Nucleic Acids

Margaret Scofield Creighton University, Omaha, United States

Nucleic acids are polymers of acidic monomeric subunits known as nucleotides. The nucleotides form a duplex, or double-stranded, molecule referred to as deoxyribonucleic acid (DNA) that stores genetic information within the cell. The genetic information in DNA is transferred to ribonucleic acid (RNA), monomeric forms of nucleic acids that are primarily single-stranded molecules. The three major RNA species differ in their composition and function. These are designated ribosomal RNA (rRNA), transfer RNA (tRNA), and messenger RNA (mRNA). The rRNA represents approximately 80% of cellular RNA, tRNA approximately 15%, and mRNA 3–5%. The RNA molecules participate in the conversion of the genetic information stored in DNA into **proteins** that are crucial for cellular function. Nucleic Acid Building Blocks: The nucleotide components of the nucleic acids include a heterocyclic nitrogenous base, a pentose sugar, and a phosphate (Fig. 1). At physiological pH, the phosphate of the nucleotide is completely ionized to the anionic form and the nitrogenous base is linked through an N-beta glycosidic bond to the 1′ carbon of the pentose sugar. By convention, the carbon atoms of the pentose sugars are numbered 1′ to 5′, with the phosphate esterified to the 5′ carbon of the sugar. The pentose sugars, deoxyribose, and ribose, are found in DNA and RNA, respectively. These sugars differ in the absence (deoxyribose) or presence (ribose) of a hydroxyl group at the 2′ carbon of the pentose (Fig. 2). The presence of the 2′ hydroxyl group is responsible for the instability of RNA molecules.

Nitrogenous Bases: There are two basic types of nitrogenous bases, pyrimidines and purines (Fig. 3). These are essentially planar, hydrophobic, weak bases. Five nitrogenous bases are found in nucleic acids (Fig. 4); adenine (A), guanine (G), and cytosine (C) are in both DNA and RNA, whereas thymine (T) is almost exclusively found in DNA, and uracil (U) almost exclusively in RN A. DNA and RNA are quantified by their absorption of UV light at 260 nm. This absorption is the result of the resonating conjugated double bonds of the purines and pyrimidines.

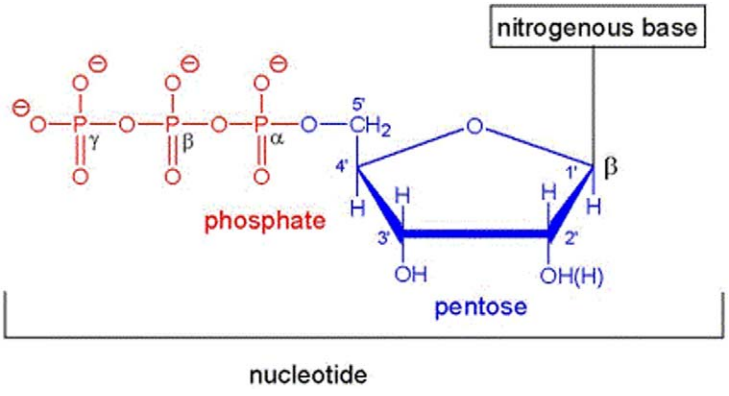

Fig.1 Nucleotide. The building block of nucleic acids, the nucleotide, is composed of a nitrogenous base, a pentose sugar, and one to three phosphates labeled as α, β, and γ.

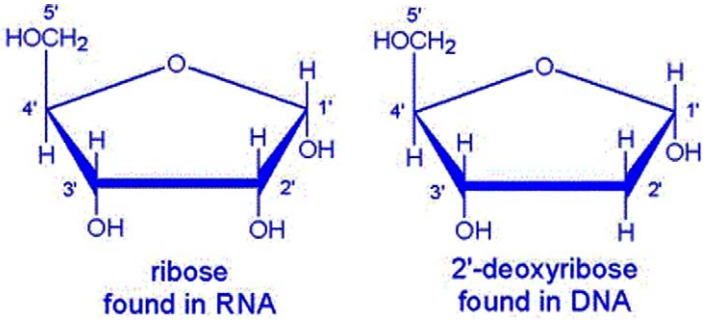

Fig. 2 Pentose sugars. Ribose and deoxyribose are found in RNA and DNA, respectively.

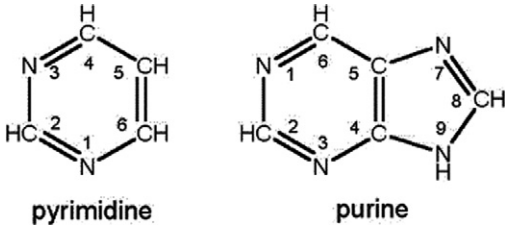

Fig. 3 Nitrogenous bases. The nitrogenous bases are classified as either purines or pyrimidines.

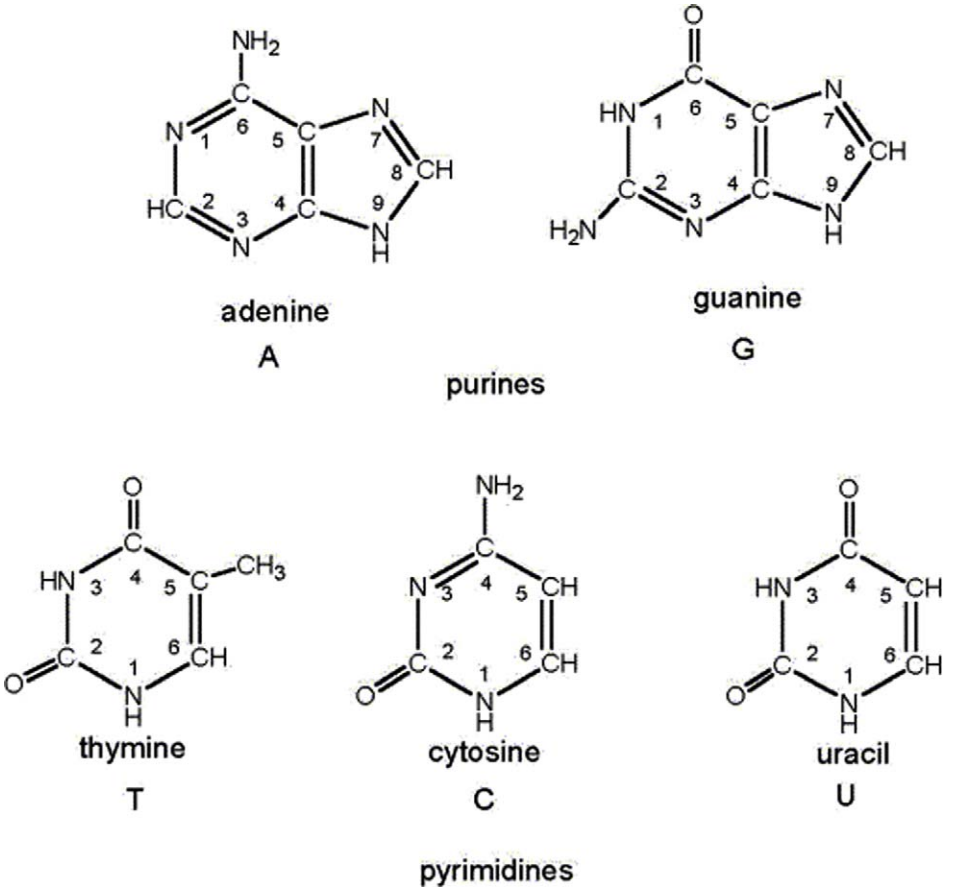

Fig. 4 Purines and pyrimidines. Adenine (A) and guanine (G) are the two most common purines. Thymine (T), cytosine (C), and uracil (U) are the three most common pyrimidines.

