

Proteins 1: Biological Function, Purification and Analysis

Shoba Ranganathan

Applied Biosciences

T: 02 9850 6262; E: shoba.ranganathan@mq.edu.au

Final reminder for Pearson text

- New PDF guide iLearn (also in Discussion Forum).
- Tutorial 1 Quiz (1%) on Pearson only
- All Tutorial Quizzes and Tests will be on Pearson only.
- All Spot Tests from L5 on Pearson only
- No L4 Spot test.

Objectives

Australian Guide to Healthy Eating

- Properties of proteins
 - What do proteins do in cells?
- Protein separation and analysis approaches

AAM: Chapter 5



Proteins: agents of biological function

- Almost every cellular activity depends on proteins
 - “**Proteome**” – studying the entire protein component
 - **Enzymes**: largest group (20%) – catalysts for specific biological reaction(s): upto 10^{16} of the uncatalyzed rate
 - **Regulatory proteins**: 14% - regulate the ability of other proteins to carry out physiological functions
 - ❖ **Hormones**: insulin, somatotropin (pituitary), thyrotropin
 - ❖ **Gene regulatory proteins**: DNA-binding proteins for transcription control
 - **Transport proteins**: 5% - trucking specific cargo from one location to another
 - ❖ **Hemoglobin** for oxygen transport, serum albumin for carrying fatty acids from adipose tissues to organs
 - ❖ **Membrane transport proteins**: facilitate metabolite transport across the membrane *via* mainly channels or pores.

Proteins: biological function - 2

- **Storage proteins:** 5% - reservoir of essential nutrients
 - ❖ **Albumin** (egg white) and **casein** (milk): N storage
 - ❖ **Zeins** (in corn) and **phaseolin** (in peas): N storage
 - ❖ **Ferritin** stores iron in animals
- **Movement proteins:** support cell division, muscle contraction and cell motility
 - ❖ **Actin and myosin:** filamentous proteins for muscle contraction
 - ❖ **Tubulin:** in microtubules (in cell division and in flagella/cilia)
 - ❖ **Dynein and kinesin:** motor proteins for movement of vesicles, granules and organelles along microtubules
- **Structural proteins:** maintain integrity of biological structures as fibres
 - ❖ **α -keratins:** hair, horn, nail
 - ❖ **Collagen:** bone, connective tissue, tendons, cartilage, hide; protective barrier in the extracellular matrix with proteoglycans
 - ❖ **Elastin:** ligaments

Proteins: biological function - 3

- **Signalling proteins:** cellular response to hormones and growth factors in complex networks: signalling pathways
 - ❖ **Hormone receptors** and **protein kinases** which add phosphate groups to other proteins
 - ❖ **Scaffold or adapter proteins:** contains specific modules for recognition and binding of specific structural elements
 - ❖ **Anchoring or targetting proteins:** bind other proteins, causing them to be localized to specific cellular structures
- **Other proteins:**
 - ❖ **Protective:**
 - *immunoglobulins, antibodies* to recognize “foreign” molecules
 - *thrombin, fibrinogen* for clotting blood after injury
 - *Antifreeze proteins*
 - ❖ **Defensive/Exploitive:** toxins and venoms
 - ❖ **Exotic:** Monellin, from an African plant is x2000 sweeter than sugar

Protein Purification & Analysis

Overview

- Environmental conditions such as pH and temperature affect a protein's stability during purification.
- An assay based on a protein's chemical or binding properties may be used to quantify a protein during purification.
- Fractionation procedures take advantage of a protein's unique structure and chemistry in order to separate it from other molecules.

Key Concepts

- A protein's ionic charge, polarity, size, and ligand-binding ability influence its chromatographic behaviour.
- Gel electrophoresis and its variations can separate proteins according to charge, size, and isoelectric point.
- The overall size and shape of macromolecules and larger assemblies can be assessed through ultracentrifugation.

Protein purification needs a strategy

- The cell contains many proteins, in organelles but also in solution and in membranes
- Need to get proteins and other biomolecules out of the cell and into solution
- The protein of interest must be kept as close to its biological environment as possible to retain function.

Factors that affect a protein's stability and function

- **pH:** buffering is usually required.
- **Temperature:** most proteins denature (i.e. become addled) at high temperatures: even a few degrees is enough! Usually 0°C chosen!
- **Degradative enzymes:** proteases and nucleases are released by destroying tissues: temperature/pH control or inhibitors required.
- **Adsorption to surfaces:** many proteins denature at the air-water interface or when in contact with glass/plastic.
- **Storage:** after purification, under inert gas and/or frozen.

Protein purification methods

Methods are based on physical properties:

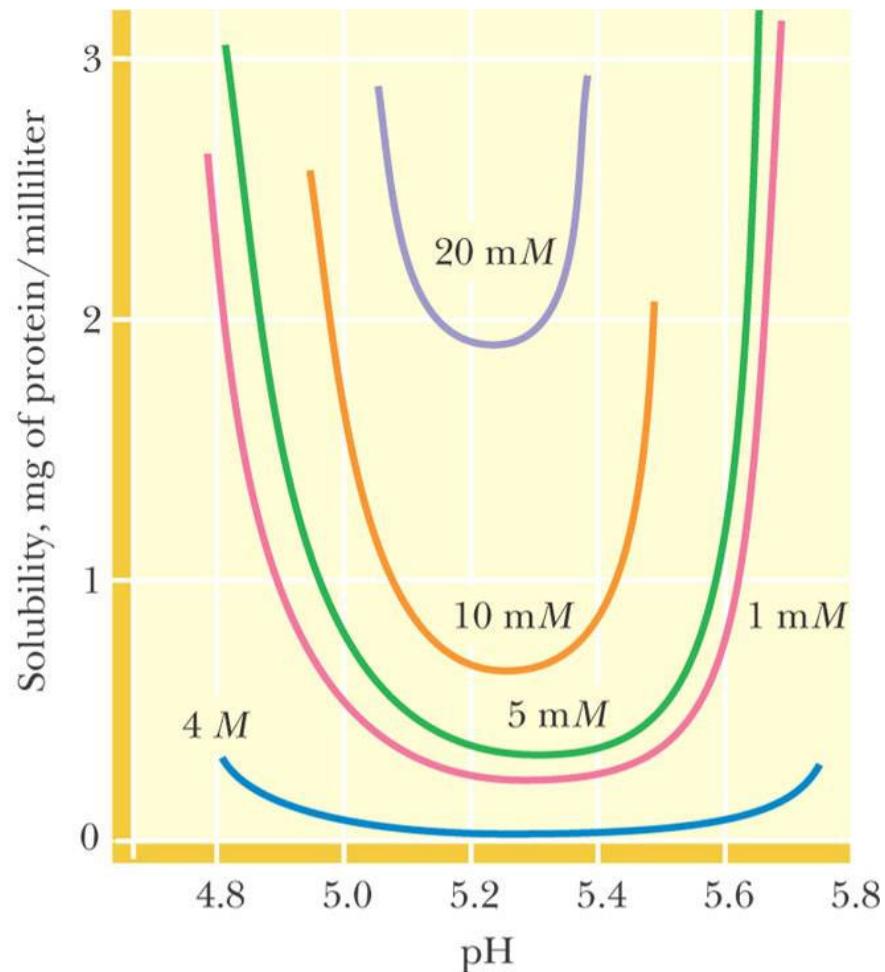
- Solubility of proteins
- Ionic Charge
- Mass
- Size
- Affinity

TABLE 5-2 Protein Purification Procedures

Protein Characteristic	Purification Procedure
Solubility	Salting out
Ionic Charge	Ion exchange chromatography Electrophoresis Isoelectric focusing
Mass	Ultracentrifugation
Size	Gel filtration chromatography SDS-PAGE
Binding Specificity	Affinity chromatography

