

COURSE CODE: BCH 307

COURSE TITLE: COURSE TITLE: *Metabolism of nucleic acids.*

NUMBER OF UNITS: 2 Units

COURSE DURATION: 2 hours/week

COURSE DETAILS:

Course Coordinator: Dosumu O.A. Mrs

Email: atdos2003@yahoo.co.uk

Office Location: Room B105, COLNAS

COURSE CONTENT:

Metabolism of purines and pyrimidines. Structure and properties of nucleotides and nucleic acids. Hydrolysis of nucleic acids. Elucidation of DNA structure and properties. Genome organization and biosynthesis of protein. Abnormalities in nucleic acid metabolism-xeroderma pigmentations and skin cancer.

COURSE REQUIREMENTS:

This is a compulsory course for all 300 level students of Biochemistry department. In view of this, students are to participate in all the course activities and have a minimum of 70% attendance to be able to write the final examination.

READING LIST:

- 1 Bohinski, R.C. 1983. Modern concepts in Biochemistry. 4th Edition. Allyn and Bacon.
- 2 Murray, R.K., Granner, D.K., Mayes, P.A. and Rodwell, V.W. 2003. Harper's Illustrated Biochemistry. 26th Edition. McGraw-Hill.

LECTURE NOTES

REMINDER: HYDROLYSIS OF NUCLEIC ACIDS.

Hydrolysis of Nucleic acids by selective methods can be achieved chemically or enzymatically.

Chemical method of hydrolysis:

ACID HYDROLYSIS:

RNA is relatively resistant to the effects of dilute acid, but gentle treatment of DNA with 1Mm Hcl leads to hydrolysis of purine glycosidic bonds and the loss of purine bases from the DNA without affecting the pyrimidine deoxyribose bonds or the phosphodiester bonds of the backbone. At other chemical conditions, selective removal of pyrimidine bases occurs. In most cases, both Nucleic acids can be hydrolysed to their constituent bases by the treatment with 72% perchloric acid (HClO_4) for 1hour. The resulting nucleic acid derivative which is devoid of purine bases is called **APURINIC ACID**; while that devoid of pyrimidine bases is called **APYRIMIDINIC ACID**.

ALKALI HYDROLYSIS:

DNA is not susceptible to alkaline hydrolysis. On the other hand, RNA is alkali labile and is readily hydrolyzed by dilute sodium hydroxide.

Enzymatic hydrolysis of Nucleic acids:

Enzymes that hydrolyse nucleic acids are called **NUCLEASES**.

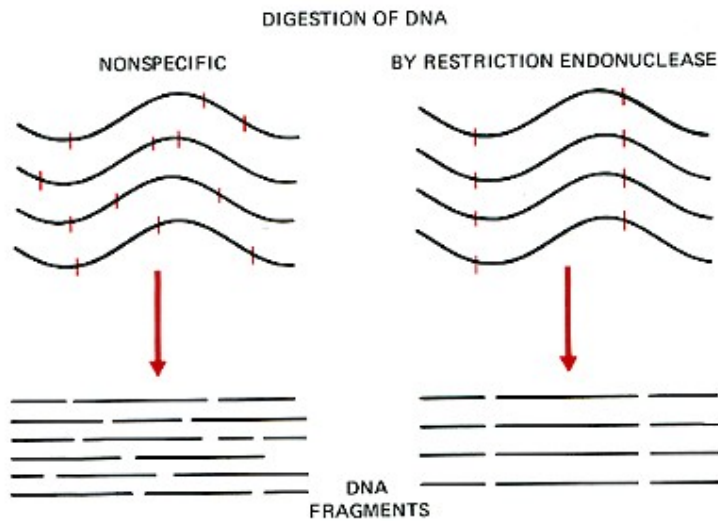
Some nucleases can hydrolyse linkages between 2 adjacent nucleotides at internal positions in the DNA or RNA strand and

proceed stepwise from that end. Such nucleases are called **ENDONUCLEASES**.

Another class of nucleases can hydrolyse only the terminal nucleotide linkage, some at the 5' and others at the 3' end; these are called **EXONUCLEASES**.

DNases (deoxyribonucleases) acts only on DNA
RNases(ribonucleases) are specific for RNA.

Restriction Enzymes.

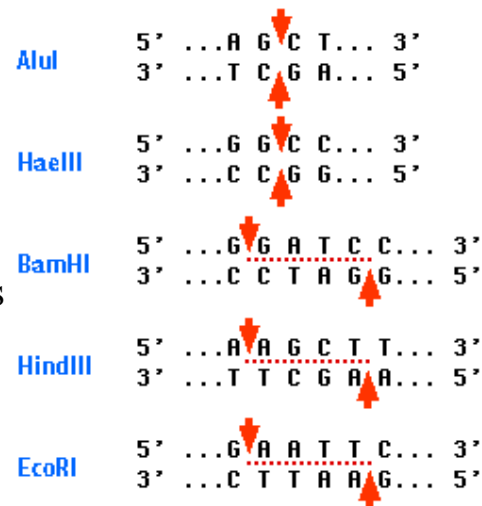


Restriction enzymes are **DNA-cutting enzymes** found in bacteria (and harvested from them for use). Because they cut within the molecule, they are often called **restriction endonucleases**.

In order to be able to sequence DNA, it is first necessary to cut it into smaller fragments. Many DNA-digesting enzymes (like those in your pancreatic fluid) can do this, but most of them are no use for sequence work because they cut each molecule randomly. This produces a heterogeneous collection of fragments of varying sizes. What is needed is a way to cleave the DNA molecule at a few precisely-located sites so that a small set of homogeneous fragments are produced. The tools for this are the restriction endonucleases. The rarer the site it recognizes, the smaller the number of pieces produced by a given restriction endonuclease.

