Chapter 2

Introduction to some basic features of genetic information: From DNA to proteins

DAVID QUIST,[1] KAARE M. NIELSEN[1, 2] AND TERJE TRAAVIK[1, 3]
[1]THE NORWEGIAN INSTITUTE OF GENE ECOLOGY (GENØK), TROMSØ, NORWAY
[2]DEPARTMENT OF PHARMACY, UNIVERSITY OF TROMSØ, NORWAY
[3]DEPARTMENT OF MICROBIOLOGY AND VIROLOGY, UNIVERSITY OF TROMSØ, NORWAY

Molecular biology is the study of *biology* at a *molecular* level, with the aim of understanding the interactions between the various systems of a cell, including the interrelationship and regulation of *DNA*, *RNA* and *protein* synthesis. In general terms, DNA (deoxyribonucleic acid) is the basic genetic information macromolecule of the cell. It provides the rudimentary instructions for all kinds of biochemical functions, from making proteins to regulatory functions. DNA is found in every cell and every cell type and organism, from single-celled organisms (*prokaryotes*, e.g. bacteria), to larger multicellular organisms (*eukaryotes*, e.g. seaweeds, fungi, plant, animals) that can have many different cell and tissue types.[1] DNA contains the genetic 'code' of information that makes each species unique. Smaller variations in the DNA can lead to minor differences among individuals of the same species. The combination of specific DNA composition, epigenetic changes (see Chapter 8) and environmental influences determine an organism's appearance and development. In this book, we discuss how the main carrier of heritable information (DNA) and the environment interact, with particular emphasis on how genetic engineering may intentionally or unintentionally affect this interaction. This chapter focuses on DNA, RNA and the concept of genes. It is structured as follows:

1. **Structure and replication of DNA**
2. **Genes as specific nucleotide compositions within DNA**
3. **RNA molecules**
4. **Genes and protein synthesis**

1. Structure and replication of DNA

The primary feature that makes DNA unique lies within its chemical structure. The information-containing properties of the nucleic acids arise from unique combinations of individual nucleotides that form long polynucleotide chains; this macromolecule is collectively called DNA. Each nucleotide consists of three parts: a nitrogen base, a pentose sugar, and a phosphate group (see Figure 2.1). DNA consists of four different base nucleotides: adenine, thymine, guanine, and cytosine (A, T, G, and C, respectively).[2] The phosphate group of one nucleotide is attached to the sugar of the adjacent nucleotide that is next in line in the chain. This results in a 'backbone' structure of alternating phosphate groups and sugar groups, from which the nucleotide bases project outward. Yet, how can so much genetic diversity come from only four basic units (nucleotides) of genetic information? This is possible because the DNA is a long strand of information, like letters in a sentence. There is almost an infinite number of combinations of nucleotides possible in a DNA macromolecule. For instance, even a short DNA molecule 10 base pairs (bp) long has 4^{10} or 1,485,576 possible combinations of bases. A bacterial gene is often 1000 bp long.

[1]Viruses form their own class of life. They may have single-stranded or double-stranded DNA or RNA as their genetic material, using the replication machinery of the organisms they infect to multiply.
[2]Note that RNA, which we will discuss later, also has four nucleotides but replaces Thymine with a Uracil (U) base.