Minor Nitrogenous Bases: While present in small amounts, other purines and pyrimidines play an important role in **regulating gene expression**, in determining the secondary structure of tRNA, and in nucleic acid metabolism. For example, the purine hypoxanthine (Fig. 5) is found in tRNA and is a product of the degradation of adenosine, the nucleoside form of adenine. Other nitrogenous bases found in tRNA include 7-methylguanine, N^2dimethylguanine, and 5, 6 dihydrouracil. 5-Methylcytosine and N^6-methyladenine are located in DNA as well as tRNA, whereas 5-hydroxymethylcytosine is found in DNA only. Some cancer chemotherapeutic agents, such as **5-fluorouracil**, are analogs of the purines and pyrimidines that interfere with the S phase of the **cell cycle**, thereby inhibiting cell replication (Fig. 6).

Nucleosides: Nucleosides are nitrogenous bases that are covalently bonded to ribose or deoxyribose by a β-*N*-glycosidic bond resistant to alkali cleavage. The 1′ carbon of the sugar is linked by a β-*N*-glycosidic bond between N-1 of the pyrimidine and N-9 of the purine (Figs. 7 and 8). The *N*-glycosidic bond of the purine nucleosides is easily hydrolyzed with weak acid, while the N-glycosidic bond of the pyrimidine nucleosides is only cleaved with strong acids, such as perchlorate. The ribonucleosides are uridine, cytidine, adenosine, and guanosine, and the deoxyribonucleosides are thymidine (or deoxythymidine), deoxycytidine, deoxyadenosine, and deoxyguanosine.

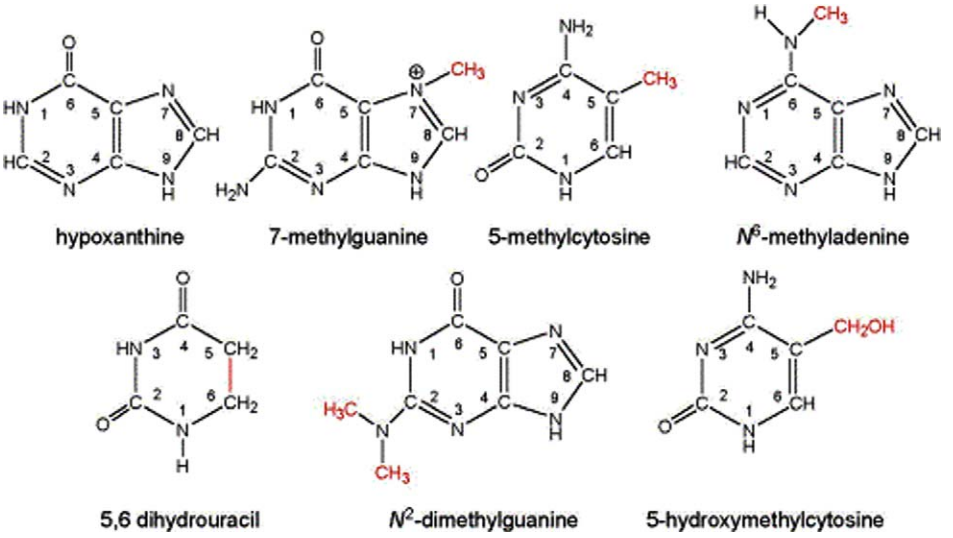

Fig. 5 Minor purine and pyrimidine bases. The common bases can be modified by the addition of methyl groups, reduction of double bonds, or removal of functional groups as shown in red.

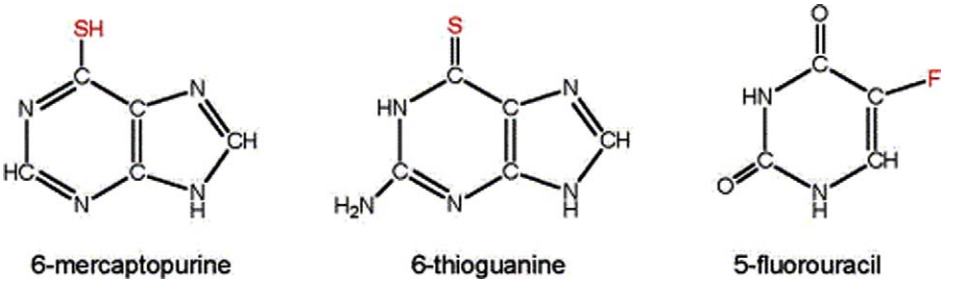

Fig. 6 Anticancer drugs. Purine or pyrimidine drug analogs where the substituent modifying the purines or pyrimidines is shown in red.

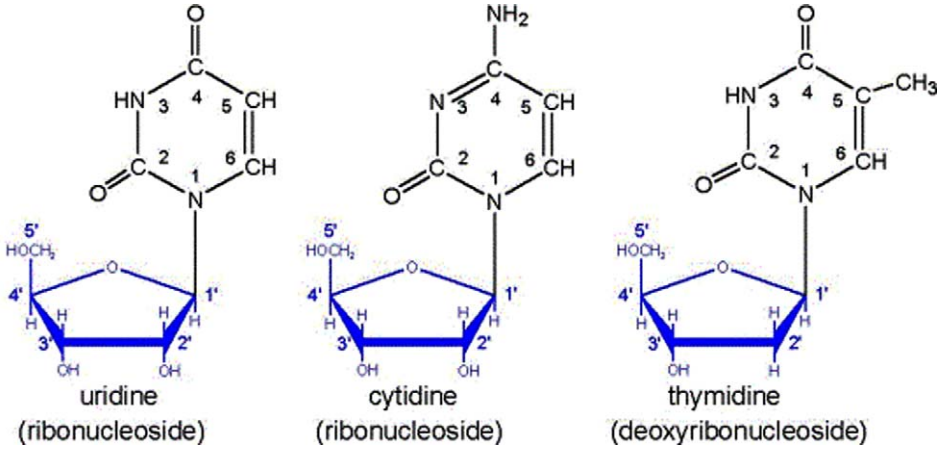

Fig. 7 Pyrimidine nucleosides.

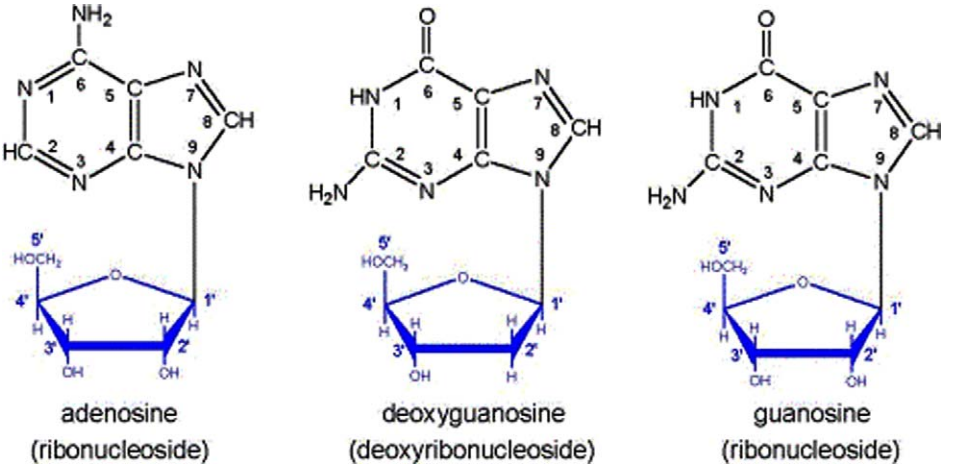

Fig. 8 Purine nucleosides.