A restriction enzyme recognizes and cuts DNA only at a particular sequence of nucleotides. For example, the bacterium

Hemophilus aegypticus produces an enzyme named **HaeIII** that cuts DNA wherever it encounters the sequence
 5'GGCC3'
 3'CCGG5'



The cut is made between the adjacent G and C. This particular sequence occurs at 11 places in the circular DNA molecule of the virus **φX174**. Thus treatment of this DNA with the enzyme produces 11 fragments, each with a precise length and nucleotide sequence. These fragments can be separated from one another and the sequence of each determined.

AluI and **HaeIII** produce blunt ends

BamHI **HindIII** and **EcoRI** produce "sticky" ends

HaeIII and AluI cut straight across the double helix producing "blunt" ends. However, many restriction enzymes cut in an offset fashion. The ends of the cut have an overhanging piece of single-stranded DNA. These are called "**sticky ends**" because they are able to form base pairs with any DNA molecule that contains the complementary sticky end. Any other source of DNA treated with the same enzyme will produce such molecules.

Mixed together, these molecules can join with each other by the base pairing between their sticky ends. The union can be made permanent by another enzyme, a **DNA ligase**, that forms covalent bonds along the backbone of each strand. The result is a molecule of **RECOMBINANT DNA**. The ability to produce recombinant DNA molecules has not only revolutionized the study of genetics, but has laid the foundation for much of the biotechnology industry. The availability of human **insulin** (for diabetics), human **factor VIII** (for

males with hemophilia A), and other proteins used in human therapy all were made possible by recombinant DNA.

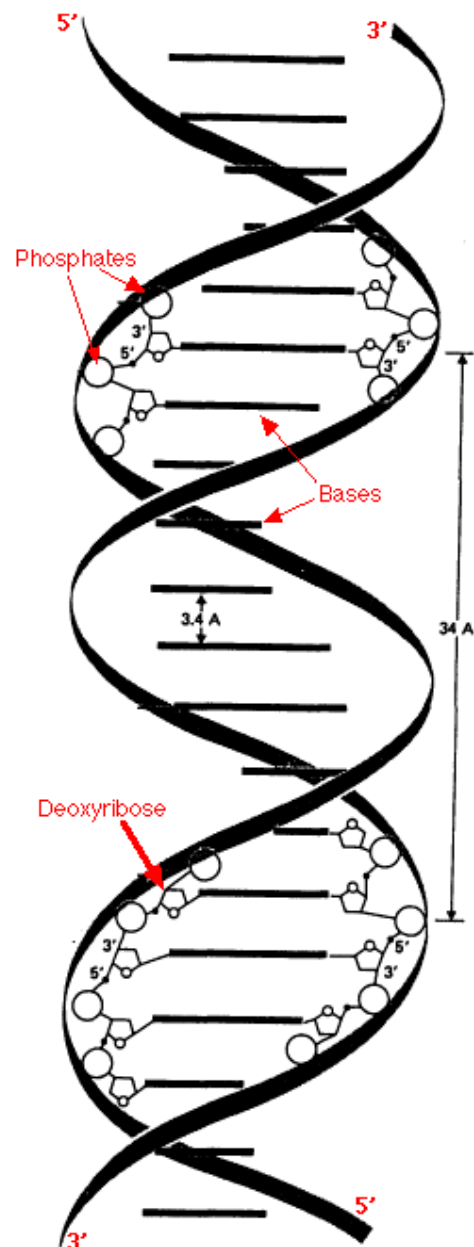
ELUCIDATION OF DNA STRUCTURE.

Double stranded DNA molecules assume one of 3 secondary structures, termed A, B and Z. Fundamentally, this DNA is a regular two-chain structure with hydrogen bonds formed between opposing bases on the 2 chains.

The Double Helix

The double helix of DNA has these features:

- It contains two polynucleotide strands wound around each other.
- The backbone of each consists of alternating deoxyribose and phosphate groups.
- The phosphate group bonded to the 5' carbon atom of one deoxyribose is covalently bonded to the 3' carbon of the next.



- The two strands are "antiparallel"; that is, one strand runs 5' to 3' while the other runs 3' to 5'.
- The DNA strands are assembled in the 5' to 3' direction and, by convention, we "read" them the same way.
- The purine or pyrimidine attached to each deoxyribose projects in toward the axis of the helix.
- Each base forms hydrogen bonds with the one directly opposite it, forming **base pairs** (also called nucleotide pairs).
- 3.4 Å separate the planes in which adjacent base pairs are located.
- The double helix makes a complete turn in just over 10 nucleotide pairs, so each turn takes a little more (35.7 Å to be exact) than the 34 Å shown in the diagram.
- There is an average of 25 hydrogen bonds within each complete turn of the double helix providing a stability of binding about as strong as what a covalent bond would provide.
- The diameter of the helix is 20 Å.
- The helix can be virtually any length; when fully stretched, some DNA molecules are as much as 5 cm (2 inches!) long.
- The path taken by the two backbones forms a major (wider) groove (from "34 Å" to the top of the arrow) and a minor (narrower) groove (the one below).

This structure of DNA was worked out by Francis Crick and James D. Watson in 1953. It is also referred to as the B-form of DNA and it is the most stable. For this epochal work, they shared a Nobel Prize in 1962.

Palindromes

A palindrome is a sequence of letters and/or words, that reads the same forwards and backwards.

"able was I ere I saw elba" is a palindrome.

Palindromes also occur in DNA.

