



Yale University School of Medicine
W.M. Keck Foundation
Biotechnology Research Laboratory



Christopher Colangelo, Ph.D.

**Quantitative Protein Profiling &
Targeted Protein Expression**

NIDA

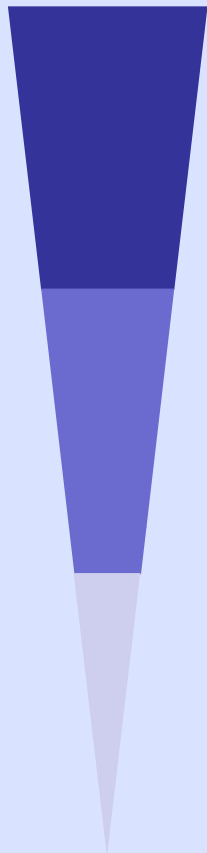
External Advisory Board (EAB) Meeting

December 3, 2008

Goals of Protein Profiling Core

- Improve and develop new protein profiling technologies to identify proteins that play key roles and/or are biomarkers in response to substance abuse
- 1. Use existing DIGE, MudPIT, iTRAQ, and SILAC pipelines on larger number of samples.
- 2. Better Protein Identification
- 3. More Accurate Protein Quantitation (replicates)
- 4. Faster Sample Turnaround
- 5. Verification of potential markers (Targeted Proteomics)

Discovery



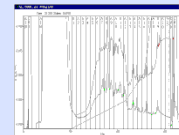
Validation



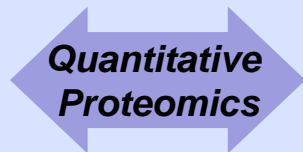
Protein ID
& PTM



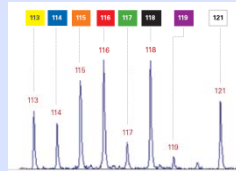
MudPIT



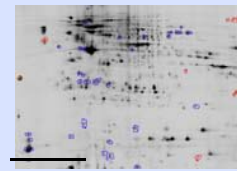
Phosphoprotein
Profiling



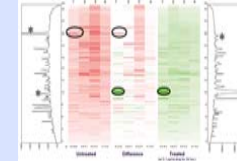
iTRAQ
ICAT



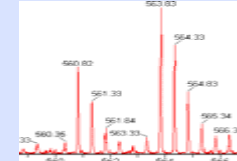
DIGE



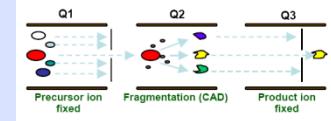
2D-Protein
Separation



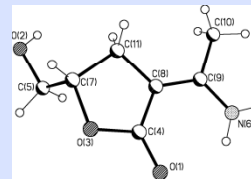
SILAC



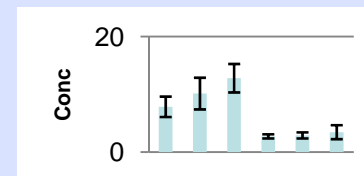
MRM
Targeted Proteomics



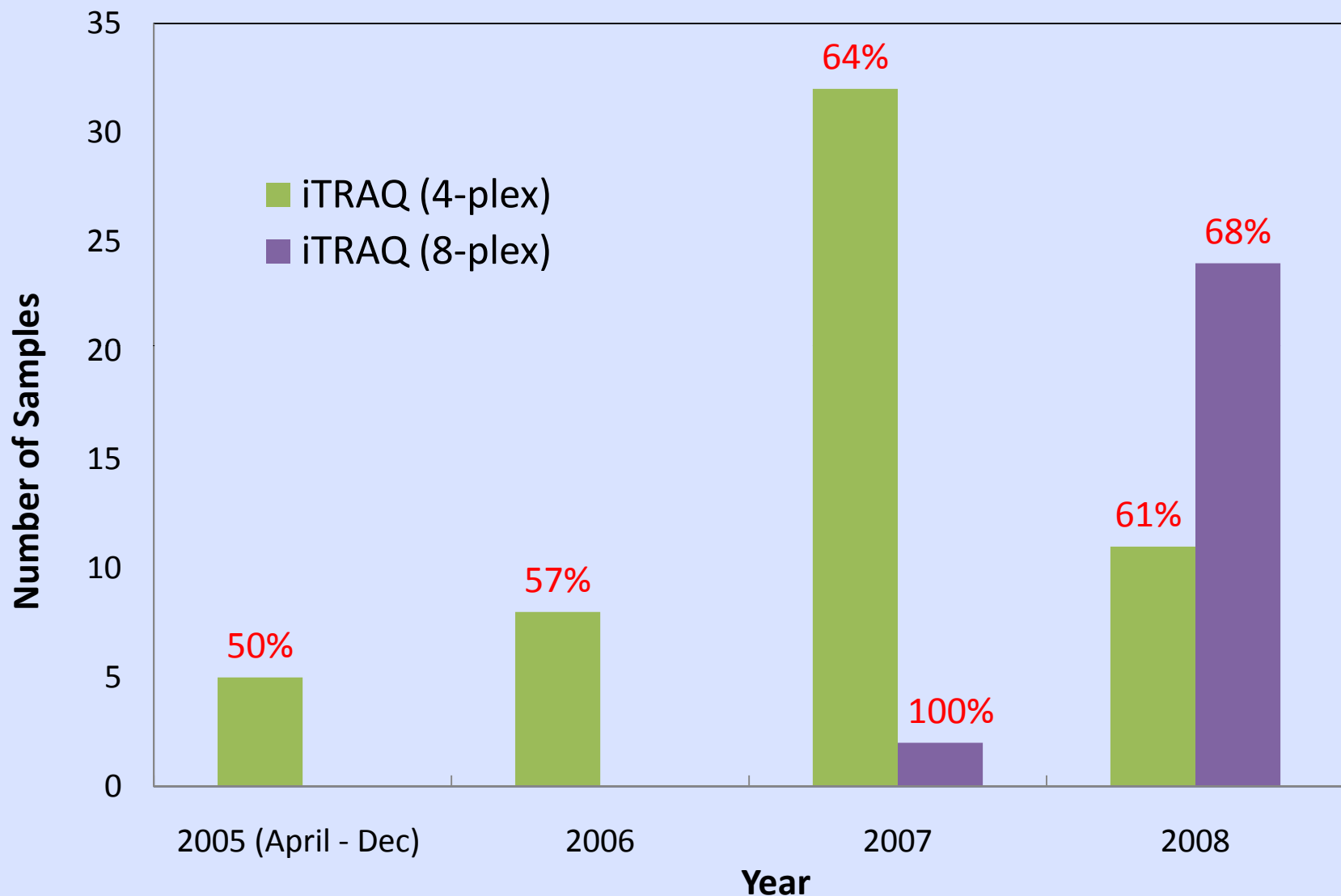
Small Molecule &
Med. Chem. MS



Bioanalytical
Quantitation
(small molecule)



Growth of NIDA iTRAQ Protein Profiling Service

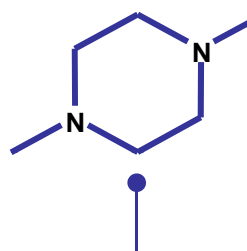


Red Values = Percentage (%) of NIDA samples/Total samples run during calendar year

