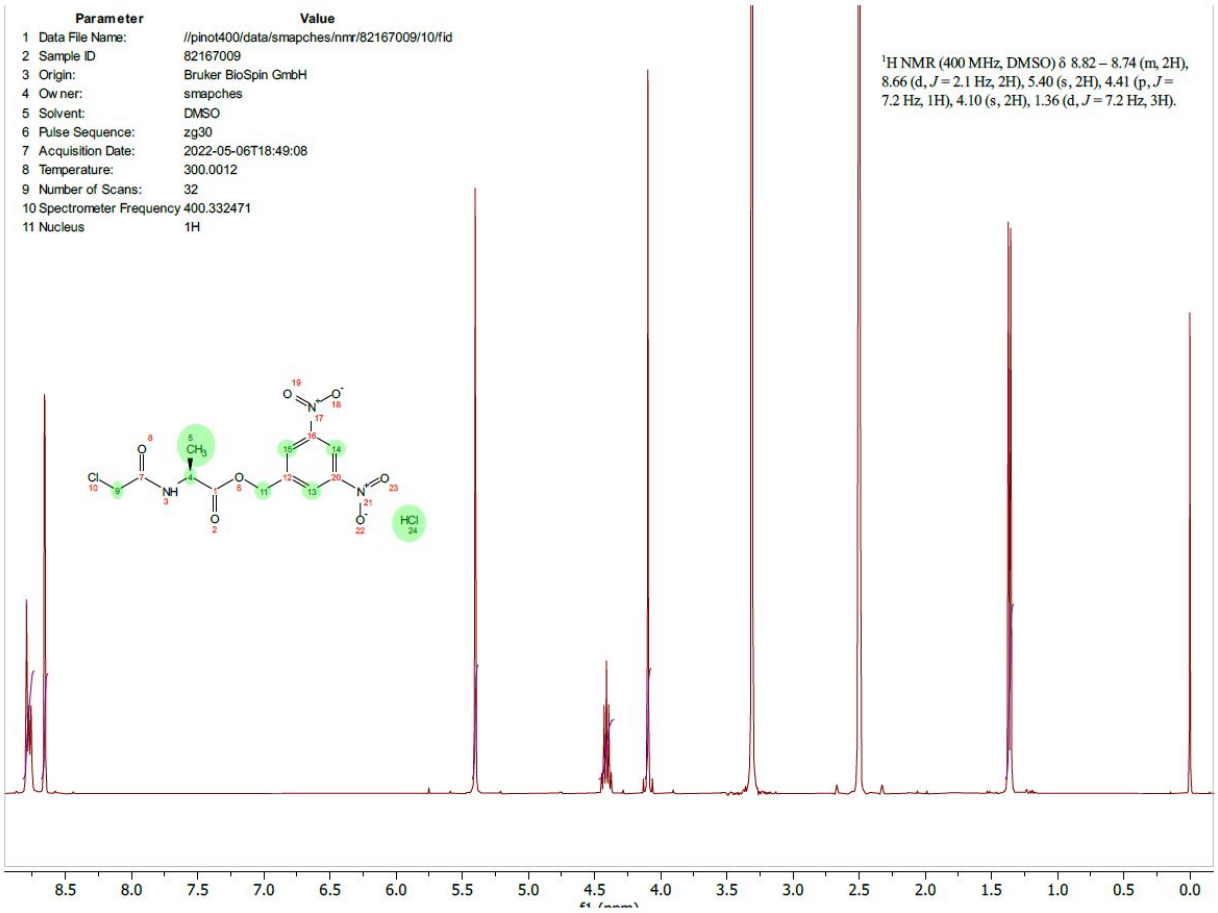
Purity 98 %



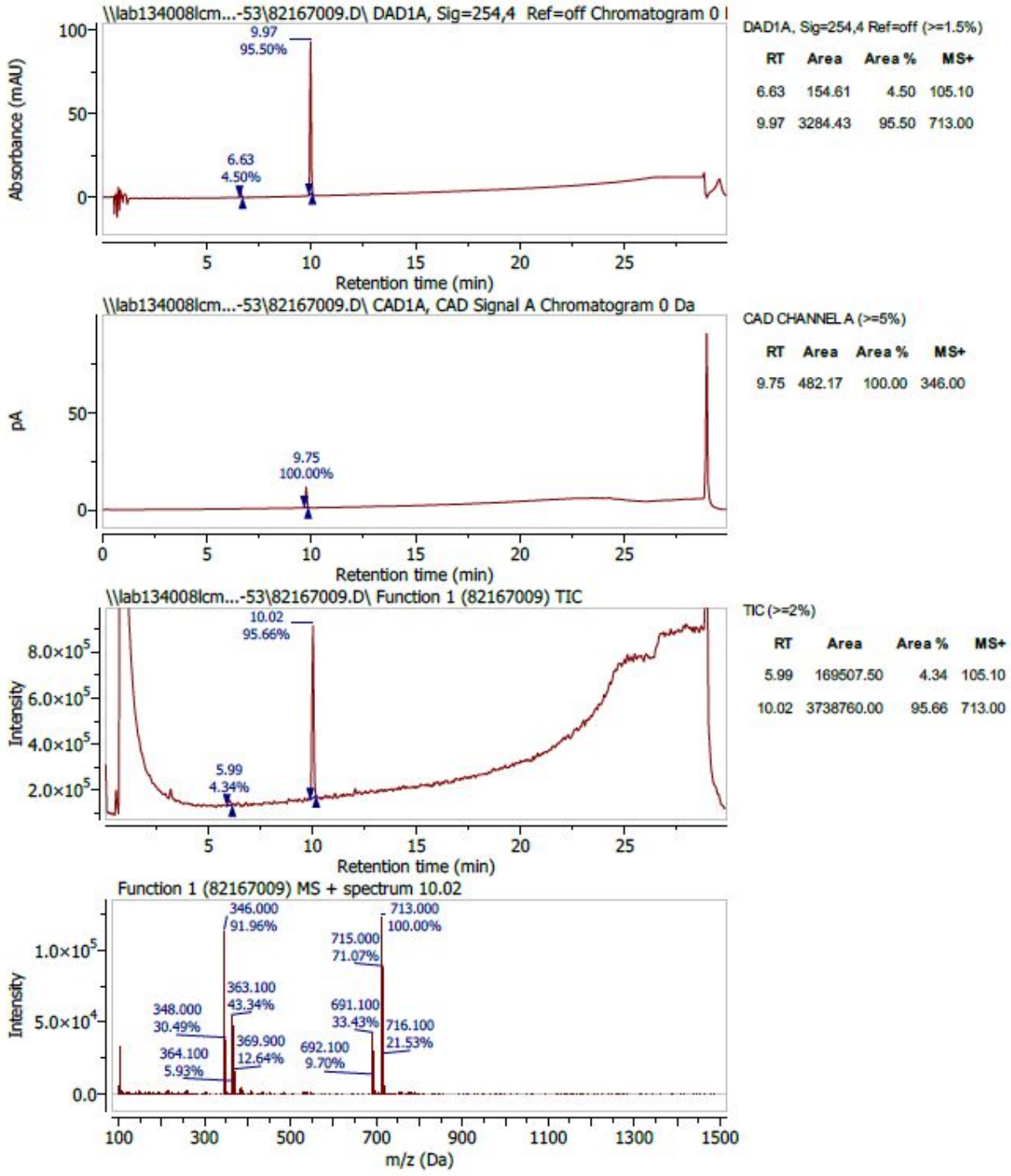
HPLC and MS spectra of **AcA-DNB**



^{1}H NMR spectrum of 3,5-dinitrobenzyl (2-chloroacetyl)carbamate hydrochloride (**ClAcA-DNB**)
1H NMR (400 MHz, DMSO) d 8.82 – 8.74 (m, 2H), 8.66 (d, J = 