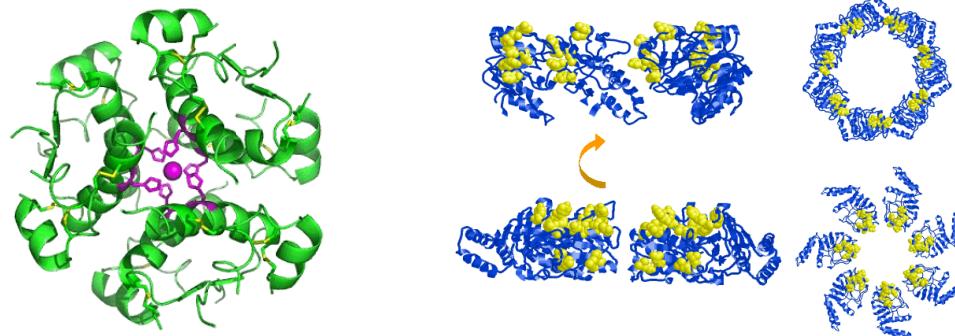


Proteins

- Proteins contain at least 40 residues; polypeptides smaller than that are simply called peptides. **Multisubunit** proteins contain several identical or non-identical chains called **subunits**.
- A few proteins contain two or more polypeptide chains linked covalently.
- Some proteins contain permanently associated chemical components in addition to amino acids; these are called **conjugated proteins**.
- The non-amino acid part of a conjugated protein is usually called its **prosthetic group**.
- Conjugated proteins are classified on the basis of the chemical nature of their prosthetic groups; for example, **lipoproteins** contain lipids, **glycoproteins** contain sugar groups, and **metalloproteins** contain a specific metal.



Properties of proteins

TABLE 3–2**Molecular Data on Some Proteins**

	Molecular weight	Number of residues	Number of polypeptide chains
Cytochrome c (human)	13,000	104	1
Ribonuclease A (bovine pancreas)	13,700	124	1
Lysozyme (chicken egg white)	13,930	129	1
Myoglobin (equine heart)	16,890	153	1
Chymotrypsin (bovine pancreas)	21,600	241	3
Chymotrypsinogen (bovine)	22,000	245	1
Hemoglobin (human)	64,500	574	4
Serum albumin (human)	68,500	609	1
Hexokinase (yeast)	102,000	972	2
RNA polymerase (<i>E. coli</i>)	450,000	4,158	5
Apolipoprotein B (human)	513,000	4,536	1
Glutamine synthetase (<i>E. coli</i>)	619,000	5,628	12
Titin (human)	2,993,000	26,926	1

Table 3-2*Lehninger Principles of Biochemistry, Fifth Edition*

© 2008 W.H. Freeman and Company

TABLE 3–4**Conjugated Proteins**

Class	Prosthetic group	Example
Lipoproteins	Lipids	β_1 -Lipoprotein of blood
Glycoproteins	Carbohydrates	Immunoglobulin G
Phosphoproteins	Phosphate groups	Casein of milk
Hemoproteins	Heme (iron porphyrin)	Hemoglobin
Flavoproteins	Flavin nucleotides	Succinate dehydrogenase
Metalloproteins	Iron	Ferritin
	Zinc	Alcohol dehydrogenase
	Calcium	Calmodulin
	Molybdenum	Dinitrogenase
	Copper	Plastocyanin

Table 3-4*Lehninger Principles of Biochemistry, Fifth Edition*

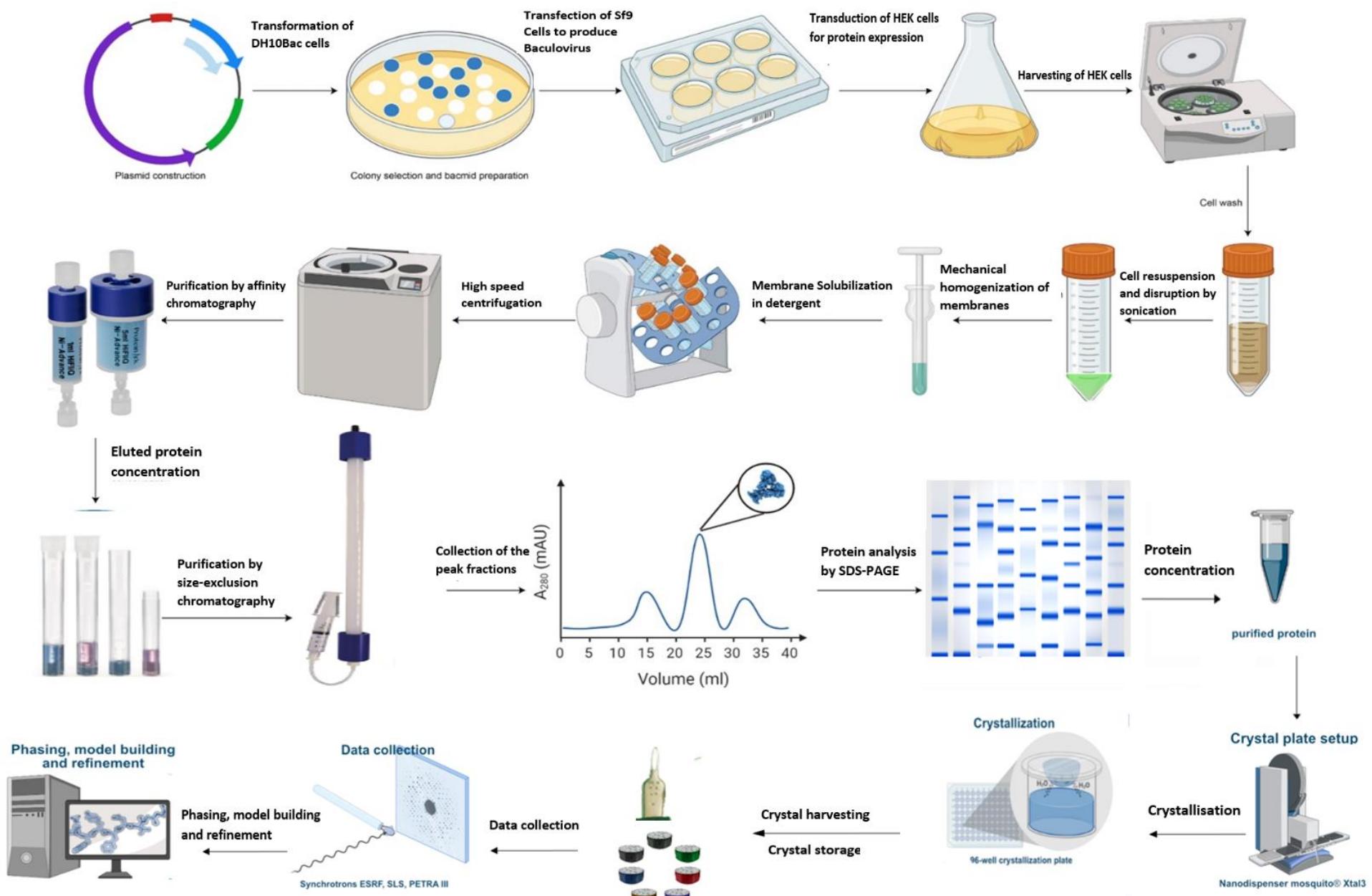
© 2008 W.H. Freeman and Company

Protein Purification and Analysis

Protein Sequencing

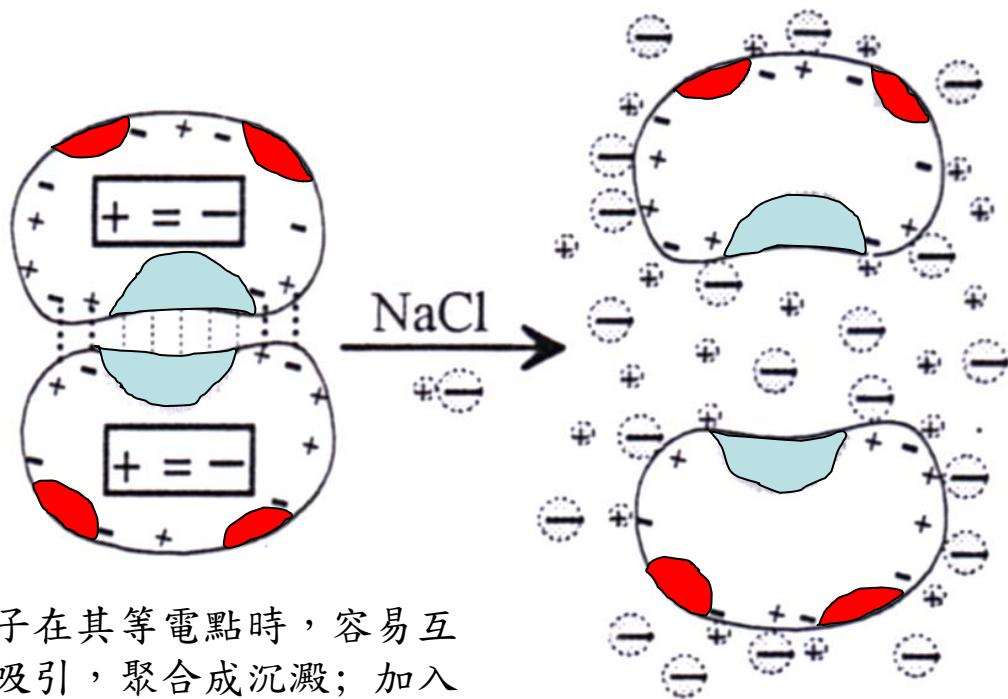
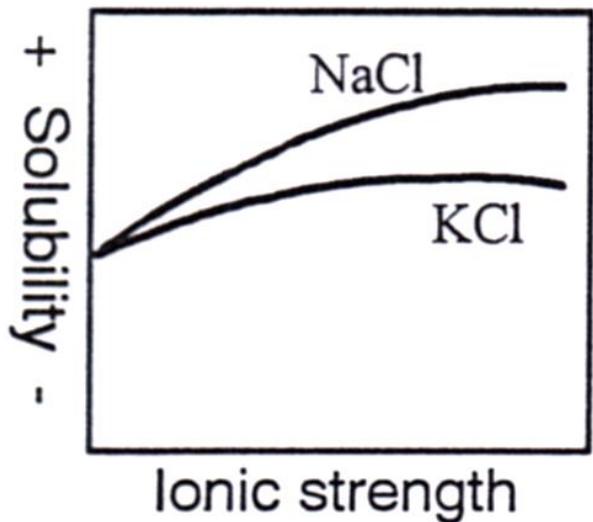
General steps in protein purification

1. The raw material
 2. Disruption:
 - Gentle: Cell lysis Erythrocytes
Enzyme digestion Lysozyme treatment
Hand homogenizer Liver tissue
Grinding Muscle
 - Vigorous: French press Bacteria, plant cells
Ultrasonication Cell suspensions
 3. Clarification: Centrifugation
 4. Precipitation:
 - Ammonium sulphate: $(\text{NH}_4)_2\text{SO}_4$
 - Organic solvent: acetone
 5. Column chromatography



Salting in: depends on surface charge distribution and polar interactions with the solvent

A. Salting-in:



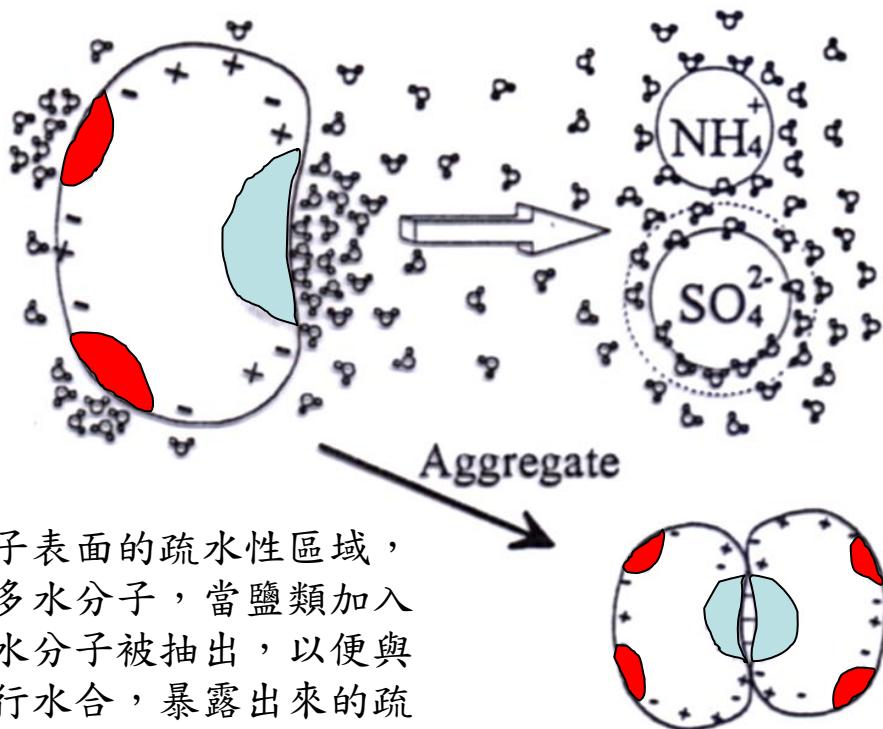
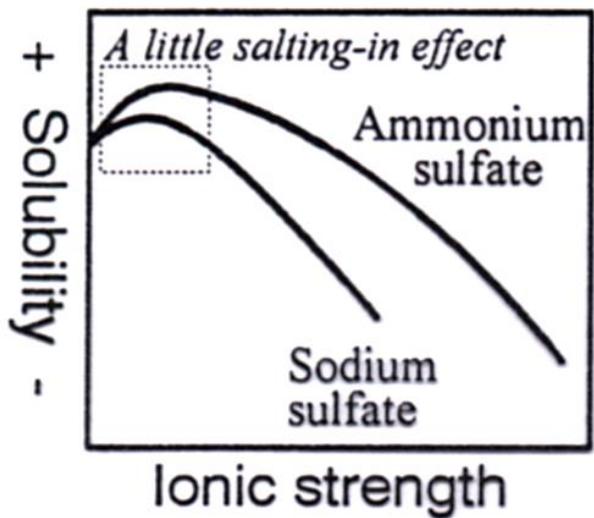
分子在其等電點時，容易互相吸引，聚合成沉澱；加入鹽離子會破壞這些吸引力，使分子散開，溶入水中。

=Hydrophilic area

=Hydrophobic area

Salting out: depends on the hydrophobicity of the protein

B. Salting-out:



蛋白質分子表面的疏水性區域，都聚集許多水分子，當鹽類加入時，這些水分子被抽出，以便與鹽離子進行水合，暴露出來的疏水性區域相互結合，形成沉澱。

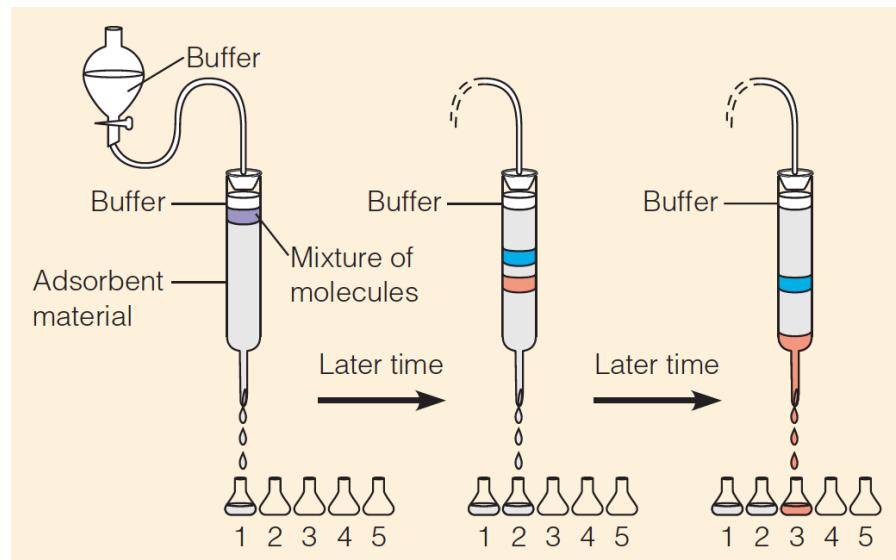
=Hydrophilic area

=Hydrophobic area

Protein Purification

Purification by chromatography:
interactions between the proteins in the cell lysate and the matrix within the column - stronger the interaction with the matrix, the later it will elute from the column.

