

## **Chapter 22. Nucleic Acids**

### **22.1 Types of Nucleic Acids**

### **22.2 Nucleotides: Building Blocks of Nucleic Acids**

### **22.3 Primary Nucleic Acid Structure**

### **22.4 The DNA Double Helix**

### **22.5 Replication of DNA Molecules**

### **22.6 Overview of Protein Synthesis**

### **22.7 Ribonucleic Acids**

### **Chemistry at a Glance: DNA Replication**

### **22.8 Transcription: RNA Synthesis**

### **22.9 The Genetic Code**

### **22.10 Anticodons and tRNA Molecules**

### **22.11 Translation: Protein Synthesis**

### **22.12 Mutations**

### **Chemistry at a Glance: Protein Synthesis**

### **22.13 Nucleic Acids and Viruses**

### **22.14 Recombinant DNA and Genetic Engineering**

### **22.15 The Polymerase Chain Reaction**

### **22.16 DNA Sequencing**

Students should be able to:

1. Relate DNA to genes and chromosomes.
2. Describe the structure of a molecule of DNA including the base-pairing pattern.
3. Describe the structure of a nucleotide of RNA.
4. Describe the structure of a molecule of RNA.
5. Describe the three kinds of RNA and construct a pictorial representation.
6. Summarize the physiology of DNA in terms of replication and protein synthesis.
7. List the sequence of events in DNA replication and explain why it is referred to as semiconservative.
8. Evaluate the process of transcription.
9. Evaluate the process of translation.
10. Given a DNA coding strand and the genetic code, determine the complementary messenger RNA strand, the codons that would be involved in peptide formation from the messenger RNA sequence, and the amino acid sequence that would be translated.
11. Define mutation.
12. Differentiate between base substitutions and base insertions and/or deletions.
13. Discuss sickle-cell anemia.
14. Describe how viruses are referenced and categorized.
15. Define bacteriophage.
16. Describe the structure and reproductive cycle(s) of viruses.
17. Analyze the HIV virus as an example of a retrovirus.
18. Evaluate the dangers associated with emerging viruses.

## **22.6 Overview of Protein Synthesis**

### **22.7 Ribonucleic Acids**

#### **Chemistry at a Glance: DNA Replication**

### **22.8 Transcription: RNA Synthesis**

### **22.9 The Genetic Code**

### **22.10 Anticodons and tRNA Molecules**

### **22.11 Translation: Protein Synthesis**

### **22.12 Mutations**

#### **Chemistry at a Glance: Protein Synthesis**

### **22.13 Nucleic Acids and Viruses**

### **22.14 Recombinant DNA and Genetic Engineering**

### **22.15 The Polymerase Chain Reaction**

### **22.16 DNA Sequencing**

## **Introduction**

A most remarkable property of living cells is their ability to produce exact replicas of themselves. This is due to the cells containing fact that all the instructions needed for making the complete organism of which they are a part. Nucleic acids are the molecules within a cell that are responsible for these amazing capabilities.

The first isolation of nucleic acid we now refer to as DNA was accomplished by Swiss physiologist Johann Friedrich Miescher circa 1870 while studying the nuclei of white blood cells. In the 1920's nucleic acids were found to be major components of chromosomes, small gene-carrying bodies in the nuclei of complex cells. Elemental analysis of nucleic acids showed the presence of phosphorus, in addition to the usual C, H, N & O. We now know that nucleic acids are found throughout a cell, not just in the nucleus, the name nucleic acid is still used for such materials.

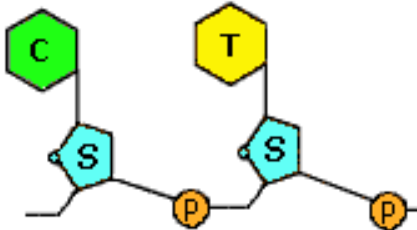
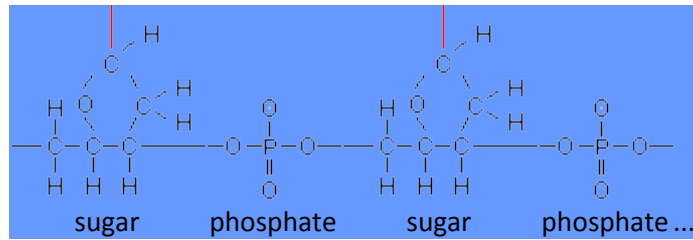
A nucleic acid is a polymer in which the monomer units are nucleotides. There are two Types of Nucleic Acids:

**DNA: Deoxyribonucleic Acid:** Found within cell nucleus for storing and transferring of genetic information that are passed from one cell to other during cell division

**RNA: Ribonucleic Acid:** Occurs in all parts of cell serving the primary function is to synthesize the proteins needed for cell functions.

## **22.1 Types of Nucleic Acids**

The nucleic acids are very large molecules that have two main parts. The backbone of a nucleic acid is made of alternating sugar and phosphate molecules bonded together in a long chain, represented below:



Each of the sugar groups in the backbone is attached (via the bond shown in red) to a third type of molecule called a nucleotide base. There are only four different nucleotide bases can occur in a nucleic acid and are classified as pyrimidine or purine bases:

### Components of Nucleic Acids

	DNA only	DNA & RNA		RNA only	
Nitrogen Bases	 Thymine	 Adenine	 Guanine	 Cytosine	 Uracil
Sugars & Phosphate	 2-Deoxyribose	 Phosphate		 Ribose	

Though only four different nucleotide bases can occur in a nucleic acid, each nucleic acid contains millions of bases bonded to it. The order in which these nucleotide bases appear in the nucleic acid is the coding for the information carried in the molecule. In other words, the nucleotide bases serve as a sort of genetic alphabet on which the structure of each protein in our bodies is encoded.

## DNA

In most living organisms (except for viruses), genetic information is stored in the molecule deoxyribonucleic acid, or DNA. DNA is made and resides in the nucleus of living cells. DNA gets its name from the sugar molecule contained in its backbone (deoxyribose); however, it gets its significance from its unique structure. Four different nucleotide bases occur in DNA: adenine (A), cytosine (C), guanine (G), and thymine (T).

