

Supplementary Materials for  
**Identification of 113 new histone marks by CHiMA, a tailored database  
search strategy**

Jinjun Gao *et al.*

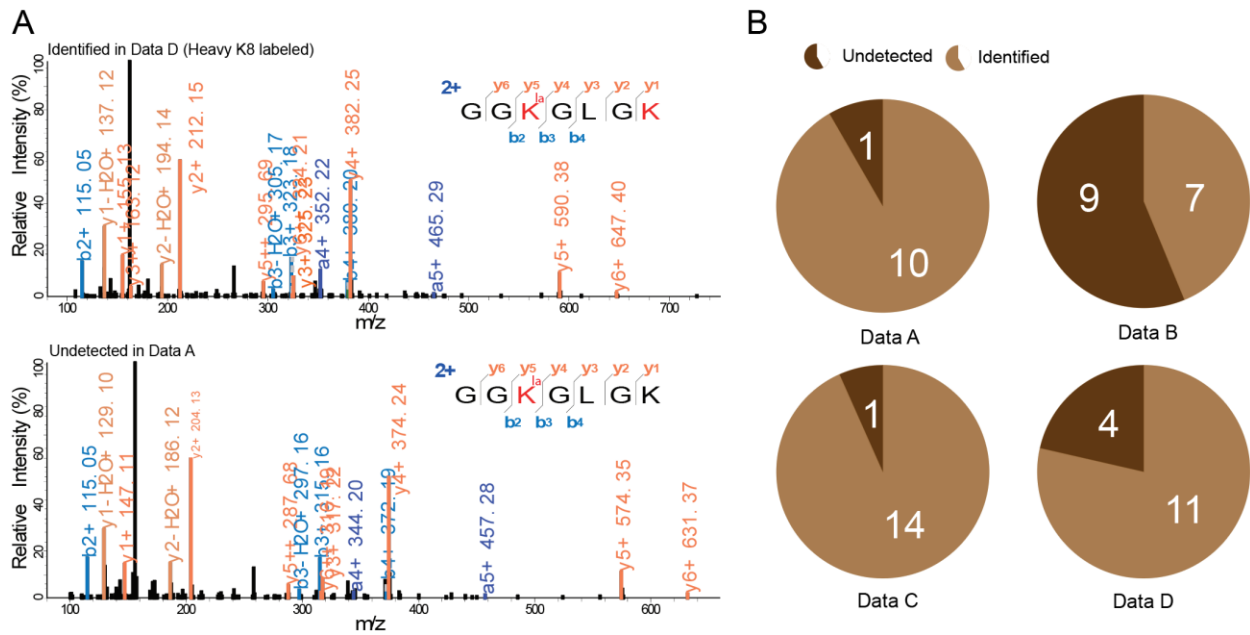
Corresponding author: Chu Wang, [chuwang@pku.edu.cn](mailto:chuwang@pku.edu.cn); Yingming Zhao, [yingming.zhao@uchicago.edu](mailto:yingming.zhao@uchicago.edu)

*Sci. Adv.* **9**, eadf1416 (2023)  
DOI: 10.1126/sciadv.adf1416

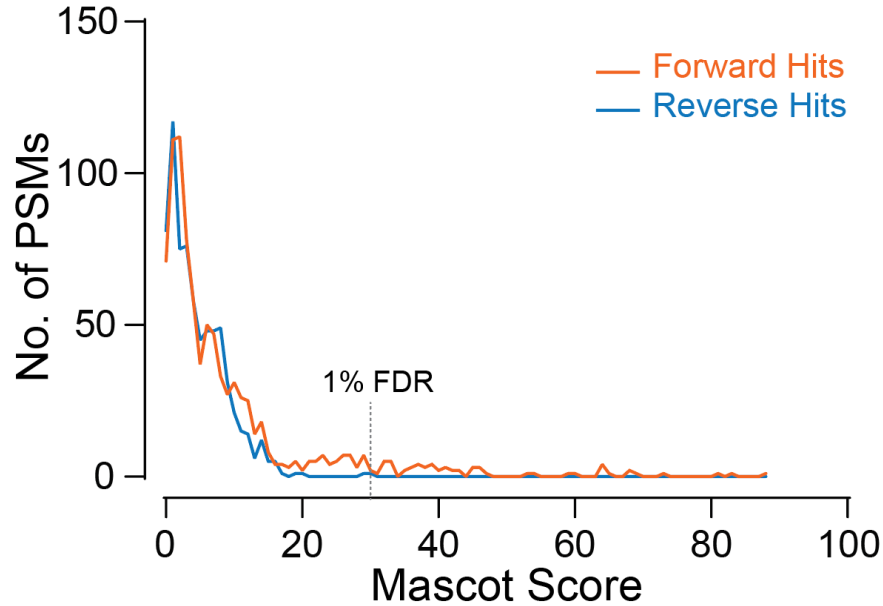
**This PDF file includes:**

Figs. S1 to S6  
Tables S1 to S4  
Data S1 to S6

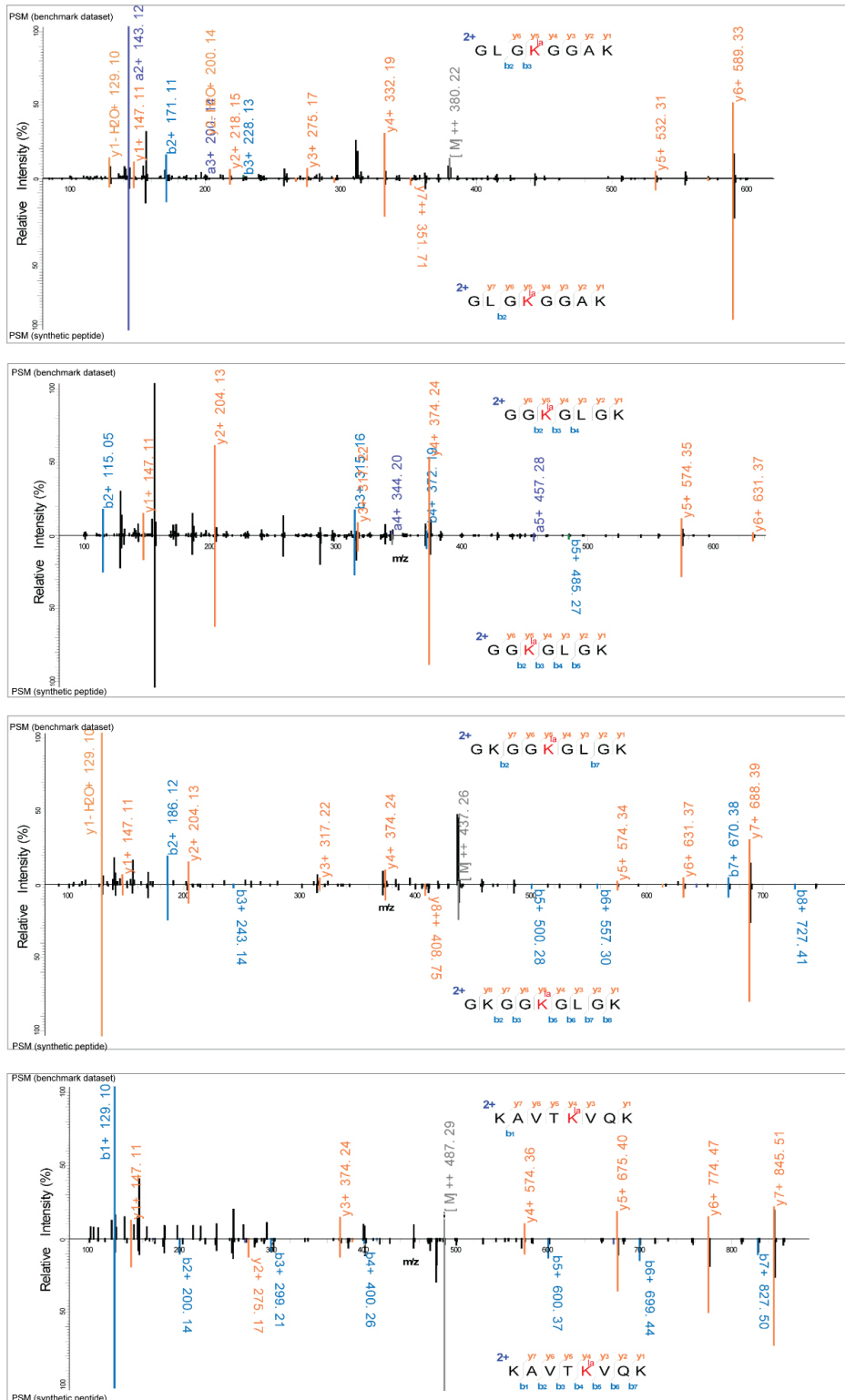
## I. Supplemental figures



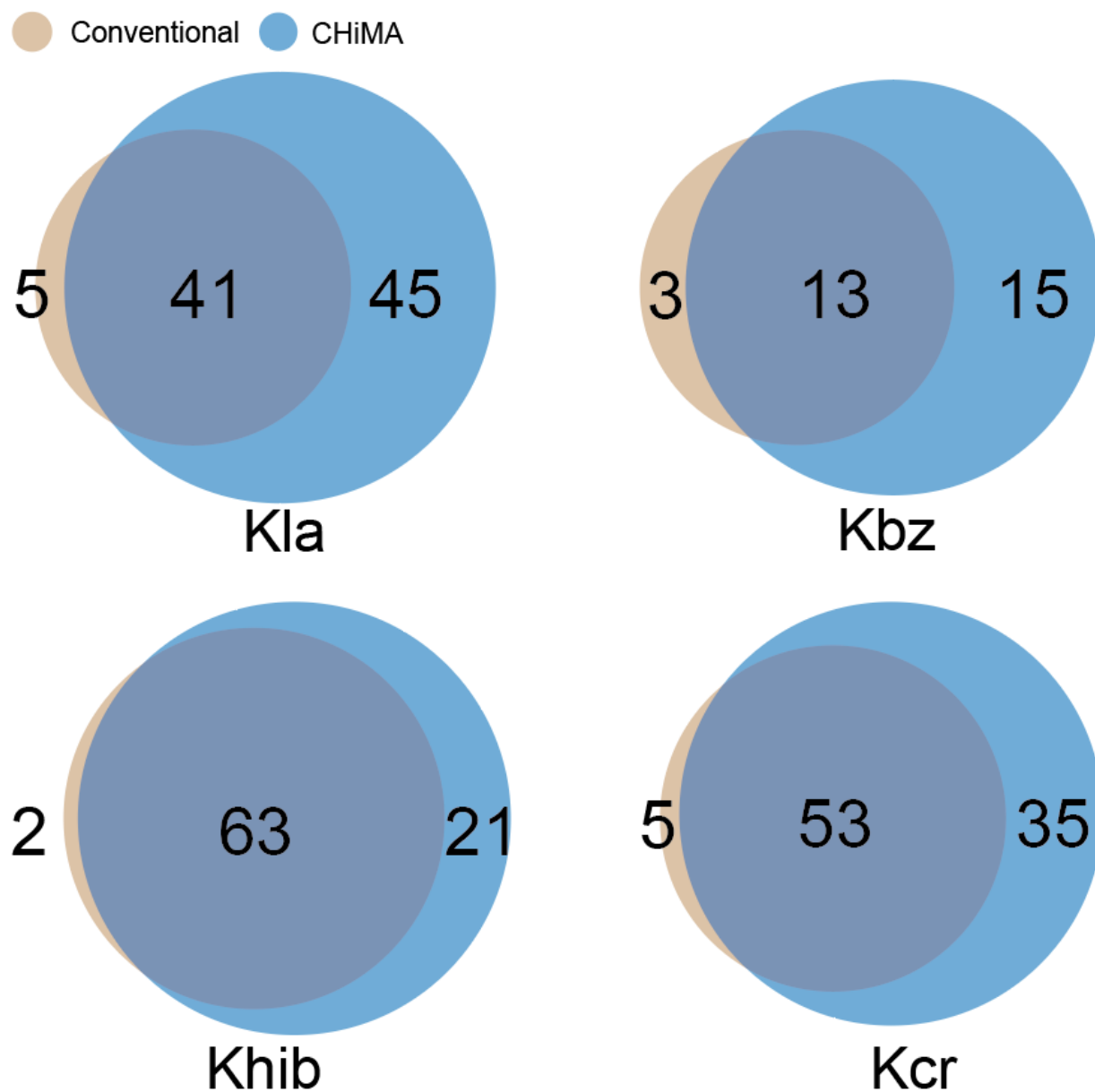
**Figure S1, related to Figure 2A. Conventional search strategies fail to detect some K1a peptides. (A)** MS/MS spectra of a representative K1a peptide that was detected in Dataset D but not in Dataset A. **(B)** Pie charts of K1a peptides in each dataset that Andromeda failed to detect.



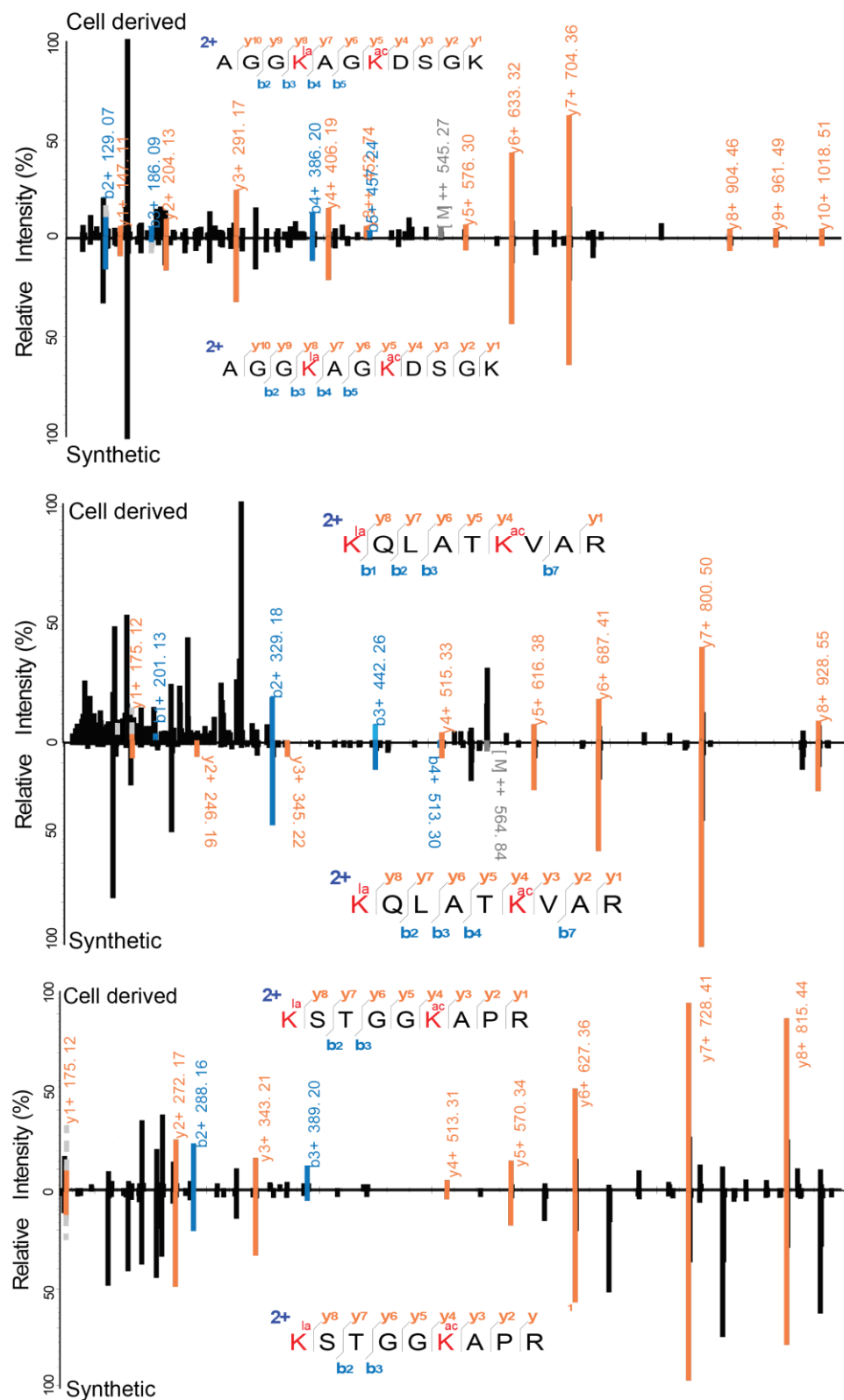
**Figure S2, related to Figure 2B. Similar score distributions of target and decoy hits for Dataset A analyzed by MASCOT.** Dataset A was searched by Mascot against a reverse-concatenated database. All PSMs were extracted by setting the significance threshold p-value to 1. The score distributions for forward (orange) and reverse PSMs (blue) are similar. The manually calculated 1% FDR threshold was indicated by a dashed gray line. The score of the top-ranked reverse peptide PSM was 30, and only 59 forward peptide PSMs scored larger than this threshold.



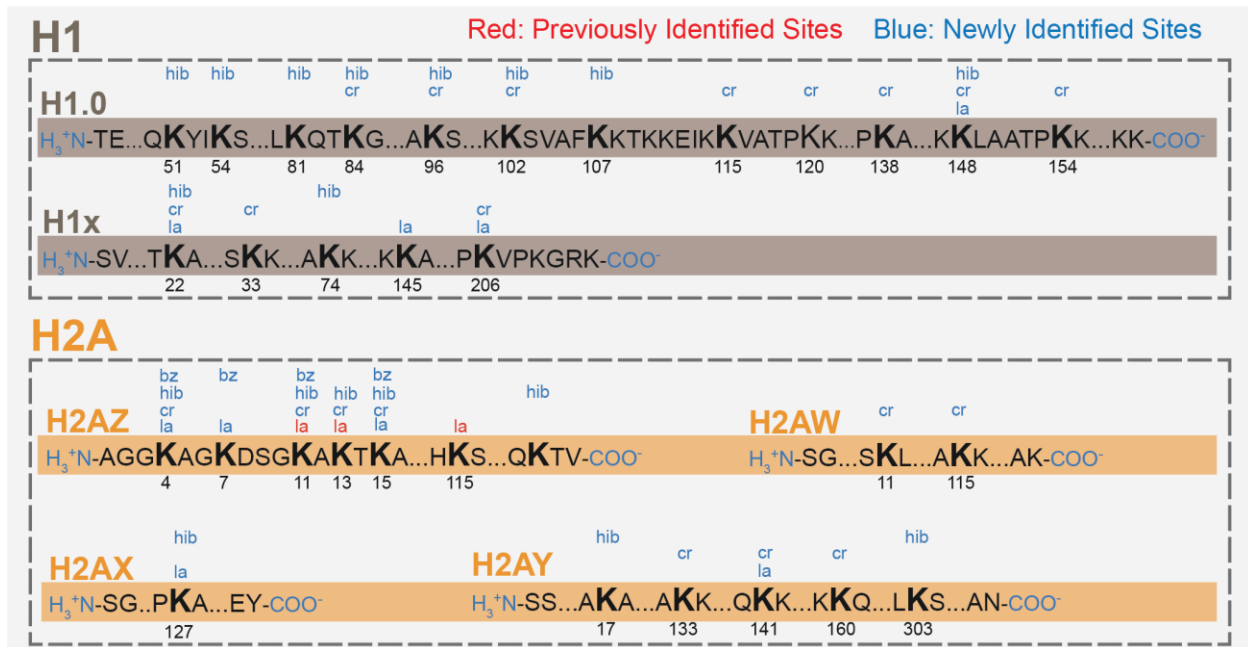
**Figure S3, related to Figure 3D. The MS/MS spectra of K1a peptides identified in Dataset A by the 50% b/y ion coverage strategy perfectly match those of the corresponding synthetic peptides.**



**Figure S4, related to Figure 6. Venn diagrams of modification sites identified by conventional strategy and CHiMA for analyzing previous Kla, Kbz, Khib, and Kcr datasets.**



**Figure S5, related to Figure 6. The MS/MS spectra of peptides identified with coexisting histone marks perfectly match those of the corresponding synthetic peptides.**



**Figure S6, related to Figure 7. A comprehensive map of Kla, Kbz, Kcr, and Khib on noncanonical histones.** A total of 54 histone marks were detected on noncanonical replication-independent histones, with 51 marks newly identified. bz: benzoylation; cr: crotonylation; hib: 2-hydroxyisobutyrylation; la: lactylation

## II. Supplemental tables

**Table S1 Xcorr and DeltaCN scores of PSMs in the four test datasets that were filtered out by false discovery rate screening.**

Sequence	Dataset	MS/MS Scan No.	Xcorr	DeltaCN
GLGK*GGAK	A	1377	1.3556	0.3489
AVTK*AQK	A	986	1.8648	0.1504
AVTK*AQK	B	965	1.2928	0.0283
GGK*GLGK	B	1259	1.9045	0.1035
KAVTK*VQK	B	1083	1.8579	0.1468
GKGGK*GLGK	B	959	1.5376	0.0138
GLGK*GGAK	B	1417	1.3659	0.052
SAPAPK*K	C	1527	1.3665	0.1915
AVTK*VQK	C	1992	1.1913	0.2776
GLGK*GGAK	C	2244	1.2545	0.3282
AVTK*AQK	C	1510	1.7691	0.1674
GGK*GLGK	D	1686	1.7433	0.1908
SAPAPK*K	D	1280	1.4292	0.231
AVTK*VQK	D	1670	1.1248	0.1848
LAHYNK*R	D	3057	1.0395	0.1842



**Table S2 The true positive and false positive rates using various cutoff values for dataset A, including FDR values from 1% to 100% and fragment coverage values from 100% to 0%.**

<b>FDR</b>	<b>True positive rate</b>	<b>False positive rate</b>
<b>0.01</b>	0.42	0
<b>0.05</b>	0.58	0
<b>0.1</b>	0.58	0
<b>0.15</b>	0.58	0.01
<b>0.2</b>	0.58	0.01
<b>0.25</b>	0.67	0.01
<b>0.3</b>	0.75	0.02
<b>0.35</b>	0.75	0.04
<b>0.4</b>	0.75	0.05
<b>0.45</b>	0.75	0.05
<b>0.5</b>	0.83	0.07
<b>0.55</b>	0.92	0.13
<b>0.6</b>	0.92	0.18
<b>0.65</b>	1.00	0.23
<b>0.7</b>	1.00	0.34
<b>0.75</b>	1.00	0.77
<b>0.8</b>	1.00	0.9
<b>0.85</b>	1.00	0.92
<b>0.9</b>	1.00	0.95
<b>0.95</b>	1.00	0.98
<b>1</b>	1.00	1

<b>FIC</b>	<b>True positive rate</b>	<b>False positive rate</b>
<b>1</b>	0	0
<b>0.95</b>	0	0
<b>0.9</b>	0.08	0
<b>0.85</b>	0.08	0
<b>0.8</b>	0.17	0
<b>0.75</b>	0.33	0
<b>0.7</b>	0.42	0
<b>0.65</b>	0.75	0
<b>0.6</b>	0.83	0.01
<b>0.55</b>	0.92	0.01
<b>0.5</b>	1	0.02
<b>0.45</b>	1	0.02
<b>0.4</b>	1	0.05
<b>0.35</b>	1	0.08
<b>0.3</b>	1	0.12
<b>0.25</b>	1	0.19
<b>0.2</b>	1	0.28
<b>0.15</b>	1	0.45
<b>0.1</b>	1	0.68
<b>0.05</b>	1	0.92
<b>0</b>	1	1

**Table S3 Chemically synthesized peptides.** Five K1a peptides (green) undetected by the target-decoy strategy in Dataset A but identified by the 50% b/y ion coverage strategy were synthesized. MS/MS spectra of these peptides are shown in Figures 2E and S3. The remaining 18 peptides (black) were identified with moderate confidence for K1a and Kcr in the other datasets. MS/MS spectra of these peptides are shown in Figures 5D and Data S5.

Sequence	Histone	mass	m/z
AVTK(la)AQK	H2B	816.4705	409.2425
GGK(la)GLGK	H4	687.3915	344.703
GKGGK(la)GLGK	H4	872.508	437.2613
GLGK(la)GGAK	H4	758.4287	380.2216
KAVTK(la)VQK	H2B	972.5968	487.3057
GTGASGSFK(la)LNK	H1	1237.63	619.8223
K(la)VAGAATPK	H1	913.5233	457.7689
PSVPK(la)VPK	H1	922.5488	462.2817
K(la)LAATPK	H1	799.4804	400.7475
K(la)TSATVGPK	H2A	959.5288	480.7717
TSATVGPK(la)APSGGK	H2A	1328.694	665.3543
K(la)ESYSVYVYK	H2B	1336.655	669.3348
VTIMPK(la)DIQLAR	H3	1455.812	728.9133
K(cr)AGGTPR	H1	753.4133	377.7139
KPVGAAGK(cr)K	H1	865.5385	433.7765
AAK(cr)SAAK	H1	713.4072	357.7109
AVKPK(cr)AAK	H1	879.5542	440.7844
K(cr)PAAAAGAK	H1	851.4865	426.7505
QTK(cr)GVGASGSFR	H1	1261.642	631.8283
VATPK(cr)K	H1	710.4327	356.2236
AGGSAALSPSK(cr)K	H1	1140.614	571.3143
PSVPK(cr)VPK	H1	918.5539	460.2842
IHPELLAK(cr)K	H2A	1115.67	558.8423

**Table S4 Co-existing histone marks.** Homologous sites of canonical histone subtypes were integrated, with their associated histone variant indicated.

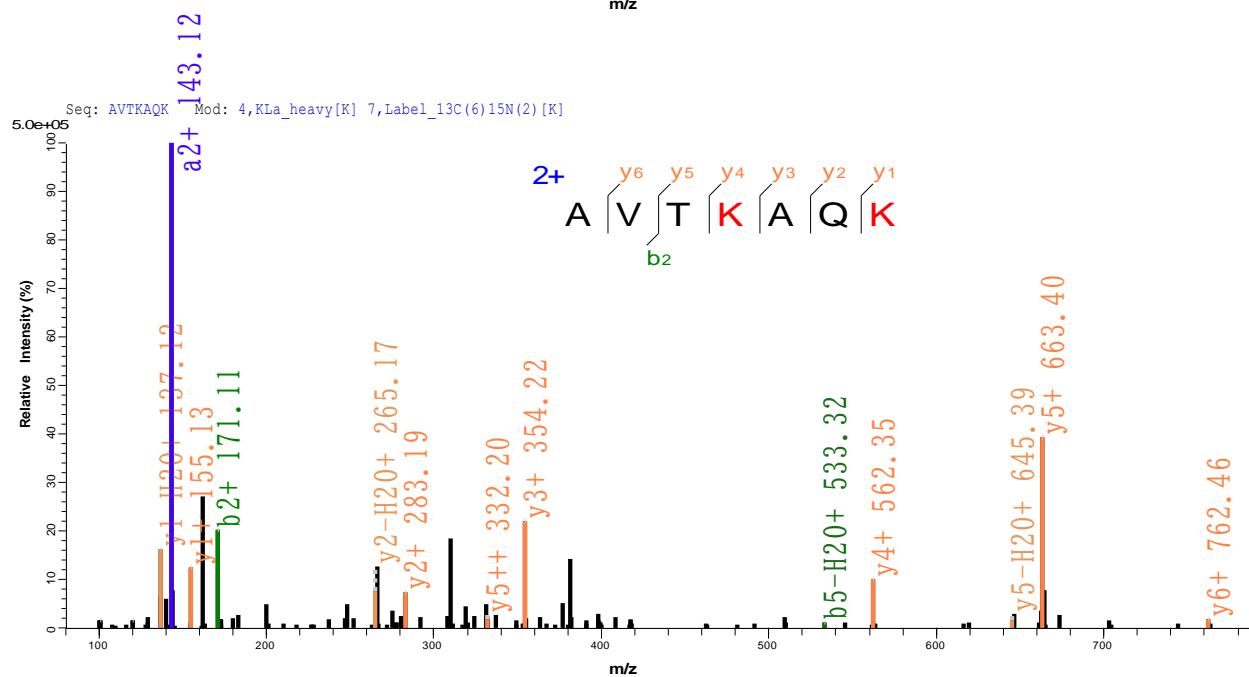
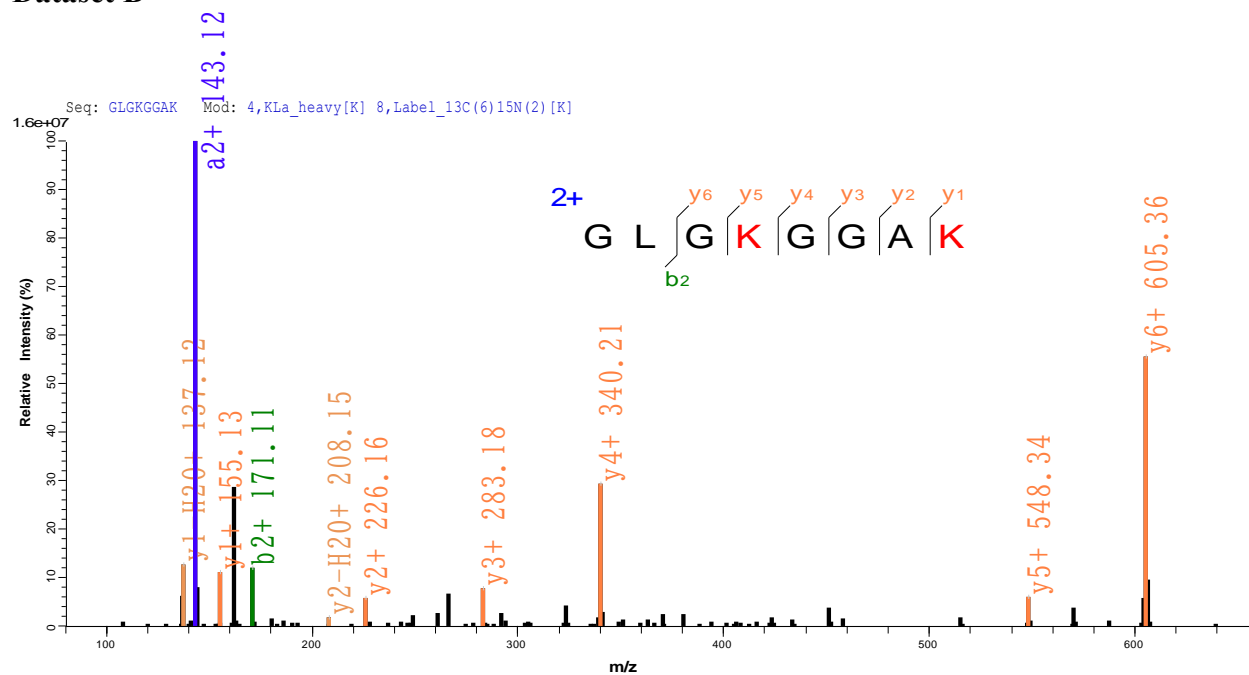
H1	H2A	H2AZ	H2B	H3	H4
K33la, R34me1	K5la, K9ac	K4la, K7ac	K5ac, K11la	K9la, K14ac	K5la, K8ac
K45la, K51me1	K95la, R99me1	K7ac, K11la	K5ac, K11la, K16ac	K9ac, K14la	K5ac, K8la
K51la, R53me1	K5ac, K9hib	K7la, K11ac	K5la, K11ac, K16ac	K27la, K36me1	K8lac, K12ac
K62la, K63me1	K5ac, K9ce	K13ac, K15la	K11la, R12me1	K18la, K23ac	K5la, K12ac
K74la, R78me1	K5ce, K9ac	K4hib, K7ac	K11ac, K12ac, K15la	K9hib, K14ac	K12ac, K16la, R17me1
K147la, K148me1		K4ac, K7hib	K15la, K16ac	K18hib, K23ac	K77la, R78me1
K167la, K168me1		K7ac, K11hib	K15ac, K16la	K18ac, K23hib	R78me1, K79la
		K11ac, K13hib	K16ac, K20la	K9me1, K14cr	K91la, R92me1
		K13ac, K15hib	K16la, K20ac	K9cr, K14ac	K5hib, K8ac
		K4ce, K7ac	K16la, K20la	K9ac, K14cr	K8hib, K12ac
		K7ac, K11cr	K20la, K23ac	K18cr, K23ac	K12hib, K16ac
		K11ac, K13cr	K20ac, K23la	K18ac, K23cr	K8cr, K8ac
			K43la, K46me1	K18cr, K23ac	K5ac, K8cr
			K85la, R86me1	K18cr, K23cr	K8cr, K12ac
			K5ac, K11hib	K9bz, K14ac	K8ac, K12cr
			K20ac, K23hib	K18bz, K23ac	K12cr, K16cr
			K15ac, K15hib		K12cr, K16cr
			K16ac, K20hib		
			K11cr, K12ac		
			K16ac, K20cr		
			K5ac, K11cr		
			K16cr, K20ac		
			K15ac, K16cr		
			K20ac, K23cr		
			K16cr, K20cr		

### III. Supplementary Data

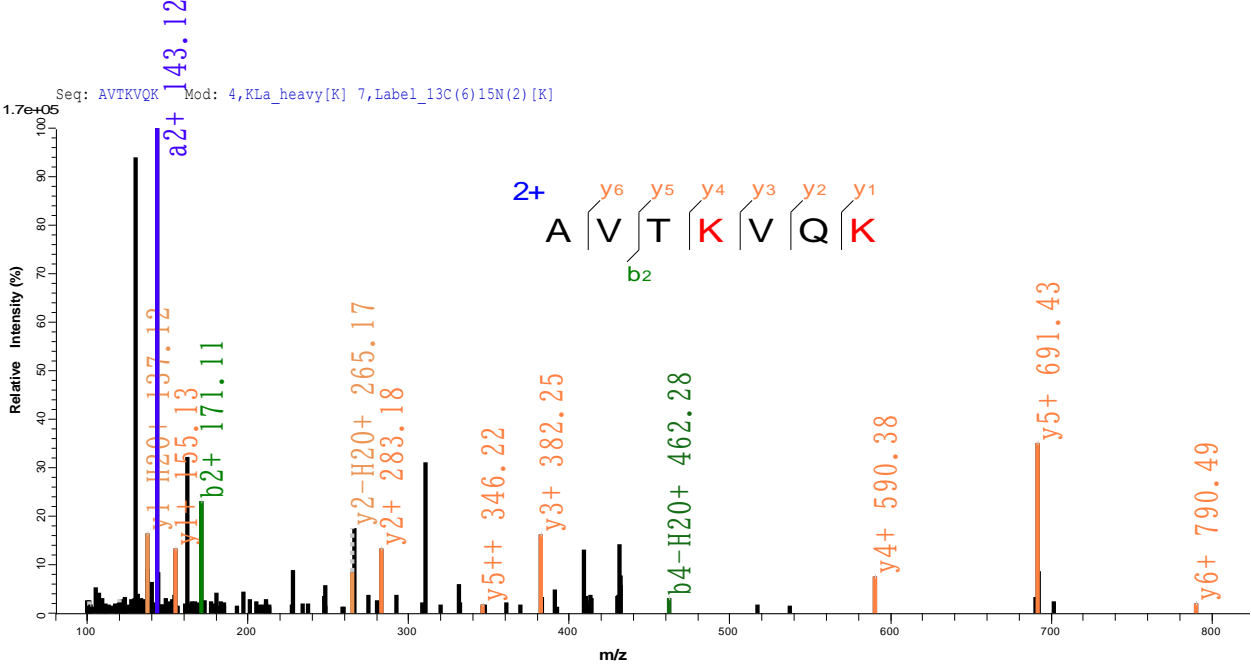
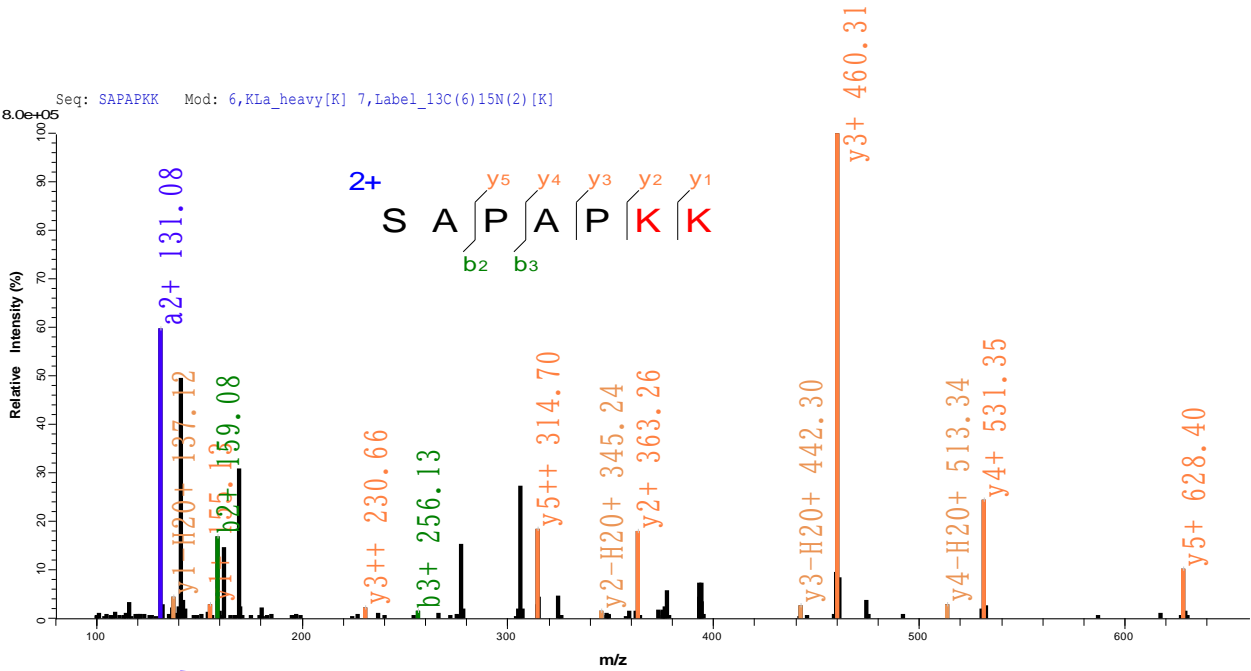
**Data S1. KLa peptide PSMs in Datasets A through E that were missed by the conventional database search strategy but identified by the strategy of 50% b/y ion coverage.**

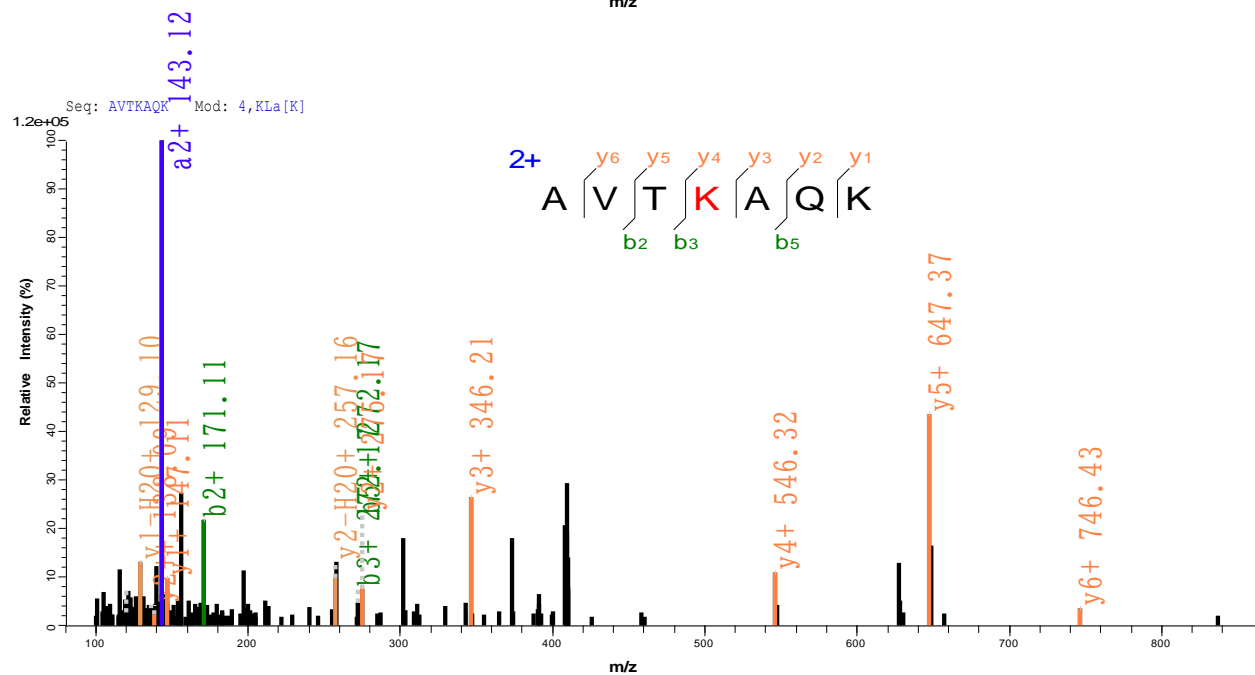
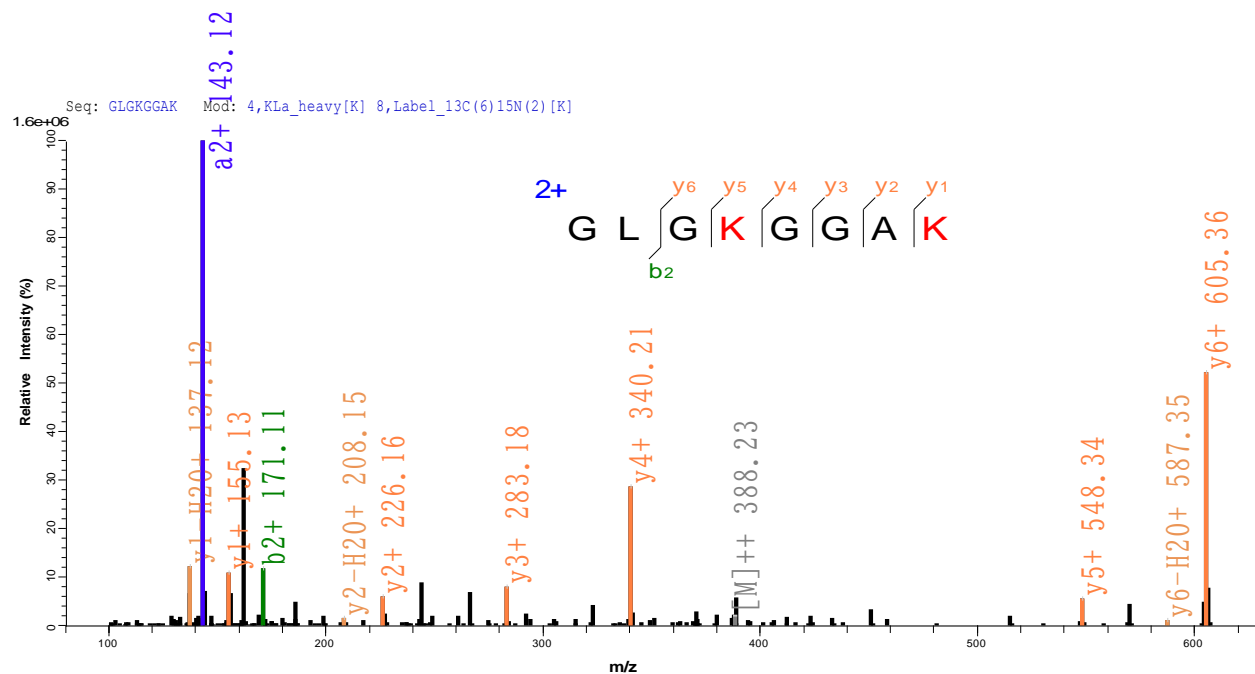
PSMs for Dataset A are shown in Fig. S3.

#### Dataset B

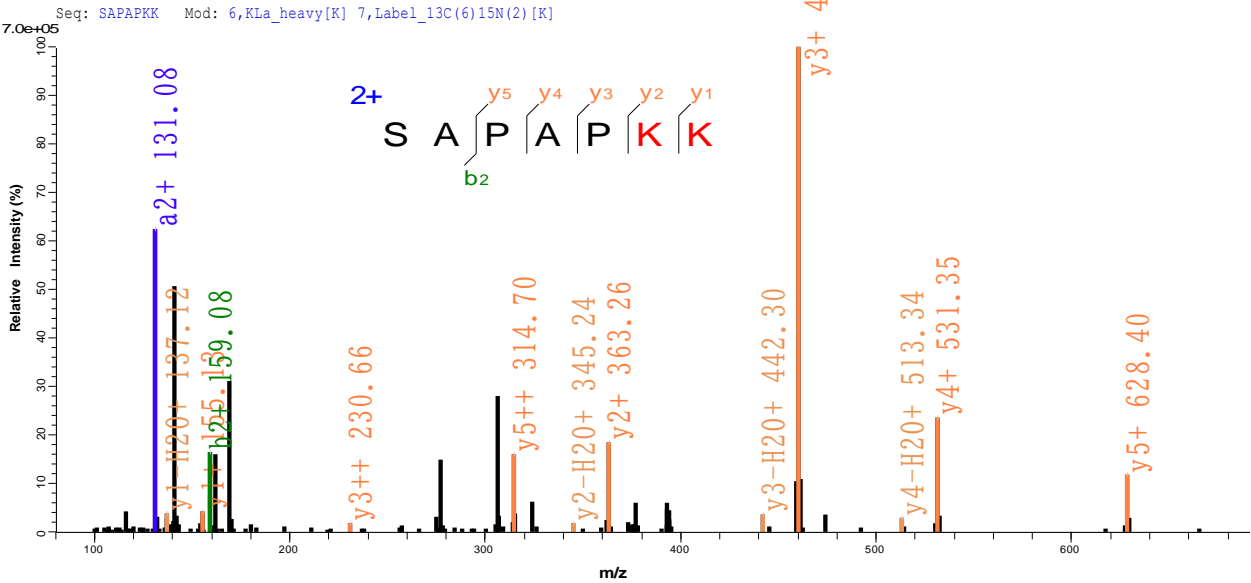
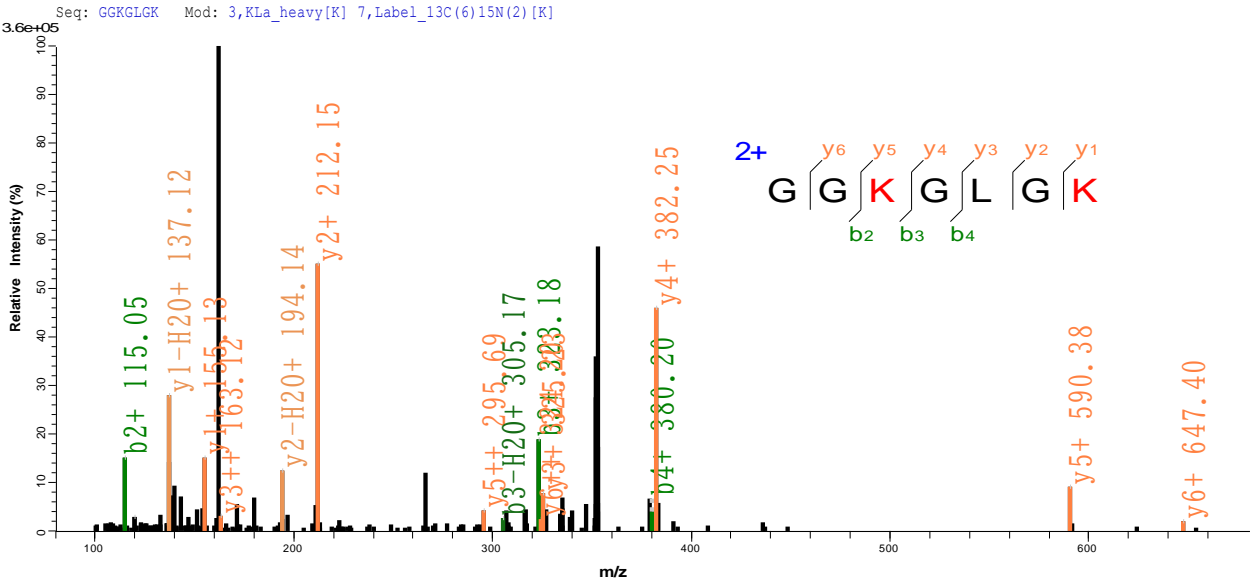


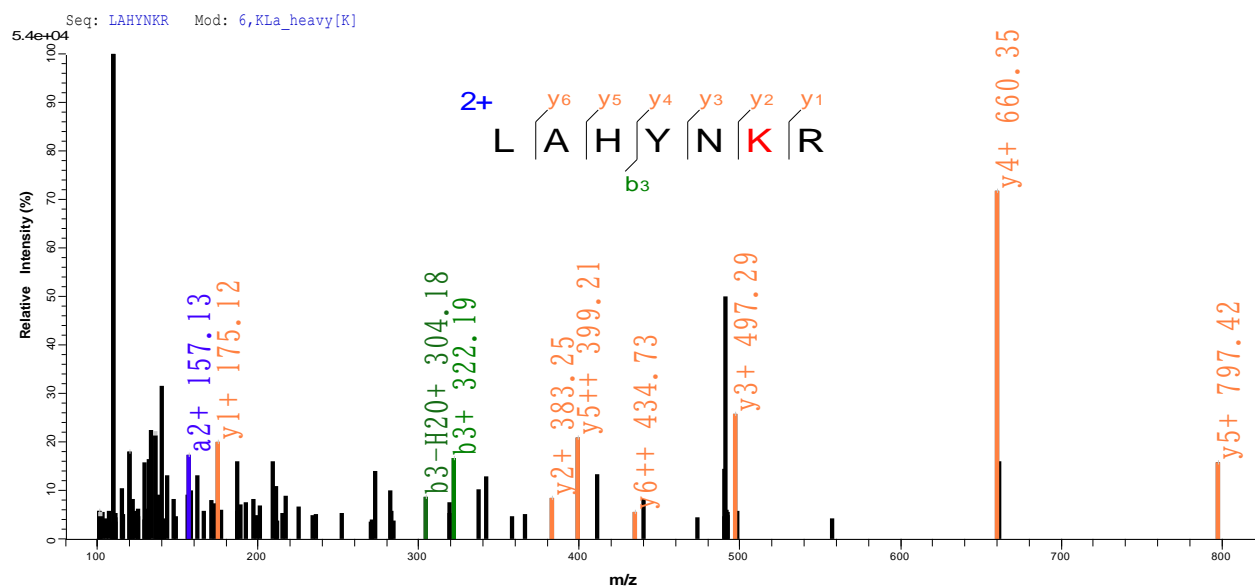
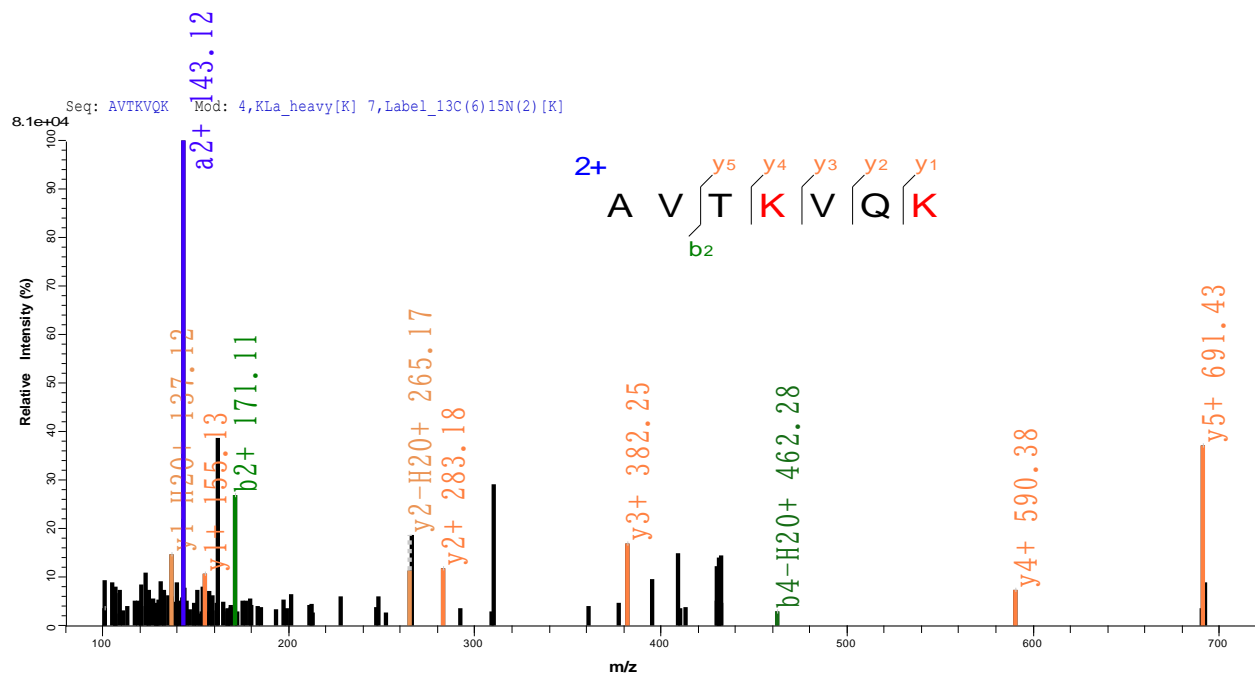
Dataset C





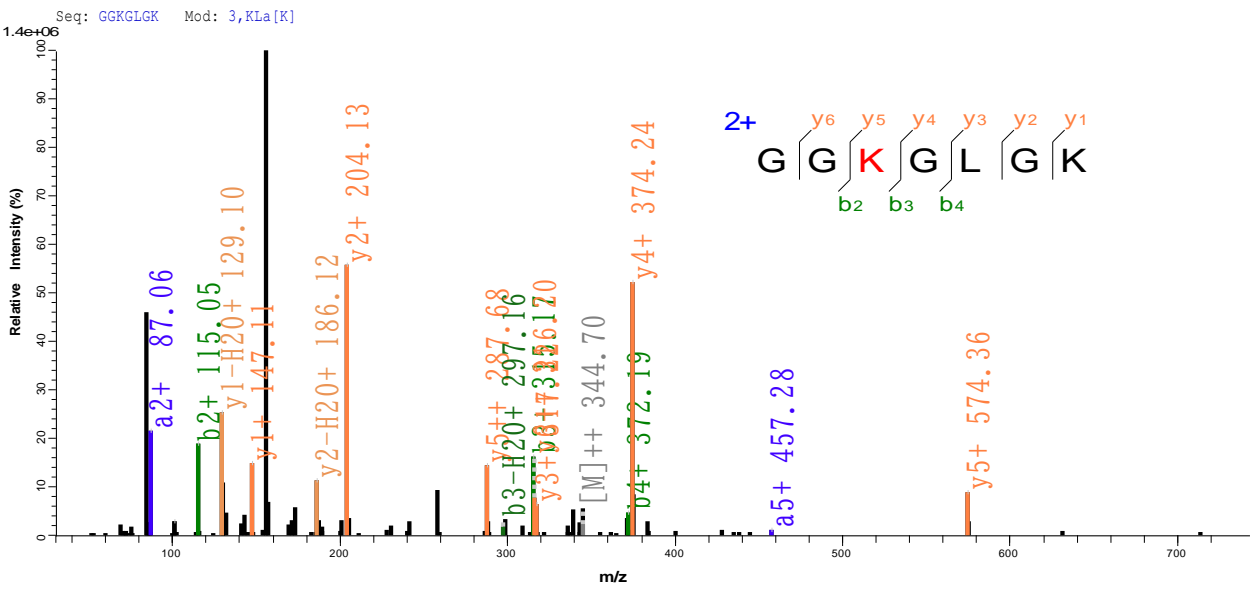
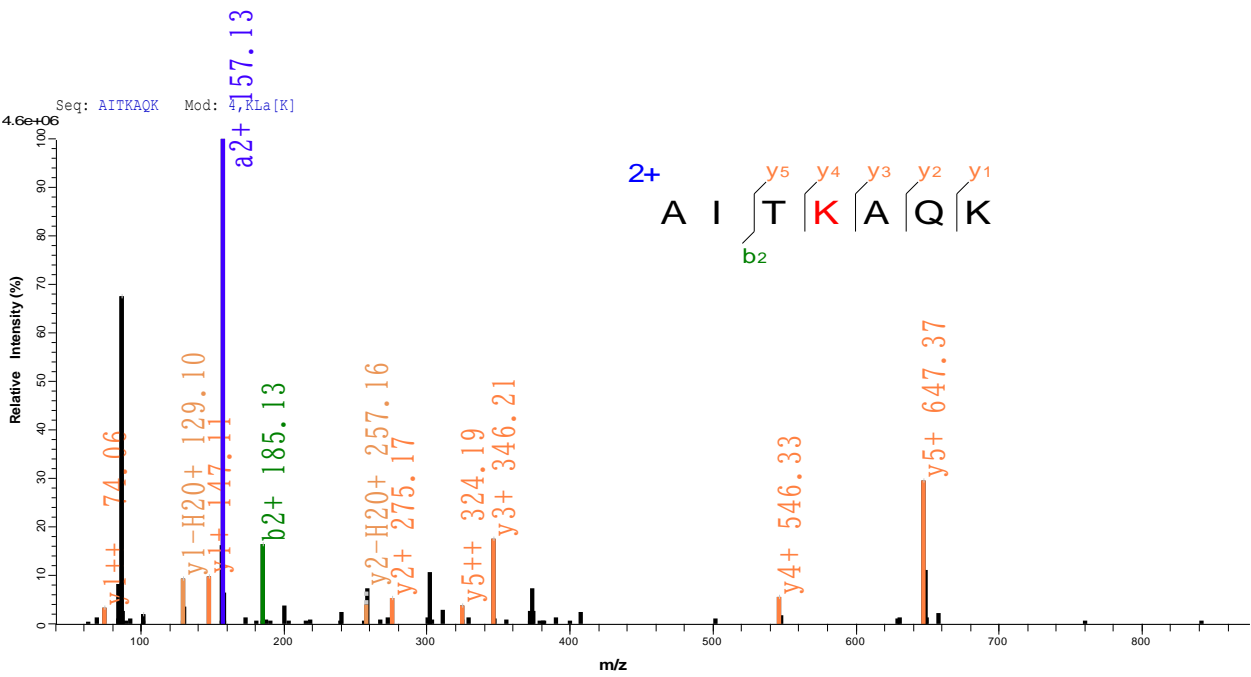
Dataset D

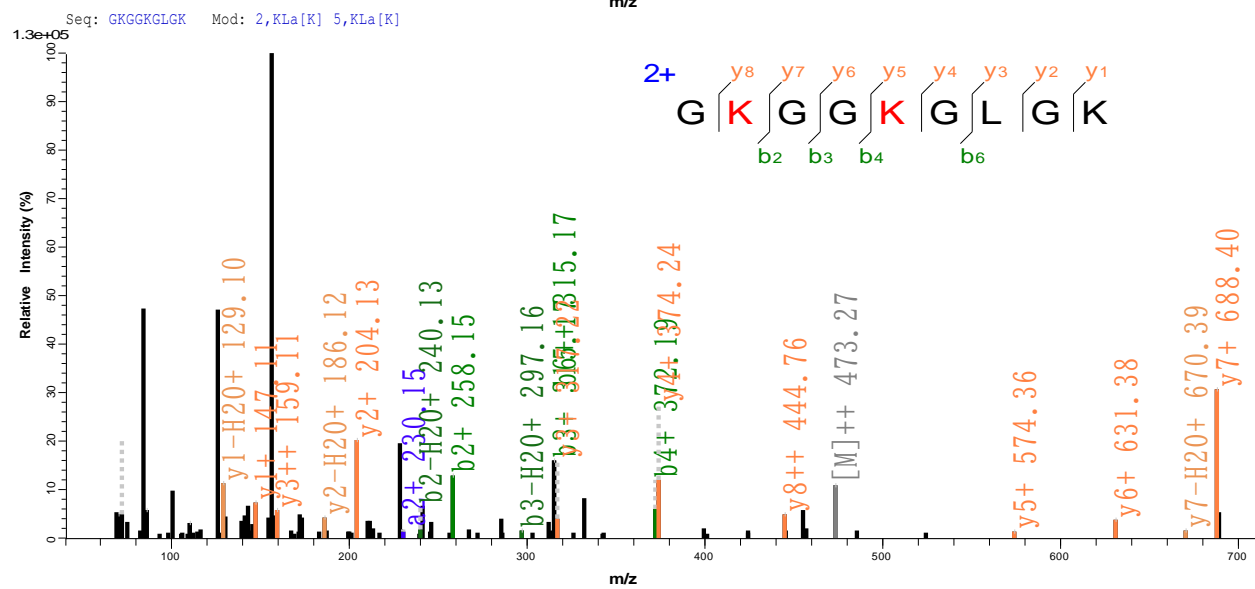
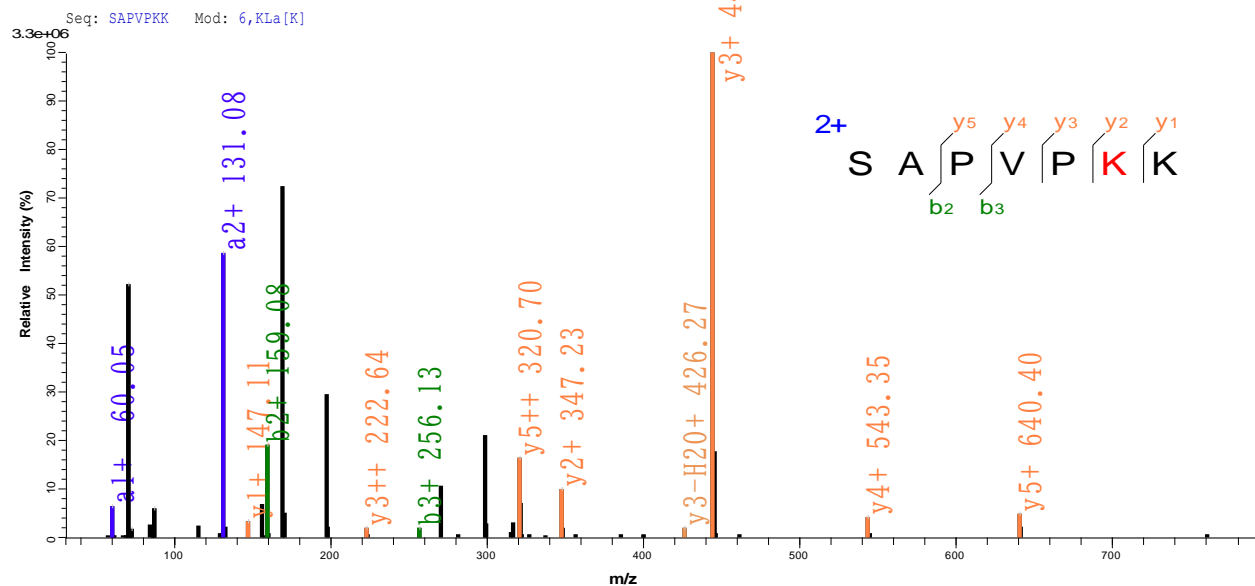


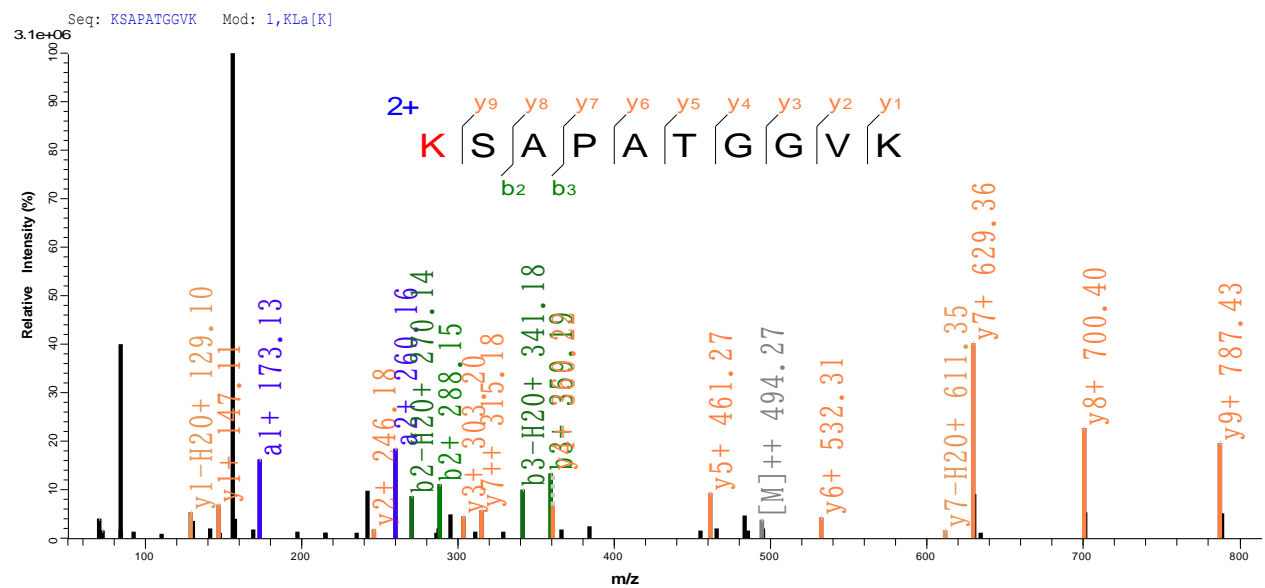




Dataset E

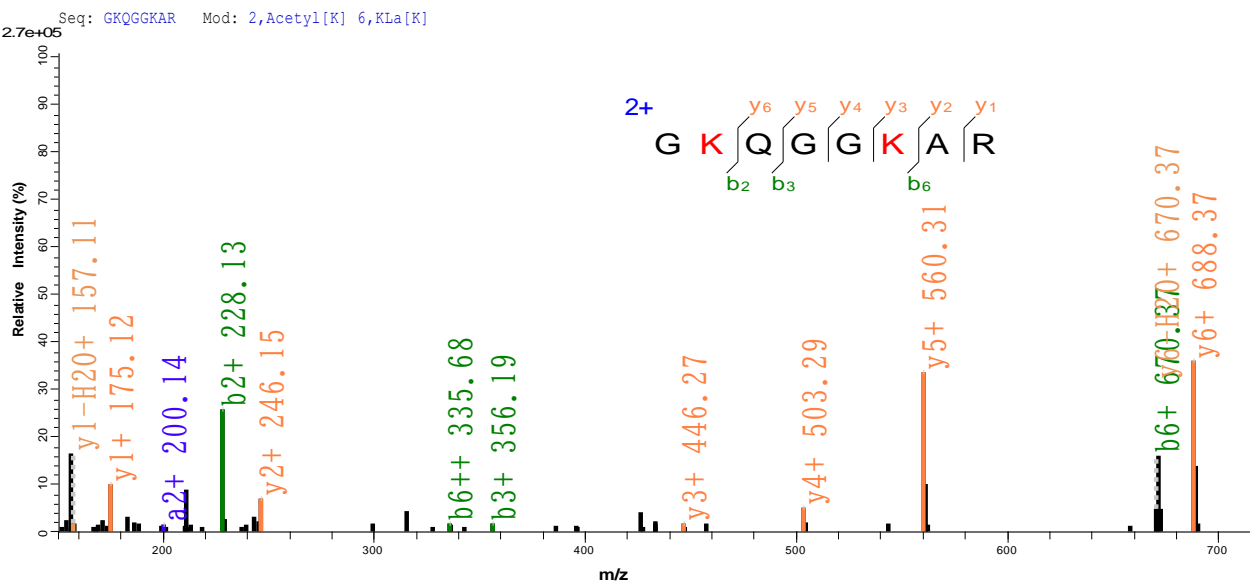
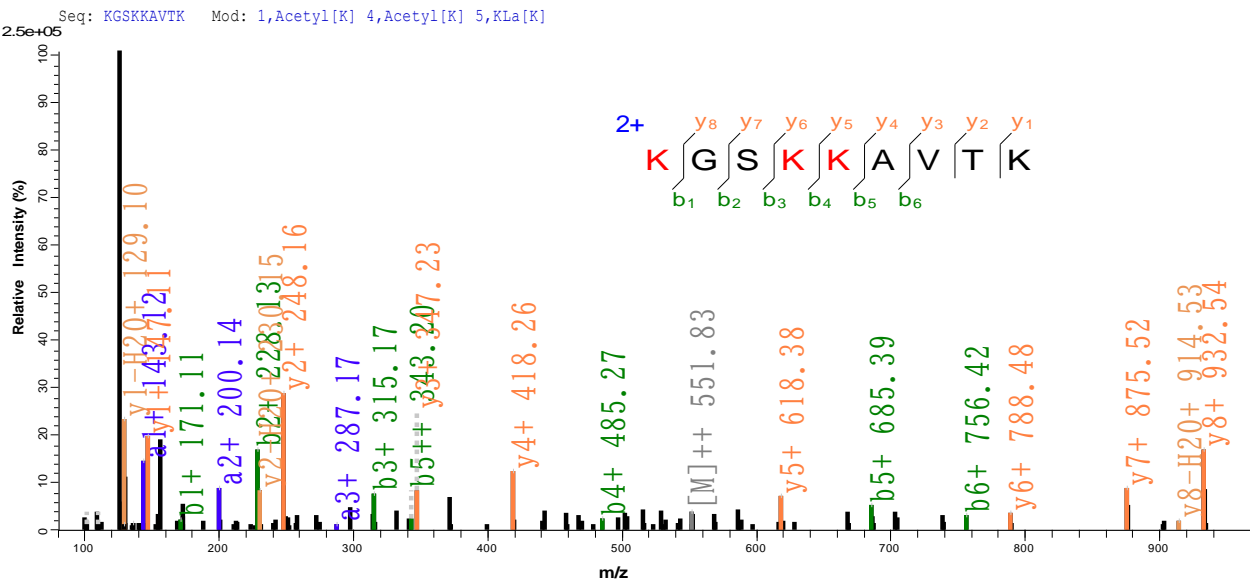


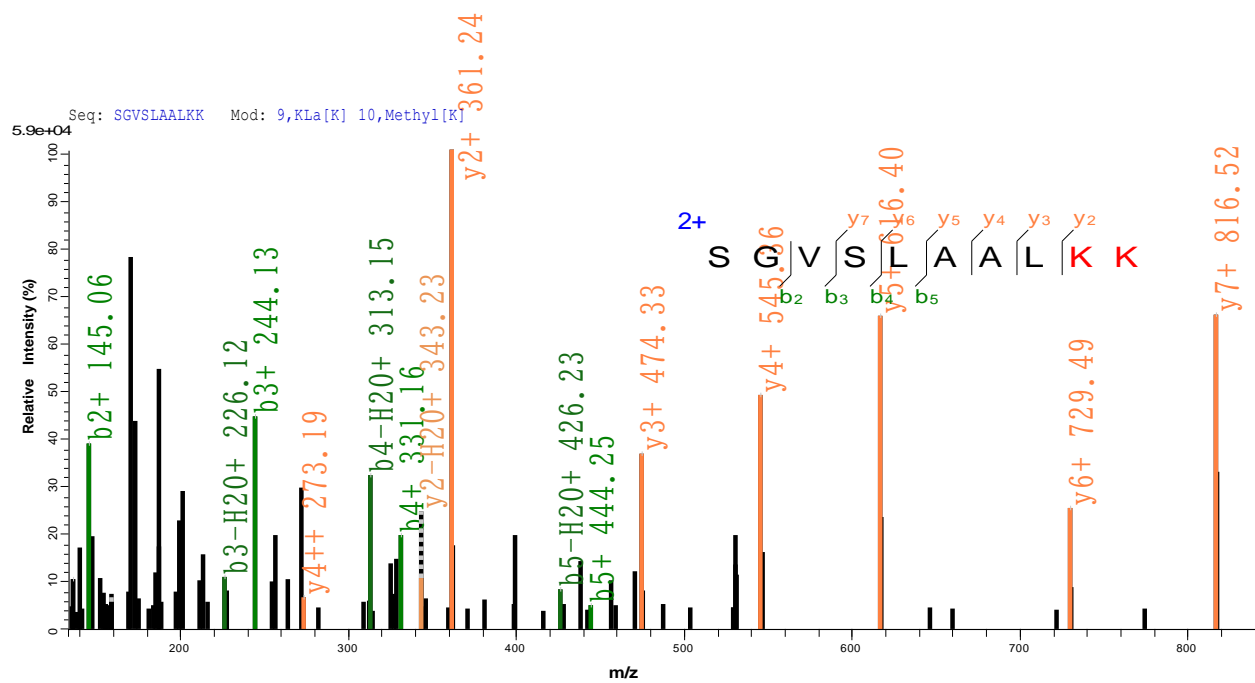
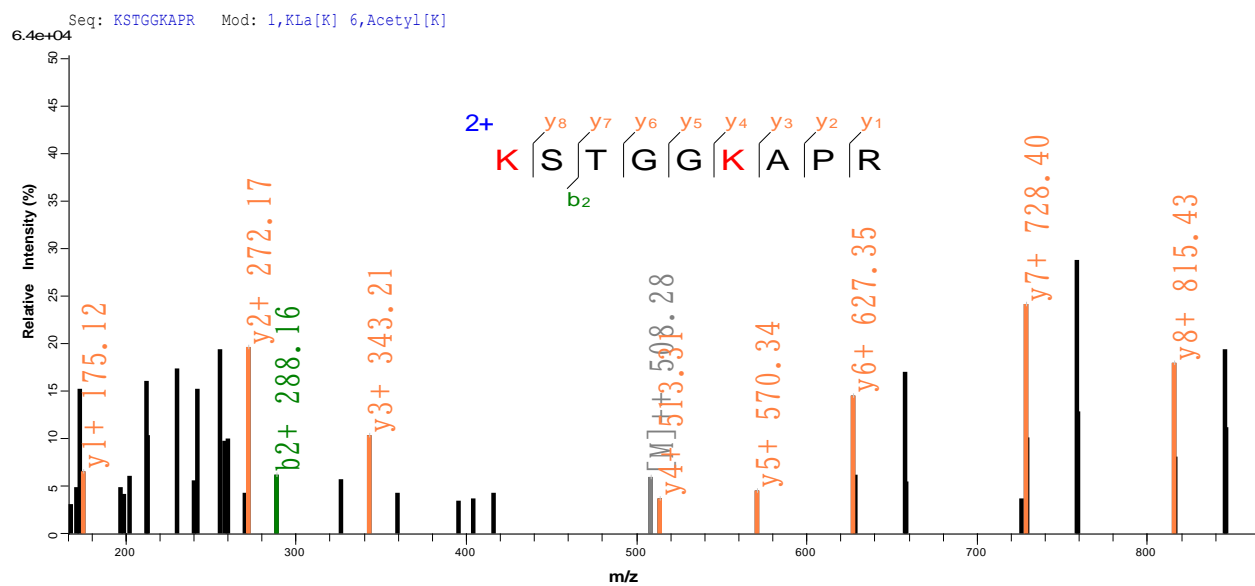


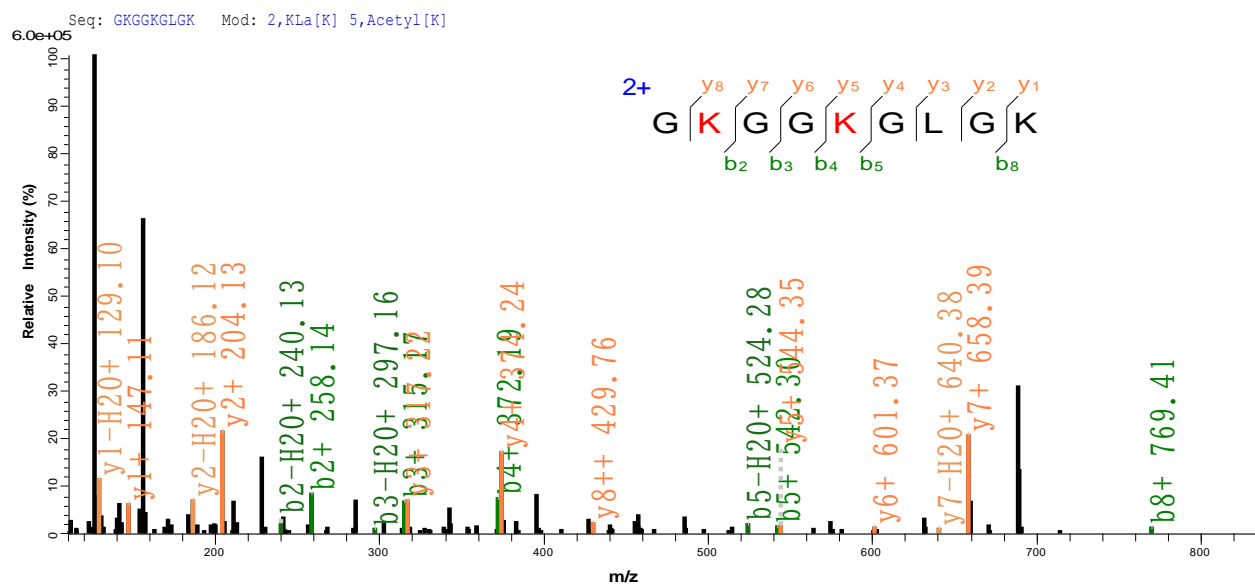


**Data S2. K1a peptide PSMs in Datasets A and E that were additionally identified by adding Kac, Kme1, and Rme1 as variable modifications.**

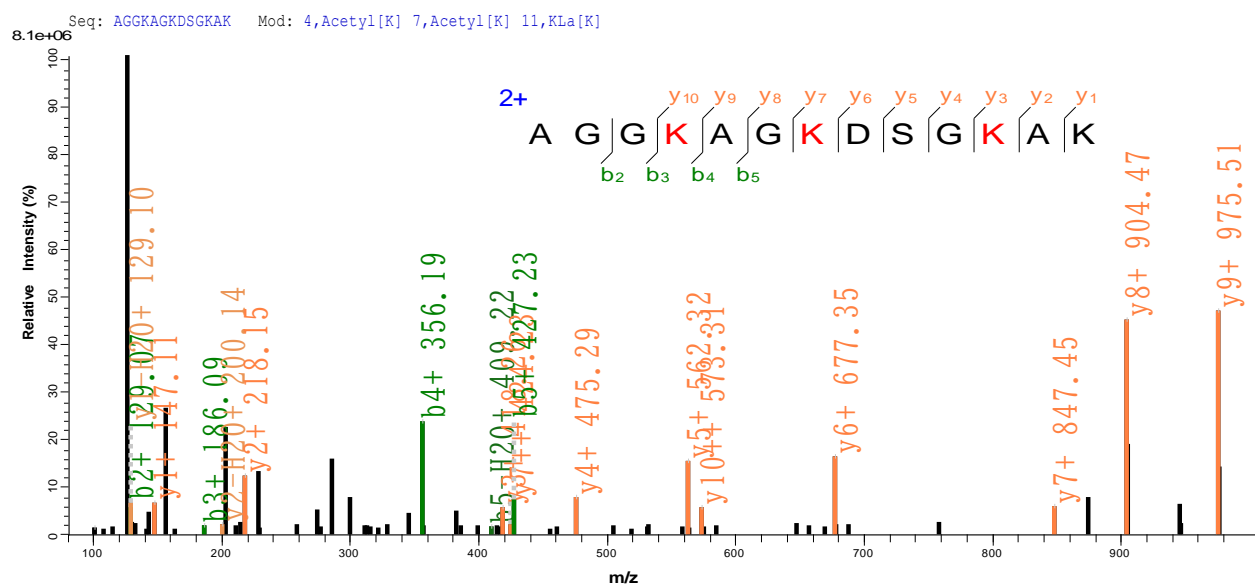
**Dataset A**

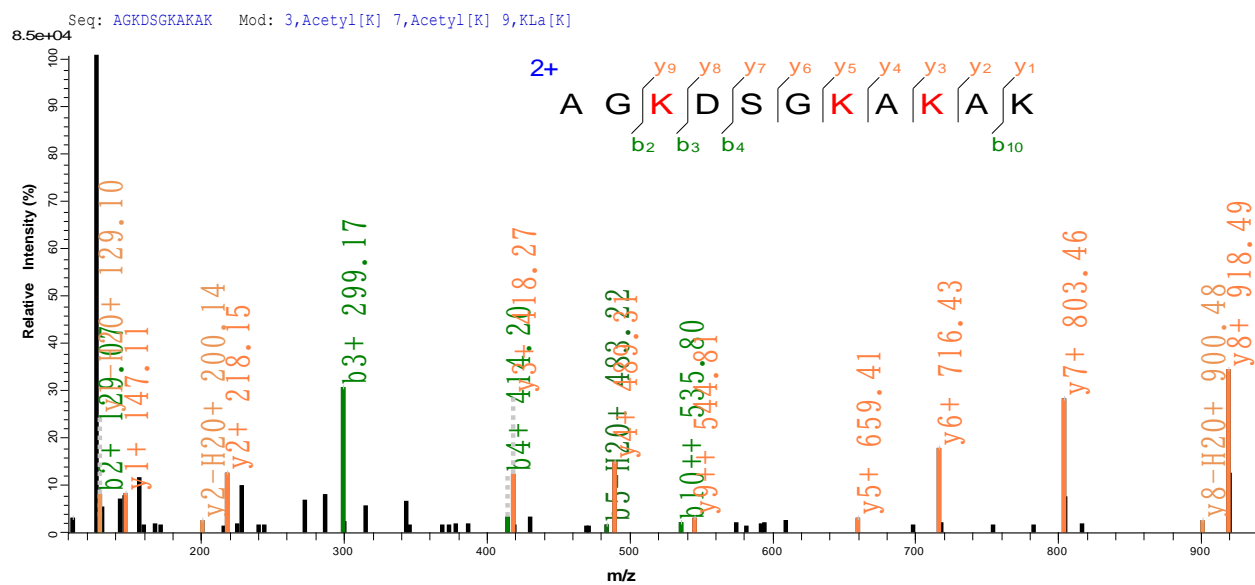
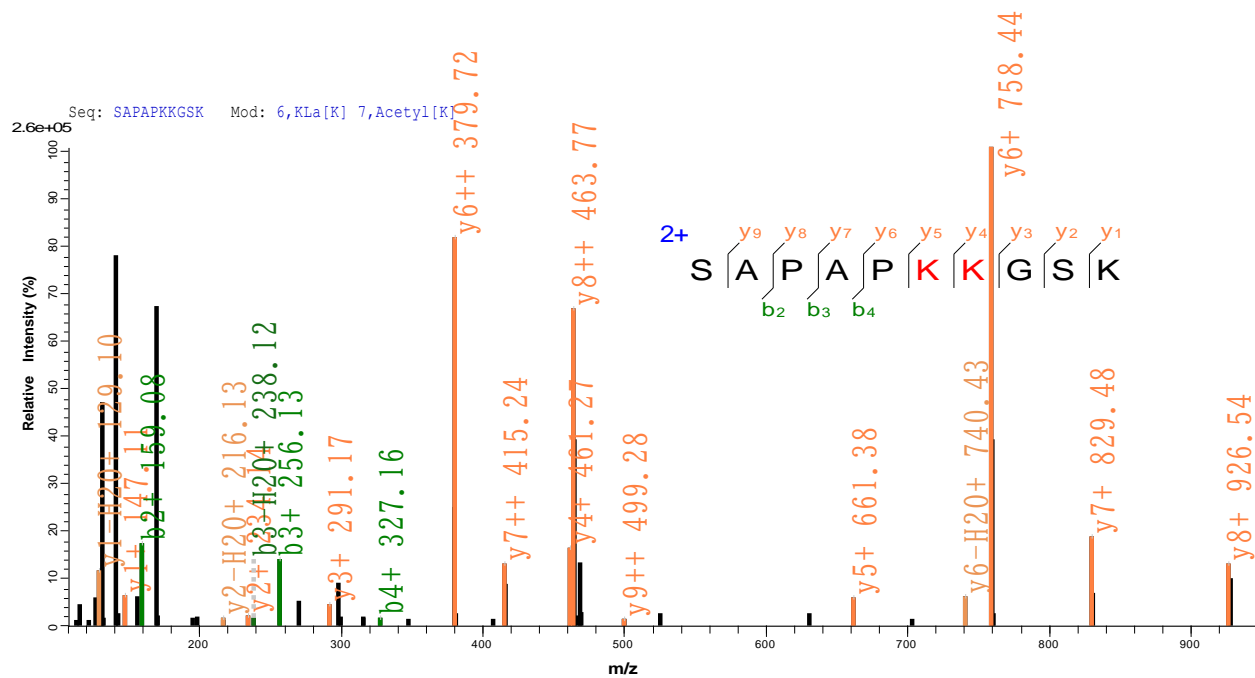


