SATHYABAMA
INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)
Accredited "A" Grade by NAAC | 12B Status by UGC | Approved by AICTE
www.sathyabama.ac.in

SCHOOL OF BIO AND CHEMICAL ENGINEERING

DEPARTMENT OF BIOTECHNOLOGY

SBTA1301 – MOLECULAR BIOLOGY AND GENETICS

SBTA1301	MOLECULAR BIOLOGY AND GENETICS	L	T	P	Credit	Total Marks
		3	*	0	3	100

COURSE OBJECTIVES

- The course aims to give an understanding on the fundamentals of conventional genetics and the intricate molecular mechanisms of heredity and variations.
- To understand storage of genetic information and its translation at molecular level in prokaryotic and eukaryotic systems.

UNIT 1 CLASSICAL GENETICS **9 Hrs.**

Fundamental principles of genetics- Mendel's principles and experiments, gene interaction, multiple alleles, complementation, linkage, sex linked, sex limited and sex influenced inheritance; Chromosomes basis of heredity- extra-chromosomal inheritance; Linkage and crossing over; Hardy-Weinberg equilibrium, Extensions of Hardy- Weinberg equilibrium.

UNIT 2 STRUCTURE AND PROPERTIES OF NUCLEIC ACIDS **9 Hrs.**

Introduction to nucleic acids: Evidence for DNA&RNA as a genetic material; Structure and physicochemical properties of elements in DNA and RNA, Biological significance of differences in DNA and RNA. Primary structure of DNA: Chemical and structural qualities of 3',5'-Phosphodiester bond. Secondary Structure of DNA: Watson & Crick model, Chargaff's rule; DNA replication- Overview of differences in prokaryotic and eukaryotic DNA replication, D-loop and rolling circle mode of replication, Telomere replication in eukaryotes; Okazaki fragments, Fidelity of DNA replication, Inhibitors of DNA replication, DNA repair- Mutagens, DNA mutations and various types of repair mechanisms.

UNIT 3 TRANSCRIPTION **9 Hrs.**

Central Dogma in molecular biology -Structure and function of mRNA, rRNA tRNA and micro RNAs. Characteristics of promoter and enhancer sequences. RNA synthesis: Initiation, elongation and termination of RNA synthesis, Proteins of RNA synthesis, Fidelity of RNA synthesis, Inhibitors of transcription, Differences in prokaryotic and eukaryotic transcription. Basic concepts in RNA world: RNA processing: 5'-Capping, Splicing-Alternative splicing, Poly 'A' tail addition and base modification.

UNIT 4 TRANSLATION **9 Hrs.**

Introduction to Genetic code: Elucidation of genetic code, Codon degeneracy, Wobble hypothesis and its importance, Prokaryotic and eukaryotic ribosomes. Steps in translation: Initiation, Elongation and termination of protein synthesis. Inhibitors of protein synthesis. Post-translational modifications and its importance.

UNIT 5 REGULATION OF GENE EXPRESSION **9 Hrs.**

Organization of genes in prokaryotic and eukaryotic chromosomes- operon concept; Gene expression and regulation- Hierarchical levels of gene regulation, Prokaryotic gene regulation –lac and trp operon, Eukaryotic gene regulation- gene silencing.

Max.45 Hrs.

COURSE OUTCOMES

On completion of the course, student will be able to

- CO1 - Explain the foundations of Mendelian genetics and chromosomal theory and apply these, with appropriate terminology, to contemporary concepts in genetics.
- CO2 - Emphasize the molecular mechanism of DNA replication and repair in various organisms
- CO3 - Explain the properties of genetic materials and storage and processing of genetic information.
- CO4 - Analyze the processes of transcription and translation in both prokaryotes and eukaryotes at molecular level.
- CO5 - Understand the redundant and universal qualities of the genetic code and how it is used to determine the amino acid sequence of a polypeptide.
- CO6 - Compare the mechanisms of gene regulation in prokaryotes and eukaryotes.

TEXT / REFERENCE BOOKS

1. Lewin B., Genes XI, International Edition, Jocelyn Krebs, Stephen Kilpatrick and Elliott Goldstein, Jones & Bartlett Learning, 2017, ISBN 978-1-4496-5985-1
2. Tropp, Burton E., Molecular Biology: Genes to Proteins, 3rd Edition, Jones and Bartlett, 2008.
3. Glick B.R. and Pasternak J.J., Molecular Biotechnology: Principles and Applications of Recombinant DNA, 4th Edition. ASM, 2010

UNIT – I – CLASSICAL GENETICS

Mendelian Inheritance

Mendelian inheritance is inheritance of biological features that follows the laws proposed by Gregor Johann Mendel in 1865 and 1866 and re-discovered in 1900. It was initially very controversial. When Mendel's theories were integrated with the Boveri–Sutton chromosome theory of inheritance by Thomas Hunt Morgan in 1915, they became the core of classical genetics while Ronald Fishercombined them with the theory of natural selection in his 1930 book *The Genetical Theory of Natural Selection*, putting evolutiononto a mathematical footing and forming the basis for Population genetics and the modern evolutionary synthesis.

History

The laws of inheritance were derived by Gregor Mendel, a nineteenth-century Austrian monk conducting hybridization experiments in garden peas (*Pisum sativum*) he planted in the backyard of the church. Between 1856 and 1863, he cultivated and tested some 5,000 pea plants. From these experiments, he induced two generalizations which later became known as *Mendel's Principles of Heredity* or *Mendelian inheritance*. He described these principles in a two-part paper, *Versuche über Pflanzen-Hybriden*(*Experiments on Plant Hybridization*), that he read to the Natural History Society of Brno on February 8 and March 8, 1865, and which was published in 1866.

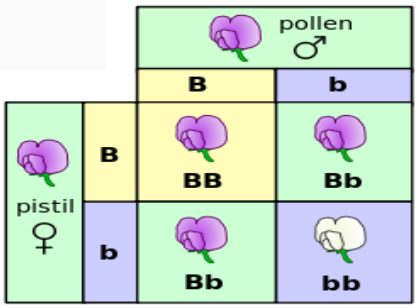
Mendel's conclusions were largely ignored. Although they were not completely unknown to biologists of the time, they were not seen as generally applicable, even by Mendel himself, who thought they only applied to certain categories of species or traits. A major block to understanding their significance was the importance attached by 19th-century biologists to the apparent blending of inherited traits in the overall appearance of the progeny, now known to be due to multigene interactions, in contrast to the organ-specific binary characters studied by Mendel. In 1900, however, his work was "re-discovered" by three European scientists, Hugo de Vries, Carl Correns, and Erich von Tschermak. The exact nature of the "re-discovery" has been somewhat debated: De Vries published first on the subject, mentioning Mendel in a footnote, while Correns pointed out Mendel's priority after having read De Vries' paper and realizing that he himself did not have priority. De Vries may not have acknowledged truthfully how much of his knowledge of the laws came from his own work, or came only after reading Mendel's paper. Later scholars have accused Von Tschermak of not truly understanding the results at all.

Regardless, the "re-discovery" made Mendelism an important but controversial theory. Its most vigorous promoter in Europe wasWilliam Bateson, who coined the terms "genetics" and "allele" to describe many of its tenets. The model of heredity was highly contested by other biologists because it implied that heredity was discontinuous, in opposition to the apparently continuous variation observable for many traits. Many biologists also dismissed the theory because they were not sure it would apply to all species. However, later work by biologists and statisticians such as Ronald Fisher showed that if multiple Mendelian factors were involved in the expression of an individual trait, they could produce the diverse results observed, and thus showed that Mendelian genetics is compatible with natural selection. Thomas Hunt Morgan and his assistants later integrated the theoretical model of Mendel with the chromosome theory of inheritance, in which the chromosomes of cells were thought to hold the actual hereditary material, and created what is now known as classical genetics, which was extremely successful and cemented Mendel's place in history.

Mendel's findings allowed scientists such as Fisher and J.B.S. Haldane to predict the expression of traits on the basis of mathematical probabilities. A large contribution to Mendel's success can be traced to his decision to start his crosses only with plants he demonstrated were true-breeding. He also only measured absolute (binary) characteristics, such as color, shape, and position of the offspring, rather than quantitative characteristics. He expressed his results numerically and subjected them to statistical analysis. His method of data analysis and his large sample size gave credibility to his data. He also had the foresight to follow several successive generations (f2, f3) of pea plants and record their variations. Finally, he performed "test crosses" (back-crossing descendants of the initial hybridization to the initial true-breeding lines) to reveal the presence and proportion of recessive characters.

Mendel's laws

A Punnett square for one of Mendel's pea plant experiments.



Mendel discovered that, when he crossed purebred white flower and purple flower pea plants (the parental or P generation), the result was not a blend. Rather than being a mix of the two, the offspring (known as the F_1 generation) was purple-flowered. When Mendel self-fertilized the F_1 generation pea plants, he obtained a purple flower to white flower ratio in the F_2 generation of 3 to 1. The results of this cross are tabulated in the Punnett square to the right.

He then conceived the idea of heredity units, which he called "factors". Mendel found that there are alternative forms of factors—now called genes—that account for variations in inherited characteristics. For example, the gene for flower color in pea plants exists in two forms, one for purple and the other for white. The alternative "forms" are now called alleles. For each biological trait, an organism inherits two alleles, one from each parent. These alleles may be the same or different. An organism that has two identical alleles for a gene is said to be homozygous for that gene (and is called a homozygote). An organism that has two different alleles for a gene is said be heterozygous for that gene (and is called a heterozygote).

Mendel also hypothesized that allele pairs separate randomly, or segregate, from each other during the production of gametes: egg and sperm. Because allele pairs separate during gamete production, a sperm or egg carries only one allele for each inherited trait. When sperm and egg unite at fertilization, each contributes its allele, restoring the paired condition in the offspring. This is called the**Law of Segregation**. Mendel also found that each pair of alleles segregates independently of the other pairs of alleles during gamete formation. This is known as the **Law of Independent Assortment**.

The genotype of an individual is made up of the many alleles it possesses. An individual's physical appearance, or phenotype, is determined by its alleles as well as by its environment. The presence of an allele does not mean that the trait will be expressed in the individual that possesses it. If the two alleles of an inherited pair differ (the heterozygous condition), then one determines the organism’s appearance and is called the dominant allele; the other has no noticeable effect on the organism’s appearance and is called the recessive allele. Thus, in the example above dominant purple flower allele will hide the phenotypic effects of the recessive white flower allele. This is known as the **Law of Dominance** but it is not a transmission law, dominance has to do with the expression of the genotype and not its transmission. The upper case letters are used to represent dominant alleles whereas the lowercase letters are used to represent recessive alleles.

Mendel's laws of inheritance

Law	Definition
Law of segregation	During gamete formation, the alleles for each gene segregate from each other so that each gamete carries only one allele for each gene.
Law of independent assortment	Genes for different traits can segregate independently during the formation of gametes.
Law of dominance	Some alleles are dominant while others are recessive; an organism with at least one dominant allele will display the effect of the dominant allele.

In the pea plant example above, the capital "P" represents the dominant allele for purple flowers and lowercase "p" represents the recessive allele for white flowers. Both parental plants were true-breeding, and one parental variety had two alleles for purple flowers (*PP*) while the other had two alleles for white flowers (*pp*). As a result of fertilization, the F_1hybrids each inherited one allele for purple flowers and one for white. All the F_1 hybrids (*Pp*) had purple flowers, because the dominant *P* allele has its full effect in the heterozygote, while the recessive *p* allele has no effect on flower color. For the F_2 plants, the ratio of plants with purple flowers to those with white flowers (3:1) is called the phenotypic ratio. The genotypic ratio, as seen in the Punnett square, is 1 *PP* : 2 *Pp* : 1 *pp*.

Law of Segregation (the "First Law")

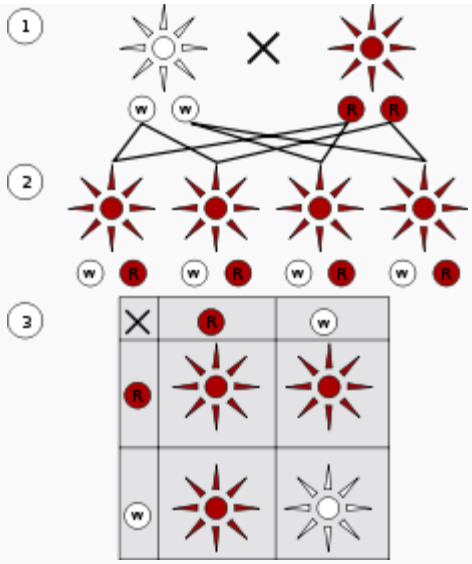


Figure 1 Dominant and recessive phenotypes.

(1) Parental generation.

(2) F_1 generation.

(3) F_2 generation. Dominant (red) and recessive (white) phenotype look alike in the F_1 (first) generation and show a 3:1 ratio in the F_2 (second) generation.

The Law of Segregation states that every individual organism contains two alleles for each trait, and that these alleles segregate (separate) during meiosis such that each gamete contains only one of the alleles. An offspring thus receives a pair of alleles for a trait by inheriting homologous chromosomes from the parent organisms: one allele for each trait from each parent.

Molecular proof of this principle was subsequently found through observation of meiosis by two scientists independently, the German botanist Oscar Hertwig in 1876, and the Belgian zoologist Edouard Van Beneden in 1883. Paternal and maternal chromosomes get separated in meiosis and the alleles with the traits of a character are segregated into two different gametes. Each parent contributes a single gamete, and thus a single, randomly successful allele copy to their offspring and fertilization.

Law of Independent Assortment (the "Second Law")

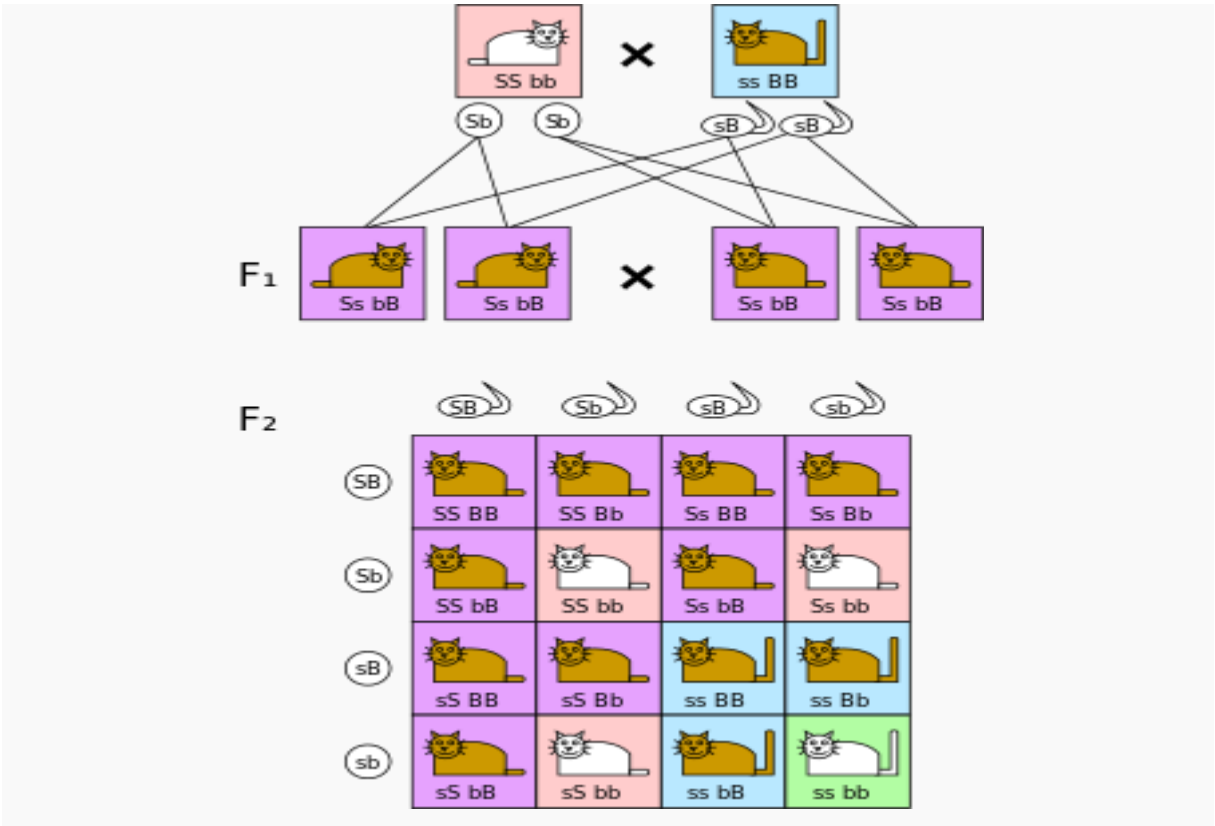


Figure 2 Dihybrid cross. The phenotypes of two independent traits show a 9:3:3:1 ratio in the F_2 generation. In this example, coat color is indicated by **B** (brown, dominant) or **b** (white), while tail length is indicated by **S** (short, dominant) or **s** (long). When parents are homozygous for each trait (**SSbb** and **ssBB**), their children in the F_1 generation are heterozygous at both loci and only show the dominant phenotypes (**SsbB**). If the children mate with each other, in the F_2 generation all combinations of coat color and tail length occur: 9 are brown/short (purple boxes), 3 are white/short (pink boxes), 3 are brown/long (blue boxes) and 1 is white/long (green box).

The Law of Independent Assortment states that alleles for separate traits are passed independently of one another from parents to offspring. That is, the biological selection of an allele for one trait has nothing to do with the selection of an allele for any other trait. Mendel found support for this law in his dihybrid cross experiments (Fig. 1). In his monohybrid crosses, an idealized 3:1 ratio between dominant and recessive phenotypes resulted. In dihybrid crosses, however, he found a 9:3:3:1 ratios (Fig. 2). This shows that each of the two alleles is inherited independently from the other, with a 3:1 phenotypic ratio for each.

Independent assortment occurs in eukaryotic organisms during meiotic prophase I, and produces a gamete with a mixture of the organism's chromosomes. The physical basis of the independent assortment of chromosomes is the random orientation of each bivalent chromosome along the metaphase plate with respect to the other bivalent chromosomes. Along with crossing over, independent assortment increases genetic diversity by producing novel genetic combinations.

The are many violations of independent assortment due to genetic linkage.

Of the 46 chromosomes in a normal diploid human cell, half are maternally derived (from the mother's egg) and half are paternally derived (from the father's sperm). This occurs as sexual reproduction involves the fusion of two haploid gametes (the egg and sperm) to produce a new organism having the full complement of chromosomes. During gametogenesis—the production of new gametes by an adult—the normal complement of 46 chromosomes needs to be halved to 23 to ensure that the resulting haploid gamete can join with another gamete to produce a diploid organism. An error in the number of chromosomes, such as those caused by a diploid gamete joining with a haploid gamete, is termed aneuploidy.

In independent assortment, the chromosomes that result are randomly sorted from all possible combinations of maternal and paternal chromosomes. Because gametes end up with a random mix

instead of a pre-defined "set" from either parent, gametes are therefore considered assorted independently. As such, the gamete can end up with any combination of paternal or maternal chromosomes. Any of the possible combinations of gametes formed from maternal and paternal chromosomes will occur with equal frequency. For human gametes, with 23 pairs of chromosomes, the number of possibilities is 2^{23} or 8,388,608 possible combinations.[4] The gametes will normally end up with 23 chromosomes, but the origin of any particular one will be randomly selected from paternal or maternal chromosomes. This contributes to the genetic variability of progeny.

Law of Dominance (the "Third Law")

Mendel's Law of Dominance states that recessive alleles will always be masked by dominant alleles. Therefore, a cross between a homozygous dominant and a homozygous recessive will always express the dominant phenotype, while still having a heterozygous genotype. Law of Dominance can be explained easily with the help of a mono hybrid cross experiment:- In a cross between two organisms pure for any pair (or pairs) of contrasting traits (characters), the character that appears in the F1 generation is called "dominant" and the one which is suppressed (not expressed) is called "recessive." Each character is controlled by a pair of dissimilar factors. Only one of the characters expresses. The one which expresses in the F1 generation is called Dominant. It is important to note however, that the law of dominance is significant and true but is not universally applicable.

According to the latest revisions, only two of these rules are considered to be laws. The third one is considered as a basic principle but not a genetic law of Mendel.

Mendelian trait

A Mendelian trait is one that is controlled by a single locus in an inheritance pattern. In such cases, a mutation in a single gene can cause a disease that is inherited according to Mendel's laws. Examples include sickle-cell anemia, Tay-Sachs disease, cystic fibrosis and xeroderma pigmentosa. A disease controlled by a single gene contrasts with a multi-factorial disease, like arthritis, which is affected by several loci (and the environment) as well as those diseases inherited in a non-Mendelian fashion.

. **Variations to Mendelian Genetics**

- Not all genetic traits strictly follow the laws discovered by Gregor Mendel.
- Some variations can be observed in all animals including humans.
- Four types:
  - Incomplete dominance
  - Codominance
  - Polygenic inheritance
  - Sex linked traits and/or Sex Influenced

Incomplete & Codominance

Incomplete Dominance

In many ways Gregor Mendel was quite lucky in discovering his genetic laws. He happened to use pea plants, which happened to have a number of easily observable traits that were determined by just two alleles. And for the traits he studied in his peas, one allele happened to be dominant for the trait & the other was a recessive form. Things aren't always so clear-cut & "simple" in the world of genetics, but luckily for Mendel (& the science world) he happened to work with an organism whose genetic make-up was fairly clear-cut & simple.

COMPLETE DOMINANCE

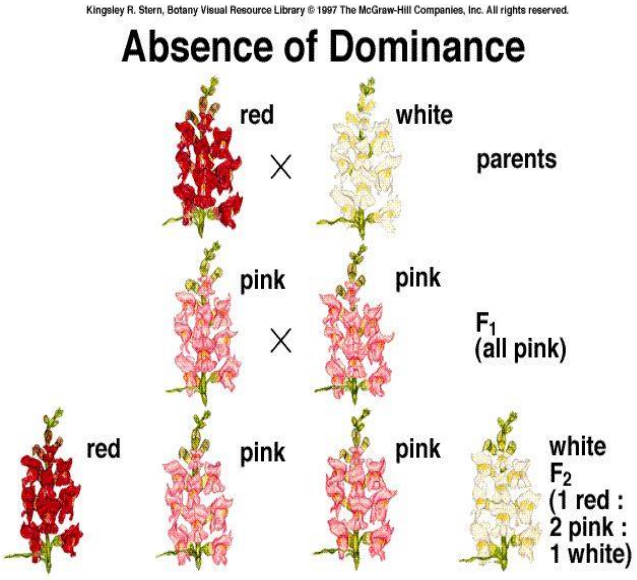
If Mendel were given a mommy black mouse & a daddy white mouse & asked what their offspring would look like, he would've said that a certain percent would be black & the others would be white. He would never have even considered that a white mouse & a black mouse could produce a *GREY* mouse! For Mendel, the phenotype of the offspring from parents with different phenotypes always resembled the phenotype of at least one of the parents. In other words, Mendel was unaware of the phenomenon of INCOMPLETE DOMINANCE.

I remember Incomplete Dominance in the form of an example like so:
RED Flower x WHITE Flower ---> PINK Flower

With incomplete dominance, a cross between organisms with two different phenotypes produces offspring with a third phenotype that is a blending of the parental traits.

It's like mixing paints, red + white will make pink. Red doesn't totally block (dominate) the pink, instead there is *incomplete*dominance, and we end up with something in-between. We can still use the Punnett Square to solve problems involving incomplete dominance. The only difference is that instead of using a capital letter for the dominant trait & a lowercase letter for the recessive trait, the letters we use are both going to be capital (because neither trait dominates the other). So the cross I used up above would look like this:

	R	r
R	R R	R r
r	R r	r r



The trick is to *recognize* when you are dealing with a question involving incomplete dominance. There are two steps to this:
1) Notice that the offspring is showing a 3rd phenotype. The parents each have one, and

the offspring are different from the parents.

2) Notice that the trait in the offspring is a blend (mixing) of the parental traits.

Codominance

The genetic gist to codominance is pretty much the same as incomplete dominance. A hybrid organism shows a third phenotype --- not the usual "dominant" one & not the "recessive" one ... but a third, *different* phenotype. With incomplete dominance we get a blending of the dominant & recessive traits so that the third phenotype is something in the middle (red x white = pink).

In COdominance, the "recessive" & "dominant" traits appear together in the phenotype of hybrid organisms.

I remember codominance in the form of an example like so:

red x white ---> red & white spotted

With codominance, a cross between organisms with two different phenotypes produces offspring with a third phenotype in which both of the parental traits appear together.

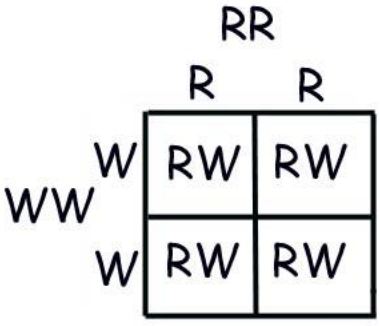
When it comes to punnett squares & symbols, it's the same as incomplete dominance. Use capital letters for the allele symbols. My example cross from above would look like so:

R = allele for red flowers

W = allele for white flowers

red x white ---> red & white spotted

RR x WW ---> 100% RW



We'll use "F" for the flower color allele.

F^R = allele for red flowers

F^W = allele for white flowers

red x white -------> red & white spotted flowers

F^RF^R x F^WF^W ----> 100% F^RF^W

The symbols you choose to use don't matter, in the end you end up with hybrid organisms, and rather than one trait (allele) dominating the other, both traits appear together in the phenotype. Wa-la, codominance.

A very very very very very common phenotype used in questions about codominance is roan fur in cattle. Cattle can be red (RR = all red hairs), white (WW = all white hairs), or roan (RW = red & white hairs together). A good example of codominance.

Another example of codominance is human blood type AB, in which two types of protein ("A" & "B") appear together on the surface of blood cells.

Multiple Alleles

It makes absolutely no sense whatsoever to continue if we don't know what the word "allele" means.

allele = (*n)* a form of a gene which codes for one possible outcome of a phenotype

For example, in Mendel's pea investigations, he found that there was a gene that determined the color of the pea pod. One form of it (one allele) creates yellow pods, & the other form (allele) creates green pods.

Get it? Two possible phenotypes of one trait (pod color) are determined by two alleles (forms) of the one "color" gene.

SOME BACKGROUND

When the gene for one trait exists as only two alleles & the alleles play according to Mendel's Law of Dominance, there are 3 possible genotypes (combination of alleles) & 2 possible phenotypes (the dominant one or the recessive one).

Using the pea pod trait as an example, the possibilities are like so:

GENOTYPES	RESULTING PHENOTYPE
Homozygous Dominant (YY)	Yellow
Heterozygous (Yy)	Yellow
Homozygous Recessive (yy)	Green

where
Y = the dominant allele for yellow & y = the recessive allele for green

---

If there are only two alleles involved in determining the phenotype of a certain trait, but there are *three* possible phenotypes, then the inheritance of the trait illustrates either incomplete dominance or codominance.

In these situations a heterozygous (hybrid) genotype produces a 3rd phenotype that is either a blend of the other two phenotypes (incomplete dominance) or a mixing of the other phenotypes with both appearing at the same time (codominance).

Here's an example with Incomplete Dominance:

GENOTYPES	RESULTING PHENOTYPE
BB = Homozygous Black	Black Fur
BW = Heterozygous	Grey Fur
WW = Homozygous White	White Fur

where
B = allele for black &
W = allele for white

And here's an example with Codominance:

GENOTYPES	RESULTING PHENOTYPE
BB = Homozygous Black	Black Fur
BW = Heterozygous	Black & White Fur
WW = Homozygous White	White Fur

where
B = allele for black &
W = allele for white

HE DEALS ON MULTIPLE ALLELES

Now, if there are *4* or more possible phenotypes for a particular trait, then more than 2 alleles for that trait must exist in the population. We call this "MULTIPLE ALLELES".

Let me stress something. There may be multiple alleles within the population, but individuals have only two of those alleles.

Why?

Because individuals have only two biological parents. We inherit half of our genes (alleles) from ma, & the other half from pa, so we end up with two alleles for every trait in our phenotype.

An excellent example of multiple allele inheritance is human blood type. Blood type exists as four possible phenotypes: A, B, AB, & O.

There are 3 alleles for the gene that determines blood type.
(Remember: You have just 2 of the 3 in your genotype --- 1 from mom & 1 from dad).

The alleles are as follows:

ALLELE	CODES FOR
I^A	Type "A" Blood
I^B	Type "B" Blood
i	Type "O" Blood

Notice that, according to the symbols used in the table above, that the allele for "O" (i) is recessive to the alleles for "A" & "B".

With three alleles we have a higher number of possible combinations in creating a genotype.

GENOTYPES	RESULTING PHENOTYPES
I^AI^A	Type A
I^Ai	Type A
I^BI^B	Type B
I^Bi	Type B
I^AI^B	Type AB
ii	Type O

Notes:

- As you can count, there are 6 different genotypes & 4 different phenotypes for blood type.
- Note that there are two genotypes for both "A" & "B" blood --- either homozygous (I^AI^A or I^BI^B) or heterozygous with one recessive allele for "O" (I^Ai or I^Bi).
- Note too that the only genotype for "O" blood is homozygous recessive (ii).

And lastly, what's the deal with "AB" blood? What is this an example of? The "A" trait & the "B" trait appear together in the phenotype. Think think think ...

Gene Interactions

Definition

With the help of lot of experiments it was found that most of the characters of living organisms are controlled / influenced / governed by a collaboration of several different genes.

This condition where a single character is governed by two or more genes and every gene affect the expression of other genes involved (means these genes affect each others expression) is known as gene interaction.

In simple way we could say that, in gene interaction, expression of one gene depends on expression (presence or absence) of another gene.

As we know, gene interactions may involve two or more pairs of genes. But all the gene interactions we have described below have the two pairs of non-allelic genes, affecting the phenotypic expression of same character. These interactions produce modified dihybrid ratios.

Types of Gene Interactions

Gene interactions can be classified as

- Allelic gene interaction
- Non-allelic gene interaction

Non-allelic gene interaction

Expression of character is produced by interaction between two or more genes. The interactions we have listed below, as inter and intra allelic are of this type.

- Inter-allelic

- Intra-allelic

Inter-allelic

- Without modification of normal F_2 ratio
- With modification of normal F_2 ratio

Such kinds of interactions modify the normal F_2 ratio (9:3:3:1). Various types of such interactions are as below.

Gene Interaction	F_2 Ratio	Test Cross Ratio
Complementary Gene Interaction	9:7	1:3
Supplementary Gene Interaction	9:3:4	1:1:2
Epistasis	12:3:1	2:1:1
Duplicate Factor	15:1	3:1
Inhibitory Factor	13:3	1:3
Polymerism or Additive Factor	9:6:1	1:2:1

Intra-allelic

- Lethal Gene

Allelic gene interaction

Expression of character is produced by interaction between alleles of a single gene.

- Complete dominance
- Incomplete dominance
- Co-dominance
- Over Dominance

Epistasis

Definition

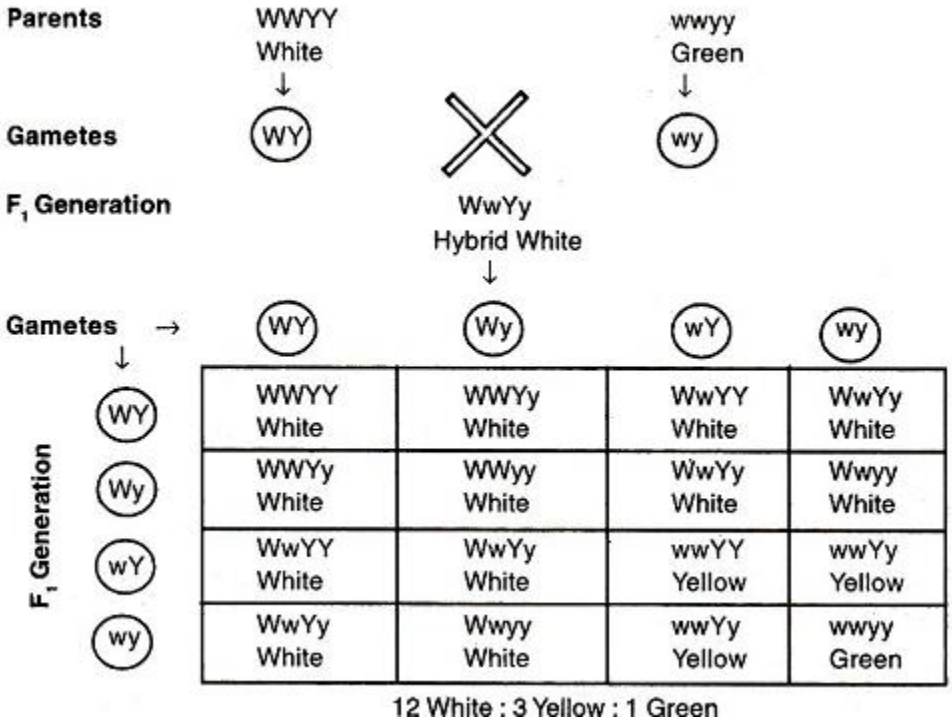
- Involves two pairs of non-allelic genes
- Both the dominant genes affect the same character
- One of them, when present alone or along with other dominant gene expresses itself
- The other gene expresses itself only when it is alone.
- The recessive forms of both the genes give rise to different phenotype

- The phenomenon of masking of effect of one dominant gene by the other gene is called as "epistasis", and hence the interaction is named as epistasis. The masking gene is epistatic gene and the masked one is hypostatic gene.

Dominant Epistasis

- Epistasis is observed in fruit color of summer squash (white, yellow and green)
  White and yellow fruit colors are monogenic dominant over green color. When plant with white fruit is crossed with that with green fruit, progeny had white fruits.
  But, in F_2 segregation in the ratio of 12white: 3yellow: 1green was observed. This behavior is based on ratio of dihybrid cross, so the trait must be governed by two pairs of genes.
  Suppose, gene W, gene Y and gene w produce white, yellow and green colors resp. Now, the genotype of pure white fruit plant must be WWYY as yellow fruited plants are produced in F_2. The gene for yellow color compulsorily should be present there. Both the recessive genes in homozygous condition produce green phenotype. So, genotype of pure breeding green plant must be wwyy. The F_2 segregation shows that the effect of gene Y is masked by gene W, when both are present together.

Cross between	**F_1**	**F_2**
WWYY X wwyy white X green	WwYy white	12 white : 3 yellow: 1 green



Recessive Episatsis

Recessive alleles at one locus (aa) mask the phenotypic expression of other gene locus (BB, Bb or bb) such epistasis is called recessive epistasis. The alleles of 'B' locus express themselves only when epistatic locus 'A' has dominant allele like AA or Aa. The phenotypic ratio is 9: 3: 4.

Example:

Complete dominance at both gene pairs, but one gene, when homozygous recessive, is epistatic to the other.

In Mouse coat colour.

Gene pair A: colour dominant over albino.

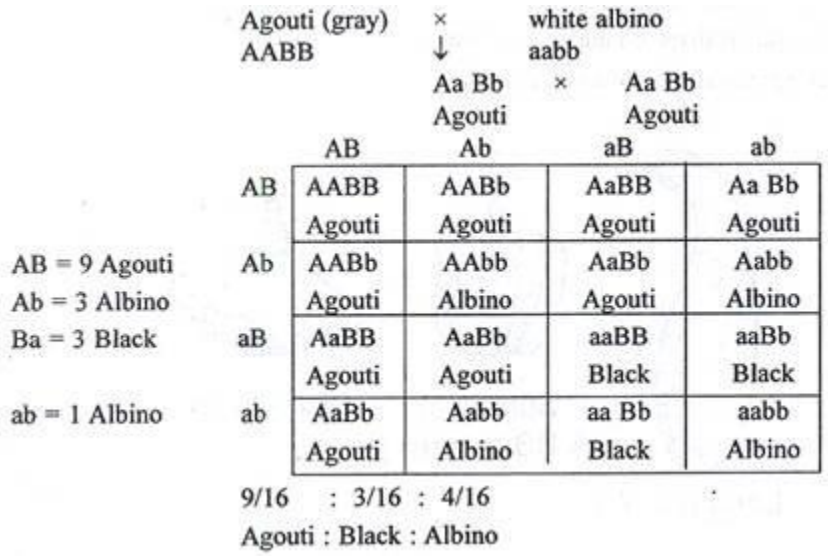
Gene pair 'B' agouti colour dominant over black.

Interaction: homozygous albino is epistatic to aguoti and black.

Agouti 9/16

Black 3/16

Albino 4/16



Duplicate Factor

Definition

- Involves two pairs of non-allelic genes
- Affect the same character
- Dominant form of both genes, either alone or in combination produce same phenotypic expression.
- Recessive forms of both the genes when together produce different phenotypic effect.
- Duplicate factor interaction is observed in awned and awnless condition in rice.

  In case of rice when pure breeding awned plant is crossed with awnless, the progeny was awned. In F_2 segregation was observed in the ratio of 15 awned: 1 awnless. This behavior is based on ratio of dihybrid cross, so the trait must be governed by two pairs of genes.

  Suppose, A1 and A2 are two duplicate factors.

  Genotype of pure breeding awned plant must be $A_1A_1A_2A_2$, as it is dominant and genotype of awnless plant must be a1a1a2a2 as it is recessive.

Cross between	**F_1**	**F_2**
$A_1A_1A_2A_2$ X $a_1a_1a_2a_2$ awned X awnless	$A_1a_1A_2a_2$ awned	15 awned : 1 awnless

Complementary Gene Interaction

Definition

- Involves two pairs of non-allelic genes
- When dominant forms of both the genes involved in complementary gene interaction are alone have same phenotypic expression
- But, if they are present in combination, yield different phenotypic effect.
- Flower color in garden pea follow this type of gene interaction.

  We have considered garden pea for the explanation of this type of gene interaction, in which it was noted for the first time. Two different varieties of garden pea produce same color flowers – white. But on crossing they yield purple color flowers. Again in F_2, 9 purple : 7 white segregation was observed. How this happened? The answer is complementary gene interaction. For the sake of understanding we will name the two varieties as variety I and II.

  Suppose, Gene W in variety I and gene C in variety II produces white flowers. But, the progeny resulting from cross between these two has purple flowers. This means both the dominant genes, W and C together produce purple color flowers.

  In the variety I, dominant C gene is absent but it posses recessive c gene. Similarly, variety II contains recessive w gene. Both recessive genes produce white flowers, when present together. As in this type of interaction, the two recessive genes complement each other, it is called as complementary gene interaction.

Cross between	**F_1**	**F_2**
WWcc X wwCC white X white	WwCc purple	9 purple : 7 white

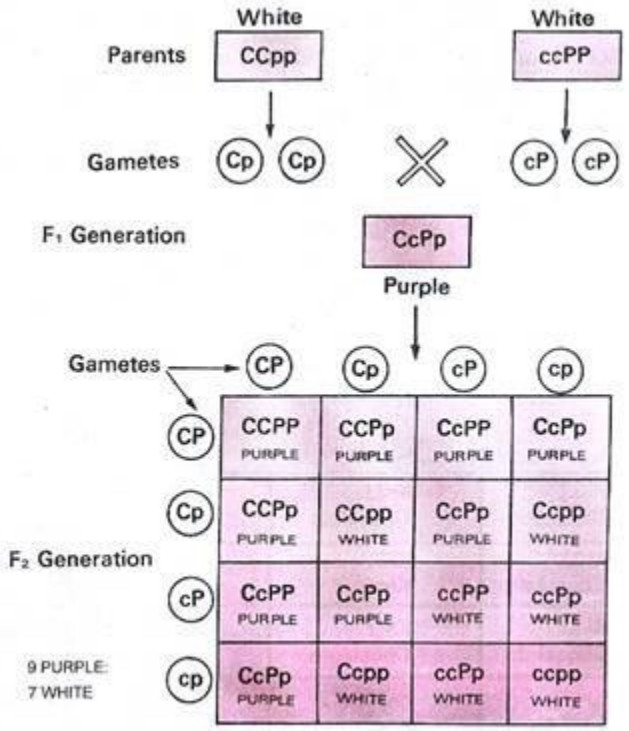


Fig. 5. 32. Inheritance of purple colour in Sweet Pea (*Lathyrus odoratus*) due to complementary and epistatic interactions of two nonallelic gene pairs.

Supplementary Gene Interaction

Definition

- Involves two pairs of non-allelic genes
- Affect the same character
- One of the dominant gene has visible effect itself
- Second dominant gene expresses itself when supplemented by the other dominant gene of a pair
- Coat color (black, albino and agouti) of mice follows supplementary gene interaction.

  In mice, black coat color is monogenically dominant over albino and agouti. The offspring resulting from the cross between black and albino has agouti coat color.

  F_2 generation shows segregation in the ratio 9 agouti: 3 black: 4 albino. This behavior is based on ratio of dihybrid cross, so the trait must be governed by two pairs of genes.

  Suppose, gene C is essential for the development of black coat color, so present in black mice and absent in albino mice. Albino mice contains only gene A, so produces albino phenotype.

  But, when gene A is present along with gene C, produces agouti phenotype. Both the genes in recessive form produce albino phenotype.

