Amino Acids & Proteins Notes

Hankertrix

April 29, 2025

Contents

1 Definitions

1.1 Proteinogenic

Proteinogenic just means "protein creating", which means an amino acid that is proteinogenic is incorporated into proteins during translation.

1.2 Physiological pH

Physiological pH is about 7.35 - 7.45.

1.3 Non-polar amino acids

Non-polar amino acids have side chains that are either aliphatic or aromatic. For example, proline has an aliphatic cyclic structure, whose amino group is a secondary amine.

1.4 Polar amino acids

Polar amino acids have polar side chains that are neutral (uncharged) neutral pH.

A few examples:

- Serine (**Ser**) and threonine (**Thr**) has a hydroxyl ($-OH$) group, which are good nucleophiles that play a role in enzymatic activity.
- Cysteine (**Cys**) has a thiol ($-SH$) group and two cysteine can oxidise to form a disulfide bond.
- Glutamine (**Gln**) and Asparagine (**Asn**) have an amide group, which do not ionize at physiological pH.

1.5 Acidic amino acids

Acidic amino acids are amino acids that have acidic side chains, which are **negatively charged** at physiological pH. Glutamic acid (**Glu**) and aspartic acid (**Asp**) are the only acidic amino acids.

1.6 Basic amino acids

Basic amino acids refer to amino acids that have basic side chains, which are **positively charged** at physiological pH. Histidine (**His**), lysine (**Lys**) and arginine (**Arg**) are the only basic amino acids. Histidine has an imidazole group, whose pK_a is ~ 6, which is very close to the physiological pH.

1.7 Amphoteric

An amphoteric substance is a substance that can react as either a base or an acid.

1.8 Ampholyte

Ampholyte refers to a molecule consisting of both positively and negatively charged components.

1.9 Zwitterion (zwitterionic form)

A zwitterion refers to a molecule or an ion having separate positively charged and negatively charged groups.

1.10 Isoelectric point (pI)

The isoelectric point refers to the pH at which a molecule has zero net charge.

$$pI = \frac{pK_{a1} + pK_{a2}}{2}$$

1.11 Proteins

Proteins are **polymers** of amino acids, where the amino acids are joined **head-to-tail** by covalent peptide bonds.

The peptide backbone of a protein consists of the repeated sequence $-H - C_\alpha - C_\circ$.

Where:

- "N" is the amide nitrogen of the amino acid
- "C_α" is the alpha carbon of the amino acid
- "$C_\circ$" is the carbonyl carbon of the amino acid

1.12 Peptide bond

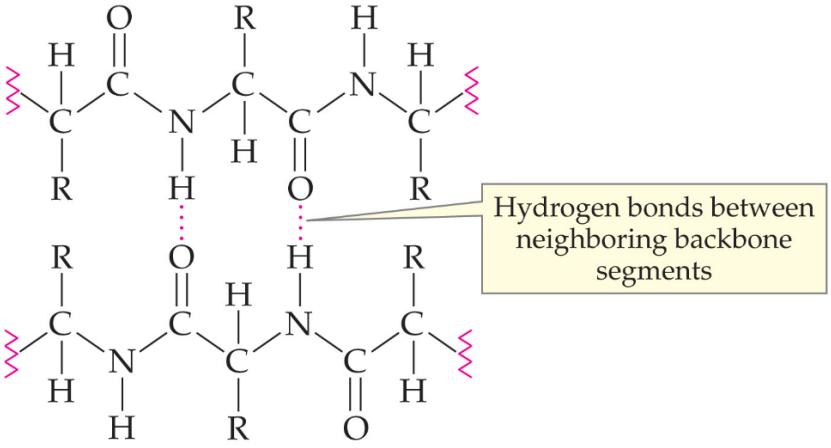
A peptide bond is an **amide** bond formed between the carboxyl group of one amino acid and the amino group of another amino acid.