2.1 Hz, 2H), 5.40 (s, 2H), 4.41 (p, J = 7.2 Hz, 1H), 4.10 (s, 2H), 1.36 (d, J = 7.2 Hz, 3H).
Purity 99 %



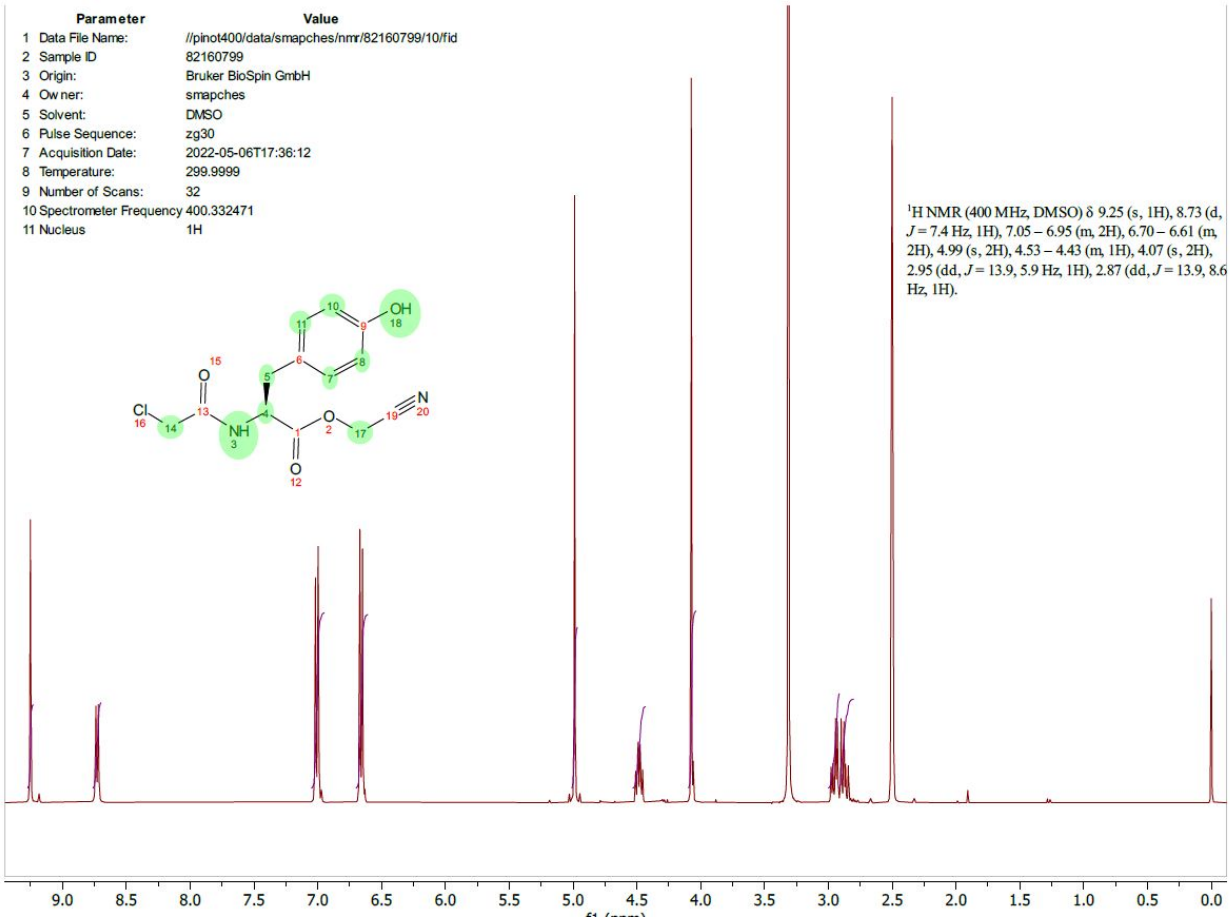
HPLC and MS spectra of **ClAcA-DNB**



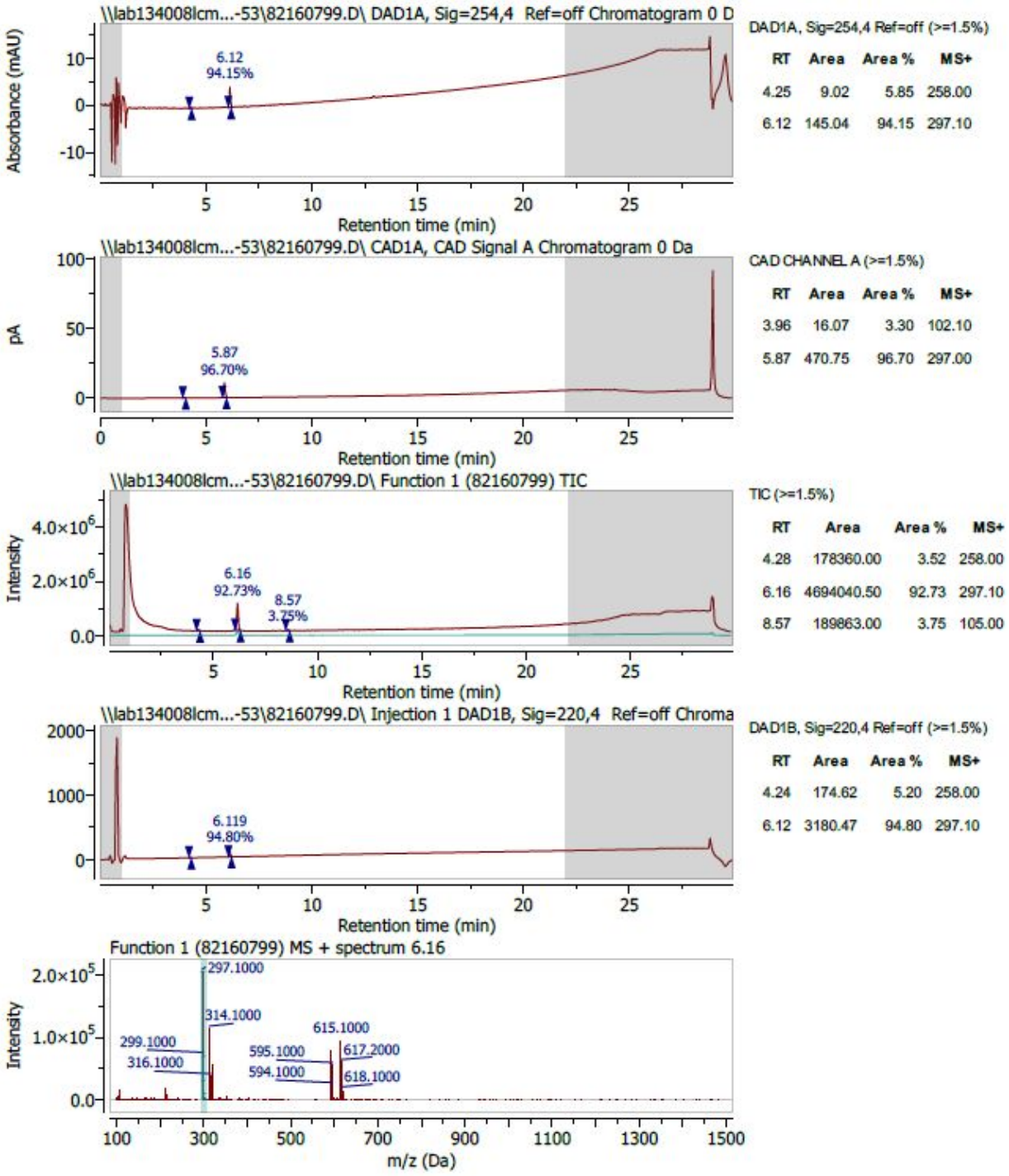
Cyanomethyl (2-chloroacetyl)-L-tyrosinate **(ClAcY-CME)**

