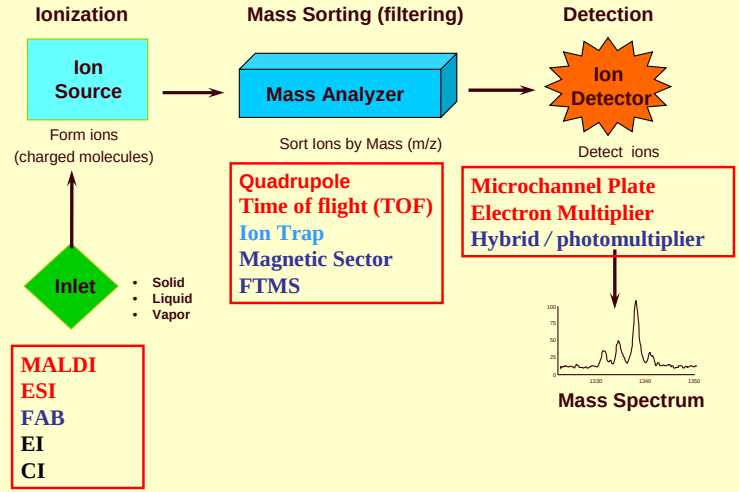
Mass Spectrometry 101

(continued)

Hackert - CH 370 / 387D

Based in part on Lecture Notes from
"An Introductory Lecture On Mass Spectrometry Fundamentals"
Presented to the Sandler Mass Spectrometry Users' Group,
University of California San Francisco, and
"Fundamentals of Mass Spectrometry – Based Proteomics"
by Doug Sheeley – Division of Biomedical Technology, National Center for Research Resources

Summary: acquiring a mass spectrum



Mass Spectrometry

Introductory Example: mass spectrum of water

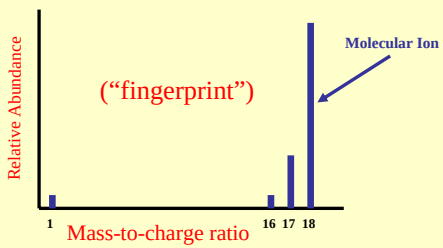
- **Electron Impact (EI) - Hard (fragments) / 1000 Da**

H_2O + fast electron → $[H_2O]+$ + 2 electrons

+ fragments ([OH]+, O+, H+)

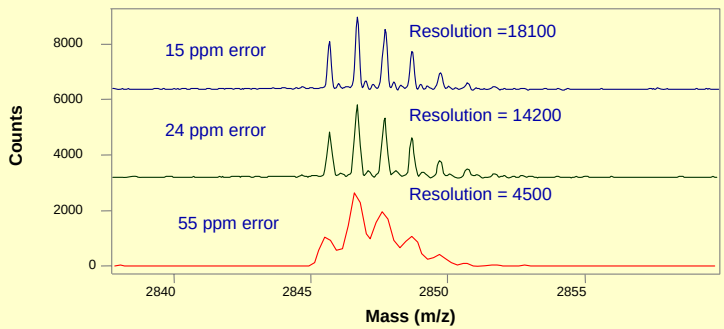
Fragmentation pattern

$[H_2O]+$	18
[OH]+	17
O+	16
H+	1

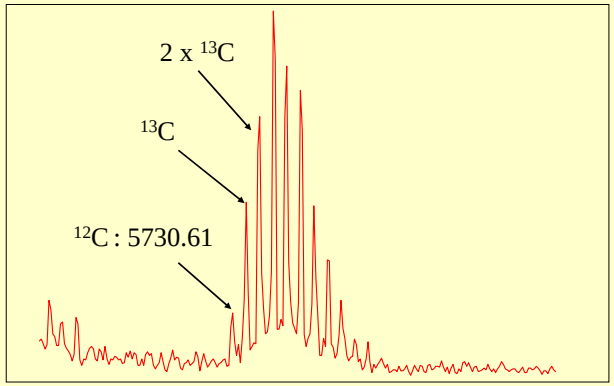


Mass measurement accuracy depends on resolution

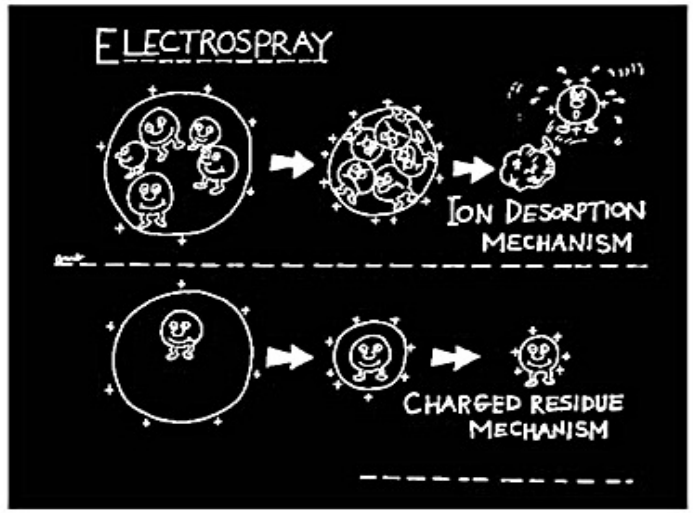
High resolution means better mass accuracy



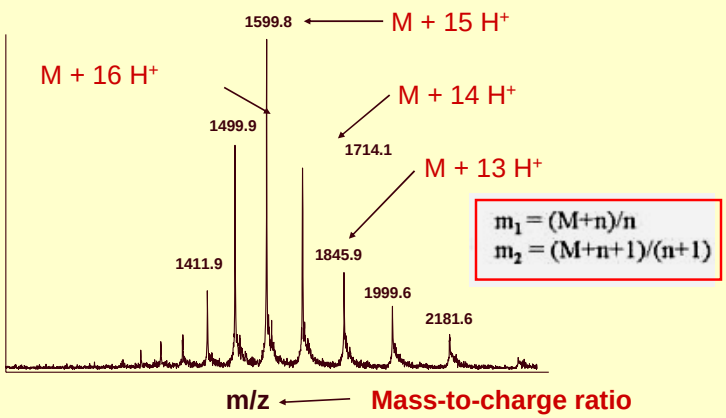
Mass spectrum of insulin



Insulin has 257 C-atoms. Above this mass, the monoisotopic peak is too small to be very useful, and the average mass is usually used.

