ABRF PSRG 2016: C-terminal identification of standard proteins by $^{18}$O labeling and bottom-up mass spectrometry

Luo X, Wood DC, Cavey GS, Feild BJ, English RD, McGrath SC, Erdjument-Bromage H
Background and objective

» Strategies for identification of protein C-terminus is desired by Core Facilities
  > Identify truncations or fusion proteins

» Incorporation of $^{18}\text{O}$ water in digestion aids determination of protein C-terminal sequence

» Approach is simple and easily adapted to a Core Laboratory operation
  > Does not require complicated chemistries
  > Fits well into routine bottom-up mass spectrometry
  > Reproducible, easy to perform, sensitive, and robust
Labeling technique: proteins are enzymatically digested in the presence of $^{18}$O water

- All specifically cleaved internal peptides will exchange $^{18}$O tag
- Peptide originating from the carboxy-terminus of the protein will NOT exchange $^{18}$O tag

Background and objective

This year’s study: Identify the C-terminus of known proteins using $^{18}$O labeling

> Myoglobin
> β-Lactoglobulin (BLG)

1. Digest protein in absence and presence of $^{18}$O water (participants’ choice of protocol)
2. Identify the $^{16}$O/$^{18}$O pairs of internal peptide fragments by bottom-up mass spectrometry
3. Report the singlet ($^{16}$O) C-terminal sequence
4. Evaluate the lowest amount of protein required to identify the protein C-terminal peptide
β-lactoglobulin digest, MALDI-TOF/TOF full scan

16O digest, all peptides appear as singlet C-terminal LSFNPTQLEEQCHI

16O/18O digest, many ion pairs observed LSFNPTQLEEQCHI still singlet
PSRG 2016 Pilot Study

Workflow A: In-Solution Digest

TPEVDDEALEKFDK (141-154)

Mass (m/z)

1635.62

LSFNPTQLEEQCHI (165-178 C-terminal)

Mass (m/z)

1658.68

16O

16O/18O
BLG C-terminal LSNFNTQLEEQCHI verified by MS/MS and database search
Total ion Chromatogram of all detected Ions (MS and MS/MS)

Control Digest, $^{16}$O water

Digest in $^{16}$O/$^{18}$O Water
**Typical Mascot MS/MS database search result:**

