CBS p. 155 - 201

# 2. NUCLEIC ACIDS

## 2.1. General Arrangement of Nucleic Acids

Nucleic acids contain four building blocks:

DNA A T G C
RNA A U G C.

Here, A is adenine, T thymin, U uracil, G guanin, and C cytosin. The building blocks are hooked onto a backbone, as shown schematically in Fig. 2.1.

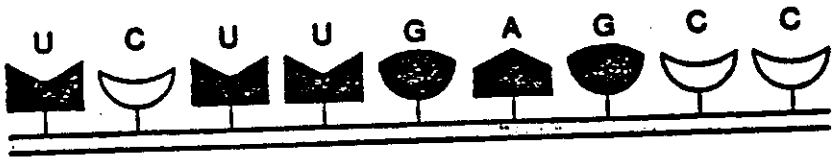



Fig. 2.1 Schematic arrangement of building blocks in nucleic acids.

The four building blocks are matched in pairs, and the interaction between the partners of a pair is stronger than between two non-partners. Matched partners are

| DNA | RNA |
|---|---|
| A=T | A=U |
| G≡C | G≡C. |

Here we have symbolically indicated that A and U are connected by two bonds, G and C by three. Matching thus is forced by the number of hydrogen bonds: A binds well to T (or U) via two bridges, G binds strongly to C via three bridges. The matching is shown schematically in Fig. 2.2.

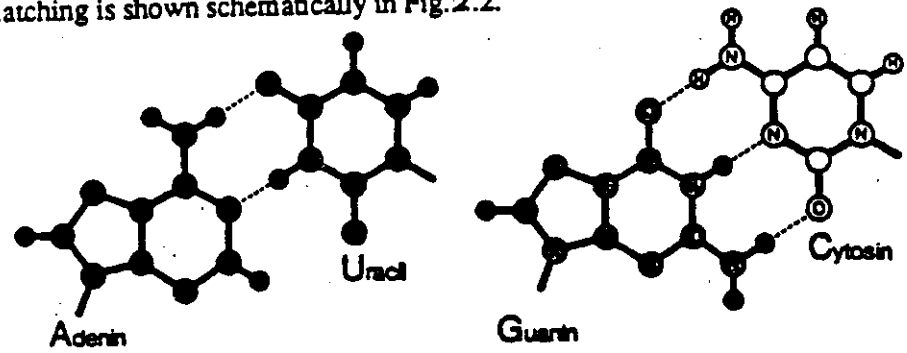



Fig. 2.2 Matching of base pairs.

Through the matching, AU and GC form base pairs that are nearly identical in size and form.

A single NA strand is most stable if it can fold such that a maximum number of matched pairs are produced. For short NA, the result is a "hairpin", for longer ones a "clover leaf". Bonding between two NA is particularly strong if each base on one

strand is bound to a complementary base on the other strand. Such an arrangement is obtained if two separate complementary chains bind. This arrangement is most stable in the famous Watson-Crick double helix. The three basic forms, hairpin, cloverleaf, and complementary chains are shown schematically in Fig. 2.3.

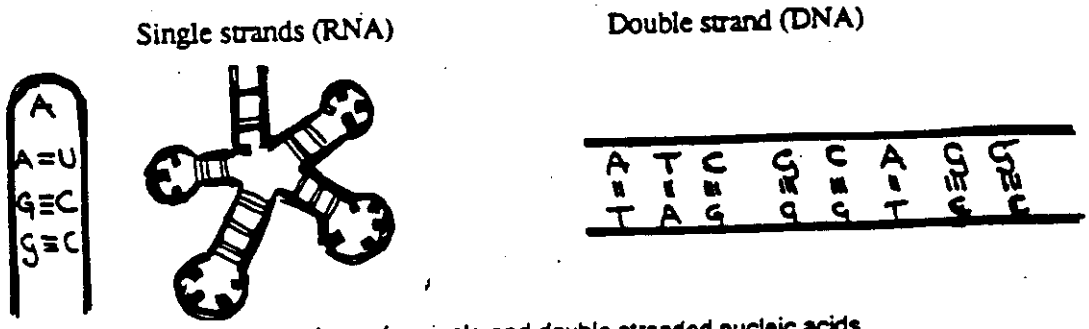



Fig. 2.3 Particularly stable forms for single and double stranded nucleic acids.

## 2.2. Mononucleotides – The Building Blocks

The four building blocks of NA are sketched schematically in Fig. 2.2 These nucleotides consist of a complex containing base + sugar + phosphate, as indicated in Fig. 2.4. The base is one of the logic building blocks and distinguishes the four units. Phosphate and sugar are nonspecific and provide the links. Sugar (carbohydrate): pentose

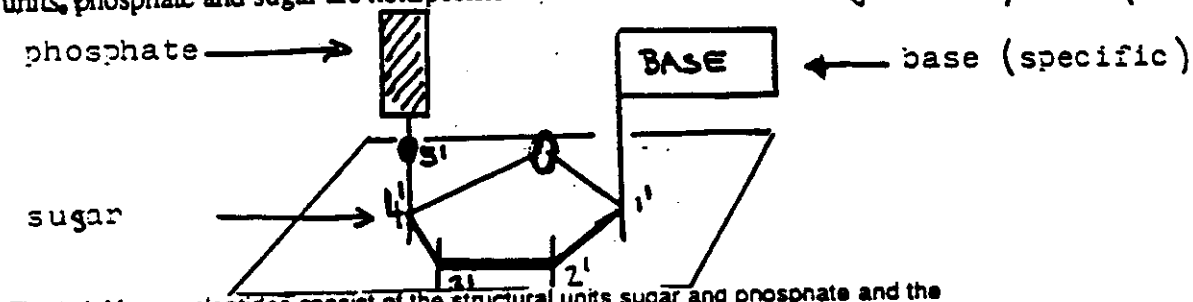



Fig. 2.4 Mononucleotides consist of the structural units sugar and phosphate and the specific base.

