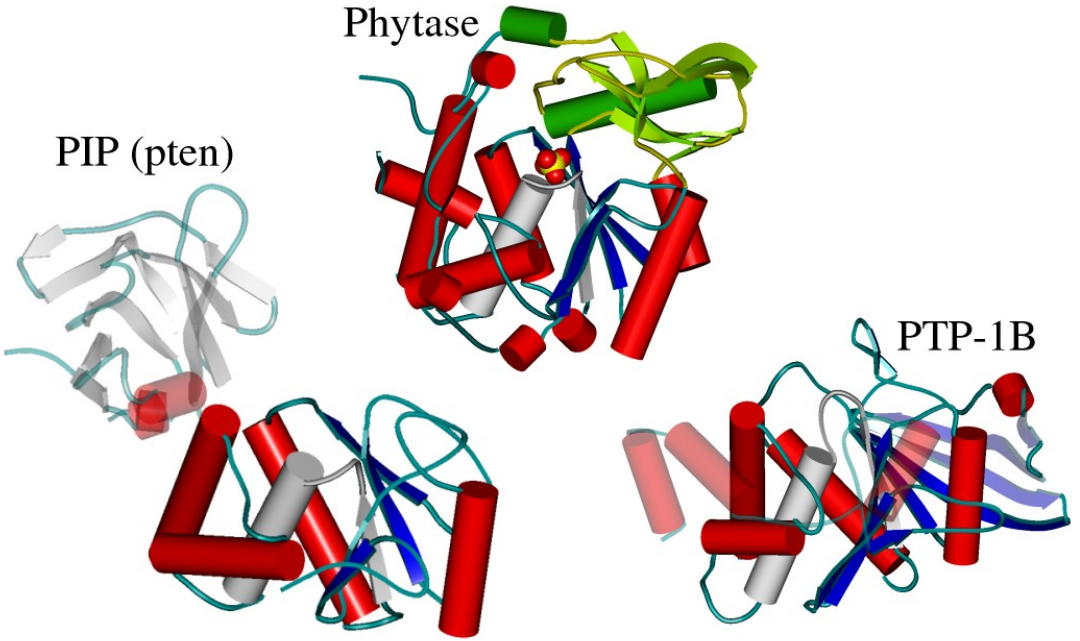
University of Lethbridge

Chapter 7: Covalent Structure of Proteins

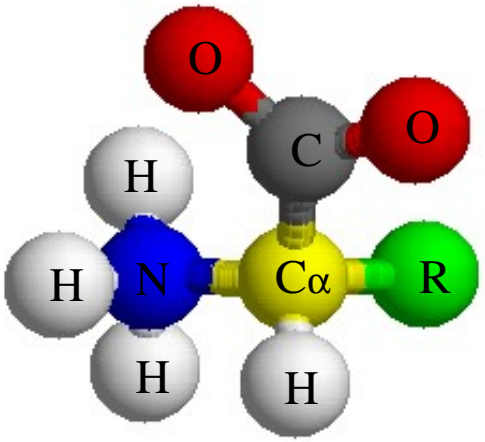
Voet & Voet: Pages 163-175

Direct sequencing of polypeptides



University of Lethbridge

One Letter Code



Amino Acid	Residue	3-letter Code	1-letter Code	R-group pK_a	Frequency (%)
Nonpolar					50.1
Glycine	Glycyl	Gly	G		7.1
Alanine	Alanyl	Ala	A		8.3
Proline	Prolyl	Pro	P		4.7
Valine	Valyl	Val	V		6.9
Leucine	Leucyl	Leu	L		9.7
Isoleucine	Isoleucyl	Ile	I		6.0
Tryptophan	Tryptophanyl	Trp	W		1.1
Phenylalanine	Phenylalanyl	Phe	F		3.9
Methionine	Methionyl	Met	M		2.4
Polar uncharged					24.0
Serine	Seryl	Ser	S		6.5
Threonine	Threonyl	Thr	T		5.3
Cysteine	Cystyl	Cys	C	8.4	1.4
Asparagine	Asparagyl	Asn	N		4.0
Glutamine	Glutaminyl	Gln	Q		3.9
Tyrosine	Tyrosyl	Tyr	Y	10.5	2.9
Polar charged					25.9
Histidine	Histidyl	His	H	6.0	2.3
Lysine	Lysyl	Lys	K	10.5	5.9
Arginine	Arginyl	Arg	R	12.5	5.5
Aspartate	Aspartyl	Asp	D	3.9	5.4
Glutamate	Glutamyl	Glu	E	4.1	6.8

University of Lethbridge

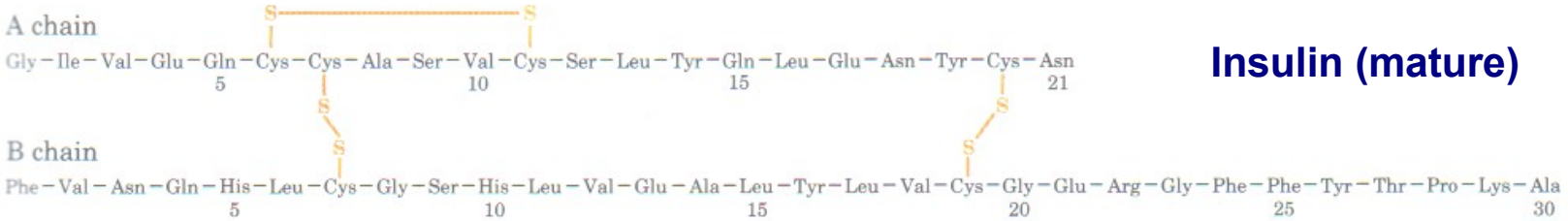
Primary Structure Determination

Primary Structure is central to the formulation of modern concepts of Biochemistry

(1) ***Structure Determination*** – primary structure (sequence) is a prerequisite for structural studies/understanding molecular mechanisms of action

(2) ***Evolutionary Relationships*** – sequence comparisons of related proteins in/between organisms shed light upon protein function and relationships between organisms

(3) ***Clinical Applications*** – many inherited diseases are caused by mutations that lead to amino acid changes in proteins; recognition of this fact has led to the development of therapies in many cases



Insulin (mature)

University of Lethbridge

Primary Sequencing Methods

Three general methods for sequencing proteins:

(1) Molecular biology methods

Convert gene to protein sequences using Universal Genetic Code – indirect determination of sequence; very cheap, rapid and easy.

