GENETIC BASIS OF FUNCTION:

One of the most important concepts to understand regarding how an organism functions is that its function is an interplay of genetics and the environment. Genes are not expressed in a vacuum but in an environment. We shall learn later in this lecture that genes are expressed as proteins. Furthermore, more than one protein can be expressed from one gene. The particular protein expressed is determined by factors that include the cell's environmental conditions (e.g., temperature, humidity) and the presence of signal transducers and activators of transcription

WHAT IS THE CENTRAL DOGMA?

As previously stated, DNA is the genetic material of nearly all life, except RNA-viruses. But just how does the information in the DNA develop into visible traits? In other words, how does the DNA function? The **central dogma of molecular biology** is the concept that information flow progresses from DNA to RNA to protein but not the reverse (see Figure 3–1). Watson and Crick (who discovered the physical structure of DNA) first stated this dogma. However, current knowledge necessitates a modification of this "old" dogma.

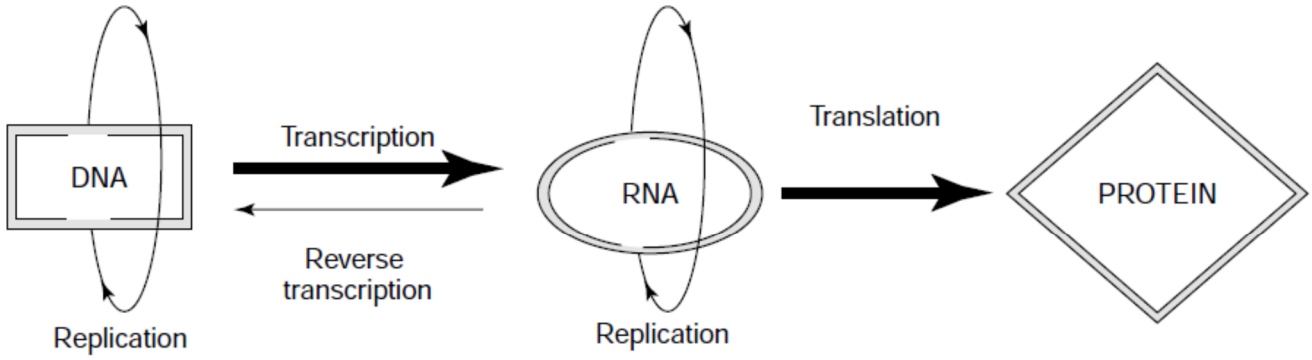

FIGURE 3–1
The central dogma of molecular biology as originally proposed by Watson and Crick.

The "new" dogma should take into account the fact that an organism's environment impacts when and how some of its genes are expressed, and also that more than one protein can be produced by a single gene.
Whereas the central dogma holds true in nature, **one of the astonishing facts of modern science is the ability of scientists to create a gene by working backward from its protein product.** The technique of reversing the central dogma is the synthesis of a **complementary DNA** (**cDNA**) using an enzyme called **reverse transcriptase**.

Three types of processes are responsible for the inheritance of genetic information and for its conversion from one form to another. These are **replication, transcription,** and **translation.** But are there significant advantages to the cell for having an intermediate between DNA and the protein it encodes? Certainly. By copying the message and taking it away to the cytoplasm where it is interpreted, the original DNA remains pristine. It is also shielded from chemicals of the cytoplasm. Another advantage to the multi-step flow of

genetic information is that the genetic information can be amplified through the production of numerous copies from just one template. Furthermore, by having multiple steps, there are more opportunities for controlling the expression of the gene under different conditions.

DNA REPLICATION:

How does DNA retain its original constitution as a cell divides and increases in number?

Replication is DNA's mode of perpetuation, whereby daughter molecules that are identical to the parent DNA molecule are created. This process of duplication fulfills the property of the genetic material to transmit information from parent to progeny. The entire genome of a cell must be replicated precisely once each time a cell undergoes **cell division.**

Replication must be executed to a high degree of fidelity in order to preserve genetic continuity between cells following cell division. To this end, the replication process incorporates editing mechanisms to correct errors. In spite of this safeguard, errors do occur, but at a low rate of 1 in 10^9 to 1 in 10^{10} base pairs.

