Comprehensive characterization of porcine and bovine trypsin digestion

Scott Walmsley*[1], Paul Rudnick[2], Yuxue Liang[2], Qian Dong[2], Stephen E. Stein[2], Alexey Nesvizhskii[1]

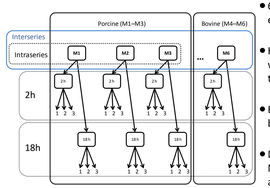
[1]University of Michigan Department of Pathology, Ann Arbor, Michigan

[2]National Institute of Standards and Technology, Gaithersburg, Maryland

Introduction

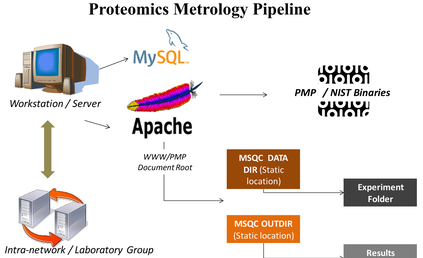
- Trypsin is an endoprotease commonly used for sample preparation in proteomics experiments.
- Protein digestion is dependent on multiple factors including trypsin origin and digestion conditions.
- We assembled a data analysis pipeline and visualization tools for quality control and characterization of variability in proteomic experiments.
- We performed a comprehensive evaluation of six commercially available proteomics-grade trypsins and their digestions of a single purified protein, human serum albumin (HSA).

Methods



- 6 trypsins from different manufacturers (M1-M3: porcine, M4-M6: bovine).
- Human serum albumin was solubilized with urea and digested according to the schema at left for 2 or 18 hours.
- Each digest was analyzed in triplicate by LC-MS/MS.
- Data were processed using the NIST – MSQC software and a custom data analysis software pipeline.

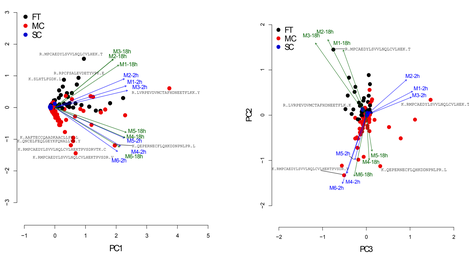
Proteomics Metrology Pipeline



- A custom data analysis pipeline was developed to interface with the NIST-MSQC binaries.
- All data were processed using NIST-MSPepSearch (v.0.9) together with a comprehensive HSA library (2782 unique ions) to identify MS/MS spectra.
- Abundances were calculated using NIST-ProMS.
- A custom PHP library was developed to perform statistical evaluation of the peptide ions and roll-up to peptide level analysis.

Results

I. PCA analysis of peptide abundances

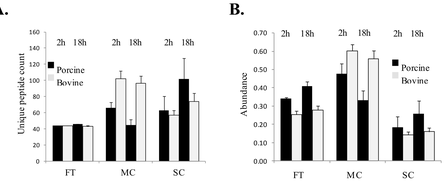


Principal components analysis was used to identify which HSA peptides contributed significantly toward the differences amongst the trypsins. Peptide abundances are a sum of ProMS abundances for each ion detected for each peptide sequence. FT: fully tryptic, MC: missed cleavage, SC: semi-tryptic cleavage, M1-M3: porcine trypsins, M4-M6: bovine trypsins, 2h and 18h: length of digest.

- PC1 summarized the differences between the intensities of different HSA peptides (caused by differences in ionization efficiency and other factors).
- PC2 aided visualization of the differences between the categories of trypsin (porcine versus bovine).
- PC3 aided detection of the differences between the 2 and 18 hour digests, especially for the porcine trypsins.

FT and MC peptides showed most differences between the bovine and porcine trypsin digests. Select sequences determined to be significantly altered between the bovine and porcine trypsins are shown in **Figure I**. Some of these sequences were nested sequence sets.

II. Peptide classification by cleavage rule



a) Peptide counts. The number of FT peptides remained constant between trypsins. There were more MC peptides produced by the bovine trypsins and more SC peptides produced by the porcine trypsins.

b) Peptide abundances. The total abundances of FT and SC peptides produced by porcine trypsins were higher compared to bovine trypsins. MC peptide abundances were favored in bovine trypsin digests.

