10 The Biological Function of Nucleic Acids

HERMANN HEUMANN and WOLFRAM ZILLIG

10.1 Introduction

10.1.1 General Remarks

The function and organization of cells and organisms are determined by the type and quantity of protein contained therein. Proteins catalyze as enzymes a large number of metabolic processes. In addition to lipids and polysaccharides, proteins are components of the cell structure.

The properties of a protein are determined by its structure. One can differentiate (see Chap. 9) between primary, secondary, tertiary, and quaternary structure. The primary structure is determined by the sequence of amino acid residues in a polypeptide chain. The secondary and tertiary structures describe the three-dimensional arrangement of the polypeptide chain. The arrangement of several different or equal polypeptide chains (subunits) in a complex is called the quaternary structure.

The secondary and tertiary structures of a polypeptide chain are determined by its primary structure. There are many examples of spontaneous renaturation of the native structure of a protein from a completely denatured and dissociated state, e.g., the pancreatic ribonuclease, the collagen, the DNA-dependent RNA polymerase and even ribosomes. Therefore, there is no reason to postulate that a particular folding mechanism controls the formation of the native structure.

An organism is not a rigid structure. Within genetically predetermined limits, the organism is capable of adapting itself to environmental changes. Such processes of adaptation, called regulatory or control processes, are reversible. The result of such a process is a change of type, quantity or activity of the protein contained in the cells. In addition, there are irreversible changes in the expression of genetic information. An example of such a change of the genetic program of a cell is the differentiation of tissue within an organism during ontogenesis. Another example is the switch of the metabolic program of the cell after infection by a virus to the new program of virus amplification. The alteration of a differentiated tissue cell into a tumor cell is a special case, where the normal control mechanisms for gene expression are defective.

Nucleic acids are present in the cell in two different forms: As deoxyribonucleic acid (DNA) and ribonucleic acid (RNA). DNA is the exclusive carrier of the genetic information. The expression of this information leads to the synthesis of the corresponding proteins. RNA can be the messenger that contains the information for the last step of gene expression, the protein biosynthesis (translation). In addition, RNA is a structural component of the apparatus for the protein synthesis, and at least adapter and vehicle for the translation.

Nucleic acids determine the functions of the organism, which are realized by proteins. Their own function is expressed by the transfer of the genetic information to the next generation (DNA replication), and by the gene expression.

10.1.2 Occurrence and Structure of Nucleic Acids

Two kinds of nucleic acids are present in the cell. DNA is contained in the nucleus, mitochondria and plastides of eukaryotic cells, and in prokaryotic cells, as carrier of the genetic information of the organism. In the case of eukaryotes RNA is synthesized in the nucleus, but its function is expressed in the cytoplasm.

The structure of nucleic acids was discussed in detail in Chap. 2. We should recall here only that nucleic acids are polynucleotide chains. In general, the DNA consists of four deoxyribonucleotides, adenyl acid (pA), thymidylic acid (pT), guanylic acid (pG), and cytidylic acid (pC). The sugar component is deoxyribose, whereas in RNA the sugar component of the corresponding ribonucleotide is ribose. In RNA, thymidylic acid is substituted by uridylic acid. The most important difference between DNA and RNA is the spatial arrangement of the polynucleotide chains. DNA is a double helix consisting of two complementary polynucleotide chains of opposite polarity (direction of the 3'→5'-phosphodiester binding). (Exceptions are certain viruses as bacteriophages φ x 174 and M 13, which contain single-stranded DNA.) RNA, in contrast, is single stranded. (An exception is rheovirus, which contains double-stranded RNA.) Partially complementary sequences within the RNA strand may be base paired. These regions are called hairpins or stems. In DNA and RNA respectively the base adenin is paired with the base thymin (uracil in RNA) via two hydrogen bondings, and guanine with cytosin via three hydrogen bondings.

The free energy, about 8 kcal/mol, is a little lower for the AT pair than for the GC pair. The stability and corresponding temperature (Tm) at which the strands are separated increase in proportion to the GC content.