Problems: Requires DNA sequence. No information regarding post-transcriptional or post-translational modifications, disulfide bridges and/or quaternary structure.

(2) Chemical methods

Original method – experimental determination of sequence; Labour, cost and time intensive

(3) Biophysical methods

Newer method – experimental determination of sequence using mass spectroscopy; Cost intensive; relatively fast

Benefits of Chemical/Biophysical methods:

Experimentally determines all types of modification, disulfide bridges and quaternary structure.

University of Lethbridge

Primary Sequencing Methods

Three general methods for sequencing proteins:

(1) Molecular biology methods

Convert gene to protein sequences using Universal Genetic Code – indirect determination of sequence; very cheap, rapid and easy.

Problems: Requires DNA sequence. No information regarding post-transcriptional or post-translational modifications, disulfide bridges and/or quaternary structure.

(2) Chemical methods

Original method – experimental determination of sequence; Labour, cost and time intensive

(3) Biophysical methods

Newer method – experimental determination of sequence using mass spectroscopy; Cost intensive; relatively fast

Benefits of Chemical/Biophysical methods:

Experimentally determines all types of modification, disulfide bridges and quaternary structure.

Post-translational modifications of Proteins

Table 1: Posttranslational protein modifications at the side chains.[a]

Residue	Reaction	Example
Asp	phosphorylation	protein tyrosine phosphatases; response regulators in two-component systems
	isomerization to isoAsp	
Glu	methylation	chemotaxis receptor proteins
	carboxylation	Gla residues in blood coagulation
	polyglycination	tubulin
	polyglutamylation	tubulin
Ser	phosphorylation	protein serine kinases and phosphatases
	O-glycosylation	notch O-glycosylation
	phosphopantetheinylation	fatty acid synthase
	autocleavages	pyruvamidyl enzyme formation
Thr	phosphorylation	protein threonine kinases/phosphatases
	O-glycosylation	
Tyr	phosphorylation	tyrosine kinases/phosphatases
	sulfation	CCR5 receptor maturation
	ortho-nitration	inflammatory responses
	TOPA quinone	amine oxidase maturation
His	phosphorylation	sensor protein kinases in two-component regulatory systems
	aminocarboxypropylation	diphthamide formation
	N-methylation	methyl CoM reductase
Lys	N-methylation	histone methylation
	N-acylation by acetyl, biotinyl, lipoyl, ubiquityl groups	histone acetylation; swinging-arm prosthetic groups; ubiquitin; SUMO (small ubiquitin-like modifier) tagging of proteins
	C-hydroxylation	collagen maturation
Cys	S-hydroxylation (S-OH)	sulfenate intermediates
	disulfide bond formation	protein in oxidizing environments
	phosphorylation	PTPases
	S-acylation	Ras
	S-prenylation	Ras
	protein splicing	intein excisions
Met	oxidation to sulfoxide	Met sulfoxide reductase
Arg	N-methylation	histones
	N-ADP-ribosylation	$G_{s\alpha}$
Asn	N-glycosylation	N-glycoproteins
	N-ADP-ribosylation	eEF-2
	protein splicing	intein excision step
Gln	transglutamination	protein cross-linking
Trp	C-mannosylation	plasma-membrane proteins
Pro	C-hydroxylation	collagen; HIF-1α
Gly	C-hydroxylation	C-terminal amide formation

[a] No modifications of Leu, Ile, Val, Ala, Phe side chains are known. A more extensive list can be found in reference [3].

Protein Posttranslational Modifications: The Chemistry of Proteome Diversifications

Christopher T. Walsh, Sylvie Garneau-Tsodikova, and Gregory J. Gatto, Jr.*

DOI: 10.1002/anie.200501023

University of Lethbridge

Chemical/Biophysical Sequencing

Four general steps for Chemical/Biophysical Sequencing:

(1) Preparing protein for sequencing

Detection of quaternary structure and disulfide bridges

Followed by separation and purification of each polypeptide chain(s)

(2) Fragmenting the polypeptide chain(s)

Cleave each chain(s) into oligopeptide fragments (< 50 residues)

Followed by separation and purification of all oligopeptide fragments from each chain

(3) Sequencing the oligopeptide fragments

Sequence each fragment using the 'Edman reagent' (Chemical) or
Mass spectroscopy (Biophysical)

(4) Organizing the sequence into a complete primary sequence

Arrange oligopeptide fragment sequences into a complete sequence

(Sometimes requires additional experiments to confirm identity of disulfide bonds)

University of Lethbridge

Step 1: Preparation for Sequencing

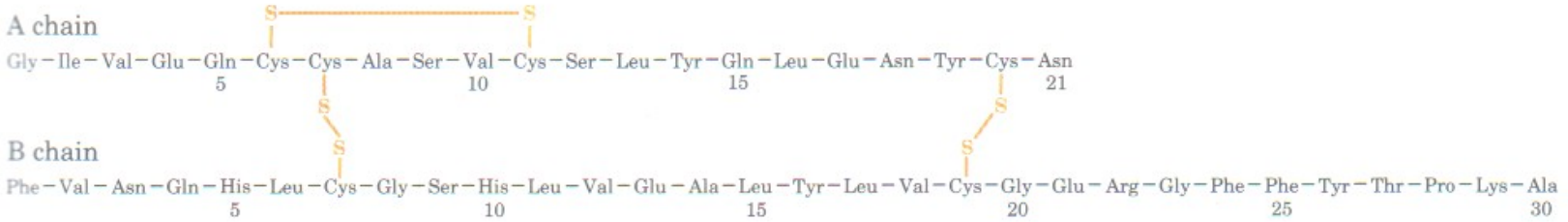
Is their quaternary structure or disulfide bridges?

