

Mass Spectrometry and Proteomics Workshop

Welcome!

OSU, July 11-13, 2016

Proteomics and Mass Spectrometry Workshop, OSU, BRT 115, July 11-13, 2016

Time	Monday, 7/11	Tuesday, 7/12	Time	Wednesday, 7/13
9:00-10:20 am	Welcome and Introduction to Proteomics (Michael Freitas)	Sample Preparation II (Cindy James); MALDI Imaging (Nan Kleinholz)	9:00-10:20 am	Mass Analyzers (Arpad Somogyi); FT-ICR Instrumentation and Techniques (Michael Freitas)
10:20-10:35 am	Break	Break	10:20-10:35 am	Break
10:35-12:00 pm	Sample Preparation I (Protein Purification) (Cindy James)	Protein Quantitation (Liwen Zhang)	10:35-12:00 pm	HPLC MS/MS and Metabolomics (Yu Cao)
12:00-12:30 pm	Lunch Break	Lunch Break	12:00-1:00 pm	Lunch Break
12:30-2:00 pm	Lab visit (Group 1)	Lab visit (Group 2)		
2:00-2:50 pm	Introduction to Mass Spec, Ionization Methods (Arpad Somogyi)	New Instruments and Service (Liwen Zhang); Introduction to Peptide Manual Sequencing (Arpad Somogyi)	1:00-2:30 pm	Small Molecule Analysis and Spectrum Interpretation (Arpad Somogyi)
2:50-3:05 pm	Break	Break	2:30-2:45 pm	Break
3:05-4:00 pm	Proteomics (Post Translational Modifications) (Liwen Zhang)	Peptide Manual Sequencing (continued) (Arpad Somogyi)	2:45-4:00 pm	GC-MS and Other Spectrum Interpretations (Arpad Somogyi) Open Discussion

Lecture only:

**boring, limited active
learning**



Goal:

**attentive class, active
learning**



Please interrupt!!

(With questions and comments)

- not with ringing cell phones



We've planned some interactive "learning checks" to keep you engaged.

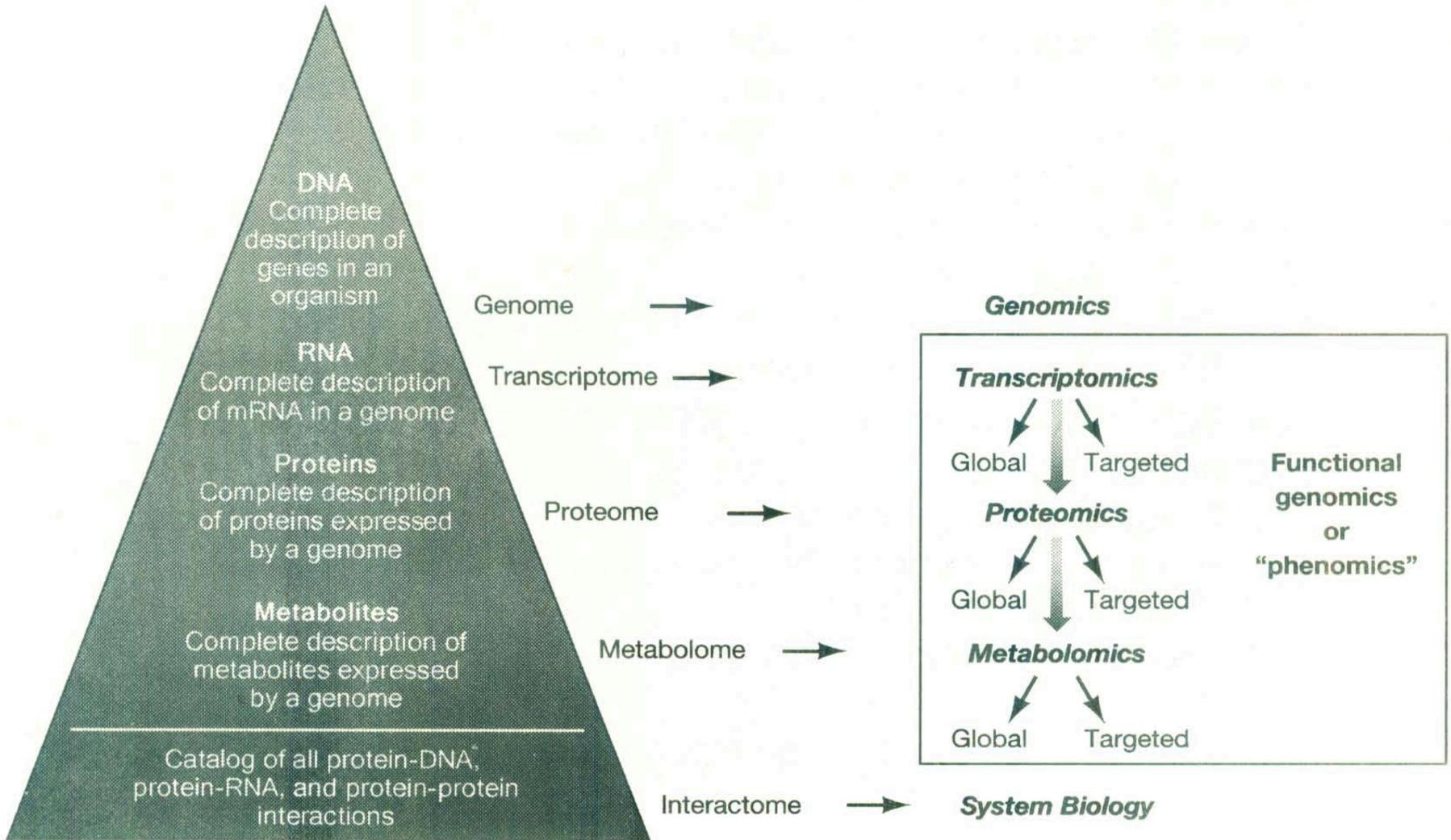
Protein Mass Spectrometry (Proteomics)

Liwen Zhang, Michael A. Freitas

Mass Spectrometry and Proteomics Facility
The Ohio State University

Summer Workshop 2015

Biology of Systems (aka Systems Biology)



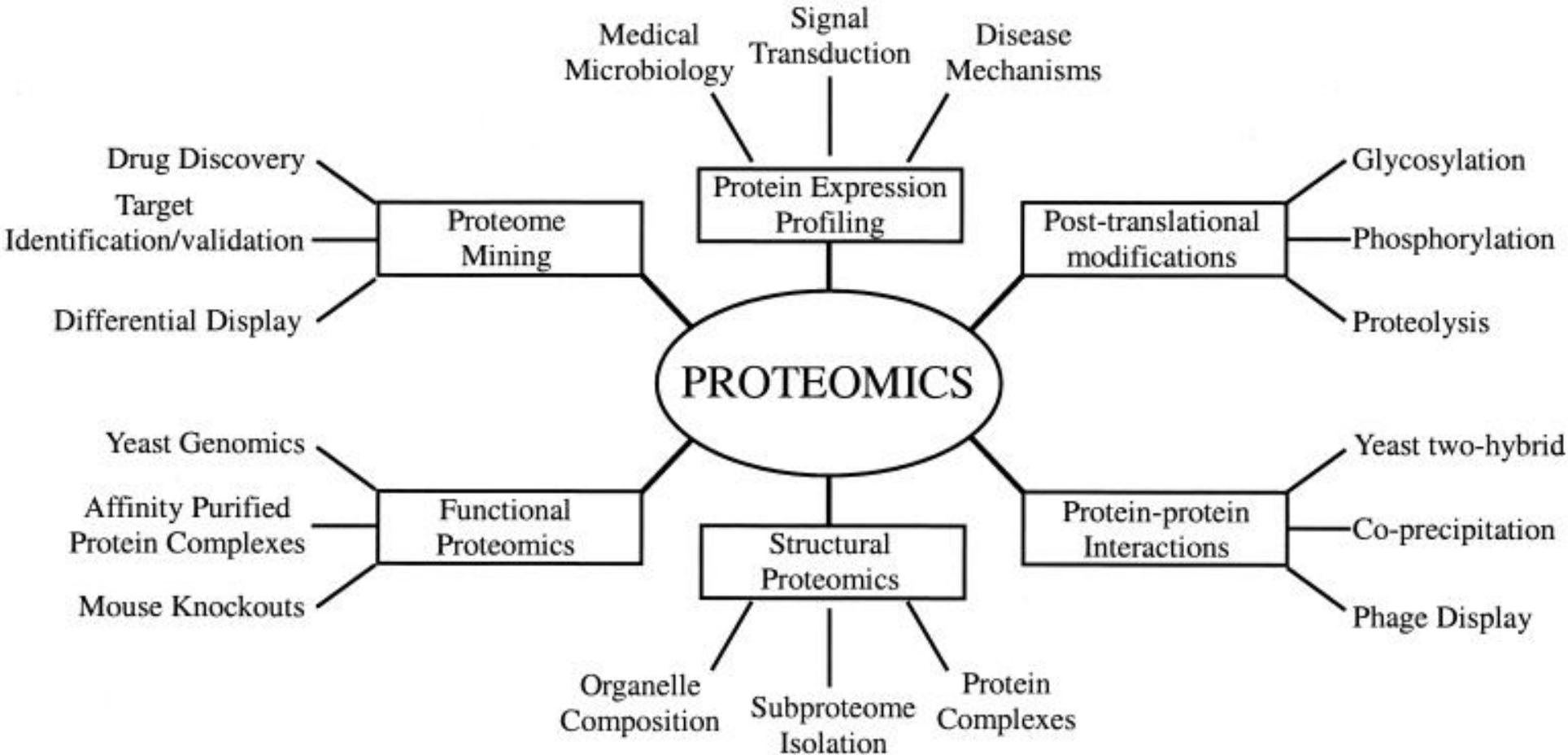
Proteomics:

The term “proteomics” was first coined in 1995

defined as the large-scale characterization of the entire protein complement of a cell line, tissue, or organism.

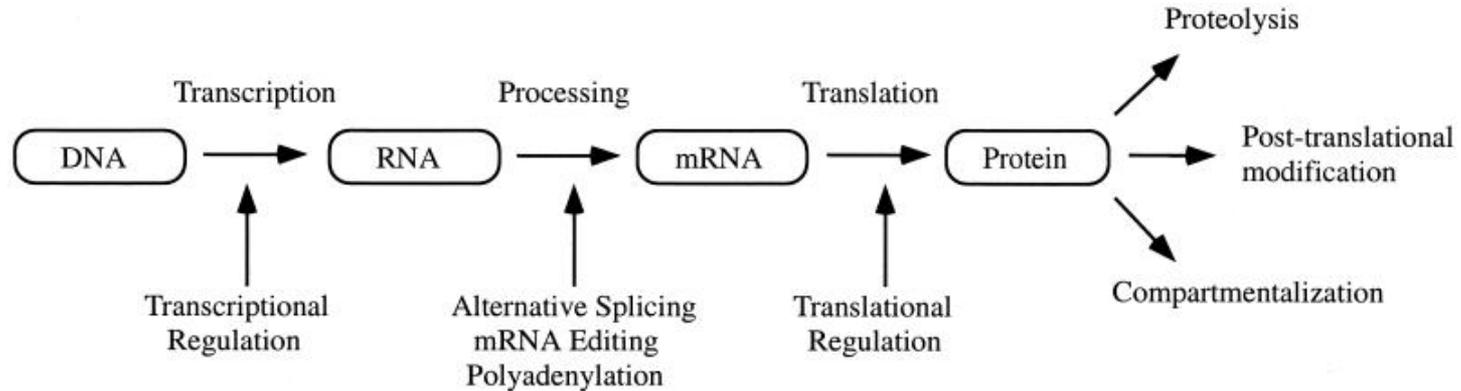
Wasinger, V. C., S. J. Cordwell, A. Cerpa-Poljak, J. X. Yan, A. A. Gooley, M. R. Wilkins, M. W. Duncan, R. Harris, K. L. Williams, and I. Humphery-Smith. 1995. Progress with gene-product mapping of the *Mollicutes: Mycoplasma genitalium*. Electrophoresis 16:1090-1094.

Proteomics:



Genomics inform proteomics:

Why not just look for genetic differences?



Genetics and transcriptomics inform what protein products may be produced in the cell.

The actual protein composition is affected by many post-transcriptional processes:

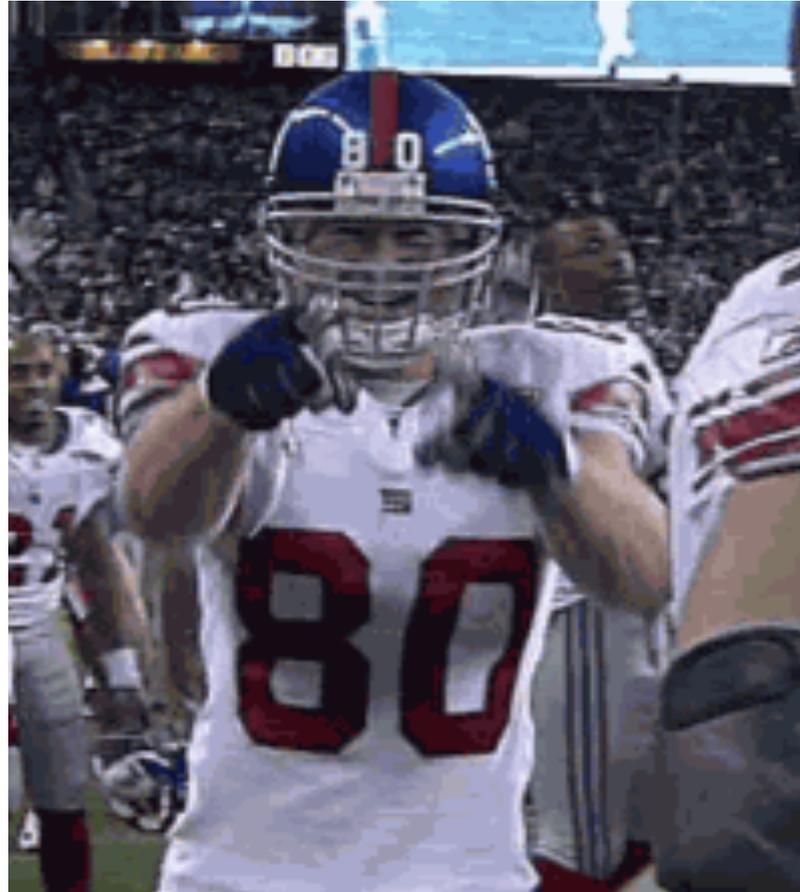
- Translational regulation due to miRNAs
- Post translational modifications
- Protein uptake from microenvironment
- Protein Stability

Proteomic is like taking a snapshot of football game and inferring the outcome.



Jeremy Shockey, celebrating a NY Giants game winning 40 yard field goal.

Proteomic is like taking a snapshot of football game and inferring the outcome.



That Missed ... Wide Left!

