Supplementary Information

SARS-CoV-2 M^{pro} inhibitors and activity-based probes for patient-sample imaging

Wioletta Rut[1*], Katarzyna Groborz[1,9], Linlin Zhang[2,3,4,9], Xinyuanyuan Sun[2,3,4], Mikolaj Zmudzinski[1], Bartlomiej Pawlik[6,7], Xinyu Wang[8], Dirk Jochmans[8], Johan Neyts[8], Wojciech Młynarski[6], Rolf Hilgenfeld[2,4,5], Marcin Drag[1*]

[1]Department of Chemical Biology and Bioimaging, Wroclaw University of Science and Technology, Wyb. Wyspianskiego 27, 50-370 Wroclaw, Poland

[2]Institute of Molecular Medicine, University of Lübeck, Ratzeburger Allee 160, 23562 Lübeck, Germany

[3]Institute of Biochemistry, Center for Structural and Cell Biology in Medicine, University of Lübeck, Ratzeburger Allee 160, 23562 Lübeck, Germany

[4]German Center for Infection Research (DZIF), Hamburg-Lübeck-Borstel-Riems Site, University of Lübeck, 23562 Lübeck, Germany

[5]Center for Brain, Behavior, and Metabolism (CBBM), University of Lübeck, Lübeck, Germany

[6]Department of Pediatrics, Oncology & Hematology, Medical University of Lodz, Sporna 36/50, 91-738 Lodz, Poland

[7]Postgraduate School of Molecular Medicine, Medical University of Warsaw, Trojdena 2a, 02-091 Warsaw, Poland

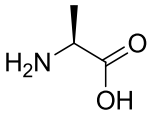
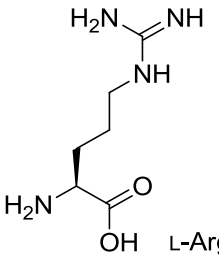
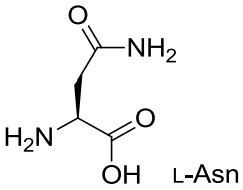
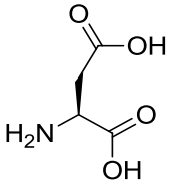
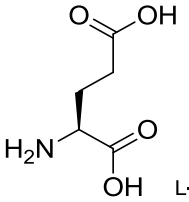
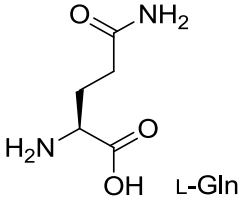
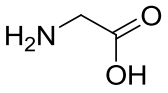
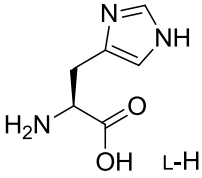
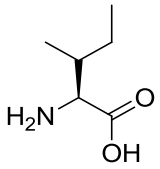
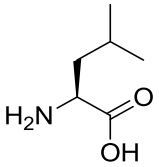
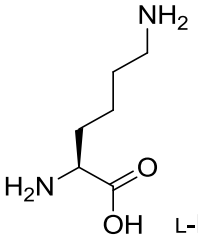
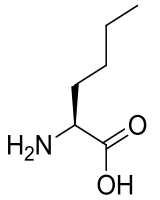
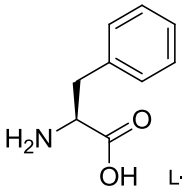
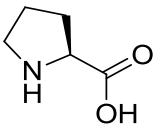
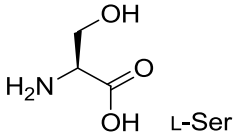
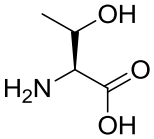
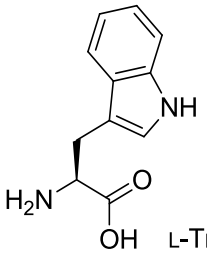
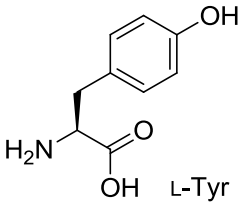
[8]Laboratory of Virology and Chemotherapy, Department of Microbiology, Immunology and Transplantation, Rega Institute, KU Leuven, 3000, Leuven, Belgium

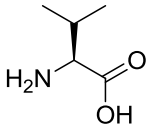
[9]Equal contribution

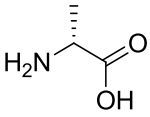
- corresponding authors: wioletta.rut@pwr.edu.pl, marcin.drag@pwr.edu.pl

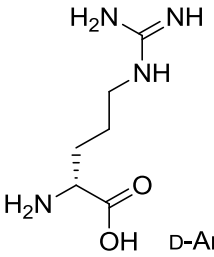
Content of Supporting Information

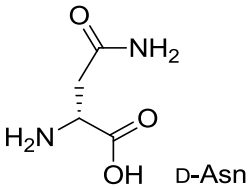
Supplementary Table 1. Structures of fixed natural and unnatural amino acids used in combinatorial library.

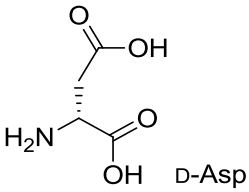
No	Structure + code	No	Structure + code	No	Structure + code
1	 L-Ala	2	 L-Arg	3	 L-Asn
4	 L-Asp	5	 L-Glu	6	 L-Gln
7	 Gly	8	 L-His	9	 L-Ile
10	 L-Leu	11	 L-Lys	12	 L-Nle
13	 L-Phe	14	 L-Pro	15	 L-Ser
16	 L-Thr	17	 L-Trp	18	 L-Tyr

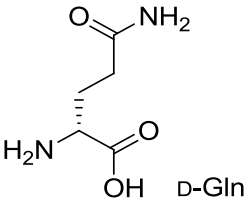
19  L-Val

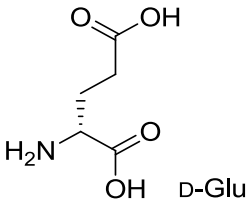
20  D-Ala

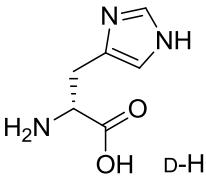
21  D-Arg

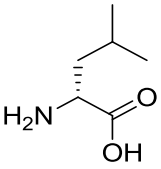
22  D-Asn

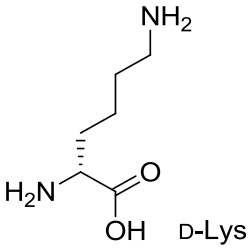
23  D-Asp

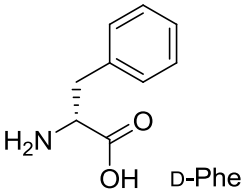
24  D-Gln

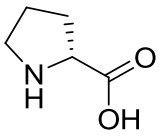
25  D-Glu

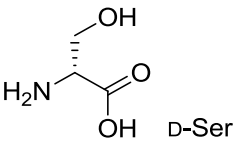
26  D-His

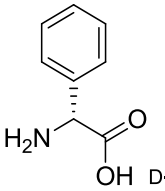
27  D-Leu

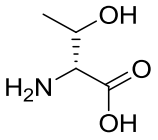
28  D-Lys

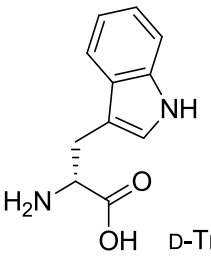
29  D-Phe

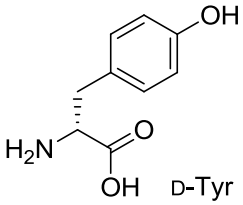
30  D-Pro

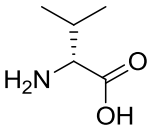
31  D-Ser

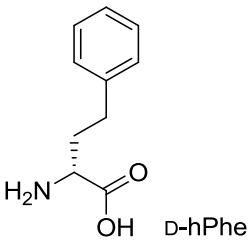
32  D-Phg

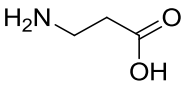
33  D-Thr

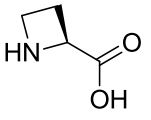
34  D-Trp

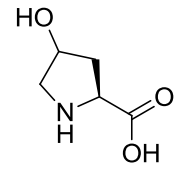
35  D-Tyr

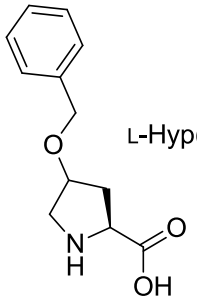
36  D-Val

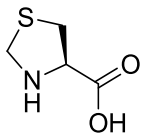
37  D-hPhe

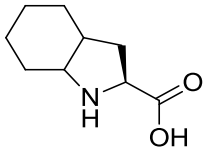
38  β–Ala

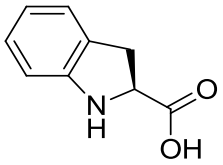
39  L-Aze

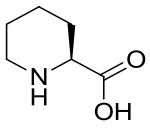
40  L-Hyp

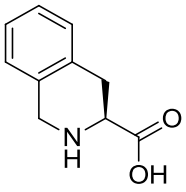
41  L-Hyp(Bzl)

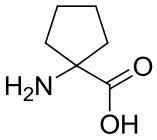
42  L-Thz

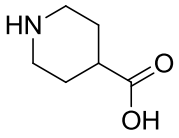
43  L-Oic

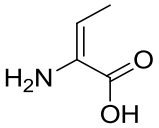
44  L-Idc

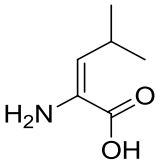
45  L-Pip

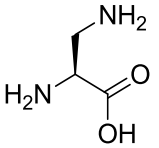
46  L-Tic

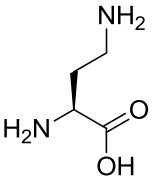
47  AC5C

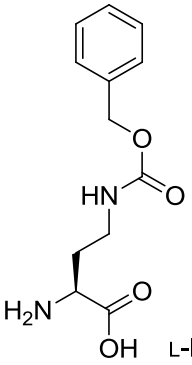
48  Inp

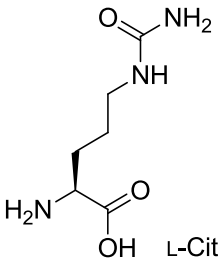
49  dhAbu

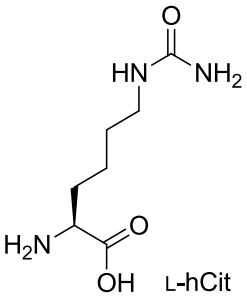
50  dhLeu

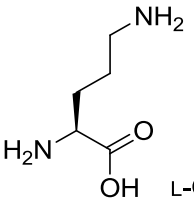
51  L-Dap

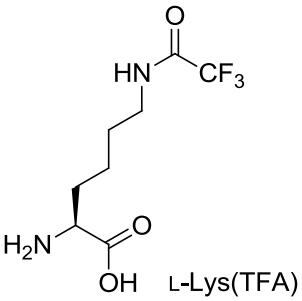
52  L-Dab

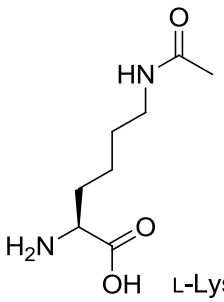
53  L-Dab(Z)

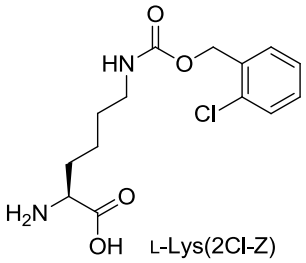
54  L-Cit

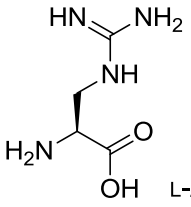
55  L-hCit

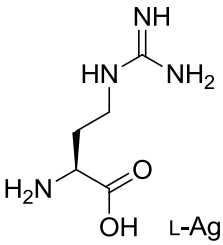
56  L-Orn

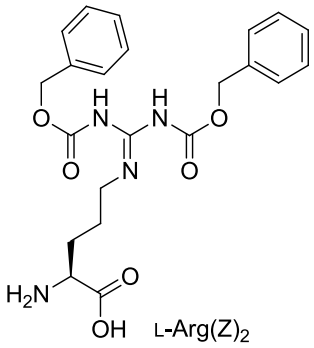
57  L-Lys(TFA)

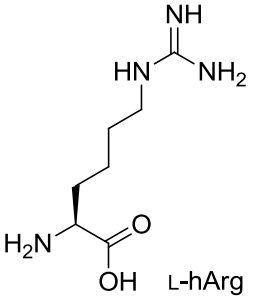
58 
L-Lys(Ac)

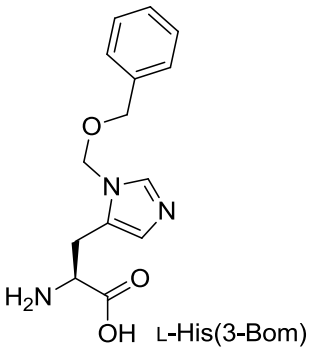
59 
L-Lys(2Cl-Z)

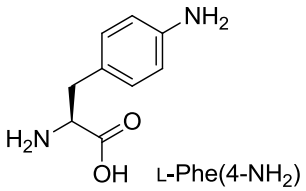
60 
L-Agp

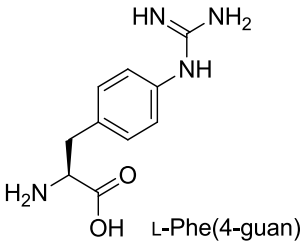
61 
L-Agb

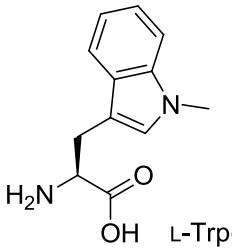
62 
L-Arg(Z)$_2$

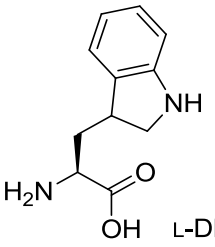
63 
L-hArg

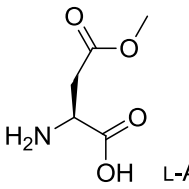
64 
L-His(3-Bom)

