Supplementary Material

A new opening for the tricky untargeted investigation of natural and modified short peptides

Andrea Cerrato[1], Sara Elsa Aita[1], Anna Laura Capriotti[1*], Chiara Cavaliere[1], Carmela Maria Montone[1], Aldo Laganà,[1,2] Susy Piovesana[1]

1. Department of Chemistry, Università degli Studi di Roma La Sapienza, Piazzale Aldo Moro 5, 00185 Rome, Italy

2. CNR NANOTEC, Campus Ecotekne, University of Salento, Via Monteroni, 73100 Lecce, Italy

*__Corresponding author__:

Prof. Anna Laura Capriotti

Department of Chemistry, Sapienza Università di Roma, Piazzale Aldo Moro 5, 00185 Rome, Italy

E-mail: annalaura.capriotti@uniroma1.it

Telephone: (+39) 06 4991 3062

Fax (+39) 06 4906 31

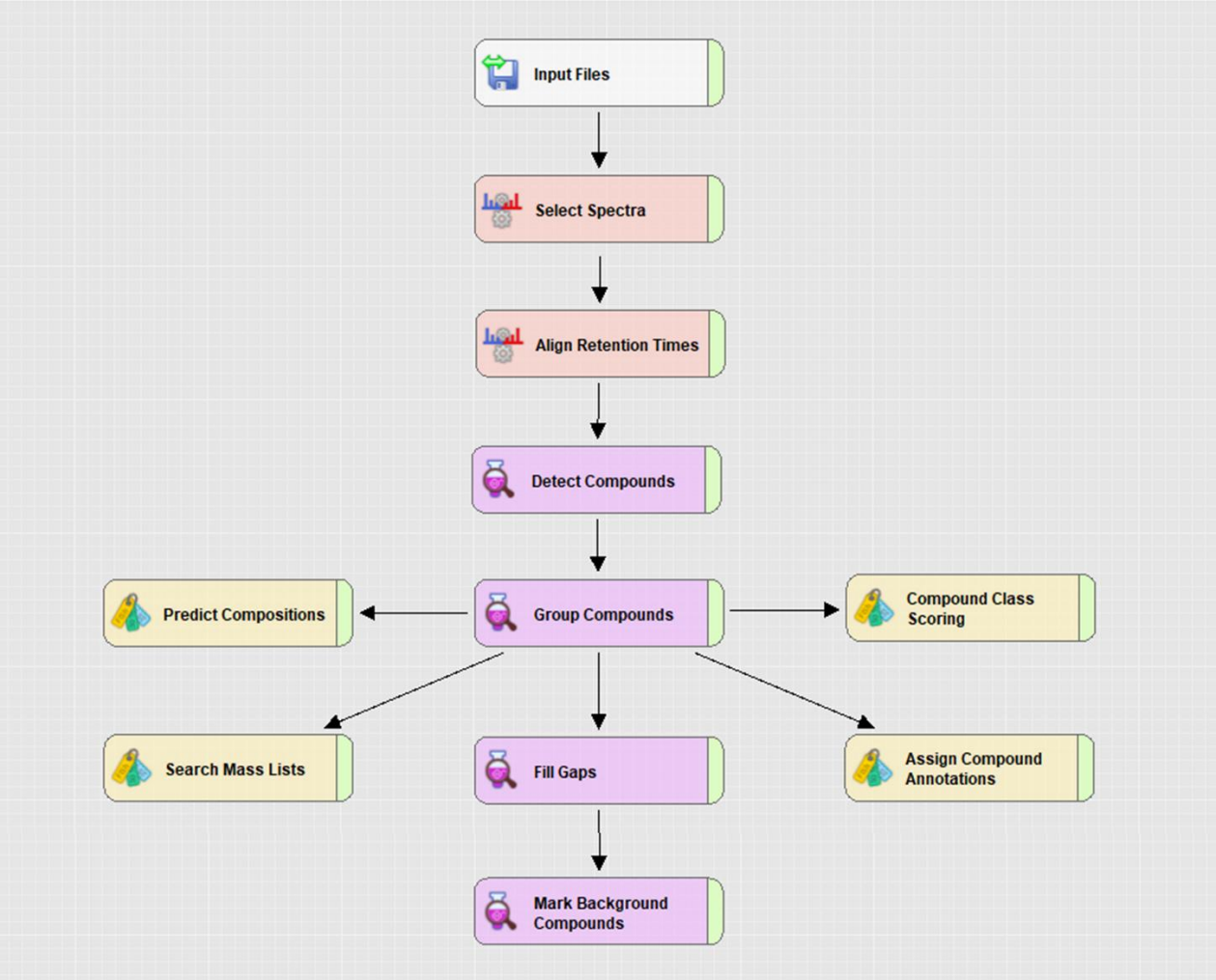

Figure S1. Exemplary data processing workflow for short peptide identification on Compound Discoverer

Select spectra	
Precursor Selection	Use MS(n-1) Precursor
Use Isotope Pattern in Precursor Reevaluation	True
Provide Profile Spectra	Automatic
Lower RT Limit	0
Upper RT Limit	25
First Scan	0
Min Precursor Mass	0 Da
Max Precursor Mass	750 Da
Minimum Peak Count	1
MS Order	Any
Polarity mode	Positive
S/N Threshold	1.5
Align Retention Time	
Alignment Model	Adaptive Curve