- SDS-PAGE (denaturing electrophoresis) followed by Size Exclusion Chromatography can be utilized to reveal quaternary structure and subunit composition (but not always)

- Denaturing 'IEF' (isoelectric Focusing) can also help as it separates polypeptides with different pl's

- 'End Group Analysis' is another method for detecting the number of polypeptide chains

  Sequencing several N- and C-terminal residues or “end groups” may establish the number of chemically distinct polypeptides



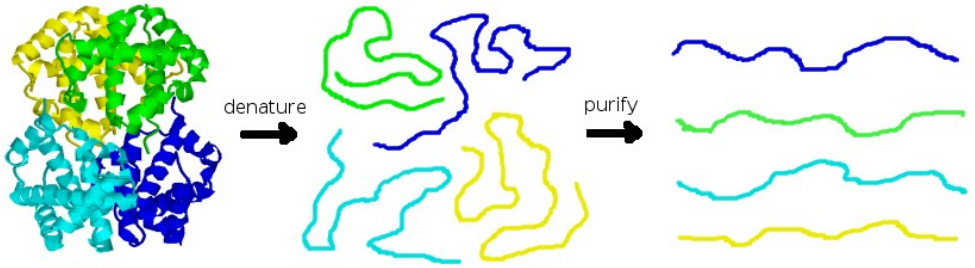
Insulin is a heterodimer with multiple disulfide bridges that are important for both structure and function

University of Lethbridge

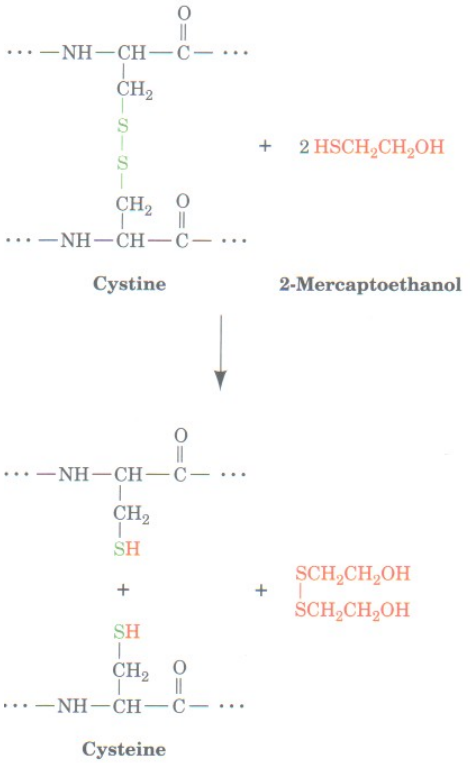
Step 1: Preparation for Sequencing

Separation and Purification of Polypeptides

Denaturation & Reduction of Disulfides are required to separate and purify individual polypeptide chains



Denaturation disrupts attractive non-covalent forces breaking apart most quaternary structures. Facilitates purification of polypeptide chains by chromatographic (ion exchange or reverse phase) methods.



Reduction of disulfide bridges breaks inter- and Intra- subunit bonds

Common denaturants include urea, guanidine HCl, detergents & heat

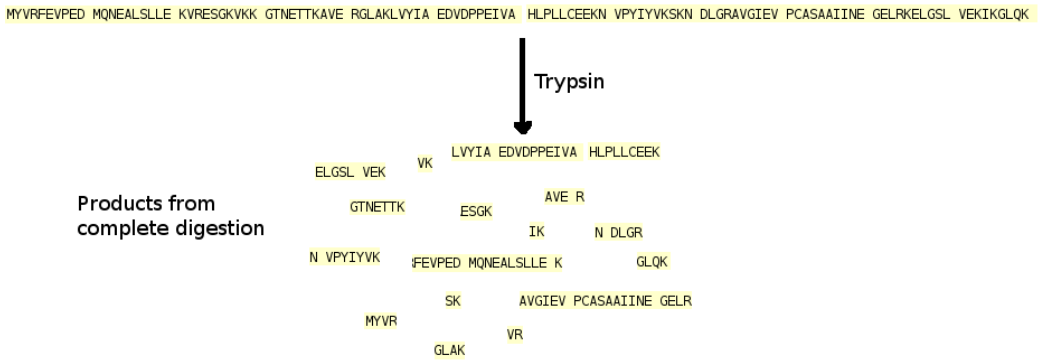
Common disulfide bridge reducing agents include 2-mercaptoethanol (BME) or dithiothreitol (DTT)

University of Lethbridge

Step 2: Fragmentation

Once each subunit is separated and purified it must be broken into fragments small enough to sequence (< 50 residues)

(1) Specific polypeptide cleavage of subunit into smaller pieces



Certain proteases and chemical reagents will only cleave proteins at specific sites (sequences)

eg. **Trypsin only hydrolyzes the peptide bond following Lys/Arg**

Chymotrypsin only hydrolyzes the peptide bond following Phe/Tyr

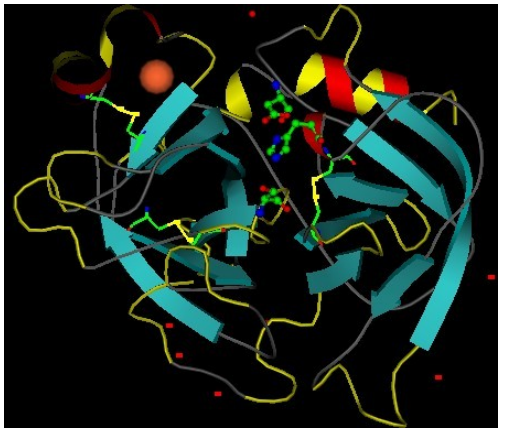
CNBr only hydrolyzes the peptide bond following Met

(2) Separation and purification of each polypeptide fragment

Reverse phase chromatography is generally sufficient to separate and purify all polypeptide fragments

University of Lethbridge

Step 2: Fragmentation (Proteases)



Trypsin: Protease of first choice due to its strict specificity for Lys/Arg residues

- Proteins with too few or too many Lys/Arg residues require the use of other enzymes

Note: Not all Lys/Arg sites are cleaved due to their accessibility. Further, the most accessible sites are cut first!

