Supporting Information

SpecOMS: a full open modification search method performing all-to-all spectra comparisons within minutes

Matthieu David[12], Guillaume Fertin[1], Hélène Rogniaux[2], Dominique Tessier[2]

[1]LS2N UMR CNRS 6004, Université de Nantes, F-44300 Nantes, France

[2]INRA, UR 1268 Biopolymères Interactions Assemblages, F-44316 Nantes

This supplementary data reports the impact of the number of the most intense masses selected in experimental spectra on the SpecOMS spectra interpretations; it also illustrates the behaviour of SpecOMS with a set of PSMs. These PSMs are shown with screenshots obtained by the interactive viewer SeeMS included in the cross-platform toolkit ProteoWizard[1]. The number of shared masses that can be seen on the screenshots sometimes slightly differs from the score given by SpecOMS because SpecOMS only considers a limited number of the most intense masses of the experimental spectra in its computations.

[1] Chambers, M. C., Maclean, B., Burke, R., Amodei, D., Ruderman, D. L., Neumann, S., Gatto, L., Fischer, B., Pratt, B., Egertson, J., Hoff, K., Kessner, D., Tasman, N., Shulman, N., Frewen, B., Baker, T. A., Brusniak, M. Y., Paulse, C., Creasy, D., Flashner, L., Kani, K., Moulding, C., Seymour, S. L., Nuwaysir, L. M., Lefebvre, B., Kuhlmann, F., Roark, J., Rainer, P., Detlev, S., Hemenway, T., Huhmer, A., Langridge, J., Connolly, B., Chadick, T., Holly, K., Eckels, J., Deutsch, E. W., Moritz, R. L., Katz, J. E., Agus, D. B., MacCoss, M., Tabb, D. L., and Mallick, P. (2012) A cross-platform toolkit for mass spectrometry and proteomics, Nat Biotechnol 30, 918-920.

1. Table. S-1: Impact of the number of selected masses on SpecOMS interpretations S-1

2. Fig. S-1: SpecOMS provides insights on the PTMs repartition in the sample S-4

3. Fig. S-2: SpecOMS identifies peptide variants S-5

4. Fig. S-3: SpecOMS identifies unusual PTMs S-6

5. Fig. S-4: SpecOMS identifies charge error assignments S-7

6. SpecOMS_results.xlsx: results produced from the analysis of the HEK293 dataset by SpecOMS S-8

1. Table. S-1: Impact of the number of selected masses on SpecOMS interpretations

Table. S-1 Evolution of the number of PSMs obtained by SpecOMS for the HEK293 dataset with a fragment mass accuracy of 0.02 Da according to k, the maximum number of selected masses in the experimental spectra. The PSMs are grouped by categories depending on the mass delta Δ_m. ST refers to PSMs with semi-tryptic peptides, and MC refers to PSMs with missed cleavage peptides. PSMs in the target database are in bold, and PSMs in the decoy database are in italics.

- k = 30:

Score	8	9	10	11	12	13	14	15	16+	Total
$\Delta_m = 0$	**1635**	**1422**	**1256**	**1092**	**790**	**485**	**307**	**146**	**98**	**7231**
	24	*12*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*36*
ST	**0**	**27**	**24**	**21**	**18**	**20**	**7**	**2**	**2**	**121**
	0	*2*	*1*	*0*	*0*	*0*	*0*	*0*	*0*	*3*
MC	**0**	**22**	**44**	**51**	**54**	**54**	**40**	**30**	**29**	**324**
	0	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*
$\Delta_m > 0$	**917**	**647**	**531**	**430**	**315**	**227**	**109**	**47**	**32**	**3255**
	536	*117*	*27*	*10*	*4*	*0*	*2*	*0*	*0*	*696*
$\Delta_m < 0$	**428**	**276**	**155**	**108**	**77**	**37**	**34**	**12**	**15**	**1142**
	172	*103*	*60*	*29*	*16*	*13*	*3*	*1*	*1*	*398*

- k = 40:

Score	8	9	10	11	12	13	14	15	16+	Total
$\Delta_m = 0$	**1448**	**1402**	**1224**	**1190**	**983**	**729**	**562**	**378**	**438**	**8354**
	51	*13*	*3*	*0*	*0*	*0*	*0*	*0*	*0*	*67*
ST	**0**	**19**	**28**	**42**	**27**	**21**	**15**	**11**	**4**	**167**
	0	*4*	*1*	*1*	*0*	*0*	*0*	*0*	*0*	*6*
MC	**0**	**18**	**45**	**77**	**89**	**102**	**80**	**52**	**111**	**574**
	0	*1*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*1*
$\Delta_m > 0$	**1329**	**885**	**667**	**600**	**453**	**359**	**251**	**170**	**160**	**4874**
	1439	*24*	*92*	*31*	*12*	*2*	*1*	*0*	*2*	*1903*
$\Delta_m < 0$	**643**	**434**	**267**	**203**	**130**	**79**	**56**	**35**	**44**	**1891**
	493	*293*	*191*	*104*	*47*	*28*	*9*	*4*	*3*	*1172*

- k = 50:

Score	8	9	10	11	12	13	14	15	16+	Total
$\Delta_m = 0$	**1322**	**1317**	**1218**	**1176**	**1056**	**870**	**660**	**493**	**873**	**8985**
	94	*27*	*7*	*0*	*0*	*0*	*0*	*0*	*0*	*128*
ST	**0**	**16**	**37**	**40**	**35**	**24**	**20**	**17**	**16**	**205**
	0	*3*	*3*	*1*	*0*	*0*	*0*	*0*	*0*	*7*
MC	**0**	**23**	**55**	**90**	**100**	**101**	**123**	**70**	**201**	**763**
	0	*1*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*1*
$\Delta_m > 0$	**1774**	**1056**	**819**	**698**	**542**	**475**	**339**	**253**	**358**	**6314**
	2703	*696*	*224*	*53*	*24*	*7*	*3*	*1*	*1*	*3712*
$\Delta_m < 0$	**865**	**684**	**421**	**317**	**203**	**110**	**88**	**49**	**82**	**2819**
	1066	*623*	*429*	*276*	*129*	*68*	*23*	*9*	*8*	*2631*

- k = 70:

Score	8	9	10	11	12	13	14	15	16+	Total
$\Delta_m = 0$	**1177**	**1199**	**1132**	**1152**	**1073**	**1009**	**808**	**600**	**1493**	**9643**
	215	*56*	*20*	*6*	*0*	*0*	*0*	*0*	*0*	*297*
ST	**0**	**28**	**34**	**55**	**40**	**37**	**33**	**19**	**38**	**284**
	0	*9*	*6*	*2*	*0*	*0*	*0*	*0*	*0*	*17*
MC	**0**	**16**	**79**	**87**	**100**	**138**	**148**	**118**	**356**	**1042**
	0	*2*	*1*	*0*	*0*	*0*	*0*	*0*	*0*	*3*
$\Delta_m > 0$	**2478**	**1393**	**1009**	**874**	**723**	**615**	**435**	**375**	**664**	**8566**
	5237	*1717*	*618*	*184*	*70*	*24*	*6*	*5*	*2*	*7863*
$\Delta_m < 0$	**1443**	**1246**	**854**	**608**	**349**	**243**	**134**	**89**	**162**	**8566**
	2401	*1898*	*1334*	*843*	*450*	*205*	*97*	*46*	*15*	*7289*

- k = 80:

Score	8	9	10	11	12	13	14	15	16+	Total
$\Delta_m = 0$	**1144**	**1151**	**1117**	**1130**	**1053**	**1047**	**860**	**625**	**1721**	**9848**
	275	*86*	*28*	*4*	*2*	*0*	*0*	*0*	*0*	*395*
ST	**0**	**31**	**32**	**58**	**50**	**36**	**38**	**23**	**46**	**314**
	0	*17*	*8*	*3*	*1*	*0*	*0*	*0*	*0*	*29*
MC	**0**	**19**	**62**	**95**	**103**	**144**	**152**	**137**	**417**	**1129**
	0	*1*	*2*	*1*	*0*	*0*	*0*	*0*	*0*	*4*
$\Delta_m > 0$	**2757**	**1507**	**1069**	**929**	**790**	**657**	**486**	**406**	**794**	**9395**
	5780	*2250*	*792*	*277*	*98*	*31*	*9*	*2*	*4*	*9243*
$\Delta_m < 0$	**1607**	**1569**	**1125**	**829**	**459**	**296**	**166**	**119**	**201**	**6371**
	2881	*2715*	*2003*	*1287*	*720*	*318*	*136*	*74*	*26*	*10160*

- k = 90:

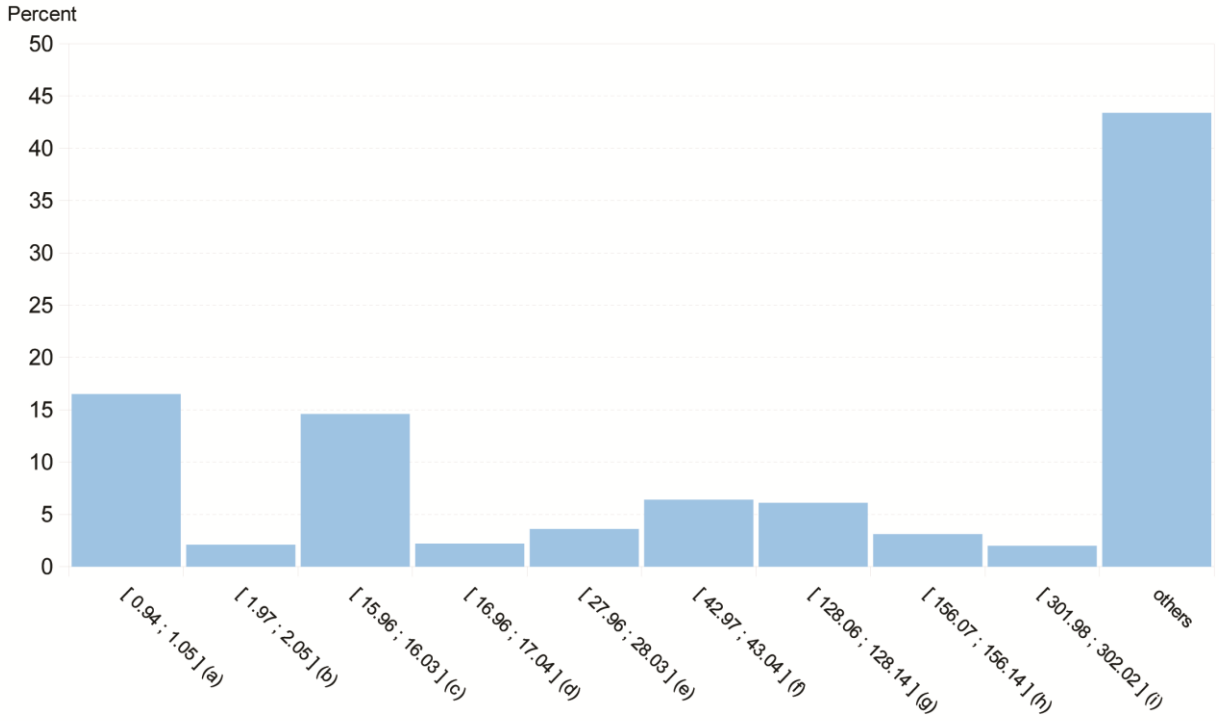
Score	8	9	10	11	12	13	14	15	16+	Total
$\Delta_m = 0$	**1103**	**1105**	**1118**	**1101**	**1051**	**1066**	**895**	**661**	**1890**	**9990**
	371	*112*	*32*	*9*	*2*	*0*	*0*	*0*	*0*	*526*
ST	**0**	**33**	**42**	**48**	**51**	**45**	**46**	**26**	**47**	**338**
	0	*23*	*8*	*6*	*1*	*0*	*0*	*0*	*0*	*38*
MC	**0**	**17**	**66**	**91**	**116**	**1133**	**153**	**165**	**485**	**1226**
	0	*4*	*1*	*1*	*1*	*0*	*0*	*0*	*0*	*7*
$\Delta_m > 0$	**2717**	**1608**	**1113**	**952**	**848**	**713**	**511**	**425**	**915**	**9802**
	5770	*2675*	*1020*	*393*	*138*	*46*	*16*	*4*	*3*	*10065*
$\Delta_m < 0$	**1616**	**1862**	**1449**	**1028**	**609**	**368**	**199**	**131**	**230**	**7492**
	2978	*3355*	*2720*	*1777*	*1010*	*500*	*208*	*105*	*41*	*12694*

