http://dx.doi.org/10.18778/8142-184-3.07

Tomasz Śliwiński

3. GENETICS

One of the discoverers of DNA's double helical structure, Francis Crick, said that "almost all aspects of life are engineered at the molecular level, and without understanding molecules, we can only have a very sketchy understanding of life itself." Others scientists, including Mendel, discovered the existence of units of heredity that were later called genes, and associated with the movement of chromosomes during mitosis and meiosis. Without knowledge of the molecules that carry genetic information, it would be impossible to understand anything about the biochemical processes through which genes determine phenotypes and pass on instructions to subsequent generations.

3.1. The structure of DNA

At the beginning of the twentieth century, there was no information about what 'genetic material' is. After 50 years of great experimental efforts, DNA was discovered to be the prime genetic molecule, carrying all of the hereditary information within chromosomes. This structure of DNA was helpful for understanding how it carries the genetic messages that are replicated when chromosomes divide to produce two identical copies of themselves. The full chemical name of DNA is deoxyribonucleic acid, reflecting three characteristics of the substance: one of its constituents is a sugar known as deoxyribose; it is found mainly in cell nuclei, and it is acidic. DNA contains only four distinct chemical building blocks linked in a long chain (Fig. 1). The four individual chemicals belong to a class of compounds known as nucleotides; the bonds joining one nucleotide to another are covalent phosphodiester bonds, and the linked chain of building block subunits is a type of polymer.

Now, over 50 years after the discovery of the double helix, this simple description of genetic material remains true and has not had to be appreciably altered to accommodate new findings. Nevertheless, we have come to realize that the structure of DNA is not quite as uniform as was first thought. The chromosomes of some small viruses have single-stranded, not double-stranded, molecules. Some DNA sequences even permit the double helix to twist in the left-handed sense (Z-DNA structure), as opposed to the right-handed sense originally formulated for DNA's general structure. Some DNA molecules are linear, whereas others are circular. Additionally, the structure of DNA varies and

these DNA variations arise from the unique physical, chemical, and topological properties of the polynucleotide chain.

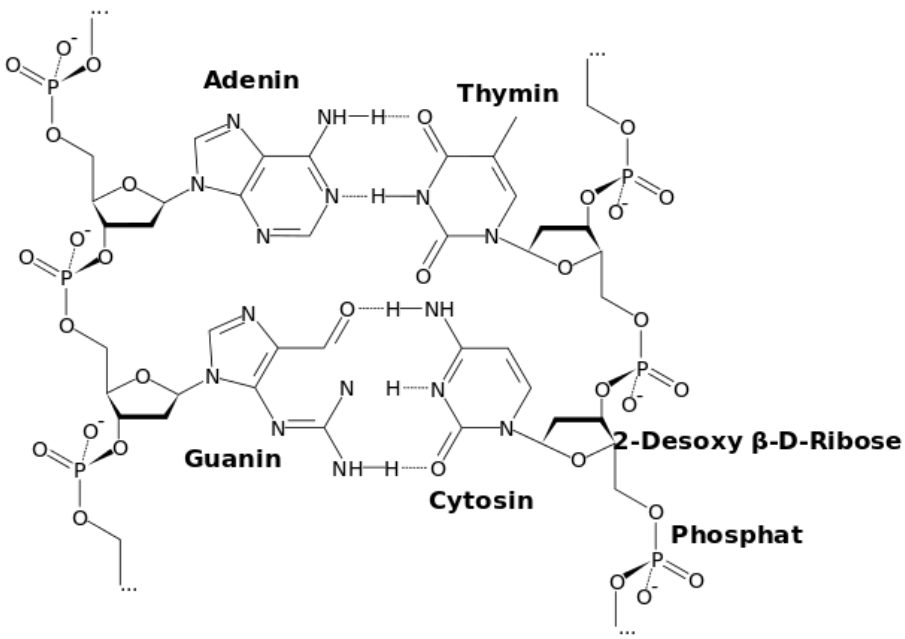

Figure 1. Chemical structure of DNA (*figure used with permission under Creative Commons license*)

DNA is composed of two polynucleotide chains twisted around each other in the form of a double helix. Figure 2 presents the structure of the double helix in a schematic form. The backbone of each strand of the helix is composed of alternating sugar and phosphate residues; the bases project inward but are accessible through the major and minor grooves (Fig. 3). The nucleotide, the fundamental building block of DNA, consists of a phosphate joined to a sugar, known as 2'-deoxyribose, to which a base is attached. The phosphate and the sugar have structures as shown in Figure 4. The sugar is called 2'-deoxyribose because there is no hydroxyl at position 2' (two hydrogens). The sugar and base alone are called a nucleoside and addition of a phosphate to a nucleoside creates a nucleotide. Making a glycosidic bond between the base and the sugar, and also making a phosphoester bond between the sugar and the phosphoric acid, creates a nucleotide. Moreover, nucleotides are joined to each other in polynucleotide chains through the 3'-hydroxyl of 2,-deoxyribose of one nucleotide and the phosphate attached to the 5,-hydroxyl of another nucleotide.