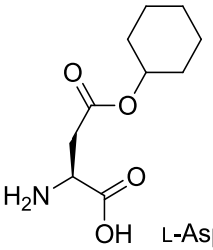
65 
L-Phe(4-NH$_2$)

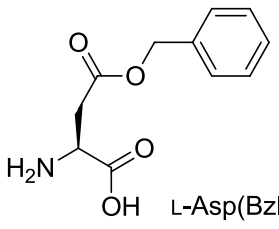
66 
L-Phe(4-guan)

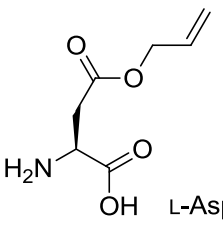
67 
L-Trp(Me)

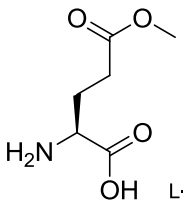
68 
L-Dht

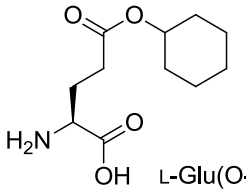
69 
L-Asp(O-Me)

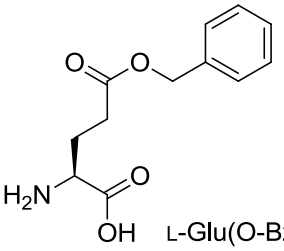
70 
L-Asp(O-cHx)

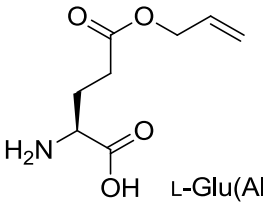
71 
L-Asp(Bzl)

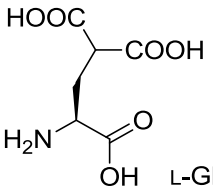
72 
L-Asp(All)

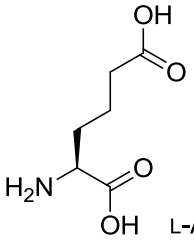
73  L-Glu(O-Me)

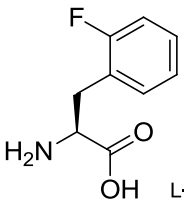
74  L-Glu(O-cHx)

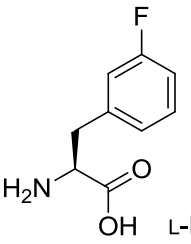
75  L-Glu(O-Bzl)

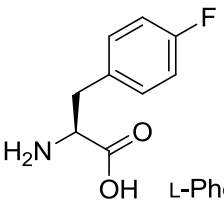
76  L-Glu(All)

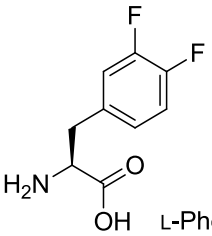
77  L-Gla

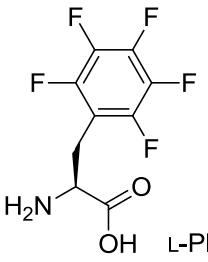
78  L-Aad

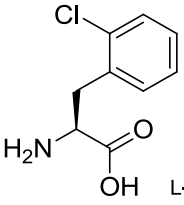
79  L-Phe(2-F)

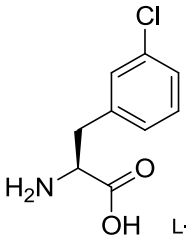
80  L-Phe(3-F)

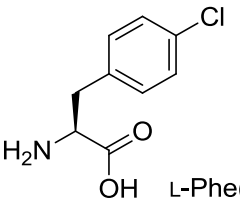
81  L-Phe(4-F)

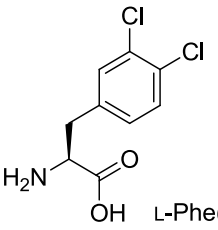
82  L-Phe(3,4-F_2)

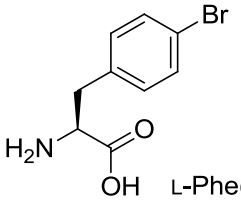
83  L-Phe(F_5)

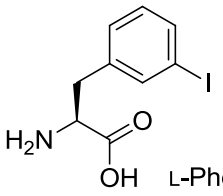
84  L-Phe(2-Cl)

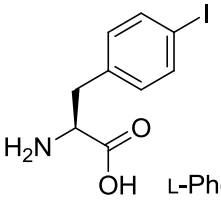
85  L-Phe(3-Cl)

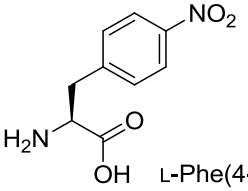
86  L-Phe(4-Cl)

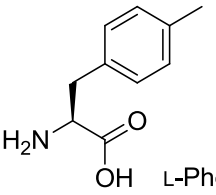
87  L-Phe(3,4-Cl_2)

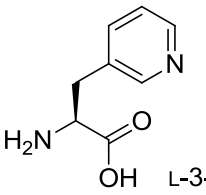
88  L-Phe(4-Br)

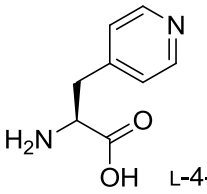
89  L-Phe(3-I)

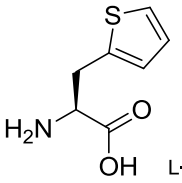
90  L-Phe(4-I)

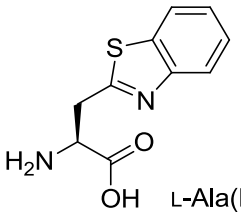
91  L-Phe(4-NO_2)

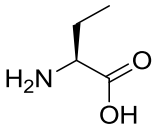
92  L-Phe(4-Me)

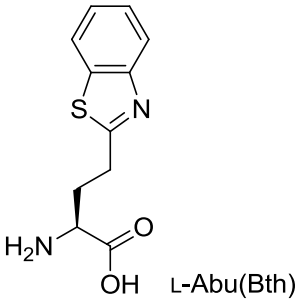
93  L-3-Pal

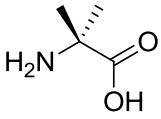
94  L-4-Pal

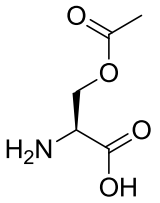
95  L-Ala(2-th)

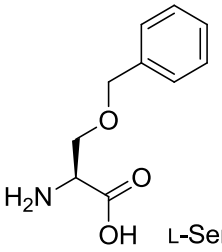
96  L-Ala(Bth)

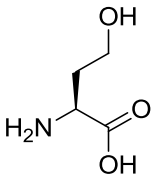
97  L-Abu

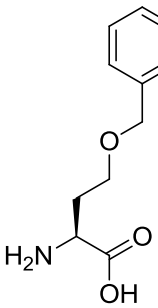
98  L-Abu(Bth)

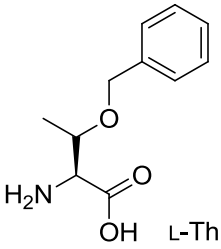
99  Aib

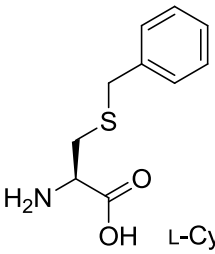
100  L-Ser(Ac)

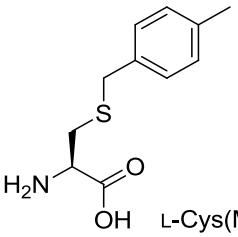
101  L-Ser(Bzl)

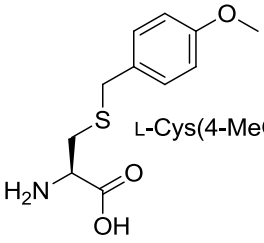
102  L-hSer

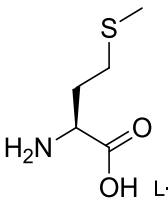
103  L-hSer(Bzl)

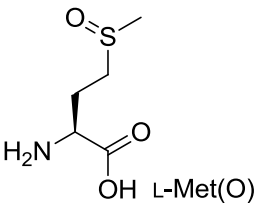
104  L-Thr(Bzl)

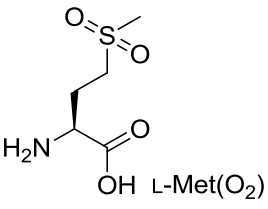
105  L-Cys(Bzl)

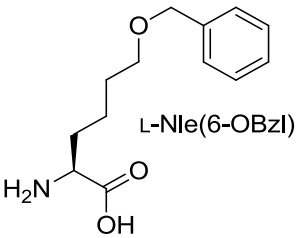
106  L-Cys(MeBzl)

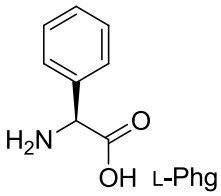
107  L-Cys(4-MeOBzl)

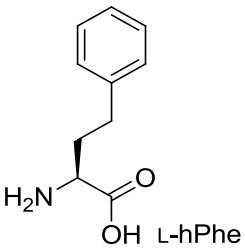
108  L-Met

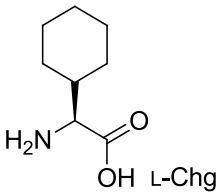
109  L-Met(O)

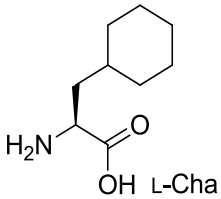
110  L-Met(O_2)

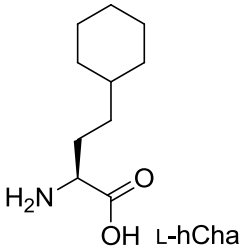
111  L-Nle(6-OBzl)

