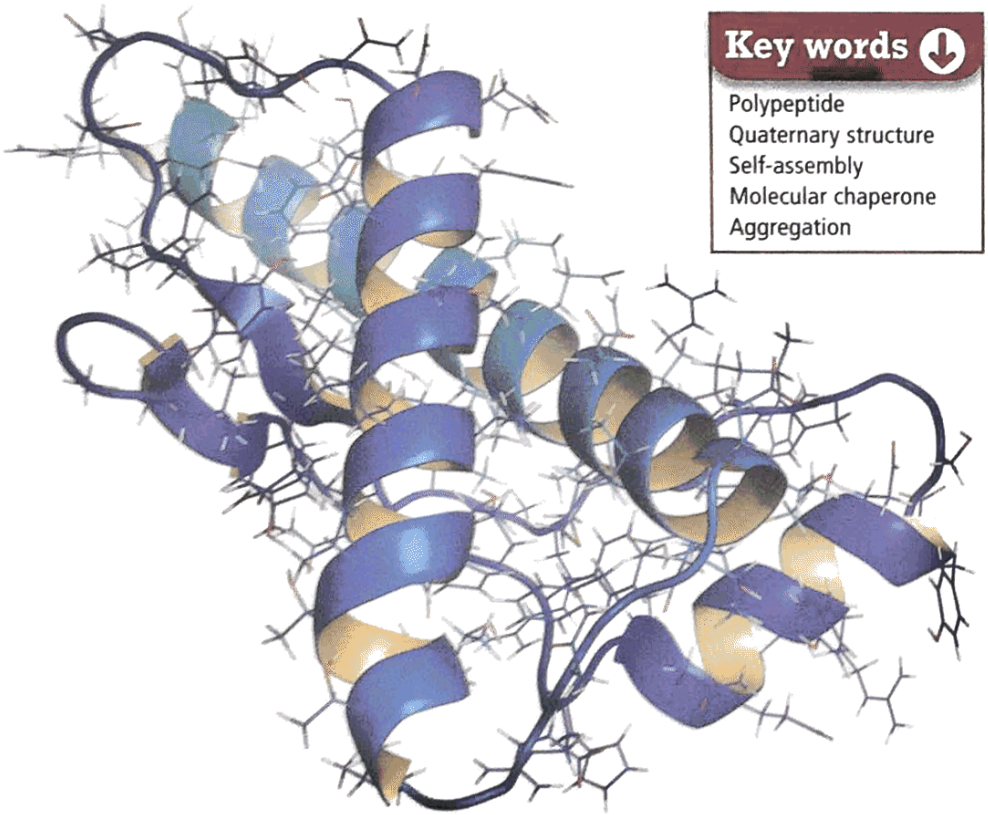
Protein folding
Principles and problems

John Ellis

Protein folding is one of the most important biological processes because it converts the linear information encoded in genes into the three-dimensional structures that give proteins their functional properties. Biochemist John Ellis explains how proteins fold, and introduces how problems with the process can lead to disease



Key words
Polypeptide
Quaternary structure
Self-assembly
Molecular chaperone
Aggregation

Proteins are the action molecules of all organisms. They are the molecules that carry out most of the processes necessary to sustain life. Genes are important because they contain the information to make these proteins. Proteins function by presenting binding sites on their surface for chemicals in their environment, including other proteins. These sites recognise and bind a wide range of chemicals in a highly specific fashion. 'Specific' means that each site binds just one type of chemical. Most proteins have more than one binding site, each specific for a different type of chemical.

Enzymes

Some proteins function as enzymes. Enzymes are defined as specific catalysts, that is, each type of enzyme promotes one type of chemical reaction, which would proceed very slowly, or not at all, in its absence. Each enzyme has its own name, which indicates the type of reaction that it catalyses. For example, hexokinase catalyses the transfer of a phosphate group from a donor molecule called ATP to a carbon atom at position six in glucose to give glucose-6-phosphate.

