



Pacific Northwest
NATIONAL LABORATORY

*Proudly Operated by **Battelle** Since 1965*

Enhanced Confidence and Quantitation of Proteoform Identifications

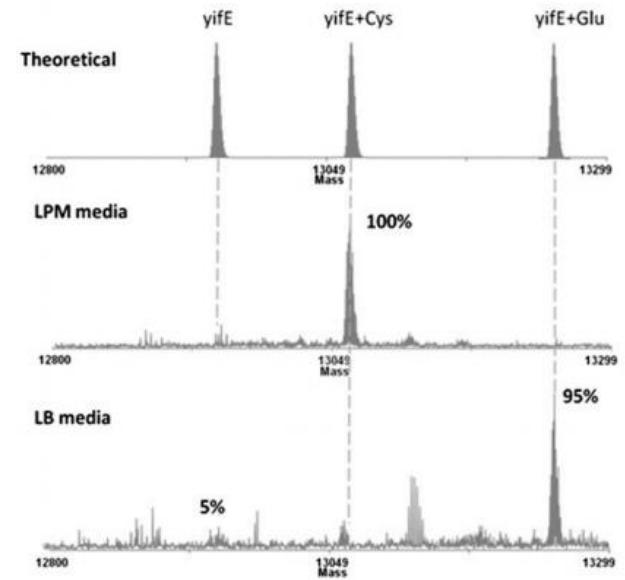
Samuel Payne

Pacific Northwest National Laboratory



Salmonella Experiment

- Proteoforms unique to infection like conditions
- Novel PTM switch



Top-down proteomics reveals a unique protein S-thiolation switch in *Salmonella* Typhimurium in response to infection-like conditions

Charles Ansong^{a,1,2}, Si Wu^{b,1}, Da Meng^c, Xiaowen Liu^{d,e}, Heather M. Brewer^b, Brooke L. Deatherage Kaiser^a, Ernesto S. Nakayasu^a, John R. Cort^a, Pavel Pevzner^f, Richard D. Smith^a, Fred Heffron^g, Joshua N. Adkins^{a,2}, and Ljiljana Paša-Tolić^{b,2}

^aBiological Sciences Division, ^bEnvironmental Molecular Sciences Laboratory, and ^cComputational Science and Mathematics Division, Pacific Northwest

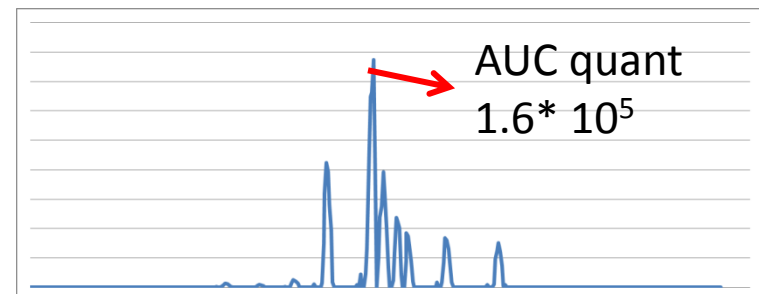
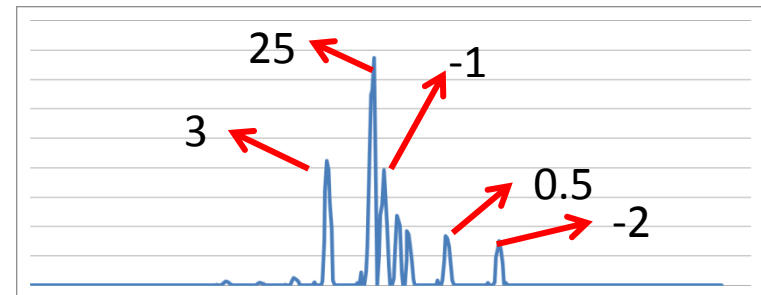
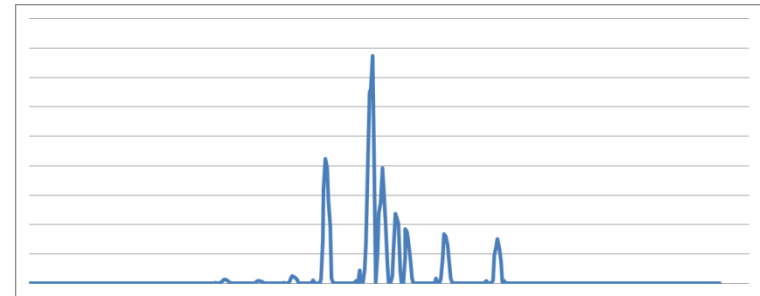
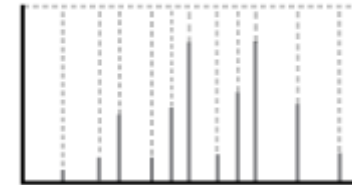
Quantitative Top Down

