

Lecture 24: Protein Identification by Mass Spectrometry

§ Identifying proteins and peptides by Mass Spectrometry

- Peptide mass fingerprint
- MS/MS based peptide ID

§ MS/MS methodology

- Quadrupole Mass Analyzer
- MS/MS based peptide ID methodology

Adapted from slides originally created by Dr. James Bruce

MS Information Analysis for Proteomics

2 Strategies for MS-based protein ID

1) Peptide mass fingerprint

2) MS/MS-based peptide ID

Both approaches are carried out with peptides generated by chemical or enzymatic cleavage of proteins

Peptide mass fingerprint

Cleave protein into a set of peptides

Measure peptide masses

Search protein database with:

- (1) measured peptide masses
- (2) Protein mass, if known
- (3) known cleavage rules

Protein Cleavage Rules for Common Proteases

Enzyme (or reagent)	Sequence Specificity
Arg-C	Cleaves after R
Asp-N	Cleaves before D
Chymotrypsin	Cleaves after F or L or M or W or Y if not followed by P
Glu-C	Cleaves after E
Lys-C	Cleaves after K
Trypsin	Cleaves after K or R if not followed by P
CNBr (Cyanogen Bromide)	Cleaves after M

Average Numbers of Peptide Fragments

§ Expected number of peptide fragments upon cleavage of a 300 amino acid protein:

Enzyme	# of peptide fragments	Average length of peptide fragments
Arg-C	17	18
Asp-N	17	18
Chymotrypsin	54	6
Glu-C	20	15
Lys-C	19	16
Trypsin	35	9
CNBr	8	38

Data from Kellner et al. (1994)

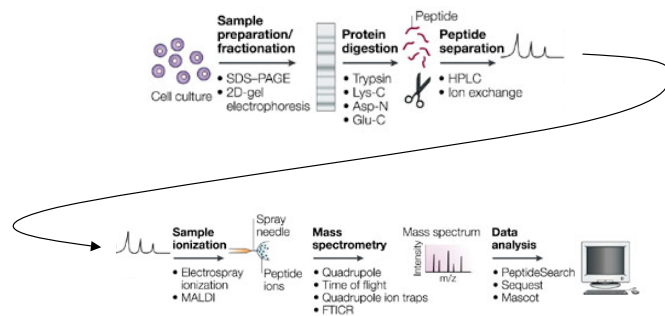
Trypsin digestion is the most commonly used cleavage method

Larger size peptides more unique mass (mass fingerprint)

Mono-isotopic peak detectable

2-3kDa peptides readily fragmented (MS/MS)

Process of Identifying Purified Proteins



Nature Reviews | Molecular Cell Biology

Images from Steen and Mann (2004) *Nature Reviews Molecular Cell Biology* 5:699-711

Ideal case: Measure peptide mass with 0 mass error

Peptide Mass List (from database)

Protein ID

- 1528.7685
- 1528.7989
- 1529.0002
- 1529.2001
- 1529.2454
- 1529.5006
- 1529.7245
- Hb-alpha 1529.7348**
- 1529.7495
- 1529.9978
- 1530.2332
- 1530.4567
- 1531.0003
- 1531.0107
- 1531.3004
- 1531.5656

Compare with database

1529.7348
Measured mass

Each peptide mass can potentially ID a protein.

Reality: Sub-PPM performance difficult to achieve

§ Mass measurement error is expressed as error / mass

- Typically, error / mass is expressed as PPM (parts per million)
- For example, a typical MS measurement may have an error of 10 PPM (i.e., $10 \times 10^{-6} = 1 \times 10^{-5}$)
- For a peak of mass 1529.7348, with accuracy of 10 PPM, then

$$10 \text{ PPM} = (\text{mass error}) / (\text{mass of peak})$$

$$10 \times 10^{-6} = (\text{mass error}) / (1529.7348)$$

$$(1 \times 10^{-5}) * (1529.7348) = \text{mass error}$$

mass error = 0.0153

- Hence, peak mass = 1529.7348 ± 0.0153
- Can't distinguish 1529.7348 from 1529.7195 or 1529.7501

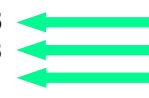
One peptide mass usually cannot uniquely identify the protein of interest

Peptide Mass List (from database)

•	
•	
Protein ID	•
•	1528.7685
•	1528.7989
•	1529.0002
•	1529.2001
•	1529.2454
•	1529.5006
LEP503?	1529.7245
Hb-alpha?	1529.7348
PHP-1?	1529.7495
•	1529.9978
•	1530.2332
•	1530.4567
•	1531.0003
•	1531.0107
•	1531.3004
•	1531.5656
•	
•	

Compare with database

1529.7348 ± 0.0153
Measured mass



Peptide mass fingerprinting: search with several peptide masses from the same protein

MS peaks (m/z)	Mass error	Number of possible proteins identified
1529.73	± 0.10	204
1529.73 1252.70	± 0.10	7
1529.73 1252.70 1833.88	± 0.10	1

Even with moderate mass accuracy ($0.1/1529 = 65\text{PPM}$), several peptide masses allow unique ID

What about mixtures of proteins??

