MACQUARIE University

Tutorial 2

Proteins: Sequencing, Separation and Structure

BMOL2201/6201

Tutorial 2 Aims

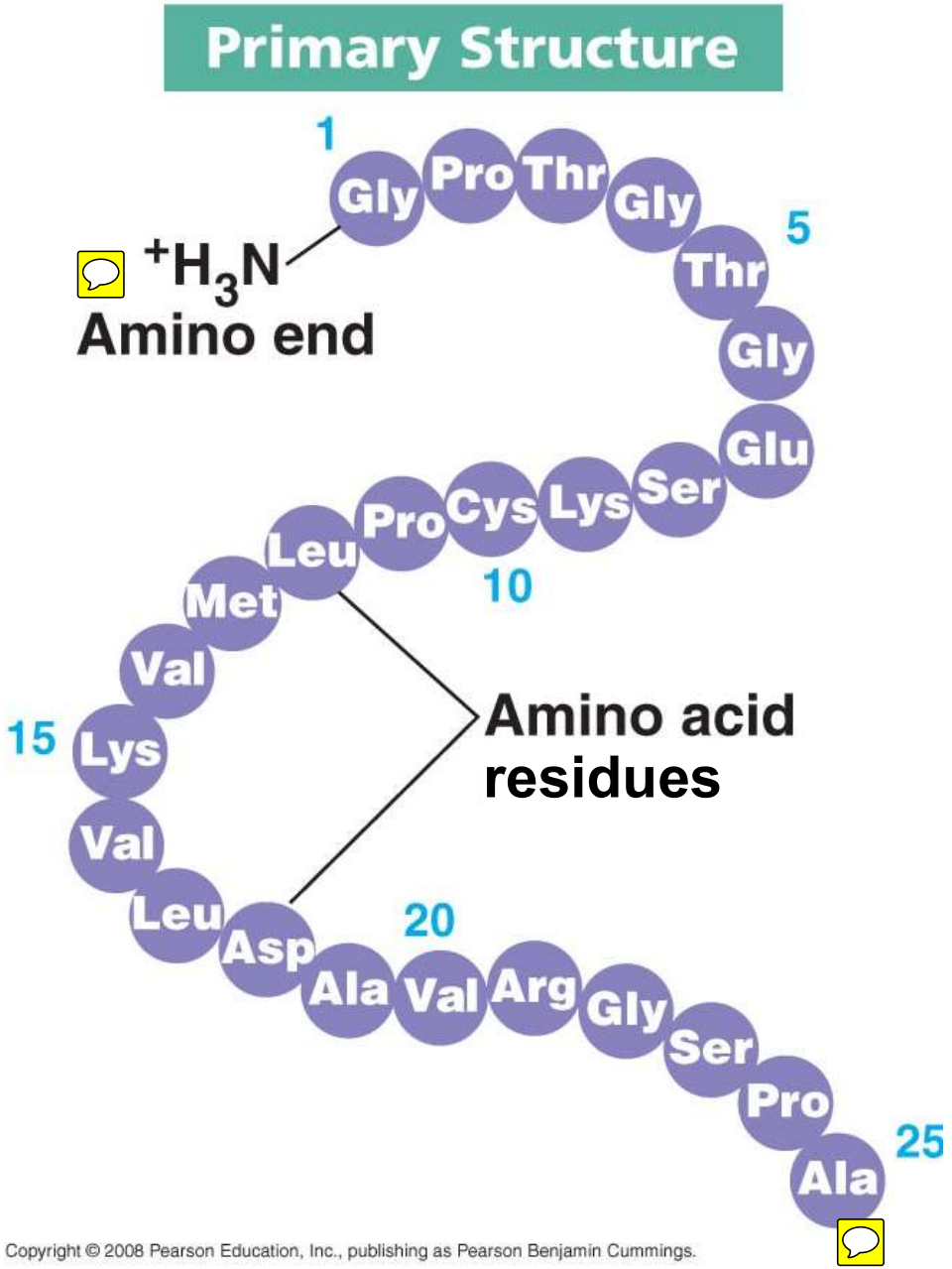
- Understand what the **primary structure** of a protein is, and how to find it by **sequencing**
- Identify what methods biochemists use to **purify** proteins, and understand the properties behind each method
- Describe the four main levels of **protein structure** including how they affect protein function

What makes a protein a protein?

- Proteins are polymers, made up of sequences of **amino acids**
- Proteins are the molecular machines of the cell and are used to carry out almost every job
- Questions we'd like to answer:
  - What makes Protein A different from Protein B?
  - How can we work this out in the lab?
  - How do two proteins made up of the same 20 amino acids do completely different things?

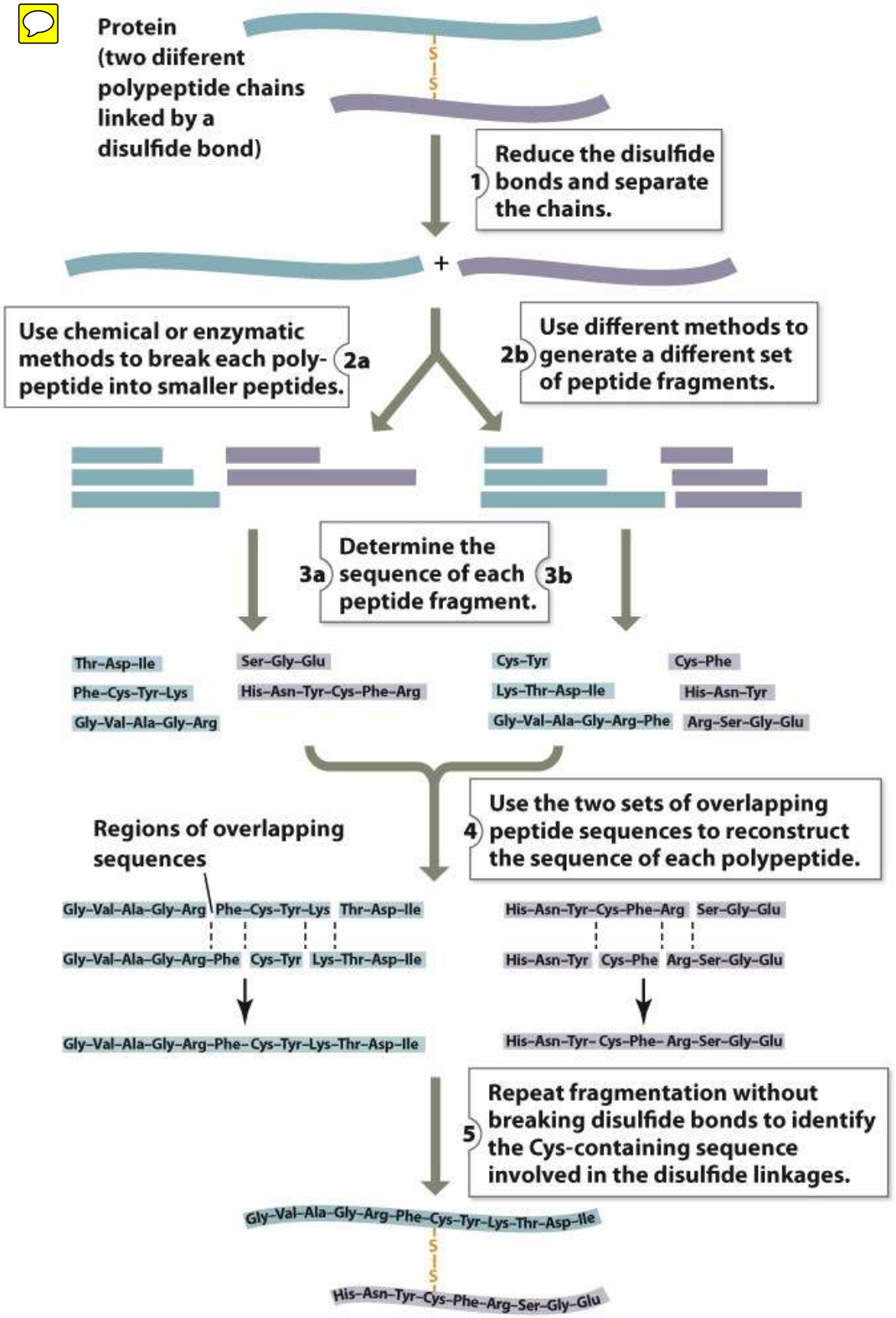
Basic level: primary structure of a protein

- Good place to start: the **type and order** of amino acids in a protein
- Direction is N (amino end) to C (carboxyl end)
- Usually call amino acids “residues” when they’re in a protein



Protein Sequencing Workflow

1. Get rid of disulfide bonds
2. Fragment peptide in at least two different ways
3. Sequence peptide fragments
4. Piece protein back together from fragments by overlapping
5. Repeat steps, leaving disulfide bonds in to work out where they go



MACQUARIE University

Amino and carboxyl terminal identification:

- Useful to have methods to remove the VERY FIRST or VERY LAST residue of a polypeptide – the amino terminus or carboxyl terminus

- Amino terminus: two popular options:
  - Dinitrofluorobenzene (DNFB) a.k.a. the Sanger method, OR
  - Dansyl Chloride method

- Carboxyl terminus: one popular option:
  - Carboxypeptidase

- Once the amino acid has been ripped off, we can use analytical chemistry (methods such as HPLC) to determine what it is

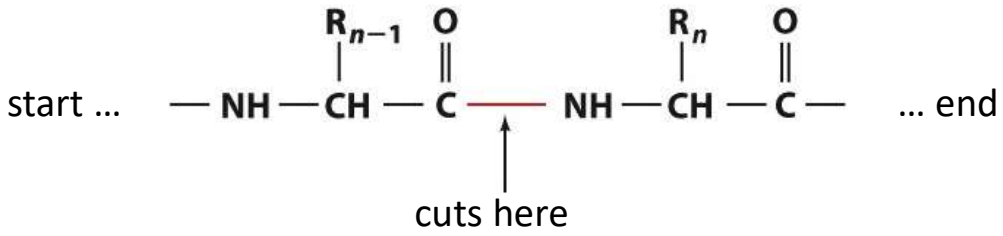
Proteases – how do we fragment peptides?

