Karpagam acadmy of higher educaton
Department of biochemistry
Molecular biology (18BCP202) MCQ
UNIT I

S.NO	UNIT	QUESTION	OPTION 1	OPTION 2	OPTION 3	OPTION 4	ANSWER
1	I	The macromolecule regarded as the reserve bank of genetic information	Protein	**DNA**	RNA	Polysaccharide	**DNA**
2	I	The total genetic information contained in a DNA is referred to as	gene	**genome**	Okazaki piece	Ribozyme	**genome**
3	I	Pseudogenes are	**Related to non functional genes**	Transcribed into mRNA	Translated in to functional proteins	Transcribed into tRNA	**Related to non functional genes**
4	I	Mobile genetic elements were visualized by	T.H Morgan	**Barbara McClintock**	G Khorana	C.B Bridge	**Barbara McClintock**
5	I	Mutation in which a purine base is substituted by a pyrimidine base or vice versa is	Transformation	Transition	**Transversion**	Transduction	**Transversion**
6	I	Mutation which results in the termination of	leaky mutation	**nonsense mutation**	point mutation	silent mutation	**nonsense mutation**
7	I	Fundamental unit of DNA organization	Replisome	**Nucleosome**	Primosome	Chromatin	**Nucleosome**
8	I	Uptake of naked DNA is called	Transduction	**Transformation**	Conjugation	Recombination	**Transformation**
9	I	Transduction experiment was conducted by	Frederick Griffith	Zinder and Ledenberg	**Hershey and Chase**	Avery and Mc Carty	**Hershey and Chase**
10	I	Histones are rich in	**Arg and Lys**	Lys & Gly	Arg & Glu	Arg & Gly	**Arg and Lys**
11	I	Which histone protein is not a part of core particle of nucleosome?	**H1**	H2a	H2b	H4	**H1**
12	I	Which histone protein is involved in the transition between the solenoid form and the extended nucleosome form?	**H1**	H2a	H2b	H4	**H1**
13	I	The molecular chaperone which causes the nucleosome assembly is	Nucleoplasmin	**Histone**	Hu protein	Ubiquitin	**Histone**
14	I	Simple sequence DNA is concentrated in	**Centromere**	Telomere	Blastomeres	Blastocyst	**Centromere**
16	I	LTRs are absent in	LINES	SINES	**LINES AND SINES**	Viral retrotransposons	**LINES AND SINES**
17	I	Chicken lysozyme gene is a good example of	**Single copy DNA**	Moderately repetitive DNA	Simple sequence DNA	Highly repetitive DNA	**Single copy DNA**
18	I	Difference in length of simple sequence tandem arrays helps to develop a technique called	Foot printing	Northern blotting	Western	**Finger printing**	**Finger printing**
19	I	Important hallmark of IS element is	**Short direct repeats**	Long direct repeats	Inverted repeats	Tandem repeats	**Short direct repeats**
20	I	IS element contain _____ enzyme	helicase	primase	**transposase**	topoisomerase	**transposase**
21	I	The length of the DNA segment present in the nucleosome core particle is	140 bp	200 bp	**166 bp**	114 bp	**166 bp**
22	I	The housekeeping genes have	**GC box**	TATA box	Pribnow box	CAAT box	**GC box**
23	I	Negative supercoils are removed by	**Topoisomerase I**	Gyrase	Helicase	Rep protein	**Topoisomerase I**
24	I	DNA chains differ from each other by one nucleotide can be revolved using	20% agarose	20% polyacrylamide	**1% agarose**	**1% agarose**	1% polyacrylamide
25	I	Intercalating agents cause	nonsense mutation	missense mutation	silent mutation	**frameshift mutation**	**frameshift mutation**
26	I	Satellite DNAs are found in the region of	Euchromatin	**Heterochromatin**	Hypervariable regions	Functional elements	**Heterochromatin**
27	I	In DNA, the genetic information residues in	Purine bases	Pyrimidine bases	**Purine and Pyrimidine bases**	Sugar	**Purine and Pyrimidine bases**
28	I	Important characteristic of satellite DNA is, they contain	Repetitive base sequence	**GC rich**	Unique sequence	AT rich	**GC rich**
29	I	The most effective *in vitro* mutagen for phages is	**EMS**	EES	HNO_2	NaN_3	**EMS**
30	I	In conjugation the integration of F factor is mediated by	transposon	**IS elements**	viruses	simple sequence DNA	**IS elements**
31	I	The DNA in the chromatin is very tightly associated with proteins called	albumin	globulin	myosin	**histones**	**histones**
32	I	The molecular weight of histones is	**between 11,000 and 21,000**	between 1,000 and 11,000	between 8,000 and 20,000	between 1,000 and 31,000	**between 11,000 and 21,000**
33	I	Histones are very rich in	acidic amino acids	**basic amino acids**	neutral amino acids	aromatic amino acids	**basic amino acids**
34	I	Number of amino acids in H4 histones is	104	108	**102**	120	**102**
35	I	The bead of each nucleosome contains	**eight histone molecules**	ten histone molecules	two histone molecules	eight histone molecules	**eight histone molecules**
36	I	The number of base pairs serves as linker DNA between nucleosome beads is	200 bp	146bp	**54 bp**	45 bp	**54 bp**
37	I	The spacing of the nucleosome beads provides a repeating unit typically of about	**200 bp**	146bp	54bp	45bp	**200 bp**
38	I	The total number of base pairs bound tightly around the eight-part histone core are	200 bp	**146 bp**	54 bp	45 bp	**146 bp**
39	I	Specific binding sites of histone core to DNA is	**A=T rich base pairs**	G≡C rich base pairs	equal ratio of both base pairs	1:2 ratio of both the base pairs	**A=T rich base pairs**
40	I	Xeroderma pigmentosum in human is associated with a mutation in	Photoreactivation	**Nucleotide excision repair**	Base excision repair	Mismatch repair	**Nucleotide excision repair**
41	I	The following disorder is a good example of a single base mutation of mitochondrial DNA	**Lebers hereditary optic neuropathy**	Xeroderma pigmentosum	Fanconis syndrome	Cockayne's syndrome	**Lebers hereditary optic neuropathy**
42	I	The replacement of a purine base by another purine base is referred to as	**transition point mutation**	transversion point mutation	frame shift mutation	spontaneous mutation	**transition point mutation**
43	I	Frame shift mutations are characterized by	insertion or deletion of one or more base pairs in DNA	insertion of one or more base pairs in DNA	deletion of one or more base pairs in DNA	**replacement of purine by a pyrimidine or vice versa**	**replacement of purine by a pyrimidine or vice versa**
44	I	Sickle cell anemia is a classical example of	**silent mutation**	frame shift mutation	nonsense mutation	missense mutation	**silent mutation**
45	I	An alteration in a nucleotide sequence that changes a triplet coding for an amino acid into a termination codon.	**Nonsense mutation**	Mutation	Missense mutation	Induced mutation	**Nonsense mutation**
46	I	A base change which causes a change in the purine-pyrimidine orientation is called	**Transversion**	Transformation	Transition	Translation	**Transversion**
47	I	Mutants of UAG are called	ochre mutations	**amber mutations**	point mutations	replacement mutations	**amber mutations**
48	I	The frameshift mutations in genes can be produced by	transition	deletion	insertion	**deletion and insertion**	**deletion and insertion**
49	I	The acridine orange produces the following type of mutation	suppressor mutation	insertion mutation	**frameshift mutation**	deletion mutation	**frameshift mutation**
50	I	Proflavin causes	**point mutation**	frameshift mutation	transverse mutation	transition mutation	**point mutation**
51	I	The type of point mutation could be detected only after nucleic acid sequencing known as	silent mutation	**missense mutation**	non-sense mutation	frameshift mutation	**missense mutation**
52	I	The enzyme responsible for containing DNA replication in prokaryotes is	DNA polymerase I	DNA polymerase III	Polymerase beta	**Primase**	**Primase**
53	I	which histone is not part of the nucleosome?	**H1**	H2A	H2B	H3	**H1**
54	I	Which amino acid residue is in abundance in histones	**Arginine**	Aspartic acid	Tryptophan	Phenyl alanine	**Arginine**
55	I	Actinomycin D is an inhibitor of	Replication	**Transcription**	Translation	All of the above	**Transcription**
56	I	Addition or deletition of a nucleotide base pair involves	point mutation	silent mutation	nonsense mutation	**frame shift mutation**	**frame shift mutation**
57	I	When a single nucleotide base pair is substituted, type of mutation occurs is	random mutation	**point mutation**	inversion	mispairing	**point mutation**
58	I	Random mutations occurs due to	radiation	chemicals	external factors	**all of above**	**all of above**
59	I	A base substitution that causes regular codon to change in to another codon that codes for different amino acid is said to be	nonsense mutation	silent mutation	**missense mutation**	none of above	**missense mutation**
60	I	Result of rapid and continual divisions in	high demand for nutrients in benign cancer patients	high demand for nutrients in malignant cancer patients	**Both A and B**	None of above	**Both A and B**

UNIT-II

SYLLABUS

DNA replication and repair: General features of chromosomal replication. Enzymology of DNA replication, DNA replication machinery.Replication in prokaryotes and eukaryotes-Initiation, elongation and termination.DNA damage-types. Repair mechanism of DNA damage-all types.

Introduction

An important property of DNA is that it can replicate, or make copies of itself. Each strand of DNA in the double helix can serve as a pattern for duplicating the sequence of bases. This is critical when cells divide because each new cell needs to have an exact copy of the DNA present in the old cell.

General features of chromosomal replication

DNA replication, the basis for biological inheritance, is a fundamental process that occurs in all living organisms that copies their DNA. This process is "replication" in that each strand of the original double-stranded DNA molecule serves as template for the reproduction of the complementary strand. Therefore, following DNA replication, two identical DNA molecules have been produced from a single double-stranded DNA molecule. Cellular proofreading and error toe-checking mechanisms ensure near perfect fidelity for DNA replication.

Semi conservative mechanism

Semi conservative replication describes the mechanism by which DNA is replicated in all known cells. This mechanism of replication was one of three models originally proposedfor DNA replication:

- Semiconservative replication would produce two copies that each contained one of the original strands and one new strand.
- Conservative replication would leave the two original template DNA strands together in a double helix and would produce a copy composed of two new strands containing all of the new DNA base pairs.

- Dispersive replication would produce two copies of the DNA, both containing distinct regions of DNA composed of either both original strands or both new strands.

Meselson–Stahl experiment

This is an experiment by Matthew Meselson and Franklin Stahl in 1958 which supported the hypothesis that DNA replication was semi conservative. Semiconservative replication means that when the double stranded DNA helix was replicated, each of the two double stranded DNA helices consisted of one strand coming from the original helix and one newly synthesized. It has been called "the most beautiful experiment in biology

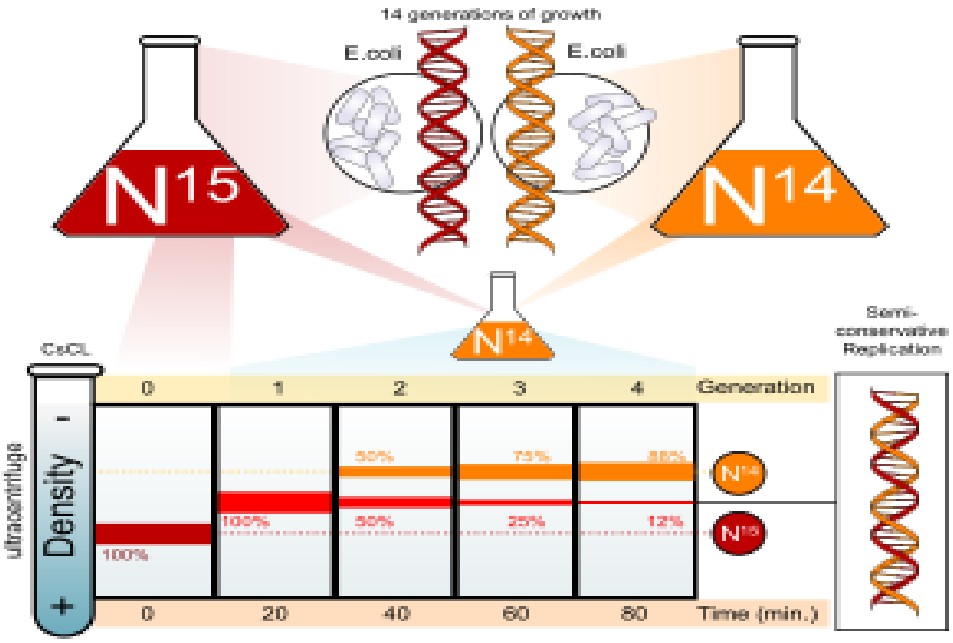
Experimental procedure

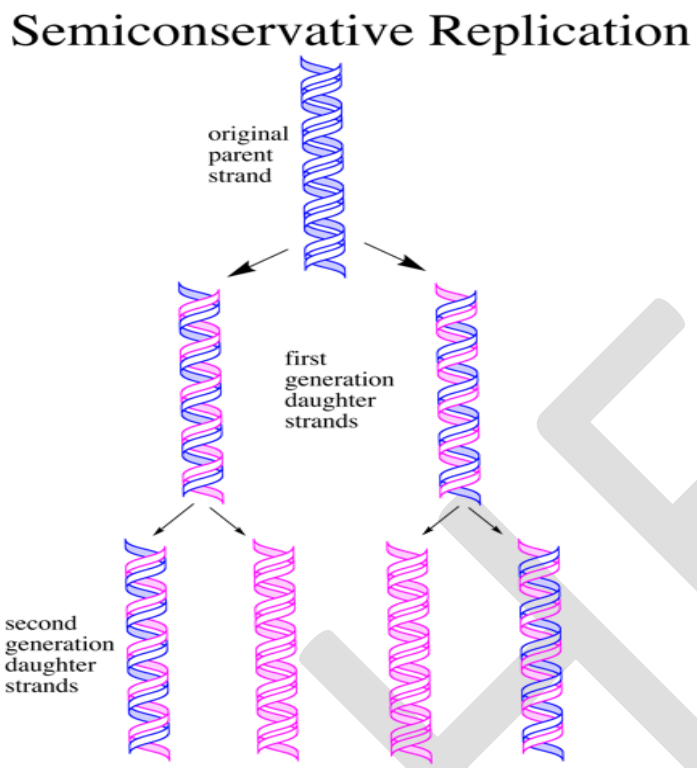
Nitrogen is a major constituent of DNA. ^{14}N is by far the most abundant isotope of nitrogen, but DNA with the heavier (but non-radioactive) ^{15}N isotope is also functional.

E. coli were grown for several generations in a medium with 15N. When DNA is extracted from these cells and centrifuged on a salt density gradient, the DNA separates out at the point at which its density equals that of the salt solution. The DNA of the cells grown in 15N medium had a higher density than cells grown in normal 14N medium. After that, E. coli cells with only 15N in their DNA were transferred to a 14N medium and were allowed to divide; the progress of cell division was monitored by measuring the optical density of the cell suspension.

DNA was extracted periodically and was compared to pure ^{14}N DNA and ^{15}N DNA. After one replication, the DNA was found to have close to the intermediate density. Since conservative replication would result in equal amounts of DNA of the higher and lower densities (but no DNA of an intermediate density), conservative replication was excluded. However, this result was consistent with both semiconservative and dispersive replication. Semiconservative replication would result in double-stranded DNA with one strand of ^{15}N DNA, and one of ^{14}N DNA, while dispersive replication would result in double-stranded DNA with both strands having mixtures of ^{15}N and ^{14}N DNA, either of which would have appeared as DNA of an intermediate density.

The authors continued to sample cells as replication continued. DNA from cells after two replications had been completed was found to consist of equal amounts of DNA with two different densities, one corresponding to the intermediate density of DNA of cells grown for only one division in ^{14}N medium, the other corresponding to DNA from cells grown exclusively in ^{14}N medium. This was inconsistent with dispersive replication, which would have resulted in a single density, lower than the intermediate density of the one-generation cells, but still higher than cells grown only in ^{14}N DNA medium, as the original ^{15}N DNA would have been split evenly among all DNA strands. The result was consistent with the semiconservative replication hypothesis.





Enzymology of DNA replication

Enzymes of DNA Replication

1.Helicase: Unwounds a portion of the DNA Double Helix

2.RNAPrimase: Attaches RNA primers to the replicating strands.

3.DNA Polymerase delta (ä): Binds to the 5' - 3' strand in order to bring nucleotides and create the daughter leading strand.

4.DNA Polymerase epsilon (å): Binds to the 3' - 5' strand in order to create discontinuous segments starting from different RNA primers.

5.Exonuclease (DNA Polymerase I): Finds and removes the RNA Primers

6.DNA Ligase: Adds phosphate in the remaining gaps of the phosphate - sugar backbone

7.Nucleases: Remove wrong nucleotides from the daughter strand.

a) primosome is a protein complex responsible for creating RNA primers on single stranded DNA during DNA replication.

The primosome consists of seven proteins: DnaGprimase, DnaB helicase, DnaC helicase assistant, DnaT, PriA, Pri B, and PriC. The primosome is utilized once on the leading strand of DNA and repeatedly, initiating each Okazaki fragment, on the lagging DNA strand.

b) Replisome is composed of the following:

2 DNA Pol III enzymes, made up of α, ε and θ subunits.

- the α subunit has polymerization activity.
- the ε subunit has proofreading activity.
- the θ subunit stimulates the ε subunit's proofreading.
- 2 β units which act as sliding DNA clamps, they keep the polymerase bound to the DNA.
- 2 τ units which connect the 2 DNA Pol III enzymes.
- 1 γ unit which acts as a clamp loader for the lagging strand Okazaki fragments, helping the two β subunits to form a unit and bind to DNA. The γ unit is made up of 5 γ subunits.

DNA Polymerases—The Enzymes of DNA Replication

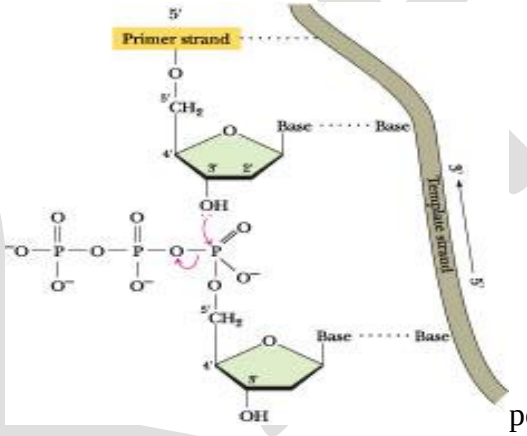
All DNA polymerases, whether from prokaryotic or eukaryotic sources, share the following properties: (a) The incoming base is selected within the DNA polymerase active site, as determined by Watson-Crick geometric interactions with its corresponding base in the template strand, (b) chain growth is in the 5' → 3' direction and is antiparallel to the template strand, and (c) DNA polymerases cannot initiate DNA synthesis de novo—all require a primer oligonucleotide with a free 3'-OH to build upon. E. *coli* DNA Polymerases Table compares the properties of the various DNA polymerases in E. *coli*. These enzymes are numbered I, II, and III, in order of their discovery. DNA polymerases I and II function principally in DNA repair; DNA polymerase III is the chief DNA-replicating enzyme of E. *coli*.

Properties of the DNA Polymerases of E. *coli*

Property	Pol I	Pol II	Pol III (core)*

Mass (kD)	103	90	130, 27.5, 8.6
Molecules/cell	400	?	40
Turnover number†	600	30	1200
Structural gene	*polA*	*polB*	*dnaE* (□ subunit) *dnaQ* (□ subunit) *holE* (□ subunit)
Polymerization 5' -□ 3'	Yes	Yes	Yes
Exonuclease 3' -□ 5'	Yes	Yes	Yes
Exonuclease 5' -□ 3'	Yes	No	No

Catalytic function



DNA polymerase I joinsdeoxynucleoside monophosphate units to the 3'-OH carries out a nucleophilic attack on the alpha-phosphoryl group of the incoming dNTP to form a phosphoester bond, and PP_i is released. The subsequent hydrolysis of PP_i by inorganic pyrophosphatase renders the reaction effectively irreversible. The reaction is

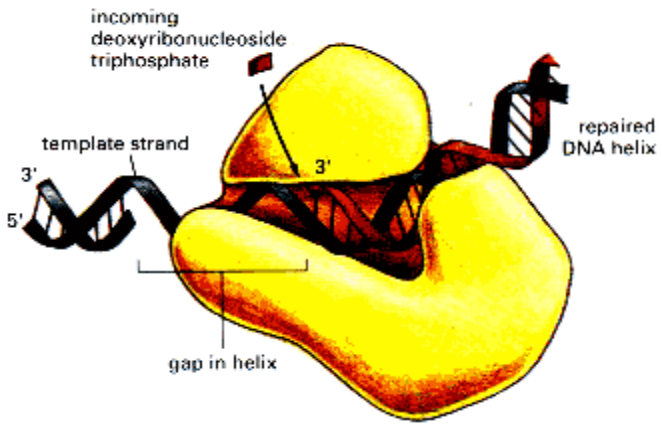
dNTP+pNpNpNpN-3'OH ------- PP_i+pNpNpNpNpN*-3'OH

Features of DNA polymerase

In 1957, Arthur Kornberg and his colleagues discovered the first DNA polymerase. A **primer** is essential because DNA polymerases can elongate only pre-existing chains; This primer must possess a free 3'-OH end to which an incoming deoxynucleoside monophosphate is added. All four dNTPs are substrates, pyrophosphate (PP_i) is released, and the dNMP is linked to the 3'-OH of the primer chain through formation of a phosphoester bond. The deoxynucleoside monophosphate to be incorporated is chosen through its geometric fit with the template base to form a Watson-Crick base pair. As DNA polymerase I catalyzes the successive addition of deoxynucleotide units to the 3'-end of the primer, the chain is elongated in the 5' - 3' direction, forming a polynucleotide sequence that runs antiparallel to the template but complementary to it. DNA polymerase I can proceed along the **template strand**, synthesizing a complementary strand of about 20 bases before it "falls off" (dissociates from) the template.

Strcuture of DNA polymerase I

E. *coli* DNA polymerase I is a 109-kD protein consisting of a single polypeptide of 928 amino acid residues. In addition to its 5' □ 3' polymerase activity, DNA polymerase I has two other catalytic functions, a 3' □ 5'*exonuclease* (*3'-exonuclease*) activity and a *5' □ 3' exonuclease (5'-exonuclease)* activity. The three distinct catalytic activities of DNA polymerase I reside in separate active sites in the enzyme.



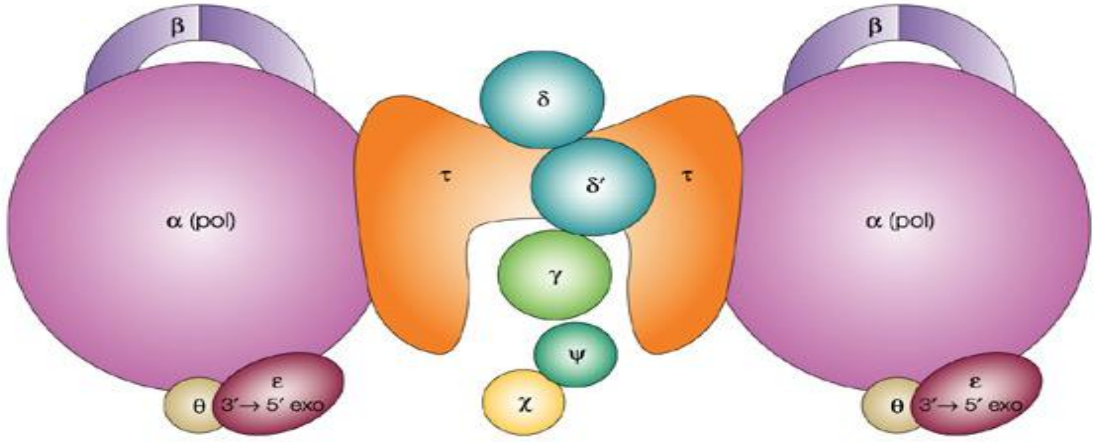
Properties of DNA polymerase I

Pol I possesses four enzymatic activities:

1. A 5' → 3' (forward) DNA-Dependent DNA polymerase activity, requiring a 3' primer site and a template strand
2. A 3' → 5' (reverse) exonuclease activity that mediates proofreading
3. A 5' → 3' (forward) exonuclease activity mediating nick translation during DNA repair.
4. A 5' → 3' (forward) RNA-Dependent DNA polymerase activity.

DNA Polymerase III

DNA polymerase III; in its holoenzyme form, responsible for replication of the E. *coli*chromosome.The DNA polymerase III holoenzyme is a multisubunit complex, which consists of 17 polypeptides. It contains four subassemblies. First, the core polymerase consists of three subunits: ∝(the polymerase); ε(the 3'–5' exonuclease); and θ(the stimulator of the 3'–5' exonuclease). Second, the τsubunit is responsible for dimerization of the core DNA polymerase. Third, the sliding clamp comprises two homodimers of the βsubunit, which provides the ring structure that is needed for processivity. Fourth, five subunits have clamp-loader functions — γ, δ, δ', χ and ψ



Nature Reviews | Molecular Cell Biology

The replisome is composed of the following:

- 2 DNA Pol III enzymes, each comprising α, ε and θ subunits.

- the α subunit has the polymerase activity.
- the ε subunit as 3'-5' exonuclease activity.
- the θ subunit stimulates the ε subunit's proofreading.
- 2 β units which act as sliding DNA clamps, they keep the polymerase bound to the DNA.
- 2 τ units which acts to dimerize two of the core enzymes (α, ε, and θ subunits).
- 1 γ unit which acts as a clamp loader for the lagging strand Okazaki fragments, helping the two β subunits to form a unit and bind to DNA. The γ unit is made up of 5 γ subunits which include 3 γ subunits, 1 δ subunit, and 1 δ' subunit. The δ is involved in copying of the lagging strand.
- X and Ψ which form a 1:1 complex and bind to τ.

DNA ligase

DNA ligase seals nicks in double-stranded DNA where a 3'-OH and a 5'-phosphate are juxtaposed. This enzyme is responsible for joining Okazaki fragments together to make the lagging strand a covalently contiguous polynucleotide chain. DNA ligase from eukaryotes and bacteriophage T4 is ATP-dependent; the E. *coli* enzyme requires NAD^+. Both types of DNA ligase act via an adenylylated ε-amino group of a Lys residue. Adenylylation of the 5'-phosphoryl group activates it for formation of a phosphoester bond with the 3'-OH, covalently sealing the sugar-phosphate backbone of DNA.

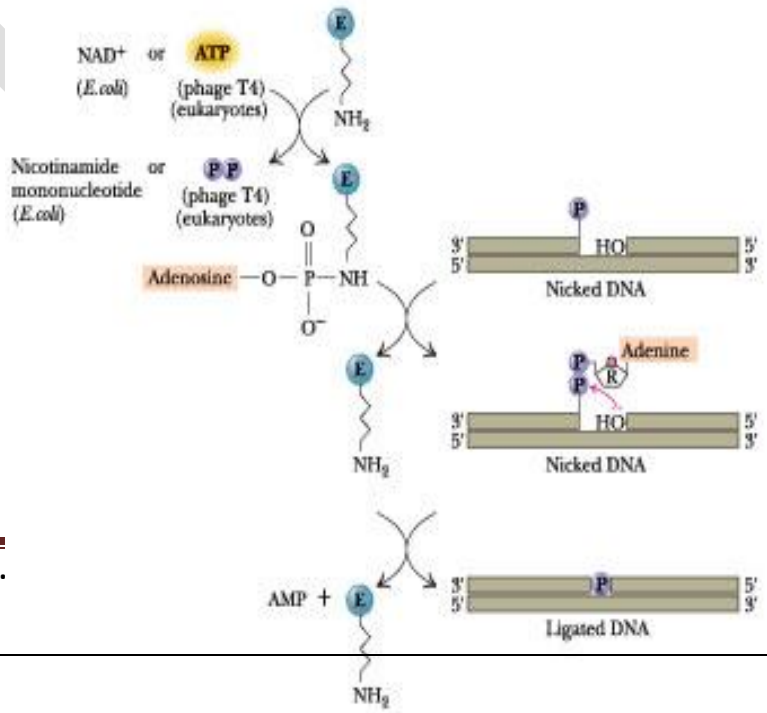


Figure: The mechanism of action of DNA ligases

General Features of a Replication Fork

We now can present a snapshot of the enzymatic apparatus assembled at a replication fork. DNA gyrase (topoisomerase) and helicase unwind the DNA double helix, and the unwound, single-stranded regions of DNA are maintained through interaction with SSB. Primase synthesizes an RNA primer on the lagging strand; the leading strand, which needs priming only once, was primed when replication was initiated. The lagging strand template is looped around, and each replicative DNA polymerase moves 5' -3' relative to its strand, copying template and synthesizing a new DNA strand. Each replicative polymerase is tethered to the DNA by its β -subunit sliding clamp. The DNA pol III complex periodically unclamps and then reclamps. Downstream on the lagging strand, DNA polymerase I excises the RNA primer and replaces it with DNA, and DNA ligase seals the remaining nick.

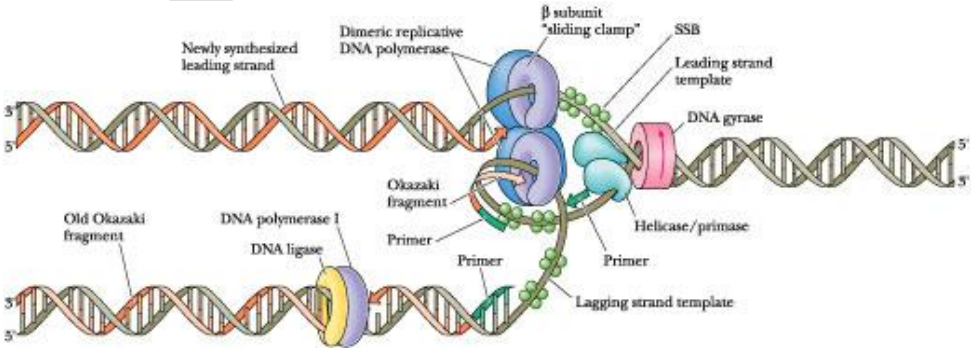


Figure :General features of a replication fork.

Proteins Involved in DNA Replication in E. *Coli*

Protein	Function
DNA gyrase	Unwinding DNA
SSB	Single-stranded DNA binding
DnaA	Initiation factor
HU	Histone-like (DNA binding)
PriA	Primosome assembly, 3' □ □5' helicase
PriB	Primosome assembly
PriC	Primosome assembly
DnaB	5' -□ 3' helicase (DNA unwinding)
DnaC	DnaB chaperone
DnaT	Assists DnaC in delivery of DnaB
Primase	Synthesis of RNA primer
DNA polymerase III holoenzyme	Elongation (DNA synthesis)
DNA polymerase I	Excises RNA primer, fills in with DNA

DNA ligase	Covalently links Okazaki fragments
Tus	Termination

Stages of replication -Initiation, elongation, Termination

Initiation

The origin of replication (also called the replication origin) is a particular sequence in a genome at which replication is initiated. This can either be DNA replication in living organisms such as prokaryotes and eukaryotes, or RNA replication in RNA viruses, such as double-stranded RNA viruses. DNA replication may proceed from this point bidirectionally or unidirectionally.The specific structure of the origin of replication varies somewhat from species to species, but all share some common characteristics such as high AT content. The origin of replication binds the pre-replication complex, a protein complex that recognizes, unwinds, and begins to copy DNA.The genome of E. coli is contained in a single circular DNA molecule of 4.6 x 106 nucleotide pairs. DNA replication begins at a single origin of replication. Replication origin is known as oriC. In E. coli, the oriC consists of three A-T rich 13-mer repeats and four 9-mer repeats. Ten to 20 monomers of the replication protein dnaA bind to the 9 mer repeats, and the DNA coils around this protein complex forming a protein core. This coiling stimulates the AT rich region in the 13 mer sequence to unwind, allowing copies of the helicase dnaB and its cofactor protein dnaC to bind to each strand of the resulting single-stranded DNA. The dnaB protein forms the basis of the replisome, a complex of enzymes that performs DNA replication.

In a cell, DNA replication begins at specific locations in the genome, called "origins". Unwinding of DNA at the origin, and synthesis of new strands, forms a replication fork. In addition to DNA polymerase, the enzyme that synthesizes the new DNA by adding nucleotides matched to the template strand, a number of other proteins like helicase, primase, SSB and topoisomerase are associated with the fork and assist in the initiation and continuation of DNA synthesis.

Origins contain DNA sequences recognized by replication initiator proteins (e.g. dnaA in E coli' and the Origin Recognition Complex in yeast).These initiator proteins recruit other proteins to separate the two strands and initiate replication forks.Initiator proteins recruit other proteins to separate the DNA strands at the origin, forming a bubble. Origins tend to be "AT-

rich" (rich in adenine and thymine bases) to assist this process.Once strands are separated, RNA primers are created on the template strands.

The initiation of DNA replication is mediated by DnaA, a protein that binds to a region of the origin known as the DnaA box. In E. coli, there are 5 DnaA boxes, each of which contains a highly conserved 9 bp consensus sequence 5' - TTATCCACA - 3'. Binding of DnaA to this region causes it to become negatively supercoiled. Following this, a region of OriC upstream of the DnaA boxes (known as DnaB boxes) become melted. There are three of these regions, and each is 13 bp long, and AT-rich (which facilitates melting because less energy is required to break the orm five DnaA dimers. DnaC is then released, and the prepriming complex is complete. In order for DNA replication to continue, SSB protein is needed to prevent the single strands of DNA from forming any secondary structures and to prevent them from reannealing, and DNA gyrase is needed to relieve the stress (by creating negative supercoils) created by the action of DnaB helicase. The unwinding of DNA by DnaB helicase allows for primase (DnaG) an RNA polymerase to prime each DNA template so that DNA synthesis can begin.

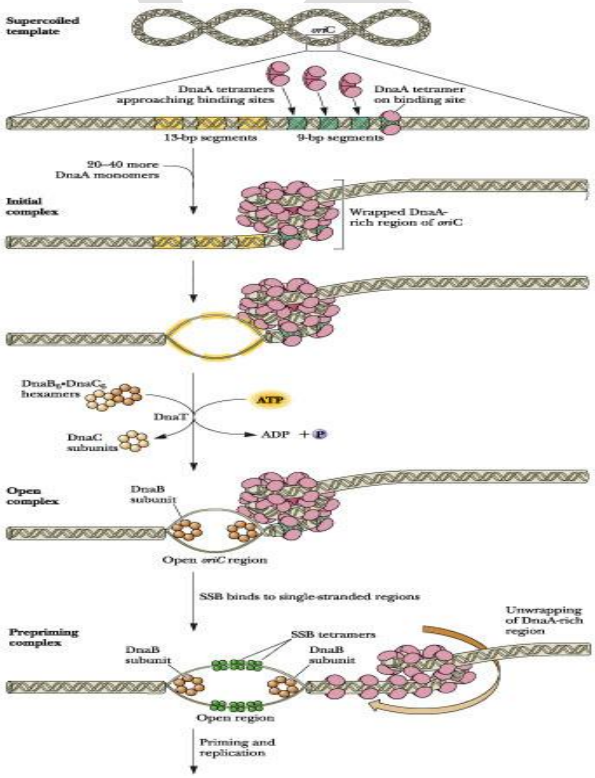


Figure- Initiation of replication

Elongation

The replication fork is a structure that forms within the nucleus during DNA replication. It is created by helicases, which break the hydrogen bonds holding the two DNA strands together. The resulting structure has two branching "prongs", each one made up of a single strand of DNA. These two strands serve as the template for the leading and lagging strands which will be created as DNA polymerase matches complementary nucleotides to the templates; The templates may be properly referred to as the leading strand template and the lagging strand template.

Leading strand

The leading strand is the template strand of the DNA double helix so that the replication fork moves along it in the 3' to 5' direction. This allows the new strand synthesized complementary to it to be synthesized 5' to 3' in the same direction as the movement of the replication fork.

On the leading strand, a polymerase "reads" the DNA and adds nucleotides to it continuously. This polymerase is DNA polymerase III (DNA Pol III) in prokaryotes and presumably Pol ε in eukaryotes.

Lagging strand

The lagging strand is the strand of the template DNA double helix that is oriented so that the replication fork moves along it in a 5' to 3' manner. Because of its orientation, opposite to the working orientation of DNA polymerase III, which moves on a template in a 3' to 5' manner, replication of the lagging strand is more complicated than that of the leading strand.

On the lagging strand, primase "reads" the DNA and adds RNA to it in short, separated segments. In eukaryotes, primase is intrinsic to Pol α. DNA polymerase III or Pol δ lengthens the primed segments, forming Okazaki fragments. Primer removal in eukaryotes is also performed

by Pol δ. In prokaryotes, DNA polymerase I "reads" the fragments, removes the RNA using its flap endonuclease domain (RNA primers are removed by 5'-3' exonuclease activity of polymerase I [weaver, 2005], and replaces the RNA nucleotides with DNA nucleotides (this is necessary because RNA and DNA use slightly different kinds of nucleotides). DNA ligase joins the fragments together.

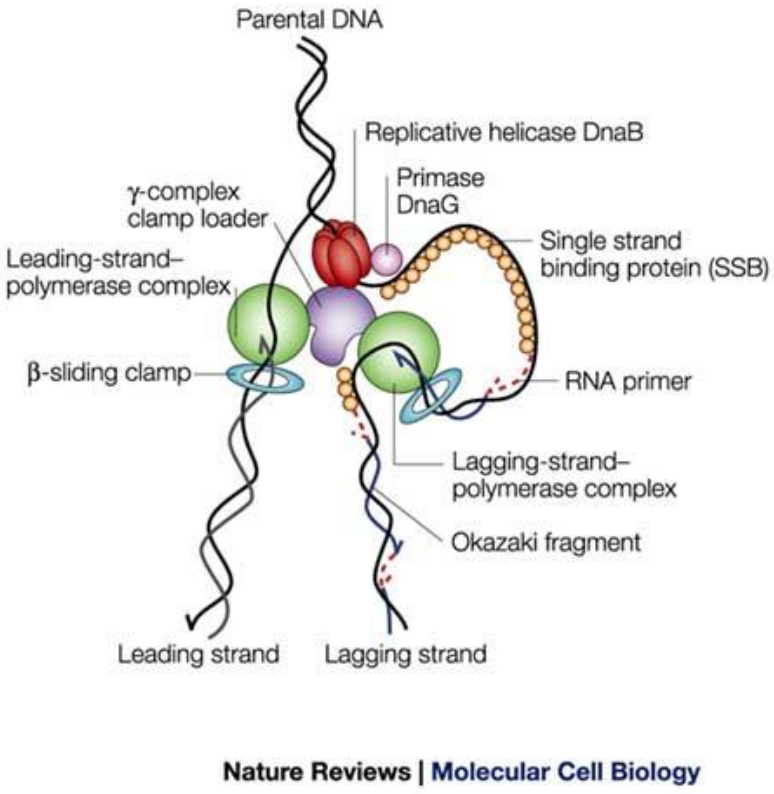


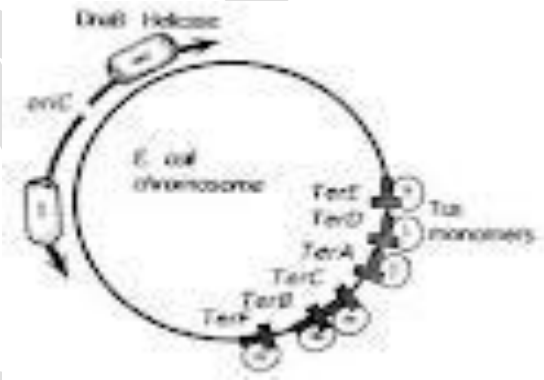
Figure-Elongation of DNA replication

Termination of replication

Located diametrically opposite from *oriC* on the E. *coli* circular map is a terminus region, the **Ter,** or t, locus. The bidirectionally moving replication forks meet here and replication is terminated. The *Ter* region contains a number of short DNA sequences containing a consensus core element 5'-GTGTGTTGT. These *Ter* sequences act as terminators; clusters of three or four *Ter* sequences are organized into two sets inversely oriented with respect to one another. One set

blocks the clockwise-moving replication fork, and its inverted counterpart blocks the counterclockwise-moving replication fork. A *Ter* sequence element will impede replication fork progression *only if oriented in the proper direction* with respect to the approaching replication fork and then only if a specific 36-kD replication termination protein, **Tus protein,** is bound to it. Tus protein is a **contrahelicase.** That is, Tus protein prevents the DNA duplex from unwinding by blocking progression of the replication fork and inhibiting the ATP-dependent DnaB helicase activity. Mutations in either the *Ter* locus or the gene encoding Tus protein do not grossly affect DNA replication, demonstrating that this termination mechanism is not essential.

Final synthesis of both duplexes is completed. Replication usually leaves the circular progeny chromosomes intertwined by 20 to 30 coils about each other, a so-called **catenated** state. In order to disengage the individual duplexes from each other prior to their distribution to daughter cells, double-stranded cuts must be made so that the double helices can pass through one another. Topoisomerase II (DNA gyrase) can catalyze this process. **Fig: terminator ssequence**



Difference between replication in eukaryotes and prokaryotes

Eukaryotes initiate DNA replication at multiple points in the chromosome, so replication forks meet and terminate at many points in the chromosome; these are not known to be regulated in any particular manner. Because eukaryotes have linear chromosomes, DNA replication often fails to synthesize to the very end of the chromosomes (telomeres), resulting in telomere shortening. This is a normal process in somatic cells — cells are only able to divide a certain number of times before the DNA loss prevents further division.

Important features

Origin-ARS

Polymerase - alpha(α,), delta and eata (-€)

Bidirectional replication

Telemere synthesis

Regulation of replication

Within eukaryotes, DNA replication is controlled within the context of the cell cycle. As the cell grows and divides, it progresses through stages in the cell cycle; DNA replication occurs during the S phase (Synthesis phase). The progress of the eukaryotic cell through the cycle is controlled by cell cycle checkpoints. Progression through checkpoints is controlled through complex interactions between various proteins, including cyclins and cyclin-dependent kinases.

The G1/S checkpoint (or restriction checkpoint) regulates whether eukaryotic cells enter the process of DNA replication and subsequent division. Cells which do not proceed through this checkpoint are quiescent in the "G0" stage and do not replicate their DNA.

DNA damage

Types of damage

DNA damage can be subdivided into two main types:

1. Endogenous damage such as attack by reactive oxygen species produced from normal metabolic byproducts (spontaneous mutation), especially the process of oxidative deamination; also includes replication errors

2. Exogenous damage caused by external agents such as

1. Ultraviolet [UV 200-300nm] radiation from the sun
2. other radiation frequencies, including x-rays and gamma rays
3. Hydrolysis or thermal disruption
4. Certain plant toxins
5. Human-made mutagenic chemicals, especially aromatic compounds that act as DNA intercalating agents

6. Cancer chemotherapy and radiotherapy
7. Viruses

The replication of damaged DNA before cell division can lead to the incorporation of wrong bases opposite damaged ones. Daughter cells that inherit these wrong bases carry mutations from which the original DNA sequence is unrecoverable (except in the rare case of a back mutation, for example, through gene conversion).

Sources of damage

There are five main types of damage to DNA due to endogenous cellular processes:

1. Oxidation of bases [e.g. 8-oxo-7,8-dihydroguanine (8-oxoG)] and generation of DNA strand interruptions from reactive oxygen species,

2. Alkylation of bases (usually methylation), such as formation of 7-methylguanine, 1-methyladenine, 6-O-Methylguanine

3. Hydrolysis of bases, such as deamination, depurination and depyrimidination.

4. “Bulky adduct formation" (i.e. benzo[a]pyrenediol epoxide-dG adduct)

5. Mismatch of bases, due to errors in DNA replication, in which the wrong DNA base is stitched into place in a newly forming DNA strand, or a DNA base is skipped over or mistakenly inserted.

Damage caused by exogenous agents comes in many forms. Some examples are:

1. UV-B light causes crosslinking between adjacent cytosine and thymine bases creating pyrimidine dimers. This is called direct DNA damage.

2. UV-A light creates mostly free radicals. The damage caused by free radicals is called indirect DNA damage.

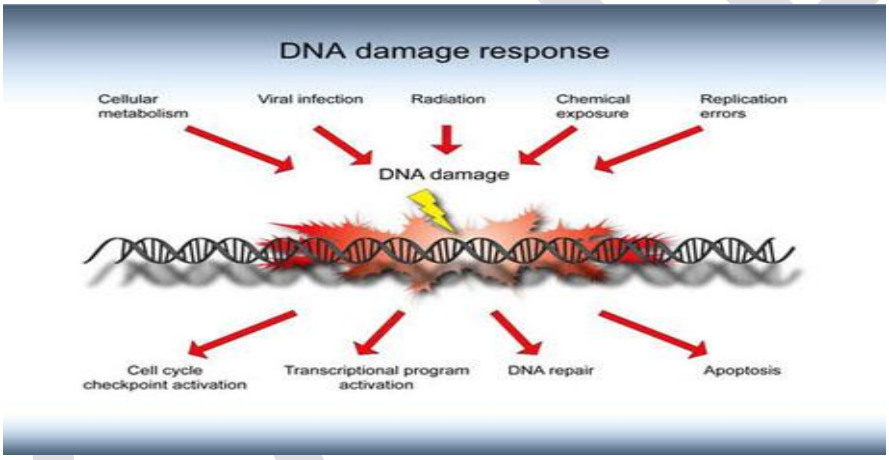
3. Ionizing radiation such as that created by radioactive decay or in cosmic rays causes breaks in DNA strands. Low level ionizing radiation may induce irreparable DNA damage (leading to replicational and transcriptional errors needed for neoplasia or may trigger viral interactions) leading to pre-mature aging and cancer.

4. Thermal disruption at elevated temperature increases the rate of depurination (loss of purine bases from the DNA backbone) and single strand breaks. For example, hydrolytic depurination is

seen in the thermophilic bacteria, which grow in hot springs at 40-80 °C. The rate of depurination (300 purine residues per genome per generation) is too high in these species to be repaired by normal repair machinery, hence a possibility of an adaptive response cannot be ruled out.

5. Industrial chemicals such as vinyl chloride and hydrogen peroxide, and environmental chemicals such as polycyclic hydrocarbons found in smoke, soot and tar create a huge diversity of DNA adducts- ethenobases, oxidized bases, alkylated phosphotriesters and Crosslinking of DNA just to name a few.

UV damage, alkylation/methylation, X-ray damage and oxidative damage are examples of induced damage. Spontaneous damage can include the loss of a base, deamination, sugar ring puckering and tautomeric shift.



DNA repair mechanisms

Repair of single strand break

When only one of the two strands of a double helix has a defect, the other strand can be used as a template to guide the correction of the damaged strand. In order to repair damage to one of the two paired molecules of DNA, there exist a number of excision repair mechanisms that remove the damaged nucleotide and replace it with an undamaged nucleotide complementary to that found in the undamaged DNA strand.

They are

1. Direct reversal
2. Base exicision repair
3. Nucleotide excision repair
4. Mismatch repair

Direct reversal

Cells are known to eliminate three types of damage to their DNA by chemically reversing it. These mechanisms do not require a template, since the types of damage they counteract can only occur in one of the four bases. Such direct reversal mechanisms are specific to the type of damage incurred and do not involve breakage of the phosphodiester backbone. The formation of pyrimidine dimers upon irradiation with UV light results in an abnormal covalent bond between adjacent pyrimidine bases. The photoreactivation process directly reverses this damage by the action of the enzyme photolyase, whose activation is obligately dependent on energy absorbed from blue/UV light (300–500 nm wavelength) to promote catalysis. Another type of damage, methylation of guanine bases, is directly reversed by the protein methyl guanine methyl transferase (MGMT), the bacterial equivalent of which is called ogt. This is an expensive process because each MGMT molecule can only be used once; that is, the reaction is stoichiometric rather than catalytic. A generalized response to methylating agents in bacteria is known as the adaptive response and confers a level of resistance to alkylating agents upon sustained exposure by upregulation of alkylation repair enzymes. The third type of DNA damage reversed by cells is certain methylation of the bases cytosine and adenine.

Base excision repair (BER)

This repairs damage to a single base caused by oxidation, alkylation, hydrolysis, or deamination. The damaged base is removed by a DNA glycosylase. The "missing tooth" is then recognised by an enzyme called AP endonuclease, which cuts the Phosphodiester bond. The missing part is then resynthesized by a DNA polymerase, and a DNA ligase performs the final nick-sealing step.

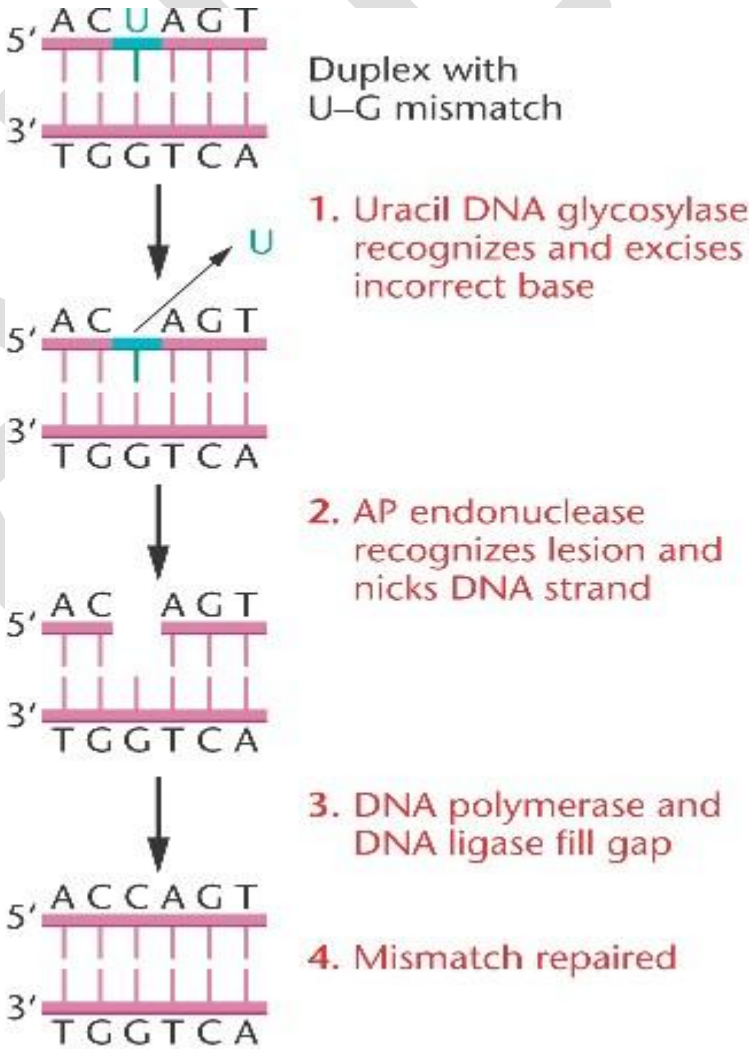
Steps in base excision repair in *E. coli*.

A base pair **mismatch** is recognized by DNA glycosylase, which removes the base from the sugar, creating an apurinic/apyrimidinic (**AP**) site.

AP endonuclease makes a cut in the phosphodiester backbone at the **AP** site.

The gap is filled by DNA polymerase I and DNA ligase

The error is corrected.



