

---

# The ABRF Protein Sequencing Research Group 2010 Study

---

N-Terminal Sequencing of an Antibody

---

## Current PSRG Members

- Peter Hunziker (Chair emeritus)
- J. Steve Smith (Chair)  
*Galveston*
- Wendy Sandoval (Chair)
- Kwasi Mawuenyega
  
- James Walters
- Bosong Xiang
- Jack Simpson (EB liaison)

*University of Zurich*

*University of Texas Medical Branch at*

*Genentech, Inc.*

*Washington University School of  
Medicine*

*Sigma-Aldrich*

*Monsanto, Co.*

*SAIC/National Cancer Institute at  
Frederick*

---

---

# ESRG 2009 Highlights

- Alternative methods for N-terminal sequence determination
    - Participants were provided with two proteins, one was commercially available and one was an N-terminally tagged recombinant protein.
    - Participants were asked to use techniques of choice to determine as far into the N-terminal sequence of both proteins as possible
  - For the first time the ESRG study was opened up to non-Edman participants.
  - ESRG was changed to PSRG to reflect the new era of terminal sequencing.
-

# ESRG 2009 Results: Sample #1

Edman

ESRG Lab #	Sample #1																											Correct Protein ID	# Correct AA Calls																						
006	G	G	S	H	H	H	H	H	G	M	A	S	M	T	G	G	Q	Q	M	G	R	D	L	Y	D	D	D	D	K	D	P	T	L	M	S	I	P	E	T	Q	K	G	V	I	F	Y	E	S	H	No	30
008	[Green]																											No	34																						
009	[Green]																											No	34																						
014	[Green]																											No	42																						
016	[Green]																											No	18																						
017	[Green]																											No	29																						
018	[Green]																											No	25																						
020	[Green]																											No	22																						
023	[Green]																											No	0																						
025	[Green]																											No	31																						
026	[Green]																											No	30																						
A	[Green]																											No	30																						
B	[Green]																											No	34																						
C	[Green]																											No	30																						
D	[Green]																											No	21																						

ISD

ESRG Lab #	Sample #1																																	Correct Protein ID	# Correct AA Calls																											
001	G	G	S	H	H	H	H	H	G	M	A	S	M	T	G	G	Q	Q	M	G	R	D	L	Y	D	D	D	D	K	D	P	T	L	M	S	I	P	E	T	Q	K	G	V	I	F	Y	E	S	H	G	K	L	E	Y	K	D	I	P	V	> 60	Yes	44
002	[Green]																																	Yes	50																											
004	[Green]																																	Yes	19																											
005	[Green]																																	Yes	16																											
007	[Green]																																	Yes	15																											
008	[Green]																																	No	18																											
010	[Green]																																	Yes	18																											
012	[Green]																																	Yes	21																											
013	[Green]																																	Yes	0																											
015	[Green]																																	Yes	16																											
019	[Green]																																	Yes	18																											
021***	[Green]																																	Yes	18																											
022	[Green]																																	Yes	31																											
024	[Green]																																	Yes	17																											
C	[Green]																																	No	9																											

Bottom Up

ESRG Lab #	Sample #1																																	Correct Protein ID	# Correct AA Calls
001	[Green]																																	Yes	13
005	[Green]																																	Yes	12
011	[Green]																																	Yes	7
016	[Green]																																	Yes	9
018	[Green]																																	Yes	11
C	[Green]																																		

# ESRG 2009 Results: Sample #2

Edman

ESRG Lab #	Sample #2																										Correct Protein ID	# Correct AA Calls																				
	V	K	G	V	N	G	F	G	R	I	G	R	L	V	T	R	A	A	F	N	S	G	K	V	D	V			V	A	I	N	D	P	F	I	D	L	H	Y	M	V	Y	M	F	Q	Y	D
006	[Green]																										Yes	30																				
008	[Green]																										Yes	35																				
009	[Green]																										Yes	49																				
014	[Green]																										Yes	78																				
016	[Green]																										Yes	30																				
017	[Green]																										Yes	41																				
018	[Green]																										Yes	33																				
020	[Green]																										Yes	24																				
023	[Green]																										Yes	28																				
025	[Green]																										Yes	40																				
026	[Green]																										Yes	28																				
A	[Green]																										Yes	64																				
B	[Green]																										Yes	24																				
C	[Green]																										Yes	37																				
D	[Green]																										Yes	18																				

ISD

ESRG Lab #	Sample #2																																Correct Protein ID	# Correct AA Calls																									
	V	K	G	V	N	G	F	G	R	I	G	R	L	V	T	R	A	A	F	N	S	G	K	V	D	V	V	A	I	N	D	P			F	I	D	L	H	Y	M	V	Y	M	F	Q	Y	D	S	T	H	G	K	F	H	G	T	V	K
001	[Green]																																Yes	50																									
002	[Green]																																Yes	54																									
004	[Green]																																Yes	28																									
005	[Green]																																Yes	32																									
007	[Green]																																Yes	17																									
008	[Green]																																Yes	9																									
010	[Green]																																Yes	30																									
012	[Green]																																Yes	28																									
013	[Green]																																Yes	22																									
015	[Green]																																Yes	32																									
019	[Green]																																Yes	25																									
021***	[Green]																																Yes	28																									
022	[Green]																																Yes	28																									
024	[Green]																																Yes	12																									
C	[Green]																																Yes	13																									

Bottom Up

ESRG Lab #	Sample #2																																Correct Protein ID	# Correct AA Calls																									
	V	K	G	V	N	G	F	G	R	I	G	R	L	V	T	R	A	A	F	N	S	G	K	V	D	V	V	A	I	N	D	P			F	I	D	L	H	Y	M	V	Y	M	F	Q	Y	D	S	T	H	G	K	F	H	G	T	V	K
001	[Green]																																Yes	8																									
005	[Green]																																Yes	9																									
010	[Green]																																Yes	9																									
011	[Green]																																Yes	10																									
018	[Green]																																Yes	8																									
C	[Green]																																Yes	9																									

# ESRG 2009 Results Comparison

	<i>Automated Edman n=15</i>	<i>ISD n=15</i>	<i>Biochemical Capture n=1</i>	<i>MAAH n=1</i>	<i>ECD/ ETD n=1</i>
<i>Average correct residues called from N-terminus</i>	<b>30 &amp; 38</b>	<b>38 &amp; 44</b>	<b>None</b>	<b>26</b>	<b>25</b>
<i>Can residues called from C-terminus?</i>	<b>No</b>	<b>Yes</b>	<b>No</b>	<b>Yes</b>	<b>Yes</b>
<i>Time to sequence 12 residues from N-terminus</i>	<b>&gt; 6 h</b>	<b>&lt; 5 min</b>	<b>6 h</b>	<b>5 - 30 min</b>	<b>30 min</b>
<i>Can identify blocked N-termini?</i>	<b>No</b>	<b>Yes</b>	<b>Yes</b>	<b>Yes</b>	<b>Yes</b>
<i>Can easily identify ragged N-termini?</i>	<b>Yes</b>	<b>No</b>	<b>Yes</b>	<b>No</b>	<b>No</b>
<i>Can differentiate between I/L &amp; K/Q?</i>	<b>Yes</b>	<b>No*</b>	<b>No</b>	<b>No</b>	<b>Yes</b>

