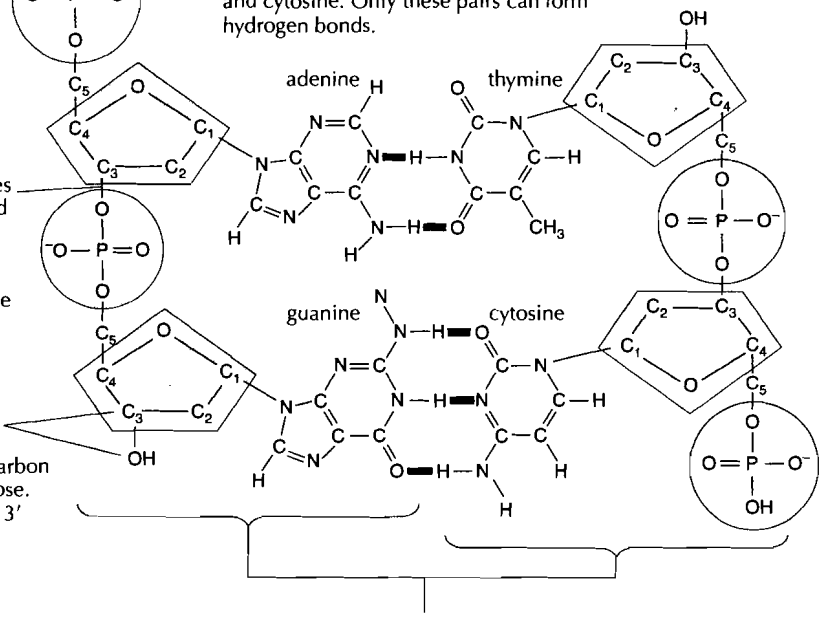
DNA STRUCTURE

At one end of each strand is a phosphate linked to carbon atom 5 of deoxyribose. This is the 5′ terminal.

Hydrogen bonds (shown as ▬) link the bases. Two bonds form between adenine and thymine and three bonds between guanine and cytosine. Only these pairs can form hydrogen bonds.

Adjacent nucleotides are linked by a bond between the phosphate group of one nucleotide and carbon atom 3 of the other nucleotide.

At one end of each strand is a hydroxyl group attached to carbon atom 3 of deoxyribose. This is therefore the 3′ terminal.



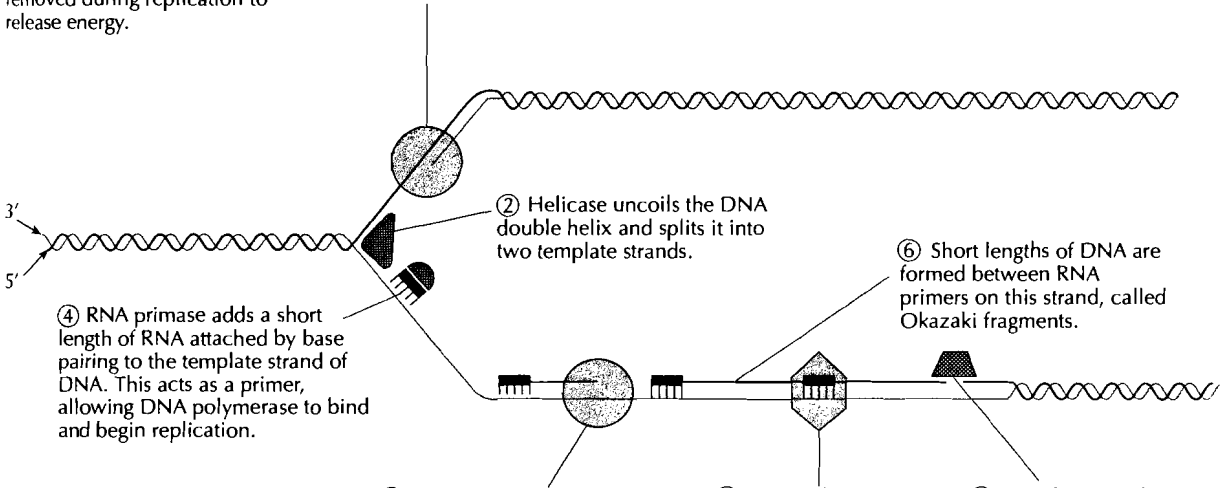
Two of the bases in DNA are purines: adenine and guanine. They have two rings in their molecules. Two of the bases in DNA are pyrimidines. Cytosine and thymine are pyrimidines. They have one ring in their molecule. Only a purine plus a pyrimidine will fit in the space between the sugar–phosphate backbones.

The two strands have their 3′ and 5′ terminals at opposite ends – they are anti-parallel. DNA replication can only occur in a 5′→3′ direction so a different method is needed for the two strands.

THE ROLE OF ENZYMES IN DNA REPLICATION

① The cell produces many free nucleotides for DNA replication. Each has three phosphate groups – they are deoxyribonucleoside triphosphates. Two phosphates are removed during replication to release energy.

③ DNA polymerase III adds nucleotides in a 5′→3′ direction. On one strand it moves in the same direction as the replication fork, close to helicase. On the other template strand it moves in the opposite direction.



② Helicase uncoils the DNA double helix and splits it into two template strands.

④ RNA primase adds a short length of RNA attached by base pairing to the template strand of DNA. This acts as a primer, allowing DNA polymerase to bind and begin replication.

⑥ Short lengths of DNA are formed between RNA primers on this strand, called Okazaki fragments.

⑤ DNA polymerase III starts replication next to the RNA primer and adds nucleotides in a 5′→3′ direction. It therefore moves *away* from the replication fork on this strand.

⑦ DNA polymerase I removes the RNA primer and replaces it with DNA. A nick is left where two nucleotides are still unconnected.

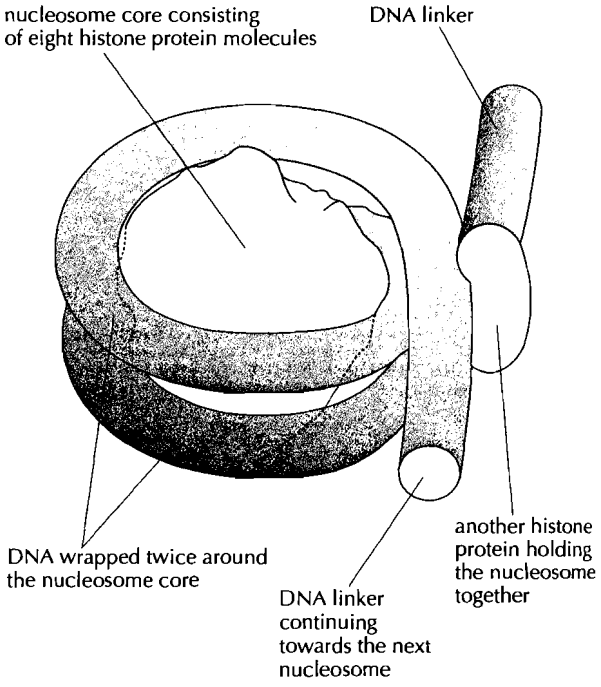
⑧ DNA ligase seals up the nick by making another sugar–phosphate bond.

DNA in eukaryotes and prokaryotes

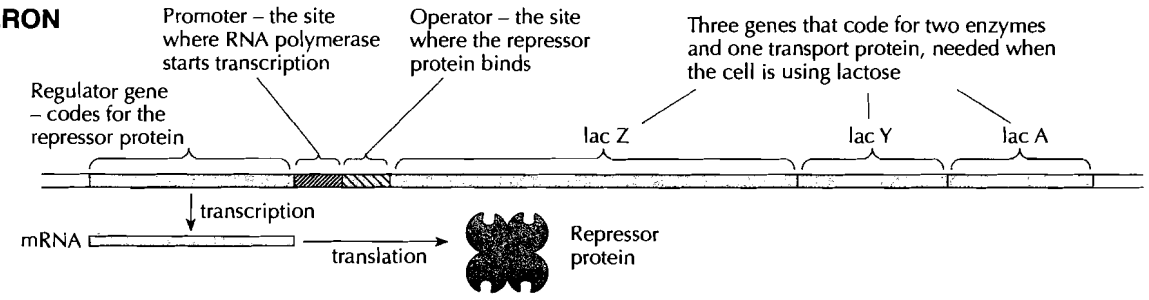
All eukaryotes and prokaryotes use DNA as their genetic material and use the same genetic code, but there are differences in the way that the DNA is used.

