

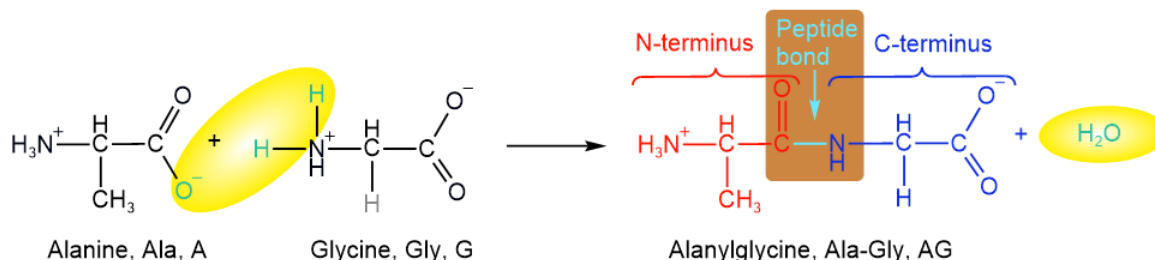
7.3: Primary structure of proteins

Learning Objectives

- Understand a peptide bond and disulfide, their nomenclature, and their characteristics.
- Define and write proteins' primary structure, importance, and related terminologies.

Peptide

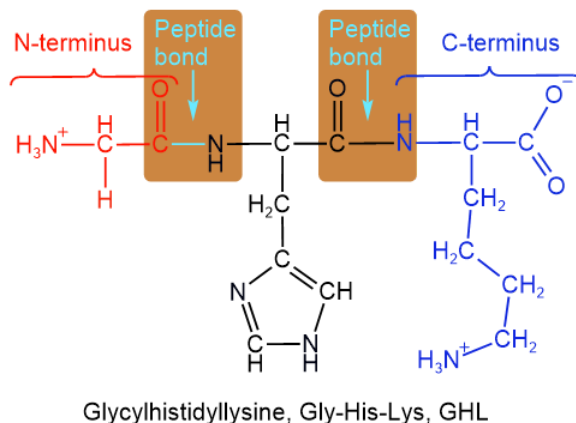
An amine reacts with a carboxylic acid and makes an amide by eliminating water. Amino acids have both an amine and a carboxylic acid group. If the amine group of one amino acid reacts with the carboxylic acid of another, the two amino acids become bonded through an amide bond. For example, alanine and glycine become bonded through an amide bond, as illustrated below.



Peptide bond

An amide bond that links two amino acids is called a **peptide bond** or **peptide linkage**. For example, a peptide bond that links alanine and glycine is highlighted in a brown box in the above illustration.

A single amino acid is also called a **monopeptide**, and two amino acids linked by a peptide bond is called a **dipeptide**. Two amino acids, alanine, and glycine, are monopeptides; their product, alanylglycine, shown above, is a dipeptide. One amino acid in a dipeptide has free ammonium ($-\text{NH}_3^+$) group that can make a peptide with $-\text{COO}^-$ of another amino acid. Similarly, one amino acid in a dipeptide has a free carboxylate ($-\text{COO}^-$) group that can make a peptide bond with $-\text{NH}_3^+$ group of another amino acid. Three amino acids linked by peptide bonds is called a **tripeptide**, four are called **tetrapeptide**, five are **pentapeptide**, six are **hexapeptide**, seven are **heptapeptide**, and so on. A tripeptide of glycine, histidine, and lysine is shown below.



Shorter chains of amino acids linked by peptide bonds are called **peptides**; longer ones are called **polypeptides**. Polypeptides and **proteins** are used interchangeably, but more than 50 amino acid chains are usually called proteins. Amino acids in peptides are usually called **residues**.

Structure of a peptide backbone

The backbone of a peptide chain is $-\text{C}-\text{C}-\text{N}-$ where the middle C is the carbon only $\text{C}=\text{O}$ and $\text{C}-\text{N}$ is the peptide bond. The peptide bond has two resonance contributors, as shown below.



Due to the resonance, the peptide bond has about 40% double bond character. There is no free rotation around the peptide bond due to its double bond character. Therefore, the four groups around the peptide bond (shown in blue color in the above structure) exist in the same plane as in the case of alkenes. Free rotation is around the other σ -bonds in the peptide backbone. Therefore, the structure of the peptide backbone is like cards connected by a swivel at the opposite corners, as shown in Figure 7.3.1.

