Protein Cross-Linking Analysis using Stable Isotope-Labeling, Mass Spectrometry, and Integrated Computational Data Processing

Jan Seebacher

Aebersold Lab, ISB, Seattle WA
In Collaboration with Michael Gelb, University of Washington
3D Structure Determination of Proteins and Protein Complexes

Molecular Biology / Affinity Purification / Mass Spectrometry

Protein Crystallography

Protein Crystallography

NMR

Sequence

Protein Complex / Protein Interactions

Candidate Structures

3D Protein Structure Prediction

Modeling
3D Structure Determination of Proteins and Protein Complexes

Chemical Cross-Linking

Constraints

Molecular Biology / Affinity Purification / Mass Spectrometry

Protein Crystallography

Protein Complex / Protein Interactions

Candidate Structures

3D Protein Structure Prediction

- NMR
- Sequence
- Modeling
Protein Cross-linking to Study 3D Protein Structure: Proof of Concept

3D Protein Structure

Protein Complex (Proteins A+B) known 3D structure, pdb.

Cross-linking Reaction

Protein Sequence A

Protein Sequence B

Cross-Linking Reagent
"Molecular Ruler"

Method Validation

Inter-Residue Distance Constraints

Protein Structure Prediction

Detection of Protein Interactions / Sites

Identification of Cross-linked AA Residues (K, n)
Method Outline: Exp. Workflow + Data Analysis

1:1 Ratio

[d_0/d_{12}]-isotope-coded Cross-Linker

18O-Isotope-Coded Buffer

[16O] [16/18O]

Protein Cross-linking

Automated MS/MS Analysis

• Peptide Identification
• Identification of Cross-Linked Residues
• Generation of “Distance Constraints” for 3D Structure Modeling, i.e. for Rosetta, NMR, etc.
• Identification of potential Interaction Sites

Automation MS Analysis:
• Isotope pattern recognition
• Spectra Alignment
• Peptide Data Base Search
• Mass Mapping → Candidates
• → MS2 Inclusion List

Robotic LC Fractionation onto MALDI Plate

Digestion, X: Proteins → Peptides

MS2 (MS/MS) Acquisition

MS1 Acquisition

ABI 4700 Proteomics Analyzer

[16O]

[16/18O]
Strategy: Protein Cross-Linking and MS Analysis

Cross-Linking Reagent: Unmodified Peptides + Mono-Links
Cross-Links

Protein Complex

Protein-reactive Group

Cross-Linking Reagent

Protease(s)

dead-end, hydrolyzed → Mono-Link

→ Identification by Mass Spectrometry

→ Digestion Products:

Unmodified Peptides + Mono-Links

Loop-Link

→ Cross-Links
MS$^1$ Analysis of isotope-coded cross-linked peptides

Cross-linking Reagent

Unmodified Peptides + Mono-Links

[16O]

[16/18O]

→ Cross-Link or Loop-Link

→ Mono-Link

[12 Da]

[2 Da]
MS\(^1\) Analysis of isotope-coded cross-linked peptides

This can be automated → development of Software Tools
MS\textsuperscript{1} Data Analysis with iXLINK (Parag Mallick)

A. Filter noise by intensity & s/n ratio

B. Save high quality doublets to inclusion list for cross-linking analysis. Save intense peaks to inclusion list for protein identification analysis.

C. Annotate doublets as cross-links, mono-links, or unknown

D. Theoretical digest and modification

E. Assign peaks to sequences by mass mapping
Example MS$^1$ spectra - d$_0$/d$_{12}$-DSS
Example MS¹ spectra - d₀/d₆-DSG

4700 Reflector Spec #1 MC[BP = 2260.1, 2788]

6 Da

ABI 4700 Proteomics Analyzer
Example MS$^1$ spectra – $d_0/d_4$-BS$^3$
Example LC- MALDI MS\textsuperscript{1} Data (iXLINK output)

![Diagram showing isotope patterns in a mass spectrum plot with LC-MS Fraction ("Retention Time") on the x-axis and m/z [Da] on the y-axis. The diagram includes two datasets: 
- [\textsuperscript{16}O] - Peptide Sample
- [\textsuperscript{16}/\textsuperscript{18}O] - Peptide Sample

