

Supplementary Materials for

Activity profiling and crystal structures of inhibitor-bound SARS-CoV-2 papain-like protease: A framework for anti-COVID-19 drug design

Wioletta Rut*, Zongyang Lv, Mikolaj Zmudzinski, Stephanie Patchett, Digant Nayak, Scott J. Snipas, Farid El Oualid, Tony T. Huang*, Miklos Bekes*, Marcin Drag*, Shaun K. Olsen*

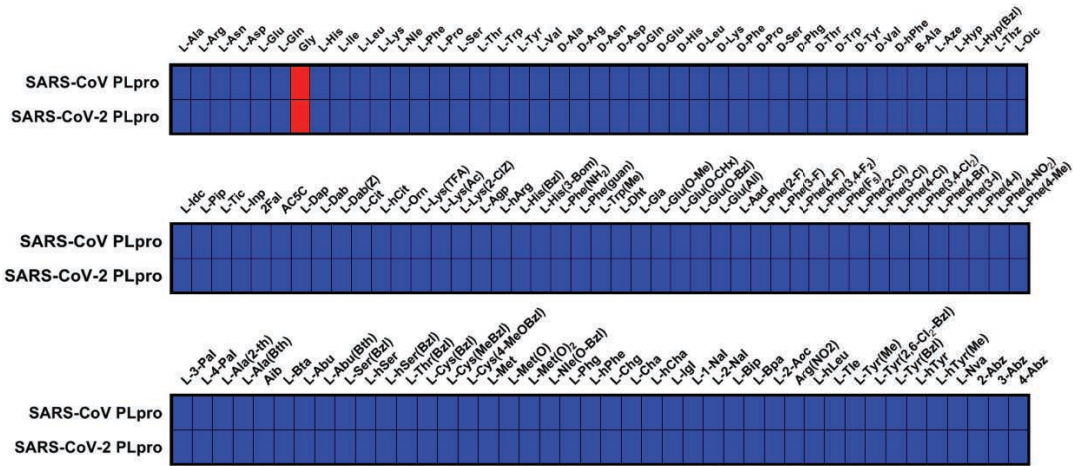
*Corresponding author. Email: wioletta.rut@pwr.edu.pl (W.R.); tony.huang@nyulangone.org (T.T.H.); miklosbekes@icloud.com (M.B.); olsens@uthscsa.edu (S.K.O.); marcin.drag@pwr.edu.pl (M.D.)

Published 16 October 2020, *Sci. Adv.* **6**, eabd4596 (2020)
DOI: 10.1126/sciadv.abd4596

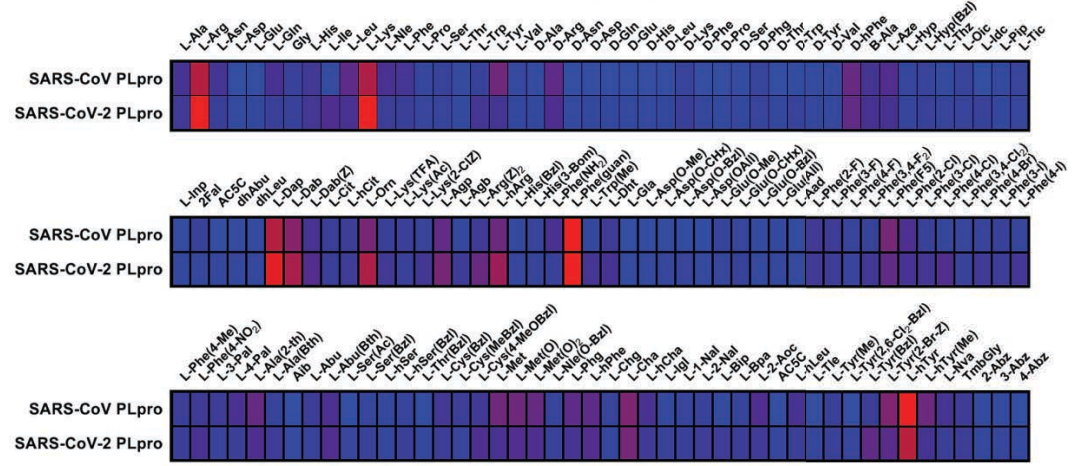
This PDF file includes:

Figs. S1 to S3
Tables S1 to S3

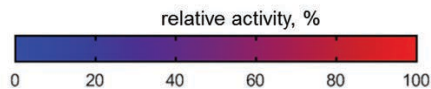
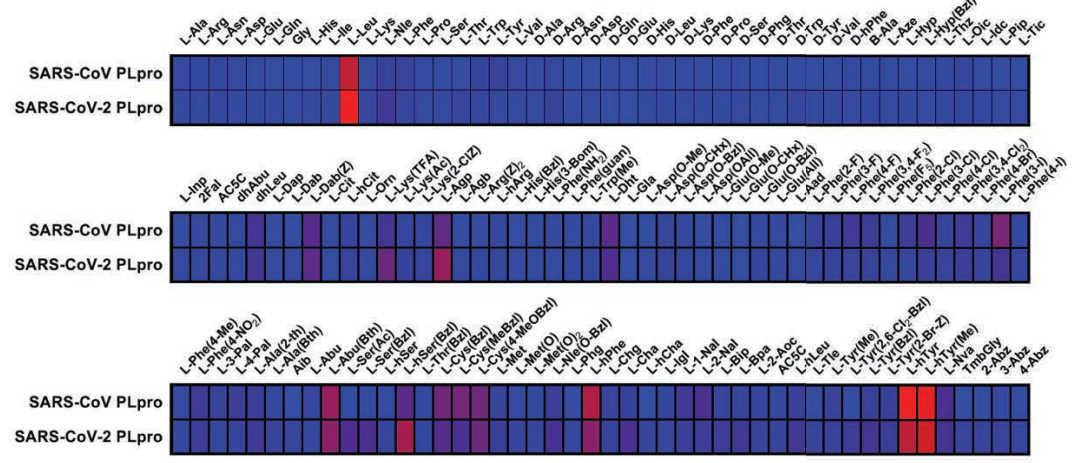
Ac-Leu-Arg-P2-Gly-ACC



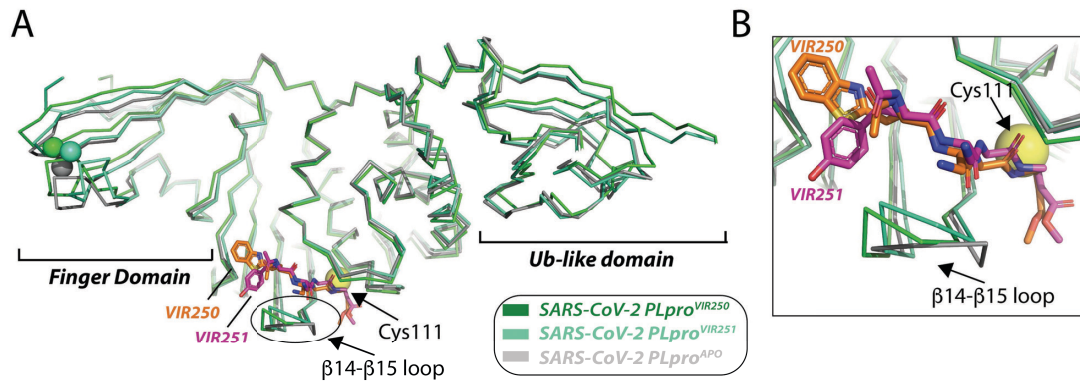
Ac-Mix-P3-Gly-Gly-ACC



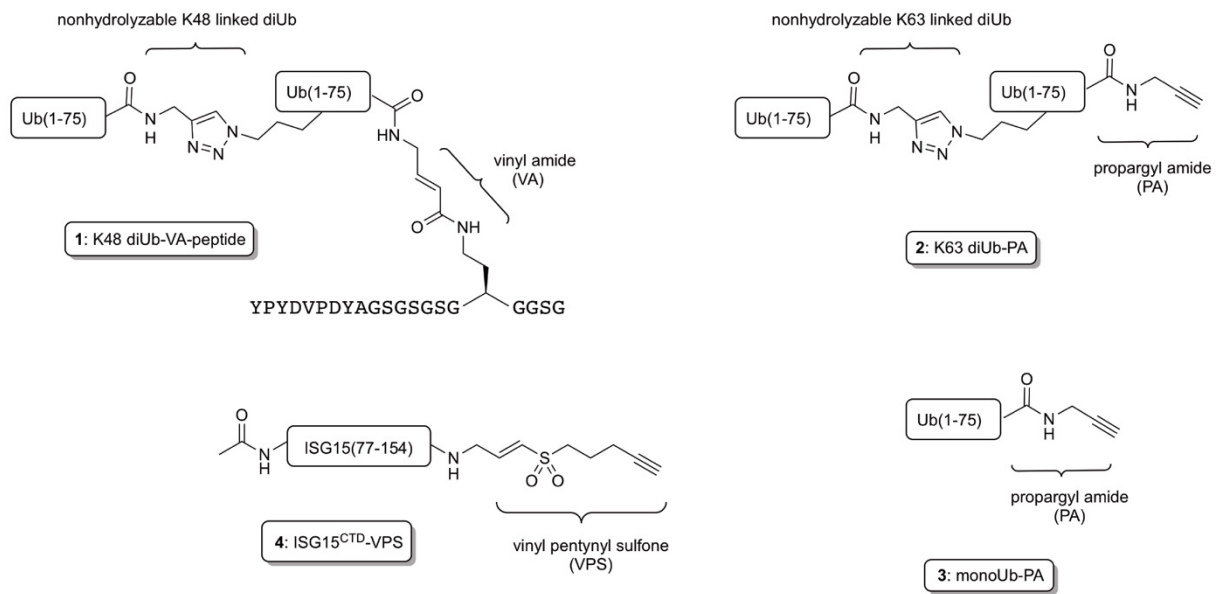
Ac-P4-Mix-Gly-Gly-ACC



Supplementary Figure 1. SARS-CoV-PLpro and SARS-CoV-2-PLpro substrate specificity profiles presented as heat maps.

