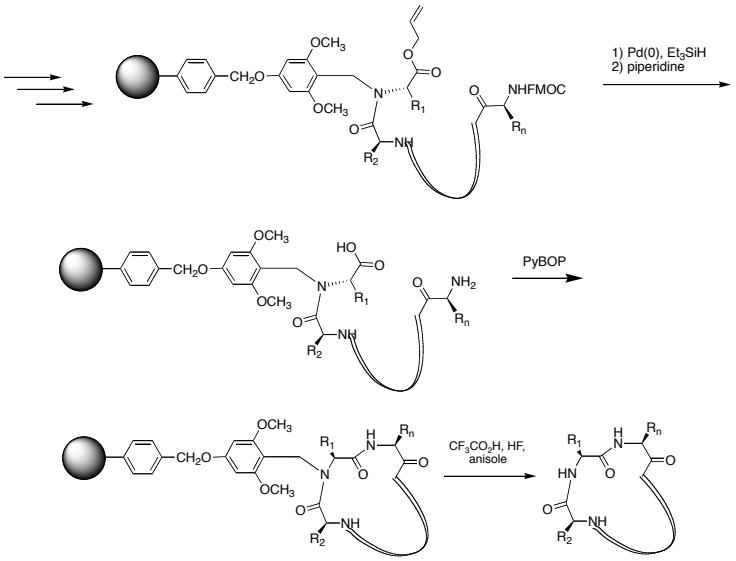
On-suppport cyclization

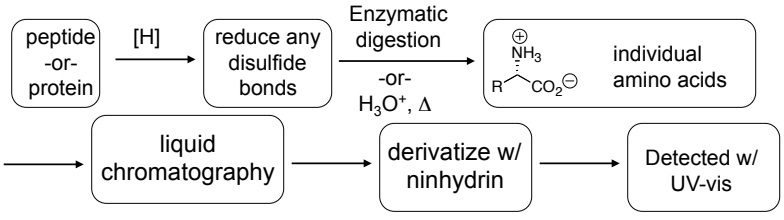


49

Peptide and Protein Analysis

Primary (1°) structure of a peptide or protein is the amino acid sequence

Amino acid analyzer- automated instrument to determine the amino acid content of a peptide or protein. Individual amino acids are separated by hplc, then detected by post-column derivatization

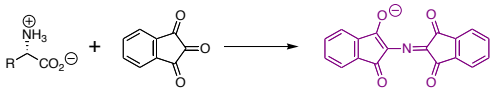


Different amino acids have different chromatographic mobilities (retention times)

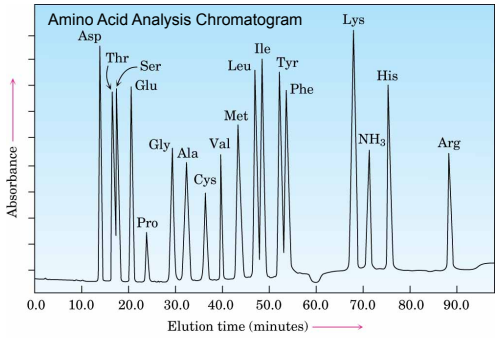
1972 Nobel Prize in Chemistry
William Stein
Stanford Moore

50

Reaction of primary amines with ninhydrin



Intense purple color



So, why is it necessary to use a post- rather than pre-column derivatization protocol?

Why are there are only 17 AA's in the chromatogram?

51

Fluorescence Detection- less background, greater sensitivity, lower detection limits

Absorption spectroscopy- wavelength that light absorbs, moloecules are in an electronically excited state

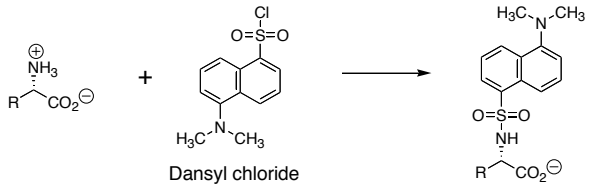
Emission spectroscopy- the excited molecules relax by emission of a photon.

Fluorescence- excitation wavelength and emission wavelength are different. Molecule will emit light at longer (lower energy) wavelength than is absorbs.