Minor Nucleosides: The less abundant tRNA nucleosides include inosine (hypoxanthine+ribose) and pseudouridine, where the ribose is attached to C-5 (Fig. 9). In addition, several nucleoside analogs are used as chemotherapeutics or antivirals (Fig. 10). These drugs are composed of a nitrogenous base covalently attached to a molecule other than ribose or deoxyribose. **Acyclovir**, a guanosine analog (acycloguanosine), is a representative antiviral agent used to treat **herpes infections**. **Zidovudine** and didanosine are **nucleoside reverse transcriptase inhibitors** used to treat **HIV infections**, and **cytarabine** is a cancer chemotherapeutic agent. In many cases, these analogs interfere with enzymes involved in DNA replication and chain elongation due to the lack of the critical 3′ hydroxyl group that must be present to synthesize DN A. In the case of cytarabine, the arabinose sugar differs only stereochemically from ribose at the hydroxyl group of C-2′, which is sufficient to inhibit DNA synthesis.

Nucleotides: Nucleotides are synthesized by kinase esterification of a phosphate on the 5′ carbon of the nucleosides to form monophosphate nucleotides. Kinases also synthesize the phosphoanhydride bonds between the α and β phosphorous to yield diphosphate

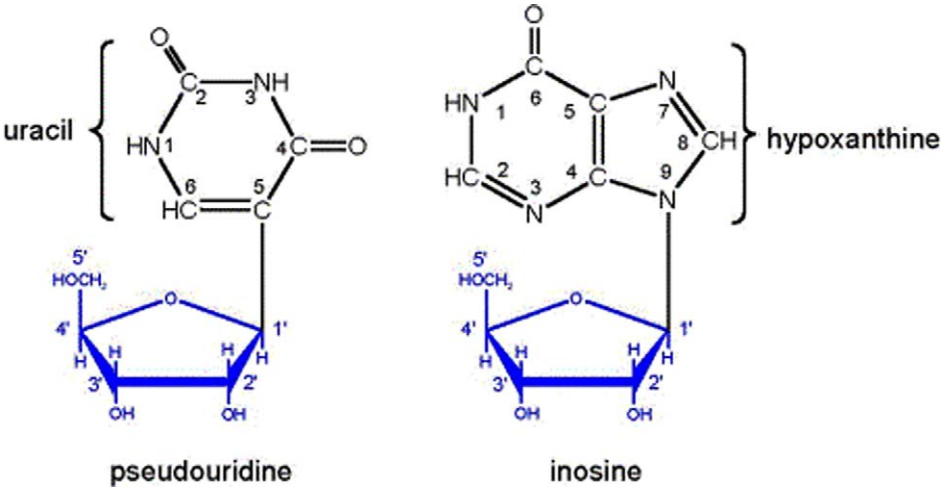

Fig. 9 Unusual nucleosides in tRNA.

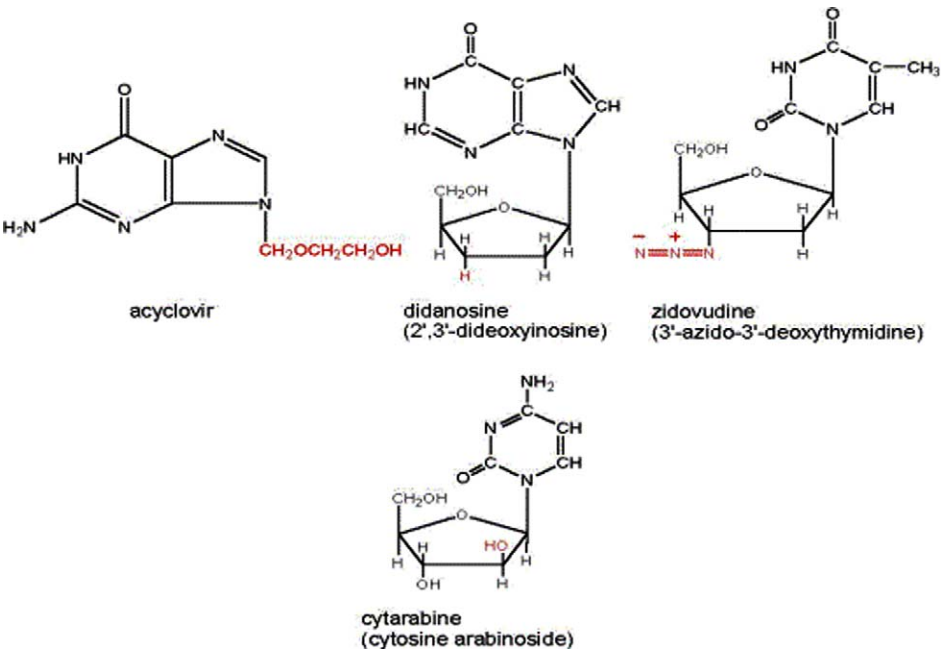

Fig. 10 Antiviral and anticancer agents. Nucleosides are modified as shown in red to form drugs to treat viral infections and cancer.

nucleotides, and between the β and γ phosphorous to produce triphosphate nucleotides. The polar phosphate on the nucleotides contributes significantly to the aqueous solubility of the nucleic acids. In addition, the energy released by cleaving nucleotide α β and β γ phosphoanhydride bonds is the driving force for many enzymatic reactions.

Nucleotide Nomenclature: The structure and nomenclature of the mono-, di-, or triphosphate nucleotides of adenine are illustrated in Fig. 11. The nomenclature and abbreviations for the remaining purine and pyrimidine nucleotides follow a similar pattern, with the name of the nucleoside (deoxyribonucleoside or ribonucleoside) followed by 5′ monophosphate, 5′ diphosphate, or 5′ triphosphate. Collectively, all the