### 22.2 Nucleotides: Building Blocks of Nucleic Acids

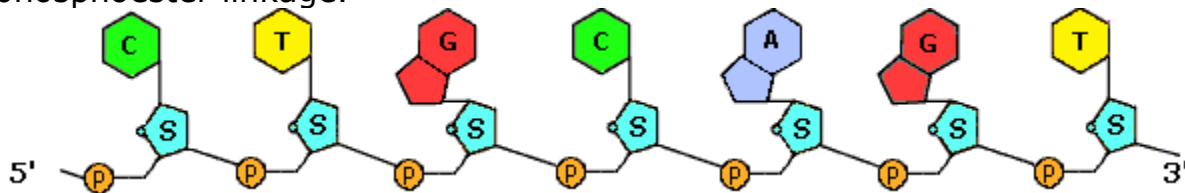
Names of DNA Nucleotide			
Name	Base	Nucleoside	5'-Nucleotide
DAMP	Adenine	2'-Deoxyadenosine	2'-Deoxyadenosine-5'-monophosphate
DCMP	Cytosine	2'-Deoxycytidine	2'-Deoxycytidine-5'-monophosphate
DGMP	Guanine	2'-Deoxyguanosine	2'-Deoxyguanosine-5'-monophosphate
DTMP	Thymine	2'-Deoxythymidine	2'-Deoxythymidine-5'-monophosphate

**RNA** has the same nucleotide structure except the thymine base is replaced by uracil.

### 22.3 Primary Nucleic Acid Structure

#### Polynucleotides

In polynucleotides, nucleotides are joined to one another by covalent bonds between the phosphate of one and the sugar of another. These linkages are called **phosphodiester** linkages. This nucleic acid found in the cell has primary structures that arise from the end-to-end polymerization of single nucleotide units. The links between each nucleotide are formed by esterification reactions between the sugar's C3' hydroxyl group and the phosphate of an incoming nucleoside triphosphate (NTP) to form a phosphoester linkage.



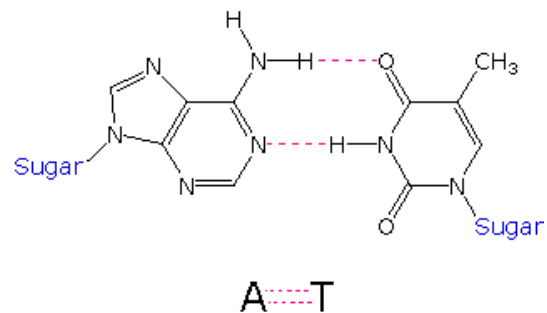
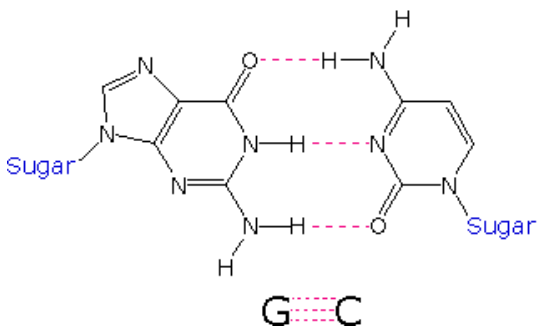
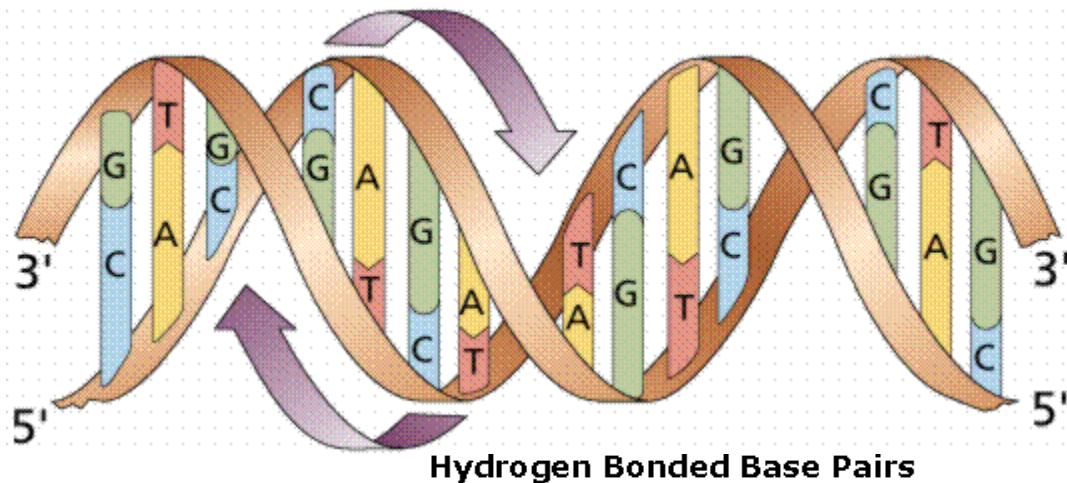
The sugar is ribose in the case of RNA, deoxyribose in DNA. This polymerization process leaves a free hydroxyl on the incoming nucleotide (on the 3' C of the sugar) to serve for the next reaction in chain elongation.

### 22.4 The DNA Double Helix

The 1962 Nobel Prize in Physiology or Medicine was awarded to Crick, Watson and Wilkins for the discovery of the molecular structure of DNA – the double helix.

### Chemical Structure of the DNA double strands

DNA (deoxyribonucleic acid) is a double-stranded molecule that is twisted into a helix like a spiral staircase. Each strand is comprised of a sugar-phosphate backbone and numerous base chemicals attached in pairs.



The four bases that make up the stairs in the spiraling staircase are adenine (A), thymine (T), cytosine (C) and guanine (G). These stairs act as the "letters" in the genetic alphabet, combining into complex sequences to form the words, sentences and paragraphs that act as instructions to guide the formation and functioning of the host cell. Maybe even more appropriately, the A, T, C and G in the genetic code of the DNA molecule can be compared to the "0" and "1" in the binary code of computer software. Like software to a computer, the DNA code is a genetic language that communicates information to the organic cell.

### Genetic code

The DNA code, like a floppy disk of binary code, is quite simple in its basic paired structure. However, it's the sequencing and functioning of that code that's enormously complex. Through recent technologies like x-ray crystallography, we now know that the cell is not a "blob of protoplasm", but

rather a microscopic marvel that is more complex than the space shuttle. The cell is very complicated, using vast numbers of phenomenally precise DNA instructions to control its every function.

