

Quantitative Analysis on the Public Protein Prospector Web Site

Peter R. Baker¹, Nicholas J. Agard²,
Alma L. Burlingame¹ and Robert J. Chalkley¹

¹ Mass Spectrometry Facility, Dept. of Pharmaceutical Chemistry, University of
California, San Francisco, USA

² Wells Lab, Dept. of Pharmaceutical Chemistry, University of California, San Francisco,
USA

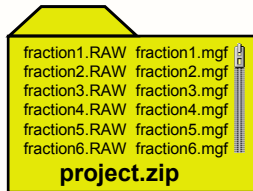
Introduction

The public web site <http://prospector.ucsf.edu> is now capable of MS based quantitative proteomics.

- All quantification types supported for Thermo XCalibur, Sciex Analyst and ABI TOFTOF Oracle software.
- iTRAQ and TMT supported for other systems.
- Quantification types include SILAC (include multiple quantification states), iTRAQ (4-plex and 8-plex), TMT (2-plex and 6-plex), ICAT, BEMAD and O16/O18.
- Peak areas/intensities can be reported for label free quantification
- Implemented as a cluster with multiprocessor LINUX computers for the database search and Windows computers for the quantitative analysis.

Data Archive

The first step is to create an archive file (zip, rar, 7z, tgz, etc) containing the centroid and raw data files. The exact contents of this file are instrument dependent¹. For example for Xcalibur files each RAW file requires a corresponding centroid file in mgf format or a set of dta files.



The mgf files must be created in the format used by PAVA² and Mascot distiller.

dta files must be have file names like those used by extract_ms/lcq_dta that indicate the RAW file name and scan number (eg fraction1.1093.1095.2.dta).

Database Search

The database search is done using the Batch-Tag Web form. Subsequent searches on the same data set should use the Batch-Tag form.

Batch-Tag Web

The labelled amino acids are selected as variable or constant modifications. Generally you would use variable modifications for MS based quantification and constant modifications for quantification based on MSMS reporter ions.

The archive file containing the centroid and raw data files are uploaded here.

Viewing the Results

Once the search has finished the results can be viewed using the Search Compare program. The Search Compare form has many options for displaying the results in different ways.

This menu includes a tab delimited text option to allow export to a spreadsheet for further analysis.

To display the results:

- 1). The Raw Type menu must be set to Quantitation.
- 2). The appropriate option must be selected on the Quantitation menu.
- 3). The resolution must be set to an appropriate value.
- 4). One of the quantitation columns must be selected.

Protein Report (SILAC L and R)

Project name derived from uploaded file name

Results name. Multiple searches can be run on the same data set with different parameters

Rank	Acc #	CAF_B18_C6toC11_QSTAR/CAF_B18_C6toC11_QSTAR_SILAC										Protein MW	Species	Protein Name	
		Num Unique	Peptide Count	% Cov	Best Expect Val	Med L/H I	Q1	Q3	Mn L/H I	-2.0σ	+2.0σ				Num
1	P13796	55	152	77.8	7.3e-10	1.58	1.35	1.88	1.63	0.895	2.96	100	70290.0	HUMAN	Plastin-2
2	P02787	40	117	64.3	7.3e-9	12.8	7.82	20.0	13.3	2.95	60.1	82	77050.6	HUMAN	Serotransferrin
3	P11142	38	87	48.8	4.2e-12	0.888	0.657	1.13	0.869	0.458	1.65	59	70898.7	HUMAN	Heat shock cognate 71 kDa protein
4	P62158	21	34	61.1	1.5e-11	1.41	1.30	1.53	1.41	1.08	1.84	16	16837.7	HUMAN	Calmodulin
5	P63104	20	56	63.3	3.4e-9	1.40	1.25	1.73	1.58	0.743	3.35	35	27745.3	HUMAN	14-3-3 protein zeta/delta
6	P07195	18	32	50.6	3.4e-8	7.84	6.97	10.6	8.62	4.49	16.5	21	36638.8	HUMAN	L-lactate dehydrogenase B chain
7	Q16531	16	16	15.4	6.4e-7	1.36	1.04	1.43	1.24	0.737	2.09	12	126969.0	HUMAN	DNA damage-binding protein 1
8	P02769	16	16	34.8	3.5e-6	12.1	5.60	18.0	10.8	2.45	47.4	15	69294.1	BOVIN	Serum albumin
9	P52566	11	12	55.7	1.5e-9	2.11	1.98	2.33	2.15	1.55	2.97	7	22988.2	HUMAN	Rho GDP-dissociation inhibitor 2
10	P30101	13	16	25.3	1.4e-7	0.838	0.719	0.905	0.790	0.534	1.17	12	56782.9	HUMAN	Protein disulfide-isomerase A3