Proteins are generally detected by UV absorbance at **280 nm** as they elute from the column.



Protein Purification

Chromatography takes advantage of the differences in a protein's physical properties to do the separation:

- Ionic charge (ion-exchange chromatography)
- Size (size-exclusion chromatography)
- Affinity to ligands (affinity chromatography)

Protein Purification

Ion-exchange chromatography separates molecules on the basis of their *electrical charge*.

The strength of interaction between a protein molecule and an ion exchange matrix depends on

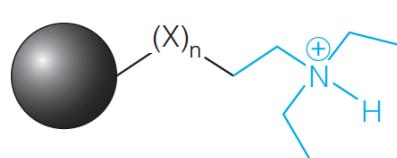
- the *charge density* on the protein
- the *ionic strength* of the mobile phase, which is always a buffered solution.

Protein Purification

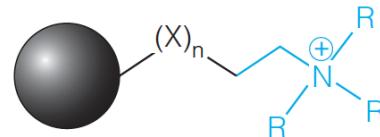
Ion-exchange chromatography

Two main types of ion exchange matrices:

Anion exchangers - carry a positive charge and bind to negatively charged proteins

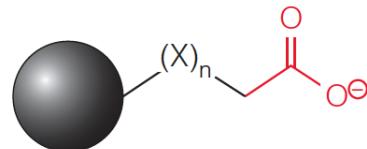


Weak anion exchanger (DEAE)

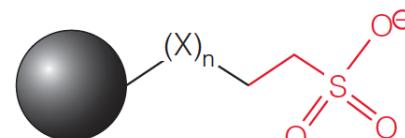


Strong anion exchanger ("Q")

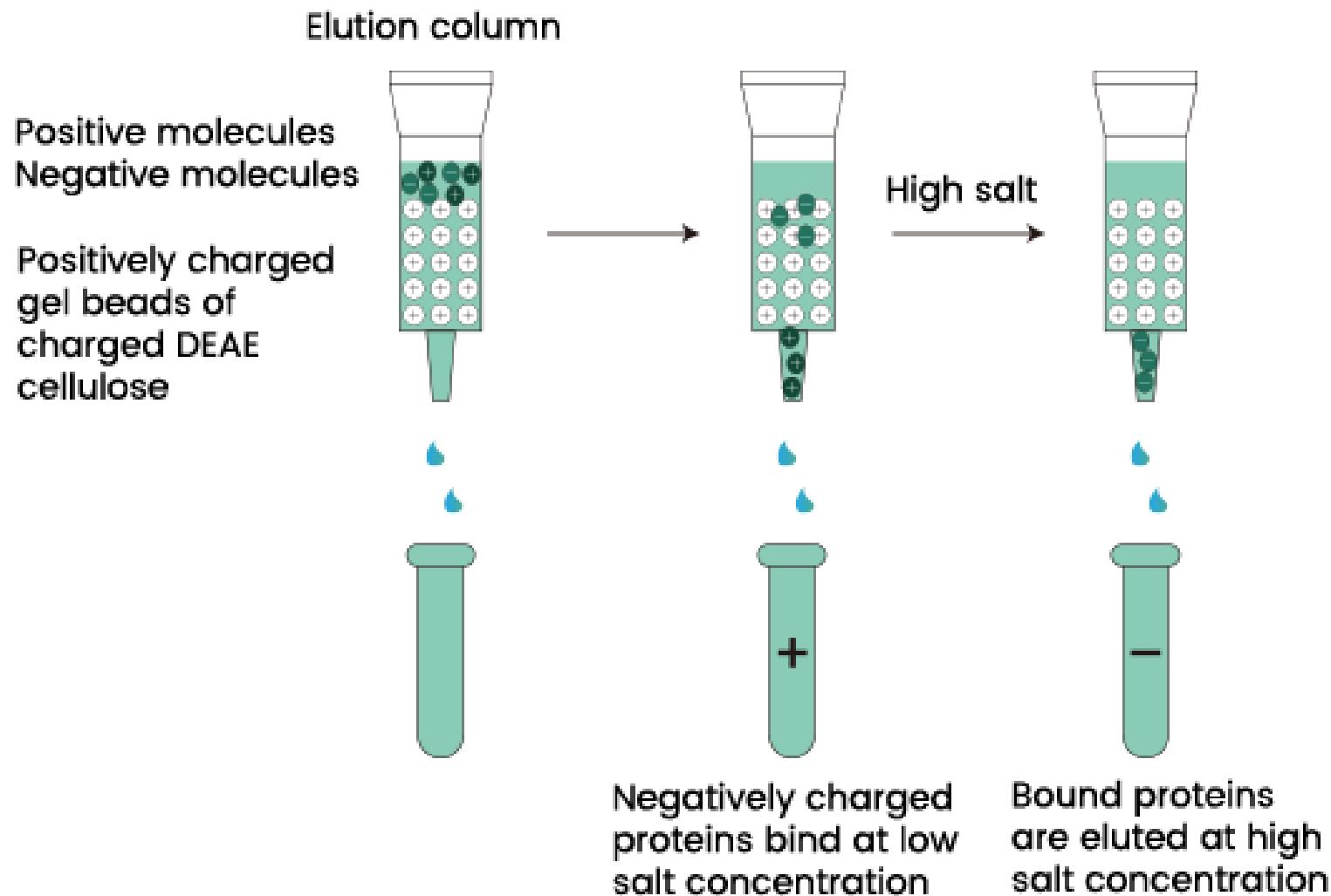
Cation exchangers - carry a negative charge and bind to positively charged proteins



Weak cation exchanger (CM)

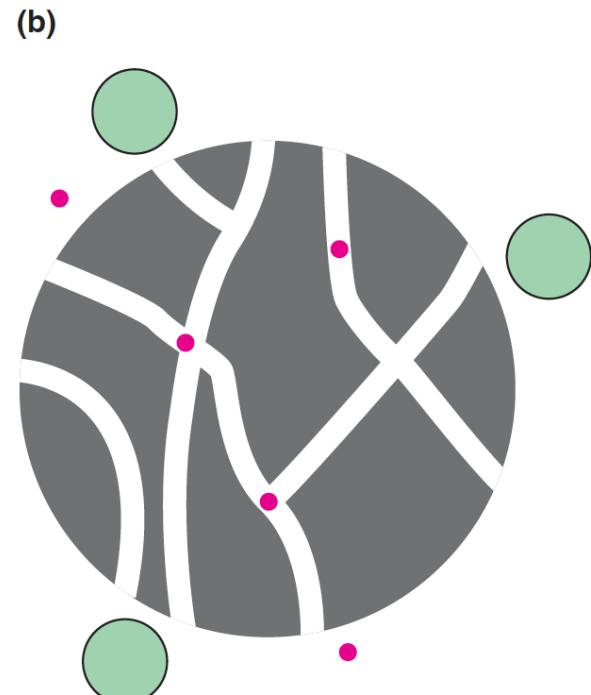
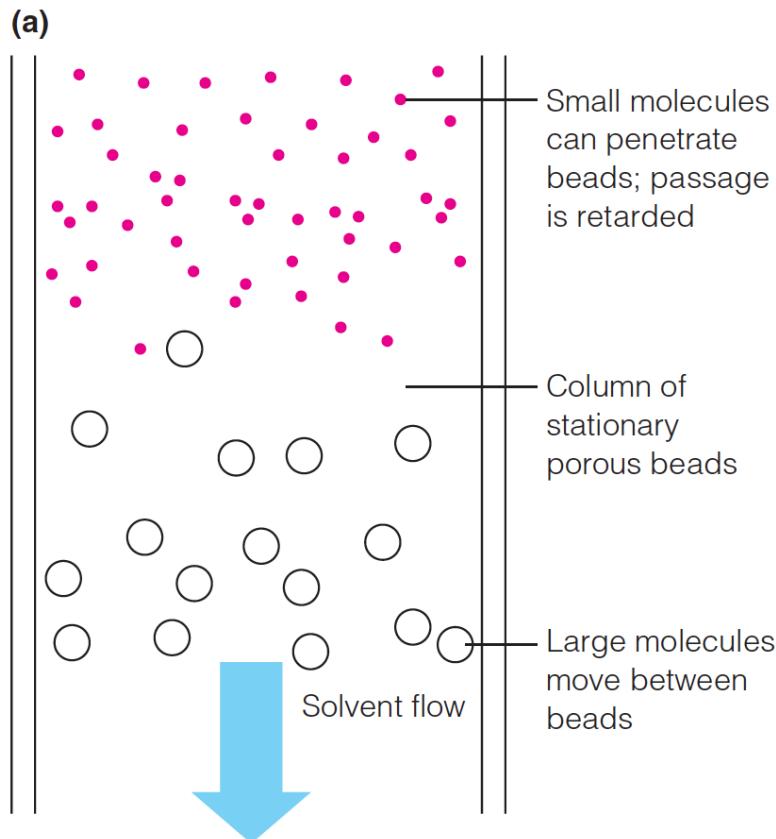


Strong cation exchanger ("S")



Protein Purification

Size-exclusion chromatography (aka gel-filtration)



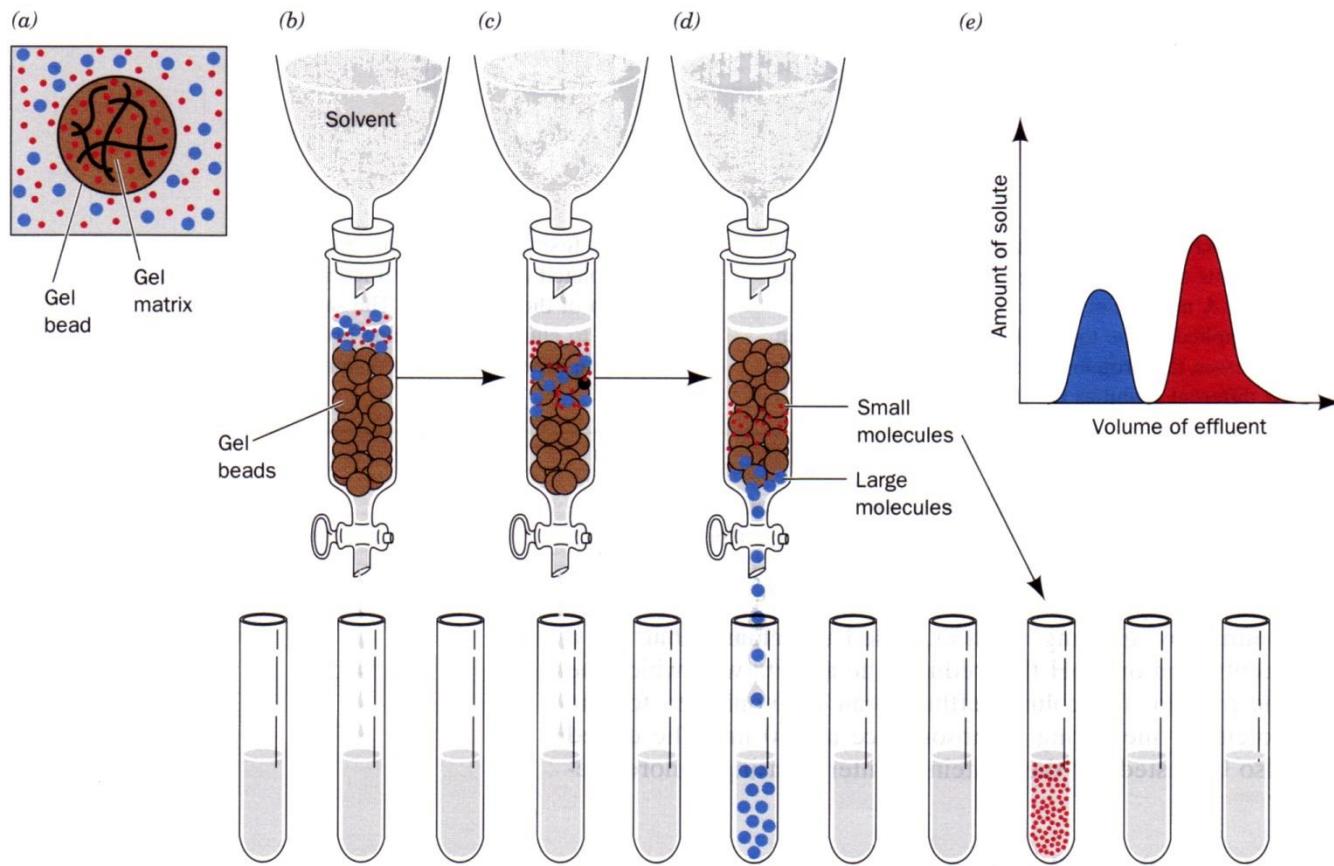
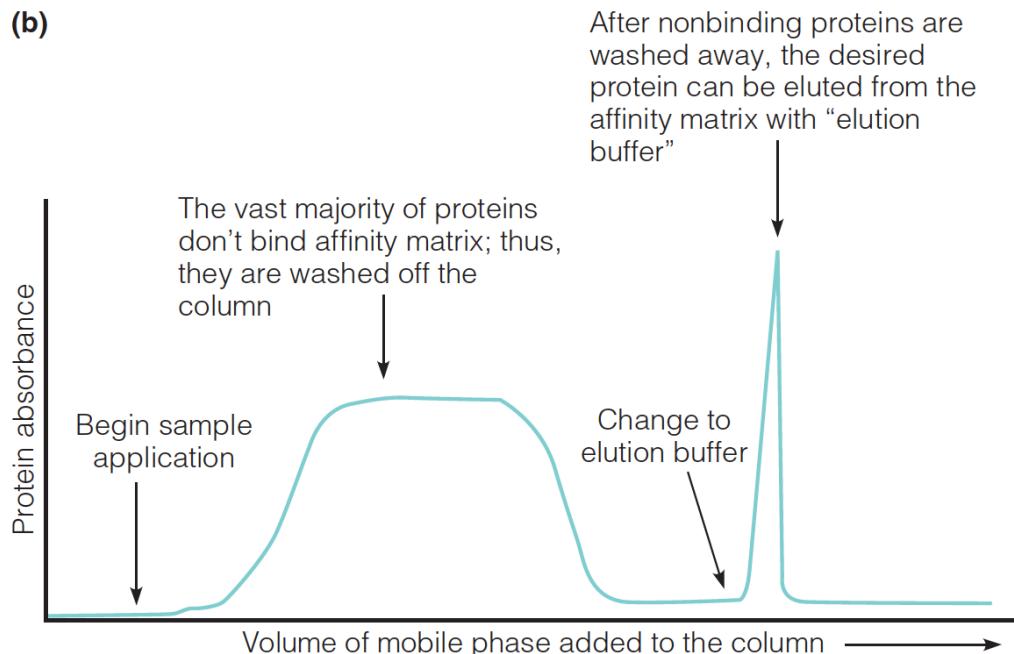
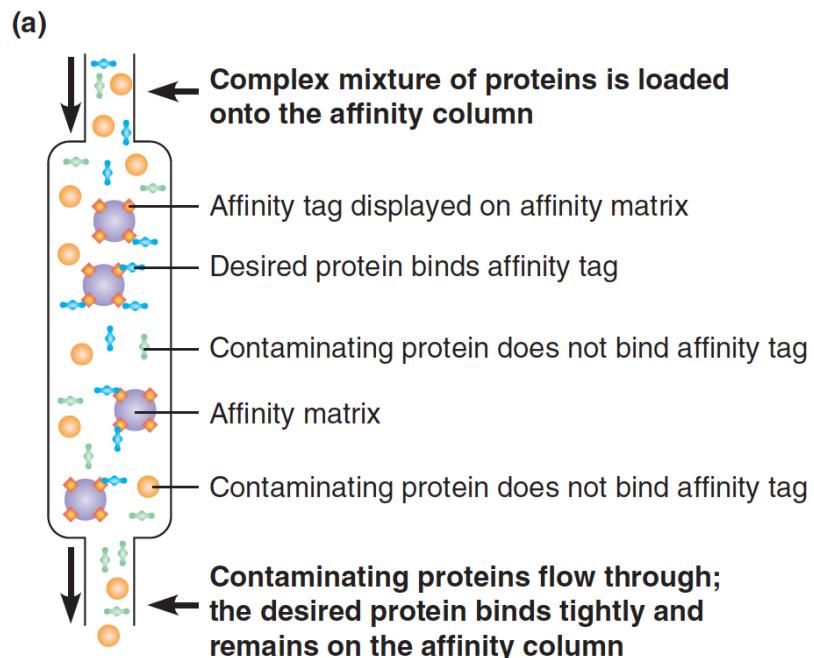