The *bases* are formed from purines and pyrimidines; the five most important ones are given in Fig. 2.5.

Purines (A,G) Pyrimidines (C;T,U)

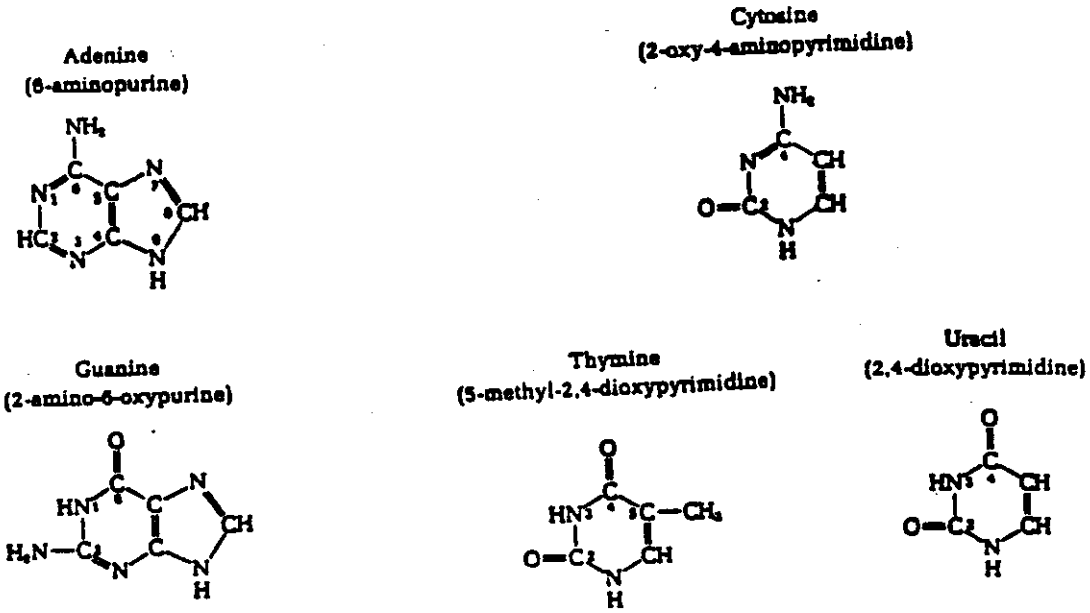



Fig. 2.5 The five important bases.

Names: A nucleoside is a purine or pyrimidine base linked to a pentose.
A nucleotide is a phosphate ester of a nucleoside.

In passing we mention that nucleotides are not only extremely important as the building blocks of RNA and DNA, but occur also in other roles: energy-carrying coenzymes, coenzymes in redox reactions and in transfer reactions.

## 2.3. Polynucleotides

DNA and RNA are linear polymers of successive mononucleotide units in which one mononucleotide is linked to the next by a phosphodiester bridge between the 3'-hydroxyl group of one nucleotide and the 5'-hydroxyl group of the next. The basic structure of such a polynucleotide is thus as in Fig. 2.6.

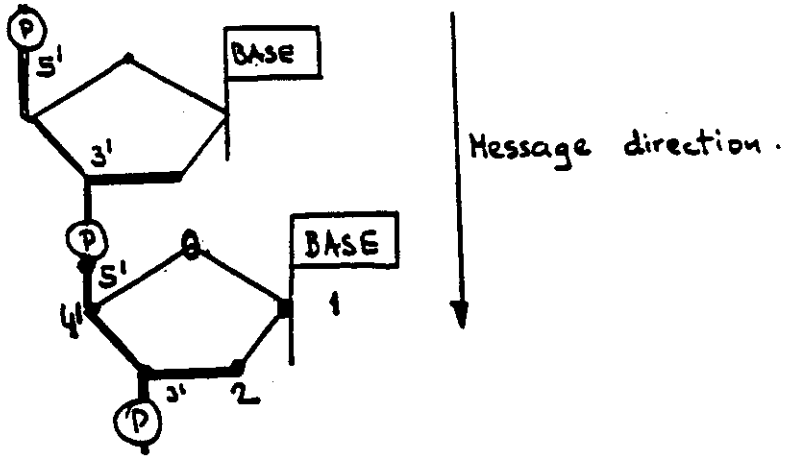



Fig. 2.6 DNA chain – basic arrangement.

Phosphates and sugars form the nonspecific structure, the bases are the specific letters of the genetic message. The message has a direction. One end of the chain has a 5'-OH group, the other a 3'-OH group, both not linked to other nucleotides. By definition, a base sequence (ACGAG..) is written in the 5' → 3' direction. (1' to 5' labels the carbon atoms.)

## 2.4. The Double Helix[5)]

Solutions of DNA are very viscous, suggesting long and rigid rather than compact and folded DNA molecules. The elucidation of the actual structure was based on X-ray data and on an understanding of the base pairing. Paired bases are already shown in Fig. 2.2; a more detailed figure is given below, in Fig. 2.7.