1H NMR (400 MHz, DMSO) d 9.25 (s, 1H), 8.73 (d, J = 7.4 Hz, 1H), 7.05 – 6.95 (m, 2H), 6.70 – 6.61 (m, 2H), 4.99 (s, 2H), 4.53 – 4.43 (m, 1H), 4.07 (s, 2H), 2.95 (dd, J = 13.9, 5.9 Hz, 1H), 2.87 (dd, J = 13.9, 8.6 Hz, 1H).

Purity 94.8 %



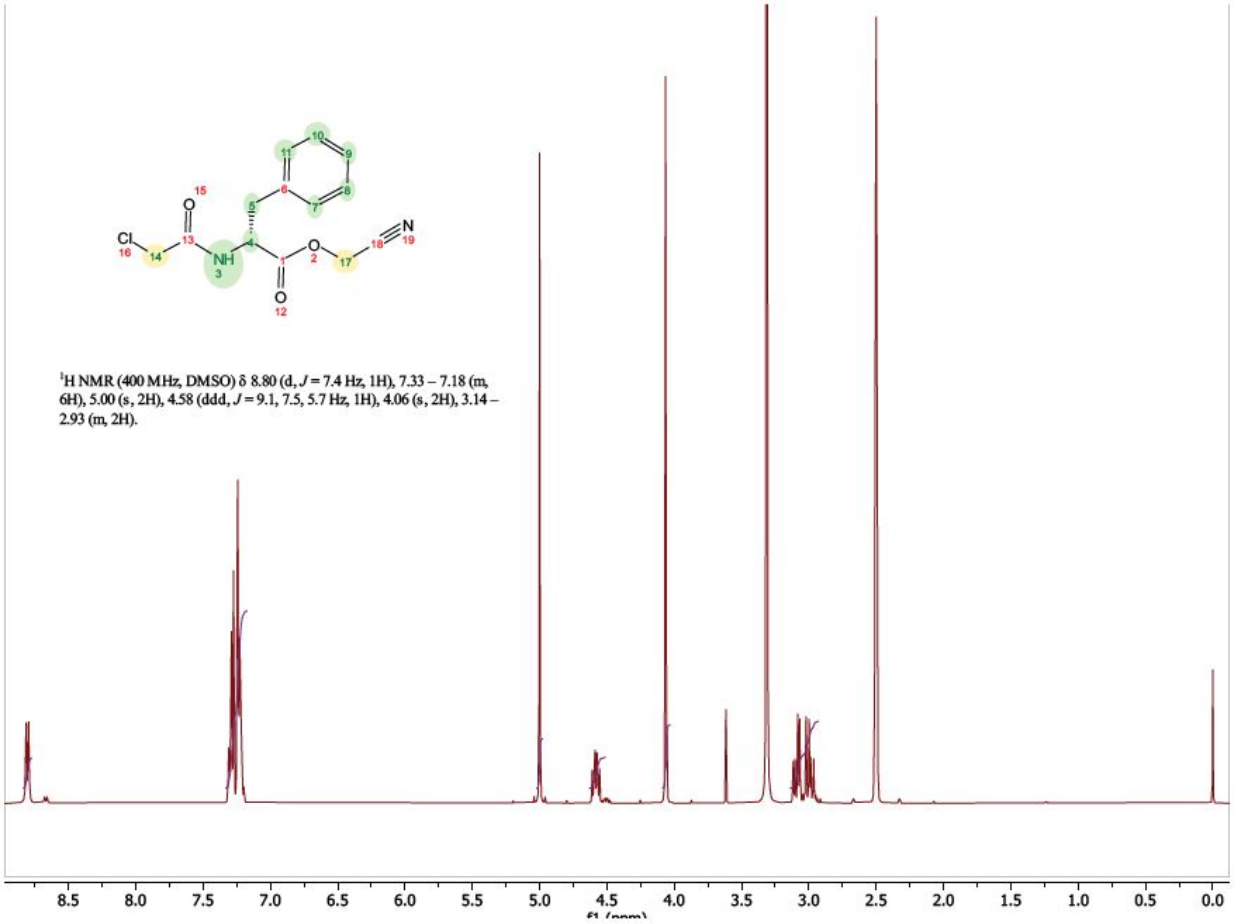
HPLC and MS spectra of **ClAcY-CME**