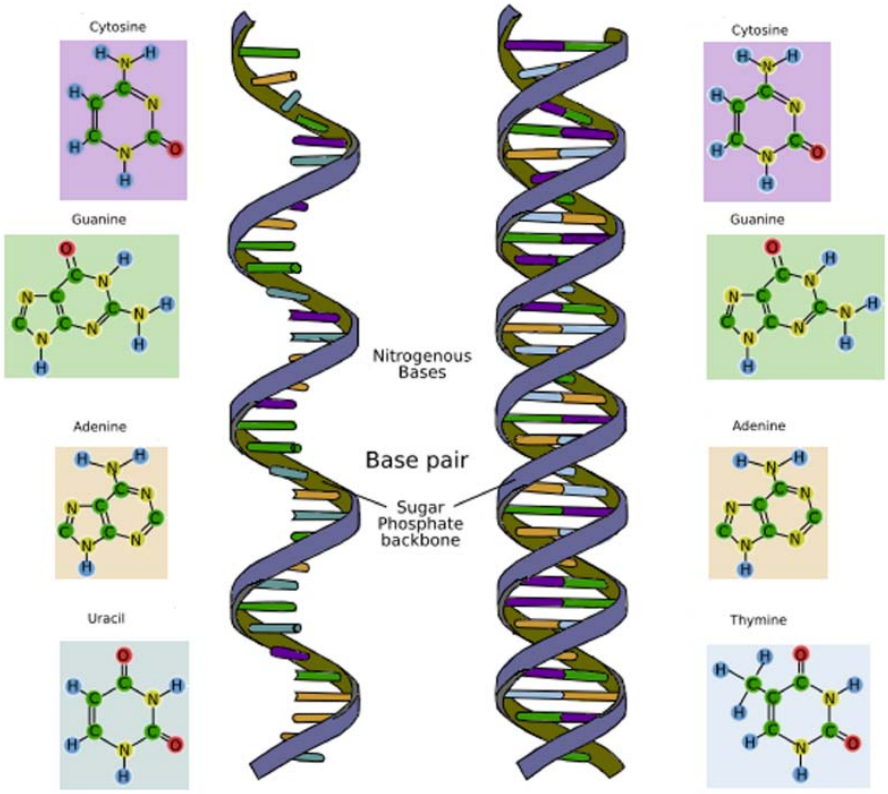

Figure 2.1. The chemical composition and structure of the DNA double helix.

DNA is a double-stranded molecule whose primary features are its *complementarity* and its *base pairing* with its sister DNA strand, forming *the double helix*. The complementarities of the nucleotide bases also facilitate replication, or copying of the genetic material. How does an organism pass this DNA to daughter cells and offspring? Inheritance, the passing of genetic information (genes) from one generation to the next, involves either i) sexual recombination (mixing of genetic information from parents via the combination of sperm and egg), or ii) through cell division that results in the inheritance of the same genetic information from the parent to the daughter cells. This is achieved by DNA replication (Figure 2.2). So each DNA strand is complementary to the other in their base pairing of nucleotides: T always pairs with A and G always pairs with C. These two complementary polynucleotide chains make a very stable spiralling structure, and form the DNA's well-described double helix.

DNA replication produces two molecules by semi-conservative replication, that is, each DNA molecule is made up of one of the original two parental strands (that make up the double helix) and one completely new synthesized strand (Figure 2.2). During replication, the DNA is unwound by enzymes, called helicases, that open up the double helix, allowing DNA replication enzymes, called DNA polymerases, to come in and synthesize a new strand of DNA. The polymerase is like a DNA copier, requiring the template (original), DNA, and the individual A, T, C, and G nucleotide units paired to its complementary base (A to T, and G to C), all one nucleotide at a time.[3] This process is thus almost identical[4] to a polymerase chain reaction (PCR) that will be described further in Chapter 33.

[3]Note that this is essentially the same biological machinery used in the laboratory to produce a Polymerase Chain Reaction (PCR), a laboratory technique that has many applications in genomic research, and is widely used as a means to detect the presence of genetically modified DNA (as described in later chapters).

[4]In PCR amplifications of DNA, a termostable polymerase is used, that allows the reaction to be repeated after heat-mediated separation of the two DNA strands.

2. Genes as specific nucleotide compositions within DNA

A gene is classically understood as a short region of DNA that encodes, for example, for the production of a particular protein product or trait. Genes are commonly described as a physical unit.[5] In such a physical conception, genes, in essence, are functional units of inheritance of DNA.