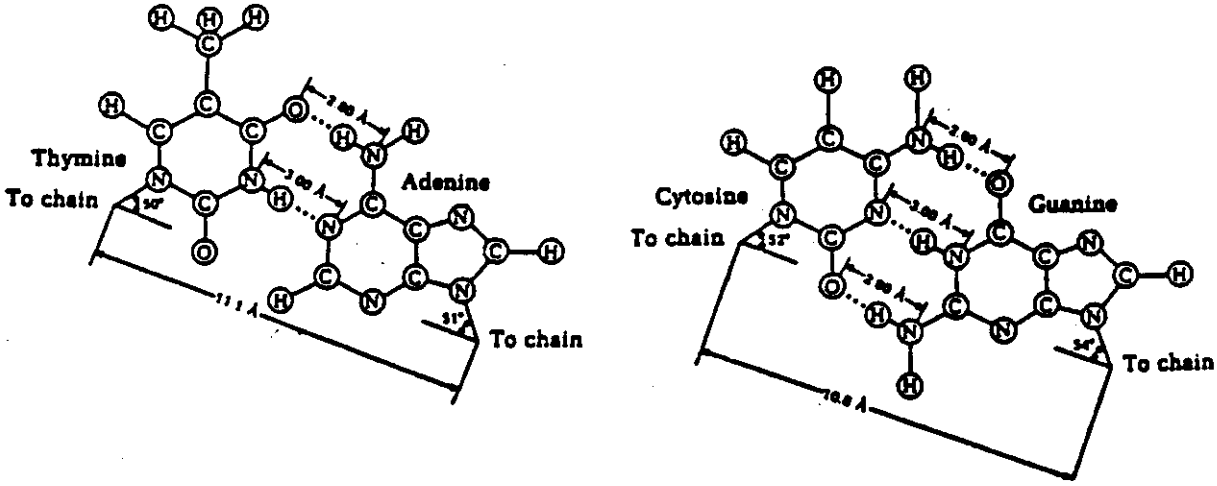
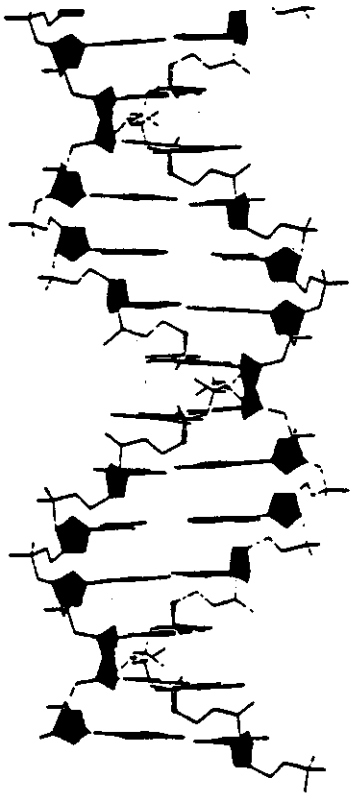



Fig. 2.7 The two base pairs.

Fig. 2.7 shows that the two base pairs have nearly identical dimensions. They can thus be arranged in an arbitrary sequence, without straining the backbone too much. X-ray data indicate a helical structure, with two periodicities, a major one of 3.4Å and a minor one of 34Å. This information is incorporated in the double helix of Watson and Crick.

The Watson-Crick model[6)] postulates two right-handed helical nucleotide chains coiled around the same axis to form a double helix. The two strands are antiparallel; their 3'-5'-phosphodiester bridges run in opposite directions. The

arrangement is shown schematically in Fig. 2.8. The coiling is such that the two chains cannot be separated unless the coils unwind. The two strands are bonded together by the hydrogen bonds of complementary bases. The 3.4Å periodicity is explained by assuming that the bases are stacked perpendicular to the long axis at a center-to-center distance of 3.4Å. The 34Å periodicity is explained by assuming that there are ten nucleotide residues in each complete turn of the helix.



Skeletal model of double-helical DNA. The structure repeats at intervals of 34 Å, which corresponds to ten residues on each chain.

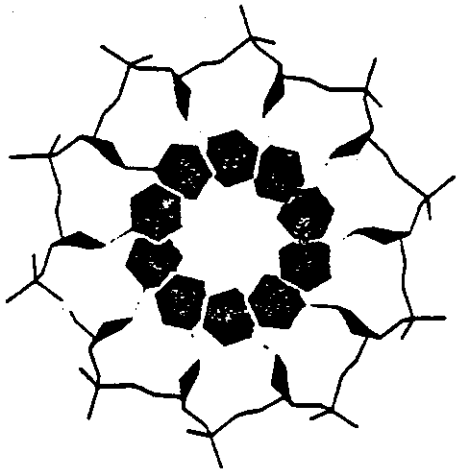


Diagram of one of the strands of a DNA double helix, viewed down the helix axis. The bases (all pyrimidines here) are inside, whereas the sugar-phosphate backbone is outside. The tenfold symmetry is evident.

Fig. 2.8 From Stryer.

Stability of the double helix is obtained through the hydrogen bonds between bases *and* through the proper arrangement of the components. The hydrophobic bases inside the double helix are shielded from the solvent water: the hydrophilic sugar residues and electrically charged phosphate groups are at the periphery, exposed to water. Initially only one form of DNA was recognized, the Watson-Crick double helix. It is now known that there are three different forms:[7,8]

*A-DNA* is double-helical DNA that contains about 11 residues per turn. The planes of the base pairs in this right-handed helix are tilted 20 degrees away from the perpendicular to the helix axis. A-DNA is formed by dehydration of B-DNA.