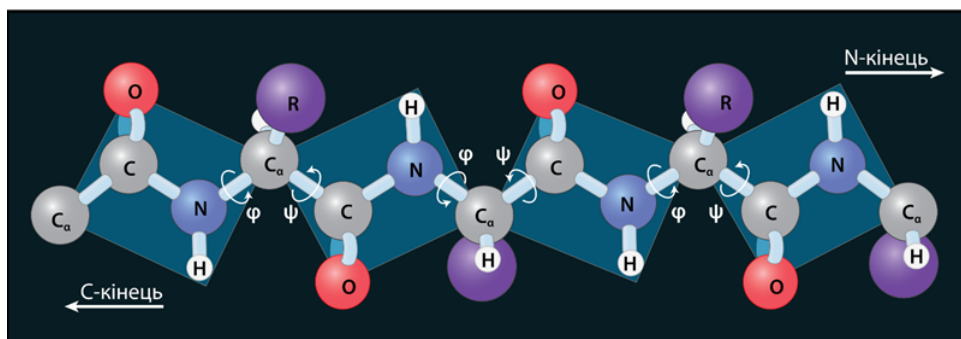
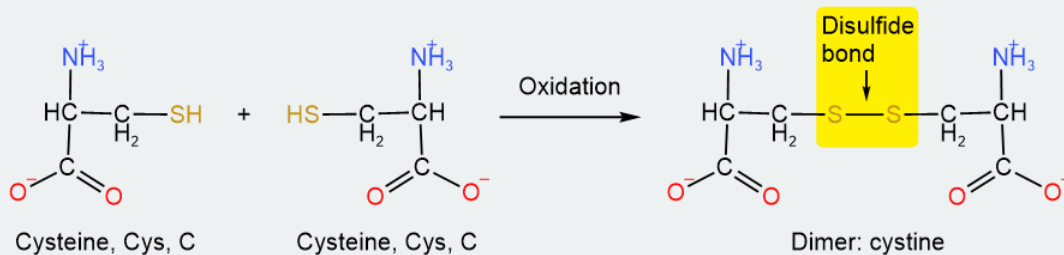


Figure 7.3.1: Illustration of planer structure around peptide bonds as card-like structures and free rotation around other σ -bonds shown by arrows. (Copyright; Zlir'a, CC0, via Wikimedia Commons)

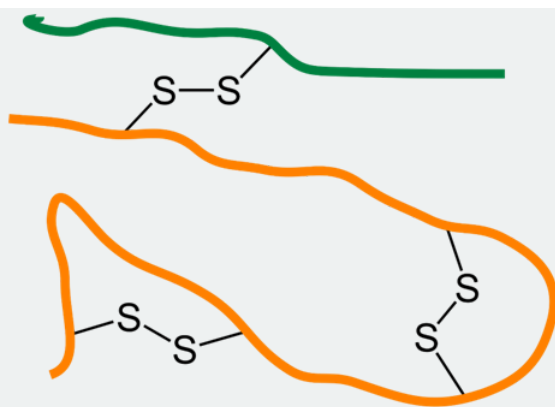
The peptide bond's rigidity limits the peptide backbone's possible orientations, affecting its secondary and tertiary structure. The N—H groups can establish hydrogen bonds with the C=O groups within the same chain or between the neighboring chains that play an essential role in determining the secondary and tertiary structures of proteins that are described in a later section.

Disulfide bond

Amino acid cysteine has a thiol (—SH) group that can easily be oxidized to **disulfide** (—S—S—) bond or **disulfide linkage** linking two cysteines into a dimer called cystine, as illustrated below.



When a cysteine residue makes a disulfide bond with another cysteine residue in the same chain or another chain, it provides a covalent linkage that binds parts of the same chain or two different chains, as shown in Figure 7.3.2. Examples of both types are found in the structure of insulin, which is composed of two polypeptides joined by sulfide linkage, as shown in Figure 7.3.2.



Human Insulin

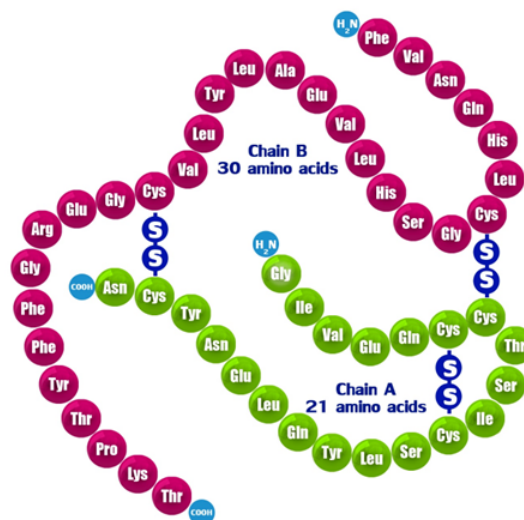


Figure 7.3.2: Illustration of disulfide linkage within the same chain and between different chains of polypeptides (left) and with the example of human insulin (right). (Copyright; Left: Jü, CC0, via Wikimedia Commons, right: Zappys Technology Solutions, CC BY 2.0, via Wikimedia Commons)

