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

Protein Sequencing Research Group

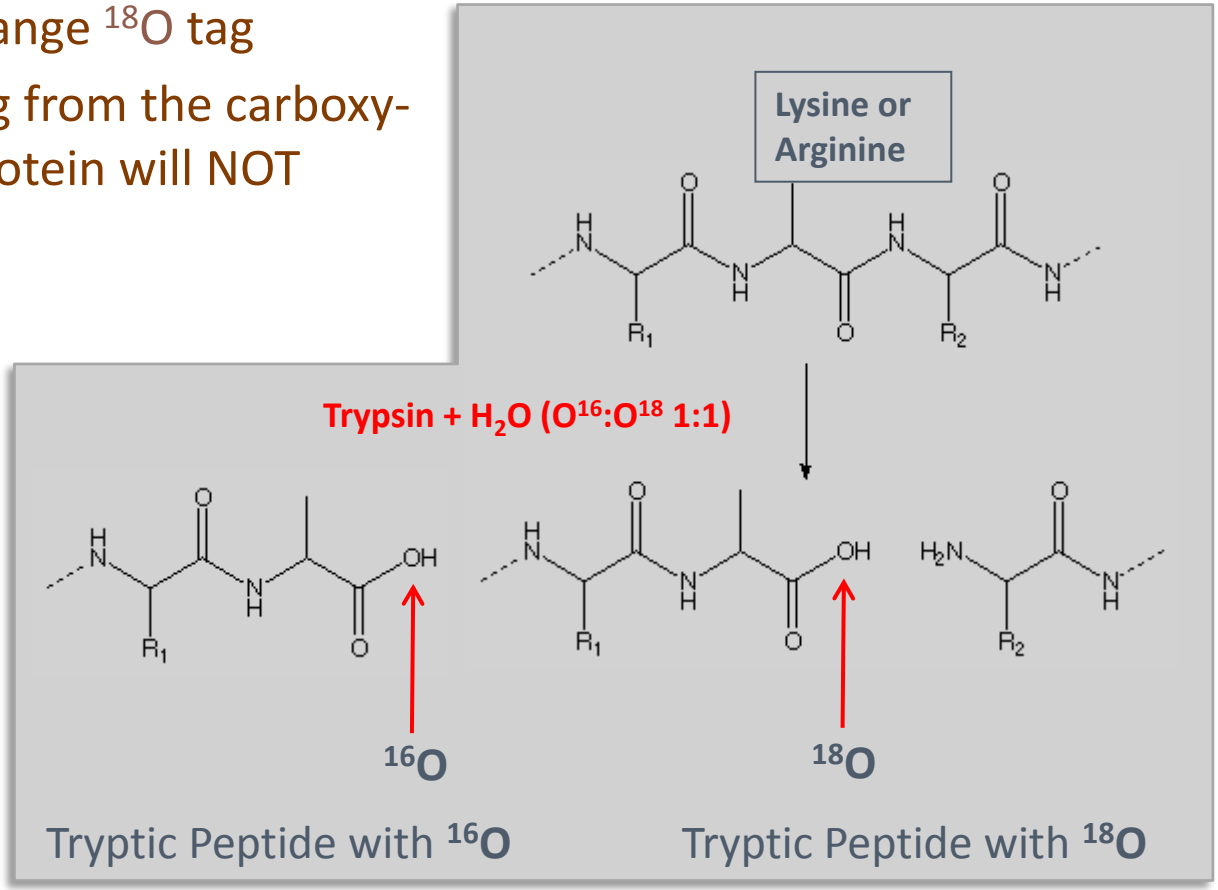
ABRF 2016 annual meeting

Background and objective

- Strategies for identification of protein C-terminus is desired by Core Facilities
  - Identify truncations or fusion proteins
- Incorporation of ^{18}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

Background and objective

- 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



Rose K, Simona M G, Offord R E, Prior C P, Otto B, Thatcher D R. *Biochem J*. **1983**;215:273–277

Background and objective

- This year's study: Identify the C-terminus of known proteins using ^{18}O labeling
  - Myoglobin
  - β-Lactoglobulin (BLG)

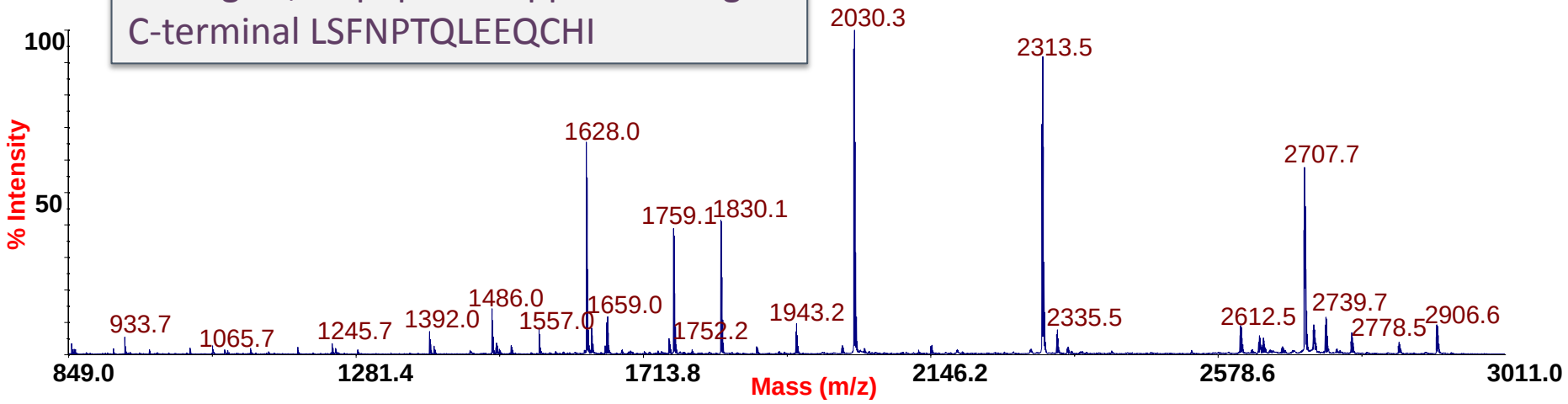
1. Digest protein in absence and presence of ^{18}O water (participants' choice of protocol)
2. Identify the $^{16}O/^{18}O$ pairs of internal peptide fragments by bottom-up mass spectrometry
3. Report the singlet (^{16}O) C-terminal sequence
4. Evaluate the lowest amount of protein required to identify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

