Development of Targeted Proteome Assays: Mouse/Rat Brain Assay for 112 Proteins

Christopher Colangelo,[1,2] Thomas Abbott,[1,2] Gordana Ivosev,[3] Lisa Chung,[1,2] Mark Shifman,[1,2] Fumika Sakaue,[2] David Cox,[3] Lyle Burton,[3] Stephen Tate,[3] Erol Gulcicek,[1,2] Ron Bonner,[3] Jesse Rinehart,[4] Angus C. Nairn,[2] and Kenneth Williams,[1,2]

[1]W.M. Keck Foundation Biotechnology Resource Laboratory, [2]Yale/NIDA Neuroproteomics Center, and [4]Department of Cellular & Molecular Physiology and Systems Biology Institute, Yale University School of Medicine, New Haven, CT;

Abstract

A comprehensive workflow has been designed for the development of large scale (>1,000 transitions/run), 90 min LC-MRM Targeted Proteome Assays that can determine the relative concentrations of 50-200 pre-selected protein biomarkers of interest. The workflow begins with an AB SCIEX TripleTOF® 5600 MS that "sequences" peptides from a tryptic digest of an extract from the sample of interest. The Yale Protein Expression Database (YPED) translates the "learned" peptide sequences into an extended LC-MRM (xMRM) assay that is run in triplicate on an AB SCIEX QTRAP® 5500 MS. The resulting LC-xMRM data are processed with MultiQuant™ software utilizing an optimized SignalFinder™ Research algorithm and exported to Excel. A suite of custom-designed bioinformatics tools then provide assay metrics, data normalization, and peptide and protein fold change calculations. The resulting data are imported into YPED where users view and download their data through a secure Web interface. Improvements in assay development, data processing, and data analysis tools were implemented that greatly increase the speed at which large scale, scheduled LC-MRM analyses can be designed and the throughput of the resulting targeted proteome assays. Our initial efforts resulted in a robust LC-MRM assay for rat/mouse brain cortex that can routinely quantify up to 112 proteins from 24 different protein classes using 15 data points/protein (3 peptides/protein x 5 transitions/peptide). This LC-MRM assay is now available as a service through the Keck Laboratory. Translation of our LC-MRM workflow into LC-SWATH on the AB SCIEX TripleTOF® 5600 MS permitted direct comparative analysis of six biological replicate samples that were each run in triplicate. For the LC-SWATH analyses the same 1,697 transitions were extracted as were used for our xMRM assay. These studies demonstrated excellent agreement between these two platforms and uncovered a potentially important experimental variable with regards to sample preparation.

Introduction

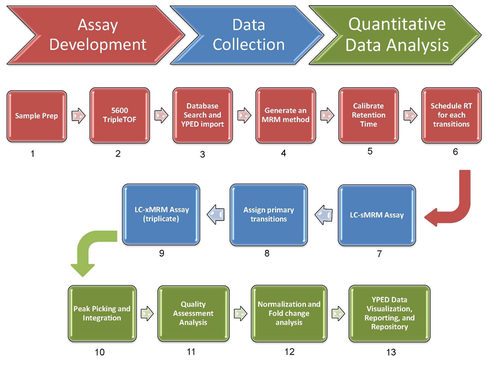
For ~15 years, large scale proteomic discovery has relied on massive LC-MS/MS to profile proteins in complex extracts. Problems with this approach are the limited dynamic range, poor run to run protein identification reproducibility, and the wide range in the number of peptides isolated from each identified protein. The latter results in MS/MS sequencing of many more peptides (>3) from some proteins than are needed to identify the parent protein. With complex mixtures this approach also must be coupled with off-line fractionation that results in numerous LC/MS/MS runs that require tens of hours of MS instrument time to detect and quantify 1,000-10,000 proteins in a complex mixture. As an example of the enormous duplication of effort with this approach, from 2007 through 5/27/2013 the MS/Proteomics Resource in the Keck Biotechnology Resource Laboratory at Yale University has sequenced and stored 47,838,039 peptides (with a FDR of 0.01) in the Yale Protein Expression Database (YPED, Shifman et al, 2007) that together represent only 3,140,349 distinct sequences or 6.6% of all YPED data. Assuming that we continue to use the same LC-MS/MS approaches, then 93.4% of our mass spectrometry instrument time will be wasted by resequencing the same peptides in each experiment. As a better approach, we are developing 90 min LC-MRM Targeted Proteome Assays (TPAs) that relatively or absolutely quantify 50-200 biomarker proteins of interest.

