

Mass Spectrometry-Based Proteomic Applications in Cell/Scaffold Products

Daniel Martin
Institute for Systems Biology

What is proteomics?

“Proteomics includes not only the identification and quantification of proteins, but also the determination of their localization, modifications, interactions, activities, and, ultimately, their function.”

-Stan Fields in *Science*, 2001.

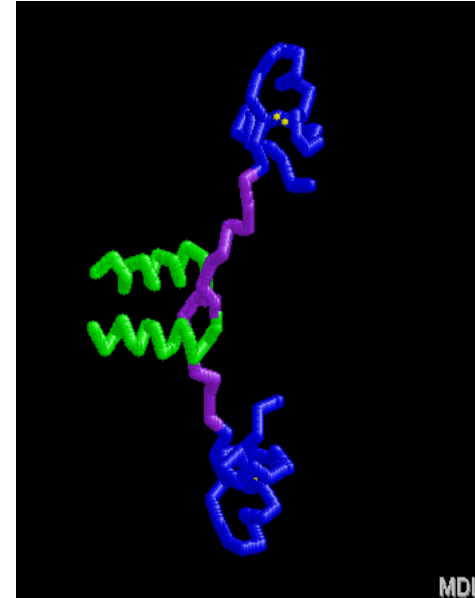
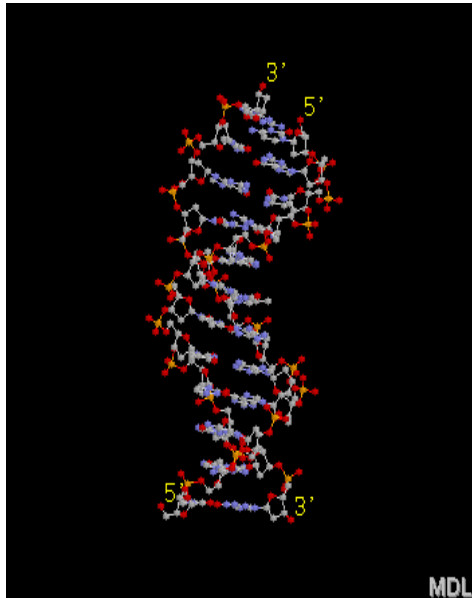
The genome and the proteome: a comparison

Genome

- static
- able to amplify (PCR)
- homogeneous
- no variability in amount

Proteome

- dynamic – condition dependent
- ***no amplification***
- non-homogenous
- high variability in amount ($>10^9$)



Proteomics technologies and methods

- Two-dimensional gel electrophoresis
- mass spectrometry
- protein chips
- yeast 2-hybrid
- phage display
- antibody engineering
- high-throughput protein expression
- high-throughput X-ray crystallography

Mass Spectrometry based proteomics: What it is and what it isn't

What it is:

- A highly powerful tool for protein identification and quantification
- Complementary to other technologies and analysis methods

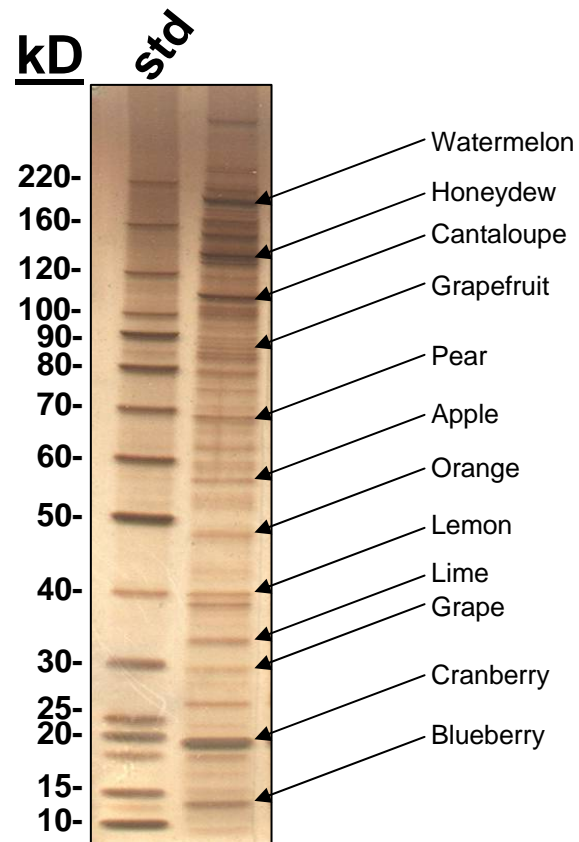
What it is not:

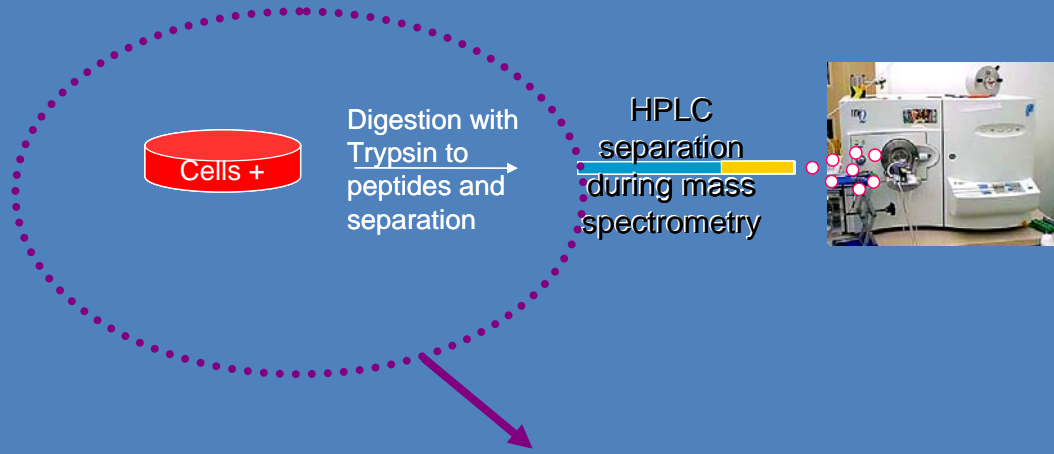
- Magic
- Able to give all the answers
- Simple (relatively speaking)
- Cheap

Mass Spectrometry based proteomics: What can we measure?

- proteins in mixtures
- quantitative analysis of protein expression
- post-translational modifications:
 - phosphorylation (a challenge)
 - glycosylation (present/absent)
- protein interactions

How to think about spectrometry based proteomics





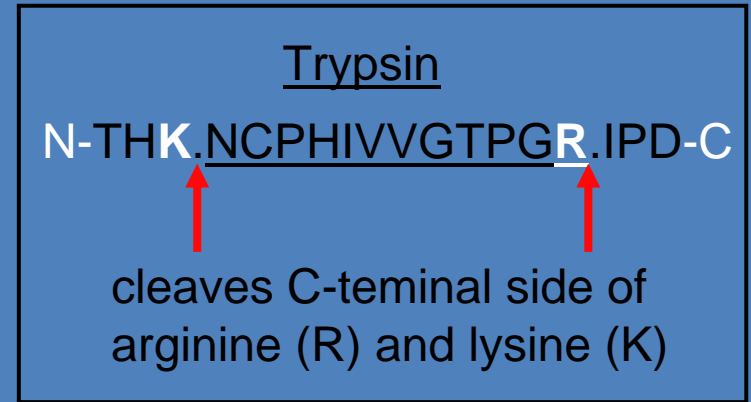
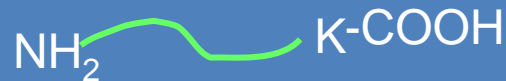
Protein, MW = 10,000 +