PSRG 2016 Pilot Study

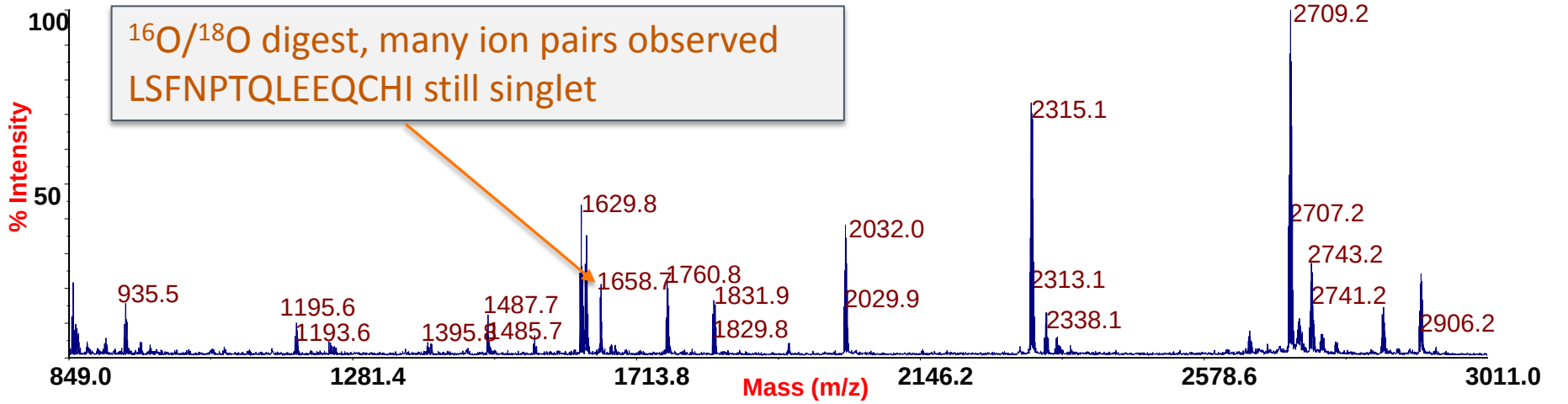
Workflow A: In-Solution Digest

β-lactoglobulin digest, MALDI-TOF/TOF full scan

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



$^{16}O/^{18}O$ digest, many ion pairs observed
LSFNPTQLEEQCHI still singlet

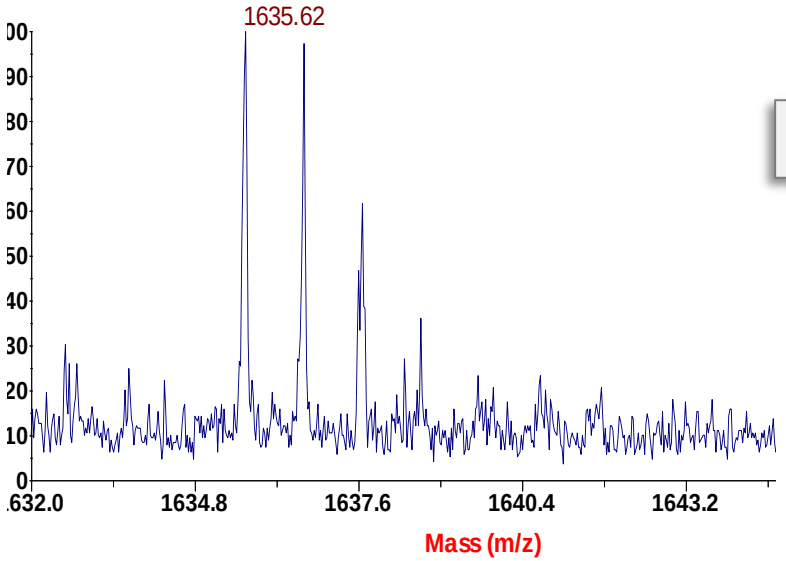


PSRG 2016 Pilot Study

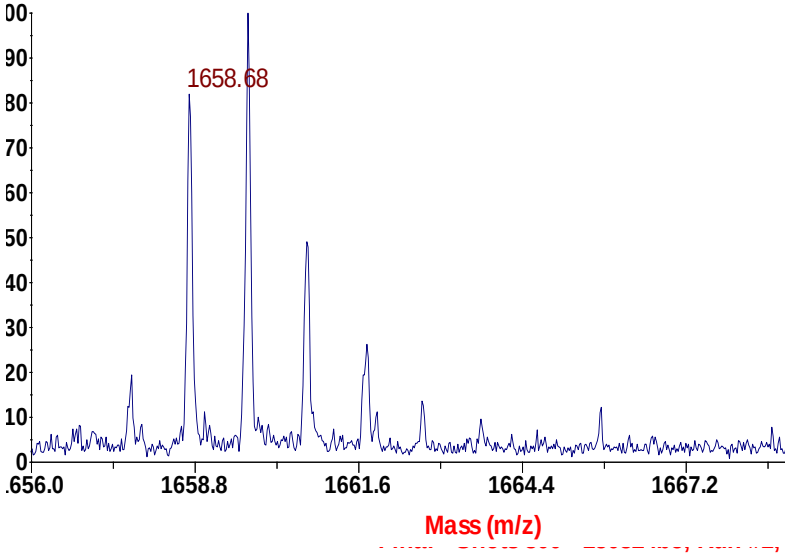
Workflow A: In-Solution Digest

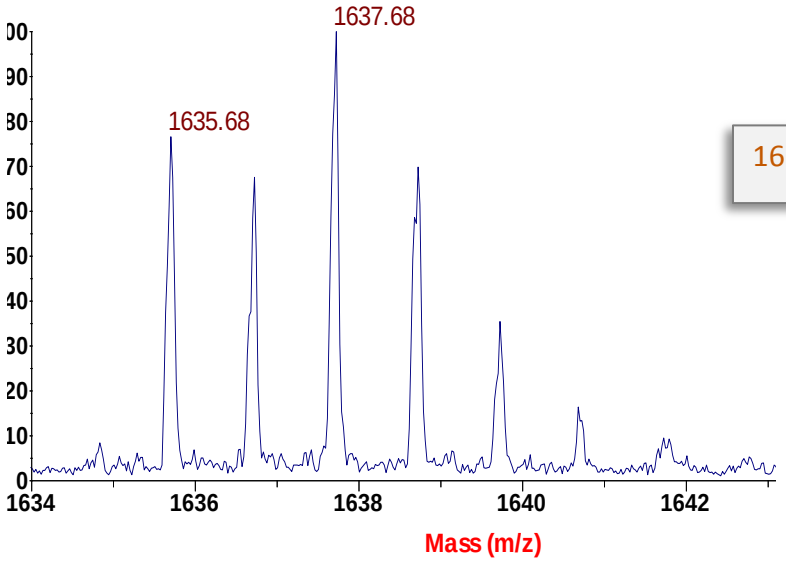
TPEVDDEALEKFDK (141-154)

LSFNPTQLEEQCHI (165-178 C-terminal)

