201: Applied Environmental Microbiology & Ecology

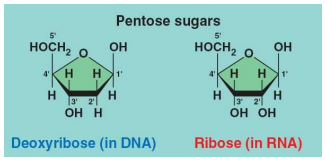
Microbial Molecular Biology

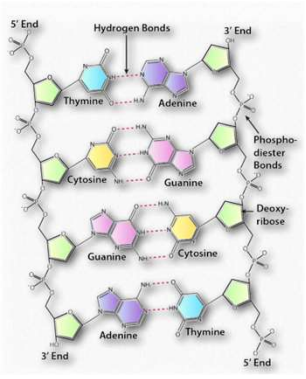
Prof. Suparna Mukherji

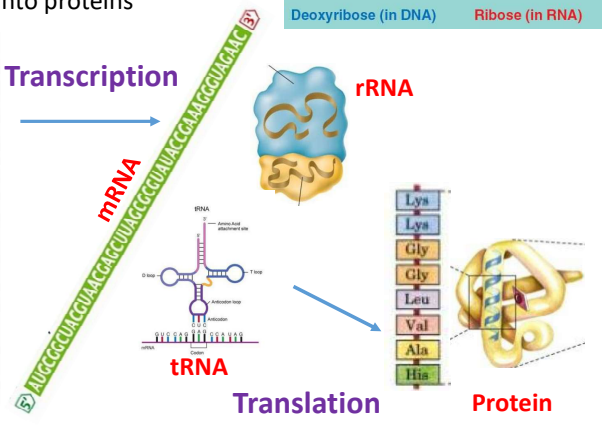
1

Introduction

- The flow of essential biological information—
  - Stable & inert DNA → proteins and enzymes
  - DNA replication
  - Transcription of DNA into RNA
  - Translation of RNA into proteins



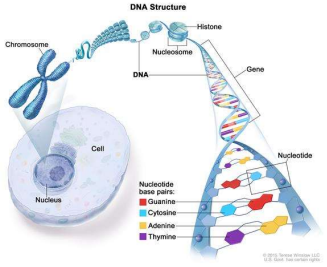




2

Macromolecules and Genes

- **Gene:** functional unit of genetic information
- **Genome**: Chromosomes or other large genetic elements containing the total complement of genetic information in the cell
- Genetic information contained in **DNA and RNA** by sequence of amino acids and proteins → **informational macromolecules**
- **DNA**: carries the genetic blueprint for the cell
- **RNA:** is an intermediary molecule that converts this blueprint into defined amino acid sequences in proteins.



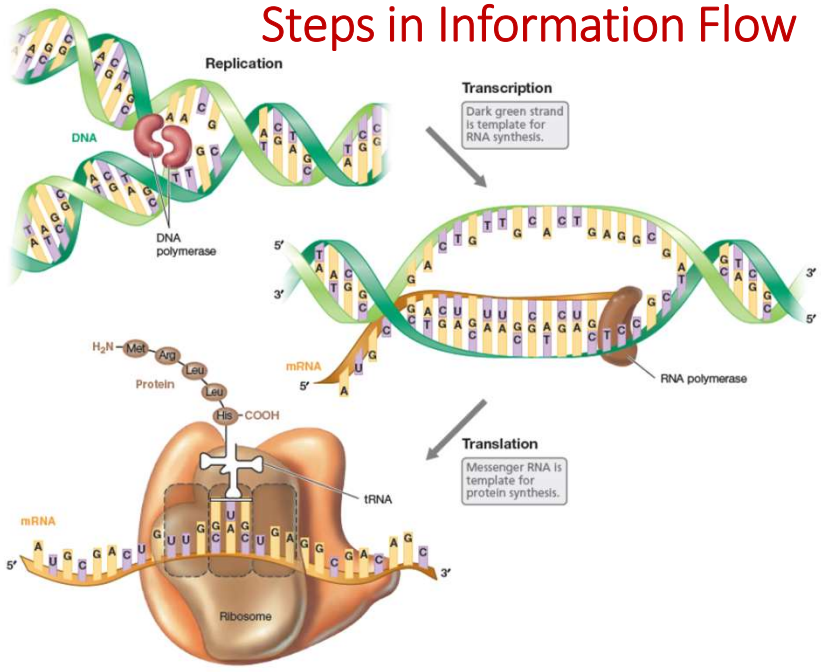
3

Steps in Information Flow

- When genes are expressed, the genetic information stored in DNA is transferred to RNA
  - **Replication**: DNA double helix is duplicated, producing two copies; Carried out *DNA polymerase*
  - **Transcription**: The transfer of genetic information from DNA to RNA Carried out by an enzyme called *RNA polymerase*
  - **Translation**: Synthesis of a protein using the genetic information in mRNA
- Many different RNA molecules can be transcribed from a relatively short region of the long DNA molecule
  - In eukaryotes, each gene is transcribed to yield a single mRNA (Monocistronic)
  - In prokaryotes, a single mRNA molecule may carry the genetic information from several genes; that is, several protein coding regions (Polycistronic)
- Each group of three bases **(codon)** on an mRNA molecule encodes a single amino acid.
  - Codons are translated polypeptides by ribosomes, tRNA, and helper proteins called *translation factors*

4

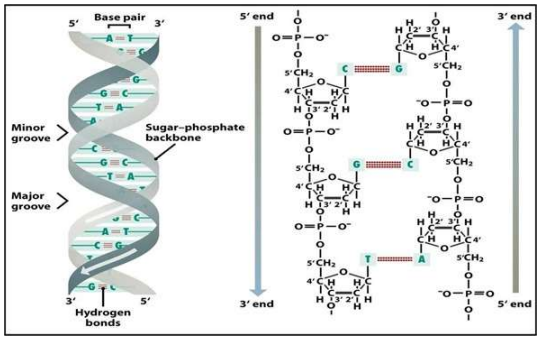
Steps in Information Flow



5

DNA- Double Helix

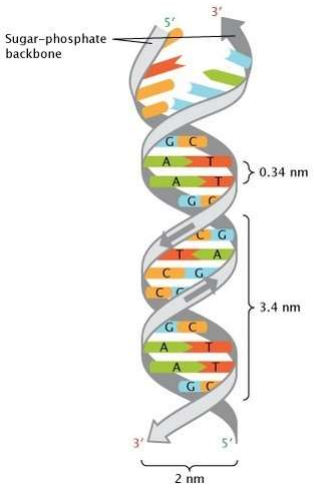
- DNA exists as a double-stranded polynucleotide strands whose base sequences with **complementary base sequences (A=T and C≡G)**.
- The two strands are arranged in an **antiparallel** fashion
- Individual hydrogen bonds are very weak → many such bonds between the base pairs confers stability to the molecule
- Two strands of DNA are wrapped around each other to form a double helix; Two distinct grooves formed: M*ajor groove* & *minor groove*→ Most proteins bind in the major groove.



6

Size and Shape of DNA Molecules

- Size of a DNA molecule → number of nucleotide bases or base pairs per molecule
  - A DNA molecule with 1000 bases is 1 kilobase (kb) of DNA
  - If the DNA is a double helix, then *kilobase pairs* (kbp) is used
  - If large genomes → *megabase pair* (Mbp) for a million base pairs is used. Genome of *E. coli* is 4.64 Mbp
- Each base pair is 0.34 nm in length along the double helix, and each turn of the helix contains approximately 10 base pairs



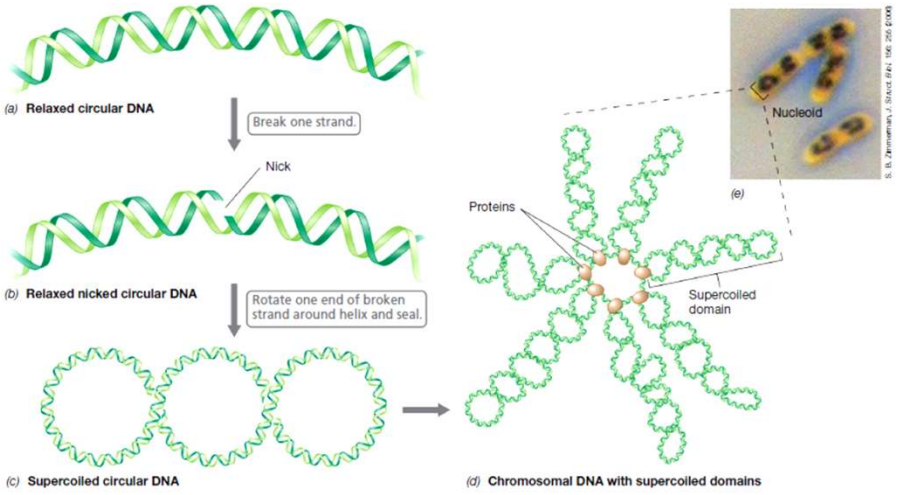
7

Supercoiling of DNA

- The twisting of DNA into folded and compact structure is called as supercoiling
- Circular DNA molecule linearization → Relaxed DNA
  - When relaxed, a DNA molecule has the exact number of turns of the helix predicted from the number of base pairs
- Supercoiled DNA: Positive/Negative supercoiled DNA
  - **Positive supercoiling**: the double helix is overwound → Contains more than the natural number of turns → common in thermophilic archaea
  - **Negative supercoiling**: the double helix is underwound (contains fewer than the natural number of turns→ More predominant
  - In the *Escherichia coli* chromosome, more than 100 supercoiled domains are thought to exist, each stabilized by specific proteins bound to the DNA.
- Negative supercoiling results when the DNA is twisted about its axis in the opposite sense from the right-handed double helix.
- Inserting supercoils into DNA requires energy from ATP, whereas releasing supercoils does not.