Figure 5-7 | Gel filtration chromatography. (a) A gel bead consists of a gel matrix (wavy solid lines) that encloses an internal solvent space. Small molecules (red dots) can freely enter the internal space of the gel bead. Large molecules (blue dots) cannot penetrate the gel pores. (b) The sample solution is applied to the top of the column (the gel beads are represented as brown spheres). (c) The small molecules can penetrate the gel and consequently migrate

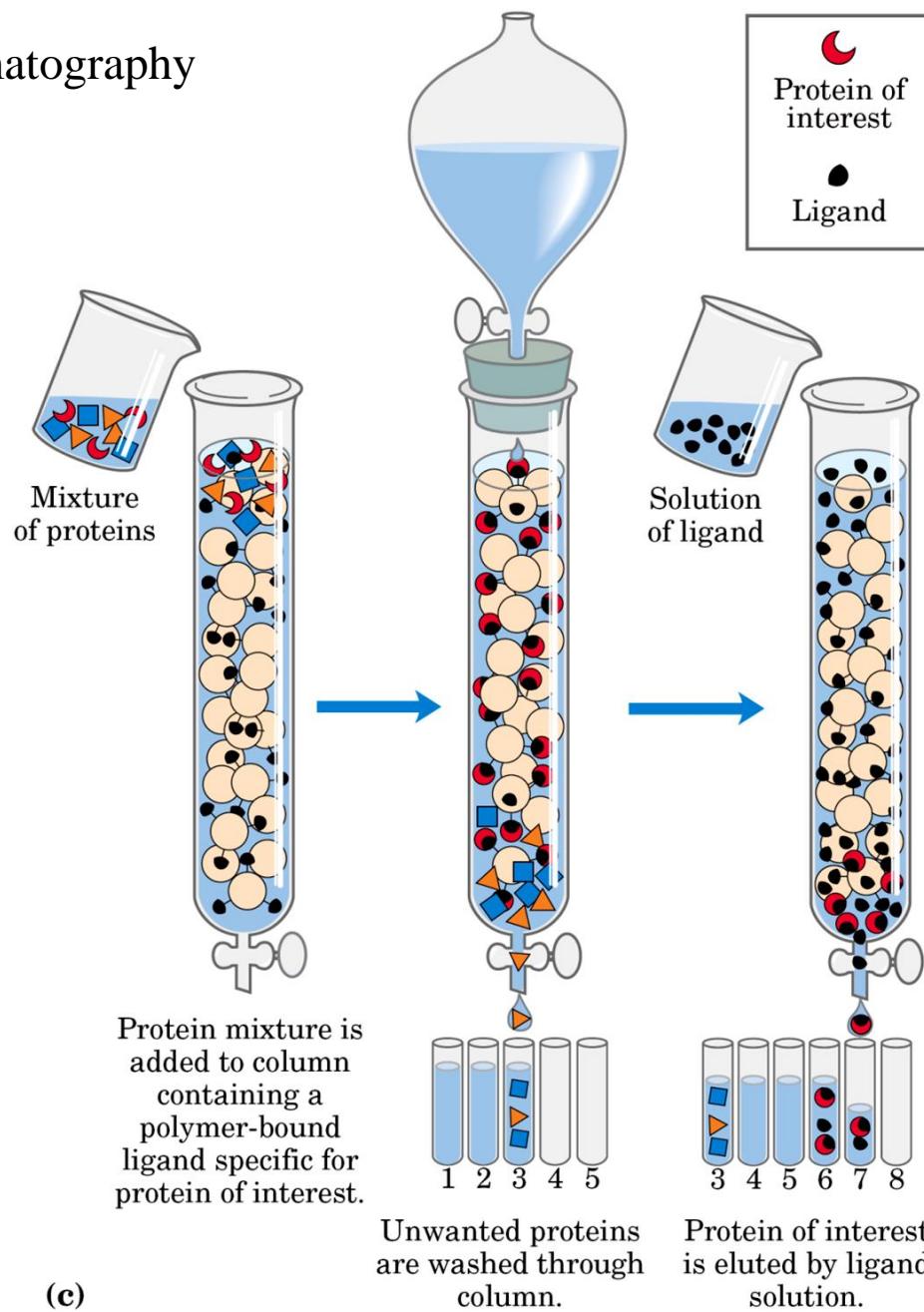
through the column more slowly than the large molecules that are excluded from the gel. (d) The large molecules elute first and are collected as fractions. Small molecules require a larger volume of solvent to elute. (e) The elution diagram, or chromatogram, indicating the complete separation of the two components.  See the **Animated Figures**.

Protein Purification

Affinity chromatography

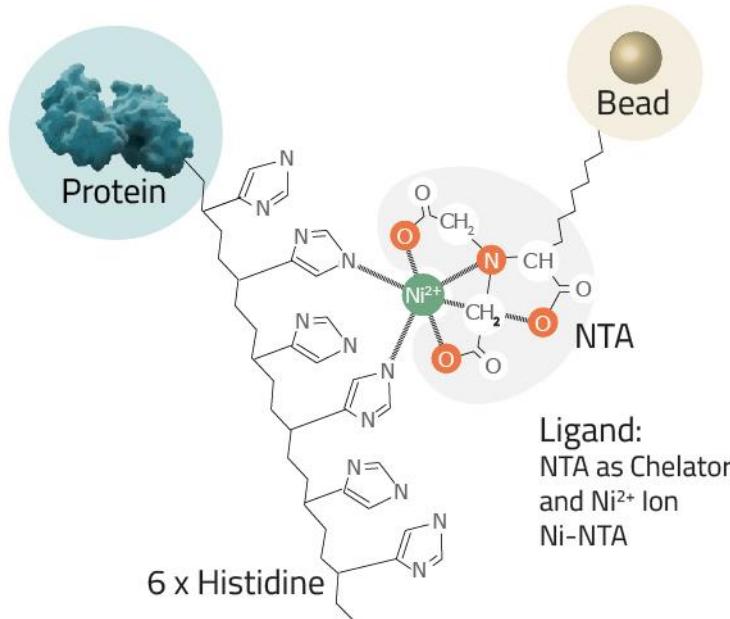
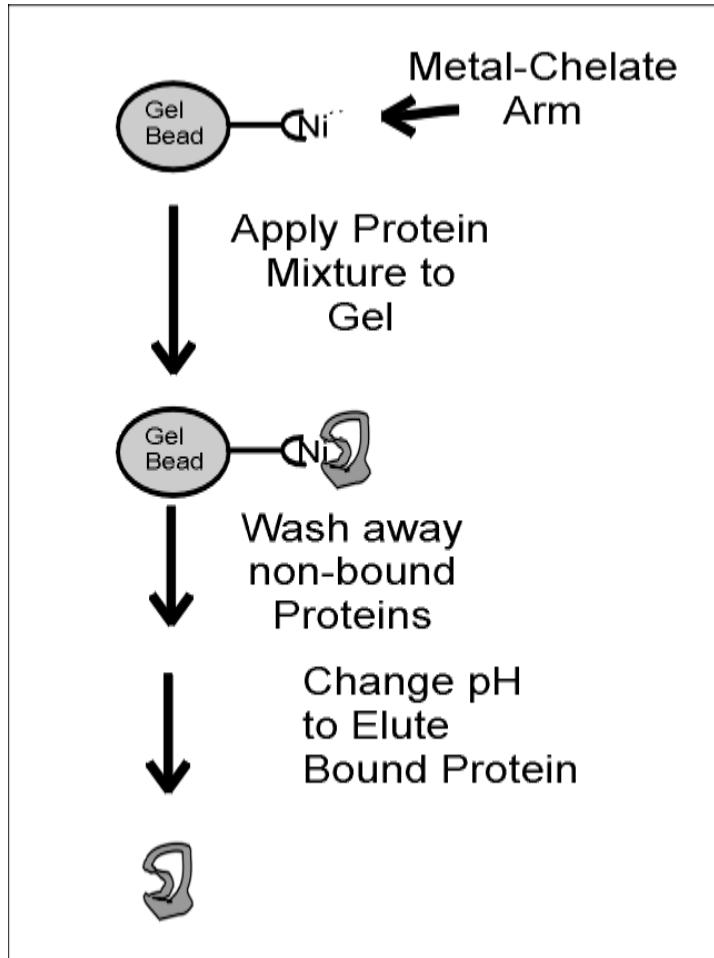


(c) affinity chromatography

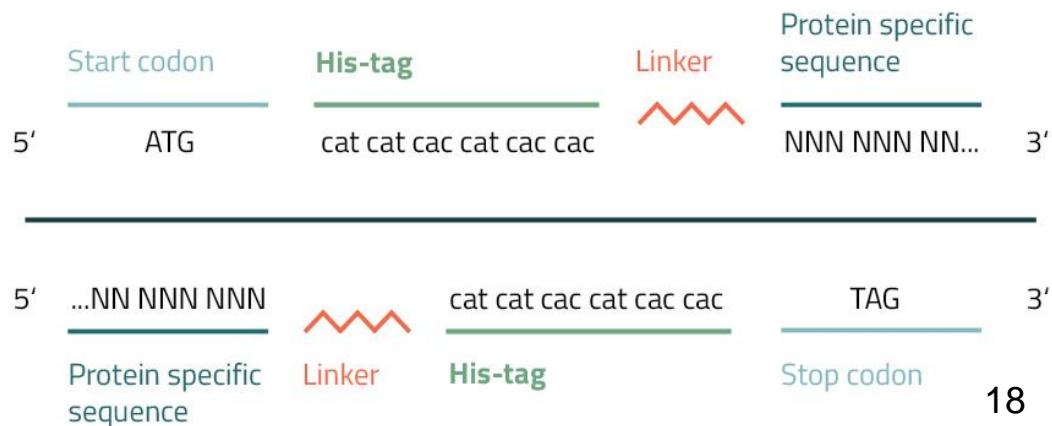


(c)