The RNA can be divided, according to its function and structure, into three different types: messenger RNA (mRNA), transfer RNA (tRNA), and ribosomal RNA (rRNA). mRNA functions as carrier of the
information for the protein biosynthesis. It is a long polynucleotide strand consisting of up to some thousand nucleotide residues. It is often polystreric, i.e., it contains the genetic information for several peptide chains. According to the variety of the proteins there are many different species of mRNA in an organism (in prokaryotes some thousand), which are present in different numbers depending on the gene dosage and efficiency of their synthesis. In eukaryotes mRNA often has a modified 5'-terminus and almost always a poly-A sequence at the 3'-terminus (except histone mRNA).

tRNA functions as adapter for the amino acid residues in the translation process. It is a short chain consisting of about 80 nucleotide residues. It has a high level of complementarity within the chain and therefore much tertiory structure. In the active state RNA always has a pCpCpA sequence at the 3'-terminus and often pGp...1 at the 5'-terminus. In the cell about 40 different species are present, some of which are rare.

Ribosomal RNA's (rRNA) are components of the ribosomes, the apparatus of translation. There are four different types of rRNA's, which are named according to their sedimentation constants: two large rRNA's, 28S and 18S in eukaryotic cells, 23S and 16S in prokaryotic cells, and two small rRNA's (5.8S and 5S in eukaryotic cells), one of which only is present in eukaryotic cells (5S RNA).

The 5S RNA and 5.8S RNA consist of about 120 nucleotide residues. The small size rRNA's and the 28S rRNA, and 23S rRNA (in prokaryotes) are components of the large 60S (50S) ribosomal subunit. The 18S rRNA (16S rRNA) is a component of the small 40S (30S) ribosomal subunit.

### 10.2 DNA Replication

The genetic information, which determines the structure and function of all proteins, is contained in the DNA base sequence. This information has a high degree of stability, which guarantees constant properties of a species over generations. But small stepwise changes of the genetic information are possible by mutations. This can lead to new, better adapted species by the selection pressure of the environment. The stability of the genetic information is due to its exact replication and equal distribution of identical replication products during cell proliferation.

#### 10.2.1 Organization of DNA in the Cell

Organisms like bacteria and blue algae, which do not have a cell nucleus divided by a membrane from the cytoplasm, are called prokaryotes. In these organisms, the main part of the genetic information is contained in covalent rings of chromosomes. These chromosomes are compact packets which are fixed to the cell membranes in an unknown way. The chromosome of *E. coli* has a molecular weight of about $2.5 \times 10^{9}$, corresponding to $4 \times 10^{6}$ base pairs and about 5000 genes. In the cell there may also be present plasmids (episomes), which exist also as covalently closed-in rings DNA double strands. Such rings contain, besides the twist characteristic for the B-structure of double-helical DNA, additional twists of opposite direction, so-called negative overtwists. By at least one nick in one or both DNA strands, the DNA can rearrange and the overtwist disappears.

Eukaryotes are organisms containing a cell nucleus. Their DNA organization is much more complex. Only a small fraction, about 10% of the whole genom is expressed. In the case of mammals, the whole genom consists of about 108 genes. Every chromosome contains a single continuous DNA double strand. Basic proteins, the histones, determine by their interaction with the DNA the structure of the chromatin. The histone protein H1 apparently connects different regions of the DNA strand. Nucleosomes (v-bodies) consist of DNA complexed with histone proteins in a certain stoichiometric ratio ($2H_{2a}$, $2H_{2b}$, $2H_{3}$, $2H_{4}$). The nucleosomes are like pearls in a chain separated from each other by small segments of free DNA of about 40 nucleotide pairs. The main part of the DNA in the nucleosome-containing segments (about 200 nucleotide pairs pro v-bodies) are in contact with the nucleosome. A model of the nucleosomes is shown in Fig. 10.1. In addition to the histones in the chromatin there are so-called nonhistone proteins, which are in general neutral or acidic. Among these proteins there are also well-known contractile proteins like actin, actomyosin, and tubulin, which might play a role in the condensation of the chromosomes during mitosis. The active segments of the chromatin, which participate in the gene expression of the so-called euchromatin, has a relatively loose structure in contrast to the strongly condensed, inactive heterochromatin.

Certain segments of some chromosomes form substructures in the nucleus, the so-called nucleoli, where the precursor of ribosomal RNA is synthesized. In eukaryotic cells, plastids and mitochondria, which are genetically partly autonomous, contain DNA. The DNA of these organelles exists in covalently closed rings like the DNA in prokaryotic cells.

#### 10.2.2 Principles of DNA Replication

In prokaryotes the DNA is replicated during the whole life of the cell, in eukaryotes, however, only during a certain period of the cell cyclus, the S phase.

The DNA is replicated semi conservatively, that means each of both daughter double strands contains one strand of the parental double strand and one complementary stand, which is newly synthesized (Fig. 10.2).