Protein Structure, Purification, Characterisation and Function Analysis

K. M. RAJKOTIA
Assistant Professor,
Chemistry Dept.
M. M. Science College, Morbi
Gujarat (India)

Abstract
Proteins are formed by the condensation of the a-amino group of one amino acid with the a-carboxyl of the adjacent amino acid. With the exception of the two terminal amino acids, therefore, the a-amino and carboxyl groups are all involved in peptide bonds and are no longer ionisable in the protein. Amino, carboxyl, imidazolyl, guanidino, phenolic and sulphydryl groups in the side chains are, however, free to ionise and of course there will be many of these. Proteins fold in such a manner that the majority of these ionisable groups are on the outside of the molecule, where they can interact with the surrounding aqueous medium. Some of these groups are located within the structure and may be involved in electrostatic attractions that help to stabilise the three-dimensional structure of the protein molecule. The relative numbers of positive and negative groups in a protein molecule influence aspects of its physical behaviour, such as solubility and electrophoresis mobility.

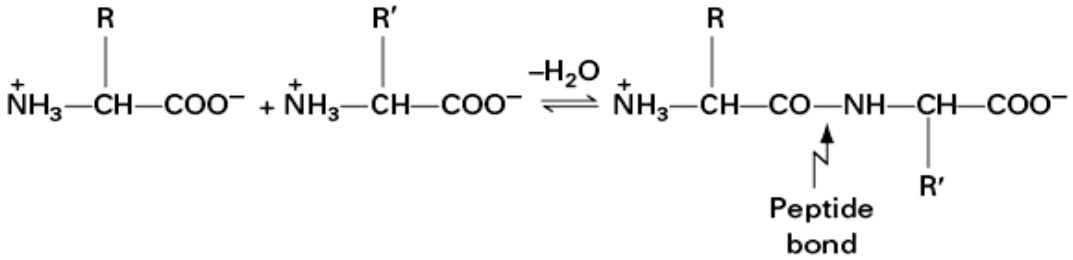
Keywords: *Protein structure, Protein purification, Protein structure determination*

1. Introduction

Twenty amino acids varying in size, shape, charge and chemical reactivity are found in proteins and each has at least one codon in the genetic code. Nineteen of the amino acids are a-amino acids (i.e. the amino and carboxyl groups are attached to the carbon atom that is adjacent to the carboxyl group) with the general formula RCH (NH2) COOH, where R is an aliphatic, aromatic or heterocyclic group. The only exception to this general formula is proline, which is an imino acid in which the -NH2 group is incorporated into a five-member ring. With the exception of the simplest amino acid glycine (R¼H), all the amino acids found in proteins contain one asymmetric carbon atom and hence are optically active and have been found to have the L configuration.

2. Protein Structure

Proteins are formed by condensing the a-amino group of one amino acid or the imino group of proline with the a-carboxyl group of another, with the concomitant loss of a molecule of water and the formation of a peptide bond.



As an alternative to possessing a second amino or carboxyl group, an amino acid side chain may contain in the R of the general formula a quite different chemical group that is also capable of ionising at a characteristic pH. Such groups include a phenolic group (tyrosine), guanidino group (arginine), imidazolyl group (histidine) and sulphydryl group (cysteine) in Table 1. It is clear that the state of ionisation of the main groups of amino acids (acidic, basic, neutral) will be grossly different at a particular pH. Moreover, even within a given group there will be minor differences due to the precise nature of the R group. These differences are exploited in the electrophoresis and ion-exchange chromatographic separation of mixtures of amino acids such as those present in a protein hydrolysate. In practice, protein molecules are always studied in buffered solutions, so it is the isoelectric point that is important. It is the pH at which, for example, the protein has minimum solubility, since it is the point at which there is the greatest opportunity for attraction between oppositely charged groups of neighbouring molecules and consequent aggregation and easy precipitation.