1.13 Primary structure of a protein (1°)

The primary structure refers to the sequence of building blocks in a polymer. For proteins, it is the **amino acid sequence** of the polypeptide chain. Along the backbone of the protein is chain of alternating **sugar and phosphate groups**. The primary structure of a protein is so crucial to the function of the protein that the change of only one amino acid can drastically alter a protein's biological properties.

1.14 Secondary structure of a protein (2°)

Secondary structures represent the **local arrangement** of the polypeptide in space. The secondary structure refers to regular and repeating structural patterns, such as the α-helix and the β-sheet. The secondary structure is formed by **hydrogen bonding** between the **backbone atoms** in the neighbouring segments of protein chains.



1.15 Tertiary structure of a protein (3°)

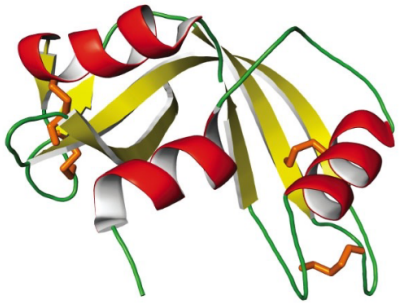
The tertiary structure is the overall **three-dimensional shape** that results from the folding of a single protein chain. The tertiary structure mainly depends on **R group interactions** that are far apart along the entire backbone.

1.15.1 Example 1

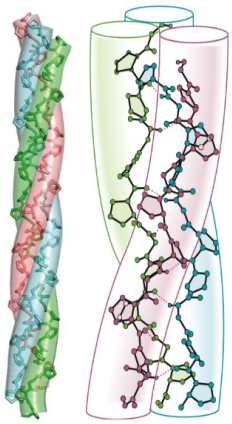
Globular protein

The globular shape has the maximum surface-to-volume ratio.

Soluble in cells.



1.15.2 Example 2



Fibrous protein

Collagen has high % of Gly and Pro.

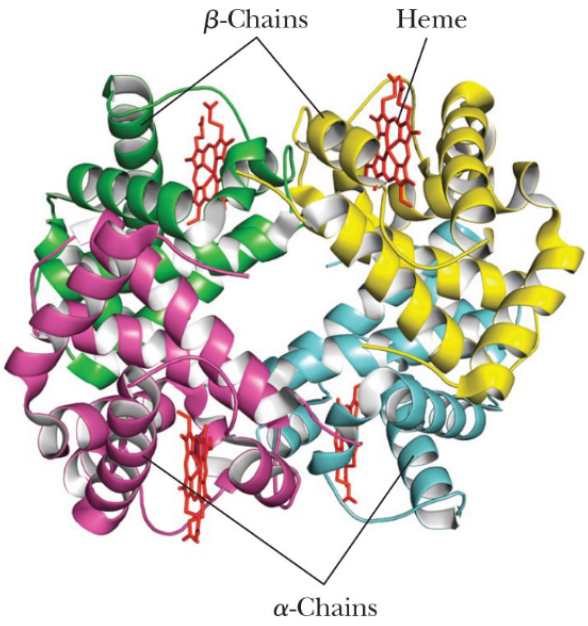
It forms a righthanded triple helix (tropocollagen), with each strand being a left-handed helix.

It differs from a-helix. It is more extended. (~3.3 residues per turn, 2.9A rise per residue)

1.16 Quaternary structure of a protein (4°) (tetramer)

The quaternary structure of a protein consist of two or more interacting polypeptide chains, each of which is referred to as a subunit of the protein.

1.16.1 Example



Hemoglobin is a tetramer consisting of two α and two β polypeptide chains.

1.17 Subunit of a protein

A subunit of a protein is a polypeptide chain that is inside a quaternary structure.

1.18 Racemisation

Racemisation is a process in which optically active compounds consisting of a single enantiomer are converted into a racemic mixture, which is an equal mixture of enantiomers with no optical activity.

1.19 Acid-labile

Acid-labile just means that a compound is easily destroyed in an acidic environment.

1.20 Acid hydrolysis

Acid hydrolysis is the method used to break peptide bonds, as it avoid racemisation and causes less destruction of certain amino acids such as **Ser**, **Thr**, **Arg**, and **Cys**.

