Supplementary Material

The active site of O-GlcNAc transferase imposes constraints on substrate sequence

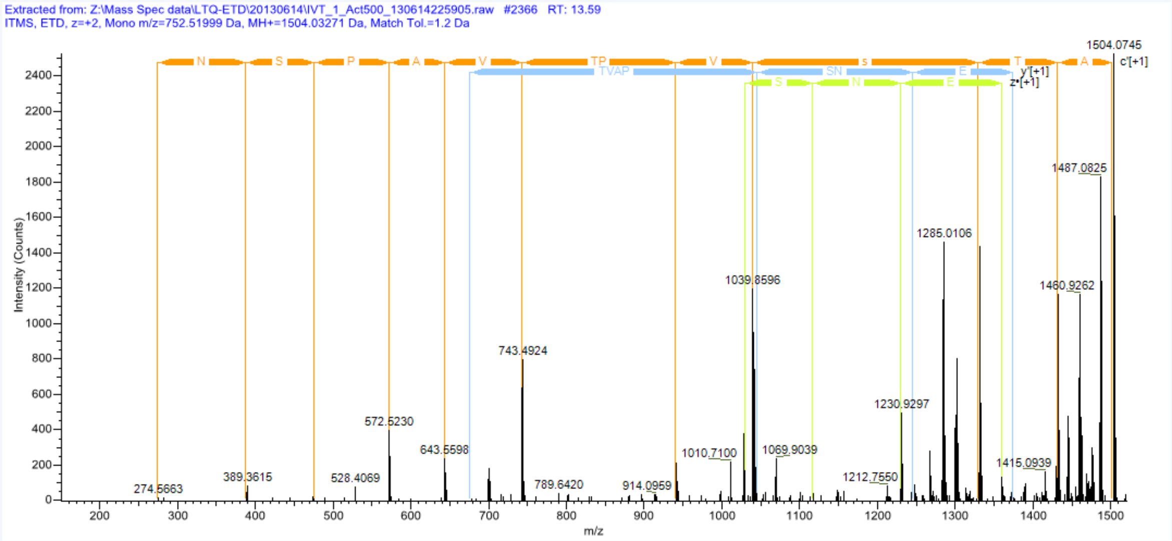
Shalini Pathak[1,a], Jana Alonso[1,a], Marianne Schimpl[1,a], Karim Rafie[a], David E. Blair[a], Vladimir S. Borodkin[a], Osama Albarbarawi[a], and Daan M. F. van Aalten[a,b,2]

Contents

Figure S1.

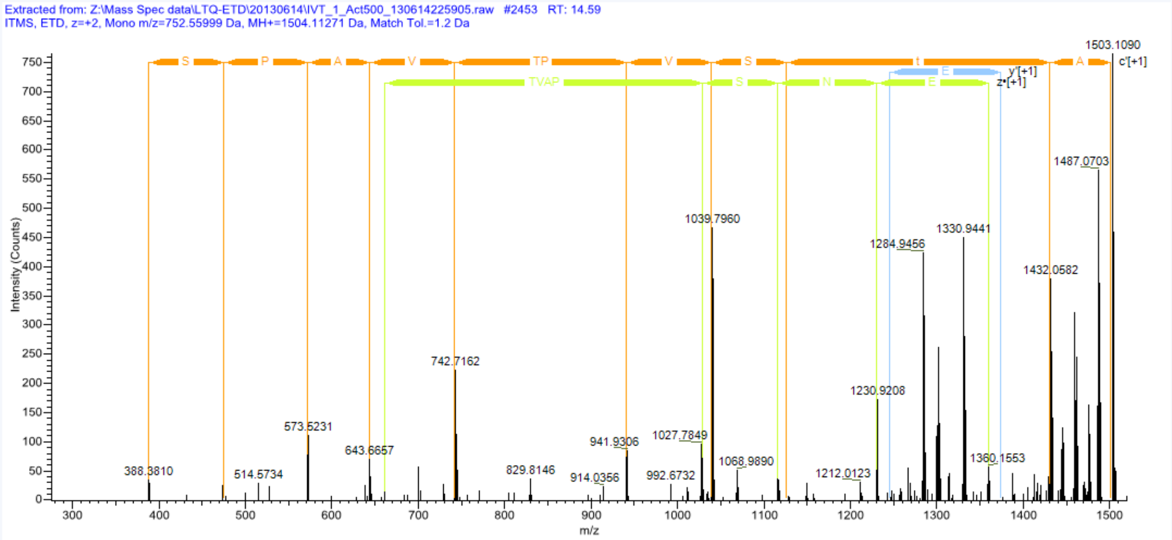
ETD fragmentation followed by Proteome Discoverer search against a generic database that contains the full-length protein and the peptides submitted to glycomapping are shown below. *c* [+1] ions are in orange, *y* [+1] are in blue, and *z* [+1] are in green. The peptide sequence with the fragments identified is showed below the spectrum. The Thr(**t**)/Ser(**s**) amino acid that carries the O-GlcNAc is in bold lowercase.

1. Protein RBL2_HUMAN



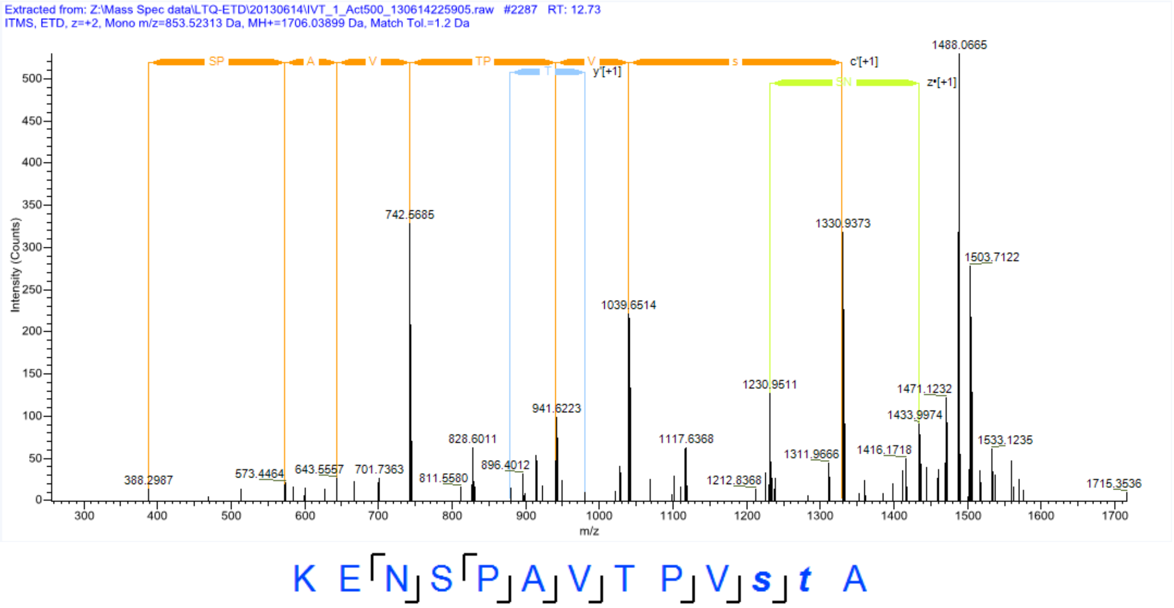
Extracted from: Z:\Mass Spec data\LTQ-ETD\20130614\IVT_1_Act500_130614225905.raw #2366 RT: 13.59
ITMS, ETD, z=+2, Mono m/z=752.51999 Da, MH+=1504.03271 Da, Match Tol.=1.2 Da

K E N S P A V T P V ***s*** T A

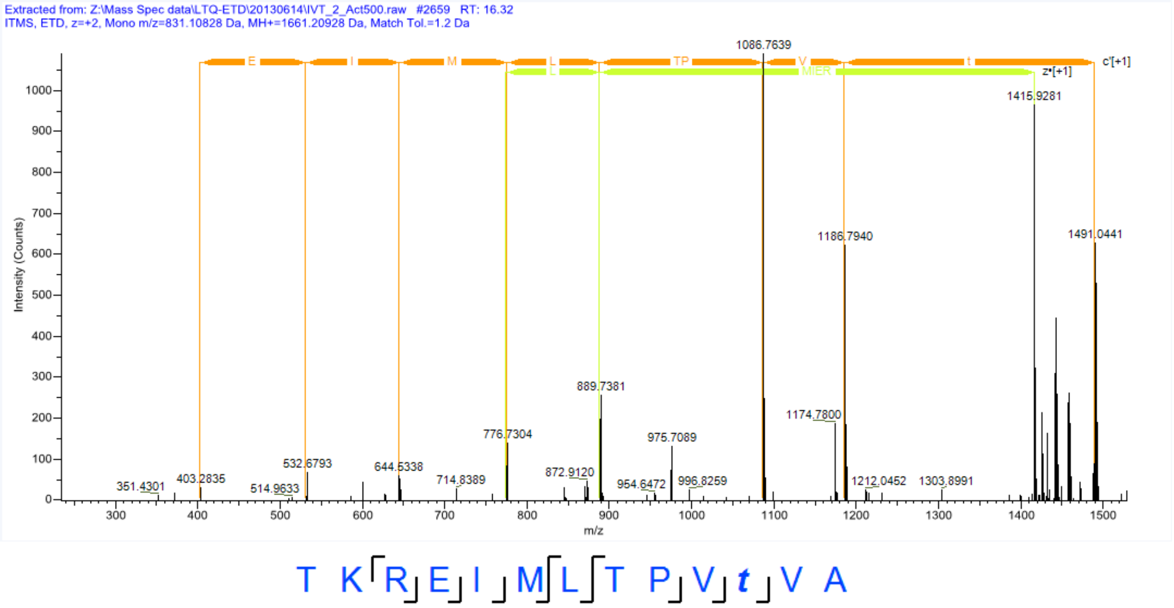


Extracted from: Z:\Mass Spec data\LTQ-ETD\20130614\IVT_1_Act500_130614225905.raw #2453 RT: 14.59
ITMS, ETD, z=+2, Mono m/z=752.55999 Da, MH+=1504.11271 Da, Match Tol.=1.2 Da