Chymotrypsin: Protease specific for Phe/Tyr residues

Proteases with broader substrate specificities may cleave proteins into fragments that are too small to effectively sequence

- Adjusting reaction times and conditions limits the number of sites cleaved (ie. limited proteolysis). Can optimize fragmentation conditions and maximize number of fragments of useful size.

University of Lethbridge

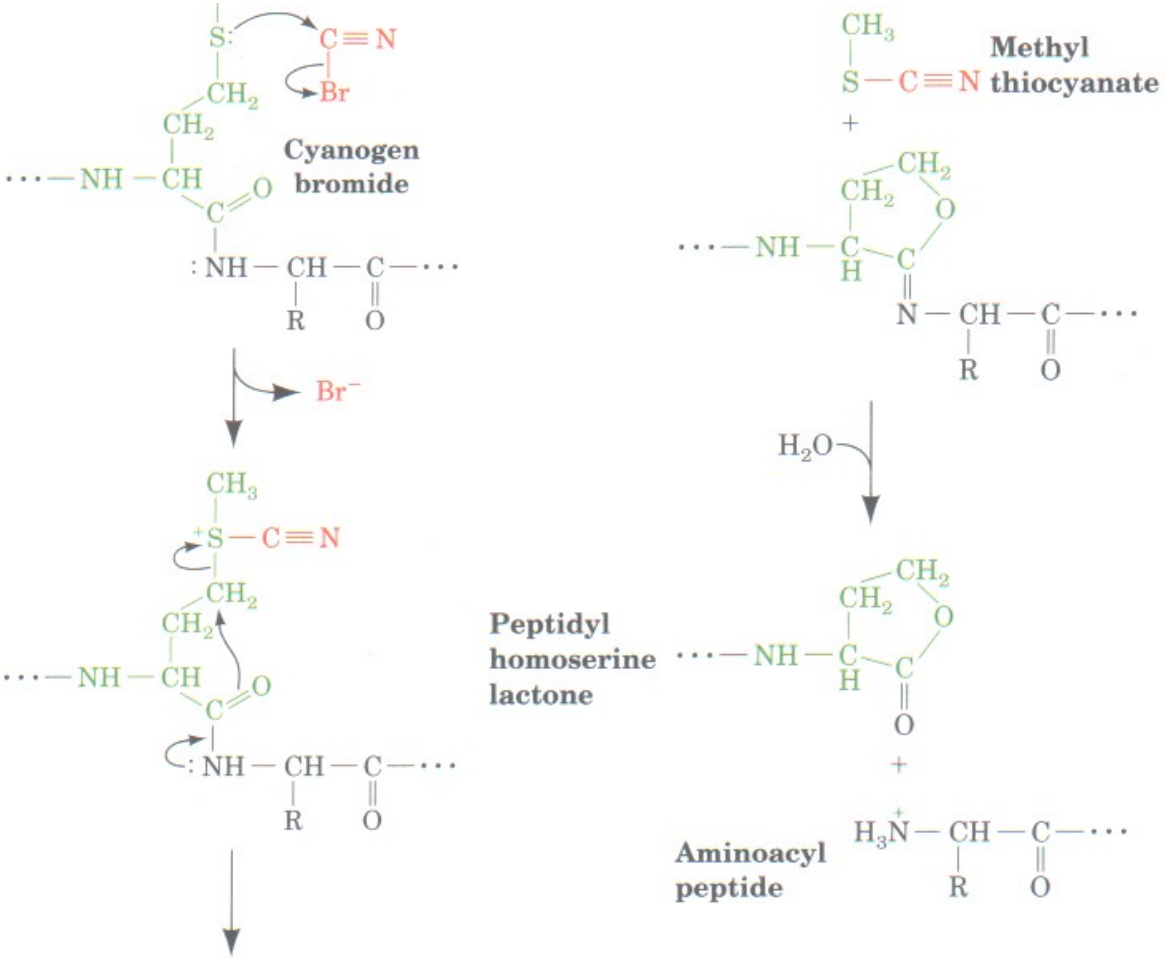
Step 2: Fragmentation (CNBr)

CNBr (cyanogen bromide)

Chemical reagent that cleaves following Met residues

Converts Met residues to homoserine lactone during hydrolysis

The only highly specific chemical reagent for peptide bond cleavage



University of Lethbridge

Step 2: Fragmentation

Separation and purification of peptide fragments achieved using several standard chromatographic techniques

(1) Ion Exchange chromatography separates based upon charge

(2) Reverse Phase chromatography separates based upon hydrophobicity

Typically, both chromatographic techniques must be used to separate and purify all peptide fragments

Summary to this point:

Preparation stage – identified, separated and purified all subunits

Fragmentation stage – each purified subunit is fragmented into peptides and all peptides are separated and purified

University of Lethbridge

Step 3: Chemical Sequencing

Each purified fragment is subjected to repeated cycles of Edman sequencing (Next Slide)

Result is a sequence for each of the isolated peptide fragments

Modern Edman sequencing has been automated: Involves immobilizing the peptide fragment on a PVDF (polyvinyldifluoride) membrane

