UTILIZING EXPERIMENTALLY-GENERATED PROTEIN ION MAPS FROM DATA-INDEPENDENT LC-MS ACQUISITIONS FOR IDENTIFYING LOW ABUNDANT PROTEINS IN COMPLEX MIXTURES

Waters
THE SCIENCE OF WHAT'S POSSIBLE.™

Roy Martin [1]; J. Will Thompson [2]; Arthur Moseley [3]; Scott Geromanos [1]
[1]Waters Corporation, Beverly, MA; [2]Duke University, Igsp, Morrisville, NC; [3]Duke University Medical Center, Durham, NC

OVERVIEW

In general most proteomic analyzes are Discovery based experiments whereby data is collected and then queried against a database.

We will illustrate the use of a novel software tool, Verify^E, for Hypothesis driven proteomics whereby the most secure precursor and product ions of previously identified proteins, or those generated in -silico from a fasta file are queried against the time -aligned lower level signals of an MS^E experiment to increase both the number of proteins identified as well as protein sequence coverage.

INTRODUCTION

It is generally accepted that:

1) a large percentage (~75%) of the ions available for LC-MS analysis from enzymatically digested complex "systems" matrices are ~2 to 2.5 orders of magnitude lower in intensity then the most intense ions available
2) in a typical LC-MS/MS analysis tens of thousands of spectra are acquired with only 10 – 20% resulting in valid peptide identifications
3) not all peptides from a protein ionize with the same efficiency.

In many cases the protein sequence coverage from these tandem MS analyzes is low relative to the number of spectra acquired resulting in ~ 40% of the protein identifications only having 1 peptide to match.

We will demonstrate the use of a software tool, Verify^E, for the extraction of protein and peptide ion maps from MS^E acquired data searched by Identity^E to increase the dynamic range of identifiable proteins in a complex protein digest by matching the ion maps to low intensity processed data.

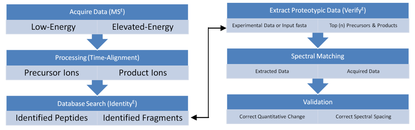
In order to graphically display MS^E data, a convention will be used that shows precursor ions as positive m/z measurements (above the zero line) and fragment ion m/z's below the line. The display size of the dots indicates relative intensity. The x-axis is chromatographic retention time (RT).

IDENTITY^E ACQUISITION METHOD

nanoACQUITY™ UPLC
nanoACQUITY™ BEH C18, 150 mm x 150 mm
Flow Rate: 1 µL/min
A: 0.1% FA in Water
B: 0.1% FA in Acetonitrile
1% B initial conditions, 1% B at 2.5 min, 5% B at 3 min, 40% B at 63 min, 85% B at 68 min, 85% B at 73 min, 1% B at 75 min
Total run time = 90 min

Q-TOF Premier™
Positive Ion V-Mode
LC-MSE
Capillary = 3 kV
Cone = 35 V
Collision Energy (CE): MS=5 volts, MSE: Ramp 15-40 volts
Scan time = 0.8 sec

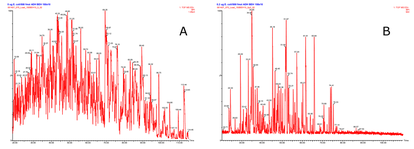
WORKFLOW

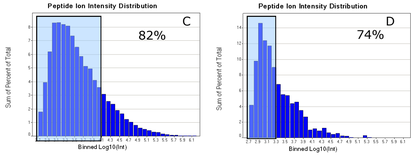


E. COLI DILUTION RESULTS

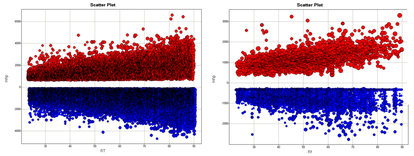
5000 ng Lysate Injection
BPI Chromatogram

200 ng Lysate Injection
BPI Chromatogram



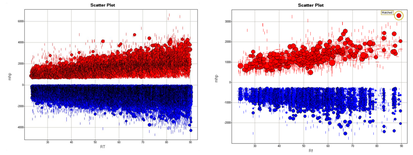


Panels A & B depict the BPI Chromatograms of the 5000 and 200 nanogram loads of digested *E.coli* . Panel C & D are bar plots illustrating the peptide ion intensity distribution. Panel C represents 167,231 *m/z* ion detections over the 90 minutes of chromatographic elution whereby Panel D (200 ngm) represents only 22,896. On average ~78% of all ion detections are 2 orders of magnitude lower in intensity than the most intense ions.



16,857 Precursors 168,857 Products

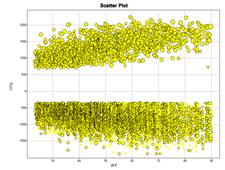
2389 Precursors 9577 Products



4295 Peptides identified to 424 Proteins

292 Peptides Identified to 58 Proteins

E. COLI CON'T



Verify^E Output
Top 5 Precursor and Product Ions from the 424 identified proteins (5000ng inj.)