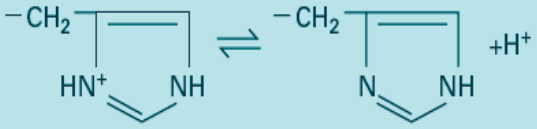
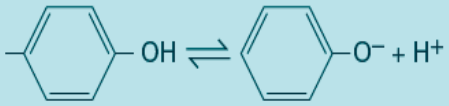
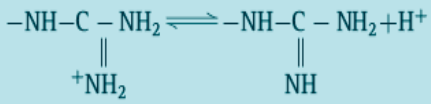
Amino acid group	pH-dependent ionisation	Approx. pK_a
N-terminal α-amino	$-NH_3 \rightleftharpoons NH_2 + H^+$	8.0
C-terminal α-carboxyl	$-COOH \rightleftharpoons COO^- + H^+$	3.0
Asp-β-carboxyl	$-CH_2COOH \rightleftharpoons CH_2COO^- + H^+$	3.9
Glu-γ-carboxyl	$-(CH_2)_2COOH \rightleftharpoons (CH_2)_2COO^- + H^+$	4.1
His-imidazolyl		6.0
Cys-sulphydryl	$-CH_2SH \rightleftharpoons -CH_2S^- + H^+$	8.4
Tyr-phenolic		10.1
Lys-ε-amino	$-(CH_2)_4\overset{+}{N}H_3 \rightleftharpoons -(CH_2)_4NH_2 + H^+$	10.3
Arg-guanidino		12.5

Table. 1 Ionisable Group Found in Proteins

Quaternary structure is restricted to oligomeric proteins, which consist of the association of two or more polypeptide chains held together by electrostatic attractions, hydrogen bonding, van der Waals forces and occasionally disulphide bridges. Thus disulphide bridges may exist within a given polypeptide chain (intra-chain) or linking different chains (inter-chain). An individual polypeptide chain in an oligomeric protein is referred to as a subunit. The subunits in a protein may be identical or different: for example, haemoglobin consists of two a- and two b-chains, and lactate dehydrogenase of four (virtually) identical chains.

The progressive condensation of many molecules of amino acids gives rise to an unbanked polypeptide chain. By convention, the N-terminal amino acid is taken as the beginning of the chain and the C-terminal amino acid as the end of the chain (proteins are biosynthesised in this direction). Polypeptide chains contain between 20 and 2 000 amino acid residues and hence have a relative molecular mass ranging between about 2 000 and 2 00 000. Many proteins have a relative molecular mass in the range 20 000 to 1 00 000. The distinction between a large peptide and a small protein is not clear. Generally, chains of amino acids containing fewer than 50 residues are referred to as peptides and those with more than 50 are referred

to as proteins. Most proteins contain many hundreds of amino acids (ribonuclease is an extremely small protein with only 103 amino acid residues) and many biologically active peptides contain 20 or fewer amino acids, for example oxytocin (9 amino acid residues), vasopressin (9), enkephalins (5), gastrin (17), somatostatin (14) and luteinising hormone (10).

3. Protein purification

At first sight, the purification of one protein from a cell or tissue homogenate that will typically contain 10 000–20 000 different proteins, seems a daunting task. However in practice, on average, only four different fractionation steps are needed to purify a given protein. Indeed, in exceptional circumstances proteins have been purified in a single chromatographic step. Since the reason for purifying a protein is normally to provide material for structural or functional studies, the final degree of purity required depends on the purposes for which the protein will be used, i.e. you may not need a protein sample that is 100% pure for your studies. Indeed, to define what is meant by an 'a pure protein' is not easy. Theoretically, a protein is pure when a sample contains only a single protein species, although in practice it is more or less impossible to achieve 100% purity. Fortunately, many studies on proteins can be carried out on samples that contain as much as 5–10% or more contamination with other proteins. This is an important point, since each purification step necessarily involves loss of some of the protein you are trying to purify. An extra (and unnecessary) purification step that increases the purity of your sample from, say, 90% to 98% may mean that you now have a more pure protein, but insufficient protein for your studies.

3.1 Lowry (Folin–Ciocalteau) method

In the past this has been the most commonly used method for determining protein concentration, although it is tending to be replaced by the more sensitive methods described below. The Lowry method is reasonably sensitive, detecting down to 10 mg cm_3 of protein, and the sensitivity is moderately constant from one protein to another. When the Folin reagent (a mixture of sodium tungstate, molybdate and phosphate), together with a copper sulphate solution, is mixed with a protein solution, a blue-purple colour is produced which can be quantified by its absorbance at 660 nm. As with most colorimetric assays, care must be taken that other compounds that interfere with the assay are not present. For the Lowry method this includes Tris, zwitterionic buffers such as Pipes and Hepes, and EDTA. The method is based on both the Biuret reaction, where the peptide bonds of proteins react with Cu2þ under alkaline conditions producing Cuþ, which reacts with the Folin reagent, and the Folin–Ciocalteau reaction, which is poorly understood but essentially involves the reduction of phosphomolybdotungstate to hetero-polymolybdenum blue by the copper-catalysed oxidation of aromatic amino acids. The resultant strong blue colour is therefore partly dependent on the tyrosine and tryptophan content of the protein sample.