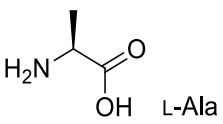
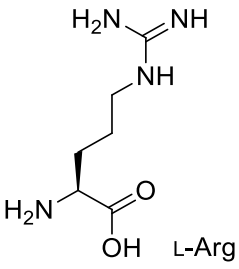
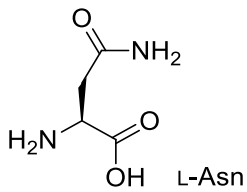
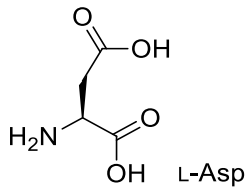
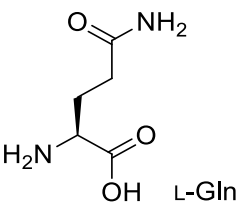
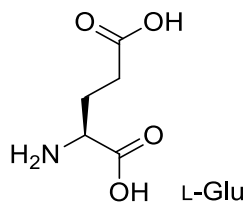
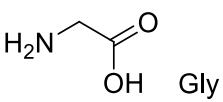
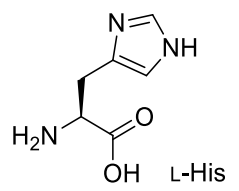
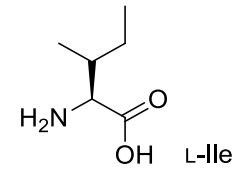
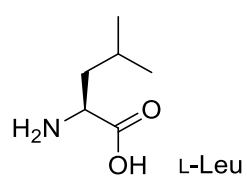
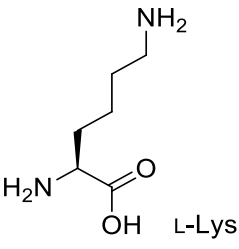
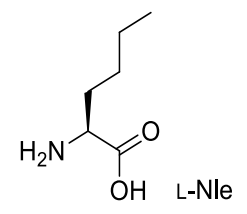
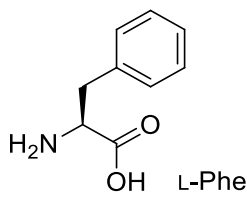
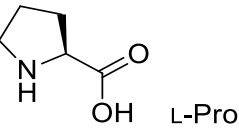
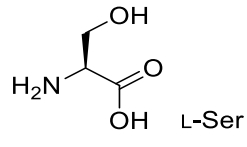
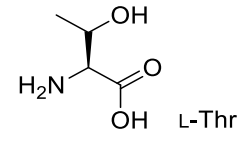
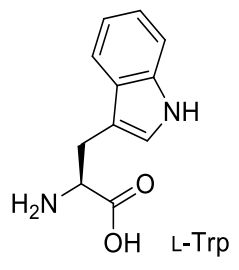
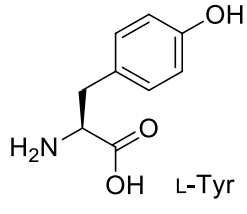


Supplementary Figure 2. Comparison of apo versus VIR250 and VIR251 bound SARS-CoV-2 PLpro. (A) The indicated structures were superimposed and are shown as ribbons. Zinc atoms are shown as spheres, colored according to their representative structure. Regions of conformational variability between apo and VIR250 and VIR251 structures are labeled including the Finger Domain, the Ub-like domain, and the β 14- β 15 loop. Catalytic cysteine 111 of SARS-CoV-2 PLpro is shown as sticks with its sg atom shown as yellow sphere. (B) Magnified view of the conformational differences around the β 14- β 15 loop of SARS-CoV-2 PLpro.

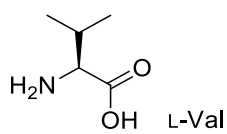


Supplementary Figure 3. Chemical structures of the Ub/Ubl activity based probes used in Figure 4B.

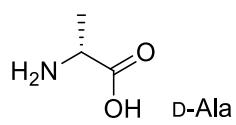
Table S1. Structures of fixed natural and unnatural amino acids used in the combinatorial library.

No	Structure + code	No	Structure + code	No	Structure + code
1	 L-Ala	2	 L-Arg	3	 L-Asn
4	 L-Asp	5	 L-Gln	6	 L-Glu
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