Metal Chelate Affinity Chromatography

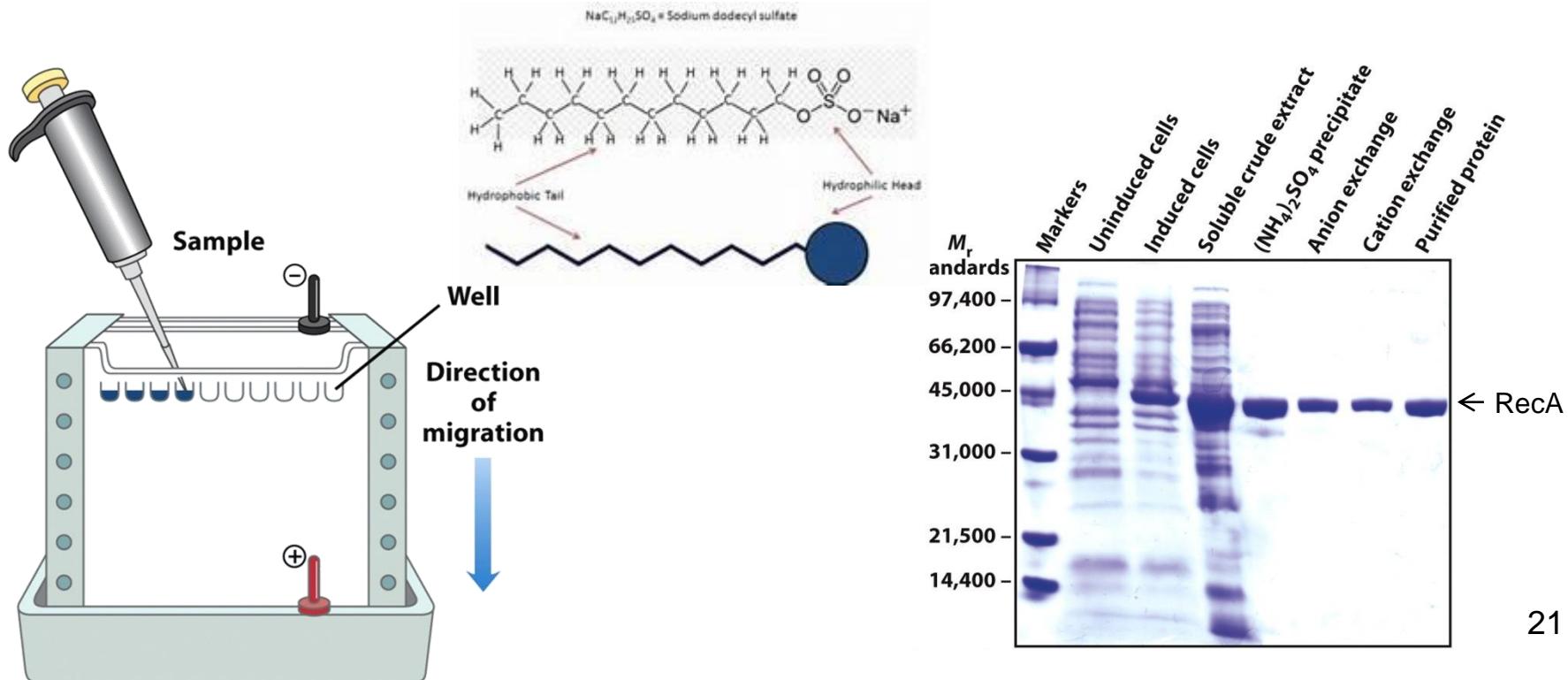


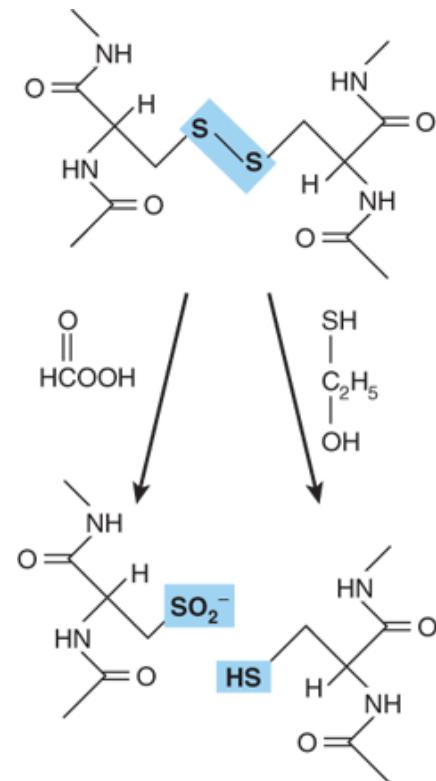
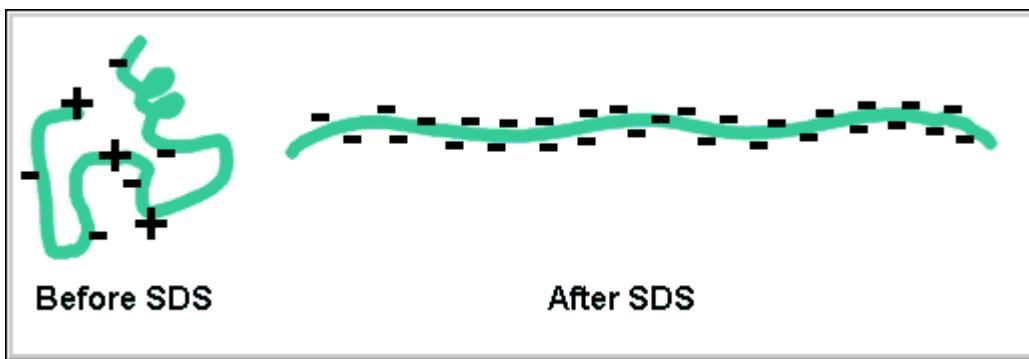
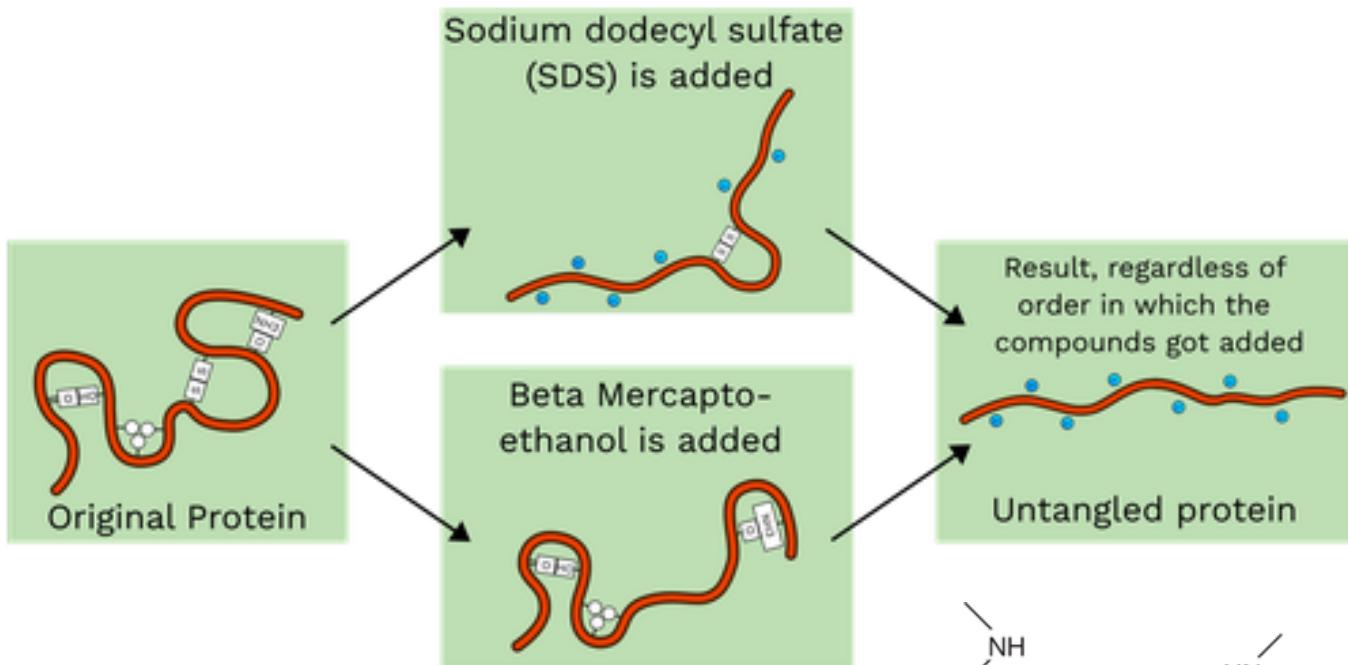
Ligand:
NTA as Chelator
and Ni²⁺ Ion
Ni-NTA



Separation and characterization of proteins by electrophoresis

- **Electrophoresis** is based on the migration of charged proteins in an electric field.
- Electrophoresis of proteins is generally carried out in gels made of the cross-linked polymer.
- **SDS-PAGE (sodium dodecyl sulfate-polyacrylamide gel electrophoresis)** provides estimation of the protein *purity* and an approximate *molecular weight*.





Estimating the molecular weight of a protein

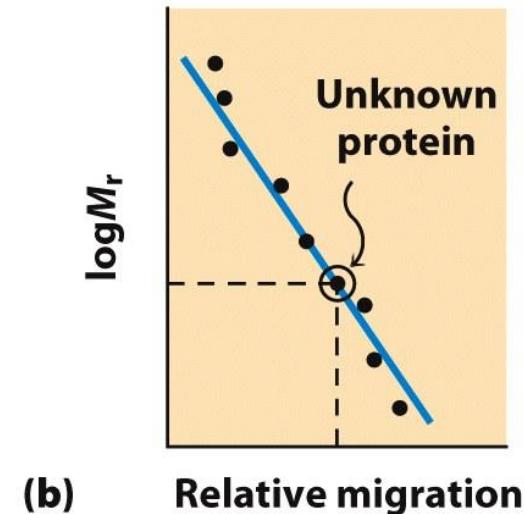
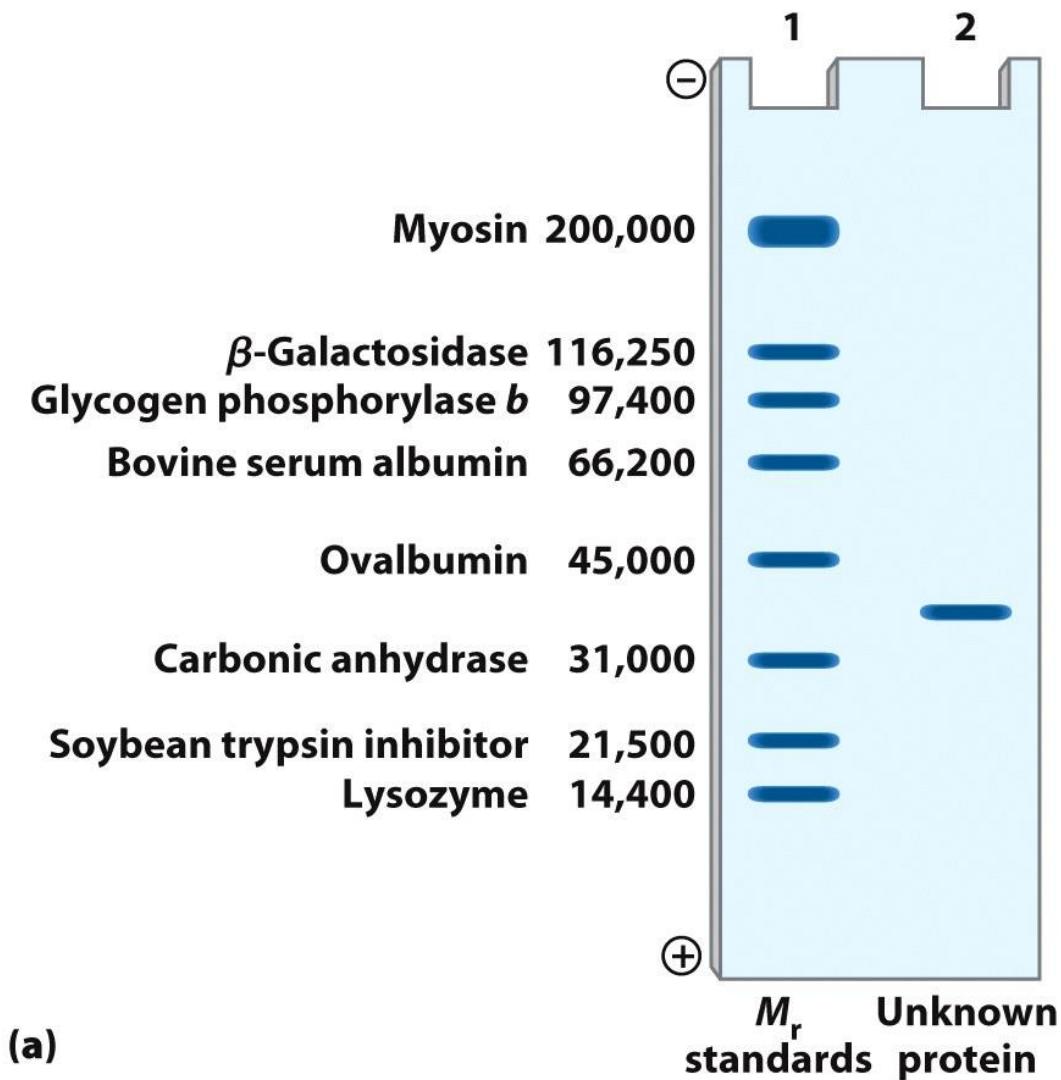


Figure 3-19

Lehninger Principles of Biochemistry, Fifth Edition

© 2008 W.H. Freeman and Company

Isoelectric focusing

- **IEF** is a procedure used to determine the isoelectric point (pl) of a protein.
- When a protein mixture is applied, each protein migrates until it reaches the pH that matches its pl.

CH₂CH—CO—NH—R, where R denotes either two different weak carboxyl groups, with pKa values of 3.6 and 4.6, or four tertiary amino groups, with pKa values of 6.2, 7.0, 8.5, and 9.3

An ampholyte solution is incorporated into a gel.

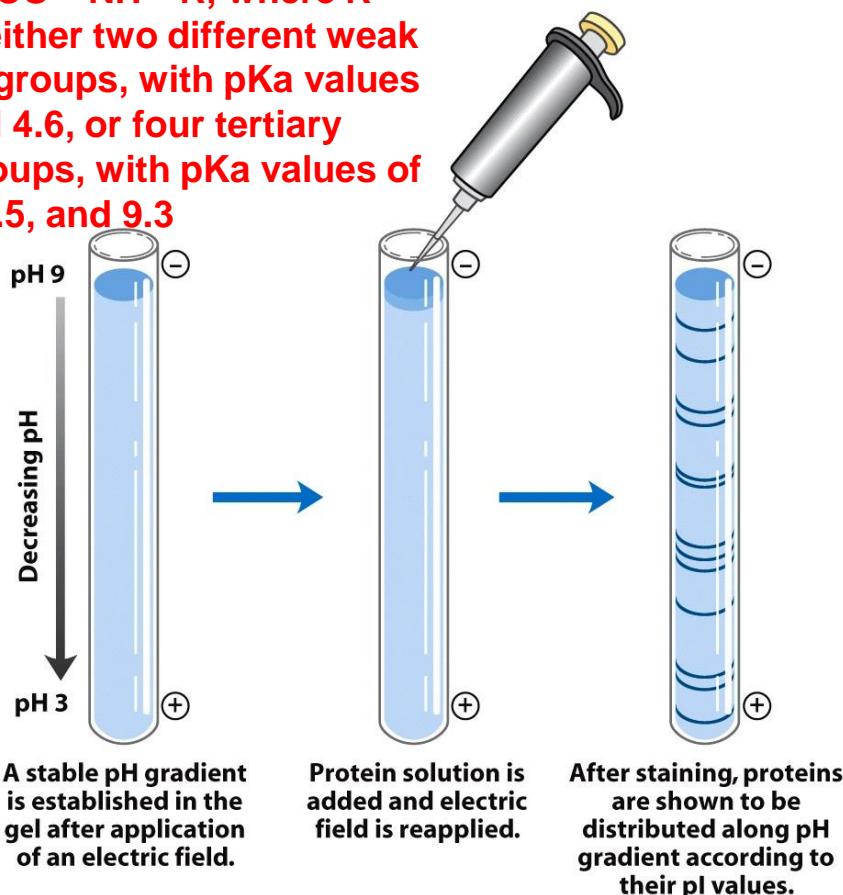
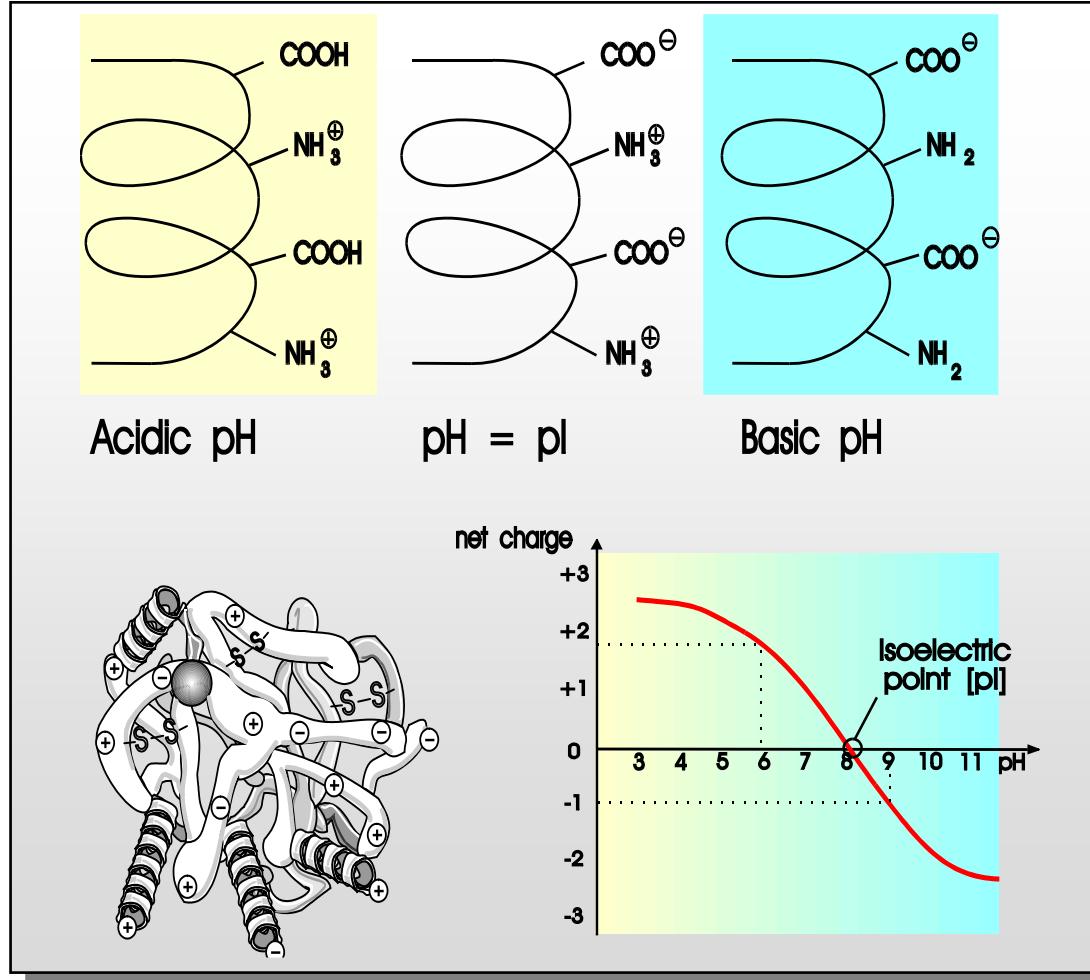