Attaching peptide fragment to a solid support (membrane) has huge advantages

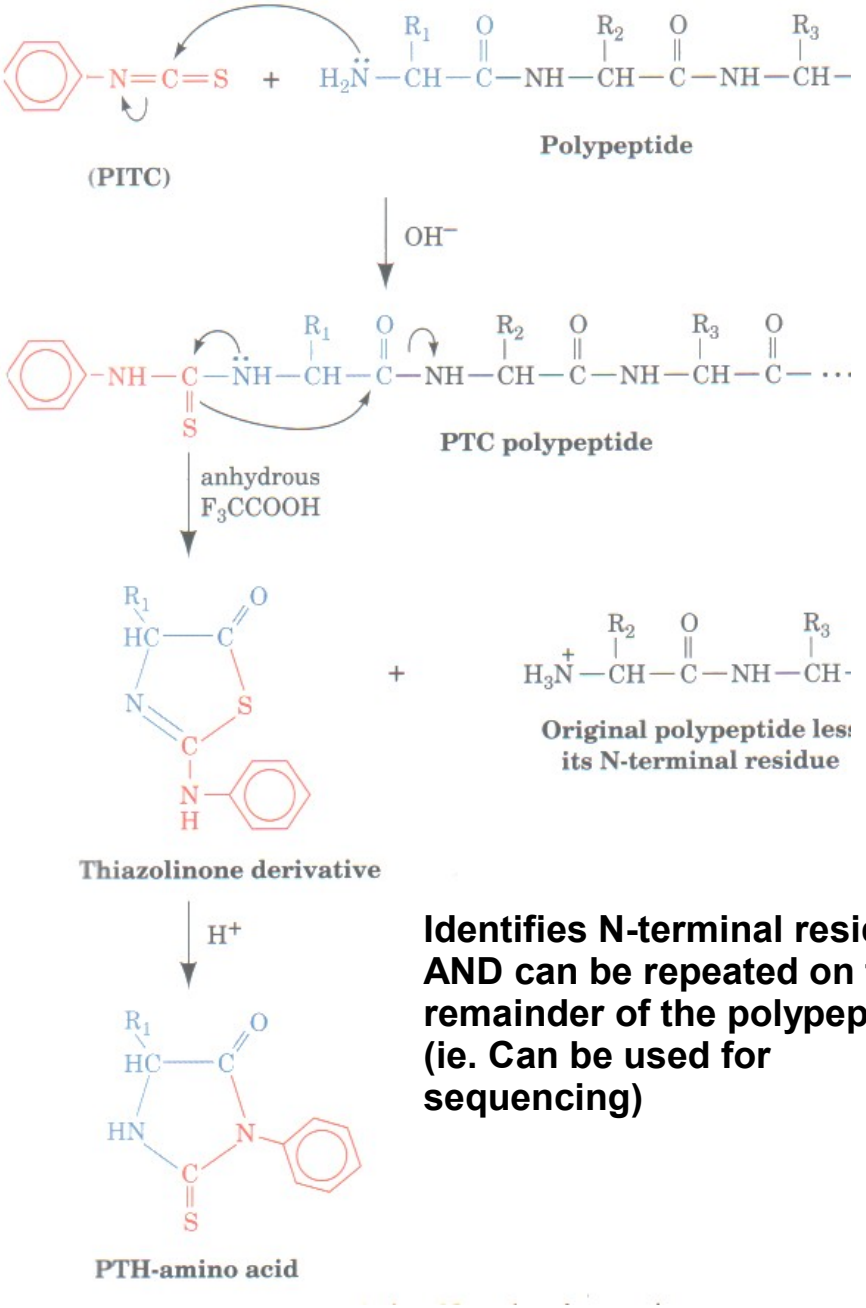
A) simplifies separation of PTH-amino acid and peptide fragment

B) simplifies identification of individual amino acids

C) greatly improved efficiency (yield per cycle)

Step 3 – Sequencing

Chemical Method:
Edman degradation
(two steps)



(1) Mild alkaline conditions

Phenylisothiocyanate (PITC) reacts with α-amino to form phenylthiocarbamyl (PTC)

(2) Anhydrous acid:

PTC further reacts with N-terminal residue leading to its removal from the polypeptide

Phenylthiohydantoin (PTH) amino acid product is identified by chromatographic methods

Comparison of unknown PTH-amino acid retention time to a set of known PTH-amino acid retention times

Identifies N-terminal residue AND can be repeated on the remainder of the polypeptide (ie. Can be used for sequencing)

University of Lethbridge

Aside – C terminal sequencing

Enzymes (exopeptidase) that remove the carboxy terminal residue are exploited

Problems:

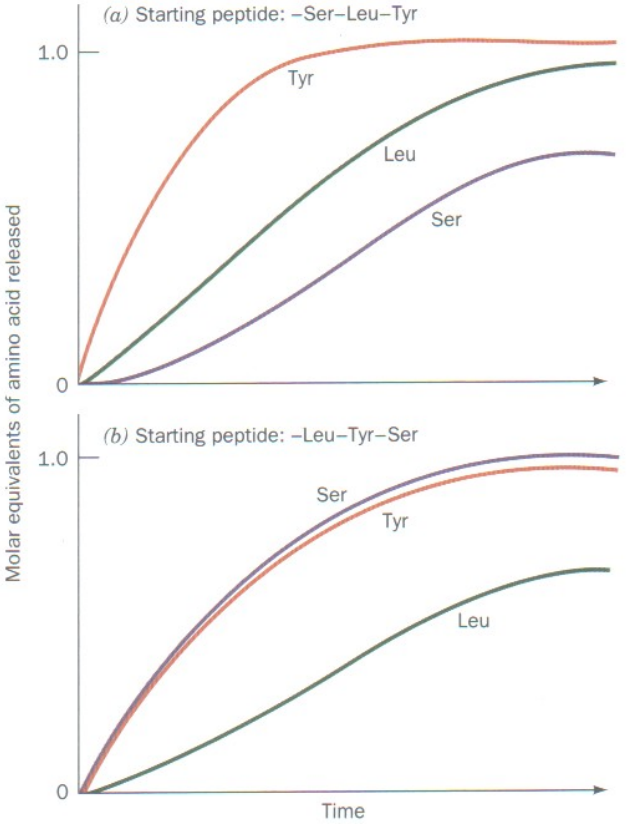
(1) enzymes remove one residue after another

Complicates detection but yields some C-terminal sequence information

(2) Different reaction rates for different residues

Can prevent accurate identification of certain residues

Normally must confirm C-terminus identification using several, different carboxypeptidase



