Introduction to proteins and amino acids

Different types of proteins. The structure and properties of amino acids. Formation of peptide bonds.

Introduction

We tend to think of protein as a mass noun: a homogeneous substance, something that your diet should contain in a certain proportion. But if you ever work in a molecular biology lab (say, for a summer internship), protein may start to look very different to you.

How so? Well, you may see first-hand that protein isn't just a single substance. Instead, there are lots and lots of different proteins in an organism, or even in a single cell. They come in every size, shape, and type you can imagine, and each one has a unique and specific job. Some are structural parts, giving cells shape or helping them move. Others act as signals, drifting between cells like messages in a bottle. Still others are metabolic enzymes, putting together or snapping apart biomolecules needed by the cell. And, odds are, one of these unique molecular players will become *yours* for the duration of your research!

Proteins are among the most abundant organic molecules in living systems and are way more diverse in structure and function than other classes of macromolecules. A single cell can contain thousands of proteins, each with a unique function. Although their structures, like their functions, vary greatly, all proteins are made up of one or more chains of amino acids. In this article, we will look in more detail at the building blocks, structures, and roles of proteins.

Types and functions of proteins

Proteins can play a wide array of roles in a cell or organism. Here, we'll touch on a few examples of common protein types that may be familiar to you, and that are important in the biology of many organisms (including us).

Enzymes

Enzymes act as catalysts in biochemical reactions, meaning that they speed the reactions up. Each enzyme recognizes one or more substrates, the molecules that serve as starting material for the reaction it catalyses. Different enzymes participate in different types of reactions and may break down, link up, or rearrange their substrates.

One example of an enzyme found in your body is salivary amylase, which breaks amylose (a kind of starch) down into smaller sugars. The amylose doesn't taste very sweet, but the smaller sugars do. This is why starchy foods often taste sweeter if you chew them for longer: you're giving salivary amylase time to get to work.

Hormones

Hormones are long-distance chemical signals released by endocrine cells (like the cells of your pituitary gland). They control specific physiological processes, such as growth, development, metabolism, and reproduction. While some hormones are steroid-based (see the article on lipids), others are proteins. These protein-based hormones are commonly called peptide hormones.

For example, insulin is an important peptide hormone that helps regulate blood glucose levels. When blood glucose rises (for instance, after you eat a meal), specialized cells in the pancreas release insulin. The insulin binds to cells in the liver and other parts of the body, causing them to take up the glucose. This process helps return blood sugar to its normal, resting level.

Some additional types of proteins and their functions are listed in the table below:

Protein types and functions

Role	Examples	Functions
Digestive enzyme	Amylase, lipase, pepsin	Break down nutrients in food into small pieces that can be readily absorbed
Transport	Haemoglobin	Carry substances throughout the body in blood or lymph
Structure	Actin, tubulin, keratin	Build different structures, like the cytoskeleton
Hormone signalling	Insulin, glucagon	Coordinate the activity of different body systems
Defence	Antibodies	Protect the body from foreign pathogens
Contraction	Myosin	Carry out muscle contraction
Storage	Legume storage proteins, egg white (albumin)	Provide food for the early development of the embryo or the seedling

Table modified from OpenStax College, Biology.

Proteins come in many different shapes and sizes. Some are globular (roughly spherical) in shape, whereas others form long, thin fibres. For example, the haemoglobin protein that carries oxygen in the blood is a globular protein, while collagen, found in our skin, is a fibrous protein.

A protein's shape is critical to its function, and, as we'll see in the next article, many different types of chemical bonds may be important in maintaining this shape. Changes in temperature and pH, as well as the presence of certain chemicals, may disrupt a protein's shape and cause it to lose functionality, a process known as **denaturation**.

Amino acids

Amino acids are the monomers that make up proteins. Specifically, a protein is made up of one or more linear chains of amino acids, each of which is called a **polypeptide**. (We'll see where this name comes from a little further down the page.) There are 20 types of amino acids commonly found in proteins.