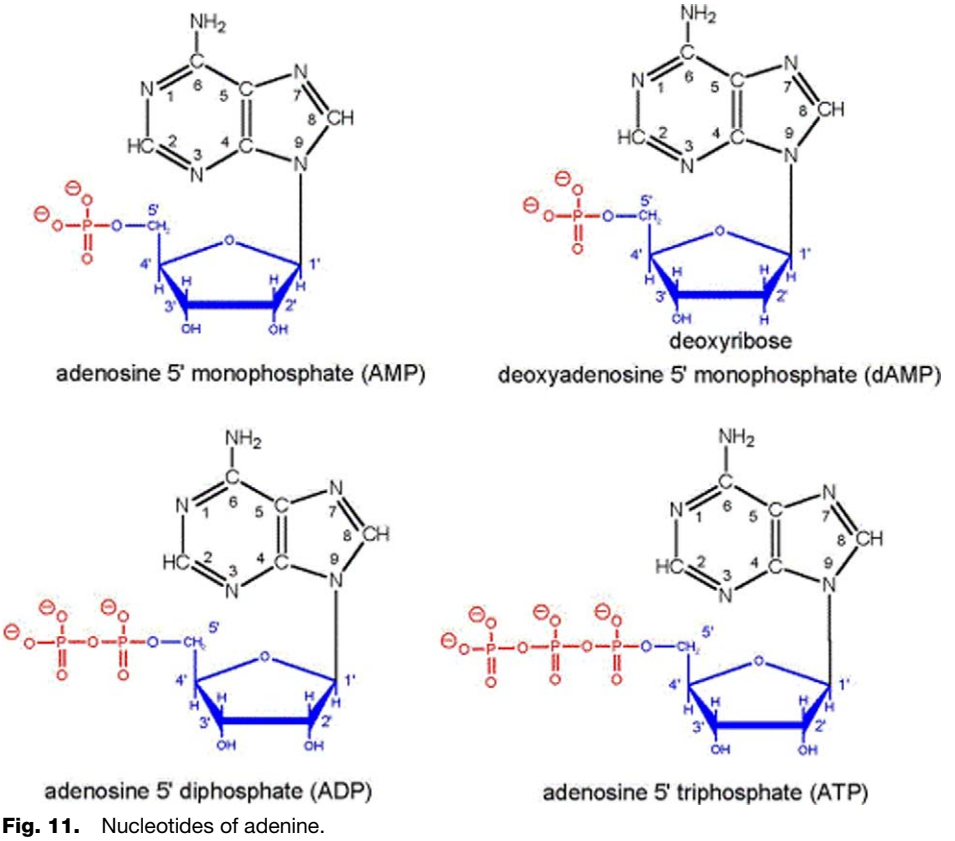

Fig. 11. Nucleotides of adenine.

common deoxyribose or ribose nucleotides are often referred to as deoxynucleoside triphosphates (dNTPs) or ribonucleoside triphosphates (NTPs), respectively. There are other important nucleotide analogs, such as cyclic AMP and cyclic GMP (Fig. 12) that are not used to synthesize nucleic acids but rather participate directly in signal transduction and regulation of cellular processes. Finally, other adenosine nucleotide derivatives are used as cofactors in enzymatic reactions. This group includes coenzyme A, flavin adenine dinucleotide (FAD), and nicotinamide adenine dinucleotide (NAD).

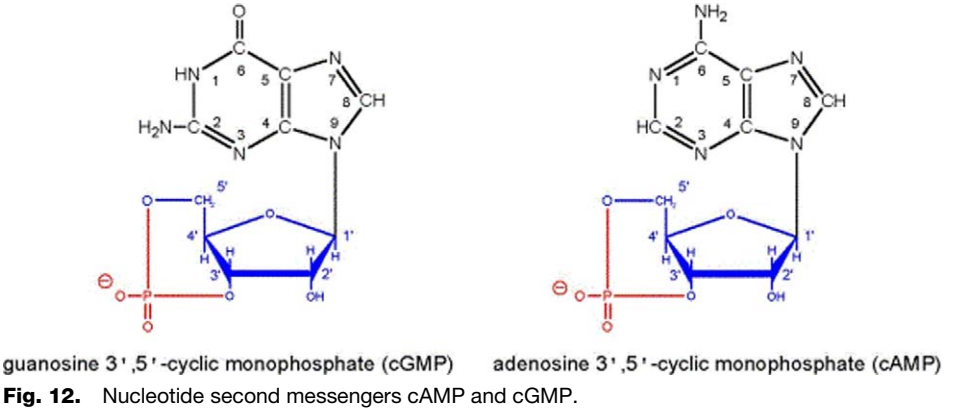

Fig. 12. Nucleotide second messengers cAMP and cGMP.

Nucleic Acid Polarity: The nucleotides are linked together by covalent phosphodiester bonds between the 3′ hydroxyl of ribose or deoxyribose, and the 5′ hydroxyl group of the next sugar to form RNA or a single strand of DNA (Fig. 13). The result is a chain, or single-stranded molecule, with a sugar-phosphate backbone supporting variable nitrogenous bases in a particular sequence, with a distinct 5′ phosphorylated carbon at one end of the strand and a free 3′ hydroxyl group on the other end. These distinct ends impart a particular orientation or polarity to the molecule. The sequence of the nitrogenous bases in the molecule shown on Fig. 13 is TACG by a convention based on the 5′ to 3′ polarity. If reversed to GCAT, the sequence would represent another distinct nucleic acid.

Nucleic Acid Synthesis: Nucleic acids are synthesized from triphosphate nucleotide precursors by DNA or RNA polymerases using DNA as a template. The monophosphorylated nucleotides are linked together by phosphodiester bonds between the 3′

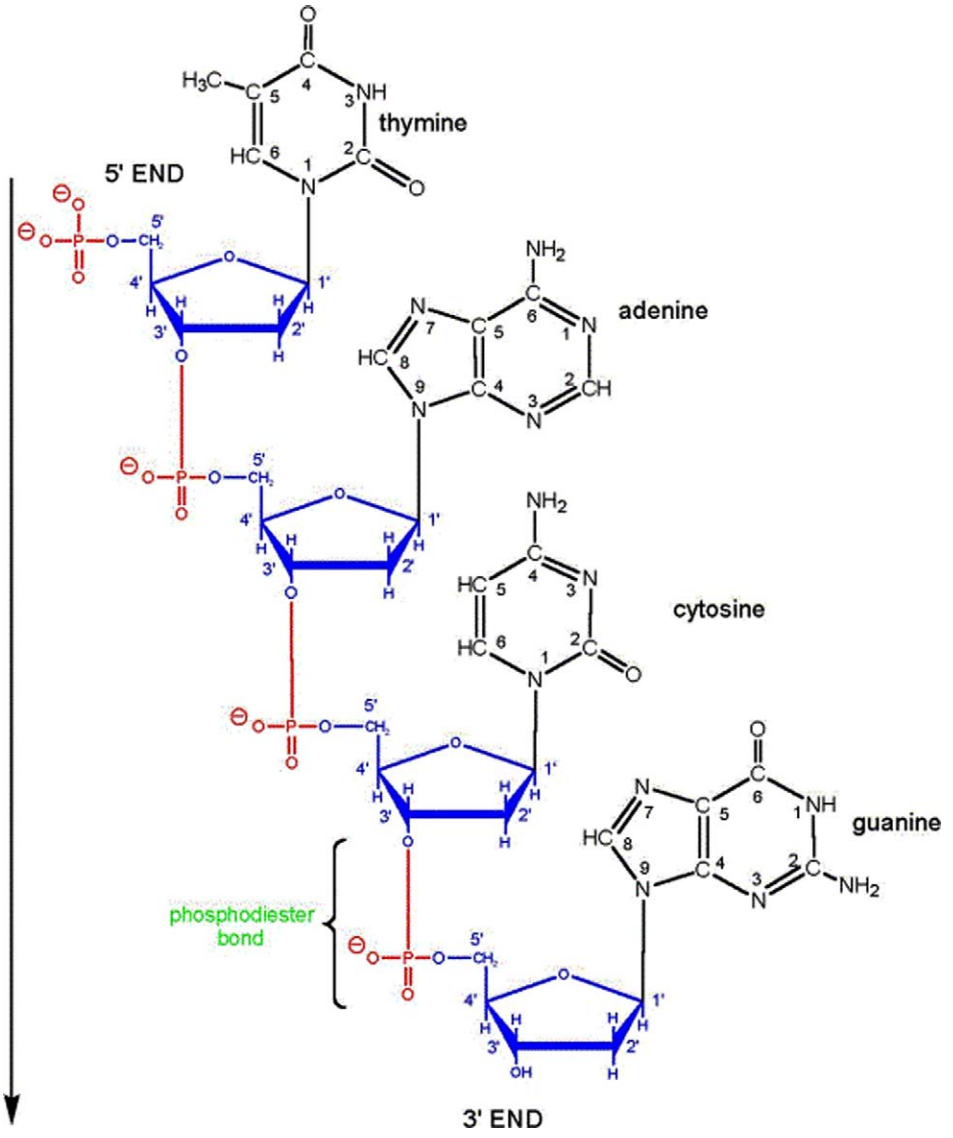

Fig. 13 DNA polarity. A single-stranded DNA molecule of the sequence TACG as written in the direction of the arrow, from the 5′ to 3′ end of the molecule.