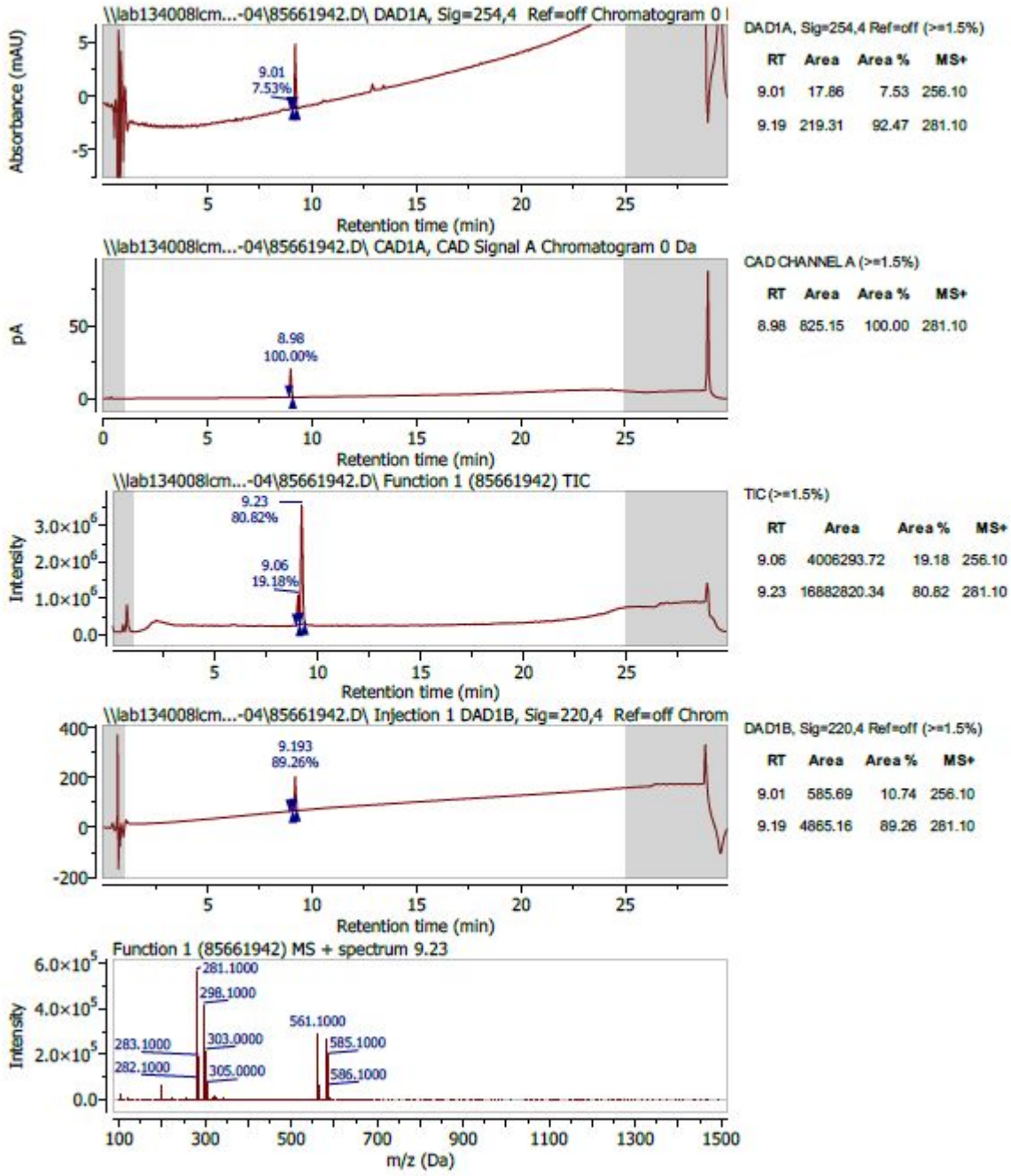
cyanomethyl (2-chloroacetyl)-D-phenylalaninate (**ClAcf-CME**)

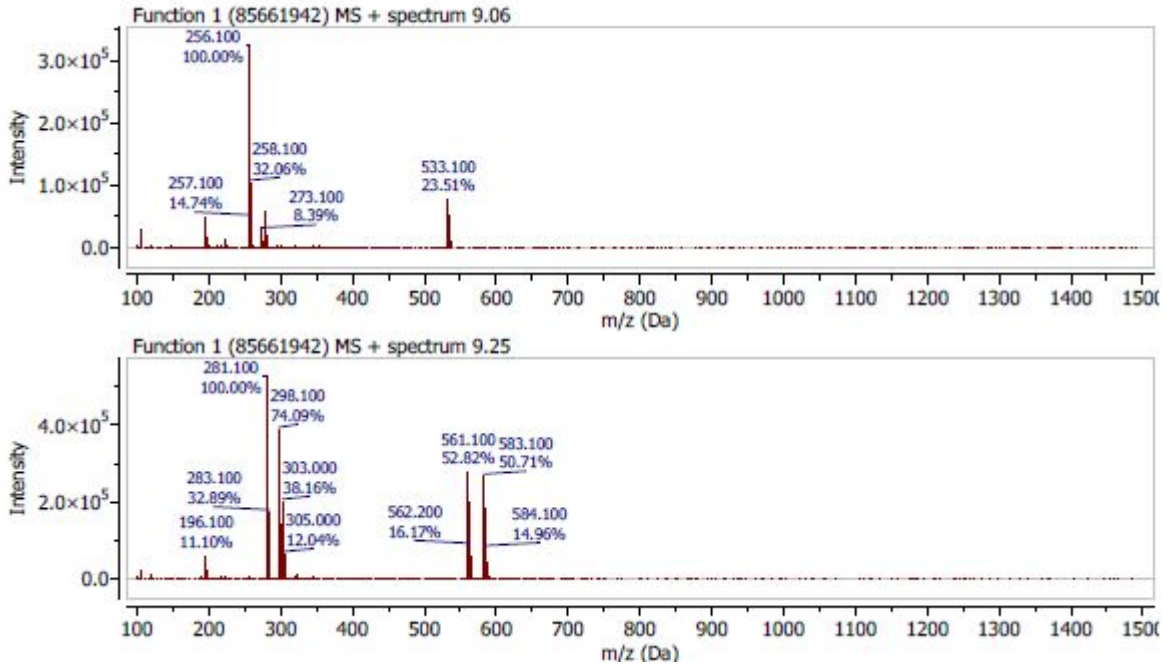
^{1}H NMR (400 MHz, DMSO) d 8.80 (d, J = 7.4 Hz, 1H), 7.33 – 7.18 (m, 6H), 5.00 (s, 2H), 4.58 (ddd, J = 9.1, 7.5, 5.7 Hz, 1H), 4.06 (s, 2H), 3.14 – 2.93 (m, 2H).