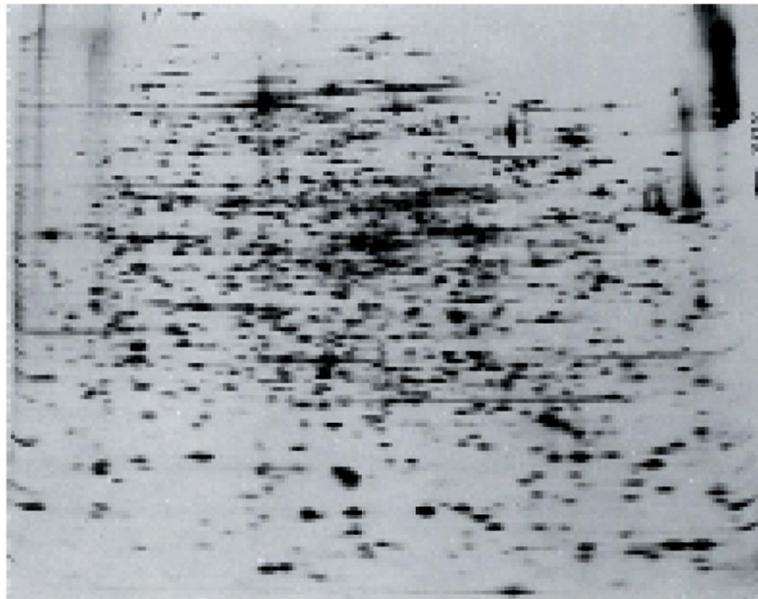
TABLE 3–6		The Isoelectric Points of Some Proteins
Protein	pl	
Pepsin	<1.0	
Egg albumin	4.6	
Serum albumin	4.9	
Urease	5.0	
β-Lactoglobulin	5.2	
Hemoglobin	6.8	
Myoglobin	7.0	
Chymotrypsinogen	9.5	
Cytochrome c	10.7	
Lysozyme	11.0	

Net Charges on Proteins

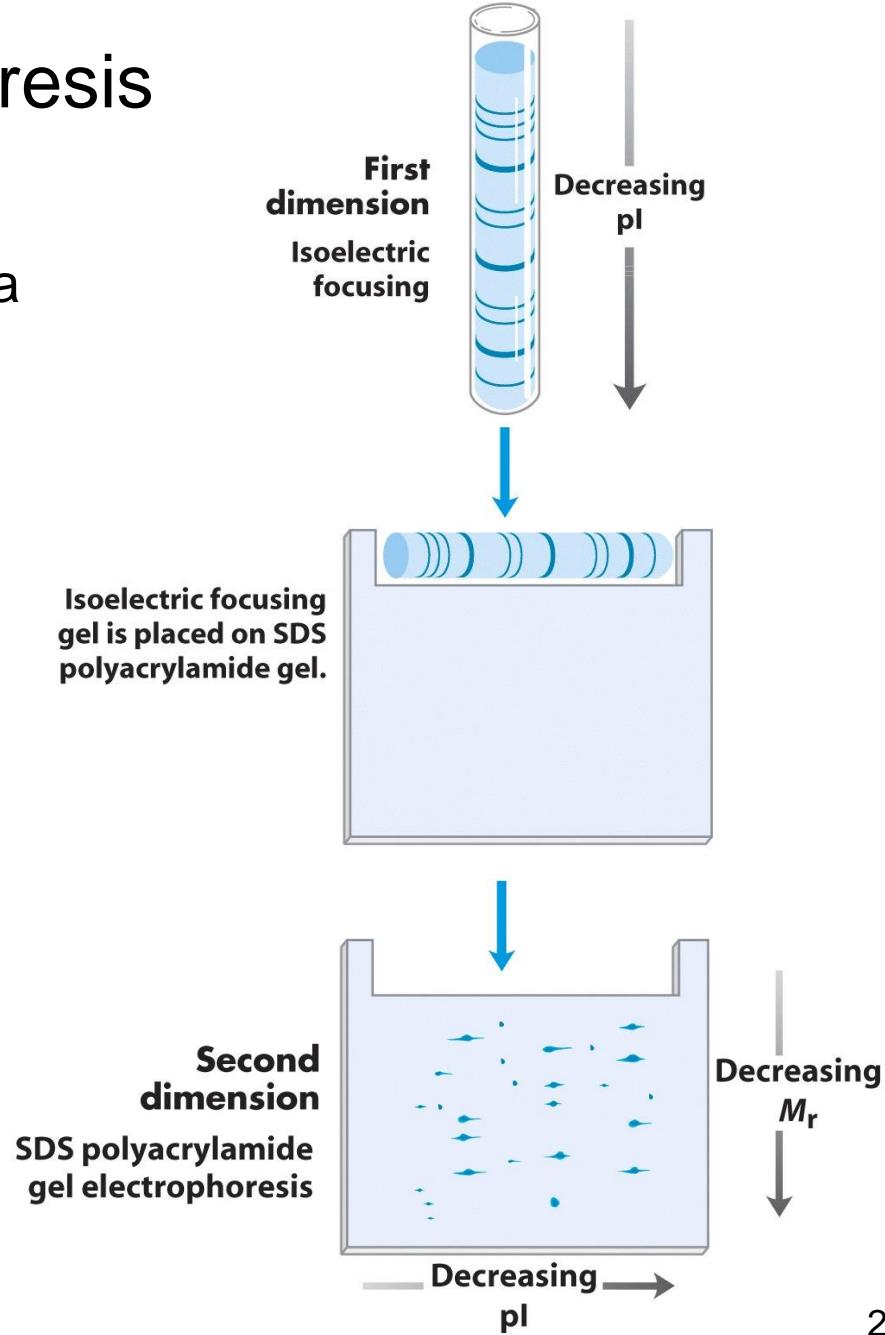


Two-dimensional electrophoresis

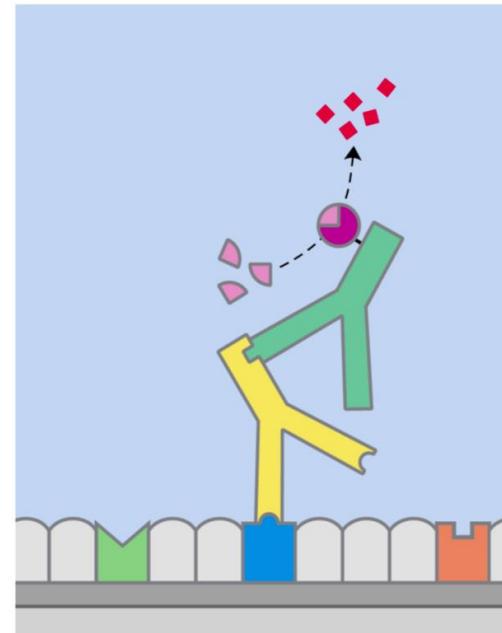
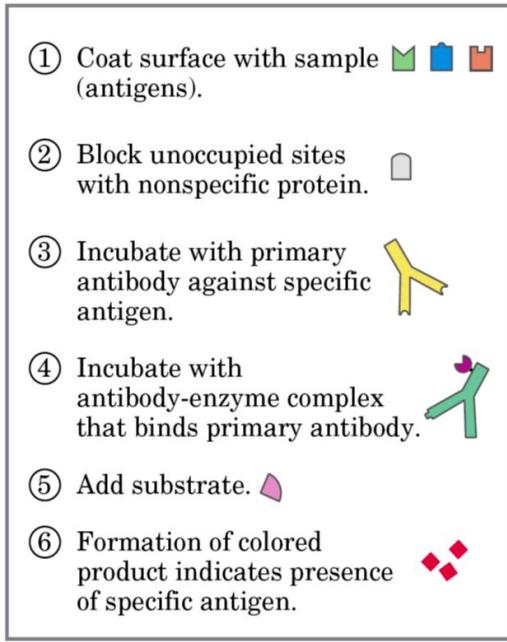
- Combining isoelectric focusing and SDS electrophoresis sequentially in a process called **two-dimensional electrophoresis** permits the resolution of complex mixtures of proteins.



More than 1,000 different proteins from *E. coli* can be resolved using this technique.



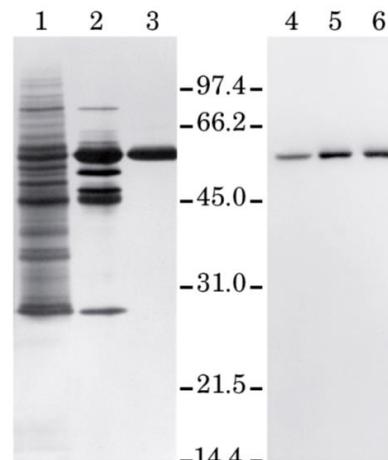
Immunoblot and ELISA (Enzyme linked immunosorbent assay)



(a)



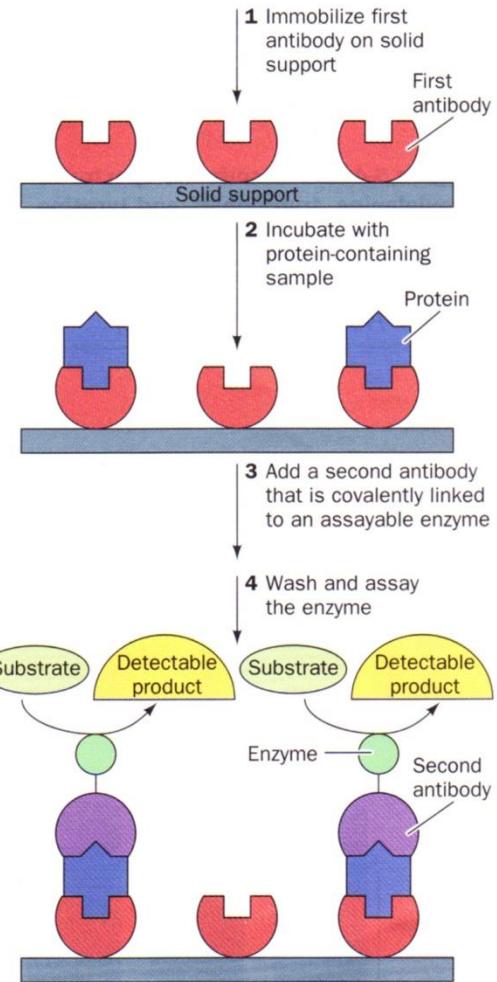
ELISA
(b)



SDS gel



Immunoblot (Western blot)
(c)



Protein sequence analysis

Step 1: determine the amino-acid composition by acid hydrolysis.
6N HCl, heat

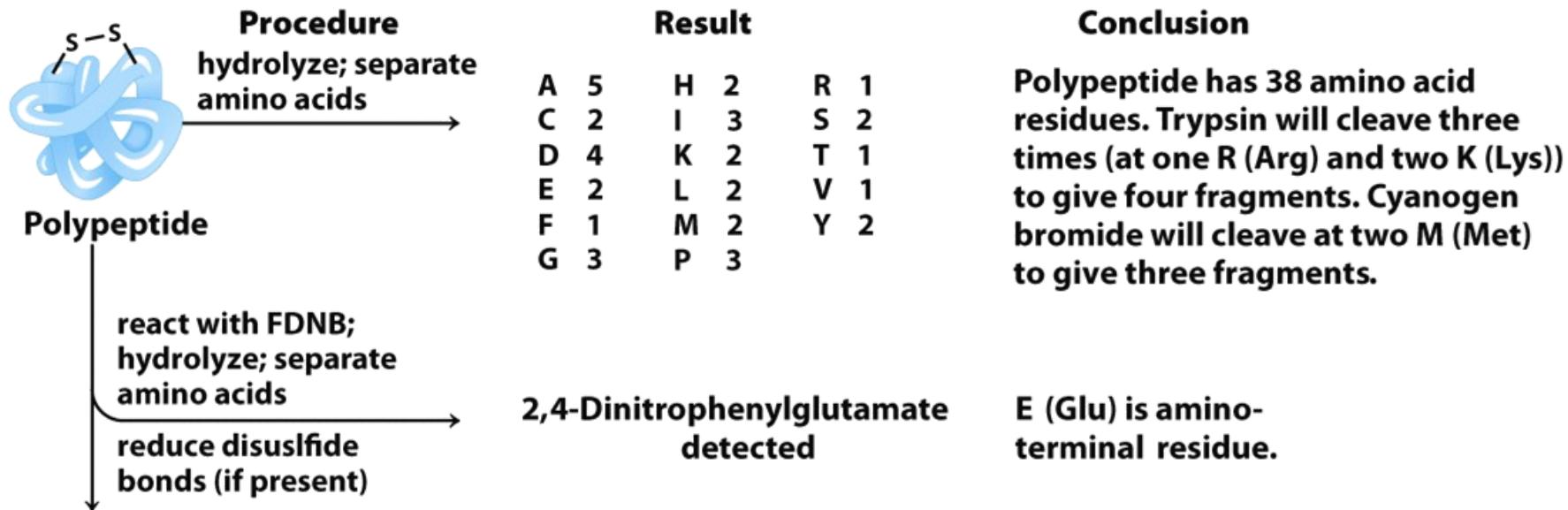
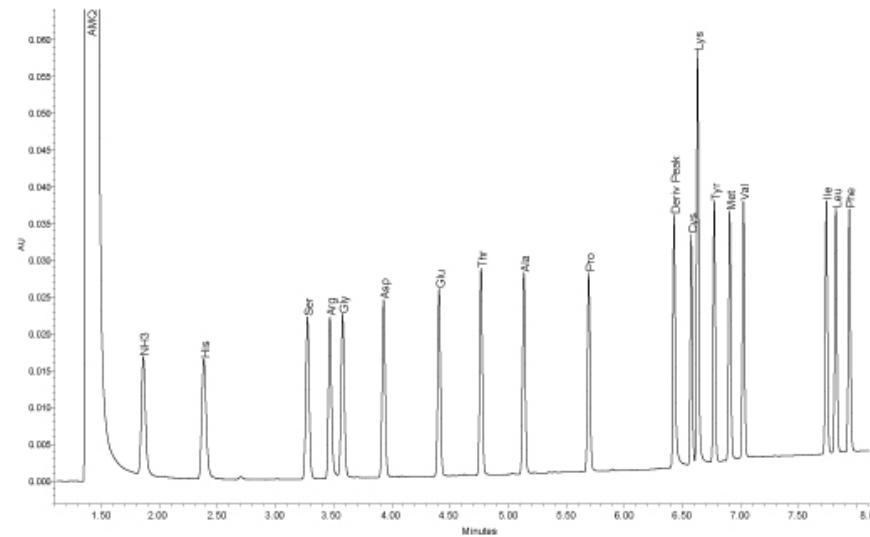