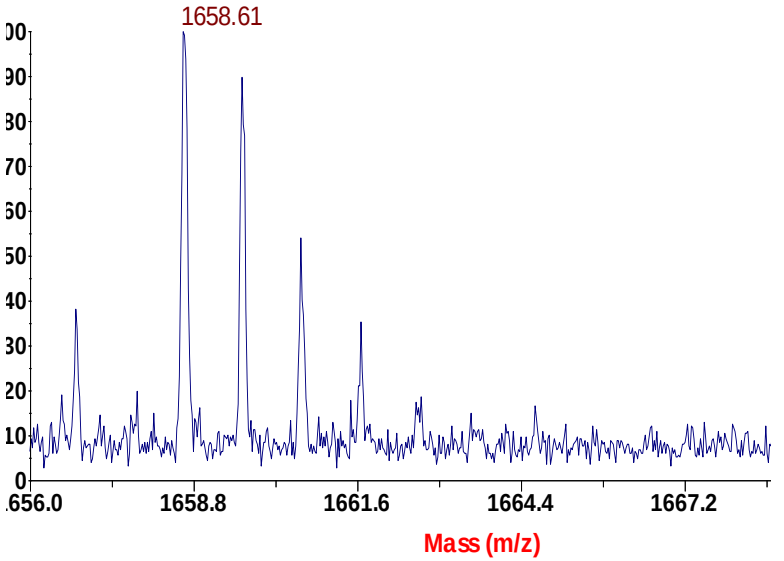


^{16}O





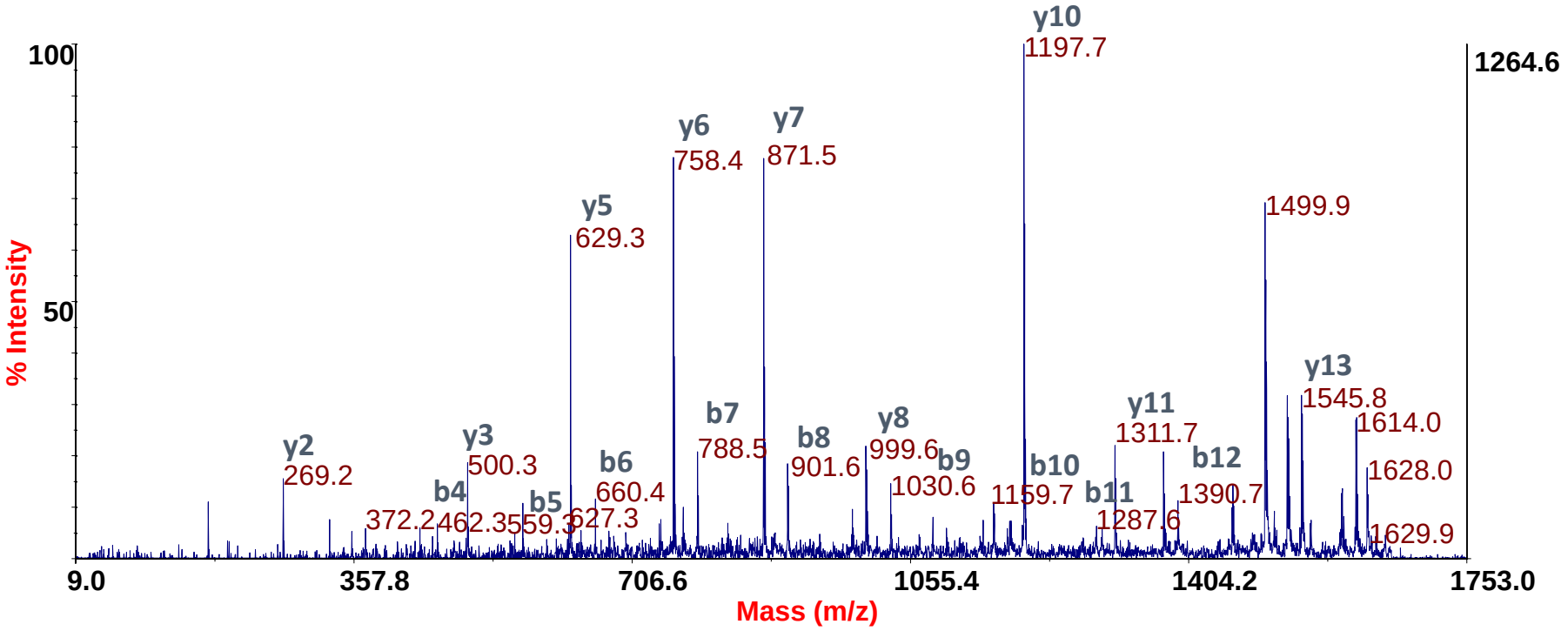
$^{16}O/^{18}O$



PSRG 2016 Pilot Study

Workflow A: In-Solution Digest

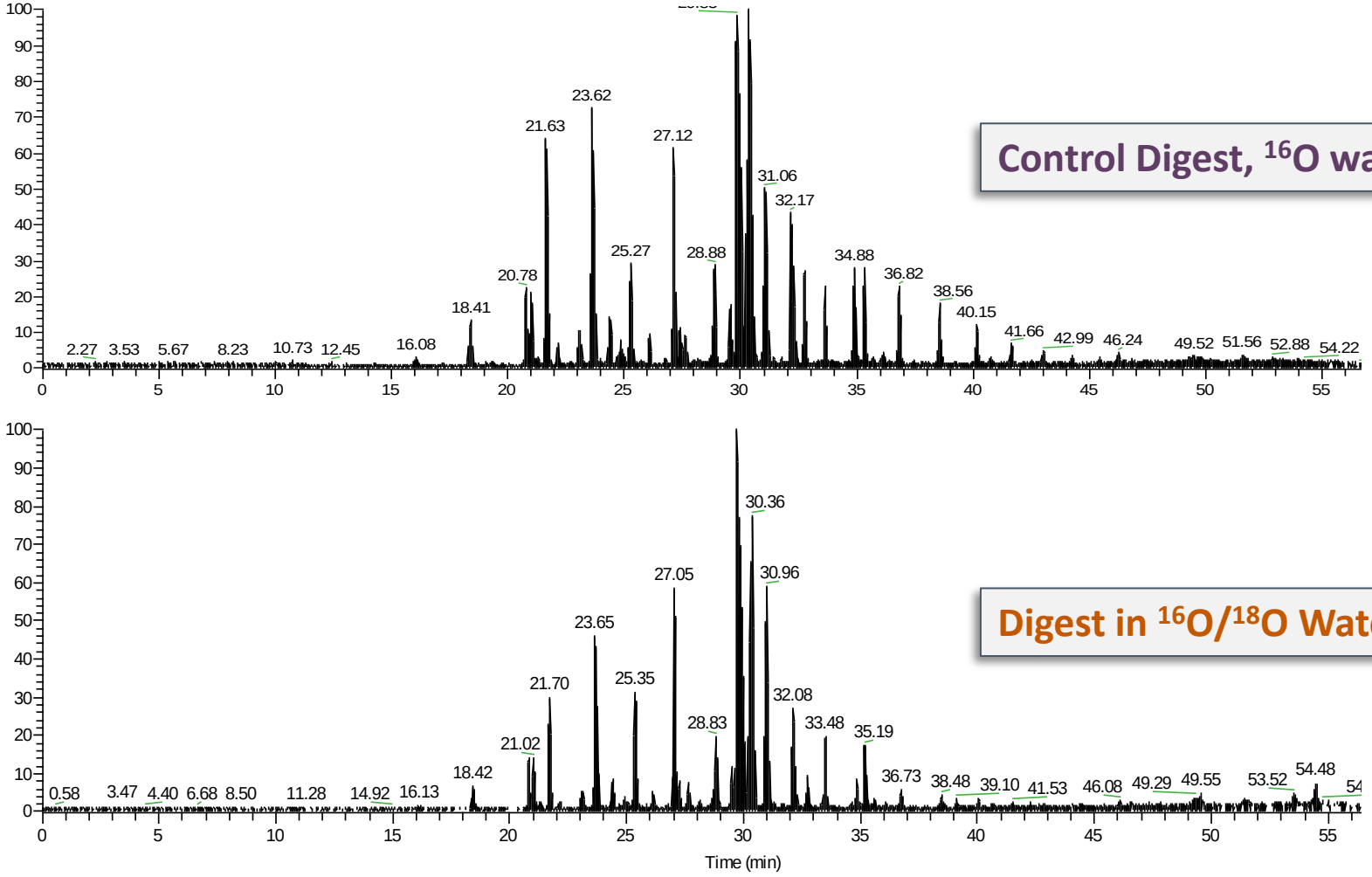
BLG C-terminal LSFNPTQLEEQCHI verified by MS/MS and database search



PSRG 2016 Pilot Study

Workflow B: In-gel Digest

Total ion Chromatogram of all detected Ions (MS and MS/MS)



PSRG 2016 Pilot Study

Workflow B: In-gel Digest

Typical Mascot MS/MS database search result:

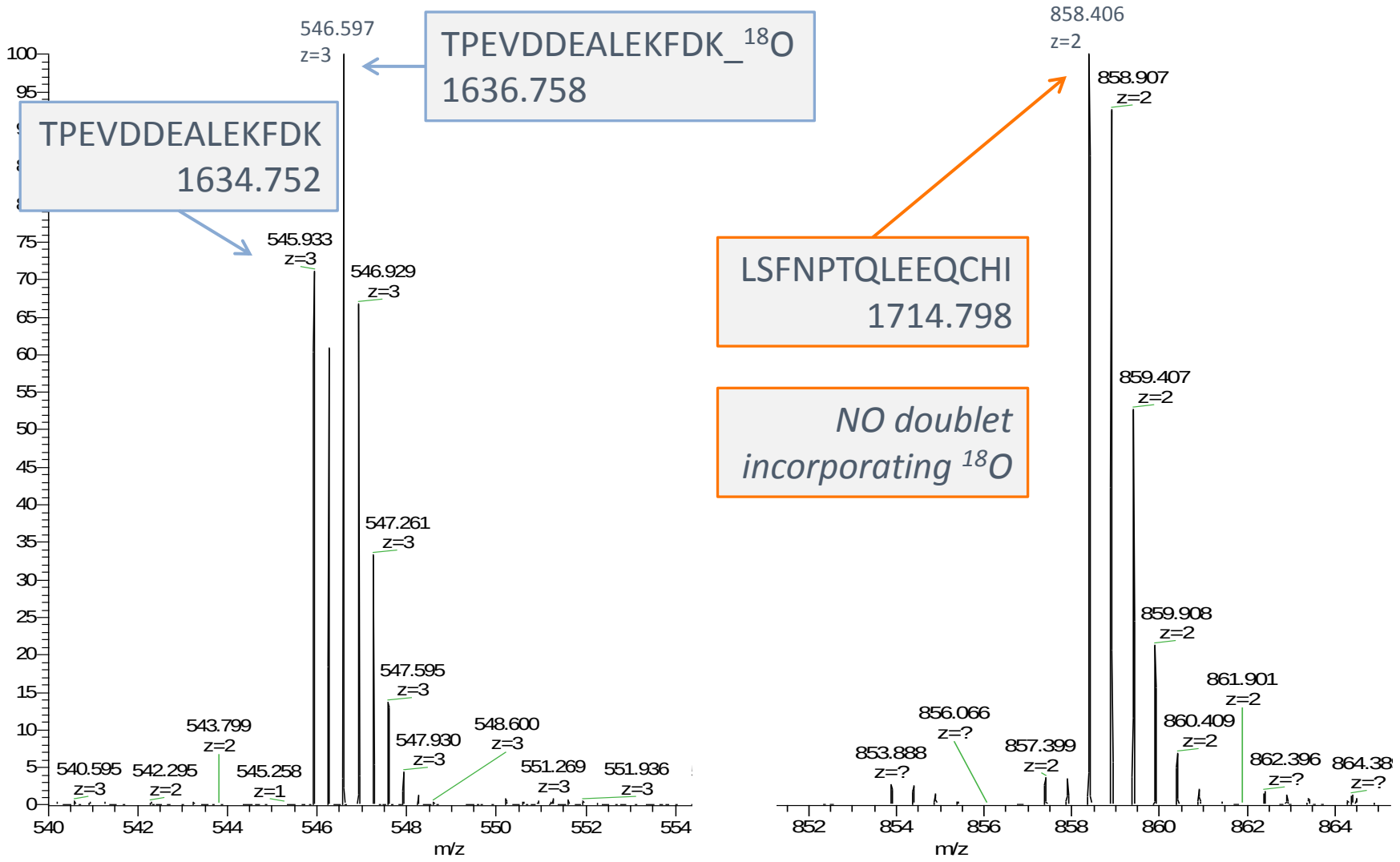
Start - End	Observed	Mr(expt)	Mr(calc)	ppm	M	Score	Expect	Rank	U	Peptide
141 - 151	623.2897	1244.5649	1244.5772	-9.88	0	71	3.9e-005	1		R.TPEVDDEALEK.F
141 - 151	624.2925	1246.5704	1246.5815	-8.86	0	50	0.00057	1		R.TPEVDDEALEK.F + Label:18O(1) (C-term)
141 - 151	624.2927	1246.5708	1246.5815	-8.57	0	66	1.8e-005	1		R.TPEVDDEALEK.F + Label:18O(1) (C-term)
141 - 151	624.2931	1246.5716	1246.5815	-7.88	0	58	0.0009	1		R.TPEVDDEALEK.F + Label:18O(1) (C-term)
141 - 151	624.2933	1246.5721	1246.5815	-7.49	0	66	3.9e-005	1		R.TPEVDDEALEK.F + Label:18O(1) (C-term)
141 - 151	624.2945	1246.5744	1246.5815	-5.63	0	52	0.00053	1		R.TPEVDDEALEK.F + Label:18O(1) (C-term)
141 - 154	545.9246	1634.7520	1634.7675	-9.48	1	62	1.2e-005	1		R.TPEVDDEALEKFDK.A
141 - 154	545.9259	1634.7559	1634.7675	-7.13	1	30	0.058	1		R.TPEVDDEALEKFDK.A
141 - 154	545.9260	1634.7563	1634.7675	-6.89	1	25	0.039	1		R.TPEVDDEALEKFDK.A
141 - 154	819.3851	1636.7557	1636.7718	-9.82	1	82	5.4e-007	1		R.TPEVDDEALEKFDK.A + Label:18O(1) (C-term)
165 - 178	858.4057	1714.7968	1714.7985	-0.94	0	79	0.0016	1	U	R.LSFNPTQLEEQCHI.- + Carbamidomethyl (C)