Purity 92.5 %



HPLC and MS spectra of **ClAcf-CME**

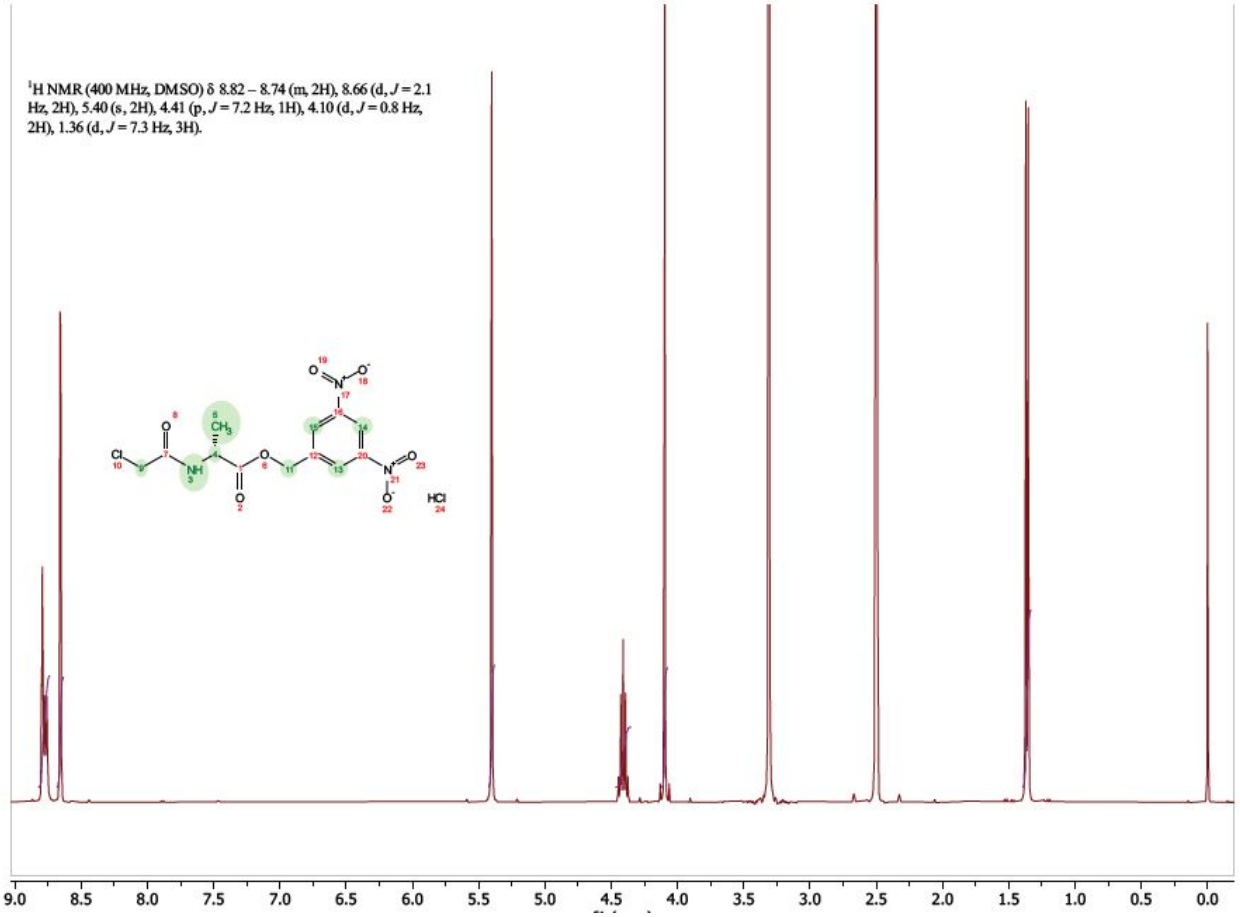



Function 1 (85661942) MS + spectrum 9.06
256.100
100.00%
258.100
32.06%
257.100
14.74%
273.100
8.39%
533.100
23.51%
Intensity
m/z (Da)
Function 1 (85661942) MS + spectrum 9.25
281.100
100.00%
298.100
74.09%
561.100
52.82%
583.100
50.71%
303.000
38.16%
283.100
32.89%
562.200
16.17%
584.100
14.96%
196.100
11.10%
305.000
12.04%
Intensity
m/z (Da)

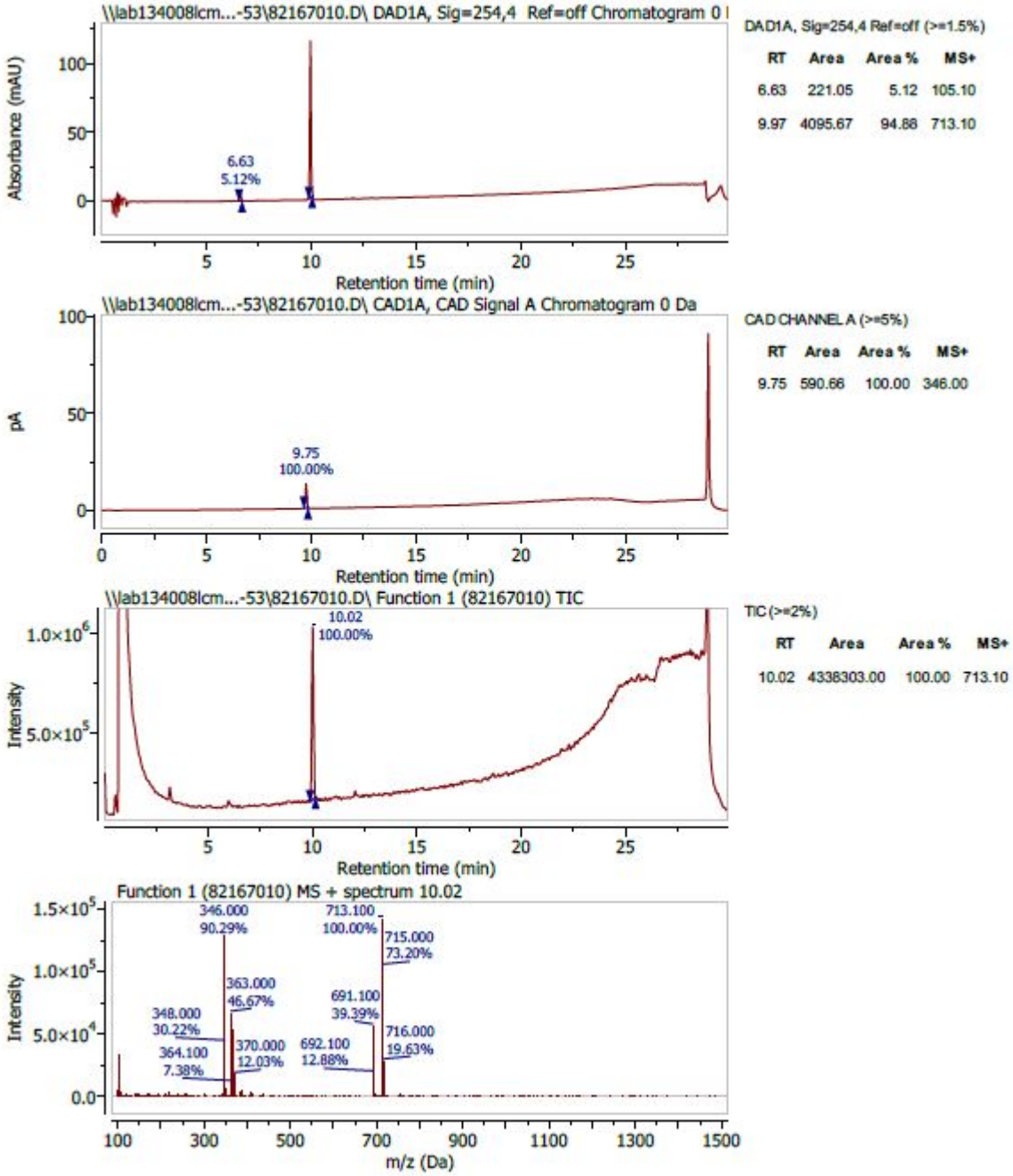
3,5-dinitrobenzyl (2-chloroacetyl)-D-alaninate hydrochloride (**ClAca-DNB**)

^{1}H NMR (400 MHz, DMSO) d 8.82 – 8.74 (m, 2H), 8.66 (d, J = 2.1 Hz, 2H), 5.40 (s, 2H), 4.41 (p, J = 7.2 Hz, 1H), 4.10 (d, J = 0.8 Hz, 2H), 1.36 (d, J = 7.3 Hz, 3H).