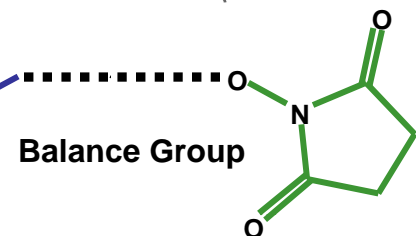
iTRAQ[®] Reagents - 8plex

Isobaric Tag: Total mass = 305

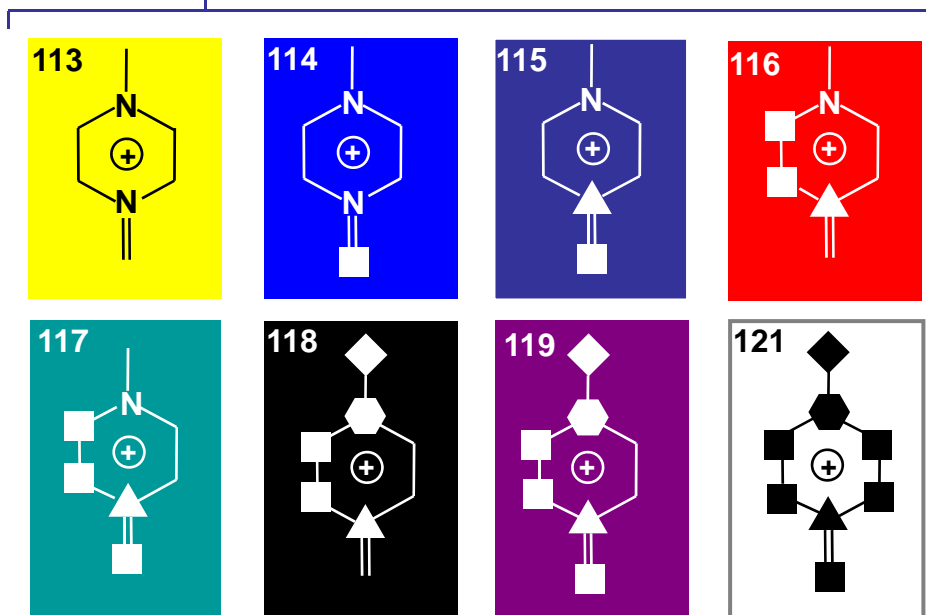
Reporter Group



Balance Group

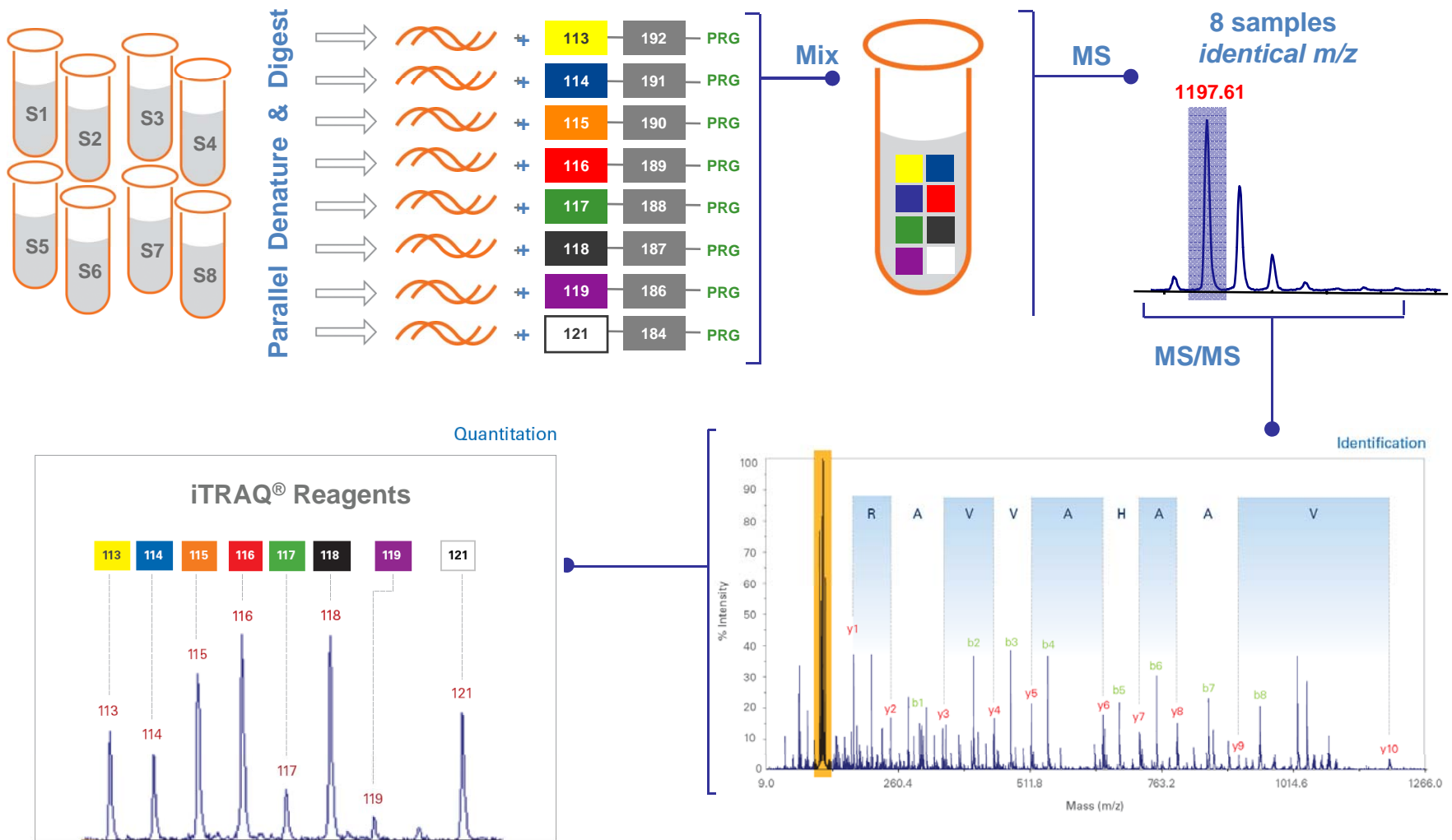


Peptide Reactive Group (PRG)