Nucleotide excision repair (NER),

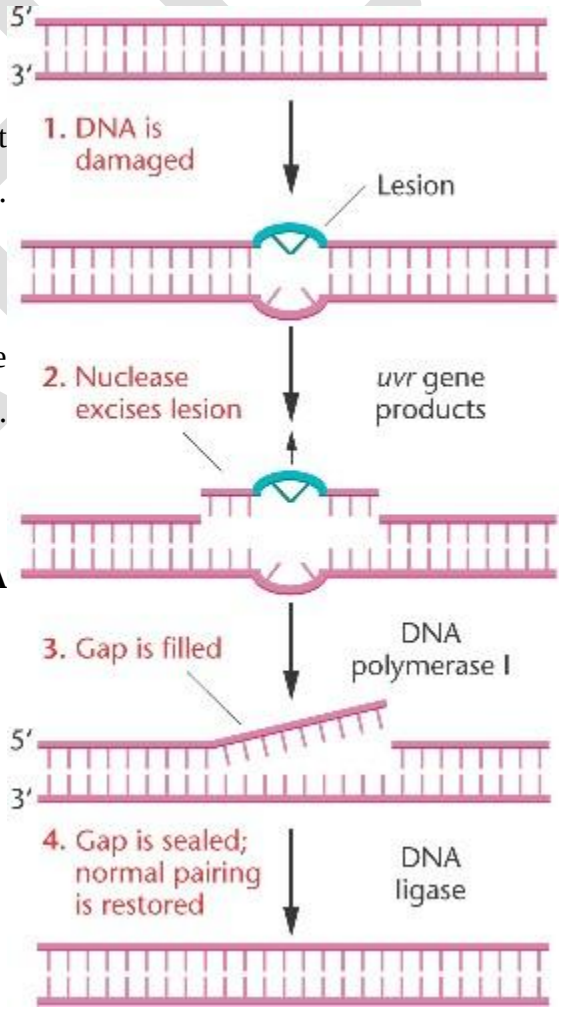
This recognizes bulky, helix-distorting lesions such as pyrimidine dimers and 6,4 photoproducts. A specialized form of NER known as transcription-coupled repair deploys NER enzymes to genes that are being actively transcribed.

1. Damaged DNA is recognized by **uvr** (ultraviolet repair) proteins.

2. A number of nucleotides is clipped out around the **lesion** by a **nuclease**.

3. The **gap** is filled by **DNA polymerase I** and **DNA ligase**

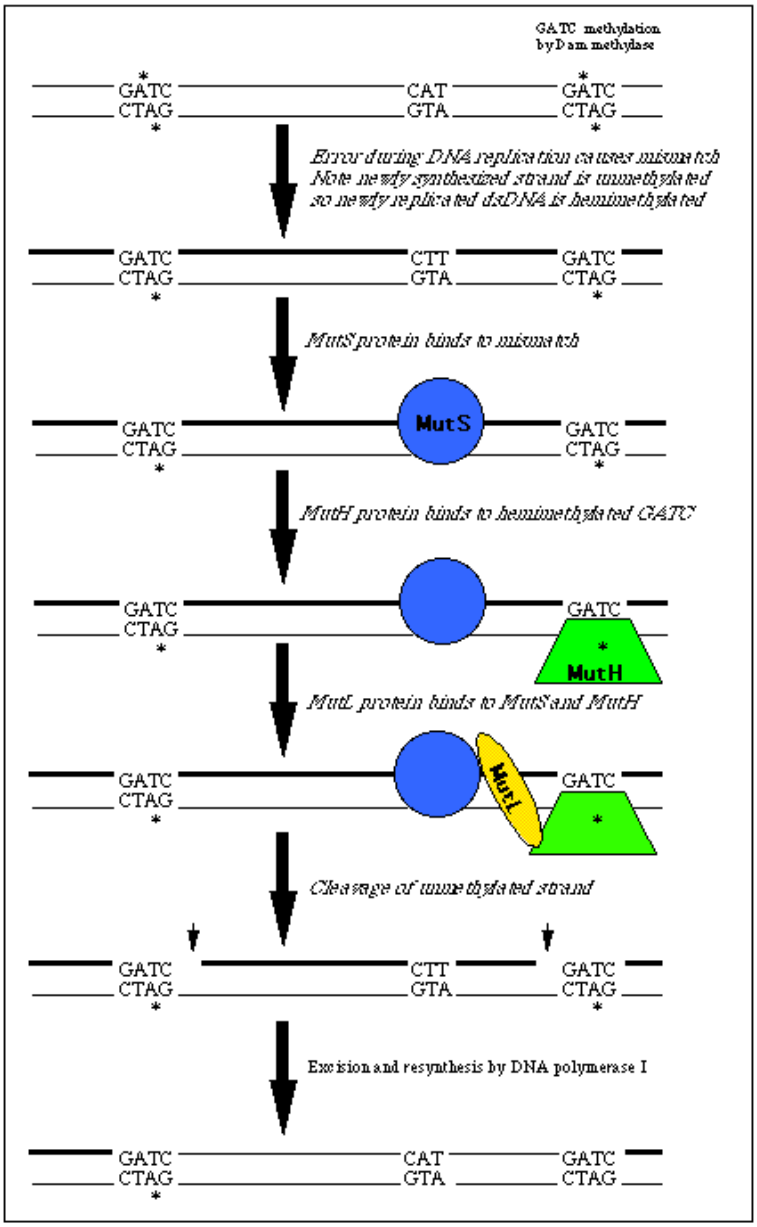
4. The error is corrected.



Mismatch repair (MMR)

Errors of DNA replication and recombination that result in mispaired (but undamaged) nucleotides are corrected by this mechnism

Mismatch repair of newly replicated DNA by the MutSLH proteins is depicted in the diagram.



Repair of double strand break

Homologous recombination requires the presence of an identical or nearly identical sequence to be used as a template for repair of the break. The enzymatic machinery responsible for this repair process is nearly identical to the machinery responsible for chromosomal crossover during meiosis. This pathway allows a damaged chromosome to be repaired using a sister chromatid (available in G2 after DNA replication) or a homologous chromosome as a template. DSBs caused by the replication machinery attempting to synthesize across a single-strand break or unrepaired lesion cause collapse of the replication fork and are typically repaired by recombination.This involves the protein Rec A and Rec BCD enzymes.

POSSIBLE QUESTIONS

UNIT II

Part-B

1. Discuss briefly about semiconservative mechanism of DNA replication
2. Explain the mechanism that seal the end of DNA
3. Explain the role of telomerase and Topoisomerase in DNA replication
4. Write short notes on okazaki fragments
5. Explain the proof reading activity of DNA polymerase
6. Explain the bidirectional mode of replication
7. Write short notes on i) primosomes ii) Replisomes
8. Explain the role of glycosylase and ligase in DNA repair
9. Explain the photo reactivation repair of DNA
10. Explain in brief about the enzymes involved in DNA replication

Part-C

1. Discuss in detail about the DNA replication in prokaryotes
2. Explain the role of various enzymes involved in DNA replication
3. Compare and contrast the eukaryotic and prokaryotic replication
4. Explain the importance and mechanism of Post replication repair and SOS response
5. Explain the mechanism of UV repair and mismatch repair

Karpagam acadmy of higher educaton
Department of biochemistry
Molecular biology (19BCP202) MCQ
UNIT II

S.NO	UNIT	QUESTION	OPTION 1	OPTION 2	OPTION 3	OPTION 4	ANSWER
1	II	Replication is characterized by the following except it is	Semiconservative	bi-directional	semidiscontinuous	**dependent on DNA primer**	**dependent on DNA primer**
2	II	The process of DNA duplication is	**Replication**	Transcription	Translation	Reverse transcription	**Replication**
3	II	An enzyme that catalyses phosphodiester linkages in DNA strands is called	DNA polymerase	**DNA ligase**	RNA polymerase	Topoisomerase	**DNA ligase**
4	II	The separation of two DNA strands for replication is brought by	DNA polymerase III	**DNA helicase**	DNA polymerase I	DNA ligase	**DNA helicase**
5	II	The synthesis of new DNA (replication) in prokaryotes is catalysed by	DNA polymerase I	DNA polymerase II	**DNA polymerase III**	DNA topoisomerase	**DNA polymerase III**
6	II	The proof reading activity of the newly synthesized DNA is present on the enzyme	DNA helicase	**DNA polymerase I**	DNA polymerase II	DNA polymerase III	**DNA polymerase I**
7	II	The problem of supercoils during DNA replication is overcome by a group of enzymes called	**DNA topoisomerases**	DNA ligases	DNA polymerases	DNA helicases	**DNA topoisomerases**
8	II	The enzyme responsible for the replication of mitochondrial DNA	DNA polymerase α	DNA polymerase β	**DNA polymerase γ**	DNA polymerase δ	**DNA polymerase γ**
9	II	Okazaki fragments are initiated with	DNA primer	**RNA primer**	DNA template	RNA template	**RNA primer**
10	II	Single strand binding protein binds to single strand DNA	to prevent replication	to repair base pairs	to initiate transcription	**to prevent reformation of duplex state**	**to prevent reformation of duplex state**
11	II	Rolling circle replication is otherwise called as	θ replication		D-loop replication	L-loop replication	
12	II	The DNA replication is discontinuous was proved by	Messelson-Stahl	**Reigi Okazaki**	Albert Lehninger	Arthur Kornberg	**Reigi Okazaki**
13	II	RNA primers are removed by	**DNA polymerase I**	DNA polymerase II	DNA polymerase III	DNA topoisomerase	**DNA polymerase I**
14	II	Primase initiates the following activities except	leading strand synthesis	replication	Okazaki fragments	**transcription**	**transcription**
16	II	Ori C is rich in _____sequence	GC	**AT**	ATGC	GATC	**AT**
17	II	Dna A protein recognize and binds to	4 9 bp repeats in ter C	13 bp AT rich segment	**4 9 bp repeats in Ori C**	RNA polymerase	**4 9 bp repeats in Ori C**
18	II	Which of the following enzymes are used to join bits of DNA?	**DNA ligase**	DNA polymerase	Primase	Endonuclease	**DNA ligase**
19	II	Semi conservative replication of DNA was first demonstrated in	***Escherichia coli***	*Streptococcus pneumoniae*	*Salmonella typhi*	*Drosophila melanogaster*	***Escherichia coli***
20	II	Mode of DNA replication in *E. coli* is	Conservative and unidirectional	Semiconservative and unidirectional	Conservative and bidirectional	**Semiconservative and bidirectional**	**Semiconservative and bidirectional**
21	II	True replication of DNA is possible due to	Hydrogen bonding	Phosphate backbone	**Complementary base pairing rule**	Phosphodiesterase	**Complementary base pairing rule**
22	II	DNA synthesis can be specifically measured by estimating the incorporation of radiolabeled	Uracil	**Thymine**	Adenine	Deoxyribose sugar	**Thymine**
23	II	The elongation of the leading strand during DNA synthesis	Progresses away from the replication fork	Occur in 3'-5' direction	Produce Okazaki fragment	**Depend on the action of DNA polymerase**	**Depend on the action of DNA polymerase**
24	II	Eukaryotes differ from prokaryote in mechanism of DNA replication due to:	Different enzyme for synthesis of lagging and leading strand	Use of DNA primer rather than RNA primer	Unidirectional rather than bidirectional replication	**Discontinuous rather than semidiscontinuous replication**	**Discontinuous rather than semidiscontinuous replication**
25	II	During the replication of DNA, the synthesis of DNA on lagging strand takes place in segments, these segments are called	Satellite segments	Double helix segments	Kornberg segments	**Okazaki segments**	**Okazaki segments**
26	II	Which of the following reactions is required for proofreading during DNA replication by DNA polymerase III?	3' - 5' exonuclease activity	5' - 3' exonuclease activity	**3' - 5' endonuclease activity**	5' - 3' endonuclease activity	**3' - 5' endonuclease activity**
27	II	In the rolling circle method of replication	the 5' tail of DNA is nicked	RNA is nicked	**one strand of DNA in the circle is nicked**	both strands of DNA in the circle are nicked	**one strand of DNA in the circle is nicked**
28	II	All of the following are differences between eukaryotic and prokaryotic DNA replication except	the type and number of polymerases involved in DNA synthesis	multiple vs. single replication origins	the rate of DNA synthesis	**the ability to form a replication fork**	**the ability to form a replication fork**
29	II	In the Meselson -Stahl experiment, which mode of replication can be eliminated based on data derived after one generation of replication?	Dispersive	Semiconservative	**Conservative**	all three modes	**Conservative**
30	II	The discovery of Okazaki fragments suggested that DNA synthesis is	**discontinuous**	continuous	3 ' to 5'	semiconservative	**discontinuous**
31	II	A replicating prokaryotic chromosome has -------- replication forks	One	Many	Three	**Two**	**Two**
32	II	Which molecule serves to destabilize the DNA helix in order to open it up, creating a replication fork?	**DNA helicase**	DNA ligase	DNA polymerase	SSBPs	**DNA helicase**
33	II	For DNA Replication, unwinding of DNA is done by	**Helicase**	Ligase	Hexonuclease	Topoisomerase	**Helicase**
34	II	In vivo synthesis of DNA is _____	3' to 5'	**5' to 3'**	both 3' to 5' and 5' to 3'	neither 3' to 5' nor 5' to 3'	**5' to 3'**
35	II	Which of the following forms of DNA can serve as a template for DNA polymerase	**Partially double stranded DNA**	Circular double stranded DNA	Intact double stranded DNA	Circular single stranded DNA	**Partially double stranded DNA**
36	II	The last DNA to be replicated in the eukaryotic chromosome is	Telomeres at the end of the chromosomes	**Heterochromatin**	Euchromatin in the arms of the chromosome	Facultative heterochromatin	**Heterochromatin**
37	II	In which phase of the cell cycle does DNA replication occur?	G0	G1	**S**	G2	**S**

38	II	The enzyme responsible for initiating DNA replication in prokaryotes is	DNA polymerase I	DNA polymerase III	Polymerase beta	**Primase**	**Primase**
39	II	The enzyme responsible for continuing DNA replication in prokaryotes, once it is initiated is:	DNA polymerase I	**DNA polymerase III**	Polymerase beta	Polymerase delta	**DNA polymerase III**
40	II	The enzyme _____ unzips and unwinds the DNA molecule.	DNA polymerase	**helicase**	primase	DNA ligase	**helicase**
41	II	DNA replication results in:	2 completely new DNA molecules	**2 DNA molecules that each contain a strand of the original**	1 new DNA molecule, 1 old molecule is conserved	1 new molecule of RNA	**2 DNA molecules that each contain a strand of the original**
42	II	During replication, what enzyme adds complimentary bases?	helicase	synthesase	replicase	**polymerase**	**polymerase**
43	II	Helicase opens up the DNA by breaking _____ bonds between the complementary strands.	hydrophobic	**hydrogen**	phosphodiester	ionic	**hydrogen**
44	II	Ligase forms a bond between	**two segments of single stranded DNA.**	two molecules of double-stranded DNA.	two segments of single stranded RNA.	two molecules of double-stranded RNA.	**two segments of single stranded DNA.**
45	II	DNA polymerase can add new nucleotides to the free _____ of a nucleotide strand.	1' end	**3' end**	5' end	1' end and 3' end	**3' end**
46	II	Which of the following possesses both 5'-3' and 3'-5' exonuclease activity?	**Kornberg enzyme**	DNA polymerase III	Taq DNA polymerase	DNA gyrase	**Kornberg enzyme**
47	II	The enzyme that catalyzes the synthesis of DNA is called	**DNA polymerase**	DNA gyrase	DNA ligase	Helicase	**DNA polymerase**
48	II	Which of the following repairs nicked DNA by forming a phosphodiester bond between adjacent nucleotides?	Helicase	DNA gyrase	Topoisomerases	**DNA ligase**	**DNA ligase**
49	II	Both strands of DNA serve as templates concurrently in	**replication**	excision repair	mismatch repair	transcription	**replication**
50	II	What is the major difference between T1 and T2 ter sequences?	Tus binds to T1 but not T2.	Tus binds to T2 but not T1.	**T1 stops counterclockwise-moving forks, T2 stops clockwise-moving forks.**	T2 stops counterclockwise-moving forks, T1 stops only clockwise-moving forks.	**T1 stops counterclockwise-moving forks, T2 stops clockwise-moving forks.**
51	II	What enzyme performs decatenation?	Polymerase	**Topoisomerase**	Telomerase	Decatenase	**Topoisomerase**
52	II	The activity of AP endonuclease activity is involved in	**Base excision repair**	Nucleotide excision repair	Mismatch repair	Double strand break repair	**Base excision repair**
53	II	Which of the following repairs nicked DNA by forming a phosphodiester bond between adjacent nucleotides?	Helicase	DNA gyrase	Topoisomerases	**DNA ligase**	**DNA ligase**
54	II	Both strands of DNA serve as templates concurrently in	**replication**	excision repair	mismatch repair	transcription	**replication**
55	II	What is the major difference between T1 and T2 ter sequences?	Tus binds to T1 but not T2.	Tus binds to T2 but not T1.	**T1 stops counterclockwise-moving forks, T2 stops**	T2 stops counterclockwise-moving forks, T1 stops only	**T1 stops counterclockwise-moving forks, T2 stops**
56	II	What enzyme performs decatenation?	Polymerase	**Topoisomerase**	Telomerase	Decatenase	**Topoisomerase**
57	II	The activity of AP endonuclease activity is involved in	**Base excision repair**	Nucleotide excision repair	Mismatch repair	Double strand break repair	**Base excision repair**
58	II	Which of the following enzyme is not involved in DNA replication?	Helicase	Primase	Topoisomerase	**DNA polymerase**	**DNA polymerase**
59	II	How many hydrogen bonds form between U and A in a Watson – crick	0	1	**2**	3	**2**
60	II	In which phase of the cell cycle does replication occur?	G0	G1	**S**	G2	**S**

UNIT-III

SYLLABUS

Transcription: prokaryotic gene transcription- Initiation, elongation and termination. Eukaryotic gene transcription- transcription unit, RNA polymerases- types, Transcription and processing of mRNA, tRNA and rRNA. Regulatory sequences in protein coding genes-TATA box, initiators, CpG island, promoter-proximal element, activators and repressors of transcription, Multiple transcription control elements. Regulation of transcription factor activity by lipid-soluble hormones.

Introduction

Transcription is the process in which messenger RNA transcripts of genetic material in prokaryotes are produced, to be translated for the production of proteins. Prokaryotic transcription occurs in the cytoplasm alongside translation. Unlike in eukaryotes, prokaryotic transcription and translation can occur simultaneously. This is impossible in eukaryotes, where transcription occurs in a membrane-bound nucleus while translation occurs outside the nucleus in the cytoplasm. In prokaryotes genetic material is not enclosed in a membrane-enclosed nucleus and has access to ribosomes in the cytoplasm.

General feature of transcription

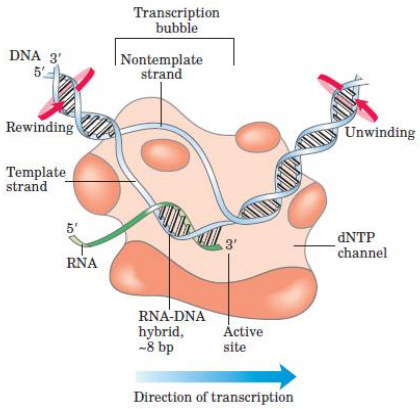
Direction : 5' → 3'

Enzyme : RNA polymerase that does not require primer but require a template

Templte strand: One of the double stranded DNA strand acts as a templte called tempalte strand or Non coding strand

Structure and functions of RNA polymerase

The RNA polymerase has two α ubunits, one β subunit and a β' subunit, a ω subunit and a σ subunit in its structure.



Functions:

- **DNA-dependent RNA polymerase** requires, in addition to a DNA template, all four ribonucleoside 5' triphosphates (ATP, GTP, UTP, and CTP) as precursors of the nucleotide units of RNA, as well as Mg^{2+}.
- The protein also binds one Zn^{2+}.
- RNA polymerase elongates an RNA strand by adding ribonucleotide units to the 3' hydroxyl end, building RNA in the 5' →3' direction. The 3' hydroxyl group acts as a nucleophile, attacking the 5'phosphate of the incoming ribonucleoside triphosphate.
- RNA polymerase requires DNA for activity and is most active when bound to a double-stranded DNA. only one of the two DNA strands serves as a template.
- The template DNA strand is copied in the 3' →5' direction (antiparallel to the new RNA strand),
- RNA polymerase does not require a primer to initiate synthesis. Initiation occurs when RNA polymerase binds at specific DNA sequences called promoters.
- The moving RNA polymerase generates waves of positive supercoils ahead of the transcriptionbubble and negative supercoils behind of the transcription bubble.

Stages of prokaryotic transcription

The transcription process consist of three stages namely inititation, elongation and termination.

Initiation

All transcription starts at a point called promoter. Prokaryotic promotershave TATA Box (-10) and -35 region

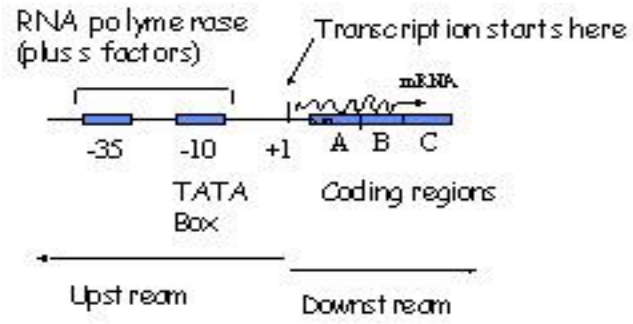
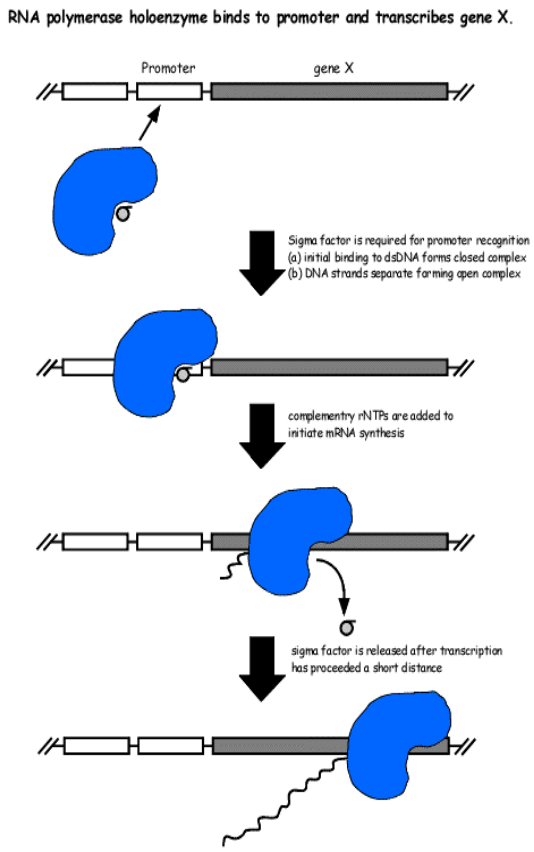


Fig: Prokaryotic promoters

The consensus -35 and -10 sequences for RNA polymerase and s^{70} binding are:

-35	-10
TTGACAT	TATAAT



The following steps occur, in order, for transcription initiation:

- RNA polymerase (RNAP) binds to one of several specificity factors, σ, to form a holoenzyme. In this form, it can recognize and bind to specific promoter regions in the DNA. The -35 region and the -10 ("Pribnow box") region comprise the basic prokaryotic promoter, and |T| stands for the terminator. The DNA on the template strand between the +1 site and the terminator is transcribed into RNA, which is then translated into protein.At this stage, the DNA is double-stranded ("closed"). This holoenzyme/wound-DNA structure is referred to as the closed complex.

- The DNA is unwound and becomes single-stranded ("open") in the vicinity of the initiation site (defined as +1). This holoenzyme/unwound-DNA structure is called the open complex.

- The RNA polymerase transcribes the DNA (the beta subunit initiates the synthesis), but produces about 10 abortive (short, non-productive) transcripts which are unable to leave the RNA polymerase because the exit channel is blocked by the σ-factor.

- The σ-factor eventually dissociates from the holoenzyme, and elongation proceeds.

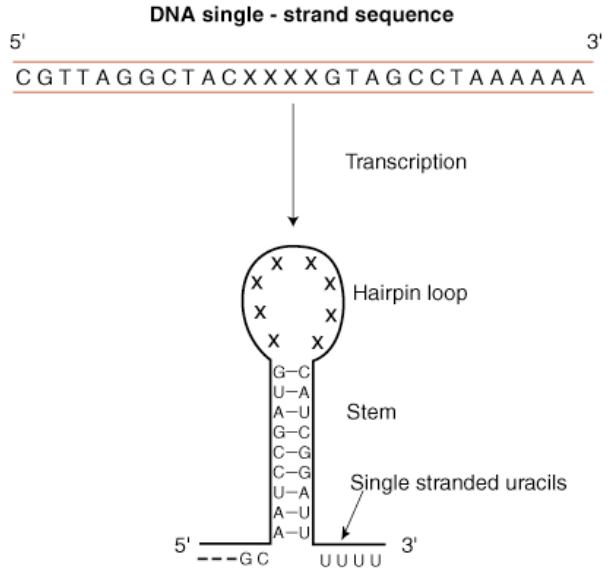
Elongation

One strand of the DNA, the template strand (or noncoding strand), is used as a template for RNA synthesis. As transcription proceeds, RNA polymerase traverses the template strand and uses base pairing complementarity with the DNA template to create an RNA copy. Although RNA polymerase traverses the template strand from 3' → 5', the coding (non-template) strand and newly-formed RNA can also be used as reference points, so transcription can be described as occurring 5' → 3'. The movement of RNA polymerase is called as inch worm movement where the back end of polymerase comprises for every base added, but front end does not move. After adding around 10 nucleotides, the front end jump several basepairs in forward direction.This produces an RNA molecule from 5' → 3', an exact copy of the coding strand (except that thymines are replaced with uracils, and the nucleotides are composed of a ribose (5-carbon) sugar where DNA has deoxyribose (one less oxygen atom) in its sugar-phosphate backbone).

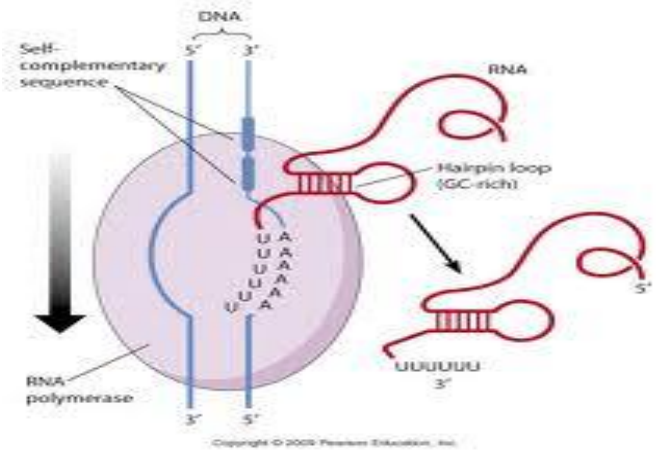
Termination

Two termination mechanisms are well known:

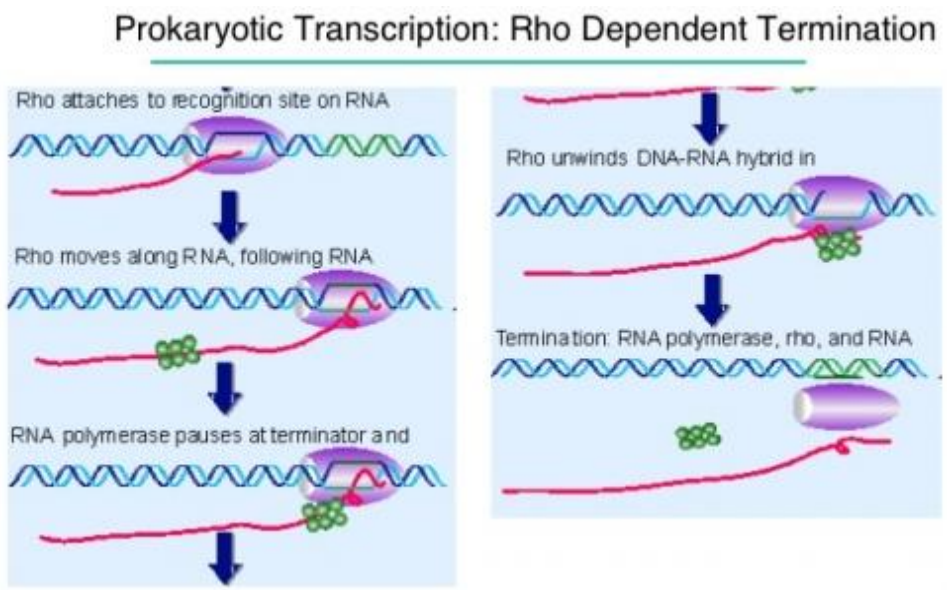
• Intrinsic termination (also called Rho-independent transcription termination) involves terminator sequences within the RNA that signal the RNA polymerase to stop. The terminator sequence is usually a palindromic sequence that forms a stem-loop hairpin structure that leads to the dissociation of the RNAP from the DNA template. Figures: Hairpin loop formation



RNA Sequence



Rho-dependent termination uses a termination factor called ρ factor(rho factor) which is a protein to stop RNA synthesis at specific sites. This protein binds at a rho utilisation site on the nascent RNA strand and runs along the mRNA towards the RNAP. A stem loop structure upstream of the terminator region pauses the RNAP, when ρ-factor reaches the RNAP, it causes RNAP to dissociate from the DNA, terminating transcription.



Eukaryotic transcription

Eukaryotic transcription is more complex than prokaryotic transcription and, until recently, it has seemed that every eukaryotic gene was unique requiring its own transcription machinery.

There are three distinct RNA polymerases in a **eukaryotic cell nucleus** which define the three major classes of eukaryotic transcription unit:

polymerase	location	type of RNA transcribed	sensitivity to □-amanitin§
I	nucleus/nucleolus	rRNA (except for 5S rRNA)	resistant
II	nucleus	hnRNA (i.e. pre-mRNA)	very sensitive
III	nucleus	small RNA such as tRNA and 5S rRNA	moderately sensitive

There may be as many as 14 subunits in an eukaryotic RNA polymerase; the total molecular weight is typically 500-700 kD.

Eukaryotic RNA polymerases cannot find or bind to a promoter by themselves. Each requires the binding of **assembly factors** and a **positional factor** to locate the promoter and to orient the polymerase correctly. As we will see, the positional factor is the same in all cases.

Eukaryotic promoter

Along with **TATA box** it contain **promoter proximal region** and **enhancer** sequences

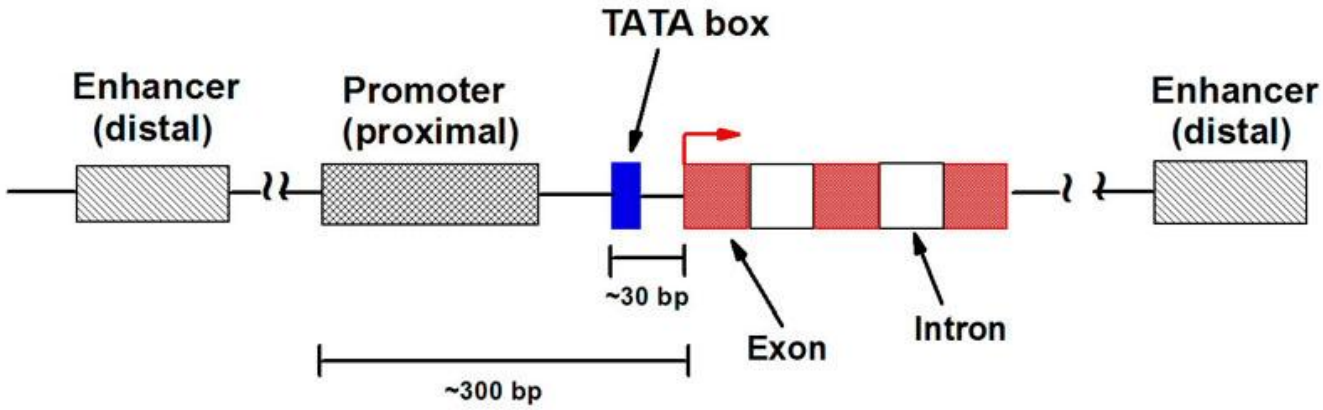


Fig: General structure of eukaryotic promoter

1. Class I Transcriptional Units

Class I genes or transcriptional units are transcribed by RNA polymerase I in the nucleolus. The best-studied examples are the rRNA transcription units:

Each transcription unit consists of 3 rRNA genes: 18S, 5.8S, and 28S; and each unit is separated by a nontranscribed spacer. Eukaryotic nucleoli typically have many hundreds of copies of these transcription units tandemly arranged.

The enzymeRNA polymerase I is a complex of 13 subunits.

The promoter

The **CORE** promoter region is located from **-31** to **+6** around the transcription startpoint. Another sequence further upstream, called the **upstream control element** (**UCE**), located from **-187** to **-107** is also required for efficient transcription.

Both elements are closely related; there is approximately 85% sequence identity between them. These elements are also unusual in that they are GC-rich. In general, sequences around the start-point of transcription tend to be AT-rich so that melting of the DNA duplex is easier.

Assembly of a transcriptional complex

Two additional transcription factors are known to be required to assist RNA polymerase I.**UBF1-** is a single polypeptide which binds to the upstream control element (**UCE**) and to the **CORE** promoter. **UBF1** recognizes a GC-rich sequence within these elements. **UBF1** is an **assembly factor**.

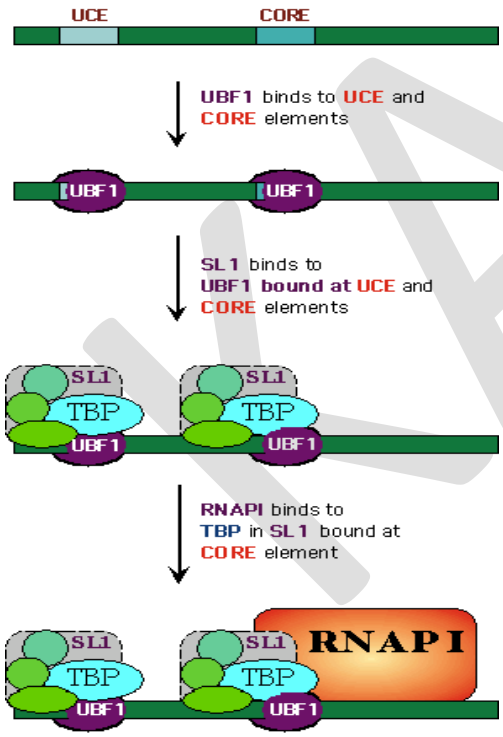


Fig: Synthesis of rRNA in eukaryotes

SL1-binds to **UBF1**. It consists of 4 proteins, one of which is TATA-box binding protein (**TBP**). **TBP** is required for the assembly of a transcriptional complex in all 3 classes of eukaryotic transcription unit. **SL1** is a **positional factor** - it targets RNA polymerase at the promoter so that it initiates transcription in the correct place.

Once **UBF1** and **SL1** have formed a complex, **RNAP I** binds to the **CORE** promoter to initiate transcription: **(Fig)**

2.**Class II Transcription Units**

All genes that are transcribed and expressed via **mRNA** are transcribed by **RNA polymerase II**. Until recently, it was common to think of eukaryotic transcription (and particularly mRNA synthesis) as taking place in discrete steps: transcription, capping, tailing, splicing and export from the nucleus for translation. The contemporary view of eukaryotic geen expression entails simultaneous transcription and processing. Recent discoveries have revealed that many of the protein factors required for these individual steps do, in fact, interact with one another. This makes sense for it allows the cell to coordinate and regulate the complete process more efficiently.

The two images below are from a recent review article. Study them carefully to see how the contemporary view of gene expression (right) contrasts with the more traditional view (left).

The enzyme

RNA polymerase II is a complex multisubunit enzyme - the yeast enzyme has 12 subunits. The largest subunit conatins the catalytic activity.

The promoter

Some of the common elements that have been described in class II eukaryotic promoters are the following:

- The **TATA Box** located approximately **25** bp upstream of the start-point of transcription is found in many promoters. The consenus sequence of this element is **TATAAAA** (so it resembles the TATAAT sequence of the prokaryotic **-10 region** but please do not mix them up). The TATA box appears to be more important for selecting the startpoint of transcription (i.e. positioning the enzyme) than for defining the promoter.
- The **Initiator** is a sequence that is found in many promoters and defines the startpoint of transcription.

- The **GC box** is a common element in eukaryotic class II promoters. Its consensus sequence is **GGGCGG**. It may be present in one or more copies which can be located between **40** and **100** bp upstream of the startpoint of transcription. The transcription factor **Sp1** binds to the **GC box**.
- The **CAAT box** - consensus sequence **CCAAT** - is also often found between **40** and **100** bp upstream of the startpoint of transcription. The transcription factor **CTF** or **NF1** binds to the **CAAT box**.
- In addition to the above elements, **Enhancers** may be required for full expression. These elements are not part of the promoter *per se*. They can be located upstream or downstream of the promoter and may be quite far away from it. The mechanism by which they work is not known. They may provide an **entry point** for RNA polymerase or they may bind other proteins that assist RNA polymerase to bind to the promoter region.

The transcriptional complex

It was found that RNA polymerase II can transcribe mRNA *in vitro* as long as a suitable template -- such as a nicked dsDNA or ssDNA -- is provided . The fact that the enzyme could not initiate transcription correctly on a dsDNA template indicated that additional transcription factors are required.At least six **general (or basal) transcription factors** (TFIIA, TFIIB, TFIID, TFIIE, TFIIF, TFIIH) have been characterized. In the presence of these transcription factors, the enzyme is able to initiate transcription at promoters correctly. However, even in the presence of transcription factors, the enzyme complex is unable to recognize and respond to regulatory signals.

In addition to the general transcription factors, the trsncriptional complex will also be affected by the presence of an promoter-proximal regulatory sequences and the presence of transcription factors that bind to those sequences. Such factors may be present in some cells/tissues but not in others. For example, the octamer motif (shown for the histone H2B gene above) binds two different transcription factors: Oct-1 and Oct-2. Oct-1 is ubiquitous but Oct-2 is expressed only in lymphoid cells where it activates immunoglobulin □ light chain gene transctiption. It has been suggested that the combination of **RNA polymerase II**, **transcription factors**, and regulatory response complexes such as the **Srb-Mediator,** is the eukaryotic equivalent of a **holoenzyme**.

Assembly of a transcriptional complex

The following image shows a model for the assembly of a Class II eukaryotic transcription complex:

The basic process is likely to include the following steps:

- TFIID recognizes and binds to the TATA box.TFIID consists of TATA box binding protein - TBP and ~10 TBP associated factors - TAFs.
- TBP is a 180 amino acid protein that consists of two very similar 66 amino acid domains separated by a short basic region. The protein has a "saddle-shaped" structure that sits astride a DNA molecule and binds to it via contacts in the minor groove. Binding also causes an 80° bend in the DNA.
- TFIID is a positional factor - it targets RNA polymerase to the promoter. In the case of class II transcriptional units, however, TBP binds directly to DNA.
- TFIIA binds and stabilizes TFIID binding.
- The RNA polymerase II holoenzyme assembles - possibly in a stepwise manner to form a preinitiation complex
- The holoenzyme consists of the RNA polymerase II complex, the regulatory complexes and the following transcription factors:
- TFIIB-TFIIB is a single polypeptide. It can bind both upstream and downstream of the TATA box (i.e. closer to the startpoint of transcription). It recruits TFIIF-RNAPII to the complex. It may interact directly with RNAP II.
- TFIIE-TFIIE is a complex of two subunits. It recruits TFIIH to the complex thereby priming the initiation complex for promoter clearance and elongation.
- TFIIF-TFIIF also has two subunits - RAP38 & RAP74. The latter has a helicase activity and may therefore be involved in melting the DNA at the promoter to expose the template strand.
- TFIIH-TFIIH is a complex of 9 subunits. One of the subunits has a kinase activity that carries out the phosphorylation that is required for promoter clearance
  The two largest subunits (**XPB** and **XPD**) have **helicase** activity; this activity of TFIIH is also required for Nucleotide Excision Repair in the cell and mutations in these subunits are associated with three genetic disorders: **Xeroderma pigmentosum**, **Cockayne's**

disease (repair defects) and **Trichothiodystrophy** (a transcription defect). Another subunit is a **cyclin** (**cdk7 - cyclinH**)

- There is some evidence that the order of assembly of transcription factors may be **TFIID ->TFIIA ->TFIIB** -> (**TFIIF** + **RNAP II**) ->**TFIIE ->TFIIH**

Finally, the various regulatory factors (**Srb-Mediator**, **Srb10-CDK** and **Swi-Snf**) bind to complete formation of the **pre-initiation complex**.

The carboxy terminal domain (CTD) of the largest subunit of RNA polymerase II is phosphorylated. This results in promoter clearance. RNA polymerase II dissociates from the Transcription factors and other protein complexes that were required for assembly.

The CTD consists of 52 repeats of the amino acid sequence Y-S-P-T-S-P-S. Ser5 is phosphorylated by tthe kinase activity of TFIIH.TFIIH phosphorylates Ser5. This serves to recruit the transcription elongation factor DSIF to the complex, which in turn recruits another elongation factor, NELF, which arrests transcription. This pause permits the capping enzymes to join and modify the 5' end of the transcript. A third elongation factor P-TEFb (a kinase) joins and phosphorylates CTD and NELF, neutralising them. P-TEFb phosphorylates CTD at Ser2.

There is also evidence that an additional transcription factor, **TFIIS**, participates in **transcription elongation.**

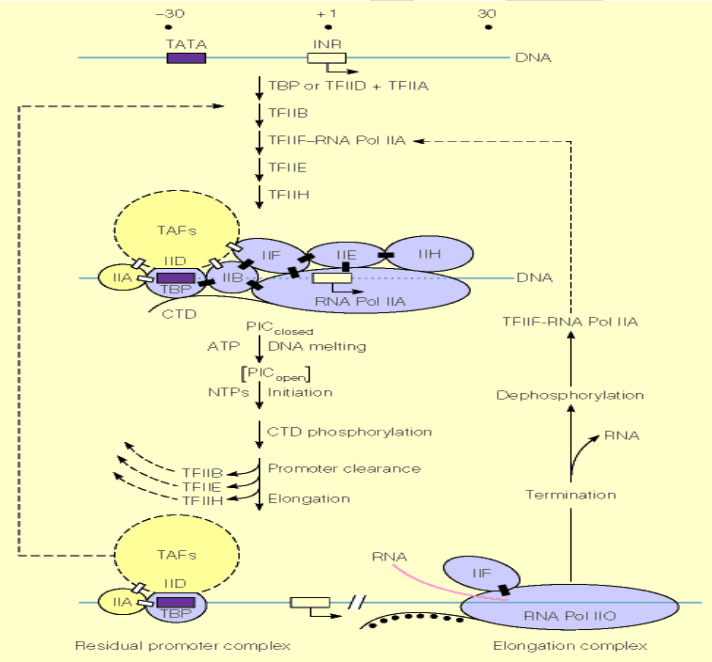
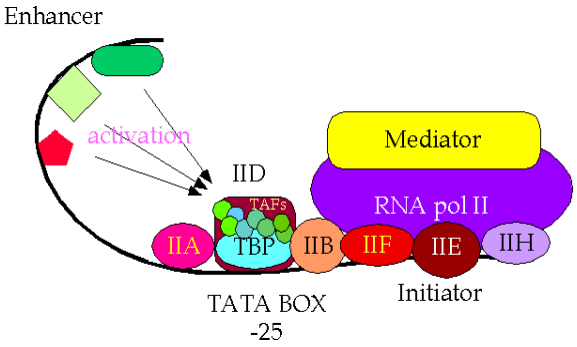


Fig: Synthesis of mRNA in eukaryotes



3.Class III Transcription Units

Class III genes are principally those for small RNA molecules in the cell. The best studied examples are the 5S rRNA gene -- which has been studied extensively in *Xenopus laevis*, and tRNA genes.

The enzyme

RNA polymerase III is the largest of the three RNA polymerases with 17 subunits and a molecular weight of over 700 kD. It is moderately sensistive to □-amanitin. It is also the most active.

The promoter

Class III promoters are distinctive because some of them are located within the gene whose transcription they direct.The promoters for 5S rRNA and tRNA genes are located within the gene. In the case of the *Xenopus laevis* 5S rRNA gene, which is 120 bp in length, it has been found that the segment from +41 to +87 is sufficient to direct transcription and therefore defines the promoter.

The promoters for snRNA genes lie upstream of the startpoint of transcription.

Assembly of a transcriptional complex

Assembly of a functional complex requires the participation of a number of additional transcription factors. Transcription of the 5S rRNA gene requires three additional factors; transcription of tRNA genes require two.

TFIIIA-This factor is required only for the transcription of 5S rRNA genes. It contains a single polypeptide with a Zn finger DNA-binding motif. It functions as an assembly factor for some class III promoters but not for all.

TFIIIB-This factor contains three subunits, one of which is TBP - TATA-box binding protein. TFIIIB is a positional factor.

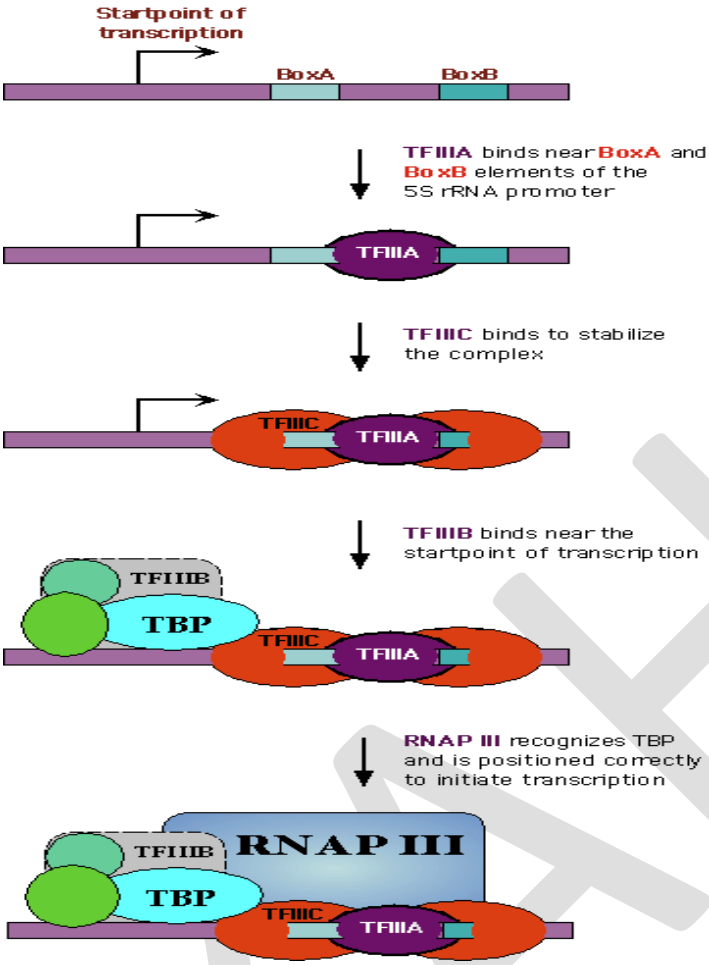
Look at Protein-DNA and Protein-Protein interactions in the TFIIIB-DNA Complex from the Hahn laboratory at the Fred Hutchinson Cancer Research Center in Seattle.

TFIIIC-TFIII consists of 6 subunits. It also functions as an assembly factor and appears to be required for all internal class III promoters.

Assembly of a transcription complex proceeds in a step-wise manner:

- TFIIIA binds to a site within the promoter region.
- TFIIIC binds to form a stable complex. TFIIIC covers the entire gene.
- TFIIB can now bind to its binding site surrounding the startpoint of transcription.
- Finally RNA polymerase III is able to bind and begin transcription of the gene.

The following diagram illustrates these steps for the 5S rRNA promoter:



Differences between transcription of prokaryotes and eukaryotes.

There are significant differences in the process of transcription in prokaryotes versus eukaryotes.

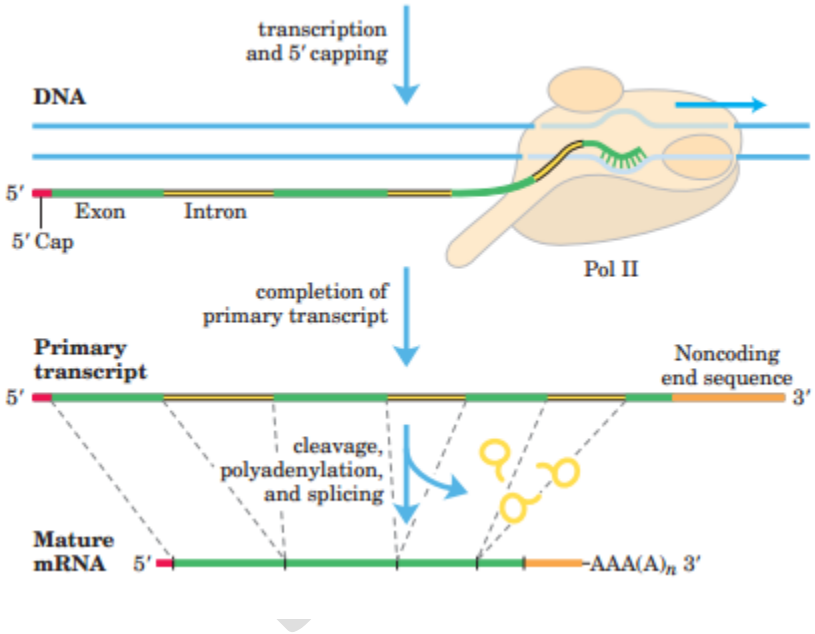
- In prokaryotes (bacteria), transcription occurs in the cytoplasm. Translation of the mRNA into proteins also occurs in the cytoplasm. In eukaryotes, transcription occurs in the cell's nucleus. mRNA then moves to the cytoplasm for translation.

- DNA in prokaryotes is much more accessible to RNA polymerase than DNA in eukaryotes. Eukaryotic DNA is wrapped around proteins called histones to form structures called nucleosomes. Eukaryotic DNA is packed to form chromatin. While RNA polymerase interacts directly with prokaryotic DNA, other proteins mediate the interation between RNA polymerase and DNA in eukaryotes.

• mRNA produced as a result of transcription is not modified in prokaryotic cells. Eukaryotic cells modify mRNA by RNA splicing, 5' end capping, and addition of a polyA tail.

Transcription and processing of mRNA

A newly synthesized RNA molecule is called a primary transcript. The most extensive processing of primary transcripts occurs in eukaryotic mRNAsand in tRNAs of both bacteria and eukaryotes. Noncoding tracts that break up the coding region of the transcript are called introns, and thecoding segments are called exons. In a processcalled splicing,the introns are removed from the primary transcript and the exons are joined to form a continuous sequence that specifies a functional polypeptide. Eukaryotic mRNAs are also modified at each end.A modified residue called a 5'cap is added atthe 5' end. The 3' end is cleaved and 80 to 250 Aresidues are added to create a poly(A) "tail." The composition of the complex changes as the primary transcript is processed,transported to the cytoplasm, and delivered to the ribosome for translation.

