Separation and purification of proteins:

1. **Crude extract preparation**: break open tissue / cells to release proteins. Differential centrifugation to isolate subcellular fractions/ organelles.

2. **Fractionation**: treatment of extract to separate proteins into different fractions based on a property (size/charge / differences in protein solubility at high salt conc.)

3. **Salting out**: addition of a salt ammonium sulfate \((\text{NH}_4)_2\text{SO}_4\) in a right amount to selectively precipitate certain proteins while others remain in soln.

4. **Dialysis**: separate proteins from solvents depending on size in a bag/tube containing a semipermeable mb. which allows exchange of solvent and salt but not protein. (good to remove ammonium sulfate from separated protein)

5. **Chromatography**: depend on differences in protein, size, charge, binding affinity.
Column chromatography:
Composed of two phases in a glass/plastic column
1- stationary phase:
Solid / porous matrix through which flows the mobile
2- mobile phase: buffered soln
Rate of protein flow through column ↓ with column length and time spent ↑
Ion exchange chromatography:
Exploit differences in *sign* and *magnitude* of net electric charges of a protein.

Cation-exchangers:
Solid matrix has –ve charge groups

Anion+exchangers:
Solid matrix has +ve charge groups

**Note:**
Elution: washing out of the column

Proteins move through the column at rates determined by their net charge at the pH being used. With cation exchangers, proteins with a more negative net charge move faster and elute earlier.
**Size exclusion chromatography:**

Solid phase: beads with pores / cavities.

Large proteins emerge sooner than small ones since they can't enter cavities so they take a short, rapid path around the beads through the column.
Affinity chromatography:

Based on binding affinity of a protein
Beads have a covalently attached chemical group = ligand
A protein with affinity to this ligand binds the beads other proteins flow through the column

Note:
Ligand: a group / molecule that binds a macromolecule such as a protein.
A refinement in chromatographic methods:

HPLC = High Performance Liquid Chromatography

Speeds the rate of protein flow through the column using:

1) high pressure pumps.
2) high quality beads that can withstand crushing due to high pressure.
Protein separation: 
thousands of published protocols → start with inexpensive protocol.

**Table 5–5**

<table>
<thead>
<tr>
<th>A Purification Table for a Hypothetical Enzyme*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Procedure or step</td>
</tr>
<tr>
<td>-------------------</td>
</tr>
<tr>
<td>1. Crude cellular extract</td>
</tr>
<tr>
<td>2. Precipitation with ammonium sulfate</td>
</tr>
<tr>
<td>3. Ion-exchange chromatography</td>
</tr>
<tr>
<td>4. Size-exclusion chromatography</td>
</tr>
<tr>
<td>5. Affinity chromatography</td>
</tr>
</tbody>
</table>

*All data represent the status of the sample after the designated procedure has been carried out. Activity and specific activity are defined on page 137.
Separation and characterization of proteins by electrophoresis:

Migration of charged proteins in an electric field.

**Advantage:**

1) a protein can be visualized as well as separated.
2) allows determination of isoelectric point (IP) and molecular weight (Mwt).
3) Degree of purity of protein sample.
Gels made of crossed linked polymer polyacrylamide. Acts as a molecular sieve. Slows proteins migration in proportion to their charge/mass ratio.

PROTEIN MIGRATION in gel depends on its size and shape.
A detergent binds to protein in proportion to its M wt.: 1 molecule SDS / 2 a.a.

SDS separates proteins relative to mass.

\[
\text{Na}^+\text{O} - \text{S} - \text{O} - \text{(CH}_2\text{)}_{11}\text{CH}_3
\]

Sodium dodecyl sulfate (SDS)
The end result:
all proteins contain only primary structure

denatured protein (reduced to its primary structure) and linearized
**Gel staining:**

After electrophoresis proteins visualized by treating the gel with a stain/dye **Coomassie blue** which binds proteins not the gel.

Useful to monitor the purification steps:
Estimate no. of different proteins, Mwt. when purification ended.

Figure on the right: RNA pol (E coli)
First lane: proteins in crude extract
Successive lanes: proteins extracted after each purification step.
Last lane: purified protein contains 4 subunits
Estimating molecular weight of a protein:
The position of a protein band in the gel provides a measure of its M wt.

![Diagram](image)

- Myosin: 200,000
- β-Galactosidase: 116,250
- Glycogen phosphorylase b: 97,400
- Bovine serum albumin: 66,200
- Ovalbumin: 45,000
- Carbonic anhydrase: 31,000
- Soybean trypsin inhibitor: 21,500
- Lysozyme: 14,400

**Relative migration**

**Log \( M_r \)**
$p_I = pH$ the net charge of a protein = zero

<table>
<thead>
<tr>
<th>Protein</th>
<th>$p_I$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pepsin</td>
<td>~1.0</td>
</tr>
<tr>
<td>Egg albumin</td>
<td>4.6</td>
</tr>
<tr>
<td>Serum albumin</td>
<td>4.9</td>
</tr>
<tr>
<td>Urease</td>
<td>5.0</td>
</tr>
<tr>
<td>$\beta$-Lactoglobulin</td>
<td>5.2</td>
</tr>
<tr>
<td>Hemoglobin</td>
<td>6.8</td>
</tr>
<tr>
<td>Myoglobin</td>
<td>7.0</td>
</tr>
<tr>
<td>Chymotrypsinogen</td>
<td>9.5</td>
</tr>
<tr>
<td>Cytochrome c</td>
<td>10.7</td>
</tr>
<tr>
<td>Lysozyme</td>
<td>11.0</td>
</tr>
</tbody>
</table>
Determination of the IP of proteins by isoelectric focusing:

1- preparation of a pH gradient by ampholytes addition in cylindrical gel.

2- addition of protein mix.