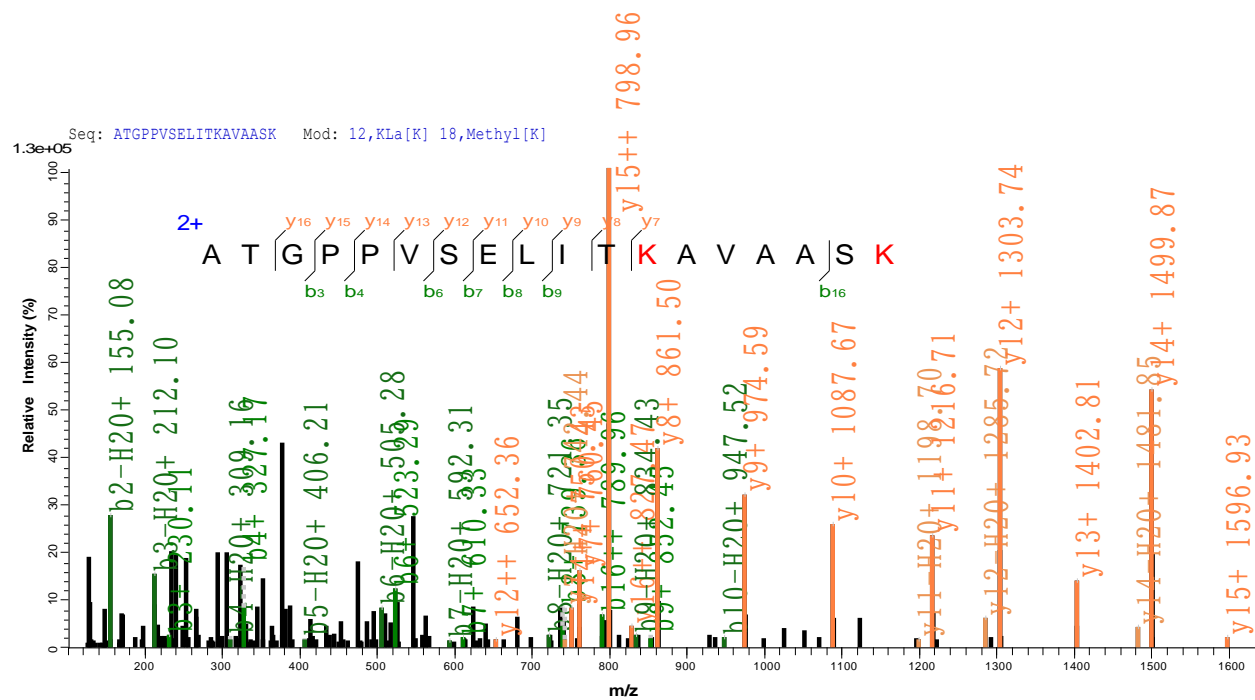
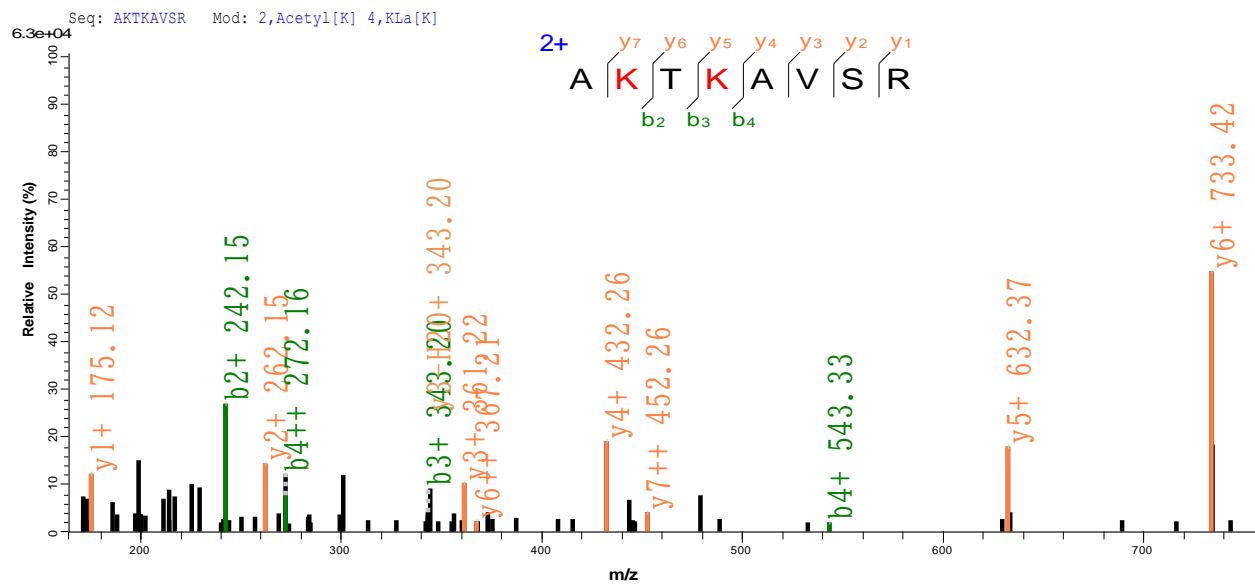




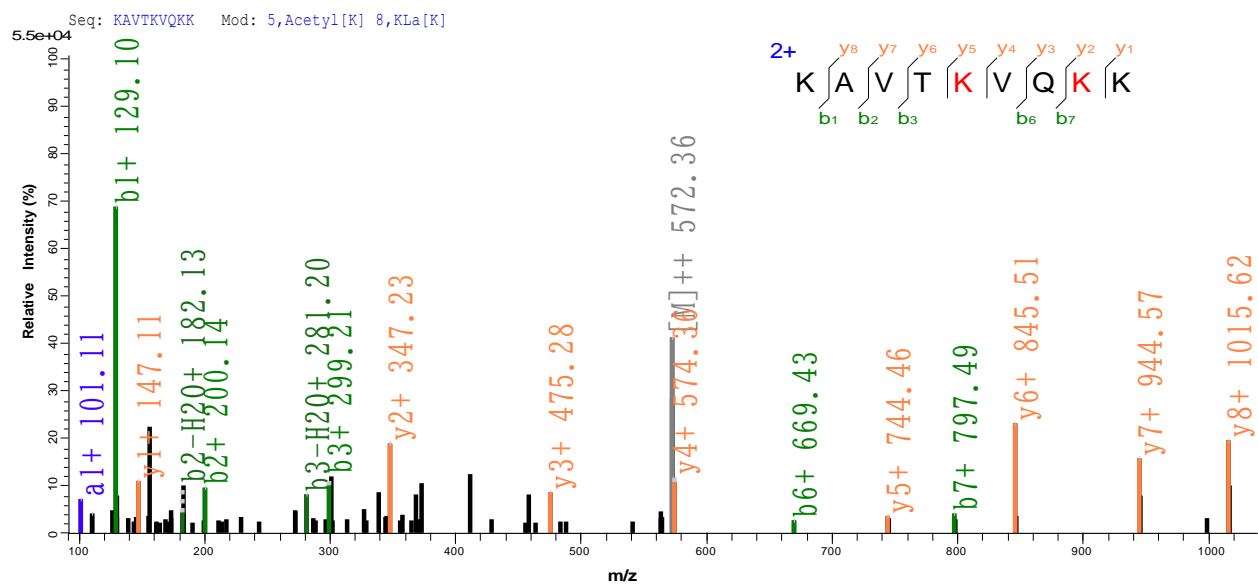
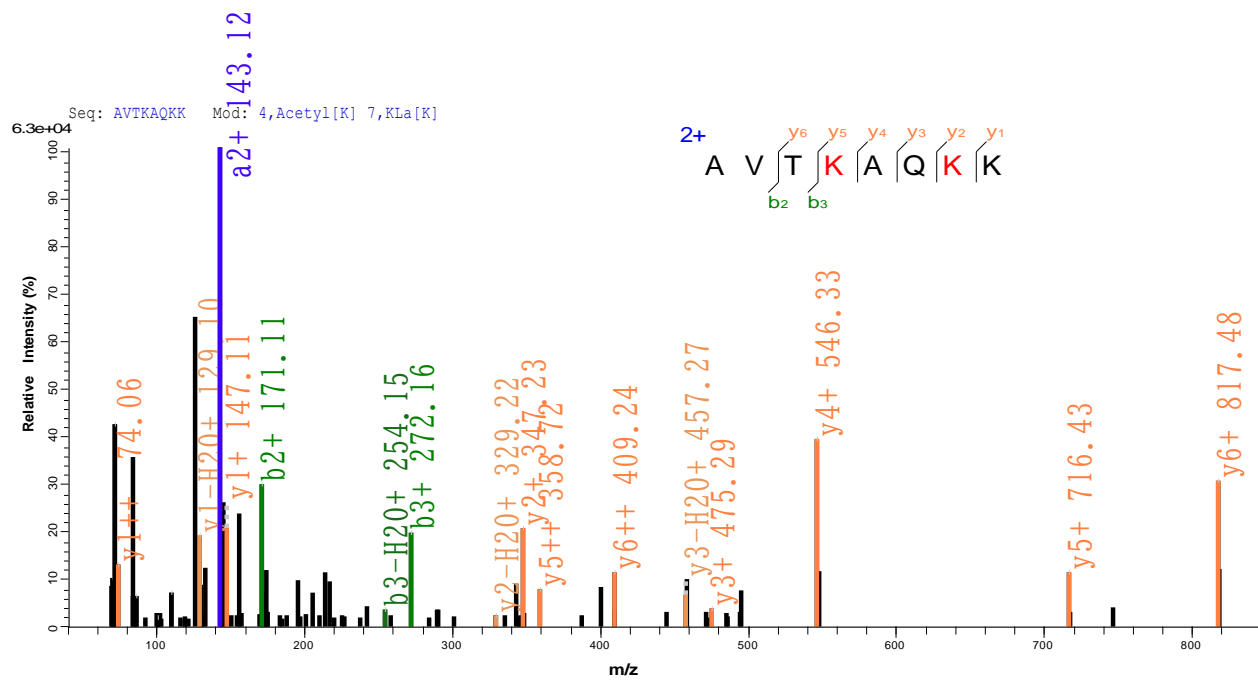
## Dataset E

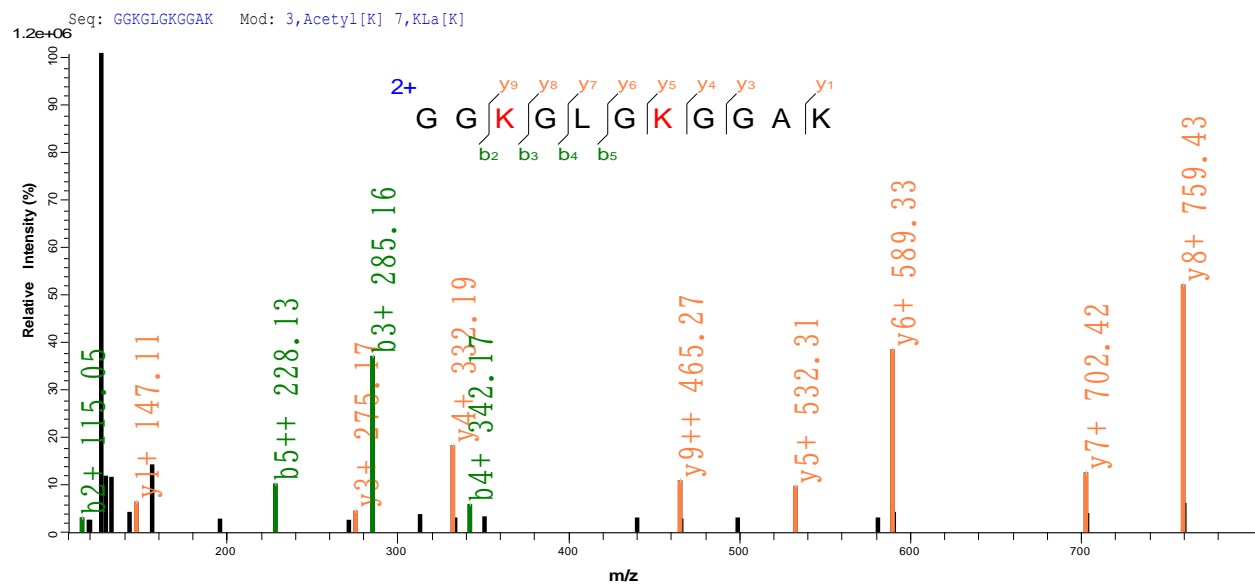
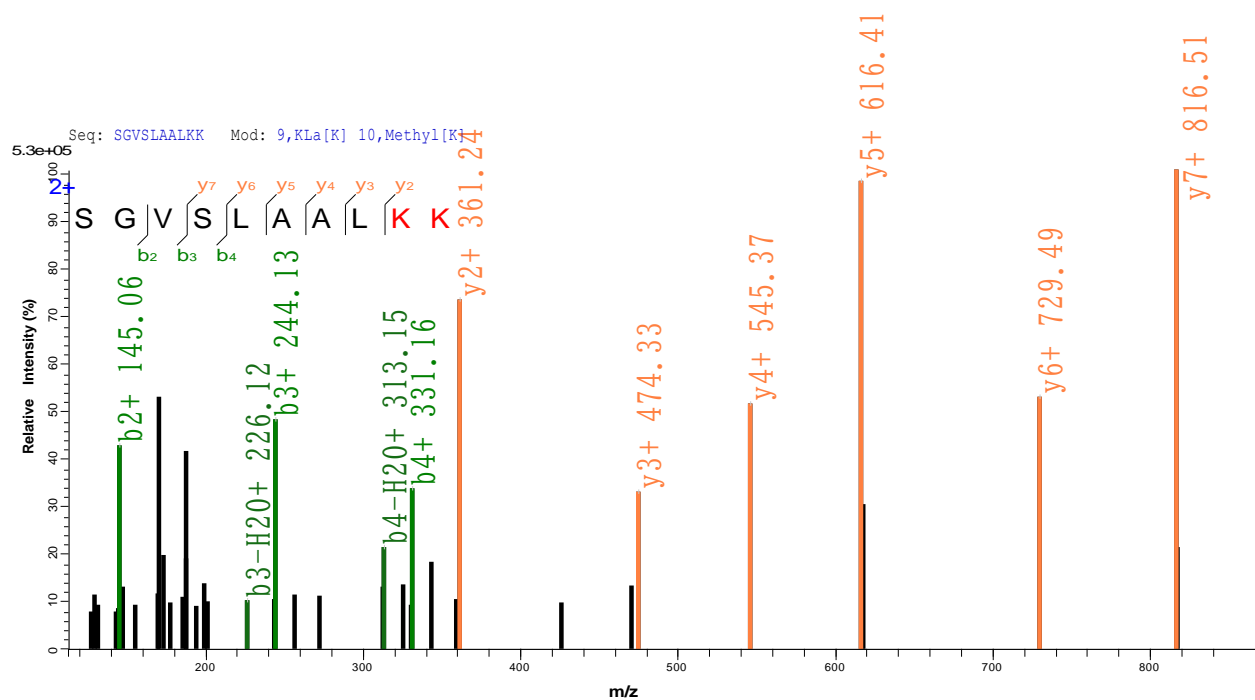


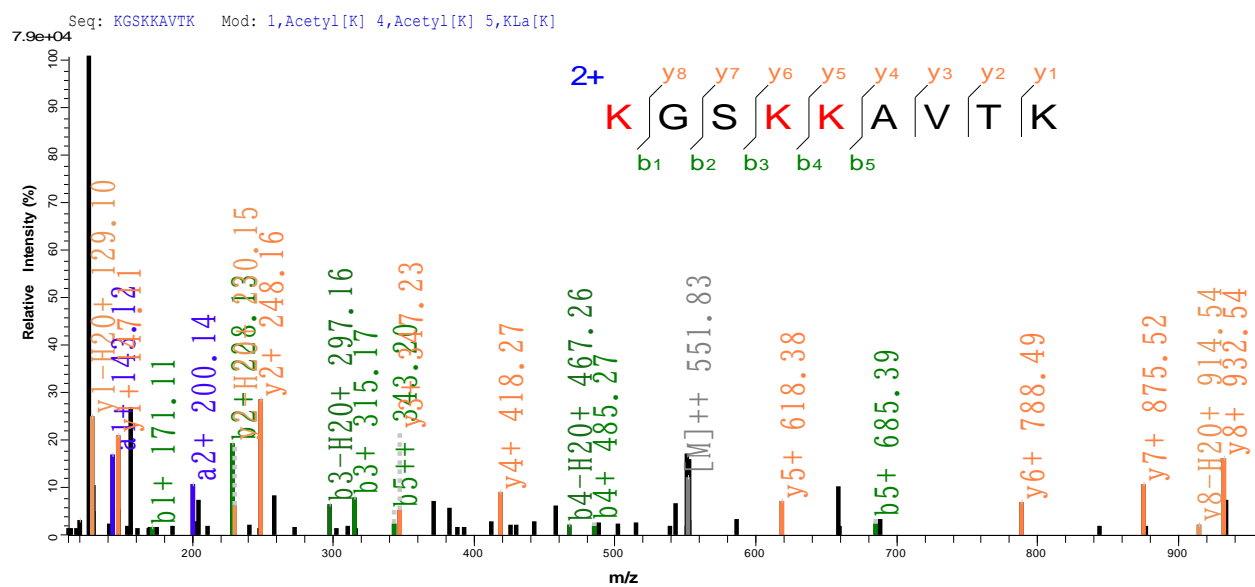
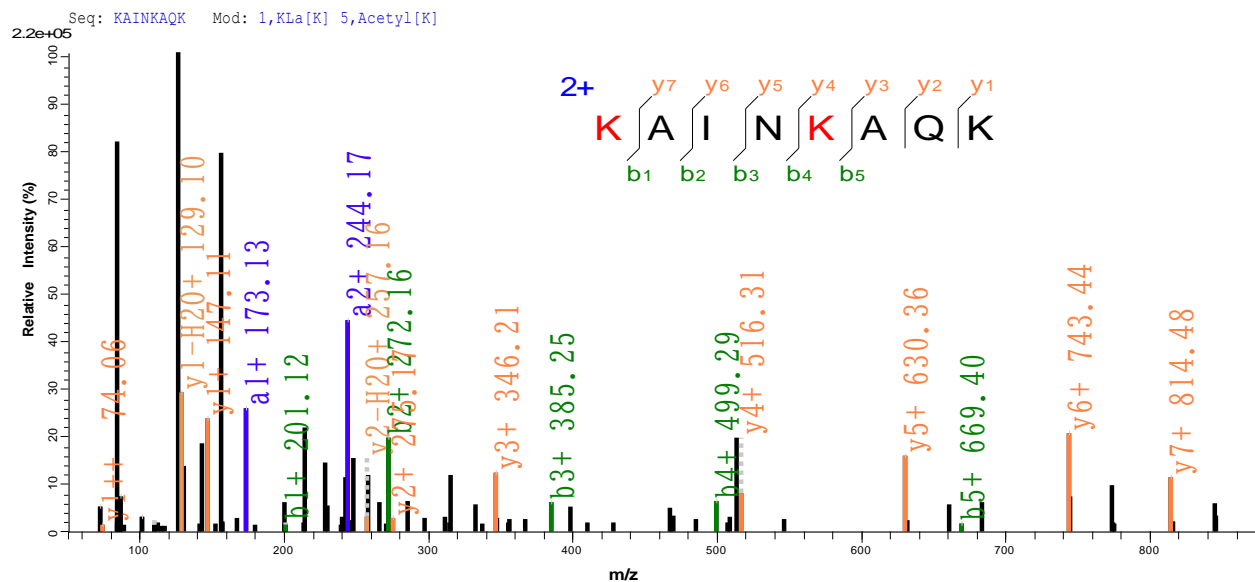


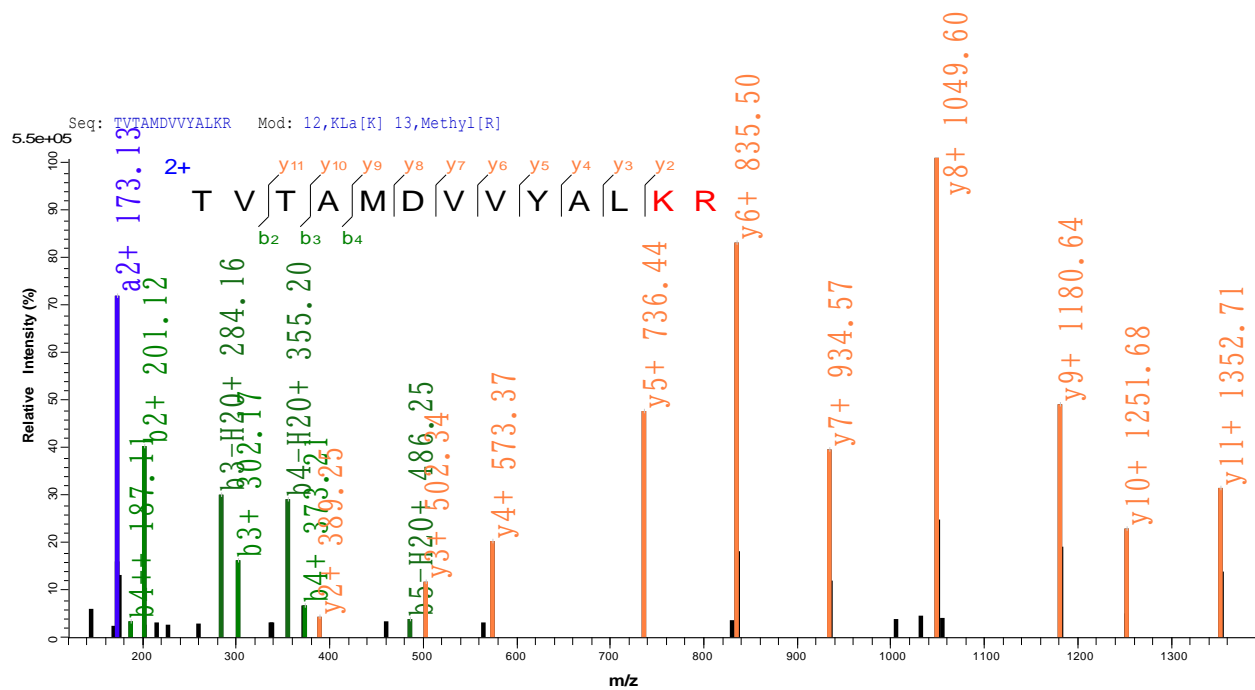
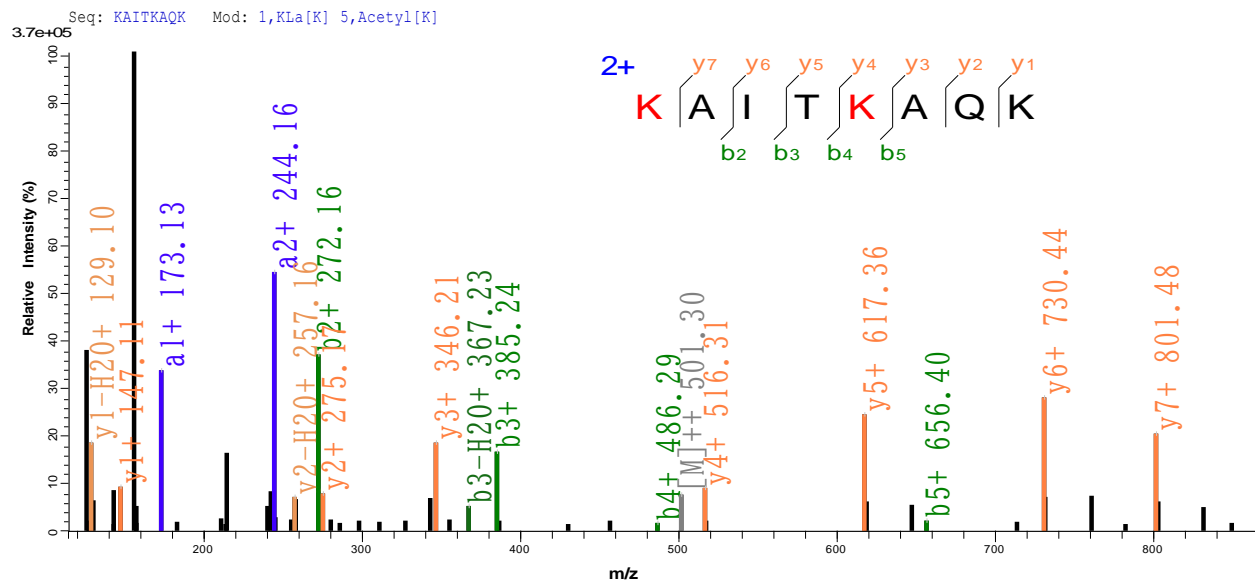


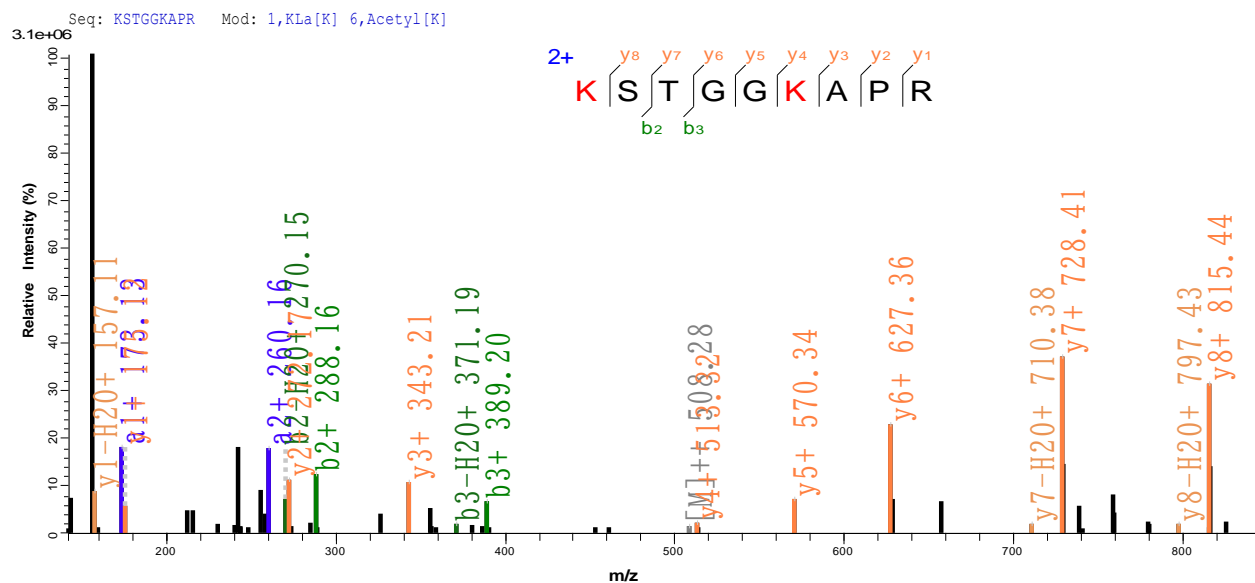
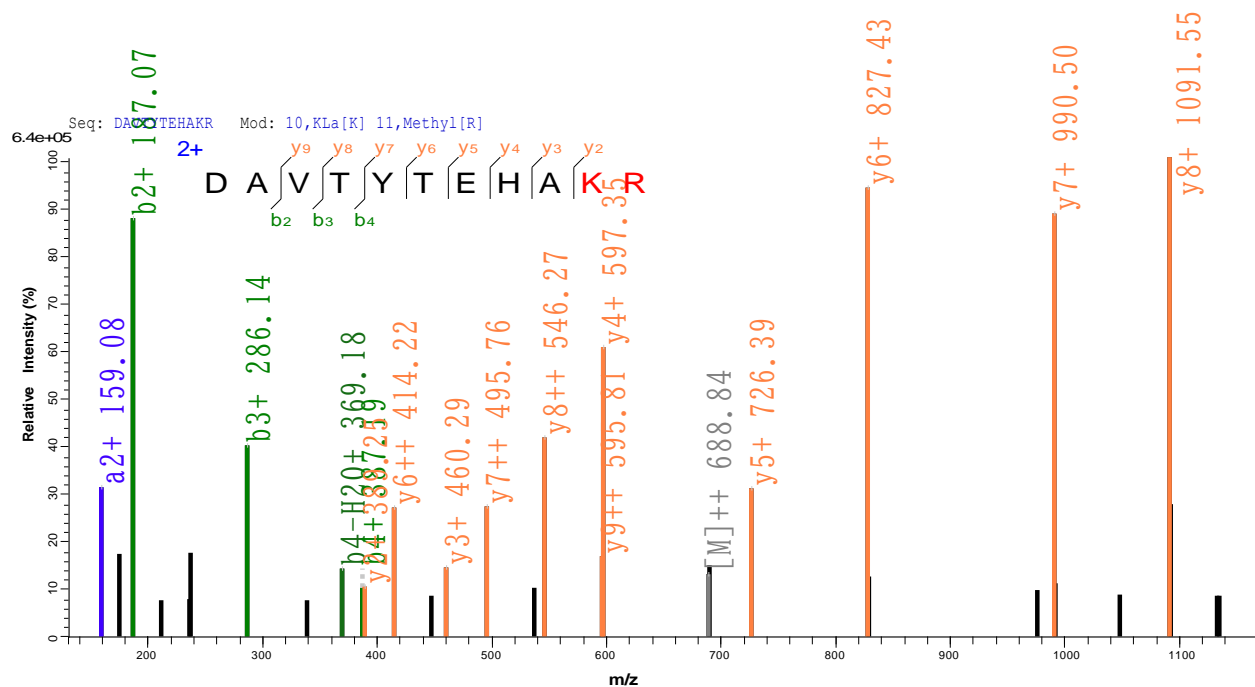






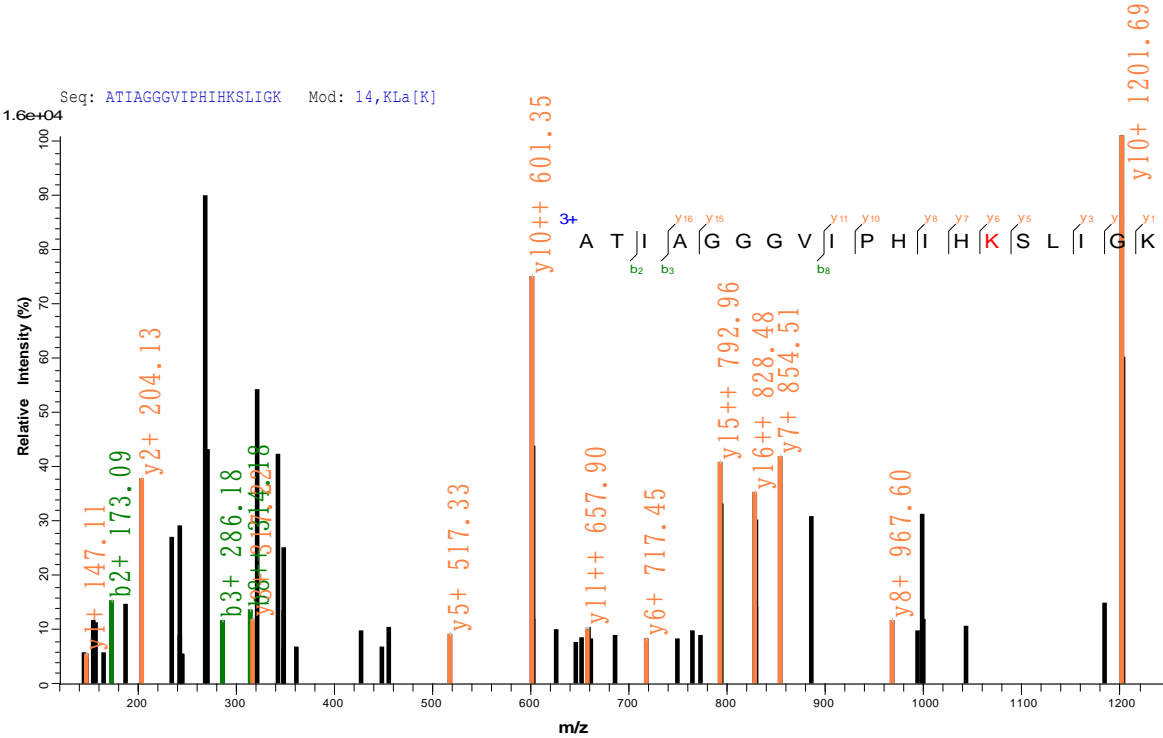
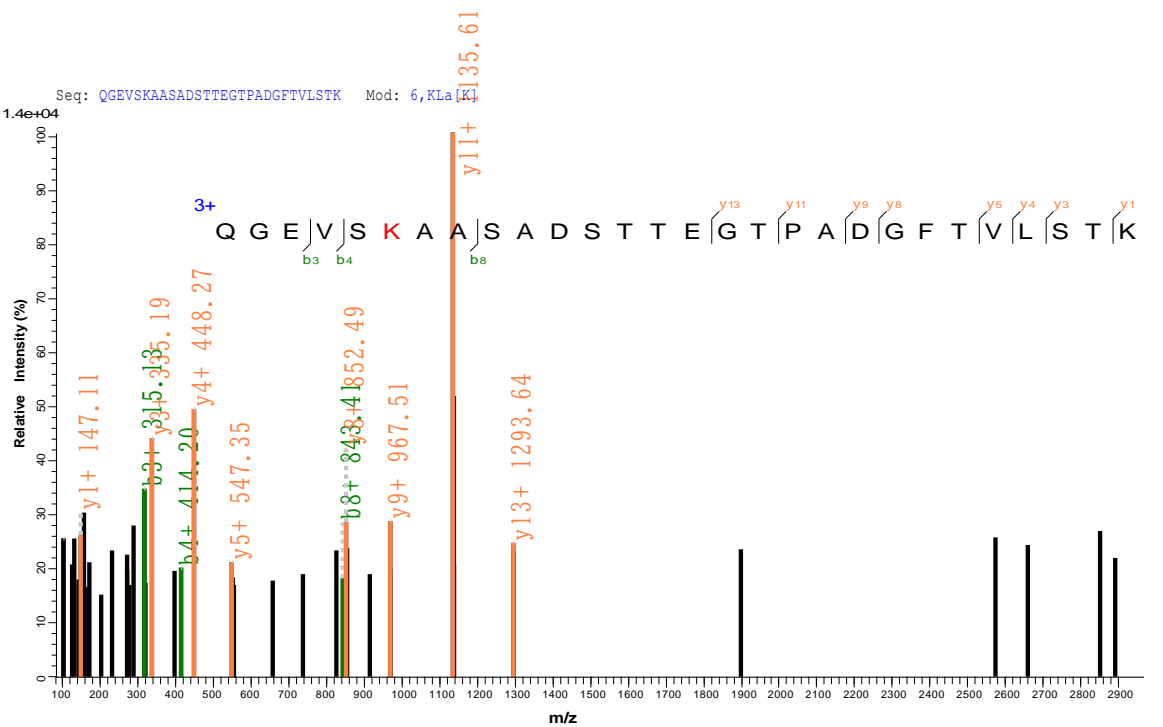


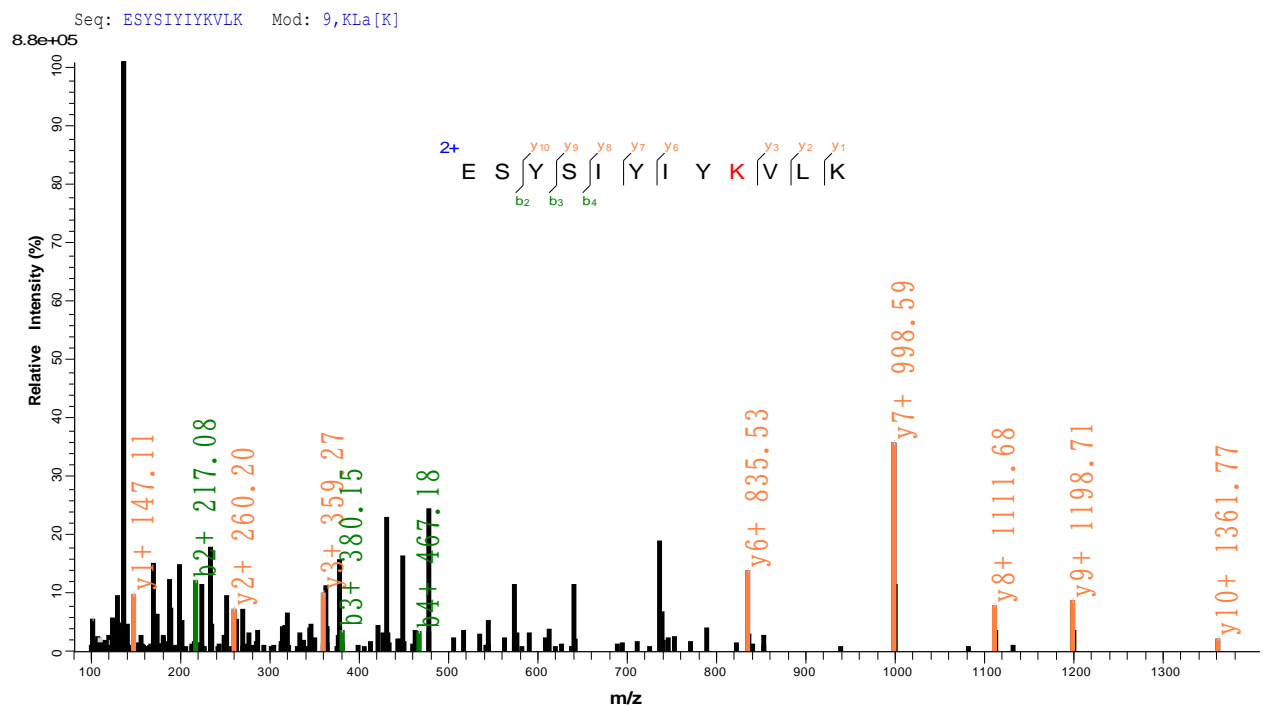
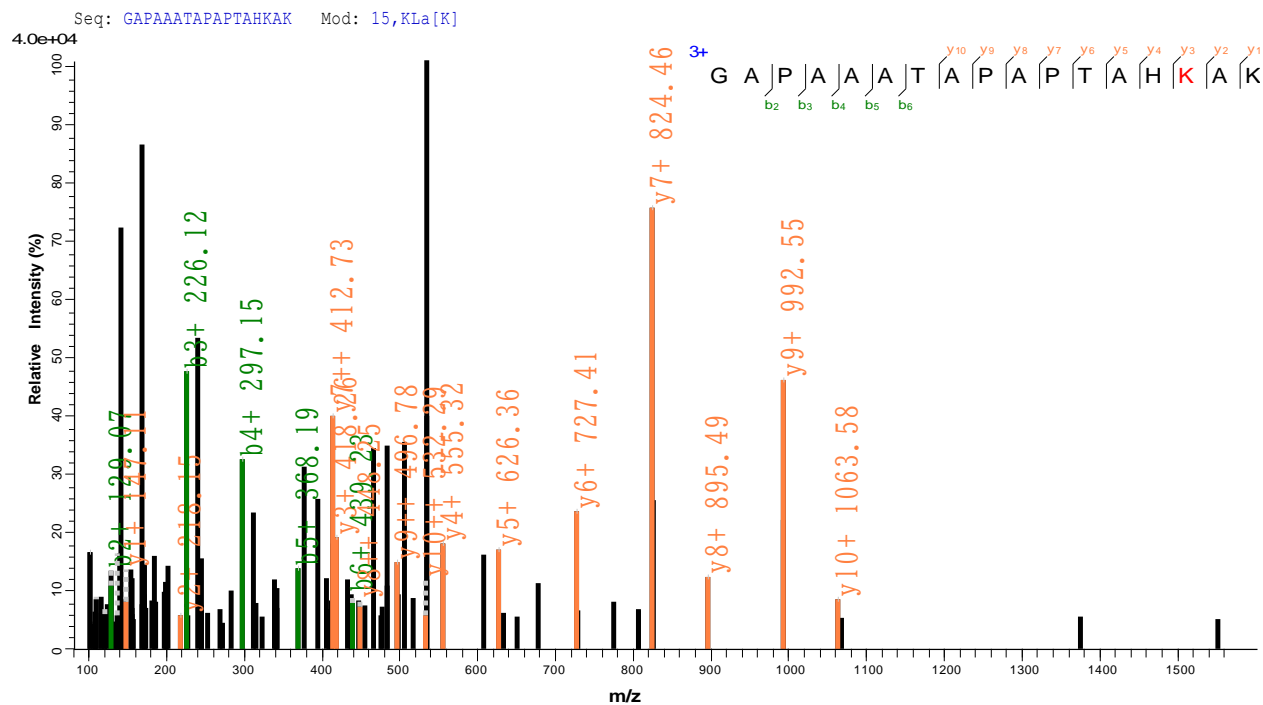


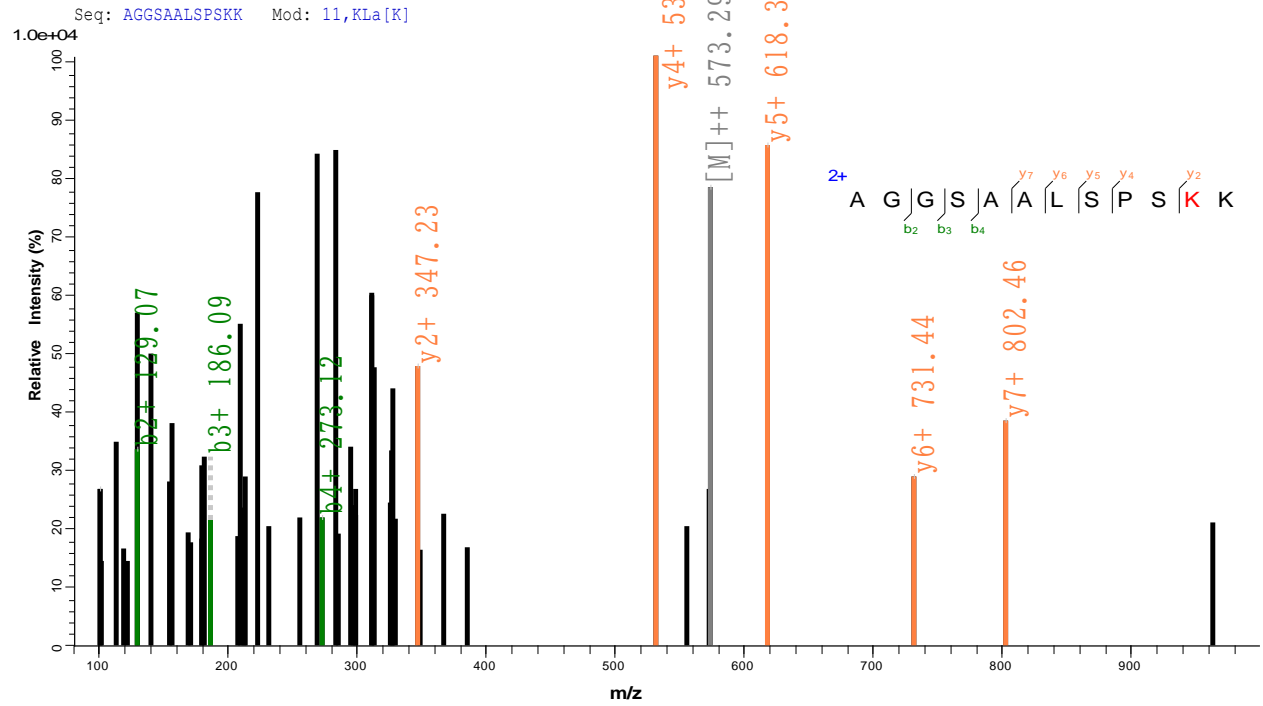


**Data S3. All PSMs identified by conventional strategy but not CHiMA when analyzing previous Kla, Kbz, Khib, and Kcr datasets.**

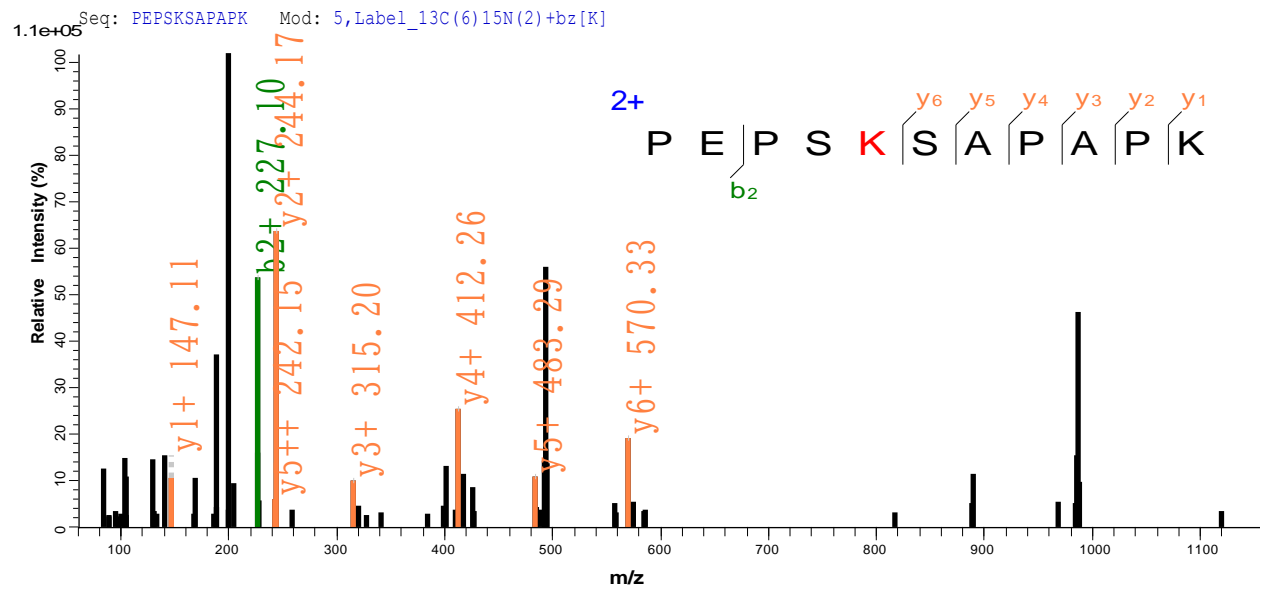
Kla



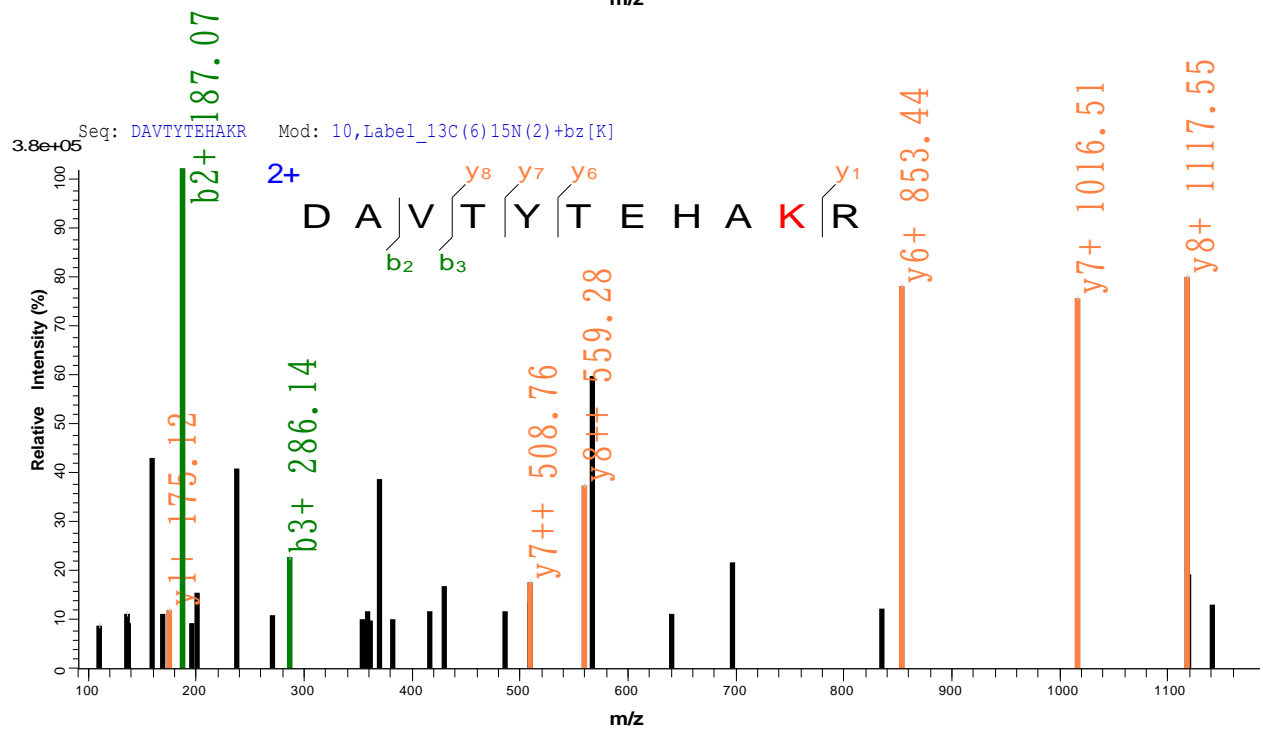
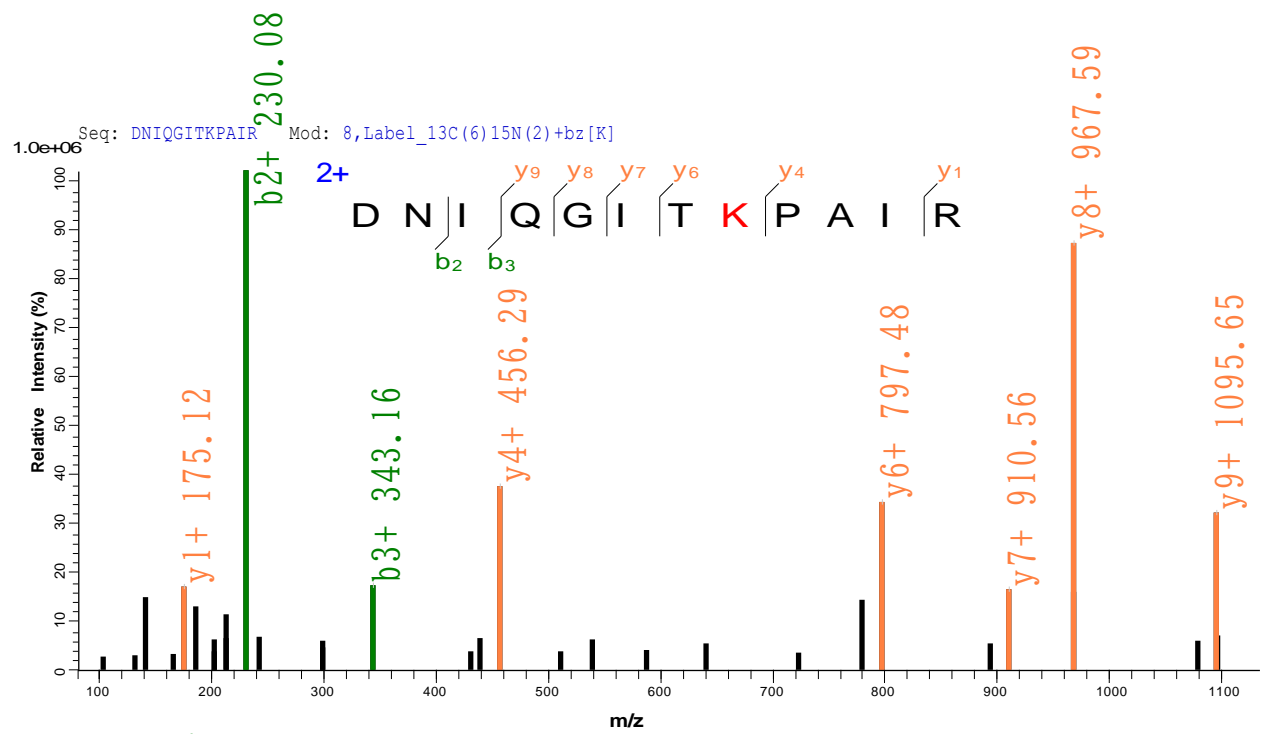




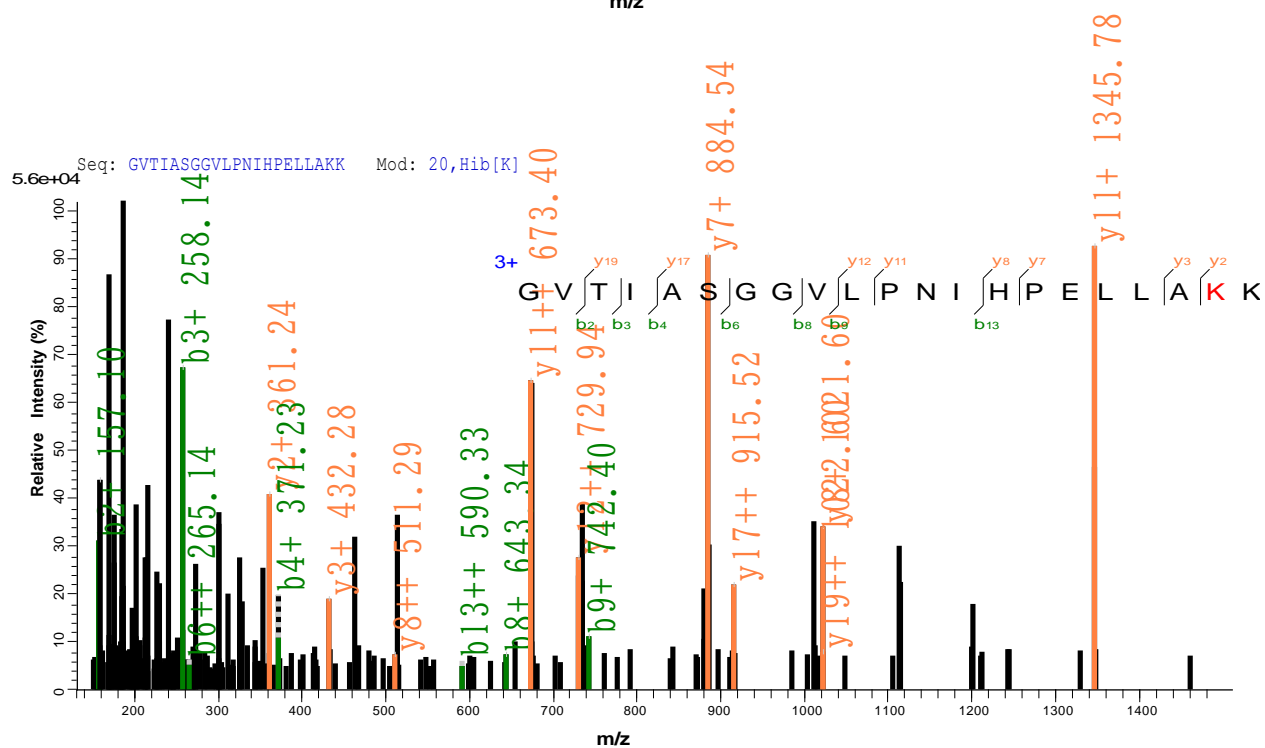
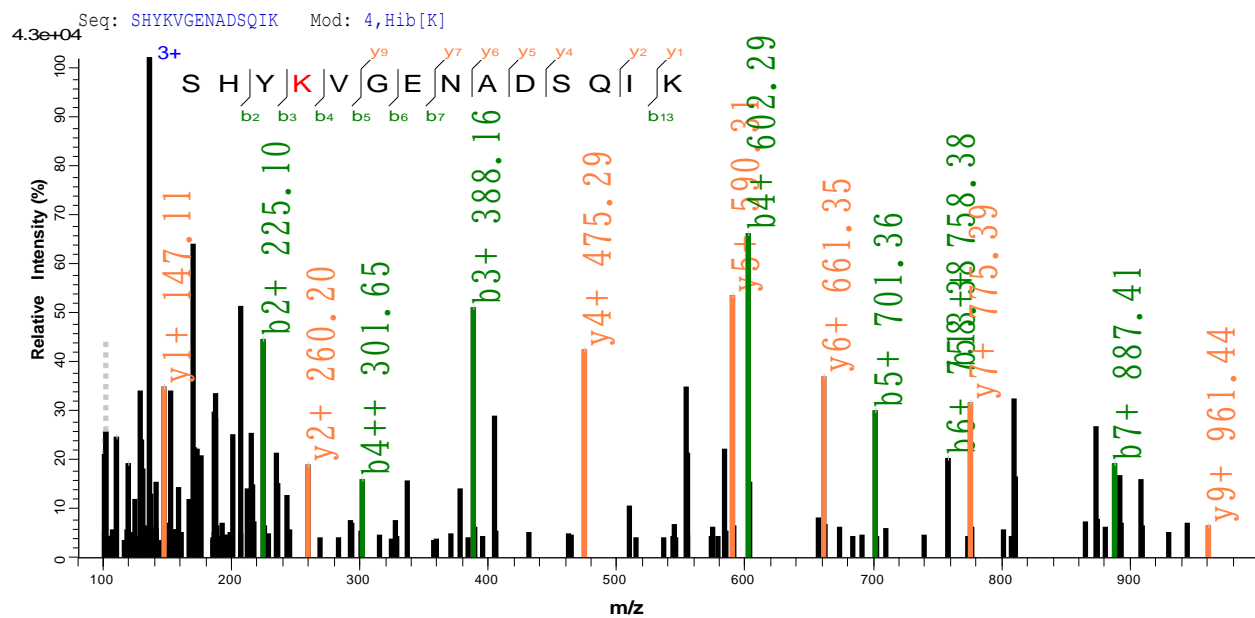
Kbz



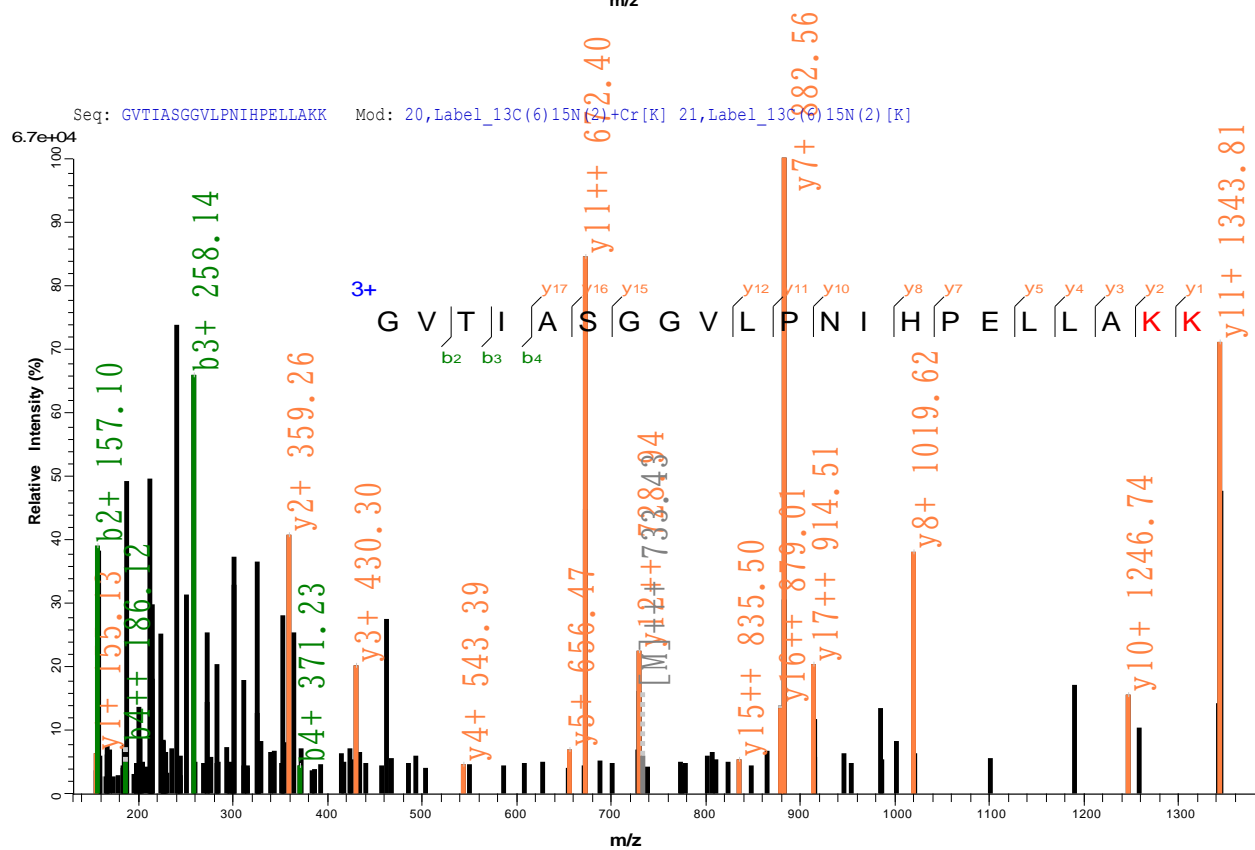
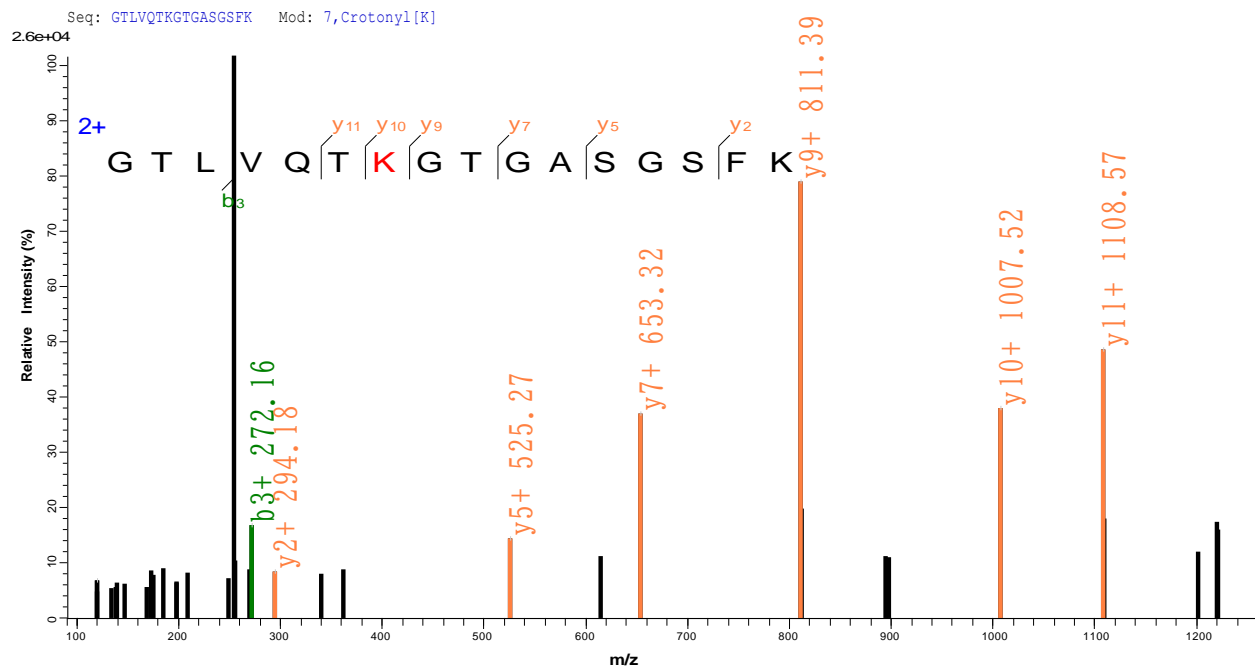


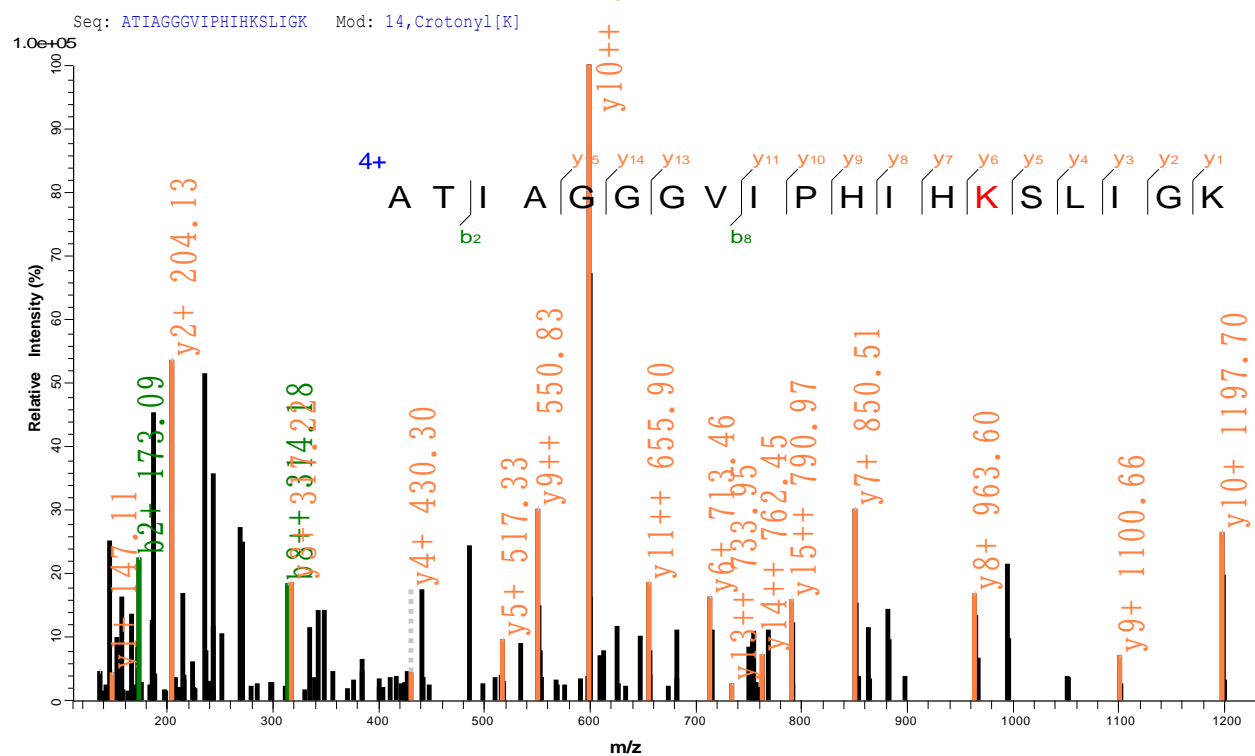
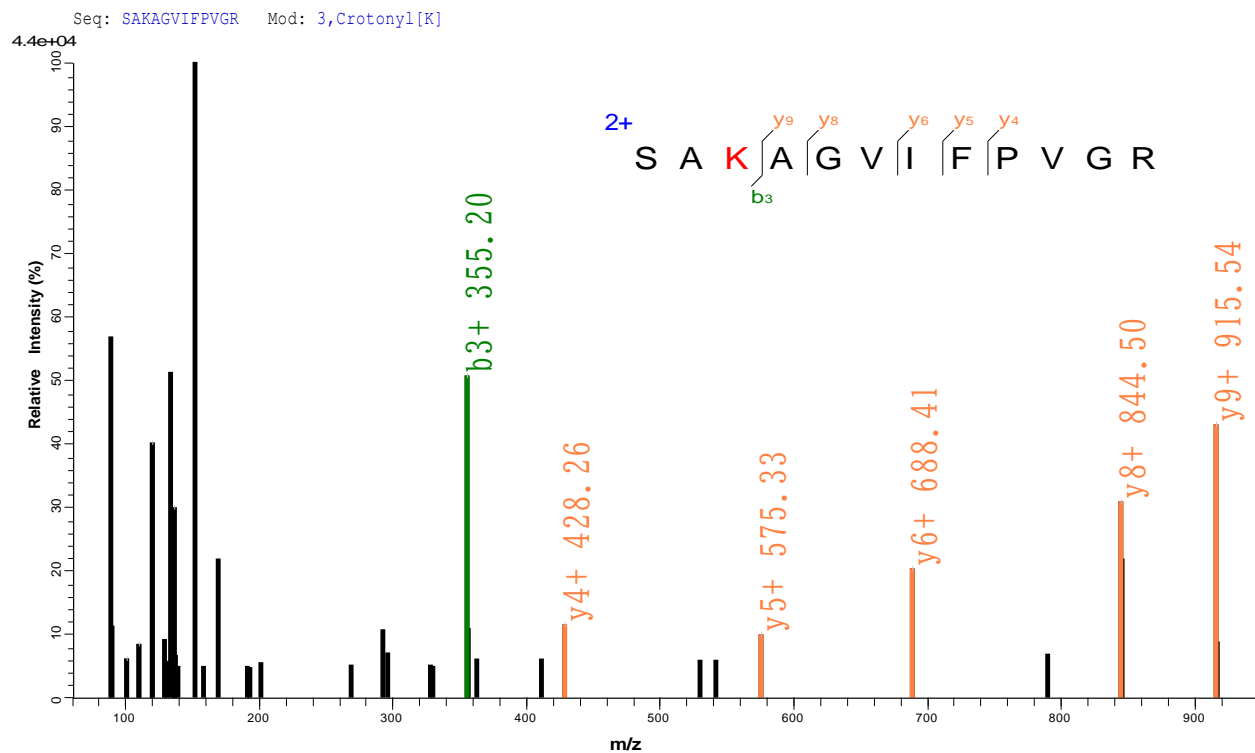


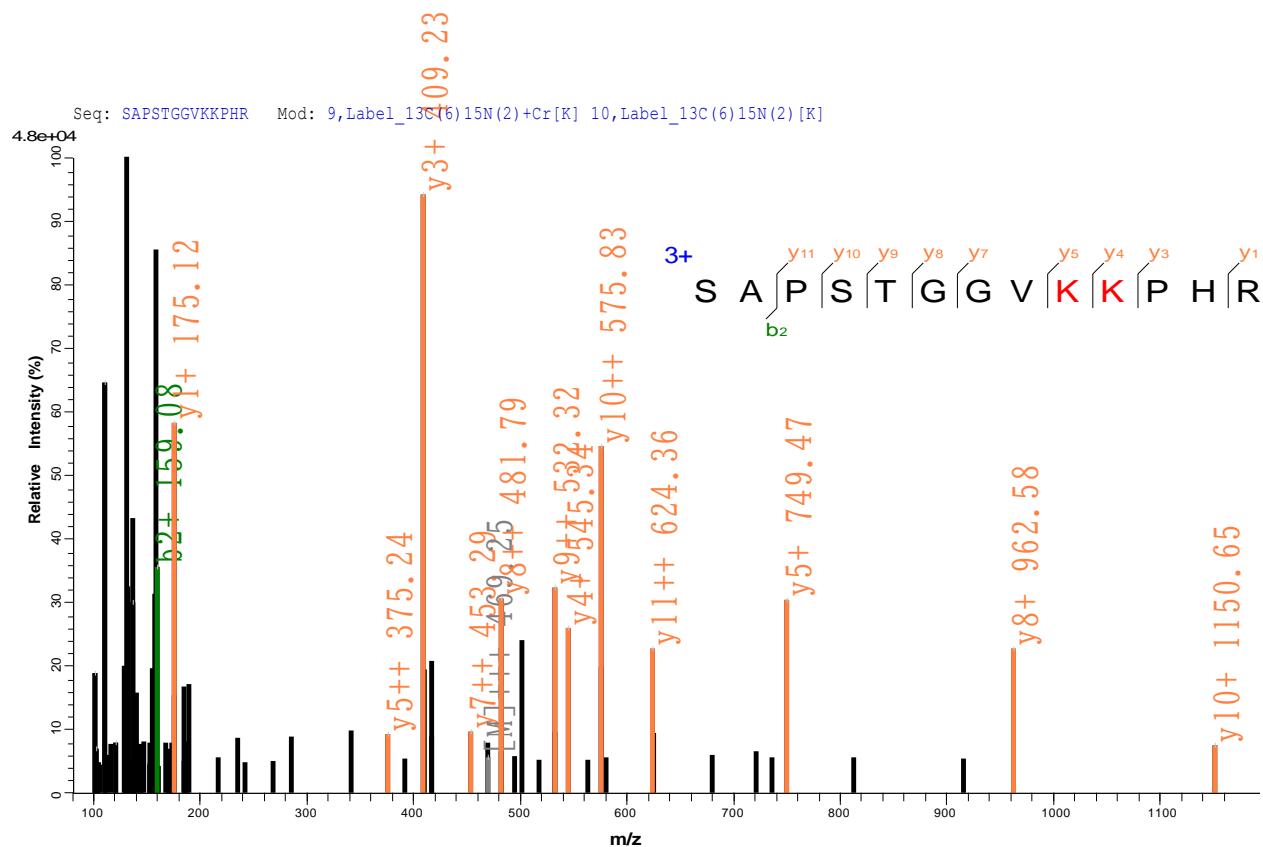
Khib



Kcr

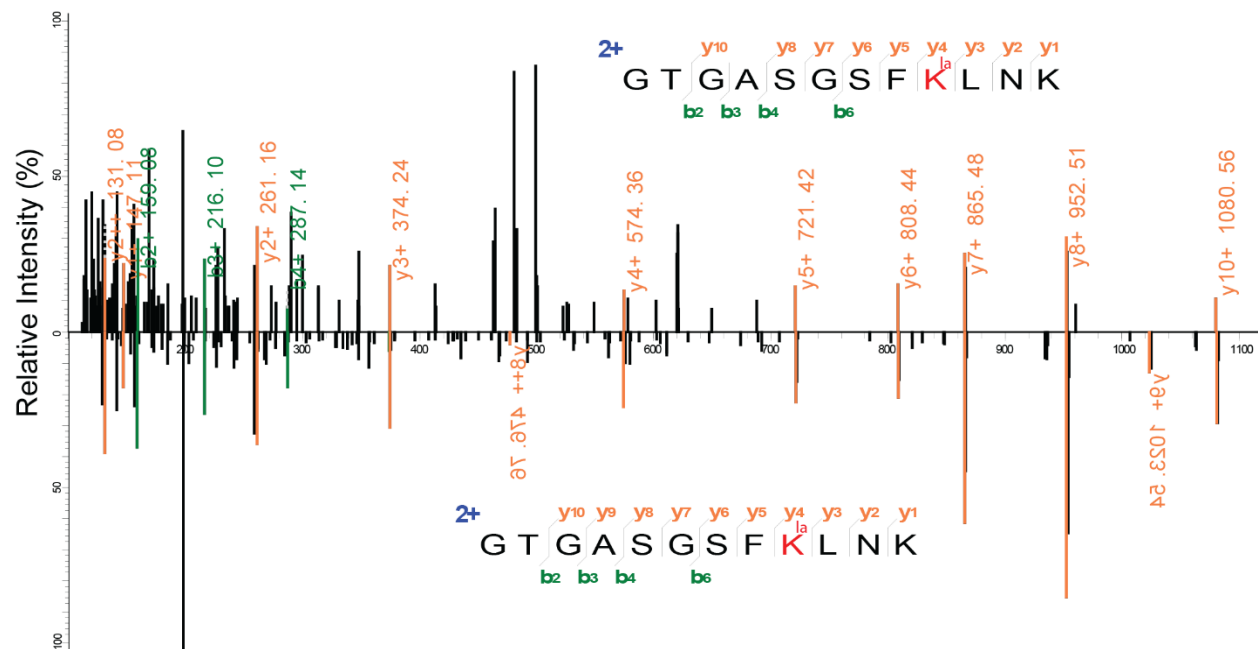


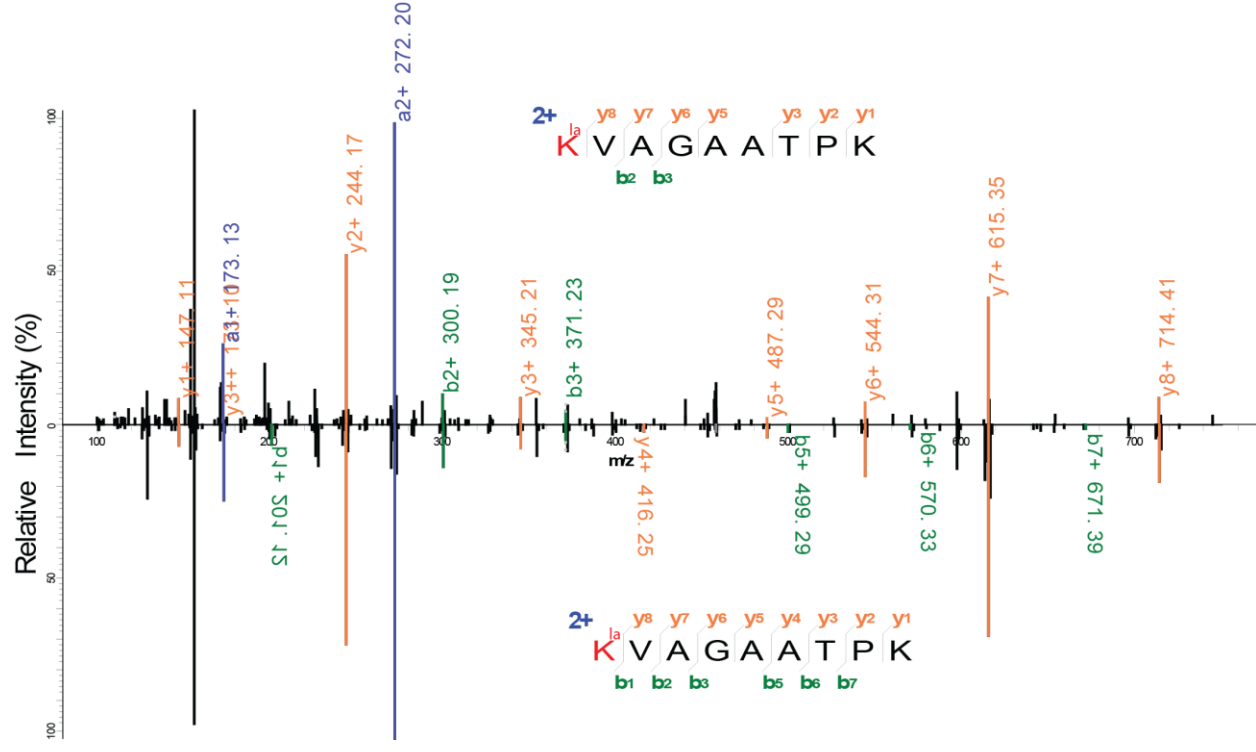
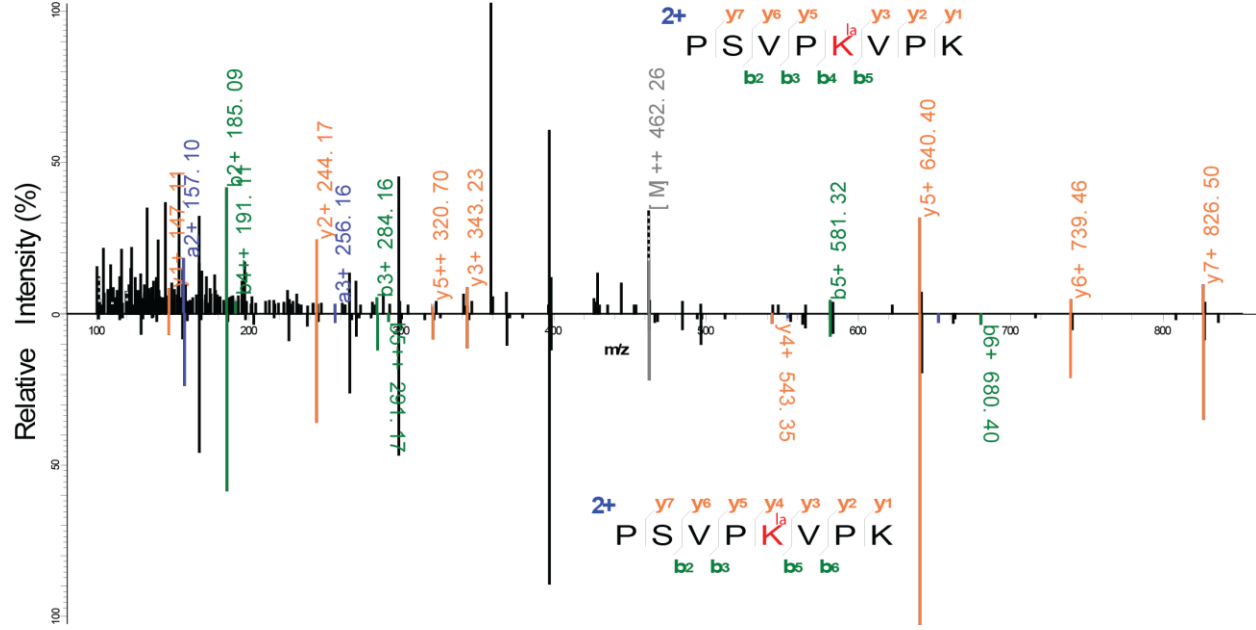