Take Home Points

Proteomics technologies are needed to characterize the gene products as follows:

A. Identification

B. Localization

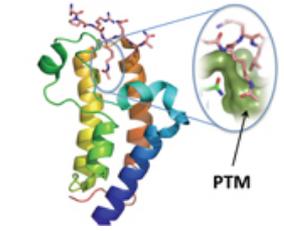
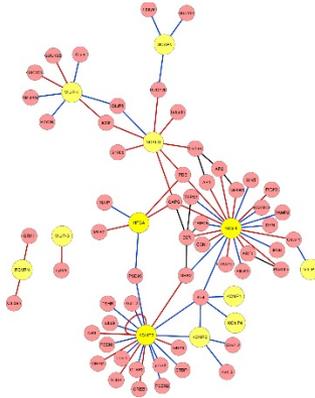
B. Characterization (modifications, stability, transport, localization, etc.)

C. Quantification

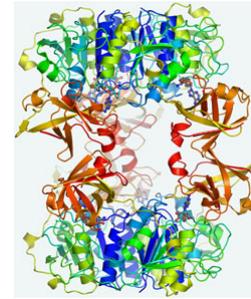
**How we measure is as important
as what we measure!**

What Can Proteomics Do

Protein Interaction Analysis

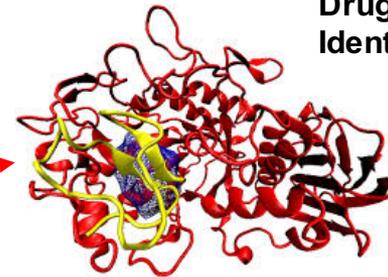


PTM Identification



Protein Crosslink Identification

Drug Binding Site Identification

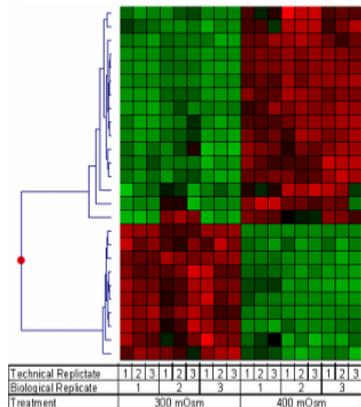
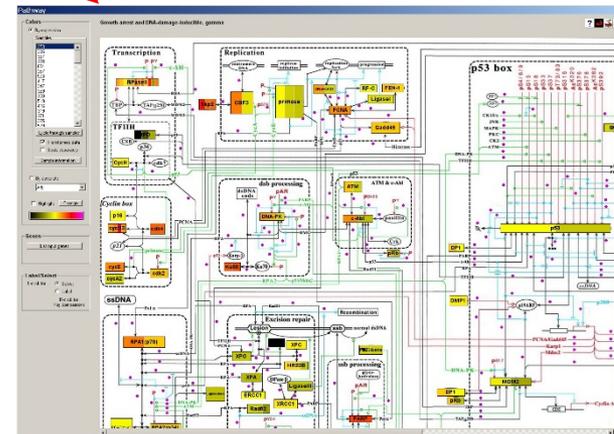


Protein(s) identification

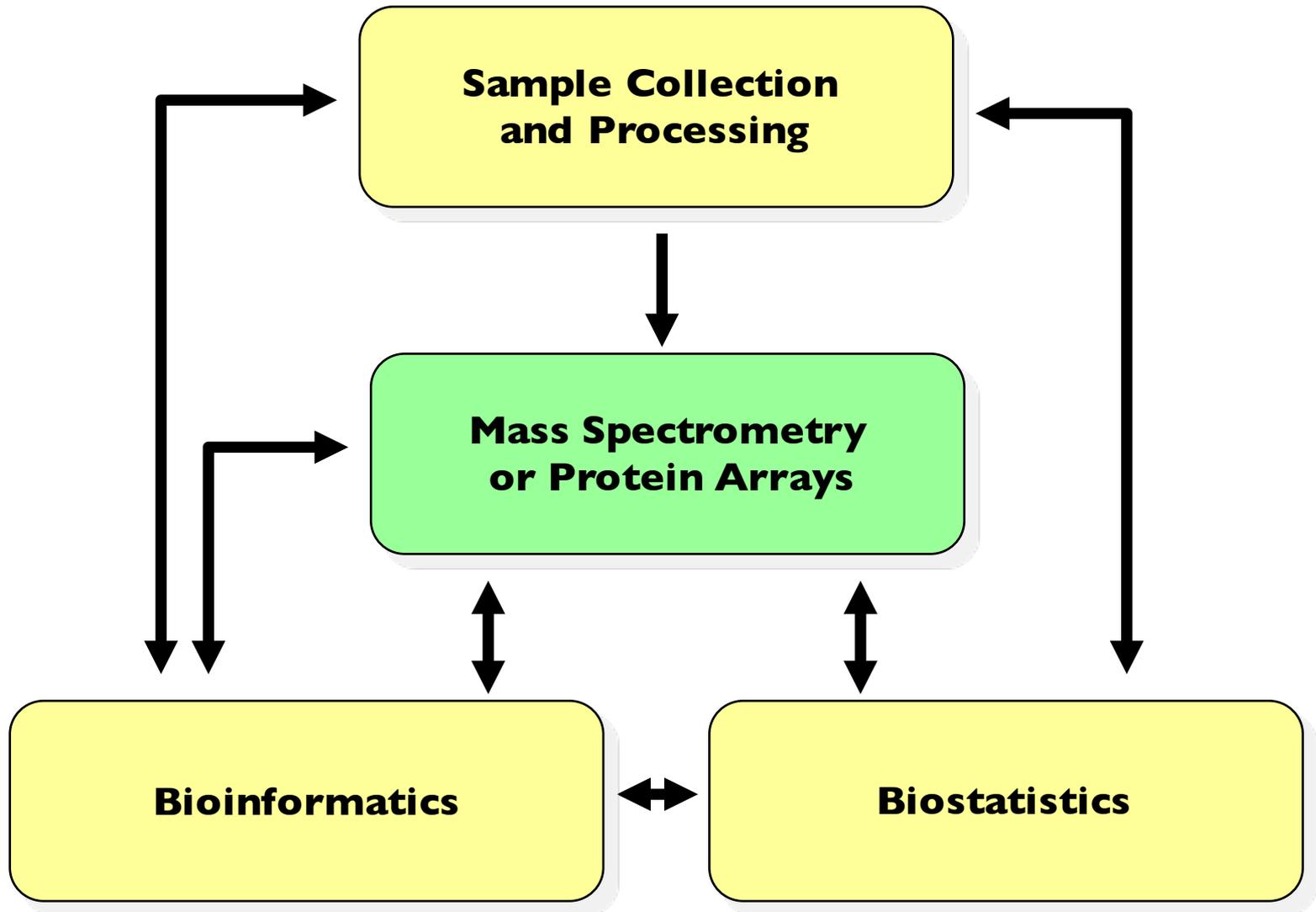
Protein Profile Analysis

Pathway Analysis

Uniprot Accession Number	Protein Name	300 mOsm (Mean x 10 ⁴)	400 mOsm (Mean x 10 ⁴)	Ratio	P-value
P26706	ATP synthase subunit alpha, mitochondrial	0.32	0.59	1.79	2.52E-08
Q99079	Histone H2B type F-S	0.47	0.81	1.68	1.21E-10
P14618	Pyruvate kinase isozymes M1/M2	1.69	2.96	1.73	1.37E-30
P15121	Aldose reductase	8.99	32.14	3.52	<1.0E-45
P15180-2	Isoform M1 of Pyruvate kinase isozymes M1/M2	23.76	40.41	1.68	<1.0E-45
P53895	Aldo-keto reductase family 1 member C2	0.38	1.89	4.93	<1.0E-45
P04179	Superoxide dismutase [Mn], mitochondrial	0.16	0.72	4.30	4.20E-45
P40939	Trifunctional enzyme subunit alpha, mitochondrial	0.05	0.12	2.61	2.34E-31
P61026	Ras-related protein Rab-10	0.37	0.58	1.56	4.34E-20
P29401	Transketolase	0.27	0.46	1.62	2.09E-21
P30613	Pyruvate kinase isozymes RA	1.48	2.49	1.65	2.03E-19
Q9N0C3	Reticulon-4	0.41	0.69	1.68	4.13E-18
Q99536	Synaptic vesicle membrane protein VAT-1 homolog	0.19	0.40	2.07	2.23E-31
Q18695	Histone H3.1	0.35	0.63	1.79	9.11E-07
P06703	Protein S100-A6	0.15	0.50	3.26	1.06E-08
Q15942	Zyxin	0.23	0.36	1.51	0.0088
P02751	Fibronectin	16.31	2.99	0.18	<1.0E-45
P19106	Myosin regulatory light chain 12A	0.73	0.46	0.63	9.90E-22
P42224	Signal transducer & activator of transcription 1- α B	0.19	0.06	0.26	<1.0E-45
Q04692	Procollagen-lysine 2-oxoaldehyde 5-dioxygenase 2	0.17	0.06	0.35	7.05E-31
P21559	5-nucleotidase	3.04	1.87	0.62	7.70E-40
P40761	Nicotinamide N-methyltransferase	0.26	0.12	0.46	4.46E-34
O15480	Profilin-4-hydroxylase subunit alpha-2	0.55	0.21	0.38	<1.0E-45
P13674	Profilin-4-hydroxylase subunit alpha-1	0.14	0.06	0.42	6.85E-27
Q9Y696	Chloride intracellular channel protein 4	0.61	0.39	0.65	6.10E-15
Q01995	Transgelin	7.32	4.34	0.60	9.63E-21



Proteomics



What is Mass Spectrometry?

Mass spectrometry is an analytical technique used to measure the mass of gas phase ions

Typical mass spectrometer consists of the following:



Source – vaporization / ionization of the molecules

Mass Analyzer - separation of gas-phase ions by mass

Detector - detection of mass separated ions

ESI and MALDI

for Ionization of Biopolymers

Q) Why are ESI and MALDI so useful in the analysis of biopolymers?

A) Because: They are “soft” ionization techniques capable of generating ions of nonvolatile biopolymers.

ESI and MALDI

The Royal Swedish Academy of Sciences has decided to award the Nobel Prize in Chemistry
for 2002



”for the development of methods for identification and structure analyses of biological macromolecules”

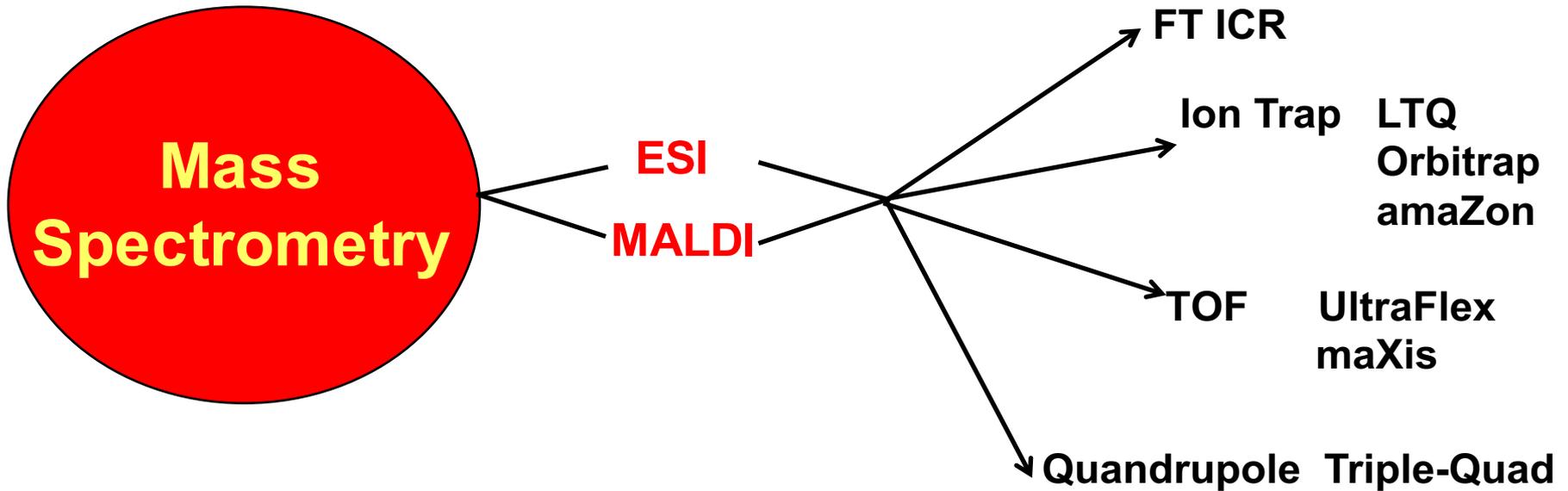
John B. Fenn, Virginia Commonwealth University, Richmond, USA, and
Koichi Tanaka, Shimadzu Corp., Kyoto, Japan

”for their development of soft desorption ionisation methods for mass spectrometric analyses of biological macromolecules”

Kurt Wüthrich, Swiss Federal Institute of Technology (ETH), Zürich, Switzerland and The Scripps Research Institute, La Jolla, USA

”for his development of nuclear magnetic resonance spectroscopy for determining the three-dimensional structure of biological macromolecules in solution”.

All types of hardware used in proteomics



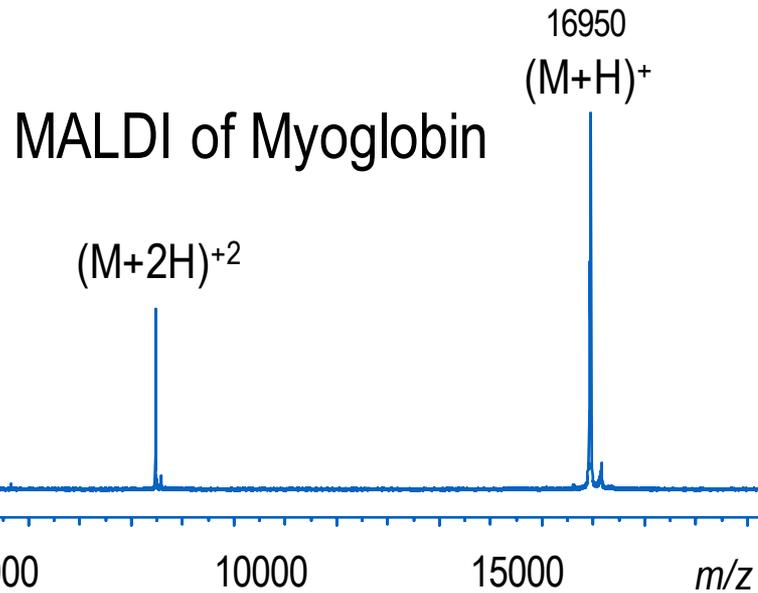
The mass to charge ratio

m/z

(Th or Thompson)

- Mass spectrometers do not directly give the mass of the ion!
- Rather the instrument determines the mass to charge ratio or m/z .
- For ions with either a +1 or -1 formal charge (as it the case most of the time) the mass of the ion is the same as the m/z ratio

