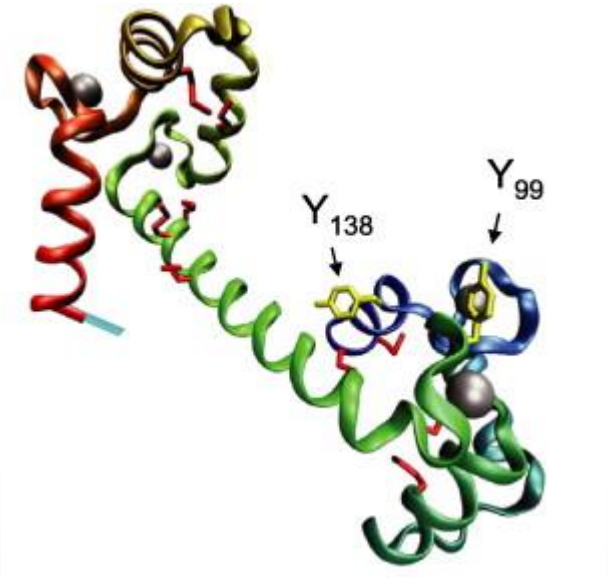
Proteoform Quantitation Through the IQ Framework

Grant Fujimoto, Sangtae Kim, Kevin Crowell, Charles Ansong, Nikola Tolic, Si Wu, Ljiljana Pasa-Tolic, Richard D Smith, Joshua N Adkins, Samuel H Payne

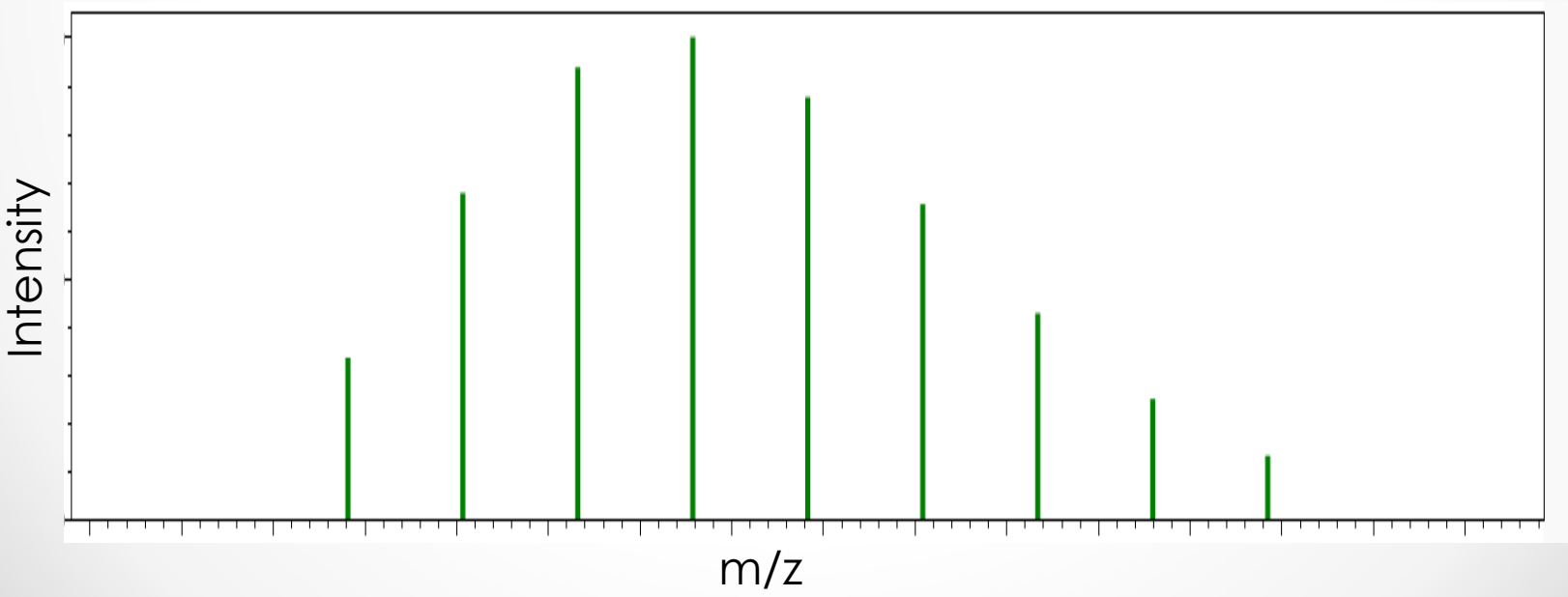
Pacific Northwest National Laboratory



Lourette et al., JASMS (2010)

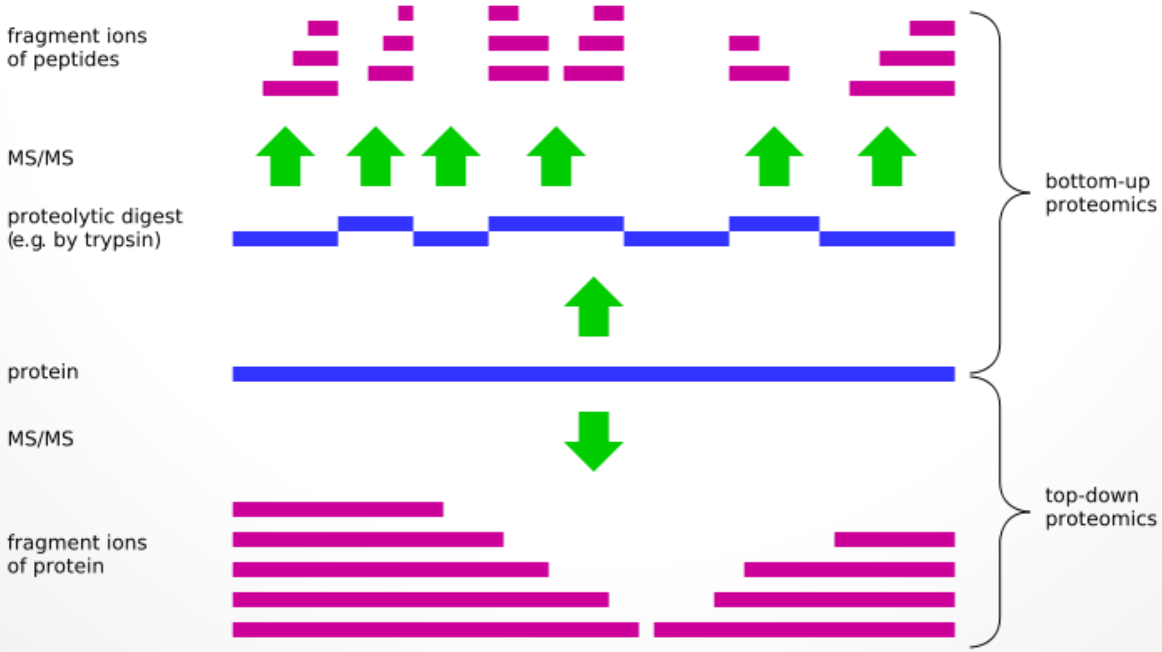
What Is IQ?