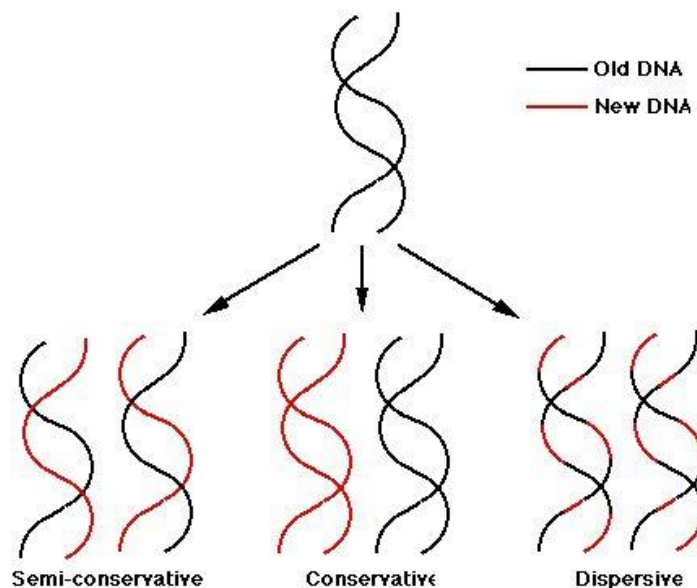
## 22.5 Replication of DNA Molecules

Before a cell divides, its DNA is replicated (duplicated.) Because the two strands of a DNA molecule have complementary base pairs, the nucleotide sequence of each strand automatically supplies the information needed to produce its partner. If the two strands of a DNA molecule are separated, each can be used as a pattern or template to produce a complementary strand. Each template and its new complement together then form a new DNA double helix, identical to the original.

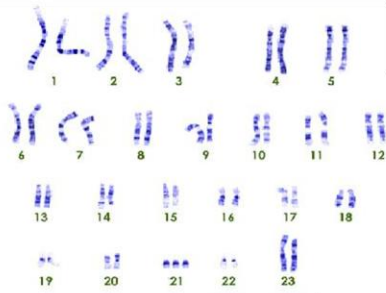
Before replication can occur, the length of the DNA double helix about to be copied must be **unwound**. In addition, the two strands must be **separated**, much like the two sides of a zipper, by breaking the weak hydrogen bonds that link the paired bases. Once the DNA strands have been unwound, they must be **held apart** to expose the bases so that new nucleotide partners can hydrogen-bond to them.

The enzyme **DNA polymerase** then moves along the exposed DNA strand, **joining newly arrived nucleotides** into a new DNA strand that is complementary to the template.

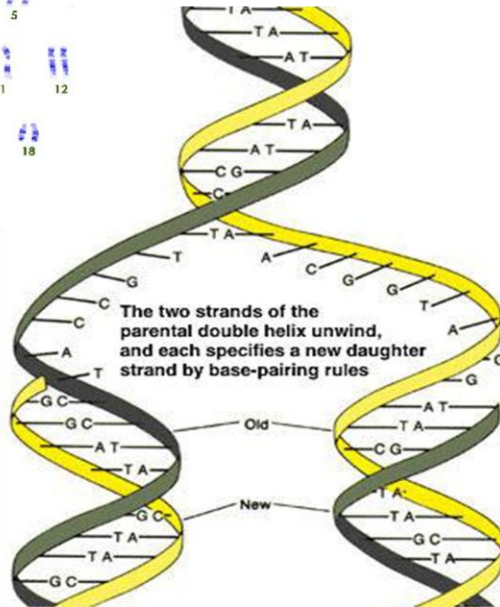
Each cell contains a **family of more than thirty enzymes** to insure the accurate replication of DNA.



# Replications



Human Chromosomes  
**Eukaryotic**



Plasmid DNA

**Prokaryotic**

## 22.6 Overview of Protein Synthesis

**DNA code:** containing specific base code is used to create a specific polypeptide—the protein containing a certain sequence of amino acids.

The genetic code present in the DNA and later transcribed into mRNA consists of **64 triplets of nucleotides**. These triplets are called **codons**. With three exceptions, each codon encodes for one of the **20 amino acids** used in the synthesis of proteins. That produces some redundancy in the code: most of the amino acids being encoded by more than one codon.

**RNA polymerase:** RNA polymerases are enzyme complexes that synthesize mRNA molecules using DNA as a template, in the process known as transcription.

Protein synthesis can be divided into two parts:

### 1. **Transcription**

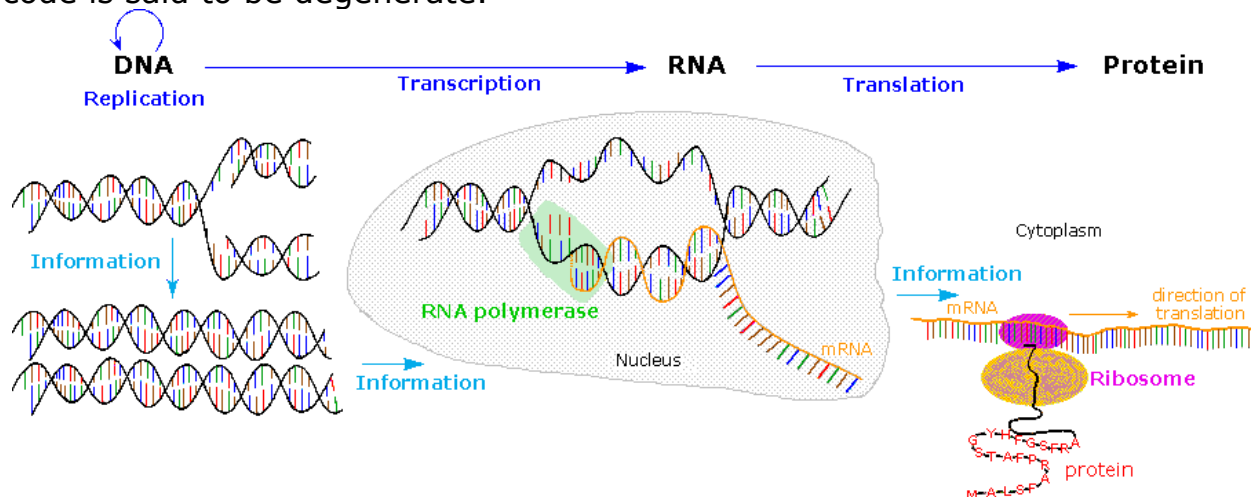
Before the synthesis of a protein begins, the corresponding RNA molecule is produced by **RNA transcription**. One strand of the DNA double helix is used as a template by the **RNA polymerase** to synthesize a **messenger RNA** (mRNA). This **mRNA** migrates from the nucleus to the cytoplasm. During this step, mRNA goes through different types of maturation including one

called **splicing** when the non-coding sequences are eliminated. The coding mRNA sequence can be described as a unit of three nucleotides called a **codon**.