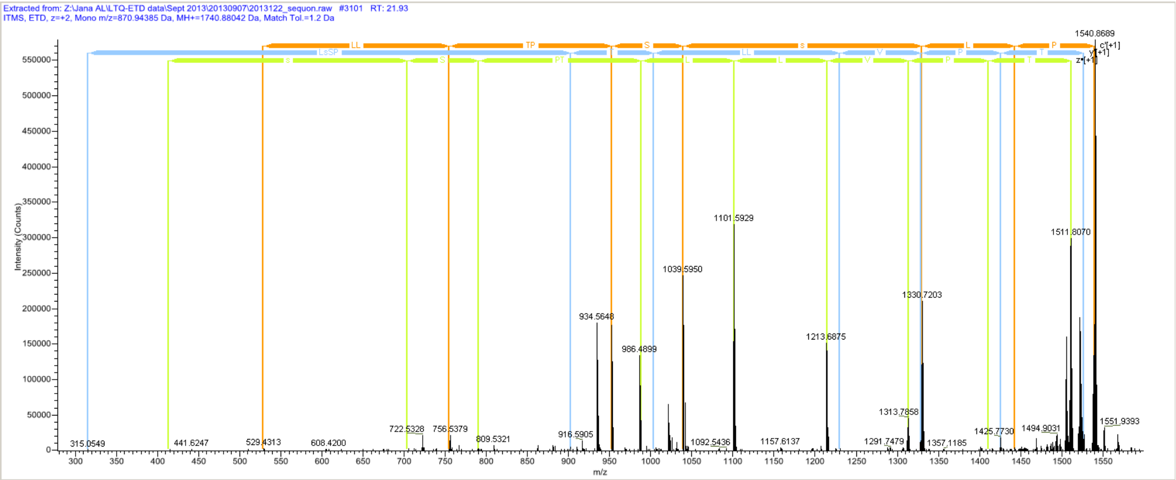
K E N S P A V T P V S ***t*** A



2. M phase phospho protein-9 (MPP9_HUMAN)

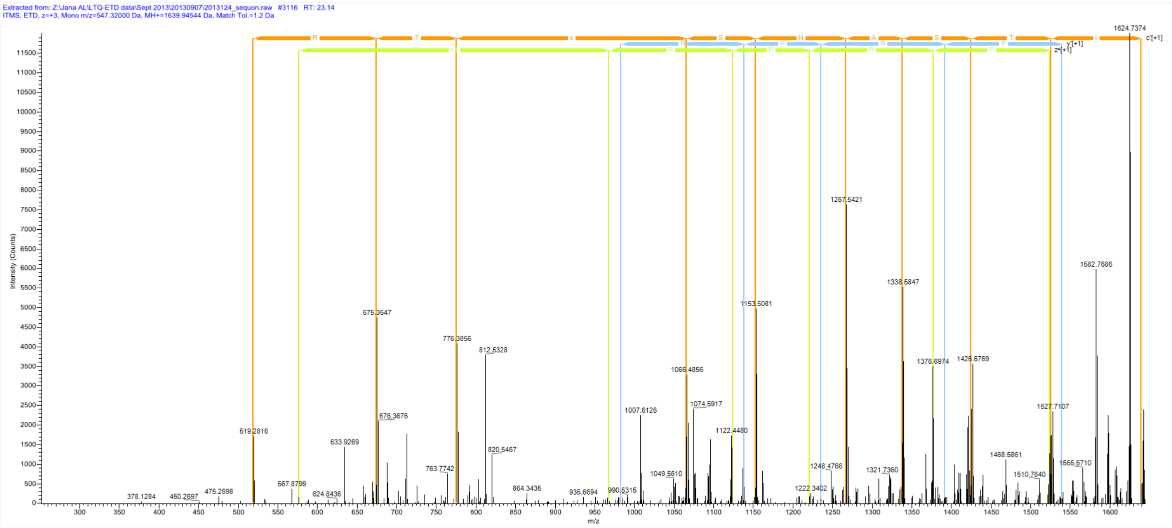


3. ELK1_HUMAN



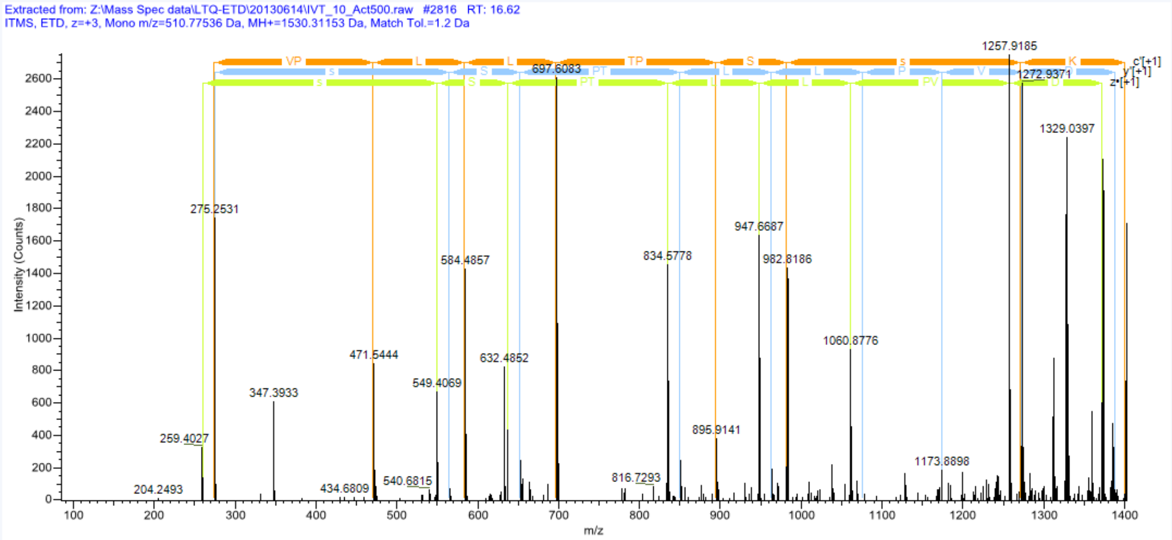
T L⌈T⌈P⌈V∫L⌈L∫T P∫S∫***s***∫L⌋P⌋S I

4. FOXO1


Extracted from: Z:\Jana AL\LTQ-ETD data\Sept 2013\20130907\2013124_sequon.raw #3116 RT: 23.14
ITMS, ETD, z=+3, Mono m/z=547.32000 Da, MH+=1639.94544 Da, Match Tol.=1.2 Da

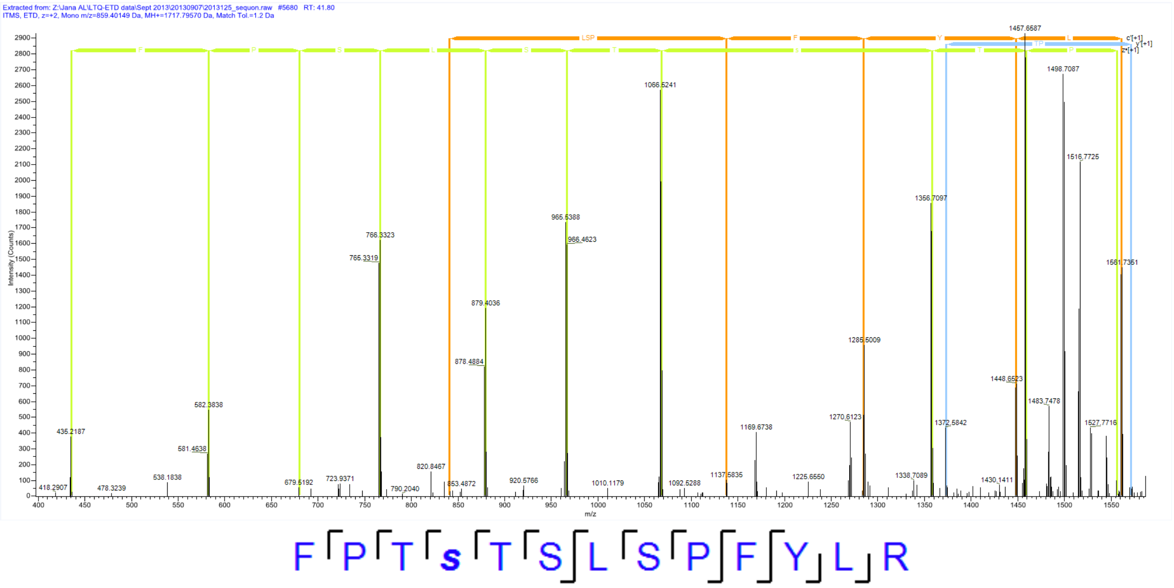
T⌈F⌈R⌈P∫R∫T⌋***s***∫S⌋N⌋A⌋S⌋T⌋I⌋

5. ETS1

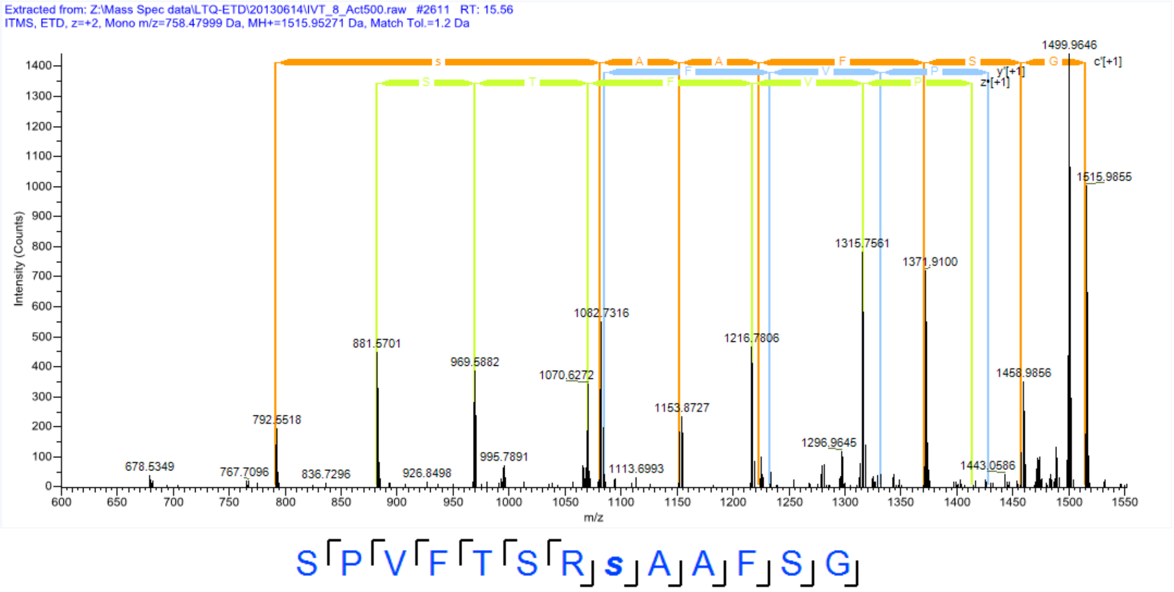

Extracted from: Z:\Mass Spec data\LTQ-ETD\20130614\IVT_10_Act500.raw #2816 RT: 16.62
ITMS, ETD, z=+3, Mono m/z=510.77536 Da, MH+=1530.31153 Da, Match Tol.=1.2 Da