Key Features of Multiple Reaction Monitoring (MRM)

- Well proven technology used for >30 years to quantify a wide range of small molecules in clinical samples.
- Wide linear dynamic range of up to 5 orders of magnitude.
- Very high precision, with a multi-site study involving 8 laboratories across the U.S. demonstrating that MRM assays are highly reproducible (Adonna et al, 2009).
- High sensitivity, detects ng/ml amounts of peptides in biological fluids and tissue extracts.
- Broad applicability, once an MRM assay has been developed for a target protein in a particular sample type it has a high probability of being applicable to the measurement of that protein in other sample types.
- In recognition of its important role in hypothesis-driven research and its increasing impact on clinical proteomics, Targeted Proteomics was chosen as the Nature Method of the Year for 2012.

Targeted Proteome Analysis Workflow

An AB SCIEX TripleTOF® 5600 MS was used to "sequence" peptides from a tryptic digest of the sample type of interest. Yale Protein Expression Database (YPED) was then used to translate the "learned" peptide sequences into a triggered LC-MRM (xMRM) assay that was run in triplicate on a AB SCIEX QTRAP® 5500 MS. The resulting LC-xMRM data was processed with MultiQuant software utilizing a newly developed SignalFinder Research algorithm and exported to Excel. Peak areas from corresponding LC-SWATH analysis were extracted in Peakview using transitions from corresponding IDA runs. A suite of bioinformatics tools then provided assay metrics, data normalization, and peptide and protein fold change calculations. The resulting data were imported into YPED where users can view and download their data through a secure Web interface.

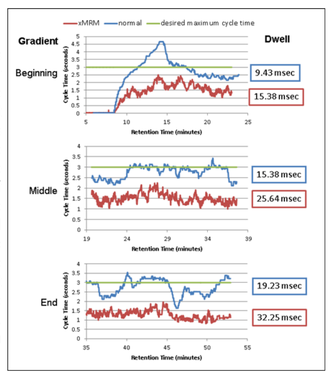


Yale Protein Expression Database (YPED)

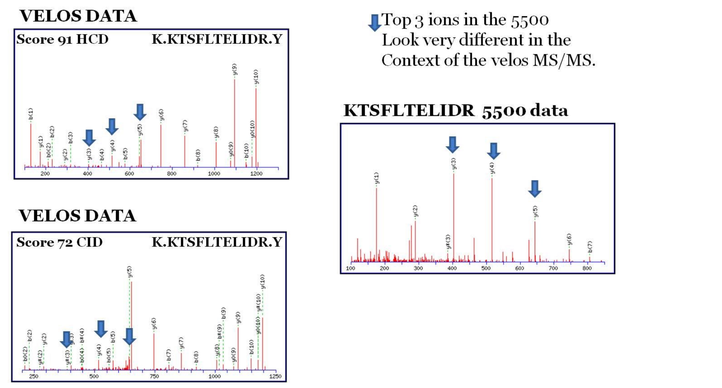
We have developed an integrated, web-accessible software system called the Yale Protein Expression Database, or YPED, to address the need for storage, retrieval, and integrated analysis of high throughput proteomic and small molecule MS analyses. The interface supports sample submission, project management, sample tracking, data import, sample administration, and user billing. For data integration, YPED handles data from: LC-MS/MS protein identifications and protein post-translational modifications (phosphorylation, ubiquitination, acetylation, methylation, and others); identification/quantitation results from label-based proteomics experiments (DIGE, iTRAQ, ICAT, and SILAC); LC-MS based label-free quantitative (LFQ) proteomics; and targeted proteomics (MRM). We developed a YPED tool to automatically transform discovery data into targeted MRM methods to construct a targeted MRM proteome assay. Data from discovery runs on a AB SCIEX TripleTOF® 5600 MS were database searched and peptide identifications were uploaded to YPED which outputs either a scheduled LC-MRM method for the AB SCIEX QTRAP® 5500 MS or a Peakview input file for SWATH acquisition on our AB SCIEX TripleTOF® 5600 MS. YPED also serves as a peptide spectral library for all our protein database search identification results. YPED contains >15,000 datasets from >1,300 users and has >3 million unique peptides identified (with a Mascot score greater than or equal to the homology score) from >650,000 unique proteins, including 20,843 human and 20,059 mouse. YPED provides a powerful resource for supporting MRM and SWATH proteome technologies and MS/MS based protein identifications.