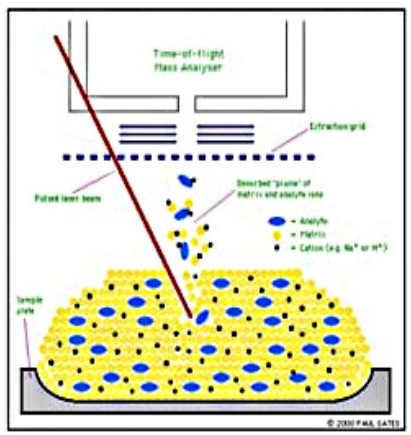


ESI Spectrum of Trypsinogen (MW 23,983)



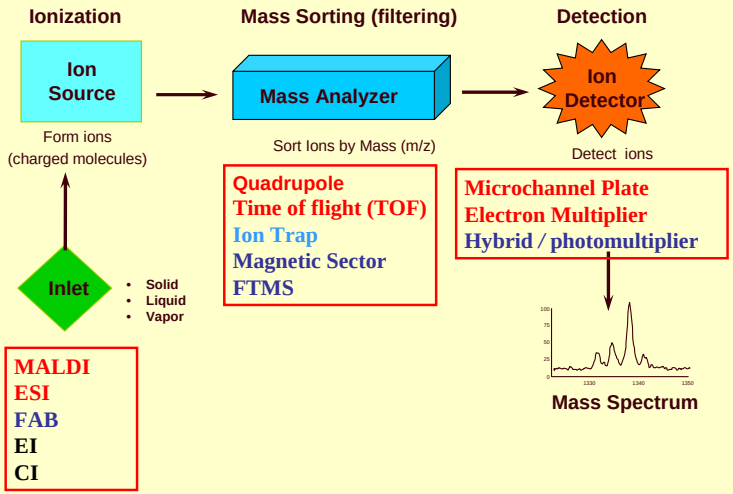
MALDI (Matrix Assisted Laser Desorption Ionization)

"Nuke It"



Analyte:
10 – 1000 fmol
1 – 500 kDa

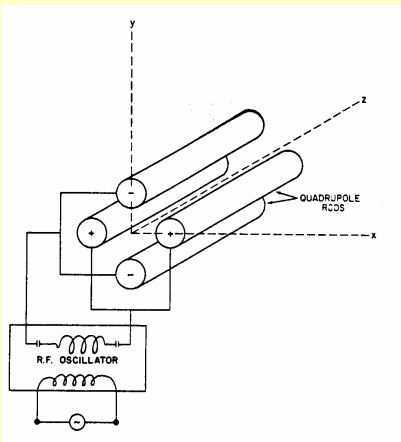
Summary: acquiring a mass spectrum



Mass analyzers separate ions based on their mass-to-charge ratio (m/z)

- Operate under **high vacuum** (keeps ions from bumping into gas molecules)
- Actually measure **mass-to-charge ratio** of ions **(m/z)**
- Key specifications are resolution, mass measurement accuracy, and sensitivity.
- Several kinds exist: for bioanalysis, quadrupole, time-of-flight and ion traps are most used.

Quadrupole Mass Analyzer



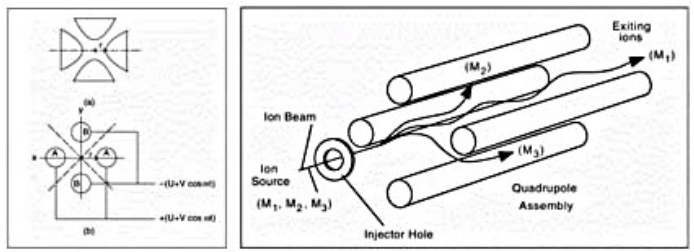
Uses a combination of RF and DC voltages to operate as a mass filter.

- Has four parallel metal rods.
- Lets one mass pass through at a time.
- Can scan through all masses or sit at one fixed mass.

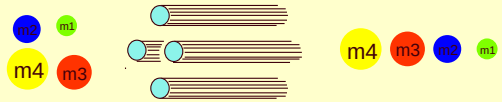
Mass Analyzers: The Quadrupole Mass Filter

A potential of ~100-1000 V is applied alternately to the opposing pairs of rods at a frequency of a few MHz. At a specific combination of DC & RF, an m/z has a stable trajectory through the rods, and all other m/z are lost. The mass range is scanned as the voltages are swept from min to max, but at constant DC/RF ratio.

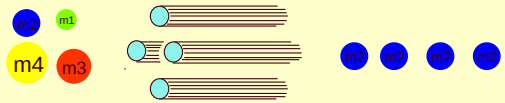
Faster Scanning than sector instruments (but not as fast as ion traps or TOF).
Mass Range generally m/z 0-2000 or 0-4000.
Facile MS/MS using Triple Quadrupole (Q-q-Q) analyzer.
Exquisitely sensitive in selected ion monitoring (both analyzers parked at one m/z).
Largely replaced by the ion trap and hybrid Q-q-TOF for biopolymer analysis.