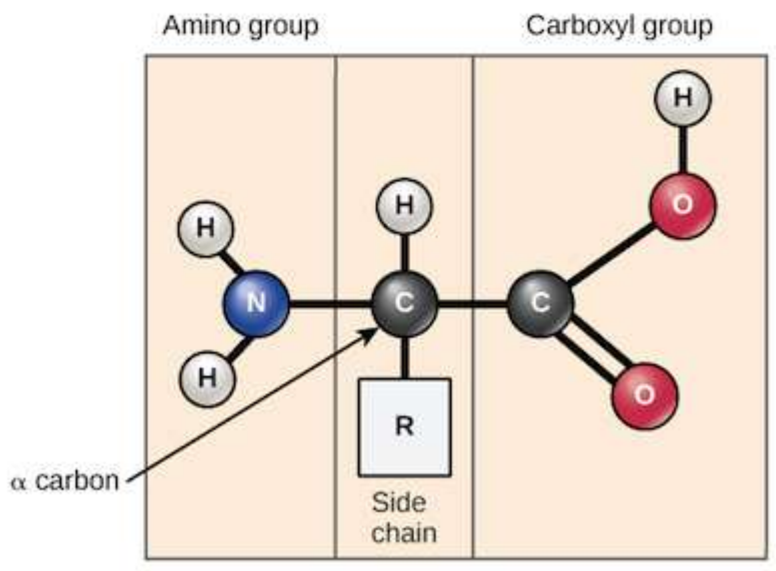


Image of an amino acid, indicating the amino group, carboxyl group, alpha carbon, and R group.

Image credit: OpenStax Biology.

Amino acids share a basic structure, which consists of a central carbon atom, also known as the alpha (α) carbon, bonded to an amino group (NH_2), a carboxyl group ($COOH$), and a hydrogen atom.

Although the generalized amino acid shown above is shown with its amino and carboxyl groups neutral for simplicity, this is not actually the state in which an amino acid would typically be found. At physiological pH (7.2- 7.4), the amino group is typically protonated and bears a positive charge, while the carboxyl group is typically deprotonated and bears a negative charge.

Every amino acid also has another atom or group of atoms bonded to the central atom, known as the R group, which determines the identity of the amino acid. For instance, if the R group is a hydrogen atom, then the amino acid is glycine, while if it's a methyl (CH_3) group, the amino acid is alanine. The twenty common amino acids are shown in the chart below, with their R groups highlighted in blue.

