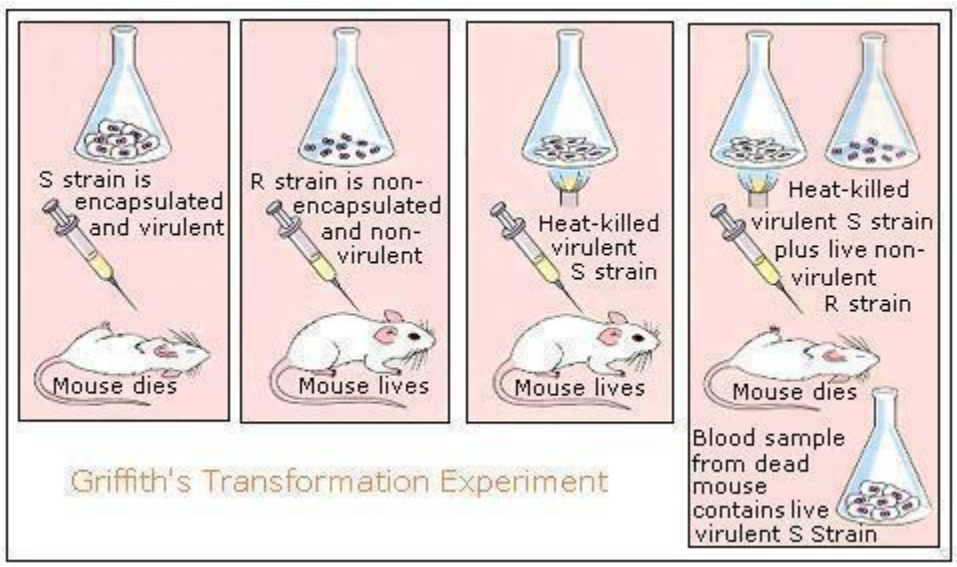
NUCLEIC ACID STRUCTURE AND FUNCTION CHAPTER 1

Scientists whose work led to the discovery of DNA as the transforming factor:



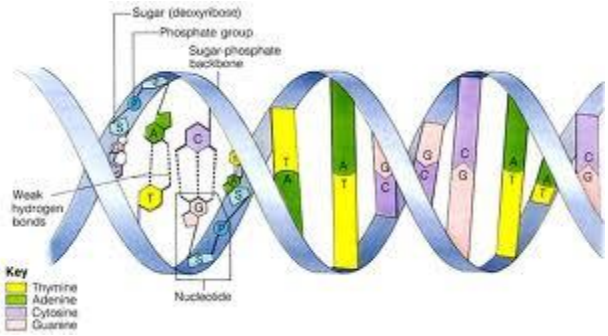
Griffith, in the 1920 experimented with the pneumococcal bacteria. The smooth strain has a polysaccharide coat, which evades phagocytosis by the host's immune system. Griffith injected smooth bacteria (S) into mice, and the mice died. He then noticed that a different strain of bacteria did not appear to have the polysaccharide coat because they had a rough appearance. He then injected these rough (r) bacteria into the mice, and the mice did not die. In his third experiment, he injected heat-killed smooth bacteria cell components into the mice, and they lived. In his fourth experiment, he injected heat-killed smooth bacteria and rough bacteria into the mice, and the mice lived. Somehow the rough bacteria were transformed into smooth bacteria.

Avery then treated Griffith's mixture of heat-treated (S) bacteria and (r) bacteria with protein destroying enzymes, and the colonies were still transformed. He then treated Griffith's mixture with DNA destroying enzymes, and the colonies were not transformed. This led him to conclude that DNA was the transforming factor or principle.

Hershey and Chase labeled one group of phage's protein coat with radioactive sulfur and another group of phages DNA with radioactive phosphorus. He allowed the two groups of phages to infect bacteria. Radioactive phosphorus was found in the cell, and radioactive sulfur was not. The conclusion is that DNA is the transforming factor.

Rosalind Franklin and Maurice Wilkins used x-ray crystallography to get pictures of DNA and determined that it had a helix structure. Watson and Crick then went on to determine that the true structure was a double helix.

