

Improve the coverage for the analysis of phosphoproteome of Hela cells by a tandem digestion approach

Supporting Figures

Figure S1. Overlap of the identified (a) phosphoproteins and (b) phosphorylation sites between the two technical replicates of Glu-C digestion of the 1D-RP-LC-MS/MS.

Figure S2. Overlap of the identified (a) phosphoproteins and (b) phosphorylation sites between the two technical replicates of the second trypsin digestion of the 1D-RP-LC-MS/MS. Overlap of the identified (c) phosphoproteins and (d) phosphorylation sites between the two technical replicates of the tandem digestion approach of the 1D-RP-LC-MS/MS.

Figure S3. Overlap of the identified (a) phosphoproteins and (b) phosphorylation sites between the trypsin and tandem digestion approach of the 2D-RP-LC-MS/MS.

Figure S4. Overlap of the identified (a) phosphoproteins and (b) phosphorylation sites between the two technical replicates of the trypsin digestion approach of the 2D-RP-LC-MS/MS. Overlap of the identified (c) phosphoproteins and (d) phosphorylation sites between the two technical replicates of the Glu-C digestion approach of the 2D-RP-LC-MS/MS. Overlap of the identified (e) phosphoproteins and (f) phosphorylation sites between the two technical replicates of the second trypsin digestion approach of the 2D-RP-LC-MS/MS. Overlap of the identified (g) phosphoproteins and (h) phosphorylation sites between the two technical replicates of the tandem digestion approach of the 2D-RP-LC-MS/MS.

Figure S5. Comparison of the identified phosphorylation sites ($\text{Ascore} \geq 19$) of (a)

tandem digestion approach and (b) trypsin digestion approach with the data set published by Olsen. et al.

Figure S6. The ratio of novel phosphorylation sites not collected in the three online phosphorylation site databases.

Figure S7a. Serine phosphorylation motifs determined using the Motif-X algorithm from trypsin digestion data set, the number of unique serine phosphorylation sites for each motif and the motif score were also displayed.

Figure S7b. Threonine phosphorylation motifs determined using the Motif-X algorithm from trypsin digestion data set, the number of unique serine phosphorylation sites for each motif and the motif score were also displayed.

Figure S7c. Serine phosphorylation motifs determined using the Motif-X algorithm from Glu-C digestion data set, the number of unique serine phosphorylation sites for each motif and the motif score were also displayed.

Figure S7d. Threonine phosphorylation motifs determined using the Motif-X algorithm from Glu-C digestion data set, the number of unique serine phosphorylation sites for each motif and the motif score were also displayed.

Figure S7e. Serine phosphorylation motifs determined using the Motif-X algorithm from second trypsin digestion data set, the number of unique serine phosphorylation sites for each motif and the motif score were also displayed.

Figure S7f. Threonine phosphorylation motifs determined using the Motif-X algorithm from second trypsin digestion data set, the number of unique serine phosphorylation sites for each motif and the motif score were also displayed.

Supplementary Tables

Table S1. Unique phosphopeptides identified from trypsin digestion, Glu-C digestion, and the second trypsin digestion.

Table S2. The *in silico* distribution of peptide length between 1 and 100 amino acid (aa).

Table S3. The *in silico* distribution of peptide molecular weight between 0 and 10500 Da.

Figure S1. Overlap of the identified (a) phosphoproteins and (b) phosphorylation sites between the two technical replicates of Glu-C digestion of the 1D-RP-LC-MS/MS.

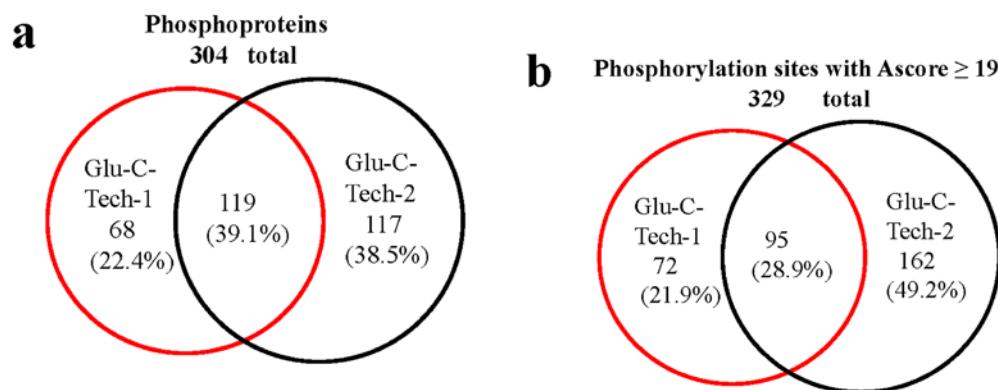


Figure S2. Overlap of the identified (a) phosphoproteins and (b) phosphorylation sites between the two technical replicates of the second trypsin digestion of the 1D-RP-LC-MS/MS. Overlap of the identified (c) phosphoproteins and (d) phosphorylation sites between the two technical replicates of the tandem digestion approach of the 1D-RP-LC-MS/MS.

