CLINICAL HEMOSTASIS REVIEW

An Update on Advances and Issues in Hemostasis

BASIC MOLECULAR BIOLOGY PART I: DNA AND RNA

Shirley A. Miller BS, M.L.T. (ASCP), CLT, (NCA)

VOLUME 19, NUMBER 2
APRIL-JUNE 2005

Objective: The reader will be able to understand the basic structures of DNA, RNA, and protein.

INTRODUCTION

This article is the first in a 2005 CHR series that will provide a basic review of molecular biology. This article will briefly describe DNA (deoxyribonucleic acid), RNA (ribonucleic acid), and the resulting protein structures and functions. The following article will provide a review discussing how mutations at the DNA level are manifest clinically. A third article in this series will present current methodologies available to laboratories for genetic analysis.

Molecular testing has become an increasingly important tool in clinical medicine. The incorporation of molecular testing within routine clinical laboratories is likely to continue to grow with time. A basic understanding of molecular biology can enhance not only the quality of molecular testing in the laboratory but also its clinical relevance and use.

Molecular biology describes cellular biology at the molecular level and encompasses studies of protein structure and function. The molecular basis of cell structure and function is DNA and RNA. Cellular synthesis of protein involves the transcription of DNA sequence information into RNA, followed by translation of the RNA sequence into an amino acid chain, and ultimately protein molecules. Any change in the DNA sequence will be transcribed into an altered RNA molecule, and may subsequently be translated into an altered amino acid. A different amino acid in the sequence may have a drastic affect on protein structure and function.

DNA STRUCTURE

DNA is a double stranded helical molecule containing the code or blueprint for the synthesis of proteins. The shape of the DNA molecule can be compared to a ladder that is twisted into a spiral. The sides of the ladder consist of chains of the pentose deoxyribose bonded together by a phosphodiester bond. The phosphodiester bond links the 3rd carbon of one sugar to the 5th carbon of the next sugar. The chain can only be added to the 3-carbon end, which makes each strand directional, 5' to 3'. The two sugar-phosphate chains making up the sides of the DNA molecular ladder are aligned in opposite directions. *(See Figure 1)*. The rungs of this ladder consist of different combinations of the 4 nucleotide bases; adenine, thymine, guanine and cytosine (A, T, G and C). Cytosine and thymine are pyrimidines, which are single ring structures, while adenine and guanine are purines, double ring structures. The nucleotide A always pairs with the nucleotide T forming a double hydrogen bond. The G always pairs with the C forming a triple hydrogen bond. When one of these bases (A, T, G or C) is bonded to the sugar phosphate, it is called a deoxyribonucleotide. *(See Figure 2)*. Nucleotides are synthesized as high-energy molecules by the addition of a triphosphate and known collectively as dNTPs (deoxyribose nucleotide triphosphates.) Once incorporated into the DNA chain as A-T or G-C bonds, they are called base pairs. The addition of a dNTP to the 3' end of a DNA chain, catalysed by the enzyme DNA polymerase, is the basic reaction in which new DNA is synthesized. *(See Figure 3)*.

During the replication process the two strands of DNA separate and each strand becomes a template for the pairing of nucleotides added in the 5' to 3' direction. Each new molecule of DNA is complementary