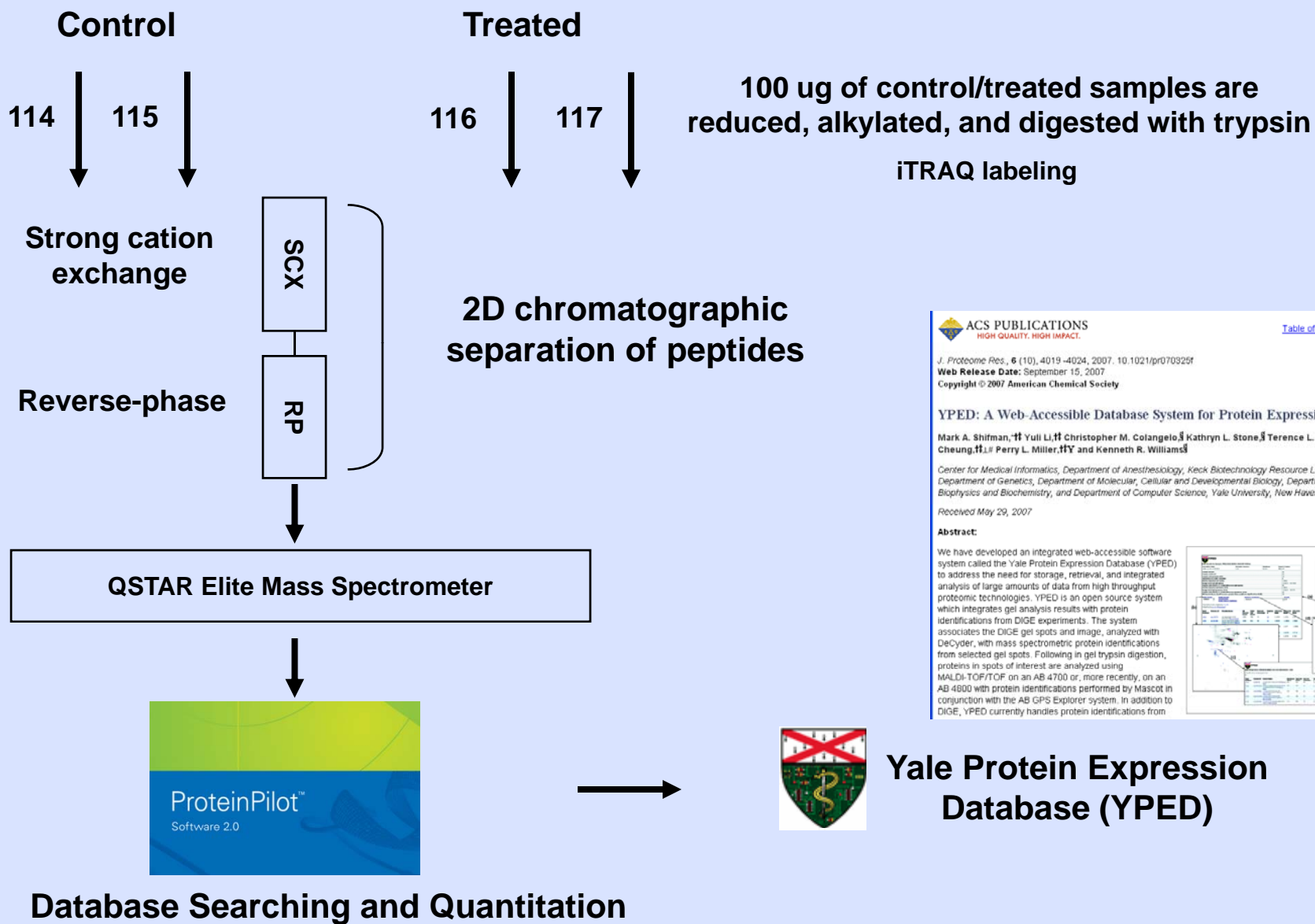
- ^{13}C (Square)
- ^{15}N (Triangle)
- ^{13}C (Diamond)
- ^{15}N (Hexagon)

iTRAQ® Reagents Workflow



[Figure used by permission from ABI]

Workflow for iTRAQ Protein Profiling Analysis



ACS PUBLICATIONS
HIGH QUALITY. HIGH IMPACT.

[Table of Contents](#) | [Journal Ho](#)

J. Proteome Res., 6 (10), 4019–4024, 2007. 10.1021/pr070325f
Web Release Date: September 15, 2007
Copyright © 2007 American Chemical Society

YPED: A Web-Accessible Database System for Protein Expression Analysis


Mark A. Shifman,†† Yuli Li,†† Christopher M. Colangelo,§ Kathryn L. Stone,§ Terence L. Wu,§ Kei-Hoi Cheung,†† Perry L. Miller,††† and Kenneth R. Williams§

Center for Medical Informatics, Department of Anesthesiology, Keck Biotechnology Resource Laboratory, Department of Genetics, Department of Molecular, Cellular and Developmental Biology, Department of Molecular Biophysics and Biochemistry, and Department of Computer Science, Yale University, New Haven, Connecticut

Received May 29, 2007

Abstract:

We have developed an integrated web-accessible software system called the Yale Protein Expression Database (YPED) to address the need for storage, retrieval, and integrated analysis of large amounts of data from high throughput proteomic technologies. YPED is an open source system which integrates gel analysis results with protein identifications from DIGE experiments. The system associates the DIGE gel spots and image, analyzed with DeCyder, with mass spectrometric protein identifications from selected gel spots. Following in gel trypsin digestion, proteins in spots of interest are analyzed using MALDI-TOF/TOF on an AB 4700 or, more recently, on an AB 4800 with protein identifications performed by Mascot in conjunction with the AB GPS Explorer system. In addition to DIGE, YPED currently handles protein identifications from





Yale Protein Expression Database (YPED): Key Features



Web-based Client Built on Open-Source Platform
Web-based viewing of results



Data and Technology Integration

Data Upload from multiple technologies, integration of the results, enables data review



Data Archive and Analysis

Intuitive Web interface for users to query, analyze and visualize data

Allows comparison of results across samples



Public Repository and Data Preservation (Sept 2008)

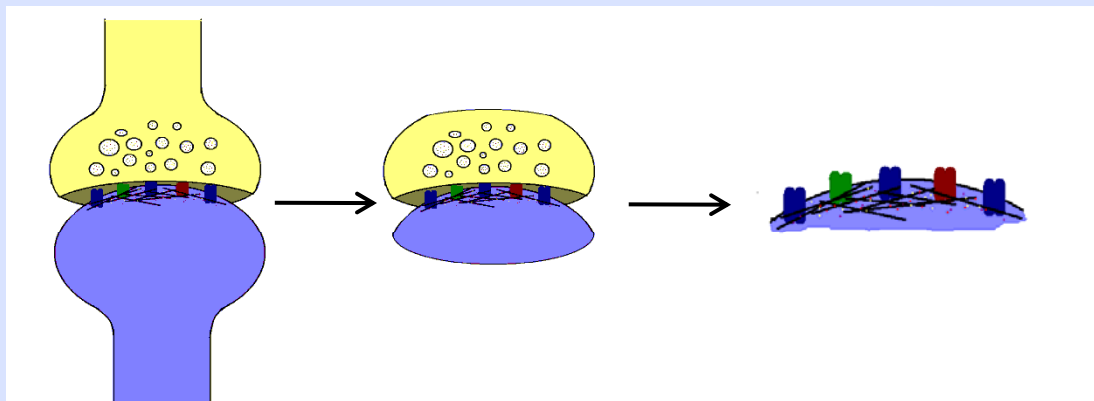
Stores and Exports data in industry standard format (MIAPE)

Chosen by Yale Digital Landscape as the best Yale scientific database for development and beta testing

“YPED: a web-accessible database system for protein expression analysis.”
Shifman, M.A., Li, Y., Colangelo, C.M., Stone, K.L., Wu, T.L., Cheung, K.H., Miller, P.L., Williams, K.R. *J Proteome Research* 2007, 6(10), 4019-4024.