hydroxyl of ribose or deoxyribose, and the 5′ hydroxyl group of the next sugar in a 5′ to 3′ direction (Fig. 14). In addition to an elongated strand, the products include the release of pyrophosphate, two linked phosphates (Fig. 14). DNA is also synthesized from RNA by reverse transcriptases. For DNA synthesis to occur, each deoxyribonucleotide must have three phosphates. In addition, each DNA molecule must have a 3′ OH group available on the deoxyribose for catalyzing the phosphodiester linkage of the next nucleotide or for further elongation of the molecule. Phosphates must also be added in vivo to activate many antiviral and cancer chemotherapeutic agents (Fig. 10) to form pharmacologically active triphosphonucleotide analogs capable of interfering with DNA and RNA synthesis.

Double-stranded DNA: The two nucleic acid chains in double-stranded DNA have opposite polarities and are held together by two hydrogen bonds between adenine and thymine, and three hydrogen bonds between cytosine and guanine (Fig. 15). Thus, "adenine and thymine" and "guanine and cytosine" represent specific purine-pyrimidine base pairings between the two DNA strands. The resulting sequences, or order of nucleotide bases on each strand, are said to be complementary to one another. This very precise pairing of larger and smaller nitrogenous bases helps maintain a constant distance across the DNA molecules and allows semi-conservative copying or duplication

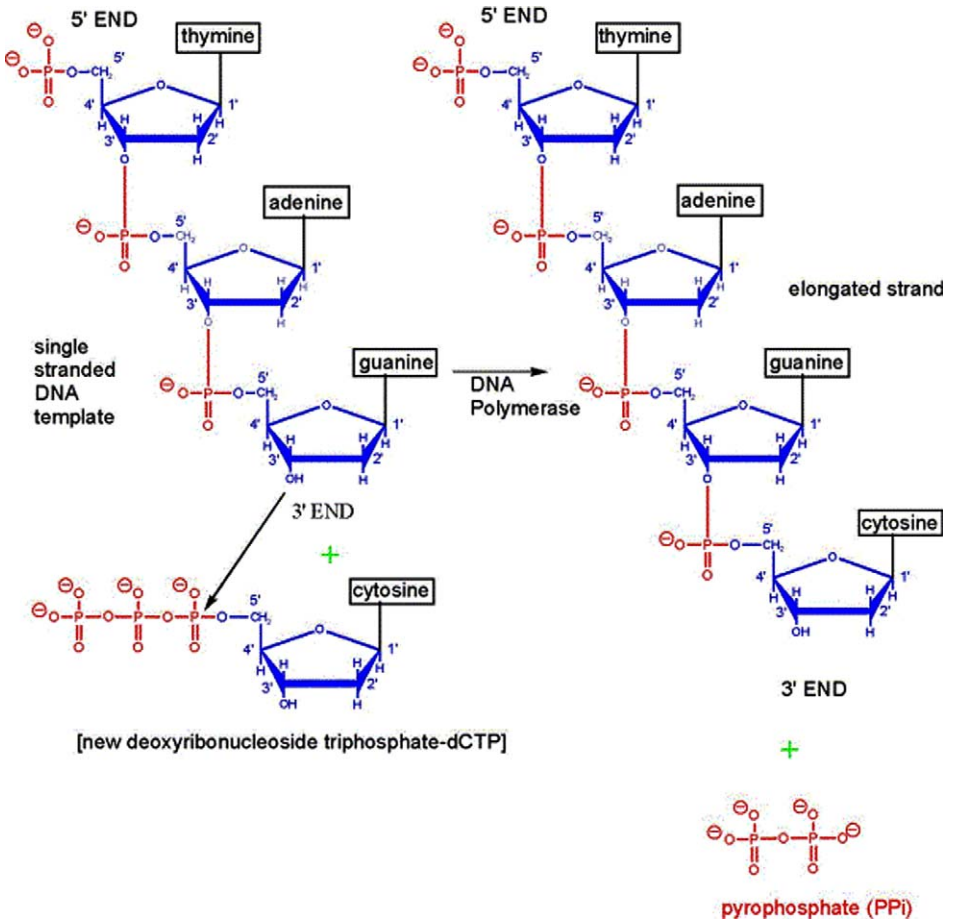

Fig. 14 Synthesis of DN A. 5′-Deoxynucleoside triphosphates are added to the terminal 3′-hydroxyl group of thegrowing DNA strand releasing one pyrophosphate molecule for each new deoxyribonucleotide added.

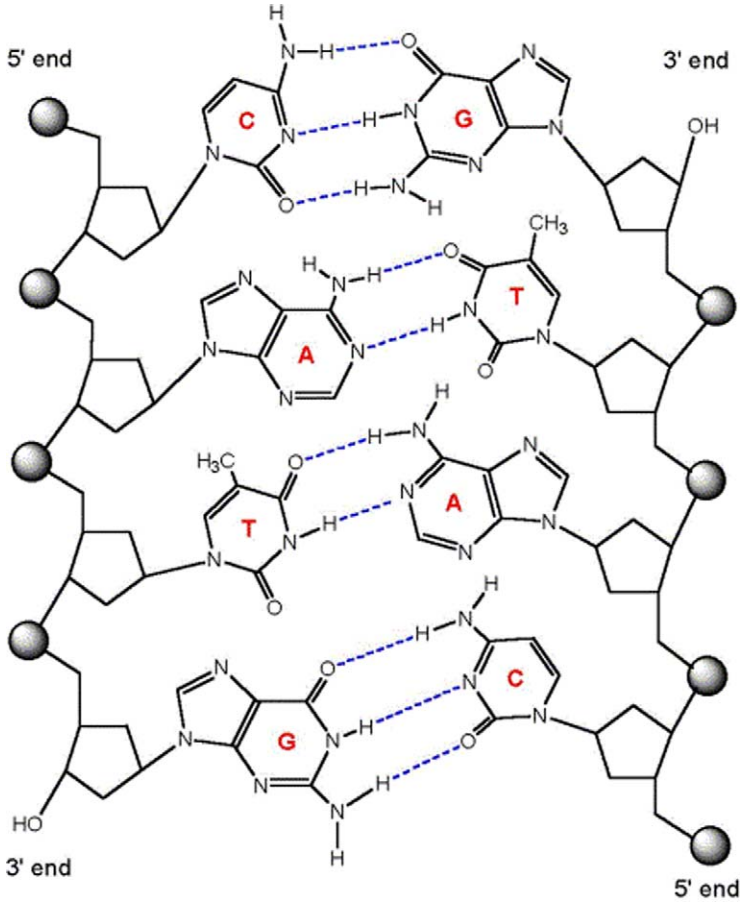

Fig. 15 Double-stranded DN A. Double-stranded DNA is held together by hydrogen bonds (dotted blue line) between G and C, and A and T. The phosphate (shaded circle) and deoxyribose (pentagon) represent the backbone.

of either strand. As a result, an A is always paired with a T, T with an A, G with a C, and C with a G, allowing the formation of two duplicate strands during replication (Fig. 16).