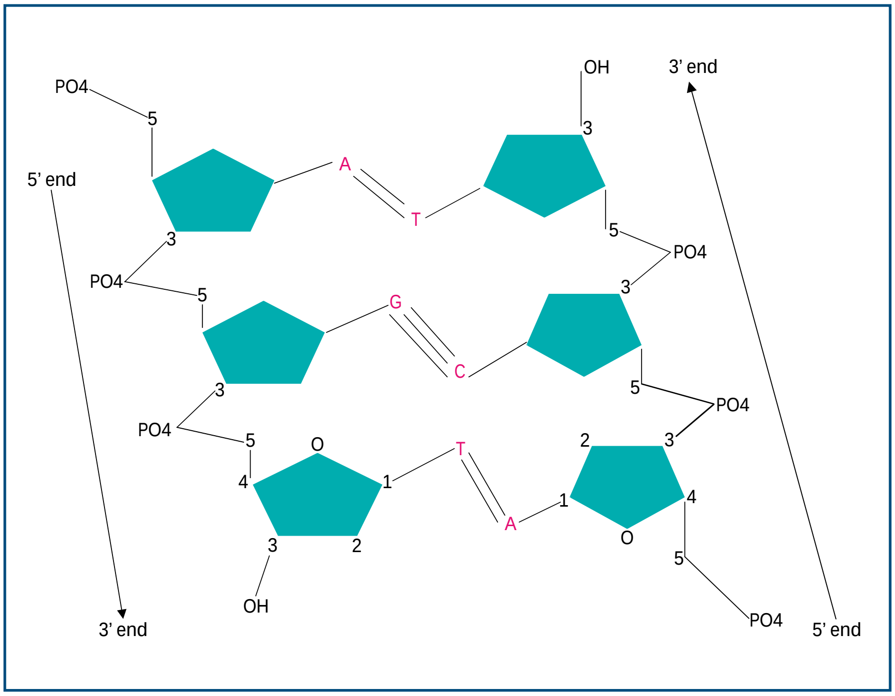

Figure 1. Double stranded DNA showing linkage, opposite directionality and base pairing

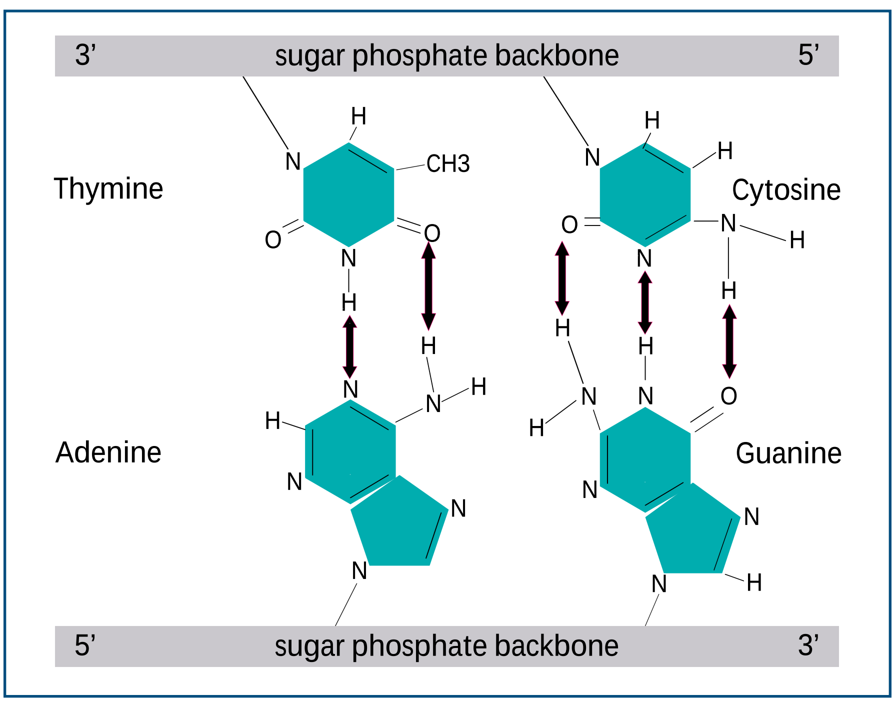

Figure 2. Base pairing of A-T and G-C bases showing hydrogen bonds.

