

Fig. 1.19. Insulin has one sulfide bridge within chain A and two sulfide bridges between chain A and chain B.

1.1.2.3 Degradation of proteins

The three-dimensional structure of a protein which is held together by hydrogen bonding, electrostatic attraction and sulfide bridges is very sensitive to its chemical and physical environment. A change in pH, temperature or ionic strength disrupts these interactions and causes the protein to unfold; this process is called *denaturation*. The protein loses activity once its normal shape is lost. In some cases, this denaturation is reversible and the protein can *renature*, although in most cases the activity loss is permanent.

1.2 Nucleic Acids

Nucleic acids are long, linear biomolecules that can have molecular weights of several million Da. There are two classes of nucleic acids, *deoxyribonucleic acid (DNA)* and *ribonucleic acid (RNA)*.

DNA contains the “code of life.” It is the hereditary molecule in all cellular life forms as it is used by cells to store and transmit genetic information. During cell division, exact copies of DNA are made. Cells use DNA to determine and control the synthesis of proteins with the help of messenger RNA (mRNA).

RNA is essential for the synthesis of proteins in the cells. Messenger RNA (mRNA) is synthesised in the cell nucleus as a transcript of a specific part of DNA.

The mRNA leaves the nucleus and enters the cell cytoplasm where it dictates the synthesis of proteins from amino acids. *Transfer RNA* (tRNA) delivers amino acids to the exact place in the cytoplasm where the proteins are synthesised.

1.2.1 The Structure of Nucleic Acids

Nucleic acids are made up from three components: *nucleobases* (usually referred to as bases), *sugars* and *phosphoric acid*. The nucleobases are derivatives of purine and pyrimidine (Figs. 1.20 and 1.21). Both DNA and RNA contain the purines Adenine (A) and Guanine (G). Of the pyrimidines, Thymine (T) and Cytosine (C) are components of DNA whereas Uracil (U) and Cytosine (C) are components of RNA. The sugar component of DNA is β -D-deoxyribose, while RNA contains β -D-ribose, (Fig. 1.22). These components are summarised in Table 1.4.

How are these linked to each other to form macromolecular DNA and RNA? A *nucleoside* is formed by one of the nucleobases covalently binding to a sugar (Fig. 1.23, left). This N-glycosidic bond is formed between the C1' atom of the

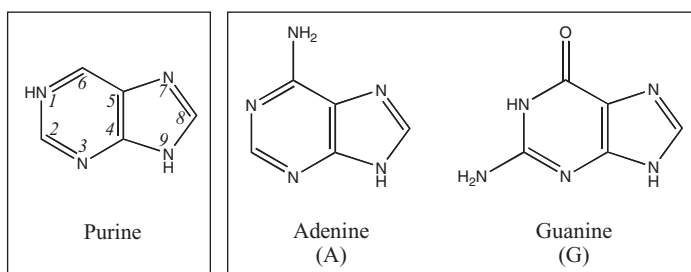


Fig. 1.20. Left: The structure of Purine, Right: The Purine derivatives Adenine and Guanine are found as bases in both DNA and RNA.

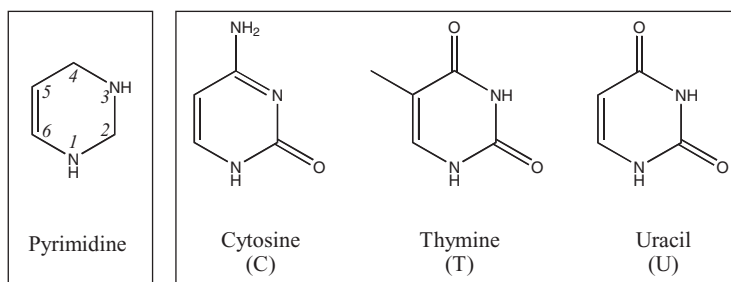


Fig. 1.21. Left: The structure of Pyrimidine. Right: Cytosine and Thymine are the Pyrimidine derivatives found in DNA, while Uracil and Thymine are found in RNA.