digest into peptides

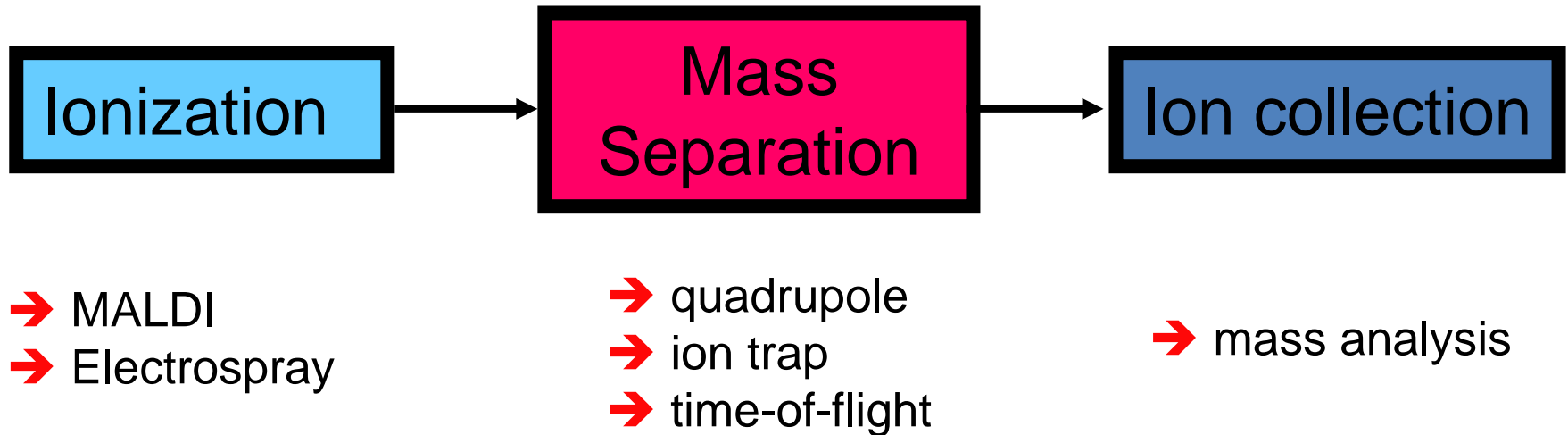


Peptides,
MW < 4,000



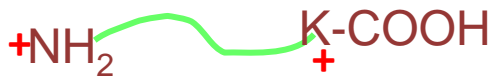
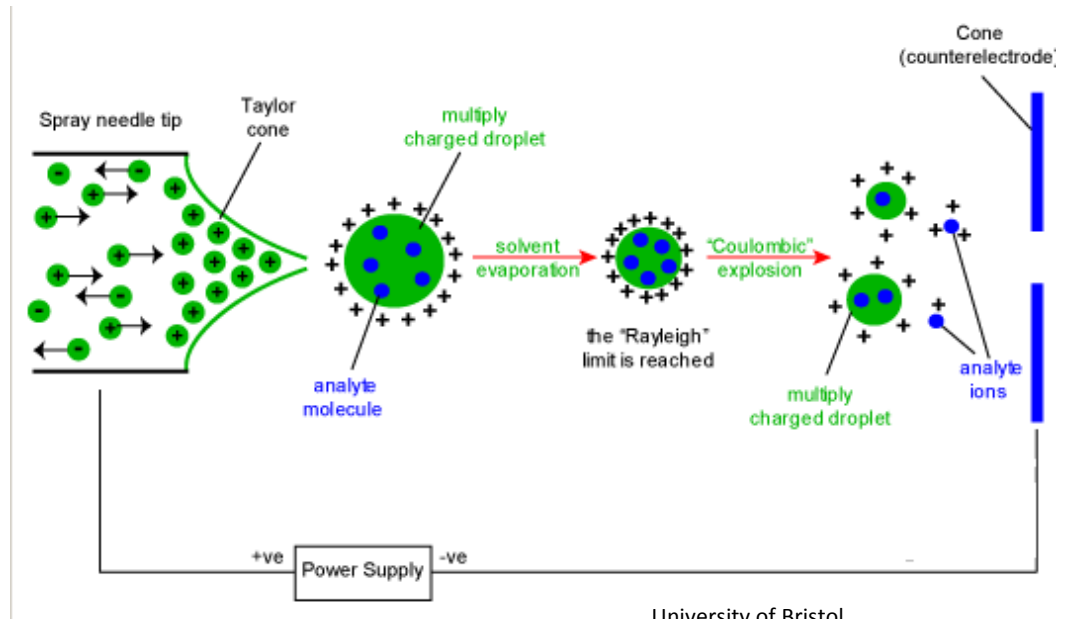
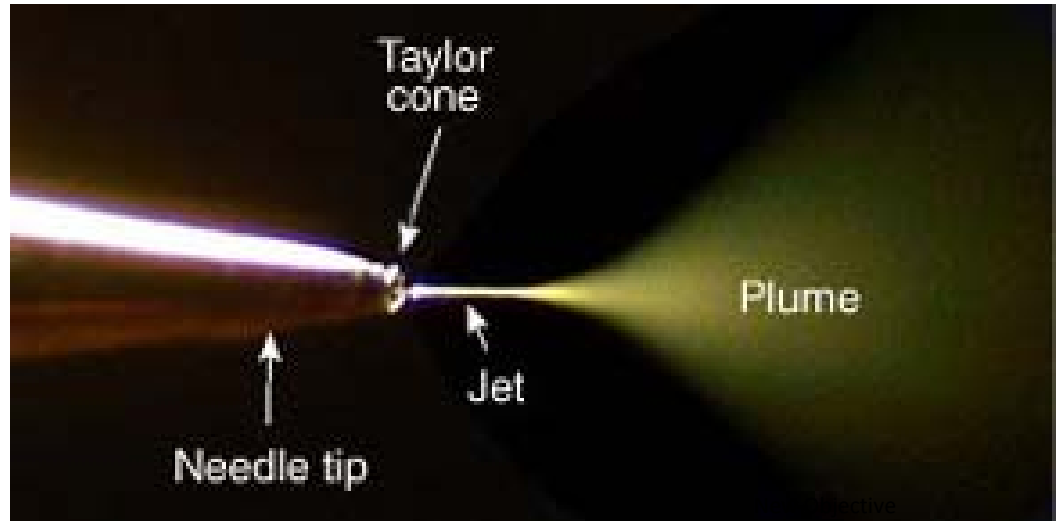
Mass Spectrometry Primer

A mass spectrometer
measures mass to charge ratio
or m/z

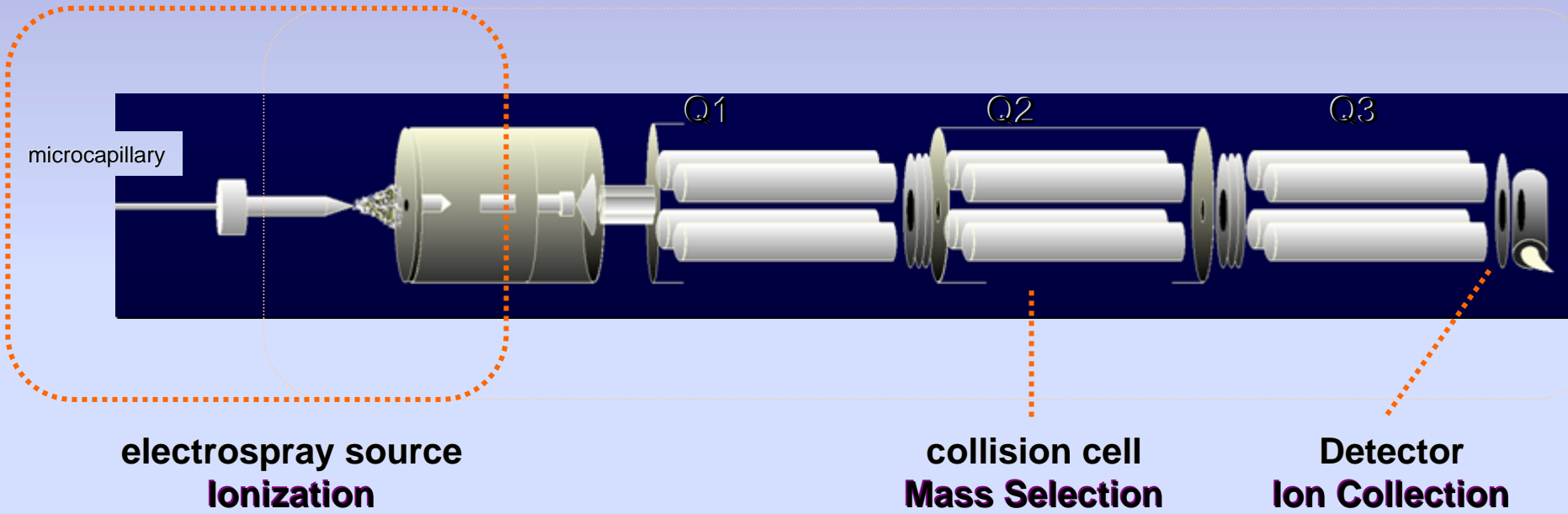


Electrospray ionization (ESI)

High voltage placed on a fused silica column causes a spray of charged droplets which evaporate leaving charged peptides

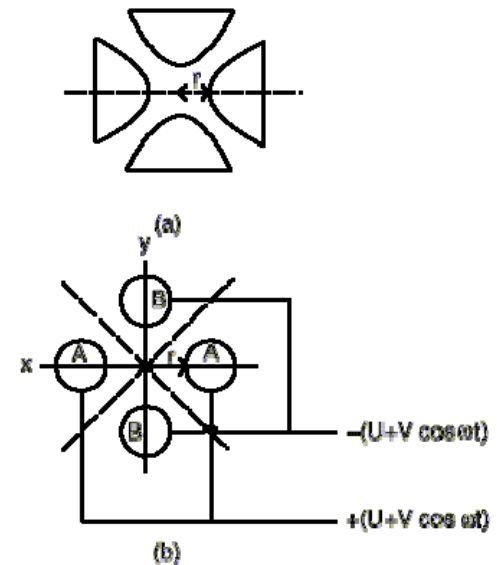
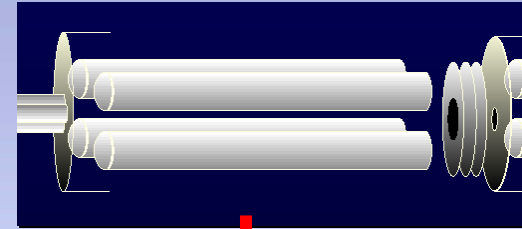