A A⌈D∫V P∫L∫L∫T P∫S∫***s***∫K⌋E

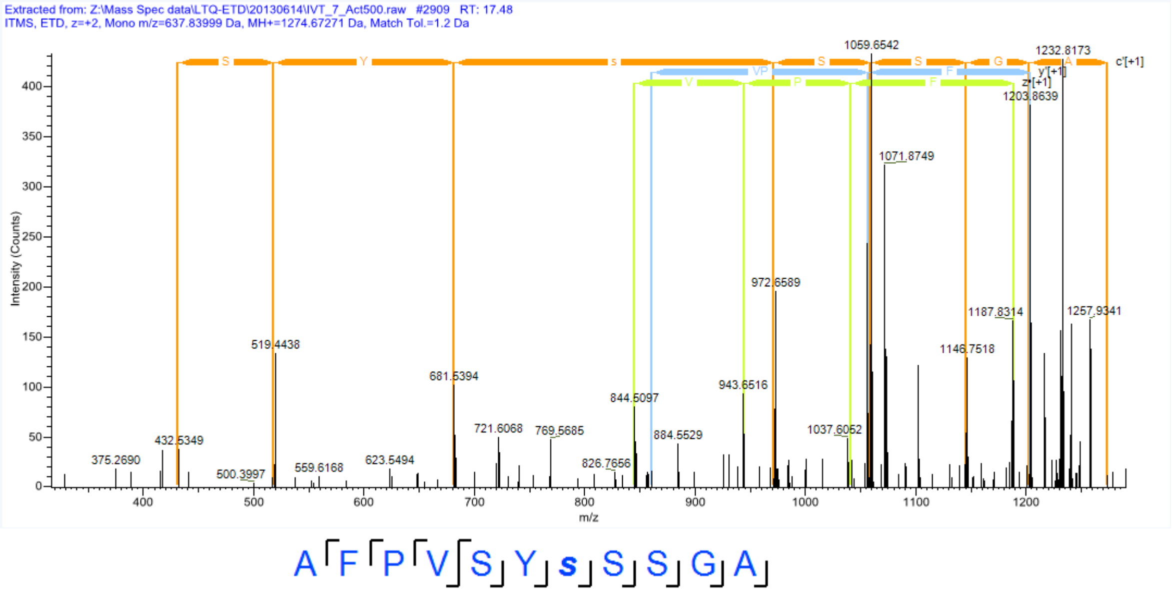
6. α-crystallin B chain



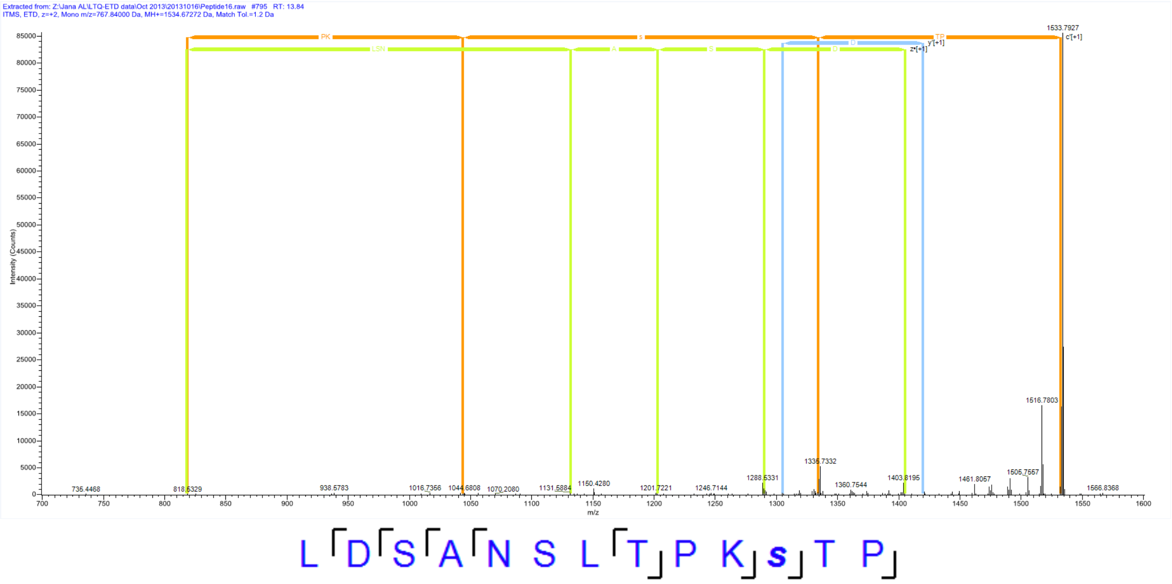
7. Keratin 7



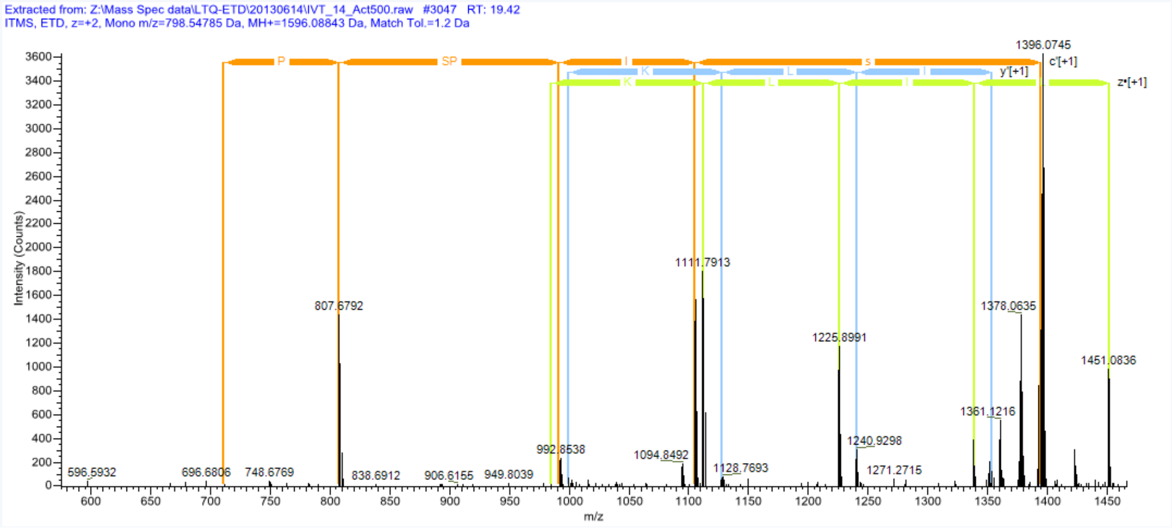
8. RET



9. MYB-related B protein

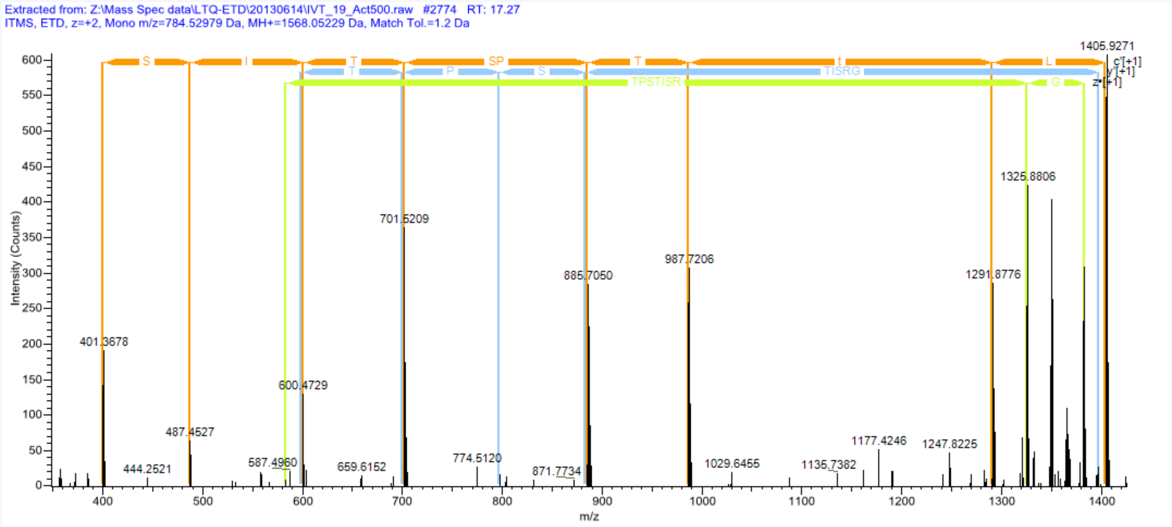


10. SCG


Extracted from: Z:\Mass Spec data\LTQ-ETD\20130614\IVT_14_Act500.raw #3047 RT: 19.42
ITMS, ETD, z=+2, Mono m/z=798.54785 Da, MH+=1596.08843 Da, Match Tol.=1.2 Da

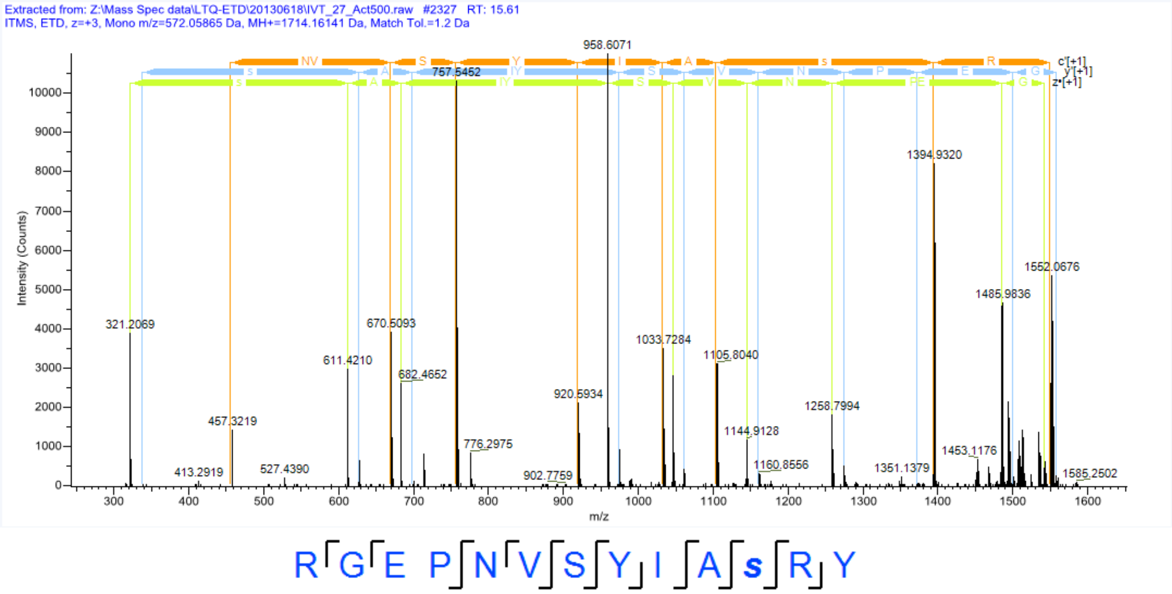
E⌈L⌈I ⌈L⌈K⌈P⌋P⌋S P⌋I ⌋***s***⌋E A

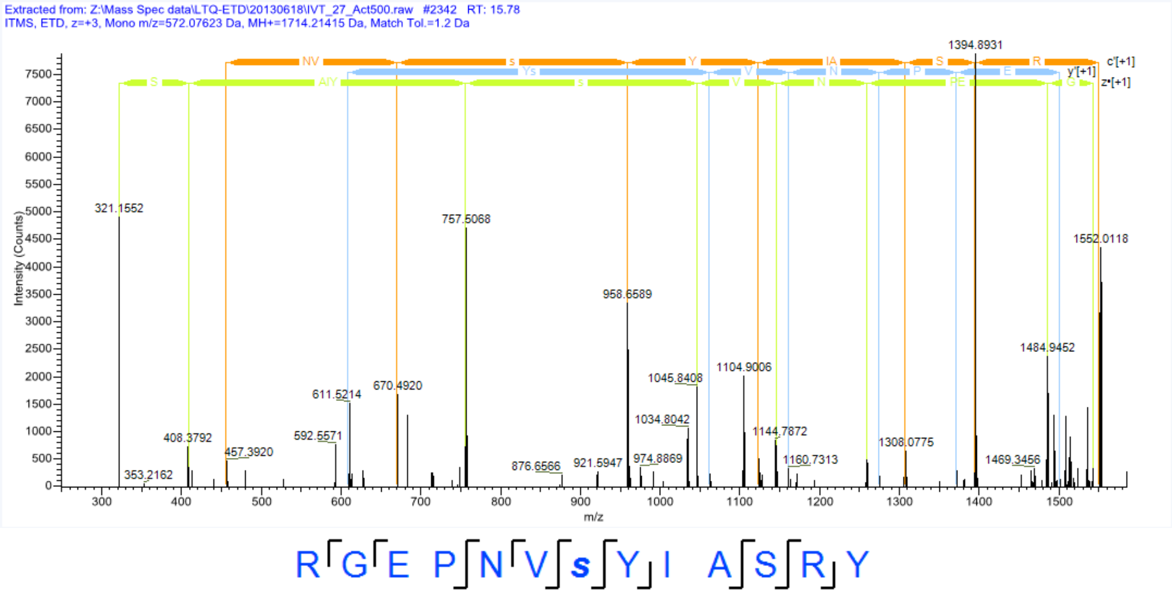
11. RBL2


Extracted from: Z:\Mass Spec data\LTQ-ETD\20130614\IVT_19_Act500.raw #2774 RT: 17.27
ITMS, ETD, z=+2, Mono m/z=784.52979 Da, MH+=1568.05229 Da, Match Tol.=1.2 Da

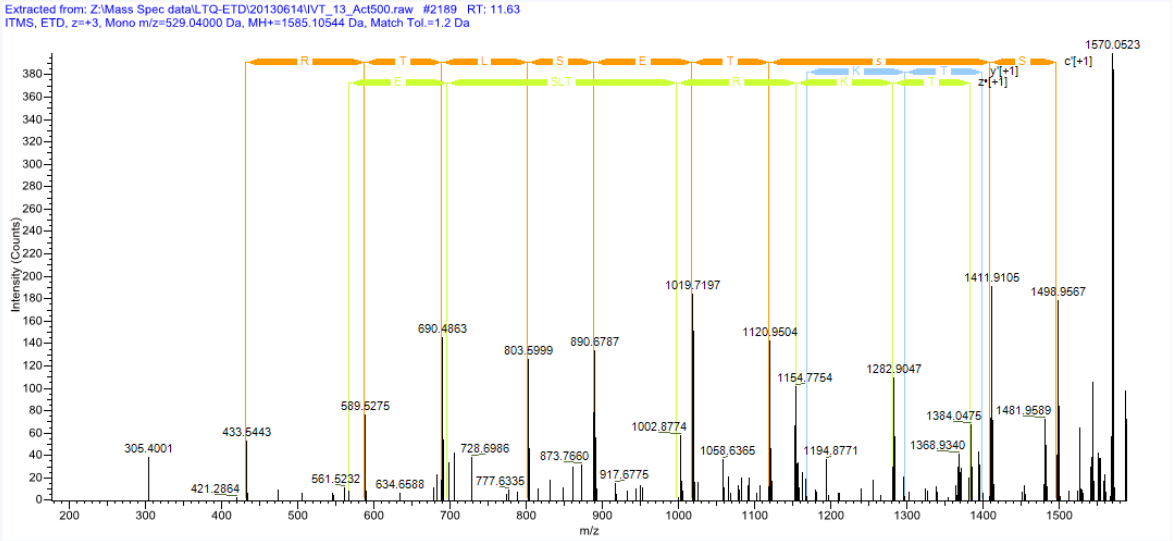
G L⌈G⌈R⌋S⌋I ⌋T⌋S P⌋T⌈***t***⌋L⌋Y

12. GSK3β


Extracted from: Z:\Mass Spec data\LTQ-ETD\20130618\IVT_27_Act500.raw #2327 RT: 15.61
ITMS, ETD, z=+3, Mono m/z=572.05865 Da, MH+=1714.16141 Da, Match Tol.=1.2 Da
c'[+1]
y'[+1]
z•[+1]
Intensity (Counts)
m/z
958.6071
757.5452
1394.9320
1552.0676
1485.9836
321.2069
670.5093
611.4210
682.4652
1033.7284
1105.8040
920.5934
457.3219
1258.7994
1144.9128
776.2975
1453.1176
1160.8556
1351.1379
413.2919
527.4390
902.7759
1585.2502
R G E P N V S Y I A s R Y

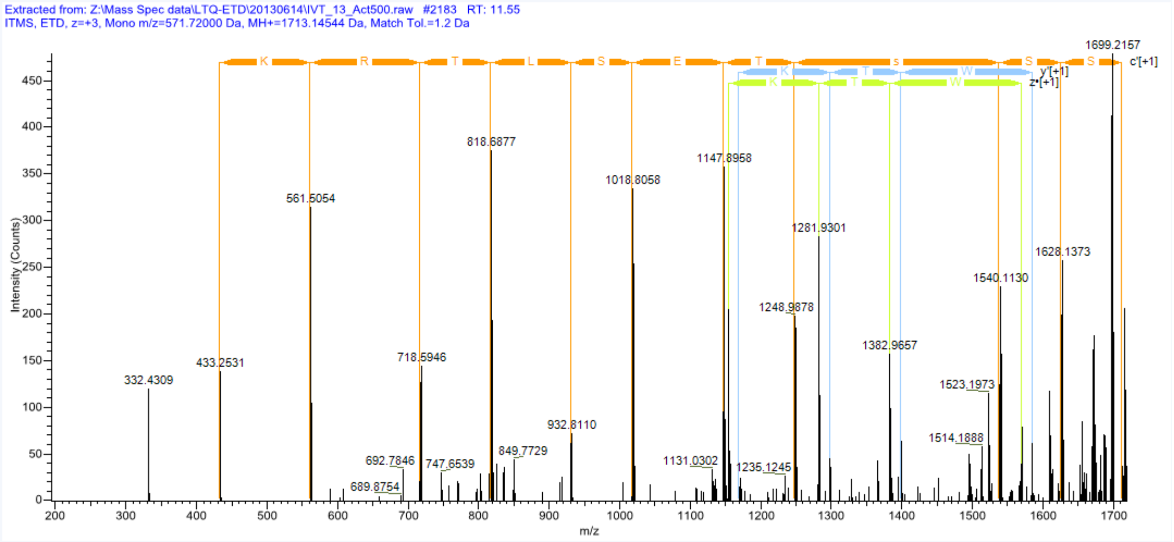

Extracted from: Z:\Mass Spec data\LTQ-ETD\20130618\IVT_27_Act500.raw #2342 RT: 15.78
ITMS, ETD, z=+3, Mono m/z=572.07623 Da, MH+=1714.21415 Da, Match Tol.=1.2 Da
c'[+1]
y'[+1]
z•[+1]
Intensity (Counts)
m/z
1394.8931
321.1552
757.5068
1552.0118
958.6589
1484.9452
1045.8408
1104.9006
611.5214
670.4920
1034.8042
408.3792
592.5571
1144.7872
1308.0775
457.3920
974.8869
1160.7313
1469.3456
921.5947
876.6566
353.2162
R G E P N V s Y I A S R Y