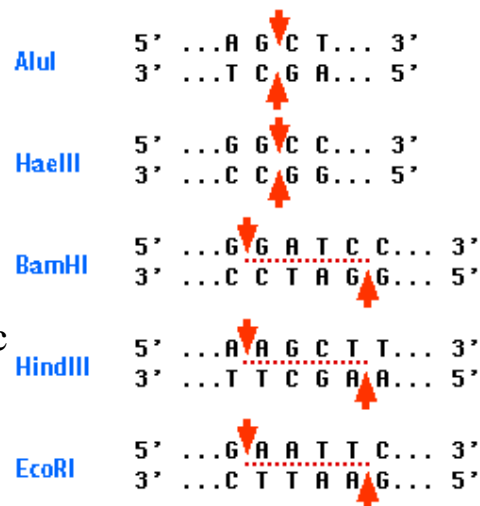
There are two types.

1. Palindromes that occur on opposite strands of the same section of DNA helix.

5' GGCC 3'
3' CCGG 5'

This type of palindrome serves as the target for most restriction enzymes.

The graphic shows the palindromic sequences "seen" by five restriction enzymes (named in blue) commonly used in recombinant DNA work.



AluI and HaeIII produce blunt ends

BamHI HindIII and EcoRI produce "sticky" ends

2. Inverted Repeats

In these cases, two different segments of the double helix read the same but in opposite directions.

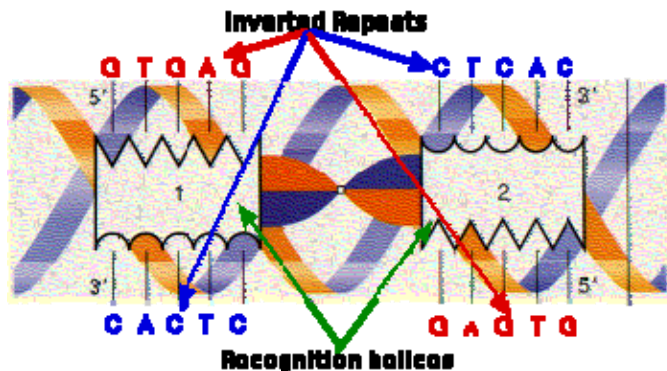
5' AGAACAnnnTGTTCT 3'
3' TCTTGTnnnACAAGA 5'

Inverted repeats are commonly found in

- The DNA to which transcription factors bind.

The DNA sequence shown above is that of the **glucocorticoid response element** where **n** represents any nucleotide.

Transcription factors are often dimers of identical proteins ("homodimers") so it is not surprising that each member of the pair needs to "see" the same DNA sequence in the same orientation



This graphic shows the "recognition helix" to which the CAP protein (a homodimer) binds in the ***lac operon*** of *E. coli*. The DNA of many transposons is flanked by inverted repeats such as this one:

5' GGCCAGTCACAATGG..~400
nt..CCATTGTGACTGGCC 3'
3' CCGGTCAGTGTTACC..~400
nt..GGTAACACTGACCGG 5'

- Inverted repeats at either end of retroviral gene sequences aid in inserting the DNA copy into the DNA of the host.
- Duplicated Genes.

The human Y chromosome contains 7 sets of genes — each set containing from 2 to 6 nearly-identical genes — oriented back-to-back or head-to-head; that is, they are inverted repeats like

the portion shown here. (The dashes represent the thousands of base pairs that separate adjacent palindromes.)

5' ...CACAATTCCCATGGGTTGTGGGAG 3' ----- 5'
 CTCCCACAACCCATGGGATTTGTG... 3'
 3' ...GTGTTAAGGGTACCCAACACCCTC 5' ----- 3'
 GAGGGTGTGGGTACCCTAACAC... 5'

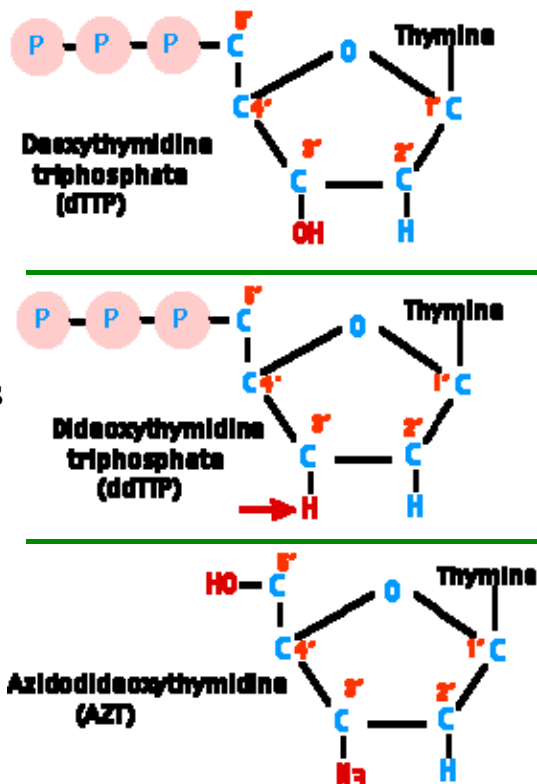
This orientation and redundancy may help ensure that a deleterious mutation in one copy of the set can be repaired using the information in another copy of that set. All that is needed is to form a loop so that the two sequences line up side-by-side. Repairs can then be made (probably by the mechanism of homologous recombination). Here, for example, the single difference in the sequences can be eliminated (red for blue or vice versa).

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).

DNA is synthesized from four deoxynucleotide triphosphates. The top formula shows one of them: deoxythymidine triphosphate



(dTTP). Each new nucleotide is added to the 3' -OH group of the last nucleotide added.