Triple Quadrupole Mass Spectrometer

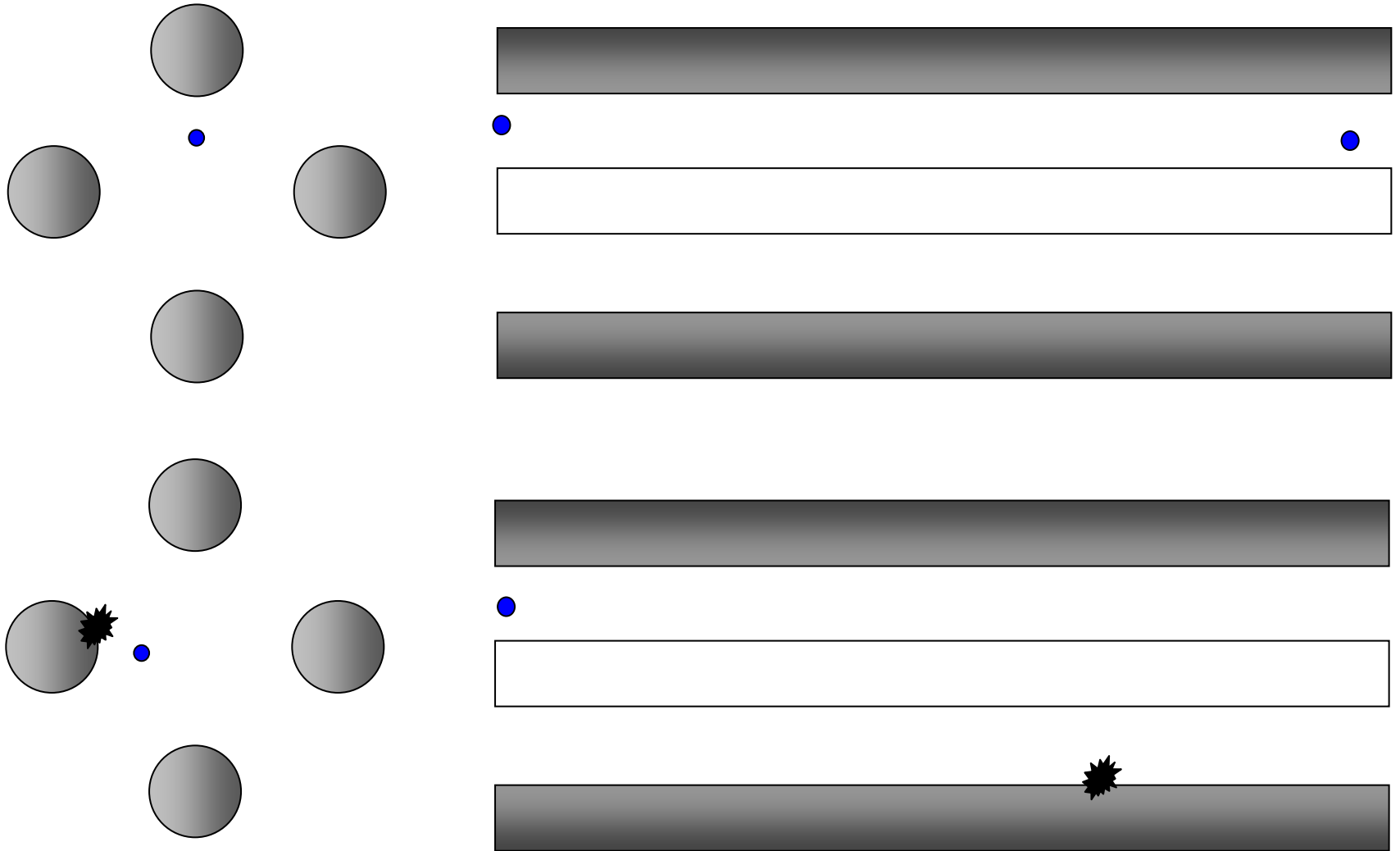


Quadrupole Optics

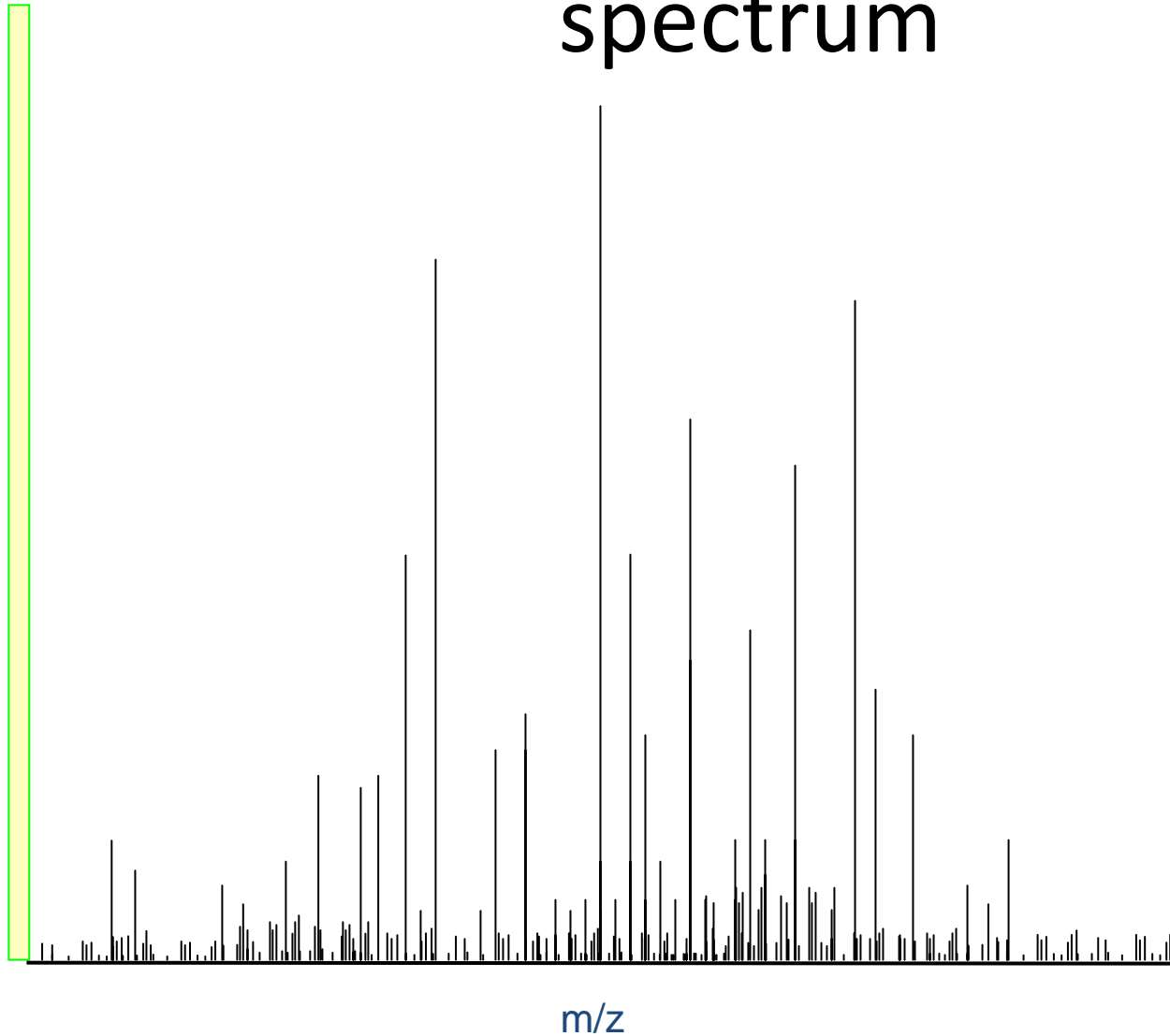
- In a quadrupole mass spectrometer four (quad) parallel rods (poles) are arranged equidistantly from a central (imaginary) axis.
- Charged ions are injected along the central axis of the quadrupole assembly.
- Static and alternating (radio frequency) electric potentials are applied to opposite pairs of rods, creating a fluctuating electric field.



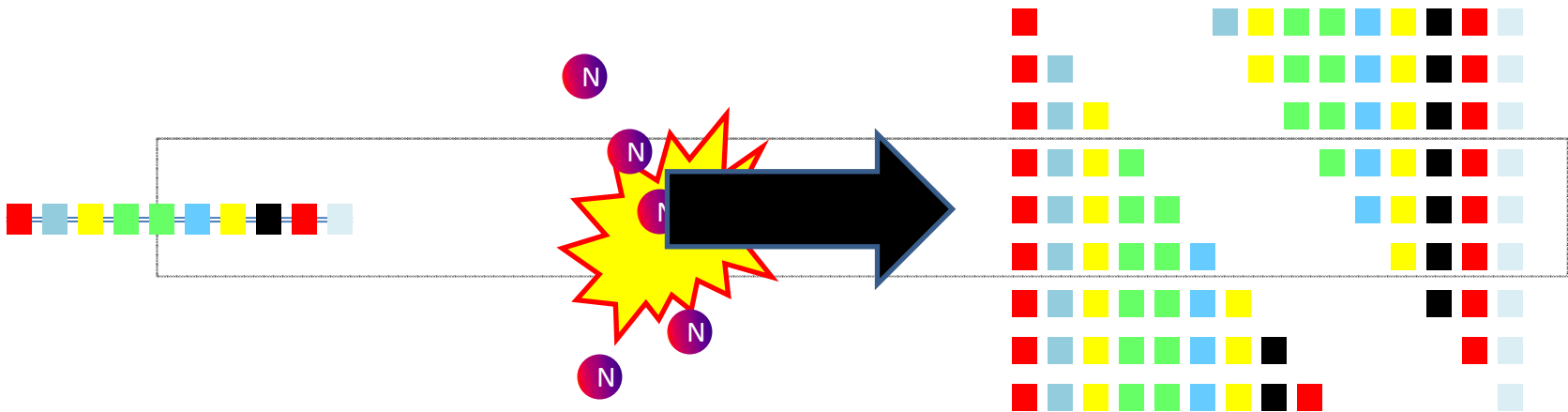
A mass spectrometer can be used as a filter or “tuner”



“Spinning the dial” generates a spectrum



Mass Spectrometers identify peptides by fragmenting along the amide backbone



Fragmentation occurs along the backbone revisited

+NSGDIVNLGSIAGR+

b

y

+N SGDIVNLGSIAGR+
+NS GDIVNLGSIAGR+
+NSG DIVNLGSIAGR+
+NSGD IVNLGSIAGR+
+NSGDIV NLGSIAGR+
+NSGDIVN LGSIAGR+
+NSGDIVNL GSIAGR+
+NSGDIVNLG SIAGR+
+NSGDIVNLGS IAGR+
+NSGDIVNLGSIA GR+
+NSGDIVNLGSIAG R+