Postsynaptic Density (PSD)

- Multi-protein complex organized into functional assemblies
- Site of synaptic plasticity
- Receptor complexes localized by scaffold proteins, which links to signaling molecules
- Proteomic studies of the rodent PSD reveals diverse classes of proteins (Li *et al.* 2004, Jordan *et al.* 2004, Yoshimura *et al.* 2002, Collins *et al.* 2005)
- Large-scale changes to the PSD after drug exposure is poorly studied



iTRAQ Workflow for Rat Brain PSD samples

Saline-treated

Cortex Striatum Hippocampus

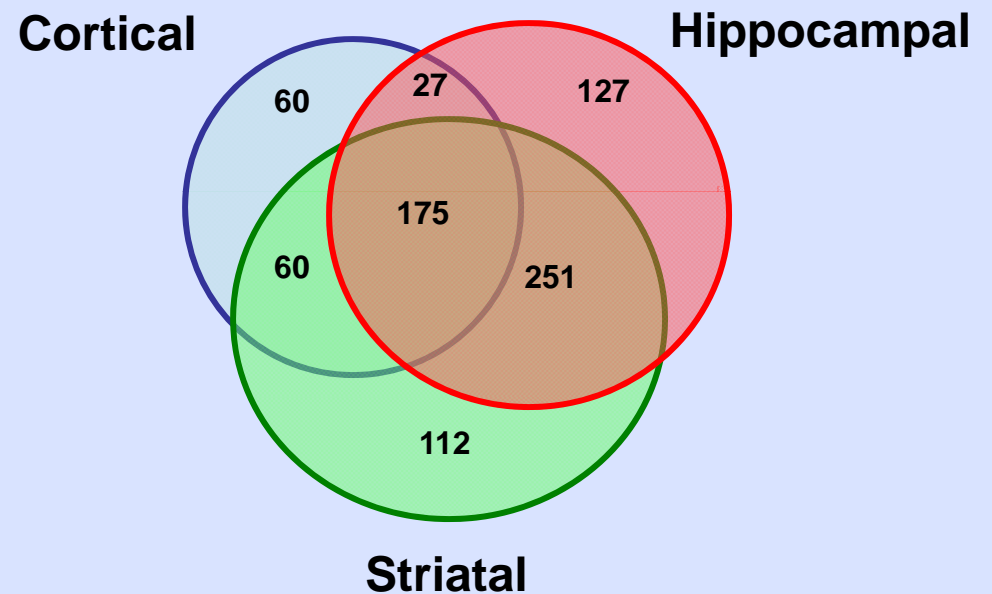
Exposed: 30 mg/kg cocaine

Cortex Striatum Hippocampus

iTRAQ Analysis

*i*Traq Summary of PSD analysis

- Cortical PSD:
 - 323 proteins identified
 - 29 proteins ↓, 63 proteins ↑
- Striatal PSD:
 - 560 proteins identified
 - 57 proteins ↓, 67 proteins ↑
- Hippocampal PSD:
 - 580 proteins identified
 - 33 proteins ↓, 45 proteins ↑



YPED Data Results 8-plex iTRAQ on Cortical, Striatal, and Hippocampal PSD from naïve rats

ITRAQ8plex Results

YPED

ITRAQ8plex Results for Sam

Execution Date
20 Sep 2007 13:49

Summary Statistics Protein

of proteins with 1 peptide
of proteins with >= 99% c
of proteins with >= 95% c
of proteins with >= 90% c

Range of 114:113 ratios
Range of 115:113 ratios
Range of 116:113 ratios
Range of 117:113 ratios
Range of 118:113 ratios
Range of 119:113 ratios
Range of 121:113 ratios

Ratio

Observed Bias Correction

[View ITRAQ Sample Inform](#)
[View PCA Image](#)

728 proteins found, displayi
[First/Prev] 1, 2, 3, 4, 5, 6, 7

| Protein Score | Protein ID | Prot |
|---------------|-----------------------------|--------------------------------------|
| 141.53 | IP100798592 | Gen spe indis |
| 89.9 | IP100319830 | Gen Isof chail |
| 71.71 | IP100752412 | Gen ATP trans poly indis |
| 71.01 | IP100649886 | Gen Isof |
| 64.37 | IP100169463 | Gen Tub |

Done

View Sample Requisition - Mozilla Firefox

https://yped.med.yale.edu:8443/yp_results/ViewSampleITRAQ8plex.do?sample_id=4133

ITRAQ8plex Results **View Sample Requisition**

YPED

View Sample Requisition PI: An

User name: Erika Andrade A

Experiment Type: * Other

Sample Name * Cortex, Striatum, Hippoca

Growth Conditions
Mouse

Sample Summary * 8-plex iTRAQ on l
Density Preparati

Label * ITRAQ8p_113

Sample Name CTX

Sample Type * normal/control

Organism * Mus musculus

Protein mg/ml * 3.9765

[Info](#)

AAA requested

Volume ul * 25.15

Total Sample ug * 397.65

[Info](#)

Sample labeled ug 50

Tissue Brain

Cell Type

Cell Line

Developmental Stage

Sex

Sample Description Cortex PSD

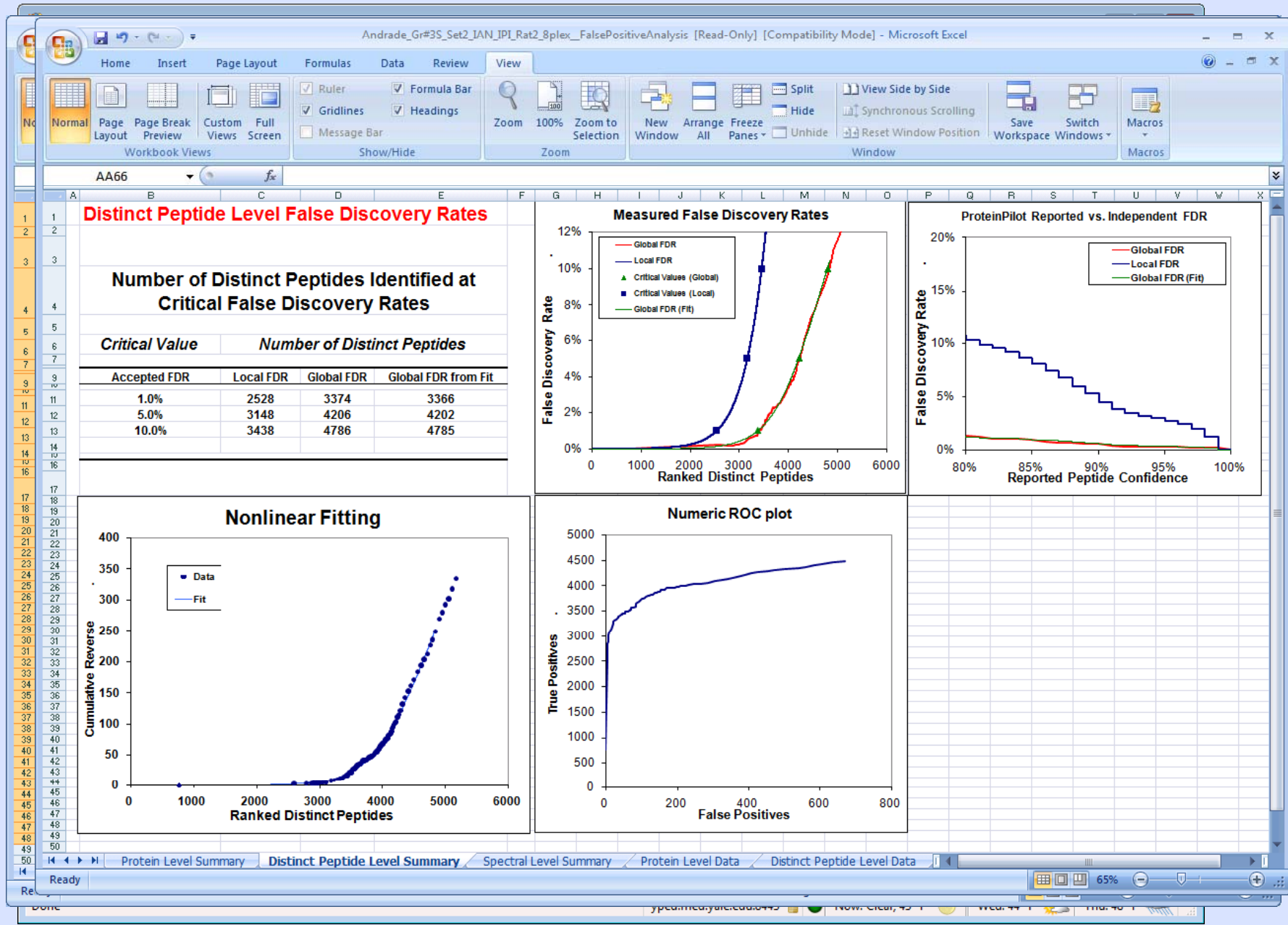
For questions or comments contact [Marl](#)