TABLE 3–3 Amino Acid Composition of Two Proteins

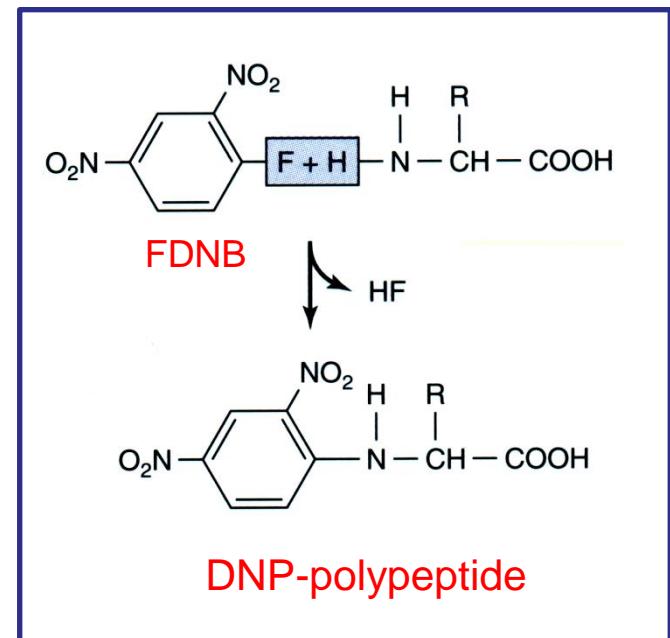
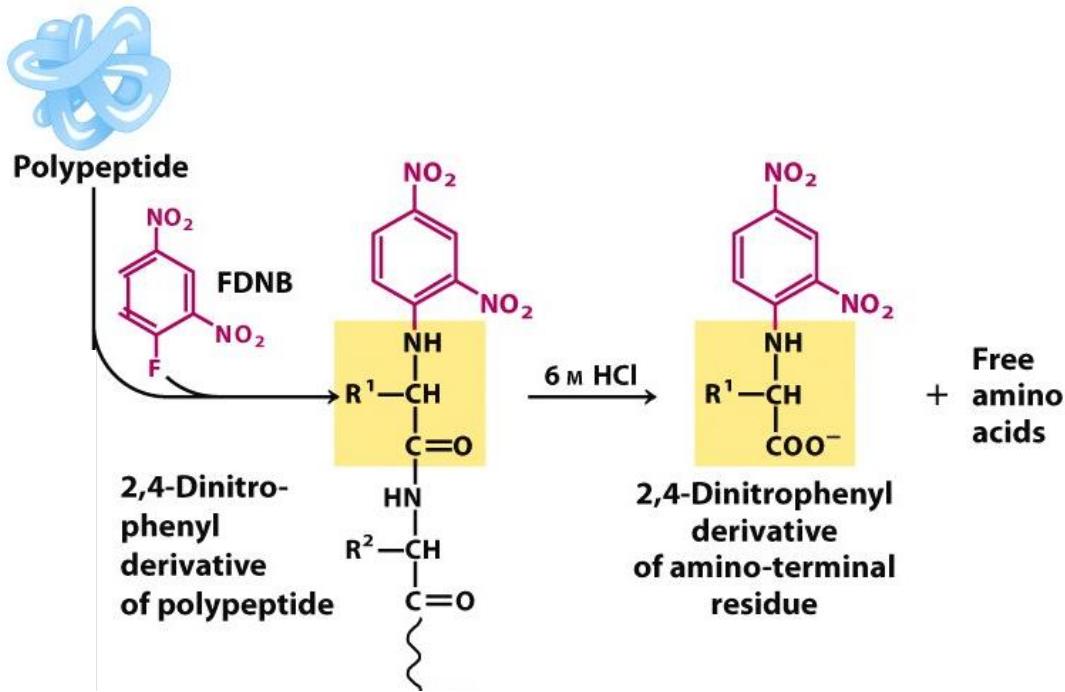
Amino acid	Number of residues per molecule of protein*	
	Bovine cytochrome c	Bovine chymotrypsinogen
Ala	6	22
Arg	2	4
Asn	5	15
Asp	3	8
Cys	2	10
Gln	3	10
Glu	9	5
Gly	14	23
His	3	2
Ile	6	10
Leu	6	19
Lys	18	14
Met	2	2
Phe	4	6
Pro	4	9
Ser	1	28
Thr	8	23
Trp	1	8
Tyr	4	4
Val	3	23
Total	104	245



Protein sequence analysis

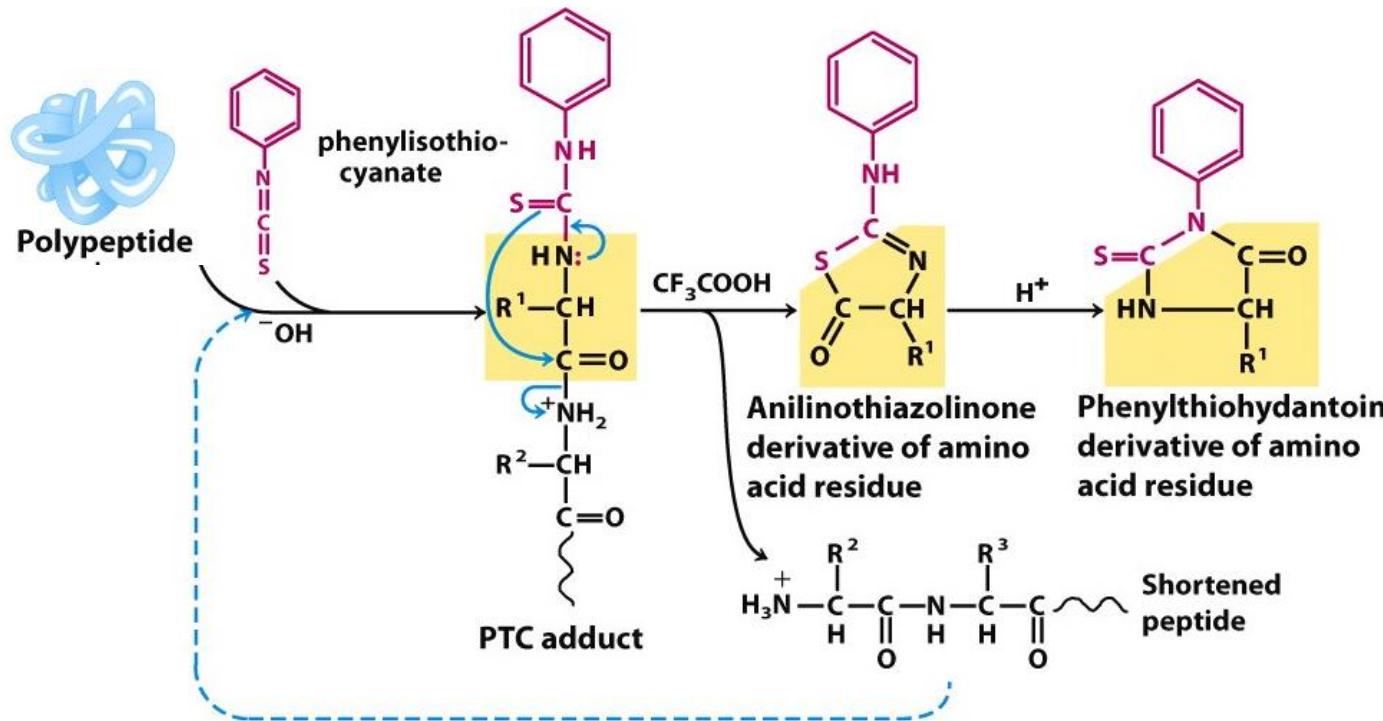
Step 2: determine the amino-terminal residue of polypeptide.

Sanger's reagent: 1-fluoro-2,4-dinitrobenzene (FDNB)



Protein sequence analysis

Step 3: determine the sequence by “Edman degradation”.



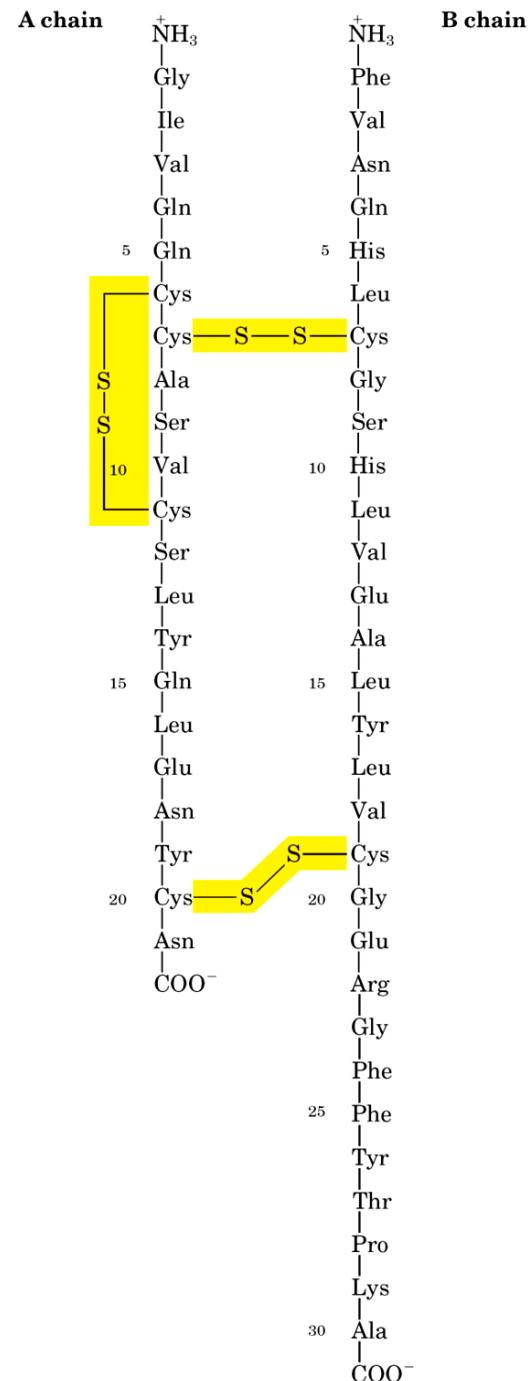
- The **Edman degradation** procedure labels and removes only the amino-terminal residue from a peptide, leaving all other peptide bonds intact.
- The Edman degradation is carried out in a machine, called a **sequencer**, that mixes reagents in the proper proportions, separates the products, identifies them, and records the results.

Protein structure analysis

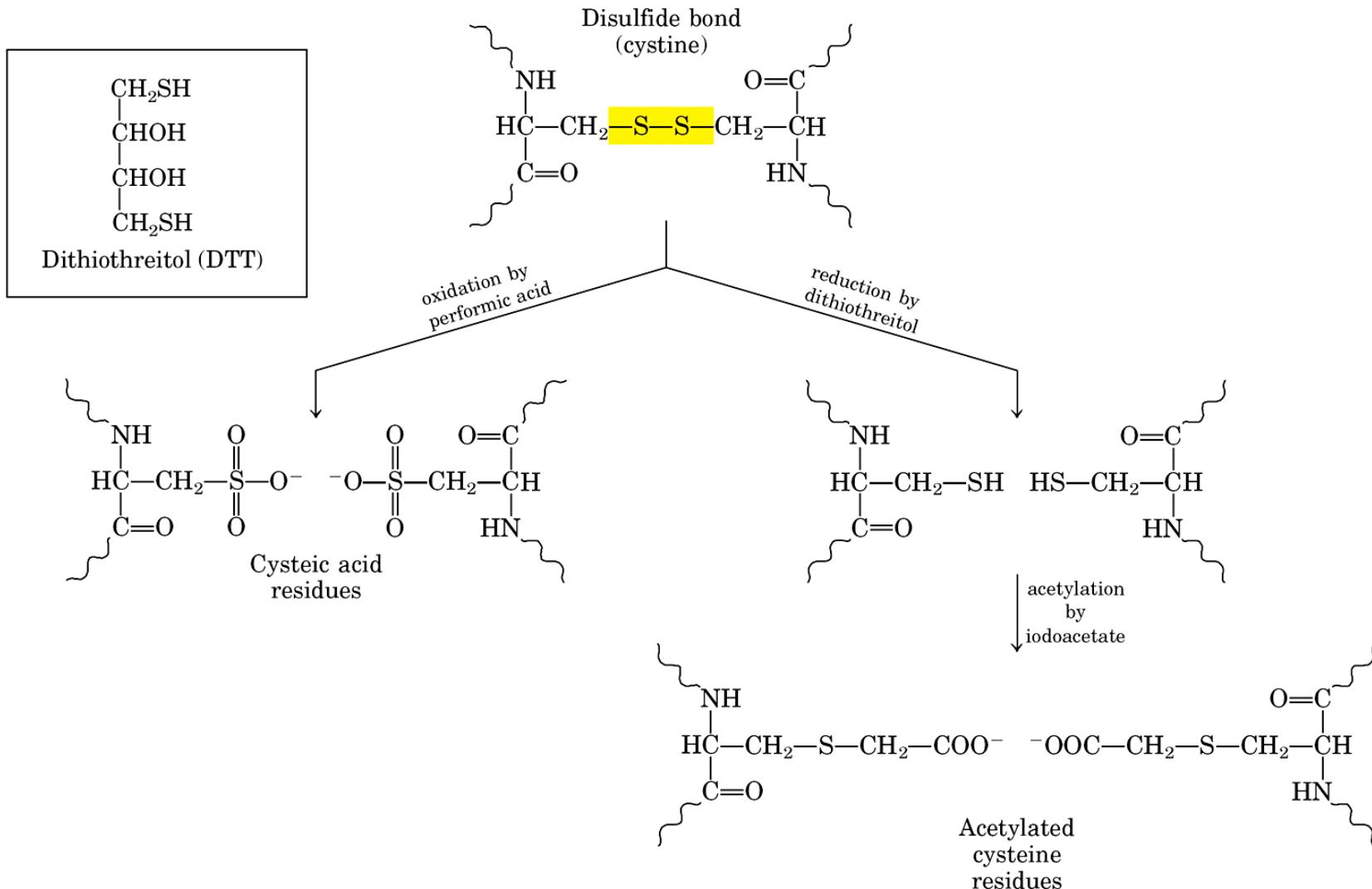
Amino acid sequence of bovine insulin



Protein sequencer



Breaking disulfide bonds in proteins



Large proteins are sequenced in smaller segments

- The overall accuracy of amino acid sequencing generally declines as the length of the polypeptide increases.
- The very large polypeptides found in proteins must be broken down into smaller pieces to be sequenced efficiently.
 - Enzymes called proteases catalyze the hydrolytic cleavage of peptide bonds.
 - Some **proteases** cleave only the peptide bond adjacent to particular amino acid residues and thus fragment a polypeptide chain in a predictable and reproducible way.
 - The number of smaller peptides produced by trypsin cleavage can thus be predicted from the total number of Lys or Arg residues in the original polypeptide, as determined by hydrolysis of an intact sample.

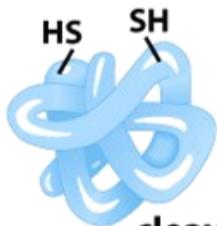
TABLE 3–7**The Specificity of Some Common Methods for Fragmenting Polypeptide Chains**

Reagent (biological source)*	Cleavage points†
Trypsin (bovine pancreas)	Lys, Arg (C)
<i>Submaxillarus</i> protease (mouse submaxillary gland)	Arg (C)
Chymotrypsin (bovine pancreas)	Phe, Trp, Tyr (C)
<i>Staphylococcus aureus</i> V8 protease (bacterium <i>S. aureus</i>)	Asp, Glu (C)
Asp-N-protease (bacterium <i>Pseudomonas fragi</i>)	Asp, Glu (N)
Pepsin (porcine stomach)	Leu, Phe, Trp, Tyr (N)
Endoproteinase Lys C (bacterium <i>Lysobacter enzymogenes</i>)	Lys (C)
Cyanogen bromide	Met (C)

*All reagents except cyanogen bromide are proteases. All are available from commercial sources.

†Residues furnishing the primary recognition point for the protease or reagent; peptide bond cleavage occurs on either the carbonyl (C) or the amino (N) side of the indicated amino acid residues.