8

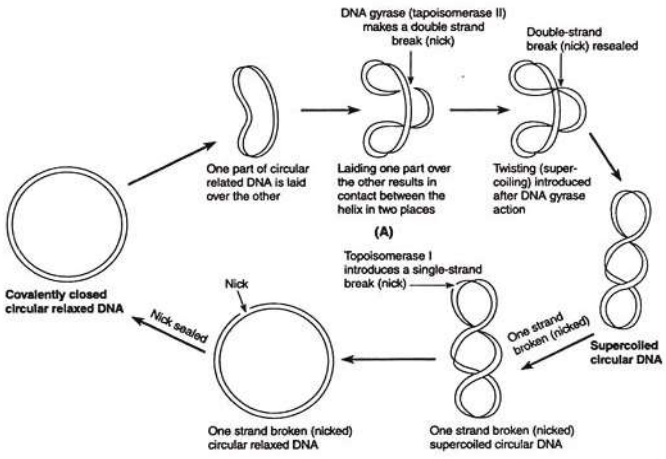
Supercoiling of DNA



9

Supercoiling of DNA

- **Supercoils are inserted or removed in DNA by enzymes called *topoisomerases***
- **DNA gyrase** is a type II topoisomerase that inserts negative supercoils



10

Genetic Elements: Chromosomes and Plasmids

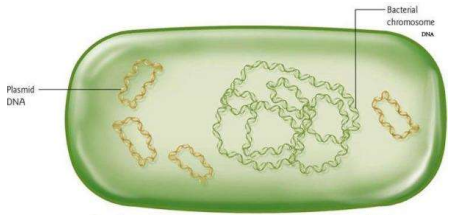
- ***Genetic elements:*** Structures containing genetic material (mostly DNA). **Chromosome:** The main genetic element in prokaryotes Other genetic elements: *Virus genomes*, *plasmids*, *organellar genomes*, and *transposable elements*
- A typical prokaryote has a single circular DNA chromosome; Sometimes 2 or 3; Eukaryotic genomes have multiple chromosomes
- DNA in all eukaryotic chromosomes is linear in contrast to most prokaryotic chromosomes, which are circular DNA molecules

Organism	*Element*	*Type of nucleic acid*	*Description*
Prokaryote	Chromosome	Double-stranded DNA	Extremely long, usually circular
Eukaryote	Chromosome	Double-stranded DNA	Extremely long, linear
All organisms	Plasmid[a]	Double-stranded DNA	Relatively short circular or linear, extrachromosomal
All organisms	Transposable element	Double-stranded DNA	Always found inserted into another DNA molecule
Mitochondrion or chloroplast	Organellar genome	Double-stranded DNA	Medium length, usually circular
Virus	Virus genome	Single- or double-stranded DNA or RNA	Relatively short, circular or linear

11

Genetic Elements

- Viruses contain genomes, of *either* DNA or RNA
- **Plasmids** replicate separately from the chromosome. **Mostly double stranded DNA; mostly circular, some are linear; typically much smaller than chromosomes**
- **Transposable elements:** segments of DNA that can move from one site on a DNA molecule to another, either on the same molecule or on a different DNA molecule. They are always found inserted into a DNA molecules, cannot exist as separate molecules of DNA
  - Chromosomes, plasmids, virus genomes, and any other type of DNA molecule may act as a host for a transposable element.



12

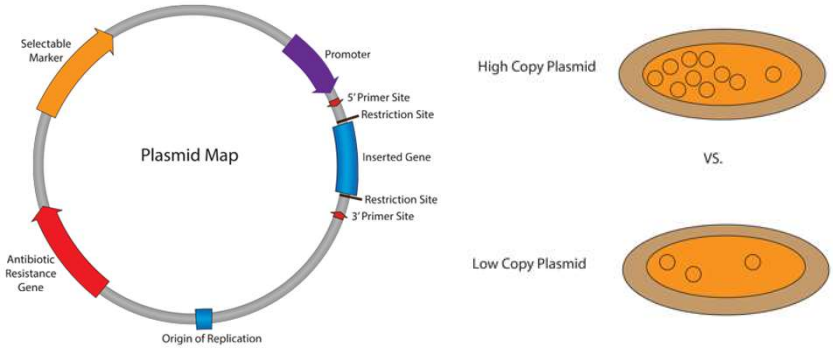
Plasmids

- They possess their own origin of replication → do rely on chromosomally encoded enzymes for their replication
- Most plasmids are usually expendable since they rarely contain genes required for growth under all conditions
- They do not have an extracellular form and exist inside cells as free DNA; Thousands of different plasmids are known
  - *E.coli* only contains 300 types of plasmids
- Naturally occurring plasmids vary in size from approximately 1 kbp to >1 Mbp. Typical less than 5% the size of the chromosome
- Some bacteria may contain several different types of plasmids.
  - *Borrelia burgdorferi* contains 17 different circular and linear plasmids

13

Plasmids

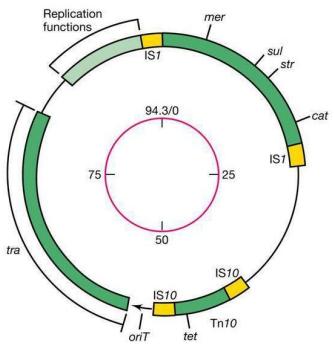
- Cellular replication enzymes also replicate plasmids. The genes encoded by a plasmid direct the initiation of replication and partitioning of replicated plasmids between daughter cells
- Different plasmids may be present in cells in different numbers called *copy number* (1-100 copy numbers).
- Copy number is controlled by genes on the plasmid and by interactions between the host and the plasmid



14

Types of Plasmids & their Function

- May carry genes that profoundly influence host cell physiology
- **Resistance plasmids (*R plasmids)* confers resistance to antibiotics** or other growth inhibitors → may encode proteins that inactivate the antibiotic
- Several antibiotic resistance genes can be encoded on a single R plasmid, while a cell with multiple resistance may contain several different R plasmids.
- Plasmid R100 (94.3 kbp) encodes resistance to sulfonamides, streptomycin, spectinomycin, fusidic acid, chloramphenicol, and tetracycline. It also encodes resistance to mercury.
- Increase in pathogenic bacteria resistant to antibiotics can be correlated with increasing use of antibiotics for treating infectious.
- Plasmid released by death & disintegration of a host may be picked up by another host; or plasmid transfer may happen through cell to cell contact, i.e, conjugation



Genetic Map of R100

15

Functions of Plasmid

- Pathogenic microorganisms possess a variety of characteristics that enable them to colonize hosts and establish infections. These are mostly plasmid encoded:
  - ✓ ability of the pathogen to attach to and colonize specific host tissue
  - ✓ production of toxins, enzymes, and other molecules that cause damage to the host
- Many bacteria also produce proteins (**bacteriocins,** analogous to antibiotics) that inhibit or kill closely related species or even different strains of the same species. Genes responsible for bacteriocins and their activity are found in plasmids
  - ✓ *E. coli* produces bacteriocins called *colicins* that bind to specific receptors on the surface of susceptible cells and disrupt their membrane function
  - ✓ Other colicins are nucleases that degrade the DNA or RNA of susceptible strains
- Plasmids can encode properties fundamental to the ecology of the bacterium.
  - ✓ *Rhizobium* requires certain plasmid functions to interact with plants and form nitrogen-fixing root nodules
- Can confer special metabolic advantage →Ability to degrade toxic pollutants