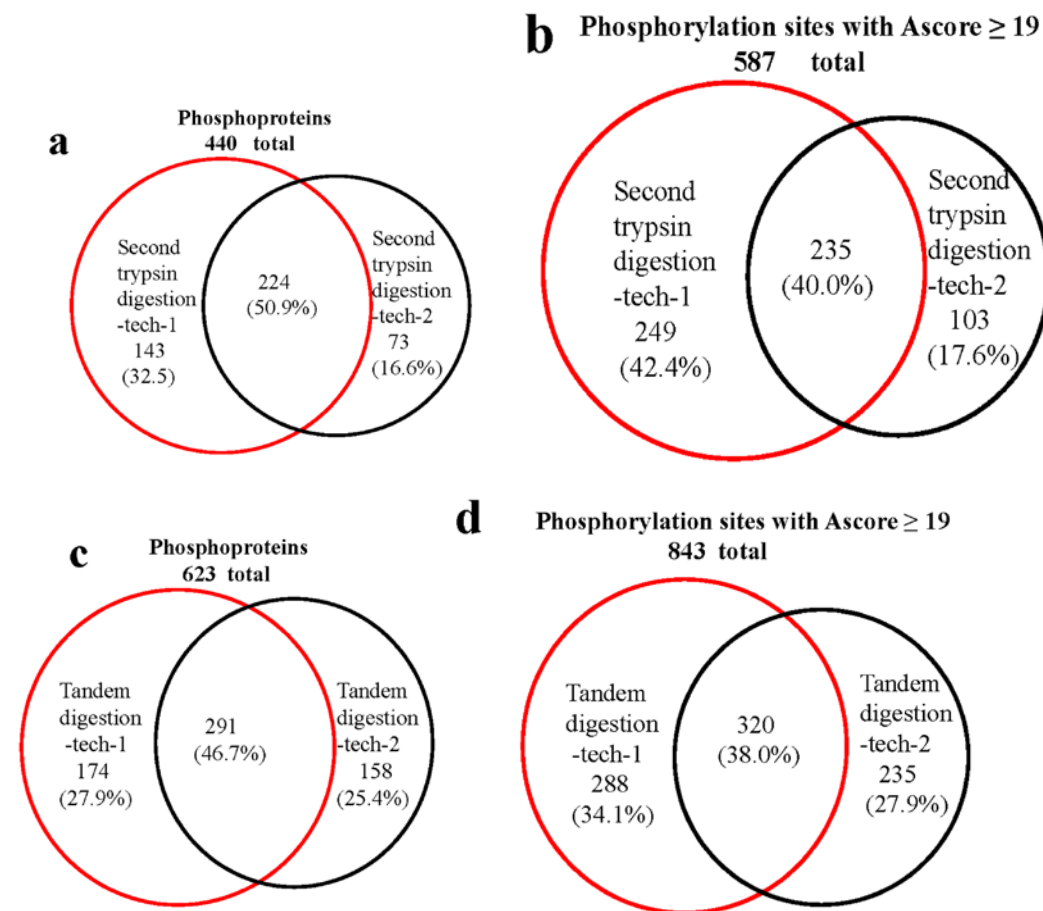


Figure S3. Overlap of the identified (a) phosphoproteins and (b) phosphorylation sites between the trypsin and tandem digestion approach of the 2D-RP-LC-MS/MS.

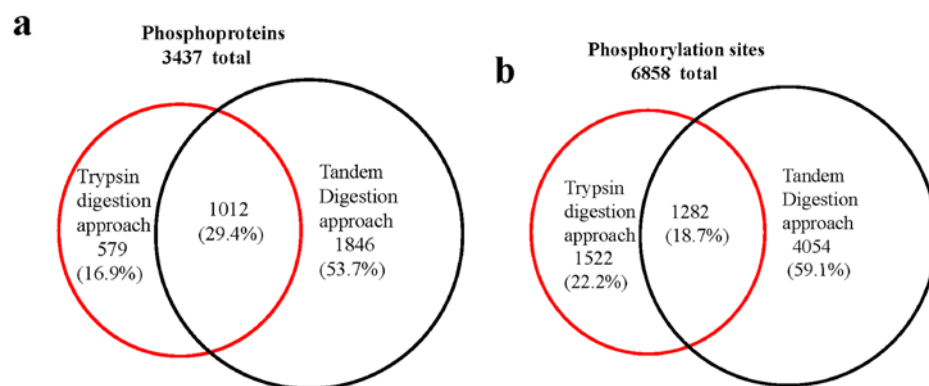
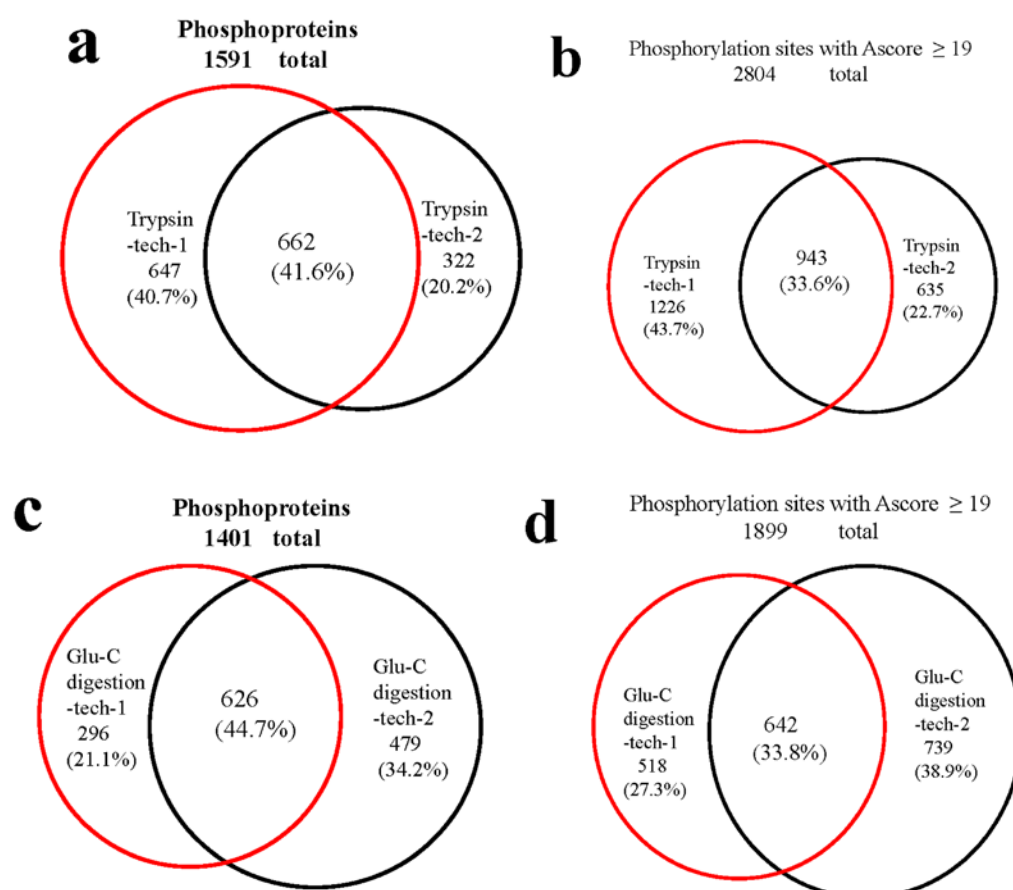