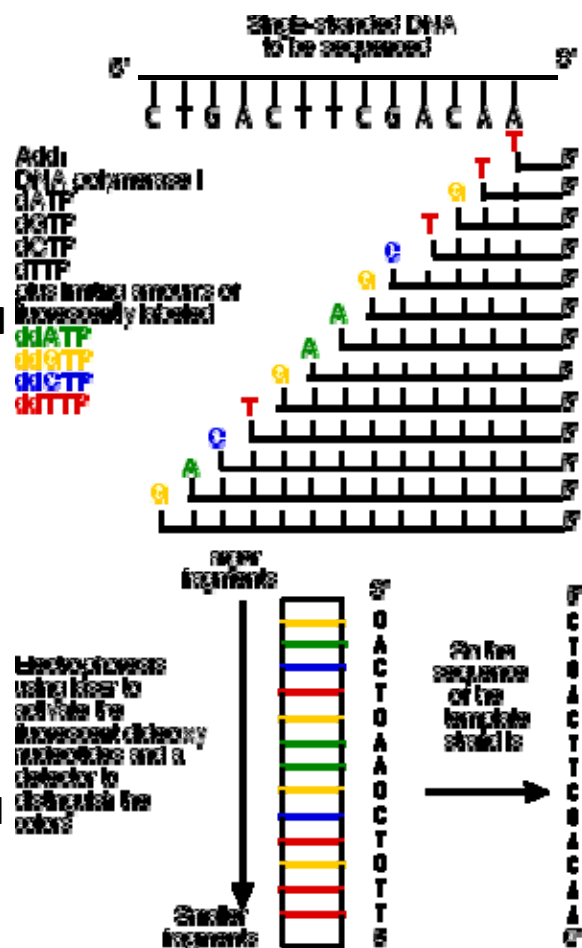
The dideoxy method gets its name from the critical role played by **synthetic** nucleotides that **lack** the -OH at the 3' carbon atom (red arrow). A dideoxynucleotide (dideoxythymidine triphosphate — ddTTP — is the one shown here) can be **added** to the growing DNA strand but when it is, **chain elongation stops** because there is no 3' -OH for the next nucleotide to be attached to. For this reason, the dideoxy method is also called the **chain termination method**.

The Procedure

The DNA to be sequenced is prepared as a single strand.

This template DNA is supplied with

- a mixture of all four **normal** (deoxy) nucleotides in ample quantities
 - dATP
 - dGTP
 - dCTP
 - dTTP
- a mixture of all four **dideoxynucleotides**, each present in limiting quantities and each labeled with a "tag" that fluoresces a different color:
 - **ddATP**
 - **ddGTP**



- ddCTP
- ddTTP
- DNA polymerase I

Because all four normal nucleotides are present, chain elongation proceeds normally until, by chance, DNA polymerase inserts a dideoxy nucleotide (shown as colored letters) instead of the normal deoxynucleotide (shown as vertical lines). If the ratio of normal nucleotide to the dideoxy versions is high enough, some DNA strands will succeed in adding several hundred nucleotides before insertion of the dideoxy version halts the process.

At the end of the incubation period, the fragments are separated by length from longest to shortest. The resolution is so good that a difference of one nucleotide is enough to separate that strand from the next shorter and next longer strand. Each of the four dideoxynucleotides fluoresces a different color when illuminated by a laser beam and an automatic scanner provides a printout of the sequence.

Biosynthesis of Nucleotide Co-Enzymes.

FAD, the pyridine nucleotides (NAD) and Co-enzyme A are derivatives of Adenylic acid.

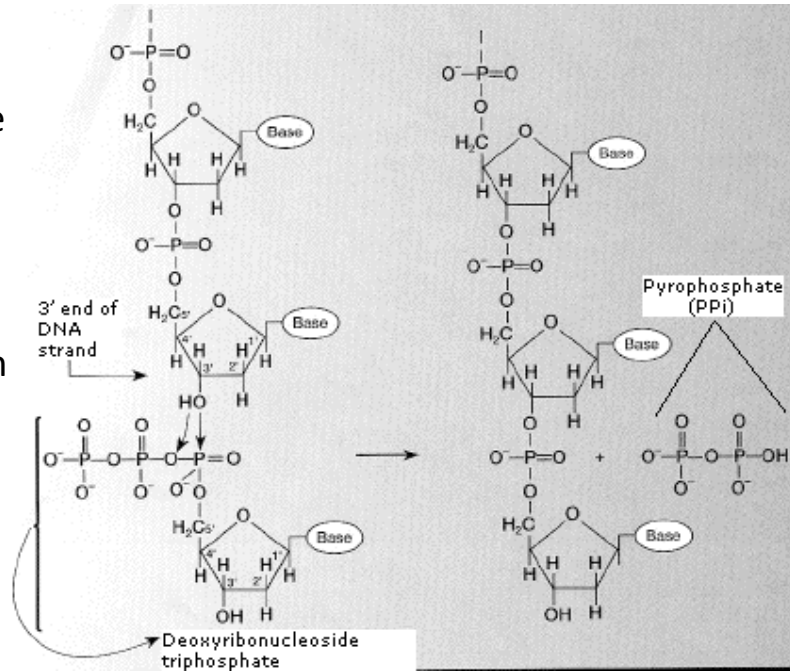
ASSIGNMENT: Write out the biosynthetic pathway of NAD^+ and Co-enzyme A.

DNA Replication

Transfer of genetic information from generation to generation requires the faithful reproduction of the parental DNA. DNA reproduction produces two identical copies of the original DNA in a process termed DNA replication.