Alignment Fallback	Use Linear Model
Minimum Shift [min]	0.2
Shift Reference File	Ture
Mass Tolerance	5 ppm
Remove Outlier	True
Detect Compounds	
Mass Tolerance [ppm]	5 ppm
Intensity Tolerance [%]	30
S/N Threshold	3
Min. Peak Intensity	100000
Ions (positive mode)	[M + H]+1
Base Ions	[M + H]+1
Min. Element Counts	C H
Max. Element Counts	C50 H60 O15 N15
Filter Peaks	True
Max. Peak Width [min]	0.5
Min #Scans per Peak	5
Min #Isotopes	1
Group Compounds	

Mass Tolerance	5 ppm
RT Tolerance	0.2
Preferred Ions	[M+H]+1
Predict Composition	
Mass Tolerance	5 ppm
Min. Element Counts	C H
Max. Elements Counts	C50 H60 O15 N15
Min. RDBE	0
Max. RDBE	30
Min. H/C	0.1
Max. H/C	4
Intensity Tolerance [%]	30
Intensity Threshold [%]	0.1
S/N Threshold	3
Min. Spectral Fit [%]	30
Min. Pattern Cov. [%]	90
Use Dynamic Recalibration	True
Search Mass List	
Use Retention Time	False
Mass Tolerance	5 ppm
Compound Class Scoring	
S/N Threshold	10
High Acc. Mass Tolerance	10 ppm
Allow AIF Scoring	True
Fill Gaps	
Mass Tolerance	5 ppm
S/N Threshold	1.5
Use Real Peak Detection	True
Mark Background Compounds	
Max. Sample/Blank	5
Hide background	True

Table S1. Detailed Compound Discoverer parameters for short peptide data processing