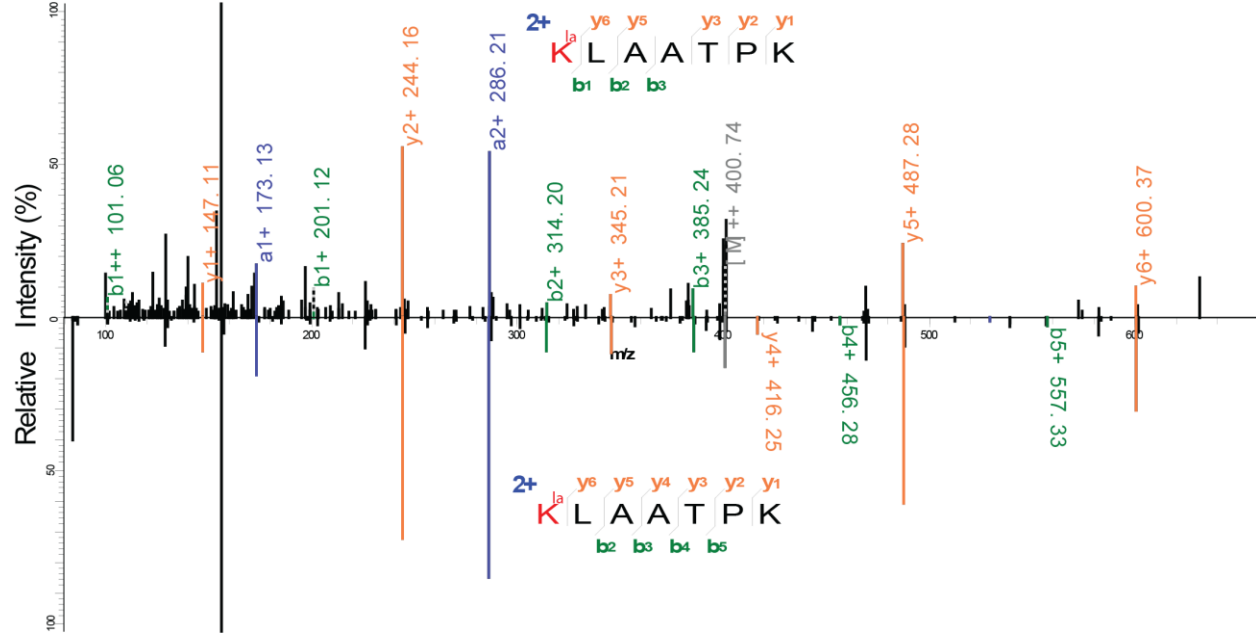
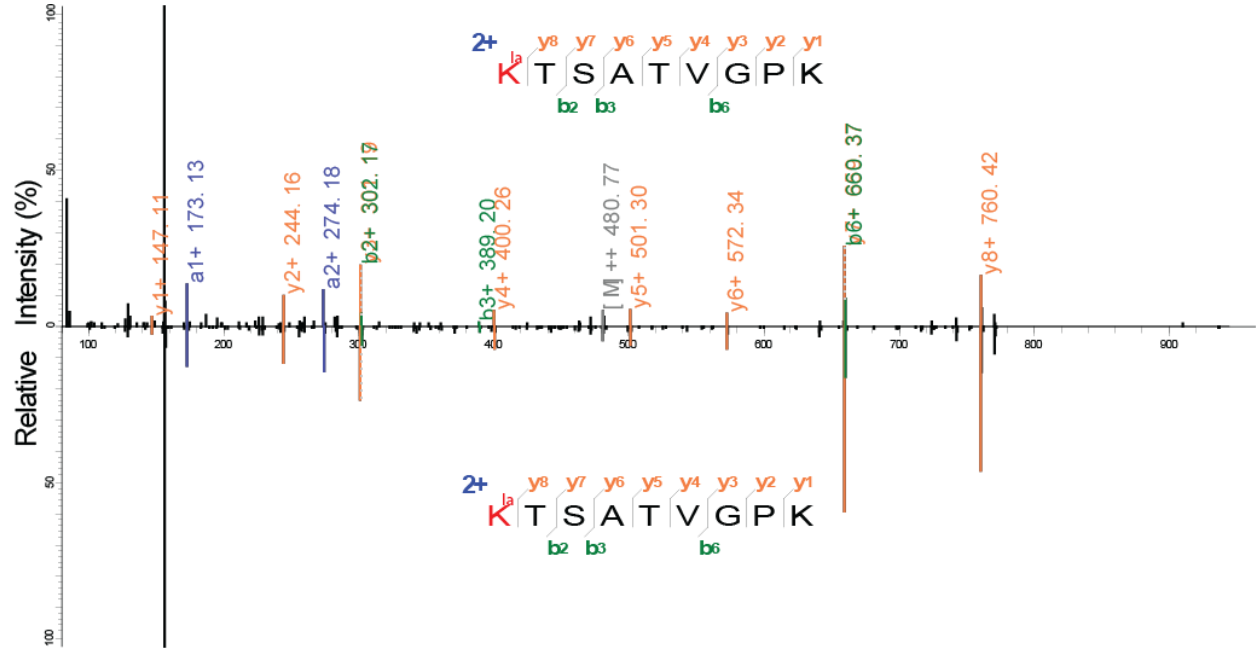


**Data S4. All moderate-confidence PSMs for newly identified K<sub>la</sub> and K<sub>cr</sub> sites were validated using synthetic peptides.**

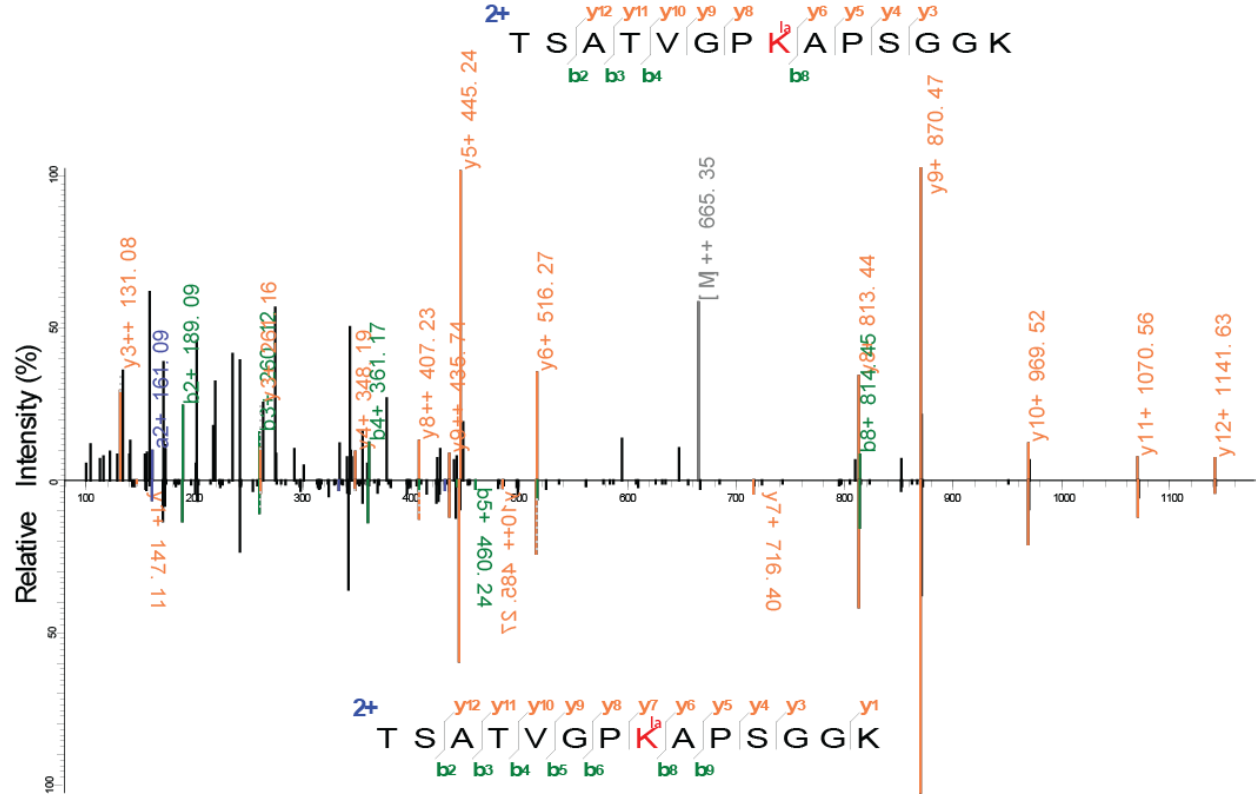
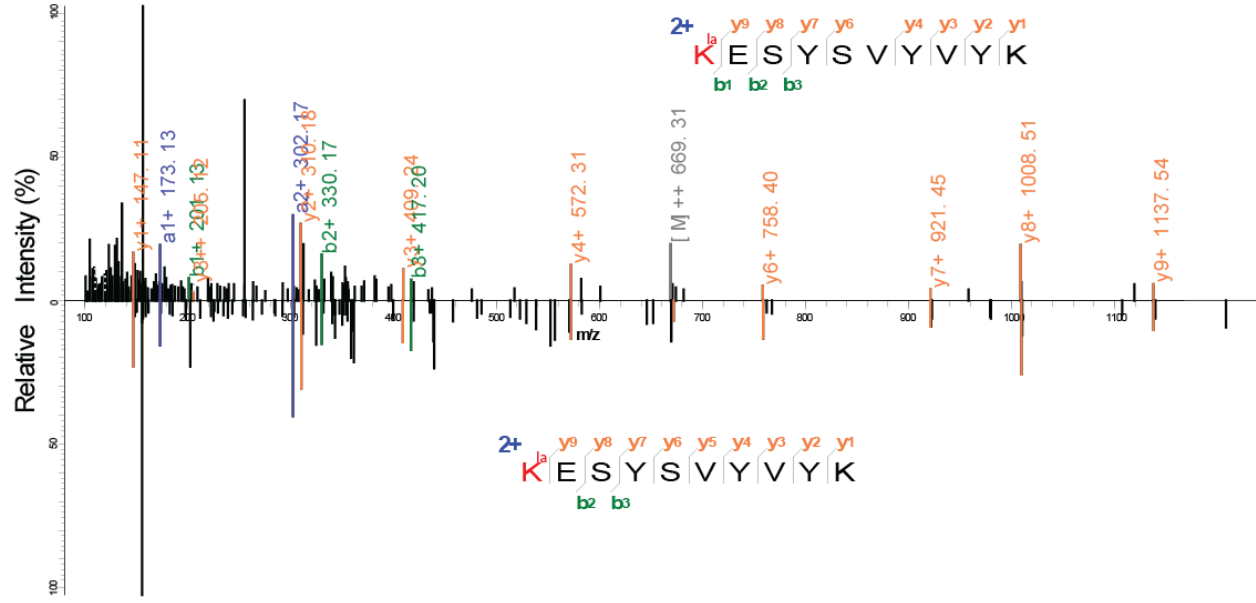
We chemically synthesized all K<sub>la</sub> and K<sub>cr</sub> peptides newly identified with moderate confidence, including 8 K<sub>la</sub> and 10 K<sub>cr</sub> peptides. The PSMs of synthetic peptides were compared to the originally identified PSMs. All paired spectra are perfectly matched to each other. For each comparison, the upper spectrum is the originally identified PSM, and the lower one is the PSM of the synthetic peptide.

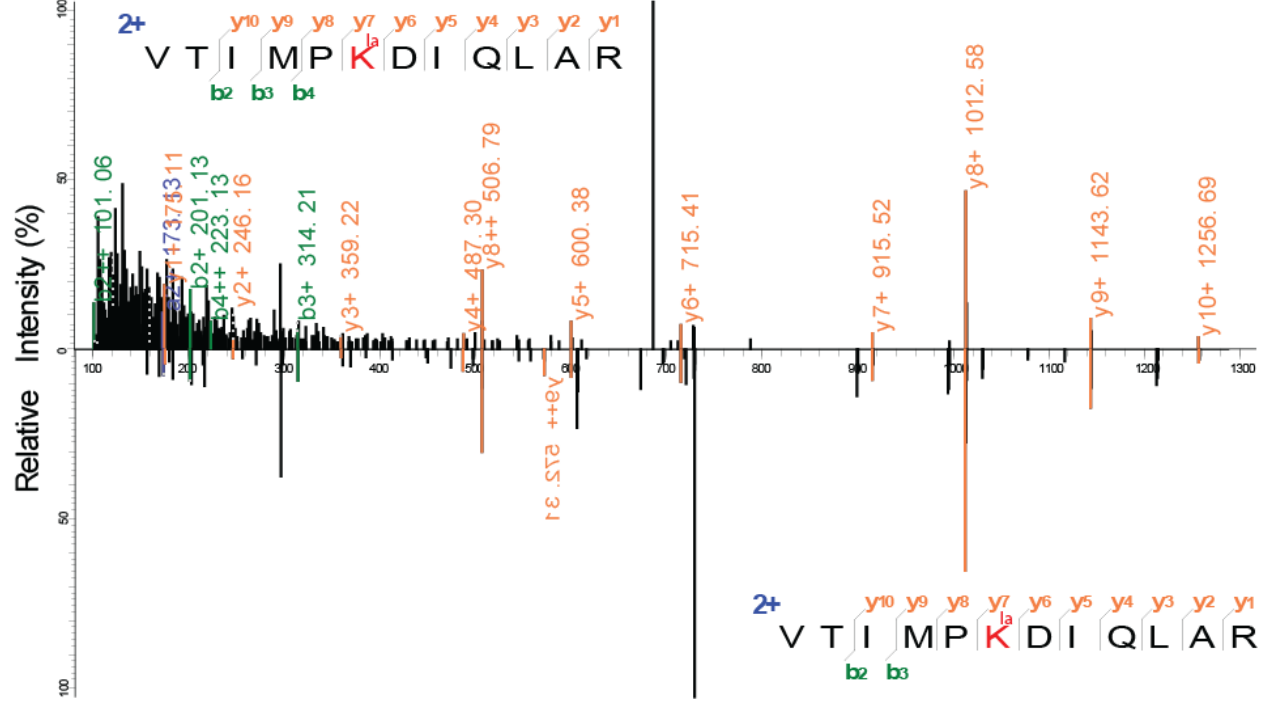
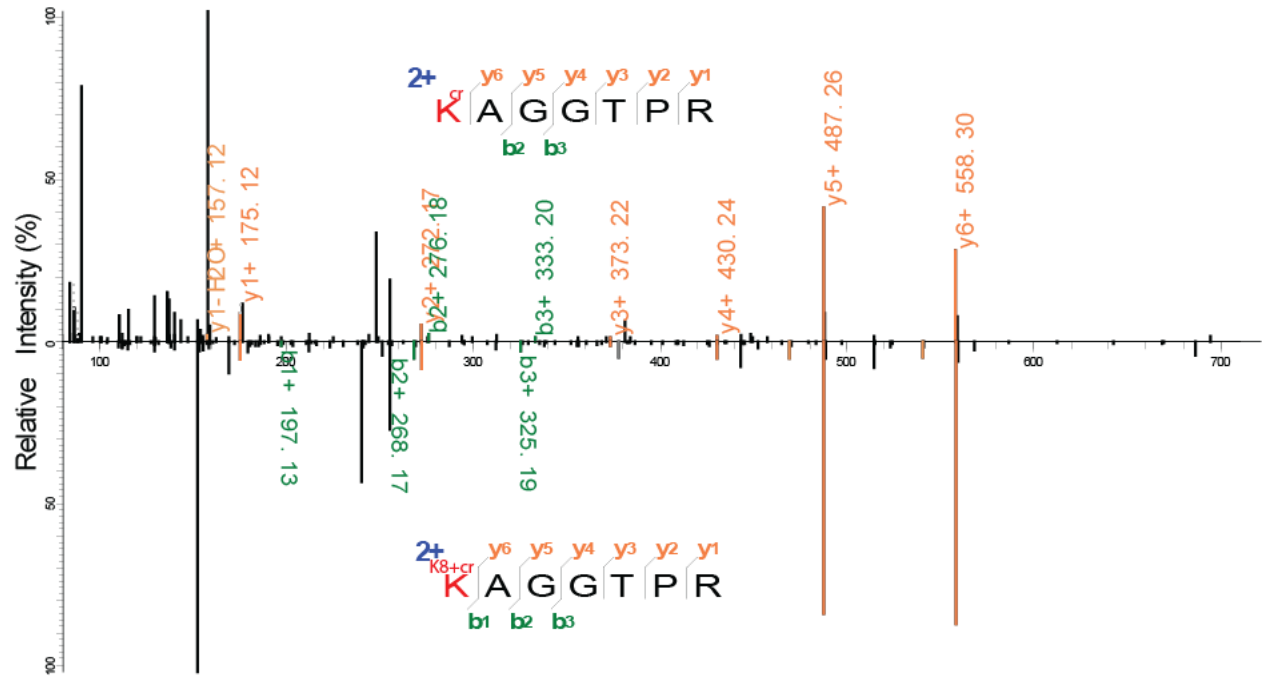


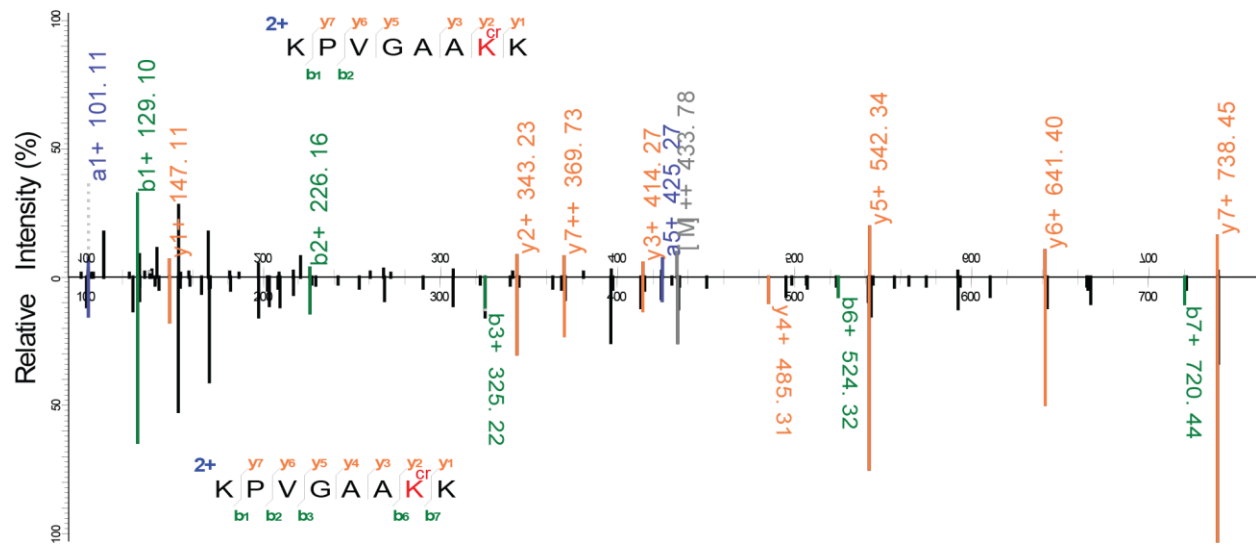
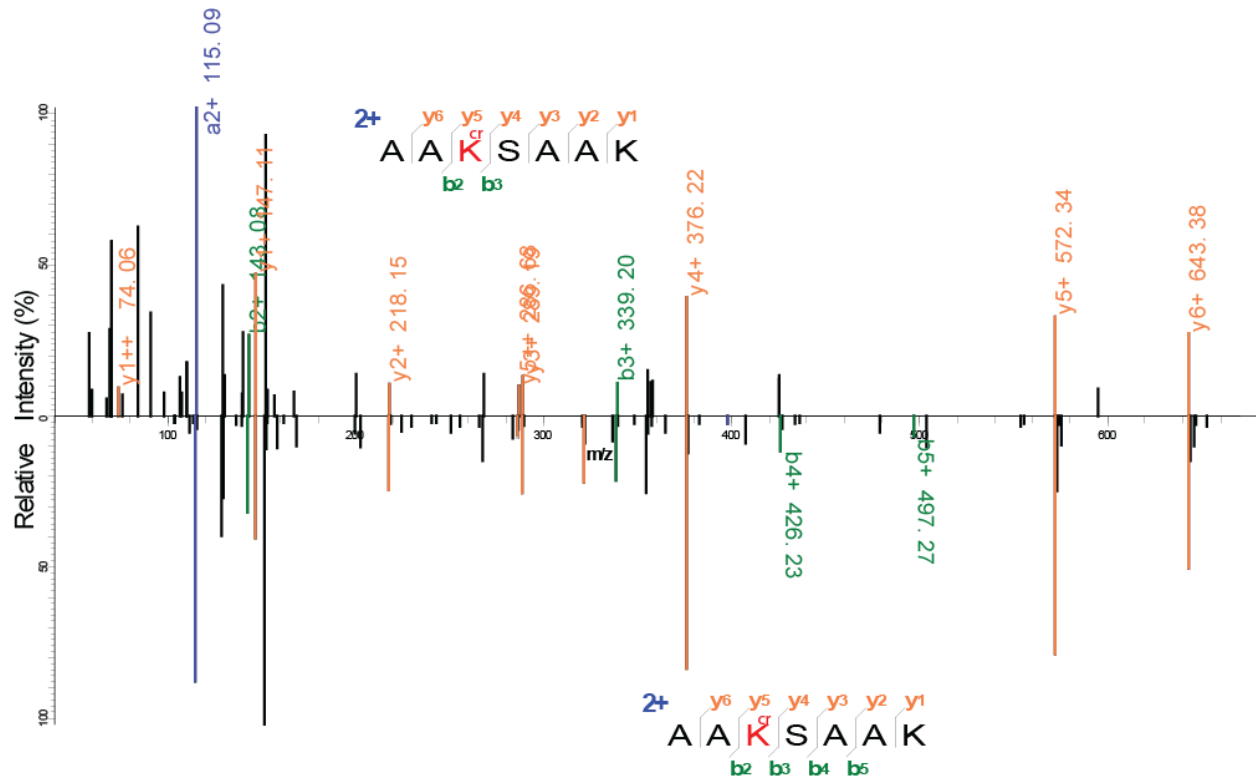


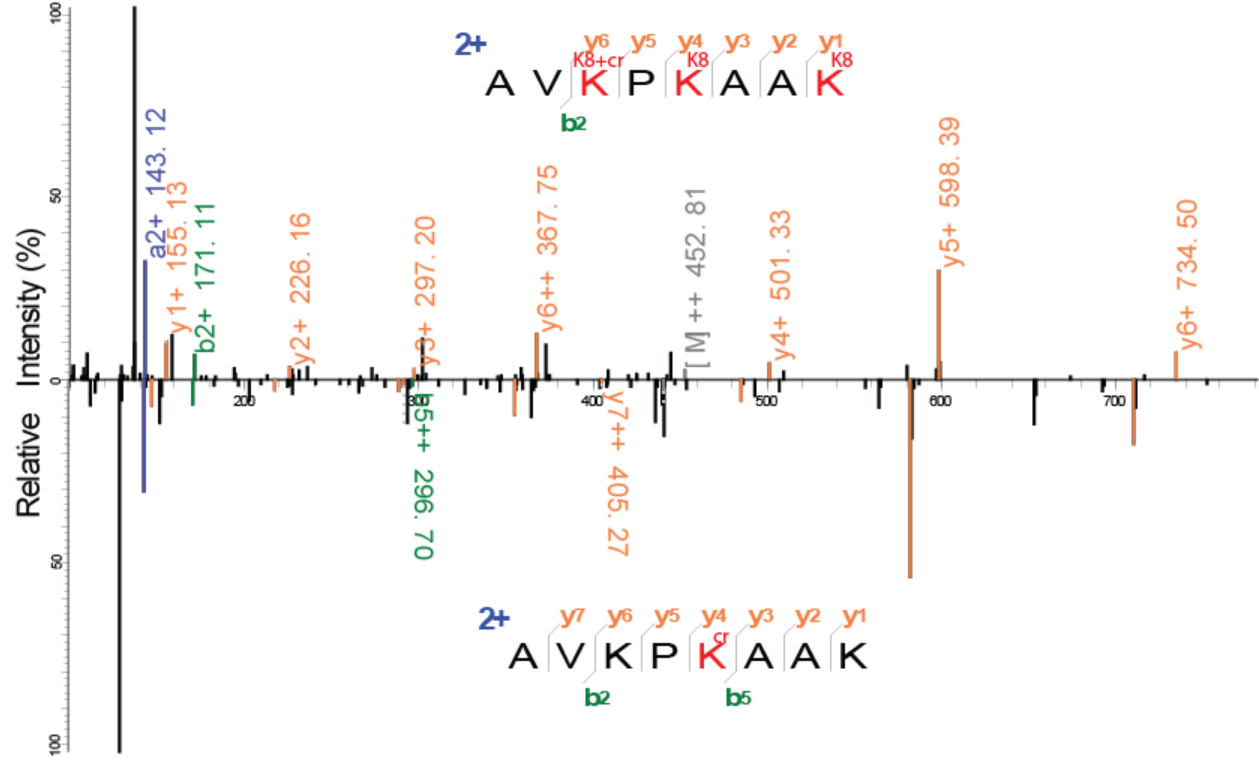
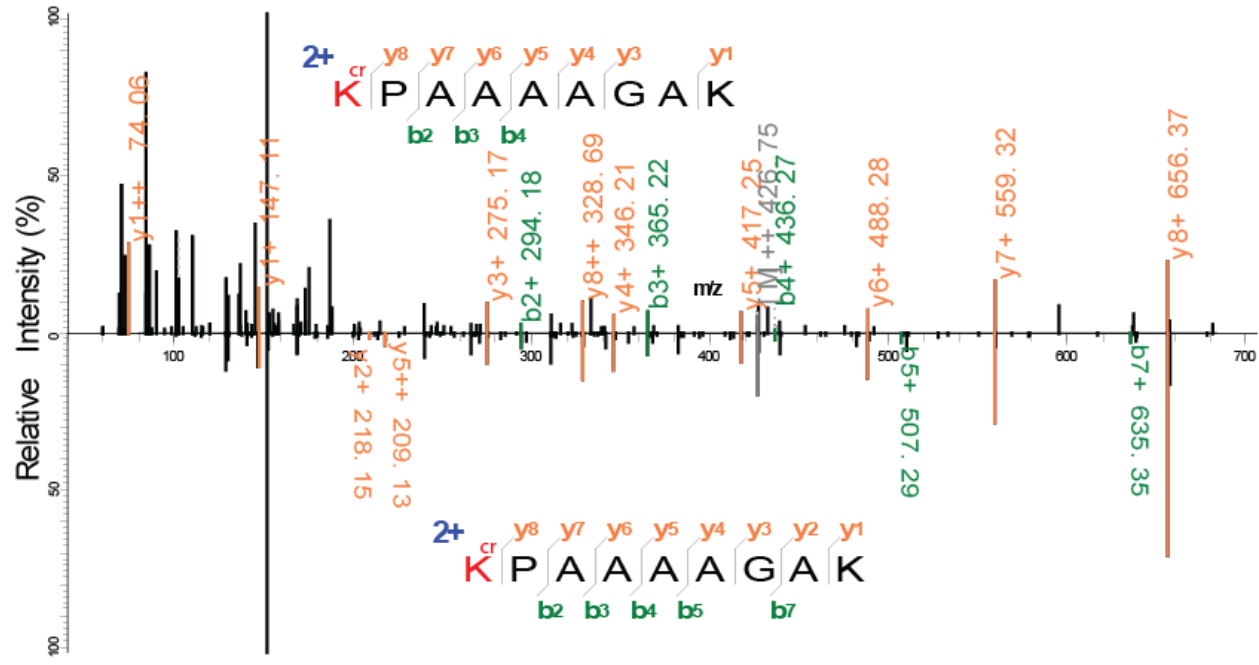


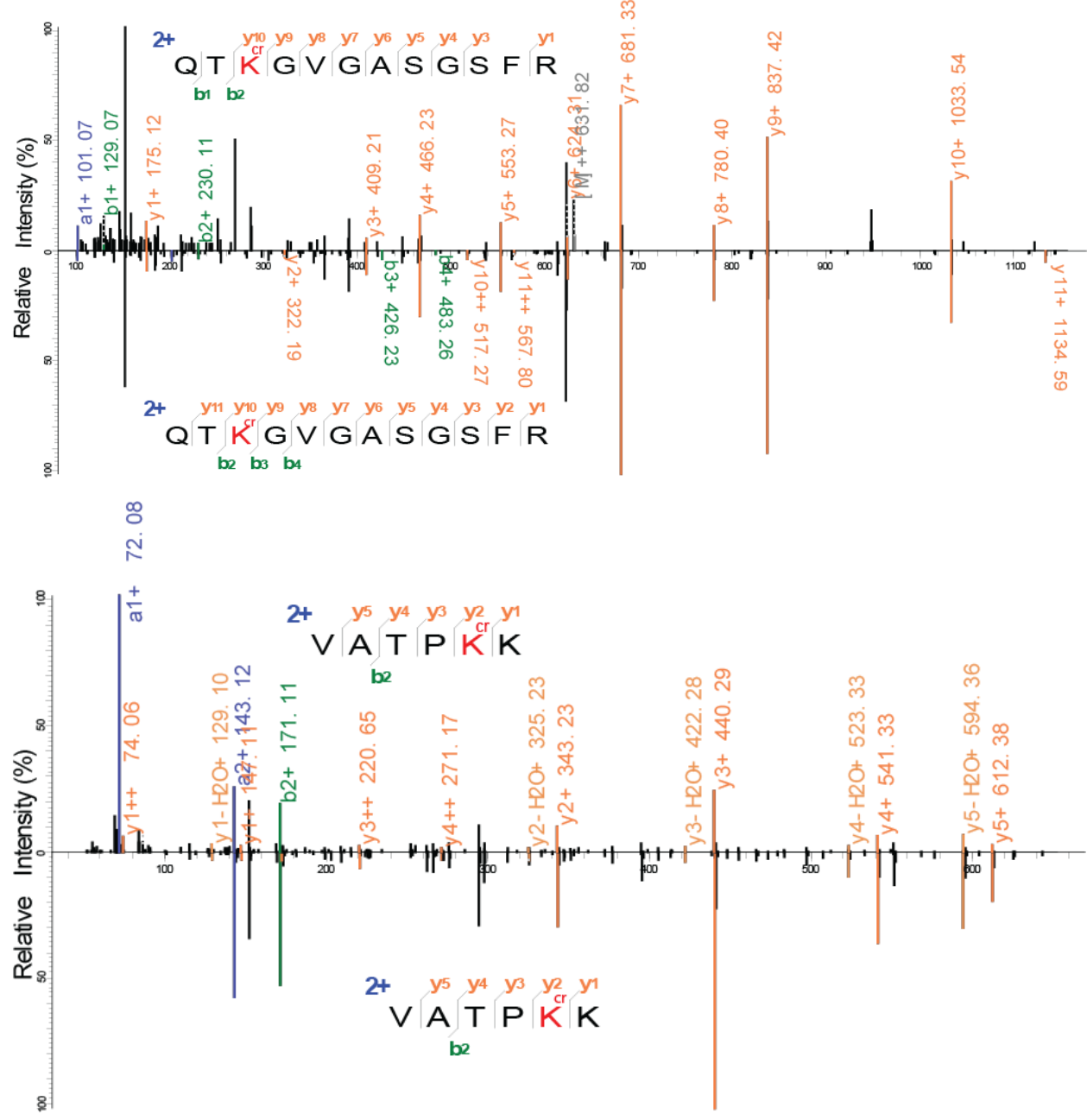


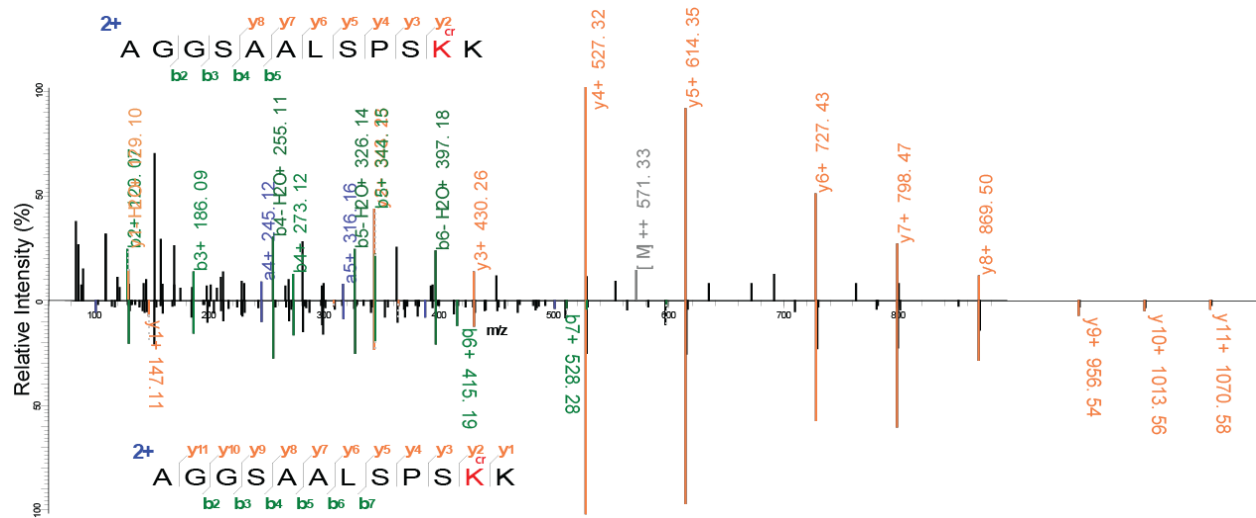
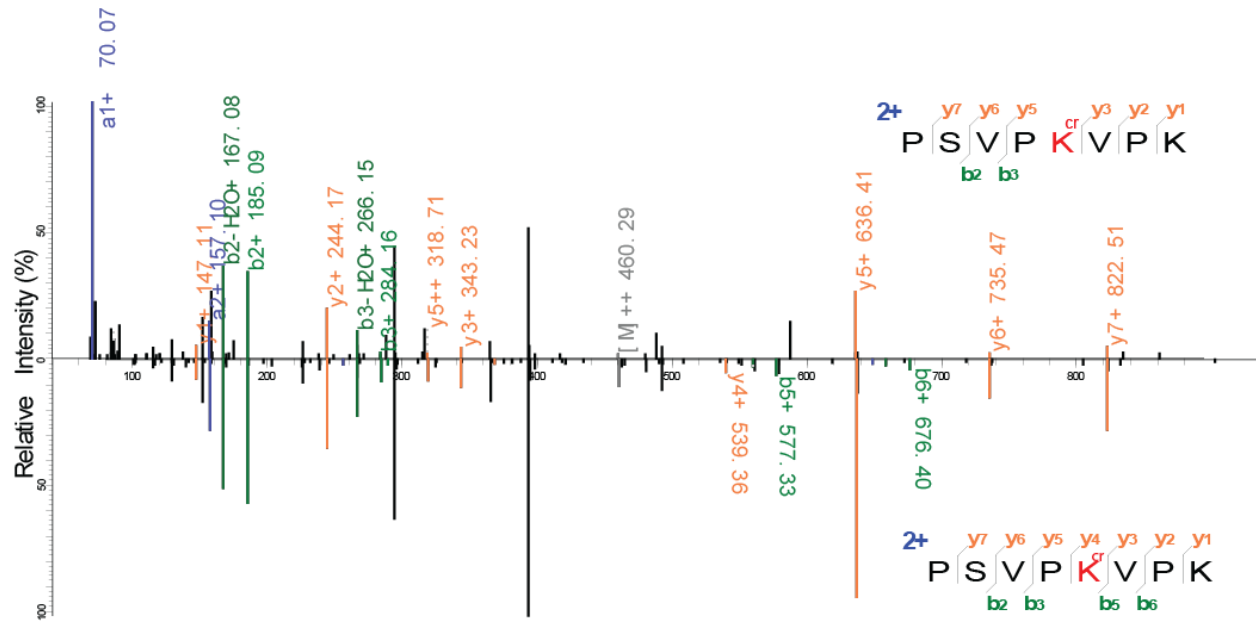


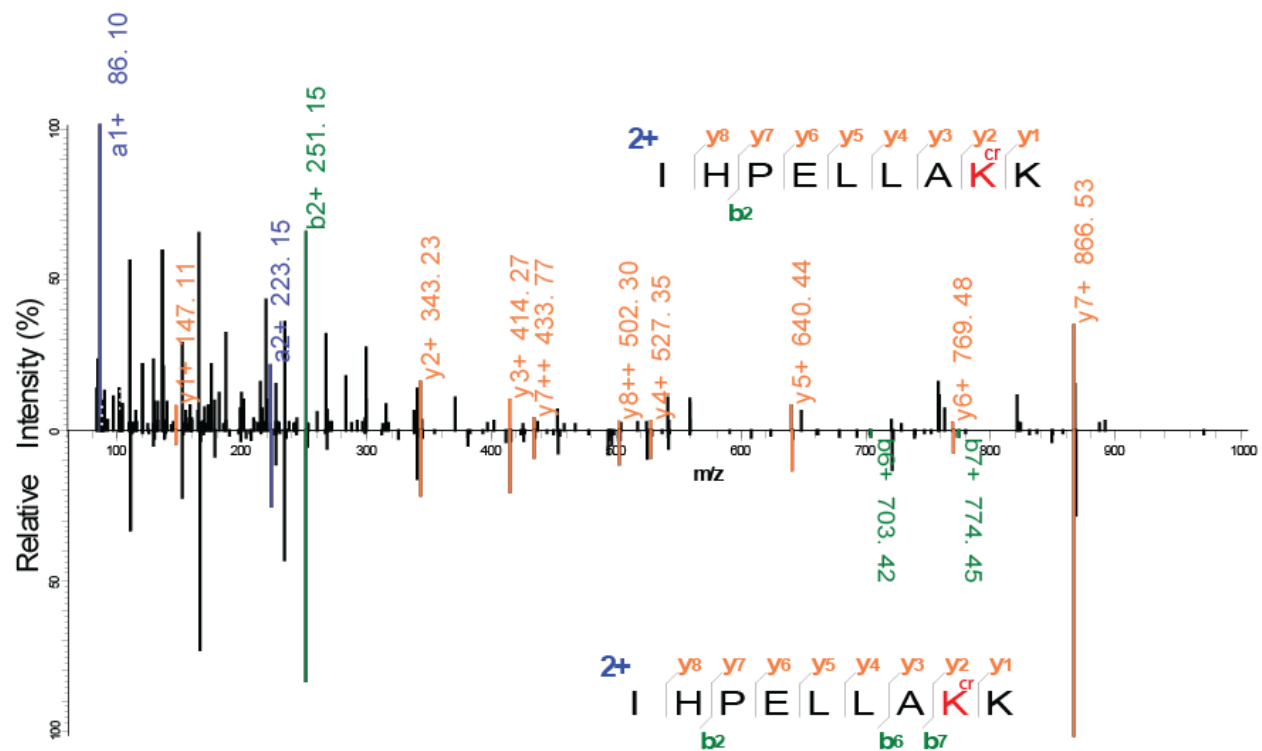




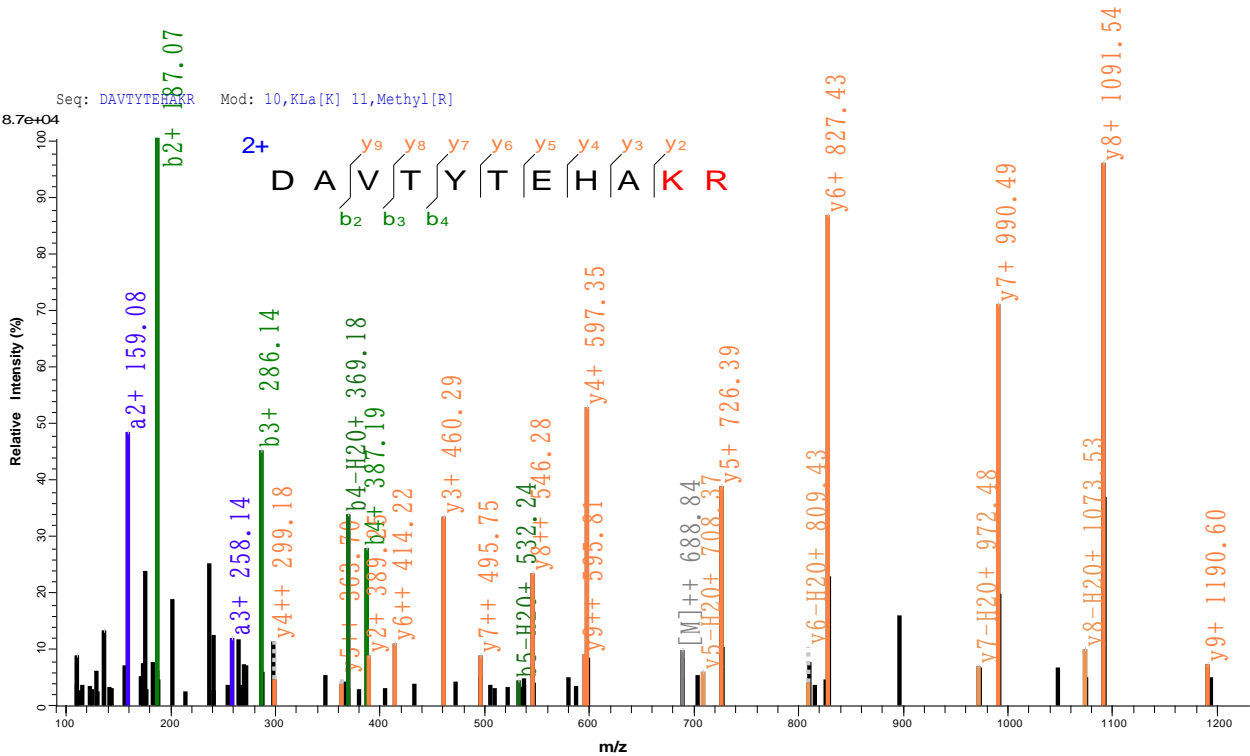
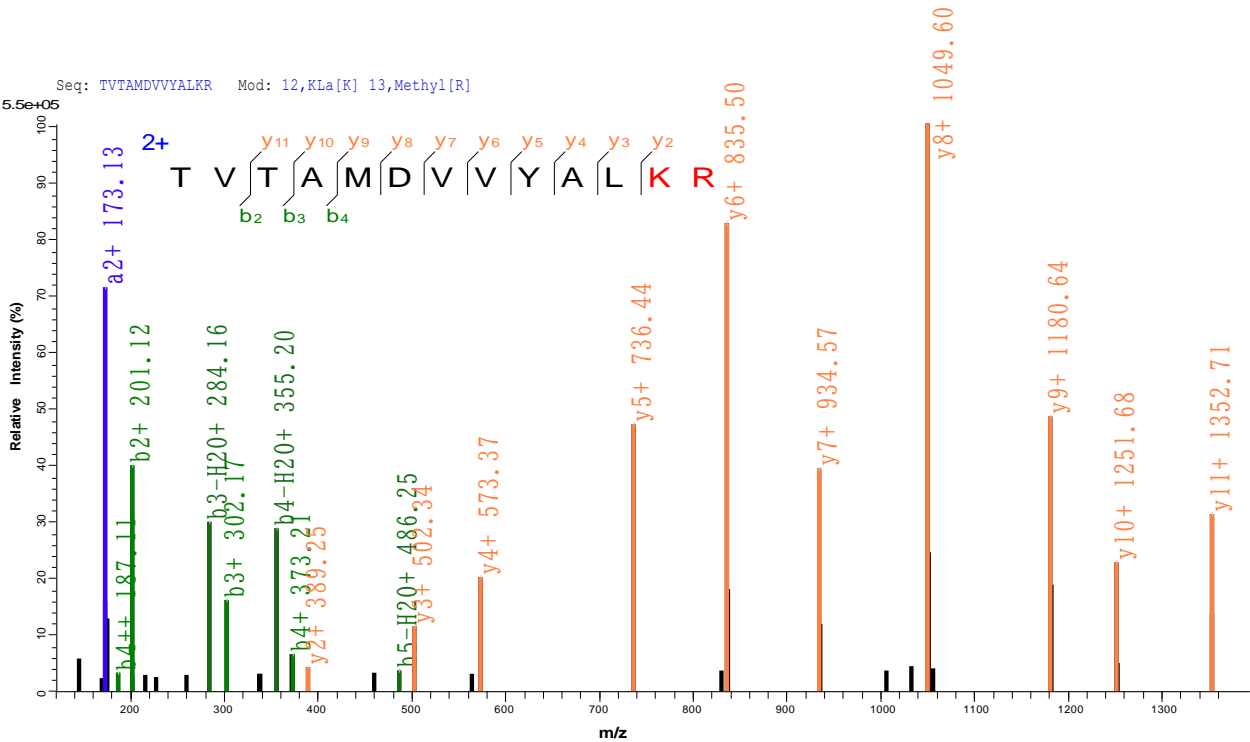




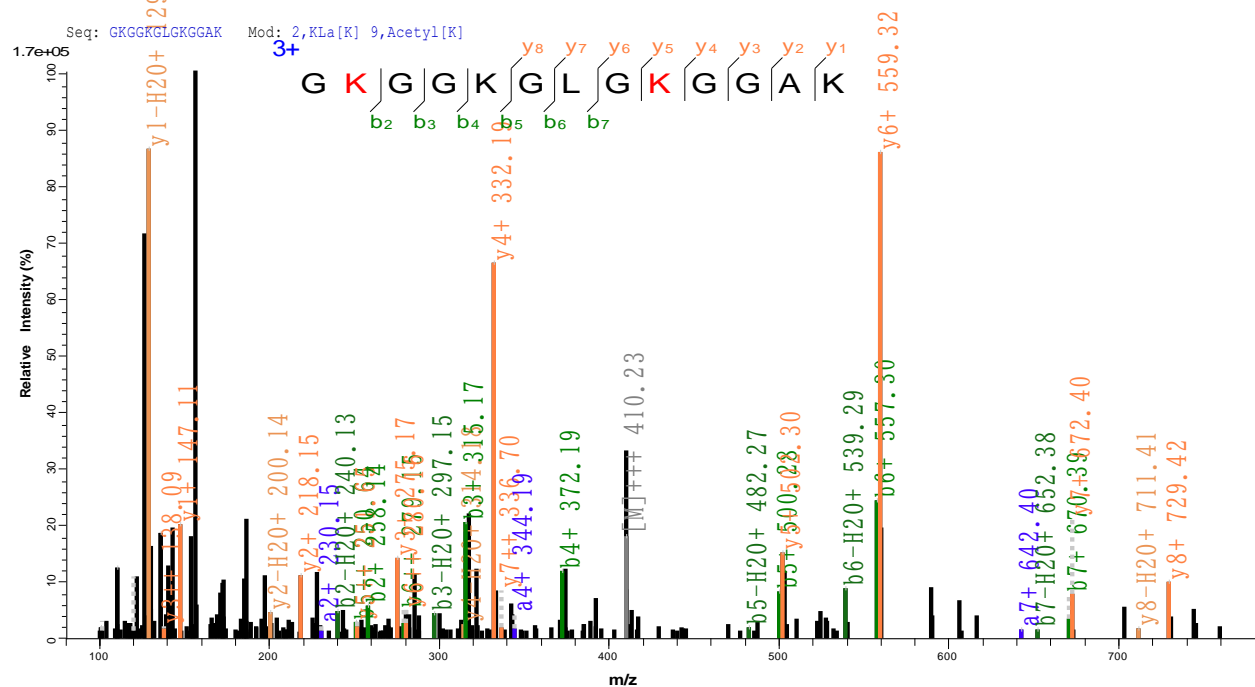
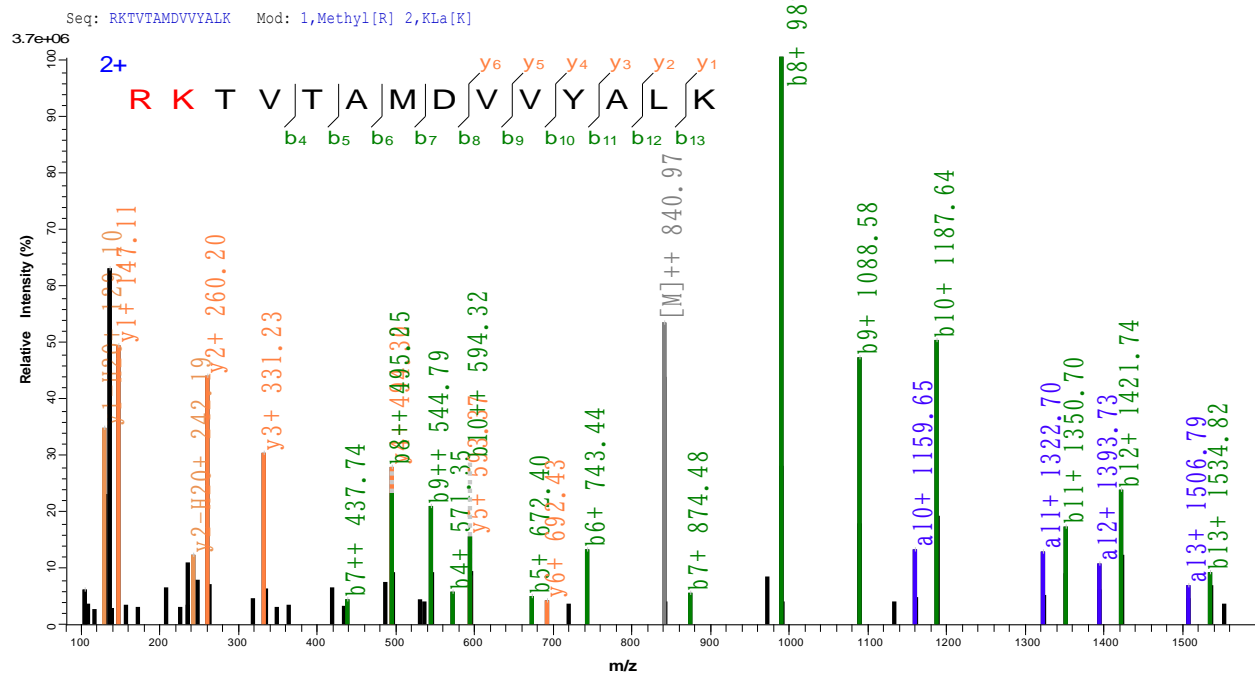


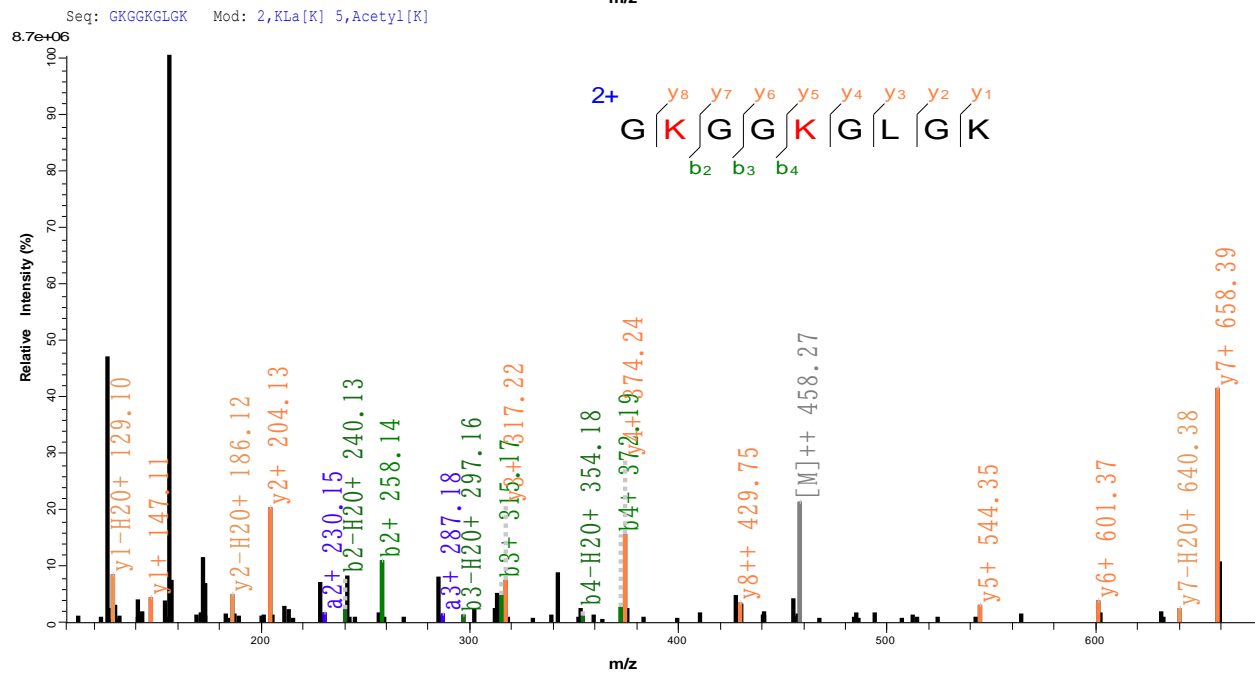
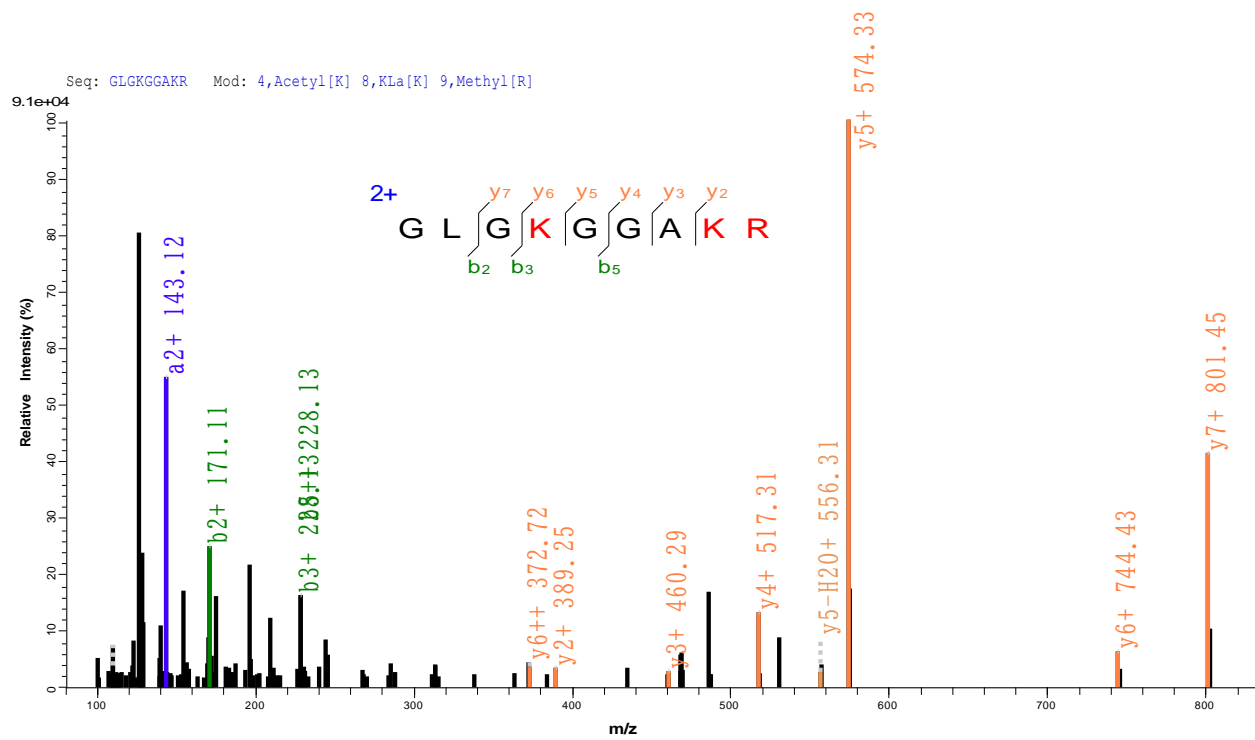


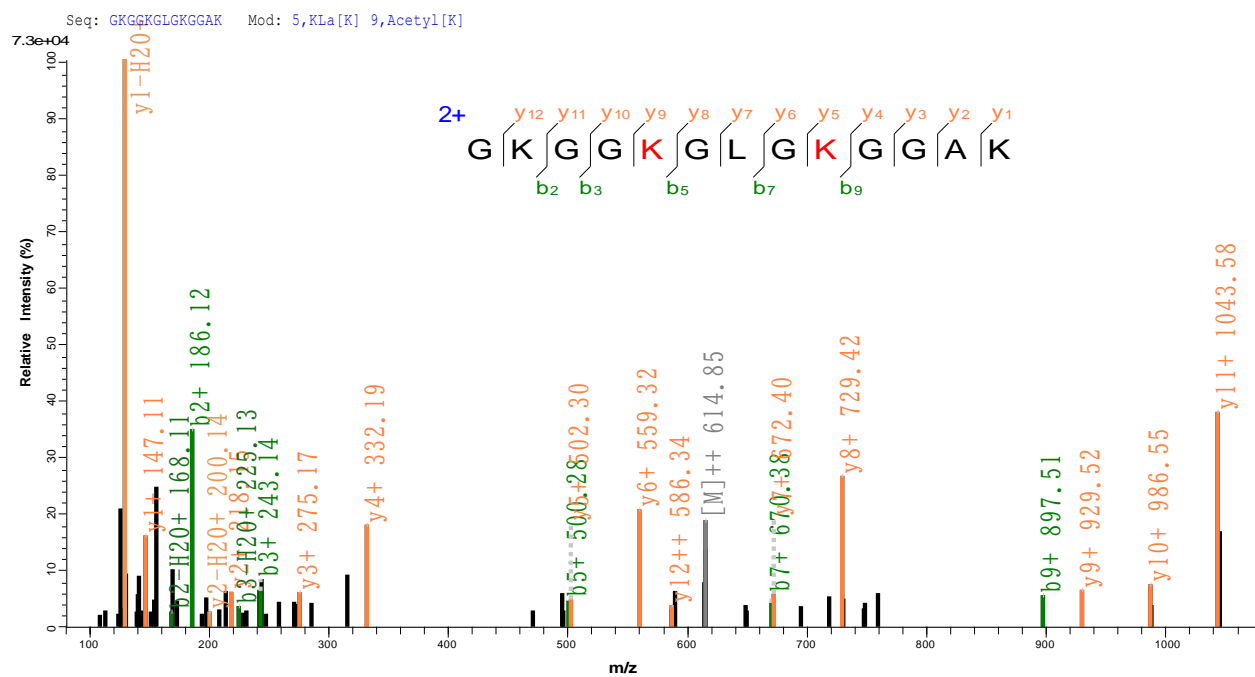
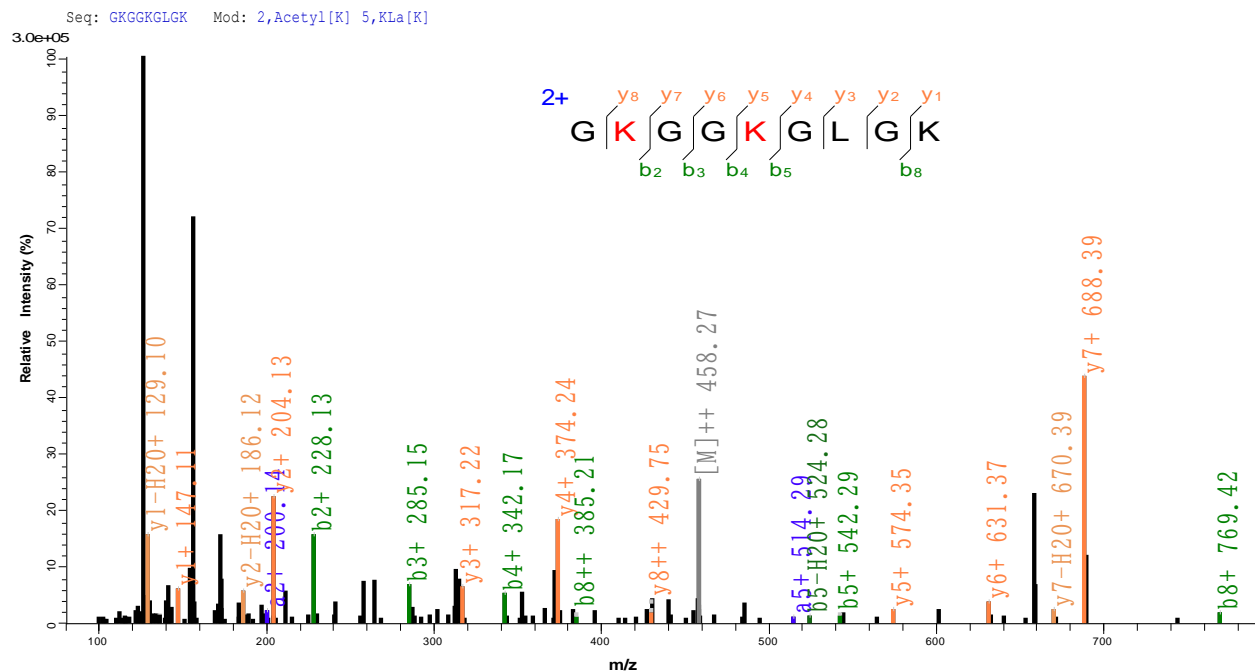
**Data S5. PSMs of peptides with co-existing histone marks.**

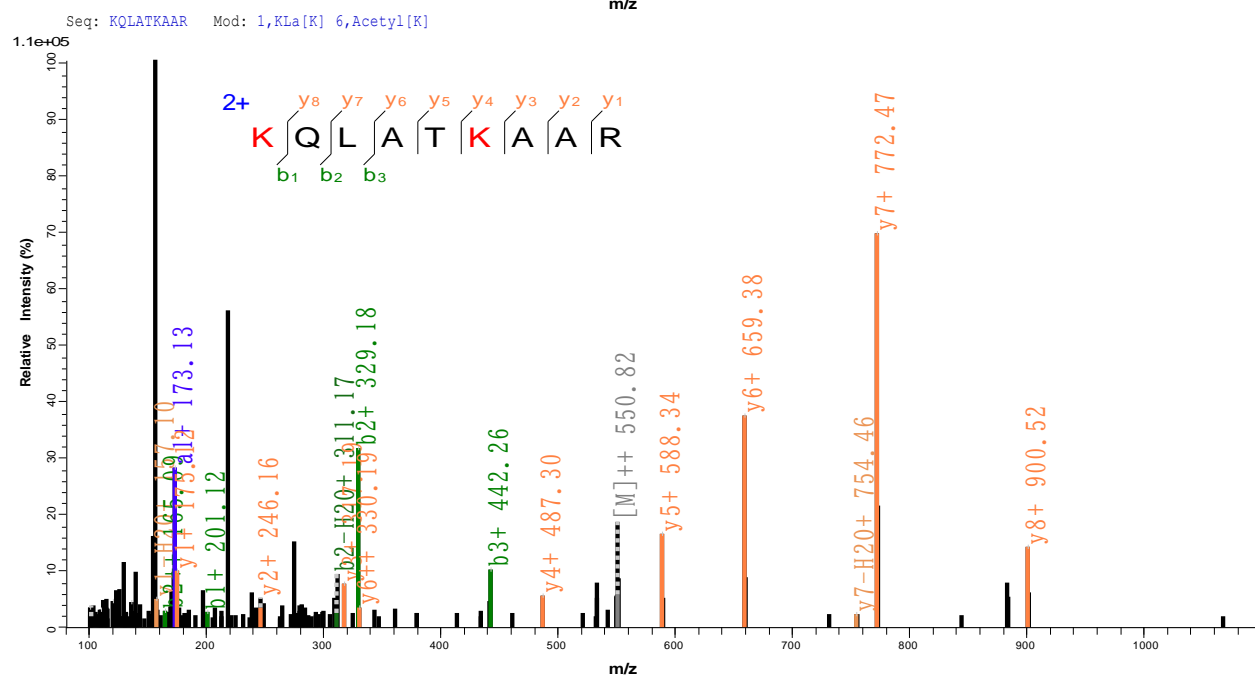
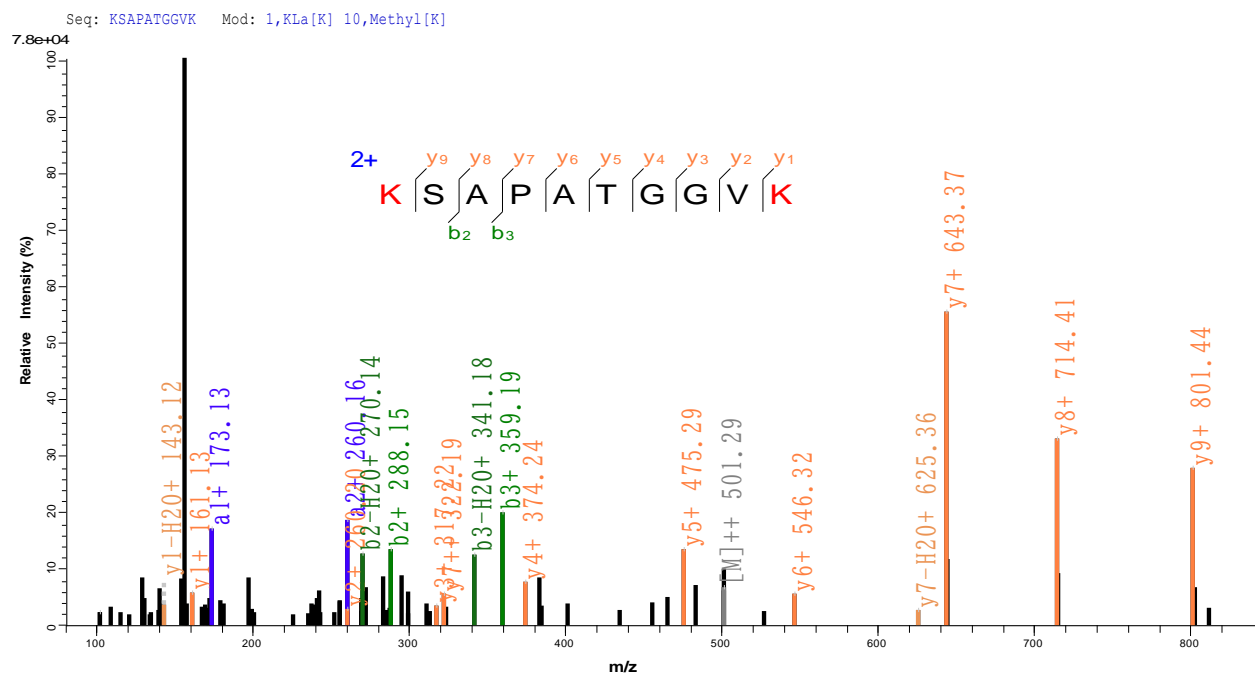


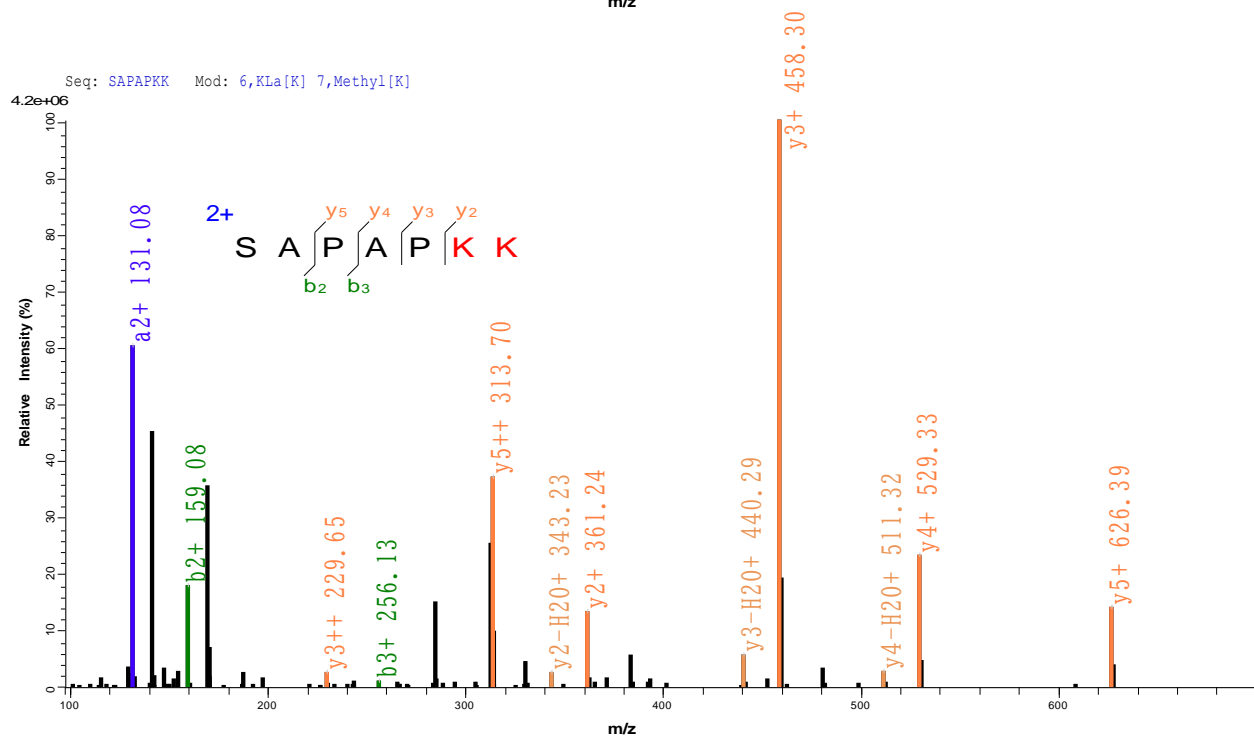
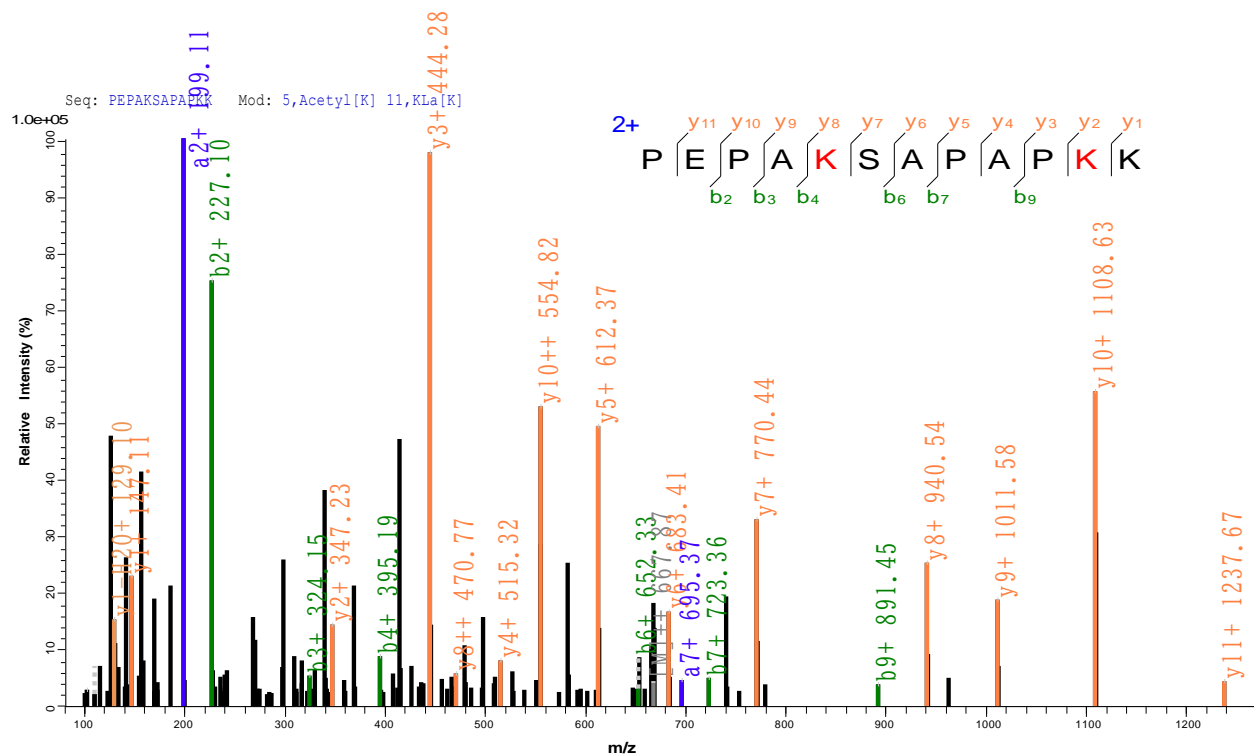


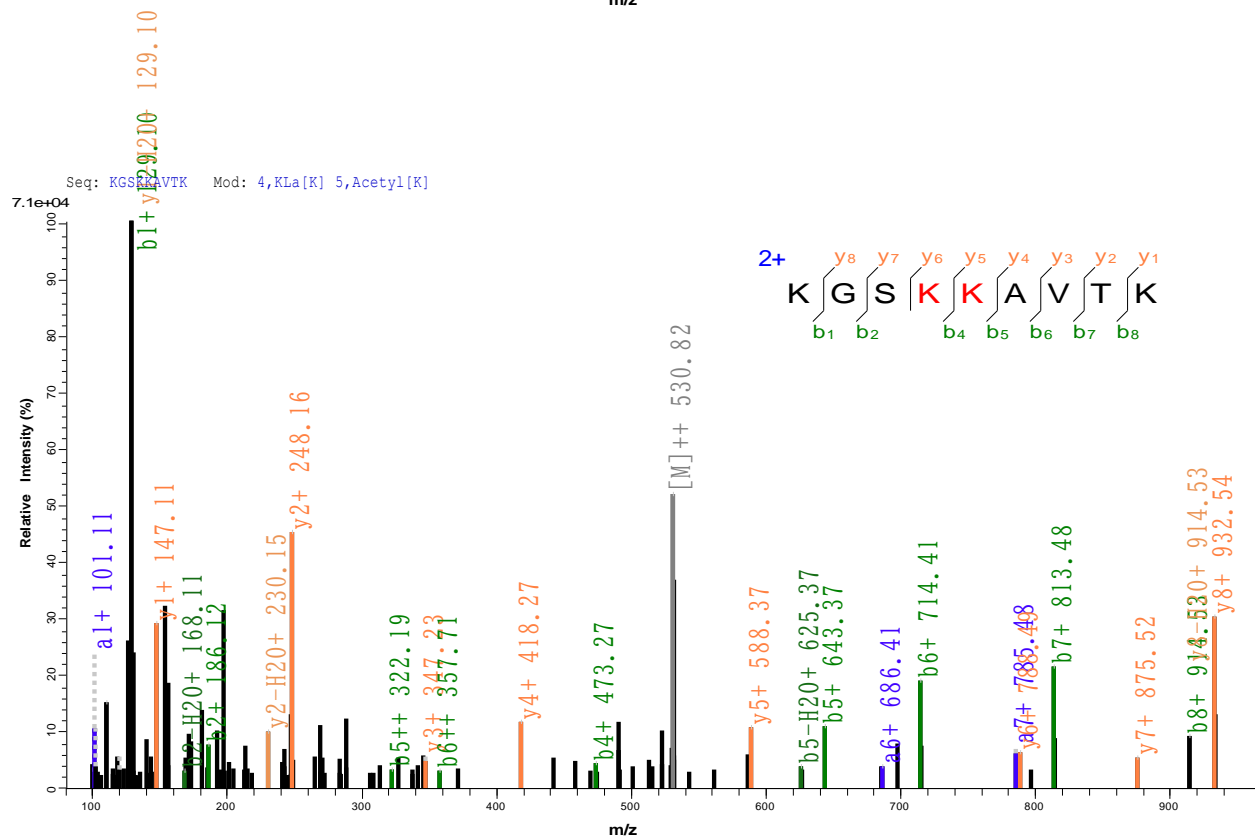
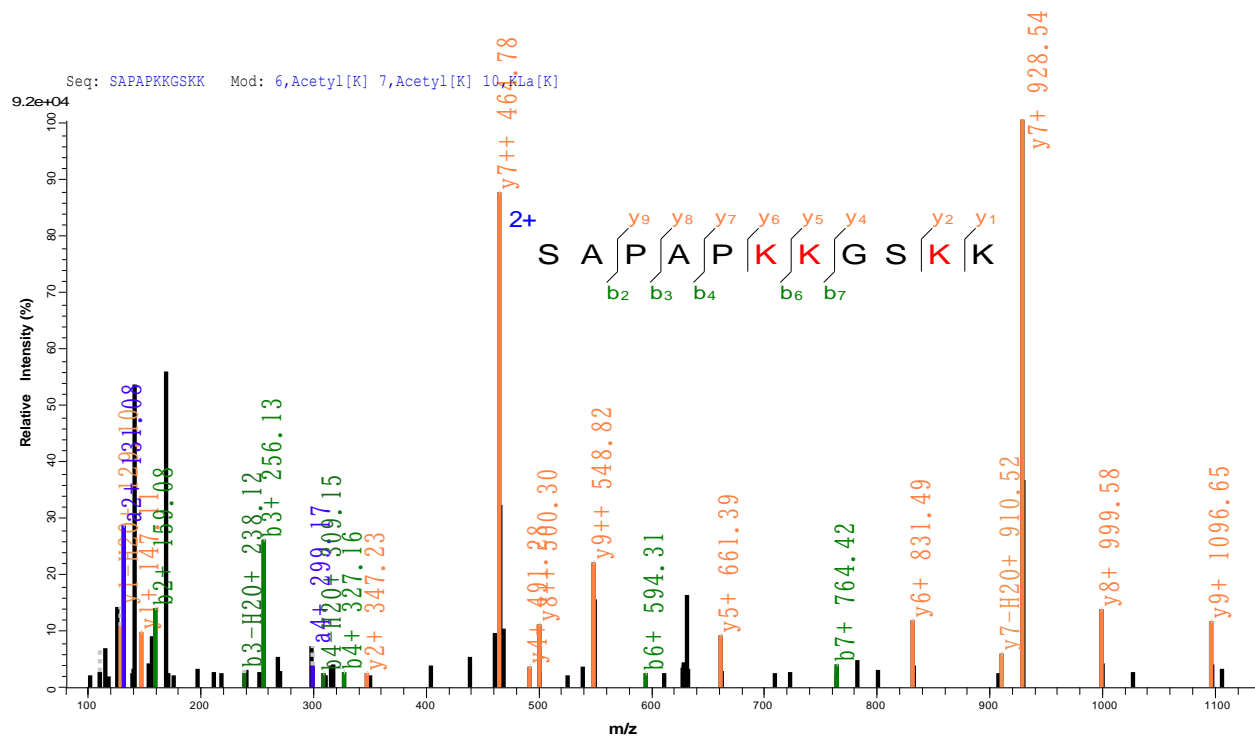