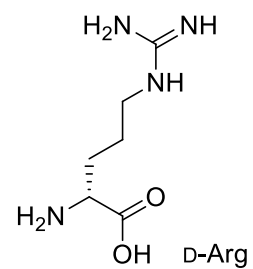
19



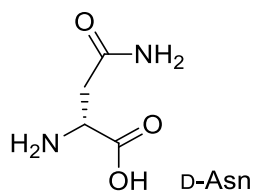
20



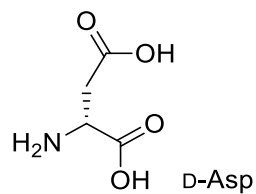
21



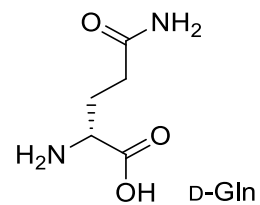
22



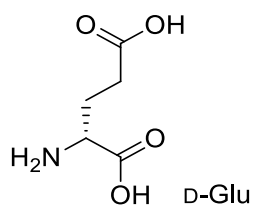
23



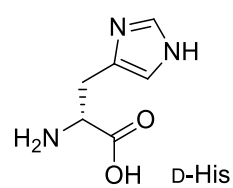
24



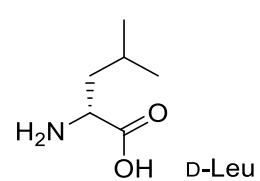
25



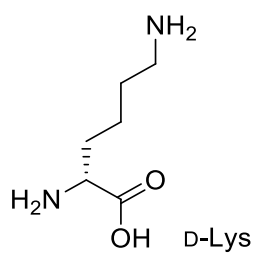
26



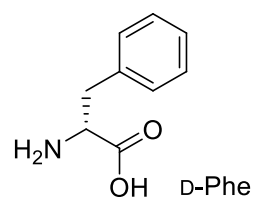
27



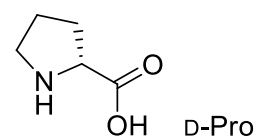
28



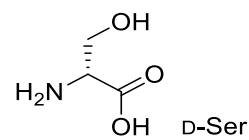
29



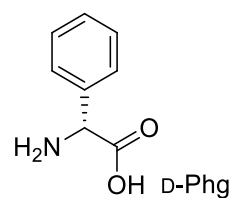
30



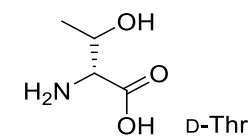
31



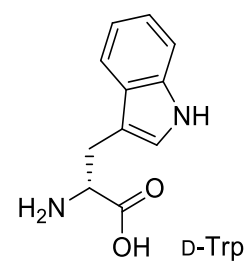
32



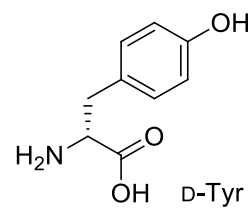
33



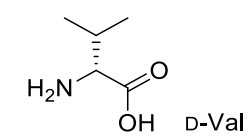
34



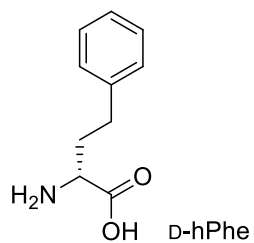
35



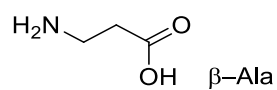
36



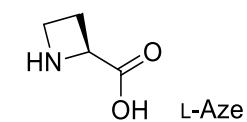
37

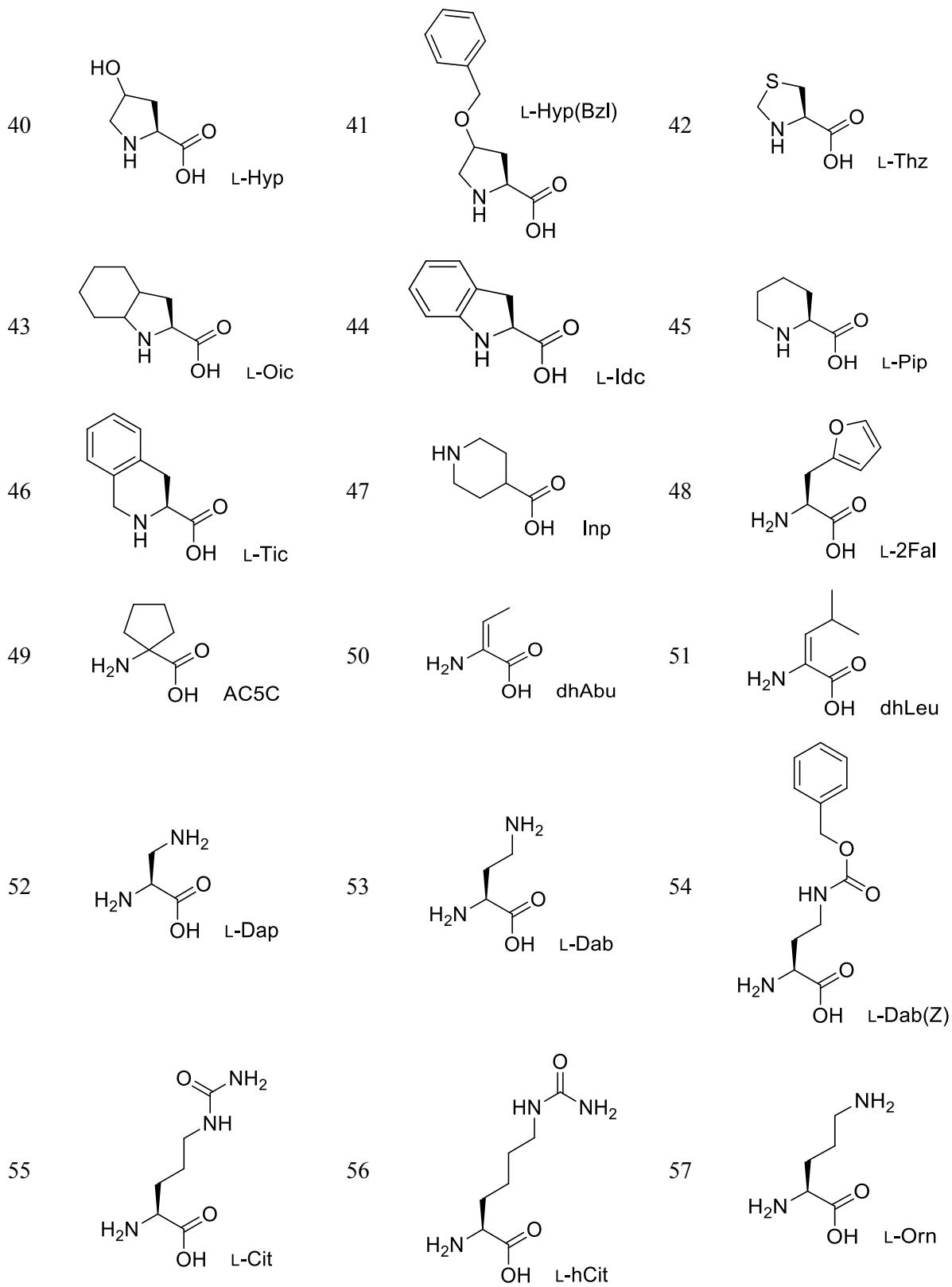


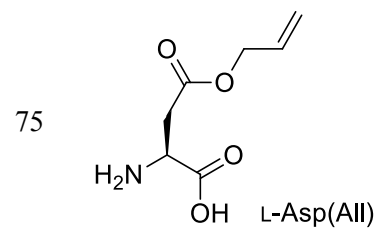
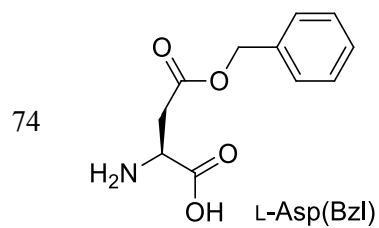
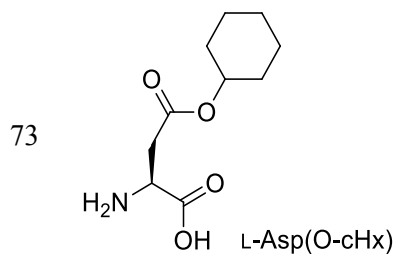
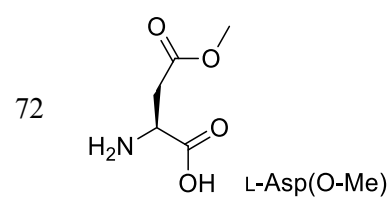
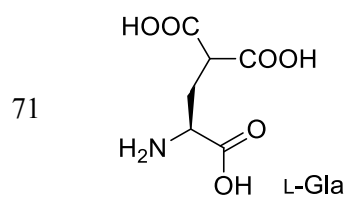
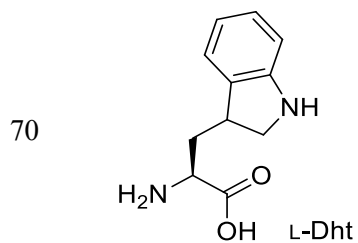
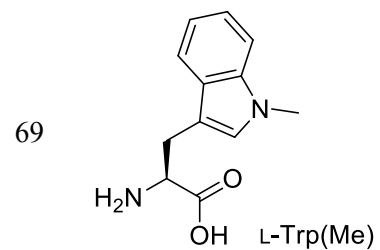
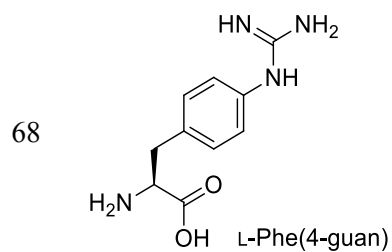
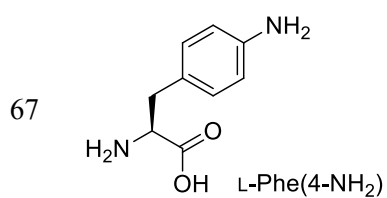
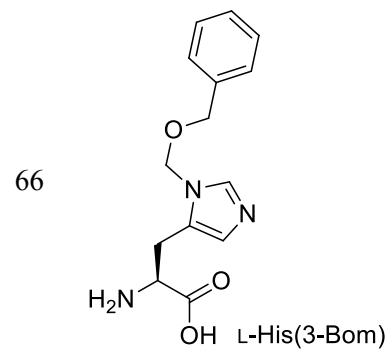
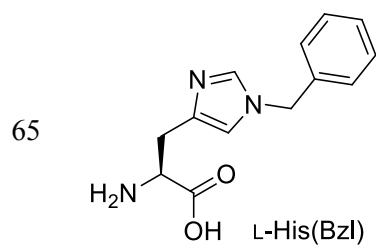
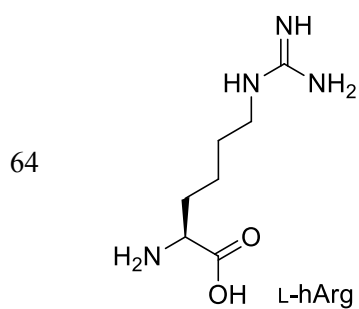
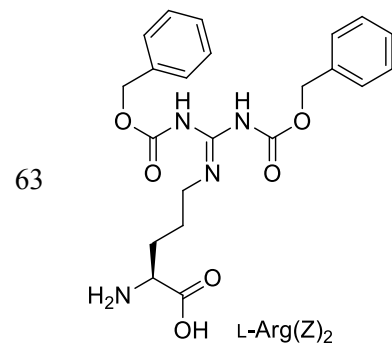
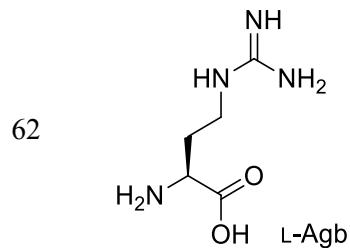
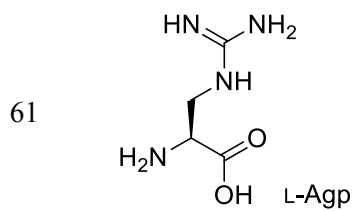
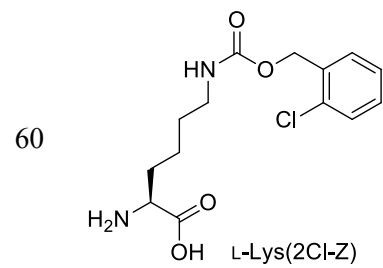
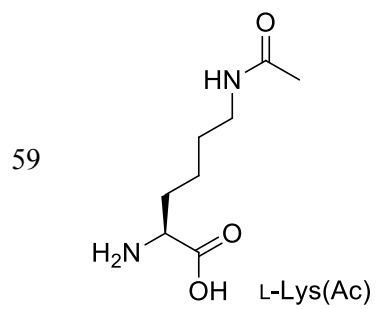
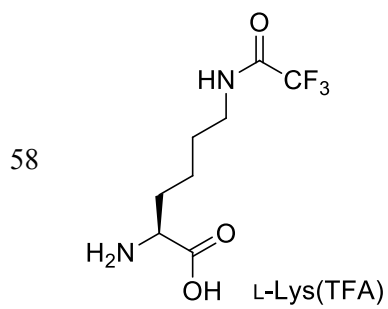
38

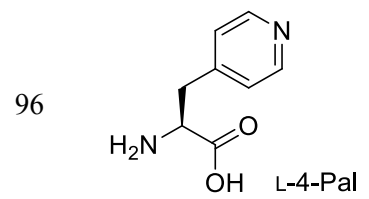
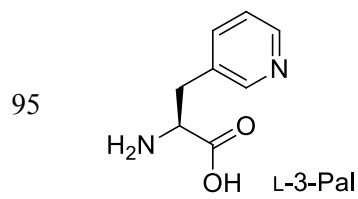
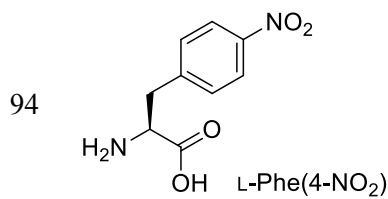
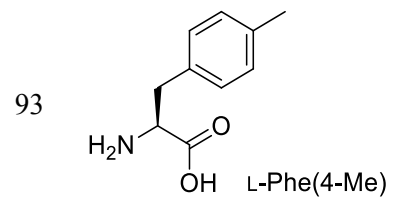
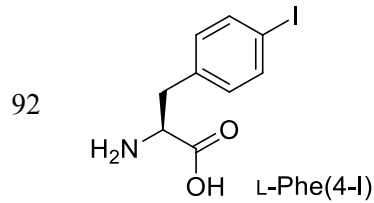
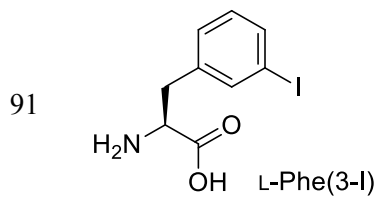
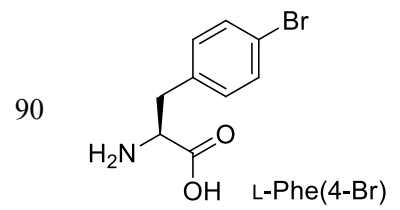
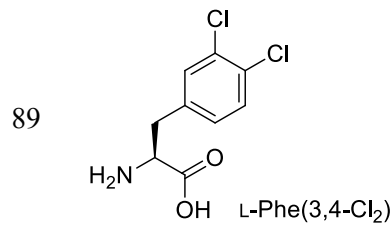
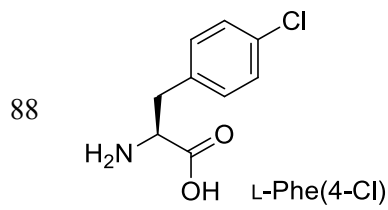
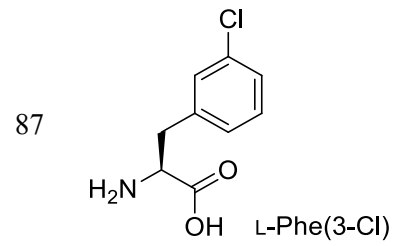
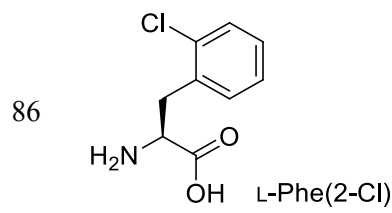
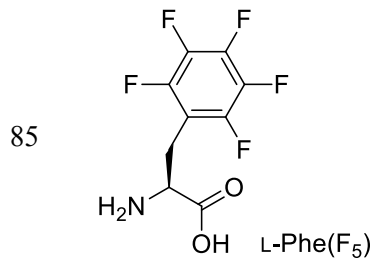
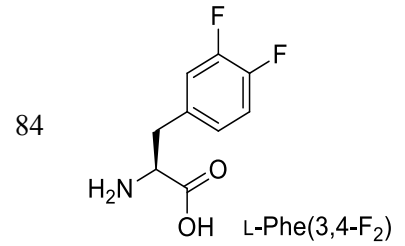
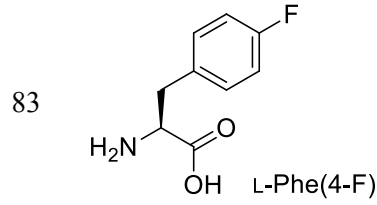
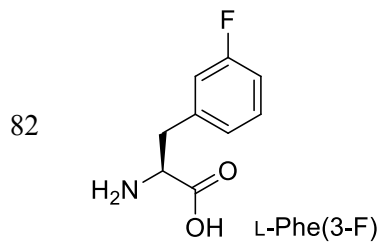
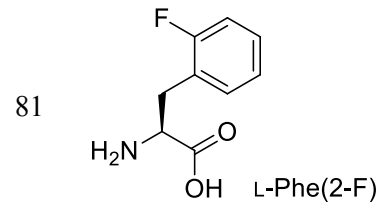
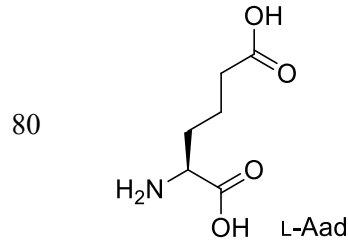
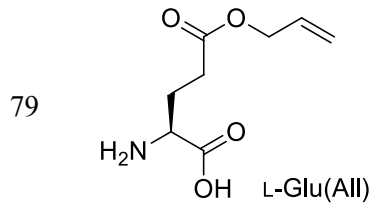
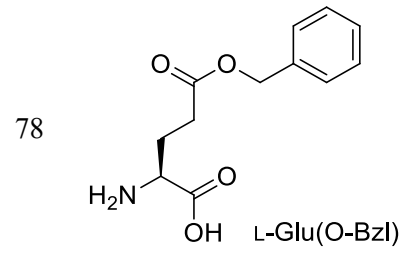
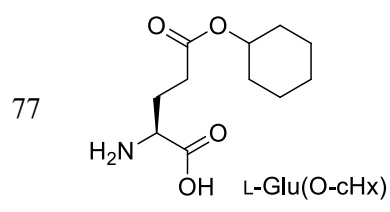
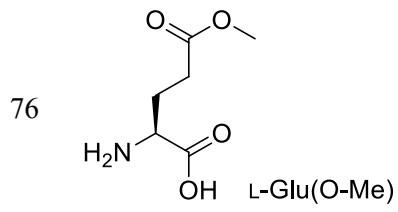