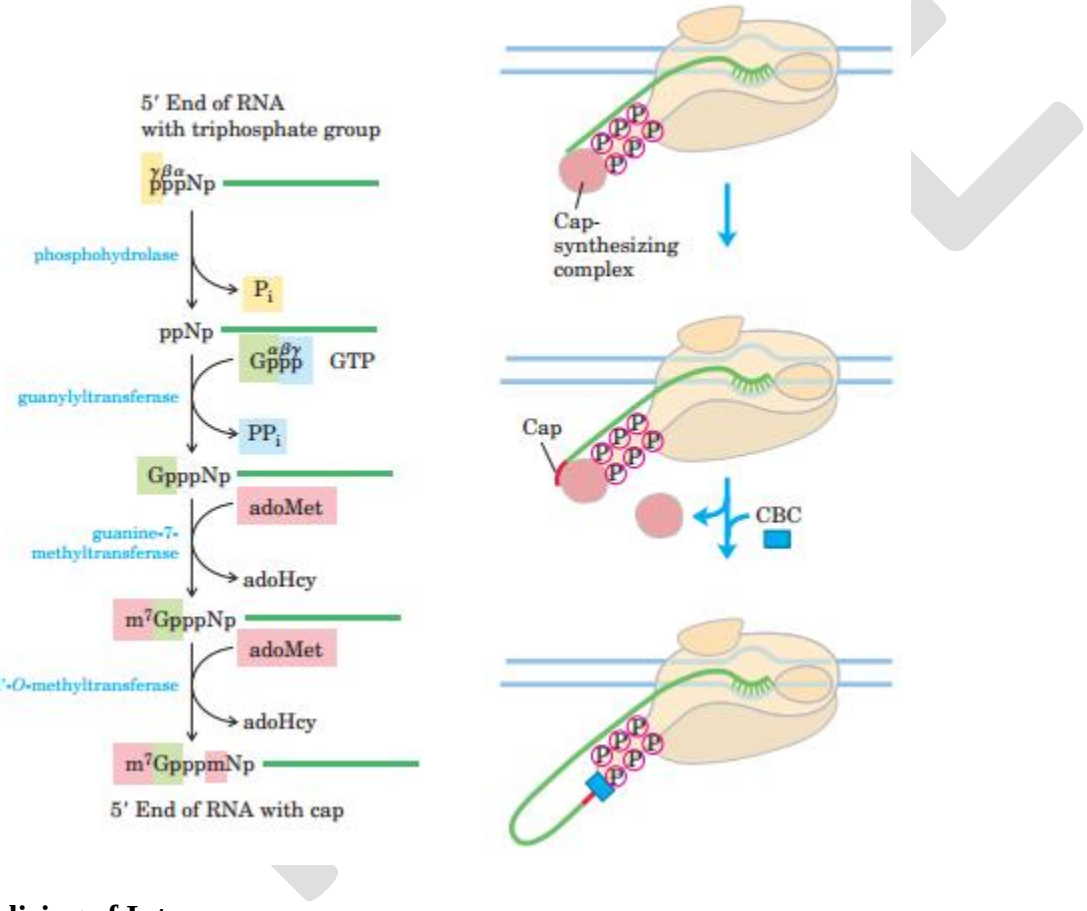


Capping of the eukaryotic mRNA at the 5' end:

Most eukaryotic mRNAs have a 5' cap,a residue of 7-methylguanosine linked to the 5'-terminal residue of themRNA through an unusual 5',5'-triphosphate linkage. The 5' cap helps

protect mRNA fromribonucleases. The cap also binds to a specific cap-binding complex of proteins and participates in bindingof the mRNA to the ribosome to initiate translation.

The 5'cap is formed by condensation of a moleculeof GTP with the triphosphate at the 5'end of the transcript. The guanine is subsequently methylated at N-7,and additional methyl groups are often added at the 2' hydroxyls of the first and second nucleotides adjacent to the cap. The methyl groups are derivedfrom S-adenosylmethionine. All these reactions occurvery early in transcription, after the first 20 to 30 nucleotides of the transcript have been added. The capped5' end is then released from the capping enzymes andbound by the cap-binding complex.



Splicing of Introns:

There are four classes of introns. The first two, thegroup I and group II intronsare self-splicing—no protein enzymes areinvolved. Group I introns are found in some nuclear, mitochondrial, and chloroplast genes coding for rRNAs,mRNAs, and tRNAs. Group II introns are generally foundin the primary transcripts of mitochondrial or chloroplast mRNAs in fungi, algae, and plants. Neither class requires a high-energy cofactor (such as ATP) for splicing.

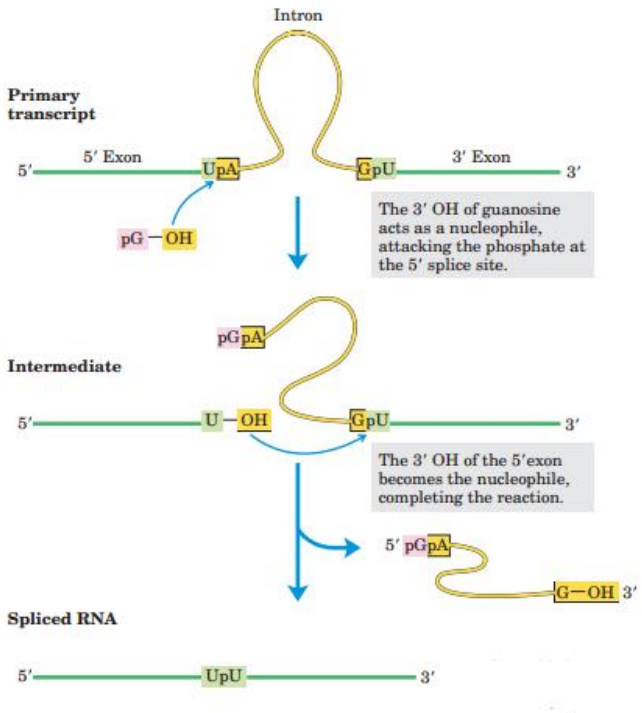
The splicingmechanisms in both groups involve two transesterification reaction steps.

Splicing of group I and II introns:

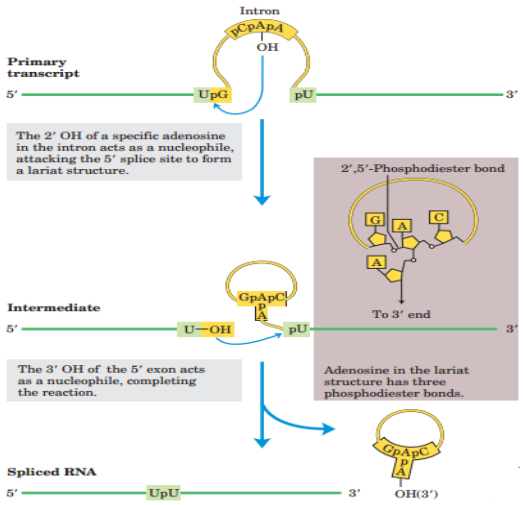
The group I splicing reaction requires a guanine nucleoside or nucleotide cofactor, but the cofactor is notused as a source of energy; instead, the 3'-hydroxylgroup of guanosine is used as a nucleophile in the firststep of the splicing pathway. The guanosine 3'-hydroxylgroup forms a normal 3',5'-phosphodiester bond withthe 5' end of the intron. The 3' hydroxylof the exon that is displaced in this step then acts as anucleophile in a similar reaction at the 3'end of the introns. The result is precise excision of the intron and ligation of the exons.

In group II introns the reaction pattern is similar except for the nucleophile in the first step, which in thiscase is the 2'-hydroxyl group of an A residue within the intron. A branched lariat structure isformed as an intermediate.

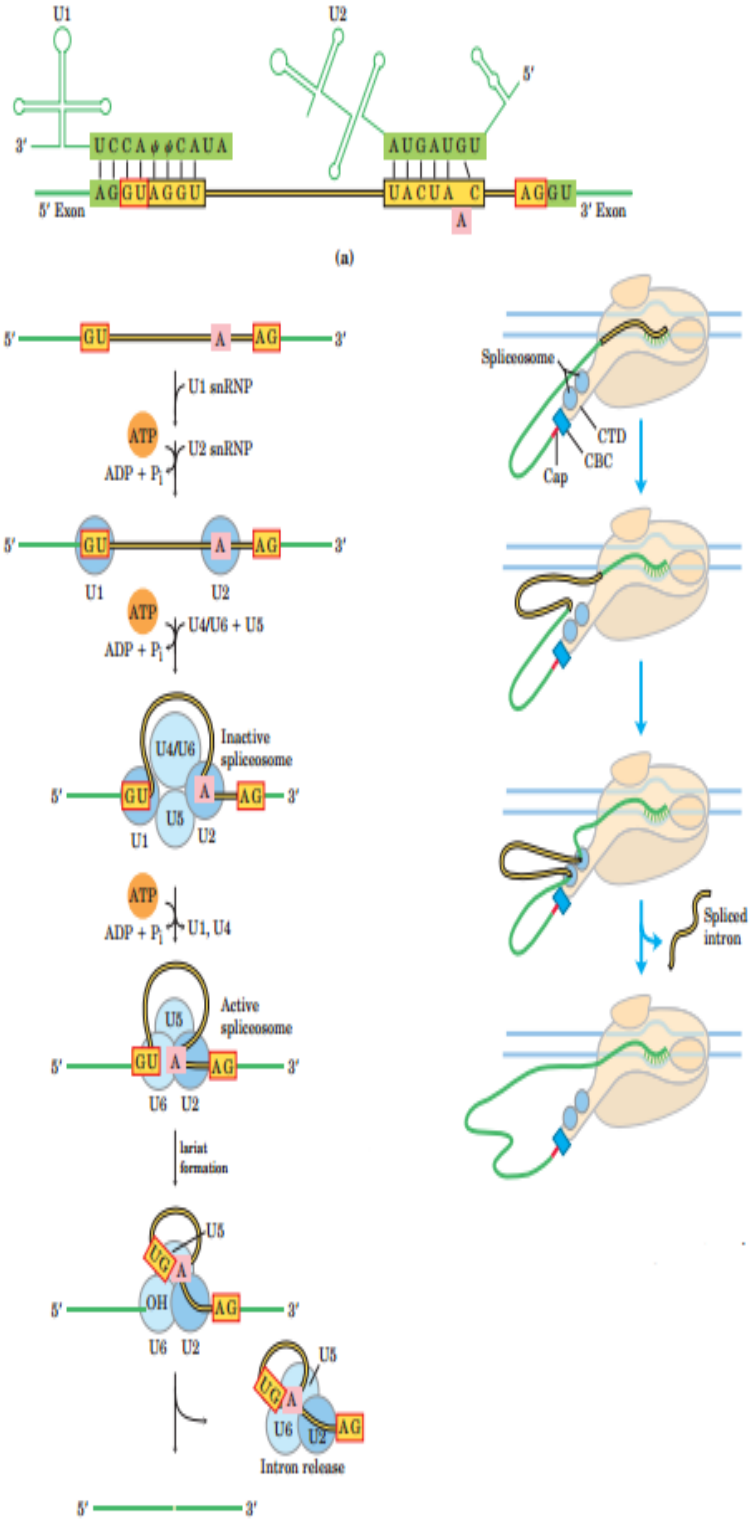
Splicing of group I introns



Splicing of group II introns



The third and largest class of introns includes those found in nuclearmRNA primary transcripts. These are called spliceosomal introns,because their removal occurs withinand is catalyzed by a large protein complex called aspliceosome.


U1
U2
5′
3′
UCCA
CAUA
AUGAUGU
AG GU AGGU
UACUA C
AG GU
A
5′ Exon
3′ Exon
(a)
5′ GU A AG 3′
U1 snRNP
ATP
U2 snRNP
ADP + Pi
U1
U2
U4/U6 + U5
U4/U6
Inactive spliceosome
U5
Spliceosome
CTD
CBC
Cap
Spliced intron
U1, U4
Active spliceosome
U6
lariat formation
OH
Intron release

Within the spliceosome, the introns undergo splicing by the same lariat-forming mechanism asthe group II introns. The spliceosome is made up of specialized RNA-protein complexes, small nuclear ribonucleoproteins (snRNPs, or "snurps").Each snRNP contains one of a class of eukaryotic RNAs,100 to 200 nucleotides long, known as small nuclearRNAs (snRNAs).Five snRNAs (U1, U2, U4, U5, and U6) involved in splicing reactions are generally found inabundance in eukaryotic nuclei. The RNAs and proteinsin snRNPs are highly conserved in eukaryotes fromyeasts to humans.

Spliceosomal introns generally have the dinucleotide sequence GU and AG at the 5'and 3'ends, respectively, and these sequences mark the sites wheresplicing occurs. The U1 snRNA contains a sequencecomplementary to sequences near the 5' splice site ofnuclear mRNA introns and the U1 snRNPbinds to this region in the primary transcript. Additionof the U2, U4, U5, and U6 snRNPs leads to formation of the spliceosome. The snRNPs togethercontribute five RNAs and about 50 proteins to thespliceosome, a supramolecular assembly nearly as complex as the ribosome. ATP isrequired for assembly of the spliceosome, but the RNAcleavage-ligation reactions do not seem to require ATP.Some mRNA introns are spliced by a less common typeof spliceosome, in which the U1 and U2 snRNPs are replaced by the U11 and U12 snRNPs. Whereas U1- andU2-containing spliceosomes remove introns with (5')GUand AG(3') terminal sequences, the U11- and U12-containing spliceosomes remove a rare class of introns that have (5')AU and AC(3')terminal sequences to mark the intronic splice sites.

The fourth class of introns, found in certain tRNAs,is distinguished from the group I and II introns in thatthe splicing reaction requires ATP and an endonuclease. The splicing endonuclease cleaves the phosphodiester bonds at both ends of the intron, and the two exons are joined by a mechanism similar to the DNA ligase reaction.

Addition of poly A tail:

At their 3' end, most eukaryotic mRNAs have a stringof 80 to 250 A residues, making up the poly(A) tail.This tail serves as a binding site for one or more specific proteins. The poly(A) tail and its associated proteins probably help protect mRNA from enzymatic destruction. Many prokaryotic mRNAs also acquirepoly(A) tails, but these tails stimulate decay of mRNA

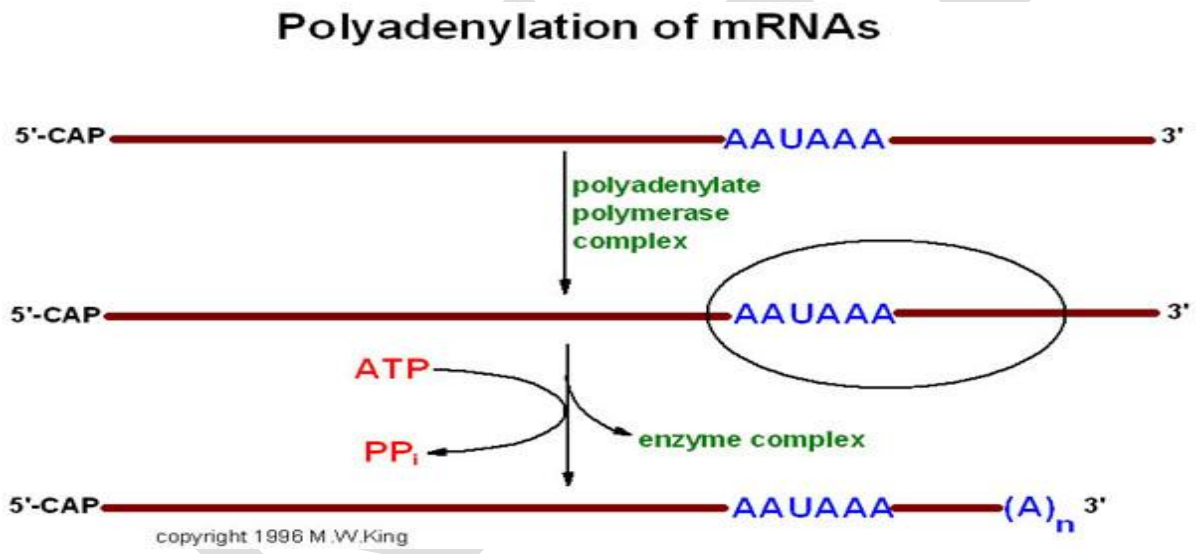
rather than protecting it from degradation.

The poly(A) tail is added in a multistep process.The transcript is extended beyond the site where thepoly(A) tail is to be added, then is cleaved at the poly(A)addition site by an endonuclease component of a largeenzyme complex, again associated with the CTD of RNA polymerase II. The mRNA site where cleavage occurs is marked by two sequence elements: thehighly conserved sequence (5')AAUAAA(3'), 10 to 30nucleotides on the 5' side (upstream) of the cleavagesite, and a less well-defined sequence rich in G and Uresidues, 20 to 40 nucleotides downstream of the cleavage site. Cleavage generates the free 3'-hydroxyl groupthat defines the end of the mRNA, to which A residuesare immediately added by polyadenylate polymerase,

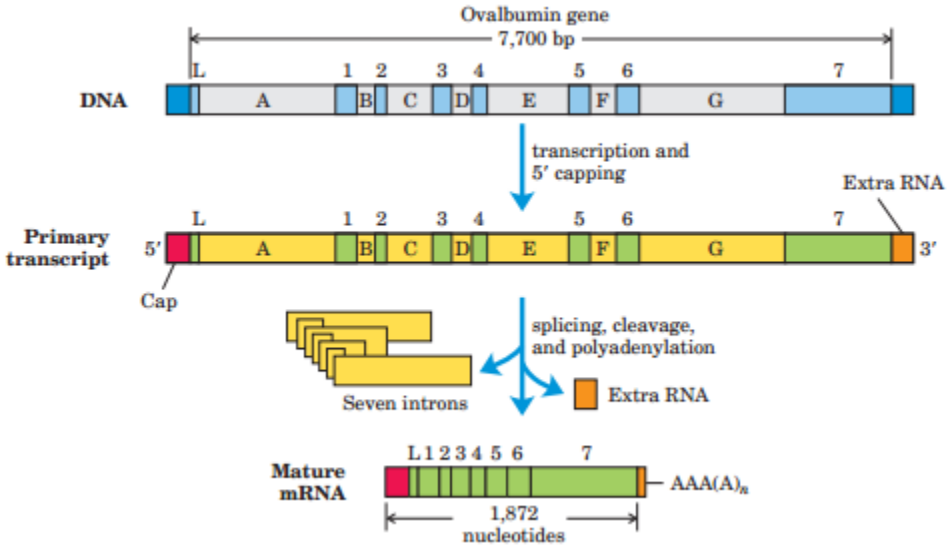
which catalyzes the reaction

$$\text{RNA} + n\text{ATP} \longrightarrow \text{RNA–(AMP)}_n + n\text{PP}_i$$

where, n = 80 to 250. This enzyme does not require atemplate but does require the cleaved mRNA as a primer.



Overall processing of mRNA



Processing of rRNA and tRNA

- Ribosomal RNAs of both prokaryotic and eukaryotic cellsare made from longer precursors called preribosomalRNAs,or pre-rRNAs, synthesized by Pol I.
- In bacteria,16S, 23S, and 5S rRNAs arise from a single30S RNA precursor of about 6,500 nucleotides.
- RNA atboth ends of the 30S precursor and segments betweenthe rRNAs are removed during processing.

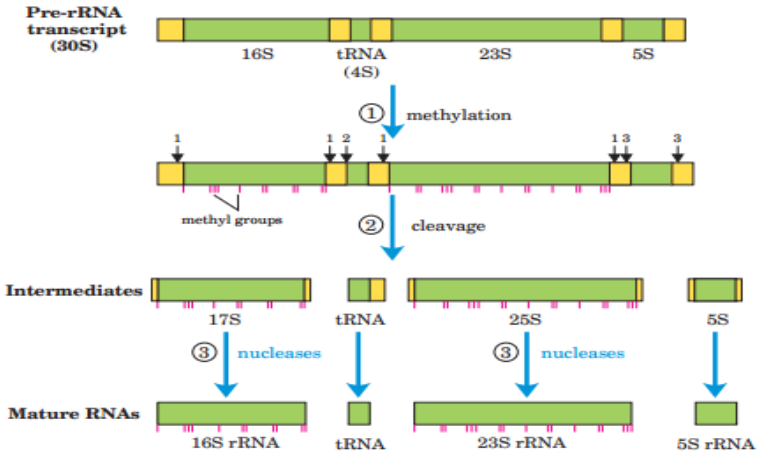


Figure: Processing of pre rRNA transcript

Fig:Processing of pre-rRNA transcripts in bacteria: Before cleavage, the 30S RNA precursor is methylated at specific bases. 2 Cleavage liberates precursors of rRNAs and tRNA(s). Cleavage at the points labeled 1, 2, and 3 is carried out by the enzymes RNase III, RNase P, and RNase E, respectively. RNase P is a ribozyme. 3 The final 16S, 23S, and 5S rRNA products result from the action of a variety of specific nucleases.

- The genome of E. coliencodes seven pre-rRNA molecules.
- All these genes have essentially identical rRNA-coding regions, but they differ in the segments betweenthese regions.
- The segment between the 16S and 23SrRNA genes generally encodes one or two tRNAs, withdifferent tRNAs arising from different pre-rRNA transcripts.
- Coding sequences for tRNAs are also found onthe 3'side of the 5S rRNA in some precursor transcripts.

Processing of pre rRNA transcript in vertebrates

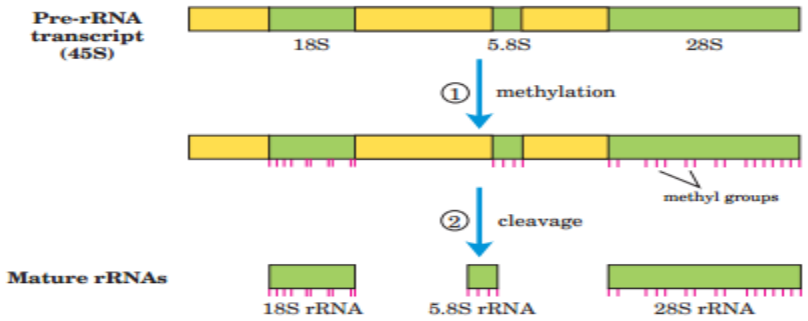


Figure:In step 1, the 45S precursor is methylated at more than 100 of its 14,000 nucleotides, mostly on the 2'-OH groups of ribose units retained in the final products. 2 A series of enzymatic cleavages produces the 18S, 5.8S, and 28S rRNAs. The cleavage reactions require RNAs found in the nucleolus, called small nucleolar RNAs (snoRNAs), within protein complexes reminiscent of spliceosomes. The 5S rRNA is produced separately.

- In eukaryotes, a 45S pre-rRNA transcript isprocessed in the nucleolus to form the 18S, 28S, and5.8S rRNAs characteristic of eukaryotic ribosomes
- The 5S rRNA of most eukaryotes is made as acompletely separate transcript by a different polymerase (Pol III instead of Pol I).
- Most cells have 40 to 50 distinct tRNAs, and eukaryotic cells have multiple copies of many of the tRNAgenes.
- Transfer RNAs are derived from longer RNA precursors by enzymatic removal of nucleotides from the5' and 3' ends.
- In eukaryotes, introns arepresent in a few tRNA transcripts and must be excised.
- Where two or more different tRNAs are contained ina single primary transcript, they are separated byenzymatic cleavage. The endonuclease RNase P, foundin all organisms, removes RNA at the 5'end of tRNAs.
- This enzyme contains both protein and RNA.
- The RNAcomponent is essential for activity, and in bacterial cellsit can carry out its processing function with precisioneven without the protein component.

- RNase P is therefore another example of a catalytic RNA.
- The 3'end of tRNAs is processedby one or more nucleases, including the exonucleaseRNase D.

3. tRNA processing

1. tRNA is transcribed by RNA polymerase III. The transcription product, the pre-tRNA, contains additional RNA sequences at both the 5' and 3'-ends. These additional sequences are removed from the transcript during processing. The additional nucleotides at the 5'-end are removed by an unusual RNA containing enzyme called **ribonuclease P** (RNase P) and **ribonuclease D** at 3' end (Ribozymes-RNA acting as enzymes).

2. Some tRNA precursors contain an intron located in the anticodon arm. These introns are spliced out during processing of the tRNA.

3. All mature tRNAs contain the trinucleotide CCA at their 3'-end. These three bases are not coded for by the tRNA gene. Instead, these nucleotides are added during processing of the pre-tRNA transcript. The enzyme responsible for the addition of the CCA-end is tRNA nucleotidyl transferase and the reaction proceeds according to the following scheme:

tRNA +CTP --> tRNA-C + PPi (pyrophosphate)

tRNA-C +CTP --> tRNA-C-C + PPi

tRNA-C-C +ATP --> tRNA-C-C-A + PPi

Mature tRNAs can contain up to 10% bases other than the usual adenine (A), guanine (G), cytidine (C) and uracil (U). These base modifications are introduced into the tRNA at the final processing step. The biological function of most of the modified bases is uncertain and the translation process seems normal in mutants lacking the enzymes responsible for modifying the bases.

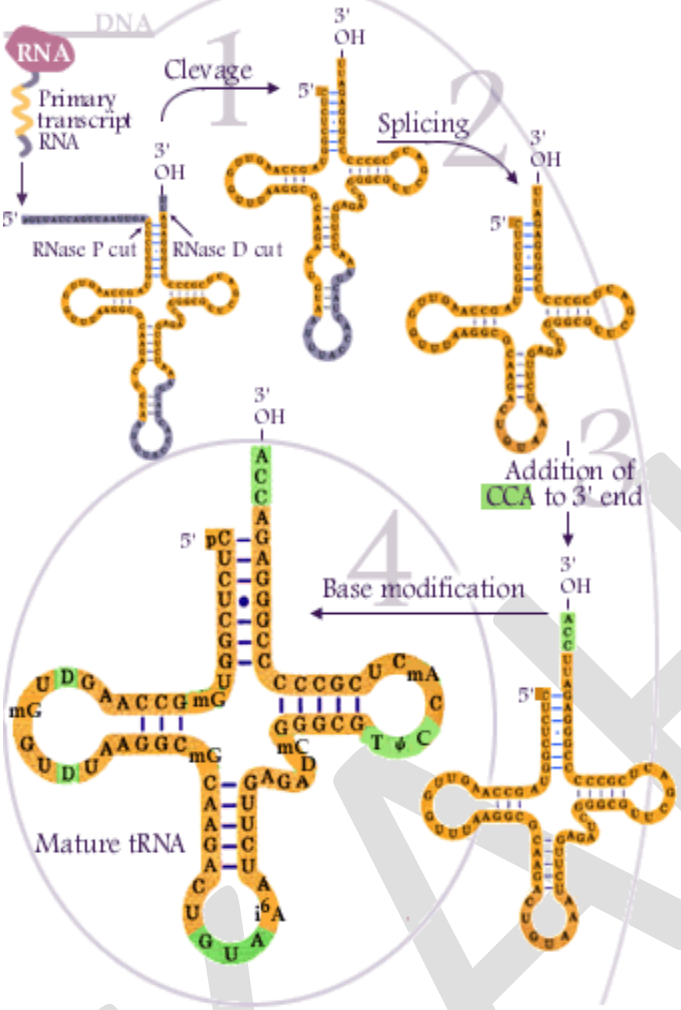


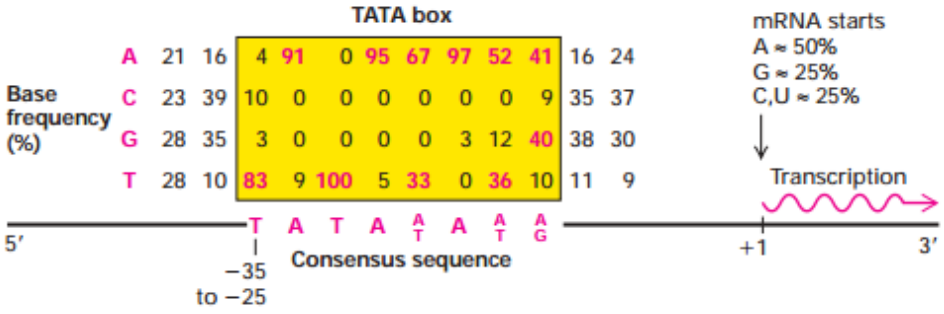
Fig: tRNA processing

Regulatory sequences in protein coding genes:

TATA box

In all the rapidly transcribing genes, a conserved sequence called the TATA boxwas found ≈25–35 base pairs upstream of the start site. Mutagenesis studieshave shown that a single-base change in this nucleotide sequence drastically decreases in vitro transcription by RNApolymerase II of genes adjacent to a TATA box. In mostcases, sequence changes between the TATA box and start sitedo not significantly affect the transcription rate. If the basepairs between the TATA box and the normal start site aredeleted, transcription of the altered, shortened template begins at a new

site ≈25 base pairs downstream from the TATAbox. Consequently, the TATA box acts similarly to an E. colipromoter to position RNA polymerase II for transcriptioninitiation.



Initiators:

Instead of a TATA box, some eukaryotic genes containan alternative promoter element called an initiator. Mostnaturally occurring initiator elements have a cytosine (C) at the -1 position and an adenine (A) residue at the transcription start site (+1). Directed mutagenesis of mammalian genes with an initiator-containing promoter hasrevealed that the nucleotide sequence immediately surrounding the start site determines the strength of such promoters. Unlike the conserved TATA box sequence, however,only an extremely degenerate initiator consensus sequencehas been defined:

(5') Y-Y-A^{+1}-N-T/A-Y-Y-Y (3')

where $A^{+1}$is the base at which transcription starts, Y is apyrimidine (C or T), N is any of the four bases, and T/A is T or A at position +3.Transcription of genes with promoters containing aTATAbox or initiator element begins at a well-defined initiation site.

CpG island:

Transcription of many protein-codinggenes has been shown to begin at any one of multiple possible sites over an extended region, often 20–200 base pairsin length. As a result, such genes give rise to mRNAs withmultiple alternative 5'ends. These genes, which generally are transcribed at low rates (e.g., genes encoding the enzymes ofintermediary metabolism, often called "housekeepinggenes"), do not contain a TATA box or an initiator. Mostgenes of this type contain a CG-rich stretch of 20–50 nucleotides within ≈100 base pairs upstream of the start-site

region. The dinucleotide CG is statistically underrepresented invertebrate DNAs, and the presence of a CG-rich region, orCpG island,just upstream from a start site is a distinctlynonrandom distribution. For this reason, the presence of aCpG island in genomic DNA suggests that it may contain a

transcription-initiation region.

Promoter proximal element:

Promoter-proximal elements are the control regions lying within 100–200 base pairs upstreamof the transcription start site. In some cases, promoter-proximal elementsare cell-type-specific; that is, they function only in specificdifferentiated cell types.

One approach frequently taken to determine the up-stream border of a transcription-control region for a mammalian gene involves constructing a set of 5'deletions. Once the 5'border of a transcription-control region is determined, analysis oflinker scanning mutationscan pinpoint the sequences withregulatory functions that lie between the border and thetranscription start site. One of the first usesof this type of analysis identified promoter-proximal elements of the thymidine kinase (tk) gene from herpes simplex virus (HSV). The results demonstrated that the DNAregion upstream of the HSV tkgene contains three separate transcription-control sequences: a TATA box in the interval from -32 to -16, and two other control elements farther upstream.

To test the spacing constraints on control elements inthe HSV tkpromoter region identified by analysis of linkerscanning mutations, researchers prepared and assayedconstructs containing small deletions and insertions between the elements. Changes in spacing between the promoter and promoter-proximal control elements of 20nucleotides or fewer had little effect. However, insertionsof 30 to 50 base pairs between a promoter-proximal element and the TATA box were equivalent to deleting the element. Similar analyses of other eukaryotic promotershave also indicated that considerable flexibility in thespacing between promoter-proximal elements is generallytolerated, but separations of several tens of base pairs maydecrease transcription.

Activators and repressors of transcription:

Activators are modular proteins composed of distinct functional domains. Studies with a yeast transcription activator called GAL4 provided early insight into the domain structure of transcription factors. The gene encoding the GAL4 protein, whichpromotes expression of enzymes needed to metabolize galactose, was identified by complementation analysis of gal4

mutants. Directed mutagenesis studies like thosedescribed previously identified UASs for the genes activatedby GAL4. Each of these UASs was found to contain one ormore copies of a related 17-bp sequence called UAS_{GAL}.DNase I footprinting assays with recombinant GAL4 proteinproduced in E. colifrom the yeast GAL4 gene showed thatGAL4 protein binds to UAS_{GAL}sequences. When a copy ofUAS_{GAL} was cloned upstream of a TATA box followed by alacZreporter gene, expression of lacZwas activated in galactose media in wild-type cells, but not in gal4mutants. Theseresults showed that UAS_{GAL} is a transcriptioncontrol element activated by the GAL4 protein in galactose media.

A remarkable set of experiments with gal4deletion mutants demonstrated that the GAL4 transcription factor iscomposed of separable functional domains: an N-terminalDNA-binding domain,which binds to specific DNA sequences, and a C-terminal activation domain,which interacts with other proteins to stimulate transcription from a nearby promoter. When the N-terminalDNA-binding domain of GAL4 was fused directly to various of its C-terminal fragments, the resulting truncated proteinsretained the ability to stimulate expression of a reporter genein an in vivo assay like that depicted in Figure 11-16. Thusthe internal portion of the protein is not required for functioning of GAL4 as a transcription factor. Similar experiments with another yeast transcription factor, GCN4, whichregulates genes required for synthesis of many amino acids,indicated that it contains an ≈60-aa DNA-binding domain atits C-terminus and an ≈20-aa activation domain near themiddle of its sequence.

Further evidence for the existence of distinct activationdomains in GAL4 and GCN4 came from experiments inwhich their activation domains were fused to a DNA-bindingdomain from an entirely unrelated E. coliDNA-binding protein. When these fusion proteins were assayed in vivo, theyactivated transcription of a reporter gene containing the cognate site for the E. coliprotein. Thus functional transcriptionfactors can be constructed from entirely novel combinationsof prokaryotic and eukaryotic elements.Studies such as these have now been carried out withmany eukaryotic activators. The structural model of eukaryotic activators that has emerged from these studies is a modular one in which one or more activation domains areconnected to a sequence-specific DNA-binding domainthrough flexible protein domains. In somecases, amino acids included in the DNA-binding domain alsocontribute to transcriptional activation. Activation domains are thought to function bybinding other proteins involved in transcription. The presence of flexible domains connecting the DNA-binding domains to activation domains may explain why alterations inthe spacing between control elements are so well toleratedin eukaryotic control regions. Thus even when the positionsof transcription factors bound to DNA are shifted relativeto each other, their activation domains may still be able to interact because they are attached to their DNA-binding domains through flexible protein regions.

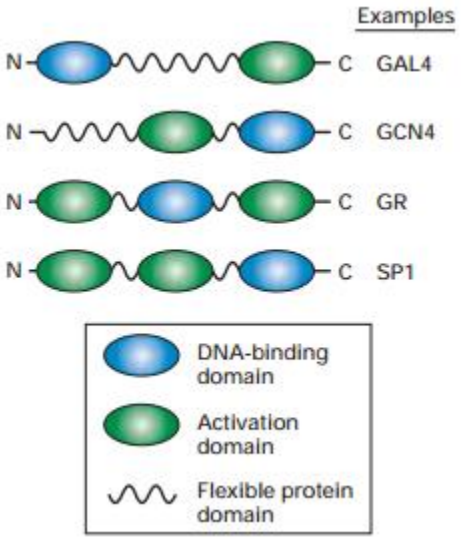


Figure: Eukaryotic transcription activators

Repressors:

Eukaryotic transcription is regulated by repressors as wellas activators. For example, geneticists have identified mutations in yeast that result in continuously high expression ofcertain genes. This type of unregulated, abnormally high expression is called constitutiveexpression and results from theinactivation of a repressor that normally inhibits the transcription of these genes. Similarly, mutants of Drosophilaand C. eleganshave been isolated that are defective in embryonic development because they express genes in embryonic cells where they are normally repressed. The mutationsin these mutants inactivate repressors, leading to abnormaldevelopment.

Repressor-binding sites in DNA have been identified bysystematic linker scanning mutation analysis. In this type of analysis, mutation ofan activator-binding site leads to decreased expression of the linked reporter gene, whereas mutation of a repressor-binding site leads to increased expression of a reporter gene.Repressor proteins that bind such sites can be purified andassayed using the same biochemical techniques describedearlier for activator proteins.

Eukaryotic transcription repressors are the functionalconverse of activators. They can inhibit transcription froma gene they do not normally regulate when their cognatebinding sites are placed within a few hundred base pairs ofthe gene's start site. Like activators, most eukaryotic repressors are modular proteins that have two functional domains:a DNA-binding domain and a repression domain.Similar toactivation domains, repression domains continue to functionwhen

fused to another type of DNA-binding domain. Ifbinding sites for this second DNA-binding domain are inserted within a few hundred base pairs of a promoter, expression of the fusion protein inhibits transcription from thepromoter. Also like activation domains, repression domainsfunction by interacting with other proteins as discussed later.The absence of appropriate repressor activity canhave devastating consequences. For instance, theprotein encoded by the Wilms' tumor (WT1) geneis a repressor that is expressed preferentially in the developing kidney. Children who inherit mutations in both the ma-ternal and paternal WT1genes, so that they produce nofunctional WT1 protein, invariably develop kidney tumorsearly in life. The WT1 protein binds to the control region ofthe gene encoding a transcription activator called EGR-1. This gene, like many other eukaryotic genes,is subject to both repression and activation. Binding by WT1represses transcription of the EGR-1gene without inhibitingbinding of the activators that normally stimulate expressionof this gene.

Multiple transcription control element:

Initially, enhancers and promoter-proximal elements werethought to be distinct types of transcription control elements. However, as more enhancers and promoter-proximalelements were analyzed, the distinctions between them became less clear. For example, both types of element generally can stimulate transcription even when inverted, and bothtypes often are cell-type-specific. The general consensus nowis that a spectrum of control elements regulates transcription by RNA polymerase II. At one extreme are enhancers,which can stimulate transcription from a promoter tens ofthousands of base pairs away (e.g., the SV40 enhancer). Atthe other extreme are promoter-proximal elements, such asthe upstream elements controlling the HSV tkgene, whichlose their influence when moved an additional 30–50 basepairs farther from the promoter. Thestart site at which transcription initiates encodes the first (5')nucleotide of the first exon of an mRNA, the nucleotide thatis capped. For many genes, especially those encoding abundantly expressed proteins, a TATA box located approximately 25–35 base pairs upstream from the start site directsRNA polymerase II to begin transcription at the proper nucleotide. Promoter-proximal elements, which are relativelyshort (≈10–20 base pairs), are located within the first ≈200base pairs upstream of the start site. Enhancers, in contrast,usually are ≈100 base pairs long and are composed of multiple elements of ≈10–20 base pairs. Enhancers may be located up to 50 kilobases upstream or downstream from thestart site or within an intron. Many mammalian genes are

controlled by more than one enhancer region.The S. cerevisiaegenome contains regulatory elementscalled upstream activating sequences (UASs),which functionsimilarly to enhancers and promoter-proximal elements inhigher eukaryotes. Most yeast genes contain only one UAS,which

generally lies within a few hundred base pairs of thestart site. In addition, S. cerevisiaegenes contain a TATA box≈90 base pairs upstream from the transcription start site

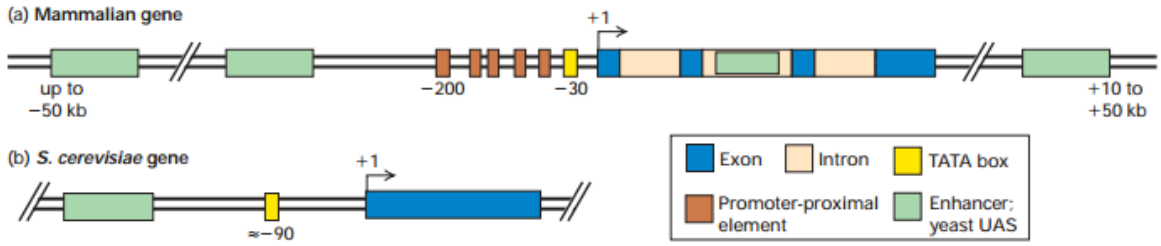


Figure: Multiple transcription control elements

Regulation of transcription factor activity by lipid-soluble hormones:

The mechanism whereby hormone binding controls the activity of nuclear receptors differs for heterodimeric and homodimeric receptors. Heterodimeric nuclear receptors (e.g.,RXR-VDR, RXR-TR, and RXR-RAR) are located exclusively in the nucleus. In the absence of their hormone ligand, they repress transcription when bound to theircognate sites in DNA. They do so by directing histonedeacetylation at nearby nucleosomes by the mechanism de-scribed earlier. As we saw earlier, in thepresence of hormone, the ligand-binding domain of theRAR monomer undergoes a dramatic conformationalchange compared with the ligand-binding domain of a nuclear receptor in the absence of hormone. In the ligand-bound conformation, heterodimericnuclear receptors containing RXR can direct hyperacetylation of histones in nearby nucleosomes, thereby reversingthe repressing effects of the free ligand-binding domain. Inthe presence of ligand, ligand-binding domains of nuclearreceptors also bind mediator, stimulating preinitiationcomplex assembly.

In contrast to heterodimeric nuclear receptors, homodimeric receptors are found in the cytoplasm in the absence of their ligands. Hormone binding to these receptorsleads to their translocation to the nucleus. The hormone-dependent translocation of the homodimeric glucocorticoid receptor (GR) was demonstrated in the transfectionexperiments shown in Figure 11-43. The GR hormone-binding domain alone mediates this transport. Subsequentstudies showed that, in the absence of hormone, GR is anchored in the cytoplasm as a large protein aggregate complexed with inhibitor proteins, including Hsp90, a proteinrelated to Hsp70, the major heat-shock chaperone in eukaryotic cells. As long as the receptor is confined to the cytoplasm, it cannot interact with target genes and hencecannot activate transcription. Hormone binding to a ho-

modimeric nuclear receptor releases the inhibitor proteins,allowing the receptor to enter the nucleus, where it canbind to response elements associated with target genes.

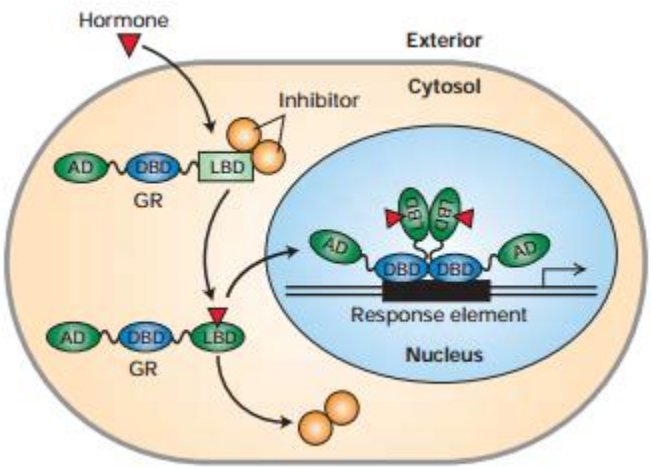


FIGURE: Model of hormone-dependent gene activation by a homodimeric nuclear receptor - In the absence of hormone, the receptor is kept in the cytoplasm by interaction between its ligand-binding domain (LBD) and inhibitor proteins. When hormone is present, it diffuses through the plasma membrane and binds to the ligand-binding domain, causing a conformational change that releases the receptor from the inhibitor proteins. The receptor with bound ligand is then translocated into the nucleus, where its DNA-binding domain (DBD) binds to response elements, allowing the ligand-binding domain and an additional activation domain (AD) at the N-terminus to stimulate transcription of target genes.

Once the receptor with bound hormonebinds to a response element, it activates transcription byinteracting with chromatin-remodeling and histone acetylase complexes and mediator.

POSSIBLE QUESTIONS

UNIT III

Part-B

1. Explain the structure of RNA polymerase and its role in transcription
2. Outline the structure and salient features of tRNA and mRNA
3. Explain the rho dependent termination of transcription

4. Explain the role of sigma factor in transcription
5. Write a short note on organization of transcription signal in eukaryotes.
6. How is mRNA processed in prokaryotes?
7. Give an account on transcription regulatory signal in eukaryotes.
8. Discuss the post transcriptional processing of mRNA
9. Differentiate promoter and enhancer.
10. Short notes on open promoter complex.

Part-C

1. Describe the events in transcription of prokaryotes
2. Discuss the influence of hormones on the eukaryotic transcription
3. Explain the processing of RNA in detail
4. Writ short notes on
    a. RNA binding motifs
    b. RNA editing
    c. Transcription factor
5. Explain the RNA polymerase and transcription factors in eukaryotes

Karpagam acadmy of higher educaton

Department of biochemistry

Molecular biology (19BCP202) MCQ

UNIT III

S.NO	UNIT	QUESTION	OPTION 1	OPTION 2	OPTION 3	OPTION 4	ANSWER
1	III	For initiation ___subunit of RNA pol required	Alpha	beta	gamma	**sigma**	**sigma**
2	III	The RNA polymerase of E. coli complex holoenzyme composed of 5 polypeptide subunits	2α,2β and one sigma factor	2α 2β and one rho factor	**2α 1β,1β and one sigma factor**	2α,1β,1β and one rho factor	**2α 1β,1β and one sigma factor**
3	III	Pribnow box consists of the following 6 nucleotide bases	**TATAAT**	TATATA	TTAAAT	TAATAT	**TATAAT**
4	III	The process of making copy of RNA from DNA is	Replication	**Transcription**	Translation	Reverse transcription	**Transcription**
5	III	RNA polymerase utilizes the following RNA triphosphates for the formation of RNA	ATP, GTP, CTP and TTP	**ATP, GTP, CTP and UTP**	ATP, CTP, UTP and UTP	ATP, CTP, GTP and TTP	**ATP, GTP, CTP and UTP**
6	III	Transfer RNA perform	Amino acid sequence by gene	**Read information in mRNA**	Synthesize proteins	Read information in tRNA	**Read information in mRNA**
7	III	RNA self splicing was discovered by	Watson & Crick	McClintock	Sanger	**Thomas Cech**	**Thomas Cech**
8	III	Termination of transcription is brought about by	sigma factor	RNA polymerase	**rho factor**	TATA box	**rho factor**
9	III	In eukaryotic cells, the synthesis of mRNA is carried out by	RNA polymerase I	**RNA polymerase II**	RNA polymerase III	DNA polymerase	**RNA polymerase II**
10	III	Ribosomal RNA perform____ function	Aminoacid sequence by gene	Read information in mRNA	**Synthesize proteins**	Read information in tRNA	**Synthesize proteins**
11	III	The enzyme responsible for reverse transcription	DNA dependent DNA polymerase	DNA dependent RNA polymerase	**RNA dependent DNA polymerase**	RNA dependent RNA polymerase	**RNA dependent DNA polymerase**
12	III	The mature 5'end of tRNA in E.coli formed as a result of endonucleolytic cleavage by	**RNase P**	RNAse D	RNase M5	RNase M13	**RNase P**
13	III	All the three RNA polymerases (I, II, III) are present only in	Prokaryotes	**Eukaryotes**	Prokaryotes and Eukaryotes	Virus	**Eukaryotes**
14	III	Rho factor is a	Dimer	Trimer	Octomer	**Tetramer**	**Tetramer**
16	III	The tRNA TΨC arm contains	Thiamin , pseudourindine and cysteine	Thiamin, uracil and cysteine	**Thiamin, pseudourindine and cytosine**	Adenine, uracil and cysteine	**Thiamin, pseudourindine and cytosine**
17	III	Transcription initiation in bacteria is inhibited by	Streptolydigin	Puromycin	**Rifampicin**	Tetracycline	**Rifampicin**
18	III	The hexa nucleotide AAU AAA provides the signal for	Capping	**Generation of eukaryotic 3'-poly A tail**	Initiation of translation	Peptide bond formation	**Generation of eukaryotic 3'- poly A tail**
19	III	Select the correct statement about spliceosomes	convert pre mRNA into mature RNA	composed of only RNA	composed of only Protein	**splice RNA in prokaryotic and eukaryotic cells**	**splice RNA in prokaryotic and eukaryotic cells**
20	III	Rho protein is involved in	transcription initiation	transcription elongation	**transcription termination at some terminators**	transcription termination at all terminators	**transcription termination at some terminators**
21	III	tRNA specifically reacts with	mRNA	nuclear RNA	**amino acyl adenylates**	ribosomal RNA	**amino acyl adenylates**
22	III	Pseudo uridine is found in	mRNA	snRNA	rRNA	**tRNA**	**tRNA**
23	III	The acceptor arm in tRNA consists of base pair stem terminates in the sequence	**CCA (5'-3')**	CAA (5'-3')	CCA (3'-5')	AAC (3'-5')	**CCA (5'-3')**
24	III	sn RNA is involved in	DNA replication	Ribosome assembly	**RNA splicing**	Initiation of translation	**RNA splicing**
25	III	Pre –r RNA are synthesized by RNA polymerase	I	II	**III**	IV	**III**
26	III	CPG islands are	**Promotors**	Operators	Silencers	Sensors	**Promotors**
27	III	rDNA is	**unique gene**	tandem gene cluster	a pseudogene	Junk DNA	**unique gene**
28	III	Rho factor is	an enzyme	**accessory protein**	rich in U residues	rich in G-C with a palindromic sequence	**accessory protein**
29	III	Cytosolic RNAs are synthysized by RNA polymerase	I	**II**	III	IV	**II**
30	III	The protein that bind to the TATA box in the promoter region are called	coregulators	coactivators	enhancers	**Transcriptional factors**	**Transcriptional factors**
31	III	Premature chain termination in prokaryotes and eukaryotes is caused by	**ricin**	abrin	erythromycin	puromycin	**ricin**
32	III	Rho dependant termination sites are rich in	only U	Only GC	**Both U and GC**	AT	**Both U and GC**
33	III	Prokaryotic transcription initiation is inhibited by	Rifamycin	**Rifampicin**	Tetracycline	Puromycin	**Rifampicin**
34	III	Who demonstrated that poly U stimulate the synthesis of poly phe	Nirenberg and Leder	Nirenberg and Ochoa	**Nirenberg and Mathaci**	H. Gobind Khorana	**Nirenberg and Mathaci**
35	III	Processing of tRNA involves adding	**CCA to the 3' end**	CCA to the 5' end	7-methyl G to the 3' end	Poly A to the 3'end	**CCA to the 3' end**
36	III	The enzyme responsible for processing the rRNAs is	**Rnase III**	Rnase H	Rnase P	Rnase F	**Rnase III**
37	III	Rho factor catalyses the unwinding of RNA-DNA double helices and	**Hydrolysis of ATP**	RNA-RNA double helices	Migrates along the RNA in 3'-5' direction	Migrates along the RNA in 5'-3' direction	**Hydrolysis of ATP**
38	III	Mature rRNAs are produced using Rnases	M16, M23 and M5	**Rnase III**	Rnase F	RNAse P	**Rnase III**
39	III	The strand that directs the synthesis of mRNA is	sense strand	coding strand	**non coding strand**	missense strand	**non coding strand**
40	III	Snurps are involved in	Poly A tail addition	**Splicing**	5' cap formation	stem loop formation	**Splicing**
41	III	Non template strand is otherwise called as	template strand	**coding strand**	non coding strand	template and non coding strand	**coding strand**
42	III	___activates transcription	**CRP**	GMP	RNase	Ribozyme	**CRP**
43	III	Promoters are recognized by	alpha subunit	gamma subunit	beta subunit	**sigma subunit**	**sigma subunit**
44	III	Enhancers are entry point on DNA for RNA polymerase	I	**II**	III	I and III	**II**
45	III	When a number of genes are transcribed as one mRNA, the mRNA is said to be	multimeric	polymeric	**polycistronic**	polyclonal	**polycistronic**
46	III	The region on tRNA recognizes the appropriate base sequence mRNA is called	codon loop	**anti codon loop**	P site	A site	**anti codon loop**
47	III	The 5' terminus of a eukaryotic mRNA molecule is	2-amino purine	5 bromo uracil	**7 methyl guanosine**	formyl methionine	**7 methyl guanosine**
48	III	Which of the following are steps in RNA synthesis?	**binding of RNA polymerase to DNA**	binding of σ factor to DNA	polymerase	binding of RNA polymerase to DNA	**binding of RNA polymerase to DNA**

49	III	The role of tRNA is	to attach the amino acids to one another	**to bring the amino acids to the correct position with respect to one another**	to increase the effective concentration of amino acids	to attach m RNA to the ribosome	**to bring the amino acids to the correct position with respect to one another**
50	III	The following are associated with transcription except	-35 sequence	pribnow box	promoter	**spacer**	**spacer**
51	III	What is the characteristic form introns have after being cut from a pre-mRNA?	Linear structure	circular form	**lariat shaped**	theta structure	**lariat shaped**
52	III	A tRNA molecule must be able to	**recognize a codon**	recognize an anticodon	distinguish one amino acid from another	recognize DNA molecules	**recognize a codon**
53	III	RNA polymerase in prokaryotes has a removable	alpha subunit	beta subunit	both (a) and (b)	sigma subunit	**sigma subunit**
54	III	The complex of RNA polymerase, DNA template and new RNA transcript is called	**transcription bubble**	replication bubble	a translation bubble	none of these	**transcription bubble**
55	III	Promoters for tRNAs are located	upstream from the start codon	**downstream from the start codon**	both (a) and (b)	none of these	**downstream from the start codon**
56	III	Rho-dependent termination of transcription in E. coli	requires ATP	requires about 50 nucleotides of uncomplexed mRNA	**both (a) and (b)**	removes mRNA and holoenzyme from the DNA	**both (a) and (b)**
57	III	**Enhancers are regions that**	bind RNA polymerase	are adjacent to the TATA box	**are CAT box binding proteins**	modulate transcription	**are CAT box binding proteins**
58	III	Transcription by E. coli polymerase occurs in	four phases known as initiation, propagation, elongation and termination	**three phases known as initiation, elongation and termination**	two phases known as initiation and termination	none of the above	**three phases known as initiation, elongation and termination**
59	III	Multiple copies of 5S genes, located at a chromosomal site distinct from the other rRNA genes	**are transcribed by rRNA polymerase III and the 5S rRNA**	are transcribed by rRNA polymerase II	are transcribed by rRNA polymerase I and the 5S rRNA	are transcribed by DNA polymerase I	**are transcribed by rRNA polymerase III and the 5S rRNA**
60	III	Each tRNA molecule has a cloverleaf secondary structure consisting of	**three stem loops**	two stem loops	four stem loops	either (a) or (b)	**three stem loops**

UNIT-IV

SYLLABUS

Translation: Deciphering genetic code, features. Wobble hypothesis. Initiation, elongation and termination of prokaryotic and eukaryotic translation. Fidelity of translation. Post translational modifications-all types; Protein targeting-Targeting protein to nucleus, ER, Golgi complex. Protein degradation- ubiquitin mediated degradation.