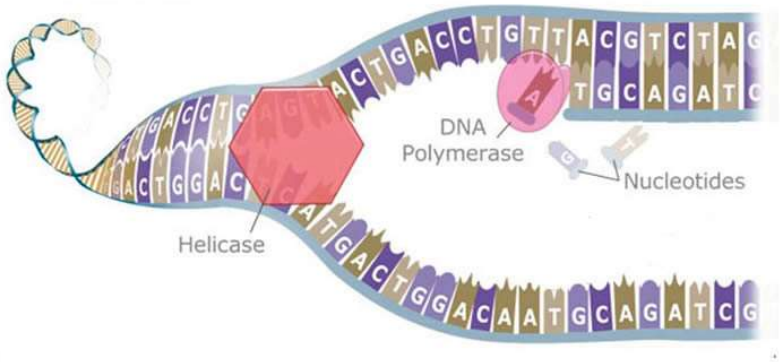
16

Phenotype class	*Organisms*[a]
Antibiotic production	*Streptomyces*
Conjugation	Wide range of bacteria
Metabolic functions	
Degradation of octane, camphor, naphthalene	*Pseudomonas*
Degradation of herbicides	*Alcaligenes*
Formation of acetone and butanol	*Clostridium*
Lactose, sucrose, citrate, or urea utilization	Enteric bacteria
Pigment production	*Erwinia, Staphylococcus*
Resistance	
Antibiotic resistance	Wide range of bacteria
Resistance to toxic metals	Wide range of bacteria
Virulence	
Tumor production in plants	*Agrobacterium*
Nodulation and symbiotic nitrogen fixation	*Rhizobium*
Bacteriocin production and resistance	Wide range of bacteria
Animal cell invasion	*Salmonella, Shigella, Yersinia*
Coagulase, hemolysin, enterotoxin	*Staphylococcus*
Toxins and capsule	*Bacillus anthracis*
Enterotoxin, K antigen	*Escherichia*

17

Transmission of Genetic Information: DNA Replication

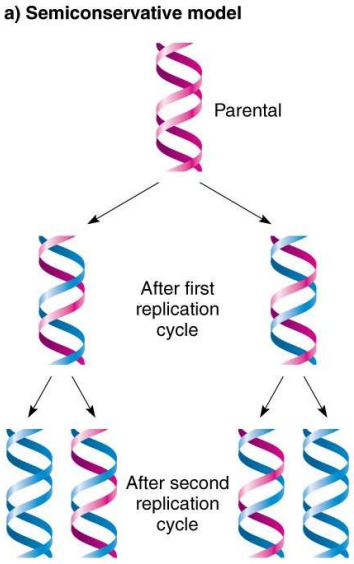
- DNA replication is necessary for cells to divide, whether to reproduce new organisms (in unicellular), or to produce new cells (multicellular).
- This process requires the activities of a host of special enzymes



18

Templates and Enzymes

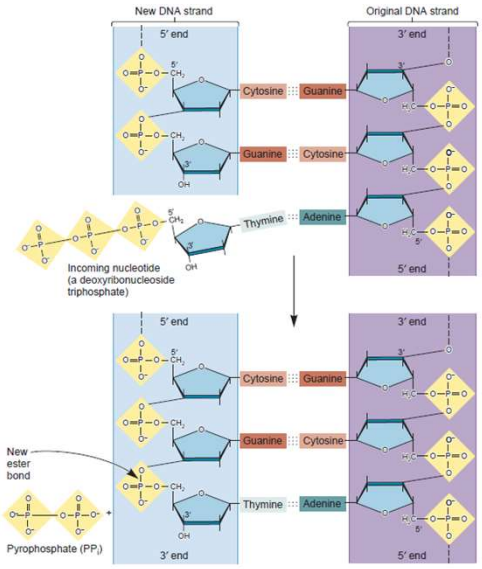
- **Semiconservative process**: Each parental strand is a template for one newly synthesized strand. Each double helices has one parental strand
- **Template strand:** The DNA strand that is used to make a complementary daughter strand
- The precursor of each new nucleotide in the DNA strand is a **deoxynucleoside 5'-triphosphate**
- During insertion, the two terminal phosphates are removed and the innermost phosphate is then covalently bound to a deoxyribose of the growing chain
- This addition of the incoming nucleotide requires the presence of a free hydroxyl group, → available only at the 3' end of the molecule.



19

DNA Replication

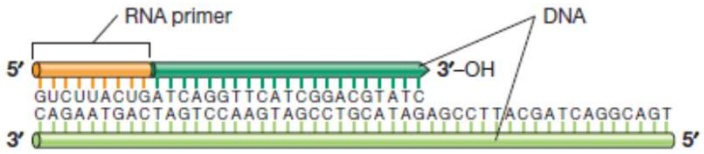
- DNA replication always proceeds ***from the 5' end to the 3' end,*** the 5'-phosphate of the incoming nucleotide being attached to the 3'-hydroxyl of the previously added nucleotide.
- Enzymes that catalyze the addition of deoxynucleotides →**DNA polymerases.**
- 5 DNA polymerases (I to V) in *Escherichia coli*
- **DNA polymerase III** (DNA Pol III) is the primary enzyme for replicating chromosomal DNA
- It cannot initiate a new chain, it can only add a nucleotide onto a preexisting 3'-OH group



20

Enzymes in DNA Replication

- **DNA polymerase I** (DNA Pol I) also participates in chromosomal replication. The other DNA polymerases only help repair damaged DNA
- To start a new chain, a **primer** (short stretch of RNA) is needed→ DNA polymerase can attach the first nucleotide to a primer
- **Primase:** RNA-polymerizing enzyme that makes the RNA primer
  - Synthesizes a short stretch (11–12 nucleotides) of RNA that is complementary in base pairing to the template strand DNA
- At the growing end of this RNA primer is a 3′-OH group to which DNA polymerase adds the first deoxyribonucleotide. The primer is eventually removed and replaced with DNA.



21

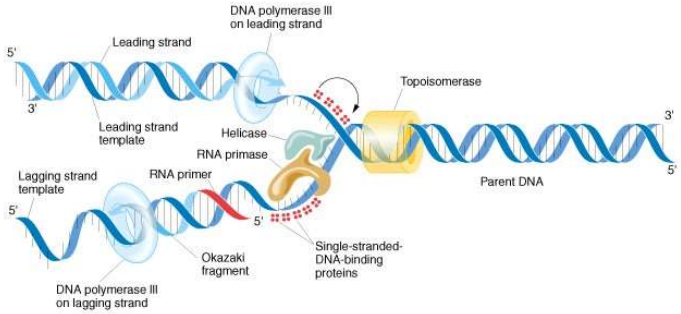
Initiation of DNA Synthesis

- The double helix must be unwound to expose the template strands. The zone of unwound DNA where replication occurs is called the **replication fork**.
- **DNA helicase** unwinds the double helix, using energy from ATP, and exposes a short single-stranded region
- Helicase moves along the DNA and separates the strands.
- The single stranded region is immediately covered with copies of single strand binding protein to stabilize the single-stranded DNA and prevent the double helix from re-forming.
- Unwinding of the double helix by helicase generates positive supercoils ahead of the advancing replication fork
- DNA gyrase travels along the DNA ahead of the replication fork and inserts negative supercoils to cancel out the positive supercoiling
- **Origin of replication (*oriC*):** Bacteria possess a single location on the chromosome where DNA synthesis is initiated (Specific sequence of ~250bp)
- **OriC** is recognized by initiation proteins DnaA, which binds to ori region and opens up the double helix.

22

Initiation of DNA Synthesis

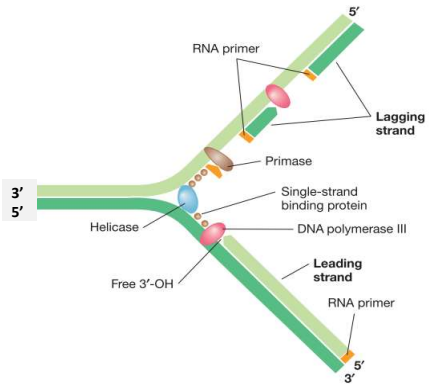
- Helicase (known as DnaB), which is helped onto the DNA by the helicase loader protein (DnaC). Two helicases are loaded, one onto each strand, facing in opposite directions
- Next, two primase and then two DNA polymerase enzymes are loaded onto the DNA behind the helicases
- Initiation of DNA replication then begins on the two single strands. As replication proceeds, the replication fork appears to move along the DNA



23

Leading and Lagging Strands

- **Leading strand:** The strand growing from 5′-PO_4^{2-} to the 3′-OH
  - DNA synthesis occurs *continuously* due to presence a free 3′-OH at the replication fork to which a new nucleotide can be added
- **Lagging strand**, DNA synthesis occurs *discontinuously* because there is no 3′-OH at the replication fork; it is at the opposite end
- **On the lagging strand, RNA primers must be synthesized by primase multiple times** to provide free 3′-OH groups for DNA Pol III. The leading strand is primed only once, at the origin
- The lagging strand is made in short segments, called ***Okazaki fragments.*** These lagging strand fragments are joined together later to yield a continuous strand of DNA



24

Enzymes Involved in DNA Replication

Prokaryotic DNA Replication: Enzymes and Their Function	
Enzyme/protein	**Specific Function**
DNA pol I	Exonuclease activity removes RNA primer and replaces with newly synthesized DNA
DNA pol II	Repair function
DNA pol III	Main enzyme that adds nucleotides in the 5'-3' direction
Helicase	Opens the DNA helix by breaking hydrogen bonds between the nitrogenous bases
Ligase	Seals the gaps between the Okazaki fragments to create one continuous DNA strand
Primase	Synthesizes RNA primers needed to start replication
Sliding Clamp	Helps to hold the DNA polymerase in place when nucleotides are being added
Topoisomerase	Helps relieve the stress on DNA when unwinding by causing breaks and then resealing the DNA
Single-strand binding proteins (SSB)	Binds to single-stranded DNA to avoid DNA rewinding back.