---

# ESRG 2009 Study Conclusions

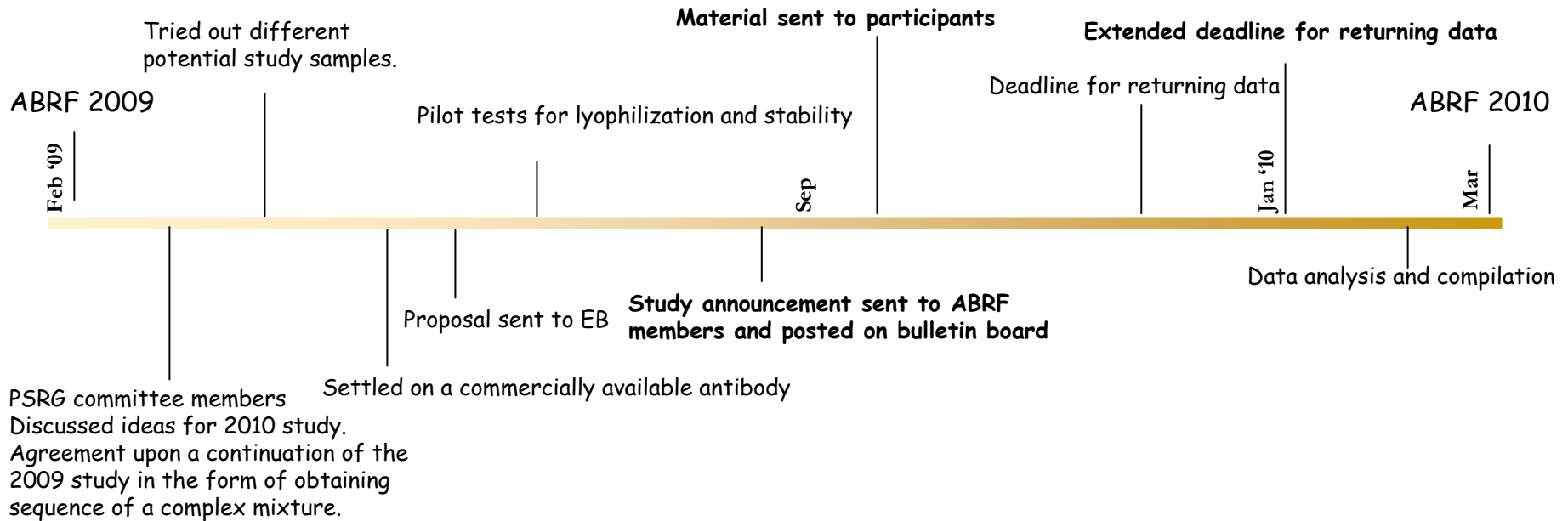
- Edman sequencing remains a reliable means for determining the *N*-terminal sequence of an unblocked protein
  - The Top down approach shows great promise for determining the *N*- (and *C*-) terminal sequence of a protein in solution
  - The bottom up approach worked well when the complete sequence of the subject protein was found in the database
  - Database searches were often relied upon to fill in missing or ambiguous mass spectral data
    - Isobaric amino acids
    - Missing low mass ions
    - Bonds with poor or no signal
-

---

# Historical Uses of Edman Sequencing

1. Cleavage site determination for proteases
2. Sequencing of MHC peptides
3. Sequencing of synthetic peptide libraries
4. Full characterization of proteins, especially recombinant proteins, that are present in large quantities
5. Stoichiometry, Edman is semi-quantitative
6. Protein identification for non-model organisms which do not have extensive DNA sequencing
7. Domain mapping
8. Confirmation of N-terminus
9. As a help for mass spectrometry sequencing to perform manual subtractions
10. Product characterization for SOPs for pharma
11. Can distinguish between the isobaric amino acids Leucine and Isoleucine
12. Clonality determination or antibody sequencing for cloning

# 2010 PSRG Study Timeline



## To obtain terminal sequence information of an antibody.

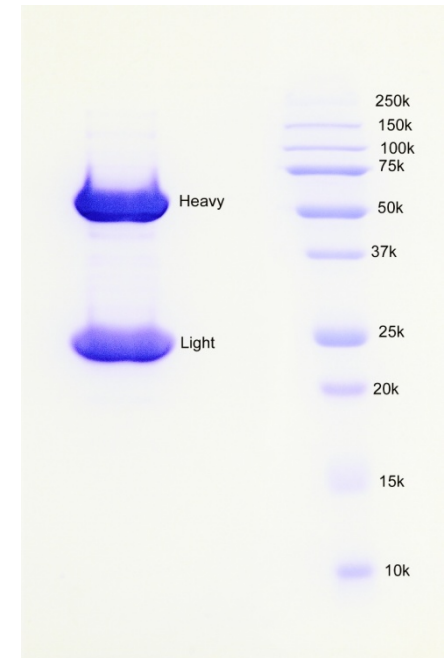
- What techniques are complementary and parallel with Edman?
- Can any technique replace Edman for de novo antibody characterization of the N-terminus?
- How much N- and C-terminal information can be obtained and from what technique?
- Is there a necessity to reduce the complexity of the mixture prior to analysis?
- How do you determine a sequence that is not in the public databases?



# Sample Preparation and Distribution

Sample was donated in liquid formulation containing glycerol.  
Clean-up and lyophilization procedure:

1. 50 $\mu$ g (50 $\mu$ L) of sample was placed in 1.5mL tube (100 tubes total)
2. Chloroform-Methanol Precipitation
  1. To each 50  $\mu$ L sample add 150  $\mu$ L Methanol
  2. Vortex. Add 200  $\mu$ L water.
  3. Vortex. Add 50  $\mu$ L chloroform
  4. Vortex and spin 13,000 x g for 2 min
  5. Two layers are formed, remove aqueous layer
  6. Add 200  $\mu$ L methanol. Protein will precipitate
  7. Vortex then spin and remove liquid. Speed vac to dryness
3. Mail to participants



*Each participant received two 50 $\mu$ g lyophilized vials of material.*

---

# PSRG 2010 Laboratory Participation

# Laboratories requesting sample (including 6 PSRG members):.....	46
# Laboratories returning data:.....	25
# Laboratories returning supplemental data:.....	5
# Laboratories who returned multiple analyses:.....	2
Techniques used:	
Edman Degradation:.....	11
Top-Down MS:	
In-Source Decay Fragmentation:.....	5
ETD/ECD:.....	0
Bottom-Up MS (Enzymatic Digestions):.....	11
Terminal labeling:.....	1
Multiple techniques reported:.....	3

---

---

# N-Terminal Techniques: Edman Degradation

---

# Edman Degradation Sample Preparation

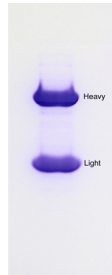
Lab	Sample Reconstitution	DTT?	PGAP?	SDS-PAGE?	HPLC?
003	100mM Tris	✓	✓	✓	no
008	0.1% TFA/ 20% ACN	not stated	no	✓	no
021	0.1% TFA/ 20% ACN	✓	✓	✓	no
024	0.1% TFA/ 20% ACN	not stated	no	✓	no
025	25mM Ammonium bicarbonate	✓	no	✓	no
034	8M Guanidine	✓	no	no	✓
038	1x Tricine Sample Buffer	✓	no	✓	no
042	0.1% TFA/ 20% ACN	no	no	no	no
043	0.1% TFA/ 20% ACN	no	no	no	no
PSRG-1	250mM Phosphate, 5mM EDTA	✓	✓	✓	no
PSRG-2	100mM Ammonium bicarbonate	✓	✓	✓	no

# Edman Workflows (11 participants)

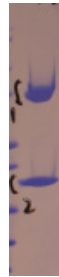


PSRG 2010 Sample

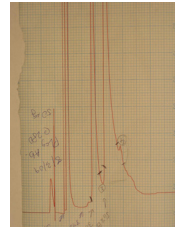
SDS-PAGE



Electroblot



HPLC



Direct sequence

Light chain only sequenced

Heavy chain de-blocked with PGAP and sequenced

Light chain sequenced

Heavy chain de-blocked with PGAP and sequenced

Light chain sequenced



*ABI Precise Instruments:*

8 - 494 HT's

1 - 492 HT

1 - 494 cLC

*Shimadzu PPSQ33-A*



# Edman degradation results

Alignment of **Heavy Chain** N-term results for **4** Edman participants:

003	1	QVQLQQS	AAELARP	GASV	KMS	-	X	AXX	-	-	-	-	25
021	1	-	-	-	-	-	GQP	TLV	-	-	-	-	7
PSRG1	1	QVQLQQS	AAELARP	GASV	KMS	XX	AX	GX	XX	FT	-	-	30
PSRG2	1	QVQLQQS	AAELARP	GASV	KMS	CK	AV	G	-	-	-	-	26

Alignment of **Light Chain** N-term results for **11** Edman participants:

003	1	DVLM	TQIP	LSLP	VSLG	GRAD	SIS	-	-	-	-	-	-	23
008	1	DVLM	TQIP	LSLP	VSLG	DQAS	IS	CRSS	SQX	IVHR	NGNT	YL	-	38
021	1	DVLM	TQIP	LSLP	VSL	-	-	-	-	-	-	-	-	15
024	1	DVLM	TQIP	LSLP	VSLG	DQAS	IS	XRSS	SQ	-	-	-	-	27
025	1	DVLM	TQIP	-	-	-	-	-	-	-	-	-	-	8
034	1	DVLM	TQIP	LSLP	VSLG	DQAS	IS	-	-	-	-	-	-	22
038	1	DVLM	TQIP	LSLP	VSLG	DQAS	IS	XRSS	SQX	IVHR	NGNT	XXE	-	39
042	1	DVLM	TQIP	LSLP	VSLG	DVRV	-	-	-	-	-	-	-	20
043	1	DVLM	TQIP	LSLP	VSLG	DQAS	IS	XRSS	SQX	IVH	XN	-	-	33
PSRG-1	1	DVLM	TQIP	LSLP	VSLG	DQAS	IS	XRSS	SQX	IVH	-	-	-	31
PSRG-2	1	DVLM	TQIP	LSLP	VSLG	DQAS	IS	CRSS	SQ	-	-	-	-	27

# PSRG 2010 Edman Conclusions & Observations

Edman sequencing allows for direct determination of the protein's *N*-terminal sequence.