Figure 1 shows an atomic model of one molecule of the enzyme hexokinase. Each sphere in this model represents the size and position of an atom of carbon, oxygen or nitrogen; hydrogen atoms, of which there are many, are omitted for clarity. The shape of hexokinase is unique to that enzyme, and every molecule of hexokinase is identical. Hexokinase has on its surface a site that specifically binds a glucose molecule.

The way in which a chemical fits into a binding site is similar to the way a key fits into a lock. However, because most proteins change their shape as they bind their substrates, the term '**induced fit**' is a more helpful description. Proteins can change their shape because the bonds that hold the shape in place are weak **non-covalent bonds**, such as hydrogen bonds, electrostatic bonds and hydrophobic interactions. Such weak bonds are easily broken and reformed under the conditions found inside organisms. Figure 1 illustrates the change in shape of hexokinase following the binding of one glucose molecule.

Protein folding

Because each organism carries out several thousand different kinds of chemical reaction, it follows that each organism contains thousands of different kinds of enzyme. When we add to their roles as catalysts the many other functions of proteins — such as acting as signalling molecules, transport systems, structural components and antibodies — it is clear that proteins are a remarkable group of chemicals. What determines the specific properties of all these different proteins?

Proteins are linear polymers of different kinds of amino acid strung together like beads on a chain. Amino acids have the same basic structure but vary in their side chains, of which there are at least 20 different types found in proteins (see Box 1). Each chain is called a polypeptide and has a unique sequence of amino acids encoded in the base sequence of its gene. This amino acid sequence is called the primary structure.

Terms explained

Dialysis A method of separating small molecules from larger ones.

Induced fit The idea that enzymes are flexible structures with binding sites that alter shape as substrates bind to them.

Non-covalent bonds Bonds that differ from covalent bonds in that they do not involve the sharing of electrons but involve weaker more dispersive electromagnetic interactions.

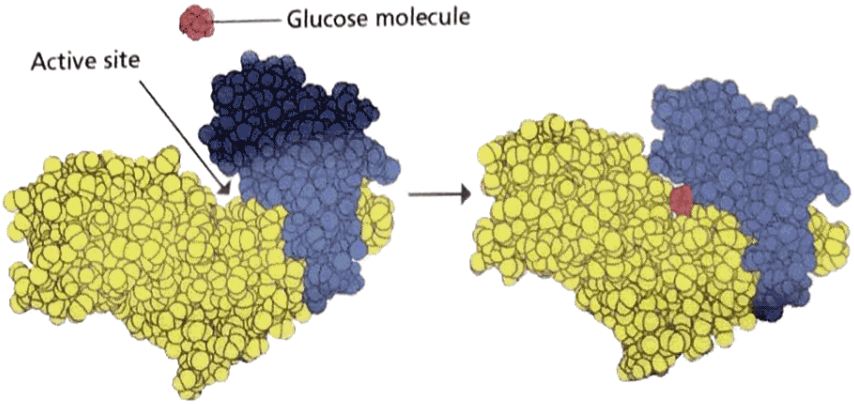


Figure 1 The enzyme hexokinase is made of two different subunits, here coloured blue and yellow. One molecule of glucose binds to its active site, and the enzyme then changes shape to enclose the glucose

A polypeptide does not exist as a straight chain for long after it is synthesised, but folds into helices and sheets, forming the secondary structure. This intermediate then folds further into a more compact stable shape, called the tertiary structure or monomer.

Protein folding is very rapid, taking only a few seconds to minutes. Most proteins consist of more than one monomer, which may be identical to, or different from, one another. These monomers assemble together into oligomers, which are said to have quaternary structure. For example, hexokinase is an oligomer of two dissimilar subunits. Figure 2 illustrates these processes, with different secondary structures depicted as blue and red ribbons.