- Needs
 - Label Free quantitation
 - Automated and confident analysis
- Challenges
 - Difficulty de-isotoping – 30-40% error rate
 - Many poor identifications
 - Identifying all charge states

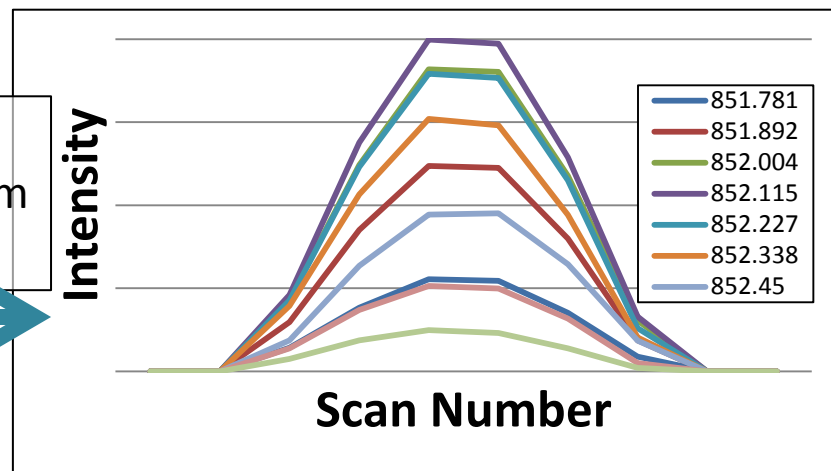
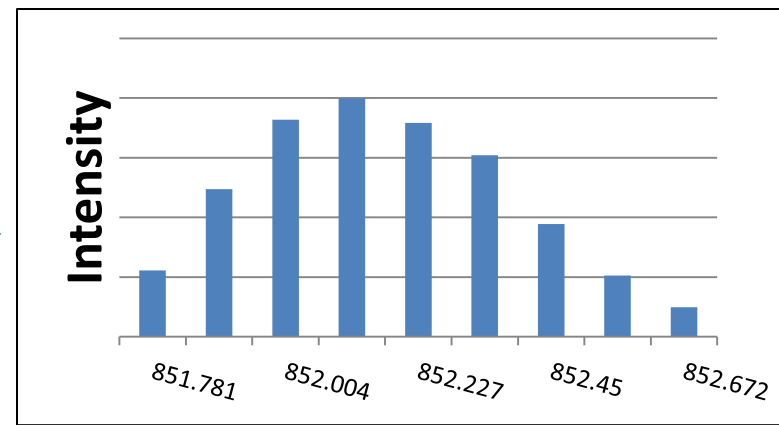
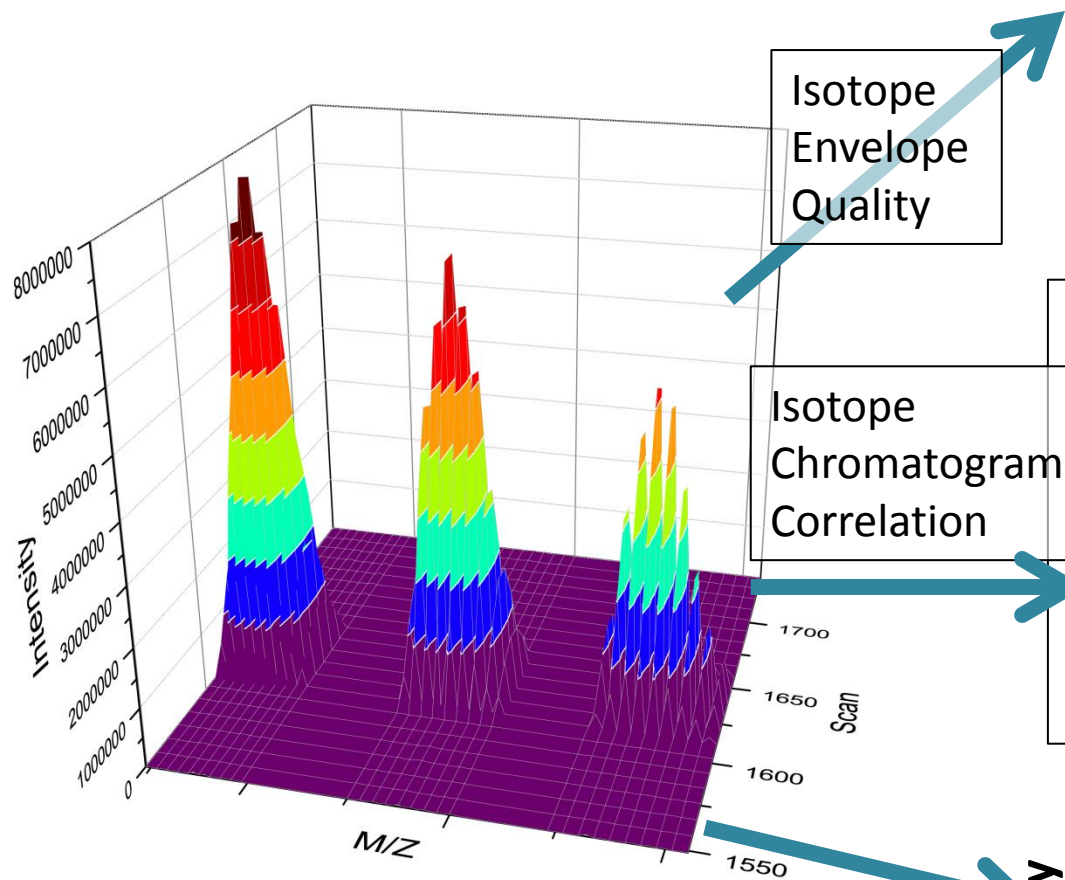
Informed Quantitation (IQ)