## 2. Translation

The **ribosome** binds to the **mRNA** at the start codon (AUG) that is recognized only by the initiator **tRNA**. The ribosome proceeds to the elongation phase of protein synthesis. During this stage, complexes, composed of an amino acid linked to tRNA, sequentially bind to the appropriate codon in mRNA by forming complementary base pairs with the **tRNA anticodon**. The ribosome moves from codon to codon along the mRNA. Amino acids are added one by one, translated into polypeptidic sequences dictated by DNA and represented by mRNA. At the end, a release factor binds to the stop codon, terminating translation and releasing the complete **polypeptide** from the ribosome.

One specific amino acid can correspond to more than one codon. The genetic code is said to be degenerate.



## 22.7 Ribonucleic Acids

One of the two main types of nucleic acid (the other being DNA), which functions in cellular protein synthesis in all living cells. Like DNA, it consists of strands of repeating nucleotides joined in chainlike fashion, but the **strands are single** and it has the nucleotide **uracil (U)** where DNA has **thymine (T)**.

### Messenger RNA

Whereas most types of RNA are the final products of their genes, mRNA is an intermediate in information transfer. It carries information from DNA to the ribosome in a genetic code that the protein-synthesizing machinery translates into protein. Specifically, mRNA sequence is recognized in a



sequential fashion as a series of nucleotide triplets by tRNAs via base pairing to the three-nucleotide anticodons in the tRNAs. There are specific triplet codons that specify the beginning and end of the protein-coding sequence. Thus, the function of mRNA involves the reading of its primary nucleotide sequence, rather than the activity of its overall structure. Messenger RNAs are typically shorter-lived than the more stable structural RNAs, such as tRNA and rRNA. See Genetic code

### **Small nuclear RNA**

Small RNAs, generally less than 300 nucleotides long and rich in uridine (U), are localized in the nucleoplasm (snRNAs) and nucleolus (snoRNAs) of eukaryotic cells. There they take part in RNA processing, such as intron removal during eukaryotic mRNA splicing and posttranscriptional modification that occurs during production of mature rRNA. See Intron

### **Catalytic RNA**

RNA enzymes, or ribozymes, are able to catalyze specific cleavage or joining reactions either in themselves or in other molecules of nucleic acid. See Catalysis, Ribozyme

## **22.8 Transcription: RNA Synthesis**

Transcription is the process of creating an equivalent RNA copy of a sequence of DNA in double helix. Both RNA and DNA have base pairs of nucleotides as a complementary language that can be converted back and forth from DNA to RNA in the presence of the correct enzymes, RNA polymerase. During transcription, a DNA sequence is read by RNA polymerase, which produces a complementary, antiparallel RNA strand. As opposed to DNA replication, transcription results in an RNA complement that includes uracil (U) in all instances where thymine (T) would have occurred in a DNA complement.

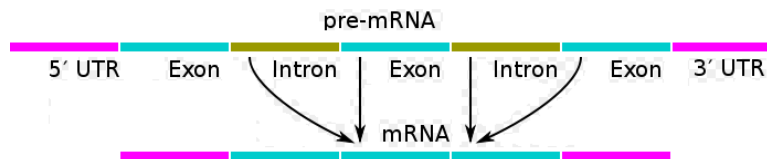
Transcription is the first step leading to gene expression. The stretch of DNA transcribed into an RNA molecule is called a transcription unit and encodes at least one gene. If the gene transcribed encodes for a protein, the result of transcription is messenger RNA (mRNA), which will then be used to create that protein via the process of translation. Alternatively, the transcribed gene may encode for either ribosomal RNA (rRNA) or transfer RNA (tRNA), other components of the protein-assembly process, or other ribozymes.

A DNA transcription unit encoding for a protein contains not only the sequence that will eventually be directly translated into the protein (the coding sequence) but also regulatory sequences that direct and regulate the synthesis of that protein. The regulatory sequence before (upstream from) the coding sequence is called the five prime untranslated region (5'UTR), and the sequence following (downstream from) the coding sequence is called the three prime untranslated region (3'UTR).

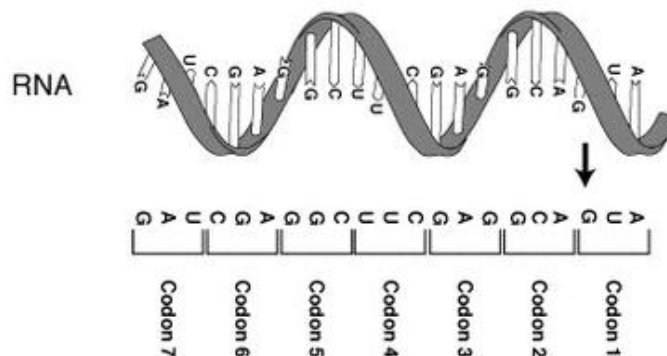
Transcription has some proofreading mechanisms, but they are fewer and less effective than the controls for copying DNA; therefore, transcription has a lower copying fidelity than DNA replication.

As in DNA replication, DNA is read from 3' → 5' during transcription. Meanwhile, the complementary RNA is created from the 5' → 3' direction. Although DNA is arranged as two antiparallel strands in a double helix, only one of the two DNA strands, called the template strand, is used for transcription. This is because RNA is only single-stranded, as opposed to double-stranded DNA. The other DNA strand is called the coding strand, because its sequence is the same as the newly created RNA transcript (except for the substitution of uracil for thymine). The use of only the 3' → 5' strand eliminates the need for the Okazaki fragments seen in DNA replication. Transcription is divided into 5 stages: pre-initiation, initiation, promoter clearance, elongation and termination.

**Splicing of mRNA:** Splicing is a modification of an RNA after transcription, in which **introns** (nonessential part of the code) are removed and **exons** (essential part of the code) are joined. Also the **UTRs**, non-coding parts of exons at the ends of the mRNA is also removed.



**Codon:** The code is defined as a mapping of three-nucleotide base sequences and the amino acids. A triplet codon in a nucleic acid sequence usually specifies a single amino acid (though in some cases the same codon triplet in different locations can code unambiguously for two different amino acids).



## 22.9 The Genetic Code

**DNA code:** containing specific base code to create a specific polypeptide- the protein containing a certain sequence of amino acids.