- Proteins with blocked N-termini cannot be sequenced unless the protein is de-blocked.
  - The PSRG-2010 sample's heavy chain was blocked with a pyroglutamic acid residue and required removal with *pyroglutamate aminopeptidase* (PGAP).
- Length of sequencing is affected by the efficiency of the instrument and size of the molecule.
- Read lengths were shorter than obtained by the MS techniques
- All commonly occurring amino acids including Ile/Leu and Lys/Gln as well as some stable PTMs can be readily identified.
- C-terminal information can not be obtained by Edman degradation
- Edman does not rely on database homology to obtain sequence information.

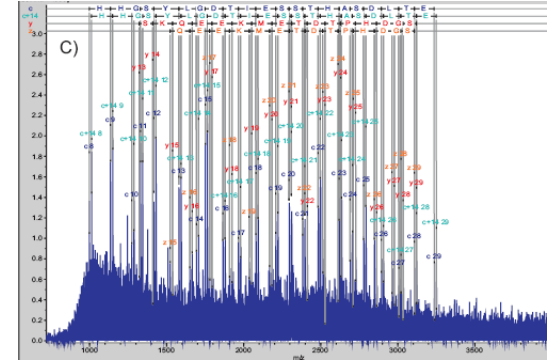
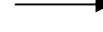
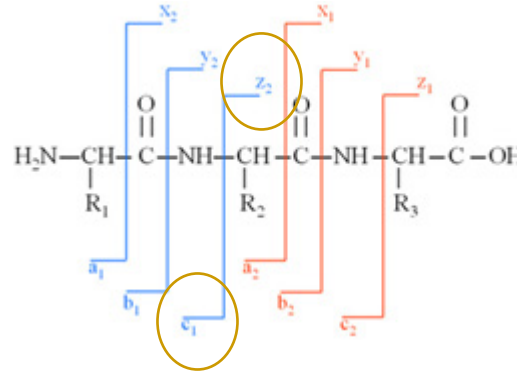
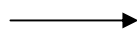
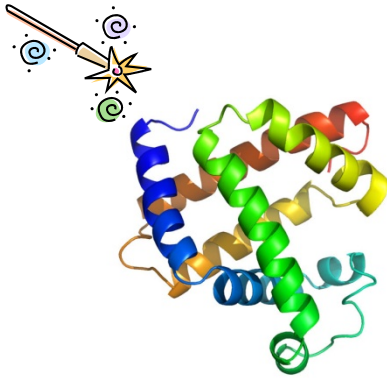
---

# N-Terminal Techniques Overview: Top-Down MS

---

In-Source Decay Fragmentation

# In-Source Decay (MALDI-ISD)



## MALDI-MS

- Analyte + matrix on metal target plate
- Spot is excited with laser, ionization occurs
- Ions are resolved by mass in TOF analyzer
- Second TOF allows for MS/MS by precursor ion fragmentation

## MALDI-ISD

- a “pseudo-MS/MS” technique
- Allows sequence determination without digestion (“Top Down”)
- Decomposition of the protein happens in the MALDI plume
- sub nsec timescale
- Ion formation due to radical transfer from matrix to analyte (Takayama, 2001)

# ISD and T3 Sequencing

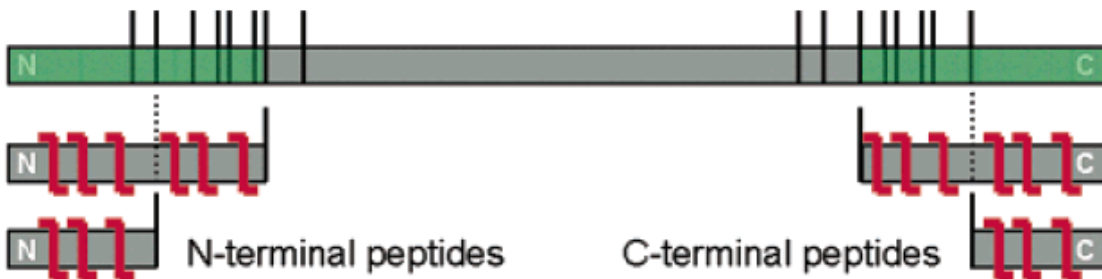
Protein sequence



ISD fragmentation (1-4 kDa)



Terminus-specific fragmentation



# ISD Experimental (8 labs)

## Sample Prep



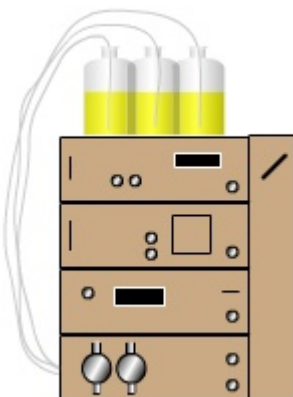
### *Recon*

2 - 20% ACN/0.1% TFA  
1 - 5% ACN/0.1% TFA  
2 - 0.1% TFA

### *Reduction*

2 - 20mM DTT  
1 - Reduction & Alkylation

## Separation



4 Labs Separated  
chains by HPLC

## ISD Instrumentation

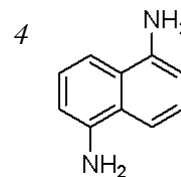


Bruker UltrafleXtreme (2)  
Bruker Ultraflex-1 (1)



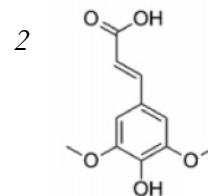
ABI 4800 MALDI-TOF/TOF (2)