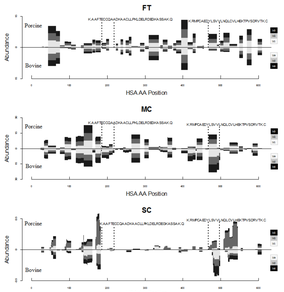
- FT abundances increased from 2 to 18 hrs. Both counts and abundances for the MC peptides decreased from 2 to 18 hrs. This indicated more complete digests at 18 hours.
- SC counts and abundances increased from 2 to 18 hrs. Some of these SC peptides were stably produced, however most were of low abundance.

III. Peptides with significant differences in porcine vs. bovine digests (18 hr)

	Flanking Residues	Category	Nested series	Position	M1	M2	M3	M4	M5	M6	PorcinetCV	BovinetCV	FC (Bov/Por)	p value
	DAHKSEVAHR	K.F	MC	-	24	1.20	0.46	0.85	3.13	2.89	3.19	00.84±0.44	03.07±0.05	3.65	0.004
	ALVLIAFAQYLQQCPFEDHVK	K.L	FT	-	44	40.83	37.31	38.87	31.06	31.46	32.56	39.00±0.05	31.69±0.02	0.81	0.009
	SLHTLFGDK	K.L	FT	-	88	20.39	25.28	22.26	1.31	2.11	0.25	22.65±0.11	01.23±0.76	0.05	0.002
	LCTVATLRETYGEMADCCAK	K.L	FT	A	97	15.36	20.28	17.91	1.68	3.28	0.87	17.85±0.14	01.94±0.63	0.11	0.002
	LCTVATLRETYGEMADCCAK	K.Q	MC	A	97	0.06			0.37	0.33	0.27	0.06	00.32±0.15	4.99	
	QEPERNECFLQHKDDNPNLPR	K.D	MC	B	117	2.51	4.31	2.82	0.25	0.66	0.35	03.21±0.30	00.42±0.50	0.13	0.033
	QEPERNECFLQHKDDNPNLPR	K.L	MC	B	117	27.59	10.76	18.63	41.23	41.51	41.48	19.00±0.44	41.41±0.00	2.18	0.044
	NECFLQHKDDNPNLPR	R.L	MC	B	122	4.84	3.45	4.78	0.58	0.63	0.42	04.36±0.18	00.54±0.20	0.12	0.013
	LVRPEVDVMCTAFHDNEETFLKK	K.Y	MC	-	138	75.36	76.96	75.79	59.32	56.58	59.16	76.03±0.01	58.35±0.03	0.77	0.000
	AAFTECCQAADK	K.A	FT	C	186	8.79	8.77	8.59	2.23	3.04	2.02	08.72±0.01	02.43±0.22	0.28	0.002
	AAFTECCQAADKCAALEPK	K.L	MC	C	186	0.28	0.18	0.15	18.94	21.11	17.51	00.20±0.33	19.19±0.09	93.77	0.003
	AACLEPK	K.L	FT	C	198	14.63	15.67	15.04	5.66	6.21	4.00	15.12±0.03	05.29±0.22	0.35	0.001
	AWAVAR	K.L	FT	-	236	10.77	10.49	10.93	8.34	9.07	8.72	10.73±0.02	08.71±0.04	0.81	0.003
	LSQHFPK	R.A	MC	-	242	0.76		0.20	6.90	5.46	7.49	00.48±0.82	06.62±0.16	13.71	0.004
	DFCKNTAEAKDVFLGMFLYEYARR	K.H	MC	D	337	0.11	0.07	0.12	13.11	19.29	17.65	00.10±0.23	16.68±0.19	165.12	0.012
	NYAEAK	K.D	FT	D	341	1.62	1.99	1.86	0.78	1.07	0.78	01.82±0.10	00.88±0.19	0.48	0.003
	NYAEAKDVFLGMFLYEYARRPDTSVTVLLR	K.L	MC	D	341	1.82	0.87	2.09	3.64	2.86	4.03	01.59±0.40	03.51±0.17	2.21	0.019
	DVFLGMFLYEYARRPDTSVTVLLR	K.L	MC	D	347	1.19	0.65	0.89	2.54	1.80	2.94	00.91±0.30	02.43±0.24	2.67	0.029
	RHPDISVTLLLR	R.L	MC	D	360	28.15	28.26	27.62	18.86	21.44	20.43	28.01±0.01	20.24±0.06	0.72	0.006
	CCAAADPHECYAK	K.V	FT	-	383	10.57	11.70	10.37	8.04	8.67	7.48	10.88±0.07	08.06±0.07	0.74	0.007
	QNCELFEQLGEYKFQNALLVR	K.F	FT	E	413	10.75	12.36	12.18	2.67	4.35	1.44	11.77±0.08	02.82±0.52	0.24	0.002
	QNCELFEQLGEYKFQNALLVR	K.Y	MC	E	413	0.21	0.18	0.09	19.96	15.45	23.17	00.16±0.39	19.52±0.20	123.29	0.013
	FQNALLVR	K.Y	FT	E	426	23.30	24.89	24.05	7.94	11.14	4.94	24.08±0.03	08.01±0.39	0.33	0.009
	KTVQFSTPTLFEYSR	K.N	MC	-	437	31.97	33.60	33.00	29.32	30.46	29.66	32.86±0.03	29.81±0.02	0.91	0.009
	RMPCAEDISVTLNQLCVLHEKTPVSDR	K.V	MC	F	468	0.19			18.90	31.06	34.58	0.19	28.18±0.29	151.51	
	MPCAEDISVTLNQLCVLHEK	R.T	FT	F	469	44.47	41.92	52.23	9.89	7.98	6.25	46.21±0.12	08.04±0.23	0.17	0.003
	MPCAEDISVTLNQLCVLHEKTPVSDR	R.V	MC	F	469	0.44		0.12	6.10	7.23	11.03	00.28±0.80	08.12±0.32	28.91	0.033
	TPVSDRVK	K.C	MC	F	490	0.08			4.62	4.08	3.55	0.08	04.08±0.13	54.09	
	QTALVELVK	K.H	FT	-	549	19.01	20.69	20.14	16.44	14.23	14.66	19.95±0.04	15.11±0.08	0.76	0.006
	ADDKETCAEEGKK	K.L	MC	-	584	15.09	16.69	14.20	10.24	8.82	7.25	15.33±0.08	08.77±0.17	0.57	0.005