Done

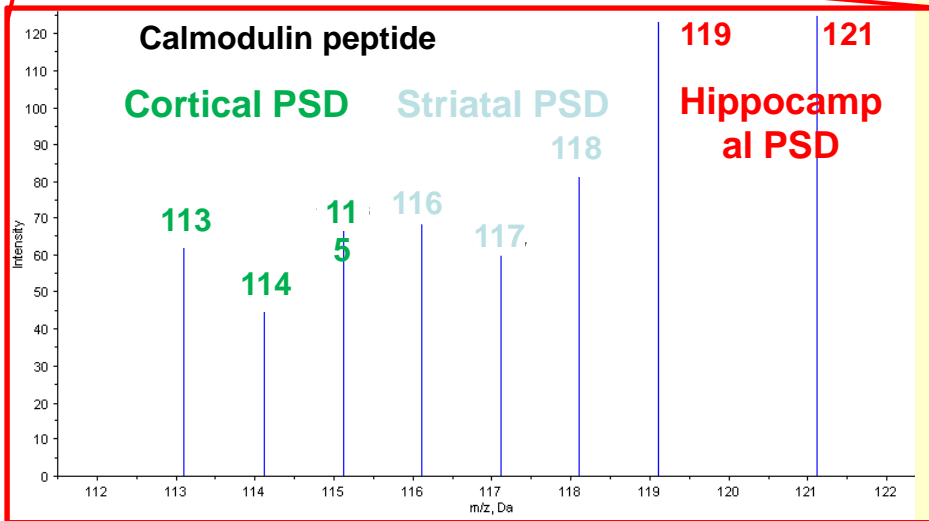
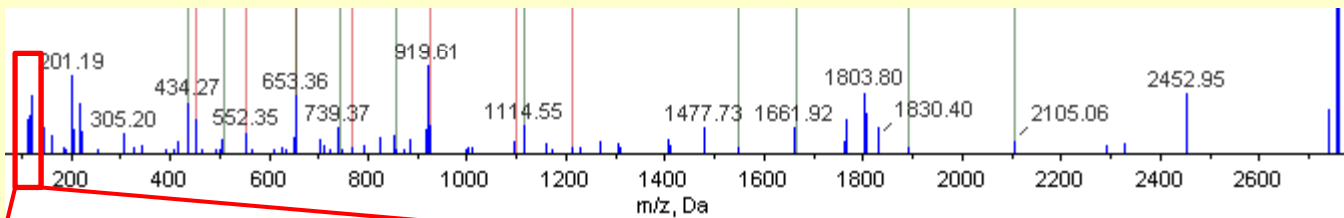
Principal Component Analysis on 8-plex iTRAQ dataset from Mouse Brain PSD from various brain regions.

| ITRAQ8p_121 |
|--------------------------|
| HIPP |
| normal/control |
| Mus musculus |
| 3.744 |
| <input type="checkbox"/> |
| 26.7 |
| 374.4 |
| 50 |
| Brain |
| <input type="checkbox"/> |
| Hippocampus PSD |

iTRAQ ProteinPilot False Discovery Rate

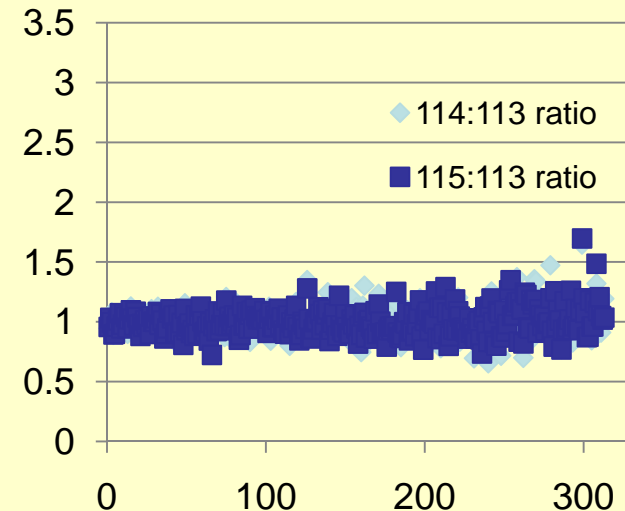


8-plex iTRAQ on Cortical, Striatal, and Hippocampal PSD from naïve rats



Additional calmodulin peptides were reviewed to substantiate this expression change.

Scatter plot for the cortical PSD replicates



Ratios shows that the expected deviation is $\pm 20\%$

Upgraded Protein Identification and Quantitation Software

- **Upgraded to MASCOT Cluster version 2.2**
 - **Offers improved Phosphopeptide Identification**
 - **Migrated MASCOT to an duo quadcore LINUX cluster**
 - **This allows processing on 8 nodes, which was needed given the additional data generated from the QSTAR Elite, LTQ-Orbitrap, and 4000 QTRAP mass spectrometers**
- **ProteinPilot 2.0.1**
 - **Protein Identification and Quantitation software**
 - **Enables simultaneous identification and quantitation for ICAT™, iTRAQ™, and SILAC™ reagents**
 - **Performs Protein Grouping and Calculates False Discovery Rates**
- **Mascot Distiller – Quantitation Toolbox (SILAC)**
 - **We were a beta test site for the MASCOT Distiller Quantitation Toolbox and helped push development for SILAC based quantitation**