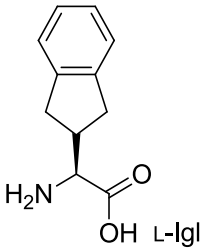
112 

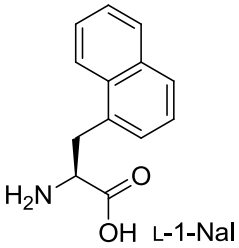
113 

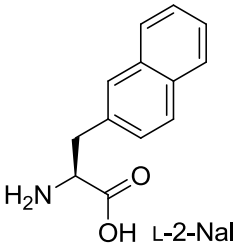
114 

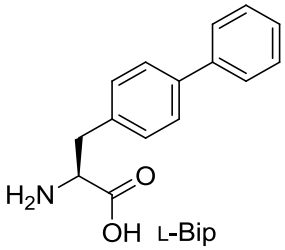
115 

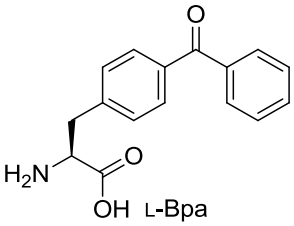
116 

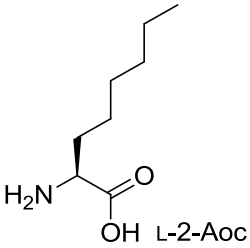
117 

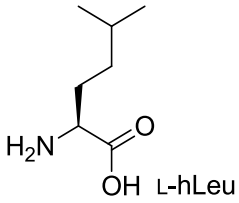
118 

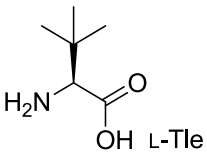
119 

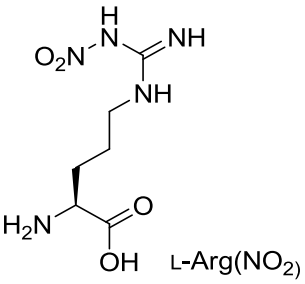
120 

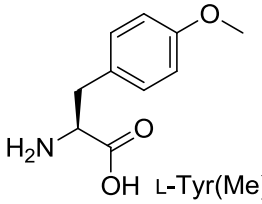
121 

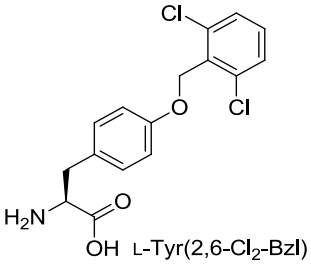
122 

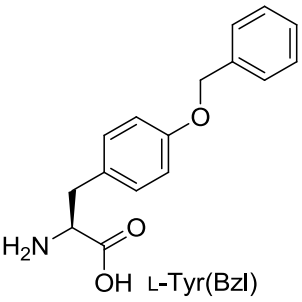
123 

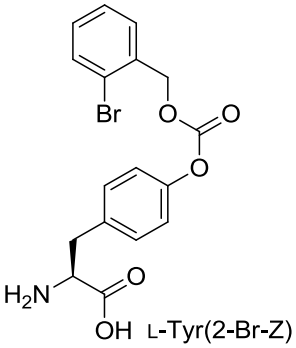
124 

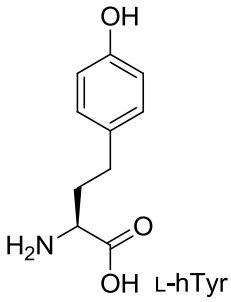
125 

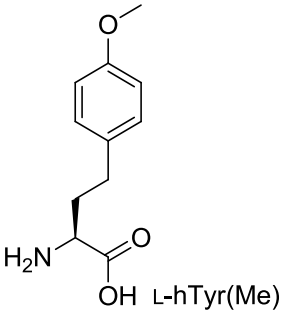
126 

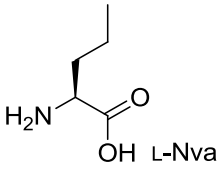
127 

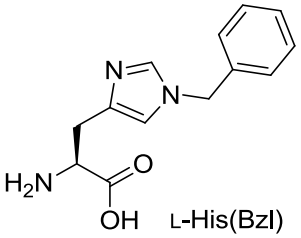
128 

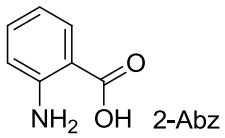
129 

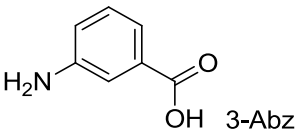
130 

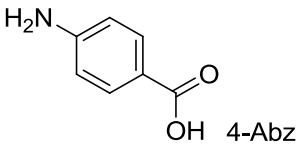
131 

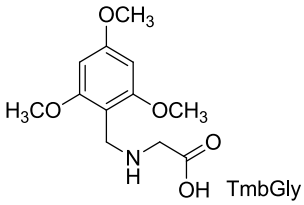
132 

133 

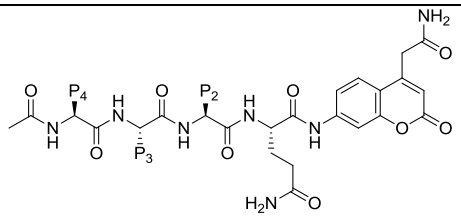
134 

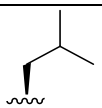
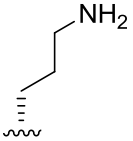
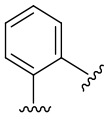
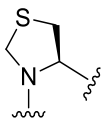
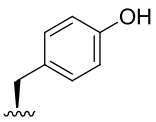
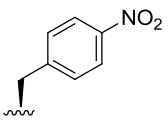
135 

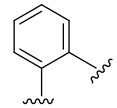
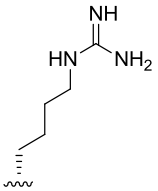
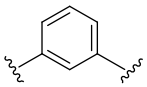
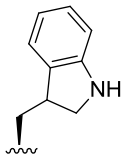
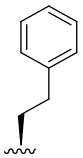
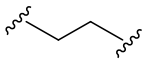
136 

137 

Supplementary Table 2. Structures of the most recognizable amino acids at each tested position by SARS-CoV-2 M^{pro}.



P4		**P3**		**P2**	
structure	Relative activity, %	structure	Relative activity, %	structure	Relative activity, %
	100		100		100
	77.5		83		54
	77		82		50

	73		78		50
	72		70		49

Supplementary Table 3. Specificity constant value for internally quenched fluorescent substrates toward SARS-CoV-2 M^{pro} (n=3, where n represents the number of independent experiments).

Substrate	k_{cat}/K_M, $M^{-1}s^{-1}$
ACC-Gly-Abu-Tle-Leu-Gln-Ser-Gly-Phe-Arg-Lys(dnp)-Lys-NH_2 **(9)**	14748 ± 684
ACC-Gly-Thz-Tle-Leu-Gln-Ser-Gly-Phe-Arg-Lys(dnp)-Lys-NH_2 **(10)**	19424 ± 1176
ACC-Gly-Val-Lys-Leu-Gln-Ser-Gly-Phe-Arg-Lys(dnp)-Lys-NH_2 **(11)**	6755 ± 208

Supplementary Table 4. Data collection and refinement statistics.

	SARS-CoV-2 M^{pro} in complex with Biotin-PEG(4)-Abu-Tle-Leu-Gln-VS
Data collection	
Space group	$P6_122$
Cell dimensions	
a, *b*, *c* (Å)	104.08, 104.08, 91.25
α, β, γ (°)	90.0, 90.0, 120.0
Resolution (Å)	45.62 - 1.70 (1.79 - 1.70)*
R_{sym} or R_{merge}	0.074 (1.574)
$I / \sigma I$	32.4 (2.8)
Completeness (%)	100.0 (100.0)
Redundancy	38.5 (38.0)
Refinement	
Resolution (Å)	1.70
No. reflections	30969
R_{work} / R_{free}	0.192/0.243
No. atoms	
Protein	2369
Ligand/ion	70
Water	232
B-factors	

Protein	34
Ligand/ion	57
Water	43
R.m.s. deviations	
Bond lengths (Å)	0.01
Bond angles (°)	1.8

*Values in parentheses are for highest-resolution shell.

Supplementary Note

Synthetic procedures

Individual substrate synthesis

ACC-labeled substrates were synthesized on the solid support according to the solid phase peptide synthesis method described elsewhere[1]. In brief, Fmoc-ACC-OH (2.5 eq) was attached to a Rink-amide resin using HOBt (2.5 eq) and DICI (2.5 eq) in DMF as coupling reagents. Then, the Fmoc protecting group was removed using 20% piperidine in DMF (three cycles: 5, 5, and 25 min). Fmoc-Gln(Trt)-OH (2.5 eq) was coupled to the H_2N-ACC-resin using HATU (2.5 eq) and 2,4,6-collidine (2.5 eq) in DMF. After Fmoc group removal, the Fmoc-P2-OH (2.5 eq) amino acid was attached (HOBt and DICI (2.5 eq) in DMF). Amino acids in P3 and P4 positions were coupled in the same manner. The free *N*-terminus was acetylated using HBTU, AcOH and DIPEA in DMF (5 eq of each reagent). Then, the resin was washed five times with DMF, three times with DCM and three times with MeOH, and dried over P_2O_5. Substrates were removed from the resin with a mixture of TFA/TIPS/H_2O (% v/v/v, 95:2.5:2.5), precipitated in Et_2O, purified on HPLC and lyophilized. The purity of each substrate was confirmed using analytical HPLC. Each substrate was dissolved in DMSO at a final concentration of 10 mM and stored at -80°C until use.

Internally quenched fluorescence substrates were synthesized on the solid support in the same manner as described previously[2].

Activity-based probe and inhibitor synthesis

The synthesis of the biotinylated activity-based probe involved three sequential steps. In the first step, the Biotin-PEG(4)-Abu-Tle-Leu-OH fragment was synthesized using a 2-chlorotrityl chloride resin. Fmoc-Leu-OH (2.5 eq) was dissolved in anhydrous DCM, pre-activated with DIPEA (3 eq) and added to the cartridge with resin (1 eq). After 3h, the mixture was filtered, washed 3 times with DCM and 3 times with DMF, and the Fmoc group was removed using 20% piperidine in DMF. Fmoc-Tle-OH, Fmoc-Abu-OH and Fmoc-PEG(4)-OH were attached

to the resin using HATU (2.5 eq) and 2,4,6-collidine (2.5 eq) in DMF as coupling reagents. The biotin tag was coupled to H_2N-PEG(4)-Abu-Tle-Leu-resin using 2.5 eq HBTU and 2.5 eq DIPEA in a DMF:DMSO mixture (1:1, v/v). After 3 h, the resin was washed 3 times with DMF, 3 times with DCM, and 3 times with MeOH and dried over P_2O_5. Finally, the peptide fragment was removed from the resin with a mixture of DCM/TFE/AcOH (v/v/v, 8:1:1). The solution was filtered and concentrated. The obtained crude peptide was dissolved in ACN:H_2O (v/v, 3:1), lyophilized, and then used without further purification (purity > 95%). In the second step, Fmoc-Gln(Trt)-VS was synthesized in the same manner as described elsewhere[3]. In the third step, the Fmoc group was removed from Fmoc-Gln(Trt)-VS using a mixture of diethylamine:ACN (1:1, v/v). Biotin-PEG(4)-Abu-Tle-Leu-OH was pre-activated with HATU (1.2 eq) and 2,4,6-collidine (3 eq) in DMF and attached to the H_2N-Gln(Trt)-VS (1 eq). The reaction was agitated for 2 h and the product was purified on HPLC. Finally, Biotin-PEG(4)-Abu-Tle-Leu-Gln(Trt)-VS was treated with a mixture of TFA/DCM/TIPS (% v/v/v/ 70:27:3) to remove the Trt group. After 40 min, solvents were evaporated and the probe was purified on HPLC.

Ac-Abu-Tle-Leu-Gln-VS, Ac-Abu-DTyr-Leu-Gln-VS, Bodipy-PEG(4)-Abu-DTyr-Leu-Gln-VS, and Cy5-PEG(4)-Abu-Tle-Leu-Gln-VS were synthesized in the same manner, but Boc-PEG(4)-Abu-Tle-Leu-OH, Boc-PEG(4)-Abu-DTyr-Leu-OH, Ac-Abu-DTyr-Leu-OH, and Ac-Abu-Tle-Leu-OH were synthesized on the 2-chlorotrityl chloride resin.

References

1. Poreba, M., Salvesen, G.S. & Drag, M. Synthesis of a HyCoSuL peptide substrate library to dissect protease substrate specificity. *Nature Protocols* **12**, 2189-2214 (2017).

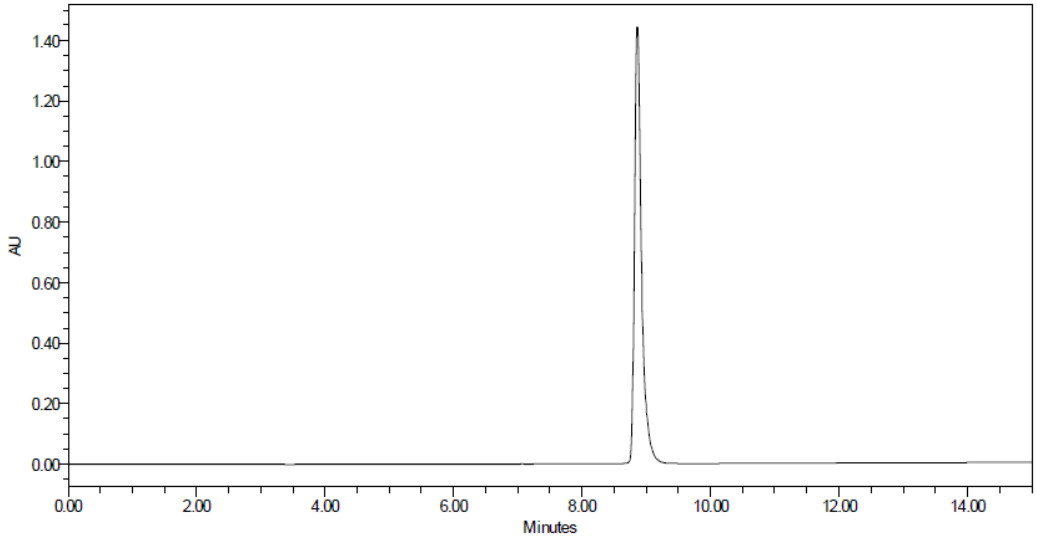
2. Poreba, M. et al. Highly sensitive and adaptable fluorescence-quenched pair discloses the substrate specificity profiles in diverse protease families. *Scientific Reports* **7** DOI: 10.1038/srep43135 (2017).

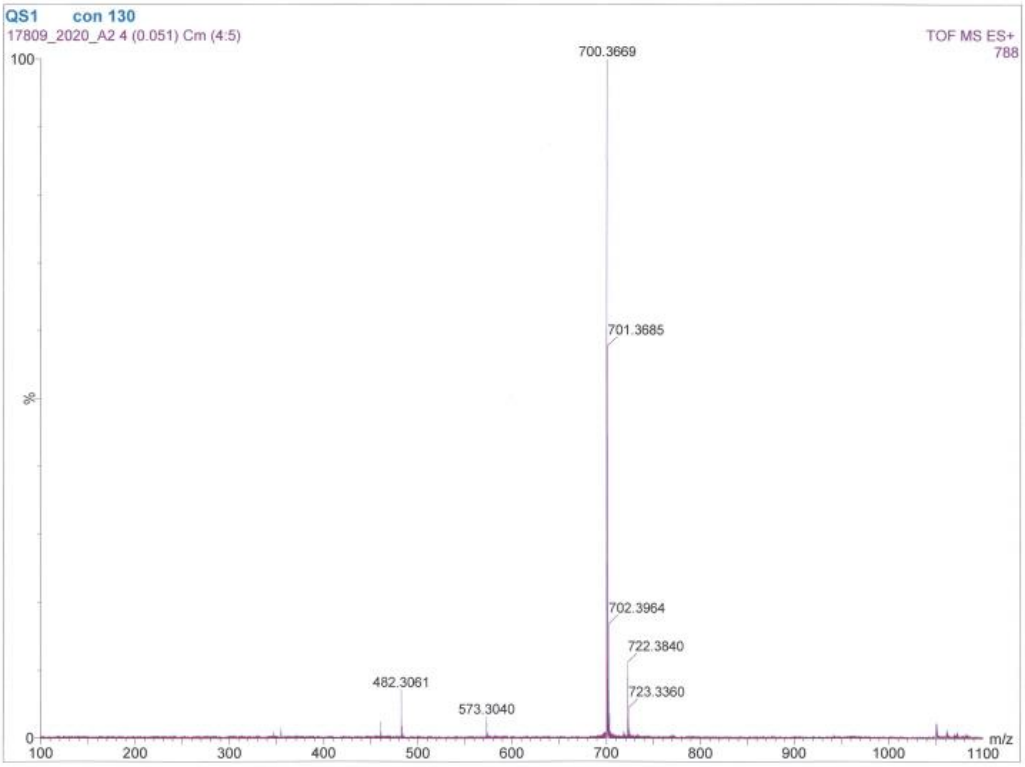
3. Rut, W., Poreba, M., Kasperkiewicz, P., Snipas, S.J. & Drag, M. Selective Substrates and Activity-Based Probes for Imaging of the Human Constitutive 20S Proteasome in Cells and Blood Samples. *Journal of Medicinal Chemistry* **61**, 5222-5234 (2018).