Y ions

+NSGDIV

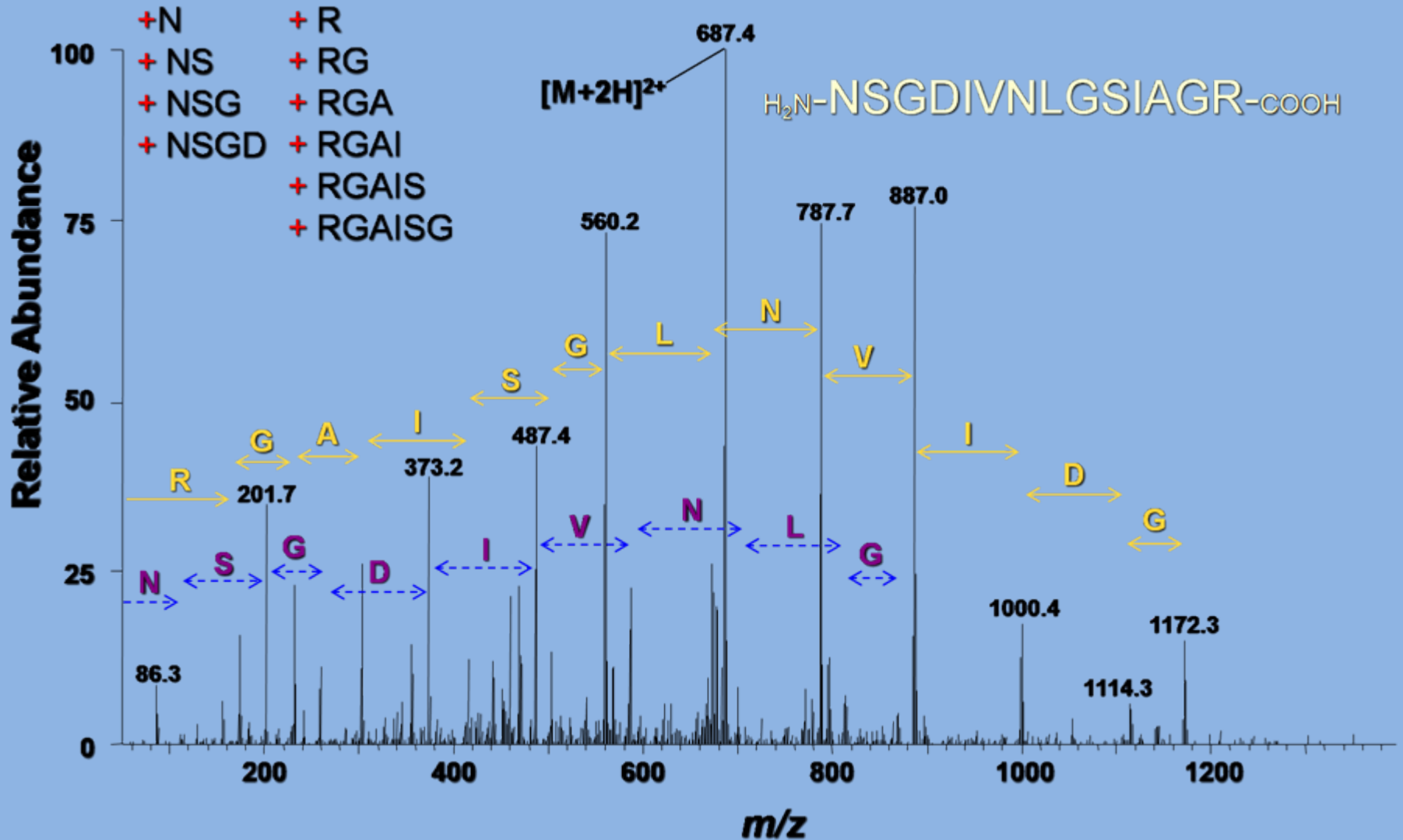
NLGSIAGR+

NLGSIAGR

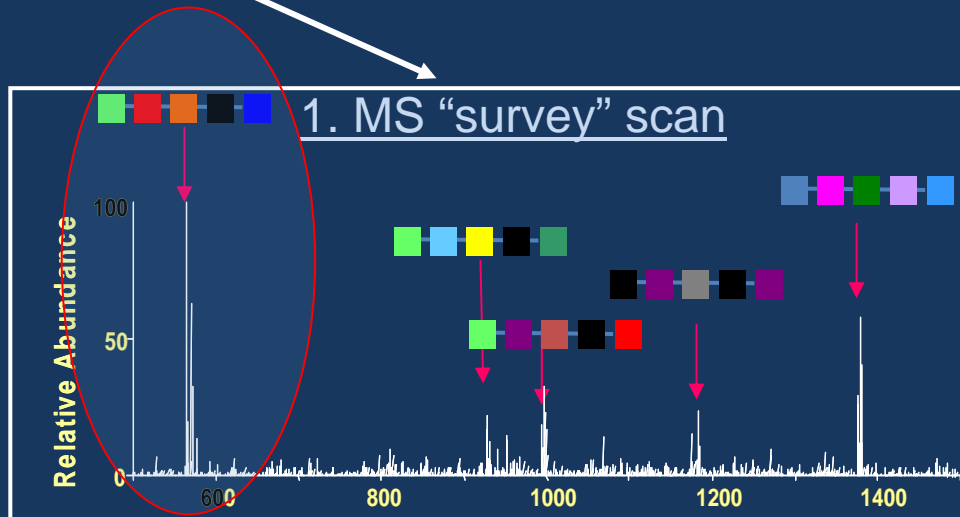
b ions
→
+NSGDIVNLGSIAGR+
←
y ions

- +N
 - + NS
 - + NSG
 - + NSGD
 - + NSGDI
 - + NSGDIV
 - + NSGDIVN
 - + NSGDIVNL
 - + NSGDIVNLG
 - + NSGDIVNLGS
 - + NSGDIVNLGSI
 - + NSGDIVNLGSIA
 - + NSGDIVNLGSIAG
 - + R
 - + RG
 - + RGA
 - + RGAI
 - + RGAIS
 - + RGAISG
 - + RGAISGL
 - + RGAISGLN
 - + RGAISGLNV
 - + RGAISGLNVI
 - + RGAISGLNVID
 - + RGAISGLNVIDG
 - + RGAISGLNVIDGS
- b ions*
- y ions*

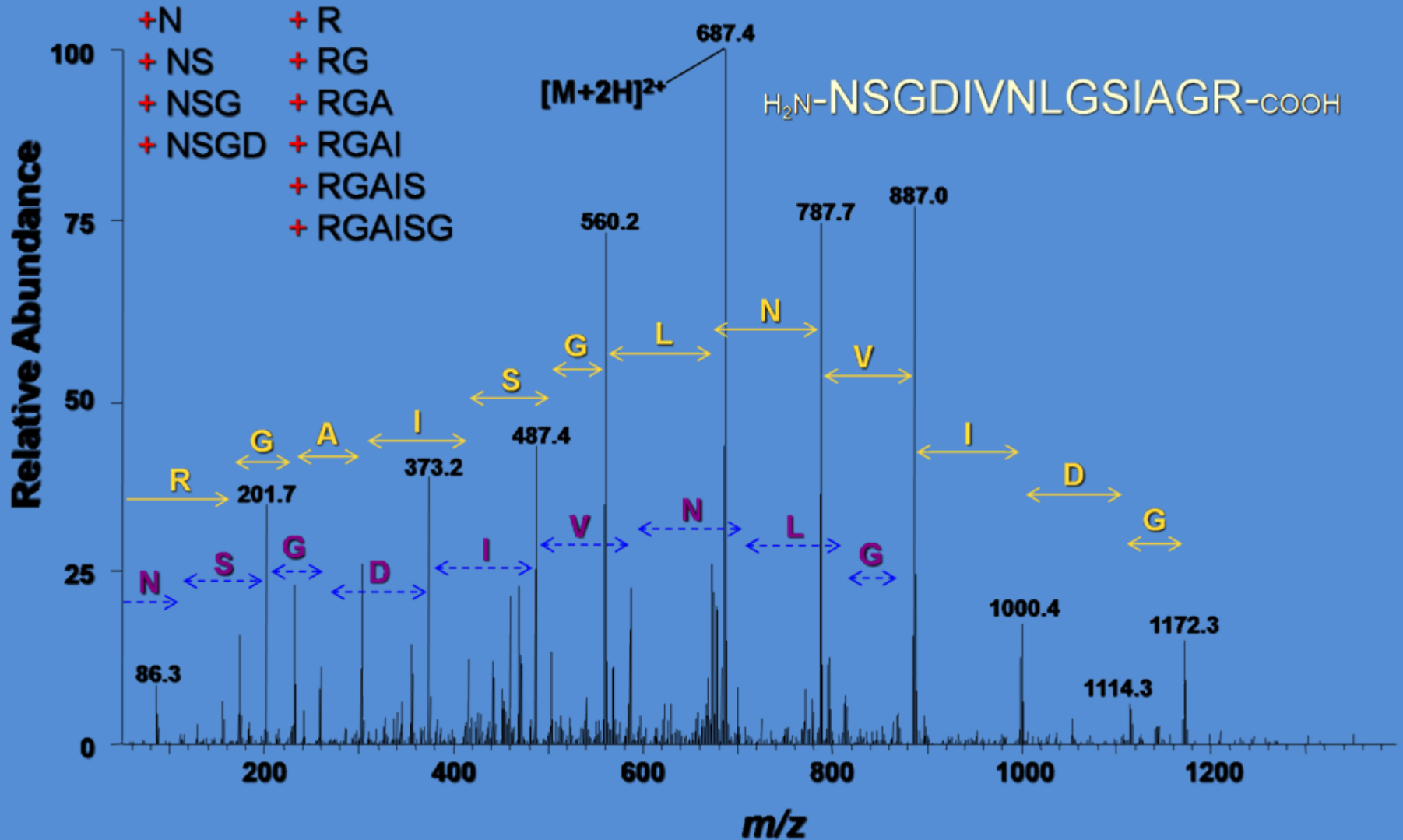
The fragment “ladder” allows identification



Tandem mass spectrometry: "Shotgun Proteomics"



The fragment “ladder” allows identification



Making an identification by database searching using SEQUEST

- SEQUEST is a search program that assigns a peptide sequence to a spectra by comparing it to virtual spectra from a protein database

SEQUEST Example

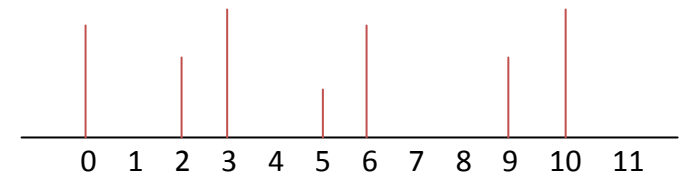
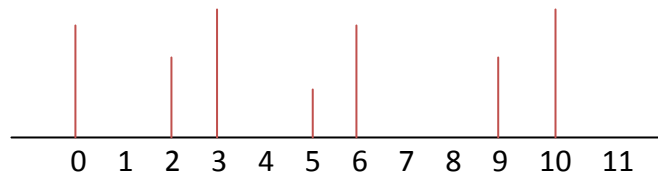
1. An MS/MS scan of m/z 750 and charge 2+ → the molecular weight is 1500 Da
2. SEQUEST searches a protein database starting at the first amino acid to find all possible peptides that weight 1500 +/- 1.5
3. SEQUEST fragments each virtually and compares to the experimental spectra.
4. For a good spectra, ***one peptide stands beats out all others***

Scoring a “match”

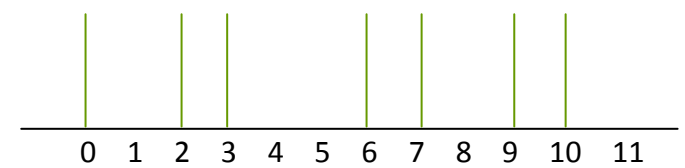
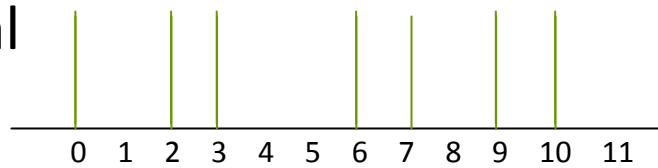
NSGDIVNLGSIAGR