- In eukaryotes the DNA is associated with proteins to form **nucleosomes**, whereas in prokaryotes the DNA is naked.
- Replication of DNA begins at special initiation points. Eukaryotes have many of these initiation points along each chromosome. Most prokaryotes have only one point on their DNA molecule where replication is initiated.
- Much of the DNA in eukaryotes consists of repetitive base sequences. These sequences are not genes. They are useful for DNA profiling, but their role in eukaryotes is uncertain. Prokaryotes do not usually have repetitive sequences.
- Many genes in eukaryotes contain **introns**. These are non-coding sequences that are transcribed but not translated. They are found in newly transcribed mRNA but are removed. Mature mRNA does not contain introns. The sequences that are not removed are called exons. Prokaryotes do not usually have introns in their genes.
- Gene expression involves transcription of a gene and translation of the mRNA produced by transcription. In any cell, at a particular time, some genes are being expressed and some genes are not. This is called gene regulation. Each gene in a eukaryote is usually regulated separately. In prokaryotes some genes are arranged in groups and are regulated together. These groups are called **operons**. The Lac Operon, found in *Escherichia coli* is an example. Figures (below) show how it works.

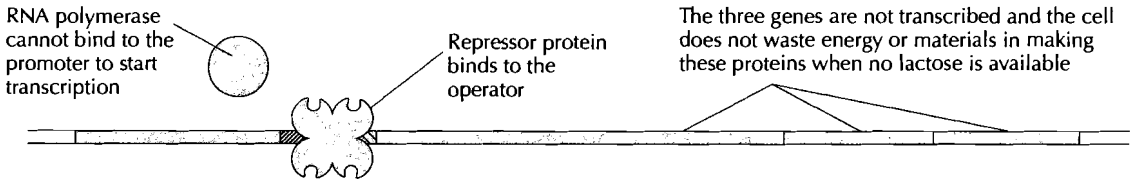
NUCLEOSOME STRUCTURE



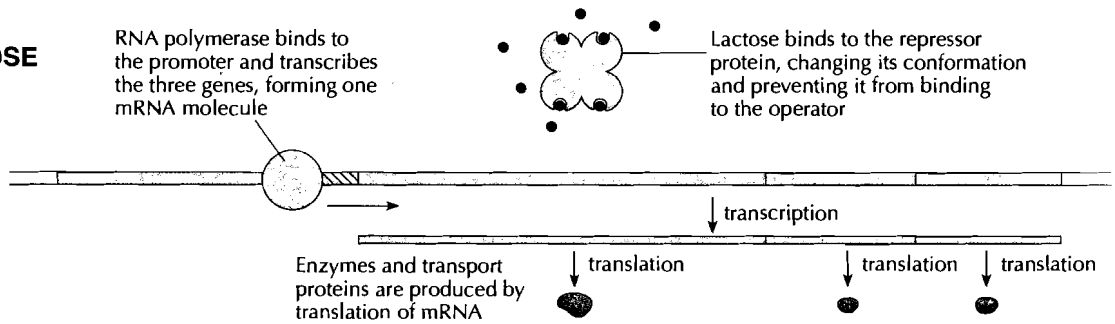
THE LAC OPERON



WITHOUT LACTOSE



WITH LACTOSE



CONTROL OF GENE EXPRESSION IN EUKARYOTES

The process by which a gene has an effect on a cell is called **gene expression**. Although a cell in a multicellular organism contains all of the organism's genes, only some of them will be expressed in that cell. This is the key to controlling the development and differentiation of cells. For example, in humans only ß islet cells in the pancreas express the genes for making insulin.

Gene expression involves several stages:

- Transcription of the gene
- Processing of the mRNA to remove introns (post-transcriptional modification)
- Translation of the mRNA to produce a protein
- Modification of the protein inside the endoplasmic reticulum or inside the Golgi apparatus (post-translational modification)

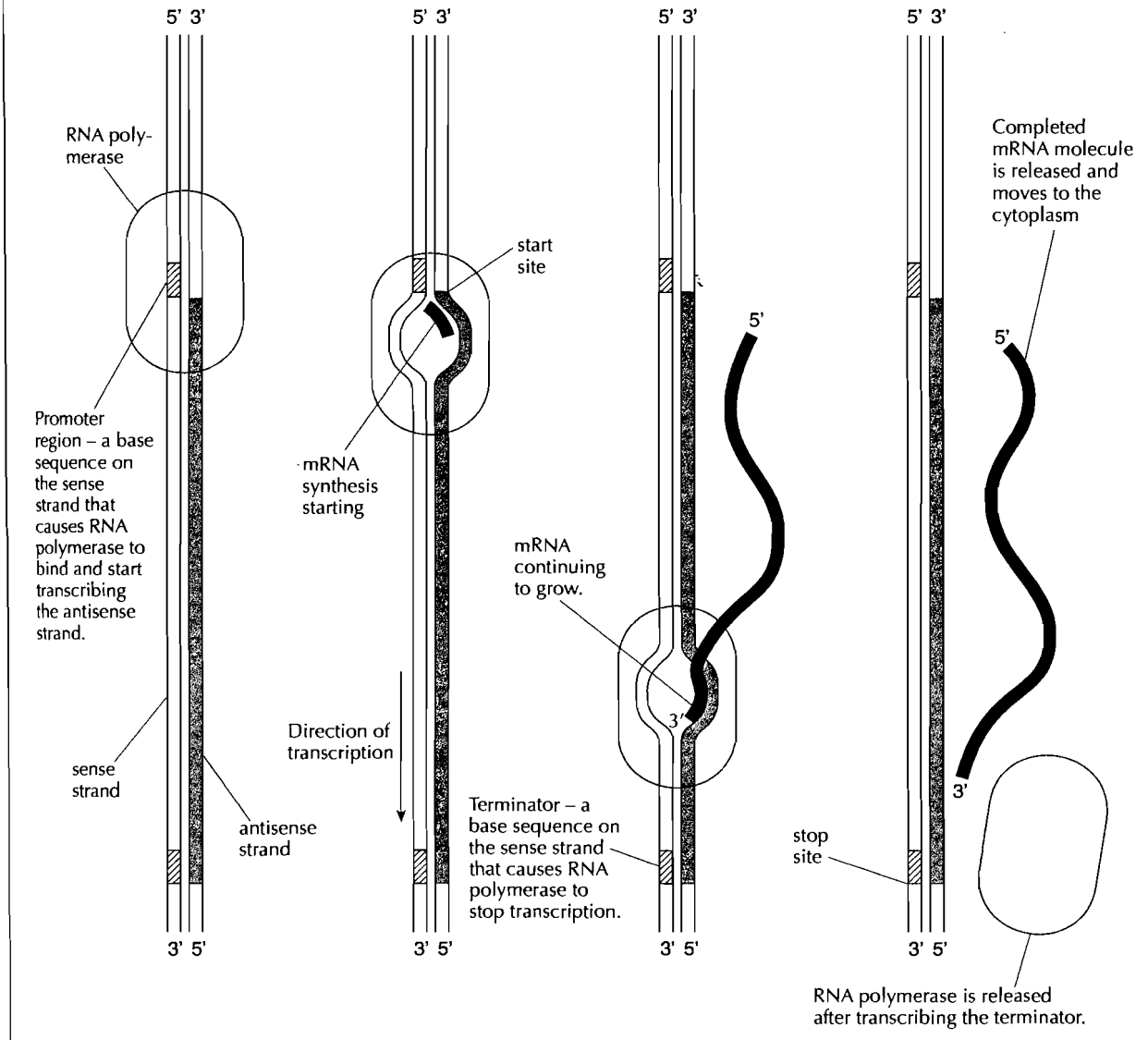
At any of these stages, the expression of genes can be regulated. Regulation of transcription is the most important stage.

REGULATION OF TRANSCRIPTION

The regulation of transcription in eukaryotic cells is a complex process.

- Genes are only transcribed if RNA polymerase binds to a region of DNA close to the start of the gene called the **promoter**.
- Most genes have several sequences of bases in their promoter that encourage binding of RNA polymerase. There is a wide variety of these sequences.
- Some base sequences always encourage binding, to allow continuous expression of genes.
- Other base sequences are sites where a regulatory protein can bind to the promoter. RNA polymerase only binds if the regulatory protein is present. Many different regulatory proteins are involved in the regulation of transcription.
- Some regulatory proteins only become active if a steroid hormone or other chemical messenger binds to them.