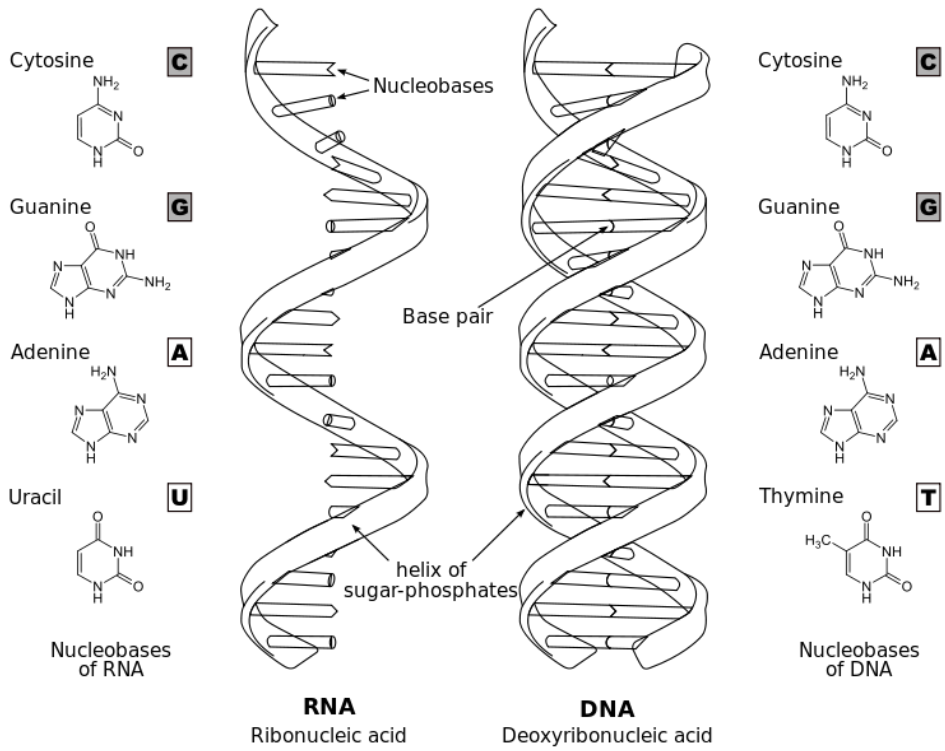

Figure 2. Structural differences between DNA and RNA (*figure used with permission under Creative Commons license*)

The double helix consists of two polynucleotide chains that are aligned in opposite orientation. The two chains have the same helical geometry but have opposite 5, to 3, orientations. That is, the 5' to 3' orientation of one chain is antiparallel to the 5' to 3' orientation of the other strand, as shown in Figures 2 and 3. The two chains interact with each other by pairing between the bases, with adenine (A) on one chain pairing with thymine (T) on the other chain and, likewise, guanine (G) pairing with cytosine (C) in anti-parallel manner. Despite large variations in the relative amounts of the bases, the ratio of A to T is not significantly different from 1:1, and the ratio of G to C is the same in every organism.

The hydrogen bonds between complementary bases are a fundamental feature of the double helix, contributing to the thermodynamic stability of the helix and the specificity of base pairing. The double helix is stabilized by hydrogen bonds and by stacking interactions between the bases. The stacked bases are attracted to each other by transient, van der Waals interactions.

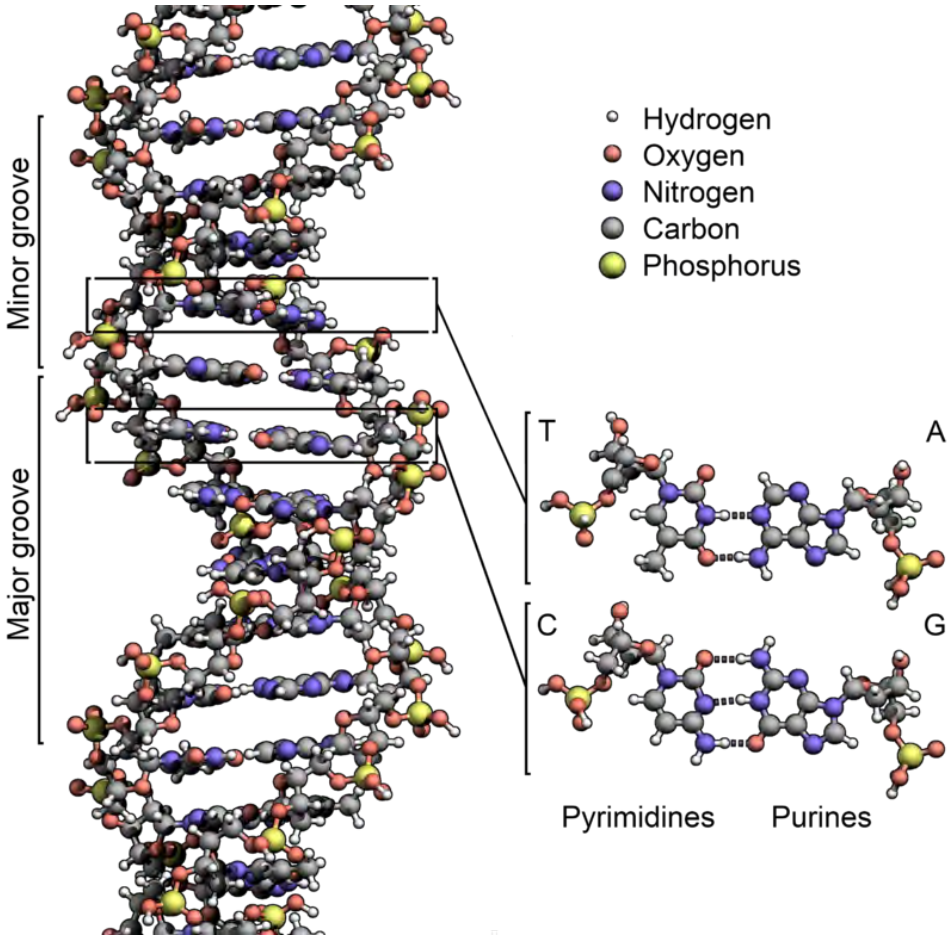

Figure 3. DNA structure (*figure used with permission under Creative Commons license*)

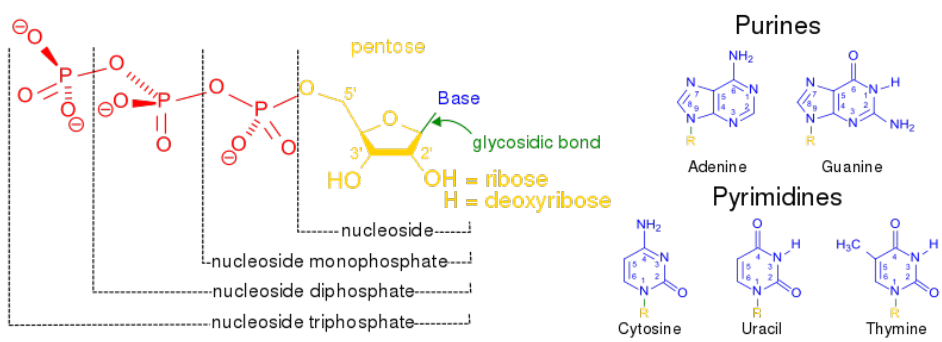

Figure 4. Structure of nucleotides (*figure used with permission under Creative Commons license*)

Each of the polynucleotide chains in the double helix is right-handed. As a result of the double-helical structure of the two chains, the DNA molecule is a long, extended polymer with two grooves that are not equal in size to each other. These different grooves are a simple consequence of the geometry of the base pair (the angle between the glycosidic bonds different for the narrow and wide angle). The two strands of the double helix are held together by relatively non-covalent (weak) forces, and can be easily separated, when a solution of DNA is heated above physiological temperatures (~ 100°C), or under conditions of high pH, and this process is known as denaturation, a reversible process. Additionally, when heated solutions of denatured DNA are slowly cooled, single strands often re-form regular double helices. This feature can be used in forming artificial hybrid DNA molecules by slowly cooling mixtures of denatured DNA from two different sources. It can also be made between complementary strands of DNA and RNA. This feature, called hybridization, is the basis for several indispensable techniques in molecular biology, such as Southern blot hybridization and DNA microarray analysis. If we plot the optical density of DNA as a function of temperature, we observe that the increase in absorption occurs abruptly over a relatively narrow temperature range. The midpoint of this transition is the melting point, or Tm (Fig. 5).