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



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

Protein Identified: sp P02754 LACB_BOVIN Beta-lactoglobulin

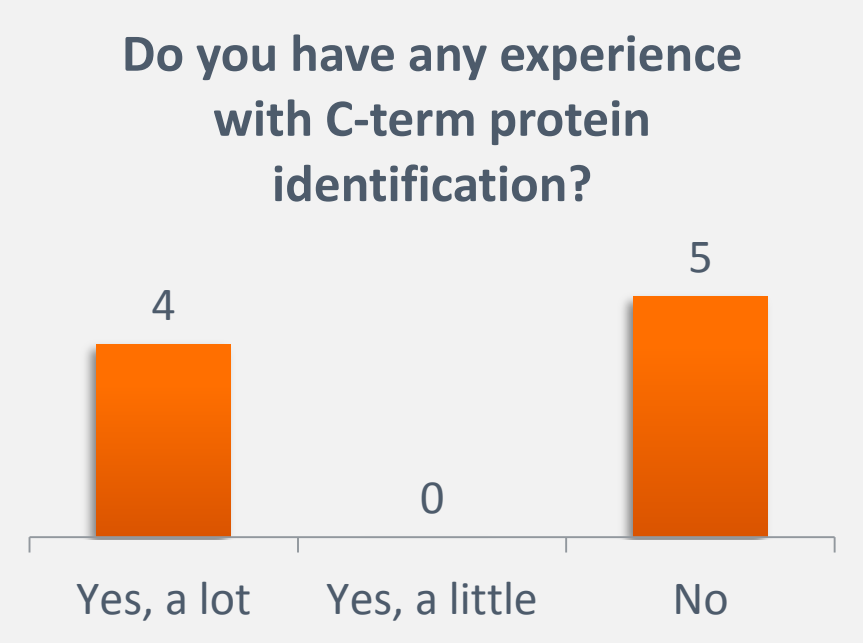
Sequence	Observed, m/z	Calculated Mr	Ratio O18/O16
TKIPAVFK	301.86	902.56	1.19
IDALNENK	458.74	915.47	0.82
LIVTQTMK	467.28	932.54	1.26
ALKALPMHIR	388.90	1148.69	0.94
VLVLDTDYKK	533.30	1192.67	0.97
TPEVDDEALEK	623.29	1244.58	0.96
TPEVDDEALEKFDK	545.93	1634.77	1.00
LSFNPTQLEEQCHI	858.41	1714.80	0.28
IDALNENKVLVLDTDYKK	697.72	2090.13	1.32
VYVEELKPTPEGDLEILLQK	771.43	2312.25	1.12

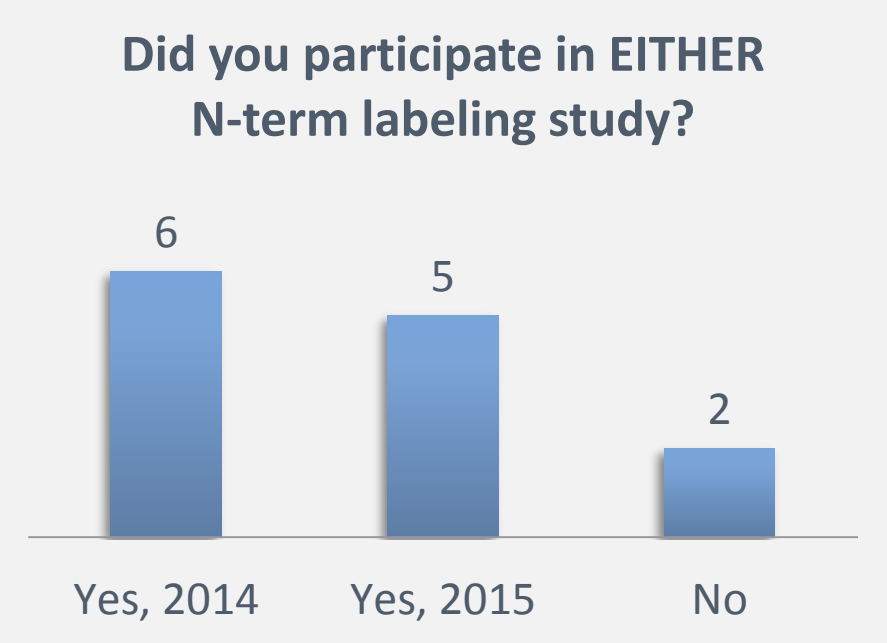
Study design

Workflow (A) In-solution digestion	Workflow (B) In-gel digestion
Solution digestion of proteins in absence and presence of O18 water	SDS-PAGE separation of proteins
Peptide Cleanup (C18 columns or similar)	In-gel digestion in absence and presence of O18 water Peptide Cleanup (C18 columns or similar)
Inspection of internal pairs of O16/O18 peptides	
MS analysis including data analysis	
Identification of singlet C-terminal peptide	

Study Demographics

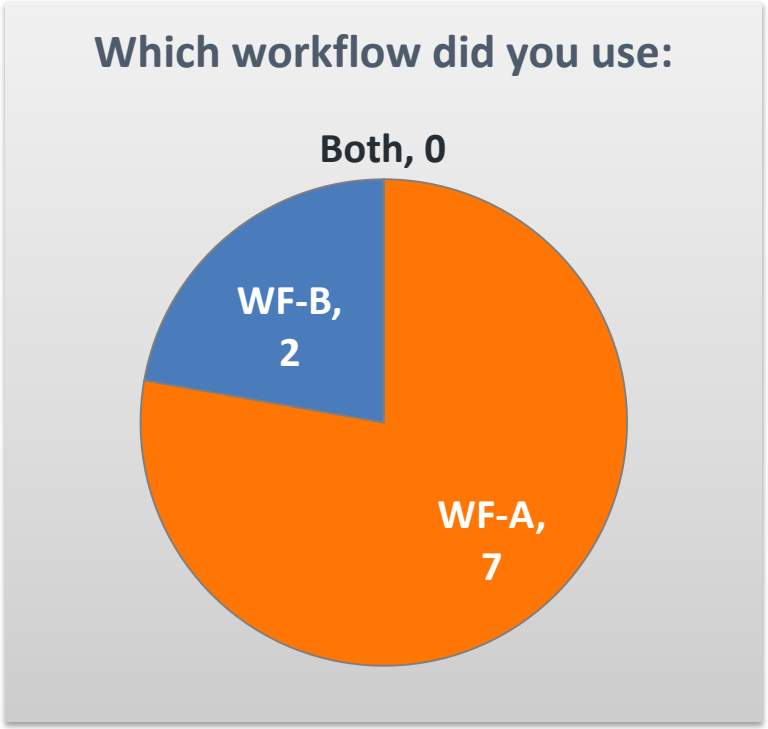
- 15 laboratories requested samples
  - 4/15 international sites
- 9 participants returned data
- Most were return participants