What determines the specific binding properties of a given protein? The answer is that it is the sequence of amino acids that determines exactly how each chain folds into the specific three-dimensional shape, called its conformation, which displays binding sites on its surface. The conformation unique to hexokinase is formed by specific interactions between the side chains of the different amino acids along the polypeptide chain. What interactions are possible, and hence the shape of the folded molecule, is determined solely by the sequence of amino acids. Once that sequence is synthesised, the chain folds into its functional conformation. The amino acid sequence of each polypeptide is encoded by the base sequence of its gene.

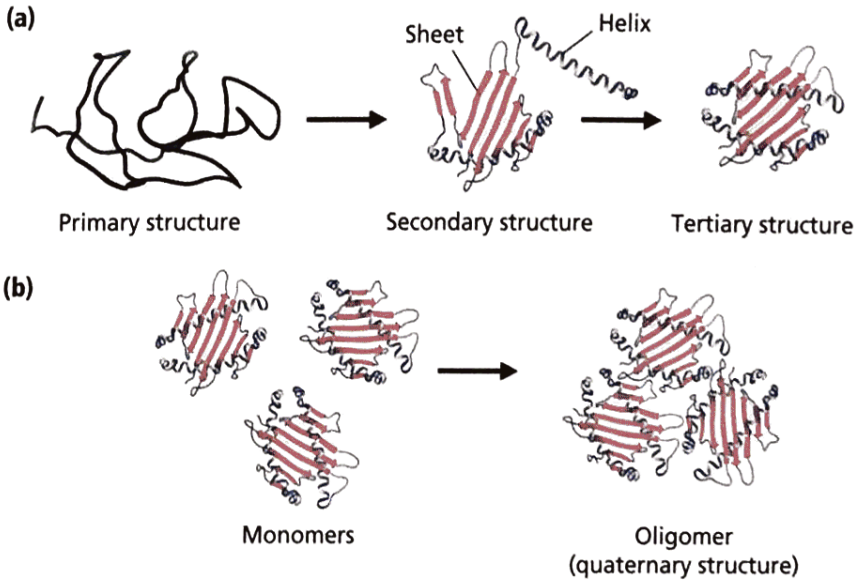


Figure 2 **(a)** Protein folding involves the rapid collapse of an extended polypeptide chain into a stable compact structure called a monomer. **(b)** Monomers may be functional on their own or assemble with identical or different monomers to form oligomers

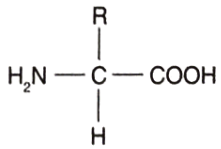
Genes in action

The biochemical process by which a gene determines the amino acid sequence of a polypeptide chain is extraordinarily complicated. This complexity is required because one chemical language — the base sequence of DNA — has to be translated into another chemical language — the amino acid sequence of polypeptide chains. It is important to realise that this process does not involve a conversion of the bases themselves into amino acids. What flows from bases to amino acids is sequence information, not material.

The mechanism of gene expression is outlined in Figure 3. DNA acts as a template that allows the

Box 1 Amino acid structure

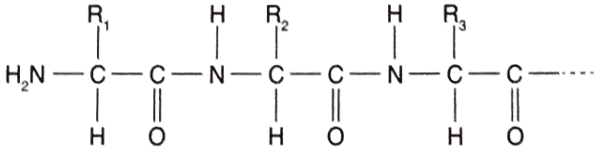
The basic structure of an amino acid is:



where R represents an amino acid side chain. For example:

- in glycine, R = H
- in alanine, R = CH_3
- in lysine, R = $CH_2CH_2CH_2CH_2NH_3^+$
- in aspartic acid, R = CH_2COO^-

Amino acids are linked together via peptide bonds to form a polypeptide.



The type and sequence of amino acid side chains (R_1, R_2, R_3 etc.) give the polypeptide chain its unique properties.