DNA Melting Temperatures: An increased frequency of nitrogenous base pairs, such as G and C with three hydrogen bonds in a DNA sequence, make it more difficult to separate or melt the two strands. The melting temperature, or T_m, is the temperature at which half the molecules are single-stranded. The reverse is true when A and T base pairs have an increased frequency in a DNA molecule, with the T_m being lower in this case. DNA strands are separated by heating to a precise temperature, or by raising the pH (>11.3) of the solution, a process known as denaturation. Single-stranded DNA also has a greater absorption at 260nm because the bases are no longer stacked in the double-stranded helix, a characteristic known as the hypochromic effect. Upon cooling a DNA solution, or returning the pH to neutral, the purines and pyrimidines eventually reform hydrogen bonds resulting in double-stranded DNA, a process that is referred to as annealing or hybridization.

DNA Conformations: DNA assumes several different conformations based on the frequency of purines or pyrimidines and its environment. In the most common structural form, or B form, the two chains run in opposite directions and the phosphodiester backbones form a right-handed double helix such that there are 10.5 base pairs per helical turn, with a diameter of 2.37 nm and a helix rise of 0.34 nm per base pair. Within the

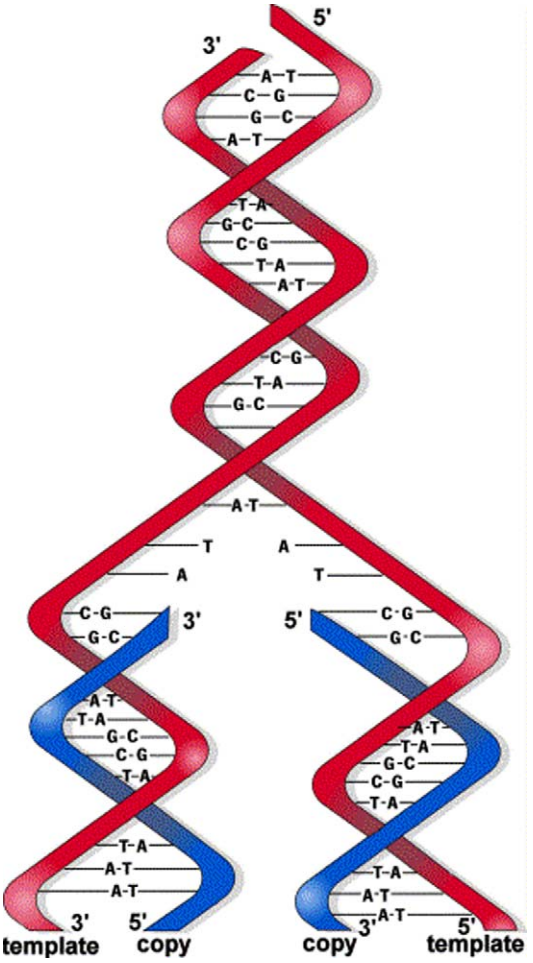

Fig. 16 Semi-conservative replication of double-helical DN A. The two DNA strands separate and each is used as a template to be copied or replicated by DNA polymerases in a 5′ to 3′ direction. Bases are added that are complementary to the template sequence, thus forming a growing strand or copy, and resulting in two new helices each with an old and new strand.

double-stranded DNA, the planar hydrophobic purine and pyrimidine base pairs stack on top of each other perpendicular to the sugar phosphate backbone lending stability to the double helix. In addition, indentations between the strands are found on the surface of the helix forming a major (wide and deep) and a minor (narrow and shallow) groove. These grooves allow DNA-binding proteins to interact directly with the DNA chain and to regulate gene activity. Another conformational form, the A-DNA, occurs when the environment is not aqueous. In this case, the wider and more compact structure has 11 base pairs per turn of a right-handed helix, the major groove is narrow and deep, and the minor groove is broad and shallow. Finally, the Z form is a conformation in which the backbone forms a zig-zag pattern and the DNA is more slender and elongated, with 12 base pairs per turn of a left-handed helix. The only groove that appears to any extent is the minor groove that is narrow and deep. The Z-DNA conformation is more likely to occur when there are long regions of C and G and 5-methylcytosine and G base pairs. Ultimately, genomic DNA may have different conformations in a variety of regions depending on continuously changing cellular conditions. Thus, DNA-binding proteins may or may not recognize sequences depending on the DNA conformation at that moment.

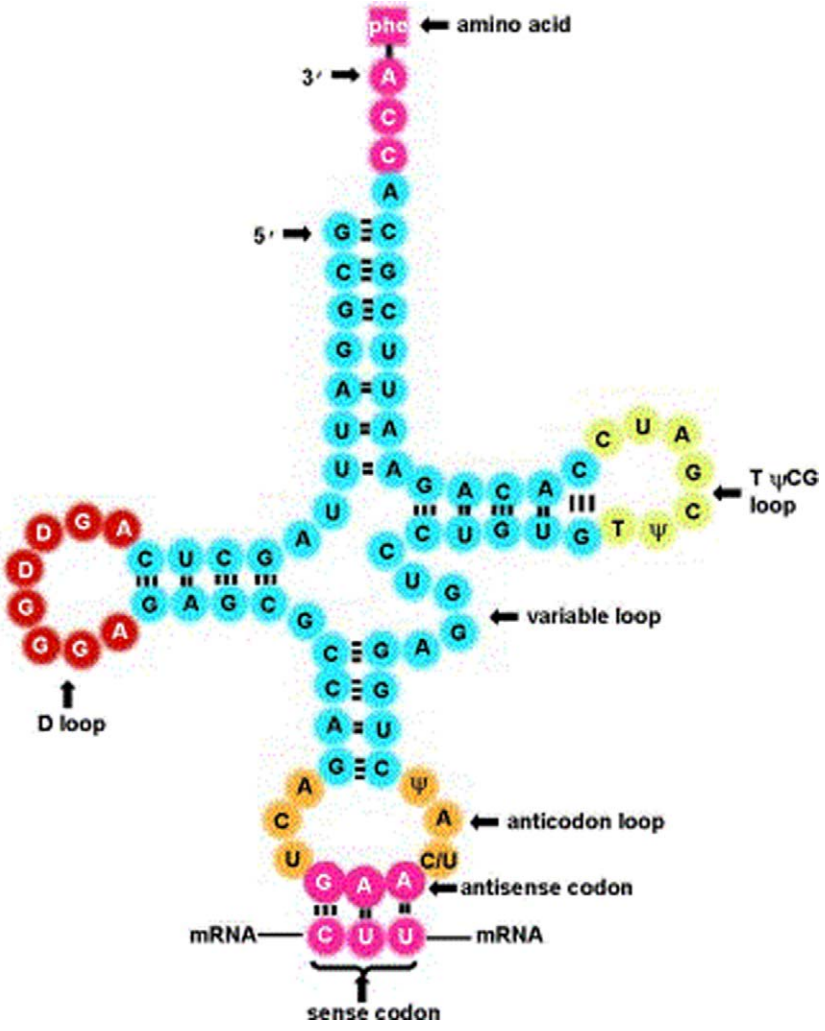

Fig. 17. Yeast phenylalanine tRN A. Intramolecular hydrogen bonding allows the formation of a cloverleaf conformation of t-RN A. Important features include the amino acid attachment site at the 3′ end and an antisense codon (GAA) that recognizes the phenylalanine sense codon (CUU) on the mRN A. Some of the modified bases are indicated as D, 5,6 dihydrouridine, and ψ, pseudouridine. tRNA molecules in general have similar sequences that result in a TψCG loop, D loop, a loop of variable size depending on the tRNA, and an anticodon loop.