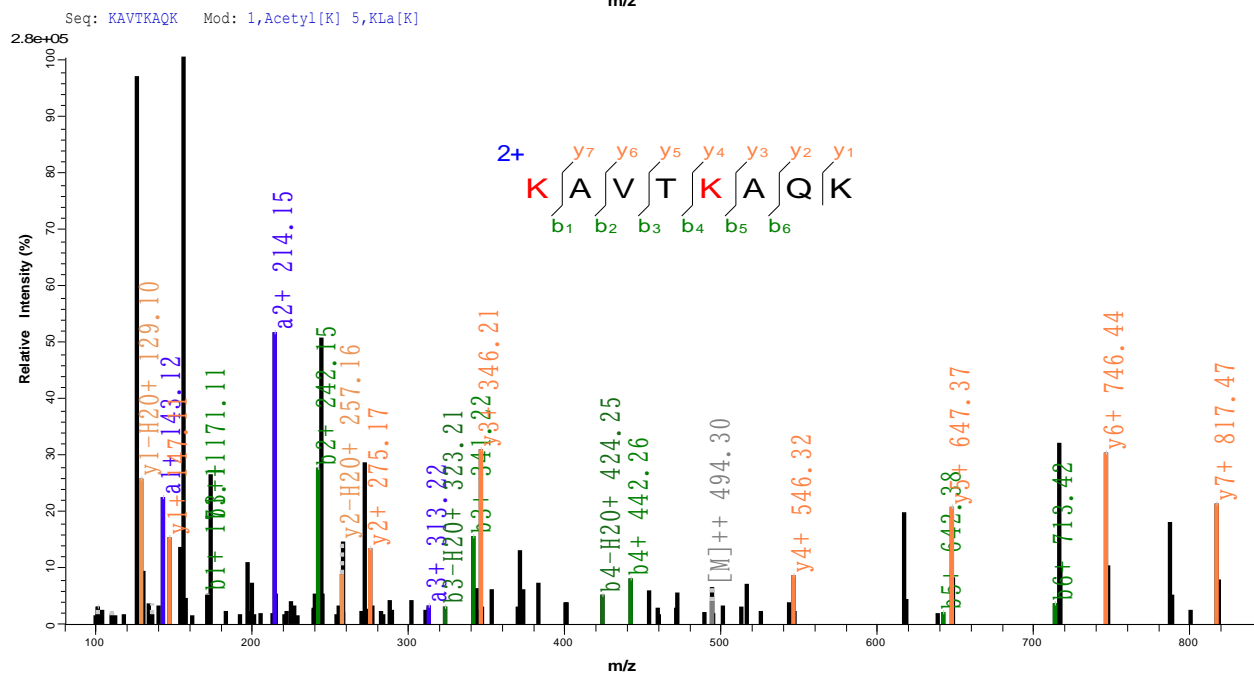
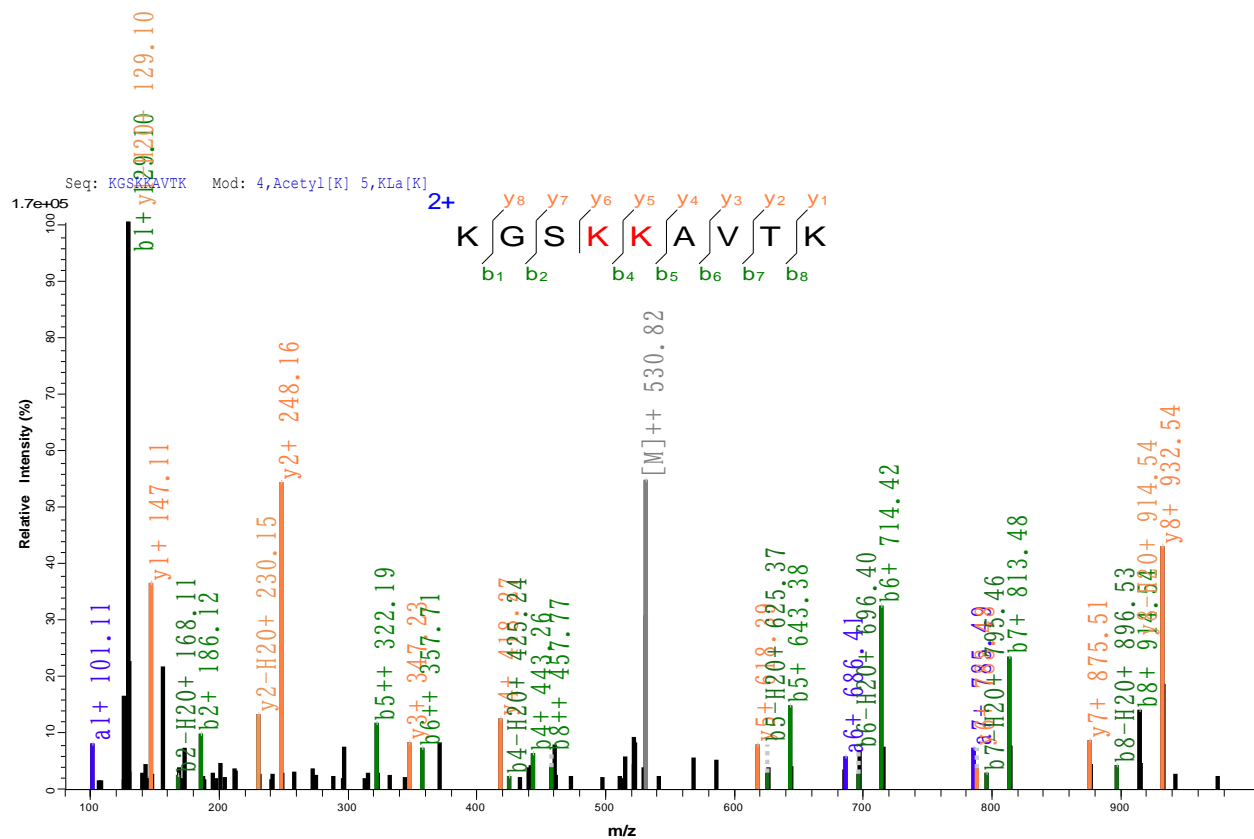


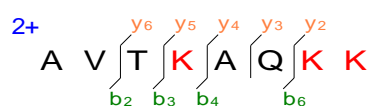
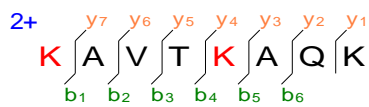




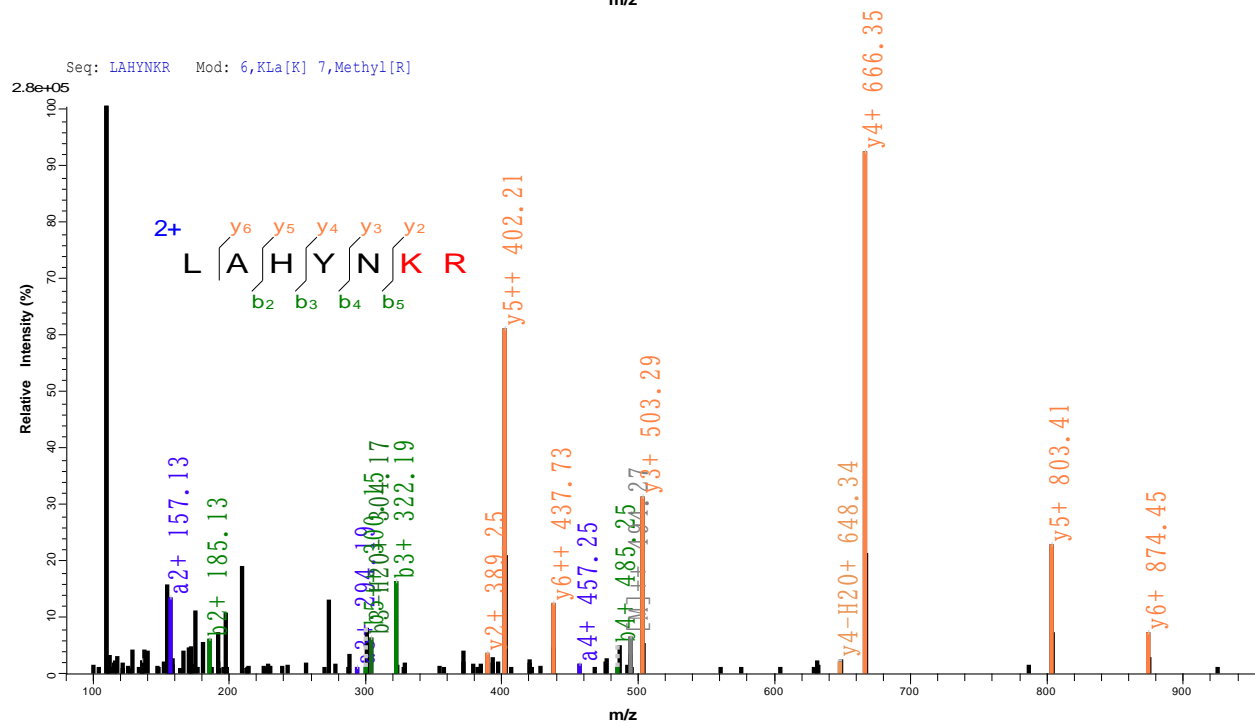
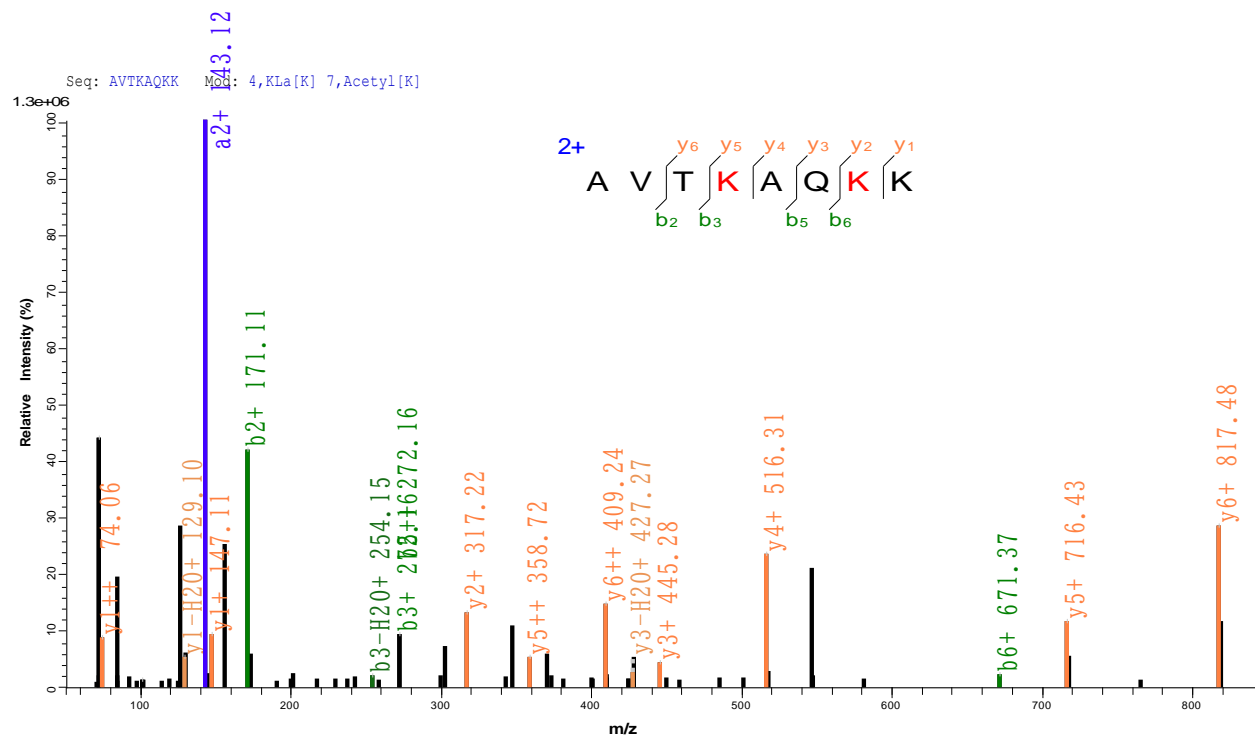


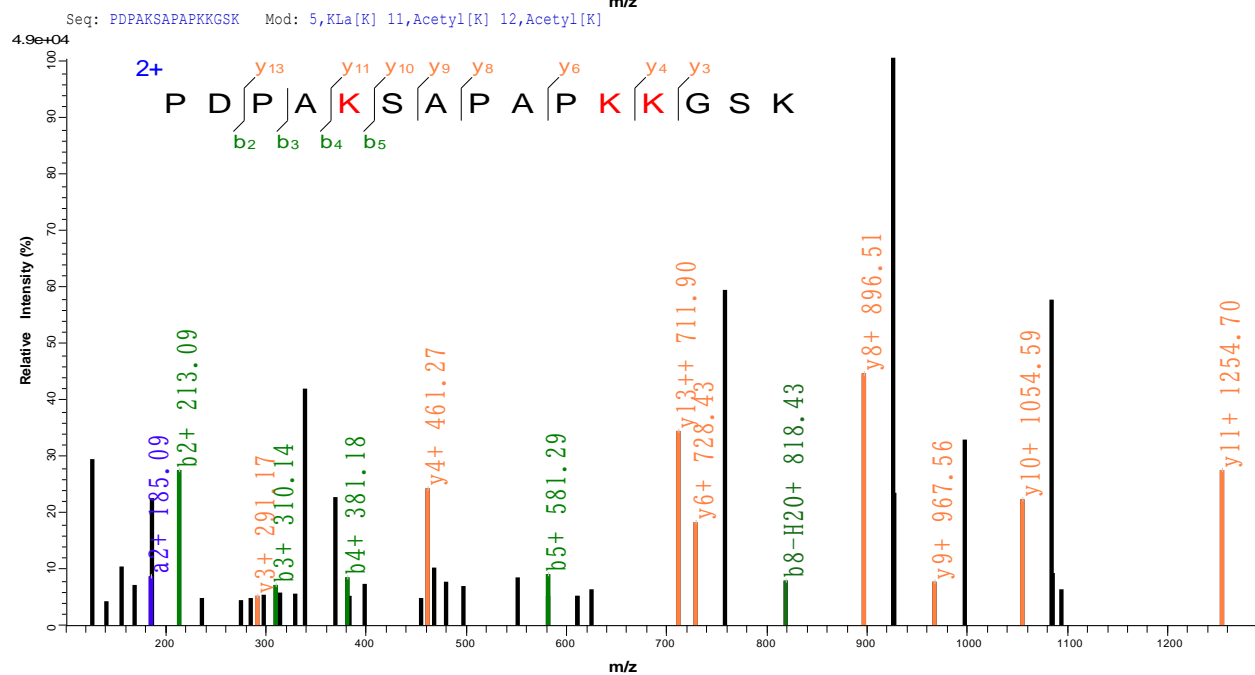
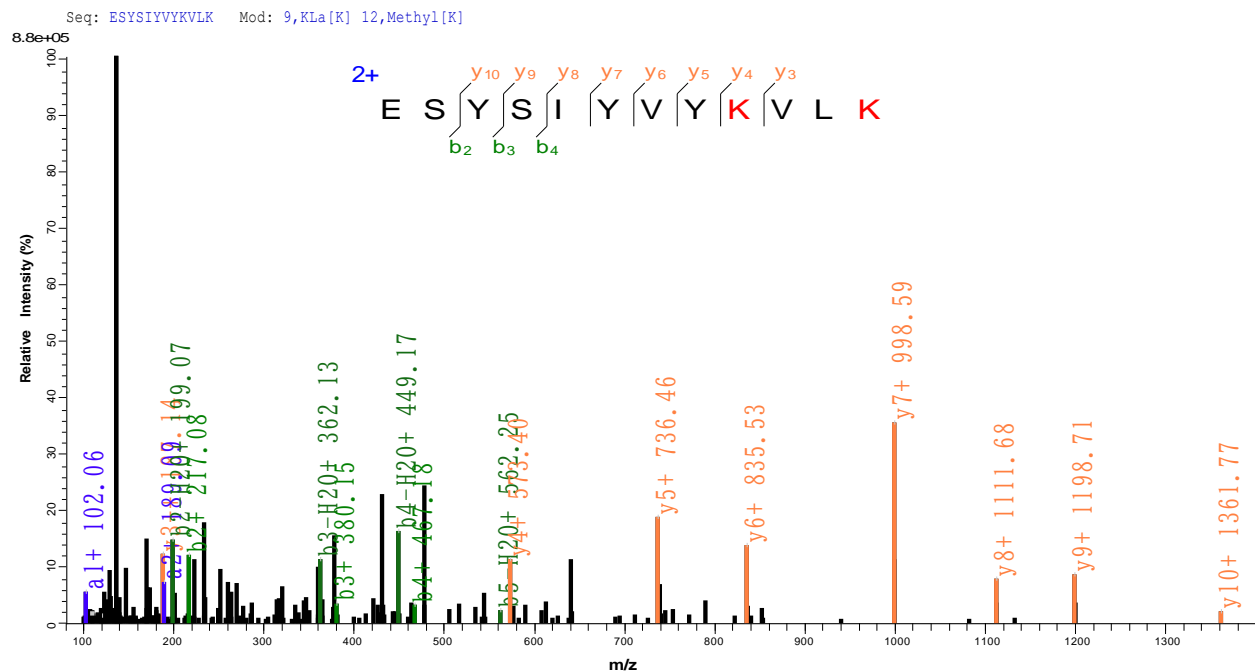


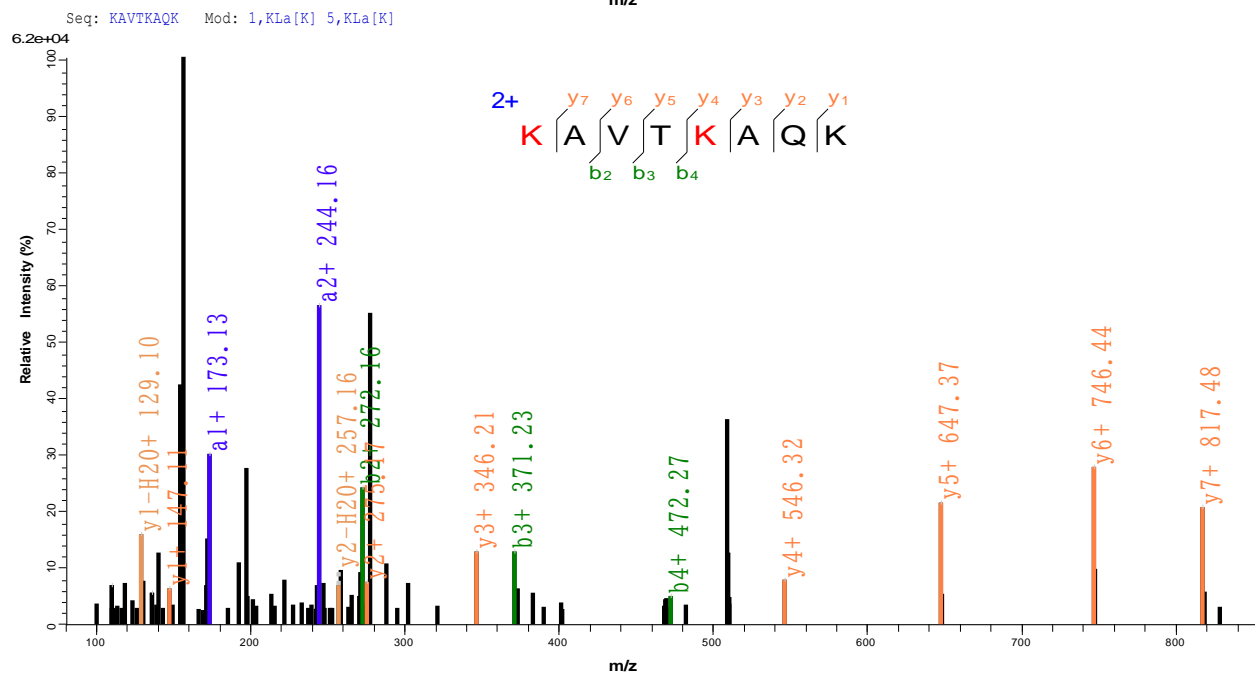
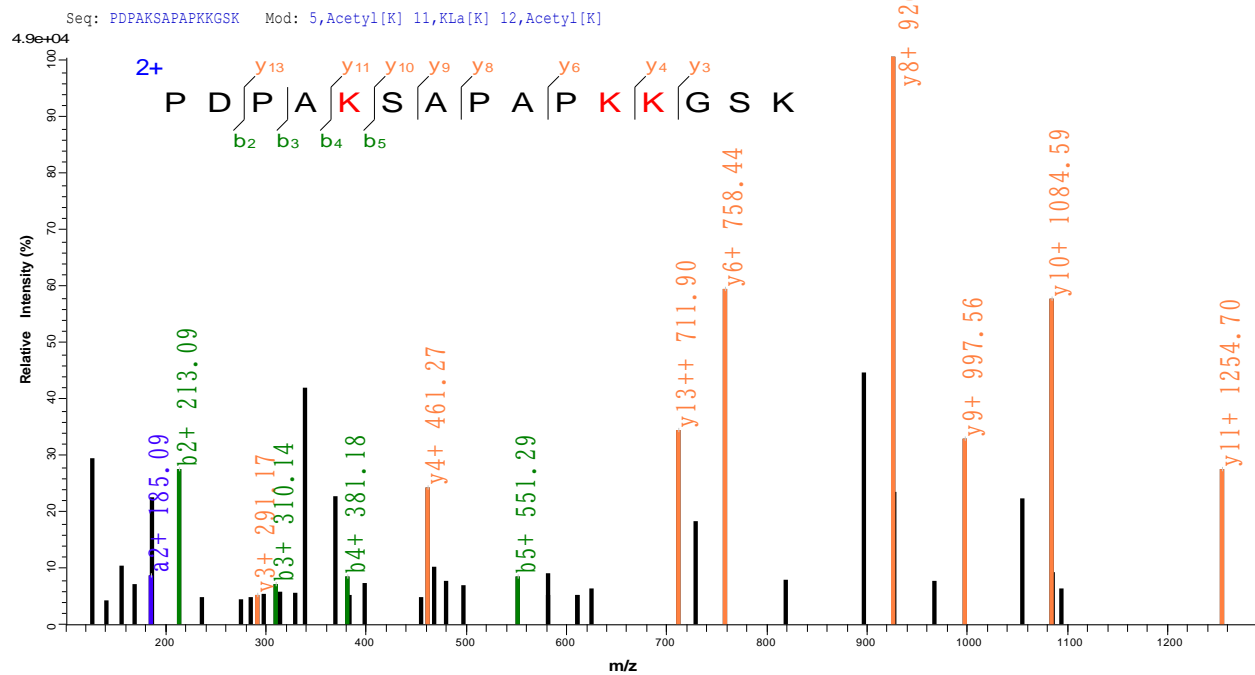


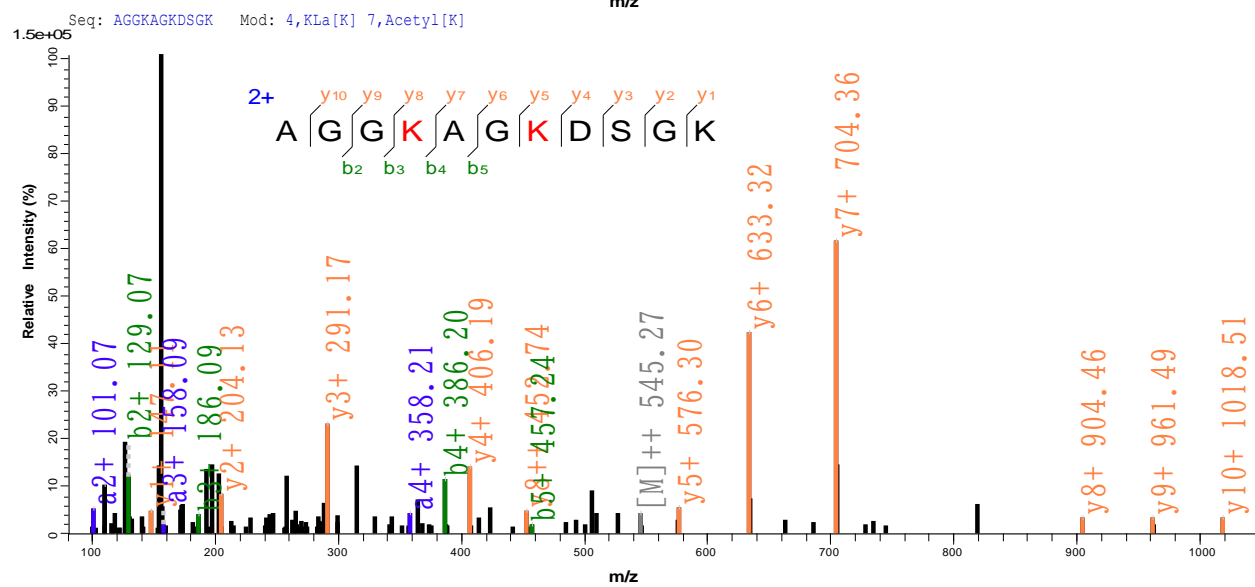
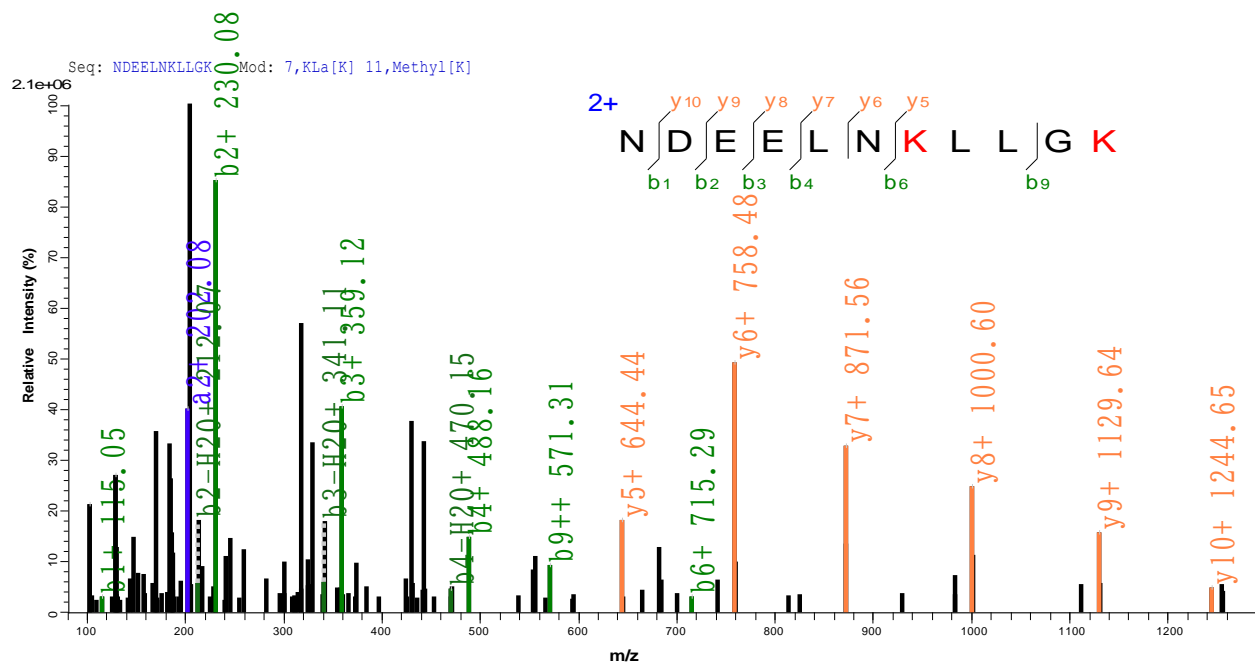


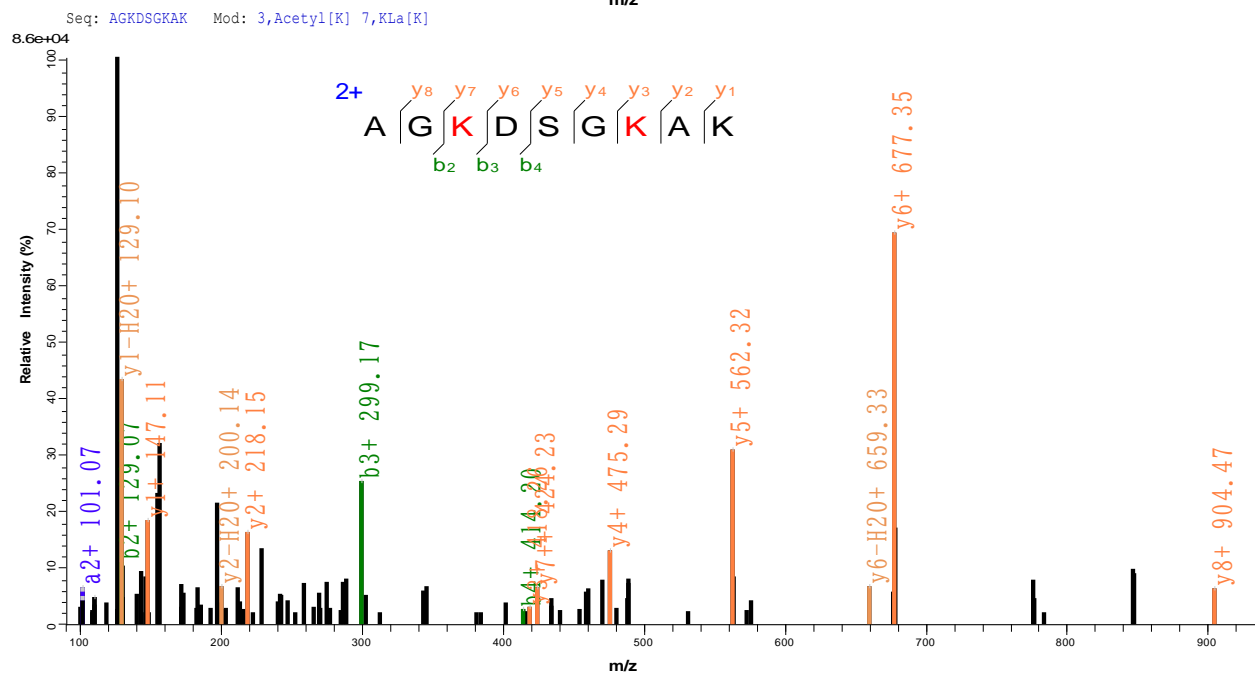
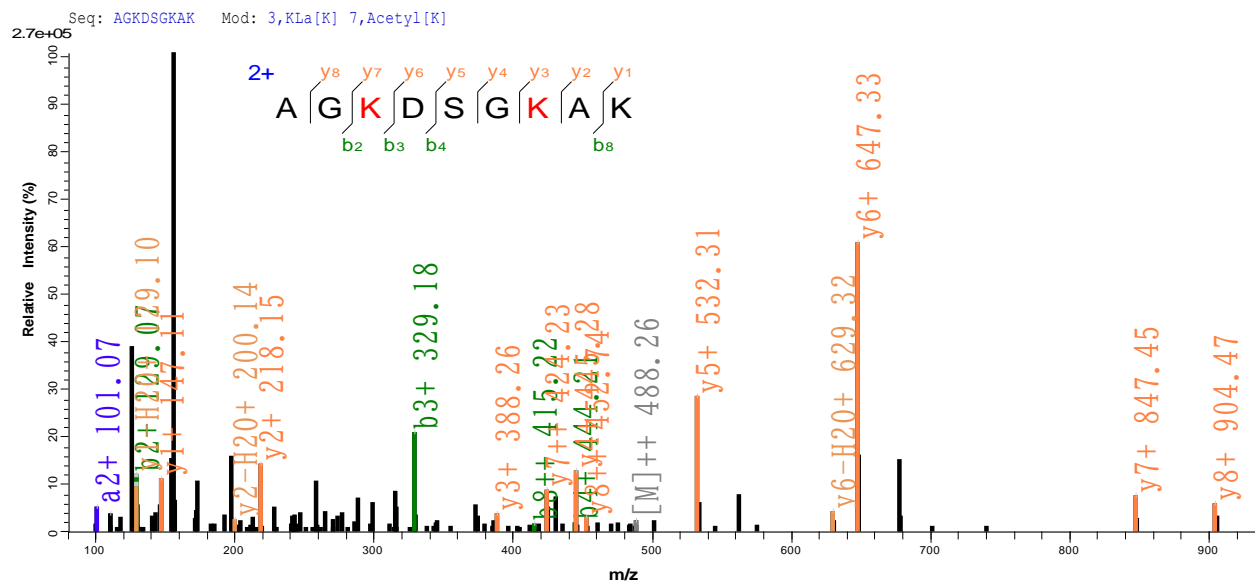


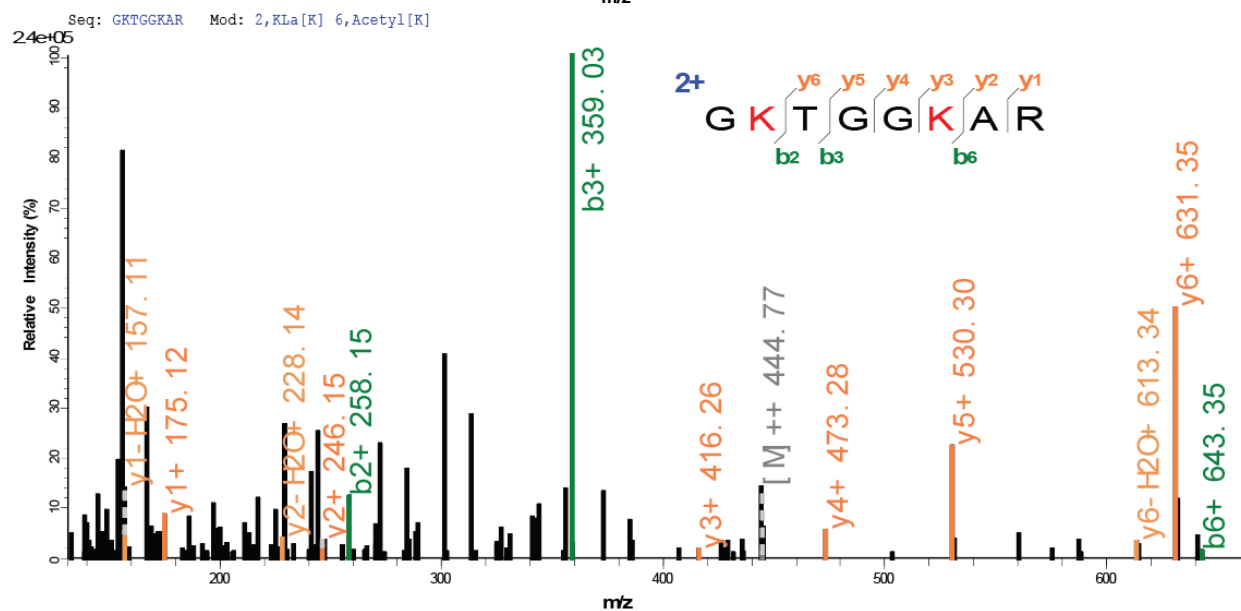
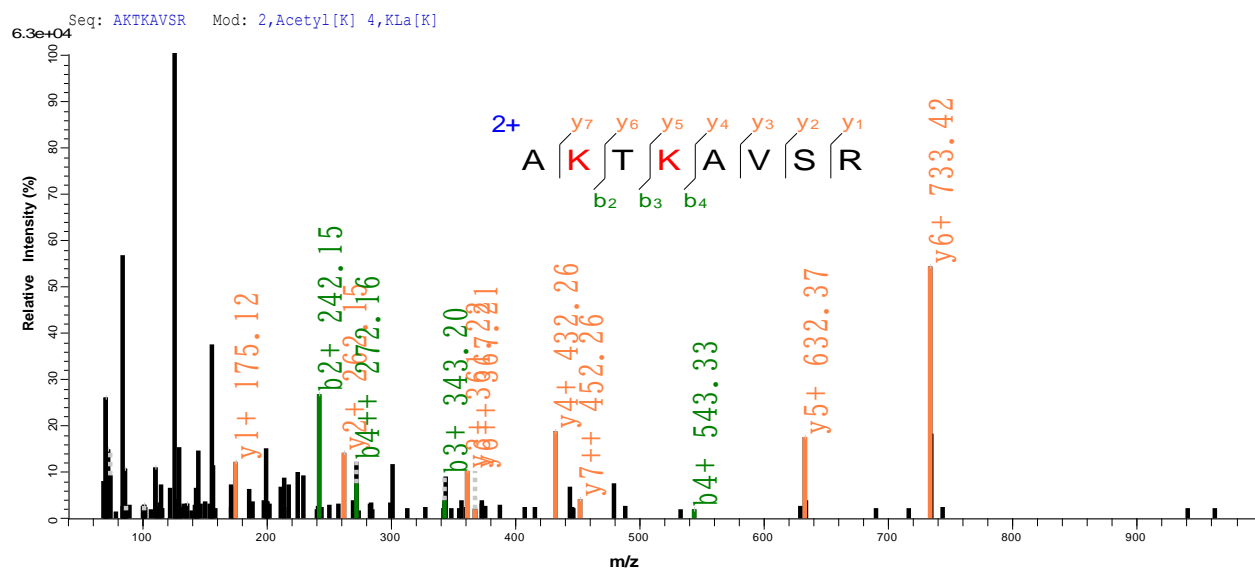


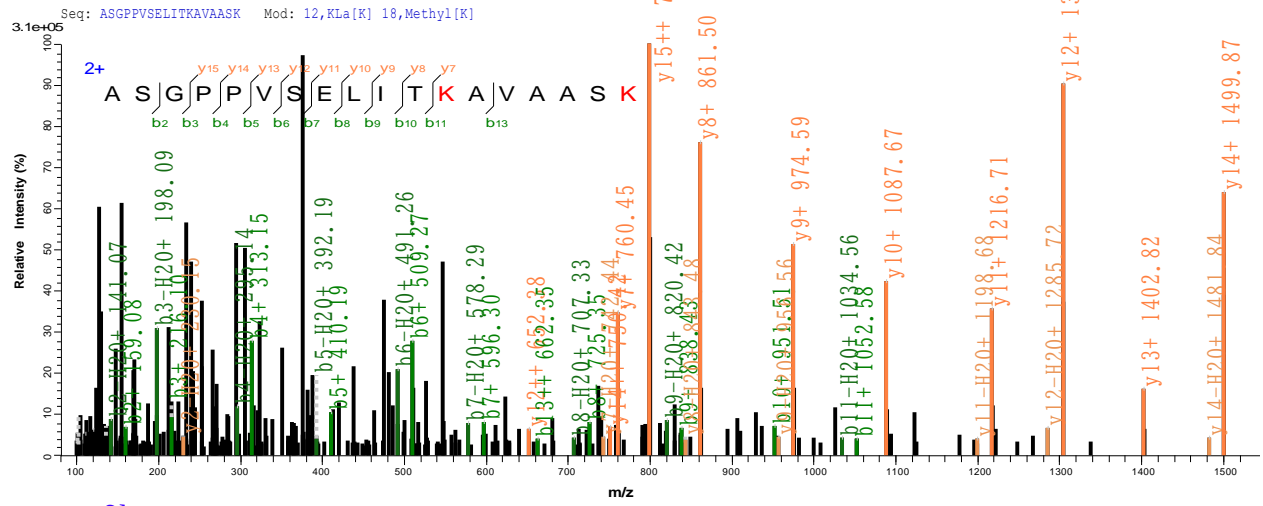


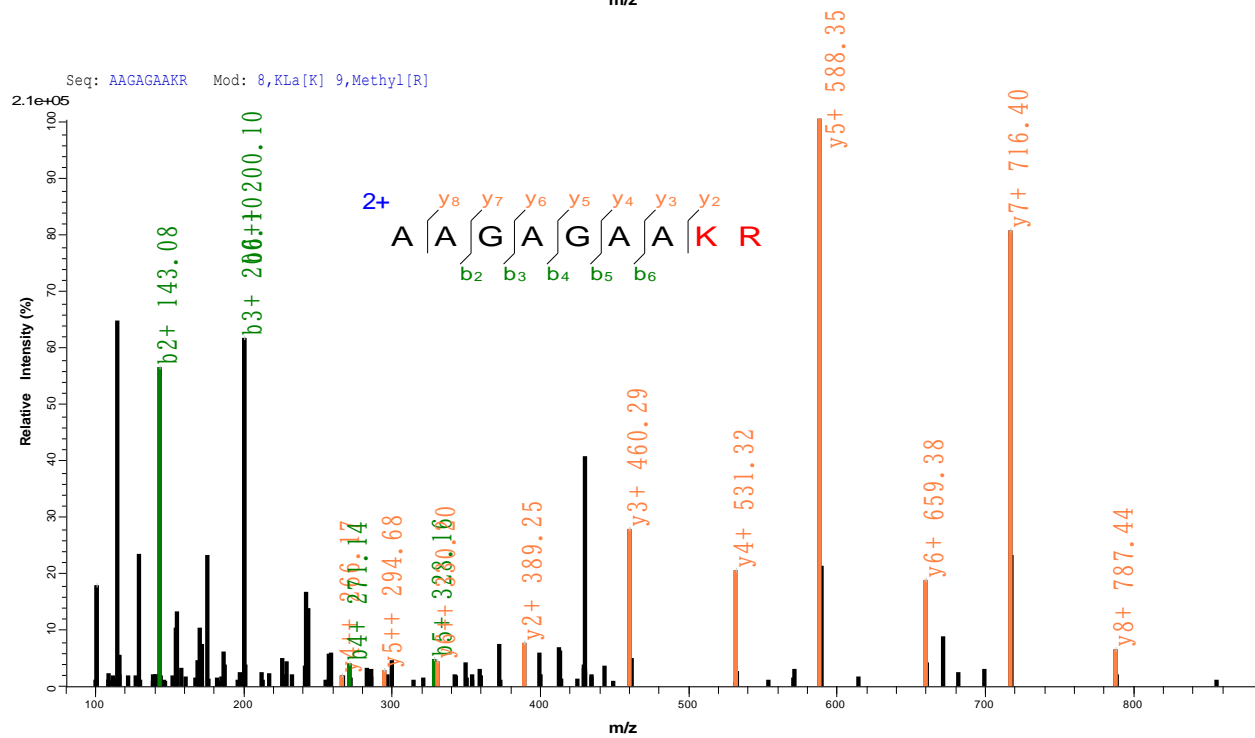
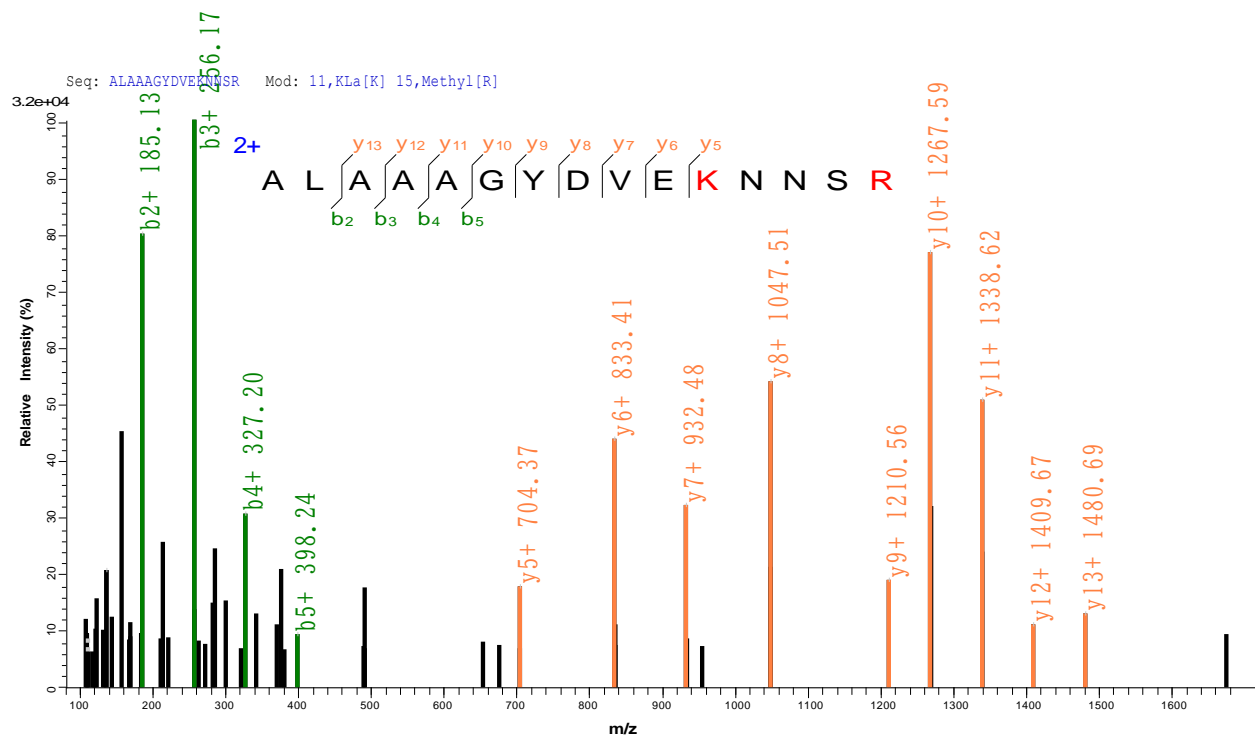




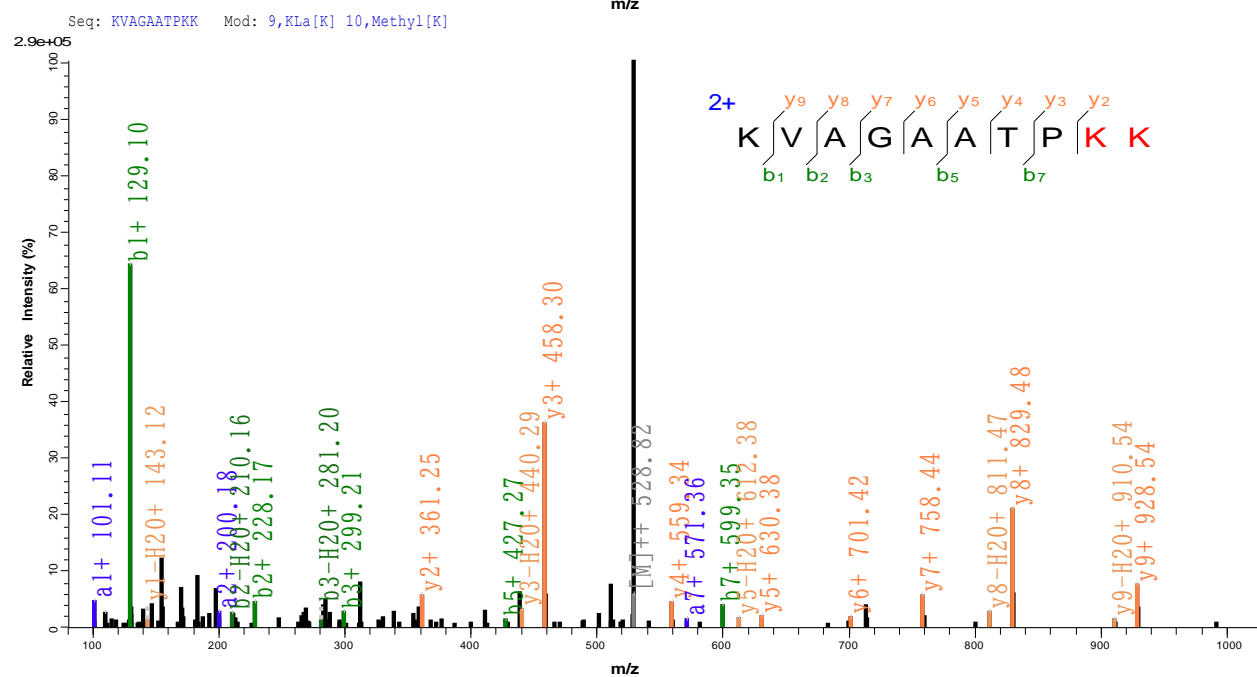
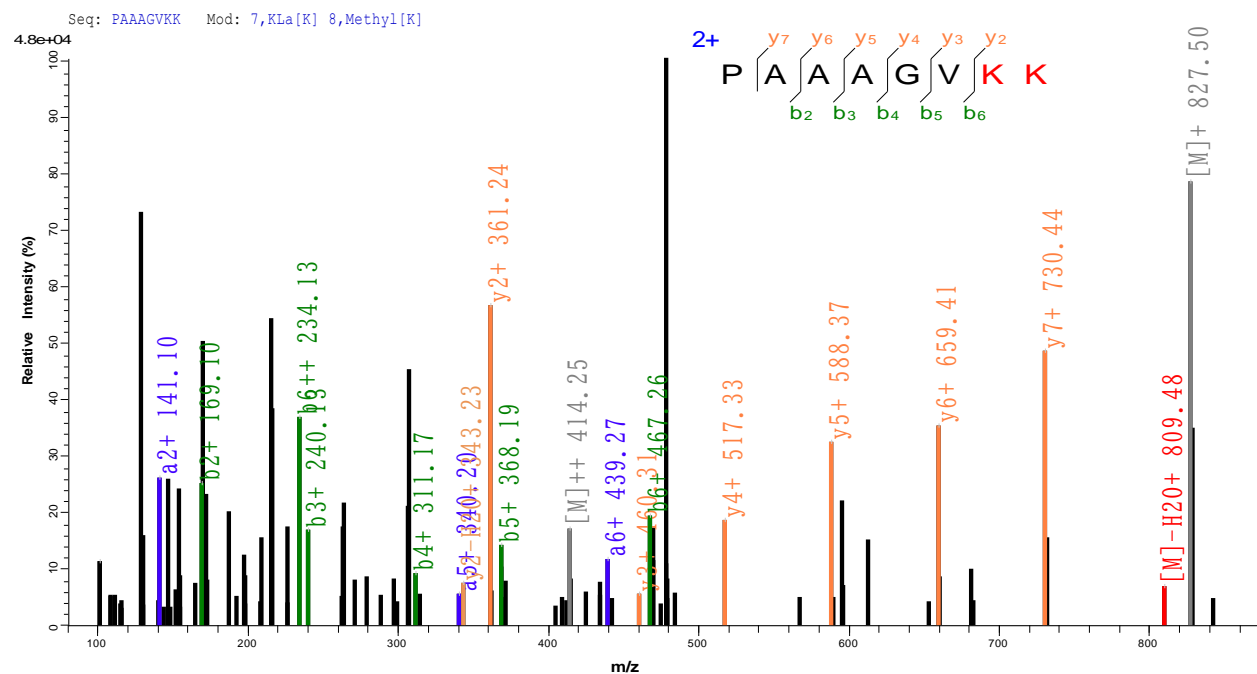


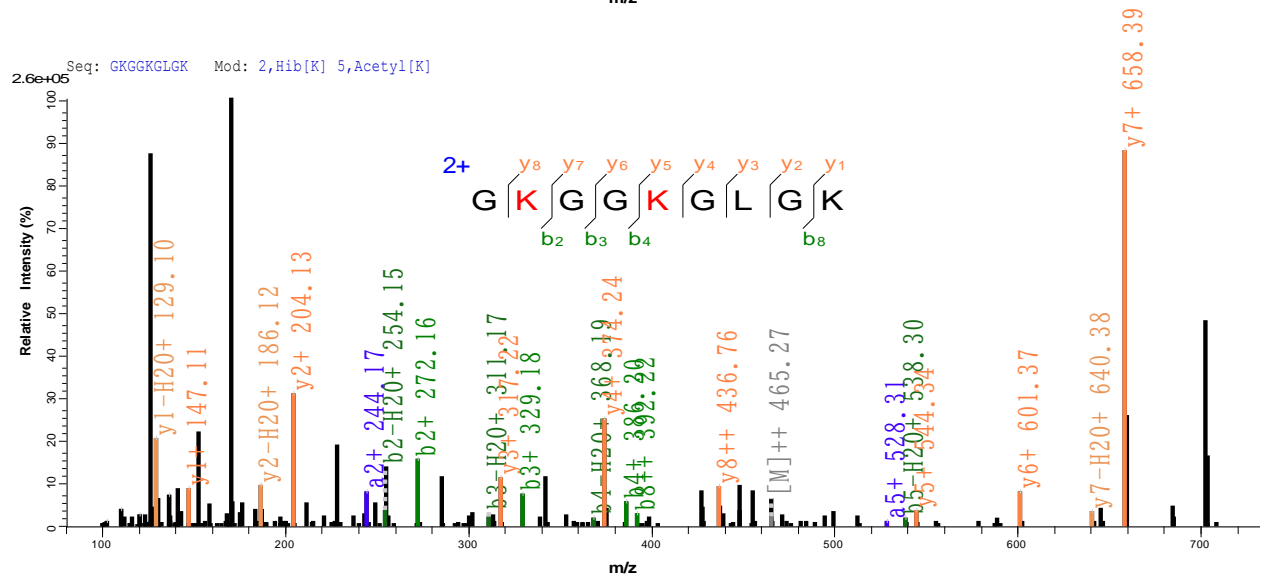
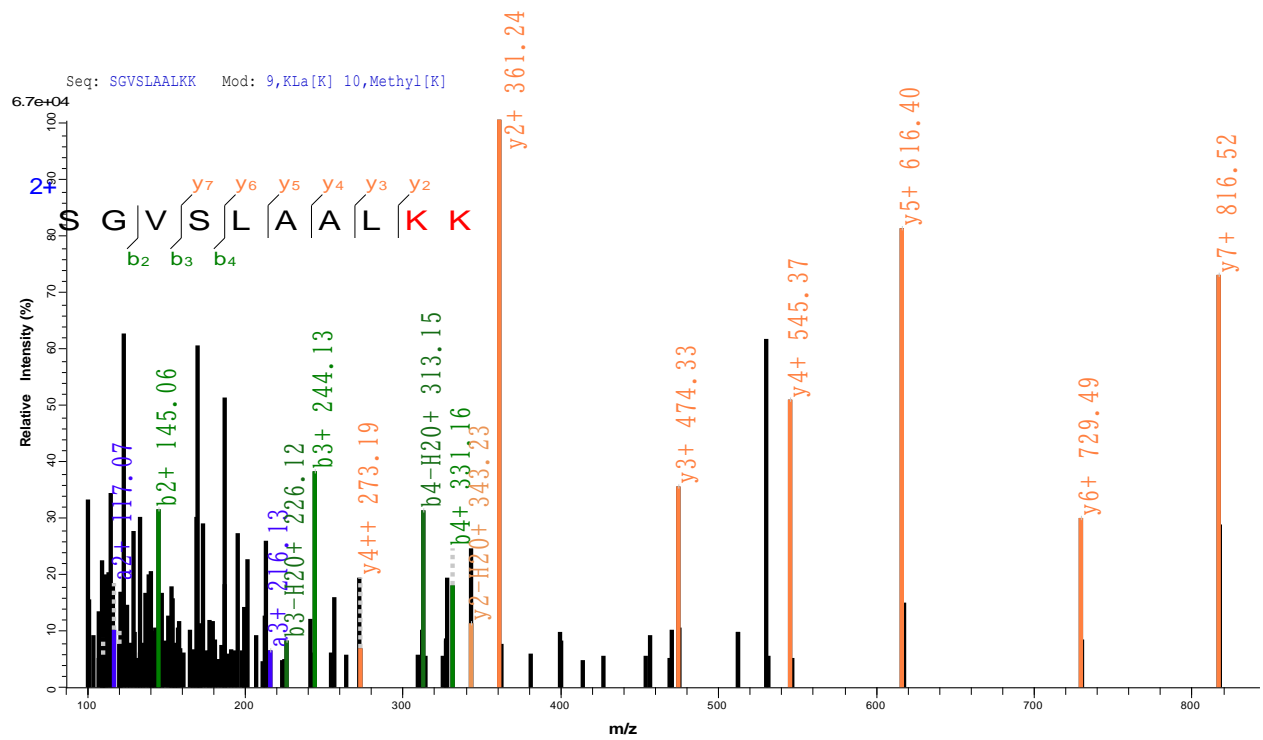


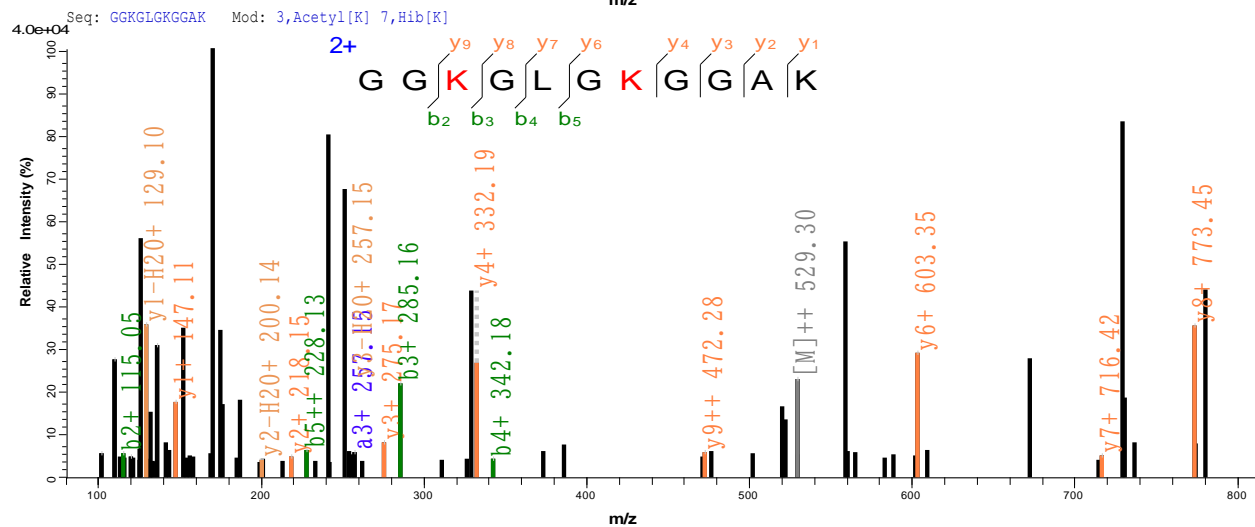
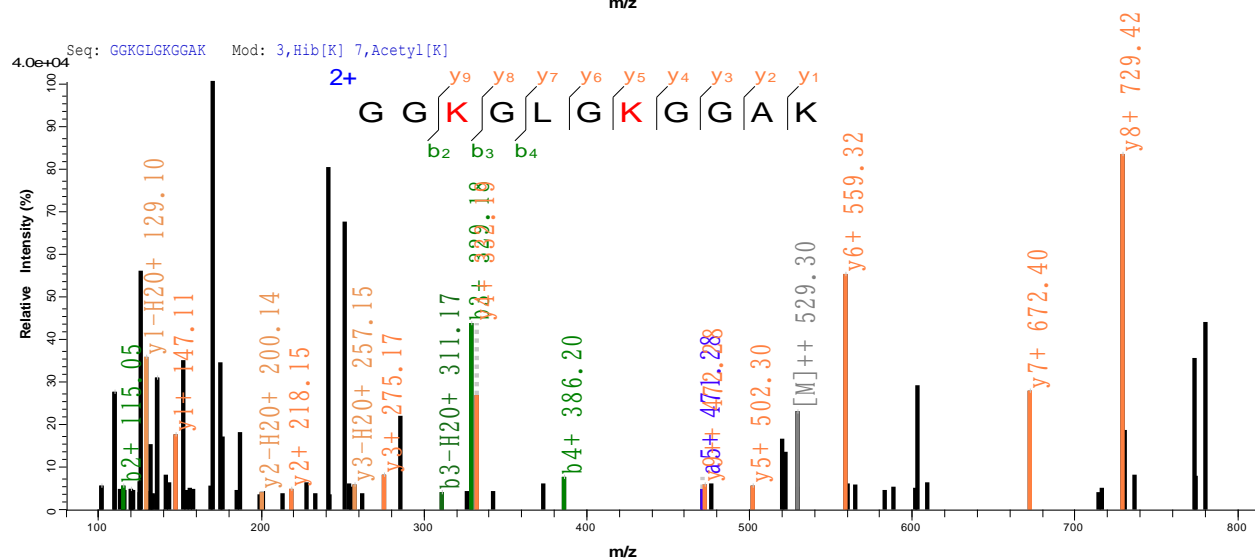
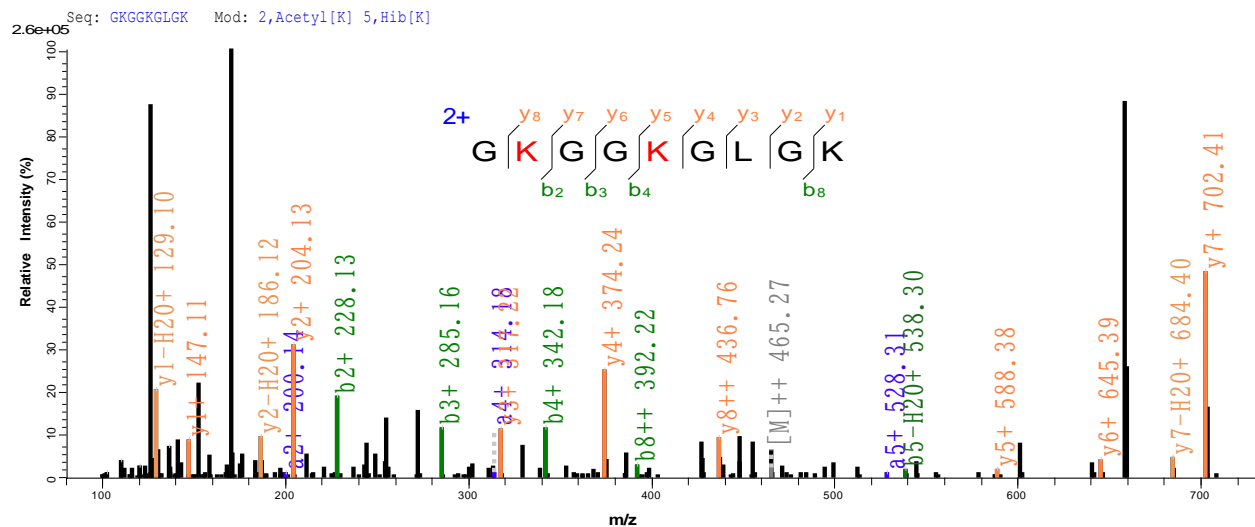


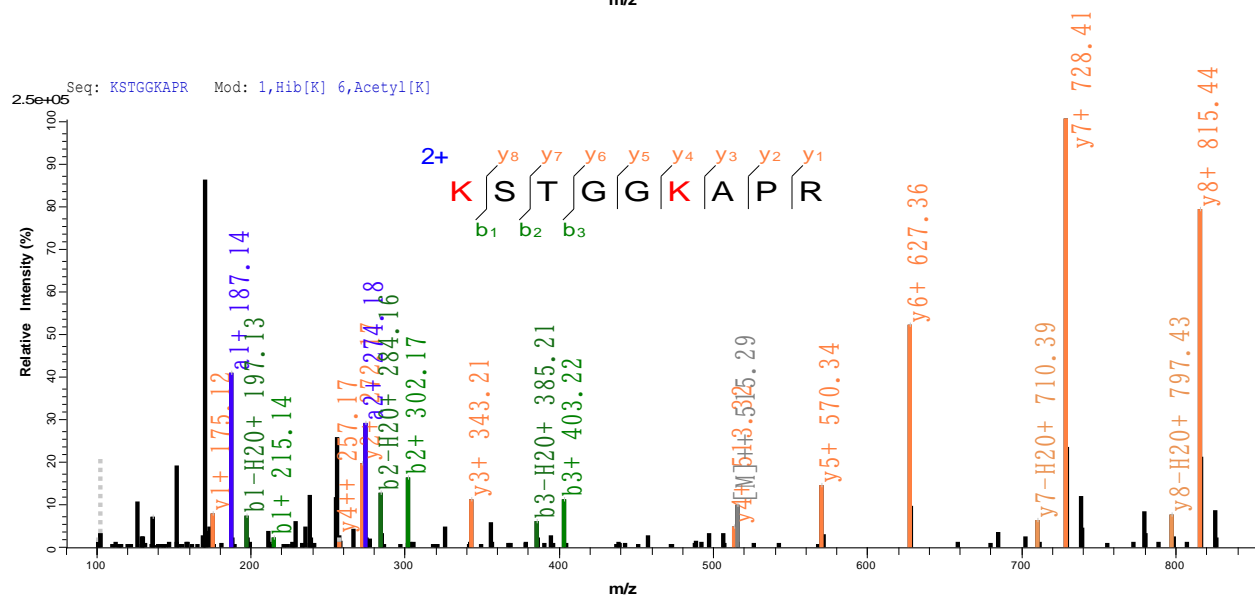
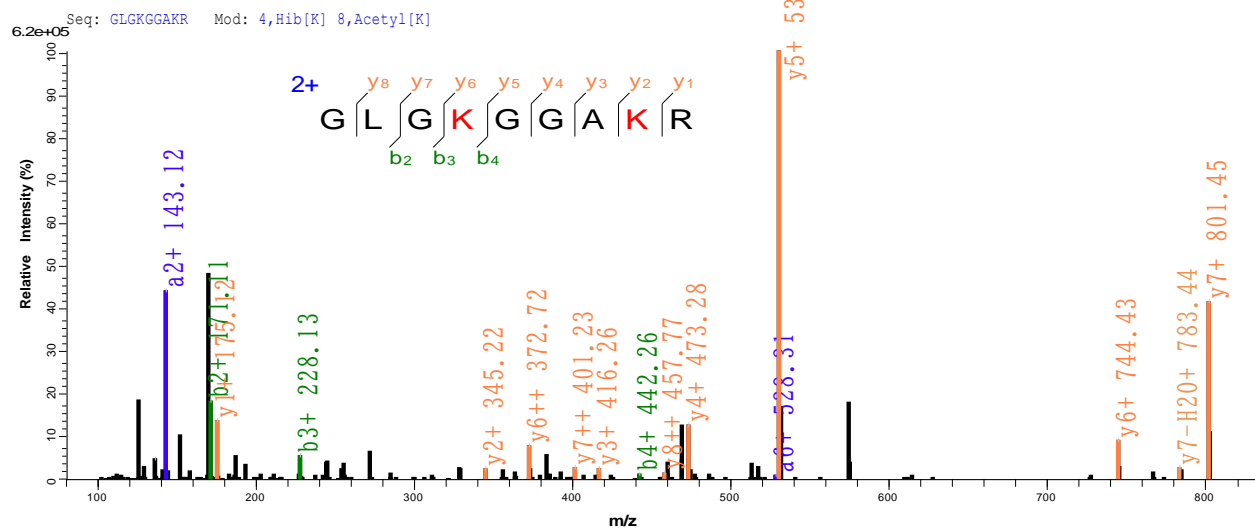


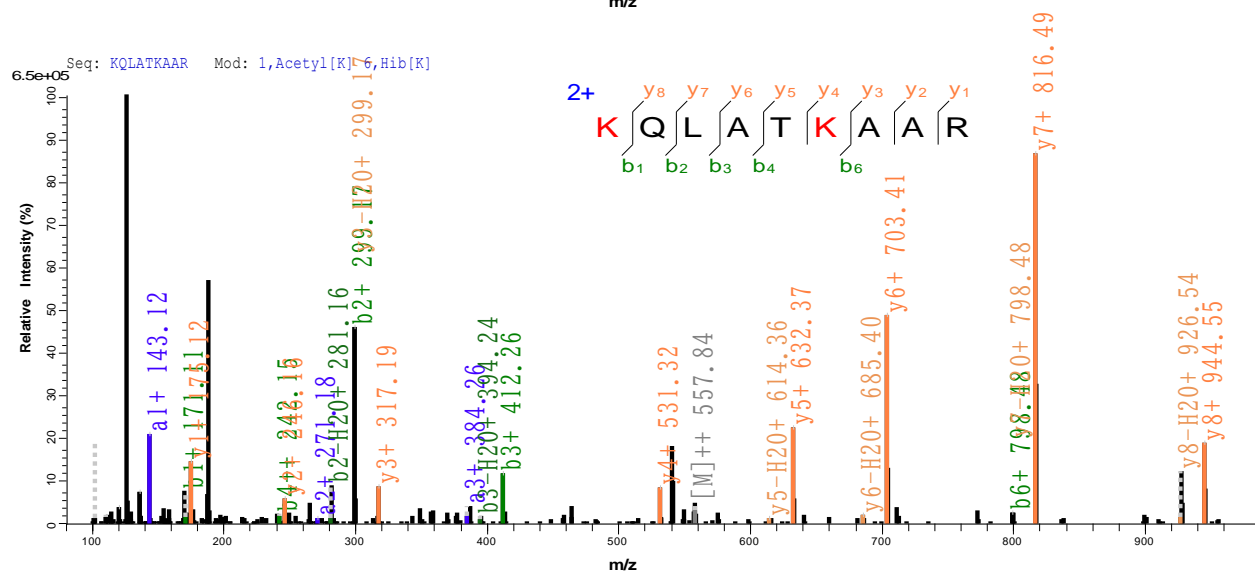
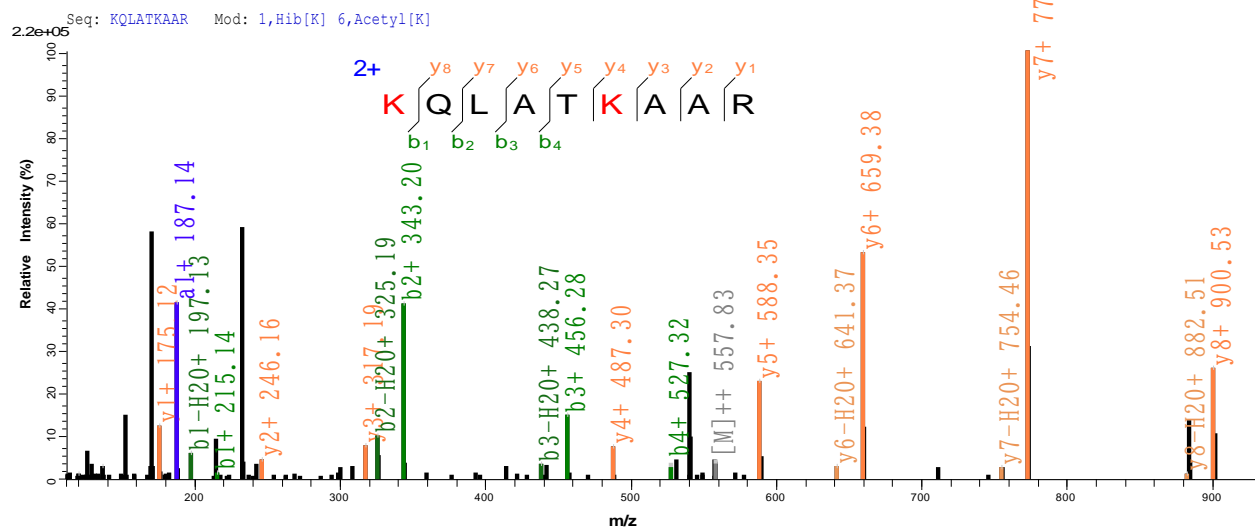


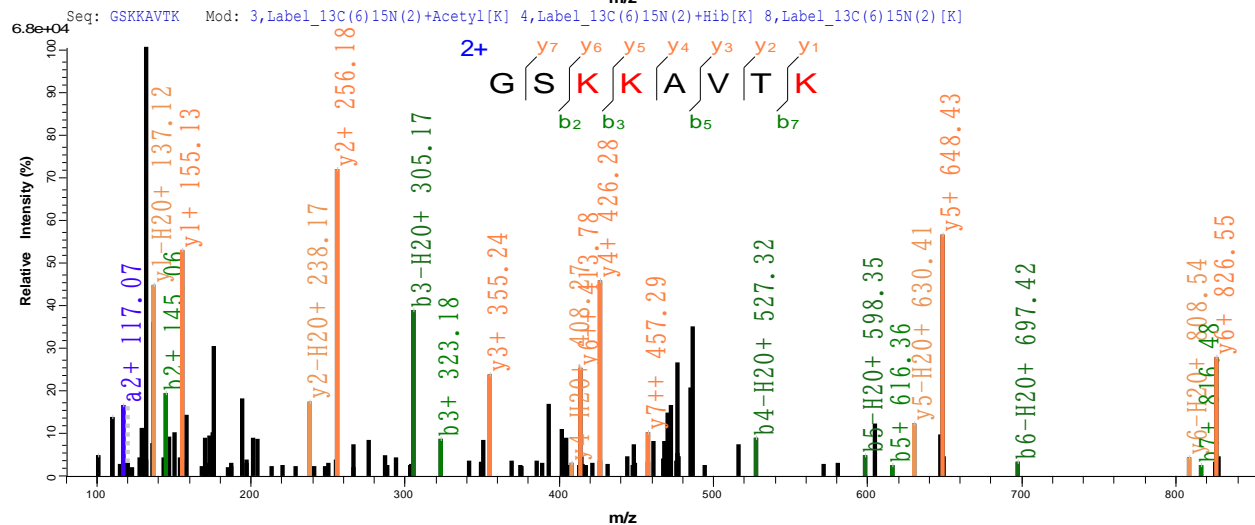
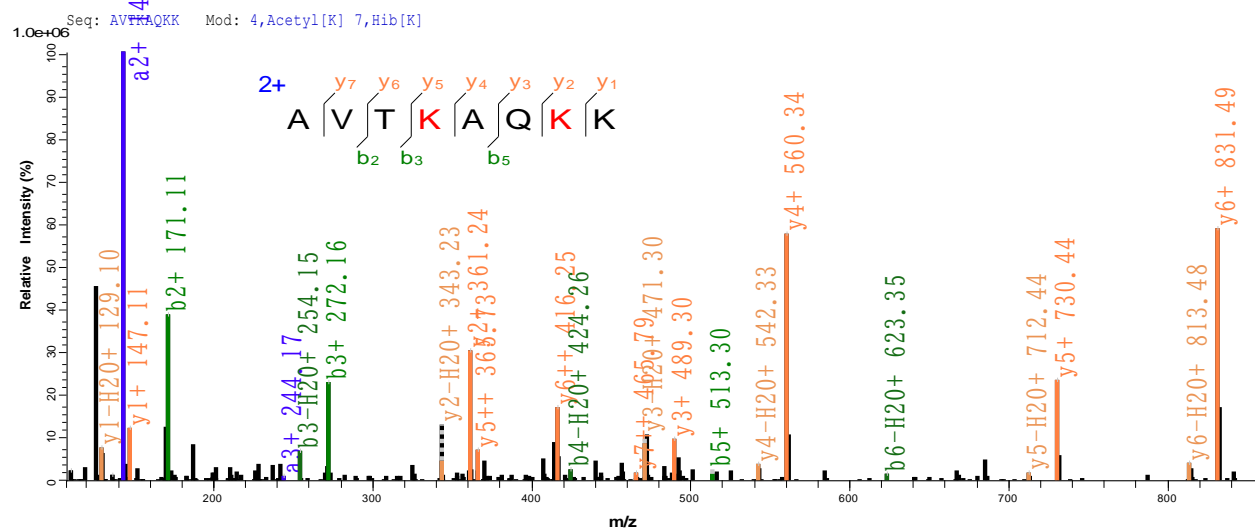
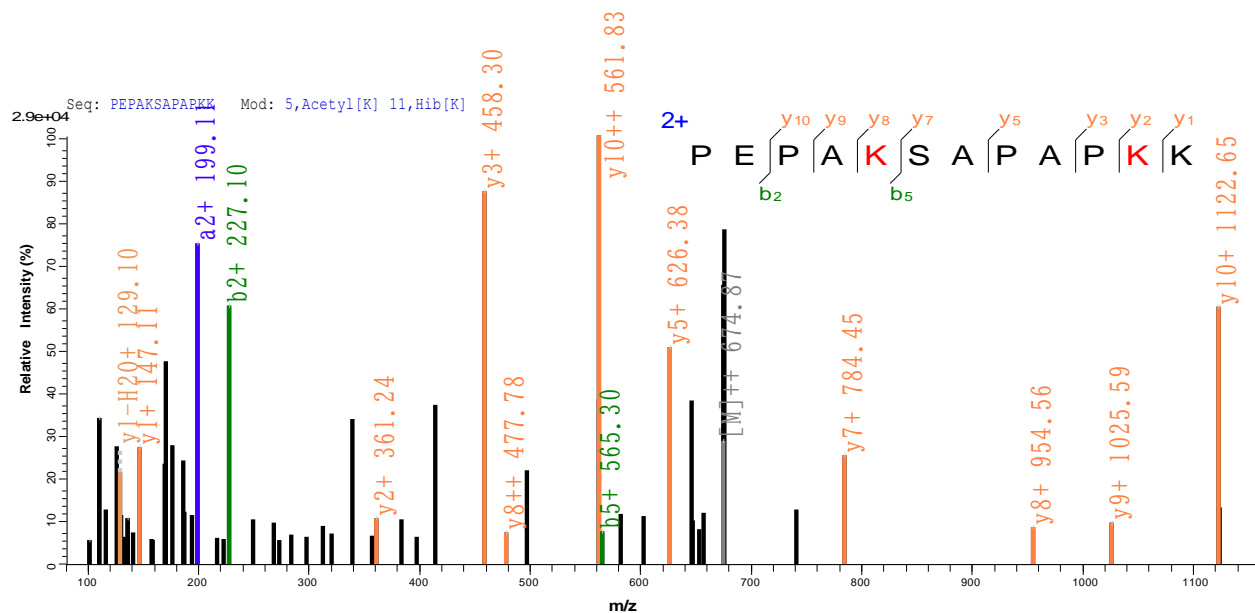


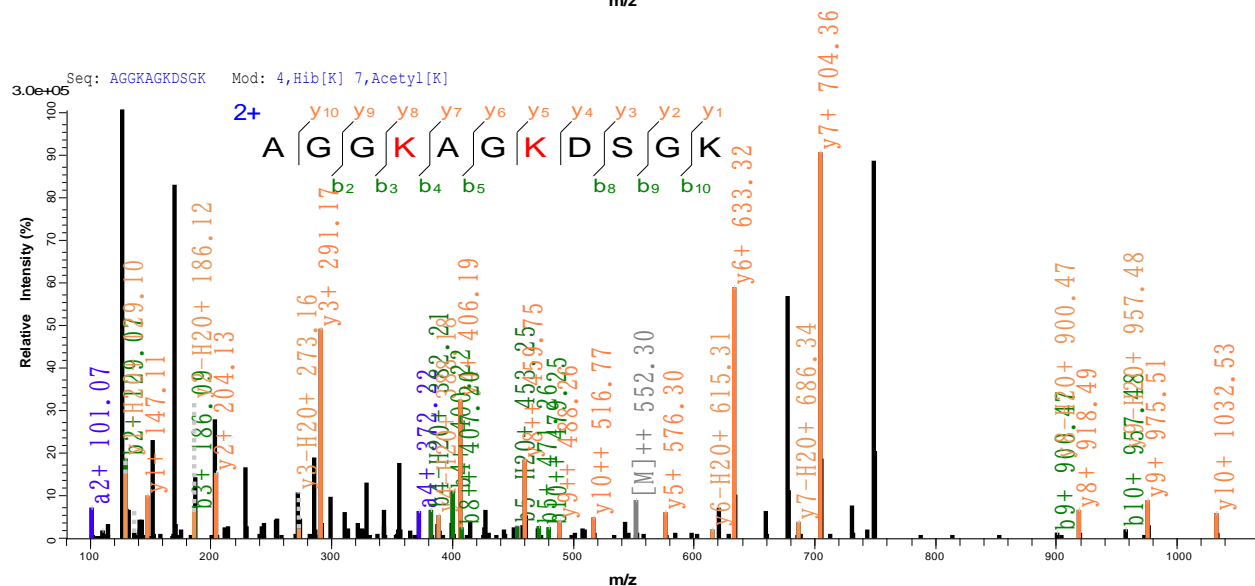
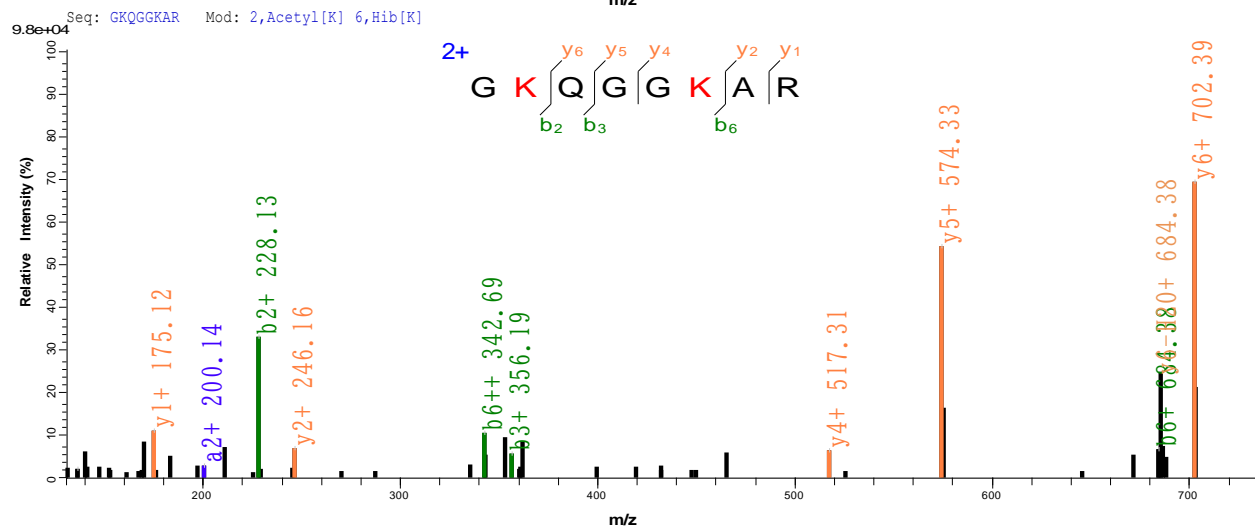
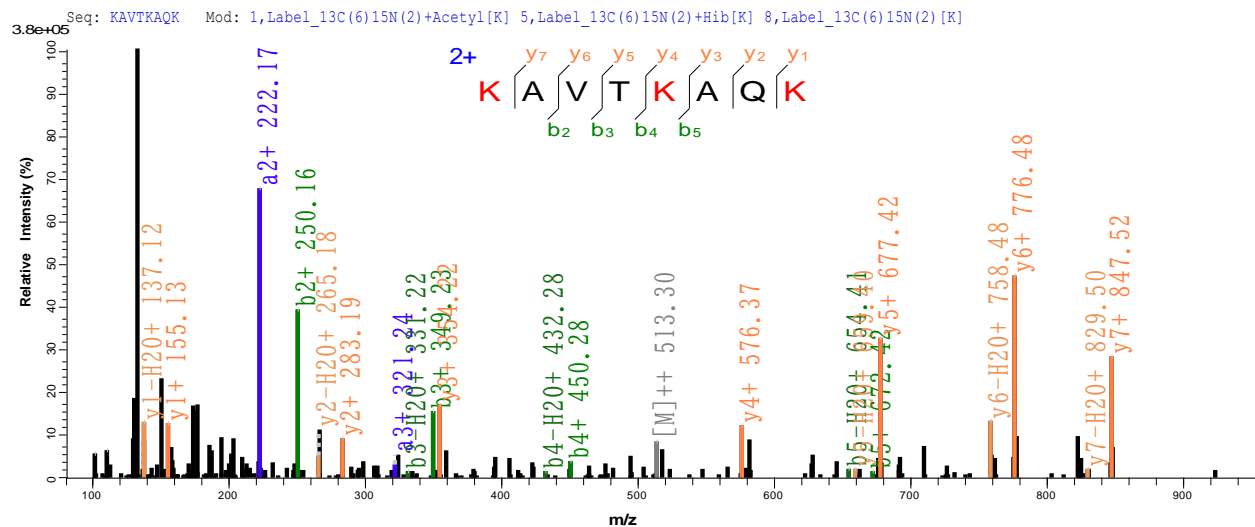