Protein solubility

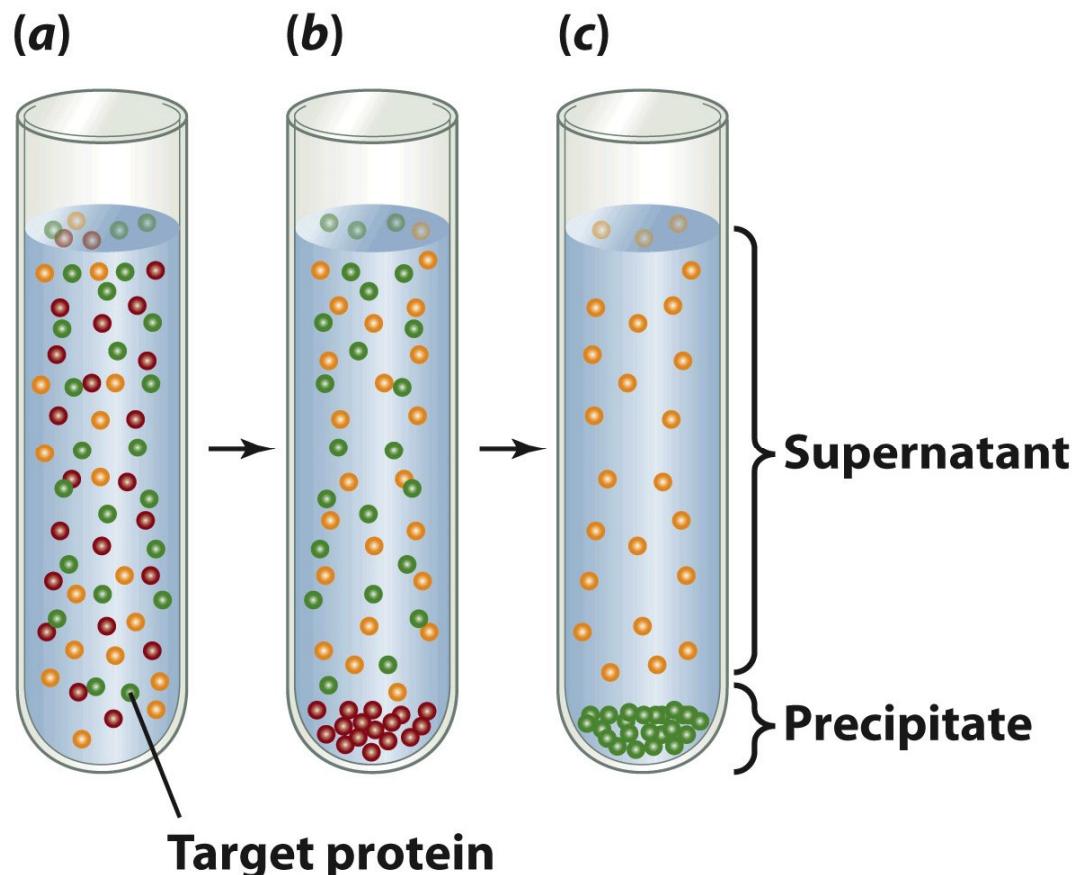
- Proteins are **least soluble at their pI**, when the net charge is zero.
 - At the pI, the electrostatic repulsion between the molecules is a minimum; protein aggregates and precipitates out.
- Generally, solubility increases with ionic strength.
 - The figure shows the solubility of a typical protein as a function of pH and various salt concentrations.



Solubility and ionic strength

- Salting-in
 - Most globular proteins are only slightly soluble in pure water
 - ❖ The **addition of salts** to increase the solubility of these proteins is called **salting-in** (typically ~20 mM)
 - ❖ The added salt induces charge interactions between the protein and the salt, rather than between protein molecules, thus preventing aggregation
- Salting-out
 - **Addition of excessive salt** ($> 1 \text{ M}$) precipitates the protein
 - ❖ The salt removes all the available water and hence the protein precipitates, which is called **salting-out**.
 - ❖ Some salts complex water better than others.
- **Increasing ionic strength at first increases the solubility of proteins (salting-in), then decreases it (salting-out).**
- Also, polyvalent ions act faster than singly charged ions.
 - ❖ Ammonium sulphate is commonly used for solubility-based separation of proteins.
- **Demo in Prac 4**

Fractionation by Salting Out



- Add salt of choice to solution containing the target protein
- Centrifuge and discard unwanted proteins (**red**)
- Add more salt to precipitate the protein of interest (**green**) and centrifuge



Electrophoresis: separation by charge and size

- A sieving method – migration of ions in an electric field.
 - Usually denatures the protein since it is coated with a detergent (sodium dodecylsulfate, SDS), to have a uniform negative charge.
 - Used for proteins as well as DNA.
- Referred to as Polyacrylamide Gel Electrophoresis (PAGE).
 - Separations based on gel filtration (size and shape) as well as electrophoretic mobility (ionic charge).
 - pH ~ 9 so that all proteins are negatively charged and will migrate to the anode, under the applied electric field. Molecules of similar size and charge will move together as a band through the gel.
 - Proteins are visualized on the gel by a stain in the gel, or by X-rays if radioactive, or by antibody binding (Western or immuno- blot).
 - SDS denatures proteins so that separation is by mass alone.

Ion Exchange Chromatography

- DEAE (diethylaminoethyl) cellulose and CM (carboxymethyl) cellulose are widely used resins
- At a pH above the pI or isoelectric point, a protein carries a **negative charge** and will bind to DEAE-cellulose
- Protein can be selectively eluted by applying a gradient of pH or increasing ionic strength

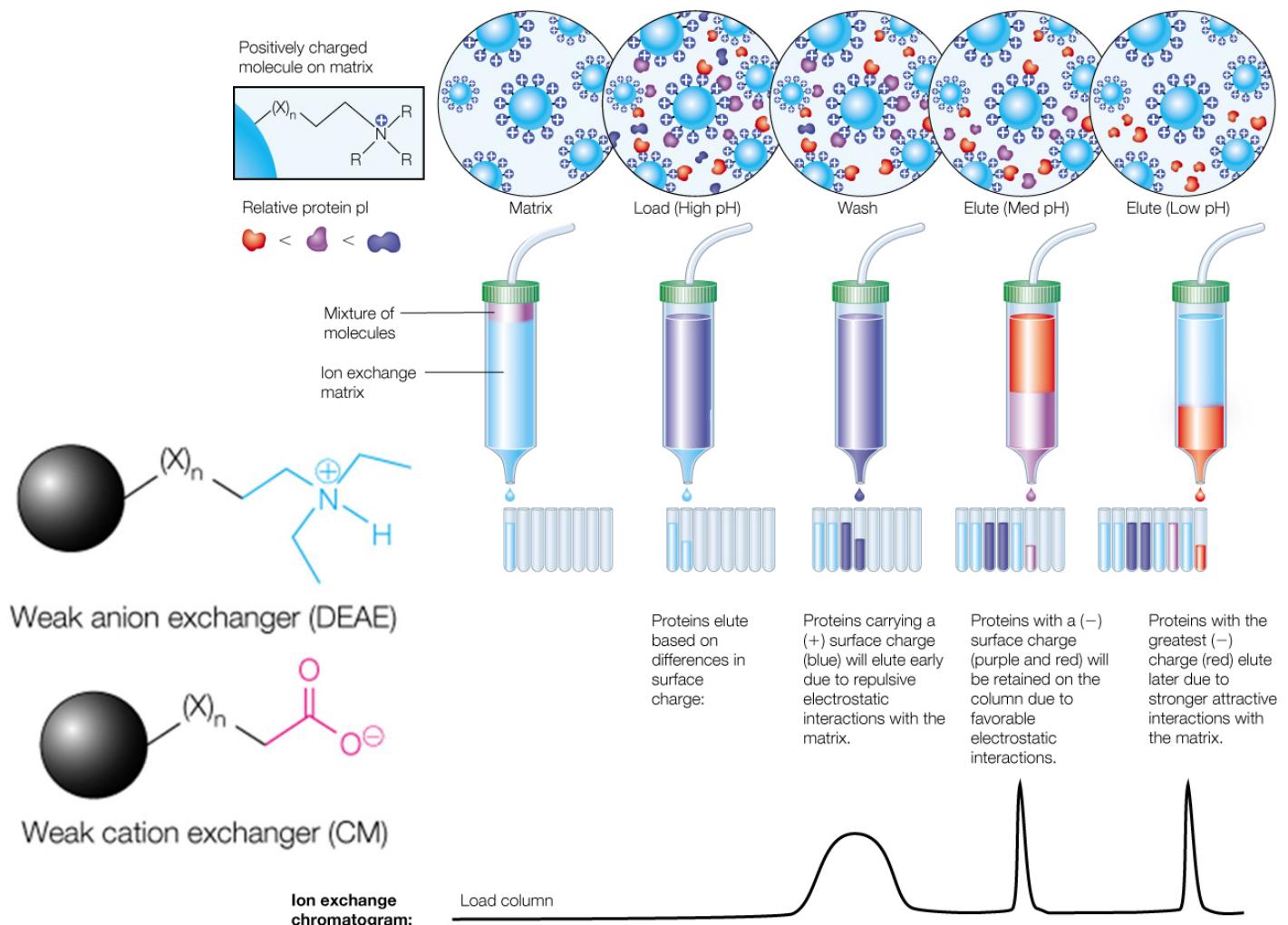


FIGURE 5A.6 An overview of ion exchange chromatography.