- k = 100:

Score	8	9	10	11	12	13	14	15	16+	Total
$\Delta_m = 0$	**1070**	**1073**	**1135**	**1097**	**1033**	**1070**	**920**	**721**	**2014**	**10133**
	461	*136*	*39*	*12*	*1*	*1*	*0*	*0*	*0*	*650*
ST	**0**	**30**	**51**	**45**	**55**	**45**	**49**	**32**	**49**	**356**
	0	*31*	*10*	*5*	*3*	*0*	*0*	*0*	*0*	*49*
MC	**0**	**14**	**61**	**96**	**120**	**160**	**160**	**168**	**537**	**1316**
	0	*6*	*1*	*1*	*1*	*0*	*0*	*0*	*0*	*9*
$\Delta_m > 0$	**2429**	**1662**	**1161**	**988**	**851**	**776**	**536**	**450**	**1013**	**9866**
	5322	*2992*	*1217*	*468*	*170*	*63*	*26*	*4*	*4*	*10266*
$\Delta_m < 0$	**1511**	**2063**	**1715**	**1294**	**759**	**444**	**231**	**164**	**262**	**8443**
	2725	*3847*	*3448*	*2362*	*1378*	*717*	*299*	*142*	*55*	*14973*

2. Fig. S-1: SpecOMS provides insights on the PTMs repartition in the sample

Fig. S-1 Distribution of the positive mass deltas obtained by SpecOMS for the HEK293 dataset. The positive mass deltas with the exclusion of missed cleavage peptides are divided into mass intervals containing at least 2% of the PTMs. These mass intervals mainly correspond to the following PTMs: (a) deamidation or 13C, (b) 13C2, (c) oxidation, (d) oxidation+13C, (e) formylation, (f) carbamylation, (g) K, (h) R and (i) others.



3. Fig. S-2: SpecOMS identifies peptide variants

Fig S-2. SpecXtract suggests peptide TFEEDPAVGAIVITGGDK for scan 60396 with a mass delta of 12.0373 Da. SpecFit computes a large increase in score if the mass delta is positioned at the N-Terminal side of the suggested peptide (position 1), from 10 to 17. In Unimod, a mass delta of 12.0373 Da can correspond to a substitution from T to I/L. The correct identification is probably I/LFEEDPAVGAIVITGGDK instead of TFEEDPAVGAIVITGGDK

