Informed Identification and Quantitation of Peptides in LC-IMS-MS Based Proteomics Data

Kevin L. Crowell, Sangtae Kim, Erin S. Baker, Richard D. Smith, Samuel H. Payne
Pacific Northwest National Laboratory, Biological Sciences Division, Richland, WA

Pacific Northwest NATIONAL LABORATORY

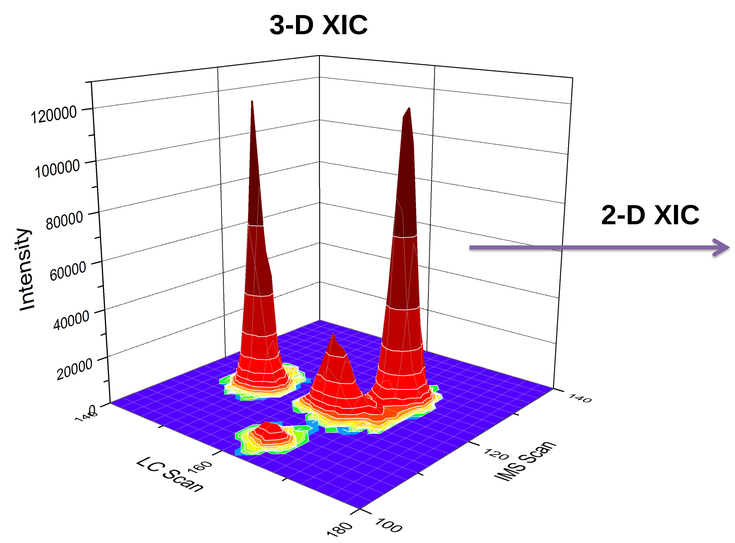
Introduction

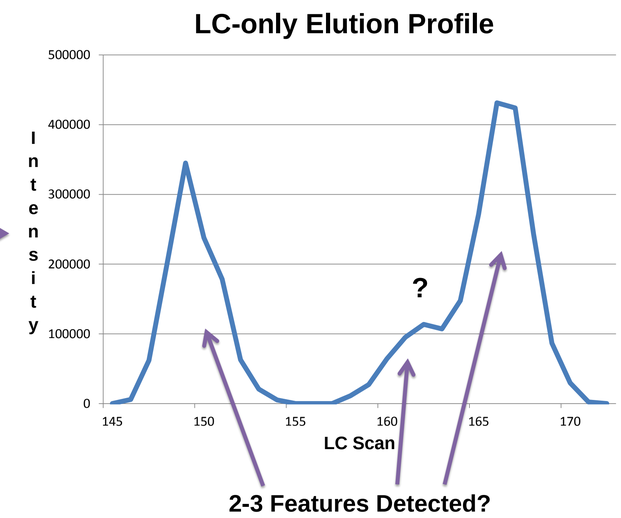
- Ion mobility provides an extra dimension of separation for greater confidence in peptide identifications
- Data processing algorithms for mass spectrometry data are numerous, slow, and prone to errors
- Informed analysis, using prior knowledge, is becoming a popular technique of data analysis
- Improvements to informed analyses such as faster processing algorithms are necessary to search all possible peptides (millions) in the genome

Algorithm Workflow

- Users enumerate target list (e.g. all tryptic peptides)
- For each peptide target
  - For each possible charge state
    - Generate theoretical isotopic profile
    - Grab 3-D Extracted Ion Chromatogram (XIC)
      - Elution time, drift time, intensity
    - Find individual 3-D features
    - For each feature
      - Find apex of LC and IMS separation and perform saturation correction if necessary
      - Extract observed isotopic profile and compare to theoretical
      - Filter bad results
        - Mass error, poor isotopic envelope quality, incorrect charge state, etc.
      - Quantify
  - Correlate charge states using elution time
  - Report results

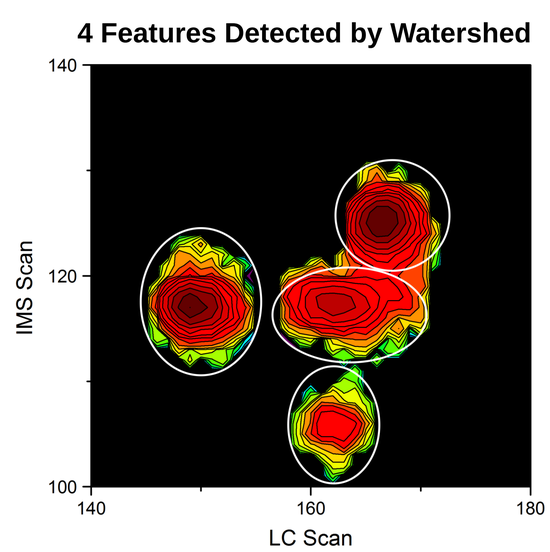
Feature Extraction and Detection





- XIC-based feature extraction instead of spectrum-based
  - Average of 0.217 ms per XIC extraction
- Uses a watershed algorithm for feature detection commonly used in image processing