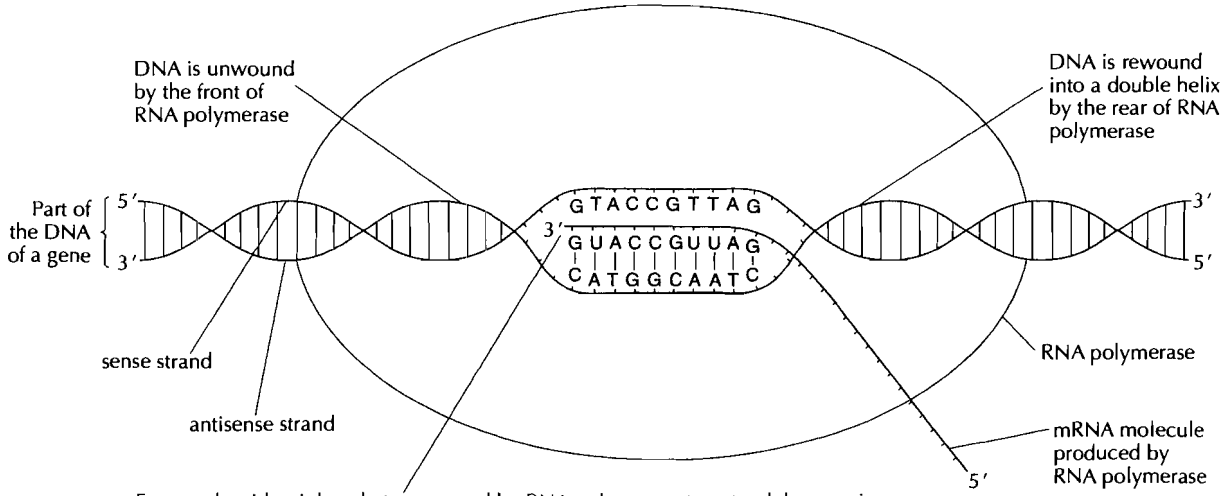
ROLE OF THE PROMOTER AND TERMINATOR IN TRANSCRIPTION



Transcription and reverse transcription

RNA POLYMERASE AND TRANSCRIPTION

DNA is split into two strands by RNA polymerase. One of these strands forms the template for transcription. The base sequence of the mRNA is complementary to it. The other strand has the same base sequence as the mRNA (except for T instead of U) and is therefore called the **sense strand**. The strand that forms the template and is transcribed is called the **antisense strand**.



Free nucleoside triphosphates are used by RNA polymerase to extend the growing mRNA molecule. Two phosphates are removed as they are linked on, converting them into RNA nucleotides. The 5′ end of the nucleotide is added to the 3′ end of the growing chain– transcription thus moves in a 5′→3′ direction.

REVERSE TRANSCRIPTASE AND ITS ROLE IN HIV

HIV and other retroviruses contain an enzyme that catalyses the production of DNA from RNA. This enzyme is called reverse transcriptase. The genome of retroviruses consists of RNA. When a retrovirus enters a host cell it makes a DNA copy of the genome using reverse transcriptase (see right). The DNA copy becomes inserted into the host cell's chromosomes.
In the case of HIV, the viral DNA can remain inactive in chromosomes of lymphocytes for long periods. When the lymphocyte replicates its DNA, the viral DNA is replicated as well. Eventually the viral DNA starts to be transcribed, producing viral mRNA that is translated to produce viral proteins. Transcription also produces copies of the entire viral genome. The viral proteins and RNA assemble to form new HIV particles, which are released from the lymphocyte. The lymphocyte subsequently dies.

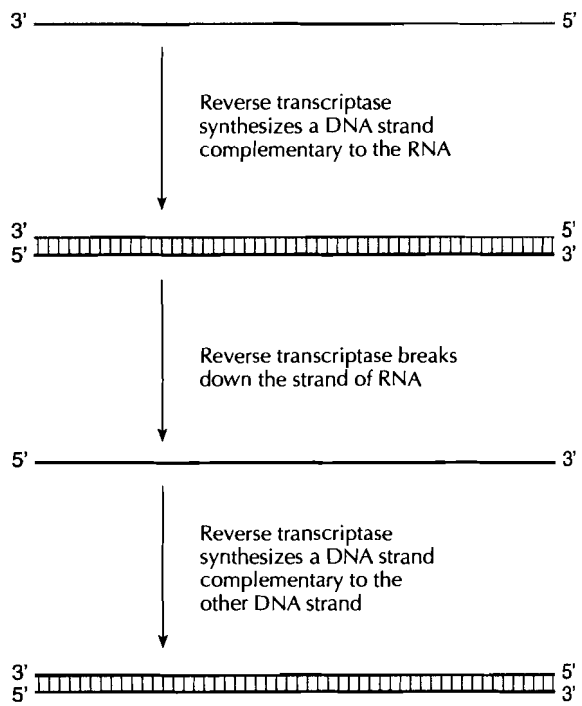
THE USE OF REVERSE TRANSCRIPTASE IN MOLECULAR BIOLOGY

Reverse transcriptase is used to obtain copies of genes for use in gene transfer.

- Cells that are transcribing the required gene are obtained and mRNA copies of the gene are extracted from them.
- Single stranded DNA copies of the mRNA are made using reverse transcriptase. This is called cDNA.
- DNA polymerase is used to convert the single-stranded DNA into double stranded DNA – the genes that can be transferred into another organisms.

The genes produced do not have introns, so if they are transferred to bacteria, which do not edit out introns, the correct protein will nonetheless be produced.

Reverse transcription in cells infected by retroviruses



The RNA of the retrovirus is the sense strand, therefore the first DNA strand is the antisense strand and the second one synthesized is the sense strand

Translating the genetic code

Messenger RNA carries the information needed for making polypeptides out from the nucleus to the cytoplasm of eukaryotic cells. The information is in a coded form, which is decoded during translation. Ribosomes, tRNA molecules and tRNA activating enzymes are needed to carry out this decoding.

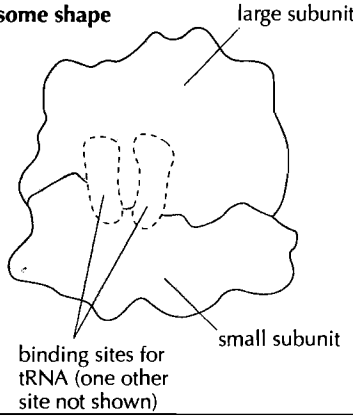
THE STRUCTURE OF RIBOSOMES

Ribosomes have a complex structure, with the following principal features.

- Proteins and ribosomal RNA molecules both form part of the structure.
- There are two subunits, one large and one small.
- There are binding sites for tRNA on the surface of the ribosome. Two tRNA molecules can bind at the same time to the ribosome.
- There is a binding site for mRNA on the surface of the ribosome.

The figure (right) shows the shape of a ribosome in outline and two tRNA binding sites.

Ribosome shape



tRNA AND tRNA ACTIVATING ENZYMES

Transfer RNA has a vital role in translating the genetic code. There are many different types of tRNA in a cell.

All tRNA molecules have:

- a triplet of bases called the anticodon, in a loop of seven nucleotides
- two other loops
- the base sequence CCA at the 3′ terminal, which forms a site for attaching an amino acid
- sections that become double stranded by base pairing.

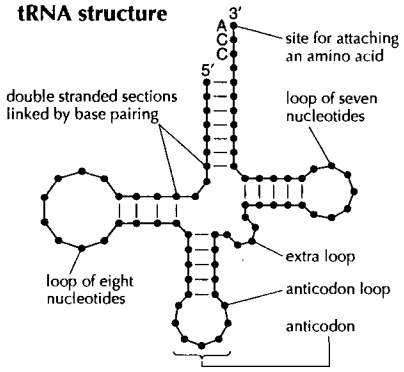
These features allow all tRNA molecules to bind to the binding sites on the ribosome and to mRNA. The base sequence of tRNA molecules varies and this causes some variable features in its structure.

- An extra small loop is sometimes present
- The base paired sections are sometimes helical.