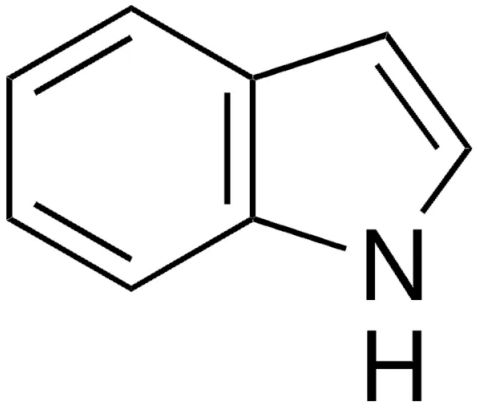
The conditions are typically 6 M HCl at 110 °C.

However, some amino acids are not compatible with acid hydrolysis and using acid hydrolysis will destroy the amino acid.

- **Trp** is not acid-compatible. UV light is used instead.
- **Asn** and **Gln** are also acid-labile, and the side chain amino nitrogen is released as mmonium. **Asn** and **Gln** are converted to **Asp** and **Glu**.
- The amount of NH_4^+ released during acid hydrolysis gives an estimation of the total amount of **Asn** and **Gln**.

1.21 Indole

The indole functional group is the one shown below. It is in the R group of tryptophan, **Tyr**.



2 Structures and properties of amino acids

- There are only 20 proteinogenic amino acids in nature.
- The general structure of amino acids has an amino group ($-NH_3^+$) and a carboxylic acid ($-COO^-$), which are bonded to the alpha carbon (C_α).
- The side chain ($-R$) of amino acid is important for its properties
- All amino acids are chiral except for glycine (G), where the side chain is a hydrogen atom ($R = H$).
- Laevorotatory amino acids are predominant in nature.

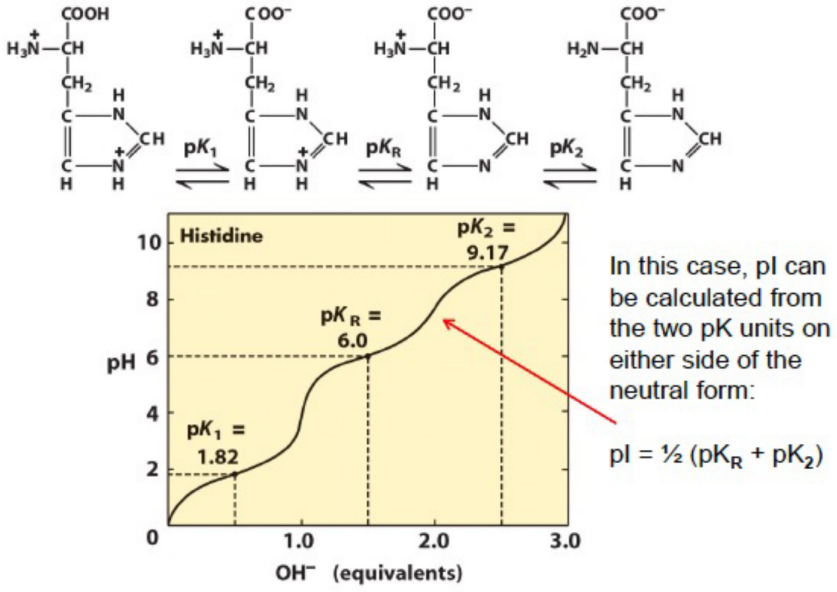
2.1 Spectroscopic properties of amino acids

- Only tryptophan (**Trp**), tyrosine (**Tyr**) and Phenylalanine (**Phe**) absorbs UV light.
- The absorbance at 280 nm is a good diagnostic device for amino acids.