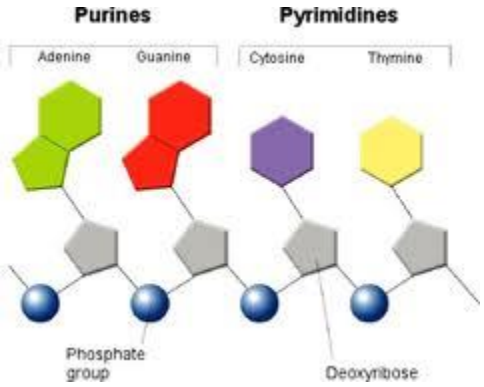
- nucleic acids are the carrier of genetic information and have a structure that is suited to that function



- there are two main types of nucleic acids**: DNA and RNA**

characteristic	DNA	RNA
sugar	deoxyribose	ribose
Number of strands	Double-stranded	Single-stranded
base	A,T,C,G	A,U,C,G

- each type of nucleic acid consists of a **sugar-phosphate backbone,** the sugars and phosphates are connected by phosphodiester bonds, and **nitrogenous base rungs**



The bases

- the nitrogenous bases are of two types **purines** and **pyrimidines**

- the two **purine bases** are adenine and guanine and are double rings

- the **pyrimidine bases** are thymine and cytosine in DNA and uracil and cytosine in RNA and are single rings

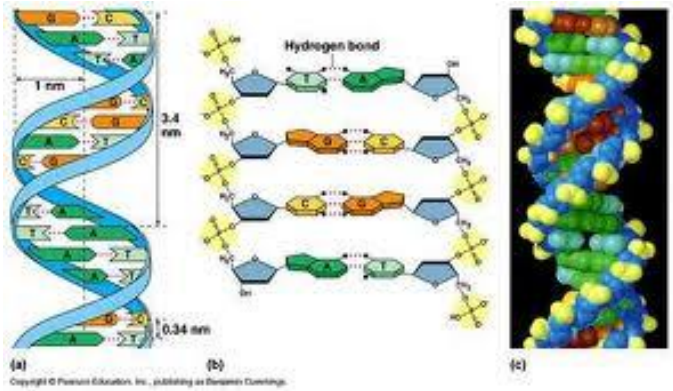
- A always pairs with T in DNA and U in RNA, C always pairs with G

- Chargaff's rule says that the ratio of A to T is 1:1 and G to C is 1:1, in humans, the ratio AT/CG is 1.4/1

Structure of DNA

-- the sugar in DNA is deoxyribose which is a five-carbon sugar, the nitrogenous bases attach at the 1' carbon and the phosphate of the same nucleotide binds to the 5' carbon, and the phosphate of the next nucleotide binds at the 3' carbon

- a molecule of DNA is composed of two nucleotide chains which are coiled clockwise around each other to form a **double helix** with ten nucleotides per turn



- the two chains run in opposite directions (antiparallel) and are held together by hydrogen bonds

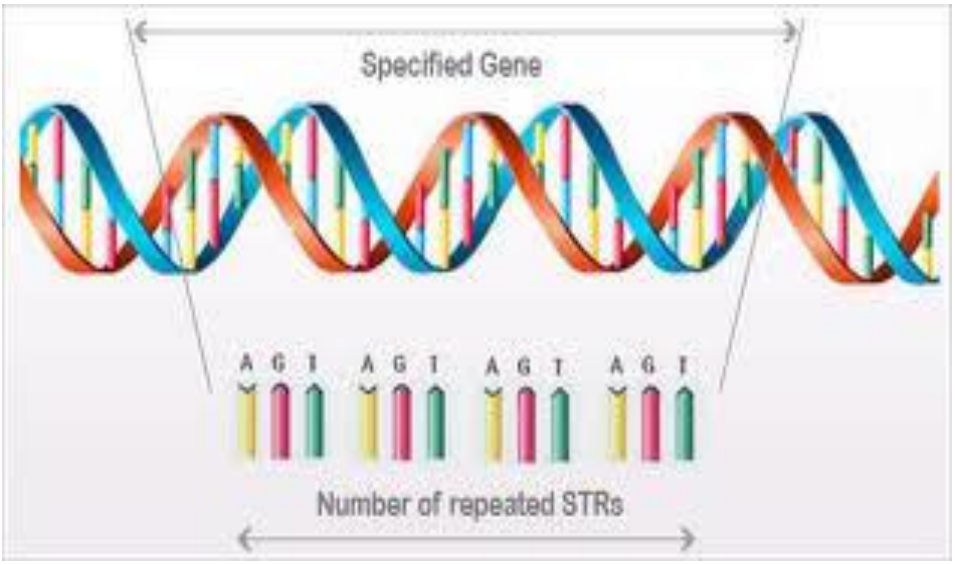
- two hydrogen bonds form between adenine and thymine and adenine and uracil; three hydrogen bonds form between guanine and cytosine, the CG bonds are more stable not because of the three hydrogen bonds, but because of intra-strand base stacking interactions

DNA

- the unit of DNA is a base pair (bp), there are 1000bp in a kilobase, the total length of DNA is 3 x 10^9 (3 billion) base pairs (haploid)