## Matrix



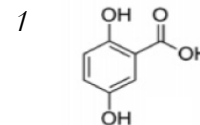
**DAN**

*1,5-diaminonaphthalene*



**SA**

*sinapinic acid*



**DHB**

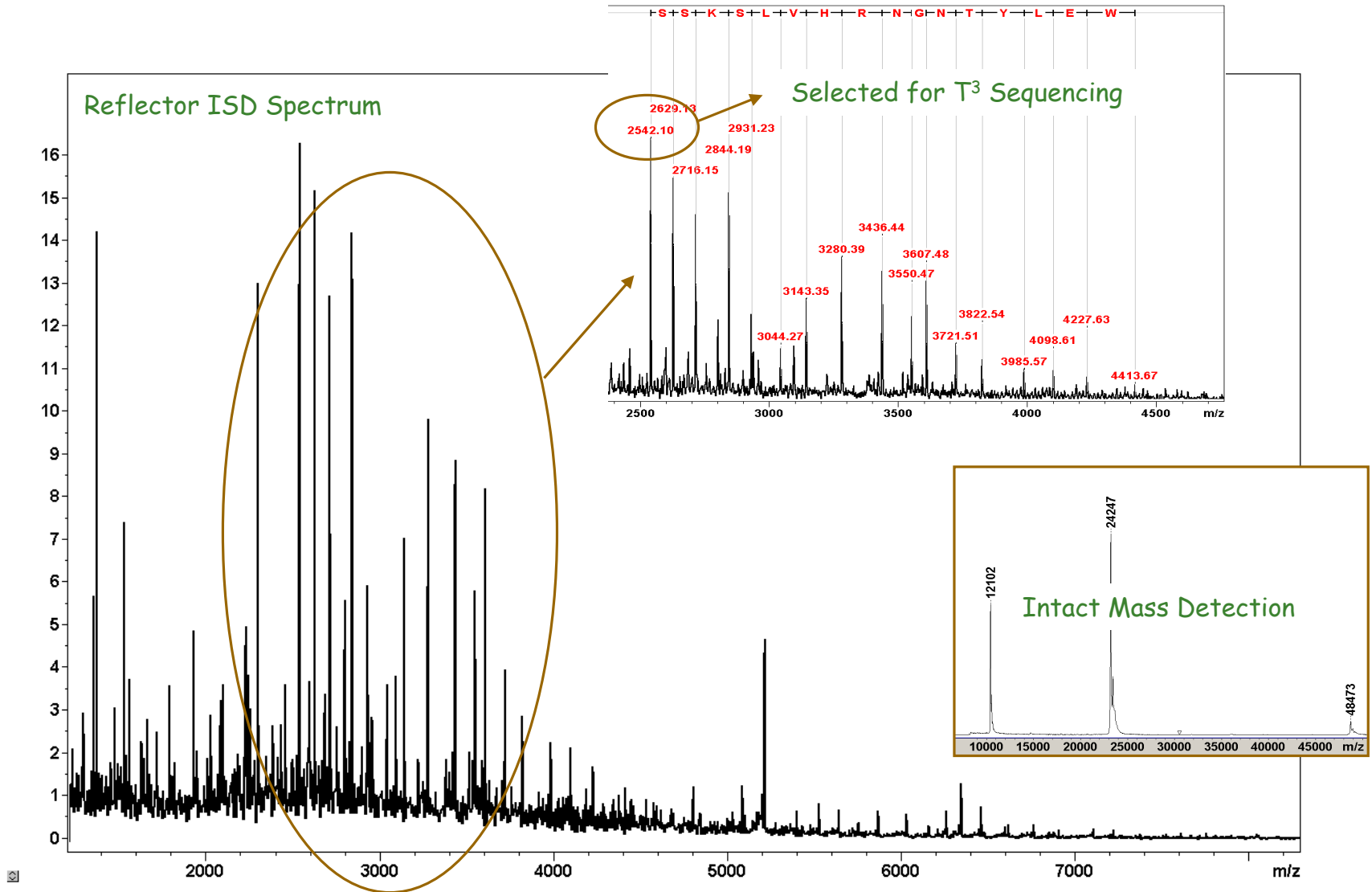
*2,5-dihydrobenzoic acid*

## Analysis

Bruker BioTools (3)  
Bruker FlexAnalysis (2)  
ABI Explorer (2)  
Mascot & p-BLAST (1)  
Manual Interpretation (2)