  So the cross will be as follows

Cross between	F_1	F_2
CCaa X ccAA black X albino	CcAa agouti	9 agouti : 3 black: 4albino

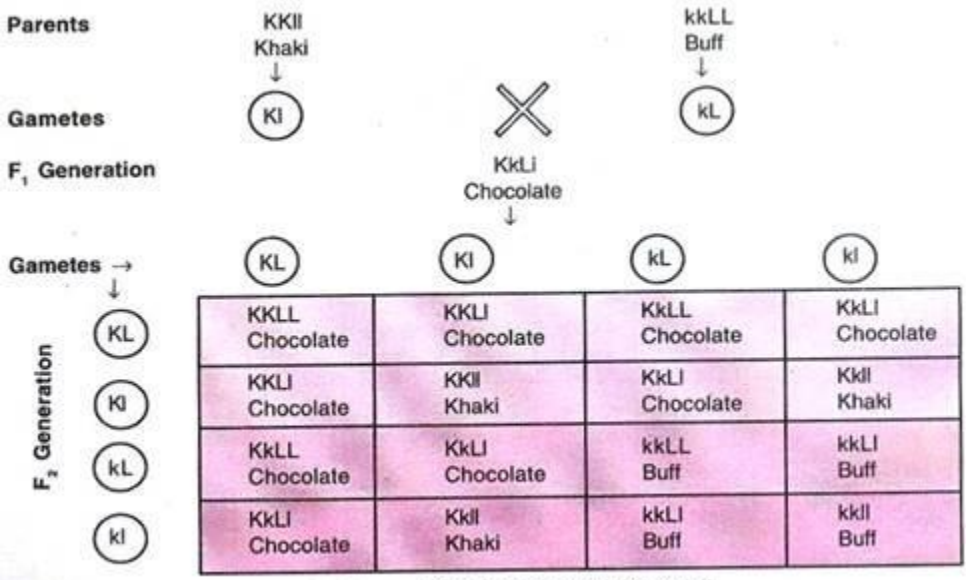


Fig. 5.33. Inheritance of seed coat colour in Lablab showing effect of supplementary gene.

Lethal Genes

Definition

Genes which result in viability reduction of individual or become a cause for death of individuals carrying them are called as lethal genes.

Phenomenon of action of lethal genes is called as **lethality**.

Some lethal genes cause death of zygote or the early embryonic stage while some express their effect in later stages of development.

Individuals carrying dominant lethal will die. Even though the dominant lethal is eliminated from the genotype of population previously carrying it, the recessive lethals are still carried in heterozygous condition. The recessive lethal in heterozygous condition reduce the viability and when they occur in homozygous condition produce lethal effect.

Usually lethal genes are dominant with respect to the phenotype they control and recessive for their lethal actions.

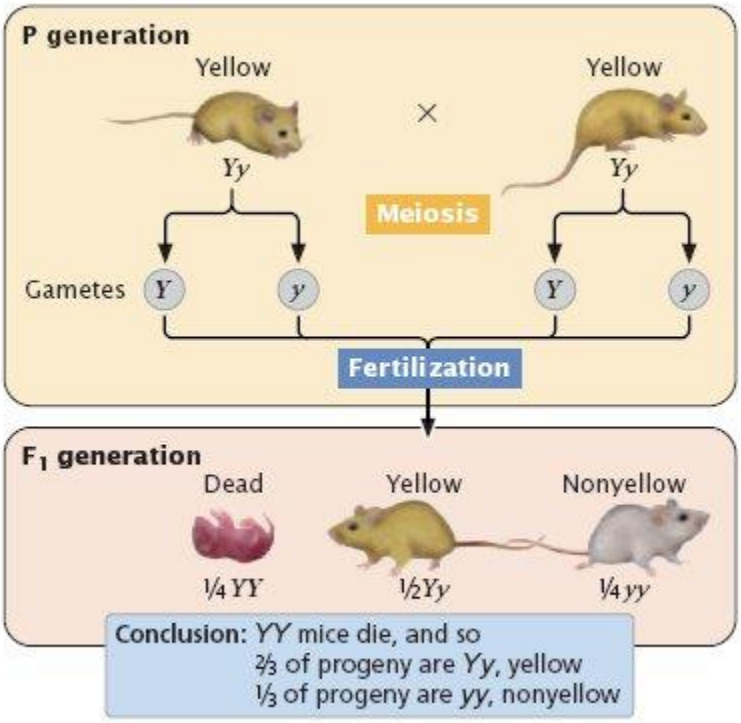
Example of Lethal Genes

In case of **mice** (By Cuenot)

In a cross of a yellow mouse with another yellow, yellow and brown mice are obtained in 2:1 ratio. True breeding yellow mice never obtained. In 1917, Stiegseder reported that 1/4th of the offspring die in embryonic condition in such crosses.

These premature dead forms are dominant homozygous.

According to Cuenot gene Y has multiple effects. It controls the yellow color of the body and affects to individual viability. It means the gene is dominant for body color and recessive with respect to lethality. The living yellow forms are heterozygous and somehow manage to survive.



In case of **maize**, albinism is example of lethal factor.

The lethal factor in heterozygous condition has no lethal effect but in homozygous condition it makes plant to die. Lethal factor in homozygous condition produces albino plant, which is unable to synthesize food due to lack of chlorophyll. Thus, the lethal factor modifies normal ratio from 3:1 to 2:1.

Chromosomal Theory of Inheritance

The Chromosomal Theory of Inheritance identified chromosomes as the genetic material responsible for Mendelian inheritance

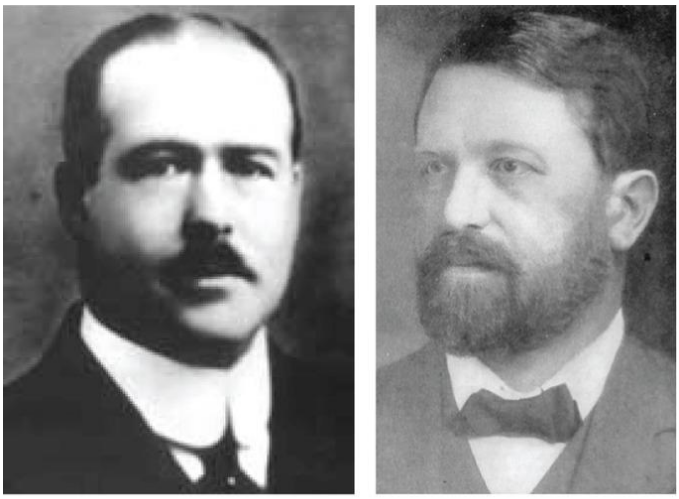
KEY POINTS

- Homologous chromosome pairs are independent of other chromosome pairs.
- Chromosomes from each homologous pair are sorted randomly into pre-gametes.
- Parents synthesize gametes that contain only half of their chromosomes; eggs and sperm have the same number of chromosomes.
- Gametic chromosomes combine during fertilization to produce offspring with the same chromosome number as their parents.
- Eye color in fruit flies was the first X-linked trait to be discovered; thus, Morgan's experiments with fruit flies solidified the Chromosomal Theory of Inheritance.

TERMS

- wild type
  the typical form of an organism, strain, gene or characteristic as it occurs in nature
- hemizygous
  having some single copies of genes in an otherwise diploid cell or organism
- autosome
  any chromosome other than sex chromosomes

The speculation that chromosomes might be the key to understanding heredity led several scientists to examine Mendel's publications and re-evaluate his model in terms of the behavior of chromosomes during mitosis and meiosis. In 1902, Theodor Boveri observed that proper embryonic development of sea urchins does not occur unless chromosomes are present. That same year, Walter Sutton observed the separation of chromosomes into daughter cells during meiosis . Together, these observations led to the development of the Chromosomal Theory of Inheritance, which identified chromosomes as the genetic material responsible for Mendelian inheritance.



(a) (b)

Sutton and Boveri

(a) Walter Sutton and (b) Theodor Boveri are credited with developing the Chromosomal Theory of Inheritance, which states that chromosomes carry the unit of heredity (genes).

The Chromosomal Theory of Inheritance was consistent with Mendel's laws and was supported by the following observations:

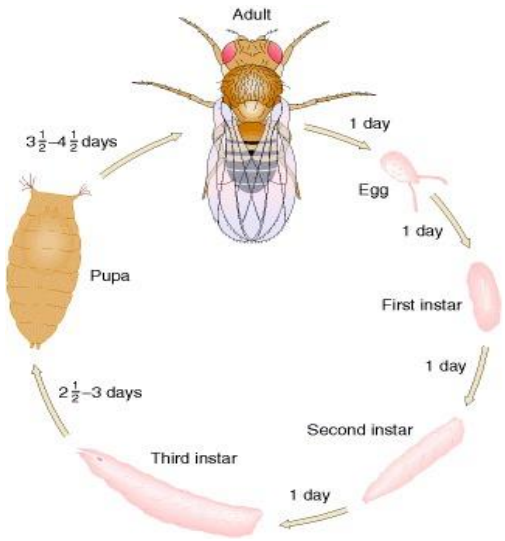
- During meiosis, homologous chromosome pairs migrate as discrete structures that are independent of other chromosome pairs.
- The sorting of chromosomes from each homologous pair into pre-gametes appears to be random.
- Each parent synthesizes gametes that contain only half of their chromosomal complement.
- Even though male and female gametes (sperm and egg) differ in size and morphology, they have the same number of chromosomes, suggesting equal genetic contributions from each parent.
- The gametic chromosomes combine during fertilization to produce offspring with the same chromosome number as their parents.

Sex chromosomes and sex-linked inheritance

Most animals and many plants show sexual dimorphism; in other words, an individual can be either male or female. In most of these cases, sex is determined by special sex chromosomes. In these organisms, there are two categories of chromosomes, **sex chromosomes** and **autosomes** (the chromosomes other than the sex chromosomes). The rules of inheritance considered so far, with the use of Mendel's analysis as an example, are the rules of autosomes. Most of the chromosomes in a genome are autosomes. The sex chromosomes are fewer in number, and, generally in diploidorganisms, there is just one pair.

Let us look at the human situation as an example. Human body cells have 46 chromosomes: 22 homologous pairs of autosomes plus 2 sex chromosomes. In females, there is a pair of identical sex chromosomes called the **X chromosomes.** In males, there is a nonidentical pair, consisting of one X and one Y. The **Y chromosome** is considerably shorter than the X. At meiosis in females, the two X chromosomes pair and segregate like autosomes so that each egg receives one X chromosome. Hence the female is said to be the homogametic sex. At meiosis in males, the X and the Y pair over a short region, which ensures that the X and Y separate so that half the sperm cells receive X and the other half receive Y. Therefore the male is called the heterogametic sex.

The fruit fly *Drosophila melanogaster* has been one of the most important research organisms in genetics; its short, simple life cycle contributes to its usefulness in this regard (Figure 2-11). Fruit flies also have XX females and XY males. However, the mechanism of sex determination in *Drosophila* differs from that in mammals. In *Drosophila,* the number of X chromosomes determines sex: two X's result in a female and one X results in a male. In mammals, the presence of the Y determines maleness and the absence of a Y determines femaleness. This difference is demonstrated by the sexes of the abnormal chromosome types XXY and XO, as shown in Table 2-3 . However, we postpone a full discussion of this topic until Chapter 23 .



Life cycle of *Drosophila melanogaster*, the common fruit fly.

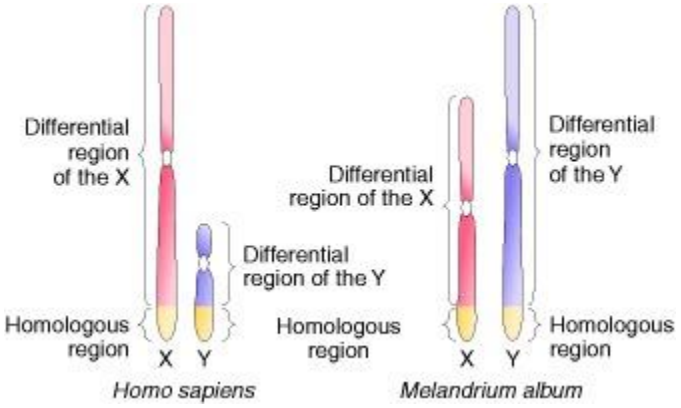
	SEX CHROMOSOMES			
Species	**XX**	**XY**	**XXY**	**XO**
Drosophila	♀	♂	♀	♂
Human	♀	♂	♂	♀

Chromosomal Determination of Sex in Drosophila and Humans.

Vascular plants show a variety of sexual arrangements. **Dioecious** species are the ones showing animal-like sexualdimorphism, with female plants bearing flowers containing only ovaries and male plants bearing flowers containing only anthers (Figure 2-12). Some, but not all, dioecious plants have a nonidentical pair of chromosomes associated with (and almost certainly determining) the sex of the plant. Of the species with nonidentical sex chromosomes, a large proportion have an XY system. For example, the dioecious plant *Melandrium album* has 22 chromosomes per cell: 20 autosomes plus 2 sex chromosomes, with XX females and XY males. Other dioecious plants have no visibly different pair of chromosomes; they may still have sex chromosomes but not visibly distinguishable types.

Cytogeneticists have divided the X and Y chromosomes of some species into homologous and nonhomologous regions. The latter are called *differential* regions (Figure 2-13). These differential regions contain genes that have no counterparts on the other sex chromosome. Genes

in the differential regions are said to be **hemizygous** ("half zygous") in males. Genes in the differential region of the X show an inheritance pattern called **X linkage;** those in the differential region of the Y show Y linkage. Genes in the homologous region show what might be called X-and-Y linkage. In general, genes on sex chromosomes are said to show sex linkage.

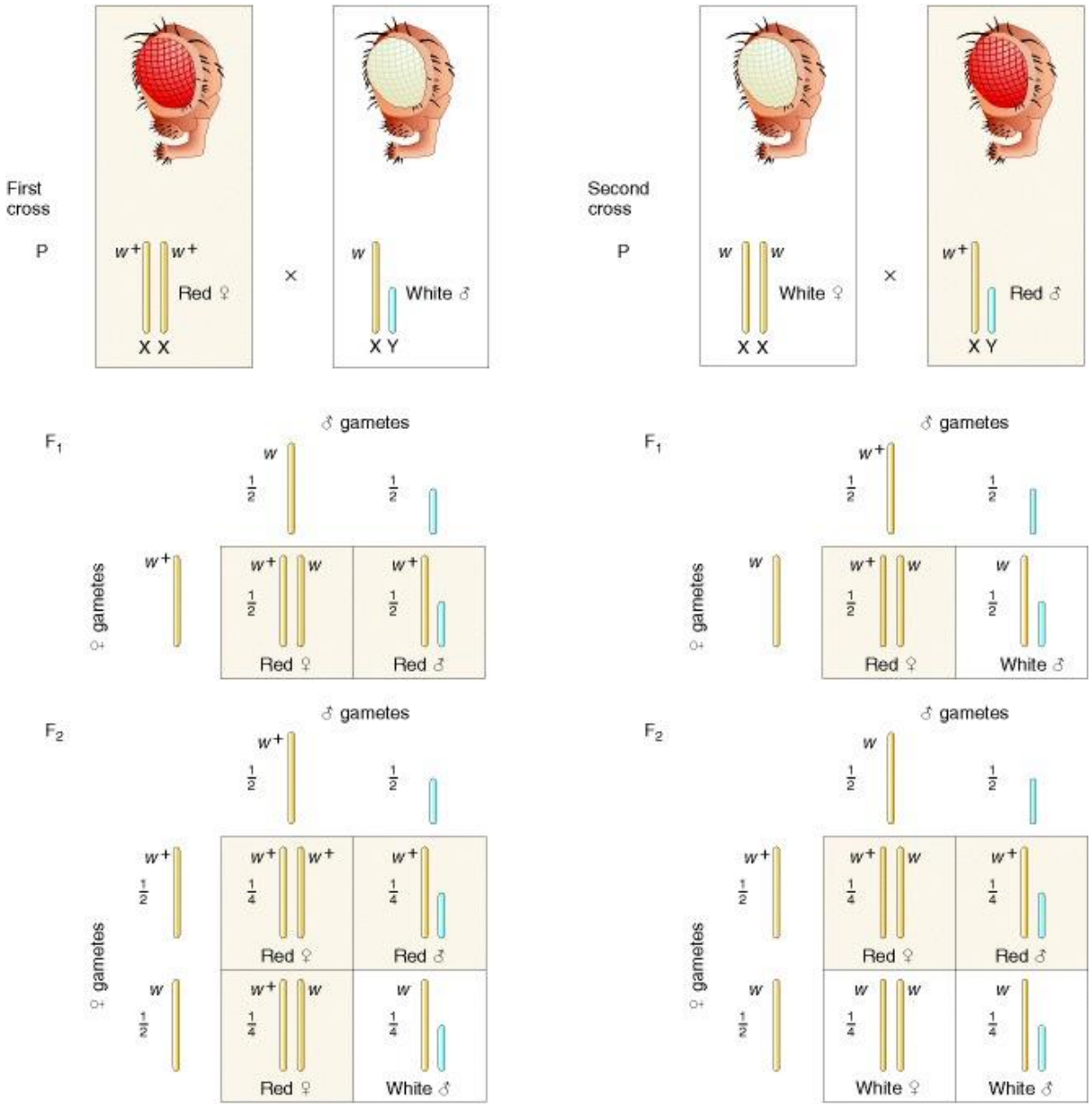


Differential and pairing regions of sex chromosomes of humans and of the plant*Melandrium album.* The regions were located by observing where the chromosomes paired up in meiosis and where they did not.

The genes on the differential regions of the sex chromosomes show patterns of inheritance related to sex. The inheritance patterns of genes on the autosomes produce male and female progeny in the same phenotypic proportions, as typified by Mendel's data (for example, both sexes might show a 3:1 ratio). However, crosses following the inheritance of genes on the sex chromosomes often show male and female progeny with different phenotypic ratios. In fact, for studies of genes of unknown chromosomal location, this pattern is a diagnostic of location on the sex chromosomes. Let's look at an example from *Drosophila.* The wild-type eye color of *Drosophila* is dull red, but pure lines with white eyes are available (Figure 2-14). This phenotypic difference is determined by two alleles of a gene located on the differential region of the X chromosome. When white-eyed males are crossed with red-eyed females, all the F_1 progeny have red eyes, showing that the allele for white is recessive. Crossing the red-eyed F_1 males and females produces a 3:1 F_2 ratio of red-eyed to white-eyed flies, but all the white-eyed flies are males. This inheritance pattern is explained by the alleles being located on the differential region of the X chromosome; in other words, by X-linkage. The genotypes are shown in Figure 2-15 . The reciprocal cross gives a different result. A reciprocal cross between white-eyed females and red-eyed males gives an F_1 in which all the females are red eyed, but all the males are white eyed.

The F_2 consists of one-half red-eyed and one-half white-eyed flies of both sexes. Hence in sex linkage, we see examples not only of different ratios in different sexes, but also of differences between reciprocal crosses.

ed-eyed and white-eyed *Drosophila.* (Carolina Biological Supply.)



Explanation of the different results from reciprocal crosses between red-eyed (red) and white-eyed (white) *Drosophila.* (In *Drosophila* and many other experimental systems, a superscript plus sign is used to designate the normal, or wild-type allele. Here w^+ = red and w = white.)

In *Drosophila,* eye color has nothing to do with sex determination, so we see that genes on the sex chromosomes are not necessarily related to sexual function. The same is true in humans, for whom pedigree analysis has revealed many X-linked genes, of which few could be construed as being connected to sexual function.

MESSAGE

Sex-linked inheritance regularly shows different phenotypic ratios in the two sexes of progeny, as well as different ratios in reciprocal crosses.

Sex Linkage

Sex linkage applies to genes that are located on the sex chromosomes. These genes are considered sex-linked because their expression and inheritance patterns differ between males and females. While sex linkage is not the same asgenetic linkage, sex-linked genes can be genetically linked (see bottom of page).

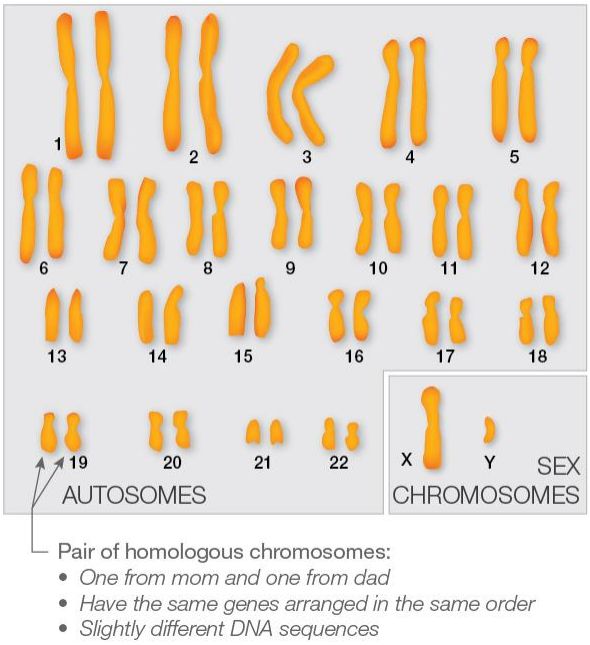
Sex Chromosomes

Sex chromosomes determine whether an individual is male or female. In humans and other mammals, the sex chromosomes are X and Y. Females have two X chromosomes, and males have an X and a Y.

Non-sex chromosomes are also called autosomes. Autosomes come in pairs of homologous chromosomes. Homologous chromosomes have the same genes arranged in the same order. So for all of the genes on the autosomes, both males and females have two copies.

A female's two X chromosomes also have the same genes arranged in the same order. So females have two copies of every gene, including the genes on sex chromosomes.

The X and Y chromosomes, however, have different genes. So for the genes on the sex chromosomes, males have just one copy. The Y chromosome has few genes, but the X chromosome has more than 1,000. Well-known examples in people include genes that control color blindness and male pattern baldness. These are sex-linked traits.



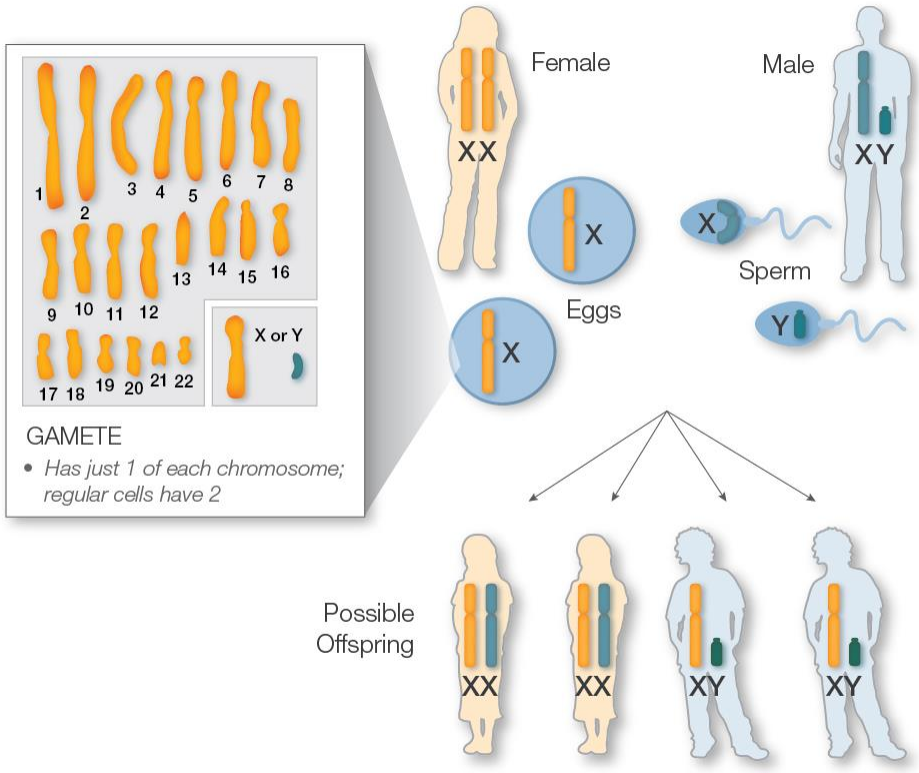
INHERITANCE OF SEX CHROMOSOMES IN MAMMALS

Meiosis is the process of making gametes, also known as eggs and sperm in most animals. During meiosis, the number of chromosomes is reduced by half, so that each gamete gets just one of each autosome and one sex chromosome.

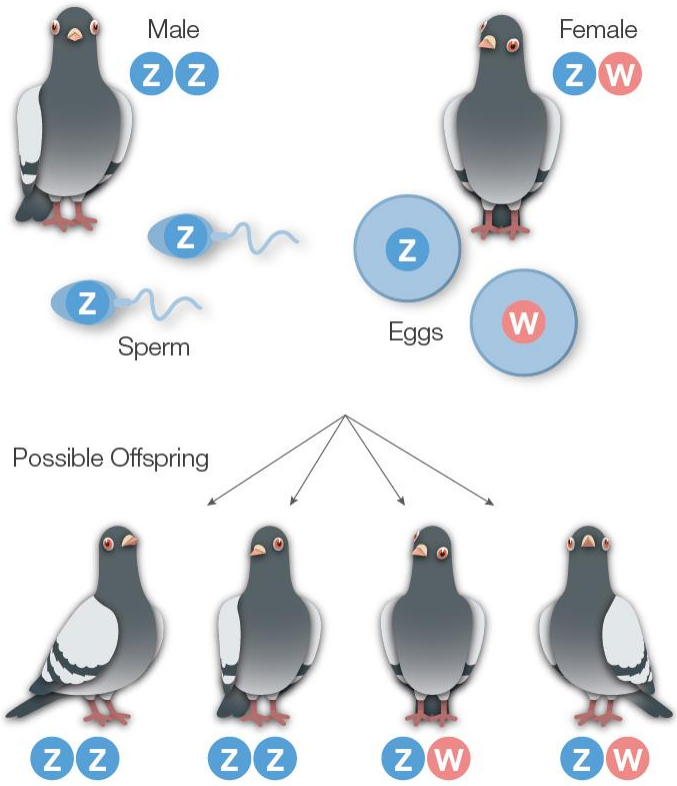
Female mammals make eggs, which always have an X chromosome. And males make sperm, which can have an X or a Y.

Egg and sperm join to make a zygote, which develops into a new offspring. An egg plus an X-containing sperm will make a female offspring, and an egg plus a Y-containing sperm will make a male offspring.

- Female offspring get an X chromsome from each parent
- Males get an X from their mother and a Y from their father
- X chromosomes never pass from father to son
- Y chromosomes always pass from father to son


1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
X or Y
GAMETE
• Has just 1 of each chromosome; regular cells have 2
Female
XX
X
Eggs
X
Male
XY
X
Sperm
Y
Possible Offspring
XX
XX
XY
XY

ex Chromosomes in Pigeons



The way sex determination works in birds is nearly the reverse of how it works in mammals. If you've played Pigeonetics, you know that the sex chromosomes in birds are Z and W. Male birds have two Z chromosomes, and females have a Z and a W. Male birds make sperm, which always have a Z chromosome. Female gametes (eggs) can have a Z or a W.

- Male offspring get a Z chromsome from each parent
- Females get a Z from their father and a W from their mother
- Z chromosomes never pass from mother to daughter
- W chromosomes always pass from mother to daughter

In birds, it's the males that have two copies of every gene, while the females have just one copy of the genes on the sex chromosomes. The W-chromosome is small with few genes. But the Z-chromosome has many sex-linked genes, including genes that control feather color and color intensity.

X & Y and Z & W are just two of the ways that sex is determined in animals. Some animals can even change from one sex to another. To learn more, visit Sex Determination.

INHERITANCE OF SEX LINKED GENE

For genes on autosomes, we all have two copies—one from each parent. The two copies may be the same, or they may be different. Different versions of the same gene are called “alleles” (uh-LEELZ). Genes code for proteins, and proteins make traits.* Importantly, it’s the two alleles working together that affect what we see—also called a “phenotype.”

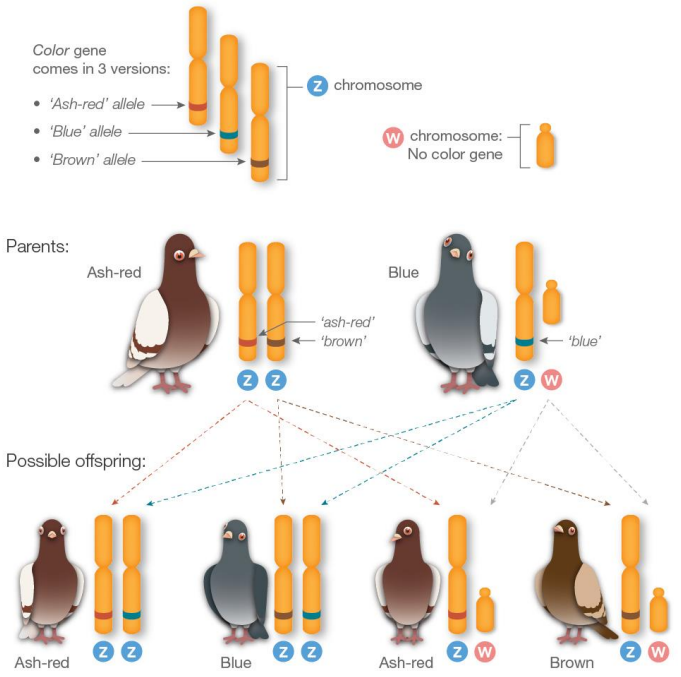
Variations in genes can affect our inherited characteristics, accounting for the differences from one individual to the next. For examples, visit Observable Human Characteristics and The Outcome of Mutation.

Female pigeons (ZW) have just one Z chromosome, and therefore just one allele for each of the genes located there. One gene on the Z chromosome affects feather color; three different alleles make feathers blue, ash-red, or brown. In a female bird (ZW), her single *color* allele determines her feather color. But in males (ZZ), two alleles work together to determine feather color according to their dominance. That is, 'ash-red' is dominant to 'blue', which is dominant to 'brown'.

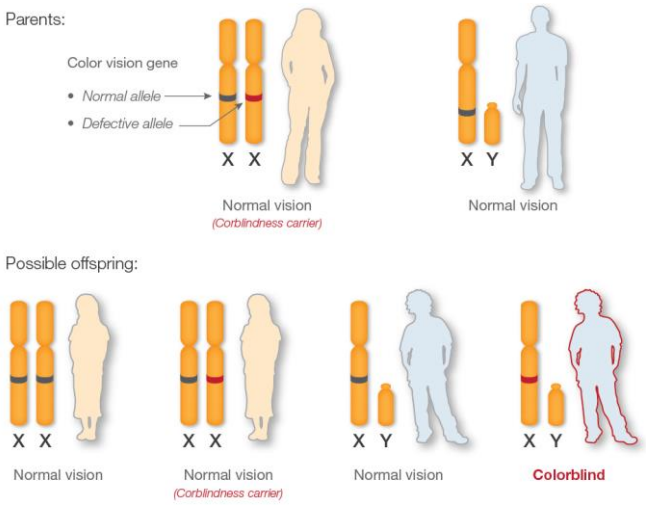
Having two copies of a gene can be important when one copy is “broken” or defective. A functional second copy can often work well enough on its own, acting as a sort of back-up to prevent problems. With sex-linked genes, male mammals (and female birds) have no back-up copy. In people, a number of genetic disorders are sex-linked, including Duchenne muscular dystrophy and hemophilia. These and other sex-inked disorders are much more common in boys than in girls.

Red/green colorblindness is also caused by a defective gene on the X-chromosome. You need at least one working copy of the gene to be able to see red and green. Since boys have just one X-chromosome, which they receive from their mother, inheriting one defective copy of the gene will render them colorblind. Girls have two X-chromosomes; to be colorblind they must inherit two defective copies, one from each parent. Consequently, red-green colorblindness is much more frequent in boys (1 in 12) than in girls (1 in 250).

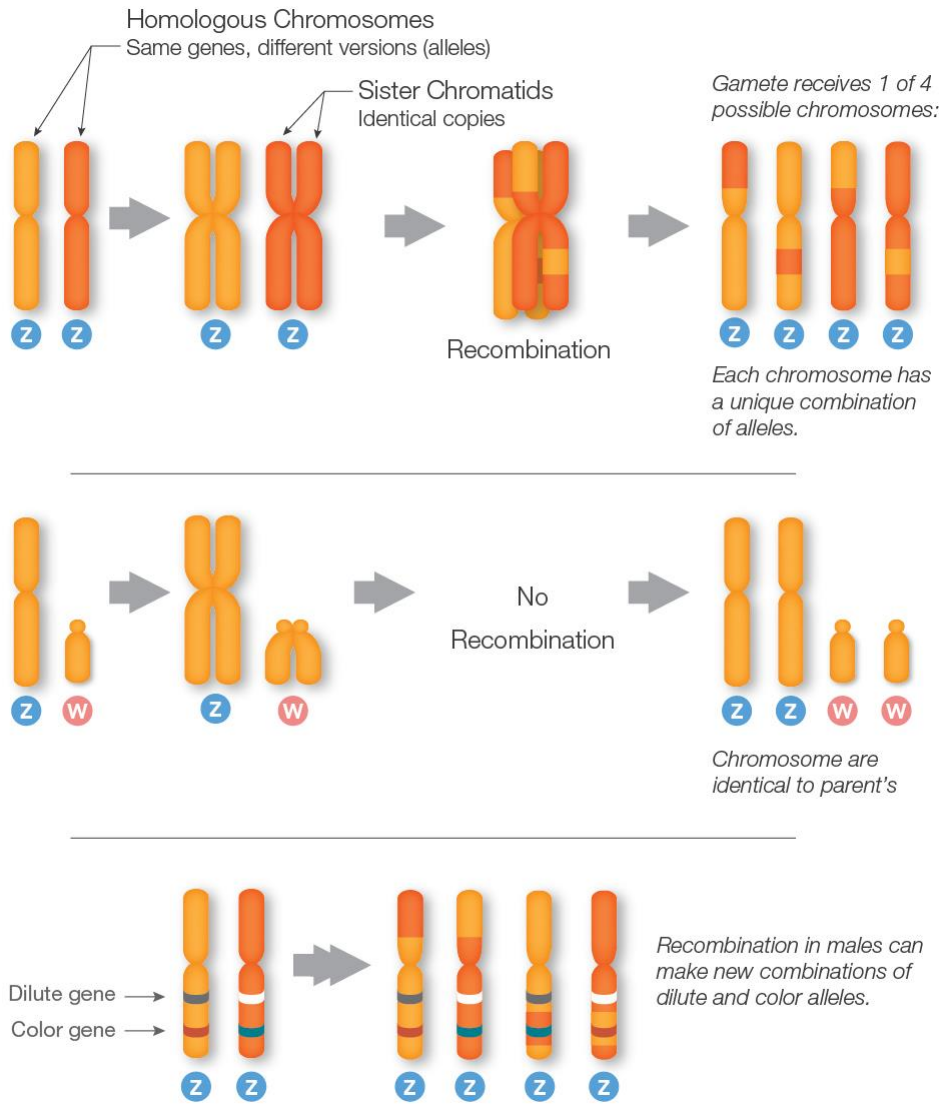
**Some genes code for functional RNAs, which also influence our traits.*



The differences in sex chromosomes between males and females leads to specific inheritance patterns for sex-linked genes. (Above) Female pigeons inherit their color allele from their father. Males inherit one allele from each parent. In humans (below), the pattern is reversed.



RECOMBINATION AND SEX LINKED GENES



When gametes (egg and sperm) form, chromosomes go through a process called recombination. During recombination, homologous chromosomes pair up and exchange stretches of DNA. Recombination makes new allele combinations, which can then be passed to offspring.

When sex chromosomes don't have a homologue (XY male mammals and ZW female birds, for instance), the sex chromosomes do not recombine.* Instead, the sex chromosomes pass unchanged from parent to offspring. But when sex chromosomes do have a homologue (as in XX female mammals and ZZ male birds), the sex chromosomes recombine to make new allele combinations.

In pigeons, color and dilute (color intensity) are controlled by two genes on the Z chromosome. In males, recombination between homologous Z chromosomes can make new combinations

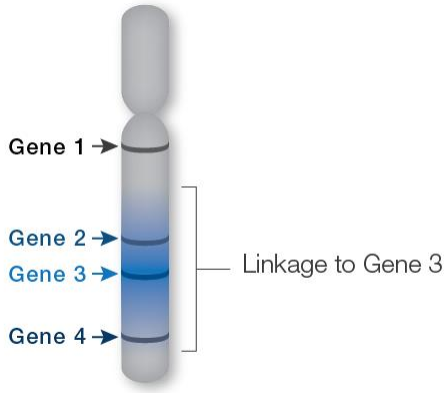
of *color* and *dilute* alleles (by chance, some offspring will still receive the same allele combination as the father). But in females, where the Z chromosome does not recombine, the two alleles always pass to offspring together.

** This isn't entirely true. Portions of the X and Y chromosomes, called the "pseudoautosomal regions," do pair up and recombine. These regions have the same genes, which are not considered sex-linked even though they're on the sex chromosomes.*

SEX LINKED GENES CAN ALSO BE GENETICALLY LINKED

In pigeons, the color and dilute genes are not only sex-linked, they are alsogenetically linked.

Unlinked genes, whether on the same or different chromosomes, are inherited separately 50% of the time. Genes that are genetically linked are inherited separately less than 50% of the time. The closer together the linked genes are, the less likely it is that a recombination event will happen between them. Color and dilute are separated by recombination about 40% of the time (in males only, of course), so they are not very close together.



Gene 3 is more closely linked to Gene 2 than to Gene 4. Gene 1 and Gene 3 are not linked, but by chance they will still stay together 50% of the time, the same as if they were on separate chromosomes.

Extrachromosomal Inheritance –

Definition, Criteria, Maternal Inheritance and Examples

"A non-mendelian pattern of inheritance governed by the DNA present in the cytoplasm is known as extrachromosomal inheritance or cytoplasmic inheritance."

The DNA is the genetic material of us and arranged on chromosomes. It helps to store and transfer information (called traits) through the process of replication, transcription and translation.

Nuclear DNA is the basis for inheritance of almost all type of phenotype of ours. It inherited in a particular fashion from parents to their offspring.

Though all genes are inherited in Mendelian style, some genes present in the cytoplasm of the cell, inherited in a non-mendelian pattern. This type of inheritance is called as extrachromosomal inheritance or cytoplasmic inheritance.

In the present article, we will discuss one of the amazing topic- "extrachromosomal inheritance."

Key Topics:

- What is extrachromosomal inheritance?
    - Definition:
    - Criteria for extrachromosomal inheritance:
- Examples:
    - Cytoplasmic male sterility in maize.
    - The maternal-effect in snail:
    - Inheritance of kappa particles in paramecium:
    - Extrachromosomal inheritance PDF, slide share and ppt:
- Conclusion

What is extrachromosomal inheritance?

The extrachromosomal inheritance also is known as cytoplasmic inheritance or non-mendelian inheritance was first reported by Boris Ephrussi in yeast during 1949.

Cytoplasmic DNA or extrachromosomal DNA is present significantly in some important organelles like chloroplast and mitochondria. It is a big mystery that how actually these organelles created their own genome.

One theory which stated that it was a symbiotic relationship. It is believed that mitochondria were once free-living bacteria. Over a period of time, it created a symbiotic relationship with

eukaryotic cells and established themselves into the cytoplasm and ultimately evolved as an organelle in living eukaryotic cell.

Similarly, the chloroplast in green plants comes from the free-living algae and established a symbiotic relationship with eukaryotic plant cells and settled into cytoplast of green plants.

Read more on chloroplast DNA

Both types of sub-genome have well-developed DNA machinery which is equipped with all the component required for central dogma. Additionally, chloroplast has antibiotic resistance genes indicate that it was derived from bacteria, previously.

The genome is made up of few genes and several thousand base pairs, still it has their own rRNA, tRNA and DNA for replication, translation and transcription.

Definition:

"The extrachromosomal DNA present in the cytoplasm and not on chromosomes which follows the non-Mendelian pattern of inheritance is known as extrachromosomal inheritance."

Criteria for extrachromosomal inheritance:

The extrachromosomal DNA follows a non-mendelian pattern of inheritance

Unlike the common Mendelian segregation pattern is not observed in the extrachromosomal DNA because it does not have the centromere it can not segregate, unlike the normal nuclear DNA.

Their own machinery for protein synthesis:

Unlike nuclear DNA, the organelle DNA or the extrachromosomal DNA has its own replication and transcription machinery. It synthesised their won DNA.

Maternal inheritance:

The extrachromosomal DNA inherited from the maternal side.

The segregation is observed in somatic cells rather than germ cells, unlike nuclear inheritance.

Examples:

Carl Correns in 1908, first reported non-mendelian inheritance in *Mirabilis Jalapa* plastid DNA. Another extrachromosomal inheritance was reported by *M M. Rhoades* in 1933. He postulated that inheritance of male sterility in maize is governed by maternal inheritance and it becomes one of the greatest discoveries in science.

Another important point that makes extrachromosomal DNA even unique is maternal inheritance. It inherits from mother to their offspring which means that only female individual from the entire population can inherit cytoplasmic DNA.

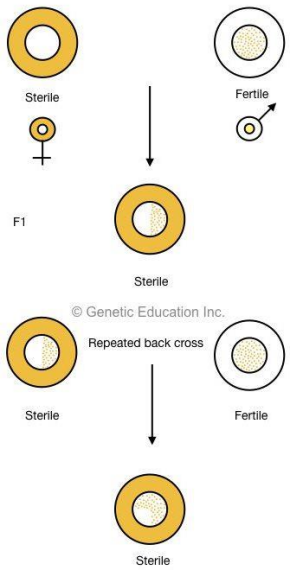
One theory suggests that female reproductive cell (ovum) is bigger, contain more cytoplasm and more organelles than male reproductive cells. This would be expected to influence non-mendelian inheritance or maternal inheritance.

One of the classical examples of maternal inheritance is :

Cytoplasmic male sterility in maize.

Here nuclear genes do not play any significant role rather, the sterility is inherited through egg cytoplasm from generation to generation.

When a male sterile plant is crossed with a normal fertile plant, all the F1 plants remain sterile. When all F1 sterile plants are backcross with a normal fertile plant, until all chromosomes from the male sterile line are exchanged to male fertile, the sterility persists in the progeny.



The image represents the inheritance of cytoplasmic male sterility in maize.