Figure S4. Overlap of the identified (a) phosphoproteins and (b) phosphorylation sites between the two technical replicates of the trypsin digestion approach of the 2D-RP-LC-MS/MS. Overlap of the identified (c) phosphoproteins and (d) phosphorylation sites between the two technical replicates of the Glu-C digestion approach of the 2D-RP-LC-MS/MS. Overlap of the identified (e) phosphoproteins and (f) phosphorylation sites between the two technical replicates of the second trypsin digestion approach of the 2D-RP-LC-MS/MS. Overlap of the identified (g) phosphoproteins and (h) phosphorylation sites between the two technical replicates of the tandem digestion approach of the 2D-RP-LC-MS/MS.



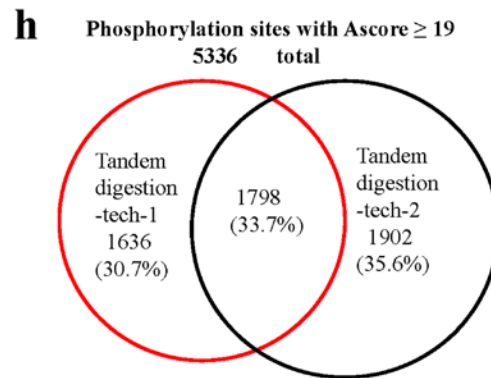
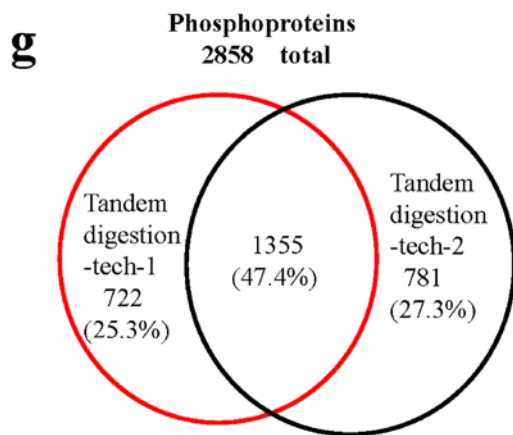
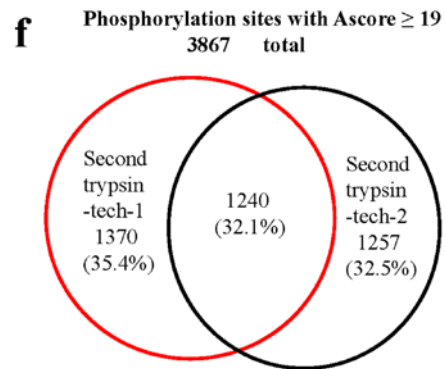
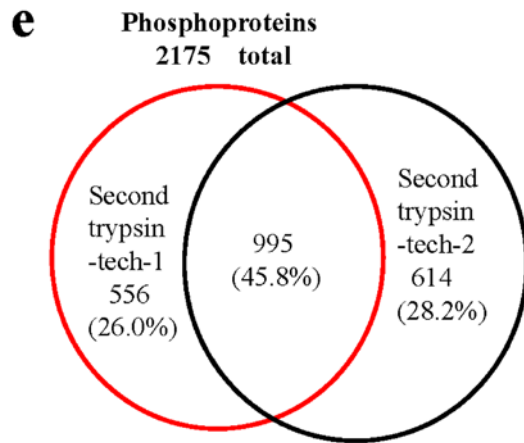


Figure S5. Comparison of the identified phosphorylation sites ($\text{Ascore} \geq 19$) of (a) tandem digestion approach and (b) trypsin digestion approach with the data set published by Olsen et al.