Molecular Weight Measurement of the Protein

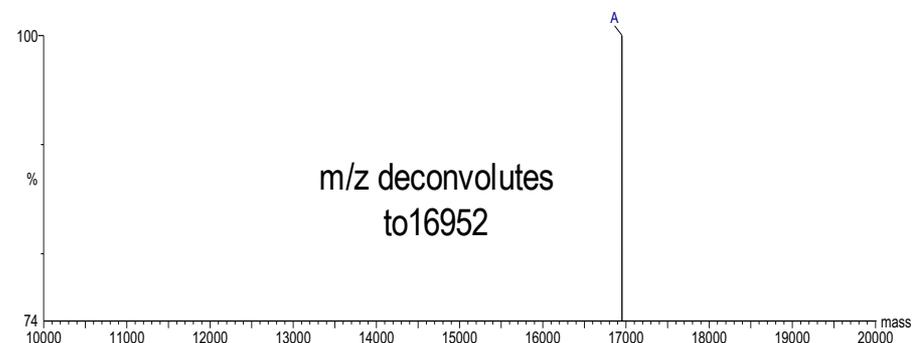
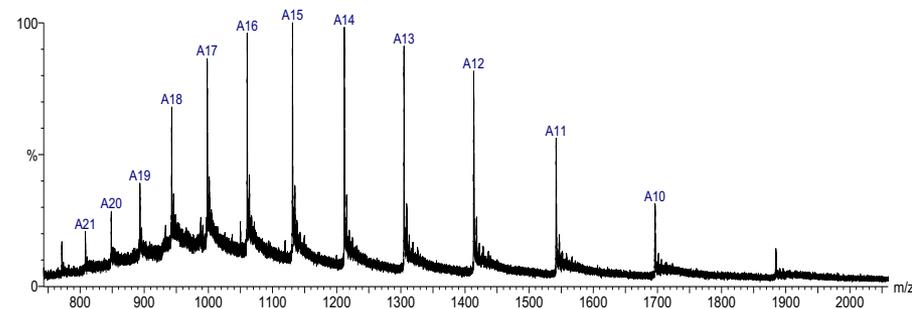


Sequence of Myoglobin

GLSDGEWQQVLNVWGKVEADIAGHGQEVLRFTGHPET
LEKFDKFKHLKTEAEMKASEDLKKHGTVVLTALGGILKKG
HHEAELKPLAQSHATKHKIPIKYLEFISDAIHVLHSHKHPGD
FGADAQGAMTKALELFRNDIAAKYKELGFQG

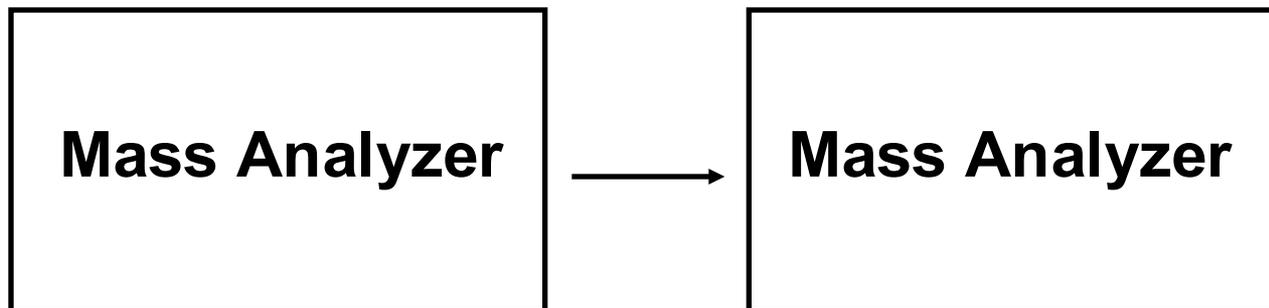
MW = 16951

Electrospray of Myoglobin



Tandem Mass Spectrometry

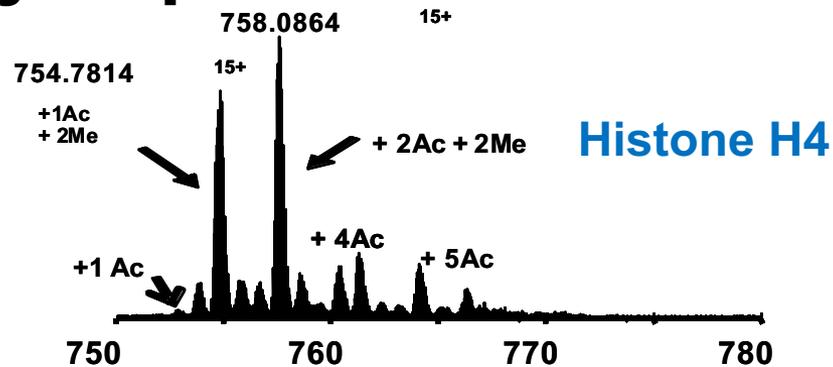
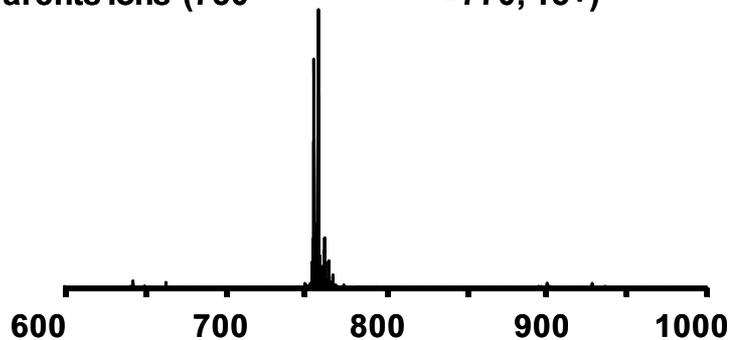
- Combines two or more mass analyzers of the same or different types
- First mass analyzer ***isolates*** the ion of interest (parent ion)
- The ions are ***fragmented*** between the first and second mass analyzer via collisions or irradiation with UV light
- The last mass analyzer obtains the mass spectrum of the fragment ions (daughter ions spectrum)



Protein Identification by Top Down MS/MS

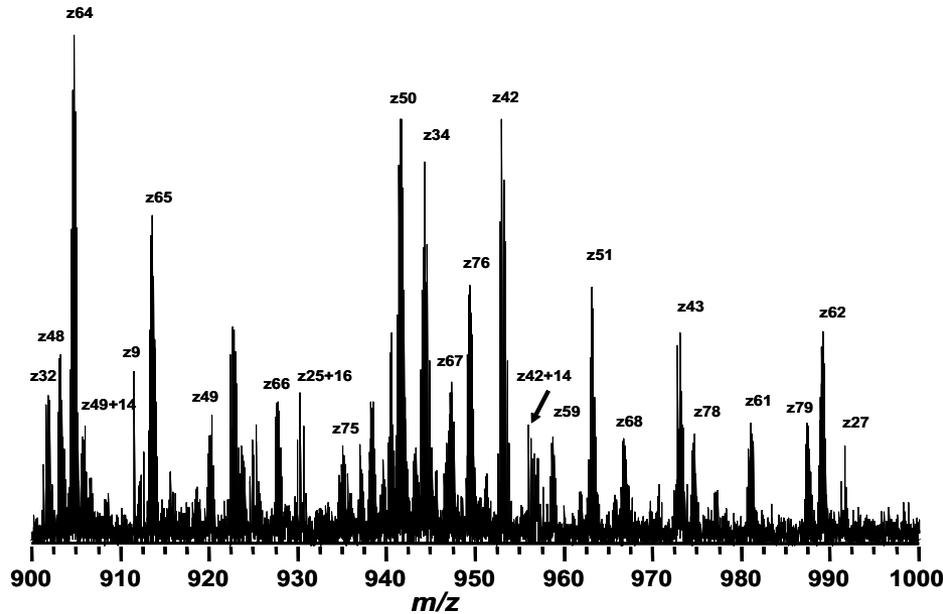
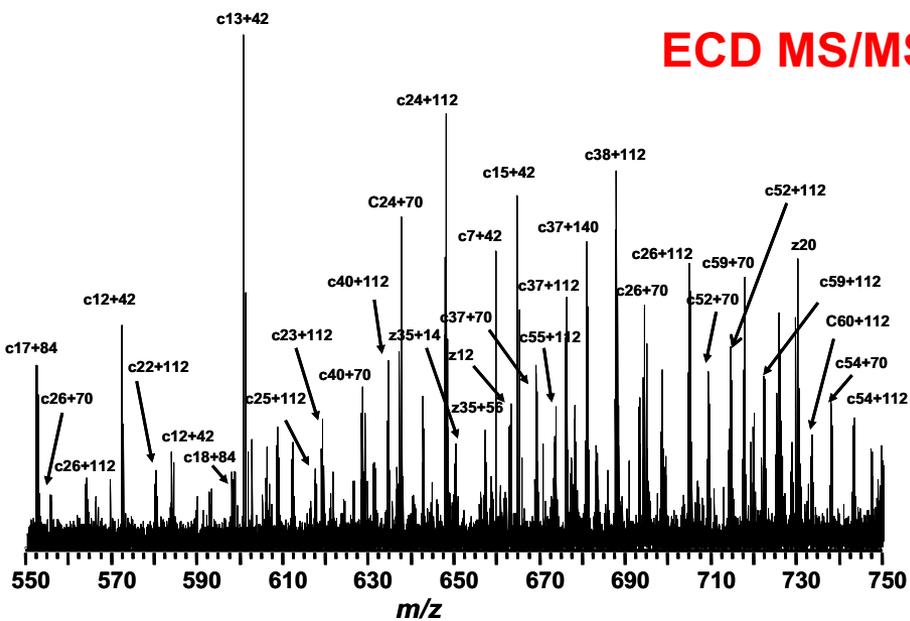
Parents ions (750

-770, 15+)

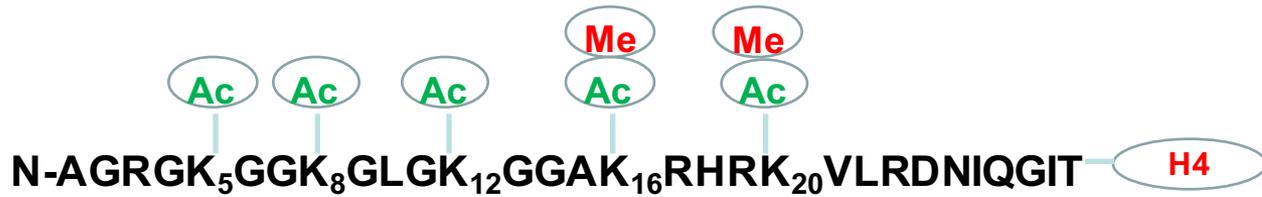


m/z

ECD MS/MS



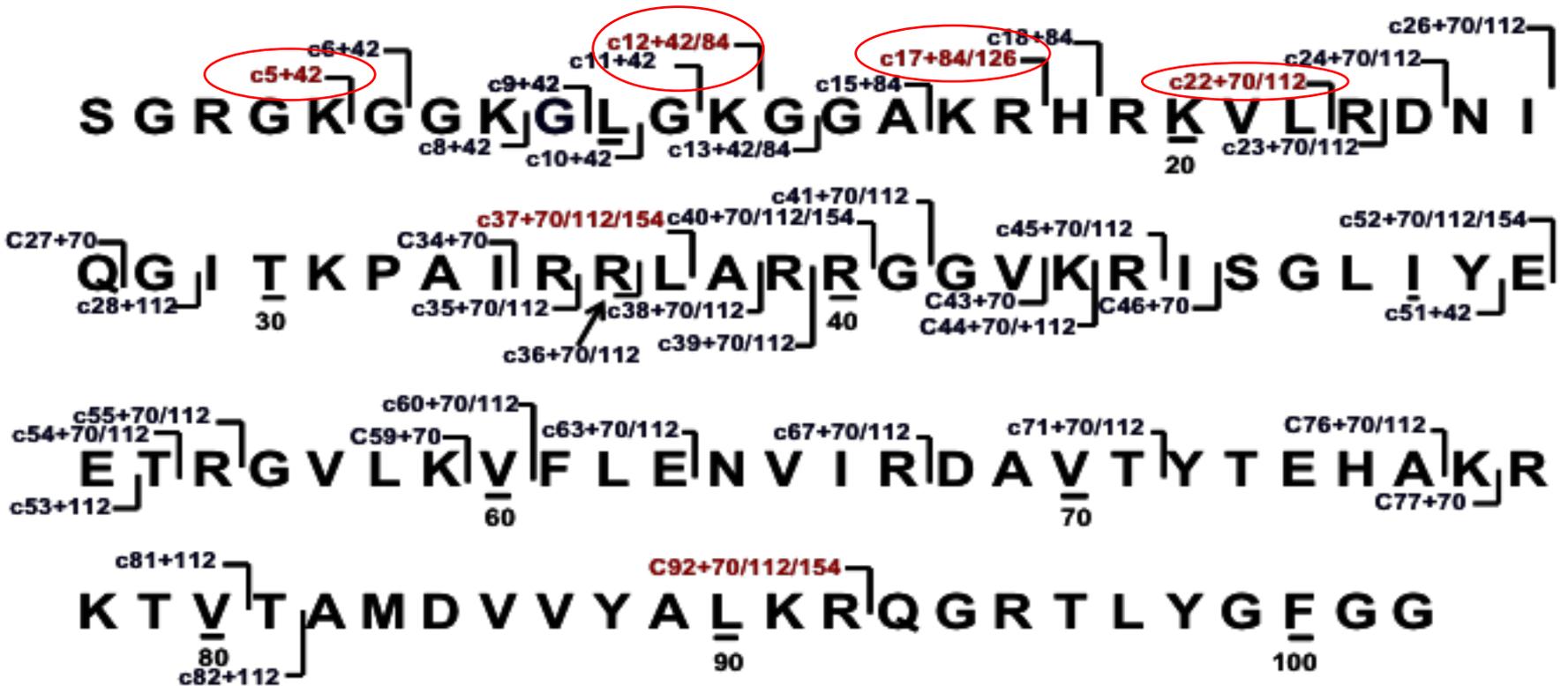
Protein Identification by Top Down MS/MS