RNA Species: Although a large portion of eukaryote DNA has no known function or significance, the rest of the DNA sequences contain information for either the regulation or synthesis of different RNA species. These include either non-coding RNA, such as ribosomal RNA or tRNA, or coding RNA or mRNA, which provides the code for the synthesis of proteins. The genetic information for protein synthesis is encoded in DNA by a particular combination of three nucleotides where each triplet designates a specific amino acid or signals the initiation or termination of protein synthesis. In general, one of the DNA strands specifies the sequence to code for rRNA, tRNA, or a protein (the coding or sense strand), whereas the other strand is complementary (antisense strand) and serves

as the template for the synthesis of RNA. Thus, the sequence of the RNA molecule is identical to the coding strand, with the exception that a T is replaced with a U and deoxyribose is replaced with ribose. rRNA and tRNA contain additional bases that are chemical modifications of common purines and pyrimidines (Figs. 5 and 9).

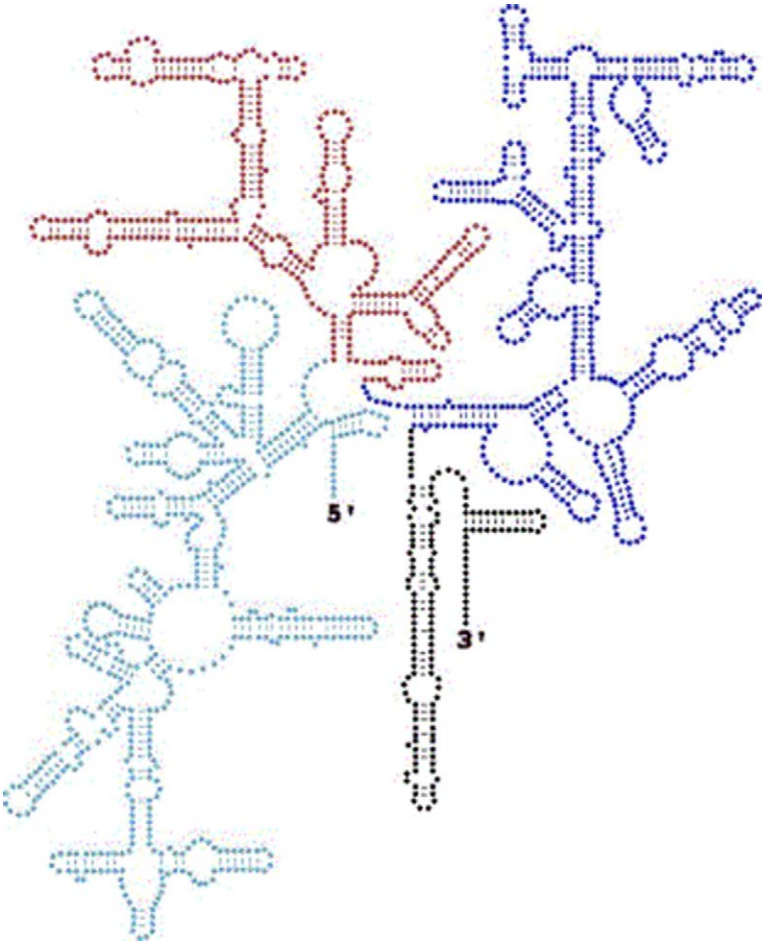

Fig. 18. E.coli 16S rRN A. Four different structural domains shown in black, dark blue, red, and light blue are shown for the secondary rRNA structure formed by intramolecular base pairing. These domains are important for the interaction of the ribosome with mRNA, tRNA, and accessory proteins and for the catalyzing translation. Specifically, the 3′ end is known to interact with mRNA during the initiation of translation.

Structure of RNA Species: In general, mRNA is a single-stranded monomer with a sequence based on nucleotides rather than base pairs. Certain regions of mRNA may form internal hydrogen bonds resulting in secondary structures that are important for regulation of the gene transcript. Secondary structures are very important for non-coding RNAs, such as tRNA and ribosomal RNA, where intramolecular hydrogen bonding forms between bases G and C, as well as others such as A and U, G and U, and modified nucleotides such as A and N^2-dimethylguanine, and G and 7-methylguanine. These regions result in a folded and more complex structure than the DNA helices and allow these molecules to perform specific functions (Figs. 17 and 18). tRNA serves as an adaptor molecule that matches the complementary nucleotide triplet to the correct amino acid.

rRNA is found in ribosomes and catalyzes the translation of RNA into protein. In eukaryotes, rRNA species are characterized by their size in Svedberg sedimentation coefficient units (S): 28S, 18S, 5.8S, and 5S, where S refers to the rate of sedimentation of the molecule during centrifugation. In prokaryotes, the rRNA species are sized at 23S, 16S, and 5 S. Other, smaller non-coding RNA species are also transcribed and include such molecules as small nuclear RNA (snRNA) or small nucleolar RNA (snoRNA), which are involved in the modification, or processing, of either mRNA or rRNA, respectively.

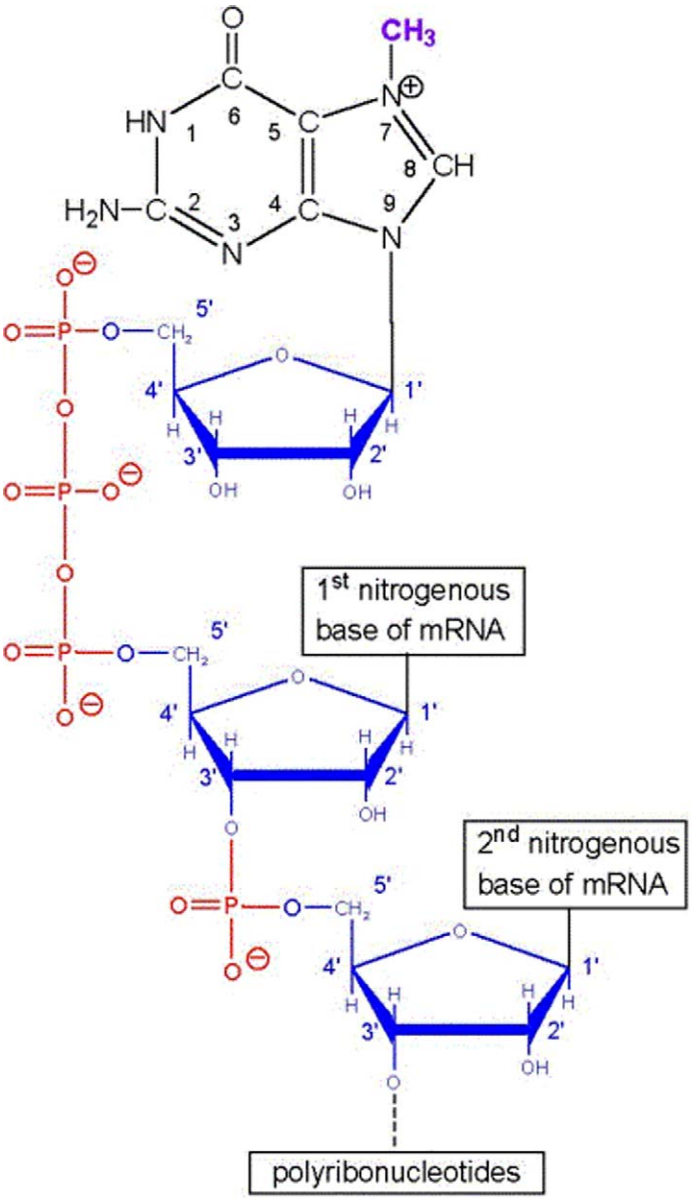

Fig. 19. mRNA cap structure. GTP is methylated and linked in a 5′ to 5′ phosphodiester linkage to the first nucleotide of mRNA.