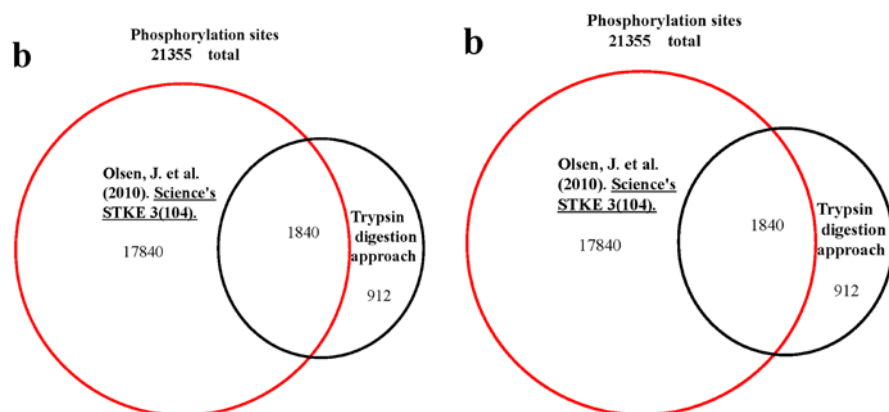


Figure S6. The ratio of novel phosphorylation sites not collected in the three online phosphorylation site databases.

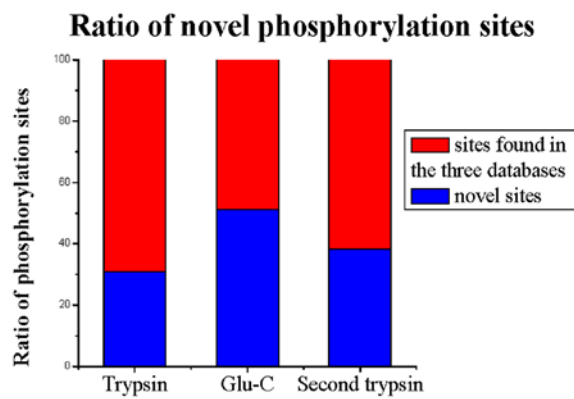
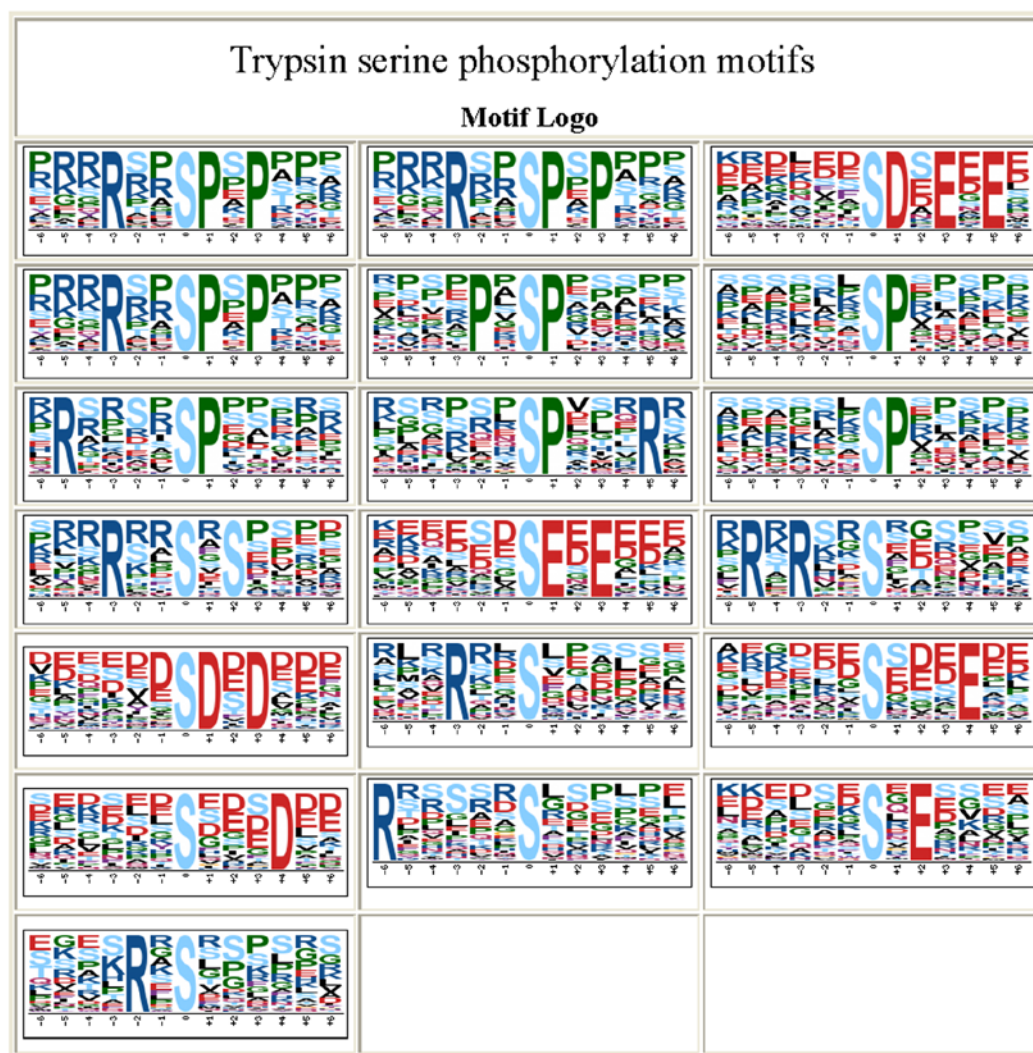


Figure S7a. Serine phosphorylation motifs determined using the Motif-X algorithm from trypsin digestion data set, the number of unique serine phosphorylation sites for each motif and the motif score were also displayed.