NSADIVNLGSIAGR

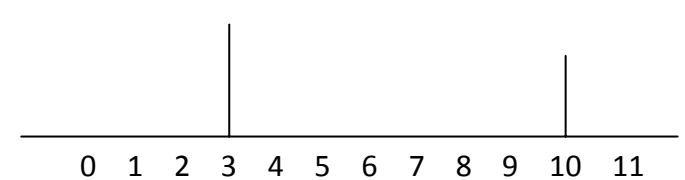
Acquired spectra



Theoretical spectra



Dot product



SEQUEST output file

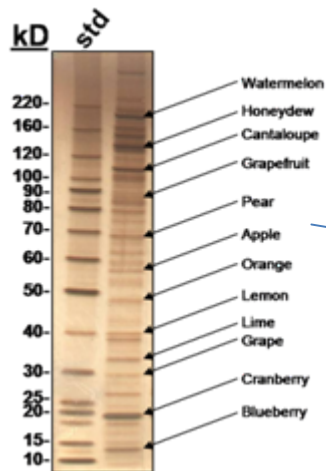
FILE: /data/search/dan/4serum/interact-data.htm J.Eng 04/2000 GO LastUndo
Strict: DelRows: RestoreOrig
Text1: Text2: XCorr: dCn: Sp: InclAA: MarkAA: SortProt
 SortPep

1408	./serum4.1554.1554.3	3333.6 (-1.2)	6.3209	0.361	1252.5	1	38/104	SW:ALBU HUMAN	+9	K.SHC*IAEVENDEMPADLP SLA
1514	./serum4.1702.1702.3	2834.2 (+0.3)	6.8561	0.473	1660.2	1	33/ 84	SW:A2MG HUMAN		R.SLFDLEAENDV LHC*VAF AVP
1008	./serum4.1219.1219.3	2430.6 (-1.2)	5.9068	0.448	1643.3	1	32/ 72	SW:TRFE HUMAN	+4	K.SDNC*EDTPEAGYFAVAVVK.K
1119	./serum4.1302.1302.3	2416.6 (+1.2)	5.3039	0.362	2199.8	1	32/ 60	SW:CFHD HUMAN	+4	K.C*YFPYLENGY QNYGR .K
1108	./serum4.1294.1294.3	2297.6 (-0.4)	4.9588	0.359	1437.1	1	31/ 68	SW:CO4 HUMAN	+6	R.GC*GEQ MIYLAPT LAASR.Y
1112	./serum4.1298.1298.2	2297.6 (-0.6)	4.8816	0.430	1407.7	1	24/ 34	SW:CO4 HUMAN	+6	R.GC*GEQ MIYLAPT LAASR.Y
1067	./serum4.1266.1266.2	2113.4 (-0.8)	4.8355	0.345	2241.4	1	22/ 28	SW:ALBU HUMAN	+11	R.PC*FSALEVD E TYVVK.E
1070	./serum4.1267.1267.3	2113.4 (+0.5)	4.6168	0.345	1260.1	1	31/ 56	SW:ALBU HUMAN	+11	R.PC*FSALEVD E TYVVK.E
1078	./serum4.1272.1272.3	2016.2 (-1.4)	4.5652	0.268*	2251.6	1	27/ 48	SW:ALBU HUMAN	+11	K.QNC*ELF E QLG EYK .F
764	./serum4.1031.1031.3	1703.9 (-0.5)	4.5637	0.305	1894.7	1	27/ 44	SW:HPT1 HUMAN	+9	K.SC*AVAEYGVVVK.V
1379	./serum4.1527.1527.2	2004.3 (+2.6)	4.5351	0.335	1023.4	1	21/ 26	SW:A1BG HUMAN		R.C*EGPI P DVTFELLR.E
1071	./serum4.1268.1268.2	2113.4 (+2.0)	4.5312	0.370	1722.4	1	21/ 28	SW:ALBU HUMAN	+11	R.PC*FSALEVD E TYVVK.E
1295	./serum4.1456.1456.2	2234.6 (-0.5)	4.4580	0.431	1087.2	1	22/ 32	SW:A1BG HUMAN		K.VTLTC*VAPLSG VDF QLR.R
985	./serum4.1202.1202.2	2422.7 (+1.9)	4.4538	0.295	1550.0	1	19/ 28	SW:TRFE HUMAN	+4	K.LC*MGSLNLC*EP NNK .E
1158	./serum4.1338.1338.2	2465.8 (+1.8)	4.3237	0.320	791.6	1	19/ 34	SW:KNH HUMAN	+2	K.LGQSLDC*NAEYV V PWEK.K
1120	./serum4.1303.1303.2	2416.6 (-0.7)	4.2952	0.335	1041.9	1	21/ 30	SW:CFHD HUMAN	+4	K.C*YFPYLENGY QNYGR .K
877	./serum4.1112.1112.2	1936.1 (+2.9)	4.2935	0.272	1246.6	1	18/ 26	SW:TRFE HUMAN	+4	R.FDEFFSEGC*APGSK.K
1166	./serum4.1344.1344.2	2441.7 (+0.4)	4.2876	0.328	652.8	1	24/ 40	SW:A2MG HUMAN	+1	K.AGAF C LSEADAGLISSTASLR
1116	./serum4.1300.1300.2	2297.6 (+2.4)	4.2768	0.386	1346.4	1	22/ 34	SW:CO4 HUMAN	+6	R.GC*GEQ MIYLAPT LAASR.Y
1306	./serum4.1468.1468.3	2814.1 (-0.7)	4.2237	0.002	658.8	1	27/ 84	SW:CERU HUMAN	+3	R.MYSVNGYTFGSLPGLSMC*AED
1126	./serum4.1307.1307.2	2416.6 (+2.3)	4.1085	0.269	1092.9	1	21/ 30	SW:CFHD HUMAN	+4	K.C*YFPYLENGY QNYGR .K
1023	./serum4.1230.1230.2	2015.3 (+2.2)	4.1073	0.396	1189.6	1	21/ 28	SW:IGJ HUMAN	+1	K.C*YTAVVPLVYGG ETK .M
1012	./serum4.1222.1222.3	2015.3 (-0.3)	4.0929	0.164	1386.5	1	30/ 56	SW:IGJ HUMAN	+1	K.C*YTAVVPLVYGG ETK .M
1291	./serum4.1454.1454.2	2234.6 (+1.7)	4.0300	0.258	882.3	1	20/ 32	SW:A1BG HUMAN		K.VTLTC*VAPLSG VDF QLR.R
1294	./serum4.1455.1455.3	2234.6 (+1.0)	4.0248	0.267	1238.3	1	29/ 64	SW:A1BG HUMAN		K.VTLTC*VAPLSG VDF QLR.R
1153	./serum4.1330.1330.3	2297.6 (+0.2)	4.0169	0.224	1330.8	1	31/ 68	SW:CO4 HUMAN	+6	R.GC*GEQ MIYLAPT LAASR.Y
1162	./serum4.1342.1342.2	2465.8 (-0.6)	3.9660	0.295	930.3	1	20/ 34	SW:KNH HUMAN	+2	K.LGQSLDC*NAEYV V PWEK.K
1013	./serum4.1223.1223.2	2430.6 (-0.4)	3.9568	0.268	631.8	1	20/ 36	SW:TRFE HUMAN	+4	K.SDNC*EDTPEAGYFAVAVVK.K
1484	./serum4.1662.1662.2	2529.9 (-0.1)	3.9566	0.205	526.7	1	18/ 36	SW:TRFE HUMAN	+2	R.SAGWNIPIGLLYC*DLPEPR.K

http://198.107.152.2/cgi-bin/displayings_html5?Dta=/data/search/dan/4serum/serum4/serum4.1662.1662.2.dta&MassType=0&NumAxis=1&D

How to think about spectrometry based proteomics

In the end you get a list!



Analysis of an IP of the DNA polymerase II gives 400 proteins

Analysis of cultured yeast yields many thousands of proteins in a day's work

The same can be said for analysis mammalian cells

Be careful what you wish for

Limitations:

- Dynamic Range 10^{3-4}
- Complexity limits analysis
 - MS based proteomics is a sampling method
- Quantification is usually relative rather than absolute

What does this mean?

Mass spectrometry based proteomics:

Is good at:

In depth analysis of pure samples across 3-4 logs of concentration

Cells, organelles, IP's

Is not good at:

Samples with dominant proteins and high dynamic ranges

Serum, plasma, csf, urine

Tissue culture media with serum added

What can mass spectrometry based
proteomics do for me?

Biomarkers!