	Ala	Arg	Asn	Asp	Cys
Formula	$C_3H_7NO_2$	$C_6H_{14}N_4O_2$	$C_4H_8N_2O_3$	$C_4H_7NO_4$	$C_3H_7NO_2S$
Molecular Weight	89.047679	174.111676	132.053493	133.037509	121.019751
$[M+H]^+$	90.054955	175.118952	133.060769	134.044785	122.027027
Iminium Ion (*a* ion)	44.0495	129.1135	87.0553	88.0393	76.0215
Im – H_2O	-	-	-	70.0287	-
Im – NH_3	-	112.0869	70.0287	-	-
b ion	72.0444	157.1084	115.0502	116.0342	104.0165
b – H_2O	-	-	-	98.0237	-
b – NH_3	55.0178	140.0818	98.0237	99.0077	86.9899
c ion	89.0709	174.1349	132.0768	133.0608	121.0430
x ion	116.0342	201.0982	159.0400	160.0241	148.0063
y ion	90.0550	175.1190	133.0608	134.0448	122.0270
y – H_2O	-	-	-	116.0342	-
y – NH_3 (*z ion*)	-	158.0924	116.0342	-	-
Others	-	70.0561; 116.0706	-	-	-
	Gly	Glu	Gln	His	Xle
Formula	$C_2H_5NO_2$	$C_5H_9NO_4$	$C_5H_9N_2O_3$	$C_6H_9N_3O_2$	$C_6H_{13}NO_2$
Molecular Weight	75.032029	147.053159	132.053493	133.037509	121.019751
$[M+H]^+$	76.039305	148.060435	133.060769	134.044785	122.027027
Iminium Ion (*a* ion)	30.0338	102.0550	101.0709	110.0713	86.0964
Im – H_2O	-	84.0444	-	-	-
Im – NH_3	-	-	84.0444	-	69.0699
b ion	58.0287	130.0499	129.0659	138.0662	114.0913
b – H_2O	-	112.0393	-	-	-
b – NH_3	-	113.0233	112.0393	121.0396	97.0648
c ion	75.0533	147.0764	146.0924	155.0927	131.1179
x ion	102.0186	174.0397	173.0557	182.0560	158.0812
y ion	76.0393	148.0604	147.0764	156.0768	132.1019

y – H_2O	-	130.0499	-	-	-
y – NH_3 (z ion)	-	-	130.0499	-	-
Others	-	-	-	-	-
	Lys	**Met**	**Phe**	**Pro**	**Ser**
Formula	$C_6H_{14}N_2O_2$	$C_5H_{11}NO_2S$	$C_9H_{11}NO_2$	$C_5H_7NO_2$	$C_3H_7NO_3$
Molecular Weight	146.105528	149.051051	165.078979	115.063329	105.042594
$[M+H]^+$	147.112804	150.058327	166.086255	116.070605	106.049870
Iminium Ion (*a* ion)	101.1073	104.0528	120.0808	70.0651	60.0444
Im – H_2O	-	-	-	-	42.0338
Im – NH_3	84.0808	-	103.0542	-	-
b ion	129.1022	132.0478	148.0757	98.0600	88.0393
b – H_2O	-	-	-	98.0237	-
b – NH_3	112.0757	-	-	99.0077	86.9899
c ion	146.1288	149.0743	165.1022	115.0866	105.0659
x ion	173.0921	176.0376	192.0655	142.0499	132.0291
y ion	147.1128	150.0583	166.0863	116.0706	106.0499
y – H_2O	129.1022	-	-	-	88.0393
y – NH_3 (z ion)	130.0863	133.0318	149.0597	-84.0808	-
Others	-	-	131.0941	-	-
	Thr	**Trp**	**Tyr**	**Val**	
Formula	$C_4H_9NO_3$	$C_{11}H_{12}N_2O_2$	$C_9H_{11}NO_3$	$C_5H_{11}NO_2$	
Molecular Weight	119.058244	204.089878	181.073894	117.078979	
$[M+H]^+$	120.06552	205.097154	182.081170	118.086255	
Iminium Ion (*a* ion)	74.0600	159.0917	136.0757	72.0808	
Im – H_2O	56.0495	-	-	-	
Im – NH_3	-	-	-	55.0542	
b ion	102.0550	187.0866	164.0706	100.0757	
b – H_2O	84.0444	-	146.0600	-	
b – NH_3	-	-	-	-	

c ion	119.0815	204.1131	181.0972	117.1022
x ion	146.0448	231.0764	208.0604	144.0655
y ion	120.0655	205.0972	182.0812	118.0863
y – H_2O	102.0550	-	-	-
y – NH_3 (z ion)	-	188.0706	165.0546	-
Others	-	170.0600; 146.0600	-	-

Table S2. Product ion deriving from natural amino acid fragmentations. Fragments in green were inserted in compound class scoring tool due to experimental data or spectra reported in literature. Fragments in red would have been inserted since were found to be typical but fall under the cut off at *m/z* 50 for MS/MS fragmentation