Image created with program “MTPeak” by Paul Loriaux]
MS Analysis Workflow

- MS^2 Spectra Alignment
- Isotope Pattern (Reporter Ions)
- iXLINK output (Peptide Candidates)
- Peptide Fragment Ion Mass Matching
- Peptide Scoring (ProBlD[^1])

doXLINK Analysis (Ning Zhang)

Mass Pairs

iXLINK Inclusion List

User Validation

A “good” MALDI MS/MS spectrum of a light/heavy cross-linked peptide pair
A “good” MALDI MS/MS spectrum of a light/heavy cross-linked peptide pair
Some typical MALDI MS/MS spectra of cross-linker-modified peptide species
Validation of Computed Cross-Linking Results with XLinkViewer (James Eddes)

<table>
<thead>
<tr>
<th>File</th>
<th>Score</th>
<th>Delta Score</th>
<th>(Loc 1) Peptide 1</th>
<th>(Loc 2) Peptide 2</th>
<th>Peptide Mass (mm)</th>
<th>Matches ions (total)</th>
<th>Matched pairs (total)</th>
<th>Reporter misses (total)</th>
<th>Assignment</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATG_MEMS_1004.mrt</td>
<td>18.82</td>
<td>0.49</td>
<td>(BS377) TGEVSQGHPR</td>
<td>(-) MONO</td>
<td>1004.50 (0.22)</td>
<td>4(16)</td>
<td>0(11)</td>
<td>0.00 (0.00)</td>
<td>correct</td>
</tr>
<tr>
<td>ATG_MEMS_1797.mrt</td>
<td>10.80</td>
<td>0.14</td>
<td>(BS377) PTVKPKQK</td>
<td>(-) MONO</td>
<td>1004.50 (0.75)</td>
<td>2(10)</td>
<td>2(10)</td>
<td>0.00 (0.00)</td>
<td>incorrect</td>
</tr>
<tr>
<td>ATG_MEMS_7766.mrt</td>
<td>14.54</td>
<td>0.00</td>
<td>(BS377) GYSGFR</td>
<td>(-) MONO</td>
<td>797.89 (0.16)</td>
<td>2(12)</td>
<td>1(1)</td>
<td>0.00 (0.00)</td>
<td>correct</td>
</tr>
<tr>
<td>ATG_MEMS_1004.mrt</td>
<td>24.33</td>
<td>0.50</td>
<td>(BS377) GYSGQR</td>
<td>(-) MONO</td>
<td>1004.50 (0.22)</td>
<td>2(16)</td>
<td>1(5)</td>
<td>0.00 (0.00)</td>
<td>correct</td>
</tr>
<tr>
<td>ATG_MEMS_1004.mrt</td>
<td>42.96</td>
<td>0.57</td>
<td>(A681) AANGKPKPK</td>
<td>(-) MONO</td>
<td>1003.55 (0.19)</td>
<td>2(10)</td>
<td>1(1)</td>
<td>186.26 (0.00)</td>
<td>correct</td>
</tr>
<tr>
<td>ATG_MEMS_1004.mrt</td>
<td>38.53</td>
<td>0.05</td>
<td>(A681) AANGKPKPK</td>
<td>(-) MONO</td>
<td>1003.53 (1.19)</td>
<td>2(10)</td>
<td>1(1)</td>
<td>186.26 (0.00)</td>
<td>correct</td>
</tr>
<tr>
<td>ATG_MEMS_1004.mrt</td>
<td>40.76</td>
<td>0.57</td>
<td>(A681) AANGKPKPK</td>
<td>(-) MONO</td>
<td>1003.53 (1.19)</td>
<td>2(10)</td>
<td>1(1)</td>
<td>186.26 (0.00)</td>
<td>correct</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>File</th>
<th>Score</th>
<th>Delta Score</th>
<th>(Loc 1) Peptide 1</th>
<th>(Loc 2) Peptide 2</th>
<th>Peptide Mass (mm)</th>
<th>Matches ions (total)</th>
<th>Matched pairs (total)</th>
<th>Reporter misses (total)</th>
<th>Assignment</th>
</tr>
</thead>
<tbody>
<tr>
<td>TQ11504R0.mrt MONO</td>
<td>1004.50</td>
<td>0.49</td>
<td>TQ11504R0</td>
<td>MONO</td>
<td>1004.50 (0.22)</td>
<td>4(16)</td>
<td>0(11)</td>
<td>0.00 (0.00)</td>
<td>correct</td>
</tr>
<tr>
<td>A8K Membrane MONO</td>
<td>1002.59</td>
<td>0.57</td>
<td>A8K Membrane</td>
<td>MONO</td>
<td>1003.53 (1.19)</td>
<td>2(10)</td>
<td>1(1)</td>
<td>186.26 (0.00)</td>
<td>correct</td>
</tr>
<tr>
<td>K8K</td>
<td>PSFIONG</td>
<td>1003.57</td>
<td>0.49</td>
<td>PSFIONG</td>
<td>1004.50 (0.22)</td>
<td>2(12)</td>
<td>1(1)</td>
<td>0.00 (0.00)</td>
<td>correct</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>cats</th>
<th>int</th>
<th>matched</th>
<th>annotation</th>
</tr>
</thead>
<tbody>
<tr>
<td>201</td>
<td>333</td>
<td>true</td>
<td>false</td>
</tr>
<tr>
<td>20.1</td>
<td>44.9</td>
<td>true</td>
<td>false</td>
</tr>
<tr>
<td>201</td>
<td>34.3</td>
<td>true</td>
<td>false</td>
</tr>
<tr>
<td>201</td>
<td>100.0</td>
<td>true</td>
<td>false</td>
</tr>
<tr>
<td>201</td>
<td>20.5</td>
<td>true</td>
<td>false</td>
</tr>
<tr>
<td>175.3</td>
<td>23.2</td>
<td>true</td>
<td>(Y)R</td>
</tr>
<tr>
<td>288.4</td>
<td>24.1</td>
<td>true</td>
<td>false</td>
</tr>
<tr>
<td>345.4</td>
<td>17.4</td>
<td>true</td>
<td>(Y)GKR</td>
</tr>
<tr>
<td>317.1</td>
<td>2.1</td>
<td>true</td>
<td>false</td>
</tr>
<tr>
<td>400.5</td>
<td>3.1</td>
<td>true</td>
<td>false</td>
</tr>
<tr>
<td>457.5</td>
<td>2.0</td>
<td>true</td>
<td>false</td>
</tr>
<tr>
<td>474.2</td>
<td>3.0</td>
<td>true</td>
<td>(Y)GKR</td>
</tr>
<tr>
<td>539.7</td>
<td>3.7</td>
<td>true</td>
<td>false</td>
</tr>
<tr>
<td>541.8</td>
<td>3.6</td>
<td>true</td>
<td>false</td>
</tr>
<tr>
<td>561.7</td>
<td>2.5</td>
<td>true</td>
<td>(Y)GKR</td>
</tr>
<tr>
<td>828.3</td>
<td>4.4</td>
<td>true</td>
<td>false</td>
</tr>
<tr>
<td>933.8</td>
<td>17.4</td>
<td>true</td>
<td>(Y)GKR</td>
</tr>
<tr>
<td>854.6</td>
<td>2.7</td>
<td>true</td>
<td>false</td>
</tr>
<tr>
<td>891.1</td>
<td>7.4</td>
<td>true</td>
<td>false</td>
</tr>
<tr>
<td>981.2</td>
<td>2.8</td>
<td>true</td>
<td>false</td>
</tr>
</tbody>
</table>
## Analysis Results