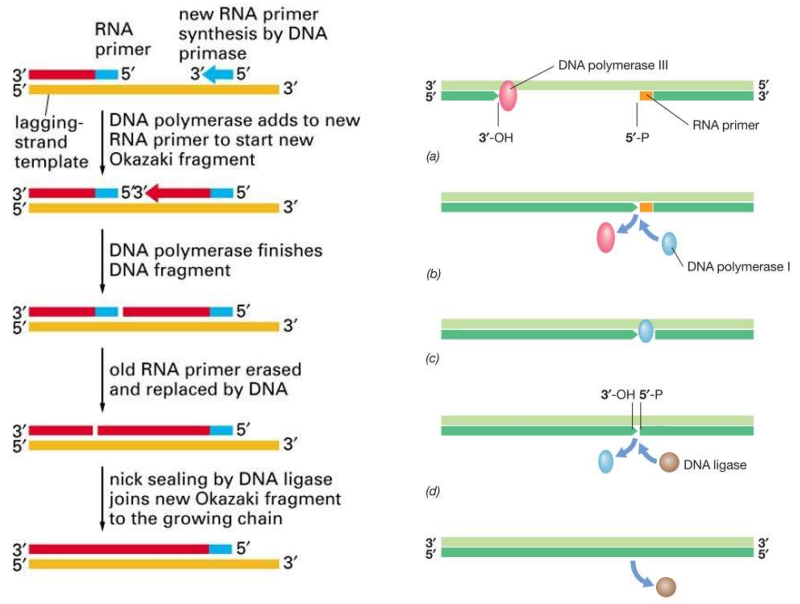
25

Synthesis of the New DNA Strands

- **After synthesizing the RNA primer, primase is replaced by DNA Pol III** which is a complex of several proteins including the polymerase core enzyme
- Each molecule of polymerase is held on the DNA by a sliding clamp, which encircles and slides along the single template strands of DNA
- **The replication fork contains two polymerase core enzymes and two sliding clamps, one set for each strand. There is only a single clamp-loader complex,** which functions to assemble the two sliding clamps onto the DNA
- On the lagging strand, DNA Pol III adds deoxyribonucleotides sequentially until it reaches the previously synthesized DNA. It then stops.
- **Besides synthesizing DNA, DNA Pol I has a 5' to 3' exonuclease activity** that removes the RNA primer preceding it. When the primer has been removed and replaced with DNA, DNA Pol I is released
- The very last phosphodiester bond is **DNA ligase**. This enzyme seals nicks in DNAs that have an adjacent 5'-PO_4^{2-} and 3'-OH and along with DNA Pol I, it also participates in DNA repair

26

Lagging Strand: Extension, Primer Removal & Resealing



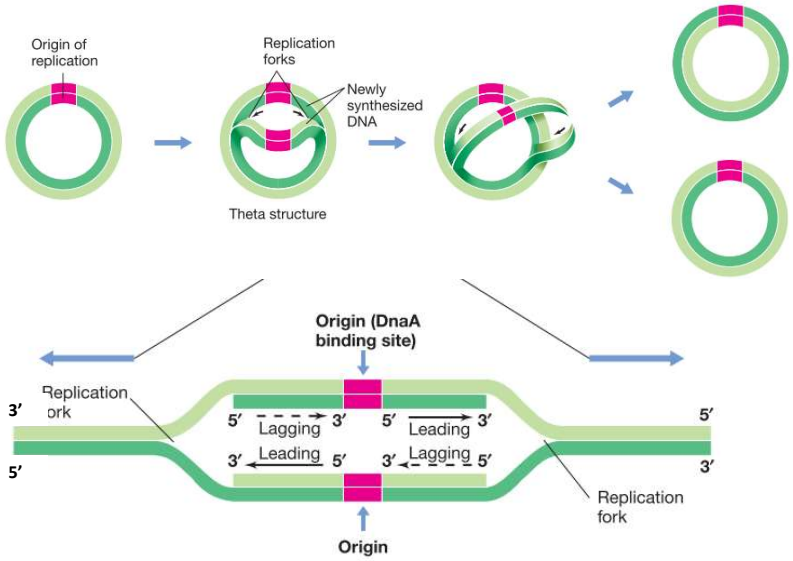
27

Bidirectional Replication

- In prokaryotes with circular chromosomes DNA replication is *bidirectional* from the origin of replication
- There are *two* replication forks on each chromosome, each moving in opposite directions. These are held together by the two **Tau protein subunits**.
- In circular DNA, bidirectional replication leads to the formation of characteristic shapes called **theta structures**.
- Synthesis occurs in both a leading and lagging fashion on each template strand, allowing DNA to replicate as rapidly as possible
- DNA Pol III can add nucleotides to a growing DNA strand at the rate of about 1000 per second, chromosome replication in *E. coli* still takes about 40 min & cell division 20 min
- If doubling time is < 40 min, chromosome may contain multiple DNA replication forks. A new round of DNA replication begins before the last round has been completed

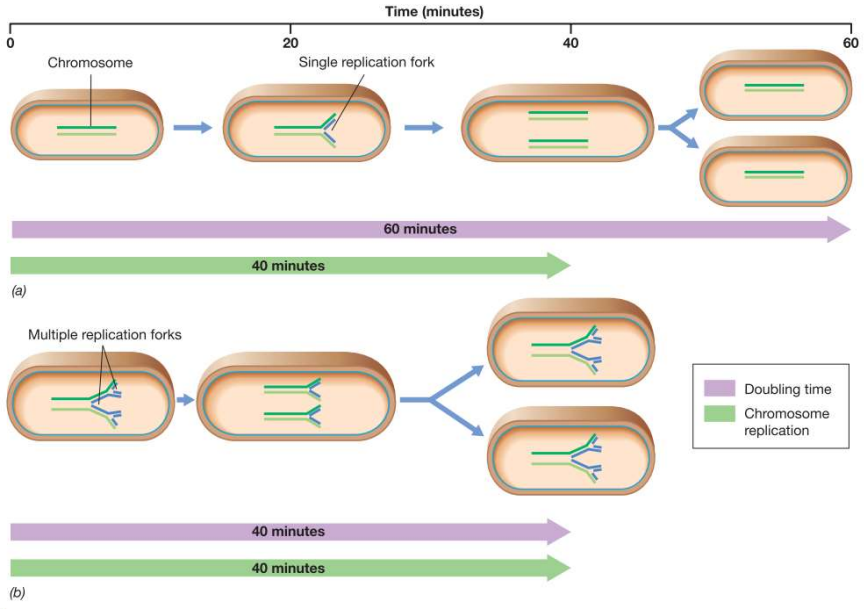
28

Bidirectional Replication



29

Cell Division Vs DNA Replication



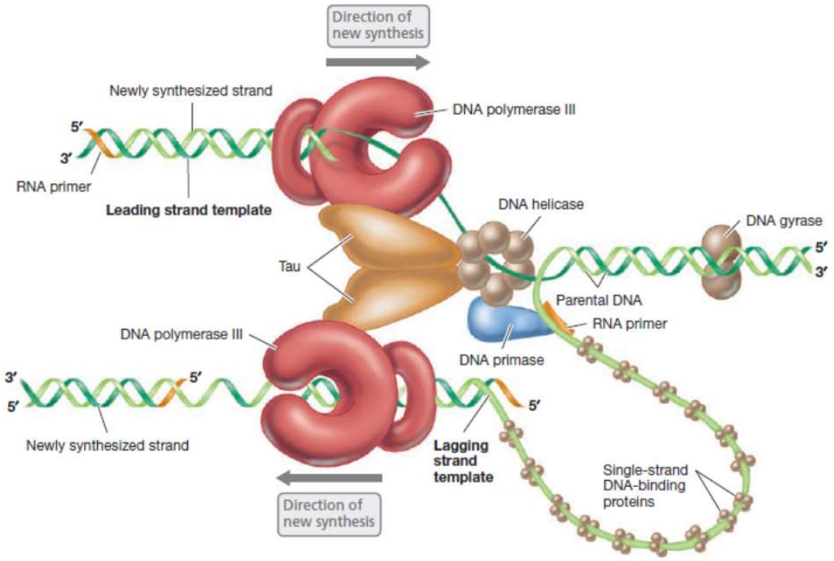
30

The Replisome

- **The enzymes involved in DNA replication does not act independently.** Replication proteins aggregate to form a large replication complex called the **replisome.**
- The lagging strand of DNA actually loops out to allow the replisome to move smoothly along both strands, and the replisome literally pulls the DNA template through it as replication occurs
- The replisome contains key replication proteins
  - ✓ DNA gyrase, which removes supercoils
  - ✓ DNA polymerase III
  - ✓ Primosome: a subcomplex containing helicase & primase → works in close association with each other
  - ✓ Single-strand binding protein, which prevents the separated template strands from re-forming a double helix