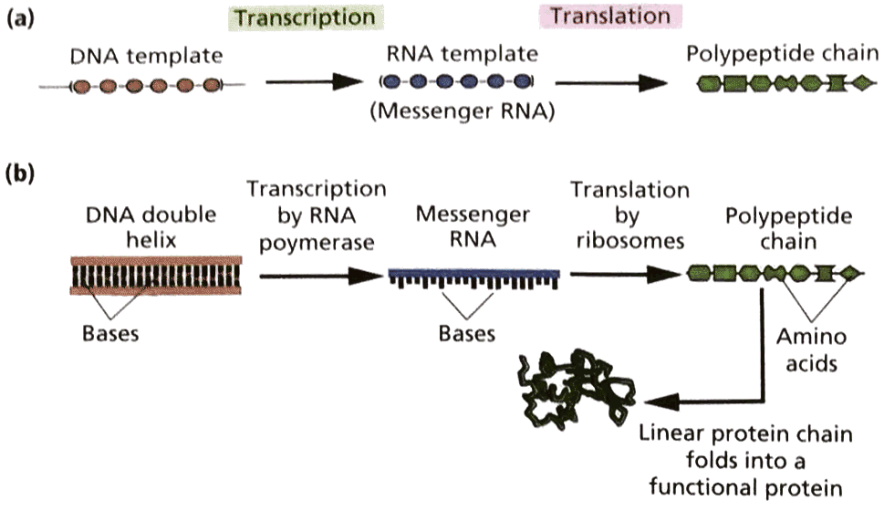


Figure 3 **(a)** Sequence information flows from DNA to protein. **(b)** Linear DNA information produces a folded polypeptide

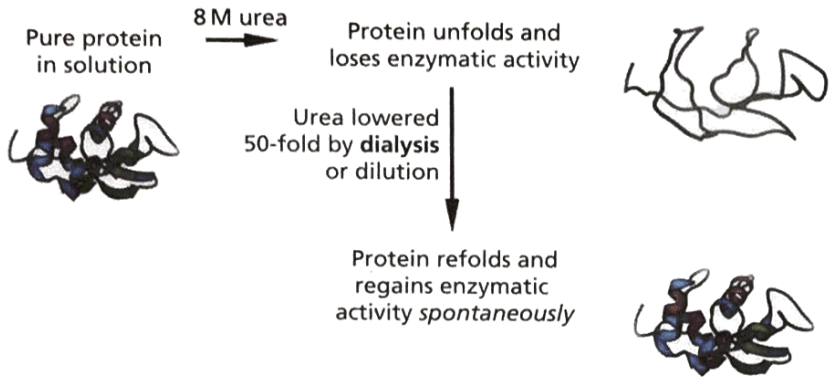


Figure 4 The classic refolding experiment of Anfinsen demonstrates the principle of protein self-assembly