Introduction

Genetic code

The genetic code is the set of rules by which information encoded in genetic material (DNA or mRNA sequences) is translated into proteins (amino acid sequences) by living cells. The code defines a mapping between tri-nucleotide sequences, called codons, and amino acids. With some exceptions, a triplet codon in a nucleic acid sequence specifies a single amino acid. The first elucidation of a codon was done by Marshall Nirenberg and Heinrich J. Matthaei in 1961 at the National Institutes of Health. They used a cell-free system to translate a poly-uracil RNA sequence (i.e., UUUUU...) and discovered that the polypeptide that they had synthesized consisted of only the amino acid phenylalanine. They thereby deduced that the codon UUU specified the amino acid phenylalanine. This was followed by experiments in the laboratory of Severo Ochoa demonstrating that the poly-adenine RNA sequence (AAAAA...) coded for the polypeptide, poly-lysine. The poly-cytosine RNA sequence (CCCCC...) coded for the polypeptide, poly-proline. Therefore the codon AAA specified the amino acid lysine, and the codon CCC specified the amino acid proline. Using different copolymers most of the remaining codons were then determined. Extending this work, Nirenberg and Philip Leder revealed the triplet nature of the genetic code and allowed the codons of the standard genetic code to be deciphered. In these experiments various combinations of mRNA were passed through a filter which contained ribosomes, the components of cells that translate RNA into protein. Unique triplets promoted the binding of specific tRNAs to the ribosome. Leder and Nirenberg were able to determine the sequences of 54 out of 64 codons in their experiments.

Subsequent work by Har Gobind Khorana identified the rest of the genetic code. Shortly thereafter, Robert W. Holley determined the structure of transfer RNA (tRNA), the adapter molecule that facilitates the process of translating RNA into protein.

Wobble Hypothesis

The genetic code is a degenerate code, meaning that a given amino acid may have more than one codon.

In the genetic code, there are 43 = 64 possible codons (tri-nucleotide sequences). For translation, each of these codons requires a tRNA molecule with a complementary anticodon. If each tRNA molecule paired with its complementary mRNA codon using canonical Watson-Crick base pairing, then 64 types (species) of tRNA molecule would be required. In the standard genetic code, three of these 64 codons are stop codons, which terminate translation by binding to release factors rather than tRNA molecules, so canonical pairing would require 61 species of tRNA. Since most organisms have fewer than 45 species of tRNA, some tRNA species must pair with more than one codon. In 1966, Francis Crick proposed the Wobble hypothesis to account for this. He postulated that the 5' base on the anticodon, which binds to the 3' base on the mRNA, was not as spatially confined as the other two bases, and could, thus, have non-standard base pairing.

As an example, yeast tRNAPhe has the anticodon 5'-GmAA-3' and can recognize the codons 5'-UUC-3' and 5'-UUU-3'. It is, therefore, possible for non-Watson–Crick base pairing to occur at the third codon position, i.e., the 3' nucleotide of the mRNA codon and the 5' nucleotide of the tRNA anticodon.

tRNA Base pairing schemes

The original wobble pairing rules, as proposed by Crick. Watson-Crick base pairs are shown in **bold**, wobble base pairs in *italic*:

tRNA 5' anticodon base	**mRNA 3' codon base**
A	**U**
C	**G**
G	**C** or *U*

U	**A** or *G*
I	*A* or *C* or *U*

Revised pairing rules

tRNA 5' anticodon base	**mRNA 3' codon base**
G	U,C
C	G
k^2C	A
A	U,C,(A),G
unmodified U	U,(C),A,G
xm^5s^2U,xm^5Um,Um,xm^5U	A,(G)
xo^5U	U,A,G
I	A,C,U

First letter of codon (5′ end)

Second letter of codon

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

The genetic code has redundancy but no ambiguity (see the codon tables above for the full correlation). For example, although codons GAA and GAG both specify glutamic acid (redundancy), neither of them specifies any other amino acid (no ambiguity). The codons encoding one amino acid may differ in any of their three positions. For example the amino acid glutamic acid is specified by GAA and GAG codons (difference in the third position), the amino acid leucine is specified by UUA, UUG, CUU, CUC, CUA, CUG codons (difference in the first or third position), while the amino acid serine is specified by UCA, UCG, UCC, UCU, AGU, AGC (difference in the first, second or third position). Degeneracy results because there are more codons than encodable amino acids. For example, if there were two bases per codon, then only 16 amino acids could be coded for (4^2=16). Because at least 21 codes are required (20 amino acids plus stop), and the next largest number of bases is three, then 4^3 gives 64 possible codons, meaning that some degeneracy must exist.

There are 4^3 = 64 different codon combinations possible with a triplet codon of three nucleotides; all 64 codons are assigned for either amino acids or stop signals during translation. If, for example, an RNA sequence, UUUAAACCC is considered and the reading frame starts with the first U (by convention, 5' to 3'), there are three codons, namely, UUU, AAA and CCC, each of which specifies one amino acid. This RNA sequence will be translated into an amino acid sequence, three amino acids long.

Translation starts with a chain initiation codon (start codon). Unlike stop codons, the codon alone is not sufficient to begin the process. Nearby sequences (such as the Shine Dalgarno sequence in E. coli) and initiation factors are also required to start translation. The most common start codon is AUG which is read as methionine or, in bacteria, as formylmethionine. Alternative start codons (depending on the organism), include "GUG" or "UUG", which normally code for valine or leucine, respectively. However, when used as a start codon, these alternative start codons are translated as methionine or formylmethionine.

The three stop codons have been given names: UAG is amber, UGA is opal (sometimes also called umber), and UAA is ochre. "Amber" was named by discoverers Richard Epstein and Charles Steinberg after their friend Harris Bernstein, whose last name means "amber" in German. The other two stop codons were named "ochre" and "opal" in order to keep the "color names"

theme. Stop codons are also called "termination" or "nonsense" codons and they signal release of the nascent polypeptide from the ribosome due to binding of release factors in the absence of cognate tRNAs with anticodons complementary to these stop signals.

The codon AUG both codes for methionine and serves as an initiation site: the first AUG in an mRNA's coding region is where translation into protein begins.

Prokaryotic translation

Translation of the genetic code involves using the genetic information to produce a polypeptide. Therefore, **translation** is synonymous with **protein synthesis**. Proteins, or to be more precise, polypeptides, are linear chains of amino acids. (You don't need to know all of the amino acids, but you should work to become more familiar with them.) As outlined in the module on the genetic code, genetic information encoded in DNA, and transferred to mRNA, is used to determine the sequence of amino acids in a polypeptide.

The following "ingredients" are required for initiation of protein synthesis:

- Aminoacylated tRNAs
- Ribosomes
- An mRNA
- Ancillary Protein "Factors"
- A special tRNA for Initiation

The actual synthesis of polypeptides is carried out by **ribosomes**.

Transfer RNA

The other player in the translation process is transfer RNA, or tRNA. tRNA serves as an adaptor or intermediary between mRNA and amino acids. tRNAs are among the best characterized RNA molecules - they are quite short (75 to 90 nucleotides long) and have nearly identical sequences in ekaryotes and prokaryotes. tRNA molecules are somewhat unique in that they contain several unusual nucleotides, such as inosine, pseudouridine, and hypoxanthine.

The sequence of each individual tRNA molecule is such that base pairing occurs between strands in different regions of the same molecule. This gives tRNA molecules a characteristic 'cloverleaf' shape. There are two main functional regions of the tRNA molecule. The middle loop of the cloverleaf contains three unpaired bases known as the **anticodon**. The anticodon base pairs with the complementary codon on mRNA during translation. Directly opposite of the anticodon is a region with no loop - it contains both ends of the linear tRNA molecule. This region, particularly the **3' end** of the tRNA is where a **specific amino acid** will bind in preparation for protein synthesis. A tRNA molecule with a particular anticodon sequence will only bind to one amino acid (for example, the tRNA with AGU as an anticodon sequence will only bind to the amino acid serine). In this way, specificity of the genetic code is maintained.

tRNA molecules are joined to their specific amino acid in a reaction known as **charging**. The 3' end of the tRNA molecule is covalently linked to the correct amino acid by an enzyme called **aminoacyl tRNA synthetase**. This enzyme recognizes the appropriate tRNA and enzyme, and uses the energy of ATP to join the two. Because the recognition of the tRNA and amino acid by the enzyme is so specific, there must be a different aminoacyl tRNA synthetase for each amino acid. Therefore, there are at least 20 different aminoacyl tRNA synthetases.

1.**Aminoacylation of Transfer RNA**

The addition of the correct amino acid to the correct tRNA is possibly the most critical step in gene expression. If this is not done correctly, then the Genetic Code would effectively be meaningless.

Once the amino acid has been added to the tRNA, it will be used for protein synthesis according to the specificity dictated by the anticodon sequence in the tRNA. This was demonstrated by the classic **Raney Nickel Experiment** carried out by **Seymour Benzer**'s group in 1962.

The aminoacylation of tRNAs is carried out by a family of enzymes called **aminoacyl tRNA synthetases** (**aaRS**). There are 21 enzymes -- one for each amino acid except lysine which has two.

There are a number of important aspects of the mechanism of action of these enzymes to consider:

- The mechanism of the addition reaction itself (aminoacylation)
- The mechanism by which the amino acyl tRNA synthetase recognizes the correct tRNA.
- The mechanism by which the amino acyl tRNA synthetase recognizes the correct amino acid.

Mechanism of Aminoacylation

Aminoacylation of the tRNA by **Aminoacyl tRNA synthetases** proceeds through a 2 step reaction mechanism that can be summarized as:

Step 1: Activation of the Amino Acid

aa + ATP <=> aa~AMP + PP_i

Step 2: Transfer of the aminoacyl group to the tRNA

aa~AMP + tRNA <=> aa-tRNA + AMP

Both steps occur in the active site of the enzyme; there is no dissociation of the aminoacyl-adenylylate intermediate from the active site during the reaction.

Mechanism of tRNA Recognition

The mechanisms by which **aaRS**s recognize their cognate tRNAs has been called a second genetic code. This is overly simplistic since the rules are clearly very complex and appear to be different for different tRNA-**aaRS** combinations.Recognition occurs through interactions with the anticodon, in which the bases of the anticodon loop are probably unstacked so that their interactions with the enzyme can be maximized.

Mechanism of Amino Acid Recognition

Recognition of the correct amino acid occurs in a manner analogous to that by which all enzymes recognize their substrates. Each amino acid will fit into an active site pocket in the **aaRS** where it will bind through a network of hydrogen bonds, electrostatic and hydrophobic interactions. Only amino acids with a sufficient number of favourable interactions will bind.

However, since some amino acids have very similar side chains, a proof-reading mechanism exists in many cases to make sure that the correct amino acid is chosen

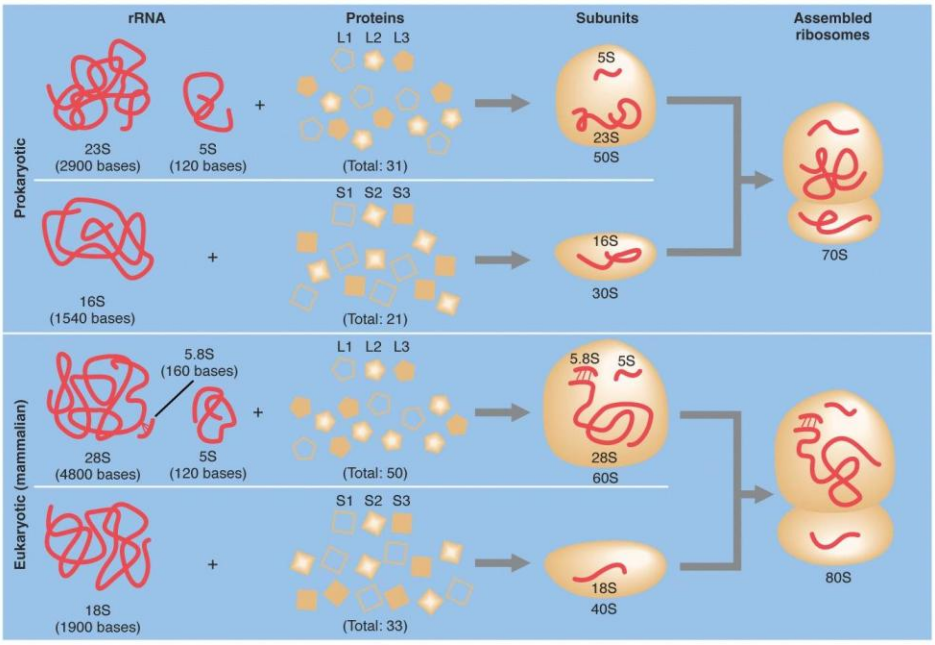
2.Ribosomes

The ribosome component synthesizes protein chains. It assembles the twenty specific amino acid molecules to form the particular protein molecule determined by the nucleotide sequence of an RNA molecule. This process is known as translation; the ribosome translates the genetic information from the mRNA into proteins. Ribosomes do this by binding to an mRNA and using it as a template for determining the correct sequence of amino acids in a particular protein. The amino acids are attached to transfer RNA (tRNA) molecules, which enter one part of the ribosome and bind to the messenger RNA sequence. The attached amino acids are then joined together by another part of the ribosome. The ribosome moves along the mRNA, "reading" its sequence and producing a corresponding chain of amino acids.

The ribosomes are complex ribonucleoprotein complexes. They consist of two ribonucleoprotein subunits: a smaller subunit and a larger subunit. he smaller subunit binds to the mRNA, while the larger subunit binds to the tRNA and the amino acids. When a ribosome finishes reading a mRNA, these two subunits split apart. The sizes and exact composition of each subunit is basically the same in all organisms though the exact details are, of course, different.

In Eukaryotes, the co-efficient of ribosomes are 80s, of which is divided into 60s for the large, and 40s for the small subunit. The 60s contain 28s rRNA,; a 5.8s, and 5sRNA. Whereas, the 40s subunit has only a single 18s rRNA along with proteins.

In prokaryotes, however, the large and small subunits are split into 50s and 30s, making a total of 70s respectively. The 50s has two types of rRNA - a 23s and a 5s Label. It also has 32 different proteins. On the other hand, the 30s contains a single 16s rRNA plus, 21 different types of proteins.



3. The messenger RNA

The mRNA must contain some feature that allows its 5' end to be recognized by and positioned correctly on the ribosome during the assembly of the protein synthesizing apparatus. In bacteria, a special **ribosome binding site** was identified by **John Shine** and **Lynn Dalgarno** in 1974. They observed that the 3' end of the **16S** rRNA is complementary to a short region just upstream of the start codon in bacterial mRNA.

The **ribosome binding site** is frequently called the **Shine-Dalgarno** sequence.

In eukaryotes, it is the 5' cap structure that is added to the mRNA which is required for correct positioning of the ribosome on the mRNA during the initiation phase of protein synthesis.

4. Ancillary Protein "Factors"

Each of the steps of protein synthesis requires the participation of a number of additional special protein factors. Each factor has a specific role to play which will be described Some of the factors are G-proteins -- they bind GTP and GTP hydrolysis is an important part of their function.

5. **A special tRNA for Initiation**

In bacteria, protein synthesis starts with a special amino acid: **N-formyl-methionine**. Addition of the formyl group to the N-terminal methionine effectively provides it with a peptide bond.

This amino acid is synthesized by modifying methionine after it has been attached to a special tRNA -- $tRNA_f^{Met}$. The same **MetRS** aminoacylates both $tRNA^{Met}$ and $tRNA_f^{Met}$.

$tRNA_f^{Met}$ is structurally different from the "regular" $tRNA^{Met}$ in a number of ways:

- It contains 3 consecutive GC base pairs in the anticodon stem.
- The terminal bases of the acceptor arm are not paired as they are in all other tRNAs.

$tRNA_f^{Met}$ is functionally different from the "regular" $tRNA^{Met}$ in two important ways:

- It is recognized by a special enzyme that will catalyse the formylation of methionyl-$tRNA_f^{Met}$ -- **transformylase**.
- **Transformylase** catalyses the formylation of the methionyl-$tRNA_f^{Met}$ but not of methionyl-$tRNA^{Met}$ or uncharged tRNA. The enzyme uses N^{10}-formyltetrahydrofolate as the formyl group donor.
- It is recognized by a special **Initiation Factor** -- **IF2** -- which recognizes fmet-$tRNA_f^{Met}$ but not Met-$tRNA^{Met}$ and brings it to the ribosome during the initiation phase of protein synthesis.

The presence of the formyl group on the methionine after it has been attached to $tRNA_f^{Met}$ serves two purposes:

- It ensures that this will be the only charged tRNA that is be positioned in the peptidyl site on the ribosome to start protein synthesis.
- It ensures that this tRNA will not be used for internal methionine codons.

The process of translation can be divided into three basic steps: **initiation**, **elongation**, and **termination**. Each of these steps will be considered in turn.

Translation takes place in five stages.

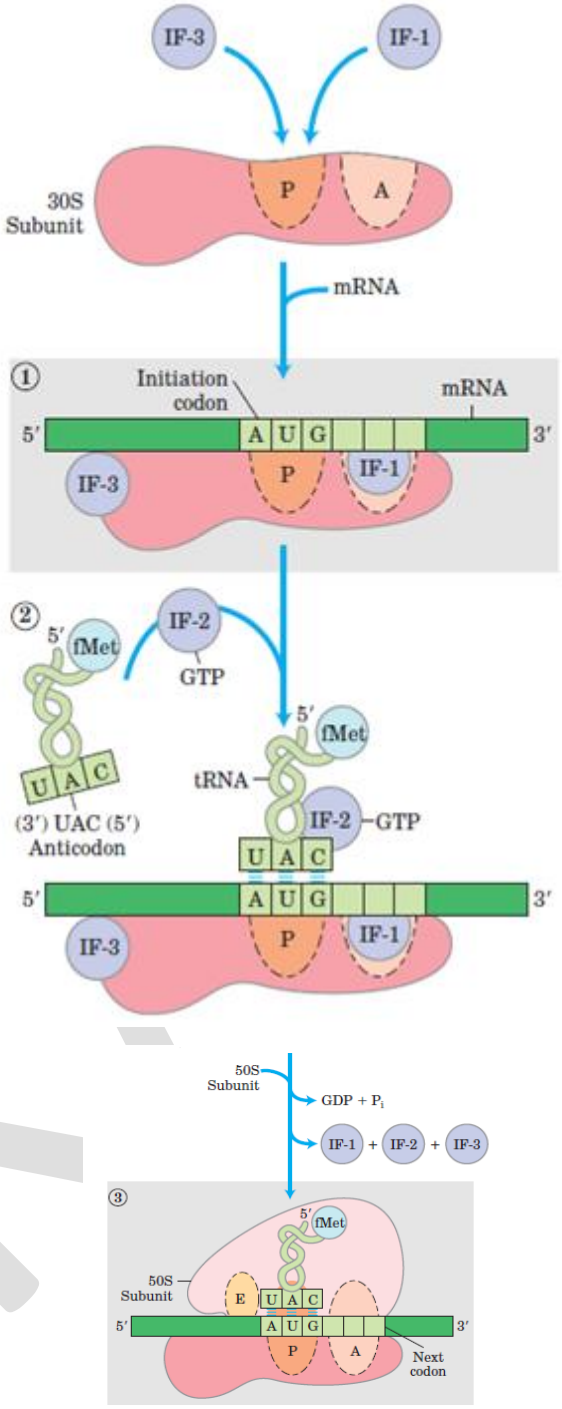
Stage 1: Activation of Amino Acids For the synthesis of a polypeptide with a defined sequence, two fundamental chemical requirements must be met: (1) the carboxyl group of each amino acid must be activated to facilitate formation of a peptide bond, and (2) a link must be established between each new amino acid and the information in the mRNA that encodes it. Both these requirements are met by attaching the amino acid to a tRNA in the first stage of protein synthesis. Attaching the right amino acid to the right tRNA is critical. This reaction takes place in the cytosol, not on the ribosome. Each of the 20 amino acids is covalently attached to a specific tRNA at the expense of ATP energy, using Mg^{2+} dependent activating enzymes known as aminoacyltRNA synthetases. When attached to their amino acid (aminoacylated) the tRNAs are said to be "charged."

$$\text{Amino acid} + \text{tRNA} + \text{ATP} \xrightarrow{Mg^{2+}} \text{aminoacyl-tRNA} + \text{AMP} + 2P_i$$

The identity of the amino acid attached to a tRNA is not checked on the ribosome, so attachment of the correct amino acid to the tRNA is essential to the fidelity of protein synthesis.

Stage 2: Initiation

Protein synthesis begins at the amino-terminal end and proceeds by the stepwise addition of amino acids to the carboxyl-terminal end of the growing polypeptide.The AUG initiation codon thus specifies an *amino-terminal* methionine residue. Although methionine has only one codon, (5')AUG, all organisms have two tRNAs for methionine. One is used exclusively when (5')AUG



is the initiation codon for protein synthesis. The other is used to code for a Met residue in an internal position in a polypeptide.

In step 1 the 30S ribosomal subunit binds two initiation factors, IF-1 and IF-3. Factor IF-3 prevents the 30S and 50S subunits from combining prematurely. The mRNA then binds to the 30S subunit. The initiating (5')AUG is guided to its correct position by the **Shine-Dalgarno sequence** in the mRNA. This consensus sequence is an initiation signal of four to nine purine residues, 8 to 13 bp to the 5' side of the initiation codon. The sequence base-pairs with a complementary pyrimidine-rich sequence near the 3' end of the 16S rRNA of the 30S ribosomal subunit. This mRNA-rRNA interaction positions the initiating (5')AUG sequence of the mRNA in the precise position on the 30S subunit where it is required for initiation of translation. The particular (5')AUG where fMet-tRNAfMet is to be bound is distinguished from other methionine codons by its proximity to the Shine-Dalgarno sequence in the mRNA.

Bacterial ribosomes have three sites that bind aminoacyl-tRNAs, the **aminoacyl (A) site,** the **peptidyl (P) site,** and the **exit (E) site.** Both the 30S and the 50S subunits contribute to the characteristics of the A and P sites, whereas the E site is largely confined to the 50S subunit. The initiating (5')AUG is positioned at the P site, the only site to which fMet- $tRNA^{fMet}$ can bind. The fMet-$tRNA^{fMet}$ is the only aminoacyl-tRNA that binds first to the P site; during the subsequent elongation stage, all other incoming aminoacyl-tRNAs (including the Met-$tRNA^{Met}$ that binds to interior AUG codons) bind first to the A site and only subsequently to the P and E sites. The E site is the site from which the "uncharged" tRNAs leave during elongation. Factor IF-1 binds at the A site and prevents tRNA binding at this site during initiation.

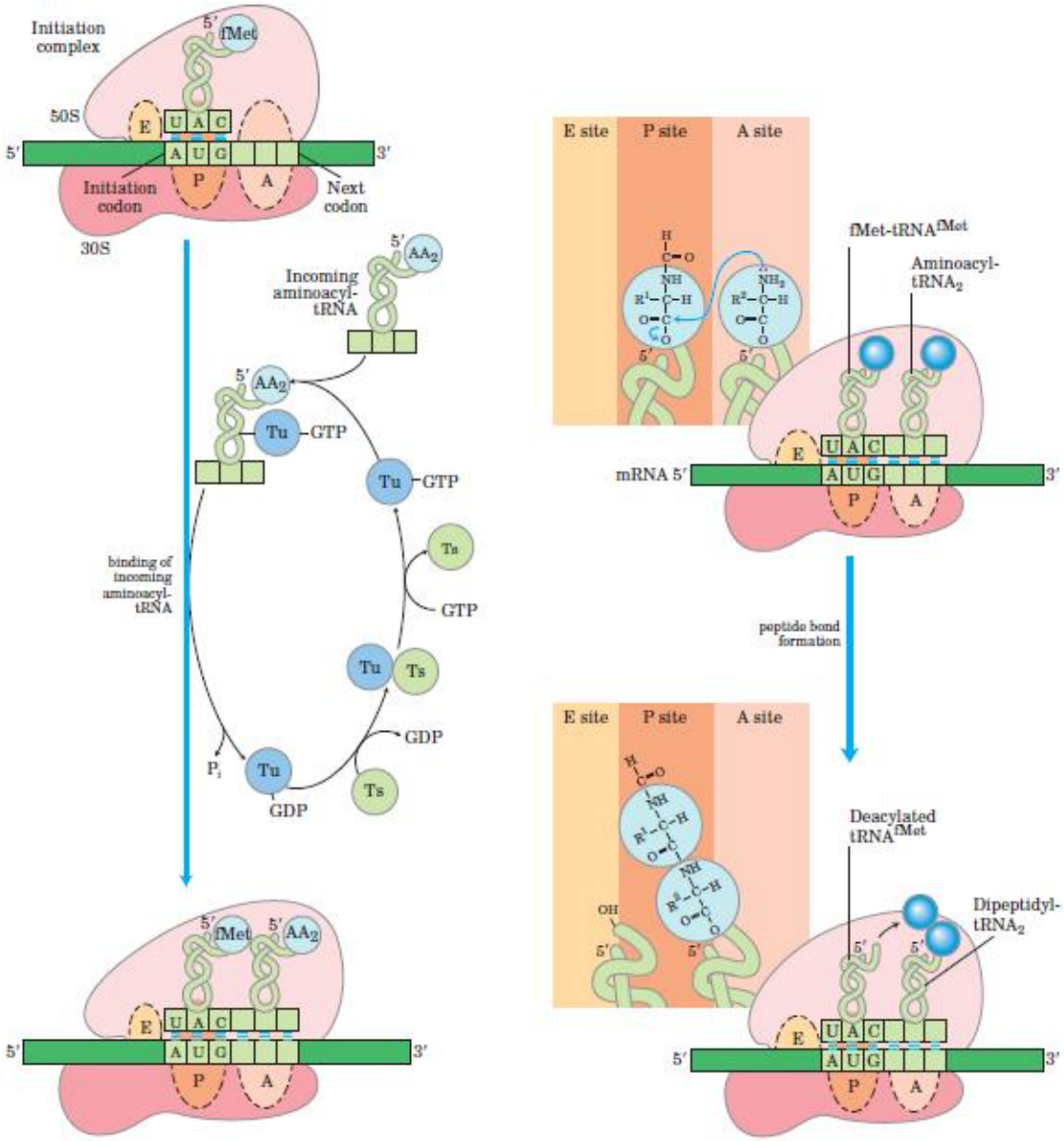
In step 2 of the initiation process, the complex consisting of the 30S ribosomal subunit, IF-3, and mRNA is joined by both GTP-bound IF-2 and the initiating fMet-$tRNA^{fMet}$. The anticodon of this tRNA now pairs correctly with the mRNA's initiation codon.

In step 3 this large complex combines with the 50S ribosomal subunit; simultaneously, the GTP bound to IF-2 is hydrolyzed to GDP and Pi, which are released from the complex. All three initiation factors depart from the ribosome at this point.

Completion of the steps produces a functional 70S ribosome called the **initiation complex,** containing the mRNA and the initiating fMettRNA^{fMet}. The correct binding of the fMet-$tRNA^{fMet}$ to the P site in the complete 70S initiation complex is assured by at least three points of recognition and attachment: the codon-anticodon interaction involving the initiation

AUG fixed in the P site; interaction between the Shine-Dalgarno sequence in the mRNA and the 16SrRNA; and binding interactions between the ribosomal P site and the fMet-tRNAfMet. The initiation complex is now ready for elongation.

Stage 3: Elongation The nascent polypeptide is lengthened by covalent attachment of successive amino acid units, each carried to the ribosome and correctly positioned by its tRNA, which base-pairs to its corresponding codon in the mRNA. Elongation requires cytosolic proteins known as elongation factors. The binding of each incoming aminoacyl-tRNA and the movement of the ribosome along the mRNA are facilitated by the hydrolysis of GTP as each residue is added to the growing polypeptide.



Elongation Step 1: In the first step of the elongation cycle, the appropriate incoming aminoacyl-tRNA binds to a complex of GTP-bound EF-Tu. The resulting aminoacyl tRNA–EF-Tu–GTP complex binds to the A site of the 70S initiation omplex. The GTP is hydrolyzed and an EF-Tu–GDP complex is released from the 70S ribosome. The EF-Tu–GTP complex is regenerated in a process involving EF-Ts and GTP.

Elongation Step 2: A peptide bond is now formed between the two amino acids bound by their tRNAs to the A and P sites on the ribosome. This occurs by the transfer of the initiating *N*-formylmethionyl group from its tRNA to the amino group of the second amino acid, now in the A site. The α amino group of the amino acid in the A site acts as a nucleophile, displacing the tRNA in the P site to form the peptide bond. This reaction produces a dipeptidyl tRNA in the A site, and the now "uncharged" (deacylated) $tRNA^{fMet}$ remains bound to the P site. The tRNAs then shift to a hybrid binding state, with elements of each spanning two different sites on the ribosome.

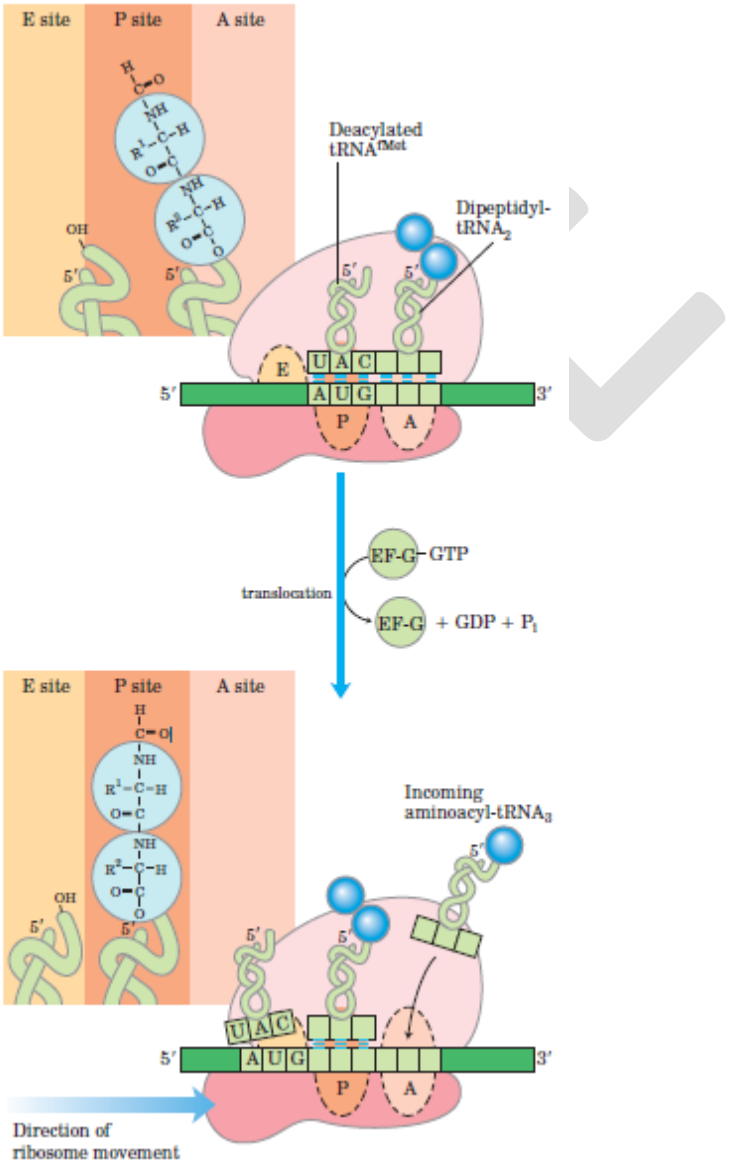
Elongation Step 3: In the final step of the elongation cycle, **translocation,** the ribosome moves one codon toward the 3' end of the mRNA. This movement shifts the anticodon of the dipeptidyl tRNA, which is still attached to the second codon of the mRNA, from the A site to the P site, and shifts the deacylated tRNA from the P site to the E site, from where the tRNA is released into the cytosol. The third codon of the mRNA now lies in the A site and the second codon in the P site. Movement of the ribosome along the mRNA requires EF-G (also known as translocase) and the energy provided by hydrolysis of another molecule of GTP.

The ribosome, with its attached dipeptidyl-tRNA and mRNA, is now ready for the next elongation cycle and attachment of a third amino acid residue. This process occurs in the same way as addition of the second residue. For each amino acid residue correctly added to the growing polypeptide, two GTPs are hydrolyzed to GDP and Pi as the ribosome moves from codon to codon along the mRNA toward the 3' end.

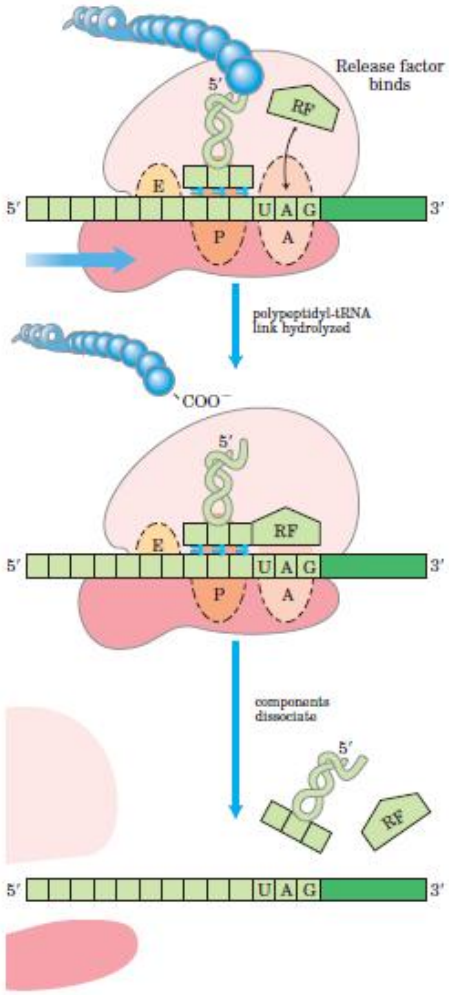
Stage 4: Termination and Release Completion of the polypeptide chain is signaled by a termination codon in the mRNA. The new polypeptide is released from the ribosome, aided by proteins called release factors.

Elongation continues until the ribosome adds the last amino acid coded by the mRNA. **Termination,** the fourth stage of polypeptide synthesis, is signaled by the presence of one of three termination codons in the mRNA (UAA, UAG, UGA), immediately following the final coded amino acid. In bacteria, once a termination codon occupies the ribosomal A site, three **termination factors,** or **release factors**—the proteins RF-1, RF-2, and RF-3— contribute to (1)

hydrolysis of the terminal peptidyltRNA bond; (2) release of the free polypeptide and the last tRNA, now uncharged, from the P site; and (3) dissociation of the 70S ribosome into its 30S and 50S subunits, ready to start a new cycle of polypeptide synthesis. RF-1 recognizes the termination codons UAG and UAA, and RF-2 recognizes UGA and UAA. Either RF-1 or RF-2



(depending on which codon is present) binds at a termination codon and induces peptidyl transferase to transfer the growing polypeptide to a water molecule rather than to another amino acid.



Fidelity in protein synthesis.

- On average, the energy derived from the hydrolysis of more than four NTPs to NDPs is required
- for the formation of each peptide bond of a polypeptide.
- This energy permits very high fidelity in the biological translation of the genetic message of mRNA into the amino acid sequence of proteins.
- The degree of fidelity in protein synthesis is sufficient to ensure that most proteins contain no mistakes and that the large amount of energy required to synthesize a protein

is rarely wasted. One defective protein molecule is usually unimportant when many correct copies of the same protein are present.

Eukaryotic Translation

As mentioned previously, eukaryotic translation is very similar overall to prokaryotic translation. There are a few notable differences, however. These include the following:

Eukaryotic mRNAs do not contain a Shine-Delgarno sequence. Instead, ribosomal subunits recognize and bind to the 5' cap of eukaryotic mRNAs. In other words, the 5' cap takes the place of the Shine-Delgarno sequence.

Eukaryotes do not use formyl methionine as the first amino acid in every polypeptide; ordinary methionine is used. Eukaryotes do have a specific initiator tRNA, however.

Eukaryotic translation involves many more protein factors than prokaryotic translation (For example, eukaryotic initiation involves at least 10 factors, instead of the 3 in prokaryotes.)

Post translational modification of proteins

Protein post-translational modification (PTM) increases the functional diversity of the proteome by the covalent addition of functional groups or proteins, proteolytic cleavage of regulatory subunits or degradation of entire proteins. These modifications include phosphorylation, glycosylation, ubiquitination, nitrosylation, methylation, acetylation, lipidation and proteolysis and influence almost all aspects of normal cell biology and pathogenesis. Therefore, identifying and understanding PTMs is critical in the study of cell biology and disease treatment and prevention.

Post-translational modification can occur at any step in the "life cycle" of a protein. For example, many proteins are modified shortly after translation is completed to mediate proper protein folding or stability or to direct the nascent protein to distinct cellular compartments (e.g., nucleus, membrane). Other modifications occur after folding and localization are completed to activate or inactivate catalytic activity or to otherwise influence the biological activity of the protein. Proteins are also covalently linked to tags that target a protein for degradation. Besides single modifications, proteins are often modified through a combination of post-translational

cleavage and the addition of functional groups through a step-wise mechanism of protein maturation or activation.

Protein PTMs can also be reversible depending on the nature of the modification. For example, kinases phosphorylate proteins at specific amino acid side chains, which is a common method of catalytic activation or inactivation. Conversely, phosphatases hydrolyze the phosphate group to remove it from the protein and reverse the biological activity. Proteolytic cleavage of peptide bonds is a thermodynamically favorable reaction and therefore permanently removes peptide sequences or regulatory domains.

Consequently, the analysis of proteins and their post-translational modifications is particularly important for the study of heart disease, cancer, neurodegenerative diseases and diabetes. The characterization of PTMs, although challenging, provides invaluable insight into the cellular functions underlying etiological processes. Technically, the main challenges in studying post-translationally modified proteins are the development of specific detection and purification methods. Fortunately, these technical obstacles are being overcome with a variety of new and refined proteomics technologies.

As noted above, the large number of different PTMs precludes a thorough review of all possible protein modifications. Therefore, this overview only touches on a small number of the most common types of PTMs studied in protein research today. They are

Phosphorylation

Reversible protein phosphorylation, principally on serine, threonine or tyrosine residues, is one of the most important and well-studied post-translational modifications. Phosphorylation plays critical roles in the regulation of many cellular processes including cell cycle, growth, apoptosis and signal transduction pathways.

Glycosylation

Protein glycosylation is acknowledged as one of the major post-translational modifications, with significant effects on protein folding, conformation, distribution, stability and activity. Glycosylation encompasses a diverse selection of sugar-moiety additions to proteins that ranges from simple monosaccharide modifications of nuclear transcription factors to highly complex branched polysaccharide changes of cell surface receptors. Carbohydrates in the form of

aspargine-linked (N-linked) or serine/threonine-linked (O-linked) oligosaccharides are major structural components of many cell surface and secreted proteins.

Ubiquitination

Ubiquitin is an 8-kDa polypeptide consisting of 76 amino acids that is appended to the Îµ-NH2 of lysine in target proteins via the C-terminal glycine of ubiquitin. Following an initial monoubiquitination event, the formation of a ubiquitin polymer may occur, and polyubiquitinated proteins are then recognized by the 26S proteasome that catalyzes the degradation of the ubiquitinated protein and the recycling of ubiquitin.

S-Nitrosylation

Nitric oxide (NO) is produced by three isoforms of nitric oxide synthase (NOS) and is a chemical messenger that reacts with free cysteine residues to form S-nitrothiols (SNOs). S-nitrosylation is a critical PTM used by cells to stabilize proteins, regulate gene expression and provide NO donors, and the generation, localization, activation and catabolism of SNOs are tightly regulated. S-nitrosylation is a reversible reaction, and SNOs have a short half life in the cytoplasm because of the host of reducing enzymes, including glutathione (GSH) and thioredoxin, that denitrosylate proteins. Therefore, SNOs are often stored in membranes, vesicles, the interstitial space and lipophilic protein folds to protect them from denitrosylation. For example, caspases, which mediate apoptosis, are stored in the mitochondrial intermembrane space as SNOs. In response to extra- or intracellular cues, the caspases are released into the cytoplasm, and the highly reducing environment rapidly denitrosylates the proteins, resulting in caspase activation and the induction of apoptosis.

S-nitrosylation is not a random event, and only specific cysteine residues are S-nitrosylated. Because proteins may contain multiple cysteines and due to the labile nature of SNOs, S-nitrosylated cysteines can be difficult to detect and distinguish from non-S-nitrosylated amino acids. The biotin switch assay, developed by Jaffrey et al., is a common method of detecting SNOs, and the steps of the assay are listed below:

- All free cysteines are blocked.
- All remaining cysteines (presumably only those that are denitrosylated) are denitrosylated.

- The now-free thiol groups are then biotinylated.
- Biotinylated proteins are detected by SDS-PAGE and Western blot analysis or mass spectrometry.

Methylation

The transfer of one-carbon methyl groups to nitrogen or oxygen (N- and O-methylation, respectively) to amino acid side chains increases the hydrophobicity of the protein and can neutralize a negative amino acid charge when bound to carboxylic acids. Methylation is mediated by methyltransferases, and S-adenosyl methionine (SAM) is the primary methyl group donor.

Methylation occurs so often that SAM has been suggested to be the most-used substrate in enzymatic reactions after ATP. Additionally, while N-methylation is irreversible, O-methylation is potentially reversible. Methylation is a well-known mechanism of epigenetic regulation, as histone methylation and demethylation influences the availability of DNA for transcription. Amino acid residues can be conjugated to a single methyl group or multiple methyl groups to increase the effects of modification.

N-Acetylation

N-acetylation, or the transfer of an acetyl group to nitrogen, occurs in almost all eukaryotic proteins through both irreversible and reversible mechanisms. N-terminal acetylation requires the cleavage of the N-terminal methionine by methionine aminopeptidase (MAP) before replacing the amino acid with an acetyl group from acetyl-CoA by N-acetyltransferase (NAT) enzymes. This type of acetylation is co-translational, in that N-terminus is acetylated on growing polypeptide chains that are still attached to the ribosome. While 80-90% of eukaryotic proteins are acetylated in this manner, the exact biological significance is still unclear.

Acetylation at the Îµ-NH2 of lysine (termed lysine acetylation) on histone N-termini is a common method of regulating gene transcription. Histone acetylation is a reversible event that reduces chromosomal condensation to promote transcription, and the acetylation of these lysine residues is regulated by transcription factors that contain histone acetyletransferase (HAT) activity. While transcription factors with HAT activity act as transcription co-activators, histone

deacetylase (HDAC) enzymes are co- repressors that reverse the effects of acetylation by reducing the level of lysine acetylation and increasing chromosomal condensation.

Sirtuins (silent information regulator) are a group of NAD-dependent deacetylases that target histones. As their name implies, they maintain gene silencing by hypoacetylating histones and have been reported to aid in maintaining genomic stability.

While acetylation was first detected in histones, cytoplasmic proteins have been reported to also be acetylated, and therefore acetylation seems to play a greater role in cell biology than simply transcriptional regulation. Furthermore, crosstalk between acetylation and other post-translational modifications, including phosphorylation, ubiquitination and methylation, can modify the biological function of the acetylated protein.

Protein acetylation can be detected by chromosome immunoprecipitation (ChIP) using acetyllysine-specific antibodies or by mass spectrometry, where an increase in histone by 42 mass units represents a single acetylation.

Lipidation

Lipidation is a method to target proteins to membranes in organelles (endoplasmic reticulum [ER], Golgi apparatus, mitochondria), vesicles (endosomes, lysosomes) and the plasma membrane. The four types of lipidation are:

- C-terminal glycosyl phosphatidylinositol (GPI) anchor
- N-terminal myristoylation
- S-myristoylation
- S-prenylation

Each type of modification gives proteins distinct membrane affinities, although all types of lipidation increase the hydrophobicity of a protein and thus its affinity for membranes. The different types of lipidation are also not mutually exclusive, in that two or more lipids can be attached to a given protein.

GPI anchors tether cell surface proteins to the plasma membrane. These hydrophobic moieties are prepared in the ER, where they are then added to the nascent protein en bloc. GPI-anchored proteins are often localized to cholesterol- and sphingolipid-rich lipid rafts, which act as signaling platforms on the plasma membrane. This type of modification is reversible, as the

GPI anchor can be released from the protein by phosphoinositol-specific phospholipase C. Indeed, this lipase is used in the detection of GPI-anchored proteins to release GPI-anchored proteins from membranes for gel separation and analysis by mass spectrometry.

N-myristoylation is a method to give proteins a hydrophobic handle for membrane localization. The myristoyl group is a 14-carbon saturated fatty acid (C14), which gives the protein sufficient hydrophobicity and affinity for membranes, but not enough to permanently anchor the protein in the membrane. N-myristoylation can therefore act as a conformational localization switch, in which protein conformational changes influence the availability of the handle for membrane attachment. Because of this conditional localization, signal proteins that selectively localize to membrane, such as Src-family kinases, are N-myristoylated.

N-myristoylation is facilitated specifically by N-myristoyltransferase (NMT) and uses myristoyl-CoA as the substrate to attach the myristoyl group to the N-terminal glycine. Because methionine is the N-terminal amino acid of all eukaryotic proteins, this PTM requires methionine cleavage by the above-mentioned MAP prior to addition of the myristoyl group; this represents one example of multiple PTMs on a single protein.

S-palmitoylation adds a C16 palmitoyl group from palmitoyl-CoA to the thiolate side chain of cysteine residues via palmitoyl acyl transferases (PATs). Because of the longer hydrophobic group, this anchor can permanently anchor the protein to the membrane. This localization can be reversed, though, by thioesterases that break the link between the protein and the anchor; thus, S-palmitoylation is used as an on/off switch to regulate membrane localization. S-palmitoylation is often used to strengthen other types of lipidation, such as myristoylation or farnesylation. S-palmitoylated proteins also selectively concentrate at lipid rafts.

S-prenylation covalently adds a farnesyl (C15) or geranylgeranyl (C20) group to specific cysteine residues within 5 amino acids from the C-terminus via farnesyl transferase (FT) or geranylgeranyl transferases (GGT I and II). Unlike S-palmitoylation, S-prenylation is hydrolytically stable. Approximately 2% of all proteins are prenylated, including all members of the Ras superfamily. This group of molecular switches is farnesylated, geranylgeranylated or a combination of both. Additionally, these proteins have specific 4-amino acid motifs at the C-terminus that determine the type of prenylation at single or dual cysteines. Prenylation occurs in

the ER and is often part of a stepwise process of PTMs that is followed by proteolytic cleavage by Rce1 and methylation by isoprenyl cysteine methyltransferase (ICMT).

Proteolysis

Peptide bonds are indefinitely stable under physiological conditions, and therefore cells require some mechanism to break these bonds. Proteases comprise a family of enzymes that cleave the peptide bonds of proteins and are critical in antigen processing, apoptosis, surface protein shedding and cell signaling.

The family of over 11,000 proteases varies in substrate specificity, mechanism of peptide cleavage, location in the cell and the length of activity. While this variation suggests a wide array of functionalities, proteases can generally be separated into groups based on the type of proteolysis. Degradative proteolysis is critical to remove unassembled protein subunits and misfolded proteins and to maintain protein concentrations at homeostatic concentrations by reducing a given protein to the level of small peptides and single amino acids. Proteases also play a biosynthetic role in cell biology that includes cleaving signal peptides from nascent proteins and activating zymogens, which are inactive enzyme precursors that require cleavage at specific sites for enzyme function. In this respect, proteases act as molecular switches to regulate enzyme activity.

Proteolysis is a thermodynamically favorable and irreversible reaction. Therefore, protease activity is tightly regulated to avoid uncontrolled proteolysis through temporal and/or spatial control mechanisms including regulation by cleavage in cis or trans and compartmentalization (e.g., proteasomes, lysosomes).

The diverse family of proteases can be classified by the site of action, such as aminopeptidases and carboxypeptidase, which cleave at the amino or carboxy terminus of a protein, respectively. Another type of classification is based on the active site groups of a given protease that are involved in proteolysis. Based on this classification strategy, greater than 90% of known proteases fall into one of four categories as follows:

- Serine proteases
- Cysteine proteases
- Aspartic acid proteases

- Zinc metalloproteases

Inhibitors of protein synthesis

It is a substance that stops or slows the growth or proliferation of cells by disrupting the processes that lead directly to the generation of new proteins.

While a broad interpretation of this definition could be used to describe nearly any antibiotic, in practice, it usually refers to substances that act at the ribosome level (either the ribosome itself or the translation factor), taking advantages of the major differences between prokaryotic and eukaryotic ribosome structures.

