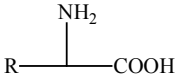
Amino Acids, Peptides and Proteins

Amino Acids

Terms:
Amino acid residue
Peptide bond
Protein
Enantiomers
Chiral center
Absolute configuration

Amino acids are important class of organic compounds that contain both the amino ($-NH_2$) and carboxyl (-COOH) groups. Of these acids, 20 serve as the building blocks of proteins. Known as the *standard*, or *alpha*, amino acids, they comprise alanine, arginine, asparagine, aspartic acid, cysteine, glutamic acid, glutamine, glycine, histidine, isoleucine, leucine, lysine, methionine, phenylalanine, proline, serine, threonine, tryptophan, tyrosine, and valine. All 20 are constructed according to a general formula:



As the formula shows, the amino and carboxyl groups are both attached to a single carbon atom, which is called the alpha carbon atom. Attached to the carbon atom is a variable group (R); it is in their R groups that the molecules of the 20 standard amino acids differ from one another (difference including **structure, size, electric charge and solubility** in water).

Please see the separate sheet for the structures of amino acids, where 22 amino acids are shown, and the sheet covers the most recent information on the structure of amino acids so far.
Absolute configuration is a term (or nomenclature) used to specify the spatial orientation of four substituents around the asymmetric carbon atoms (which is the chiral center).
Traditionally, the absolute configurations of simple sugars and amino acids are specified by the D & L system based on the absolute configuration of the three carbon sugar glyceraldehydes. However, do not forget the R&S system.

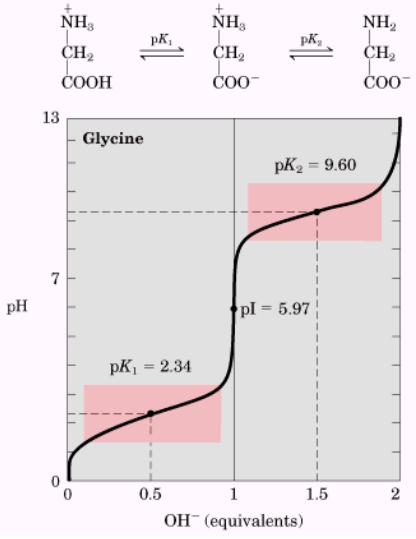
The amino acid residues in proteins are L stereoisomers
The amino acids found in naturally occurring proteins are exclusively L stereoisomers, this is because all biomolecules synthesized in cell are catalyzed by enzymes, and enzymes are highly asymmetric to contain their specific conformation and confer the efficient catalytic property. D-amino acid only exists in some small peptide of some bacteria cell wall and peptide antibiotics.

The properties of standard amino acids are shown in following table

Properties and Conventions Associated with the Standard Amino Acids

Amino acid	Abbreviated names		M_r	pK_1 (—COOH)	pK_2 ($—NH_3^+$)	pK_R (R group)	pI	Hydropathy index*	Occurrence in proteins (%)†
Nonpolar, aliphatic R groups									
Glycine	Gly	G	75	2.34	9.60		5.97	−0.4	7.2
Alanine	Ala	A	89	2.34	9.69		6.01	1.8	7.8
Valine	Val	V	117	2.32	9.62		5.97	4.2	6.6
Leucine	Leu	L	131	2.36	9.60		5.98	3.8	9.1
Isoleucine	Ile	I	131	2.36	9.68		6.02	4.5	5.3
Methionine	Met	M	149	2.28	9.21		5.74	1.9	2.3
Aromatic R groups									
Phenylalanine	Phe	F	165	1.83	9.13		5.48	2.8	3.9
Tyrosine	Tyr	Y	181	2.20	9.11	10.07	5.66	−1.3	3.2
Tryptophan	Trp	W	204	2.38	9.39		5.89	−0.9	1.4
Polar, uncharged R groups									
Serine	Ser	S	105	2.21	9.15		5.68	−0.8	6.8
Proline	Pro	P	115	1.99	10.96		6.48	1.6	5.2
Threonine	Thr	T	119	2.11	9.62		5.87	−0.7	5.9
Cysteine	Cys	C	121	1.96	10.28	8.18	5.07	2.5	1.9
Asparagine	Asn	N	132	2.02	8.80		5.41	−3.5	4.3
Glutamine	Gln	Q	146	2.17	9.13		5.65	−3.5	4.2
Positively charged R groups									
Lysine	Lys	K	146	2.18	8.95	10.53	9.74	−3.9	5.9
Histidine	His	H	155	1.82	9.17	6.00	7.59	−3.2	2.3
Arginine	Arg	R	174	2.17	9.04	12.48	10.76	−4.5	5.1
Negatively charged R groups									
Aspartate	Asp	D	133	1.88	9.60	3.65	2.77	−3.5	5.3
Glutamate	Glu	E	147	2.19	9.67	4.25	3.22	−3.5	6.3

- Nonpolar, Aliphatic R Groups
  - Glycine
  - Alanine
  - Valine
  - Leucine
  - Isoleucine
  - Methionine
- Aramotic R Groups
  - Phenylalanine
  - Tyrosine
  - Tryptophan
- Polar, Uncharged R Groups
  - Serine
  - Threonine
  - Cysteine
  - Proline
  - Asparagine
  - Glutamine
- Pisitively Charged (Basic) R Groups
  - Lysine
  - Arginine
  - Histidine
- Negatively Charged (Acidic) R Groups
  - Aspartate
  - Glutamate



Amino acids have their characteristic titration curves

Amino acid can be classified by R group

As mentioned earlier, although 20 standard amino acids are quite different from each other in terms of their structure, size, solubility in water and electric charge, these amino acids can be classified and grouped into different categories.