- Informed Quantitation
  - High-throughput quantitation of mass spectrometry data
- Utilize prior information to guide data processing
  - Proteoform atomic composition
  - Observed LC elution time



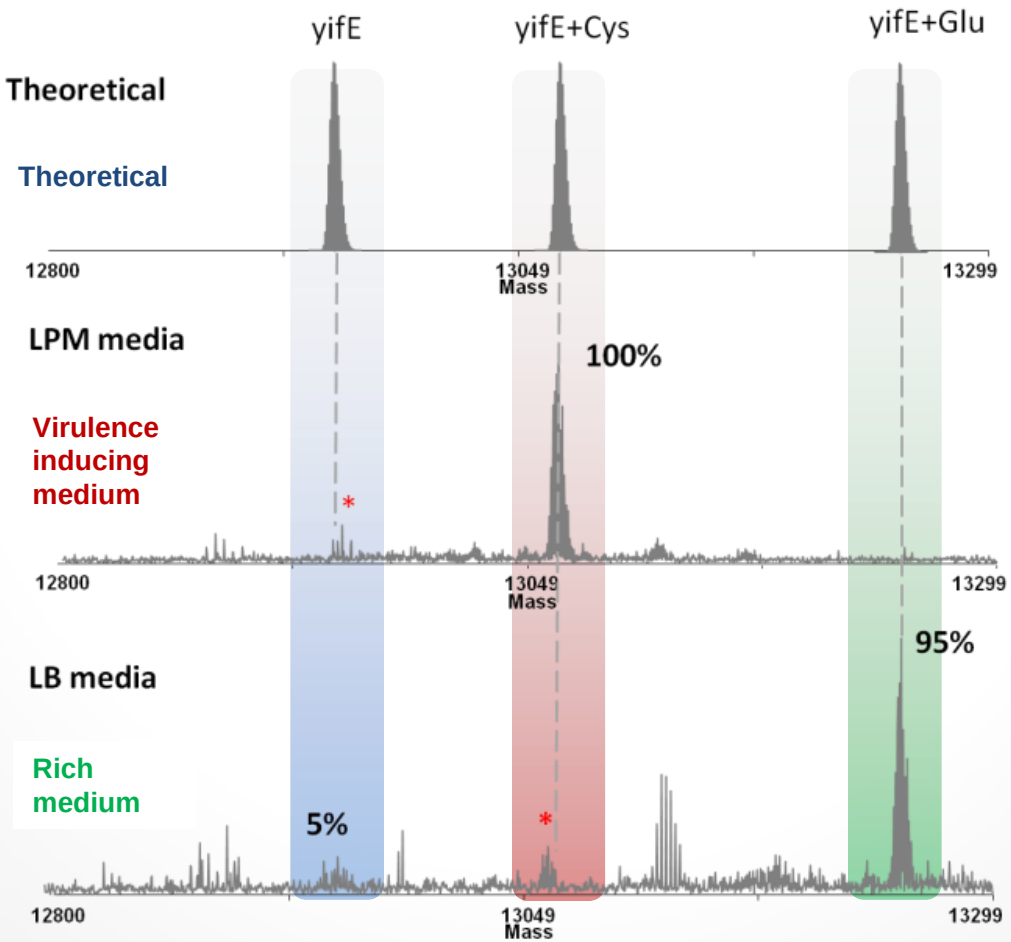
Why Quantify Proteoforms?

- Proteoform: specific molecular form of a protein including splicing and post-translational modifications
  - Top-Down: The analysis of proteoforms
  - Bottom-Up: Analysis of digested peptides
- Protein Inference complicates accurate proteoform quantitation



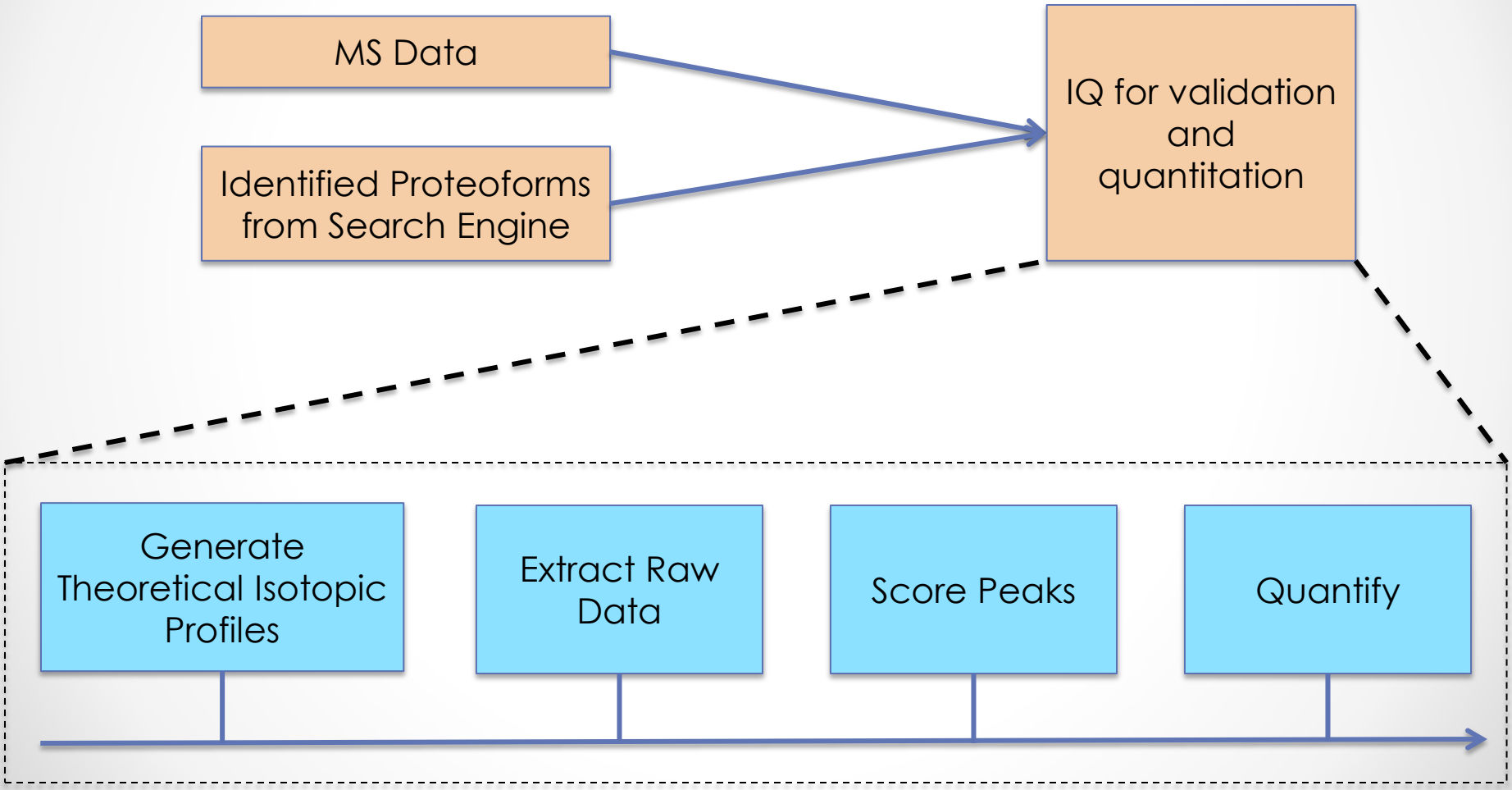
Top-Down Analysis of Salmonella

- PTM exclusive for infection-like condition
  - Cysteine glutathionylation in LB media
  - Cysteine cysteinylation in LPM media (infection like)