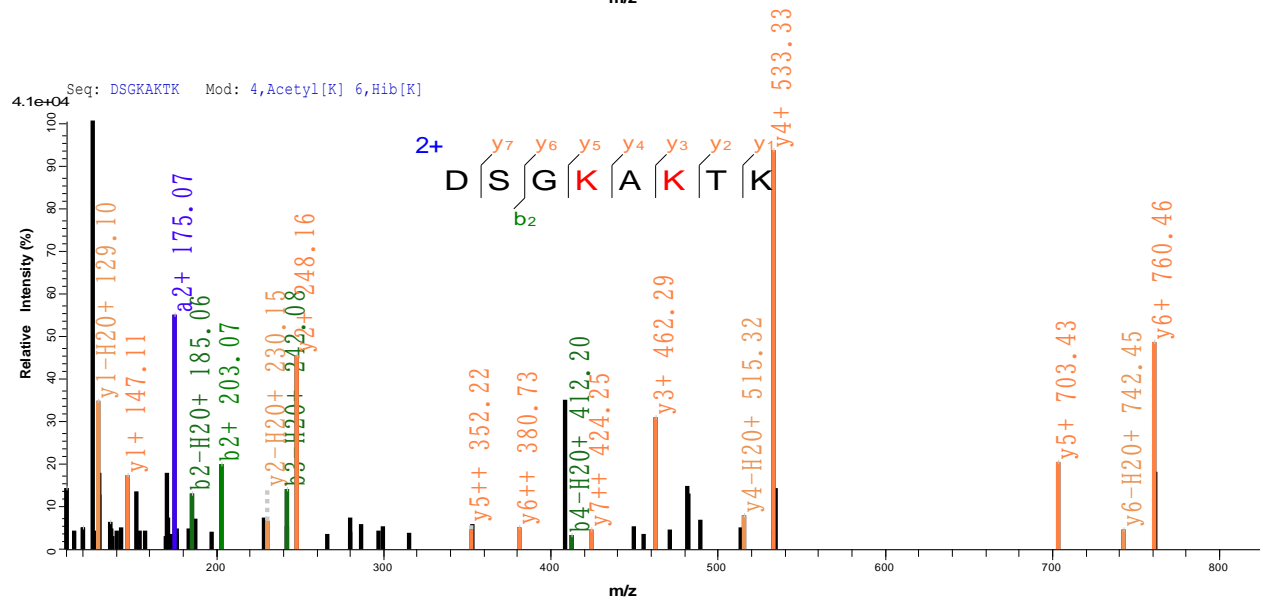
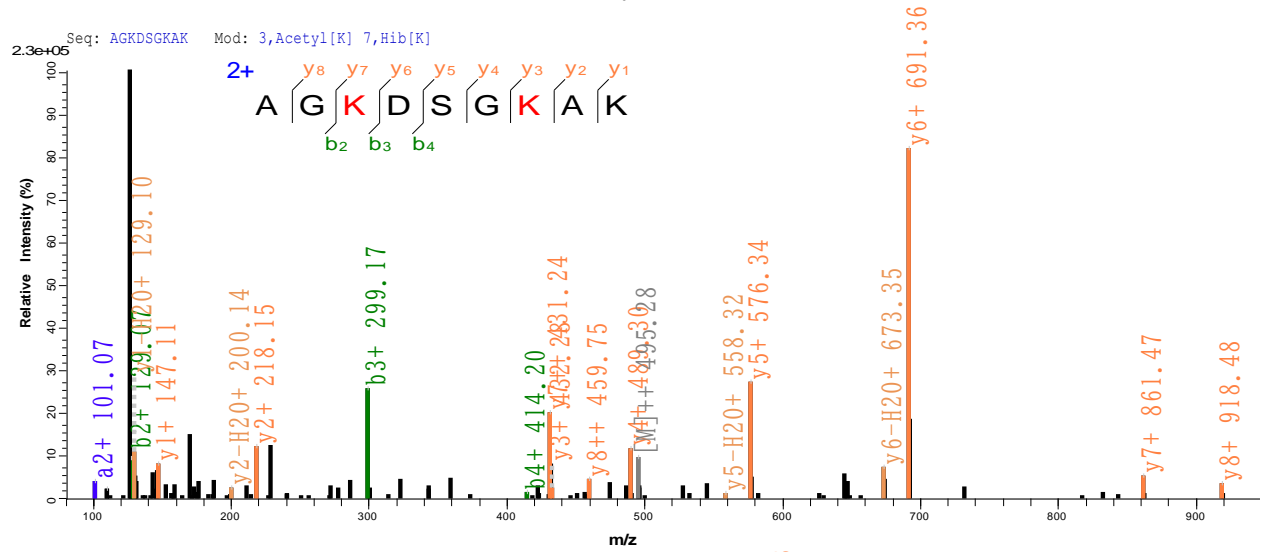
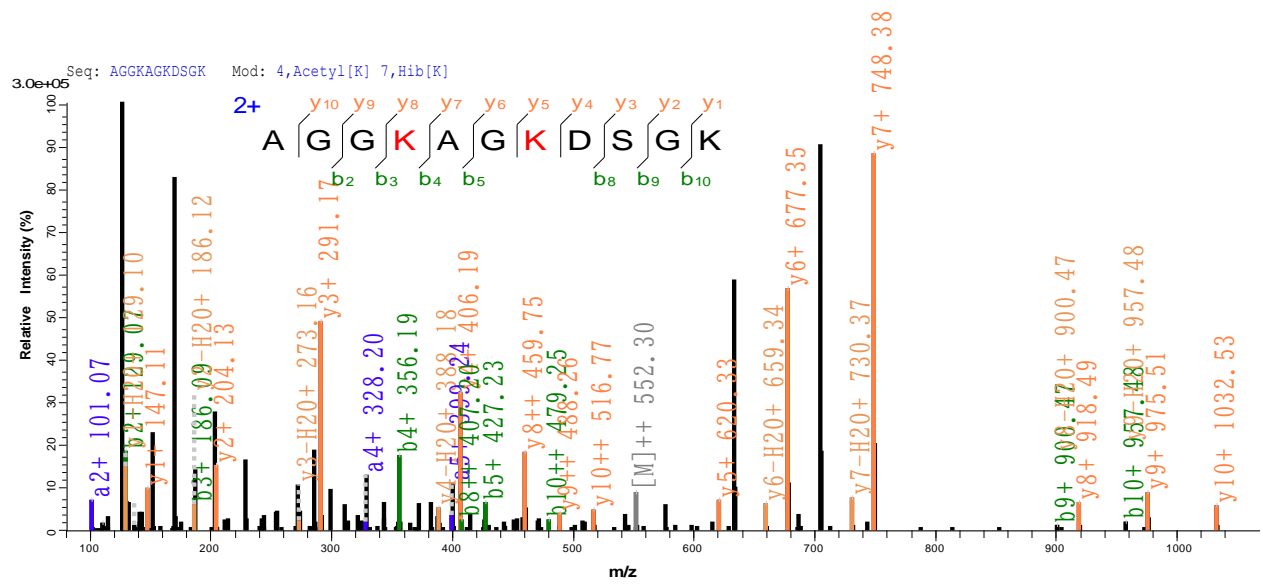




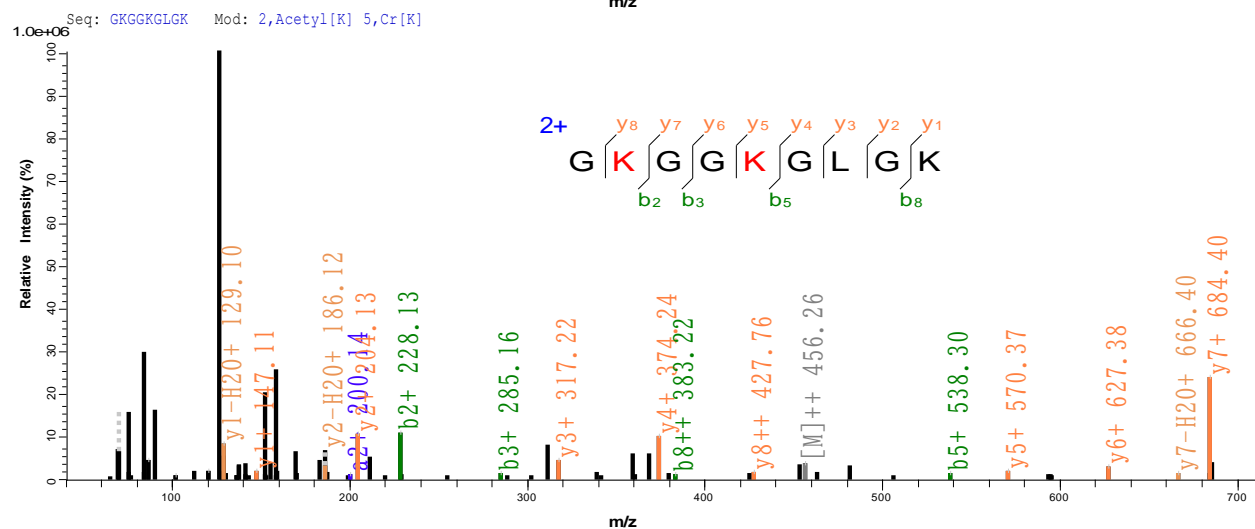
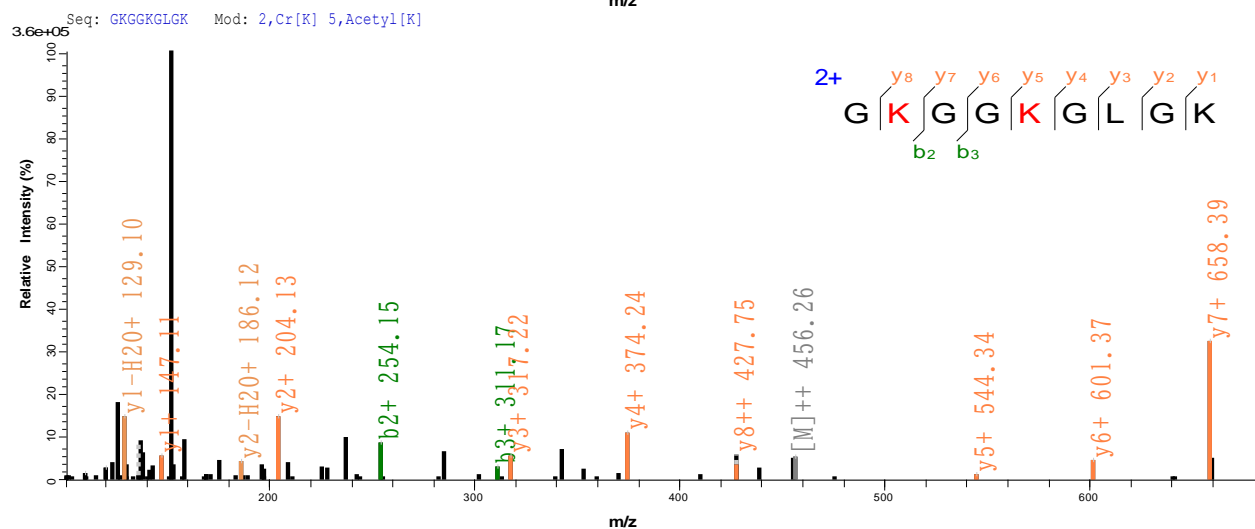
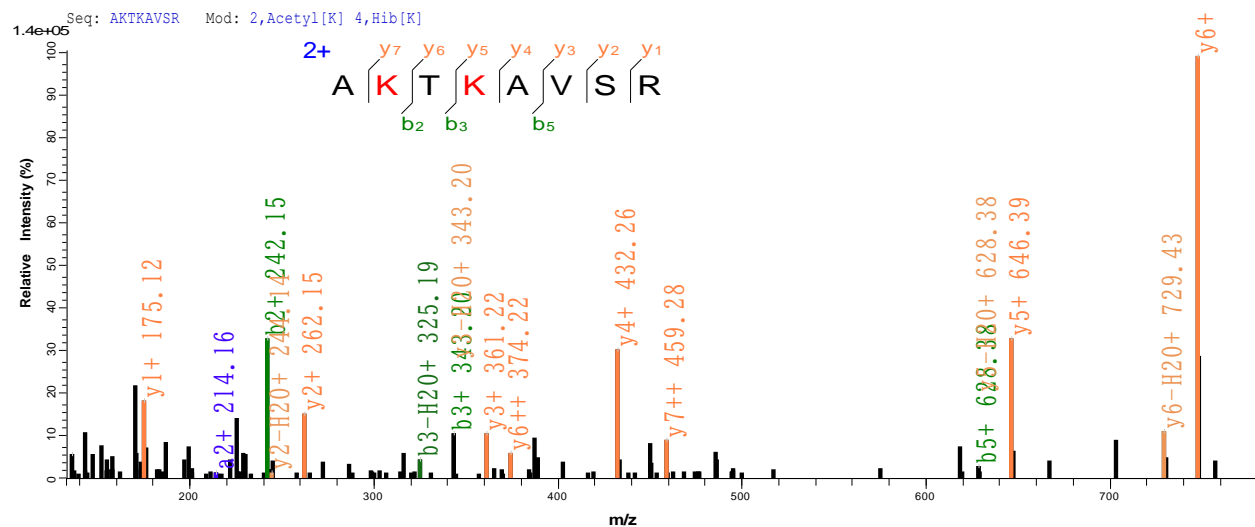


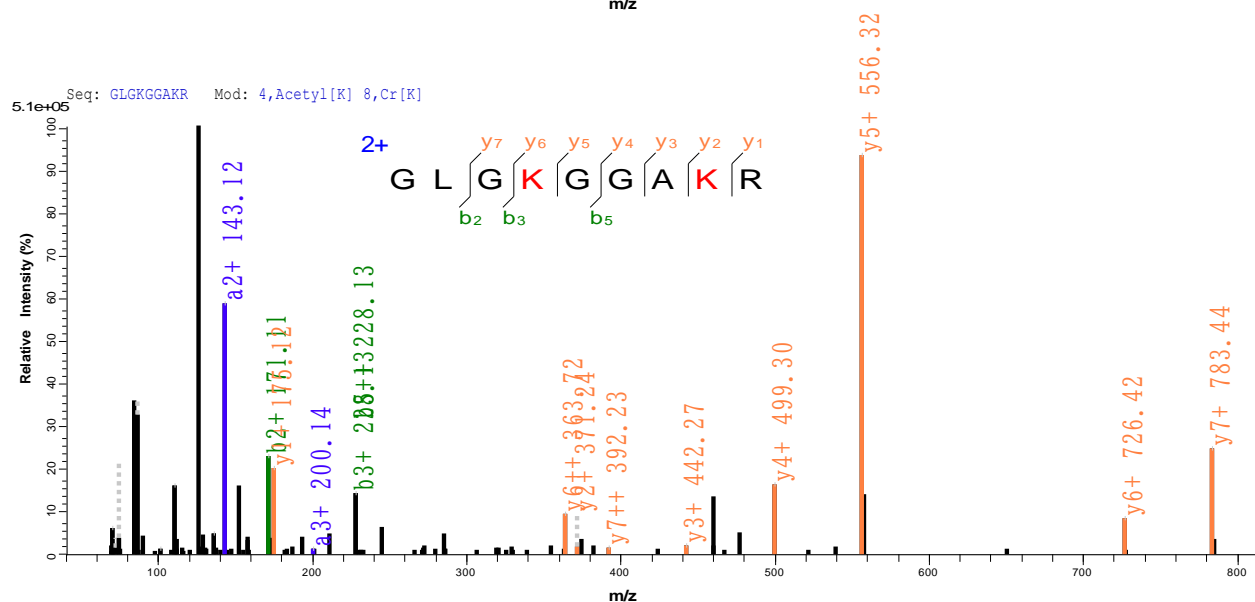
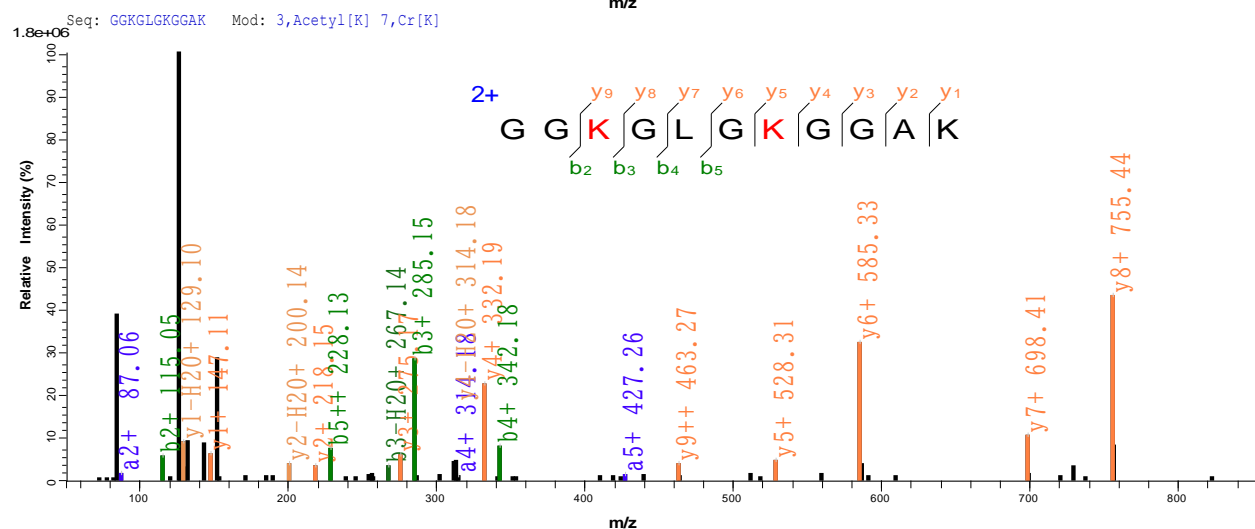
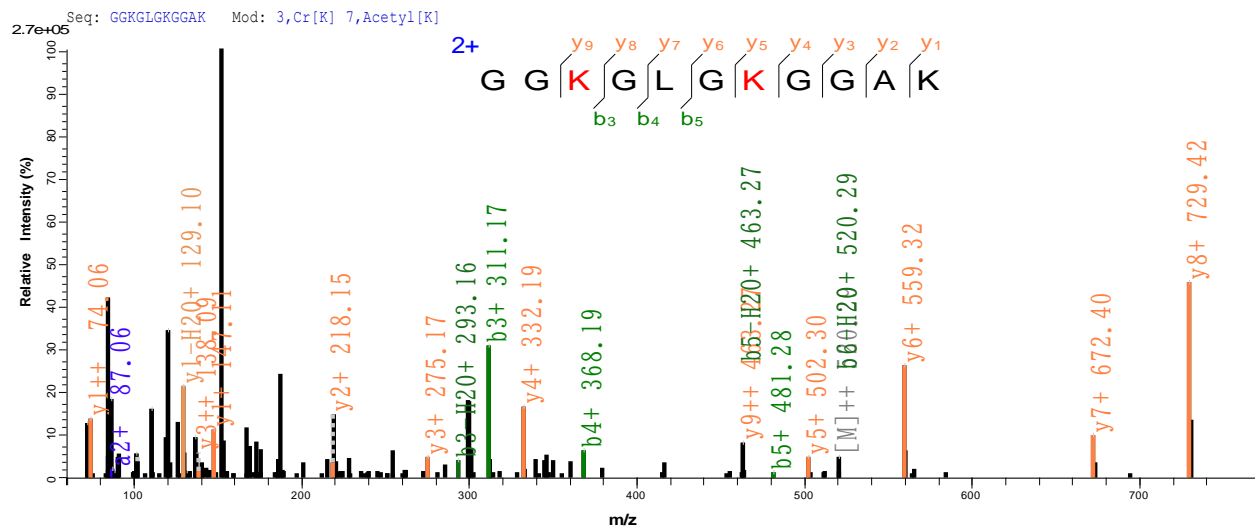


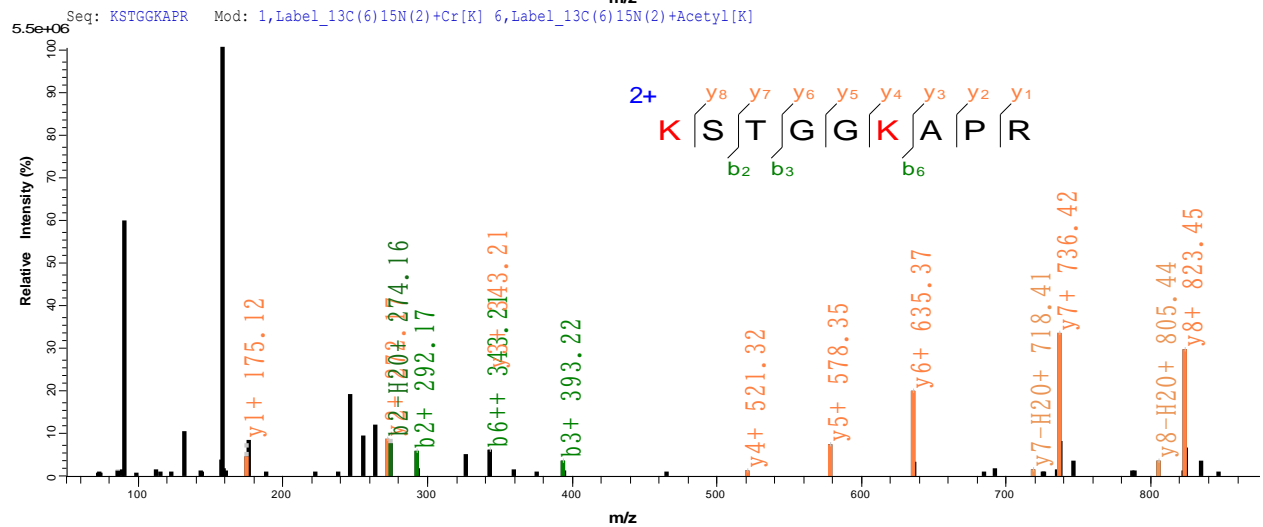
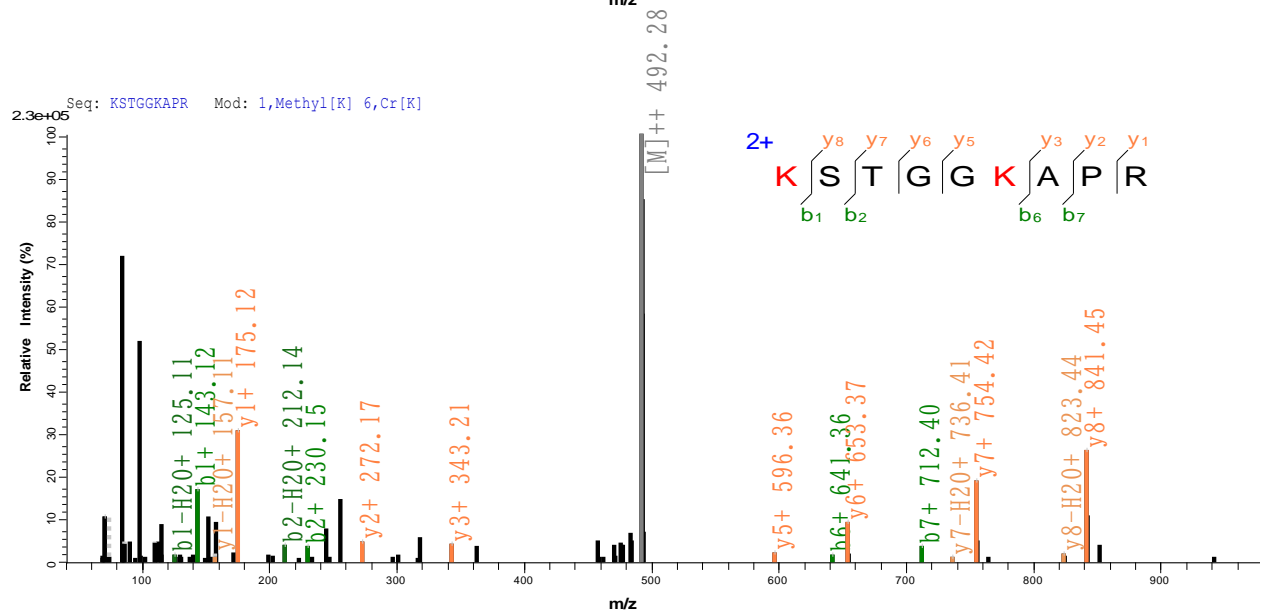
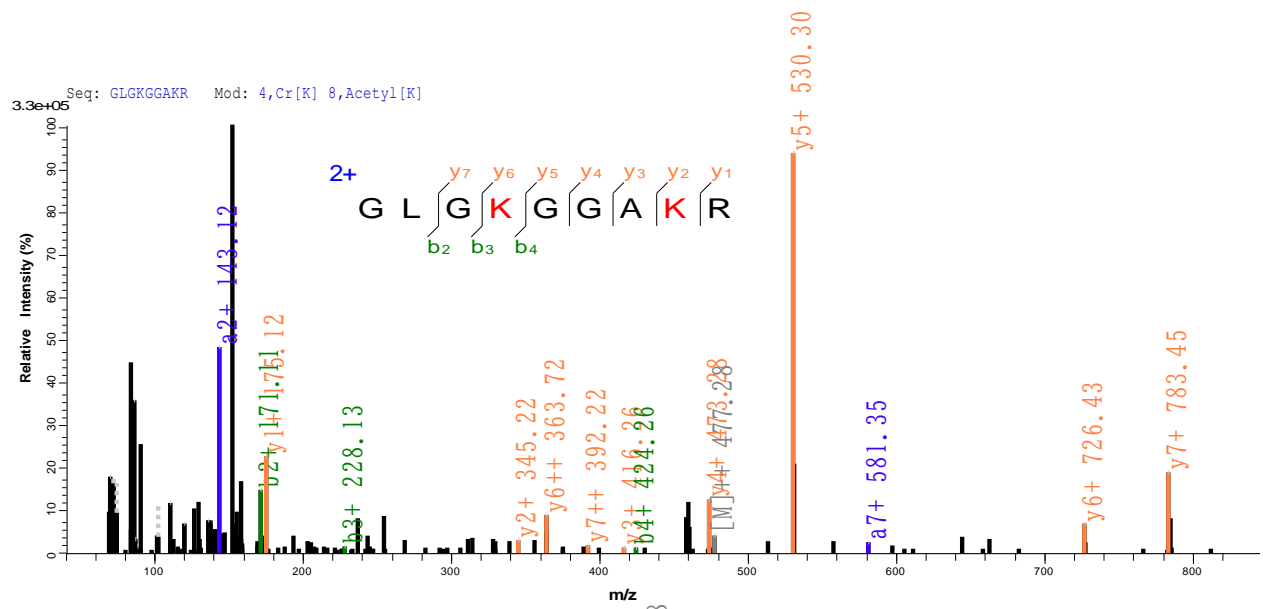


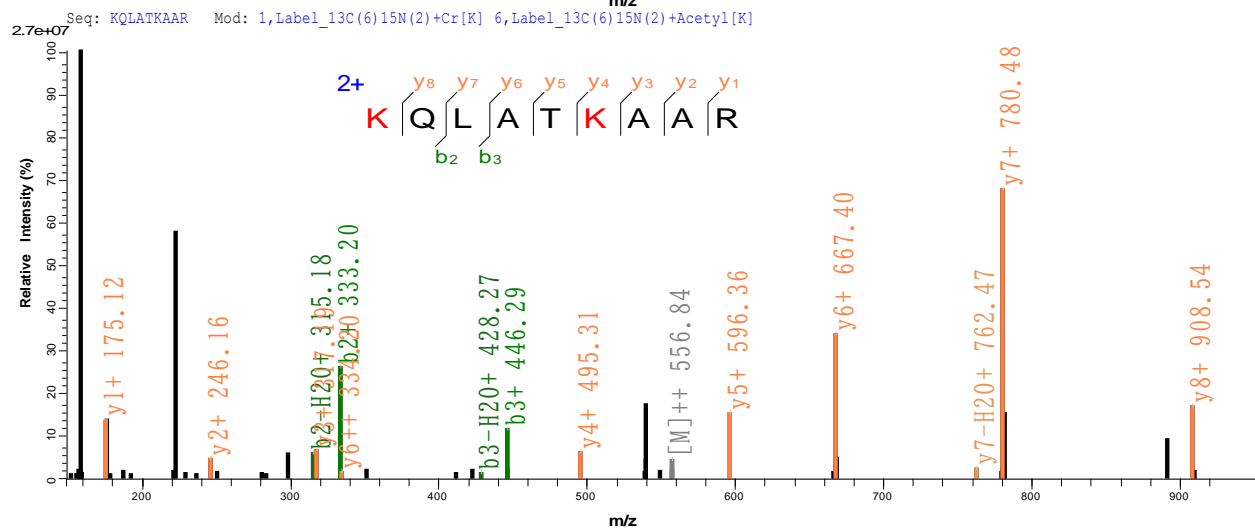
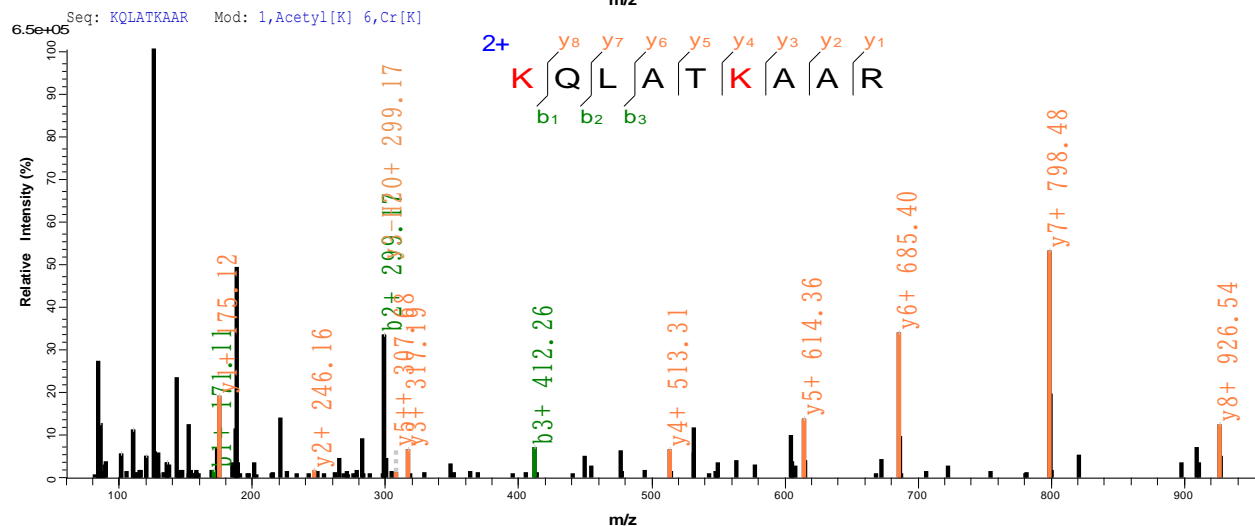
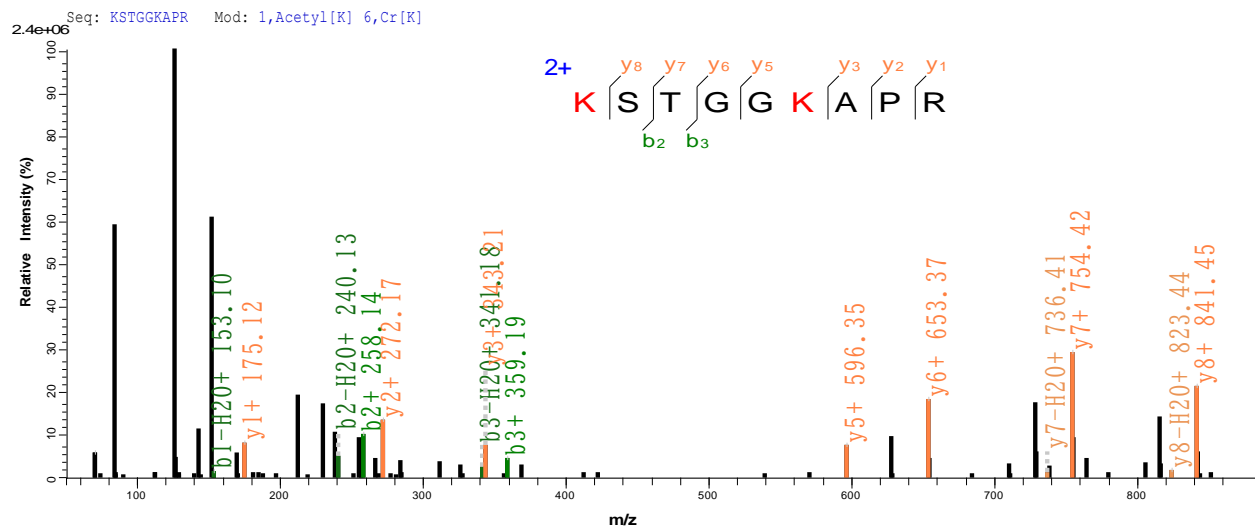


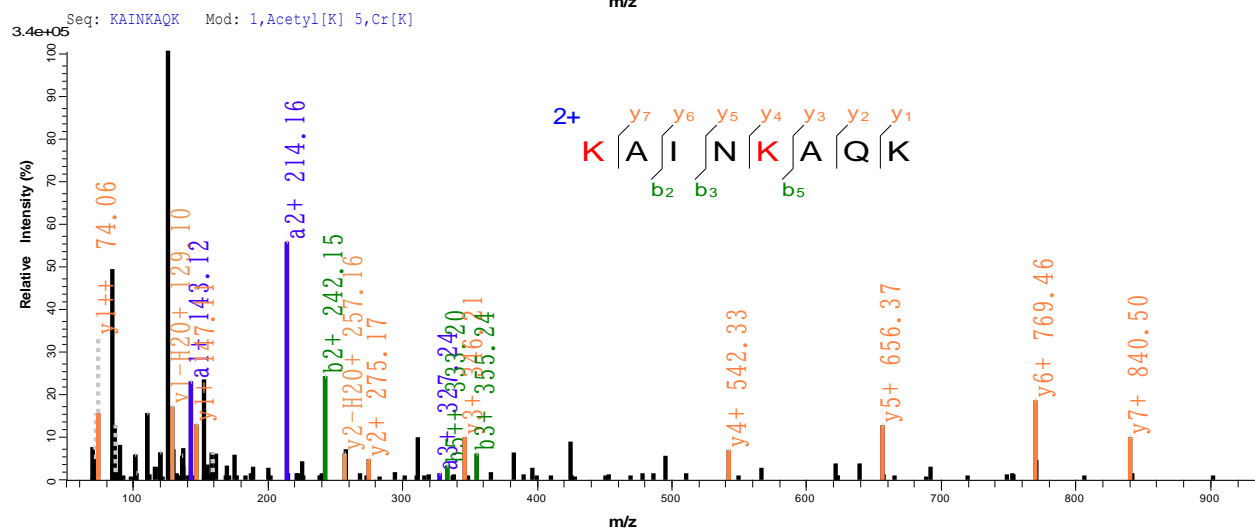
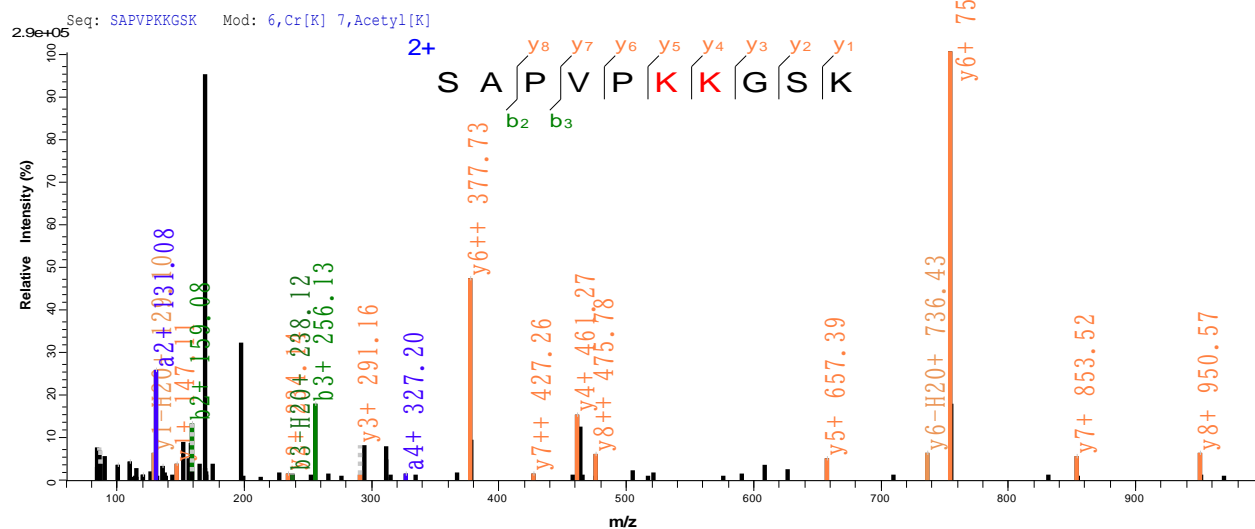
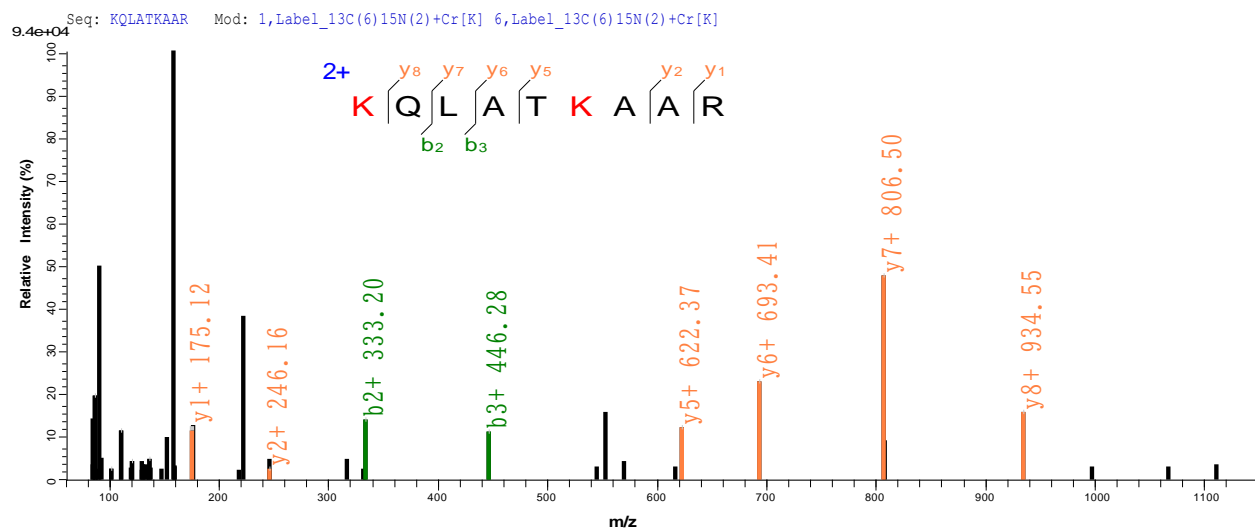


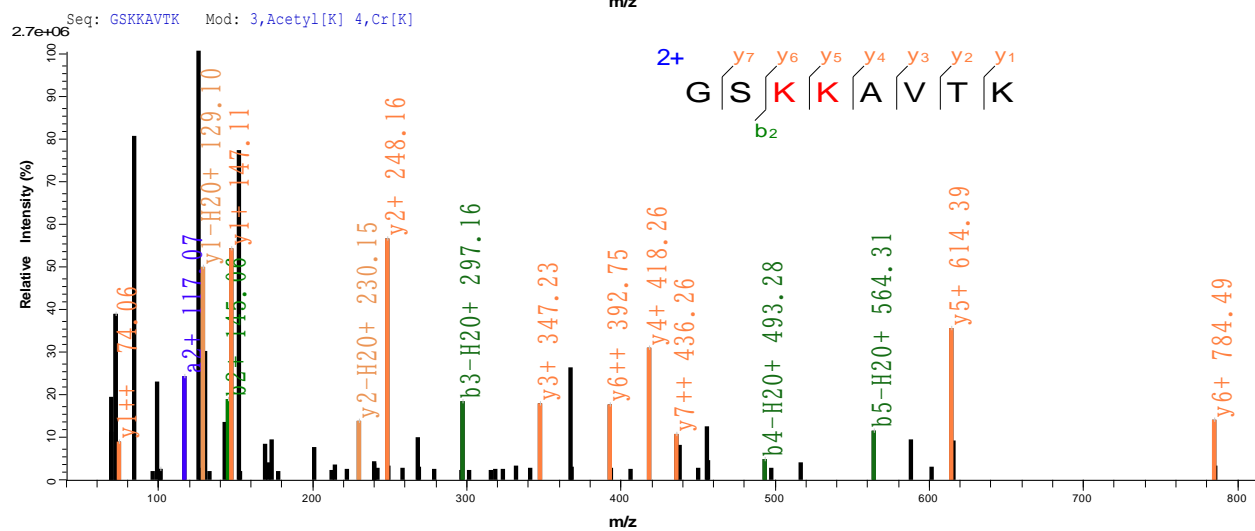
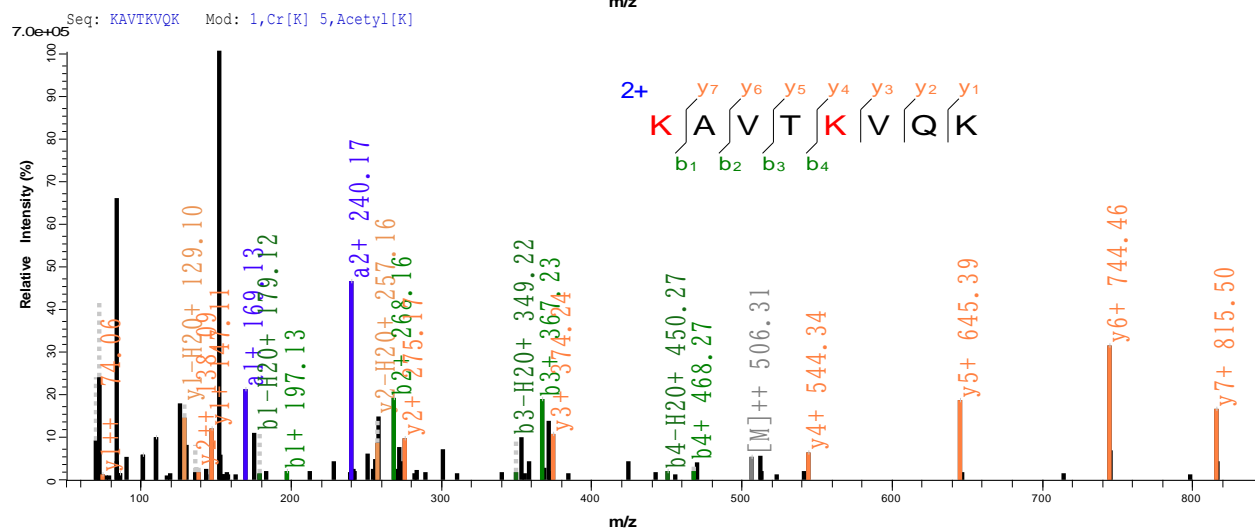
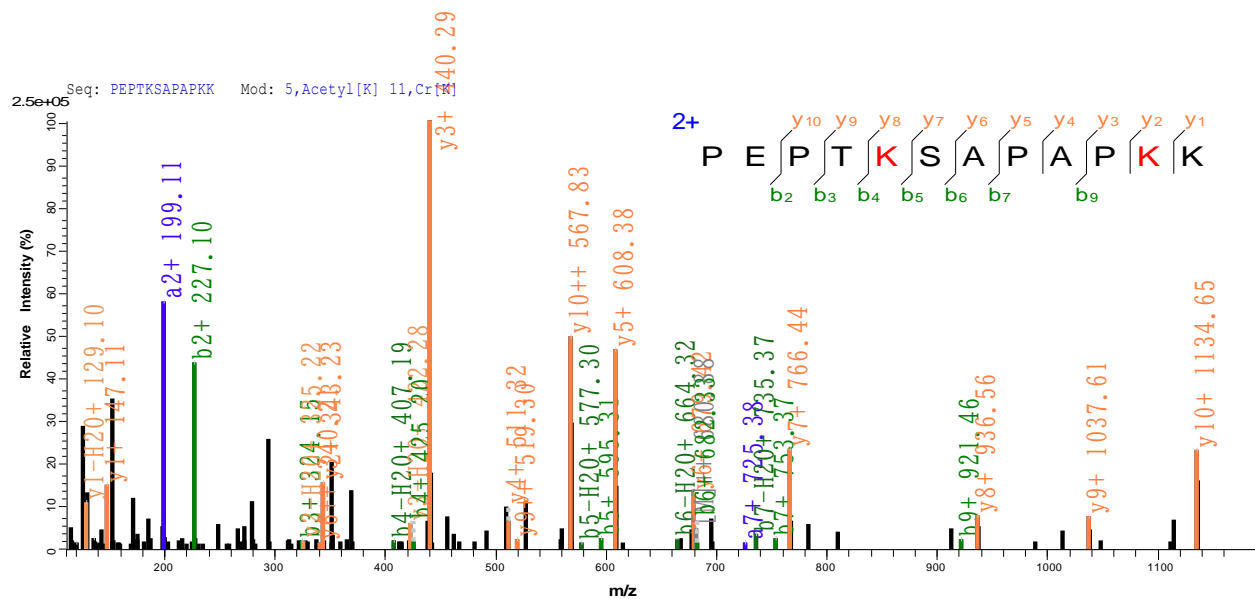


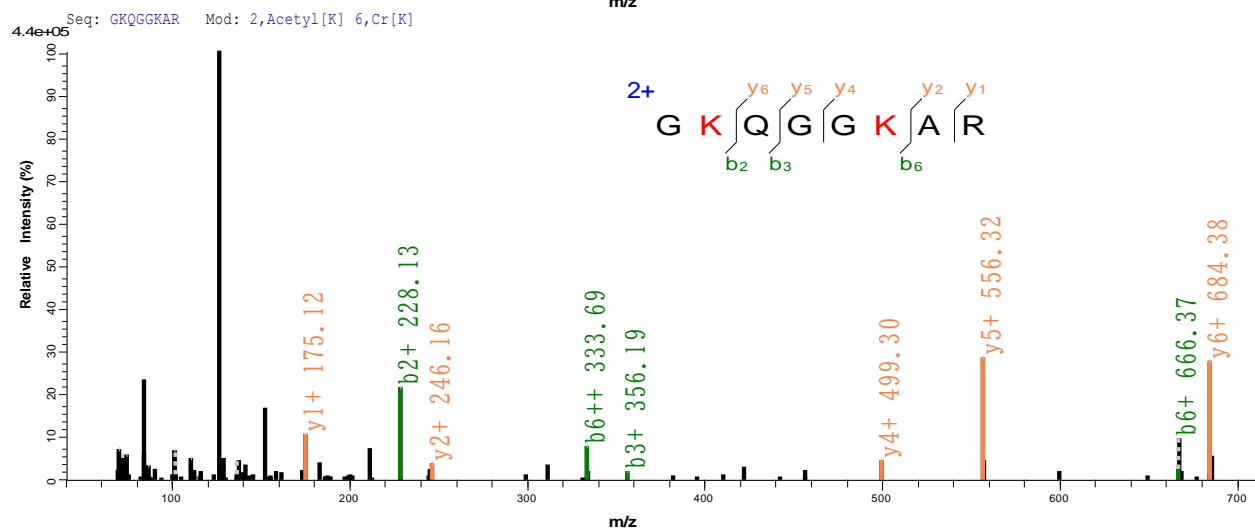
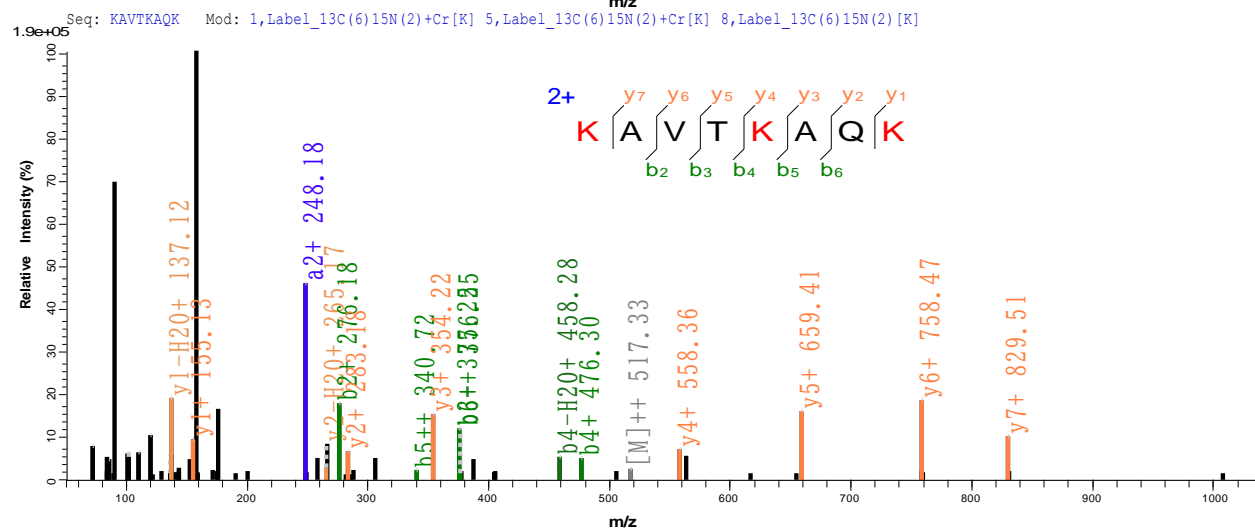
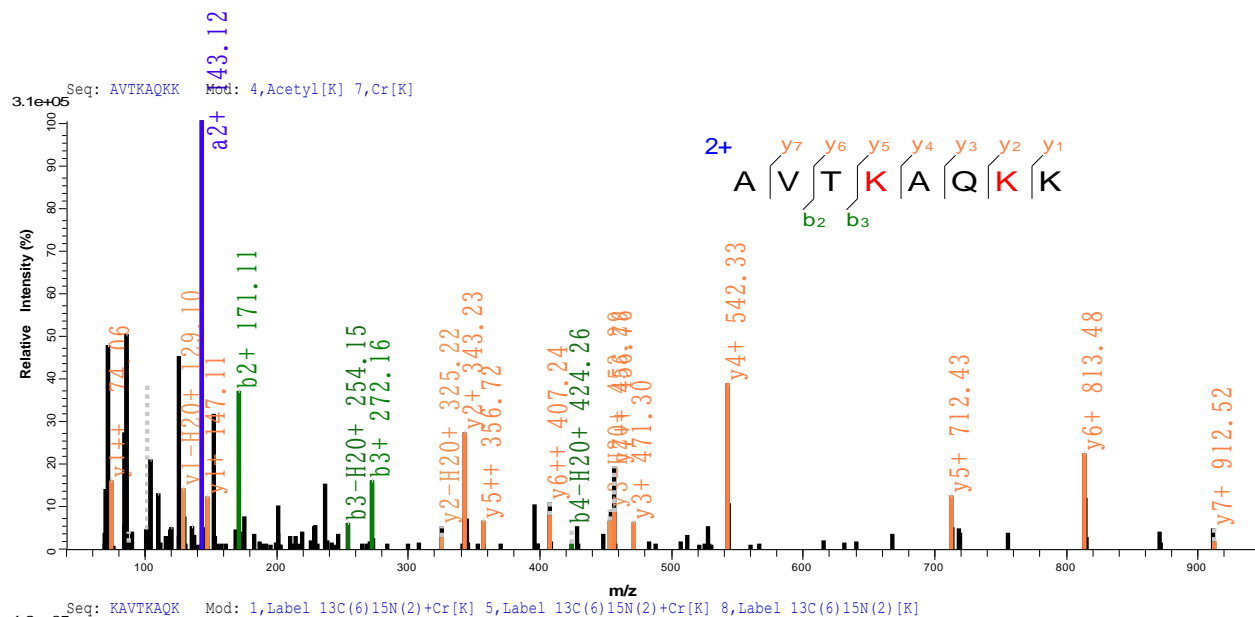


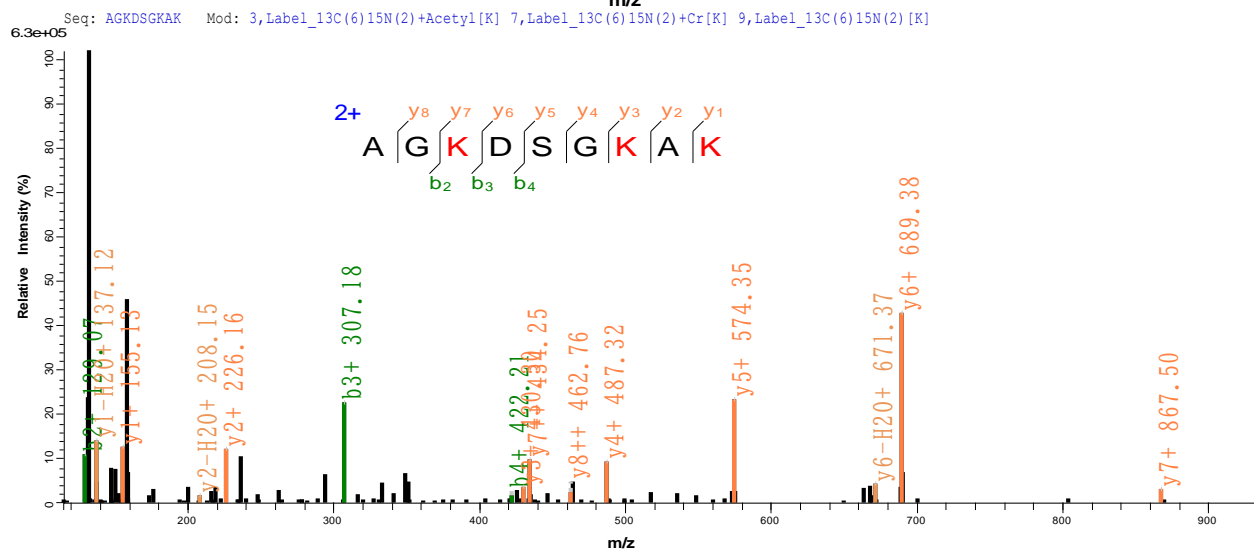
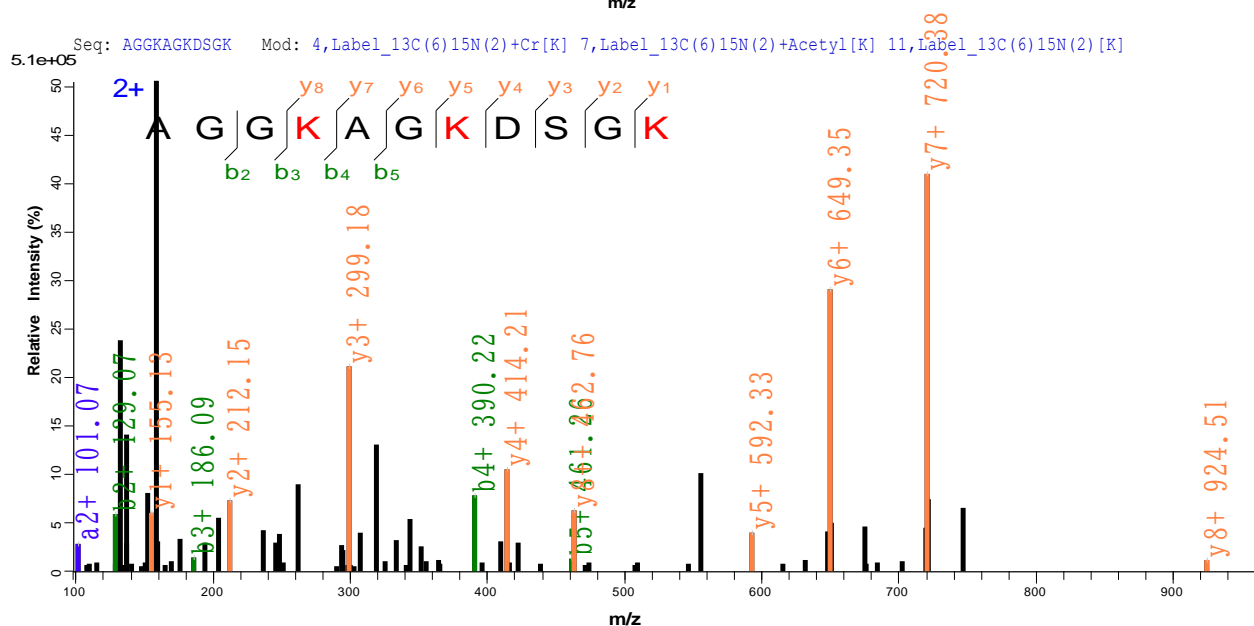
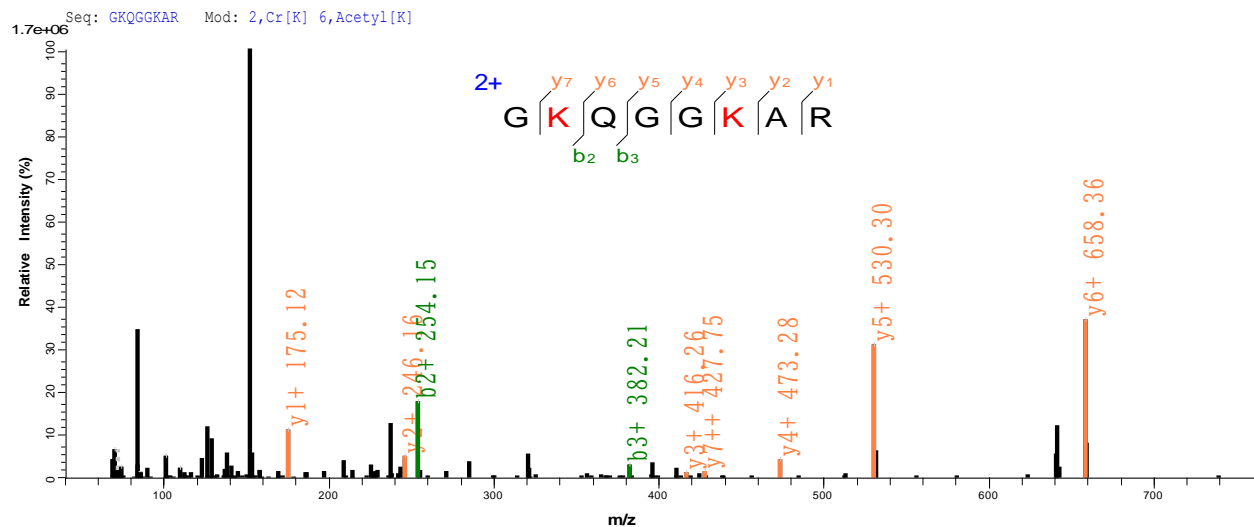




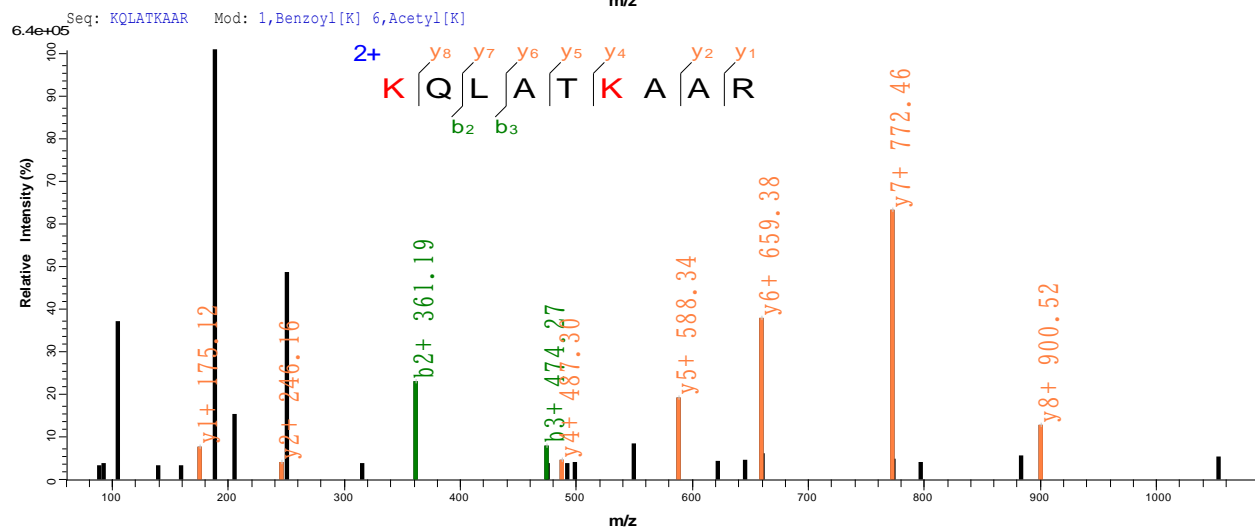
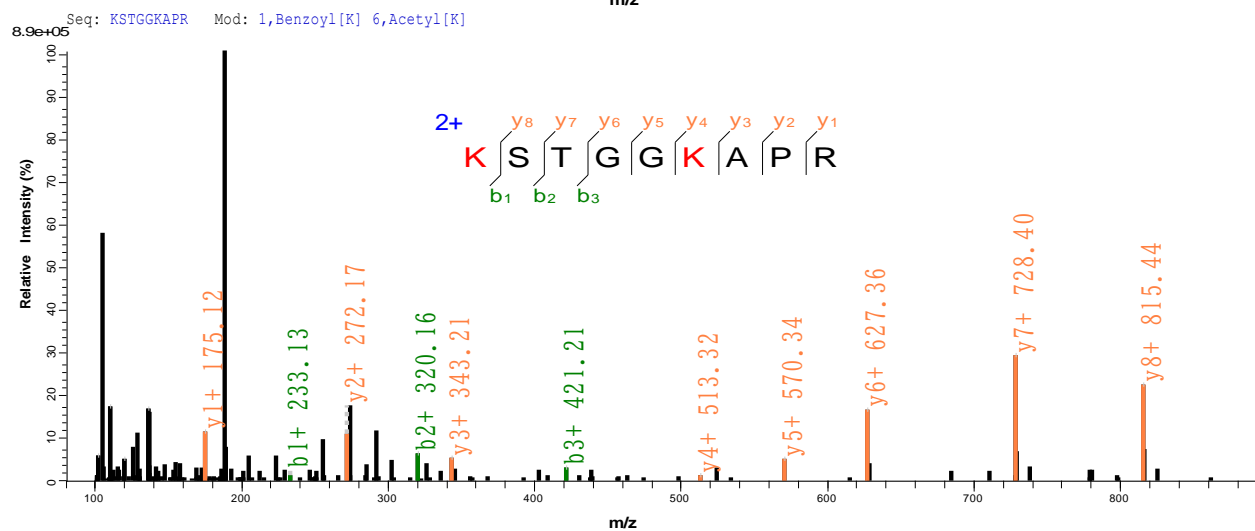
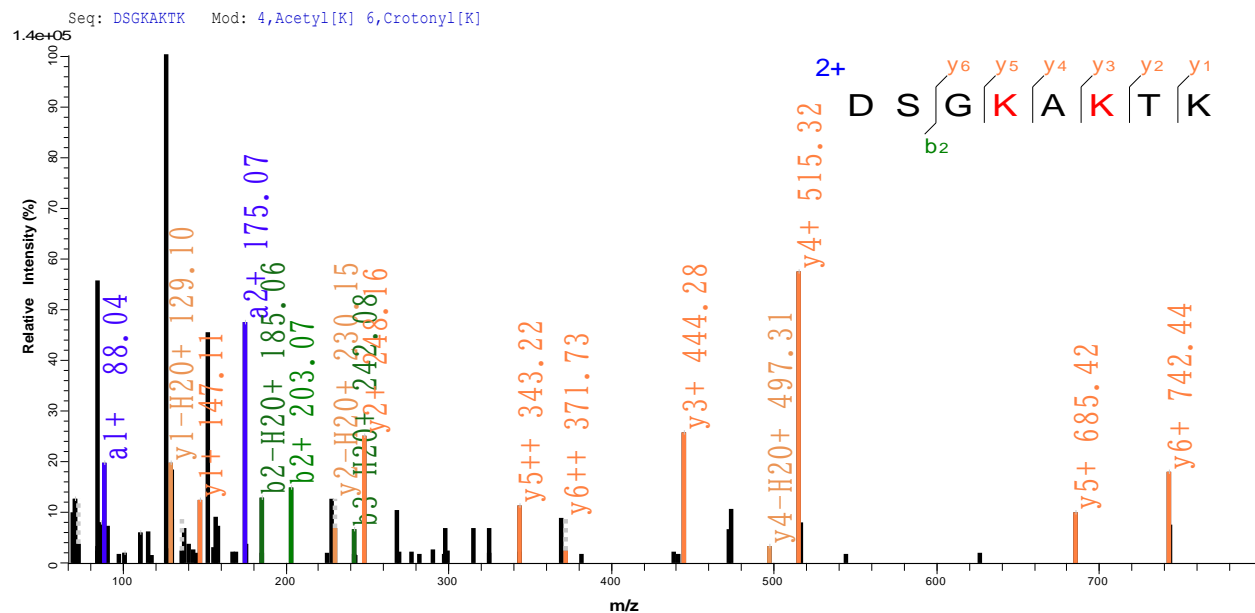








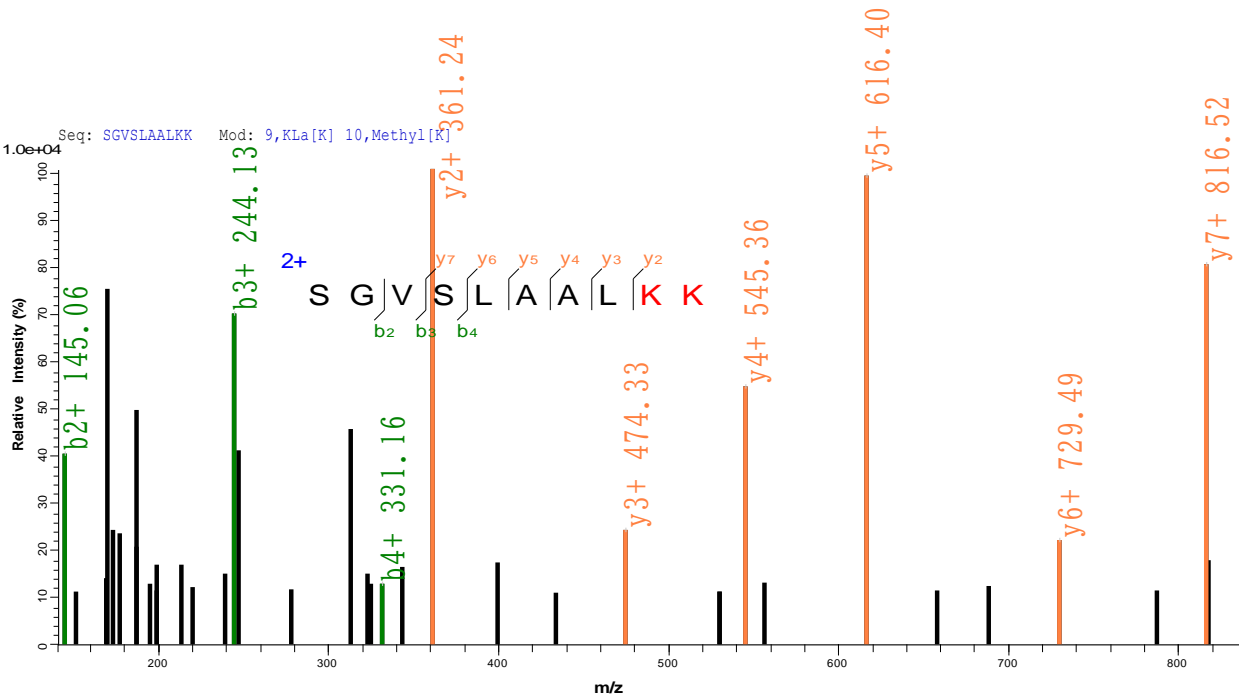
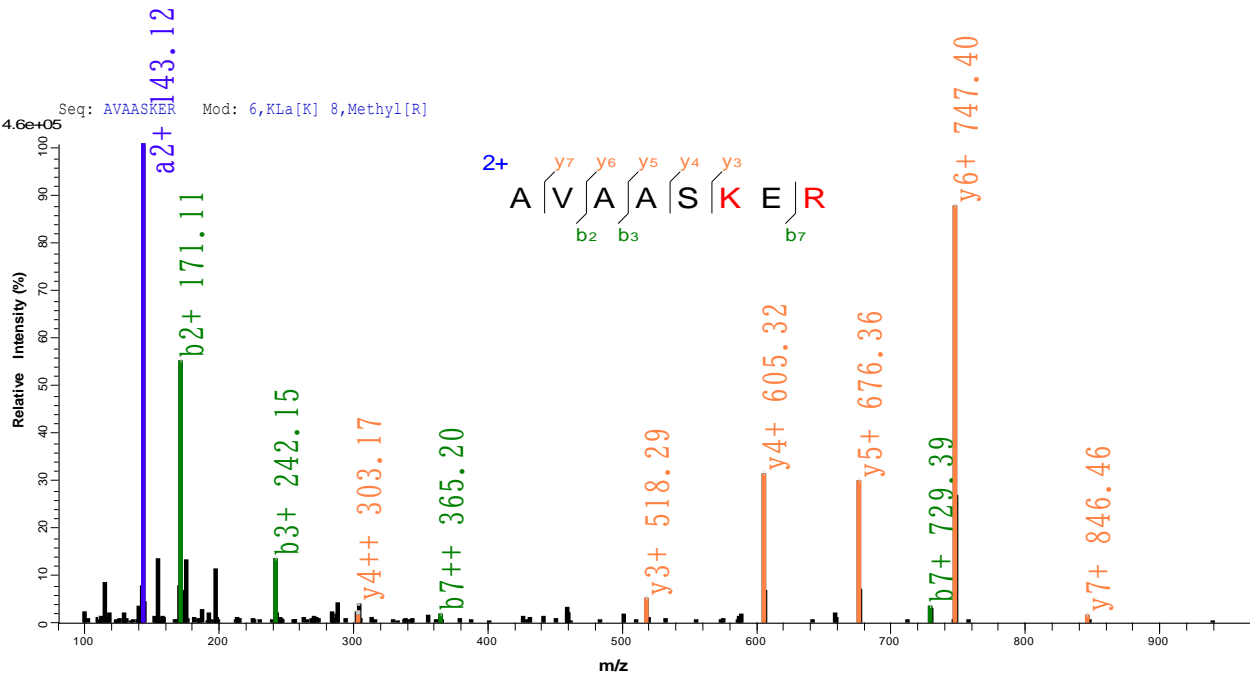


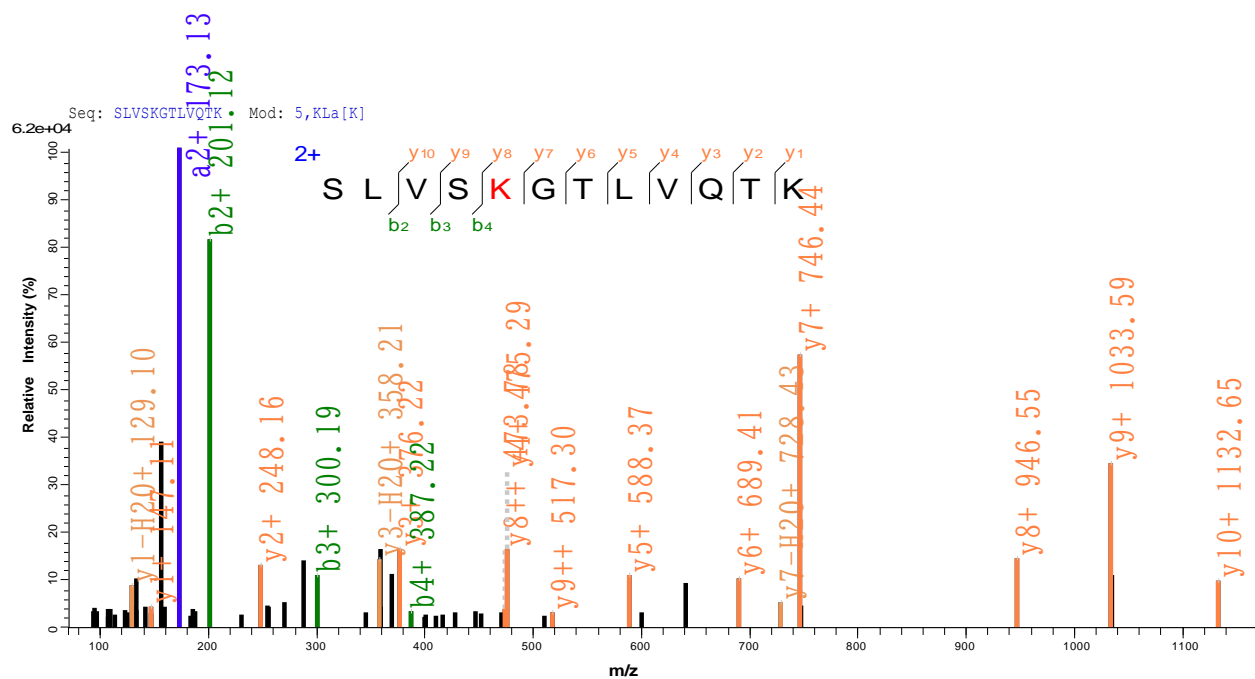
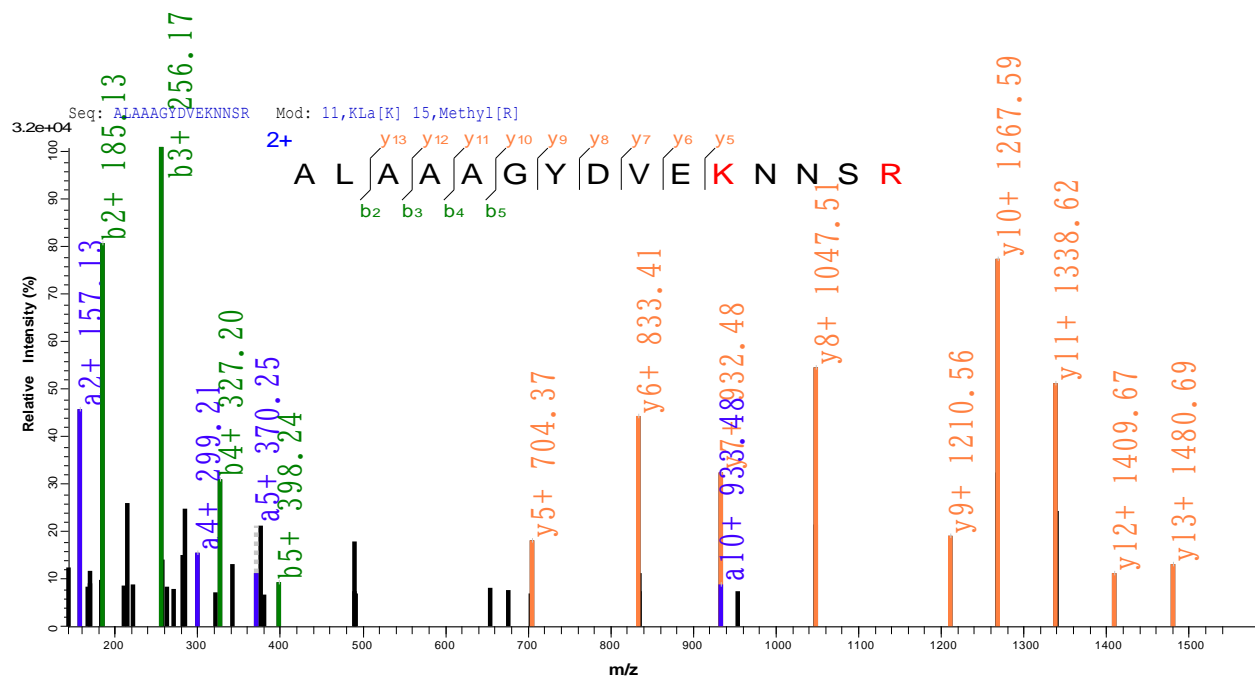


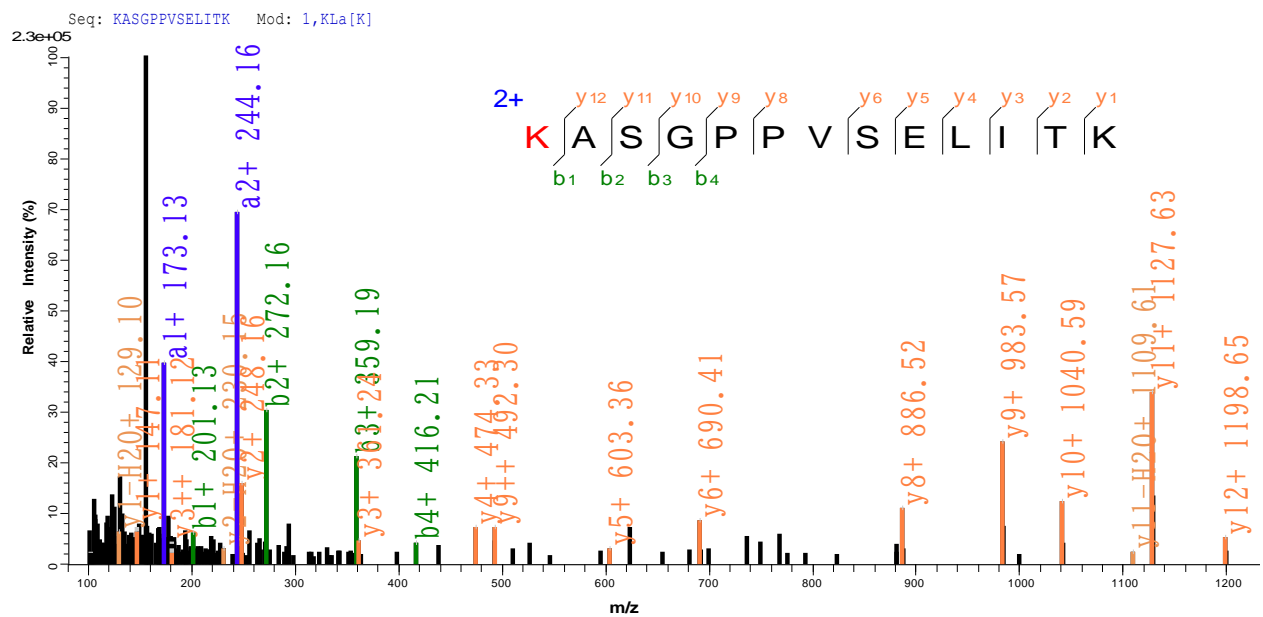
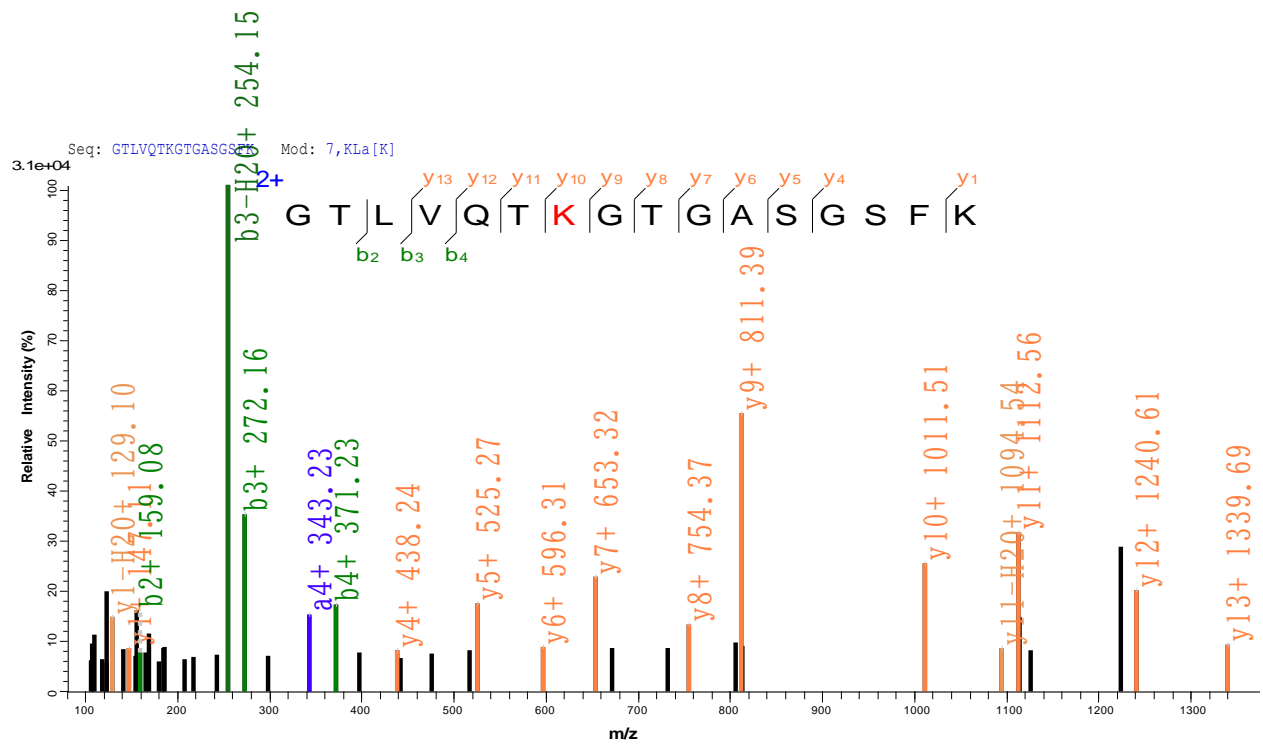
Data S6. PSMs for all newly identified histone marks.