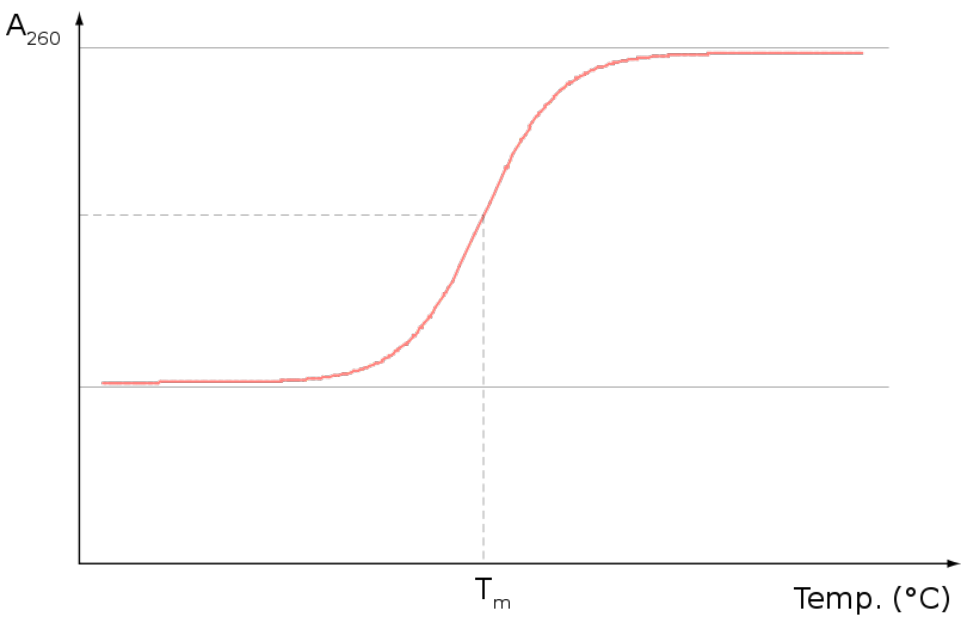

Figure 5. Hyperchromicity of DNA (*figure used with permission under Creative Commons license*)

DNA melts because of the transition from a highly ordered double-helical structure to a much less ordered structure of individual strands. Renaturation occurs by a slow nucleation process in which a relatively small stretch of bases on one strand finds and pairs with their complement on the complementary strand. Tm of DNA is largely determined by the G:C content of the DNA and the ionic strength of the solution. A larger amount of the percent of G:C base pairs in the DNA than A:T pairs (three hydrogen bonds versus two) gives the higher melting point. Likewise, the higher the salt concentration of the solution, the greater is the temperature at which the DNA denatures. Additionally, it happens because the stacking interactions of G:C base pairs with adjacent base pairs are more favourable than the corresponding interactions of A:T base pairs with their neighbouring base pairs. Moreover, another fundamental feature of DNA is the effect of ionic strength when the backbones of DNA strands contain negative charged phosphoryl groups, which are close enough across the two strands that, if not shielded, they tend to cause the strands to repel each other, facilitating their separation. At high ionic strength, the negative charges, shielded by cations, stabilize the helix, and at low ionic strength the unshielded negative charges make the helix less stable.

The majority of naturally-occurring DNA molecules have the Watson and Crick configuration, known as B-form DNA, right-handed. DNA structure also exists in left-handed form, known as a Z form, in which the helix spirals to the left and the backbone takes on a zigzag shape.

The information content of DNA resides in the sequence of its four bases which are like the letters of an alphabet and different sequences of them spell out different 'words', which possess their own phenotype effect, for example, GGTCCA and ATTGCCA mean different things. In all cellular forms of life as Prokaryotes and Eukaryotes and many viruses, DNA carries the genetic information. By contrast, some viruses, including retroviruses such as HIV, use RNA as their genetic material.

3.2. The structure and function of RNA

RNA is principally found as a single-stranded molecule. An intrastrand base pairing, RNA exhibits extensive double-helical character and is capable of folding into a wealth of diverse tertiary structures. Most remarkable of all, some RNA molecules are enzymes, one of which performs a reaction that is at the core of information transfer from nucleic acid to protein and hence is of profound evolutionary significance.

There are three major chemical differences between RNA and DNA: First, RNA possesses sugar ribose instead of deoxyribose found in DNA, (Fig. 2), second, RNA contains the base uracil (U) instead of the base thymine (T); U,

like T, base pairs with A (Fig. 6) and third, RNA molecules mostly are single-stranded and they are much shorter than the very long DNA molecules found in nuclear chromosomes. Within a single-stranded RNA molecule, can form a short double-stranded, base paired stretch. RNA, in comparison to double-helical shape of a DNA molecule has a complicated structure of short double-stranded segments interspersed with single-stranded loops. RNA has the same ability as DNA to carry information in the sequence such as in some viruses, but is much less stable than DNA. Additionally, RNA fulfils several vital functions in all cells, such as gene expression and protein synthesis and plays a significant role in DNA replication.