*B-DNA* is the classical Watson-Crick double helix with about 10 residues per turn. The helix is right-handed and the base planes are perpendicular to the helix axis.

*Z-DNA* is a left-handed helix containing about 12 base pairs per turn.

## 2.5. Some Properties of DNA

DNA is a remarkable biomolecule. While the diameter is always fixed to about 2 nm by the base pairing, the length depends on the number of base pairs and can exceed cm. DNA thus is extremely asymmetric. In Table 2.1, a few data are collected.

Table 2.1 Properties of some DNA

| Source | Number of base pairs | Length | Molecular weight | Information content (bit) |
|---|---|---|---|---|
| Polyoma virus | 4600 | 1.6 μm | $3 \times 10^6$ | $9 \times 10^3$ |
| T2 phage | 185000 | 63 μm | $122 \times 10^6$ | $3.7 \times 10^5$ |
| E coli bact. | $3.4 \times 10^6$ | 1.2 mm | $2.3 \times 10^9$ | $6.8 \times 10^6$ |
| Drosophila | $6 \times 10^7$ | 2.1 cm | $43 \times 10^9$ | $1.2 \times 10^8$ |
| Human cell | $\approx 3 \times 10^9$ | | $2 \times 10^{12}$ | $6 \times 10^9$ |

DNA molecules have been seen directly by electron microscopy. Many form closed loops. Some DNA molecules interconvert between linear and circular forms.

## 2.6. Replication

The Watson-Crick model leads directly to an understanding of the way in which information is replicated.[6] The bases on one strand unambiguously determine the bases on the complementary strand. A DNA molecule thus can duplicate as sketched in Fig. 2.9.

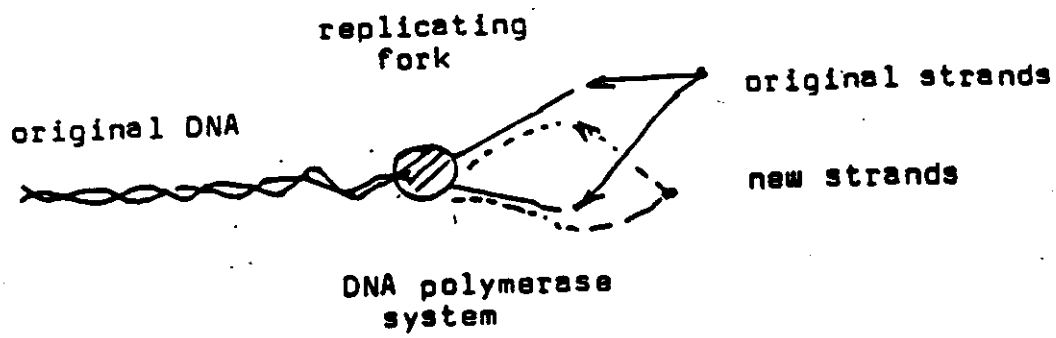



Fig. 2.9 DNA replication.

In a protein system (DNA polymerase), the original double strand is unwound. Each single strand then serves as a template for a new complementary strand. The complete system after duplication thus consists of two double strands, each identical (apart from replicating mistakes) to the original DNA. Actually, replication occurs in fragments which are then joined by DNA ligase.
Replication involves a proofreading mechanism.[9,10]

## 2.7. Storage of the Information

Anyone who has worked with a tangled climbing rope or a "nested" HiFi tape realizes how difficult it must be to store a DNA tape that is 1 m long but only 1 nm in diameter. Moreover, information from the tape should be extracted with a minimum of delay. How can the cell package the DNA helix into a chromosome so that retrieval, reading, replication, and repackaging is possible?

Some of the present ideas can be summarized as follows:[11-13] The basic organizers are the *histones* (proteins) with 5 classes known. They contain between 100 and 200 amino acids and have a spherical unit with about 2.5 nm diameter. In the first packaging step, the histones organize the DNA double helix into *nucleosomes*. These are particles about 10 nm in diameter, with histones on the inside and about 200 base pairs of DNA mainly on the outside. In the higher steps, the nucleosomes are organized into chromatin fibers. The organization continues through different molecular levels up to the chromosomes visible by microscope. In Fig. 2.10, an example of the lowest order of organization is shown.