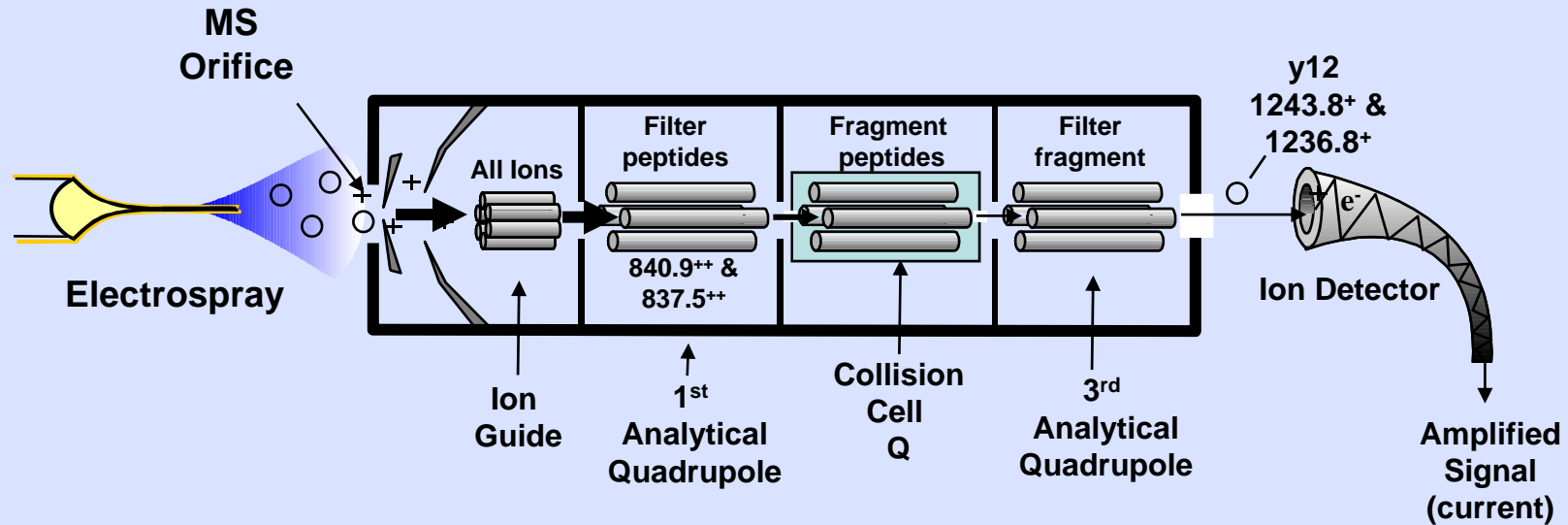
Targeted Proteomics Platform



**ABI 4000 QTRAP and
Waters NanoAcquity UPLC**

- Utilizes Multiple Reaction Monitoring (MRM)
- Uses
 - Utilizes information from MS based discovery
 - Rapid develop of an assay to monitor protein dynamics (eg. Expression changes or PTMs in experimental systems)
- Advantages
 - High dynamic ranges (10^5)
 - Hundreds of samples
 - Quantitation – Relative or Absolute (heavy idiotypic peptides)

Schematic of the ESI Triple Quadrupole MS System in MRM Operation Mode



Select Peptide



Fragment peptide



Select Fragment



Detect Fragment



- Highest specificity and sensitivity for detecting components in a complex mixture
- Largest linear dynamic range for quantitation
- Requires triple quadrupole MS capability
- Well accepted as the MS technique for quantitation (Pharmaceutical Industry)

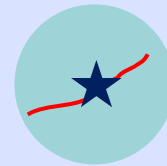
Targeted Proteomics Workflow



Protein mixture



Enzymatic Digest



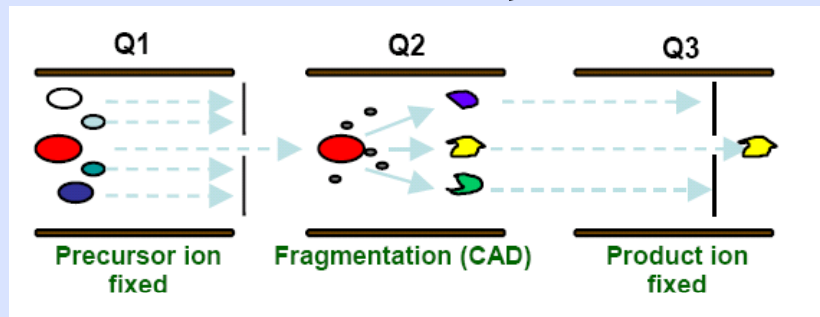
Heavy idiotype peptide



Peptide synthesis
HPLC purification
Amino acid analysis

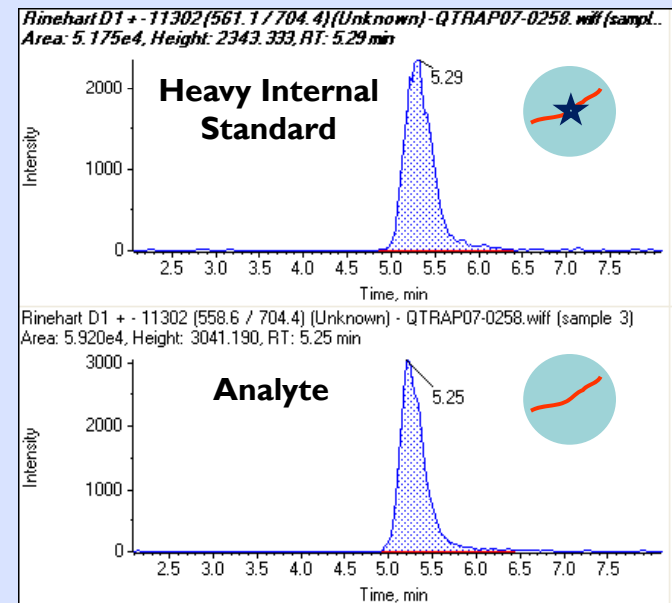
Mix

MRM Analysis




Schematic of Multiple Reaction Monitoring Scan (MRM)

LC-MRM Chromatogram



YPED Synthetic Peptide Database




Synthetic Peptide Search

6 results found, displaying all results.

| Synthetic Peptide ID | Sequence |
|--------------------------|------------------------------|
| WIL11612 | GL(13C6)(15N1)SS(P |
| WIL11613 | L(13C6)(15N1)ES(PO |
| WIL11880 | ELLSNVDQDVHEL(15N1)EK |
| WIL11614 | EAQEL(13C6)(15N1)GS(PO4)PEDR |
| WIL11615 | QIL(13C6)(15N1)SQSTDS(PO4) |
| WIL11616 | S(PO4)L(13C6)(15N1)PAGDALYLS |

Export options: [CSV](#) | [Excel](#)

For questions or comments contact [Mark Shifman](#)
Updated 02 Dec 2008 14:48



Edit Synthetic Peptide

| Synthetic Peptide ID | Sequence | Length | HPLC RT | XLS |
|----------------------|---------------------------|------------------|---------|---------------------------|
| WIL11613 | L(13C6)(15N1)ES(PO4)FESLR | 8 | 16.5 | 11613.xls |
| M+H mono | M+2H mono | M+3H mono | | |
| 1067.522489 | 534.265157 | 356.512713 | | |

| | MRM Transitions | | | | |
|----------------------------------|-----------------|----|----|----|-----|
| | Q1 | Q3 | CE | DP | Ion |
| MS RT | 21.8 | | | | |
| AAA conc | 3.5 | | | | |
| Protein ID | NP_001012079 | | | | |
| Protein Name | Rho GEF2 LFC | | | | |
| Opt Internal Std Conc(fm) | 250 | | | | |
| Organism | | | | | |
| Comment | | | | | |