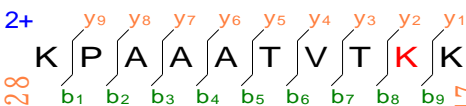
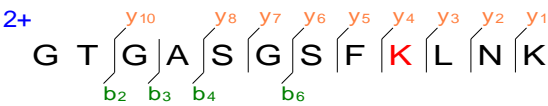
A. K1a

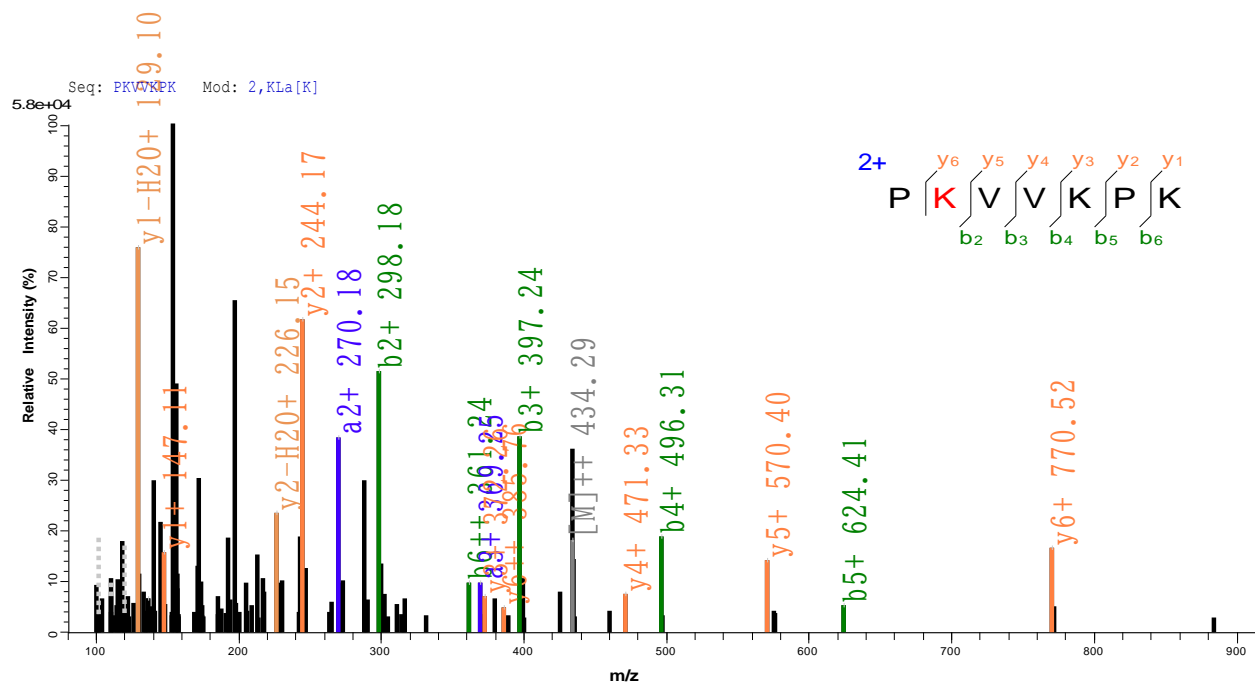
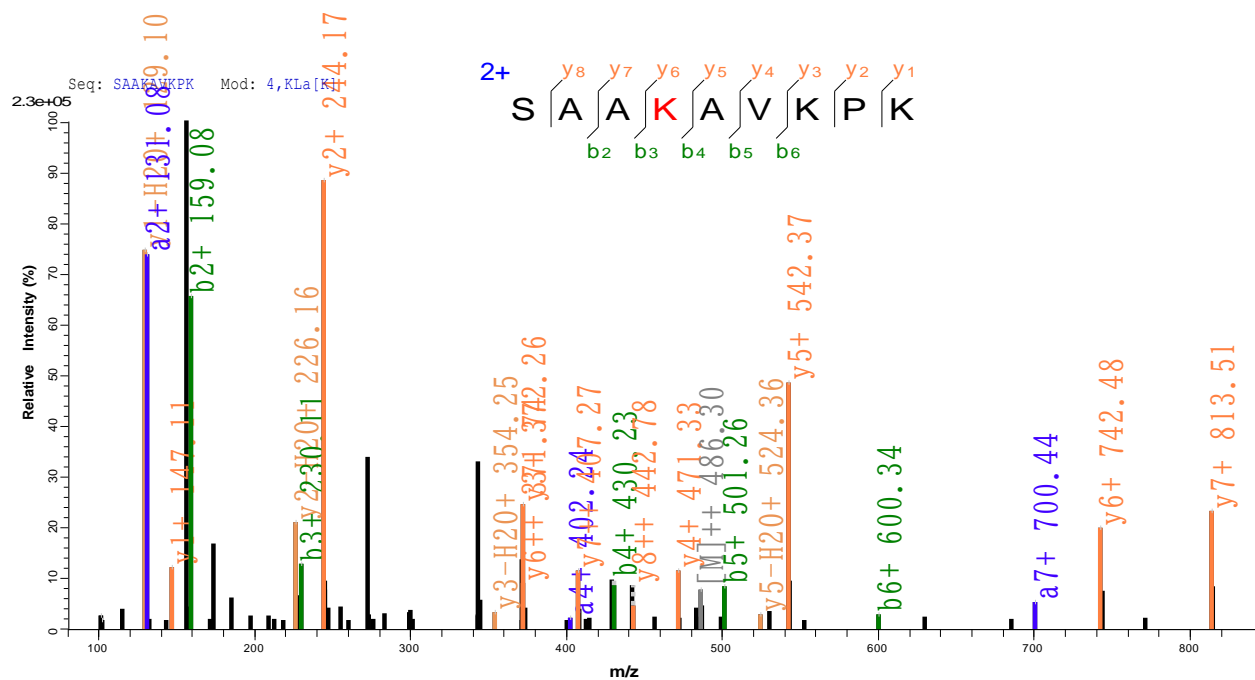
H1

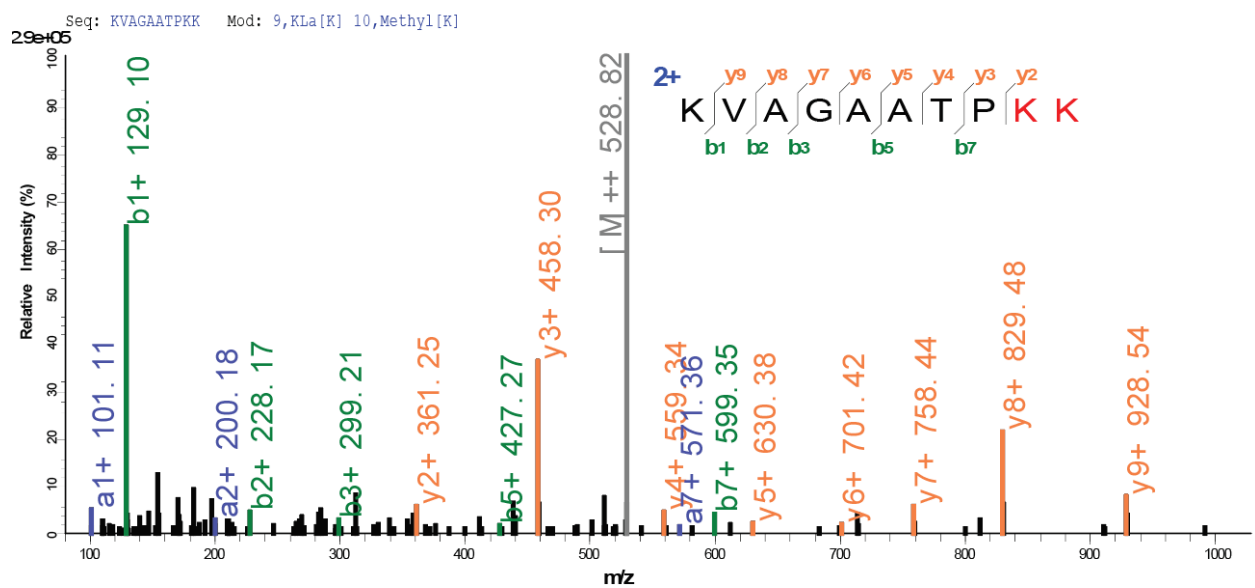
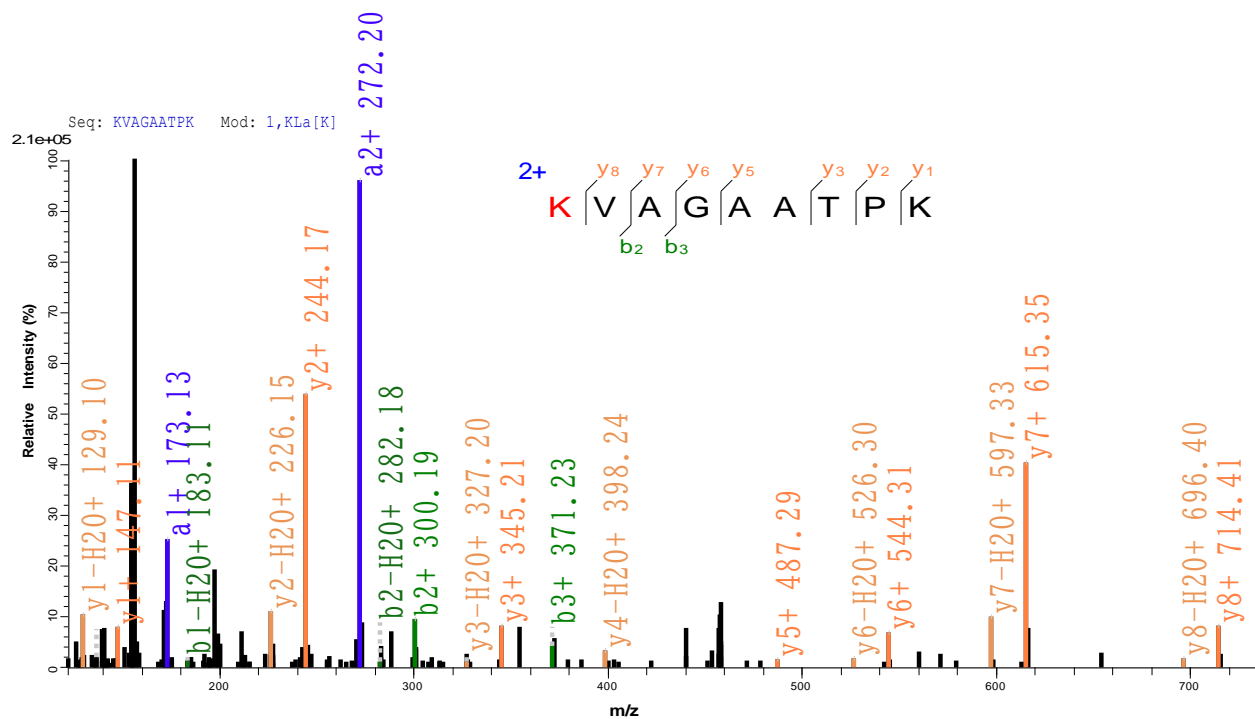


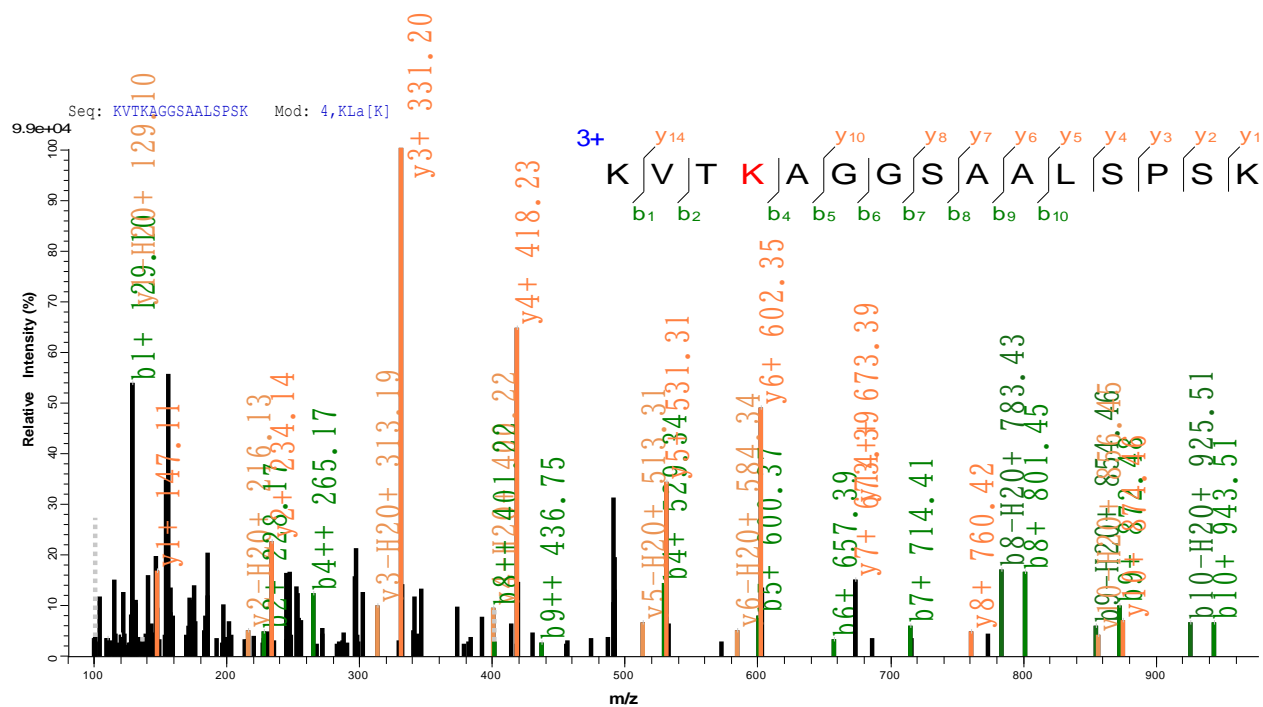
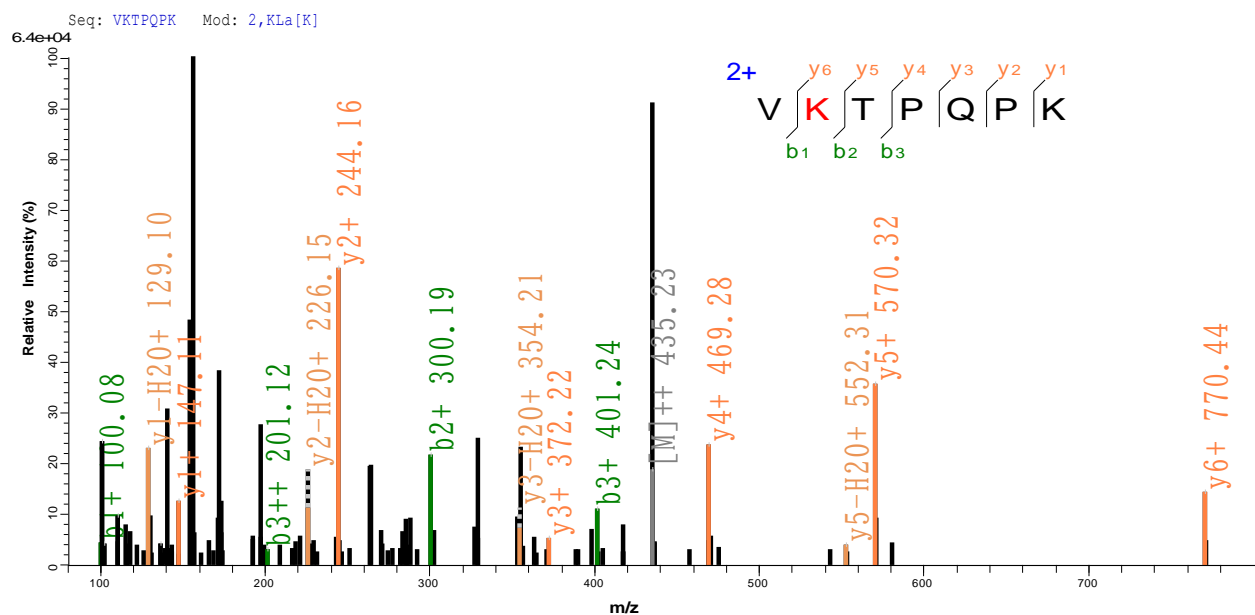




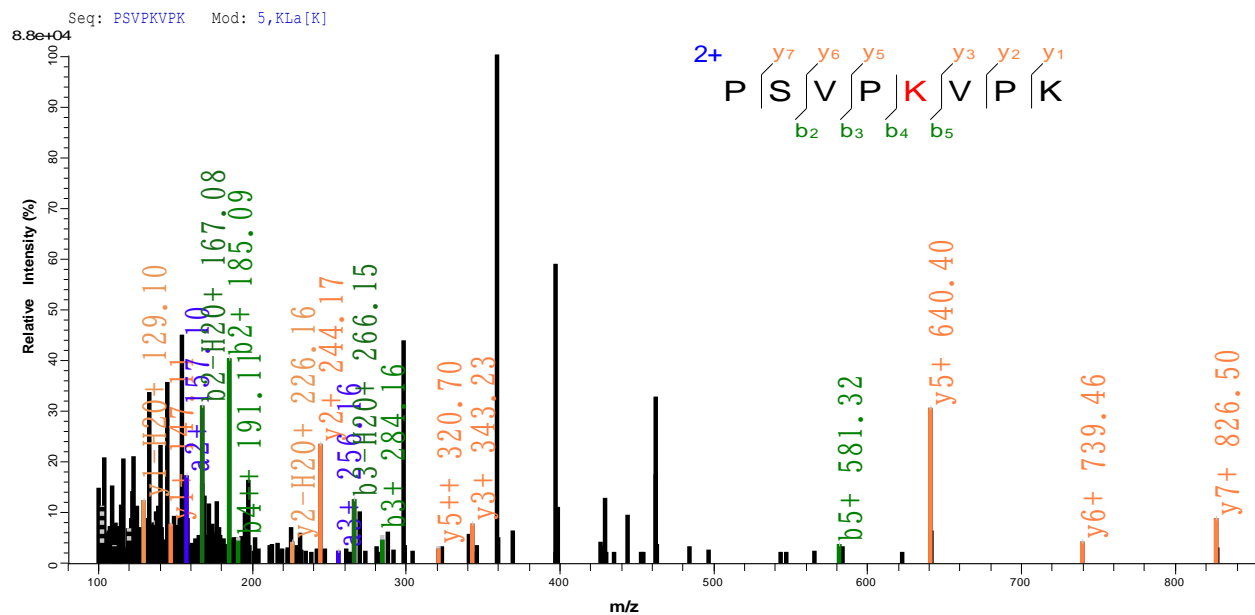
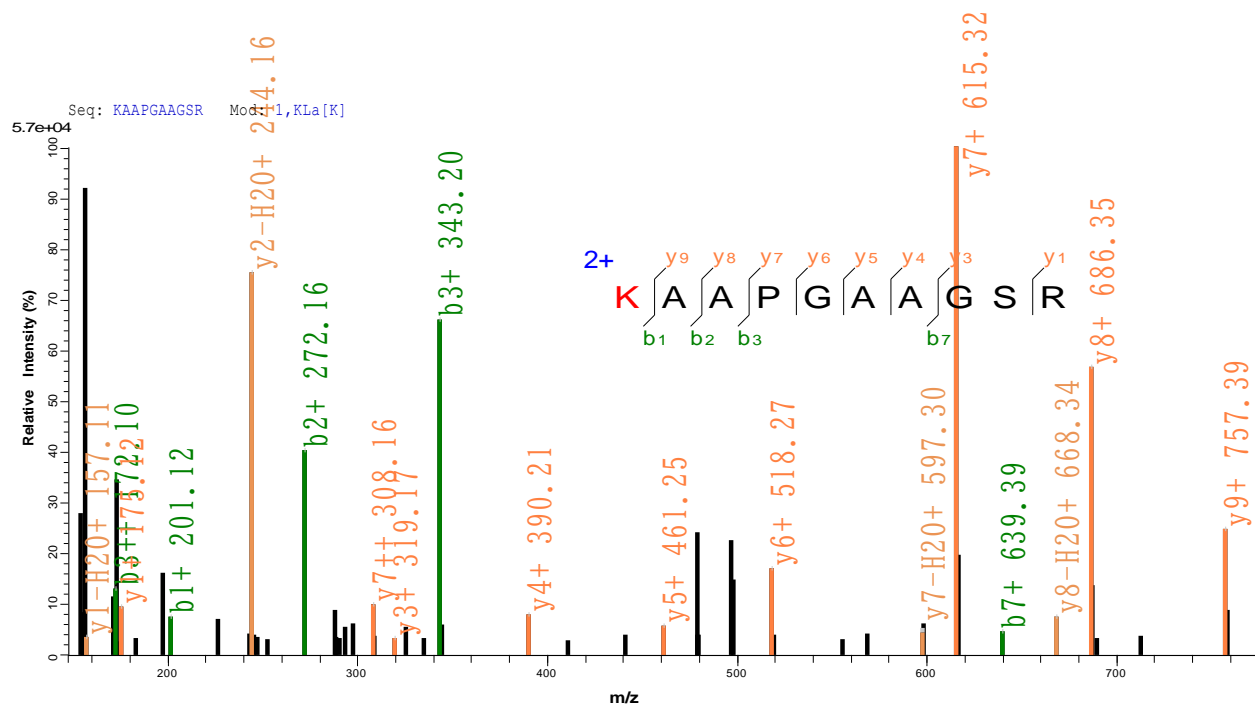


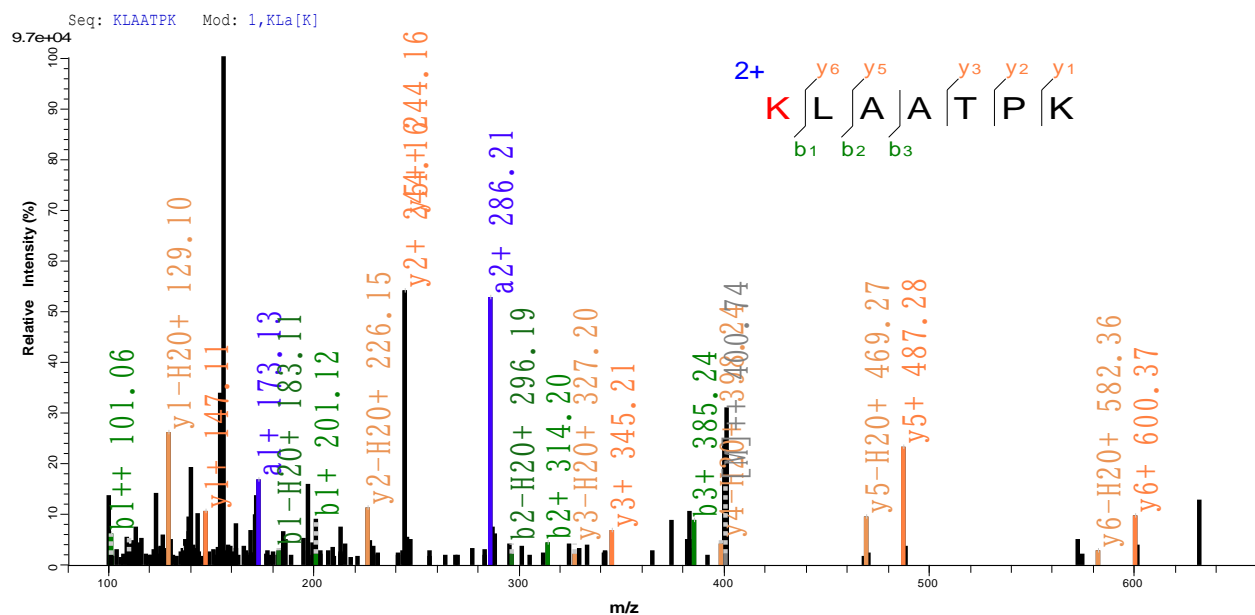




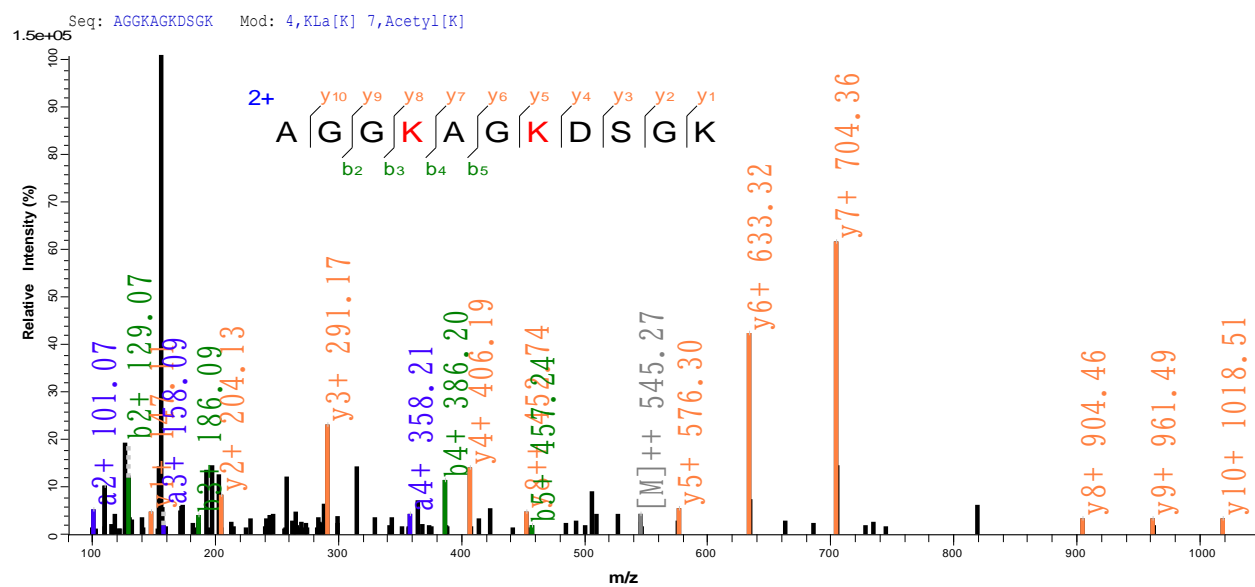


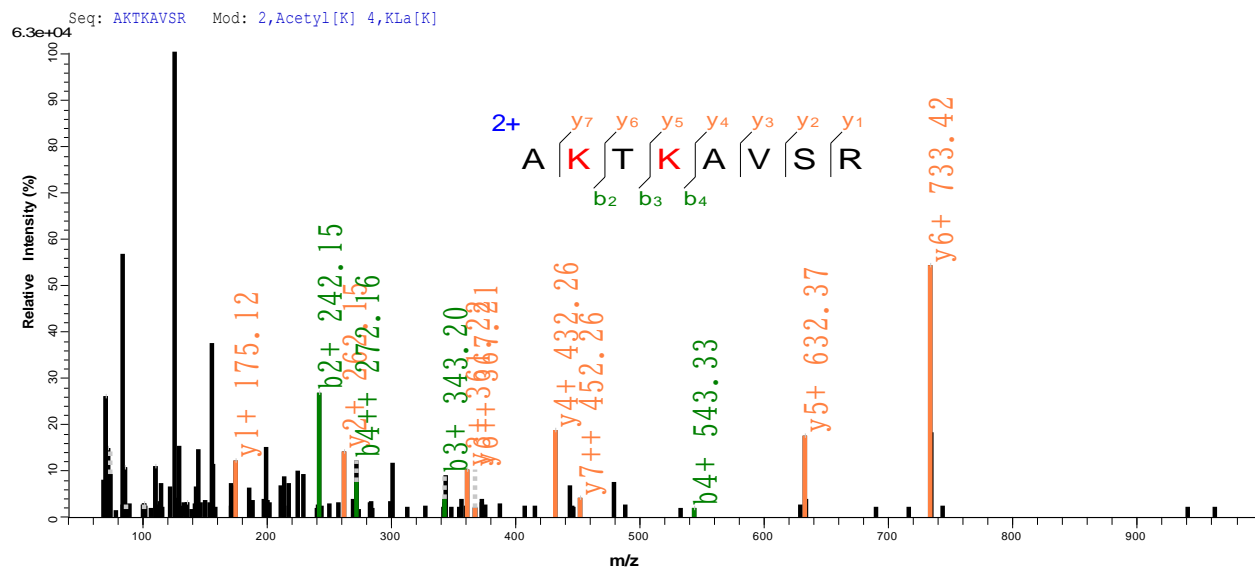
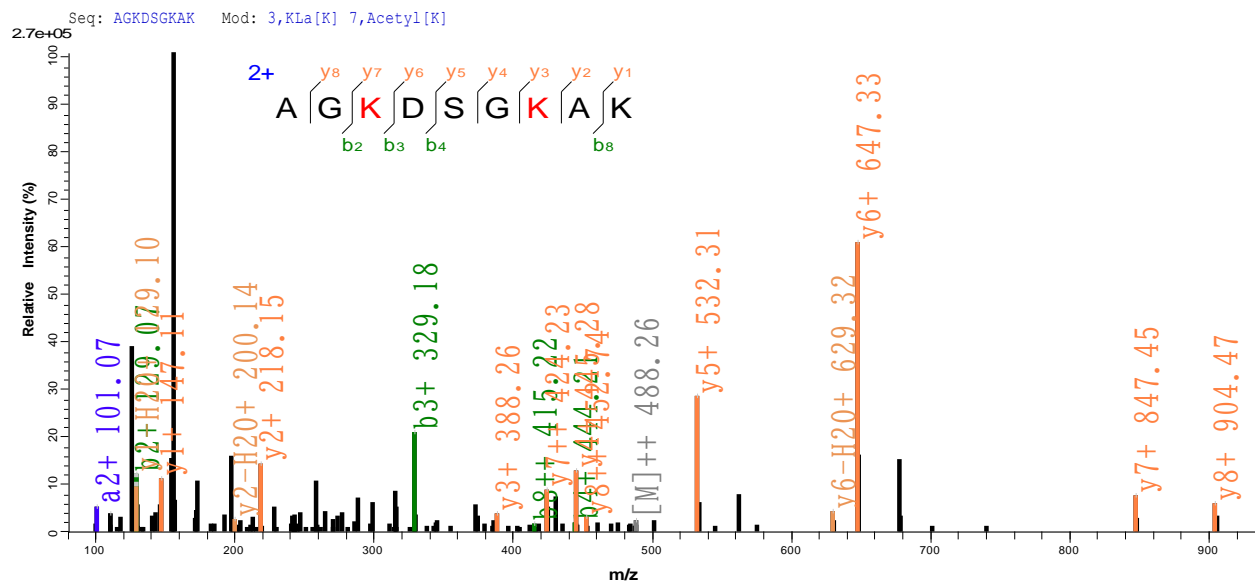


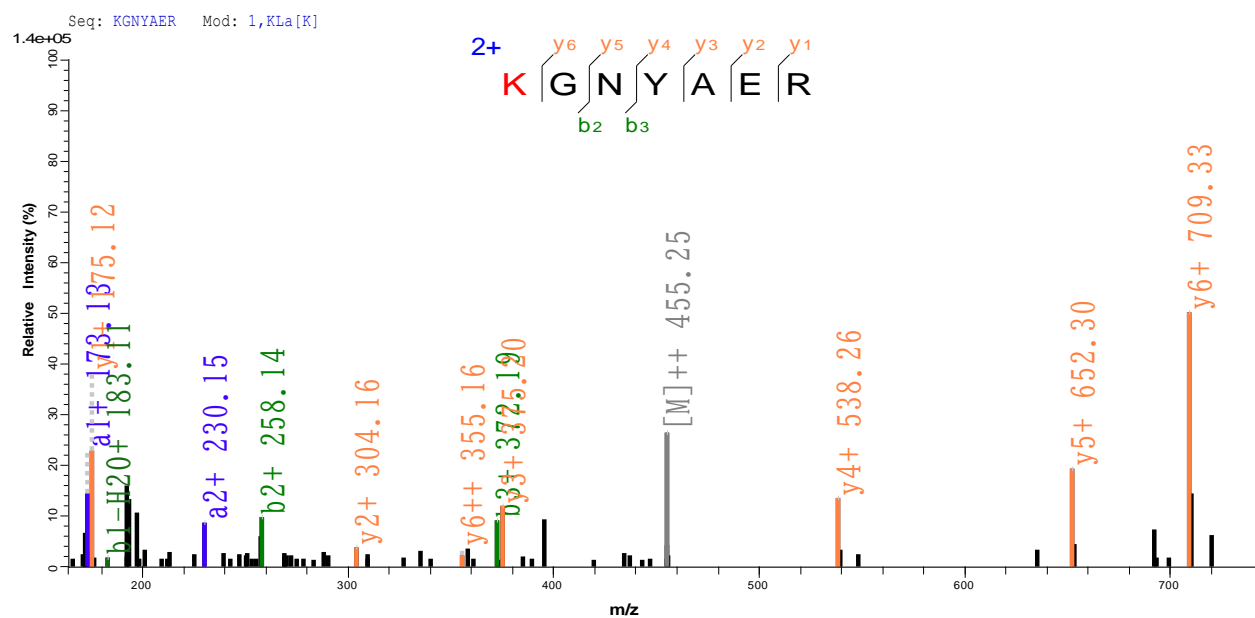
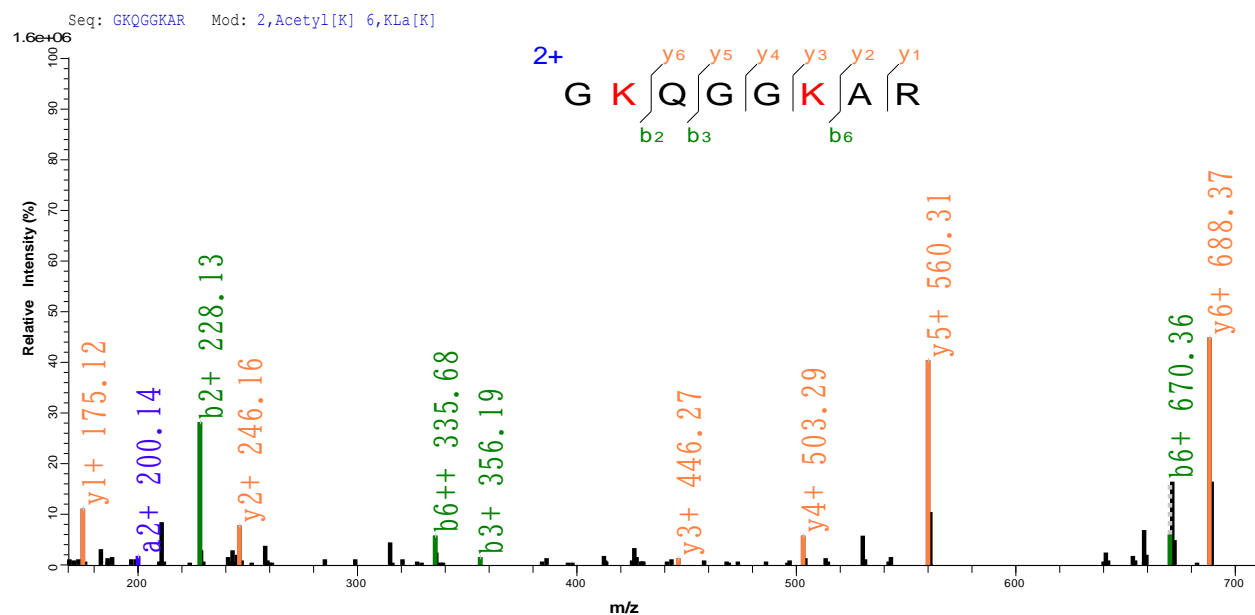


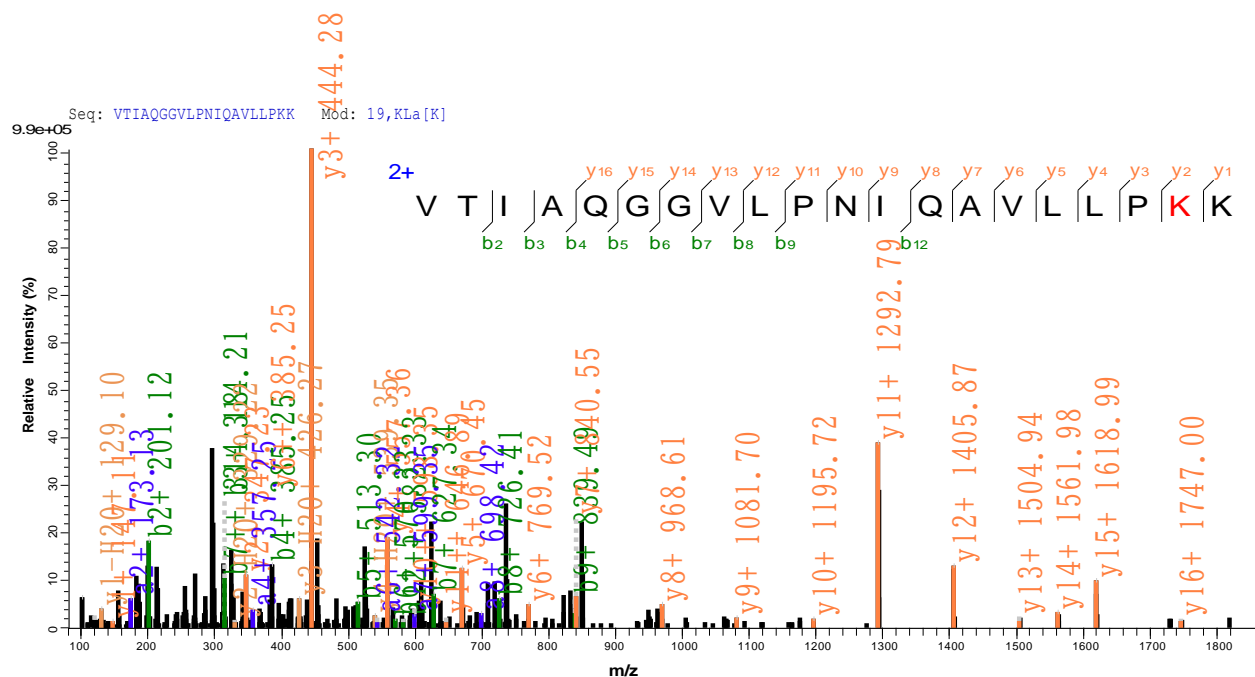
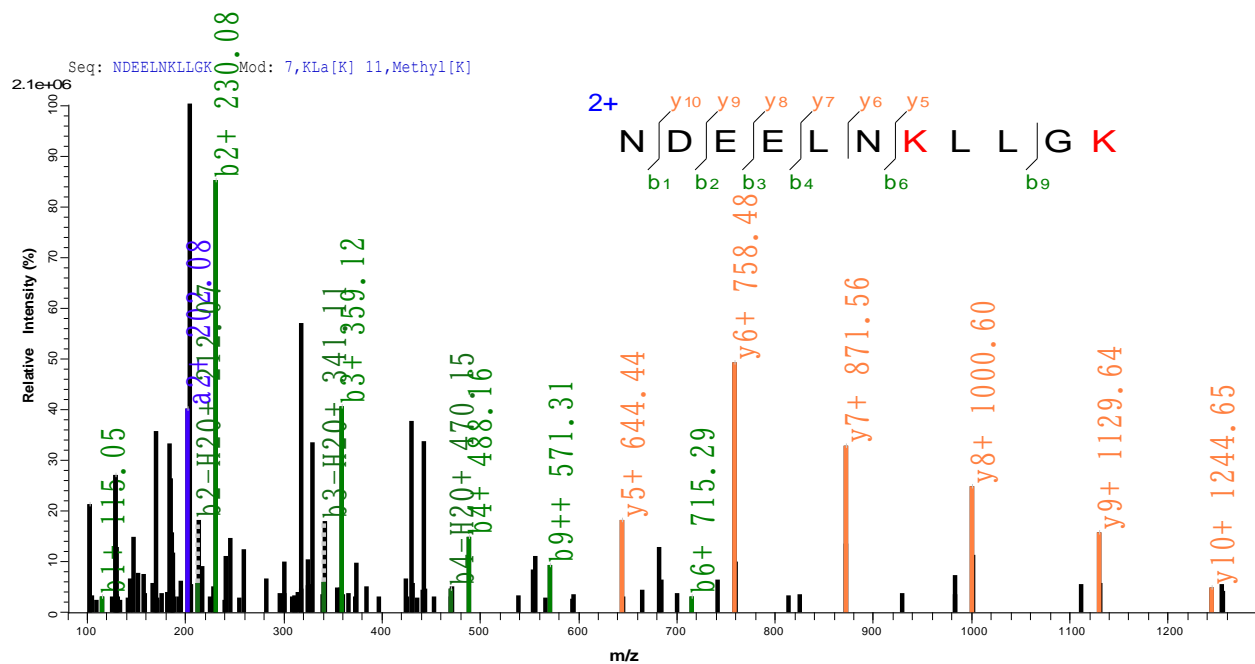


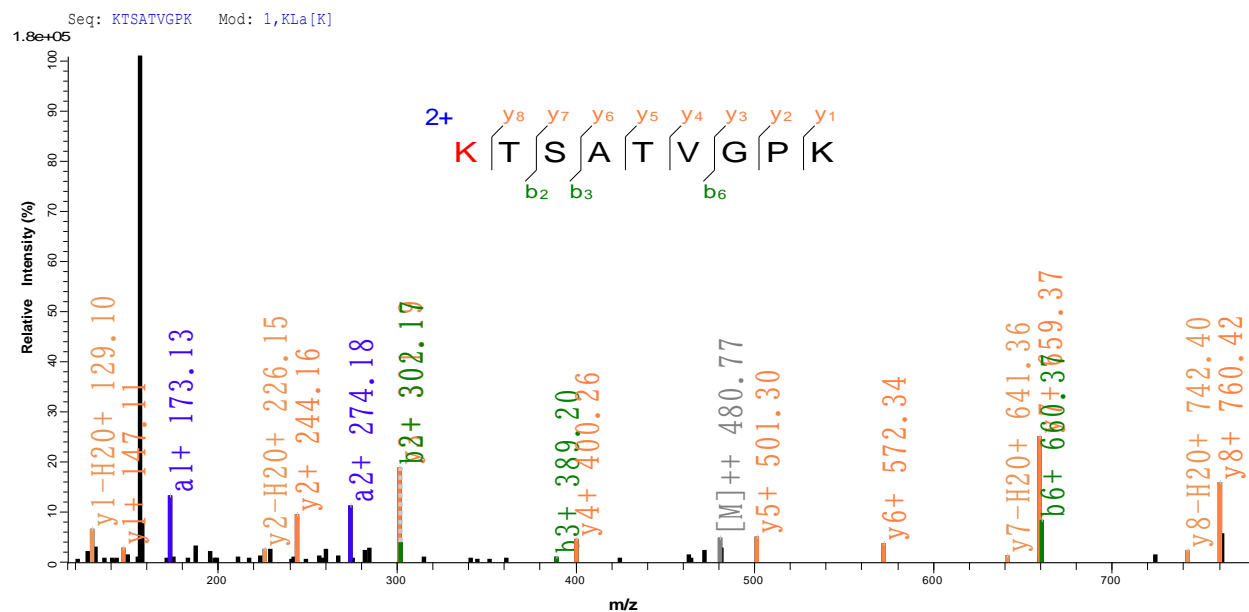
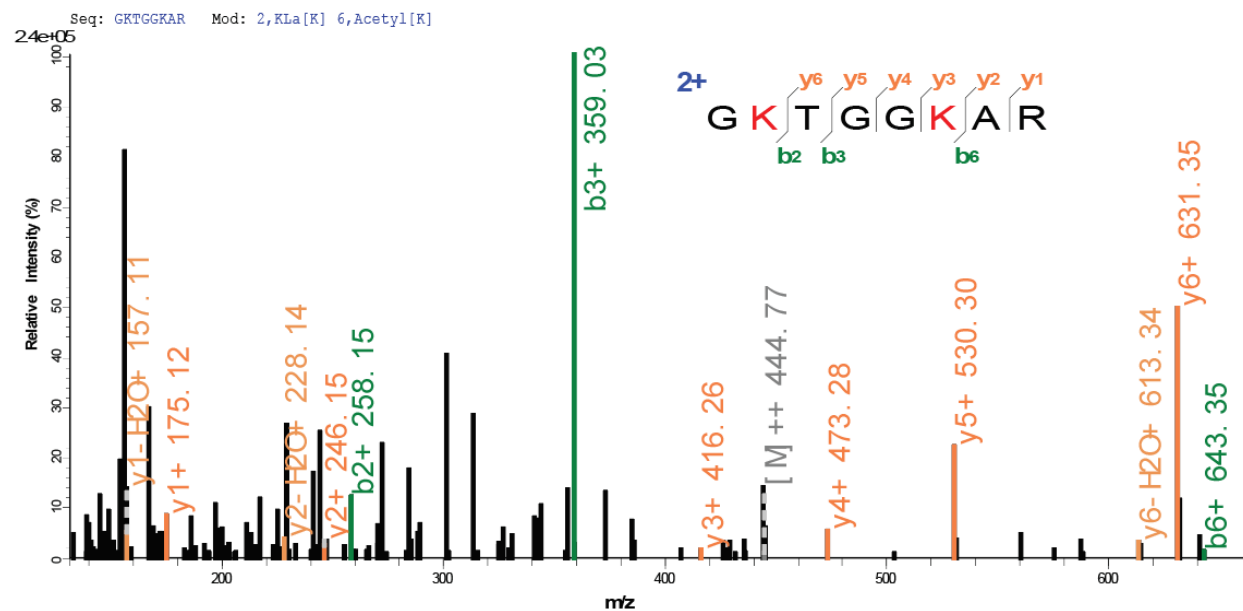
## H2A

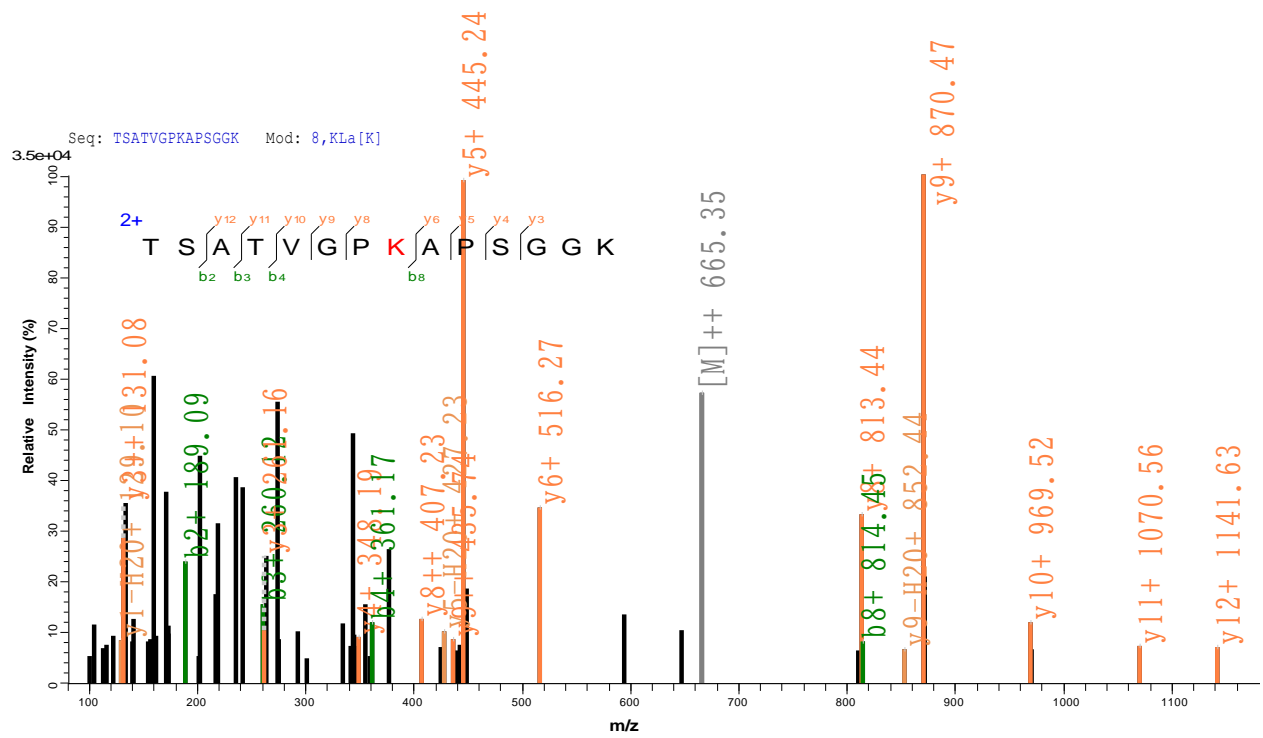




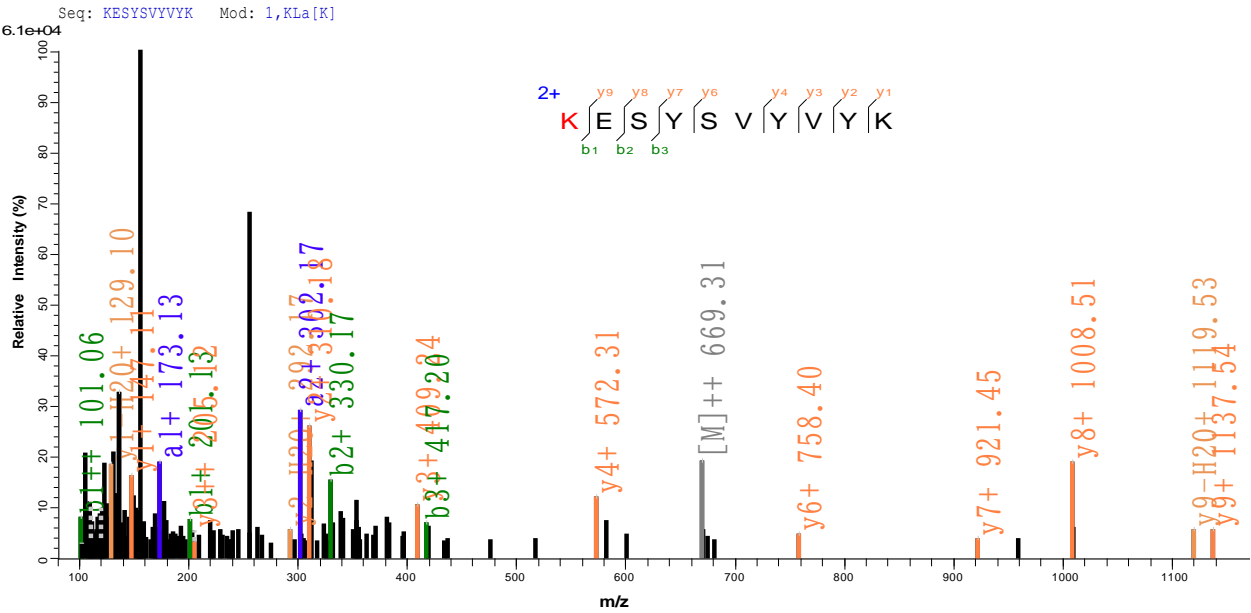




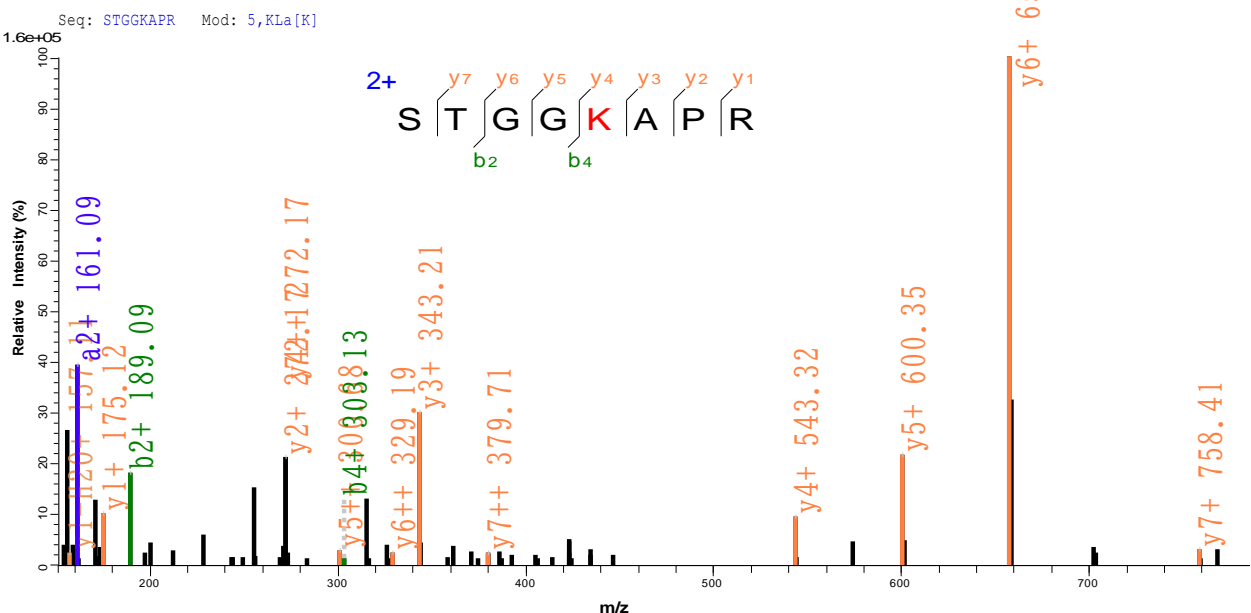




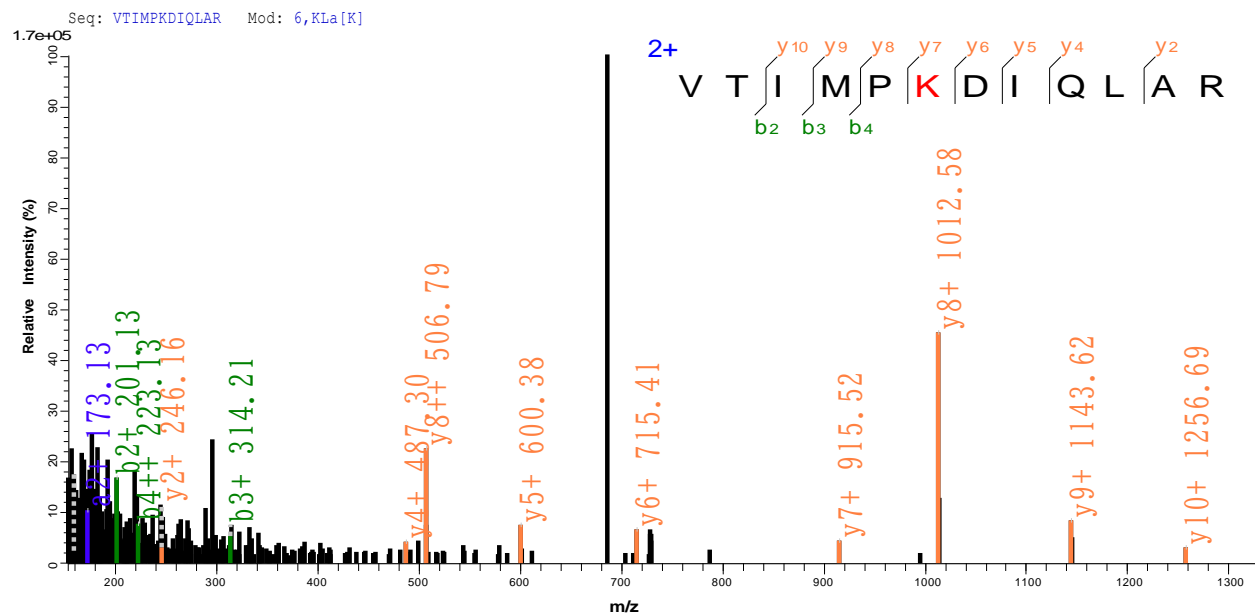
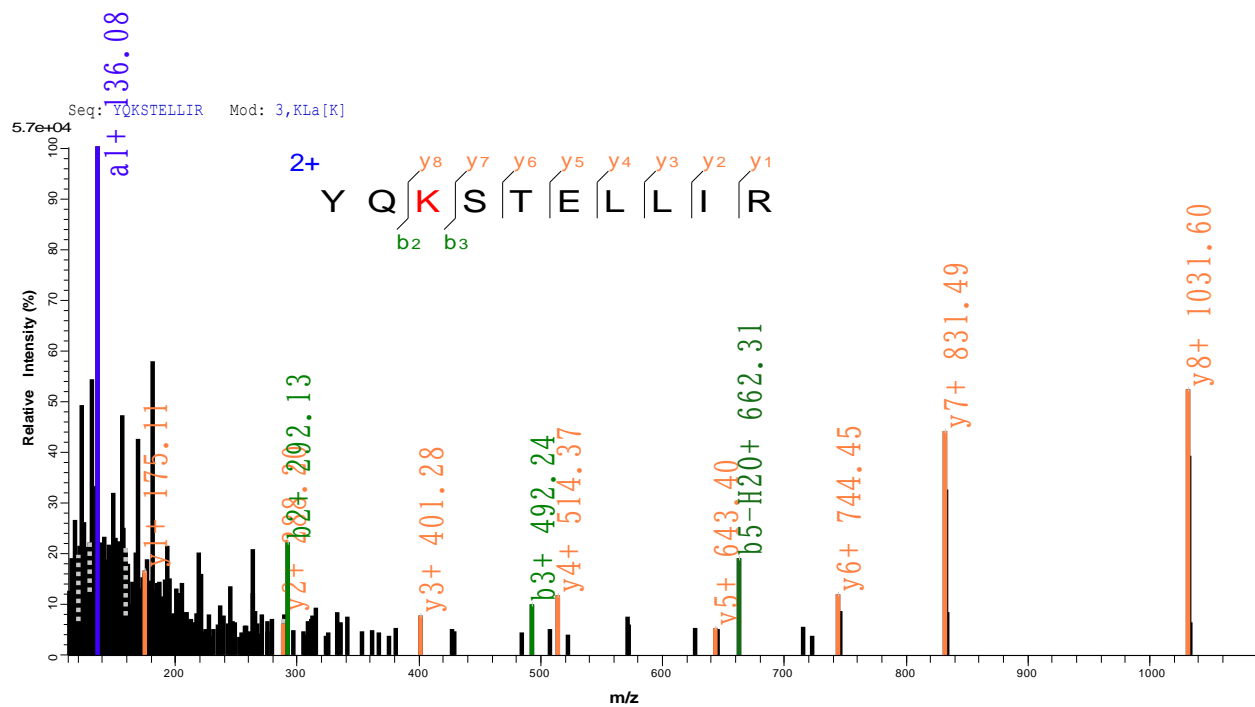
H2B



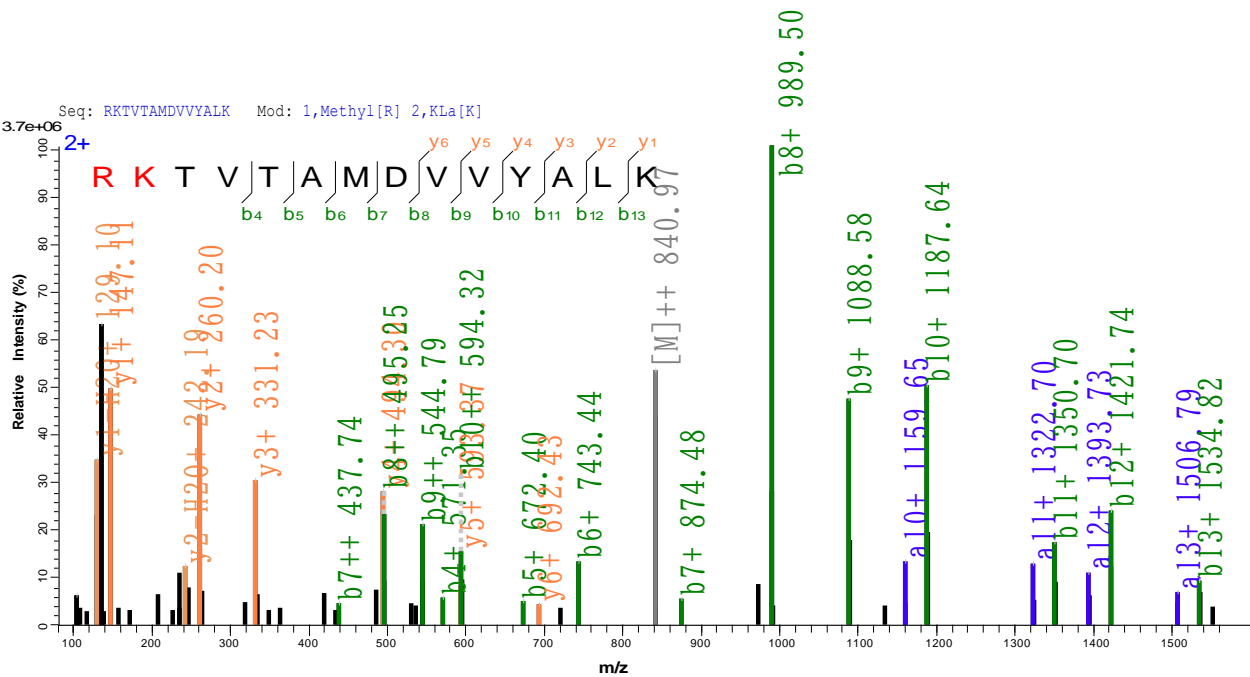
H3





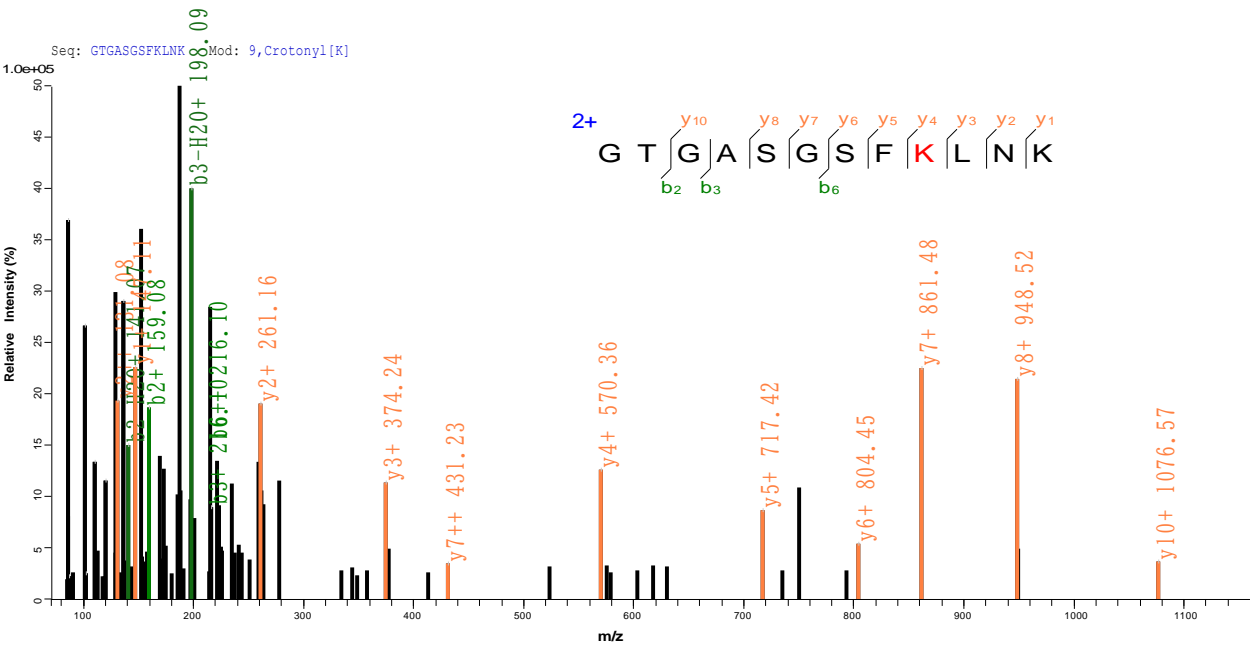


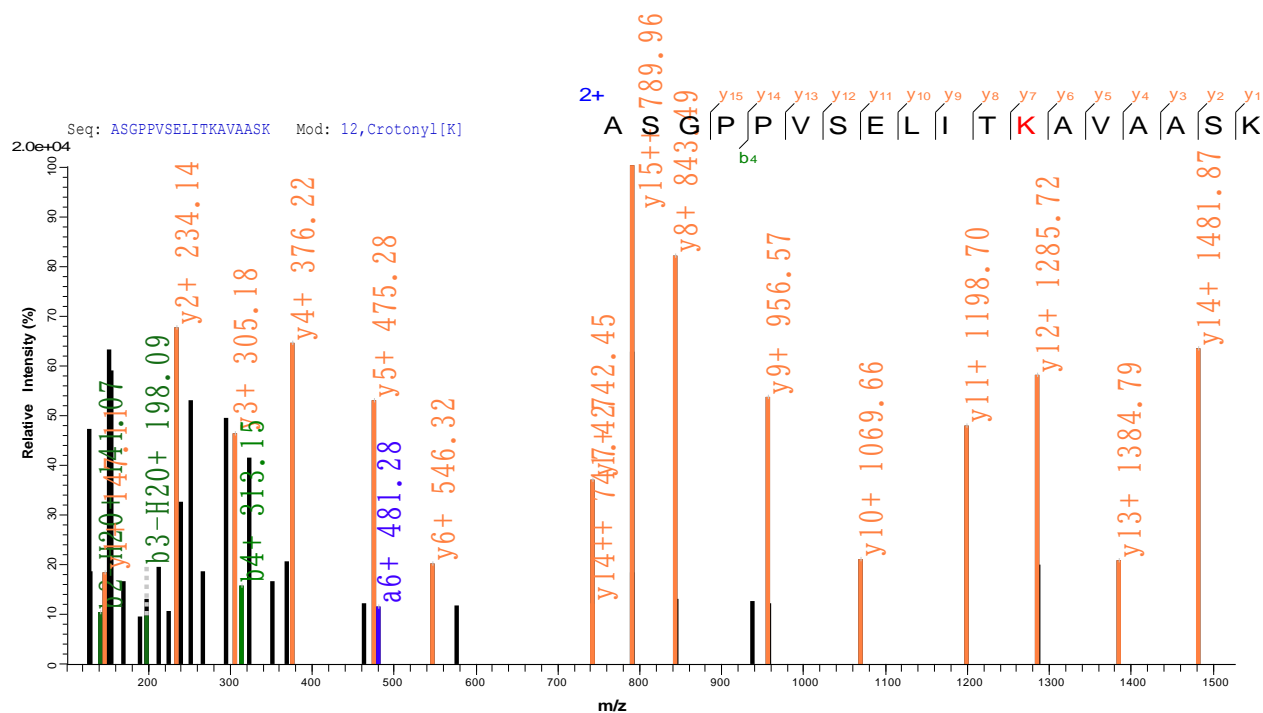
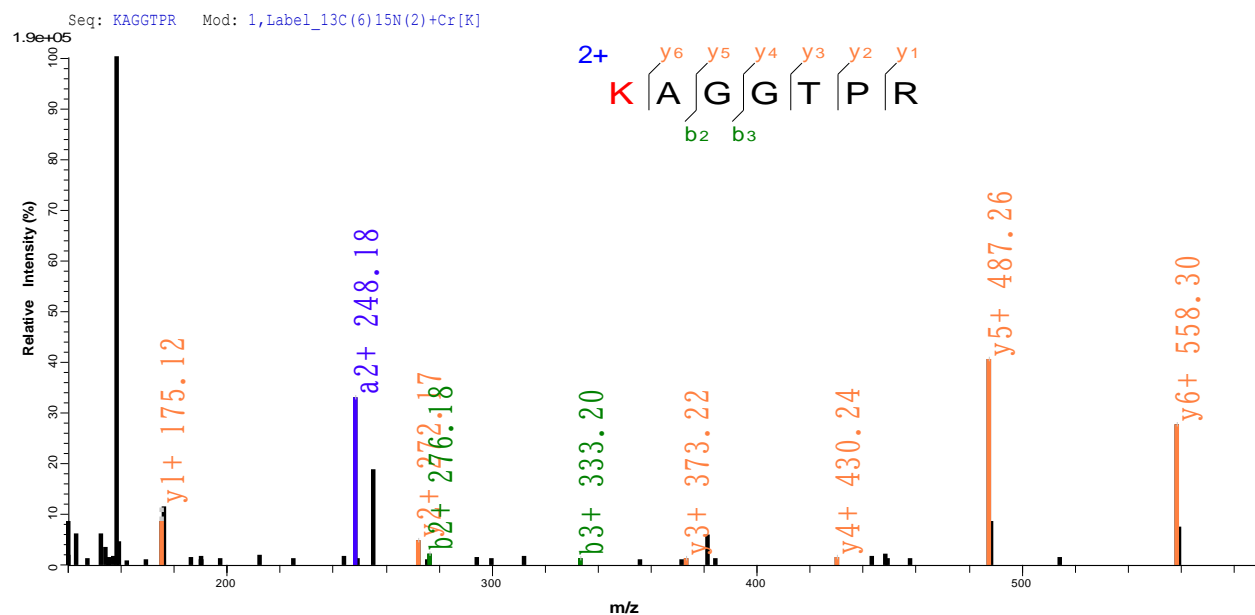
H4

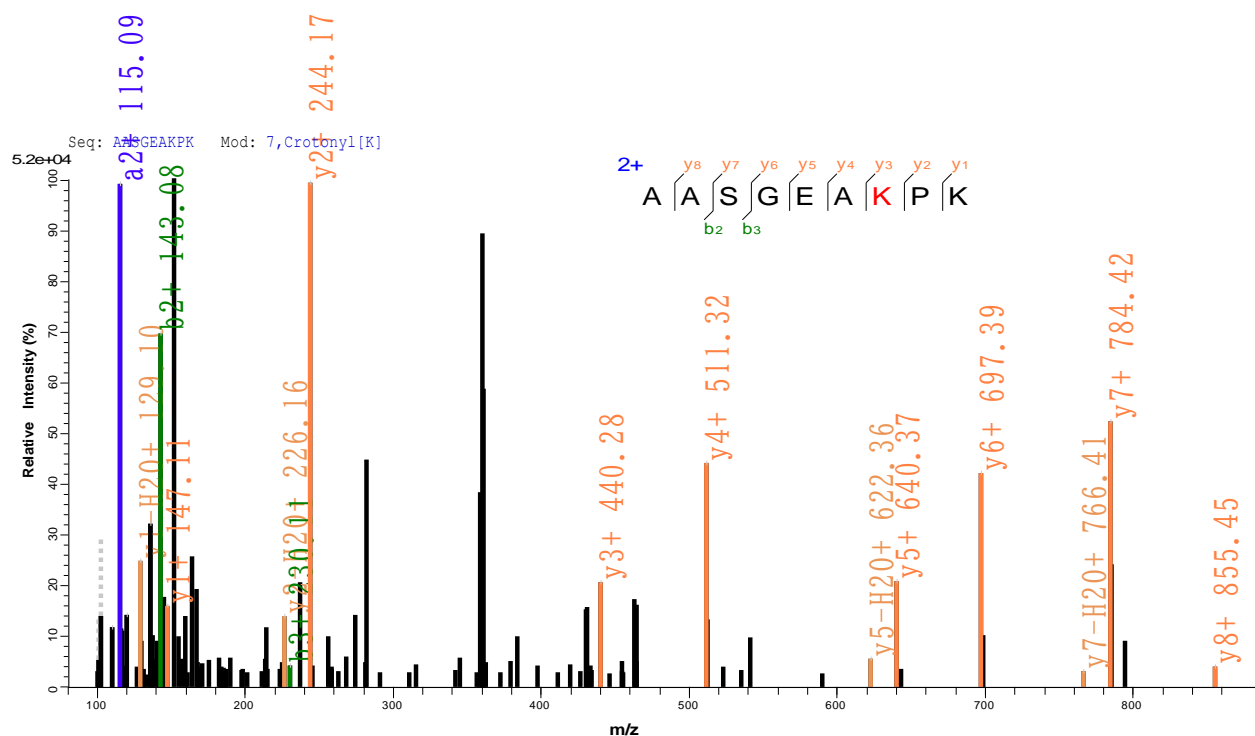
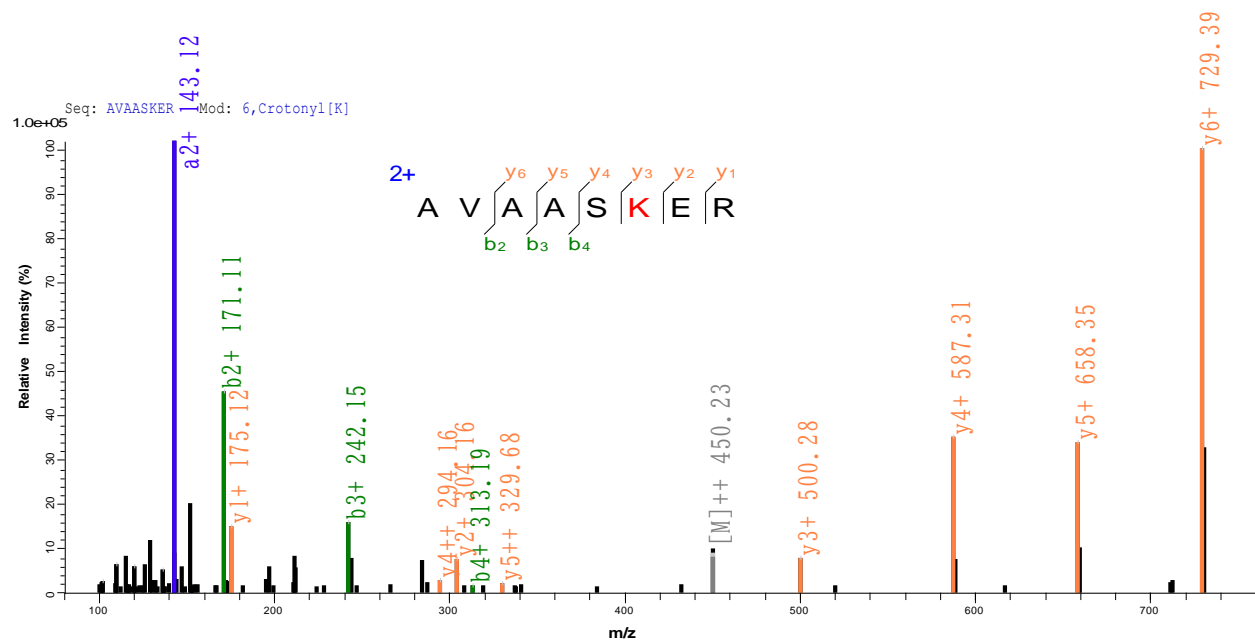


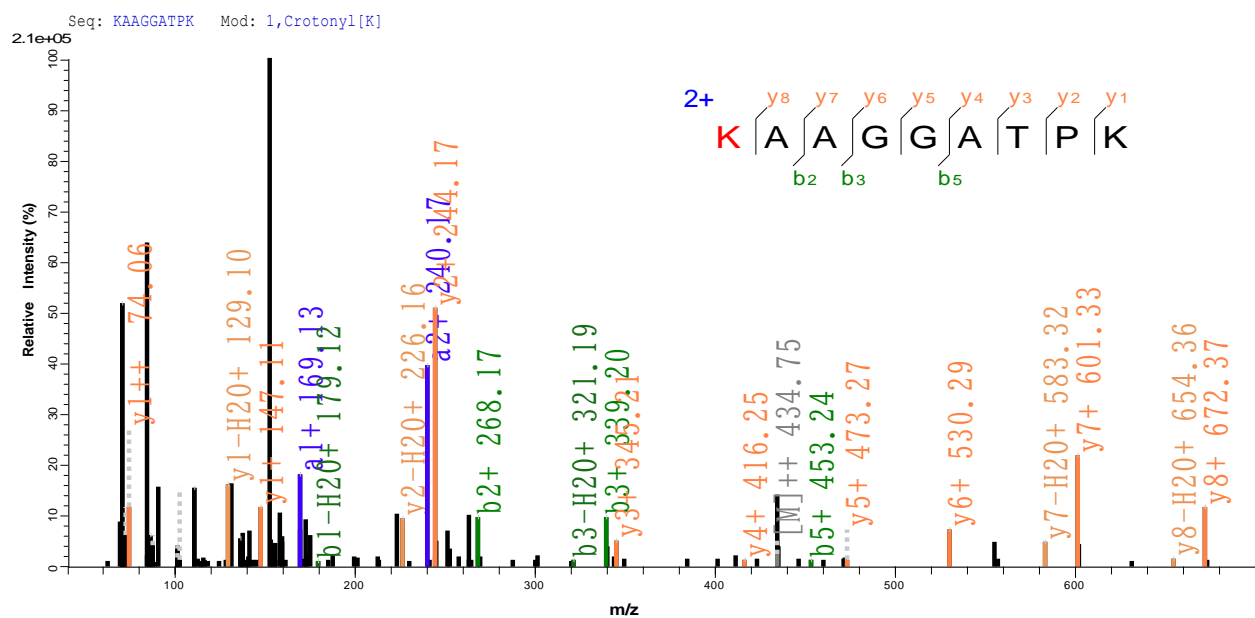
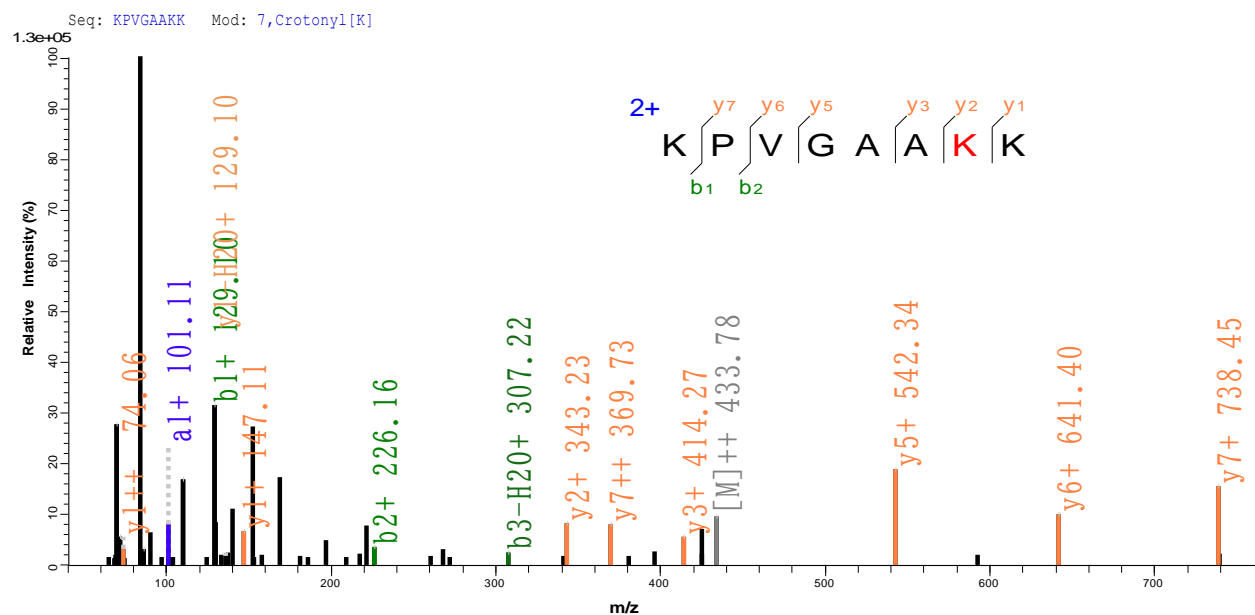
B. Ker