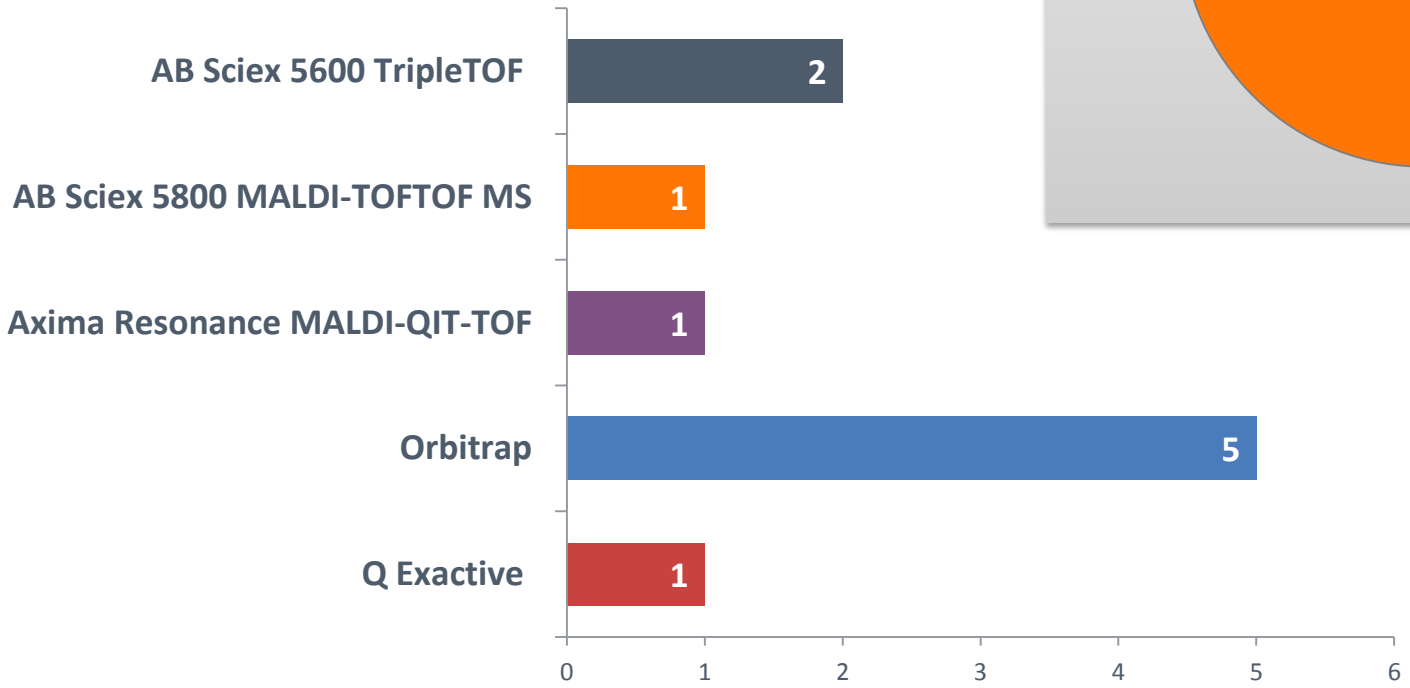


Study Demographics

- Many vendors and instrument platforms were represented in the study



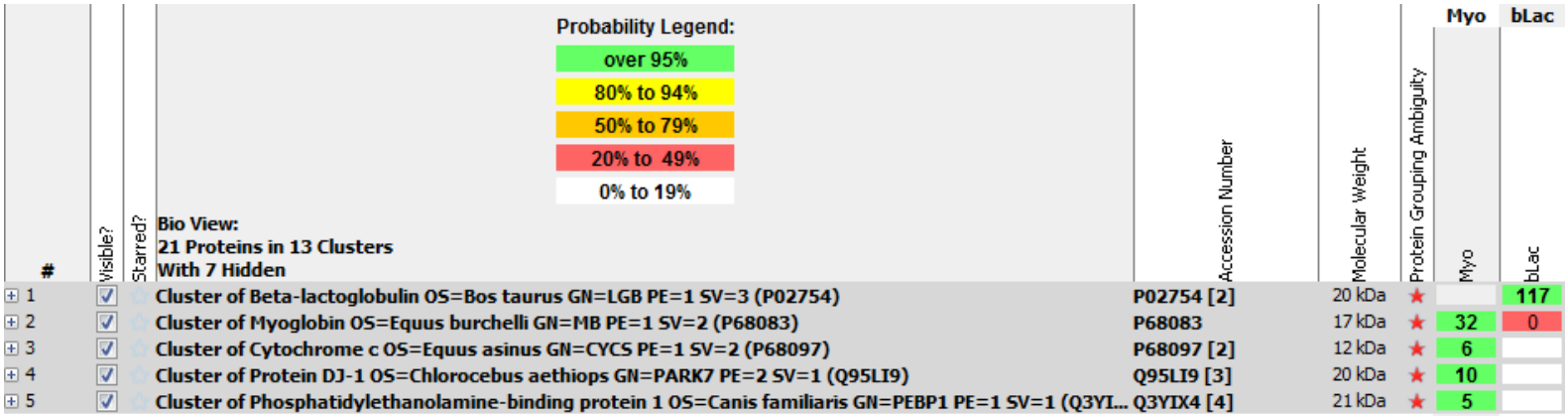
What instrumentation did you use?



Participant Data: Lab 31U

Workflow A: In-Solution Digest

	Correct ID?	Sequence coverage
Myoglobin	✓	92%
b-lactoglobulin	✓	76%



Probability Legend:
over 95%
80% to 94%
50% to 79%
20% to 49%
0% to 19%

Bio View:
21 Proteins in 13 Clusters
With 7 Hidden

#	Visible?	Starred?		Accession Number	Molecular Weight	Protein Grouping Ambiguity	Myo	bLac
							Myo	bLac
1	☑	☆	Cluster of Beta-lactoglobulin OS=Bos taurus GN=LGB PE=1 SV=3 (P02754)	P02754 [2]	20 kDa	★		117
2	☑	☆	Cluster of Myoglobin OS=Equus burchelli GN=MB PE=1 SV=2 (P68083)	P68083	17 kDa	★	32	0
3	☑	☆	Cluster of Cytochrome c OS=Equus asinus GN=CYCS PE=1 SV=2 (P68097)	P68097 [2]	12 kDa	★	6	
4	☑	☆	Cluster of Protein DJ-1 OS=Chlorocebus aethiops GN=PARK7 PE=2 SV=1 (Q95LI9)	Q95LI9 [3]	20 kDa	★	10	
5	☑	☆	Cluster of Phosphatidylethanolamine-binding protein 1 OS=Canis familiaris GN=PEBP1 PE=1 SV=1 (Q3YI...	Q3YIX4 [4]	21 kDa	★	5	

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

MKCLLLALALTCGAQALIVTQTMKGLDIQKVAGTWYSLAMAASDISLLDAQSAPLRVYVEELKPTPEGDLEILLQKWEN
GECAQKKIIAEKTKIPAVFKIDALNENKVLVLDTDYKKYLLFCMENSAEPEQSLACQCLVRTPEVDDEALEKFDKALKA
LPMHIRLSFNPTQLEEQCHI

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

MGLSDGEWQQVLNVWGKVEADIAGHGQEVLIRLFTGHPETLEKFDKFKHLKTEAEMKASEDLKKHGTVVLTALGGILKK
KGHHEAELKPLAQSHATKHKIPIKYLEFISDAIIHVLHSKHPGDFGADAQGAMTKALELFRNDIAAKYKELGFQG