31

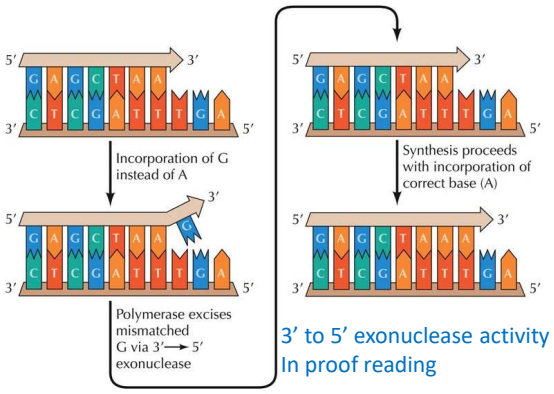
The Replisome



32

Fidelity of DNA Replication: Proofreading

- DNA replicates with a remarkably low error rate. Errors in DNA replication introduce *mutations*, changes in DNA sequence
- This accuracy can be due to complementary base pairing by DNA Pol III or due to second enzymatic activity of both DNA Pol I and Pol III, called *proofreading.*
- In DNA Pol III, a separate protein subunit, DnaQ, performs the proofreading function, whereas in DNA Pol I, a single protein performs both polymerization and proofreading
- Proofreading activity occurs because incorrect base pairing causes a slight distortion in the double helix

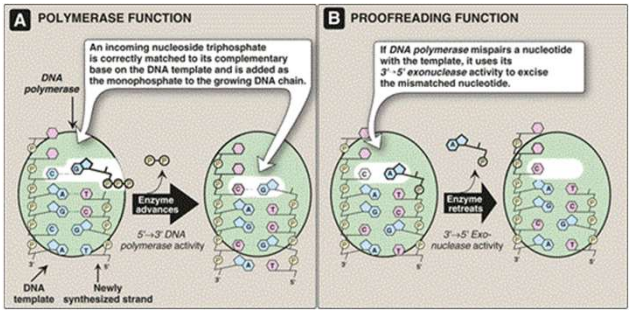


3’ to 5’ exonuclease activity In proof reading

This is distinct from the 5’ to 3’ exonuclease activity of DNA Pol I that removes primer

33

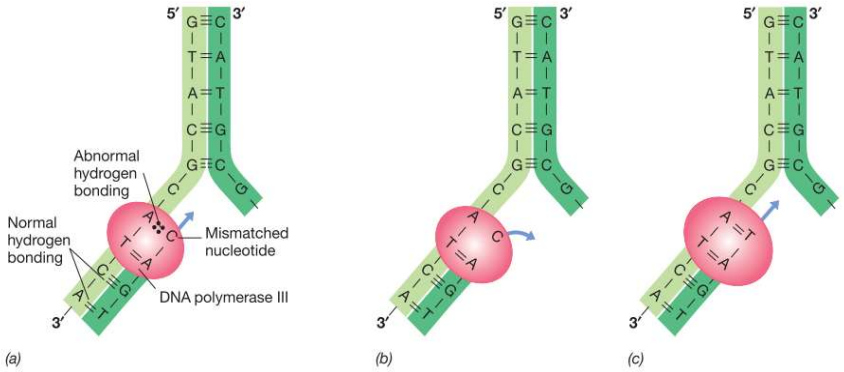
Fidelity of DNA Replication: Proofreading



- DNA Pol I and Pol III possess a **3’ to 5’ exonuclease activity** that can remove such wrongly inserted nucleotides
- After the removal of a mismatched nucleotide, the polymerase then gets a second chance to insert the correct nucleotide
- The proofreading exonuclease activity is distinct from the 5’ to 3’ exonuclease activity of DNA Pol I that removes the RNA primer from both the leading and lagging strands
- Exonuclease proofreading occurs in prokaryotes, eukaryotes, and viral DNA replication systems

34

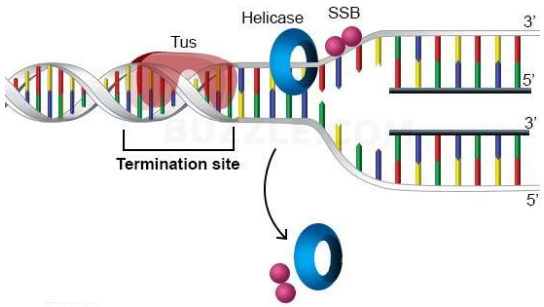
3’ to 5’ Exonuclease Activity



35

Termination of Replication

- On the opposite side of the circular chromosome from the origin is a site called the ***terminus of replication***
- The two replication forks collide as the new circles of DNA are completed
- In the terminus region several DNA sequences ***Ter*** **sites are recognized by a protein Tus, whose function is to block progress of the replication forks**
- When replication of the circular chromosome is complete, the two circular molecules are linked together
- After DNA replication, the DNA is partitioned so that each daughter cell receives a copy of the chromosome, with the help of cell division protein FtsZ.



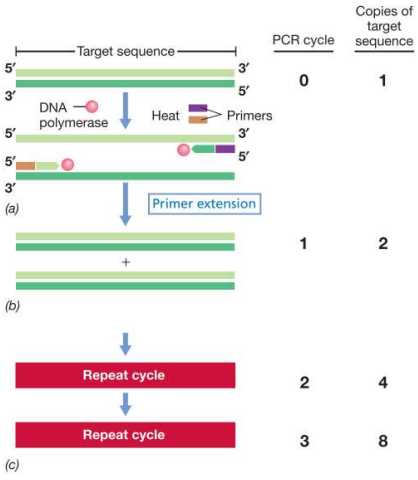
36

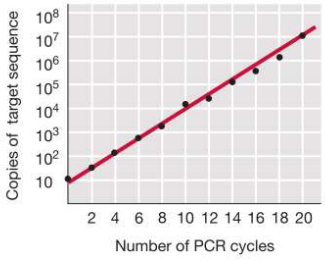
Polymerase Chain Reaction (PCR)

- In-vitro DNA replication conceived by Kary Mullins
- Copies a segment of DNA (the target of up to a few 1000 bp) up to a billion fold in a test tube→ Amplification→ Can produce a large amount of a specific gene
- DNA polymerase naturally copies the template DNA, Artificially synthesized primers initiate the process (composed of DNA not RNA)
- Steps in PCR
  - The template DNA is denatured by heating
  - Two artificial DNA oligonucleotide primers (flanking the target DNA) are provided in excess (the primers should attach to the template rather than to each other as cooling occurs)
  - DNA polymerase (Thermophiles aquaticus) extends the primer using the template DNA
  - After some time of incubation, mixture is heated again to separate the strands
  - Mix is cooled to allow the primer to hybridize with complementary regions of newly synthesized DNA
  - Steps are repeated
  - **20-30 cycles cause 10^6 to 10^9 fold increase in the target sequence**

37

Polymerase Chain Reaction (PCR)





38

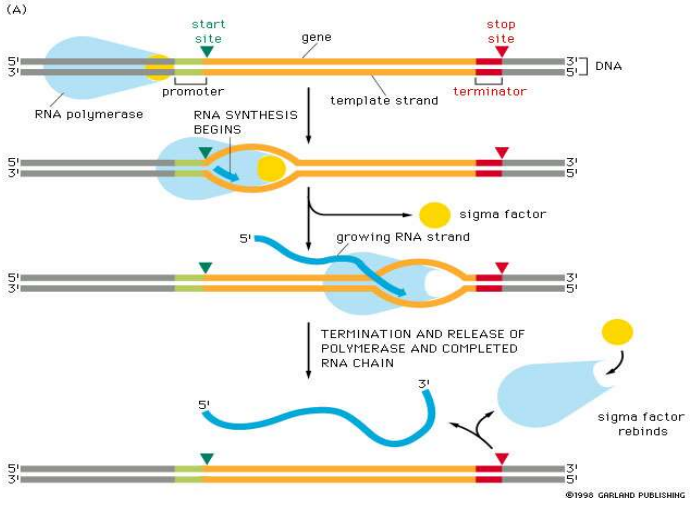
RNA Synthesis: Transcription

- Transcription is the synthesis of ribonucleic acid (RNA) using DNA as a template. Enzymes that act on DNA usually have no effect on RNA, and vice versa
- Transcription is catalyzed by the enzyme **RNA polymerase**. The mechanism of is similar; Phosphodiester bond formed between ribonucleotides.
- During elongation, ribonucleoside triphosphates are added to the 3'-OH of the preceding nucleotide. Thus chain growth is 5' to 3' and the newly synthesized strand of RNA is antiparallel to the DNA template strand it is transcribed from
- RNA polymerase uses DNA as a template, but only one of the two strands is transcribed for any given gene. DNA sequences on both strands are transcribed, although at different locations
- **RNA polymerase can initiate new strands of RNA on its own; no primer sequence is necessary**
- As the newly made RNA dissociates from the DNA, the opened DNA closes back into the original double helix
- Transcription stops at specific sites called *transcription terminators*

39

RNA Synthesis: Transcription

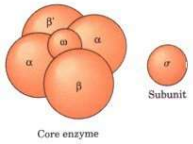
- Transcription copies much smaller units of DNA, often as little as a single gene. It allows the cell to transcribe different genes at different frequencies, depending on the needs of the cell for different proteins.