Known Modification



Modification Detected

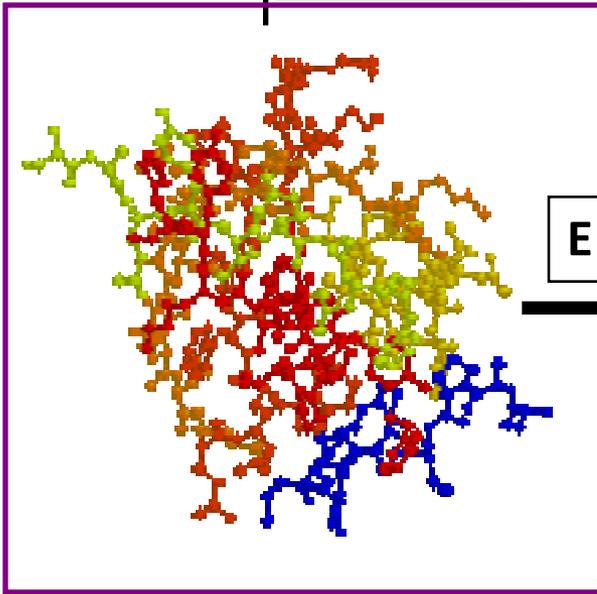


TOP DOWN

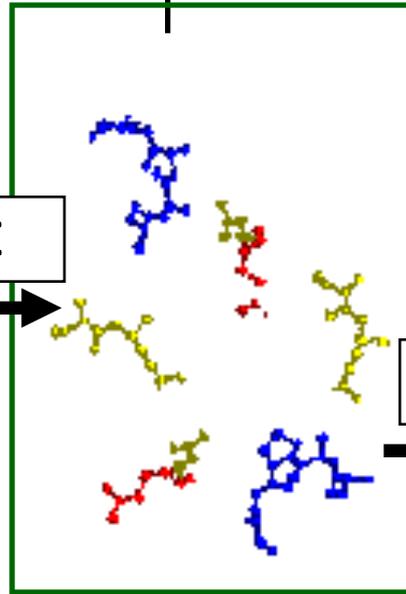
MS/MS

Peptide Fingerprinting

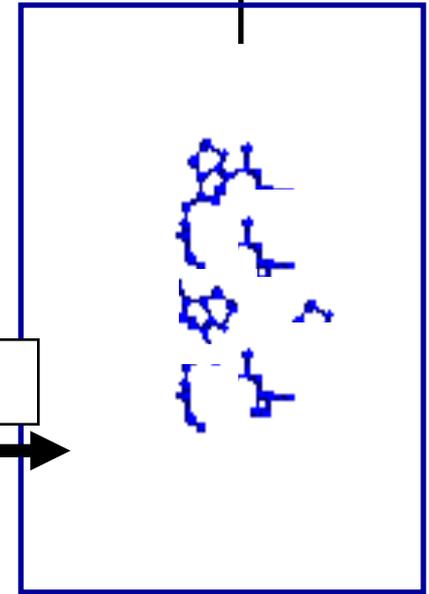
Middle Down
Bottom Up



ENZYME



MS/MS



Measure protein

Measure peptides

Measure fragments

Positives:

fast experiment, observe changes in protein mass

Negatives:

not good for protein id due to resolution & PTMs, requires clean sample

Positives:

fast experiment, identify proteins & PTMs

Negatives:

not good for mixtures of proteins, confidence in identification not ideal

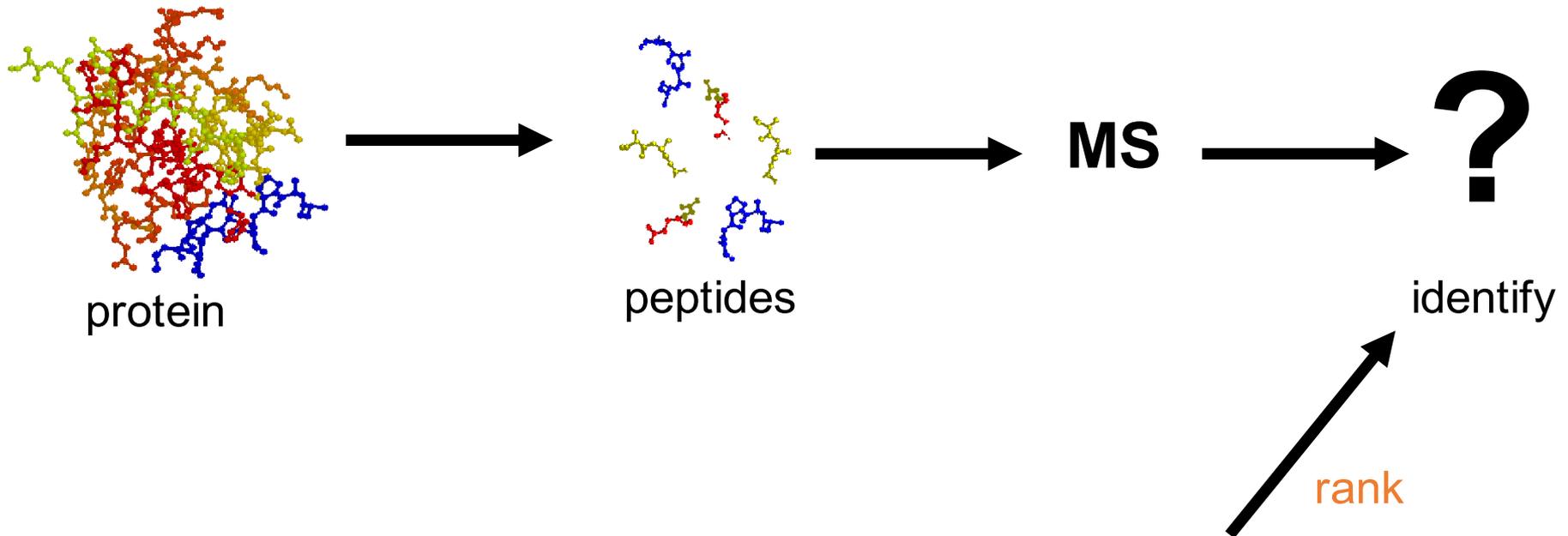
Positives:

identify protein & PTMs, high result confidence.

Negatives:

longer instrument time, longer database search

Peptide Mass Mapping



**MS Peptide MW
Found in Selected
Databases**

NDALYFPT...
SWDLTAL...
PTDLVSY...

- **Important data**
 - multiple peaks
 - mass accuracy
 - confirming information (pI, approx. mass, organism, etc.)

Cut protein into manageable pieces for the mass spectrometer to analyze

Bovine Albumin

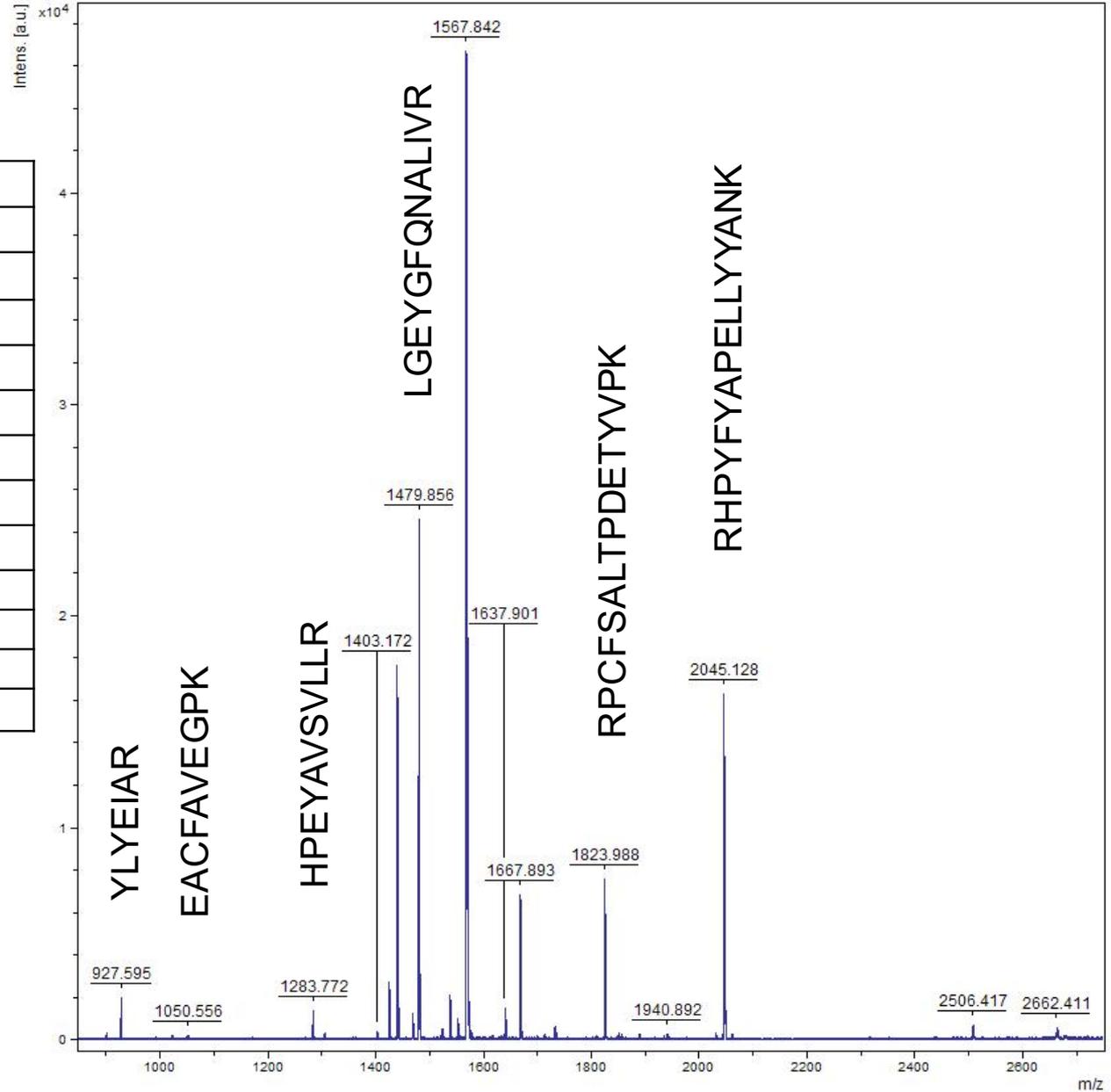
```

1 MKWVTFISLL LFFSSAYSRG VFRRDTHKSE IAHRFKDLGE EHFKGLVLIA FSQYLQQCPF DEHVKLVNEL
71 TEFAKTCVAD ESHAGCEKSL HTLFGDELCK VASLRETYGD MADCCEKQEP ERNECFLSHK DDSPDLPKLK
141 PDPNTLCDEF KADEKKFWGK YLYEIIARRHP YFYAPELLYY ANKYNGVFQE CCQAEDKGAC LLPKIETMRE
211 KVLTSARQR LRCASIQKFG ERALKAWSVA RLSQKFPKAE FVEVTKLVTD LTKVHKECCH GDLLECADDR
281 ADLAKYICDN QDTISSKLKE CCDKPLLEKS HCIAEVEKDA IPENLPPLTA DFAEDKDVCK NYQEAKDAFL
351 GSFLYEYSRR HPEYAVSVLL RLAKYEATL EECCAADDPH ACYSTVFDKL KHLVDEPQNL IKQNCDFEKE
421 LGEYGFQNAL IVRYTRKVPQ VSTPTLVEVS RSLGKVGTRC CTKPESERMP CTEDYLSLIL NRLCVLHEKT
491 PVSEKVTKCC TESLVNRRPC FSALTPDETY VPKAFDEKLF TFHADICTLP DTEKQIKKQT ALVELLKHKP
561 KATEEQLKTV MENFVAFVVK CCAADDKEAC FAVEGPKLVV STQTALA
    
```

mass	position	peptide sequence
927.486	161-167	YLYEIIAR
1050.485	588-597	EACFAVEGPK
1282.703	361-371	HPEYAVSVLLR
1305.708	402-412	HLVDEPQNLIK
1439.504	360-371	RHPEYAVSVLLR
1478.788	421-433	LGEYGFQNALIVR
1567.735	347-359	DAFLGSFLYEYSR
1639.931	437-451	KVPQVSTPTLVEVSR
1667.893	469-482	MPCTEDYLSLILNR
1823.892	508-523	RPCFSALTPDETYVPK
2045.021	168-183	RHPYFYAPELLYYANK
2506.243	469-489	M _{OX} PCTEDYLSLILNRLCVLHEK

Peptide Mass Mapping

927.486	161-167	YLYEIAR
1050.485	588-597	EACFAVEGPK
1283.703	361-371	HPEYAVSVLLR
1305.708	402-412	HLVDEPQNLIK
1439.504	360-371	RHPEYAVSVLLR
1479.788	421-433	LGEYGFQNALIVR
1567.735	347-359	DAFLGSFLYEYSR
1639.931	437-451	KVPQVSTPTLVEVSR
1667.893	469-482	MPCTEDYLSLILNR
1823.892	508-523	RPCFSALTPDETYVPK
2045.021	168-183	RHPYFYAPELLYYANK
2506.243	469-489	M _{ox} PCTEDYLSLILNRLCVLHEK



Human Albumin

MKWVTFISLL FLFSSAYSRG VFRRDAHKSE VAHRFKDLGE ENFKALVLIA FAQYLQQCPF EDHVKLVNEV TEFAKTCVAD
 ESAENCCKSL HTLFGDKLCT VATLRETYGE MADCCAKQEP ERNECFLOHK DDNPNLPRLV RPEVDVMCTA FHDNEETFLLK
 KYLYEIARRH PYFYAPELLE FAKRYKAAFT ECCQAADKAA CLLPKLDEL R DEGKAS SAKQ RLK CASLQKF GERA FKAWAV
 ARLSQRFPKA EFAEVSKLVT DLT KVHTECC HGDLLECADD RADLAKYICE NQDSISSKLLK ECCEKPLLEK SHCIAEVEND
 EMPADLPSLA ADFVESKDVC KNYAEAKDVF LGMFLYEYAR RHPDYSVLL LRLAKTYETT LEKCCAAADP HECYAKVFDE
 FKPLVEEPQN LIKQNCLEFE QLGEYKFQNA LLVRYTKKVP QVSTPTLVEV SRNLGKVGSK CCKHPEAKRM PCAEDYLSVV
 LNQLCVLHEK TPVSDRVTKC CTE SLVNRPP CFSALEVDET YVPKEFNAET FTFHADICTL SEKERQIKKQ TALVELVKHK
 PKATKEQLKA VMDDFAAFVE KCKKADKET CFAEEGKKLV AASQAALGL

m/z	Missed Cleavages	Sequence
1013.4244	0	⁵⁸⁹ <u>ETCFAEEGK</u> ⁵⁹⁷
1013.5990	0	⁵⁹⁹ <u>LVAASQAALGL</u> ⁶⁰⁹
1141.5194	1	⁵⁸⁹ <u>ETCFAEEGKK</u> ⁵⁹⁸
1141.6939	1	⁵⁹⁸ <u>KLVAASQAALGL</u> ⁶⁰⁹
1149.5759	1	²⁵ <u>DAHKSEVAHR</u> ³⁴
1149.6150	0	⁶⁶ <u>LVNEVTEFAK</u> ⁷⁵
1352.6661	1	⁴⁹⁷ <u>VTKCCTESLVNR</u> ⁵⁰⁸
1352.7685	1	⁴²⁷ <u>FQNALLVRYTK</u> ⁴³⁷
1623.7876	0	³⁴⁸ <u>DVFLGMFLYEYAR</u> ³⁶⁰
1623.9581	1	³⁶² <u>HPDYSVLLLLRLAK</u> ³⁷⁵
1898.9952	1	¹⁶⁹ <u>RHPYFYAPELFFAK</u> ¹⁸³
1898.9952	1	¹⁷⁰ <u>HPYFYAPELFFAKR</u> ¹⁸⁴