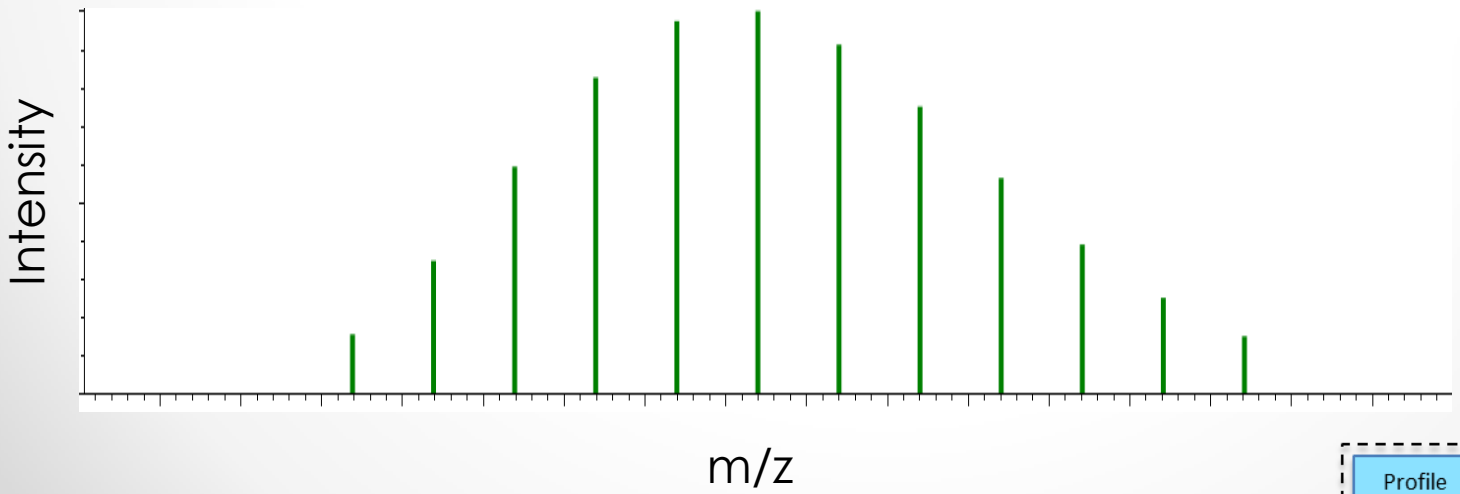
Ansong et al., *PNAS (2013)*

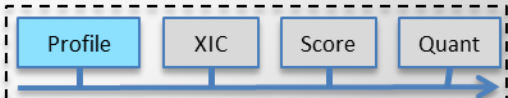
IQ Workflow


MS Data
Identified Proteoforms from Search Engine
IQ for validation and quantitation
Generate Theoretical Isotopic Profiles
Extract Raw Data
Score Peaks
Quantify

Generate Theoretical Isotopic Profiles

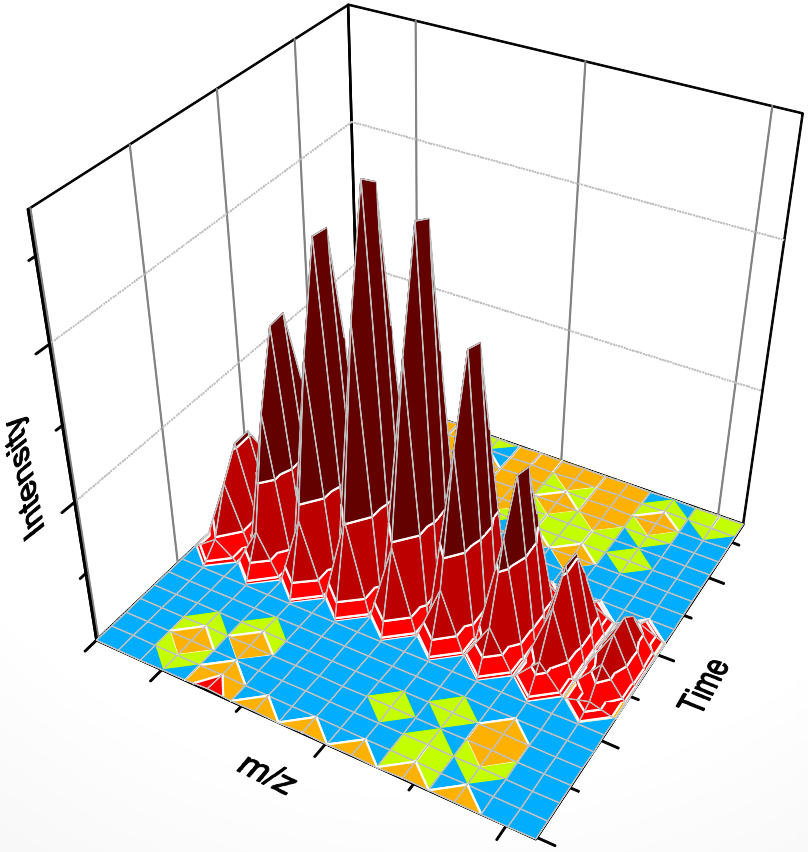
1. Get Identifications from PrSM (MS-Align+, ProSight)
2. Convert proteoforms sequence to chemical composition ($C_xH_xN_xO_xS_x$)
3. Calculate relative intensities and m/z values for isotopes in each charge state

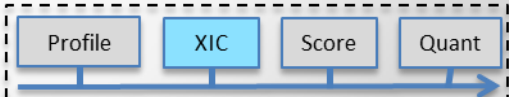




Extract Raw Data

For each m/z (all charges, all isotopes), we generate an extracted ion chromatogram