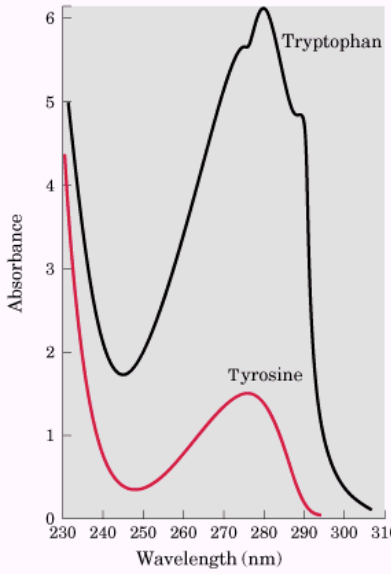
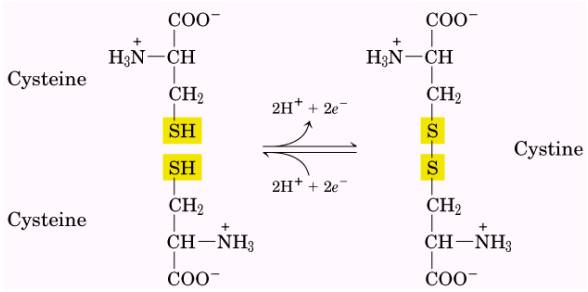


Fig. 1 The absorbance of ultraviolet light by aromatic amino acids



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

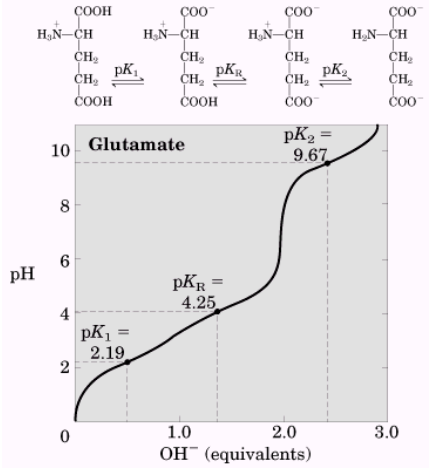
Among the 20 amino acids, tryptophan and tyrosine have the UV maximum absorption around 280 nm, and the extinction coefficient of tryptophan is much higher than tyrosine. No other amino acids have absorption around 280 nm (see Fig 1), therefore, the absorption around 280 nm can be used to detected the presence of tryptophan and tyrosine. In addition, as the percentage of tryptophan and tyrosine in protein is fixed at certain level, the absorption at 280 nm can also be used to detect the concentration of protein in aqueous solution. Besides the absorption at 280 nm, protein has another maximum absorption around 250 nm, and that absorption arises from the aromatic ring of phenylalanine, histidine.

Amino acids can act as acids and bases

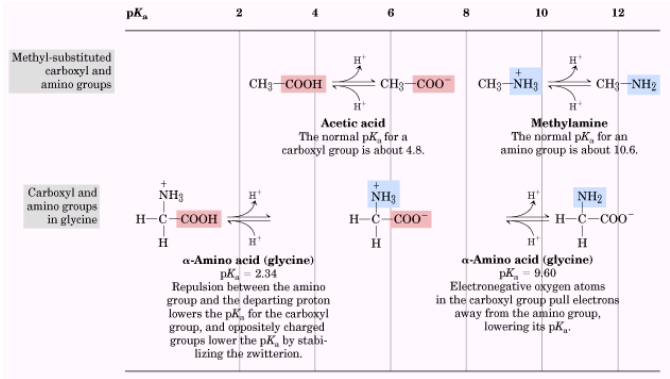
When amino acid is dissolved in water, it will become dipolar, or zwitterions, because carboxylic group is likely to give out its proton, and form anion, and amino group is likely to obtain a proton, and form cation, when both amino group and carboxylic group change to ionized forms, they have higher chance to associate with water molecules, because they can be hydrated by water molecules. This kind of molecule is amphoteric and is also called ampholytes. One example is shown below for Glycine

Titration
- The titration gives a quantitative measurement of the pKa of two ionizable groups (the pKa of any functional group is greatly affected by its chemcal environment)
- Amino acid has two regions of buffering power (amino acids with charged groups have complicated titration curves)
- Titration can tell the isoelectric point of isoelectric pH (pI), where the net electric charge of amino acid is zero.

Amino acids with single amino and carboxyl group have similar pKa values
- COOH: $1.8 < pKa < 2.4$
- NH_3^+: $8.8 < pKa < 11.0$



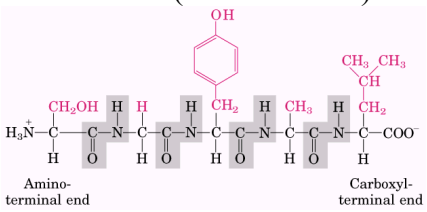
The titration curve for Glutamate



The acidity of functional group is affect by its chemical environment

Peptides and Proteins

Two amino acids can be linked together by a bond called **peptide bond**, where the amino group of one amino acid is connected to the carboxylic group of the other amino acid. This particular compound is called **dipeptide**. If less than 10 amino acid are linked together in this fashion, the resulted molecule is called **oligopeptide**, if more than 10 amino acids are linked together, the resulted molecule is called **polypeptide**; further linked by amino acids will give **protein** or enzyme. Peptide is one of a group of organic chemicals found in most living tissues, with a wide range of biological functions. Hormones such as ACTH and vasopressin are important polypeptides. In a peptide, the amino acid residue at the end with a free a-amino group is the **amino-terminal** (or N-terminal) residue, the residue at the other end, which has a free carboxyl group, is the **carboxyl-terminal** (C-terminal) residue.