HRMS and analytical chromatograms of synthesized substrates

1 Ac-Abu-Tle-Leu-Gln-ACC

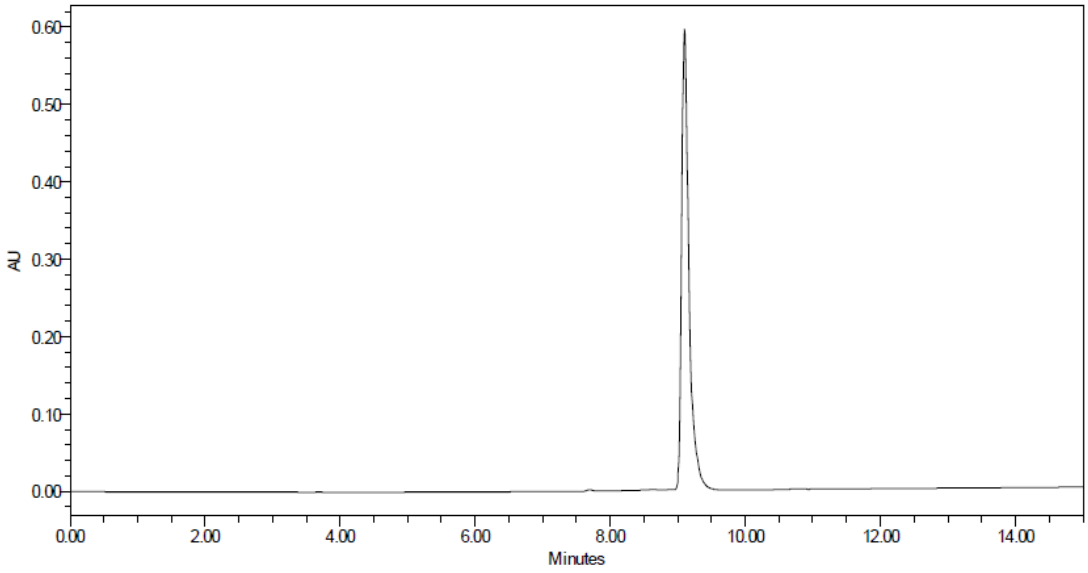
m/z: $[M+H]^+$ for $C_{34}H_{49}N_7O_9$ (m/z_{calcd} = 700.3665; m/z_{found} = 700.3669)

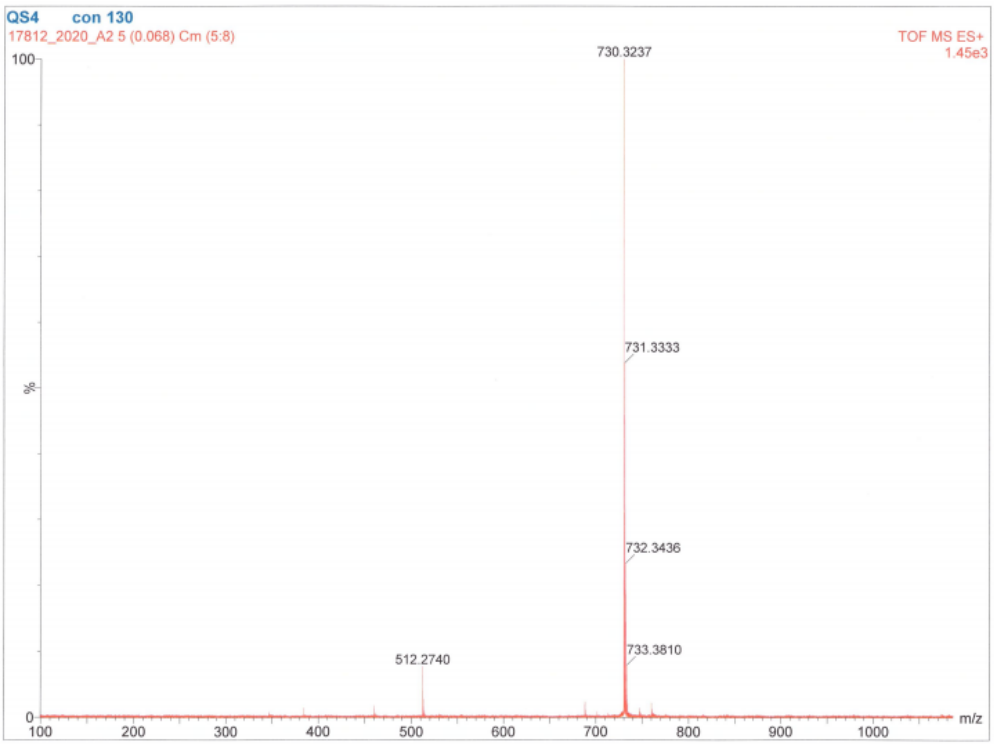




2 Ac-Thz-Tle-Leu-Gln-ACC

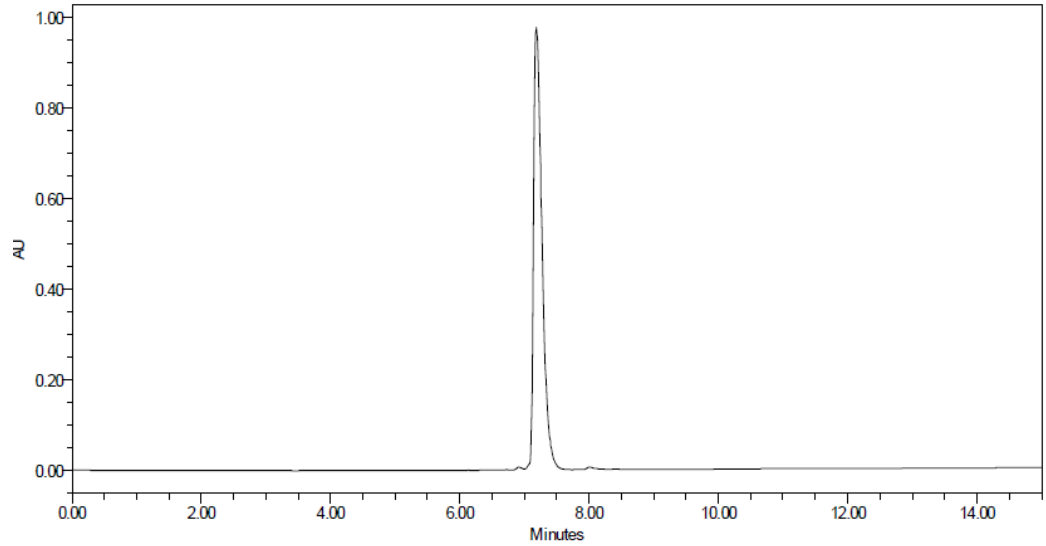
m/z: $[M+H]^+$ for $C_{34}H_{47}N_7O_9S$ (m/z_{calcd} = 730.3229; m/z_{found} = 730.3237)

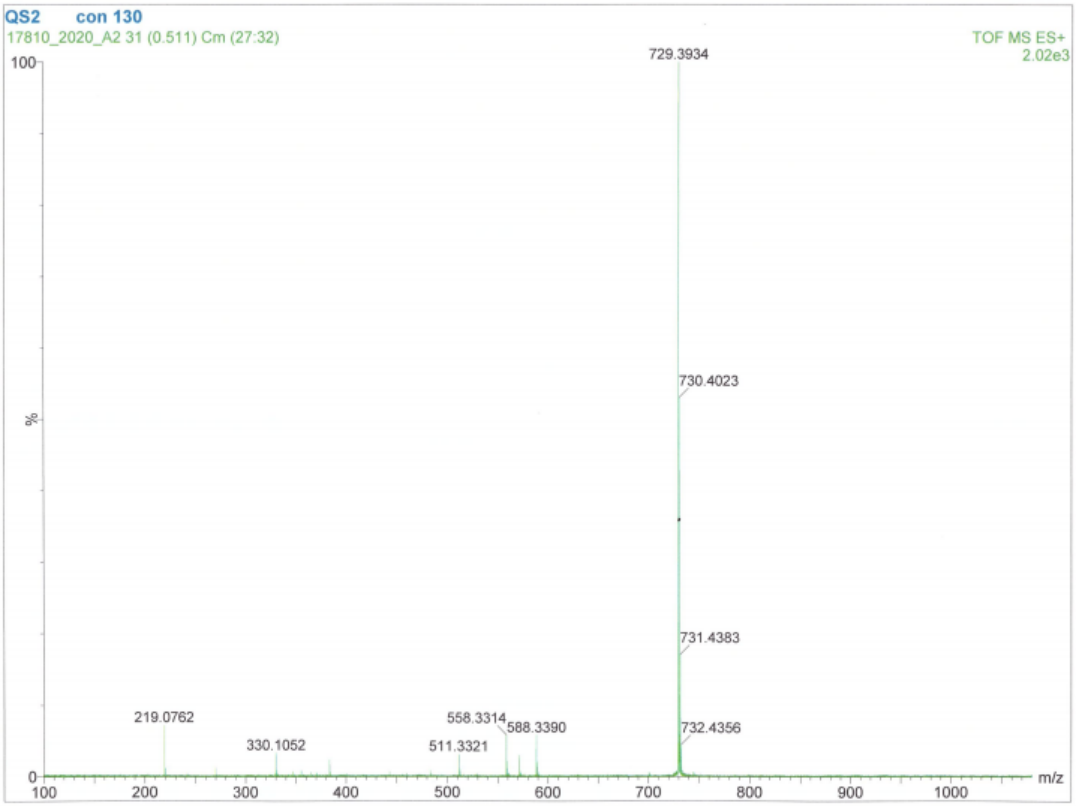




3 Ac-Val-Lys-Leu-Gln-ACC

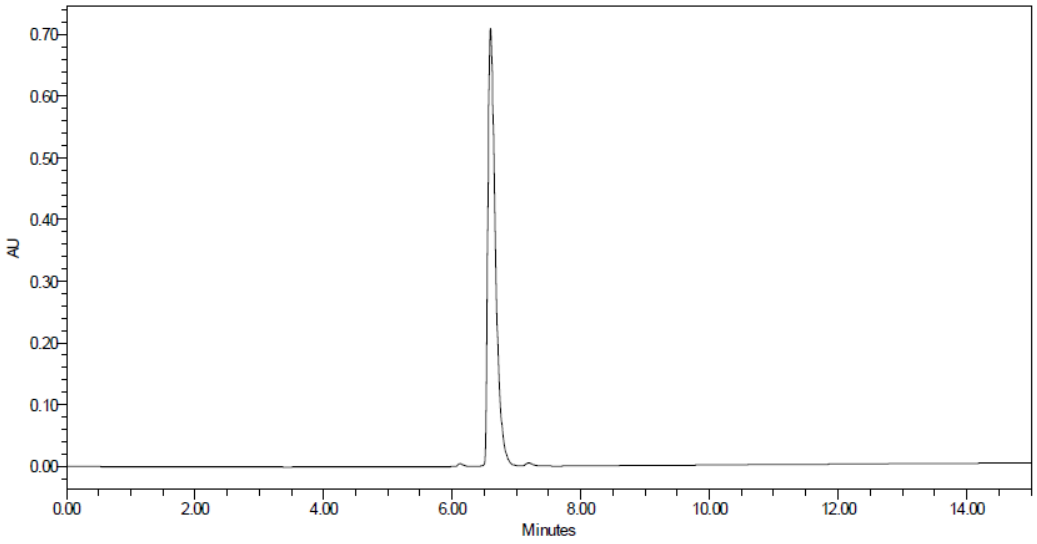
m/z: $[M+H]^+$ for $C_{35}H_{52}N_8O_9$ (m/z_{calcd} = 729.3931; m/z_{found} = 729.3934)