40

RNA Polymerases

- The RNA polymerase from *Bacteria* has five different subunits, β, β', α (2 copies), ω, and σ. The β and β' subunits are similar but not identical
- The subunits interact to form the active enzyme, called the RNA polymerase holoenzyme.
- σ is not as tightly bound as the others and easily dissociates, leading to the formation of the RNA polymerase core enzyme, $\alpha_2\beta\beta'\omega$
- The core enzyme alone synthesizes RNA while σ functions to recognize the appropriate site (promoter) on the DNA (to form RNA polymerase–DNA complex) for RNA synthesis to begin
- ω is needed for assembly of the core enzyme but not for RNA synthesis
- σ dissociates from the bacterial RNA polymerase holoenzyme once a short stretch of RNA has been formed
- Elongation of the RNA molecule is then catalyzed by the core enzyme alone

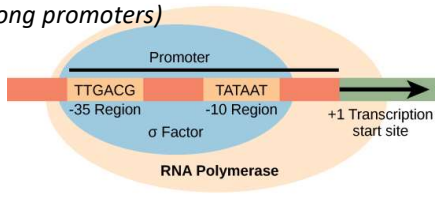


41

Promoters

- In *Bacteria*, promoters are recognized by the sigma subunit of RNA polymerase
- In this process, the DNA double helix at the promoter is opened up by the RNA polymerase to form a **transcription bubble**
- As the polymerase moves, it unwinds the DNA in short segments
- This transient unwinding exposes the template strand and allows it to be copied into the RNA complement
- If DNA has two nearby promoters pointing in opposite directions, then transcription from one will proceed in one direction (on one DNA strand) while from other promoter will proceed in the opposite direction (on the other strand)
- In *E. coli*, promoters that are most like the consensus sequence are usually more effective in binding RNA polymerase (*strong promoters)*

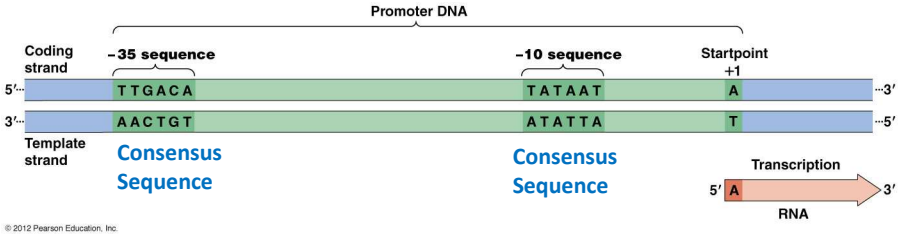
Promoters are specific DNA sequences that binds RNA polymerase



42

Sigma Factors and Consensus Sequences

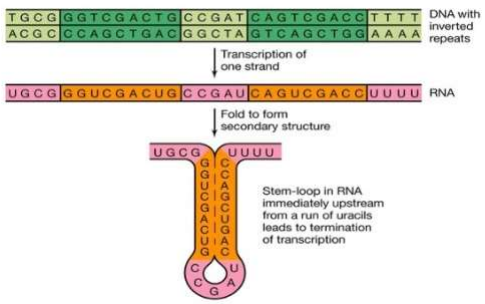
- Sigma (σ^{70} , 70 kDa protein) recognizes two highly conserved shorter sequences within the promoter
- One is 10 bases before the transcription start, the **-10 region, or *Pribnow box***, gives the consensus sequence: TATAAT
- The second conserved region is about 35 bases upstream, -35 region is TTGACA
- **Several alternative sigma factors exist that recognize different consensus sequences**
- Each alternative sigma factor is specific for a group of genes required under special circumstances and thus essential for regulating gene expression.

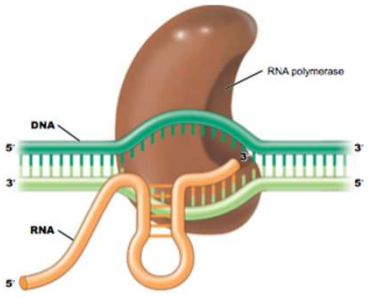


43

Termination of Transcription

- **Termination** of RNA synthesis is governed by specific base sequences on the DNA (a GC-rich sequence containing an inverted repeat with a central nonrepeating segment)
- **When such a DNA sequence is transcribed, the RNA forms a stem–loop structure by intra-strand base pairing**
- **Stem–loops followed by a run of adenines in the DNA template** are effective transcription terminators due to the formation of a stretch of U:A base pairs which are very weak and thus dissociates



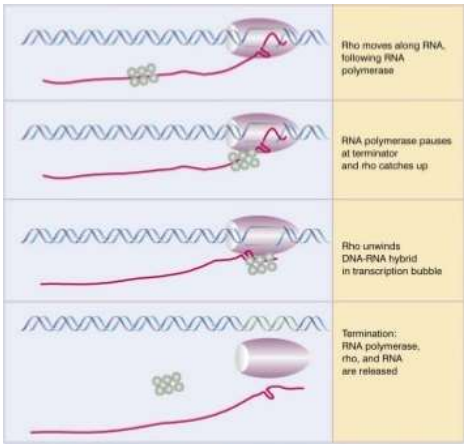


44

Termination of Transcription

- Transcription termination in *Bacteria* using a **specific protein factor, known as Rho.**

- Rho binds tightly to RNA and moves down the chain toward the RNA polymerase–DNA complex.
- RNA polymerase pauses at a Rho-dependent termination site (a specific sequence on the DNA template)
- Rho causes both the RNA and RNA polymerase to be released from the DNA, thus terminating transcription.



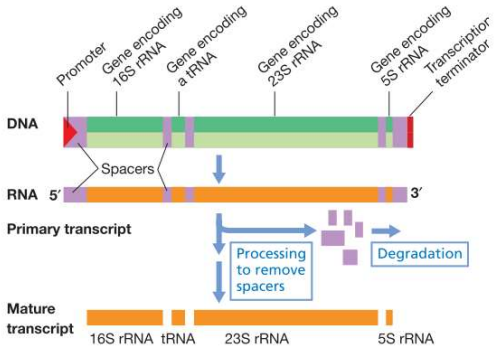
45

The Unit of Transcription

- **Transcriptional units** are segments of DNA that are transcribed into a single RNA molecule
- **Each is bound by transcription initiation & termination sites**
- **Codes for a single gene or two or more genes**, that are *co-transcribed*, to form a single RNA molecule
- Most genes encode proteins, but others encode non-translated RNAs, such as ribosomal RNA or transfer RNA
- *Bacteria* and *Archaea* produce 3 types of rRNA: 16S rRNA, 23S rRNA, and 5S rRNA. Transcription unit contains one gene for each of these rRNAs which are co-transcribed
- In prokaryotes tRNA genes are also often co-transcribed with each other or with genes for rRNA
- These transcripts are processed by specific proteins in the cell that cut them into individual units, yielding mature/functional rRNAs or tRNAs

46

The Unit of Transcription

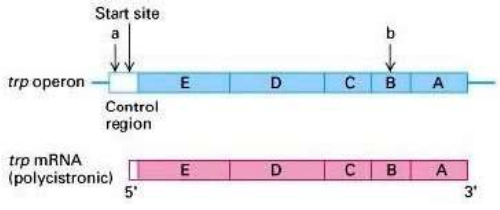


- In prokaryotes, most mRNAs have a short half-life after which they are degraded by enzymes called ***ribonucleases***
- rRNA and tRNA are stable RNAs, due to the highly folded secondary structures
- mRNA does not form such structures and is susceptible to ribonuclease attack
- The rapid turnover of mRNAs permits the cell to quickly adapt to new environmental conditions

47

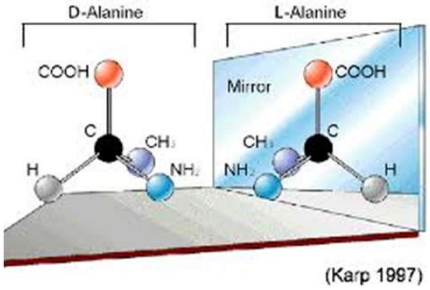
Polycistronic mRNA and the Operon

- In prokaryotes, genes that encode several enzymes of a particular metabolic pathway are often clustered together
- RNA polymerase transcribes the entire group of genes into a single, long mRNA molecule called as *polycistronic mRNA*
- Polycistronic mRNAs contain multiple *open reading frames* (that encodes amino acids). When this mRNA is translated, several polypeptides are synthesized sequentially by the same ribosome
- Operon: A group of related genes that are transcribed together to give a single polycistronic mRNA→ Allows coordinated expression of the genes together

