Counting Strategies for Differential Characterization Of Post-Translational Modifications

McComb ME[1], Spencer JL[1], Bhatia VN[1], Whelan SA[1], Kehasse A[1], Perlman DH[2], Heckendorf CF[1], and Costello CE[1]

[1]Boston University School of Medicine, Boston, MA 02118

[2]Princeton University, Princeton, NJ 08544

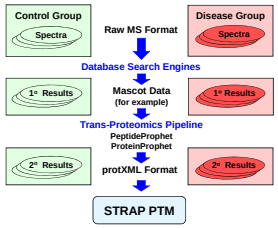
Introduction

There is increasing interest in the identification and characterization of protein post-translational modifications (PTMs) as the field of mass spectrometry-based proteomics begins to mature. However, the vast amount of information obtained within a typical differential proteomics study makes the facile measurement of PTMs quite challenging. In addition to label-free approaches, we have begun to explore counting methods for differential analysis of PTM changes in proteomes. Our novel approach has resulted in the development of a software program, the **Software Tool for Rapid Annotation of Proteins: Post-Translation Modification edition (STRAP PTM)**. STRAP PTM uses a new counting-based PTM scoring algorithm to facilitate multi-sample PTM comparison through collation and visualization. Here we demonstrate the utility of STRAP PTM across different PTM/proteomics experiments.

Methods

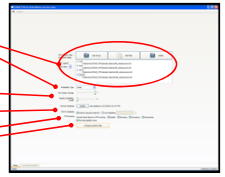
STRAP PTM: PTM Counting

Workflow

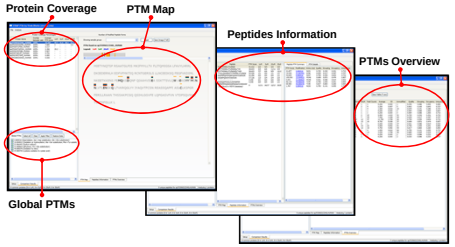


Setup Window

1. Enter protXML files per group.
2. Select ProteinProphet probability type.
3. Choose minimum protein overlap and peptide probability cutoff.
4. Indicate FASTA database.
5. Check factors for PTM scoring.
6. Execute comparison.



Comparison of Results Windows



Methods

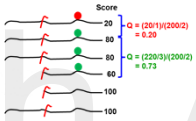
STRAP PTM: PTM Scoring

Algorithm	Expression
PTM Score	$S_{mip} = 100\frac{Q_{mip}\gamma_{mip}W_{mip}U_{mp}}{(\max S)_p}$

- Based on four factors (Q = quality, γ = grouping, W = occupancy, and U = uniqueness) and normalized by the highest score in the data set (max PTM score = 100).
- Calculated for a specific PTM (m = 1, …, M) on a specific site (i = 1, …, I) of a specific protein (p) across all groups or data sets.

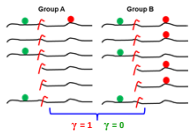
Factor Name	Expression
Quality	$Q_{mip} = \frac{\langle B_{mip}\rangle}{\langle B^0_{ip}\rangle}$

- Defined as the ratio of (the average score of peptides having a specific PTM on a specific site) to (the average score of all unmodified peptides having the specific site).
- Determines the quality of the database search results based on MS/MS ID scores (B).
- Higher score indicates positive ID (max = 1).



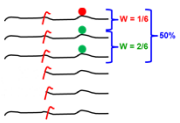
Factor Name	Expression
Grouping	$\gamma_{mip} = \frac{\sigma_{mip}}{(\max \sigma)_p}$

- Defined as the standard deviation (σ) of a specific PTM on a specific site across groups and normalized by the highest σ for the specific protein.
- Higher score indicates group specificity for the specific PTM and site (max = 1).



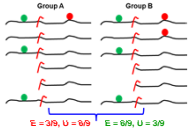
Factor Name	Expression
Occupancy	$W_{mip} = \frac{N_{mip}}{(\sum_{m=1}^{M} N_{mip}) + (N^0{}_{ip})}$

- Defined as the degree of modification of a specific site with a specific PTM on a specific protein.
- W is the ratio of (a specific PTM on a specific site) to (the sum of all PTMs and vacancies on the site).
- Higher score indicates more of the specific PTM at the site (max = 1).
- Total site occupancy equals the sum of specific PTM occupancies (susceptibility to be modified).



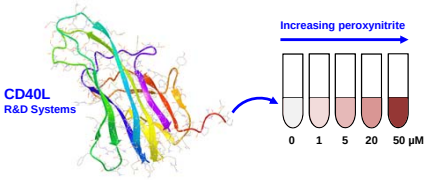
Factor Name	Expression
Uniqueness	$U_{mp} = 1 - E_{mp} = 1 - \frac{\sum_{i=1}^{I} N_{mip}}{\sum_{m=1}^{M}\sum_{i=1}^{I} N_{mip}}$

- Defined as the enhancement (E) of a specific PTM on a specific protein subtracted from 1.
- E is the ratio of (a specific PTM on all sites) to (all PTMs on all sites).
- Higher score indicates more specificity in the PTM (max = 1).