13. Kv2.1



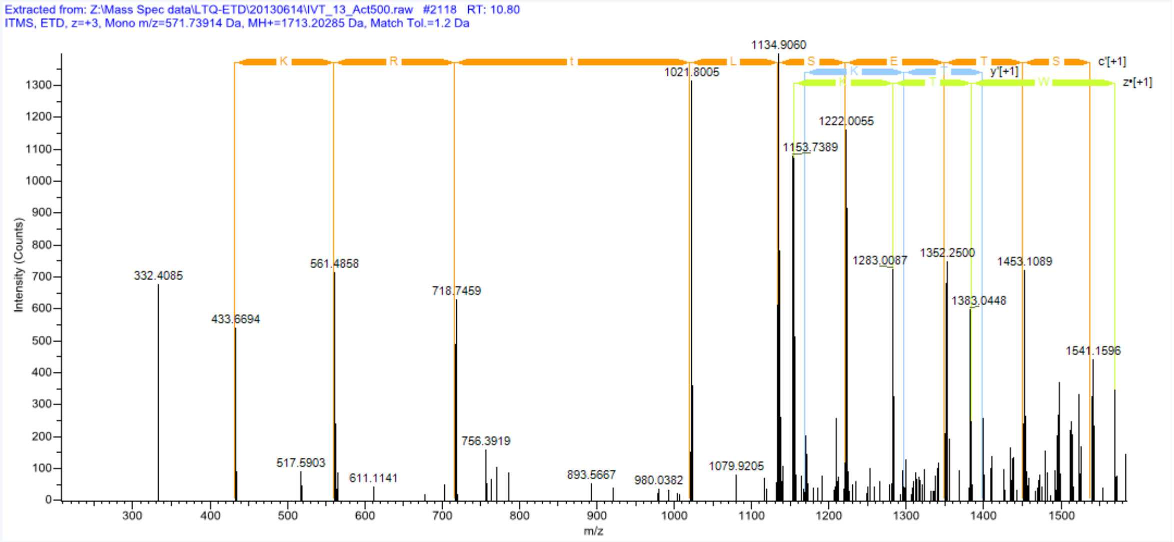
Extracted from: Z:\Mass Spec data\LTQ-ETD\20130614\IVT_13_Act500.raw #2189 RT: 11.63
ITMS, ETD, z=+3, Mono m/z=529.04000 Da, MH+=1585.10544 Da, Match Tol.=1.2 Da

W T K R T L S E T ***s*** S S



Extracted from: Z:\Mass Spec data\LTQ-ETD\20130614\IVT_13_Act500.raw #2183 RT: 11.55
ITMS, ETD, z=+3, Mono m/z=571.72000 Da, MH+=1713.14544 Da, Match Tol.=1.2 Da

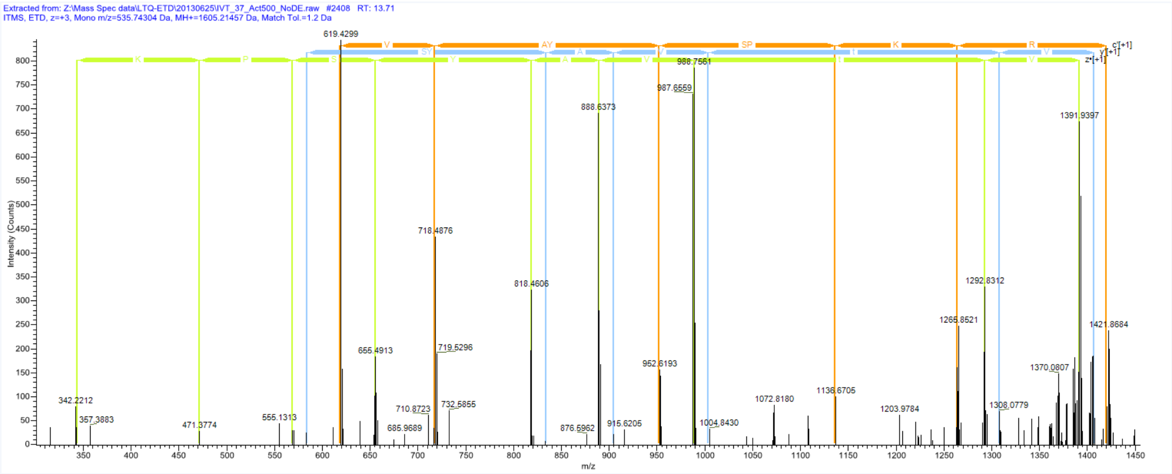
K W T K R T L S E T ***s*** S S



Extracted from: Z:\Mass Spec data\LTQ-ETD\20130614\IVT_13_Act500.raw #2118 RT: 10.80
ITMS, ETD, z=+3, Mono m/z=571.73914 Da, MH+=1713.20285 Da, Match Tol.=1.2 Da

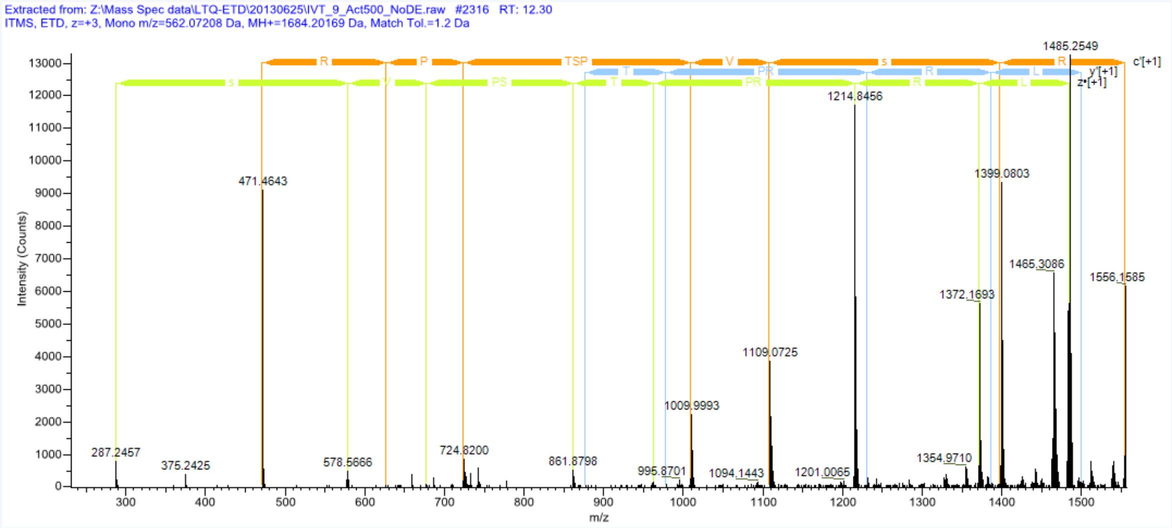
K⌈W⌋T⌋K⌋R⌋ ***t*** ⌋L⌋S⌋E⌋T⌋S⌋S S

14. M phase phospho protein-9


Extracted from: Z:\Mass Spec data\LTQ-ETD\20130625\IVT_37_Act500_NoDE.raw #2408 RT: 13.71
ITMS, ETD, z=+3, Mono m/z=535.74304 Da, MH+=1605.21457 Da, Match Tol.=1.2 Da

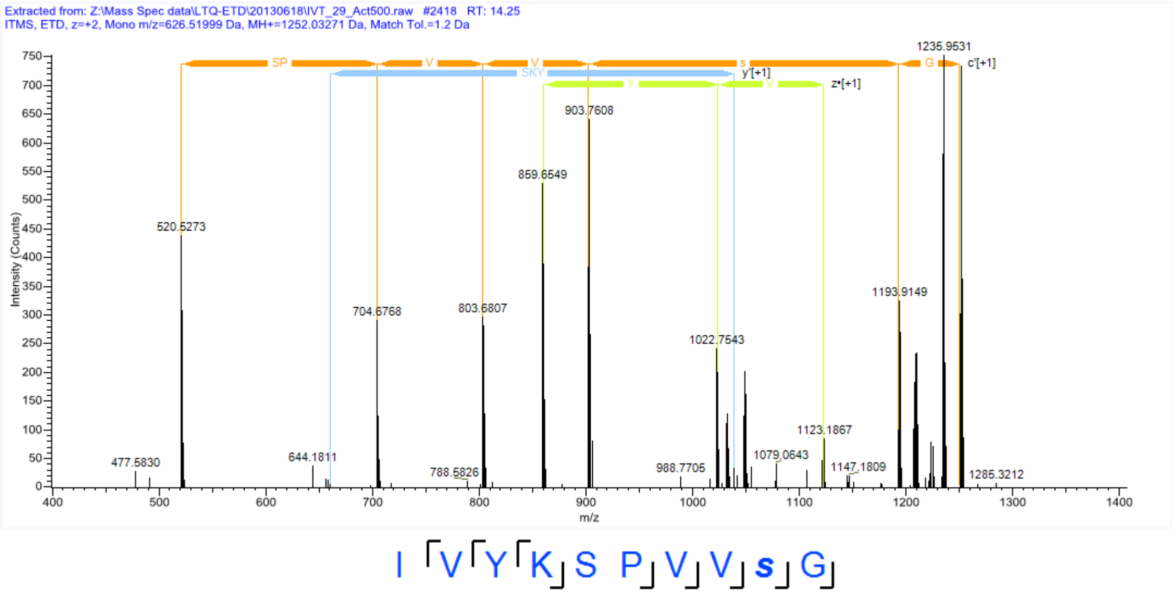
T P⌈V⌋ ***t*** ⌋V⌋A⌈Y⌋S⌈P⌋K⌋R⌋S P

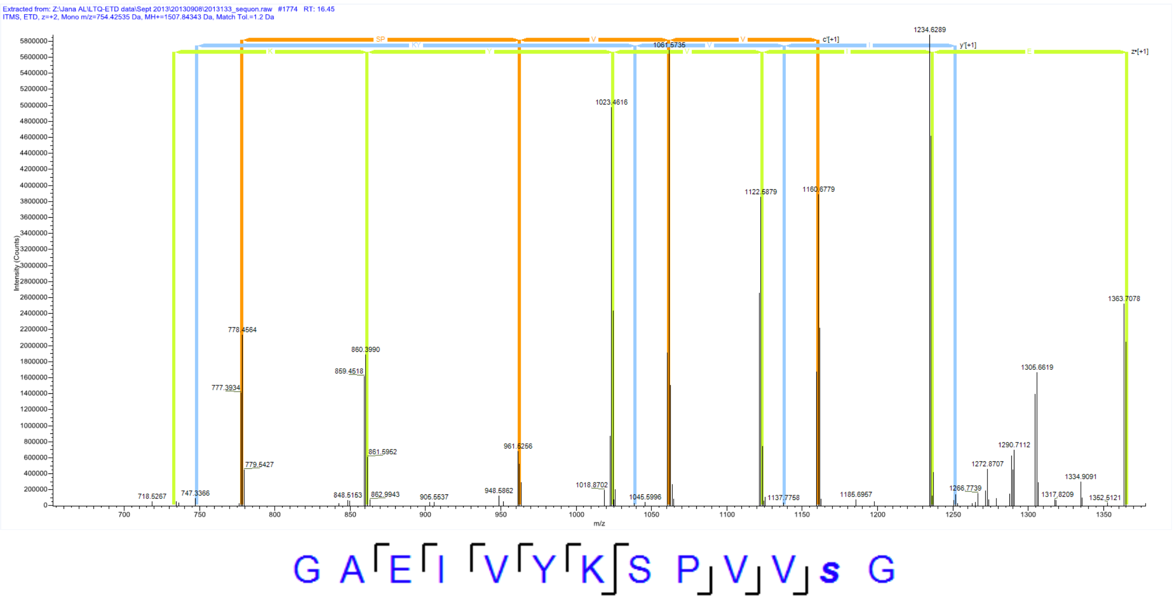
15. Early E1A32Kd


Extracted from: Z:\Mass Spec data\LTQ-ETD\20130625\IVT_9_Act500_NoDE.raw #2316 RT: 12.30
ITMS, ETD, z=+3, Mono m/z=562.07208 Da, MH+=1684.20169 Da, Match Tol.=1.2 Da