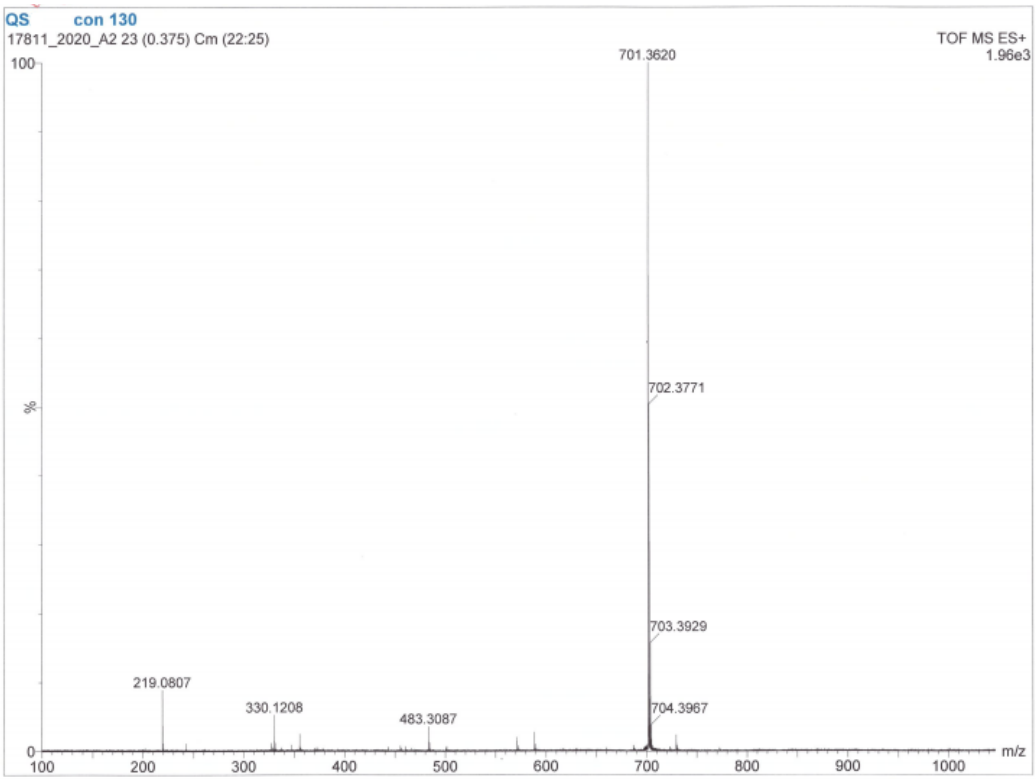




4 Ac-Ala-Lys-Leu-Gln-ACC

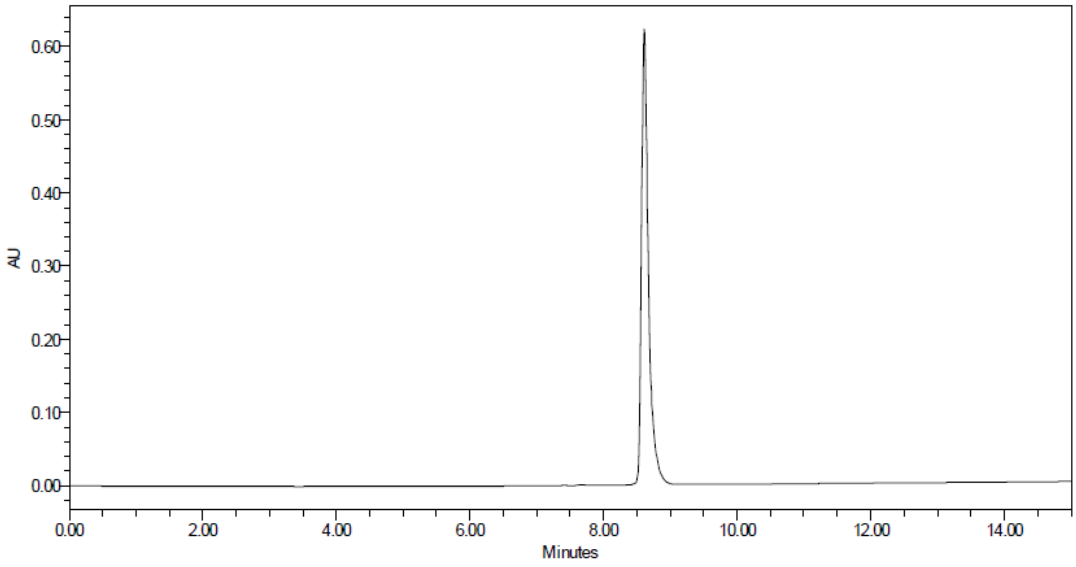
m/z: $[M+H]^+$ for $C_{33}H_{48}N_8O_9$ (m/z_{calcd} = 701.3618; m/z_{found} = 701.3620)

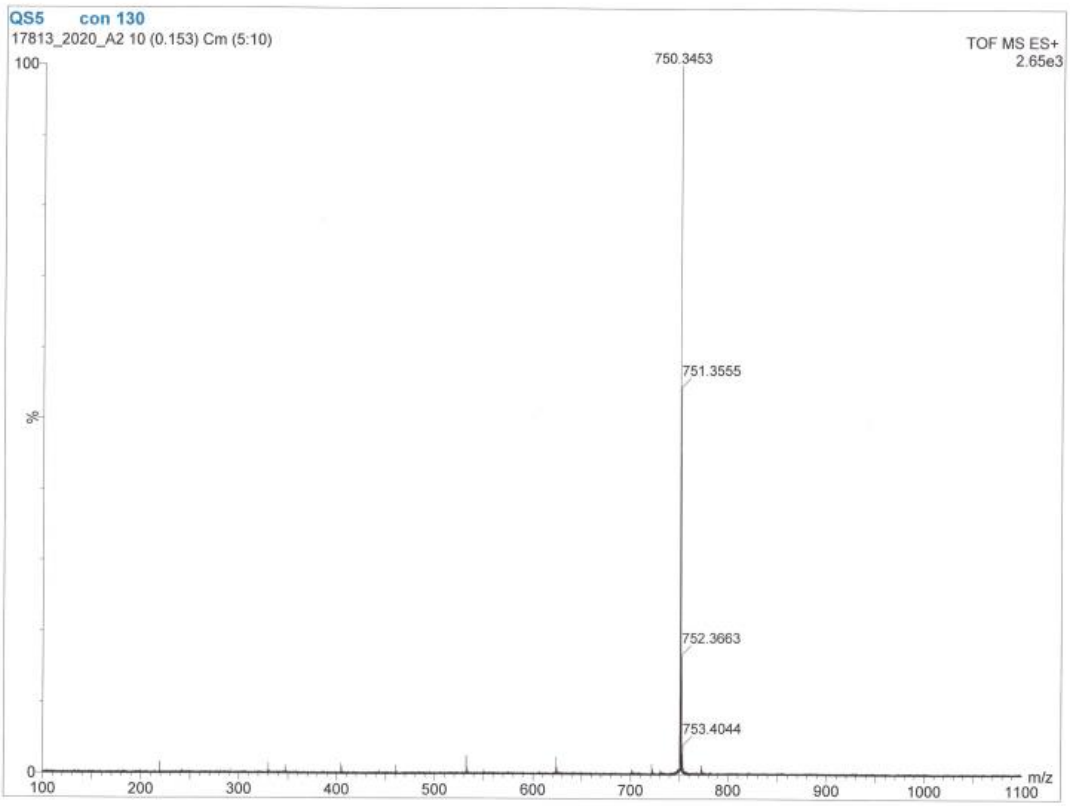




5 Ac-Abu-DTyr-Leu-Gln-ACC

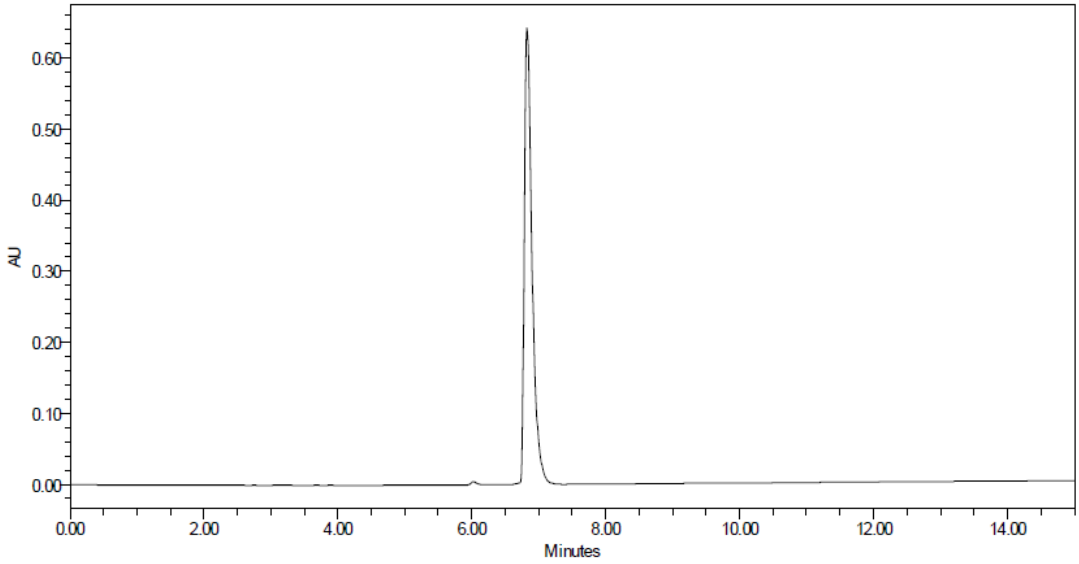
m/z: $[M+H]^+$ for $C_{37}H_{47}N_7O_{10}$ (m/z_{calcd} = 750.3458; m/z_{found} = 750.3453)

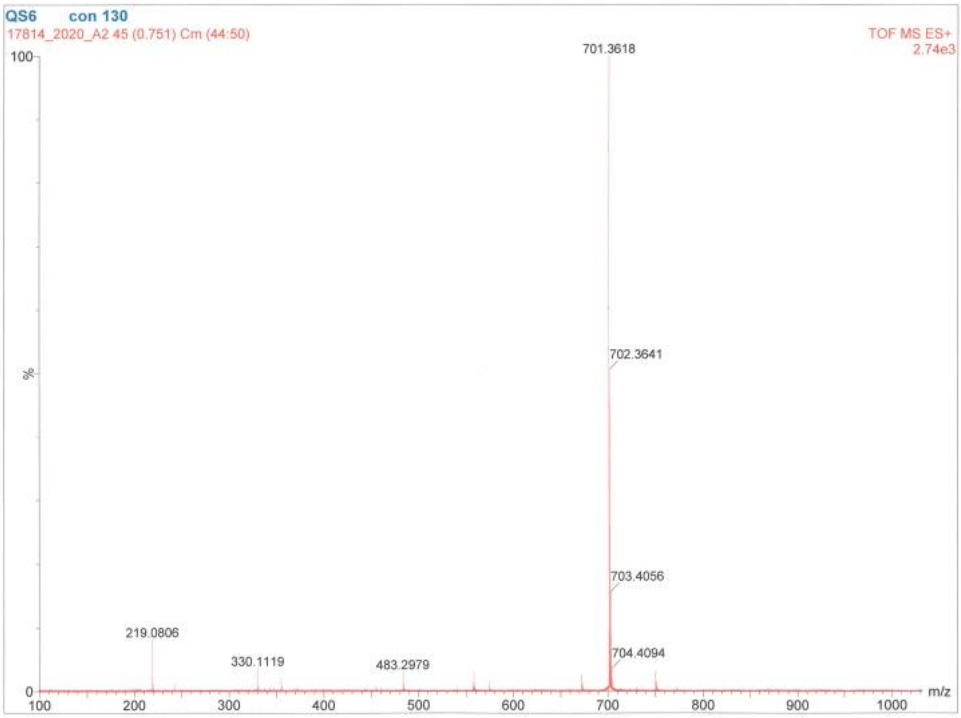




6 Ac-Abu-Orn-Leu-Gln-ACC

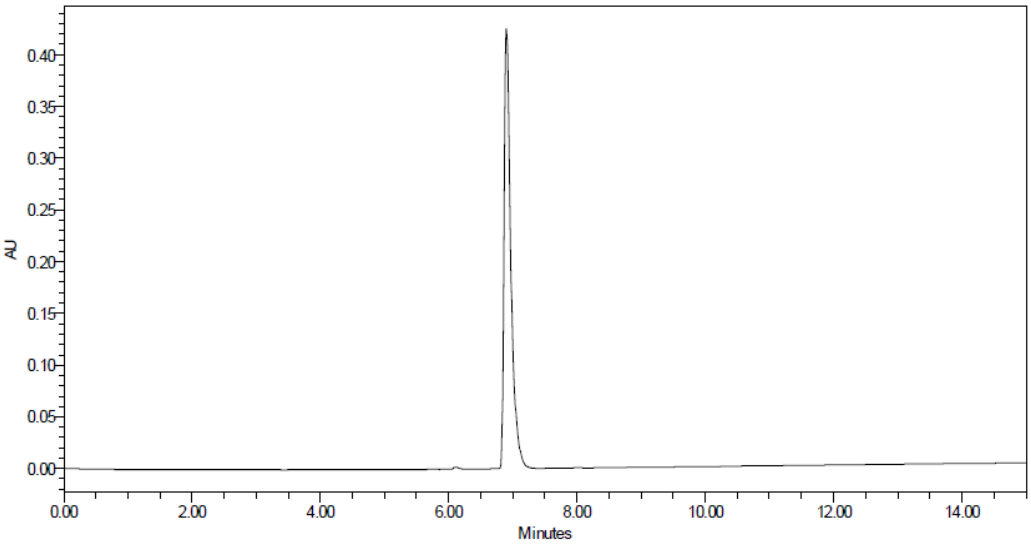
m/z: $[M+H]^+$ for $C_{33}H_{48}N_8O_9$ (m/z_{calcd} = 701.3618; m/z_{found} = 701.3618)