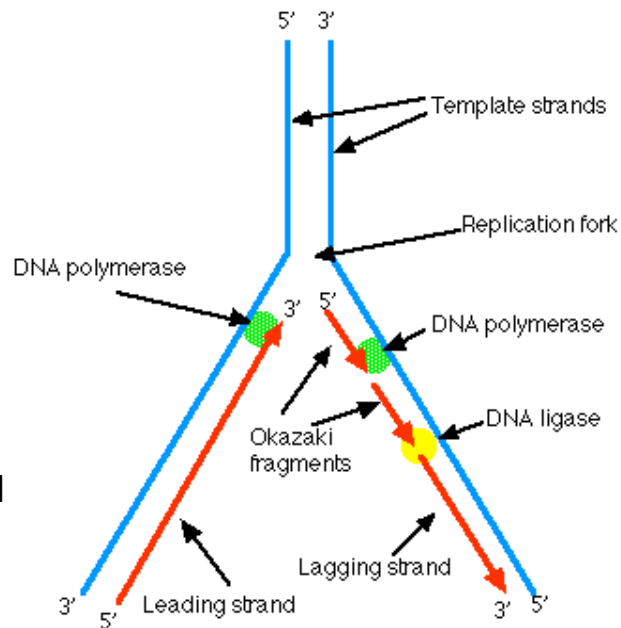
The Biochemical Reactions

- DNA replication begins with the "unzipping" of the parent molecule as the hydrogen bonds between the base pairs are broken.
- Once exposed, the sequence of bases on each of the separated strands serves as a template to guide the insertion of a complementary set of bases on the strand being synthesized.
- The new strands are assembled from deoxynucleoside triphosphates.
- Each incoming nucleotide is covalently linked to the "free" 3' carbon atom on the pentose (figure) as
- the second and third phosphates are removed together as a molecule of **pyrophosphate (PPi)**.
- The nucleotides are assembled in the order that complements the order of bases on the strand serving as the template.
- Thus each C on the template guides the insertion of a G on the new strand, each G a C, and so on.
- When the process is complete, two DNA molecules have been formed identical to each other and to the parent molecule.



The Enzymes

- A portion of the double helix is unwound by a **helicase**.
- A molecule of a **DNA polymerase** binds to one strand of the DNA and begins moving along it in the 3' to 5' direction, using it as a template for assembling a **leading strand** of nucleotides and reforming a double helix. In eukaryotes, this molecule is called DNA polymerase delta (δ).



- Because DNA synthesis can only occur 5' to 3', a molecule of a second type of DNA polymerase (epsilon, ϵ , in eukaryotes) binds to the other template strand as the double helix opens. This molecule must synthesize discontinuous segments of polynucleotides (called Okazaki fragments). Another enzyme, **DNA ligase I** then stitches these together into the **lagging strand**.

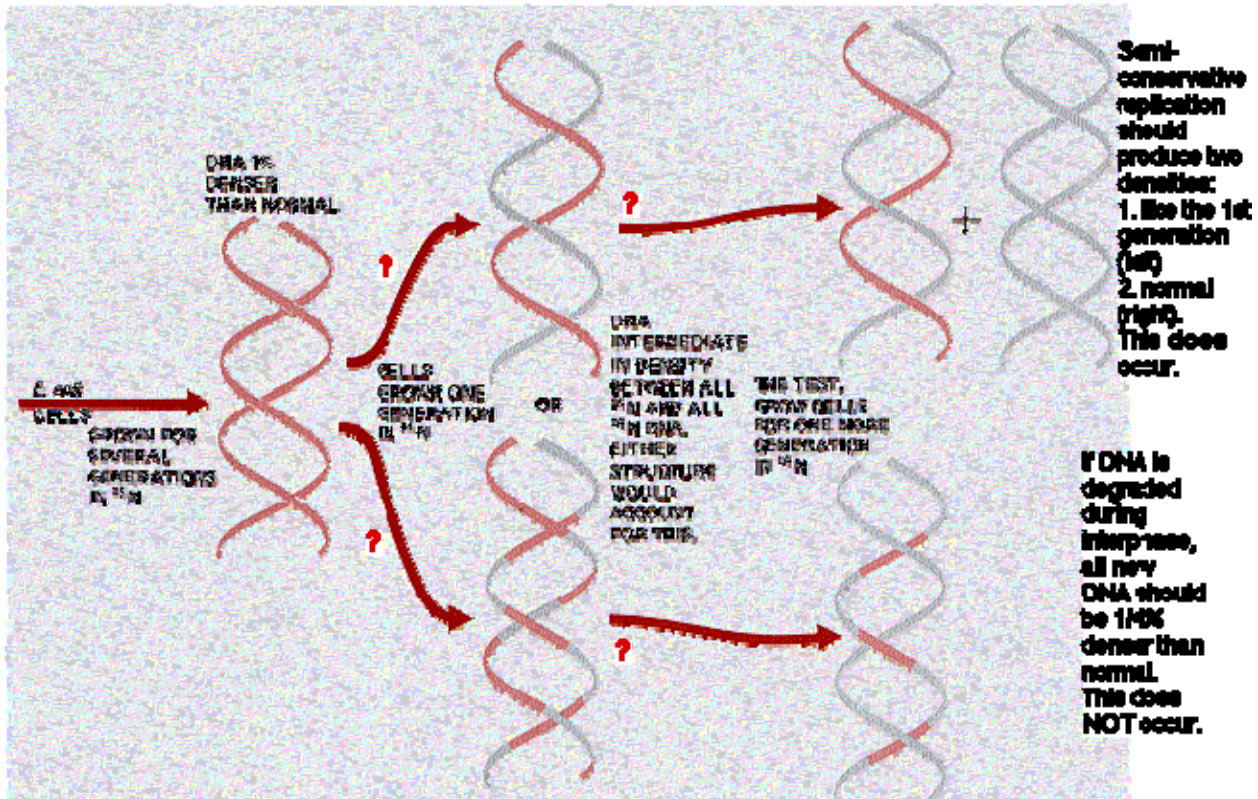
DNA Replication is Semiconservative

When the replication process is complete, two DNA molecules — identical to each other and identical to the original — have been produced. Each strand of the original molecule has

- remained intact as it served as the template for the synthesis of
- a complementary strand.

This mode of replication is described as semi-conservative: one-half of each new molecule of DNA is old; one-half new.

Watson and Crick had suggested that this was the way the DNA would turn out to be replicated. Proof of the model came from the



experiments of Meselson and Stahl.