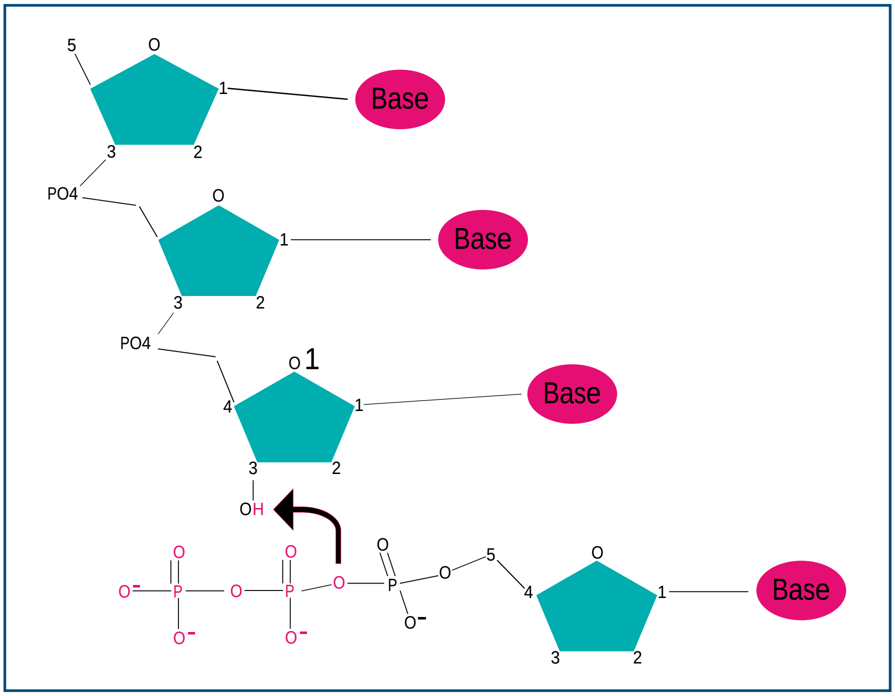

Figure 3. Incoming dNTP attaching to the 3' end of DNA chain forming a phosphodiester bond.

to the nucleotides in the parent strand. The genetic information is duplicated and two complete strands of DNA are formed with the identical sequence of the parent templates.

The human genome contains over 3 billion base pairs of DNA within the 23 chromosomes. Less than 2% of these 3 billion units actually code for proteins. The coding regions are called genes. The genes are separated into conserved domains called exons which contain the information for coding proteins, and the non-coding regions separating them are called introns or intervening sequences. A gene may have many exons and introns.

RNA STRUCTURE

DNA is informational only, and by itself, cannot make protein. Information is transcribed from DNA by another nucleic acid molecule called RNA, and this molecule directs the synthesis of protein. RNA differs from DNA in several ways. The 5-carbon sugar in the backbone of RNA is ribose instead of deoxyribose. RNA is single stranded, rather than double stranded, and contains the pyrimidine uracil instead of thymine. Therefore, the RNA chain is made up of sequences of A, G, U and C and pair up with T, C, A and G respectively on the DNA strand.

During the transcription process, the DNA molecule opens up in a small section where a specific sequence, called the promoter, signals the starting point of RNA synthesis. The enzyme RNA polymerase binds to the promoter, moves along the DNA strand in the 5' to 3' direction and catalyses the addition of rNTPs (ribose nucleotide triphosphates) one at a time to produce a strand of RNA that is complementary to the DNA sequence. When the RNA reaches a termination sequence, it detaches itself. The newly formed RNA strand is then released from the DNA template. Multiple copies of RNA can be synthesized at the same time from the open frame of DNA.

The newly transcribed RNA chain will contain both introns and exons. Since only the exons actually code for amino acids, the introns must be cut out of the RNA before it can leave the nucleus of the cell and be translated into protein. The process by which the introns are removed and the RNA resealed into an uninterrupted coding sequence is called RNA splicing. It was once thought that one gene made one protein. However, it is now known that there are more that 100,000 proteins derived from less than 25,000 genes. This is possible because of alternative splicing. The exons can be selectively spliced back together in different combinations to produce a multitude of proteins from a single gene. After the introns are removed, the spliced RNA, now referred to as mRNA (messenger RNA), is able to leave the nucleus and enter the cytoplasm of the cell.

PROTEIN SYNTHESIS

Protein synthesis is accomplished through the translation of mRNA into an amino acid sequence. The A, U, G and C sequences of the mRNA, are translated three nucleotides at a time. Each nucleotide triad, called a codon, codes for an amino acid. There are 64 possible combinations of the four bases, taken 3 at a time (4x4x4). Three of the 64 are stop codes to signal the end of translation. The remaining 61 code for amino acids. There are only 20 amino acids found in proteins. Most of the amino acids have multiple codons, except for methionine (AUG) and tryptophan (UGG), which have one codon each. When an amino acid has multiple codons, the difference between codons occurs in the third base. For example, alanine is coded by the four triplets: GCA, GCU, GCG and GCC. The first two bases of each codon determine its specificity. The third base position has less specificity and is called "wobble".

The translation of mRNA into amino acids involves two other kinds of RNA, tRNA (transfer RNA) and rRNA (ribosomal RNA). Each amino acid has its own tRNA, which is actually shaped like the letter "t". The tRNA functions as an adaptor to identify a specific mRNA codon and connect the correct amino acid to the mRNA sequence. The amino acid covalently bonds to the top arm of the tRNA, while the bottom arm of the tRNA contains the anticodon sequence which is a complementary nucleotide sequence. For example, the mRNA codon for tryptophan is UGG, so the corresponding tRNA anticodon is the complementary ACC.

