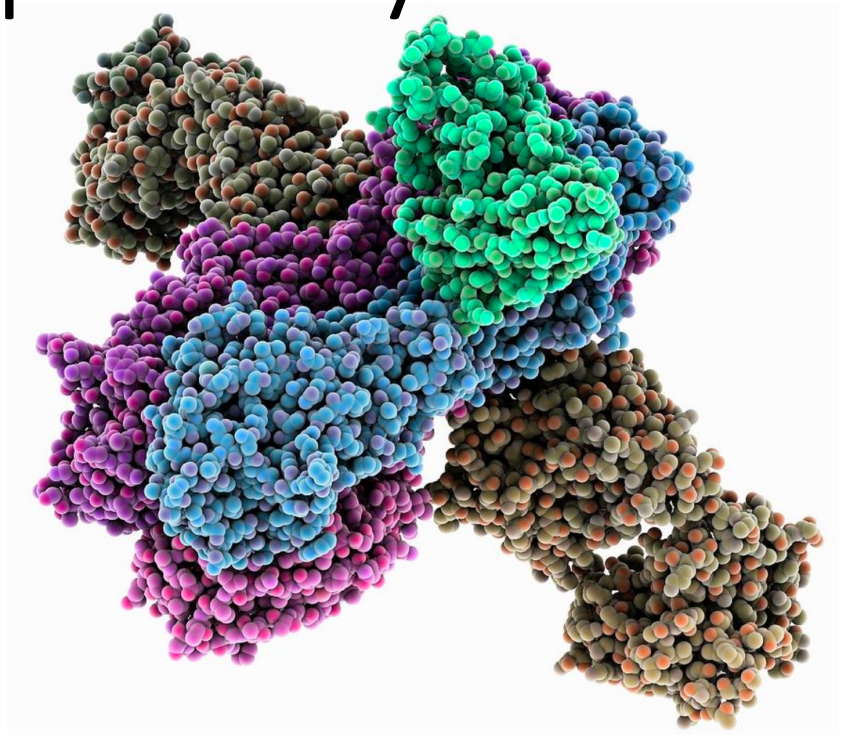
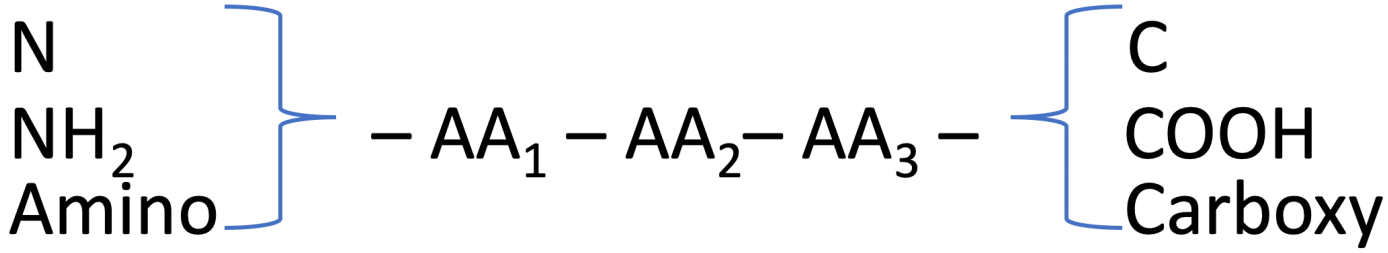
Lesson 4
Protein polarity and structure



Protein polarity

- Protein = amino acid polymer



Protein polarity

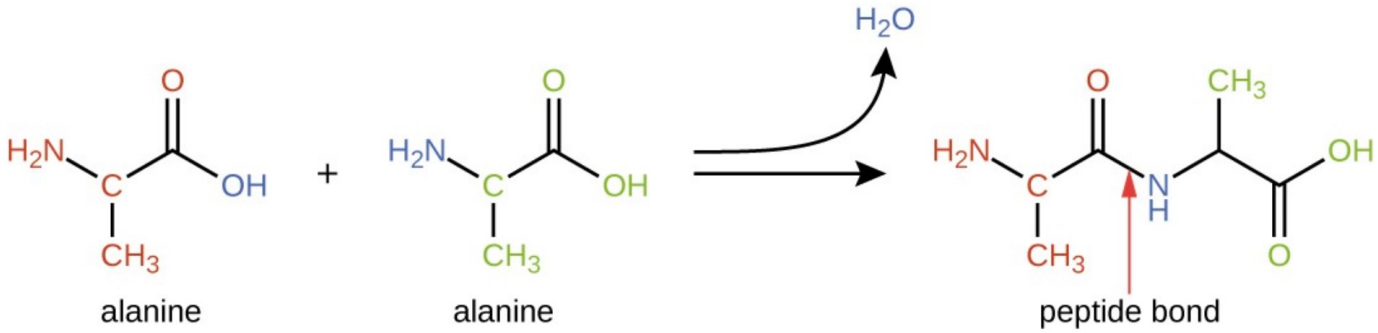
- Protein = amino acid polymer

N
NH_2
Amino

$-AA_1-AA_2-AA_3-$

C
COOH
Carboxy

- Peptide bond = covalent bond between NH_2 of AA_n and COOH of AA_{n+1}



Protein polarity

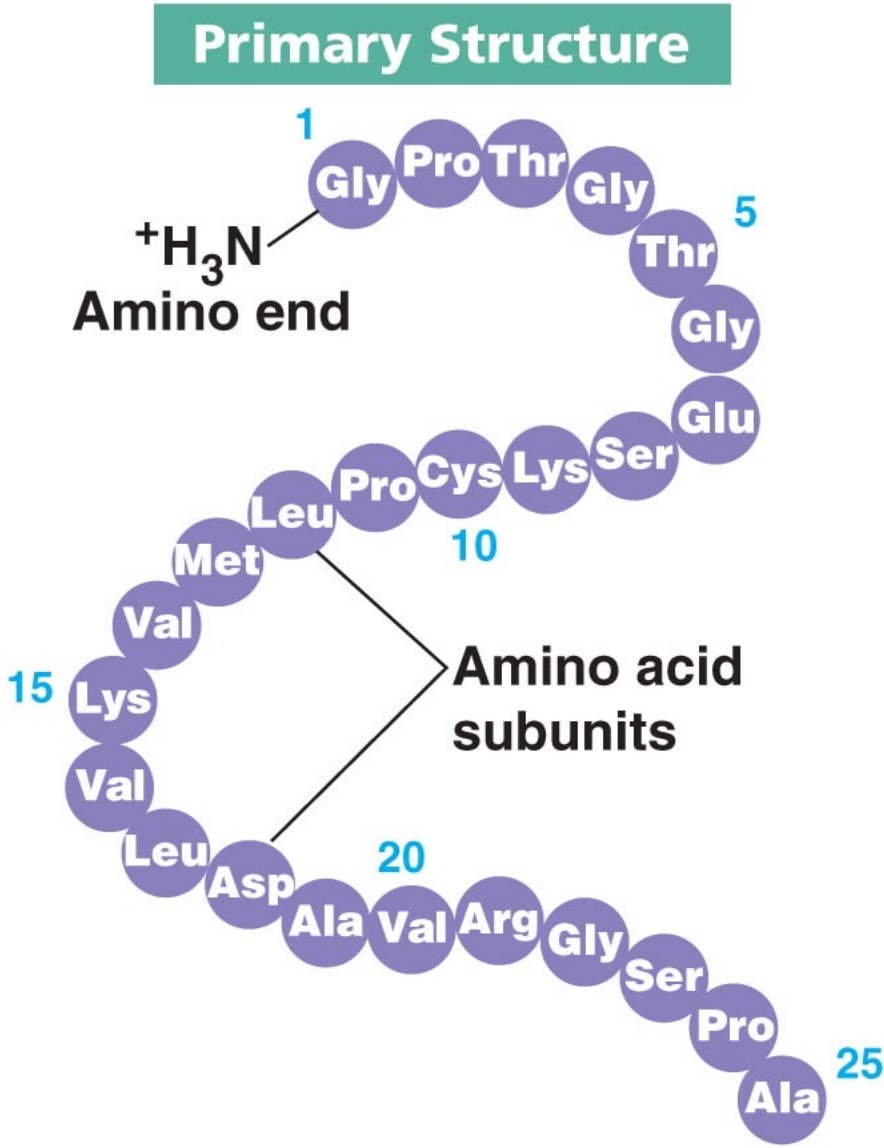
- Protein = amino acid polymer

N
NH_2
Amino

$$- AA_1 - AA_2 - AA_3 -$$

C
COOH
Carboxy

- Peptide bond = covalent bond between NH_2 of AA_n and COOH of AA_{n+1}
- POLARITY = amino (N) and carboxy (C) ends
- INFORMATION = amino acids order