Peptides can be distinguished by their ionization behavior

Although peptides contain only one free a-amino group and one carboxyl group at the ends of peptides, these groups will ionize as they do in free amino acids (the rest of amino groups and carboxyl groups are used to form peptide bonds, and cannot ionize anymore). Besides these, other R groups on peptides might be ionizable, and they will affect the total pKa of a peptide, and also the isoelectric pH (pI) values, at this pH value, the peptide will not move in electric field.

Term: protomers

The number of amino acids in peptides or proteins can be roughly determined by the molecular weight, the average contribution of amino acid to molecular weight is 110.

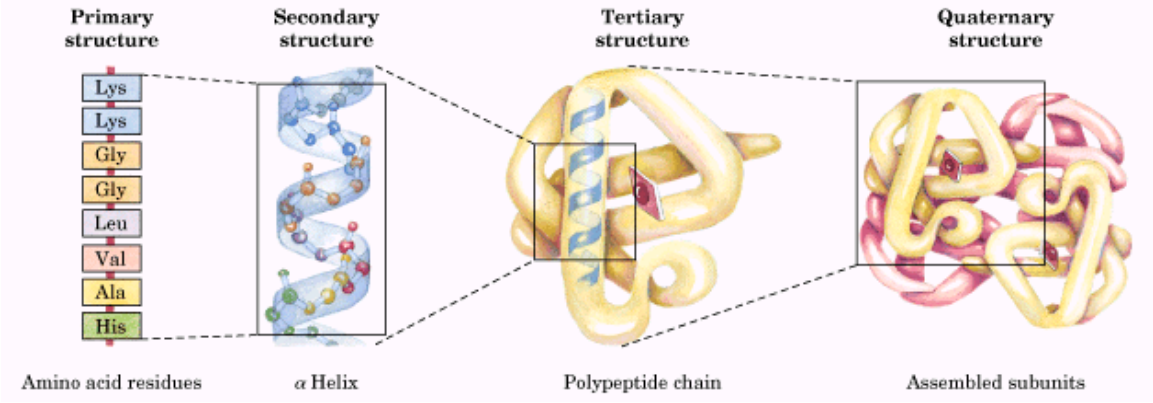
Complex proteins

Some proteins, besides their amino acids constituents, they also contain other chemical components, these kinds of proteins are called conjugated proteins. The non-amino acid part of a conjugated protein is called its prosthetic group. For examples, lipoproteins, glycoproteins, metalloproteins. (See table)

Conjugated Proteins

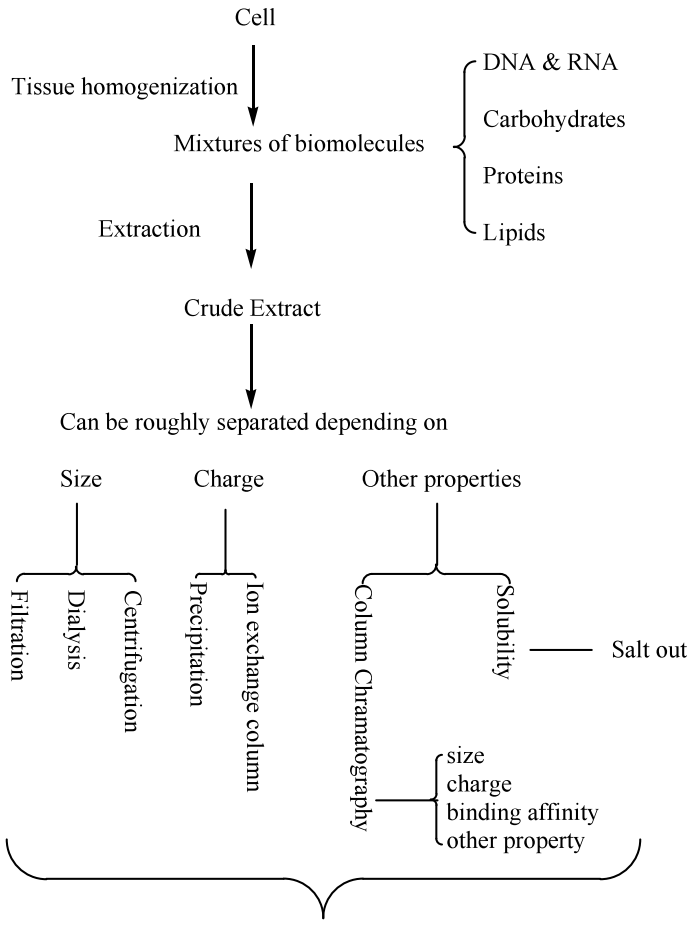
Class	Prosthetic group(s)	Example
Lipoproteins	Lipids	β_1-Lipoprotein of blood
Glycoproteins	Carbohydrates	Immunoglobulin G
Phosphoproteins	Phosphate groups	Casein of milk
Hemoproteins	Heme (iron porphyrin)	Hemoglobin
Flavoproteins	Flavin nucleotides	Succinate dehydrogenase
Metalloproteins	Iron	Ferritin
	Zinc	Alcohol dehydrogenase
	Calcium	Calmodulin
	Molybdenum	Dinitrogenase
	Copper	Plastocyanin

Protein has four levels of structure constitutions; they are primary structure, secondary structure, tertiary structure and quaternary structure.