- Peptide mass fingerprinting becomes problematic
- Relationship between individual peptides is unclear
- Need additional information

Peptide Mass Fingerprint

Advantages

Fast
Sensitive
Not prone to failure from unexpected modifications

Disadvantages

Prone to failure with protein mixtures
Generally requires separation by gels
Higher false positive

Quadrupole Mass Spectrometry

Quadrupolar oscillating electric field

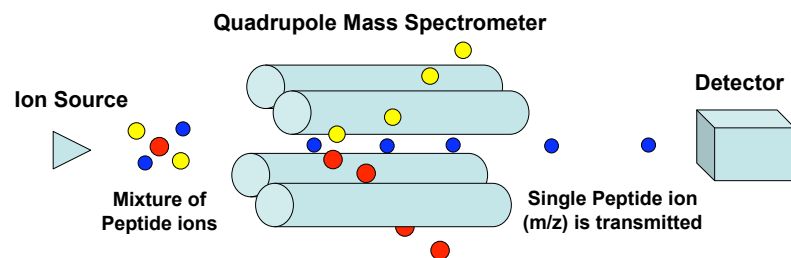
“Mass Filter” can selectively transmit single m/z

Can be connected in series - MS/MS

Commonly coupled with ESI and chromatographic separations

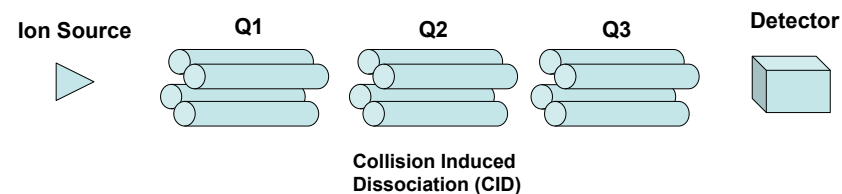
Generally, allows better quantitation than MALDI-TOF

Quadrupole Mass Spectrometry: How does it work



Based on figure from Glish and Vachet (2003) Nature Reviews Drug Discovery 2:140-150

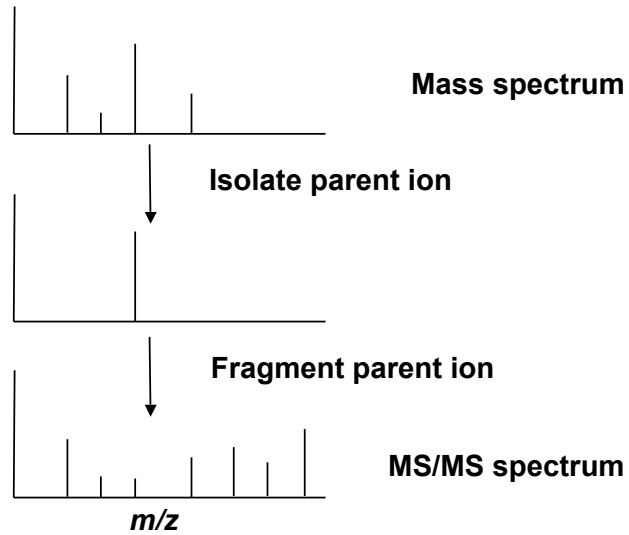
Triple quadrupole mass spectrometer



Each set of quads can be operated independently

Allows tandem mass spec – MS/MS

What is MS/MS?



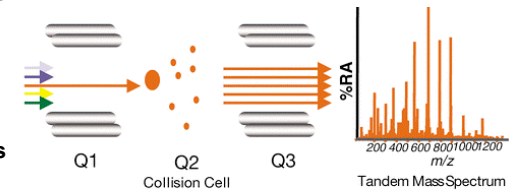
MS/MS-based peptide ID

Isolate a single peptide from a mixture

Fragment in mass spectrometer

Measure fragment masses

Compare with database for protein ID



In this case, each peptide can lead to protein ID

Image from Goodlett and Li (2002) *Funct Integr Genomics* 2:138-153

Collision Induced Dissociation of Peptides

- An peptide ion of particular m/z is selected in Quadropole 1 (Q1)
- Quadropole 2 (Q2) is flooded with gas to induce CID
- The m/z ratios of resultant peptide fragment ions (b and y) are measured in Quadropole (Q3)