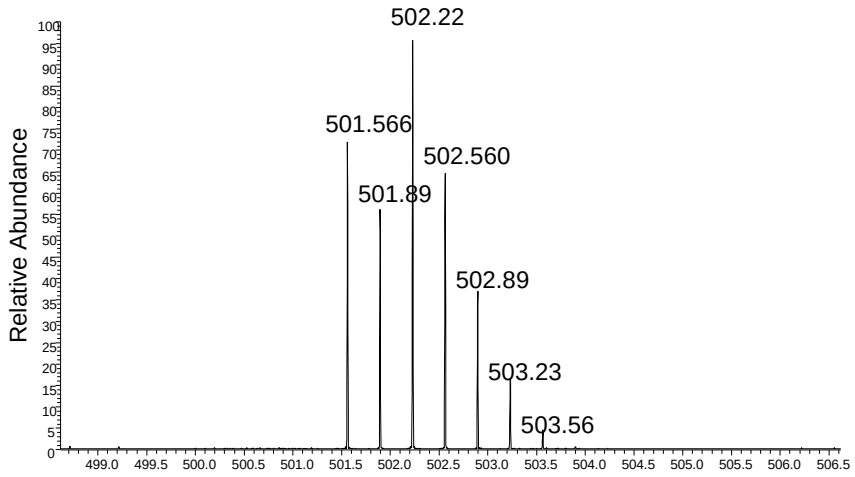
Participant Data: Lab 31U

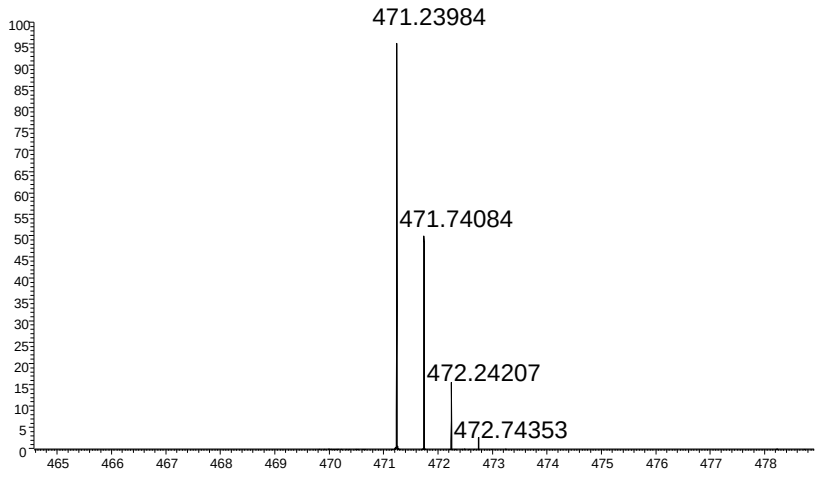
Workflow A: In-Solution Digest

Internal Peptide = Doublet

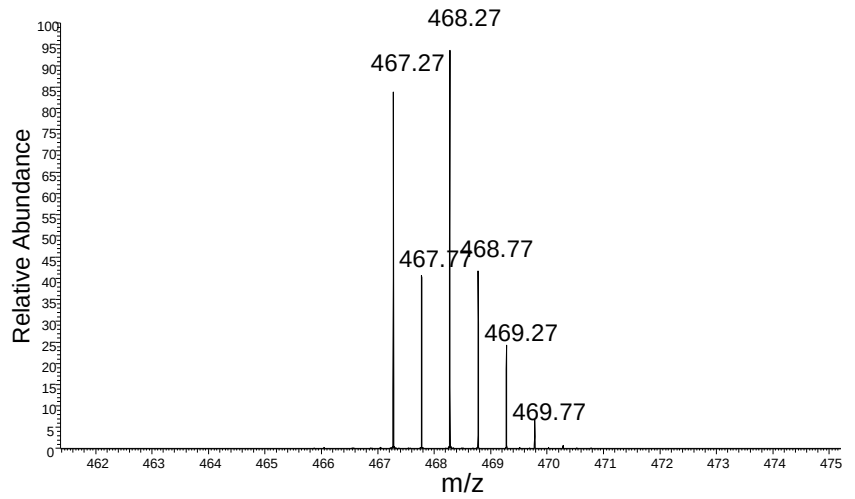
C-Terminal Peptide= Singlet

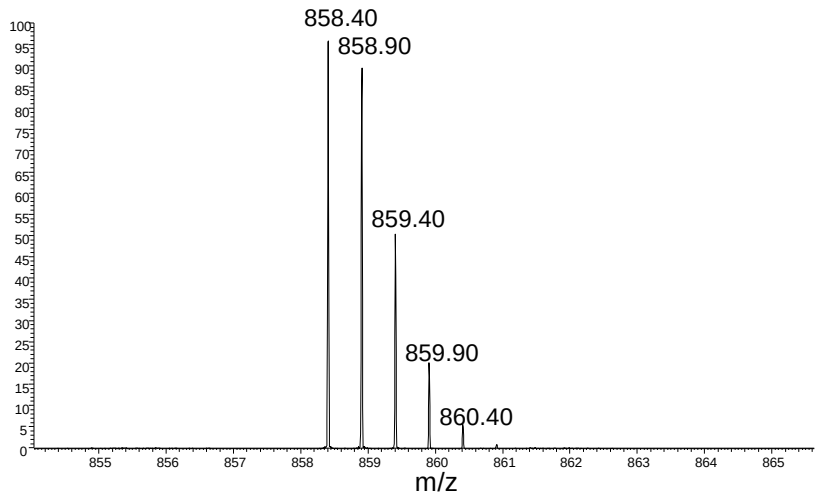
Myoglobin





β-Lactoglobulin

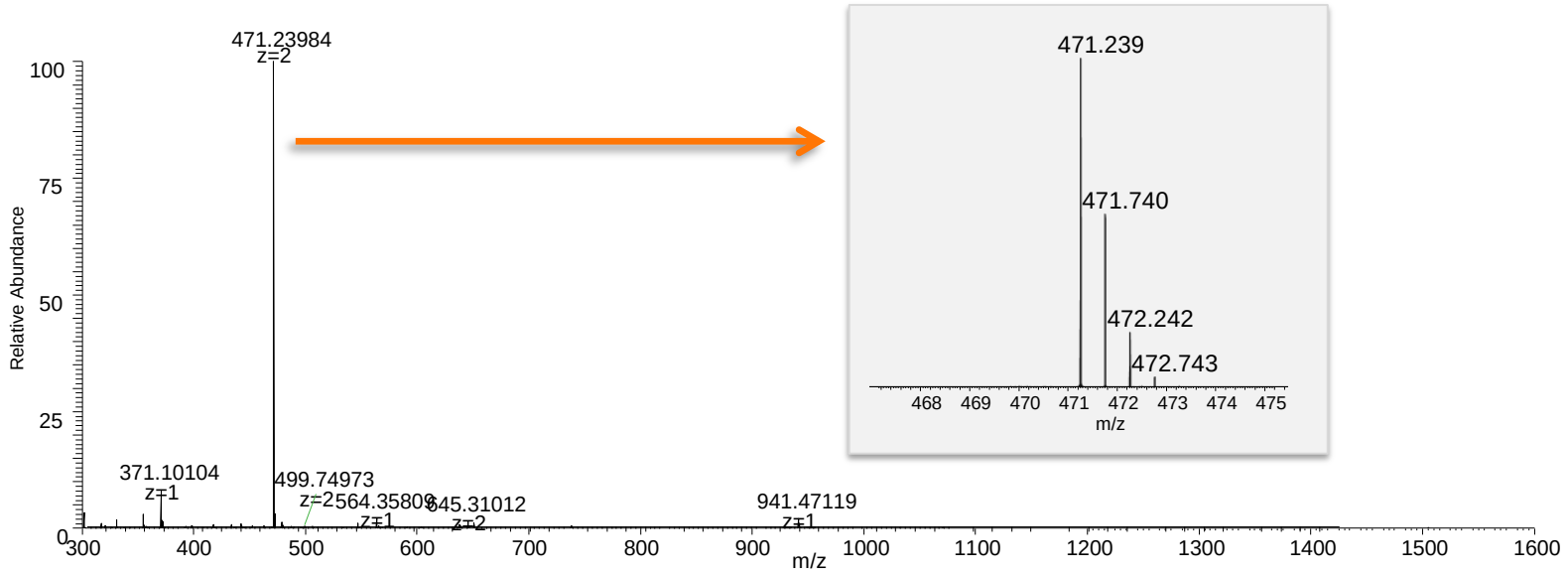


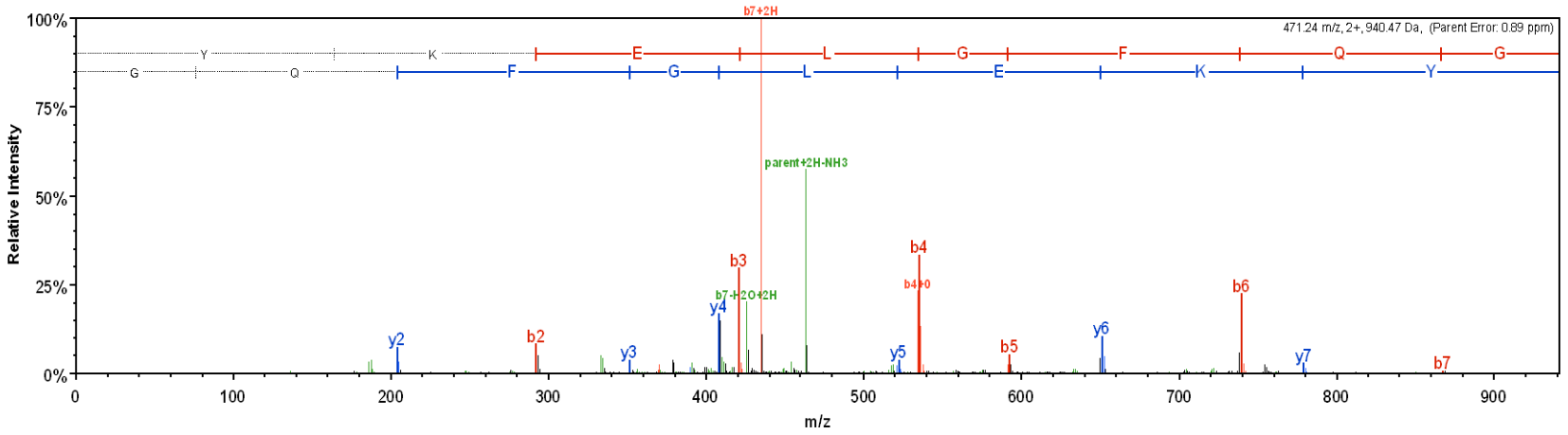


Participant Data: Lab 31U

Workflow A: In-Solution Digest