Term:
Primary structure: the description of all covalent bonds linking amino acid residues in a polypeptide chain
Secondary structure: particular stable arrangements of amino acid residues giving rise to recurring structural patterns.
Tertiary structure: description of all aspects of the three-dimensional folding of a polypeptide.
Quaternary structure: the description of spatial arrangements of polypeptide subunits.

The separation and purification of proteins
Term: crude extract
Fractionation
Salt out
Dialysis

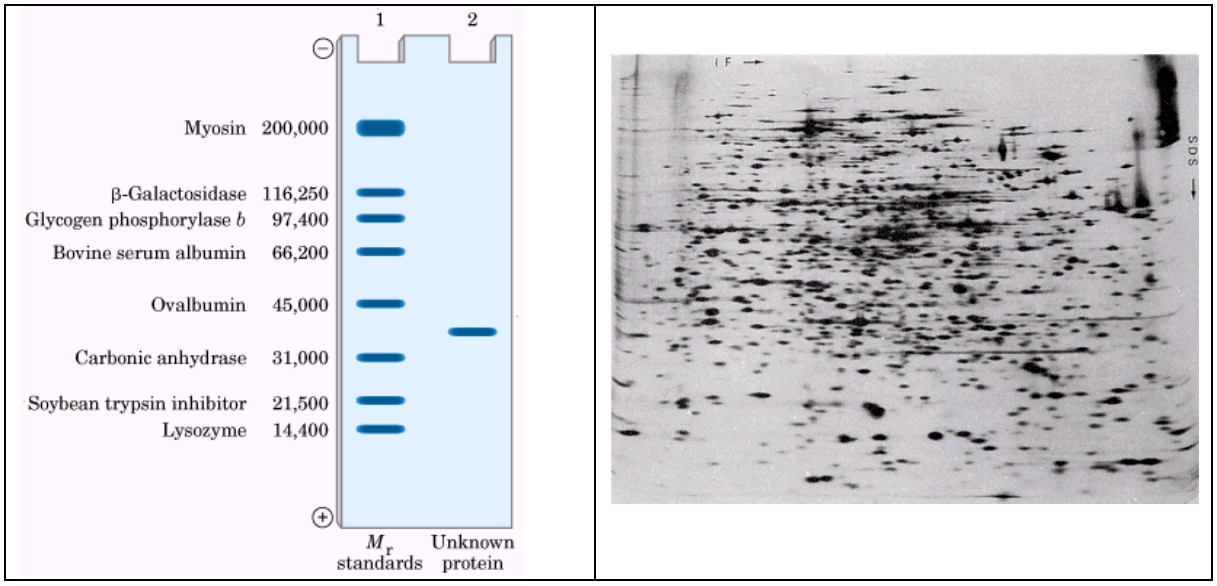


Electrophoresis: The process used for the separation of proteins based on the migration of charged proteins in an electric field. It can determine the isoelectric point and rough molecular weight. The migration of protein in electric field will depend on its size and shape. It is good for analysis but not for separation, as after the electrophoresis, some proteins might have been denatured already.

$$\mu = \frac{V}{E} = \frac{Z}{f}$$

μ refers to electrophoretic mobility, V is the velocity of particle in electric field, E stands for electrical potential, Z is the net charge of the molecule, *and f* is the frictional coefficient.

During the electrophoresis, the detergent SDS (sodium dodecyl sulfate) was mixed with protein, and SDS will bind to protein. Roughly about one SDS will bind two amino acid residues. After the binding, the resulting particle will have many charges on them, and the charge is proportional to the molecular weight of protein, in addition, after the binding, the conformation of protein will change, and the whole mixture will take the spherical shape. Because of these changes, different proteins will have similar shapes, and can be separated from each other according to their molecular weight. After the electrophoresis, the protein can be detected through binding with some dye, such as Coomassie blue. Comparing the position on the gel with other protein with known molecular weight, the molecular weight of unknown can be accurately measured. See the following figure.



Also, when protein is suspended on gel with pH gradient, the protein will move in electrical field until it reaches a specific position, where the pH is identical to its pI value. This process is called isoelectric focusing. Combining this with electrophoresis, the mixture of protein can be completely separated and analyzed; this is one of the hottest areas currently in science, called **proteomics**.

The covalent structure of proteins

The function of proteins depend on their amino acid sequence

	Different amino acid sequence	⟶	Different protein function
Evidence:	Defective proteins	⟶	Human genetic diseases
	Proteins with similar functions in different species	⟶	Similar amino acid sequence

Exception: polymorphic, proteins with similar function but with the variation in their amino acid sequence, these variations play less role on the protein functions.

Protein homology among species

Homologous proteins are proteins that are evolutionarily related. They usually perform the same function in different species. Homologous proteins from different species may have polypeptide chains that are identical or nearly identical in length. Many positions in amino acid sequence are occupied by the same residue in all species and are thus called invariant residues. Other positions show considerable variation in the amino acid residue from one species to another, these are called variable residues.

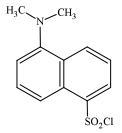
Term: conservative substitutions & nonconservative substitutions

Short polypeptides are sequenced using automated procedures

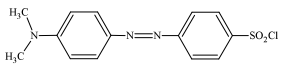
Peptide can be hydrolyzed in concentrated HCl solution, where monomeric amino acids will form. So the constitution of this peptide can be known by analyzing the resulting amino acid concentration. Different peptide will have different amino acid constitution, thus this constitution can be used for finger print peptide.