Enzyme	Source	Specificity[a]
Carboxypeptidase A	Bovine pancreas	$R_n \neq$ Arg, Lys, Pro; $R_{n-1} \neq$ Pro
Carboxypeptidase B	Bovine pancreas	$R_n =$ Arg, Lys; $R_{n-1} \neq$ Pro
Carboxypeptidase C	Citrus leaves	All free C-terminal residues; pH optimum = 3.5
Carboxypeptidase Y	Yeast	All free C-terminal residues, but slowly with $R_n =$ Gly
Leucine aminopeptidase	Porcine kidney	$R_1 \neq$ Pro
Aminopeptidase M	Porcine kidney	All free N-terminal residues

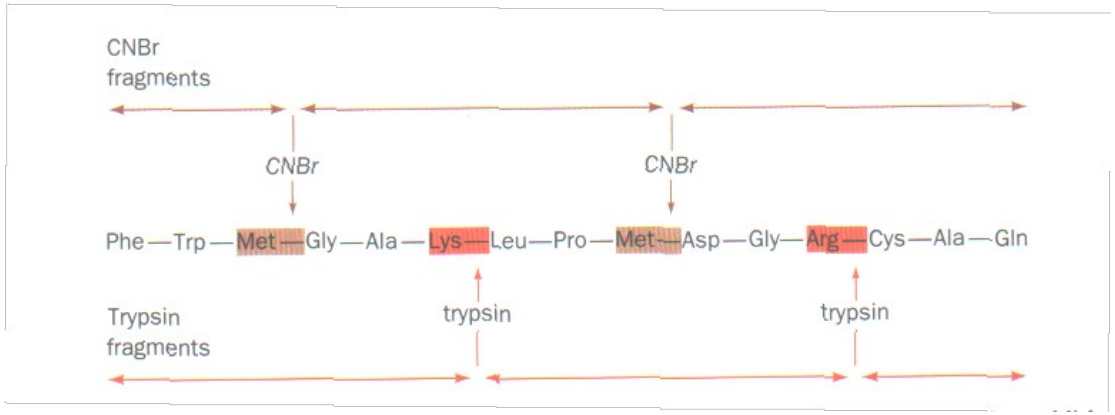
University of Lethbridge

Step 4: Organizing the Sequence

How is the sequence information for each fragment reassembled into a protein sequence? How do we know the sequence order?

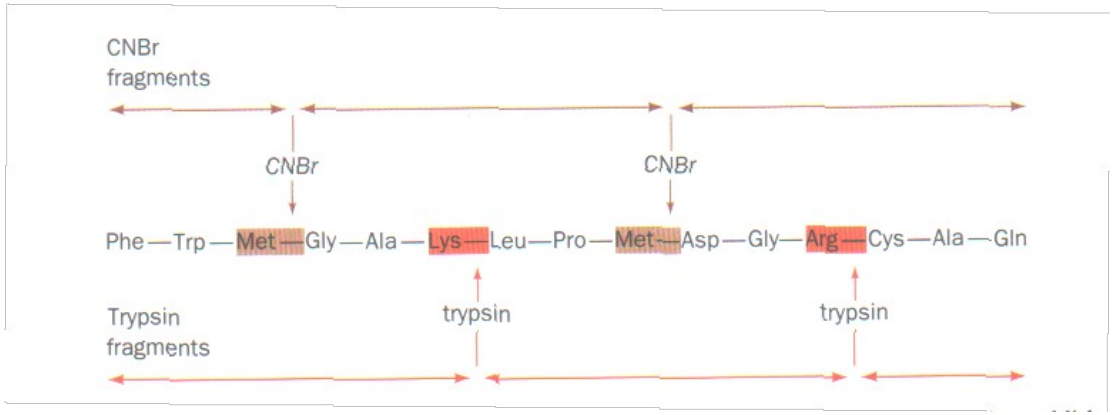
Requires a second independent set of peptide fragments that overlap the first

Sequencing must be performed a second time using a different fragmentation step that generates a different sets of protein cleavage products !!



University of Lethbridge

Step 4: Organizing the sequence (simple example)



CNBr fragments

Phe-Trp-Met

Gly-Ala-Lys-Leu-Pro- Met

Asp-Gly-Arg-Cys-Ala-Gln

Trypsin fragments

Phe-Trp-Met-Gly-Ala-Lys

Leu-Pro- Met-Asp-Gly-Arg

Cys-Ala-Gln

Combine information from two sets of fragments into complete sequence

Phe-Trp-Met-Gly-Ala-Lys-Leu-Pro-Met-Asp-Gly-Arg-Cys-Ala-Gln

University of Lethbridge

Step 4: Locating Disulfide Bonds

Typically requires an additional experiment(s) !!

Step 1:Prepare protein sample without disrupting disulfide bonds

Step 2:Repeat fragmentation reaction

Step 3: Separate and purify peptide fragments

Step 4: Analysis

Analysis includes comparing the separated and purified fragments from the fragmentation

A) Protein sample without disulfide bonds

B) Protein sample with disulfide bonds intact

Differences are due to the presence of disulfide bond(s).