- Several peptides were produced in significantly altered quantities in the porcine vs. bovine trypsin digests.
- There were sets of nested sequences identified (see **Figure V** for examples).
- For these nested sets, MC subsequences were of greater abundance when using bovine trypsins, whereas nested FT peptides were of greater abundance when using porcine trypsins.

IV. Quantitative sequence maps of HSA



- Quantitative sequence maps indicated that abundances of peptides were generally reproducible between different trypsins.
- However, there were regions of significantly altered abundances between the bovine and porcine trypsin digests. Examples are shown in the figure.
- These regions of altered abundance contained nested set of FT, MC, and SC peptides. (see **Figure V**).

Conclusions

- We performed a statistical analysis of peptide abundances from different trypsin digests and visualized data using PCA and quantitative protein "sequence maps".
- The performance of individual trypsins across digests was reproducible, however significant differences were observed depending on the origin of the trypsin.
- Bovine trypsins produced a higher number of peptides containing missed cleavages, whereas porcine trypsins produced more semi-tryptic peptides.
- Overall, this work illustrates effects of an often neglected source of variability in proteomics experiments: the origin of trypsin

This work was funded in part by the project "Understanding measurement variability in proteomics" (12012NIST-MSE-01).

V. Human Serum Albumin peptide heatmap.

Total abundances detected for each identified sequence are shown. Cleavage rule for each sequence is color coded at left. Areas of interest (nested sequence sets) identified first by PCA and shown in table III are labeled along the left side of the map.