More importantly, to analyze peptide is not only to determine which amino acid exists in peptide, but also the amino acid sequence, as mentioned previously, the amino acid control the function of peptide.

Frederick Sanger used 1-fluoro-2,4-dinitrobenzene to determine the N-terminal amino acid, more powerful reagents are:



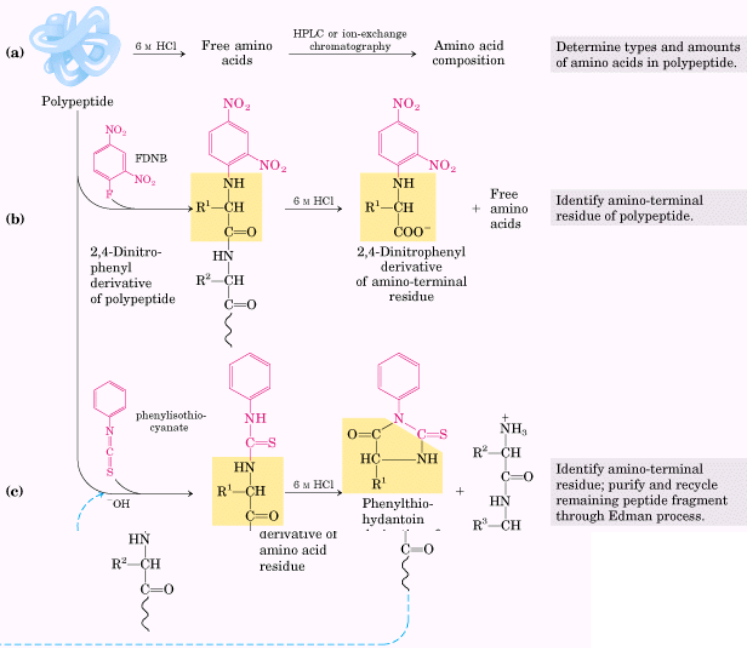
Dansyl chloride



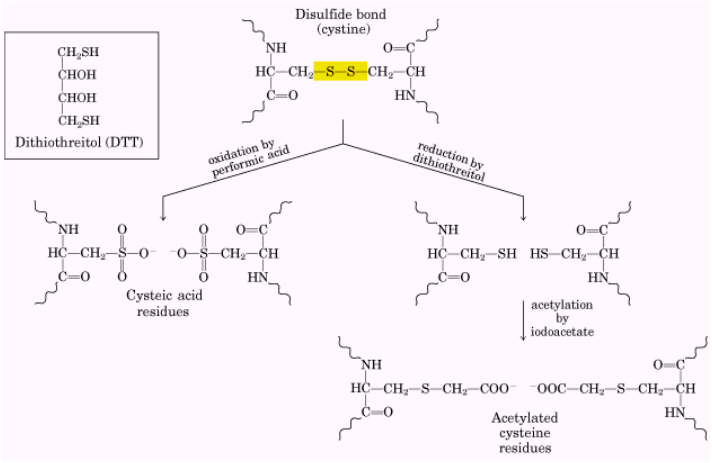
Dabsyl chloride

Edman degradation and automated sequence

The Edman degradation procedure labels and removes only the N-terminal residue from a peptide, leaving all other peptide bonds intact. The peptide is reacted with phenylisothiocyanate, and the N-terminal residue is ultimately removed as a phenylthiohydantoin derivative. After removal and identification of the N-terminal residue, the new N-terminal residue is exposed, and can be labeled, removed and identified through the same series of reaction. This procedure can be done totally by machine automatically.



However, for protein with more than hundreds of amino acids, will be impossible to be sequenced in this manner, because of experimental error. So, protein must be broken into a few pieces, and each piece is sequenced individually, and the final amino acid sequence will be obtained by linking the amino acid sequence of these pieces. The protein can be broken down to pieces by either chemical or enzymatic methods. For examples, disulfide bond can be broken down by either oxidation and reduction method shown as follows. Cyanogen bromide can be used to cleave only those peptide bonds in which the carbonyl group is contributed by Met.