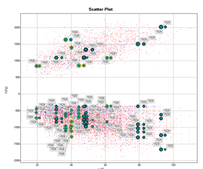
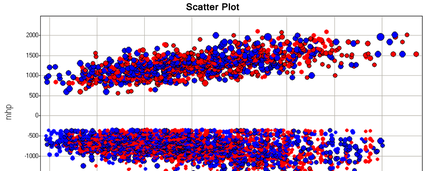
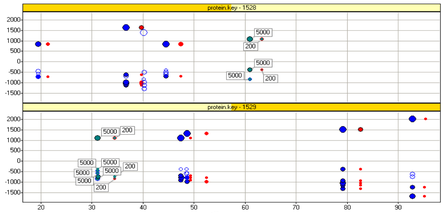


Illustration of Top 5 Matched to lower abundant *E.coli* proteins OPD1 (1528) & ODO2 (1529)



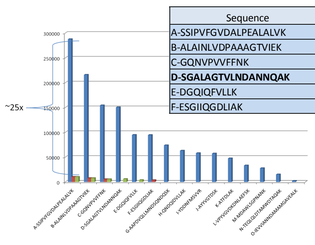
Matching Extracted Precursors and Products (Red = 200 ngm, Blue 5000 ngm) added 661 peptides, 1596 Product Ions to an additional 156 Proteins.



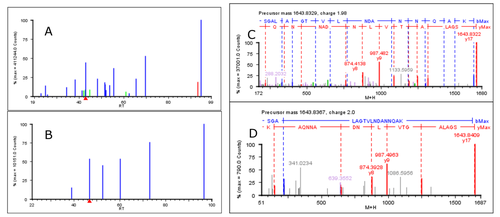
Match Status	Ion Type	mhp	retT	intensity	Loading	protein.key	peptide.seq	Entry Name
Matched	Product	-833.4234	60.92	677	5000	1528	ESVGFLVTIK	ODO2_ECOLI
Matched	Product	-373.1762	60.97	1626	5000	1528	ESVGFLVTIK	ODO2_ECOLI
Matched	Product	-361.2418	60.95	4104	5000	1528	ESVGFLVTIK	ODO2_ECOLI
Extracted	Product	-361.2416	63.30	152	200	1528	ESVGFLVTIK	ODO2_ECOLI
Matched	Precursor	1092.6261	60.93	27983	5000	1528	ESVGFLVTIK	ODO2_ECOLI
Extracted	Precursor	1092.6272	63.30	1139	200	1528	ESVGFLVTIK	ODO2_ECOLI
Matched	Product	-835.3973	31.14	3137	5000	1529	TYVPADDYR	ODP1_ECOLI
Extracted	Product	-835.3914	34.44	124	200	1529	TYVPADDYR	ODP1_ECOLI
Matched	Product	-736.3292	31.15	7119	5000	1529	TYVPADDYR	ODP1_ECOLI
Extracted	Product	-736.3292	34.44	274	200	1529	TYVPADDYR	ODP1_ECOLI
Matched	Product	-568.2458	31.11	576	5000	1529	TYVPADDYR	ODP1_ECOLI
Matched	Product	-453.2129	31.14	1303	5000	1529	TYVPADDYR	ODP1_ECOLI
Matched	Product	-364.1787	31.14	275	5000	1529	TYVPADDYR	ODP1_ECOLI
Matched	Precursor	1099.5028	31.14	25421	5000	1529	TYVPADDYR	ODP1_ECOLI
Extracted	Precursor	1099.5094	34.43	908	200	1529	TYVPADDYR	ODP1_ECOLI

The scatter plot above illustrates the ions matched to the OPD1 and ODO2 proteins from *e. Coli*. Filled blue circles are those ions generated by the Verify^E tool, open blue circles are those ions not found in the dilute sample, and red circles are those ions matched in the low level. The table illustrates an intensity differential commensurate with the loading. The fragmentation pattern, and therefore the peptide identity, is confirmed by relative intensity.

Sequence	5000ng	200ng	Extracted
A-SSIPVFGVDALPEALALVK	287654	10151	10151
B-ALAINLVDPAAAGTVIEK	215870	7709	7709
C-GQNVPVVFFNK	154321	5449	5449
D-SGALAGTVLNDANNQAK	150461		5989
E-DGQIQFVLLK	94695		3808
F-ESGIIIQGDLIAK	94460	3541	