Protein level quantification statistics

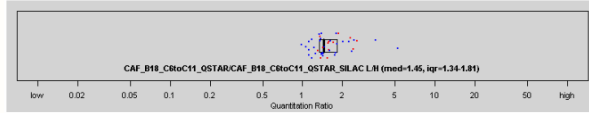
Peptide count column can be used in conjunction with the keep replicates option for spectrum counting

Link to single protein peptide report

Peptide Report (SILAC L and R)

Num Unique	% Cov	Best Expect Val	Med L/H I	Q1	Q3	Mn L/H I	-2.0σ	+2.0σ	Num
55	77.8	7.3e-10	1.45	1.34	1.81	1.60	0.834	3.08	44

Protein level quantification statistics



Scatterplot of quantification ratios

m/z	z	ppm	Peptide	Fraction	RT	Int Pk 1	Int Pk 2	L/H Intensity	Score	Expect	# in DB
824.0327	3	-24	EGIC(Carbamidomethyl)AIGGTSEOSSVVGTOHSYSEEEK	X7030611	21.22	149.8	149.8		54.5	7.3e-10	2
838.3996	2	-29	FSLVGGGODLNEGNR	X7030611	40.28	755.9	586.4	1.59	59.1	1.6e-8	1
691.7876	2	-18	GSVSDDEEMMELR	X7030611	36.35	547.9	459.1	1.37	48.7	7.1e-8	2
699.7916	2	-8.5	GSVSDDEEMM(Oxidation)ELR	X7030611	22.05	84.08	69.85	1.38	48.5	8.6e-8	2
880.0833	3	-29	HVIPMNPNTNDFLNAVGGGIVLC(Carbamidomethyl)K	X7030611	43.09	291.1	217.0	1.54	53.2	9.0e-8	1
789.8947	4	-15	AYYHLLLEOVAPKGEEGVPAVIDMSGRLR	X7030611	44.36	94.03	95.37	1.30	51.4	1.7e-7	1
720.9849	3	-32	WDTDGNGYISFENLNDLFK	X7030611	49.79	420.6	358.5	1.35	48.5	4.1e-7	1
477.8221	3	-7.9	AYYHLLLEOVAPK	X7030611	36.63	1224	1041	1.35	46.9	4.4e-7	2
626.6065	3	-26	EITENLMATGDLDDQGR	X7030611	37.96	439.1	341.8	1.58	53.1	4.5e-7	1
378.2160	3	-21	IKVPVDWNR(Label:13C(6))	X7030611	32.63				37.0	5.9e-7	2
699.7916	2	-8.5	GSVSDDEEM(Oxidation)MELR	X7030611	22.05	84.08	69.85	1.38	43.9	9.0e-7	2
702.3199	2	-16	MI(Oxidation)YMTVFAC(Carbamidomethyl)LMGK	X7030611	46.11	504.4	488.5	1.11	34.2	2.6e-6	3
442.2278	3	-4.2	IGNFSTDIKDSK	X7030612	21.33	242.9	242.9		51.5	2.7e-6	1
793.3976	2	-37	VYALPEDLVEVNPK	X7030611	40.7	1160	824.1	1.62	46.2	4.4e-6	2
1349.6995	2	-37	IISTSLPVLDLIDAIOPGSIYDILLK	X7030611	58.82	122.3	86.57	2.01	46.8	5.8e-6	2
581.9475	3	-17	GDEEGVPAVIDMSGRLR	X7030612	42.69	71.51	48.24	1.70	35.3	1.2e-5	1

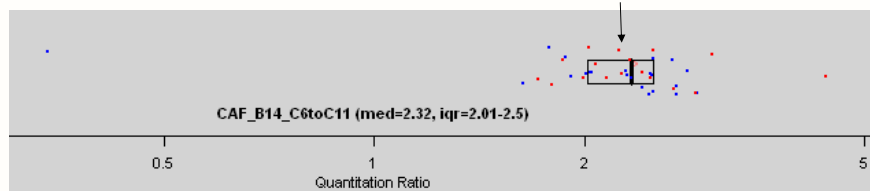
Link to raw data display

Number of times the peptide occurs in the database

Details of Quantification Ratio Scatterplot

In quantification experiments the ratios measured for a given protein often vary widely for a number of reasons. Scatterplots are one way of showing how reliable the protein level quantification ratio is.