3- migration of proteins until each reaches a pH equivalent to its PI.
Two dimensional electrophoresis:
SDS polyacrylamide + isoelectric focusing.

More sensitive analytical method than any alone.
Separates proteins with similar PI but different M wt., or vice versa.
Two dimensional electrophoresis > 1000 protein separated from E.coli
Separation of a protein that is an Enzyme:
The amount in a given soln. / tissue extract expressed as enzyme activity =
The rate at which the substrate is converted $\rightarrow$ rxn products.

So we need to know the following:

1) The overall equation of rxn.
2) Analytical procedure for product appearance / substrate disappearance.
3) Requirement for cofactors metal ions/ coenzymes.
4) Dependence on substrate conc.
5) Optimum pH.
6) Temperature zone at which enzyme is stable and active, usually (25-38 °C).
Activity versus specific activity:

1 unit of E activity = amount of E transforming 1.0 umol substrate / min. at 25°C

Activity = total unit of E in a soln.

Specific activity = no. E units / mg total protein.

Specific activity measure of E purity, ↑ during purification procedures.
Maximal +constant when E is pure.

Both has same no. of red marbles. But different marbles of other colors.
Both has same activity of the protein (red). Specific activity differs.
The primary structure of a protein determines how it folds into its three-dimensional structure → determines protein function.

Relationship between a.a sequence and function:

1) Proteins with different functions has different a.a sequence.
2) Human genetic: diseases linked to defective proteins (in one third of these cases a single a.a is changed).
3) Functionally similar proteins from different species have similar a.a sequences.
   e.g. ubiquitin (76 a.a. protein) identical fruit flies → humans.
The a.a sequence is not completely fixed, but flexible. 20-30% of human proteins are polymorphic = proteins that contain a.a variations in human population.

Proteins contain crucial regions with conserved sequences essential to their function.
Two major discoveries in biochem history 1953:

1) Watson Crick and DNA double helix.
2) Frederick Sanger and insulin sequence.

Two polypeptide chains joined by disulfide linkage.
Similarity in sequence bw. human, dog, horse, cow, pig, and other animals.
Polypeptide sequencing of primary structure:

Labeling and identification of the N-terminal a.a residue:

Sanger developed (FDNB) 1-fluro-2,4-dinitrobenzene. Other reagents: dansyl chloride, dabsyl chloride.

All label the N-terminal residue.

To identify *only* this residue the protein is hydrolyzed \(\rightarrow\) a.a and is destroyed.

No. of polypeptides identified in protein e.g. phe and Gly for insulin.

\[
\begin{align*}
\text{Dansyl chloride} & & \text{Dabsyl chloride} \\
\text{CH}_3 & \text{CH}_3 & \text{CH}_3 \\
\text{CH}_3 & \text{SO}_2\text{Cl} & \text{N} = \text{N} & \text{SO}_2\text{Cl}
\end{align*}
\]
Dansyl / dabsyl chloride

• a reagent that reacts with primary amino groups in a.a. → blue or blue green fluorescent sulfonamide adducts.

• Dansyl chloride is widely used in protein sequencing and amino acid analysis.
**Edman degradation:** To sequence the entire polypeptide:

Labels and removes only the amino terminal residue from a peptide, leaving all other bonds intact.

1) The peptide + phenylthiocarbamoyl (PTC) in **alkaline conditions**.
2) Cleavage of peptide bond next to PTC in **trifluoroacetic acid** with removal of terminal a.a in **acidic conditions**.

Each a.a is labeled, removed and identified through this sequence of rxns.

The procedure is repeated until the entire sequence is determined.

- Edman degradation carried out in a machine called **Sequencer** mixes reagents in proper proportions, separates products, identifies and records the results. (from few micrograms of protein)
Sangers method identifies amino terminal residue.

Edman degradation reveals the entire sequence of peptide.
Sequencing of Large polypeptides:

1) Disulfide bonds cleaved.

2) Protein cleaved by chemical / enzymatic methods.

3) Each fragment purified and sequenced by Edman procedure.

4) Peptide fragments ordered.

5) Disulfide bonds located.
Cleaving polypeptide chain:

1) Proteases: catalyze hydrolytic cleavage of peptide bonds.

The no. of smaller peptides produced by trypsin can be predicted from no. of total Lys and Arg residues in original polypeptide.

2) Chemicals:

Can function in a specific manner

<table>
<thead>
<tr>
<th>Treatment*</th>
<th>Cleavage points†</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trypsin</td>
<td>Lys, Arg (C)</td>
</tr>
<tr>
<td>Submaxillar&lt;sub&gt;us&lt;/sub&gt; protease</td>
<td>Arg (C)</td>
</tr>
<tr>
<td>Chymotrypsin</td>
<td>Phe, Trp, Tyr (C)</td>
</tr>
<tr>
<td>Staphylococcus aureus V8 protease</td>
<td>Asp, Glu (C)</td>
</tr>
<tr>
<td>Asp-N-protease</td>
<td>Asp, Glu (N)</td>
</tr>
<tr>
<td>Pepsin</td>
<td>Phe, Trp, Tyr (N)</td>
</tr>
<tr>
<td>Endoproteinase Lys C</td>
<td>Lys (C)</td>
</tr>
<tr>
<td>Cyanogen bromide</td>
<td>Met (C)</td>
</tr>
</tbody>
</table>

*All 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 protein and sequencing and ordering the peptide fragments:

1) Determine the a.a composition and terminal residue.
2) Break disulfide bonds
3) Fragmentation by two methods.
4) Overlap sequences.
Determination of the polypeptide sequence from DNA sequence:

Development of DNA sequencing methods.

When the gene is available easier to sequence DNA than protein.

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

DNA sequence (gene)  CAGTATCCTACGATTTTG