Protein polarity

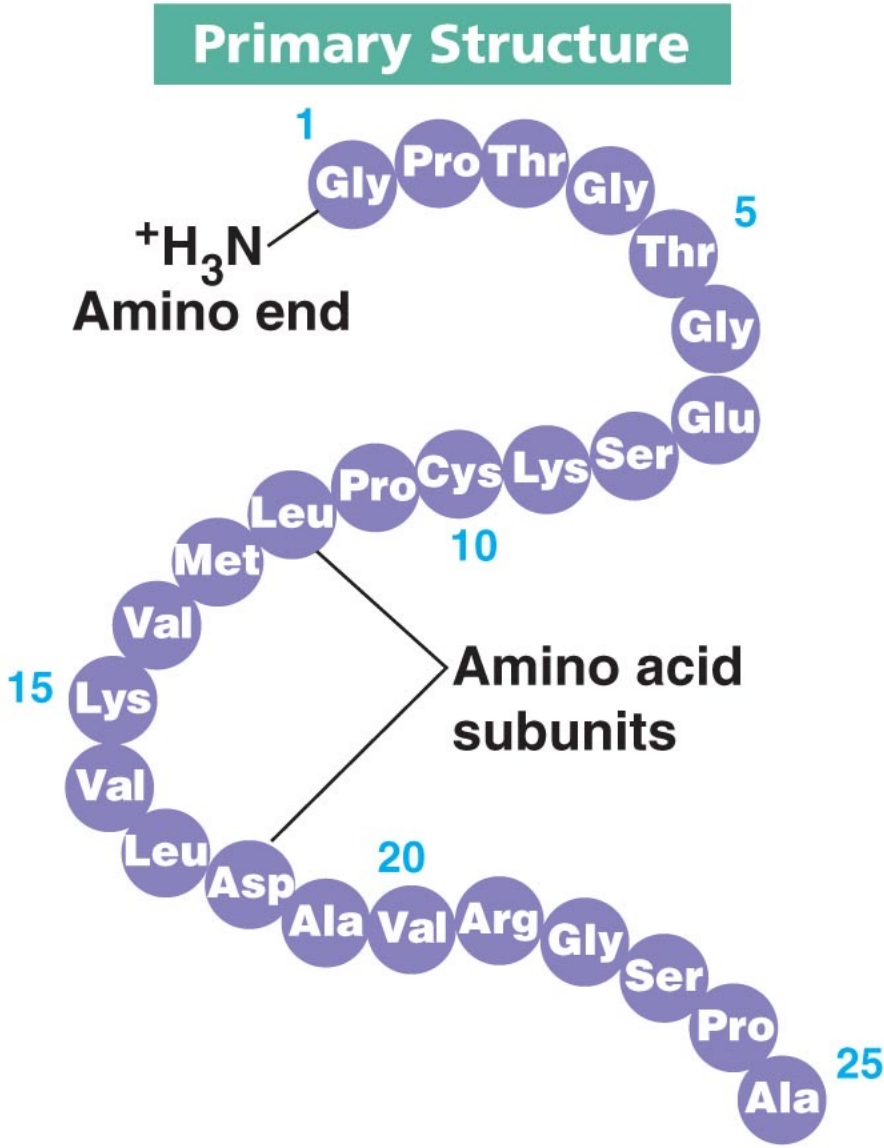
- Protein = amino acid polymer

N
NH_2
Amino

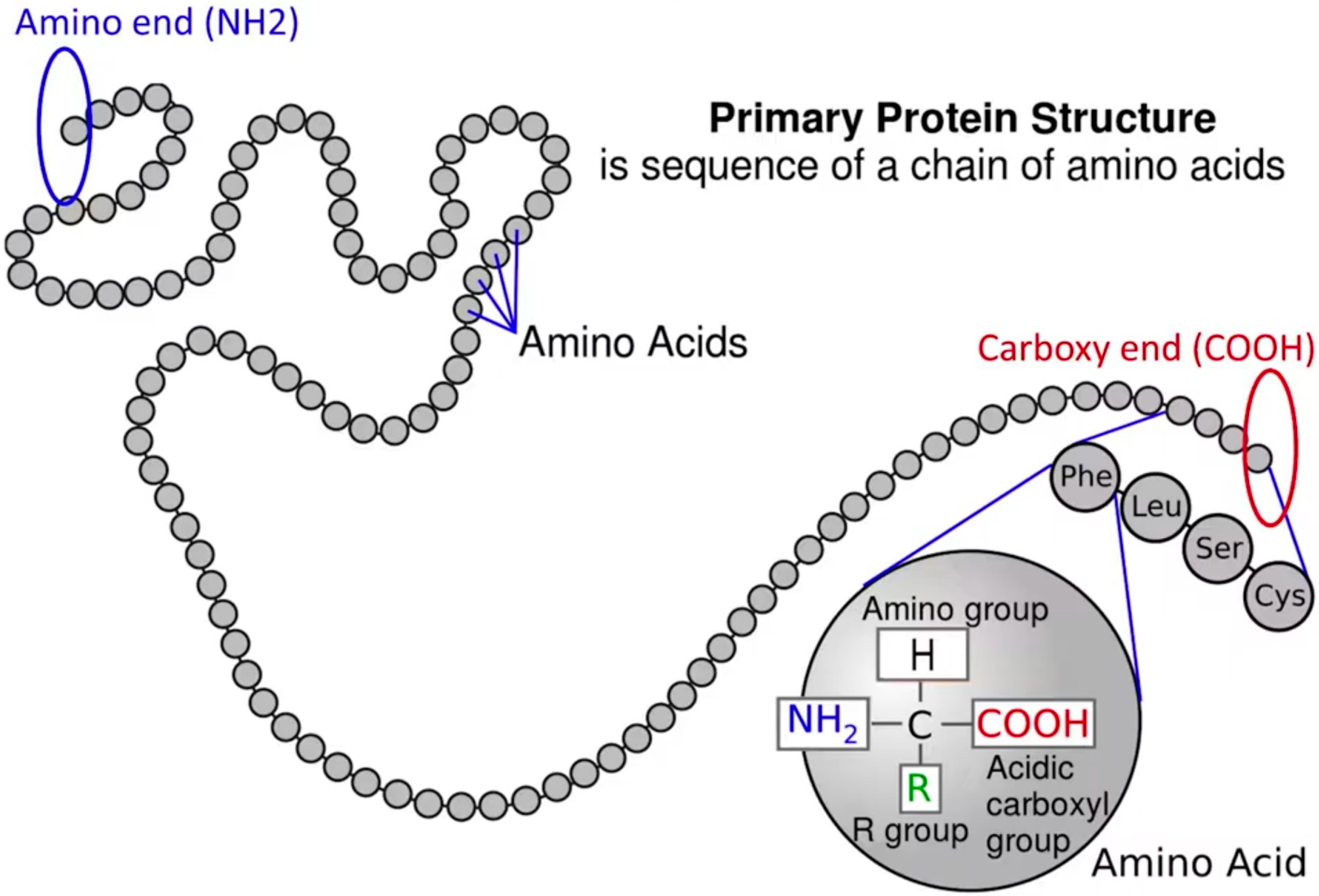
$- AA_1 - AA_2 - AA_3 -$

C
COOH
Carboxy

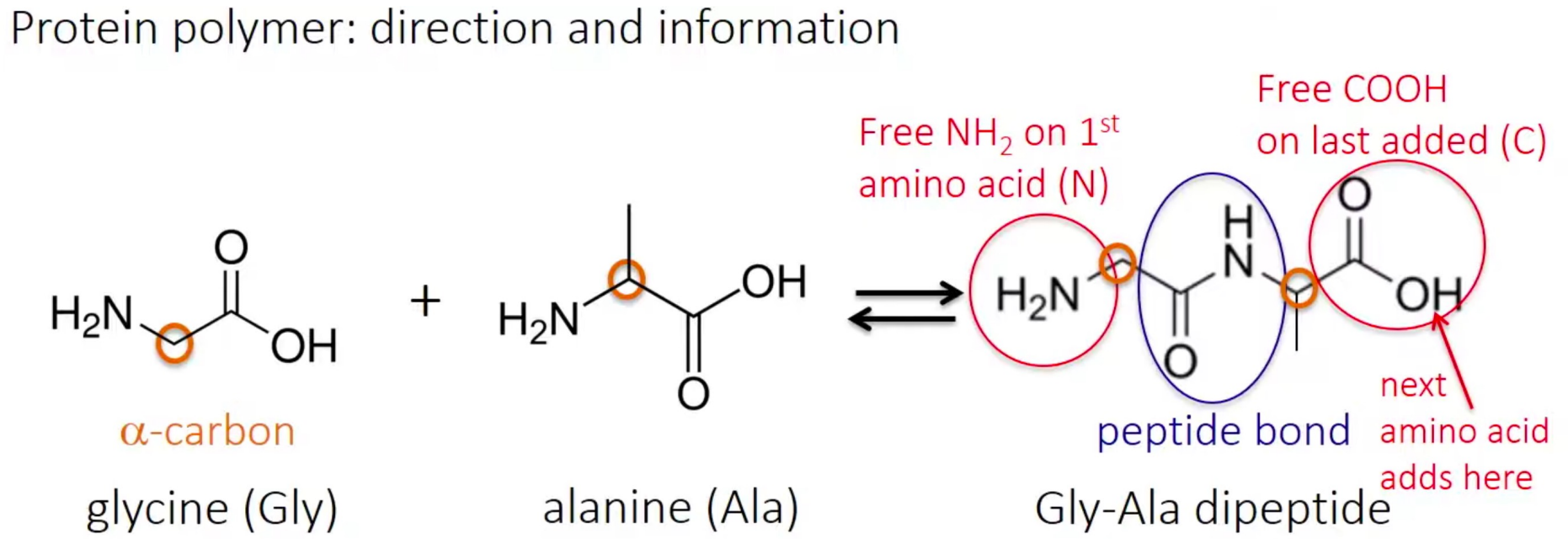
- Peptide bond = covalent bond between NH_2 of AA_n and COOH of AA_{n+1}
- POLARITY = amino (N) and carboxy (C) ends
- INFORMATION = amino acids order
- AA_3 is the last amino acid added
- Next AA adds to COOH



Protein polarity



Protein polarity



- Proteins are written with three or 1 letter amino acid code (*e.g.*, VAL or V)
- **ALWAYS write N and C at the beginning and at the end of a protein sequence**