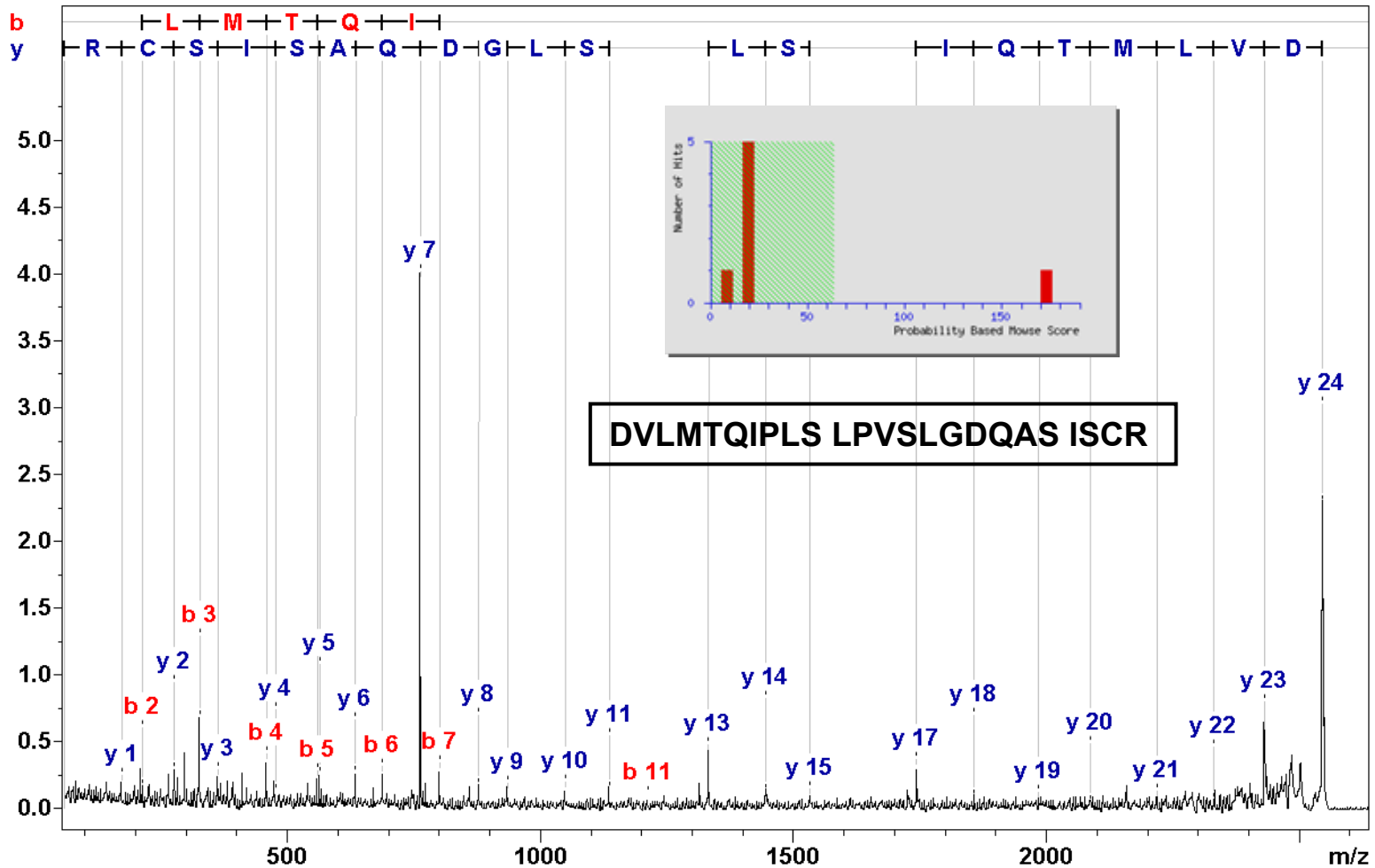


# Light Chain Analysis: LC-ISD Example



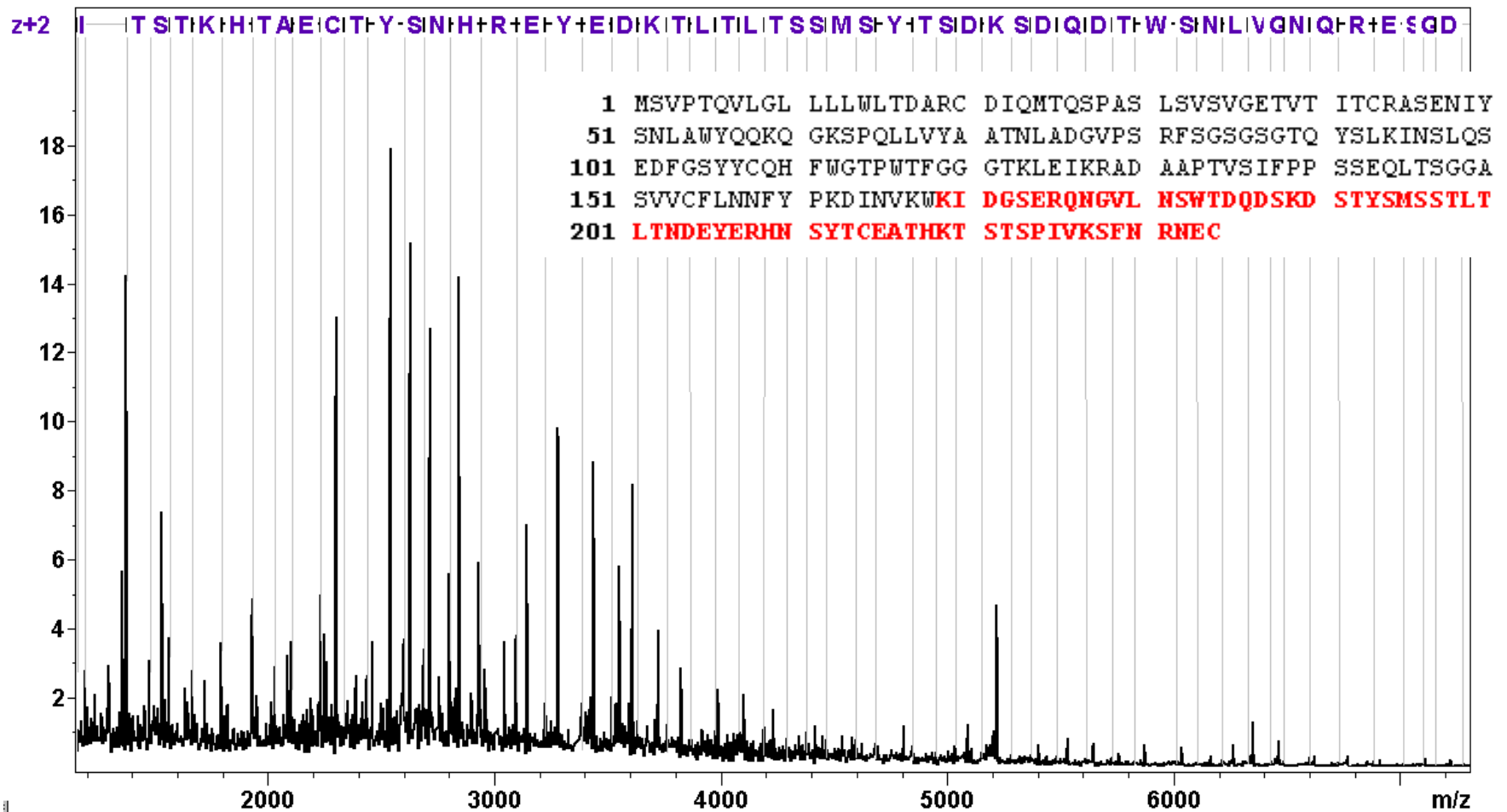
# Light Chain Analysis: Lab 10

## T<sup>3</sup>- Spectrum of ISD Fragment m/z 2542.3

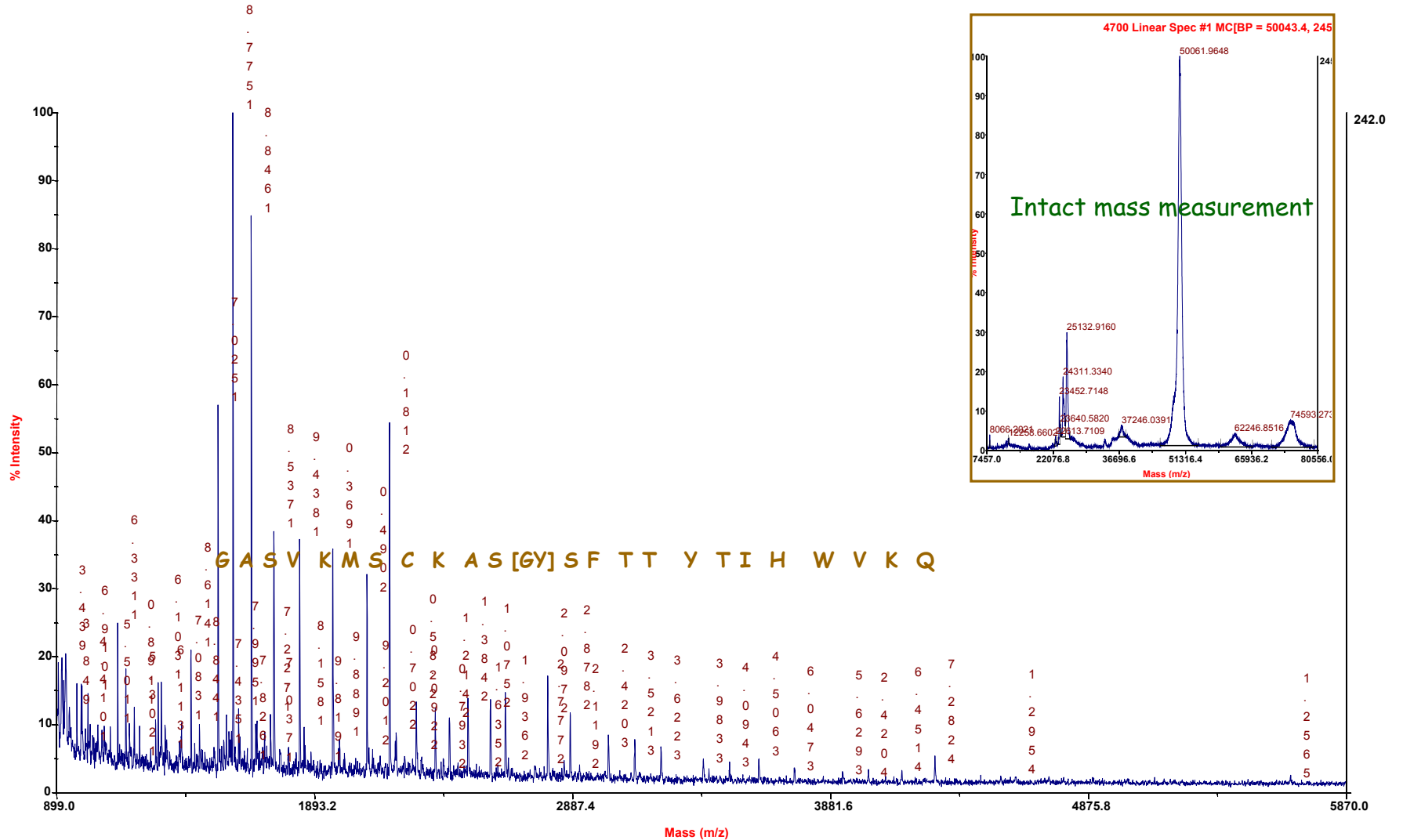


# Light Chain C-Terminal Analysis

*Mascot through BioTools*



# Heavy Chain ISD Spectrum with Manual Data Analysis







# Top-Down MS ISD Study Conclusions

In-source decay is a top down technique whereby fragment ions from both termini are generated upon laser ablation.

- Read lengths varied from 10 to over 70 amino acids from the N-terminus
- To obtain true N-terminal information it was necessary to use T<sup>3</sup> sequencing.
- Blocked N-termini are not prohibitive to ISD analyses.
- C-terminal sequence information can be obtained.
- It was necessary for ISD to use T3 sequencing to obtain the terminal sequence.
- The entire ion series representing the termini may not be present and extrapolations of the ISD spectra were not sufficient to obtain the terminal sequence.
- ISD is unable to differentiate between the isobaric amino acids Ile and Leu and has difficulty differentiating between nominal mass difference of Lys and Gln.
- Bioinformatics was relied upon to report an amino acid when in question.

---

# N-Terminal Techniques Overview: Bottom-Up MS Techniques

---

Enzymatic Digestion

# Bottom-Up MS Experimental – LC Systems

All Labs used LC separation prior to peptide analysis.

11 Labs returned Bottom up data for the Study



*Waters nanoAcquity - 5*



*Waters capLC - 1*



*Dionex nanoLC  
UltiMate 3000 - 1*



*Michrom Paradigm  
nanoCapillary MDLC  
MS4 - 1*

# Bottom-Up Experimental – MS Instrumentation



*Thermo LTQ XL - 3*



*Waters Q-TOF micro - 2*



*Bruker OTOF-Q II - 1*



*Thermo LTQ-Orbitrap XL - 2*



*Waters Synapt HDMS - 1*

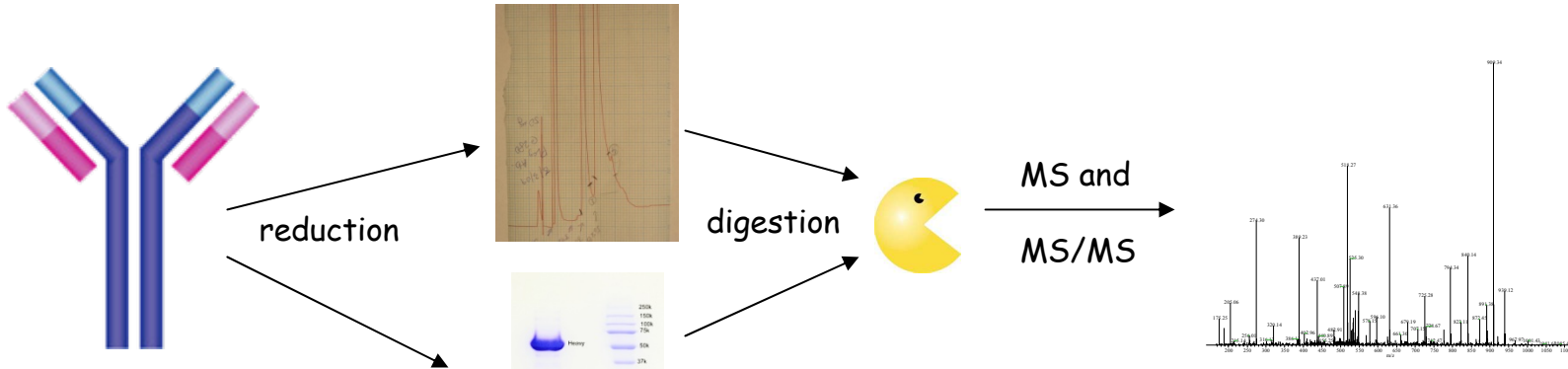


*AB Sciex TOF/TOF 5800 - 1*



*Thermo LTQ-FT - 1*

# Bottom up Sample Preparation



## Reconstitution Buffer

(dependent on separation method)