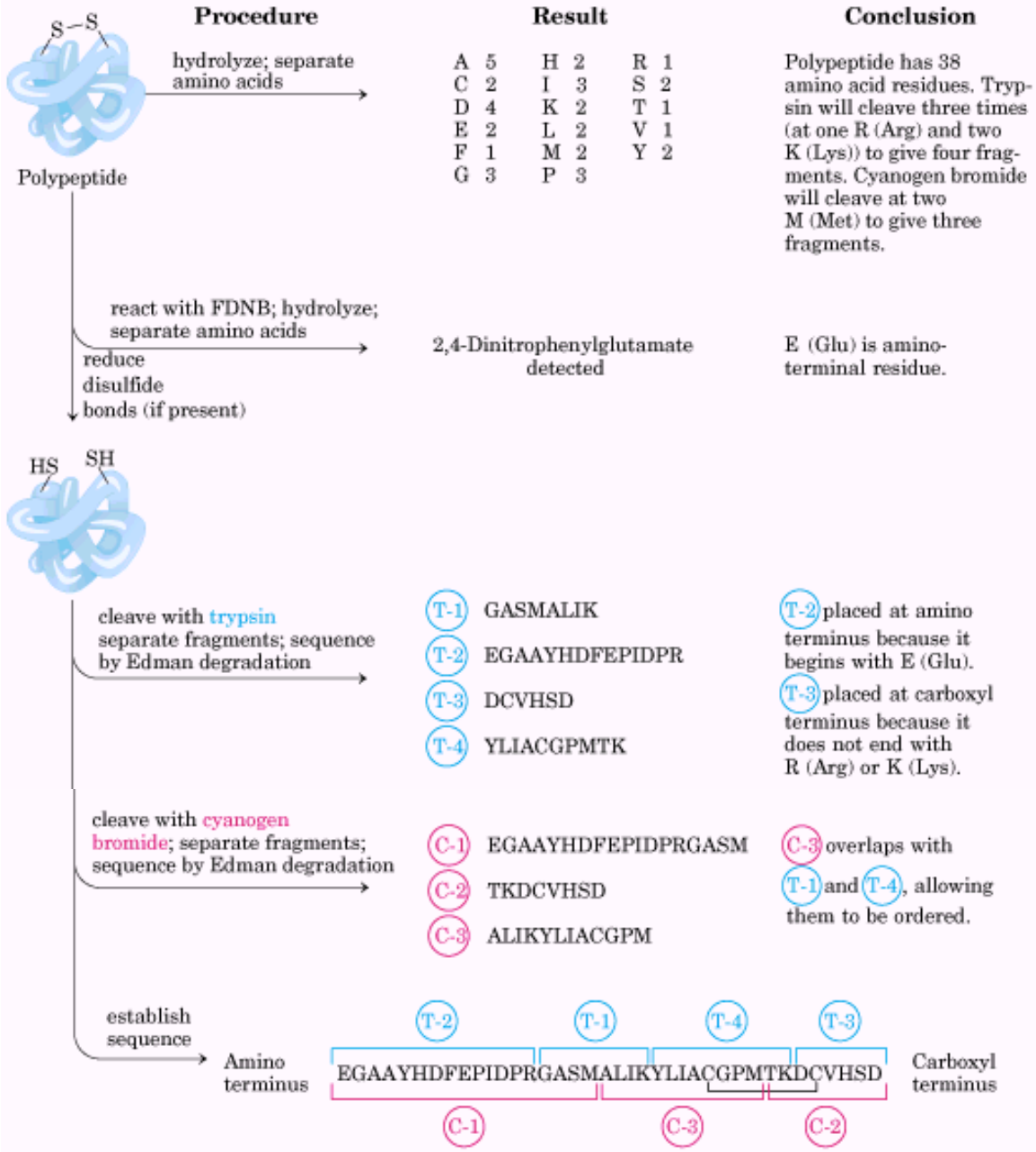
The protein can be broken down by enzyme also, this kind of enzymes are called proteases, which catalyze the hydrolytic cleavage of peptide bonds. Some proteases cleave only the peptide bond adjacent to particular amino acid residues. These proteases are listed in the following table.

The Specificity of Some Common Methods for Fragmenting Polypeptide Chains

Treatment*	Cleavage points†
Trypsin	Lys, Arg (C)
Submaxillarus protease	Arg (C)
Chymotrypsin	Phe, Trp, Tyr (C)
Staphylococcus aureus V8 protease	Asp, Glu (C)
Asp-*N*-protease	Asp, Glu (N)
Pepsin	Phe, Trp, Tyr (N)
Endoproteinase Lys C	Lys (C)
Cyanogen bromide	Met (C)

*All except cyanogen bromide are proteases. All are available from commercial sources.
†Residues furnishing the primary recognition point for the protease or reagent; peptide bond cleavage occurs on either the carbonyl (C) or the amino (N) side of the indicated amino acid residues.

After the degradation, and the sequenation of different pieces, then the question is how to link these amino acid sequences and connect to the whole protein. Another degradation must be down by different enzyme to cleave at different peptide bond, and the resulting pieces will be sequenced again. The information obtained from two series experiments will be combined together to achieve the whole picture of the protein, see detail as shown in follows.



Amino acid sequences can also be deduced by other methods

Genomics ⟶ Gene ⟶ genetic code ⟶ amino acid sequence in protein

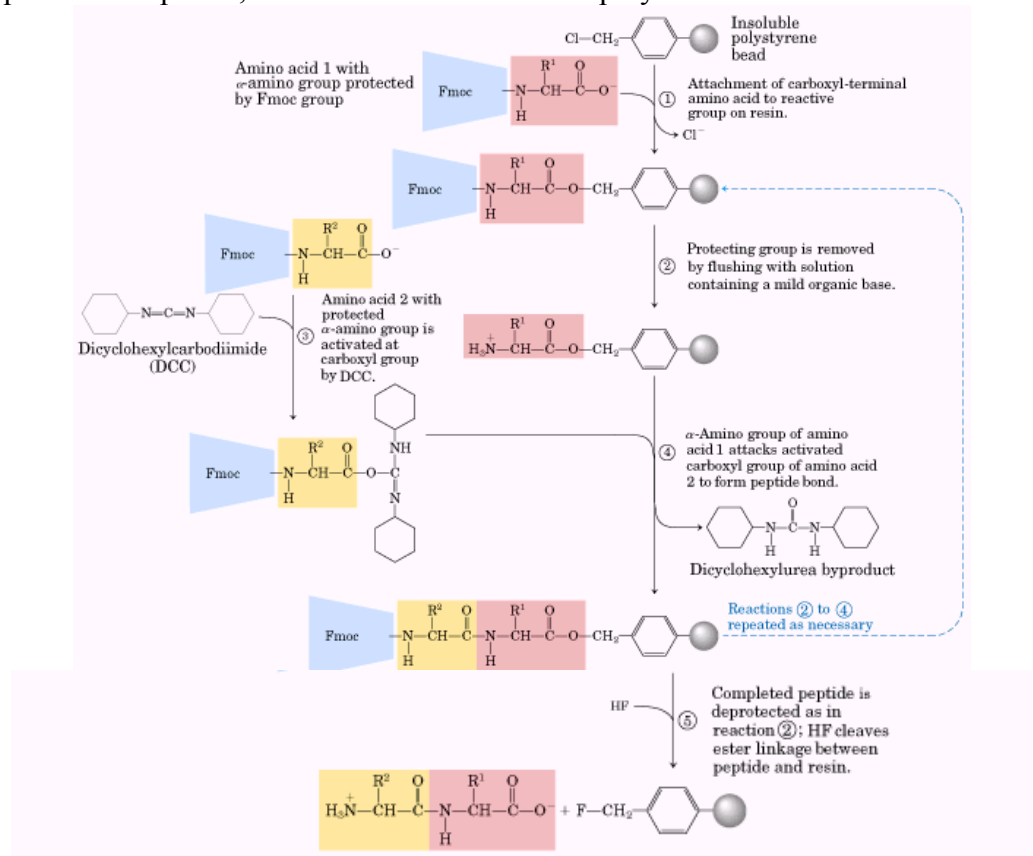
Proteomics ⟶ 2D electrophoresis ⟶ MALDI-TOF/ESI MS/MS ⟶ amino acid sequence in protein