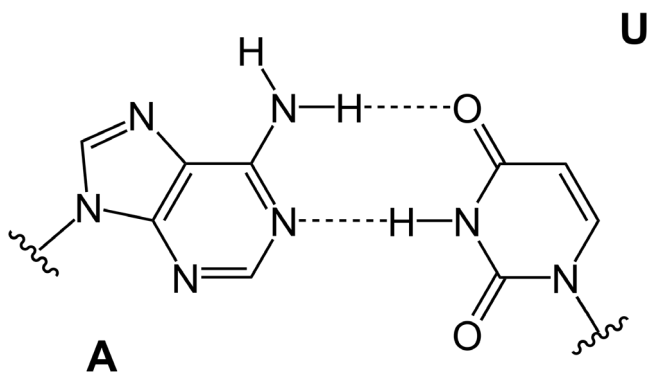

Figure 6. Base pair A:U (*figure used with permission under Creative Commons license*)

RNA functions as the intermediate, the messenger RNA (mRNA), between the gene and the protein-synthesizing machinery. Another function of RNA is as an adaptor, the transfer RNA (tRNA), between the codons in the mRNA and amino acids. RNA can also play a structural role, as in the case of the RNA components of the ribosome (rRNA). Finally, some RNAs are enzymes that catalyse essential reactions in the cell. RNA is capable of forming long double helices, but these are very rare in nature. Despite being single-stranded, RNA molecules can form a double-helix form, the two stretches of complementary sequence are near each other, the RNA may form a stem-loop structure, internal loops (unpaired nucleotides on either side of the stem), bulges (an unpaired nucleotide on one side of the bulge), or junctions.

The central dogma of molecular biology states that DNA maintains the information to encode all of our proteins, and that three different types of RNA convert this code into polypeptides. But with the advent of new techniques, scientists have revealed that RNA does much more than simply play a role in

protein synthesis, such as catalytic reactions. Generally, this type of RNSs are named non-coding RNAs and play complex regulatory roles in cells.

In eukaryotes, non-coding RNA comes in several varieties, such as tRNA and rRNA, involved in the translation of mRNA to proteins, but there are also a lot of non-coding RNAs assist in many essential functions in eukaryotic cells and new ones are still being discovery. Various regulatory non-coding RNAs exert their effects through a combination of complementary base pairing, complexing with proteins, and their own enzymatic activities, such as small nuclear RNAs, microRNAs, small interfering RNAs, small nucleolar RNAs or riboswitches.

This complexity and diversity of RNA forms and activities in cells lends credence to the so-called 'RNA world' hypothesis, which states that RNA may have evolved prior to DNA and protein, and it may have played the roles of both of these molecules in the earliest life-forms. This hypothesis is supported by example of existing RNAs having both coding and catalytic capacity. Nevertheless, the full contribution of RNA to the life of the cell may still be unknown.

3.3. The gene expression process as a flow of information from DNA through RNA to protein

The Central Dogma maintains that genetic information flows in two distinct stages (Fig. 7). The information of genes written in the language of nucleic acids is subsequently transcribed to the same instructions written in the RNA dialect. The conversion of DNA to RNA is called transcription. In the next stage of gene expression, the cellular machinery translates mRNA into its polypeptide equivalent in the language of amino acids. This decoding of nucleotide information to a sequence of amino acids is known as translation and is done using ribosomes, which are composed of proteins and rRNAs. Decoding depends on the genetic code that defines each amino acid in terms of specific sequences of three nucleotides, as well as on tRNAs, small RNA adaptor molecules that place specific amino acids in the correct position in a growing polypeptide chain. The Central Dogma does not explain the behaviour of all genes, when a large subset of genes is transcribed into RNAs that are never translated into proteins, because they encode rRNAs and tRNAs.

In addition, scientists later found that certain viruses contain an enzyme that can reverse the DNA-to-RNA flow of information by copying RNA to DNA in a process called reverse transcription. The language of nucleic acids is written in four nucleotides – A, G, C, and T in the DNA dialect; A, G, C, and U in the RNA dialect – while the language of proteins is written in amino acids. Now we know the list of the 20 amino acids that are genetically encoded by DNA or RNA.

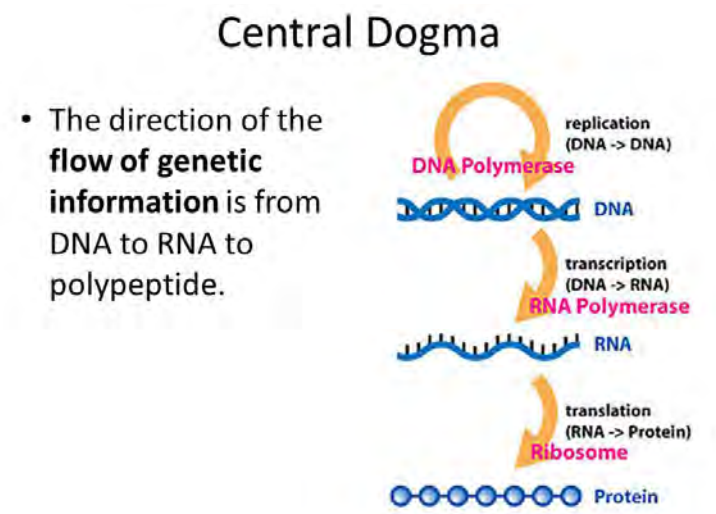

Figure 7. The central dogma of molecular biology (*figure used with permission under Creative Commons license*)

3.4. Transcription: From DNA to RNA

Transcription is the process by which the polymerization of ribonucleotides guided by complementary base pairing produces an RNA transcript of a gene and is catalysed by RNA polymerase. The template for the RNA transcript is one strand of that portion of the DNA double helix that composes the gene. This process is divided into three phases: initiation, elongation, and termination. Transcription is begun by RNA polymerase at the DNA sequences near the beginning of genes, called promoters. After that, RNA polymerase adds nucleotides to the growing RNA polymer in the 5' to 3' direction. Transcription uses ribonucleotide triphosphates (ATP, CTP, GTP, and UTP) instead of deoxyribonucleotide triphosphates. Transcription is terminated at the sequences in the RNA products, known as terminators, where RNA polymerase stops this process. In eukaryotic organisms most primary transcripts undergo processing in the nucleus before they migrate to the cytoplasm to direct protein synthesis, and this process plays a fundamental role in the evolution of complex organisms. Some RNA processing in eukaryotes modifies only the 5' or 3' ends of the primary transcript, leaving the information content of the rest of the mRNA untouched. Other processing deletes information from the middle of the primary transcript, leaving the content of the mature mRNA related, but not identical, to the complete set of DNA nucleotide pairs in the original gene. Enzymes known

as methyl transferases add methyl ($-CH_3$) groups to G nucleotide at the 5' end of a eukaryotic mRNA, forming a so-called methylated cap. Like the 5' methylated cap, the 3' end of most eukaryotic mRNAs is not encoded directly by the gene and in a large majority of eukaryotic mRNAs, the 3' end consists of 100–200 base pairs, referred to as a poly-A tail. Both the methylated cap and the poly-A tail are critical for the efficient translation of the mRNA into protein.