- there are an estimated 20,000-25,000 structural genes encoded in human DNA; structural genes usually are single copy, the single-copy genes and their regulatory sequences account for 70% of the DNA, the remaining 30% is repetitive and has no proven function

- DNA is about 2 meters long; it is wound around a histone protein to form a nucleosome, the nucleosome is coiled to form a helical solenoid, the solenoids are organized in loops which are attached to protein scaffolds

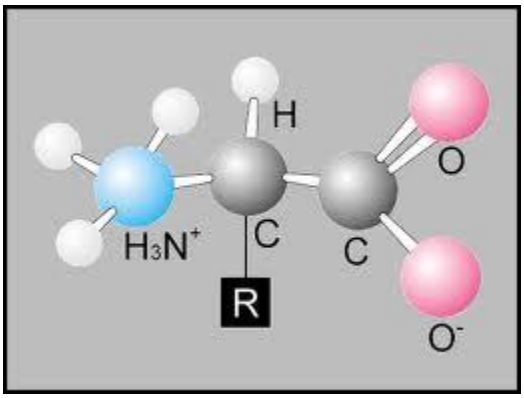


- repetitive DNA is subdivided into tandem repeats (satellite DNA) and interspersed repeats

Nucleic Acid function

- nucleic acids have two functions: the direction of protein synthesis and transmission of this information from one generation to the next

- proteins whether structural components, enzymes, carrier molecules, hormones or receptors are all composed of amino acids



- twenty amino acids are known in humans

- the sequence of amino acids determines the form and function of the resulting proteins

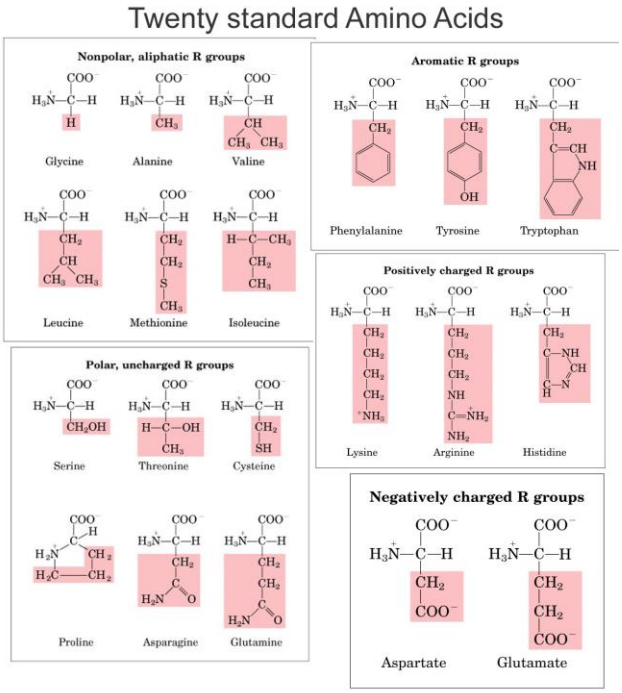
- all proteins are encoded in DNA; proteins are the only organic compound encoded by DNA, protein enzymes manage the construction of the other organic compounds

- the unit of DNA that codes for a protein is **a gene**

- a set of three DNA base pairs is called a **triplet** and codes for an amino acid, a set of three bases on mRNA is called a **codon** and three bases on tRNA are called an **anticodon**

- since there are four bases and three bases in a triplet, there are 64 possible combinations or codons (4^3= 64)

- by convention, each codon is shown in terms of the messenger DNA and the strand of DNA that is the same as the mRNA except having T instead of U is called the sense strand and the strand of DNA that acts as a template is called the antisense strand



- all amino acids except methionine and tryptophan are coded for by more than one codon (a degenerate code), each tRNA carries only one amino acid, but may be complementary to more than one codon on mRNA

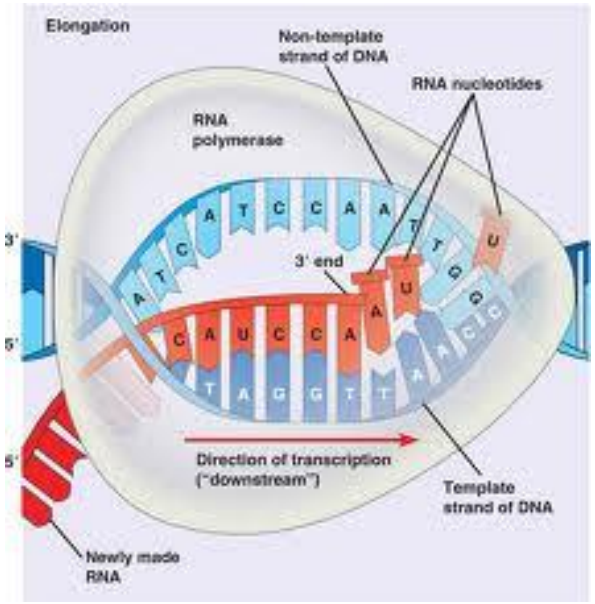
- **wobble hypothesis**- the interaction between the codon in the mRNA and the anticodon in the tRNA need to be exact in 2 out of 3 nucleotide positions; it did not in the 3rd position which is 5' on the anticodon and 3' on the codon