For questions or comments contact [Mark Shifman](#)
Updated 02 Dec 2008 14:48

Done

Done

yped.med.yale.edu:8443

Now: Clear, 40° F

Wed: 44° F

Thu: 48° F

yped.med.yale.edu:8443

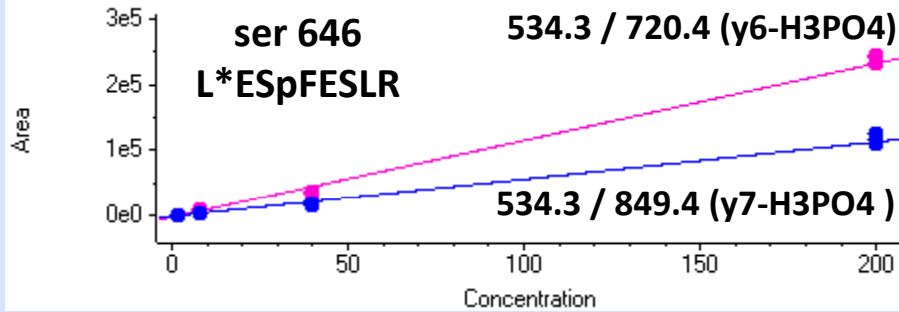
Now: Clear, 40° F

Wed: 44° F

Thu: 48° F

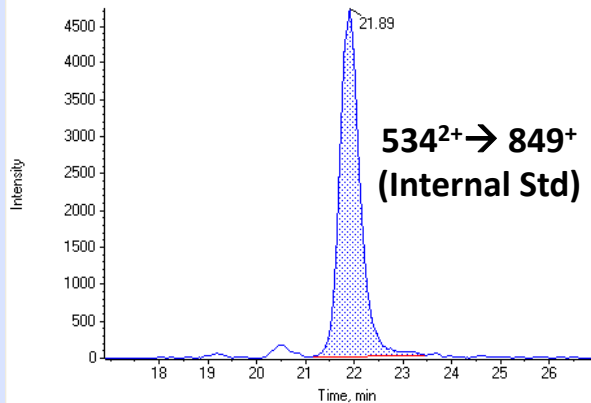
MRM targeted Proteomics of Rho GEF LFC Ser646 (WIL11613)

● Calibration for WIL11613 (534.3 / 849.4): $y = 570.05093x + -1199.33455$ ($r = 0.99432$)
 ● Calibration for WIL11613 (534.3 / 720.4): $y = 1173.00769x + -2372.37216$ ($r = 0.99604$)

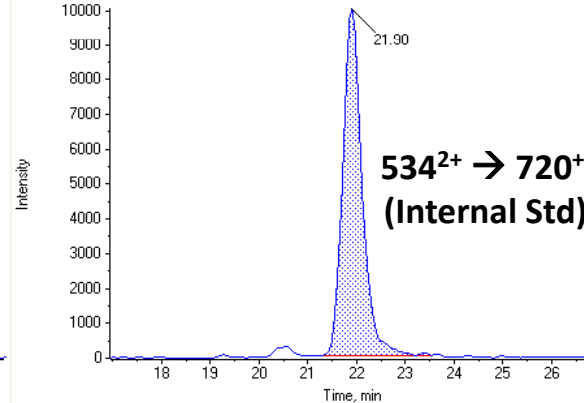


Heavy internal standard peptide calibration curve for the 534.3/849.4 and 534.3/720.4 MRM transitions of ser 646 peptide L*ESpFESLR (L* = Heavy Leucine Amino Acid ($^{13}\text{C}_6$, ^{15}N), Sp = phosphoserine)

Stipanovich 1 cntl enriched, 40 fmole/ul heavy, 1 - WIL11613 (534.3 / 849.4) (Unknown)...
 Area: 1.378e5, Height: 4719.710, RT: 21.89 min

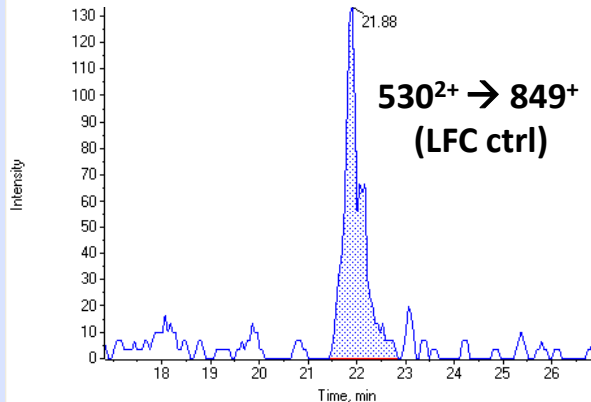


Stipanovich 1 cntl enriched, 40 fmole/ul heavy, 1 - WIL11613 (534.3 / 720.4) (Unknown)...
 Area: 2.761e5, Height: 9999.773, RT: 21.90 min

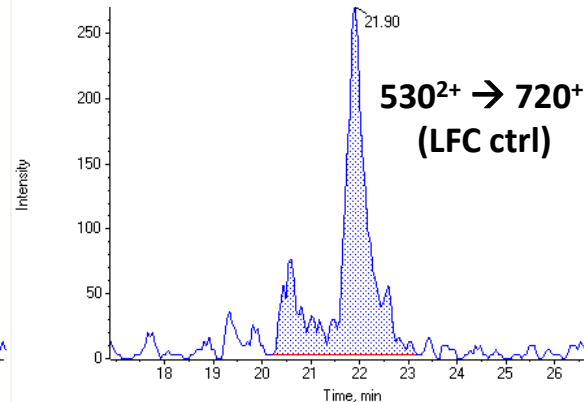


Heavy Internal Standard
 L*ESpFESLR

Stipanovich 1 cntl enriched, 40 fmole/ul heavy, 1 - WIL11613 (530.7 / 849.4) (Unknown)...
 Area: 3246.240, Height: 133.333, RT: 21.88 min



Stipanovich 1 cntl enriched, 40 fmole/ul heavy, 1 - WIL11613 (530.7 / 720.4) (Unknown)...
 Area: 10138.007, Height: 266.667, RT: 21.90 min



LFC control mouse brain sample

ser 646
 L^{*}ESpFESLR

Integration of MultiQuant Results in YPED

The screenshot displays the YPED web application interface, showing the integration of MultiQuant results. The interface is divided into three main panels:

- Left Panel (SuperUser Functions):** Shows a list of projects. The project "TargetedProteomics 062708-LFC" is highlighted with a red box.
- Middle Panel (View Targeted Proteomics):** Shows a list of samples. The sample "Stipanovich D1 enriched" is highlighted with a red box.
- Right Panel (Targeted Proteomics Results):** Displays a table of results for the selected sample. The table includes columns for Protein Name, Sequence, Component, Calc Conc, Area Ratio, Corrected Area, Quality, IS Quality, and Signal Noise.