3.5. Translation: From mRNA to Protein

Translation is the process by which the sequence of nucleotides in a messenger RNA directs the assembly of the correct sequence of amino acids in the corresponding polypeptide and takes place on ribosomes that coordinate the movements of transfer RNAs carrying specific amino acids with the genetic instructions of an mRNA. Ribosomes facilitate polypeptide synthesis by recognizing mRNA features that signal the start of translation: helping to ensure accurate interpretation of the genetic code by stabilizing the interactions between tRNAs and mRNAs *via* codon-anticodon recognition; supplying the enzymatic activity that links the amino acids in a growing polypeptide chain, and helping end polypeptide synthesis by dissociating both from the mRNA directing polypeptide construction and from the polypeptide product itself. The small subunit is the part of the ribosome that initially binds to mRNA and the larger subunit contributes an enzyme known as peptidyl transferase, which catalyses formation of the peptide bonds joining adjacent amino acids (Fig. 8). Both the small and the large subunits contribute to three distinct tRNA binding areas known as the aminoacyl (A) site, the peptidyl (P) site, and the exit (E) site. This process is divided into three phases: initiation, elongation, and termination.

AUG is the first (initiation) codon to be translated at the 5' end of the gene's reading frame, and is recognized by initiating tRNAs carrying a modified form of methionine called N-formylmethionine (fMet). The ribosome moves along the mRNA in the 5' to 3' direction. At each step of translation, the polypeptide grows by the addition of the next amino acid in the chain to its C terminus. Translation terminates when the ribosome reaches a UAA, UAG, or UGA nonsense codon at the 3' end of the gene's reading frame.

Protein structure is not irrevocably fixed upon completion of translation (Fig. 9). Several different processes may subsequently modify a polypeptide's structure, such as cleavage and removal of amino acids, such as the N-terminal fMet, from a polypeptide, or generation of smaller polypeptides from one larger product. The addition of chemical constituents, such as phosphate groups, methyl groups, or carbohydrates, to specific amino acids can also modify a polypeptide after translation and this process is referred to as post-translational modifications, very important or biochemical function of many enzymes.

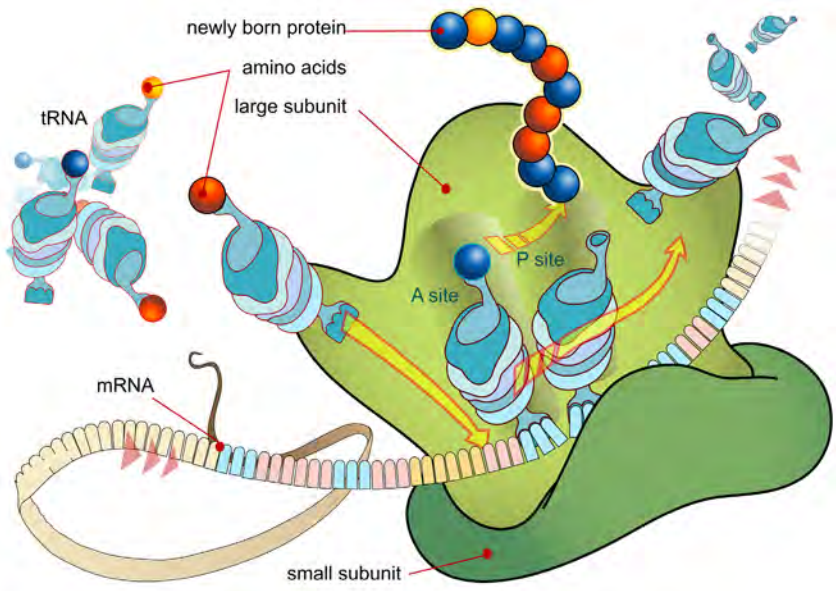

Figure 8. Translation process (*figure used with permission under Creative Commons license*)

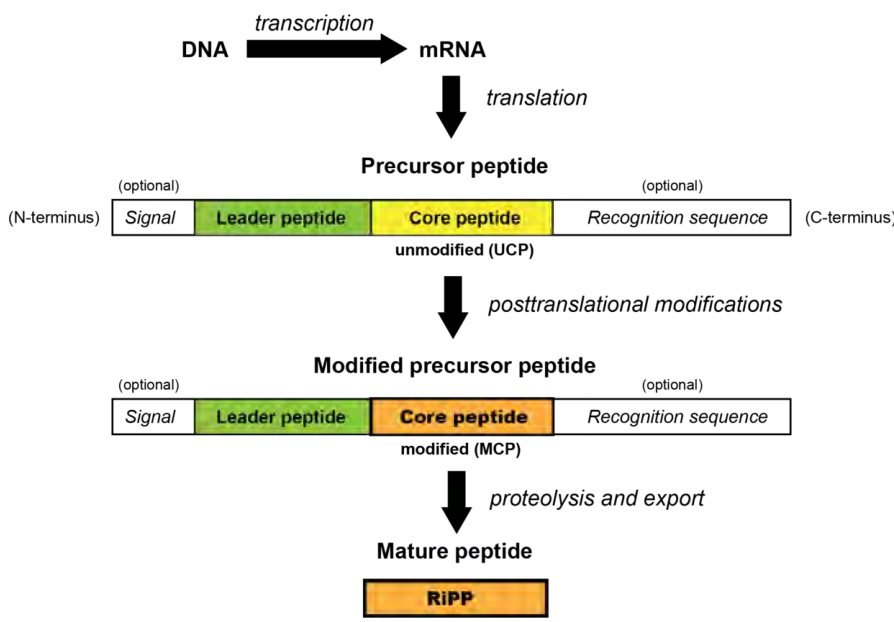

Figure 9. Biosynthesis of protein (*figure used with permission under Creative Commons license*)

3.6. Prokaryotic and Organelle Genetics

The advent of recombinant DNA technology in the 1970s and '80s facilitated an understanding of genes, chromosomes, and restriction enzymes in bacteria. The bacterial chromosome is the essential component of a typical bacterial genome and is a single molecule of double-helical DNA arranged in a circle with 4–5 Mb long in most of the commonly studied species. Inside the cell, the long, circular DNA molecule condenses by supercoiling and looping into a densely packed nucleoid body.