The genetic code consists of **64 triplets of nucleotides**. These triplets are called **codons**. With three exceptions, each codon encodes for one of the 20 amino acids used in the synthesis of proteins. That produces some redundancy in the code: most of the amino acids being encoded by more than one codon.

One codon, **AUG** serves two related functions:

- it signals the start of translation
- it codes for the incorporation of the amino acid methionine (Met) into the growing polypeptide chain

The genetic code can be expressed as either **RNA codons** or **DNA codons**. RNA codons occur in messenger RNA (**mRNA**) and are the codons that are actually "read" during the synthesis of polypeptides during **translation**. But each mRNA molecule acquires its sequence of nucleotides by **transcription** from the corresponding gene. Because DNA sequencing has become so rapid and because most genes are now being discovered at the level of DNA before they are discovered as mRNA or as a protein product, it is extremely useful to have a table of codons expressed as DNA. So here are both.

Note that for each table, the left-hand column gives the first nucleotide of the codon, the 4 middle columns give the second nucleotide, and the last column gives the third nucleotide.

### The RNA Codons

		Second nucleotide				
		<b>U</b>	<b>C</b>	<b>A</b>	<b>G</b>	
<b>U</b>	UUU <b>Phenylalanine</b> (Phe)	UCU <b>Serine</b> (Ser)	UAU <b>Tyrosine</b> (Tyr)	UGU <b>Cysteine</b> (Cys)	<b>U</b>	
	UUC Phe	UCC Ser	UAC Tyr	UGC Cys	<b>C</b>	
	UUA <b>Leucine</b> (Leu)	UCA Ser	UAA <b>STOP</b>	UGA <b>STOP</b>	<b>A</b>	
	UUG Leu	UCG Ser	UAG <b>STOP</b>	UGG <b>Tryptophan</b> (Trp)	<b>G</b>	
<b>C</b>	CUU <b>Leucine</b> (Leu)	CCU <b>Proline</b> (Pro)	CAU <b>Histidine</b> (His)	CGU <b>Arginine</b> (Arg)	<b>U</b>	

	CUC Leu	CCC Pro	CAC His	CGC Arg	<b>C</b>
	CUA Leu	CCA Pro	CAA <b>Glutamine</b> (Gln)	CGA Arg	<b>A</b>
	CUG Leu	CCG Pro	CAG Gln	CGG Arg	<b>G</b>
<b>A</b>	AUU <b>Isoleucine</b> (Ile)	ACU <b>Threonine</b> (Thr)	AAU <b>Asparagine</b> (Asn)	AGU <b>Serine</b> (Ser)	<b>U</b>
	AUC Ile	ACC Thr	AAC Asn	AGC Ser	<b>C</b>
	AUA Ile	ACA Thr	AAA <b>Lysine</b> (Lys)	AGA <b>Arginine</b> (Arg)	<b>A</b>
	AUG <b>Methionine</b> (Met) or <b>START</b>	ACG Thr	AAG Lys	AGG Arg	<b>G</b>
<b>G</b>	GUU <b>Valine</b> Val	GCU <b>Alanine</b> (Ala)	GAU <b>Aspartic acid</b> (Asp)	GGU <b>Glycine</b> (Gly)	<b>U</b>
	GUC (Val)	GCC Ala	GAC Asp	GGC Gly	<b>C</b>
	GUA Val	GCA Ala	GAA <b>Glutamic acid</b> (Glu)	GGA Gly	<b>A</b>
	GUG Val	GCG Ala	GAG Glu	GGG Gly	<b>G</b>

### Genome:

the **genome** is the entirety of an organism's hereditary information. It is encoded either in DNA or, for many types of virus, in RNA.

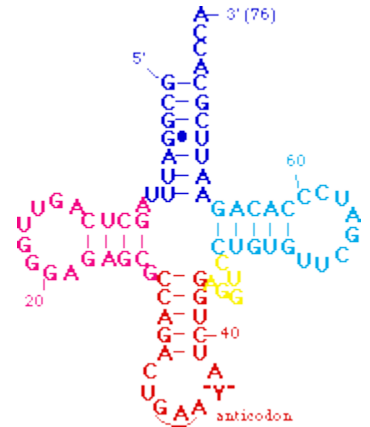
**The Human Genome Project** - the entire human genome is currently being decoded by labs around the world. The project, which started in 1990, aims to have the complete 3.2 billion base pair genome completed in a high quality form in 2003, at a final cost of over 3 billion dollars. Recently (1998) a private company, Celera Genomics, has amassed enough high speed automated DNA sequencers and computing power (second only to the Pentagon)

## 22.10 Anticodons and tRNA Molecules

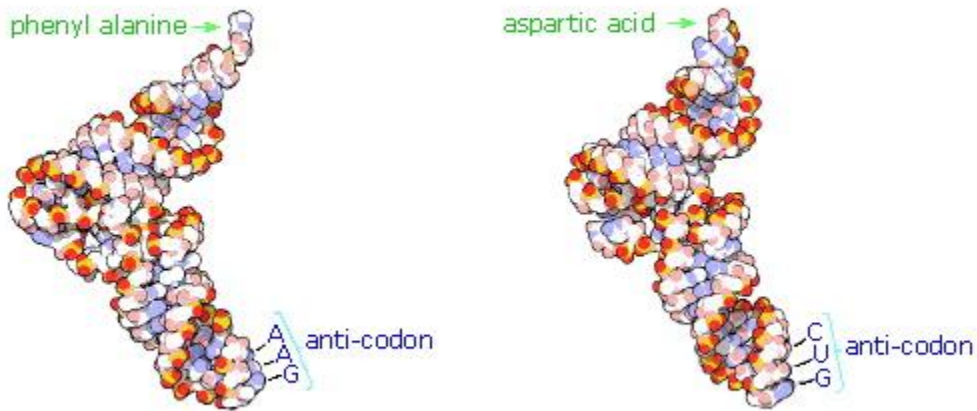
These small RNAs (70–90 nucleotides) that act as adapters to translate the nucleotide sequence of mRNA into protein sequence. They do this by carrying the appropriate amino acid to the ribosome during the process of protein synthesis. Each cell contains at least one type of tRNA specific for each of the 20 amino acids, and usually several types. The base sequence in the mRNA directs the appropriate amino acid-carrying tRNAs to the ribosome to ensure that the correct protein sequence is made.