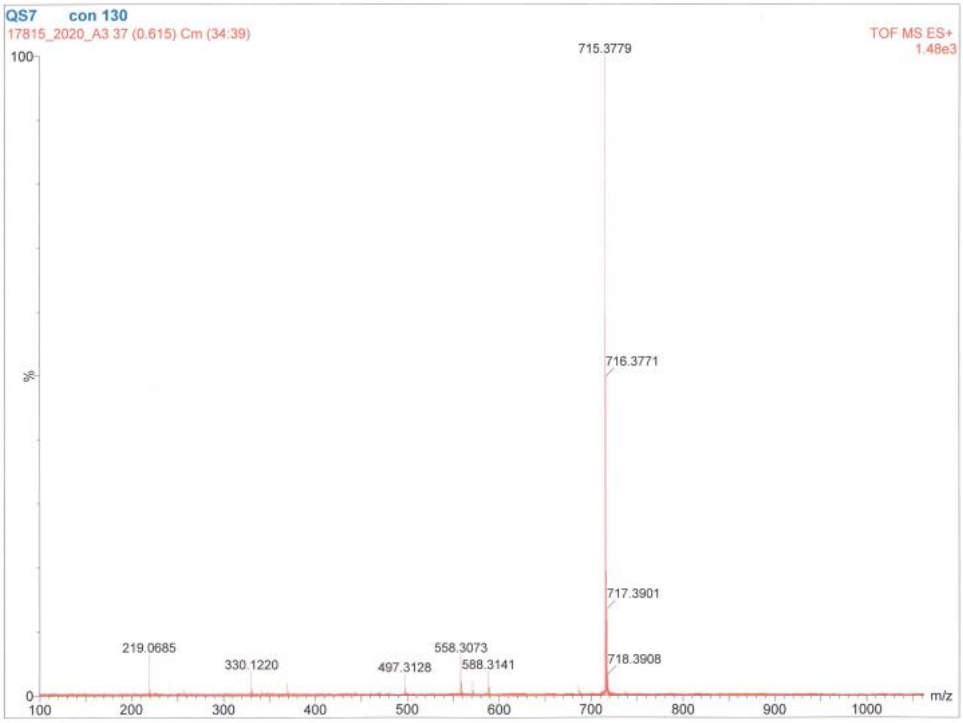




7 Ac-Abu-Lys-Leu-Gln-ACC

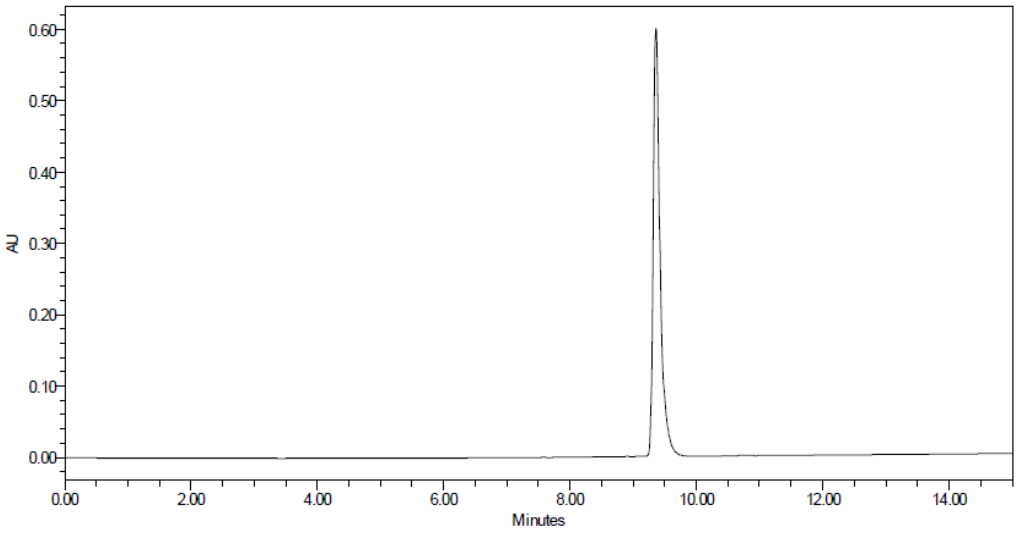
m/z: $[M+H]^+$ for $C_{34}H_{50}N_8O_9$ (m/z_{calcd} = 715.3774; m/z_{found} = 715.3779)

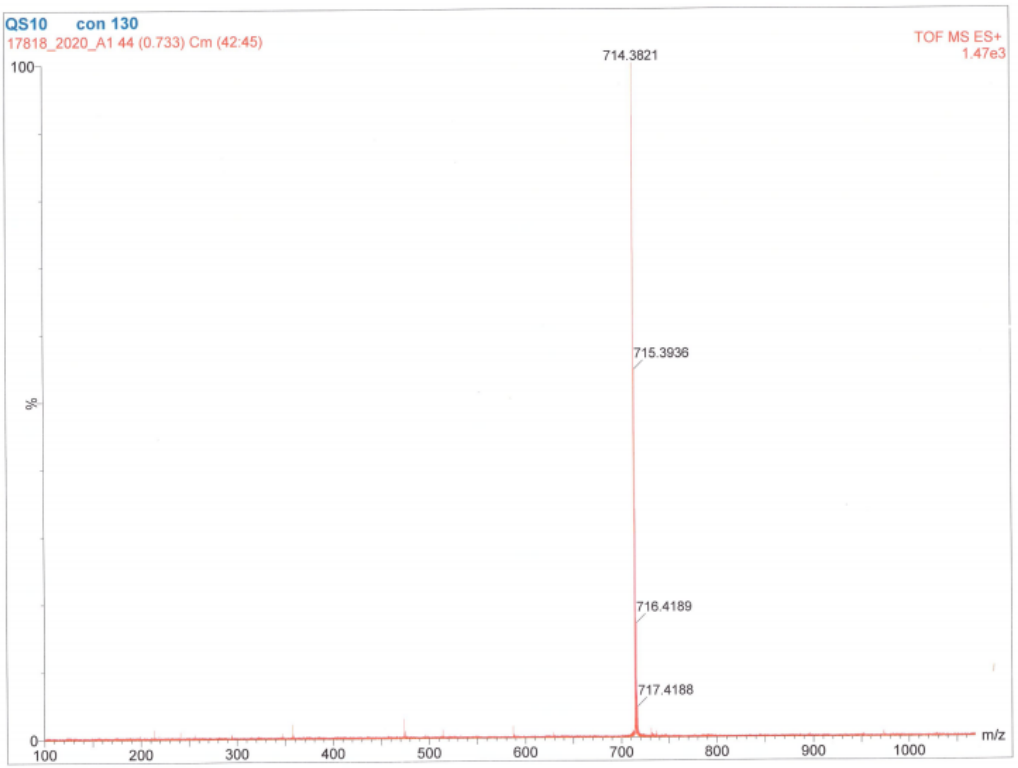




8 Ac-Abu-Tle-hLeu-Gln-ACC

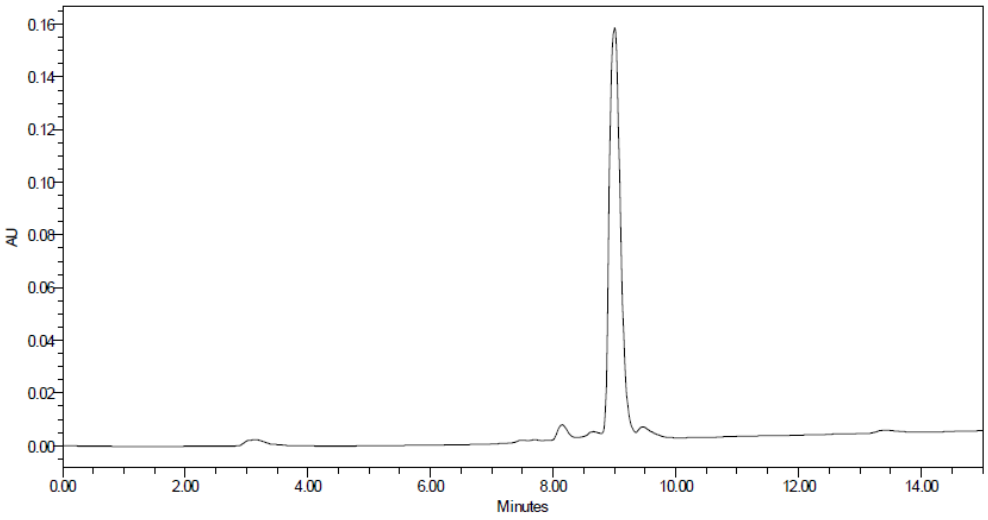
m/z: $[M+H]^+$ for $C_{35}H_{51}N_7O_9$ (m/z_{calcd} = 714.3822; m/z_{found} = 714.3821)

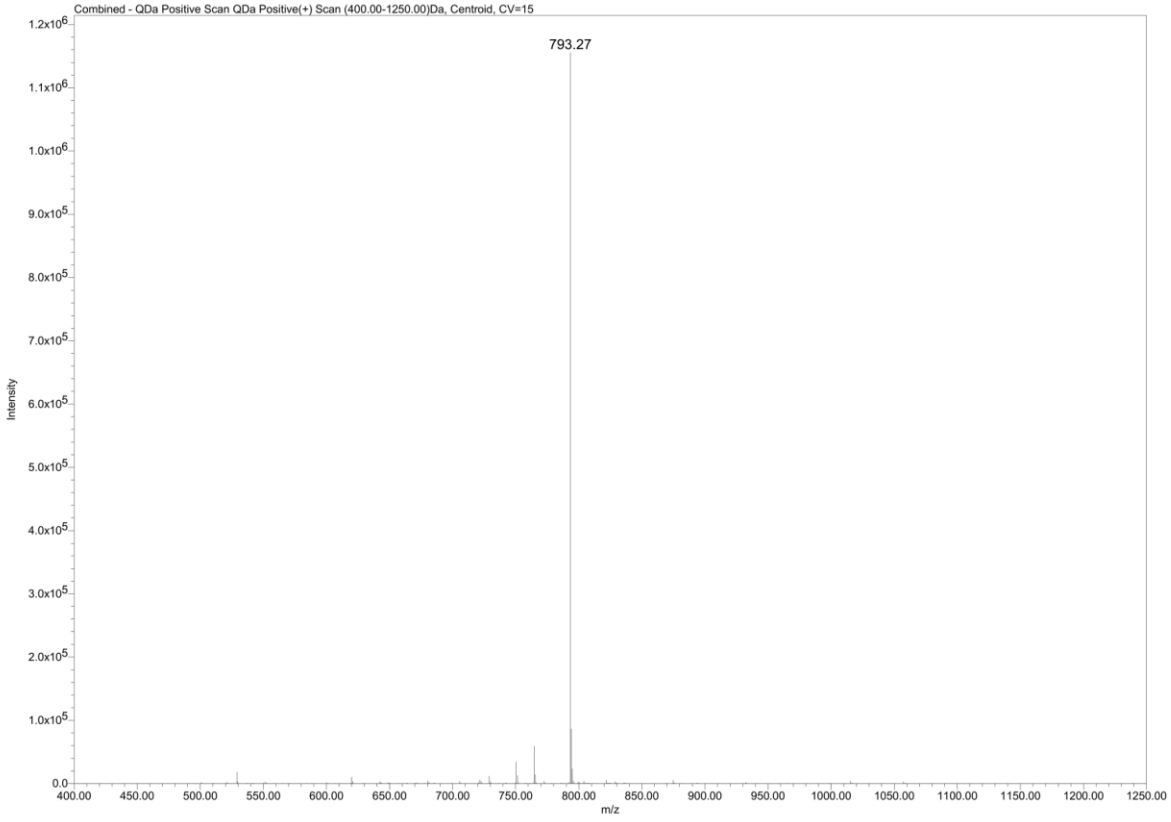




9 ACC-Gly-Abu-Tle-Leu-Gln-Ser-Gly-Phe-Arg-Lys(dnp)-Lys-NH_2

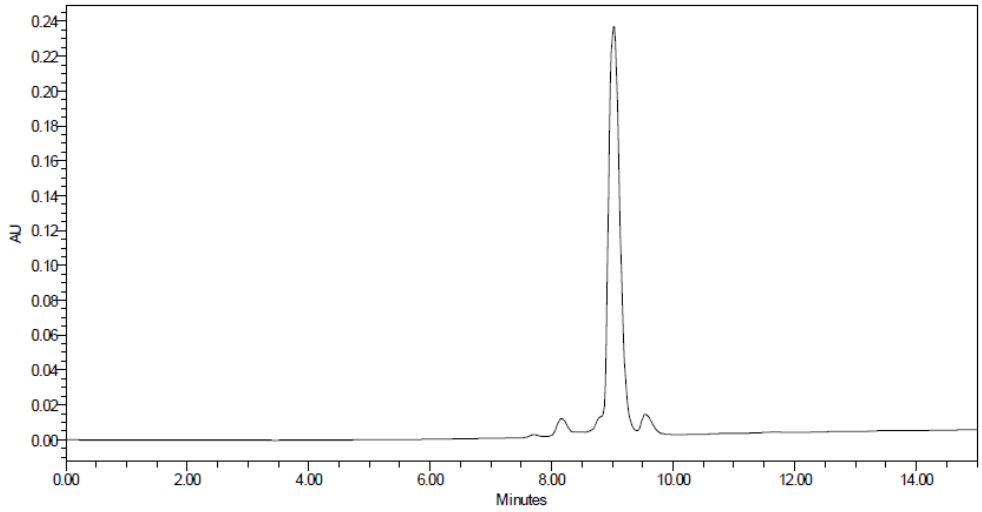
m/z: $[M+H]^{2+}$ for $C_{72}H_{105}N_{21}O_{20}$ (m/z_{calcd} = 792.89; m/z_{found} = 793.27)

