Approach for large-scale identification of linked peptides from tandem mass spectra

Jian Wang¹, Veronica Anania², Jeff Knoll³, John Rush⁴, Jennie R Li⁵, Philip E Bourne⁶ and Nuno Bandeira⁷,⁸

1. Bioinformatics Program, UCSD, La Jolla, CA
2. Protein Chemistry Department, Genentech, Inc., 2880 Woodside, South San Francisco, California
3. Cell Signaling Technologies, Danvers, MA.
4. Skaggs School of Pharmacy and Pharmaceutical Sciences, UCSD, La Jolla, CA
5. Center for Computational Mass Spectrometry, UCSD, La Jolla, CA
6. Computer Science and Engineering, UCSD, La Jolla, CA

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

Introduction:
Chemical cross-linking and mass spectrometry have been shown to constitute a powerful tool to study protein-protein interactions and to help elucidate the structure of large protein complexes. However, computational methods to interpret the complex MS/MS spectra from linked peptides are still in their infancy, thus making the high-throughput application of this approach largely impractical. Here we use disulfide-linked peptides as an example to describe a generic procedure to: i) efficiently generate large mass spectral reference data for linked peptides and ii) use this data to systematically train an algorithm that can efficiently and accurately identify linked peptides from MS/MS spectra.

Challenge of identification of linked peptides:
1. Linked Peptide has substantially different fragmentation pattern than unlinked linear peptide.
2. MS/MS spectra from linked peptide contain a mixture of fragments from more than one peptide.
3. Lack of a large and reliable annotated dataset to learn the fragmentation pattern of linked peptides.

Method:
1. Generating large training datasets using combinatorial synthetic peptide libraries: Building and accurate scoring model for peptide identification usually requires a large set of reliably annotated spectra. However, such datasets are usually hard to procure to obtain without having a deep understanding of the links and non-links between a pair of peptides. Here we address the ‘chicken and egg’ problem. We build the cycle using combinatorial synthetic peptide libraries.

2. Developing scoring function for linked peptides:
   a. Non-linked fragments of linked peptides, although share some characteristics with fragments from unlinked peptides, their intensity are generally suppressed (panel a); b. Linked fragments also have very different fragmentation compare to those of unlinked peptides, particularly high-charge fragments are very prominent (panel c).
   c. We develop a rigorous probabilistic models that capture the specific fragmentation patterns of linked peptides.

3. Workflow of database search method for linked peptides
   a. We show that this new approach can identify thousands of MS/MS spectra from linked peptides.
   b. We developed a rigorous probabilistic models that capture the specific fragmentation patterns of linked peptides.
   c. We propose that this approach can identify and predict thousands of MS/MS spectra from disulfide-bridged peptide pairs.

Results:

<table>
<thead>
<tr>
<th>Library</th>
<th>Interlinked peptides</th>
<th>Disulfide-bridged peptides</th>
</tr>
</thead>
<tbody>
<tr>
<td>Library 1</td>
<td>1077</td>
<td>791</td>
</tr>
<tr>
<td>Library 2</td>
<td>2636</td>
<td>965</td>
</tr>
<tr>
<td>Library 3</td>
<td>1077</td>
<td>791</td>
</tr>
</tbody>
</table>

Examples of identified MS/MS spectra from linked peptides

Acknowledgements:
Current database search methods do not try to capture the specific fragmentation of linked peptides because there are limited number of annotated data to learn fragmentation patterns of linked peptides.

Using disulfide-linked peptides as an example, we demonstrate that the use of combinatorial synthetic peptide libraries is an effective way to generate a large and reliable reference MS/MS dataset for linked peptides.

We developed a rigorous probabilistic models that capture the specific fragmentation patterns of linked peptides.

We propose that this new approach can identify and predict thousands of MS/MS spectra from disulfide-bridged peptide pairs.

Our approach can be generalized to identify peptides with other linked bonds.