Toxins such as ricin also function via protein synthesis inhibition. Ricin acts at the eukaryotic 60S.Examples:

- Neomycin
- Geneticin, also called G418

Mechanism

In general, protein synthesis inhibitors work at different stages of prokaryotic mRNA translation into proteins, like initiation, elongation (including aminoacyl tRNA entry, proofreading, peptidyl transfer, and ribosomal translocation) and termination:

Earlier stages

- Rifampicin inhibits prokaryotic DNA transcription into mRNA by inhibiting DNA-dependent RNA polymerase by binding its beta-subunit.

Initiation

Linezolid acts at the initiation stage, probably by preventing the formation of the initiation complex, although the mechanism is not fully understood.

Aminoacyl tRNA entry

- Tetracyclines and Tigecycline (a glycylcycline related to tetracyclines) block the A site on the ribosome, preventing the binding of aminoacyl tRNAs.

Proofreading

- Aminoglycosides, among other potential mechanisms of action, interfere with the proofreading process, causing increased rate of error in synthesis with premature termination.

Peptidyl transfer

- Chloramphenicol blocks the peptidyl transfer step of elongation on the 50S ribosomal subunit in both bacteria and mitochondria.
- Macrolides (as well as inhibiting ribosomal translocation and other potential mechanisms) bind to the 50s ribosomal subunits, inhibiting peptidyl transfer.
- Quinupristin/dalfopristin act synergistically, with dalfopristin, enhancing the binding of quinupristin, as well as inhibiting peptidyl transfer. Quinupristin binds to a nearby site on the 50S ribosomal subunit and prevents elongation of the polypeptide, as well as causing incomplete chains to be released.

Ribosomal translocation

- Macrolides, clindamycin and aminoglycosides (with all these three having other potential mechanisms of action as well), have evidence of inhibition of ribosomal translocation.
- Fusidic acid prevents the turnover of elongation factor G (EF-G) from the ribosome.

Termination

- Macrolides and clindamycin (both also having other potential mechanisms) cause premature dissociation of the peptidyl-tRNA from the ribosome.
- Puromycin has a structure similar to that of the tyrosinyl aminoacyl-tRNA. Thus, it binds to the ribosomal A site and participates in peptide bond formation, producing peptidyl-puromycin. However, it does not engage in translocation and quickly dissociates from the ribosome, causing a premature termination of polypeptide synthesis.
- Streptogramins also cause premature release of the peptide chain.

Protein synthesis inhibitors of unspecified mechanism

- Retapamulin

Binding site

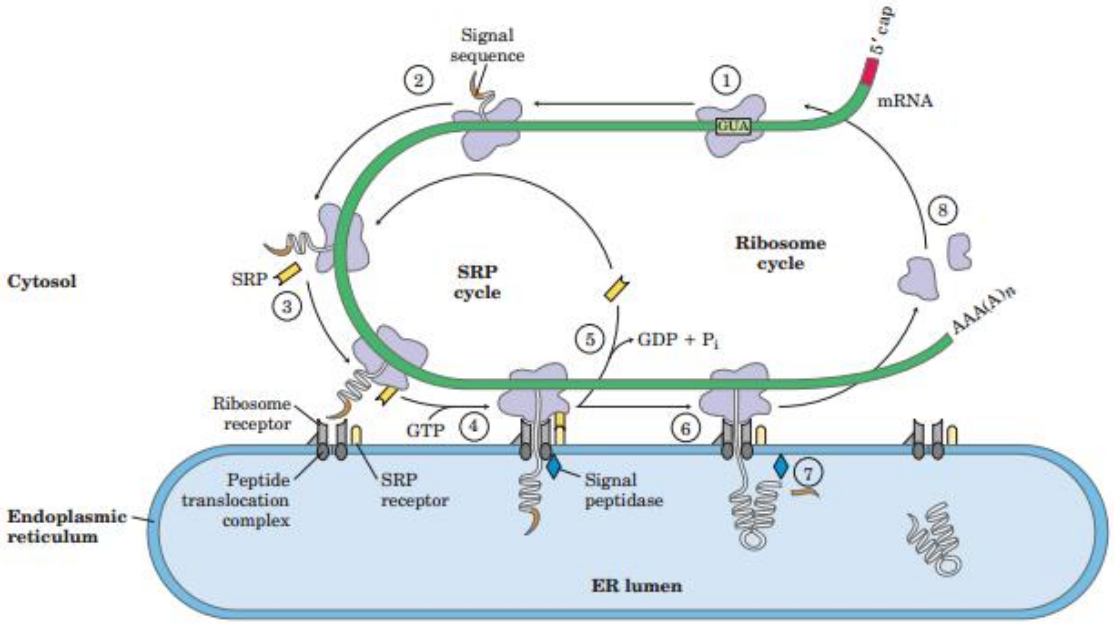
The following antibiotics bind to the 30S subunit of the ribosome:

- Aminoglycosides
- Tetracyclines

The following antibiotics bind to the 50S ribosomal subunit:

- Chloramphenicol
- Erythromycin
- Clindamycin
- Linezolid
- Telithromycin
- Streptogramins
- Retapamulin

Directing eukaryotic proteins with the appropriate signals to the endoplasmic reticulum

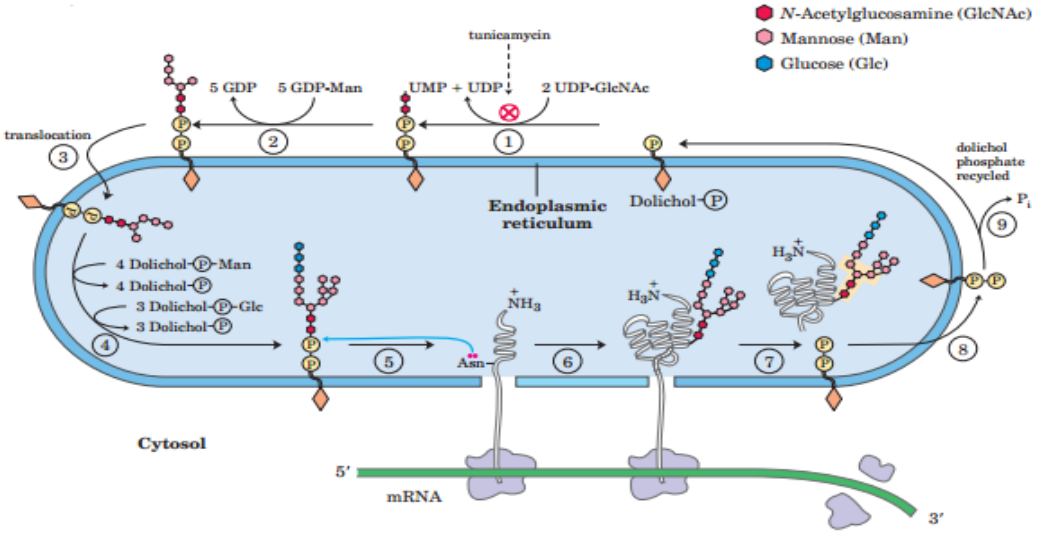


Posttranslational mdification of many eukaryotic proteins begins in the endoplasmic reticulum. Most lysosomal, membrane, or secreted proteins have an amino-terminal signal sequence that marks them for translocation into the lumen of the ER; The carboxyl terminus of the signal sequence is defined by a cleavage site, where protease action removes the sequence after the protein is imported into the ER. Signal sequences vary in length from 13 to 36 amino acid residues, but all have the following features: (1) about 10 to 15 hydrophobic amino acid residues; (2) one or more positively charged residues, usually near the amino terminus, preceding the hydrophobic sequence; and (3) a short sequence at the carboxyl terminus (near the cleavage site) that is relatively polar, typically having amino acid residues with short side chains (especially Ala) at the positions closest to the cleavage site. Proteins with these signal sequences

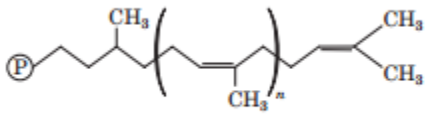
are synthesized on ribosomes attached to the ER. The signal sequence itself helps to direct the ribosome to the ER.

1. The targeting pathway begins with initiation of protein synthesis on free ribosomes.
2. The signal sequence appears early in the synthetic process, because it is at the amino terminus, which is synthesized first.
3. As it emerges from the ribosome, the signal sequence and the ribosome itself are bound by the large signal recognition particle (SRP); SRP then binds GTP and halts elongation of the polypeptide when it is about 70 amino acids long and the signal sequence has completely emerged from the ribosome.
4. The GTP-bound SRP now directs the ribosome (still bound to the mRNA) and the incomplete polypeptide to GTP-bound SRP receptors in the cytosolic face of the ER; the nascent polypeptide is delivered to a peptide translocation complex in the ER, which may interact directly with the ribosome.
5. SRP dissociates from the ribosome, accompanied by hydrolysis of GTP in both SRP and the SRP receptor.
6. Elongation of the polypeptide now resumes, with the ATP-driven translocation complex feeding the growing polypeptide into the ER lumen until the complete protein has been synthesized.
7. The signal sequence is removed by a signal peptidase within the ER lumen
8. The ribosome dissociates and is recycled.

Role of glycosylation in protein targeting



In the ER lumen, following the removal of signal sequences, polypeptides are folded, disulfide bonds formed, and many proteins glycosylated to form glycoproteins. In many glycoproteins the linkage to their oligosaccharides is through Asn residues. A 14 residue core oligosaccharide is built up in a step-wise fashion, and then transferred from a dolichol phosphate donor molecule to certain Asn residues in the protein. The transferase is on the luminal face of the ER and thus cannot catalyze glycosylation of cytosolic proteins. After transfer, the core oligosaccharide is trimmed and elaborated in different ways on different proteins, but all N-linked oligosaccharides retain a pentasaccharide core derived from the original 14 residue oligosaccharide.



Antibiotics such as tunicamycin act by interfering with one or more steps in this process. It mimics the structure of UDP-N-acetylglucosamine and blocks the first step of the process. A few proteins are O-glycosylated in the ER, but most O-glycosylation occurs in the Golgi

complex or in the cytosol (for proteins that do not enter the ER). Suitably modified proteins can now be moved to a variety of intracellular destinations. Proteins travel from the ER to the Golgi complex in transport vesicles. In the Golgi complex, oligosaccharides are O-linked to some proteins, and N-linked oligosaccharides are further modified. By mechanisms not yet fully understood, the Golgi complex also sorts proteins and sends them to their final destinations. The processes that segregate proteins targeted for secretion from those targeted for the plasma membrane or lysosomes must distinguish among these proteins on the basis of structural features other than signal sequences, which were removed in the ER lumen.

This sorting process is best understood in the caseof hydrolases destined for transport to lysosomes. The three-dimensional structure of the hydrolase is recognized by a phosphotransferase, which phosphorylates certain mannose residues in the oligosaccharide. The presence of one or more mannose 6-phosphate residues in its N-linked oligosaccharide is the structural signal that targets the protein to lysosomes. A receptor protein in the membrane of the Golgi complex recognizes the mannose 6-phosphate signal and binds the hydrolase so marked. Vesicles containing these receptor-hydrolase complexes bud from the trans side of the Golgi complex and make their way to sorting vesicles. Here, the receptor-hydrolase complex dissociates in a process facilitated by the lower pH in the vesicle and by phosphatase-catalyzed removal of phosphate groups from the mannose 6-phosphate residues. The receptor is then recycled to the Golgi complex, and vesicles containing the hydrolases bud from the sorting vesicles and move to the lysosomes. In cells treated with tunicamycin hydrolases that should be targeted for lysosomes are instead secreted, confirming that the N-linked oligosaccharide plays a key role in targeting these enzymes to lysosomes.

The pathways that target proteins to mitochondria and chloroplasts also rely on amino-terminal signal sequences. Although mitochondria and chloroplasts contain DNA, most of their proteins are encoded by nuclear DNA and must be targeted to the appropriate organelle. Unlike other targeting pathways, however, the mitochondrial and chloroplast pathways begin only after a precursor protein has been completely synthesized and released from the ribosome. Precursor proteins destined for mitochondria or chloroplasts are bound by cytosolic chaperone proteins and delivered to receptors on the exterior surface of the target organelle. Specialized translocation

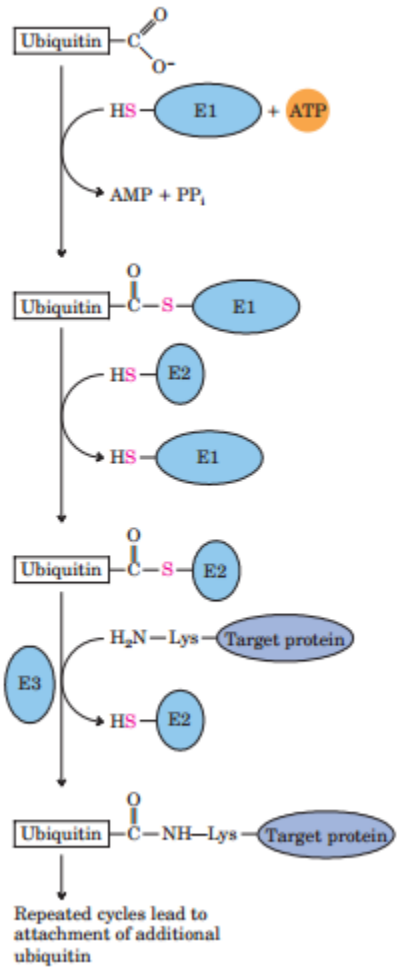
mechanisms then transport the protein to its final destination in the organelle, after which the signal sequence is removed.

Protein degradation-Ubiquitin mediated pathway:

Protein degradation prevents the buildup of abnormal or unwanted proteins and permits the recycling of amino acids. The half-lives of eukaryotic proteins vary from 30 seconds to many days. Most proteins turn over rapidly relative to the lifetime of a cell, although a few (such as hemoglobin) can last for the life of the cell (about 110 days for an erythrocyte). Rapidly degraded proteins include those that are defective because of incorrectly inserted amino acids or because of damage accumulated during normal functioning. And enzymes that act at key regulatory points in metabolic pathways often turn over rapidly.

Defective proteins and those with characteristically short half-lives are generally degraded in both bacterial and eukaryotic cells by selective ATP-dependent cytosolic systems. A second system in vertebrates, operating in lysosomes, recycles the amino acids of mem-brane proteins, extracellular proteins, and proteins with characteristically long half-lives. In E. coli, many proteins are degraded by an ATP-dependent protease called Lon (the name refers to the "long form" of proteins, observed only when this pro-tease is absent). The protease is activated in the presence of defective proteins or those slated for rapid turnover; two ATP molecules are hydrolyzed for every peptide bond cleaved. The precise role of this ATP hydrolysis is not yet clear. Once a protein has been reduced to small inactive peptides, other ATP-independent proteases complete the degradation process.

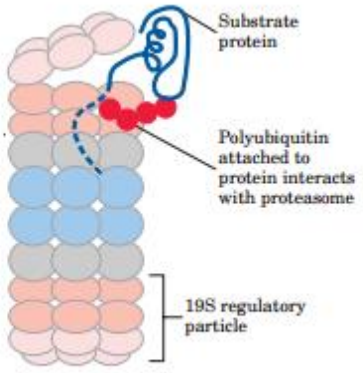
The ATP-dependent pathway in eukaryotic cells is quite different, involving the protein ubiquitin,which, as its name suggests, occurs throughout the eukaryotic kingdoms. One of the most highly conserved proteins known, ubiquitin (76 amino acid residues) is essentially identical in organisms as different as yeasts and humans. Ubiquitin is covalently linked to proteins slated for de-struction via an ATP-dependent pathway involving three separate enzymes (E1, E2, and E3).



Ubiquitinated proteins are degraded by a large complex known as the 26S proteasome. The proteasome consists of two copies each of at least 32 different subunits, most of which are highly conserved from yeasts to humans. The proteasome contains two main types of subcomplexes, a barrel-like core particle and regulatory particles on either end of the barrel. The 20S core particle consists of four rings; the outer rings are formed from seven α subunits, and the inner rings from seven β subunits. Three of the seven subunits in each β ring have protease activities, each with different substrate specificities. The stacked rings of the core particle form the barrel-like structure within which target proteins are degraded. The 19S regulatory particle on each end of the core particle con-tains 18 subunits, including some that recognize and bind to

ubiquitinated proteins. Six of the subunits are ATPases that probably function in unfolding the ubiquitinated proteins and translocating the unfolded polypeptide into the core particle for degradation.

Although we do not yet understand all the signals that trigger ubiquitination, one simple signal has been found. For many proteins, the identity of the first residue that remains after removal of the amino-terminal Met residue, and any other posttranslational proteolytic processing of the amino-terminal end, has a profound influence on half-life. These amino-terminal signals have been conserved over billions of years of evolution, and are the same in bacterial protein degradation systems and in the human ubiquitination pathway. More complex signals, are also being identified. Ubiquitin-dependent proteolysis is as important for the regulation of cellular processes as for the elimination of defective proteins. Many proteins required at only one stage of the eukaryotic cell cycle are rapidly degraded by the ubiquitin-dependent pathway after completing their function. The same pathway also processes and presents class I MHC antigens. Ubiquitin dependent destruction of cyclin is critical to cell-cycle regulation. The E2 and E3 components of the ubiquitination cascade pathway are in fact two large families of proteins. Different E2 and E3 enzymes exhibit different specificities for target proteins and thus regulate different cellular processes. Some E2 and E3 enzymes are highly localized in certain cellular compartments, reflecting a specialized function.



Complete proteasome

Not surprisingly, defects in the ubiquitination pathway have been implicated in a wide range of disease states. An inability to degrade certain proteins that activate cell division (the products of oncogenes) can lead to tumor formation, whereas a too-rapid degradation of proteins

that acts as tumor suppressors can have the same effect. The ineffective or overly rapid degradation of cellular proteins also appears to play a role in a range of other conditions: renal diseases, asthma, neurodegenerative disorders such as Alzheimer's and Parkinson's diseases (associated with the formation of characteristic proteinaceous structures in neurons), cystic fibrosis (caused in some cases by a too-rapid degradation of a chloride ion channel, with resultant loss of function; Liddle's syndrome (in which a sodium channel in the kidney is not degraded, leading to excessive Na^+ absorption and early-onset hypertension)—and many other disorders. Drugs designed to inhibit proteasome function are being developed as potential treatments for some of these conditions. In a changing metabolic environment, protein degradation is as important to a cell's survival as is protein synthesis, and much remains to be learned about these interesting pathways.

POSSIBLE QUESTIONS

UNIT IV

Part-B

1. Explain the role of ribosomal RNA in translation
2. Describe the action of any three inhibitors in protein synthesis
3. Explain the wobble hypothesis
4. How does amino acyl tRNA recognize its corresponding tRNA?
5. Write the composition of eukaryotic and prokaryotic ribosomes
6. Explain the proof reading activity of amino acyl tRNA synthetase
7. What are the post translational modifications of proteins?
8. How are aminoacid activated?
9. Write in brief about the termination process of translation
10. How proteins of endoplasmic reticulum reach their destination after the synthesis.

Part-C

1. Describe the protein synthesis in prokaryotes
2. Give an account on inhibitors of protein synthesis
3. Explain the targeting of proteins to
   a) Nucleus
   b) Endoplasmic reticulum
   c) Golgi complex
4. List out the post translational modification of proteins
5. Explain about the fidelity of transcription

Karpagam acadmy of higher educaton
Department of biochemistry
Molecular biology (19BCP202) MCQ
UNIT IV

S.NO	UNIT	QUESTION	OPTION 1	OPTION 2	OPTION 3	OPTION 4	ANSWER
1	IV	Translation refers to synthesis of	DNA from RNA	RNA from DNA	protein from DNA	**protein from RNA**	**protein from RNA**
2	IV	The process of synthesizing protein from RNA	Replication	Transcription	**Translation**	Reverse transcription	**Translation**
3	IV	Translation occurs in the	Nucleus	**Cytoplasm**	Nucleolus	Lysosome	**Cytoplasm**
4	IV	Which is the energy rich molecule required for initiation of translation?	ATP	**GTP**	CTP	AMP	**GTP**
5	IV	The enzyme involved in amino acid activation is	ATP synthetase	**Aminoacyl tRNA synthetase**	Aminoacyl mRNA synthetase	Aminoacyl rRNA synthetase	**Aminoacyl tRNA synthetase**
6	IV	Genetic code is the dictionary of the following nucleotide bases	A, G, C and T	**A, G, C and U**	A, G, T and U	A, G, T and T	**A, G, C and U**
7	IV	The initiating codon in protein synthesis	UAA	UAG	UGA	**AUG**	**AUG**
8	IV	The following are the termination codons except	**AUG**	UAA	UAG	UGA	**AUG**
9	IV	The codon (of mRNA) and anticodon (of tRNA) recognize each other by pairing	5'-3' of mRNA with 5'-3' of tRNA	3'-5' of mRNA with 3'-5' of tRNA	**5'-3' of mRNA with 3'-5' of tRNA**	5'-3' of tRNA with 3'-5' of mRNA	**5'-3' of mRNA with 3'-5' of tRNA**
10	IV	Wobble hypothesis explained by one of the following characteristics of the genetic code	Universality	Specificity	non overlapping	**degenerate**	**degenerate**
11	IV	Wobble hypothesis is characterized by recognition of	a single codon by a single tRNA	**more than one codon by a single tRNA**	a single codon by more than one tRNA	a single codon by a single mRNA	more than one codon by a single tRNA
12	IV	The specific information required for the sequence amino acids in protein is located on	rRNA	**mRNA**	tRNA	hnRNA	**mRNA**
13	IV	The factories for protein synthesis	Mitochondria	nucleus	**ribosomes**	golgi bodies	**ribosomes**
14	IV	During translation mRNA is read in 5'-3' direction &polypeptide synthesis proceeds	C-terminal end to C-terminal end	C-terminal end to N terminal end	**N-terminal end to C-terminal end**	N-terminal end to N-terminal end	**N-terminal end to C-terminal end**
16	IV	The chain initiating amino acid in protein biosynthesis in E.Coli is	Methionine	Cysteine	**N formyl methionine**	Homocystine	**N formyl methionine**
17	IV	Rifampicin inhibit protein synthesis by binding to	**beta subunit of RNA pol**	delta subunit of RNA pol	gamma subunit of RNA pol	alpha subunit of RNA pol	**beta subunit of RNA pol**
18	IV	Nucleic acid concerned with the protein synthesis is	DNA	mRNA	**ribosomal RNA**	tRNA	**ribosomal RNA**
19	IV	During translation, proteins are synthesized	by ribosomes using the information on DNA	by lysosome using the information on DNA	**by ribosomes using the information on mRNA**	by ribosomes using the information on rRNA	**by ribosomes using the information on mRNA**
20	IV	During translation, the role of enzyme peptidyl transferase is	transfer of phosphate group	amino acid activation	**peptide bond formation between adjacent amino acids**	binding of ribosome subunits to mRNA	**peptide bond formation between adjacent amino acids**
21	IV	Which of the following are termination codons	**UAG, UAA, UGA**	UUA, AUC, GUC	UAG, AGG,UGA	UAG, GAA, GUC	**UAG, UAA, UGA**
22	IV	In prokaryotes protein synthesis is carried out by ribosomes of	80s,30s, and 50s	80s,40s, and 50s	**70s,30s, and 50s**	70s,40s, and 60s	**70s,30s, and 50s**
23	IV	The stop codon UAA is otherwise called as	Amber	**Ochre**	Opal	Umber	**Ochre**
24	IV	The number of 'primordial' amino acids are	15	17	18	**20**	**20**
25	IV	The binding of $tRNA^{fmet}$ to the P site is inhibited by	**Streptomycin**	Chloramphenicol	Erythromycin	Lincomycin	**Streptomycin**
26	IV	Translocase is an enzyme required in the process of	DNA replication	RNA synthesis	Initiation of protein synthesis	**Elongation of peptides**	**Elongation of peptides**
27	IV	Which one of the following post translational modifications occurs in clotting factors?	Hydroxylation	methylation	**carboxylation**	phosphorylation	**carboxylation**
28	IV	Diphtheria toxin inhibits elongation step in translation through	Glycosylation	Methylation	**ADP- ribosylation**	Ubiquitinylation	**ADP- ribosylation**
29	IV	The Shine Dalgarno sequence responsible for initiation of protein synthesis is found in	**bacterial mRNA**	bacterial & eukaryotic mRNA	Bacterial rRNA	bacterial & eukaryotic rRNA	**bacterial mRNA**
30	IV	Selection of translational initiation site is carried out by	**Base pairing between mRNA and 16rRNA**	initiation codon	stop codon	initiation and stop codon	**Base pairing between mRNA and 16rRNA**
31	IV	How many polypeptide chains can be formed simultaneously by a given ribosome?	**One**	about a dozen	up to 30	variable depending on mRNA	**One**
32	IV	RF1 recognises	**UAA**	GUG	UGA	AUG	**UAA**
33	IV	The antibiotic cycloheximide	blocks chain elongation during protein synthesis	blocks glycosylation of proteins	resembles aminoacyl tRNA	**inhibits the formation of peptide bond**	**inhibits the formation of peptide bond**
34	IV	Larger subunit of ribosome is present in eukaryotes	**60s**	40s	20s	80s	**60s**
35	IV	Large subunit of ribosome present in prokaryotes is	**50s**	40s	20s	30s	**50s**
36	IV	Sequence responsible for translation process	AT rich sequence	**Shine dalgarno sequence**	GC rich sequence	AT and GC rich sequence	**Shine dalgarno sequence**
37	IV	In protein, aminoacids are joined by _____ bond	**Peptide**	Phosphodiester	Hydrogen	Covalent	**Peptide**
38	IV	Which of the following Binds to 50S subunit and inhibits translocation	Chloroamphenical	Cycloheximide	**Erythromycin**	Puromycin	**Erythromycin**
39	IV	Eukaryotic ribosomal RNA was synthesised by	**DNA Pol I**	RNA Pol I	RNA Pol II	RNA Pol III	**DNA Pol I**
40	IV	Principle function of RNA pol II in eukaryote is	**Synthesis mRNA**	Synthesis tRNA	Synthesis rRNA	Synthesis DNA	**Synthesis mRNA**
41	IV	Chlroamphenicol inhibits	**Peptidyl transferase**	Binding of amino acyl Trna	Translocation	Transversion	**Peptidyl transferase**
42	IV	The dissociation of 70S ribosome is carried out by	IF 1	IF 2	**IF 3**	IF 4	**IF 3**
43	IV	Premature termination of poly peptide is caused by	Erythromycin	**Puromycin**	Streptomycin	Ricin	**Puromycin**
44	IV	The codon that terminates protein synthesis is	UGG	**UGA**	UGC	UGU	**UGA**
45	IV	Which of the following antibiotics mimics amino acyl tRNA	Ricin	**Puromycin**	Streptomycin	Erythromycin	**Puromycin**

46	IV	The drug chloramphenicol blocks	cell wall formation	transcription	translation termination release factors	**polypeptide chain elongation**	**polypeptide chain elongation**
47	IV	The Wobble hypothesis accounts for the	triple nature of codon	**degeneracy of codon**	universality of the codon	doublet nature of the codon	**degeneracy of codon**
48	IV	Shine Dalgarno sequence is	rich in pyrimindine	**rich in purine**	centered 100 nucleotide upstream from the start codon	centered 100 nucleotide downstream from the start codon	**rich in purine**
49	IV	Which of the following amino acids is specified by only a single codon?	Glutamine	Tryptophan	Asparagine	**Isoleucine**	**Isoleucine**
50	IV	The enzyme responsible for the cleavage of the bond connecting amino acid and tRNA is	amino acyl-tRNA synthetase	peptidyl transferase	aminoacyl transferase	**tRNA deacylase**	**tRNA deacylase**
51	IV	Coupled transcription and translation occurs	only in eukaryotes	**only in prokaryotes**	both in prokaryotes and in eukaryotes	neither in prokaryotes nor in eukaryotes	**only in prokaryotes**
52	IV	Which of the following antibiotics is an inhibitor of protein synthesis	**Chloramphenicol**	Penicillin	Bacitracin	Cephalosporin	**Chloramphenicol**
53	IV	The amino acyl synthetase must	recognize the codon	recognize the anticodon	**be able to distinguish one amino acid from another**	recognize both codon and anticodon	**be able to distinguish one amino acid from another**
54	IV	tRNA, 5S RNA, Special RNA s are synthesised by RNA Pol____	I	II	**III**	IV	**III**
55	IV	Catalytic RNA is	Abzyme	**Ribozyme**	Enzyme	Protein	**Ribozyme**
56	IV	One of the following is distinctive structure of eukaryotic mRNA	**poly A tail**	intron	exon	poly G tail	**poly A tail**
57	IV	Large protein complex that involved in intron removal is	Replisome	Primosome	**Spliceosome**	Ribosome	**Spliceosome**
58	IV	5' cap in mRNA is done by	ATP	**GTP**	CTP	UTP	**GTP**
59	IV	Tetracycline blocks protein synthesis by	**Inhibiting binding of aminoacyl tRNA to ribosome**	Inhibiting initiation of translation	Inhibiting peptidyl transferase	Inhibiting translocase enzyme	**Inhibiting binding of aminoacyl tRNA to ribosome**
60	IV	In eukaryotes, translation is initiated by binding of ribosome to the	Pribnows box	Hogness box	**5'cap**	Poly A tail	**5'cap**

UNIT-V

SYLLABUS

Prokaryotic gene regulation: Operon model, Lac, trp and ara operons.Regulatory proteins-DNA binding domain, protein- protein interaction domain.Recombination- holiday model, Rec BCD enzymes, Rec A protein, MesselsonRadding model, site- specific recombination.Antisense RNA technology.

Eukaryotic gene regulation: Transcriptionally active chromatin, chromatin remodeling, DNA binding transactivators and coactivators. Regulation of gene expression by intracellular and intercellular signal, RNAi.

Introduction

In genetics, an operon is a functioning unit of genomic material containing a cluster of genes under the control of a single regulatory signal or promoter. The genes are transcribed together into an mRNA strand and either translated together in the cytoplasm, or undergo trans-splicing to create monocistronic mRNAs that are translated separately, i.e. several strands of mRNA that each encode a single gene product. The result of this is that the genes contained in the operon are either expressed together or not at all. Originally operons were thought to exist solely in prokaryotes but since the discovery of the first operons in eukaryotes in the early 1990s, more evidence has arisen to suggest they are more common than previously assumed. Several genes must be both co-transcribed and co-regulated to define an operon.

Structure

This is the general structure of an operon:

- Promoter – a nucleotide sequence that enables a gene to be transcribed. The promoter is recognized by RNA polymerase, which then initiates transcription. In RNA synthesis, promoters indicate which genes should be used for messenger RNA creation – and, by extension, control which proteins the cell manufactures.

- Operator – a segment of DNA that a regulator binds to. It is classically defined in the lac operon as a segment between the promoter and the genes of the opero. In the case of a repressor, the repressor protein physically obstructs the RNA polymerase from transcribing the genes.

- Structural genes – the genes that are co-regulated by the operon.

Regulation

Control of an operon is a type of gene regulation that enables organisms to regulate the expression of various genes depending on environmental conditions. Operon regulation can be either negative or positive by induction or repression.[6]

Negative control involves the binding of a repressor to the operator to prevent transcription.

- In negative inducible operons, a regulatory repressor protein is normally bound to the operator and it prevents the transcription of the genes on the operon. If an inducer molecule is present, it binds to the repressor and changes its conformation so that it is unable to bind to the operator. This allows for expression of the operon.

- In negative repressible operons, transcription of the operon normally takes place. Repressor proteins are produced by a regulator gene but they are unable to bind to the operator in their normal conformation. However certain molecules called corepressors are bound by the repressor protein, causing a conformational change to the active state. The activated repressor protein binds to the operator and prevents transcription.

Operons can also be positively controlled. With positive control, an activator protein stimulates transcription by binding to DNA (usually at a site other than the operator).

- In positive inducible operons, activator proteins are normally unable to bind to the pertinent DNA. When an inducer is bound by the activator protein, it undergoes a change in conformation so that it can bind to the DNA and activate transcription.

- In positive repressible operons, the activator proteins are normally bound to the pertinent DNA segment. However, when a corepressor is bound by the activator, it is prevented from binding the DNA. This stops activation and transcription of the system.

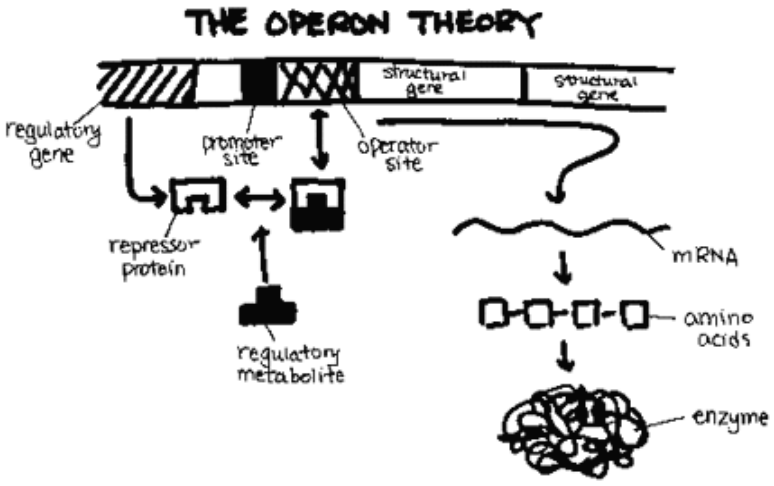


Figure 5.1 The operon theory

Inducible genes - The operon model

Definition

An inducible gene is a gene that is expressed in the presence of a substance (an inducer) in the environment. This substance can control the expression of one or more genes (structural genes) involved in the metabolism of that substance. For example, lactose induces the expression of the *lac* genes that are involved in lactose metabolism. An certain antibiotic may induce the expression of a gene that leads to resistance to that antibiotic.

Induction is common in metabolic pathways that result in the catabolism of a substance and the inducer is normally the substrate for the pathway

Lactose Operon

Structural genes

The lactose operon (figure 1) contains three structural genes that code for enzymes involved in lactose metabolism.

- The *lac z* gene codes for β-galactosidase, an enzyme that breaks down lactose into glucose and galactose
- The *lac y* gene codes for a permease, which is involved in uptake of lactose
- The *lac a* gene codes for a galactosetransacetylase.

These genes are transcribed from a common promoter into a polycistronic mRNA, which is translated to yield the three enzymes.

Regulatory gene

The expression of the structural genes is not only influenced by the presence or absence of the inducer, it is also controlled by a specific regulatory gene. The regulatory gene may be next to or far from the genes that are being regulated. The regulatory gene codes for a specific protein product called a REPRESSOR.

Operator

The repressor acts by binding to a specific region of the DNA called the operator which is adjacent to the structural genes being regulated. The structural genes together with the operator region and the promoter are called an OPERON. However, the binding of the repressor to the operator is prevented by the inducer and the inducer can also remove repressor that has already

bound to the operator. Thus, in the presence of the inducer the repressor is inactive and does not bind to the operator, resulting in transcription of the structural genes. In contrast, in the absence of inducer the repressor is active and binds to the operator, resulting in inhibition of transcription of the structural genes. This kind of control is referred to a NEGATIVE CONTROL since the function of the regulatory gene product (repressor) is to turn off transcription of the structural genes.

Inducer

Transcription of the lac genes is influenced by the presence or absence of an inducer (lactose or other β-galactosides) (Figure 2).

e.g	+ inducer	expression
	- inducer	no expression

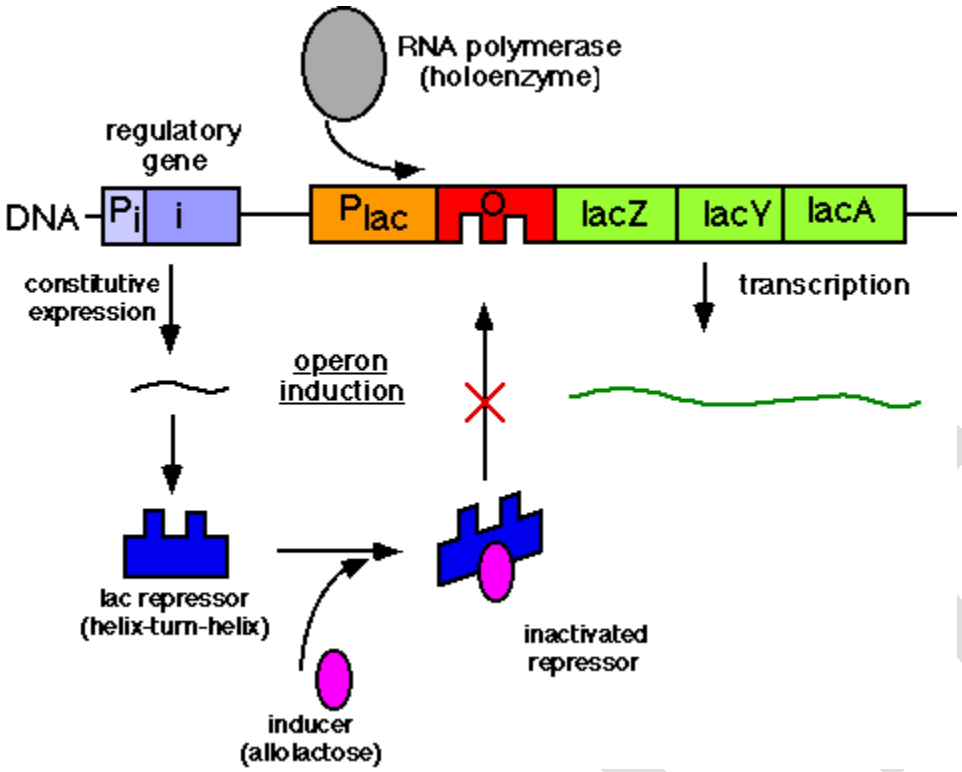


Figure 5.2 Lac operon

Catabolite repression (Glucose Effect)

Many inducible operons are not only controlled by their respective inducers and regulatory genes, but they are also controlled by the level of glucose in the environment. The ability of glucose to control the expression of a number of different inducible operons is called

Catabolite Repression

Catabolite repression is generally seen in those operons which are involved in the degradation of compounds used as a source of energy. Since glucose is the preferred energy source in bacteria, the ability of glucose to regulate the expression of other operons ensures that bacteria will utilize glucose before any other carbon source as a source of energy.

Mechanism

There is an inverse relationship between glucose levels and cyclic AMP (cAMP) levels in bacteria. When glucose levels are high cAMP levels are low and when glucose levels are low cAMP levels are high. This relationship exists because the transport of glucose into the cell inhibits the enzyme adenylcyclase which produces cAMP. In the bacterial cell cAMP binds to a cAMP binding protein called CAP or CRP. The cAMP-CAP complex, but not free CAP protein,

binds to a site in the promoters of catabolite repression-sensitive operons. The binding of the complex results in a more efficient promoter and thus more initiations of transcriptions from that promoter as illustrated in Figures 4 and 5.Since the role of the CAP-cAMP complex is to turn on transcription this type of control is said to be POSITIVE CONTROL. The consequences of this type of control is that to achieve maximal expression of a catabolite repression sensitive operon glucose must be absent from the environment and the inducer of the operon must be present. If both are present, the operon will not be maximally expressed until glucose is metabolized. Obviously, no expression of the operon will occur unless the inducer is present.

The Tryptophan Operon

Repressible genes - The operon model

Definition

Repressible genes are those in which the presence of a substance (a co-repressor) in the environment turns off the expression of those genes (structural genes) involved in the metabolism of that substance.

e.g., Tryptophan represses the expression of the trp genes.

Repression is common in metabolic pathways that result in the biosynthesis of a substance and the co-repressor is normally the end product of the pathway being regulated.

Structural genes

The tryptophan operon (figure 6) contains five structural genes that code for enzymes involved in the synthesis of tryptophan. These genes are transcribed from a common promoter into a polycistronic mRNA, which is translated to yield the five enzymes.

Regulatory gene

The expression of the structural genes is not only influenced by the presence or absence of the co-repressor, it is also controlled by a specific regulatory gene. The regulatory gene may be next to or far from the genes that are being regulated. The regulatory gene codes for a specific protein product called an REPRESSOR (sometimes called an apo-repressor). When the repressor is synthesized it is inactive. However, it can be activated by complexing with the co-repressor (*i.e.* tryptophan).

Operator

The active repressor/co-repressor complex acts by binding to a specific region of the DNA called the operator which is adjacent to the structural genes being regulated. The structural genes

together with the operator region and the promoter is called an OPERON. Thus, in the presence of the co-repressor the repressor is active and binds to the operator, resulting in repression of transcription of the structural genes. In contrast, in the absence of co-repressor the repressor is inactive and does not bind to the operator, resulting in transcription of the structural genes. This kind of control is referred to a NEGATIVE CONTROL since the function of the regulatory gene product (repressor) is to turn off transcription of the structural genes.

Co-repressor

Transcription of the tryptophan genes is influenced by the presence or absence of a co-repressor (tryptophan) (Figure 7).

e.g.	+co-repressor	no expression
	-co-repressor	expression

Attenuation

In many repressible operons, transcription that initiates at the promoter can terminate prematurely in a leader region that precedes the first structural gene. (*i.e.* the polymerase terminates transcription before it gets to the first gene in the operon). This phenomenon is called ATTENUATION; the premature termination of transcription. Although attenuation is seen in a number of operons, the mechanism is best understood in those repressible operons involved in amino acid biosynthesis. In these instances attenuation is regulated by the availability of the cognate aminoacylated t-RNA.

Mechanism

When transcription is initiated at the promoter, it actually starts before the first structural gene and a leader transcript is made. This leader region contains a start and a stop signal for protein synthesis. Since bacteria do not have a nuclear membrane, transcription and translation can occur simultaneously. Thus, a short peptide can be made while the RNA polymerase is transcribing the leader region. The test peptide contains several tryptophan residues in the middle of the peptide. Thus, if there is a sufficient amount of tryptophanyl-t-RNA to translate that test peptide, the entire peptide will be made and the ribosome will reach the stop signal. If, on the other hand, there is not enough tryptophanyl-t-RNA to translate the peptide, the ribosome will be arrested at the two tryptophan codons before it gets to the stop signal.

The sequence in the leader m-RNA contains four regions, which have complementary sequences (Figure 9). Thus, several different secondary stem and loop structures can be formed. Region 1 can only form base pairs with region 2; region 2 can form base pairs with either region 1 or 3; region 3 can form base pairs with region 2 or 4; and region 4 can only form base pairs with region 3. Thus three possible stem/loop structures can be formed in the RNA.

region 1:region 2

region 2:region 3

region 3:region 4

One of the possible structures (region 3 base pairing with region 4) generates a signal for RNA polymerase to terminate transcription (*i.e.* to attenuate transcription). However, the formation of one stem and loop structure can preclude the formation of others. If region 2 forms base pairs with region 1 it is not available to base pair with region 3. Similarly if region 3 forms base pairs with region 2 it is not available to base pair with region 4.

The ability of the ribosomes to translate the test peptide will affect the formation of the various stem and loop structures Figure 10. If the ribosome reaches the stop signal for translation it will be covering up region 2 and thus region 2 will not available for forming base pairs with other regions. This allows the generation of the transcription termination signal because region 3 will be available to pair with region 4. Thus, when there is enough tryptophanyl-t-RNA to translate the test peptide attenuation will occur and the structural genes will not be transcribed. In contrast, when there is an insufficient amount of tryptophanyl-t-RNA to translate the test peptide no attenuation will occur. This is because the ribosome will stop at the two tryptophan codons in region 1, thereby allowing region 2 to base pair with region 3 and preventing the formation of the attenuation signal (*i.e.* region 3 base paired with region 4). Thus, the structural genes will be transcribed.

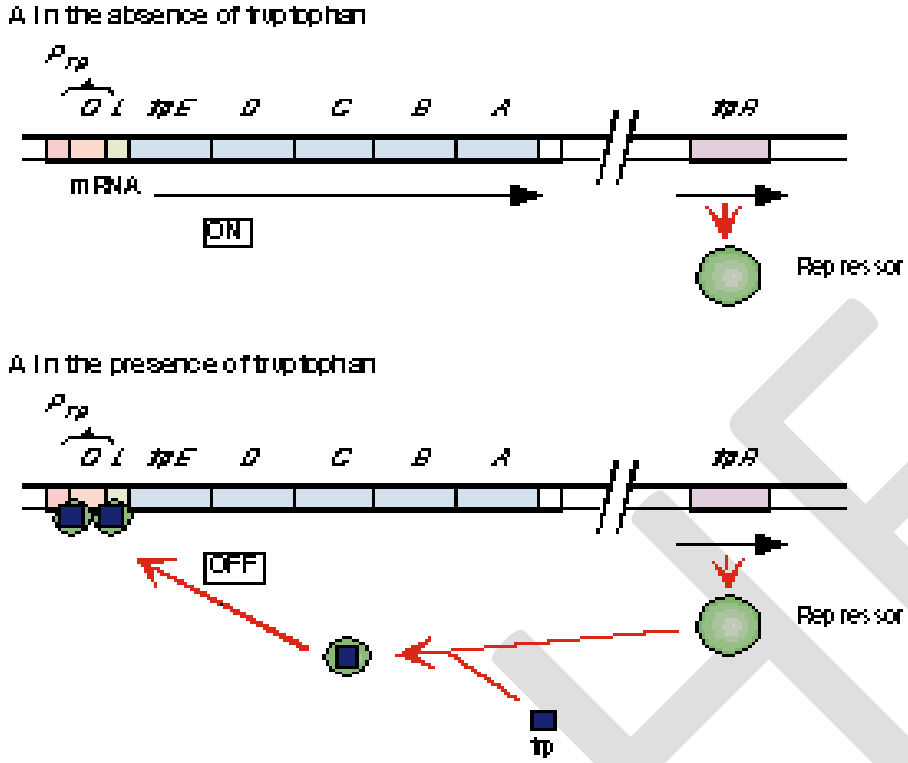
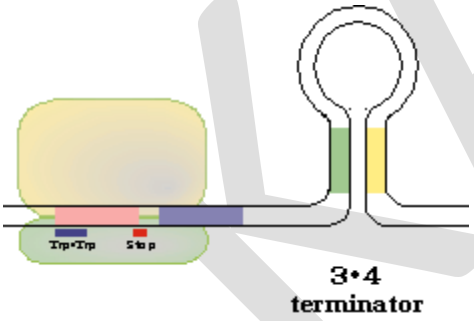


Figure 5.3 Trp operon



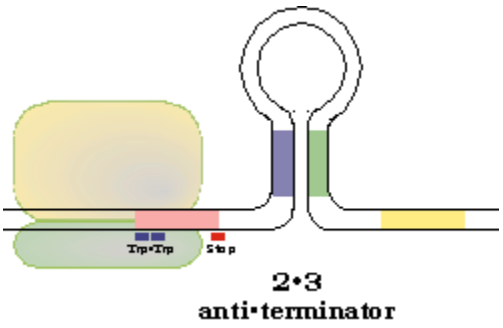


Figure 5.4 Termination process

Attenuation in other operons

Attenuation as a means of regulating expression occurs in a number of other amino acid biosynthetic operons. In all cases, the leader region is rich in codons for the particular amino acids that are synthesized by the enzymes encoded by the particular operon and it can form two alternative stem-loop structures, one of which is a transcription terminator.

The Arabinose Operon

The ***ara*** operon codes for three enzymes that are required to catalyze the metabolism of arabinose.

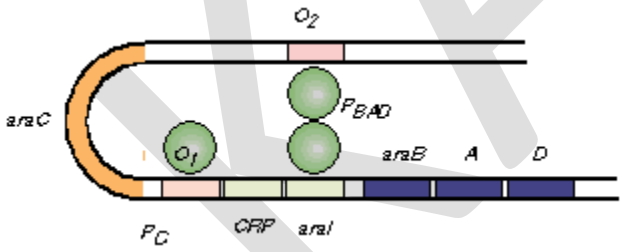
- Arabinose isomerase - encoded by ***araA*** - coverts arabinose to ribulose
- Ribulokinase - encoded by ***araB*** -- phosphorylates ribulose
- Ribulose-5-phosphate epimerase - encoded by ***araD*** -- converts ribulose-5-phosphate to xylulose-5-phosphate which can then be metabolized via the pentose phosphate pathway.

The three structural genes are arranged in an operon that is regulated by the ***araC***gene product. There are four important regulatory sites as shown in the following diagram:

- ***araO$_1$*** is an operator site. **AraC** binds to this site and represses its own transcription from the **P$_C$** promoter. In the presence of arabinose, however, **AraC** bound at this site helps to activate expression of the **P$_{BAD}$**promoter.
- ***araO$_2$***is also an operator site. **AraC** bound at this site can simultaneously bind to the ***araI*** site to repress transcription from the **P$_{BAD}$** promoter
- ***araI***is also the inducer site. **AraC** bound at this site can simultaneously bind to the ***araO$_2$*** site to repress transcription from the **P$_{BAD}$** promoter. In the presence of arabinose, however, **AraC** bound at this site helps to activate expression of the **P$_{BAD}$** promoter.
- **CRP** binds to the **CRP** binding site. It does not directly assist RNA polymerase to bind to the promoter in this case. Instead, in the presence of arabinose, it promotes the rearrangement of **AraC** when arabinose is present from a state in which it represses transcription of the **P$_{BAD}$** promoter to one in which it activates transcription of the **P$_{BAD}$** promoter.

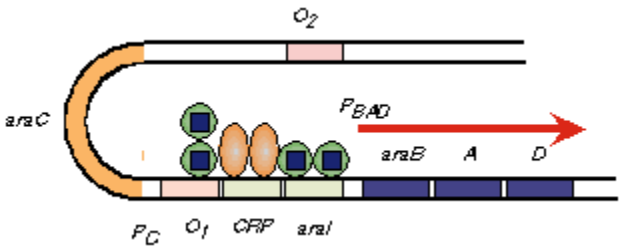
Regulation of the arabinose operon is, clearly, much more complex than the lactose operon.

When arabinose is absent, there is no need to express the structural genes. **AraC** does this by binding simultaneously to ***araI*** and ***araO$_2$***. As a result the intervening DNA is **looped**. These two events block access to the **P$_{BAD}$** promoter which is, in any case, a very weak promoter (unlike the ***lac*** promoter):

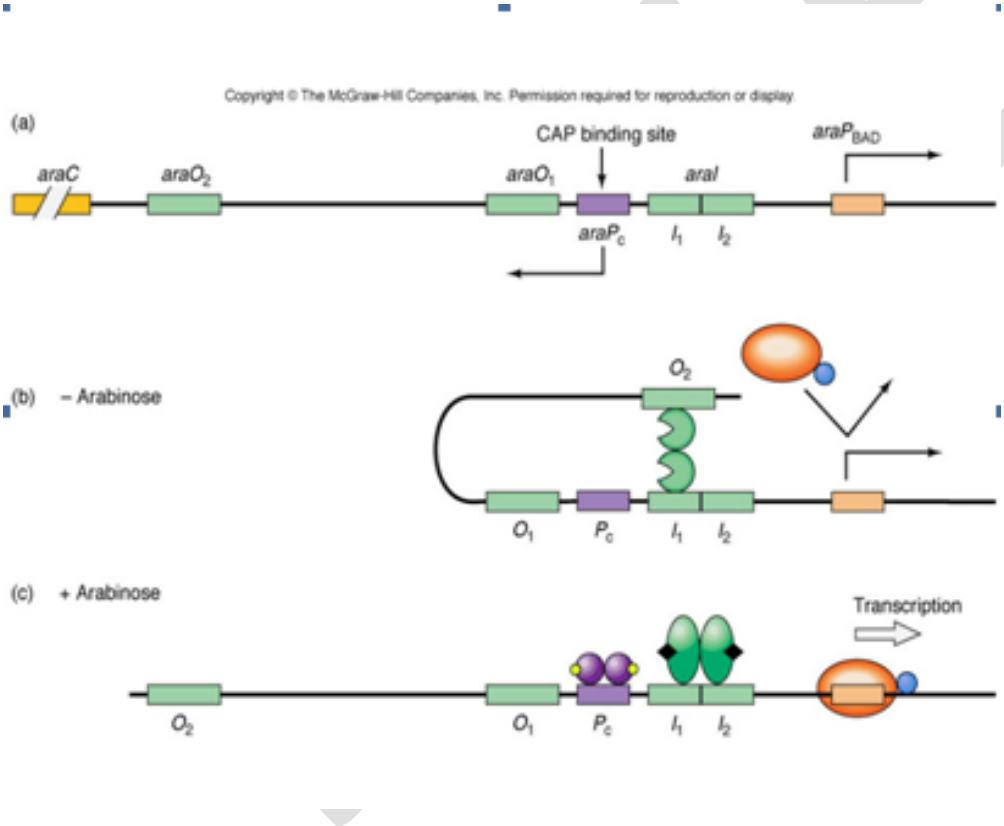


AraC also prevents its own expression. Thus, it is an autoregulator of its own expression. This makes sense; there is no need to over-express **AraC**. If the concentration falls too low then transcription of ***araC*** resumes until the amount of **AraC** is sufficient to prevent more transcription again.