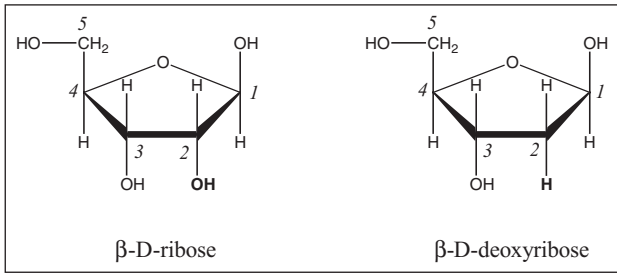


Fig. 1.22. The pentose β -D-ribose occurs in RNA. β -D-deoxyribose is the sugar component in DNA.

Table 1.4. Components of nucleic acids.

| | | DNA | RNA |
|--------------------|--------------------|------------------------|--------------------|
| Nucleobases | Purines | Adenine Guanine | Adenine Guanine |
| | Pyrimidines | Thymine Cytosine | Uracil Cytosine |
| Sugar | | β -D-deoxyribose | β -D-ribose |
| Phosphate | | phosphoric acid | phosphoric acid |

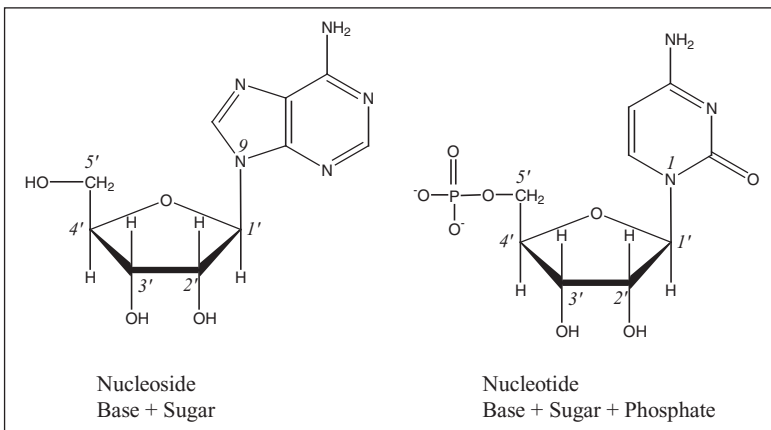


Fig. 1.23. Left: A nucleoside consisting of the base Adenine (A) and ribose. Right: A nucleotide consisting of Cytosine, ribose and one phosphate group.

sugar and either the N9 atom of a purine base or the N1 atom of a pyrimidine base. A *nucleotide* is a nucleoside with one or more phosphate groups covalently attached to the 5'-hydroxylic group of the sugar (Fig. 1.23, right). At pH 7, the acidic phosphate groups are negatively charged.

Nucleotide units can connect to each other to form a chain, a *polynucleotide*. The phosphate residue at the 5' position of one sugar bonds to the 3'-hydroxyl group of another sugar (Fig. 1.24). Thus, a strand is formed with the sugar and