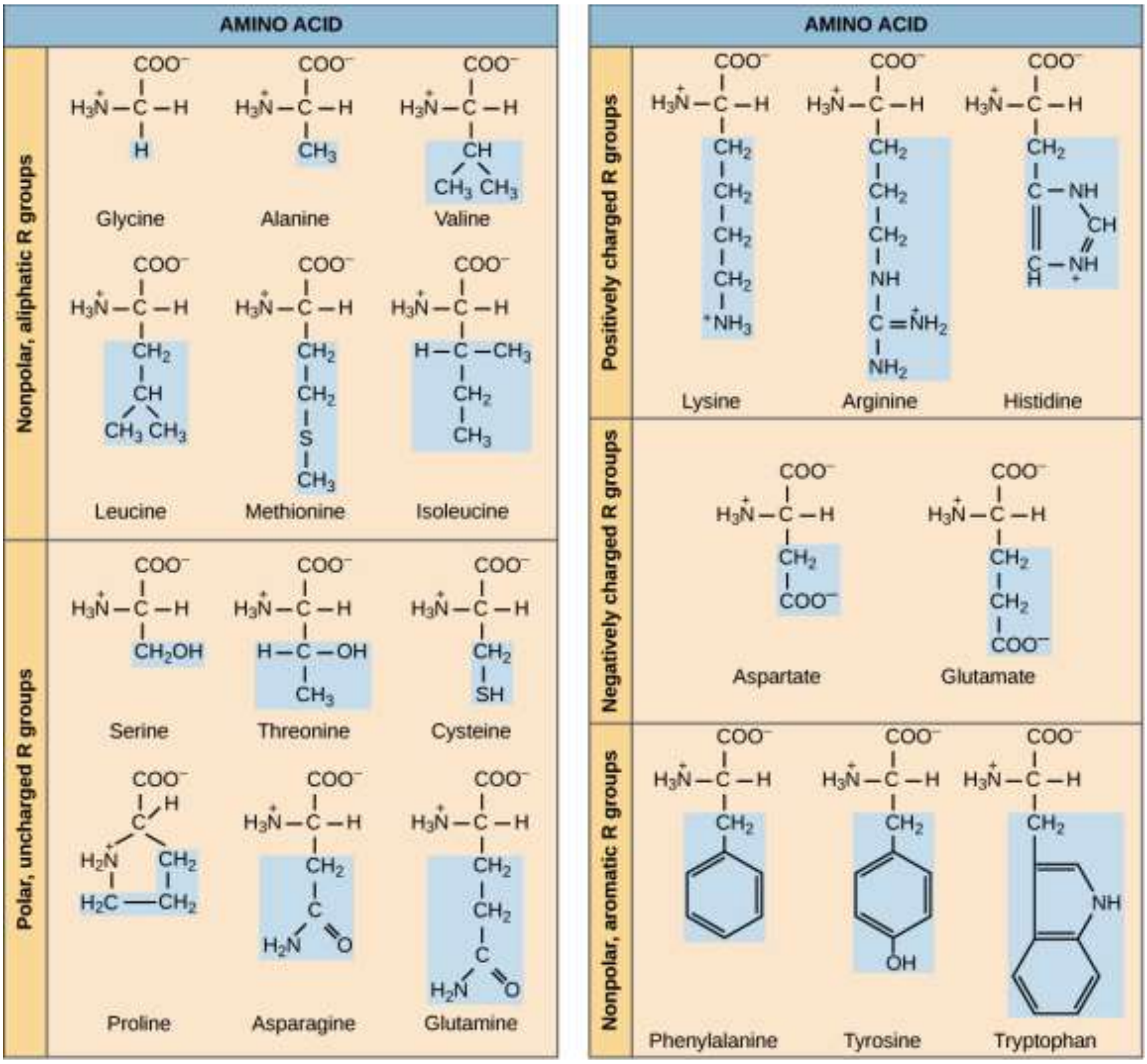


Chart depicting the 20 common amino acids in their predominant protonation forms at physiological pH (7.2-7.4).

Image modified from OpenStax Biology.

The properties of the side chain determine an amino acid's chemical behaviour (that is, whether it is considered acidic, basic, polar, or nonpolar). For example, amino acids such as valine and leucine are nonpolar and hydrophobic, while amino acids like serine and glutamine have hydrophilic side chains and are polar. Some amino acids, such as lysine and arginine, have side chains that are positively charged at physiological pH and are considered basic amino acids. (Histidine is sometimes put in this group too, although it is mostly deprotonated at physiological pH.) Aspartate and glutamate, on the other hand, are negatively charged at physiological pH and are considered acidic.

A few other amino acids have R groups with special properties, and these will prove to be important when we look at protein structure:

- Proline has an R group that's linked back to its own amino group, forming a ring structure. This makes it an exception to the typical structure of an amino acid, since it no longer has the standard NH_3 amino group. If you think that ring structure looks a little awkward, you're right: proline often causes bends or kinks in amino acid chains.

- Cysteine contains a thiol (-SH) group and can form covalent bonds with other cysteine's. We'll see why this is important to protein structure and function in the article on orders of protein structure

Finally, there are a few other "non-canonical" amino acids that are found in proteins only under certain conditions.