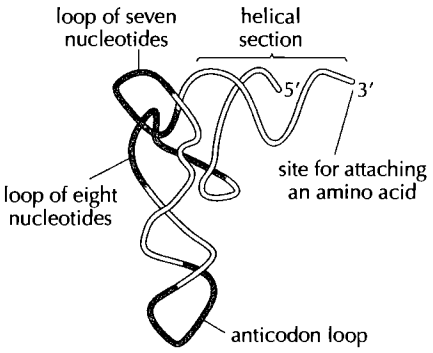
The variable features give each type of tRNA a distinctive three-dimensional shape and distinctive chemical properties. This allows the correct amino acid to be attached to the 3′ terminal by an enzyme called a **tRNA activating enzyme**. There are 20 different tRNA activating enzymes – one for each of the 20 different amino acids. Each of these enzymes attaches one particular amino acid to all of the tRNA molecules that have an anticodon corresponding to that amino acid. The tRNA activating enzymes recognize these tRNA molecules by their shape and chemical properties. The figure (right) is a two-dimensional view of the structure of tRNA molecules and the figure (lower right) shows an example of the three-dimensional structure formed when a tRNA molecule folds up.

Energy from ATP is needed for the attachment of amino acids. A high-energy bond is created between the amino acid and the tRNA. Energy from this bond is later used to link the amino acid to the growing polypeptide chain during translation.

tRNA structure



Three-dimensional view of tRNA



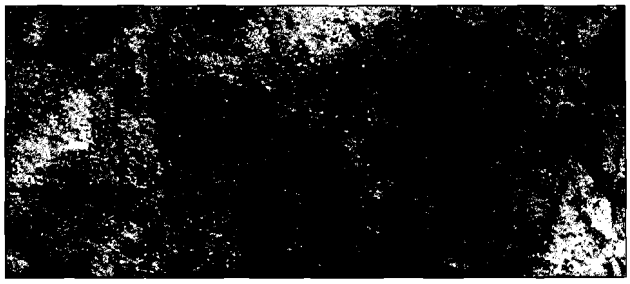
THE GENETIC CODE

Although the genetic code appears completely random at first, there are some rules, which are always or nearly always followed (right).

First base of codon (5′ end)	Second base of codon on messenger RNA				Third base of codon (3′ end)
	U	C	A	G	
U	Phenylalanine	Serine	Tyrosine	Cysteine	U
	Phenylalanine	Serine	Tyrosine	Cysteine	C
	Leucine	Serine	STOP	STOP	A
	Leucine	Serine	STOP	Tryptophan	G
C	Leucine	Proline	Histidine	Arginine	U
	Leucine	Proline	Histidine	Arginine	C
	Leucine	Proline	Glutamine	Arginine	A
	Leucine	Proline	Glutamine	Arginine	G
A	Isoleucine	Threonine	Asparagine	Serine	U
	Isoleucine	Threonine	Asparagine	Serine	C
	Isoleucine	Threonine	Lysine	Arginine	A
	Methionine / START	Threonine	Lysine	Arginine	G
G	Valine	Alanine	Aspartic acid	Glycine	U
	Valine	Alanine	Aspartic acid	Glycine	C
	Valine	Alanine	Glutamic acid	Glycine	A
	Valine	Alanine	Glutamic acid	Glycine	G

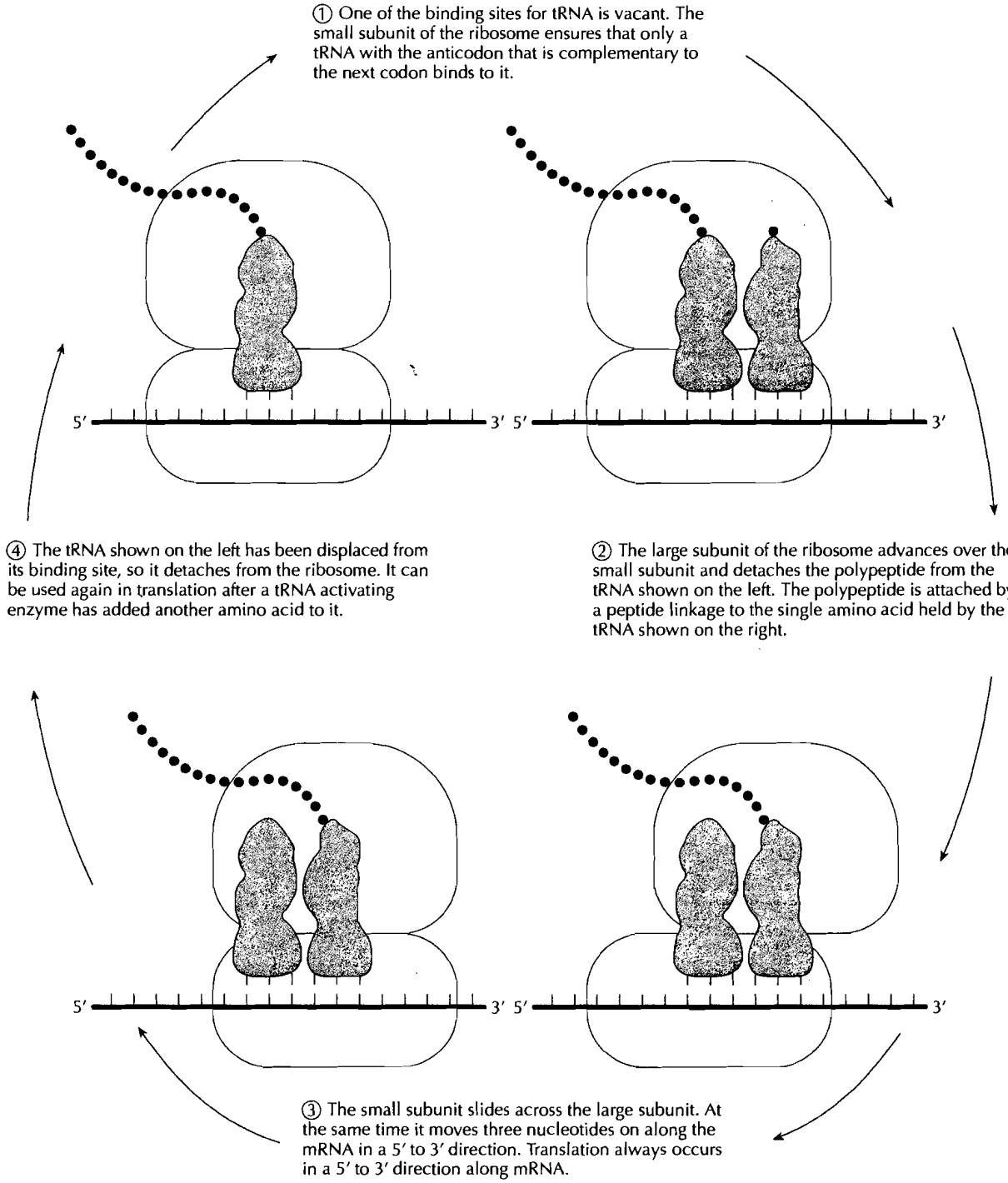
Polysomes and polypeptide elongation

The figure (right) is an electron micrograph showing groups of ribosomes called **polysomes** (or polyribosomes). A polysome is a group of ribosomes moving along the same mRNA, as they simultaneously translate it. Each ribosome follows a series of steps that is repeated many times to translate the mRNA. One amino acid is added to the elongating polypeptide each time the cycle of steps is repeated (see below). As ribosomes move along the mRNA towards the 3′ end, the polypeptide is gradually elongated.



(× 180 000)