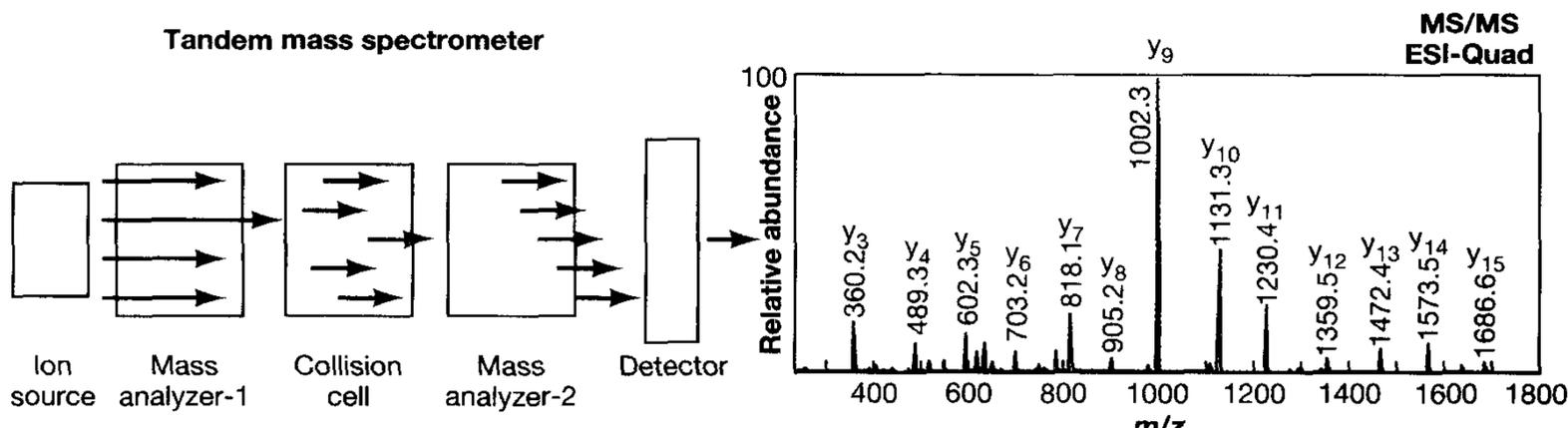
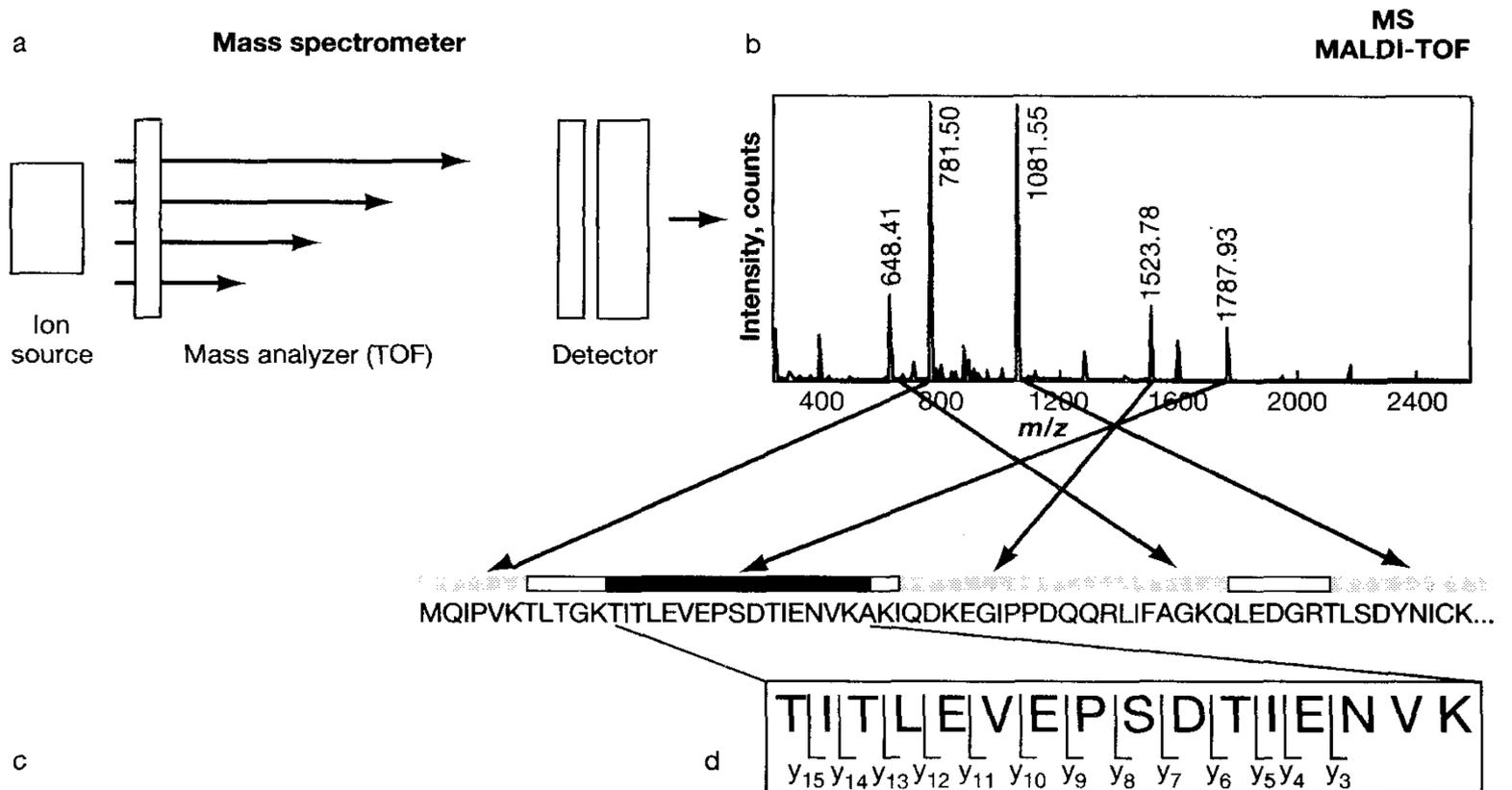
MS alone **CAN NOT** distinguish peptides of similar molecular weight.

- post-translational modification.
- sequence coverage determinations difficult.
- the identification of more than one protein in a sample impossible.
- Mass accuracy requirements

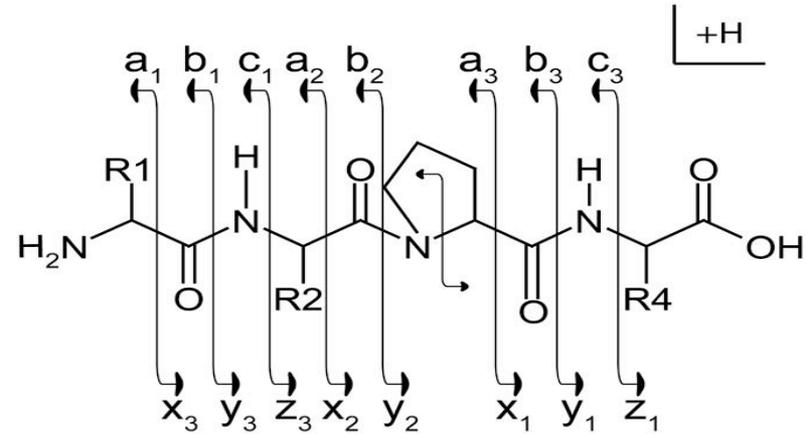
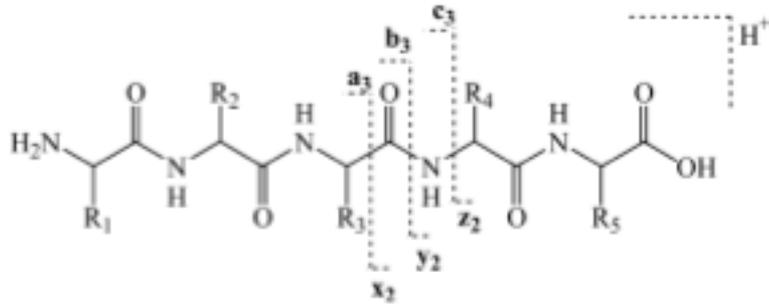
Peptide Fragmentation

- By use of Tandem Mass Spectrometry we can selectively fragment a peptide/protein.
- CID - Collision Induced Dissociation is due to high energy collisions with an inert gas.
- ETD and ECD - Electron Transfer/Capture Dissociation is due to an energetic transfer of an electron to the positively charged peptide or protein.
- Peptide and Proteins follow rules for how they fragment.

Basic Elements of MS and MS/MS analysis of proteins



Peptide Fragmentation



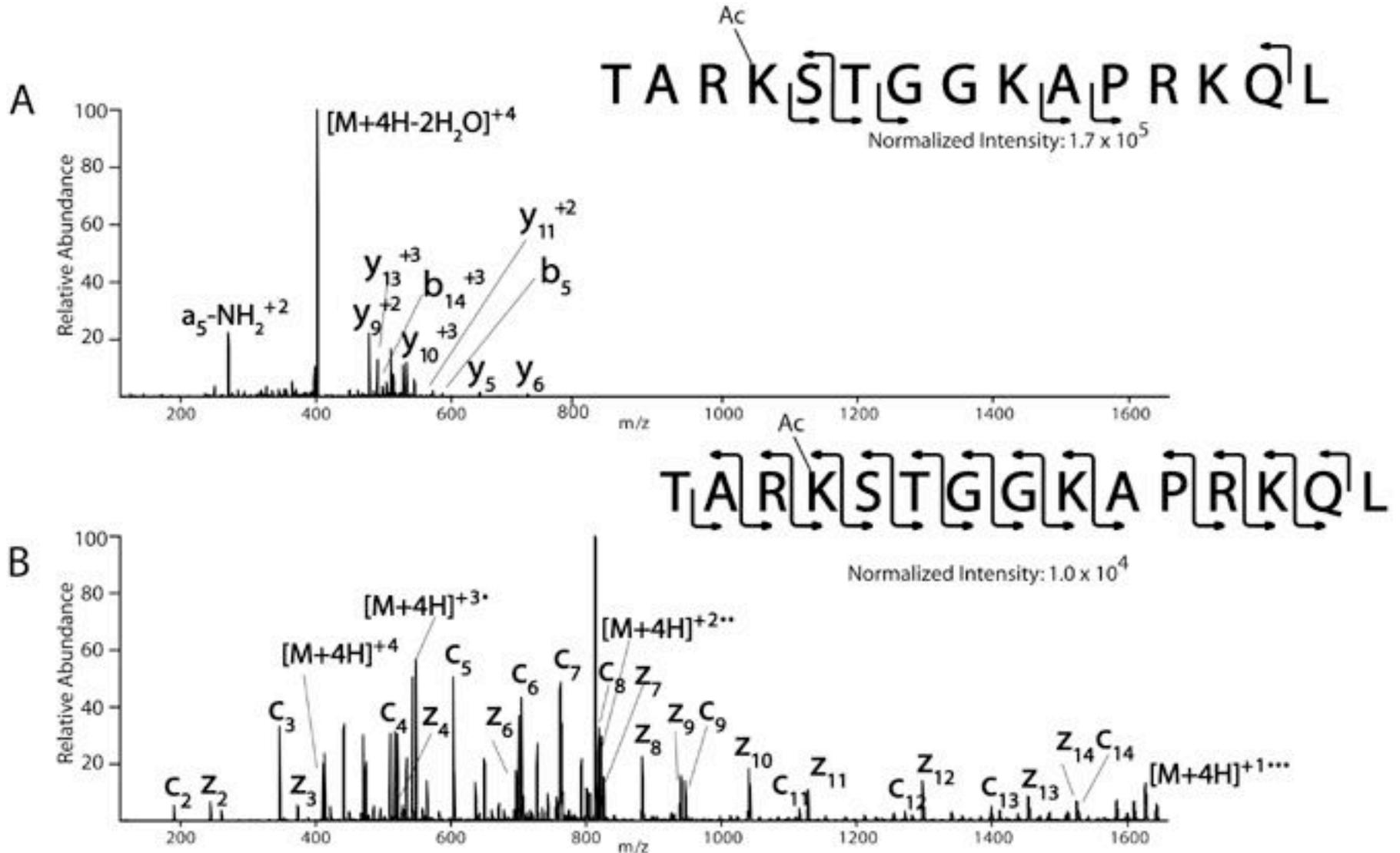
Roepstorff Nomenclature Scheme

Fragment ions form from the backbone cleavage of protonated peptides.

Fragment ions retaining the positive charge on the amino terminus are termed a-, b-, or c-type ions.

Fragment ions retaining the positive charge on the carboxy terminus are termed x-, y-, or z-type ions.

Peptide Fragmentation



Comparison of CID and ETD

LC/MSMS

#	b	Seq.	y	#
1	58.0287	G		13
2	171.1128	L	1362.7515	12
3	284.1969	L	1249.6674	11
4	413.2395	E	1136.5834	10
5	528.2664	D	1007.5408	9
6	641.3505	L	892.5138	8
7	698.3719	G	779.4298	7
8	861.4353	Y	722.4083	6
9	976.4622	D	559.3450	5
10	1075.5306	V	444.3180	4
11	1174.5990	V	345.2496	3
12	1273.6674	V	246.1812	2
13		K	147.1128	

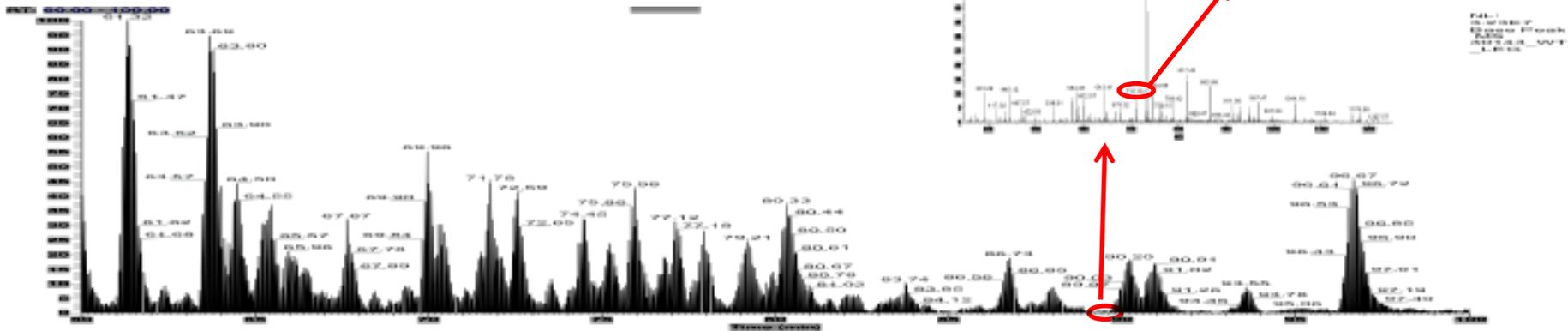
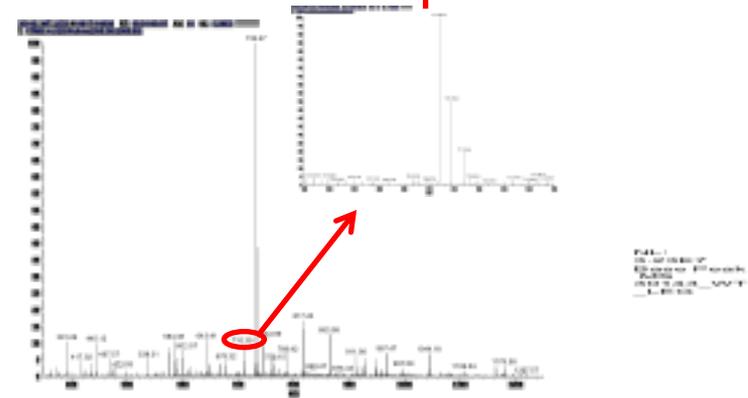
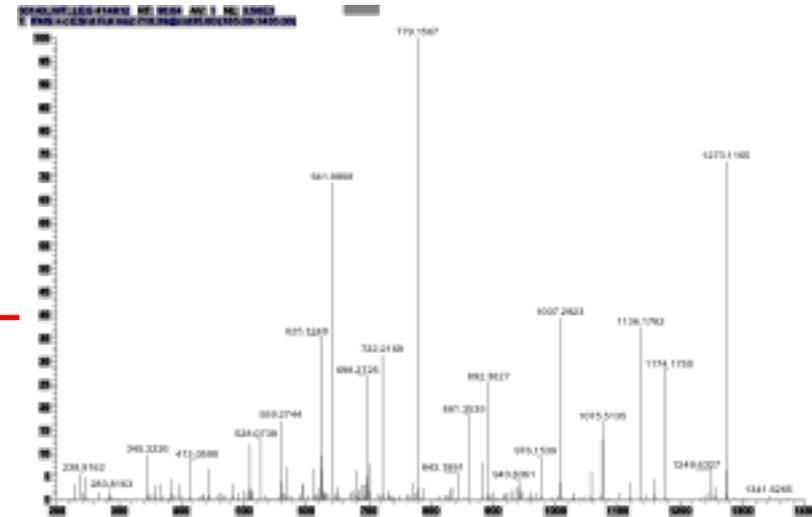
GLLEDLGYDVVVK