N-Gly-Ala-Val-Ser-C or NH_2-G-A-V-S-COOH

1st → last, next adds here

Polymerization direction

Protein polarity vs nucleic acid polarity

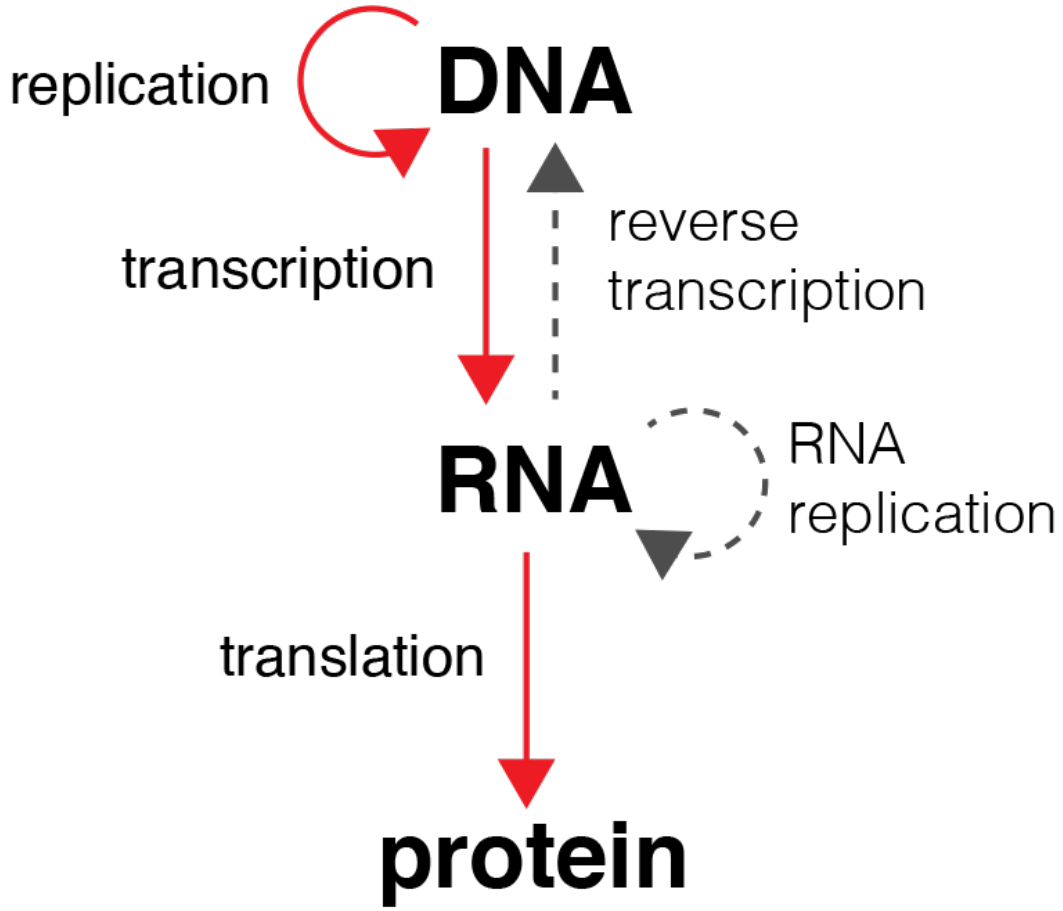
PROTEINS

- Amino acid order = INFORMATION
- Polarity = N and C ends: shows
  - First to last amino acid added
  - Direction to read information

NUCLEIC ACIDS

- Base order = INFORMATION
- Polarity = 5’ and 3’ ends: shows
  - First to last nucleotide added
  - Direction to read information

The central dogma of molecular biology

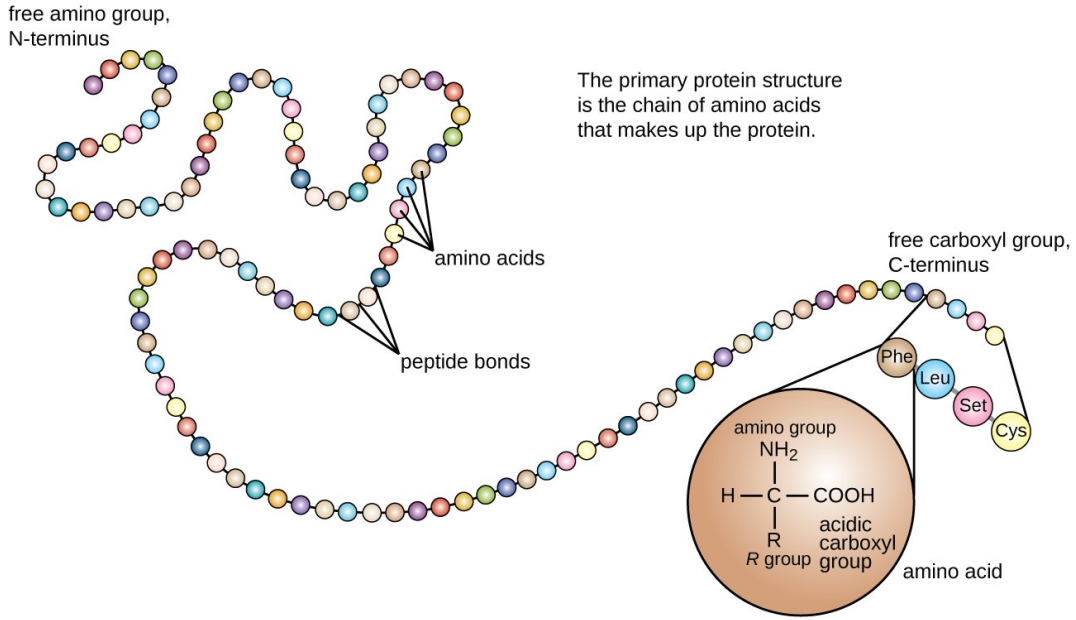


Protein structure

- Protein structure is categorized in four levels: primary, secondary, tertiary, and quaternary