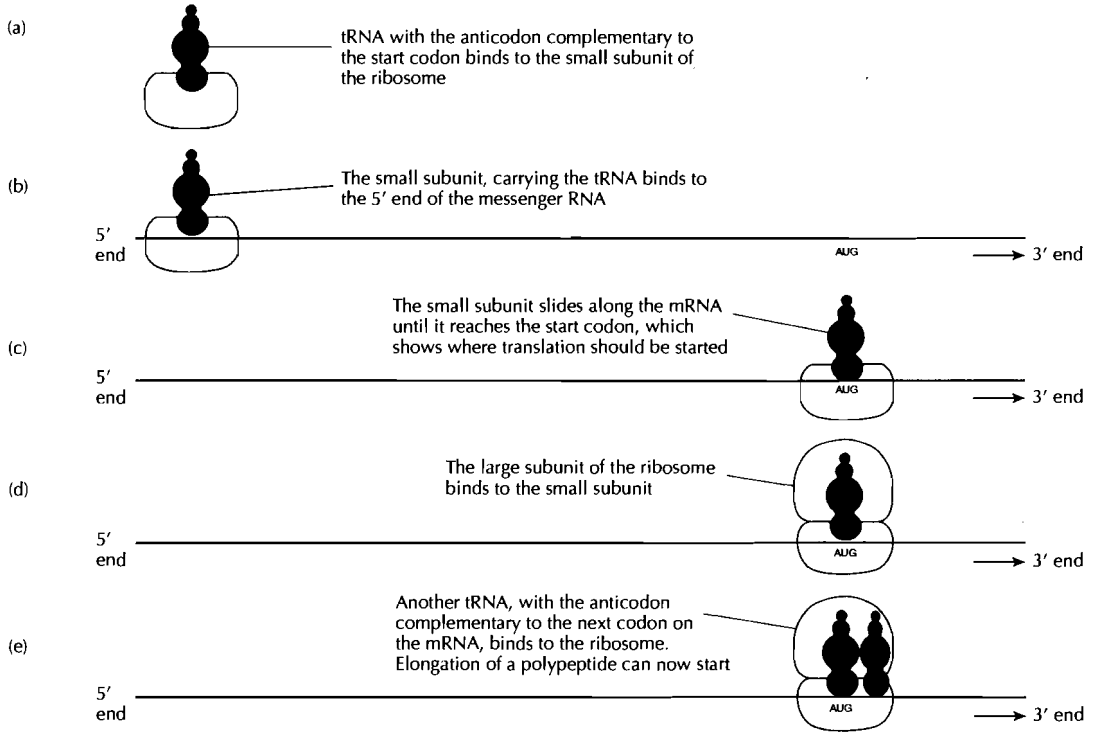
POLYPEPTIDE ELONGATION



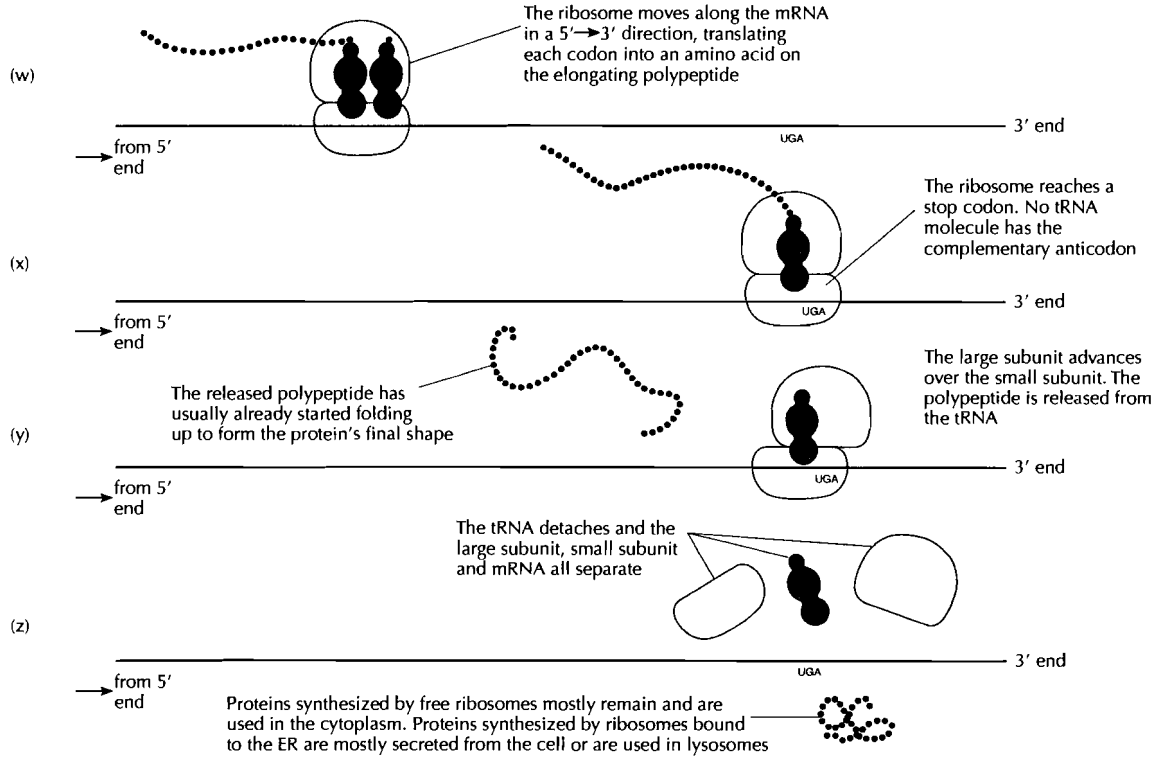
Starting and stopping translation

Special steps are needed to start the process of translation and to stop it. These steps are called **initiation** and **termination**. The three stages of translation are thus initiation, elongation and termination.

INITIATION OF TRANSLATION

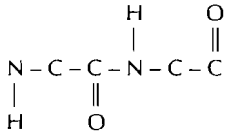


TERMINATION OF TRANSLATION



Intramolecular bonding in proteins

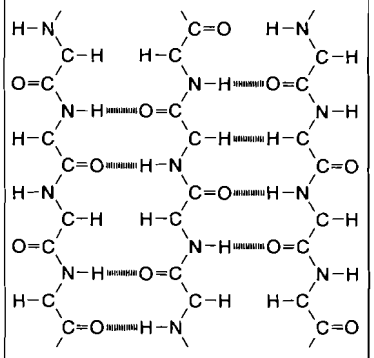
Polypeptides have a main chain consisting of a repeating sequence of covalently bonded carbon and nitrogen atoms: N – C – C – N – C – C, and so on. Each nitrogen atom has a hydrogen atom bonded to it (N – H). Every second carbon atom has an oxygen atom bonded to it (C = O).

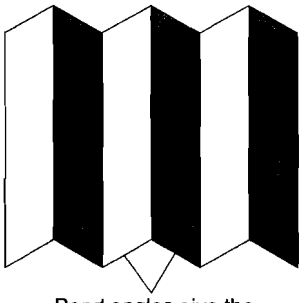


Hydrogen bonds can form between N – H and C = O groups, if they are brought close together. For example, if sections of polypeptide run parallel, hydrogen bonds can form between them. The structure that develops is called a ß-pleated sheet. If the polypeptide is wound into a right-handed helix, hydrogen bonds can form between adjacent turns of the helix. The structure that develops is called an α-helix. Because the groups forming hydrogen bonds are regularly spaced, secondary structures always have the same dimensions.

In addition to the hydrogen bonding in ß-pleated sheets and α-helices, there are many other types of bonding. Most of these involve the R groups of the amino acids. The figure (below) shows some of these bonds.