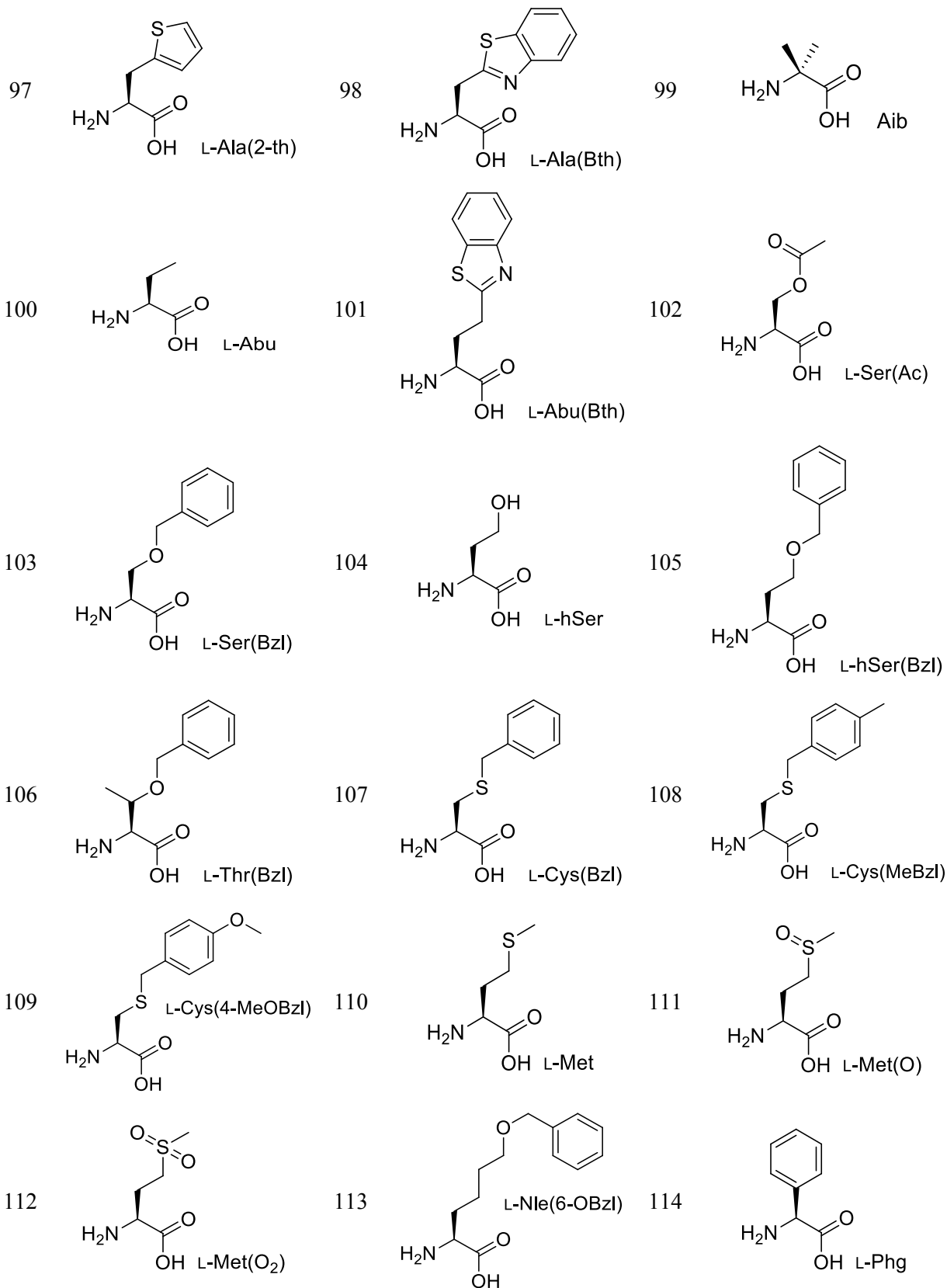
39

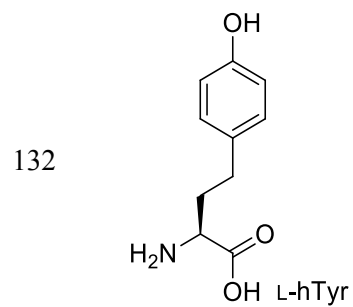
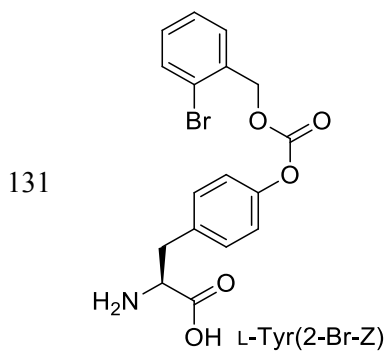
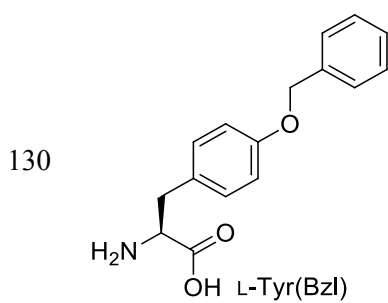
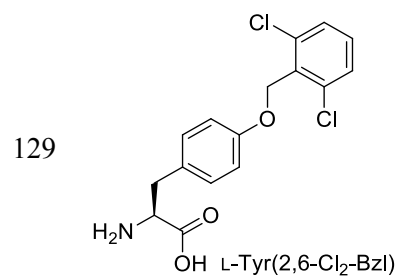
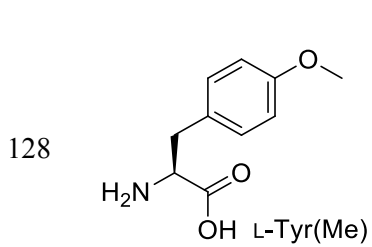
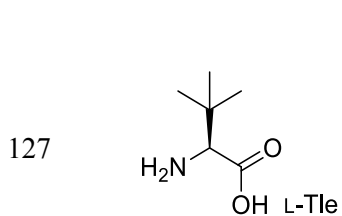
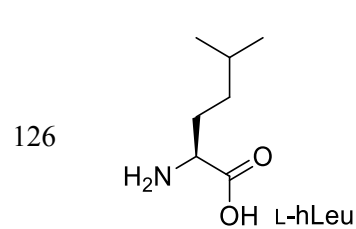
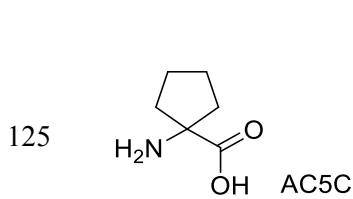
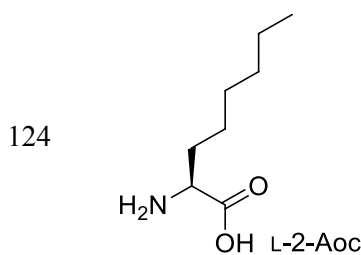
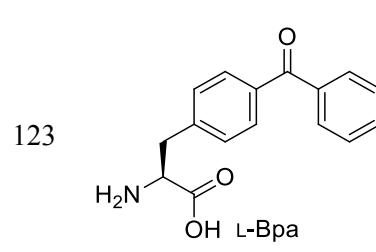
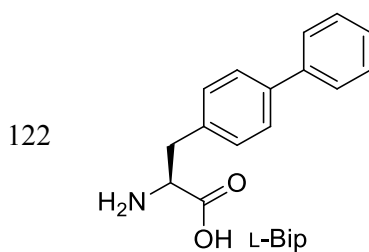
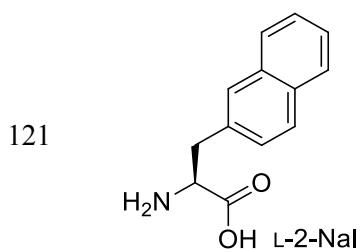
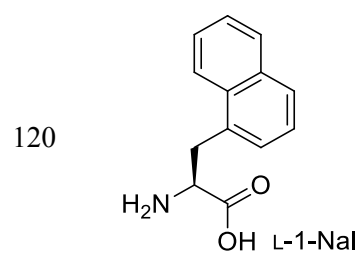
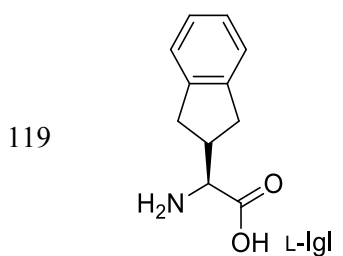
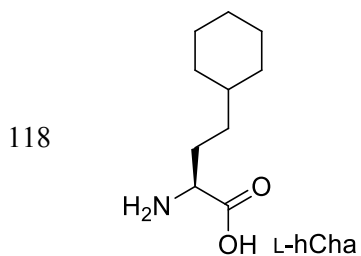
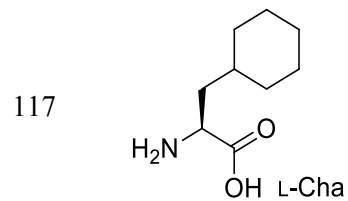
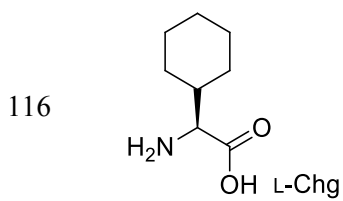
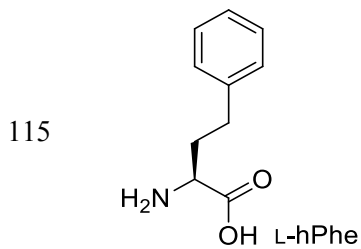


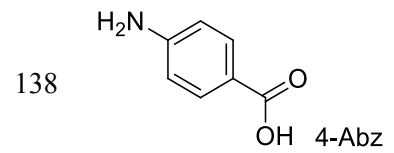
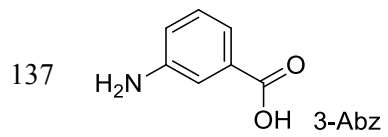
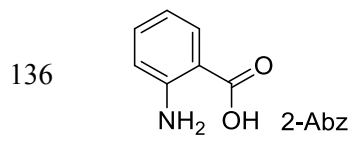
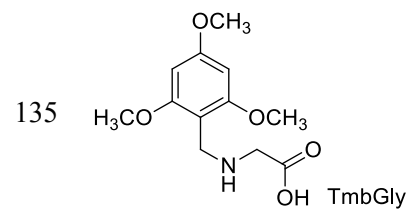
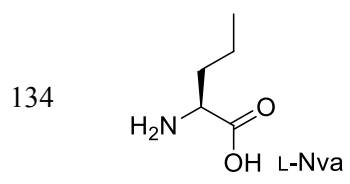
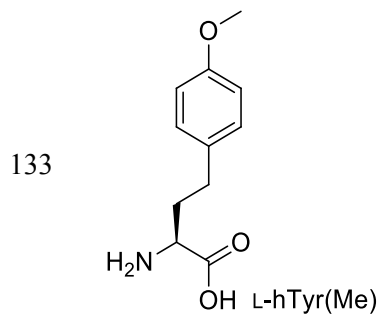












Supplementary Table 2. Crystallographic Data and Refinement Statistics

	SARS-CoV-2 PLpro/VIR250 complex	SARS-CoV-2 PLpro/VIR251 complex
PDB ID	6WUU	6WX4
Source	APS 24 IDC	APS 24 IDC
Wavelength (Å)	0.979	0.979
Resolution Limits (Å)	94.9-2.79 (2.93-2.79)	90.7-1.65 (1.68-1.65)
Space Group	P2 ₁	I222
Unit Cell (Å) <i>a, b, c</i>	58.4, 189.7, 63.1	44.9, 113.5, 151.1
Unit Cell (°) α, β, γ	90, 98.7, 90	90, 90, 90
Number of observations	237255	611490
Number of reflections	33589 (4431)	46470 (2152)
Completeness (%)	99.6 (99.1)	99.5 (91.0)
Mean $I/\sigma I$	10.1 (1.1)	15.5 (1.2)
$CC_{1/2}$	0.996 (0.437)	0.998 (0.531)
R_{merge}^a	0.158 (1.79)	0.094 (1.73)
R_{pim}	0.064 (0.753)	0.027 (0.513)
Refinement Statistics		
Resolution Limits (Å)	94.9-2.79 (2.86-2.79)	90.7-1.65 (1.70-1.65)
# of reflections (work/free)	31489 (1992)	44301 (2000)
Completeness (%)	99.3 (98.0)	99.5 (94.0)
Protein/solvent/ligand atoms	10122/8/144	2545/194/35
R_{cryst}^b	0.195 (0.299)	0.170 (0.241)
R_{free}	0.230 (0.333)	0.196 (0.256)
Bonds (Å)/ Angles (°)	0.004/0.737	0.006/0.844
B-factors: protein/solvent/ligand (Å ²)	81.2/84.8/78.2	37.7/43.5/30.6
Ramachandran plot statistics (%)		
favored	95.3	97.0
allowed	4.6	3.0
outliers	0.1	0
MolProbity score	1.48- 100 th percentile (N=4482, 2.79 Å ± 0.25Å)	1.19- 99 th percentile (N=789, 1.65 Å ± 0.25Å)

Parentheses indicate statistics for the high-resolution data bin for x-ray data.

a. $R_{\text{merge}} = \frac{\sum hkl \sum i |I(hkl)_i - \langle I(hkl) \rangle|}{\sum hkl \sum i \langle I(hkl) \rangle}$.

b. $R_{\text{cryst}} = \frac{\sum hkl |F_o(hkl) - F_c(hkl)|}{\sum hkl |F_o(hkl)|}$, where F_o and F_c are observed and calculated structure factors, respectively.