<table>
<thead>
<tr>
<th>Spectrum</th>
<th>PrecursorMH</th>
<th>CalculatedMH</th>
<th>Error</th>
<th>Da</th>
<th>Error_ppm</th>
<th>Peptide 1</th>
<th>Peptide 2</th>
<th>Location 1</th>
<th>Location 2</th>
<th>Score</th>
<th>DeltScore</th>
<th>TotalPepInDB</th>
<th>Assignment</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1349.57</td>
<td>1349.57</td>
<td>0.87</td>
<td>64</td>
<td>0.07</td>
<td>nMnMr</td>
<td>A61</td>
<td>B431</td>
<td>22.37</td>
<td>0.12</td>
<td>2</td>
<td>2</td>
<td>unassigned</td>
</tr>
<tr>
<td>3</td>
<td>1349.57</td>
<td>1349.58</td>
<td>0.11</td>
<td>81</td>
<td>1.51</td>
<td>TQDVSQKRR</td>
<td>nMnMr</td>
<td>B537</td>
<td>19.68</td>
<td>0.12</td>
<td>2</td>
<td>2</td>
<td>correct</td>
</tr>
<tr>
<td>4</td>
<td>1349.59</td>
<td>1349.60</td>
<td>0.1</td>
<td>74</td>
<td>0.01</td>
<td>nMnMr</td>
<td>a61</td>
<td>B431</td>
<td>23.67</td>
<td>0.13</td>
<td>2</td>
<td>2</td>
<td>unassigned</td>
</tr>
<tr>
<td>5</td>
<td>1349.58</td>
<td>1349.67</td>
<td>0.1</td>
<td>6522</td>
<td>0.08</td>
<td>A61</td>
<td>B431</td>
<td>19.47</td>
<td>0.16</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>unassigned</td>
</tr>
<tr>
<td>6</td>
<td>1045.49</td>
<td>1045.45</td>
<td>0.16</td>
<td>1004</td>
<td>3.10</td>
<td>TQDVSQKRR</td>
<td>MONO</td>
<td>B637</td>
<td>25.34</td>
<td>0.61</td>
<td>2</td>
<td>2</td>
<td>correct</td>
</tr>
<tr>
<td>7</td>
<td>1045.46</td>
<td>1045.46</td>
<td>0.16</td>
<td>1004</td>
<td>3.50</td>
<td>nMnMr</td>
<td>A81</td>
<td>B431</td>
<td>6.88</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>unassigned</td>
</tr>
<tr>
<td>8</td>
<td>1117.42</td>
<td>1117.6</td>
<td>0.01</td>
<td>150</td>
<td>0.01</td>
<td>PGKATMRGK</td>
<td>nMnMr</td>
<td>B431</td>
<td>10.08</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>unassigned</td>
</tr>
<tr>
<td>9</td>
<td>1000.38</td>
<td>999.88</td>
<td>0.05</td>
<td>799</td>
<td>0.05</td>
<td>PGKATMRGK</td>
<td>MONO</td>
<td>B431</td>
<td>15.7</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>unassigned</td>
</tr>
<tr>
<td>10</td>
<td>1000.38</td>
<td>999.88</td>
<td>0.1</td>
<td>799</td>
<td>0.1</td>
<td>PGKATMRGK</td>
<td>MONO</td>
<td>B431</td>
<td>15.7</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>unassigned</td>
</tr>
<tr>
<td>11</td>
<td>1002.42</td>
<td>1002.51</td>
<td>0.12</td>
<td>1197</td>
<td>0.51</td>
<td>NKK6FGK</td>
<td>nMnMr</td>
<td>B431</td>
<td>23.52</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>correct</td>
</tr>
<tr>
<td>12</td>
<td>1028.38</td>
<td>1027.48</td>
<td>0.89</td>