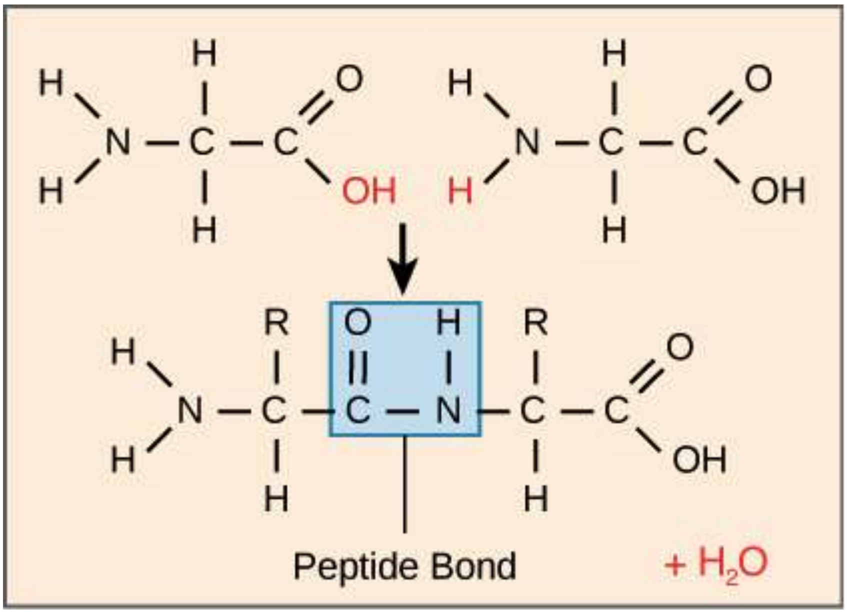
You may not need to know about these for your biology class, but they're pretty darn interesting!

- One is ***N*-formylmethionine** (abbreviated fMet), which is often the first amino acid of protein chains produced by bacteria. It's similar in structure to methionine, but with an oxygen-containing group added to the amino group[1].
- Another is **pyrrolysine**, which looks similar to lysine but has an extra ring-containing group attached to the end of the R group. Pyrrolysine is found mostly in the proteins of microorganisms called methanogenic (methane-producing) archaea[2]Finally, **selenocysteine** is found in many different types of organisms, including humans, and is used to build selenium-containing proteins. Selenium is highly reactive, so this amino acid must be handled with care: it is specified in the genetic code, and added to protein chains, in a different way than a normal amino acid. Structurally, selenocysteine looks just like cysteine, but with a –SeH group in place of the –SH group[4].

Peptide bonds

Each protein in your cells consists of one or more polypeptide chains. Each of these polypeptide chains is made up of amino acids, linked together in a specific order. A polypeptide is kind of like a long word that is "spelled out" in amino acid letters[4]. The chemical properties and order of the amino acids are key in determining the structure and function of the polypeptide, and the protein its part of. But how are amino acids actually linked together in chains?

The amino acids of a polypeptide are attached to their neighbours by covalent bonds known as a **peptide bonds**. Each bond forms in a dehydration synthesis (condensation) reaction. During protein synthesis, the carboxyl group of the amino acid at the end of the growing polypeptide chain reacts with the amino group of an incoming amino acid, releasing a molecule of water. The resulting bond between amino acids is a peptide bond



Peptide bond formation between two amino acids. In a peptide bond, the carbonyl C of one amino acid is connected to the amino N of another.

Image modified from OpenStax Biology.

Because of the structure of the amino acids, a polypeptide chain has **directionality**, meaning that it has two ends that are chemically distinct from one another. At one end, the polypeptide has a free amino group, and this end is called the **amino terminus** (or N-terminus). The other end, which has a free carboxyl group, is known as the **carboxyl terminus** (or C-terminus). The N-terminus is on the left and the C-terminus is on the right for the very short polypeptide shown above.

How do we go from the amino acid sequence of a polypeptide to the three-dimensional structure of a mature, functional protein? To learn how interactions between amino acids cause a protein to fold into its mature shape, I highly recommend the video on orders of protein structure.