- Enumeration of targets
 - Results from PrSMs
 - Atomic composition
 - Elution time
 - Charge state(s) identified
- Retrieval
 - Generate extracted ion chromatogram
- Score
 - Isotope Envelop Quality
 - Chromatogram Correlation
- Quantify the best peak

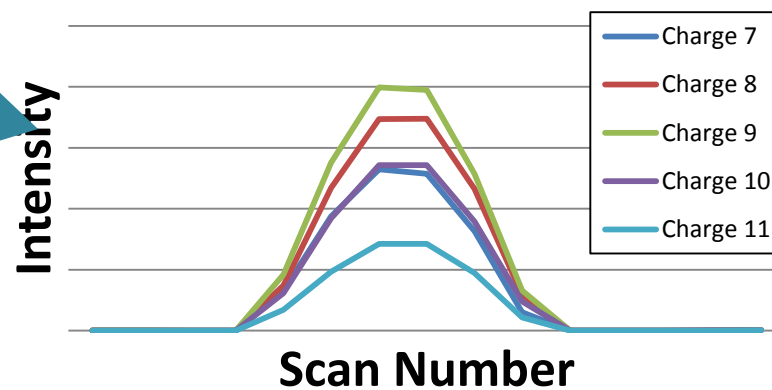
PEPTIDE ID



Retrieving Metrics

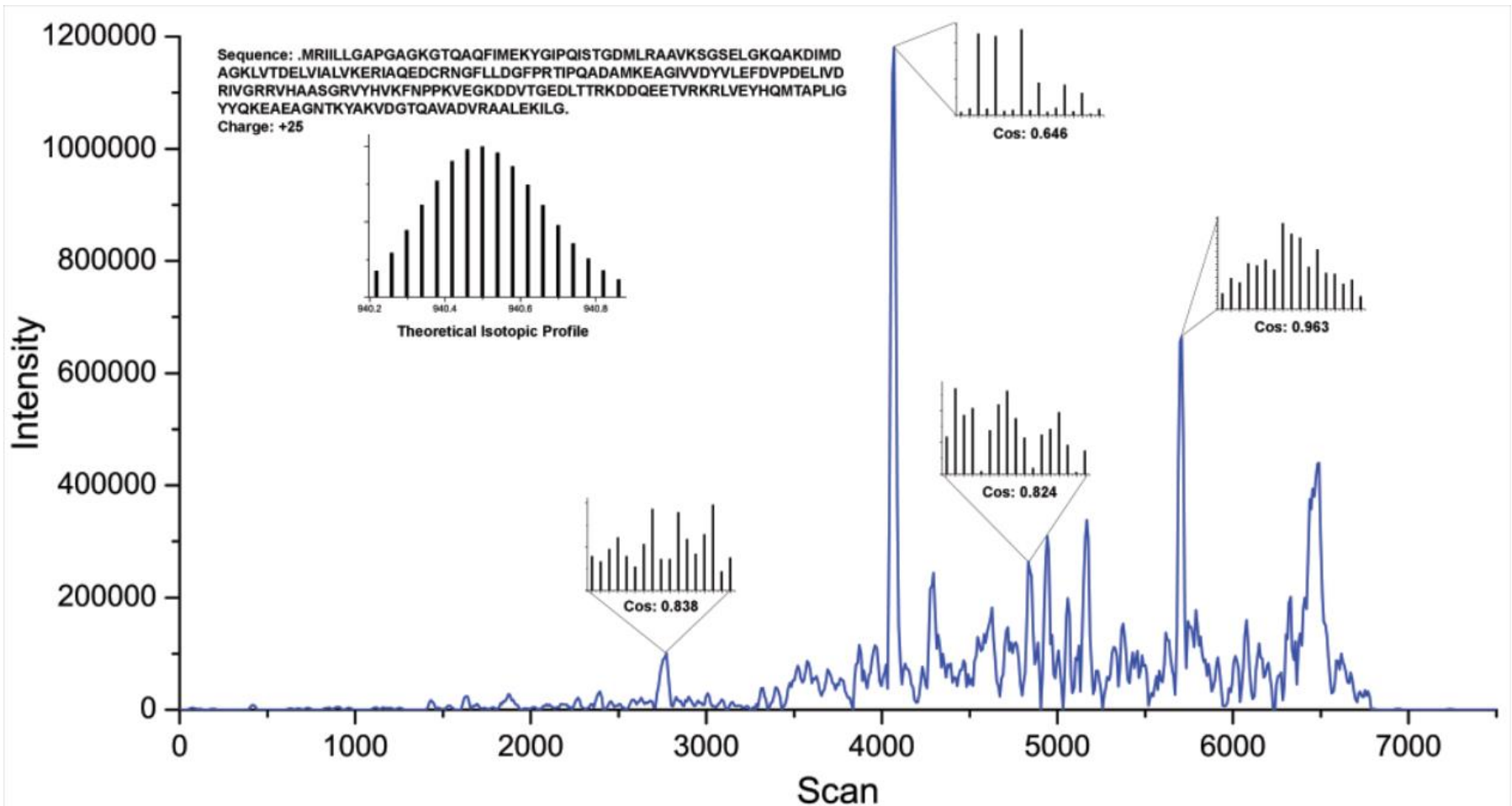


Charge State Chromatogram Correlation



Calculating Probability

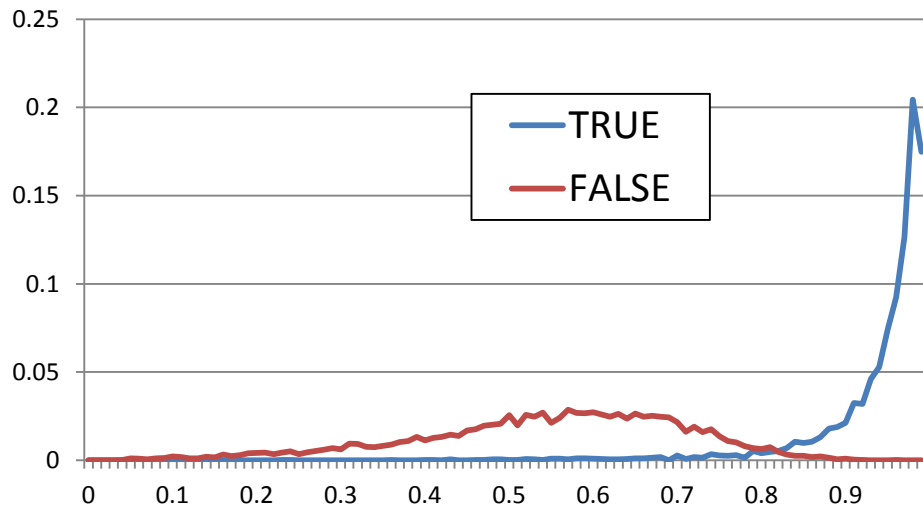