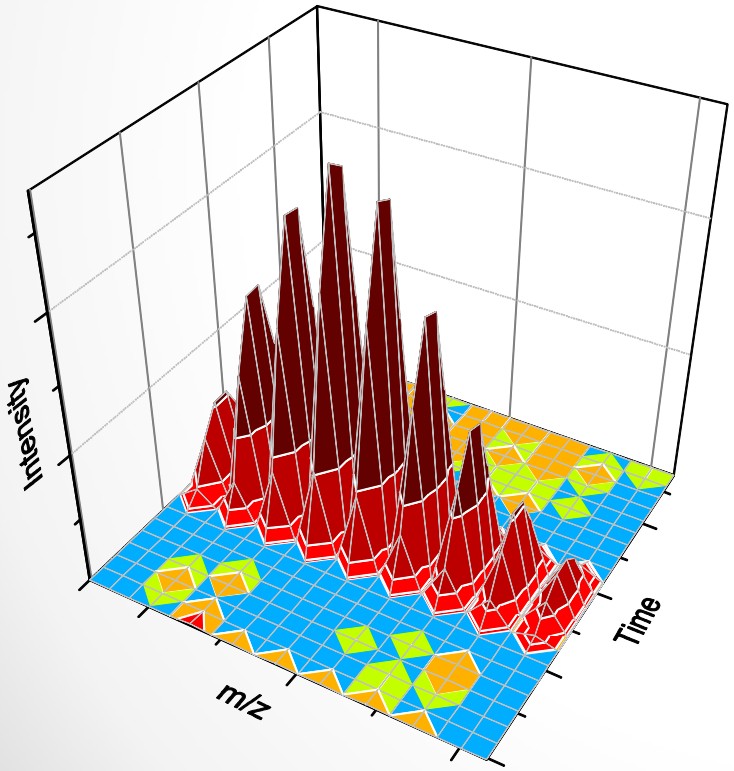


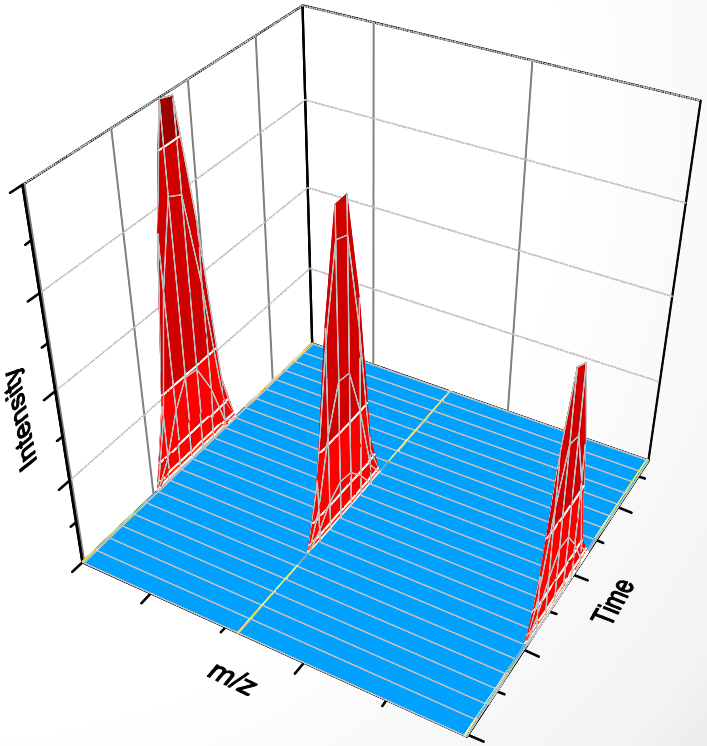


IQ Scoring Metrics

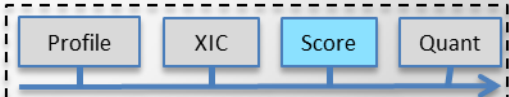
- Isotope Envelope Quality Score
- Isotopic Envelope Correlation
- Charge State Correlation

Charge View





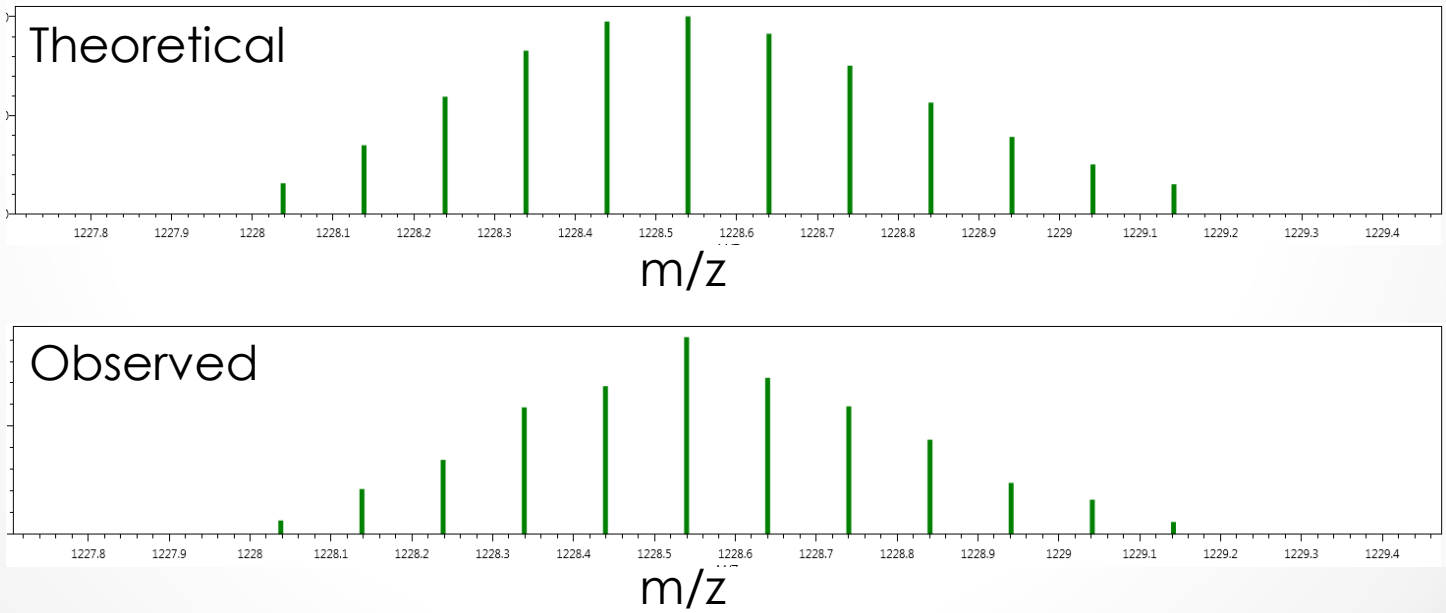
Isotopic Profile View



Isotopic Envelope Quality

How well does the observed mass spectrum match the theoretical?

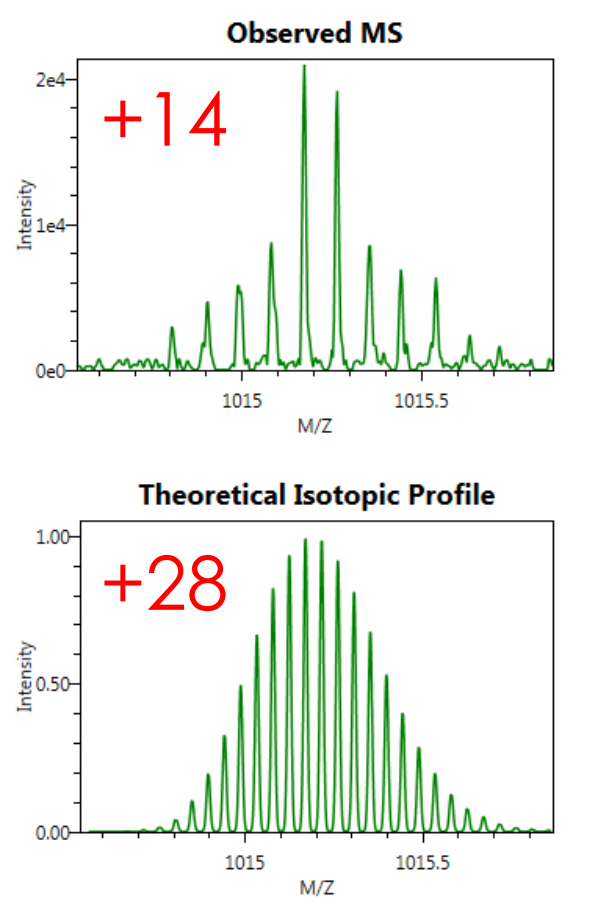
- Matches the spacing and relative intensity of peaks using Pearson correlation algorithm