Cleaving proteins and sequencing and ordering the peptide fragments



cleave with **trypsin**; separate fragments; sequence by Edman degradation

- (T-1) GASMALIK
- (T-2) EGAAYHDFEPIDPR
- (T-3) DCVHSD
- (T-4) YLIACGPMTK

(T-2) placed at amino terminus because it begins with E (Glu).
(T-3) placed at carboxyl terminus because it does not end with R (Arg) or K (Lys).

cleave with **cyanogen bromide**; separate fragments; sequence by Edman degradation

- (C-1) EGAAYHDFEPIDPRGASM
- (C-2) TKDCVHSD
- (C-3) ALIKYLIACGPM

(C-3) overlaps with (T-1) and (T-4), allowing them to be ordered.

establish sequence

Amino terminus

EGAAAYHDFEPIDPRGASMALIKYLIACGPMTKDCVHSD

Carboxyl terminus

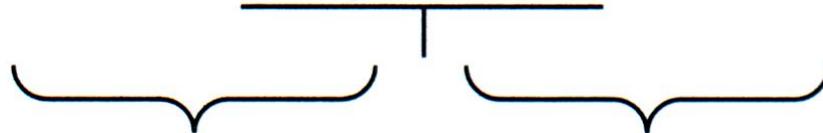
(C-1)

(C-3)

(C-2)

Peptide X Peptide Y

Peptide Z



Carboxyl terminal portion of peptide X Amino terminal portion of peptide Y

Amino acid sequences can also be deduced by other methods

- Deduce the sequence of a polypeptide by determining the sequence of nucleotides in the gene that codes for it.

Amino acid sequence (protein)	Gln – Tyr – Pro – Thr – Ile – Trp
DNA sequence (gene)	CAGTATCCTACGATTGG

- New methods based on mass spectrometry permit the sequencing of short polypeptides in just a few minutes.
- The entire protein complement encoded by an organism's DNA is termed **proteome**.

Proteins: Polypeptides of Defined Sequence

- Every protein has a defined number and order of amino acid residues.
- As with the nucleic acids, this sequence is referred to as the ***primary structure*** of the protein.

Key:	
	Identical amino acids
	Conservative substitutions
	Nonconservative substitutions

Number	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15			
Human	G	L	S	D	G	E	W	Q	L	V	L	N	V	W	G			
Whale	V	L	S	E	G	E	W	Q	L	V	L	H	V	W	A			
Number	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30			
Human	K	V	E	A	D	I	P	G	H	G	Q	E	V	L	I			
Whale	K	V	E	A	D	V	A	G	H	G	Q	D	I	L	I			
Number	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45			
Human	R	L	F	K	G	H	P	E	T	L	E	K	F	D	K			
Whale	R	L	F	K	S	H	P	E	T	L	E	K	F	D	R			
Number	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60			
Human	F	K	H	L	K	S	E	D	E	M	K	A	S	E	D			
Whale	F	K	H	L	K	T	E	A	E	M	K	A	S	E	D			
Number	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75			
Human	L	K	K	H	G	A	T	V	L	T	A	L	G	G	I			
Whale	L	K	K	H	G	V	T	V	L	T	A	L	G	A	I			
Number	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90			
Human	L	K	K	K	G	H	H	E	A	E	I	K	P	L	A			
Whale	L	K	K	K	G	H	H	E	A	E	L	K	P	L	A			
Number	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105			
Human	Q	S	H	A	T	K	H	K	I	P	V	K	Y	L	E			
Whale	Q	S	H	A	T	K	H	K	I	P	I	K	Y	L	E			
Number	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120			
Human	F	I	S	E	C	I	I	Q	V	L	Q	S	K	H	P			
Whale	F	I	S	E	A	I	I	H	V	L	H	S	R	H	P			
Number	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135			
Human	G	D	F	G	A	D	A	Q	G	A	M	N	K	A	L			
Whale	G	N	F	G	A	D	A	Q	G	A	M	N	K	A	L			
Number	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153
Human	E	L	F	R	K	D	M	A	S	N	Y	K	E	L	G	F	Q	G
Whale	E	L	F	R	K	D	I	A	A	K	Y	K	E	L	G	Y	Q	G

Sequence homology between myoglobin in humans vs. whales.

Proteomics

- The proteins present in a given cell make up the so-called ***proteome*** of that cell.
- ***Proteomics*** is the field of study that attempts to understand the complex relationships between proteins and cell function through *global analysis* of the proteome
- Proteomics includes, among other things, efforts to understand how protein expression and/or post-translational modification levels change in cells, and the consequences of such changes.

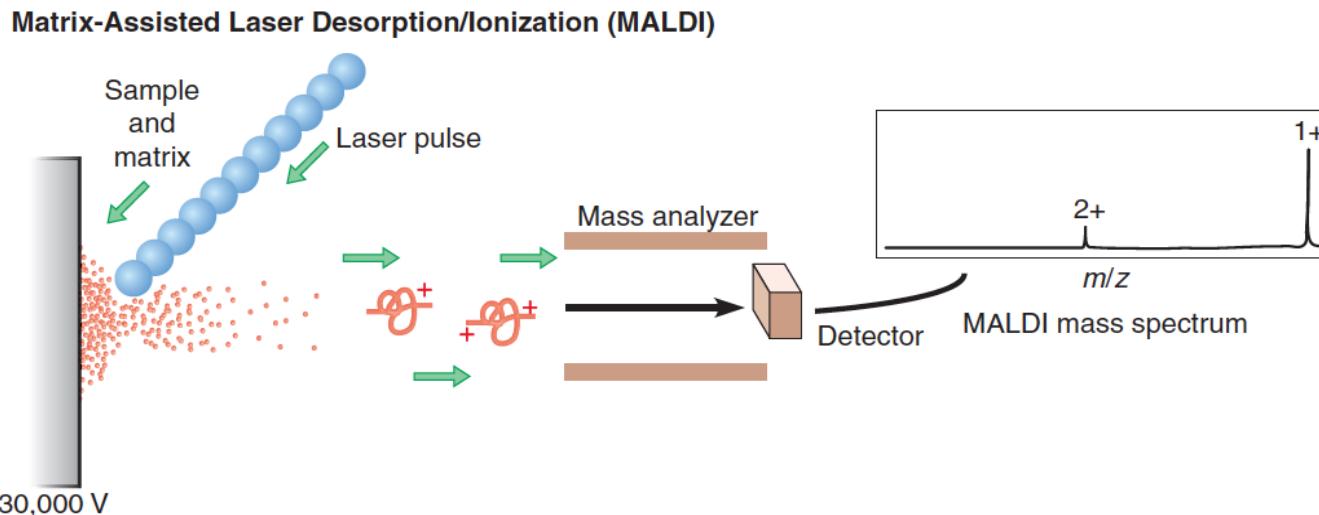
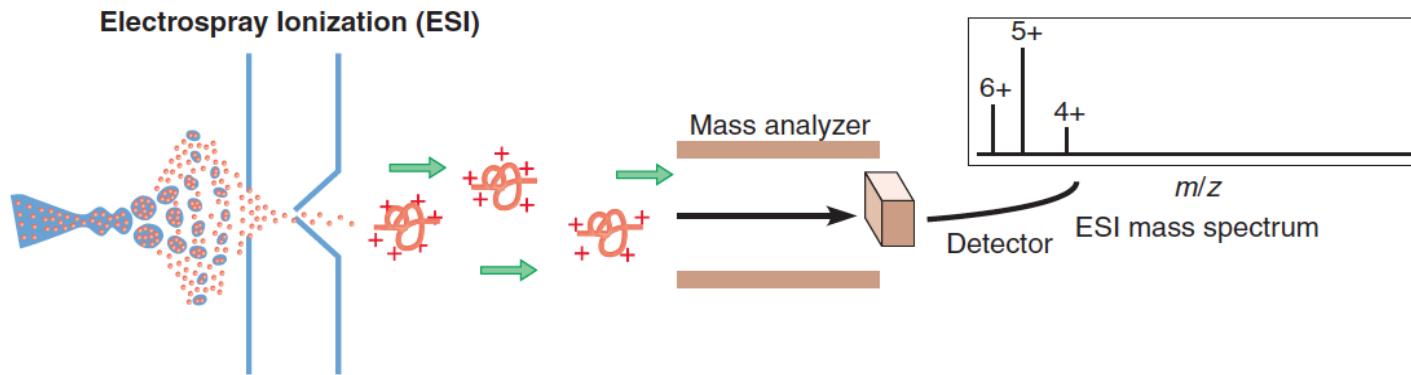
Proteomics

A typical proteomics experiment includes the following steps:

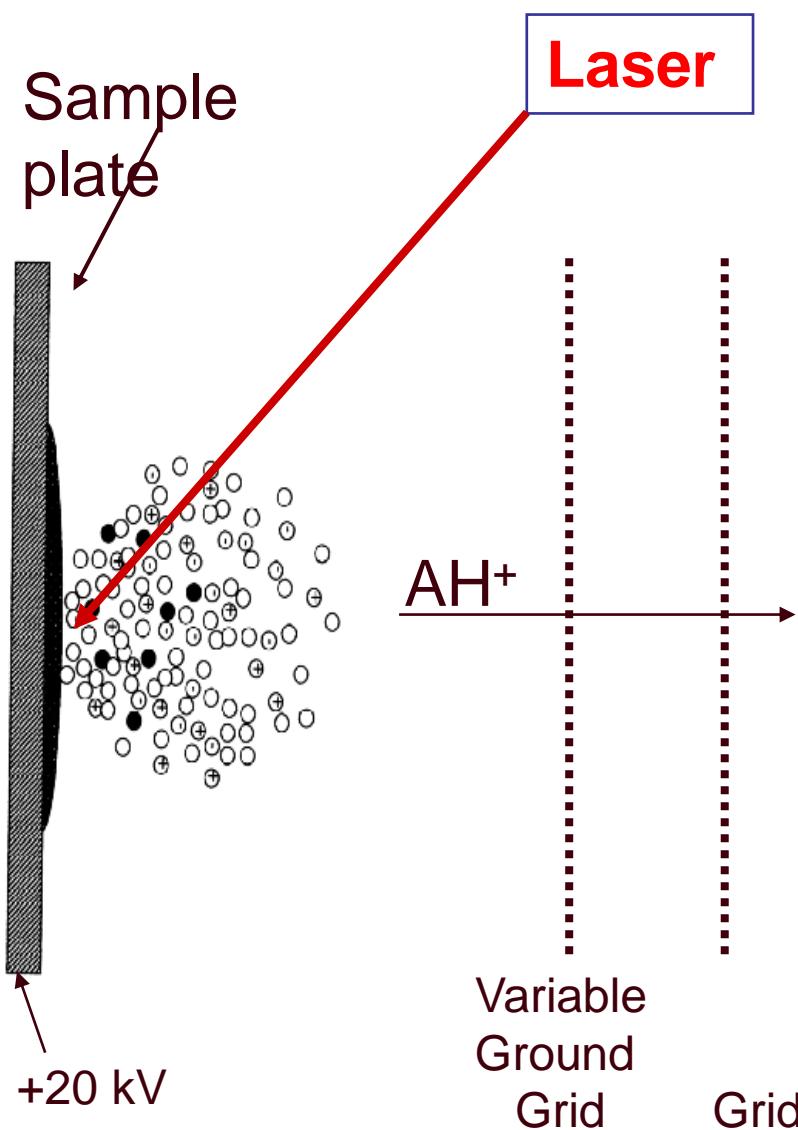
1. Separation and isolation of proteins, or protein fragments, from cells or an organism.
2. Identification by MS-MS sequencing of a particular protein within the complex mixture.
3. Database searching to identify the target protein, and its putative function.

Mass Spectrometry

Mass spectrometry (MS) provides the most accurate mass measurements of large biomolecules.