Fig S2_A Scan 60396, the PSM before SpecFit

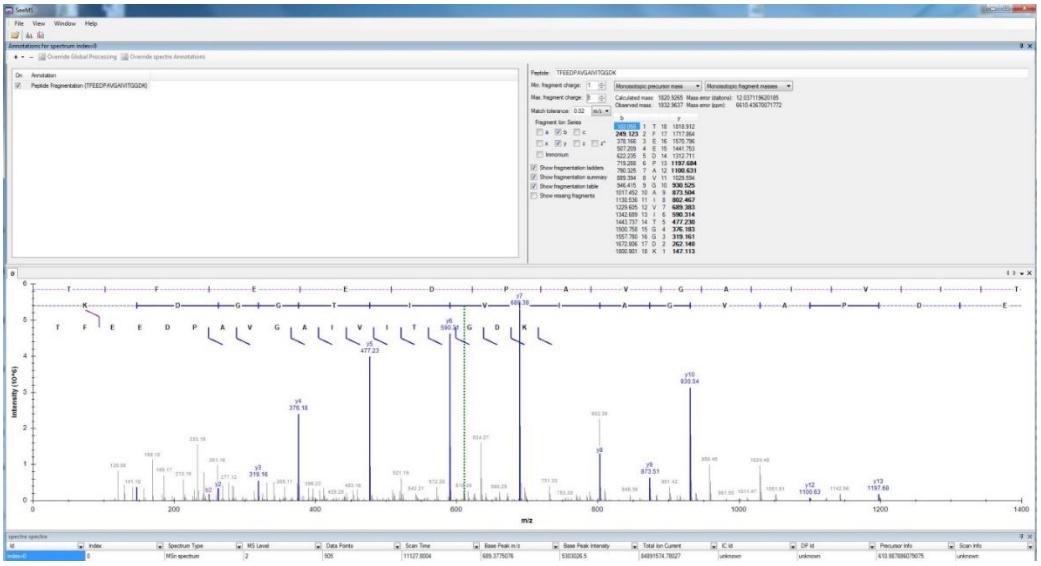
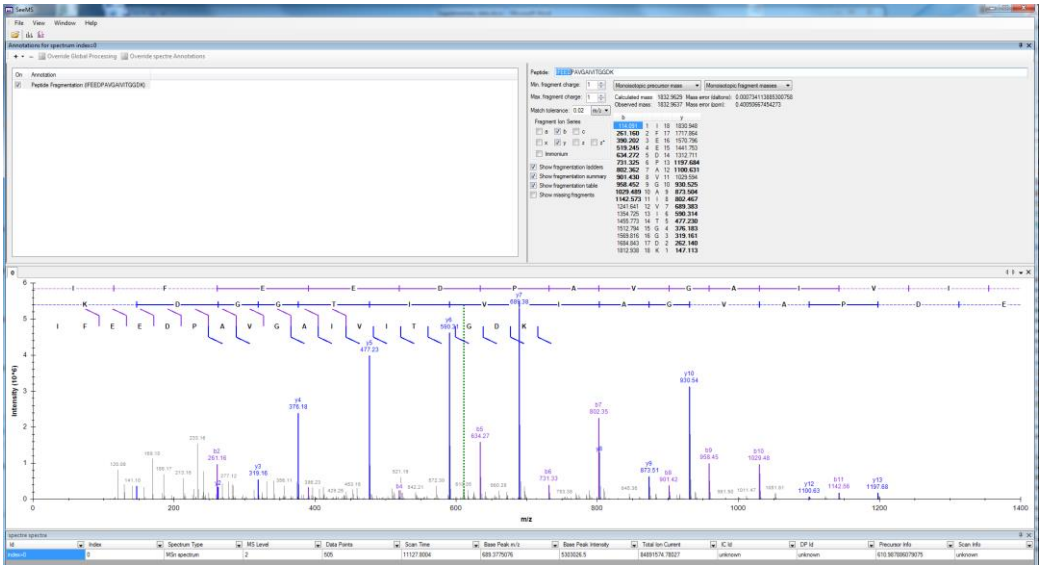


Fig S2_B Scan 60396, the PSM after SpecFit



4. Fig. S-3: SpecOMS identifies unusual PTMs

Fig S-3. SpecXtract suggests peptide NNASTDYDISDK for the scan 11241 with a mass delta of 21.977 Da. SpecFit computes a large increase in score - from 8 to 15- (mass delta is positioned at the position 5 of the peptide). In Unimod, a mass delta of 21.977 Da can correspond to a sodium adduct on D or E.