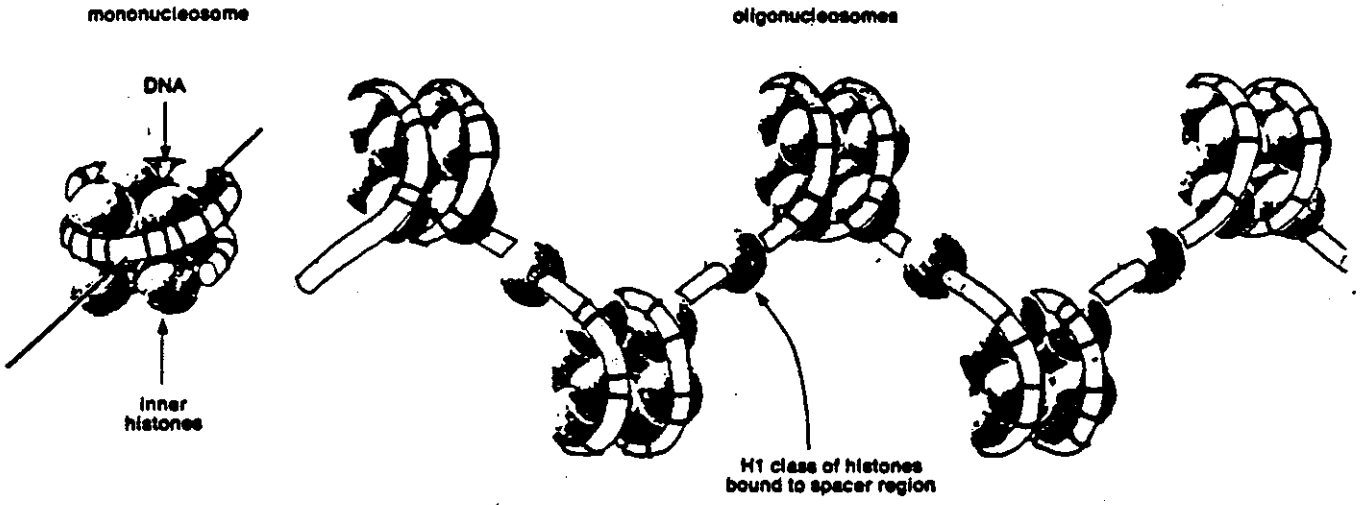



**Figure 3.** The current model of the core nucleosome particle proposes a disk-shaped structure, the inner histones close-packed in the center and DNA wrapped around the outside, with a dyad axis (*black line*). The zig-zag organization of the model for the oligonucleosomes is often seen in electron micrographs and probably represents a vestige of the higher-order organization.

Fig. 2.10 From Olins and Olins, Ref. 11.


1. Lehninger, Chapter 12.
2. Stryer, Part IV.
3. J. D. Watson, *Molecular Biology of the Gene*, Benjamin.
4. M. Eigen and L. de Maeyer, Naturwiss 53, 50 (1966).
5. The double helix has inspired literature: J. H. Watson, *The Double Helix*. R. Olby, *The Path to the Double Helix*. H. Judson, *The Eight Day of Creation* (Simon and Schuster, 1979). F. Crick, *How to Live with a Golden Helix*. The Sciences, Sept. 1979.
6. J. D. Watson and F. H. C. Crick, Nature 171, 964-967 (1953).
7. A. H-J. Wang, G. J. Quigley, F. J. Kolpak, J. L. Crawford, J. H. van Boom, G. van der Marel, and A. Rich, Nature 282, 680 (1979).
8. R. E. Dickerson, H. R. Drew, B. N. Conner, R. M. Wing, A. V. Fratini, and M. L. Kopka, *The Anatomy of A-, B-, and Z-DNA*, Science 216, 475-485 (1982).
9. J. J. Hopfield, Proc. Natl. Acad. Sci. USA 71, 4135-4139 (1974).
10. M. Gueron, *Enhanced Selectivity of Enzymes by Kinetic Proofreading*, Amer. Scientist 66, 202-208 (1978).
11. D. E. Olins and A. L. Olins, *Nucleosomes, The Structural Quantum in Chromosomes*, American Scientist, November-December 66, 704-711 (1978).
12. M. Botchan and J. D. Watson, eds. *Chromatin*, Cold Spring Harbor Symposium on Quantitative Biology, XLII, 1978.
13. C. A. Nicolini, *Chromatin Structure and Function*, 2 Vols., Plenum, 1979.

## 2.8. Transcription and Translation

The information contained in DNA is not used directly to control biosynthesis; the flow of information occurs as follows:

$$\text{DNA} \xrightarrow[\text{transcription}]{} \text{RNA} \xrightarrow[\text{translation}]{} \text{protein.}$$

In transcription, a mRNA (messenger RNA) acts as a template and takes the information required for the synthesis of a particular protein from the master, the DNA. In translation, the information on the mRNA is used to synthesize the protein; the action occurs at the ribosome, where the rRNA (ribosomal RNA) and the tRNA (transfer RNA) play the major roles.

Transcription occurs at the double-stranded DNA, but only one strand is used. The DNA thus unwinds partially so that the information contained on a certain length is transcribed onto the particular mRNA that corresponds to the protein to be produced.

## 2.9. The Code

The information on the DNA is coded in terms of the four bases A, T, C, and G. A given amino acid must be described by more than one base (letter), otherwise only four amino acids could be targeted. Two-letter words, such as AT or AG, would suffice for at most 16 amino acids. Since at least 20 must be specified, and start and stop signals are also needed, the code must contain at least three-letter words. Indeed, a series of brilliant experiment in the early 1960's established that the genetic code uses three-letter words, with the dictionary given in Table 2.2.

Table 2.2

The Genetic Code

| *First position (5' end)* | *Second position* | | | | *Third position (3' end)* |
|---|---|---|---|---|---|
| | U | C | A | G | |
| U | Phe | Ser | Tyr | Cys | U |
| | Phe | Ser | Tyr | Cys | C |
| | Leu | Ser | Stop | Stop | A |
| | Leu | Ser | Stop | Trp | G |
| C | Leu | Pro | His | Arg | U |
| | Leu | Pro | His | Arg | C |
| | Leu | Pro | Gln | Arg | A |
| | Leu | Pro | Gln | Arg | G |
| A | Ile | Thr | Asn | Ser | U |
| | Ile | Thr | Asn | Ser | C |
| | Ile | Thr | Lys | Arg | A |
| | Met | Thr | Lys | Arg | G |
| G | Val | Ala | Asp | Gly | U |
| | Val | Ala | Asp | Gly | C |
| | Val | Ala | Glu | Gly | A |
| | Val | Ala | Glu | Gly | G |

Table 2.2 shows that the code is degenerate; most amino acids correspond to more than one triplet. Degenerate triplets are called synonyms. Most synonyms differ only in the last base of the triplet. In particular XYC and XYU always, XYG and XYA usually code for the same amino acid. Degeneracy minimizes the effect of deleterious mutations: If the code were not degenerate, 20 codons would specify amino acids, and 44 would signify chain terminations (stop). Most mutations would lead to chain terminations and thus to inactive proteins.