Triggered xMRM

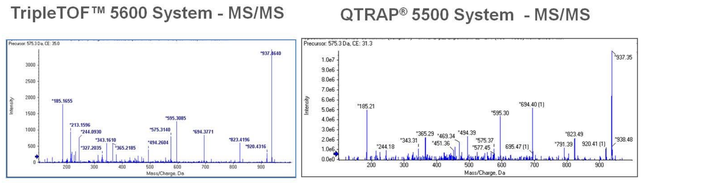
Triggered xMRM was used to improve quantitation. With this approach designated primary MRM are monitored throughout their entire scheduled window, while secondary MRM for each peptide are only monitored if the primary MRM exceeds a preset threshold. This approach reduced the number of MRM transitions being monitored at any given time, thus improving dwell time while decreasing cycle time. The figure on the right shows the cycle times for three transitions from the beginning, middle, and end of the LC-MRM gradient. The red line is the cycle time using the xMRM assay and the blue line is from the normal scheduled MRM (sMRM) assay. The plots show that by using xMRM the cycle time for each transition significantly decreases which increases the number of data points collected across each peak. On the right hand side of the figure is the dwell time for each of these transitions in either xMRM (red box) or sMRM mode (blue box). xMRM increased the dwell time for each transition by 67% which then provides better signal to noise measurements.



Challenges in Translating Information Dependent Acquisition (IDA) Datasets



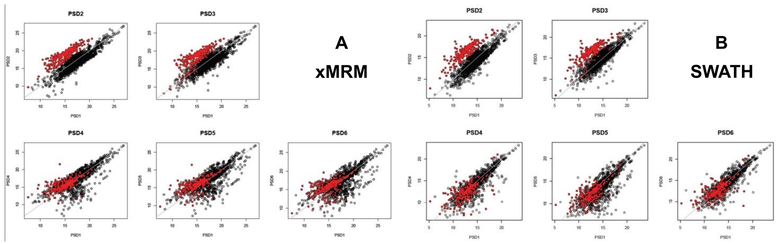
As shown above, there are substantial differences in relative peak intensities for IDA MS/MS data acquired on the LTQ-Orbitrap Velos (left panels) as compared to the AB SCIEX QTRAP® 5500 MS. As a result, it is difficult to translate IDA data acquired on the Velos into a scheduled LC-MRM analysis on the 5500 Q-Trap. In contrast, since both the AB SCIEX TripleTOF® 5600 MS and the AB SCIEX QTRAP® 5500 MS have similar sources and identical collision cells, the MS/MS peptide fragmentation patterns are very similar on both these platforms (see below) – which greatly facilitates choosing the most intense transitions to interrogate each peptide of interest.



Peptide LVNEVTEFAK; *m/z* 575.3 [2+]

X-MRM vs SWATH Fold Change Comparisons on Rat Brain PSD Samples

Tryptic digests of three rat brain cortex Post-Synaptic Density (PSD) fractions were run on a AB SCIEX TripleTOF® 5600 MS using 180 min LC-MS runs. MASCOT database searching identified 1,574 unique rodent proteins. Using this list one scheduled MRM assay of 1,697 transitions was generated for 112 proteins from 337 peptides. Six PSD cortex biological replicates were then analyzed in triplicate across the 1,697 transitions in the PSD MRM proteome assay for a total of 30,546 transitions. The same six PSD cortex biological replicates were also run in triplicate with SWATH acquisition. For comparative analysis with our xMRM assay, the same 1697 transitions were extracted in the SWATH assay. The resulting data was analyzed and a variety of graphical plots were produced using our Matlab and "R" fold change analysis tools. Figure A and B are xMRM and SWATH log 2 scatter plots between PSD 1 and the other five PSD biological replicates. The red dots indicate transitions that are four-fold "up-regulated" in PSD 2 and PSD 3 vs PSD 1. A log 2 fold change scatter plot between xMRM and SWATH for PSD2 vs PSD 1 (data not shown) gave a correlation of 0.90 which demonstrated that the two methods produce consistent fold-changes between each other.