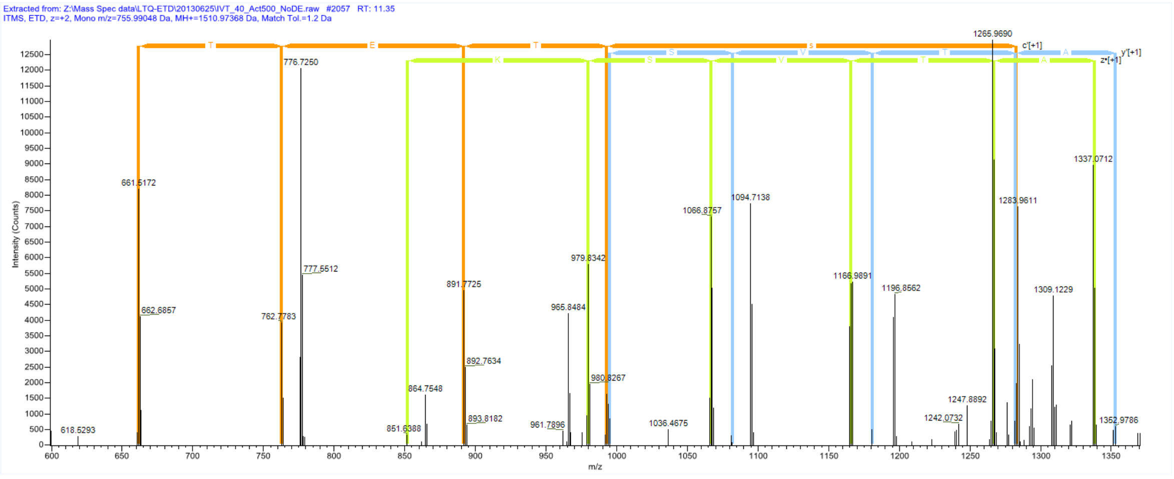
A I ⌈L⌈R⌋R⌋P⌋T⌈S P⌋V⌋ ***s*** ⌋R⌋E

16. Tau

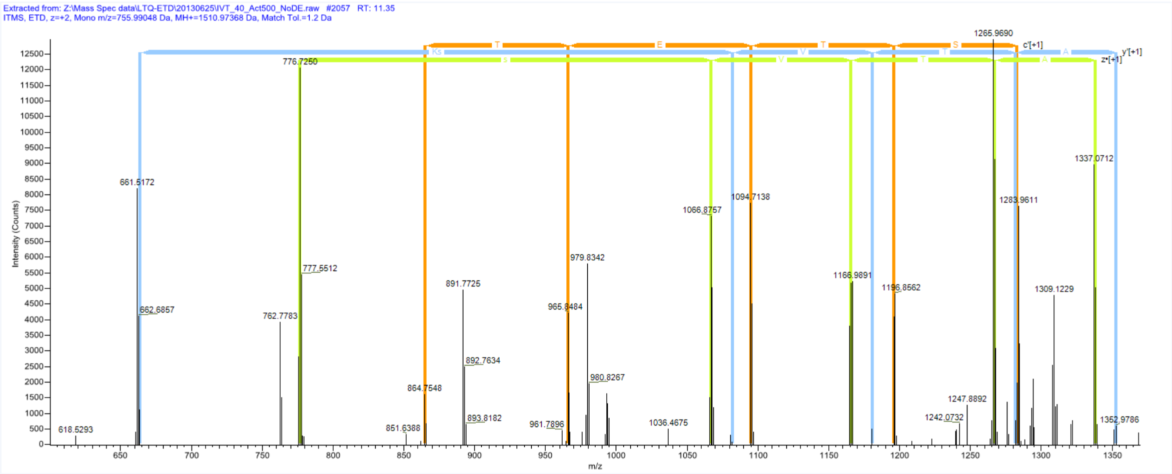




17. Rhodopsin-OPSD

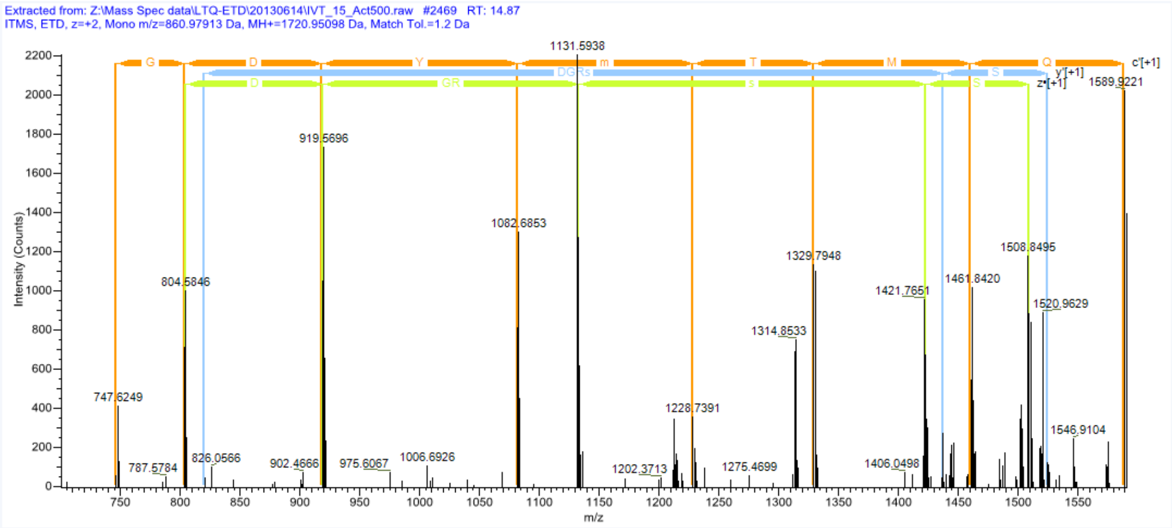


A S⌈A⌈T⌈V⌈S⌈K⌋⌈T⌋E⌋T⌋***s***⌋Q V

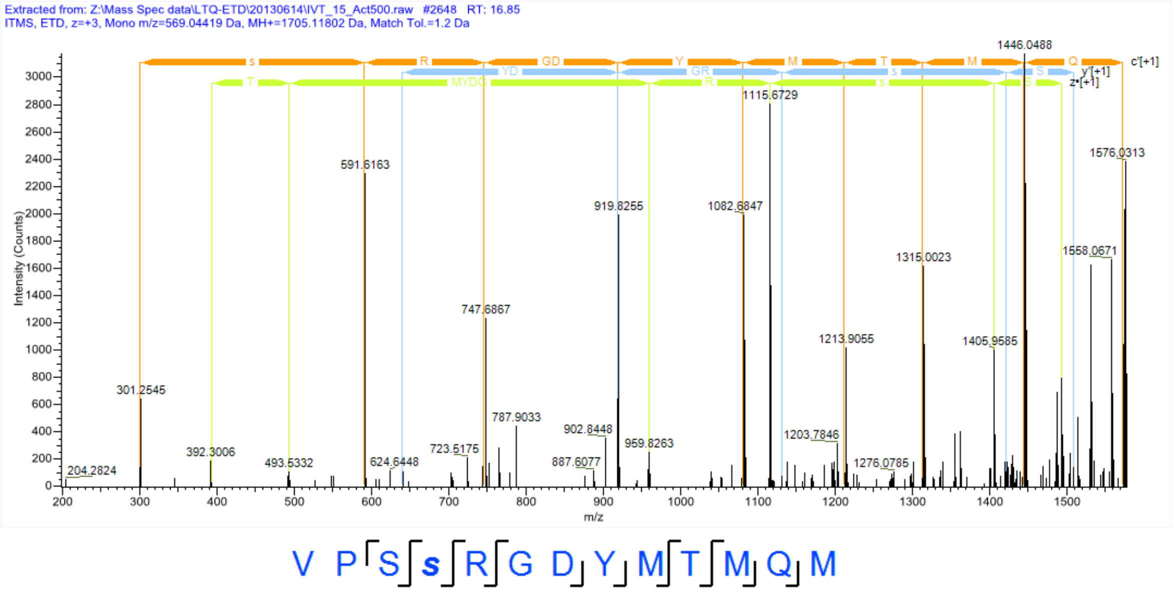


A S⌈A⌈T⌈V⌈***s***⌈K⌋T⌋E⌋T⌋S⌋Q V

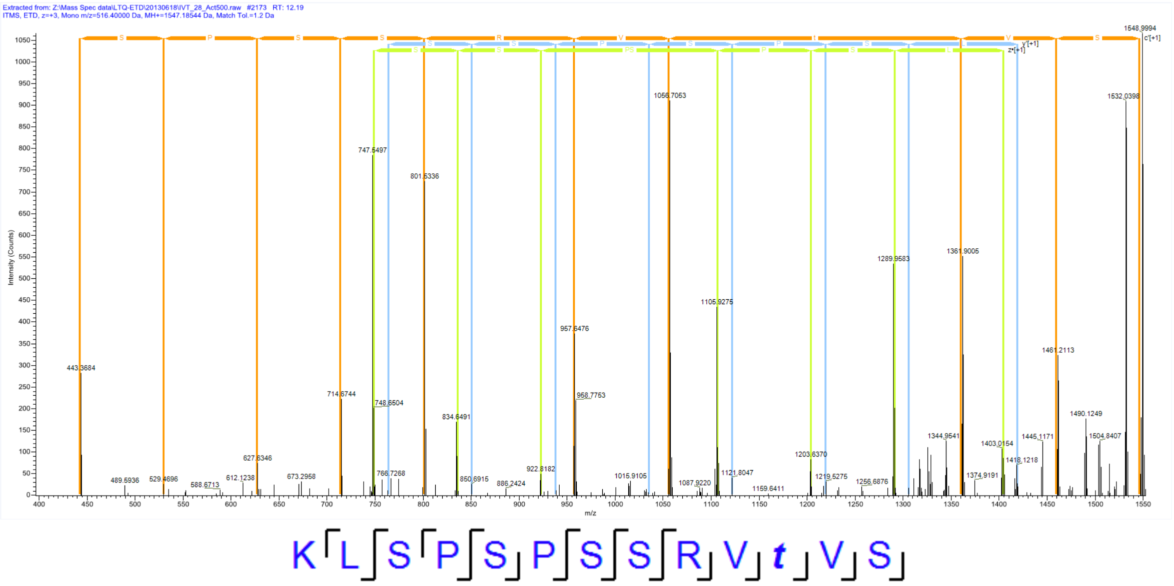
18. Insulin receptor substrate-1



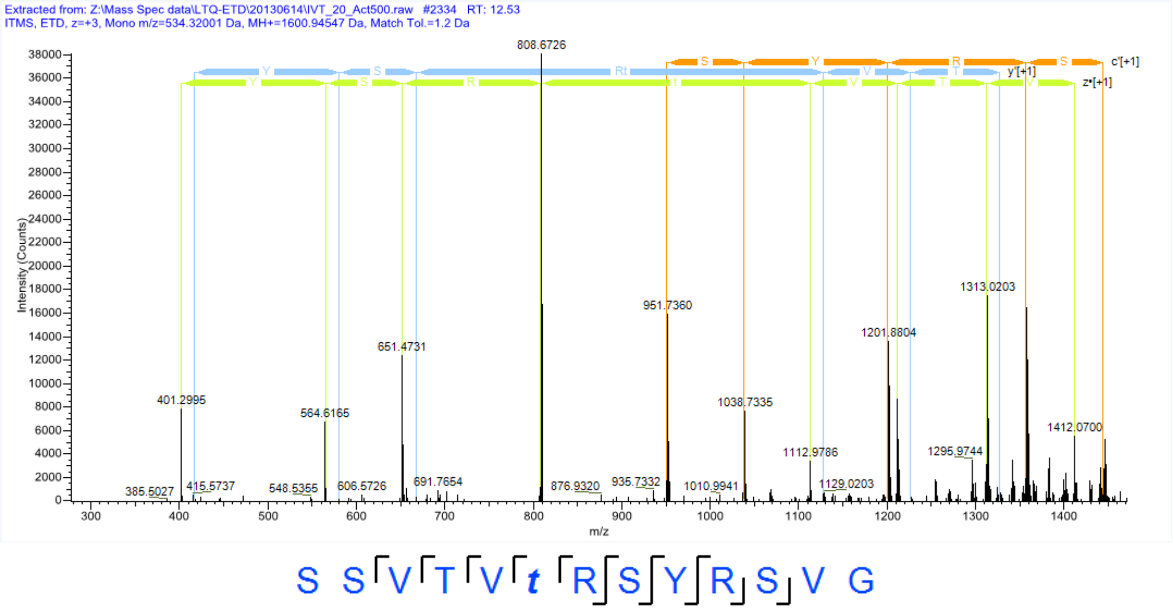
V P⌈S⌈***s***⌈R⌋G⌋⌈D⌋⌈Y⌋***m***⌋T⌋M⌋Q⌋M

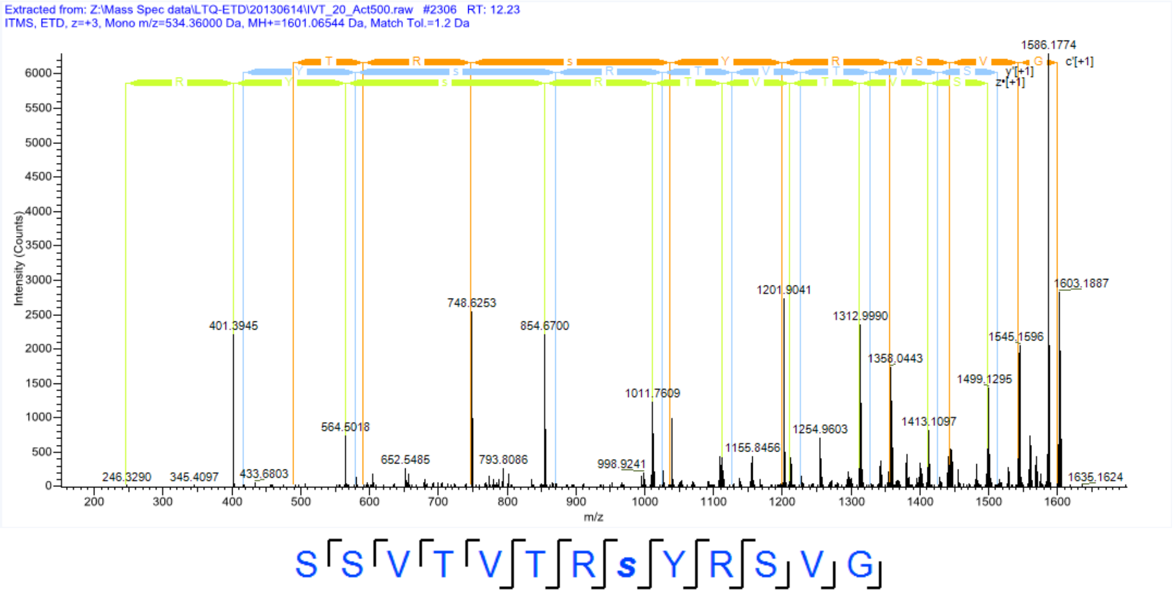

Extracted from: Z:\Mass Spec data\LTQ-ETD\20130614\IVT_15_Act500.raw #2648 RT: 16.85
ITMS, ETD, z=+3, Mono m/z=569.04419 Da, MH+=1705.11802 Da, Match Tol.=1.2 Da
Intensity (Counts)
m/z
c'[+1]
y'[+1]
z•[+1]
s
R
GD
Y
M
T
M
Q
YD
GR
s
S
T
MYDG
R
s
S
1446.0488
1115.6729
1576.0313
591.6163
919.8255
1082.6847
1315.0023
1558.0671
747.6867
1213.9055
1405.9585
301.2545
787.9033
902.8448
1203.7846
723.5175
959.8263
392.3006
624.6448
887.6077
1276.0785
493.5332
204.2824
V P S s R G D Y M T M Q M

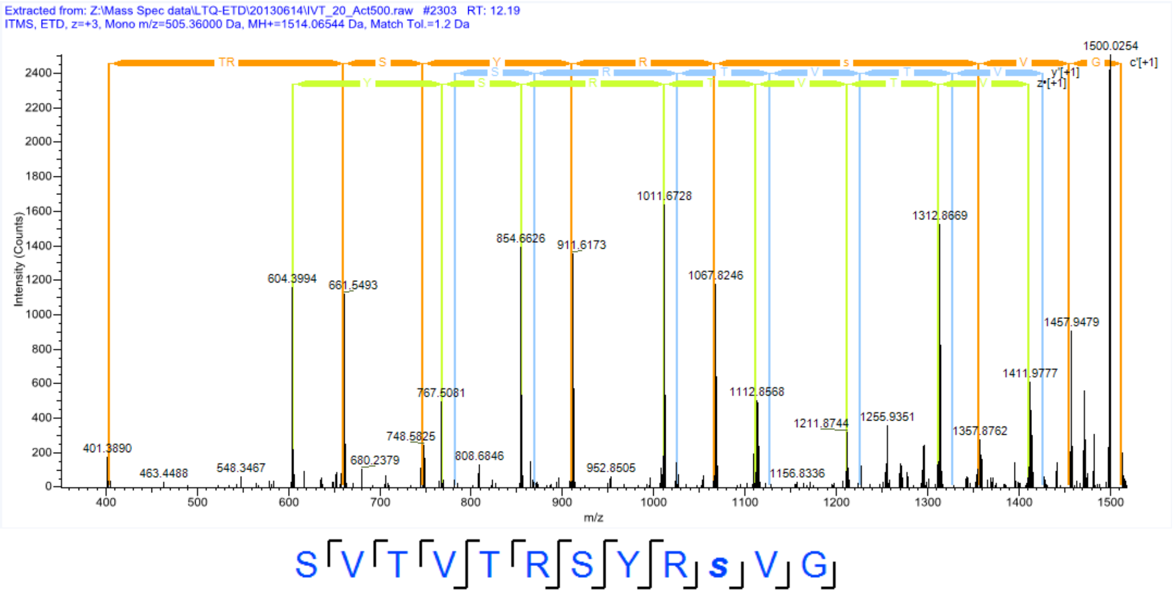