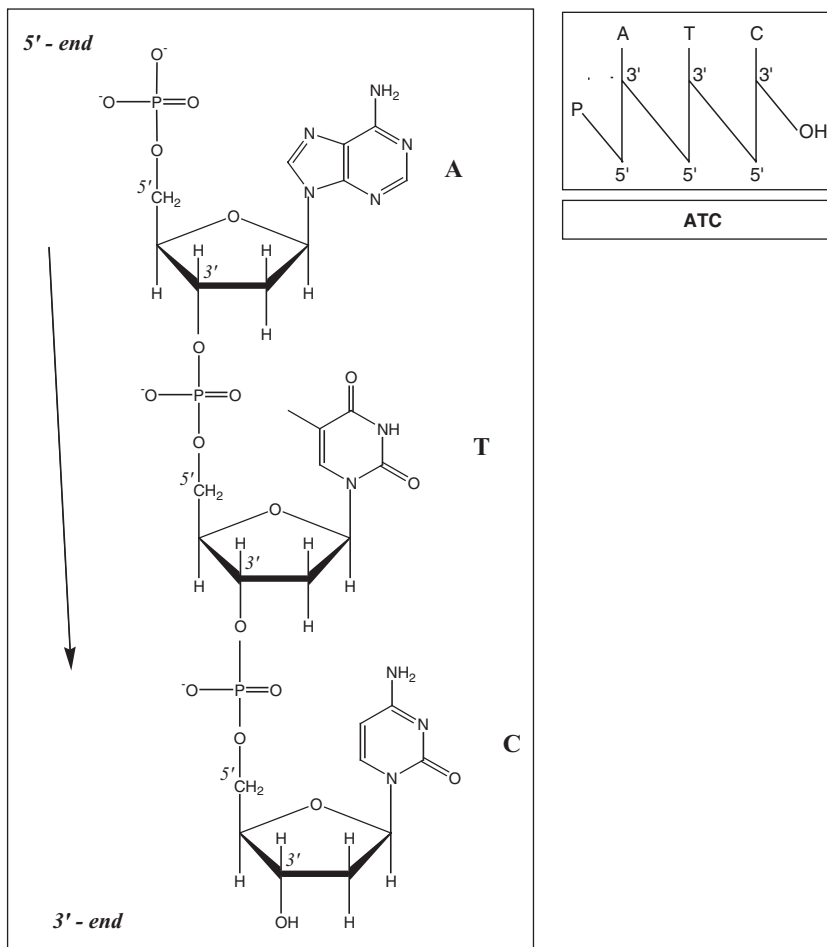


Fig. 1.24. Left: structure of a polynucleotide with the bases Adenine (A), Thymine (T) and Cytosine (C). Right: Abbreviated writing of the same polynucleotide sequence, as bar structure and as listing of nucleobases with hyphens.

phosphate units as a *backbone* and the nucleobases as *side groups*. In higher life forms, these strands can consist of millions of nucleotides.

To give the sequence a start and an end point, it is read from the nucleotide with the free phosphate group (the 5'-end) to the nucleotide with the free 3'-hydroxyl-group (the 3'-end).

The chemical structure of the polynucleotide can be described as shown in the left hand side of Fig. 1.24 or by using short forms with bars or by just listing the nucleobases in the sequence and omitting the sugar and phosphate.

1.2.1.1 3D structure of DNA

The three-dimensional structure of DNA was discovered by Francis Crick and James Watson in 1953. DNA as found in the cell nucleus has the shape of a *right twisted double helix* consisting of two polynucleotide strands twisted around each other (Fig. 1.25, left). The hydrophilic backbones composed of the sugar and phosphate groups are on the outside of the helix, while the hydrophobic bases are on the inside. The bases are connected to each other by weak hydrogen bonds to form *base pairs*. (Fig. 1.25, middle and right). The two strands run in opposite directions so that the 3'-end of one strand and the 5'-end of the other strand are linked.

Due to steric reasons, only two combinations of base pairs are possible: *Adenine with Thymine (A-T or T-A)* and *Guanine with Cytosine (G-C or C-G)* (Fig. 1.26).

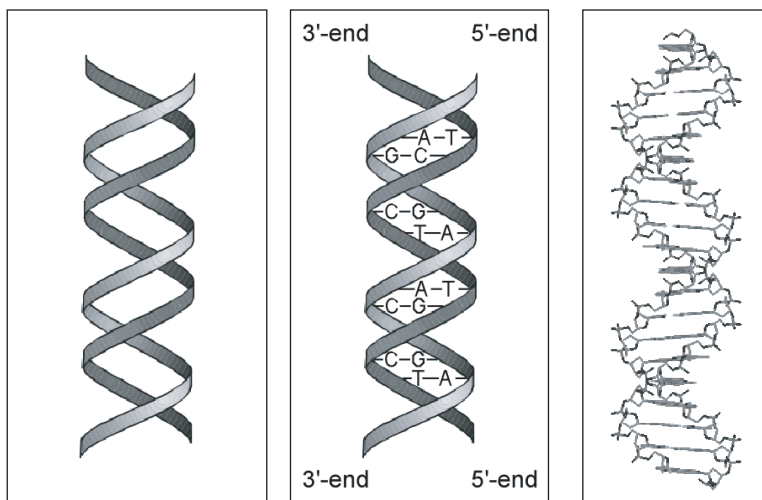


Fig. 1.25. Left: Schematic of a double helix. Middle: A DNA double helix with base pairs and sequence direction. Right: Structure of a DNA helix showing the parallel base pairs inside the helix.