<table>
<thead>
<tr>
<th>Start - End</th>
<th>Observed</th>
<th>Mr(expt)</th>
<th>Mr(calc)</th>
<th>ppm</th>
<th>M Score</th>
<th>Expect</th>
<th>Rank</th>
<th>U</th>
<th>Peptide</th>
</tr>
</thead>
<tbody>
<tr>
<td>141 - 151</td>
<td>623.2897</td>
<td>1244.5649</td>
<td>1244.5772</td>
<td>-9.88</td>
<td>0</td>
<td>71</td>
<td>3.9e-005</td>
<td>1</td>
<td>R.TPEVDDEALEK.F</td>
</tr>
<tr>
<td>141 - 151</td>
<td>624.2925</td>
<td>1246.5704</td>
<td>1246.5815</td>
<td>-8.86</td>
<td>0</td>
<td>50</td>
<td>0.00057</td>
<td>1</td>
<td>R.TPEVDDEALEK.F + Label:180(1) (C-term)</td>
</tr>
<tr>
<td>141 - 151</td>
<td>624.2927</td>
<td>1246.5708</td>
<td>1246.5815</td>
<td>-8.57</td>
<td>0</td>
<td>66</td>
<td>1.8e-005</td>
<td>1</td>
<td>R.TPEVDDEALEK.F + Label:180(1) (C-term)</td>
</tr>
<tr>
<td>141 - 151</td>
<td>624.2931</td>
<td>1246.5716</td>
<td>1246.5815</td>
<td>-7.88</td>
<td>0</td>
<td>58</td>
<td>0.0009</td>
<td>1</td>
<td>R.TPEVDDEALEK.F + Label:180(1) (C-term)</td>
</tr>
<tr>
<td>141 - 151</td>
<td>624.2933</td>
<td>1246.5721</td>
<td>1246.5815</td>
<td>-7.49</td>
<td>0</td>
<td>66</td>
<td>3.9e-005</td>
<td>1</td>
<td>R.TPEVDDEALEK.F + Label:180(1) (C-term)</td>
</tr>
<tr>
<td>141 - 151</td>
<td>624.2945</td>
<td>1246.5744</td>
<td>1246.5815</td>
<td>-5.63</td>
<td>0</td>
<td>52</td>
<td>0.00053</td>
<td>1</td>
<td>R.TPEVDDEALEK.F + Label:180(1) (C-term)</td>
</tr>
<tr>
<td>141 - 154</td>
<td>545.9246</td>
<td>1634.7520</td>
<td>1634.7675</td>
<td>-9.48</td>
<td>1</td>
<td>62</td>
<td>1.2e-005</td>
<td>1</td>
<td>R.TPEVDDEALEKFDK.A</td>
</tr>
<tr>
<td>141 - 154</td>
<td>545.9259</td>
<td>1634.7559</td>
<td>1634.7675</td>
<td>-7.13</td>
<td>1</td>
<td>30</td>
<td>0.058</td>
<td>1</td>
<td>R.TPEVDDEALEKFDK.A</td>
</tr>
<tr>
<td>141 - 154</td>
<td>545.9260</td>
<td>1634.7563</td>
<td>1634.7675</td>
<td>-6.89</td>
<td>1</td>
<td>25</td>
<td>0.039</td>
<td>1</td>
<td>R.TPEVDDEALEKFDK.A</td>
</tr>
<tr>
<td>141 - 154</td>
<td>819.3851</td>
<td>1636.7557</td>
<td>1636.7718</td>
<td>-9.82</td>
<td>1</td>
<td>82</td>
<td>5.4e-007</td>
<td>1</td>
<td>R.TPEVDDEALEKFDK.A + Label:180(1) (C-term)</td>
</tr>
<tr>
<td>165 - 178</td>
<td>858.4057</td>
<td>1714.7968</td>
<td>1714.7985</td>
<td>-0.94</td>
<td>0</td>
<td>79</td>
<td>0.0016</td>
<td>1</td>
<td>R.LSFHPTQLEEQCHI.- + Carbamidomethyl (C)</td>
</tr>
</tbody>
</table>

Internal peptides identified with and without $^{18}$O label

C-terminal peptides only appear without $^{18}$O label pair
Visual inspection of MS1 data

Two tryptic peptides from B-lactoglobulin

TPEVDDEALEKFDK
1634.752

LSFNPTQLEEQCHI
1714.798

NO doublet incorporating $^{18}$O
MaxQuant Analysis of Peptides

- Average H/L ratio of internal peptides generated in presence of $^{18}$O water: $^{18}$O:$^{16}$O = 1.06
- C-terminal peptide $^{18}$O:$^{16}$O = ratio low

<table>
<thead>
<tr>
<th>Protein Identified: sp P02754 LACB_BOVIN Beta-lactoglobulin</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sequence</strong></td>
</tr>
<tr>
<td>TKIPAVFK</td>
</tr>
<tr>
<td>IDALNENK</td>
</tr>
<tr>
<td>LIVTQTMK</td>
</tr>
<tr>
<td>ALKALPMHJR</td>
</tr>
<tr>
<td>VLVLDTDYKK</td>
</tr>
<tr>
<td>TPEVDDEALEK</td>
</tr>
<tr>
<td>TPEVDDEALEKFDK</td>
</tr>
<tr>
<td>LSFNPTQLEEQCHI</td>
</tr>
<tr>
<td>IDALNENKVLVLDTDYKK</td>
</tr>
<tr>
<td>VYVEELKPTPEGDLEILLQK</td>
</tr>
</tbody>
</table>
## Study design