Transcription and Processing: RNA synthesis, or transcription, is catalyzed by DNA-dependent RNA polymerases (I, II, or III) and other proteins that bind as a complex to a

specific DNA sequence, or promoter. The antisense DNA strand serves as a template for the synthesis of the RNA transcript. The promoter region may be either internal or upstream (5′) to the start site of transcription. At the start site on the DNA template, complementary RNA is synthesized by polymerizing ribonucleotide triphosphates to form phosphodiester bonds between phosphates and the ribose sugar. In prokaryotes, the coding RNA, or mRNA, serves as a direct template for protein synthesis, or translation. However, in eukaryotes, the mRNA is modified, or processed, prior to protein synthesis. The simplest processing that occurs in eukaryotes, such as yeast, includes the capping of the 5′ end of mRNA where GTP forms a 5′-5′ bond and guanine is methylated to form 7-methylguanosine (Fig. 19). The capping of the primary transcript helps in the efficient translation of the message. Further processing must occur when genes have additional or intervening sequences (introns) that are initially transcribed into RNA, forming a primary or pre-mRNA structure (Fig. 20). These introns are not translated into proteins and must be removed in the final transcript by a process known as splicing. Those nucleic acid regions that remain in the transcript are referred to as exons. Within the same gene, alternate introns might be removed such that a multitude of proteins can be formed from the same gene. As a result, one gene can code for more than one protein. tRNA and rRNA can also have introns that must be spliced out to form mature molecules. Finally, eukaryotic mRNA molecules also have approximately 250 or more adenosine monophosphate bases added to the 3′ end by poly A polymerase, an RNA polymerase that does not require a template. This process, which is known as polyadenylation, may protect the transcript from cellular degradation and help to initiate translation. During the process of translation, a ribosome complex comprised of proteins and different species of rRNA catalyzes the synthesis of proteins from the processed mRNA, tRNAs, and amino acids. Ultimately, the 5′ end of the coding region of the transcript codes for the amino terminus of the protein, while the 3′ end codes for the carboxyl terminus.

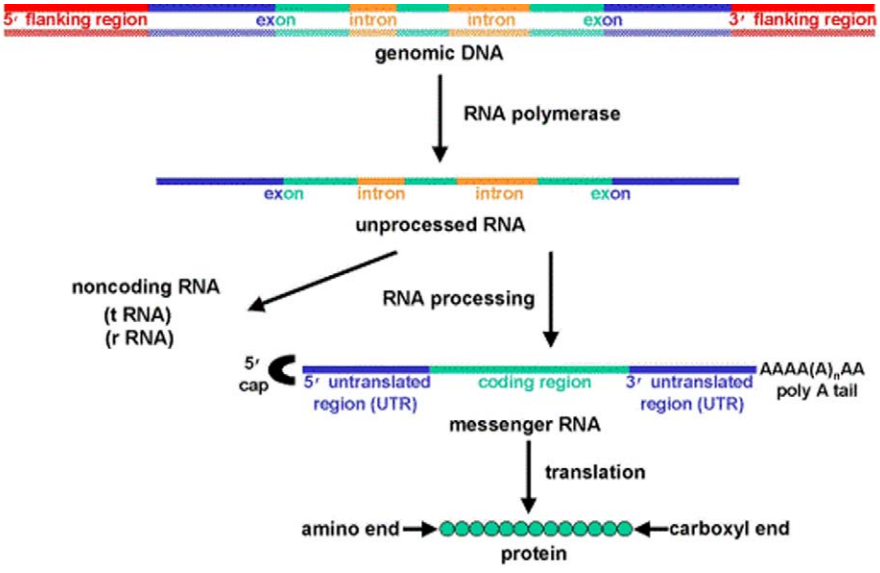

Fig. 20. Transcription of RNA and RNA processing. An unprocessed RNA molecule is transcribed from DNA by RNA polymerase. During transcription, the antisense strand (stippled genomic DNA) acts as a template and all the sequences in between the two flanking regions (red) are copied into RN A. During RNA processing, introns are removed and the remaining exons are spliced together. mRNA is further processed with the

addition of a cap and adenine nucleotides. From the final transcript, the coding region is translated into protein while the 5′ and 3′ untranslated regions are not translated.

Other Information – Web Sites

http://www.biosci.ohio-state.edu/~mgonzalez/Micro521/01.html Some historical aspects of the discovery of nucleic acids.
http://members.aol.com/logan20/nucleic.html Basic nucleic acid information including structures of coenzymes.
http://www.med.unibs.it/~marchesi/nucleic.html#intro General nucleic acid chemistry and structure of DNA helices.
http://www.bact.wisc.edu/MicrotextBook/BacterialStructure/NucleicAcids.html General information on nucleic acids.
http://cmgm.stanford.edu/biochem118/DNA-stucture.html Structure of A, B, and Z DNA, and historical perspectives of Rosalind Franklin and Watson and Crick.
http://www.colorado.edu/MCDB/MCDB1150/lectures/stahdiag.html DNA-RNA-Protein Lectures on DNA replication, transcription, and translation.
http://www.pbs.org/wgbh/nova/genome/sequencer.html Determining the genetic code of DNA by sequencing.
http://www.biology.washington.edu/fingerprint/dnaintro.html DNA fingerprinting principles.
http://www.ornl.gov/TechResources/Human_Genome/ The Human Genome Project.
http://anx12.bio.uci.edu/~hudel/bs99a/lecture22/index.html Ribosomal RNA.
http://anx12.bio.uci.edu/~hudel/bs99a/lecture21/index.html Transfer RNA.
http://anx12.bio.uci.edu/~hudel/bs99a/lecture20/index.html The Genetic Code.
http://arethusa.unh.edu/bchm752/ppthtml/april27/april27/sld003.htm Secondary structure of tRNA and the genetic code.
http://www.blc.arizona.edu/marty/411/Modules/mod15.html RNA splicing.
http://www.scri.sari.ac.uk/Biomech/Geneexp/RNAproc/Proces.htm Pre-rRNA processing and snoRNA genes.
http://molbiol.ru/eng/scripts/01_03.html Spectrophotometric measurements of nucleic acids concentrations.

Further Reading

Stryer, L., Biochemistry, 4th ed., Freeman and Company, San Francisco, CA, 1995.
Brown, T.A., Genomes 2, 2nd ed., John Wiley & Sons, New York, 2002.
Lewin, B., Genes VII, Oxford University Press, Oxford, England, 2000.
Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K. and Walter, P., Molecular Biology of the Cell, 4th ed., Garland Science, New York, 2002.
Nelson, D. L. and Cox, M.M., Lehninger Principles of Biochemistry, 3rd ed., Worth Publishers, New York, 2000.
Devlin, T. M. (ed.), Textbook of Biochemistry with Clinical Correlations, 4th ed., John Wiley & Sons, New York, 1997.
Lodish, H., Berk, A., Zipursky, S. L. Matsudaira, P., Baltimore, D. and Darnell, J., Molecular Cell Biology, 4th ed., Freeman and Company, San Francisco, CA, 2000.