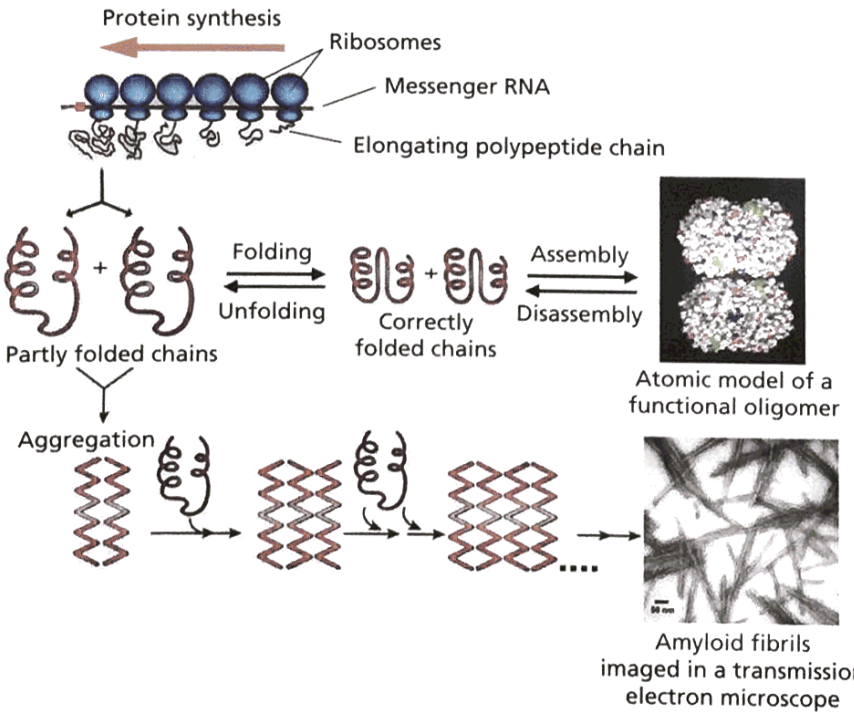


Figure 5 Newly synthesised chains can either fold correctly to produce functional proteins or aggregate together into non-functional fibrillar complexes. These complexes are called amyloid fibrils because when first discovered they were thought (incorrectly) to be made of carbohydrate

sequence of bases in one strand of the double helix of each gene to be copied in the form of RNA. This process is called transcription, and it produces from each gene many copies of an RNA molecule called messenger RNA (mRNA). Each molecule of mRNA is about the same length as the gene that acts as a template for its synthesis, and the enzyme that carries out this synthesis is called RNA polymerase. The order of bases in mRNA is the same as the order of bases in one strand of the double helix of DNA from which it is copied.

The mRNA in turn acts as a template for a second process called translation, in which amino acids are joined together by ribosomes in a linear order determined by the order of bases in the mRNA. Each molecule of mRNA can programme the synthesis of many molecules of the protein chain that it encodes.

The principle of self-assembly

Experiments with isolated proteins show that protein folding is spontaneous — no other source of information or energy is required for folding to occur once the sequence has been made. This fact is sometimes called the 'principle of self-assembly'.

This principle was established by experiments carried out by Christian Anfinsen, which earned him the Nobel prize in chemistry in 1972. He purified the enzyme ribonuclease, which hydrolyses RNA, until no other proteins were present. He then incubated the pure enzyme in aqueous solution with a high concentration of urea. This treatment breaks the weak non-covalent bonds holding the tertiary structure together. The protein thus unfolds and loses its enzymatic activity because it no longer has binding sites for its RNA substrate. Such an unfolded protein is said to be denatured. If the urea concentration is diluted by a large factor so that it is no longer effective as a denaturing agent, the protein then refolds spontaneously into an active enzyme (see Figure 4). This type of refolding experiment has been carried out successfully with many isolated proteins. But inside organisms the situation is not so simple.

Problems with folding

Mutations that change the primary structure of a polypeptide may result in the failure of protein folding to produce a functional protein (see Biological Sciences Review, Vol. 26, No. 3, pp. 22–25). However, mutations are rare — every time a human cell divides, only approximately three base pairs in the entire genome of three billion base pairs are changed. But there is another folding problem that potentially affects every polypeptide chain, whether mutated or not. This is the problem of protein aggregation.

Tremor in the hand of an elderly patient with Parkinson's disease. Tremors are the result of reduced numbers of functional neurones in the brain