Protein structure

- Protein structure is categorized in four levels: primary, secondary, tertiary, and quaternary
- The **primary structure** is simply the sequence of amino acids that make up the polypeptide chain

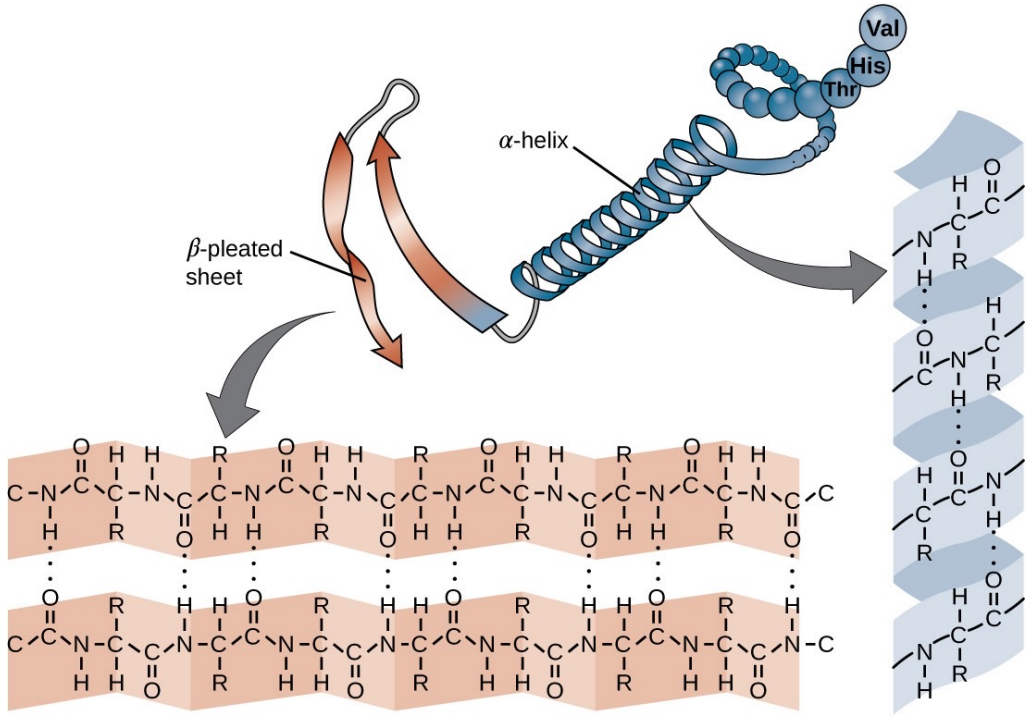


Protein structure

- The chain of amino acids that defines a protein’s primary structure is flexible

Protein structure

- The chain of amino acids that defines a protein's primary structure is flexible
- When the chain is sufficiently long, H-bonds may occur between NH_2 and COOH groups along the backbone → localized folding of chain into **helices** and **sheets**
- These shapes constitute a protein's **secondary structure**
  - the most common secondary structures are the **α-helix** and **β-pleated sheets**



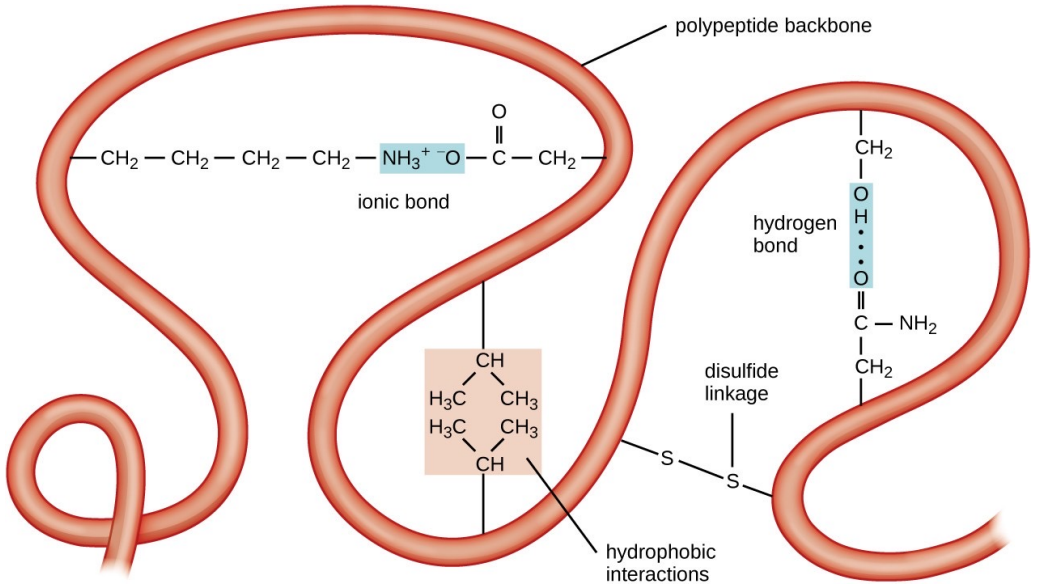
Protein structure

- The next level of protein organization is the **tertiary structure** or the large-scale three-dimensional shape of a single polypeptide chain

Protein structure

- The next level of protein organization is the **tertiary structure** or the large-scale three-dimensional shape of a single polypeptide chain

- Tertiary structure is determined by interactions between amino acid residues that are far apart in the chain:
  - disulfide bridges, which are bonds between the sulfhydryl (–SH) functional groups on amino acid side groups
  - hydrogen bonds
  - ionic bonds
  - and hydrophobic interactions between nonpolar side chains

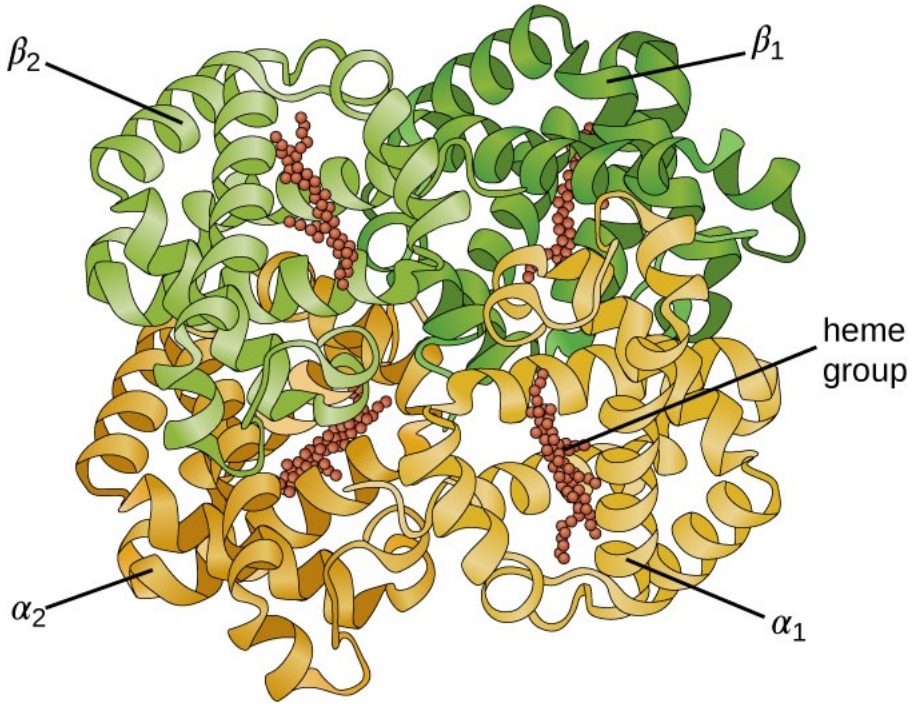


Protein structure

- Some proteins are assemblies of protein subunits
  - These proteins function adequately only when all subunits are present and appropriately configured