UAA, UAG, and UGA designate chain termination. AUG or GUG is part of the initiation signal and the conformation around the corresponding mRNA determines whether the signal is to be read as a codon for an amino acid or as a start signal.

## 2.10. Codon and Anticodon; tRNA

A triplet on mRNA specifying a particular amino acid is called a codon. The codon is recognized by a corresponding anticodon on tRNA. The tRNA is the adaptor molecule: at least one tRNA exists for each amino acid. Each contains an amino acid attachment site and a template recognition site, the anticodon. A tRNA carries a specific amino acid to the site of protein synthesis.

A schematic of protein synthesis is shown in Fig. 2.11. DNA can either replicate, or direct protein synthesis via mRNA. The synthesis occurs at the ribosome, a nucleoprotein consisting of nucleic acids and proteins. The tRNA's pick up the required amino acids and transfer them to the ribosome. There, the anticodons on the tRNA match the codons on the mRNA to determine the order on the polypeptide chain. In all operations, proteins are involved. Nucleic acids and proteins thus are fully interwoven.

In 1968, it was discovered that tRNA molecules could be crystallized. It took another three years, however, before crystals of sufficient quality were produced for X-ray work. By now, the main features of tRNA are clear.[1,2]

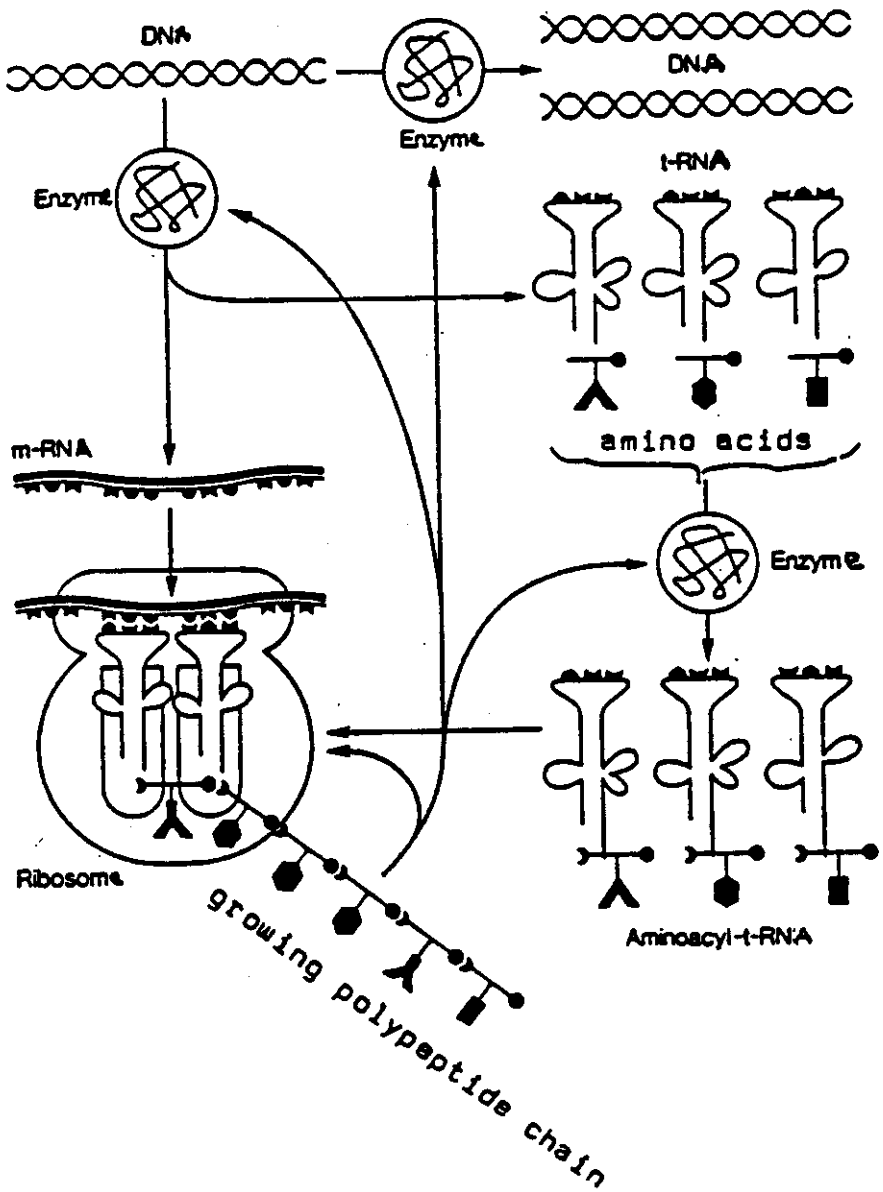
The schematic arrangement of a tRNA molecule is shown in Fig. 2.12. Similar to studies on proteins, heavy-atom derivatives are needed for the phase determination.[3] With refinement, the structure has been determined to a resolution of 2.5 Å. The structure is shown in Fig. 2.13.