#	Motif	Motif Score	Foreground Matches	Foreground Size	Background Matches	Background Size	Fold Increase
1.	<u>...R..SP.P...</u>	43.54	53	2287	741	1094911	34.24
2.	<u>.....SP.K...</u>	32.00	107	2234	3463	1094170	15.13
3.	<u>.....SP.R...</u>	29.61	108	2127	5189	1090707	10.67
4.	<u>....P.SP.....</u>	29.29	145	2019	8868	1085518	8.79
5.	<u>.R....SP.....</u>	29.08	77	1874	3952	1076650	11.19
6.	<u>.....SP...R.</u>	25.22	61	1797	3742	1072698	9.73
7.	<u>.....SD.E.E.</u>	42.91	58	1736	630	1068956	56.69

8.	SP.....	16. 00	368	1678	58380	1068326	4. 01
9.	SD. E...	32. 00	113	1310	4006	1009946	21. 75
10.	...R..S.S....	32. 00	96	1197	7503	1005940	10. 75
11.	SE. E...	32. 00	104	1101	6204	998437	15. 20
12.	.R. R..S.....	27. 10	51	997	4525	992233	11. 22
13.	SD. D...	32. 00	61	946	3131	987708	20. 34
14.	...R..S.....	16. 00	158	885	48140	984577	3. 65
15.	S...E..	16. 00	118	727	58666	936437	2. 59
16.	S...D..	14. 07	77	609	41348	877771	2. 68
17.	R.....S.....	8. 74	66	532	46759	836423	2. 22
18.	S. E....	8. 25	61	466	46362	789664	2. 23
19.	R. S.....	6. 63	48	405	39465	743302	2. 23

Parameters for this run.

fgtype = 'prealigned'

fgcentralres = 'S'

width = '13'

occurrences = '46'

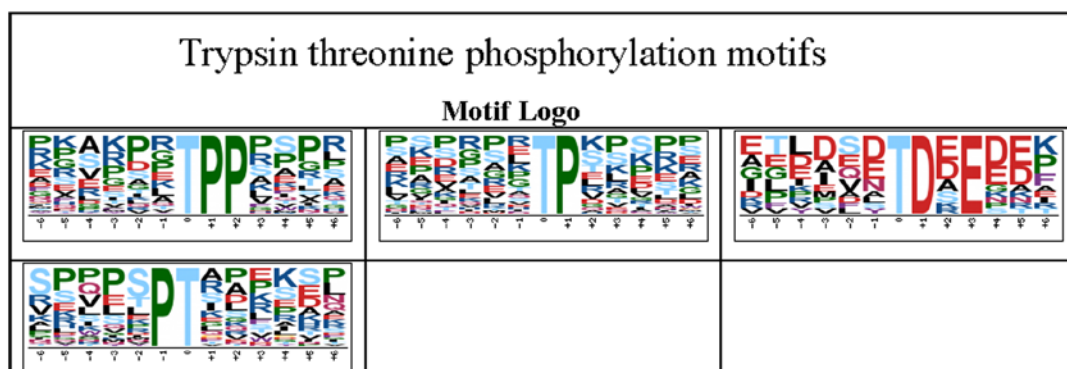
significance = '0.000001'

bgdb = 'ipi.HUMAN.fasta'

bgtype = 'fasta'

bgcentralres = 'S'

Figure S7b. Threonine phosphorylation motifs determined using the Motif-X algorithm from trypsin digestion data set, the number of unique serine phosphorylation sites for each motif and the motif score were also displayed.



#	Motif	Motif Score	Foreground Matches	Foreground Size	Background Matches	Background Size	Fold Increase
1.	TPP....	29.85	46	374	4213	687043	20.06
2.	TP.....	16.00	104	328	44417	682830	4.87
3.	TD.E...	15.03	13	224	2314	638413	16.01
4.	PT.....	6.36	33	211	37982	636099	2.62

Parameters for this run.

fgtype = 'prealigned'

fgcentralres = 'T'

width = '13'

occurrences = '8'