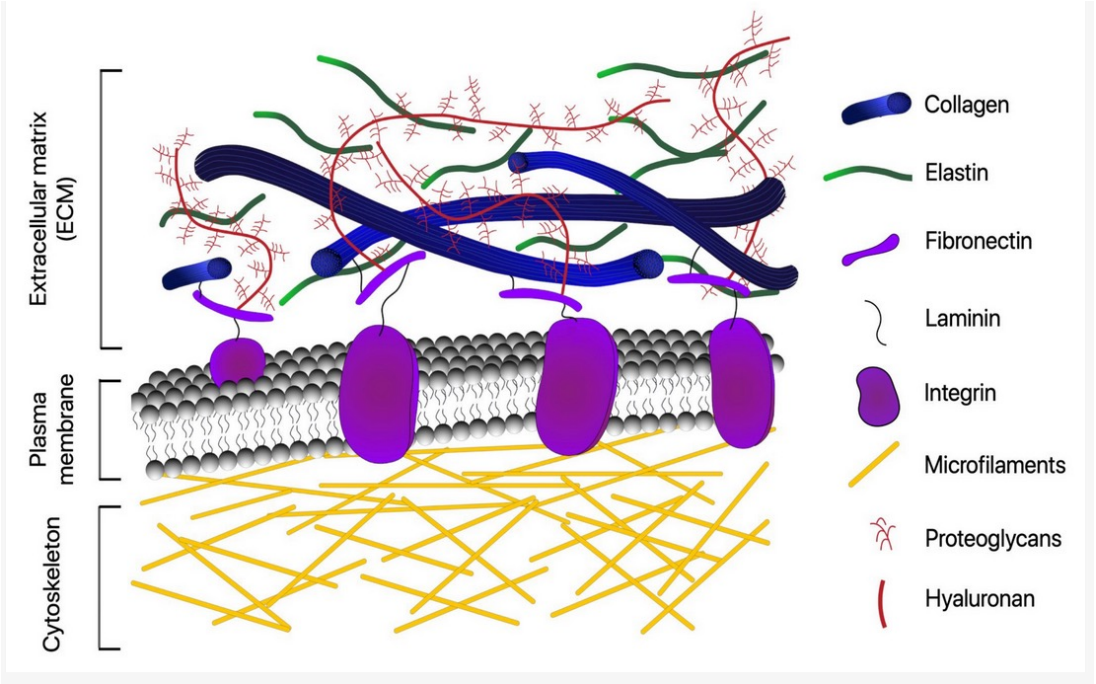
Protein structure

- Some proteins are assemblies of protein subunits
  - These proteins function adequately only when all subunits are present and appropriately configured
- The interactions that hold these subunits together constitute the **quaternary structure** of the protein
  - The overall quaternary structure is stabilized by relatively weak interactions
  - Hemoglobin is a prototypical example



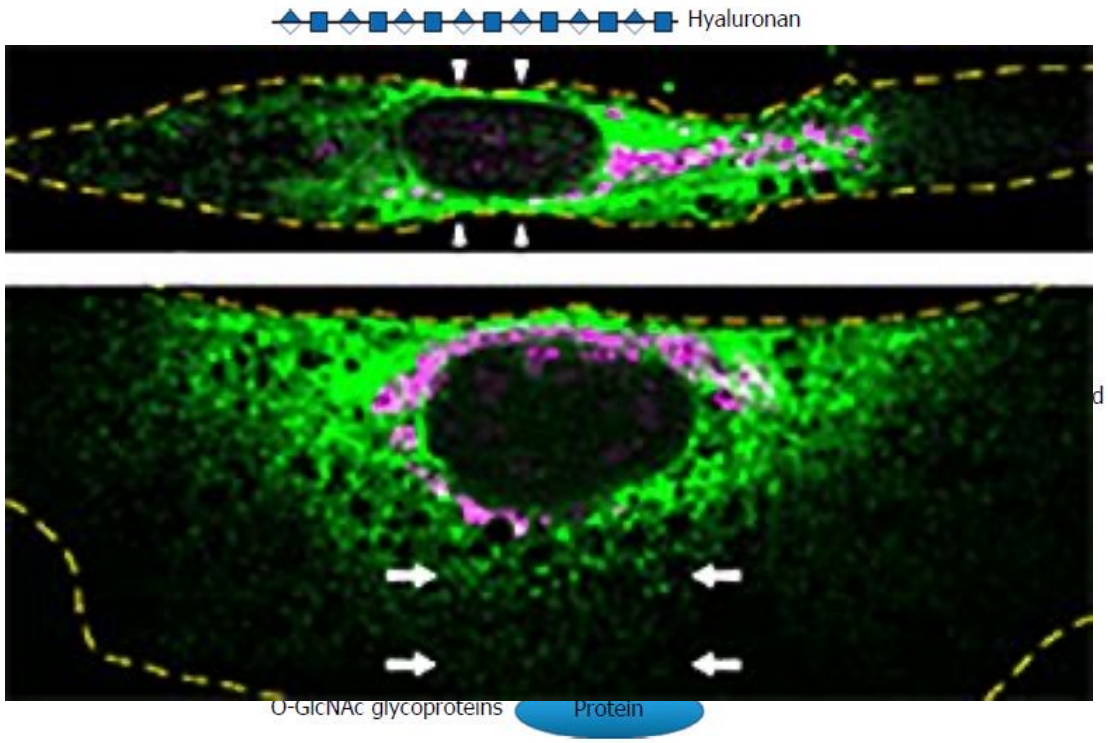
Protein functions

- Proteins reinforce structures
  - Part of the plasma membrane structure
  - Cytoskeletal proteins reinforce the cell's internal structure
  - Extracellular proteins act as cell support