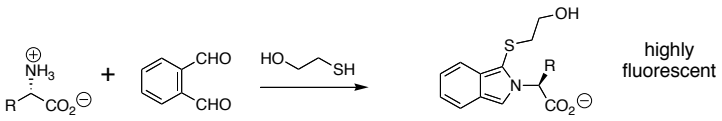
52

Fluorescent tags

Dansyl- detected by UV or fluorescence

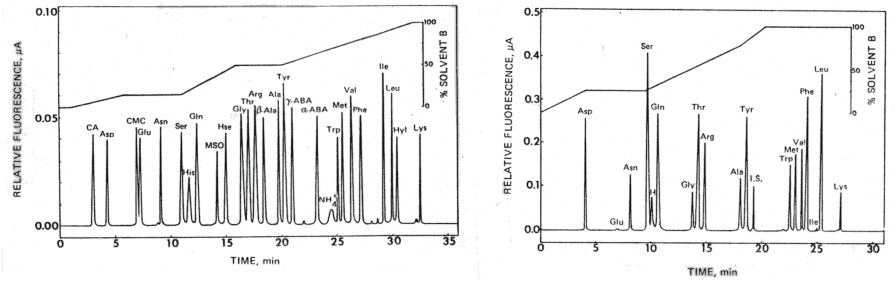


OPA (o-phthalaldehyde)- detected by fluorescence



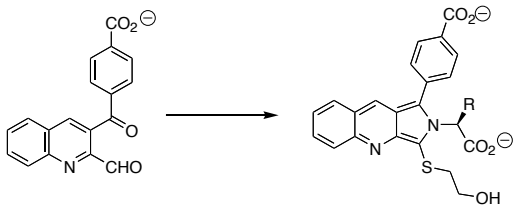
53

Reversed-phase (C-18) HPLC Trace
5 pmols amino acids w/ OPA, $HOCH_2CH_2SH$



54

Attomol detection w/ laser induced fluorescence



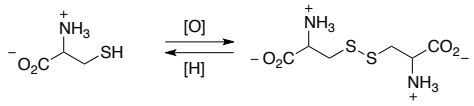
excitation: 488 nm
emission: 560 nm

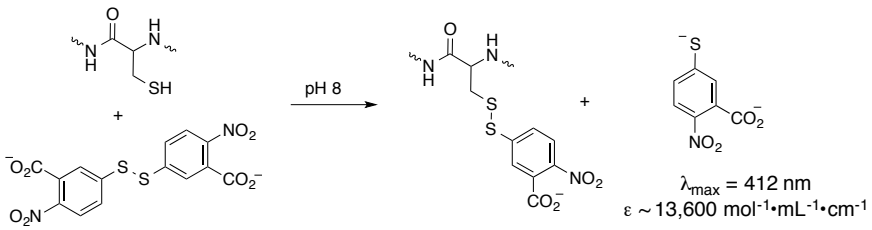
10^{-3}	milli
10^{-6}	micro
10^{-9}	nano
10^{-12}	pico
10^{-15}	fempto
10^{-18}	atto
10^{-21}	zepto

Avagadro's number 10^{23}

55

Cysteine vs. Cystine





Ellman's Reagent

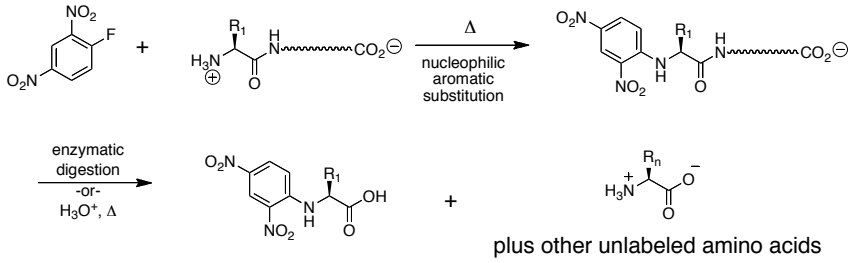
λ_{max} = 412 nm
$\varepsilon \sim 13{,}600\ mol^{-1} \cdot mL^{-1} \cdot cm^{-1}$

56

Peptide and Protein Sequences:

primary (1°) structure- amino acid sequence

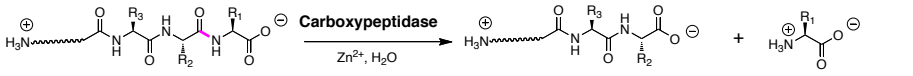
N-labeling with Sanger's reagent: Sanger's (2,4-dinitrofluorobenzene) reagent reacts with the N-terminal amino group and has a diagnostic UV absorbance that is detected after enzymatic digestion and amino acid analysis



N-terminal amino acid is specifically labeled with a unique UV chromophore

C-terminal sequencing:

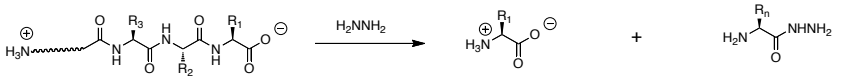
Carboxypeptidase- enzyme that hydrolyzed amide bonds of a peptide or protein starting from the C-termial end (exopeptidase)



peptide has a new C-terminal AA

derivatize and identify by HPLC

Hydrolyze peptide with hydazine (H_2N-NH_2)

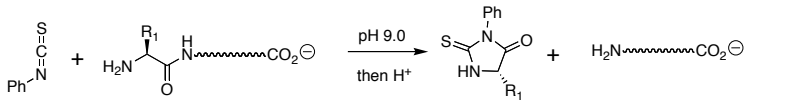


C-terminal AA is still an amino acid

All other AA's are converted to the hydrazides

Edman Degradation: chemical method for the sequential cleavage and identification of the amino acids of a peptide, one at a time starting from the N-terminus.