- Majority of proteases work by cutting AFTER a specific type of residue (usually the most popular choices):
  - **Trypsin**: cuts AFTER lysine and arginine (i.e. basic residues)
  - **Chymotrypsin**: cuts AFTER phenylalanine, tyrosine and tryptophan (i.e. residues with aromatic rings)
  - **Cyanogen bromide**: cuts AFTER methionine (usually starts proteins!)
- Note: if there are multiple sites, protease can cut at **any or ALL**, so need to look at **all the different combinations**

- Also other proteases, some of which work by cutting BEFORE a residue – see next slide for a reference table

- Need to remember the most popular three (above) at least

Endopeptidase Specificity



Enzyme	Specificity	Comments
Trypsin **	**AFTER** Lys **(K)**, Arg **(R)** (**basic** residues)	Doesn't work if next residue is proline
Chymotrypsin **	**AFTER** Phe **(F)**, Trp **(W)**, Tyr **(Y)** (**aromatic rings**)	Doesn't work if next residue is proline
Elastase	AFTER Ala, Gly, Ser, Val (**small neutral** residues)	Doesn't work if next residue is proline
Thermolysin	BEFORE Ile, Met, Phe, Trp, Tyr, Val	Doesn't work if previous residue is proline
Pepsin	BEFORE Leu, Phe, Trp, Tyr	Doesn't work if previous residue is proline
Endopeptidase V8	BEFORE Glu	
Cyanogen bromide **	**AFTER** Met **(M)**	(Not an endopeptidase)

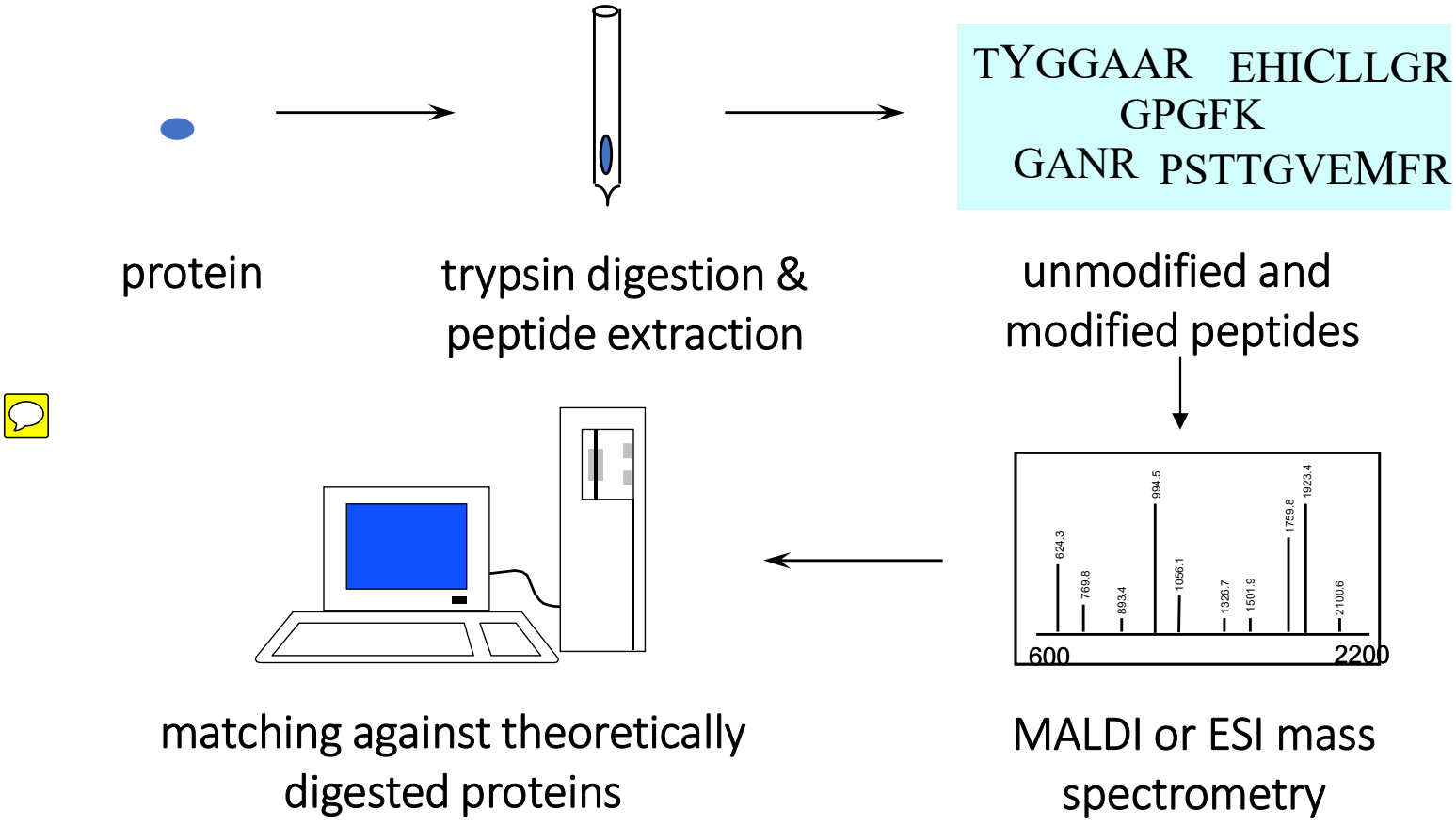
** most commonly used

Edman Degradation for N-terminal sequencing of peptides

- Sequencing method that takes one residue off at a time and determines its identity using analytical chemistry.
- Once first one is known, identify second, third, fourth etc.
- Good for short peptides, but not great for long polypeptides.
- Still used in the lab.

Peptide Masses from Mass Spectrometry

Mass spectrometers accurately measure peptide (fragment) masses:



- Fragments act as a “fingerprint” for a protein, which can be checked against an online database using sophisticated search algorithms.
- Best method currently available and faster than Edman.

Question 1: Protein Sequencing

(Example of short question for final assessment)

a) What is the primary structure of a protein? Does the order of amino acids matter?

b) I have a peptide with the following residues:

AGTPVKVGGFRAPAWGG

If I digest it with **trypsin**, how many fragments will I get?

c) What is the main method used to identify the fragments of proteins?

Protein purification methods

Methods are based on physical properties:

- **Size**
- **Charge**
- **Affinity**
- **Solubility**
- **Mass**

Remember the first listed method is the best one!

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

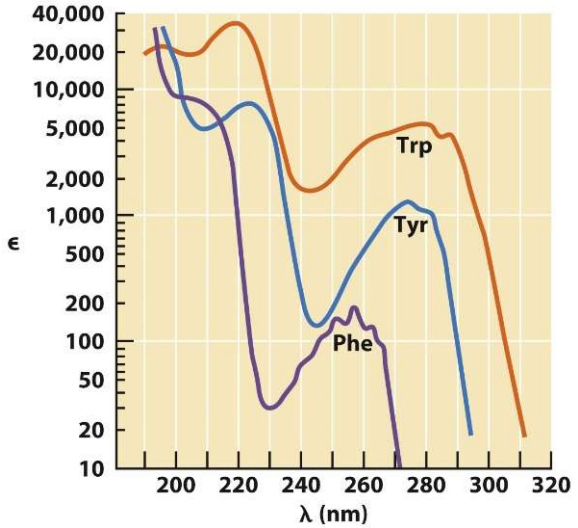
Chromatography: Overview

- Used to separate a mixture of substances, dissolved in a **liquid** (the "mobile" phase), based on their attraction or affinity to 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 monitored by collecting fractions of the mobile phase off the bottom, and determining the protein concentration by absorption spectrophotometry at 280 nm.



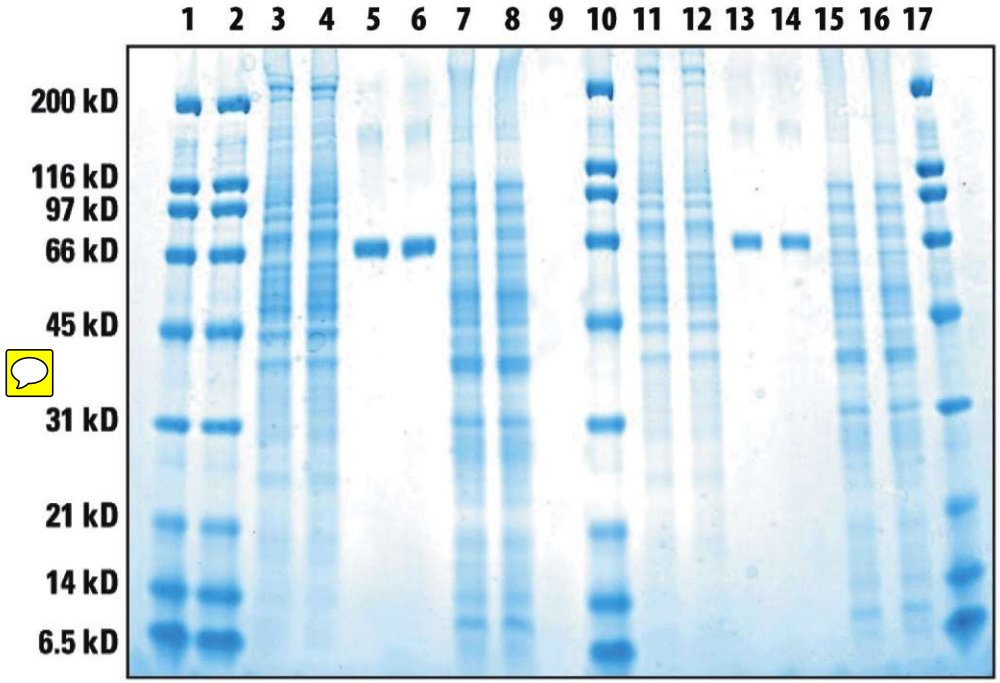
Column chromatography: In detail

- What solid matrix do we use to separate?
  - Porous gel beads for **size** separation
    - Called **gel filtration** or size exclusion chromatography
    - Will do this one in **Prac 4**!
    - **Small** molecules get TRAPPED by pores, meaning they go SLOWLY and come out **last** – **big** molecules speed past, not entering the pores, and come out **first**
  - Ion exchange for **charge** separation
    - Positive column if we want negatively charged proteins etc.
  - Specialised material for **affinity** separation
    - Groups are present on the solid matrix that specifically bind to the protein of interest such as antibody columns
  - Hydrophobic material for **hydrophobicity** separation (HIC)
    - Protein binds under high salt concentration and is eluted with low salt concentration