Quadrupoles have variable ion transmission modes

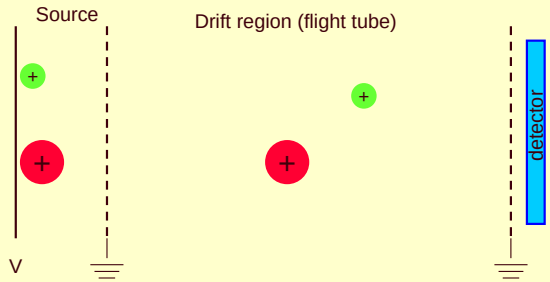


mass scanning mode

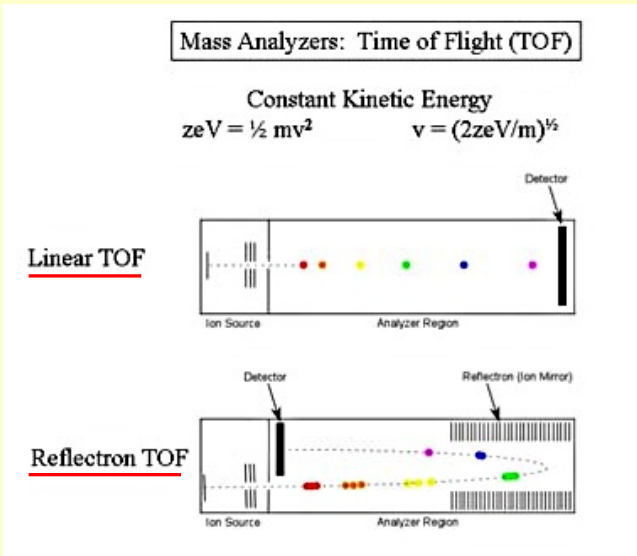


single mass transmission mode

Time-of-flight (TOF) Mass Analyzer

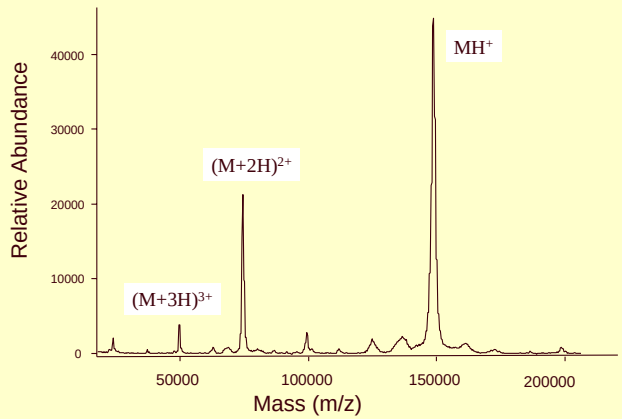


- Ions are formed in pulses.
- The drift region is field free.
- Measures the time for ions to reach the detector.
- Small ions reach the detector before large ones.



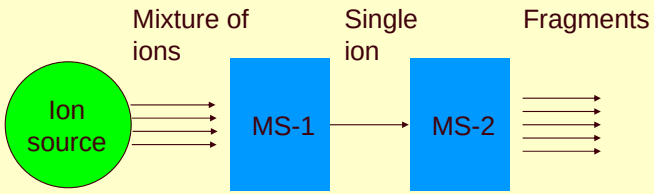
The mass spectrum shows the results

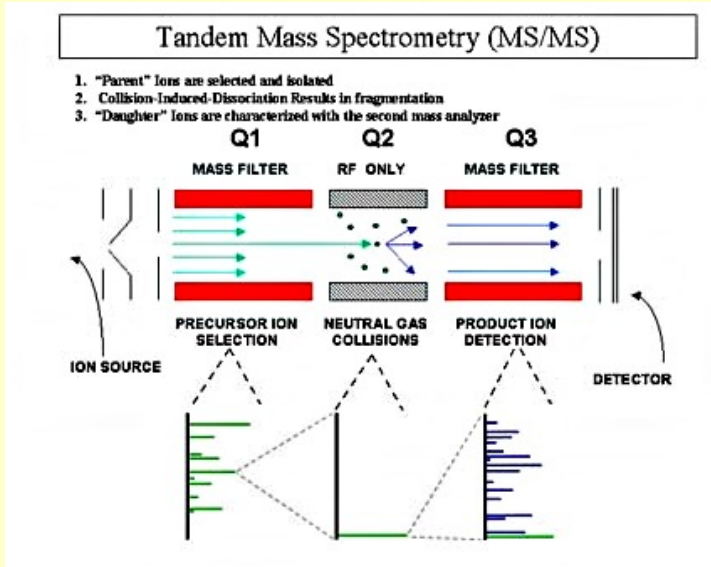
MALDI TOF spectrum of IgG

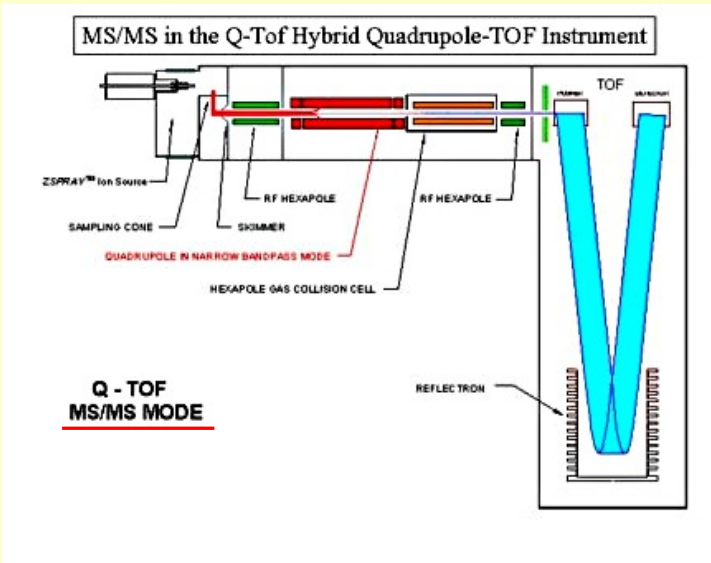


What is MSMS?

MS/MS means using two mass analyzers (combined in one instrument) to select an analyte (ion) from a mixture, then generate fragments from it to give structural information.