- 4- Sample Buffer with DTT
- 3- 25mM Ammonium bicarbonate
- 2- 20% ACN/0.1% TFA
- 1- Performic acid
- 1- Modified RIPA buffer

## Sample Separation

- 8- SDS-PAGE
- 3- HPLC

## Digestion Enzymes

- 11- Trypsin
- 3- Multiple Digestions
  - Trypsin, Pepsin, Elastase*
  - Trypsin, Chymotrypsin, PNGaseF*
  - Trypsin + 6 others*



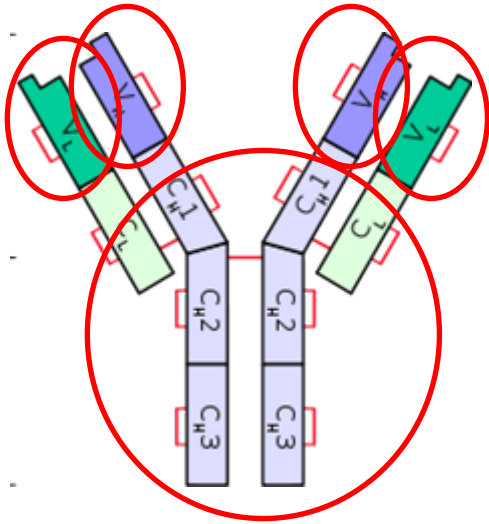
## Analysis

- 6- Mascot
- 1- BioTools
- 1- Protein Pilot
- 1- In house analysis tool

## Database

- 2- NCBIInr
- 2- SwissProt
- 7- Unspecified

# Bottom up Strategies



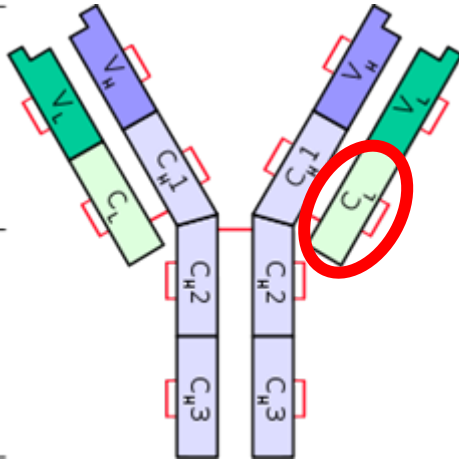
1. Use database search to obtain Ab constant regions for light and heavy chains
2. Use homology searches to reveal individual peptide hits in variable regions
3. Piece together peptides from different database antibody sequences and/or  
Piece together overlapping peptides from various digests
4. Fill in missing regions with *de novo* sequencing and manually confirm database peptide hits

## Example:

From an NCBI Mascot search of a single in gel tryptic digestion followed by LC-MS/MS, one participant obtained complete coverage of the constant region and 42% coverage of variable region of Light chain.

# Digestion database searches were used to identify the constant regions

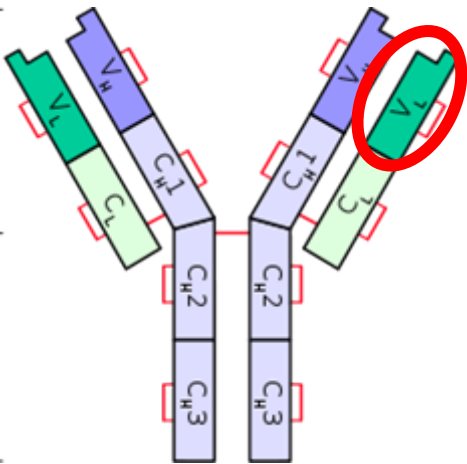
## IGKC\_MOUSE – 100% Coverage Ig kappa chain C region Mus musculus



1 ADAAPTVSIF PPSSEQLTSG GASVVCFLNN FYPKDINVKW KIDGSERQNG  
51 VLNSWTDQDS KDSTYSMSST LTLTKDEYER HNSYTCEATH KTSTSPIVKS  
101 FNRNEC

Start - End	Observed	Mr (expt)	Mr (calc)	ppm	Miss Sequence
1 - 34	1191.2364	3570.6874	3570.7341	-13	0 -.ADAAPTVSIFPPSSEQLTSGGASVVCFLNNFYPK.D ( <a href="#">Ions score 92</a> )
1 - 39	1036.0043	4139.9881	4140.0514	-15	1 -.ADAAPTVSIFPPSSEQLTSGGASVVCFLNNFYPKDINVK.W ( <a href="#">Ions score</a> )
<a href="#">66</a>					
18 - 27	492.7198	983.4250	983.4383	-13	0 L.TSGGASVVCFL.L ( <a href="#">Ions score 18</a> )
28 - 40	825.9246	1649.8346	1649.8566	-13	2 F.LNNFYPKDINVKW.K ( <a href="#">Ions score 40</a> )
29 - 40	513.2566	1536.7480	1536.7725	-16	1 L.NNNFYPKDINVKW.K ( <a href="#">Ions score 42</a> )
35 - 41	451.7521	901.4897	901.5021	-14	1 K.DINVKW.K.I ( <a href="#">Ions score 22</a> )
40 - 47	495.7381	989.4616	989.4930	-32	1 K.WKIDGSER.Q ( <a href="#">Ions score 40</a> )
40 - 61	642.0439	2564.1464	2564.1779	-12	2 K.WKIDGSERQNGVLNSWTDQDSK.D 2 Deamidated (NQ) ( <a href="#">Ions score 27</a> )
41 - 52	658.8419	1315.6692	1315.6732	-3	0 W.KIDGSERQNGVL.N Deamidated (NQ) ( <a href="#">Ions score 46</a> )
41 - 55	568.6115	1702.8126	1702.8274	-9	1 W.KIDGSERQNGVLNSW.T Deamidated (NQ) ( <a href="#">Ions score 30</a> )
41 - 65	948.4320	2842.2742	2842.3006	-9	2 W.KIDGSERQNGVLNSWTDQDSKDY.S ( <a href="#">Ions score 72</a> )
42 - 61	750.6699	2248.9878	2249.0196	-14	1 K.IDGSERQNGVLNSWTDQDSK.D Deamidated (NQ) ( <a href="#">Ions score</a> )
<a href="#">53</a>					
48 - 61	796.8535	1591.6925	1591.7114	-12	0 R.QNGVLNSWTDQDSK.D Deamidated (NQ) ( <a href="#">Ions score 129</a> )
48 - 75	1036.4745	3106.4017	3106.4401	-12	1 R.QNGVLNSWTDQDSKDYMSSTLTLTK.D ( <a href="#">Ions score 28</a> )
48 - 80	950.6733	3798.6640	3798.7166	-14	2 R.QNGVLNSWTDQDSKDYMSSTLTLTKDEYER.H ( <a href="#">Ions score 89</a> )
53 - 65	773.8094	1545.6042	1545.6219	-12	1 L.NSWTDQDSKDY.S ( <a href="#">Ions score 95</a> )
62 - 75	767.8620	1533.7094	1533.7232	-9	0 K.DSTYSMSSTLTLTK.D ( <a href="#">Ions score 100</a> )
62 - 80	1114.0068	2225.9990	2225.9998	-0	1 K.DSTYSMSSTLTLTKDEYER.H ( <a href="#">Ions score 134</a> )
72 - 84	828.3692	1654.7238	1654.7587	-21	2 L.TLTKDEYERHNSY.T ( <a href="#">Ions score 63</a> )
81 - 91	449.8374	1346.4905	1346.5673	-57	0 R.HNSYTCEATH.K.T ( <a href="#">Ions score 73</a> )
81 - 99	541.0027	2159.9819	2160.0270	-21	1 R.HNSYTCEATHKTSTSPIVK.S ( <a href="#">Ions score 56</a> )
85 - 101	947.4650	1892.9155	1892.9302	-8	0 Y.TCEATHKTSTSPIVKSFN.N ( <a href="#">Ions score 71</a> )
85 - 106	642.5449	2566.1507	2566.1904	-15	1 Y.TCEATHKTSTSPIVKSFNNEC.- ( <a href="#">Ions score 51</a> )
92 - 99	416.7128	831.4110	831.4702	-71	0 K.TSTSPIVK.S ( <a href="#">Ions score 43</a> )
92 - 103	465.2448	1392.7125	1392.7361	-17	1 K.TSTSPIVKSFN.N Carbamidomethyl (K) ( <a href="#">Ions score 40</a> )
92 - 106	599.6154	1795.8243	1795.8523	-16	2 K.TSTSPIVKSFNNEC.- Carbamidomethyl (K) ( <a href="#">Ions score 40</a> )
100 - 106	463.6736	925.3327	925.3712	-42	1 K.SFNNEC.- ( <a href="#">Ions score 20</a> )
100 - 106	463.6746	925.3346	925.3712	-40	1 K.SFNNEC.- ( <a href="#">Ions score 19</a> )