<table>
<thead>
<tr>
<th>Workflow (A)</th>
<th>Workflow (B)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>In-solution digestion</strong></td>
<td><strong>In-gel digestion</strong></td>
</tr>
<tr>
<td>Solution digestion of proteins in absence and presence of O18 water</td>
<td>SDS-PAGE separation of proteins</td>
</tr>
<tr>
<td>Peptide Cleanup (C18 columns or similar)</td>
<td>In-gel digestion in absence and presence of O18 water</td>
</tr>
<tr>
<td></td>
<td>Peptide Cleanup (C18 columns or similar)</td>
</tr>
<tr>
<td></td>
<td>Inspection of internal pairs of O16/O18 peptides</td>
</tr>
<tr>
<td></td>
<td>MS analysis including data analysis</td>
</tr>
<tr>
<td></td>
<td>Identification of singlet C-terminal peptide</td>
</tr>
</tbody>
</table>
Study Demographics

» 15 laboratories requested samples
  > 4/15 international sites
» 9 participants returned data
» Most were return participants

Do you have any experience with C-term protein identification?

- Yes, a lot: 4
- Yes, a little: 0
- No: 5

Did you participate in EITHER N-term labeling study?

- Yes, 2014: 6
- Yes, 2015: 5
- No: 2
Many vendors and instrument platforms were represented in the study.

What instrumentation did you use?

- AB Sciex 5600 TripleTOF: 2
- AB Sciex 5800 MALDI-TOF TOF MS: 1
- Axima Resonance MALDI-QIT-TOF: 1
- Orbitrap: 5
- Q Exactive: 1

Which workflow did you use:
- WF-A: 7
- WF-B: 2
- Both: 0
Participant Data: Lab 31U

Workflow A: In-Solution Digest

<table>
<thead>
<tr>
<th></th>
<th>Correct ID?</th>
<th>Sequence coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Myoglobin</td>
<td>✓</td>
<td>92%</td>
</tr>
<tr>
<td>β-lactoglobulin</td>
<td>✓</td>
<td>76%</td>
</tr>
</tbody>
</table>

>sp|P02754|LACB_BOVIN Beta-lactoglobulin-136/178 amino acids (76% coverage)

MKCLLALALTCGAQALIVTQT<KGLDIQKVAGTWYSLAAASDSSLDAQSAPLRVVEELKPTPEGDEIIKLQKENGEAQKIIIAEKTKIPAVFKIDALNENKVLVLDYKHLLFCMENSAEPEQSLACQCLVRTEVDEALEKFDKALKALP|HIRLSFNPTQLEEOH|

>sp|P68083|MYG_EQUBU Myoglobin-141/154 amino acids (92% coverage)

MGLSDGEWQQVLYNGWVKVEADIAGHGQEVGLIRFLGPHPETLEKFDFKHKLTEAEKASEDLKKHGTVVLTAALGGILKKGHHEAEELKPLAQSHATKHPIKYLEFISDAIIHVNLHSKHPDFGADAOGAKSALELFRNDIAAKYKELGFQCG
**Participant Data: Lab 31U**

**Internal Peptide = Doublet**

- **Myoglobin**
  - m/z values: 462, 463, 464, 465, 466, 467, 468, 469, 470, 471, 472, 473, 474, 475, 476, 477, 478
  - Relative Abundance values: 0, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100

- **β-Lactoglobulin**
  - m/z values: 855, 856, 857, 858, 859, 860, 861, 862, 863, 864, 865
  - Relative Abundance values: 0, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100

**C-Terminal Peptide = Singlet**

- **Myoglobin**
  - m/z values: 501.566, 502.560, 502.89, 503.23, 503.56
  - Relative Abundance values: 0, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100

- **β-Lactoglobulin**
  - m/z values: 471.23984, 471.74084, 472.24207, 472.74353
  - Relative Abundance values: 0, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100

**Workflow A:**

In-Solution Digest
Participant Data: Lab 31U

C-Terminal Peptide Example: Myoglobin – Correctly Identified YKELGFQG

Workflow A: In-Solution Digest

- Workflow A: In-Solution Digest
C-Terminal Peptide Example: BLG – Correctly Identified LSFNPTQLEEIQCHI

Workflow A: In-Solution Digest
MSMS of 2313.3 Da from β-Lactoglobulin (1.5 pmol on spot)