The rRNA makes up subunits of a complex molecule called a ribosome. The translation process involves the ribosome binding to a start sequence on the mRNA. The mRNA sequence is translated 3 bases (one codon) at a time, as the ribosome moves in the 5' to 3' direction along the mRNA. The tRNA, with the appropriate anticodon, transfers its attached amino acid to the ribosomal complex, forming a peptide bond. This reaction is catalysed by the enzyme peptidyl transferase, which is also bound to the ribosome. After discharging its amino acid, the tRNA detaches itself, the ribosome moves to the next codon, and another tRNA adds its appropriate amino acid to the growing chain. When the ribosome reaches the stop codon, translation ceases and the newly formed polypeptide is released.

PROTEIN STRUCTURE AND FUNCTION

The original information in the in the DNA is transcribed, spliced and transported as mRNA, translated into amino acid chains, and finally processed into a protein molecule. Therefore, the primary structure of the protein depends on the sequence of amino acids as defined by the information contained in the DNA. The secondary structure of protein is defined by how the amino acid sequence is shaped and takes up space, for example, alpha-helix or beta-sheet formation. The tertiary structure refers to the folding of the protein molecule upon itself, forming disulfide bonds. This post-translational modification enables the molecule to form active sites capable of catalytic function. The quaternary structure refers to aggregation of several subunits into a functioning protein complex, like the four polypeptide subunits that make up a hemoglobin molecule.

Part 2 of this series will discuss how mutations at the DNA level can produce altered proteins with clinical manifestations.

REFERENCES:

1. *Alberts B, Bray D, Lewis J, Raff M, Roberts K, and Watson J. The Molecular Biology of the Cell. New York: Garland Publishing, Inc. 1983.*
2. *Lehninger AL. Principles of Biochemistry. New York: Worth Publishers, Inc. 1982.*
3. *Guttmacher AE, Collins FS. Genomic Medicine-A Primer. N Engl J Med , 2002 Vol 347 (19) 1512-1520.*

QUESTIONS CHR 2005 ISSUE 2

1. The portion of the gene called the exon:

a. contains the non-vital component of the genetic code
b. contains the information necessary to be translated into a protein
c. contains nonsense sequences that allow proteins to form tertiary structures
d. controls the base pair sequencing enzymes

2. RNA differs from DNA in the following ways:

a. RNA is double stranded, contains the nucleotide uracil instead of thymine and the sugar backbone ribose rather than deoxyribose
b. RNA is single stranded, contains the nucleotide uracil instead of thymine and the sugar backbone ribose rather than deoxyribose
c. RNA is single stranded, contains the nucleotide adenine instead of thymine and the sugar backbone ribose rather than deoxyribose
d. RNA is single stranded, contains the nucleotide uracil instead of thymine and the sugar backbone deoxyribose rather than ribose

3. The process by which multiple different proteins can be coded for by a single gene:

a. Occurs because the DNA is a very large molecule and codes for multiple genes in a sequential fashion
b. Involves the transcription of a common portion of the gene that is necessary to manufacture all proteins
c. Occurs through the process of RNA splicing
d. Does not occur, one gene is responsible for the production of one protein only

4. The process by which proteins are manufactured involves the following chain of events:

a. DNA transcription, RNA translation and post translational protein modification
b. RNA transcription, DNA translation and post translational protein modification
c. RNA translation, DNA transcription, and post translational protein modification
d. Post translational protein modification, DNA transcription and RNA translation

5. Which of the following facts about DNA structure are true:

a. Cytosine pairs with thymine and guanine pairs with adenine
b. DNA contains bases classified as purine and pyrosines
c. Cytosine pairs with guanine and adenine with thymine
d. Uracil is found in all DNA

CLINICAL HEMOSTASIS REVIEW
3176 S. Peoria Court, Aurora, CO 80014
www.esoterix.com

EDITOR
Dorothy M. Adcock, M.D.

EDITORIAL ASSISTANT
Monica Thibault

Clinical Hemostasis Review is published by Esoterix and is circulated to selected physicians. Copyright 2005. Esoterix is a leading laboratory services company providing esoteric testing in numerous disease corridors. The opinions expressed in the articles are those of the author(s) and do not necessarily reflect the opinions or recommendations of the advertisers, editors, or publisher.

All correspondence should be directed to the attention of the Editor, Clinical Hemostasis Review, 3176 S. Peoria Ct, Aurora, CO 80014.