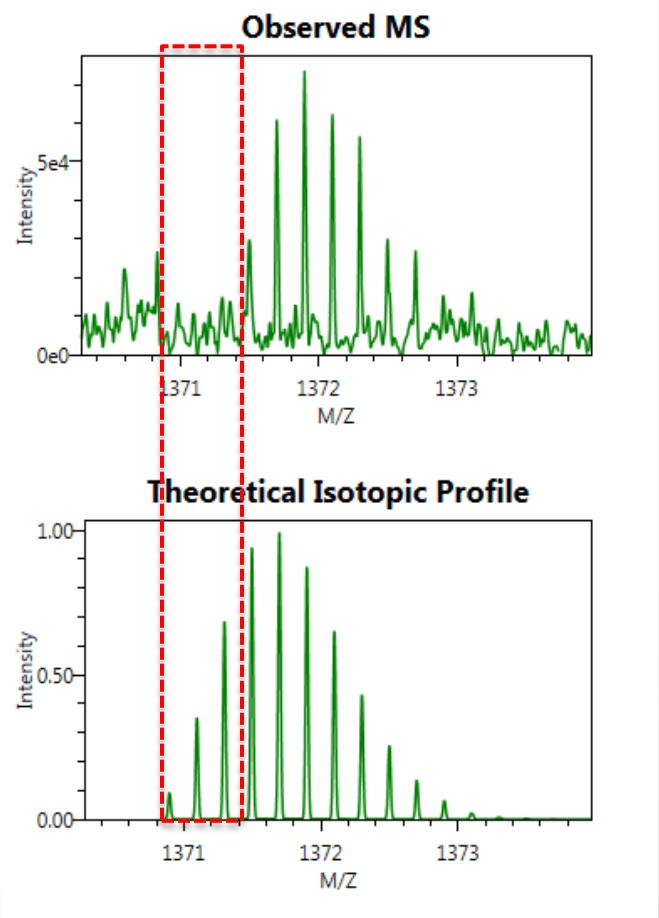
Isotopic Quality Score: 0.986

Validating Identifications

- Wrong charge determination

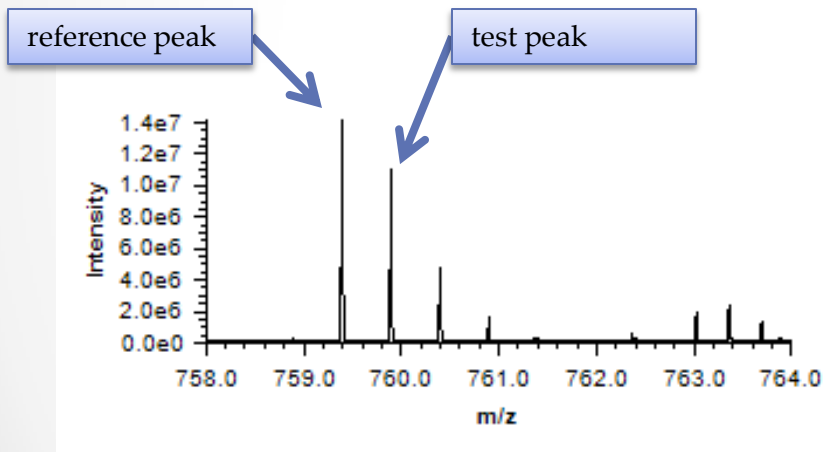


- Wrong isotopic peak selected as mono-isotope

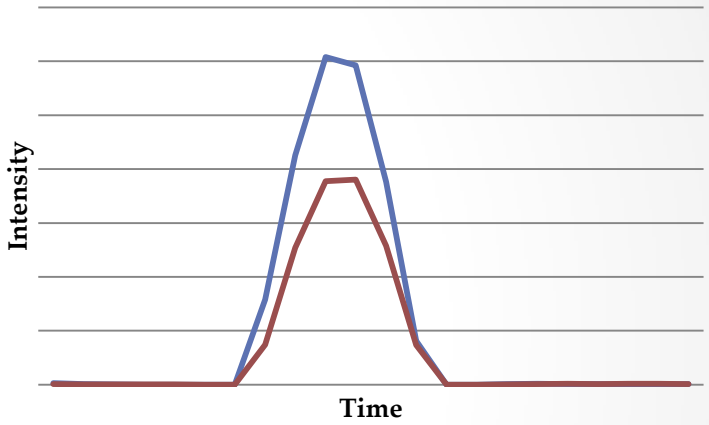


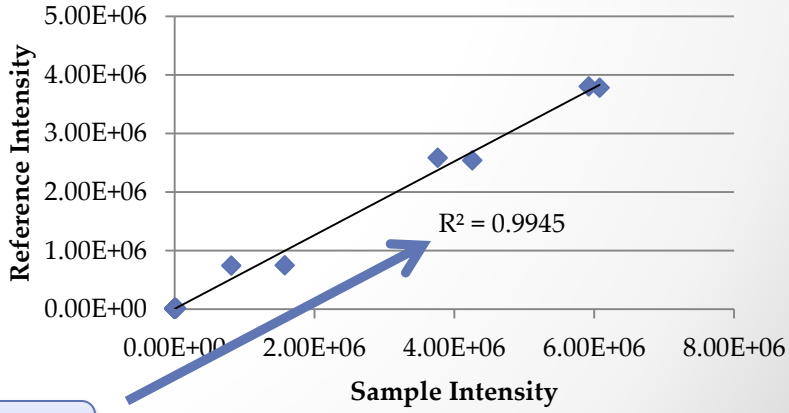
Chromatographic Profile Correlation