What is the exact mechanism by which replication occurs? To duplicate a double stranded DNA molecule, the double helix first **unwinds** so that each strand serves as a template for the synthesis of its complement. This means each replicated DNA molecule would comprise one new strand and one parent strand. This mode of replication is called **semiconservative replication.**

DNA replication starts at an **origin of replication (ori)**. Once started, it continues until the entire genome has been duplicated. The unit or length of DNA that is replicated following one initiation event at one origin of replication is called a **replicon.** Whereas the bacterial chromosome constitutes one replicon, eukaryotic chromosomes contain numerous replicons (e.g., 35,000 in faba bean (*Vicia faba)*). The point on the DNA at which replication is occurring is called the **replication fork.** Once replication begins, it may proceed either in one direction (**unidirectional**) or two directions (**bidirectional**) (see Figure 3–2).

DNA replication is underlain by complex enzymatic processes involving a battery of bioactive proteins, notably **DNA polymerases I and III.** It is important to mention that DNA polymerase I and III cannot initiate synthesis of a new strand on a bare single strand.

They require a **primer** (an oligonucleotide that is hydrogen bonded to the template strand) with a 3' -OH group onto which a dNTP deoxy ribonucleotide triphosphate (a precursor molecule) can attach. Chain elongation occurs only in the 5' to 3' direction by the addition of nucleotides to the growing 3' terminus.

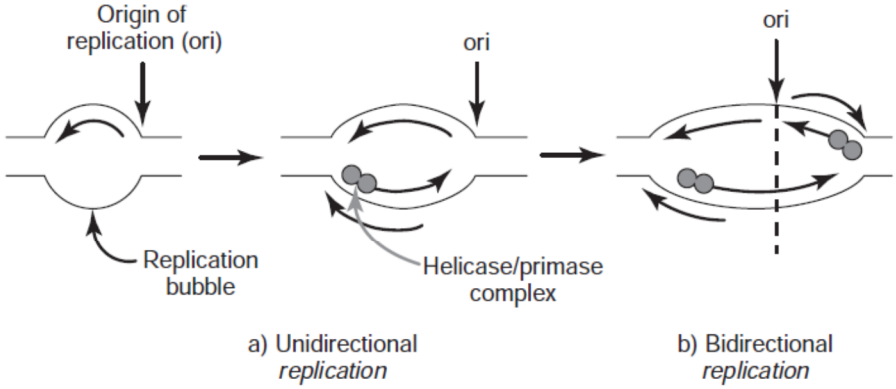

FIGURE 3–2
DNA replication starts at an origin of replication and may proceed in one or two directions on opposite sides of the origin of replication. A bubble is formed as replication continues.

DNA polymerase I has both 3' - 5' and 5' - 3' exonuclease activity, which is the ability to pause in the middle of polymerization and remove or excise nucleotides just added to the growing strand. This proofreading or editing function ensures fidelity of replication. In the event an incorrect nucleotide is added that cannot base-pair with the nucleotide in the template, polymerase I is able to remove the unpaired base. Conversely, nucleotides can be removed one at a time, from the 5' -P terminus and replaced by a process called **nick translation.**

Replication of the duplex DNA may be divided into three stages: **initiation, elongation,** and **termination.** First, the origin of replication must be recognized. The origin of replication of *E. coli* (*ori C*) consists of 245 base pairs. Before replication can start, the double helix will have to unwind and separate. The complex of proteins associated with the elongation stage of replication is called a **replisome.** As this complex moves along the DNA, the two parental strands unwind while new (daughter) strands are synthesized. As already mentioned, the DNA strands are antiparallel. Consequently, as the replication fork moves from 5' to 3' on one strand, it moves in the opposite direction (3' to 5') on the other. However, nucleic acids are synthesized in one direction only: 5' to 3'. Therefore one strand, called the **leading strand,** is synthesized continuously from 5' to 3' while the other strand, **lagging strand,** is synthesized discontinuously (see Figure 3–3). Discontinuous synthesis involves short fragments that are synthesized in the reverse direction (relative to the fork movement). These fragments are eventually connected to form a continuous strand. The fragments are called **Okazaki fragments,** after their discoverer. This mode of DNA replication is called **discontinuous replication.** Each Okazaki fragment starts with a primer. The primers are removed and replaced by appropriate nucleotides before linking. The nicks between adjacent fragments are sealed by the enzyme DNA ligase. The mode of DNA replication in prokaryotes is generally applicable to eukaryotes. However, there are significant differences stemming largely from the relatively high complexity of eukaryotic DNA organization in the chromosome, and the enormous amount of DNA per cell (50 times that in prokaryotes). As previously stated,

because of the enormous amount of DNA, eukaryotic cells have numerous initiation sites. *Drosophila* has about 3,500 replicons per genome, while mammals have about 25,000 replicons per genome, with sizes ranging between 100 to 200 kb. Another distinction between prokaryotic and eukaryotic DNA replication is that, sometimes, the replication of the leading strand is not continuous. This mode is called semi discontinuous replication.