<td>876.5</td>
<td>0.18</td>
<td>nMnMr</td>
<td>A61</td>
<td>B431</td>
<td>9.87</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>unassigned</td>
</tr>
<tr>
<td>13</td>
<td>1046.4</td>
<td>1046.54</td>
<td>0.16</td>
<td>1374</td>
<td>1.05</td>
<td>TQDVSQKRR</td>
<td>MONO</td>
<td>B637</td>
<td>8.26</td>
<td>0.78</td>
<td>2</td>
<td>0</td>
<td>unassigned</td>
</tr>
<tr>
<td>14</td>
<td>1046.4</td>
<td>1046.54</td>
<td>0.16</td>
<td>1374</td>
<td>1.05</td>
<td>nMnMr</td>
<td>A61</td>
<td>B431</td>
<td>8.26</td>
<td>0.78</td>
<td>2</td>
<td>0</td>
<td>unassigned</td>
</tr>
<tr>
<td>15</td>
<td>1058.35</td>
<td>1059.54</td>
<td>0.11</td>
<td>1005</td>
<td>7.50</td>
<td>TQDVSQKRR</td>
<td>MONO</td>
<td>B637</td>
<td>3.39</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>unassigned</td>
</tr>
<tr>
<td>16</td>
<td>1090.36</td>
<td>1091.61</td>
<td>0.02</td>
<td>1262</td>
<td>1.16</td>
<td>LK6KGF</td>
<td>nMnMr</td>
<td>B431</td>
<td>9.34</td>
<td>0.55</td>
<td>2</td>
<td>0</td>
<td>unassigned</td>
</tr>
<tr>
<td>17</td>
<td>1112.33</td>
<td>1113.66</td>
<td>0.13</td>
<td>1165</td>
<td>1.17</td>
<td>IA61K6RD</td>
<td>MONO</td>
<td>B431</td>
<td>4.14</td>
<td>0.55</td>
<td>2</td>
<td>0</td>
<td>unassigned</td>
</tr>
<tr>
<td>18</td>
<td>1117.43</td>
<td>1117.6</td>
<td>0.01</td>
<td>152</td>
<td>0.01</td>
<td>PGKATMRGK</td>
<td>nMnMr</td>
<td>B431</td>
<td>16.04</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>unassigned</td>
</tr>
<tr>
<td>19</td>
<td>1202.54</td>
<td>1202.64</td>
<td>0.01</td>
<td>836</td>
<td>1.05</td>
<td>TQDVSQKRR</td>
<td>MONO</td>
<td>B637</td>
<td>22.46</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>correct</td>
</tr>
<tr>
<td>20</td>
<td>1218.51</td>
<td>1218.72</td>
<td>0.12</td>
<td>153</td>
<td>0.12</td>
<td>nMnMr</td>
<td>A61</td>
<td>B431</td>
<td>15.34</td>
<td>0.42</td>
<td>3</td>
<td>0</td>
<td>unassigned</td>
</tr>
<tr>
<td>21</td>
<td>1218.51</td>
<td>1218.72</td>
<td>0.12</td>
<td>153</td>
<td>0.12</td>
<td>nMnMr</td>
<td>A61</td>
<td>B431</td>
<td>15.34</td>
<td>0.42</td>
<td>3</td>
<td>0</td>
<td>unassigned</td>
</tr>
<tr>
<td>22</td>
<td>1333.46</td>
<td>1333.71</td>
<td>0.16</td>
<td>594.6</td>
<td>0.28</td>
<td>A61</td>
<td>B431</td>
<td>5.45</td>
<td>0.62</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>unassigned</td>
</tr>
<tr>
<td>23</td>
<td>1333.46</td>
<td>1333.71</td>
<td>0.16</td>
<td>594.6</td>
<td>0.28</td>
<td>nMnMr</td>
<td>A61</td>
<td>B431</td>
<td>5.45</td>
<td>0.62</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