- Testing for LC peak shape



Analytical Chemistry
MacCoss & Yates 2003

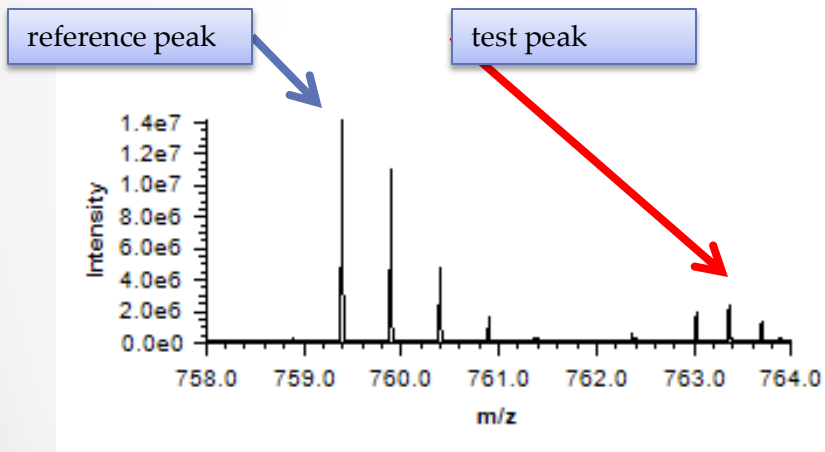




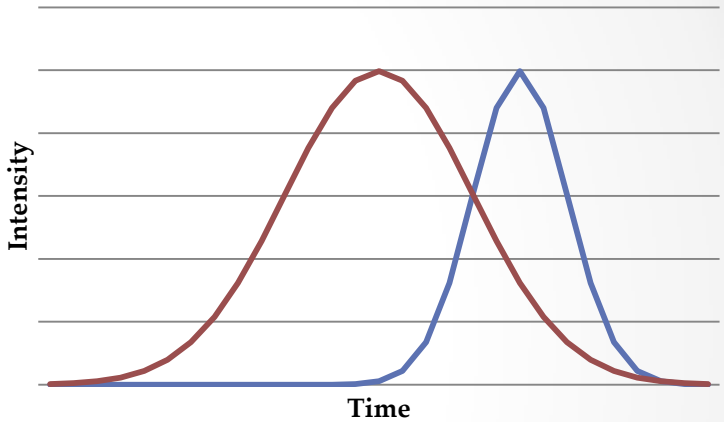
Correlation Value

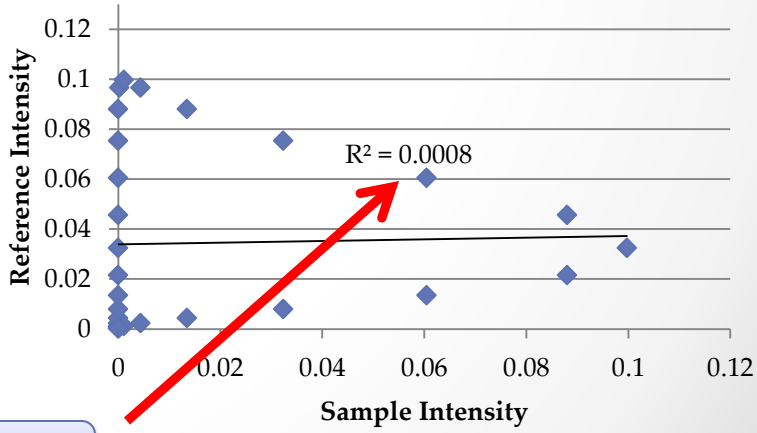
Chromatographic Profile Correlation

- Testing for LC peak shape



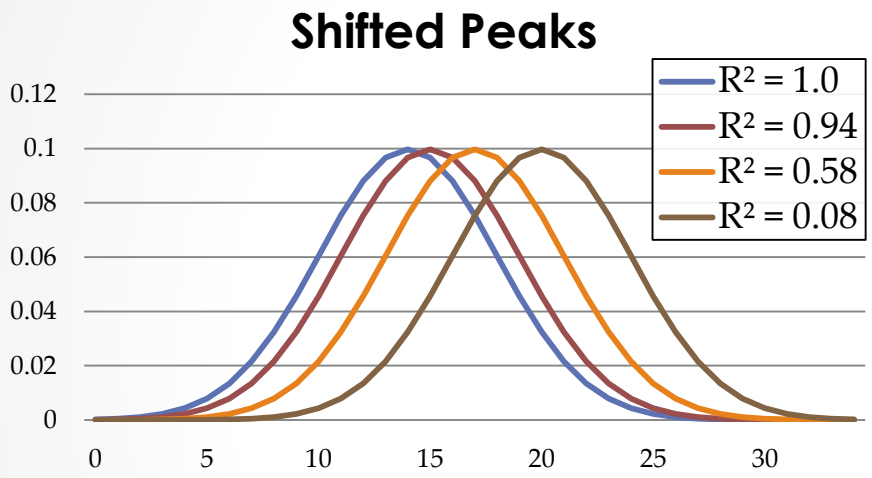
Analytical Chemistry
MacCoss & Yates 2003