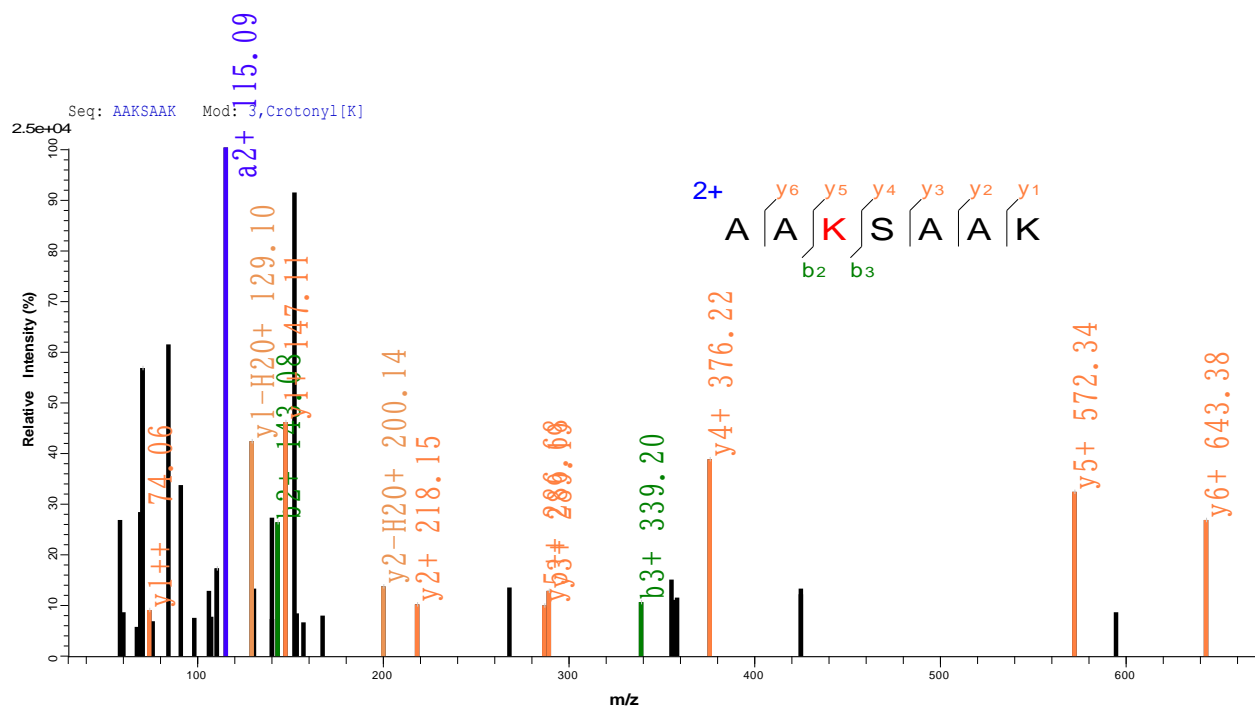
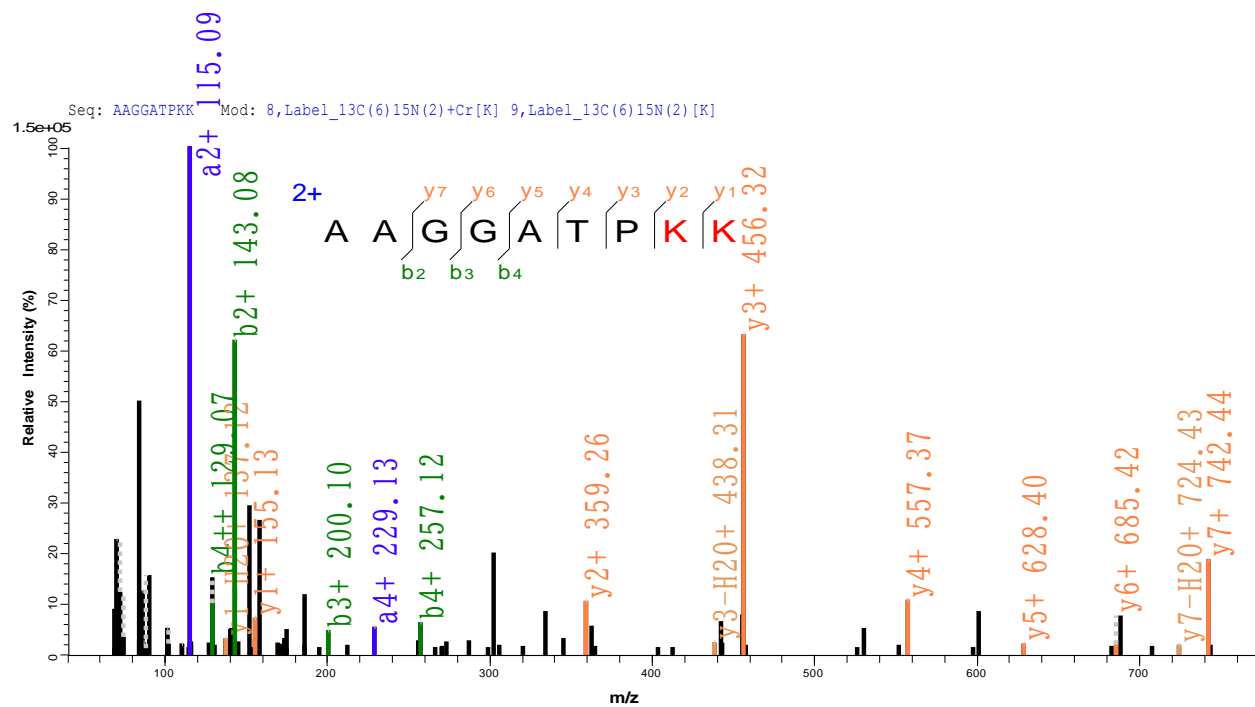
H1

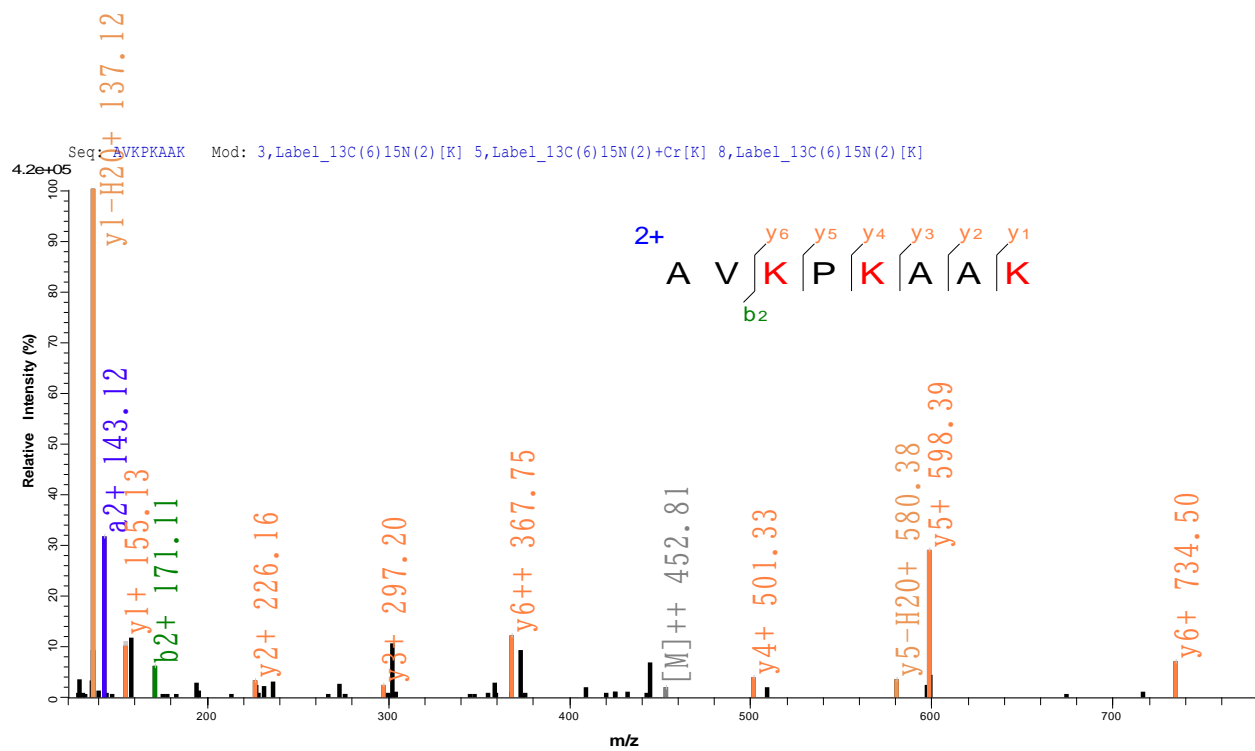
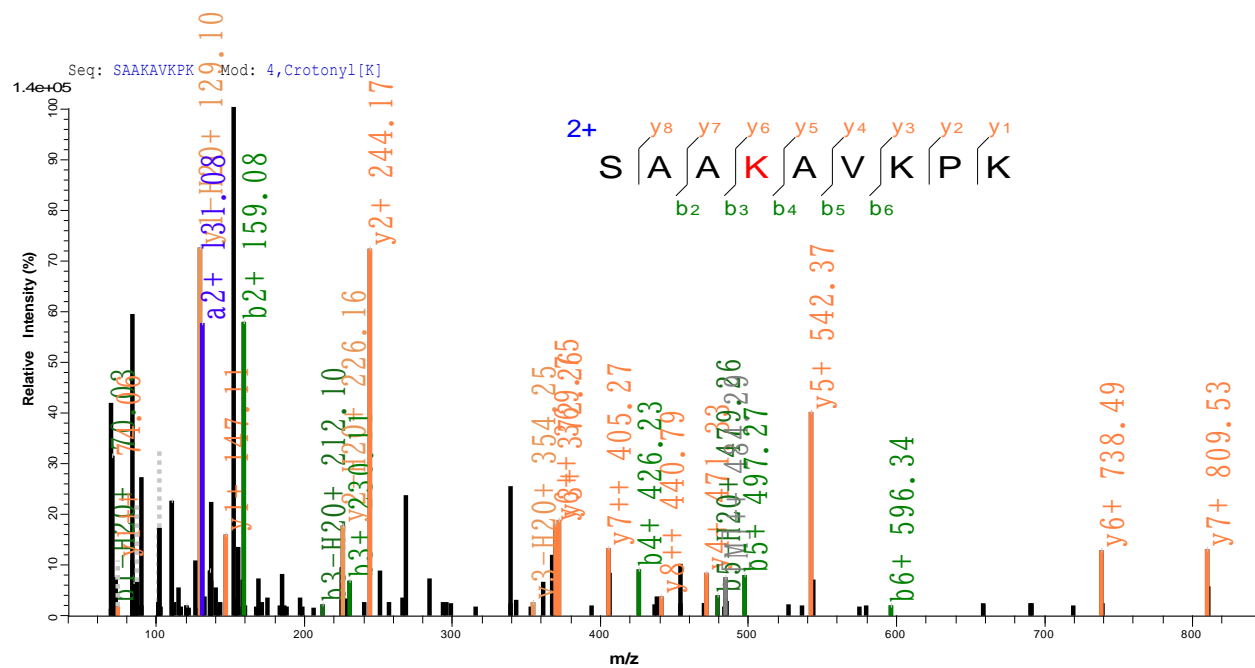


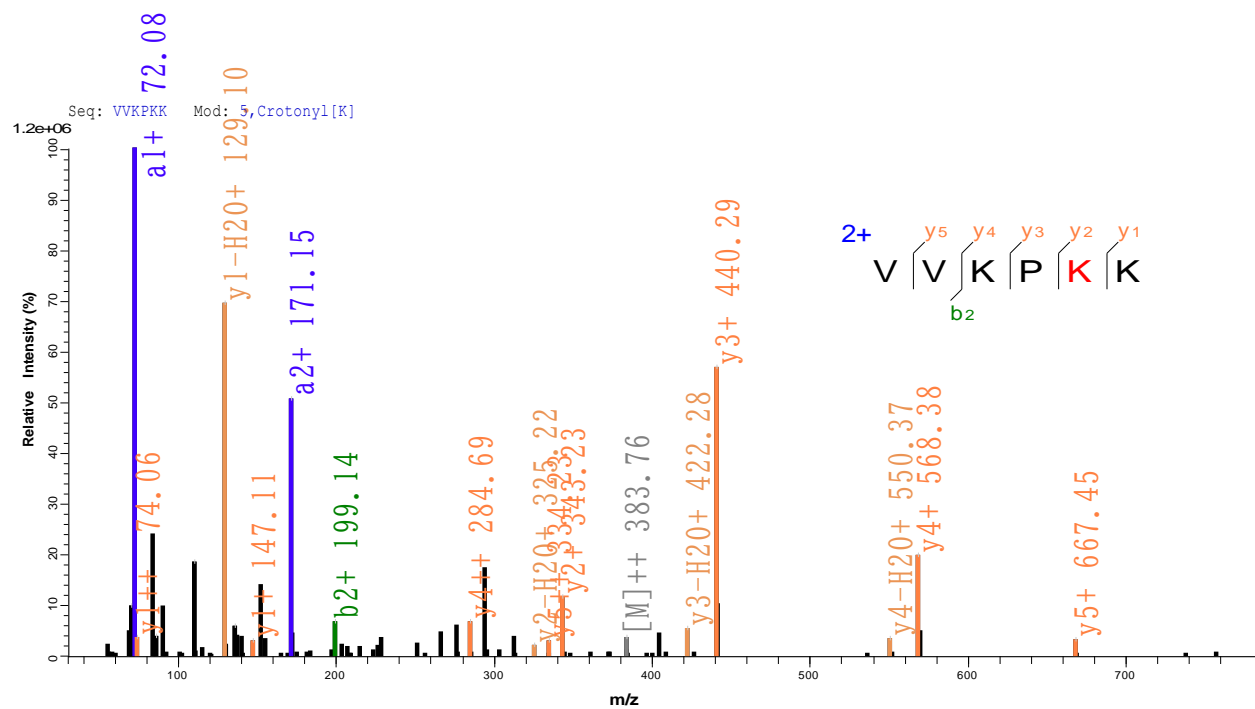
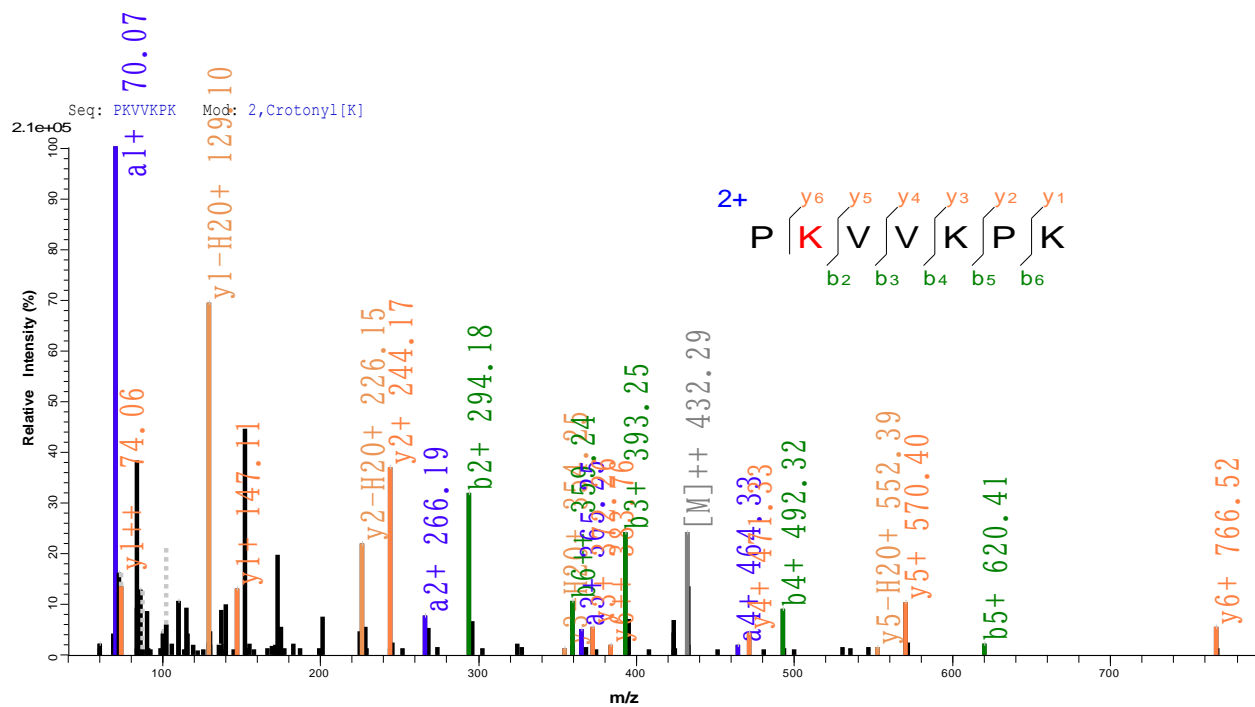




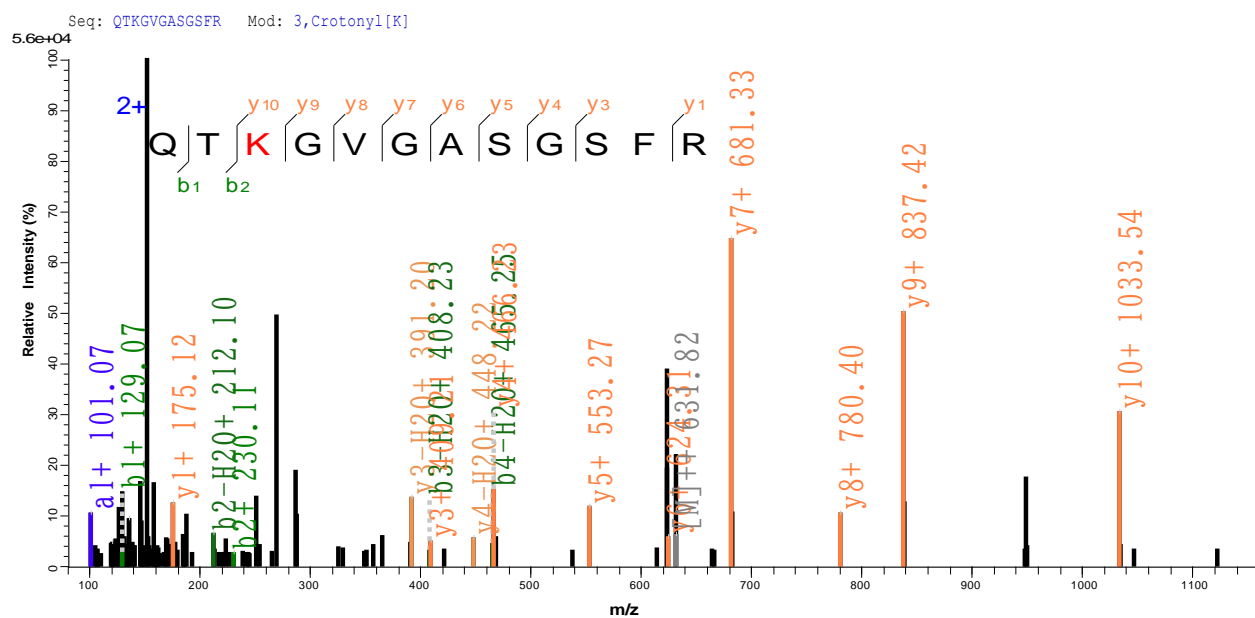
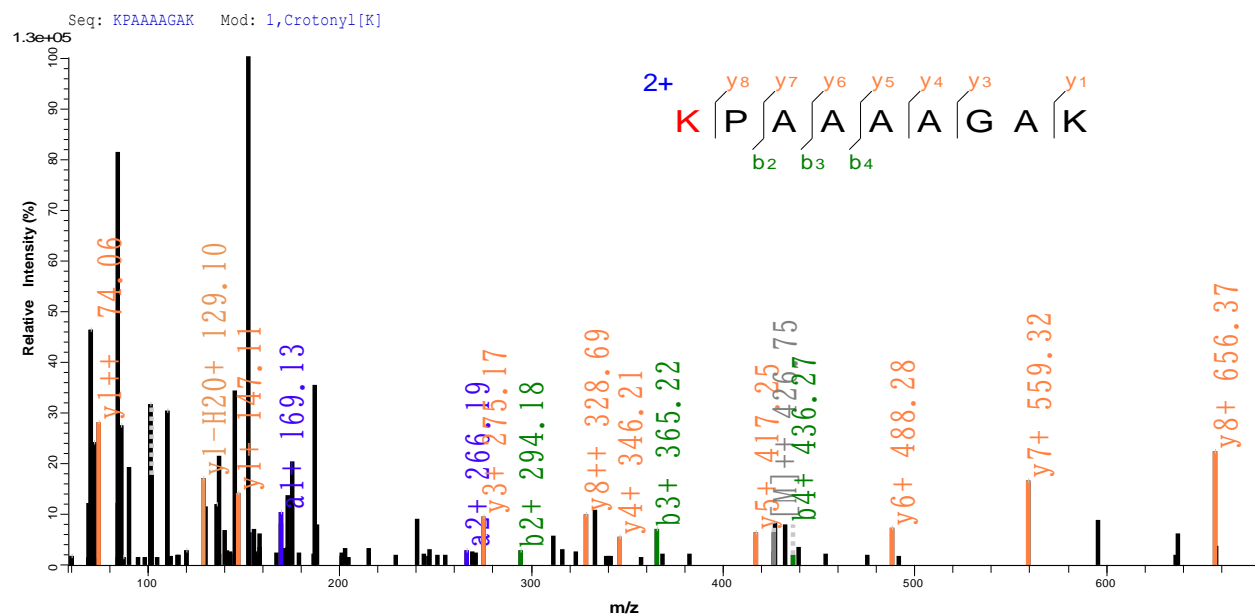


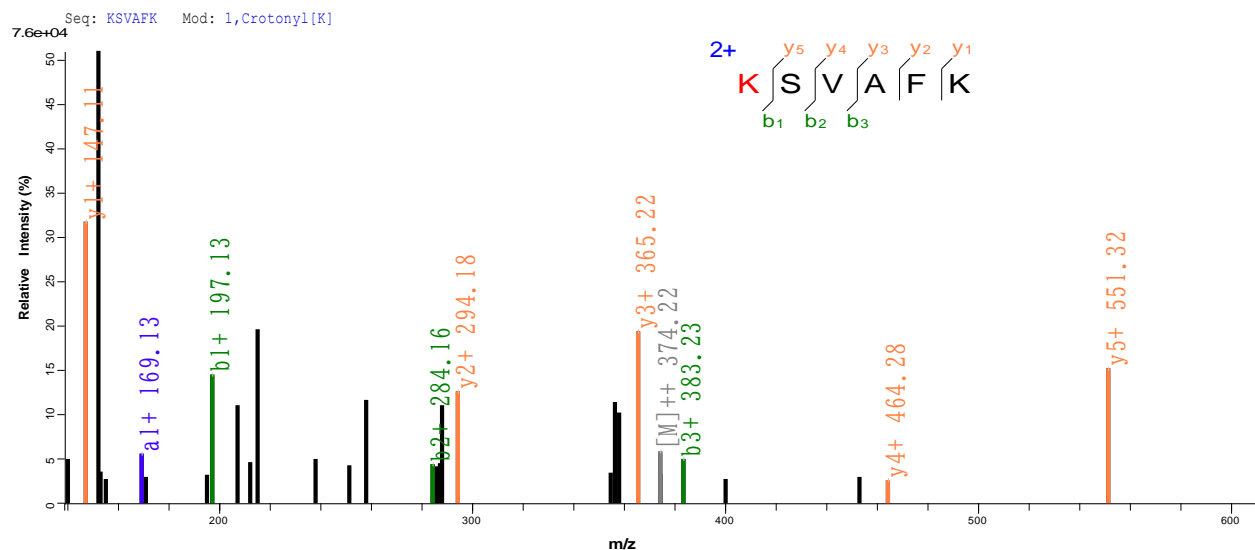
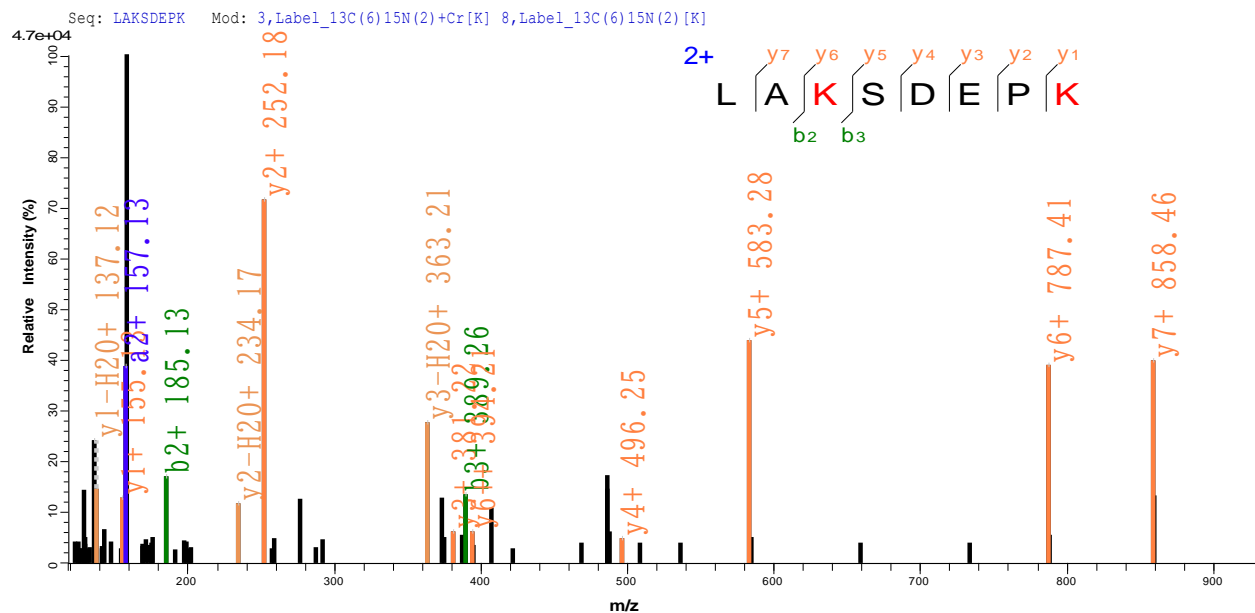


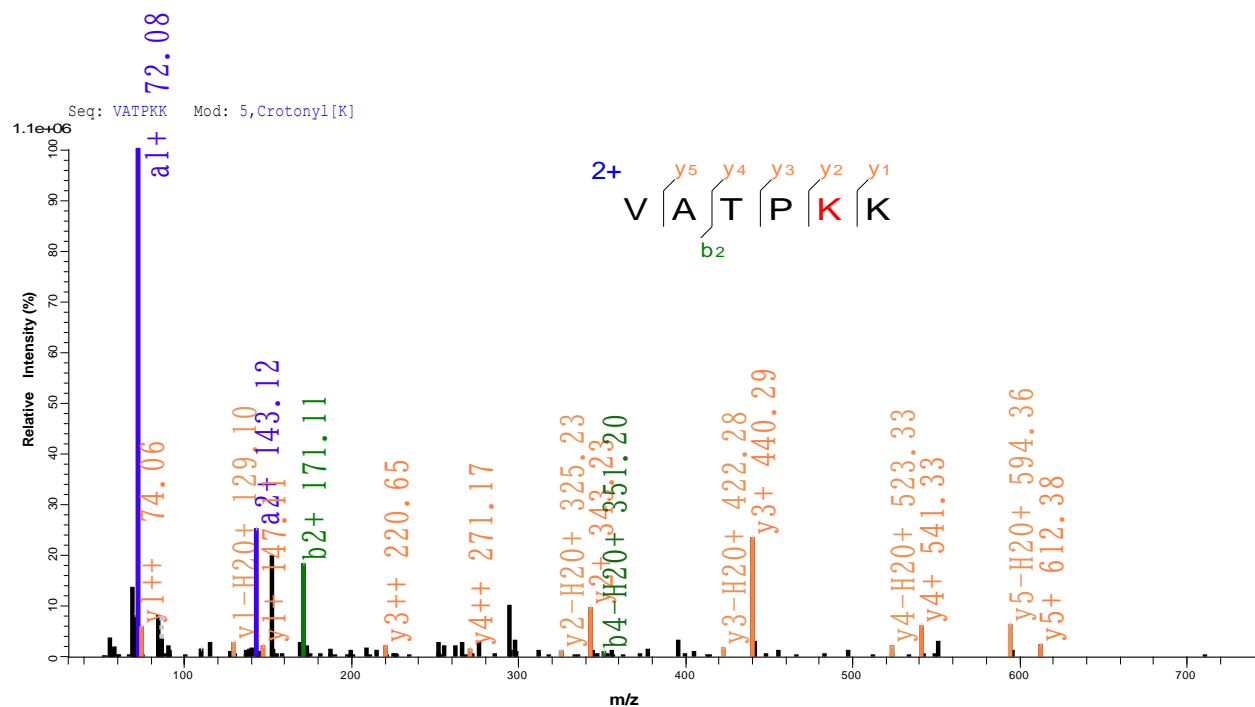
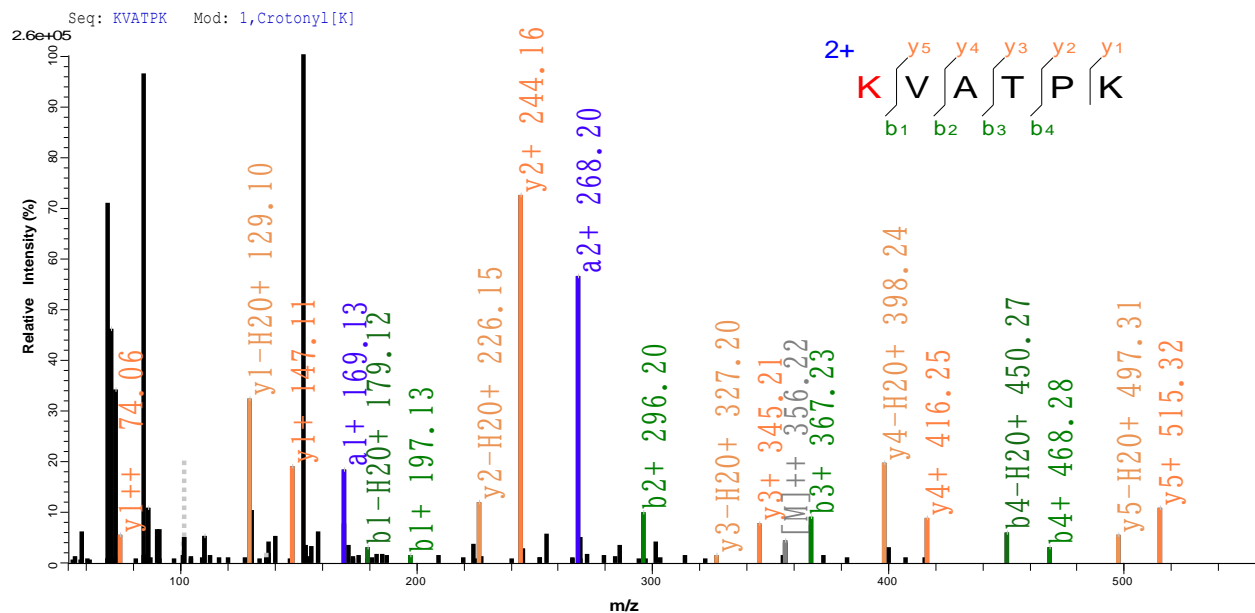


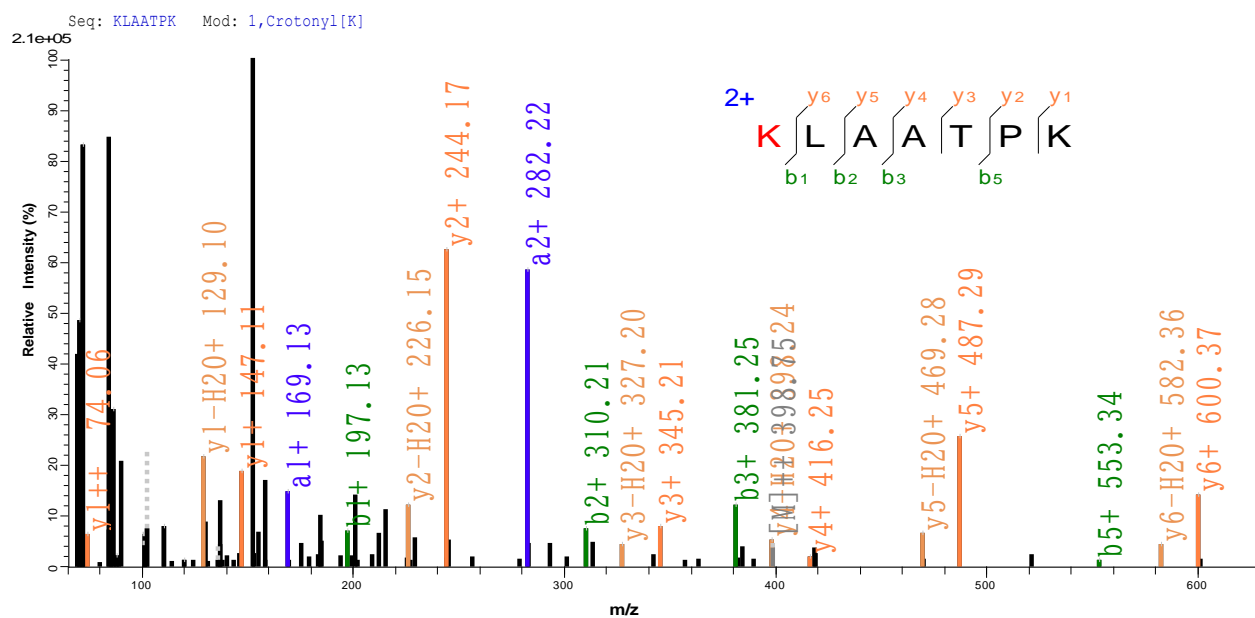
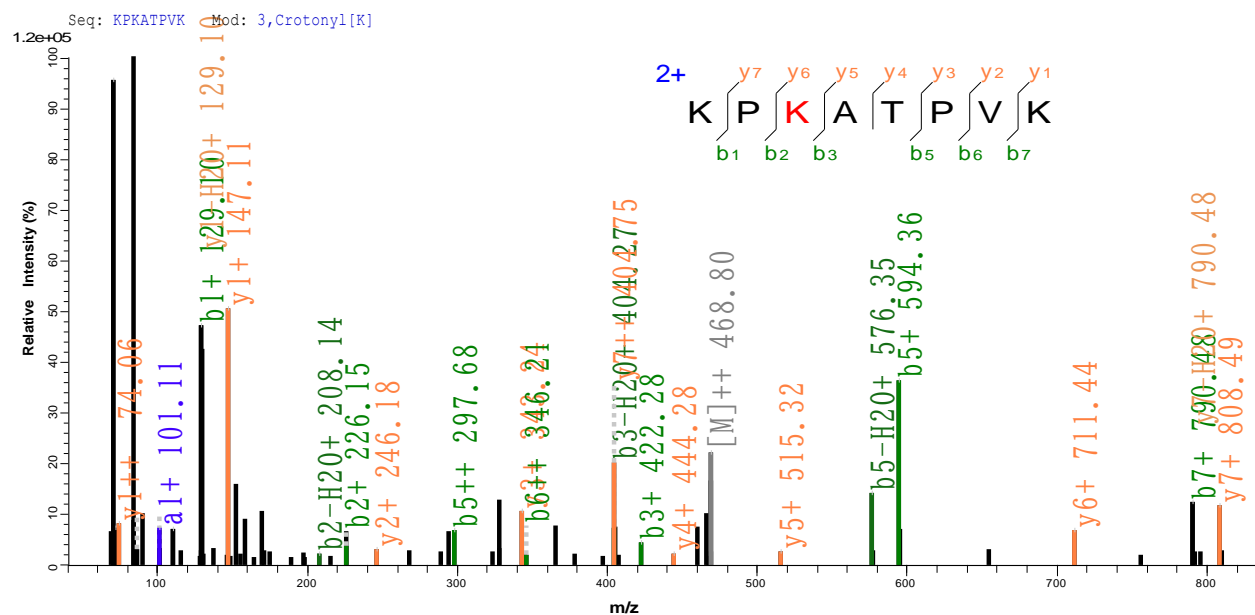


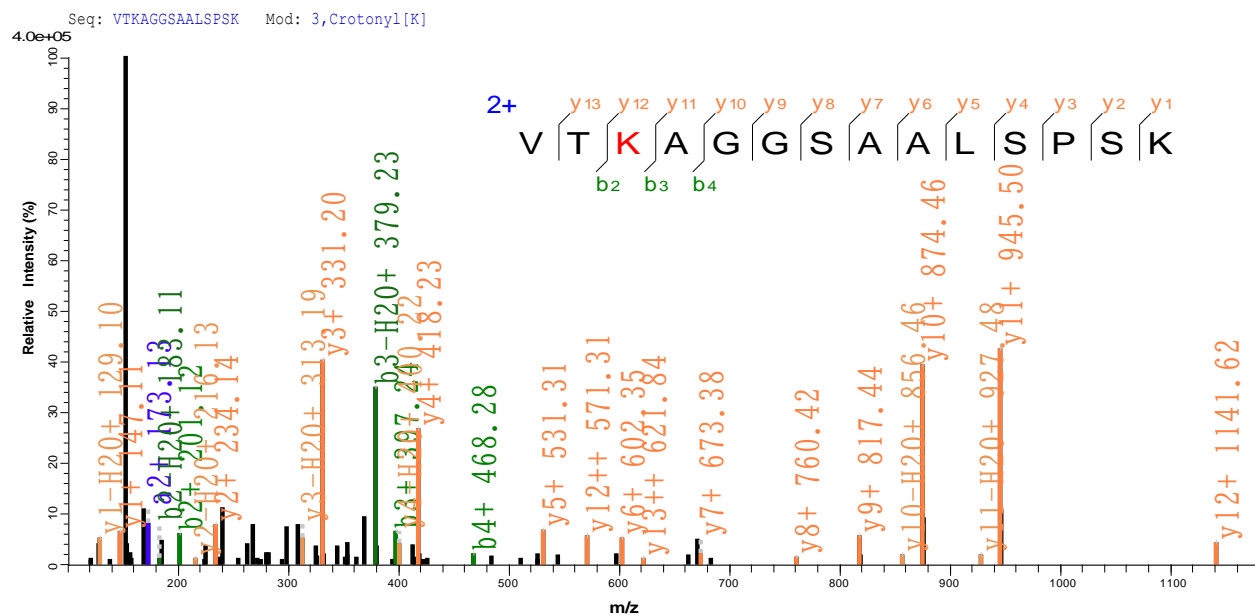
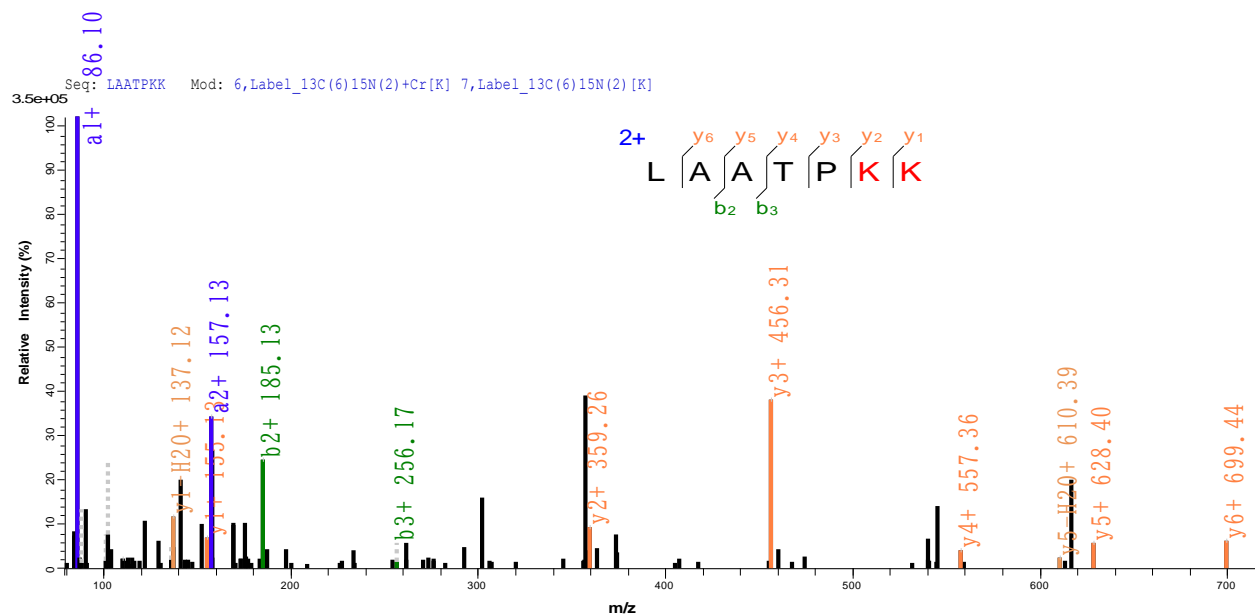


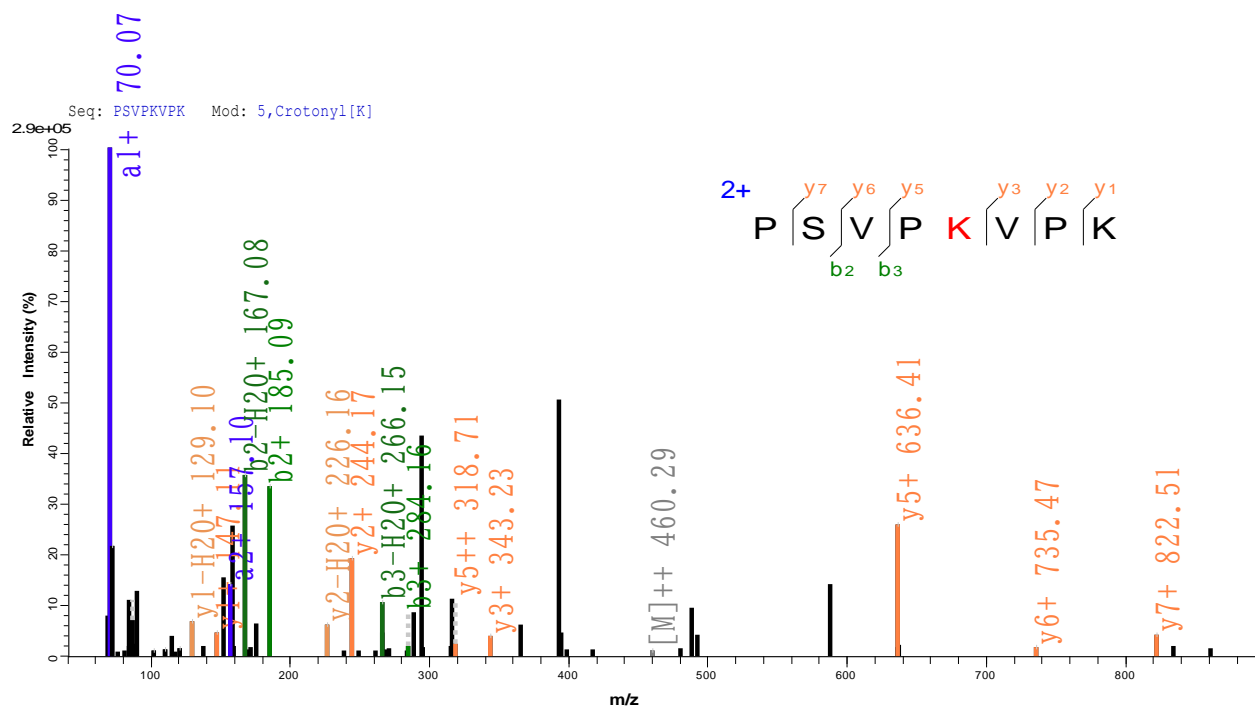
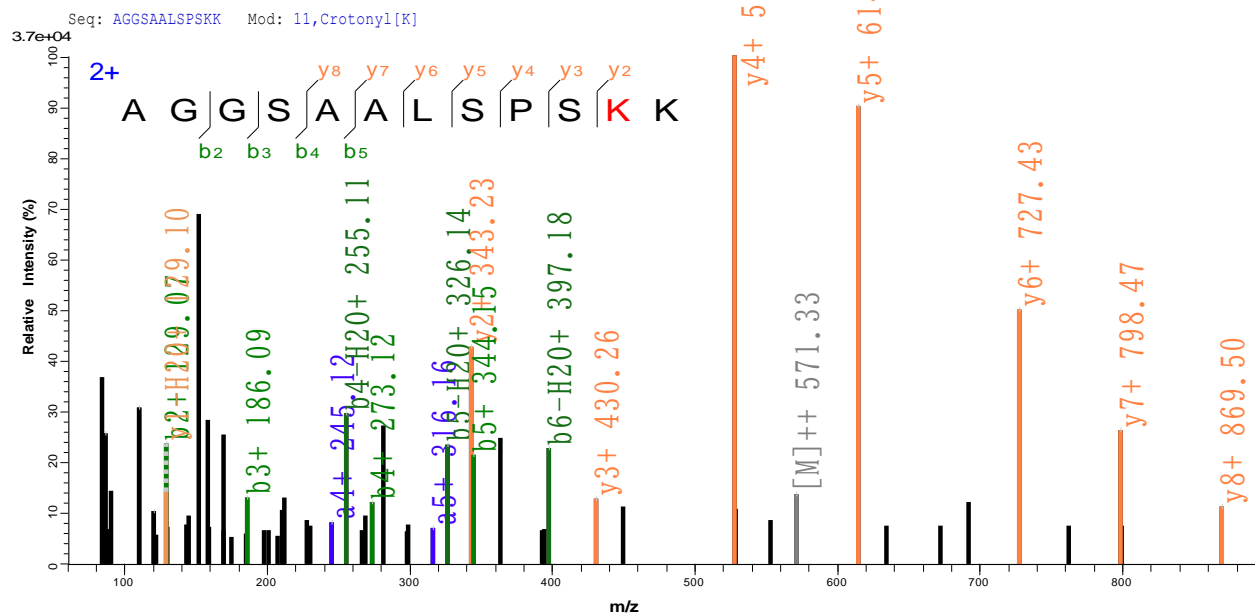




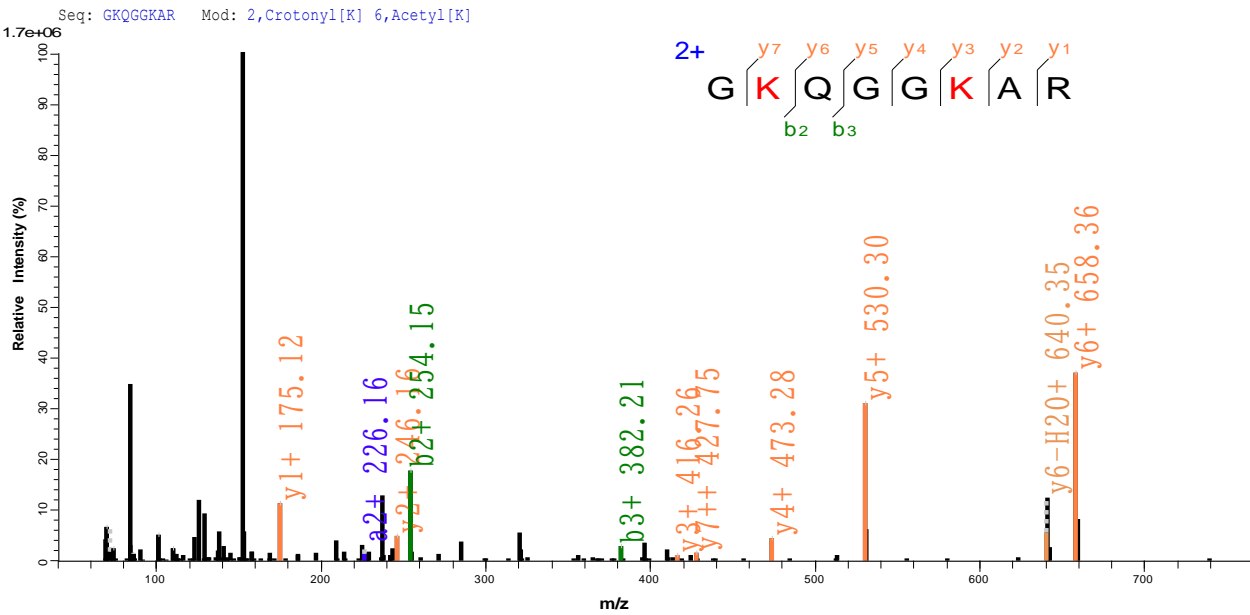
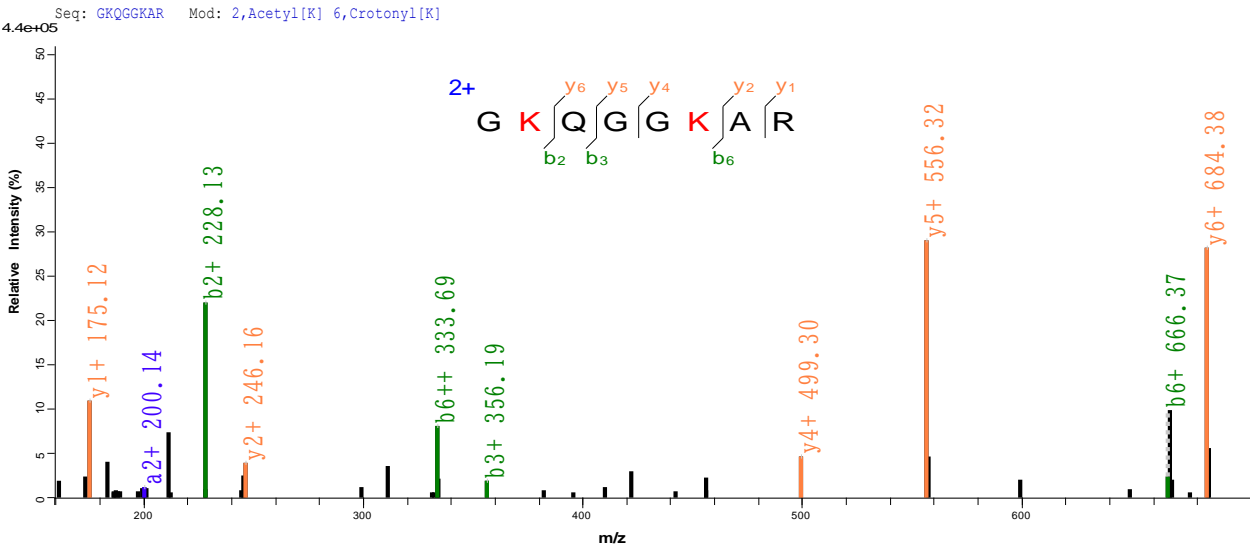


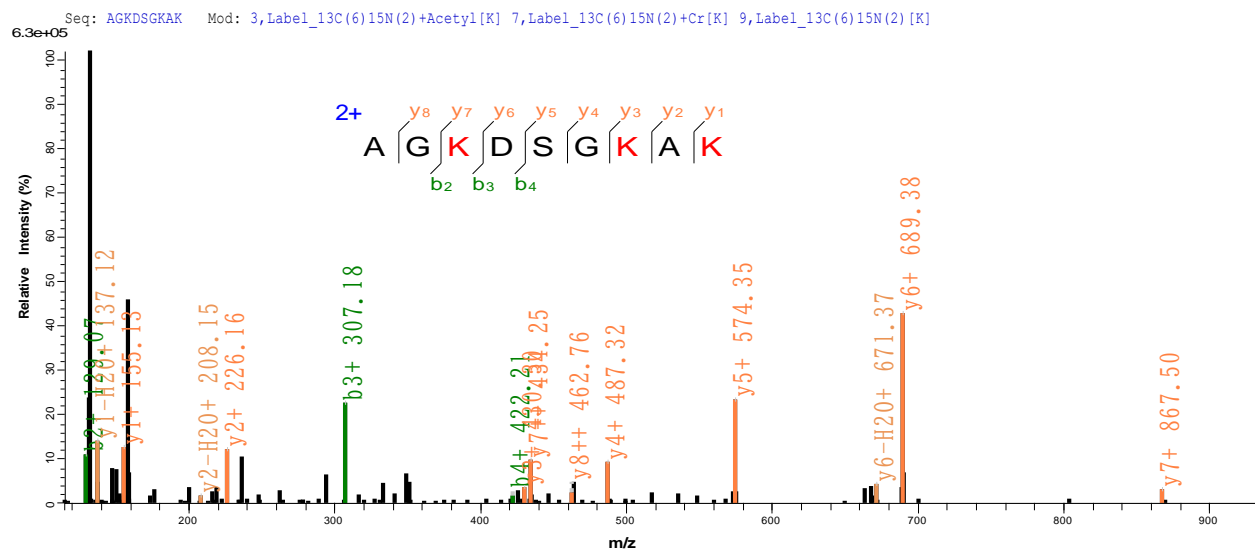
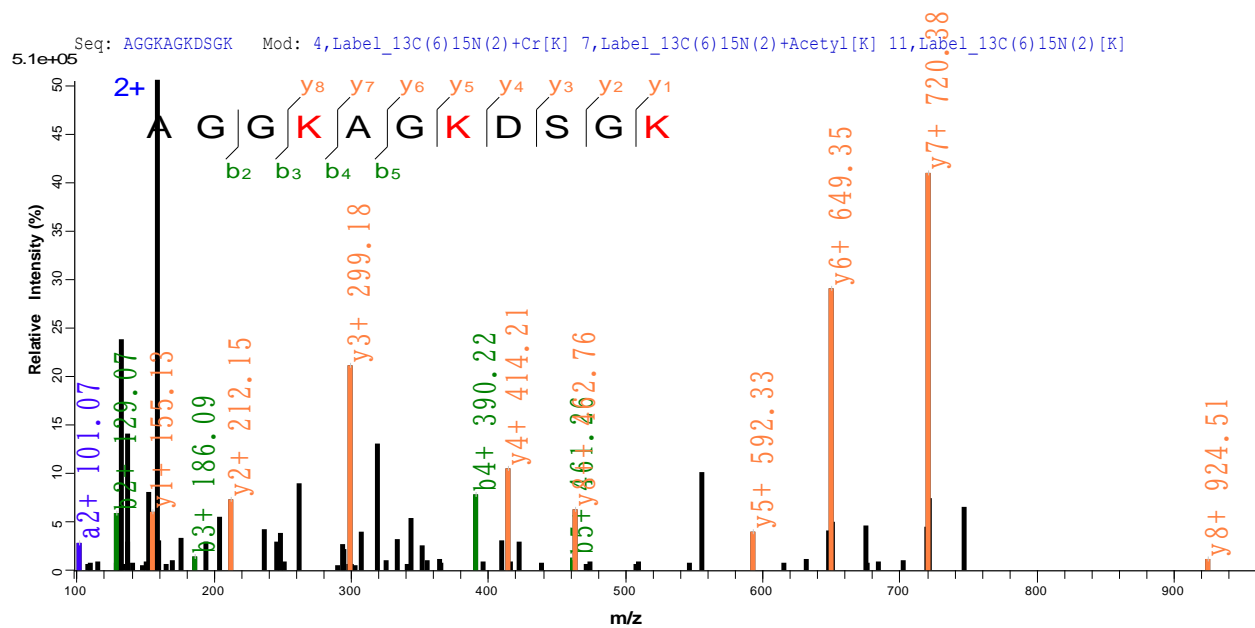




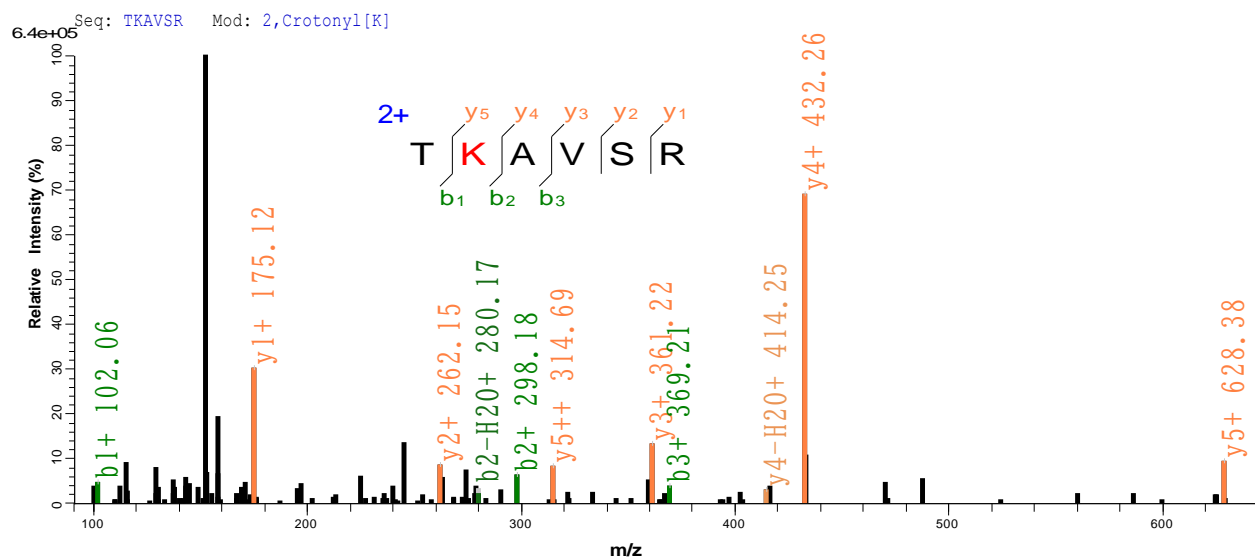
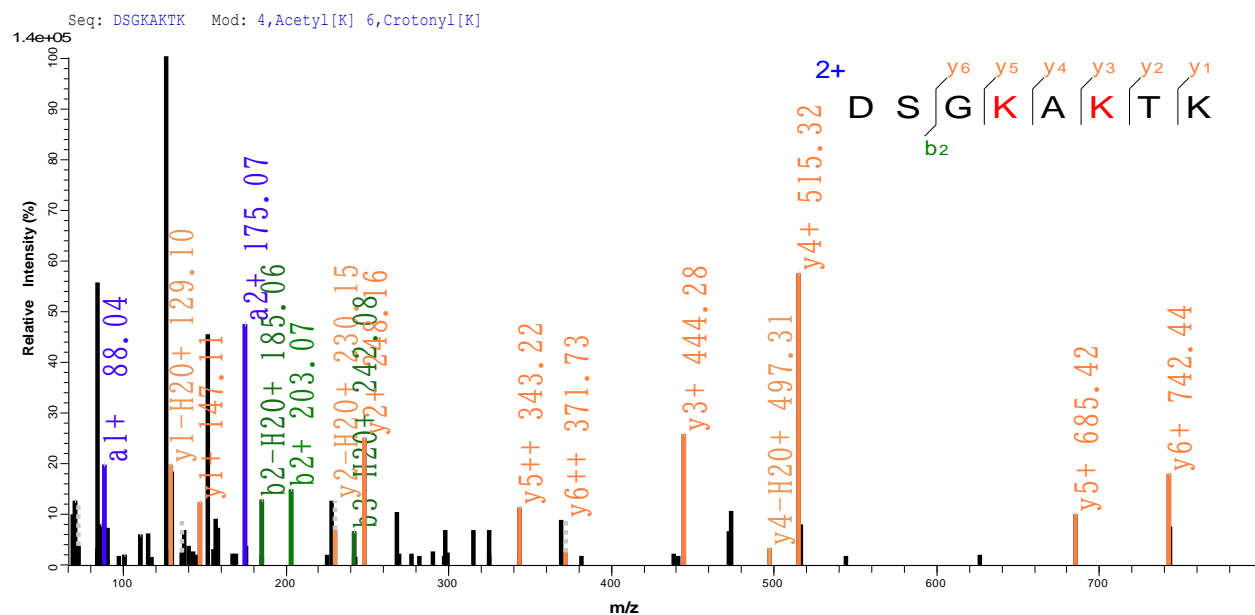


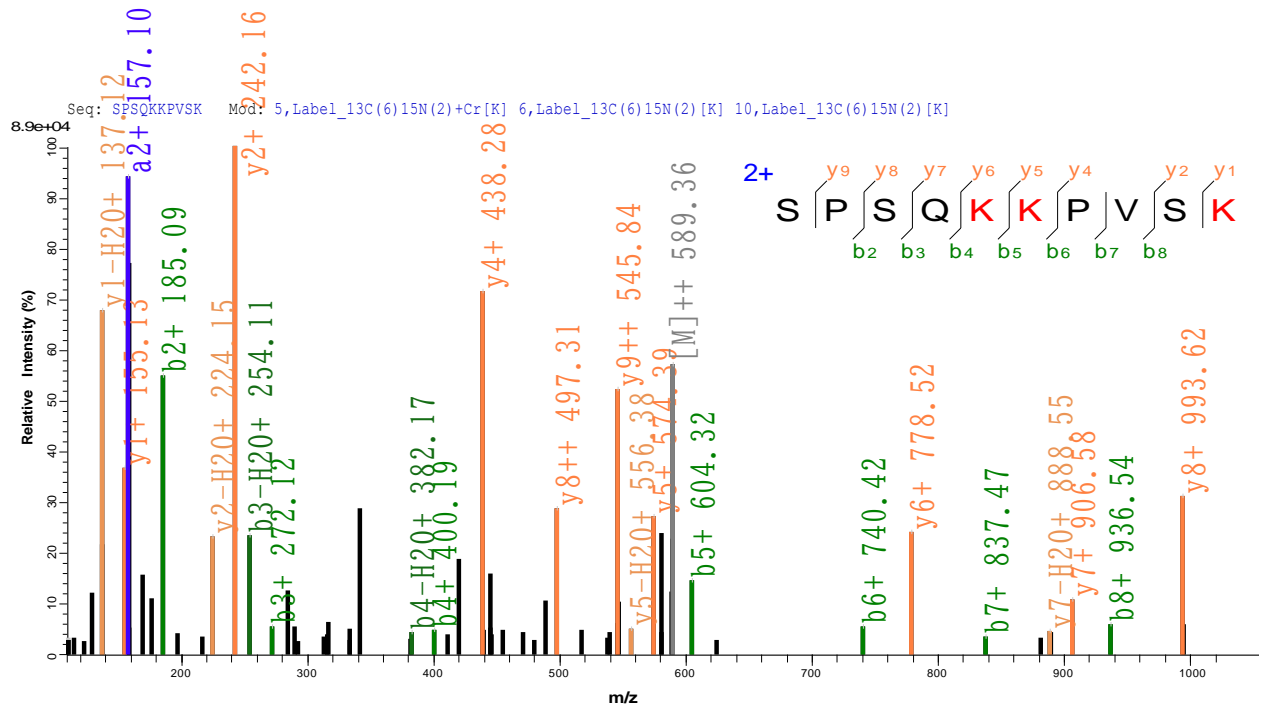
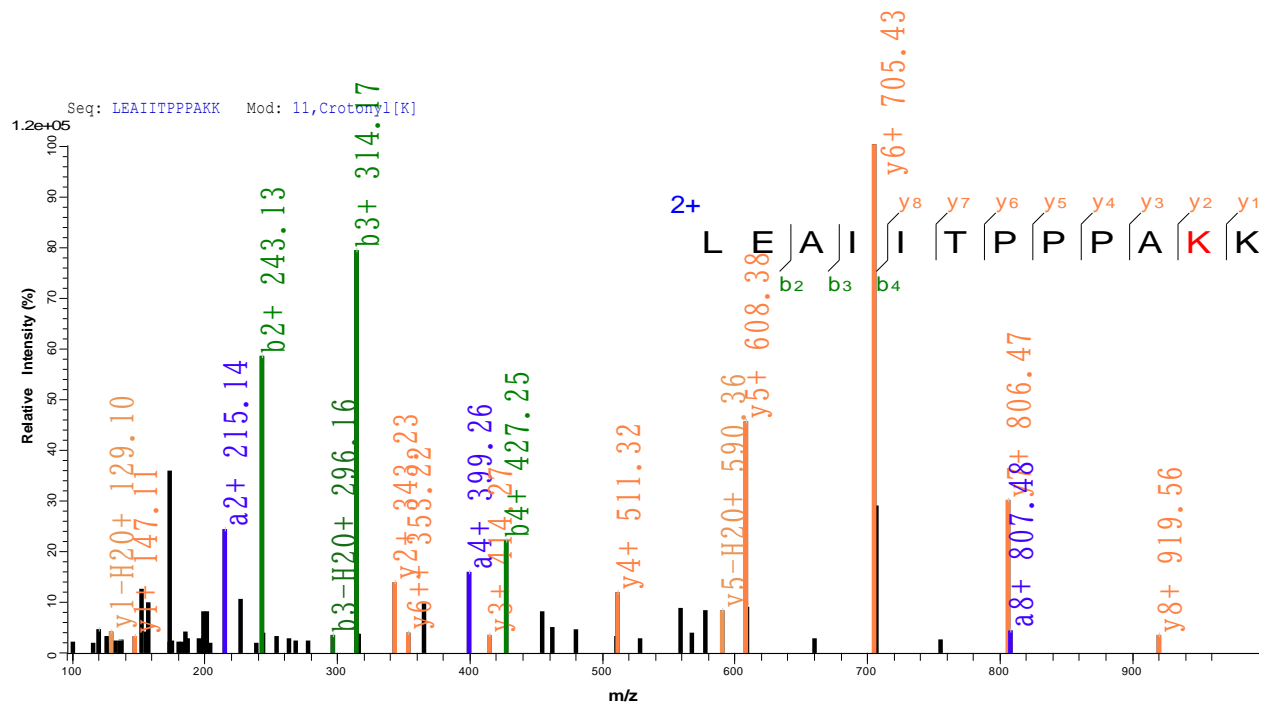
H2A

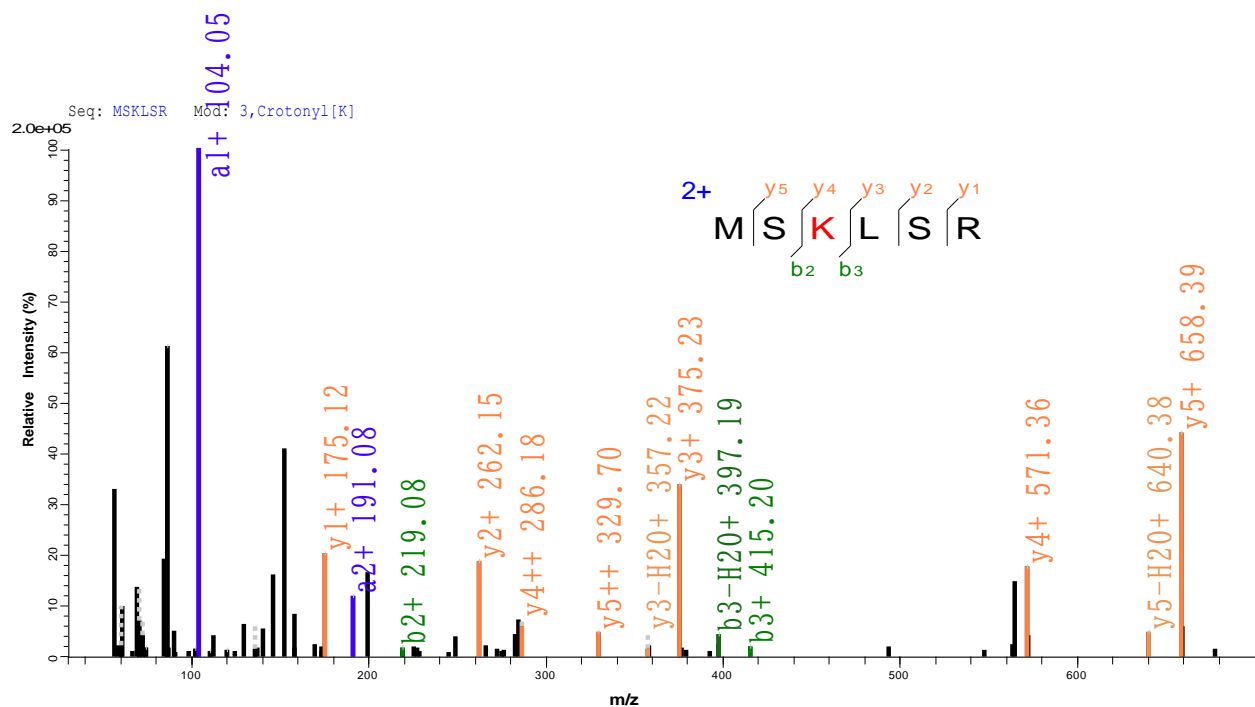
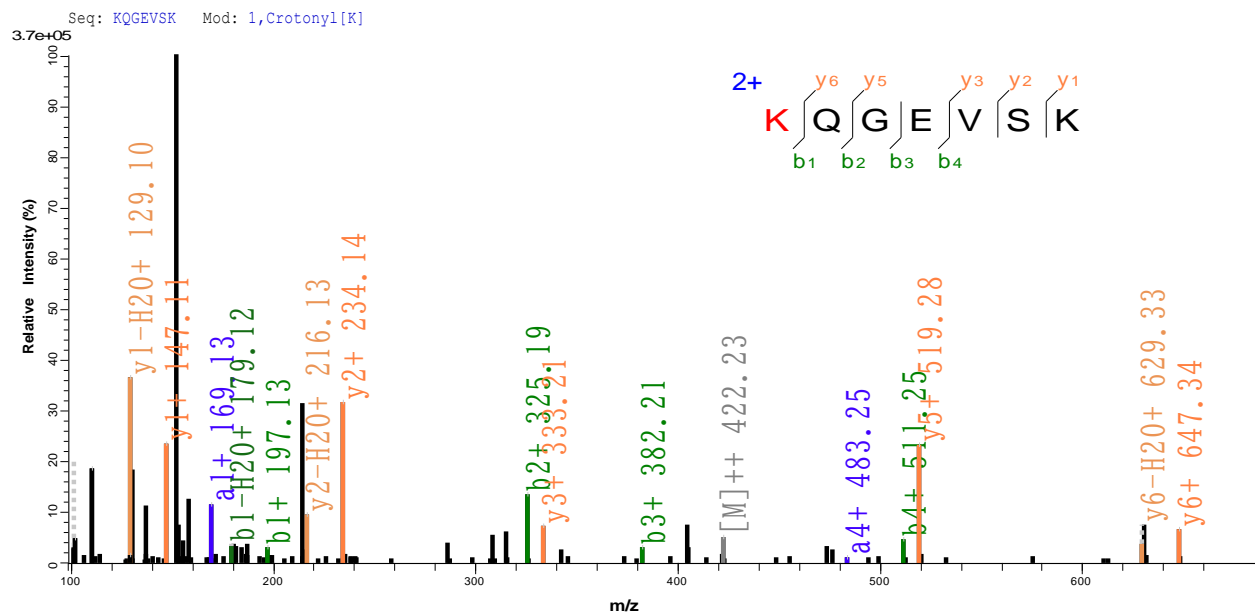


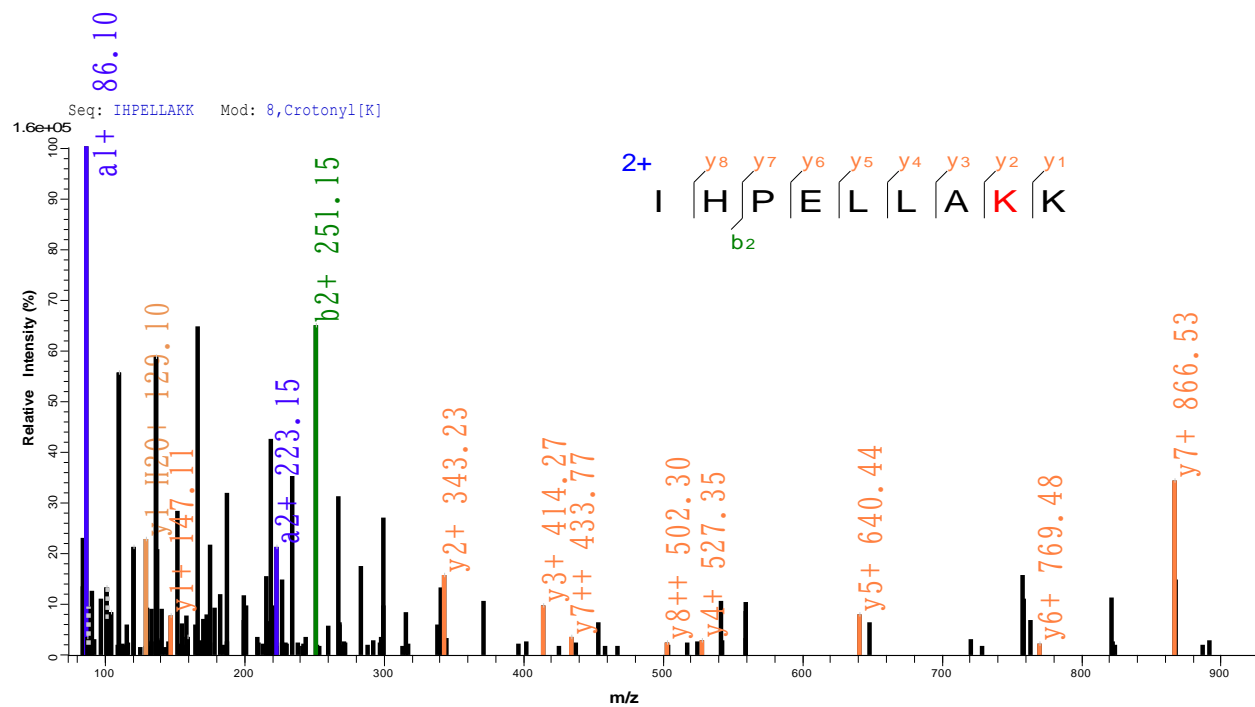




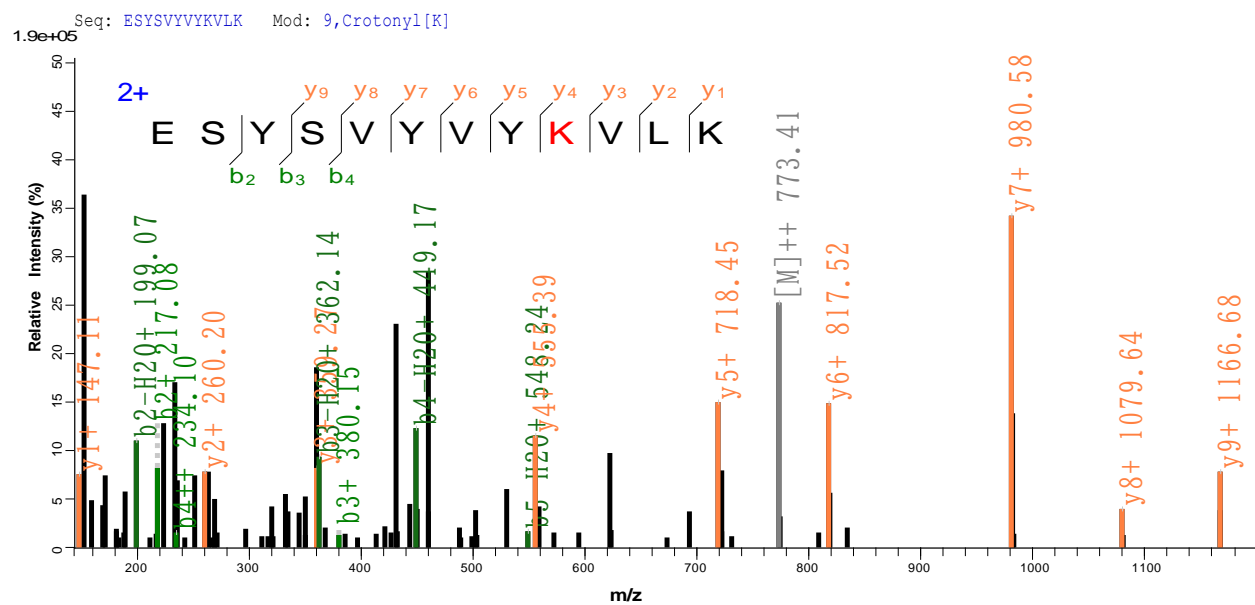


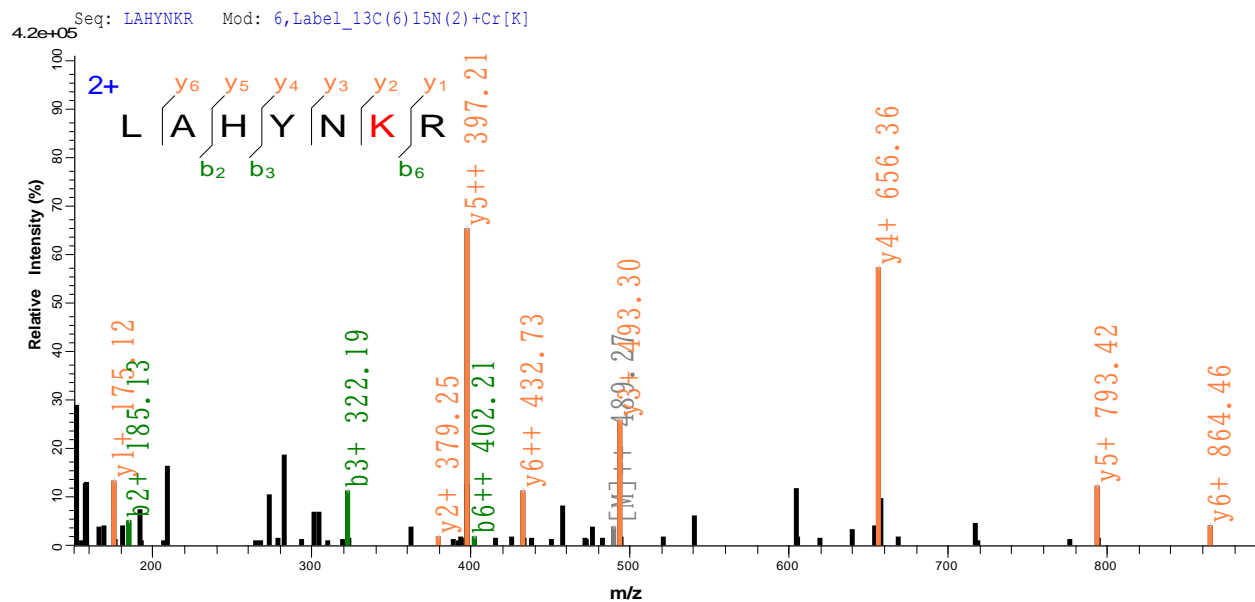
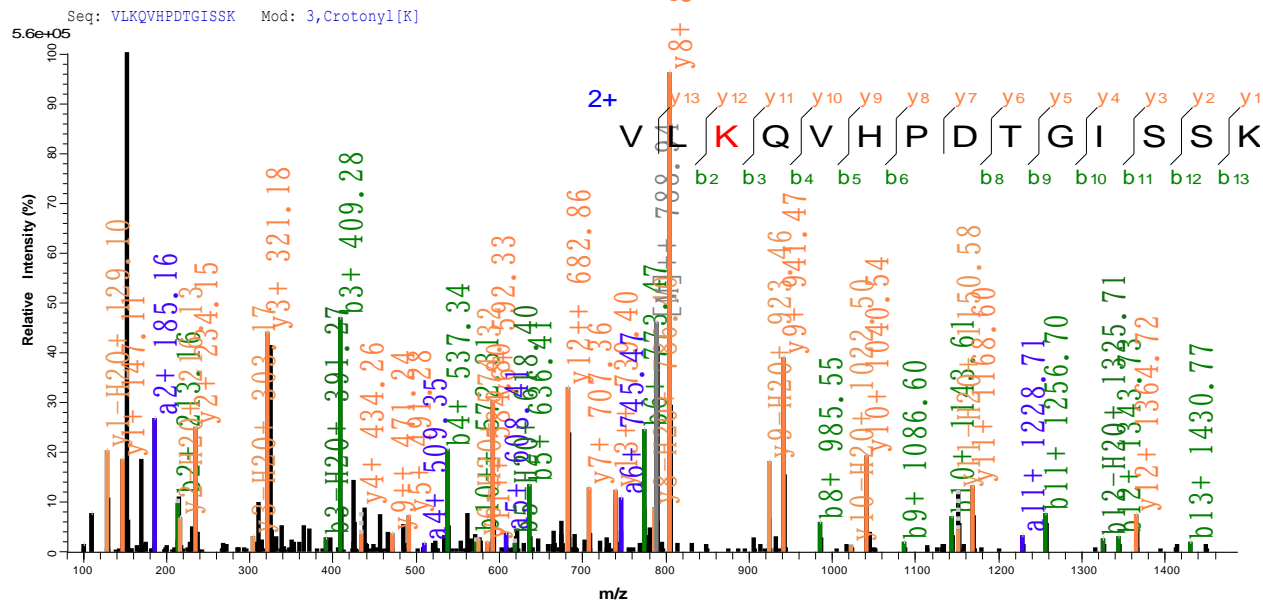