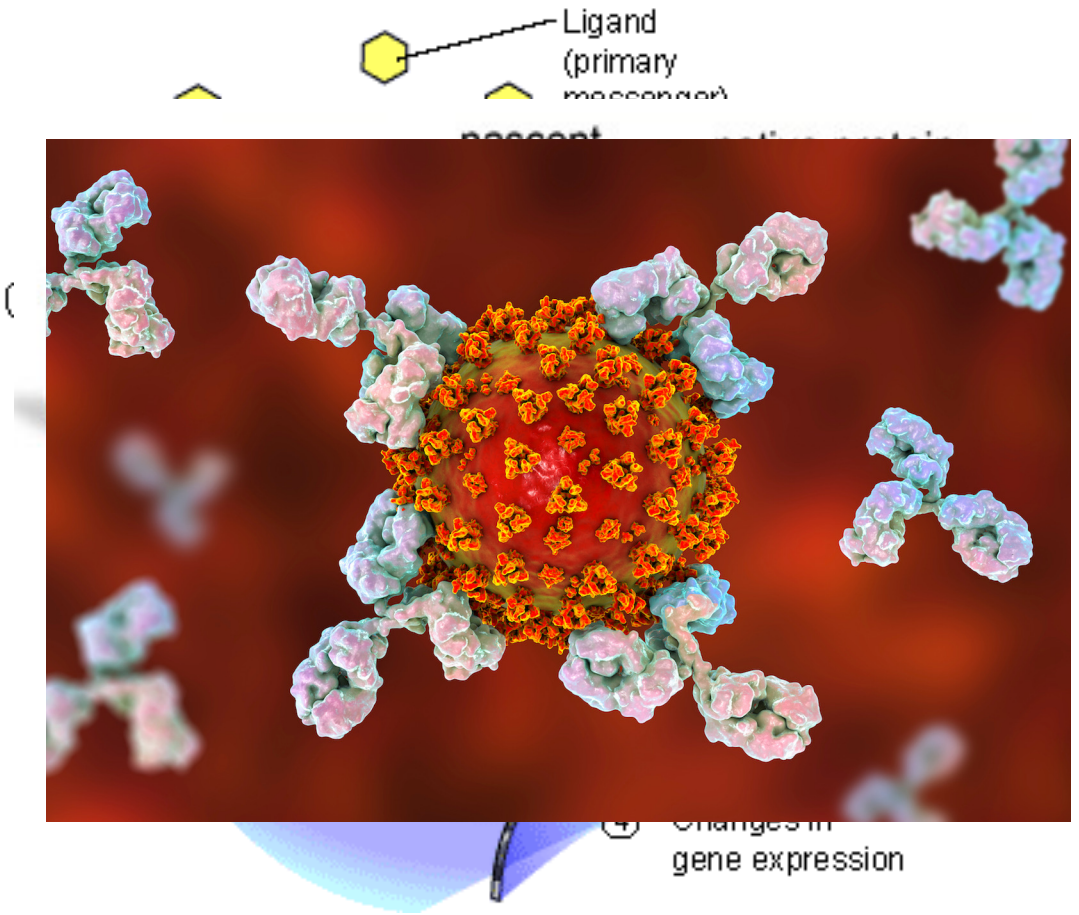
Protein functions

- Proteins transport materials in and out the cell
  - Membrane proteins, pores and channels
- Proteins are involved in cellular identity
  - Glycoproteins on cell surface act as markers that identify cells
- Proteins help cell to move
  - Cytoskeletal proteins empower flagella movement and allow cells to move like amoebae



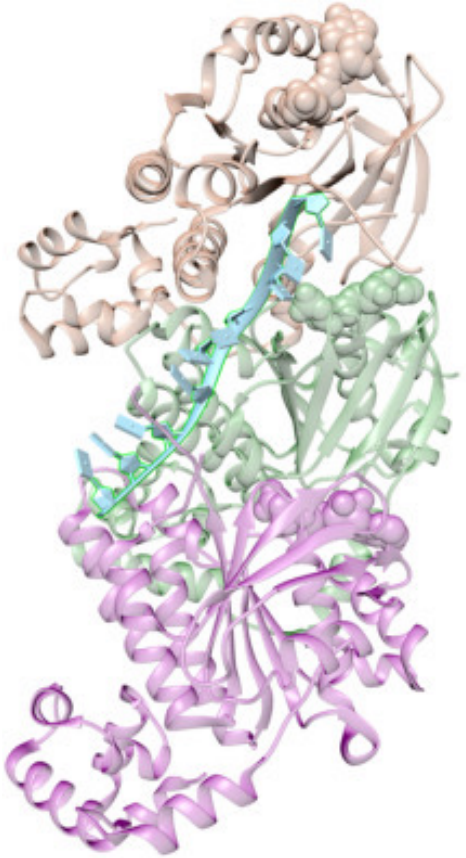
Protein functions

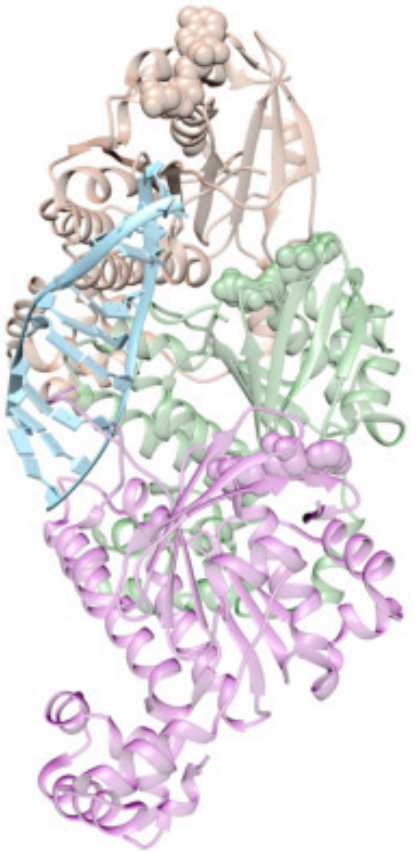
- Proteins help cells to communicate
  - Send and receive signals to and from cells
- Proteins organize molecules within a cell
  - Chaperone proteins assist folding of new proteins and guide them to precise cell locations
- Proteins help defend the body against pathogens
  - Antibodies are key players in the immune systems, helping target bacteria and viruses for destruction (more later)



Protein functions

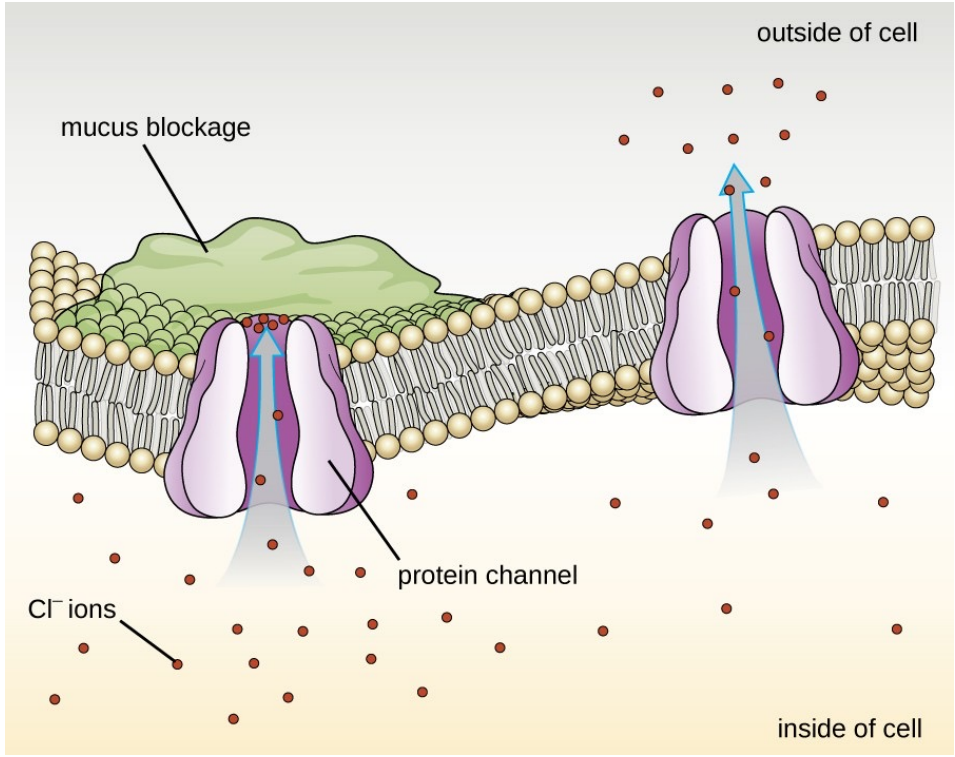
- Proteins regulate how DNA is used by the cell
  - DNA-binding proteins control which sections of DNA are to be used by the cell and which must be kept silent
    (gene expression)





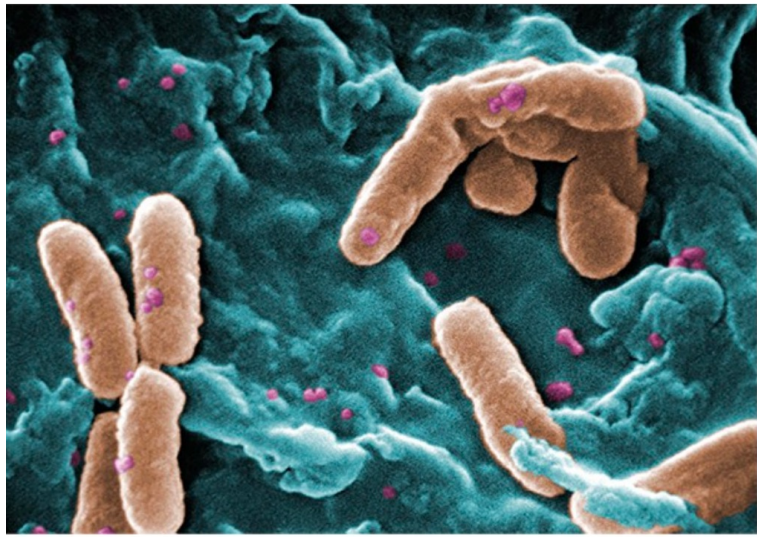
Protein structure and diseases: the CF example