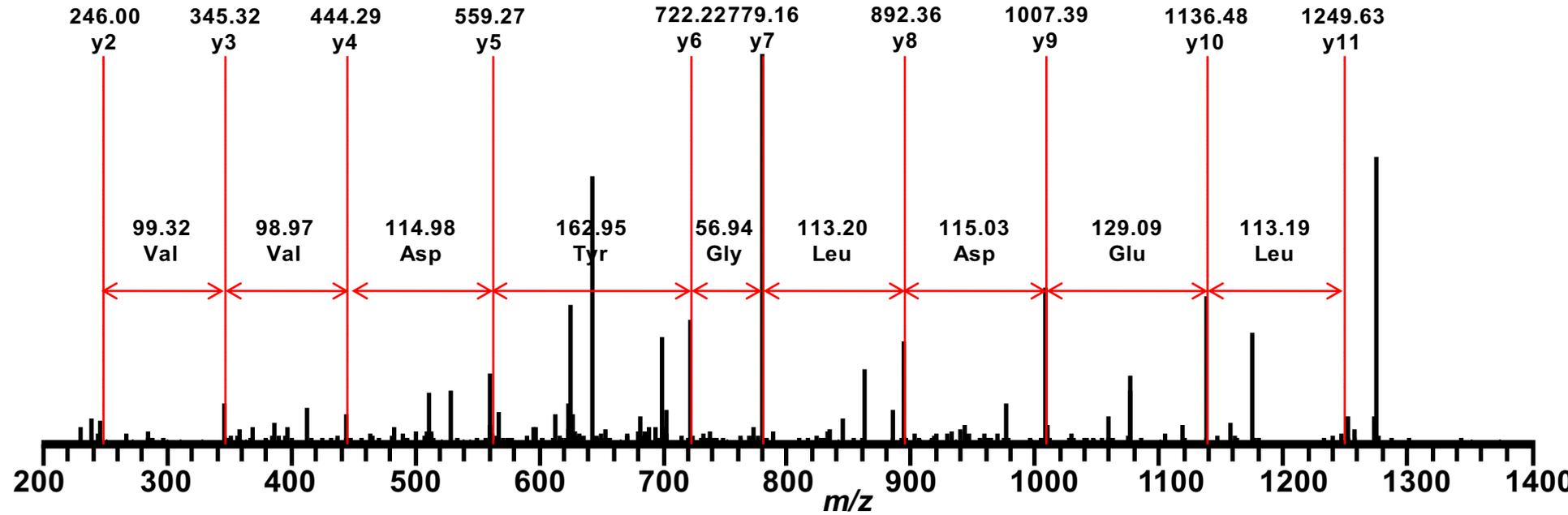
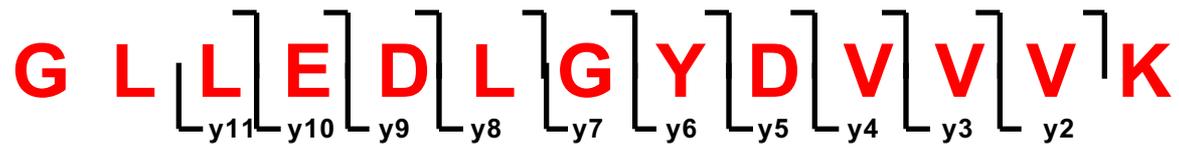


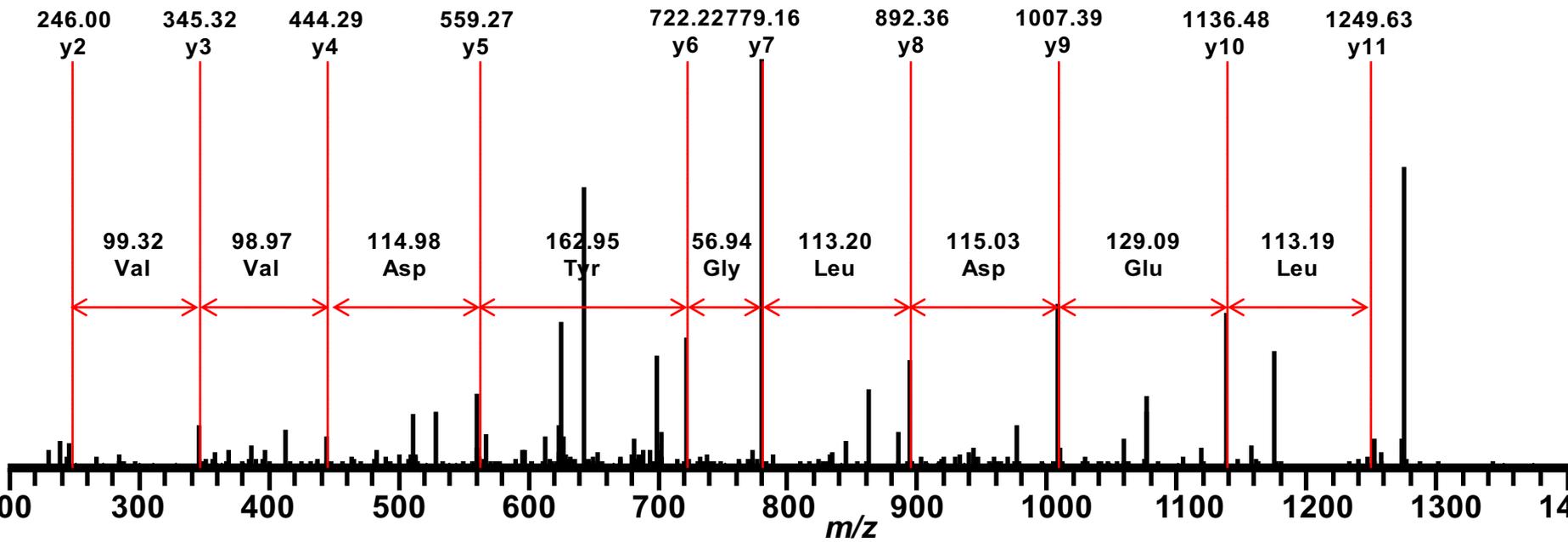
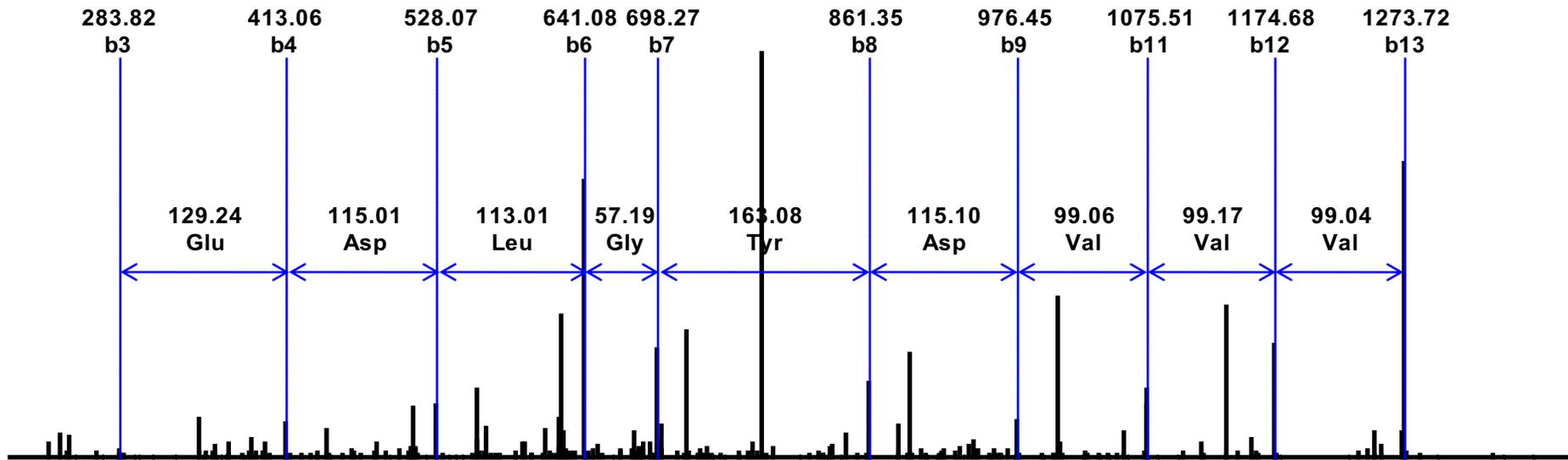
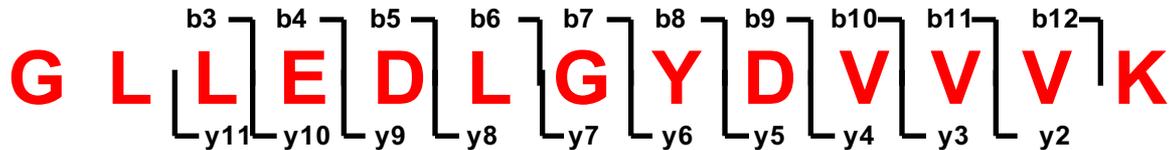
CASP4_MOUSE

Caspase-4

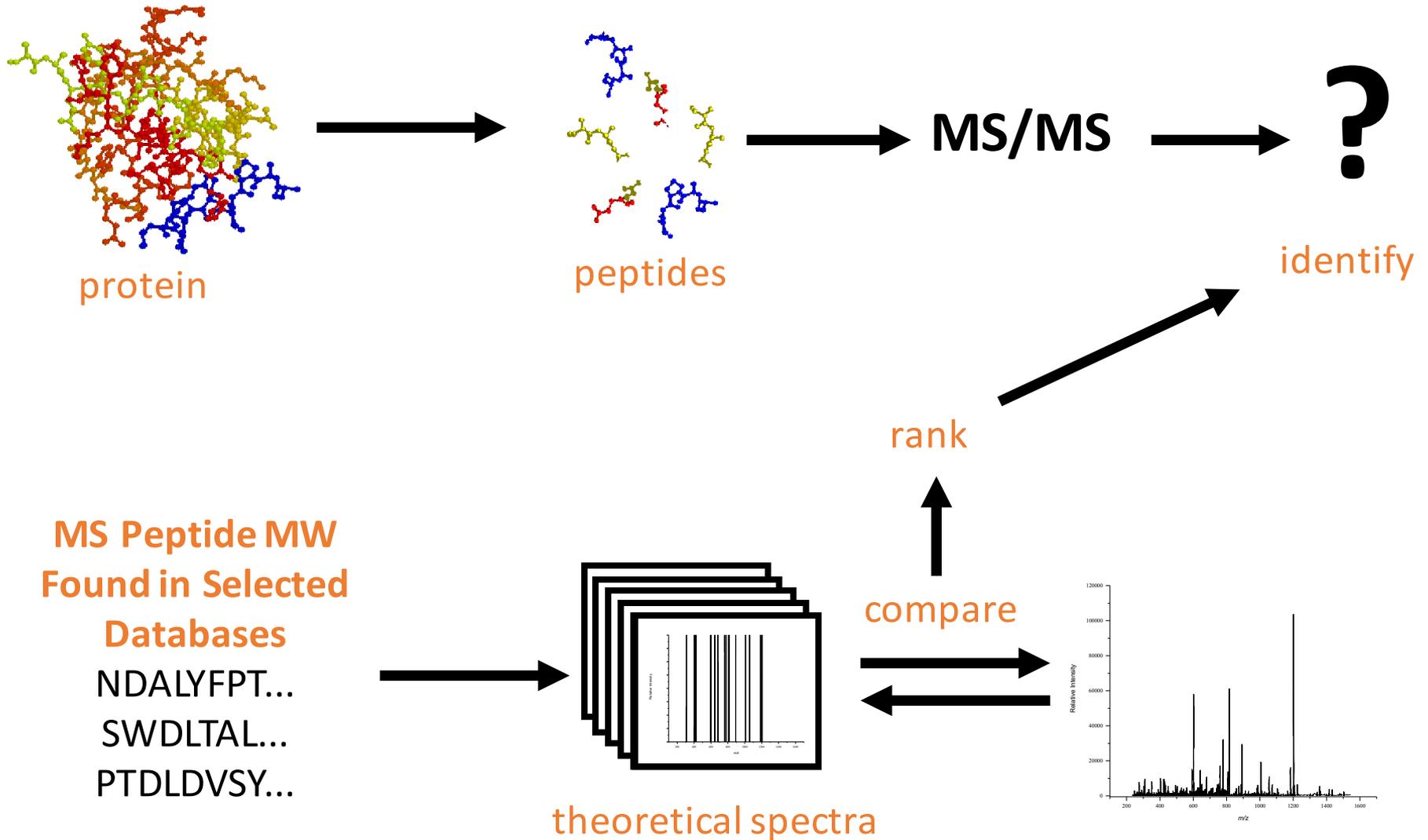
OS=Mus musculus







Data Analysis for MS/MS Method



Database Searching

.Mascot

- 16 node cluster for high-speed data processing

.Protein sequences are digested and fragmented *In Silico* which produces an enormous peak list

.Raw MS/MS data is converted to a peak list and compared against the *In Silico* peak list.