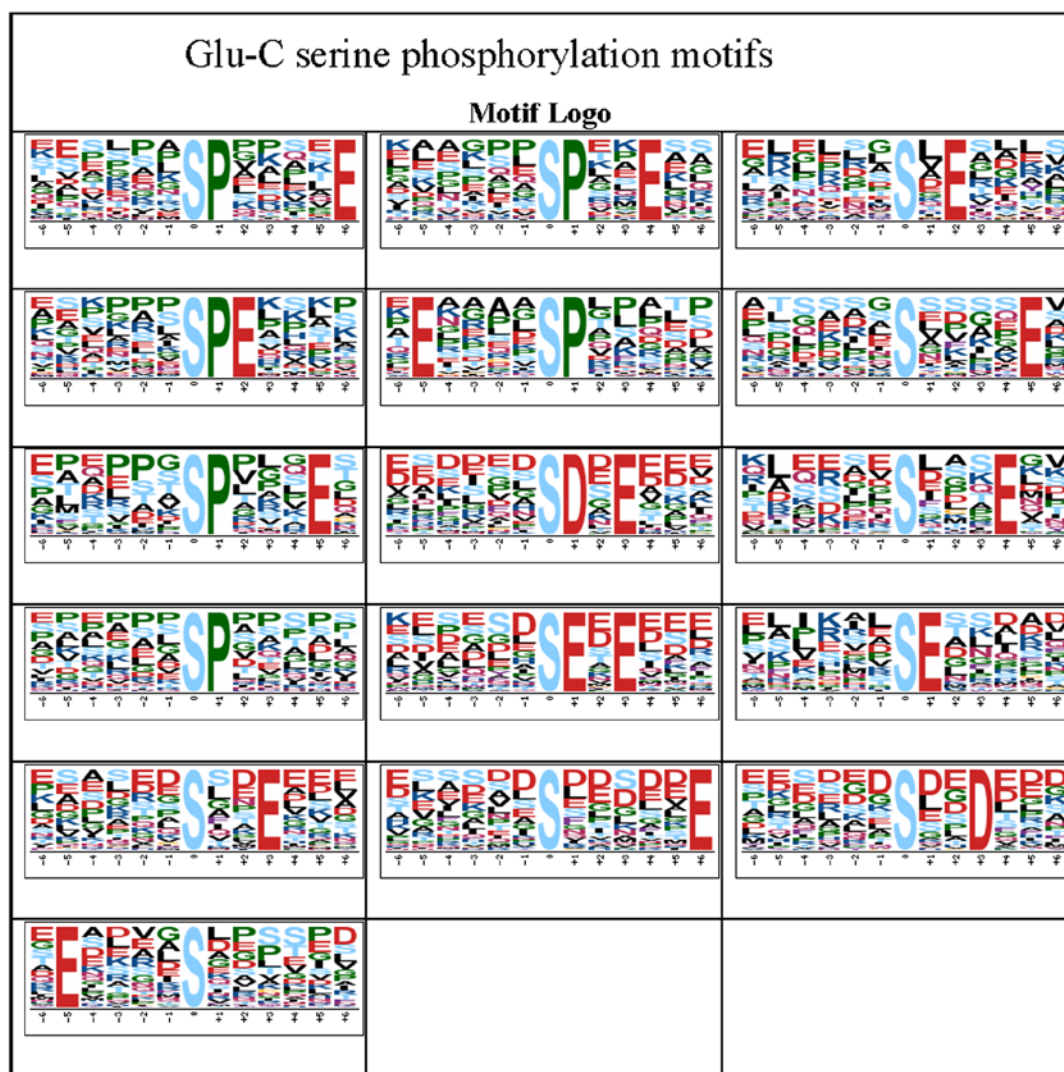
significance = '0.000001'

bgdb = 'ipi.HUMAN.fasta'

bgtype = 'fasta'

bgcentralres = 'T'

Figure S7c. Serine phosphorylation motifs determined using the Motif-X algorithm from Glu-C digestion data set, the number of unique serine phosphorylation sites for each motif and the motif score were also displayed.



#	Motif	Motif Score	Foreground Matches	Foreground Size	Background Matches	Background Size	Fold Increase
1.	SP....E	32.00	99	1472	4607	1094911	15.98
2.	SP..E..	32.00	91	1373	4174	1090304	17.31
3.	SPE....	27.61	67	1282	4843	1086130	11.72
4.	.E....SP.....	25.84	50	1215	3763	1081287	11.82
5.	SP...E.	24.59	41	1165	3309	1077524	11.46
6.	SD.E...	32.00	87	1124	4636	1074215	17.93
7.	SP.....	16.00	252	1037	63639	1069579	4.08

8.	SE.E...	30.95	55	785	6231	1005940	11.31
9.	S..E...	16.00	117	730	57138	999709	2.80
10.	S.....E	16.00	100	613	57400	942571	2.68
11.	S..D...	16.00	78	513	43006	885171	3.13
12.	.E....S.....	12.07	66	435	48312	842165	2.64
13.	S.E....	11.36	56	369	42912	793853	2.81
14.	S....E	10.60	50	313	41838	750941	2.87
15.	S...E..	6.69	37	263	39147	709103	2.55
16.	SE.....	6.53	33	226	36591	669956	2.67

Parameters for this run.

fgtype = 'prealigned'

fgcentralres = 'S'

width = '13'

occurrences = '30'

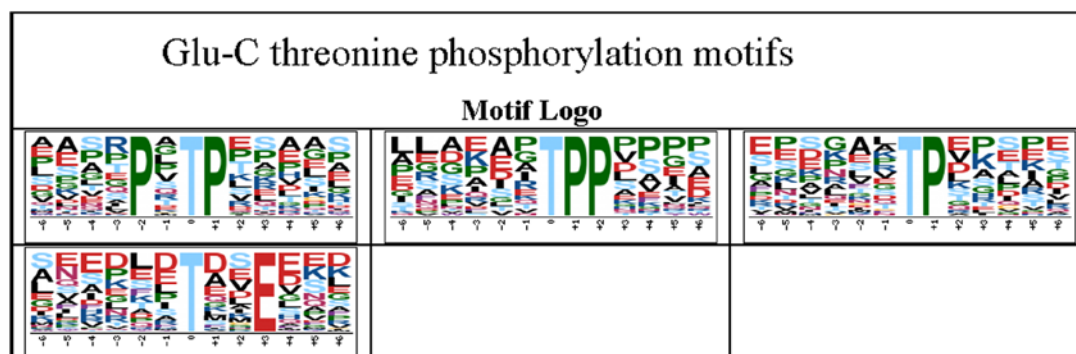
significance = '0.000001'

bgdb = 'ipi.HUMAN.fasta'

bgtype = 'fasta'

bgcentralres = 'S'

Figure S7d. Threonine phosphorylation motifs determined using the Motif-X algorithm from Glu-C digestion data set, the number of unique serine phosphorylation sites for each motif and the motif score were also displayed.