Reagent: Ph-N=C=S, phenylisothiocyanate



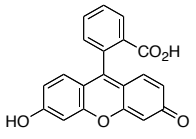
N-phenylthiohydantoin: separated by HPLC, detected by UV-vis

-1 peptide with a new N-terminal amino acid (repeat degradation cycle)

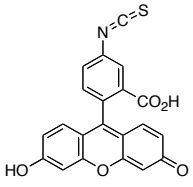
59

Peptide sequencing by Edman degradation: Monitor the appearance of N-phenylthiohydantoin over time to get the peptide sequence. Good for peptides up to ~ 25 amino acids long. Longer peptides and proteins must be cut into smaller fragments before Edman sequencing

Fluorescent Edman sequencing reagent



Fluorescein
(a common fluorescent dye)



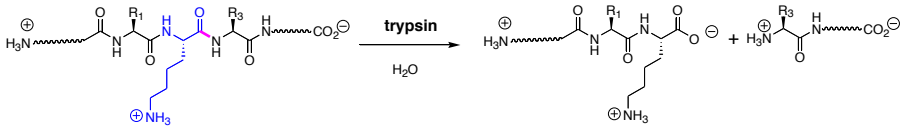
Fluorescein Isothiocyanate
(a fluorescent Edman reagent)

60

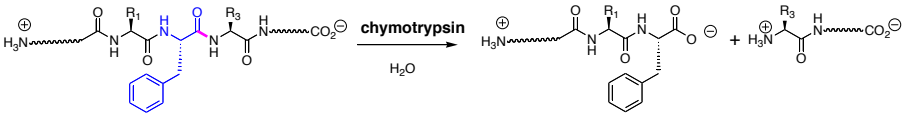
Enzymatic and chemical cleavage of peptides and proteins at defined sites

Enzymatic

- trypsin: cleaves at the C-terminal side of basic residues, Arg, Lys but not His

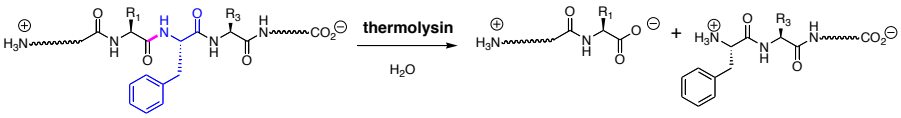


- chymotrypsin: cleaves at the C-terminal side of aromatic residues Phe, Tyr, Trp



61

- thermolysin: cleaves at the N-terminal side of hydrophobic residues Phe, Trp, Leu



Chymotrypsin cleavage products	Trypsin cleavage products
Tyr	Arg
Asp-Asn-Gln	Leu-Lys
Gly-Gly-Phe	Ile-Arg-Pro-Lys
Leu-Arg-Arg-Ile-Arg-Pro-Lys-Leu-Lys-Trp	Tyr-Gly-Gly-Phe-Leu-Arg
	Trp-Asp-Asn-Gln

Trypsin:

Tyr-Gly-Gly-Phe-Leu-Arg Arg Ile-Arg-Pro-Lys Leu-Lys Trp-Asp-Asn-Gln

Chymotrypsin:

Tyr Gly-Gly-Phe Leu-Arg-Arg-Ile-Arg-Pro-Lys-Leu-Lys-Trp Asp-Asn-Gln

62

Other Commonly Used Protein Digest Reagents

Glu-C – cleaves to the C-terminal side of Glu residues
(cleavage at Asp is 100-300 times slower)

Asp-N – cleaves to the N-terminal side of Asp residues
and cysteic acid

Lys-N – cleaves to the N-terminal side of Lys residues

Lys-C – cleaves to the C-terminal side of Lys residues

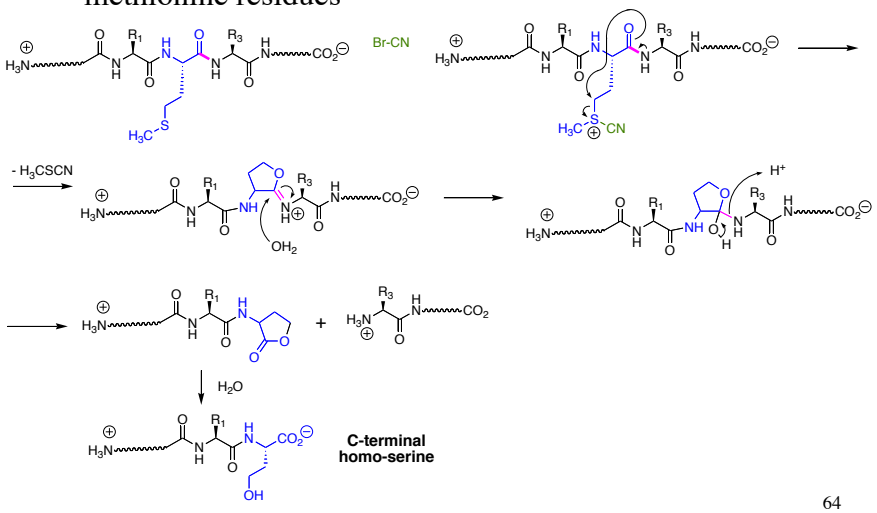
Arg-C – cleaves to the C-terminal side of Arg residues

Non-specific proteases: pepsin, proteinase K, subtilysin

63

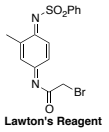
Chemical cleavage of peptides and proteins at defined sites