## tRNA Secondary Structure

The translation process is fundamentally straightforward. The **mRNA** strand bearing the transcribed code for synthesis of a protein interacts with relatively small RNA molecules (about 70-nucleotides) to which individual amino acids have been attached by an ester bond at the 3'-end. These **transfer RNA's** (tRNA) have distinctive three-dimensional structures consisting of loops of single-stranded RNA connected by double stranded segments. This cloverleaf secondary structure is further wrapped into an "L-shaped" assembly, having the amino acid at the end of one arm, and a characteristic **anti-codon** region at the other end. The anti-codon consists of a nucleotide triplet that is the complement of the amino acid's codon(s). Models of two such tRNA molecules are shown to the right. When read from the top to the bottom, the **anti-codons** depicted here should complement a codon in the previous table.



### Transfer RNA Molecules



## 22.11 Translation: Protein Synthesis

This process involves following components:

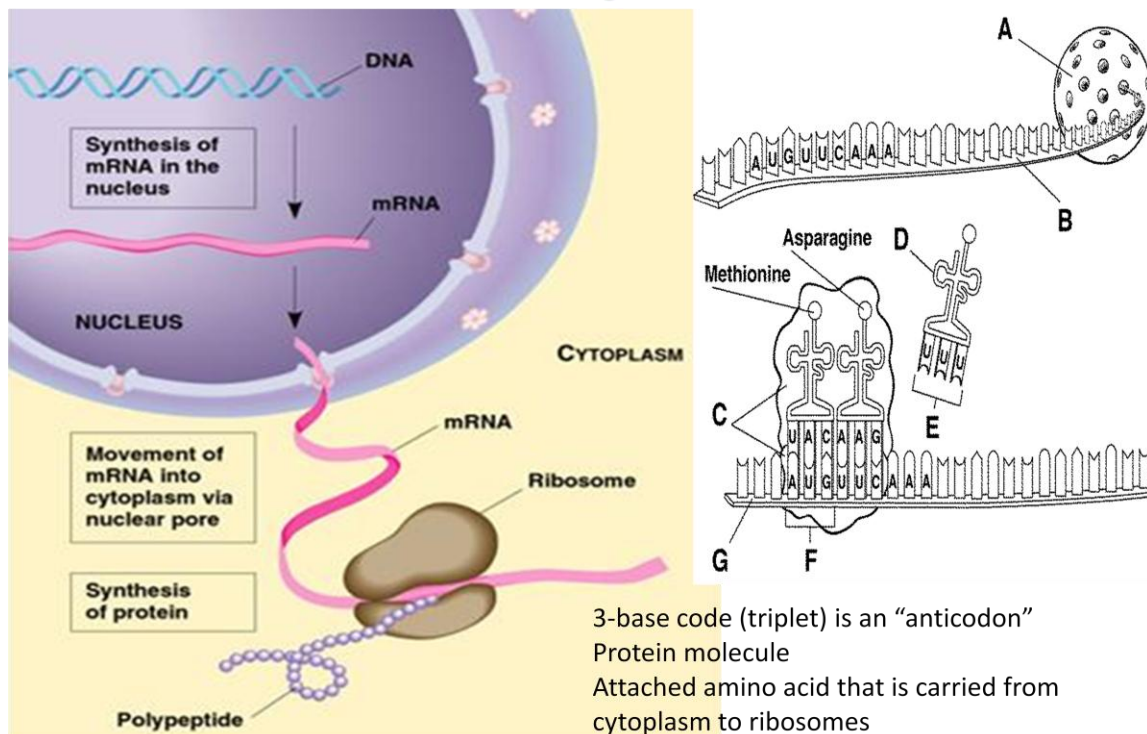
**mRNA:** Messenger RNA which is a single stranded copy of a DNA double helix base pair sequence with uracil in the places where thymine was.

## Ribosome

Ribosomes are the components of cells that make proteins from amino acids. Ribosomes then read the information in the mRNA and use the codons to produce proteins.

**tRNA:** The genetic code is read during translation using transfer-RNA, tRNAs that have 3-base anticodons complementary to codons in mRNA.

# Protein Synthesis



## The Steps of Translation

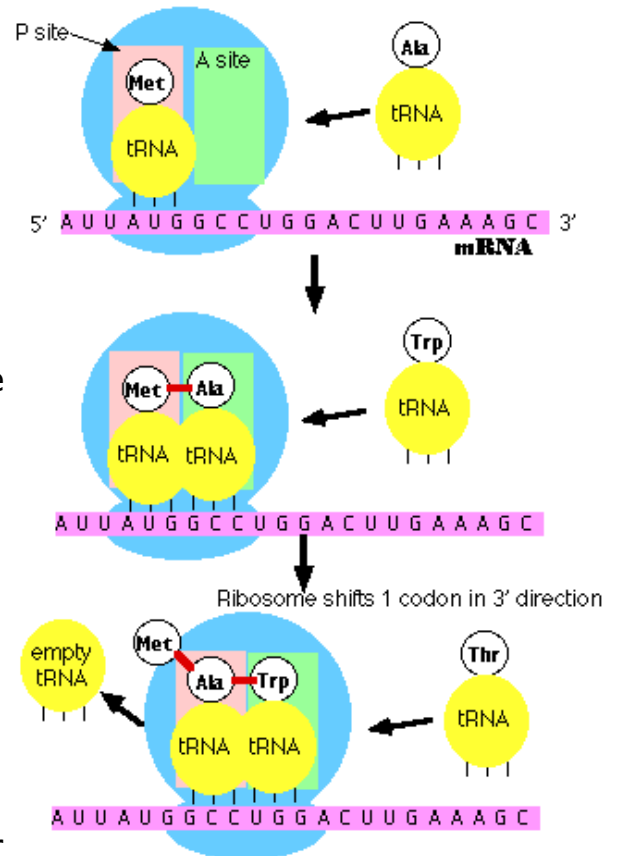
### 1. Initiation

- The **small subunit** of the ribosome binds to a site "upstream" (on the 5' side) of the start of the message.
- It proceeds downstream (5' → 3') until it encounters the start codon **AUG**. (The region between the cap and the AUG is known as the 5'-untranslated region [**5'-UTR**].)
- Here it is joined by the **large subunit** and a special **initiator tRNA**.
- The initiator tRNA binds to the **P site** (shown in pink) on the ribosome.

- In eukaryotes, initiator tRNA carries **methionine** (Met). (Bacteria use a modified methionine designated **fMet**.)