As we know, every three genetic code will be corresponding to one amino acid in protein, so if the entire genetic sequence of gene which control that particular protein, then the amino acid sequence of that protein will be deduced from the genetic code. And this is the indirect method but used mostly to obtain the amino acid sequence for protein, because sequencing gene is much easier than sequence protein.

Other method used currently is from proteomics, where the key roles are two-dimensional gel electrophoresis and high efficient mass spectroscopy. The protein mixtures from the cell or tissue are completely separated by 2D gel electrophoresis, and the individual spot can be subjected for MS analysis. Nowadays, the analysis of protein through MS is achievable, as new techniques are introduced into MS, there MALDI-TOF and ESI play important roles.

Small peptides and proteins can be chemically synthesized

The small peptides and proteins cannot only be sequenced by chemical methods, but also they can be synthesized by chemical methods. The major improvement in peptide and protein synthesis is the solid-state synthesis, where the first amino acid is grafted onto the polymer bead, which is insoluble in aqueous solution. When amino acid is grafted onto this polymer, only one side can form the peptide with another amino acid, say the amino group of the amino acid on polymer. The carboxyl group of other amino acids will be activated and react with the free amino group, and form the peptide bonds. When the desired protein or peptide is completed, it will be cleaved from the polymer. The detail in follows.



More facts about peptides and proteins

A protein may be formed of a single polypeptide chain, or it may consist of several such chains held together by weak molecular bonds. Each protein is formed according to a precise set of instructions contained within the nucleic acid (see Nucleic Acids), which is the genetic material of the cell. These instructions determine which of the 20 standard amino acids are to be incorporated into the protein, and in what sequence. The R groups of the amino acid subunits determine the final shape of the protein and its chemical properties; an extraordinary variety of proteins can be produced from the same 20 subunits.
The standard amino acids serve as raw materials for the manufacture of many other cellular products, including hormones and pigments. In addition, several of these amino acids are key intermediates in cellular metabolism.

In addition to the amino acids that form proteins, more than 150 other amino acids have been found in nature, including some that have the carboxyl and amino groups attached to separate carbon atoms. These unusually structured amino acids are most often found in fungi and higher plants.

Complex proteins are absorbed from the digestive tract and are broken down into about 20 amino acids needed for cellular anabolism. Amino acids may undergo further chemical change to form such internal secretions as hormones and digestive enzymes. Amino acids in excess of those required to replenish body cells and fluids are catabolized in two steps. The first is deamination, in which the nitrogen-containing part of the molecule is removed and united with carbon and oxygen to form urea, ammonia, and uric acid- the nitrogenous products of protein metabolism. Following deamination, each of the remaining amino acids undergoes further chemical breakdown to form other compounds, which are then still further catabolized, often by pathways common to those of similar products from the catabolism of carbohydrates and fat. The end products of these protein portions are carbon dioxide and water.
If an enzyme is lacking because of some hereditary defect, the chemical transformation in which it would participate is blocked. As a result, cell products fail to be synthesized or catabolized, too much of a metabolic product accumulates, causing injury to tissues, or intracellular materials fail to cross cell membranes.
Although the effects of some metabolic errors are manifested in early infancy, others may appear only in adulthood. Some inborn errors may be fatal, some may have no apparent harmful effects, and some may persist. A result of error in amino acid metabolism is phenylketonuria (PKU). This occurs in infants when metabolism of the amino acid phenylalanine is blocked; the accumulated metabolic products may cause brain damage. In carbohydrate metabolism, one error results in galactosemia, in which the enzyme required to convert galactose to glucose is absent. The consequent inability to metabolize milk sugar results in the accumulation of galactose in the blood, sometimes with damage to the brain and liver and the development of cataracts and mental retardation.

The word protein is coined from the Greek proteios , or "primary."
Protein molecules range from the long, insoluble fibers that make up connective tissue and hair to the compact, soluble globules that can pass through cell membranes and set off metabolic reactions. They are all large molecules, ranging in molecular weight from a few thousand to more than a million, and they are specific for each species and for each organ of each species. Humans have an estimated 30,000 different proteins, of which only about 2 percent have been adequately described. Proteins in the diet serve primarily to build and maintain cells, but their chemical breakdown also provides energy, yielding close to the same 4 calories per gram as do carbohydrates.
Besides their function in growth and cell maintenance, proteins are also responsible for muscle contraction. The digestive enzymes are proteins, as are insulin and most other hormones. The antibodies of the immune system (q.v.) are proteins, and proteins such as hemoglobin carry vital substances throughout the body. Proteins also transmit all hereditary characteristics in the form of genes.