.Critical point of database searching

- The enzyme must work properly (no non-specific cleavage and missed cuts)
- The mass accuracy limits must be set appropriately
- Any modifications must be accounted for (modified cysteine)
- The database must contain the protein

Peptide Sequence and MS/MS Information

{MATRIX} Mascot Search Results

Peptide View

MS/MS Fragmentation of **HGTVVLTALGGILK**

Found in **MYG_EQUBU** in **SwissProt_ID**, Myoglobin OS=Equus burchelli GN=MB PE=1 SV=2

Match to Query 1630: 1377.846448 from(689.930500,2+) index(2183)

Title: STD_071614_B.3138.3138.2.dta

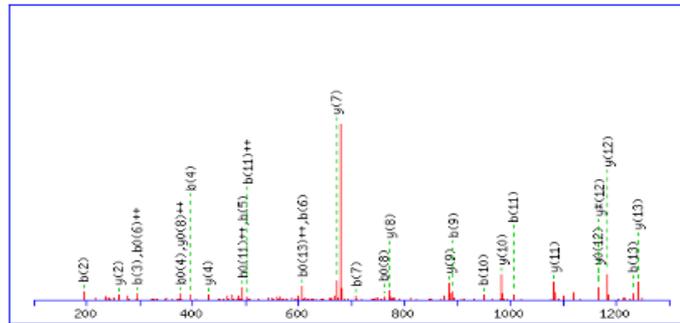
Data file \\filesrv\Mass_Spec\Data_New\Orbitrap TEST\071614\STD_071614_B_xml.mgf

Click mouse within plot area to zoom in by factor of two about that point

Or, Plot from to Da

Label all possible matches Label matches used for scoring

Show Y-axis



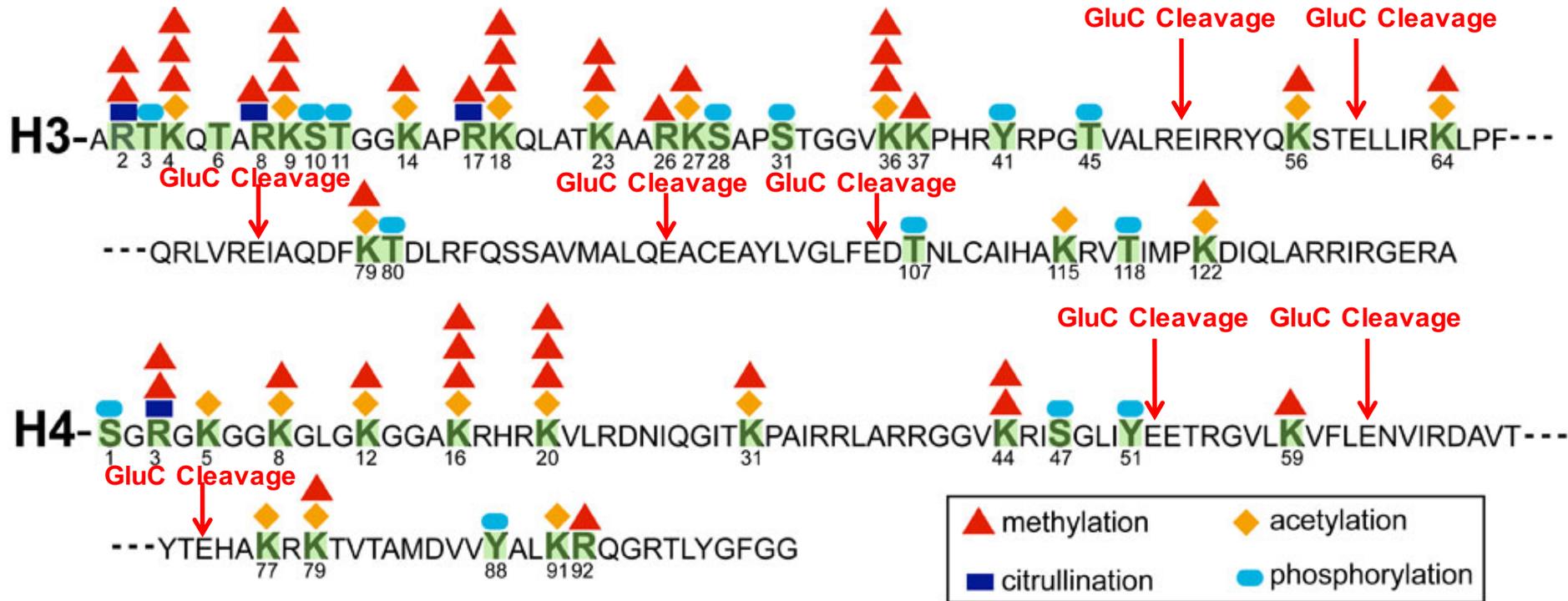
Monoisotopic mass of neutral peptide $M_r(\text{calc})$: 1377.8344

Ions Score: 100 Expect: $3.2e-010$

Matches : 28/112 Fragment ions using 33 most intense peaks [\(help\)](#)

#	b	b ⁺⁺	b ⁰	b ⁰⁺⁺	Seq.	y	y ⁺⁺	y [*]	y ⁺⁺⁺	y ⁰	y ⁰⁺⁺	#
1	138.0662	69.5367			H							14
2	195.0877	98.0475			G	1241.7828	621.3950	1224.7562	612.8817	1223.7722	612.3897	13
3	296.1353	148.5713	278.1248	139.5660	T	1184.7613	592.8843	1167.7347	584.3710	1166.7507	583.8790	12
4	395.2037	198.1055	377.1932	189.1002	V	1083.7136	542.3604	1066.6871	533.8472	1065.7030	533.3552	11
5	494.2722	247.6397	476.2616	238.6344	V	984.6452	492.8262	967.6186	484.3130	966.6346	483.8210	10
6	607.3562	304.1817	589.3457	295.1765	L	885.5768	443.2920	868.5502	434.7788	867.5662	434.2867	9
7	708.4039	354.7056	690.3933	345.7003	T	772.4927	386.7500	755.4662	378.2367	754.4822	377.7447	8
8	779.4410	390.2241	761.4305	381.2189	A	671.4450	336.2262	654.4185	327.7129			7
9	892.5251	446.7662	874.5145	437.7609	L	600.4079	300.7076	583.3814	292.1943			6
10	949.5465	475.2769	931.5360	466.2716	G	487.3239	244.1656	470.2973	235.6523			5
11	1006.5680	503.7876	988.5574	494.7824	G	430.3024	215.6548	413.2758	207.1416			4
12	1119.6521	560.3297	1101.6415	551.3244	I	373.2809	187.1441	356.2544	178.6308			3
13	1232.7361	616.8717	1214.7256	607.8664	L	260.1969	130.6021	243.1703	122.0888			2
14					K	147.1128	74.0600	130.0863	65.5468			1

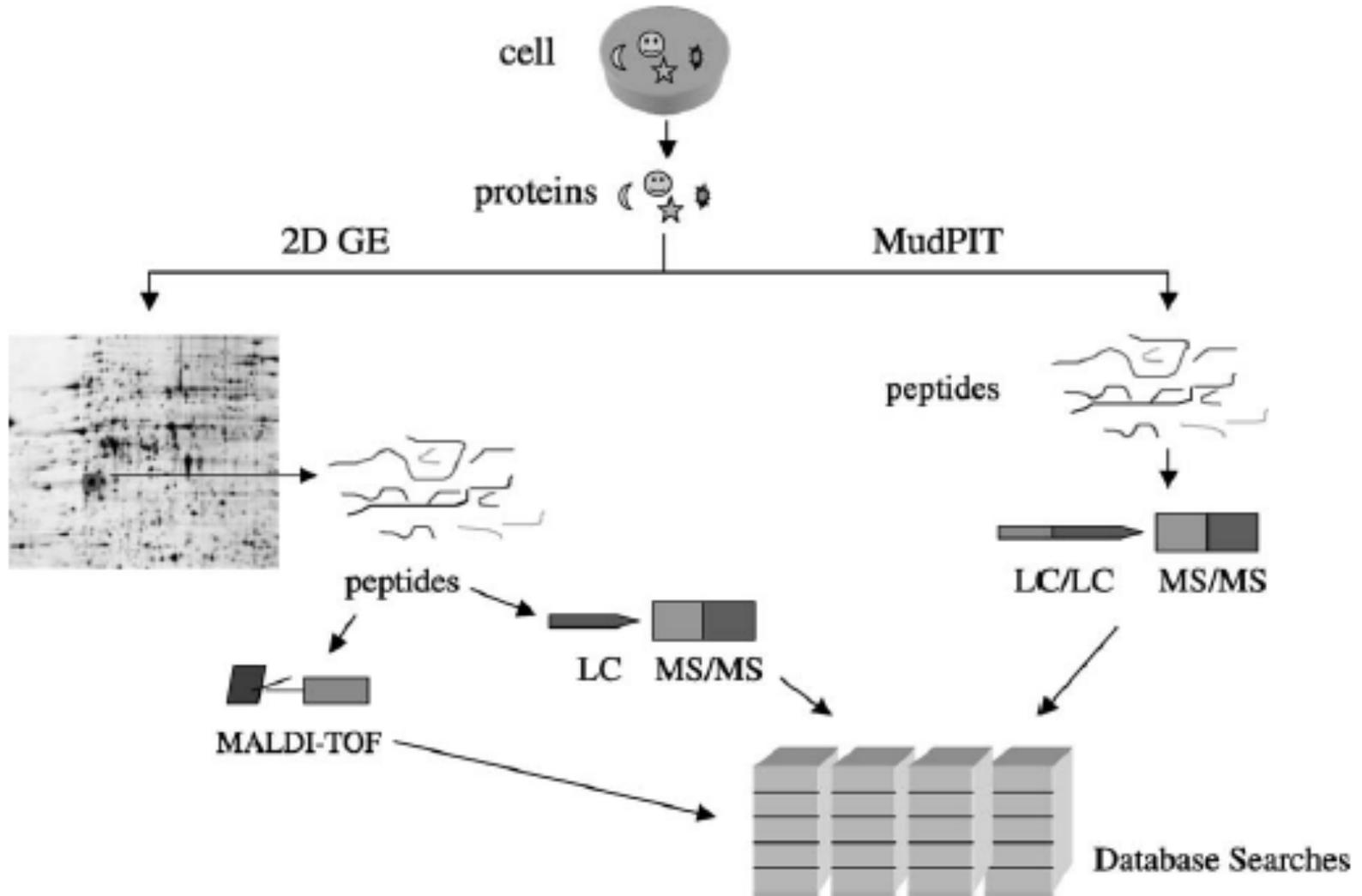
Middle Down



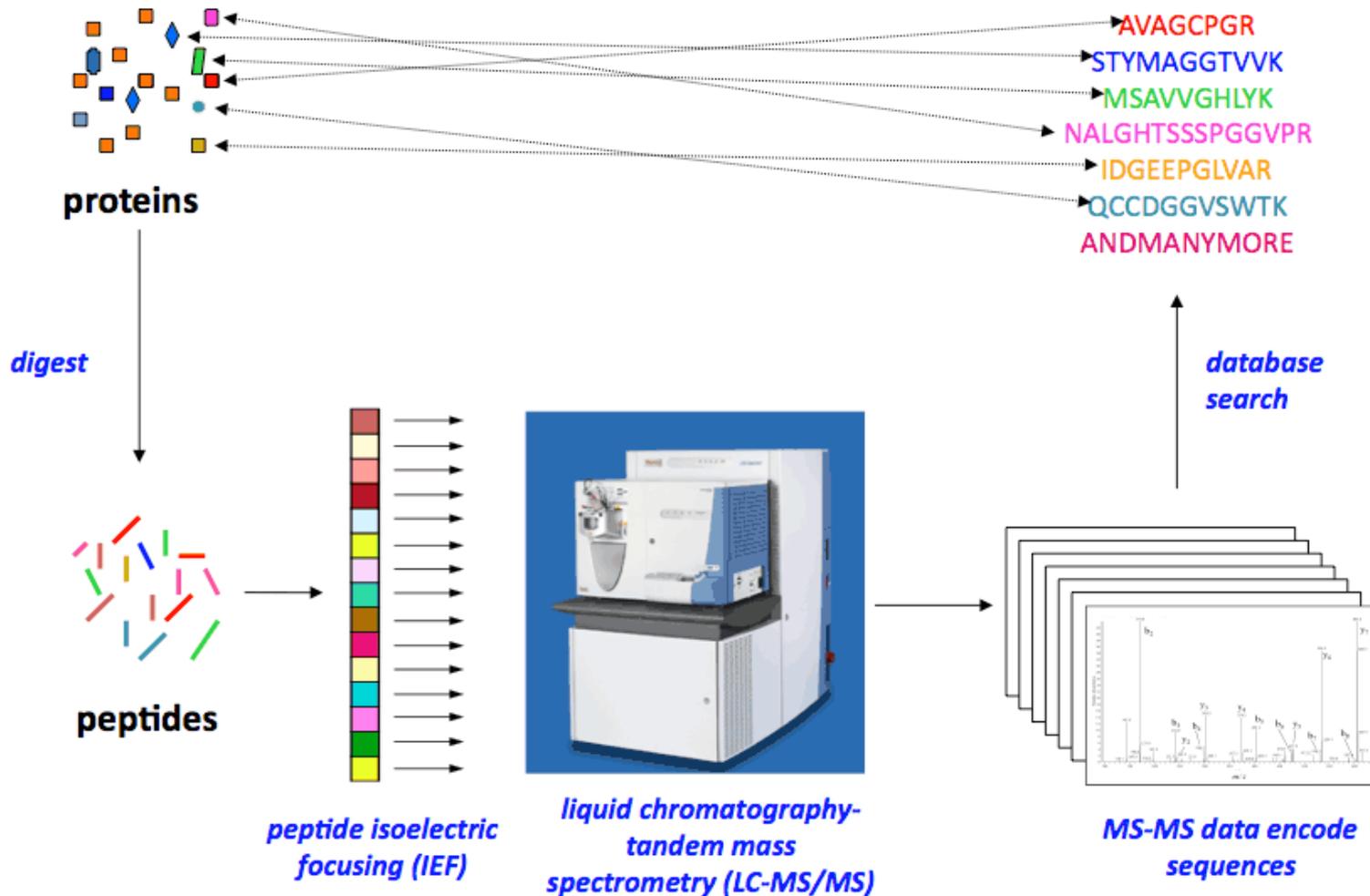
Why Middle Down:

simplifies the complex mixtures offered by bottom up strategies
 while avoiding the diminished performance of top down experiments