The sum of an organism's genetic information is what is generally referred to as its *genome*. Understanding the function of genes and other parts of the genome is known as *functional genomics*. The genome of an organism consists of very long strands of DNA molecules, usually packaged with specific proteins into *chromosomes.*

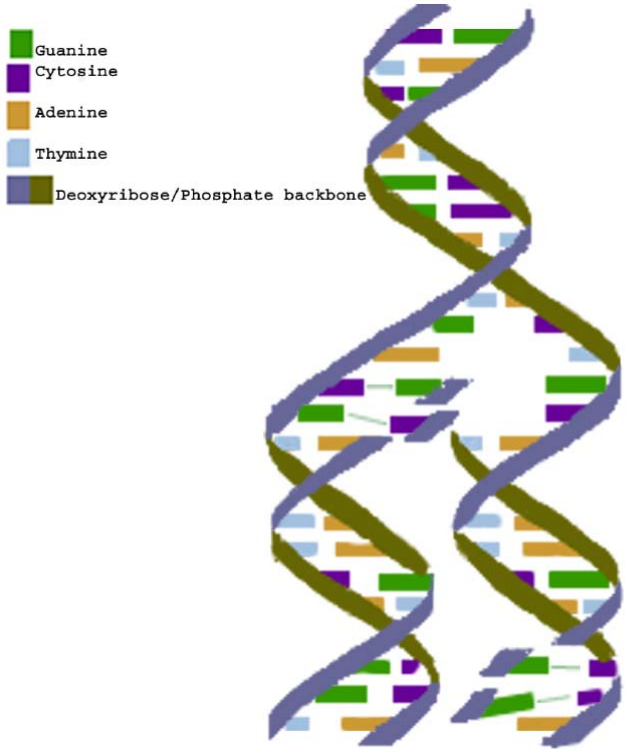

Figure 2.2. DNA replication is semi-conservative, with one of each parental strand serving as template for each newly synthesized complement.

Different organisms have different sized genomes (see Figure 2.3), though the size of an organism's genome does not necessarily correlate with its complexity. It has been demonstrated that only a very small percentage of the DNA in the whole genome actually encodes for a protein (only *c.*5% in humans, for example). Thus, the remaining DNA may have important genome stability, and developmental and regulatory functions. The large regions of DNA not encoding proteins were earlier termed 'junk DNA'.[6]
The DNA is tightly wound around a series of proteins (e.g. histones) that have both DNA packaging and regulatory functions.[7] These protein complexes are further wound to produce

[5]While this can be true in a most reduced sense, genes and genomes are really much more than that, as they participate in interactive layered biological networks of metabolic regulation with the cell, tissue and organism. The concept of a gene, and the genome itself, is therefore not as straightforward as it may seem at first.
[6]In later sections, we will see that this 'junk DNA' is now known to have important regulatory functions.
[7]The nucleus of a single diploid human cell contains approximately 6 x 109 bp of DNA. This enormous degree of packaging is achieved by wrapping up the DNA with proteins called histones. In vertebrates, there are five

chromosomes (in eukaryotes). Chromosomes are amazingly long (stretched out, the DNA of just one human cell would be almost two metres long) and hence need to be compacted within the cell. In the case of humans, we have 23 chromosomes, with two copies per cell (one from each of the sexes). In eukaryotes, the majority of the genetic information is compartmentalized in the cell's nucleus (mitochondria, and chloroplasts in plants, also contain functional DNA from their former lives as free living organisms). In prokaryotes, genetic information is more loosely compacted in a single circular chromosome within the organism.