19. Lamin B1



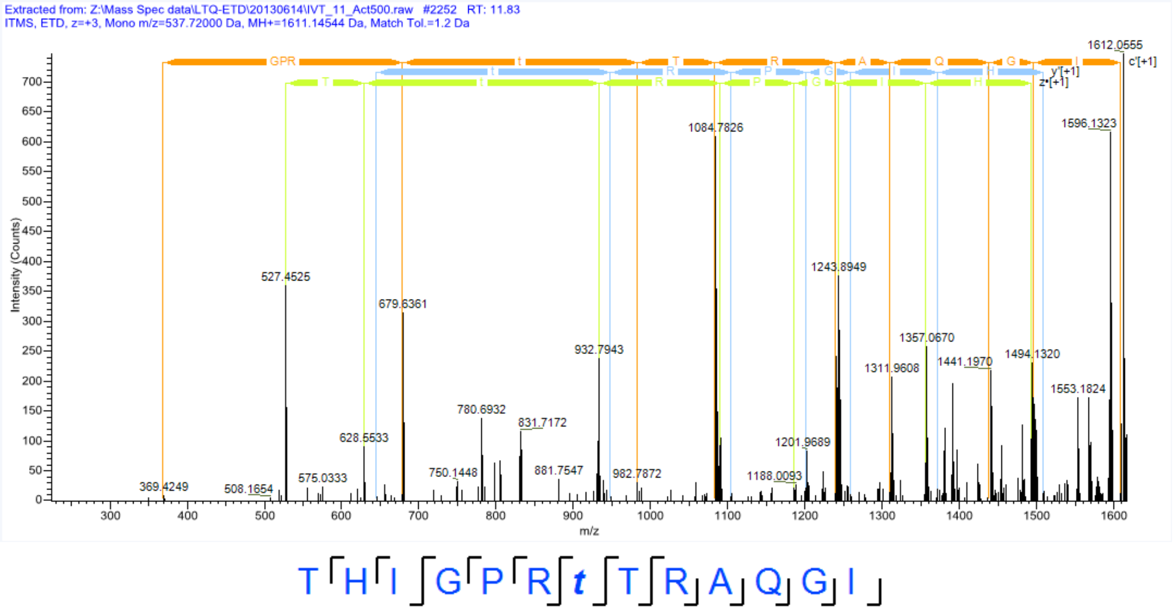
20. Lamin A



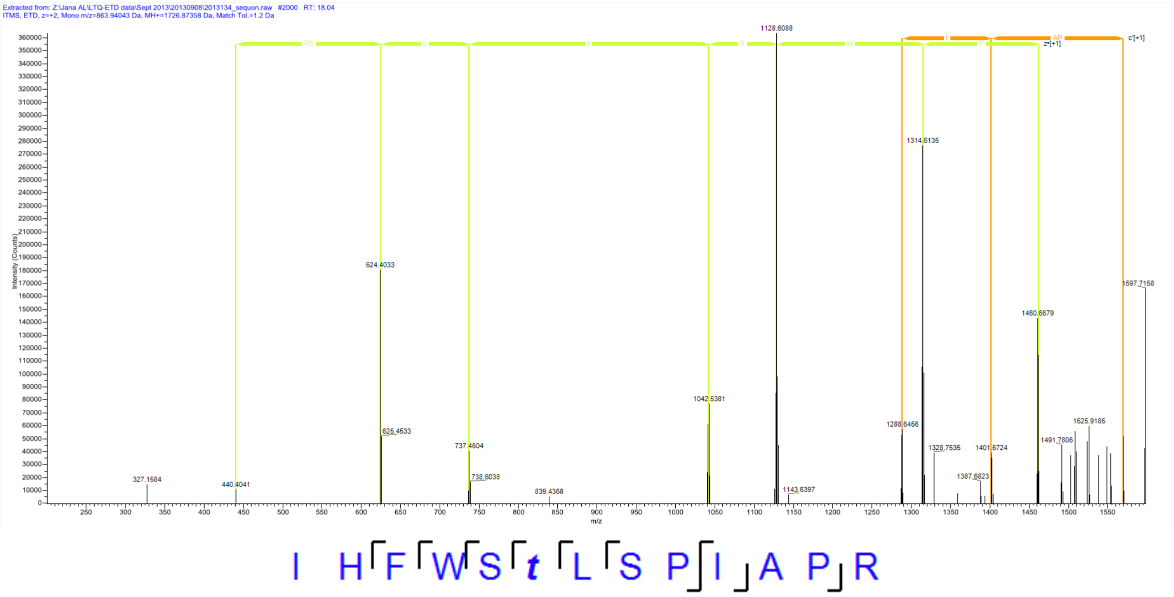

Extracted from: Z:\Mass Spec data\LTQ-ETD\20130614\IVT_20_Act500.raw #2306 RT: 12.23
ITMS, ETD, z=+3, Mono m/z=534.36000 Da, MH+=1601.06544 Da, Match Tol.=1.2 Da
Intensity (Counts)
m/z
c'[+1]
y[+1]
z•[+1]
1586.1774
1603.1887
1201.9041
748.6253
401.3945
854.6700
1312.9990
1545.1596
1358.0443
1499.1295
1011.7609
1413.1097
1254.9603
564.5018
1155.8456
652.5485
793.8086
998.9241
246.3290
345.4097
433.6803
1635.1624
S S V T V T R s Y R S V G


Extracted from: Z:\Mass Spec data\LTQ-ETD\20130614\IVT_20_Act500.raw #2303 RT: 12.19
ITMS, ETD, z=+3, Mono m/z=505.36000 Da, MH+=1514.06544 Da, Match Tol.=1.2 Da
Intensity (Counts)
m/z
c'[+1]
y[+1]
z•[+1]
1500.0254
1011.6728
1312.8669
854.6626
911.6173
604.3994
661.5493
1067.8246
1457.9479
1411.9777
767.5081
1112.8568
1211.8744
1255.9351
1357.8762
748.5825
808.6846
680.2379
401.3890
463.4488
548.3467
952.8505
1156.8336
S V T V T R S Y R s V G