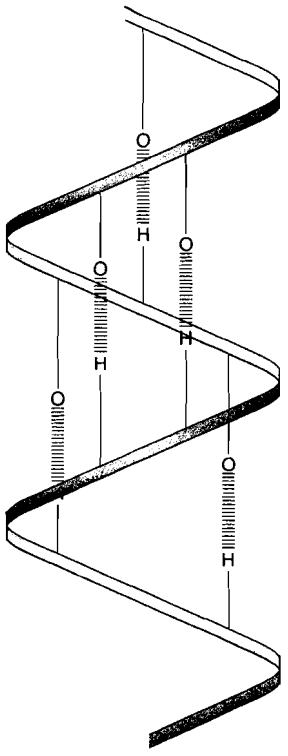
β-PLEATED SHEET





Bond angles give the sheet a pleated shape

α -HELIX



Types of intramolecular bond in proteins

Ionic bonds can form between positively and negatively charged R groups

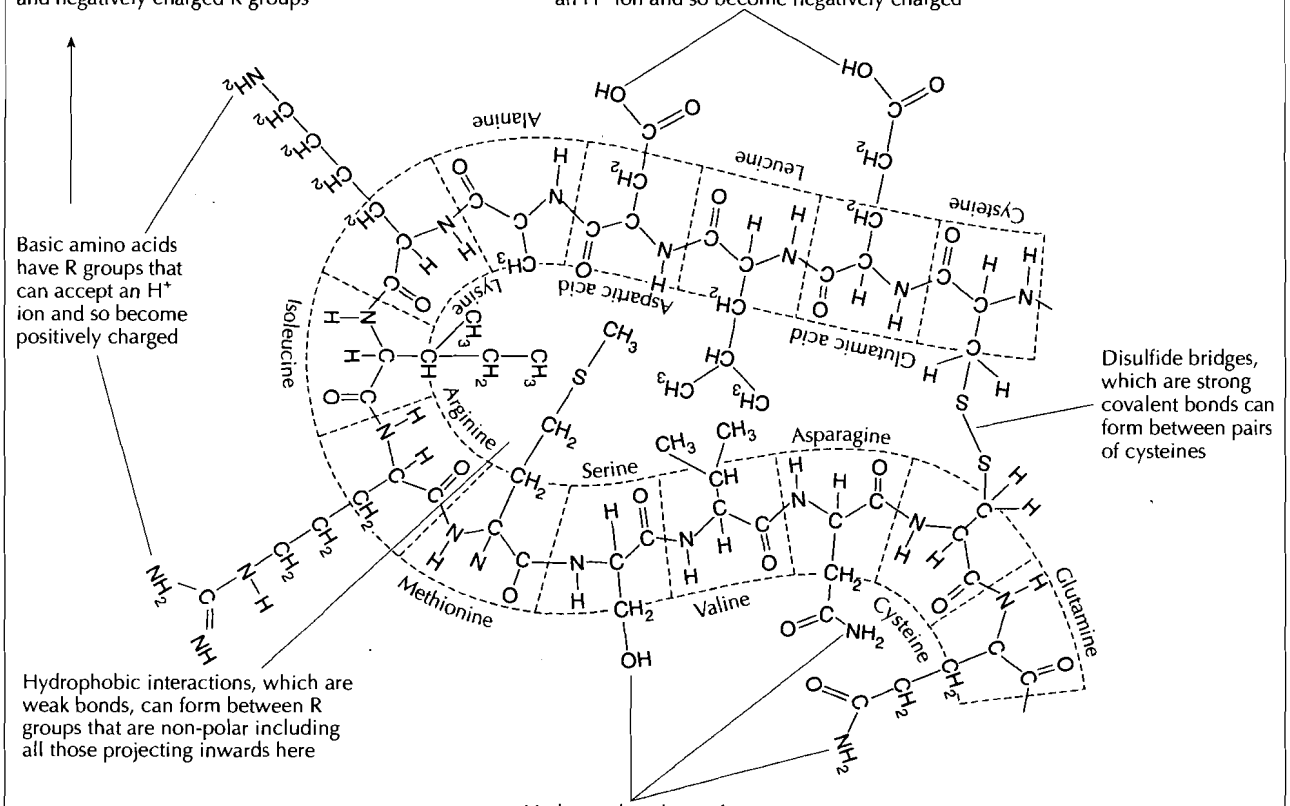
Acidic amino acids have R groups that can lose an H^+ ion and so become negatively charged

Basic amino acids have R groups that can accept an H^+ ion and so become positively charged

Disulfide bridges, which are strong covalent bonds can form between pairs of cysteines

Hydrophobic interactions, which are weak bonds, can form between R groups that are non-polar including all those projecting inwards here

Hydrogen bonds can form between some R groups



Protein structure

Proteins have a complex structure, which can be explained by defining four levels of structure, primary, secondary, tertiary and quaternary structure.

PRIMARY STRUCTURE

Primary structure is the number and sequence of amino acids in a polypeptide. Most polypeptides consist of between 50 and 1000 amino acids. The primary structure is determined by the base sequence of the gene that codes for the polypeptide. The figure (below) shows the primary structure of ß-endorphin, a protein consisting of a single polypeptide of 31 amino acids that acts as a neurotransmitter in the brain.

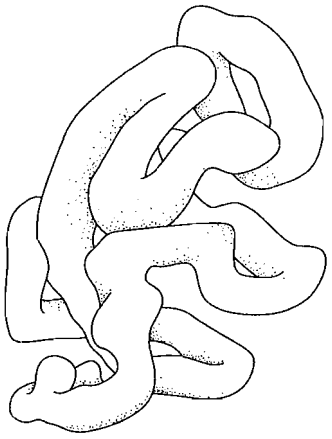
Primary structure of ß-endorphin

Alanine → Isoleucine → Isoleucine → Lysine
Asparagine → Alanine → Histidine → Lysine
Lysine → Glycine → Glutamine → Tyrosine
Glycine → Glycine → Phenylalanine → Methionine
Threonine → Serine → Glutamic acid → Lysine
Serine → Glutamine → Threonine → Proline
Leucine → Valine → Threonine → Leucine
Phenylalanine → Lysine → Asparagine

TERTIARY STRUCTURE

Tertiary structure is the three-dimensional conformation of a polypeptide. It is formed when the polypeptide folds up after being produced by translation. The conformation is stabilized by intramolecular bonds that form between amino acids in the polypeptide, especially between their R groups. These include ionic bonds, hydrogen bonds, hydrophobic interactions and disulfide bridges. The intramolecular bonds are often formed between amino acids that are widely separated in the primary structure of the polypeptide, but which are brought together during the folding process.
The figure below shows the tertiary structure of lysozyme using the sausage model.

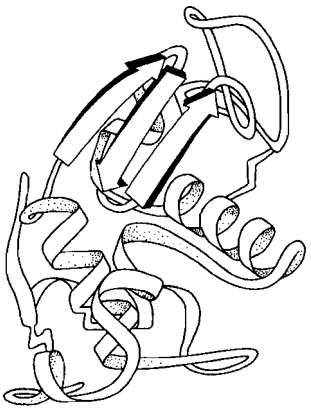
Sausage model of lysozyme



SECONDARY STRUCTURE

Secondary structures are regular repeating structures, including ß-pleated sheets and α-helices stabilized by hydrogen bonds between groups in the main chain of the polypeptide. In many proteins, parts of the polypeptide form secondary structures and other parts do not. In some proteins secondary structures do not form at all. In a few proteins almost all of the polypeptide forms secondary structures. For example almost all of myosin molecules is α-helix and almost all of fibroin (silk protein) is β-pleated sheet.
The figure (below) shows the position of secondary structures in lysozyme, using the ribbon model. Sections of α-helix are represented by helical ribbons and sections of β-pleated sheet are represented by arrows.

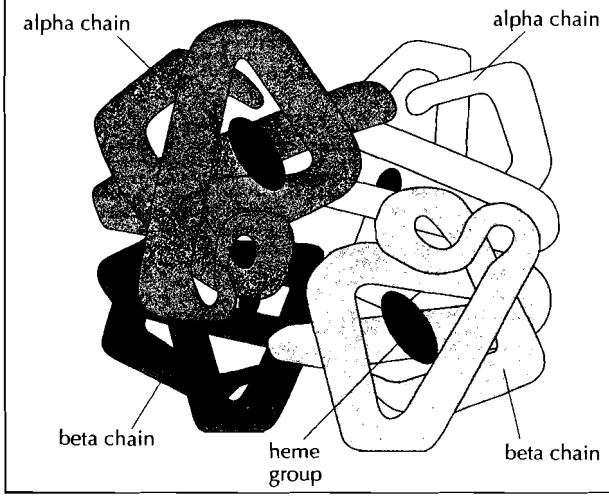
Ribbon model of lysozyme



QUATERNARY STRUCTURE

Quaternary structure is the linking together of two or more polypeptides to form a single protein. For example, insulin consists of two polypeptides linked together, collagen consists of three polypeptides and hemoglobin consists of four.
In some cases proteins also contain a non-polypeptide structure called a **prosthetic group**. Each of the four polypeptides in hemoglobin is linked to a heme group, which is not made of amino acids. Proteins with a prosthetic group are called **conjugated proteins**.
The figure (below) shows the quaternary structure of hemoglobin.

Sausage model of hemoglobin



Protein functions

Proteins have a huge range of functions in living organisms. Some proteins are located in membranes. The functions of membrane proteins are listed on page 7. Six of the functions of non-membrane proteins are listed (below). Proteins can also be used as food stores, for example, casein in milk, as pigments, for example, opsin in the retina and as toxins as in some snake venom.