Note: Works in most but not all cases

University of Lethbridge

Biophysical Sequencing (mass spectrometry)

Mass Spectrometry accurately (0.01%) measures the mass-to-charge ratio (m/z) of ions in the gas phase

Can be used for:

1) Mass determination

2) Sequencing polypeptide fragments (< 25 residues)

Requires method for producing gas phase ions of polypeptide fragments

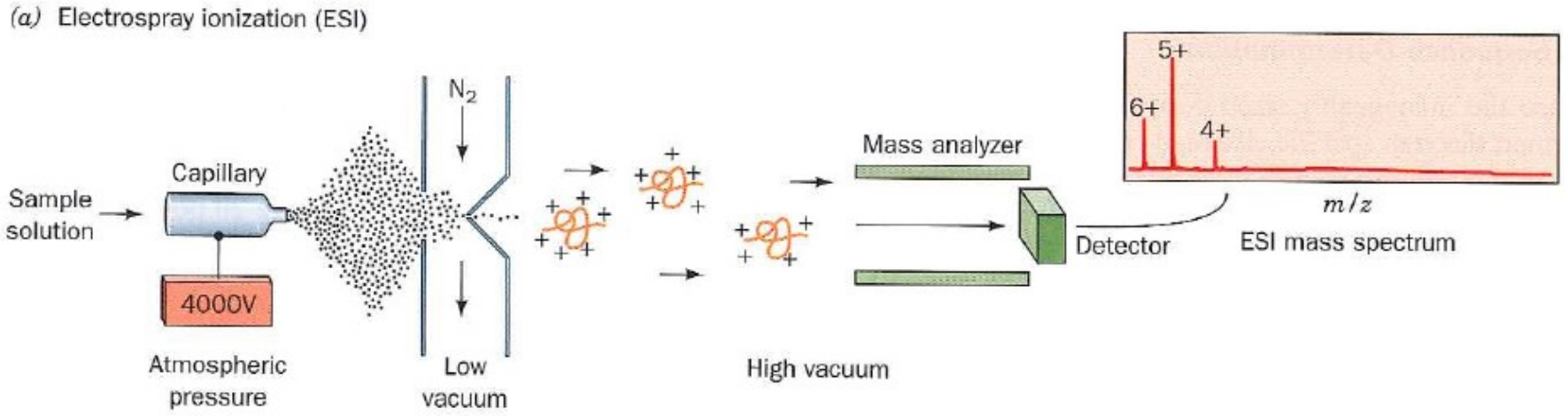
Two methods for generating gas phase ions are applicable to macromolecules

A) Electrospray Ionization (ESI)

B) Matrix-assisted laser desorption / ionization (MALDI)

University of Lethbridge

Electrospray Ionization (mass spectrometry)



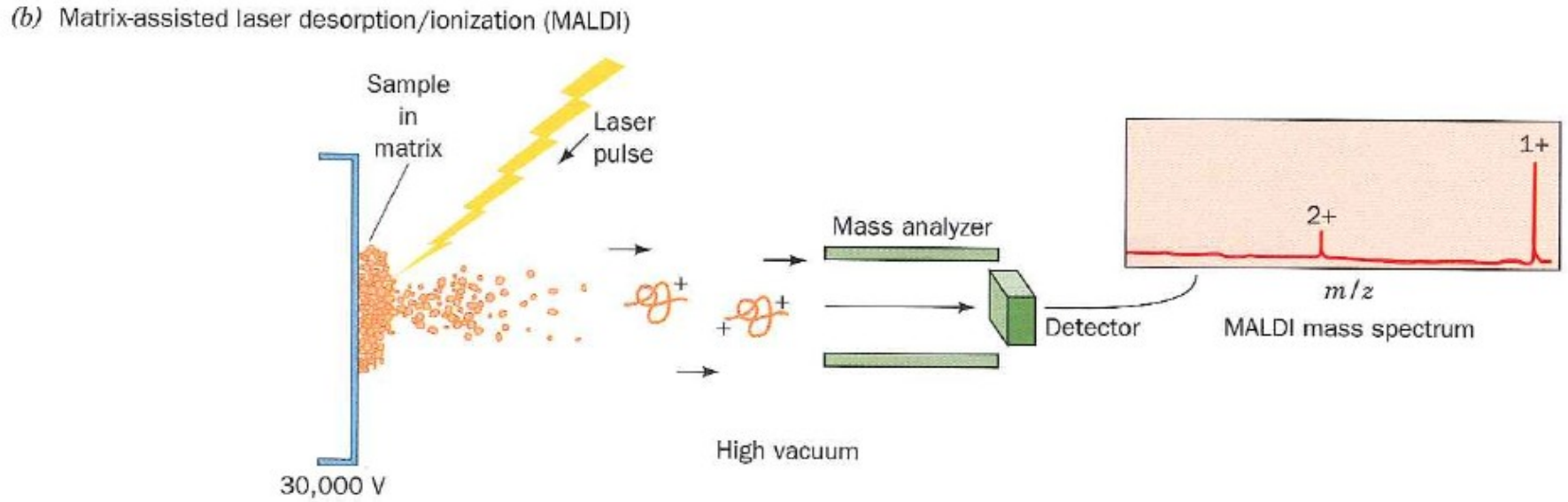
Liquid sample is sprayed through narrow capillary maintained at high voltage (adds average charge of +0.5 - +2.0 / kilodalton)

Tiny droplets of liquid rapidly evaporate as they pass into a strong, constant electromagnetic field which accelerates the gas phase ions towards a detector

Ions are separated based upon time (Time of Flight (TOF)) or curvature (as they pass through a separate magnetic field)

University of Lethbridge

MALDI (mass spectrometry)



Sample is embedded in a dry, gel-like matrix and ejected from the matrix upon exposure to an intense laser pulse (adds only a few charges)

Ejected gas phase ions pass into a strong, constant electromagnetic field which accelerates the gas phase ions towards a detector

Ions are separated based upon time (Time of Flight (TOF)) or curvature (as they pass through a separate magnetic field)

University of Lethbridge

Obtaining Mass Information (mass spectrometry)

How are the charges (and masses) of each peak determined?