- Creating a decoy for MS1 features



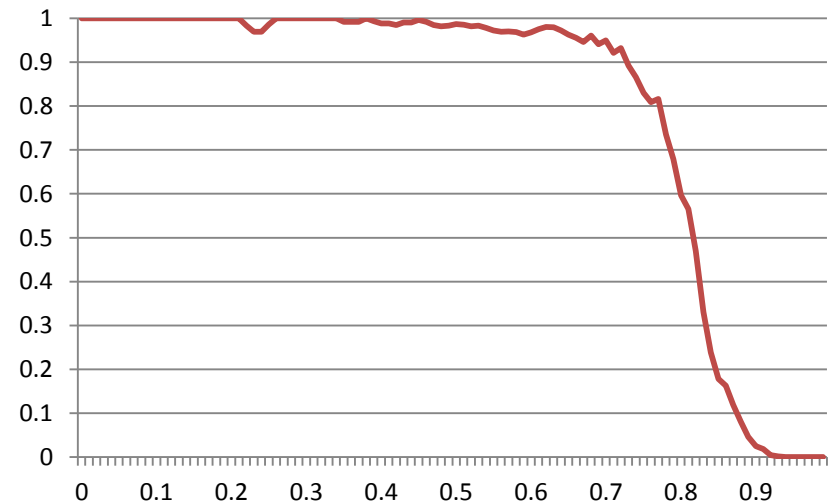
Calculating Probability

- Score metrics need an estimate of truth
 - i.e. how good is a cosine similarity of 0.91?

Cosine Similarity of Isotopic Envelope

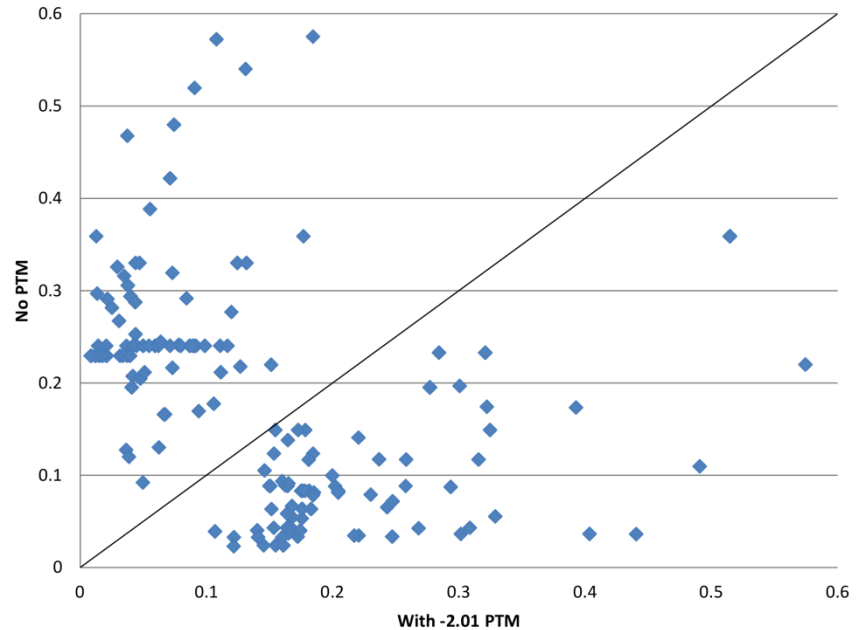
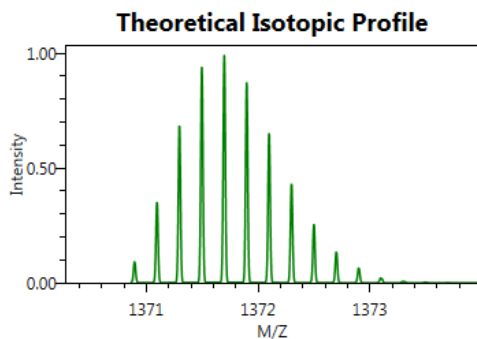
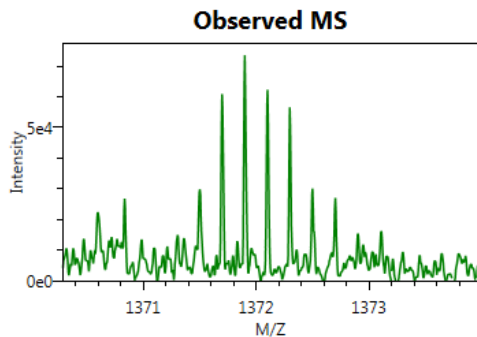


Posterior Error Probability



Validating TopDown IDs

- Many proteoforms contain phantom PTMs
 - Driven by errors in deisotoping (~30-40%)



High Throughput Operation

- Runtime: ~ 1 minute per LC-MS run
 - Abundance values for ~1000 proteoforms
- Most time is spent in correcting PrSM identifications

Acknowledgements

- NIH/NIGMS (P41 GM103493)



National Institute of
General Medical Sciences

- DOE  U.S. DEPARTMENT OF
ENERGY



Sangtae Kim
Grant Fujimoto
Kevin L. Crowell
Charles Ansong
Si Wu
Richard D. Smith
Ljiljana Pasa-Tolic
Joshua N. Adkins