Naming peptides

When amino acids combine by peptide bonds, one amino acid on one terminal has free ammonium ($-\text{NH}_3^+$) group, an amino acid on the other terminal has free carboxylate ($-\text{COO}^-$) group, as highlighted in the structures of dipeptide and tripeptides shown above.

N-Terminus and C-terminus of a peptide

- An amino acid in a peptide that has free ammonium ($-\text{NH}_3^+$) group is called **N-terminus**. For example, glycine in the tripeptide shown above is N-terminus.
- An amino acid that has a free carboxylate ($-\text{COO}^-$) group is called **C-terminus**. For example, lysine in the tripeptide shown above is C-terminus.

Amino acids in a peptide are written horizontally from left to right, where N-terminus is the leftmost amino acid, and C-terminus is the rightmost amino acid.

A **peptide is named** by listing the names of its constituent amino acids in a sequence from N-terminus to C-terminus, with the last syllable changed to yl, except for the C-terminus. For example, the dipeptide of alanine and glycine is alanyl^{yl}glycine, and the tripeptide of glycine, histidine, and lysine is glycyl^{yl}histidyl^{yl}lysine.

Often three-letter abbreviations of the amino acids in a peptide are written in a sequence from N-terminus to C-terminus, separated by hyphens. For example, the dipeptide alanyl^{yl}glycine can be written as Ala-Gly, and the tripeptide glycyl^{yl}histidyl^{yl}lysine as Gly-His-Lys. For polypeptides, one-letter abbreviations of the amino acid residues are usually written in a sequence from N-terminus to C-terminus. For example, dipeptide Ala-Gly is AG, and tripeptide Gly-His-Lys is GHL.

What is the primary structure of proteins?

Primary structure of proteins

The primary structure of peptides or proteins is the sequence of amino acids linked together by peptide bonds. For example, the primary structures of the dipeptide and tripeptides shown above are Ala-Gly and Gly-His-Lys.

The primary structure of a protein is shown as a sequence of amino acids written from the N-terminus to the C-terminus. When the sequence of amino acids is known, three-letter abbreviations are separated by hyphens, e.g., Gly-His-Lys. When the sequence of amino acids in a peptide is not known, the three-letter abbreviations of the constituent amino acid are listed, separated by commas. For example, Ala, Gly could mean Ala-Gly or Gly-Ala, which are different compounds with different properties related to each other as constitutional isomers. Similarly, Gly, His, Lys could mean any one of the following six constitutional isomers: Gly-His-Lys, Gly-Lys-His, His-Lys-Gly, His-Gly-Lys, Lys-Gly-His, or Lys-His-Gly.

The number of constitutional isomers increases exponentially as the number of amino acids in the peptide increases. Constitutional isomers of a polypeptide of n amino acids chosen from 20 amino acids commonly found in proteins are given by 20^n . For example, a polypeptide containing 60 amino acids selected from 20 amino acids found in proteins may have 20^{60} , i.e., 10^{78} , which is an enormous number of possibilities. This analysis shows that the number of proteins that can be synthesized using 20 amino acids is enormously large. An analog is the entire English language composed of letters from 26 different alphabets.

When the primary structure of a polypeptide is modified, its function is affected. The extent of the effect depends on the number of amino acids replaced and their nature. For example, human insulin comprises two peptides: chain A of 21 amino acids and chain B of 30 amino acids, two chains joined by disulfide linkages, as shown in Figure 7.3.1. The amino acids at positions 8, 9, and 10 in chain A (-Thr-Ser-Ile-) and position 30 in chain B (-Thr) in human insulin are replaced with -Ala-Ser-Val- and -Ala, respectively, in bovine insulin, but the two perform the same function. Humans can use bovine insulin, though it is less effective in humans and sometimes causes an allergic reaction.

As shown below, vasopressin and oxytocin are two nonapeptides that differ in two amino acid residues at positions 3 and 8. Their cysteine residues form a disulfide bond, and the carboxylate group ($-\text{COO}^-$) on C-terminus is converted to a primary amide ($-\text{CONH}_2$).