## 2. Elongation

- An **aminoacyl-tRNA** (a tRNA covalently bound to its amino acid) able to base pair with the next codon on the mRNA arrives at the **A site** (green) associated with:
  - an **elongation factor** (called EF-Tu in bacteria)
  - GTP** (the source of the needed energy)
- The preceding amino acid (Met at the start of translation) is covalently linked to the incoming amino acid with a peptide bond (shown in red).
- The initiator tRNA is released from the P site.
- The ribosome moves one codon downstream.
- This shifts the more recently-arrived tRNA, with its attached peptide, to the P site and opens the A site for the arrival of a new aminoacyl-tRNA.
- This last step is promoted by another protein **elongation factor** (called EF-G in bacteria) and the energy of another molecule of **GTP**.



Note: the initiator tRNA is the only member of the tRNA family that can bind directly to the P site. The P site is so-named because, with the exception of initiator tRNA, it binds only to a **peptidyl-tRNA** molecule; that is, a tRNA with the growing peptide attached.

The A site is so-named because it binds only to the incoming **aminoacyl-tRNA**; that is the tRNA bringing the next amino acid. So, for example, the tRNA that brings Met into the interior of the polypeptide can bind only to the A site.

### 3. Termination

- The end of translation occurs when the ribosome reaches one or more **STOP** codons (**UAA, UAG, UGA**). (The nucleotides from this point to the poly(A) tail make up the 3'-untranslated region [**3'-UTR**] of the mRNA.)
- There are no tRNA molecules with anticodons for STOP codons.
- However, protein release factors recognize these codons when they arrive at the A site.
- Binding of these proteins —along with a molecule of **GTP** — releases the polypeptide from the ribosome.
- The ribosome splits into its subunits, which can later be reassembled for another round of protein synthesis.

### 22.12 Mutations

A mutation is a permanent change in the DNA sequence of a gene or the genetic code. Mutations in a gene's DNA sequence can alter the amino acid sequence of the protein encoded by the gene.

How does this happen? Like words in a sentence, the DNA sequence of each gene determines the amino acid sequence for the protein it encodes. The DNA sequence is interpreted in groups of three nucleotide bases, called codons. Each codon specifies a single amino acid in a protein.

Mutations can lead to changes in the structure of an encoded protein or to a decrease or complete loss in its expression. Because a change in the DNA sequence affects all copies of the encoded protein, mutations can be particularly damaging to a cell or organism. In contrast, any alterations in the sequences of RNA or protein molecules that occur during their synthesis are less serious because many copies of each RNA and protein are synthesized.

Geneticists often distinguish between the **genotype** and **phenotype** of an organism. Strictly speaking, the entire set of genes carried by an individual is its genotype, whereas the function and physical appearance of an individual is referred to as its phenotype.

### Chemistry at a Glance: Protein Synthesis

#### 22.13 Nucleic Acids and Viruses

All viruses have genes made from either **DNA** or **RNA**, long molecules that carry genetic information; all have a protein coat that protects these genes;



and some have an envelope of fat that surrounds them when they are outside a cell. **Viroids** do not have a protein coat and **prions** contain no RNA or DNA. Viruses vary from simple helical and icosahedral shapes, to more complex structures. Most viruses are about one hundred times smaller than an average bacterium. The origins of viruses in the evolutionary history of life are unclear: some may have evolved from **plasmids**—pieces of DNA that can move between cells—while others may have evolved from bacteria. In evolution, viruses are an important means of **horizontal gene transfer**, which increases genetic diversity.

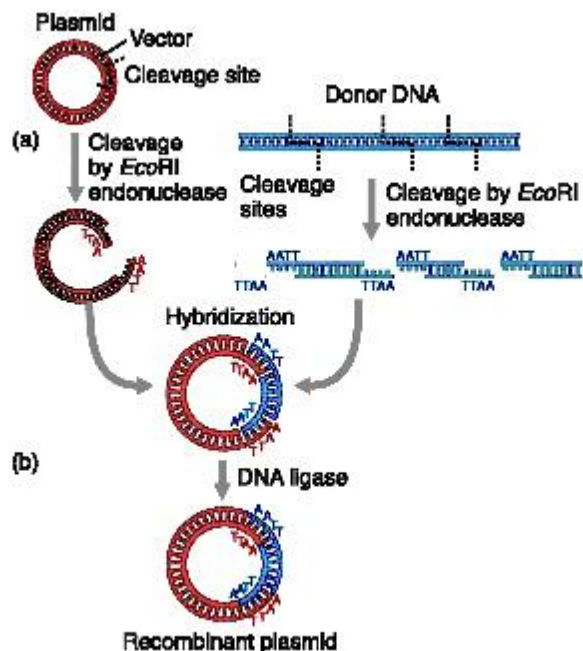
## 22.14 Recombinant DNA and Genetic Engineering

### 1. Isolate Gene of Interest

The gene for producing a protein is isolated from a cell. The gene is on the DNA in a chromosome. Special DNA cutting proteins are used to cut out certain sections of DNA. The gene can be isolated and then copied so that many genes are available to work with.

### 2. Prepare Target DNA

In 1973, two scientists named Boyer and Cohen developed a way to put DNA from one organism into the DNA of bacteria. This process is called recombinant DNA technology. First, a circular piece of DNA called a plasmid is removed from a bacterial cell. Special proteins are used to cut the plasmid ring to open it up.



### 3. Insert

#### DNA into Plasmid

The host DNA that produces the wanted protein is inserted into the opened plasmid DNA ring. Then special cell proteins help close the plasmid ring.

#### **4. Insert Plasmid back into cell**

The circular plasmid DNA that now contains the host gene is inserted back into a bacteria cell. The plasmid is a natural part of the bacteria cell. The bacteria cell now has a gene in it that is from a different organism, even from a human. This is what is called recombinant DNA technology

#### **5. Plasmid multiplies**

The plasmid that was inserted into the bacteria cell can multiply to make several copies of the wanted gene. Now the gene can be turned on in the cell to make proteins.

#### **6. Target Cells Reproduce**

Many recombined plasmids are inserted into many bacteria cells. While they live, the bacteria's cell processes turn on the inserted gene and the protein is produced in the cell. When the bacterial cells reproduce by dividing, the inserted gene is also reproduced in the newly created cells.

#### **7. Cells Produce Proteins**

The protein that is produced can be purified and used for a medicine, industrial, agricultural, or other uses. Check out the Uses section to see how **GE** is used.