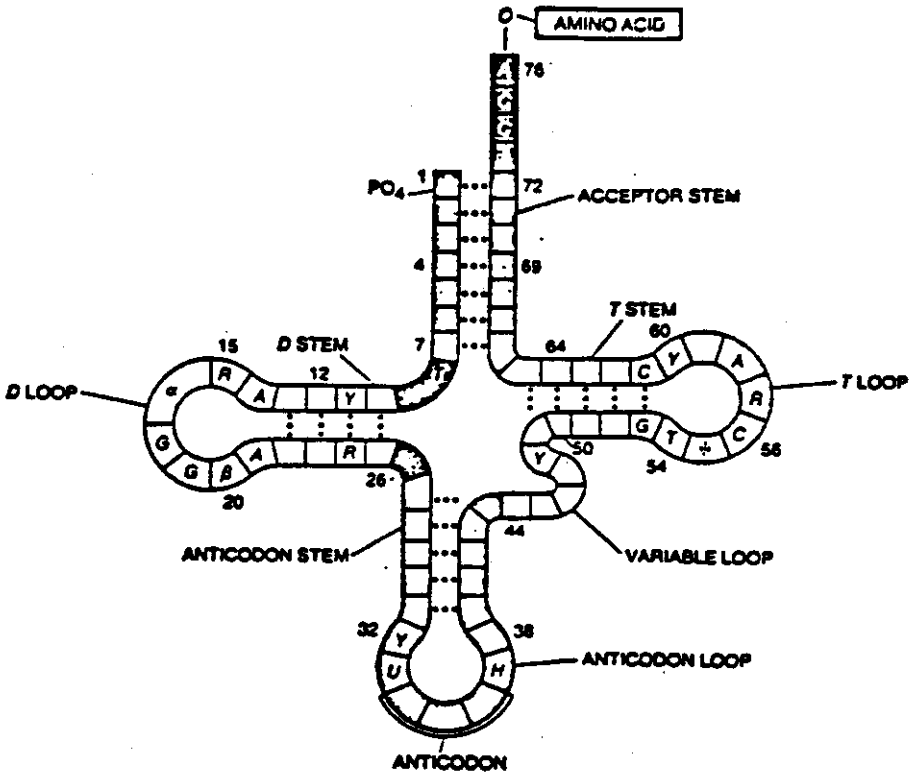
The dominant features of the tRNA structure are:

1. The molecule is L-shaped; the lengths of the two arms are about 7.3 and 7.0 nm. The thickness is about 2.2 nm.

1. J. L. Sussman and S.-H. Kim, Science 192, 853-858 (1976).
2. G. J. Quigley and A. Rich, Science 194, 796-806 (1976).
3. S. H. Kim, G. J. Quigley, F. L. Suddath, A. McPherson, D. Sneden, J. J. Kim, J. Weinzierl., P. Blattman, and A. Rich, PNAS 69, 3746 (1972).

Fig. 2.11 DNA replication and direction of protein synthesis via mRNA (simplified).






**CLOVERLEAF DIAGRAM** is the two-dimensional folding pattern of the transfer-RNA (tRNA) molecule, which was first deduced in 1965 from the sequence of nucleotide building blocks in yeast alanine tRNA. Since then the diagram has been found to fit the nucleotide sequences of about 100 tRNA's isolated from plant, animal and bacterial cells. Nucleotide bases found in the same positions in all tRNA sequences are indicated. The ladderlike stems are made up of complementary bases in different parts of the polynucleotide chain that pair up and form hydrogen bonds, causing the chain to fold back on itself. The number of nucleotides in the various stems and loops is generally constant except for two parts of the *D* loop designated α and β (which consist of from one to three nucleotides in different tRNA's) and the variable loop (which usually has four or five nucleotides but may have as many as 21). Abbreviations are *A* (adenosine), *G* (guanosine), *C* (cytidine), *U* (uridine), *R* (adenosine or guanosine), *Y* (cytidine or uridine), *T* (ribothymidine), ψ (pseudouridine), *H* (modified adenosine or guanosine).

Fig. 2.12

2. At one end of the L is the anticodon, at the other the 3' acceptor end. One end thus contacts the amino acid, the other the codon of the mRNA. The distance between the two points is rather long, about 7.5 nm.