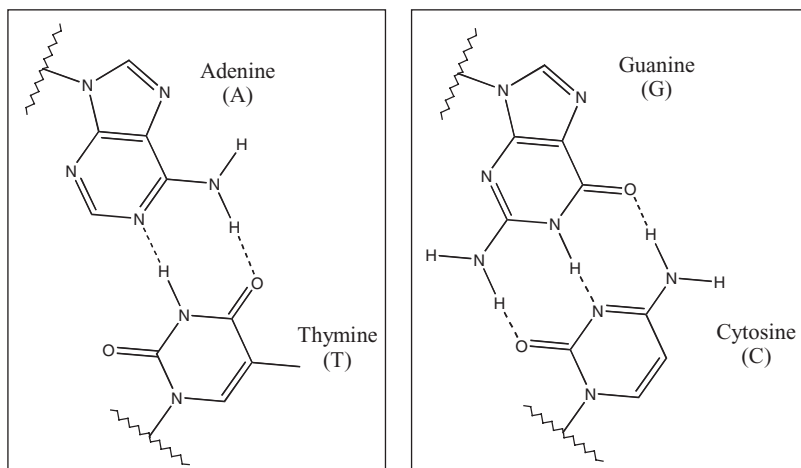


Fig. 1.26. Base pairs are formed between Adenine and Thymine (A-T) and Guanine and Cytosine (G-C).

The aromatic rings of the base pairs are parallel to each other, forming a twisted ladder-like structure. The diameter of a DNA helix molecule is about 2 nm. With millions of nucleotides in the strand, the length of a DNA molecule when laid out straight can measure several centimetres.

The two strands of the double helix are *complementary* to each other. The nucleobase-sequence of one strand unambiguously determines the sequence of the other strand.

Similar to the three-dimensional structure of proteins, three levels of organisation can be distinguished for DNA. The *primary structure* is determined by the sequence of nucleotides, usually written as the sequence of bases they contain. The *secondary structure* is given by the shape of the double stranded helix. This helical chain does not exist as a straight, long molecule. It forms turns and twists and folds. This coiling is referred to as the *tertiary structure* of DNA.

In comparison to proteins, DNA is a fairly simple molecule. It consists of only four different bases, which are repeated throughout the whole structure and the double helix is its only structural component.

When heated e.g. to 95°C or when deviating from physiological conditions, the hydrogen bonds between the two DNA strands are cleaved and the strands are separated from each other to form *single stranded DNA (ssDNA)*. This *denaturation* is usually a reversible process. When reverting to lower temperatures or to physiological conditions, the two strands can link back together to reform the double helix. Denaturation into ssDNA is a necessary step for the replication of DNA. Once cleaved, complementary daughter strands can be formed, which are

an exact copy of the original strand. Denaturation into ssDNA is also essential for the synthesis of mRNA in cell nuclei and for the polymerase chain reaction (PCR), which is used for DNA amplification (see section 6.2).

1.2.1.2 3D-structure of RNA

RNA is also a polynucleotide, but with ribose as the sugar component and Uracil (U) as a base instead of Thymine (T). Due to the additional –OH group on the ribose sugar molecule, steric hindrance is too great to allow for the formation of a double strand. Hence, RNA can only exist as a *single stranded* molecule. This strand can fold and loop and form base pairing with itself in certain places.

1.2.2 Synthesis of Proteins

One of the main tasks of the DNA is to initiate the synthesis of proteins as and when they are needed. Proteins are synthesised in the ribosomes of the cell cytoplasm. DNA, however, is found in the cell nucleus. So how is the information contained in the DNA passed out of the cell nucleus and into the cytoplasm? First, the DNA helix unfolds, and, in a process called *transcription*, a complementary strand of RNA is synthesised along a crucial part of one of the single DNA strands. This is the *messenger RNA (mRNA)* which leaves the cell nucleus and is transported into the manufacturing centres for proteins, the ribosomes. In the ribosome, *transfer RNA (tRNA)* delivers the amino acids required for polypeptide synthesis. The sequence of each group of three bases on the mRNA determines which amino acid is next in the peptide sequence. For example, the sequence AGC in the mRNA specifies the incorporation of the amino acid serine. This process is referred to as *translation* (Fig. 1.27). The genetic code, i.e. which sequence of bases in the DNA strand refers to which amino acid is given in Table 1.5.