C-Terminal Peptide Example: Myoglobin – Correctly Identified YKELGFQG

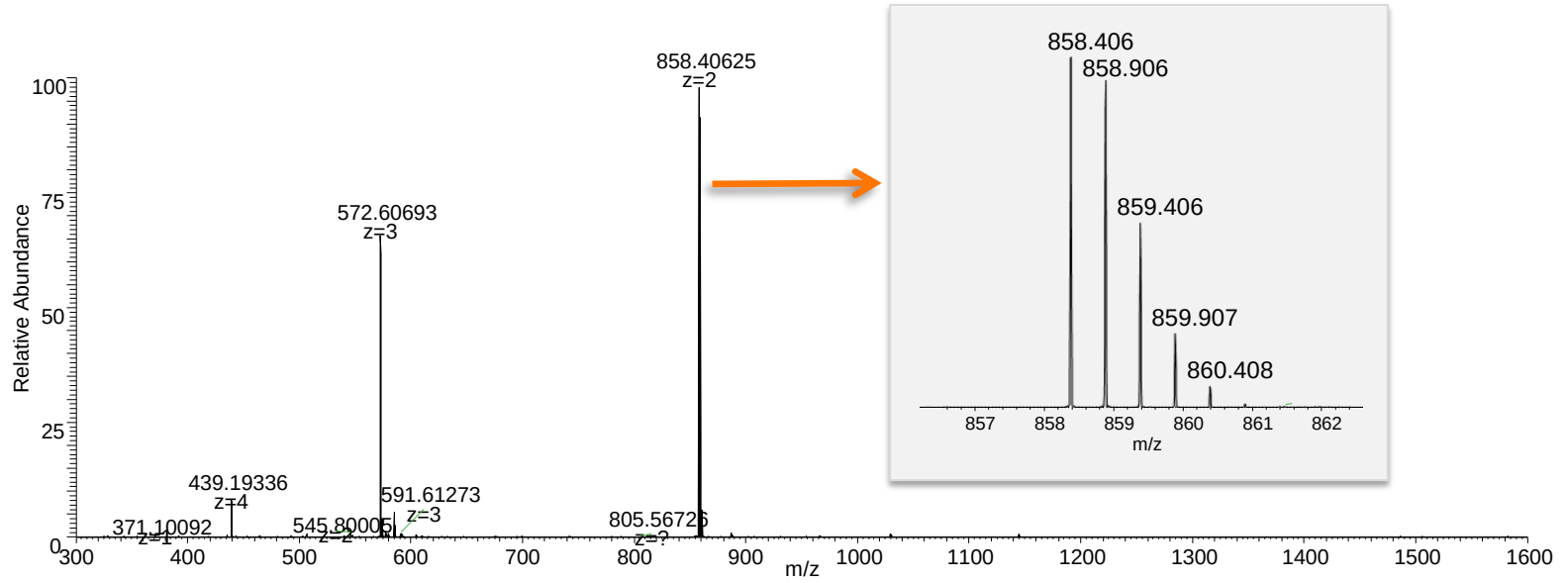


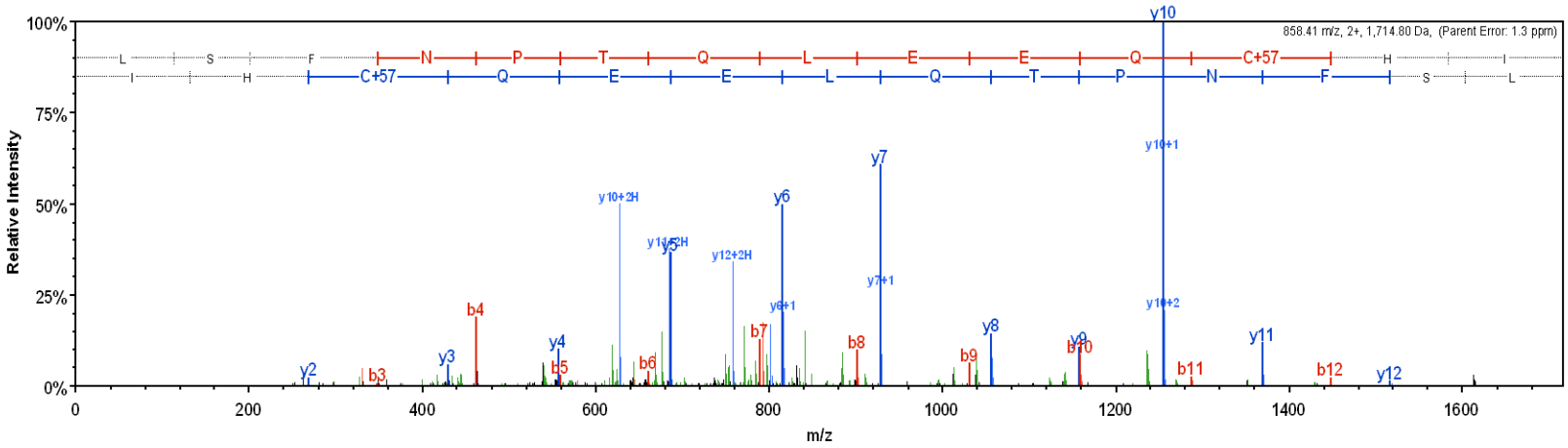


Participant Data: Lab 31U

Workflow A: In-Solution Digest

C-Terminal Peptide Example: BLG – Correctly Identified LSFNPTQLEEQCHI



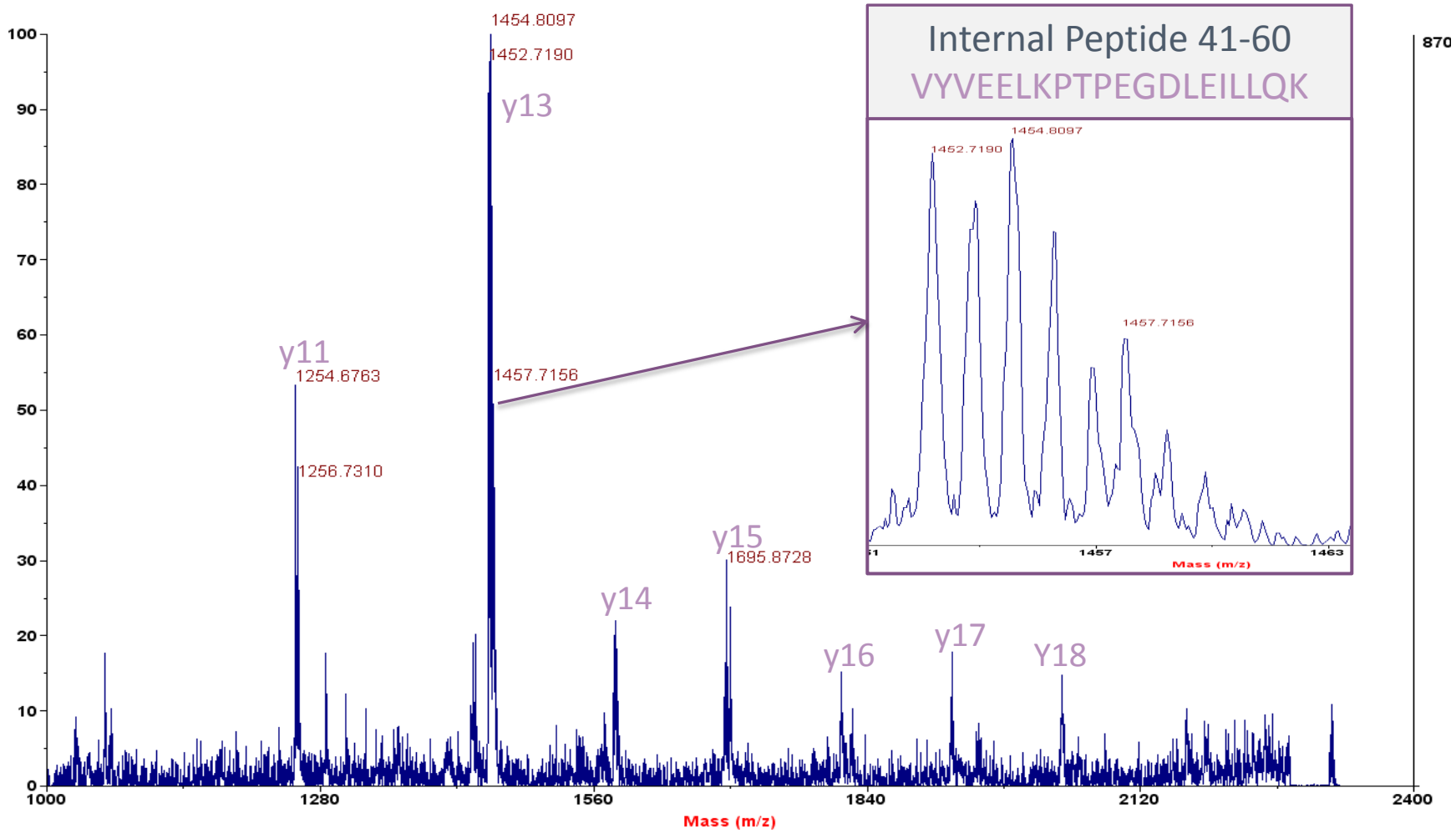


Participant Data: Lab 37C

Workflow B: In-Gel Digest

MSMS of 2313.3 Da from β-Lactoglobulin (1.5 pmol on spot)