Table S3. Codon optimized DNA and protein sequences of CoV-2 PLpro constructs used in this study.

Codon optimized SARS CoV-2 PLpro [N-terminal GST tagged]- DNA sequence

GGATCCCGTGAAGTTCGTACCATTAAGGTGTTTACCACCGTTGATAATATTAATCTGCATACCCAGGT
GGTGGATATGAGCATGACCTATGGTCAGCAGTTTGGTCCGACCTATCTGGATGGTGCCGATGTTACC
AAAATTAAGCCGCATAATAGCCATGAAGGTAAAACCTTTTATGTGCTGCCGAATGATGATACCCCTGCG
TGTGGAAGCCTTTGAATATTATCATACCACCGATCCGAGTTTTCTGGGCCGTTATATGAGTGAAGTGA
ATCATACCAAAAAATGGAAATATCCGCAGGTTAATGGTCTGACCAGTATTAAGTGGGCAGATAATAAT
TGTTACCTGGCAACCGCCCTGCTGACCCTGCAGCAGATTGAACTGAAATCAATCCGCCGCCCTG
CAGGATGCCTATTATCGCGCCCGCGCAGGTGAAGCCGCCAATTTTTGCGCCCTGATTCTGGCATATT
GTAATAAGACCGTTGGTGAAGTGGGTGACGTGCGTGAAACCATGAGTTATCTGTTTCAGCATGCCAA
TCTGGATAGTTGTAACGCGTGCTGAATGTTGTTTGCAAAACCTGTGGCCAGCAGCAGACCACCCTG
AAAGGTGTGGAAGCAGTGATGTATATGGTACCCTGAGCTATGAACAGTTTAAAAAAGGTGTGCAGA
TTCCGTGTACCTGTGGCAAAACAGGCAACCAAATATCTGGTTCAGCAGGAAAGCCCGTTTGTATGAT
GAGTGCCCCGCCGCCAGTATGAACTGAAACATGGCACCTTTACCTGTGCAAGTGAATATACCGG
TAATTATCAGTGTGGTCAATATAAACATATCACCAGTAAAGAAACCCTGTATTGATTGATGGCGCACT
GCTGACCAAAAAGCAGCGAATATAAAGGCCCGATTACCGATGTTTTCTATAAAGAAAATAGCTACACCA
CCACCATTAAGCCGTAAGCGGCCG

Codon optimized SARS CoV-2 PLpro [uncleavable C-terminal 6x His tag]- protein sequence after Precision Protease cleavage of GST tag

GPLGSREVRTIKVFTTVDNINLHTQVVDMSMTYQQFGPTYLDGADVTKIKPHNSHEGKTFYVLPNDLTL
RVEAFEYHHTDPSFLGRYMSALNHTKKWKYPQVNGLTSIKWADNNCYLATALLTQQIELKFNPPALQD
AYRRARAGEAANFCALILAYCNKTVGELGDVRETMSYLFQHANLDSCKRVLNVVCKTCGQQQTLLKQVE
AVMYMGTLSEYQFKKGVQIPCTCGKQATKYLQQESPVMMSAPPAQYELKHGTFTCASEYTGNYQCG
HYKHITSKETLYCIDGALLTKSSEYKGPITDVFYKENSYTTTIKP

Codon optimized SARS CoV-2 PLpro [uncleavable C-terminal 6x His tag]- DNA sequence

ATGCGTGAAGTTCGTACCATTAAGGTGTTTACCACCGTTGATAATATTAATCTGCATACCCAGGTGGT
GGATATGAGCATGACCTATGGTCAGCAGTTTGGTCCGACCTATCTGGATGGTGCCGATGTTACCAAA
ATTAAGCCGCATAATAGCCATGAAGGTAAAACCTTTTATGTGCTGCCGAATGATGATACCCCTGCGTGT
GGAAGCCTTTGAATATTATCATACCACCGATCCGAGTTTTCTGGGCCGTTATATGAGTGAAGTGA
ATACCAAAAAATGGAAATATCCGCAGGTTAATGGTCTGACCAGTATTAAGTGGGCAGATAATAATTGT
TACCTGGCAACCGCCCTGCTGACCCTGCAGCAGATTGAACTGAAATCAATCCGCCGCCCTGCGAG
GATGCCTATTATCGCGCCCGCGCAGGTGAAGCCGCCAATTTTTGCGCCCTGATTCTGGCATATTGTA
ATAAGACCGTTGGTGAAGTGGGTGACGTGCGTGAAACCATGAGTTATCTGTTTCAGCATGCCAATCT
GGATAGTTGTAACGCGTGCTGAATGTTGTTTGCAAAACCTGTGGCCAGCAGCAGACCACCCTGAAA
GGTGTGGAAGCAGTGATGTATATGGGTACCCTGAGCTATGAACAGTTTAAAAAAGGTGTGCAGATTC
CGTGTACCTGTGGCAAAACAGGCAACCAAATATCTGGTTCAGCAGGAAAGCCCGTTTGTATGATGAG
TGCCCCGCCGCCAGTATGAACTGAAACATGGCACCTTTACCTGTGCAAGTGAATATAACCGGTAAT
TATCAGTGTGGTCAATATAAACATATCACCAGTAAAGAAACCCTGTATTGATTGATGGCGCACTGCT
GACCAAAAAGCAGCGAATATAAAGGCCCGATTACCGATGTTTTCTATAAAGAAAATAGCTACACCACCA
CCATTAAGCCGCTCGAGCACCACCACCACCACCTGA

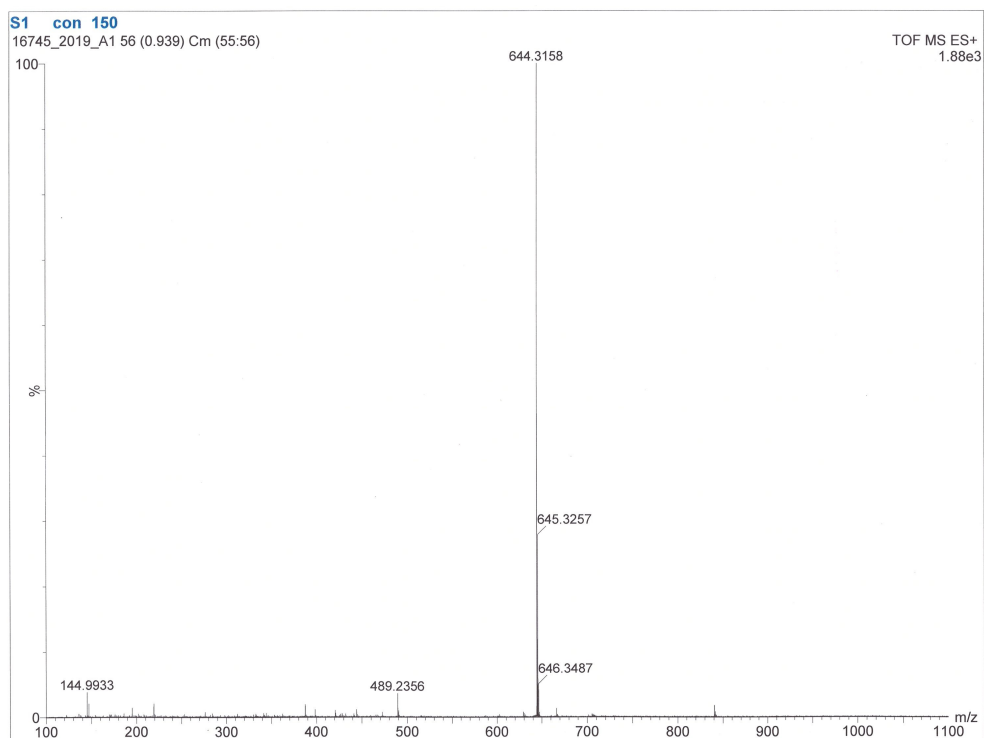
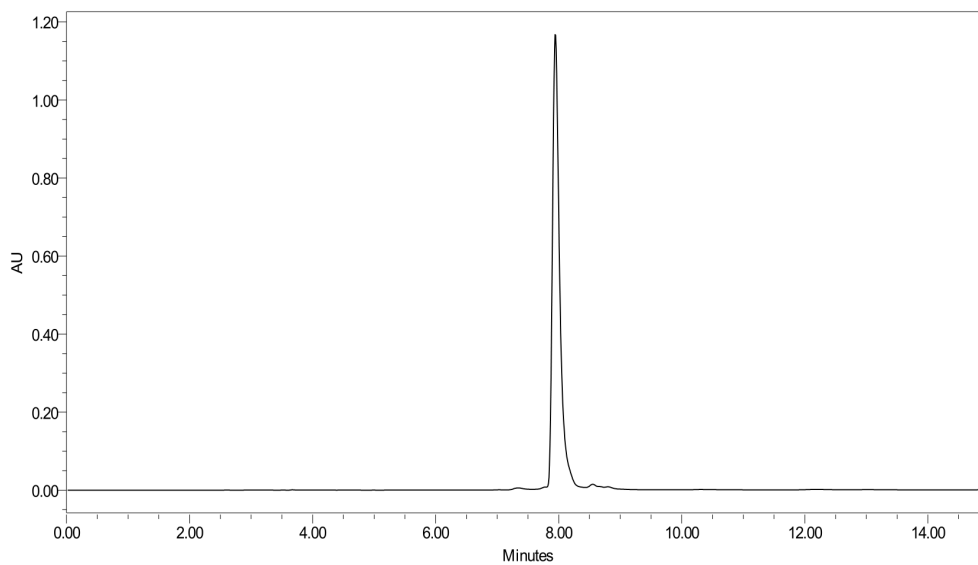
Codon optimized SARS CoV-2 PLpro [uncleavable C-terminal 6x His tag]- protein sequence

MREVRTIKVFTTVDNINLHTQVVDMSMTYQQFGPTYLDGADVTKIKPHNSHEGKTFYVLPNDLTLRVEA
FEYHHTDPSFLGRYMSALNHTKKWKYPQVNGLTSIKWADNNCYLATALLTQQIELKFNPPALQDAYYR
ARAGEAANFCALILAYCNKTVGELGDVRETMSYLFQHANLDSCKRVLNVVCKTCGQQQTLLKQVEAVMY
MGTLSEYQFKKGVQIPCTCGKQATKYLQQESPVMMSAPPAQYELKHGTFTCASEYTGNYQCGHYKHI
TSKETLYCIDGALLTKSSEYKGPITDVFYKENSYTTTIKPLEHHHHHH