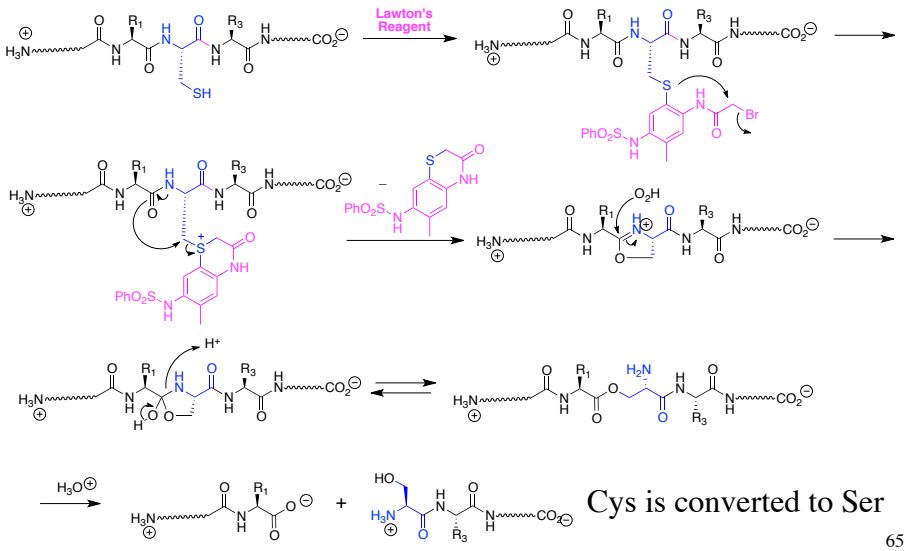
- Cyanogen bromide (Br-CN): cleaves to the C-terminal side of methionine residues



64

• Lawton's Reagent:
cleavage at N-terminal side of cysteine residues





Cys is converted to Ser

65

EPIDERMAL GROWTH FACTOR (EGF)

H_2N-ASN1•SER2•TYR3•PRO4•GLY5•CYS6•PRO7•SER8•SER9•**TYR10**•
ASP11•GLY12•**TYR13**•CYS14•LEU15•ASN16•GLY17•GLY18•VAL19•
CYS20•**MET21**•HIS22•ILE23•GLU24•SER25•LEU26•ASP27•SER28•
TYR29•THR30•CYS31•ASN32•CYS33•VAL34•ILE35•GLY36•**TYR37**•
SER38•GLY39•ASP40•**ARG41**•CYS42•GLN43•THR44•**ARG45**•ASP46•
LEU47•**ARG48**•TRP49•TRP50•GLU51•LEU52•ARG53-CO_2H

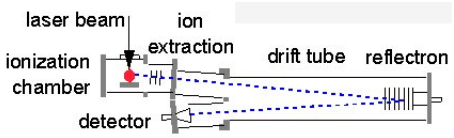
Trypsin **Chymotrypsin** **Cyanogen Bromide**

Disulfides bridges at: **Cys6 – Cys20**
Cys14 – Cys31
Cys33 – Cys42

S. Cohen et al. ***J. Biol. Chem.*** **1972, *247*, 5928-5934**
1972, *247*, 7612-7621
1973, *248*, 7669-7672

66

Peptide sequencing by tandem mass spectrometry
Ionization: SIMS (secondary ion mass spectrometry)
Time-of-flight (TOF) mass spectrometer



Methods to get large, polar molecules into the gas phase for MS analysis

FAB: Fast Atom Bombardment
MALDI: Matrix-Assisted Laser Desorption Ionization
ESI: Electrospray Ionization

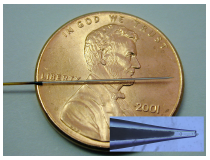
Mass spectrometry gives mass/charge (*m/z*) ratio

"Introduction to Proteomics: Tools for the New Biology," Liebler, D. C., Humana Press: **2002**

67

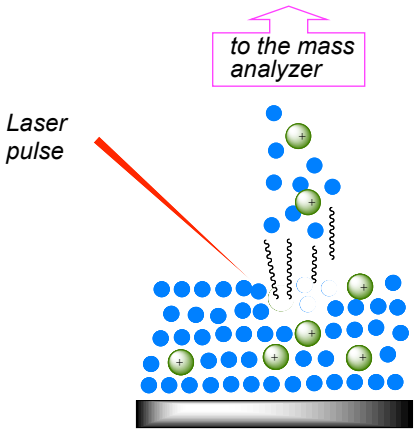
Mass spectrometry is a gas phase technique. Peptides (and proteins) are charged, polar, high molecular weight molecules (ions). How can peptides and proteins be coaxed into the gas phase?

Electrospray ionization (ESI): analyte is introduced into the mass spectrometer as an aerosol.

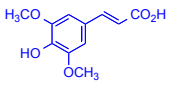


68

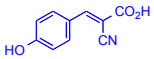
MALDI ionization (matrix-assisted laser desorption): analyte is co-crystallized with an organic molecule that has an intense UV absorption. A laser that is tuned to the absorption of the matrix, is "pulsed" at the MALDI matrix and energy is indirectly transferred to the analyte.