Electrophoresis: Separation based on charge **and** size

- Ions moving in an electric field
  - Usually **denatures** the protein since it is coated with a detergent (sodium dodecylsulfate, SDS), to have a uniform negative charge
- Usually called 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 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

SDS-PAGE: Example gel

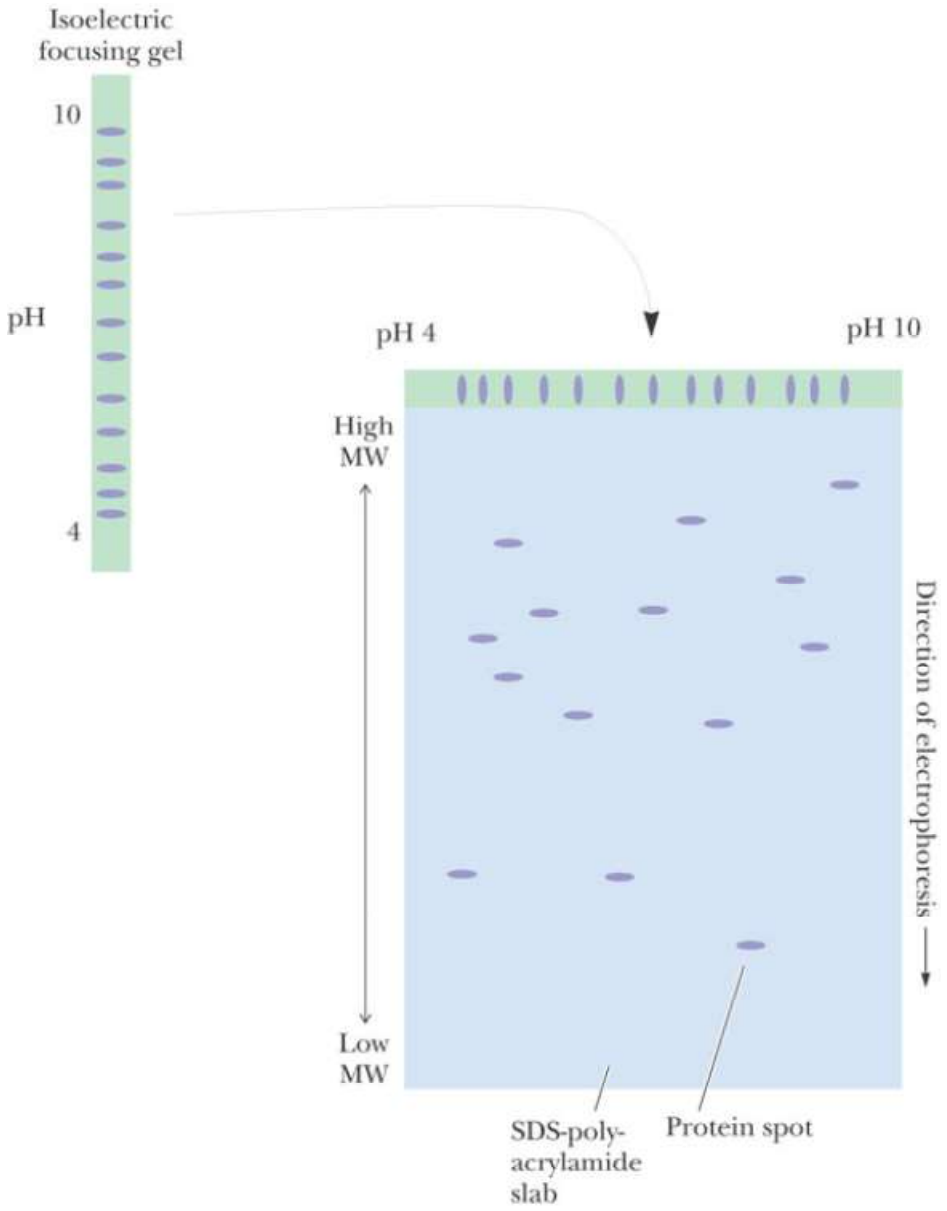


- Proteins separated from top to bottom in parallel lanes and stained
- Lanes 1, 2, 10 and 17 contain mass standards as shown (kind of a molecular ruler!)
- Lighter molecules move faster than heavier ones

Two-Dimensional (2D) Gel Electrophoresis

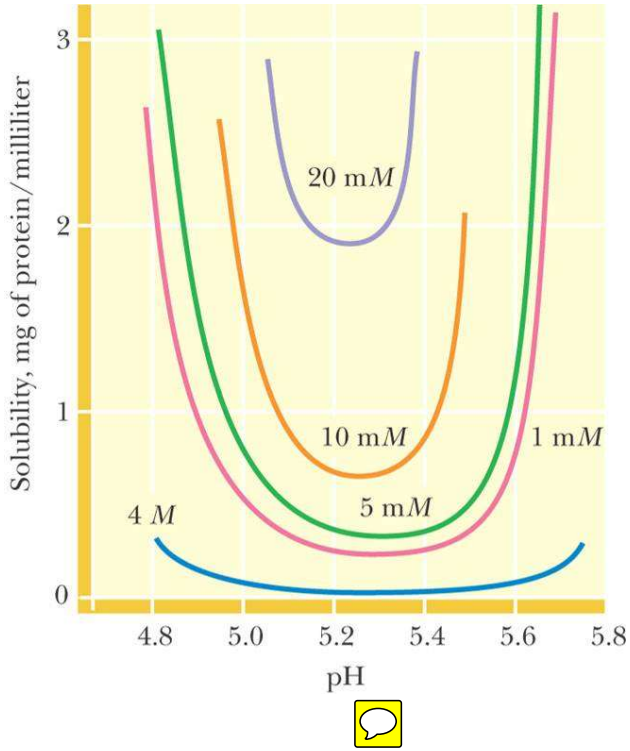
- Two-step process:

1. Separate based on pI using isoelectric focusing
2. Separate based on mass using electrophoresis



Protein solubility

- Proteins are **least soluble at their pI** (when the net charge is zero)
- 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



- **Addition of excessive salt** (> 1 M) precipitates the protein and is called **salting-out**
  - The salt removes all the available water and hence the protein precipitates
- **Prac 4**

Centrifugation: Two types

- **Differential centrifugation**
  - Regular type of centrifugation done in the lab – black-and-white split between dissolved and precipitated
  - Usually separate organelles from cytosol and/or membranes (and therefore membrane bound proteins) from soluble proteins

- **Analytical ultracentrifugation**
  - Proteins sediment (precipitate) out of solution at different rates based on their mass, size, and shape
  - Can use these to separate based on mass and size
  - **The smaller the particle and the less spherical its shape, the more slowly it sediments in a centrifuge**

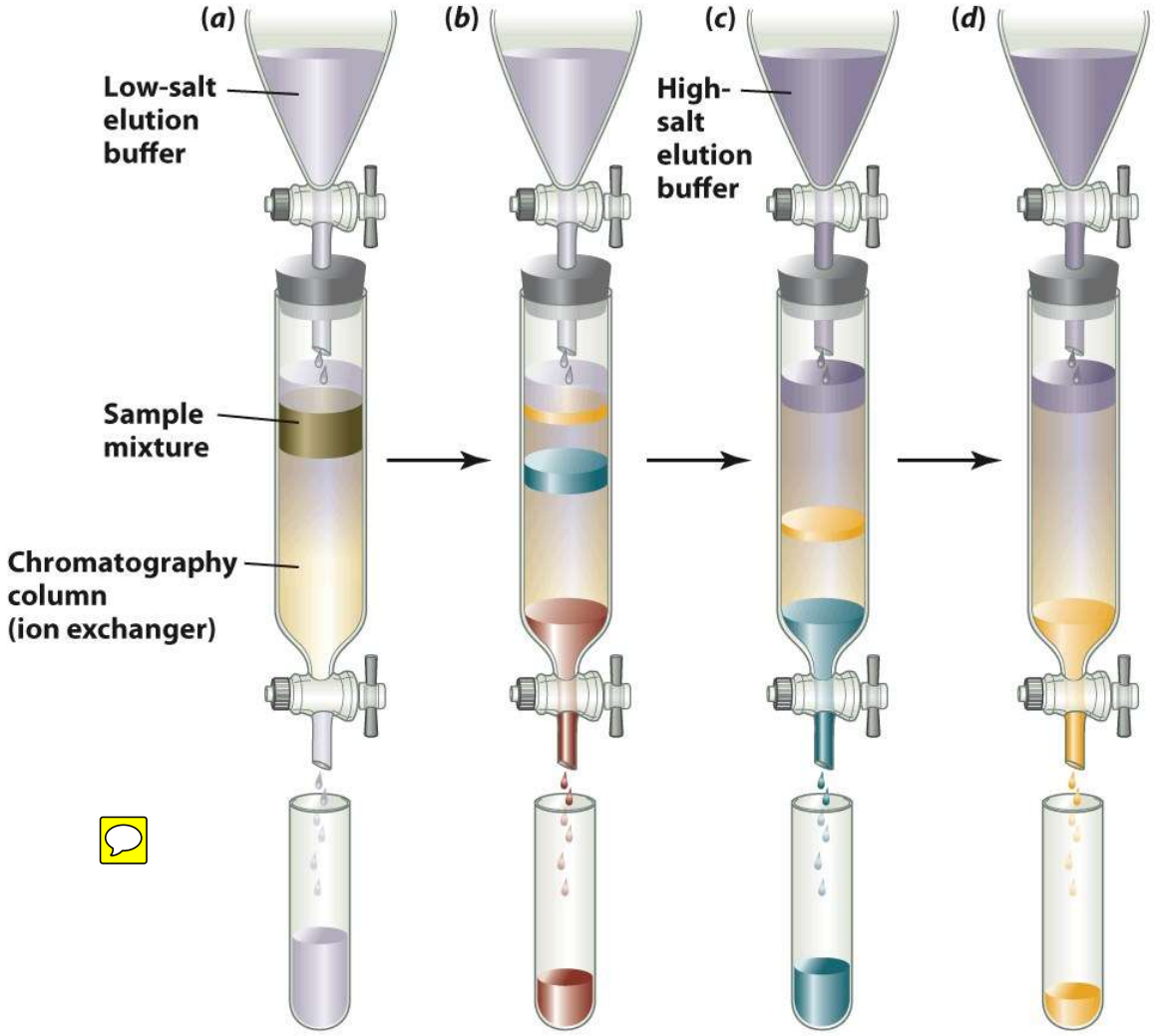
Question 2: Purification Methods

a) How does size exclusion chromatography (a.k.a. gel filtration) work?

b) If I have a big, bulky protein A and a short, small peptide B on the same gel filtration column, which will elute (come out) first?

c) I have two proteins with vastly different charges – which purification method could I use to separate them?