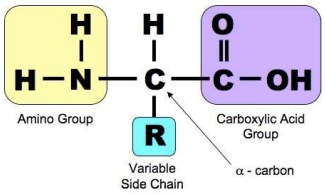


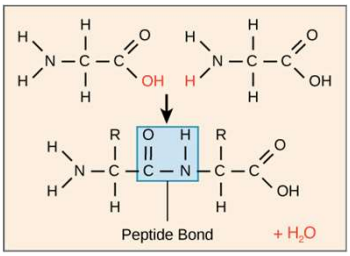
48

Amino Acids



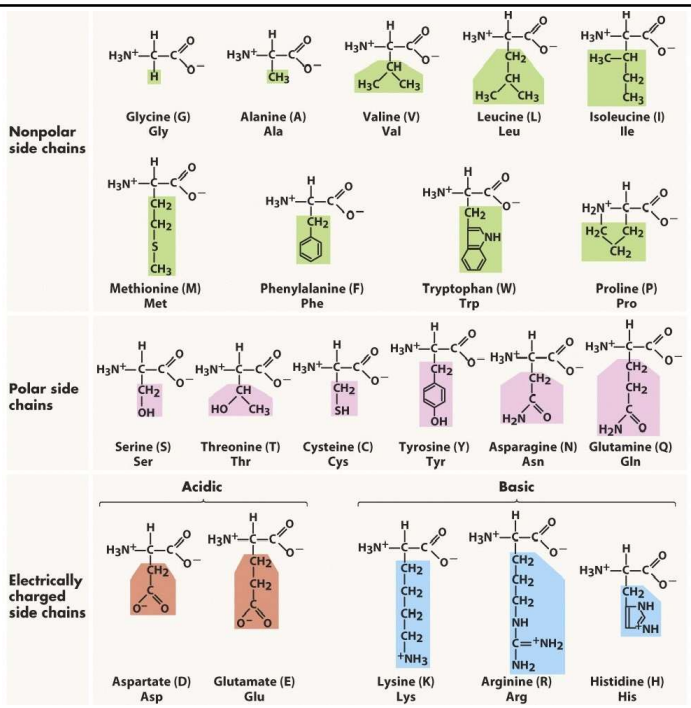
(Karp 1997)





- ✓ Most natural proteins have L-amino acids
- ✓ Side chain R containing –COOH groups are acidic amino acids (aspartic acid, glutamic acid
- ✓ Those with hydrophobic side chains→ Nonpolar amino acids
- ✓ Cysteine contains sulfhydryl group→ two cysteine can form disulphide linkage

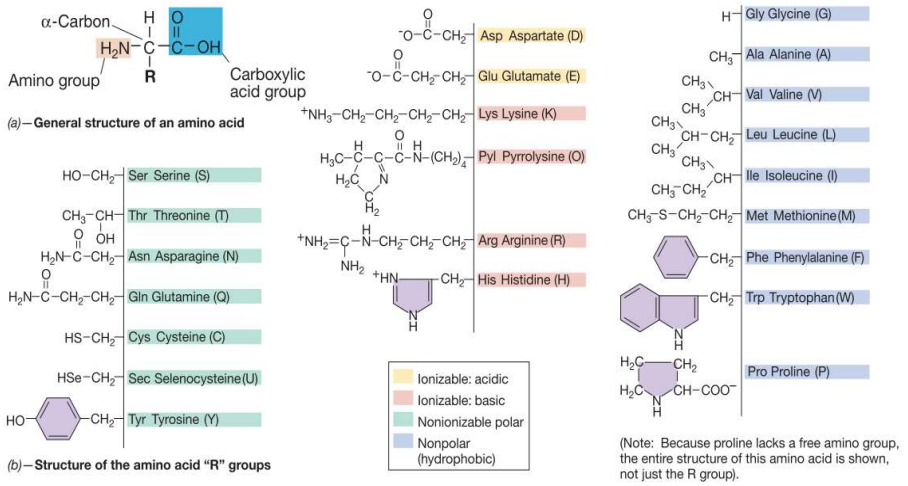
49



50

Structure of Genetically Encoded Amino Acids

- ✓ **Structure of the 22 genetically encoded amino acids**
- ✓ An average protein contains 300 amino acids → Enormous number of distinct proteins
- ✓ Polypeptide: C Terminus: Carboxylic acid & N Terminus: Amino group



51

Protein Synthesis: Translation

- Process involves formation of a polypeptide using a nucleic acid as template

- **Codon: An RNA triplet of three bases;** encodes a specific amino acid. The genetic code is RNA since, it is mRNA that is actually translated
- There are ~ 64 possible codons (4 bases taken three at a time = 4^3).
- Specific codons are used for starting and stopping translation

Second letter

First letter	U	C	A	G	Third letter
U	UUU Phe UUC Phe UUA Leu UUG Leu	UCU Ser UCC Ser UCA Ser UCG Ser	UAU Tyr UAC Tyr UAA Stop UAG Stop	UGU Cys UGC Cys UGA Stop UGG Trp	U C A G
C	CUU Leu CUC Leu CUA Leu CUG Leu	CCU Pro CCC Pro CCA Pro CCG Pro	CAU His CAC His CAA Gln CAG Gln	CGU Arg CGC Arg CGA Arg CGG Arg	U C A G
A	AUU Ile AUC Ile AUA Ile AUG Met	ACU Thr ACC Thr ACA Thr ACG Thr	AAU Asn AAC Asn AAA Lys AAG Lys	AGU Ser AGC Ser AGA Arg AGG Arg	U C A G
G	GUU Val GUC Val GUA Val GUG Val	GCU Ala GCC Ala GCA Ala GCG Ala	GAU Asp GAC Asp GAA Glu GAG Glu	GGU Gly GGC Gly GGA Gly GGG Gly	U C A G

52

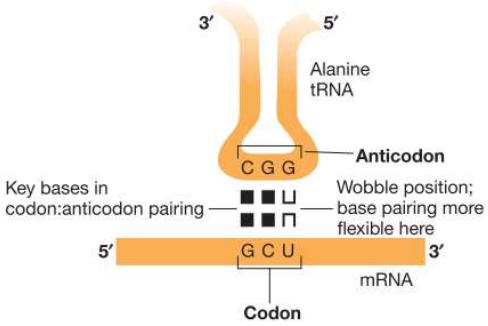
Properties of the Genetic Code

- 64 codons in mRNA codes for 22 amino acids; Consequently, amino acids can be encoded by more than 1 codon
  - **Knowing codon at a given location unambiguously identifies the corresponding amino acid, the reverse is not true**
- *Degenerate code*: When many codons code for single amino acid
- A codon is recognized by specific base pairing with a complementary sequence of three bases called the **anticodon**, which is part of tRNAs;
- **A specific t-RNA would be required to recognize each codon**; *E coli* has separate t-RNAs for the 5 Leucine codons.
- Out of codons that encode the same amino acid→ organism-specific **codon bias seen;** some codons are preferred over others

53

The Wobble Concept

Wobble: tRNA molecules form standard base pairs at the first two positions of the codon while tolerating irregular base pairing at the third position

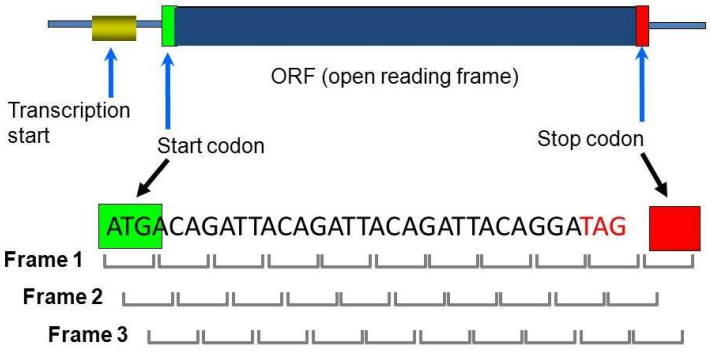


Some tRNAs can recognize more than one codon. There are two lysine codons in *E. coli*, there is only one lysyl tRNA, whose anticodon can base-pair with either AAA or AAG.