Conventional Proteomics

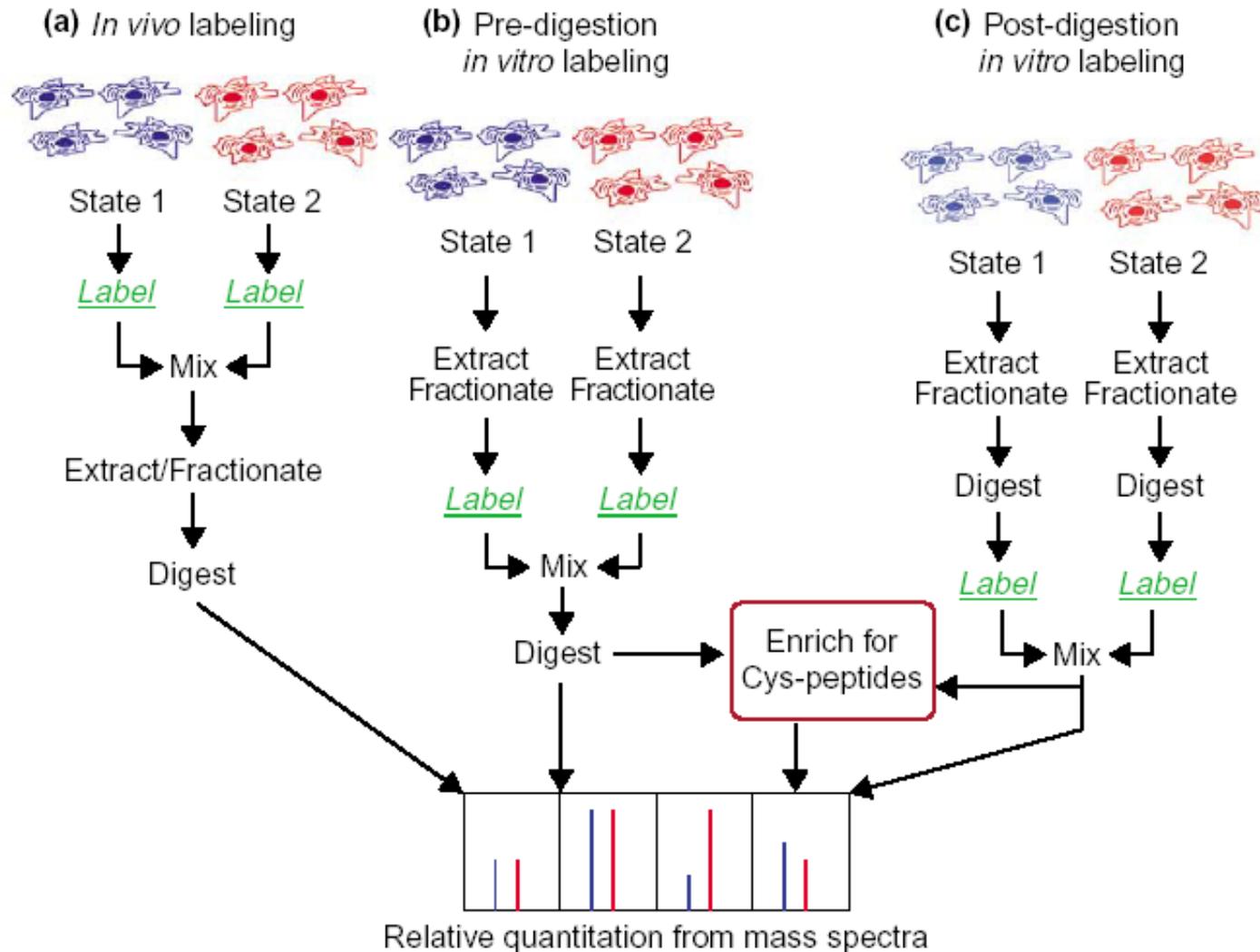


Shotgun Proteomics, aka LC-MS/MS



Differential Expression Profiling

Stable Isotope Labeling



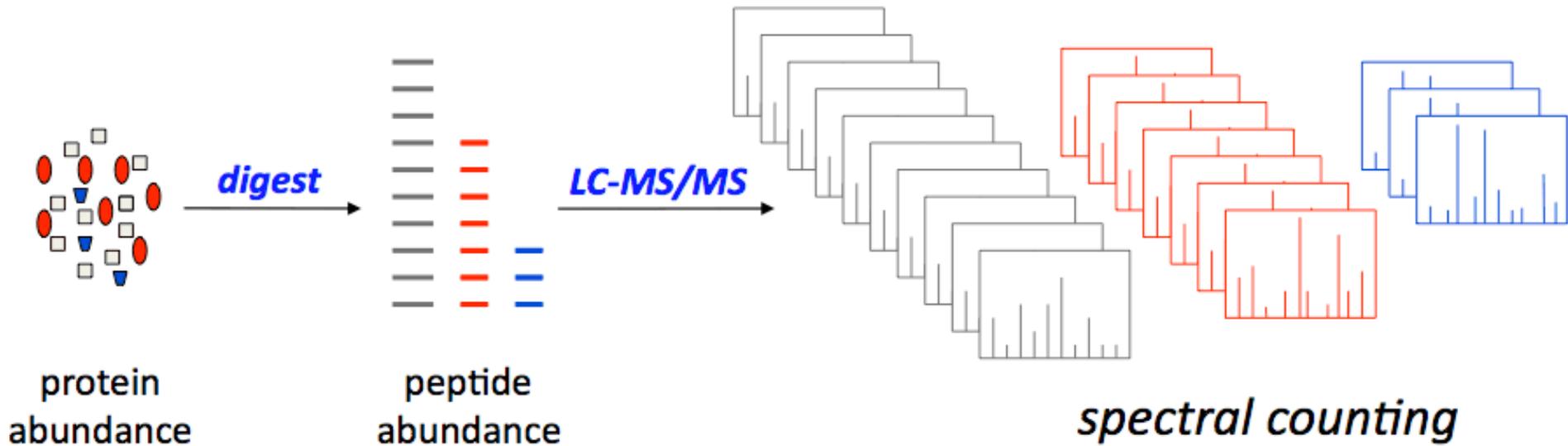
Stable Isotope Labeling

Summary of pros and cons for quantitative proteomics using MS*.

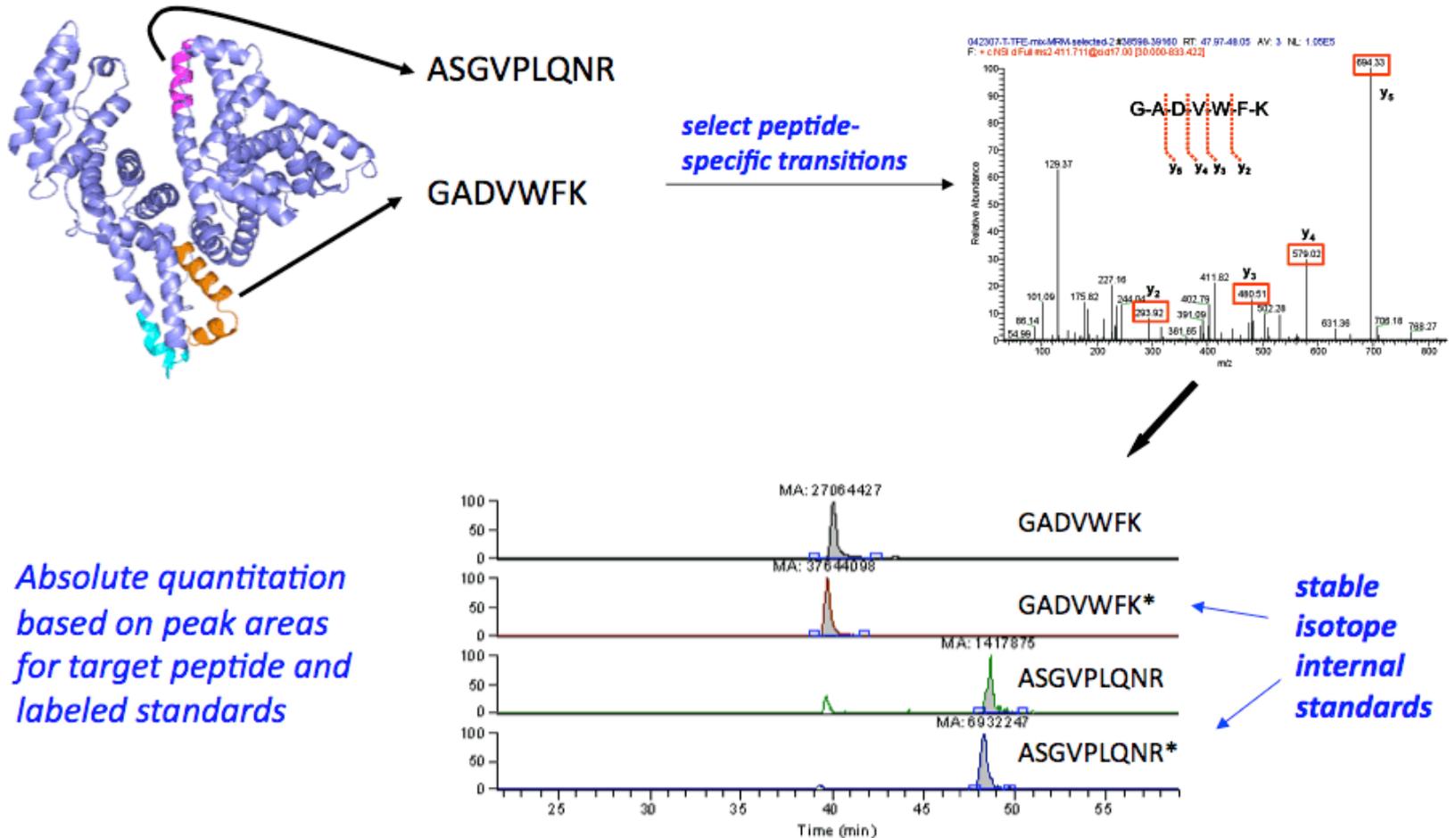
Method	Pros	Cons
<i>In vivo</i> labeling	Simple and accurate Labeling of all peptides is achieved	Limited to cells that can be grown in culture
ICAT	Reduced complexity by enrichment of cysteine-containing peptides Applicable to all types of sample	Limited to the analysis of proteins containing cysteine Complex MS/MS spectra due to the presence of the affinity label
Cleavable ICAT	Reduced complexity by enrichment of cysteine-containing peptides. Applicable to all types of sample	Limited to the analysis of proteins containing cysteine
Labeling with acrylamide	Simple derivatization and procedure Applicable to all types of sample	Requires more steps than ICAT Limited to the quantification of proteins containing cysteine
Proteolysis in the presence of $^{18}\text{O}/^{16}\text{O}$	Easy derivatization Applicable to all types of sample Labeling of all peptides is achieved	Complex analysis. Possible loss or incomplete incorporation of the label
Esterification	Easy derivatization Applicable to all types of sample Labeling of all peptides is achieved	Derivatization efficiency might not be high
Derivatization of primary amines	Applicable to all types of sample	Requires multiple steps for derivatization increasing the possibility of errors and decreasing the accuracy Derivatization efficiency might not be high
MCAT	Applicable to all types of sample Inexpensive, no need for stable isotopes	Requires multiple steps for derivatization Low accuracy
Site-specific photo-cleavable method	Reduced complexity by enrichment of cysteine-containing peptides Applicable to all types of sample	Limited to the analysis of proteins containing cysteine

* The most common methods used in quantitative proteomics are reported (left column). None of the above procedures works for all cases and they all have some pros (center column) or cons (right column) that are briefly described.

Label Free Spectral Counting

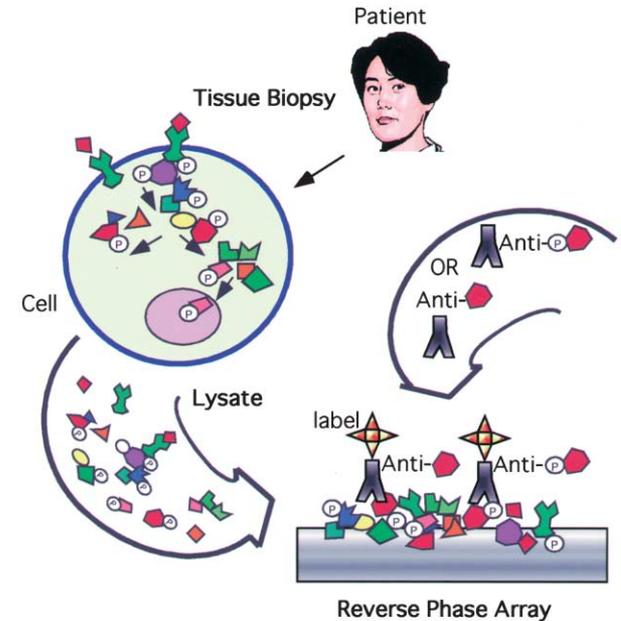
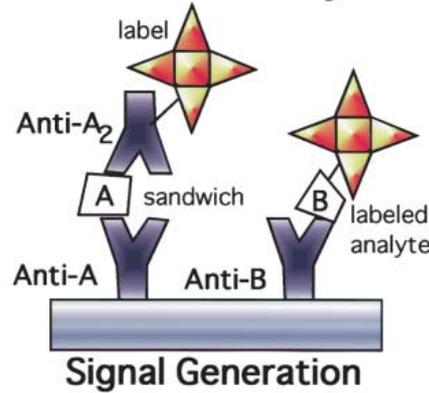
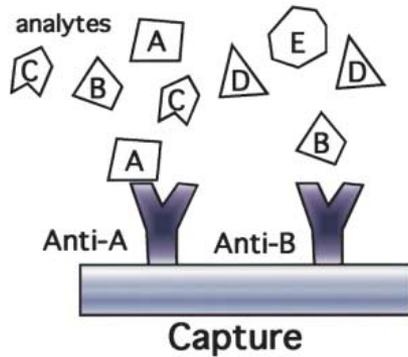


Targeted Proteomics - Multiple Reaction Monitoring

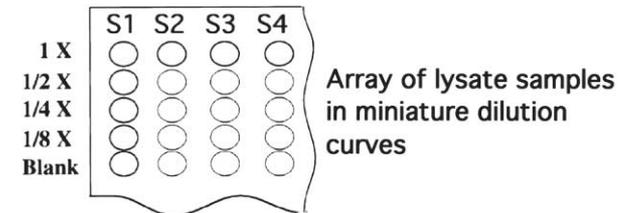
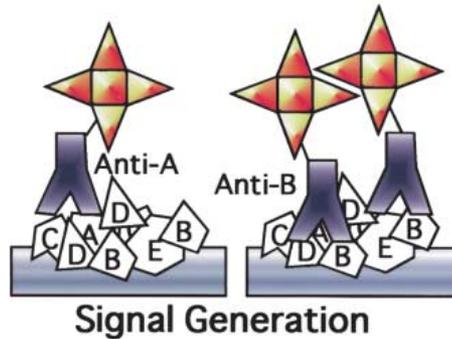
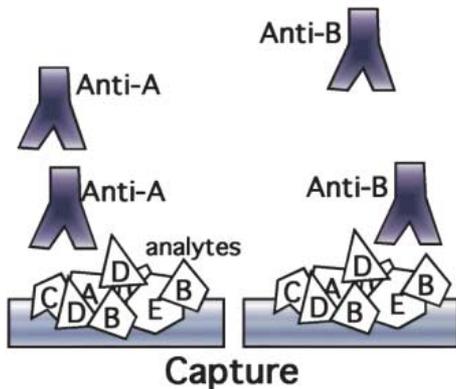


Protein Microarray Technologies: Overview

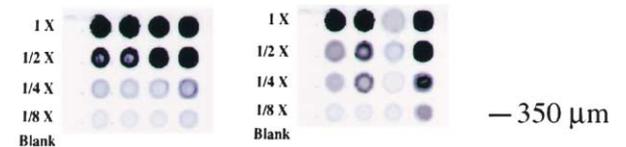
Forward Phase Protein Microarray



Reverse Phase Protein Microarray



Anti - Protein N Anti - Phospho-Protein N



Protein Microarray Technologies: Challenges

- **Antibody Specificity:**

For example, given an antibody with a specificity of a detection ~99%. A crossreacting protein that exists in a thousand fold (or greater) excess will lead to an unacceptable background signal.

- **Dynamic range:**

The dynamic range of detection is many orders of magnitude lower than the dynamic range of protein abundance in cells/serum.

- **Sensitivity:**

Low abundant substrates will require higher amounts of tissue in order to observe signal