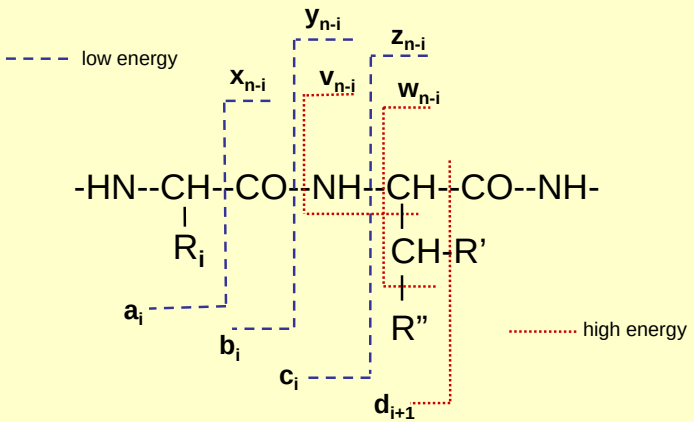




Tandem Mass Spectrometry (MS/MS) is the Method of Choice for Sequence Analysis of Peptides

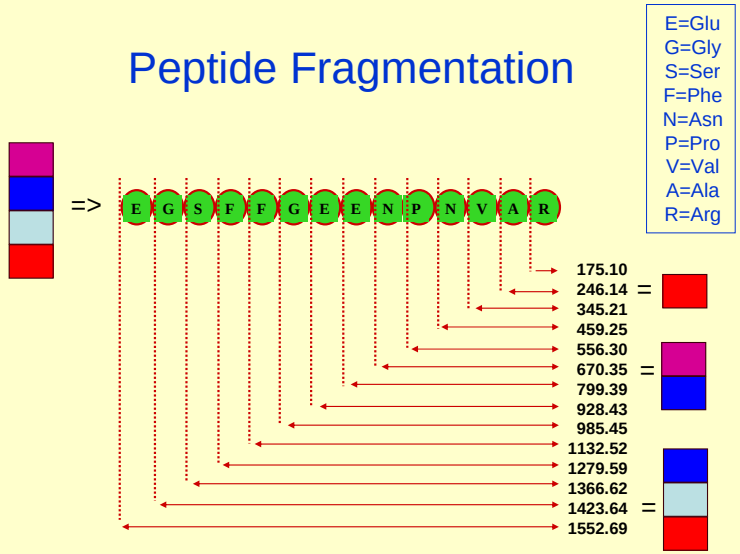
Speed
Sensitivity
Tolerance for Amino-terminal Blocking Groups
High Specificity for Protein Identification

Cleavages Observed in MS/MS of Peptides

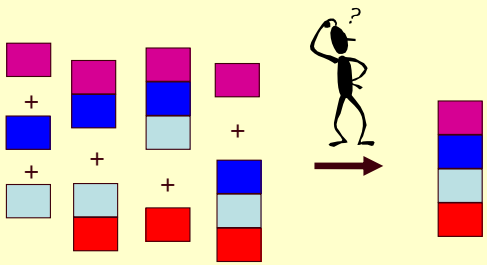


CID (Collision InDuced) Spectra – adds **sequence data** to **mass mapping** for improved database identification!

Peptide Fragmentation

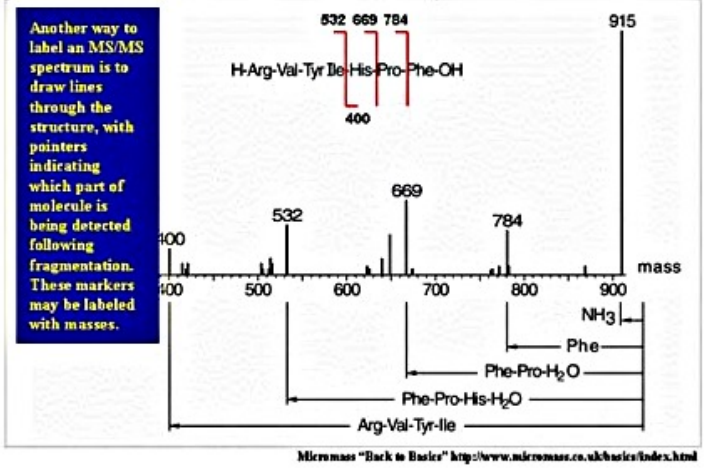


Interpretation of an MSMS spectrum to derive structural information is analogous to solving a puzzle



Use the fragment ion masses as specific pieces of the puzzle to help piece the intact molecule back together

MS/MS of Angiotensin III:
selection and fragmentation of the $(M+H)^+$ molecular ion at m/z932



Mass Spec

Applications in Proteomics and Systems Biology

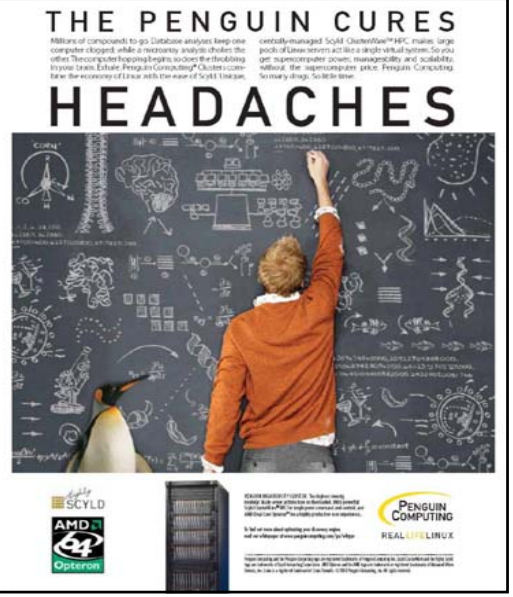
Genomics

Proteomics

Interactomics

Systems Biology –

None of these fields of research would be possible without **Bioinformatics,** which would not be possible with lots of **computing power!**

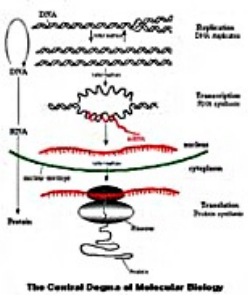


The Proteome

All an organism's cells carry the same Genome, and it is Static. Genomes do not describe function. They are a parts list.