Transcription and Translation

The majority of genes are expressed as the proteins they encode. The process occurs in two steps:

- **Transcription = DNA → RNA**
- **Translation = RNA → protein**

**Taken together, they make up the "central dogma" of biology:
DNA → RNA → protein.**

Gene Transcription: DNA → RNA

DNA serves as the template for the synthesis of RNA much as it does for its own replication.

The Steps

- Some 50 different protein **transcription factors** bind to **promoter** sites, usually on the 5' side of the gene to be transcribed.
- An enzyme, an **RNA polymerase**, binds to the complex of transcription factors.
- Working together, they open the DNA double helix.
- The RNA polymerase proceeds to "read" one strand moving in its 3' → 5' direction.
- In eukaryotes, this requires — at least for protein-encoding genes — that the nucleosomes in front of the advancing RNA polymerase (**RNAP II**) be removed. A complex of proteins is responsible for this. The same complex replaces the nucleosomes after the DNA has been transcribed and RNAP II has moved on.
- As the RNA polymerase travels along the DNA strand, it assembles **ribonucleotides** (supplied as triphosphates, e.g., **ATP**) into a strand of RNA.
- Each ribonucleotide is inserted into the growing RNA strand following the rules of **base pairing**. Thus for each C encountered on the DNA strand, a G is inserted in the RNA; for each G, a C; and for each T, an A. However, each A on the DNA guides the insertion of the **pyrimidine** uracil (**U**, from uridine triphosphate, UTP). There is no T in RNA.

Quality control. Occasionally RNA polymerase will select and insert an incorrect, mismatched, ribonucleotide. When this

occurs in bacteria (and perhaps in all organisms), the enzyme backs up, removes the incorrect nucleotide (and the one preceding it) and tries again. (Described by Zenkin *et al.*, in the 28 July 2006 issue of **Science**.)

- Synthesis of the RNA proceeds in the 5' → 3' direction.
- As each nucleoside triphosphate is brought in to add to the 3' end of the growing strand, the two terminal phosphates are removed.
- When transcription is complete, the transcript is released from the polymerase and, shortly thereafter, the polymerase is released from the DNA.

Note that at any place in a DNA molecule, either strand may be serving as the template; that is, some genes "run" one way, some the other (and in a few remarkable cases, the same segment of double helix contains genetic information on both strands!). In all cases, however, RNA polymerase transcribes the DNA strand in its 3' → 5' direction.

Gene Translation: RNA -> Protein

How does a particular sequence of nucleotides specify a particular sequence of amino acids?

The answer: by means of **transfer RNA** molecules, each specific for one amino acid and for a particular **triplet** of nucleotides in messenger RNA (mRNA) called a **codon**. The family of tRNA molecules enables the codons in a mRNA molecule to be **translated** into the sequence of amino acids in the protein. At least one kind of tRNA is present for each of the 20 amino acids used in protein synthesis. (Some amino acids employ the services of two or three different tRNAs, so most cells contain as many as 32 different kinds of tRNA.) The amino acid is attached to the appropriate tRNA by an

activating enzyme (one of 20 **aminoacyl-tRNA synthetases**) specific for that amino acid as well as for the tRNA assigned to it.

Each kind of tRNA has a sequence of 3 unpaired nucleotides — the **anticodon** — which can bind, following the rules of base pairing, to the complementary triplet of nucleotides — the **codon** — in a **messenger RNA (mRNA)** molecule. Just as DNA replication and transcription involve base pairing of nucleotides running in opposite direction, so the reading of codons in mRNA (5' -> 3') requires that the anticodons bind in the opposite direction.

Anticodon: 3' CGA 5'
Codon: 5' GCU 3'

The RNA Codons

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

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

Note:

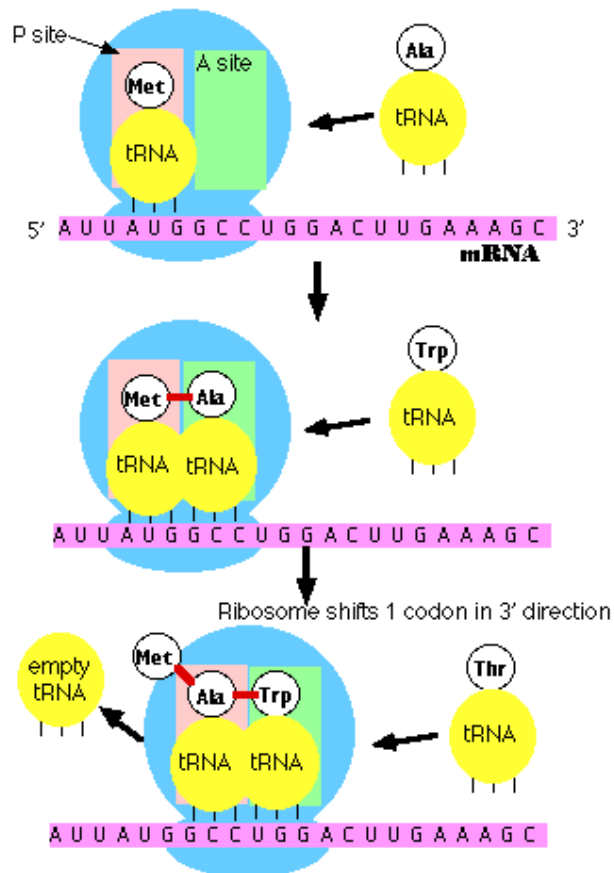
- Most of the amino acids are encoded by synonymous codons that differ in the third position of the codon.