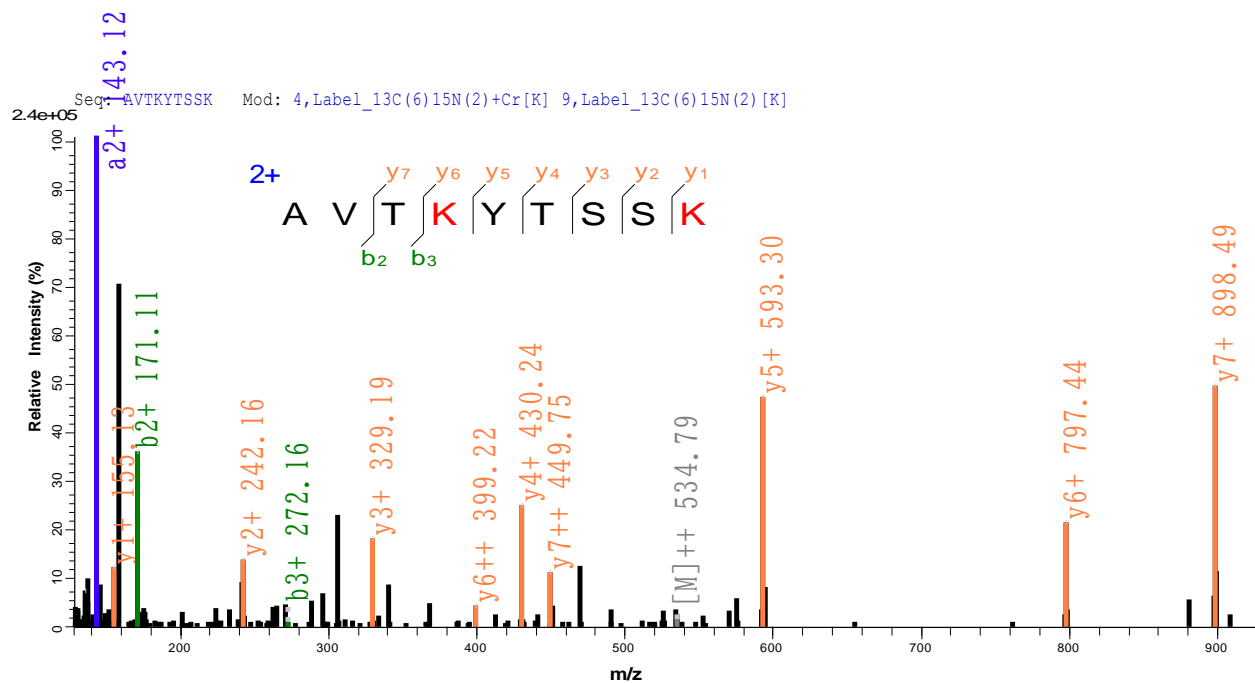




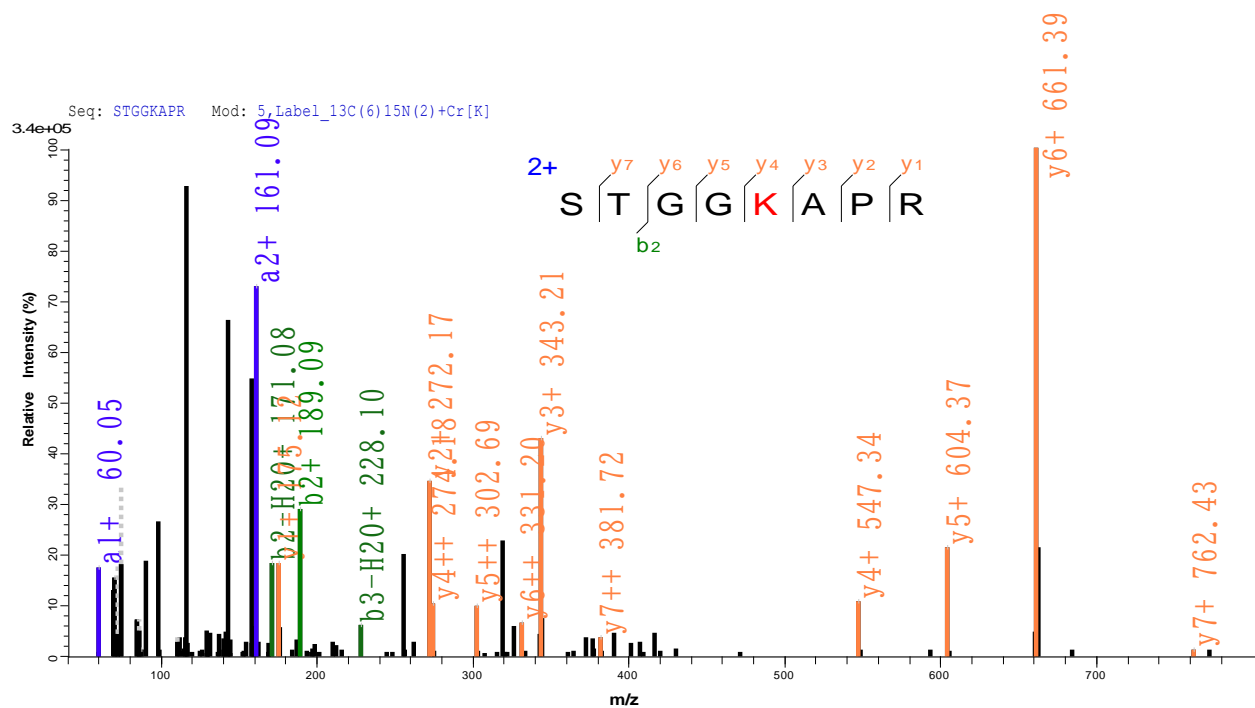
## H2B

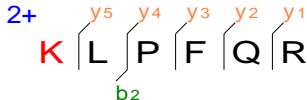


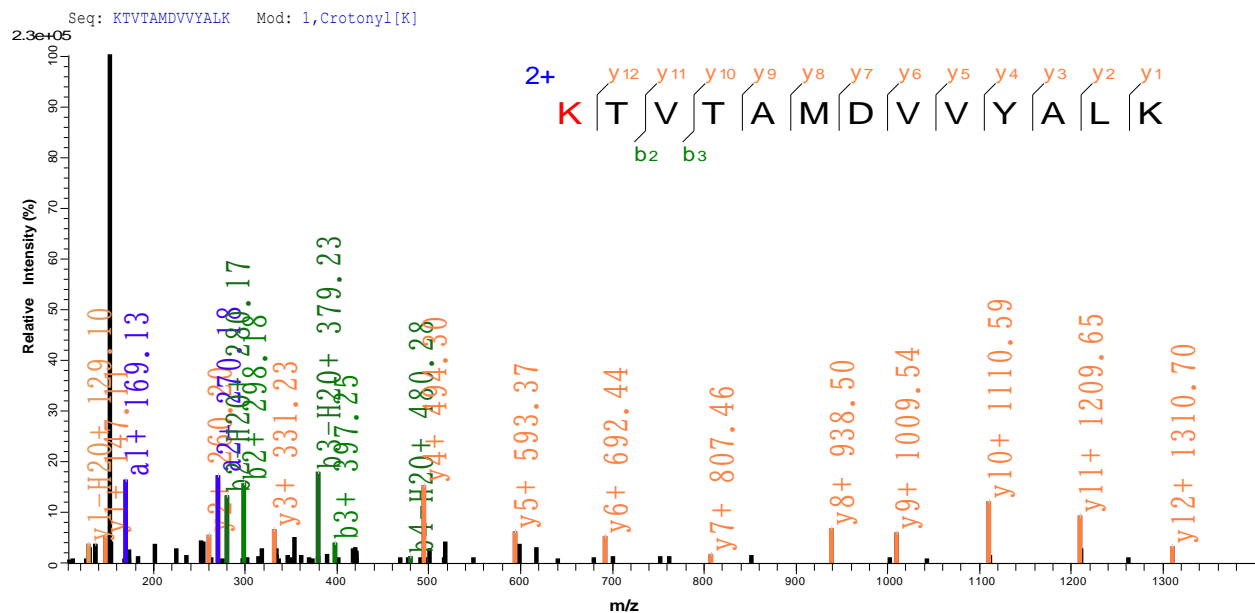




H3

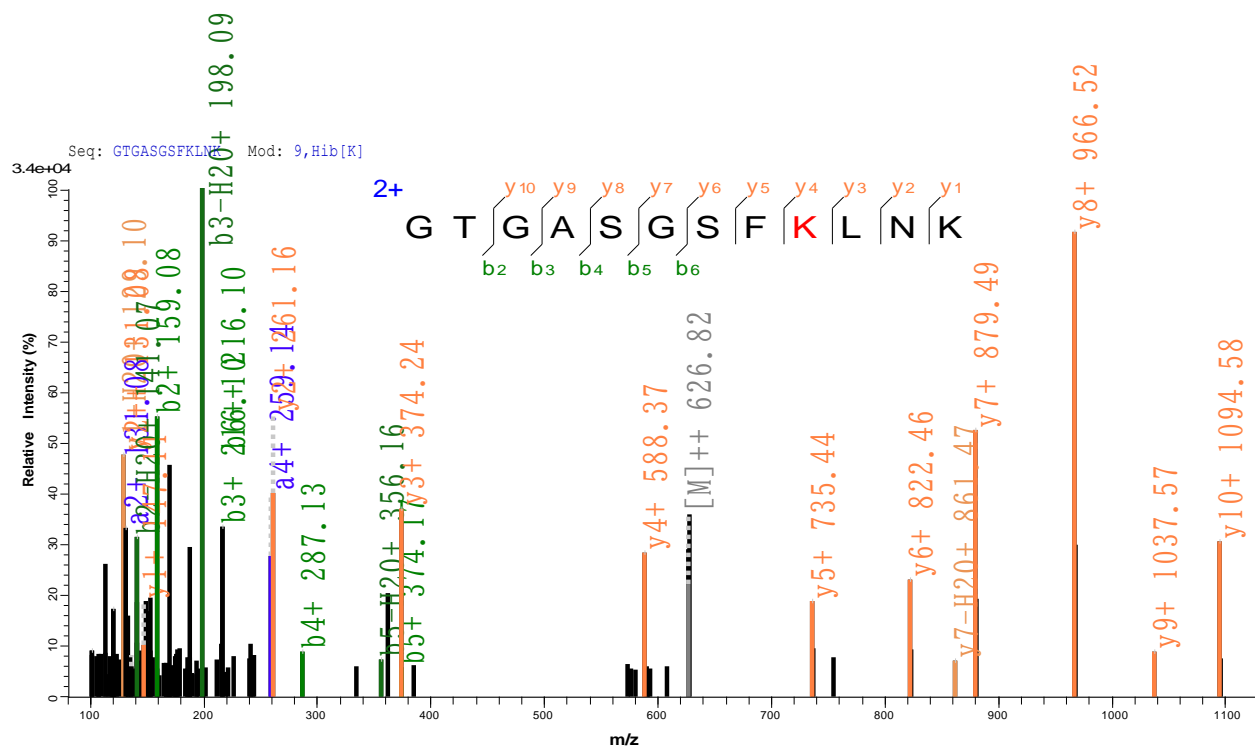




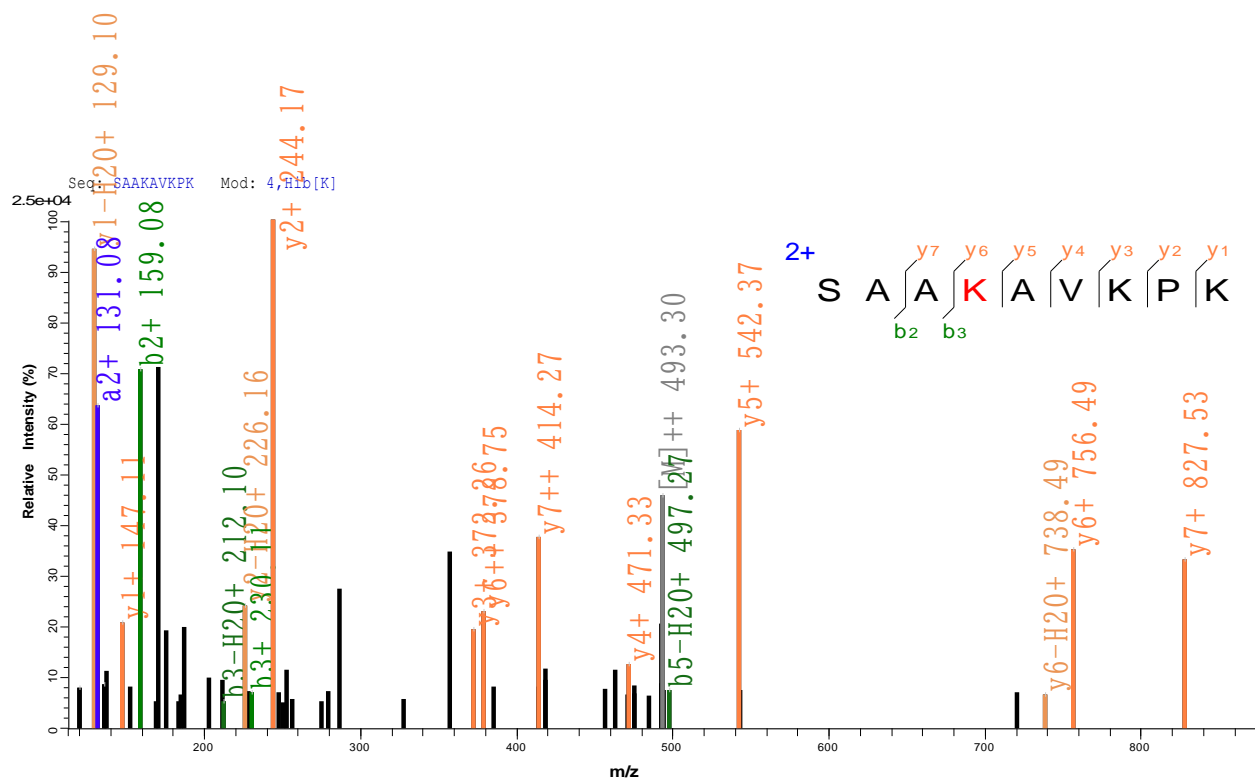
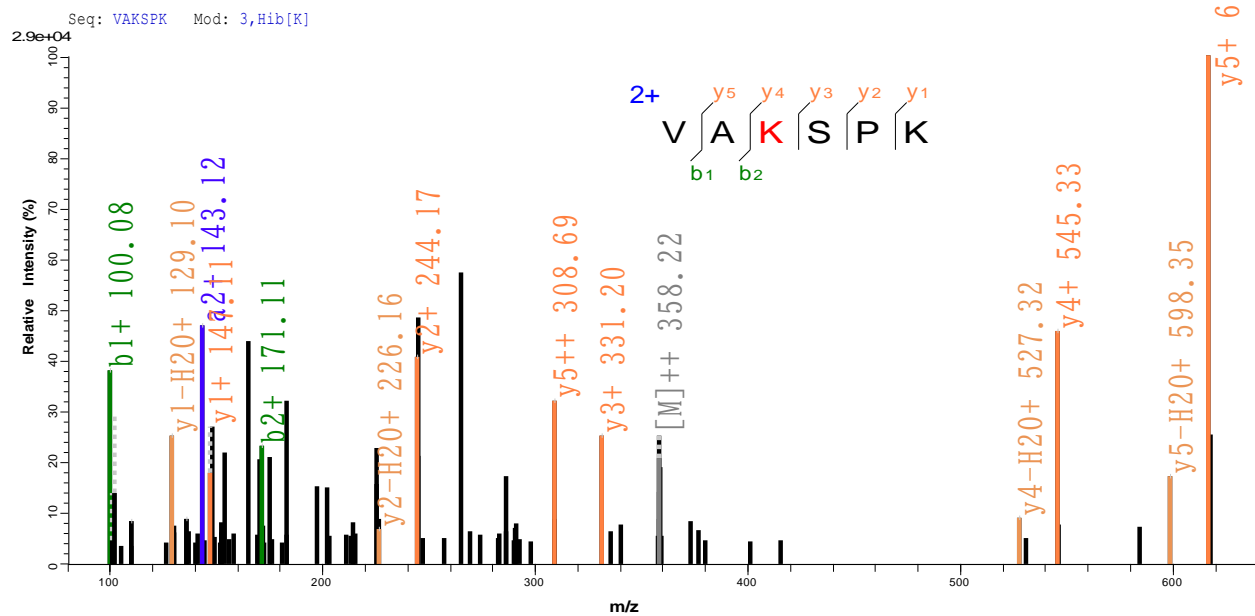


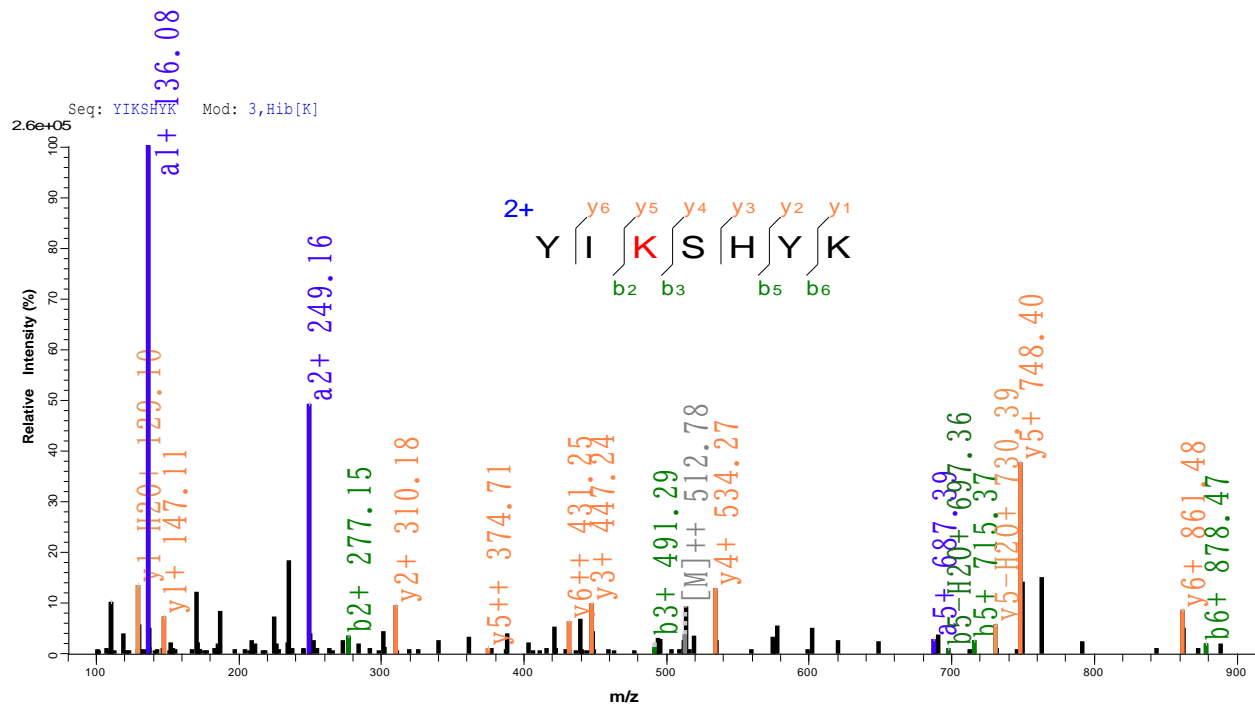
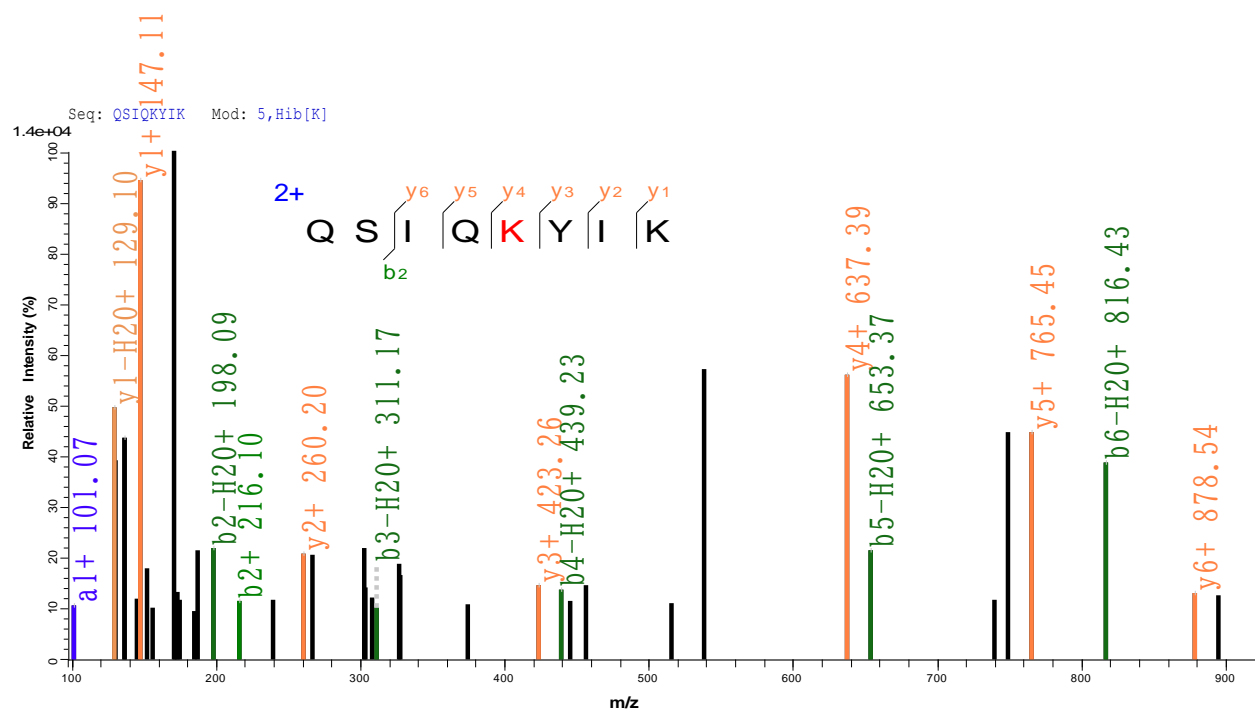
C. Khib

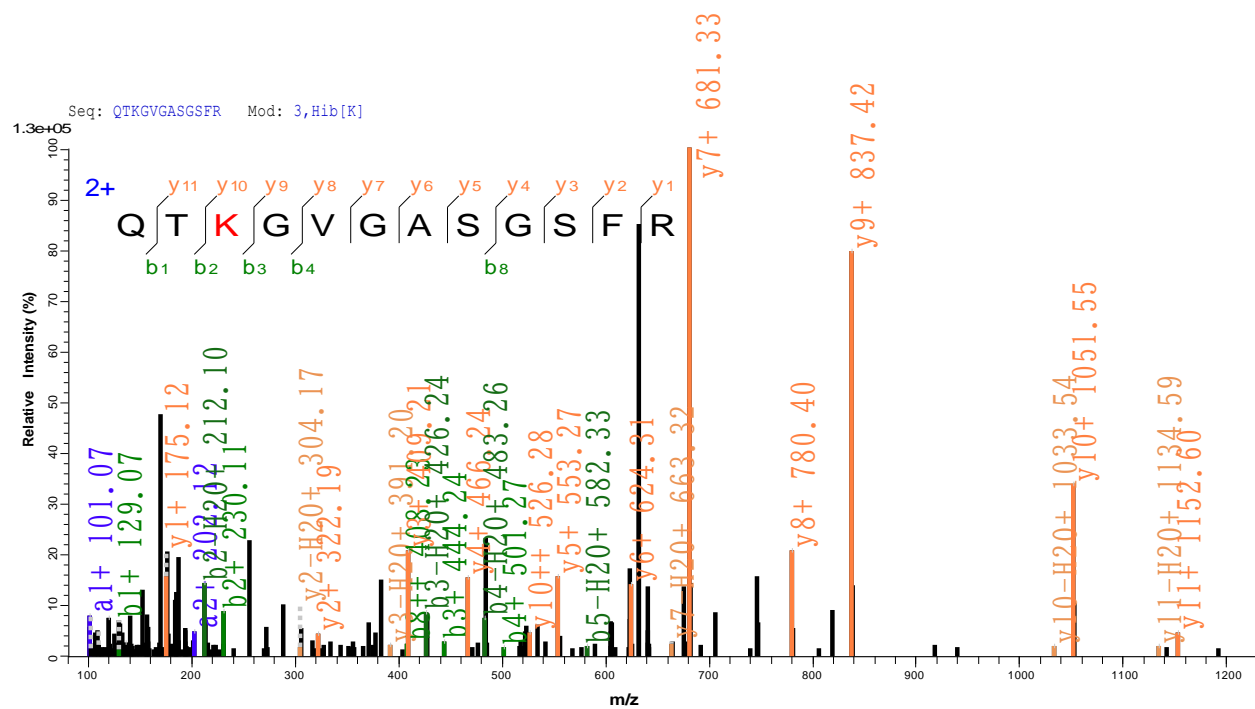
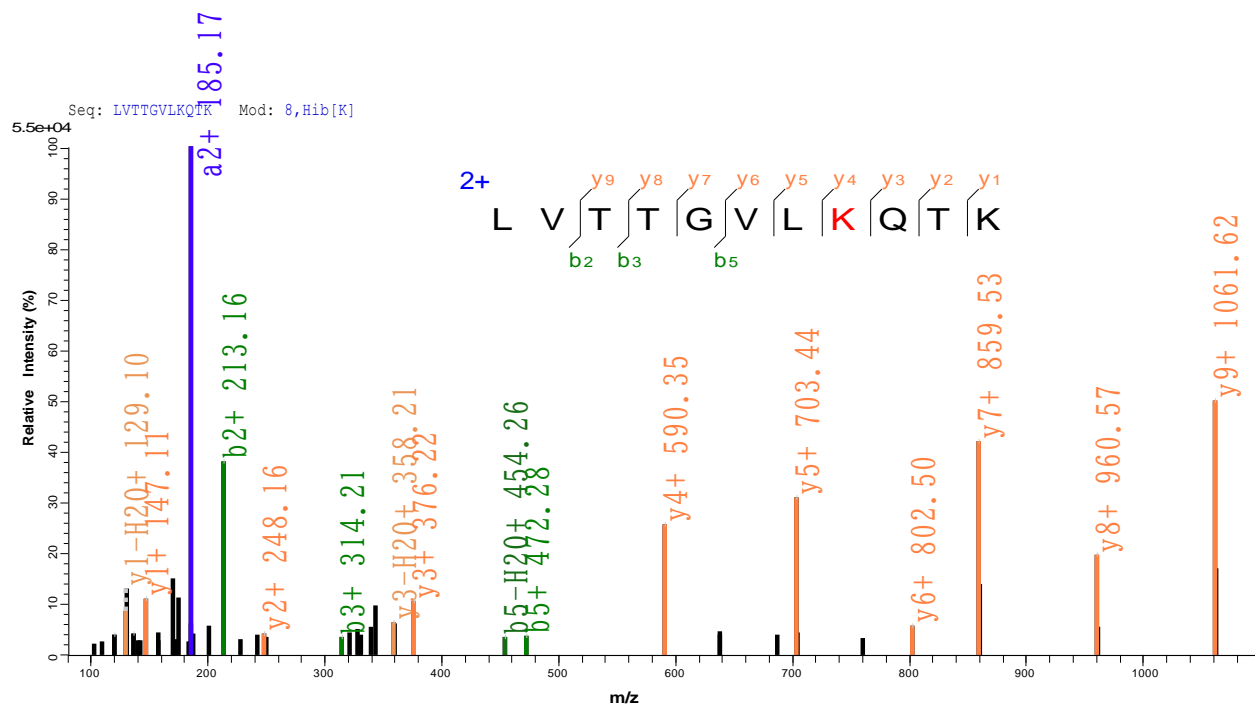
H1

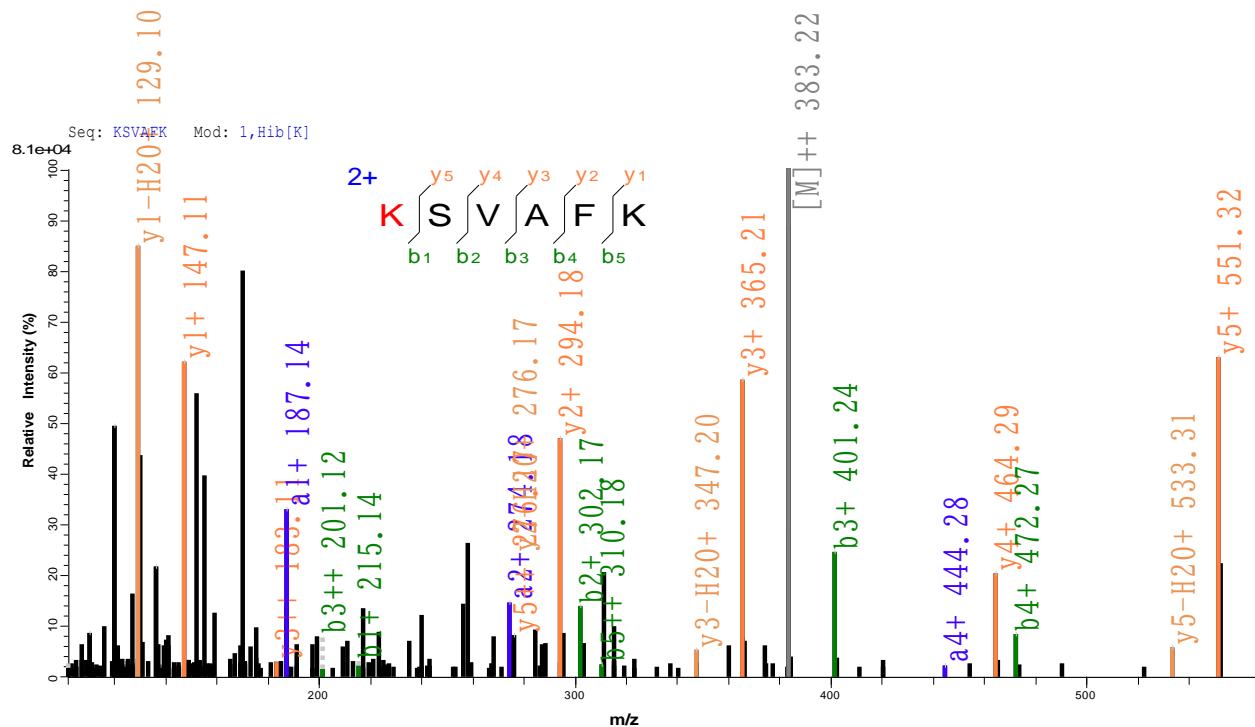
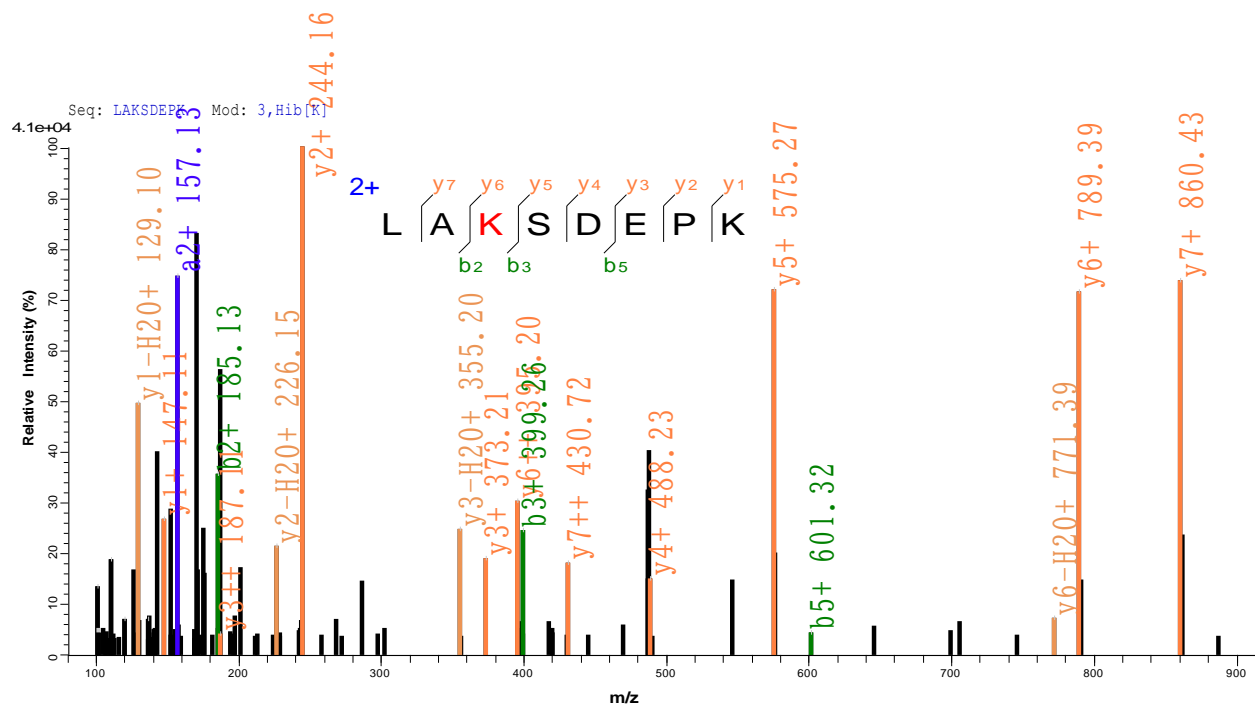


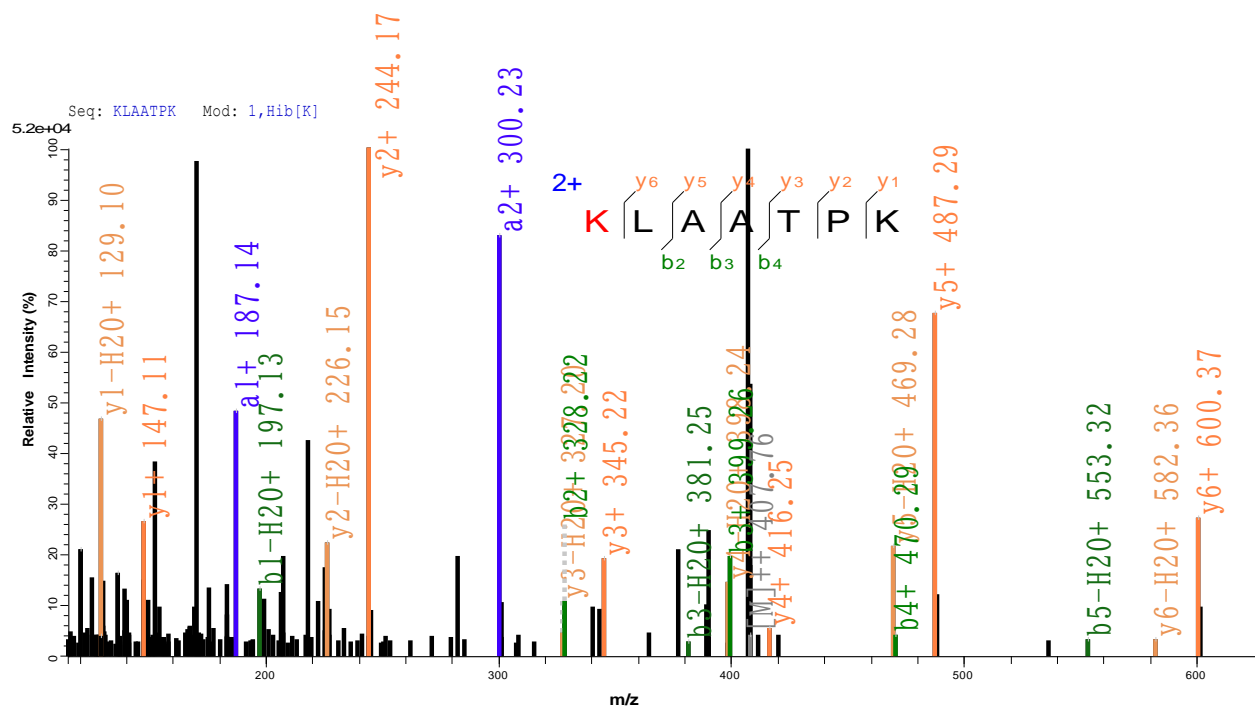
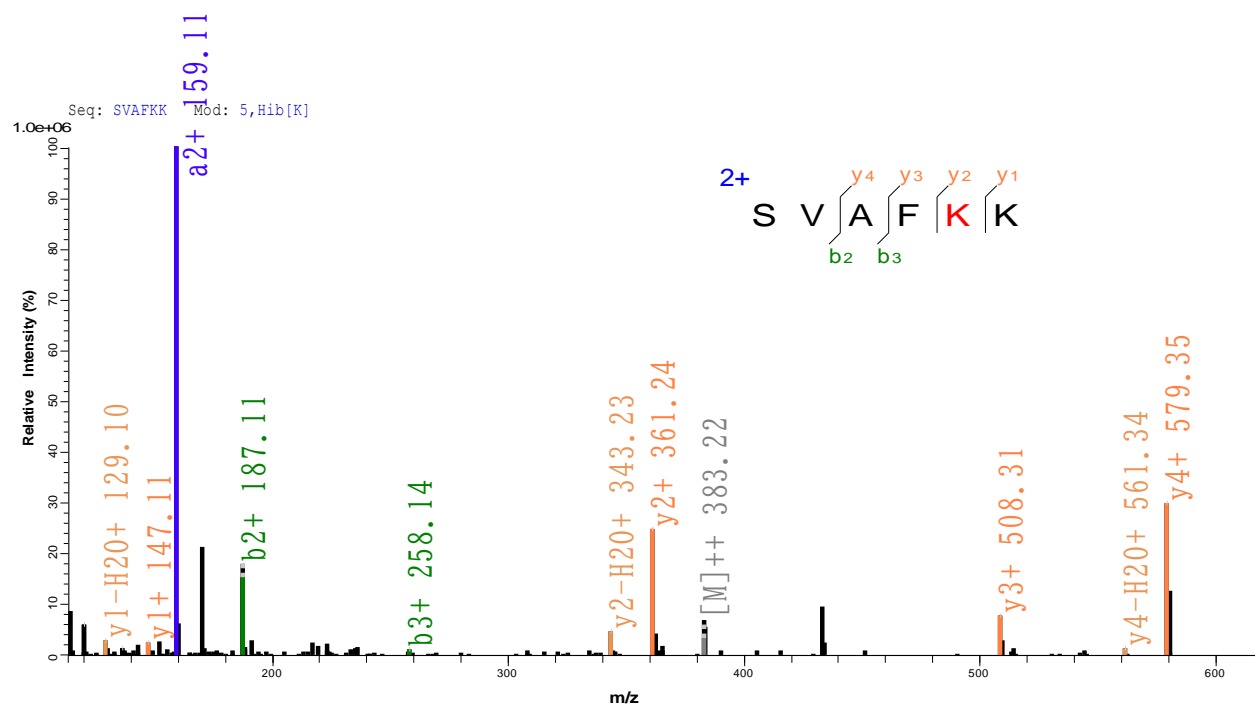


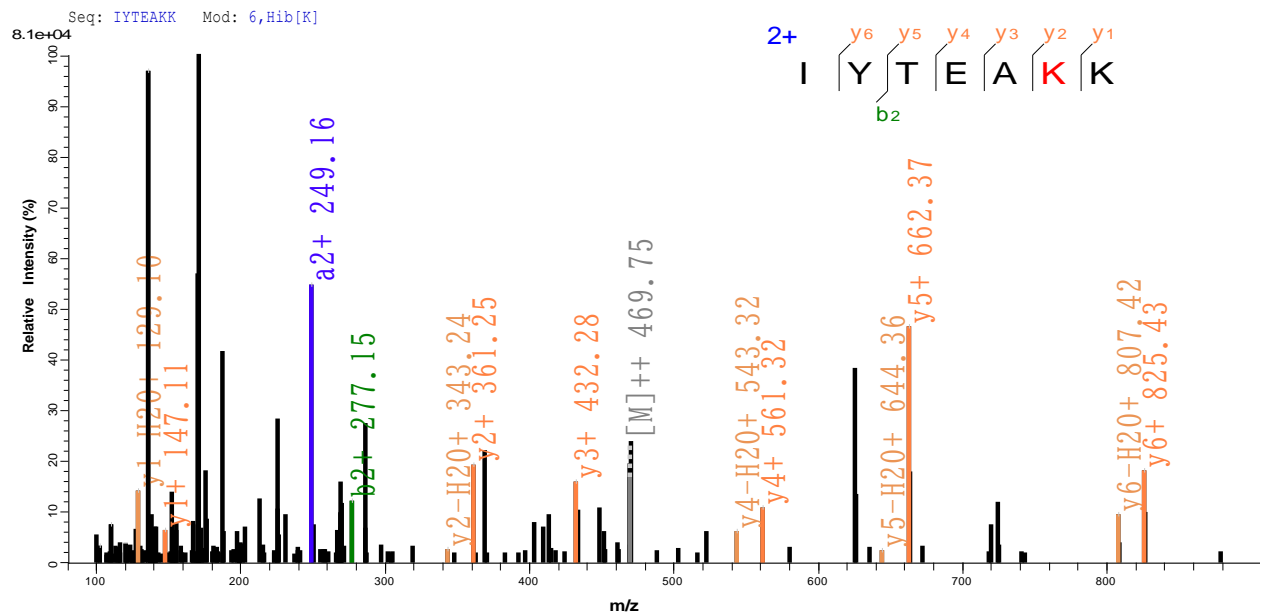
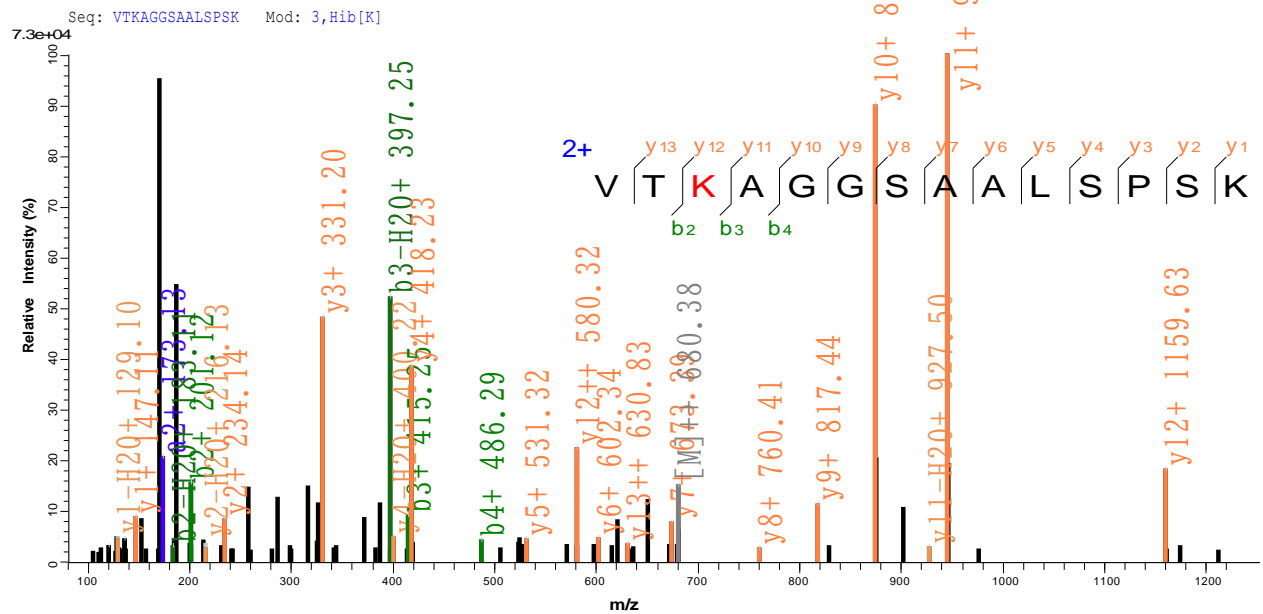




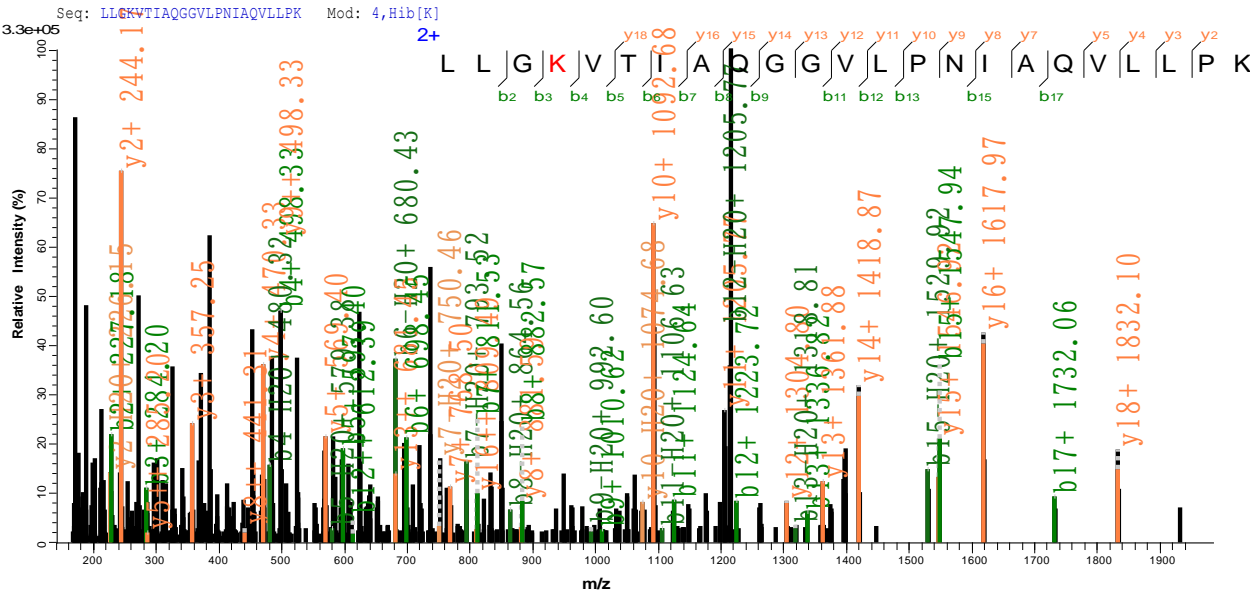


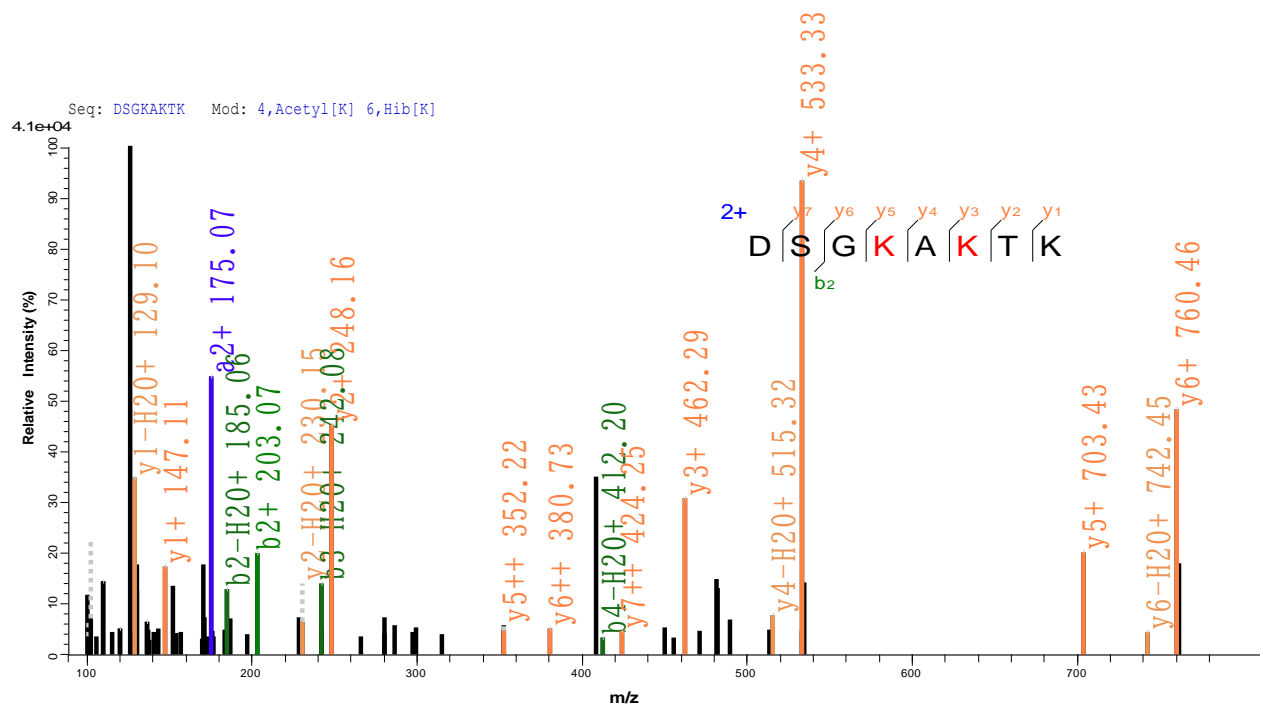
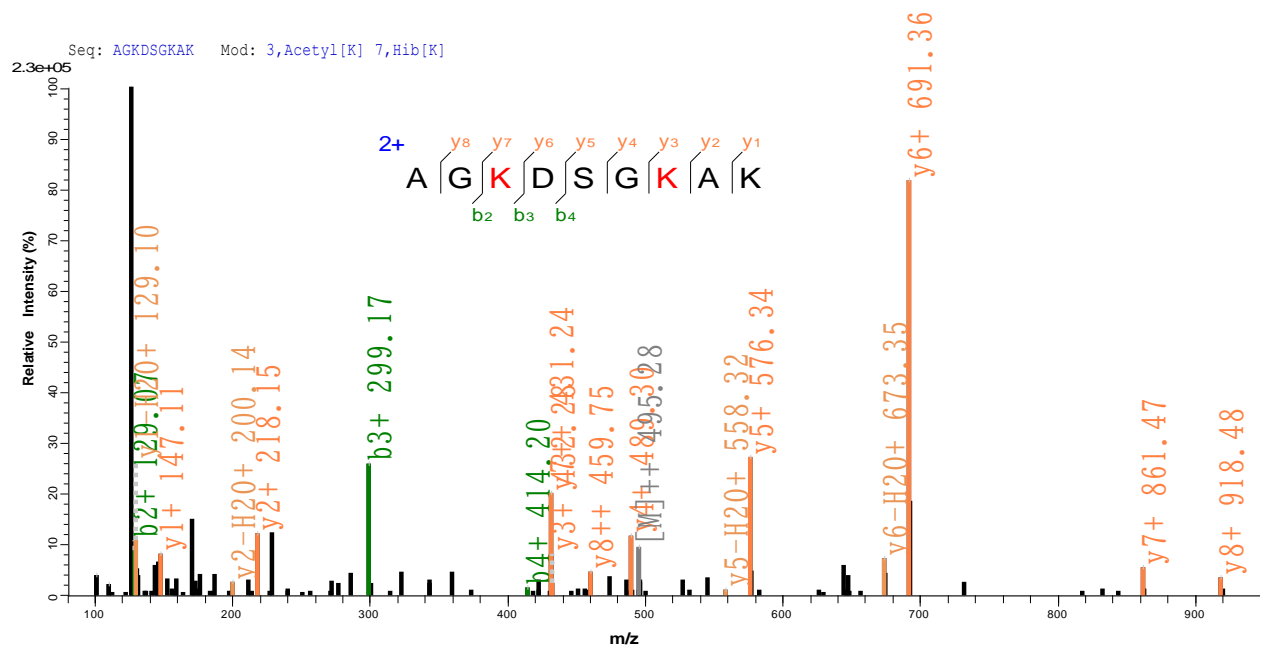




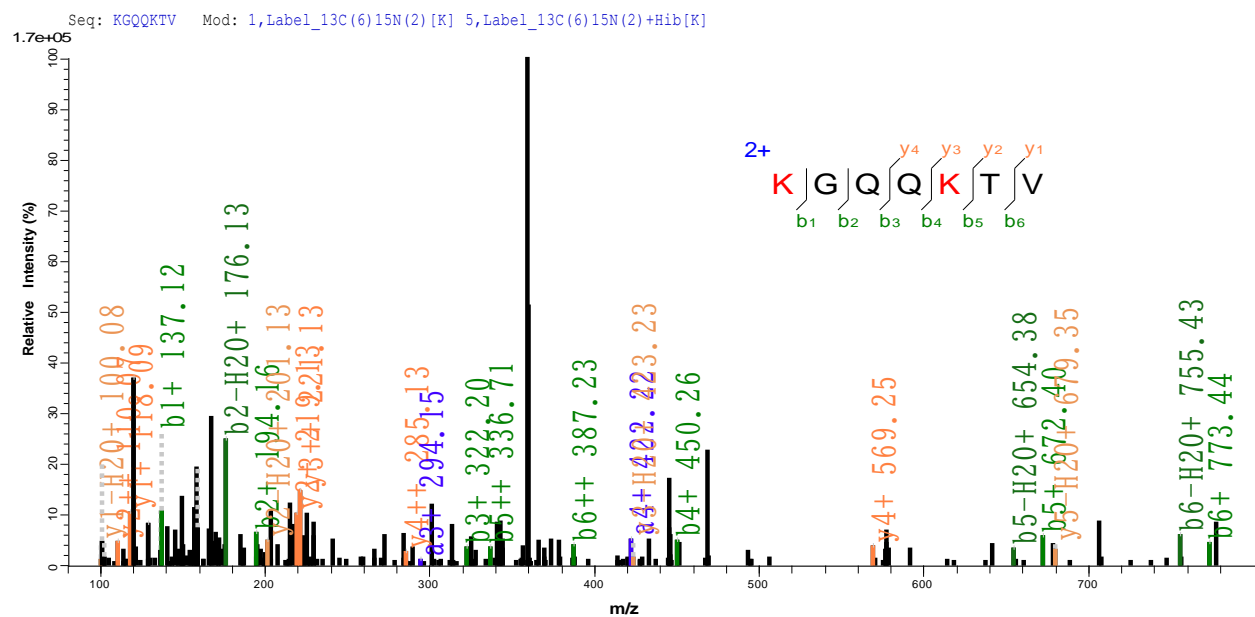
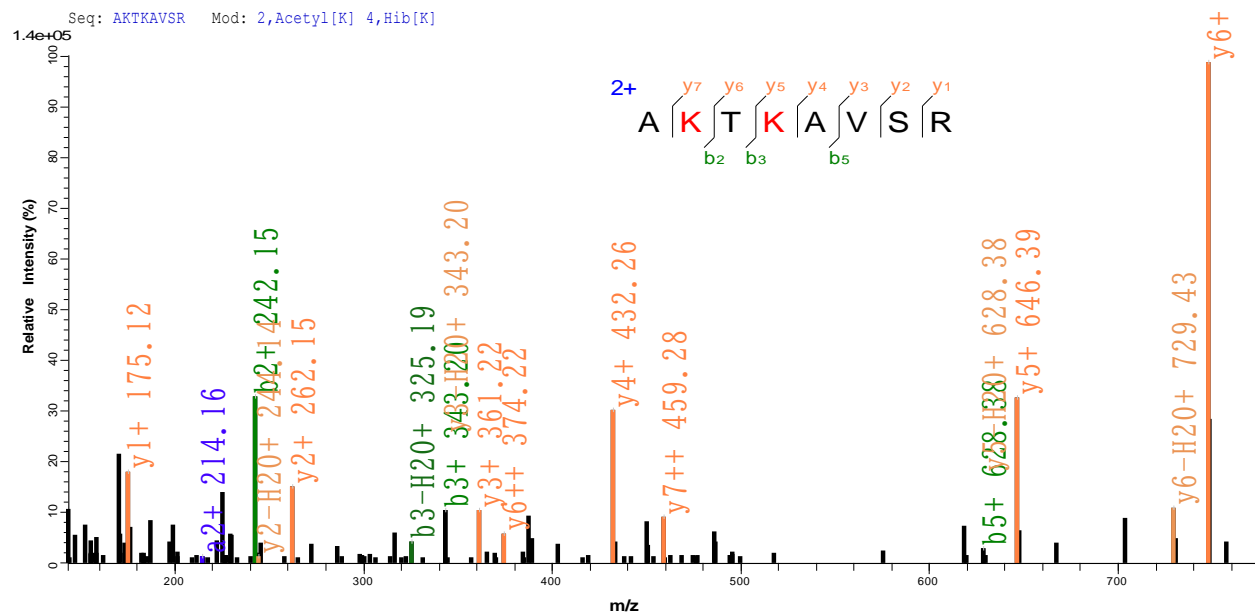


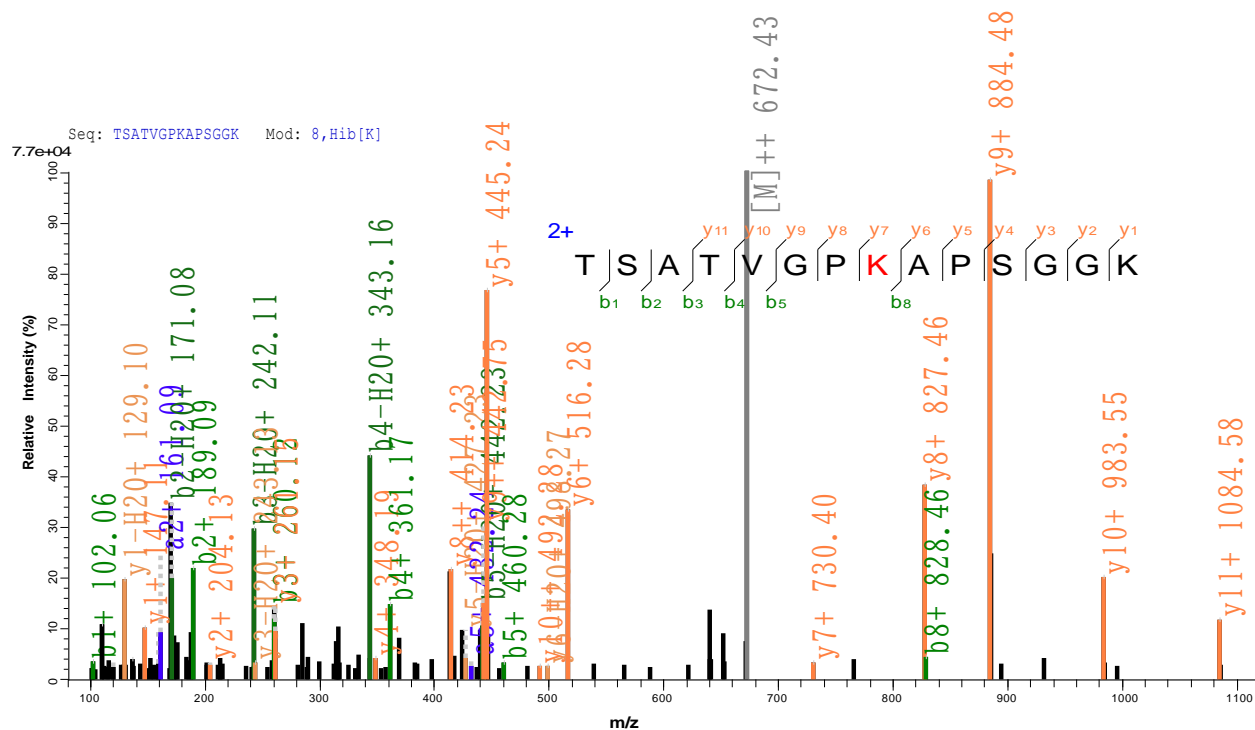
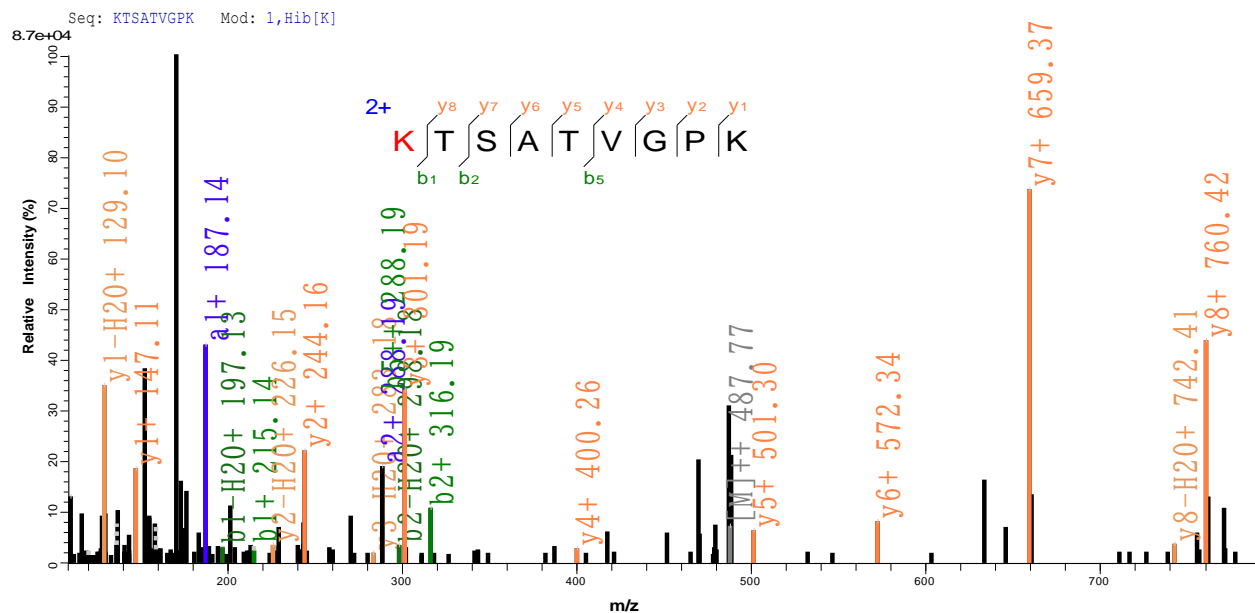
H2A

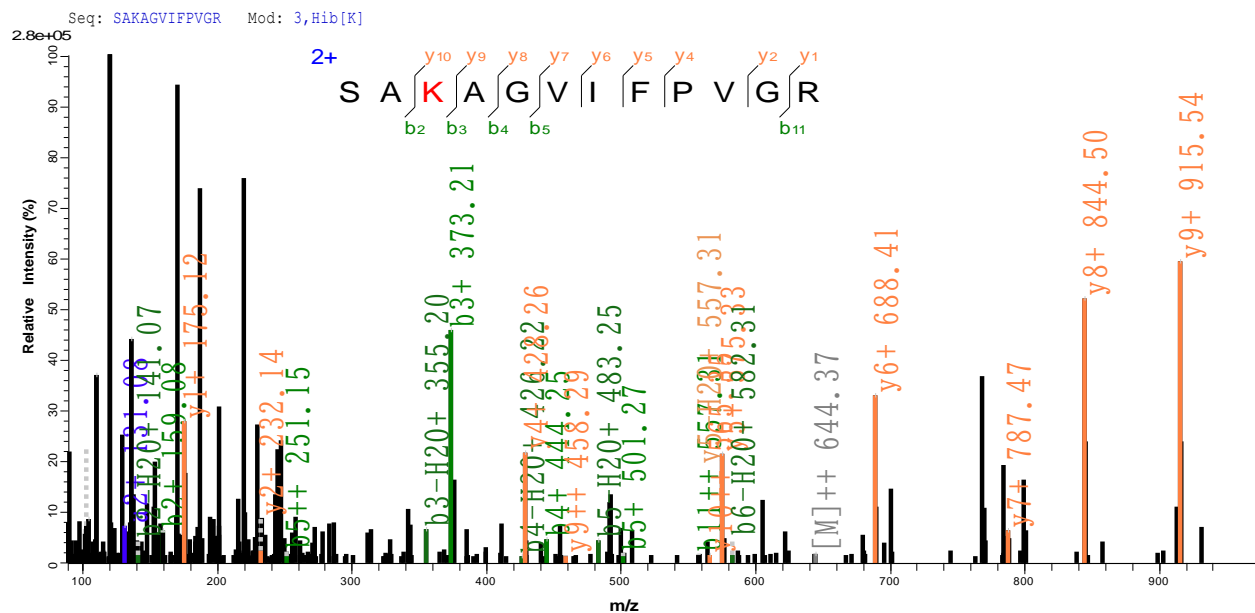




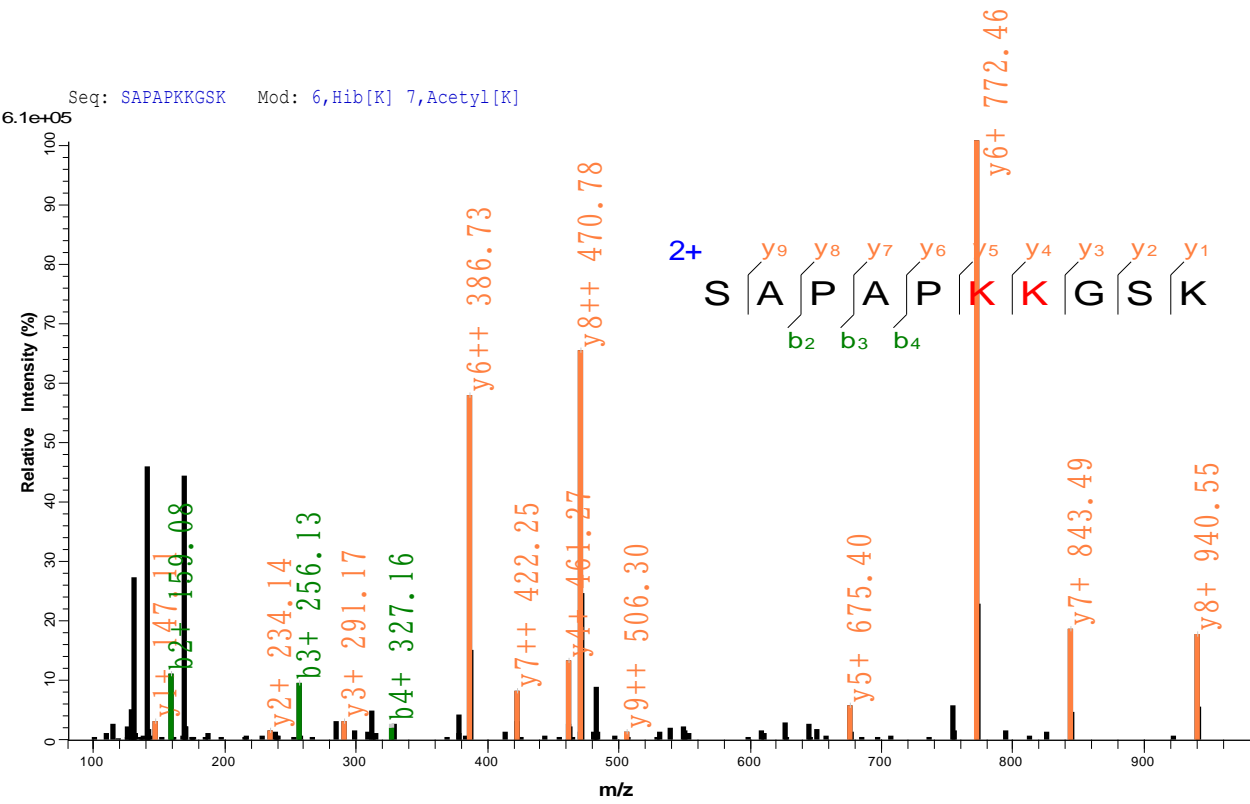
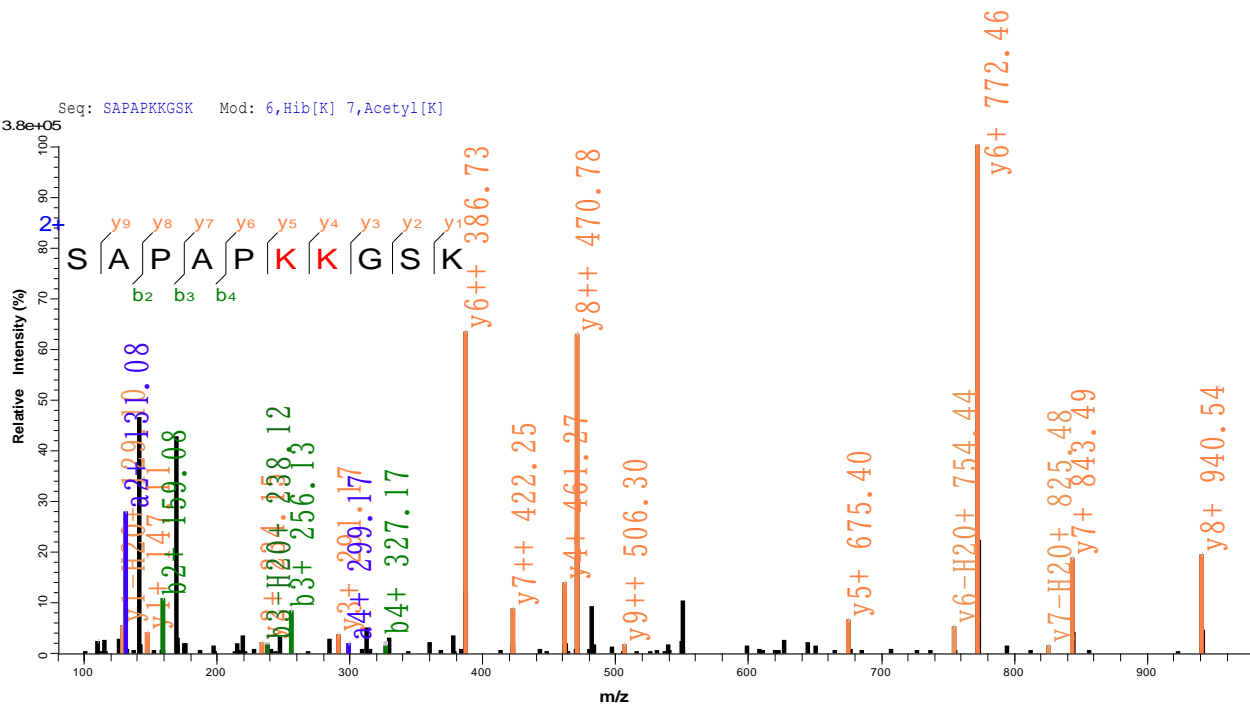


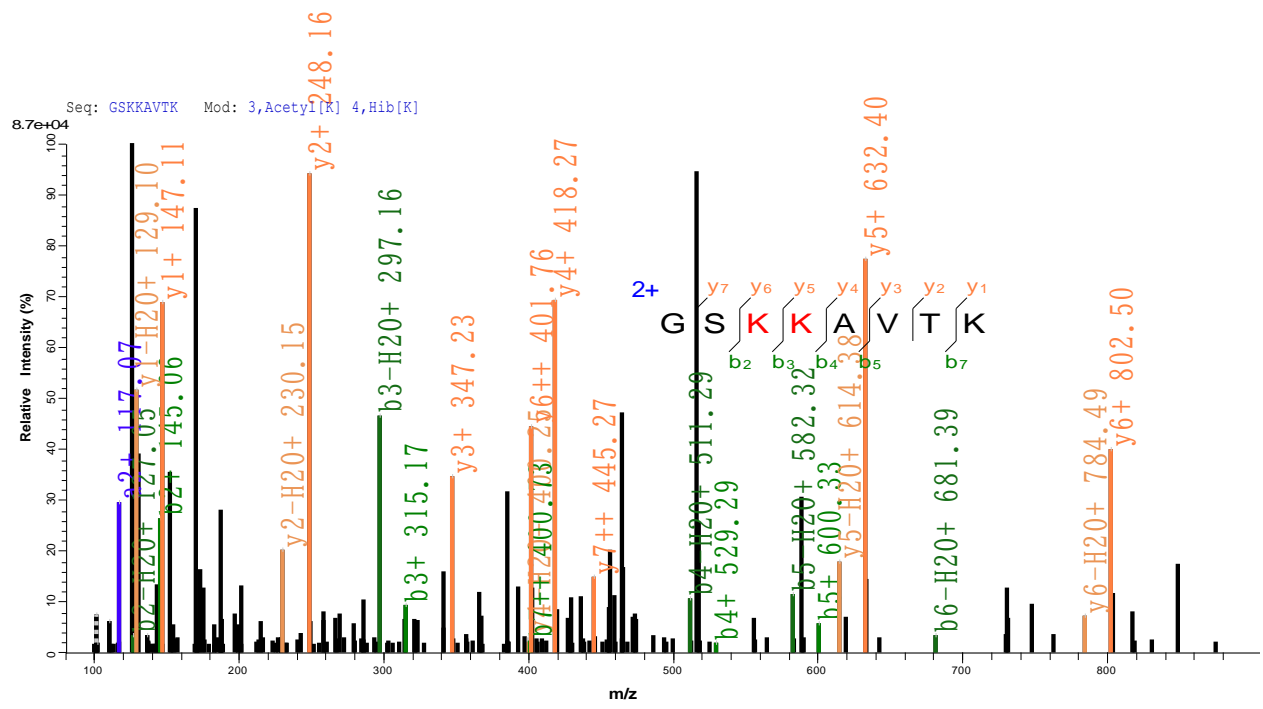






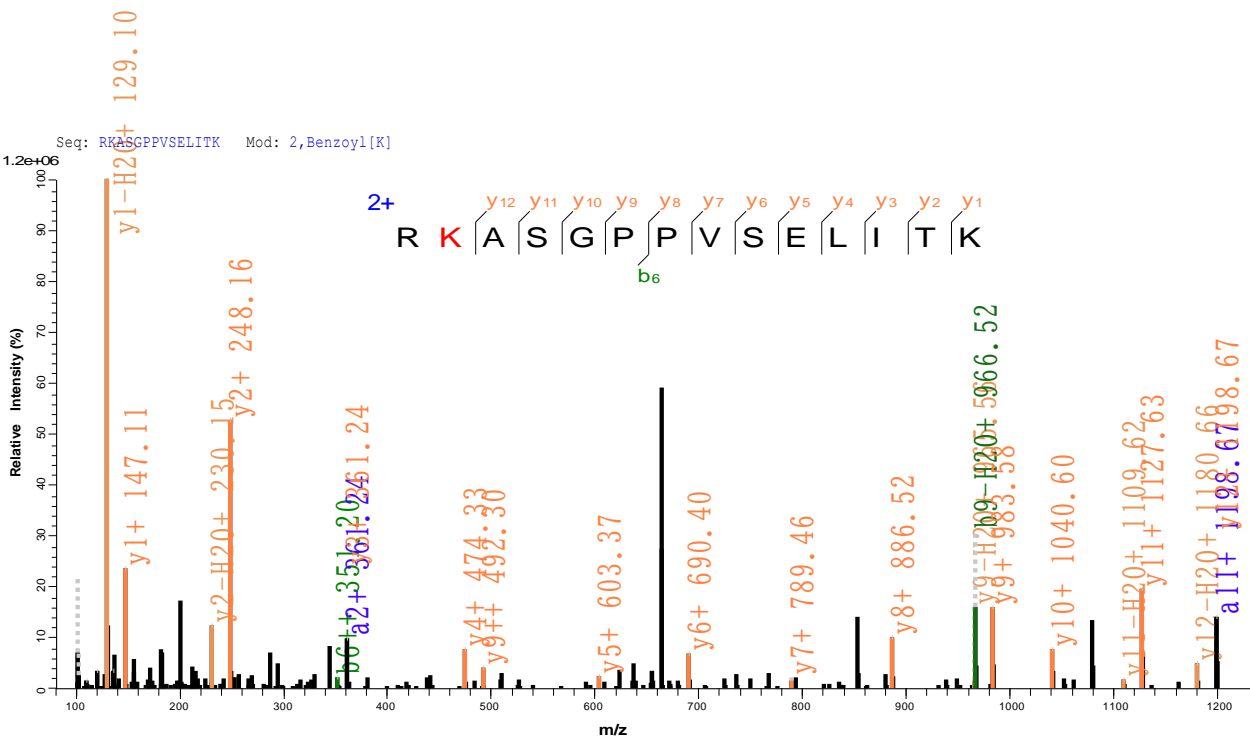
H2B



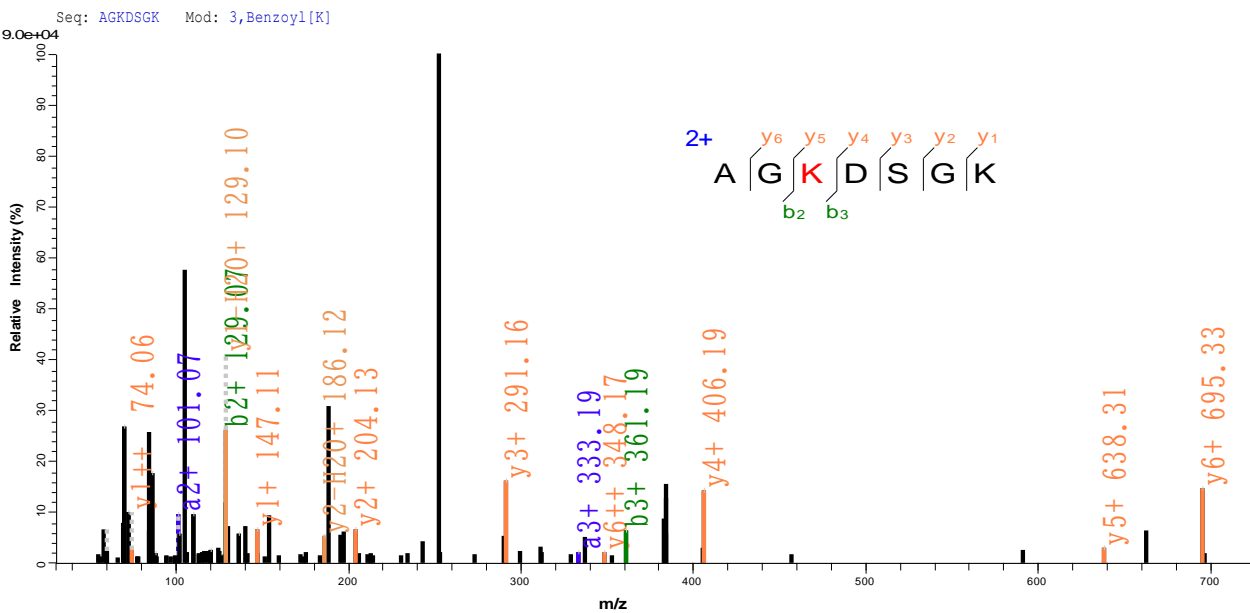


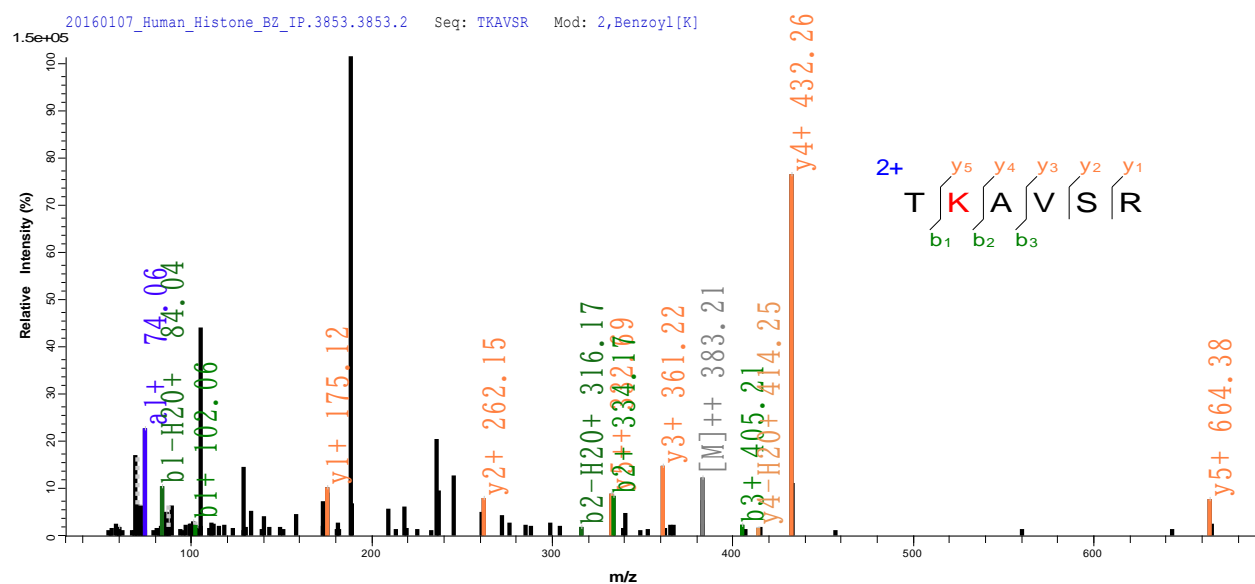
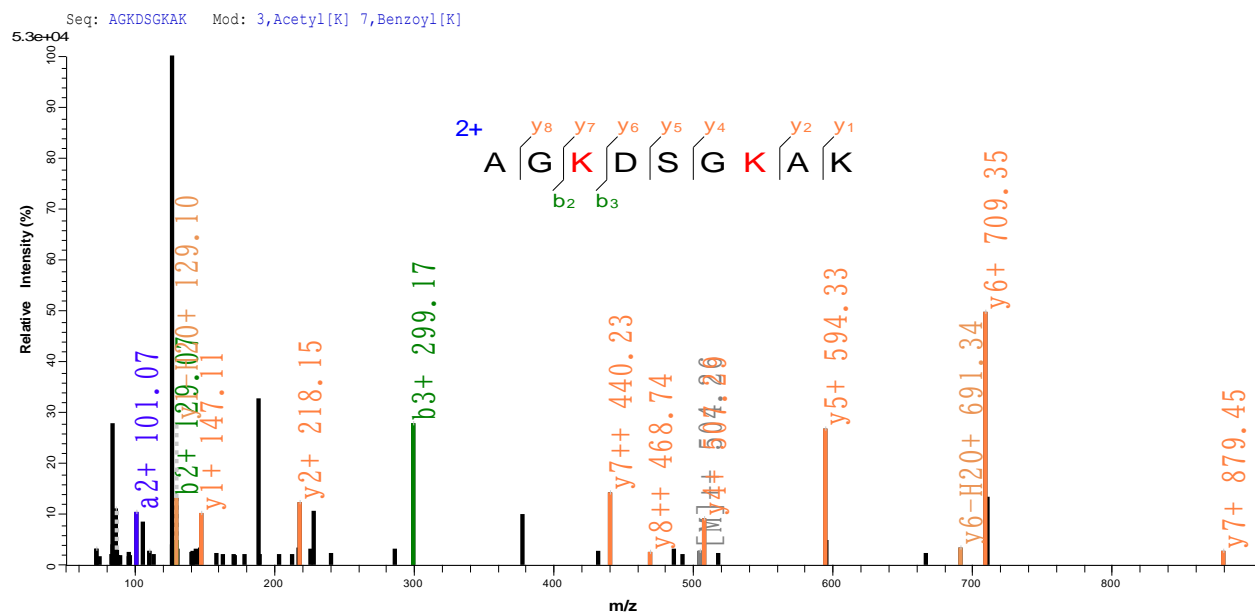
D. Kbz

H1

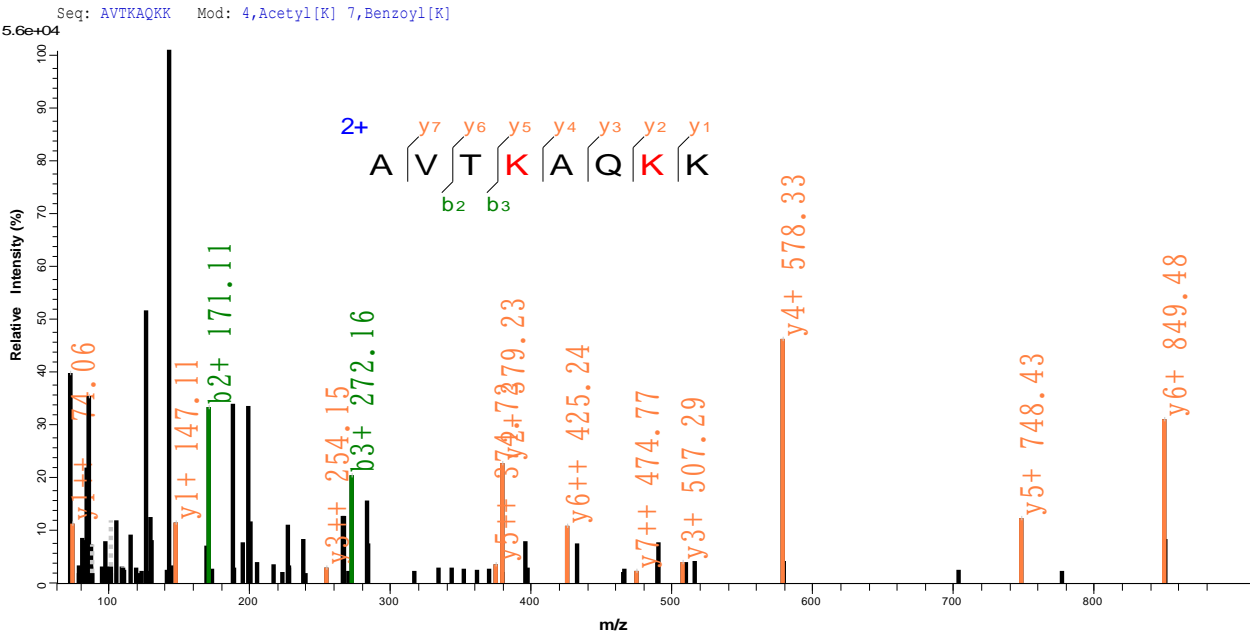


H2A





H2B



H3