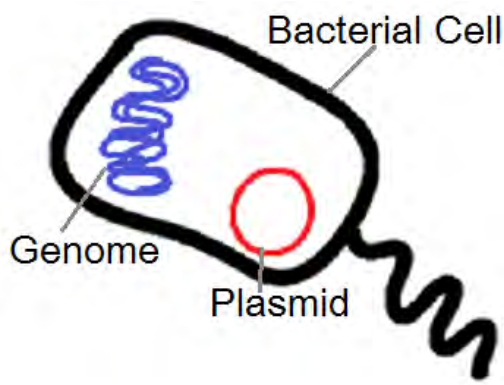

Figure 10. Genetic information within bacterial cell (*figure used with permission under Creative Commons license*)

Bacteria carry the genes necessary for their growth and reproduction in their large circular chromosome. In addition, some bacteria carry genes not needed for growth and reproduction under normal conditions in smaller circles of double-stranded DNA known as plasmids. These elements carry genes that protect their hosts against toxic metals, encoding resistance to antibiotics or contribute to their pathogenicity.

3.7. Analysis of molecular information

3.7.1. Sequence-specific DNA Fragmentation

A restriction enzyme recognizes a specific sequence of bases anywhere within the genome and then severs two covalent bonds (one in each strand) in the sugar-phosphate backbone at particular positions within or near that sequence. This digestion process generates well-defined restriction fragments

suitable for manipulation and characterization. Restriction enzymes originate in and can be purified from bacterial cells and protect them from viral infection by digesting viral DNA. Bacteria shield their DNA from digestion by their own restriction enzymes through the selective addition of methyl groups to the restriction recognition sites in their DNA. Restriction enzymes recognize target sequences of 4–8 bp in DNA isolated from any other organism and cut the DNA at or near these sites. For the majority of these enzymes, the recognition site contains 4–6 base pairs and exhibits a kind of palindromic symmetry in which the base sequences of each of the two DNA strands are identical when read in the 5'-to-3' direction. Most enzymes make their cuts in one of two ways: either straight through both DNA strands right at the line of symmetry to produce fragments with blunt ends, or displaced equally in opposite directions from the line of symmetry by one or more bases to generate fragments with single-stranded ends-sticky ends (easy to ligation process, because they are free to base pair with a complementary sequence from the DNA of any organism cut by the same restriction enzyme) (Figs. 11 and 12). They are an important molecular tool to create unique DNA fragments and also as an analytic tool to create restriction maps of viral genomes and other purified DNA fragments showing the relative order and distances between multiple restriction sites, which thus act as landmarks along a DNA molecule.

Figure 11. Restriction enzyme recognition site leaving sticky ends (*figure used with permission under Creative Commons license*)

Figure 12. Restriction enzyme recognition site leaving blunt ends (*figure used with permission under Creative Commons license*)

One of the most commonly used methods for creating a restriction map involves digestion fragments made by multiple restriction enzymes, followed by gel electrophoresis to visualize the produced fragments.

Restriction enzyme digestion and standard agarose gel electrophoresis allow only for analysis of simple DNA molecules, in up to 50,000 base pairs (50 kbp). The genomes of animals, plants, and even microorganisms are far too large to be analysed in this way. Only purified and then amplified original DNA molecules can be treated with chemical and physical techniques to analyse the isolated DNA fragment.

3.7.2. Cloning Fragments of DNA

There are the two strategies to accomplish the purification and amplification of individual fragments, such as molecular cloning, which replicates individual fragments of previously uncharacterized DNA, and the polymerase chain reaction (PCR), which can purify and amplify a previously sequenced genomic region.

Molecular cloning, which consists of two basic steps, is the process that takes a complex mixture of restriction fragments and uses living cells to purify and make many exact replicas of just one fragment at a time (Fig. 13). In the first step, DNA fragments are inserted into specialized carriers called vectors necessary for the proper transport, replication, and purification of individual inserts. In the second step, vector harbouring insert is transported into living cell, where many identical copies of this molecule (DNA clones) are made (Fig. 14). These DNA clones may be purified for immediate study or stored within cells as libraries for future analysis. When the inserted DNA fragment and DNA vector are from two different origins, the spliced molecule creates a recombinant DNA.

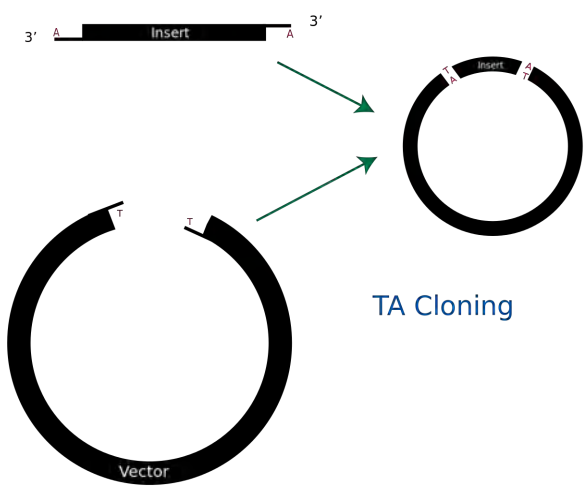

Figure 13. Introducing DNA insert into DNA vector (*figure used with permission under Creative Commons license*)

A vector must possess distinguishing features, such as size or shape, by which it can be purified away from the host cell's genome. Creating single-stranded sticky ends is a basis for the efficient production of a vector-insert recombinant. Moreover, the ends produced with the same enzyme are complementary in sequence and are available for base pairing, and no matter what the origin of the DNA. Such digested DNA mixtures in the presence of DNA ligase allows for creating completed, stable recombinant DNA molecule.