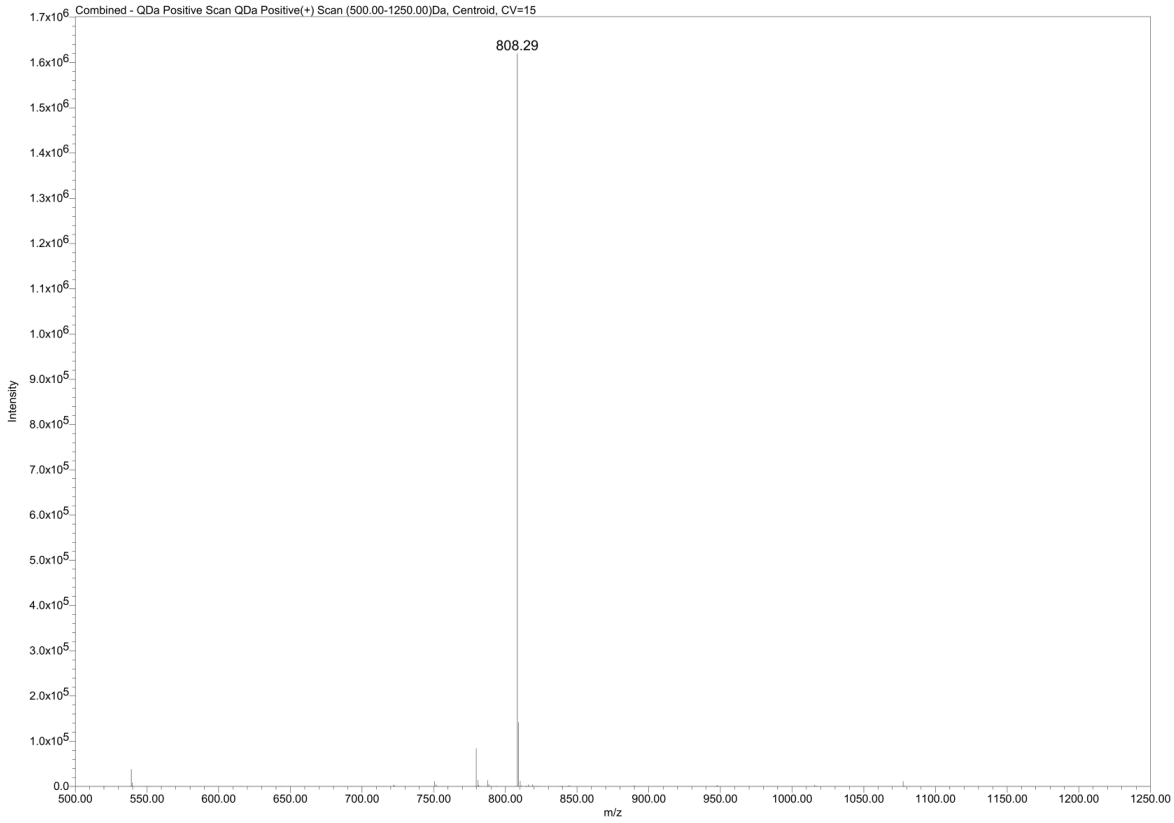




10 ACC-Gly-Thz-Tle-Leu-Gln-Ser-Gly-Phe-Arg-Lys(dnp)-Lys-NH_2

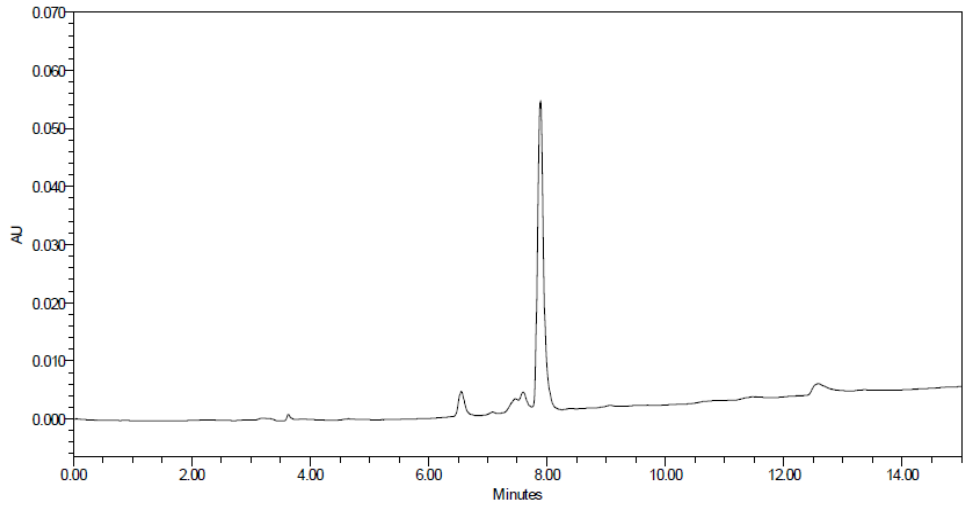
m/z: $[M+H]^{2+}$ for $C_{72}H_{103}N_{21}O_{20}S$ (m/z_{calcd} = 807.87; m/z_{found} = 808.29)

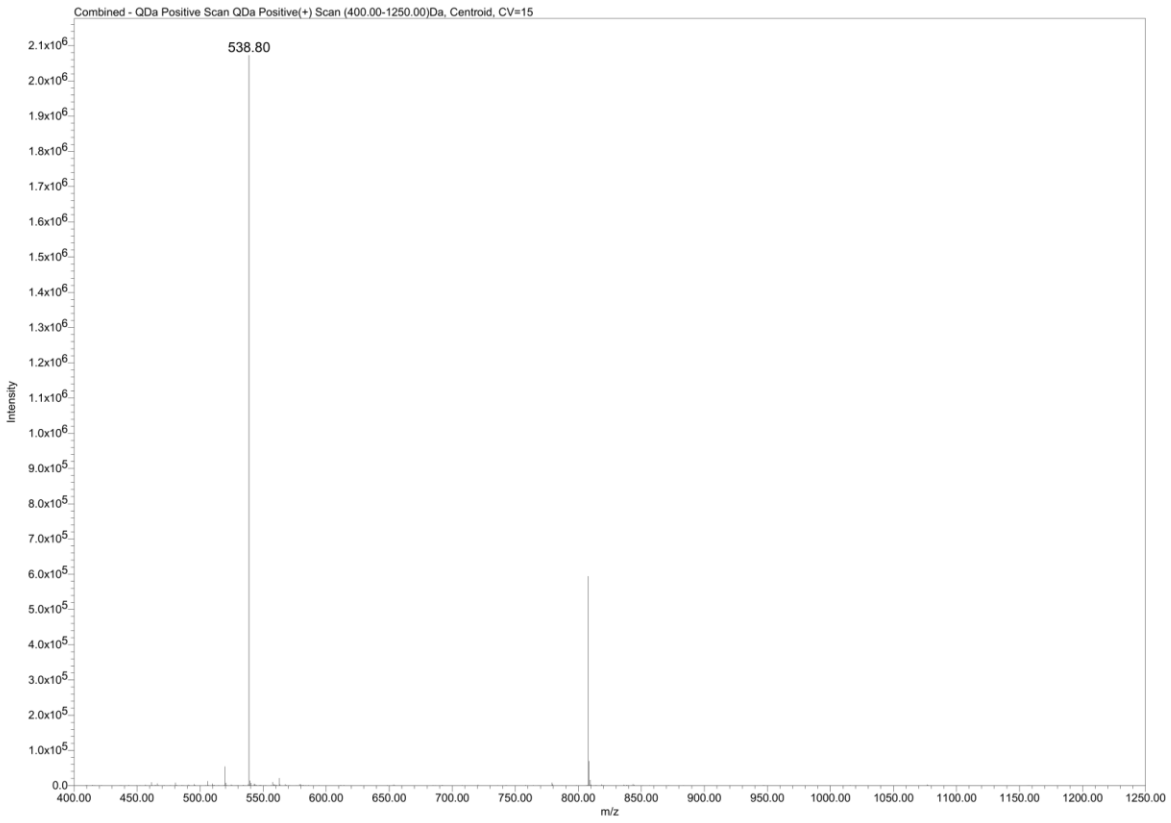




11 ACC-Gly-Val-Lys-Leu-Gln-Ser-Gly-Phe-Arg-Lys(dnp)-Lys-NH_2

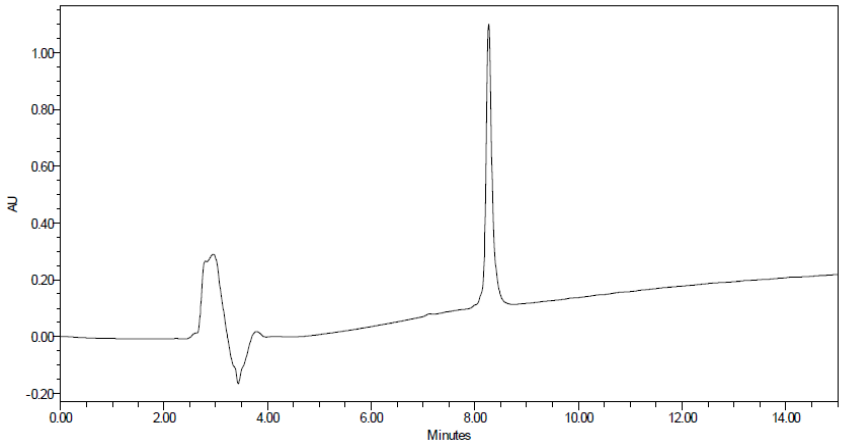
m/z: $[M+H]^{3+}$ for $C_{73}H_{108}N_{22}O_{20}S$ (m/z_{calcd} = 538.60; m/z_{found} = 538.80)

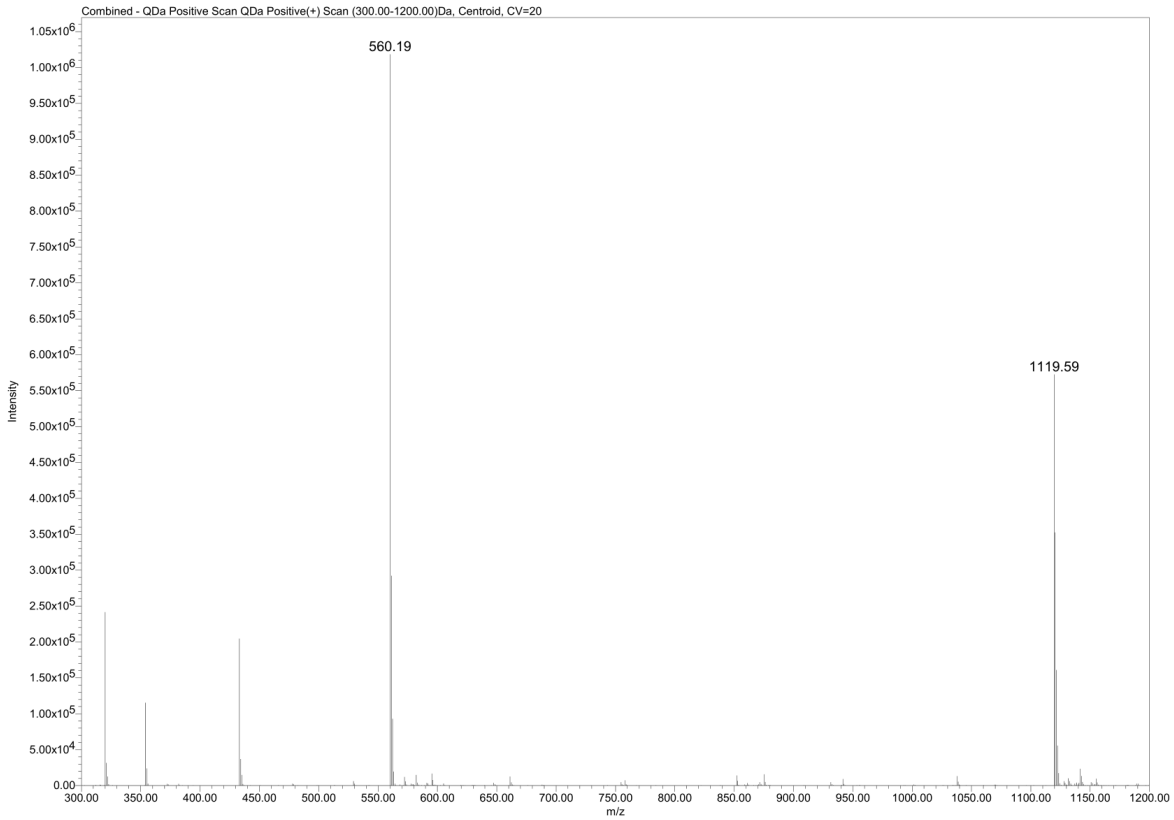




12 Ac-Abu-Tle-Leu-Gln-VS

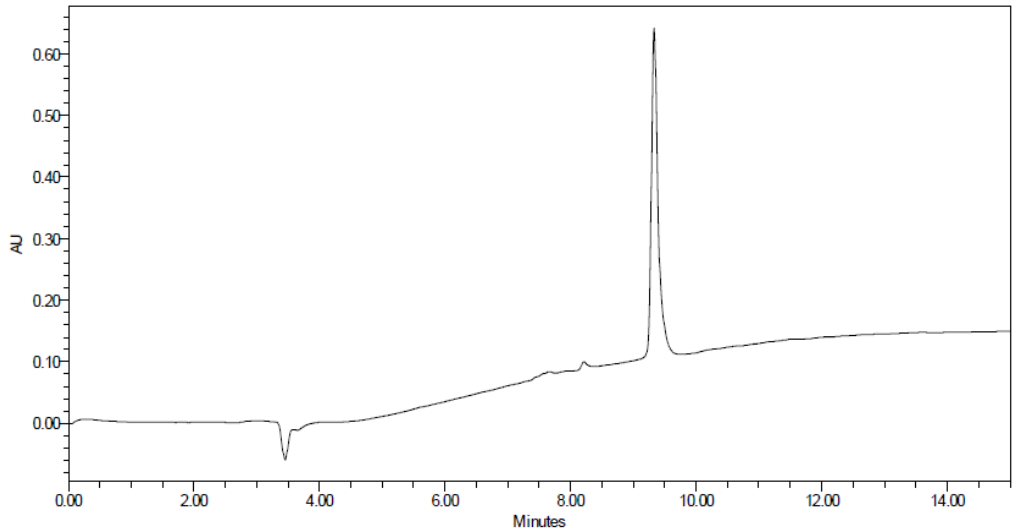
m/z: $[M+H]^+$ for $C_{25}H_{45}N_5O_7S$ (m/z_{calcd} = 560.31; m/z_{found} = 560.19)