Generally, male-sterile lines are denoted as ***tcs,*** T (Texas), C (Cytoplasmic), S (Sterility). It was believed that T (Texas) cytoplasm is associated with susceptibility against several types of disease like leaf blight disease and yellow blight disease in maize.

This result indicates that chromosomal nuclear DNA does not have any significant role in male sterility (particularly in maize). Furthermore, most of the cytoplasm and organelles are inherited

from the maternal side. From the scientific findings, it is confirmed that the sterility is inherited from the cytoplasm.

This discovery becomes a crucial milestone in crop improvement. Hybrid sterile maize plant becomes more popular as the corn of maize developed uniformly. The hybrid seed becomes more popular for mass production of maize.

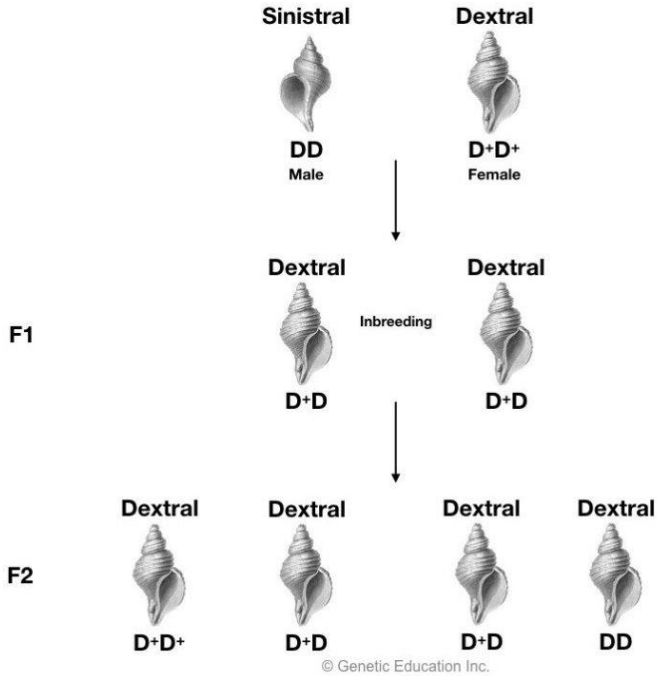
Though maternal inheritance may be extrachromosomal or chromosomal, it is one of the miracle events in nature. Here genetic compositions of maternal side influence several phenotypes of offspring.

In some organism, not only maternal inheritance rather the genotype of the maternal side has great influence on the phenotype of offspring. Here phenotype of mother does not have any role in the development of phenotype in offspring.

The maternal-effect in snail:

The character of coiling in snail is governed by maternal inheritance. Snail, *Limnaea peregra*, has two types of shell coiling phenotypes: one is dextral shells which coil for the right side and another is a sinistral shell which coils for the left side.

Here, the mother's genotype (not a phenotype) is exclusively responsible for the development of coiling style. Assume that D^+ genotype codes for dextral (right side) coiling and D is codes for sinistral coiling. The reciprocal cross of D^+ and D is shown in the figure:

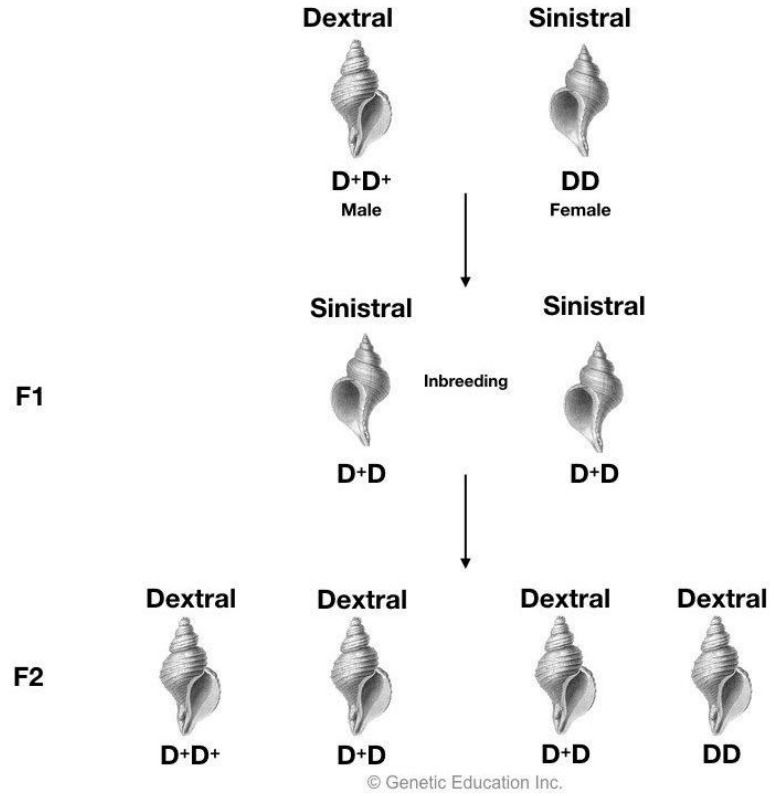


The image represents the maternal effect of the snail. Here crossing between dextral female and sinistral male results in dextral offspring in F1, while inbreeding results in all dextral progenies in the F2 generation.

Crossing between D^+D^+ female and DD male, all the F1, as well as F2 progeny, become dextral as the mother is D^+D^+ dextral, here the DD recessive phenotype is not expressed and typical Mendelian 3:1 ratio is not obtained (all four are dextral).

In another condition when DD sinistral female is crossed with D^+D^+ dextral male, F1 offspring become sinistral with genotype D^+D, here mentioning genotype is important because the inheritance is governed by genotype not by phenotype.

When this F1 progeny is inbred (D^+D * D^+D) all the F2 progeny become dextral and coil for the right side. This results indicated that phenotype of parents do not have any influence on the phenotype of progeny because although all of the F1 progeny are sinistral, all F2 offspring becomes dextral.



The image represents the maternal effect in the snail. Crossing between sinistral female and dextral male results in sinistral F1 progenies. Though all F1 progenies are sinistral, all F2 progenies become dextral.

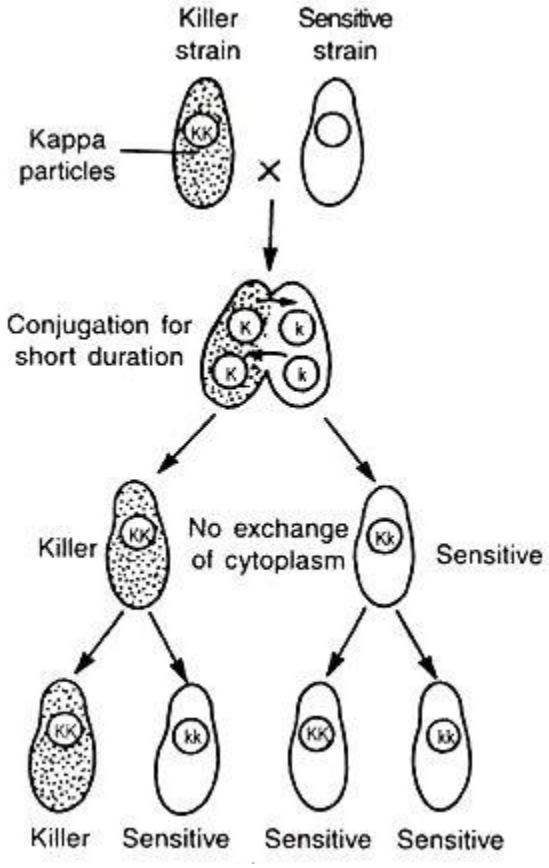
Detailed investigation shows that spindle formed during the second metaphase division decides the direction of coiling. The spindles of dextral snail are tipped to right and vice verse for sinistral. Interestingly, spindle arrangement in metaphase in controlled by maternal genes.

So the actual phenotype of "type of coiling" in snail is governed by maternal genes and it does not depend on the phenotype of any of parent.

In some of the organism, the amount of exchange of cytoplasm plays a crucial role in the inheritance of phenotype.

Inheritance of kappa particles in paramecium:

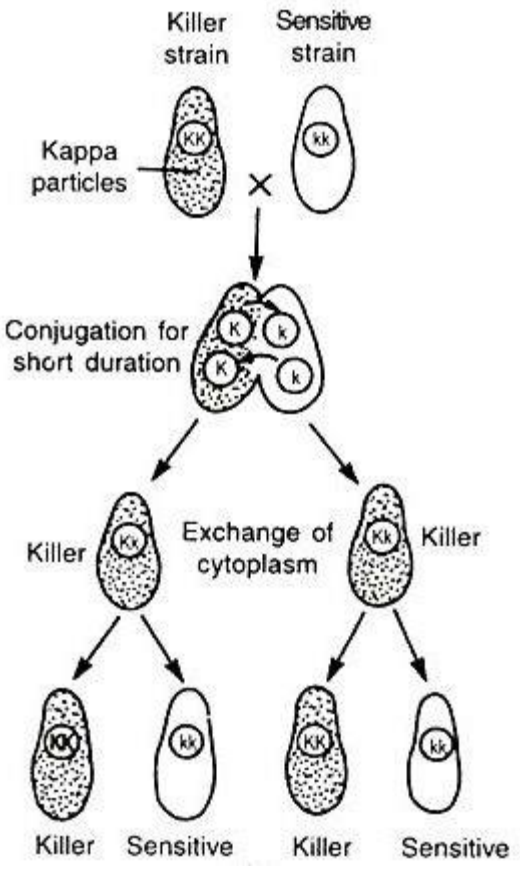
Paramecin is a substance found in some of the killer strain of paramecium which kills the sensitive strains. Paramecin production is governed by the kappa particles present in the cytoplasm of the paramecium.



The image represents inheritance of kappa particles in paramecium at a shorter period of conjugation.

Here KK gene is responsible for the production of kappa particle which is dominant over kk gene. In case of inheritance of kappa particle, cytoplasmic exchange during conjugation plays a crucial role.

When KK killer strains are crossed with kk strains by conjugation, all the progeny obtained are heterozygous with genotype Kk but the phenotype of paramecium depends on presence or absence of kappa particles and it will be influenced by time of conjugation.



The image represents inheritance of kappa particles in paramecium at the longer period of conjugation which results in the exchange of cytoplasm.

If both are conjugates for a shorter period of time, in F1 generation killer strains remain killer and non-killer remain non-killer in the heterozygous condition. Here only nuclear genes are transferred but the cytoplasm is not exchanged between both strains.

In another condition, if killer and non-killer strains are conjugated for a longer period of time, due to the exchange of kappa particles, sensitive strain receives kappa particles through cytoplasmic exchange and sensitive strains become killer in F1 generation.

Linkage and crossing over

The genomes of eukaryotic organisms contain hundreds to thousands of genes (an estimated 30,000-50,000 in humans). Yet there are only a handful of chromosomes. Thus, each chromosome in a eukaryotic genome must contain a large number of genes.

The transmission of genes located on the same chromosome may violate Mendel's Law of Independent Assortment, particularly if they are located very close together along the same arm of a chromosome.

This set of lecture notes will explain why, and provide the theoretical basis for mapping genes along a chromosome by following the degree to which they violate Mendel's Law of Independent Assortment during genetic crosses.

I. Linkage and Crossing Over

In eukaryotic species, each linear chromosome contains a long piece of DNA

- A typical chromosome contains many hundred or even a few thousand different genes
- The term "linkage" has two related meanings
    - 1. Two or more genes can be located on the same chromosome
    - 2. Genes that are close together tend to be transmitted as a unit

Chromosomes are called **linkage groups**

- They contain a group of genes that are linked together
- The number of linkage groups is the number of types of chromosomes of the species
    - For example, in humans
        - 22 autosomal linkage groups
        - An X chromosome linkage group
        - A Y chromosome linkage group

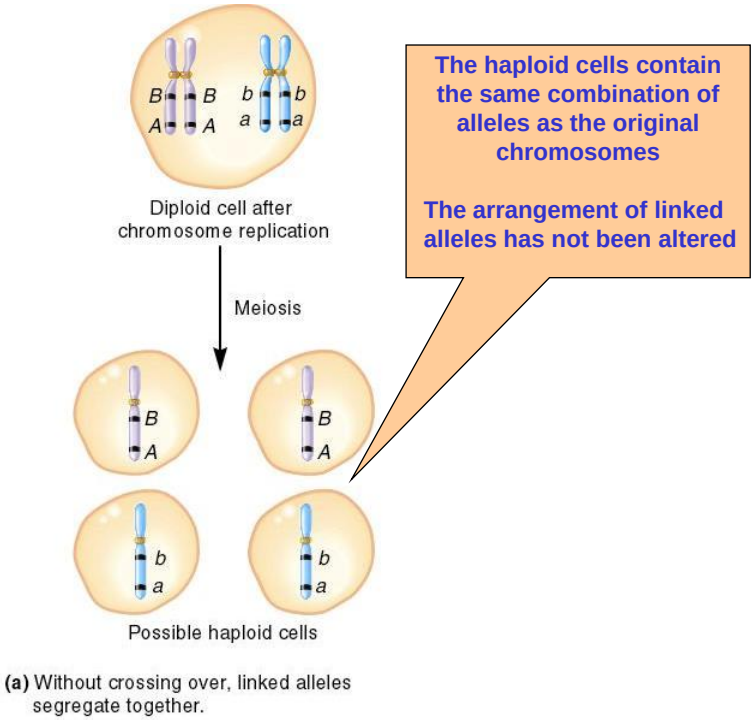
Genes that are far apart on the same chromosome may independently assort from each other due to **crossing-over** during meiosis.

- Occurs during prophase I of meiosis

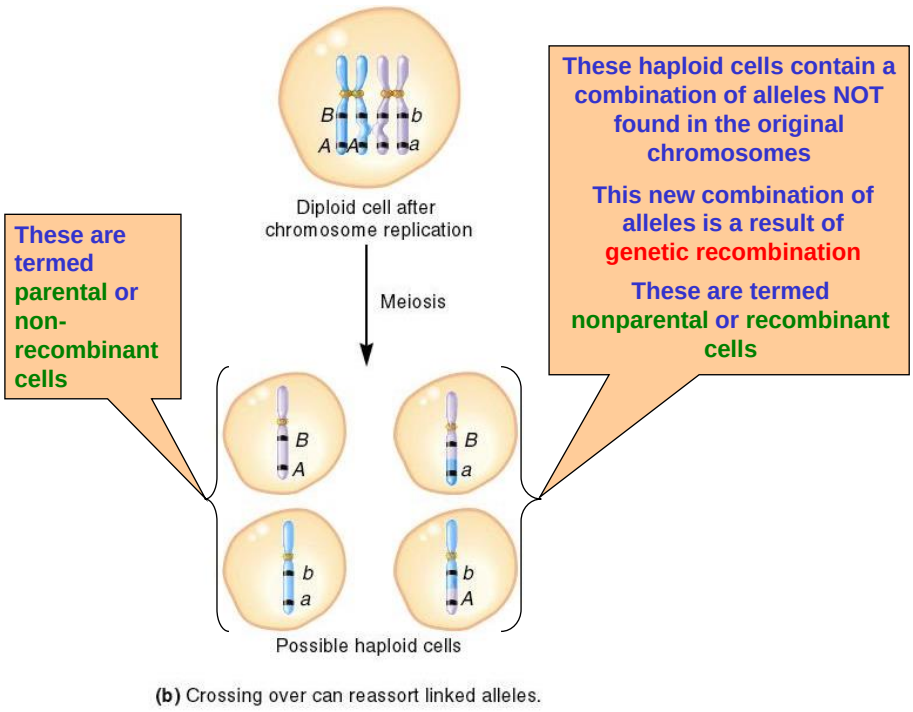
- Homologous chromosomes exchange DNA segments

Crossing Over May Produce Recombinant Phenotypes

- In diploid eukaryotic species, linkage can be altered during meiosis as a result of crossing over
  - Crossing over Occurs
    - during prophase I of meiosis at the bivalent stage
    - Non-sister chromatids of homologous chromosomes exchange DNA segments
- Figure illustrates the consequences of crossing over during meiosis



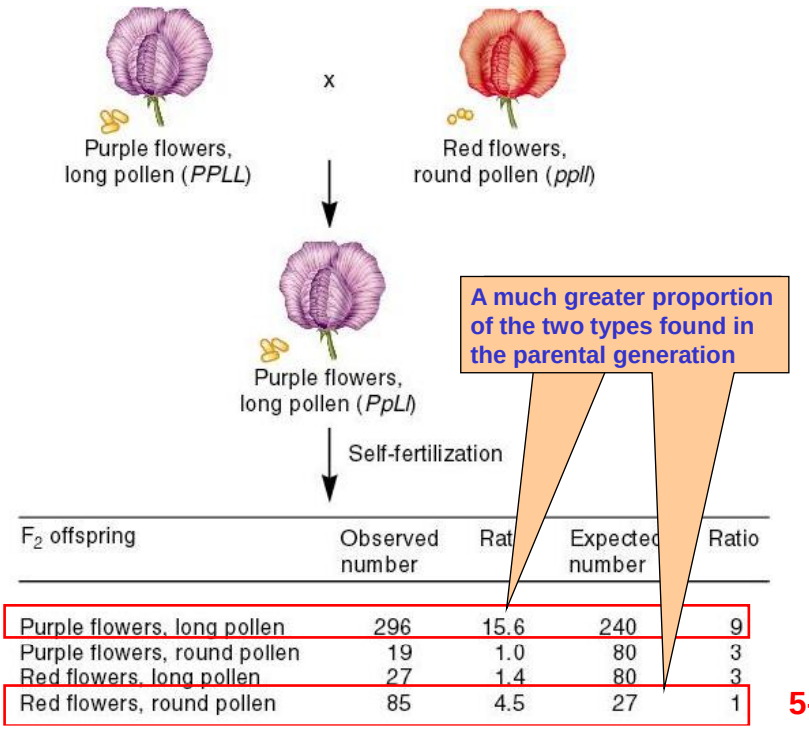
(a) Without crossing over, linked alleles segregate together.



(b) Crossing over can reassort linked alleles.

Bateson and Punnett Discovered Two Traits That Did Not Assort Independently

- In 1905, William Bateson and Reginald Punnett conducted a cross in sweet pea involving two different traits
  - Flower color and pollen shape
- This is a dihybrid cross that is expected to yield a 9:3:3:1 phenotypic ratio in the F_2 generation
  - However, Bateson and Punnett obtained surprising results
- Refer to Figure



F_2 offspring	Observed number	Rat	Expected number	Ratio
Purple flowers, long pollen	296	15.6	240	9
Purple flowers, round pollen	19	1.0	80	3
Red flowers, long pollen	27	1.4	80	3
Red flowers, round pollen	85	4.5	27	1

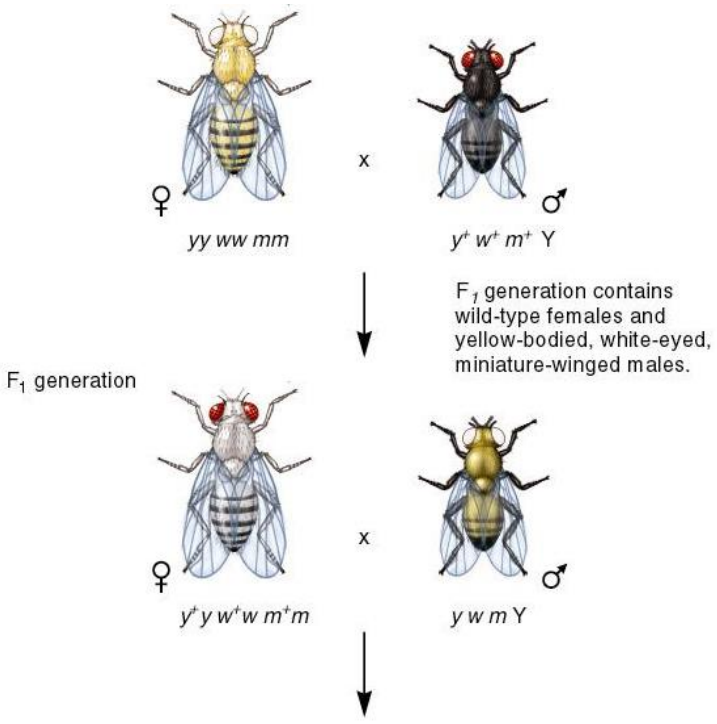
5-9

Bateson and Punnett Discovered Two Traits That Did Not Assort Independently

- They suggested that the transmission of the two traits from the parents was somehow coupled
  - The two traits are not easily assorted in an independent manner
- However, they did not realize that the coupling is due to the linkage of the two genes on the same chromosome

Morgan Provided Evidence for the Linkage of Several X-linked Genes

- The first direct evidence of linkage came from studies of Thomas Hunt Morgan
- Figure illustrates an experiment involving three traits
  - Body color
  - Eye color



- Wing

F_2 generation	Females	Males	Total
Gray body, red eyes, normal wings	439	319	758
Gray body, red eyes, miniature wings	208	193	401
Gray body, white eyes, normal wings	1	0	1
Gray body, white eyes, miniature wings	5	11	16
Yellow body, red eyes, normal wings	7	5	12
Yellow body, red eyes, miniature wings	0	0	0
Yellow body, white eyes, normal wings	178	139	317
Yellow body, white eyes, miniature wings	365	335	700

- Morgan observed a much higher proportion of the combinations of traits found in the parental generation
- Morgan's explanation:
  - All three genes are located on the X chromosome
  - Therefore, they tend to be transmitted together as a unit

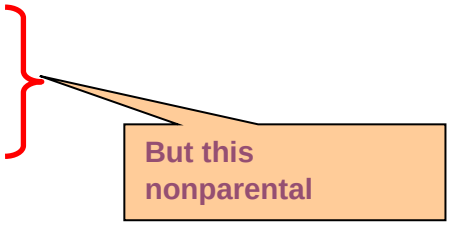
Morgan Provided Evidence for the Linkage of Several X-linked Genes

- However, Morgan still had to interpret two key observations

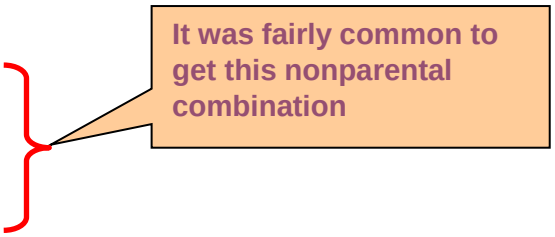
- 1. Why did the F_2 generation have a significant number of nonparental combinations?
- 2. Why was there a quantitative difference between the various nonparental combinations?

Let's reorganize Morgan's data by considering the pairs of genes separately

Gray body, red eyes	1,159
Yellow body, white eyes	1,017
Gray body, white eyes	17
Yellow body, red eyes	12
Total	2,205



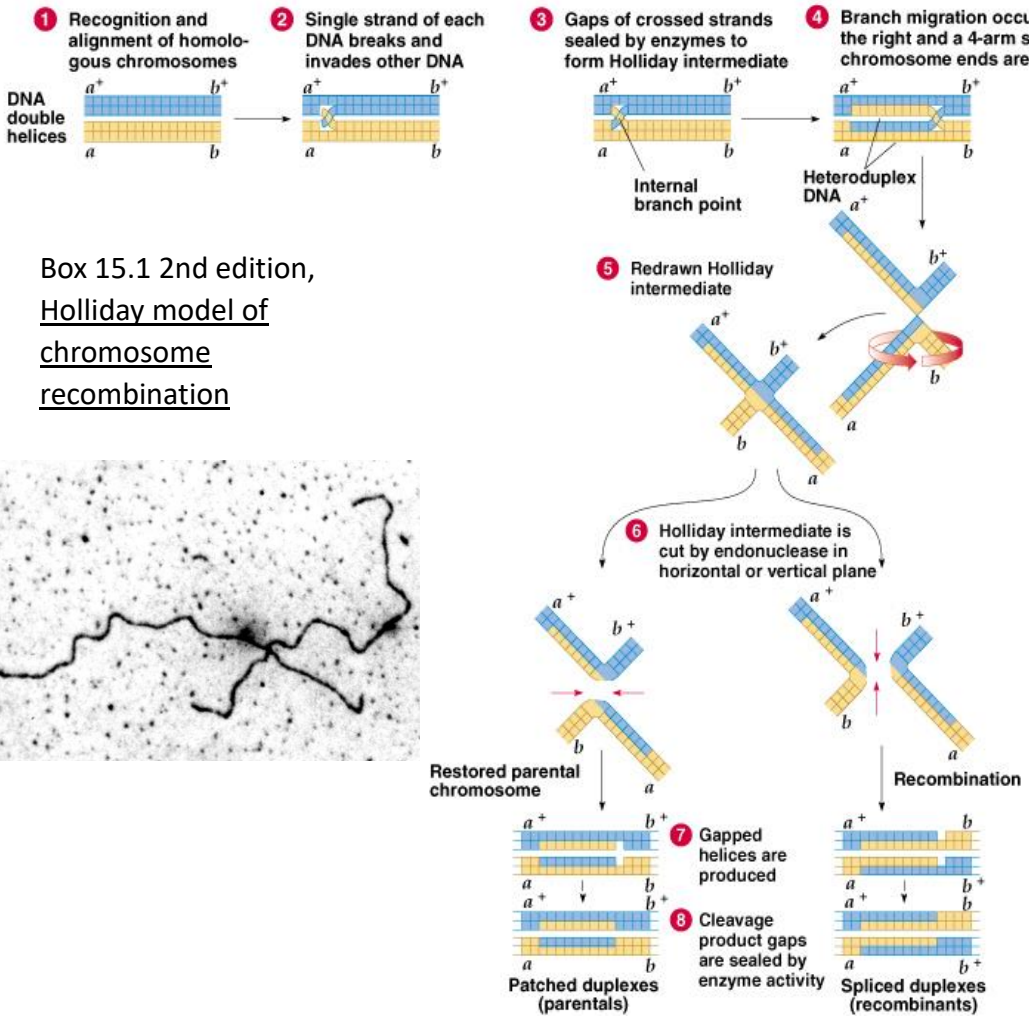
Red eyes, normal wings	770
White eyes, miniature wings	716
Red eyes, miniature wings	401
White eyes, normal wings	318
Total	2,205



Molecular mechanism for crossing-over (Robin Holliday, 1960s):

1. Homologous chromosomes “recognize” and align.
2. Single strands of each DNA (one on each chromosome) break and anneal to the opposite chromosome forming Holliday intermediate.
3. As chromosome ends pull apart, branch point migrations occur to create a 4-arm intermediate structure.
4. 4-arm intermediate is cut by endonucleases in one of 2 planes.
5. Model predicts that physical exchange between two gene loci at the ends of the chromosomes should occur about 50% of the time.

- One pattern (intermediate cut in one plane) yields the parental arrangement.
- The other (cut in the other plane) is recombinant.



Box 15.1 2nd edition, Holliday model of chromosome recombination

Hardy-Weinberg equilibrium

The Hardy-Weinberg equilibrium is a principle stating that the genetic variation in a population will remain constant from one generation to the next in the absence of disturbing factors.

In population genetics, the **Hardy–Weinberg principle**, also known as the **Hardy–Weinberg equilibrium, model, theorem**, or **law**, states

When mating is random in a large population with no disruptive circumstances, the law predicts that both **genotype** and **allele frequencies** will remain constant because they are in equilibrium.

The Hardy-Weinberg equilibrium can be disturbed by a number of forces

Mutations disrupt the equilibrium of allele frequencies by introducing new alleles into a population.

Natural selection and **nonrandom mating** disrupt the Hardy-Weinberg equilibrium because they result in changes in gene frequencies. This occurs because certain alleles help or harm the reproductive success of the organisms that carry them.

Genetic drift, which occurs when allele frequencies grow higher or lower by chance and typically takes place in small populations.

Gene flow, which occurs when breeding between two populations transfers new alleles into a population, can also alter the Hardy-Weinberg equilibrium.

Because all of these disruptive forces commonly occur in nature, the Hardy-Weinberg equilibrium rarely applies in reality. Therefore, the Hardy-Weinberg equilibrium describes an idealized state, and genetic variations in nature can be measured as changes from this equilibrium state.

The seven assumptions underlying Hardy–Weinberg equilibrium are as follows:

- organisms are diploid
- only sexual reproduction occurs
- generations are nonoverlapping
- mating is random
- population size is infinitely large

- allele frequencies are equal in the sexes
- there is no migration, gene flow, admixture, mutation or selection

Violations of the Hardy–Weinberg assumptions can cause deviations from expectation.

- Random mating When the random mating assumption is violated, the population will not have Hardy–Weinberg proportions. A common cause of non-random mating is inbreeding, which causes an increase in homozygosity for all genes.

If a population violates one of the following four assumptions, the population may continue to have Hardy–Weinberg proportions each generation, but the allele frequencies will change over time.

- Selection, in general, causes allele frequencies to change, often quite rapidly.

While directional selection eventually leads to the loss of all alleles except the favored one (unless one allele is dominant, in which case recessive alleles can survive at low frequencies),

some forms of selection, such as balancing selection, lead to equilibrium without loss of alleles.

- Mutation will have a very subtle effect on allele frequencies. Mutation rates are of the order 10^{-4} to 10^{-8}, and the change in allele frequency will be, at most, the same order.

Recurrent mutation will maintain alleles in the population, even if there is strong selection against them.

- Migration genetically links two or more populations together. In general, allele frequencies will become more homogeneous among the populations.
- Small population size can cause a random change in allele frequencies. This is due to a sampling effect, and is called genetic drift. Sampling effects are most important when the allele is present in a small number of copies.

In the simplest case of a single locus with two alleles denoted *A* and *a*

Frequencies $f(\mathrm{A}) = p$ and $f(\mathrm{a}) = q$

ALLELE FREQUENCY = p + q = 1

Expected genotype frequencies under random mating are

GENOTYPE FREQUENCY = $(p + q)^2 = p^2 + 2pq + q^2$

$f(AA) = p^2$ for the AA homozygotes,

$f(aa) = q^2$ for the aa homozygotes, and

$f(Aa) = 2pq$ for the heterozygotes.

Population	**Gene Pool**	**Allele Frequency**	**Genotype Frequency**
5 individuals TT TT Tt Tt tt	T = 6 t = 4	Frequency of Dominant Allele p p= 6/10 = 0.6 Frequency of Recessive Allele q q= 4/10 = 0.4	Punnett Square F(TT) = pp = p2 =(0.6)2 = 0.36 Homozygous Dominant F(tt) = qq = q2 = (0.4)2 = 0.16 Homozygous Recessive F(Tt) = pq + pq = 2pq =2x0.6x0.4=0.48 Heterozygous

REFERENCES

GENETICS: A conceptual approach, 4th Edition, Benjamin A. Pierce, W. H. Freeman and company England; 2006

Cell Biology, Genetics, Molecular Biology, Evolution and Ecology, P.S. Verma, V.K. Agarwal, S. Chand & Company Ltd, 2005

Principles of Molecular Biology, Veer Bala Rastogi, Medtech, 2016

Molecular Biology of the Cell. 4th edition. Alberts B, Johnson A, Lewis J, et al. New York: Garland Science; 2002.

http://www.ncbi.nlm.nih.gov/books/NBK22104/

https://www.google.co.in/search?q=dna+as+genetic+material&biw=1280&bih=890&tbm=isch&imgil=TUQEOSzOlQrxuM%253A%253B5_P9-NeMFLTYMM%253Bhttps%25253A%25252F%25252Fwww.youtube.com%25252Fwatch%25253Fv%2525253DATtWiTqn1Ok&source=iu&pf=m&fir=TUQEOSzOlQrxuM%253A%252C5_P9-NeMFLTYMM%252C_&dpr=1&usg=__t7Dy9JbSGZodbUp0eyZeNP5aO7w%3D&ved=0ahUKEwj3-vLAyqPKAhUVCY4KHTtrBisQyjcIUA&ei=k5SUVrflMpWSuAS71pnYAg#imgrc=TUQEOSzOlQrxuM%3A&usg=__t7Dy9JbSGZodbUp0eyZeNP5aO7w%3D

https://en.wikipedia.org/wiki/Mendelian_inheritance

https://www.google.co.in/webhp?sourceid=chrome-instant&ion=1&espv=2&ie=UTF-8#

http://www.hobart.k12.in.us/jkousen/Biology/multalle.htm

http://theagricos.com/genetics/gene-interactions/

http://www.yourarticlelibrary.com/biology/notes-on-duplicate-polymeric-complementary-and-supplementary-genes/12042/

http://www.isogg.org/wiki/Lethal_alleles

http://www.nature.com/scitable/topicpage/mendelian-ratios-and-lethal-genes-557

https://geneticeducation.co.in/extrachromosomal-inheritance-definition-criteria-maternal-inheritance-and-examples/

SATHYABAMA
INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)
Accredited "A" Grade by NAAC | 12B Status by UGC | Approved by AICTE
www.sathyabama.ac.in

SCHOOL OF BIO AND CHEMICAL ENGINEERING

DEPARTMENT OF BIOTECHNOLOGY

SBTA1301 – MOLECULAR BIOLOGY AND GENETICS

SBTA1301	MOLECULAR BIOLOGY AND GENETICS	L	T	P	Credit	Total Marks
		3	*	0	3	100

COURSE OBJECTIVES

- The course aims to give an understanding on the fundamentals of conventional genetics and the intricate molecular mechanisms of heredity and variations.
- To understand storage of genetic information and its translation at molecular level in prokaryotic and eukaryotic systems.

UNIT 1 CLASSICAL GENETICS **9 Hrs.**

Fundamental principles of genetics- Mendel's principles and experiments, gene interaction, multiple alleles, complementation, linkage, sex linked, sex limited and sex influenced inheritance; Chromosomes basis of heredity- extra-chromosomal inheritance; Linkage and crossing over; Hardy-Weinberg equilibrium, Extensions of Hardy- Weinberg equilibrium.

UNIT 2 STRUCTURE AND PROPERTIES OF NUCLEIC ACIDS **9 Hrs.**

Introduction to nucleic acids: Evidence for DNA&RNA as a genetic material; Structure and physicochemical properties of elements in DNA and RNA, Biological significance of differences in DNA and RNA. Primary structure of DNA: Chemical and structural qualities of 3',5'-Phosphodiester bond. Secondary Structure of DNA: Watson & Crick model, Chargaff's rule; DNA replication- Overview of differences in prokaryotic and eukaryotic DNA replication, D-loop and rolling circle mode of replication, Telomere replication in eukaryotes; Okazaki fragments, Fidelity of DNA replication, Inhibitors of DNA replication, DNA repair- Mutagens, DNA mutations and various types of repair mechanisms.

UNIT 3 TRANSCRIPTION **9 Hrs.**

Central Dogma in molecular biology -Structure and function of mRNA, rRNA tRNA and micro RNAs. Characteristics of promoter and enhancer sequences. RNA synthesis: Initiation, elongation and termination of RNA synthesis, Proteins of RNA synthesis, Fidelity of RNA synthesis, Inhibitors of transcription, Differences in prokaryotic and eukaryotic transcription. Basic concepts in RNA world: RNA processing: 5'-Capping, Splicing-Alternative splicing, Poly 'A' tail addition and base modification.

UNIT 4 TRANSLATION **9 Hrs.**

Introduction to Genetic code: Elucidation of genetic code, Codon degeneracy, Wobble hypothesis and its importance, Prokaryotic and eukaryotic ribosomes. Steps in translation: Initiation, Elongation and termination of protein synthesis. Inhibitors of protein synthesis. Post-translational modifications and its importance.

UNIT 5 REGULATION OF GENE EXPRESSION **9 Hrs.**

Organization of genes in prokaryotic and eukaryotic chromosomes- operon concept; Gene expression and regulation- Hierarchical levels of gene regulation, Prokaryotic gene regulation –lac and trp operon, Eukaryotic gene regulation- gene silencing.

Max.45 Hrs.

COURSE OUTCOMES

On completion of the course, student will be able to

- CO1 - Explain the foundations of Mendelian genetics and chromosomal theory and apply these, with appropriate terminology, to contemporary concepts in genetics.
- CO2 - Emphasize the molecular mechanism of DNA replication and repair in various organisms
- CO3 - Explain the properties of genetic materials and storage and processing of genetic information.
- CO4 - Analyze the processes of transcription and translation in both prokaryotes and eukaryotes at molecular level.
- CO5 - Understand the redundant and universal qualities of the genetic code and how it is used to determine the amino acid sequence of a polypeptide.
- CO6 - Compare the mechanisms of gene regulation in prokaryotes and eukaryotes.

TEXT / REFERENCE BOOKS

1. Lewin B., Genes XI, International Edition, Jocelyn Krebs, Stephen Kilpatrick and Elliott Goldstein, Jones & Bartlett Learning, 2017, ISBN 978-1-4496-5985-1
2. Tropp, Burton E., Molecular Biology: Genes to Proteins, 3rd Edition, Jones and Bartlett, 2008.
3. Glick B.R. and Pasternak J.J., Molecular Biotechnology: Principles and Applications of Recombinant DNA, 4th Edition. ASM, 2010

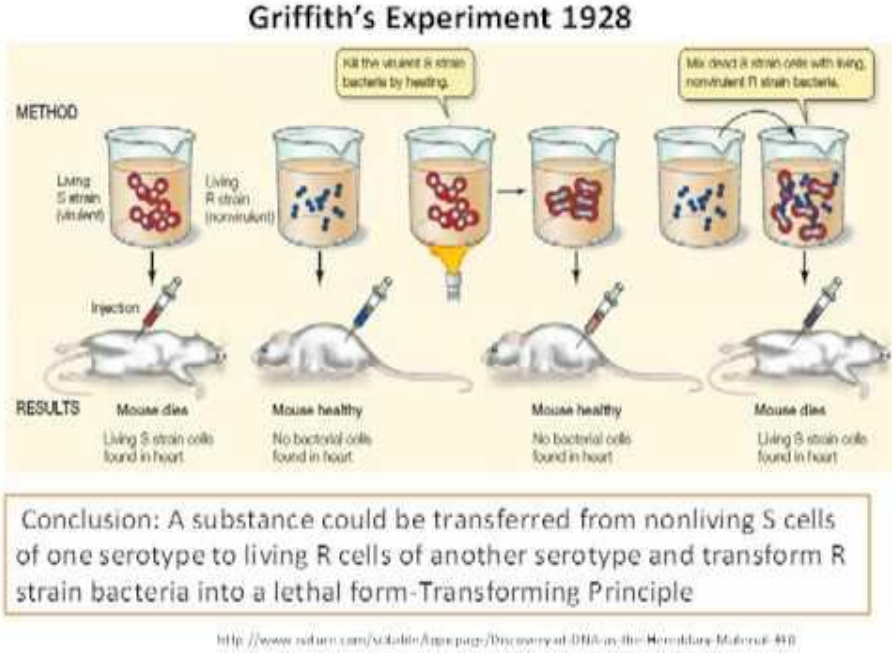
UNIT – II – STRUCTURE AND FUNCTIONS OF NUCLEIC ACIDS

DNA: The genetic material

The physical nature of the gene fascinated scientists for many years. A series of experiments beginning in the 1920s finally revealed that DNA was the genetic material.

Discovery of transformation

A puzzling observation was made by Frederick Griffith in the course of experiments on the bacterium *Streptococcus pneumoniae* in 1928. This bacterium, which causes pneumonia in humans, is normally lethal in mice. However, different strains of this bacterial species have evolved that differ in virulence (in the ability to cause disease or death). In his experiments, Griffith used two strains that are distinguishable by the appearance of their colonies when grown in laboratory cultures. In onestrain, a normazl virulent type, the cells are enclosed in a polysaccharide capsule, giving colonies a smooth appearance; hence, this strain is labeled *S.* In Griffith's other strain, a mutant nonvirulent type that grows in mice but is not lethal, the polysaccharide coat is absent, giving colonies a rough appearance; this strain is called *R.*



Griffith killed some virulent cells by boiling them and injected the heat-killed cells into mice. The mice survived, showing that the carcasses of the cells do not cause death. However, mice injected with a mixture of heat-killed virulent cells and live nonvirulent cells did die. Furthermore, live cells could be recovered from the dead mice; these cells gave smooth colonies and were virulent on subsequent injection. Somehow, the cell debris of the boiled S cells had converted the live R

cells into live S cells. The pr0cess is called transformation. Griffith's experiment is summarized in Figure .

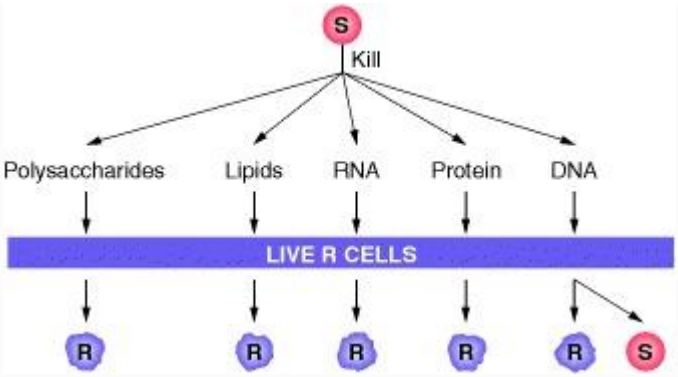
Figure

The first demonstration of bacterial transformation. (a) Mouse dies after injection with the virulent S strain. (b) Mouse survives after injection with the R strain. (c) Mouse survives after injection with heat-killed S strain. (d) Mouse dies after injection with a mixture of heat-killed S strain and live R strain. The heat-killed S strain somehow transforms the R strain into virulence. Parts a, b, and c act as control experiments for this demonstration. .

This same basic technique was then used to determine the nature of the *transforming principle*—the agent in the cell debris that is specifically responsible for transformation. In 1944, Oswald Avery, C. M. MacLeod, and M. McCarty separated the classes of molecules found in the debris of the dead S cells and tested them for transforming ability, one at a time. These tests showed that the polysaccharides themselves do not transform the rough cells. Therefore, the polysaccharide coat, although undoubtedly concerned with the pathogenic reaction, is only the phenotypic expression of virulence. In screening the different groups, Avery and his colleagues found that only one class of molecules, DNA, induced the transformation of R cells (Figure given below). They deduced that DNA is the agent that determines the polysaccharide character and hence the pathogenic character (see pages 219–220 for a description of the mechanism of transformation). Furthermore, it seemed that providing R cells with S DNA was tantamount to providing these cells with S genes.