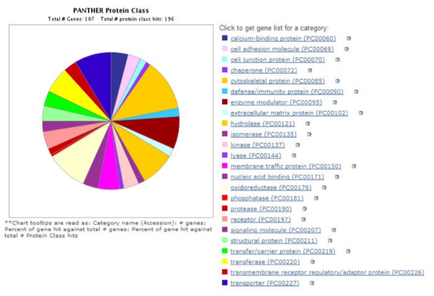


Mouse/Rat Brain Targeted Proteome Assay (MBA/112)

The Mouse/Rat Brain Proteome Assay interrogates the relative level of expression of 112 proteins (see Table) from 24 different protein classes (Figure below). The assay was developed and optimized for Rat Cortex Post-synaptic density (PSD) preparations. We have also utilized this assay for Cortex Synaptoneurosomes and Striatal PSD. When this assay is used on other brain fractions the number of quantified proteins may be somewhat less than 112.

List of Mouse/Rat Brain TPA Proteins

ADT1	CMTD1	DLGP4	IMMT	M2OM	NDUS1	NMDZ1	PPIA	SHSA7	TBB2C	VISL1
AT1A3	CNTN1	DPYL2	IQEC1	MBP	NDUS2	NRX1A	QCR1	SPTA2	TBB3	VPP1
ATPA	COX5A	DYN1	IQEC2	MPCP	NDUS3	NSF	QCR2	SPTB2	TBB5	
ATPB	CTNA2	ERC2	KCC2A	MTCH1	NDUS7	ODO1	RAB35	SRCN1	TENR	
BAIP2	CTRO	FLOT2	KCC2B	MYH10	NDUS8	ODP2	RIMS1	STX1A	TOM40	
BASP1	DLG2	GDL1	KCC2D	MYO5A	NDUV1	ODPB	RL9	STX1B	TOM70	
BEGIN	DLG3	GDAP1	KCC2G	MYPR	NDUV2	OPA1	RPGF4	STXB1	VAID1	
BSN	DLG4	GLPK	KCD16	NDUA7	NEUM	PHB	SEPT8	SYN2	VAPA	
CI172	DLGP1	GNAZ	KIF2A	NDUA8	NFL	PHB2	SFXN3	SYNPO	VATB2	
CLH	DLGP2	HSP7C	LIPA2	NDUA9	NFM	PKP4	SHAN2	SYT1	VDAC1	
CMC1	DLGP3	HXK1	LRRC7	NDUAC	NMDE2	PLEC	SHAN3	TBA1B	VDAC2	

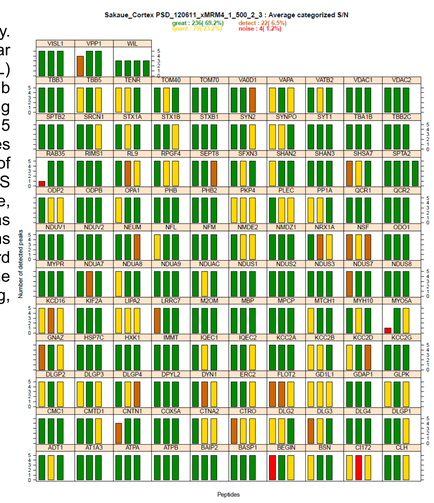


Peptide Signal/Noise Metric Plot of Data from an LC-MRM Analysis

Metric Bar Graph of PSD xMRM assay quality. The plot on the right is broken into 113 sets of bar graphs for each protein and internal standard (WIL) in the PSD assay. Each bar graph is further sub divided into individual bars with each representing the data quality of a single peptide with 5 transitions/peptide (except the WIL std peptides which have 3 transitions/peptide). The height of each bar corresponds to the number of MS/MS transitions observed for the corresponding peptide, with all proteins potentially having 5 transitions as compared to the maximum of 3 transitions monitored for each of the internal standard peptides. The color of each bar depicts the average signal/noise ratio for the underlying, usually, 5 transitions as described in Table 2.