Nutrition.

Whether found in humans or in single-celled bacteria, proteins are composed of units of about 20 different amino acids (q.v.) , which, in turn, are composed of carbon, hydrogen, oxygen, nitrogen, and sometimes sulfur. In a protein molecule these acids form peptide bonds- bonds between amino and carboxyl (COOH) groups-in long strands (polypeptide chains). The almost numberless combinations in which the acids line up, and the helical and globular shapes into which the strands coil, help to explain the great diversity of tasks that proteins perform in living matter.
For adults, the Recommended Dietary Allowance (RDA) for protein is 0.79 g per kg (0.36 g per lb) of body weight each day. For children and infants this RDA is doubled and tripled, respectively, because of their rapid growth.

Structure of Proteins.

The simplest protein, called a primary structure, is a linear sequence of amino acids. Different sequences of the acids along a chain, however, affect the structure of a protein molecule in different ways. Forces such as hydrogen bonds, disulfied bridges, attractions between positive and negative charges, and hydrophobic ("water-fearing") and hydrophilic ("water-loving") linkages cause a protein molecule to coil

or fold into a secondary structure, examples of which are the so-called α-helix and the β-pleated sheet. When forces cause the molecule to become even more compact, as in globular proteins, a protein of a tertiary structure is formed. When a protein is made up of more than one polypeptide chain, as in hemoglobin and some enzymes, it is called a quaternary structure.

Interaction with Other Proteins.

Polypeptide chains are sequenced and coiled in such a way that the hydrophobic amino acids usually face inward, giving the molecule stability, and the hydrophilic amino acids face outward, where they are free to interact with other compounds and especially other proteins. Globular proteins, in particular, can join with a specific compound such as a vitamin derivative and form a coenzyme, or join with a specific protein and form an assembly of proteins needed for cell chemistry or structure.

Fibrous Proteins.

The major fibrous proteins, described below, are collagen, keratin, fibrinogen, and muscle proteins.
Collagen, which makes up bone, skin, tendons, and cartilage, is the most abundant protein found in vertebrates. The molecule usually contains three very long polypeptide chains, each with about 1000 amino acids, that twist into a regularly repeating triple helix and give tendons and skin their great tensile strength. When long collagen fibrils are denatured by boiling, their chains are shortened to form gelatin.

Keratin

Keratin, which makes up the outermost layer of skin and the hair, scales, hooves, nails, and feathers of animals, twists into a regularly repeating coil called an α-helix. Serving to protect the body against the environment, keratin is completely insoluble in water. Its many disulfide bonds make it an extremely stable protein, able to resist the action of proteolytic (protein-hydrolyzing) enzymes. In beauty treatments, human hair is set under a reducing agent, such as thioglycol, to reduce the number of disulfide bonds, which are then restored when the hair is exposed to oxygen.

Fibrinogen

Fibrinogen is a blood plasma protein responsible for blood clotting. With the catalytic action of thrombin, fibrinogen is converted into molecules of the insoluble protein fibrin, which link together to form clots.

Muscle proteins.

Myosin, the protein chiefly responsible for muscle contraction, combines with actin, another muscle protein, forming actomyosin, the different filaments of which shorten, causing the contracting action.

Globular Proteins

Unlike fibrous proteins, globular proteins are spherical and highly soluble. They play a dynamic role in body metabolism. Examples are albumin, globulin, casein, hemoglobin, all of the enzymes, and protein hormones. The albumins and globulins are classes of soluble proteins abundant in animal cells, blood serum, milk, and eggs. Hemoglobin is a respiratory protein that carries oxygen throughout the body and is responsible for the bright red color of red blood cells. More than 100 different human hemoglobins have been discovered, among which is hemoglobin S, the cause of sickle-cell anemia, a hereditary disease suffered mainly by blacks.

Enzymes

All of the enzymes are globular proteins that combine rapidly with other substances, called substrate, to catalyze the numerous chemical reactions in the body. Chiefly responsible for metabolism and its regulation, these molecules have catalytic sites on which substrate fits in a lock-and-key manner to trigger and control metabolism throughout the body.

Protein hormones

These proteins, which come from the endocrine glands, do not act as enzymes. Instead they stimulate target organs that in turn initiate and control important activities-for example, the rate of metabolism and the production of digestive enzymes and milk. Insulin, secreted by the islands of Langerhans, regulates carbohydrate metabolism by controlling blood glucose levels. Thyroglobulin, from the thyroid gland, regulates overall metabolism; calcitonin, also from the thyroid, lowers blood calcium levels. Angiogenin, a protein structurally determined in the mid-1980s, directly induces the growth of blood vessels in tissues.

Antibodies

Also called immunoglobulins, antibodies make up the thousands of different proteins that are generated in the blood serum in reaction to antigens (body-invading substances or organisms). A single antigen may elicit the production of many antibodies, which combine with different sites on the antigen molecule, neutralize it, and cause it to precipitate from the blood.

Microtubules

Globular proteins can also assemble into minute, hollow tubes that serve both to structure cells and to conduct substances from one part of a cell to another. Each of these microtubules, as they are called, is made up of two types of nearly spherical protein molecules that pair and joins onto the growing end of the microtubule, adding on length as required. Microtubules also make up the inner structure of cilia, the hairlike appendages by which some microorganisms propel themselves.