DNA TRANSCRIPTION:

DNA replication is the duplication of a sequence or strand of DNA. Transcription is the process by which the genetic information stored or encoded in the DNA is copied, with the copy being in the form of an RNA molecule.

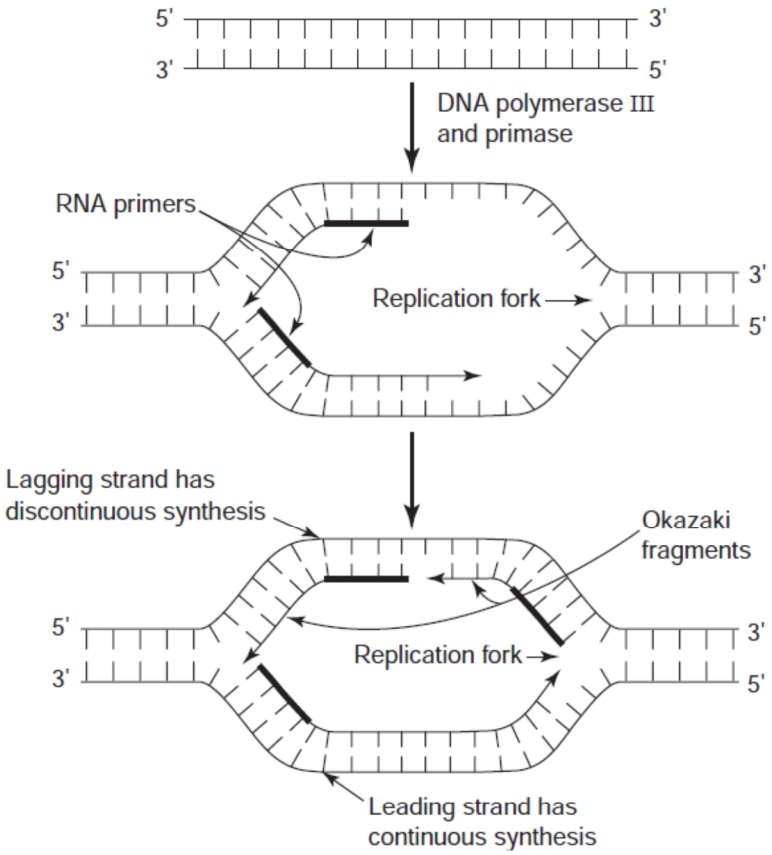

FIGURE 3–3
Replication of the two strands in each bubble proceeds in opposite directions. The leading strand is continuously extended in the direction of the replication fork, while the lagging strand is replicated in a discontinuous fashion, with each fragment (Okazaki fragment) being initiated by a primer.

Which Strand Is Transcribed?

To accomplish transcription, the RNA molecule is synthesized on a DNA template by an enzymatic process that involves several distinct events. First, transcription starts at a specific site to which the appropriate enzyme

binds. After this, polymerization is initiated, followed by chain elongation. Finally, the synthesis of the chain comes to an end (chain termination), after which the product (RNA molecule) is released. Binding sites on the DNA molecule are called promoters. These are base sequences of about 40 bp in length.

The DNA, as described previously, is double stranded. However, only one strand is a **template strand.** Nonetheless, both strands are involved in the recognition site interactions.