	Hydroxyproline (Hyp)	Hydroxylysine (Hyl)	Pyroglutamic acid (Pyr)	Citrullin (Cit)	Methylarginine (MeArg)
Formula	$C_5H_9NO_3$	$C_6H_{14}N_2O_3$	$C_5H_7NO_3$	$C_6H_{13}N_3O_3$	$C_7H_{16}N_4O_2$
Molecular Weight	131.058244	162.100442	129.042594	175.095691	188.127326
$[M+H]^+$	132.065520	163.107718	130.049870	176.102967	189.134602
Iminium Ion (*a* ion)	86.0600	117.1022	84.0444	130.0975	143.1291
Im – H_2O	68.0495	99.0917	-	112.0869	-
Im – NH_3	-	100.0757	-	113.0709	126.1026
b ion	114.0550	145.0972	112.0393	158.0924	171.1240
b – H_2O	96.0444	127.0866	-	140.0818	-
b – NH_3	97.0284	128.0706	-	141.0659	154.0975
c ion	131.0815	162.1237	129.0659	175.1190	188.1506
x ion	158.0448	189.0870	-	202.0822	215.1139
y ion	132.0655	163.1077	-	176.1030	189.1346
y – H_2O	114.0550	145.0972	-	158.0924	-
y – NH_3 (z ion)	-	146.0812	-	159.0764	172.1081
Others	-	-	-	-	112.0869; 70.0561
	Methyllysine (MeLys)	Dimethyllysine (DiMeLys)	Trimethyllysine (TriMeLys)	Acetyllysine (AcLys)	Succinyllysine (SucLys)

Formula	$C_7H_{16}N_2O_2$	$C_8H_{18}N_2O_2$	$C_9H_{21}N_2O_2^+$	$C_8H_{16}N_2O_3$	$C_{10}H_{18}N_2O_5$
Molecular Weight	160.121178	174.136828	189.159754	188.116093	246.121573
$[M+H]^+$	161.128454	175.144104	189.159754	189.123369	247.128849
Iminium Ion (*a* ion)	115.1230	129.1386	143.1543	143.1179	201.1234
Im – H_2O	-	-	-	-	-
Im – NH_3	98.0964	112.1121	126.1277	126.0913	184.0968
b ion	143.1179	157.1335	171.1492	171.1128	229.1183
b – H_2O	-	-	-	-	-
b – NH_3	126.0913	140.1070	154.1226	154.0863	212.0917
c ion	160.1444	174.1601	188.1757	188.1394	246.1448
x ion	187.1077	201.1234	215.1390	188.1394	246.1448
y ion	161.1285	175.1441	189.1598	189.1234	247.1288
y – H_2O	-	-	-	-	-
y – NH_3 (*z ion*)	-	-	-	-	-
Others	130.0863; 84.0808	130.0863; 84.0808	130.0863; 84.0808	130.0863; 84.0808	130.0863; 84.0808
	Methionine sulfoxide (Mes)	**Tyrosine O-sulfate (sTyr)**	**Serine O-phosphate (pSer)**		
Formula	$C_5H_{11}NO_3S$	$C_9H_{11}NO_6S$	$C_3H_8NO_6P$		
Molecular Weight	165.045966	261.030711	185.008890		
$[M+H]^+$	166.053242	262.037987	186.016166		
Iminium Ion (*a* ion)	120.0478	216.0325	140.0107		
Im – H_2O	-	-	-		
Im – NH_3	103.0212	-	-		
b ion	148.0427	244.0274	168.0056		
b – H_2O	-	-	-		
b – NH_3	131.0161	-	-		
c ion	165.0692	261.0540	185.0322		
x ion	192.0325	288.0173	211.9954		
y ion	166.0532	262.0380	186.0162		
y – H_2O	-	-	-		

y – NH_3 (z ion)	149.0267	245.0114	168.9896
Others	56.0495; 84.0444; 102.0550	136.0757; 165.0546; 182.0812	60.0444; 106.0499

Table S3. Product ion deriving from modified amino acid fragmentations. Fragments in green were inserted in compound class scoring tool due to experimental data or spectra reported in literature.