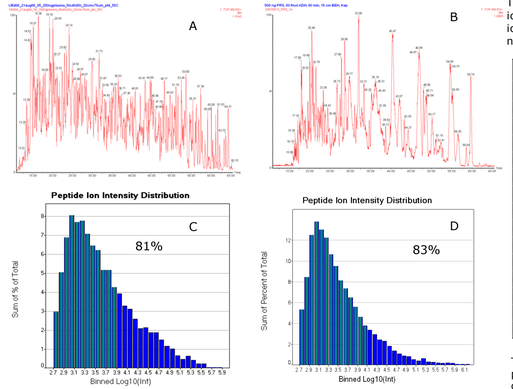


The Bar chart illustrates the peptide ionization distribution of the peptides assigned to ODO2. The intensity ratio of matched pairs is proportional to the differences in the loading

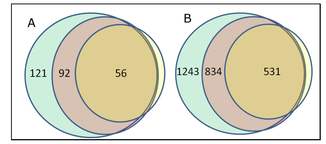


Panels A & B represent the identified (5000ng) and extracted (200ng) peptides to ODO2. Panels C & D illustrate the MS^E spectra of the peptide SGALAGTVLNDANNQAK from the matched high level and extracted low level data. The intensity ratio of the fragment ions between the two levels is proportional to the differences in the loadings, and is internally consistent for the peptide.

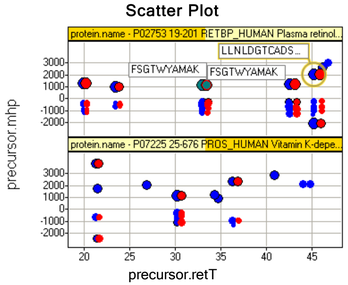
DEPLETED VS. NON-DEPLETED SERUM



Panels A & B depict BPI Chromatograms of Depleted and Non-Depleted Serum respectfully. Panel C & D are bar plots illustrating the peptide ion intensity distribution. Panel C represents 112,231 m/z ion detections over the 55 minutes of chromatographic elution whereby Panel D (Non-Depleted) represents 80,569. On average 82% of the precursor Ions are 2 orders of magnitude lower than the most intense signal.



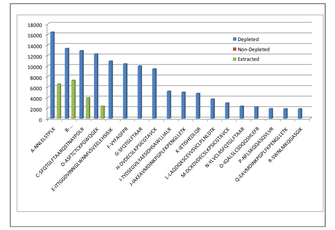
The above Venn diagrams illustrate the number of peptides and proteins identified in the depleted (green), non-depleted (yellow) and extracted datasets (rose). Querying the Ion Maps from the lower abundant proteins in the depleted serum analysis resulted in a 64% and 57% increase in the number of proteins and peptides identified in the non-depleted analysis.



The scatter plot above shows the comparison of the matched ions from proteins in the non-depleted serum to the extracted ions in the depleted serum sample. The extracted proteins did not pass identification criteria during discovery.

Sample	Ion Type	mhp	RT	Int
Depleted	Extracted	1161.538	32.97	42305
Depleted	Extracted	-235.116	32.99	3121
Depleted	Extracted	-349.19	32.96	1839
Depleted	Extracted	-583.287	32.95	1288
Depleted	Extracted	-1161.53	32.93	1209
Depleted	Extracted	-769.381	32.96	936
Depleted	Matched	-1014.49	32.92	514
Depleted	Matched	-218.148	32.96	476
Depleted	Matched	-420.232	32.97	257
Non-Depleted	Matched	1161.547	33.46	27498
Non-Depleted	Matched	-235.119	33.62	2434
Non-Depleted	Matched	-349.194	33.59	1269
Non-Depleted	Matched	-583.294	33.58	863
Non-Depleted	Matched	-1161.55	33.56	870
Non-Depleted	Matched	-769.39	33.59	730

The above Table illustrates the extracted ions matched to the peptide sequence FSGTWYAMAK from RETBP_HUMAN. The ratio of matched precursor and product ion intensities confirm a correct match.



The blue bars in the above bar chart illustrates the peptide ionization distribution of the peptides assigned to RETBP_HUMAN from the Depleted Serum sample. The green bars depict the distribution of the extracted peptides.

DISCUSSION

- Identity^E generates high fidelity protein identification based on a reproducible pattern of accurate mass and retention time correlated peptide fragment ions.
- We have shown that these patterns are also reproducible by intensity; the relative intensity of peptides are consistent, and the relative intensity of their fragments are consistent.
- These patterns, consisting of a RT, precursor, and 5 fragment ions, can be used as a "spectral library" to query other data sets.

CONCLUSION

- **Once a protein is securely identified extracting the best ionizing peptides and preferred product ions provide a means for identifying that protein at concentration levels below that necessary for database search algorithms.**
- **Extracted protein identifications can be validated by comparing the ion intensity ratios of "matched" precursor and product ions.**
- **Using fragmentation and retention-time modeling preferred precursor and product ions generated from an *in-silico* digested can be queried against a spectra database.**
- **The VerifyE tool can also be used to generate methods for subsequent analyses by MRM techniques.**

TO DOWNLOAD A COPY OF THIS POSTER, VISIT WWW.WATERS.COM/POSTERS

720002698EN