Polyacrylamide gel electrophoresis

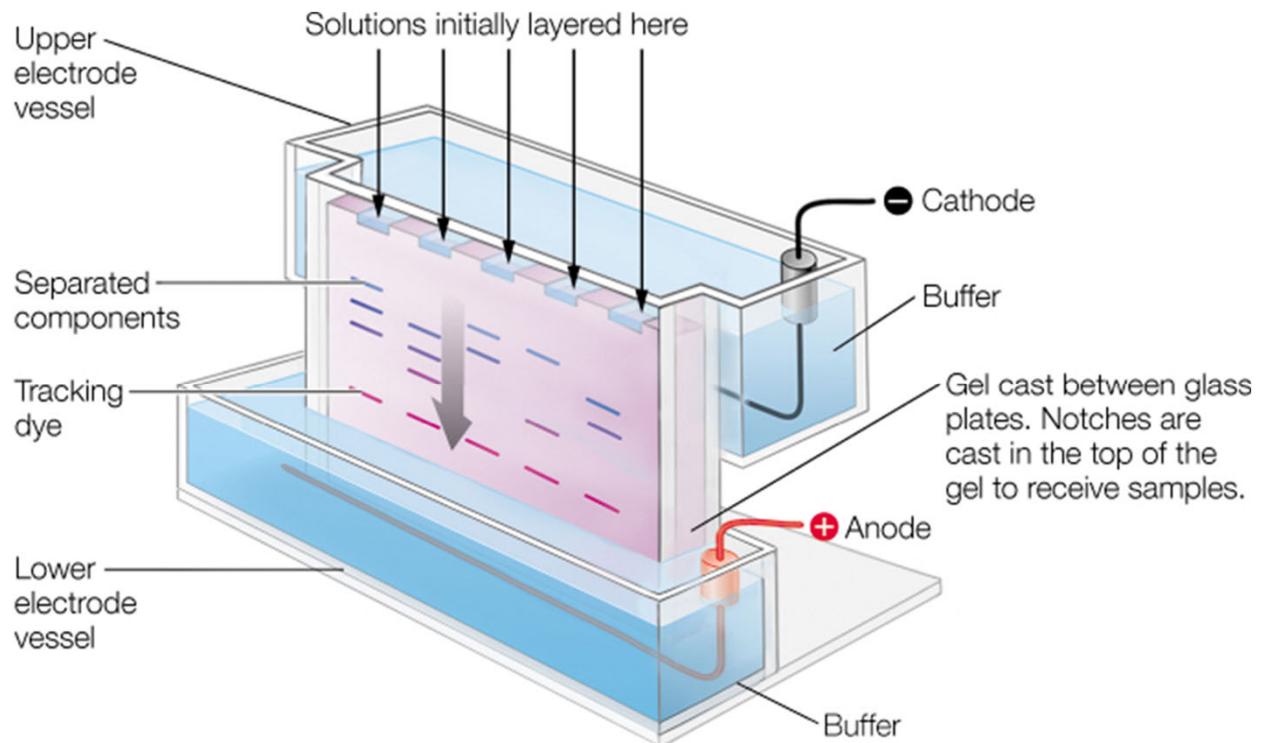


FIGURE 2A.2 Gel electrophoresis.

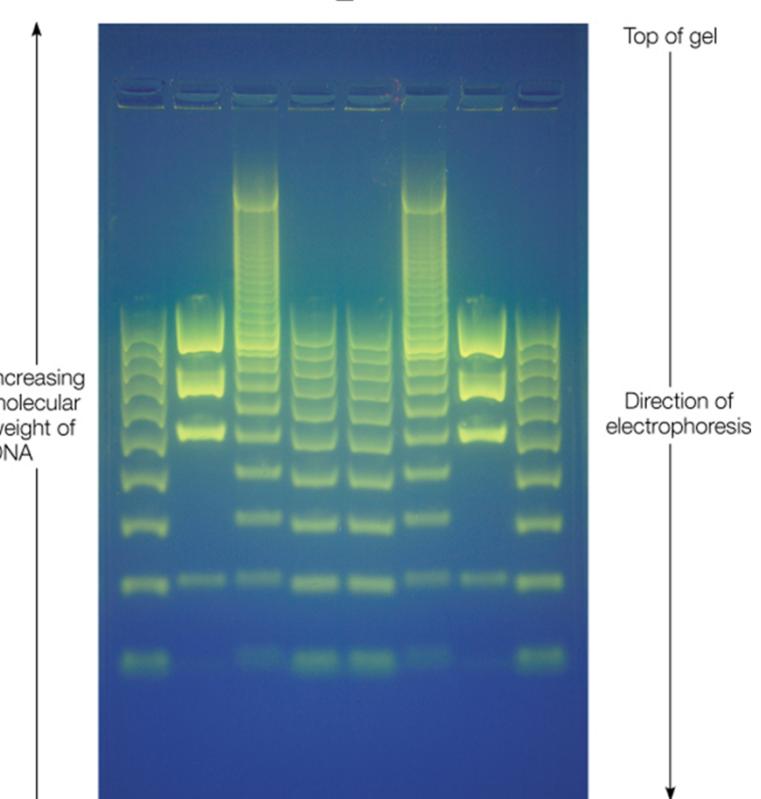


FIGURE 2A.3 ⁺Gel showing separation of DNA fragments.

Isoelectric Focussing

- While agarose and polyacrylamide gel electrophoreses (as used in biochemical laboratories) separate macroions based on their sizes (molecular weights), isoelectric focusing separates macroions based on their surface charge-associated isoelectric point (pl)
 - Macroions move in the electric field until the pH is reached where they have no net charge

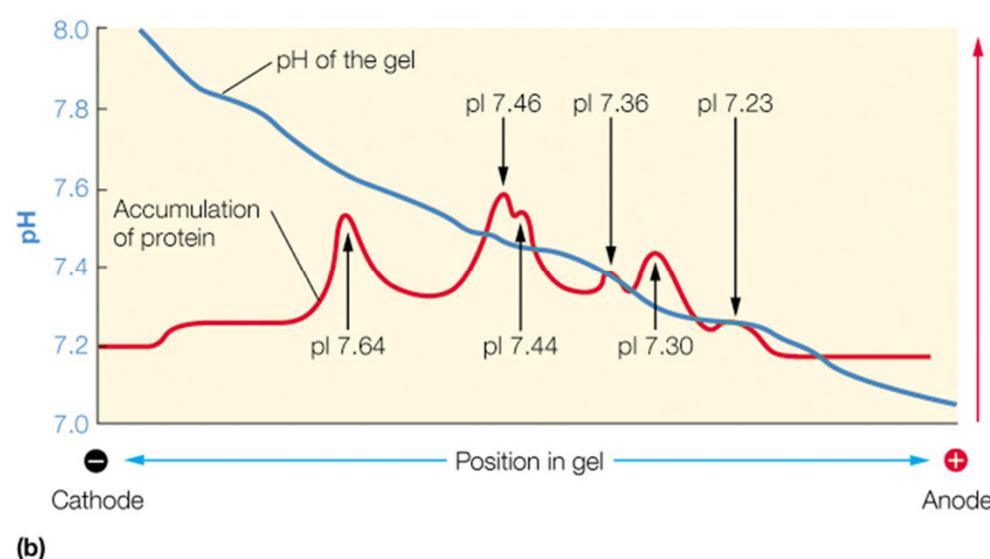
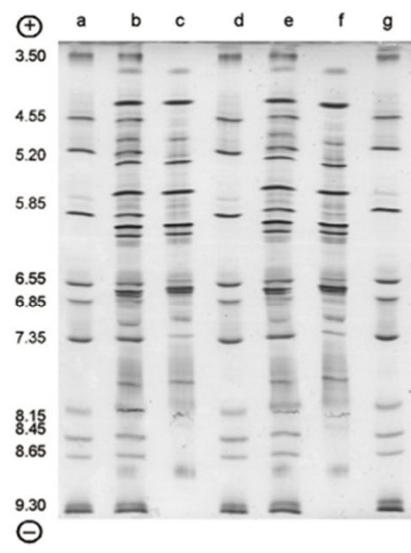