Ion Exchange Chromatography: separates cations and anions. Property used: ionic charge (Lecture 4)

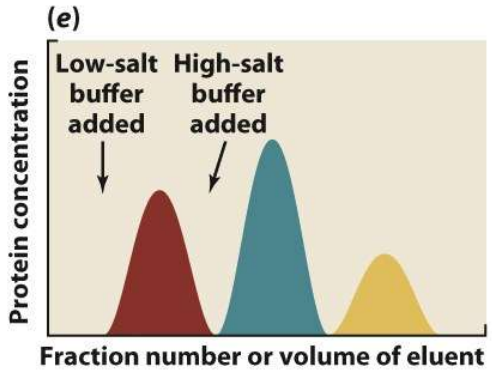


(a) A mixture of proteins dissolved in a buffer (eluant) is added to the top of the matrix in the column.

(b) During *elution*, the proteins separate into discrete bands due to their different affinities for the column. The first protein (*red*) has passed through the column and has been isolated as a separate fraction. The other proteins (*blue*, *yellow*) pass through the column more slowly.

(c) & (d) The salt concentration in the eluant is increased to elute the remaining proteins.

(e) The elution diagram of the protein mixture from the column.

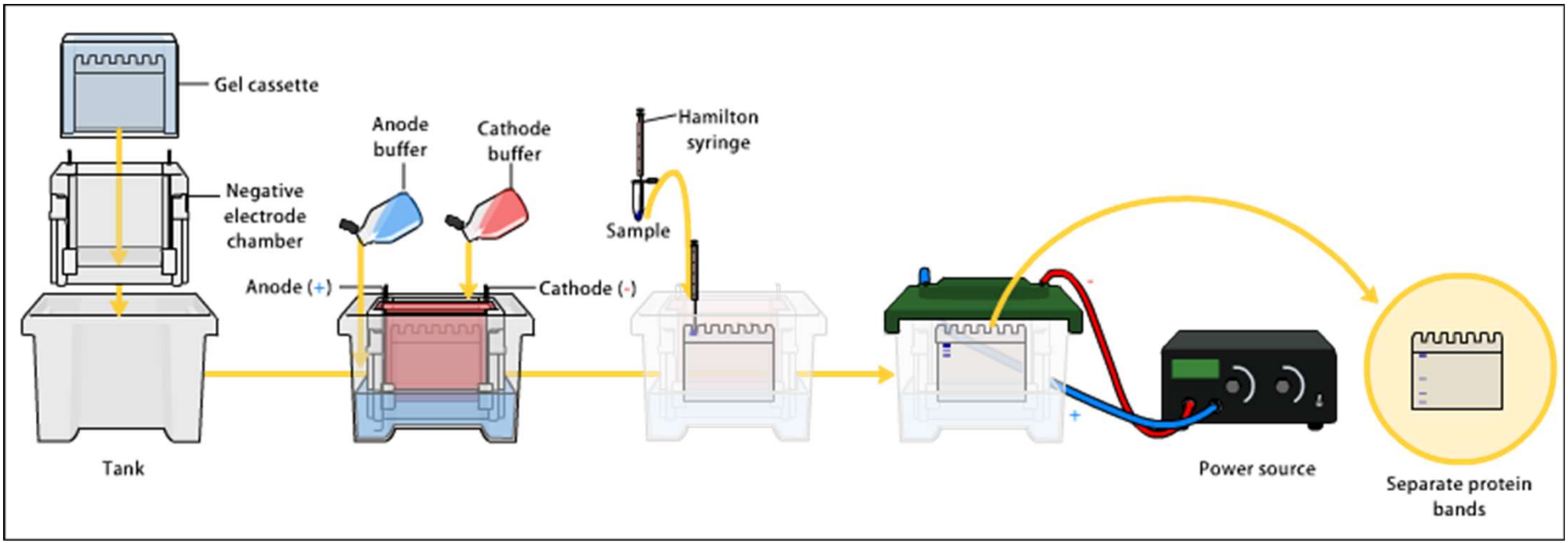


- The protein that binds better is released later. The first one to come out is the worst binder.

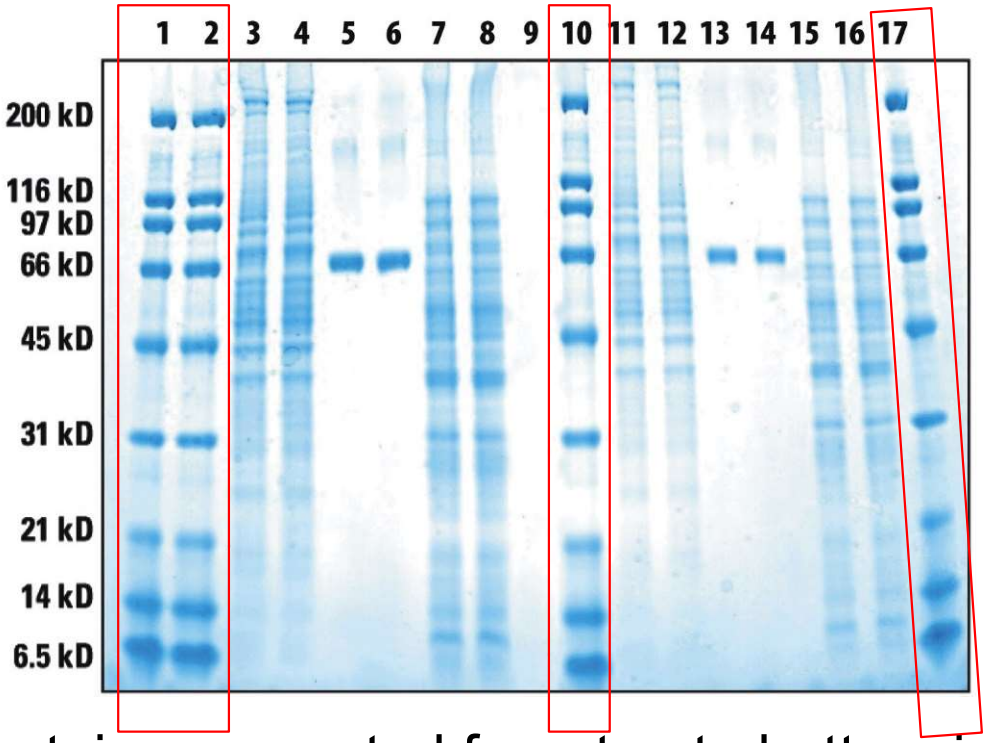
MACQUARIE University

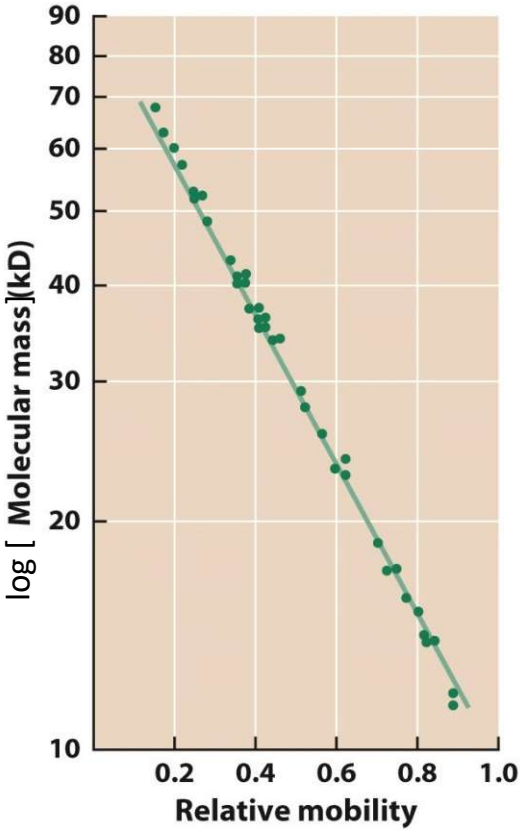
Charge of the protein can be modified by changing pH and by adding salt

Gel electrophoresis



Protein mixtures separated based on mass and charge





- Proteins separated from top to bottom in parallel lanes and stained with a dye: Coomassie blue.
- Lanes 1, 2, 10 and 17 contain **mass standards** as shown.
- The **relative mobility** is related to the **log** of the **molecular mass**: lighter molecules move faster than heavier ones.


Case Study 2: Proteases!
Team work time!

