Characterization of Post-Translational Modifications Using Counting Approaches

Jean L. Spencer, Vivek N. Bhatia, Amanuel Kehasse, Stephen A. Whelan, Christian F. Heckendorf, Catherine E. Costello, and Mark E. McComb

Boston University School of Medicine, Boston, MA 02118

Introduction

There is an increasing interest in the identification and characterization of protein post-translational modifications (PTMs). The vast amount of information obtained within a typical differential proteomics study makes the measure of PTMs challenging. We have begun to explore counting methods for differential analysis of PTM changes in proteomes and have developed a software program to aid in this approach. The **Software Tool for Rapid Annotation of Proteins: Post-Translation Modification** edition (STRAP PTM) uses a new counting-based PTM scoring algorithm to facilitate multi-sample PTM comparison through collation and visualization. The application imports parsed prot.XML file data and ranks peptides based on differential PTMs. Here we show the utility of STRAP PTM across data sets with increasing complexity and compare the results with label-free analysis using Progenesis LC-MS (NonLinear Dynamics).

Results Results System 1: Oxidation of CD40L System 2: PTM Peptides in Plasma B CD40L **R&D Systems** PTM Standards (PS)



Methods

STRAP PTM: Counting and Scoring Workflow



		T X		5			N.					Co	onc (µ	ıM)	
ΡΤΙ	Ν	Мар													
Lege	nd	l: 1 µl	Μ	5	ŏμ	Μ	2	0 µ	Μ	50 µm					
MIET	YN	QTSP	RS	SA	ΑΤ	Gl	PIS	6 M	KIF	MYLLT	V FI	LITQ	MIG	SA	
LFAV	YLI	HRRL	Dł	KIE	D	ER	NLI	H EI	DF\	/FMKTI	QR	CNT	GER	SLS	
LLNC	EE		FE	G	=V	KC		L N	KE	ETKKEI	NS I	FEM	QKG	DQI	NP
	HV	ISEA S	SS	KT	TS	SV	LQV	NA	EK	GYYTM	SN	NLV	TLE	NGK	Q
LTVK	RQ	GLYY	IY	AG	۷۲	ΓF	CSI	NR	EA	SSQAP	FI A	SLC	LKS	SPGF	R
FERIL	_LF	RAAN ⁻	ТΗ	ISS	SA	KF	CG	Q (QSI	HLGG	FE	LQP	GAS	SVFV	/N
VTDP	SQ	VSHG	т	GF	TS	SF	GLL	K		(Y145)	(146		2		
										()	C194	4) S(D3H
ΡΤΙ	Ν	Sco	ri	in	g									-	
Position	AA	Mod Mass	1 1	Conc 5	; (μΜ 20) 50	(Total	Counts Avg	s SD	Unmodified Forms	Q	γ	w	U	PTM Score
148	М	15.994915	3	7	10	15	35	8.75	5.06	37	0.698	0.872	0.479	0.807	100.0
145	Y	44.985078	0	1	11	10	22	5.50	5.80	51	0.755	1.000	0.301	0.850	82.1
146	Y	44.985078	0	1	8	10	19	4.75	4.99	54	0.817	0.860	0.260	0.850	66.0
157	Ν	0.984016	11	10	14	12	47	11.75	1.71	26	0.759	0.294	0.644	0.715	43.6
113	М	15.994915	0	0	0	1	1	0.25	0.50	0	0.984	0.086	1.000	0.807	29.0
117	D	21.981943	1	3	3	2	9	2.25	0.96	5	0.802	0.165	0.643	0.727	26.3
114	Q	0.984016	0	0	0	1	1	0.25	0.50	0	0.984	0.086	1.000	0.715	25.7
151	Ν	0.984016	7	7	5	3	22	5.50	1.92	51	0.725	0.330	0.301	0.715	21.9
160	Q	0.984016	4	1	5	3	13	3.25	1.71	45	0.885	0.294	0.224	0.715	17.7
	_		I _	-	-	_				1		–	–		