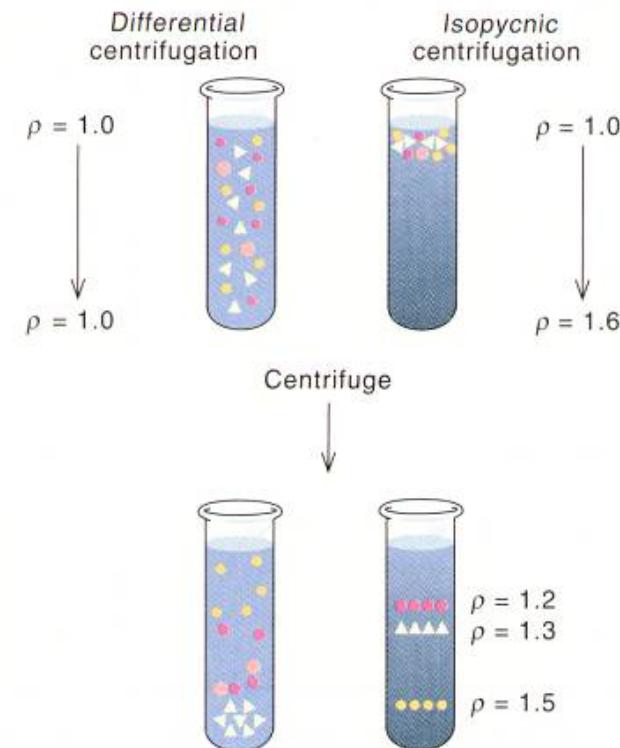
FIGURE 2A.4 Isoelectric focusing of proteins.

- pH gradient used from 3.5 (anode end) to 9.3 (cathode end).
- Schematic of where proteins with indicated pl values (peaks in red) would appear in the pH gradient (in blue) gel.

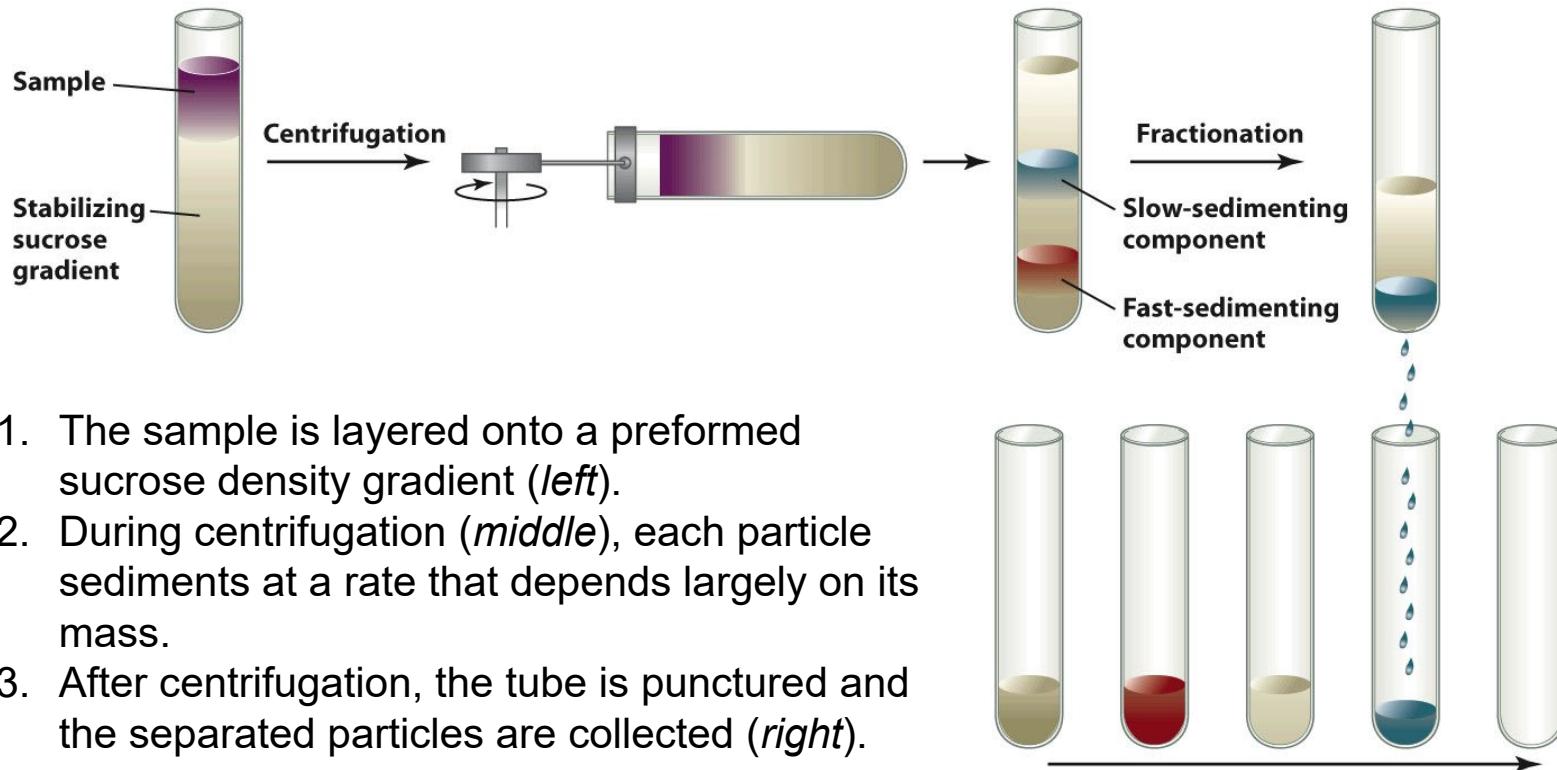


Centrifugation as a separation method

- Differential
 - constant density
 - size dependent
 - time and g force dependent
- Analytical – using an isopycnic solution (with a density gradient)
 - Separation based on density
 - components move to regions of similar density
 - *aka* ultracentrifugation



Ultracentrifugation



1. The sample is layered onto a preformed sucrose density gradient (*left*).
2. During centrifugation (*middle*), each particle sediments at a rate that depends largely on its mass.
3. After centrifugation, the tube is punctured and the separated particles are collected (*right*).

Ultrafiltration membranes are used in dialysis

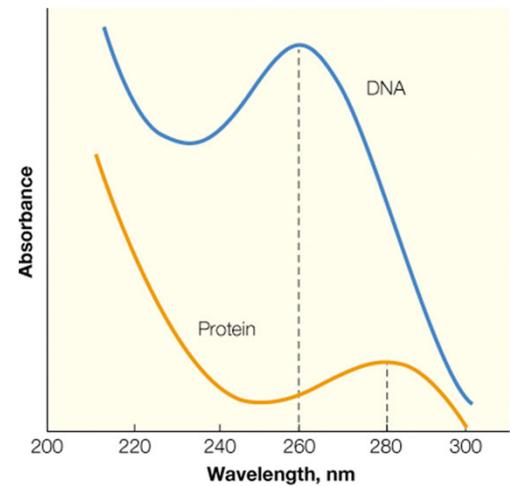
Chromatography: “colour writing” for separation

In general, chromatography is used to separate a mixture of substances, dissolved in a liquid (the “mobile” phase), based on their attraction or affinity for a solid (the “stationary” phase) – originally used to separate coloured substances, hence the name.