To obtain its biological activity, the synthesised protein must fold into its native structure. Disulfide bridges must be formed. If the protein has a quaternary structure, then the different peptide chains must combine. Often, the protein undergoes a number of *post-translational modifications* to gain its full activity. The most common post-translational modification is the cleavage of one or several amino groups from the N- or C-terminus of the peptide chain. The side chains of the amino acids can undergo chemical modifications such as phosphorylation, acetylation, and methylation. *Glycoproteins* are synthesised by glycosylation, i.e. the addition of an oligosaccharide to the peptide chain. Similarly, *lipoproteins* can be formed. More than 150 such post-translational modifications are known. None of these is determined by the DNA sequence. However, these modifications are crucial for the biological activity of the proteins.

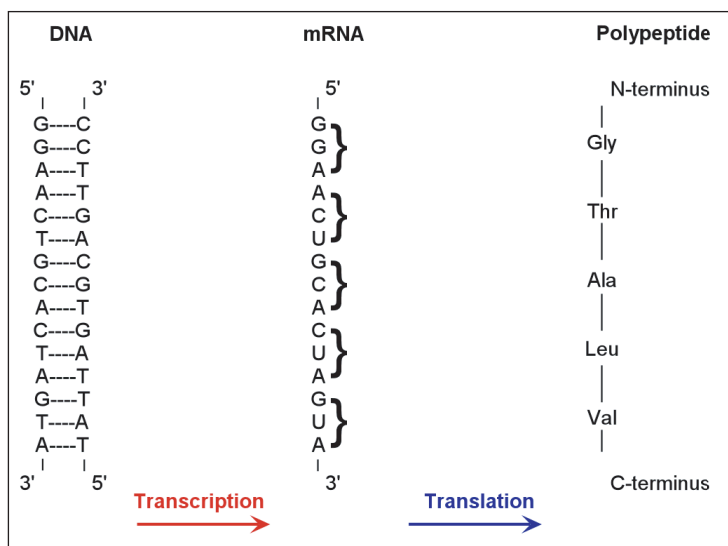


Fig. 1.27. Transcription of a DNA sequence into mRNA followed by translation into a polypeptide sequence.

Table 1.5. Genetic code. Each sequence of three bases in the mRNA determines which amino acid is used in the polypeptide (refer to Table 1.1 for amino acid abbreviations).

| First position 5' end | Second position | | | | Third position 3' end |
|--------------------------|-----------------|---------|----------|----------|--------------------------|
| | U | C | A | G | |
| U | UUU Phe | UCU Ser | UAU Tyr | UGU Cys | U |
| | UUC Phe | UCC Ser | UAC Tyr | UGC Cys | C |
| | UUA Leu | UCA Ser | UAA Stop | UGA Stop | A |
| | UUG Leu | UCG Ser | UAG Stop | UGG Trp | G |
| U | CUU Leu | CCU Pro | CAU His | CGU Arg | U |
| | CUC Leu | CCC Pro | CAC His | CGC Arg | C |
| | CUA Leu | CCA Pro | CAA Gln | CGA Arg | A |
| | CUG Leu | CCG Pro | CAG Gln | CGG Arg | G |
| A | AUU Ile | ACU Thr | AAU Asn | AGU Ser | U |
| | AUC Ile | ACC Thr | AAC Asn | AGC Ser | C |
| | AUA Ile | ACA Thr | AAA Lys | AGA Arg | A |
| | AUG Met | ACG Thr | AAG Lys | AGG Arg | G |
| G | GU Val | GCU Ala | GAU Asp | GGU Gly | U |
| | GUC Val | GCC Ala | GAC Asp | GGC Gly | C |
| | GUA Val | GCA Ala | GAA Glu | GGA Gly | A |
| | GUG Val | GCG Ala | GAG Glu | GGG Gly | G |