- three of the codons are stop codons: **UAA, UGA, UAG,** and one codon. **AUG** is a start signal.

- the first stage of protein synthesis is transcription

Transcription

- the process of transferring information from a DNA's gene base sequence to the complementary base sequence of an mRNA molecule, DNA cannot code directly to protein



- one strand, which is the template strand and is called the **antisense strand**, is read from the 3' to 5' direction, and the mRNA is synthesized in the 5' to 3' direction with a complementary sequence under the influence of RNA polymerase II. The mRNA sequence is identical to that of the DNA that does not serve as the template which is the sense strand, except for the substitution of uracil for thymine.

- the 5' direction is considered upstream, and the 3' is considered downstream

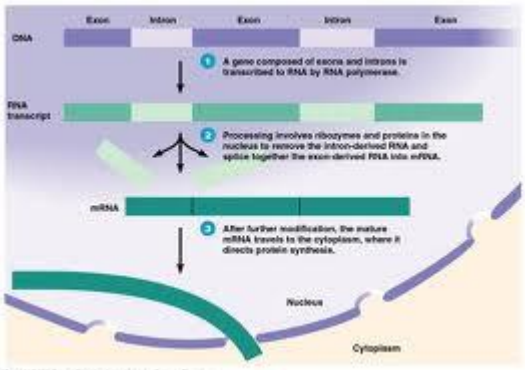
- a **transcription factor** binds to a particular DNA site adjacent to the "start" sequence called the **promoter**. For some human genes several different promoters exist and are in different parts of the genome. This can result in variations in the protein produced. The transcription factor determines which genes are to be decoded and when.

- the transcription factor mediates the binding of **RNA polymerase II**, the enzyme that oversees the synthesis of mRNA, the promoter sequence orients the RNA polymerase and determines which strand of DNA will be transcribed, RNA polymerase moves in the 3' to 5' direction and assembles the complementary mRNA in the 5' to 3' direction

- the RNA polymerase opens the DNA helix and the DNA segment coding for the protein is uncoiled by the enzyme helicase.

- soon after RNA synthesis begins, the 5' end of the growing RNA molecule is "capped" by the addition of a chemically modified guanine nucleotide. This **5' cap** appears to help prevent the RNA molecule from being degraded during synthesis, and later it helps to indicate the starting position for translation of the mRNA molecule into a protein.

- transcription proceeds at a rate of about 30 nucleotides per second, and the signal to terminate transcription is A-rich sequences on the template strand called **a poly-A tail** and contains 100-200 adenine bases. This poly-A tail may be involved in stabilizing the mRNA molecule, so it is not degraded when it reaches the cytoplasm. The introns are then removed. The strand of mRNA, if lined up with the corresponding DNA, would be shorter, so DNA would



make loops to match the correct DNA sequence with the remaining exons.

- after modification, the mRNA molecule diffuses to the cytoplasm, and the DNA strands reassociate

Introns and Exons

- human genes are larger than expected because the coding **exons** are interrupted by the non-coding **introns** whose function is not known

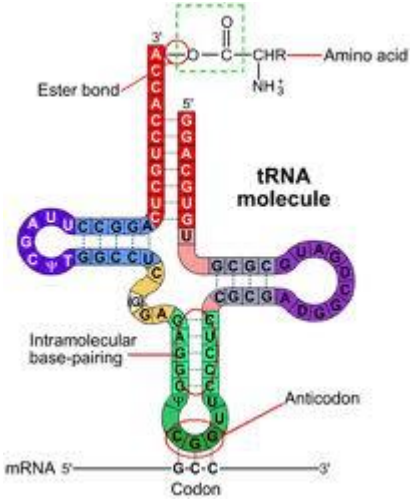
- the introns are removed by splicing before the mRNA moves to the cytoplasm. The splicing enzymes are directed to the correct location by DNA sequences known as **consensus sequences** (common in all eukaryotic organisms)

- the sequences around the intron/exon junctions serve as recognition sites for the splicing enzymes

- introns usually start with GT and end in AG, mutations can interfere with mRNA splicing and can cause genetic disease

- the initial mRNA may be 2-3 times the length as the definitive mRNA. Possible explanations for the presence of introns can be by lengthening genes. They encourage the shuffling of genes when homologous chromosomes exchange material during meiosis. Introns may also serve to modify the amount of time required for DNA replication and transcription.