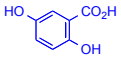
2002 Nobel Prize in Chemistry
John Fenn (ESI)
Koichi Tanaka (MALDI)



Sinapinic acid



α-cyano-4-hydroxycinnamic acid (CHCA)



2,5-dihydroxybenzoic acid (DHB)

69

Mass Spectrometry (MS): measures the mass to charge ratio (*m/z*)

Dalton (Da) or mass unit (u) = units for measuring molecular masses.
One Da. = 1/12 the mass of the ^{12}C atom

Monoisotopic mass – sum of the exact masses of the most abundant isotope of each element in a molecule

Average mass – sum of the averaged masses of each element in a molecules, weighted according to isotopic abundance.

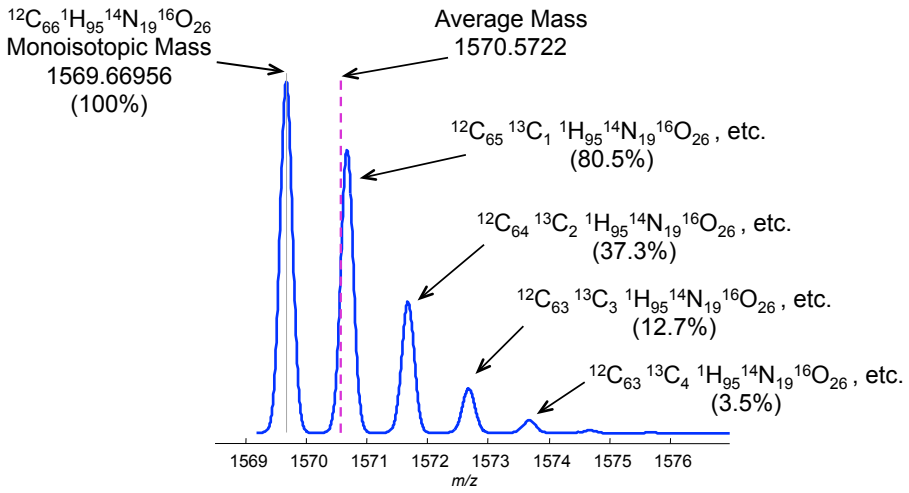
Nominal mass – mass calculated using the integer mass of the most abundant isotope for each element (H=1, C=12, O=16, N=14, etc.)

Isotope	Mass	Natural Abundance
^{1}H	1.0078	99.99%
^{2}H	2.0141	0.015
^{12}C	12	98.89
^{13}C	13.0034	1.11
^{14}N	14.0031	99.64
^{15}N	15.0001	0.36
^{16}O	15.9949	99.76
^{17}O	16.9991	0.04
^{18}O	17.9992	0.2

Isotope	Mass	Natural Abundance
^{31}P	30.9737	100
^{32}S	31.9721	95
^{33}S	32.9715	0.76
^{34}S	33.9679	4.22
^{36}S	35.9671	0.02

70

Glu-Gly-Val-Asn-Asp-Asn-Glu-Glu-Gly-Phe-Phe-Ser-Ala-Arg
(EGVNDNEEGFFSAR)
$C_{66}H_{95}N_{19}O_{26}$

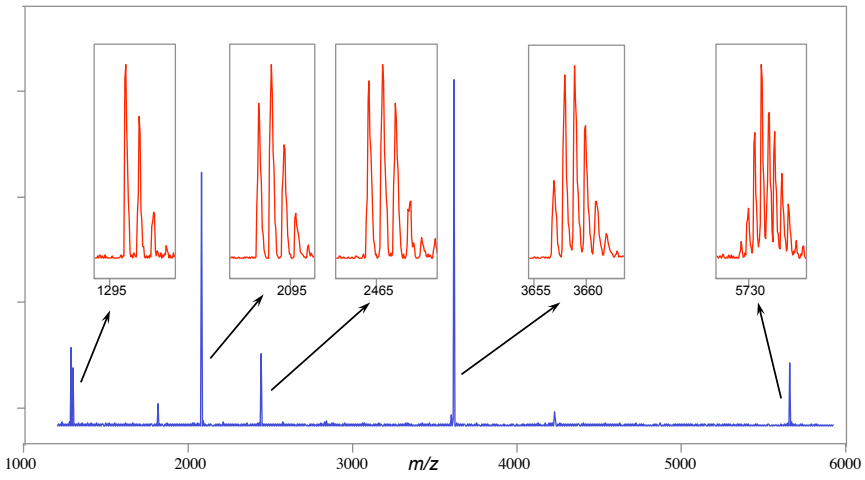


Slide adapted from lecture material developed by A. Burlingame, S. Guan, and M. Baldwin (UCSF) entitled "*Mass Spectrometry and Proteomics*"

http://www-personal.umich.edu/~junhuay/pattern.htm

71

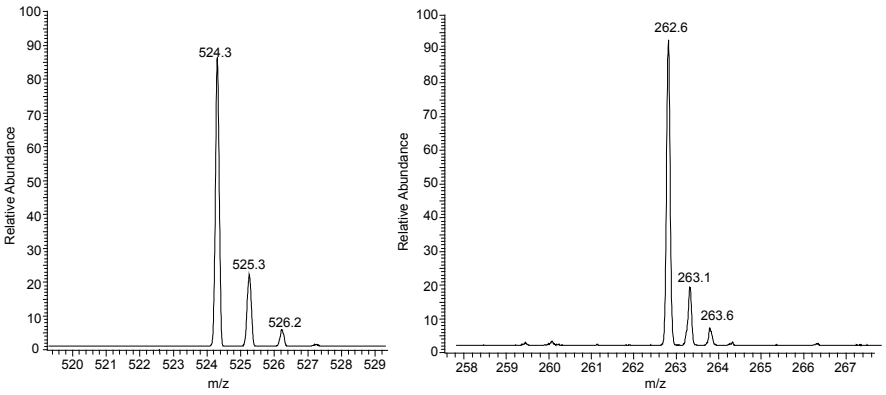
As the number of atoms in the molecule increases, the pattern of masses due to the presence of isotopes will change