2.2 Acid-base properties of amino acids (amphoteric)

- The pK_a of the carboxylic acid group is about 2
- The pK_a of the amino group is about 2
- Therefore, at physiological pH, both the carboxylic acid group and the amino group will be ionised. This ionised form is called the **zwitterion**.
- **Amino acids are typically written in their zwitterionic form.**

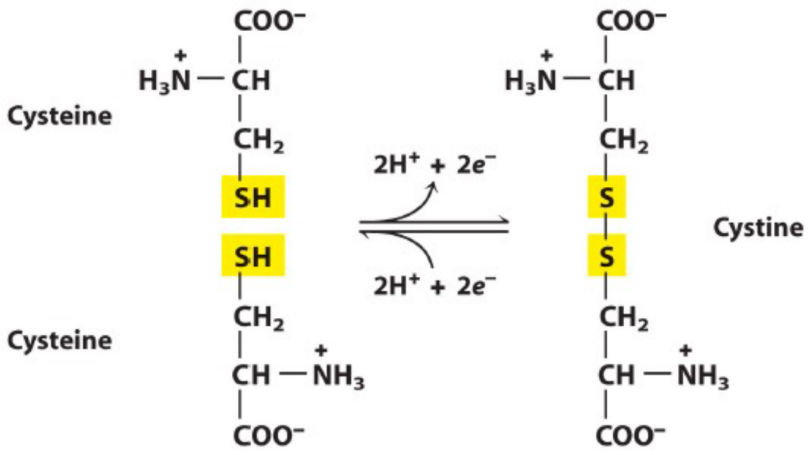
3 Titration of an amino acid



3.1 A useful formula

$$pH = pK + \log\left(\frac{[A^-]}{[HA]}\right)$$

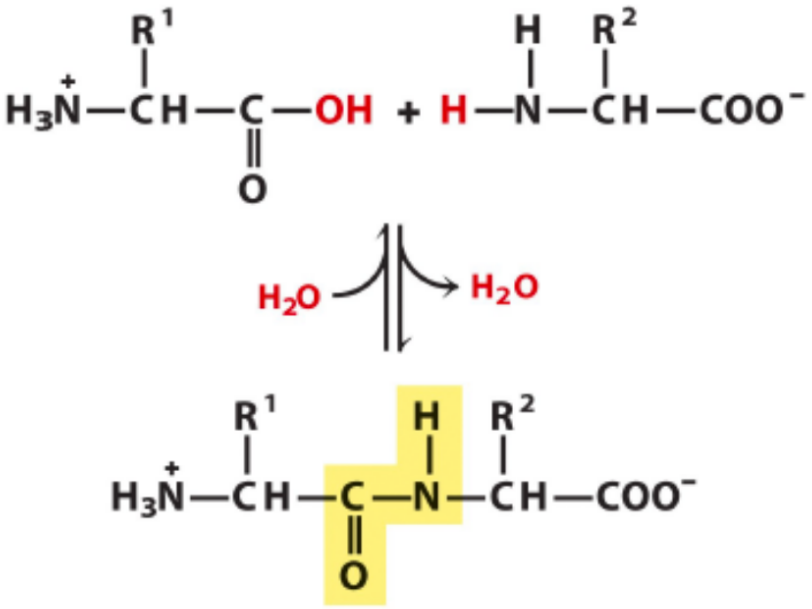
4 Disulfide bond formation by two cysteines



Reversible formation of a disulfide bond by the oxidation of two molecules of cysteine. Disulfide bonds between Cys residues stabilize the structures of many proteins

5 Peptide bonds

5.1 Peptide bond formation



5.2 Properties

- Peptide bonds are **planar** with partial double bond character.
- Due to the **partial double bond character**, the bond length is in between the typical single bond double bond, and the six atoms of the peptide bond groups are always in the **same plane**.
- They are usually found in the *trans* configuration as the *cis* configuration has **steric hindrance**. The major exception is the peptide bonds in the sequence $X - Pro$ where X is any other amino acid. Here, the *cis* configuration is sometimes preferred, but the trans configuration is still favoured overall, with ratio of 4:1.
- The amide $N - H$ group is **partially positive**, and the carbonyl oxygen is partially negative, which results in a net dipole moment. Thus, the peptide bond is polar.