Figure

Demonstration that DNA is the transforming agent. DNA is the only agent that produces smooth (S) colonies when added to live rough (R) cells.



MESSAGE

The demonstration that DNA is the transforming principle was the first demonstration that genes are composed of DNA.

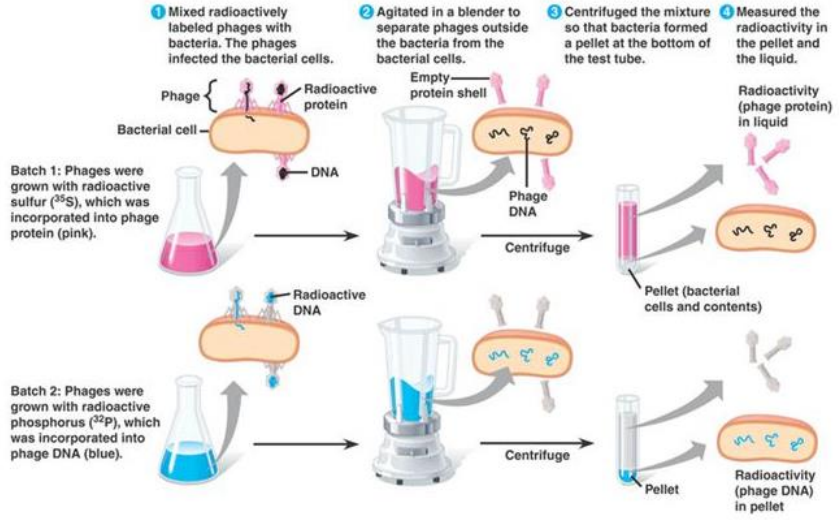
Hershey-Chase experiment

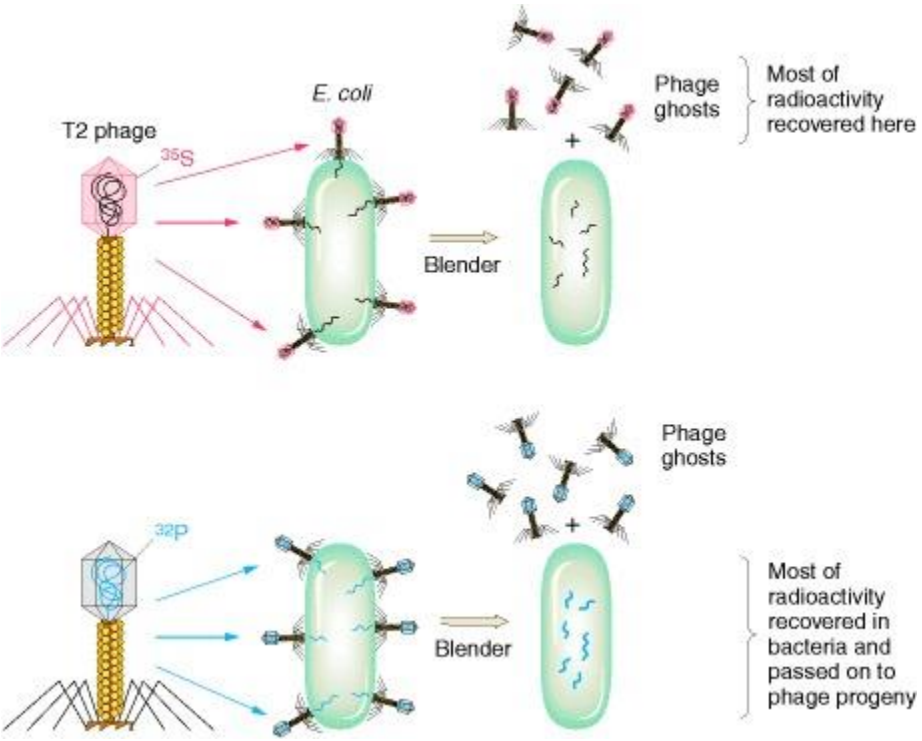
The experiments conducted by Avery and his colleagues were definitive, but many scientists were very reluctant to acceptDNA (rather than proteins) as the genetic material. The clincher was provided in 1952 by Alfred Hershey and Martha Chase with the use of the phage (virus) T2. They reasoned that phage infection must entail the introduction (injection) into the bacterium of the specific information that dictates viral reproduction. The phage is relatively simple in molecular constitution. Most of its structure is protein, with DNA contained inside the protein sheath of its "head."

Phosphorus is not found in proteins but is an integral part of DNA; conversely, sulfur is present in proteins but never in DNA. Hershey and Chase incorporated the radioisotope of phosphorus (^{32}P) into phage DNA and that of sulfur (^{35}S) into the proteins of a separate phage culture. They then used each phage culture independently to infect *E. coli* with many virus particles per cell. After sufficient time for injection to take place, they sheared the empty phage carcasses (called *ghosts*) off the bacterial cells by agitation in a kitchen blender. They used centrifugation to separate the bacterial cells from the phage ghosts and then measured the radioactivity in the two fractions. When the ^{32}P-labeled phages were used, most of the radioactivity ended up inside the bacterial cells, indicating that the phage DNA entered the cells. ^{32}P can also be recovered from phage progeny. When the ^{35}S-labeled phages were used, most of the radioactive material ended up in the phage ghosts, indicating that the phage protein never entered the bacterial cell (Figure 8-3). The conclusion is inescapable: DNA is the hereditary material; the phage proteins are mere structural packaging that is discarded after delivering the viral DNA to the bacterial cell.

The Hershey-Chase experiment

- *Analysis of the Hershey-Chase experiment provided evidence that DNA is the genetic material*





Figure

The Hershey-Chase experiment, which demonstrated that the genetic material of phage is DNA, not protein. The experiment uses two sets of T2 bacteriophages. In one set, the protein coat is labeled with radioactive sulfur (^{35}S), not found in DNA. ther set, the DNA is labeled with radioactive phosphorus (^{32}P), not found in protein. Only the ^{32}P is injected into the *E. coli*, indicating that DNA is the agent necessary for the production of new phages.

Why such reluctance to accept this conclusion? DNA was thought to be a rather simple chemical. How could all the information about an organism's features be stored in such a simple molecule? How could such information be passed on from one generation to the next? Clearly, the genetic material must have both the ability to encode specific information and the capacity to duplicate that information precisely. What kind of structure could allow such complex functions in so simple a molecule?

RNA as genetic material

RNA is the genetic material in viruses was demonstrated in 1956 with the experiments conducted on tobacco plant by A.Gierer and G.Schramm.

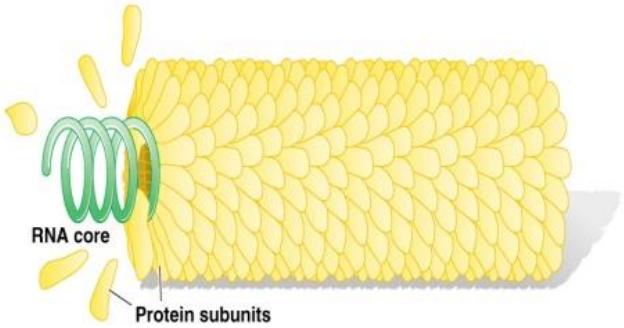
All viruses are not limited to bacterial hosts. Viruses that infect and parasite plant cells, some animal cells contain RNA only. In these viruses RNA act as genetic material. One plant virus, Tobacco mosaic virus(TMV), that ontains RNA, not DNA was an important tool for genetic Experiments. TMV infects tobacco, causing the infected regions on leaves to become discoloured and bristled. Different strains of TMV produce clearly different inherited lesions on the infected leaves.The common virus produces a green mosaic disease, but a variant Holmes rib grass(TMV-HR), produces ring spot lesions. Moreover , the amino acid compositions of the proteins of these two strains differ.

H.Fraenkel-Conrat and B.Singer first developed the techniques for separating TMV particles into RNA and protein. They found that virus could be broken into component parts and they could again be reassembled or reconsitituted to form functional virus.From the two strains of TMV they were able to reconstitute viruses with the RNA from TMV common enclosed in TMV-HR protein and TMV-HR RNA with TMV common protein. When these reassembled viruses were used to infect tobacco leaves, the progency viruses produced were always found to be phenotypically and genotypically identical to the parent strain from which

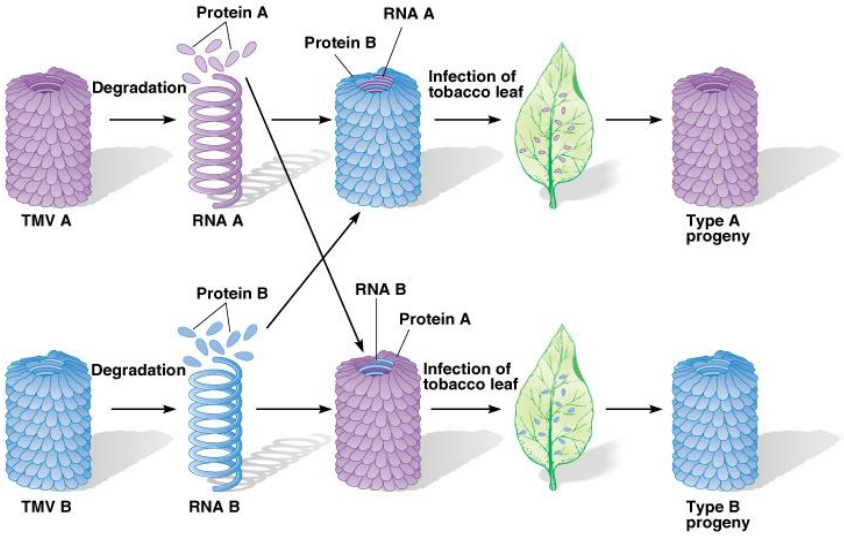
the RNA had been obtained. The reassembled viruses with the TMV-common RNA and TMV-HR protein produced a green mosaic disease characteristic of TMV-common. Recovered virus had protein characteristic of TMV common.This proved that specificity of virus proteins was determined by RNA alone and that proteins carried no genetic information. Hence RNA carries genetic information not proteins.

The genetic RNA is usually found to be single stranded but in some it is double stranded as in reovirus, wound tumor virus.

Typical tobacco mosaic virus (TMV) particle

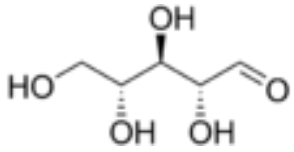


Demonstration that RNA is the genetic material in tobacco mosaic virus (TMV)



SUGARS INNUCLEIC ACIDS

Ribose



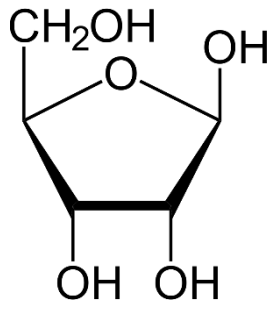
Ribose is a simple sugar and carbohydrate with molecular formula $C_5H_{10}O_5$ the linear-form composition $H{-}(C{=}O){-}(CHOH)_4{-}H$.

In its linear form, ribose can be recognised as the pentose sugar with all of its hydroxyl functional groups on the same side in its Fischer projection.

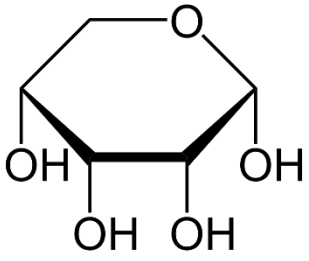
Fischer chose the name "ribose" as it is a partial rearrangement of the name of another sugar, arabinose, of which ribose is an epimer at the 2' carbon;

both names also relate to gum arabic, from which arabinose was first isolated and from which they prepared l-ribose.

Cyclisation of ribose



β-D-ribofuranose



α-D-ribopyranose

In each case, there are two possible geometric outcomes, named as α- and β- and known as anomers, depending on the stereochemistry at the hemiacetal carbon atom (the "anomeric carbon")

At room temperature, about 76% of d-ribose is present in pyranose forms (α:β = 1:2) and 24% in the furanose forms (α:β = 1:3), with only about 0.1% of the linear form present.

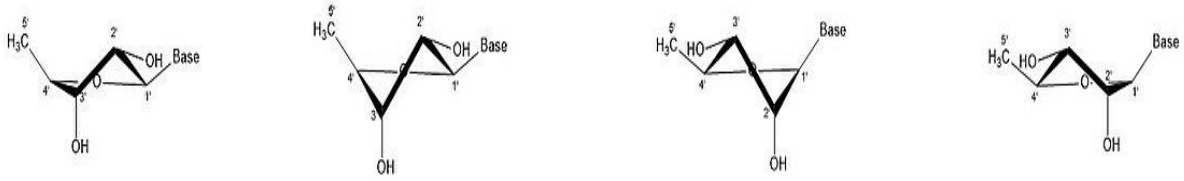
- The ribonucleosides adenosine, cytidine, guanosine, and uridine are all derivatives of β-d-ribofuranose.
- Metabolically-important species that include phosphorylated ribose include ADP, ATP, coenzyme A, and NADH.
- cAMP and cGMP serve as secondary messengers in some signaling pathways and are also ribose derivatives.
- The ribose moiety appears in some pharmaceutical agents, including the antibiotics neomycin and paromomycin.

Structure

- Ribose is an aldopentose (a monosaccharide containing five carbon atoms) that, in its open chain form, has an aldehyde functional group at one end.
- In the conventional numbering scheme for monosaccharides, the carbon atoms are numbered from C1' (in the aldehyde group) to C5’.
- The deoxyribose derivative found in DNA differs from ribose by having a hydrogen atom in place of the hydroxyl group at C2'. This hydroxyl group performs a function in RNA splicing.
- Like many monosaccharides, ribose exists in an equilibrium among 5 forms—the linear form H−(C=O)−$(CHOH)_4$–H and either of the two ring forms: *α- or β-ribofuranose* ("C3'-endo"), with a five-membered tetrahydrofuran ring, and *α- or β-ribopyranose* ("C2'-endo"), with a six-membered tetrahydropyran ring.

- The “d-” in the name d-ribose refers to the stereochemistry of the chiral carbon atom farthest away from the aldehyde group (C4'). In d-ribose, as in all d-sugars, this carbon atom has the same configuration as in d-glyceraldehyde.

Different pucker configurations of Ribose



2' endo 2' endo 3' exo 3' endo 2' exo 3' endo

A ribose molecule is typically represented as a planar molecule on paper. Despite this, it is typically non-planar in nature.

Even between hydrogen atoms, the many constituents on a ribose molecule cause steric hindrance and strain between them.

To relieve this crowding and ring strain, the ring puckers, i.e. becomes non-planar.

This puckering is achieved by displacing an atom from the plane, relieving the strain and yielding a more stable configuration.

When only a single atom is displaced, it is referred to as an "envelope" pucker.

When two atoms are displaced, it is referred to as a "twist" pucker, in reference to the zigzag orientation

In an "endo" pucker, the major displacement of atoms is on the β-face, the same side as the C4'-C5' bond and the base.

In an "exo" pucker, the major displacement of atoms is on the α-face, on the opposite side of the ring.

The major forms of ribose are the 3'-endo pucker (commonly adopted by RNA and A-form DNA) and 2'-endo pucker (commonly adopted by B-form DNA).

FUNCTIONS IN BIOCHEMISTRY

- Ribose plays many important roles in metabolism, which means that it is involved in a lot of biochemistry.
- Ribose is used as a building block for a lot of the signals and products throughout the metabolic pathway.
- One of the most important products of the metabolic pathway is adenosine triphosphate (ATP), which provides energy that drives processes in cells.
- ATP is derived from ribose; it contains one ribose, three phosphate groups, and an adenine base.

- ATP is created during cellular respiration from adenosine diphosphate (ATP with one less phosphate group).

Signalling Pathway

Ribose also plays a major role in signaling pathways because it is a building block in secondary signaling molecules such as cyclic adenosine monophosphate (cAMP) which is derived from ATP.

One specific case in which cAMP is used is in cAMP-dependent signaling pathways. In cAMP signaling pathways, either a stimulative or inhibitory hormone receptor is activated by a signal molecule.

These receptors are linked to a stimulative or inhibitory regulative G-protein. When a stimulative G-protein is activated, adenylyl cyclase catalyzes ATP into cAMP by using Mg^{2+} or Mn^{2+}. cAMP, a secondary messenger, then goes on to activate protein kinase A, which is an enzyme that regulates cell metabolism.

Protein kinase A regulates metabolic enzymes by phosphorylation which causes a change in the cell depending on the original signal molecule. The opposite occurs when an inhibitory G-protein is activated; the G-protein inhibits adenylyl cyclase and ATP is not converted to cAMP.

Metabolism

- Ribose is referred to as the "molecular currency" because of its involvement in intracellular energy transfers. For example, nicotinamide adenine dinucleotide (NAD), flavin adenine dinucleotide (FAD), and nicotinamide adenine dinucleotide phosphate (NADP) all contain the d-ribofuranose moiety.
- They can each be derived from d-ribose after it is converted to d-ribose 5-phosphate by the enzyme ribokinase.
- NAD, FAD, and NADP act as electron acceptors in biochemical redox reactions in major metabolic pathways including glycolysis, the citric acid cycle, fermentation, and the electron transport chain.

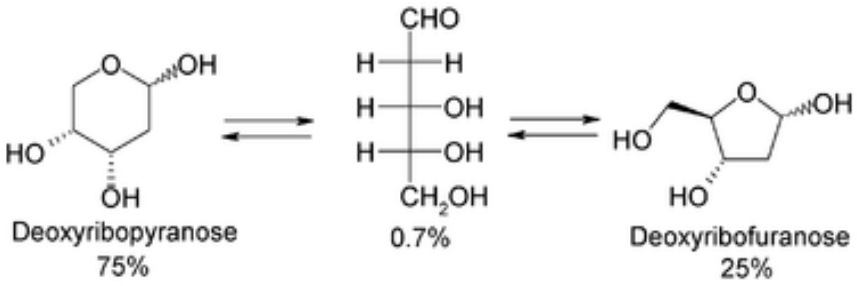
DEOXYRIBOSE

Deoxyribose, or more precisely 2-deoxyribose, is a monosaccharide with idealized formula $H{-}(C{=}O){-}(CH_2){-}(CHOH)_3{-}H$.

Its name indicates that it is a deoxy sugar, meaning that it is derived from the sugar ribose by loss of an oxygen atom.

Since the pentose sugars arabinose and ribose only differ by the stereochemistry at C2′, 2-deoxyribose and 2-deoxyarabinose are equivalent, although the latter term is rarely used because ribose, not arabinose, is the precursor to deoxyribose.

STRUCTURE



Several isomers exist with the formula $H{-}(C{=}O){-}(CH_2){-}(CHOH)_3{-}H$, but in deoxyribose all the hydroxyl groups are on the same side in the Fischer projection.

The term "2-deoxyribose" may refer to either of two enantiomers:

- the biologically important d-2-deoxyribose and
- to the rarely encountered mirror image l-2-deoxyribose.

d-2-deoxyribose is a precursor to the nucleic acid DNA. 2-deoxyribose is an aldopentose, that is, a monosaccharide with five carbon atoms and having an aldehyde functional group.

In aqueous solution, deoxyribose primarily exists as a mixture of three structures: the linear form $H{-}(C{=}O){-}(CH_2){-}(CHOH)_3{-}H$ and two ring forms, deoxyribofuranose ("C3′-endo"), with a five-membered ring, and deoxyribopyranose ("C2′-endo"), with a six-membered ring.

BIOLOGICAL IMPORTANCE

As a component of DNA, 2-deoxyribose derivatives have an important role in biology.

The DNA (deoxyribonucleic acid) molecule, which is the main repository of genetic information in life, consists of a long chain of deoxyribose-containing units called nucleotides, linked via phosphate groups.

In the standard nucleic acid nomenclature, a DNA nucleotide consists of a deoxyribose molecule with an organic base (usually adenine, thymine, guanine or cytosine) attached to the 1′ ribose carbon. The 5′ hydroxyl of each deoxyribose unit is replaced by a phosphate (forming a nucleotide) that is attached to the 3′ carbon of the deoxyribose in the preceding unit.

The absence of the 2′ hydroxyl group in deoxyribose is apparently responsible for the increased mechanical flexibility of DNA compared to RNA, which allows it to assume the double-helix conformation, and also (in the eukaryotes) to be compactly coiled within the small cell nucleus.

The double-stranded DNA molecules are also typically much longer than RNA molecules. The backbone of RNA and DNA are structurally similar, but RNA is single stranded, and made from ribose as opposed to deoxyribose.

Other biologically important derivatives of deoxyribose include mono-, di-, and triphosphates, as well as 3′-5′ cyclic monophosphates.

NITROGENOUS BASES

ADENINE

Adenine is a nucleobase (a purine derivative).

It is one of the four nucleobases in the nucleic acid of DNA that are represented by the letters G–C–A–T.

The three others are guanine, cytosine and thymine.

Its derivatives have a variety of roles in biochemistry including cellular respiration, in the form of both the energy-rich adenosine triphosphate (ATP) and the cofactors nicotinamide adenine dinucleotide (NAD) and flavin adenine dinucleotide (FAD).

It also has functions in protein synthesis and as a chemical component of DNA and RNA.

The shape of adenine is complementary to either thymine in DNA or uracil in RNA.

When connected into DNA, a covalent bond is formed between deoxyribose sugar and the bottom left nitrogen (thereby removing the existing hydrogen atom).

Adenosine is adenine reacted with ribose, as used in RNA and ATP;

deoxyadenosine is adenine attached to deoxyribose, as used to form DNA.

Function

Adenine is one of the two purine nucleobases (the other being guanine) used in forming nucleotides of the nucleic acids.

In DNA, adenine binds to thymine via two hydrogen bonds to assist in stabilizing the nucleic acid structures.

In RNA, which is used for protein synthesis, adenine binds to uracil.

Adenine forms adenosine, a nucleoside, when attached to ribose, and deoxyadenosine when attached to deoxyribose.

It forms adenosine triphosphate (ATP), a nucleoside triphosphate, when three phosphate groups are added to adenosine.

Adenosine triphosphate is used in cellular metabolism as one of the basic methods of transferring chemical energy between chemical reactions.

GUANINE

Guanine is one of the four main nucleobases found in the nucleic acids DNA and RNA, the others being adenine, cytosine, and thymine (uracil in RNA).

In DNA, guanine is paired with cytosine. The guanine nucleoside is called guanosine.

With the formula $C_5H_5N_5O$, guanine is a derivative of purine, consisting of a fused pyrimidine-imidazole ring system with conjugated double bonds.

This unsaturated arrangement means the bicyclic molecule is planar.

Properties

Guanine, along with adenine and cytosine, is present in both DNA and RNA, whereas thymine is usually seen only in DNA, and uracil only in RNA. Guanine has two tautomeric forms, the major keto form and rare enol form.

It binds to cytosine through three hydrogen bonds. In cytosine, the amino group acts as the hydrogen bond donor and the C-2 carbonyl and the N-3 amine as the hydrogen-bond acceptors.

Guanine has the C-6 carbonyl group that acts as the hydrogen bond acceptor, while a group at N-1 and the amino group at C-2 act as the hydrogen bond donors.

Guanine can be hydrolyzed with strong acid to glycine, ammonia, carbon dioxide, and carbon monoxide.

First, guanine gets deaminated to become xanthine.

Guanine oxidizes more readily than adenine, the other purine-derivative base in DNA.

Its high melting point of 350 °C reflects the intermolecular hydrogen bonding between the oxo and amino groups in the molecules in the crystal.

Because of this intermolecular bonding, guanine is relatively insoluble in water, but it is soluble in dilute acids and bases.

Other occurrences and biological uses

Guanine crystals are rhombic platelets composed of multiple transparent layers, but they have a high index of refraction that partially reflects and transmits light from layer to layer, thus producing a pearly luster. It can be applied by spray, painting, or dipping.

It may irritate the eyes.

Guanine has a very wide variety of biological uses that include a range of functions ranging in both complexity and versatility. These include camouflage, display, and vision among other purposes.

Spiders, scorpions, and some amphibians convert ammonia, as a product of protein metabolism in the cells, to guanine, as it can be excreted with minimal water loss.

Guanine is also found in specialized skin cells of fish called iridocytes (e.g., the sturgeon), as well as being present in the reflective deposits of the eyes of deep-sea fish and some reptiles, such as crocodiles.

On 8 August 2011, a report, based on NASA studies with meteorites found on Earth, was published suggesting building blocks of DNA and RNA (guanine, adenine and related organic molecules) may have been formed extra-terrestrially in outer space.

CYTOSINE

Cytosine C is one of the four main bases found in DNA and RNA, along with adenine, guanine, and thymine (uracil in RNA).

It is a pyrimidine derivative, with a heterocyclic aromatic ring and two substituents attached (an amine group at position 4 and a keto group at position 2).

The nucleoside of cytosine is cytidine.

In Watson-Crick base pairing, it forms three hydrogen bonds with guanine.

Chemical Reactions

Cytosine can be found as part of DNA, as part of RNA, or as a part of a nucleotide.

As cytidine triphosphate (CTP), it can act as a co-factor to enzymes, and can transfer a phosphate to convert adenosine diphosphate (ADP) to adenosine triphosphate (ATP).

In DNA and RNA, cytosine is paired with guanine.

However, it is inherently unstable, and can change into uracil (spontaneous deamination). This can lead to a point mutation if not repaired by the DNA repair enzymes such as uracil glycosylase, which cleaves a uracil in DNA.

Cytosine can also be methylated into 5-methylcytosine by an enzyme called DNA methyltransferase or be methylated and hydroxylated to make 5-hydroxymethylcytosine.

The difference in rates of deamination of cytosine and 5-methylcytosine (to uracil and thymine) forms the basis of bisulfite sequencing.

Biological function

When found third in a codon of RNA, cytosine is synonymous with uracil, as they are interchangeable as the third base.

When found as the second base in a codon, the third is always interchangeable. For example, UCU, UCC, UCA and UCG are all serine, regardless of the third base.

Active enzymatic deamination of cytosine or 5-methylcytosine by the APOBEC family of cytosine deaminases could have both beneficial and detrimental implications on various cellular processes as well as on organismal evolution.

The implications of deamination on 5-hydroxymethylcytosine, on the other hand, remains less understood.

THYMINE

Thymine T is one of the four nucleobases in the nucleic acid of DNA that are represented by the letters G–C–A–T.

The others are adenine, guanine, and cytosine.

Thymine is also known as **5 -methyluracil,** a pyrimidine nucleobase.

In RNA, thymine is replaced by the nucleobase uracil.

Thymine was first isolated in 1893 by Albrecht Kossel and Albert Neumann from calves' thymus glands, hence its name.

Derivation

As its alternate name (5-methyluracil) suggests, thymine may be derived by methylation of uracil at the 5th carbon.

In RNA, thymine is replaced with uracil in most cases.

In DNA, thymine (T) binds to adenine (A) via two hydrogen bonds, thereby stabilizing the nucleic acid structures.

Thymine combined with deoxyribose creates the nucleoside deoxythymidine, which is synonymous with the term thymidine.

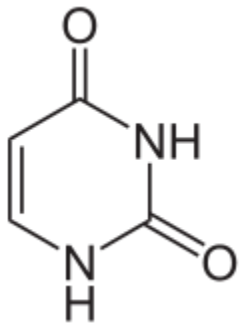
Thymidine can be phosphorylated with up to three phosphoric acid groups, producing dTMP (**d**eoxy**t**hymidine **m**ono**p**hosphate), dTDP, or dTTP (for the **d**i- and **t**ri- phosphates, respectively).

One of the common mutations of DNA involves two adjacent thymines or cytosine, which, in presence of ultraviolet light, may form thymine dimers, causing "kinks" in the DNA molecule that inhibit normal function.

Thymine could also be a target for actions of 5-fluorouracil (5-FU) in cancer treatment. 5-FU can be a metabolic analog of thymine (in DNA synthesis) or uracil (in RNA synthesis). Substitution of this analog inhibits DNA synthesis in actively dividing cells.

Thymine bases are frequently oxidized to hydantoins over time after the death of an organism.

URACIL



Uracil, U is one of the four nucleobases in the nucleic acid RNA that are represented by the letters A, G, C and U.

The others are adenine (A), cytosine (C), and guanine (G).

In RNA, uracil binds to adenine via two hydrogen bonds.

In DNA, the uracil nucleobase is replaced by thymine.

Uracil is a demethylated form of thymine.

Uracil is a common and naturally occurring pyrimidine derivative.

Properties

In RNA, uracil base-pairs with adenine and replaces thymine during DNA transcription. Methylation of uracil produces thymine.

In DNA, the evolutionary substitution of thymine for uracil may have increased DNA stability and improved the efficiency of DNA replication (discussed below). Uracil pairs with adenine through hydrogen bonding.

When base pairing with adenine, uracil acts as both a hydrogen bond acceptor and a hydrogen bond donor.

In RNA, uracil binds with a ribose sugar to form the ribonucleoside uridine. When a phosphate attaches to uridine, uridine 5'-monophosphate is produced.

Reactions

Uracil readily undergoes regular reactions including oxidation, nitration, and alkylation. While in the presence of phenol (PhOH) and sodium hypochlorite (NaOCl), uracil can be visualized in ultraviolet light.

Uracil also has the capability to react with elemental halogens because of the presence of more than one strongly electron donating group.

Uses

Uracil's use in the body is to help carry out the synthesis of many enzymes necessary for cell function through bonding with riboses and phosphates.

Uracil serves as allosteric regulator and coenzyme for reactions in animals and in plants.

UMP controls the activity of carbamoyl phosphate synthetase and aspartate transcarbamoylase in plants, while UDP and UTP requlate CPSase II activity in animals.

UDP-glucose regulates the conversion of glucose to galactose in the liver and other tissues in the process of carbohydrate metabolism.

Uracil is also involved in the biosynthesis of polysaccharides and the transportation of sugars containing aldehydes.

Uracil is important for the detoxification of many carcinogens, for instance those found in tobacco smoke.

PHOSPHODIESTER BOND

A **phosphodiester bond** occurs when exactly two of the hydroxyl groups in phosphoric acid react with hydroxyl groups on other molecules to form two ester bonds.

Phosphodiester bonds are central to all life on Earth as they make up the backbone of the strands of nucleic acid. In DNA and RNA, the phosphodiester bond is the linkage between the 3' carbon atom of one sugar molecule and the 5' carbon atom of another, deoxyribose in DNA and ribose in RNA.

Strong covalent bonds form between the phosphate group and two 5-carbon ring carbohydrates (pentoses) over two ester bonds.

The phosphate groups in the phosphodiester bond are negatively charged.

Because the phosphate groups have a pK_a near 0, they are negatively charged at pH 7.

This repulsion forces the phosphates to take opposite sides of the DNA strands and is neutralized by proteins (histones), metal ions such as magnesium, and polyamines.

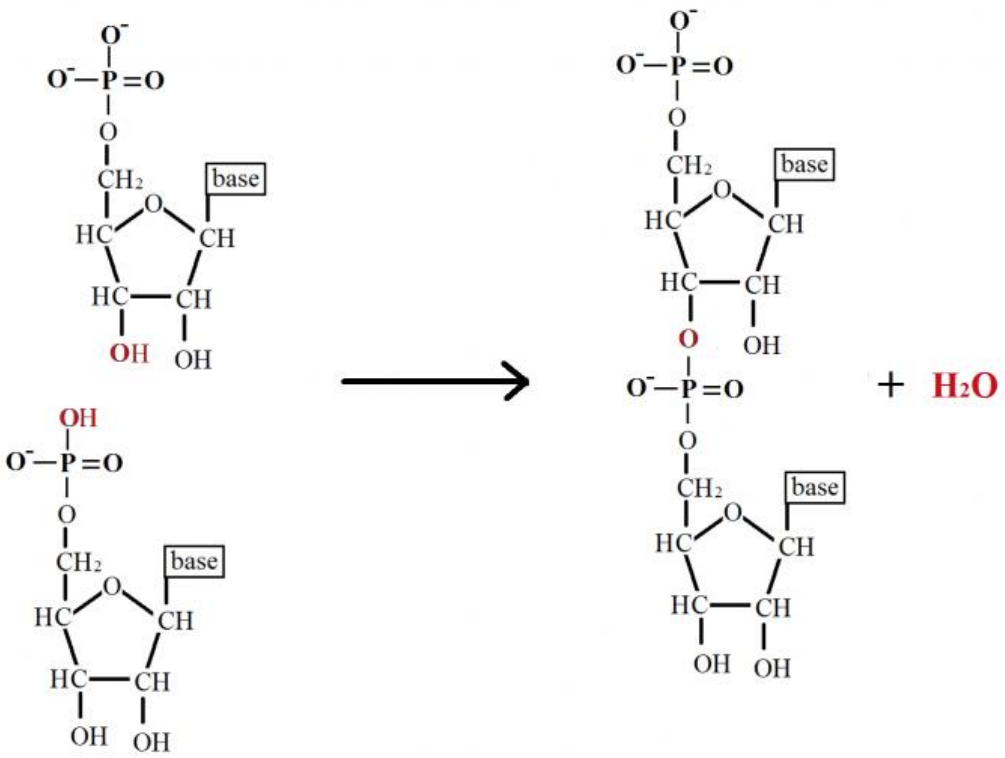
In order for the phosphodiester bond to be formed and the nucleotides to be joined, the tri-phosphate or di-phosphate forms of the nucleotide building blocks are broken apart to give off energy required to drive the enzyme-catalyzed reaction.

When a single phosphate or two phosphates known as pyrophosphates break away and catalyze the reaction, the phosphodiester bond is formed.

Hydrolysis of phosphodiester bonds can be catalyzed by the action of phosphodiesterases which play an important role in repairing DNA sequences.

The phosphodiester linkage between two ribonucleotides can be broken by alkaline hydrolysis, whereas the linkage between two deoxyribonucleotides is more stable under these conditions.

The relative ease of RNA hydrolysis is an effect of the presence of the 2' hydroxyl group.



Phosphodiester Bond Formation

In phosphodiester formation, two hydroxyl (OH) groups on the phosphate molecule bind to the 3' and 5' carbons on two independent pentose sugars.

These are two condensation reactions, so two molecules of water are produced.

The phosphate is then bonded to the sugars by two ester bonds, hence the nomenclature of phosphodiester bond. This reaction is catalysed by ligases, such as DNA ligase during DNA replication.

NUCLEOTIDES

Nucleic acids are linear, unbranched polymers of **nucleotides**.

1.

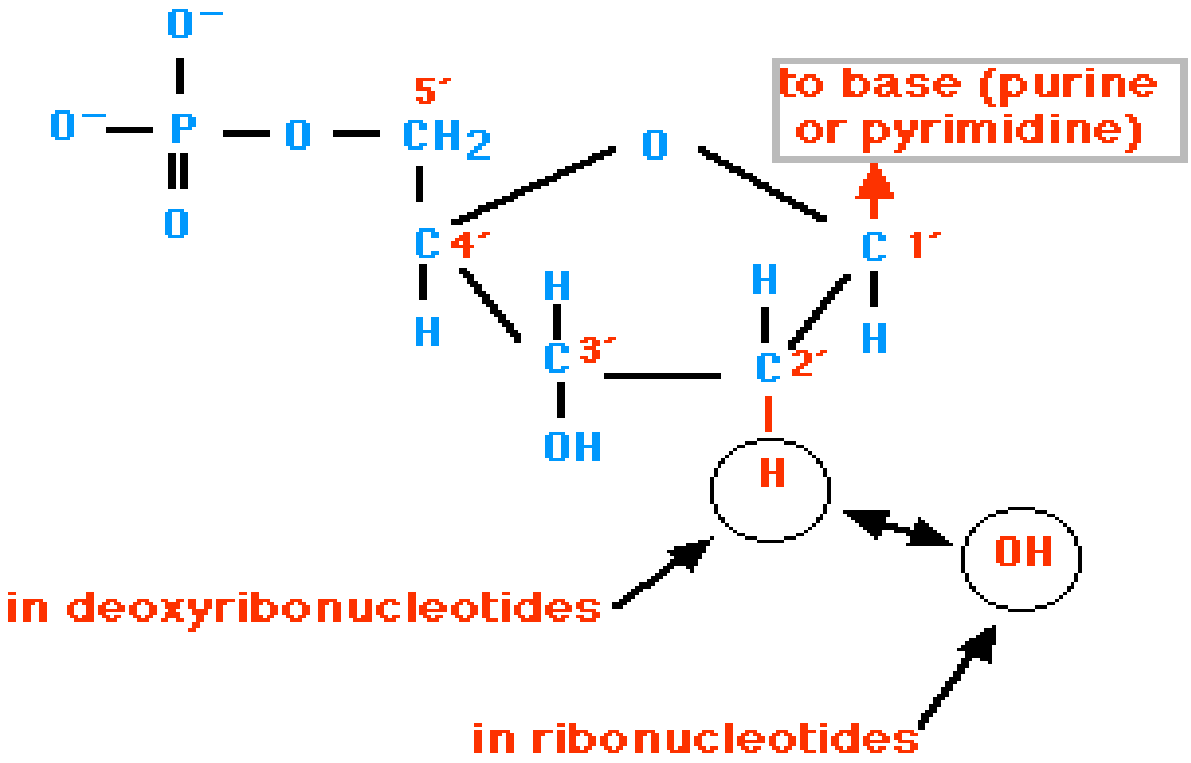
A five-carbon sugar (hence a **pentose**). Two kinds are found:

deoxyribose, which has a hydrogen atom attached to its #2 carbon atom (designated 2'), and

ribose, which has a hydroxyl group there.

Deoxyribose-containing nucleotides, the **deoxyribonucleotides**, are the monomers of deoxyribonucleic acids (**DNA)**.

Ribose-containing nucleotides, the **ribonucleotides**, are the monomers of ribonucleic acids (**RNA)**.



2.

A nitrogen-containing ring structure called a **nucleobase** (or simply a base). The nucleobase is attached to the 1' carbon atom of the pentose. In **DNA**, four different nucleobases are found:

two **purines**, called **adenine** (**A**) and **guanine** (**G**)

two **pyrimidines**, called **thymine** (**T**) and **cytosine** (**C**)

RNA contains:

The same purines, **adenine** (**A**) and **guanine** (**G**).

RNA also uses the pyrimidine **cytosine** (**C**), but instead of thymine, it uses the pyrimidine **uracil** (**U**).

The combination of a nucleobase and a pentose is called a **nucleoside**.

3.

One (as shown in the first figure), two, or three **phosphate** groups. These are attached to the 5' carbon atom of the pentose. The product in each case is called a **nucleotide**.

Both DNA and RNA are assembled from **nucleoside triphosphates**.

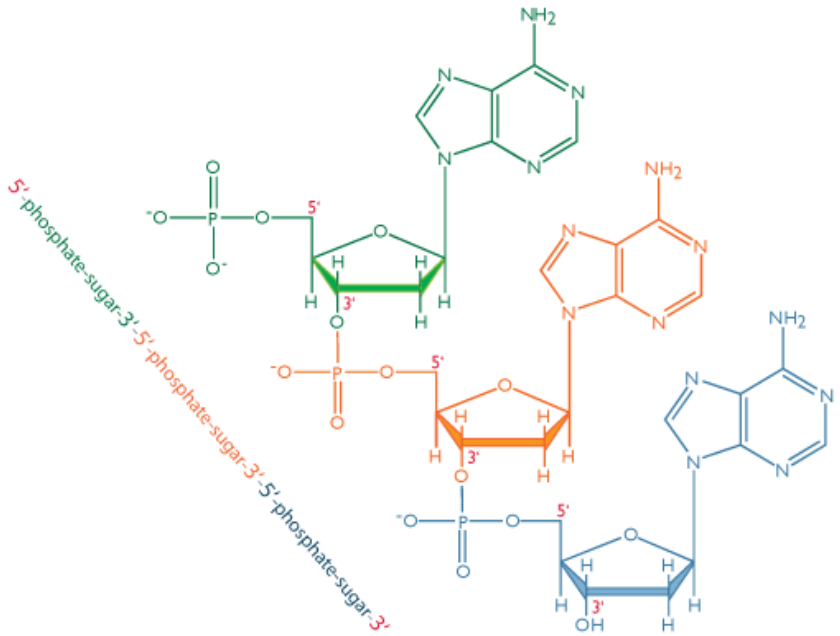
For **DNA**, these are **dATP**, **dGTP**, **dCTP**, and **dTTP**.

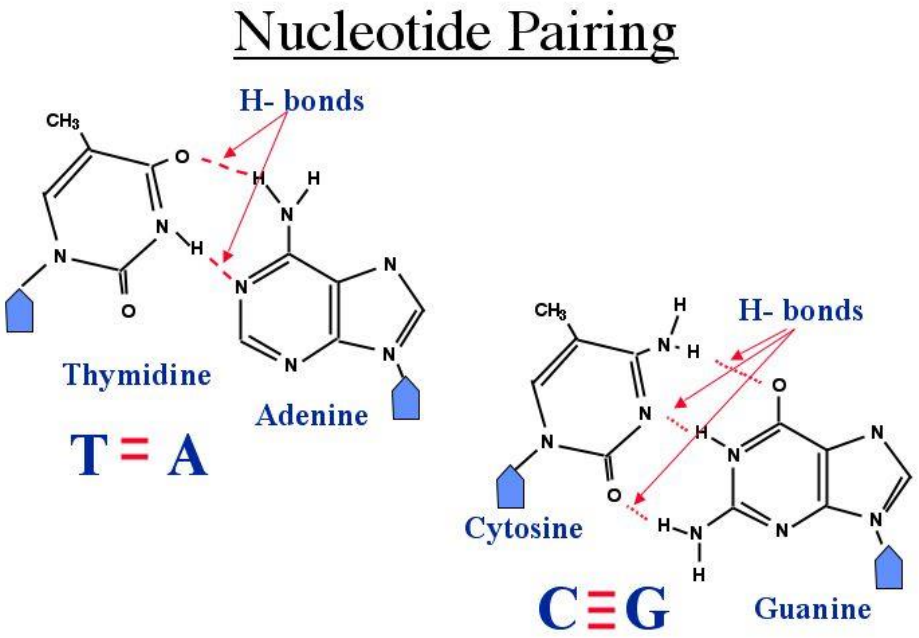
For **RNA**, these are **ATP**, **GTP**, **CTP**, and **UTP**.