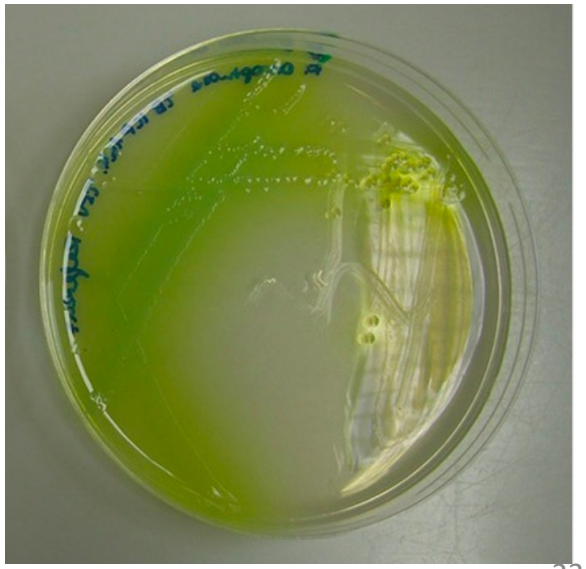
- Cystic fibrosis (CF) is a human **genetic disease**
- CF affects mostly the lungs but may also pancreas, liver, kidneys, and intestine
- CF is caused by **the loss** of the amino acid **phenylalanine** in the primary sequence of the **cystic fibrosis transmembrane protein (CFTR)**
- This **MUTATION changes the primary structure of CFTR** that normally helps transport salt and water in and out of cells
- The change in the primary structure prevents the protein from functioning properly
  - the body produces unusually thick mucus that
    - clog the lungs
    - obstructs the pancreas and stops natural enzymes from helping the body break down food and absorb vital nutrients



Protein structure and diseases: the CF example

- The altered CF mucus provides an environment where bacteria can thrive
- This colonization leads to the formation of **biofilms** in the small airways of the lungs
- The most common pathogens found in the lungs of patients with cystic fibrosis are ***Pseudomonas aeruginosa*** and ***Burkholderia cepaci***
- *Pseudomonas* differentiates within the biofilm in the lung and forms large colonies, called "mucoid" *Pseudomonas*.
  - The colonies have a unique pigmentation that shows up in laboratory tests
    - provides physicians with the first clue that the patient has CF (such colonies are rare in healthy individuals)



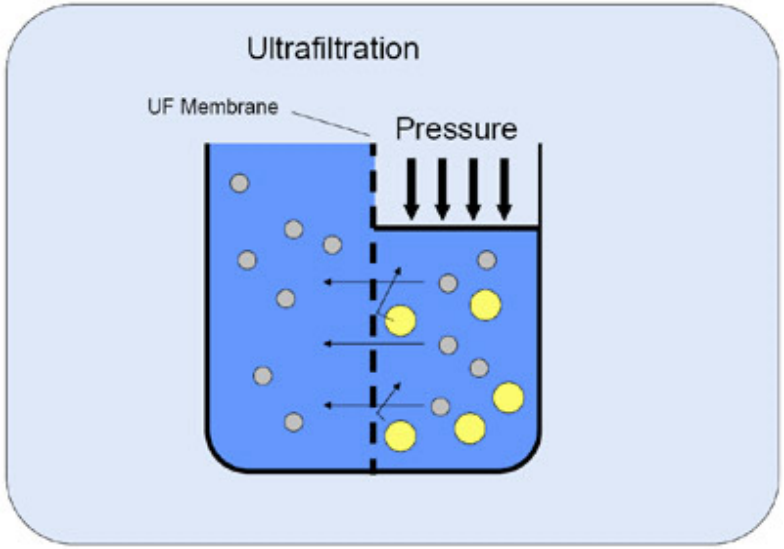


Dissecting a Protein for Study

- A cell has thousands of different proteins
  - Require protein separation
    - Methods applicable in general to other biochemicals
- Two key protein separation methods
  - By size and mass
  - By charge
- Protein separation by size and mass
  - Ultrafiltration
  - Ultracentrifugation
  - Molecular (size) exclusion chromatography

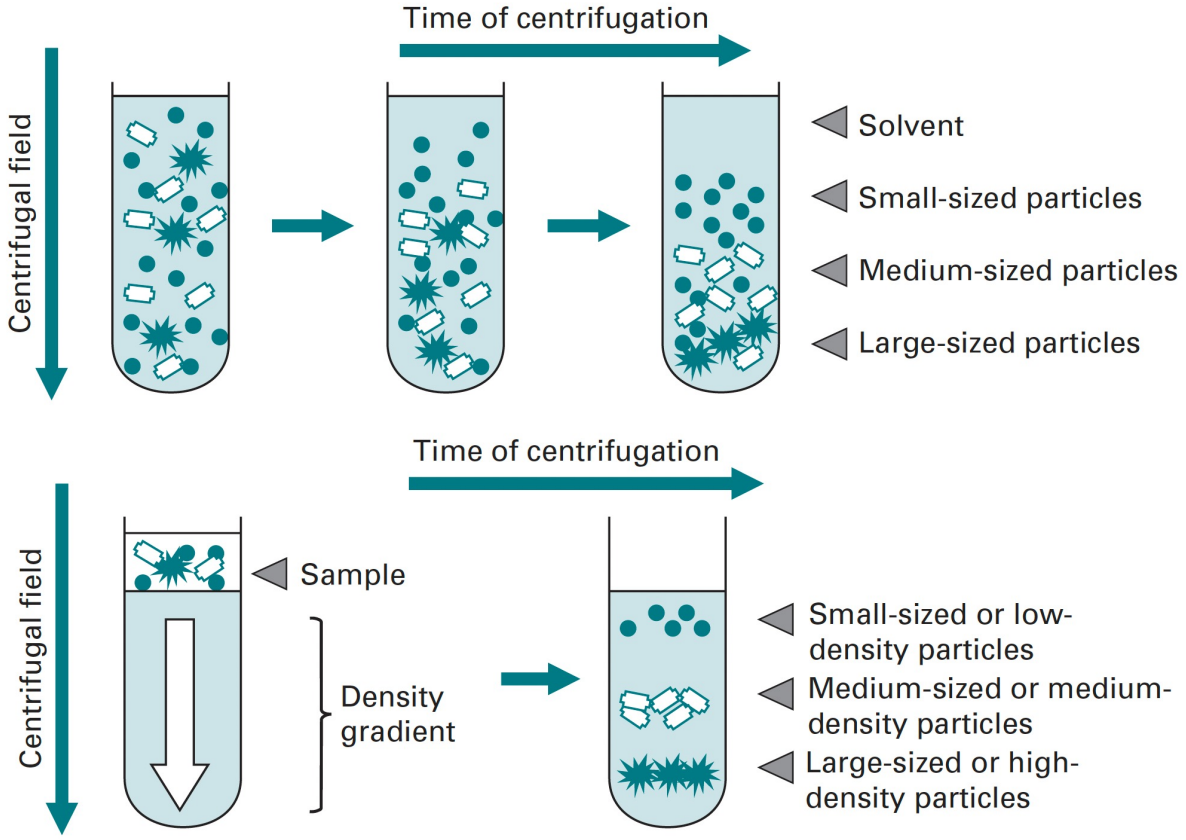
Dissecting a Protein for Study

- Protein separation by size and mass
  - Ultrafiltration
    - The method has limited resolving power
    - Useful when the protein of interest is either particularly large or particularly small