LC-MS analysis of synthesized chemical compounds

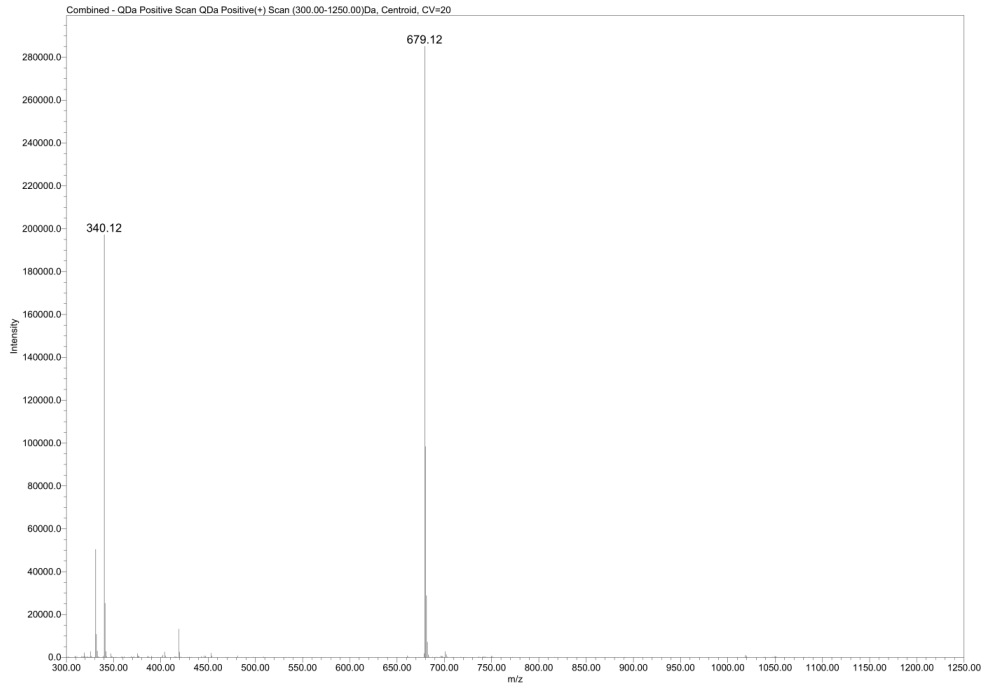
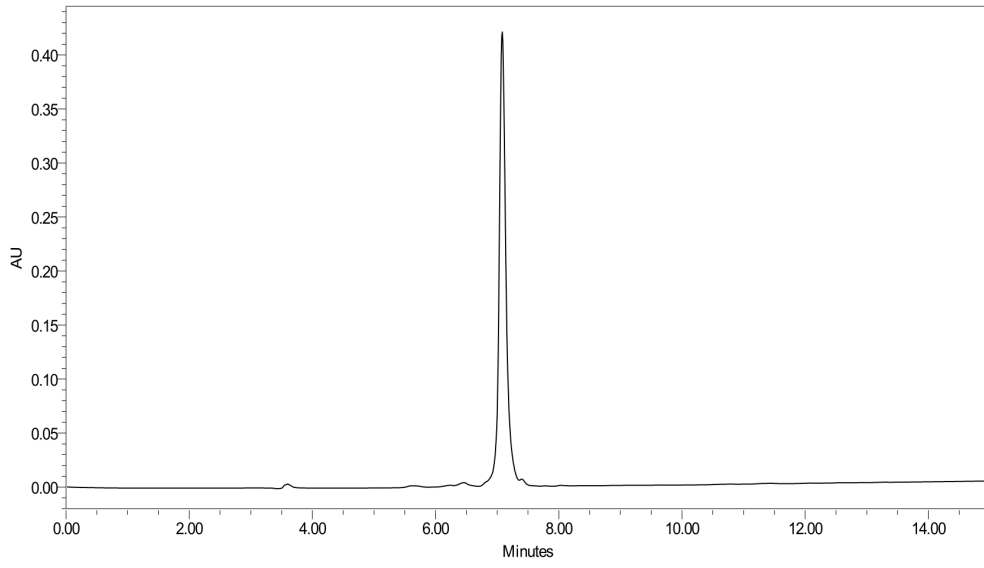
Ac-Leu-Arg-Gly-Gly-ACC

HRMS for $C_{29}H_{41}N_9O_8$ ($m/z_{\text{calcd}} = 644.3151$; $m/z_{\text{found}} = 644.3158$)



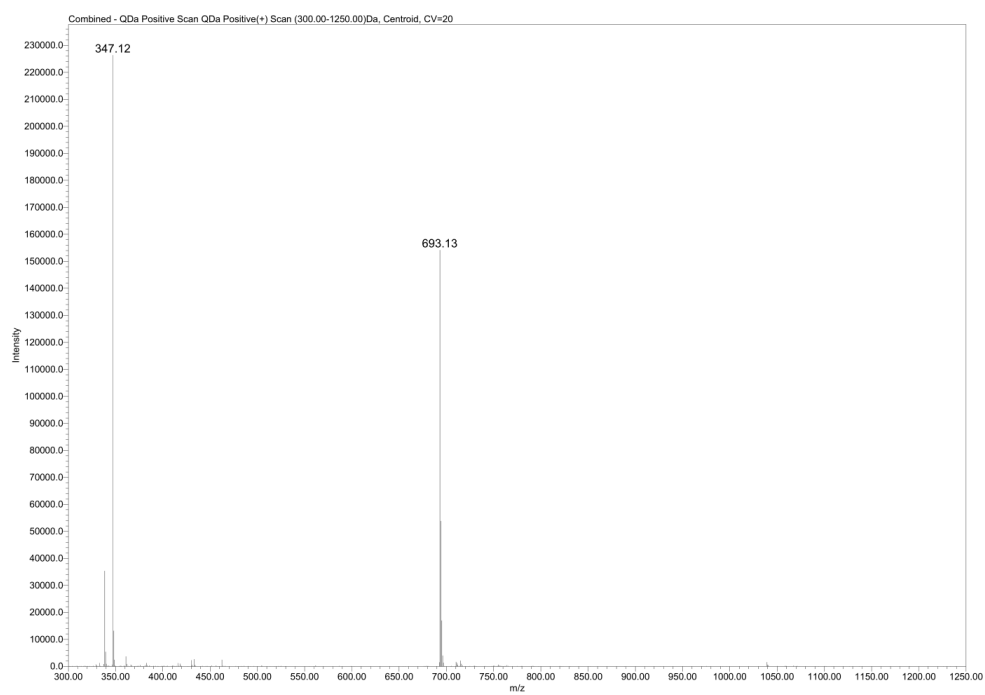
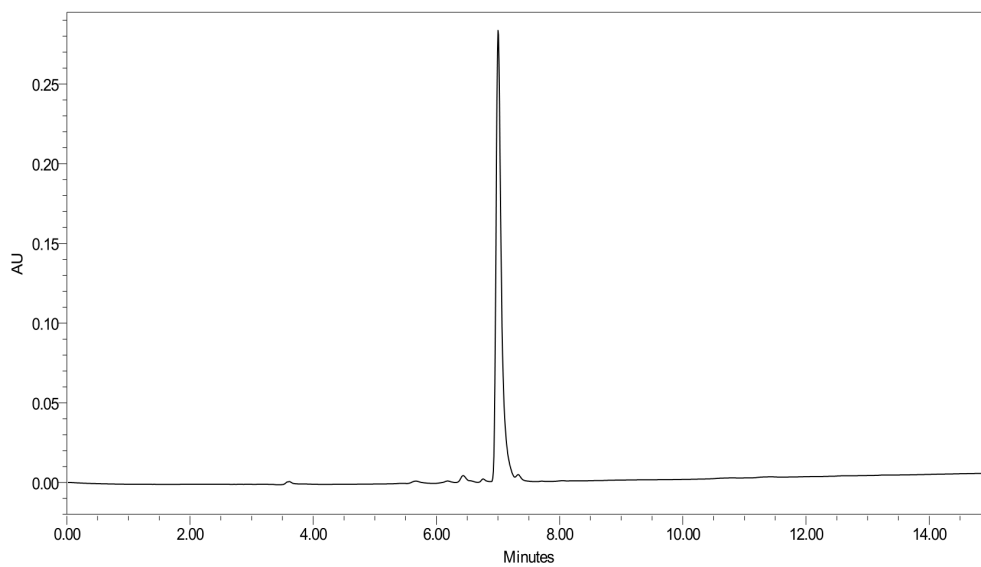
Ac-Abu(Bth)-Dap-Gly-Gly-ACC

HRMS for $C_{31}H_{34}N_8O_8S$ ($m/z_{\text{calcd}} = 679.23$; $m/z_{\text{found}} = 679.12$)



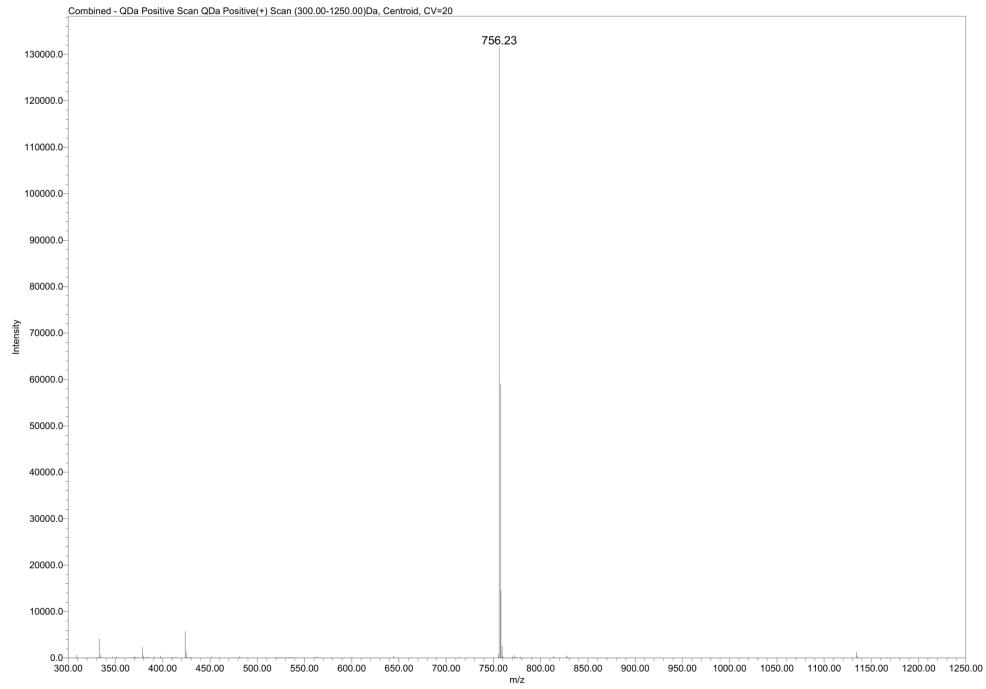
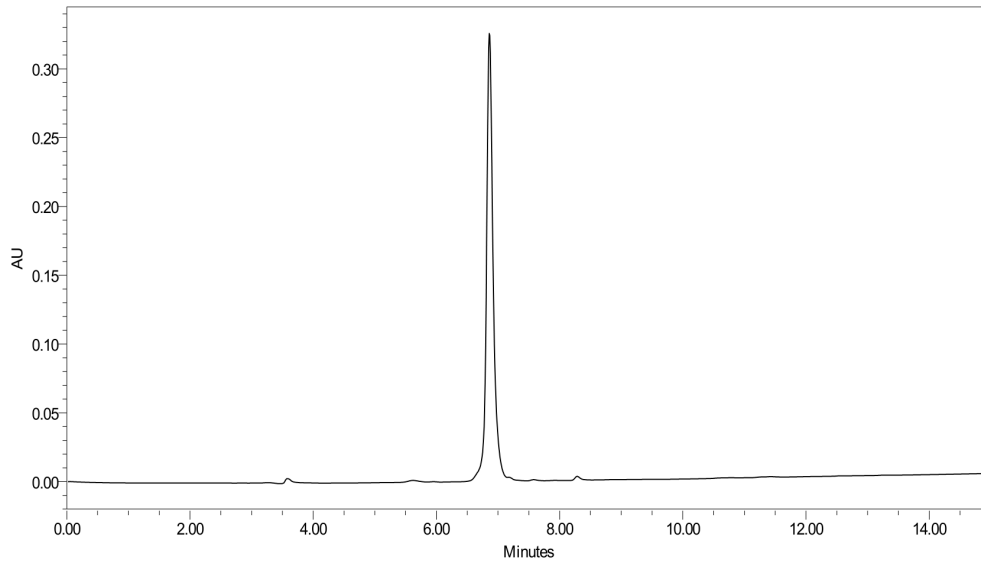
Ac-Abu(Bth)-Dab-Gly-Gly-ACC