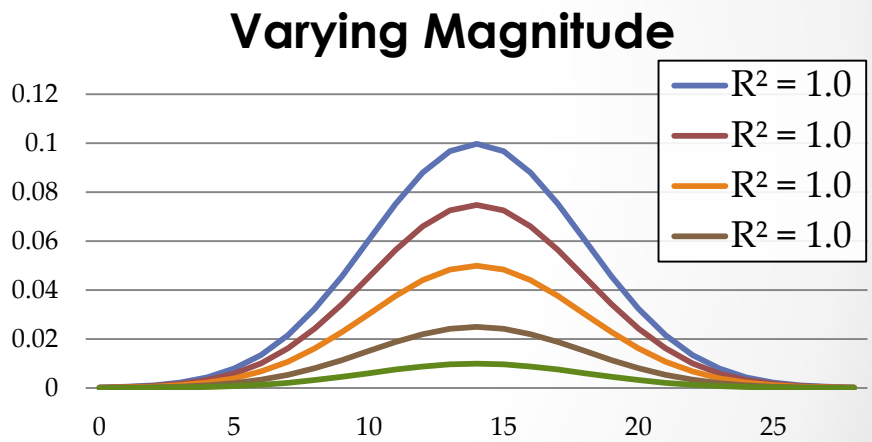


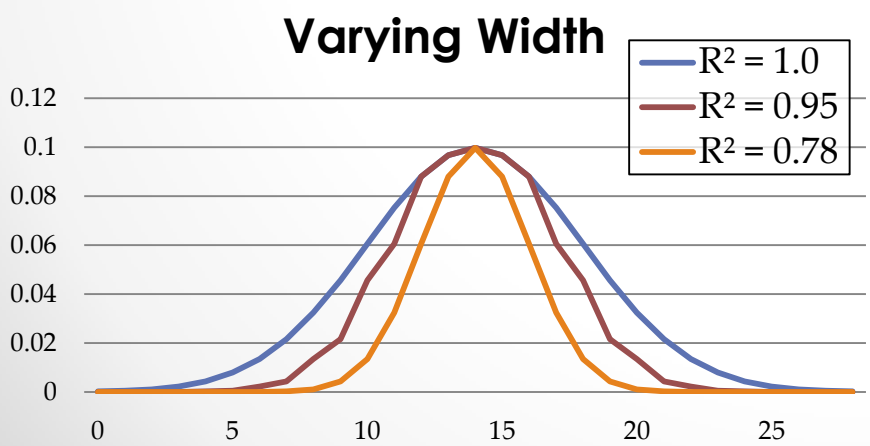


Correlation Value

Chromatographic Profile Correlation

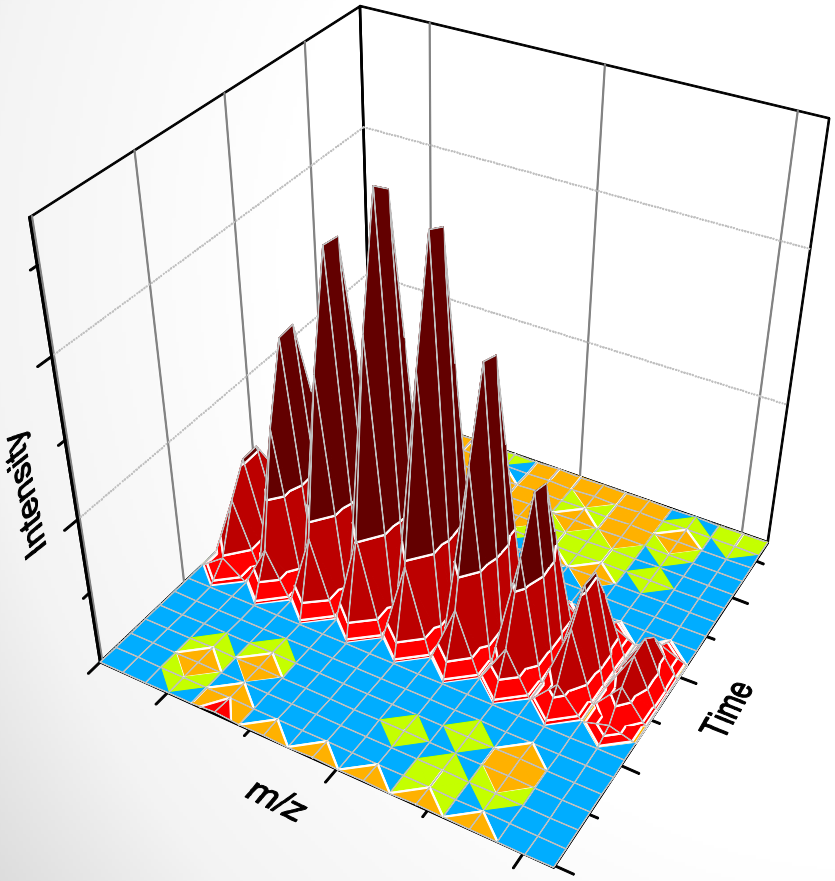


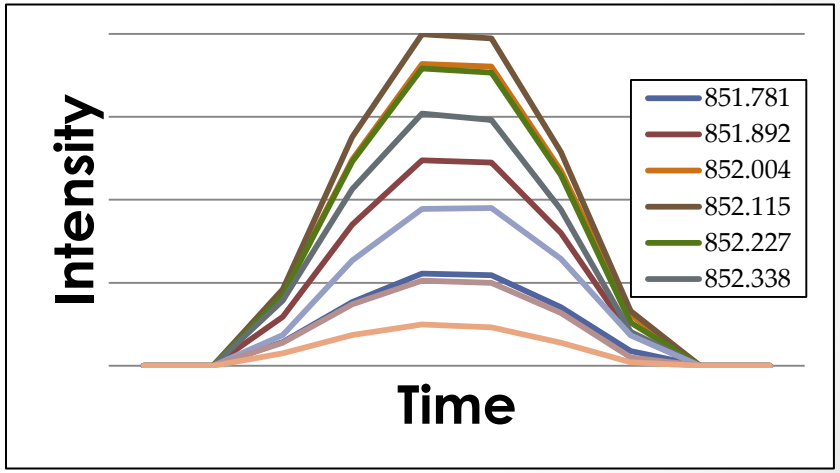




Isotopic Envelope Correlation

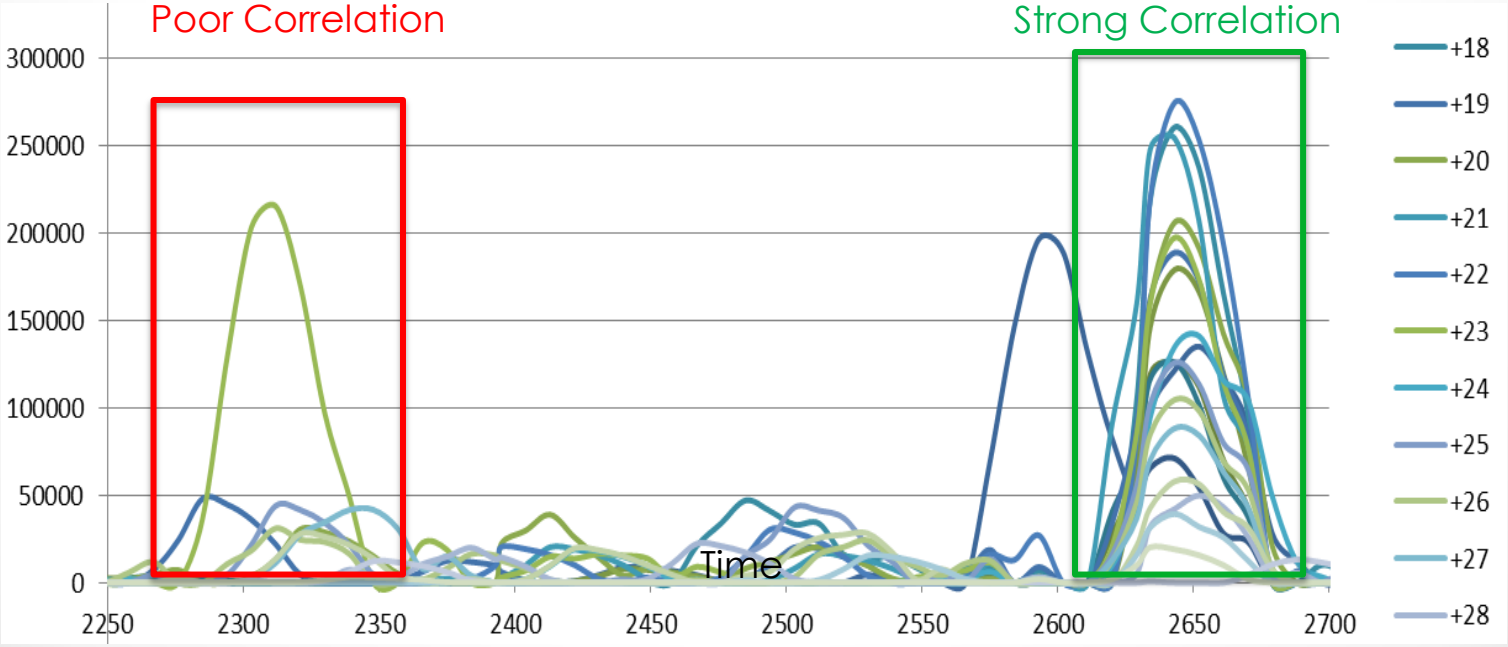
Isotopes of the same species should co-elute