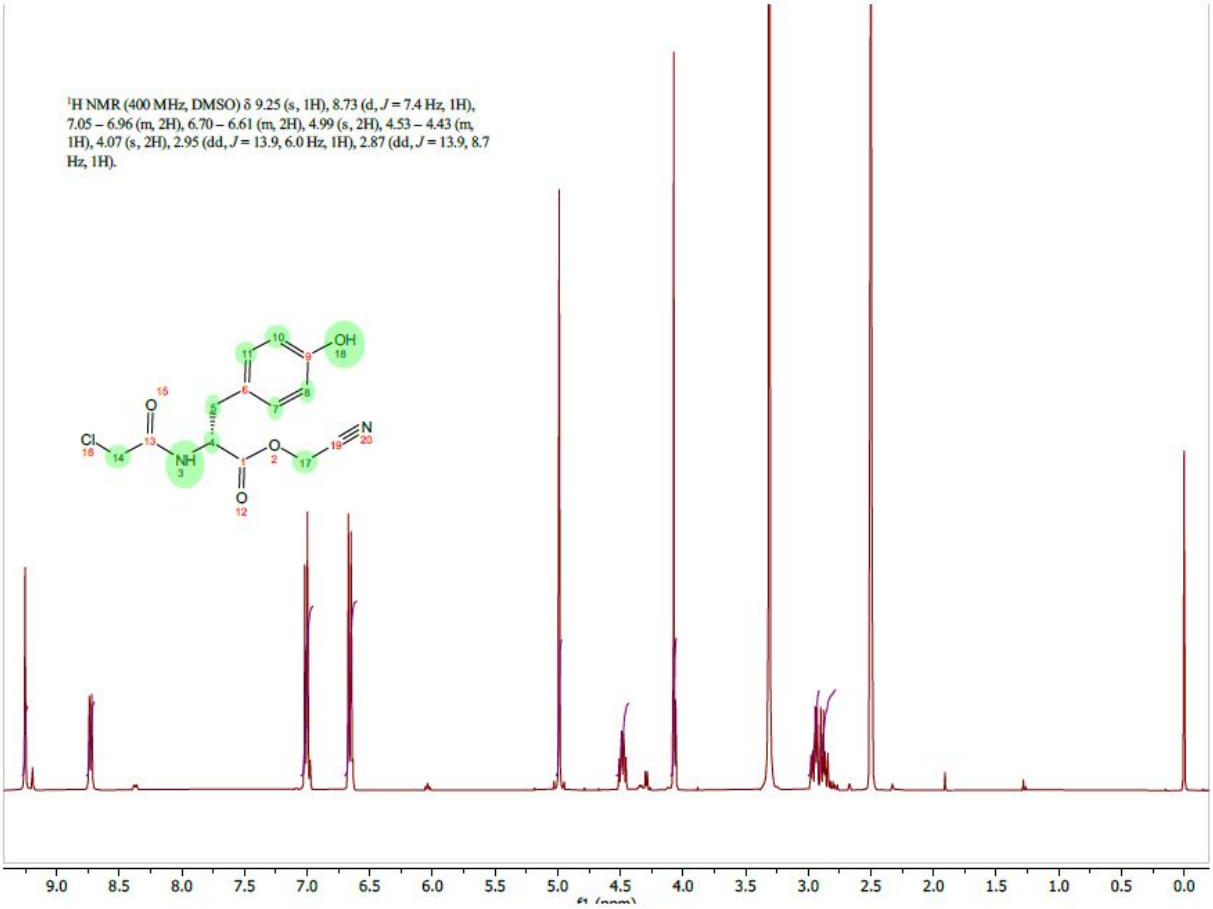
Purity 94 %



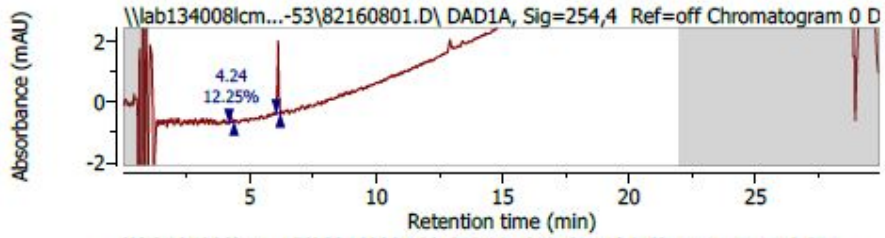
HPLC and MS spectra of (**ClAca-DNB**)



^{1}H NMR spectrum of cyanomethyl (2-chloroacetyl)-D-tyrosinate (**ClAcy-CME**)
^{1}H NMR (400 MHz, DMSO) d 9.25 (s, 1H), 8.73 (d, J = 7.4 Hz, 1H), 7.05 – 6.96 (m, 2H), 6.70 – 6.61 (m, 2H), 4.99 (s, 2H), 4.53 – 4.43 (m, 1H), 4.07 (s, 2H), 2.95 (dd, J = 13.9, 6.0 Hz, 1H), 2.87 (dd, J = 13.9, 8.7 Hz, 1H).
Purity 93 %

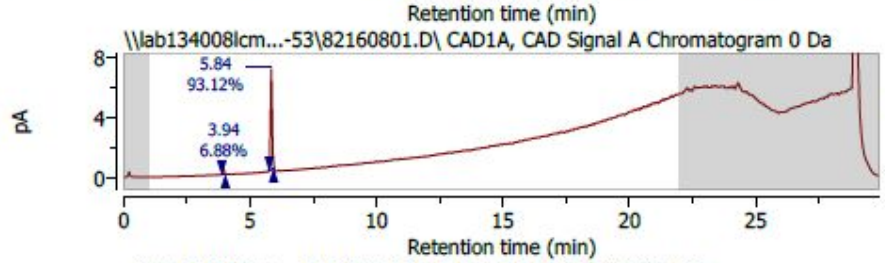


HPLC and MS spectra of **ClAcy-CME**



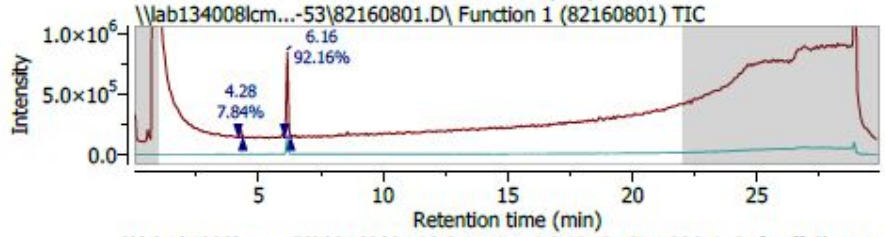
DAD1A, Sig=254,4 Ref=off (>=1.5%)