### **22.15 The Polymerase Chain Reaction**

The polymerase chain reaction is a technique for quickly "cloning" a particular piece of DNA in the test tube (rather than in living cells like *E. coli*). Thanks to this procedure, one can make virtually unlimited copies of a single DNA molecule even though it is initially present in a mixture containing many different DNA molecules.

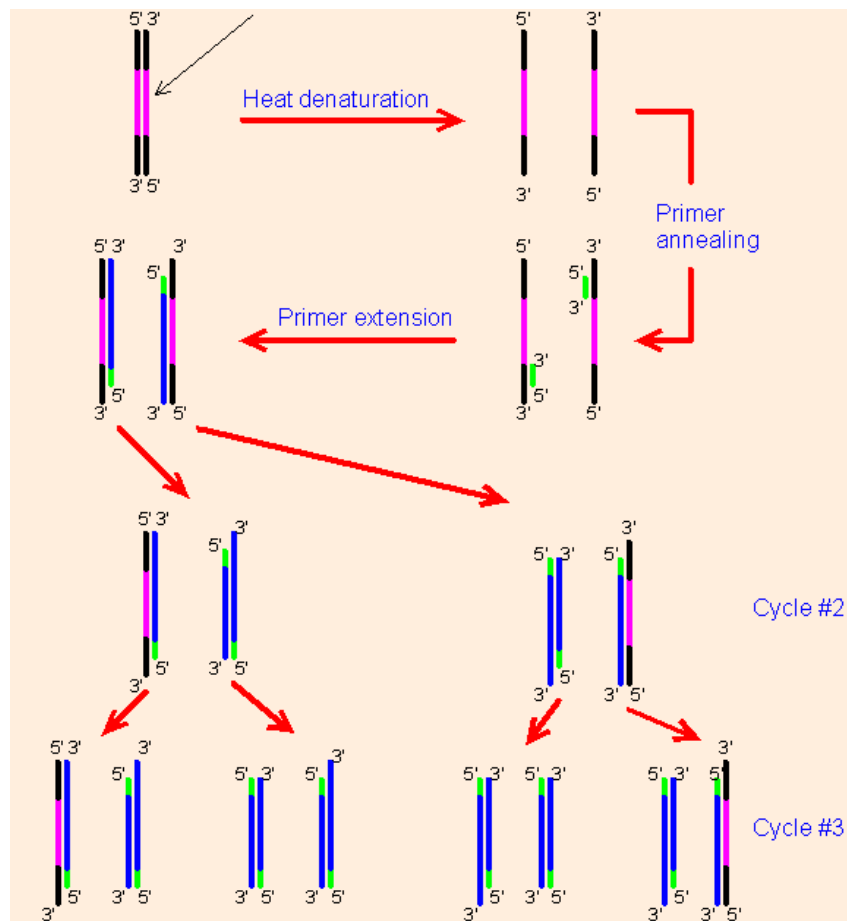
The technique was made possible by the discovery of **Taq polymerase**, the **DNA polymerase** that is used by the bacterium ***Thermus aquaticus*** that was discovered in hot springs. This DNA polymerase is stable at the high temperatures need to perform the amplification, whereas other DNA polymerases become denatured.

Since this technique involves amplification of DNA, the most obvious application of the method is in the detection of minuscule amounts of specific DNAs. This important in the detection of low level **bacterial infections** or rapid changes in transcription at the single cell level, as well as the detection of a **specific individual's DNA in forensic science** (like in the O.J. trial). It can also be used in DNA sequencing, **screening for genetic disorders**, site specific **mutation of DNA**, or **cloning or subcloning of cDNAs**.

There are three basic steps in PCR. First, the target genetic material must be denatured-that is, the strands of its helix must be unwound and separated-

by heating to 90-96°C. The second step is hybridization or annealing, in which the primers bind to their complementary bases on the now single-stranded DNA. The third is DNA synthesis by a polymerase. Starting from the primer, the polymerase can read a template strand and match it with complementary nucleotides very quickly. The result is two new helices in place of the first, each composed of one of the original strands plus its newly assembled complementary strand.

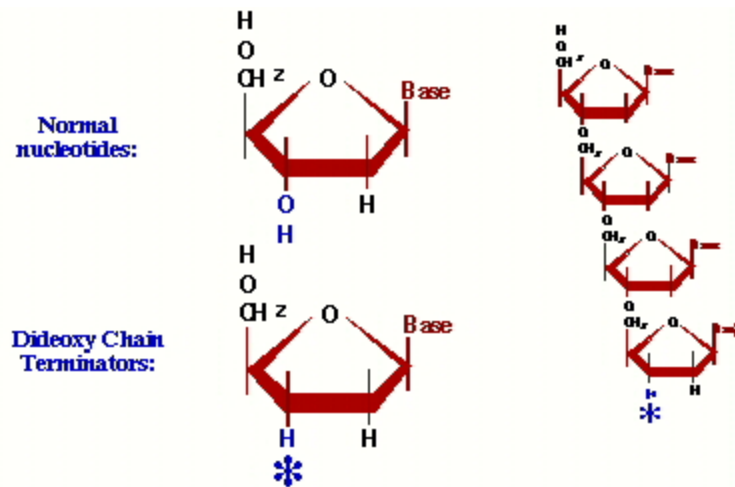
All PCR really requires in the way of equipment is a reaction tube, reagents, and a source of heat. But different temperatures are optimal for each of the three steps, so machines now control these temperature variations automatically.



## 22.16 DNA Sequencing

DNA sequencing is the determination of the precise sequence of nucleotides in a sample of DNA.

The most popular method for doing this is called the **dideoxy method** or Sanger method (named after its inventor, Frederick Sanger, who was awarded the 1980 Nobel prize in chemistry [his second] for this achievement).



### Replicating a DNA strand in the presence of dideoxy-T

MOST of the time when a 'T' is required to make the new strand, the enzyme will get a good one and there's no problem. MOST of the time after adding a T, the enzyme will go ahead and add more nucleotides. However, 5% of the time, the enzyme will get a dideoxy-T, and that strand can never again be elongated. It eventually breaks away from the enzyme, a dead end product.

Sooner or later ALL of the copies will get terminated by a T, but each time the enzyme makes a new strand, the place it gets stopped will be random. In millions of starts, there will be strands stopping at every possible T along the way.

ALL of the strands we make started at one exact position. ALL of them end with a T. There are billions of them ... many millions at each possible T position. To find out where all the T's are in our newly synthesized strand, all we have to do is find out the sizes of all the terminated products!