Translation



- each mRNA molecule becomes attached to one or more **ribosomes**

- the ribosome moves along mRNA from the 5' to the 3' end, and each mRNA is recognized by a complementary tRNA

- tRNA has the **anticodon,** which is complementary to the codon, at one end and the amino acid is carried at the tail

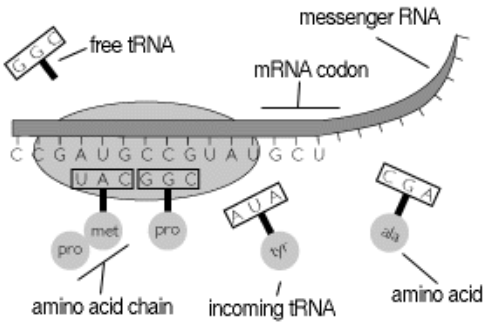
- the attachment process is controlled by a synthetase enzyme which is activated by ATP

- anticodons form hydrogen bonds with complementary codons, so the tiny tRNA links the languages of nucleic acids and proteins

- translation begins when a particular "initiator" tRNA binds to an mRNA molecule and a small ribosomal subunit

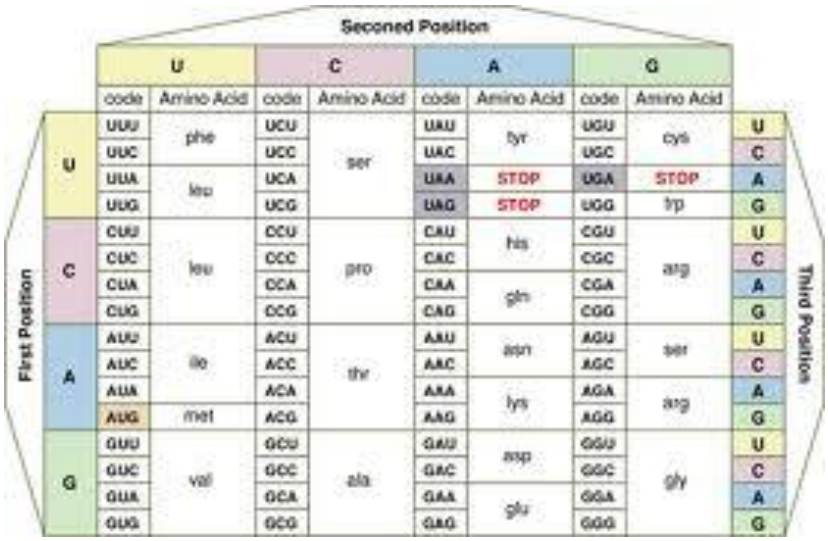
- downstream on the mRNA is the "start" codon (AUG) which is recognized by the tRNA anticodon (UAC) which carries the amino acid methionine

- **peptide bonds** (covalent) form between the carboxyl carbon atom of one amino acid and the core amino nitrogen atom of another



- the mRNA strand continues to be read in the 5' to 3' direction (the amino end of the mRNA is the 5' end, and the carboxyl end is the 3' end)until it reaches one of the stop codons (UGA, UAA or UAG)

- the average protein contains about 300 amino acids, which would be coded by 900bp of DNA



	U		C		A		G		
	code	Amino Acid	code	Amino Acid	code	Amino Acid	code	Amino Acid	
U	UUU	phe	UCU	ser	UAU	tyr	UGU	cys	U
	UUC		UCC		UAC		UGC		C
	UUA	leu	UCA		UAA	STOP	UGA	STOP	A
	UUG		UCG		UAG	STOP	UGG	trp	G
C	CUU	leu	CCU	pro	CAU	his	CGU	arg	U
	CUC		CCC		CAC		CGC		C
	CUA		CCA		CAA	gln	CGA		A
	CUG		CCG		CAG		CGG		G
A	AUU	ile	ACU	thr	AAU	asn	AGU	ser	U
	AUC		ACC		AAC		AGC		C
	AUA		ACA		AAA	lys	AGA	arg	A
	AUG	met	ACG		AAG		AGG		G
G	GUU	val	GCU	ala	GAU	asp	GGU	gly	U
	GUC		GCC		GAC		GGC		C
	GUA		GCA		GAA	glu	GGA		A
	GUG		GCG		GAG		GGG		G