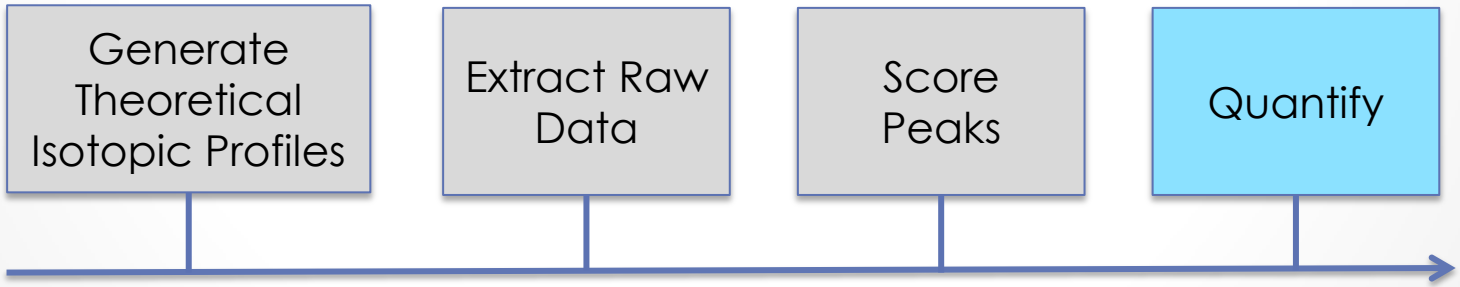
Charge State Chromatogram Correlation

Different charge states of the same species have identical LC elution profiles



Proteoform Quantitation

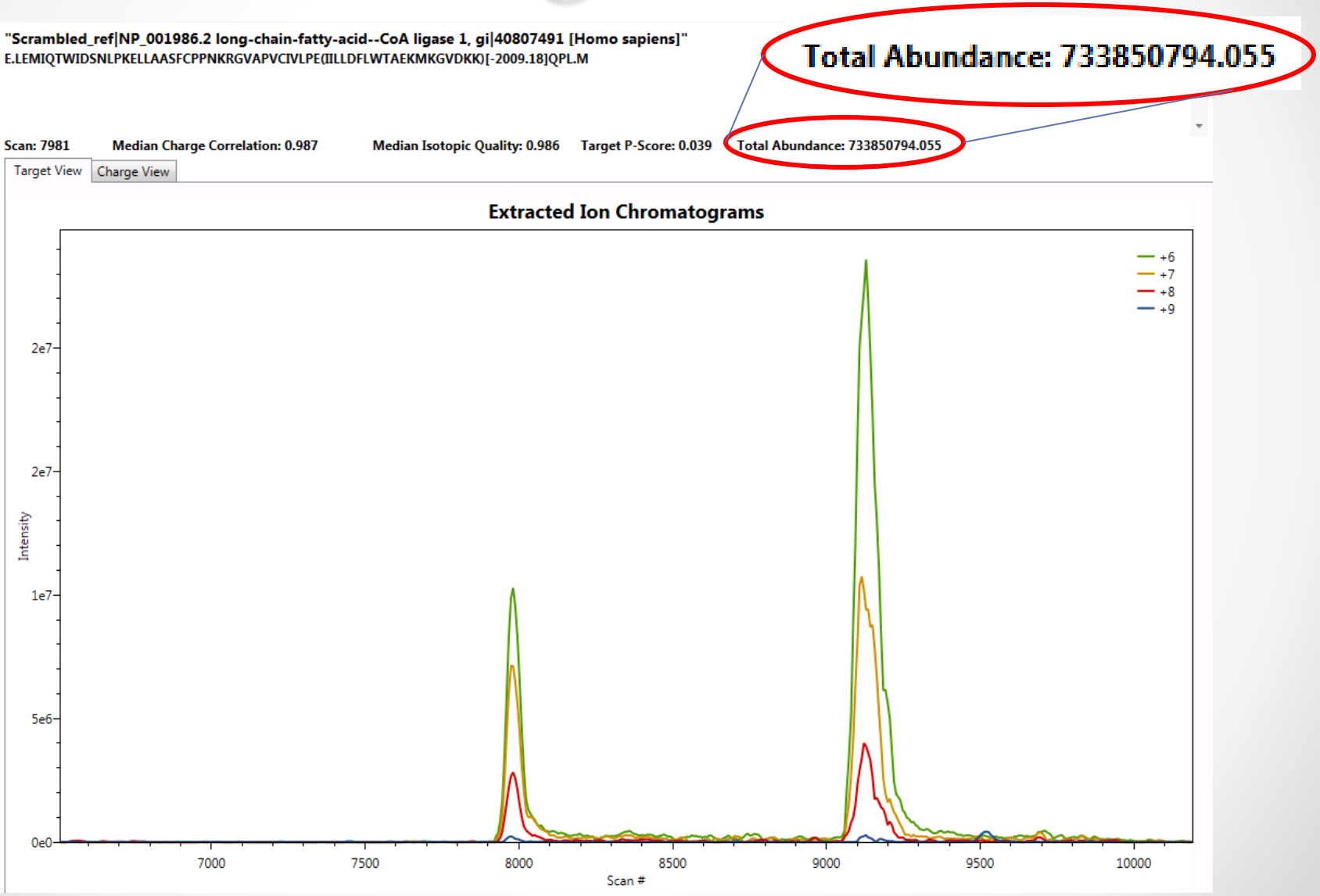
- Chromatographic peaks that meet scoring thresholds are grouped together by elution time(one peak per charge state)
- Chromatogram peaks in each group are summed to report quantitative values at a proteoform level



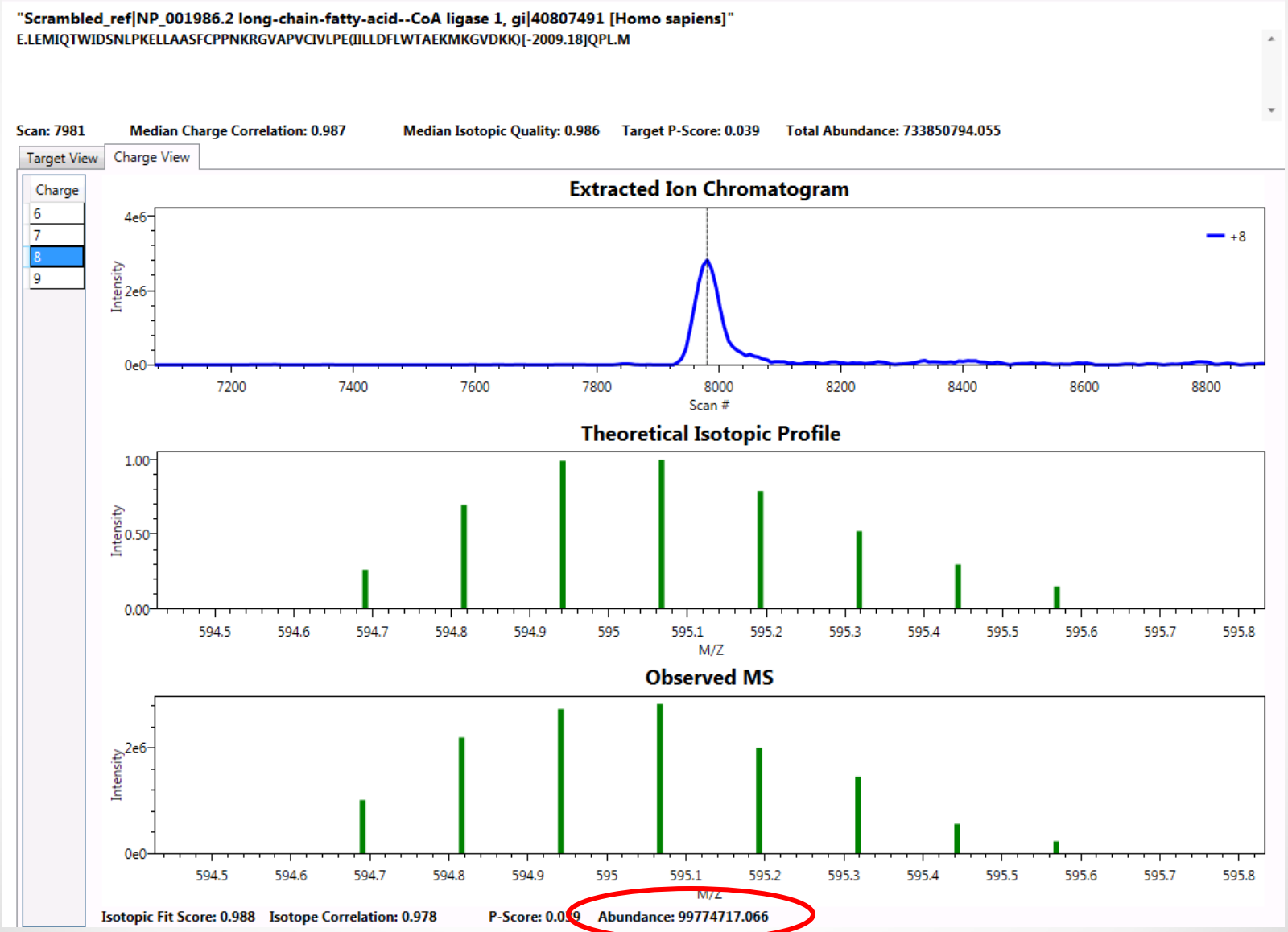
High Throughput Implementation

- IQ software
  - C# open source
  - Graphical and command-line interface
- Automated quantitation analysis for LC-MS runs
  - Input: MS data file and Protein Target file (e.g. MS-Align+)
- Runtime: ~ 1 minute per LC-MS run
  - Generates ~ 1,000,000 XICs
  - Calculates scoring metrics for ~1000 proteoforms

IQ Target View



IQ Charge View



Summary

- IQ utilizes proteoform atomic compositions and LC elution time to guide data processing.
- Top-Down IQ software package provides high-throughput quantitation of top-down data
- **Beta version AVAILABLE NOW! http://omics.pnl.gov/software**

Acknowledgements

- NIH/NIGMS P41 GM103493
- NIH/NIAID Y1-AI-8401
- DOE/BER Early Career Award
- Environmental Molecular Science Laboratory at PNNL