- In some cases, a single tRNA can recognize two or more of these synonymous codons.
- Example: phenylalanine tRNA with the anticodon 3' AAG 5' recognizes not only UUC but also UUU.
- The violation of the usual rules of base pairing at the third nucleotide of a codon is called "wobble"
- The codon **AUG** serves two related functions
 - It begins every message; that is, it signals the **start of translation** placing the amino acid **methionine** at the amino terminal of the polypeptide to be synthesized.
 - When it occurs within a message, it guides the incorporation of methionine.
- Three codons, **UAA**, **UAG**, and **UGA**, act as signals to terminate translation. They are called **STOP** codons.

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 mRNA 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; EF-1 in eukaryotes)
 - **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; EF-2 in eukaryotes) 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.

DNA Repair

Importance

DNA in the living cell is subject to many chemical alterations (a fact often forgotten in the excitement of being able to do DNA sequencing on dried and/or frozen specimens. If the genetic information encoded in the DNA is to remain uncorrupted, any chemical changes must be corrected.

A failure to repair DNA produces a mutation.

The recent publication of the human genome has already revealed 130 genes whose products participate in DNA repair. More will probably be identified soon.

Agents that Damage DNA

- Certain wavelengths of **radiation**
 - ionizing radiation such as gamma rays and X-rays

- **ultraviolet rays**, especially the UV-C rays (~260 nm) that are absorbed strongly by DNA but also the longer-wavelength UV-B that penetrates the ozone shield
- Highly-reactive **oxygen radicals** produced during normal cellular respiration as well as by other biochemical pathways.
- Chemicals in the **environment**
 - many hydrocarbons, including some found in cigarette smoke
 - some plant and microbial products, e.g. the aflatoxins produced in moldy peanuts
- Chemicals used in **chemotherapy**, especially chemotherapy of cancers

Types of DNA Damage

1. All four of the bases in DNA (A, T, C, G) can be covalently modified at various positions.
 - One of the most frequent is the loss of an amino group ("deamination") — resulting, for example, in a **C** being converted to a **U**.
2. **Mismatches** of the normal bases because of a failure of proofreading during DNA replication.
 - Common example: incorporation of the pyrimidine U (normally found only in RNA) instead of **T**.
3. **Breaks** in the backbone.
 - Can be limited to one of the two strands (a single-stranded break, **SSB**) or
 - on **both strands** (a **double-stranded break (DSB)**).
 - Ionizing radiation is a frequent cause, but some chemicals produce breaks as well.
4. **Crosslinks** Covalent linkages can be formed between bases
 - on the same DNA strand ("intrastrand") or
 - on the opposite strand ("interstrand").

Several chemotherapeutic drugs used against cancers crosslink DNA.

Repairing Damaged Bases

Damaged or inappropriate bases can be repaired by several mechanisms:

- **Direct chemical reversal** of the damage
- **Excision Repair**, in which the damaged base or bases are removed and then replaced with the correct ones in a localized burst of DNA synthesis. There are three modes of excision repair, each of which employs specialized sets of enzymes.
 1. **Base Excision Repair (BER)**
 2. **Nucleotide Excision Repair (NER)**
 3. **Mismatch Repair (MMR)**

Direct Reversal of Base Damage

Perhaps the most frequent cause of point mutations in humans is the spontaneous addition of a methyl group (CH₃-) (an example of alkylation) to Cs followed by deamination to a T. Fortunately, most of these changes are repaired by enzymes, called glycosylases, that remove the mismatched T restoring the correct C. This is done without the need to break the DNA backbone (in contrast to the mechanisms of excision repair described below).

Some of the drugs used in cancer chemotherapy ("chemo") also damage DNA by alkylation. Some of the methyl groups can be removed by a protein encoded by our **MGMT** gene. However, the protein can only do it once, so the removal of each methyl group requires another molecule of protein.

This illustrates a problem with direct reversal mechanisms of DNA repair: they are quite wasteful. Each of the myriad types of chemical alterations to bases requires its own mechanism to correct. What the

cell needs are more general mechanisms capable of correcting all sorts of chemical damage with a limited toolbox. This requirement is met by the mechanisms of **excision repair**.

Base Excision Repair (BER)

The steps and some key players:

1. removal of the damaged base (estimated to occur some 20,000 times a day in each cell in our body!) by a DNA glycosylase. We have at least 8 genes encoding different DNA glycosylases each enzyme responsible for identifying and removing a specific kind of base damage.
2. removal of its deoxyribose phosphate in the backbone, producing a gap. We have two genes encoding enzymes with this function.
3. replacement with the correct nucleotide. This relies on **DNA polymerase beta**, one of at least 11 DNA polymerases encoded by our genes.
4. ligation of the break in the strand. Two enzymes are known that can do this; both require ATP to provide the needed energy.

Nucleotide Excision Repair (NER)

NER differs from BER in several ways.

- It uses different enzymes.
- Even though there may be only a single "bad" base to correct, its nucleotide is removed along with many other adjacent nucleotides; that is, NER removes a large "patch" around the damage.

The steps and some key players:

1. The damage is recognized by one or more protein factors that assemble at the location.

2. The DNA is unwound producing a "bubble". The enzyme system that does this is **Transcription Factor IIH, TFIIH**, (which also functions in normal transcription).
3. Cuts are made on both the 3' side and the 5' side of the damaged area so the tract containing the damage can be removed.
4. A fresh burst of DNA synthesis — using the intact (opposite) strand as a template — fills in the correct nucleotides. The DNA polymerases responsible are designated polymerase **delta** and **epsilon**.
5. A **DNA ligase** covalently inserts the fresh piece into the backbone.

Xeroderma Pigmentosum (XP)

XP is a rare **inherited** disease of humans which, among other things, predisposes the patient to

- pigmented lesions on areas of the skin exposed to the sun and
- an elevated incidence of skin cancer.