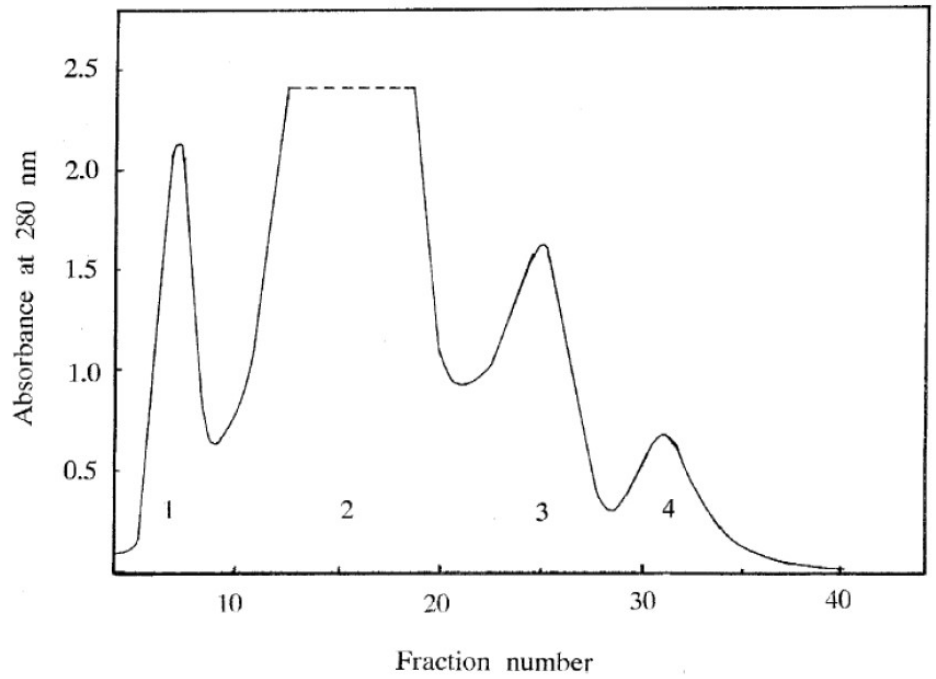
Case Study Q&A!

Case Study Question 1:

What is the property on which separation by a gel filtration column is based?

Case Study Question 2:

How is the BN protein different from the other proteins found in the *B. nigra* seeds from the elution profile shown?

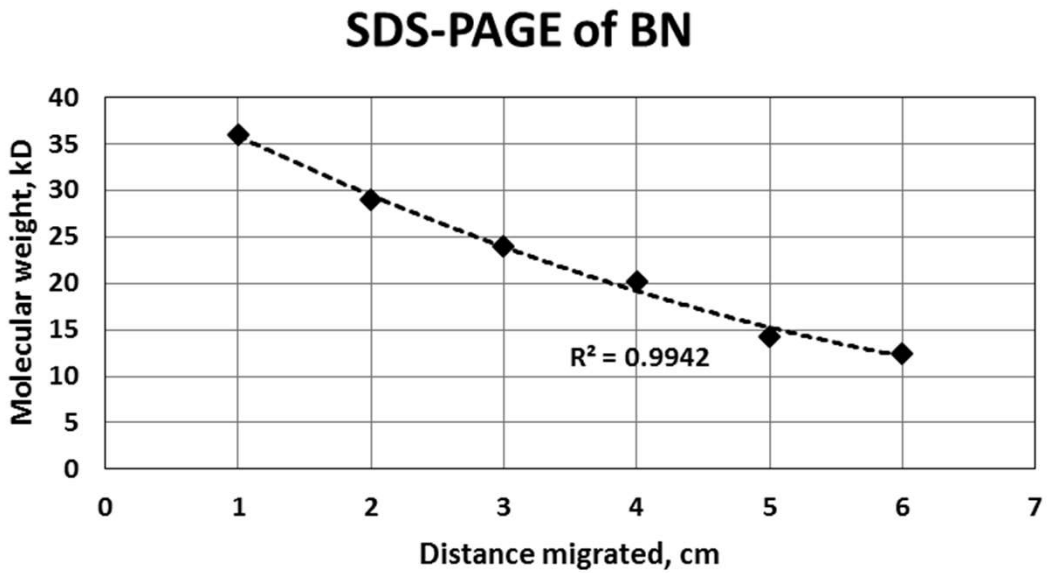


MACQUARIE University

Case Study Question 3:

What is the molecular weight of the BN protein from the standard curve?

Molecular weight, kD	Distance migrated, cm
36	0.3
29	1.4
24	2.6
20.1	3.5
14.2	5.4
12.5	6.1
BN protein	5.0



Note: The relationship is logarithmic and usually log(Molecular weight) is plotted. For this tutorial, we are using a simplified graph, with a log curve fitted to the points, to get a quick answer.

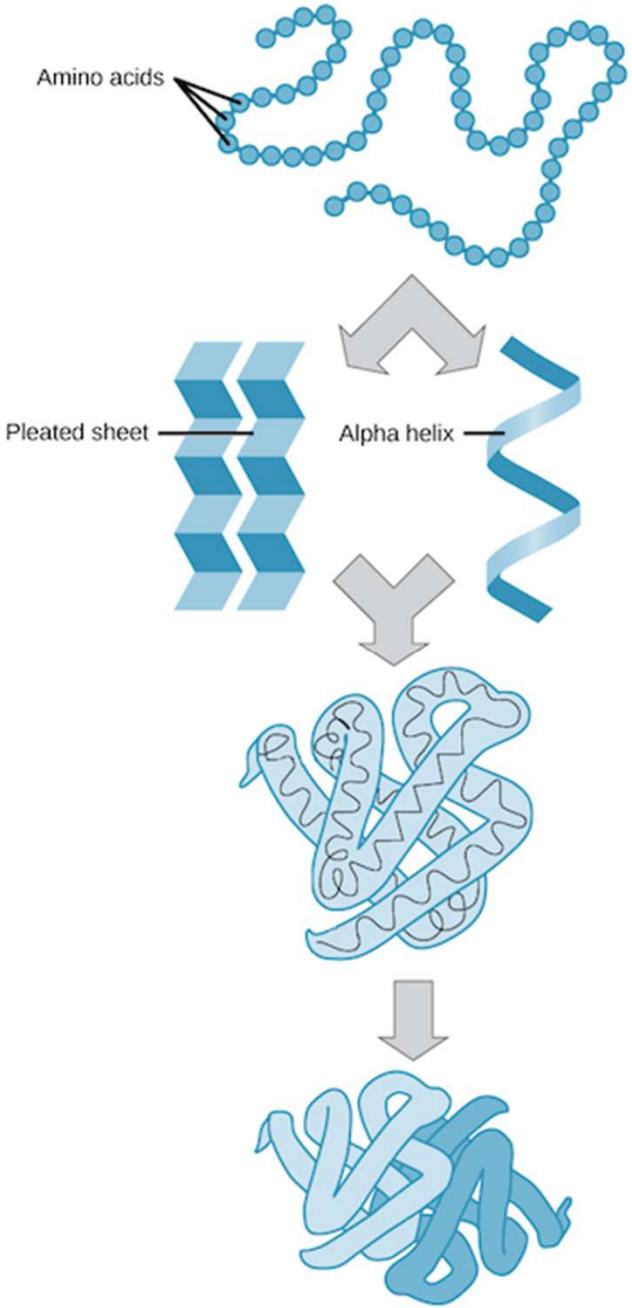
Case Study Question 4:

If *N*-acetyl serine is the amino terminal amino acid, why would sequencing using the Edman method be unsuccessful?

Case Study Question 5:

What are some other ways investigators could try and sequence the protein?

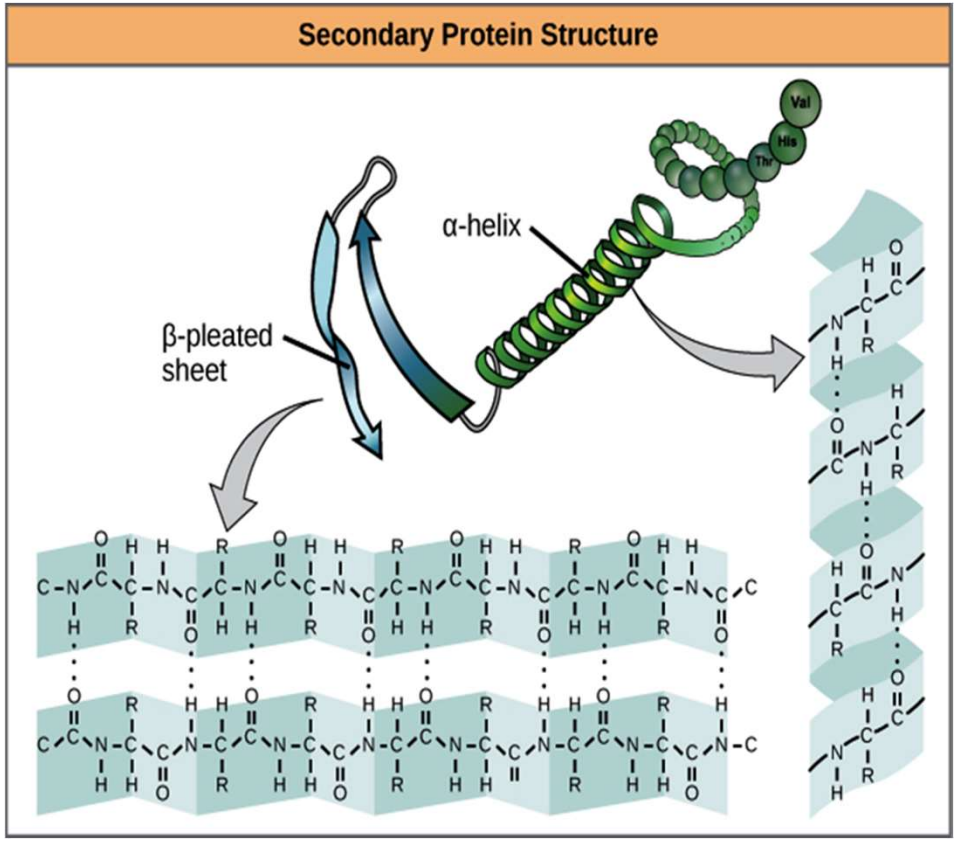
Protein structure: Definition of terms (Prac 3)