^{18}O isotopic pattern for an internal peptide

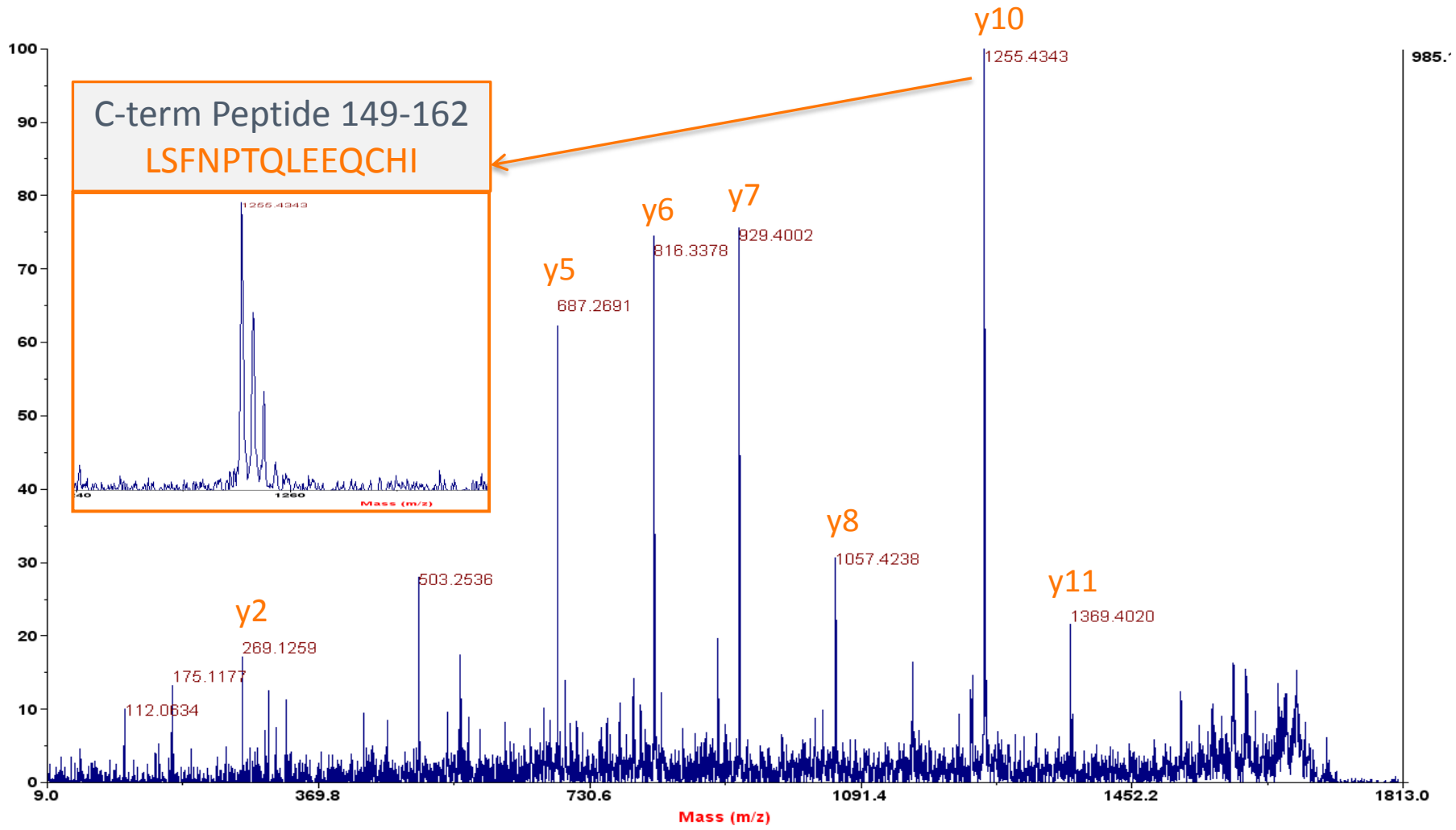


Participant Data: Lab 37C

Workflow B: In-Gel Digest

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

Typical isotopic pattern for a C-terminal peptide

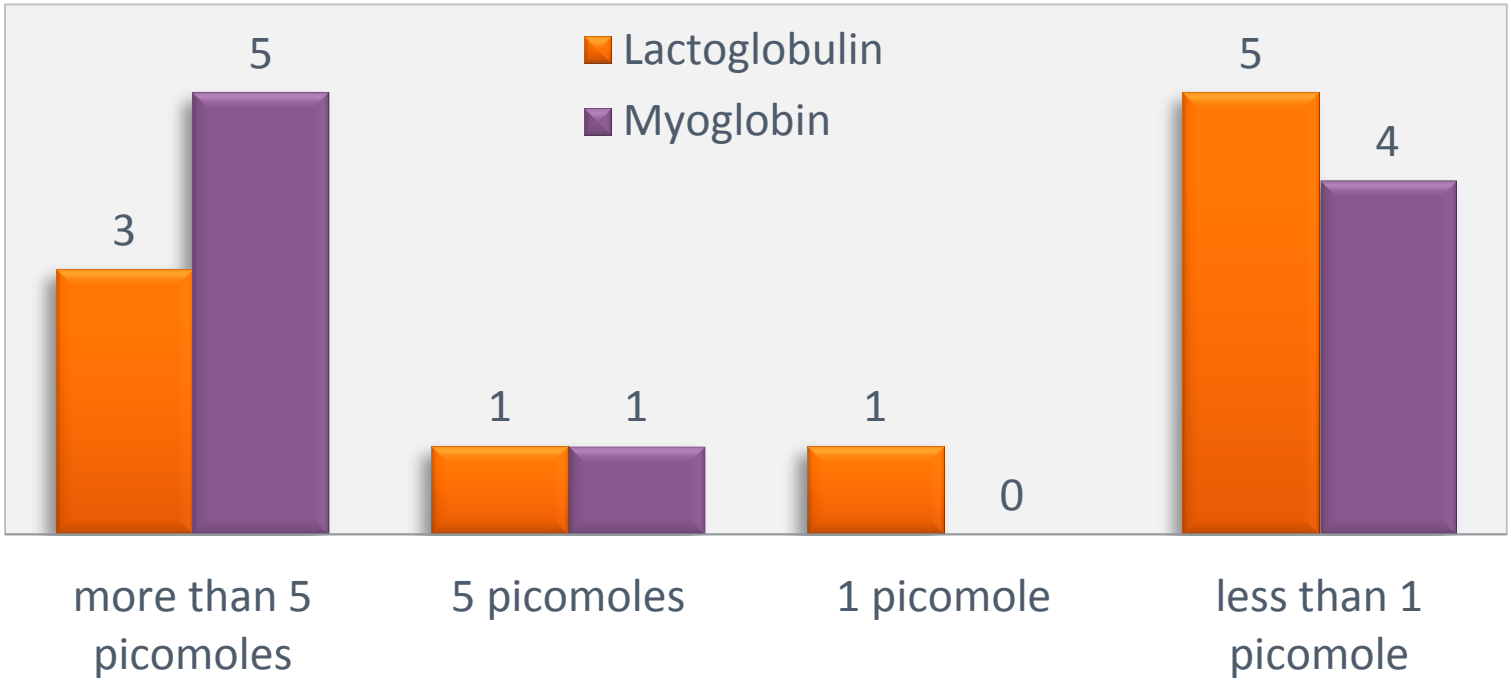


Lessons from Study Design

- 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

Sensitivity of the Technique

- PSRG goal: help Cores estimate the quantity of protein needed from investigators
- Sensitivity correlates with experience level of participants



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 $^{16}O/^{18}O$ 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)

Strategy for use in a Core

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 2nd 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!!!!!!

References

- A new mass-spectrometric C-terminal sequencing technique finds a similarity between γ-interferon and α_2-interferon and identifies a proteolytically clipped γ-interferon that retains full antiviral activity. Rose K, Simona M G, Offord R E, Prior C P, Otto B, Thatcher D R.; *Biochem J.* 1983;215:273–277

- A method for C-terminus determination of Proteins and Peptides using Protease Catalyzed 18O incorporation. Rusnak J, Davis, MT, Swiderek, KM. ABRF 1998

- Hepatitis A virus capsid protein VP1 has a heterogeneous C terminus. Graff J, Richards OC, Swiderek KM, Davis MT, Rusnak F, Harmon SA, Jia XY, Summers DF, Ehrenfeld E.; *J Virol.* 1999 Jul;73(7):6015-23

- Detection of C-terminal peptide of proteins using isotope coding strategies. Julka S, Dielman, D, Young, SA. J Chromatogr B Analyt Technol; *Biomed Life Sci*, 2008; 874:101-110

- In-gel digestion for mass spectrometric characterization of proteins and proteomes. Shevchenko A, Tomas H, Havlis J, Olsen JV, Mann M, *Nature Protocols*, **1**, *6,* p 2856-2860 (2006)

- Universal sample preparation method for proteome analysis. Wiśniewski JR, Zougman A, Nagaraj N, Mann M, *Nature Methods* **6**, 359-362 (2009)