Fig S3_A Scan 11241, the PSM before SpecFit

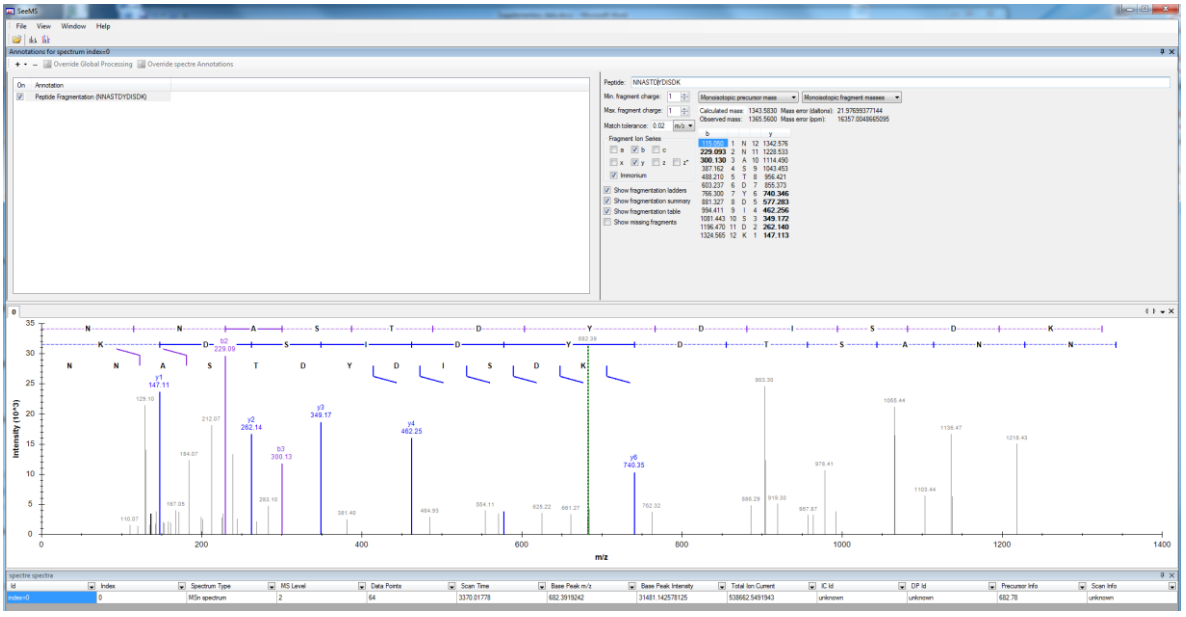
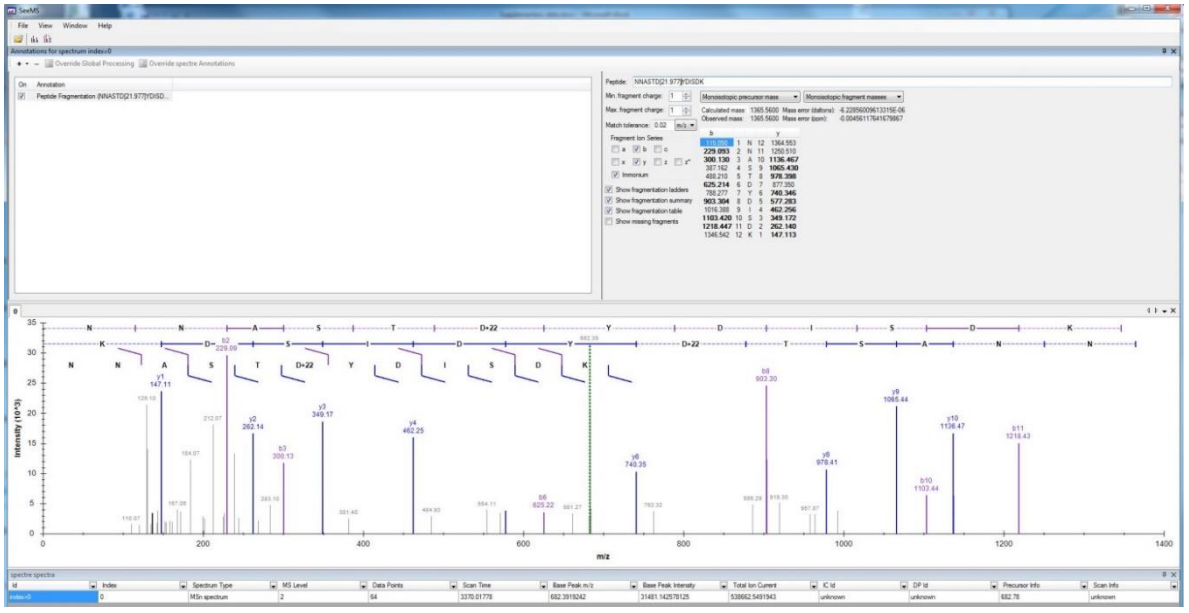