3. Each of the arms of the L forms a double helix.

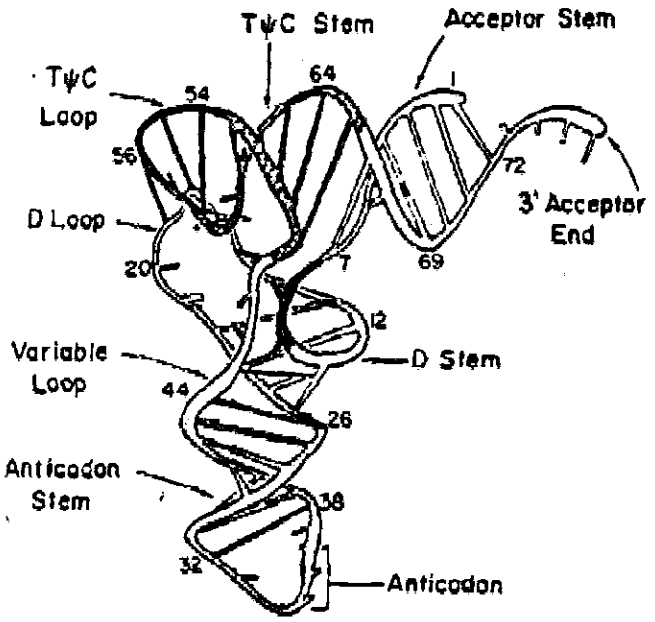
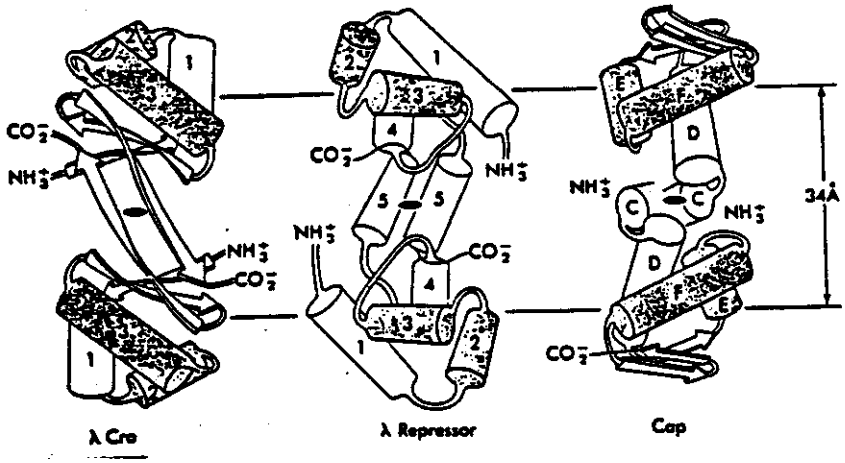
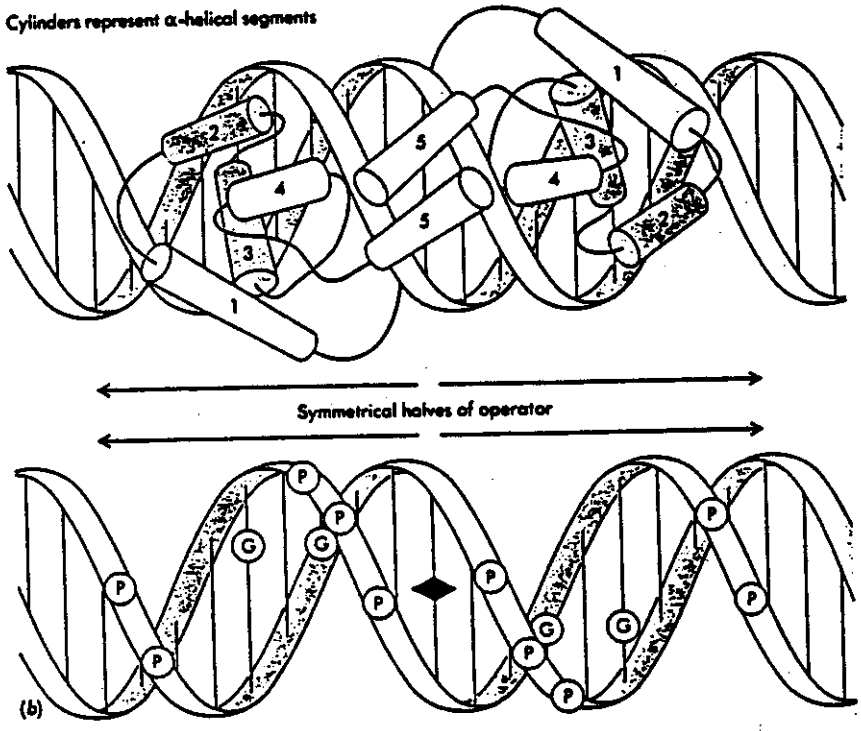



Fig. 2.13 Structure of yeast tRNA.[2] Shown is the ribose phosphate backbone. Hydrogen bonds between bases are shown as rungs.

Recently, there has been much interest in complexes of regulatory proteins with DNA. The structures of several of these proteins have been determined by X-ray diffraction, and a common pattern has emerged as illustrated in Fig. 2.14 for λ Cro, λ repressor and Cap (catabolite gene activator protein of E. coli). We note that all these proteins form dimers with a similar arrangement of two helical segments that are shaded in Fig. 2.14 The longer one of these fits into the major groove of B DNA while the shorter one contacts the phosphate groups. The two subunits of the protein are related by a 180° rotation about an axis perpendicular to the axis of the DNA, and the helical segments mentioned before are positioned to fit into adjacent grooves of the double helix. It is believed that other DNA-binding proteins "use" similar structural principles to recognize and bind to specific base pair sequences.




(c) The DNA binding faces of three DNA binding regulatory proteins. In each case, the number 3 α helices (named "F" in CAP protein) in the dimer are separated by 34 Å, the length of one turn of B-form DNA. (Courtesy of T. A. Steitz and I. T. Weber.)




(b) How two λ repressor monomers fit onto the operator. The diagram on top shows the portion of each polypeptide chain at the operator binding surface. There are two polypeptides, placed symmetrically on th halves of the operator. Cylinders ar helices, and they are connected by less structured segments of polypeptide. The amino terminus of the polypeptide is behind the DNA and connects to α helix 1. The bottom diagram shows which bases and backbone phosphates contact the repressor, as determined by chemical protection experiments. [After C. Pabo and R. Sauer, *Ann. Rev. Biochem.* 53 (1984):300, with permission.]

Fig. 2.14 Regulatory protein interacting with DNA