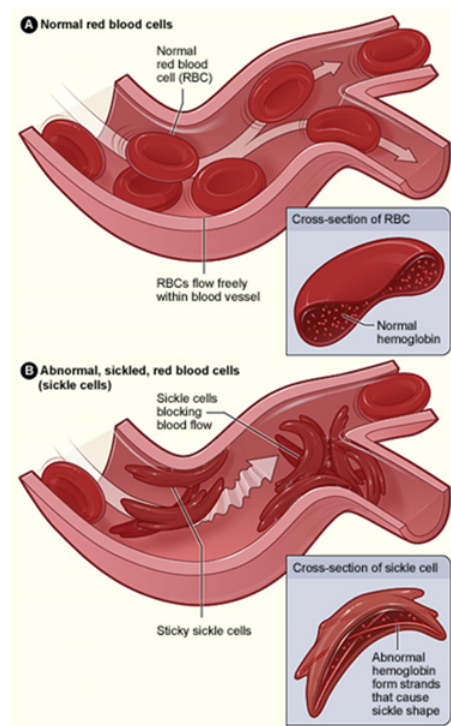


Both vasopressin and oxytocin are hormones the pituitary gland produces but have different functions. Vasopressin is an antidiuretic hormone that regulates blood pressure by adjusting the water reabsorbed by the kidneys. Oxytocin stimulates uterine contractions in labor.

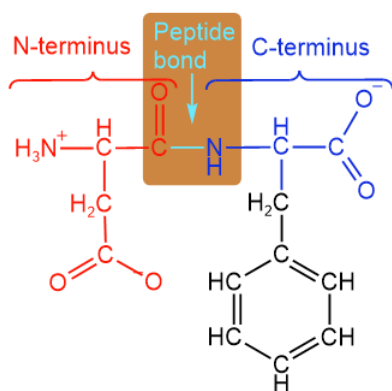
Another example where a slight change in the primary structure of a protein alters its function significantly is sickle cell anemia. Hemoglobin in the red blood cell is responsible for carrying oxygen. Hemoglobin comprises four polypeptides, two alpha chains, and two beta chains containing 574 amino acid residues. A change of glutamic acid (a hydrophilic amino acid) with valine (a hydrophobic amino acid) in the sixth position of the two beta-chains changes the structure of hemoglobin so much that it causes the red blood cells to change from a rounded shape to a sickle body. Sickled cells do not function properly and block the blood flow - a medical condition called sickle cell anemia, illustrated in The Figure on the right.

Some examples and uses of peptides

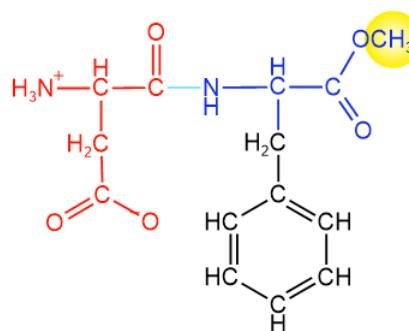
Peptides have different functions in the body of living things. Some of them are commercially important also. For example, dipeptide Ala-Gly shown is a dietary supplement. Aspartame, 200 times sweeter than sucrose and used as a sugar substitute, is a methyl ester of a dipeptide Asp-Phe, shown below.



The National Heart, Lung, and Blood Institute (NHLBI), Public domain, via Wikimedia Commons



Aspartylphenylalanine, Asp-Phe, DF



Asparame

The tripeptide Gly-His-Lys shown before is a human copper-binding peptide with wound healing and skin remodeling activity. Met-enkephalin (Tyr-Gly-Gly-Phe-Met) is a pentapeptide related to pain signals in the body.

Vasopressin and oxytocin are two nonapeptide hormones produced by the pituitary gland. Vasopressin is an antidiuretic hormone that regulates blood pressure by adjusting the water reabsorbed by the kidneys. Oxytocin stimulates uterine contractions in labor.

Substance P (SP) is an undecapeptide (Arg-Pro-Lys-Pro-Gln-Gln-Phe-Phe-Gly-Leu-Met) that is a neuropeptide acting as a neurotransmitter and as a neuromodulation. Insulin is a combination of two polypeptides linked by disulfide linkage that regulates glucose in the blood. Malfunctioning of insulin causes diabetes.

This page titled [7.3: Primary structure of proteins](#) is shared under a [Public Domain](#) license and was authored, remixed, and/or curated by [Muhammad Arif Malik](#).