Boxplot indicating median and interquartile range



Red: Peptide is only found in this database accession
 Blue: Peptide is found in multiple database accessions
 <>: 1 measurement peak is below a defined SNR

3-Way SILAC - Methods

This project³ compared three different ways of inducing pyroptosis (a form of programmed necrosis). The medium and heavy samples are LPS and ATP treated or monosodium urate treated respectively- both of which induce pyroptosis through a protein complex termed the NALP3 inflammasome. The light samples were transfected with dsDNA (poly dA:dT), inducing pyroptosis through an AIM2 inflammasome. We examined caspase-1 mediated proteolysis in these samples using a previously published N-terminal isolation technique⁴. Interestingly, the ratios of the caspase-cleaved peptides in the NALP3 inflammasome samples remained constant, but the ratio of the NALP3 inflammasome to the AIM2 peptides varied from peptide to peptide. This suggested that the nature of the inflammasome regulated caspase substrate profiles.

As we were looking at semitryptic N-termini rather than tryptic peptides from a full protein, there's no reason to expect that multiple peptides in the same protein would be present at the same ratio. The samples had systematic bias with the light peptides being most abundant.

3-Way SILAC - Results

1 Acc. #: [P04692](#) Gene: [CPNS1_HUMAN](#) Species: HUMAN Name: Calpain small subunit 1

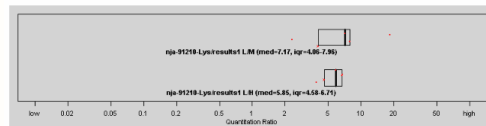
Protein MW: 28316.0 Protein pI: 5.0 Protein Length: 268

```

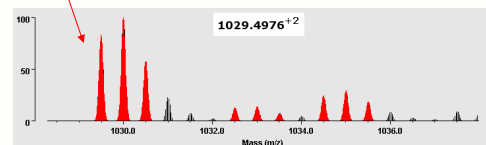
1  NPLVNSPLNG GGGGGGGGGG LGGGLGHWLG GLISGAGGGG GGGGGGGGGG GGGGGGTAMR ILGGVTSATG KARAGYNSPP
01  RPPRTHTYND KANREIVYRQ PEPFLAAGLD DQREVSATKL HPLLNQVYTS HPLDRTD-GFG IDTCREHVAV HSDYTFKRLG
141  PEPKRYLQND EHWALQNG PGTGGTTC GSEPLGATKA ADFLREHET HMLIRYDDE SDMDYNNFI SCLYPLDAMF
241  RARFSLKNDG TQDQYDQDE WLQLTETS
    
```

Coverage map shows the identified peptides.

Num Unique	% Cov	Best Disc Score	Best Expect Val
6	7.8	2.64	7.0e-5

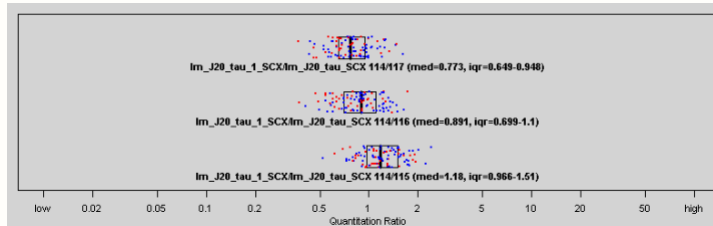


m/z	z ppm	Peptide	Fraction	RT	L/M Intensity	L/H Intensity	Expect
1029.4976	2	SerTyr-AISEAAADYNPEPPPPR	T9121012	31.5527	7.17	3.92	7.0e-5
805.414	3	SerTyr-QVISEAAADYNPEPPPPR	T9121012	47.8567	>4.06	>4.58	4.0e-4
745.8513	3	SerTyr-QVNPPEPPPPR(Label:13(6))	T9121013	22.1376	2.33	6.77	4.1e-4
793.8723	2	SerTyr-AADYNPEPPPPR	T9121013	23.0233	18.6	6.71	8.0e-4
756.3643	2	SerTyr-AADYNPEPPPPR	T9121013	22.8590	7.96	5.85	0.0013
1024.5812	2	SerTyr-AISEAAADYNPEPPPPR(Label:13(6))SN(4)	T9121012	31.6368			0.0017