RT	Area	Area %	MS+
4.24	8.93	10.38	258.00
6.11	77.11	89.62	297.10



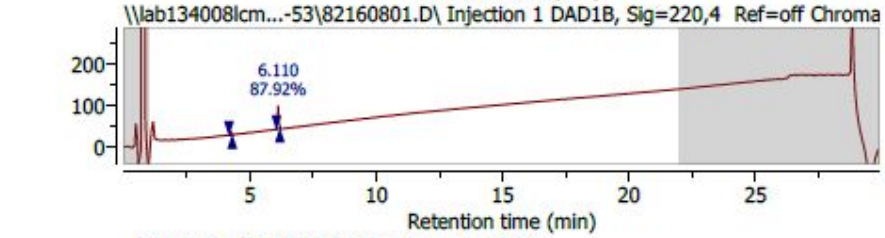
CAD CHANNEL A (>=1.5%)

RT	Area	Area %	MS+
3.94	20.43	6.88	105.10
5.84	276.45	93.12	105.10



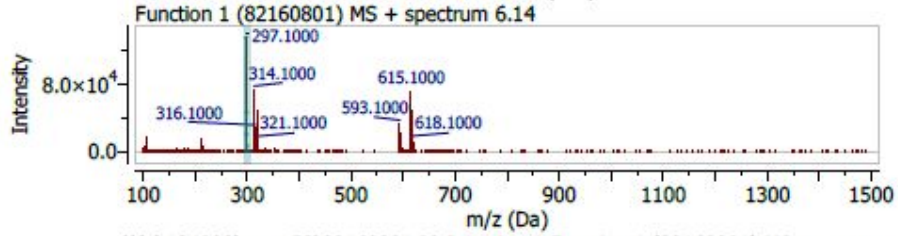
TIC (>=1.5%)

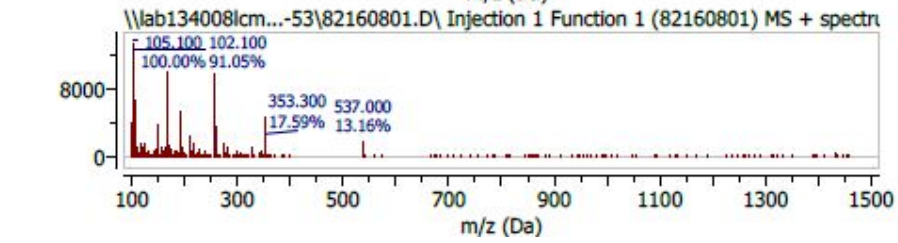
RT	Area	Area %	MS+
4.28	291100.50	7.84	258.00
6.16	3422350.52	92.16	297.10

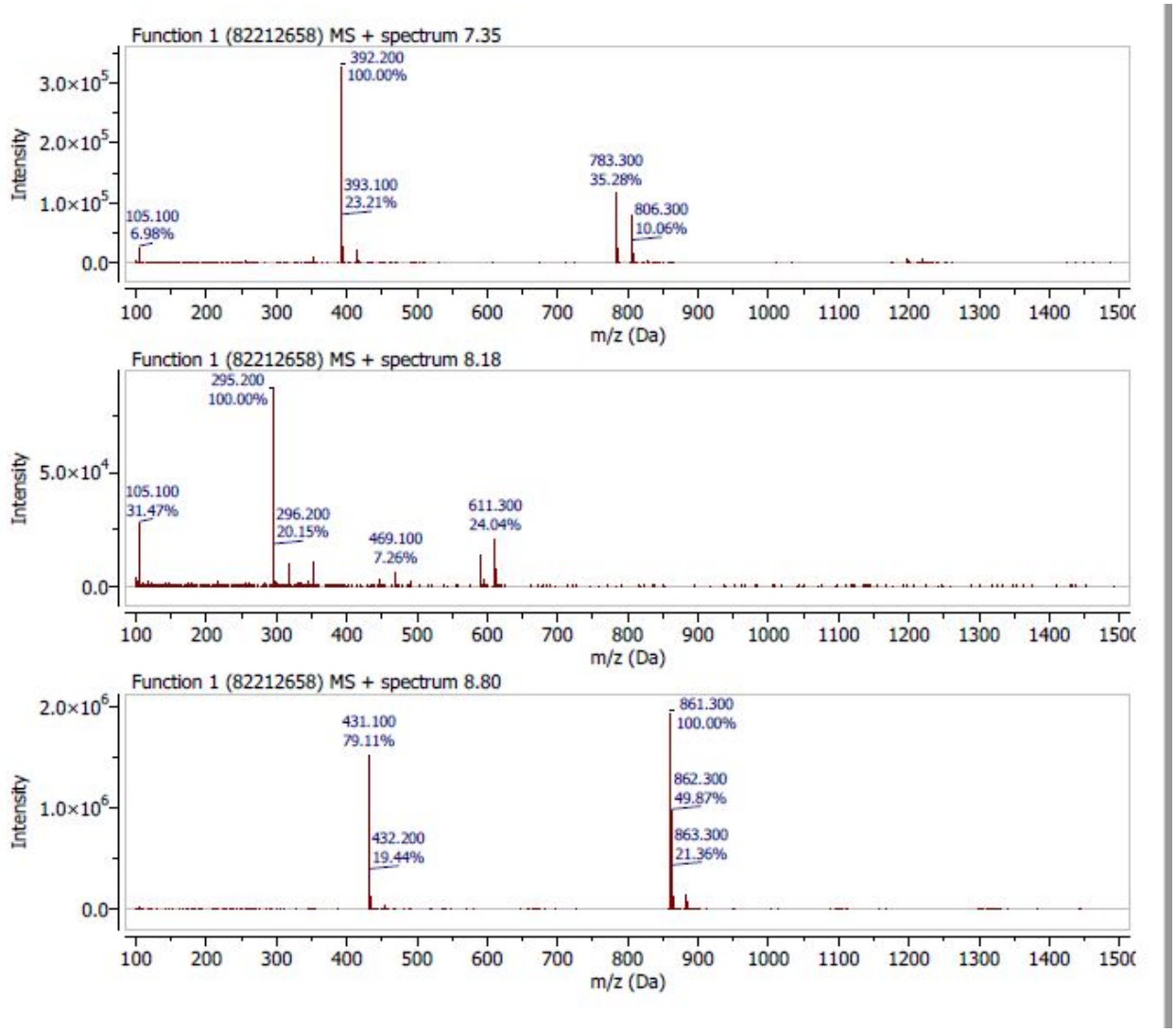


DAD1B, Sig=220,4 Ref=off (>=1.5%)