1. Paper chromatography uses filter paper as stationary phase and a mixture of solvent and buffer as the mobile phase (**Prac 5 demo**).

2. Column chromatography uses porous substances as stationary phase (called the matrix). There are different column types depending on the nature of substances to be separated.

- Presence of proteins can be monitored by collecting fractions of the mobile phase and determining the protein concentration by **absorption spectrophotometry at 280 nm**.
- Plotting absorbance at 280 nm gives graph (chromatogram) with **peaks for each protein** separated by the column.



Column Chromatography

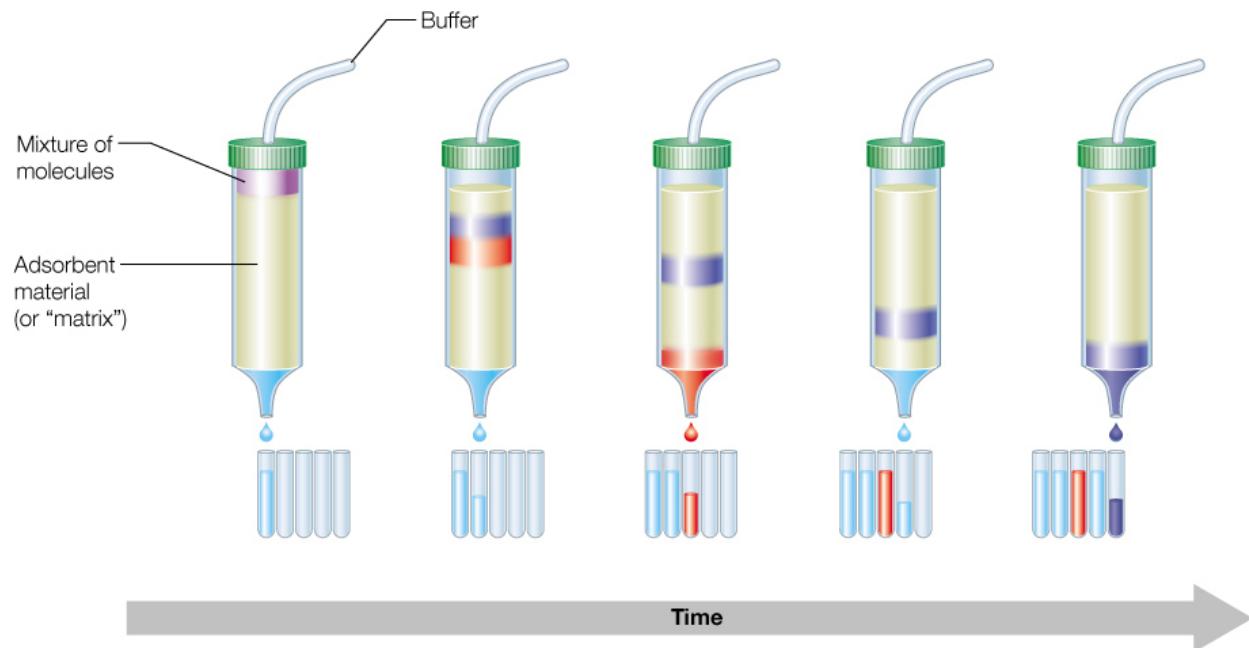


FIGURE 5A.2 The principle of column chromatography.

- A **mixture** of (cellular) proteins is separated as a result of differential interactions with the column **matrix**.
- The **buffer elutes** (or moves) the components of the mixture
- The more a protein interacts with the column material or **matrix**, the later it will elute from the column

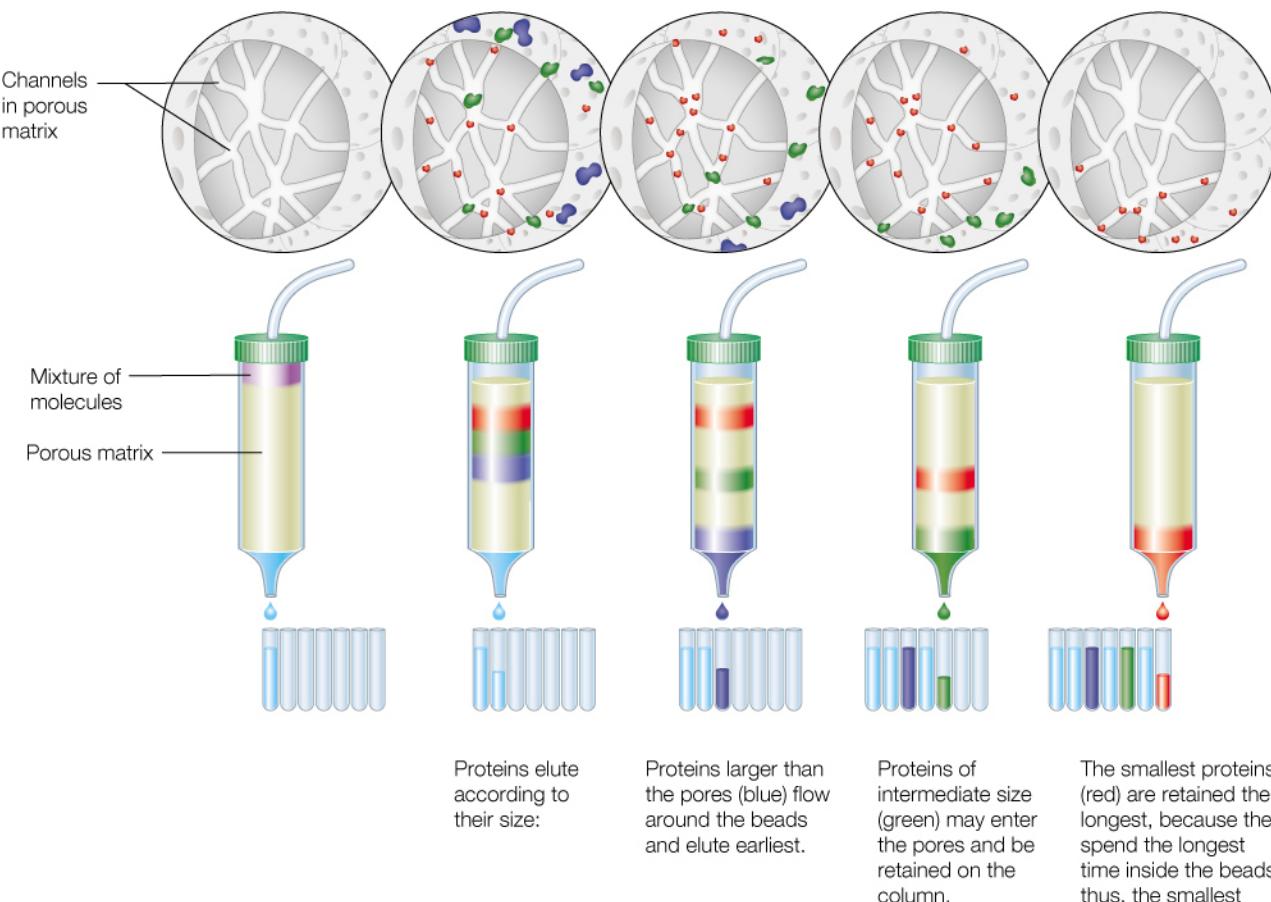


FIGURE 5A.7 An overview of size exclusion chromatography.

Size Exclusion Chromatography

- Separation of proteins is based on the apparent sizes of the molecules
- Larger proteins elute earlier because they are excluded from the interior volume of the chromatography matrix
- Smaller proteins elute later because they are retained within the pores of the chromatography matrix, resulting in a longer residence time on the column
- Prac 4 Theory**

Affinity Chromatography

- Selective adsorption of a protein to a natural or synthetic ligand, usually a substrate or inhibitor
 - The matrix contains the covalently bound ligand
- The protein interacts strongly with the **affinity matrix** and will elute last
- For recombinant proteins, a short His tag of 6 His residues is added in, which will bind preferentially to metal ions in an immobilized metal affinity chromatography (**IMAC**) column.