Table 2: Criteria for Categorizing Quantitative Protein MRM Data

S/N	S/N Category		
	Description	Bar Graph Color	Number Designation
Greater than 10	Great	Green	5
Between 5 and 10	Quant	Yellow	4
Between 3 and 5	Detect	Orange	3
Below 3	Noise	Red	2
Missing peak	N/A	N/A	1



Hierarchical Clustering Analysis of Six Biological Replicate Rat Brain PSD Preparations

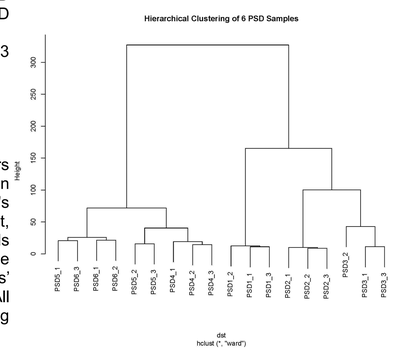
A 'bottom-up' approach was used to determine if there was any "Day Effect" that resulted from the 6 biological replicate samples of rat brain cortex PSD fractions being prepared on 3 different days (i.e, PSD 1/Day 1; PSD 2,3/Day 2; PSD 4,5,6/Day 3). Briefly:

1) Each of 18 samples (i.e., 6 biological x 3 technical replicates each) forms an initial cluster.

2) Euclidian distance is calculated between any pair of clusters.

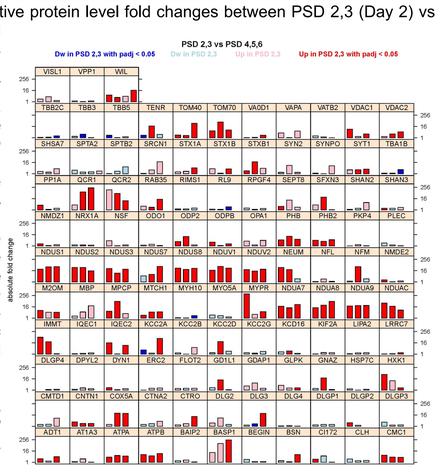
3) The pair with the smallest distance is joined together to form a new cluster.

4) Steps 2) and 3) are repeated until all the clusters are merged into one. Step 3) is performed based on one of the various linkage criteria. For example, Ward's minimum variance method aims at finding compact, spherical clusters. The "complete linkage" method finds similar clusters based on the maximum distance. The "single linkage" method adopts a 'friends of friends' clustering strategy (based on minimum distance). All approaches resulted in PSD 4,5,6 vs PSD 2,3 being grouped together.



Observation of a "Day" Effect When Comparing PSD Cortex Preparations

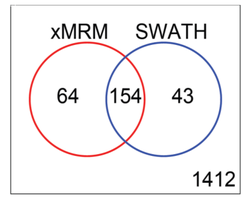
The bar chart on the right summarizes the relative protein level fold changes between PSD 2,3 (Day 2) vs PSD 4,5,6 (Day 3) samples. Across 3 technical runs, the median log2 (normalized) peak area was computed resulting in 6 observed values (PSD 1 through 6) per transition. The data has 1697 transitions, 337 peptides. For each peptide, a linear mixed model was fitted to determine the significant difference between PSD 2, 3 vs PSD 4,5,6. A group effect p-value was then calculated for each peptide that provides a measure of the significance of the difference between PSD 2,3 vs PSD 4,5,6. In this chart, each peptide has its own bar. The height of the bar indicates the fold change (PSD 2,3/PSD 4,5,6). If it is less than 1, the value is inversed. The color of the bar represents the direction of the fold change (red for PSD 2,3 > PSD 4,5,6 and blue otherwise) and the magnitude of the adjusted p-value (dark if the values < 0.05 = significant). Based on this analysis several mitochondrial proteins (e.g., NDUS1, NDUS2, NDUS3 in the 6th row from the top) are present in about 10x higher amounts in PSD 2,3 (Day 2) as compared to PSD4,5,6 (Day 3), with the "Day" referring to the Day each PSD preparation was carried out.