When arabinose is present, it binds to **AraC** and allosterically induces it to bind to ***araI*** instead ***araO$_2$***. If **glucose** is also absent, then the presence of **CRP** bound to its site between ***araO$_1$*** and ***araI*** helps to break the DNA loop and also helps **AraC** to bind to ***araI***:



The ***ara*** operon demonstrates both negative and positive control. It shows a different function for **CRP**. It also shows how a protein can act as a switch with its activity being radically altered upon the binding of a small molecule.



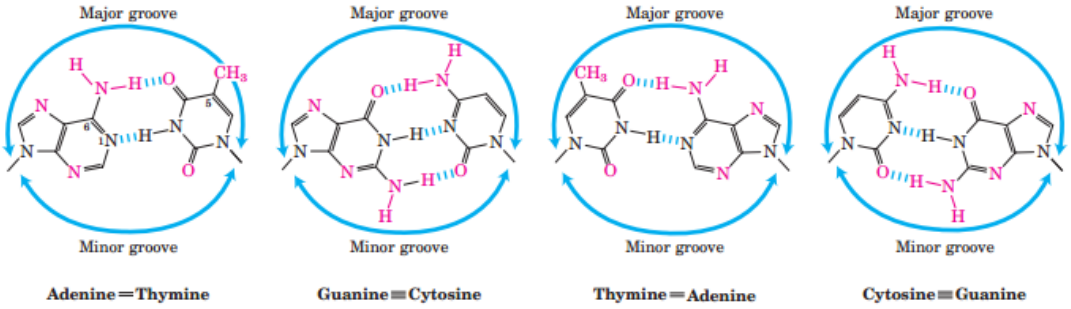
Regulatory proteins

DNA binding domain

Regulatory proteins generally bind to specific DNA sequences. Their affinity for these target sequences is roughly 10^4 to 10^6 times higher than their affinity for any other DNA sequences. Most regulatory proteins have discrete DNA-binding domains containing sub-structures that

interact closely and specifically with the DNA. These binding domains usually include one or more of a relatively small group of recognizable and characteristic structural motifs.

To bind specifically to DNA sequences, regulatory proteins must recognize surface features on the DNA. Most of the chemical groups that differ among the four bases and thus permit discrimination between base pairs are hydrogen-bond donor and acceptor groups exposed in the major groove of DNA , and most of the protein-DNA contacts that impart specificity are hydrogen bonds.



A notable exception is the nonpolar surface near C-5 of pyrimidines, where thymine is readily distinguished from cytosine by its protruding methyl group.

Protein-DNA contacts are also possible in the minor groove of the DNA, but the hydrogen-bonding patterns here generally do not allow ready discrimination between base pairs. Within regulatory proteins, the amino acid side chains most often hydrogen-bonding to bases in the DNA are those of Asn, Gln, Glu, Lys, and Arg residues. Is there a simple recognition code in which a particular amino acid always pairs with a particular base? The two hydrogen bonds that can form between Gln or Asn and the N6 and N-7 positions of adenine cannot form with any other base. And an Arg residue can form two hydrogen bonds with N-7 and O6 of guanine.

Examination of the structures of many DNA-binding proteins, however, has shown that a protein can recognize each base pair in more than one way, leading to the conclusion that there is no simple amino acid–base code. For some proteins, the Gln-adenine interaction can specify AUT base pairs, but in others a van der Waals pocket for the methyl group of thymine can recognize AUT base pairs. To interact with bases in the major groove of DNA, a protein requires a relatively small structure that can stably protrude from the protein surface. The DNA-binding domains of regulatory proteins tend to be small (60 to 90 amino acid residues), and the structural mo-tifs within these domains that are actually in contact with the DNA are smaller still. Many small proteins are unstable because of their limited capacity to form layers of structure to bury hydrophobic groups.

The DNA-binding motifs provide either a very compact stable structure or a way of allowing a segment of protein to protrude from the protein surface. The DNA-binding sites for regulatory proteins are often inverted repeats of a short DNA sequence (a palindrome) at which multiple (usually two) subunits of a regulatory protein bind cooperatively. The Lac repressor is unusual in that it functions as a tetramer, with two dimers tethered together at the end distant from the DNA-binding sites. An E. coli cell normally contains about 20 tetramers of the Lac repressor. Each of the tethered dimers separately binds to a palindromic operator sequence, in contact with 17 bp of a 22 bp region in the lac operon. And each of the tethered dimers can independently bind to an operator sequence, with one generally binding to $O_1$and the other to O_2or O_3. The symmetry of the O1 operator sequence corresponds to the twofold axis of symmetry of two paired Lac repressor subunits. The tetrameric Lac repressor binds to its operator sequences in vivo with an estimated dissociation constant of about 10^{-10} M. The repressor discriminates between the operators and other sequences by a factor of about 106, so binding to these few base pairs among the 4.6 million or so of the E. coli chromosome is highly specific. Several DNA-binding motifs have been described, but here we focus on two that play prominent roles in the binding of DNA by regulatory proteins: the helix-turn-helix and the zinc finger. We also consider a type of DNA-binding domain—the homeodomain—found in some eukaryotic proteins.

Helix-Turn-Helix

This DNA-binding motif is crucial to the interaction of many prokaryotic regulatory proteins with DNA, and similar motifs occur in some eukaryotic regulatory proteins. The helix-turn-helix motif comprises about 20 amino acids in two short α-helical segments, each seven to nine amino acid residues long, separated by a β turn. This structure generally is not stable by itself; it is simply the reactive portion of a somewhat larger DNA-binding domain. One of the two α-helical segments is called the recognition helix, be-cause it usually contains many of the amino acids that interact with the DNA in a sequence-specific way. This α helix is stacked on other segments of the protein structure so that it protrudes from the protein surface. When bound to DNA, the recognition helix is positioned in or nearly in the major groove. The Lac repressor has this DNA-binding motif

Zinc Finger

In a zinc finger, about 30 amino acid residues form an elongated loop held together at the base by a single Zn2+ ion, which is coordinated to four of the residues (four Cys, or two Cys and two His). The zinc does not itself interact with DNA; rather, the coordination of zinc with the amino acid residues stabilizes this small structural motif. Several hydrophobic side chains in the core of the structure also lend stability.

Many eukaryotic DNA-binding proteins contain zinc fingers. The interaction of a single zinc finger with DNA is typically weak, and many DNA-binding proteins, like Zif268, have multiple zinc fingers that substantially enhance binding by interacting simultaneously with the DNA. One DNA-binding protein of the frog Xenopus has 37 zinc fingers. There are few known examples of the zinc finger motif in prokaryotic proteins.

Some zinc fingers contain the amino acid residues that are important in sequence discrimination, whereas others appear to bind DNA nonspecifically (the amino acids required for specificity are located elsewhere in the protein). Zinc fingers can also function as RNA-binding motifs—for example, in certain proteins that bind eukaryotic mRNAs and act as translational repressors.

Homeodomain

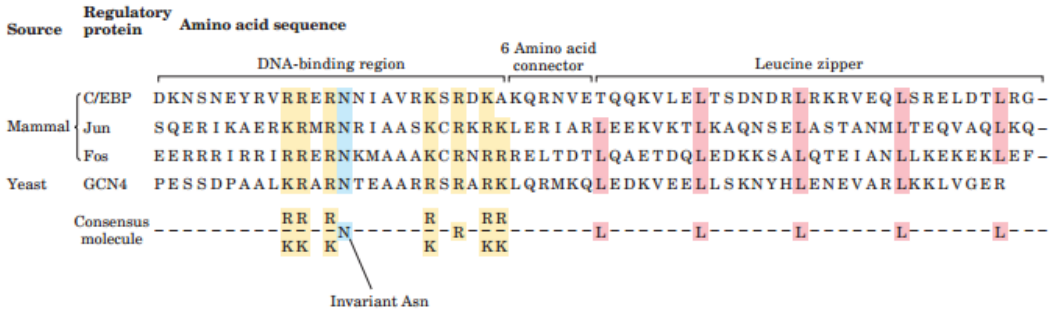
Another type of DNA-binding domain has been identified in a number of proteins that function as transcriptional regulators, especially during eukaryotic development. This domain of 60 amino acids—called the homeodomain, because it was discovered in homeotic genes (genes that regulate the development of body pat-terns)—is highly conserved and has now been identified in proteins from a wide variety of organisms, including humans DNA-binding segment of the domain is related to the helix-turn-helix motif. The DNA sequence that encodes this domain is known as the homeobox.

Protein protein interaction domain

Regulatory proteins contain domains not only for DNA binding but also for protein-protein interactions—with RNA polymerase, other regulatory proteins, or other sub-units of the same regulatory protein. Examples include many eukaryotic transcription factors that function as gene activators, which often bind as dimers to the DNA, using DNA-binding domains that contain zinc fingers. Some structural domains are devoted to the interactions required for dimer formation, which is generally a pre-requisite for DNA binding. Like DNA-binding motifs, the structural motifs that mediate protein-protein interactions tend to fall within one of a few common categories.

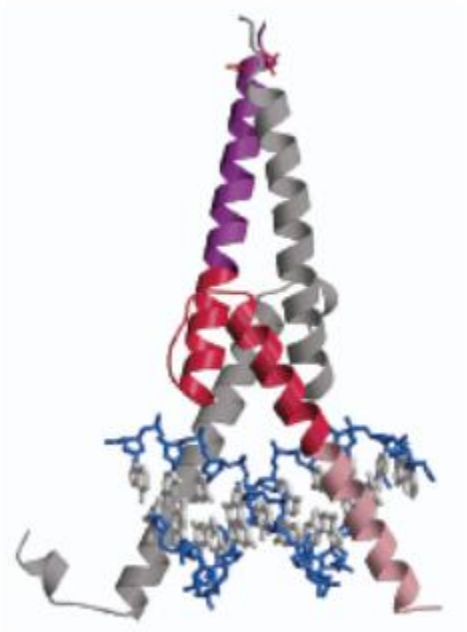
Two important examples are the leucine zipper and the basic helix-loop-helix. Structural motifs such as these are the basis for classifying some regulatory proteins into structural families.

Leucine Zipper This motif is an amphipathic α helix with a series of hydrophobic amino acid residues concentrated on one side, with the hydrophobic surface forming the area of contact between the two polypeptides of a dimer.



A striking feature of these α helices is the occurrence of Leu residues at every seventh position, forming a straight line along the hydrophobic surface. Although researchers initially thought the Leu residues interdigitated (hence the name "zipper"), we now know that they line up side by side as the interacting α helices coil around each other forming a coiled coil. Regulatory proteins with leucine zippers often have a separate DNA-binding domain with a high concentration of basic (Lys or Arg) residues that can interact with the negatively charged phosphates of the DNA backbone. Leucine zippers have been found in many eukaryotic and a few prokaryotic proteins.

Basic Helix-Loop-Helix Another common structural motif occurs in some eukaryotic regulatory proteins implicated in the control of gene expression during the development of multicellular organisms. These proteins share a conserved region of about 50 amino acid residues important in both DNA binding and protein dimerization. This region can form two short amphipathic α helices linked by a loop of variable length, the helix-loop-helix (distinct from the helix-turn-helix motif associated with DNA binding). The helix-loop-helix motifs of two polypeptides interact to form dimmers.



In these proteins, DNA binding is mediated by an adjacent short amino acid sequence rich in basic residues, similar to the separate DNA-binding region in proteins containing leucine zippers.

Recombination

Rec A protein

RecA protein is required for homologous recombination mediated by both the RecBCD and RecFOR pathways in E. coli. In addition to promoting genetic recombination, RecA plays an important role in the recombinational repair of DNA damage, and the induction of the SOS system in response to DNA damage. The E. coli RecA protein has multiple biochemical functions: it binds to both single- and double-stranded DNA, it promotes pairing and exchange of homologous DNA, it catalyzes ATP hydrolysis, and it promotes the proteolytic cleavage of several proteins including LexA, UmuD, and the lambda cI repressor. RecA is highly conserved between bacteria. Homologs of RecA have also been found in Archae (RadA and RadB), Eukarya (Rad51, Rad57, Rad55 and DMC1), and phage (e.g., the phage T4 UvsX protein). The RecA protein family website provides links to recA sequences from a variety of organisms.

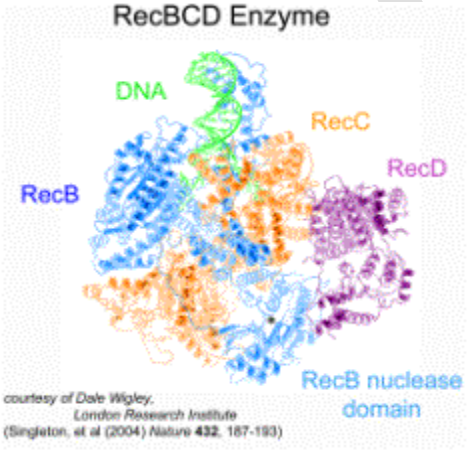
RecBCDEnzyme

RecBCD, also known as Exonuclease V, is an enzyme of the E. coli bacterium that initiates recombinational repair from potentially lethal double strand breaks in DNA which may result from ionizing radiation, replication errors, endonucleases, oxidative damage, and a host of other factors.[1] The RecBCD enzyme is both a helicase that unwinds, or separates the strands of, DNA and a nuclease that makes single-stranded nicks in DNA.

Structure

The enzyme complex is composed of three different subunits called RecB, RecC, and RecD and hence the complex is named RecBCD (Figure 1). Before the discovery of the *recD* gene,[3] the enzyme was known as "RecBC." Each subunit is encoded by a separate gene:

gene	chain	protein function
RecB	beta	P08394 3'-5' helicase, nuclease
RecC	gamma	P07648 Likely recognizes Chi (*c*rossover *h*otspot *i*nstigator)
RecD	alpha	P04993 5'-3' helicase



Function

Both the RecD and RecB subunits are helicases, i.e., energy-dependent molecular motors that unwind DNA (or RNA in the case of other proteins). The RecB subunit in addition has a

nuclease function.[4] Finally, RecBCD enzyme (perhaps the RecC subunit) recognizes a specific sequence in DNA, 5'-GCTGGTGG-3', known as Chi (sometimes designated with the Greek letter χ).

RecBCD is unusual amongst helicases because it has two helicases that travel with different rates[5] and because it can recognize and be altered by the Chi DNA sequence.[6][7]RecBCD avidly binds an end of linear double-stranded (ds) DNA. The RecD helicase travels on the strand with a 5' end at which the enzyme initiates unwinding, and RecB on the strand with a 3' end. RecB is slower than RecD, so that a single-stranded (ss) DNA loop accumulates ahead of RecB (Figure 2). This produces DNA structures with two ss tails (a shorter 3' ended tail and a longer 5' ended tail) and one ss loop (on the 3' ended strand) observed by electron microscopy.[8] The ss tails can anneal to produce a second ss loop complementary to the first one; such twin-loop structures were initially referred to as "rabbit ears."

Holliday model for recombination

The widely accepted model for DNA crossover was first proposed by Robin Holliday in 1964. It involves several steps as illustrated in the following figure.

The Holliday model of DNA crossover.

(a) Two homologous DNA molecules line up (e.g., two nonsister chromatids line up during meiosis).

(b) Cuts in one strand of both DNAs.

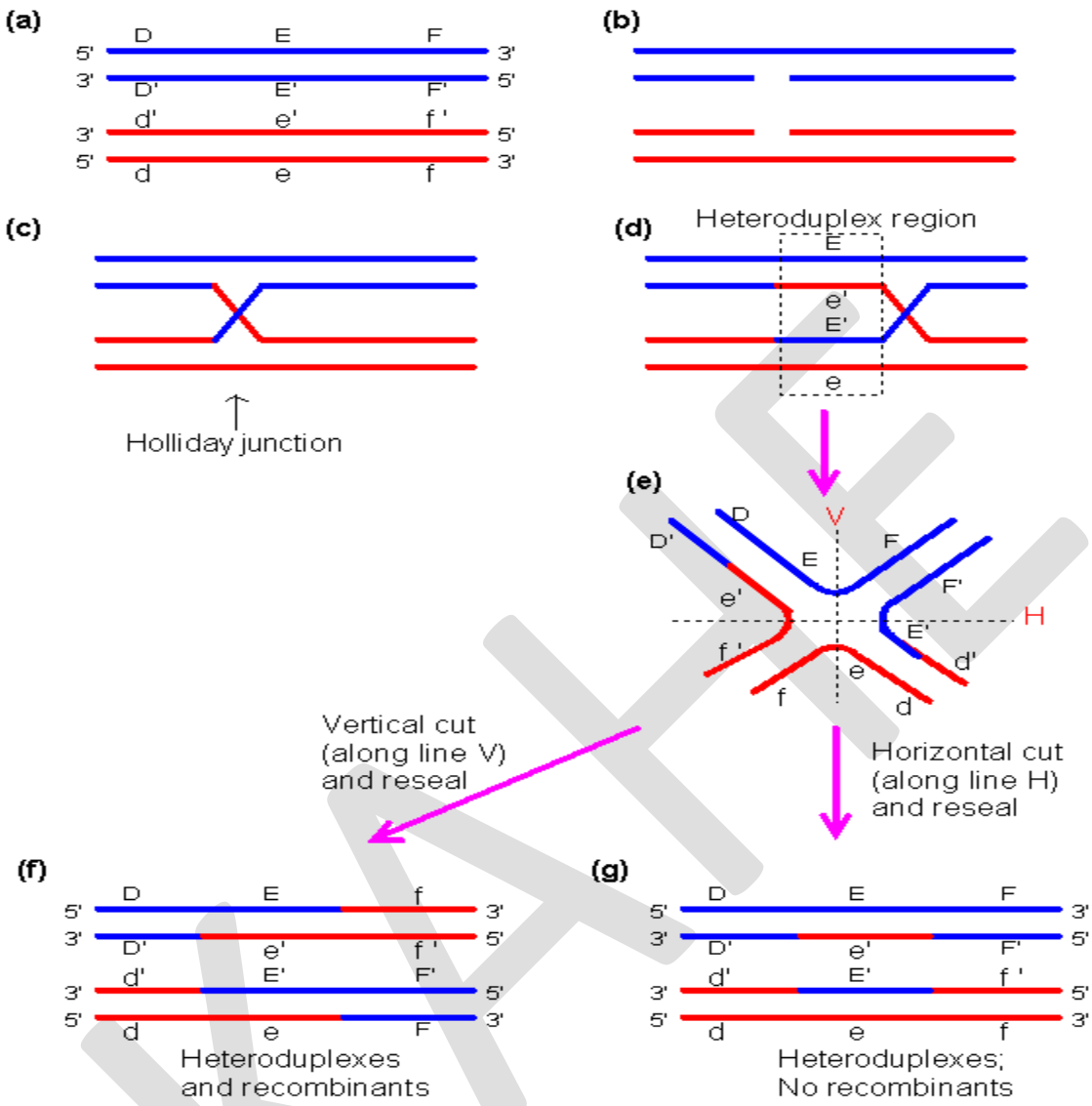
(c) The cut strands cross and join homologous strands, forming the Holliday structure (or Holliday junction).

(d) Heteroduplex region is formed by branch migration.

(e) Resolution of the Holliday structure. Figure 8-D-2e is a different view of the Holliday junction than Figure 8-D-2d. DNA strands may be cut along either the vertical line or horizontal line.

(f) The vertical cut will result in crossover between f-f' and F-F' regions. The heteroduplex region will eventually be corrected by mismatch repair.

(g) The horizontal cut does not lead to crossover after mismatch repair. However, it could cause gene conversion.



The detailed mechanism of homologous recombination was mainly obtained from the study of E. coli. Although bacteria do no undergo meiosis, homologous recombination could occur during or immediately after DNA replication. It may also occur in a mating process called conjugation.

In E. coli, the recombination is initiated by the enzyme RecBCD, consisting of three subunits: RecB, RecC and RecD. This enzyme has both helicase and nuclease activities. The enzyme first uses its helicase activity to unwind DNA. When it hits the Chi site (with sequence GCTGGTGG), one of the exposed strand will be cut by its nuclease activity. The reason why this special site is called the "Chi site" is because the Greek letter □ (chi) looks like a crossover point. The Chi site is the position of the Holliday junction and also the position of chiasma.

After DNA strands are cut by RecBCD, the strand invasion is catalyzed by RecA proteins, which can wrap around single stranded DNA and direct it to form the Holliday structure.

Finally, the branch migration is catalyzed by RuvA and RuvB. The Holliday structure is resolved by the protein RuvC.

Meselsonrading model

Meselson And Radding Model - The original Meselson and Radding model (1975) for the mechanisms of strand transfer bas been elaborated by Radding (1978). The mechanisms involved are strand displacement, uptake, loop cleavage, assimilation, isomerization and branch migration.

The Meselson-Radding model of meiotic recombination explains how the regions of heteroduplex DNA on both sides of the initial recombination event are not necessarily the same length, providing an advantage over the Holiday model of recombination.

Firstly, a nick is made in one strand of one of the chromosomes (in the following, each line represents a single strand of a double helix)

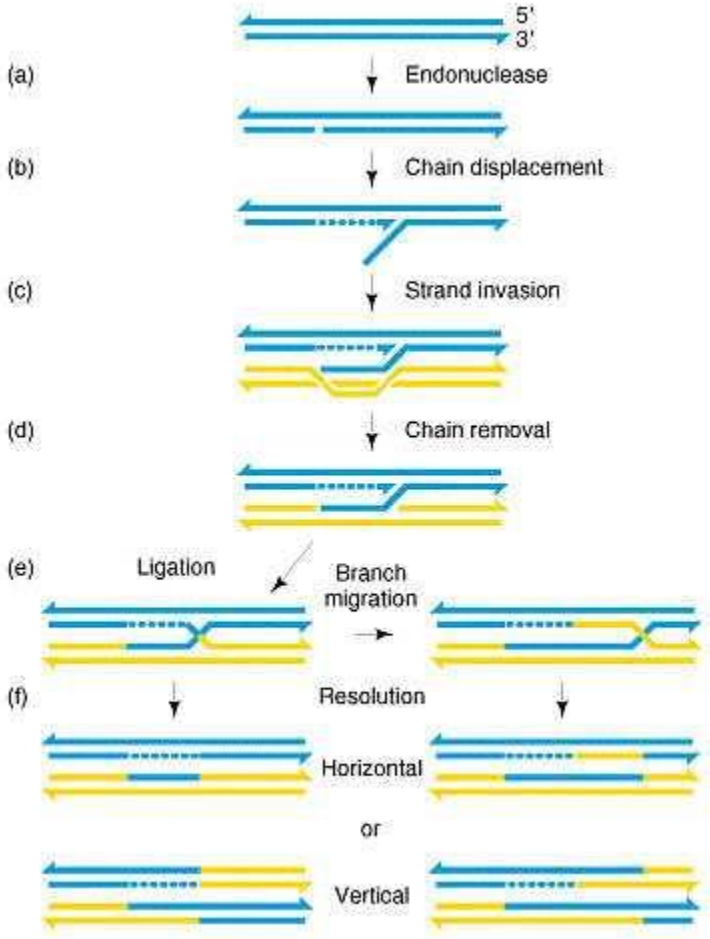
Secondly, a new strand is produced, displacing the original one on one side of the nick (the new strand is represented by c, but is identical to a)

Thirdly, the displaced a strand invades the other chromosome (this requires RecA)

Strand migration occurs

The junction may be resolved in one of two ways. One produces non-recombinant chromosomes with a heteroduplex region. The other produces recombinant chromosomes.

(Type 1 - non-recombinant)



Site specific recombination

Site-specific recombination differs from general recombination in that short specific sequences which are required for the recombination, are the only sites at which recombination occurs. These reactions invariably require specialized proteins to recognize these sites and to catalyze the recombination reaction at these sites.

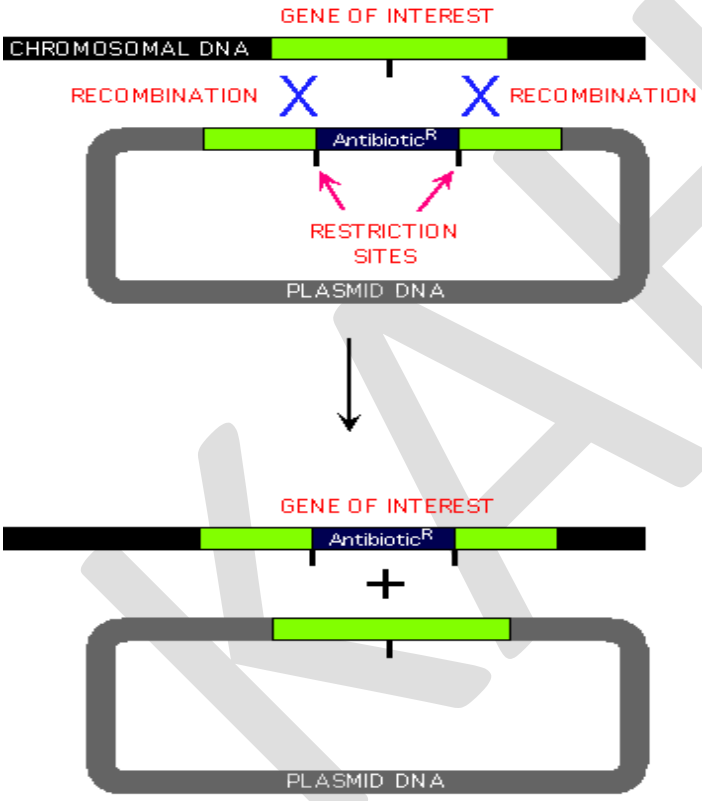
Because they involve specific sites, there are really only two types of site-specific recombination reaction.

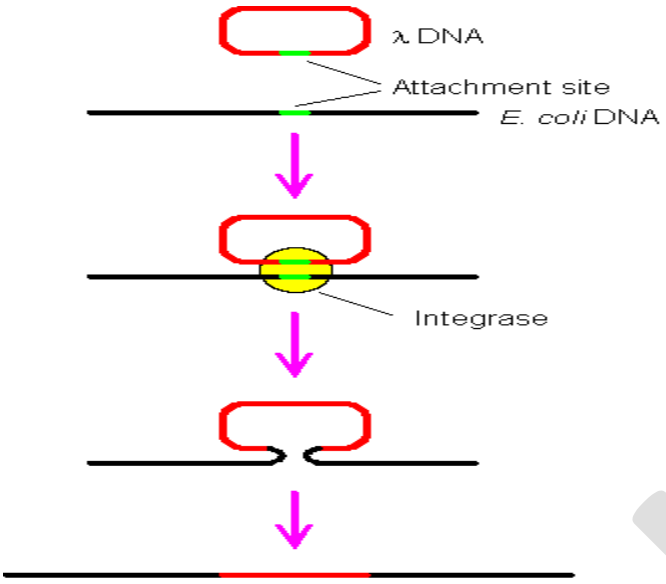
Site-specific recombination occurs at a specific DNA sequence. The first example was found in the integration between DNA and E. coli DNA. Both of them contain a sequence, 5'-TTTATAC-3', called the attachment site, which allows the two DNA molecules to attach together by base pairing. Once attached, the enzyme integrase catalyzes two single strand breaks

as in the Holliday model. After a short branch migration, the integrase exerts a second strand cuts on two other strands. Resolution of two Holliday junctions completes the integration process.

The steps and features of the general recombination reaction, however, still apply:

- strand exchange
- formation of a Holliday intermediate
- branch migration
- resolution.





Using General Recombination in a Site-specific Event

The construction of interposon mutants

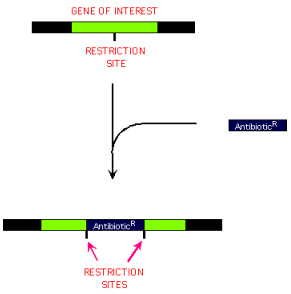
Whenever one is studying any particular gene, one invariably wants to know what happens if the gene cannot be expressed *in vivo*. To do this, one could use site-specific mutagenesis to place very specific base pair changes within a coding region. However, this is both difficult and requires a lot of information about the system to know exactly what sort of chnages must be made.

A much easier way to do this is simply to insert a large piece of DNA in the middle of the gene. This should block expression of the gene since transcription is unlikely to proceed all the way through the inserted DNA, and, even if it did, the mRNA that is made would not be translated into a functional protein.

If you're going to do this, then you need a way to know that the piece of DNA which you wish to insert really has been inserted. If the piece of DNA to be inserted codes for an antibiotic resistance, then you can select for its presence quite easily.

The following diagrams outline the procedure that is followed:

Cloning the antibiotic resistance piece of DNA into the middle of his gene was straightforward and only involved standard cloning techniques.



If we now transform or conjugate the plasmid into the bacterium of choice, at this point we have two copies of the target gene in the cell, but

- one of them has an antibiotic resistance element in the middle of it, and
- if, we are studying an organism other than *E. coli*, it is on a plasmid that cannot replicate.

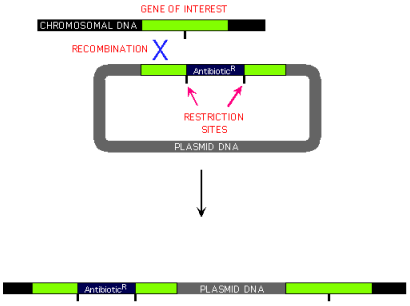
(Note that it is not always possible to transform bacteria with DNA. If transformation is not possible then conjugation or electroporation can be used to get DNA into the bacterium.)

Now if we select for cells that grow on media containing the antibiotic, they cannot do so unless the antibiotic resistance gene is maintained in the cell. However, if it is located on a plasmid that cannot replicate, it must integrate into the chromosome in order to be maintained and expressed.

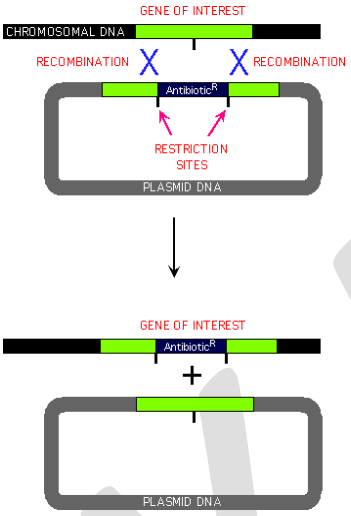
This can only happen if there is a region of homology in the plasmid which can recombine with the host chromosome.

There are, in fact, two such regions - one on either side of the antibiotic resistance gene, i.e. the two halves of the original target gene.

A single recombination on one side only will result in the integration of the whole plasmid containing the interrupted target gene. This would leave us with two copies of the RNA-binding protein gene in the cell:



Recombination on both sides would get rid of the wild-type copy of the gene. However, this can be difficult to find unless specially designed plasmid vectors are used that force the second recombination event to take place.



So this Interposon mutagenesis requires a site-specific recombination event in the sense that the recombination events must occur only in the range of places determined by a cloned piece of DNA. However, the mechanism by which this recombination takes place is still general homologous recombination.

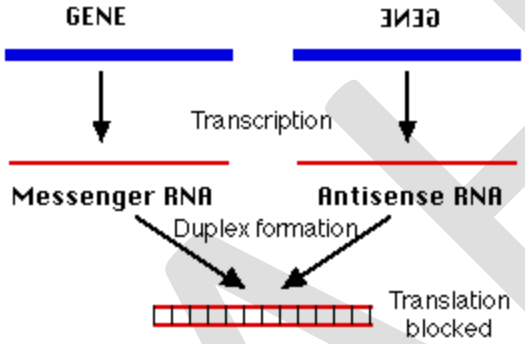
Antisense RNA technology

Antisense RNA

Messenger RNA (mRNA) is single-stranded. Its sequence of nucleotides is called "sense" because it results in a gene product (protein). Normally, its unpaired nucleotides are "read" by transfer RNA anticodons as the ribosome proceeds to **translate** the message.

However, RNA **can** form duplexes just as DNA does. All that is needed is a second strand of RNA whose sequence of bases is complementary to the first strand; e.g.,

5´ C A U G 3´ mRNA
3´ G U A C 5´ Antisense RNA



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

The second strand is called the antisense strand because its sequence of nucleotides is the complement of message sense. When mRNA forms a duplex with a complementary antisense RNA sequence, translation is blocked. This may occur because

- the ribosome cannot gain access to the nucleotides in the mRNA or
- duplex RNA is quickly degraded by ribonucleases in the cell (see RNAi below).

Example: the FlavrSavr tomato

Most tomatoes that have to be shipped to market are harvested before they are ripe. Otherwise, ethylene synthesized by the tomato causes them to ripen and spoil before they reach the customer. Transgenic tomatoes have been constructed that carry in their genome an artificial gene (DNA) that is transcribed into an antisense RNA complementary to the mRNA for an enzyme involved in ethylene production. These tomatoes make only 10% of the normal amount of the enzyme. The goal of this work was to provide supermarket tomatoes with something closer to the appearance and taste of tomatoes harvested when ripe. However, these tomatoes often became damaged during shipment and handling and have been taken off the market.

Regulation of gene expression in eukaryotes

Transcriptionally active chromatin

About 10% of the chromatin in a typical eukaryotic cell is in a more condensed form than the rest of the chromatin. This form, heterochromatin, is transcriptionally inactive. Heterochromatin is generally associated with particular chromosome structures—the centromeres, for example. The remaining, less condensed chromatin is called euchromatin.

Transcription of a eukaryotic gene is strongly repressed when its DNA is condensed within heterochromatin. Some, but not all, of the euchromatin is transcriptionally active. Transcriptionally active chromosomal regions can be detected based on their increased sensitivity to nuclease-mediated degradation. Nucleases such as DNase I tend to cleave the DNA of carefully isolated chromatin into fragments of multiples of about 200 bp, reflecting the regular repeating structure of the nucleosome. In actively transcribed regions, the fragments produced by nuclease activity are smaller and more heterogeneous in size. These regions contain hypersensitive sites, sequences especially sensitive to DNase I, which consist of about 100 to 200 bp within the 1,000 bp flanking the 5’ ends of transcribed genes. In some genes, hypersensitive sites are found farther from the 5’ end, near the 3’ end, or even within the gene itself. Many hypersensitive sites correspond to binding sites for known regulatory proteins, and the relative

absence of nucleosomes in these regions may allow the binding of these proteins. Nucleosomes are entirely absent in some regions that are very active in transcription, such as the rRNA genes.

Transcriptionally active chromatin tends to be deficient in histone H1, which binds to the linker DNA between nucleosome particles. Histones within transcriptionally active chromatin and heterochromatin also differ in their patterns of covalent modification. The core histones of nucleosome particles (H2A, H2B, H3, H4) are modified by irreversible methylation of Lys residues, phosphorylation of Ser or Thr residues, acetylation , or attachment of ubiquitin . Each of the core histones has two distinct structural domains. A central domain is involved in histone-histone interaction and the wrapping of DNA around the nucleosome. A second, lysine-rich amino-terminal domain is generally positioned near the exterior of the assembled nucleosome particle; the covalent modifications occur at specific residues concentrated in this amino-terminal domain. The patterns of modification have led some researchers to propose the existence of a histone code, in which modification patterns are recognized by enzymes that alter the structure of chromatin. Modifications associated with transcriptional activation would be recognized by enzymes that make the chromatin more accessible to the transcription machinery.

5-Methylation of cytosine residues of CpG sequences is common in eukaryotic DNA but DNA in transcriptionally active chromatin tends to be undermethylated. Furthermore, CpG sites in particular genes are more often undermethylated in cells from tissues where the genes are expressed than in those where the genes are not expressed. The overall pattern sug-gests that active chromatin is prepared for transcription by the removal of potential structural barriers.

Chromatin remodeling

The detailed mechanisms for transcription-associated structural changes in chromatin, called chromatin remodeling, are now coming to light, including identification of a variety of enzymes directly implicated in the process. These include enzymes that covalently modify the core histones of the nucleosome and others that use the chemical energy of ATP to remodel nucleosomes on the DNA. The acetylation and deacetylation of histones figure prominently in the processes that activate chromatin for transcription. As noted above, the amino-terminal domains of the core histones are generally rich in Lys residues. Particular Lys residues are acetylated by histone acetyltransferases (HATs). Cytosolic (type B) HATs acetylate newly synthesized histones before the histones are imported into the nucleus. The subsequent assembly of the histones into chromatin is facilitated by additional proteins: CAF1 for H3 and H4, and NAP1 for H2A and H2B.

Where chromatin is being activated for transcription, the nucleosomal histones are further acetylated by nuclear (type A) HATs. The acetylation of multiple Lys residues in the amino-terminal domains of histones H3 and H4 can reduce the affinity of the entire nucleosome for

DNA. Acetylation may also prevent or promote interactions with other proteins involved in transcription or its regulation. When transcription of a gene is no longer required, the acetylation of nucleosomes in that vicinity is reduced by the activity of histone deacetylases, as part of a general gene-silencing process that restores the chromatin to a transcriptionally inactive state. In addition to the removal of certain acetyl groups, new covalent modification of histones marks chromatin as transcriptionally inactive. As an example, the Lys residue at position 9 in histone H3 is often methylated in heterochromatin.

Chromatin remodeling also requires protein complexes that actively move or displace nucleosomes, hydrolyzing ATP in the process. The enzyme complex SWI/SNF found in all eukaryotic cells, contains 11 polypeptides that together create hypersensitive sites in the chromatin and stimulate the binding of transcription factors. SWI/SNF is not required for the transcription of every gene. NURF is another ATP-dependent enzyme complex that remodels chromatin in ways that complement and overlap the activity of SWI/SNF. These enzyme complexes play an important role in preparing a region of chromatin for active transcription.

DNA binding transactivator and coactivator

To continue our exploration of the regulation of gene expression in eukaryotes, we return to the interactions between promoters and RNA polymerase II (Pol II), the enzyme responsible for the synthesis of eukaryotic mRNAs. Although most (but not all) Pol II promoters include the TATA box and Inr (initiator) sequences, with their standard spacing, they vary greatly in both the number and the location of additional sequences required for the regulation of transcription. These additional regulatory sequences are usually called enhancers in higher eukaryotes and upstream activator sequences (UASs) in yeast. A typical enhancer may be found hundreds or even thousands of base pairs upstream from the transcription start site, or may even be downstream, within the gene itself. When bound by the appropriate regulatory proteins, an enhancer increases transcription at nearby promoters regardless of its orientation in the DNA. The UASs of yeast function in a similar way, although generally they must be positioned upstream and within a few hundred base pairs of the transcription start site. An average Pol II promoter may be affected by a half-dozen regulatory sequences of this type, and even more complex promoters are quite common.

Successful binding of active RNA polymerase II holoenzyme at one of its promoters usually requires the action of other proteins. These are of three types:

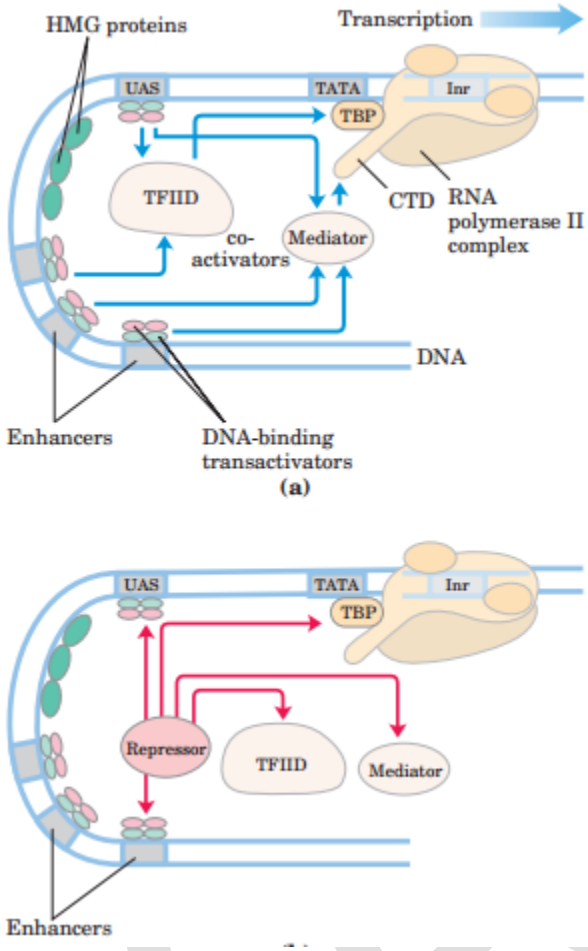


Fig: Eukaryotic promoters and regulatory proteins

(1)basal transcription factors required at every Pol II promoter;

(2) DNA-binding transactivators, which bind to enhancers or UASs and facilitate transcription; and (3) coactivators.The latter group act indirectly—not by binding to the DNA—and are required for essential communication between the DNA-binding transactivators and the complex composed of Pol II and the general transcription factors.

Furthermore, a variety of repressor proteins can interfere with communication between the RNA polymerase and the DNA-binding transactivators, resulting in repression of transcription.

TATA-Binding Protein - The first component to bind in the assembly of a preinitiation complex at the TATA box of a typical Pol II promoter is the TATA-binding protein (TBP). The complete complex includes the basal (or general) transcription factors TFIIB, TFIIE, TFIIF, TFIIH; Pol II; and perhaps TFIIA (not all of the factors are shown in Fig.). This minimal preinitiation complex, however, is often insufficient for the initiation of transcription and generally does not form at all if the promoter is obscured within chromatin. Positive regu-lation leading to transcription is imposed by the trans-activators and coactivators.

DNA-Binding Transactivators - The requirements for trans-activators vary greatly from one promoter to another. A few transactivators are known to facilitate transcription at hundreds of promoters, whereas others are specific for a few promoters. Many transactivators are sensitive to the binding of signal molecules, providing the capacity to activate or deactivate transcription in response to a changing cellular environment. Some enhancers bound by DNA-binding transactivators are quite distant from the promoter's TATA box. How do the transactivators function at a distance? The answer in most cases seems to be that, as indicated earlier, the intervening DNA is looped so that the various protein complexes can interact directly. The looping is promoted by certain non-histone proteins that are abundant in chromatin and bind nonspecifically to DNA. This high mobility group (HMG) proteins (Fig "high mobility" refers to their electrophoretic mobility in polyacrylamide gels) play an important structural role in chromatin remodeling and transcriptional activation.

Coactivator Protein Complexes - Most transcription requires the presence of additional protein complexes. Some major regulatory protein complexes that interact with Pol II have been defined both genetically and biochemically. These coactivator complexes act as intermediaries between the DNA-binding transactivators and the Pol II complex.

The best-characterized coactivator is the transcription factor TFIID. In eukaryotes, TFIID is a large complex that includes TBP and ten or more TBP-associated factors (TAFs). Some TAFs resemble histones and may play a role in displacing nucleosomes during the activation of transcription. Many DNA-binding transactivators aid in transcription initiation by interacting with one or more TAFs. The requirement for TAFs to initiate transcription can vary greatly from one gene to another. Some promoters require TFIID, some do not, and some require only subsets of the TFIID TAF subunits.

Another important coactivator consists of 20 or more polypeptides in a protein complex called mediator(Fig); the 20 core polypeptides are highly conserved from fungi to humans. Mediator

binds tightly to the carboxyl-terminal domain (CTD) of the largest subunit of Pol II. The mediator complex is required for both basal and regulated transcription at promoters used by Pol II, and it also stimulates the phosphorylation of the CTD by TFIIH. Both mediator and TFIID are required at some promoters. As with TFIID, some DNA-binding transactivators interact with one or more components of the mediator complex. Coactivator complexes function at or near the promoter's TATA box.

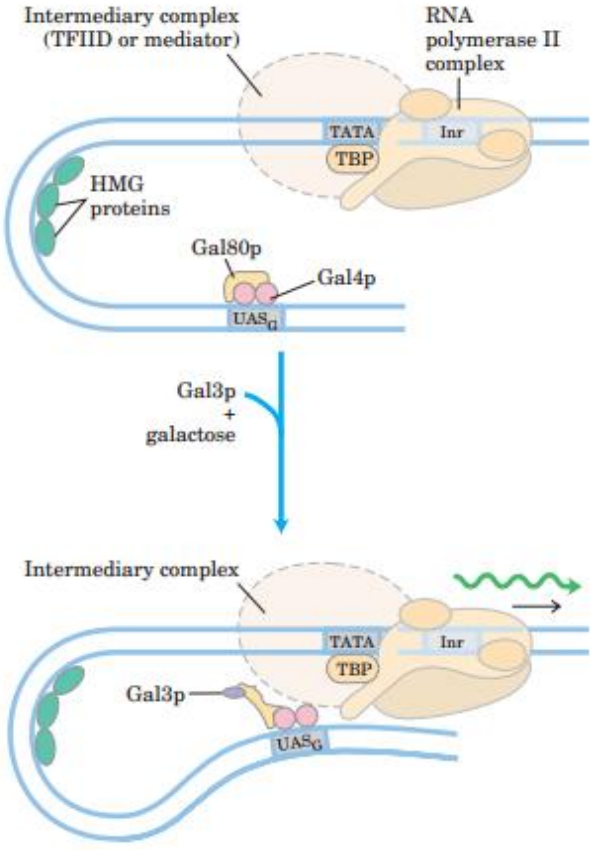
Choreography of Transcriptional Activation - We can now begin to piece together the sequence of transcriptional activation events at a typical Pol II promoter. First, crucial remodeling of the chromatin takes place in stages. Some DNA-binding transactivators have significant affinity for their binding sites even when the sites are within condensed chromatin. Binding of one transacti-vator may facilitate the binding of others, gradually dis-placing some nucleosomes. The bound transactivators can then interact directly with HATs or enzyme complexes such as SWI/SNF (or both), accelerating the remodeling of the surrounding chromatin. In this way a bound transactivator can draw in other components necessary for further chromatin remodeling to permit transcription of specific genes. The bound transactivators, generally acting through complexes such as TFIID or mediator (or both), stabilize the binding of Pol II and its associated transcription factors and greatly facilitate formation of the preinitiation transcription complex. Complexity in these regulatory circuits is the rule rather than the exception, with multiple DNA-bound transactivators promoting transcription.

The script can change from one promoter to another, but most promoters seem to require a precisely ordered assembly of components to initiate transcription. The assembly process is not always fast. At some genes it may take minutes; at certain genes in higher eukaryotes the process can take days.

Reversible Transcriptional Activation - Although rarer, some eukaryotic regulatory proteins that bind to Pol II promoters can act as repressors, inhibiting the formation of active preinitiation complexes. Some transactivators can adopt different conformations, enabling them to serve as transcriptional activators or repressors. For example, some steroid hormone receptors function in the nucleus as DNA-binding transactivators, stimulating the transcription of certain genes when a particular steroid hormone signal is present. When the hormone is absent, the receptor proteins revert to a repressor conformation, prevent-ing the formation of preinitiation complexes. In some cases this repression involves interaction with histone deacetylases and other proteins that help restore the surrounding chromatin to its transcriptionally inactive state.

Both Positive and Negative Regulation Of The Genes of Galactose Metabolism in Yeast

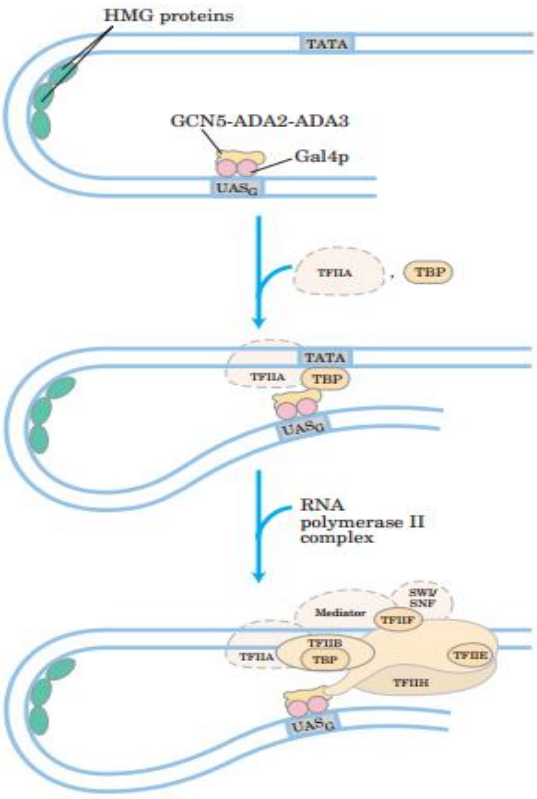
The enzymes required for the importation and metabolism of galactose in yeast are encoded by genes scattered over several chromosomes. Each of theGALgenes is transcribed separately, and yeast cells have no operons like those in bacteria. However, all the GALgenes have similar pro-moters and are regulated coordinately by a common set of proteins. The promoters for the GALgenes consist of the TATA box and Inr sequences, as well as an up-stream activator sequence (UASG) recognized by a DNA-binding transcriptional activator known as Gal4 protein (Gal4p). Regulation of gene expression by galac-tose entails an interplay between Gal4p and two other proteins, Gal80p and Gal3p (Fig. 28–28). Gal80p forms a complex with Gal4p, preventing Gal4p



from functioning as an activator of the GALpromoters. When galactose is present, it binds Gal3p, which then interacts with Gal80p, allowing Gal4p to function as an activator at the various GALpromoters.

Other protein complexes also have a role in activating transcription of the GALgenes. These may include the SAGA complex for histone acetylation, the SWI/SNF complex for nucleosome remodeling, and the mediator complex. The following figure provides an idea of the complexity of protein interactions in the overall process of transcriptional activation in eukaryotic cells.

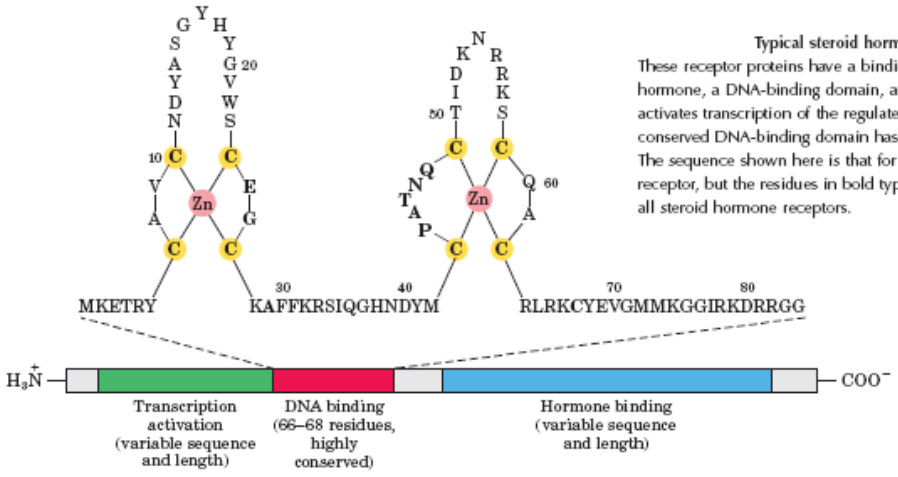
Glucose is the preferred carbon source for yeast, as it is for bacteria. When glucose is present, most of the GALgenes are repressed—whether galactose is present or not. The GALregulatory system described above is effectively overridden by a complex catabolite repression system that includes several proteins (not depicted in the fig).