Slide adapted from lecture material developed by A. Burlingame, S. Guan, and M. Baldwin (UCSF) entitled "*Mass Spectrometry and Proteomics*"

72

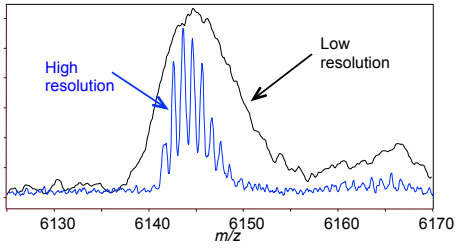
What does the isotopic distributions tell us?



Slide adapted from lecture material developed at the University of Lund

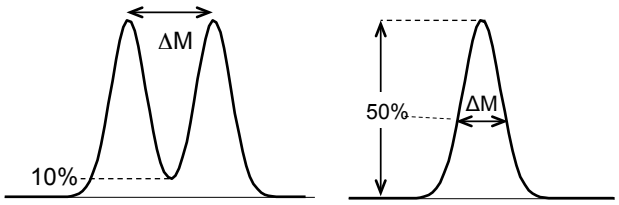
73

Resolution and Resolving Power (RP): terms used interchangably



The smallest mass difference (ΔM) between peaks such that the valley between them is a specified fraction of the peak height

Full Width Half Maximum (FWHM): Width of a single peak measured at 50% peak apex.



Slide adapted from lecture material developed by A. Burlingame, S. Guan, and M. Baldwin (UCSF) entitled "*Mass Spectrometry and Proteomics*"

74

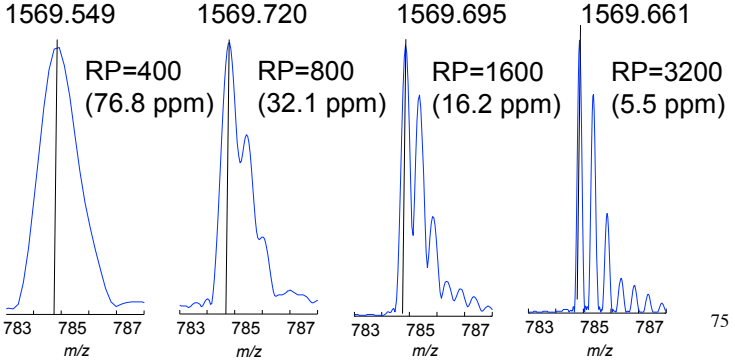
Mass Accuracy (MA) - the difference between the experimental mass (M_{exp}) and the theoretical value (M_{calc}), calculated from elemental composition.

$$MA = \frac{M_{\exp} - M_{calc}}{M_{calc}}$$ (ppm for high resolution MS)

M_{exp} = 1569.684, M_{calc} = 1569.6696 accuracy = 9.2 ppm

High resolution means better mass accuracy

m/z:	784.775	784.860	784.848	784.830
M:	1569.549	1569.720	1569.695	1569.661



Slide adapted from lecture material developed by A. Burlingame, S. Guan, and M. Baldwin (UCSF) entitled "*Mass Spectrometry and Proteomics*"

75

Peptide Mass Fingerprinting : Proteins (or peptides) are digested in a predictable way and the masses of the resulting peptide fragments are unique enough to identify the protein.

Requires a database of known sequences and search software to compare (score) the experimentally observed masses with the calculated masses in the database .

m/z =	1529 ± 1 Da	478	peptide fragments from
	1529.7 ± 0.1	164	mouse/human genome
	1529.73 ± 0.01	25	
	1529.7340 ± 0.001	4	
	1529.7348 ± 0.0001	2	

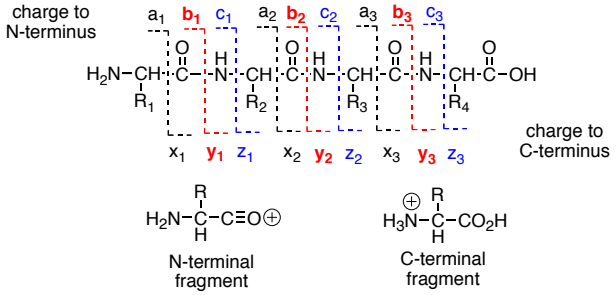
76

Many peptides and proteins give multiply charged ions

CID: collision induced dissociation

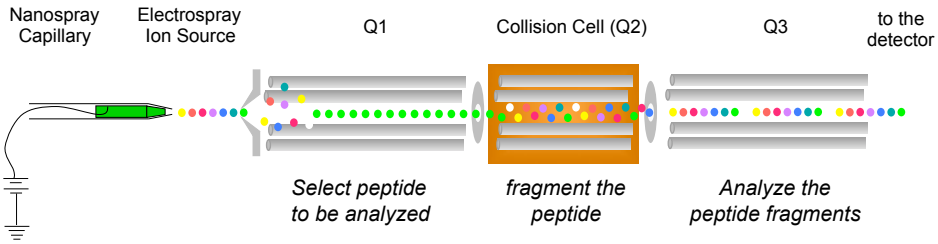
$[M + H]^+$ → accelerated into MS → collision cell (He, Ar, Xe) → analyze fragments to get sequence