HRMS for $C_{32}H_{36}N_8O_8S$ ($m/z_{\text{calcd}} = 693.25$; $m/z_{\text{found}} = 693.13$)



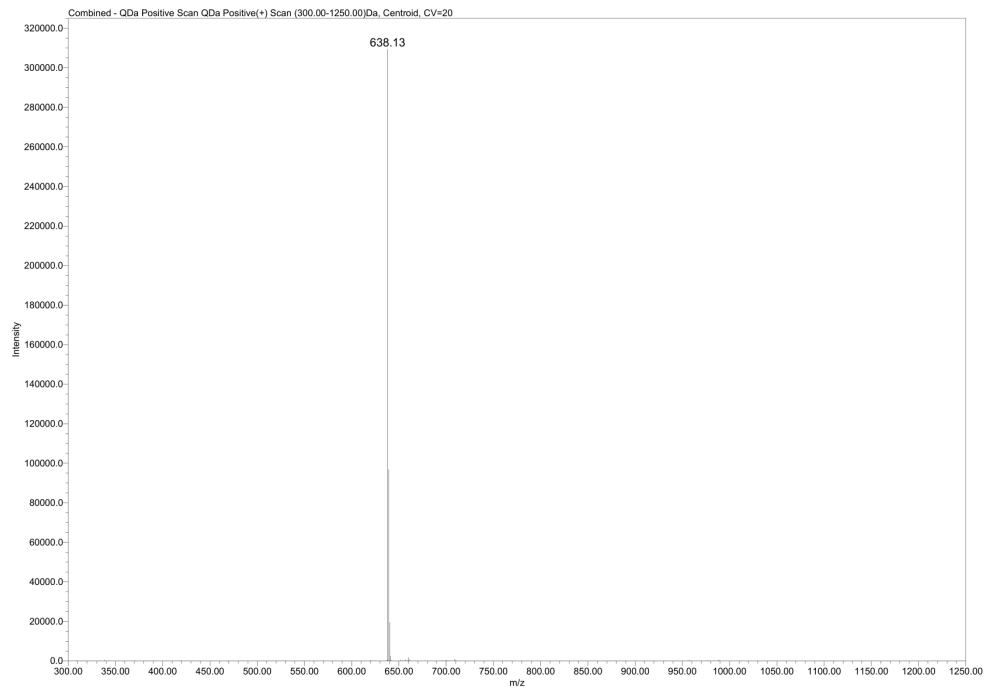
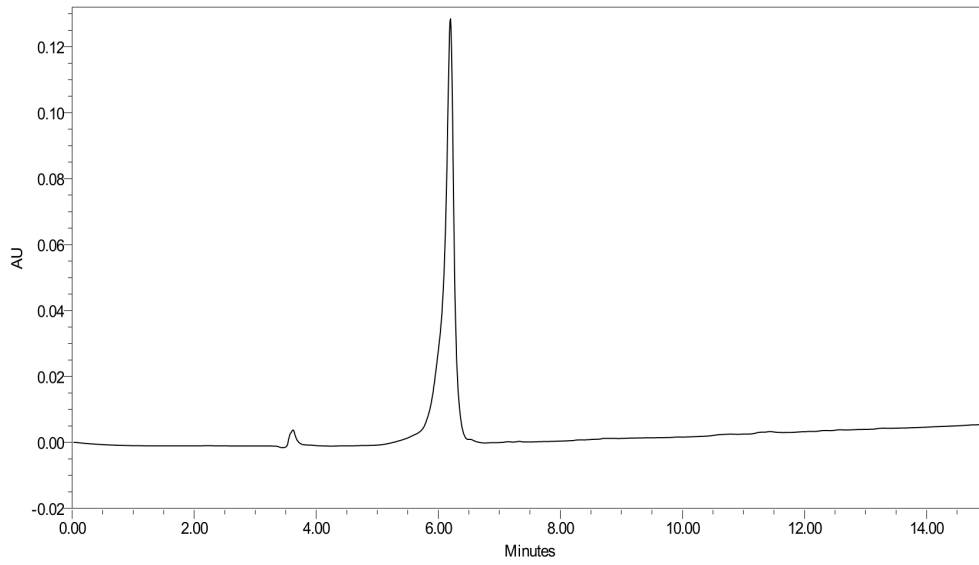
Ac-hTyr-Phe(guan)-Gly-Gly-ACC

HRMS for $C_{37}H_{41}N_9O_9$ ($m/z_{\text{calcd}} = 756.31$; $m/z_{\text{found}} = 756.23$)



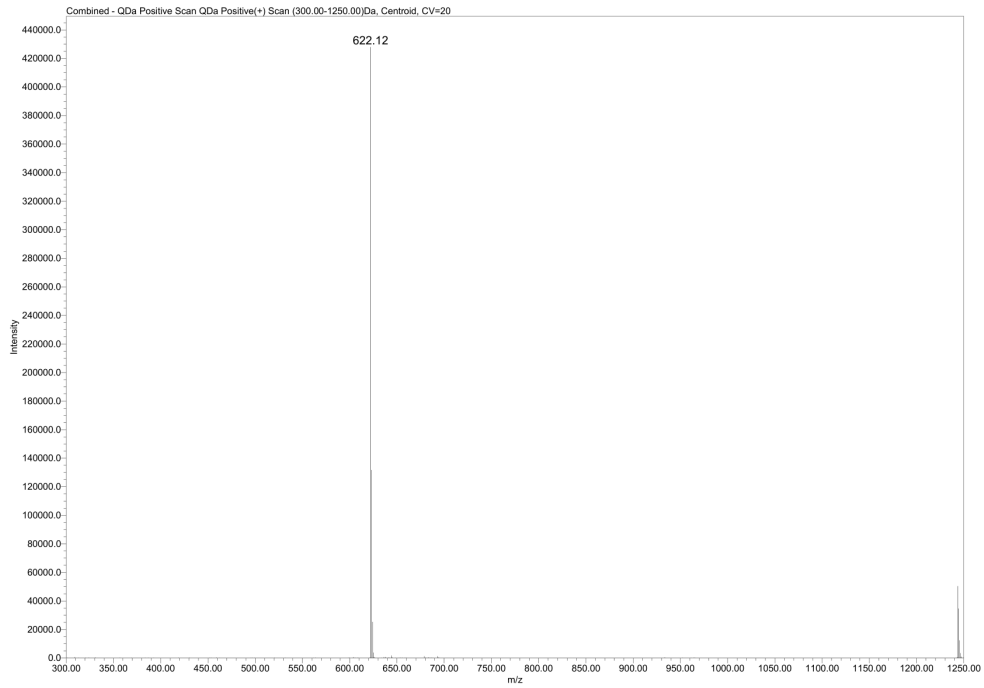
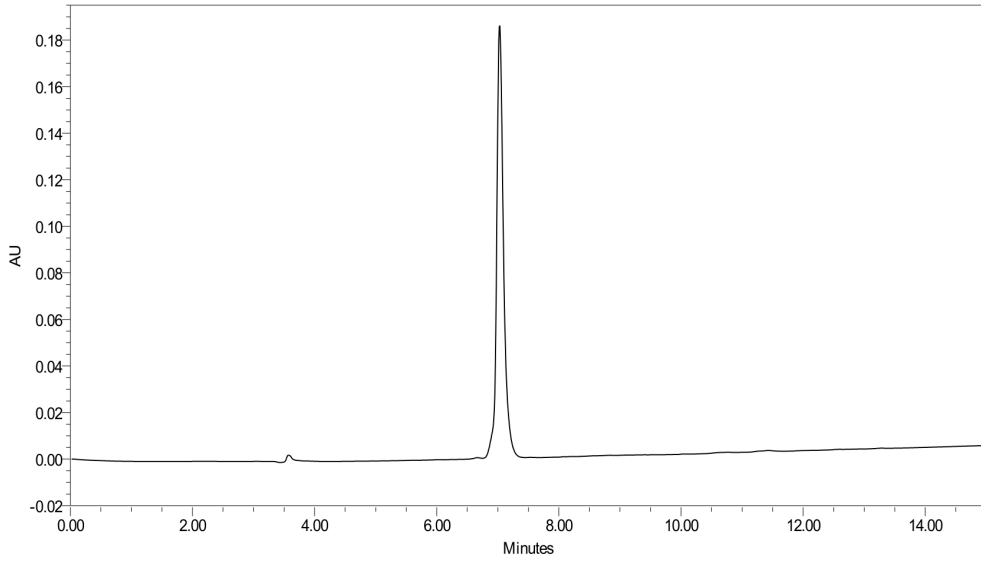
Ac-hTyr-Dap-Gly-Gly-ACC

HRMS for $C_{30}H_{35}N_7O_9$ ($m/z_{\text{calcd}} = 638.26$; $m/z_{\text{found}} = 638.13$)