Targeted Proteomics Results Table:

| Protein Name | Sequence | Component | Calc Conc | Area Ratio | Corrected Area | Quality | IS Quality | Signal Noise |
|--------------|--------------------------|------------------------------------|-----------|------------|----------------|---------|------------|--------------|
| | GLSS(PO4)LSLAK | WIL11612-LIGHT (478.3 / 418.3) | 17.135 | 0.0674 | 38662.37 | 0.5261 | 0.974 | 235.7797 |
| | GLSS(PO4)LSLAK | WIL11612-LIGHT (478.3 / 698.3) | 2.4293 | 0.0217 | 3061.3367 | 0.8942 | 0.9963 | 36.2655 |
| | GLSS(PO4)LSLAK | WIL11612-LIGHT (478.3 / 687.4) | -0.4163 | 0.0016 | 1283.8067 | 1 | 0.9852 | 24.2559 |
| | GLSS(PO4)LSLAK | WIL11612-LIGHT (478.3 / 600.4) | -0.3801 | 0.0016 | 148.1333 | | 0.9838 | 2.5409 |
| | LES(PO4)FESLR | WIL11613-LIGHT (530.8 / 504.3) | -0.2487 | 0.0116 | 3925.4133 | 0.5058 | 0.9774 | 24.3446 |
| | LES(PO4)FESLR | WIL11613-LIGHT (530.8 / 651.3) | 2.7274 | 0.0234 | 3382.305 | 0.988 | 0.9798 | 31.3997 |
| | LES(PO4)FESLR | WIL11613-LIGHT (530.8 / 849.4) | -0.7578 | 0.003 | 938.1633 | | 0.9858 | 19.3226 |
| | LES(PO4)FESLR | WIL11613-LIGHT (530.8 / 720.4) | 1.2241 | 0.0128 | 6814.0433 | 1 | 0.9853 | 84.6551 |
| | LES(PO4)FESLR | WIL11613-LIGHT (530.8 / 504.3 (2)) | -0.2042 | 0.0127 | 4279.295 | 0.6303 | 0.9774 | 26.4839 |
| | EAQELGS(PO4)PEDR | WIL11614-LIGHT (655.8 / 642.3) | 6.0253 | 0.0214 | 10961.705 | 0.9933 | 0.981 | 126.2977 |
| | EAQELGS(PO4)PEDR | WIL11614-LIGHT (655.8 / 853.3) | 6.5598 | 0.0102 | 345.6367 | | 1 | 9.6486 |
| | EAQELGS(PO4)PEDR | WIL11614-LIGHT (655.8 / 755.4) | 8.1802 | 0.027 | 4435.6933 | 1 | 0.9824 | 69.4719 |
| | EAQELGS(PO4)PEDR | WIL11614-LIGHT (655.8 / 740.3) | 9.1024 | 0.0369 | 2024.455 | 1 | 0.994 | 30.8298 |
| | QILSQSTDS(PO4)LNMR | WIL11615-LIGHT (794.9 / 932.3) | 10.4225 | 0.027 | 296.25 | 0.7659 | 0.9026 | 8.2763 |
| | QILSQSTDS(PO4)LNMR | WIL11615-LIGHT (794.9 / 1019.4) | 0.0262 | 0.0027 | 98.7517 | 0.2308 | 0.9736 | 2.5523 |
| | QILSQSTDS(PO4)LNMR | WIL11615-LIGHT (794.9 / 1234.5) | 3.3001 | 0.013 | 395.0133 | | 1 | 7.3327 |
| | QILSQSTDS(PO4)LNMR | WIL11615-LIGHT (794.9 / 831.3) | 0.7234 | 0.0048 | 691.275 | | 0.9798 | 7.8804 |
| | QILSQSTDS(PO4)LNMR | WIL11615-LIGHT (794.9 / 921.4) | 3.8708 | 0.0173 | 296.2617 | 1 | 0.9607 | 5.2416 |
| | QILSQSTDS(PO4)LNMR | WIL11615-LIGHT (794.9 / 834.4) | 7.5797 | 0.0214 | 148.1333 | | 0.92 | 2.5342 |
| | S(PO4)LPAGDALYLSFNPPQPSR | WIL11616-LIGHT (1055.5 / 1142.5) | | 0.3239 | 3481.0617 | 1 | 1 | 38.4795 |

Small Molecule Bioanalytical Chemistry

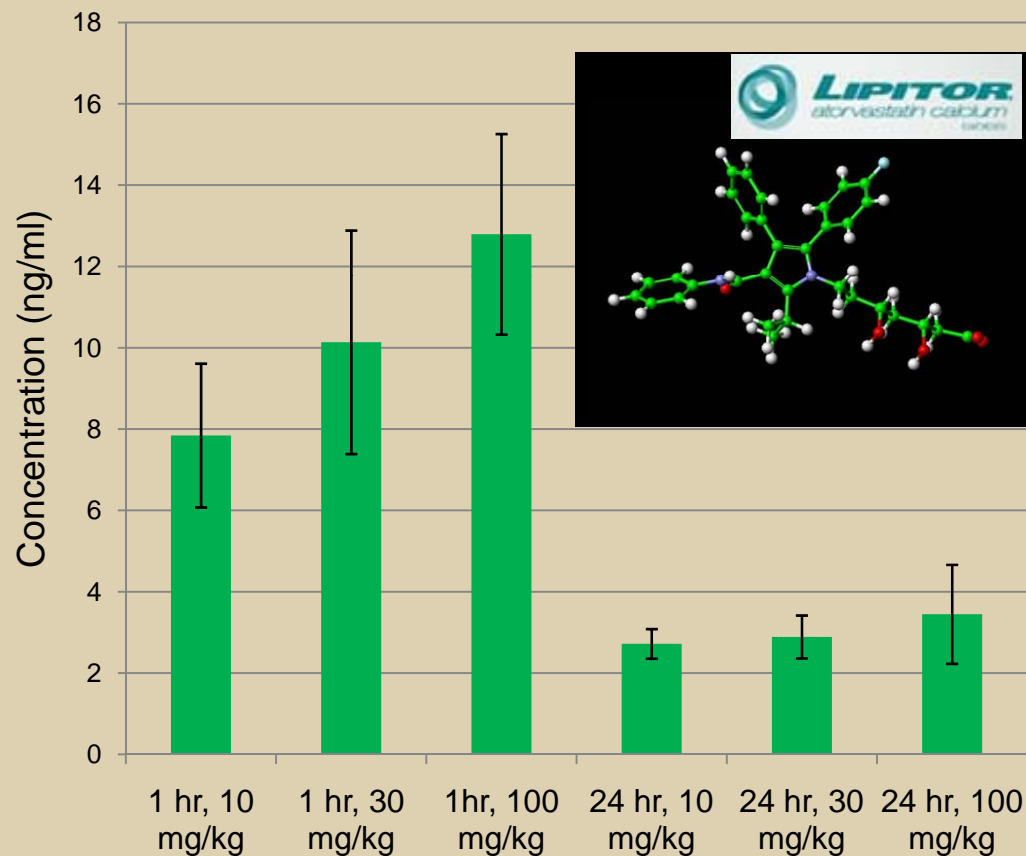


PE Series 275 HRes LC System



API 5000

Plasma levels of Lipitor in Mouse Plasma



Yi T, Rao DA, Tang PC, Wang Y, Cuchara LA, Bothwell AL, Colangelo CM, Tellides G, Pober JS, Lorber MI. *Transplantation*. 2008 Sep 15;86(5):719-27.

Acknowledgements

Ken Williams (Director)
Kathy Stone
Erol Gulcicek
TuKiet Lam
Terence Wu
Mary Lopresti
Tom Abbott
Jean Kanyo
Kathrin Wilczak-Havill
Edward Voss
Matt Berberich
Mark Shifman
Ji Young Lee
Hongyu Zhao
Erika Andrade
Alex Stipanovich