Regulation of gene expression by intracellular and intercellular signal

The effects of steroid hormones (and of thyroid and retinoid hormones, which have the same mode of action) provide additional well studied examples of the modulation of eukaryotic regulatory proteins by direct interaction with molecular signals. Unlike other types of hormones, steroid hormones do not have to bind to plasma membrane receptors. Instead, they can interact with intracellular receptors that are themselves transcriptional transactivators. Steroid hormones too hydrophobic to dissolve readily in the blood (estrogen, progesterone, and cortisol, for example) travel on specific carrier proteins from their point of release to their target tissues. In

the target tissue, the hormone passes through the plasma membrane by simple diffusion and binds to its specific receptor protein in the nucleus. The hormone-receptor complex acts by binding to highly specific DNA sequences called **hormone response elements (HREs),** thereby altering gene expression. Hormone binding triggers changes in the conformation of the receptor proteins so that they become capable of interacting with additional transcription factors. The bound hormone-receptor complex can either enhance or suppress the expression of adjacent genes. The DNA sequences (HREs) to which hormonereceptor complexes bind are similar in length and arrangement, but differ in sequence, for the various steroid hormones. Each receptor has a consensus HRE sequence to which the hormone-receptor complex binds well, with each consensus consisting of two six-nucleotide sequences, either contiguous or separated by three nucleotides, in tandem or in a palindromic arrangement. The hormone receptors have a highly conserved DNA-binding domain with two zinc fingers.



Typical steroid hormone receptors. These receptor proteins have a binding site for the hormone, a DNA-binding domain, and a region that activates transcription of the regulated gene. The highly conserved DNA-binding domain has two zinc fingers. The sequence shown here is that for the estrogen receptor, but the residues in bold type are common to all steroid hormone receptors.

The hormone-receptor complex binds to the DNA as a dimer, with the zinc finger domains of each monomer recognizing one of the six-nucleotide sequences. The ability of a given hormone to act through the hormone-receptor complex to alter the expression of a specific gene depends on the exact sequence of the HRE, its position relative to the gene, and the number of HREs associated with the gene.

Unlike the DNA-binding domain, the ligand-binding region of the receptor protein—always at the carboxyl terminus—is quite specific to the particular receptor. In the ligand-binding region, the glucocorticoid receptor is only 30% similar to the estrogen receptor and 17% similar to the thyroid hormone receptor. The size of the ligand- binding region varies dramatically; in the vitamin D receptor it has only 25 amino acid residues, whereas in the mineralocorticoid receptor it has 603 residues. Mutations that change one amino acid in these regions can result in loss of

responsiveness to a specific hormone. Some humans unable to respond to cortisol, testosterone, vitamin D, or thyroxine have mutations of this type.

Regulation Can Result from Phosphorylation of Nuclear Transcription Factors

We noted in Chapter 12 that the effects of insulin on gene expression are mediated by a series of steps leading ultimately to the activation of a protein kinase in the nucleus that phosphorylates specific DNA-binding proteins and thereby alters their ability to act as transcription factors (see Fig. 12–6). This general mechanism mediates the effects of many nonsteroid hormones. For example, the _-adrenergic pathway that leads to elevated

levels of cytosolic cAMP, which acts as a second messenger in eukaryotes as well as in prokaryotes (see Figs 12–12, 28–18), also affects the transcription of a set of genes, each of which is located near a specific DNA sequence called a cAMP response element (CRE). The catalytic subunit of protein kinase A, released when cAMP levels rise (see Fig. 12–15), enters the nucleus and phosphorylates a nuclear protein, the CRE-binding protein (CREB). When phosphorylated, CREB binds to CREs near certain genes and acts as a transcription factor, turning on the expression of these genes.

RNAi.

Posttranscriptional Gene Silencing Is Mediated by RNA Interference

In higher eukaryotes, including nematodes, fruit flies, plants, and mammals, a class of small RNAs has been discovered that mediates the silencing of particular genes. The RNAs function by interacting with mRNAs, often in the 3UTR, resulting in either mRNA degrada-tion or translation inhibition. In either case, the mRNA, and thus the gene that produces it, is silenced. This form of gene regulation controls developmental timing in at least some organisms. It is also used as a mechanism to protect against invading RNA viruses (particularly important in plants, which lack an immune system) and to control the activity of transposons. In addition, small RNA molecules may play a critical (but still undefined) role in the formation of heterochromatin. The small RNAs are sometimes called micro-RNAs (miRNAs). Many are present only transiently during development, and these are sometimes referred to as small temporal RNAs (stRNAs). Hundreds of different miRNAs have been identified in higher eukaryotes. They are transcribed as precursor RNAs about 70 nucleotides long, with internally complementary sequences that form hairpinlike structures. The precursors are cleaved by endonucleases to form short duplexes about 20 to 25 nucleotides long. The best-characterized nuclease goes by the delightfully suggestive name Dicer; endonucleases in the Dicer family are widely distributed in higher

eukaryotes. One strand of the processed miRNA is transferred to the target mRNA (or to a viral or transposon RNA), leading to inhibition of translation or degradation of the RNA. This gene regulation mechanism has an interesting and very useful practical side. If an investigator intro-duces into an organism a duplex RNA molecule corre-sponding in sequence to virtually any mRNA, the Dicer endonuclease cleaves the duplex into short segments, called small interfering RNAs (siRNAs). These bind to the mRNA and silence it. The process is known as RNA interference(RNAi).In plants, virtu-ally any gene can be effectively shut down in this way.

In nematodes, simply introducing the duplex RNA into the worm's diet produces very effective suppression of the target gene. The technique has rapidly become an important tool in the ongoing efforts to study gene func-tion, because it can disrupt gene function without cre-ating a mutant organism. The procedure can be applied to humans as well. Laboratory-produced siRNAs have already been used to block HIV and poliovirus infections in cultured human cells for a week or so at a time. Al-though this work is in its infancy, the rapid progress makes RNA interference a field to watch for future med-ical advances.

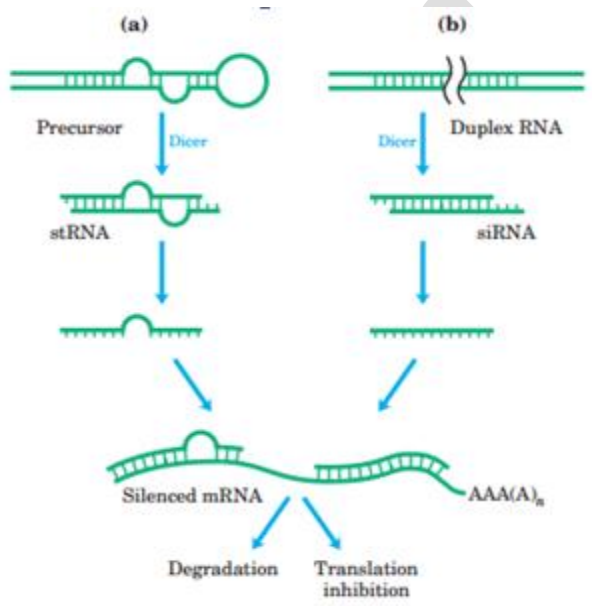


Figure Gene Silencing by RNA interference. (a) Small temporal RNAs (SiRNAs) are generated by Dicer-mediated cleavage of longer precursors that fold to create duplex regions. The SiRNAs then bind to mRNAs leading to degradation of mRNA or inhibition of translation. (b) Double-stranded RNAs can be constructed and introduced into a cell. Dicer processes the duplex RNAs into small interfering RNAs, which interact with the target mRNA. Again, the mRNA is either degraded or its translation inhibited.

POSSIBLE QUESTIONS

UNIT V

Part-B

1. What is Rec A protein ? Explain its role in various process
2. Explain the Holliday model of recombination
3. Narrate the mechanism in Messelson and Radding model of recombination
4. Write about site specific recombination?
5. What are all the outcome of recombination?
6. What is antisense RNA ? How will you prepare it?
7. Explain the applications of antisense RNA technology in science
8. What is attenuation? Explain with reference to try operon.
9. What is operon? Briefly explain its positive and negative regulation
10. Explain the control mechanism of Lac operon.

Part-C

1. Explain the Holliday model of recombination in detail
2. With a neat diagram discuss the events in site specific recombination
3. Explain the positive and negative regulation of lac operon
4. Explain the following
   a)Ara operon

b)Attenuation

5. Discuss the techniques and applications of antisense RNA technology

Karpagam acadmy of higher educaton
Department of biochemistry
Molecular biology (19BCP202) MCQ
UNIT V

S.NO	UNIT	QUESTION	OPTION 1	OPTION 2	OPTION 3	OPTION 4	ANSWER
1	V	The role of Rec A gene in homologous recombination is	**DNA pairing**	Nicking of strands	To provide energy	DNA synthesis	**DNA pairing**
2	V	Rec A protein has an important function in	SOS repair	Recombination	**SOS repair and Recombination**	Base excision repair	**SOS repair and Recombination**
3	V	Establishment of lysogeny in phage lambda is triggered by	**gpc II**	Cro protein	gpb I	gpb II	**gpc II**
4	V	The minus strand of of θx174 is called	sense strand	**antisense strand**	missense strand	nonsense strand	**antisense strand**
5	V	Rec BCD initiates recombination by making	**ss nicks**	ds nicks	exchange process	mispairing	**ss nicks**
6	V	The Holliday model provides a molecular basis for	**the association between aberrant segregation and crossing over**	the activation of transcription	the repression of early transcription	DNA repair	**the association between aberrant segregation and crossing over**
7	V	Rec BCD protein is a multifunctional protein with an	**endonuclease activity**	exonuclease activity	polymerase activity	ligase activity	**endonuclease activity**
8	V	Rec A performs all of the following except	it coats the ss region and polymerizes	it catalyses the *in vitro* formation of Holliday structures	It participates in DNA repair by regulating SOS response	**It is involved in the DNA replication**	**It is involved in the DNA replication**
9	V	Rec BCD is also known as	Exonuclease N	Exonuclease B	**Exonuclease** V	Exonuclease D	**Exonuclease** V
10	V	Rotation of the Holliday structure at the crossover site forms a	**rotational isomer**	Circular isomer	Crosslinking isomer	exchange isomer	**rotational isomer**
11	V		The type of mutation caused by nitrous acid in DNA is	**Transition**	Transversion	Insertion	Deletion
12	V	**Transition**	Rotational isomer is also called as	**Holliday structure**	Radding structure	Circular structure	Rotational structure
13	V	**Holliday structure**	Who gave basis of most popular current models for molecular events of recombination?	**Robin Holliday**	Matthew Messelson and Charles Radding	Arthur Kornberg	Messelson Stahl
14	V	Who suggested mechanism for creating cross-strand Holliday structure?	Matthew Meselson and charles Radding	**Robin Holliday**	Arthur Kornberg	Messelson Stahl	**Robin Holliday**
16	V	Hydroxylamine causes	**GC base pairing to AT**	AT base pairing to GC	GC base pairing to TA	AT base pairing to CG	**GC base pairing to AT**
17	V	A base change that does not change the purine – pyrimidine orientation is called	Transversion	Transformation	**Transition**	Translation	**Transition**
18	V	Cyclobutane pyrimidine dimers can be monomerized again by _____ in the presence of visible light	Exonuclease	**DNA photolyases**	Transposase	DNA polymerase	**DNA photolyases**
19	V	Mutational hot spots are sites containing	MeG	MeA	**MeC**	MeB	**MeC**
20	V	In base excision repair, the lesion is removed by	**DNA glycosylase**	Excisionase	Transposase	DNA polymerase	**DNA glycosylase**
21	V	Which of the following leads to frame shift mutation	UV light	Alkylation	**Intercalation**	Spontaneous	**Intercalation**
22	V	Which of the following is an mutagen?	**EMS**	EES	HCl	APS	**EMS**
23	V	Which of the following mutation leads to the death of cell?	Frameshift	Missense	**Lethal**	Point	**Lethal**
24	V	The type of mutation caused by nitrous acid in DNA is	**Transition**	Transversion	Insertion	Deletion	**Transition**
25	V	The enzyme primarily involved in DNA repair in bacteria is	**DNA pol I**	DNA pol II	DNA pol III	DNA pol α	**DNA pol I**
26	V	E. Coli DNA ligase utilizes	**NAD+**	ATP	GTP	GDP	**NAD+**
27	V	Altered bases in DNA are removed by	Photolyase	**Glycosylase**	Methylase	UvrABC endonuclease	**Glycosylase**
28	V	DNA helicase enzyme involved in base excision repair mechanism is	DNA helicase I	DNA helicase II	**DNA helicase III**	DNA helicase IV	**DNA helicase III**
29	V	Rec A protein can bind to	**ssDNA**	ds DNA	ssDNA and dsDNA	RNA	**ssDNA**
30	V	Nucleotide excision repair is carried by	**Uvr ABC endonuclease**	AP endonuclease	Methyl transferase	Glycosylase	**Uvr ABC endonuclease**
31	V	Rec A protein has an important function in	SOS repair	Recombination	**SOS repair and Recombination**	Replication	**SOS repair and Recombination**
32	V	The enzyme of Escherichia coli is a nuclease that initiates the repair of double-stranded DNA breaks by homologous recombination.	RNA polymerase	DNA polymerase	DNA ligase	**RecBCD**	**RecBCD**
33	V	The type of point mutation which detected only after nucleic acid sequencing is known as	**silent mutation**	missense mutation	non-sense mutation	point mutation	**silent mutation**
34	V	Daughter strand repair is the other name for	SOS repair	excision repair	**recombination repair**	photoreactivation	**recombination repair**
35	V	Mutation in which there is an amino acid substitution is called	point mutation	nonsense mutation	**missense mutation**	silent mutation	**missense mutation**
36	V	Post dimer initiation is responsible for	**recombination repair**	sos repair	excision repair	photoreactivation	**recombination repair**
37	V	The helicase active in E. coli DNA	**Rep protein**	Polymerase I	Polymerase II	Polymerase III	**Rep protein**
38	V	Which of the following enzyme selectively degrades single stranded nucleic acid?	DNA ligase	DNA polymerase	Ribonuclease	**S1 nuclease**	**S1 nuclease**
39	V	5-bromo uracil damages DNA. It is an analog of	Uracil	**Thymine**	Cytosine	Guanine	**Thymine**

40	V	DNA repair mechanism is absent in	Nuclear genome	Mitochondrial genome	Chloroplast genome	**Mitochondrial and chloroplast genome**	**Mitochondrial and chloroplast genome**
41	V	umu C, umu D gene family and Rec A proteins are involved in	BER	NER	**SOS repair**	Recombinational repair	**SOS repair**
42	V	Which of the following molecule has self-repair mechanisms?	DNA, RNA and protein	DNA and RNA	**DNA only**	DNA and protein	**DNA only**
43	V	Which DNA polymerase is involved in BER?	DNA polymerase α	**DNA polymerase β**	DNA polymerase θ	DNA polymerase £	**DNA polymerase β**
44	V	Which of the following is a base repair system?	BER	NER	**SOS**	Recombinational repair	**SOS**
45	V	eRF1 is the release factor in eukaryotes that requires	ATP for its binding to ribosome	**GTP for its binding to ribosome**	ATP and GTP for its binding to ribosome	Mn^{2+} for its binding to ribosome	**GTP for its binding to ribosome**
46	V	The role of Rec A gene in homologous recombination is	**DNA pairing**	Nicking of strands	To provide energy	DNA synthesis	**DNA pairing**
47	V	Rec A protein has an important function in	SOS repair	Recombination	**SOS repair and Recombination**	Base excision repair	**SOS repair and Recombination**
48	V	Establishment of lysogeny in phage lambda is triggered by	**gpc II**	Cro protein	gpb I	gpb II	**gpc II**
49	V	The minus strand of of θx174 is called	sense strand	**antisense strand**	missense strand	nonsense strand	**antisense strand**
50	V	Rec BCD initiates recombination by making	**ss nicks**	ds nicks	exchange process	mispairing	**ss nicks**
51	V	Outcomes of point mutations include	missense mutation	nonsense mutation	silent mutation	**all of above**	**all of above**
52	V	When a single nucleotide base pair is substituted, type of mutation occurs is	random mutation	**point mutation**	inversion	mispairing	**point mutation**
53	V	Type of RNA which carries genetic information copied from DNA in form of series of three base codes is	**mRNA**	tRNA	rRNA	all of above	**mRNA**
54	V	Adding a phosphate group to proteins is called	replication	hydrogenation	**phosphorylation**	carbation	**phosphorylation**
55	V	Small subunit of ribosome binds to 5' end of	**mRNA**	tRNA	rRNA	DNA	**mRNA**
56	V	Ribosomes are made up of	RNA	rRNA	proteins	**both B and C**	**both B and C**
57	V	A phosphate group is added by protein enzyme called	**kinase**	helicase	gyrase	polymerase	**kinase**
58	V	Actual genetic code used by cells is	double code	**triplet code**	single code	tetra code	**triplet code**
59	V	hosphatase is a protein enzyme that removes	sulphate group	amino group	**phosphate group**	hydroxyl group	**phosphate group**
60	V	To begin translation initiation complex include	ribosomes	mRNA	tRNA	**all of above**	**all of above**

19BCP202 MOLECULAR BIOLOGY 4H-4C

Instruction hours per week:L: 4 T: 0 P:0 **Marks:** Internal**:40**External**:60**Total**: 100**

End Semester Exam: 3 hours

Course objectives

The course aims to provide students with a basic understanding of

- Organization of DNA in a genome and the mechanism behind replication, transcription and translation.
- Regulation of gene expression in prokaryotes and Eukaryotes.

Course outcome

At the end of the course, student will be able to

- Understand the structure of nucleic acids and the DNA replication process
- Learn about the process of transcription
- Understand the mechanism of translation
- Learn about gene regulation in prokaryotes
- Learn about gene regulation in eukaryotes

UNIT I: Molecular structure of genes

Molecular definition of gene, chromosomal organization of genes and non-coding DNA, protein coding genes, tandomly repeated genes, single sequence DNA. Structural organization of eukaryotic chromosomes- histone proteins, chromatin, functional elements. Mobile DNA elements- bacterial IS elements, transposons, viral transposons and non- viral transposons. Mutation- types.

UNIT II: DNA replication andrepair

General features of chromosomal replication. Enzymology of DNA replication, DNA replication machinery. Replication in prokaryotes and eukaryotes- Initiation, elongation and termination. DNA damage-types. Repair mechanism of DNA damage-all types.

UNIT III: Transcription

Prokaryotic gene transcription- Initiation, elongation and termination. Eukaryotic gene transcription- transcription unit, RNA polymerases- types, Transcription and processing of mRNA, tRNA and rRNA. Regulatory sequences in protein coding genes-TATA box, initiators, CpG island, promoter-proximal element, activators and repressors of transcription, Multiple transcription control elements. Regulation of transcription factor activity by lipid-soluble hormones.

UNIT IV: Translation

Deciphering genetic code, features. Wobble hypothesis. Initiation, elongation and termination of prokaryotic and eukaryotic translation. Fidelity of translation. Post translational modifications-all types; Protein targeting-Targeting protein to nucleus, ER, Golgi complex. Protein degradation-ubiquitin mediated degradation.

UNIT V: Prokaryotic gene regulation

Operon model, Lac, trp and ara operons.Regulatory proteins-DNA binding domain, protein-protein interaction domain. Recombination- holiday model, Rec BCD enzymes, Rec A protein, MesselsonRadding model, site- specific recombination. Antisense RNA technology.

Eukaryotic gene regulation: Transcriptionally active chromatin, chromatin remodeling, DNA binding transactivators and coactivators. Regulation of gene expression by intracellular and intercellular signal, RNAi.

SUGGESTED READINGS

1. Watson, J. D., Hopkins, N. H., Roberts, J. W., Steitz, J. A., and Weiner, A. M., (2005)Molecular biology of the gene, The Benjamin/Cummings publishingcompanies,Inc,California.

2. Lewin, B., (2008) GenesIX,Oxford University Press, 9th Edition, Oxford, London,

3. Weaver, R. F., (2008) Molecular biology, WCB McGraw-Hill companies, 6thEdition.Inc,NewYork.

4. Lodish, H., Berk, A., Kaiser, C.A., and Krieger, M.,. (2012). MolecularCell Biology, 7th edition. W.H. Freeman & Company,

5. Lehninger, L., Nelson, D.L., and Cox, M.M., (2012). Principles of Biochemistry, WHFreeman and Company, 6th Edition, New York.

6. Kornberg, A., Baker, A., (2005). DNA replication, W.H. Freeman and Co, USA.

7. Cooper, G.M., and Hausman, R.E.,(2013).Cell-A Molecular Approach, 6thEdition. Sinauer Associates. USA

KARPAGAM ACADEMY OF HIGHER EDUCATION

(Deemed to be University)
(Established Under Section 3 of UGC Act 1956)
Coimbatore - 641021.
(For the candidates admitted from 2018 onwards)
DEPARTMENT OF BIOCHEMISTRTY

SUBJECT : MOLECULAR BIOLOGY

SEMESTER : II

SUBJECT CODE: 19BCP202 **CLASS : I M. Sc. BC**

LECTURE PLAN
DEPARTMENT OF BIOCHEMISTRTY

S.No	Lecture Duration Hour	Topics to be Covered	Support Material/Page Nos
		UNIT-I	
1	1	Molecular definition of gene, chromosomal organization of genes and non-coding DNA,	T1: 406-408
2	1	Protein coding genes	T2:409-411
3	1	Tandomly repeated genes,	T1: 411-412
4	1	Single sequence DNA.	T1: 413
5	1	Structural organization of eukaryotic chromosomes- histone proteins,	T1:424-425
6	1	Chromatin, functional elements.	T1:425-428
7	1	Mobile DNA elements- bacterial IS elements, Transposons, viral transposons and non- viral transposons.	T1:414-423
8	1	Mutation- types.	T2: 91-94
	Total No Of Hours Planned For Unit 1=8		
		UNIT-II	
1	1	General features of chromosomal replication.	T2:73-75
2	1	Enzymology of DNA replication.	T2:75-78
3	1	DNA replication machinery.	T3:78-82
4	1	Replication in prokaryotes - Initiation, elongation and termination.	T1:82-86

5	1	Replication in eukaryotes- Initiation, elongation and termination.	T2:86-90
6	1	DNA damage-types. Repair mechanism of DNA damage-all types.	T2:95-100
	Total No Of Hours Planned For Unit II=06		
		UNIT-III	
1	1	Prokaryotic gene transcription- Initiation, elongation and termination.	T2:185-193
2	1	Eukaryotic gene transcription- transcription unit	T3:241-244
3	1	RNA polymerases- types, Transcription and processing of mRNA, tRNA and rRNA.	T1:211-217
4	1	Regulatory sequences in protein coding genes- TATA box, initiators, CpG island	T2:217-221
5	1	Promoter-proximal element, activators and repressors of transcription,	T2:221-223
6	1	Multiple transcription control elements.	T1:486-489
7	1	Regulation of transcription factor activity by lipid-soluble hormones.	T1:481-484
	Total No Of Hours Planned For Unit III=07		
		UNIT-IV	
1	1	Deciphering genetic code, features.	T4:1065-1066
2	1	Wobble hypothesis.	T4:1066-1072
3	1	Initiation, elongation and termination of prokaryotic translation.	T2:273-278
4	1	Initiation, elongation and termination of eukaryotic translation.	T2:279-282
5	1	Fidelity of translation.	T4:1095-1098
6	1	Post translational modifications-all types	T3:283-286
7	1	Protein targeting-Targeting protein to nucleus.	T2:286-287
8	1	Protein targeting- ER.	T4:1100-1101
9	1	Protein degradation- Golgi complex.	T4:1101-
10	1	Protein degradation- ubiquitin mediated degradation.	T1: 71-72
	Total No Of Hours Planned For Unit IV=10		
		UNIT-V	

1	1	Operon model – Lac operon	T5:1294-1296
2		Operon model – trp and ara operons.	T5:1291-1301
3	1	Regulatory proteins-DNA binding domain, protein- protein interaction domain.	T5:1288-1290
4	1	Recombination- holiday model, Rec BCD enzymes, Rec A protein, Messelson Radding model, site- specific recombination.	T5:1226-1252
5	1	Antisense RNA technology	T5:1323-1325
6	1	Transcriptionally active chromatin, chromatin remodeling,	T1:476-478
7	1	DNA binding transactivators and coactivators.	T1:478-481
8	1	Regulation of gene expression by intracellular and intercellular signal,	T1:481-486
9	1	RNAi.	T1:518-520
	Total no of Hours Planned for unit V=09		
Total Planned Hours		**40**	

References:

T1: Lodish, H., Berk, A., Kaiser, C.A., and Krieger, M., (2012). Molecular Cell Biology, 7th edition. W.H. Freeman & Company,

T2: Turner, P., McLennan, A., Bates, A., and White, M., (2006) Instant Notes in Molecular Biology 3nd Edition. Publisher: Taylor and Francis Group.

T3: Lewin, B., (2008) Genes IX, Oxford University Press, 9th Edition, Oxford, London.

T4: Lehninger, L., Nelson, D.L., and Cox, M.M., (2012). Principles of Biochemistry, WH Freeman and Company, 6th Edition, New York.

T5: Donald Voet and Judith Voet; 2012, Biochemistry, 4th Edition, John Wiley and Sons. Inc

Signature of the Staff

UNIT-I

SYLLABUS

Molecular Structure of genes - Molecular definition of gene, chromosomal organization of genes and non-coding DNA, protein coding genes, tandomly repeated genes, single sequence DNA. Structural organization of eukaryotic chromosomes- histone proteins, chromatin, functional elements. Mobile DNA elements- bacterial IS elements, transposons, viral transposons and non- viral transposons. Mutation- types

Molecular Definition of Gene:

In molecular terms, a gene commonly is defined as the entire nucleic acid sequence that is necessary for the synthesis of a functional gene product (polypeptide or RNA). According to this definition, a gene includes more than the nucleotides encoding the amino acid sequence of a protein, referred to as the coding region. A gene also includes all the DNA sequences required for synthesis of a particular RNA transcript. In eukaryotic genes, transcription-control regions known as **enhancers** can lie 50 kb or more from the coding region. Other critical noncoding regions in eukaryotic genes are the sequences that specify 3' cleavage and polyadenylation, known as poly(A) sites, and splicing of primary RNA transcripts, known as splice sites. Mutations in these RNA-processing signals prevent expression of a functional mRNA and thus of the encoded polypeptide.

Although most genes are transcribed into mRNAs, which encode proteins, clearly some DNA sequences are transcribed into RNAs that do not encode proteins (e.g., tRNAs and rRNAs). However, because the DNA that encodes tRNAs and rRNAs can cause specific phenotypes when it is mutated, these DNA regions generally are referred to as tRNA and rRNAgenes, even though the final products of these genes are RNA molecules and not proteins. Many other RNA molecules are also transcribed from non-protein-coding genes.

Monocistronic mRNAs Produced by Eukaryotic Genes and their Introns

Many bacterial mRNAs are polycistronic; that is, a single mRNA molecule (e.g., the mRNA encoded by the trpoperon) includes the coding region for several proteins that function

together in a biological process. In contrast, most eukaryotic mRNAs are monocistronic; that is, each mRNA molecule encodes a single protein.This difference between polycistronic and monocistronic mRNAs correlates with a fundamental difference in their translation.

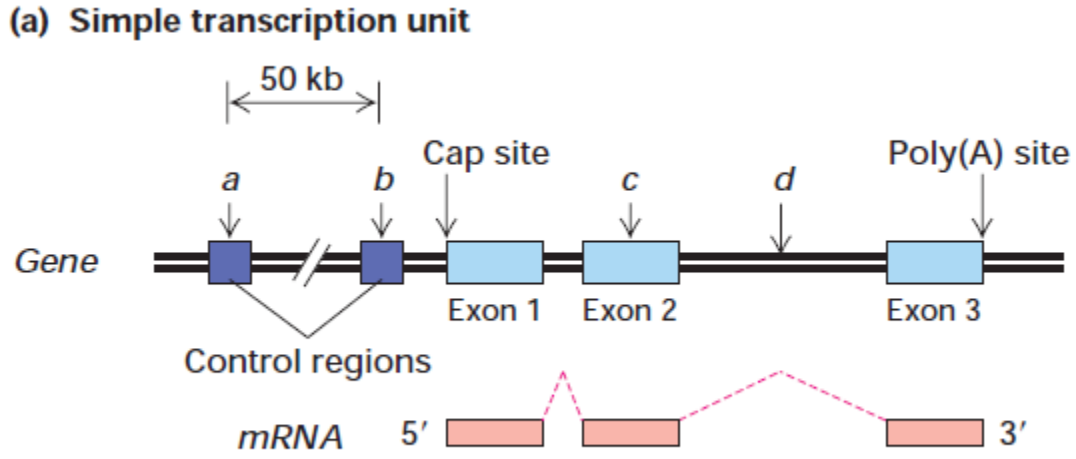
Within a bacterial polycistronic mRNA a ribosome binding site is located near the start site for each of the protein coding regions, or cistrons, in the mRNA. Translation initiation can begin at any of these multiple internal sites, producing multiple proteins. In most eukaryotic mRNAs, however, the 5'-cap structure directs ribosome binding, and translation begins at the closest AUG start codon. As a result, translation begins only at this site. In many cases, the primary transcripts of eukaryotic protein-coding genes are processed into a single type of mRNA, which is translated to give a single type of polypeptide.

Unlike bacterial and yeast genes, which generally lack introns, most genes in multicellular animals and plants contain introns, which are removed during RNA processing. In many cases, the introns in a gene are considerably longer than the exons. For instance, of the ≈50,000 base pairs composing many human genes encoding average-size proteins, more than 95 percent are present in introns and noncoding 5' and 3' regions. Many large proteins in higher organisms have repeated domains and are encoded by genes consisting of repeats of similar exons separated by introns of variable length. An example of this is fibronectin, a component of the extracellular matrix that is encoded by a gene containing multiple copies of three types of exons.

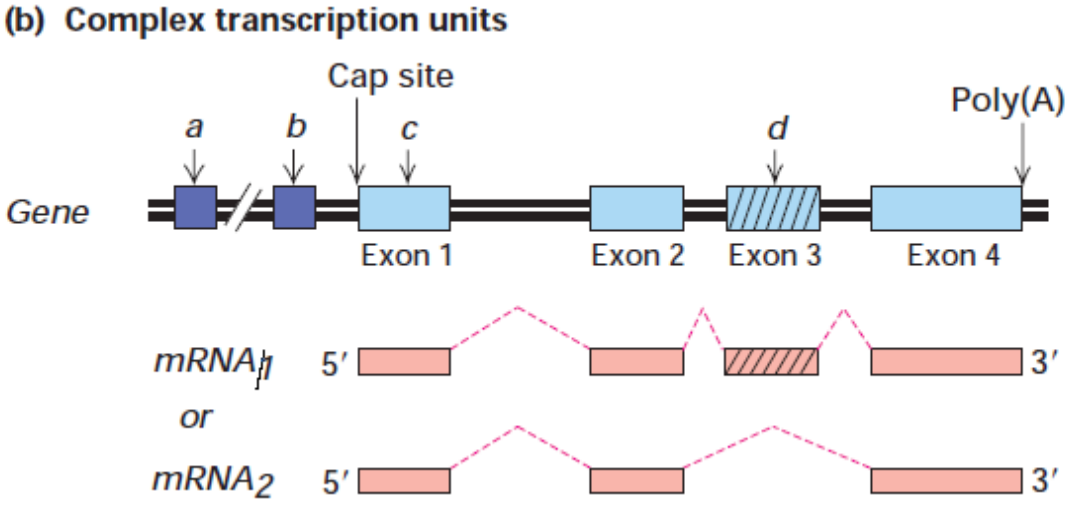
Simple and Complex Transcription Units in Eukaryotic Genomes

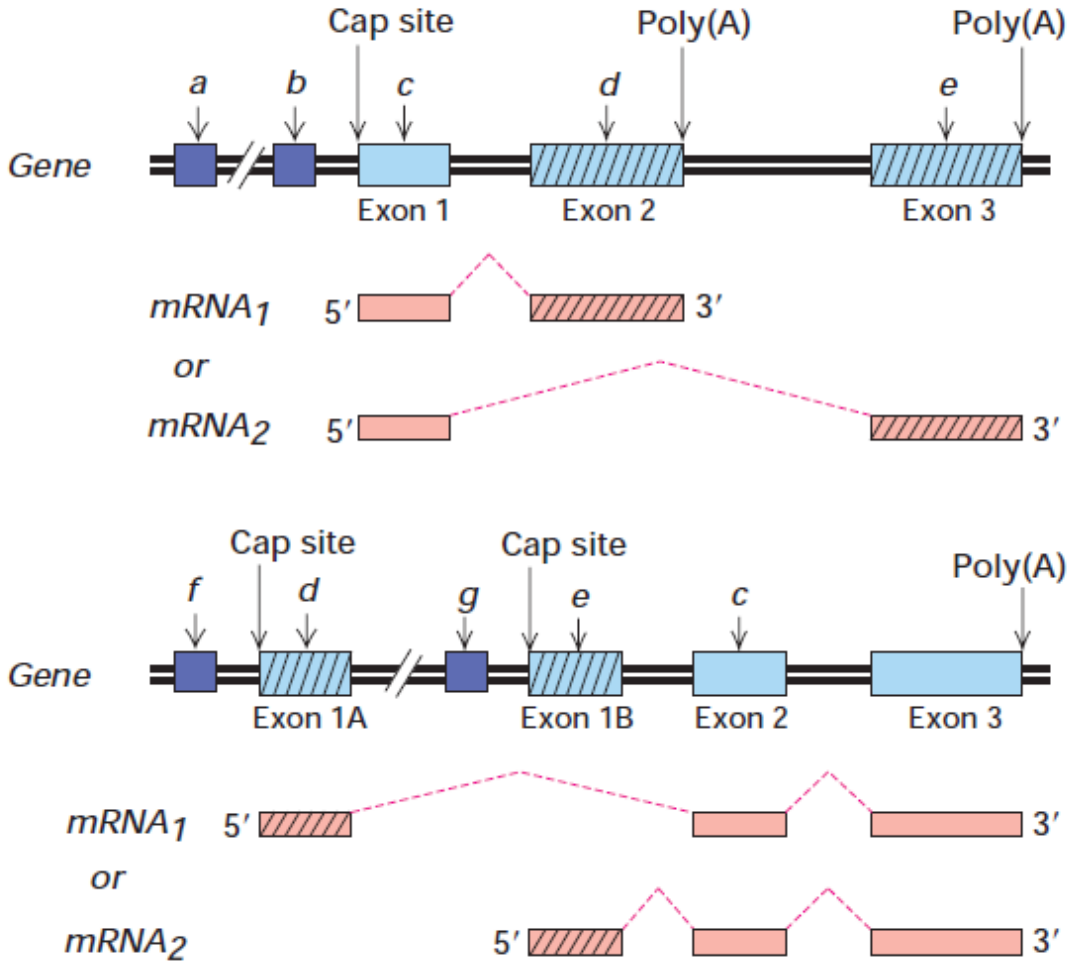
The cluster of genes that form a bacterial operon comprises a single **transcription unit,** which is transcribed from a particular promoter into a single primary transcript. In other words, genes and transcription units often are distinguishable in prokaryotes. In contrast, most eukaryotic genes and transcription units generally are identical, and the two terms commonly are used interchangeably. Eukaryotic transcription units, however, are classified into two types, depending on the fate of the primary transcript. The primary transcript produced from a simple transcription unit is processed to yield a single type of mRNA, encoding a single protein. Mutations in exons, introns, and transcription-control regions all may influence expression

of the protein encoded by a simple transcription unit (Figure 10-2a).



In the case of complex transcription units, which are quite common in multicellular organisms, the primary RNA transcript can be processed in more than one way, leading to formation of mRNAs containing different exons. Each mRNA, however, is monocistronic, being translated into a single polypeptide, with translation usually initiating at the first AUG in the mRNA. Multiple mRNAs can arise from a primary transcript in three ways (Figure 10-2b):





1. Use of different splice sites, producing mRNAs with the same 5' and 3' exons but different internal exons. Figure 10-2b (top) shows one example of this type of alternative RNA processing, exon skipping.

2. Use of alternative poly(A) sites, producing mRNAs that share the same 5' exons but have different 3' exons (Figure 10-2b [middle]).

3. Use of alternative promoters, producing mRNAs that have different 5' exons and common 3' exons. A gene expressed selectively in two or more types of cells is often transcribed from distinct cell-type-specific promoters (Figure 10-2b [bottom]).

Examples of all three types of alternative RNA processing occur during sexual differentiation in Drosophila. Commonly, one mRNA is produced from a complex transcription

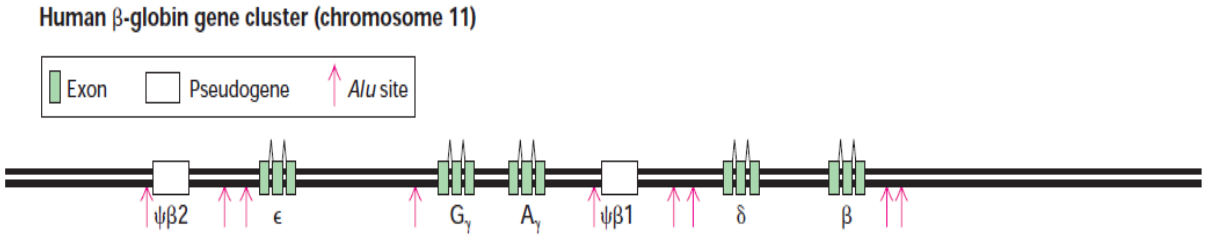
unit in some cell types, and an alternative mRNA is made in other cell types. For example, differences in RNA splicing of the primary fibronectin transcript in fibroblasts and hepatocytes determines whether or not the secreted protein includes domains that adhere to cell surfaces. The relationship between a mutation and a gene is not always straightforward when it comes to complex transcription units. A mutation in the control region or in an exon shared by alternative mRNAs will affect all the alternative proteins encoded by a given complex transcription unit. On the other hand, mutations in an exon present in only one of the alternative mRNAs will affect only the protein encoded by that mRNA. However, in the complex transcription unit shown in Figure –Complex transcription unit (middle), mutations d and e would complement each other in a genetic complementation test, even though they occur in the same gene. This is because a chromosome with mutation d can express a normal protein encoded by mRNA2 and a chromosome with mutation e can express a normal protein encoded by mRNA1. However, a chromosome with mutation c in an exon common to both mRNAs would not complement either mutation d or e. In other words, mutation c would be in the same complementation groups as mutations d and e, even though d and e themselves would not be in the same complementation group.

Chromosomal Organization of Genes and Noncoding DNA

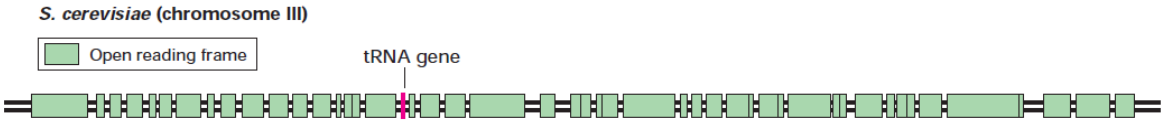
Introduction

- The vertebrates with the greatest amount of DNA per cell are amphibians, which are surely less complex than humans in their structure and behavior.
- The unicellular protozoal species *Amoeba dubia*has 200 times more DNA per cell than humans.
- Many plant species also have considerably more DNA per cell than humans have. For example, tulips have 10 times as much DNA per cell as humans.
- The DNA content per cell also varies considerably between closely related species.
- All insects or all amphibians would appear to be similarly complex, but the amount of haploid DNA in species within each of these phylogenetic classes varies by a factor of 100.

- The genomes of higher eukaryotes contain large amounts of noncoding DNA. For example, only a small portion of the β-globin gene cluster of humans, about 80 kb long, encodes protein.



- Moreover, compared with other regions of vertebrate DNA, the β-globin gene cluster is unusually rich in protein-coding sequences, and the introns in globin genes are considerably shorter than those in many human genes.
- In contrast, a typical 80-kb stretch of DNA from the yeast *S. cerevisiae*, a single-celled eukaryote (Figure 10-3b) contains many closely spaced protein-coding sequences without introns and relatively much less noncoding DNA.



- Of the 94 percent of human genomic DNA that has been sequenced, only ≈1.5 percent corresponds to protein-coding sequences (exons).
- Most human exons contain 50–200 base pairs, although the 3' exon in many transcription units is much longer.
- Human introns vary in length considerably. Although many are ≈90 bp long, some are much longer; their median length is 3.3 kb. Approximately one-third of human genomic DNA is thought to be transcribed into pre-mRNA precursors, but some 95 percent of these sequences are in introns, which are removed by RNA splicing.
- Different selective pressures during evolution may account, at least in part, for the remarkable difference in the amount of nonfunctional DNA in unicellular and multicellular organisms.

- For example, microorganisms must compete for limited amounts of nutrients in their environment, and metabolic economy thus is a critical characteristic.
- Since synthesis of nonfunctional (i.e., noncoding) DNA requires time and energy, presumably there was selective pressure to lose nonfunctional DNA during the evolution of microorganisms.
- On the other hand, natural selection in vertebrates depends largely on their behavior.
- The energy invested in DNA synthesis is trivial compared with the metabolic energy required for the movement of muscles; thus there was little selective pressure to eliminate nonfunctional DNA in vertebrates.

Protein coding genes

There are two groups of protein coding genes. (i) Solitary genes (ii) Duplicated genes

(i) Solitary genes

- In multicellular organisms, roughly 25–50 percent of the protein-coding genes are represented only once in the **haploid genome** and thus are termed *solitary* genes.
- A well studied example of a solitary protein-coding gene is the chicken lysozyme gene.
- The 15-kb DNA sequence encoding chicken lysozyme constitutes a simple transcription unit containing four exons and three introns.
- The flanking regions, extending for about 20 kb upstream and downstream from the transcription unit, do not encode any detectable mRNAs.
- Lysozyme, an enzyme that cleaves the polysaccharides in bacterial cell walls, is an abundant component of chicken egg-white protein and also is found in human tears.
- Its activity helps to keep the surface of the eye and the chicken egg sterile.

(ii) Duplicated genes

- Duplicated genes constitute the second group of protein coding genes.
- These are genes with close but nonidentical sequences that generally are located within 5–50 kb of one another. In vertebrate genomes, duplicated genes probably constitute half the protein-coding DNA sequences.

- A set of duplicated genes that encode proteins with similar but nonidentical amino acid sequences is called a **gene family;** the encoded, closely related, homologous proteins constitute a **protein family.**
- A few protein families, such as protein kinases, transcription factors, and vertebrate immunoglobulins, include hundreds of members.
- Most protein families, however, include from just a few to 30 or so members; common examples are cytoskeletal proteins, 70-kDa heat-shock proteins, the myosin heavy chain, chicken ovalbumin, and the α- and β-globins in vertebrates.
- The genes encoding the β-like globins are a good example of a gene family.
- The β-like globin gene family contains five functional genes designated β, δ, Aγ, Gγ, and ε; the encoded polypeptides are similarly designated.
- Two identical β-like globin polypeptides combine with two identical α-globin polypeptides (encoded by another gene family) and four small heme groups to form a hemoglobin molecule All the hemoglobins formed from the different β-like globins carry oxygen in the blood, but they exhibit somewhat different properties that are suited to specific roles in human physiology.
- For example, hemoglobins containing either the AγorGγ polypeptides are expressed only during fetal life.
- Because these fetal hemoglobins have a higher affinity for oxygen than adult hemoglobins, they can effectively extract oxygen from the maternal circulation in the placenta.
- The lower oxygen affinity of adult hemoglobins, which are expressed after birth, permits better release of oxygen to the tissues, especially muscles, which have a high demand for oxygen during exercise.
- The different β-globin genes probably arose by duplication of an ancestral gene, most likely as the result of an "unequal crossover" during meiotic recombination in a developing germ cell (egg or sperm).

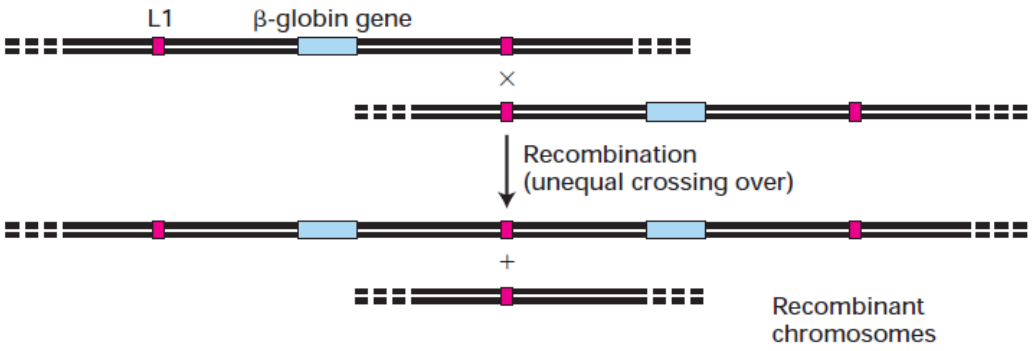


Figure: Gene duplication resulting from unequal crossing over

- Over evolutionary time the two copies of the gene that resulted accumulated random mutations; beneficial mutations that conferred some refinement in the basic oxygen-carrying function of hemoglobin were retained by natural selection, resulting in *sequence drift.*
- Repeated gene duplications and subsequent sequence drift are thought to have generated the contemporary globin-like genes observed in humans and other complex species today.
- Two regions in the human β-like globin gene cluster contain nonfunctional sequences, called **pseudogenes,** similar to those of the functional β-like globin genes.
- Sequence analysis shows that these pseudogenes have the same apparent exon-intron structure as the functional β-like globin genes, suggesting that they also arose by duplication of the same ancestral gene.
- However, sequence drift during evolution generated sequences that either terminate translation or block mRNA processing, rendering such regions nonfunctional even if they were transcribed into RNA.
- Because such pseudogenes are not deleterious, they remain in the genome and mark the location of a gene duplication that occurred in one of our ancestors.
- Other nonfunctional gene copies can arise by reverse transcription of mRNA into cDNA and integration of this intron-less DNA into a chromosome.
- Several different gene families encode the various proteins that make up the cytoskeleton.

- These proteins are present in varying amounts in almost all cells.
- In vertebrates, the major cytoskeletal proteins are the actins, tubulins, and intermediate filament proteins like the keratins.
- We examined the origin of one such family, the tubulin family, in the last chapter (see Figure 9-32). Although the physiological rationale for the cytoskeletal protein families is not as obvious as it is for the globins, the different members of a family probably have similar but subtly different functions suited to the particular type of cell in which they are expressed.

Tandemly repeated genes

- In vertebrates and invertebrates, the genes encoding rRNAs and some other noncoding RNAs such as some of the snRNAs involved in RNA splicing occur as tandemly repeated arrays.
- These are distinguished from the duplicated genes of gene families in that the multiple tandemly repeated genes encode identical or nearly identical proteins or functional RNAs.
- Most often copies of a sequence appear one after the other, in a head-to-tail fashion, over a long stretch of DNA.
- Within a tandem array of rRNA genes, each copy is exactly, or almost exactly, like all the others. Although the transcribed portions of rRNA genes are the same in a given individual, the nontranscribed spacer regions between the transcribed regions can vary.
- The tandemly repeated rRNA, tRNA, and histone genes are needed to meet the great cellular demand for their transcripts.
- To understand why, consider that a fixed maximal number of RNA copies can be produced from a single gene during one cell generation when the gene is fully loaded with RNA polymerase molecules.
- If more RNA is required than can be transcribed from one gene, multiple copies of the gene are necessary.

- For example, during early embryonic development in humans, many embryonic cells have a doubling time of ≈24 hours and contain 5–10 million ribosomes.
- To produce enough rRNA to form this many ribosomes, an embryonic human cell needs at least 100 copies of the large and small subunit rRNA genes, and most of these must be close to maximally active for the cell to divide every 24 hours (Table 10-2). That is, multiple RNA polymerases must be loaded onto and transcribing each rRNA gene at the same time.
- All eukaryotes, including yeasts, contain 100 or more copies of the genes encoding 5S rRNA and the large and small subunit rRNAs.
- The importance of repeated rRNA genes is illustrated by Drosophila mutants called bobbed (because they have stubby wings), which lack a full complement of the tandemly repeated **pre-rRNA**genes.
- A bobbed mutation that reduces the number of pre-rRNA genes to less than ≈50 is a recessive lethal mutation.
- Multiple copies of tRNA and histone genes also occur, often in clusters, but generally not in tandem arrays.

Simple sequence DNA

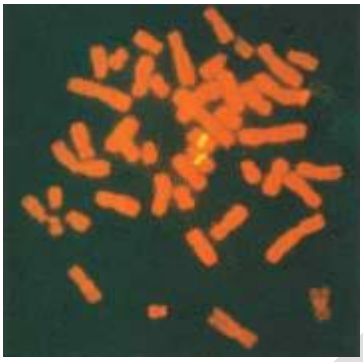
- Besides duplicated protein-coding genes and tandemly repeated genes, eukaryotic cells contain multiple copies of other DNA sequences in the genome, generally referred to as repetitious DNA
- Of the two main types of repetitious DNA, the less prevalent is **simple-sequence DNA**, which constitutes about 3 percent of the human genome and is composed of perfect or nearly perfect repeats of relatively short sequences.
- Simple-sequence DNA is commonly called satellite DNA because in early studies of DNAs from higher organisms using equilibrium buoyant-density ultracentrifugation some simple-sequence DNAs banded at a different position from the bulk of cellular DNA.

These were called satellite bands to distinguish them from the main band of DNA in the buoyant-density gradient.

- Simple-sequence DNAs in which the repeats contain 1–13 base pairs are often called microsatellites.
- Most have repeat lengths of 1–4 base pairs and usually occur in tandem repeats of 150 base pairs or fewer.
- Microsatellites are thought to have originated by "backward slippage" of a daughter strand on its template strand during DNA replication so that the same short sequence is copied twice.
- Microsatellites occasionally occur within transcription units.
- Some individuals are born with a larger number of repeats in specific genes than observed in the general population, presumably because of daughter-strand slippage during DNA replication in a germ cell from which they developed.
- Such expanded microsatellites have been found to cause at least 14 different types of neuromuscular diseases, depending on the gene in which they occur.
- In some cases expanded microsatellites behave like a recessive mutation because they interfere with the function or expression of the encoded gene.
- But in the more common types of diseases associated with expanded microsatellite repeats, myotonic dystrophy and spinocerebellar ataxia, the expanded repeats behave like dominant mutations because they interfere with RNA processing in general in the neurons where the affected genes are expressed.

Simple sequence DNA as a marker

- Human metaphase chromosomes stained with a fluorescent dye were hybridized in situ with a particular simple-sequence DNA labeled with a fluorescent biotin derivative.
- When viewed under the appropriate wavelength of light, the DNA appears red and the hybridized simple-sequence DNA appears as a yellow band on chromosome 16, thus locating this particular simple sequence to one site in the genome.