Function	Example	Details	Shape
Enzymes	Catalase	The function of catalase is to catalyse the conversion of hydrogen peroxide, a toxic waste product of metabolism, into water and oxygen	Globular
Structural	Collagen	The function of collagen is to strengthen bone, tendon and skin. These tissues all produce tough collagen fibres in the spaces between their cells	Fibrous
Transport	Hemoglobin	The function of hemoglobin is to bind oxygen in the lungs and to transport it to respiring tissues	Globular
Movement	Myosin	The function of myosin (with another protein called actin) is to cause contraction in muscle fibres and as a result cause movement in animals	Fibrous
Hormones	Insulin	The function of insulin is to bind to receptors in the plasma membranes of target cells and stimulate them to remove glucose from the blood	Globular
Defence	Immunoglobulin	The function of immunoglobulin is to act as antibodies. Part of the immunoglobulin molecule can be varied, so that an almost endless variety of different antibodies can be produced	Globular

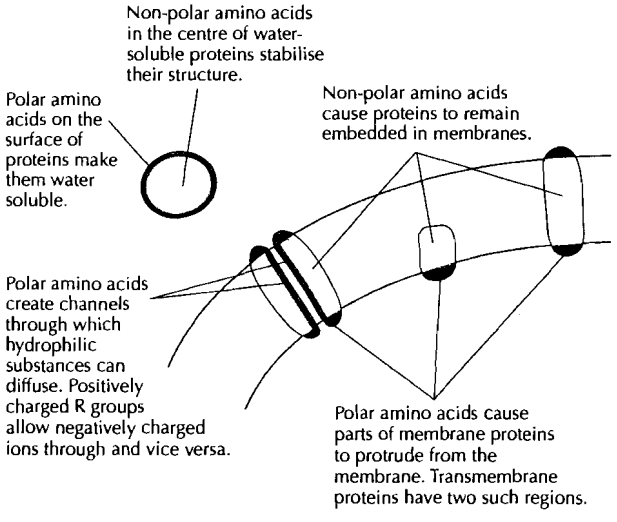
FIBROUS AND GLOBULAR PROTEINS

The table (above) indicated the shape of each of the named proteins. Proteins can be divided into two types according to their shape, fibrous or globular. Fibrous proteins have a long and narrow shape. They are mostly insoluble in water. Globular proteins have a rounded shape. They are mostly soluble in water.

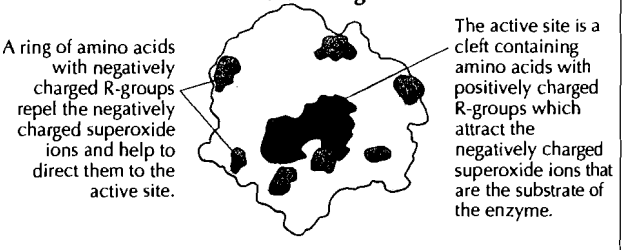
POLAR AND NON-POLAR AMINO ACIDS IN PROTEINS

Amino acids can be divided into two types according to the chemical characteristics of their R group. Polar amino acids have hydrophilic R groups and non-polar amino acids have hydrophobic R groups. The distribution of polar and non-polar amino acids in a protein molecule influence where the protein is located in a cell and what function it can carry out. The figures (below) show examples of this.

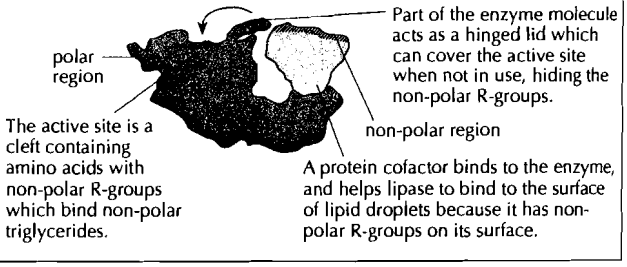
Positions of proteins in and out of membranes



Superoxide dismutase – an enzyme found in all aerobic organisms



Lipase – an enzyme that works in the small intestine



ENERGY CHANGES DURING CHEMICAL REACTIONS

During chemical reactions, reactants are converted into products. Before a molecule of the reactant can take part in the reaction it has to gain some energy. This is called the **activation energy** of the reaction. The energy is needed to break bonds within the reactant. Later, during the progress of the reaction, energy is given out as new bonds are made. In **exergonic** reactions this amount of energy is greater than the activation energy. In **endergonic** reactions it is less. Enzymes reduce the activation energy of the reactions that they catalyse and therefore make it easier for reactions to occur.

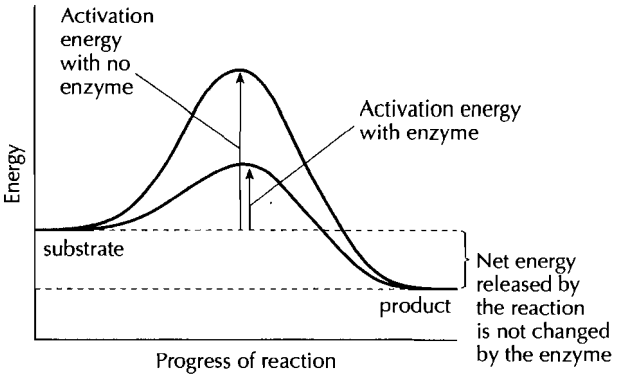
The chemical environment provided by the active site for the substrate causes changes within the substrate molecule, which weakens its bonds. The substrate is changed into a transition state, which is different from the transition state during the reaction when an enzyme is not involved. The transition state achieved during binding to the active site has less energy and this is how enzymes are able to reduce the activation energy of reactions.

THE INDUCED FIT MODEL OF ENZYME ACTIVITY

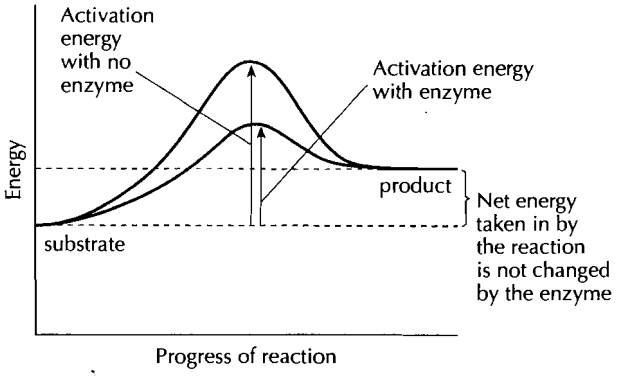
Biochemists have investigated many enzymes and found that the lock and key model does not fully explain the binding of the substrate to the active site. Until the substrate binds, the active site does not fit the substrate precisely. As the substrate approaches the active site and binds to it, the shape of the active site changes and only then does it fit the substrate. The substrate induces the active site to change, weakening bonds in the substrate during the process and thus reducing the activation energy. The figure (below) shows the induced fit model of enzyme activity.

Some enzymes can have quite broad specificity, for example some proteases. The induced fit model explains this better than the lock and key model – if the shape of an active site alters when substrates bind, several different but similar substrates could easily bind successfully to it.

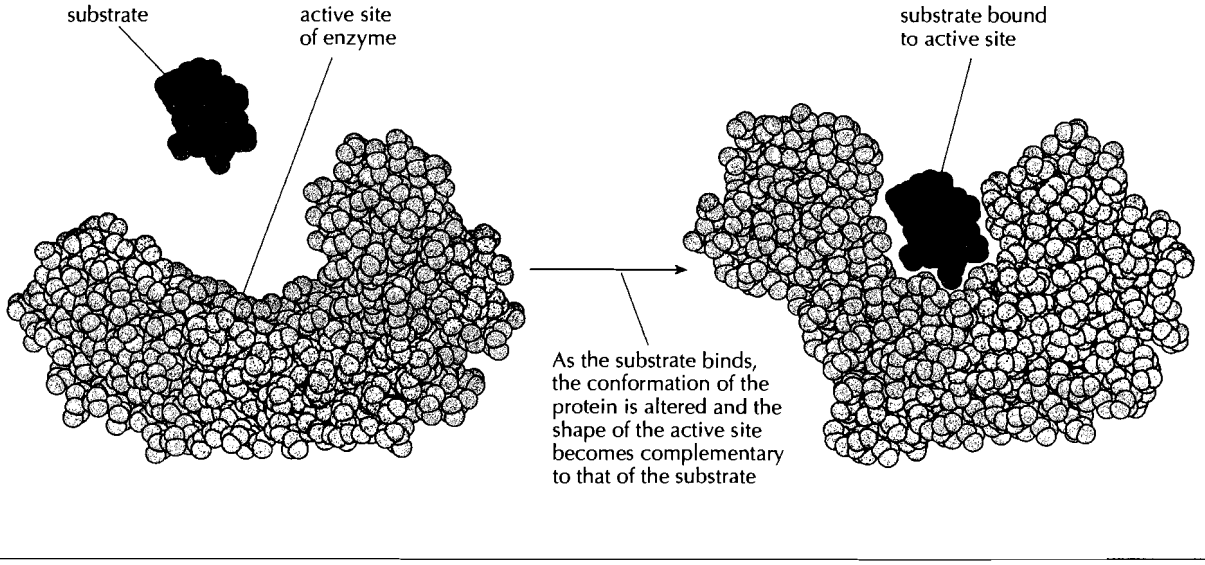
Energy changes during exergonic reactions



Energy changes during endergonic reactions



In living organisms, endergonic reactions are coupled with exergonic reactions, for example hydrolysis of ATP. The endergonic reaction can then occur more easily.