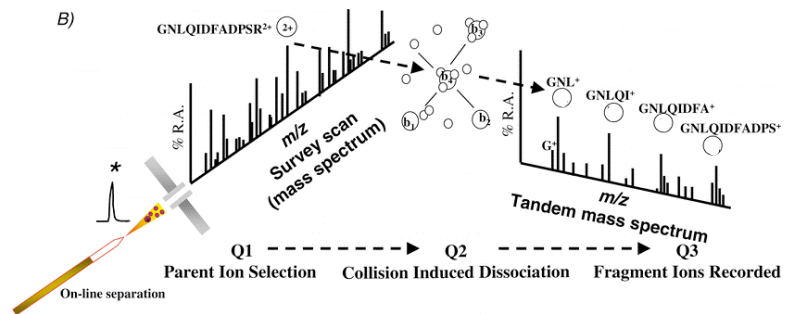


Image from Goodlett and Li (2002) *Funct Integr Genomics* 2:138-153

Determining Sequence of Peptide from Fragment Ions

- Fragment ions are predominately due to cleavage of the amide bond in peptide
- b-ions contain N-terminus of peptide; y-ions contain C-terminus of peptide
- Subscript indicates the number of amino acids included in peptide

B) $y_{14} y_{13} y_{12} y_{11} y_{10} y_9 y_8 y_7 y_6 y_5 y_4 y_3 y_2 y_1$
 H G V T V L T A L G A I L K
 b1 b2 b3 b4 b5 b6 b7 b8 b9 b10 b11 b12 b13 b14

Seq #	b	y	+1
H	1	138.1	- 14
G	2	195.1	1269.8 13
V	3	294.1	1212.8 12
T	4	395.2	1113.7 11
V	5	494.3	1012.7 10
L	6	607.3	913.6 9
T	7	708.4	800.5 8
A	8	779.4	699.5 7
L	9	892.5	628.4 6
G	10	949.5	515.4 5
A	11	1020.6	458.3 4
I	12	1133.7	387.3 3
L	13	1246.7	274.2 2
K	14	-	161.1 1

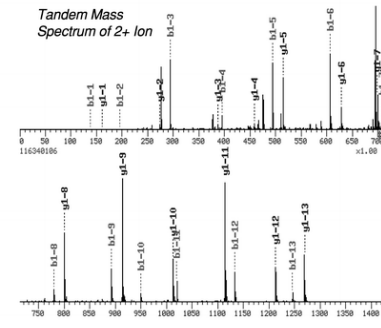


Image from Goodlett and Li (2002) *Funct Integr Genomics* 2:138-153

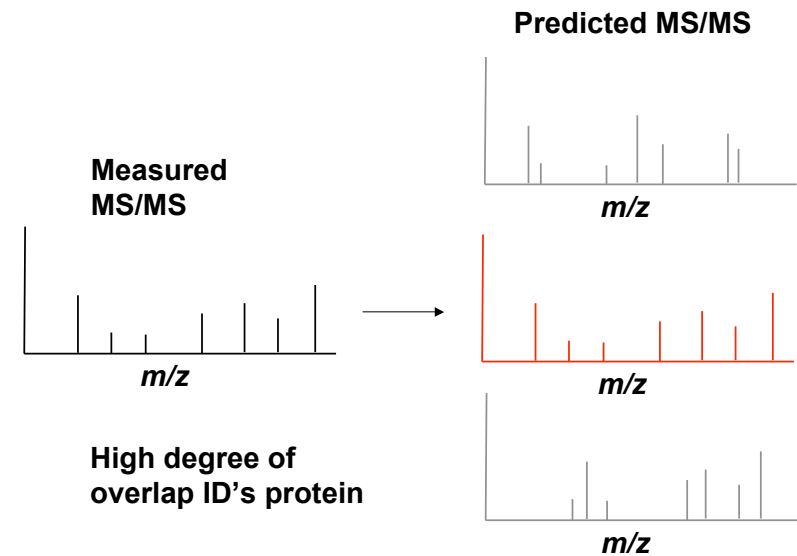
Protein ID for each peptide is done by:

measure MS/MS spectrum for a peptide

find all possible peptides in database that have same mass as measured peptide

calculate and compare hypothetical MS/MS spectra with measured MS/MS spectrum

correct peptide (protein) will have the greatest similarity



MS/MS-Based Protein ID

Advantages

Can ID proteins from mixtures

ID further strengthened by multiple peptides

Can be done without gel separation

Disadvantages

Demands much more skill and sophisticated instrument
on-line chromatography

tandem MS capabilities

More prone to failure from unexpected modifications