3. RNA molecules

RNA molecules, like DNA, are made up of nucleotides, except that the thymine (T) nucleotide is replaced with a uracil (U) nucleotide, which is not found in DNA. Due to this small but important difference, a double helix structure does not form easily, but instead, RNA remains single stranded (ss).[8] SsRNA serves various functions in the cell, such as messenger RNA (mRNA) and transfer RNA (tRNA), two types of RNA that are required for protein synthesis. Other RNAs serve regulatory functions. The role of RNA within the cell is explained in greater detail in Chapter 3.

histones, H1, H2A, H2B, H3, and H4. The basic packaging unit, or nucleosome, is an octamer composed of two molecules of each of the histones H2A, H2B, H3, and H4, forming a disc-shaped structure. Exactly 146 bp of DNA are wound around the disc, like a thread on a spool, making slightly less than two complete turns. The gap between neighbouring nucleosides is approximately 50 bp in length, and one molecule of histone HI binds in this linker region. In transcriptionally inactive chromatin there is a further order of packaging to form a structure known as the solenoid, comprising nucleosomes wrapped around a multimeric rod of H1 subunits. The solenoid is 30 nm in diameter and each turn contains six nucleosomes and six H1 molecules.

[8]Double-stranded (ds) RNAs do, however, make up the genomes of some virus families (e.g. Reoviridae), and are also important in the regulation of gene expression (see Chapter 3).