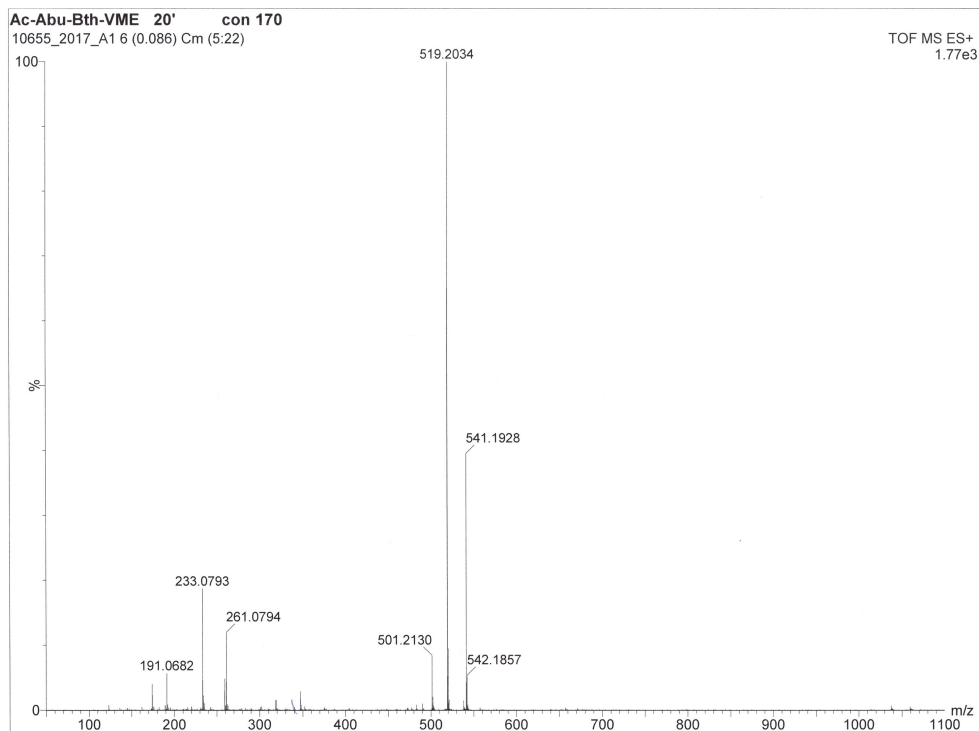
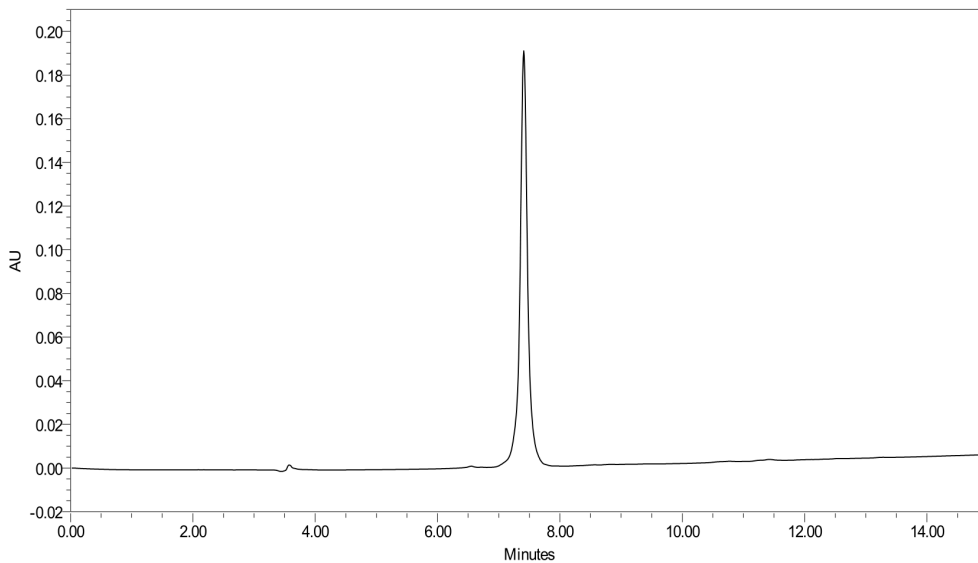


Ac-hPhe-Dap-Gly-Gly-ACC

HRMS for $C_{30}H_{35}N_7O_8$ ($m/z_{\text{calcd}} = 622.26$; $m/z_{\text{found}} = 622.12$)

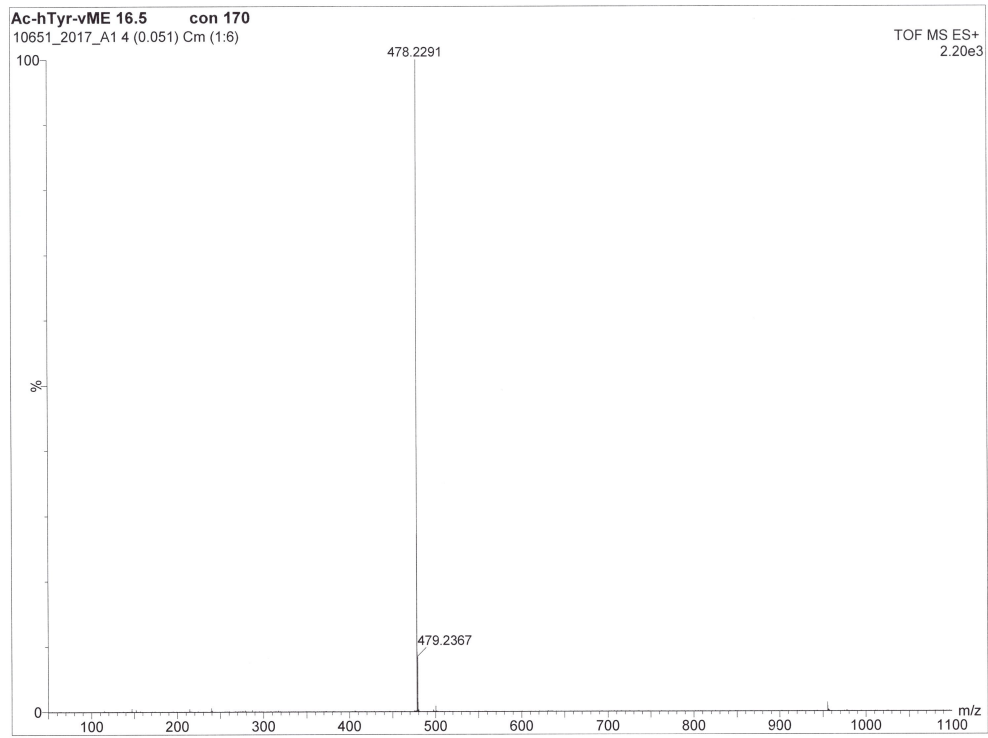
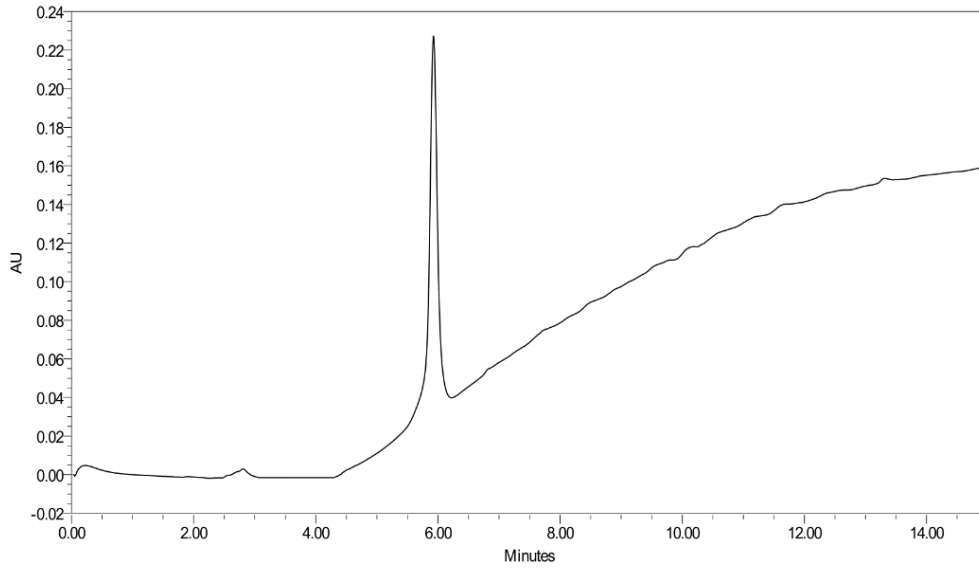


Ac-Abu(Bth)-Dap-Gly-Gly-VME
HRMS for $C_{23}H_{30}N_6O_6S$ ($m/z_{\text{calcd}} = 519.2026$; $m/z_{\text{found}} = 519.2034$)



Ac-hTyr-Dap-Gly-Gly-VME

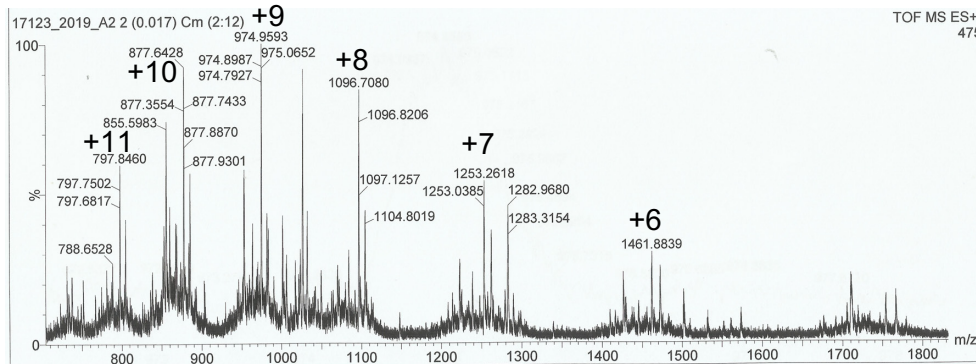
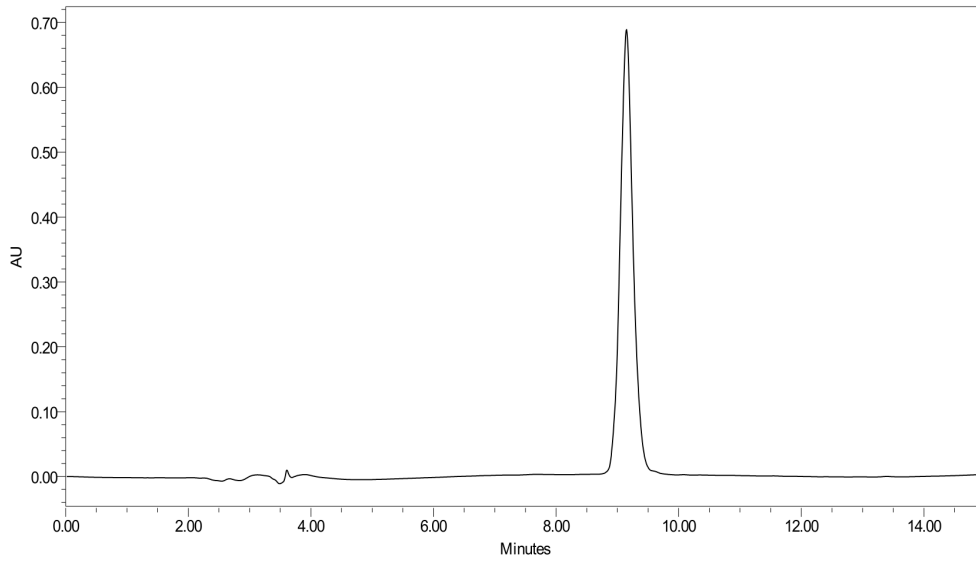
HRMS for $C_{22}H_{31}N_5O_7$ ($m/z_{\text{calcd}} = 478.2302$; $m/z_{\text{found}} = 478.2291$)



Ub-ACC

HRMS for C₃₈₉H₆₃₇N₁₀₇O₁₂₀S

charge	m/z _{calcd}	m/z _{found}
+6	1461.6221	1461.8839
+7	1252.9628	1253.2618
+8	1096.4684	1096.7080
+9	974.7505	974.9593
+10	877.3762	877.6428
+11	797.7063	797.8460



Biot-6-Ahx-Ub-VME

HRMS for $C_{397}H_{658}N_{108}O_{121}S_2$

charge	m/z _{calcd}	m/z _{found}
+6	1491.4778	1491.8063
+7	1278.5535	1278.8268
+8	1118.8602	1119.2485
+9	994.6543	994.8820
+10	895.2896	895.3971
+11	813.9912	814.0791
+12	746.2426	746.3508