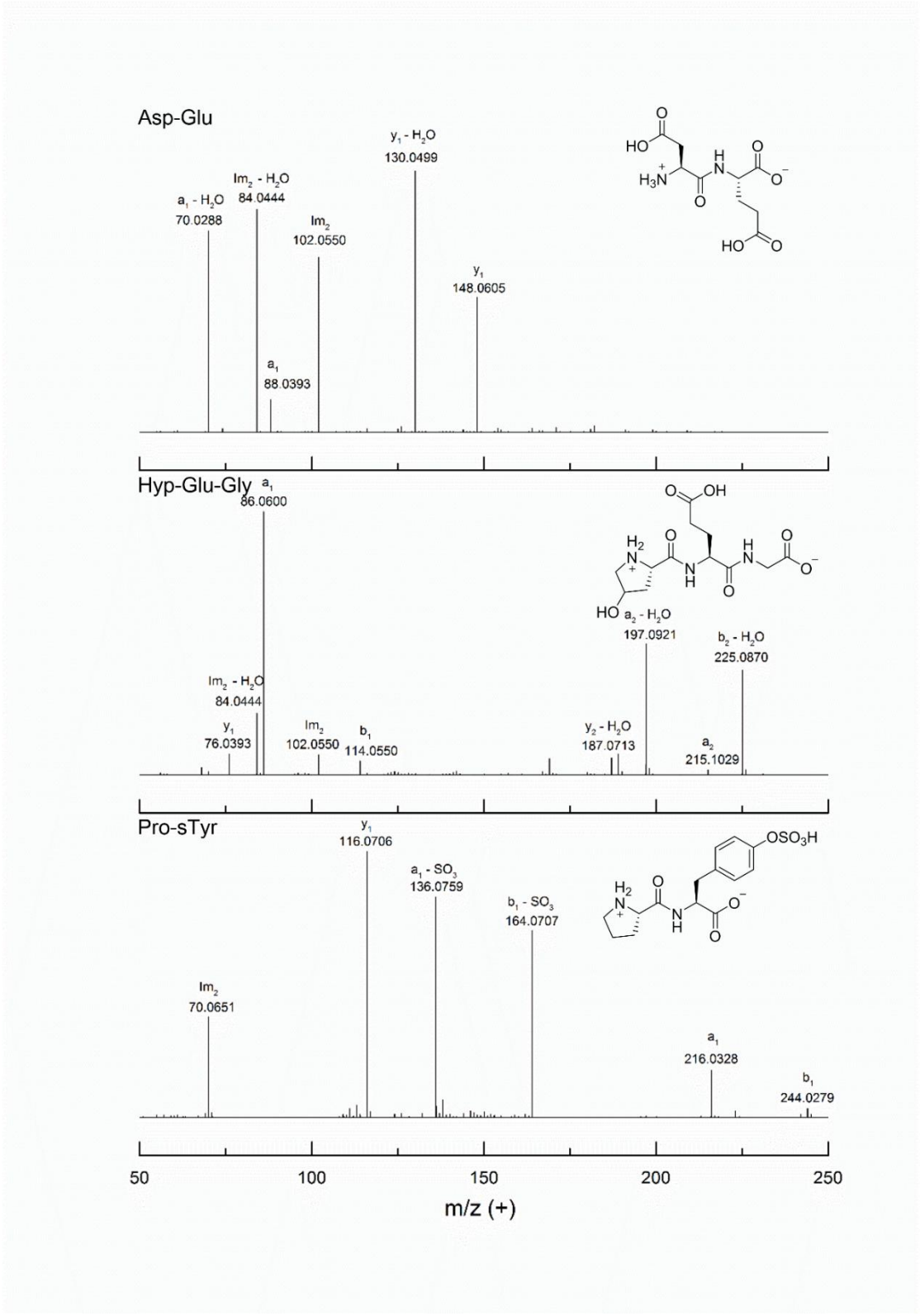

Figure S2. Exemplary fragmentation spectra of natural and modified short peptides identified in urine samples

Table S4. Natural short peptides identified by suspect and untargeted raw files data analysis by the customized data processing workflow on Compound Discoverer and by pNovo.

Table S5. Modified short peptides identified by suspect and untargeted raw files data analysis by the customized data processing workflow on Compound Discoverer and by pNovo.