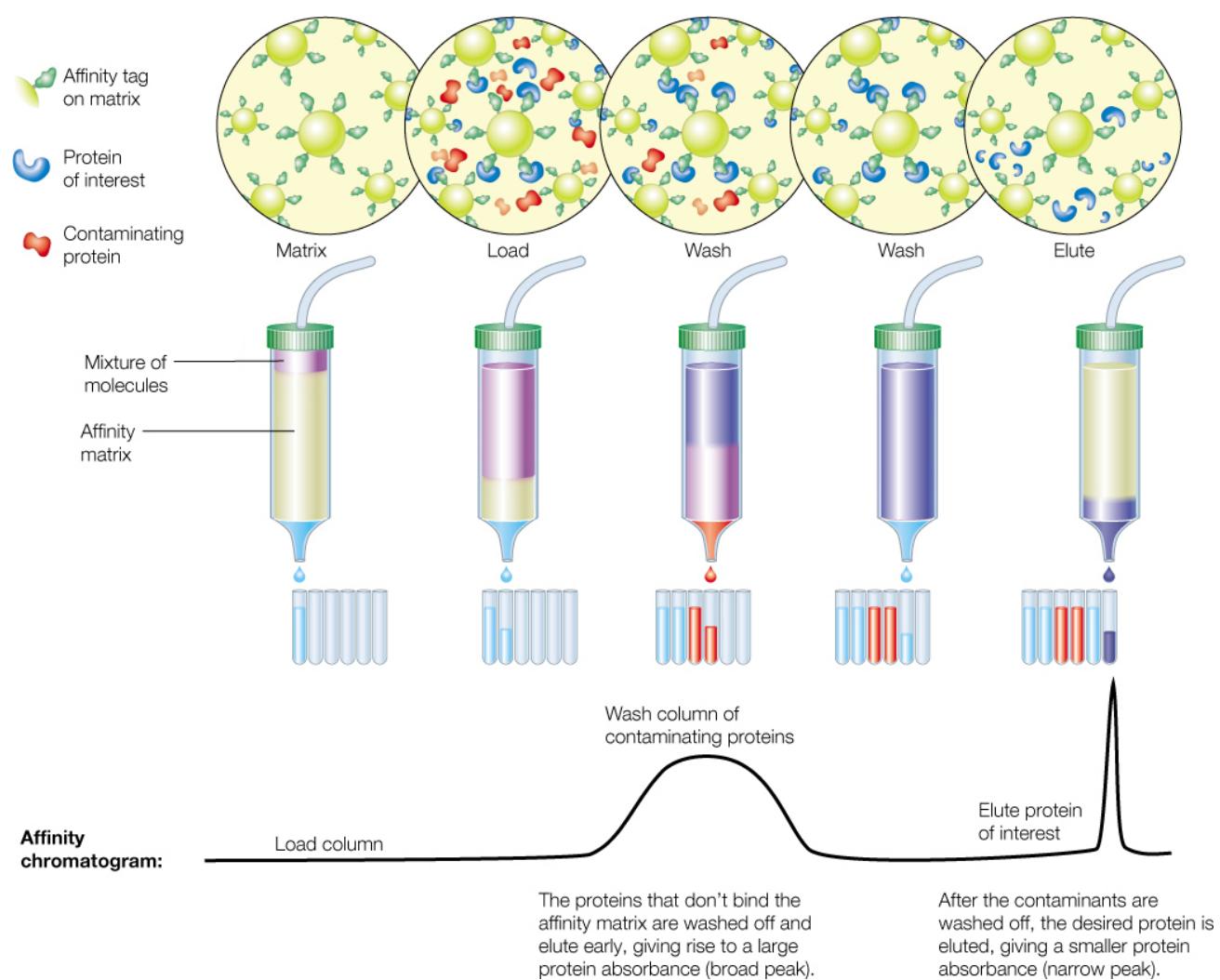


FIGURE 5A.3 An overview of affinity chromatography.

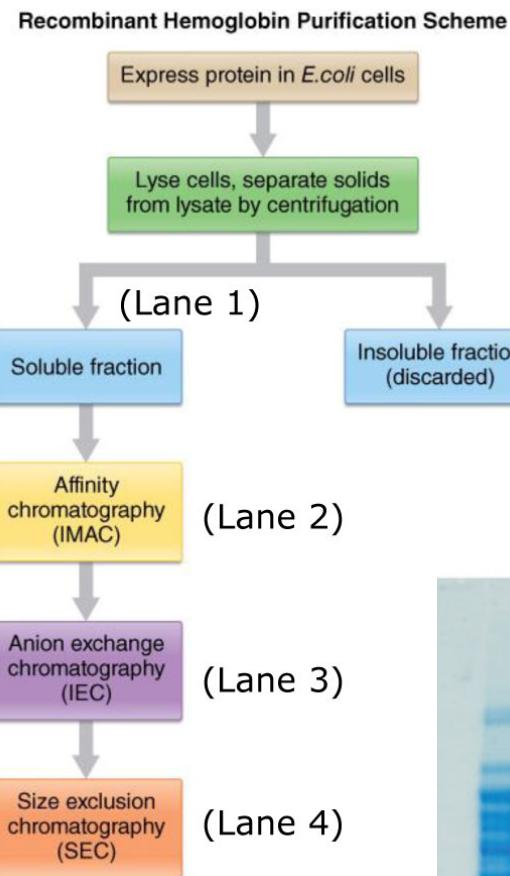


FIGURE 5A.8 Flowchart for the purification of recombinant hemoglobin.



MACQUARIE
University

Example: Purifying a Recombinant Hemoglobin Mutant

Haemoglobin (Hb) has two polypeptide chains: α -globin and β -globin. This mutant has an 8-amino acid insertion in the β -globin sequence. After protein production, the cells are lysed and the contents of the cytoplasm released into a buffered solution.

Centrifugation: removes the insoluble material (e.g., membranes and precipitated protein aggregates) from soluble fraction.

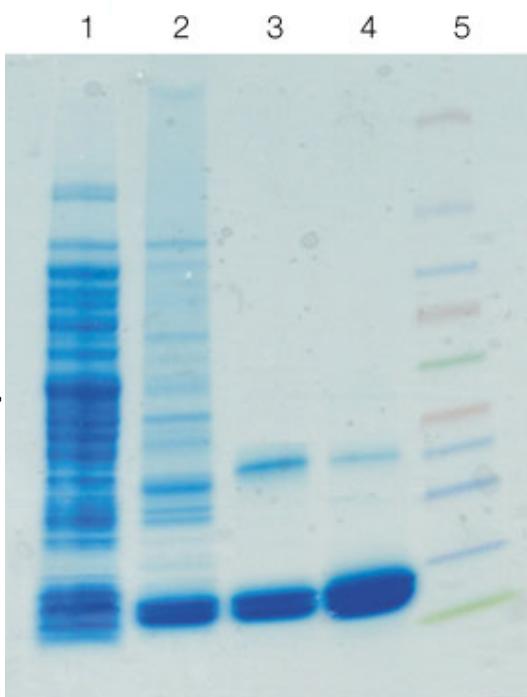
Lane 1: The resulting supernatant is a complex mixture of nucleic acids and proteins.

Lane 2: an IMAC purification step significantly purifies the mutant Hb.

Lane 3: several contaminants are removed by anion exchange on Q resin.

Lane 4: Finally, SEC is used to separate the Hb from the contaminant at 35 kDa. Lane 4 is overloaded in order to detect the presence of impurities in the final preparation.