- many proteins are not in their final form after ribosomal translation

- each kind of amino acid has a unique side part that sticks out. This protrusion may be long, short, or shaped like a fat pillow. In addition, this side part might be greasy and therefore easily slide next to a nearby protein that has greasy amino acids or harbors a positive or negative charge that could make it happy to reside in the watery cytoplasm

- the protein is then folded into its three-dimensional shape with the help of protein **chaperones** with acidic and essential amino acids facing out and hydrophobic bases facing inward

- a mutation of an amino acid can affect not only the active site but also the three-dimensional shape of the protein.

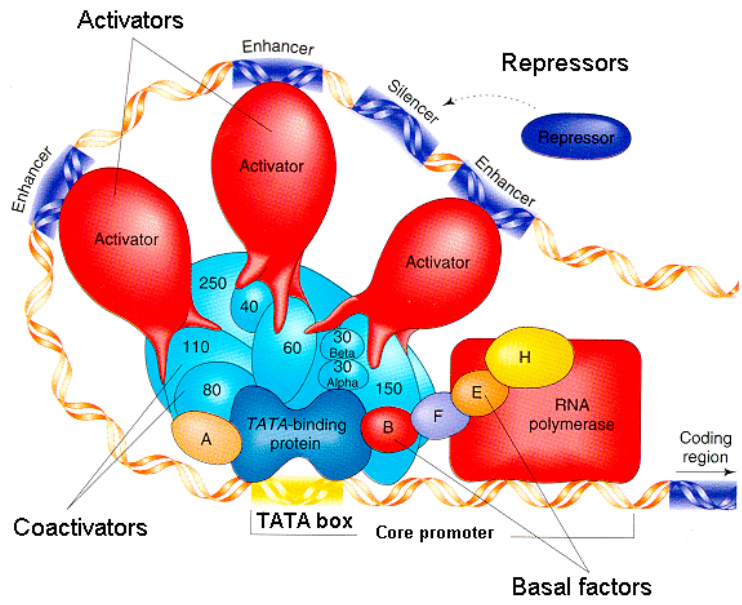
- some genes are transcribed in all cells of the body. These **housekeeping genes** encode products that are required for a cell's maintenance and metabolism.

-most genes are transcribed only in specific tissues at specific points in time. Therefore, in most cells, only a small proportion of genes are actively transcribed. This explains why different cells can produce different protein products even though they all have the same DNA.

Gene regulation

- all nucleated cells of an individual have an identical genome. However, only a small fraction of the total genome is expressed

- the pattern of gene expression varies not only for the initial differentiation of the cells and tissue, but also to meet the fluctuating demands of the cells for certain proteins



- immediately upstream of the gene is the **promoter** which is involved in the attachment of RNA polymerase II to the DNA template

- promoters for RNA polymerase II are usually several nucleotides long and contain a consensus sequence 5'-TATA-3' (the **TATA box**)

- the TATA sequence binds a series of general transcription factors to initiate transcription of nearly all mRNA. It is surrounded by GC rich sequences as well as the BRE sequence (transcription factor recognition element), the INR (initiator sequence), the DPE (downstream promoter element), and further upstream the CAAT box to which several transcription factors bind.

- **enhancers** bind to specific transcription factors and modulate (increase or decrease) the activity of the promoters, most enhancers are active only in specific cells, **silencers** help to repress the transcription of genes.

- DNA –binding motifs allow transcription factors to find specific DNA sequences. These DNA binding motifs, which are configurations in the transcription-factor protein, allow it to fit snugly and stably into a unique part of the DNA helix. These DNA –binding motifs may even bend DNA so that distant enhancer sequences can interact with target genes.

- a **mutation** can result in no gene product, abnormal persistence of a fetal gene product, or anomalous patterns of gene expression