In both cases, as each nucleotide is attached, the second and third phosphates are removed.

The nucleosides and their mono-, di-, and triphosphates

	Nucleobase	**Nucleoside**	**Nucleotides**		
DNA	Adenine (A)	Deoxyadenosine	dAMP	dADP	dATP
	Guanine (G)	Deoxyguanosine	dGMP	dGDP	dGTP
	Cytosine (C)	Deoxycytidine	dCMP	dCDP	dCTP
	Thymine (T)	Deoxythymidine	dTMP	dTDP	dTTP
RNA	Adenine (A)	Adenosine	AMP	ADP	ATP
	Guanine (G)	Guanosine	GMP	GDP	GTP
	Cytosine (C)	Cytidine	CMP	CDP	CTP
	Uracil (U)	Uridine	UMP	UDP	UTP





The nucleic acids, both DNA and RNA, consist of polymers of nucleotides. The nucleotides are linked covalently between the 3' carbon atom of the pentose and the phosphate group attached to the 5' carbon of the adjacent pentose.

Most intact DNA molecules are made up of **two** strands of polymer, forming a "double helix".

RNA molecules, while single-stranded, usually contain regions where two portions of the strand twist around each other to form helical regions

The two strands of DNA and the helical regions of RNA are held together by base pairing.

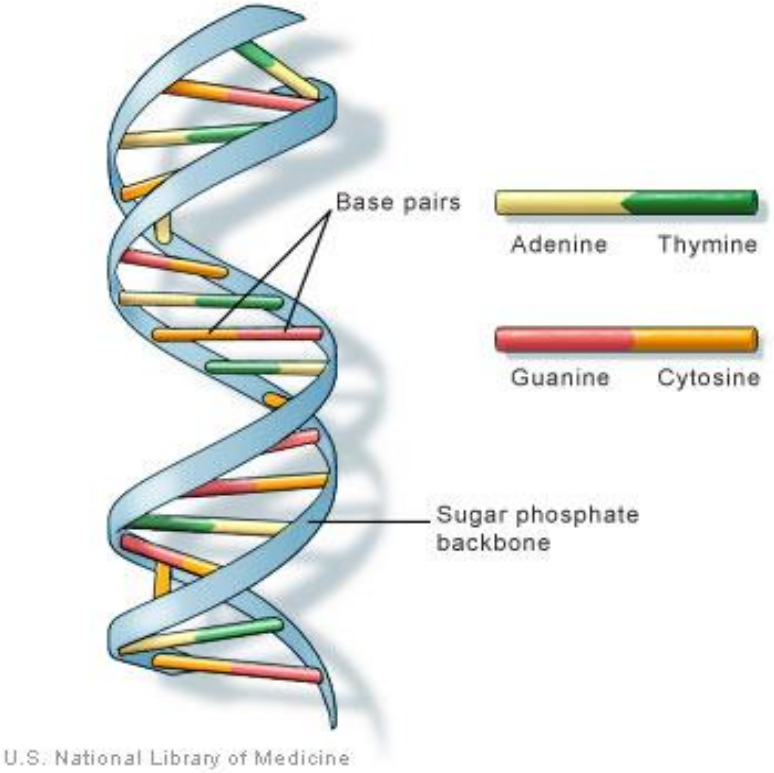
DOUBLE HELIX

The double helix of DNA has these features:

- It contains two polynucleotide strands wound around each other.
- The backbone of each consists of alternating deoxyribose and phosphate groups.
- The phosphate group bonded to the 5' carbon atom of one deoxyribose is covalently bonded to the 3' carbon of the next.
- The two strands are "antiparallel"; that is, one strand runs 5′ to 3′ while the other runs 3′ to 5′.
- The DNA strands are assembled in the 5′ to 3′ direction [More] and, by convention, we "read" them the same way.
- The purine or pyrimidine attached to each deoxyribose projects in toward the axis of the helix.
- Each base forms hydrogen bonds with the one directly opposite it, forming **base pairs** (also called nucleotide pairs).
- 3.4 Å separate the planes in which adjacent base pairs are located.
- The double helix makes a complete turn in just over 10 nucleotide pairs, so each turn takes a little more (35.7 Å to be exact) than the 34 Å shown in the diagram.
- There is an average of 25 hydrogen bonds within each complete turn of the double helix providing a stability of binding about as strong as what a covalent bond would provide.
- The diameter of the helix is 20 Å.

- The helix can be virtually any length; when fully stretched, some DNA molecules are as much as 5 cm (2 inches!) long.
- The path taken by the two backbones forms a major (wider) groove (from "34 A" to the top of the arrow) and a minor (narrower) groove (the one below).

This structure of DNA was worked out by **Francis Crick** and **James D. Watson** in 1953.



DNA REPLICATION

It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material' – **Watson & Crick** ***Nature (1953)***

Process of duplication of the entire genome prior to cell division

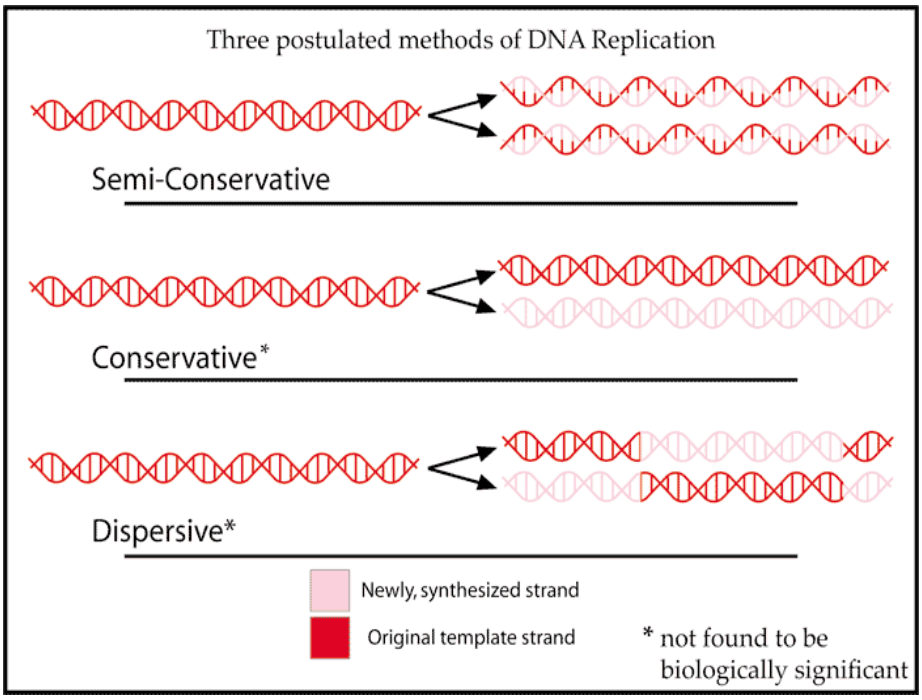
Biological significance

- extreme accuracy of DNA replication is necessary in order to preserve the integrity of the genome in successive generations
- In eukaryotes , replication only occurs during the S phase of the cell cycle.
- Replication rate in eukaryotes is slower resulting in a higher fidelity/accuracy of replication in eukaryotes

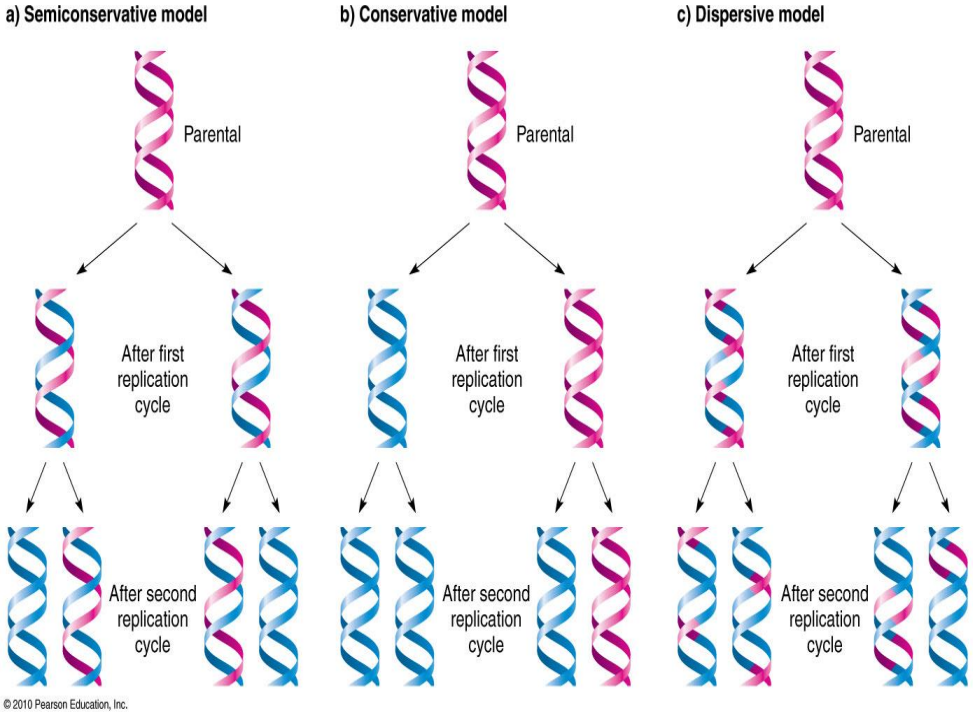
BASIC RULES OF REPLICATION

A. Semi-conservative
B. Starts at the 'origin'
C. Synthesis always in the 5-3' direction
D. Can be uni or bidirectional
E. Semi-discontinuous
F. RNA primers required

A. Semiconservative



3 MODES OF REPLICATION

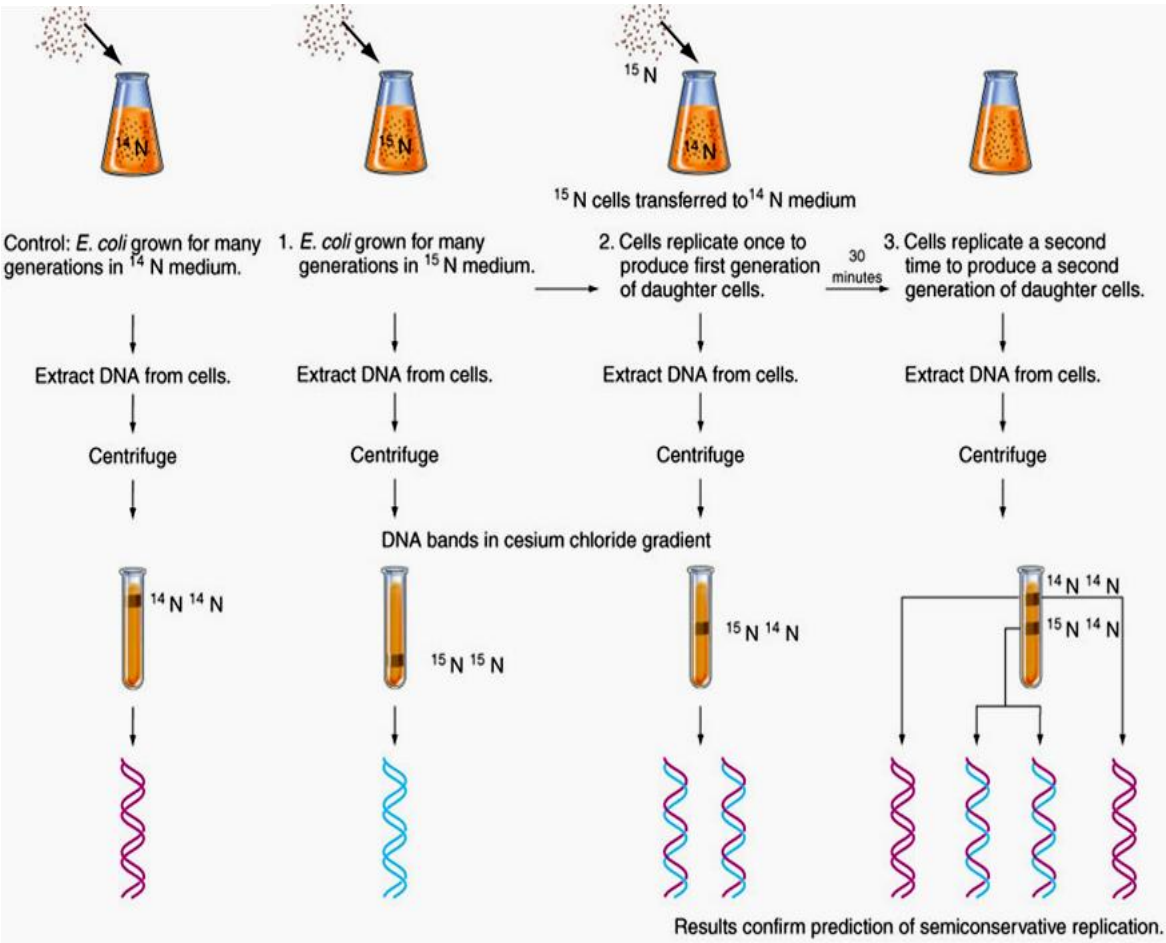


Meselson and Stahl's experiment. M. Meselson and **F.W. Stahl** (1958) verified the semiconservative nature of DNA replication in a series of elegant experiments using isotopically labelled DNA and a form of isopycnic density gradient centrifugation. They cultured *Escherichai coli* cells in a medium in which the nitrogen was 15N (a 'heavy' isotope of nitrogen, but not a radioisotope) instead of commonly occurring and lighter 14N. In time, the purines and pyrimidines of DNA in new cells contained 15N (where 14N normally occurs) and, thus, the DNA molecules were denser. DNA in which the nitrogen atoms are heavy (15N) can be distinguished from DNA containing light nitrogen (14N), because during isopycnic centrifugation, the two different DNAs band at different density positions in the centrifuge tube.

Depending on its content of 15N and 14N, the DNA bands at a specific position in the density gradient. Because the DNA synthezised by *E.coli.* cells grown in 15N would be denser than 14Ncontaining DNA, it would band further down the tube. *E. coli* cells grown for sometime in the presence of 15N-medium were washed free of the medium and transferred to 14N-containing medium and allowed to continue to grow for specific lengths of time (*i.e*., for various numbers of generation time). DNA isolated from cells grown for one generation of time in the 14N medium had a density intermediate to that of the DNA from cells grown only in 15N-containing medium (identified as *generation O;* Fig. 4.3) and that of DNA from cells grown only in 14N-containing medium (*the controls*). Such a result immediately ruled out the possibility that DNA replication was conservative, because the conservative replication would have yielded two DNA bands in the density gradient for **generation 1** (*i.e.,* F1 cells).

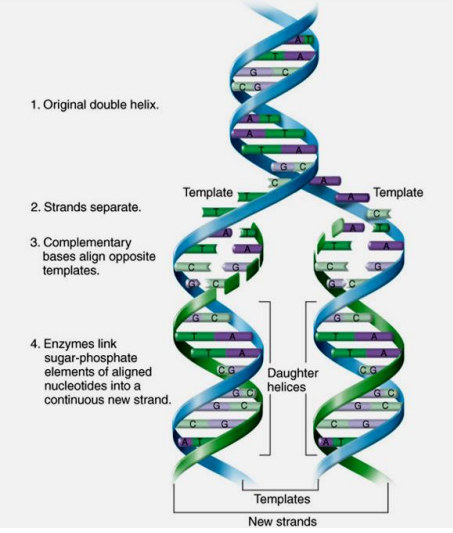
The single band of intermediate density (identified as ''**hybrid DNA'')** consisted of DNA molecules in which one strand contained 15N and the other contained 14N. When the incubation in the 14N-medium was carried out for two generations of time (*i.e*. **generation 2**), two DNA bands were formed — one at the same density position as the DNA from cells grown exclusively in 14N medium (*i.e.,* light controls) and the other of intermediate density. Subsequent generations produced greater numbers of DNA molecules

MESELSON – STAHL EXPERIMENT



Semi-conservative replication:

One strand of duplex passed on unchanged to each of the daughter cells. This 'conserved' strand acts as a template for the synthesis of a new, complementary strand by the enzyme DNA polymerase



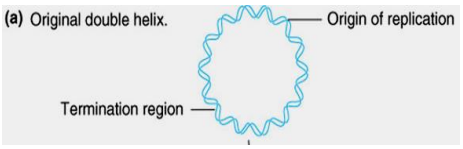
B. STARTS AT ORIGIN OF REPLICATION

Initiator proteins identify specific base sequences on DNA called sites of origin

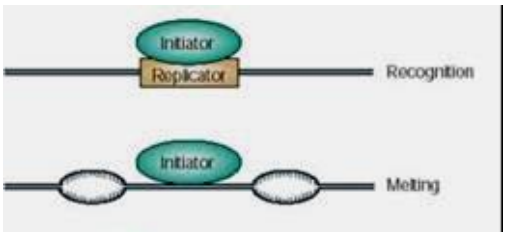
Prokaryotes – single origin site E.g *E.coli* - *oriC*

Eukaryotes – multiple sites of origin (replicator)

E.g. yeast - ARS (autonomously replicating sequences)

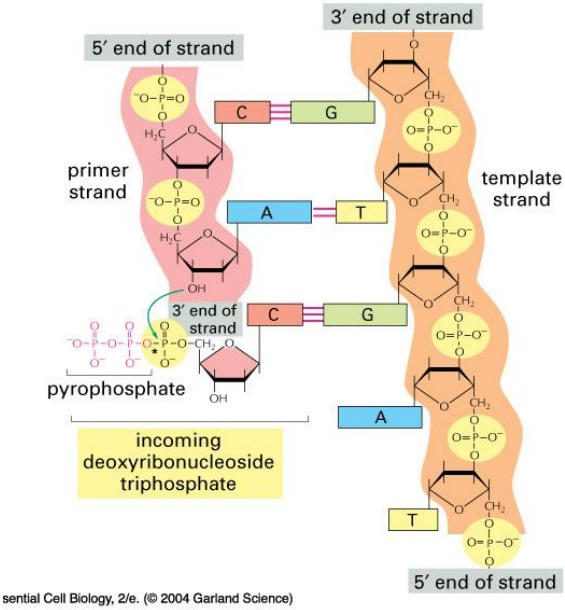


Prokaryotes



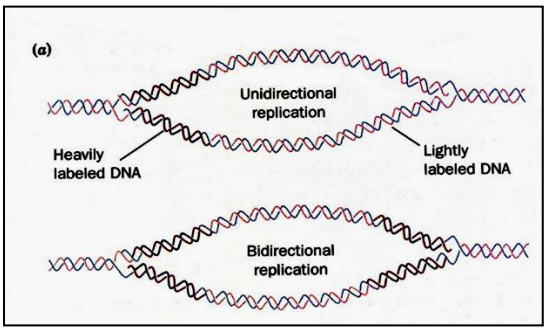
Eukaryotes

C. SYNTHESIS ALWAYS IN 5'-3'DIRECTION



sential Cell Biology, 2/e. (© 2004 Garland Science)

D. UNI OR BIDIRECTIONAL

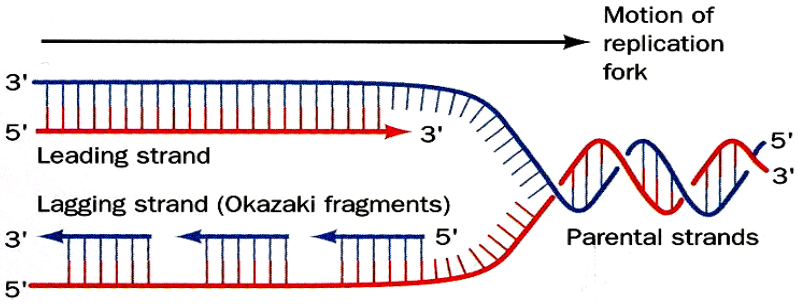


E. SEMI DISCONTINUOUS REPLICATION

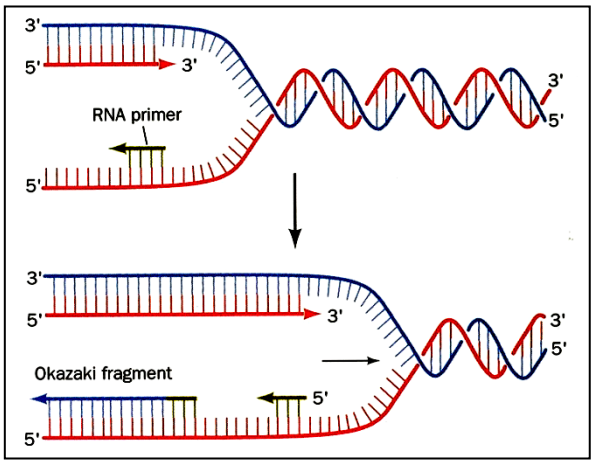
Anti parallel strands replicated simultaneously

Leading strand synthesis **continuously** in 5'– 3'

Lagging strand synthesis in **fragments** in 5'-3'



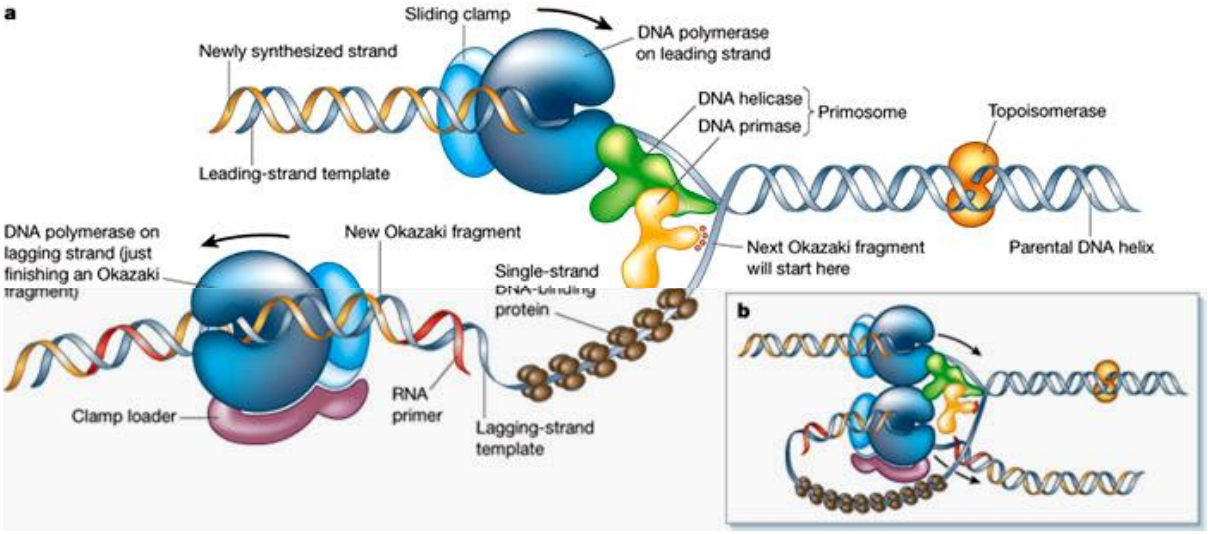
F. RNA PRIMERS REQUIRED



THE MECHANISM OF DNA REPLICATION

Arthur Kornberg, a Nobel prize winner and other biochemists deduced steps of replication

- Initiation
  - Proteins bind to DNA and open up double helix
  - Prepare DNA for complementary base pairing
- Elongation
  - Proteins connect the correct sequences of nucleotides into a continuous new strand of DNA
- Termination
  - Proteins release the replication complex



Bacterial Replication Requires a Large Number of Enzymes and Proteins

Replication takes place in four stages: initiation, unwinding, elongation, and termination. The following discussion of the process of replication will focus on bacterial systems, where replication has been most thoroughly studied and is best understood. Although many aspects of replication in eukaryotic cells are similar to those in prokaryotic cells, there are some important differences. We will compare bacterial and eukaryotic replication later in the chapter.

Initiation

The circular chromosome of *E. coli* has a single replication origin (*oriC*). The minimal sequence required for *oriC* to function consists of 245 bp that contain several critical sites. An **initiator protein** (known as DnaA in *E. coli*) binds to *oriC* and causes a short section of DNA to unwind. This unwinding allows helicase and other single- strand-binding proteins to attach to the polynucleotide strand.

Unwinding

Because DNA synthesis requires a single-stranded template and because double-stranded DNA must be unwound before DNA synthesis can take place, the cell relies on several proteins and enzymes to accomplish the unwinding.

DNA helicase A **DNA helicase** breaks the hydrogen bonds that exist between the bases of the two nucleotide strands of a DNA molecule. Helicase cannot *initiate* the unwinding of double-stranded DNA; the initiator protein first separates DNA strands at the origin, providing a short stretch of single-stranded DNA to which a helicase binds. Helicase binds to the lagging-strand template at each replication fork and moves in the 5'→3′ direction along this strand, thus also moving the replication fork.

Single-strand-binding proteins After DNA has been unwound by helicase, **single-strand-binding proteins** (SSBs) attach tightly to the exposed single-stranded DNA. These proteins protect the single-stranded nucleotide chains and prevent the formation of secondary structures such as hairpins (see Figure 10.17) that interfere with replication.

Unlike many DNA-binding proteins, SSBs are indifferent to base sequence: they will bind to any single-stranded DNA. Single- strand-binding proteins form tetramers (groups of four); each tetramer covers from 35 to 65 nucleotides.

DNA gyrase Another protein essential for the unwinding process is the enzyme **DNA gyrase**, a topoisomerase. Topoisomerases control the supercoiling of DNA. In replication, DNA gyrase reduces the torsional strain (torque) that builds up ahead of the replication fork as a result of unwinding. It reduces torque by making a double-stranded break in one segment of the DNA helix, passing another segment of the helix through the break, and then resealing the broken ends of the DNA. This action removes a twist in the DNA and reduces the supercoiling.

A group of antibiotics called 4-quinolones kill bacteria by binding to DNA gyrase and inhibiting its action. The inhibition of DNA gyrase results in the cessation of DNA synthesis and bacterial growth. Many bacteria have acquired resistance to quinolones through mutations in the gene for DNA gyrase.

Elongation

During the elongation phase of replication, single-stranded DNA is used as a template for the synthesis of DNA. This process requires a series of enzymes.

Synthesis of primers All DNA polymerases require a nucleotide with a 3′-OH group to which a new nucleotide can be added. Because of this requirement, DNA polymerases cannot initiate DNA synthesis on a bare template; rather, they require a primer—an existing 3′-OH group—to get started. How, then, does DNA synthesis begin?

An enzyme called **primase** synthesizes short stretches of nucleotides, or **primers**, to get DNA replication started. Primase synthesizes a short stretch of RNA nucleotides (about 10–12 nucleotides long), which provides a 3′-OH group to which DNA polymerase can attach DNA nucleotides. (Because primase is an RNA polymerase, it does not require a 3′-OH group to which nucleotides can be added.) All DNA molecules initially have short RNA primers embedded within them; these primers are later removed and replaced by DNA nucleotides. On the leading strand, where DNA synthesis is continuous, a primer is required only at the 5′ end of the newly synthesized strand. On the lagging strand, where replication is discontinuous, a new primer must be generated at the beginning of each Okazaki fragment. Primase forms a complex with helicase at

the replication fork and moves along the template of the lagging strand. The single primer on the leading strand is probably synthesized by the primase–helicase complex on the template of the lagging strand of the *other* replication fork, at the opposite end of the replication bubble.

Two of them, DNA polymerase I and DNA polymerase III, carry out DNA synthesis in replication; the other three have specialized functions in DNA repair. **DNA polymerase III** is a large multiprotein complex that

acts as the main workhorse of replication. DNA polymerase III synthesizes nucleotide strands by adding new nucleotides to the 3′ end of a growing DNA molecule. This enzyme has two enzymatic activities (see Table 12.3). Its 5′→3′ polymerase activity allows it to add new nucleotides in the 5′→3′ direction. Its 3′→5′ exonuclease activity allows it to remove nucleotides in the 3′→5′ direction, enabling it to correct

errors. If a nucleotide having an incorrect base is inserted into the growing DNA molecule, DNA polymerase III uses its 3′→5′ exonuclease activity to back up and remove the incorrect nucleotide. It then resumes its 5′→3′ polymerase activity. These two functions together allow DNA polymerase III to efficiently and accurately synthesize new DNA molecules.

DNA polymerase III has high processivity, which means that it is capable of adding many nucleotides to the growing DNA strand without releasing the template: it normally holds on to the template and continues synthesizing DNA until the template has been completely replicated. The high processivity of DNA polymerase III is ensured by one of the polypeptides that constitutes the enzyme. This polypeptide, termed the β subunit, serves as a clamp for the polymerase enzyme: it encircles the DNA and keeps the DNA polymerase attached to the template strand during replication. DNA polymerase III adds DNA nucleotides to the primer, synthesizing the DNA of both the leading and the lagging strands.

The first *E. coli* polymerase to be discovered, **DNA polymerase I**, also has 5′→3′ polymerase and 3′→5′ exonuclease activities (see Table 12.3), permitting the enzyme to synthesize DNA and to correct errors. Unlike DNA polymerase III, however, DNA polymerase I also possesses 5′→3′ exonuclease activity, which is used to remove the primers laid down by primase and to replace them with DNA nucleotides by synthesizing in a 5′→3′ direction. DNA polymerase I has lower processivity than DNA polymerase III. The removal and replacement of primers appear to constitute the main function of DNA polymerase I. After DNA polymerase III has initiated synthesis at the primer and moved downstream, DNA polymerase I removes the RNA nucleotides of the primer, replacing them with DNA nucleotides. DNA polymerases II, IV, and V function in DNA repair.

Despite their differences, all of *E. coli*'s DNA polymerases

1. synthesize any sequence specified by the template strand;

2. synthesize in the 5′→3′ direction by adding nucleotides to a 3′-OH group;

3. use dNTPs to synthesize new DNA;

4. require a primer to initiate synthesis;

5. catalyze the formation of a phosphodiester bond by joining the 5′-phosphate group of the incoming nucleotide to the 3′-OH group of the preceding nucleotide on the growing strand, cleaving off two phosphates in the process;

6. produce newly synthesized strands that are complementary and antiparallel to the template strands; and

7. are associated with a number of other proteins.

DNA ligase After DNA polymerase III attaches a DNA nucleotide to the 3′-OH group on the last nucleotide of the RNA primer, each new DNA nucleotide then provides the 3′-OH group needed

for the next DNA nucleotide to be added. This process continues as long as template is available. DNA polymerase I follows DNA polymerase III and, using its 5′→3′ exonuclease activity, removes the RNA primer.

It then uses its 5′→3′ polymerase activity to replace the RNA nucleotides with DNA nucleotides. DNA polymerase I attaches the first nucleotide to the OH group at the 3′ end of the preceding Okazaki fragment and then continues, in the 5′→3′ direction along the nucleotide strand, removing and replacing, one

at a time, the RNA nucleotides of the primer.

After polymerase I has replaced the last nucleotide of the RNA primer with a DNA nucleotide, a nick remains in

the sugar–phosphate backbone of the new DNA strand. The 3′-OH group of the last nucleotide to have been added by DNA polymerase I is not attached to the 5′-phosphate group of the first nucleotide added by DNA polymerase III. This nick is sealed by the enzyme **DNA ligase**, which catalyzes the formation of a phosphodiester bond without adding another nucleotide to the strand. Some of the major enzymes and proteins required for prokaryotic DNA replication are summarized in **Table**.

Elongation at the replication fork

Now that the major enzymatic components of elongation—DNA polymerases, helicase, primase, and ligase—have been introduced, let's consider how these components interact at the replication fork. Because the synthesis of both strands takes place simultaneously, two units of DNA polymerase III must be present at the replication fork, one for each strand. In one model of the replication process, the two units of DNA polymerase III are connected; the lagging-strand template loops around so that it is in position for 5′→3′ replication. In this way, the DNA polymerase III complex is able to carry out 5′→3′ replication simultaneously on both templates, even though they run in opposite directions. After about 1000 bp of new DNA has been synthesized, DNA polymerase III releases the lagging strand template, and a new loop forms (see Figure 12.15).

Primase synthesizes a new primer on the lagging strand and DNA polymerase III then synthesizes a new Okazaki fragment.

In summary, each active replication fork requires five basic components

1. helicase to unwind the DNA,

2. single-strand-binding proteins to protect the single nucleotide strands and prevent secondary structures,

3. the topoisomerase gyrase to remove strain ahead of the replication fork,

4. primase to synthesize primers with a 3′-OH group at the beginning of each DNA fragment, and

5. DNA polymerase to synthesize the leading and lagging nucleotide strands.

Termination

In some DNA molecules, replication is terminated whenever two replication forks meet. In others, specific termination sequences block further replication. A termination protein, called Tus in *E. coli*, binds to these sequences. Tus blocks the movement of helicase, thus stalling the replication fork and preventing further DNA replication.

The Fidelity of DNA Replication

Overall, the error rate in replication is less than one mistake per billion nucleotides. How is this incredible accuracy achieved?

DNA polymerases are very particular in pairing nucleotides with their complements on the template strand. Errors in nucleotide selection by DNA polymerase arise only about once per 100,000 nucleotides. Most of the errors that do arise in nucleotide selection are corrected in a second process called **proofreading**. When a DNA polymerase inserts an incorrect nucleotide into the growing strand, the 3′-OH group of the mispaired nucleotide is not correctly positioned in the active site of the DNA polymerase for accepting the next nucleotide. The incorrect positioning stalls the polymerization reaction, and the 3′→5′ exonuclease activity of DNA polymerase removes the incorrectly paired nucleotide. DNA polymerase then inserts the correct nucleotide. Together, proofreading and nucleotide selection result in an error rate of only one in 10 million nucleotides.

A third process, called **mismatch repair** (discussed further in Chapter 18), corrects errors after replication is complete. Any incorrectly paired nucleotides remaining after replication produce a deformity in the secondary structure of the DNA; the deformity is recognized by enzymes that excise an incorrectly paired nucleotide and use the original nucleotide strand as a template to replace the incorrect nucleotide. Mismatch repair requires the ability to distinguish between the old and the new strands of DNA, because the enzymes need some way of determining which of the two

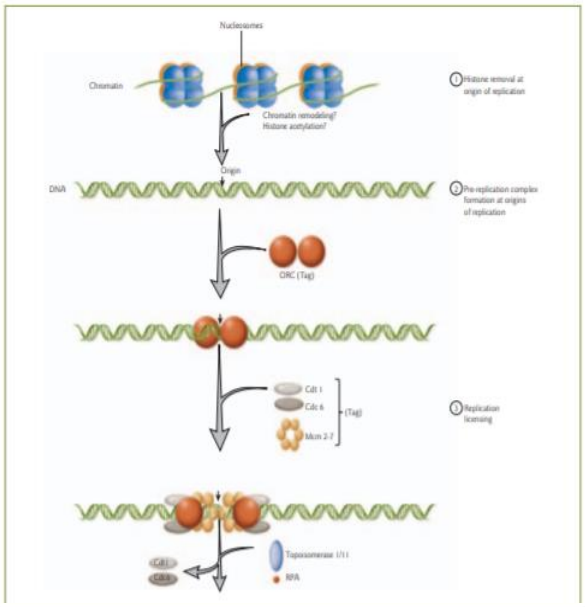
incorrectly paired bases to remove. In *E. coli*, methyl groups (−CH3) are added to particular nucleotide sequences, but only *after* replication. Thus, immediately after DNA synthesis, only the old DNA strand is methylated. It can therefore be distinguished from the newly synthesized strand, and mismatch repair takes place preferentially on the unmethylated nucleotide strand. No single process could produce this level of accuracy; a series of processes are required, each process catching errors missed by the preceding ones.

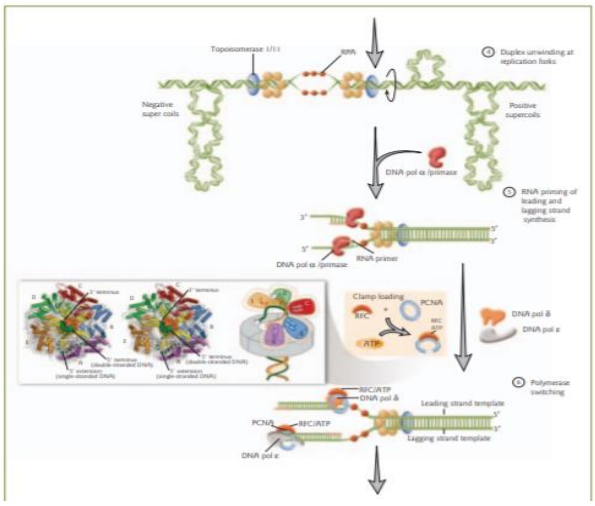
The Basic Rules of Replication

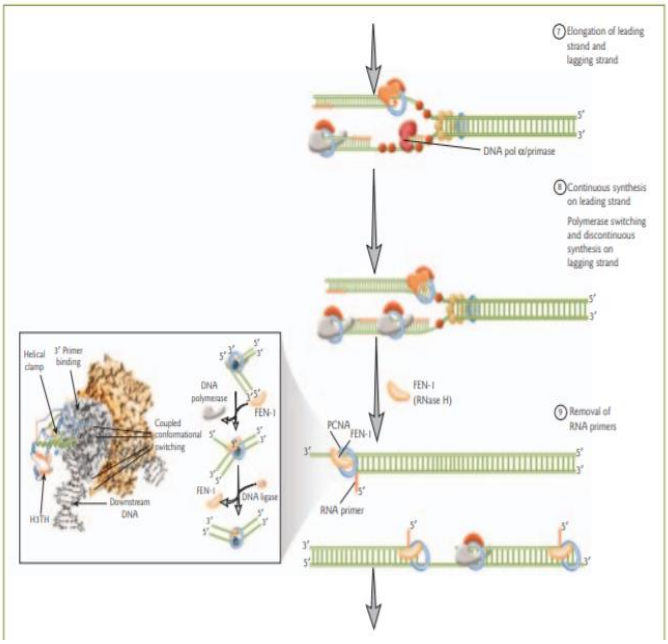
Bacterial replication requires a number of enzymes , proteins, and DNA sequences that function together to synthesize a new DNA molecule. These components are important, but we must not become so immersed in the details of the process that we lose sight of the general principles of replication.

1. Replication is always semiconservative.
2. Replication begins at sequences called origins.
3. DNA synthesis is initiated by short segments of RNA called primers.
4. The elongation of DNA strands is always in the 5′→3′ direction.
5. New DNA is synthesized from dNTPs; in the polymerization of DNA, two phosphate groups are cleaved from a dNTP and the resulting nucleotide is added to the 3′-OH group of the growing nucleotide strand.
6. Replication is continuous on the leading strand and discontinuous on the lagging strand.
7. New nucleotide strands are complementary and antiparallel to their template strands.
8. Replication takes place at very high rates and is astonishingly accurate, thanks to precise nucleotide selection, proofreading, and repair mechanisms.

EUKARYOTIC REPLICATION


Nucleosomes
Chromatin
1 Histone removal at origin of replication
Chromatin remodeling?
Histone acetylation?
Origin
DNA
2 Pre-replication complex formation at origins of replication
ORC (Tag)
Cdt 1
Cdc 6
(Tag)
Mcm 2-7
3 Replication licensing
Topoisomerase I/II
RPA


Topoisomerase I/II
RPA
4 Duplex unwinding at replication forks
Negative super coils
Positive supercoils
DNA pol α /primase
5 RNA priming of leading and lagging strand synthesis
3'
5'
DNA pol α /primase
RNA primer
Clamp loading
RFC
PCNA
ATP
DNA pol δ
DNA pol ε
6 Polymerase switching
RFC/ATP
DNA pol δ
Leading strand template
PCNA
RFC/ATP
Lagging strand template
DNA pol ε



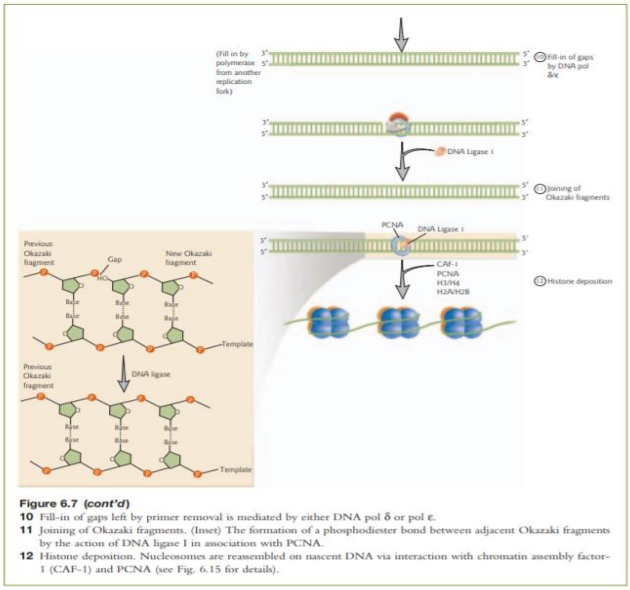


Figure 6.7 (*cont'd*)
10 Fill-in of gaps left by primer removal is mediated by either DNA pol δ or pol ε.
11 Joining of Okazaki fragments. (Inset) The formation of a phosphodiester bond between adjacent Okazaki fragments by the action of DNA ligase I in association with PCNA.
12 Histone deposition. Nucleosomes are reassembled on nascent DNA via interaction with chromatin assembly factor-1 (CAF-1) and PCNA (see Fig. 6.15 for details).