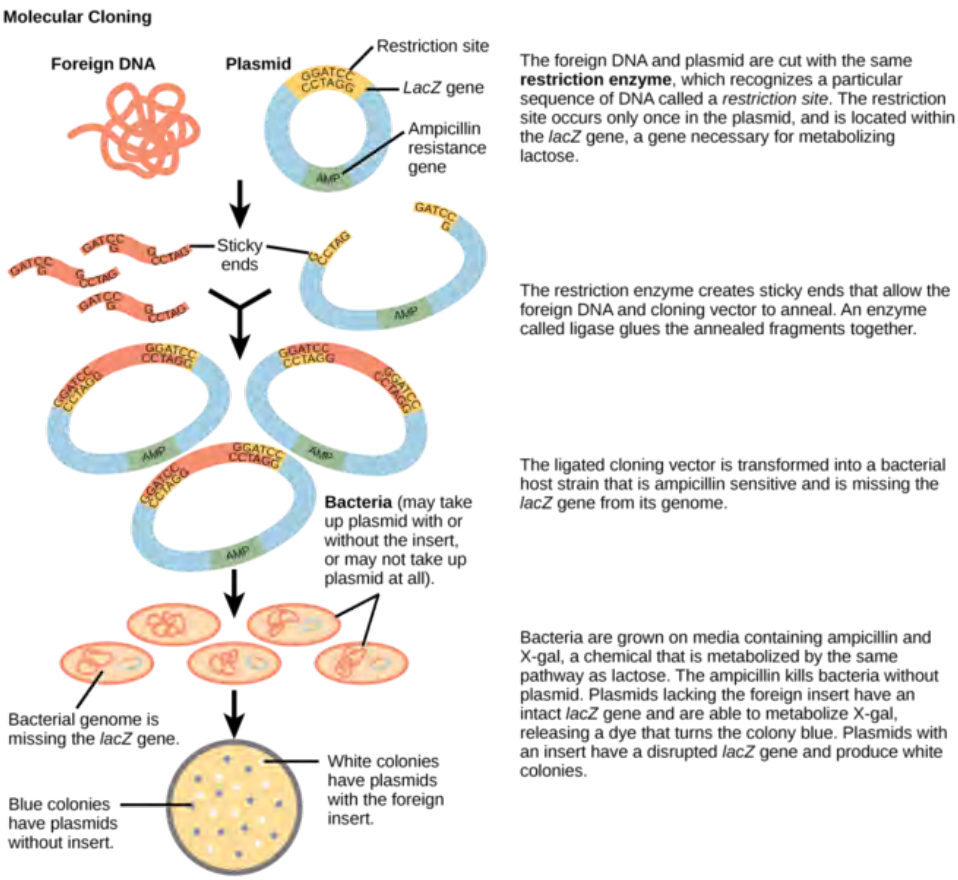

Figure 14. Molecular cloning (*figure used with permission under Creative Commons license*)

The simplest DNA vector is a plasmid vector. Each plasmid vector carries an origin of replication and a gene for resistance to a specific antibiotic necessary for proper selection of positive clone (cells harbouring a DNA vector). For the efficient cloning very important is distinguishing insert-containing recombinant

molecules from vectors without inserts. Figure 14 illustrates the three-part process with a plasmid vector containing an origin of replication, the gene for resistance to ampicillin (amp R), and the *E. coli* lacZ gene, which encodes the enzyme β-galactosidase. It is possible to insert foreign DNA into the gene at that location and then use the disruption of the lacZ gene function to distinguish insert-containing recombinant molecules from vectors without inserts. Taking up a foreign DNA molecule by bacterial cell, changing the genetic characteristics of that cell is known as transformation.

3.7.3. DNA Sequence Analysis

The DNA sequence of a genome provides a lot of practical information, such as the evolution of genomes when the DNA sequences between genes are compared. Additionally, comparison of genomic and cDNA sequences immediately shows how a gene is divided into exons and introns, and can suggest whether alternative splicing of the gene's primary transcript is occurring.

The Sanger method

There are two steps to the Sanger method of sequencing. The first step is the generation of a complete series of single-stranded sub-fragments complementary to a portion of the DNA template under analysis, which differs in length by a single nucleotide from the preceding and succeeding fragments, distinguishable according to their terminal 3' base. In the second step of the sequencing process, the mixture of DNA sub-fragments through polyacrylamide gel electrophoresis is analysed, under conditions that allow the separation of DNA molecules differing in length by just a single nucleotide.

In this method he single strands are mixed in a solution with DNA polymerase, the four deoxynucleotide triphosphates, and a radioactively labelled oligonucleotide primer complementary to DNA adjacent to the 3' end of the template strand under analysis. The solution is next divided into four aliquots. To each one, a small amount of a single type of a nucleotide triphosphate lacking the 3'-hydroxyl group (dideoxynucleotide), which is critical for the formation of the phosphodiester bonds that lead to chain extension is added. Dideoxynucleotide comes in four forms: ddTTP, ddATP, ddGTP, or ddCTP. In each sample reaction tube, the oligonucleotide primer hybridizes at the same location on the template DNA strand. As a primer, it will supply a free 3' end for DNA chain extension by DNA polymerase by adding the nucleotides to the growing strand that are complementary to those of the sample's template strand. The addition of nucleotides continues until, by chance, a dideoxynucleotide is incorporated instead of a normal nucleotide. The absence of a 3'-hydroxyl group in the dideoxynucleotide prevents the DNA

polymerase from forming a phosphodiester bond with any other nucleotide, ending the polymerization for that new strand of DNA. After enough time for the polymerization of all molecules to reach completion, templates may be released from the newly synthesized strands by denaturing the DNA at high temperature. Each sample tube now holds a whole collection of single-stranded radioactive DNA chains as well as the nonradioactive single strands of the template DNA. The lengths of the radioactive chains reflect the distance from the 5' end of the oligonucleotide primer to the position in the sequence at which the specific dideoxynucleotide present in that particular tube was incorporated into the growing chain. The samples in the four tubes are now electrophoresed in adjacent lanes on a polyacrylamide gel, and the gel is subjected to a system that detects the presence of the radioactive label. Each band represents a chain that is one nucleotide longer than the chain of the band below.