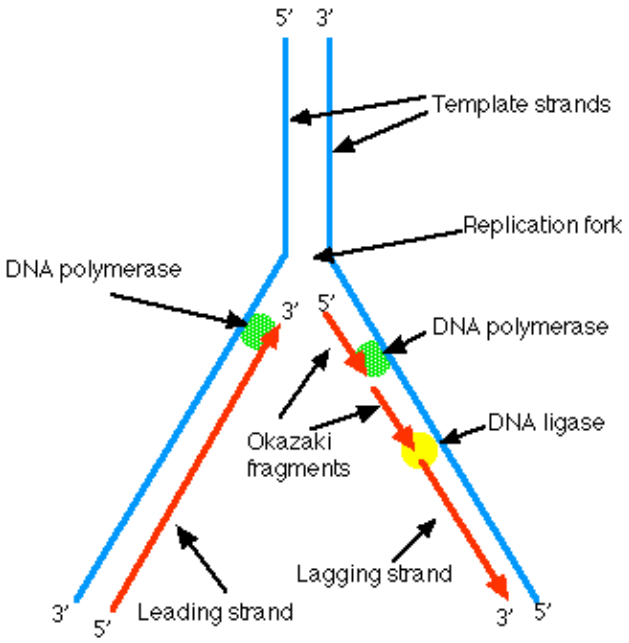
-gene activity can also be related to patterns of chromatin coiling or condensation. Chromatin is the combination of **DNA and the histone proteins** around which the DNA is wound)

-decondensed, or open, chromatin is called euchromatin and are characterized by histone acetylation, acetylation of histones reduces their binding to histones, so it is more accessible to transcription factors

- heterochromatin is usually less acetylated and more methylated, more condensed and is transcriptionally inactive

DNA replication

- Mendelson and Stahl designed an experiment to show the DNA replication is semiconservative. They put cesium chloride in a centrifuge until it layered out by density, They then grew E. coli in ^{15}N media, and because it was a heavy isotope of nitrogen, the DNA grew heavier, they then put it into ^{14}N media. When they put the DNA in the cesium chloride where it would layer out by density, the new strands of DNA were intermediate in density between ^{15}N and ^{14}N, proving the replication was semiconservative. This method is called **semiconservative** since each new DNA has one original strand and one new one if a chromosome's termini are not completely replicated then the chromosome gets shorter with each replication cycle



- the two strands of DNA separate at a number of points (up to 100 per chromosome) and each strand can serve as a template, **helicase** is the enzyme that unwinds the DNA

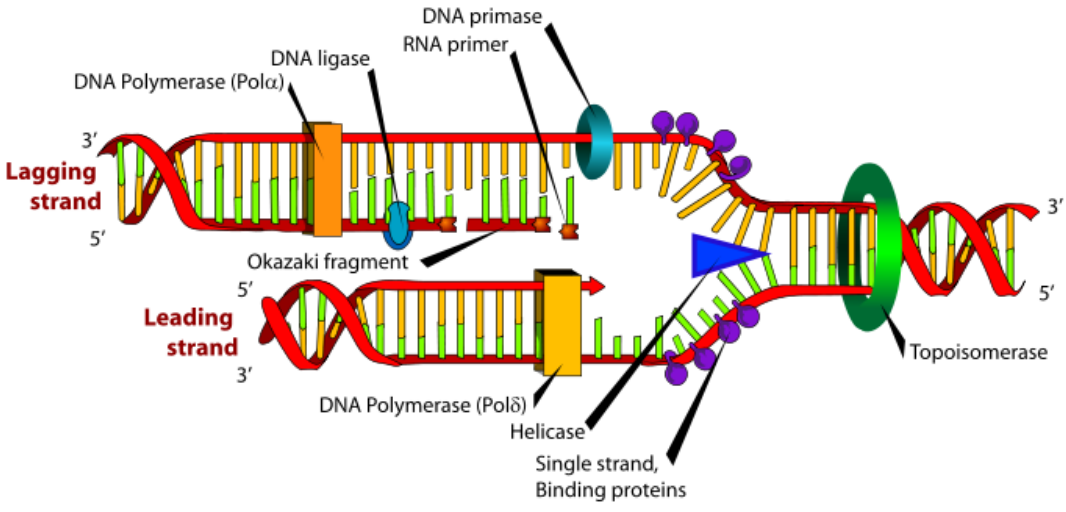
- the site of separation is called the replication bubble and the Y-shaped region is called the replication fork

- the DNA building blocks are nucleotides and the substrates for DNA synthesis are DNA **nucleoside** triphosphates, which have three phosphate groups, as each DNA nucleoside triphosphate joins the nucleotide chain, hydrolysis of the terminal two phosphates provides the energy to drive the polymerization process

- DNA polymerase can only work in one direction (it can only add nucleotides at the 3' end), so the leading strand follows the replication fork, and the lagging strand is constructed in short segments in the opposite direction called **Okazaki fragments**. DNA can only be read in the 3' to 5' direction

- In order for the Okazaki fragments to be synthesized, a **primer** must be created, either RNA polymerase or primase makes the primer, the bases are then added in the 5' to 3' direction, the primers are then removed, and the fragments are joined together **by DNA ligase**

- When referring to the orientation of sequences along a gene, the 5' direction is termed "upstream," and the 3' direction is termed "downstream."