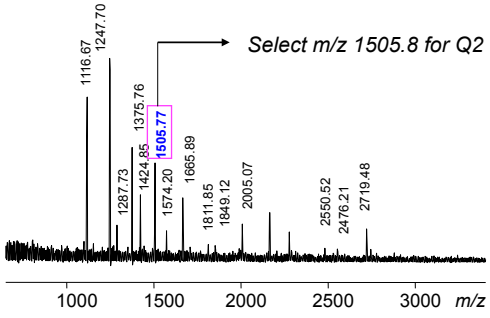
Collision of the [M+H]+ ion with the gas causes it to fragment, analysis of these fragments ions gives sequence information



77

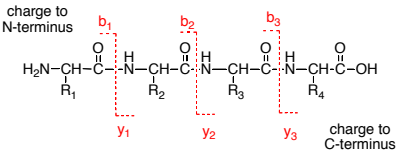
Peptide sequencing by tandem mass spectrometry





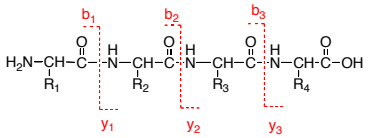
Select m/z 1505.8 for Q2

Peptides fragment in a predictable manner



78

Amino Acids Sorted by Mass



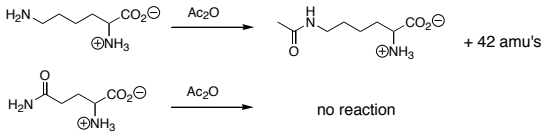
		average	exact	- HN-CHR-CO
Glycine	G	75.07	75.03	57.1
Alanine	A	89.10	89.05	71.1
Serine	S	105.09	105.04	87.1
Proline	P	115.13	115.05	97.1
Valine	V	117.15	117.08	99.1
Threonine	T	119.12	119.06	101.1
Cysteine	C	121.16	121.02	103.1
Isoleucine	I	131.18	131.09	113.2
Leucine	L	131.18	131.09	113.2
Asparagine	N	132.12	132.05	114.1
Aspartic Acid	D	133.11	133.04	115.1
Glutamine	Q	146.15	146.07	128.2
Lysine	K	146.19	146.11	128.1
Glutamic Acid	E	147.13	147.13	129.1
Methionine	M	149.21	149.05	131.2
Histidine	H	155.16	155.02	137.1
Phenylalanine	F	165.19	165.19	147.2
Arginine	R	174.20	174.11	156.2
Tyrosine	Y	181.19	181.07	163.2
Tryptophan	W	204.23	204.09	186.2

79

Some ambiguities with MS sequencing

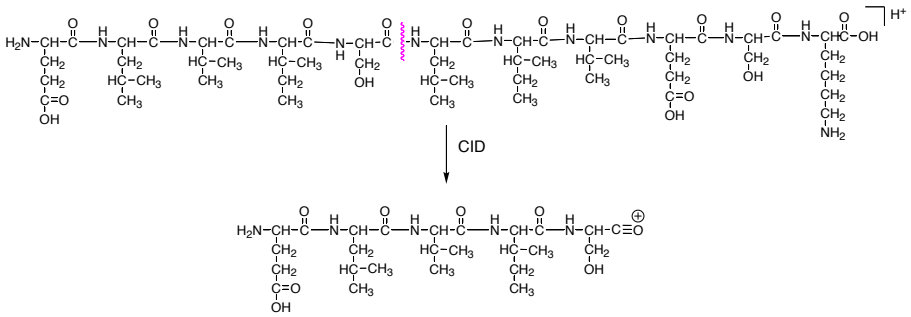
leucine (L) vs isoleucine (I): difficult to distinguish, must look at fragmentation of the sidechain

lysine (K, m/z=128.09) vs glutamine (Q, m/z = 128.06)



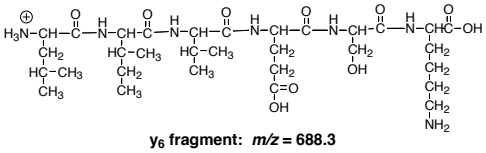
gly (G) + gly (G) = 114.04 = asn (N)= 114.04
ala (A) + gly (G) = 128.06 = gln (Q) = 128.06 = lys (K)= 128.09
gly (G) + val (V) = 156.09 = arg (R) = 156.10
ala (A) + asp (D) = glu (E) + gly (G) = 186.06 = trp (W) = 186.08
ser (S) + val (V) = 186.1 = trp (W) = 186.08

