

2: Proteins Structure: from Amino Acid Sequence to Three Dimensional Structure

Proteins are macromolecules. They are constructed from one or more unbranched chains of amino acids; that is, they are polymers. An average eukaryotic protein contains around 500 amino acids but some are much smaller (the smallest are often called **peptides**) and some much larger (the largest to date is titin a protein found in skeletal and cardiac muscle; one version contains 34,350 amino acids in a single chain!).

Every function in the living cell depends on proteins.

- Motion and locomotion of cells and organisms depends on proteins. [Examples: Muscles, Cilia and Flagella]
- The catalysis of all biochemical reactions is done by **enzymes**, which contain protein.
- The structure of cells, and the extracellular matrix in which they are embedded, is largely made of protein. [Examples: Collagens] (Plants and many microbes depend more on carbohydrates, e.g., cellulose, for support, but these are synthesized by enzymes.)
- The transport of materials in body fluids depends of proteins.
- The receptors for hormones and other signaling molecules are proteins.
- Proteins are an essential nutrient for heterotrophs.
- The transcription factors that turn genes on and off to guide the differentiation of the cell and its later responsiveness to signals reaching it are proteins.
- and many more — **proteins are truly the physical basis of life.**

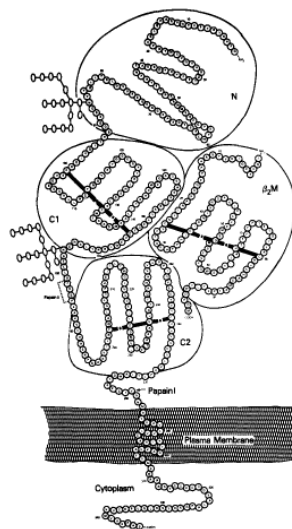


Figure 2.1: Structure of protein molecules

The protein consists of two polypeptide chains, a long one on the left of 346 amino acids — it is called the **heavy chain** — and a short one on the right of 99 amino acids. The heavy chain is shown as consisting of 5 main regions or **domains**:

- three extracellular domains, designated here as N (includes the **N-terminal**), C1, and C2;
- a transmembrane domain where the polypeptide chain passes through the plasma membrane of the cell;
- a cytoplasmic domain (with the **C-terminal**) within the cytoplasm of the cell.

Because it is anchored in the plasma membrane of the cell, the heavy chain is called an **integral membrane protein**.

To the right is the protein molecule called **beta-2 microglobulin**. It is not attached to the heavy chain by any covalent bonds, but rather by a number of noncovalent interactions like hydrogen bonds. Proteins associated non-covalently with integral membrane proteins are called **peripheral membrane proteins**.

The dark bars represent disulfide (S—S) bridges linking portions of each external domain (except the N domain). However, the bonds in S—S bridges are no longer than any other covalent bond, so if this molecule could be viewed in its actual tertiary (3D) configuration, we would find that the portions of the polypeptide chains containing the linked Cys are actually close together.

The two objects on the left of the image that look like candlesticks represent short, branched chains of sugars. The base of each is attached to an **asparagine (N)**. Proteins with covalently linked carbohydrate are called **glycoproteins**. When the carbohydrate is linked to asparagine, it is said to be "**N-linked**". The presence of sugars on the molecule makes this region hydrophilic.

The amino acids exposed at the surface of the extracellular domains tend to be **hydrophilic** as well. However, most of the amino acids in the transmembrane domain are **hydrophobic**, and the amino acids in the cytoplasmic domain are hydrophilic, which is appropriate for the aqueous medium of the cytosol, but carbohydrate is not found in the intracellular domains of integral membrane proteins.

The regions marked "Papain" represent the places on the long chain that are attacked by the proteinase papain (and made it possible to release the extracellular domains from the plasma membrane for easier analysis). This molecule represents a "single-pass" transmembrane protein; the polypeptide chain traverses the plasma membrane once only. However, many transmembrane proteins pass through several, but always a precisely defined number, of times.

Contributors

Contributed by [John W. Kimball](#)

Professor (retired) at [Tufts University & Harvard](#)

Contributed by [Henry Jakubowski](#)

Professor (Chemistry) at [College of St. Benedict/St. John's University](#)

This page titled [2: Proteins Structure: from Amino Acid Sequence to Three Dimensional Structure](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Henry Jakubowski](#).