Results

Model System: In Vitro Oxidation of CD40L



PTM Map

Legend: 1 μM 5 μM 20 μM 50 μm

MIETYNQTSP RSAATGLPIS MKIFMYLLTV FLITQMIGSA LFAVYLHRRL

DKIEDERNLH EDFVFMKTIQ RCNTGERSLS LLNCEEIKSQ FEGFVKDIML

NKEETKKENS FEMQKGDQNP QIAAHVISEA SSKTTSVLQW AEKGYYTMSN

NLVTLENGKQ LTVKRQGLYY IYAQVTFCSN REASSQAPFI ASLCLKSPGR

FERILLRAAN THSSAKPCGQ QSIHLGGVFE LQPGASVFVN VTDPSQVSHG

TGFTSFGLLK L

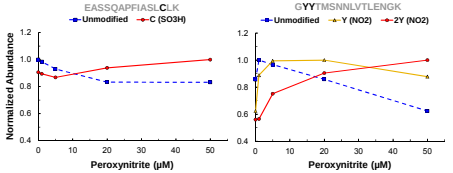
(Y145, Y146) NO2

(C194) SO3H

PTM Scoring

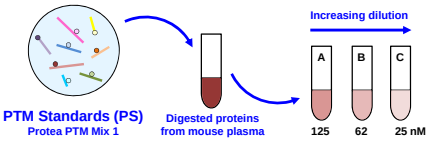
Position	AA	Index	Mod Mass	1uM	5uM	20uM	50uM	Total Counts	Average Counts	SD	Unmodified Forms	Quality	Grouping	Occupancy	Uniqueness	PTM Score
148	M		15.99492	3	7	10	15	35	8.750	5.058	37	0.698	0.872	0.479	0.807	100.000
145	Y		44.98508	0	1	11	10	22	5.500	5.802	51	0.755	1.000	0.301	0.850	82.110
146	Y		44.98508	0	1	8	10	19	4.750	4.922	54	0.617	0.860	0.260	0.858	65.083
157	N		0.984016	11	10	14	12	47	11.750	1.708	26	0.759	0.294	0.644	0.715	43.629
113	M		15.99492	0	0	0	1	1	0.250	0.500	0	0.984	0.086	1.000	0.857	29.044
117	D		21.98194	1	3	3	2	9	2.250	0.957	5	0.802	0.165	0.843	0.727	26.286
114	Q		0.984016	0	0	0	1	1	0.250	0.500	0	0.984	0.086	1.000	0.715	25.728
151	N		0.984016	7	7	5	3	22	5.500	1.915	51	0.725	0.330	0.301	0.715	21.877
160	Q		0.984016	4	1	5	3	13	3.250	1.708	45	0.885	0.294	0.224	0.715	17.726
156	E		21.98194	7	6	9	7	29	7.250	1.258	44	0.539	0.217	0.397	0.727	14.325
129	E		21.98194	0	2	1	1	4	1.000	0.816	10	0.644	0.141	0.286	0.727	7.986
194	C		47.98475	0	0	0	1	1	0.250	0.500	1	0.245	0.086	0.500	0.858	4.349
150	N		0.984016	0	2	1	1	4	1.000	0.816	69	0.755	0.141	0.055	0.715	1.767
140	W		15.99492	0	0	0	1	1	0.250	0.500	2	0.030	0.086	0.333	0.857	0.293
142	E		21.98194	0	0	0	1	1	0.250	0.500	2	0.030	0.086	0.333	0.727	0.264
148	M		31.98983	0	0	1	0	1	0.250	0.500	37	0.001	0.086	0.014	0.932	0.001

Label-Free Comparison: Progenesis LC-MS



Results

Model System: PTM Peptides in Plasma



PTM Map

Legend: A (125 nM) B (62 nM) C (25 nM)

PS_280 SYSMEHFRWG (M4) O

PS_321 DRVYIHPFHL (Y4) NO2

PS_500 EAISPPDAAS AAPLR (S4, S10) GalNAc

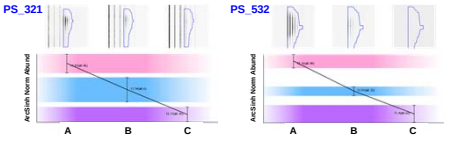
PS_532 DFNKFHTFPQ TAIGV (K4) Ac

PTM Scoring

Peptide	Position	AA Index	Mod Mass	A	B	C	Total Counts	Average Counts	SD	Unmodified Forms	Quality	Grouping	Occupancy	Uniqueness	PTM Score
PS_321	4	Y	44.98508	8	5	3	16	5.333	2.517	0	0.899	1.000	1.000	0.885	100.000
PS_532	4	K	42.01057	4	1	0	5	1.667	2.082	0	0.899	0.827	1.000	0.617	57.859
PS_500*	10	S	203.0794	4	3	0	7	2.333	2.082	12	0.899	0.827	0.368	0.821	28.252
PS_280	4	M	15.99492	8	8	7	23	7.667	0.577	0	0.899	0.229	1.000	0.839	21.706

* Peptide PS_500 was also observed as a native mouse plasma peptide and thus was scored with unmodified forms.

Label-Free Comparison: Progenesis LC-MS



Progenesis LC-MS Values

Peptide	Confidence Score	Anova (p) x 10²	Max Fold Change
PS_321	22.9	58.7	12.9
PS_532	17.7	1.98	41.7
PS_500	37.5	5.84	11.2
PS_280	80.3	161	11.0

Conclusions

- PTM counting is a powerful technique for rapid semi-quantitative integration of large mass spectrometry-based proteomics data sets.
- STRAP PTM represents a novel counting approach that uses a new scoring algorithm to rank PTMs in differential proteomics experiments.
- STRAP PTM software is easy to implement on a PC and provides fast turnaround for even large data sets (on the order of minutes).
- Select STRAP PTM results correlate well with label-free results from both simple and complex model data sets.

Acknowledgments: Funding was provided by NIH-NCRR grants P41 RR010888/GM104603, S10 RR015942, S10 RR020946, S10 RR025082 and NIH-NHLBI contract HHSN268201000031C.