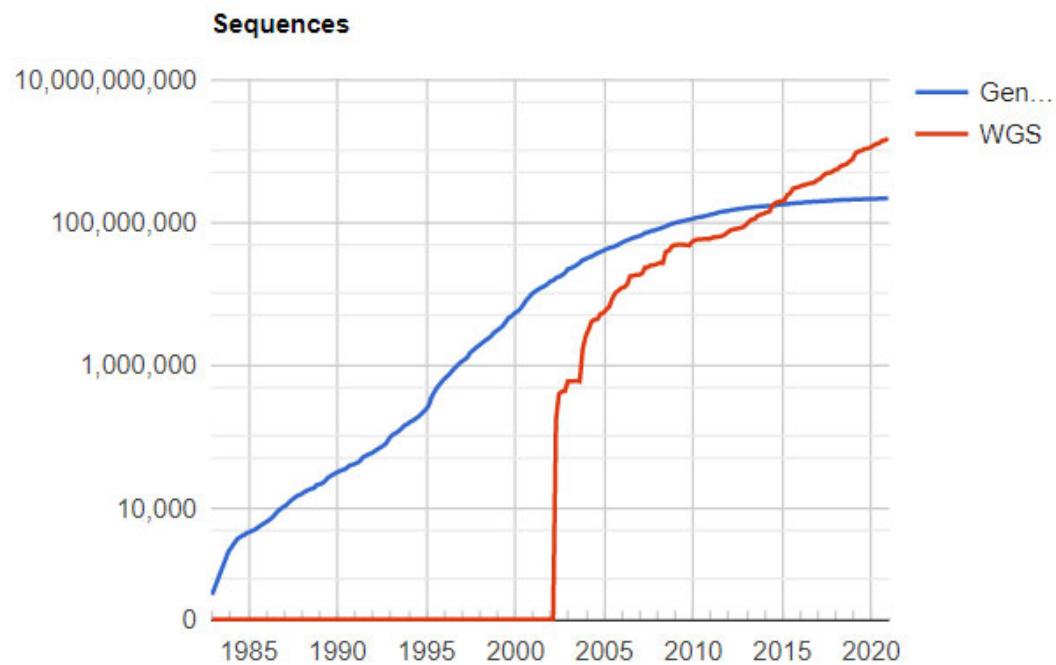
(Lane 5 contains a series of known mass standards)



17 kDa In this case, the hemoglobin mutant is greater than 95% of the total protein in the purified sample.

From Gene Sequence to Protein Function

- Genes code for proteins and
- Late 1970s: recombinant DNA technology became available to express many copies of a single gene which then lead to protein expression.
 - New field of Systems Biology
 - Millions of DNA sequences now deposited in databanks: GenBank now has 221,467,827 sequences
 - Whole genome shotgun (WGS) sequencing has led to 1,517,995,689 sequences! (as of Dec. 2020)



However, Biochemistry is interested in what these sequences do! i.e. their function!

Using sequences to find protein function

- Bioinformatics to the rescue!

- Gene sequences can be translated on a computer and then matched against these millions of database sequences to pick up functional clues
- If a new sequence matches the human myoglobin, it is possibly a globin?
- Let's see how well myoglobin and haemoglobin match up (they both contain ***heme*** – a ligand)

Score = 30.8 bits (68), **Expect = 6e-06**, Method: Compositional matrix adjust.

Identities = 32/133 (25%), Positives = 48/133 (37%), Gaps = 40/133 (30%)

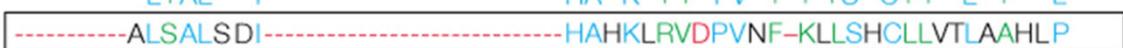
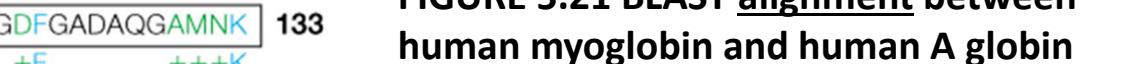
Human Mb	2		61
Human α	2		46
Human Mb	62		120
Human α	47		82
Human Mb	121		133
Human α	83		95

FIGURE 5.21 BLAST alignment between human myoglobin and human A globin sequences.



How similar are 2 sequences?

- Identities: 25% i.e. same residues
- Positives: this includes chemically similar side chains and so the score goes up to 37%
- Gaps: 30% - not so good, but maybe the sequences diverged a lot during evolution
- Expect = 6e-06: this is the expectation value for finding this match by chance: probability score and the smaller it is, the better.

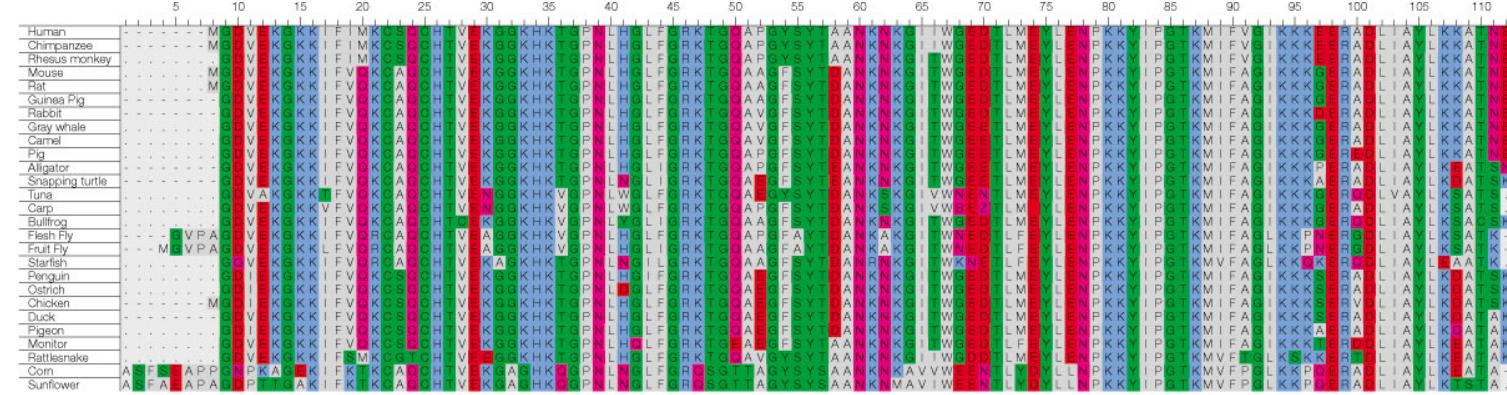
Score = 30.8 bits (68), **Expect = 6e-06**, Method: Compositional matrix adjust.
Identities = 32/133 (25%), Positives = 48/133 (37%), Gaps = 40/133 (30%)

Human Mb	2	LSDGEWQLV LNV WGKV EADIPG HGQEVL RLF KGH PET LEK FDK F KHL KSE DEM KASE DL	61
Human α	2	LS PAD KTN V KAA WGKV GA HAGE YGA EA LER MF LS F PTT KTY FPHF -----	46
Human Mb	62	KKHGATV LT ALGG I LKKKGH HEAEIK PLA QSH A TKHKI - PV KYLE FISE CI IQ VLQ SKHP	120
Human α	47	L + AL I HA K ++ PV + + + S C ++ L + L ----- ALSA L SDI ----- HA HKL RV DP VN F KLL SH C LL VT LA AH LP	82
Human Mb	121	GDF GADA QGAM NK + F + + + K	133
Human α	83	A EFT PAV H AS LDK	95

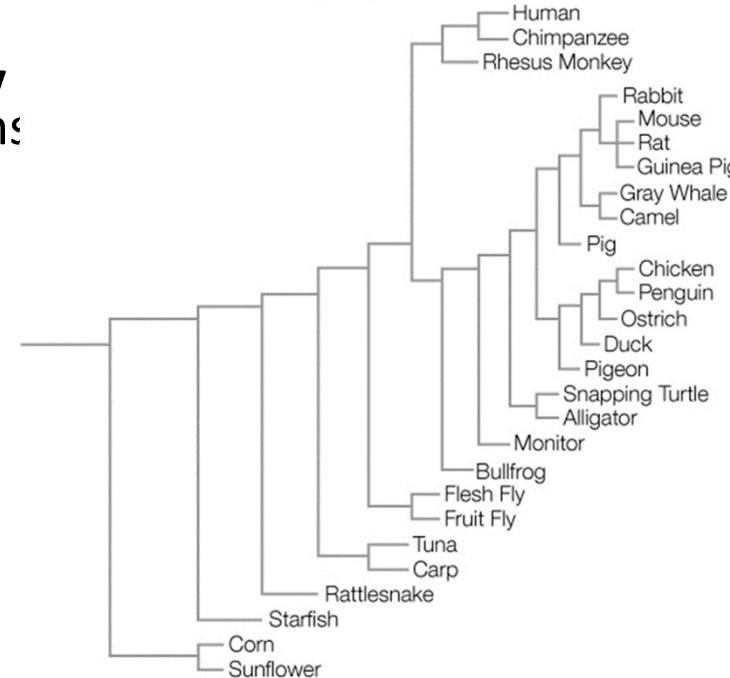
1. So 25% is good and if two sequences are 25% identical, they are considered homologous.
 - Likely to have the same structure, i.e. folded shape and
 - The same function!
2. The probability score helps us to find homologous sequences (< 1e-05).