6 Intermolecular interactions in protein structures

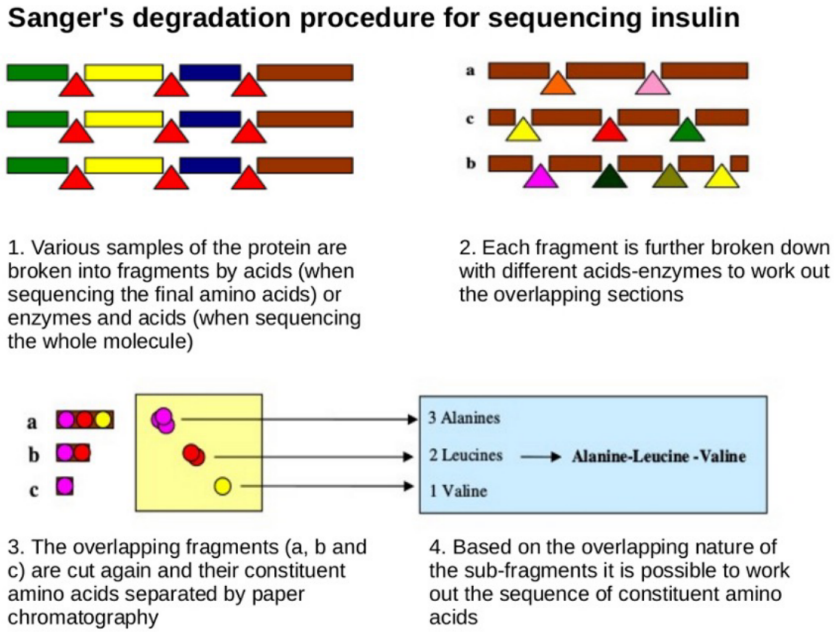
Non-covalent interactions stabilise the higher levels of the protein structure, like the secondary, tertiary and quaternary structures.

- **Hydrogen bonds** are formed whenever possible. These interactions are found on the peptide backbone and the polar residues.
- **Hydrophobic interactions** drive protein folding, and they are usually found on the interior of the proteins and non-polar residues.
- **Ionic interactions** usually occur on the protein surface. An example is the electrostatic interactions between opposite charges or repulsion between like charges.
- **Van der waals** interactions are ubiquitous. An example is the instantaneous dipole-induced dipole interactions.

7 Protein characteristics

- The unique characteristics of each protein is the **distinctive sequence of amino acid residues in its polypeptide chains**.
- The **primary sequence** of proteins is encoded by the nucleotide sequence in DNA.
- A polypeptide chain has two ends, the **N-terminus** and the **C-terminus**.

8 Protein sequencing



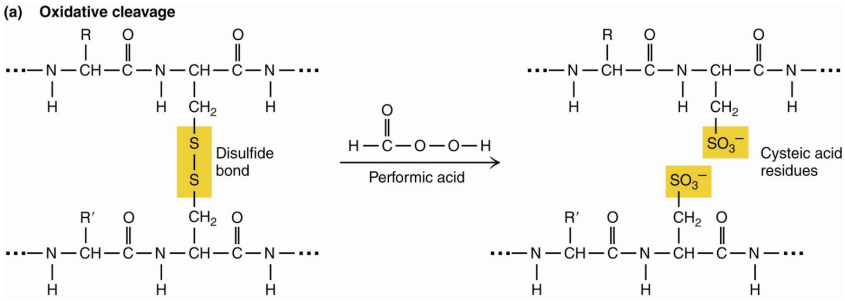
8.1 Step 1

- The interactions between protein subunits depend on weak forces, which are interactions that are not covalent, like instantaneous dipole-induced dipole interactions, permanent dipole-permanent dipole interactions and hydrogen bonding.
- Hence, separation is achieved with:
  - Extreme pH
  - 8 M urea
  - 6 M guanidine HCl
  - High salt concentration, usually ammonium sulfate