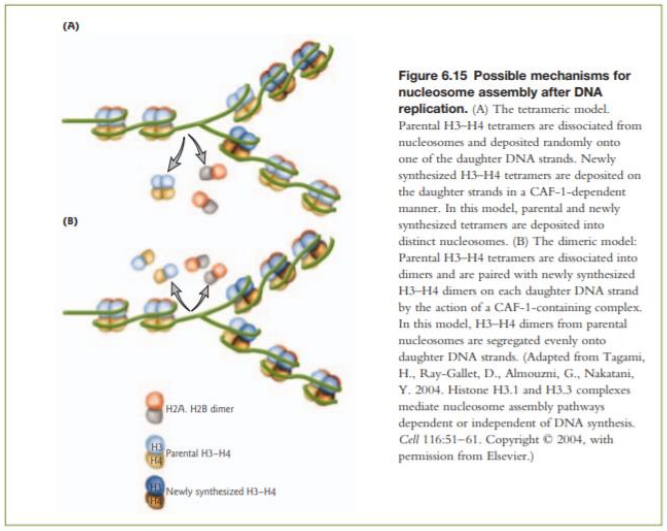


Figure 6.15 Possible mechanisms for nucleosome assembly after DNA replication. (A) The tetrameric model. Parental H3–H4 tetramers are dissociated from nucleosomes and deposited randomly onto one of the daughter DNA strands. Newly synthesized H3–H4 tetramers are deposited on the daughter strands in a CAF-1-dependent manner. In this model, parental and newly synthesized tetramers are deposited into distinct nucleosomes. (B) The dimeric model: Parental H3–H4 tetramers are dissociated into dimers and are paired with newly synthesized H3–H4 dimers on each daughter DNA strand by the action of a CAF-1-containing complex. In this model, H3–H4 dimers from parental nucleosomes are segregated evenly onto daughter DNA strands. (Adapted from Tagami, H., Ray-Gallet, D., Almouzni, G., Nakatani, Y. 2004. Histone H3.1 and H3.3 complexes mediate nucleosome assembly pathways dependent or independent of DNA synthesis. *Cell* 116:51–61. Copyright © 2004, with permission from Elsevier.)

DNA Replication in Eukaryotes

Eukaryotic DNA replication requires two different DNA polymerase enzymes, namely **DNA polymerase** α and **DNA polymerase** δ. DNA polymerase δ synthesizes the DNA on the leading strand (continuous DNA synthesis), whereas DNA polymerase α synthesizes the DNA on the lagging strand (discontinuous DNA synthesis). Besides these two enzymes, six more factors are involved in eukaryotic DNA replication: (1) T antigen; (2) replication factor A or RF-A (also called RP-A or eukaryotic SSB); (3) topoisomerase I; (4) topoisomerase II; (5) proliferating - cell nuclear antigen (PCNA, also called cyclin), and (6) replication factor Cor RF-C.

The process of eukaryotic DNA replication involves the following steps:

1. Before the onset of DNA synthesis, there is a presynthetic stage of 8-1 0 minutes duration for the formation of unwound DNA complex. This step needs only three purified proteins, namely T antigen (T-ag or tumour antigen), RF-A and topiosomerases I and II.
2. The T-antigen, using its DNA-binding domain, forms a multi-subunit complex with site I and site II in the presence of A TP and caused local unwinding.
3. More extensive duplex unwinding occurs due to association ofRF-A and a topoisomerase with the help of DNA helicase component ofT-ago Topoisomerases help in unwinding of DNA by altering topology of DNA at the replication fork.
4. RF-A or SSB proteins bind to unwound single stranded DNA.
5. The primer RNA synthesis is performed by primase which is tightly associated with DNA Polymerase α.
6. DNA polymerase α helps in synthesis of an okazaki fragment in 5' to 3' direction.

7. Replication factor C (or RF-C) and PCNA (cyclin) help in switching of DNA polymerases so that pol α is replaced by pol δ which then continuously synthesized DNA on the leading strand.
8. Another okazaki fragment is then synthesized from the replication fork on the lagging strand by pol α - primase complex and this step is repeated again and again, till the .entire DNA molecule is covered.
9. The RNA primers are removed and the gaps are filled as in prokaryotic DNA replication. Recently, role of DNA polymerase E in DNA replication has been stressed upon, so that three DNA polymerases (α, β and ε) are now known to be involved in eukaryotic DNA replication.
A. Sugino and coworkers have proposed that DNA polymerase a might function at both the leading and lagging strands (since polymerase a has a primase activity), whereas polymerase E and polymerase δ are involved in elongation of the leading and lagging-strands respectively.

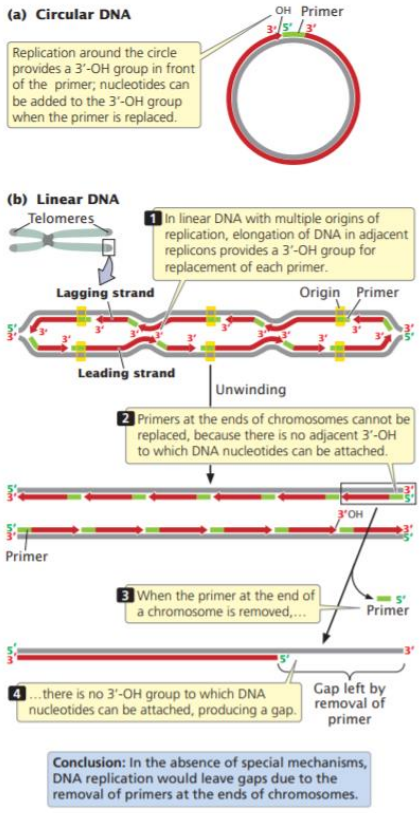
Replication at the Ends of Chromosomes

A fundamental difference between eukaryotic and bacterial replication arises because eukaryotic chromosomes are linear and thus have ends.

As already stated, the 3′-OH group needed for replication by DNA polymerases is provided at the initiation of replication by RNA primers that are synthesized by primase.

This solution is temporary because, eventually, the primers must be removed and replaced by DNA nucleotides. In a circular DNA molecule, elongation around the circle eventually provides a 3′-OH group immediately in front of the primer.

After the primer has been removed, the replacement DNA nucleotides can be added to this 3′-OH group.



12.18 DNA synthesis at the ends of circular and linear chromosomes must differ.

The end-replication problem

In linear chromosomes with multiple origins, the elongation of DNA in adjacent replicons also provides a 3′-OH group preceding each primer (Figure 12.18b).

At the very end of a linear chromosome, however, there is no adjacent stretch of replicated DNA to provide this crucial 3′-OH group. When the primer at the end of the chromosome has been removed, it cannot be replaced by DNA nucleotides, which produces a gap at the end of the chromosome, suggesting that the chromosome should become progressively shorter with each round of replication.

Chromosome shortening would mean that, when an organism reproduced, it would pass on shorter chromosomes than it had inherited.

Chromosomes would become shorter with each new generation and would eventually destabilize. This situation has been termed the end-replication problem.

Chromosome shortening does in fact take place in many somatic cells but, in single-celled organisms, germ cells, and early embryonic cells, chromosmes do not shorten and self-destruct. So how are the ends of linear chromosomes replicated?

Telomeres

To prevent the loss of genes as chromosome ends wear down, the tips of eukaryotic chromosomes have specialized DNA “caps” called **telomeres**. Telomeres consist of hundreds or thousands of repeats of the same short DNA sequence, which varies between organisms but is 5'-TTAGGG-3' in humans and other mammals.

Telomeres protect the end of the chromosome from DNA damage or from fusion with neighbouring chromosomes.

The ends of chromosomes— the telomeres—possess several unique features, one of which is the presence of many copies of a short repeated sequence.

In the protozoan Tetrahymena, this telomeric repeat is TTGGGG, with this G-rich strand typically protruding beyond the C-rich strand: toward ← 5′–TTGGGGTTGGGG–3′ → end of centromere 3′–AACCCC–5′ chromosome

Telomeres need to be protected from a cell's DNA repair systems because they have single-stranded overhangs, which "look like" damaged DNA. The overhang at the lagging strand end of the chromosome is due to incomplete end replication.

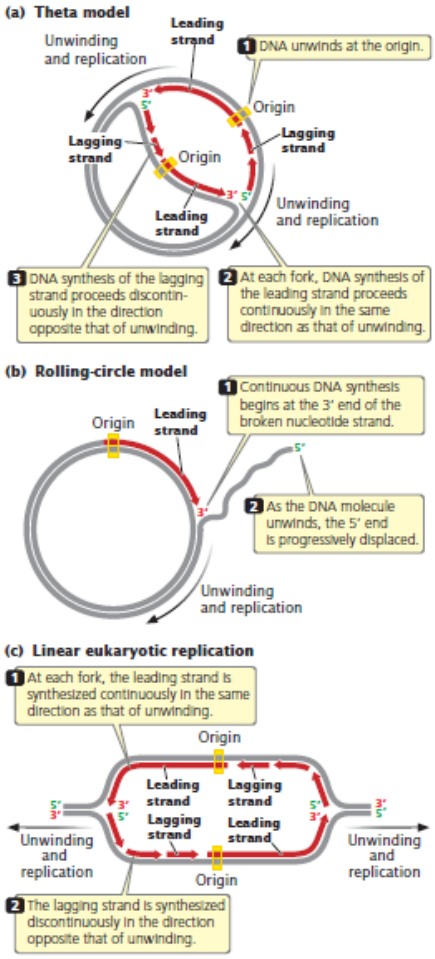
A telomere is a region of repetitive sequences at each end of the chromosomes of most eukaryotes.

In some species (including humans), the single-stranded overhangs bind to complementary repeats in the nearby double-stranded DNA, causing the telomere ends to form protective loops ^33 cubed.

Proteins associated with the telomere ends also help protect them and prevent them from triggering DNA repair pathways.

The single-stranded protruding end of the telomere, known as the G overhang can be extended by telomerase, an enzyme with both a protein and an RNA component (also known as a ribonucleoprotein).

MODELS OF DNA REPLICATION



The Direction of Replication in Different Models of Replication

In the theta model, the DNA unwinds at one particular location, the origin, and a replication bubble is formed. If the bubble has two forks, one at each end, synthesis takes place simultaneously at both forks (bidirectional replication). At each fork, synthesis on one of the template strands proceeds in the same direction as that of unwinding; this newly replicated strand is the leading strand with continuous replication. On the other template strand, synthesis proceeds in the direction opposite that of unwinding; this newly synthesized strand is the lagging strand with discontinuous replication. Focus on just one of the template strands within the bubble. Notice that synthesis on this template strand is continuous at one fork but discontinuous at the other. This difference arises because DNA synthesis is always in the same direction (5′→ 3′), but the two forks are moving in opposite directions. Replication in the rolling-circle model (**Figure 12.10b**) is somewhat different, because there is no replication bubble. Replication begins at the 3′ end of the broken nucleotide strand. Continuous replication takes place on the circular template as new
nucleotides are added to this 3′ end.

The replication of linear molecules of DNA, such as those found in eukaryotic cells, produces a series of replication bubbles. DNA synthesis in these bubbles is the same as that in the single replication bubble of the theta model; it begins at
the center of each replication bubble and proceeds at two forks, one at each end of the bubble. At both forks, synthesis of the leading strand proceeds in the same direction as that of unwinding, whereas synthesis of the lagging strand proceeds in the direction opposite that of unwinding.

MUTATION AND REPAIR OF DNA

Most biological molecules have a limited lifetime. Many proteins, lipids and RNAs are degraded when they are no longer needed or damaged, and smaller molecules such as sugars are metabolized to compounds to make or store energy. In contrast, DNA is the most stable biological molecule known, befitting its role in storage of genetic information. The DNA is passed from one generation to another, and it is degraded only when cells die. However, it can change, i.e. it is mutable. **Mutations**, or changes in the nucleotide sequence, can result from errors during DNA

replication, from covalent changes in structure because of reaction with chemical or physical agents in the environment, or from transposition. Most of the sequence alterations are **repaired** in cells. Some of the major avenues for changing DNA sequences and repairing those mutations will be discussed in this chapter.

Sequence alteration in the genomic DNA is the fuel driving the course of evolution. Without such mutations, no changes would occur in populations of species to allow them to adapt to changes in the environment. Mutations in the DNA of germline cells fall into three categories with respect to their impact on evolution. Most have no effect on phenotype; these include sequence changes in the large portion of the genome that neither codes for protein, or is involved in gene regulation or any other process. Some of these **neutral** mutations will become prevalent in a population of organisms (or **fixed**) over long periods of time by stochastic processes. Other mutations do have a phenotype, one that is advantageous to the individuals carrying it. These mutations are fixed in populations rapidly (i.e. they are subject to **positive selection**). Other mutations have a detrimental phenotype, and these are cleared from the population quickly. They are subject to **negative** or **purifying selection**.

Whether a mutation is neutral, disadvantageous or useful is determined by where it is in the genome, what the type of change is, and the particulars of the environmental forces operating on the locus. For our purposes, it is important to realize that sequence changes are a natural part of DNA metabolism. However, the amount and types of mutations that accumulate in a genome are determined by the types and concentrations of mutagens to which a cell or organism is exposed, the efficiency of relevant repair processes, and the effect on phenotype in the organism.

Mutations and mutagens

Types of mutations

Mutations commonly are **substitutions**, in which a single nucleotide is changed into a different nucleotide. Other mutations result in the loss (**deletion**) or addition (**insertion**) of one or more nucleotides. These insertions or deletions can range from one to tens of thousands of nucleotides. Often an insertion or deletion is inferred from comparison of two homologous

sequences, and it may be impossible to ascertain from the data given whether the presence of a segment in one sequence but not another resulted from an insertion of a deletion. In this case, it can be referred to as an **indel**. One mechanism for large insertions is the **transposition** of a sequence from one place in a genome to another.

Nucleotide substitutions are one of two classes. In a **transition**, a purine nucleotide is replaced with a purine nucleotide, or a pyrimidine nucleotide is replaced with a pyrimidine nucleotide. In other words, the base in the new nucleotide is in the same chemical class as that of the original nucleotide. In a **transversion**, the chemical class of the base changes, i.e. a purine nucleotide is replaced with a pyrimidine nucleotide, or a pyrimidine nucleotide is replaced with a purine nucleotide.

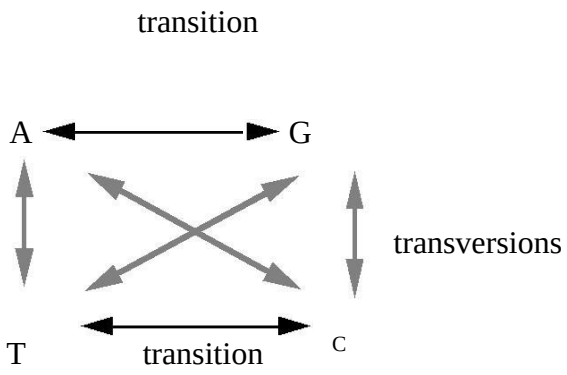


Figure 1. Diagram of the types of substitutions: transitions and transversions.

Comparison of the sequences of homologous genes between species reveals a pronounced preference for transitions over transversions (about 10-fold), indicating that transitions occur much more frequently than transversions.

Errors in Replication

Despite effective proofreading functions in many DNA polymerases, occasionally the wrong nucleotide is incorporated. It is estimated that *E. coli* DNA polymerase III holoenzyme (with a fully functional proofreading activity) uses the wrong nucleotide during elongation about 1 in 10^8

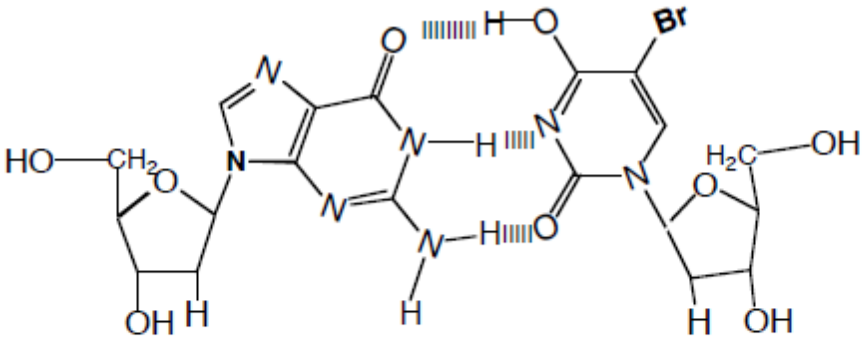
times. It is more likely for an incorrect pyrimidine nucleotide to be incorporated opposite a purine nucleotide in the template strand, and for a purine nucleotide to be incorporated opposite a pyrimidine nucleotide. Thus these misincorporations resulting in a transition substitution are more common. However, incorporation of a pyrimidine nucleotide opposite another pyrimidine nucleotide, or a purine nucleotide opposite another purine nucleotide, can occur, albeit at progressively lower frequencies. These rarer misincorporations lead to transversions.

Question 1. If a dCTP is incorporated into a growing DNA strand opposite an A in the template strand, what mutation will result? Is it a transition or a transversion?

Question 2. If a dCTP is incorporated into a growing DNA strand opposite a T in the template strand, what mutation will result? Is it a transition or a transversion?

A change in the isomeric form of a purine or pyrimidine base in a nucleotide can result in a mutation. The base-pairing rules are based on the hydrogen-bonding capacity of nucleotides with their bases in the *keto* tautomer. A nucleotide whose base is in the *enol* tautomer can pair with the "wrong" base in another nucleotide. For example, a T in the rare *enol* isomer will pair with a *keto* G (Fig. 2), and an *enol* G will pair with a *keto* T.

Working with Molecular Genetics



dG

5-Br-dU

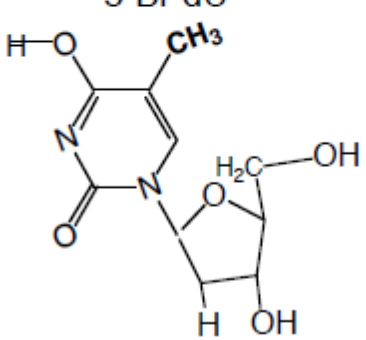


Figure 2. Illustration of the nucleoside*enol* 5-bromodeoxyuridine (or 5-BrdU, an analog of thymidine) paired with the nucleoside *keto* deoxyguanidine. 5-BrdU shifts into the *enol* tautomer more readily than thymidine does.

The *enol* tautomers of the normal deoxynucleotides guanidylate and thymidylate are rare, meaning that a single molecule is in the *keto* form most of the time, or within a population of molecules, most of them are in the *keto* form. However, certain nucleoside and base analogs adopt these alternative isomers more readily. For instance 5-bromo-deoxyuridine (or 5-BrdU) is an analog of deoxythymidine (dT) that is in the *enol* tautomer more frequently than dT is (although most of the time it is in the *keto* tautomer).

Thus the frequency of misincorporation can be increased by growth in the presence of base and nucleoside analogs. For example, growth in the presence of 5-BrdU results in an increase in the incorporation of G opposite a T in the DNA, as illustrated in Fig. 3. After cells take up the nucleoside 5-BrdU, it is converted to 5-BrdUTP by nucleotide salvage enzymes that add phosphates to its 5' end. During replication, 5-BrdUTP (in the *keto* tautomer) will incorporate opposite an A in DNA. The 5-BrdU can shift into the *enol* form while in DNA, so that when it serves as a template during the next round of replication (arrow 1 in the diagram below), it will direct incorporation of a G in the complementary strand. This G will in turn direct incorporation of a C in the top strand in the next round of replication (arrow 2). This leaves a C:G base pair where there was a T:A base pair in the parental DNA. Once the pyrimidine shifts back to the

favored *keto* tautomer, it can direct incorporation of an A, to give the second product in the diagram below (with a BrU-A base pair).

Question 3. Where are the hydrogen bonds in a base pair between *enol* –guanidine and *keto*-thymidine in DNA?

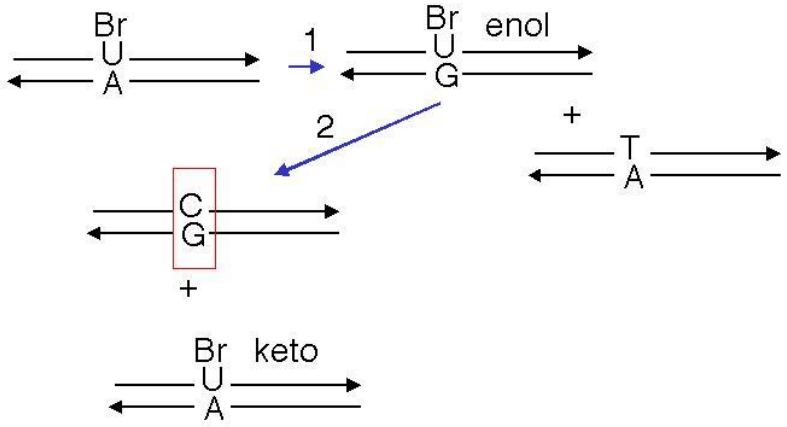


Figure 3. Replication of a misincorporated nucleotide (or nucleotide analog) will leave a mutation.

Likewise, misincorporation of A and C can occur when they are in the rare *imino* tautomers rather than the favored *amino* tautomers. In particular, *imino* C will pair with *amino* A, and *imino* A will pair with *amino* C (Fig. 4).

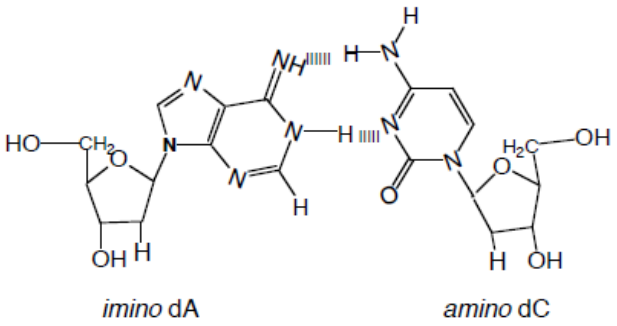


Figure 4. An A in the rare *imino* tautomer will pair with *amino* C. This can cause an A:T to G:C transition.

Misincorporation during replication is the major pathway for introducing *transversions* into DNA. Normally, DNA is a series of purine:pyrimidine base pairs, but in order to have a transversion, a pyrimidine has to be paired with another pyrimidine, or a purine with a purine. The DNA has to undergo local structural changes to accommodate these unusual base pairs. One way this can happen for a purine-purine base pair is for one of the purine nucleotides to shift from the preferred *anti* conformation to the *syn* conformation. Atoms on the "back side" of the purine nucleotide in the *syn*-isomer can form hydrogen bonds with atoms in the rare tautomer of the purine nucleotide, still in the preferred *anti* conformation. For example, an A nucleotide in the *syn*-, *amino*- isomer can pair with an A nucleotide in the *anti*-, *imino*- form (Fig. 5). Thus the transversion required a shift in the tautomeric form of the base in one nucleotide as well as a change in the base-sugar conformation (*anti* to *syn*) of the other nucleotide.

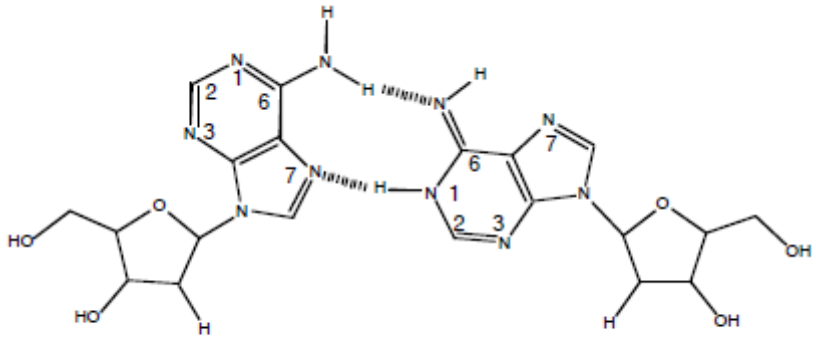


Figure 5. A base pair between a *syn*-, *amino*- isomer of A and the *anti*-, *imino*- form of A.

Question 4. Why does the shift of a purine nucleotide from *anti* to *syn* help allow a purine:purine base pair? Is this needed for a pyrimidine:pyrimidine base pair?

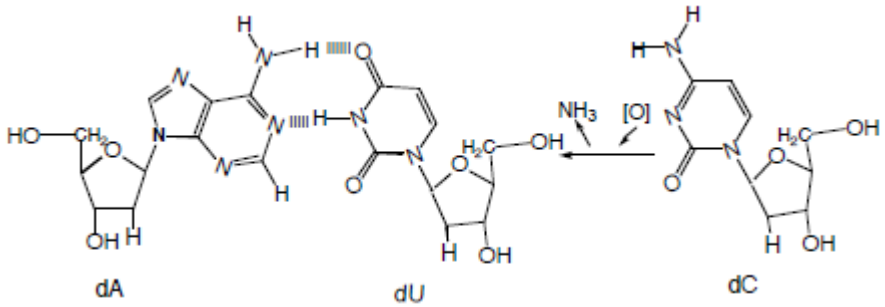
Errors in replication are not limited to substitutions. **Slippage errors** during replication will add or delete nucleotides. A DNA polymerase can insert additional nucleotides, more commonly when tandem short repeats are the template (e.g. repeating CA dinucleotides). Sometimes the template strand can loop out and form a secondary structure that the DNA polymerase does not read. In this case, a deletion in the nascent strand will result. The ability of intercalating agents to increase the frequency of such deletions is illustrated in Fig. 10.B. (see below).

Reaction with mutagens

Many mutations do not result from errors in replication. Chemical reagents can oxidize and alkylate the bases in DNA, sometimes changing their base-pairing properties. Radiation can also damage DNA. Examples of these mutagenic reactions will be discussed in this section.

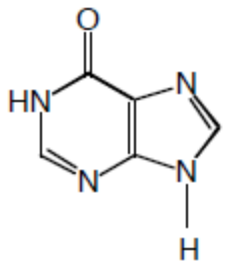
Chemical modification by oxidation

When the amino bases, adenine and cytosine, are oxidized, they also lose an amino group. Thus the amine is replaced by a keto group in the product of this oxidative deamination reaction. For instance, oxidation of cytosine produces uracil, which base



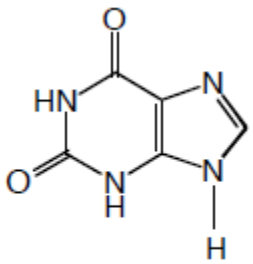
pairs with adenine (shown for deoxycytidine in Fig..6). Likewise, oxidation of adenine yields hypoxanthine, which base pairs with cytosine (Fig.7.A). Thus the products of these chemical reactions will be mutations in the DNA, if not repaired. Oxidation of guanine yields xanthine (Fig. 7.B). In DNA, xanthine will pair with cytosine, as does the original guanine, so this particular alteration is not mutagenic.

Figure 6. Oxidative deamination of deoxycytidine yields deoxyuridine. The deoxyuridine in DNA would direct pairing with dA after replication.



hypoxanthine

Figure 7.A. Structure of hyp oxanthine, the product of oxidation deamination of adenine.



xanthine

Figure 7.B. Structure of xanthine, the product of oxidative deamination of guanine.

Question 5. Both hypoxanthine and xanthine can base pair with cytosine in DNA. Why is this?

Oxidation of C to U occurs spontaneously at a high rate. The frequency is such that 1 in 1000 Cs in the human genome would become Us during a lifetime, if they were not repaired. As will be discussed later, repair mechanisms have evolved to replace a U in DNA with a T.

Methylation of C prior to its oxidative deamination will effectively mask it from the repair processes to remove U's from DNA. This has a substantial impact on the genomes of organisms that methylate C. In many eukaryotes, including vertebrates and plants (but not yeast or *Drosophila*), the principal DNA methyl transferase recognizes the dinucleotide CpG in DNA as the substrate, forming 5-methyl-CpG (Fig. 8). When 5-methyl cytosine undergoes oxidative deamination, the result is 5-methyl uracil, which is the same as thymine. The surveillance system that recognizes U's in DNA does nothing to the T, since it is a normal component of DNA. Hence the oxidation of 5-methyl CpG to TpG, followed by a round of replication, results in a C:G to T:A transition at former CpG sites (Fig. 7.8). This spontaneous deamination is quite frequent; indeed, C to T transitions at CpG dinucleotides are the most common mutations in humans. Since this transition is not repaired, over time the number of CpG dinucleotides is greatly diminished in the genomes of vertebrates and plants.

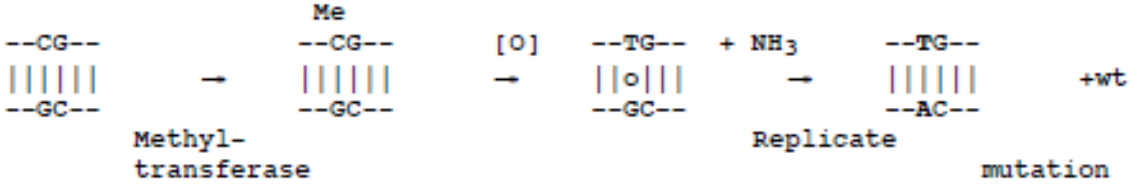


Figure 8. Methylation of CpG dinucleotides followed by oxidative deamination results in TpG dinucleotides.

Some regions of plant and vertebrate genomes do not show the usual depletion of CpG dinucleotides. Instead, the frequency of CpG approaches that of GpC or the frequency expected from the individual frequency of G and C in the genome. One working definition of these **CpG islands** is that they are segments of genomic DNA at least 100 bp long with a CpG to GpC ratio of at least 0.6. These islands can be even longer and have a CpG/GpC > 0.75. They are distinctive regions of these genomes and are often found in promoters and other regulatory regions of genes. Examination of several of these CpG islands has shown that they are not methylated in any tissue, unlike most of the other CpGs in the genome. Current areas of research include investigating how the CpG islands escape methylation and their role in regulation of gene expression.

Question 6. If a CpG island were to be methylated in the germ line, what would be consequences be over many generations?

The rate of oxidation of bases in DNA can be increased by treating with appropriate reagents, such as nitrous acid (HNO_2). Thus treatment with nitrous acid will increase the oxidation of C to U, and hence lead to C:G to T:A transitions in DNA. It will also increase the oxidation of adenine to hypoxanthine, leading to A:T to G:C transitions in DNA.

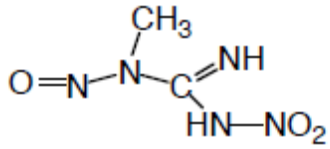
Chemical modification by alkylation

Many mutagens are **alkylating agents**. This means that they will add an alkyl group, such as methyl or ethyl, to a base in DNA. Examples of commonly used alkylating agents in laboratory work are N-methyl-nitrosoguanidine and N-methyl-N'-nitro-nitrosoguanidine (MNNG, Fig.

9.A.). The chemical warfare agents sulfur mustard and nitrogen mustard are also alkylating agents.

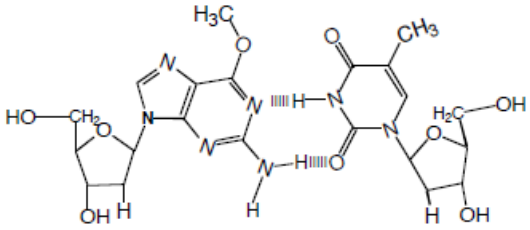
N-methyl-nitrosoguanidine and MNNG transfer a methyl group to guanine (e.g. to the O^6 position) and other bases (e.g. forming 3-methyladenine from adenine). The additional methyl (or other alkyl group) causes a distortion in the helix. The distorted helix can alter the base pairing properties. For instance, O^6-methylguanine will sometimes base pair with thymine (Fig. 9.B.).

A. N-methyl-N'-nitro-N-nitrosoguanidine (MNNG)



MNNG

B. 6-O-methyl-G will pair with T



6-*O*-methyl-dG dT

Figure 9.A. Structure of MNNG and the base pair between O^6-methyl G and T

The order of reactivity of nucleophilic centers in purines follows roughly this series:

$$N^7\text{-G} >> N^3\text{-A} > N^1\text{-A} \cong N^3\text{-G} \cong O^6\text{-G}.$$

A common laboratory reagent for purines in DNA is dimethylsulfate, or DMS. The products of this reaction are primarily N^7 -guanine, but N^3-adenine is also detectable. This reaction is used to identify protein-binding sites in DNA, since interaction with a protein can cause decreased reactivity to DMS of guanines within the binding site but enhanced reactivity adjacent to the site.

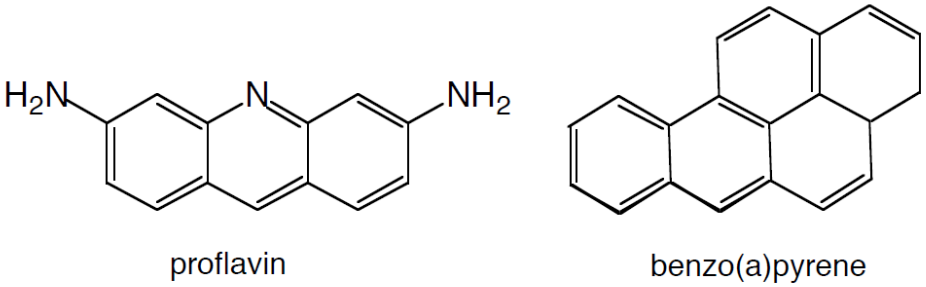
Methylation to form N^7-methyl-guanine does not cause miscoding in the DNA, since this modified purine still pairs with C.

Chemicals that cause deletions

Some compounds cause a loss of nucleotides from DNA. If these deletions occur in a protein-coding region of the genomic DNA, and are not an integral multiple of 3, they result in a frameshift mutation. These are commonly more severe loss-of-function mutations than are simple substitutions. Frameshift mutagens such as proflavin or ethidium bromide have flat, polycyclic ring structures (Fig. 10.A.). They may bind to and **intercalate** within the DNA, i.e. they can insert between stacked base pairs. If a segment of the template DNA is the looped out, DNA polymerase can replicate past it, thereby generating a deletion. Intercalating agents can stabilize secondary structures in the loop, thereby increasing the chance that this segment stays in the loop and is not

copied during replication (Fig. 10.B.) Thus growth of cells in the presence of such intercalating agents increase the probability of generating a deletion.

A.



B.

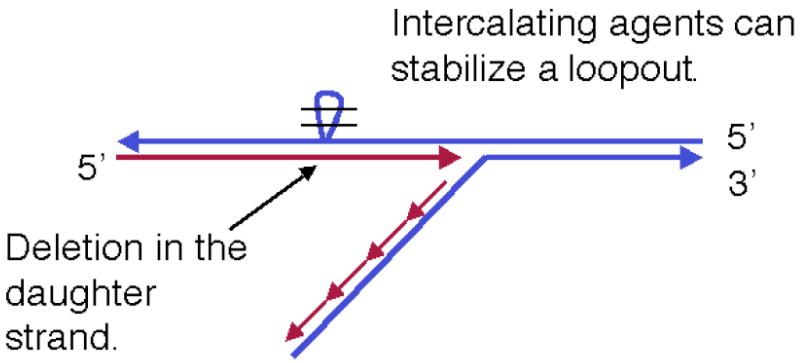


Figure 10. Two intercalating agents (A) and their ability to stabilize loops in the template, leading to deletions in the nascent DNA strand (B). Benz(a)pyrenes are present in soot.

Ionizing radiation

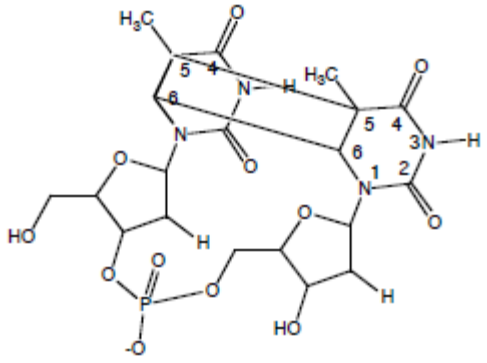
High energy radiation, such as X-rays, γ-rays, and β particles (or electrons) are powerful mutagens. Since they can change the number of electrons on an atom, converting a compound to an ionized form, they are referred to as **ionizing radiation**. They can cause a number of chemical changes in DNA, including directly break phosphodiester backbone of DNA, leading to deletions.

Ionizing radiation can also break open the imidazole ring of purines. Subsequent removal of the damaged purine from DNA by a glycosylase generates an apurinic site.

Ultraviolet radiation

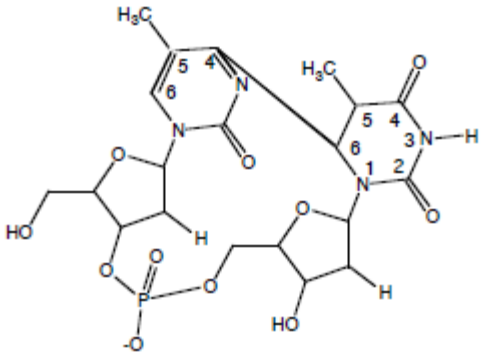
Ultraviolet radiation with a wavelength of 260 nm will form pyrimidine dimers between adjacent pyrimidines in the DNA. The dimers can be one of two types (Fig. 11). The major product is a cytobutane-containing thymine dimer (between C5 and C6 of adjacent T's). The other product has a covalent bond between position 6 on one pyrimidine and position 4 on the adjacent pyrimidine, hence it is called the "6-4" photoproduct.

A.



cyclobutane thymine dimers in DNA

B.



6-4 photoproducts of thymine dimers in DNA

6-4 photoproducts of thymine dimers in DNA

Figure 11. Pyrimidine dimers formed by UV radiation, illustrated for adjacent thymidylates on one strand of the DNA. (A) Formation of a covalent bond between the C atoms at position 5 of each pyrimidine and between the C atoms at position 6 of each pyrimidine makes a cyclobutane ring connecting the two pyrimidines. The bases are stacked over each other, held in place by the cyclobutane ring. The C-C bonds between the pyrimidines are exaggerated in this drawing so that the pyrimidine ring is visible. (B) Another photoproduct is made by forming a bond between the C atom at position 6 of one pyrimidine and position 4 of the adjacent pyrimidine, with loss of the O previously attached at position 4.

The pyrimidine dimers cause a distortion in the DNA double helix. This distortion blocks replication and transcription.

Question 7. What is the physical basis for this distortion in the DNA double helix?

Summary: Causes of transitions and transversions

Table 1 lists several causes of mutations in DNA, including mutagens as well as mutator strains in bacteria. Note that some of these mutations lead to mispairing (substitutions), others lead to distortions of the helix, and some lead to both.

Transitions can be generated both by damage to the DNA and by misincorporation during replication. Transversions occur primarily by misincorporation during replication. The frequency of such errors is greatly increased in mutator strains, e.g. lacking a proofreading function in the replicative DNA polymerase. Also, after a bacterial cell has sustained sufficient damage to induce the SOS response, the DNA polymerase shifts into a an error-prone mode of replication. This can also be a source of mutant alleles.

Table. 1. Summary of effects of various agents that alter DNA sequences (mutagens and mutator genes)

Agent (mutagen, etc.)	Example	Result
Nucleotide analogs	BrdUTP	transitions, e.g. A:T to G:C
Oxidizing agents	nitrous acid	transitions, e.g. C:G to T:A
Alkylating agents	nitrosoguanidine	transitions, e.g. G:C to A:T
Frameshift mutagens	Benz(a)pyrene	deletions (short)
Ionizing radiation	X-rays, γ-rays	breaks and deletions (large)
UV	UV, 260 nm	Y-dimers, block replication
Misincorporation:		
Altered DNA Pol III	*mutD=dnaQ*; ε subunit	transitions, transversions and
	of DNA PolIII	frameshifts in mutant strains
Error-prone repair	Need UmuC, UmuD,	transitions and transversions in
	DNA PolIII	wild-type during SOS
Other mutator genes	*mutM, mutT, mutY*	transversions in the mutant
		strains

Repair mechanisms

The second part of this chapter examines the major classes of DNA repair processes. These are:

- reversal of damage,
- nucleotide excision repair,
- base excision
- repair, mismatch repair,
- recombinational repair, and
- error-prone repair.

Many of these processes were first studies in bacteria such as *E. coli*, however only a few are limited to this species. For instance, nucleotide excision repair and base excision repair are found in virtually all organisms, and they have been well characterized in bacteria, yeast, and mammals. Like DNA replication itself, repair of damage and misincorporation is a very old process.

Reversal of damage

Some kinds of covalent alteration to bases in DNA can be directly reversed. This occurs by specific enzyme systems recognizing the altered base and breaking bonds to remove the adduct or change the base back to its normal structure.

Photoreactivation is a light-dependent process used by bacteria to reverse pyrimidine dimers formed by UV radiation. The enzyme photolyase binds to a pyrimidine dimer and catalyzes a second photochemical reaction (this time using visible light) that breaks the cyclobutane ring and reforms the two adjacent thymidylates in DNA. Note that this is not formally the reverse of the reaction that formed the pyrimidine dimers, since energy from visible light is used to break the bonds between the pyrimidines, and no UV radiation is released. However, the result is that the DNA structure has been returned to its state prior to damage by UV. The photolyase enzyme has two subunits, which are encoded by the *phrA* and *phrB* genes in *E. coli*.

A second example of the reversal of damage is the **removal of methyl groups**. For instance, the enzyme O^6-methylguanine methyltransferase, encoded by the *ada* gene in *E. coli*, recognizes O^6-methylguanine in duplex DNA. It then removes the methyl group, transferring it to an amino acid of the enzyme. The methylated enzyme is no longer active, hence this has been referred to as a suicide mechanism for the enzyme.