#	Motif	Motif Score	Foreground Matches	Foreground Size	Background Matches	Background Size	Fold Increase
1.	P. TP.....	26.07	47	328	5654	687043	17.41
2.	TPP.....	24.58	29	281	3574	681389	19.68
3.	TP.....	16.00	73	252	39402	677815	4.98
4.	T..E...	9.40	37	179	41705	638413	3.16

Parameters for this run.

fgtype = 'prealigned'

fgcentralres = 'T'

width = '13'

occurrences = '7'

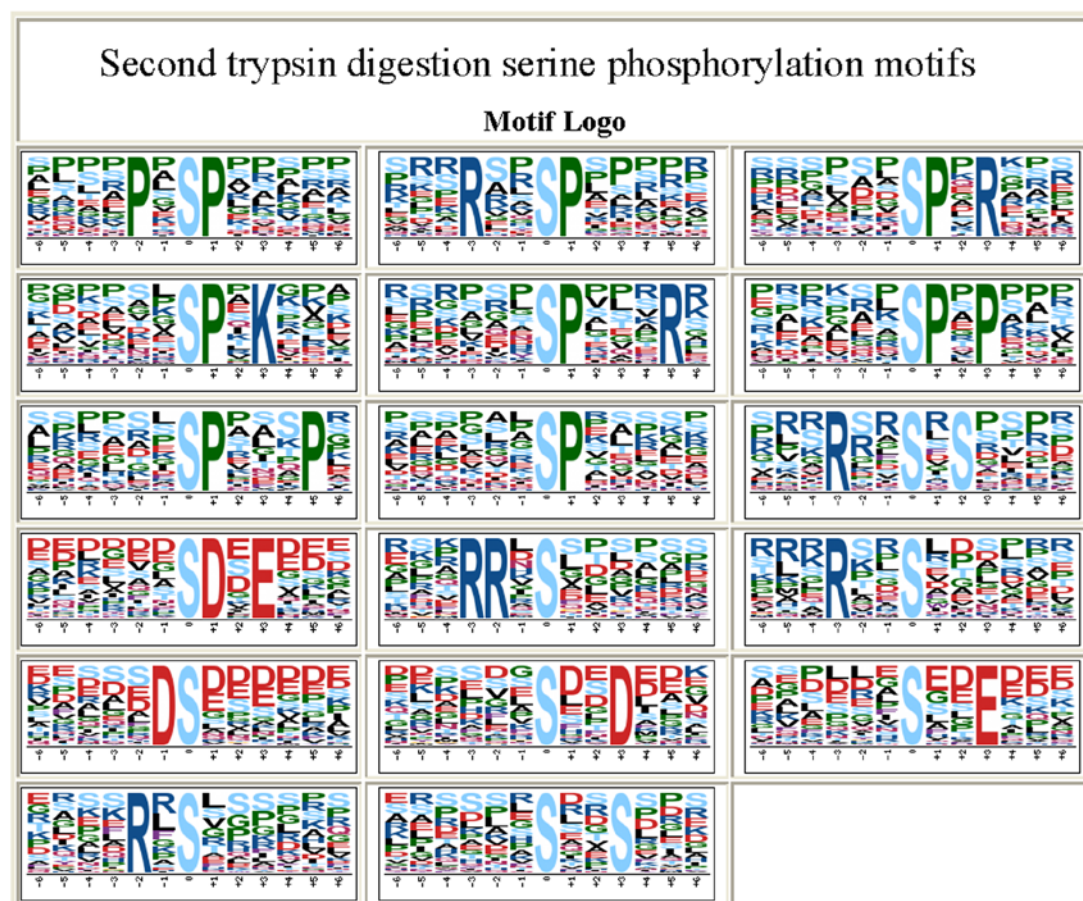
significance = '0.000001'

bgdb = 'ipi.HUMAN.fasta'

bgtype = 'fasta'

bgcentralres = 'T'

Figure 7e. Serine phosphorylation motifs determined using the Motif-X algorithm from second trypsin digestion data set, the number of unique serine phosphorylation sites for each motif and the motif score were also displayed.