- Most satellite DNA is composed of repeats of 14–500 base pairs in tandem repeats of 20–100 kb. In situ hybridization studies with metaphase chromosomes have localized these satellite DNAs to specific chromosomal regions.
- In most mammals, much of this satellite DNA lies near **centromeres,** the discrete chromosomal regions that attach to spindle microtubules during mitosis and meiosis.
- Satellite DNA is also located at **telomeres,** the ends of chromosomes, and at specific locations within chromosome arms in some organisms.
- These latter sequences can be useful for identifying particular chromosomes by **fluorescence in situ hybridization (FISH)**.
- Simple-sequence DNA located at centromeres may assist in attaching chromosomes to spindle microtubules during mitosis.
- As yet, however, there is little clear-cut experimental evidence demonstrating any function for most simple sequence DNA, with the exception of the short repeats at the very ends of chromosomes.

DNA Fingerprinting

- DNA fingerprinting depends on differences in length of simple-sequence DNAs
- Within a species, the nucleotide sequences of the repeat units composing simple-sequence DNA tandem arrays are highly conserved among individuals.
- In contrast, differences in the number of repeats, and thus in the length of simple-sequence tandem arrays containing the same repeat unit, are quite common among individuals.

- These differences in length are thought to result from unequal crossing over within regions of simple-sequence DNA during meiosis.
- As a consequence of this unequal crossing over, the lengths of some tandem arrays are unique in each individual.
- In humans and other mammals, some of the satellite DNA exists in relatively short 1- to 5-kb regions made up of 20–50 repeat units, each containing 15 to about 100 base pairs.
- These regions are called minisatellitesto distinguish them from the more common regions of tandemly repeated satellite DNA, which are ≈20–100 kb in length.
- They differ from microsatellites mentioned earlier, which have very short repeat units.
- Even slight differences in the total lengths of various minisatellites from different individuals can be detected by **Southern blotting** of cellular DNA treated with a restriction enzyme that cuts outside the repeat sequence.
- The polymerase chain reaction (PCR), using primers that hybridize to the unique sequences flanking each minisatellite, also can detect differences in minisatellite lengths between individuals.
- These DNA polymorphisms form the basis of DNA fingerprinting, which is superior to conventional fingerprinting for identifying individuals.

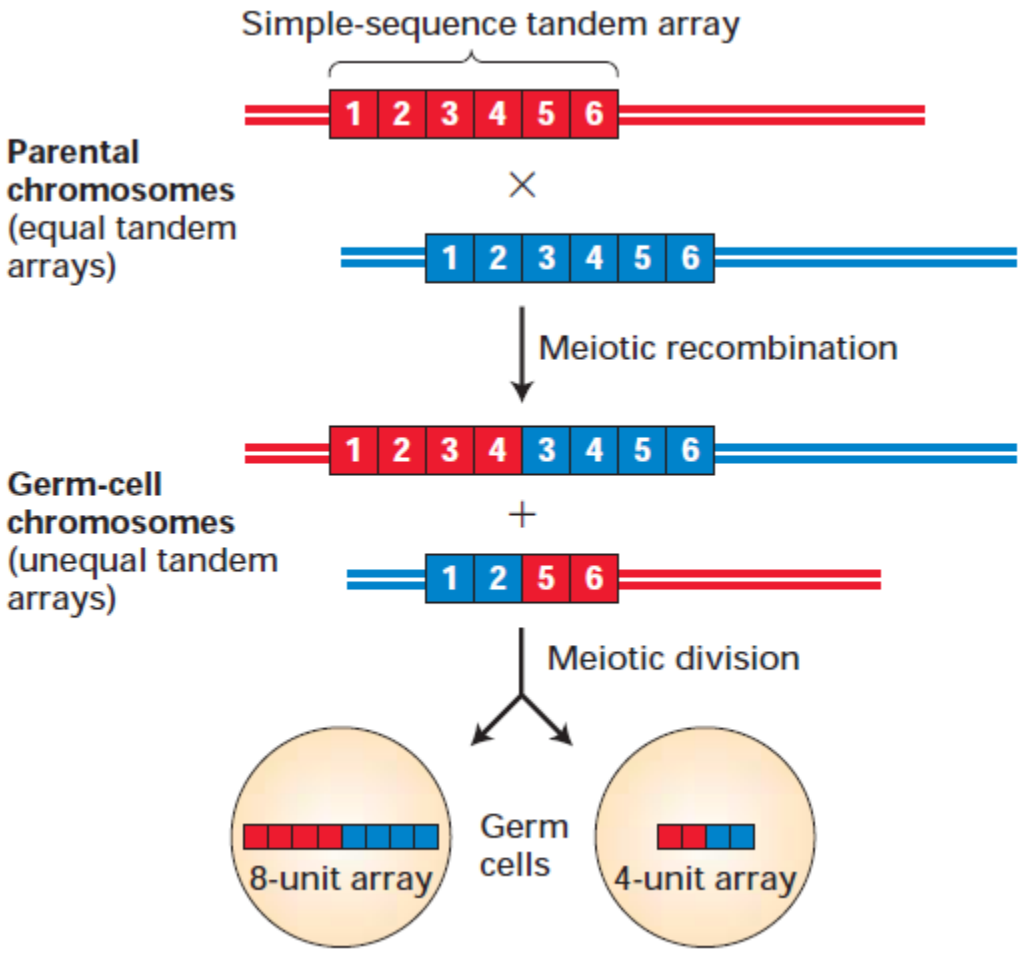


Figure: Generation of differences in lengths of a simple-sequence DNA by unequal crossing over during meiosis. In this example, unequal crossing over within a stretch of DNA containing six copies (1–6) of a particular simple sequence repeat unit yields germ cells containing either an eight unit or a four-unit tandem array.

Functional elements

Any eukaryotic chromosome must contain three functional elements in order to replicate and segregate correctly:

(1) Replication origins at which DNA polymerases and other proteins initiate synthesis of DNA

(2) the Centro mere, and

(3) The telomeres or the two ends

These three are very much important for chromosome function.

Structural organisation of eukaryotic chromosomes

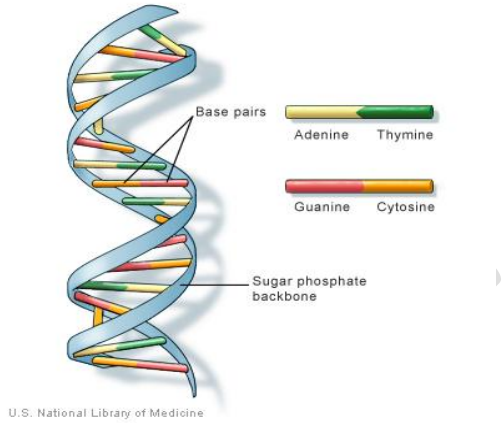
DNA, or deoxyribonucleic acid, is the hereditary material in humans and almost all other organisms carrying the genetic information in genes. Most DNA is located in the cell nucleus (where it is called nuclear DNA), but a small amount of DNA can also be found in the mitochondria (where it is called mitochondrial DNA or mtDNA). DNA is one of the three major macromolecules that are essential for all known forms of life.

DNA consists of two long polymers of simple units called nucleotides, with backbones made of sugars and phosphate groups joined by ester bonds. These two strands run in opposite directions to each other and are therefore anti-parallel. Attached to each sugar is one of four types of molecules called nucleobases (informally, *bases*). It is the sequence of these four nucleobases along the backbone that encodes information. This information is read using the genetic code, which specifies the sequence of the amino acids within proteins. The code is read by copying stretches of DNA into the related nucleic acid RNA in a process called transcription.

Within cells DNA is organized into long structures called chromosomes. During cell division these chromosomes are duplicated in the process of DNA replication, providing each cell its own complete set of chromosomes. Eukaryotic organisms (animals, plants, fungi, and protists) store most of their DNA inside the cell nucleus and some of their DNA in organelles, such as mitochondria or chloroplasts. In contrast, prokaryotes (bacteria and archaea) store their DNA only in the cytoplasm. Within the chromosomes, chromatin proteins such as histones compact and organize DNA. These compact structures guide the interactions between DNA and other proteins, helping control which parts of the DNA are transcribed.

Structure of DNA

DNA is a long polymer made from repeating units called nucleotides. As first discovered by James D. Watson and Francis Crick, the structure of DNA of all species comprises two helical chains each coiled round the same axis, and each with a pitch of 34 Angstroms (3.4 nanometers) and a radius of 10 Angstroms (1.0 nanometers).



DNA is a double helix formed by base pairs attached to a sugar-phosphate backbone

The backbone of the DNA strand is made from alternating phosphate and sugar residues. The sugar in DNA is 2-deoxyribose, which is a pentose (five-carbon) sugar. The sugars are joined together by phosphate groups that form phosphodiester bonds between the third and fifth carbon atoms of adjacent sugar rings. These asymmetric bonds mean a strand of DNA has a direction. In a double helix the direction of the nucleotides in one strand is opposite to their direction in the other strand: the strands are *antiparallel*. The asymmetric ends of DNA strands are called the 5′ (*five prime*) and 3′ (*three prime*) ends, with the 5' end having a terminal phosphate group and the 3' end a terminal hydroxyl group. One major difference between DNA and RNA is the sugar, with the 2-deoxyribose in DNA being replaced by the alternative pentose sugar ribose in RNA.

The four bases found in DNA are adenine (abbreviated A), cytosine (C), guanine (G) and thymine (T). These four bases are attached to the sugar/phosphate to form the complete nucleotide, as shown for adenosine monophosphate. The nucleobases are classified into two types: the purines, A and G, being fused five- and six-membered heterocyclic compounds, and the pyrimidines, the six-membered rings C and T. A fifth pyrimidine nucleobase, uracil (U), usually takes the place of thymine in RNA and differs from thymine by lacking a methyl group on its ring. Uracil is not usually found in DNA, occurring only as a breakdown product of cytosine. In addition to RNA and DNA a large number of artificial nucleic acid analogues have also been created to study the proprieties of nucleic acids, or for use in biotechnology.

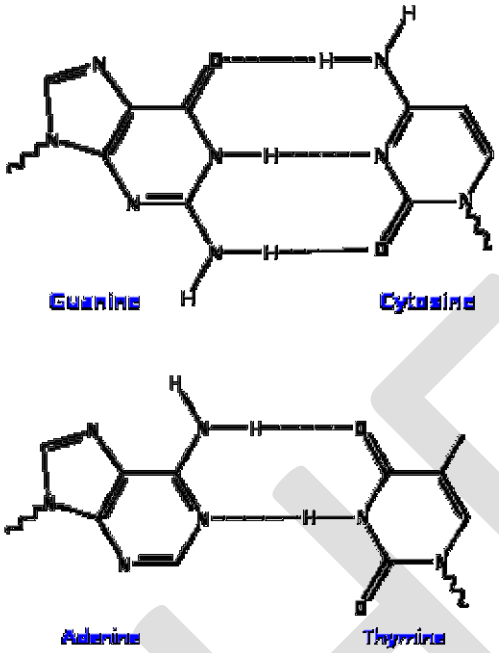
Grooves

Twin helical strands form the DNA backbone. Another double helix may be found by tracing the spaces, or grooves, between the strands. These voids are adjacent to the base pairs and may provide a binding site. As the strands are not directly opposite each other, the grooves are unequally sized. One groove, the major groove, is 22 Å wide and the other, the minor groove, is 12 Å wide. The narrowness of the minor groove means that the edges of the bases are more accessible in the major groove. As a result, proteins like transcription factors that can bind to specific sequences in double-stranded DNA usually make contacts to the sides of the bases exposed in the major groove.

Base pairing

In a DNA double helix, each type of nucleobase on one strand normally interacts with just one type of nucleobase on the other strand. This is called complementary base pairing. Here, purines form hydrogen bonds to pyrimidines, with a bonding only to T, and C bonding only to G. This arrangement of two nucleotides binding together across the double helix is called a base pair. As hydrogen bonds are not covalent, they can be broken and rejoined relatively easily. The two strands of DNA in a double helix can therefore be pulled apart like a zipper, either by a mechanical force or high temperature. As a result of this complementarity, all the information in the double-stranded sequence of a DNA helix is duplicated on each strand, which is vital in DNA replication. Indeed, this reversible and specific interaction between complementary base pairs is critical for all the functions of DNA in living organisms.

The two types of base pairs form different numbers of hydrogen bonds, AT forming two hydrogen bonds, and GC forming three hydrogen bonds (see figures, right). DNA with high GC-content is more stable than DNA with low GC-content. Long DNA helices with a high GC-content have stronger-interacting strands, while short helices with high AT content have weaker-interacting strands.



Top, a GC base pair with three hydrogen bonds. Bottom, an AT base pair with two hydrogen bonds. Non-covalent hydrogen bonds between the pairs are shown as dashed lines.

Basic Structure of a Protein-Coding Gene

A protein-coding gene consists of a promoter followed by the coding sequence for the protein and then a terminator. The promoter is a base-pair sequence that specifies where transcription begins. The coding sequence is a base-pair sequence that includes coding information for the polypeptide chain specified

Chromatin

Chromatin is the combination of DNA and proteins that make up the contents of the nucleus of a cell. The primary functions of chromatin are to package DNA into a smaller volume to fit in the cell, to strengthen the DNA to allow mitosis and meiosis and prevent DNA damage, and to control gene expression and DNA replication. The primary protein components of chromatin are histones that compact the DNA. Chromatin is only found in eukaryoticcells.

Prokaryotic cells have a very different organization of their DNA which is referred to as a genophore (a chromosome without chromatin).

The structure of chromatin depends on several factors. The overall structure depends on the stage of the cell cycle: during interphase the chromatin is structurally loose to allow access to RNA and DNA polymerases that transcribe and replicate the DNA. The local structure of chromatin during interphase depends on the genes present on the DNA: DNA coding genes that are actively transcribed ("turned on") are more loosely packaged and are found associated with RNA polymerases (referred to as euchromatin) while DNA coding inactive genes ("turned off") are found associated with structural proteins and are more tightly packaged (heterochromatin). Epigenetic chemical modification of the structural proteins in chromatin also alters the local chromatin structure, in particular chemical modifications of histone proteins by methylation and acetylation. As the cell prepares to divide, i.e. enters mitosis or meiosis, the chromatin packages more tightly to facilitate segregation of the chromosomes during anaphase. During this stage of the cell cycle this makes the individual chromosomes in many cells visible by optical microscope.

Chromatin organization:

In general terms, there are three levels of chromatin organization:

1. DNA wraps around histone proteins forming nucleosomes; the "beads on a string" structure (euchromatin).
2. Multiple histones wrap into a 30 nm fibre consisting of nucleosome arrays in their most compact form (heterochromatin).
3. Higher-level DNA packaging of the 30 nm fibre into the metaphase chromosome (during mitosis and meiosis)

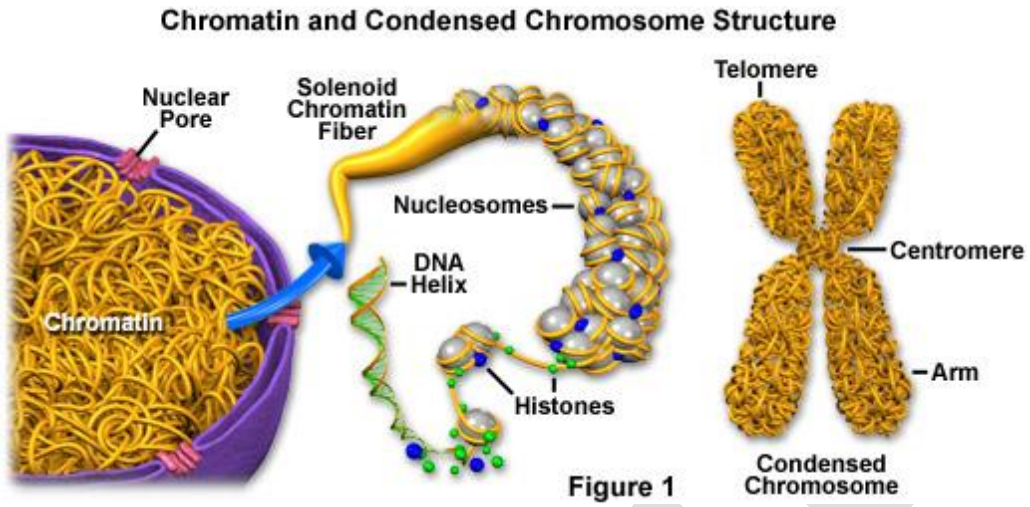


Figure 1

Components in chromatin organization

Histones

Histones are the major structural proteins of chromosomes. The DNA molecule is wrapped twice around a Histone Octamer to make a Nucleosome. Six Nucleosomes are assembled into a Solenoid in association with H1 histones. The solenoids are in turn coiled onto a Scaffold, which is futher coiled to make the chromosomal matrix.

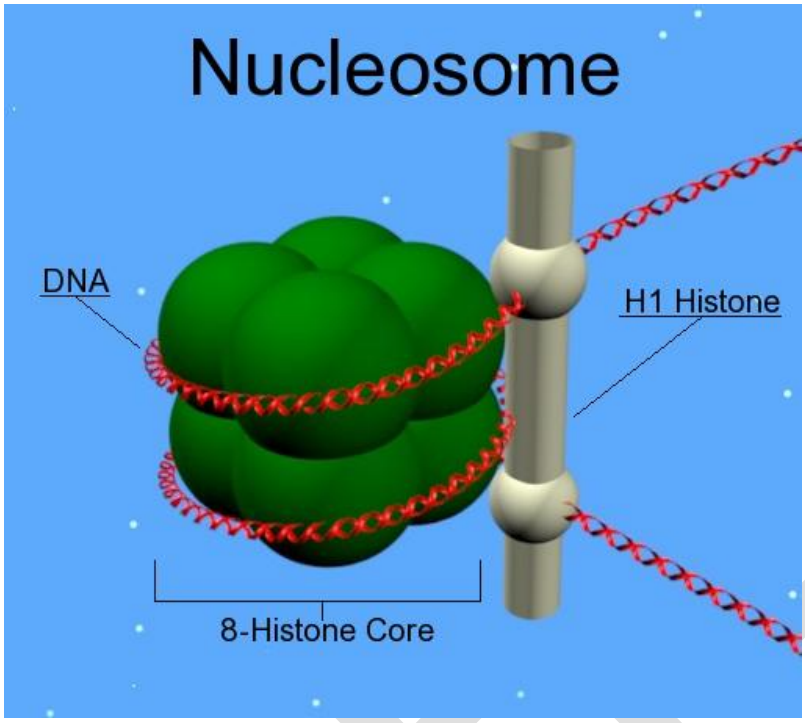
Histones are highly alkaline proteins found in eukaryotic cell nuclei that package and order the DNA into structural units called nucleosomes. They are the chief protein components of chromatin, acting as spools around which DNA winds, and play a role in gene regulation. Without histones, the unwound DNA in chromosomes would be very long (a length to width ratio of more than 10 million to one in human DNA). For example, each human cell has about 1.8 meters of DNA, but wound on the histones it has about 90 micrometers (0.09 mm) of chromatin, which, when duplicated and condensed during mitosis, result in about 120 micrometers of chromosomes.

Classes

Five major families of histones exist: H1/H5, H2A, H2B, H3, and H4. Histones H2A, H2B, H3 and H4 are known as the core histones, while histones H1 and H5 are known as the linker histones.

1. Nucleosome and "beads-on-a-string"

The basic repeat element of chromatin is the nucleosome, interconnected by sections of linker DNA. The nucleosome core is formed of two H2A-H2B dimers and a H3-H4 tetramer, forming two nearly symmetrical halves by tertiary structure (C2 symmetry; one macromolecule is the mirror image of the other) and 147 base pairs of DNA wrap around this core particle 1.65 times in a left-handed super-helical turn. The linker histone H1 binds the nucleosome and the entry and exit sites of the DNA, thus locking the DNA into place and allowing the formation of higher order structure. The most basic such formation is the 10 nm fiber or beads on a string conformation. This involves the wrapping of DNA around nucleosomes with approximately 50 base pairs of DNA separating each pair of nucleosomes (also referred to as linker DNA). The assembled histones and DNA is called chromatin. Higher-order structures include the 30 nm fiber (forming an irregular zigzag) and 100 nm fiber, these being the structures found in normal cells. During mitosis and meiosis, the condensed chromosomes are assembled through interactions between nucleosomes and other regulatory proteins.



Function:

1. Compacting DNA strands

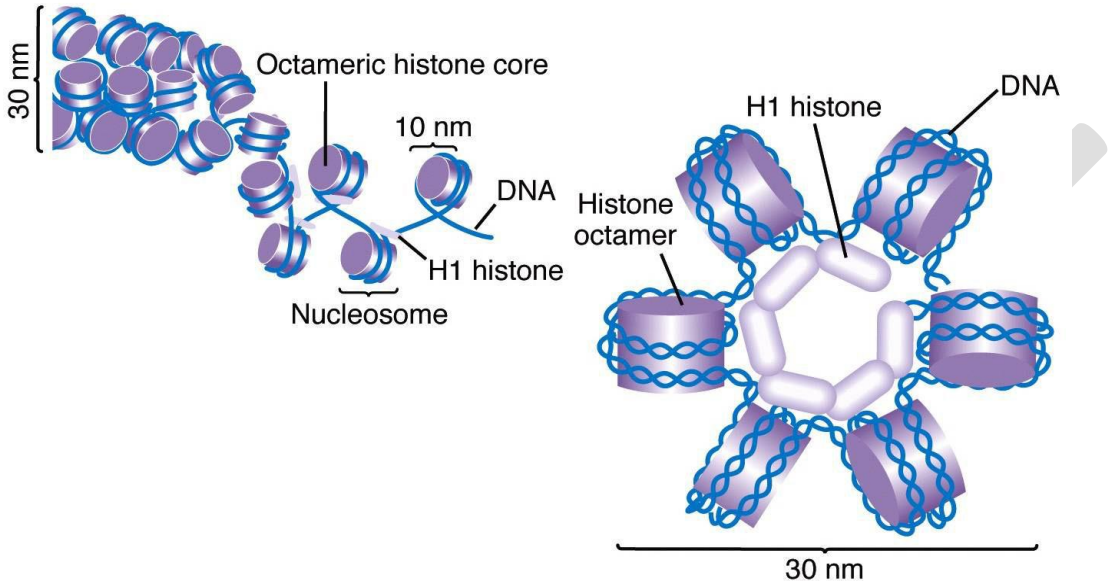
Histones act as spools around which DNA winds. This enables the compaction necessary to fit the large genomes of eukaryotes inside cell nuclei the compacted molecule is 40,000 times shorter than an unpacked molecule.

2. Chromatin regulation

Histones undergo posttranslational modifications that alter their interaction with DNA and nuclear proteins. The H3 and H4 histones have long tails protruding from the nucleosome, which can be covalently modified at several places. Modifications of the tail include methylation, acetylation, phosphorylation, ubiquitination, SUMOylation, citrullination, and ADP-ribosylation. The core of the histones H2A, H2B, and H3 can also be modified. Combinations of modifications are thought to constitute a code, the so-called "histone code". Histone modifications act in diverse biological processes such as gene regulation, DNA repair, chromosome condensation (mitosis) and spermatogenesis (meiosis).

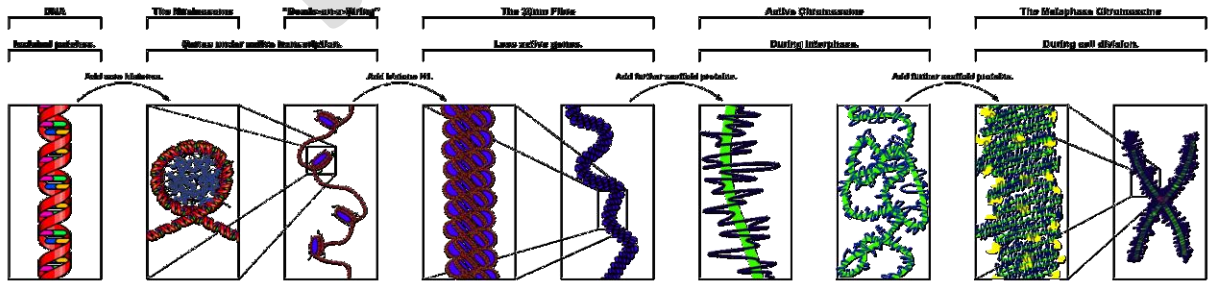
2. 30 nm chromatin fibre

With addition of H1, the "beads-on-a-string" structure in turn coils into a 30 nm diameter helical structure known as the 30 nm fibre or filament. This level of chromatin structure is thought to be the form of euchromatin, which contains actively transcribed genes. Electron microscopic studies have demonstrated that the 30 nm fibre is highly dynamic such that it unfolds into a 10 nm fiber ("beads-on-a-string") structure when transversed by an RNA polymerase engaged in transcription.



DNA packaging

Prokaryotes do not possess nuclei. Instead, their DNA is organized into a structure called the nucleoid.The nucleoid is a distinct structure and occupies a defined region of the bacterial cell. This structure is, however, dynamic and is maintained and remodeled by the actions of a range of histone-like proteins, which associate with the bacterial chromosome.



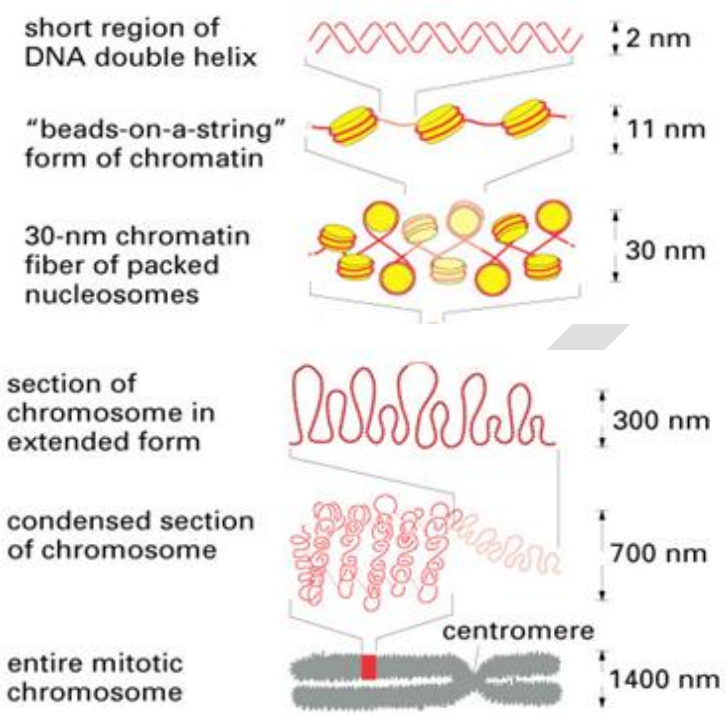


Fig : The major structures in DNA compaction; DNA, the nucleosome, the 10nm "beads-on-a-string" fibre, the 30nm fibre and the metaphase chromosome

Net result: each DNA molecule has been packaged into a mitotic chromosome that is 10,000 fold shorter than its extended length

Mobile DNA elements

Introduction

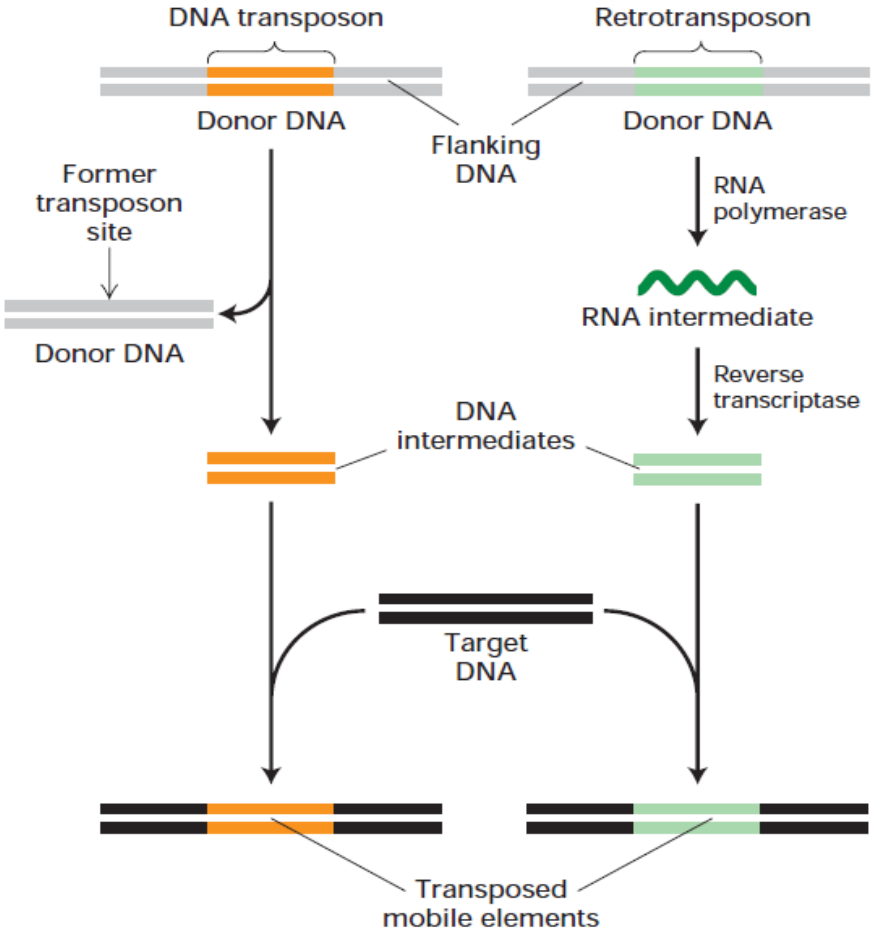
- The second type of repetitious DNA in eukaryotic genomes, termed interspersed repeats (also known as moderately repeated DNA, or intermediate-repeat DNA).
- They are composed of a very large number of copies of relatively few sequence families.
- These sequences, which are interspersed throughout mammalian genomes, make up ≈25–50 percent of mammalian DNA (≈45 percent of human DNA).
- Because moderately repeated DNA sequences have the unique ability to "move" in the genome, they are called **mobile DNA elements** (or transposable elements).

- Although mobile DNA elements, ranging from hundreds to a few thousand base pairs in length, originally were discovered in eukaryotes, they also are found in prokaryotes.
- The process by which these sequences are copied and inserted into a new site in the genome is called **transposition.**
- Mobile DNA elements (or simply mobile elements) are essentially molecular symbionts and have no specific function in the biology of their host organisms, but exist only to maintain themselves.
- For this reason, Francis Crick referred to them as "selfish DNA."
- When transposition of eukaryotic mobile elements occurs in germ cells, the transposed sequences at their new sites can be passed on to succeeding generations.
- In this way, mobile elements have multiplied and slowly accumulated in eukaryotic genomes over evolutionary time.
- Since mobile elements are eliminated from eukaryotic genomes very slowly, they now constitute a significant portion of the genomes of many eukaryotes.
- Transposition also may occur within a somatic cell; in this case the transposed sequence is transmitted only to the daughter cells derived from that cell.
- In rare cases, this may lead to a somatic-cell mutation with detrimental phenotypic effects, for example, the inactivation of a tumor suppressor gene.

Movement of mobile elements

- Barbara McClintock discovered the first mobile elements in maize (corn) during the 1940s.
- She characterized genetic entities that could move into and back out of genes, changing the phenotype of corn kernels.
- Her theories were very controversial until similar mobile elements were discovered in bacteria, where they were characterized as specific DNA sequences, and the molecular basis of their transposition was deciphered.
- Mobile elements are classified into two categories: (1) those that transpose directly

- as DNA and (2) those that transpose via an RNA intermediate transcribed from the mobile element by an RNA polymerase and then converted back into double-stranded DNA by a **reverse transcriptase** .
- Mobile elements that transpose through a DNA intermediate are generally referred to as **DNA transposons.**
- Mobile elements that transpose to new sites in the genome via an RNA intermediate are called **retrotransposons**because their movement is analogous to the infectious process of retroviruses. Indeed, retroviruses can be thought of as retrotransposons that evolved genes encoding viral coats, thus allowing them to transpose between cells.



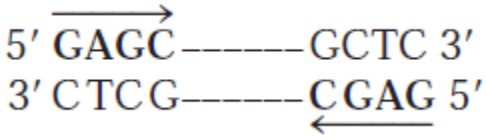
Major types of mobile elements

Mobile elements that move as DNA

- Most mobile elements in bacteria transpose directly as DNA.
- Most mobile elements in eukaryotes are retrotransposons, but eukaryotic DNA transposons also occur.

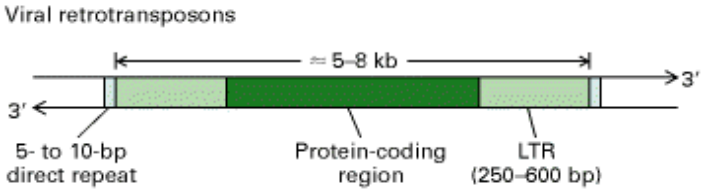
Bacterial Insertion Sequences

- Certain *E.coli*mutations caused by the spontaneous insertion of a DNA sequence, ≈1–2 kb long, into the middle of a gene.
- These inserted stretches of DNA are called *insertion sequences,* or *IS elements.*
- So far, more than 20 different IS elements have been found in *E. coli* and other bacteria.
- Transposition of an IS element is a very rare event, occurring in only one in 10^5–10^7 cells per generation, depending on the IS element.
- Many transpositions inactivate essential genes, killing the host cell and the IS elements it carries.
- Therefore, higher rates of transposition would probably result in too great a mutation rate for the host cell to survive.
- However, since IS elements transpose more or less randomly, some transposed sequences enter nonessential regions of the genome (e.g., regions between genes), allowing the cell to survive.
- At a very low rate of transposition, most host cells survive and therefore propagate the symbiotic IS element.
- IS elements also can insert into plasmids or lysogenic viruses, and thus be transferred to other cells.
- When this happens, IS elements can transpose into the chromosomes of virgin cells.
- An inverted repeat, usually containing ≈50 base pairs, invariably is present at each end of an insertion sequence.
- In an inverted repeat the 5'→ 3'sequence on one strand is repeated on the other strand, as:



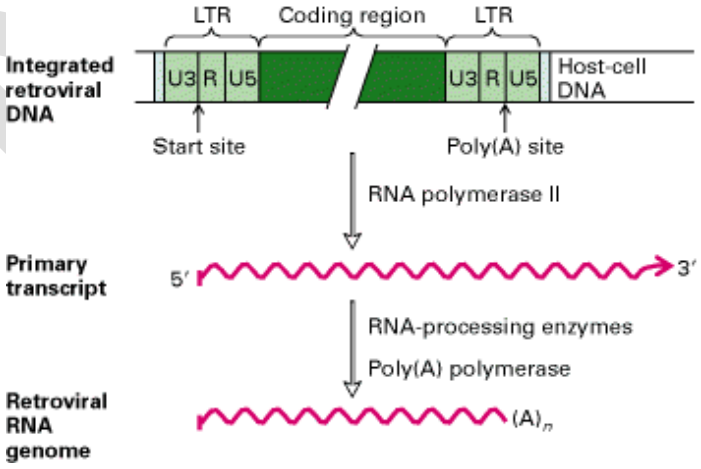
Between the inverted repeats is a region that encodes transposase, an enzyme required for transposition of the IS elements

The general structure of viral retrotransposons found in eukaryotes



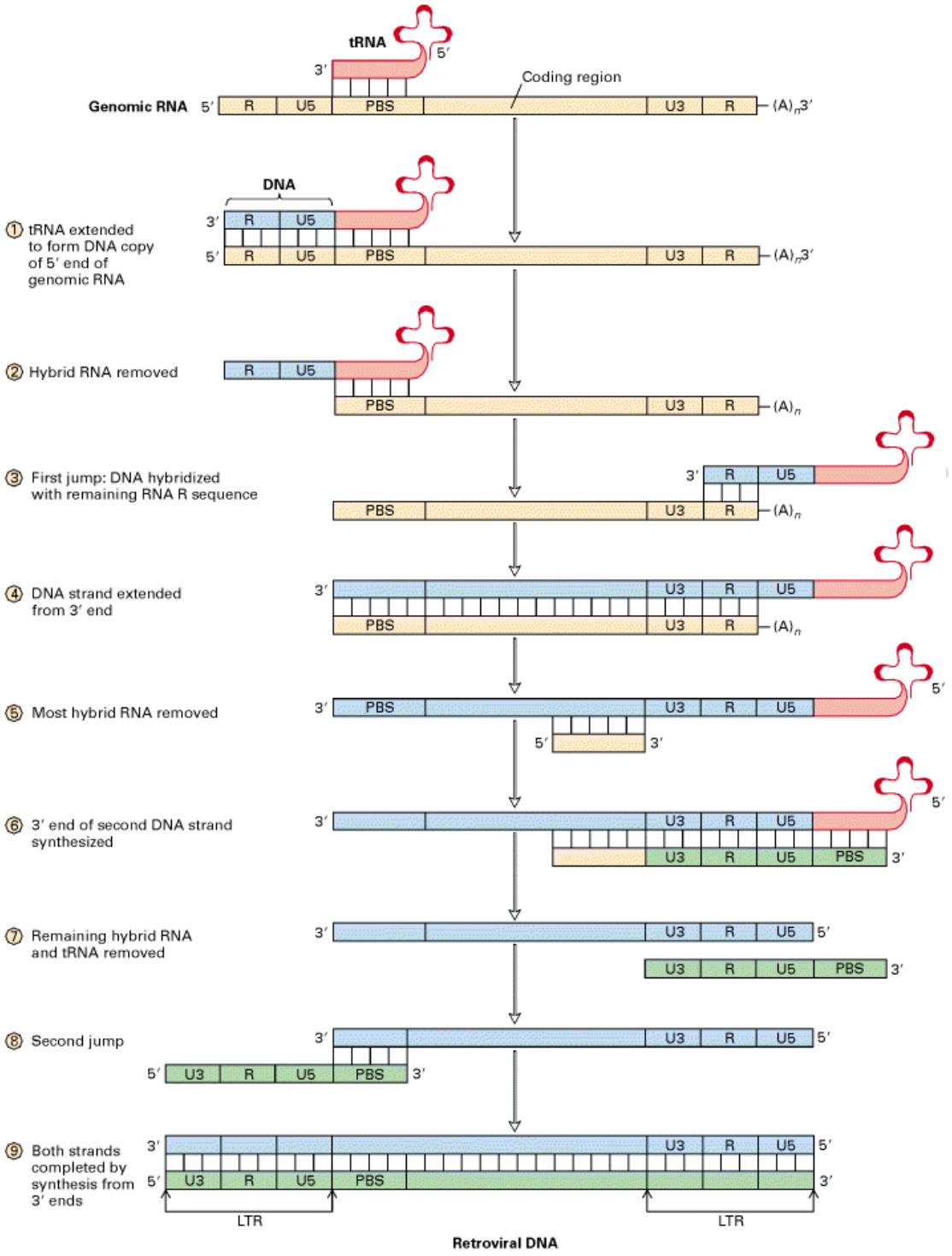
General structure of eukaryotic viral retrotransposon

- 5 and 3 direct repeats typical of all mobile elements.
- The central protein-coding region is flanked by two long terminal repeats (LTRs) 250- to 600-bp, which are element-specific direct repeats.
- LTRs, the hallmark of these mobile elements, also are present in retroviral DNA.
- LTRs are characteristic of integrated retroviral DNA and are critical to the life cycle of retroviruses.
- Like other mobile elements, integrated retrotransposons have short target-site direct repeats at their 3 and 5 ends.
- The protein-coding region constitutes 80 percent or more of a retrotransposon.



Generation of retroviral genomic RNA from integrated retroviral DNA

- The short direct repeat sequences (light blue) of target-site DNA are generated during integration of the retroviral DNA into the host-cell genome.
- The left LTR directs cellular RNA polymerase II to initiate transcription at the first nucleotide of the left R region. The resulting primary transcript extends beyond the right LTR.
- The right LTR, now present in the RNA primary transcript, directs cellular enzymes to cleave the primary transcript at the last nucleotide of the right R region and to add a poly(A) tail, yielding a retroviral RNA genome.
- The genomic RNA is packaged in the virion with a retrovirus-specific cellular tRNA hybridized to a complementary sequence near its 5 end called the primer-binding site (PBS).
- The retroviral RNA has a short direct-repeat terminal sequence (R) at each end. The overall reaction is catalyzed by reverse transcriptase, which catalyzes polymerization of deoxyribonucleotides and digestion of the RNA strand in a DNA-RNA hybrid.
- The entire process yields a double-stranded DNA molecule that is longer than the template RNA and has a long terminal repeat (LTR) at each end.
- The PBS and R regions are actually much shorter than the U5 and U3 regions, and the central coding region is very much longer (7500 nucleotides) than the other regions.

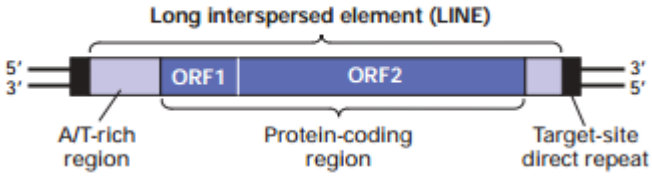


Generation of LTRs during reverse transcription of retroviral genomic RNA

NonviralRetrotransposons

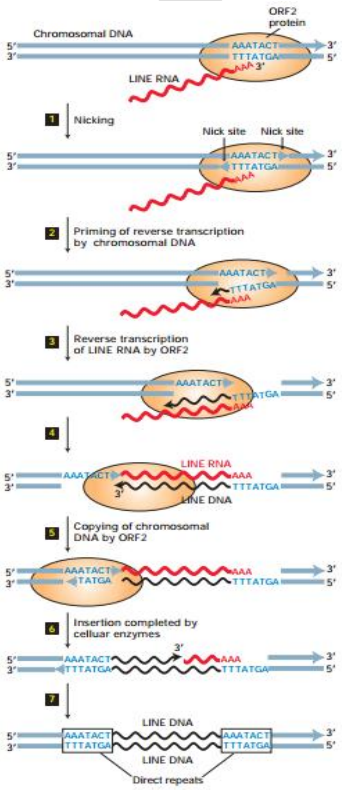
- The most abundant mobile elements in mammals are retrotransposons that lack LTRs, sometimes called nonviral retro-transposons.
- There are two classes of nonviral retro transposons.
  (i) long interspersed elements (LINEs)and
  (ii) short interspersed elements (SINEs).
- In humans, full-length LINEs are ≈6 kb long, and SINEs are ≈300 bp long.
- Repeated sequences with the characteristics of LINEs have been observed in protozoans, insects, and plants, and they are particularly abundant in the genomes of mammals.
- SINEs also are found primarily in mammalian DNA.
- They transpose through an RNA intermediate.
- Human DNA contains three major families of LINE sequences that are similar in their mechanism of transposition, but differ in their sequences: L1, L2, and L3. Only members of the L1 family transpose in the contemporary human genome.
- LINE sequences are present at ≈900,000 sites in the human genome, accounting for a staggering 21 percent of total human DNA.

The general structure of a complete LINE



- LINEs usually are flanked by short direct repeats, and contain two long open reading frames (ORFs).
- ORF1, ≈1 kb long, encodes an RNA-binding protein.
- ORF2, ≈4 kb long, encodes a protein that has a long region of homology with the reverse transcriptases of retroviruses and viral retrotransposons, but also exhibits DNA endonuclease activity.

- Since LINEs do not contain LTRs, their mechanism of transposition through an RNA intermediate differs from that of LTR retrotransposons.
- ORF1 and ORF2 proteins are translated from a LINE RNA.
- Transcription by RNA polymerase II is directed by promoter sequences at the left end of integrated LINE DNA.
- LINE RNA is polyadenylated by the same post-transcriptional mechanism that polyadenylates other mRNAs.
- The LINE RNA then is transported into the cytoplasm, where it is translated into ORF1 and ORF2 proteins.
- Multiple copies of ORF1 protein then bind to the LINE RNA, and ORF2 protein binds to the poly(A) tail.
- The LINE RNA is then transported back into the nucleus as a complex with ORF1 and ORF2. ORF2 then makes staggered nicks in chromosomal DNA on either side of any A/T-rich sequence in the genome (step 1).



- Reverse transcription of LINE RNA by ORF2 is primed by the single-stranded T-rich sequence generated by the nick in the bottom strand, which hybridizes to the LINE poly(A) tail (step 2). ORF2 then reverse-transcribes the LINE RNA (step 3) and then continues this new DNA strand, switching to the single-stranded region of the upper chromosomal strand as a template (steps 4 and 5).
- Cellular enzymes then hydrolyze the RNA and extend the 3' end of the chromosomal DNA top strand, replacing the LINE RNA strand with DNA (step 6).
- Finally, 5'and 3'ends of DNA strands are ligated, completing the insertion (step 7).
- These last steps (6 and 7) probably are catalyzed by the same cellular enzymes that remove RNA primers and ligate Okazaki fragments during DNA replication.
- The complete process results in insertion of a copy of the original LINE retrotransposon into a new site in chromosomal DNA.
- A short direct repeat is generated at the insertion site because of the initial staggered cleavage of the two chromosomal DNA strands (step 1).

SINEs

- The second most abundant class of mobile elements in the human genome, SINEs constitute ≈13 percent of total human DNA.
- Varying in length from about 100 to 400 base pairs, these retrotransposons do not encode protein, but most contain a 3' A/T-rich sequence similar to that in LINEs.
- SINEs are transcribed by RNA polymerase III, the same nuclear RNA polymerase that transcribes genes encoding tRNAs, 5S rRNAs, and other small stable RNAs
- Most likely, the ORF1 and ORF2 proteins expressed from full length LINEs mediate transposition of SINEs by the retrotransposition mechanism.
- SINEs occur at about 1.6 million sites in the human genome.
- Of these, ≈1.1 million are Alu elements, so named because most of them contain a single recognition site for the restriction enzyme AluI.

- Alu elements exhibit considerable sequence homology with and may have evolved from 7SL RNA, a component of the signal-recognition particle.
- This abundant cytosolic ribonucleoprotein particle aids in targeting certain polypeptides, as they are being synthesized, to the membranes of the endoplasmic reticulum.
- Alu elements are scattered throughout the human genome at sites where their insertion has not disrupted gene expression: between genes, within introns, and in the 3'untranslated regions of some mRNAs.
- For instance, nine Alu elements are located within the human β-globin gene cluster.
- The overall frequency of L1 and SINE retrotranspositions in humans is estimated to be about one new retro-transposition in very eight individuals, with ≈40 percent being L1 and 60 percent SINEs, of which ≈90 percent are Alu elements.
- Similar to other mobile elements, most SINEs have accumulated mutations from the time of their insertion in the germ line of an ancient ancestor of modern humans.
- Like LINEs, many SINEs also are truncated at their 5'end.

Mutation-Definition

A mutation is a permanent change in the sequence of DNA. In order for an observable effect, mutations must occur in gene exons or regulatory elements. Changes in the non-coding regions of DNA (introns and junk DNA) generally do not affect function.

Causes of mutations

Mutations can be caused by external (exogenous) or endogenous (native) factors or they may be caused by errors in the cellular machinery. Physical or chemical agents that induce mutations in DNA are called mutagens and are said to be mutagenic.

Exogenous factors: environmental factors such as sunlight, radiation, and smoking can cause mutations.

Endogenous factors: errors during DNA replication can lead to genetic changes as can toxic by-products of cellular metabolism.

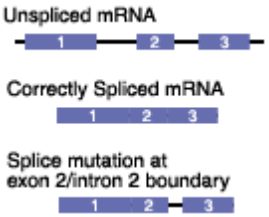
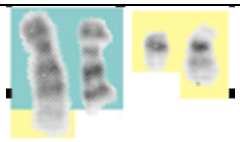
Consequences of mutations

Mutations can be advantageous and lead to an evolutionary advantage of a certain genotype. Mutations can also be deleterious, causing disease, developmental delays, structural abnormalities, or other effects.

Types of Mutations:

There are several classes of mutations

Type	Description	Example
1. Deletion	Genetic material is removed or deleted. A few bases can be deleted (as shown on the left) or it can be complete or partial loss of a chromosome (shown on right).	...TCGGAATCGC... ...TCGCGC...
2. Frameshift	The insertion or deletion of a number of bases that is not a multiple of 3. This alters the reading frame of the gene and frequently results in a premature stop codon and protein truncation.	...ACT TTT CAT AGT... ...Thr Phe His Ser... ...ACT TTT TCA TAG T... ...Thr Phe Ser Stop
3. Insertion	When genetic material is put into another region of DNA. This may be the insertion of 1 or more bases, or it can be part of one chromosome being inserted into another, non-homologous chromosome.	...TTGAAAACGCTG... ...TTGAAAAACGCTG...
4. Missense	A change in DNA sequence that changes the codon to a different amino acid. Not all missense mutations are deleterious, some changes can have no effect. Because of the ambiguity of missense mutations, it is often difficult to interpret the consequences of these mutations in causing disease.	...ACT CAG AAC... ...Thr Gln Asn... ...ACT CGG AAC... ...Thr Arg Asn...
5. Nonsense	A change in the genetic code that results in the coding for a stop codon rather than an amino acid. The shortened protein is generally non-function or its	...ATA CGA GCT... ...Ile Arg Ala... ...ATA TGA GCT... ...Ile Stop

	function is impeded.	
6. Point	A single base change in DNA sequence. A point mutation may be silent, missense, or nonsense.	...CGTAATCCTCGA... ...CGTAGTCCTCGA...
7. Silent	A change in the genetic sequence that does not change the protein sequence. This can occur because of redundancy in the genetic code where an amino acid may be encoded for by multiple codons.	...TTC TGT AGT GGT... ...Phe Cys Ser Gly... ...TTC TGC AGT GGT... ...Phe Cys Ser Gly...
8. Splice Site	A change in the genetic sequence that occurs at the boundary of the exons and introns. The consensus sequences at these boundaries signal where to cut out introns and rejoin exons in the mRNA. A change in these sequences can eliminate splicing at that site which would change the reading frame and protein sequence.	
9. Translocation	A structural abnormality of chromosomes where genetic material is exchanged between two or more non-homologous chromosomes.	

POSSIBLE QUESTIONS

UNIT I

Part-B

1. Write a note on Repetitive DNA
2. Explain the Single copy DNA
3. Narrate the mobile DNA elements
4. Explain the transposition mechanism ?
5. Write short note on viral trasposans.
6. Explain the bacterial IS elements
7. Add a note on functional elements in genome
8. What are SINES and LINES . Mention their biological significance.
9. What is the basic unit of genome organization. Explain it.
10. What are intercalating agents? Describe their role in mutation

Part-C

1. Explain the structural organization of eukaryotic chromosomes.
2. How mutations occur? Explain the various types of mutation.
3. What are transposons?Explain the mechanism of transposition
4. Give an account on IS elements, Tn transposons and composite transposons
5. Write short notes on
   i) Single copy DNA
   ii) Tandemly repeated genes
   iii) Pseudogenes

POSSIBLE QUESTIONS

UNIT-I

PART-A (20 MARKS)

(Q.NO 1 TO 20 Online Examination)

PART-B (2 MARKS)

1. Define the term precision.

2. Define the term accuracy.

3. Write a note on Quality Assurance.

4. Define Quality control.

5. Define Trueness.

PART-C (6 MARKS)

1. Explain about the Quality control in clinical biochemistry and its classifications.

2. What are the safety regulations carried out in biochemistrylaboratories.

3. Write about the various different methods for collection of bloodand how they are preserved.

4. Explain the pre-analytical phase of laboratory diagnostic process.

5. Describe about the analytical phase of laboratory diagnostic process

6. Derive the relationship between precision and Trueness of analytical methods.

7. Write about the post-analytical phase of laboratory diagnostic process.

8. Write about the differences between Quality control and Quality assurance.

9. Explain in detail about the Diagnostic sensitivity and specificity methods forlaboratory screening.