Different cells express different proteins. The type and quantity of this expression changes.

The Proteome is Dynamic. It is the total of all proteins expressed by a particular *cell* at a given *time*, under specific *conditions*.



A Proteome cannot be studied the way a Genome is sequenced. There has to be a specific biological question behind an experiment. The questions may be either very broad or strictly defined.

Fundamentals of Mass Spectrometry – Based Proteomics

Doug Sheeley
Division of Biomedical Technology
National Center for Research Resources

Fundamentals of Mass Spectrometry – Based Proteomics

Purpose:

To convey basic concepts in proteomics and biological mass spectrometry, in order to build a working vocabulary and a basis for further study

Outline:

1. Proteomics
2. Mass Spectrometry- ion sources and mass analyzers
3. Protein Chemistry in the context of proteomics
4. MALDI-TOF/MS for Peptide Mass Mapping
5. LC/MS/MS for Peptide Sequence Analysis

Report

Practical Proteomics 1-2/2006

Proteomics Education, an Important Challenge for the Scientific Community: Report on the Activities of the EuPA Education Committee

European Proteomics Association (EuPA)

EuPA Tutorial Program (preliminary draft)
Fundamentals and Core Techniques

Protein Chemistry	
	Amino acid chemistry/functionality
	PTM natural chemical/enzymatic modifications
	PTM un-natural chemical/enzymatic modifications
	Protein function families: E.C: GO classification
	X-ray principles
	NMR principles
	Protein substructure principles
	Protein structure families
	Membrane protein structure/function
	Extracellular protein structure/function
Protein-protein Interaction	
	Protein complex isolation & examples
	MS-TAP approach to complexes
	Two-hybrid approach
	Biacore, microcalorimetry & CD, FT, ...
DNA/RNA Techniques	
	DNA cloning & sequencing
	RNA structure determination
	Microarray formats
	SAGE
	SNP, methylation, CGH analysis
Separation Science	
	Affinity chromatography
	Free flow electrophoresis
	CZE
	Centrifugation
	HPLC
	2D-PAGE
Protein Expression	
	Antibody generation and use
	Phage display
	Protein arrays
	Tissue arrays
	HT cloning & expression library structure
	HT crystallisation

MS Basics	
	MALDI ionisation
	ESI ionisation
	TOF
	Quads
	Ion-trap, linear & 3D
	FT-ICR, Orbitrap
	Detectors
	Scan modes
Metabolomics	
	GC-MS approaches & derivatisation chemistry
	ESI-MS approaches & derivatisation chemistry
	NMR approaches
	Pathway analysis & modelling EcoCYC
Applied Technologies	
	Microfluidics
	Automation
	Fluorescent labeling, DNA sequencing, microarr
Bioinformatics/Systems Biology	
	Sequence homology searching
	Protein id by MALDI
	Protein id by MS/MS
	ID verification principles, Prophet, *etc.*
	Array analysis
	Database structure
	Relevant stat applications
	Advanced data mining techniques
	Web databases
	Experimental design principles

insight review articles

NATURE | VOL 422 | 13 MARCH 2003 | www.nature.com/nature

Mass spectrometry-based proteomics

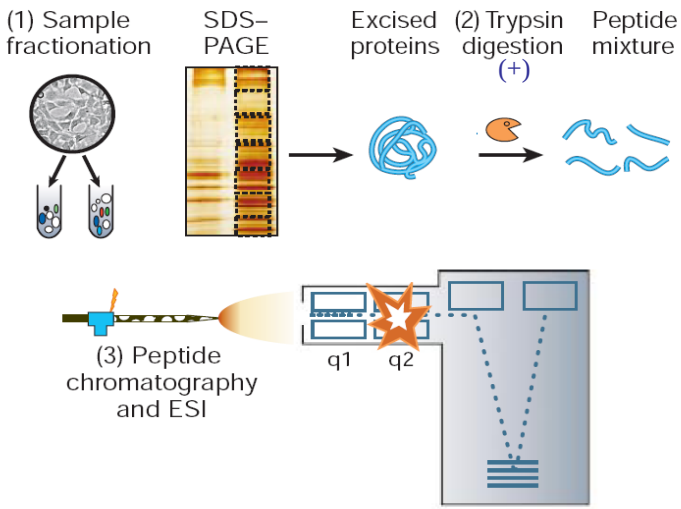
Ruedi Aebersold* & Matthias Mann†

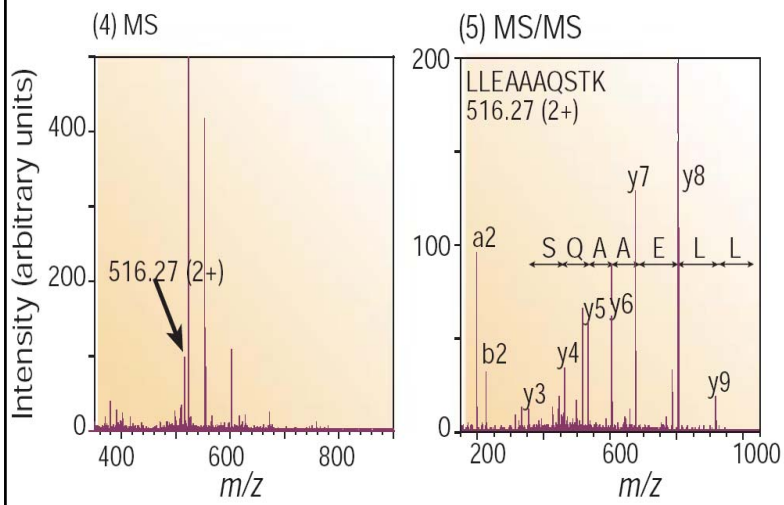
*Institute for Systems Biology, 1441 North 34th Street, Seattle, Washington 98103-8904, USA (e-mail: raebersold@systemsbiology.org)
†Center for Experimental BioInformatics (CEBI), Department of Biochemistry and Molecular Biology, University of Southern Denmark, Campusvej 55, DK-5230 Odense M, Denmark (e-mail: mann@bmb.sdu.dk)