Homology search

- Matching to an entire database finds homologous sequences from different organisms
- Sequence alignment of cytochrome c, a **mitochondrial respiratory protein**, from 27 organisms
- The more similar two homologous protein sequences are, the more closely they are related evolutionarily
- Lining them up (i.e. their alignment) can help us trace evolutionary history of a protein/gene.



(a) Alignment of cytochrome c sequences from 27 organisms, where hydrophobic amino acids are highlighted in gray, basic amino acids in blue, acidic amino acids in red, and polar uncharged amino acids in green (except for Asn and Gln, which are magenta).



(b) A phylogenetic tree for the sequences shown in (a). Branches indicate points of evolutionary divergence based on differences in the amino acid sequences of the aligned proteins.

FIGURE 5.22 Sequence alignment and a phylogenetic tree for cytochromes c from different organisms.

- Mitochondrial respiration uses homologous proteins in multiple organisms!
- Sequences that are very similar are from closely related organisms – lots of use in biochemistry, biotechnology and of course, biology!
- The function of a protein is conserved even though they look slightly different!

Consensus Sequences from an alignment

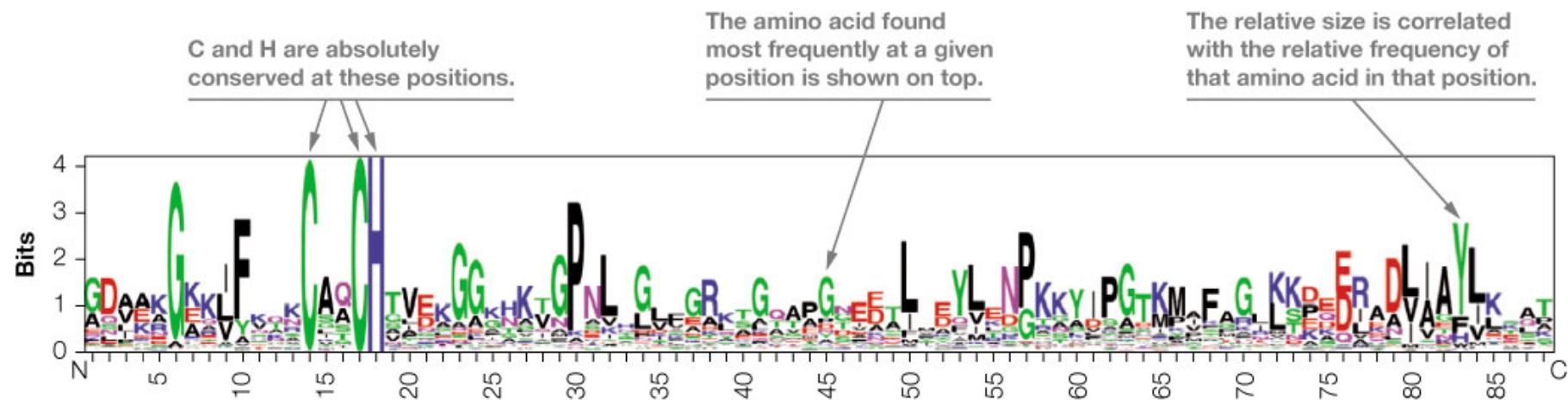
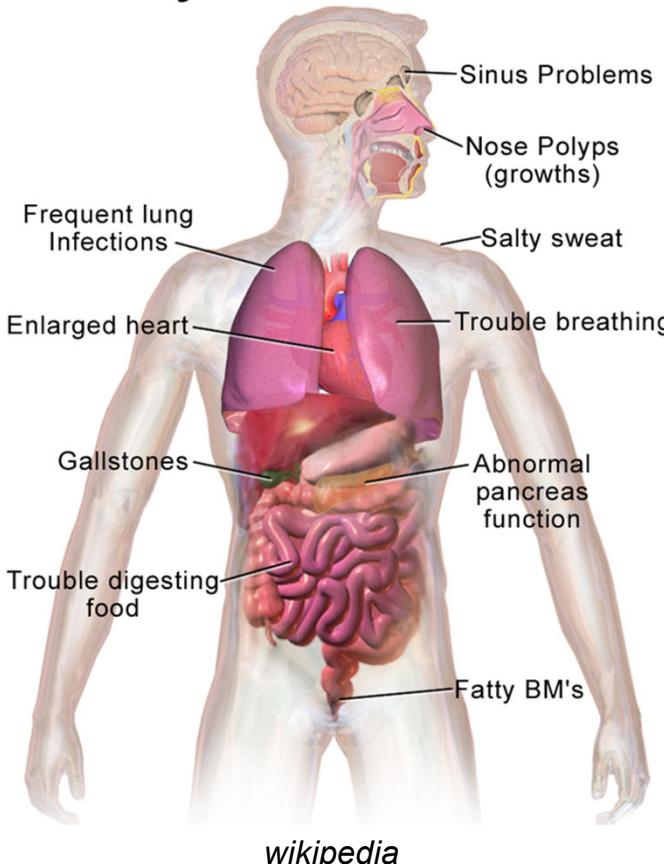


FIGURE 5.23 Sequence logo from the alignment of 412 sequences from the cytochrome c family.

- A consensus sequence (here for cytochrome c) allows us to identify the most highly conserved amino acid residues – these may be in separate clusters!
- These are critical for function and mutating them might completely destroy biological activity
- Naturally occurring mutations or variations have led to the discovery of genetic diseases

Example: Cystic Fibrosis (CF)

Health Problems with Cystic Fibrosis



- Genetic disorder that affects mostly the lungs but also the pancreas, liver, kidneys and intestine. Long-term issues include difficulty in breathing and coughing up sputum as a result of frequent lung infections.
- Recognized as a specific disease by Dorothy Andersen in 1938, with descriptions that fit the condition occurring at least as far back as 1595!
- CF is caused by a mutation in the gene cystic fibrosis transmembrane conductance regulator (CFTR). The most common mutation, **ΔF508**, is a deletion (Δ signifying deletion) of three nucleotides that results in a loss of the amino acid phenylalanine (F) at the 508th position on the protein.
- $\Delta F508$ -CFTR, which occurs in >75% of patients globally, creates a protein that does not fold normally and is degraded by the cell.



Summary

- Proteins are polymers of α -amino acids
 - twenty common amino acids (and two rare), distinguished by their side chains (R groups), are incorporated into proteins
- Proteins are produced by condensation of amino acids via peptide bond formation
- The unique defined sequence of amino acids constitutes the primary structure of proteins
- In cells, genes are transcribed into messenger RNA (mRNA), which is then translated into a polypeptide strand at the ribosomes

Summary

- Using recombinant expression techniques, proteins can be produced in non-native organisms at very high yields
 - Genes of interest are incorporated into these non-native organisms, and then transcribed and translated by these organisms
- Depending on the certain properties of the produced proteins, they can be purified at high purity by various chromatographic techniques, and the protein masses and sequences can be determined.

Reminders

- Pracs and Tutorials start this week
- Classes 1-3: Pracs this week
 - Pl. do your pre-lab
 - Pl. come to the lab on time, with your lab coat and fully enclosed shoes!
- Classes 4-6: Tutorials this week
 - Class 7: online tutorial
- Please sign up at the text book site from iLearn
 - *Tutorial 1 Quiz for all classes on the textbook site.*