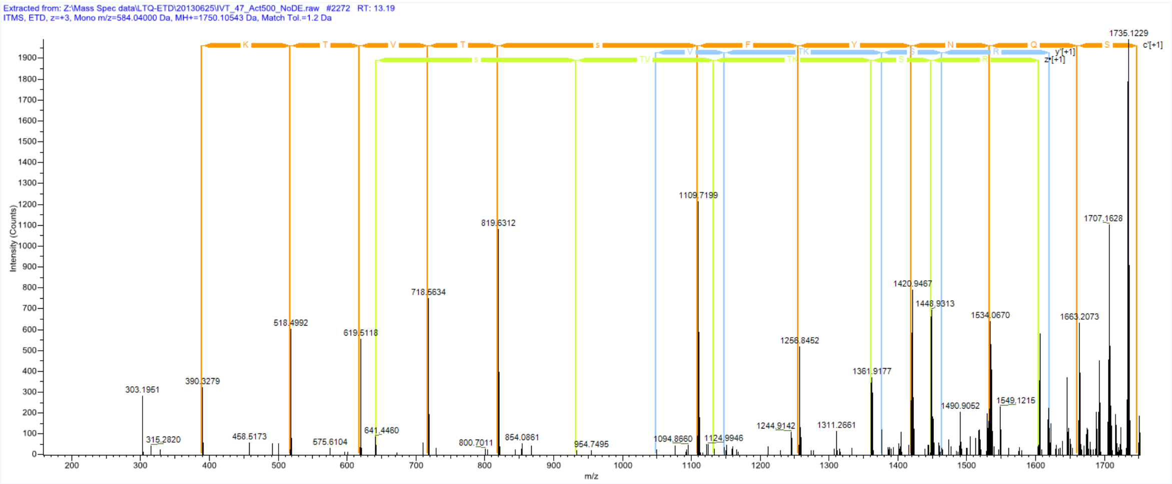
21. CGKI



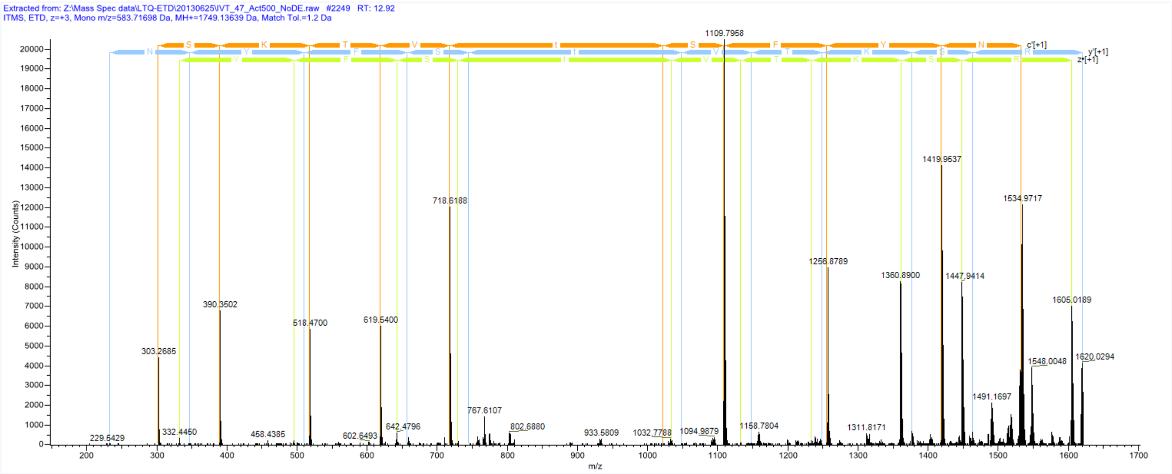
22. ELK1



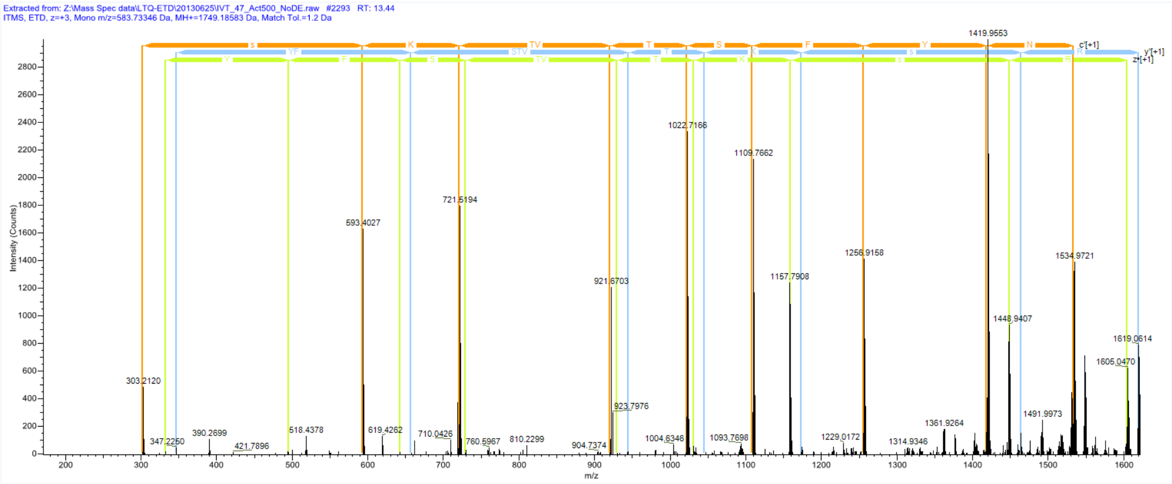
23. BCKD kinase



E R S K T V T ***s*** F Y N Q S

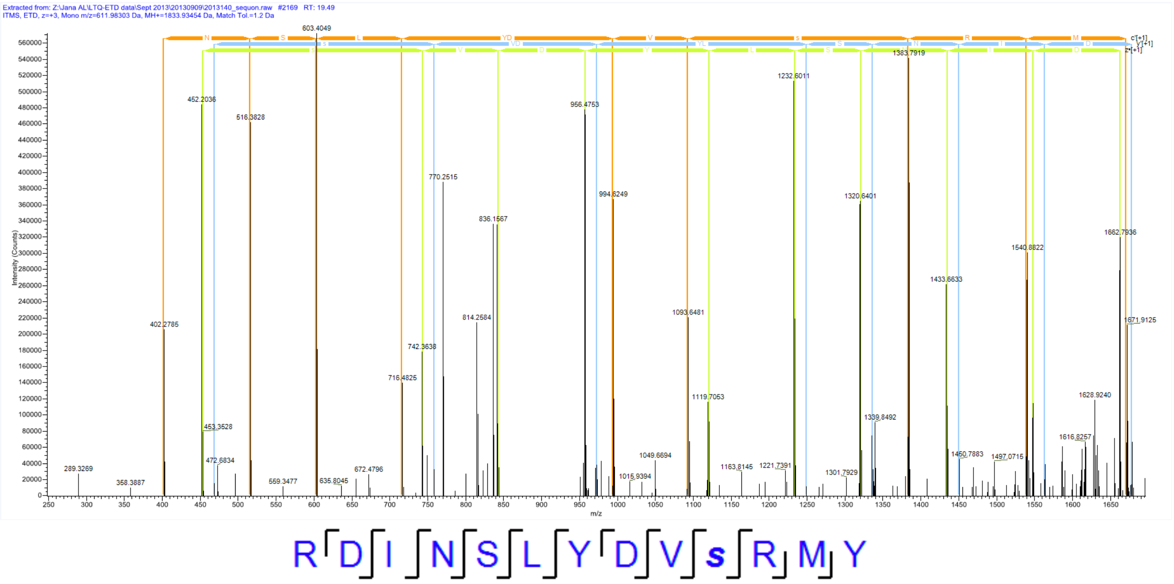


E R S K T V ***t*** S F Y N Q S

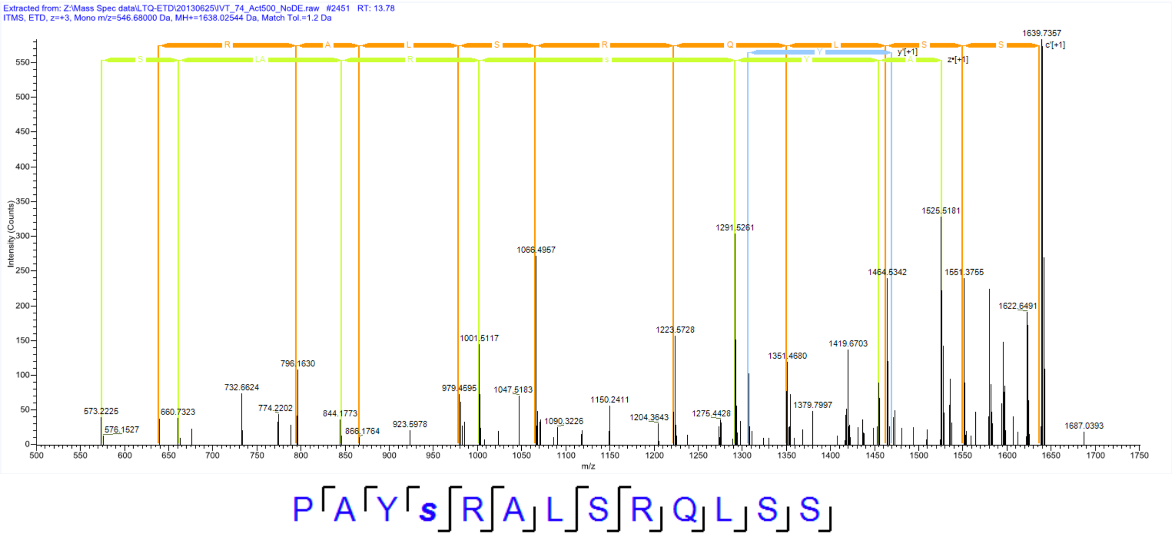


E R ***s*** K T V T S F Y N Q S

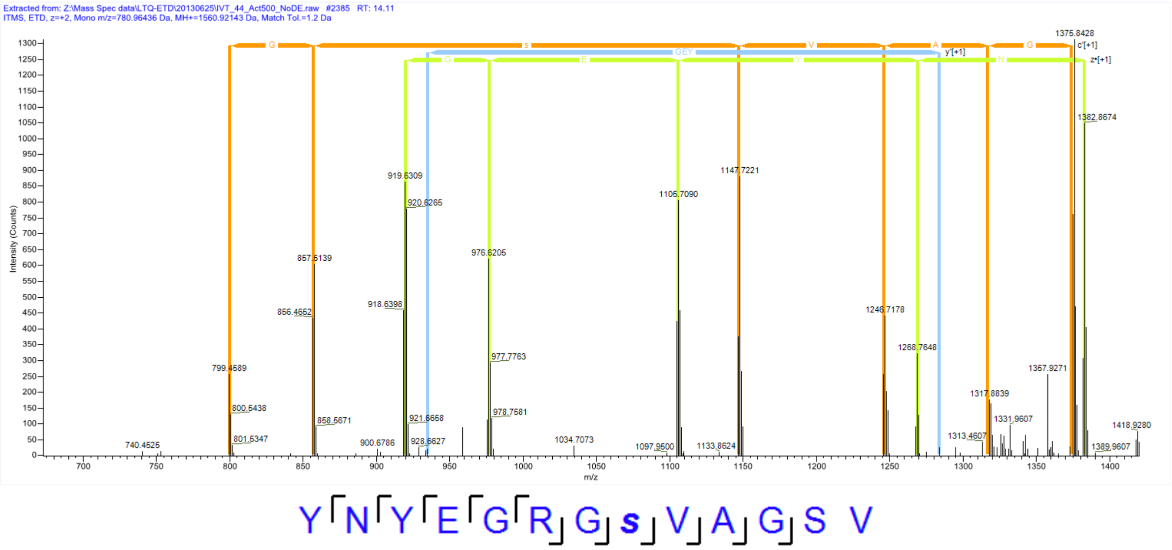
24. Phospholipase C-γ-2



25. HSP27



26. Desmocollin-3


Extracted from: Z:\Mass Spec data\LTQ-ETD\20130625\IVT_44_Act500_NoDE.raw #2385 RT: 14.11
ITMS, ETD, z=+2, Mono m/z=780.96436 Da, MH+=1560.92143 Da, Match Tol.=1.2 Da