#	Motif	Motif Score	Foreground Matches	Foreground Size	Background Matches	Background Size	Fold Increase
1.	P.SP....	32.00	347	3093	10017	1094911	12.26
2.	...R..SP.....	32.00	162	2746	4813	1084894	13.30
3.	SP.R...	32.00	142	2584	4047	1080081	14.67
4.	SP.K...	32.00	117	2442	2908	1076034	17.73
5.	SP...R.	29.39	106	2325	3764	1073126	13.00
6.	SP.P...	25.21	143	2219	7272	1069362	9.48
7.	SP...P.	23.02	97	2076	5274	1062090	9.41
8.	SP.....	16.00	453	1979	46240	1056816	5.23
9.	...R..S.S....	32.00	129	1526	7535	1010576	11.34

10.	SD.E...	32. 00	139	1397	4604	1003041	21. 68
11.	...RR.S.....	32. 00	71	1258	4361	998437	12. 92
12.	...R..S.....	16. 00	196	1187	48761	994076	3. 37
13.	DS.....	16. 00	164	991	44894	945315	3. 48
14.	S..D...	16. 00	100	827	41133	900421	2. 65
15.	S..E...	16. 00	112	727	56402	859288	2. 35
16.	R.S.....	12. 05	80	615	43629	802886	2. 39
17.	S..S...	8. 52	113	535	92035	759257	1. 74

Parameters for this run.

fgtype = 'prealigned'

fgcentralres = 'S'

width = '13'

occurrences = '63'

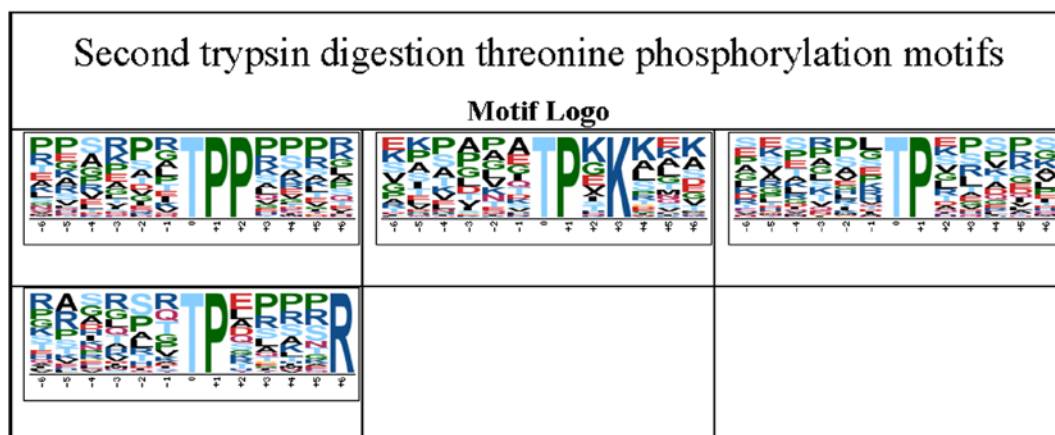
significance = '0.000001'

bgdb = 'ipi.HUMAN.fasta'

bgtype = 'fasta'

bgcentralres = 'S'

Figure S7f. Threonine phosphorylation motifs determined using the Motif-X algorithm from second trypsin digestion data set, the number of unique serine phosphorylation sites for each motif and the motif score were also displayed.



#	Motif	Motif Score	Foreground Matches	Foreground Size	Background Matches	Background Size	Fold Increase
1.	TPP.....	32.00	101	603	4213	687043	27.31
2.	TP.K...	27.53	39	502	2003	682830	26.48
3.	TP....R	23.01	32	463	2419	680827	19.45
4.	TP.....	16.00	165	431	39995	678408	6.49

Parameters for this run.

fgtype = 'prealigned'

fgcentralres = 'T'

width = '13'

occurrences = '12'

significance = '0.000001'

bgdb = 'ipi.HUMAN.fasta'

bgtype = 'fasta'

bgcentralres = 'T'

Table S1. Unique phosphopeptides identified from trypsin digestion, Glu-C digestion, and the second trypsin digestion.

		Unique phospho- peptides	Unique Phospho- proteins	Unique Phospho- rylation sites	Unique phosphorylation sites with Ascore ≥ 19
Trypsin digestion	Replicates-1	713	390	858	593
	Replicates-1	680	389	795	555
	Two replicates	965	503	1147	790
Glu-C digestion	Replicates-1	293	187	326	167
	Replicates-1	390	236	446	257
	Two replicates	516	304	595	329
Second trypsin digestion	Replicates-1	622	367	701	484
	Replicates-1	446	297	506	338
	Two replicates	762	440	867	587
Tandem digestion approach	Two replicates	1243	623	1336	843

Table S2. The *in silico* distribution of peptide length between 1 and 100 amino acid (aa).

Peptide length	Trypsin	Glu-C	Second trypsin
≤10 aa	37.89%	21.44%	39.13%
≤20 aa	66.00%	41.66%	70.72%
≤30 aa	81.95%	57.76%	87.09%
≤40 aa	90.51%	69.59%	94.40%
≤50 aa	94.90%	78.32%	97.46%
≤60 aa	97.17%	84.69%	98.75%
≤70 aa	98.41%	89.12%	99.36%
≤80 aa	99.07%	92.28%	99.64%
≤90 aa	99.43%	94.53%	99.78%
≤100 aa	99.63%	96.10%	99.86%

Table S3. The *in silico* distribution of peptide molecular weight between 0 and 10500 Da.

Molecular weight	Trypsin	Glu-C	Second trypsin
≤1500 Da	46.03%	26.75%	48.40%
≤3000 Da	76.75%	51.42%	82.33%
≤4500 Da	90.45%	68.78%	94.49%
≤6000 Da	95.94%	80.19%	98.15%
≤7500 Da	98.21%	87.63%	99.29%
≤9000 Da	99.17%	92.23%	99.69%
≤10500 Da	99.58%	95.13%	99.85%