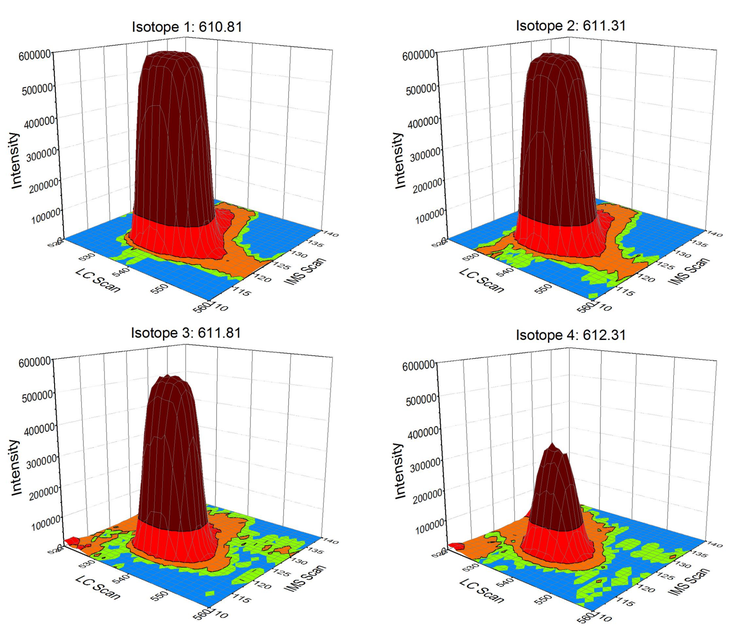
- Unable to detect all features using LC dimension
- Ion mobility dimension allows for detection of 4 features

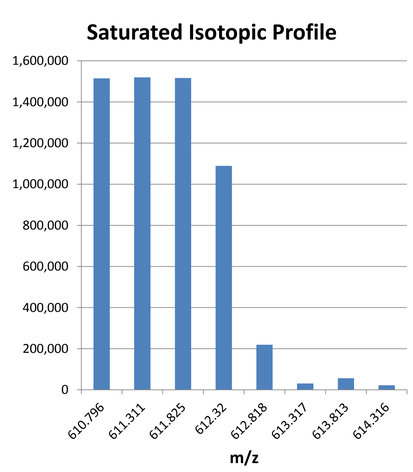


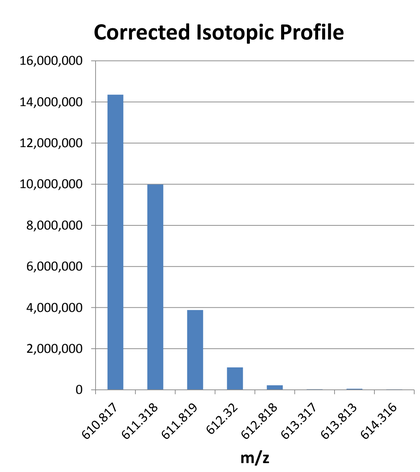
Watershed Algorithm

- The watershed algorithm is extremely fast
  - Average of 0.255 ms per 3-D XIC
- Commonly used for edge detection in images
  - Imagine 3-D features as an overhead heat map
- How to use the algorithm:
  - First apply Savitzky-Golay smoothing
  - Filter out low-level noise features if desired

Saturation Correction







- Find isotope that is not saturated (4th isotope in above case)
- Use unsaturated isotope to find accurate LC and IMS apex

- Saturation correction increases dynamic range
  - 1 order of magnitude added in above case
- Allows for accurate quantitation of very abundant peptides

Conclusions

- Algorithms developed and implemented for informed analysis of LC-IMS-MS data
- Uses prior knowledge about each peptide target to extract evidence for the peptide
- Replaces multiple levels of processing normally executed globally during data analysis such as deisotoping, feature finding, etc.
- Fast processing allows for targeting millions of peptides instead of thousands
- Saturation correction provides greater dynamic range
  - Essential for ToF-based instruments
- Similar approaches can be adopted to traditional LC-MS data

U.S. DEPARTMENT OF ENERGY

Acknowledgements
Portions of this research were supported by the National Institutes of Health P41 GM103493, R01 ES022190, R21 GM103497, and by the Department of Energy Office of Biological and Environmental Research Genome Sciences Program under the Pan-omics project. Significant portions of the work were performed in the Environmental Molecular Science Laboratory, a DOE/BER national scientific user facility at Pacific Northwest National Laboratory (PNNL) in Richland, Washington.

Contact Information
Kevin L. Crowell
Kevin.Crowell@pnnl.gov

omics.pnl.gov