ogond	٨	(125 p	RЛ`		D	164	2 0	(л)	C (2)		Л			
.egena	A	(125 11	IVI)	D	(0/	2 111	VI)	C (2;	5 11	VI)			
PS_280	SY	SYSMEHFRWG								(M4) O				
PS_321	DF	DRVYIHPFHL								(Y4) NO2				
PS_500	EA		A/	S	A	AP	LR			(S4, S10) GalNAc				
										(K4) Ac				
PS_532	DF	orin	9]	PG			GV		Inmodified	<mark>(K4</mark>	<mark>.) Ac</mark>			DTM
PS_532	DF SC on AA	NKFH Oring Mod Mass	9	B	C	TAI	GV Counts Avg	SD L	nmodified Forms	(K4	. <mark>) Αс</mark>	w	U	PTM Score
PS_32 Peptide Posit	DF SC on AA	NKFH1	9 A 8	B 5	C .	TAI C Total	GV Counts Avg	SD (L 2.52	Inmodified Forms	(K4 Q 0.899	•) Αc γ 1.000	w 1.000	U 0.885	PTM Score
PS_321 4 PS_532 4	DF SC on AA	NKFH7	9 A 8 4	РС В 5 1	C .	Total	GV Counts Avg 5.33 2 1.67 2	SD 2.52 2.08	Inmodified Forms	(K4 Q 0.899 0.899	 Ac γ 1.000 0.827 	W 1.000 1.000	U 0.885 0.617	PTM Score 100.0 57.7
PS_532 Peptide Posit PS_532 4 PS_532 4 PS_500* 10	DF SC on AA Y K s	NKFH1 Orin(Mod Mass 44.985078 42.010565 203.079373	9 A 4 4	РС В 1 3	C .	TAI C Total	GV Counts Avg 1.67 2 2.33 2	SD 2.52 2.08 2.08	nmodified Forms	(K4 Q 0.899 0.899	 Ac γ 1.000 0.827 0.827 	W 1.000 1.000 0.368	U 0.885 0.617 0.821	PTM Score 100.0 57.7 28.3

Epidermal growth factor receptor	Human corneal-limbal epithelial cells	Phosphoe IP / I	nrichment MAC
PTM Map			
Legend: Cont	trol ATP EGF		
MRPSGTAGAA	··· WMIDADSRPK F	RELIIEFSK M	ARDPQRYLV
	PTDSNFYRA LMDI	EEDMDDV VD	ADEYLIPQ
QGFFSSPSTS R	TPLLSSLSA TSNN	STVACI DRNG	LQSCPI
KEDSFLQRYS S	SDPTGALTED SIDD	TFLPVP EYING	QSVPKR
PAGSVQNPVY	INQPLNPAPS RDP	HYQDPHS TA	/GNPEYLN
TVQPTCVNST F	DSPAHWAQK GSH		QDFFPKE
AKPNGIFKGS T		SEFIGA	\backslash
(Y1110) PO4	(Y1197) PO4 (Y9	98) PO4 (Y1	172) PO4

PTM Scoring

Position	AA	Mod Mass	Tr Ctrl	eatme ATP	ent EGF	(Total	Counts Avg	s SD	Unmodified Forms	Q	γ	W	U	PTM Score
987	М	15.994915	0	4	6	10	3.33	3.06	7	0.401	0.661	0.588	0.920	56.0
1172	Υ	79.966331	0	1	2	3	1.00	1.00	0	0.878	0.217	1.000	0.265	19.6
998	Y	79.966331	0	8	8	16	5.33	4.62	16	0.369	1.000	0.500	0.265	19.0
1110	Υ	79.966331	0	1	0	1	0.33	0.58	2	0.983	0.125	0.333	0.265	4.2
1104	S	79.966331	0	1	1	2	0.67	0.58	1	0.109	0.125	0.667	0.265	0.9
993	т	79.966331	0	2	0	2	0.67	1.16	22	0.267	0.250	0.083	0.265	0.6
995	S	79.966331	0	2	2	4	1.33	1.16	21	0.083	0.250	0.160	0.265	0.4

Results Windows









Uniqueness Occupancy Grouping $\gamma_{mip} = \frac{\sigma_{mip}}{\sigma_{mip}}$ $\mathbf{Q_{mip}} = \frac{\langle \mathbf{B_{mip}} \rangle}{\langle \mathbf{B^0}_{ip} \rangle}$ $\mathbf{W}_{mip} = \frac{\mathbf{W}_{mip}}{\left(\sum_{m=1}^{M} \mathbf{N}_{mip}\right) + \left(\mathbf{N}^{0}_{ip}\right)}$ $U_{mp} = 1 - 1$ $\overline{\sum_{m=1}^{M} \sum_{i=1}^{I} \mathbf{N}_{mip}}$ • Rarity of specific PTM on specific protein. Degree of modification of specific site • Quality of database search based on Standard deviation (σ) of specific PTM with specific PTM on specific protein. on specific site across groups; normalized MS/MS ID scores (B). • Ratio of (specific PTM on all sites) to (all by highest σ across proteins. • Ratio of (specific PTM on specific site) PTMs on all sites) subtracted from 1. • Ratio: (avg score of peptides with specific to (all PTMs and vacancies on site). PTM on specific site) to (avg score of • High score \rightarrow group specificity for • High score \rightarrow less enrichment of specific specific PTM and site (max = 1). unmodified peptides with specific site). • High score → more susceptible to PTM (max = 1).

Conclusions

- > PTM counting is a powerful technique for rapid semi-quantitative integration of large MS-based proteomics data sets.
- > STRAP PTM is a novel counting approach with 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 large data sets.



> Select STRAP PTM results correlate well with label-free results from both simple and complex data sets.

Acknowledgments: Funding was provided by NIH-NCRR grants P41 RR010888/GM104603, S10 RR015942,

S10 RR020946, S10 RR025082 and NIH-NHLBI grant R21 HL107993 and contract HHSN268201000031C.