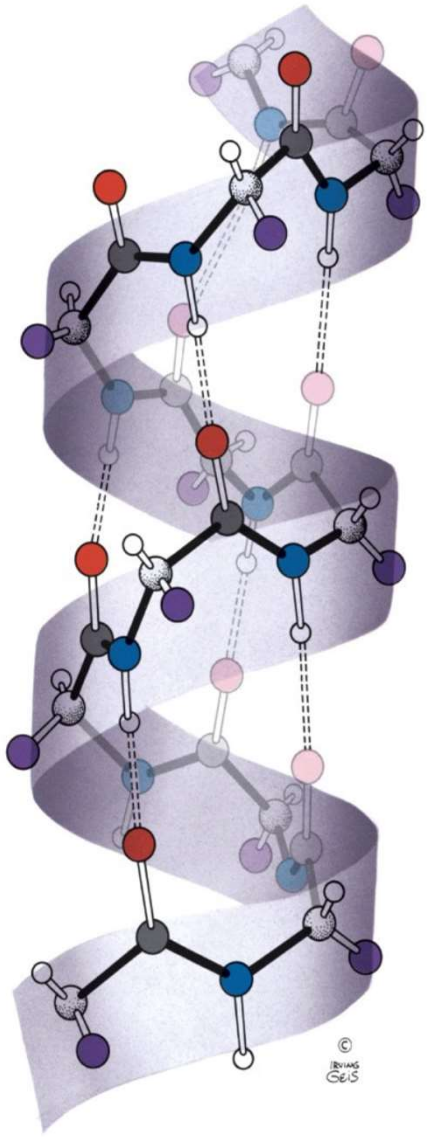
- Primary structure
  - Sequence of amino acids in a chain (in order!)
- Secondary structure
  - Structural elements in proteins that are primarily formed through peptide backbone interactions
  - H-bonding causes amino acids to fold into a repeating pattern
- Tertiary structure
  - The overall three-dimensional structure of a protein
- Quaternary Structure
  - Arrangement of subunits within a multi-subunit protein

Secondary Structure

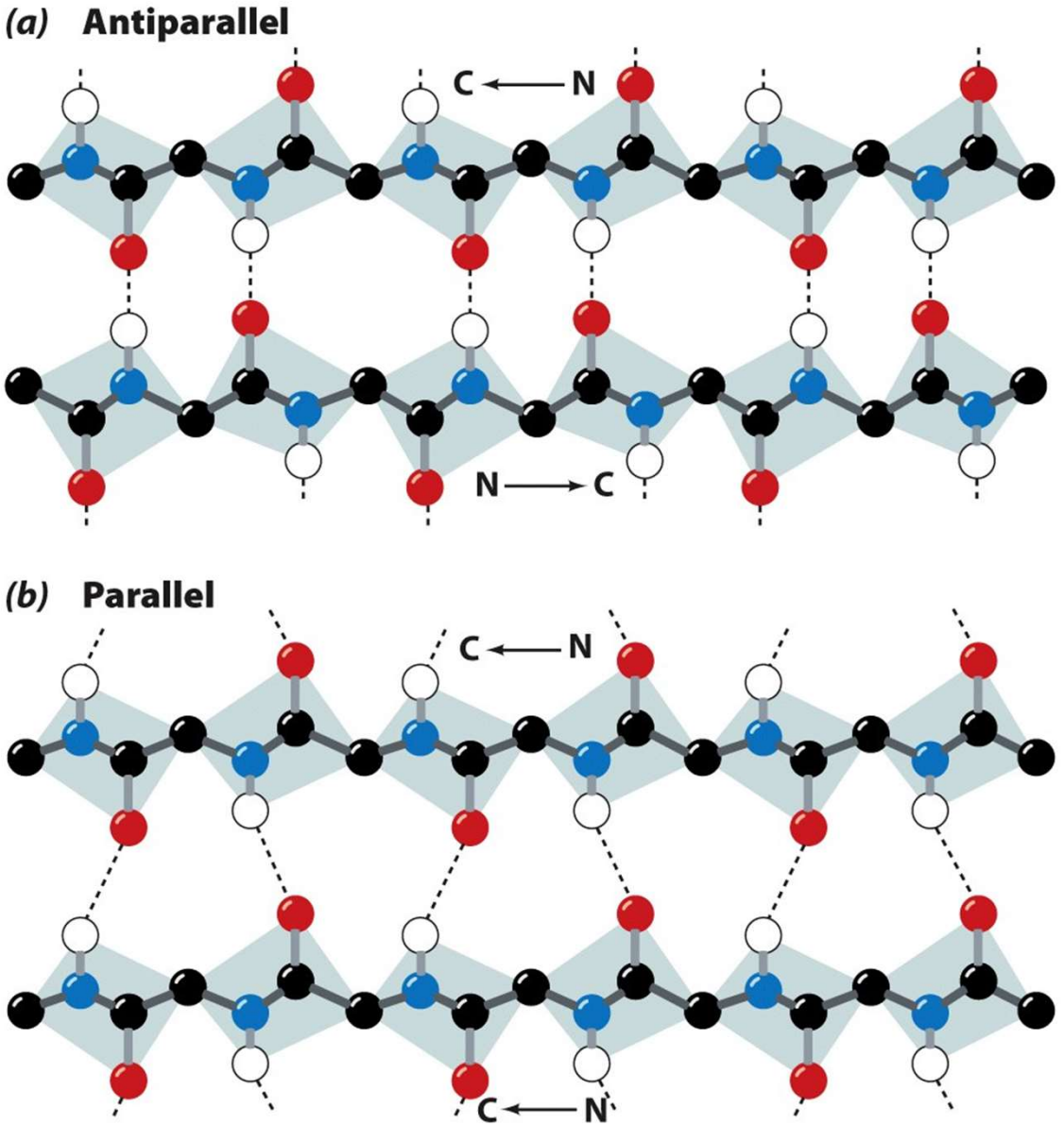
- Backbone of a protein tends to form hydrogen bonds with itself.
  - Backbone just refers to the **amino acid chain apart from the R groups**
- Hydrogen bonds formed between any N-H (amino) and C=O groups (carbonyl) that are close together in space
- Two main types of 2° structures:
  - α-helices
  - β-sheets, made up of β-strands
- H-bonding is the stabilising force of 2° protein structure.



Hydrogen bonding patterns in α-helices and β-sheets



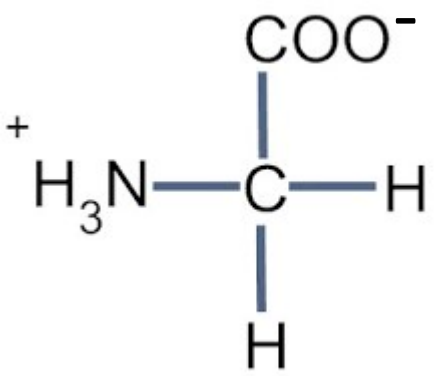
α-helix



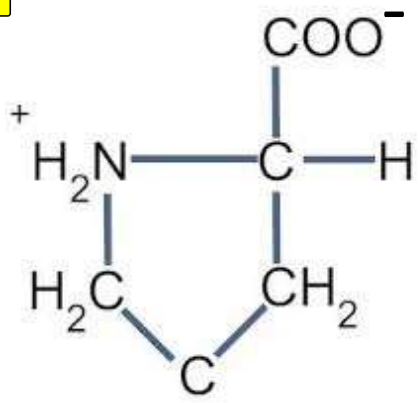
β-sheets

MACQUARIE University

Glycine and proline in peptides and proteins



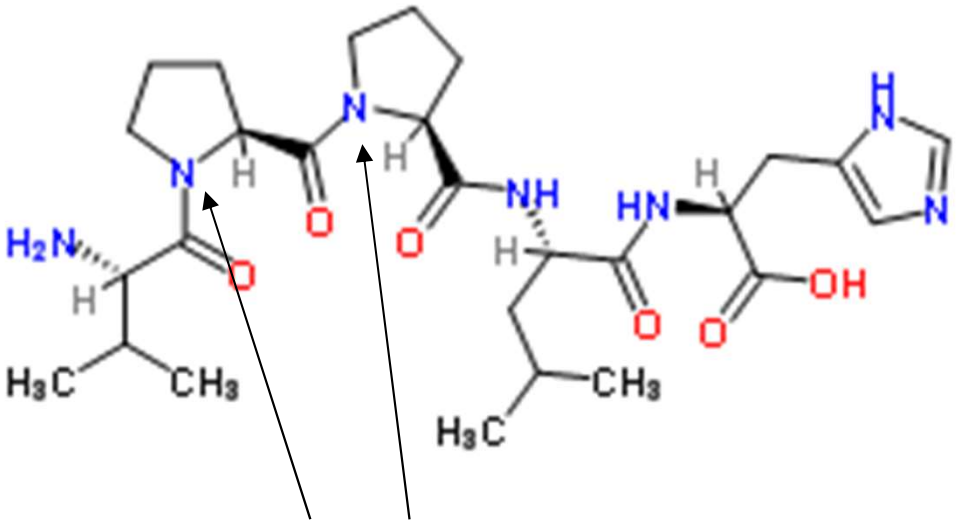
Glycine



Proline

Glycine:

- **Has no** sidechain which makes this the most flexible amino acid
- Rarely found in a helix, as it tends to bend too freely and deforms the helix
- Commonly found in **turns**



Proline

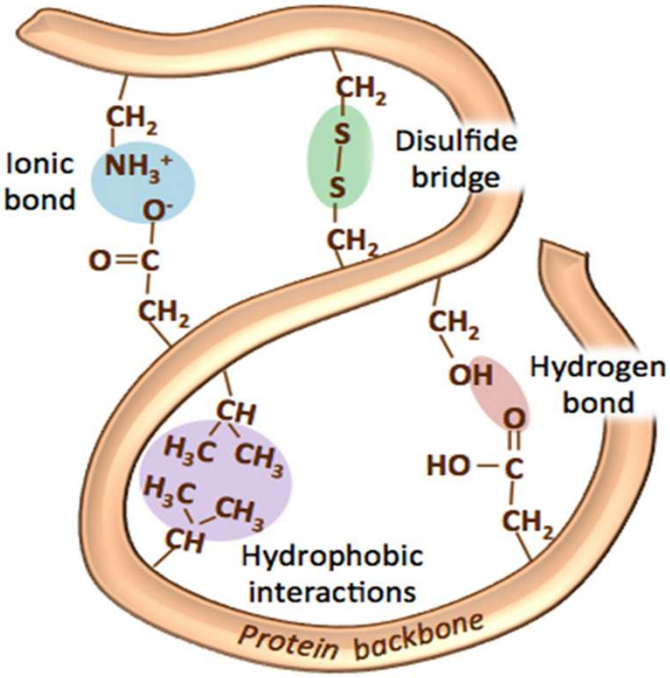
Proline:

- **Does not** have a H atom on its amino group, so it cannot donate a hydrogen bond to stabilise α-helices or β-sheets
- **Rarely found in a helix**
- Causes a bend in the helix
- Seen at the end of helix or in turns or loops

MACQUARIE University

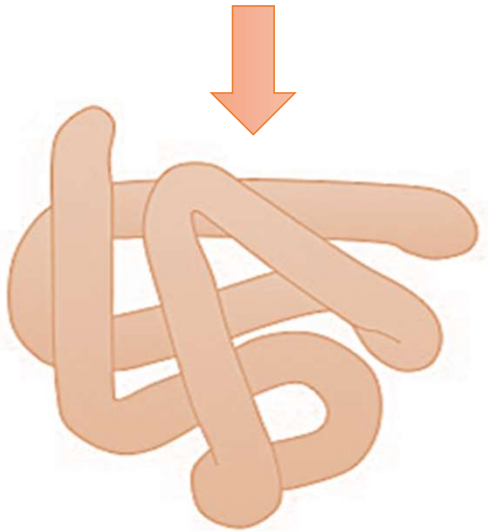
Tertiary Structure

- Overall three–dimensional structure of a polypeptide
- Formed primarily due to interactions between the R side chains of the amino acids in the protein
- Important elements may be **far apart in the primary structure** but **really close together in the tertiary structure**, such as **binding pockets**



Where do the R side chains go? Based on polarity:

- Non-polar (Val, Leu, Ile, Met, Phe):
  - Usually in the hydrophobic interior away from solvent
- Charged polar (Asp, Glu, His, Lys, Arg):
  - Usually on the surface in contact with solvent
- Uncharged polar (Ser, Thr, Asn, Gln, Tyr):
  - usually on the surface in contact with solvent



One polypeptide
(tertiary structure)

Tertiary structure brings distant catalytic residues together in 3D space!

e.g. Serine proteases: digestive enzymes **trypsin**, **chymotrypsin** and **elastase**

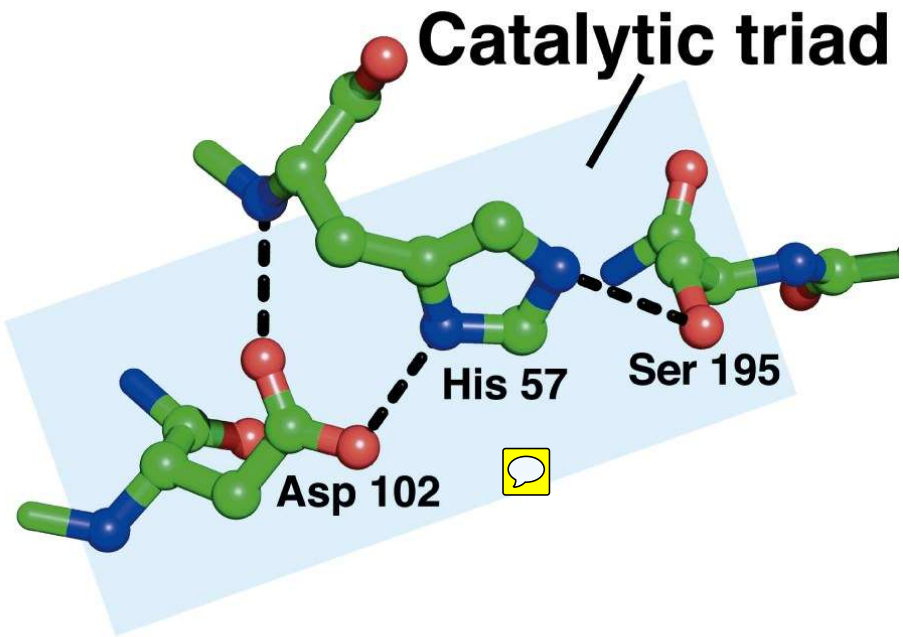
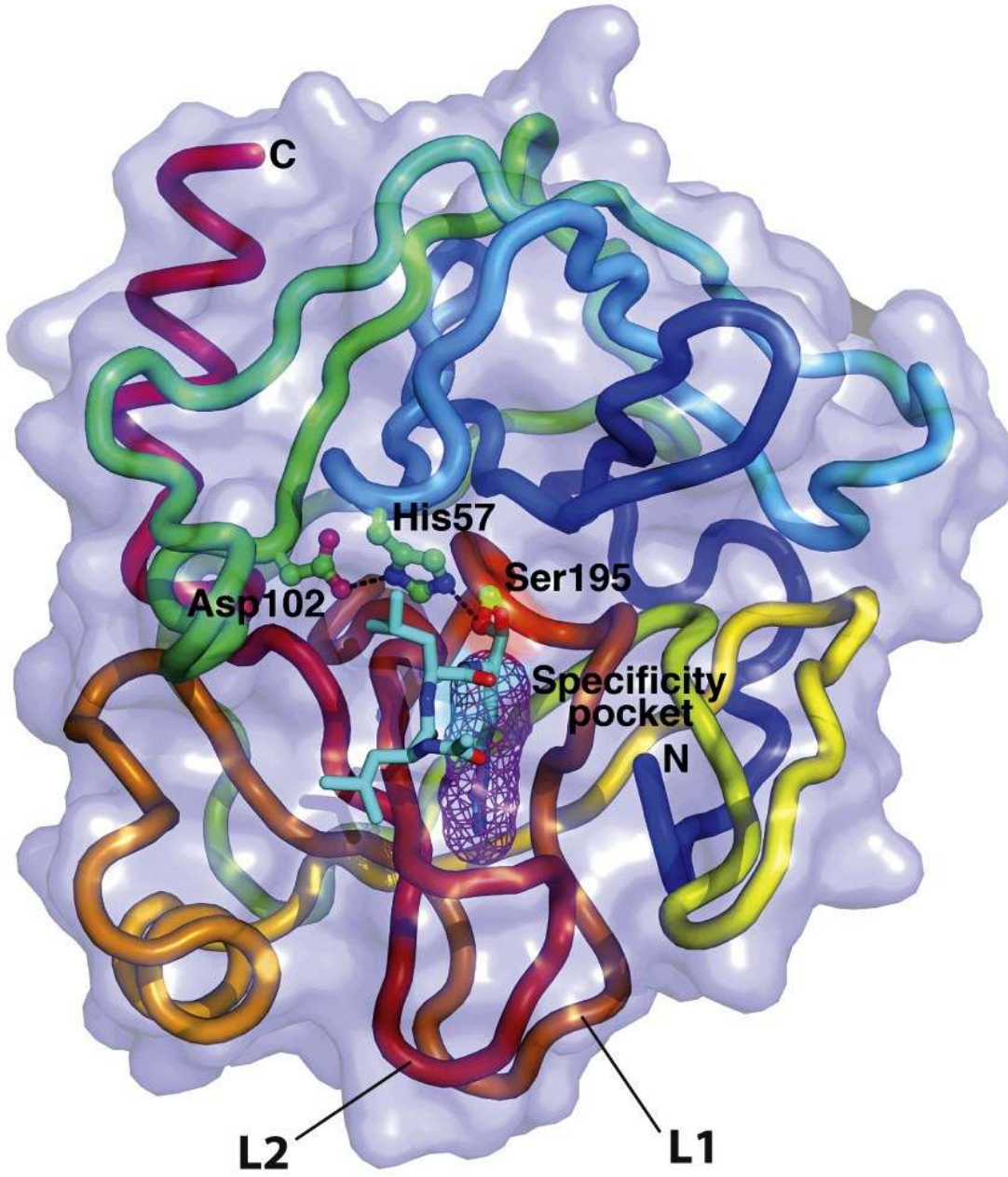


Figure 11-26

Bovine trypsin with inhibitor leupeptin (PDB: 2AGI)



MACQUARIE University

Quaternary Structure

- Most proteins are made of a **single polypeptide chain** and have only three levels of structure (1°, 2° & 3°)
- Some proteins are made up of **multiple polypeptide chains**, also known as **subunits**
- Arrangement of these subunits into a bigger multisubunit structure = **quaternary structure**