# Peptides from Various Database sequences were Combined for partial coverage of the variable region



## KV2A7\_MOUSE Ig kappa chain V-II region

```
1 DVVMTQTPLS LPVSLGDQAS ISCRSSQSLV HSNGNTYLNW YLQKAGQSPK
51 LLIYKVSNRF SGVPDRFSGS GSGTDFTLKI SRVEAEDLGI YFCSQTTHVP
101 PTFGGGTKLE IKR
```

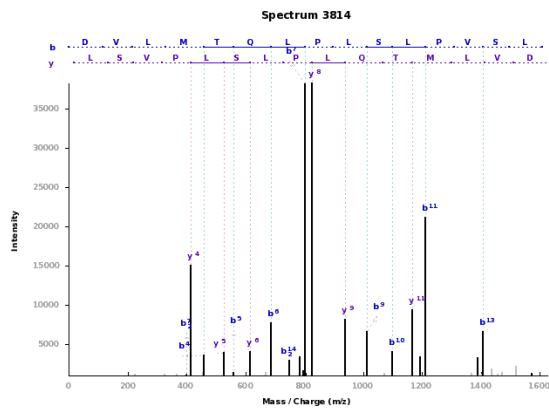
## KV2A4\_MOUSE Ig kappa chain V-II region 2S1.3

```
1 DIVMTQAAFS NPVTLGTSAS FSCRSSKSLQ QSKGITYLYW YLQKPGQSPQ
51 LLIYQMSNLA SGVPDRFSGS GSGTDFTLRI SRVEAEDVGV YYCANLQELP
101 YTFGGGTKLE IK
```

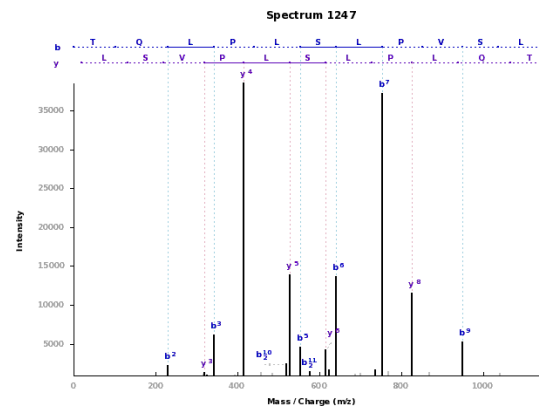
## COMBINED RESULT – 60% coverage

```
1 DVVMTQTPLS LPVSLGDQAS ISCRSSQSLV HSNGNTYLNW YLQKPGQSPQ
51 LLIYKVSNRF SGVPDRFSGS GSGTDFTLKI SRVEAEDLGI YFCSQTTHVP
101 PTFGGGTKLE IKR
```

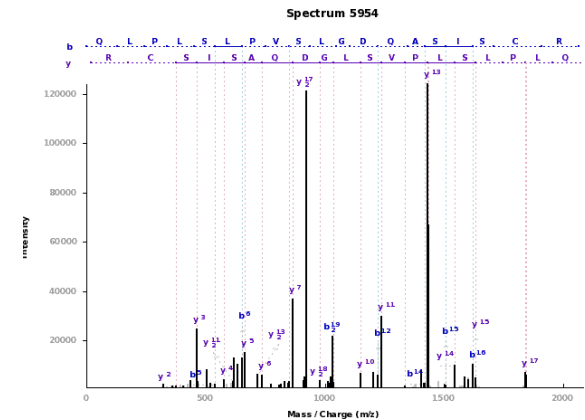
# Use of overlapping spectra and informatics to extend sequence



MS/MS spectrum LC N-terminus  
Residues 1-15

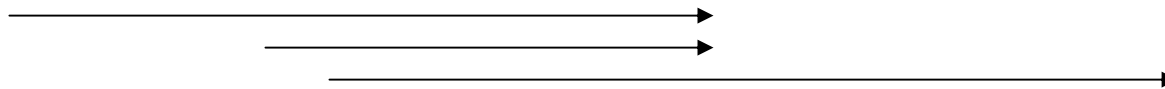


MS/MS spectrum LC N-terminus  
Residues 5-15

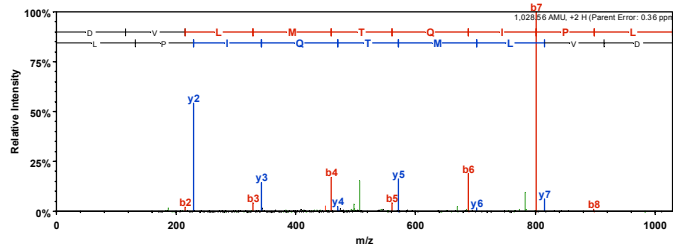


MS/MS spectrum LC  
Residues 6-24

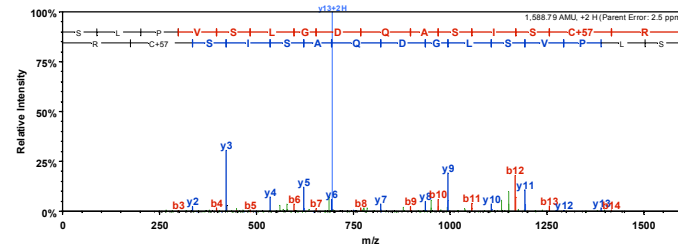
DVLMT QLPLS LPVSL GDQAS ISCR



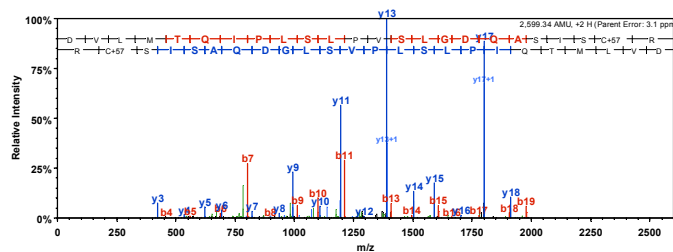
# Multiple Digestions and Peptide Assembly based on homology