**Matched Peptides**

**“Score”**

**Error [Da/ppm]**

**Locations of cross-linked Residues in Protein-Complex**

**User-Validation**
Final Results – 3D Structure (1ujz.pdb) – two cross-linking reagents

Protein cross-linking protocol was developed using isotope-labeling and LC-MALDI mass spectrometry.

Three integrative software tools were developed to analyze protein cross-linking MS and MS/MS data: iXLINK, doXLINK and XLinkViewer.

All software is available for download from the SPC website.

23 cross-linked amino acid residues were identified in the Colicin E7 DNAse/Im7 Protein complex (with d\textsubscript{0/12}-DSS and d\textsubscript{0/d6}-DSG cross-linking reagents).

All lysine residues within < 22.1 Å distance of each other → distance constraints (as expected from the cross-linker chain length).

5-10 ug of protein was used per experiment.

Compatible with various common isotope-coded bis-NHS ester cross-linking reagents (readily synthesized, or i.e. [d\textsubscript{0/d4}]-BS\textsuperscript{2}G and BS\textsuperscript{3} from Pierce).

Method has been validated with additional 5 single proteins (1 week analysis time).

Experiment + analysis can be automated.

Potential for high-throughput protein structure analysis.

Conclusions from this Study
References

Web-Links:
____________2006 ASMS poster
http://tools.proteomecenter.org/XLink.php  
_______download software + manual

Original Publication:
Seebacher, J., Mallick, P., Zhang, N., Eddes, J. S., Aebersold, R., Gelb, M. H.
"Protein Cross-linking Analysis Using Mass Spectrometry, Isotope-Coded Cross-linkers, and Integrative Computational Data Processing"

Heavy and light crosslinker pairs developed for MS applications.

http://www.piercenet.com
Acknowledgements

- Parag Mallick (Cedars-Sinai)
- Ning Zhang
- James S. Eddes
- Nichole King
- Richard Bonneau (NYU)
- Simon Letarte
- Sheng Pan
- Bernd Wollscheid (IMSB)

- Lars Malmström
- David Baker
- Brian Smart (UCB)
- Michael Gelb

- Ruedi Aebersold

- NHLBI, SPC