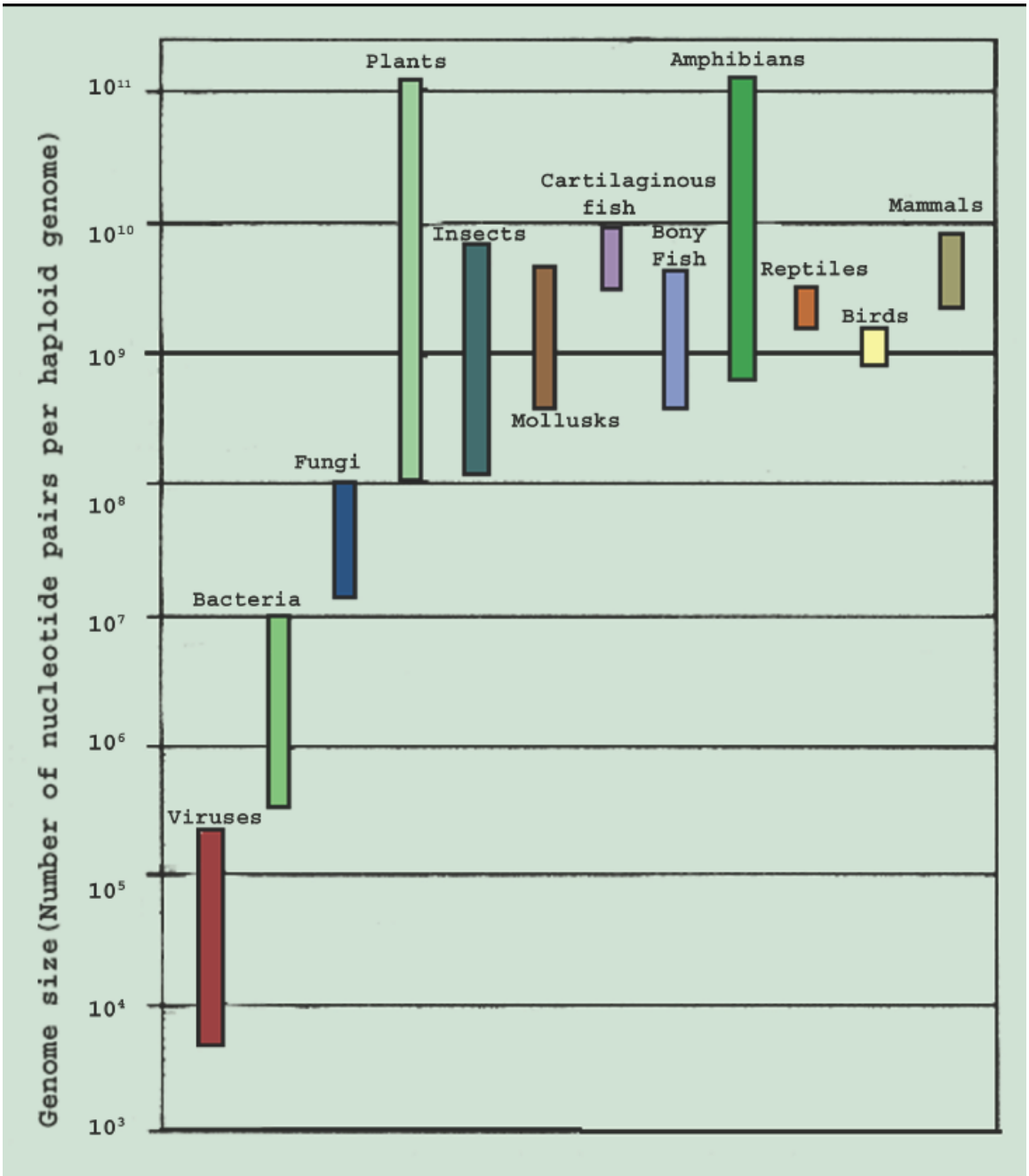

Figure 2.3. Genome size variations in different categories of organisms.

4. Genes and protein synthesis

One theory in molecular biology has, for the last half century, been the guiding principle for understanding how genetic information is processed in the cell. This theory, called The Central Dogma of Molecular Biology, states that genetic information that instructs protein synthesis flows in one direction, from DNA (via transcription) to RNA to protein (via translation) (see Figure 2.4). This dogma is the guiding principle of genetic engineering, suggesting that genes are independent modules that can function equally well in different organisms where the gene is in command regardless of its cellular (biological) and environmental context.

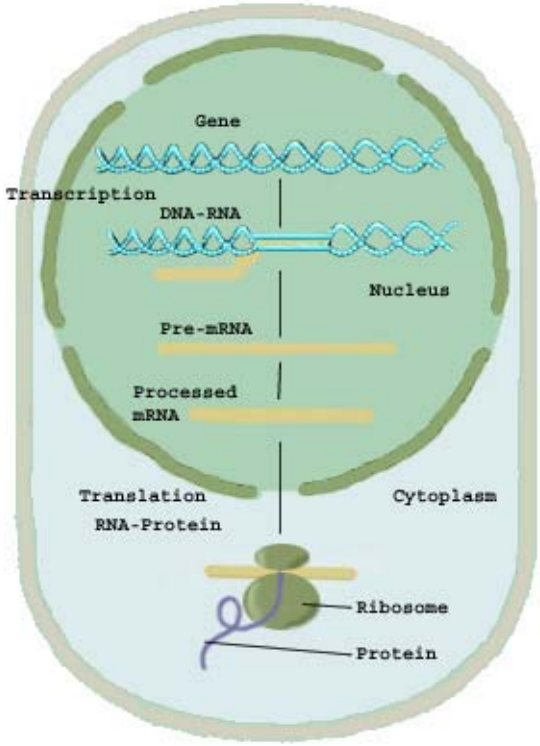

Figure 2.4. Simplified illustration of the Central Dogma assumed information flow from gene to protein in a prokaryotic cell.

The next chapter (Chapter 3) examines in more detail how genetic information contributes to the synthesis of a particular gene (protein) product and discusses how DNA is only part of a two-way regulatory network influenced by both abiotic and biotic factors at complex levels of organization within an organism.

Further Reading

Avise, J. (2001) Evolving genomic metaphors: A new look at the language of DNA. Science 294: 86-87.

Brown T.A. (2002) Genomes. Second Edition, Chapter 1. BIOS Scientific Publishers Ltd, (ISBN 1-85996-029-4)

Dennis, C. *et al*. (2003) The double helix – 50 years. Nature 421: 395-453.

Karp, G. (2005) Cell and Molecular Biology, Concepts and Experiments. Fourth Edition, Chapter 10. John Wiley and Sons Inc., (ISBN 0-471-65665-8).

Pääbo, S. (2003) The mosaic that is our genome. Nature 421: 409-411.