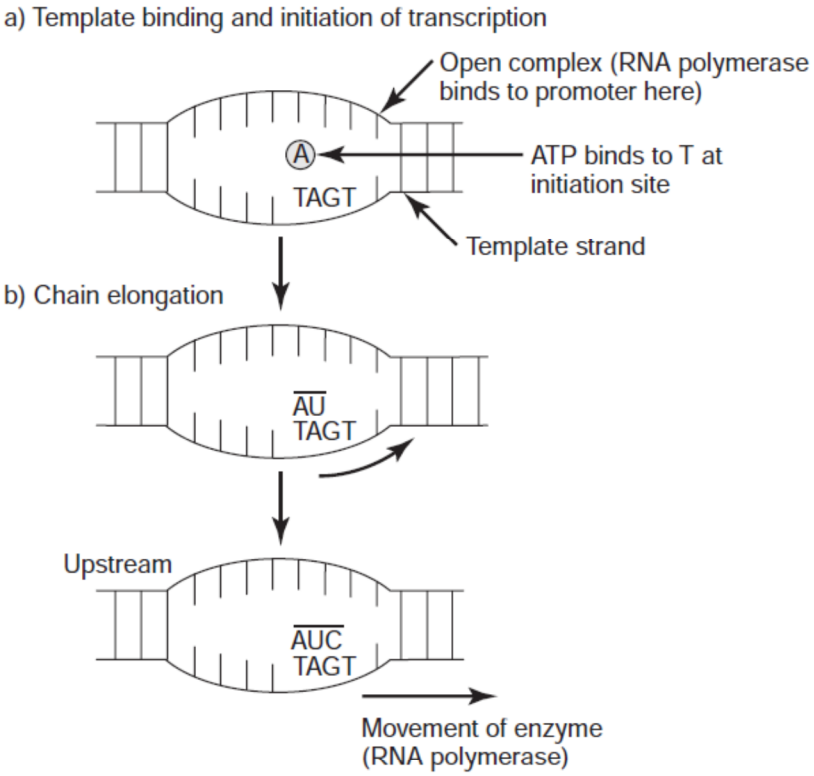

FIGURE 3–4
Transcription starts with the binding of RNA polymerase to the promoter, subsequently creating an open complex. Only one of the two strands (the template strand) is transcribed. ATP binds to T on the template strand. RNA synthesis proceeds in the direction of movement of the RNA polymerase, with complementary nucleotide triphosphates being added as the enzyme moves.

In order for the promoter to function properly, the RNA polymerase must recognize and bind to it properly. The first nucleotide triphosphate to be positioned is usually a purine (ATP or GTP). Consequently, the first base to be transcribed is usually T (thymine) or C (cytosine) (see Figure 3–4). Since only one of the two DNA strands is transcribed, the strand that has base pairs complementary to the first nucleotide triphosphate to reach the template strand (also called the **antisense strand**) is selected for transcription.

The other strand is the **sense** or **coding strand.**

The Product of Transcription:

Messenger RNA is the product of transcription, that is, the RNA synthesized from the template DNA. It is the mRNA that is decoded to determine the amino acid sequence of a DNA strand. The term **cistron** is used

to refer to a segment of DNA corresponding to a polypeptide, including the start and stop sequences. When an mRNA codes for one polypeptide, it is called a **monocistronic mRNA.** Frequently, prokaryotic mRNA is **polycistronic,** encoding several different polypeptide chains or different genes. When this happens, the polycistronic mRNA may be interspersed by sequences called **spacers.** Furthermore, prokaryotic mRNA is short lived, degrading within a few minutes after synthesis.

Transcription in eukaryotes is complicated by their genetic organization. Eukaryotic cells have compartmentalized and membrane-bound subunits called organelles, one of the most distinct being the nucleus. The DNA in the chromosomes is tightly bound to nucleoproteins (histones), forming a complex structure called **chromatin.** This structure is altered to permit the transcription of specific segments. Another significant difference is that transcription in eukaryotes occurs in the nucleus, and then the mRNA is transported out of the nucleus into the cytoplasm for translation.

Regulation of Transcription:

Two types of regulatory sequences located upstream from the point of initiation are involved in the stimulation and initiation of gene transcription in eukaryotes. Specific proteins called **transcriptional factors** that facilitate the binding of RNA polymerase II recognize these sequences (promoters and enhancers). These proteins are indispensable because, unlike prokaryotes, RNA polymerase cannot bind directly to eukaryotic promoters. Most promoters (sequences that literally promote and regulate DNA transcription or expression of the gene) have a sequence comprised of repeats of thymine (T) and adenine (A) nucleotides (called the **TATA box**) that is usually located about –25 base pairs upstream of the start point. However, many housekeeping genes do not always contain TATA boxes (or **Goldberg–Hogness boxes,** after their discoverers). There are other promoter modules such as the CCAAT and GC boxes (which contain the sequence GGGCGG). It should be mentioned that none of these sequences is uniquely essential for promoter function. However, the choice of starting point depends on the TATA box, and promoters that lack TATA boxes usually lack unique start points. Promoters and their associated transcriptional factors control the degree of transcription initiation and consequently the amount of transcription of the corresponding gene. **Enhancers** (sequences that increase the transcriptional activity of genes) interact with promoters to increase the rate of transcription initiation. Enhancers vary in their location, and may be found upstream, downstream, or even within the gene. A promoter, however, may be operationally defined as a sequence (or sequences) of DNA that must be in a relatively fixed location with regard to the start point of transcription.