Recent successes illustrate the role of mass spectrometry-based proteomics as an indispensable tool for molecular and cellular biology and for the emerging field of systems biology. These include the study of protein–protein interactions via affinity-based isolations on a small and proteome-wide scale, the mapping of numerous organelles, the concurrent description of the malaria parasite genome and proteome, and the generation of quantitative protein profiles from diverse species. The ability of mass spectrometry to identify and, increasingly, to precisely quantify thousands of proteins from complex samples can be expected to impact broadly on biology and medicine.

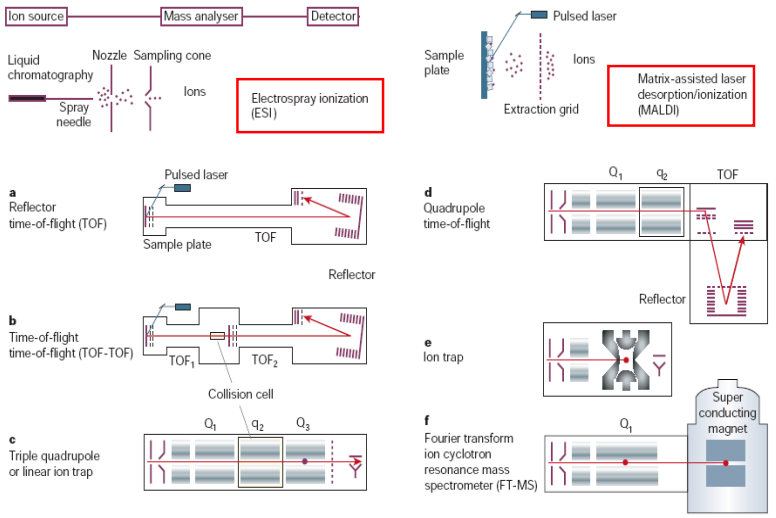
Note: HT Proteomics is restricted to those species where a sequence database exists!

Generic Mass Spectrometry-based Proteomics





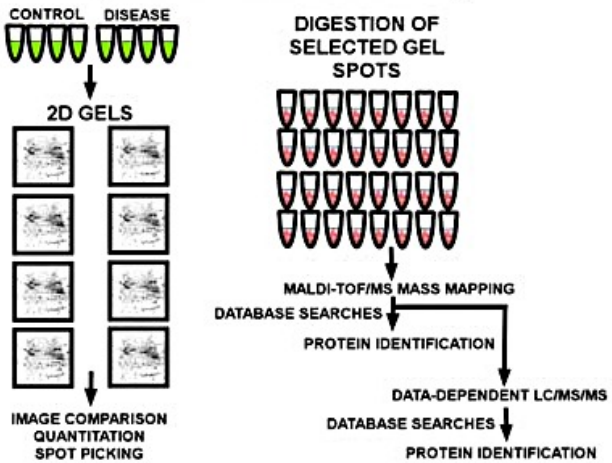
Common MSs Used in Proteomics Research



With the new genomic data bases of model species, such as *Esherichia coli, Saccharomyces cerevisae,* mouse, and human, the sequences of many/most proteins of biological interest will in principle be known, and the problem of characterizing a protein primary structure will be reduced to identifying it in the data base.

Within the past few years research groups have demonstrated how MS can be used for identification of proteins in sequence data bases. One approach is to **cleave the protein with a sequence-specific proteolytic enzyme**, **measure molecular weight** values for the resulting peptide mixture by mass spectrometry, and **search a sequence data base for proteins that should yield these values**. **Search algorithms** can utilize low resolution tandem mass spectra of selected peptides (<3 kDa) from the protein degradation. Yates and coworkers compared the MS/MS sequence data to the sequences predicted for each of the peptides that would be generated from each protein in the data base. **In the PEPTIDESEARCH sequence tag approach of Mann and Wilm, a partial sequence of 2–3 amino acids is assigned from the fragment mass differences in the MS/MS spectrum.** This partial sequence and its mass distance from each end of the peptide (based on the masses of the fragment and molecular ions) are used for the data base search. Often, **a single sequence tag retrieved only the correct protein from the data base**.

Differential Expression Proteomics



Two Dimensional Gel Electrophoresis

Isoelectric focusing is performed on precast gel strips using commercial instruments. Many pH ranges are available. Multiple strips can be run in parallel.

An immobilized pH gradient is created in a polyacrylamide gel strip by incorporating a gradient of acidic and basic buffering groups when the gel is cast.

Resolution is determined by the slope of the pH gradient and the field strength.

Loading capacity depends on gel size and thickness.

In 2D IEF/PAGE, the gel strip from IEF is loaded into a single large well.

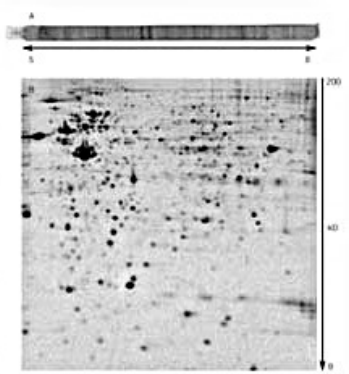
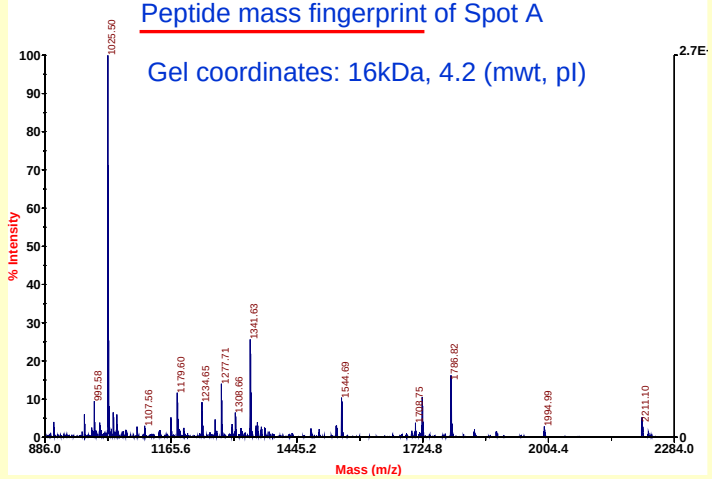