Biomarkers

- Cellular markers that will tell us a story
 - Markers of desired tissues-identity
 - Markers of scaffold health/consistency
 - Markers of host response

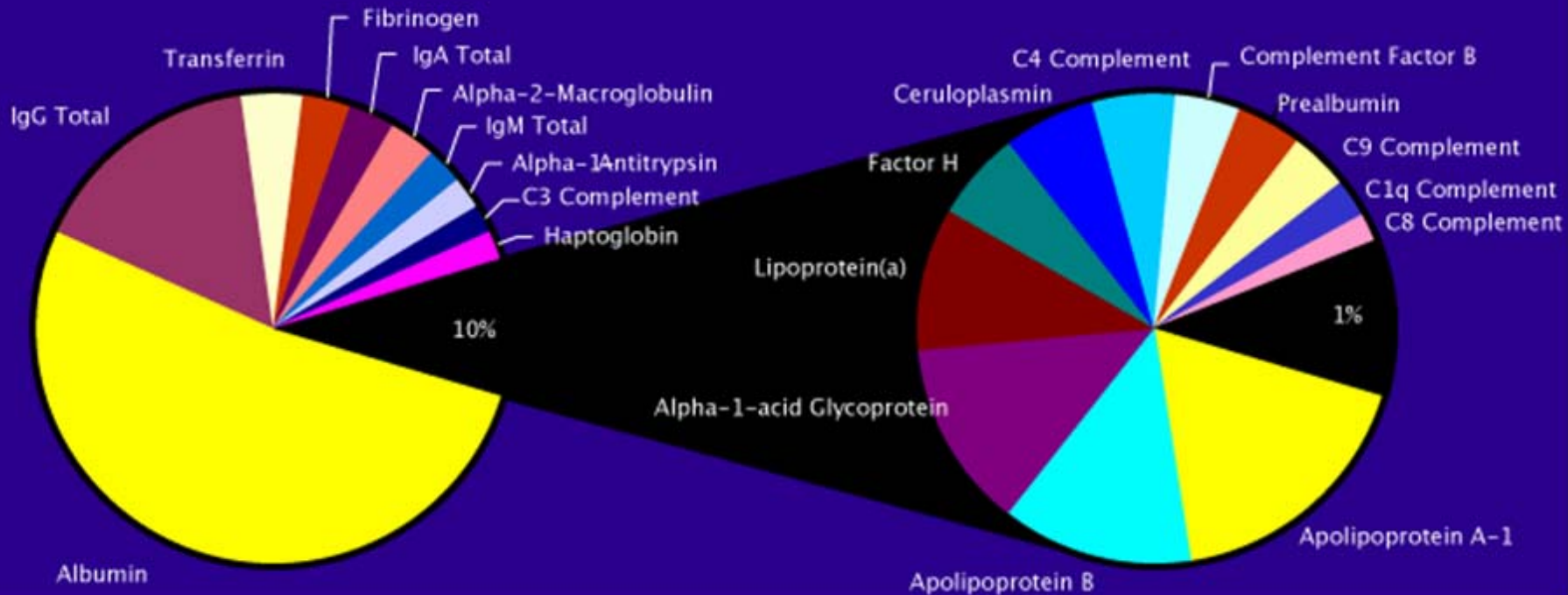
Proteomics and Biomarkers

Expectations have been high

Proteomics is being used to discover new biomarkers

- “Clinical” based discovery strategies
 - Paired clinical samples are probed to determine disease specific differences
 - Example: Compare serum, urine, CSF
- “Target based” validation strategies
 - Tissue or cells are studied to identify targets for later validation in clinical samples or model systems

The Serum Challenge I: A Few Very Abundant Proteins



Serum albumin represents >50% total serum protein itself
22 most abundant serum proteins represent 99% total protein

“Discovery” of low abundance biomarkers is a big challenge!

Analysis of Cells

Protein lists can be assembled-
(you got your wish)

- Many thousands will be seen
- All are potential markers
- This may satisfy some (based on the presentations yesterday)

- If I don't know which proteins are the markers, how many can I follow through time?

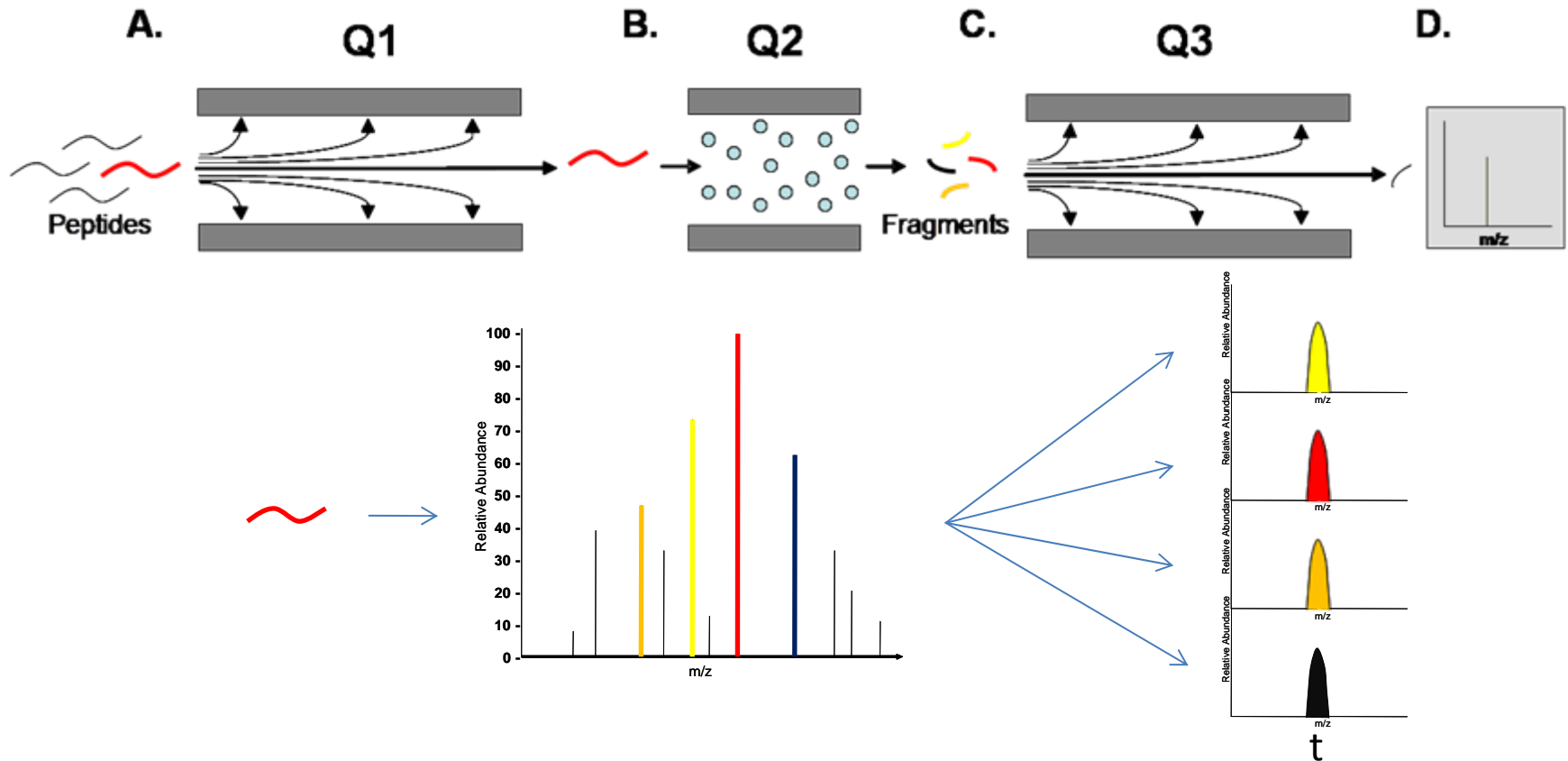
Targeted Proteomics

I have a candidate list

Atlas based proteomics

- The goal: generate and measure a minimal set of “peptide transitions” that completely and non-redundantly represent a proteome
- Uses a variant of mass spectrometry based proteomics called Multiple Reaction Monitoring (MRM)
 - Higher dynamic range
 - Higher sampling speed
 - It is still not magic

Multiple Reaction Monitoring Mass Spectrometry (MRM-MS)

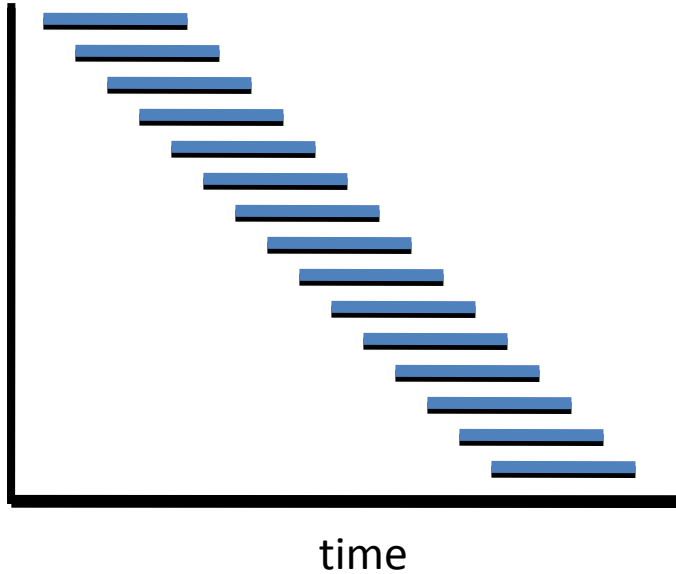


Each measurement “transition” can be made in 10 msec

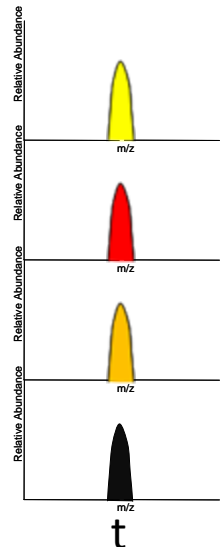
A list of 100 can be cycled once per second