18O isotopic pattern for an internal peptide

Internal Peptide 41-60
VYVEELKPTPEGDLEILLQK
Participant Data: Lab 37C

MSMS of 1716.8 Da from β-Lactoglobulin (1.5 pmol on spot)

Typical isotopic pattern for a C-terminal peptide

C-term Peptide 149-162
LSFNPTQLEEQCHI

Workflow B: In-Gel Digest
Lessons from Study Design

» Several participants performed the technique well

» Many participants returned results without data demonstrating use of technique
  > Protein coverage, match scores not sufficient
  > $^{16}$O/$^{18}$O internal peptide pairs should be shown

» Unclear if participants could effectively use $^{18}$O labeling with an unknown/modified protein
PSRG goal: help Cores estimate the quantity of protein needed from investigators

Sensitivity correlates with experience level of participants

- More than 5 picomoles: 3
- 5 picomoles: 1 (Lactoglobulin), 1 (Myoglobin)
- 1 picomole: 1
- Less than 1 picomole: 5 (Lactoglobulin), 4 (Myoglobin)

Proteins: Lactoglobulin, Myoglobin
Summary: Technique

- PSRG pilot study demonstrated that the technique:
  - Is easy to use
  - Requires minimum derivatization/chemistry steps
  - Does not yield side reactions
  - Is sensitive (sub-picomole of protein)
  - Does not interfere with downstream bottom-up MS

- Isotopic envelopes of tryptic peptides
  - Observed ~50% incorporation of $^{18}$O for internal peptides
  - C-terminal peptide was observed only as a singlet

- Success is limited by the generation and recovery of the C-terminal-most peptide
  - Similar to N-terminal peptide ID using bottom-up MS
Summary: Lessons

» This technique relies on inspection of MS1 data
  > Verify that $^{16}\text{O}/^{18}\text{O}$ internal peptide pairs present
  > Look for internal peptide preceding putative C-term peptide

» Many participants used typical proteomics/database search pipeline
  > May not return correct result
  > Especially critical for fusion proteins or truncations

» Different instruments, software can be beneficial
  > MALDI-TOF
    + Good for identification of doublet pairs
    + Difficulty with identification of myoglobin C-term
  > Software for quantitative analysis of isotopes (MaxQuant)
1. Digest and analyze a known or unknown (truncated) protein
   > Identify the protein
   > Identify the C-terminal-most peptide

2. Verify that the C-terminal peptide is present in MS spectra
   > Manual inspection or database searching
   > Suggestion for calculation software?

3. Digest and analyze a 2\textsuperscript{nd} aliquot of protein in buffer with 50\% $^{18}$O water
   > Manually inspect chromatogram for the C-terminal peptide candidate found in step 1 (at same retention time)
   > If this peptide is observed as a singlet without $^{18}$O label it can be implicated as the C-terminal peptide of the protein.

4. Consider an alternative protease
   > If the C-terminal amino acid is lysine or arginine
   > If the C-terminal peptide is too large or does not ionize well
Suggestions for Future Studies

» Sample preparation techniques to improve protein detection at low quantity
  > Minimize loss of material
  > Desalting/removal of excess reagent

» Disulfide mapping in proteins

» Quantitative glycoproteomics
  > Possible joint study with gPRG

» Please send us your suggestions
Please Join Us!

» We are always looking for new PSRG members!

» If you have interest in protein sequencing, and skills with either Edman or mass spectrometry, please contact one of our current members

» Robert English – Shimadzu Scientific Instruments
» Sara McGrath – FDA Center for Food Safety and Applied Nutrition
» Greg Cavey – Launch MI Lab, Southwest Michigan Innovation Center
» Hediye Erdjument-Bromage – NYU School of Medicine
» Xuemei Luo – University of Texas Medical Branch
» David Wood – St. Louis University
» Brian Field – Shimadzu Scientific Instruments
Acknowledgments

» Sponsors of study proteins and reagents:
  
  ABRF
  
  SIGMA-ALDRICH

» Anonymizer/Sample Coordinator:
  
  Sara McGrath, FDA/Center for Food Safety and Applied Nutrition

» ........and study participants!!!!!!