Protein Quantitation with SWATH

From six LC-SWATH PSD Cortex samples (18 runs) we extracted 56,000 transitions (>1,200 proteins) in each run for a total of 1,000,800 data points. After minimum variance normalization and fold change analysis of PSD 2 vs PSD 1, we were able to expand the number of >4 fold up-regulated PSD cortex proteins from 16 with the LC-xMRM assay to 101 proteins with LC-SWATH, for a 6-fold increase. The Venn diagram shows that 75% of the transitions with >4-fold up-regulation overlap when the xMRM and SWATH results were compared. The 1412 remaining transitions in the assay were below 4-fold in both xMRM and SWATH. These results demonstrate good correlation between these two technologies. As a result of these and the above findings we carefully re-examined the protocols used to prepare the 6 biological replicate control samples of PSD cortex and found an experimental variable that may explain the apparent "up-regulation" of proteins in PSD 2,3. We hypothesize that the difference between these two groups may result from the 8hrs that the PSD 2,3 (Day 2) samples sat on ice prior to Percoll gradient centrifugation.

Venn Diagram of Transitions >4-Fold Up in



Mouse/Rat Brain Proteome (MBA/112 Protein) Assay Details

Targeted Proteome Assays (TPAs) from the Keck Laboratory include tryptic digestion, C18 clean-up, and 90 min LC-MRM runs on an AB SCIEX 5500 QTRAP mass spectrometer. Quantitation of each targeted protein is based on up to 15 data points (3 peptides x 5 transitions/peptide). A minimum of 10µg protein/sample is needed and we recommend that sample protein concentrations are based on hydrolysis/amino acid analysis. We inject 1µg of protein digest for a single injection and for triplicate injections the technical replicates are run in a block randomized order. Data analysis includes:

- Peak picking and Integration
- Quality Assessment Analysis
- Fold-change analysis and Statistical Inference
- Deposition of the resulting data into YPED
- As soon as the analyses are complete the data may be retrieved using an individual, password-protected account from the web-based Yale Protein Expression Database: (http://yped.med.yale.edu)

Mouse/Rat Targeted Proteome Assay Service Charges

Service	Yale	Non-Profit
Single Analysis	$299	$359
Triplicate Analysis	$549	$609

Conclusions

- An optimized pipeline has been developed that includes Targeted Assay Development, Data Processing, and highly customized Data Analysis tools for both LC-MRM and LC-SWATH assays.
- Triggered xMRM was used to improve dwell time while also decreasing cycle time during LC-MRM runs.
- Signal Finder Research (SF2) significantly improves peak integration: SF2 reduced the % of peaks that were incorrectly integrated to 2.5%, as compared to 6% for MultiQuant (version 4).
- Data metrics and automated R plots have been developed including the Signal/Noise Peptide Metric Plot.
- Robust normalization algorithms and confidence weighted fold-change analysis has been implemented for analysis of transition to peptide to protein level LC-MRM data.
- Virtually identical fold-change values were obtained when the Pipeline was used to analyze both LC-MRM and LC-SWATH data from the Rat PSD Cortex samples
- Demonstrated the ability of SWATH to expand our LC-MRM Targeted Proteome Assays (TPA) 6-fold.
- The LC-MRM assay is sufficiently robust that it identified an experimental variable that may account for the Day Effect observed in the six PSD biological replicates—with samples #(2,3/Day 2) versus (1,4,5,6/Day 3) representing two groups of "similar" PSD samples that differ with respect to their day of preparation and a slight variation in the PSD sample preparation protocol. That is, samples #2 and #3 sat on ice for about 8 hrs prior to the Percoll gradient centrifugation step in the procedure used to prepare the rat brain cortex Post-Synaptic Density (PSD) preparation that was the starting material for the Rat Brain Proteome Assay.

References

Addona, T., Abbatiello, S., Schilling, B., et al (2009) Multi-site assessment of the precision and reproducibility of multiple reaction monitoring-based measurements of proteins in plasma, Nat Biotechnol 27, 633-641.

Shifman, M. Y. Li, C. Colangelo, K.S. KL, T. Wu, K. Cheung, P. Miller, K. Williams, YPED: a web-accessible database system for protein expression analysis, J. Proteome Res., 6 (2007) 4019-4024.

Funding

Yale NIDA Proteomics Center: P30 DA018343 (NIDA/NIH), Yale Center for Clinical Investigation: ULRR024139 (CTSA/NCRR/NIH), NIH/NIDDK K01 DK089006