Figure from BioRad Product Literature

Tryptic Digest of ADH: Expected Peptides vs. Those Detected

Peptide mass fingerprint of Spot A

Gel coordinates: 16kDa, 4.2 (mwt, pI)



MS-Fit (by Peter Baker and Karl Clauser) Instructions

A peptide-mass fingerprinting tool from the UCSF Mass Spectrometry Facility that tries to fit a user's mass spectrometry data to a protein sequence in an existing database and thus suggest the identity of the user's protein. The MS input data should be generated by analyzing the peptides produced by the enzymatic digestion of a user's protein.

ProteinProspector Home | MS-Tag | MS-Seq | MS-Edman | MS-Fit at UCSF (San Francisco)
MS-Digest | MS-Product | MS-Comp | DB-Stat | MS-Isotope

Start Search

Database: SwissProt,012601 Instrument: MALDI-TOF Report MOWSE Scores: Pfactor: 0.4
DNA Frame translation: 3
Search Hits: From: msfit Filename: lastres
Save Hits to file: Filename: lastres
Species: All
MW of Protein: (from 1000 Da to 150000 Da) All
Protein pI: (from 3.0 to 10.0) All
Digest: Trypsin
Max. # of missed cleavages: 1
Cysteines modified by: unmodified
N terminus: Hydrogen C terminus: Free Acid
Sample ID (comment): Unknown A
Max. Reported Hits: 15
Possible Modifications Mode (default): Peptide N-terminal Gln to pyroGlu; Oxidation of M; Protein N-terminus Acetylated; Acrylamide Modified Cys
User Defined Modification 1: Phosphorylation of S, T and Y
OR
Homology Mode (select any mode but identity)
Search mode: identity
Min. # matches with NO AA substitutions: 1
Peptide Mass shift: +/- 40.1 Da

Peptide masses are: monoisotopic
Min. # peptides required to match: 3

Peptide Masses
mass tolerance: +/- 15 ppm
Mass (m/z) Charge (z)
905.6874
973.5103
989.6093
995.5787
1007.4948
1024.4374
1025.4959
1025.7433
1037.5184
1045.5657
1090.5471
1106.5649
1139.5205
1164.5909
1165.5664
1179.6002
1184.5950
1193.6111
1233.5911
1234.6510
1263.6850
1267.7091
1277.7065

Mass accuracy tolerance = 15 ppm

This means that the mass is within 0.015 Da at m/z 1000

MS-Fit Search Results

Press stop on your browser if you wish to abort this MS-Fit search prematurely.

Sample ID (comment): **Unknown A**
Database searched: **SwissProt.012601**
Molecular weight search (**1000 - 150000 Da**) selects **90539** entries.
Full pI range: **92236** entries.
Combined molecular weight and pI searches select **90539** entries.
MS-Fit search selects **858** entries (results displayed for top 15 matches).

Considered modifications: | **Peptide N-terminal Gln to pyroGlu** | **Oxidation of M** | **Protein N-terminus Acetylated** | **Acrylamide Modified Cys** |

Min. # Peptides to Match	Peptide Mass Tolerance (+/-)	Peptide Masses are	Digest Used	Max. # Missed Cleavages	Cysteines Modified by	Peptide N terminus	Peptide C terminus	Input # Peptide Masses
3	15.000 ppm	monoisotopic	Trypsin	1	unmodified	Hydrogen (H)	Free Acid (O H)	46