RT	Area	Area %	MS+
4.24	242.91	12.08	258.00
6.11	1768.66	87.92	297.10



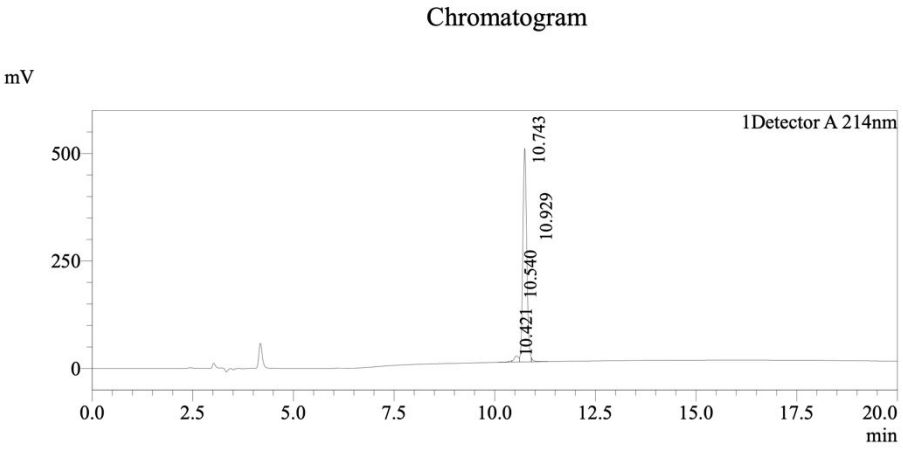



Function 1 (82212658) MS + spectrum 7.35
392.200
100.00%
393.100
23.21%
783.300
35.28%
806.300
10.06%
105.100
6.98%
Intensity
m/z (Da)
Function 1 (82212658) MS + spectrum 8.18
295.200
100.00%
105.100
31.47%
296.200
20.15%
469.100
7.26%
611.300
24.04%
Intensity
m/z (Da)
Function 1 (82212658) MS + spectrum 8.80
861.300
100.00%
431.100
79.11%
862.300
49.87%
432.200
19.44%
863.300
21.36%
Intensity
m/z (Da)

Peptide Synthesis

P1E (Ac-FGGHGGHHGGGGGSSKKSGWRLF-acid). HPLC purity 95.3%. ESI-MS $[M+2H]^{2+}$ calcd. 1155.1. found 1155.6.

Analytical HPLC spectrum of **P1E**



Peak Table

Detector A 214nm

Peak#	Ret. Time	Area	Height	Area%
1	10.421	22469	3152	0.638
2	10.540	112782	14316	3.204
3	10.743	3354567	496631	95.297
4	10.929	30306	6560	0.861
Total		3520123	520659	100.000

ESI-MS spectrum of **P1E**

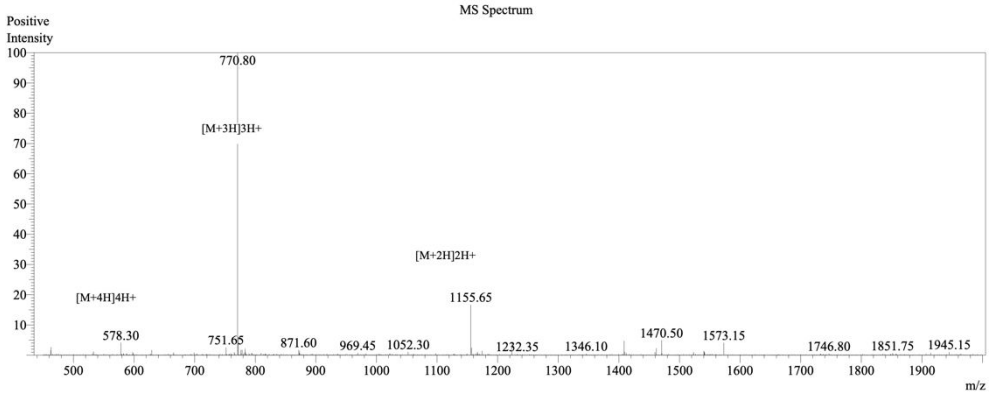


Table S5. Oligonucleotide sequences.

Name	DNA Sequence (5' → 3')
FW	CTAGTAATACGACTCACTATAGGGTTAACTTTAAGAAGGAGATATACATATG
S1A-re	CGAAGCTTAGCTAATCTTCTTGAACAGCCGGCAGCCGCTCACCTTCTTACTAGATCCACCACC TCC
S1B-re	CGAAGCTTACAGAATCTCCTCGAACAGCCGGTAGCCGGTCACCTTCTTACTAGATCCACCACC TCC
S1C-re	CGAAGCTTACTTCTTGAACAGCCGGCAGCCGCTCTTCTTACTAGATCCACCACCTCC
S1D-re	CGAAGCTTAGCTAATCTTCTTGAACAGCCGCTTCTTACTAGATCCACCACCTCC
S1E-re	CGAAGCTTAGAACAGCCGGCAGCCGCTCTTCTTACTAGATCCACCACCTCC
FLAG re	CGAAGCTTACTTGTCGTCGTCGTCCTTGTAGTCCTTCTTACTAGATCCACCACCTCC
S1-temp	TAAGAAGGAGATATACATATGGGTGGCCACGGTGGCCATCACGGCGGAGGTGGTGGATCT AGTAAGAAG
S2-temp	TAAGAAGGAGATATACATATGGCCGGCCACGGTGGCCATCACGGCGGAGGTGGTGGATCT AGTAAGAAG
S3-temp	TAAGAAGGAGATATACATATGGGTGCCCACGGTGGCCATCACGGCGGAGGTGGTGGATCT AGTAAGAAG
S4-temp	TAAGAAGGAGATATACATATGGGTGGCGCCGGTGGCCATCACGGCGGAGGTGGTGGATCT AGTAAGAAG
S5-temp	TAAGAAGGAGATATACATATGGGTGGCCACGCCGGCCATCACGGCGGAGGTGGTGGATCT AGTAAGAAG
S6-temp	TAAGAAGGAGATATACATATGGGTGGCCACGGTGCCCATCACGGCGGAGGTGGTGGATCT AGTAAGAAG
S7-temp	TAAGAAGGAGATATACATATGGGTGGCCACGGTGGCGCCCACGGCGGAGGTGGTGGATCT AGTAAGAAG
S8-temp	TAAGAAGGAGATATACATATGGGTGGCCACGGTGGCCATGCCGGCGGAGGTGGTGGATCT AGTAAGAAG
S9-temp	TAAGAAGGAGATATACATATGGGTGGCCACGGTGGCCATCACGCCGGAGGTGGTGGATCT AGTAAGAAG
L8-temp	TAAGAAGGAGATATACATATGNNTNNTNNTNNTNNTNNTTGGGGAGGTGGTGGAAGTAG C
L14-temp	TAAGAAGGAGATATACATATGNNTNNTNNTNNTNNTNNTNNTNNTNNTNNTNNTTG GGGAGGTGGTGGAAGTAGC
Lib HiBit217 re	CCCGCCTCCCGCCCCCCGTCCTAGAACAGCCGGCAGCCGCTCTTCTTGCTACTTCCACCACCT CCCCA