Fig S3_B Scan 11241, the PSM after SpecFit

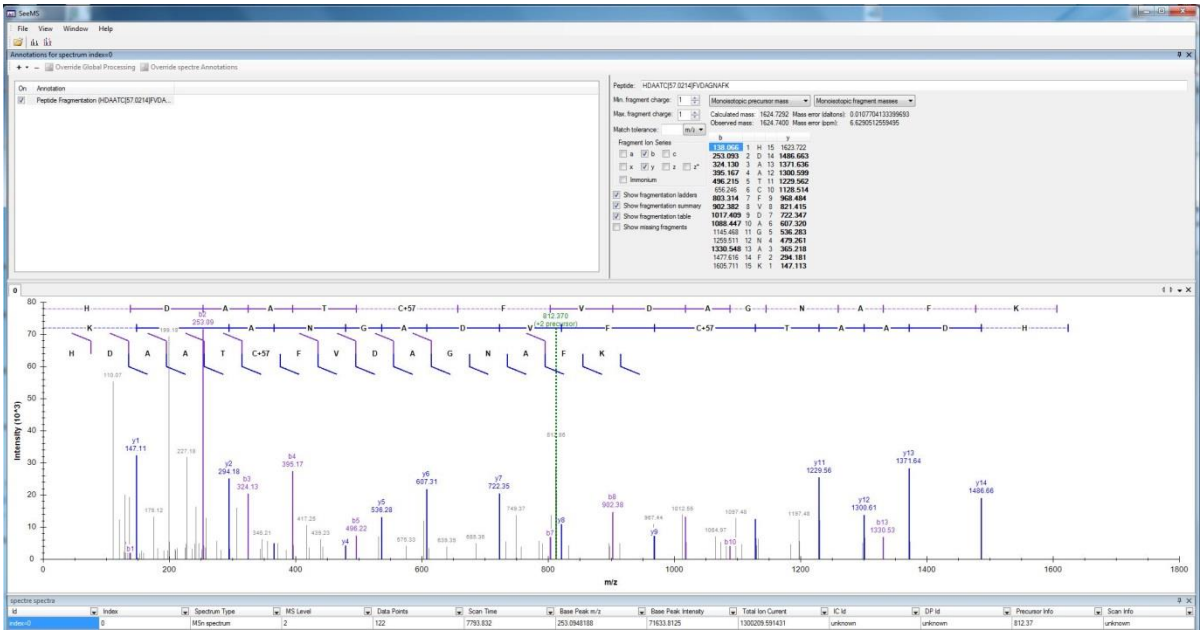


5. Fig. S-4: SpecOMS identifies charge error assignments

Fig S-4. SpecXtract suggests peptide HDAATCFVDAGNAFK for scan 38793 with a mass delta of 811.3735 Da. The peak picking made by RawConverter attributes a charge of 3 to this spectrum. However, a correction in the charge state of the spectrum explains the mass delta. In this case, SpecFit does not improve the score of 18 because there is no modification, but it suggests the correct PSM, even though the mass delta is very large.

Fig S4_A Scan 38793, the PSM with a spectrum charge = 3

Fig S4_B Scan 38793, the PSM with a spectrum charge = 2 (modification of the spectrum mgf file)



6. SpecOMS_results.xlsx: results produced from the analysis of the HEK293 dataset by SpecOMS

SpecOMS_results.xlsx. Provided as a separate file.