8.2 Step 2

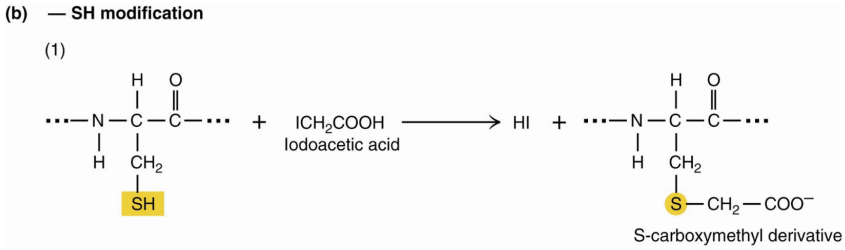
Cleavage of disulfide bridges.

8.2.1 Oxidation using performic acid



8.2.2 Sulfhydryl reducing agnents

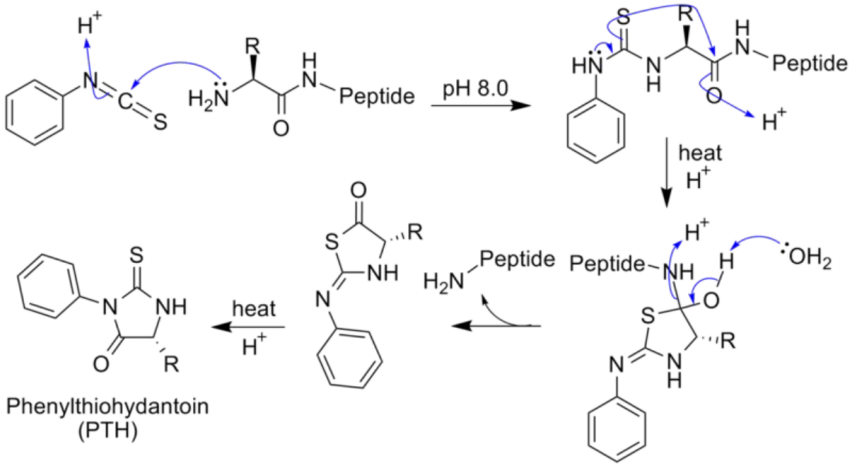
- Examples include **mercaptoethanol**, **bME** or **dithiothreitol**, **DTT**.
- To prevent recombination, alkylating agent like iodoacetate is used.



8.3 Step 3

8.3.1 N-terminal analysis

- Edman's reagent (phenylisothiocyanate, PITC)
- The reagent combines with the N-terminus of a protein, forming derivative phenylthiohydantoin derivative (PTH derivative)
- Sequential Edman degradation is also possible



8.3.2 C-terminal analysis

- Enzymatic analysis (carboxypeptidase)
- **Carboxypeptidase A** cleaves the N-side of any residue at the C-terminal position except **Pro**, **Arg**, **Lys**, **Glu**, **Asp**.
- **Carboxypeptidase B** (hog pancreas) only works on the N-side of **Arg** and **Lys** at the C-terminus.
- **Carboxypeptidase Y** (yeast) works with any residue.

8.3.3 Fragmentation of polypeptide chains

1. Enzymatic fragmentation

    - This is done using **trypsin**, **chymotrypsin**, **clostripain**, **staphylococcal protease**.
    - Trypsin cleaves on the C-side of Lys and Arg.
    - Chymotrypsin cleaves on the C-side of Phe, Tyr and Trp.
    - Clostripain only cleaves on the C-side of Arg.
    - Staphylococcal protease cleaves on the C-side of Asp and Glu.

2. Chemical fragmentation

    - **Cyanogen bromide (CNBr)** acts on the C-side of methionine residues.
    - It is useful as proteins usually only have a few Met residues.
    - The use of cyanogen bromide is indicated by a homoserine lactone at the C-terminal of the peptide. Homoserine lactone refers to the ring structure in the product in the picture below. The presence of that structure means that methionine is present and cyanogen bromide was used.