Elution over 20 seconds will have 20 points to generate a quantifiable peak

Scheduled MRMs using a Waters Triple Quadrupole



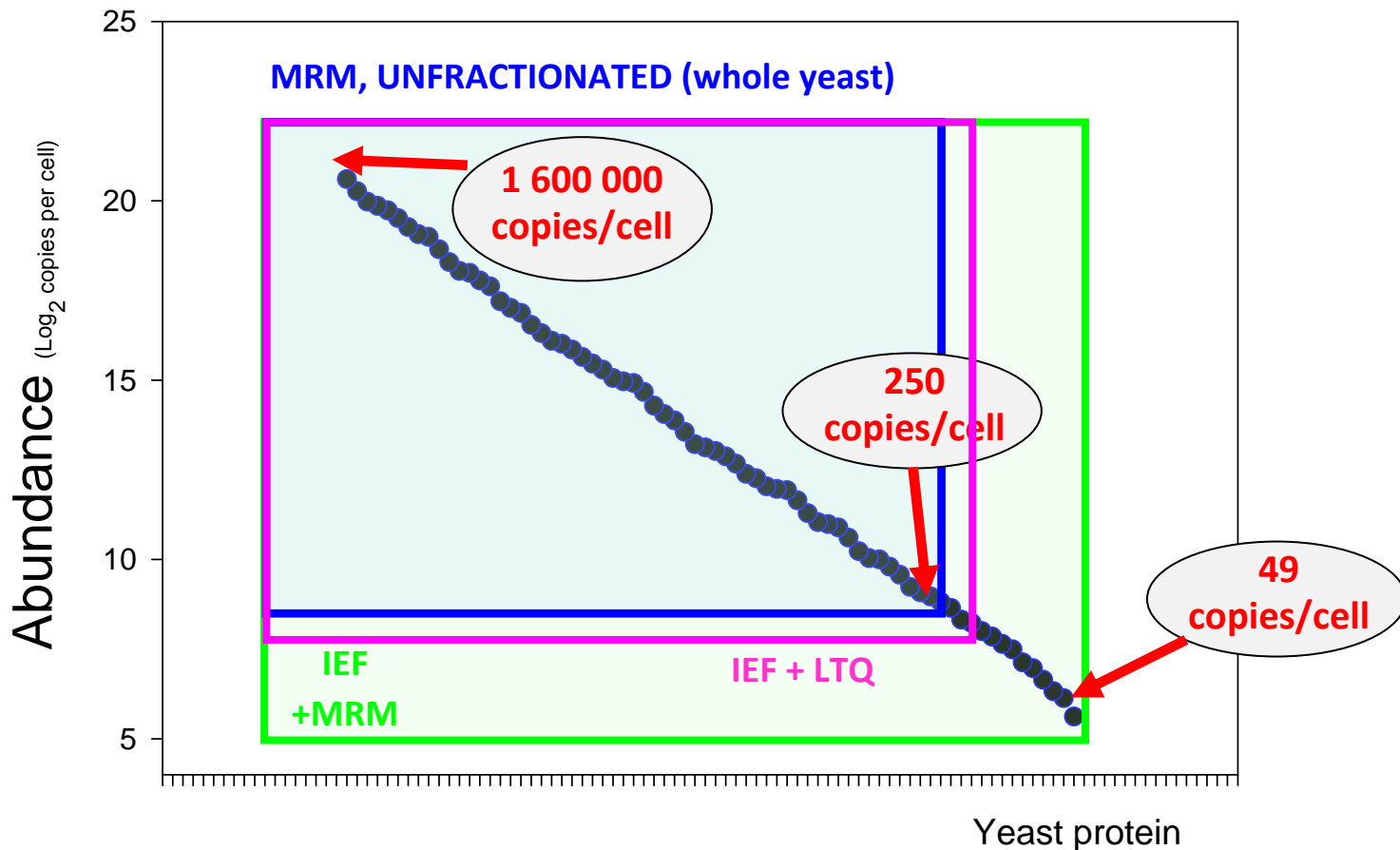
- ◆ 32 segments over a 30 minute run
- ◆ 32 ions / seg
- ◆ **1,024** transitions



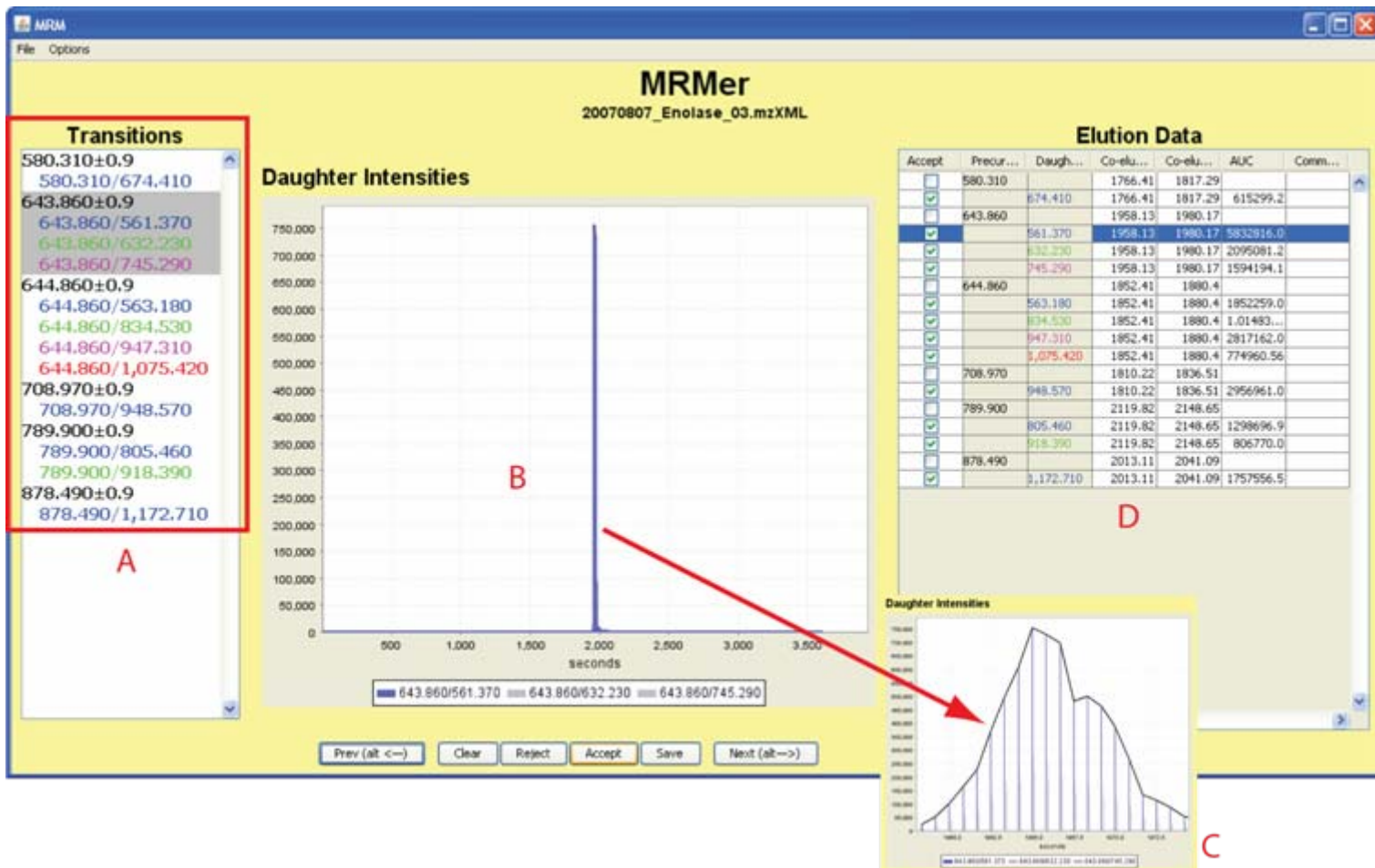
100 proteins per run at a coverage of three peptides per protein and three fragments per peptide.

1 run/hr -> full coverage of 6000 proteins in 3 days

Targeted Analysis of Yeast Proteins by MRM improves sensitivity

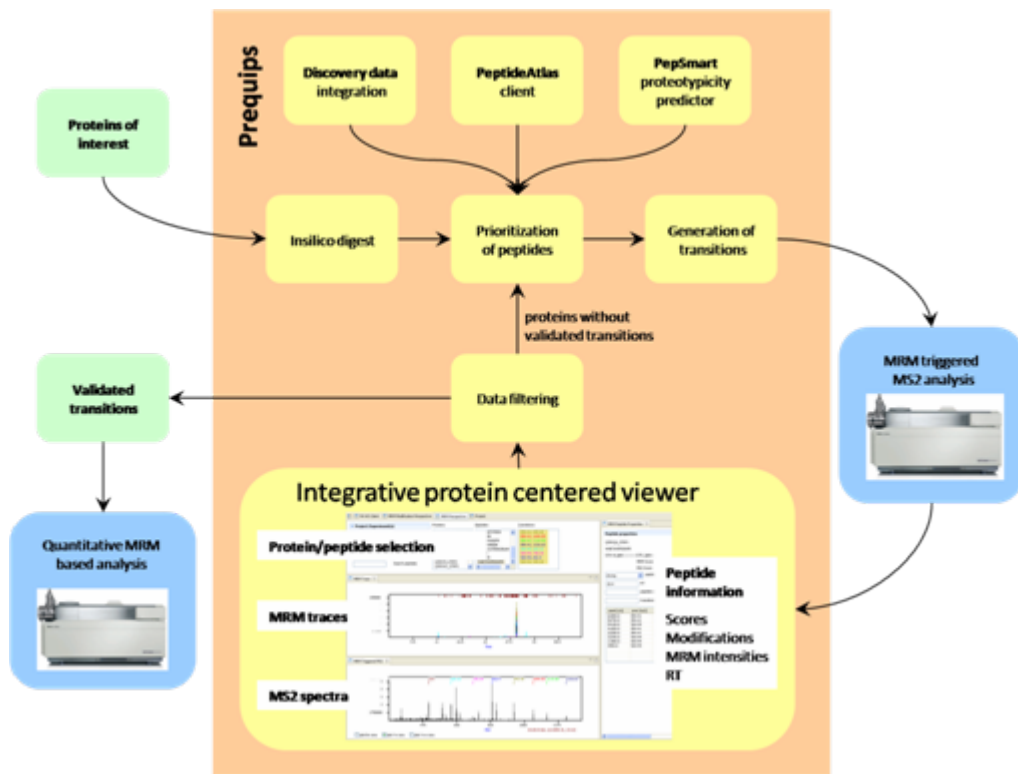


We have developed software to deal with large MRM analysis



PeptideAtlas: Different builds

Build	# Exps	# MS Runs	Searched Spectra	ID P>0.9	Distinct Peptides	Distinct Proteins
Human All	90	1517	3.3 M	334 k	35,391	8000
Drosophila	>100	>1500	~10 M	480k	100,000	> 10.000
Human Plasma	40	39,659	>14 M	660 k	31,953	~3000
Yeast	46	1326	4.1 M	536 k	35 k	> 4000



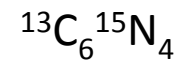
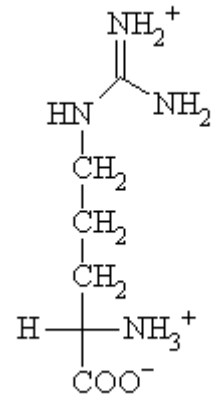
In the near future, we will take the atlases and build (and verify) peptides that represent the entire proteome to generate a “MRM transition atlas”

Implications:

- Shotgun proteomics, with dependency on duty cycle and inherent dynamic range limits, may be superseded by targeted studies
- This may increase the ability to interrogate for a large number of targets in a biological sample
- This may help (but not solve) issues of measuring rare proteins in biofluids (serum, urine, CSF)