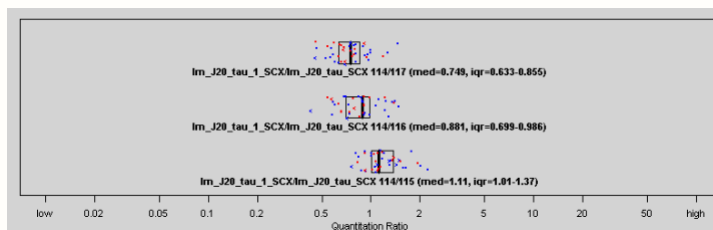


The raw data can be viewed by clicking on the masses in the m/z column.

iTRAQ Quantification



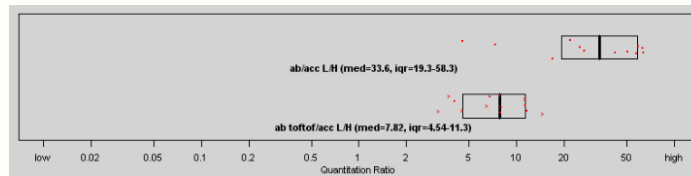
Scatterplot showing ratios for all identified peptides from a protein



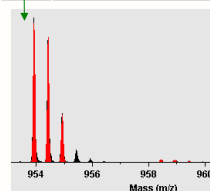
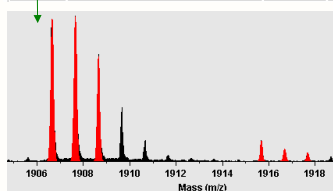
Scatterplot with low intensity peptides eliminated

Comparing Searches

Searches can be compared with each other. A TOF/TOF and QSTAR analysis of a sample reveals a greater dynamic range in the QSTAR data.



ab tof/acc				ab/acc			
m/z	Peptide	L/H Intensity	Expect	m/z	Peptide	L/H Intensity	Expect
1306.6483	STEYGEYAC(IKAT-C)DIDLR	7.82	2.4e-8	63.7	STEYGEYAC(IKAT-C)DIDLR	63.7	1.2e-10
1349.5642	IAETC(IKAT-C)YVPLR	11.7	3.1e-4	58.8	IAETC(IKAT-C)YVPLR	58.8	1.1e-7
1361.5508	MI(Oxidation)AETC(IKAT-C)YVPLR	>6.48	0.0052	50.6	MI(Oxidation)AETC(IKAT-C)YVPLR	50.6	1.8e-7
				57.7	IEC(IKAT-C)IGANDM(Oxidation)K	2.4e-7	2.4e-7
1223.5250	IEC(IKAT-C)IGANDM(Oxidation)K	>14.7	6.6e-6	57.7	IEC(IKAT-C)IGANDM(Oxidation)K	57.7	9.1e-7



Conclusions

- Protein Prospector quantification capabilities now implemented on the public web site
- Support for all major isotopic labelling strategies
- Approach has been tested on very large multi-fraction datasets
- Quantification information available at the protein and peptide label
- Scatterplot allows easy assessment of protein level quantification
- Easy access to raw data to detect and eliminate outliers
- Facilities for investigating and correcting for incomplete isotopic labelling
- Tab delimited output allows data to be exported to external program
- Data is kept on the server allowing subsequent database search and comparison of data sets

Acknowledgements

This work was supported by NIH NCRR grant RR001614, SIG RR019934 and the Biotechnology and Biological Sciences Research Council of the UK. We also thank the Vincent Coates Foundation for support.

This software is freely available on the web at <http://prospector.ucsf.edu>

References

1. http://prospector.ucsf.edu/prospector/html/instruct/batchtagman.htm#vendor_specific
2. Lynn A. J., Chalkley R. J., Baker P. R., Medzihradszky K. F., Guan S. and Burlingame A. L., The Effect of Peaklist Generation Software on Database Search Results, **56th ASMS Conference of Mass Spectrometry and Allied Topics**, Denver, Colorado June 1st - June 5th 2008
3. Agard N. J., Maltby D. and Wells J. A., Inflammatory Stimuli Regulate Caspase Substrate Profiles, **Molecular and Cellular Proteomics**, Vol. 9, No. 5, Pp. 880-893 (2010)
4. Mahrus S., Trinidad J. C., Barkan D. T., Sali A., Burlingame A. L., Wells J. A., Global Sequencing of Proteolytic Cleavage Sites in Apoptosis by Specific Labeling of Protein N Termini, **Cell**, Vol. 134, No. 4, 868-876 (2008)