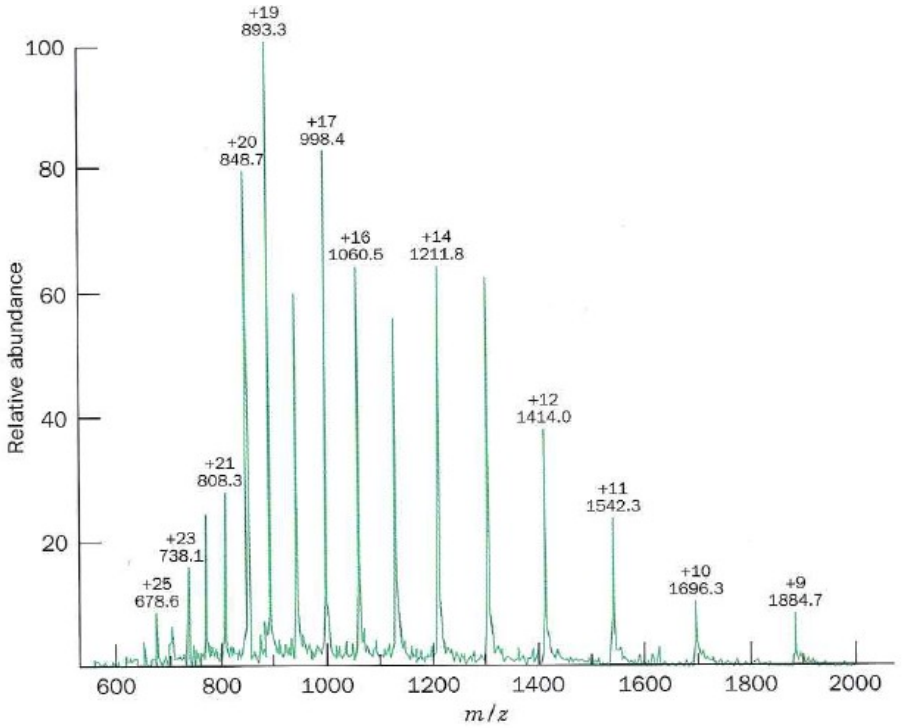
Any two adjacent peaks in the spectrum differ in charge by +1.

For adjacent peaks the following relations hold:

$$P_1 = (M + z_1) / z_1$$

$$P_2 = (M + z_1 - 1) / (z_1 - 1)$$

where P_1 and P_2 are m/z values for each peak (from the spectrum). M is unknown mass and z_1 is charge associated with P_1.



ESI-MS spectrum of apomyoglobin

University of Lethbridge

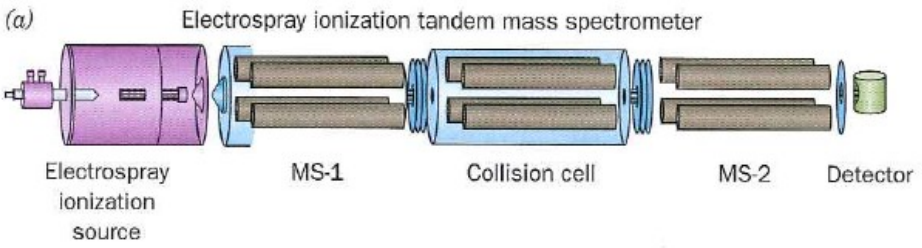
Tandem MS

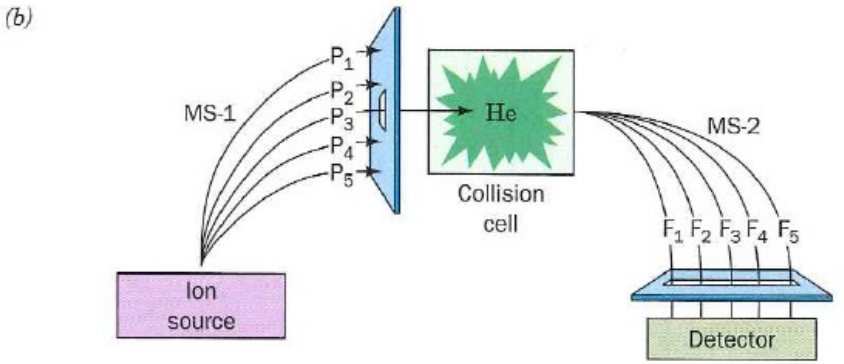
Sequence information is generally obtained from Tandem MS

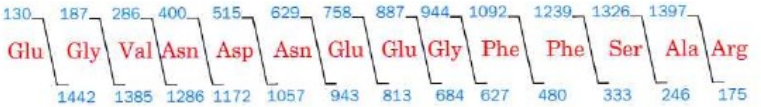
MS-1: A small peptide (<25 residues) in the gas phase is forced to 'collide' with He atoms and breaks into fragments

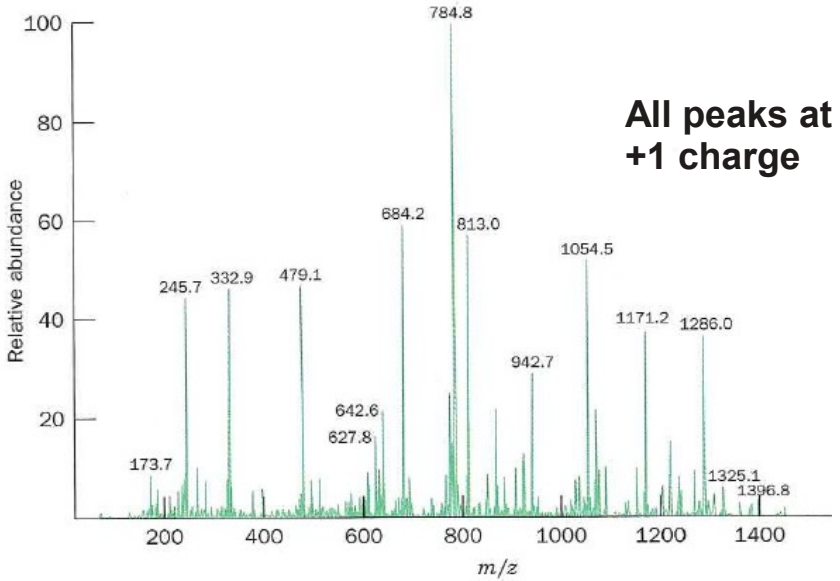
MS-2: As each fragment is already in the gas phase they accelerated towards a detector where the m/z values for each fragment are determined

Note: Tandem MS experiments can obtain sequence information from a mixture of different peptide fragments as MS-1 can be used to select for a single peptide from the mixture









All peaks at +1 charge