It turns out that XP can be caused by mutations in any one of several genes — all of which have roles to play in **NER**. Some of them:

- **XPA**, which encodes a protein that binds the damaged site and helps assemble the other proteins needed for NER.
- **XPB** and **XPD**, which are part of TFIIH. Some mutations in XPB and XPD also produce signs of premature aging.
- **XPF**, which cuts the backbone on the 5' side of the damage
- **XPG**, which cuts the backbone on the 3' side.

Transcription-Coupled NER

Nucleotide-excision repair proceeds most rapidly

- in cells whose genes are being actively transcribed

- on the DNA strand that is serving as the template for transcription.

This enhancement of NER involves XPB, XPD, and several other gene products. The genes for two of them are designated **CSA** and **CSB** (mutations in them cause an inherited disorder called **Cockayne's syndrome**).

The CSB product associates in the nucleus with **RNA polymerase II**, the enzyme responsible for synthesizing **messenger RNA** (mRNA), providing a molecular link between transcription and repair.

One plausible scenario: If RNA polymerase II, tracking along the template (antisense) strand), encounters a damaged base, it can recruit other proteins, e.g., the CSA and CSB proteins, to make a quick fix before it moves on to complete transcription of the gene.

Mismatch Repair (MMR)

Mismatch repair deals with correcting mismatches of the **normal bases**; that is, failures to maintain normal Watson-Crick base pairing (A•T, C•G)

It can enlist the aid of enzymes involved in both base-excision repair (BER) and nucleotide-excision repair (NER) as well as using enzymes specialized for this function.

- **Recognition** of a mismatch requires several different proteins including one encoded by **MSH2**.
- **Cutting** the mismatch out also requires several proteins, including one encoded by **MLH1**.

Mutations in either of these genes predisposes the person to an inherited form of colon cancer. So these genes qualify as tumor suppressor genes.

Synthesis of the repair patch is done by **DNA polymerase delta**.

Cells also use the MMR system to enhance the fidelity of recombination; i.e., assure that only homologous regions of two DNA molecules pair up to crossover and recombine segments (e.g., in meiosis).

Repairing Strand Breaks

Ionizing radiation and certain chemicals can produce both single-strand breaks (**SSBs**) and double-strand breaks (**DSBs**) in the DNA backbone.

Single-Strand Breaks (SSBs)

Breaks in a single strand of the DNA molecule are repaired using the same enzyme systems that are used in Base-Excision Repair (BER).

Double-Strand Breaks (DSBs)

There are two mechanisms by which the cell attempts to repair a complete break in a DNA molecule:

- **Direct joining** of the broken ends. This requires proteins that recognize and bind to the exposed ends and bring them together for ligating. They would prefer to see some complementary nucleotides but can proceed without them so this type of joining is also called **Nonhomologous End-Joining (NHEJ)**.
- Errors in direct joining may be a cause of the various **translocations** that are associated with cancers.
- Examples:
 - Burkitt's lymphoma
 - the Philadelphia chromosome in chronic myelogenous leukemia (CML)
 - B-cell leukemia
- **Homologous Recombination**. Here the broken ends are repaired using the information on the intact

- **sister chromatid** (available in G_2 after chromosome duplication), **or** on the
- **homologous chromosome** (in G_1 ; that is, before each chromosome has been duplicated). This requires searching around in the nucleus for the homolog — a task sufficiently uncertain that G_1 cells usually prefer to mend their DSBs by **NHEJ**. **or** on the
- **same chromosome** if there are duplicate copies of the gene on the chromosome oriented in opposite directions (head-to-head or back-to-back).

Two of the proteins used in homologous recombination are encoded by the genes **BRCA1** and **BRCA2**. Inherited mutations in these genes predispose women to breast and ovarian cancers.

Cancer Chemotherapy

- The hallmark of all cancers is continuous cell division.
- Each division requires both
 - the replication of the cell's DNA (in S phase) and
 - transcription and translation of many genes needed for continued growth.
- So, any chemical that damages DNA has the potential to inhibit the spread of a cancer.
- Many (but not all) drugs used for cancer therapy do their work by damaging DNA.

The table lists (by trade name as well as generic name) some of the anticancer drugs that specifically target DNA.

6-mercaptopurine	Purinethol®	<u>purine</u> analog. One effect: substitutes for G, inducing abortive <u>MMR</u> and strand breaks
Gemcitabine	Gemzar®	<u>pyrimidine</u> analog substitutes for C

		blocking strand elongation
Cyclophosphamide	Cytoxan®	<u>alkylating agents</u> ; form interstrand and/or intrastrand crosslinks
Melphalan	Alkeran®	
Busulfan	Myleran®	
Chlorambucil	Leukeran®	
Mitomycin	Mutamycin®	
Cisplatin	Platinol®	forms crosslinks
Bleomycin	Blenoxane®	cuts DNA strands between GT or GC
Irinotecan	Camptosar®	inhibit the proper functioning of enzymes (topoisomerases) needed to unwind DNA for replication and transcription
Mitoxantrone	Novantrone®	
Dactinomycin	Cosmegen®	inserts into the double helix preventing its unwinding

Sadly,

- the cancer patient has many other cell types that are also proliferating rapidly, e.g., cells of the
 - intestinal endothelium
 - bone marrow
 - hair follicles

and anticancer drugs also damage these — producing many of the unpleasant side effects of "chemo".

- Agents that damage DNA are themselves carcinogenic, and chemotherapy poses a significant risk of creating a new cancer, often a leukemia.

Here is an overview.

5' ...A T G G C C T G G A C T T C A... 3' **Sense strand of DNA**
3' ...T A C C G G A C C T G A A G T... 5' **Antisense strand of DNA**

↓ **Transcription of antisense strand**

5' ...A U G G C C U G G A C U U C A... 3' **mRNA**

↓ **Translation of mRNA**

Met—Ala—Trp—Thr—Ser— **Peptide**