Deoxyhemoglobin
PDB: 2DHB

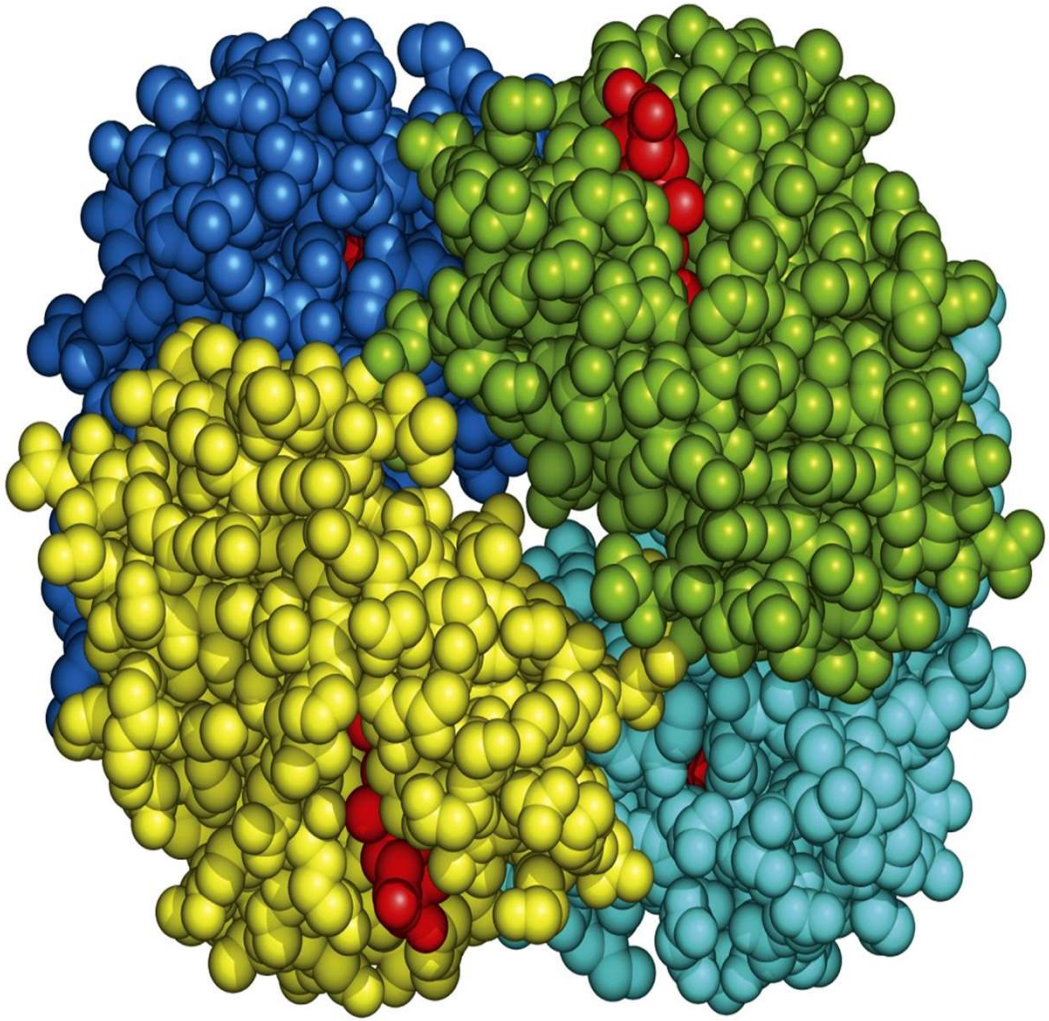
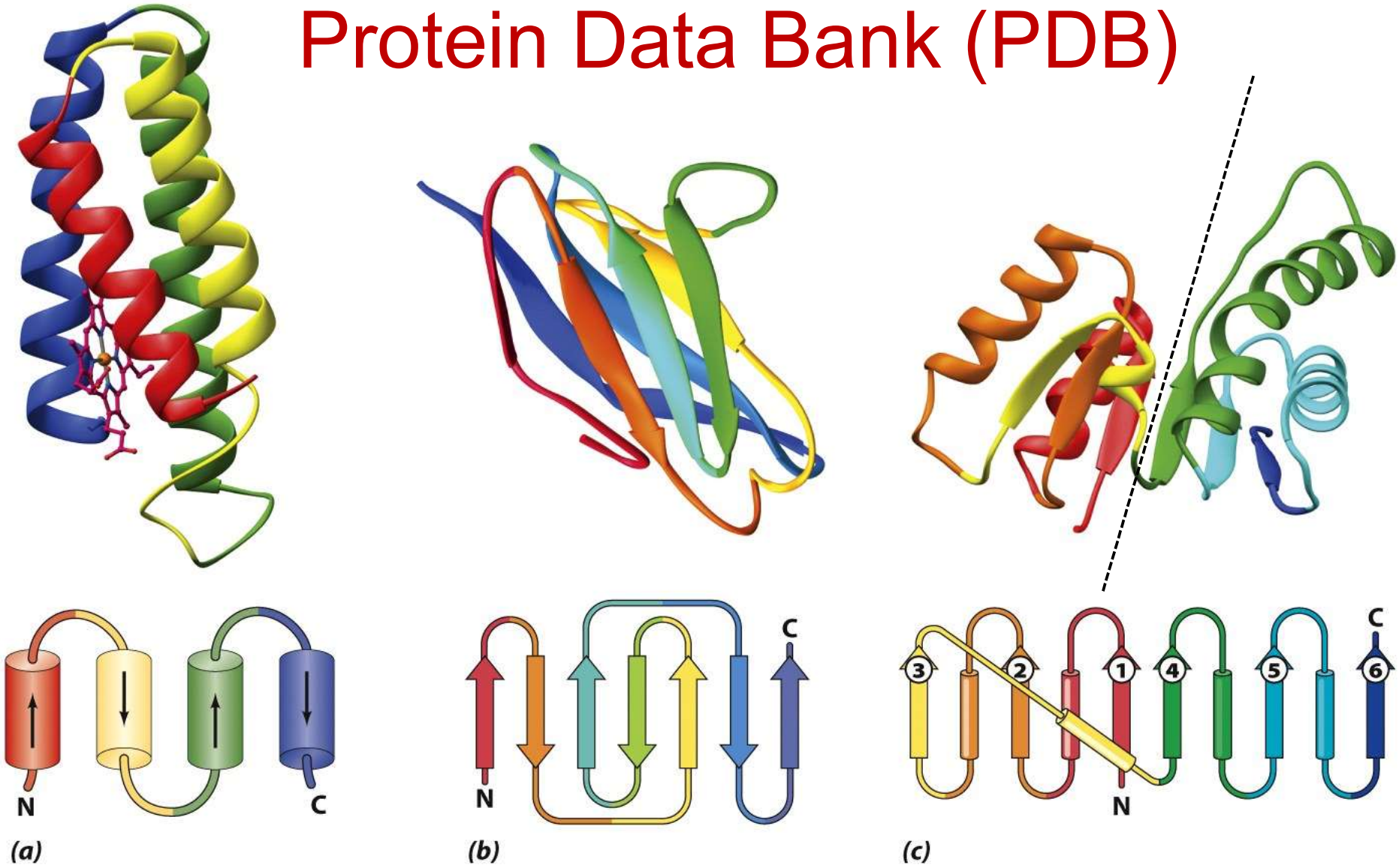


Figure 6-33

Example Protein Structures from Protein Data Bank (PDB)



Cytochrome *b*562
PDB 256B

Human immunoglobulin fragment
PDB 7FAB

Dogfish lactate dehydrogenase
PDB 6LDH

MACQUARIE University

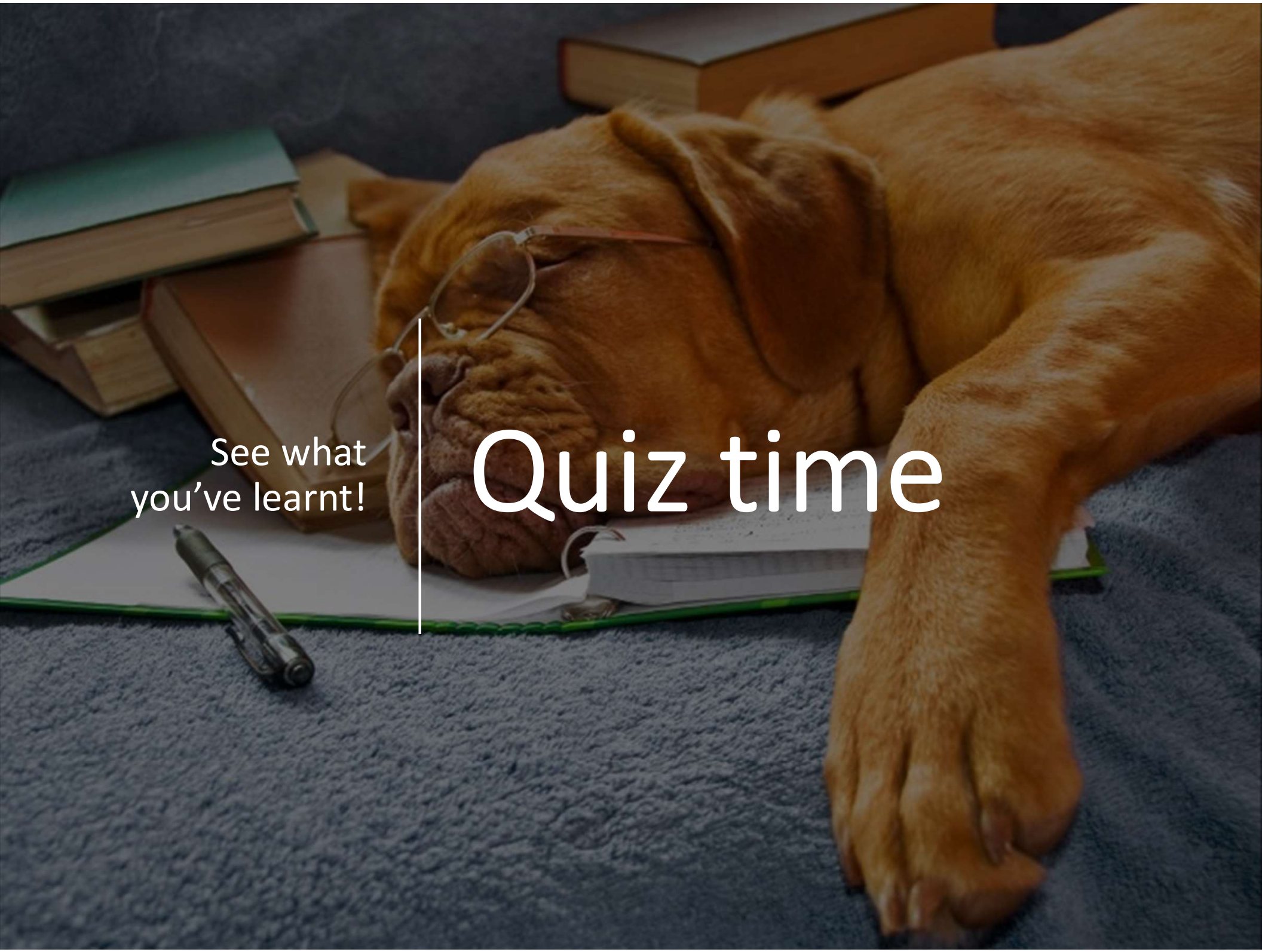
Question 3: Secondary and Tertiary Structure

a) Which residue typically breaks α-helices?

i. Leu ii. Pro iii. Glu

b) True or false: the catalytic residues in the binding pocket of an enzyme are usually next to each other in the primary sequence.

c) What is the main force that holds secondary structure elements (α-helices and β-sheets) together?


See what
you've learnt!
Quiz time