54

Open Reading Frames

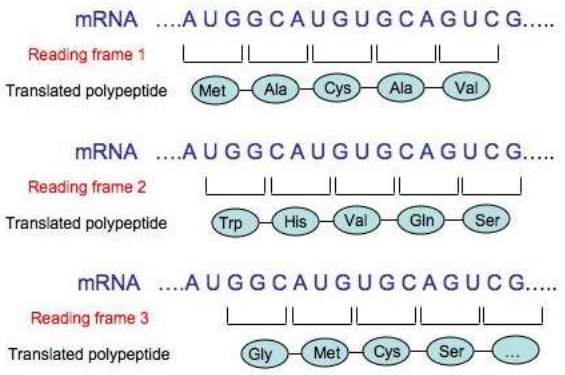
- **Open reading frames (ORFs)**: If an mRNA can be translated, it contains a start codon (typically AUG) followed by codons and then a stop codon in the same reading frame as the start codon.
- Only ORFs long enough to encode a functional polypeptide are accepted as true coding sequences, while a few protein hormones and regulatory peptides are much shorter



55

Start and Stop Codons

- In bacteria mRNA is translated beginning with its **start codon** AUG which encodes a chemically modified methionine: *N-formylmethionine*
- AUG at the *beginning* encodes *N*-formylmethionine, while *within* the coding region encodes methionine. Two different tRNAs are involved
- *Archaea* and *Eukarya* insert a regular methionine as the first amino acid in a polypeptide
- It is essential for ribosome to find the correct start codon to begin translation, if it does not, the whole reading frame of the mRNA will be shifted and thus an entirely different protein will be made



56

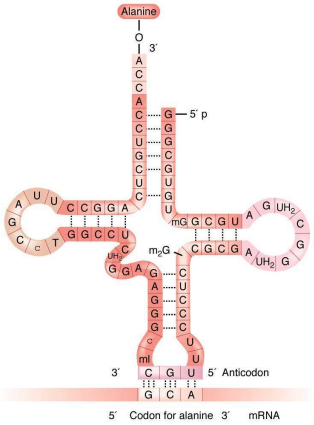
Start and Stop Codons

- **Reading frame fidelity: interactions between mRNA and rRNA within the ribosome**
- In prokaryotes, **rRNA recognizes a specific AUG on the mRNA as a start codon with the aid of an upstream sequence in the mRNA called the *ribosome-binding site* (RBS) or Shine–Dalgarno sequence**
- Some mRNAs can use other start codons, such as GUG, but still they incorporate ***N*-formylmethionine as the initiator amino acid**
- Few codons (**Stop codons, UAA, UAG, and UGA**) do not encode any amino acid, and they signal the termination of translation of a protein-coding sequence on the mRNA
- Stop codons are also called **nonsense codons**
  - Certain microorganism use certain nonsense codons to encode amino acids **(Example: animal mitochondria use the codon UGA to encode tryptophan instead of using it as a stop codon).**
  - Both selenocysteine and pyrrolysine (unusual and rare amino acids) are encoded by stop codons (UGA and UAG, respectively). **This switch is controlled by a recognition sequence just downstream of the now coding stop codon**

57

Transfer RNA

- A transfer RNA carries the **anticodon** that base-pairs with the codon on mRNA
- Each tRNA is specific for the amino acid that corresponds to its own anticodon **(its *cognate* amino acid)**
- **Aminoacyl-tRNA synthetases: Enzymes that ensure that each tRNA receives its correct amino acid**
  - For each amino acid, a separate aminoacyl-tRNA synthetase exists that specifically binds to both the amino acid and the tRNA with corresponding anticodons
  - they must recognize both a specific tRNA and its cognate amino acid
  - There are about 60 different tRNAs in bacterial cells and 100– 110 in mammalian cells



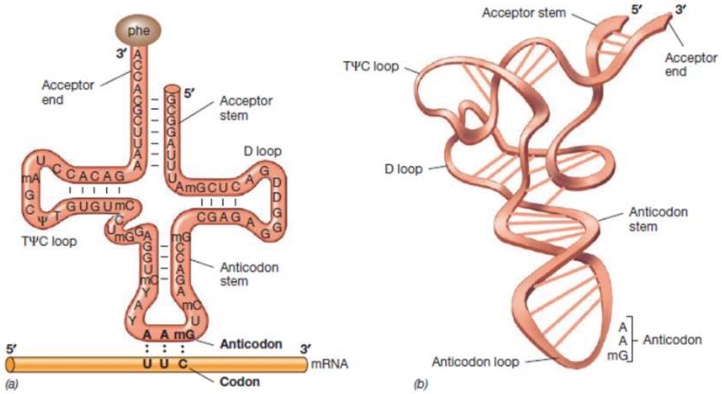
58

Structure of tRNA

- tRNA: Short, single-stranded molecules that contain extensive secondary structure; 73–93 nucleotides long
- Certain bases and secondary structures are constant for all tRNAs, whereas other parts are variable
- They contain some modified purine and pyrimidine bases: pseudouridine (ψ), inosine, dihydrouridine (D), ribothymidine, methyl guanosine, dimethyl guanosine, and methyl inosine
- The mature active tRNA **also contains extensive double-stranded regions within the molecule formed by internal base pairing** → tRNA can have a structure of cloverleaf
- Some regions of tRNA secondary structure are named after the modified bases found there **(for example, the TψC and D loops) or after their functions (for example, the anticodon loop and acceptor stem).**

59

Structure of tRNA



60

The Anticodon & Amino Acid Binding Site

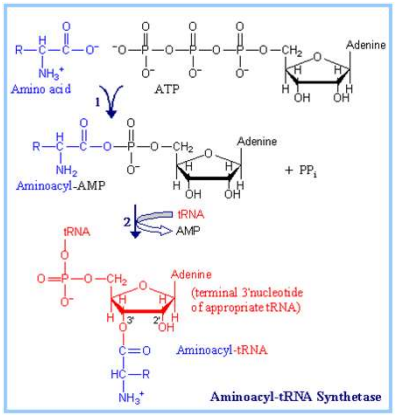
- *Anticodon*: the group of 3 bases that recognizes the codon on the mRNA
- Other portions of the tRNA interact with both the rRNA and ribosomal protein, non-ribosomal translation protein & the **aminoacyl-tRNA synthetase enzyme**
- At the **3'end (acceptor stem)** of all tRNAs are three unpaired nucleotides (CCA), which is essential for its function
- In most organisms the 3' CCA is not encoded in the tRNA gene on the chromosome, each nucleotide is added one by one by a protein called *CCA adding enzyme*, using CTP and ATP as substrates
- **The cognate amino acid is then covalently attached to the terminal adenosine of the CCA end of its corresponding tRNA by an ester linkage to the ribose sugar→ This is an activated tRNA**

61

Recognition, Activation & Charging

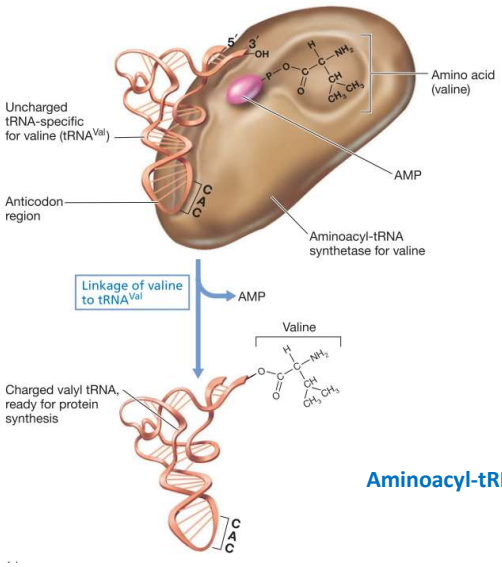
- The anticodon of the tRNA is important for recognition by the synthetase
- The other key recognition nucleotides are often part of the acceptor stem or D loop of the tRNA

- The **aminoacyl-AMP intermediate** formed normally remains bound to the tRNA synthetase until collision with the appropriate tRNA molecule
- The pyrophosphate (PPi) formed in the first reaction is split by a pyrophosphatase, giving two molecules of inorganic phosphate
- After activation and charging the aminoacyl-tRNA leaves the synthetase until it is bound by a ribosome where actual polypeptide synthesis occurs



62

Charging of tRNA

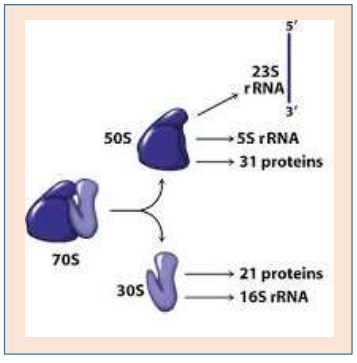


Aminoacyl-tRNA synthetase enzyme

63

Ribosomes in Prokaryotes

- **Ribosomes** are the sites of protein synthesis
- Each ribosome consists of two subunits, 30S & 50S subunits that yield the intact 70S ribosomes
- The S-values are *Svedberg units*→ sedimentation coefficients of the subunits or intact ribosomes when subjected to centrifugal force in an ultracentrifuge
- The ribosome subunits alternately associate and dissociate and also interact with many other proteins
- Several proteins (translation factors) that are essential for ribosome function interact with the ribosome at various stages of translation



64