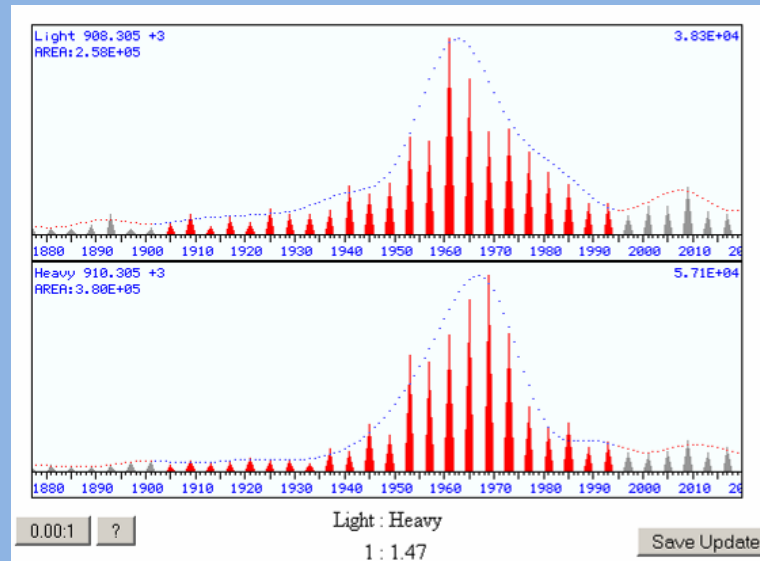
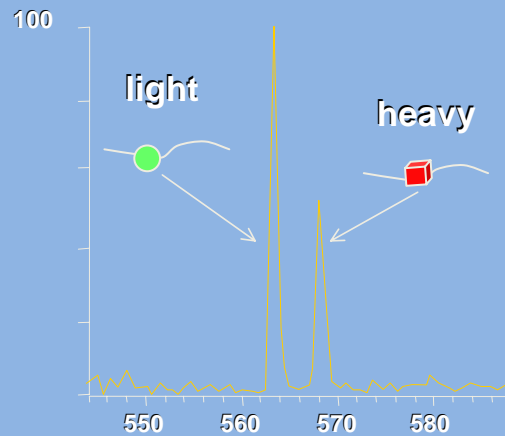
Quantitative Implications

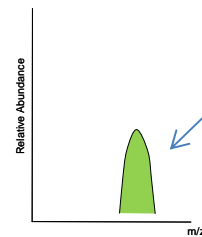
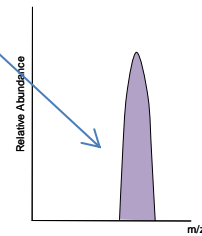
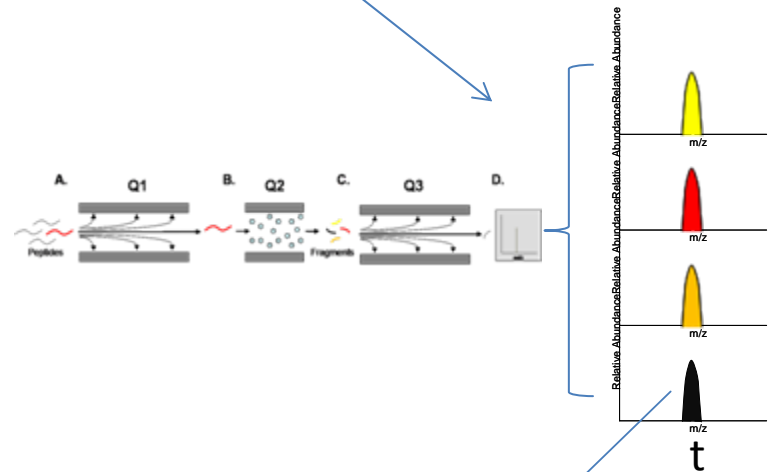
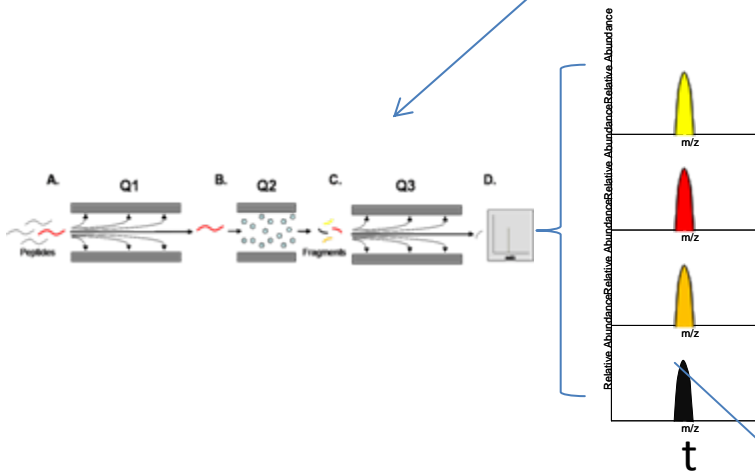
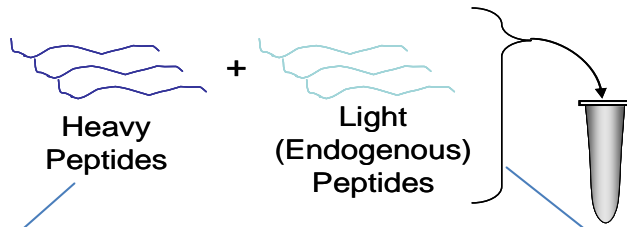
NSGDIVNLGSIAGR



Isotopic modification strategies

- Mass spectrometers measure mass-hence isotopically different peptides can be compared.
- Single ion chromatograms for each isotope can be compared to determine relative quantities of each.



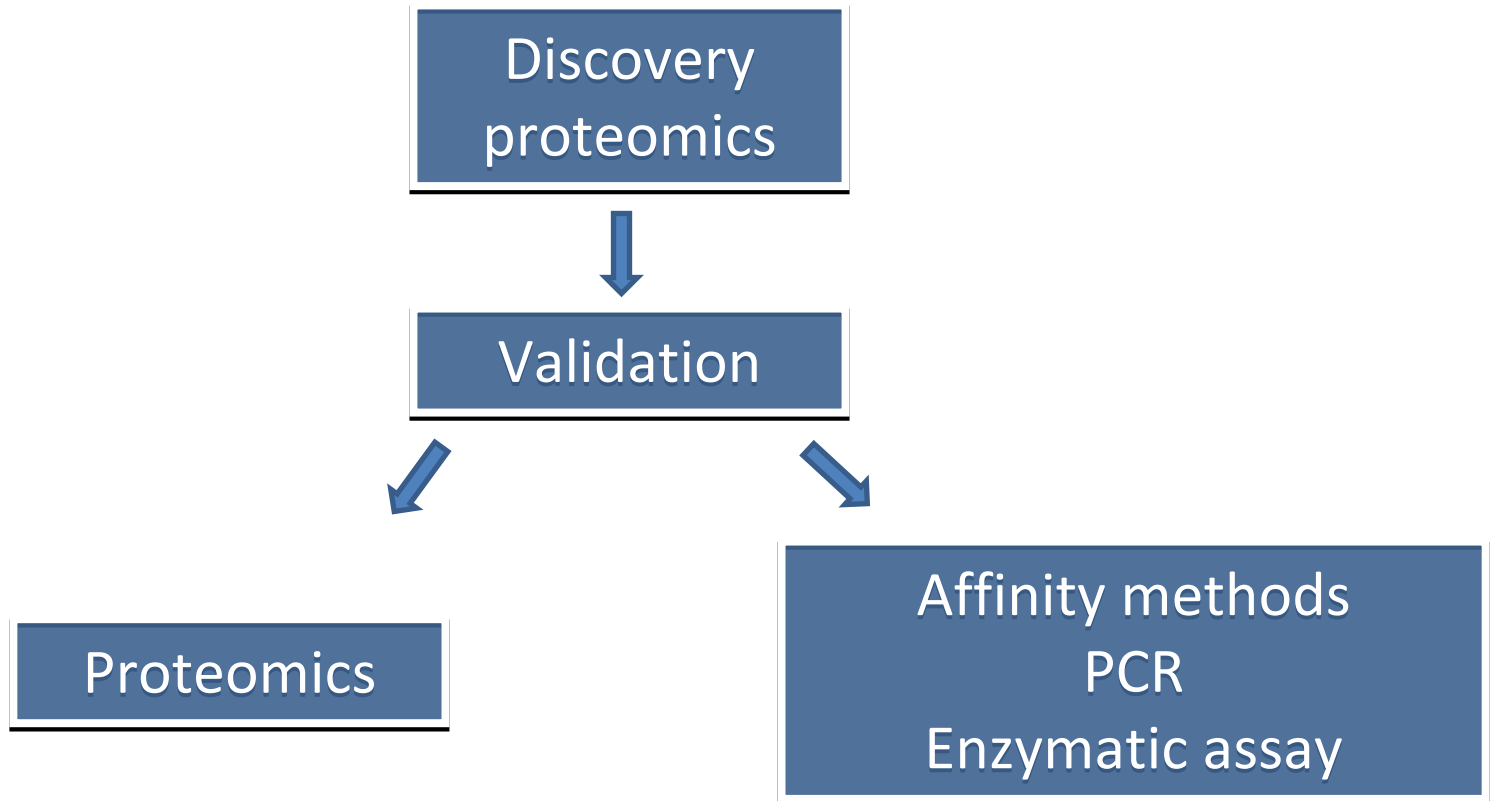


Ratio = 2.0
All fragments for the same sample give the same ratio

How will this play out?

- There may come a day when every one of the unique peptide targets can be paired with an isotopically heavy peptide to determine absolute concentration.

Mass Spectrometry-Based Proteomic Applications in Cell/Scaffold Products



Mass Spectrometry-Based Proteomic Applications in Cell/Scaffold Products

- Discover proteins for in-vitro assays during product development and during production
 - Monitor cells
- Monitor processes at the protein level with targeted mass spectrometry (MRM) or multi-dimensional affinity reagent panel (antibody chip)
 - Monitor growth media during production with targeted analysis/affinity reagent
- Monitor for target protein presence after implantation (likely via affinity reagents)

Proteomics is just a tool

- You can't screw a lightbulb in with a hammer
- There were a lot of if's yesterday
 - If I could measure xxxxxx