Result Summary

Rank	MOWSE Score	# (%) Masses Matched	Protein MW (Da)/pI	Species	SwissProt.012601 Accession #	Protein Name
1	2.86e+005	9/46 (19%)	16930.2 / 4.56	HUMAN	P16475	MYOSIN LIGHT CHAIN ALKALI, NON-MUSCLE ISOFORM (MLC3NM) (LC17A) (LC17-NM)
2	2.86e+005	9/46 (19%)	16961.2 / 4.46	HUMAN	P24572	MYOSIN LIGHT CHAIN ALKALI, SMOOTH-MUSCLE ISOFORM (MLC3SM) (LC17B) (LC17-GI)
3	2.86e+005	9/46 (19%)	16975.3 / 4.46	RAT	Q64119	MYOSIN LIGHT CHAIN ALKALI, SMOOTH-MUSCLE ISOFORM (MLC3SM)
4	1.77e+004	7/46 (15%)	15730.9 / 4.80	MOUSE	Q60605	MYOSIN LIGHT CHAIN ALKALI, NON-MUSCLE ISOFORM (MLC3NM)
5	1.41e+004	7/46 (15%)	66018.0 / 8.16	HUMAN	P04264	KERATIN, TYPE II CYTOSKELETAL 1 (CYTOKERATIN 1) (K1) (CK 1) (67 KDA CYTOKERATIN) (HAIR ALPHA PROTEIN)
6	1.19e+003	4/46 (8%)	15282.4 / 6.10	STRPU	P32006	PROFILIN
7	420	5/46 (10%)	16983.3 / 4.63	CHICK	P09296	MYOSIN LIGHT CHAIN ALKALI, NON-MUSCLE ISOFORM (FIBROBLAST) (G2 CATALYTIC) (LC17-NM)
8	419	5/46 (10%)	16987.4 / 4.52	CHICK	P02607	MYOSIN LIGHT CHAIN ALKALI, SMOOTH-MUSCLE ISOFORM (GIZZARD) (G2 CATALYTIC) (LC17-GI)
9	391	4/46 (8%)	38545.3 / 8.59	XENLA	P27006	ANNEXIN II TYPE I (LIPOCORTIN II) (CALPACTIN I HEAVY CHAIN) (CHROMOBINDIN 8) (P36) (PROTEIN I) (PLACENTAL ANTICOAGULANT PROTEIN IV) (PAP-IV)
10	286	5/46 (10%)	22156.3 / 5.03	RAT	P16409	MYOSIN LIGHT CHAIN 1, SLOW-TWITCH MUSCLE B/VENTRICULAR ISOFORM
11	262	3/46 (6%)	19590.2 / 9.34	BGMV	P05174	AL2 PROTEIN (19.6 KD PROTEIN)
12	220	5/46 (10%)	21932.2 / 5.03	HUMAN	P08590	MYOSIN LIGHT CHAIN 1, SLOW-TWITCH MUSCLE B/VENTRICULAR ISOFORM (MLC1SB) (ALKALI)
13	211	3/46 (6%)	16990.5 / 6.92	ECOLI	P37052	HYPOTHETICAL 17.0 KDA PROTEIN IN HNR-PURU INTERGENIC REGION
14	202	3/46 (6%)	17947.3 / 5.24	ARATH	P25855	GLYCINE CLEAVAGE SYSTEM H PROTEIN 1, MITOCHONDRIAL PRECURSOR
15	186	3/46 (6%)	16613.9 / 4.63	RAT	P02601	MYOSIN LIGHT CHAIN 3, SKELETAL MUSCLE ISOFORM (A2 CATALYTIC) (ALKALI) (MLC3F)

Detailed Results

1. 9/46 matches (19%). 16930.2 Da, pI = 4.56. Acc. # P16475. HUMAN. MYOSIN LIGHT CHAIN ALKALI, NON-MUSCLE ISOFORM (MLC3NM) (LC17A) (LC17-NM).

m/z submitted	MH+ matched	Delta ppm	start	end	Peptide Sequence (Click for Fragment Ions)	Modifications
995.5787	995.5890	-10.3014	111	119	(R)HVLVTLGEK(M)	
1025.4959	1025.5056	-9.4785	14	21	(K)EAFQLFDR(T)	
1233.5911	1233.5898	1.0857	99	110	(K)EGNGTVMGAEIR(H)	
1354.7187	1354.7331	-10.5955	38	50	(R)ALGQNPTNAEVLK(V)	
1544.6928	1544.6869	3.8248	82	94	(K)DQGTYEDYVEGLR(V)	
1722.8598	1722.8485	6.5620	95	110	(R)VFDKEGNGTVMGAEIR(H)	
1786.8229	1786.8248	-1.0535	80	94	(K)NKDQGTYEDYVEGLR(V)	
1888.0274	1888.0043	12.2526	64	79	(K)VLDPEHPLPMLQTVAK(N)	
2226.1294	2226.1552	-11.6082	99	119	(K)EGNGTVMGAEIRHVLVTLGEK(M)	1Met-ox

37 unmatched masses: 905.6874 973.5183 989.6093 1007.4948 1024.4374 1025.7433 1037.5184 1045.5657 1090.5471 1106.5649 1139.5205 1164.5909 1165.5664 1179.6002 1184.5958 1193.6111 1234.6510 1263.6858 1267.7091 1277.7065 1300.5432 1307.6644 1308.6596 1340.6612 1341.6288 1357.6707 1373.6434 1475.7257 1493.7172 1532.6160 1699.8525 1707.7833 1716.8276 1723.8256 1838.9438 1993.9497 2211.1041

The matched peptides cover **50%** (77/151 AA's) of the protein.
Coverage Map for This Hit (MS-Digest index #): 11572

2. 9/46 matches (19%). 16961.2 Da, pI = 4.46. Acc. # P24572. HUMAN. MYOSIN LIGHT CHAIN ALKALI, SMOOTH-MUSCLE ISOFORM (MLC3SM) (LC17B) (LC17-GI).

m/z submitted	MH+ matched	Delta ppm	start	end	Peptide Sequence (Click for Fragment Ions)	Modifications
995.5787	995.5890	-10.3014	111	119	(R)HVLVTLGEK(M)	
1025.4959	1025.5056	-9.4785	14	21	(K)EAFQLFDR(T)	
1233.5911	1233.5898	1.0857	99	110	(K)EGNGTVMGAEIR(H)	
1354.7187	1354.7331	-10.5955	38	50	(R)ALGQNPTNAEVLK(V)	
1544.6928	1544.6869	3.8248	82	94	(K)DQGTYEDYVEGLR(V)	
1722.8598	1722.8485	6.5620	95	110	(R)VFDKEGNGTVMGAEIR(H)	
1786.8229	1786.8248	-1.0535	80	94	(K)NKDQGTYEDYVEGLR(V)	
1888.0274	1888.0043	12.2526	64	79	(K)VLDPEHPLPMLQTVAK(N)	
2226.1294	2226.1552	-11.6082	99	119	(K)EGNGTVMGAEIRHVLVTLGEK(M)	1Met-ox

37 unmatched masses: 905.6874 973.5183 989.6093 1007.4948 1024.4374 1025.7433 1037.5184 1045.5657 1090.5471 1106.5649 1139.5205 1164.5909 1165.5664 1179.6002 1184.5958 1193.6111 1234.6510 1263.6858 1267.7091 1277.7065 1300.5432 1307.6644 1308.6596 1340.6612 1341.6288 1357.6707 1373.6434 1475.7257 1493.7172 1532.6160 1699.8525 1707.7833 1716.8276 1723.8256 1838.9438 1993.9497 2211.1041

Stable-isotope Protein Labelling for Proteomics