Enzyme inhibition

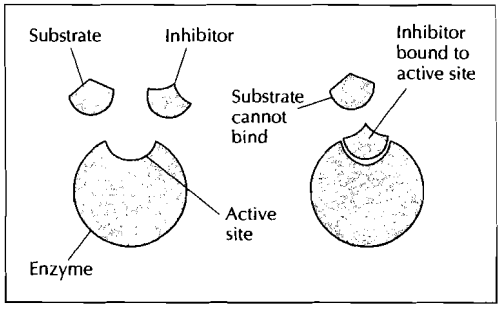
Some chemical substances reduce the activity of enzymes or even prevent it completely. These substances are called enzyme inhibitors. Some enzyme inhibitors are **competitive** and some are **non-competitive**. Figures below are a comparison of these types of inhibitor, with an example of each.

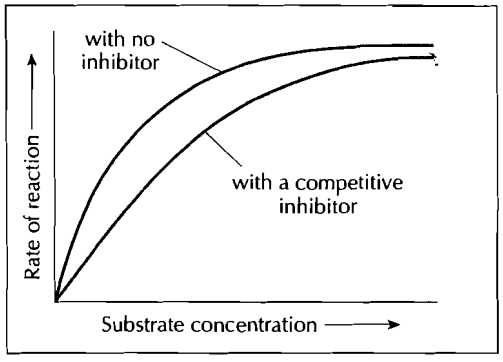
Competitive inhibition

The substrate and inhibitor are chemically very similar

The inhibitor binds to the active site of the enzyme

While the inhibitor occupies the active site, it prevents the substrate from binding and so the activity of the enzyme is prevented until the inhibitor dissociates

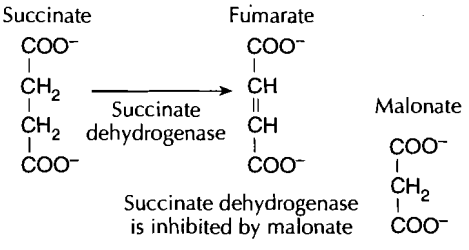




With a fixed low concentration of inhibitor, increases in the substrate concentration gradually reduce the effect of the inhibitor.
The inhibitor and substrate compete for the active site. When the substrate binds to the active site, the inhibitor cannot bind, so the proportion of enzyme molecules that are inhibited becomes less and less. When there are many more substrate molecules than inhibitor molecules, the substrate always wins the competition and binds to the active site. The same maximum enzyme activity rate is then reached as when there is no inhibitor.

EXAMPLE

Succinate → Fumarate (Succinate dehydrogenase)

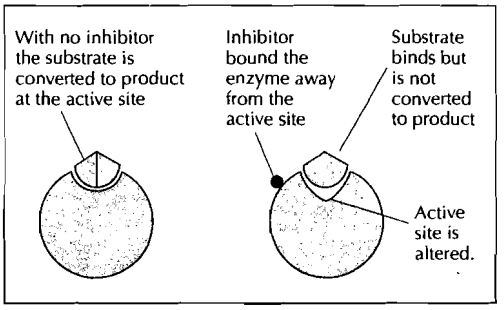


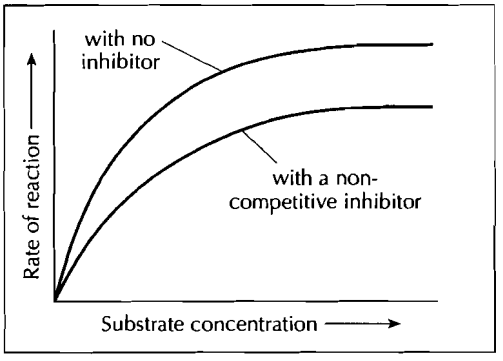
Non-competitive inhibition

The substrate and active site are not similar

The inhibitor binds to the enzyme at a different site from the active site

The inhibitor changes the conformation of the enzyme. The substrate may still be able to bind, but the active site does not catalyse the reaction, or catalyses it at a slower rate





With a fixed low concentration of inhibitor, increases in substrate concentration increase enzyme activity. However, the substrate and inhibitor are not competing for the same site. The substrate cannot prevent the binding of the inhibitor, even at very high substrate concentrations. Some of the enzyme molecules therefore remain inhibited and the maximum enzyme activity rate reached is lower than when there is no inhibitor

EXAMPLE

Metal ions including copper (Cu^{2+}), mercury (Hg^{2+}) and silver (Ag^{+}) act as non-competitive inhibitors of many enzymes by binding reversibly to the –SH groups of cysteine – the amino acid that forms disulfide bridges. This disrupts the structure of the enzyme:

$$-SH + Ag^{+} \longrightarrow -S - Ag + H^{+}$$

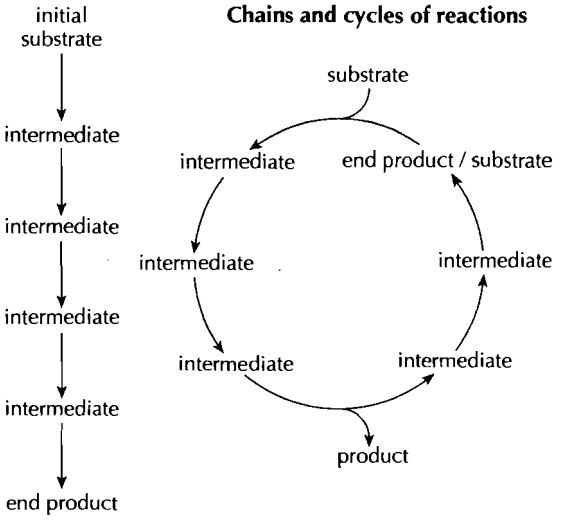
METABOLIC PATHWAYS

Metabolic pathways have these features:

- They consist of many chemical reactions that are carried out in a particular sequence.
- An enzyme catalyses each reaction.
- All the reactions occur inside cells.
- Some pathways build up organic compounds (anabolic pathways) and some break them down (catabolic pathways).
- Some metabolic pathways consist of chains of reactions. Glycolysis is an example of a chain of reactions–a chain of ten enzyme-controlled reactions converts glucose into pyruvate.
- Some metabolic pathways consist of cycles of reactions, where a substrate of the cycle is continually regenerated by the cycle. The Krebs cycle is an example.

The figure (opposite) shows the general pattern of reactions in a chain and a cycle.

Chains and cycles of reactions



ALLOSTERY AND THE CONTROL OF METABOLIC PATHWAYS

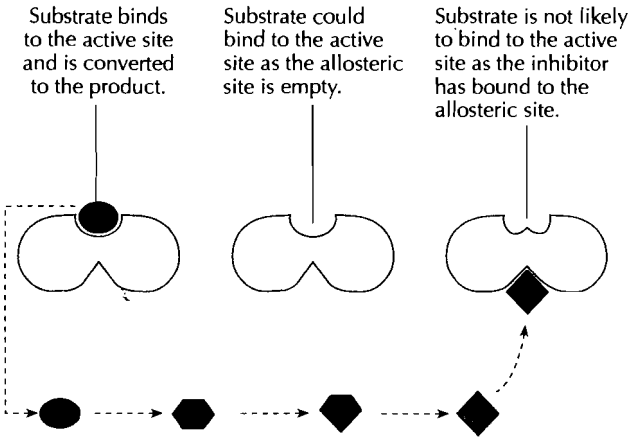
In many metabolic pathways, the product of the last reaction in the pathway inhibits the enzyme that catalyses the first reaction. This is called **end-product inhibition**.

The enzyme that is inhibited by the end products is an example of an **allosteric** enzyme. Allosteric enzymes have two non-overlapping binding sites. One of these is the active site. The other is the allosteric site.

In this case the allosteric site is a binding site for the end product. When it binds, the structure of the enzyme is altered so that the substrate is less likely to bind to the active site. This is how the end-product acts as an inhibitor. Binding of the inhibitor is reversible and if it detaches, the enzyme returns to its original conformation, so the active site can bind the substrate easily again (right).

The advantage of this method of controlling metabolic pathways is that if there is an excess of the end-product the whole pathway is switched off and intermediates do not build up. Conversely, as the level of the end-product falls, more and more of the enzymes that catalyse the first reaction will start to work and the whole pathway will become activated. End product inhibition is an example of negative feedback (see example below).

End-product inhibition



The substrate of the first enzyme in the metabolic pathway is converted by the pathway into an inhibitor of the enzyme.

An example of end product inhibition