27. DOUBLIN

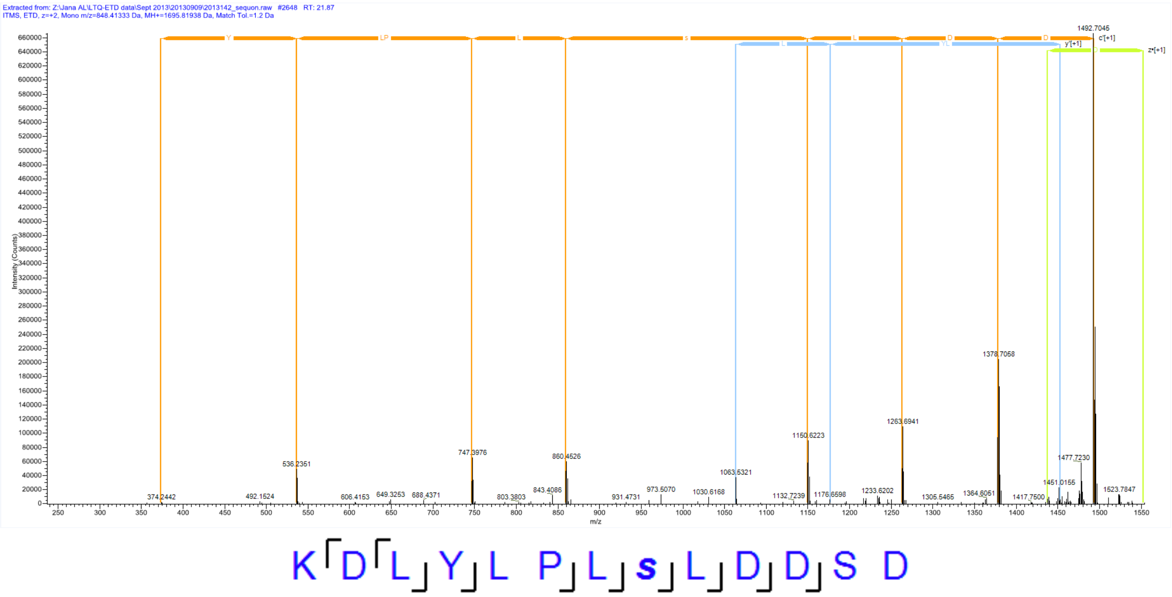
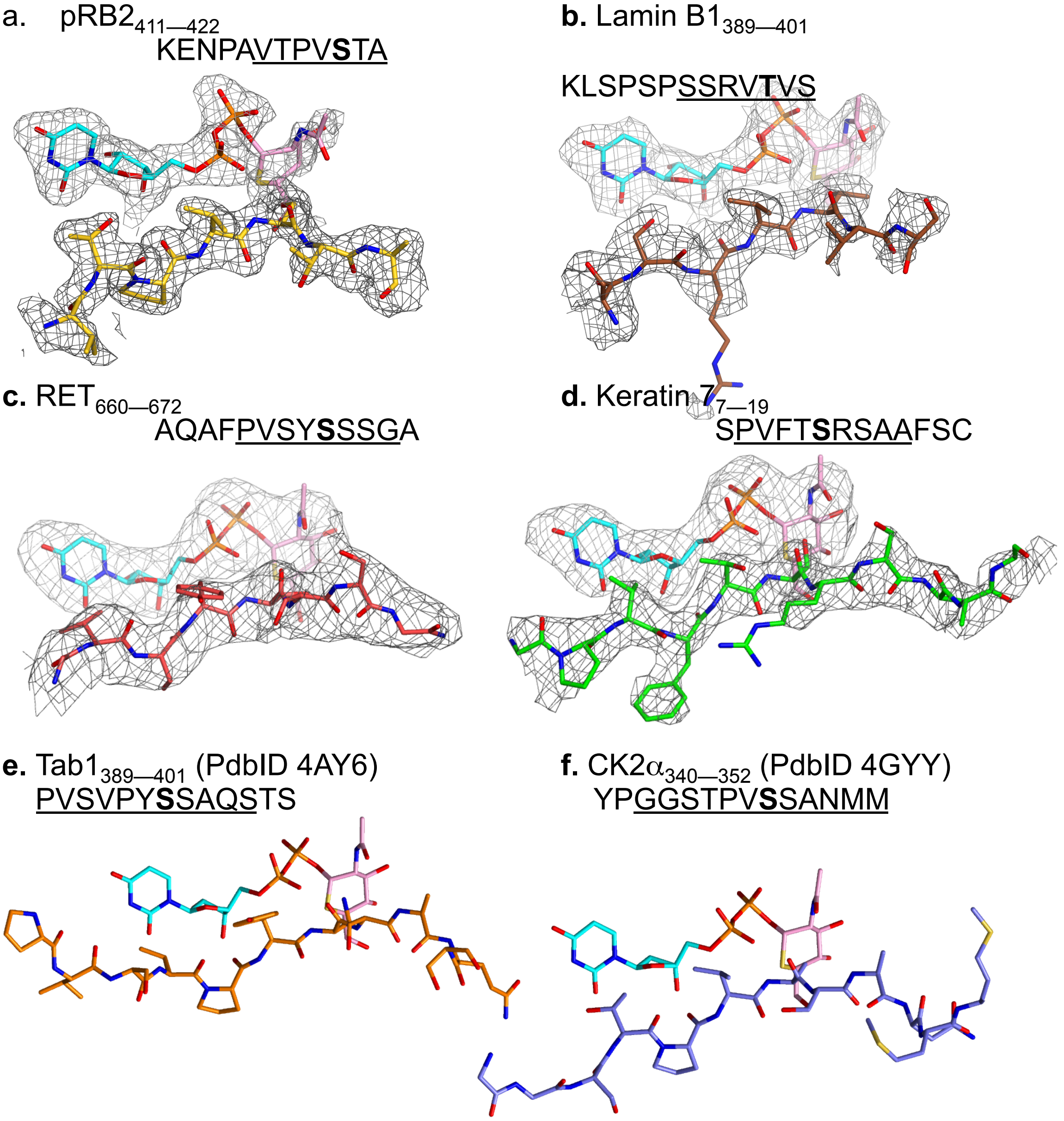

Extracted from: Z:\Jana AL\LTQ-ETD data\Sept 2013\20130909\2013142_sequon.raw #2648 RT: 21.87
ITMS, ETD, z=+2, Mono m/z=848.41333 Da, MH+=1695.81938 Da, Match Tol.=1.2 Da

Table S2.

X-ray diffraction data collection and structure refinement statistics. Values for the highest resolution shell are given in brackets.

	hOGT + UDP-5SGlcNAc + RB-like 2 (411—422)	hOGT + UDP-5SGlcNAc + keratin-7 (7—19)	hOGT + UDP-5SGlcNAc + Ret (660—672)	hOGT + UDP-5SGlcNAc + lamin B1 (179—191)
Data collection				
Beamline, wavelength	ID30A-3, 0.9677 Å	ID23-2, 0.873 Å	I04-1, 0.922 Å	I04-1, 0.922 Å
Space group	*F*222	*P*321	*P*321	*F*222
Cell dimensions (Å)	*a*=138.53, *b*=151.61, *c*=200.43	*a*=*b*=275.1, *c*=143.1	*a*=*b*=274.6, *c*=142.3	*a*=138.18, *b*=150.18, *c*=199.24
Resolution (Å)	46.20-2.05 (2.16-2.05)	25.00-3.15 (3.32-3.15)	91.26-3.38 (3.56-3.38)	30.00-2.40 (2.53-2.40)
R_{merge}	0.070 (0.826)	0.122 (0.508)	0.144 (0.674)	0.104 (0.719)
$I/\sigma I$	17.4 (2.5)	5.5 (1.6)	7.4 (2.1)	11.8 (2.4)
$CC_{1/2}$	0.99 (0.85)	0.99 (0.91)	0.97 (0.60)	0.99 (0.82)
R_{meas}	0.076 (0.897)	0.172 (0.714)	0.197 (0.887)	0.114 (0.78)
R_{pim}	0.029 (0.349)	0.121 (0.502)	0.100 (0.449)	0.044 (0.30)
Completeness (%)	100 (100)	99.4 (99.4)	99.4 (99.1)	99.9 (100)
Redundancy	6.7 (6.6)	1.7 (1.7)	3.4 (3.0)	6.5 (6.7)
Refinement	46.20—2.05 Å	25.00—3.15 Å	30.00—3.38 Å	30.00-2.40
No. total reflections	443970	183993	284765	263631
No. unique reflections	65809	106474	84183	40463
R_{work}, R_{free}	0.195 / 0.229	0.190 / 0.217	0.193 / 0.222	0.187 / 0.235
No. atoms				
Protein	5516	22140	22056	5489
Nucleotide sugar	39	156	156	39
Peptide	47	292	205	50
B-factor average				
Protein	44.75	54.43	70.84	45.17
Nucleotide sugar	36.11	46.01	66.81	31.43
Peptide	44.13	69.76	95.12	60.43
R.m.s. deviations				
Bond lengths (Å)	0.010	0.095	0.075	0.012
Bond angles (°)	1.381	1.31	1.19	1.49
Pdb ID	4XIE	4XIF	4XI9	5BNW

Figure S2. a-b Unbiased F_o-F_c difference electron density for ligands (UDP-5S-GlcNAc and peptide) contoured at 2.25 σ. **c-d** Unbiased, NCS-averaged F_o-F_c difference electron density for ligands (UDP-5S-GlcNAc and peptide) contoured at 3.5 σ. **e,f** Previously reported OGT substrate complexes. The entire sequence of the peptides used in the study is given; underlined residues are represented in the final model.


a. pRB2$_{411—422}$ KENPAVTPV**S**TA

b. Lamin B1$_{389—401}$ KLSPSPSSRV**T**VS

c. RET$_{660—672}$ AQAFPVSY**S**SSGA

d. Keratin 7$_{7—19}$ SPVFT**S**RSAAFSC

e. Tab1$_{389—401}$ (PdbID 4AY6) PVSVPY**S**SAQSTS

f. CK2$\alpha_{340—352}$ (PdbID 4GYY) YPGGSTPV**S**SANMM

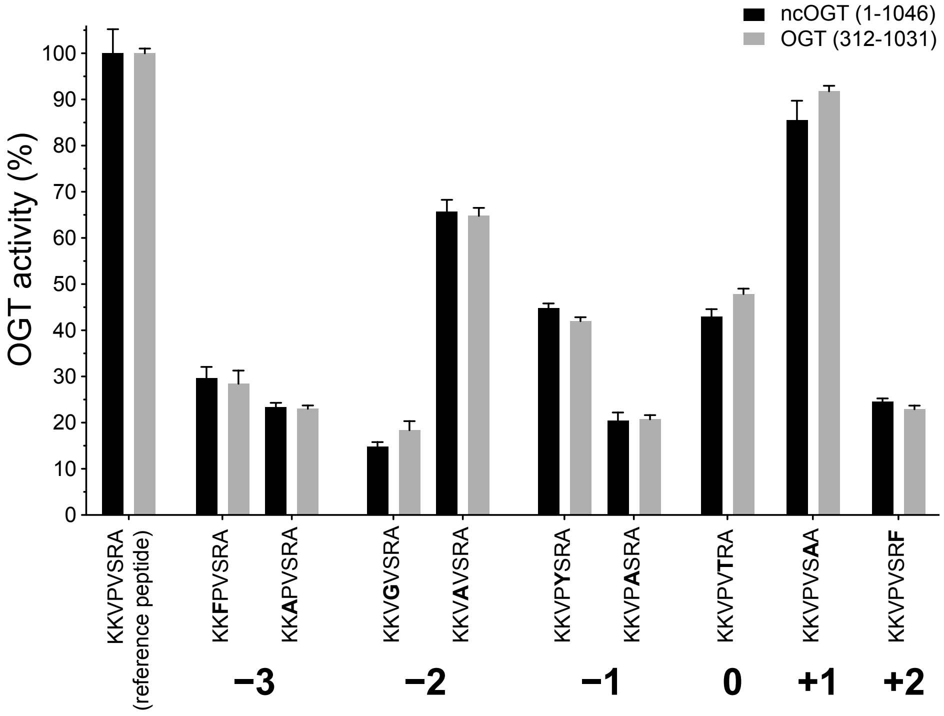
Figure S3.

List of 32 hexapeptides derived from the peptide library hits used to generate the sequon for Fig. 4.

Protein	Hexapeptide																			
RBL2_HUMAN	K	E	N	S	P	A	V	T	P	V	S	T	A							
MPP9_HUMAN	T	K	R	E	I	M	L	T	P	V	T	V	A							
ELK1_HUMAN	T	L	T	P	V	L	L	T	P	S	S	L	P	S	I					
FOXO1					T	F	R	P	R	T	S	S	N	A	S	T	I			
ETS1	A	A	D	V	P	L	L	T	P	S	S	K	E							
α-crystallin B chain								F	P	T	S	T	S	L	S	P	F	Y	L	R
Keratin 7				S	P	V	F	T	S	R	S	A	A	F	S	G				
RET			A	Q	A	F	P	V	S	Y	S	S	S	G	A					
MYB-related B protein	L	D	S	A	N	S	L	T	P	K	S	T	P							
SCG	E	L	I	L	K	P	P	S	P	I	S	E	A							
RBL2_HUMAN	G	L	G	R	S	I	T	S	P	T	T	L	Y							
GSK3β	R	G	E	P	N	V	S	Y	I	A	S	R	Y							
					R	G	E	P	N	V	S	Y	I	A	S	R	Y			
Kv2.1	K	W	T	K	R	T	L	S	E	T	S	S	S							
						K	W	T	K	R	T	L	S	E	T	S	S	S		
MPP9_HUMAN								T	P	V	T	V	A	Y	S	P	K	R	S	P
Early E1A32kD	A	I	L	R	R	P	T	S	P	V	S	R	E							
Rhodopsin-OPSD	A	S	A	T	V	S	K	T	E	T	S	Q	V							
						A	S	A	T	V	S	K	T	E	T	S	Q	V		
Insulin receptor substrate-1								V	P	S	S	R	G	D	Y	M	T	M	Q	M
Lamin B1	K	L	S	P	S	P	S	S	R	V	T	V	S							
Lamin A						S	S	V	T	V	T	R	S	Y	R	S	V	G		
				S	S	V	T	V	T	R	S	Y	R	S	V	G				
	S	S	V	T	V	T	R	S	Y	R	S	V	G							
CGKI					T	H	I	G	P	R	T	T	R	A	Q	G	I			
ELK1						I	H	F	W	S	T	L	S	P	I	A	P	R		
BCKD kinase				E	R	S	K	T	V	T	S	F	Y	N	Q	S				
					E	R	S	K	T	V	T	S	F	Y	N	Q	S			
Phospholipase C-γ-2		R	D	I	N	S	L	Y	D	V	S	R	M	Y						
HSP27								P	A	Y	S	R	A	L	S	R	Q	L	S	S
Desmocollin-3				Y	N	Y	E	G	R	G	S	V	A	G	S	V				
DOUBLIN				K	D	L	Y	L	P	L	S	L	D	D	S	D				

Figure S4.

Tolerance of different OGT isoforms for single amino acid substitutions.



OGT activity on the reference peptide KKVPVSRA was measured with two different constructs of the enzyme possessing a different number of TPR repeats. Nucleocytoplasmic OGT (ncOGT) is the longest (full length) natural OGT isoform, whereas the truncated construct, OGT (312—1031), was used for crystallographic studies and library screening due to its increased stability. The reference peptide KKVPVSRA represents the optimal OGT hexapeptide sequon except for position −3, where Val was used in order to avoid a potential second O-GlcNAc acceptor. Two N-terminal Lys residues were added, in order to aid peptide solubility. Assay details are given in the Online Methods section. The average of three measurements is shown, with error bars depicting the s.e.m. Activity for each enzyme isoform was normalized to the reference peptide.