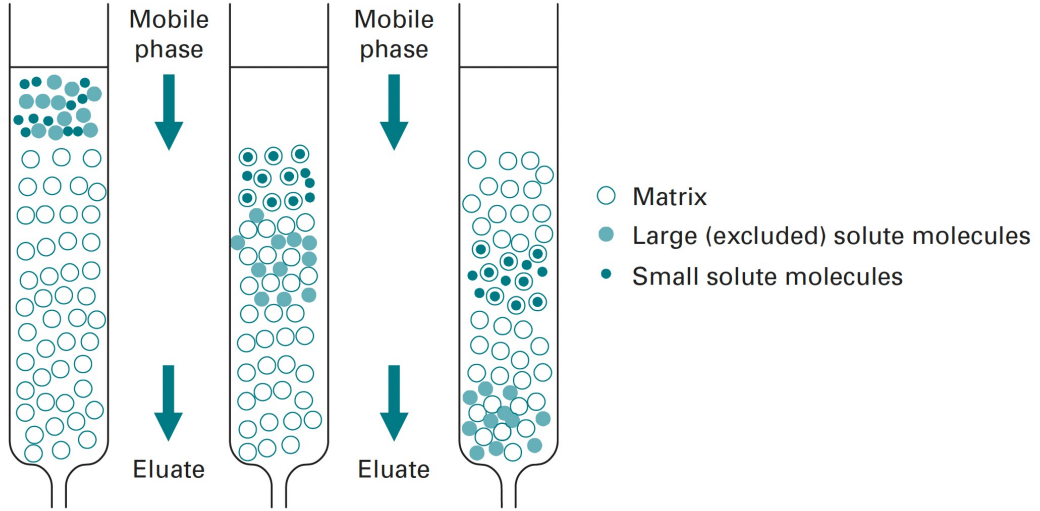
Dissecting a Protein for Study

- Protein separation by size and mass
- Ultracentrifugation
  - Differential and density-gradient sedimentation
  - Heavier or denser macromolecules will sink faster
  - Can be used to determine a protein's molecular mass



Dissecting a Protein for Study

- Protein separation by size and mass
- Molecular (size) exclusion chromatography
  - Based on molecular size and shape
  - Exploits the molecular sieve properties of a variety of porous particles
    - Large molecules that are completely excluded from the pores will pass through the interstitial matrix spaces
      - will appear first in the eluate
    - Smaller analytes will be distributed within the matrix
      - will appear appearing last in the eluate

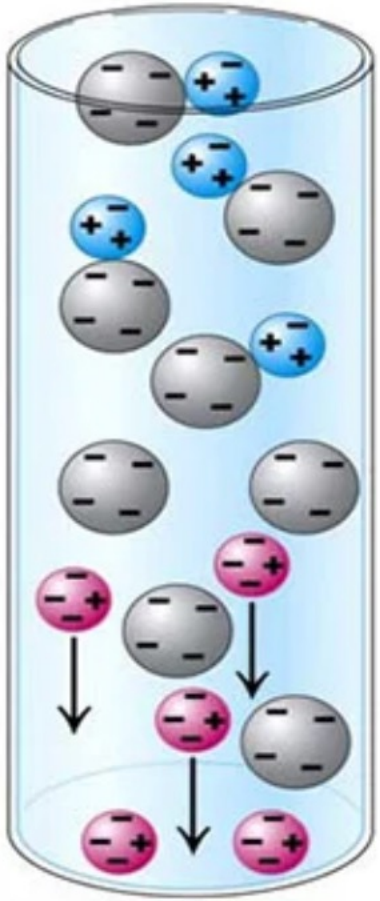


Dissecting a Protein for Study

- Protein separation by charge
  - Ion exchange chromatography
  - Electrophoresis
- All of these methods is pH dependent

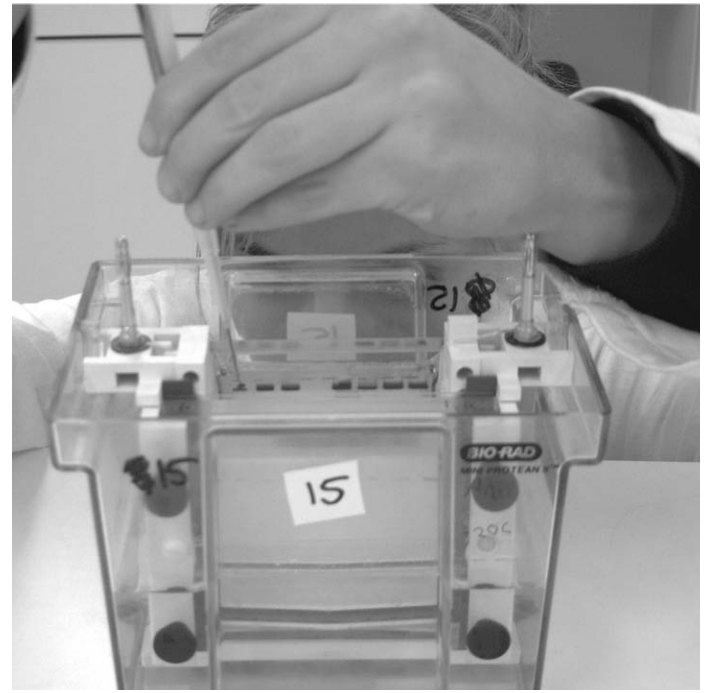
Dissecting a Protein for Study

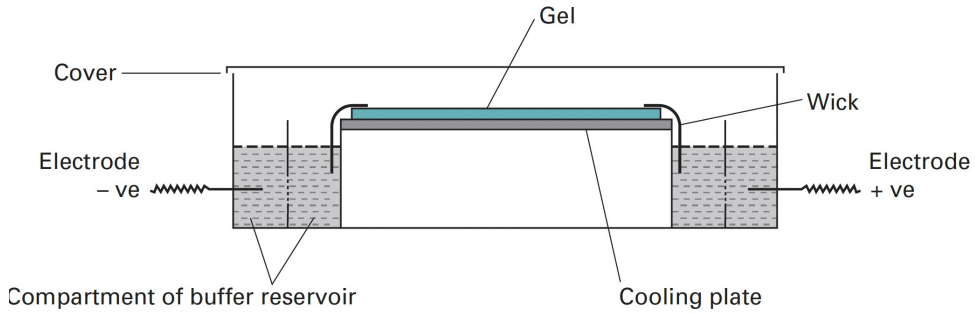
- Protein separation by charge
- Ion exchange chromatography
  - Relies on the attraction between oppositely charged stationary phase (ion exchanger) and analyte
  - High resolving power and high capacity
  - Cation exchangers
    - Possess negatively charged groups and these will attract positively charged cations
  - Anion exchangers
    - Have positively charged groups that will attract negatively charged anions



Dissecting a Protein for Study

- Protein separation by charge
- Electrophoresis
  - Relies on the migration of a charged particle under the influence of an electric field E
  - Under the influence E, these charged particles will migrate either to the cathode or to the anode, depending on the nature of their net charge





Digging into the details: uncovering a protein's primary sequence

- Pure sample protein available (also valid for other biologics)
  1. Separation and purifying the polypeptide chains
     - *e.g.*, proteins with quaternary structures
  2. Cleaving intrachain disulfide bridges
     - Reduction to –SH followed by alkylation to -SR in order to prevent S-S- reformation
  3. Determining amino acid concentration of the protein chain
     - Automatic amino acid analyzer (< 1hr, 1 nmol protein
     - Output → % of each amino acid in the primary structure, NOT the sequence
  4. Identifying the terminal amino acids (many methods)
     - N-terminal → compound (molecule or enzyme) that specifically reacts with the N-terminal, tags it and hydrolyzes the full protein
     - C-terminal → same as above

Digging into the details: uncovering a protein's primary sequence

5. Cleaving polypeptide chain into smaller fragments
   - Use specific enzymes to break each chain into fragments up to 50 amino acids long
   - Separate and purify the fragments
   - Determine the sequence of each fragment via automatic protein sequenator (sequencer)
   - Repeat with a different pattern of cleavage

Digging into the details: uncovering a protein's primary sequence

6. Combining information to get the total sequence

For example: octapeptide

Complete hydrolysis (step 4) → Ala, Asp, Gly, Lys, Phe, Val, and 2 Cys

Partial hydrolysis (step 5) → Gly-Cys, Phe-Val-Gly, Cys-Asp, Lys-Cys, Cys-Asp-Lys, and Cys-Ala

Fragment matching →

Cys-Asp-Lys-Cys

Gly-Cys-Asp-Lys-Cys-Ala

Phe-Val-Gly-Cys-Asp-Lys-Cys-Ala