3.2 The bicinchoninic acid method:

This method is similar to the Lowry method in that it also depends on the conversion of Cu2þ to Cuþ under alkaline conditions. The Cuþ is then detected by reaction with bicinchoninic acid (BCA) to give an intense purple colour with an absorbance maximum at 562 nm. The method is more sensitive than the Lowry method, being able to detect down to 0.5 mg protein cm_3, but perhaps more importantly it is generally more tolerant of the presence of compounds that interfere with the Lowry assay, hence the increasing popularity of the method.

3.3 The Bradford method

This method relies on the binding of the dye Coomassie Brilliant Blue to protein. At low pH the free dye has absorption maxima at 470 and 650 nm, but when bound to protein has an absorption maximum at 595nm. The practical advantages of the method are that the reagent is simple to prepare and that the colour develops rapidly and is stable. Although it is sensitive down to 20 mg protein cm_3, it is only a relative

method, as the amount of dye binding appears to vary with the content of the basic amino acids arginine and lysine in the protein. This makes the choice of a standard difficult. In addition, many proteins will not dissolve properly in the acidic reaction medium.

4. Protein Structure Determination

There are three methods available for determining protein relative molecular mass, Mr, frequently referred to as molecular weight. The first two described here are quick and easy methods that will give a value to_5–10%.

4.1 SDS-polyacrylamide gel electrophoresis (SDS-PAGE)

This form of electrophoresis separates proteins on the basis of their shape (size), which in turn relates to their relative molecular masses. A series of proteins of known molecular mass (molecular weight markers) are run on a gel on a track adjacent to the protein of unknown molecular mass. The distance each marker protein moves through the gel is measured and a calibration curve of log Mr versus distance moved is plotted. The distance migrated by the protein of unknown Mr is also measured, and from the graph its log Mr and hence Mr is calculated. The method is suitable for proteins covering a large Mr range (10 000–300 000). The method is easy to perform and requires very little material. If silver staining is used, as little as 1 ng of protein is required. In practice SDS–PAGE is the most commonly used method for determining protein Mr Values.

4.2 Molecular exclusion (gel filtration) chromatography

The elution volume of a protein from a molecular exclusion chromatography column having an appropriate fractionation range is determined largely by the size of the protein such that there is a logarithmic relationship between protein relative molecular mass and elution volume. By calibrating the column with a range of proteins of known Mr, the Mr of a test protein can be calculated. The method is carried out on HPLC columns (_1_30 cm) packed with porous silica beads. Flow rates are about 1 cm3 min_1, giving a run time of about 12 min, producing sharp, well resolved peaks. A linear calibration line is obtained by plotting a graph of log Mr versus Kd for the calibrating proteins. Kd is calculated from the following equation:

$$K_d = \frac{(V_e - V_0)}{(V_t - V_0)}$$

Where Vo is the volume in which molecules that are wholly excluded from the column material emerge (the excluded volume), Vt is the volume in which small molecules that can enter all the pores emerge (the included volume) and Ve is the volume in which the marker protein elutes. This method gives values that are accurate to_10%.

4.3 Mass spectrometry

Using either electrospray ionisation (ESI) or matrix-assisted laser desorption ionisation (MALDI) intact molecular ions can be produced for proteins and hence their masses accurately measured by mass spectrometry. ESI produces molecular ions from molecules with molecular masses up to and in excess of 100 kDa, whereas MALDI produces ions from intact proteins up to and in excess of 200 kDa. In either case, only low picomole quantities of protein are needed. For example, ab2 crystalline gave a molecular mass value (20 200_0.9), in excellent agreement with the deduced mass of 20 201. However, in addition about 10% of the analysed material produced an ion of mass 20 072.2. This showed that some of the purified protein molecules had lost their N-terminal amino acid (lysine).

Reference

1. Anonymous (2007). Guidelines on the Practice of Ethics Committees in Medical Research with Human Participants, 4th edn. London: Royal College of Physicians.
2. Crooke, S. T. (ed.) (2008). Antisense Drug Technology: Principles, Strategies and Applications, 2nd edn. New York: CRC Press. (Gives a review of the mechanisms of action, pharmacokinetics and therapeutic potential of antisense drugs in the context of a wide range of clinical conditions.)
3. Kshirsagar, T. (ed.) (2008). High-Throughput Lead Optimization in Drug Discovery. New York: CRC Press. (Uses real examples to illustrate the application of the technique in modern drug development.)
4. Rang, H. P. (ed.) (2006). Drug Discovery and Development: Technology in Transition. London: Elsevier. (An in-depth coverage written in an authoritative style by experts from the pharmaceutical industry.)