Excision repair

The most common means of repairing damage or a mismatch is to cut it out of the duplex DNA and recopy the remaining complementary strand of DNA, as outlined in Fig. 7.12. Three different types of excision repair have been characterized: nucleotide excision repair, base excision repair, and mismatch repair. All utilize a **cut, copy, and paste** mechanism. In the **cutting** stage, an enzyme or complex removes a damaged base or a string of nucleotides from the DNA. For the **copying**, a DNA polymerase (DNA polymerase I in *E. coli*) will copy the template to replace the excised, damaged strand. The DNA polymerase can initiate synthesis from 3' OH at the single-strand break (nick) or gap in the DNA remaining at the site of damage after excision. Finally, in the **pasting** stage, DNA ligase seals the remaining nick to give an intact, repaired DNA.

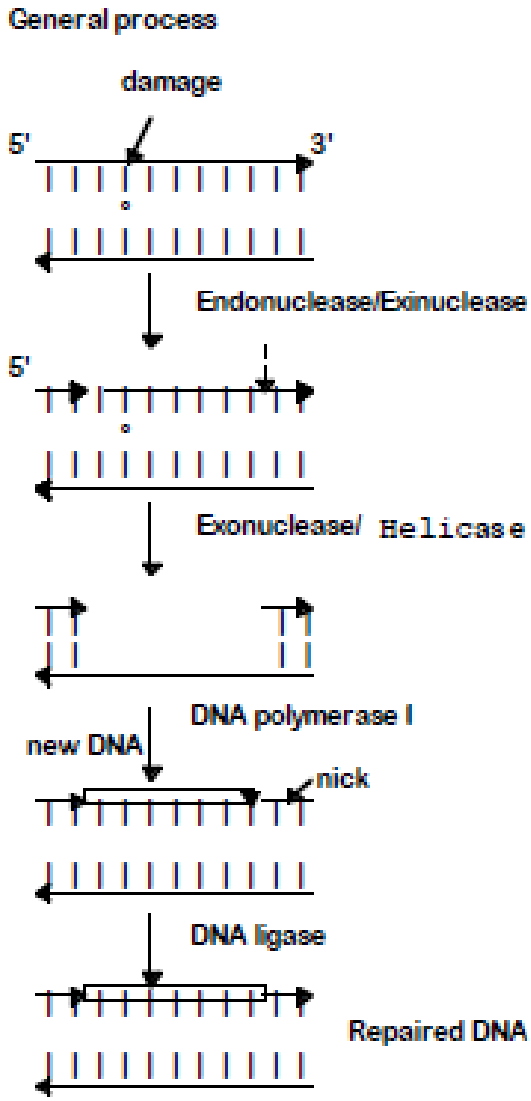


Figure 12. A general scheme for excision repair, illustrating the cut (steps 1 and 2), copy (step 3) and paste (step 4) mechanism.

Nucleotide excision repair

In **nucleotide excision repair** (**NER**), damaged bases are cut out within a string of nucleotides, and replaced with DNA as directed by the undamaged template strand. This repair

system is used to remove pyrimidine dimers formed by UV radiation as well as nucleotides modified by bulky chemical adducts. The common feature of damage that is repaired by nucleotide excision is that the modified nucleotides cause a significant distortion in the DNA helix. NER occurs in almost all organisms examined.

Some of the best-characterized enzymes catalyzing this process are the UvrABC excinuclease and the UvrD helicase in *E. coli*. The genes encoding this repair function were discovered as mutants that are highly sensitive to UV damage, indicating that the mutants are defective in UV repair.

The enzymes encoded by the *uvr* genes have been studied in detail. The polypeptide products of the *uvrA*, *uvrB*, and *uvrC* genes are subunits of a multisubunit enzyme called the **UvrABC excinuclease**. UvrA is the protein encoded by *uvrA*, UvrB is encoded by *uvrB*, and so on. The UvrABC complex recognizes damage-induced structural distortions in the DNA, such as pyrimidine dimers. It then cleaves on both sides of the damage. Then UvrD (also called helicase II), the product of the *uvrD* gene, unwinds the DNA, releasing the damaged segment. Thus for this system, the UvrABC and UvrD proteins carry out a series of steps in the cutting phase of excision repair. This leaves a gapped substrate for copying by DNA polymerase and pasting by DNA ligase.

The UvrABC proteins form a dynamic complex that recognizes damage and makes endonucleolytic cuts on both sides. The two cuts around the damage allow the single-stranded segment containing the damage to be excised by the helicase activity of UvrD. Thus the UvrABC dynamic complex and the UvrBC complex can be called **excinucleases**. After the damaged segment has been excised, a gap of 12 to 13 nucleotides remains in the DNA. This can be filled in by DNA polymerase and the remaining nick sealed by DNA ligase. Since the undamaged template directs the synthesis by DNA polymerase, the resulting duplex DNA is no longer damaged.

In more detail, the process goes as follows (Fig. 14). $UvrA_2$ (a dimer) and Uvr B recognize the damaged site as a $(UvrA)_2UvrB$ complex. $UvrA_2$ then dissociates, in a step that requires ATP hydrolysis. This is an autocatalytic reaction, since it is catalyzed by UvrA, which is itself an

ATPase. After UvrA has dissociated, UvrB (at the damaged site) forms a complex with UvrC. The UvrBC complex is the active nuclease. It makes the incisions on each side of the damage, in another step that requires ATP. The phosphodiester backbone is cleaved 8 nucleotides to the 5' side of the damage and 4-5 nucleotides on the 3' side. Finally, the UvrD helicase then unwinds DNA so the damaged segment is removed. The damaged DNA segment dissociates attached to the UvrBC complex. Like all helicase reactions, the unwinding requires ATP hydrolysis to disrupt the base pairs. Thus ATP hydrolysis is required at three steps of this series of reactions.

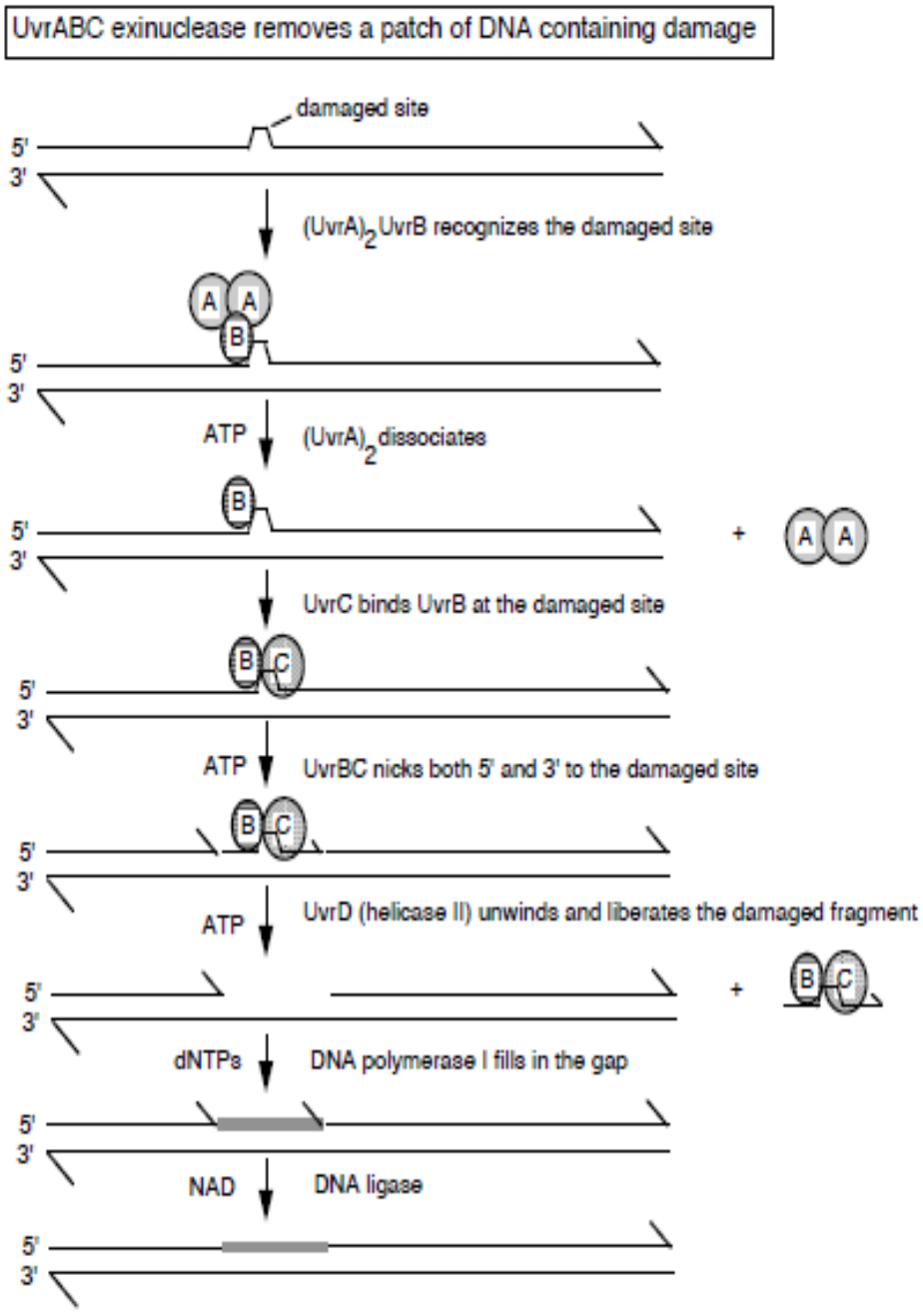


Figure 14. The Uvr(A)BC excinuclease of *E. coli* recognizes AP sites, thymine dimers, and other structural distortions and makes nicks on both sides of the damaged region. The 12-13 nucleotide-long fragment is released together with the excinuclease by helicase II action.

Question 8. How does an excinuclease differ from an exonuclease and an endonuclease?

NER occurs in two modes in many organisms, including bacteria, yeast and mammals. One is the global repair that acts throughout the genome, and the second is a specialized activity is that is coupled to transcription. Most of the XP gene products listed in Table 2 function in both modes of NER in mammalian cells. However, XPC (acting in a complex with another protein called hHR23B) is a DNA-damage sensor that is specific for global genome NER. In transcription coupled NER, the elongating RNA polymerase stalls at a lesion on the template strand; perhaps this is the damage recognition activity for this mode of NER. One of the basal transcription factors that associates with RNA polymerase II, TFIIH, also plays a role in both types of NER. A rare genetic disorder in humans, Cockayne syndrome (CS), is associated with a defect specific to transcription coupled repair. Two complementation groups have been identified, *CSA* and *CSB*. Determination of the nature and activity of the proteins encoded by them will provide additional insight into the efficient repair of transcribed DNA strands. The phenotype of CS patients is pleiotropic, showing both photosensitivity and severe neurological and other developmental disorders, including premature aging. These symptoms are more severe than those seen for XP patients with no detectable NER, indicating that transcription-coupled repair or the CS proteins have functions in addition to those for NER.

Ataxia telangiectasia, or AT, illustrates the effect of alterations in a protein not directly involved in repair, but perhaps signaling that is necessary for proper repair of DNA. AT is a recessive, rare genetic disease marked by uneven gait (ataxia), dilation of blood vessels (telangiectasia) in the eyes and face, cerebellar degeneration, progressive mental retardation, immune deficiencies, premature aging and about a 100-fold increase in susceptibility to cancers. That latter phenotype is driving much of the interest in this locus, since heterozygotes, which comprise about 1% of the population, also have an increased risk of cancer, and may account for as much as 9% of breast cancers in the United States. The gene that is mutated in AT (hence called "ATM") was isolated in 1995 and localized to chromosome 11q22-23.

The ATM gene does not appear to encode a protein that participates directly in DNA repair (unlike the genes that cause XP upon mutation). Rather, AT is caused by a defect in a cellular signaling pathway. Based on homologies to other proteins, the ATM gene product may be involved in the regulation of telomere length and cell cycle progression. The C-terminal domain is homologous to phosphatidylinositol-3-kinase (which is also a Ser/Thr protein kinase) - hence the connection to signaling pathways. The ATM protein also has regions of homology to DNA-dependent protein kinases, which require breaks, nicks or gaps to bind DNA (via subunit Ku); binding to DNA is required for the protein kinase activity. This suggests that ATM protein could be involved in targeting the repair machinery to such damage.

Base excision repair

Base excision repair differs from nucleotide excision repair in the types substrates recognized and in the initial cleavage event. Unlike NER, the base excision machinery recognizes damaged bases that do not cause a significant distortion to the DNA helix, such as the products of oxidizing agents. For example, base excision can remove uridines from DNA, even though a G:U base pair does not distort the DNA. Base excision repair is versatile, and this process also can remove some damaged bases that do distort the DNA, such as methylated purines. In general, the initial recognition is a specific damaged base, not a helical distortion in the DNA. A second major difference is that the initial cleavage is directed at the glycosidic bond connecting the purine or pyrimidine base to a deoxyribose in DNA. This contrasts with the initial cleavage of a phosphodiester bond in NER.

Cells contain a large number of specific **glycosylases** that recognize damaged or inappropriate bases, such as uracil, from the DNA. The glycosylase removes the damaged or inappropriate base by catalyzing cleavage of the N-glycosidic bond that attaches the base to the sugar-phosphate backbone. For instance, uracil-N-glycosylase, the product of the *ung* gene, recognizes uracil in DNA and cuts the N-glycosidic bond between the base and deoxyribose (Fig. 7.15). Other glycosylases recognize and cleave damaged bases. For instance methylpurine glycosylase removes methylated G and A from DNA. The result of the activity of these

glycosylases is an apurinic/apyrimidinic site, or AP site (Fig. 15). At an AP site, the DNA is still an intact duplex, i.e. there are no breaks in the phosphodiester backbone, but one base is one.

Next, an **AP endonuclease** nicks the DNA just 5’ to the AP site, thereby providing a primer for DNA polymerase. In *E. coli*, the 5' to 3' exonuclease function of DNA polymerase I removes the damaged region, and fills in with correct DNA (using the 5' to 3' polymerase, directed by the sequence of the undamaged complementary strand).

Additional mechanisms have evolved for keeping U’s out of DNA. *E. coli* also has a dUTPase, encoded by the *dut* gene, which catalyzes the hydrolysis of dUTP to dUMP. The product dUMP is the substrate for thymidylate synthetase, which catalyzes conversion of dUMP to dTMP. This keeps the concentration of dUTP in the cell low, reducing the chance that it will be used in DNA synthesis. Thus the combined action of the products of the *dut* + *ung* genes helps prevent the accumulation of U's in DNA.

Question 9. In base excision repair, which enzymes are specific for a particular kind of damage and which are used for all repair by this pathway?

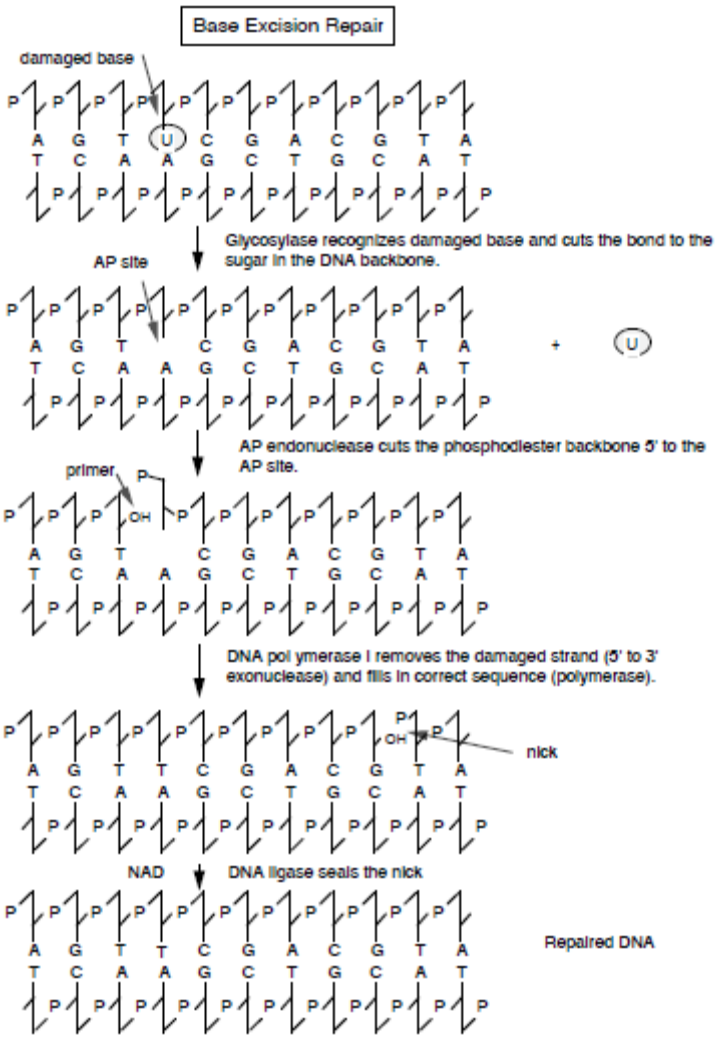


Figure 15. Base excision repair is initiated by a glycosylase that recognizes and removes chemically damaged or inappropriate bases in DNA. The glycosylase cleaves the glycosidic bond between the base and the sugar, leaving an apurinic/apyrimidinic site. The AP endonuclease can then nick the phosphodiester backbone 5’ to the AP site. When DNA polymerase I binds the free primer end at the nick, its 5'-3' exonuclease activity cuts a few nucleotides ahead of the missing base, and its polymerization activity fills the entire gap of several nucleotides.

Mismatch repair

The third type of excision repair we will consider is **mismatch repair**, which is used to repair errors that occur during DNA synthesis. Proofreading during replication is good but not perfect. Even with a functional ε subunit, DNA polymerase III allows the wrong nucleotide to be incorporated about once in every 10^8 bp synthesized in *E. coli*. However, the measured mutation rate in bacteria is as low as one mistake per 10^{10} or 10^{11} bp. The enzymes that catalyze **mismatch repair** are responsible for this final degree of accuracy. They recognize misincorporated nucleotides, excise them and replace them with the correct nucleotides. In contrast to nucleotide excision repair, mismatch repair does not operate on bulky adducts or major distortions to the DNA helix. Most of the mismatches are substitutes within a chemical class, e.g. a C incorporated instead of a T. This causes only a subtle helical distortions in the DNA, and the misincorporated nucleotide is a normal component of DNA. The ability of a cell to recognize a mismatch reflects the exquisite specificity of **MutS**, which can distinguish normal base pairs from those resulting from misincorporation. Of course, the repair machinery needs to know which of the nucleotides at a mismatch pair is the correct one and which was misincorporated. It does this by determining which strand was more recently synthesized, and repairing the mismatch on the nascent strand.

In *E. coli*, the methylation of A in a GATC motif provides a covalent marker for the parental strand, thus methylation of DNA is used to discriminate parental from progeny strands. Recall that the ***dam* methylase** catalyzes the transfer of a methyl group to the A of the pseudopalindromic sequence GATC in duplex DNA. Methylation is delayed for several minutes after replication. IN this interval before methylation of the new DNA strand, the mismatch repair system can find mismatches and direct its repair activity to nucleotides on the unmethylated, newly replicated strand. Thus replication errors are removed preferentially.

The enzyme complex MutH-MutL-MutS , or MutHLS, catalyzes mismatch repair in *E. coli*. The genes that encode these enzymes, *mutH*, *mutL* and *mutS*, were discovered because strains carrying mutations in them have a high frequency of new mutations. This is called a **mutator phenotype**, and hence the name *mut* was given to these genes. Not all mutator genes are involved in mismatch repair; e.g., mutations in the gene encoding the proofreading enzyme

of DNA polymerase III also have a mutator phenotype. This gene was independently discovered in screens for defects in DNA replication (*dnaQ*) and mutator genes (*mutD*). Three complementation groups within the set of mutator alleles have been implicated primarily in mismatch repair; these are *mutH*, *mutL* and *mutS*.

MutS will recognize seven of the eight possible mismatched base pairs (except for C:C) and bind at that site in the duplex DNA (Fig. 16). **MutH** and **MutL** (with ATP bound) then join the complex, which then moves along the DNA in either direction until it finds a hemimethylated GATC motif, which can be as far a few thousand base pairs away. Until this point, the nuclease function of MutH has been dormant, but it is activated in the presence of ATP at a hemimethylated GATC. It cleaves the unmethylated DNA strand, leaving a nick 5' to the G on the strand containing the unmethylated GATC (i.e. the new DNA strand). The same strand is nicked on the other side of the mismatch. Enzymes involved in other processes of repair and replication catalyze the remaining steps. The segment of single-stranded DNA containing the incorrect nucleotide is to be excised by UvrD, also known as helicase II and MutU. SSB and exonuclease I are also involved in the excision. As the excision process forms the gap, it is filled in by the concerted action of DNA polymerase III (Fig. 16.).

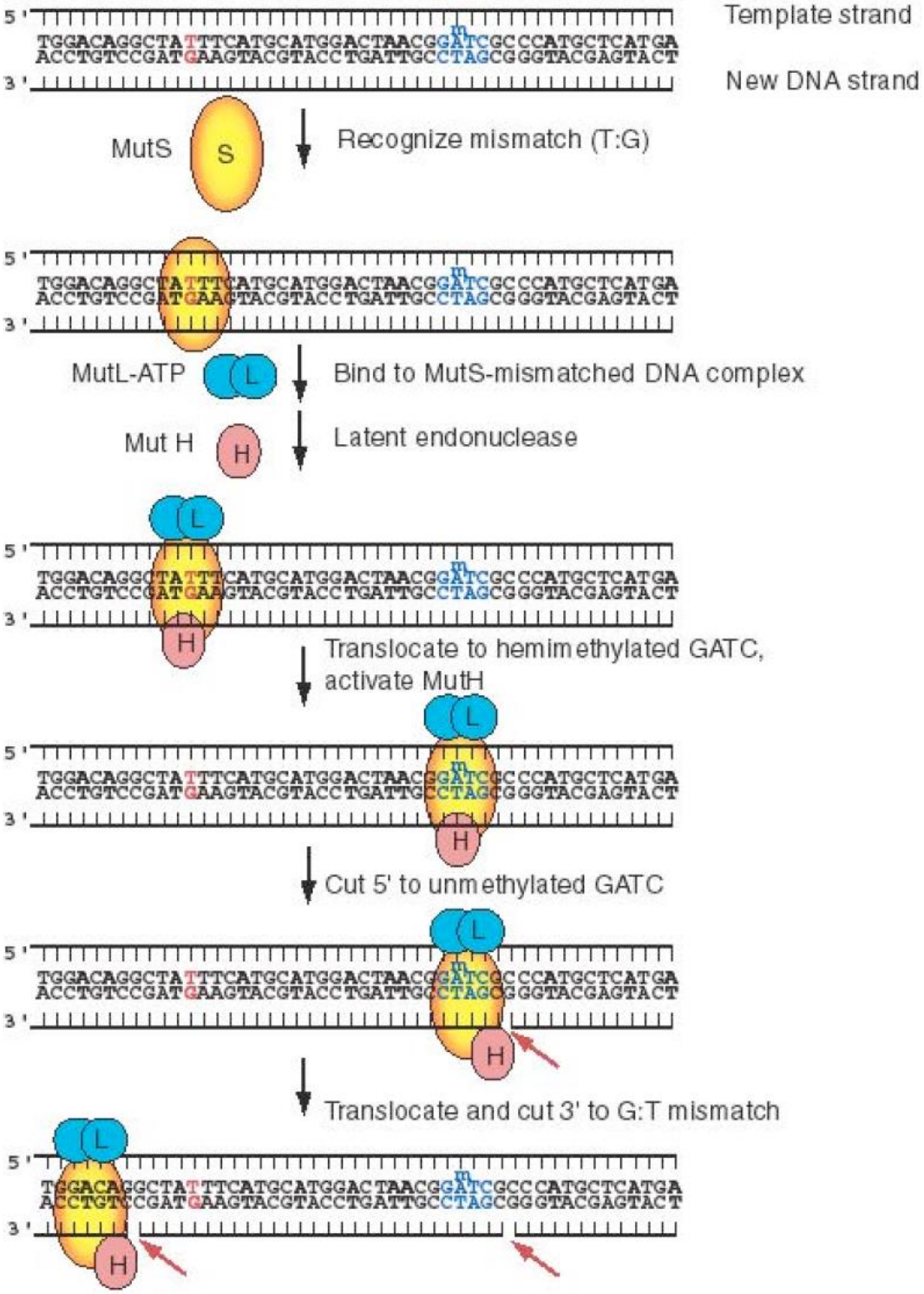

5'
TGGACAGGCTATTTCATGCATGGACTAACGGATCGCCCATGCTCATGA
ACCTGTCCGATGAAGTACGTACCTGATTGCCTAGCGGGTACGAGTACT
3'
Template strand
New DNA strand
MutS
S
Recognize mismatch (T:G)
MutL-ATP
L
Bind to MutS-mismatched DNA complex
Mut H
H
Latent endonuclease
Translocate to hemimethylated GATC,
activate MutH
Cut 5' to unmethylated GATC
Translocate and cut 3' to G:T mismatch

Figure 16 (part 1). Mismatch Repair by MutHLS: recognition of mismatch (shown in red), identifying the new DNA strand (using the hemimethylated GATC shown in blue) and cutting to encompass the unmethylated GATC and the misincorporated nucleotide (red G).

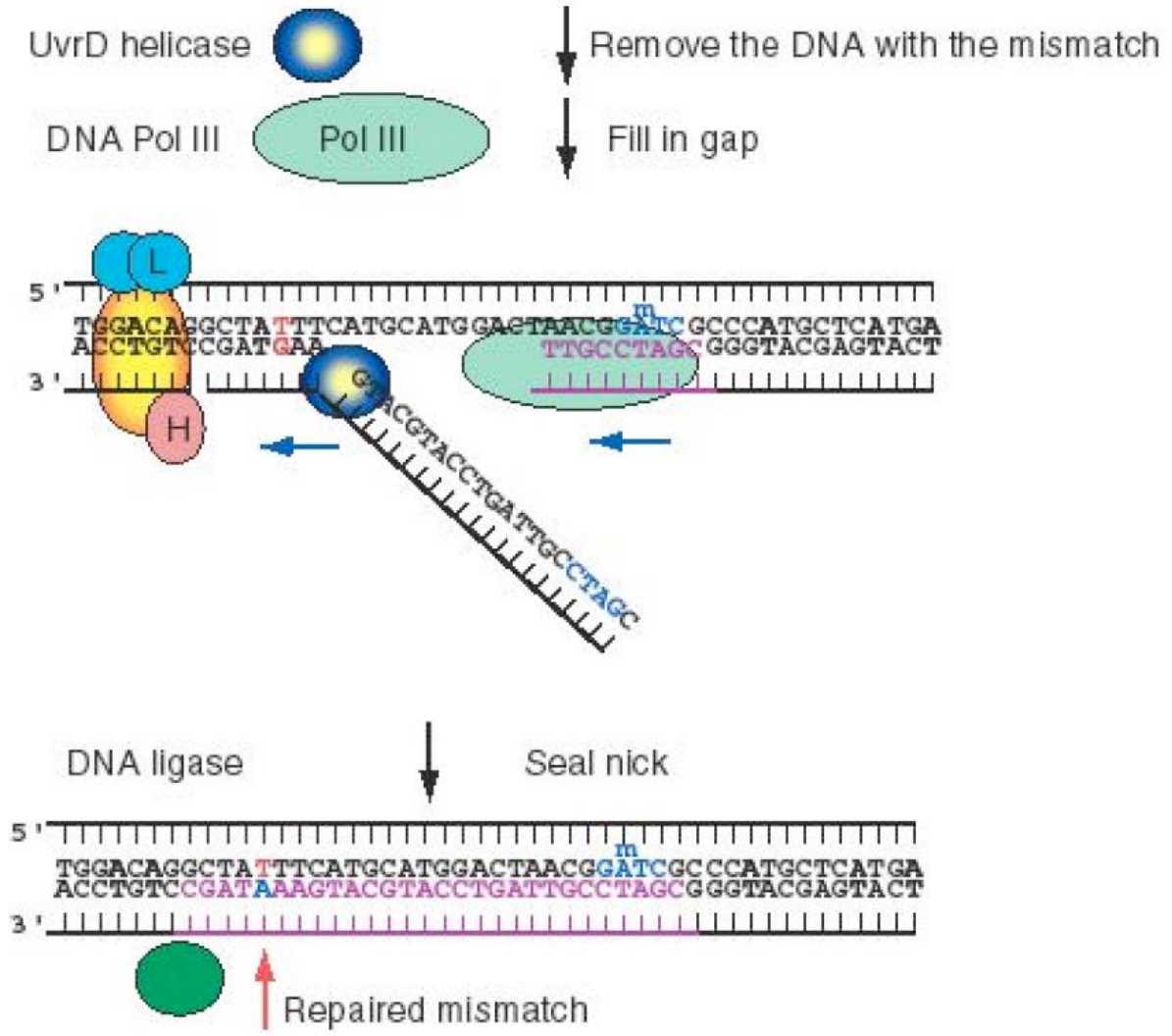


Figure 16 (part 2). Mismatch Repair: excision of the DNA with the misincorporated nucleotide bu Uvr D (aided by exonuclease I and SSB), gap filling by DNA polymerase III and ligation.

Mismatch repair is highly conserved, and investigation of this process in mice and humans is providing new clues about mutations that cause cancer. Homologs to the *E. coli* genes *mutL* and *mutS* have been identified in many other species, including mammals. The key breakthrough came from analysis of mutations that cause one of the most common hereditary cancers, *hereditary nonpolyposis colon cancer* (HNPCC). Some of the genes that, when mutated, cause this disease encode proteins whose amino acid sequences are significantly similar to those of two of the *E. coli* mismatch repair enzymes. The human genes are called *hMLH1* (for human *mutL* homolog 1), *hMSH1*, and *hMSH2* (for human *mutS* homolog 1 and 2, respectively). Subsequent work has shown that these enzymes in humans are involved in mismatch repair. Presumably the increased frequency of mutation in cells deficient in mismatch repair leads to the accumulation of mutations in proto-oncogenes, resulting in dysregulation of the cell cycle and loss of normal control over the rate of cell division.

Question 10. The human homologs to bacterial enzymes involved in mismatch repair are also implicated in homologous functions. Given the human homologs discussed above, which enzymatic functions found in bacterial mismatch repair are also found in humans?

What functions are missing, and hence are likely carried out by an enzyme not homologous to those used in bacterial mismatch repair?

Recombination repair (Retrieval system)

In the three types of excision repair, the damaged or misincorporated nucleotides are cut out of DNA, and the remaining strand of DNA is used for synthesis of the correct DNA sequence. However, this complementary strand is not always available. Sometimes DNA polymerase has to synthesize past a lesion, such as a pyrimidine dimer or an AP site. One way it can do this is to stop on one side of the lesion and then resume synthesis about 1000 nucleotides further down. This leaves a gap in the strand opposite the lesion (Fig. 17).

The information needed at the gap is retrieved from the normal daughter molecule by bringing in a single strand of DNA, using RecA-mediated recombination. This fills the gap opposite

the dimer, and the dimer can now be replaced by excision repair (Fig. 17). The resulting gap in the (previously) normal daughter can be filled in by DNA polymerase, using the good template.

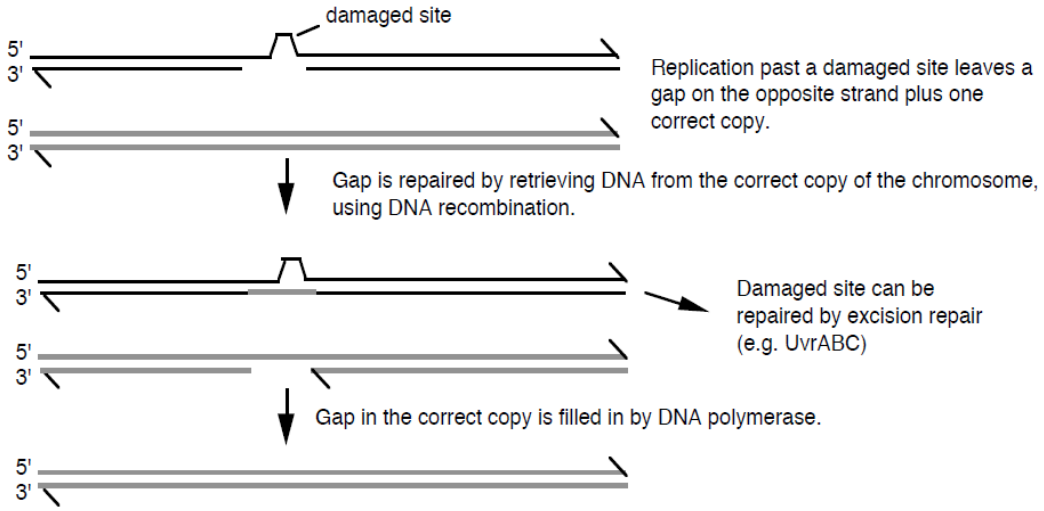


Figure 17. Recombination repair, a system for retrieval of information

Translesion synthesis

As just described, DNA polymerase can skip past a lesion on the template strand, leaving behind a gap. It has another option when such a lesion is encountered, which is to synthesis DNA in a non-template directed manner. This is called **translesion synthesis**, bypass synthesis, or error-prone repair. This is the last resort for DNA repair, e.g. when repair has not occurred prior to replication. In translesion replication, the DNA polymerase shifts from template directed synthesis to catalyzing the incorporation of random nucleotides. These random nucleotides are usually mutations (i.e. in three out of four times), hence this process is also designated error-prone repair.

Translesion synthesis uses the products of the *umuC* and *umuD* genes. These genes are named for the UV nonmutable phenotype of mutants defective in these genes.

Question 11. Why do mutations in genes required for translesion synthesis (error prone repair) lead to a *non*mutable phenotype?

UmuD forms a homodimer that also complexes with UmuC. When the concentration of single-stranded DNA and RecA are increased (by DNA damage, see next section), RecA stimulates an autoprotease activity in $UmuD_2$ to form $UmuD'_2$. This cleaved form is now active in translesional synthesis. UmuC itself is a DNA polymerase. A multisubunit complex containing UmuC, the activated $UmuD'_2$ and the α subunit of DNA polymerase III catalyze translesional synthesis. Homologs of the UmuC polymerase are found in yeast (RAD30) and humans (XP-V).

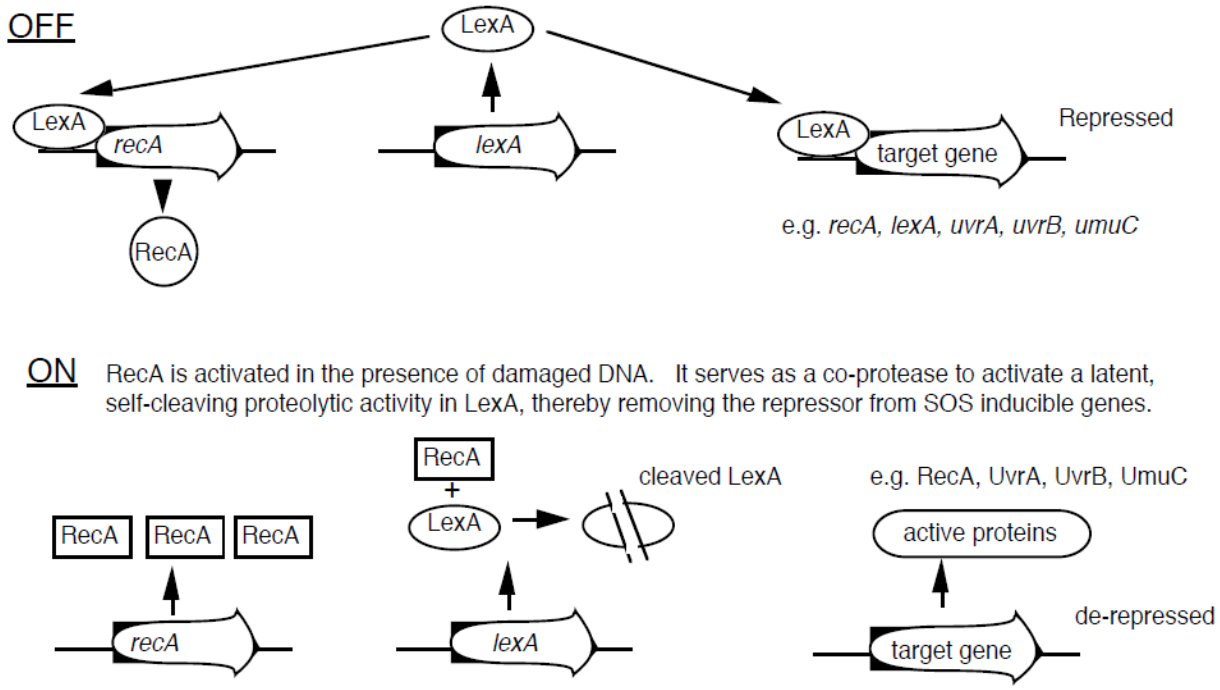
The SOS response

A coordinated battery of responses to DNA damage in *E. coli* is referred to as the SOS response. This name is derived from the maritime distress call, “SOS” for "Save Our Ship".

Accumulating damage to DNA, e.g. from high doses of radiation that break the DNA backbone, will generate single-stranded regions in DNA. The increasing amounts of single-stranded DNA induce SOS functions, which stimulate both the recombination repair and the translesional synthesis just discussed.

Key proteins in the SOS response are **RecA** and **LexA**. RecA binds to single stranded regions in DNA, which activates new functions in the protein. One of these is a capacity to further activate a latent proteolytic activity found in several proteins, including the LexA repressor, the **UmuD** protein and the repressor encoded by bacteriophage lambda (Fig. 18). RecA activated by binding to single-stranded DNA is not itself a protease, but rather it serves as a co-protease, activating the latent proteolytic function in LexA, UmuD and some other proteins.

In the absence of appreciable DNA damage, the LexA protein represses many operons, including several genes needed for DNA repair: *recA, lexA, uvrA, uvrB, and umuC.* When the activated RecA stimulates its proteolytic activity, it cleaves itself (and other proteins), leading to coordinate induction of the SOS regulated operons (Fig. 18).



Restriction/Modification systems

The DNA repair systems discussed above operate by surveillance of the genome for damage or misincorporation and then bring in enzymatic machines to repair the defects. Other systems of surveillance in bacterial genomes are **restriction/modification systems**. These look for foreign DNA that has invaded the cell, and then destroy it. In effect, this is another means of protecting the genome from the damage that could result from the integration of foreign DNA.

These systems for safeguarding the bacterial cell from invasion by foreign DNA use a combination of covalent modification and restriction by an endonuclease. Each species of bacteria modifies its DNA by **methylation** at specific sites (Fig. 19). This protects the DNA from cleavage by the corresponding **restriction endonuclease**. However, any foreign DNA (e.g. from an infecting bacteriophage or from a different species of bacteria) will not be methylated at that site, and the restriction endonuclease will cleave there. The result is that invading DNA will be cut up and inactivated, while not damaging the host DNA.

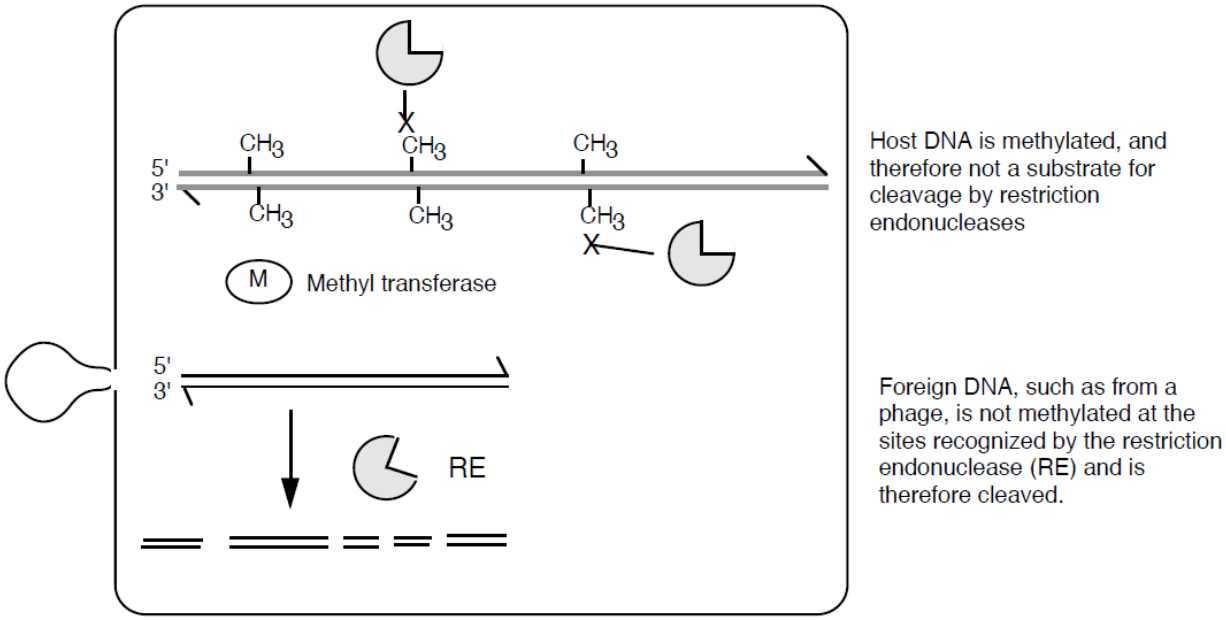


Figure 19. Restriction/modification systems in bacteria.

Any DNA that escapes the restriction endonuclease will be a substrate for the methylase. Once methylated, the bacterium now treats it like its own DNA, i.e. does not cleave it. This process can be controlled genetically and biochemically to aid in recombinant DNA work. Generally, the restriction endonuclease is encoded at the *r* locus and the methyl transferase is encoded at the *m* locus. Thus passing a plasmid DNA through an r^-m^+ strain (defective in restriction but competent for modification) will make it resistant to restriction by strains with a wildtype r^+ gene. For some restriction/modification systems, both the endonuclease and the methyl transferase are available commercially. In these cases, one can modify the foreign DNA (e.g. from humans) prior to ligating into cloning vectors to protect it from cleavage by the restriction endonucleases it may encounter after transformation into bacteria.