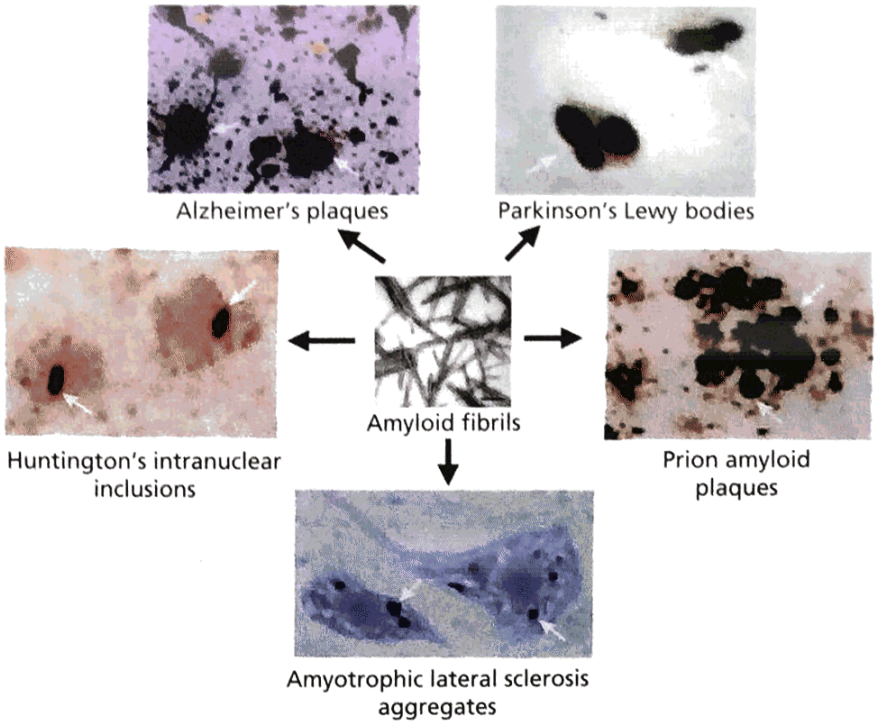


Figure 6 Amyloid plaques in the brain are characteristic of most human neurodegenerative diseases

Protein folding is rapid, but not fast enough to completely prevent partly folded chains from interacting prematurely with one another to form incorrectly folded structures called aggregates (see Figure 5). Aggregation is highly specific — only chains that are identical or nearly identical in sequence will form aggregates, because the structures that aggregate have some secondary and tertiary structure. It is also a progressive process — more and more partly folded chains are added together to form growing fibrils.

Fibrils that are large enough to become insoluble are called amyloid plaques. Such amyloid plaques are characteristic of human neurodegenerative diseases such as Alzheimer's, Parkinson's and Huntington's disease, where they accumulate in the brain (see Figure 6). It is thought that it is the smaller, soluble aggregates that cause neurones to become damaged and die.

Chaperones

Protein aggregation is a universal cellular problem, so it is not surprising that all cells have evolved means to combat this problem. A major mechanism is the existence of proteins that recognise partly folded polypeptides and bind to them transiently. This binding prevents the partly folded chains from aggregating with one another. The binding is then reversed under conditions that favour correct folding. These protective proteins are called molecular chaperones. The term 'molecular chaperone' was coined because their properties are a precise molecular analogy of the role of human chaperones. The traditional role of human chaperones is to prevent incorrect interactions between pairs of humans.

Unlike other proteins, most chaperones have broad specificity and bind to a wide range of different partly folded polypeptides. Some chaperones bind to monomers released by the disassembly of oligomers, and prevent them from aggregating. Most chaperones use ATP as an energy source to perform their function.

The discovery of chaperones in the 1980s resulted in the replacement of the original principle of spontaneous self-assembly by the current principle of assisted self-assembly. Note that the basic concept of self-assembly is retained, but modified to include assistance by other proteins (molecular chaperones) that minimise incorrect folding. This discovery has enabled researchers to devise novel strategies to develop chemicals that can mimic molecular chaperones, thereby providing treatments for diseases caused by protein aggregation.

John Ellis is an emeritus professor in life sciences at the University of Warwick. Visit **http://en.wikipedia.org/wiki/R._John_Ellis** for downloadable lectures and a book called *How Science Works: Evolution*

Key points

- Proteins are the action molecules of life, and they function by binding biochemicals in a highly specific manner.
- Proteins are made as linear chains of amino acids but then fold rapidly into three-dimensional shapes with specific binding sites on their surfaces.
- Partly folded chains may aggregate together to form non-functional structures that can be toxic.
- Aggregation is prevented by some proteins acting as molecular chaperones that bind transiently to partly folded chains.