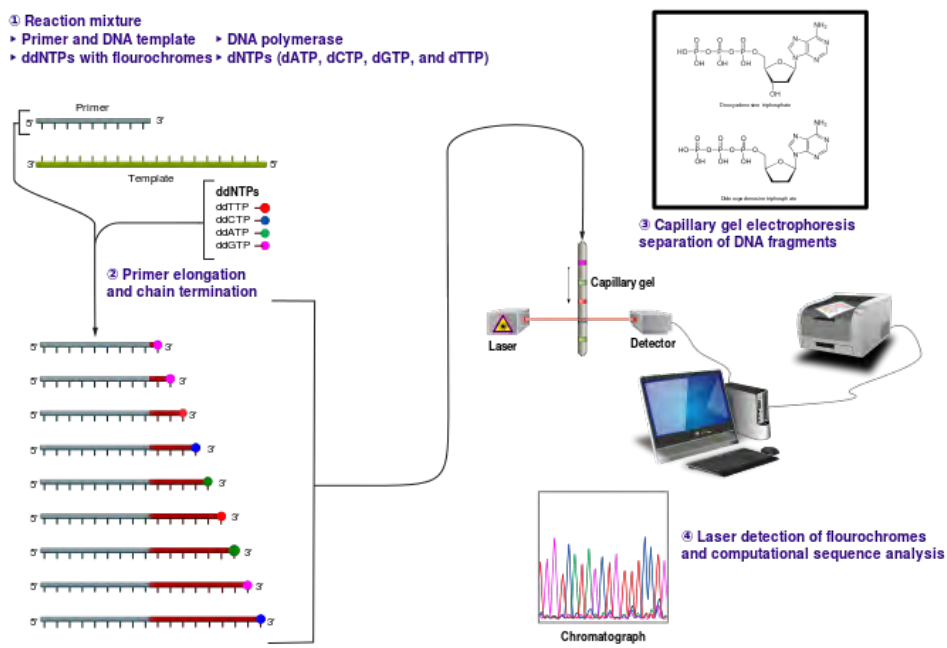

Figure 15. Automated Sanger sequencing (*figure used with permission under Creative Commons license*)

To automate the DNA sequencing process, the method of labelling the newly formed complementary DNA strands has changed; each of the four chain-terminating dideoxynucleotides is labelled with a different colour fluorescent dye. A DNA sequencing machine follows the DNA chains of each

length in the ascending series through a special detector that can distinguish the different colours associated with each terminating dideoxynucleotide (Fig. 15).

High-throughput sequencing instruments can analyse up to 384 DNA fragments at a time. New technologies and strategies allow the sequencing of 100 billion bp in 96 hours using highly-parallel simultaneous analyses of millions of DNA fragments, and facilitate sequencing of individual human genomes.

References

Atkins J.F., Gesteland R.F., Cech T.R., *RNA worlds: From life's origins to diversity in gene regulation*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York 2011.

Bloomfield V.A., Crothers D.M., Tinoco I. Jr., Heast J.E., *Nucleic acids: Structures, properties, and functions*, University Science Books, Sausalito, California 2000.

Chambers D.A. (ed.), DNA: *The double-helix – Perspective and prospective at forty years*, Ann. N.Y. Acad. Sci. 1995, no. 758.

Clark B., Petersen H. (eds.), *Gene expression: The translational step and its control*, Alfred Benzon Symposium, vol. 19, Copenhagen, Munksgaard 1984.

Crick F.H.C., Barnett L., Brenner S., Watts-Tobin R.J., *General nature of the genetic code for proteins*, Nature 1961, no. 192.

Darnell J.E. Jr., *RNA*, Sci Am. 1985, no. 253.

Dickerson R.E., *The DNA helix and how it is read*, Sci. Am. 1983, no. 249.

Dreyfuss G., Kim V.N., Kataoka N., *Messenger-RNA-binding proteins and the messages they carry*, Nat. Rev. Mol. Cell Biol. 2002, no. 3.

Frazer K.A., Pachter L., Poliakov A., Rubin E.M., Dubchak I., *VISTA: Computational tools for comparative genomics*, Nucleic Acids Res. 2004, no. 32.

Gebauer F., Hentze M.W., *Molecular mechanisms of translational control*, Nat. Rev. Mol. Cell Biol. 2004, no. 5.

Gesteland R.F., Cech T.R., Atkins J.F. (eds.), *The RNA world*, (3rd ed.), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York 2006.

Green M., Sambrook J., *Molecular cloning: Alaboratory manual*, (4th ed.), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York 2012.

Griffiths A.J.F., Gelbart W.M., Lewontin R.C., Miller J.H., *Modern genetic analysis*, (2nd ed.), W.H. Freeman, New York 2002.

Hartwell L., Hood L., Goldberg M.L., Reynolds A.E., Silver L.M., Veres R.C., *Genetics: From genes to genomes*, (2nd ed.), McGraw-Hill, New York 2003.

Herbert K.M., Greenleaf W.J., Block S.M., *Single-molecule studies of RNA polymerase: Motoring along*, Annu. Rev. Biochem. 2008, no. 77.

Human Genome, Science 2001, no. 291.

Human Genome, Nature 2001, no. 409.

The Human Genome at Ten, Nature 2011, no. 464.

International Human Genome Sequencing Consortium, Finishing the euchromatic sequence of the human genome, Nature 2004, no. 431

Isaacs F.J., Carr P.A.,Wang H.H., Lajoie M.J., Sterling B., Kraal L., Tolonen A.C., Gianoulis T.A., Goodman D.B., Reppas N.B., et al., *Precise manipulation of chromosomes in vivo enables genome-wide codon replacement*, Science 2011, no. 333.

Kornberg R.D., Thomas J.O., *Chromatin structure; oligomers of the histones*, Science 1974 May 24, no. 184(4139).

Ling J., Reynolds N., Ibba I., *Aminoacyl-tRNA synthesis and translational quality control*, Annu. Rev. Microbiol. 2009, no. 63.

Maniatis T., Goodbourn S., Fischer J.A., *Regulation of inducible and tissue-specific gene expression*, Science 1987, Jun 5, no. 236(4806).

Maniatis T., Reed R., *An extensive network of coupling among gene expression machines*, Nature 2002, no. 416.

Mathews M.B., Sonenberg N., Hershey J.W.B., *Translational control in biology and medicine*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York 2007.

McManus C.J., Graveley B.R., *RNA structure and the mechanisms of alternative splicing*, Curr. Opin. Genet. Dev. 2011, no. 21.

Osoegawa K., Mammoser A.G., Wu C., Frengen E., Zeng C., Catanese J.J., de Jong P.J., *A bacterial artificial chromosome library for sequencing the complete human genome*, Genome Res. 2001, no. 11.

Shatkin A.J., *Capping of eucaryotic mRNAs*, Cell. 1976, Dec. 9(4 PT 2).

Snustad D.P., Simmons M.J., *Principles of genetics*, (3rd ed.), Wiley, New York 2002.

Watson J.D., Crick F.H.C., *Molecular structure of nucleic acids: A structure for deoxyribonucleic acids*, Nature 1953a, no. 171; *Genetical implications of the structure of deoxyribonucleic acids*, Nature 1953b, no. 171.

Watson J.D. (ed.), *Structures of DNA*, Cold Spring Harbor Symposium on Quantitative Biology, Vol. 47, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York 1982.

Yates J.R. III, Gilchrist A., Howell K.E., Bergeron J.J., *Proteomics of organelles and large cellular structures*, Nat. Rev. Mol. Cell. Biol. 2005, no. 6.