80



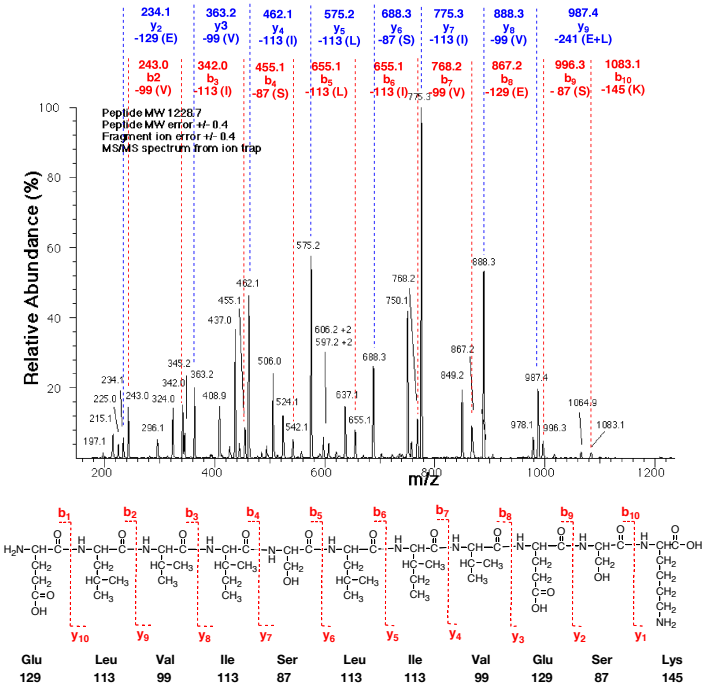
b_5 fragment: *m/z* = 542.1

-or-

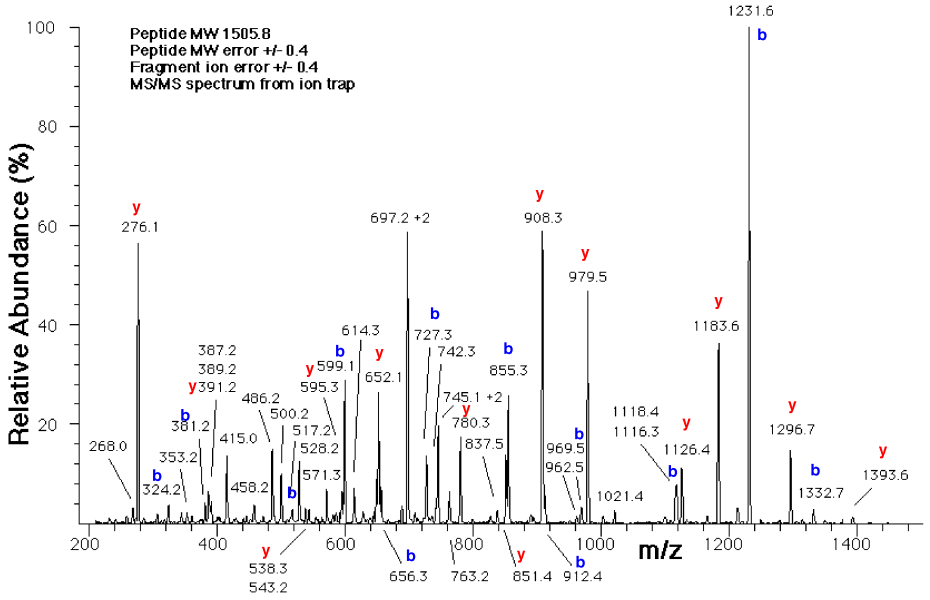


y_6 fragment: *m/z* = 688.3

81



82



y-ions: 1393.6, 1296.7, 1183.6, 1126.4, 979.5, 908.3, 851.4, 780.3, 652.1, 595.3, 538.3, 391.2, 276.1
b-ions: 1332.7, 1231.6, 1116.3, 969.5, 912.4, 855.3, 727.3, 656.3, 599.1, 528.2, 381.2, 324.2

Central Dogma

DNA ⟶ mRNA ⟶ protein ⟶ post-translational modifications

genome transcriptome proteome

THE STANDARD GENETIC CODE

UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys
UUC	Phe	UCC	Ser	UAC	Tyr	UGC	Cys
UUA	Leu	UCA	Ser	UAA	Stop	UGA	Stop
UUG	Leu	UCG	Ser	UAG	Stop	UGG	Trp
CUU	Leu	CCU	Pro	CAU	His	CGU	Arg
CUC	Leu	CCC	Pro	CAC	His	CGC	Arg
CUA	Leu	CCA	Pro	CAA	Gln	CGA	Arg
CUG	Leu	CCG	Pro	CAG	Gln	CGG	Arg
AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser
AUC	Ile	ACC	Thr	AAC	Asn	AGC	Ser
AUA	Ile	ACA	Thr	AAA	Lys	AGA	Arg
AUG	Met	ACG	Thr	AAG	Lys	AGG	Arg
GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly
GUC	Val	GCC	Ala	GAC	Asp	GGC	Gly
GUA	Val	GCA	Ala	GAA	Glu	GGA	Gly
GUG	Val	GCG	Ala	GAG	Glu	GGG	Gly

AUG is part of the initiation signal, as well as being the codon for internal methionine.

84