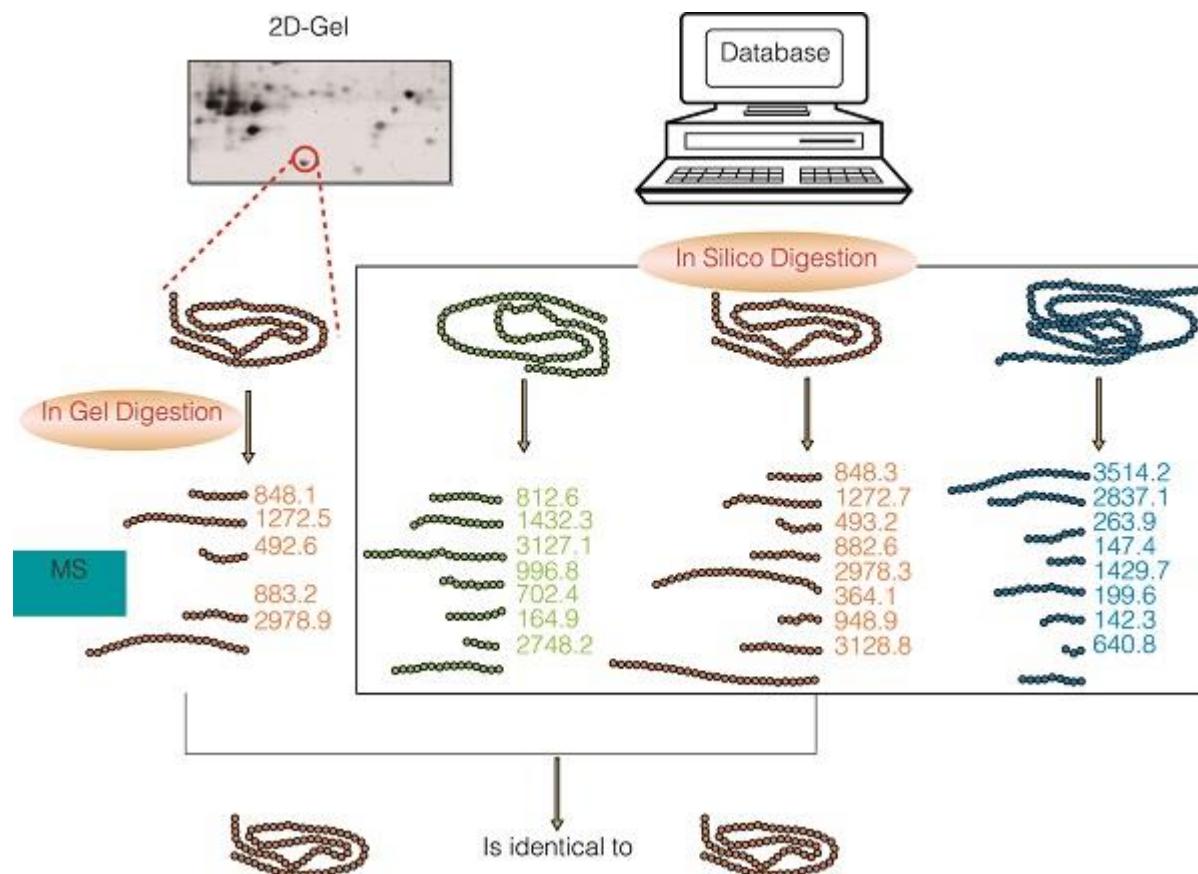


MALDI: Matrix Assisted Laser Desorption Ionization

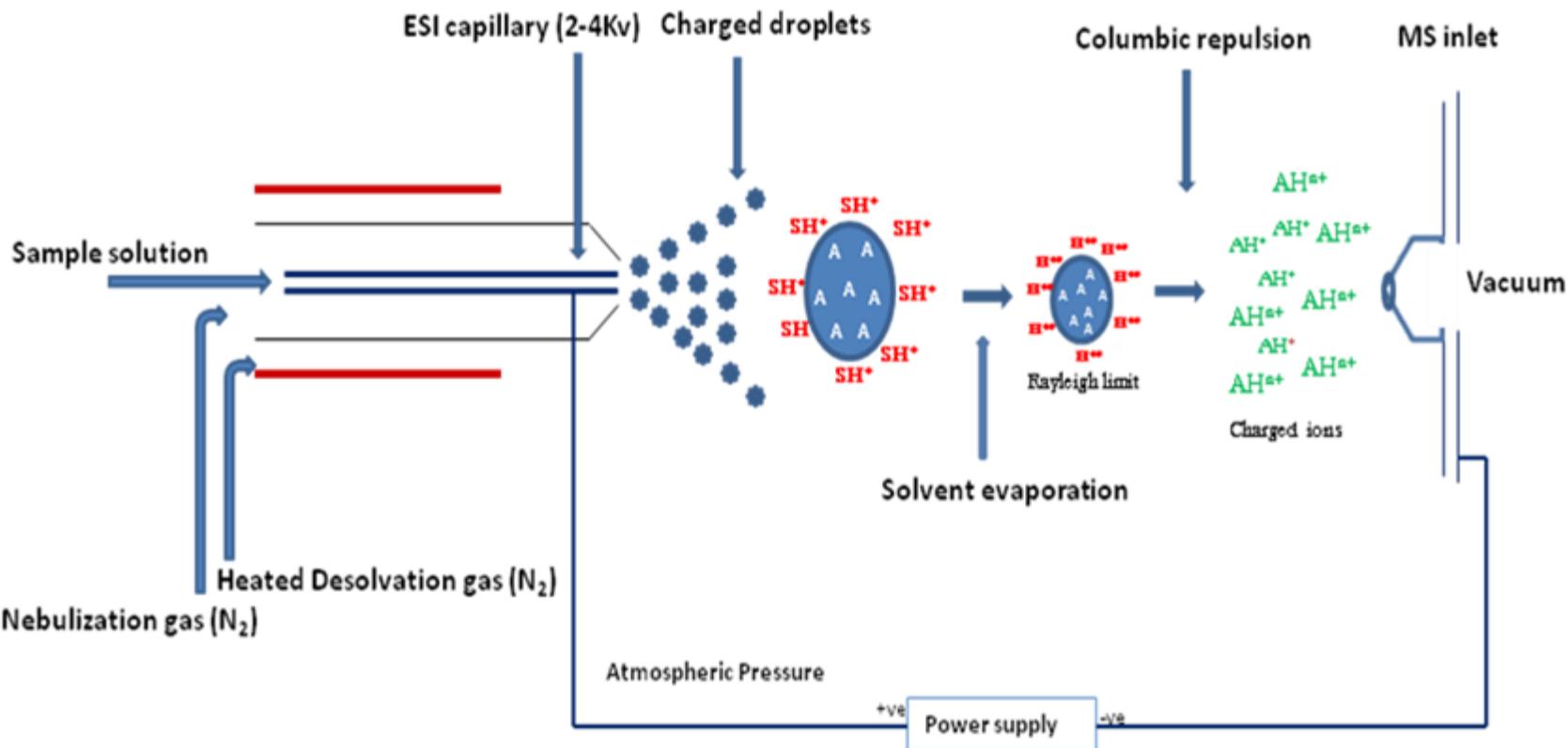


1. Sample (A) is mixed with excess matrix (M) and dried on a MALDI plate.
2. Laser flash ionizes matrix molecules.
3. Sample molecules are ionized by proton transfer from matrix:
 $MH^+ + A \rightarrow M + AH^+$.

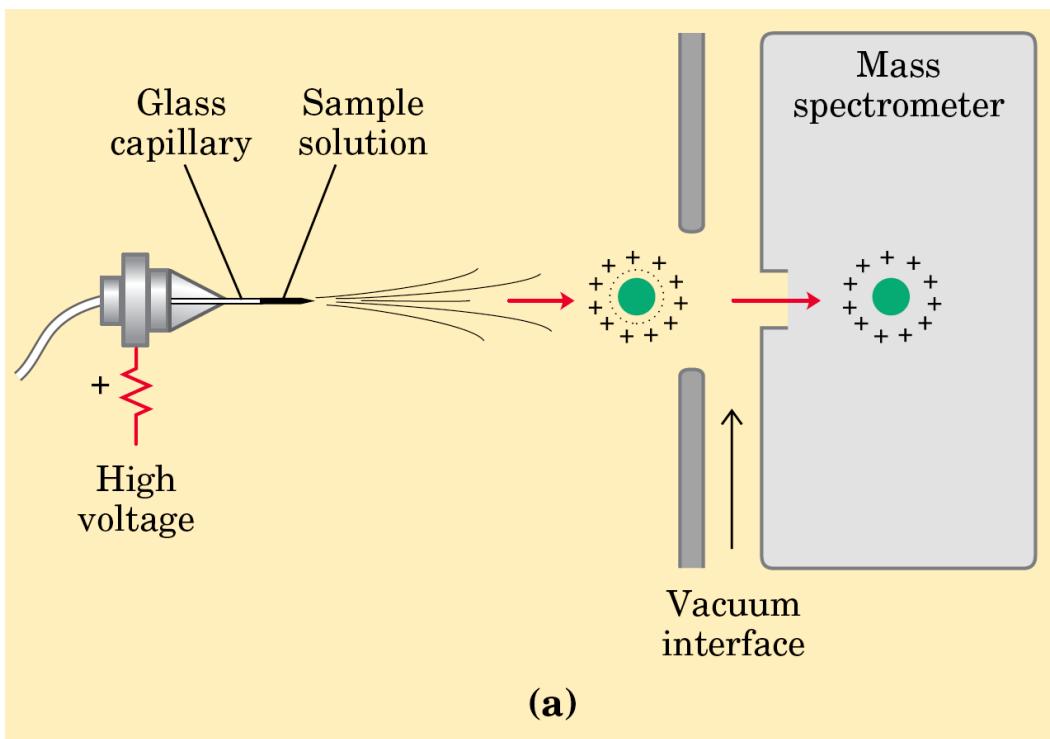
Peptide Mass Fingerprinting with MALDI-TOF



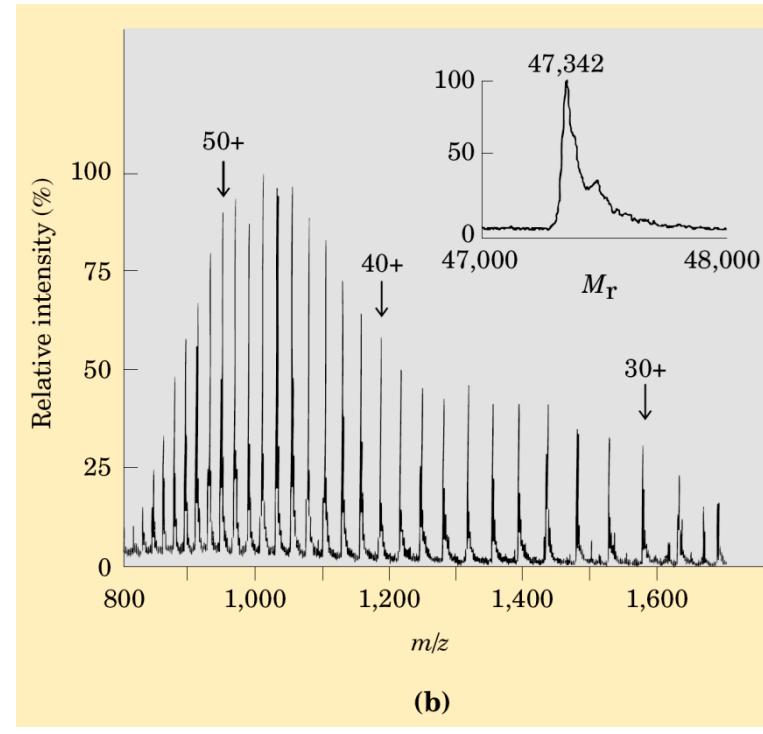
Electrospray Ionization (ESI)



Electrospray ionization mass spectrometry of a protein

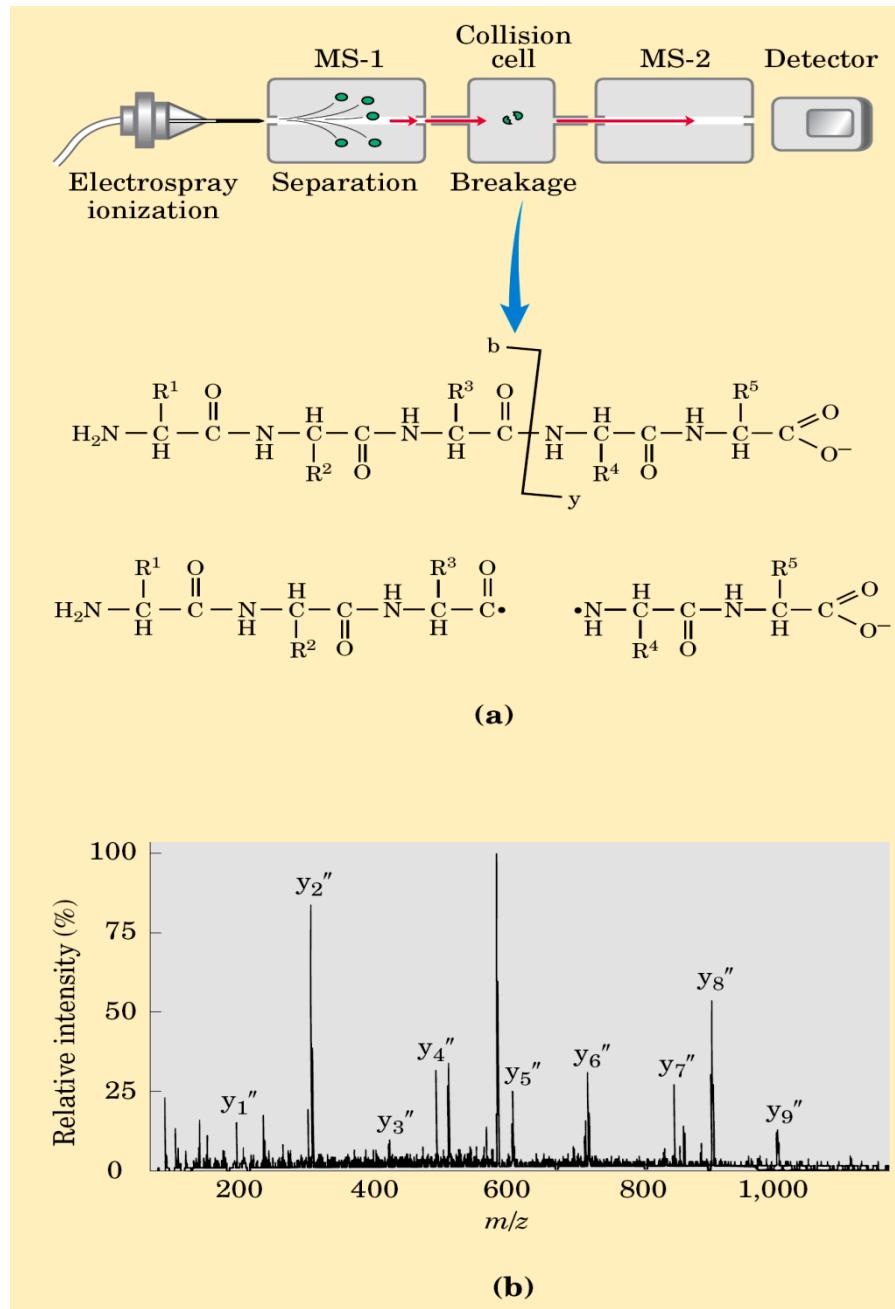


(Mass Spectrum)
荷質比大小排列之圖譜

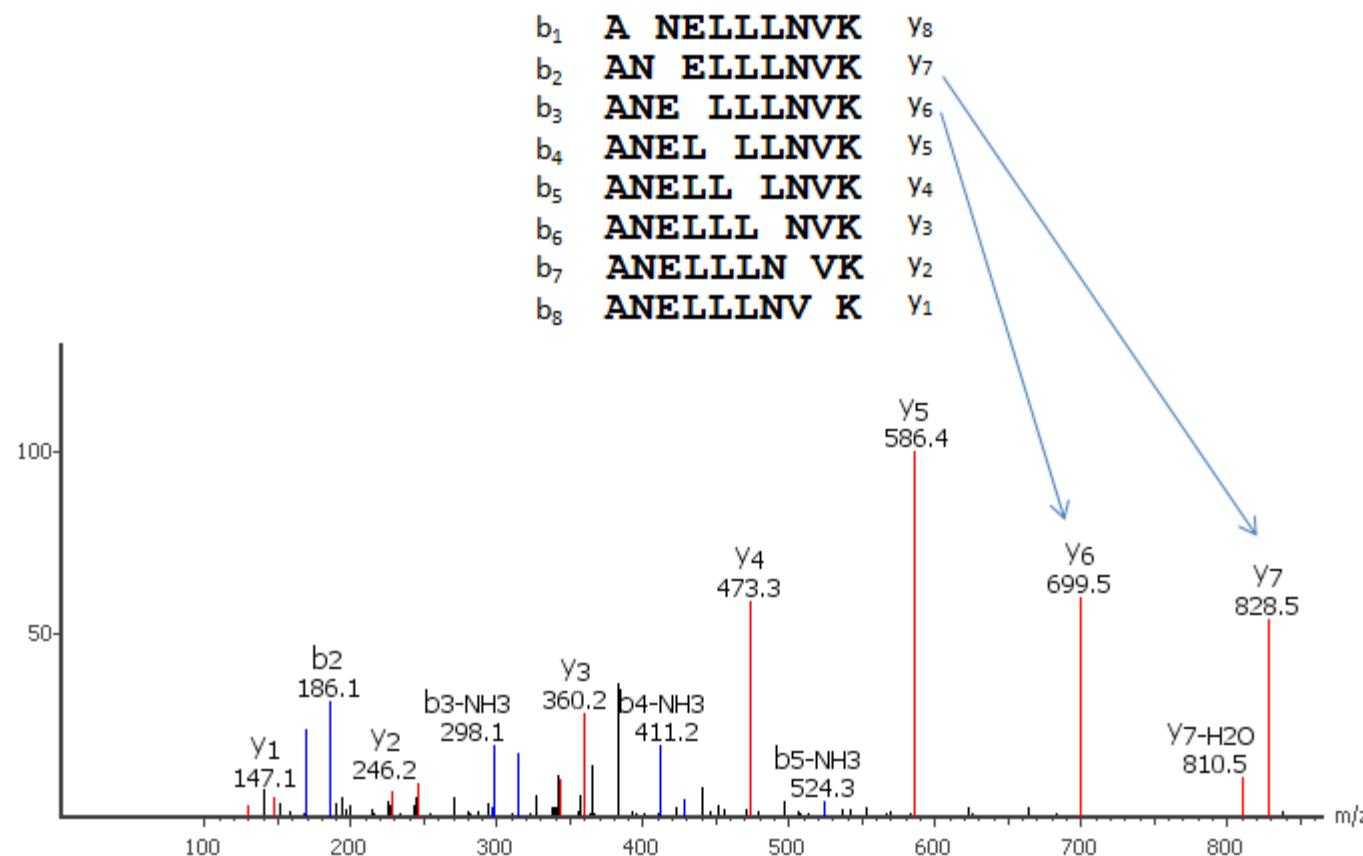


m/z : 質荷比
離子的質量與電荷的比值

Obtaining protein sequence information with tandem mass spectrometry



MS/MS Spectrum of Peptide



Amino acid sequence read as differences in weight between ions in series

Mass increases resulting from common post-translational modifications

Modification	Mass Increase (Da)
Phosphorylation	80
Hydroxylation	16
Methylation	14
Acetylation	42
Myristylation	210
Palmitoylation	238
Glycosylation	162

General approach used by MS technologies for the identification of proteins in complex mixtures