MS/MS spectrum LC N-terminus  
Residues 1-9

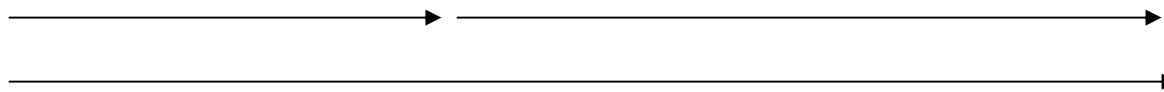


MS/MS spectrum LC N-terminus  
Residues 10-24

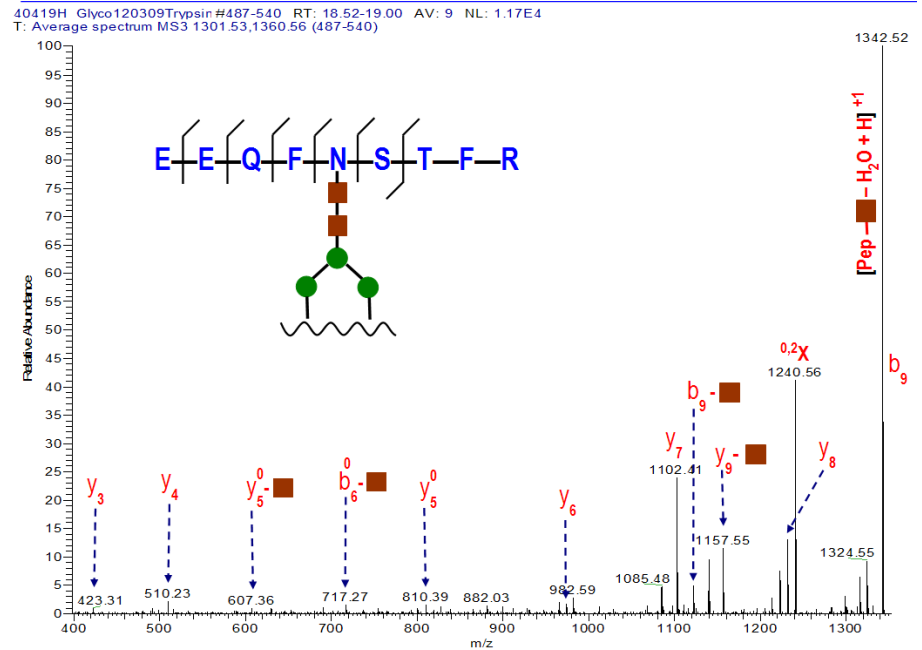
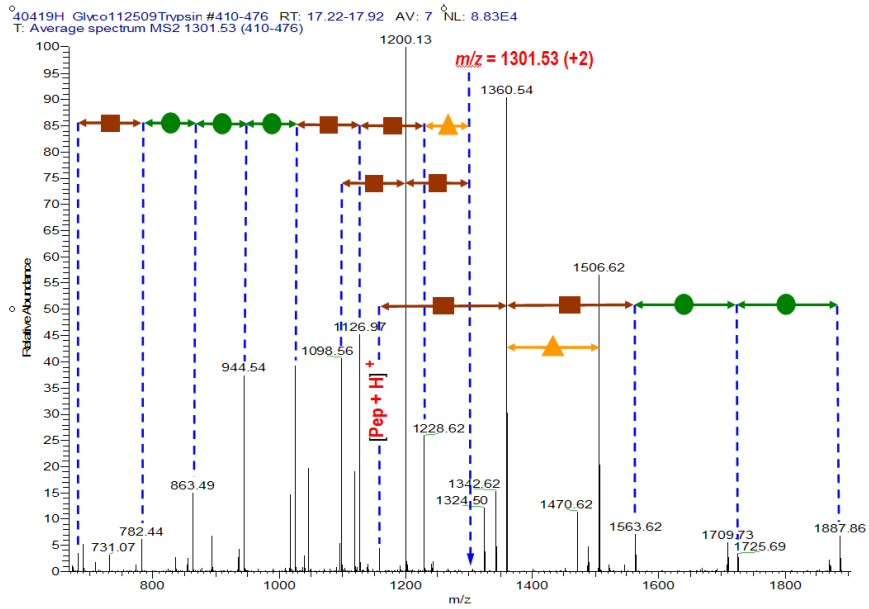


MS/MS spectrum LC N-terminus  
Residues 1-24

DVLMT QLPLS LPVSL GDQAS ISCR



# CHO Characterization







# Bottom-Up Conclusions

Bottom up analysis involves enzymatic or chemical cleavage of the protein followed by MS/MS analysis of the peptide mixture.

- Small (6-25aa) fragments are generated that usually do not cover the complete protein sequence and may not include the terminal fragments.
- Successful bottom up analyses utilized multiple enzymes and relied heavily on bioinformatics to assemble and fill in sequence gaps.
- Blocked N-termini are not an issue for bottom up techniques.
- C-terminal sequence information can be obtained.
- MS/MS is unable to differentiate between the isobaric amino acids Ile and Leu and has difficulty differentiating between nominal mass differences of Lys and Gln.

*2 Labs reported complete de novo sequencing of both light and heavy chains using bottom up approach and informatics*

*Bottom up was only method where complete sequencing was obtained*

---

---

# No data is still important data

- 5 labs who attempted bottom up analyses said they got “no data”
    - Were there experimental/instrumental issues? (none reported)
    - Was this an exceptionally hard project?
    - Is that because it is an antibody or because the answer was not in the database?
    - What could have helped these labs get results?
    - Is protein separation an uncommon technique in a proteomics laboratory?
-

# Final conclusions

- Three techniques were successfully employed in this study to obtain N-terminal sequence information of an antibody not present in public databases.
    - Edman Degradation
    - In Source Decay Fragmentation
    - Enzymatic Digestions
  - It was necessary for ISD participants to use T<sup>3</sup> sequencing to obtain true terminal information, extrapolation of standard ISD spectra were not sufficient to call sequence at the termini.
  - Edman analyses required deblocking of the heavy chain using PGAP before sequencing could proceed.
    - Some labs did not perform deblocking due to a) cost, or b) time
  - The most complete *de novo* sequencing was obtained by bottom up participants.
    - Successful bottom up analyses utilized multiple proteases and relied heavily on bioinformatics.
  - All MS analyses were unable to differentiate between isobaric Ile and Leu amino acids, and many had trouble identifying Lys versus Gln.
    - Many participants relied on homology searching to determine which residue to report when in question. Edman participants had no trouble with nominal mass or isobaric residues.
  - Read length of Edman analyses was shorter for light and heavy chains than corresponding MS analyses, especially after heavy chain deblocking.
  - Separation of the antibody chains was necessary prior to most analyses for complexity reduction.
-

---

# Comments from Participants

## *About the sample...*

“... the sample was difficult to resolubilize...”

“...the Light chain sequence was not observed using Trypsin.”

*or* “...unable to detect N-terminal tryptic peptide of light chain sequence.”

“Did not directly find N-terminal peptide from bottom-up analysis. By knowing the N-terminus of the both the heavy and light chain from Edman sequencing, I did an extracted-ion-chromatogram of the N-terminal tryptic peptides for confirmation.”

“... it is our experience that it is much easier to obtain light chain ISD data from intact or reduced antibody samples. With this sample it was much harder to find light chain fragments.”

“There were issues with data processing. After all the modifications done to the protein, it became really time consuming trying to do the data processing when I don't have an in-house search engine. Time was the limiting factor here.”

---

---

# Participant Comments

## *About the study...*

“For de-novo-sequencing of mAbs, a separation of the chains prior to ISD analysis is essential. Possible separation approaches like LC or electroelution are crucial regarding capacity and recovery of the sample. The MS specialists have to develop those methods if they are not routine in their lab. This is time and labor consuming and may keep participants from attending the study. The success of the study depends a lot on the experience handling Abs and not necessarily on analysis methods to be evaluated, could be avoided in future studies.”

“I would like to get more information of the sample before I start the study.”

“I would like to see the correct sequence of this mAb.”

“Differentiation between L and I and between K and Q is not possible using ISD alone. It can be obtained by bottom-up heCID on the same TOF/TOF though, which was not attempted in this study.”

“This was a very good sample to show the advantages and disadvantages of Edman vs MS N-terminal sequencing.”

“Hopefully a robust MS method will be identified as the one to use in the future.”

“I really enjoyed trying to piece together peptides to get the variable region both in Light and Heavy chains. I would like to see the correct sequence of this mAb to evaluate the approach we adopted for mAb characterization. I thank you everybody in the committee for organizing this study.”

---

---

# Suggestions from Participants for PSRG 2011 Study

- Analysis of termini of a complex mixture
  - “N-linked Glycopeptide sequencing, small glycoproteins”
  - “Determine the sensitivity for obtaining N-terminal sequence of a protein. Provide 2 or three amounts of the same protein to compare sensitivity of the various techniques.”
  - “I'd like to see if MS can do as good analyzing a sample from a gel or blot.”
  - “Provide the same peptide on various supports (ie. solution, gel and blot), and ask participants to determine N-terminal sequence by various techniques from each support”
-

---

# Acknowledgements

- *Dr. Robert English* *University of Texas, Medical Branch*
    - Accumulation & anonymization of data
  - *Sigma-Aldrich*
    - Donation of material
  - *Executive Board*
    - For support and scrutiny of study proposal
  
  - ***Participating labs!!!!!!***
-