Peptide identification using spectral library search

Jian Wang1, Monika Tuchela2, Brett Larsen3, Stephen Tate2, Anne-Claude Grega2 and Nuno Bandeira4,4

1. Bioinformatics Program, UCSD, La Jolla, CA
2. Center for System Biology, Samuel Lunenfeld Research Institute, Toronto, Ontario, Canada
3. AB Sites, Toronto, Canada
4. Skaggs School of Pharmacy and Pharmaceutical Science, UCSD, La Jolla, CA
5. Center for Computational Mass Spectrometry, UCSD, La Jolla, CA

Contact: jwang06@ucsd.edu; bandeira@ucsd.edu

Introduction
The complexity of proteomics samples continues to pose significant challenges for traditional data dependent acquisition (DDA) workflows. Data independent acquisition (DIA) strategies have been used to increase the reproducibility and comprehensiveness of data collection. Recent developments such as the SWATH MS method make it possible to perform DIA at high speed, high mass accuracy and with sensitivity and dynamic range comparable to selected reaction monitoring assays (SRMs). While SWATH was proposed as a targeted approach for peptide quantification, all peptides in the sample are analyzed in every run and thus creates the possibility of extending this method for peptide identification from the resulting multiplex spectra. Here we evaluate the feasibility of using spectral library search for peptide identification in SWATH data.

Data-independent acquisition
Data dependent acquisition (DDA) selects the most intense precursor ions in MS scan for SRM analysis. Limitation only a small fraction of abundant peptide precursors are selected for MS/MS analysis and quantitation.

Data-independent acquisition (DIA): cycle through wide selections of precursor ions. Segment and analyze everything in the elution window (SWATH) thereby the wide window can be a series of 1280 scans/12.2 kHz. 1280 peaks.

Methods Overview

Traditional DIA approach: spectral library search for peptide identification in SWATH data. Peptide identification from mixture tandem mass spectra Data independent acquisition

Simple example: A SWATH spectrum M is a mixture of two spectra A (red peaks) and B (blue peaks), where A and B are single peptide spectra from two different peptides.

Spectral similarity is computed using cosine:

$$\text{Cosine}(A, B) = \frac{A \cdot B}{||A|| \cdot ||B||}$$

We evaluate how a SWATH spectrum M is similar to the spectra A and B using the cosine of their projections to the A and B axes.

Dependent library matches are library spectra that are very similar to each other. Here, we evaluate the feasibility of using spectral library search for peptide identification in SWATH data.

Identification of SWATH spectra: 1280 DDA runs can be considered a mixture of a set of spectra from DDA runs M = A1 + A2 + ... + An where A is spectra from DDA runs.

We assigned UPS standard proteins recycled with E. coli. Cell-based backbone using multiple runs of DDA in order to build spectral library and assess characteristics of QTOF spectra. The library was then used to identify SWATH spectra.

UPS peptides = E.coli peptides in spectral libraries has lower spectral quality which affect their identifiability in UPS spectra.

Increasing background complexity has small effect on the statistics of UPS peptides indicating it is not the main reasons peptides not identified. In general UPS peptides have good quality spectra in the library. This is also indicated by the two different distribution of signal intensity and peptide abundance vs location of the peptide in SWATH data.

For peptide at retention time, their relative intensity is usually low or obtained or not detected for spectral matching. Spectrum with more noise peaks are of better identifiability. Severe peptide peaks are also used in other spectral match to other spectra from the same peptide (defined by database search results).

Peptide identification

Example spectra from the same peptides at different position of the elution profile.

Spectral similarity of QTOF spectra

Spectral similarity of QTOF spectra can be used to match a SWATH spectrum. Which can create false positive matches.

Peptide abundance vs spectral quality

Example spectra from the same peptide at different position of the elution profile. Scans in which indicate they are not possible presented in the graph.

Spectral similarity of UPS spectra

Example spectra from the same peptides at different position of the elution profile. Scans in which indicate they are not possible presented in the graph.

Peptide identification in SWATH data

Spectral abundance vs spectral quality

Understanding peptide identifiability in SWATH data

We evaluate how a library spectrum M is similar to the spectra A and B using the cosine of their projections to the A and B axes. Here, we evaluate the feasibility of using spectral library search for peptide identification in SWATH data.

Library search

LCMS map

LCMS analytics

DIA

Parallel analysis

Spectral library search

Peptide identifications

DDA

Build spectral library

- No QTOF E.coli or UPS spectral library available
- Build appropriate spectral library from DDA runs
- Use parallel DDA runs to compare with SWATH data

Identification of SWATH spectra:

<table>
<thead>
<tr>
<th>Sample</th>
<th>UPS peptides</th>
<th>UPS E.coli</th>
<th>UPS E.coli</th>
<th>UPS E.coli E.coli lysate</th>
<th>UPS E.coli E.coli lysate</th>
</tr>
</thead>
<tbody>
<tr>
<td>UPS1_450pm</td>
<td>1280</td>
<td>1124</td>
<td>1380</td>
<td>1385</td>
<td>1280</td>
</tr>
<tr>
<td>UPS1_450pm_Ecoli</td>
<td>1107</td>
<td>6772</td>
<td>1014</td>
<td>3002</td>
<td>4936</td>
</tr>
<tr>
<td>UPS2_45pm</td>
<td>725</td>
<td>599</td>
<td>761</td>
<td>761</td>
<td>761</td>
</tr>
<tr>
<td>UPS2_45pm_Ecoli</td>
<td>590</td>
<td>3788</td>
<td>506</td>
<td>1775</td>
<td>2956</td>
</tr>
</tbody>
</table>

Spectral similarity

When peaks of the peaks with lowest intensity are removed, the intensity of the main peaks Signal peaks = Annexated peaks. Ten are 20%, 60%, 80%, 100%.

For peaks at retention time, their relative intensity is usually low or obtained or not detected for spectral matching. Spectrum with more noise peaks are of better identifiability. Severe peptide peaks are also used in other spectral match to other spectra from the same peptide (defined by database search results).

DIA...