- DNA replication occurs at a rate of 40-50 nucleotides per second. The human genome is 250 million nucleotides. This would take two months except

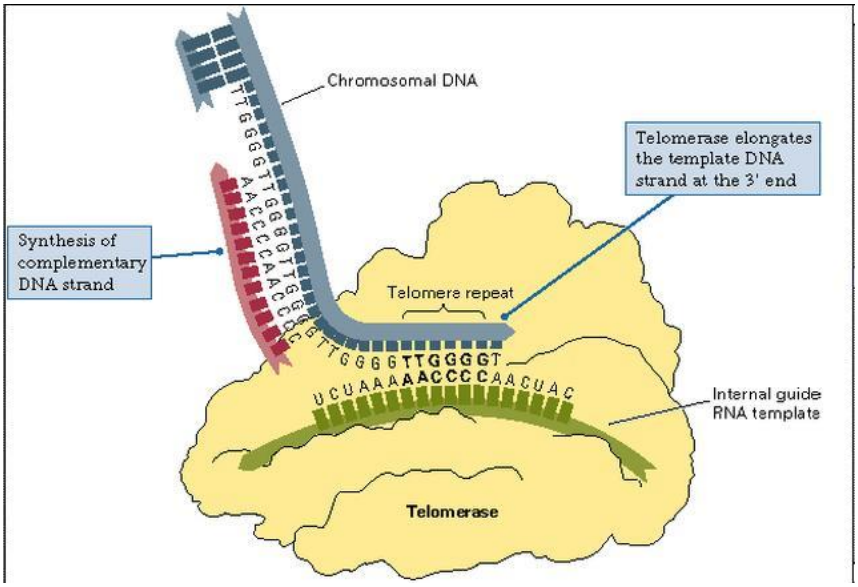
that there are many different replication points along a single chromosome.

- replication occurs in both directions from each initiation point until the two new strands of DNA are complete

- after replication histones synthesized in the cytoplasm associate with the DNA

- the ends of the chromosomes get shorter every time the cell divides, because part the bases are used to template off themselves.

-The cells do have a way of extending the length of telomeres, an enzyme, telomerase. Which can make telomeres longer by synthesizing the repeating sequence 5'-GGGTT-3'. It contains a necessary piece of RNA that serves as a template for synthesizing the new strand that does not depend on DNA. Telomerase enzyme is only active in embryonic cells and cancer cells



-In cancer cells, the telomerase gene is over-expressed. Early results of clinical trials show inhibiting the telomerase protein can slow or completely stop the growth of some cancers. Could it also delay aging?

http://www.cbs.dtu.dk/dtucourse/cookbooks/dave/Chromosome-anat.html

Types of DNA

- 1% (34 million) of the 3 billion nucleotide pairs in the human genome encode for proteins. 21 million nucleotides are transcribed into mRNA that is not translated into proteins

- single-copy DNA sequences are seen only once or possibly a few times in the genome. Single-copy DNA comprises 45% of the genome and includes the protein-coding genes.

- Most of single-copy DNA is found in introns or in DNA sequences that lie between genes.

Repetitive DNA

- repetitive DNA, which are sequences that can be repeated thousands of times. There are two main classes of repetitive DNA, dispersed repetitive DNA, and satellite (tandem repeats) DNA.

- **Satellite sequences** are clustered together and are in tandem (next to each other) and compose about 10% of the genome, alpha-satellites (also called macro) are tandem repeats of a 171-bp sequence that can be repeated to several million pairs and are found near the centromeres and the

telomeres of the chromosomes, minisatellites are blocks of tandem repeats (14 to 500 bl long) which extend to a few thousand base pairs. Microsatellites, which are 1 to 13 bp long and extend to a few hundred base pairs Mini and microsatellites are useful for gene mapping because they vary in length among individuals, mini and microsatellites may also be a hotspot for recombination. They extend the length of the chromosome, which facilitates crossing over during meiosis.

- satellite DNA sequences, owing to their composition can be easily separated by centrifugation in a cesium chloride density gradient

- **Dispersed repetitive DNA** are scattered singly throughout the genome and can be divided into short interspersed elements (SINEs) and long interspersed elements (LINEs). These are transposable elements.

- the density of transposable elements in humans is much higher and contains ancient transposons which may give an evolutionary advantage

- one of the most critical SINES are the Alu repeats, the Alu repeats are a family of genes, meaning they all have highly similar DNA sequences, about 1 million Alu repeats are scattered throughout the genome, Alu repeats contain their internal RNA polymerase and promoter sequence so they can replicate themselves and insert themselves into the genome