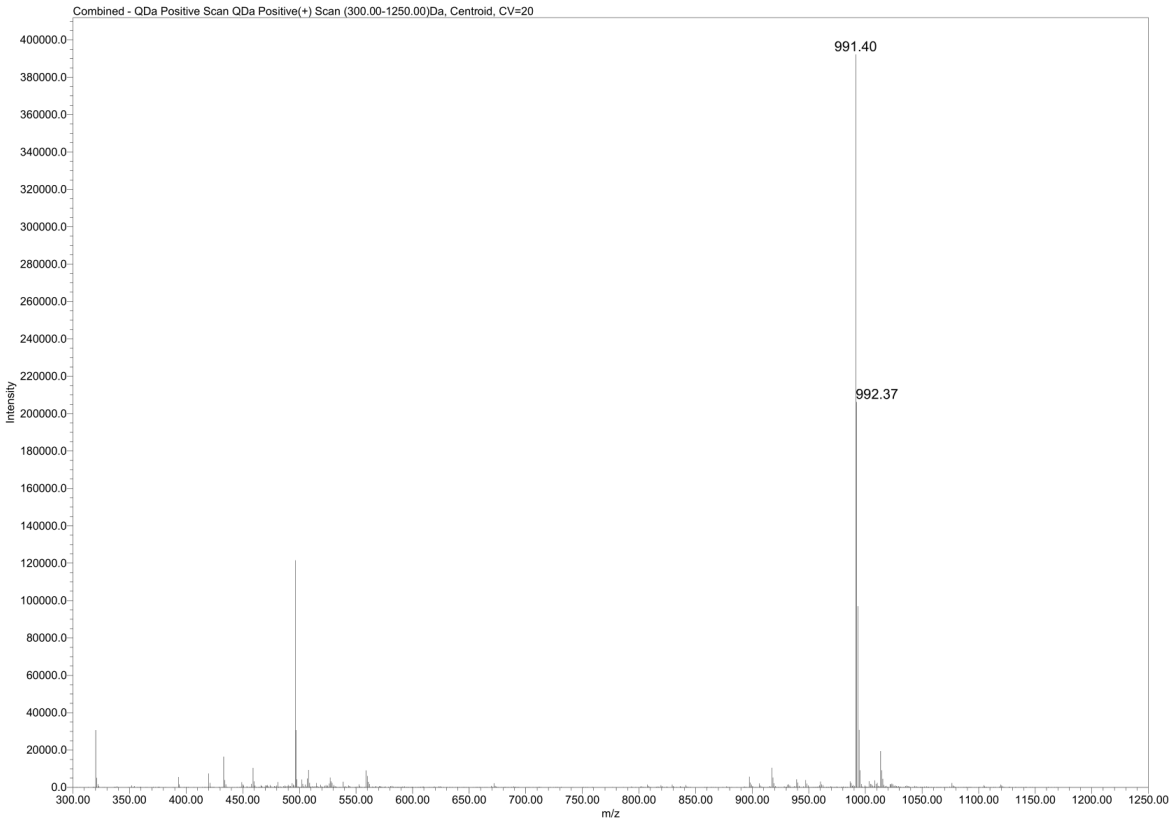




13 Biotin-PEG(4)-Abu-Tle-Leu-Gln-VS

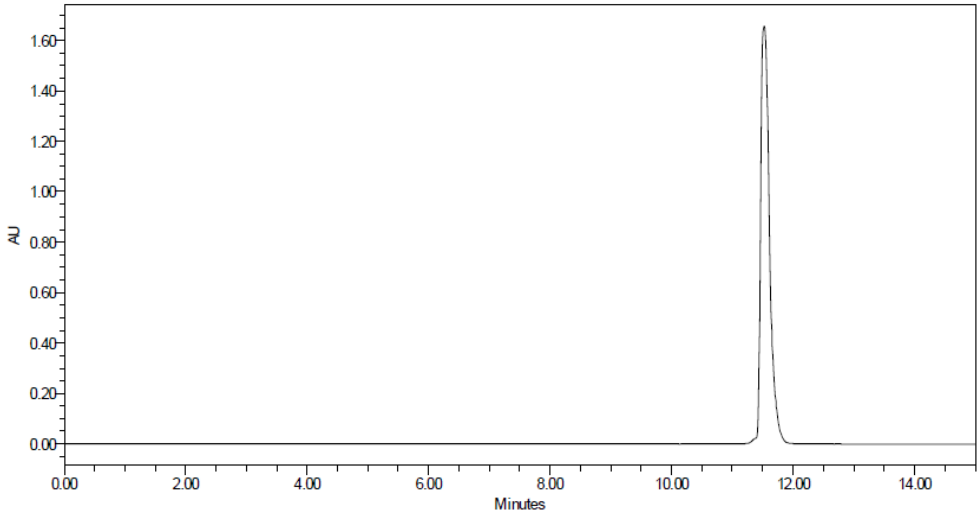
m/z: $[M+H]^+$ for $C_{44}H_{78}N_8O_{13}S_2$ (m/z_{calcd} = 991.52; m/z_{found} = 991.40)

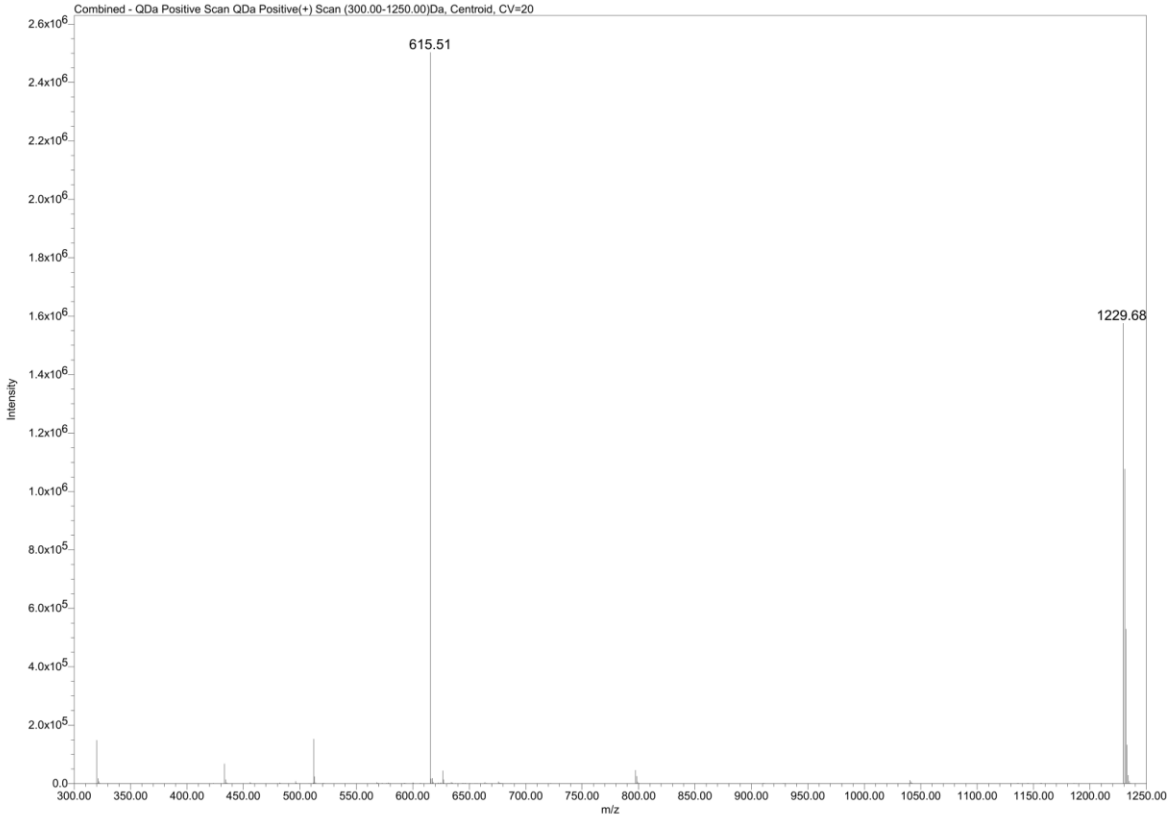




14 Cy5-PEG(4)-Abu-Tle-Leu-Gln-VS

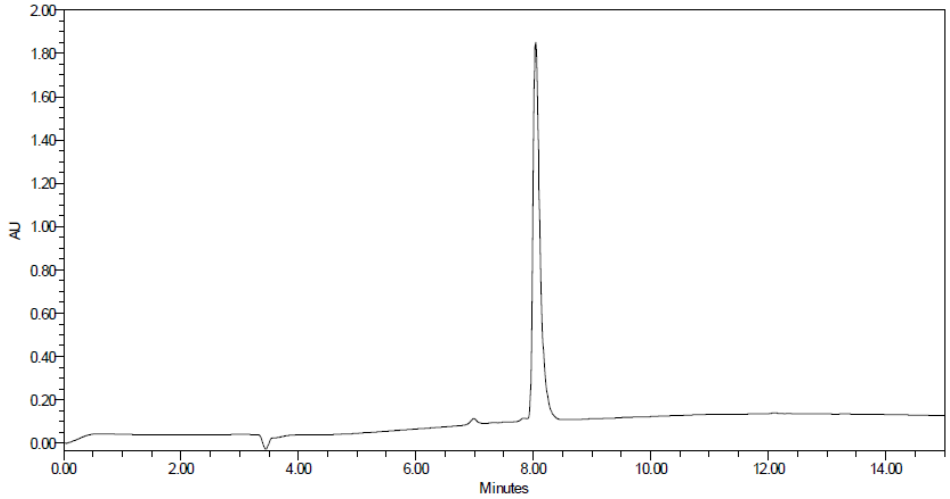
m/z: $[M+H]^+$ for $C_{66}H_{101}N_8O_{12}S$ (m/z_{calcd} = 1229.73; m/z_{found} = 1229.68)

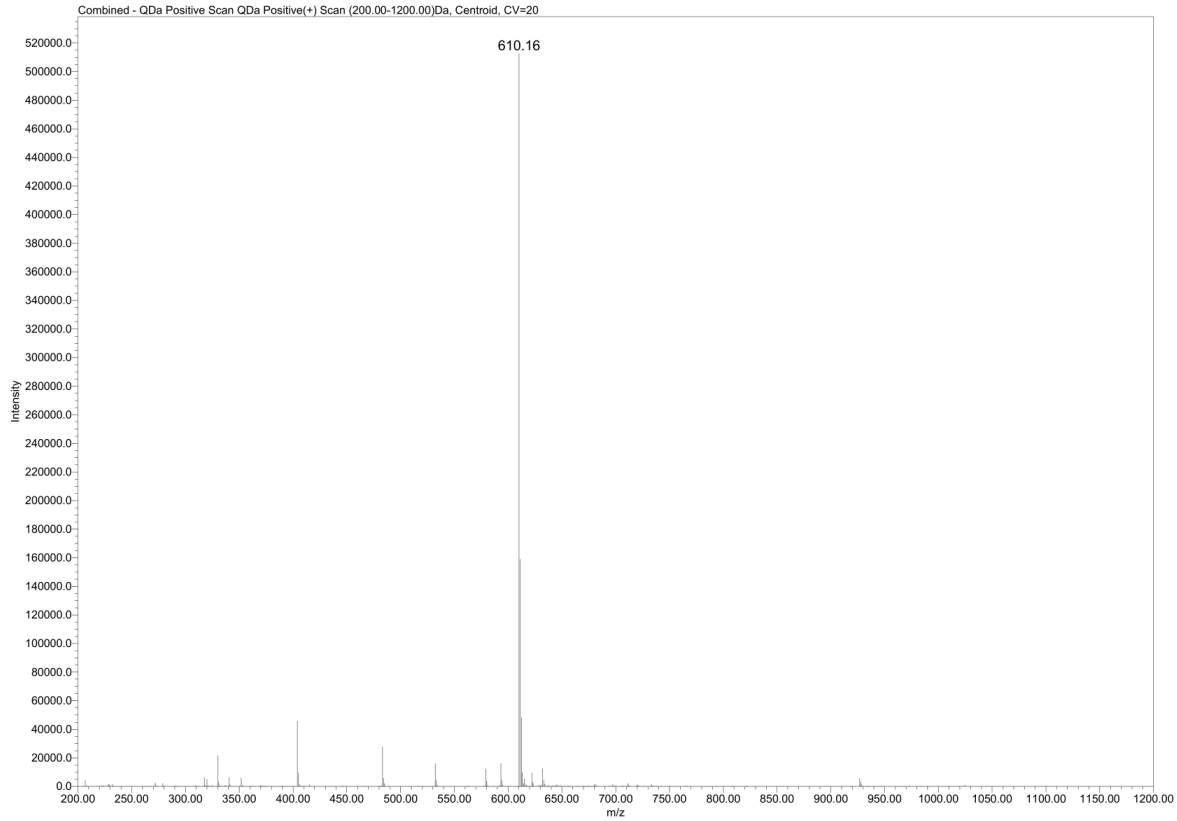




15 Ac-Abu-ᴅTyr-Leu-Gln-VS, Ac-QS5-VS

m/z: $[M+H]^+$ for $C_{28}H_{43}N_5O_8S$ (m/z_{calcd} = 610.29; m/z_{found} = 610.16)





16 Bodipy-PEG(4)-Abu-ᴅTyr-Leu-Gln-VS, Bodipy-QS5-VS

m/z: $[M+H]^+$ for $C_{51}H_{75}BF_2N_8O_{13}S$ (m/z_{calcd} = 1089.52; m/z_{found} = 1089.59)