For the type II restriction/modification systems, the methylation and restriction occurs at the same, pseudopalindromic site. These are the most common systems, with a different sequence specificity for each bacterial species. This has provided the large variety of restriction endonucleases that are so commonly used in molecular biology.

REFERENCES

GENETICS: A conceptual approach, 4th Edition, Benjamin A. Pierce, W. H. Freeman and company England; 2006

Cell Biology, Genetics, Molecular Biology, Evolution and Ecology, P.S. Verma, V.K. Agarwal, S. Chand & Company Ltd, 2005

Principles of Molecular Biology, Veer Bala Rastogi, Medtech, 2016

Molecular Biology of the Cell. 4th edition. Alberts B, Johnson A, Lewis J, et al. New York: Garland Science; 2002.

http://www.ncbi.nlm.nih.gov/books/NBK22104/

https://www.google.co.in/search?q=dna+as+genetic+material&biw=1280&bih=890&tbm=isch&imgil=TUQEOSzOlQrxuM%253A%253B5_P9-NeMFLTYMM%253Bhttps%25253A%25252F%25252Fwww.youtube.com%25252Fwatch%25253Fv%2525253DATtWiTqn1Ok&source=iu&pf=m&fir=TUQEOSzOlQrxuM%253A%252C5_P9-NeMFLTYMM%252C_&dpr=1&usg=__t7Dy9JbSGZodbUp0eyZeNP5aO7w%3D&ved=0ahUKEwj3-vLAyqPKAhUVCY4KHTtrBisQyjcIUA&ei=k5SUVrflMpWSuAS71pnYAg#imgrc=TUQEOSzOlQrxuM%3A&usg=__t7Dy9JbSGZodbUp0eyZeNP5aO7w%3D

http://biologyntechnology.blogspot.in/2009/10/rna-as-genetic-material.html

www.bio.miami.edu/mccracken/genetics/lectures-2013/chapter_02.ppt

http://www.bx.psu.edu/~ross/workmg/RepairDNACh7.pdf

SATHYABAMA
INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)
Accredited "A" Grade by NAAC | 12B Status by UGC | Approved by AICTE
www.sathyabama.ac.in

SCHOOL OF BIO AND CHEMICAL ENGINEERING

DEPARTMENT OF BIOTECHNOLOGY

SBTA1301 – MOLECULAR BIOLOGY AND GENETICS

SBTA1301	MOLECULAR BIOLOGY AND GENETICS	L	T	P	Credit	Total Marks
		3	*	0	3	100

COURSE OBJECTIVES

- The course aims to give an understanding on the fundamentals of conventional genetics and the intricate molecular mechanisms of heredity and variations.
- To understand storage of genetic information and its translation at molecular level in prokaryotic and eukaryotic systems.

UNIT 1 CLASSICAL GENETICS **9 Hrs.**

Fundamental principles of genetics- Mendel's principles and experiments, gene interaction, multiple alleles, complementation, linkage, sex linked, sex limited and sex influenced inheritance; Chromosomes basis of heredity- extra-chromosomal inheritance; Linkage and crossing over; Hardy-Weinberg equilibrium, Extensions of Hardy- Weinberg equilibrium.

UNIT 2 STRUCTURE AND PROPERTIES OF NUCLEIC ACIDS **9 Hrs.**

Introduction to nucleic acids: Evidence for DNA&RNA as a genetic material; Structure and physicochemical properties of elements in DNA and RNA, Biological significance of differences in DNA and RNA. Primary structure of DNA: Chemical and structural qualities of 3',5'-Phosphodiester bond. Secondary Structure of DNA: Watson & Crick model, Chargaff's rule; DNA replication- Overview of differences in prokaryotic and eukaryotic DNA replication, D-loop and rolling circle mode of replication, Telomere replication in eukaryotes; Okazaki fragments, Fidelity of DNA replication, Inhibitors of DNA replication, DNA repair- Mutagens, DNA mutations and various types of repair mechanisms.

UNIT 3 TRANSCRIPTION **9 Hrs.**

Central Dogma in molecular biology -Structure and function of mRNA, rRNA tRNA and micro RNAs. Characteristics of promoter and enhancer sequences. RNA synthesis: Initiation, elongation and termination of RNA synthesis, Proteins of RNA synthesis, Fidelity of RNA synthesis, Inhibitors of transcription, Differences in prokaryotic and eukaryotic transcription. Basic concepts in RNA world: RNA processing: 5'-Capping, Splicing-Alternative splicing, Poly 'A' tail addition and base modification.

UNIT 4 TRANSLATION **9 Hrs.**

Introduction to Genetic code: Elucidation of genetic code, Codon degeneracy, Wobble hypothesis and its importance, Prokaryotic and eukaryotic ribosomes. Steps in translation: Initiation, Elongation and termination of protein synthesis. Inhibitors of protein synthesis. Post-translational modifications and its importance.

UNIT 5 REGULATION OF GENE EXPRESSION **9 Hrs.**

Organization of genes in prokaryotic and eukaryotic chromosomes- operon concept; Gene expression and regulation- Hierarchical levels of gene regulation, Prokaryotic gene regulation –lac and trp operon, Eukaryotic gene regulation- gene silencing.

Max.45 Hrs.

COURSE OUTCOMES

On completion of the course, student will be able to

- CO1 - Explain the foundations of Mendelian genetics and chromosomal theory and apply these, with appropriate terminology, to contemporary concepts in genetics.
- CO2 - Emphasize the molecular mechanism of DNA replication and repair in various organisms
- CO3 - Explain the properties of genetic materials and storage and processing of genetic information.
- CO4 - Analyze the processes of transcription and translation in both prokaryotes and eukaryotes at molecular level.
- CO5 - Understand the redundant and universal qualities of the genetic code and how it is used to determine the amino acid sequence of a polypeptide.
- CO6 - Compare the mechanisms of gene regulation in prokaryotes and eukaryotes.

TEXT / REFERENCE BOOKS

1. Lewin B., Genes XI, International Edition, Jocelyn Krebs, Stephen Kilpatrick and Elliott Goldstein, Jones & Bartlett Learning, 2017, ISBN 978-1-4496-5985-1
2. Tropp, Burton E., Molecular Biology: Genes to Proteins, 3rd Edition, Jones and Bartlett, 2008.
3. Glick B.R. and Pasternak J.J., Molecular Biotechnology: Principles and Applications of Recombinant DNA, 4th Edition. ASM, 2010

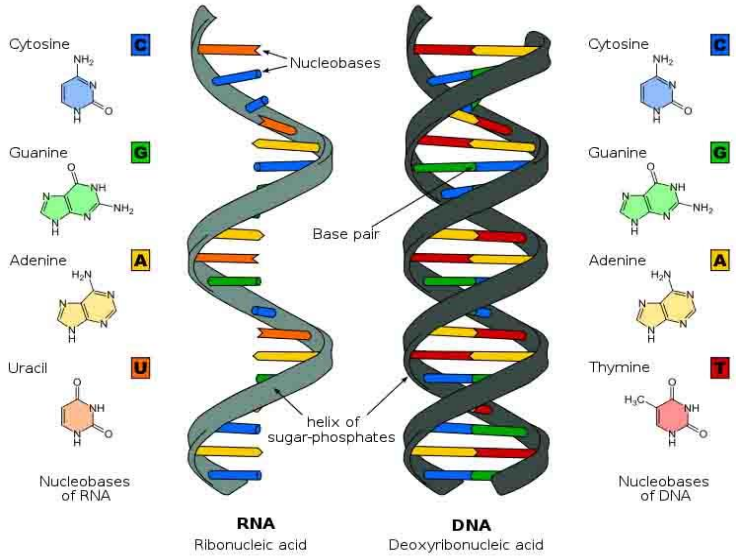
UNIT – III – TRANSCRIPTION

RIBONUCLEIC ACID

- RNA is a polymer of ribonucleotides linked together by 3'-5' phosphodiester linkage.
- The major role of RNA is to participate in protein synthesis, which requires three classes of RNA.

STRUCTURE OF RNA

- Back bone is sugar and phosphate group
- Nitrogenous bases linked to sugar moiety project from the backbone
- Nitrogenous bases are linked to pentose sugar through N-glycosidic linkage to form a nucleoside
- Phosphate group is linked with 3'OH of nucleoside through phosphoester linkage
- 2 nucleotides are linked through 3'-5'-phosphodiester linkage to form a dinucleotide
- More and more such groups will be linked to form a poly nucleotide chain
- Such a polymer has a free phosphate moiety at 5' end of ribose sugar and it is called as 5'-end of polynucleotide chain
- At other end, ribose has free 3'-OH group which is called as the 3'-end of polynucleotide chain
- In RNA, every nucleotide has an additional-OH present at 2'-position of ribose



- ❖ Synthesis of RNA is usually catalyzed by an enzyme—RNA polymerase
- ❖ By using DNA as a template
- ❖ The process is known as transcription
- ❖ There are also a number of RNA-dependent RNA polymerases that use RNA as their template for synthesis of a new strand of RNA
- ❖ A number of RNA viruses (such as poliovirus) use this type of enzyme to replicate their genetic material

BASIC PRINCIPLES OF TRANSCRIPTION AND TRANSLATION

- RNA is the bridge between genes and the proteins for which they code
- Transcription is the synthesis of RNA using information in DNA
- Transcription produces messenger RNA (mRNA)
- Translation is the synthesis of a polypeptide, using information in the mRNA
- Ribosomes are the sites of translation
- In prokaryotes, translation of mRNA can begin before transcription has finished
- In a eukaryotic cell, the nuclear envelope separates transcription from translation
- Eukaryotic RNA transcripts are modified through RNA processing to yield the finished mRNA
- A primary transcript is the initial RNA transcript from any gene prior to processing
- The *central dogma* is the concept that cells are governed by a cellular chain of command: DNA → RNA → protein

TYPES OF RNA

In all prokaryotic and eukaryotic organisms, three main classes of RNA molecules exist-

1) Messenger RNA(m RNA)
2) Transfer RNA (t RNA)
3) Ribosomal RNA (r RNA)

The other are –

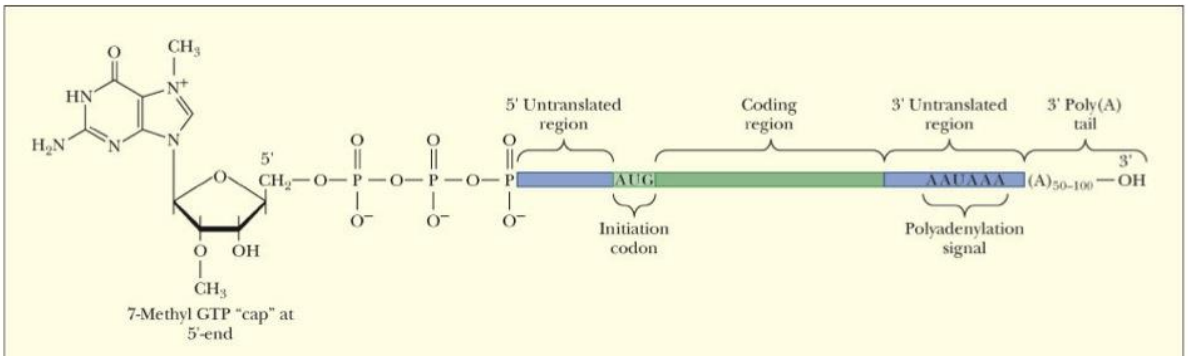
- small nuclear RNA (SnRNA),
- micro RNA(mi RNA) and
- small interfering RNA(Si RNA) and
- heterogeneous nuclear RNA (hnRNA)

Heterogeneous nuclear RNA (hnRNA)

- In mammalian nuclei , hnRNA is the immediate product of gene transcription
- The nuclear product is heterogeneous in size (Variable) and is very large.
- Molecular weight may be more than $10^{7,}$ while the molecular weight of m RNA is less than $2x\ 10^{6}$
- 75 % of hnRNA is degraded in the nucleus, only 25% is processed to mature m RNA

Messenger RNA

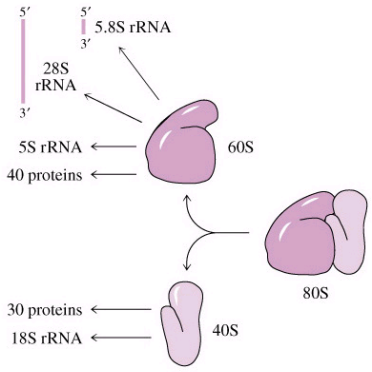
- Messenger RNA (mRNA) carries information about a protein sequence to the ribosomes, the protein synthesis factories in the cell
- It is coded so that every three nucleotides (a codon) correspond to one amino acid
- . In eukaryotic cells, once precursor mRNA (pre-mRNA) has been transcribed from DNA, it is processed to mature mRNA
- This removes its introns—non-coding sections of the pre-mRNA
- The mRNA is then exported from the nucleus to the cytoplasm, where it is bound to ribosomes and translated into its corresponding protein form with the help of tRNA



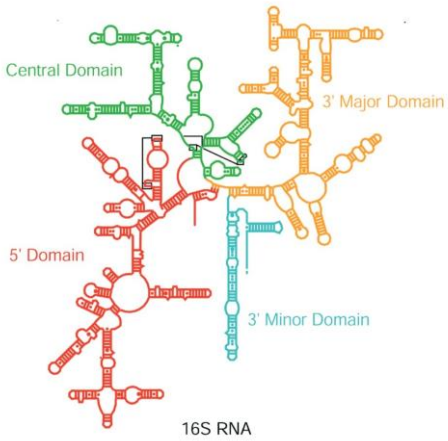
Ribosomal RNA

The mammalian ribosome contains two major nucleoprotein subunits—a larger one with a molecular weight of 2.8 x 10^6 (60S) and a smaller subunit with a molecular weight of 1.4 x 10^6 (40S).

- Is a structural and functional component of ribosomes w/ are "platforms" on which protein synthesis occur.
- Consist of about 35% protein and 65% ribosomal RNA.
- Complexed w/ proteins, the rRNA forms the cellular structures called the ribosomes.
- Ribosomal RNA (rRNA) is the catalytic component of the ribosomes
- Eukaryotic ribosomes contain four different rRNA molecules: 18S, 5.8S, 28S and 5S rRNA
- Three of the rRNA molecules are synthesized in the nucleolus, and one is synthesized elsewhere.
- In the cytoplasm, ribosomal RNA and protein combine to form a nucleoprotein called a ribosome
- The ribosome binds mRNA and carries out protein synthesis
- Several ribosomes may be attached to a single mRNA at any time.
- Nearly all the RNA found in a typical eukaryotic cell is rRNA.



Eukaryote

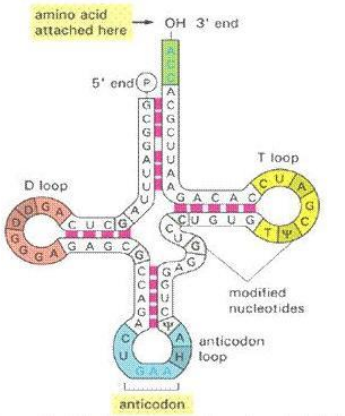


16S RNA

Transfer RNA

1) Primary structure- The nucleotide sequence of all the t RNA molecules allows extensive intrastand complimentarity that generates a secondary structure.

2) Secondary structure- Each single t- RNA shows extensive internal base pairing and acquires a clover leaf like structure. The structure is stabilized by hydrogen bonding between the bases and is a consistent feature.



- ❑ The L shaped tertiary structure is formed by further folding of the clover leaf due to hydrogen bonds between T and D arms.
- ❑ The base paired double helical stems get arranged in to two double helical columns, continuous and perpendicular to one another.

Small nuclear RNA (SiRNA)

- ❑ Most of these molecules are complexed with proteins to form small nuclear ribonucleoproteins particles{snurps} and are distributed in the nucleus, in the cytoplasm, or in both.
- ❑ They range in size from 20 to 300 nucleotides and are present in 100,000–1,000,000 copies per cell.

Function:

- ✓ Help w/ the processing of the initial mRNA into mature form that is ready for export out of the nucleus-{ splicing}.
- ❑ Of the several snRNAs, U1, U2, U4, U5, and U6 are involved in intron removal and the processing of hnRNA into mRNA
- ❑ The U7 snRNA is involved in production of the correct 3' ends of histone mRNA—which lacks a poly(A) tail.

MicroRNA, mi RNA, Small interferingRNA, siRNA

- ❑ These two classes of RNAs represent a subset of small RNAs; both play important roles in gene regulation.
- ❑ miRNAs and siRNAs cause inhibition of gene expression by decreasing specific protein production albeit apparently via distinct mechanisms

Five different types of RNA, each encoded by different genes:

1. mRNA	Messenger RNA, encodes the amino acid sequence of a polypeptide.
2. tRNA	Transfer RNA, transports amino acids to ribosomes during translation.
3. rRNA	Ribosomal RNA, forms complexes called ribosomes with protein, the structure on which mRNA is translated.
4. snRNA	Small nuclear RNA, forms complexes with proteins used in eukaryotic RNA processing (e.g., exon splicing and intron removal).
5. miRNA/siRNA	Micro RNA/small interfering RNA, short ~22 nt RNA sequences that bind to 3' UTR target mRNAs and result in gene silencing.

Transcription

When a protein is needed by a cell, the genetic code for that protein must be read from the DNA and processed.

A two step process:

1. Transcription = synthesis of a single-stranded RNA molecule using the DNA template (1 strand of DNA is transcribed).
2. Translation = conversion of a messenger RNA sequence into the amino acid sequence of a polypeptide (i.e., protein synthesis)

✓ Both processes occur throughout the cell cycle. Transcription occurs in the nucleus, whereas translation occurs in the cytoplasm.

Materials required

✓ The enzyme RNA polymerase or DNA directed RNA polymerase

✓ DNA template – the transcription unit

✓ All the four types of ribonucleoside triphosphates (ATP, CTP, GTP and UTP)

- ✓ Divalent metal ions Mg2+ or Mn2+ as a cofactor
- ✓ No primer is needed for RNA synthesis

Transcription: How is an RNA strand synthesized?

1. **Regulated by gene regulatory elements within each gene.**
2. **DNA unwinds next to a gene.**
3. **RNA is transcribed 5' to 3' from the template (3' to 5').**
4. **Similar to DNA synthesis, except:**
    - ✓ **NTPs instead of dNTPs (no deoxy-)**
    - ✓ **No primer**
    - ✓ **No proofreading**
    - ✓ **Adds Uracil (U) instead of thymine (T)**
    - ✓ **RNA polymerase**

RNA POLYMERASE

α - Two α polypeptide chains, coded by gene rpoA, assembly of core enzyme & and help in the probably in the recognition of promoter

β – one copy of β subunit, , coded by gene rpoB, binds with the incoming nucleotides & helps in the formation of the first phosphodiester bond

β' – one copy of β' subunit, , coded by gene rpoC, binds with the template strand or antisense DNA strand

σ - Single polypeptide chain, loosely attached to the core enzyme.

σ Subunit

- Recognises the start signal on DNA molecules and directs the core enzyme of RNA polymerase to bind to the promoter region upstream of initiation codon
- Recognises two special sequences of bases in the promoter region of the coding strand (i.e. Antitemplate strand) of DNA - -10 sequence and -35 sequence
- Recognises of promoter sequence, facilitates opening or melting of DNA helix

- Separates from core enzyme once about 10nt are joined to initiate RNA synthesis

FUNCTIONS OF RNA POLYMERASE

- Unwinds about 15 bases of DNA around the initiation site to form an open promoter – DNA complex and provides single strand of DNA to act as template for transcription
- Catalyses the formation of phosphodiester bonds between successive nt of a polynucleotide chain during synthesis of RNA
- Lacks proof reading 3'-5' exonuclease activity
- Therefore, one error for every 10^4 to 10^5 ribonucleotides incorporated is introduced during RNA transcription
- But the mistake during transcription is not serious because of its high turnover and Wobble pairing during translation

Three Steps to Transcription:

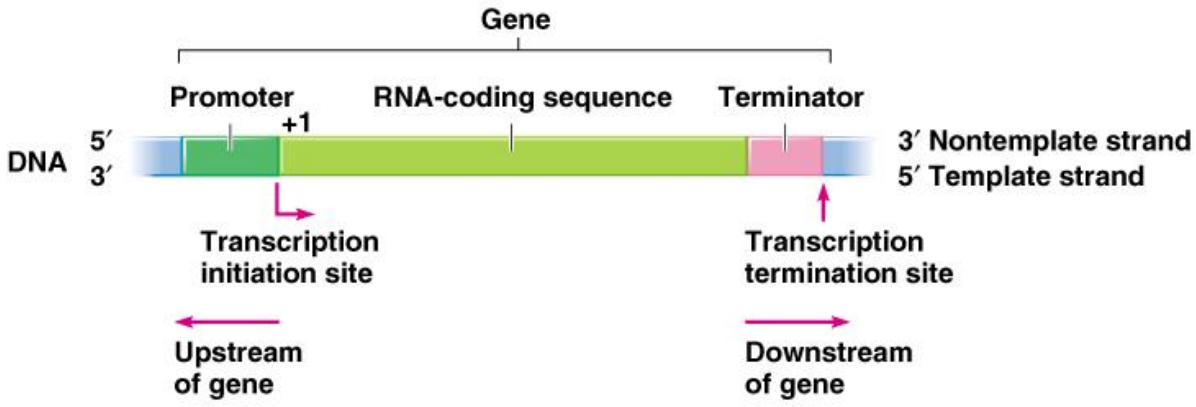
1. **Initiation**
2. **Elongation**
3. **Termination**

✓ **Occur in both prokaryotes and eukaryotes.**

✓ **Elongation is conserved in prokaryotes and eukaryotes.**

✓ **Initiation and termination proceed differently.**

INITIATION



Each gene has three regions:

1. 5’ Promoter, attracts RNA polymerase

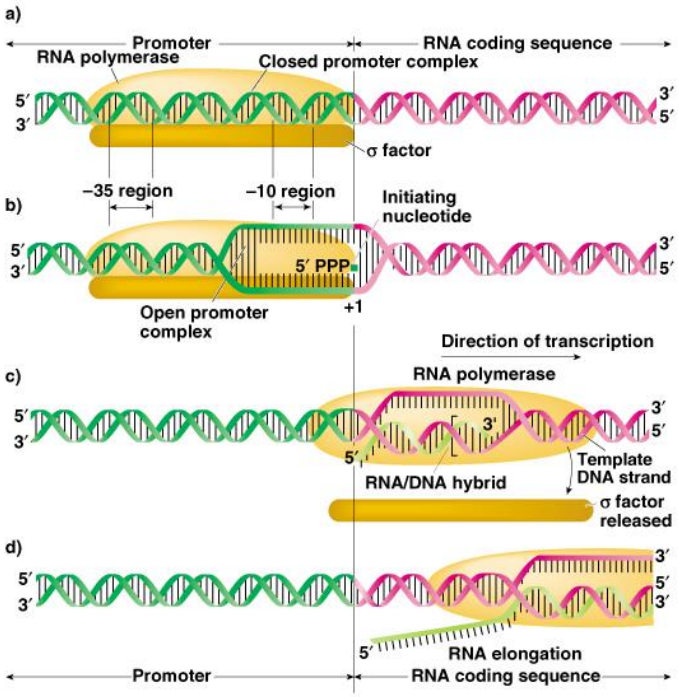
   -10 bp 5’-TATAAT-3’

   -35 bp 5’-TTGACA-3’

2. Transcribed sequence (transcript) or RNA coding sequence
3. 3’ Terminator, signals the stop point

Step 1-Initiation, *E. coli* model:

1. RNA polymerase combines with sigma factor (a polypeptide) to create RNA polymerase holoenzyme
   - ✓ Recognizes promoters and initiates transcription.
   - ✓ Sigma factor required for efficient binding and transcription.
   - ✓ Different sigma factors recognize different promoter sequences.
2. RNA polymerase holoenzyme binds promoters and untwists DNA
   - ✓ Binds loosely to the -35 promoter (DNA is d.s.)
   - ✓ Binds tightly to the -10 promoter and untwists
3. Different types and levels of sigma factors influence the level and dynamics of gene expression (how much and efficiency).



Step 2-Elongation, *E. coli* model:

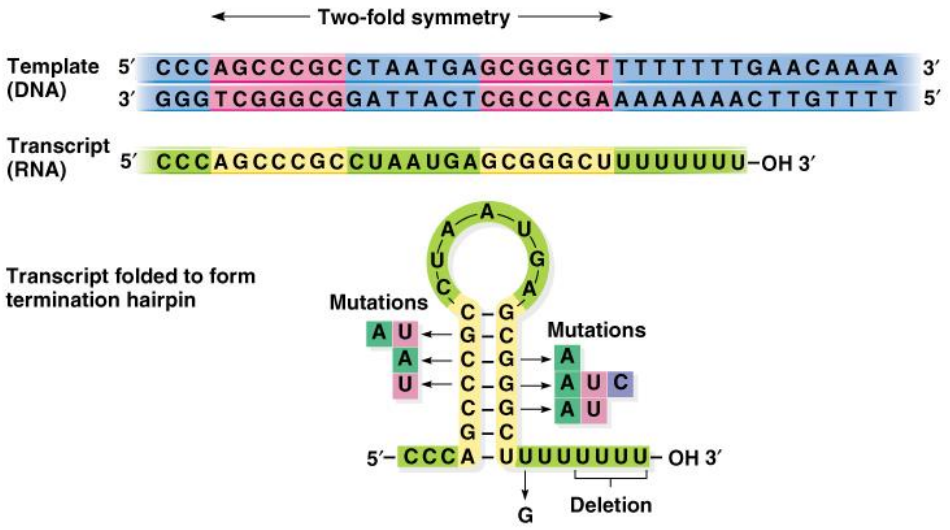
1. After 8-9 bp of RNA synthesis occurs, sigma factor is released and recycled for other reactions.
2. RNA polymerase completes the transcription at 30-50 bp/second (and order of magnitude slower than DNA polymerase).
3. DNA untwists rapidly, and re-anneals behind the enzyme.
4. Part of the new RNA strand is hybrid DNA-RNA, but most RNA is displaced as the helix reforms.

Step 3-Termination, *E. coli* model:

Two types of terminator sequences occur in prokaryotes:

1. Type I (ρ-independent)

Palindromic, inverse repeat forms a hairpin loop and is believed to physically destabilize the DNA-RNA hybrid.



2. Type II (ρ-dependent)

Involves ρ factor proteins that break the hydrogen bonds between the template DNA and RNA.

The Basic Rules of Transcription

Before we examine the process of eukaryotic transcription, let's pause to summarize some of the general principles of bacterial transcription.

1. Transcription is a selective process; only certain parts of the DNA are transcribed at any one time.

2. RNA is transcribed from single-stranded DNA. Within a gene, only one of the two DNA strands—the template strand—is normally copied into RNA.

3. Ribonucleoside triphosphates are used as the substrates in RNA synthesis. Two phosphate groups are cleaved from a ribonucleoside triphosphate, and the resulting nucleotide is joined to the 3′-OH group of the growing RNA strand.

4. RNA molecules are antiparallel and complementary to the DNA template strand. Transcription is always in the 5′→3′ direction, meaning that the RNA molecule grows at the 3′ end.

5. Transcription depends on RNA polymerase—a complex, multimeric enzyme. RNA polymerase consists of a core enzyme, which is capable of synthesizing RNA, and other subunits that may join transiently to perform additional functions.

6. A sigma factor enables the core enzyme of RNA polymerase to bind to a promoter and initiate transcription.

7. Promoters contain short sequences crucial in the binding of RNA polymerase to DNA; these consensus sequences are interspersed with nucleotides that play no known role in transcription.

8. RNA polymerase binds to DNA at a promoter, begins transcribing at the start site of the gene, and ends transcription after a terminator has been transcribed.

EUKARYOTIC TRANSCRIPTION

Table 5-3. Proper.ties of three different eukaryotic nuclear RNA polymerases.

Enzymes	Location	Product and abundance	Sensitivity to α - amanitin
1. RNA polymerase I	Nucleolus	rRNA (50–70%) (except 5S rRNA)	Not sensitive
2. RNA polymerase II	Nucleoplasm	hnRNA (mRNA) (20–40%)	Sensitive
3. RNA polymerase III	Nucleoplasm	tRNA(~ 10%) (and 5S rRNA)	Inhibited in animals at high levels; not in yeast and insects.

Most eukaryotic cells
possess three distinct types of RNA polymerase, each of
which is responsible for transcribing a different class of
RNA: **RNA polymerase I** transcribes rRNA; **RNA polymerase II** transcribes pre-mRNAs, snoRNAs, some miRNAs,
and some snRNAs; and **RNA polymerase III** transcribes
other small RNA molecules—specifically tRNAs, small
rRNA, some miRNAs, and some snRNAs (**Table 13.3**). RNA
polymerases I, II, and III are found in all eukaryotes. Two
additional RNA polymerases, **RNA polymerase IV** and **RNA polymerase V**, have been found in plants. RNA polymerases
IV and V transcribe RNAs that play a role in DNA methylation
and chromatin structure.
All eukaryotic polymerases are large, multimeric enzymes,
typically consisting of more than a dozen subunits. Some sub- units are common to all RNA polymerases, whereas others
are limited to one of the polymerases. As in bacterial cells, a

number of accessory proteins bind to the core enzyme and affect its function.

Table 11.2 Eukaryotic promoter elements.

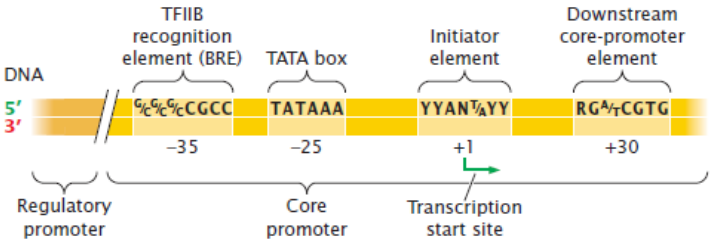
Promoter	Position	Transcription factor	Consensus sequence
Upstream core promoter elements			
TFIIB recognition element (BRE)	−37 to −32	TFIIB	(G/C)(G/C)(G/A)CGCC
TATA box	−31 to −26	TBP	TATA(A/T)AA(G/A)
Initiator (Inr)	−2 to +4	TAF1 ($TAF_{II}250$) TAF2 ($TAF_{II}150$)	$PyPyA_{+1}N(T/A)PyPy$
Downstream core promoter elements			
Motif ten element (MTE)	+18 to +27	TFIID	C(G/A)A(A/G)C(G/C) (C/A/G)AACG(G/C)
Downstream promoter element (DPE)	+28 to +32	TAF9 ($TAF_{II}40$) TAF6 ($TAF_{II}60$)	(A/G)G(A/T)(C/T)(G/A/C)
Proximal promoter elements			
CAAT box	−200 to −70	CBF, NF1, C/EBP	CCAAT
GC box	−200 to −70	Sp1	GGGCGG

Most, but not all, CAAT and GC boxes are located between −200 and −70.
CBF, CAAT-binding protein; C/EBP, CAAT/enhancer-binding protein; N, any (A, T, C, or G); Py, pyrimidine (C or T).

General transcription machinery	RNA pol II (pol II, RNAPII)	RNA polymerase II	Catalysis of RNA synthesis
	General transcription factors:		
	TFIIB	Transcription factor for RNA polymerase II B	Stabilization of TBP–DNA interactions, recruitment of RNA pol II–TFIIF, start site selection by RNA pol II
	TFIID:		
	TBP	TATA-binding protein	Core promoter recognition, TFIIB recruitment
	TAF	TBP-associated factor	Core promoter recognition/selectivity
	TFIIE	Transcription factor for RNA polymerase II E	TFIIH recruitment
	TFIIF	Transcription factor for RNA polymerase II F	Recruitment of RNA pol II to promoter DNA–TBP–TFIIB complex
	TFIIH	Transcription factor for RNA polymerase II H	Promoter melting, helicase, RNA pol II CTD kinase
	Mediator	Mediator	Transduces regulatory information from activator and repressor proteins to RNA pol II

Table 11.1 Proteins that regulate transcription.

Category	Acronym	Derivation of name	Function
Transcription factors (activators or repressors)	Some examples mentioned in text:		
	CBF	CAAT binding factor	Binds CAAT box
	C/EBP	CAAT/enhancer-binding protein	Binds CAAT box
	CREB	cAMP response element-binding protein	Binds the cAMP response element
	CTCF	CCCTC-binding factor	Binds insulator element (CCCTC) and mediates enhancer blocking activity
	FOG-1	Friend of GATA-1	Required for developmental expression of β-globin genes
	GATA-1	GATA-binding protein	Required for developmental expression of β-globin genes
	NF-E2	Nuclear factor erythroid-derived 2	Required for developmental expression of β-globin genes
	NF-κB	Nuclear factor of kappa light polypeptide enhancer in B cells	Central mediator of human stress and immune responses
	USF1, USF2	Upstream stimulatory factor 1 and 2	Bind insulator element, recruit chromatin-modifying enzymes
	SATB1	Special AT-rich binding protein 1	Matrix attachment region (MAR) binding protein required for T-cell-specific gene regulation
	Sp1	SV40 early and late promoter-binding protein 1	Binds GC box
General transcription machinery	RNA pol II (pol II, RNAPII)	RNA polymerase II	Catalysis of RNA synthesis
	General transcription factors:		
	TFIIB	Transcription factor for RNA polymerase II B	Stabilization of TBP–DNA interactions, recruitment of RNA pol II–TFIIF, start site selection by RNA pol II
	TFIID:		
	TBP	TATA-binding protein	Core promoter recognition, TFIIB recruitment
	TAF	TBP-associated factor	Core promoter recognition/selectivity
	TFIIE	Transcription factor for RNA polymerase II E	TFIIH recruitment
	TFIIF	Transcription factor for RNA polymerase II F	Recruitment of RNA pol II to promoter DNA–TBP–TFIIB complex
	TFIIH	Transcription factor for RNA polymerase II H	Promoter melting, helicase, RNA pol II CTD kinase
	Mediator	Mediator	Transduces regulatory information from activator and repressor proteins to RNA pol II
Coactivators and corepressors	Chromatin modification complexes:		
	HAT	Histone acetyltransferase	Acetylates histones
	HDAC	Histone deacetylase	Deacetylates histones
	CBP	CREB-binding protein	HAT activity
	HMT	Histone methyltransferase	Methylates histones
	LSD1	Lysine-specific demethylase 1	Demethylates histones
	Chromatin remodeling complexes:		
	SWI/SNF	Mating-type switching defective/sucrose nonfermenters	ATP-dependent chromatin remodeling (sliding and disassembly)
	ISWI	Imitation Swi2	ATP-dependent chromatin remodeling (sliding)
	SWR1	Swi2/Snf2-related 1	ATP-dependent chromatin remodeling (histone replacement)
Elongation factors	FACT	Facilitates chromatin transcription	Transcription-dependent nucleosome alterations
	Elongator	Elongator	Exact function in elongation unknown
	TFIIS	Transcription factor for RNA polymerase II S	Facilitates RNA pol II passage through regions that cause transcriptional arrest



13.15 The promoters of genes transcribed by RNA polymerase II consist of a core promoter and a regulatory promoter that contain consensus sequences. Not all the consensus sequences shown are found in all promoters.

A significant difference between bacterial and eukaryotic transcription is the existence of three different eukaryotic RNA polymerases, which recognize different types of promoters. In bacterial cells, the holoenzyme (RNA polymerase plus the sigma factor) recognizes and binds directly to sequences in the promoter. In eukaryotic cells, promoter recognition is carried out by accessory proteins that bind to the promoter and then recruit a specific RNA polymerase (I, II, or III) to the promoter.

One class of accessory proteins comprises **general transcription factors**, which, along with RNA polymerase, form the **basal transcription apparatus**—a group of proteins that assemble near the start site and are sufficient to initiate minimal levels of transcription. Another class of accessory proteins consists of **transcriptional activator proteins**, which bind to specific DNA sequences and bring about higher levels of transcription by stimulating the assembly of the basal transcription apparatus at the start site.

A promoter for a gene transcribed by RNA polymerase II typically consists of two primary parts: the core

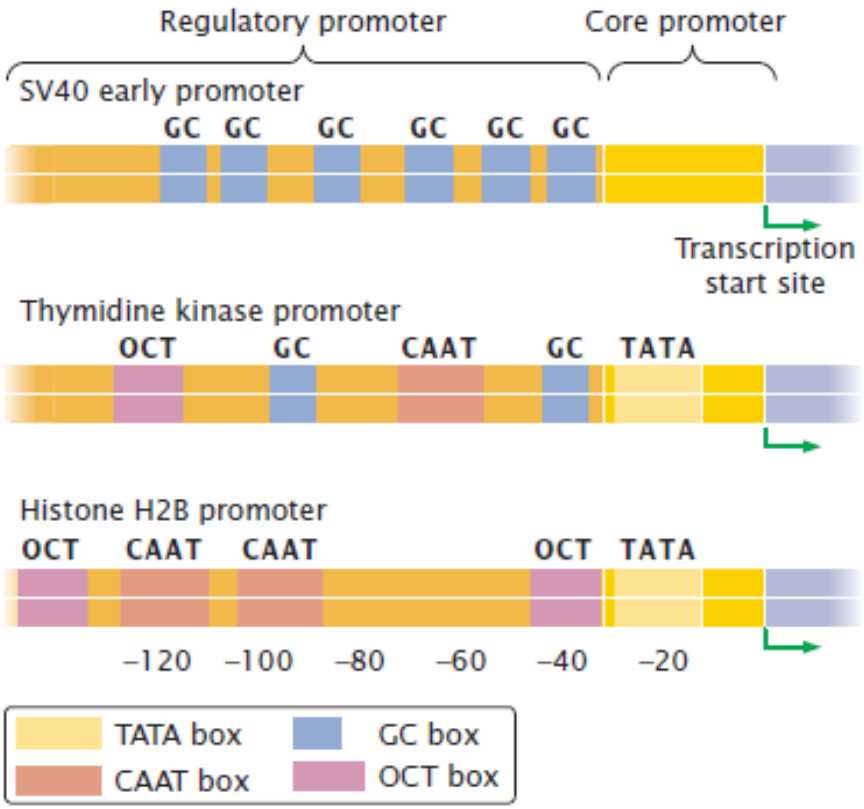
promoter and the regulatory promoter.

Core promoter The **core promoter** is located immediately upstream of the gene and is the site to which the basal transcription apparatus binds. The core promoter typically includes one or more consensus sequences. One of the most common of these sequences is the **TATA box**, which has the consensus sequence TATAAA and is located from −25 to −30 bp upstream of the start site. Additional consensus sequences that may be found in the core promoters of genes transcribed by RNA polymerase II are shown in Figure 13.15. These consensus sequences are recognized by transcription factors that bind to them and serve as a platform for the assembly of the basal transcription apparatus.

Regulatory promoter The **regulatory promoter** is located immediately upstream of the core promoter. A variety of different consensus sequences can be found in the regulatory promoters, and they can be mixed and matched in different combinations. Transcriptional activator proteins bind to these sequences and either directly or indirectly make contact with the basal transcription apparatus and affect the rate at which transcription is initiated.

Transcriptional activator proteins also regulate transcription by binding to more-distant sequences called **enhancers**. The DNA between an enhancer and the promoter loops out, and so transcriptional activator proteins bound to the enhancer can interact with the basal transcription machinery at the core promoter.

Polymerase I and III promoters RNA polymerase I and RNA polymerase III each recognize promoters that are distinct from those recognized by RNA polymerase II. For example, promoters for small rRNA and tRNA genes, transcribed by RNA polymerase III, contain **internal promoters** that are downstream of the start site and are transcribed into the RNA.



13.16 Consensus sequences in the promoters of three eukaryotic genes. These promoters illustrate the principle that consensus sequences can be mixed and matched in different combinations to yield a functional eukaryotic promoter.

Initiation of Eukaryotic Transcription

For the eukaryotic transcription the regulatory DNA sequences (such as promoters, enhancers and silencers) for genes transcribed by each of the three RNA polymerases differ. Various transcription factors are also involved in the formation of a transcription complex which are needed

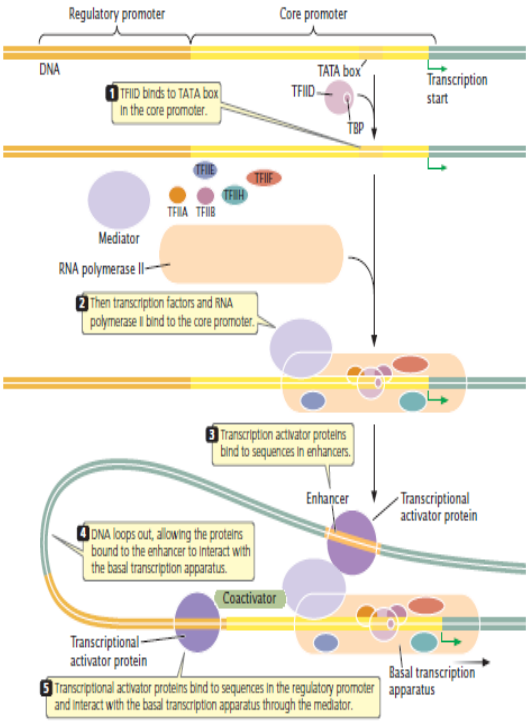
for initiation of transcription. Generally, each of RNA polymerase is believed to have its own set of transcription factors, however, TF II D or a part of it (*e.g.*, TBP=TATA binding protein) is required for all the three RNA polymerases.

The **transcription factors** (**TFs**) can be defined as proteins, which are needed for initiation of transcription, but are not part of the RNA polymerase. They help in DNA binding of a RNA polymerase to constitute the so-called **pre-initiation complex** or **transcription complex**. After the formation of this complex initiation of transcription occurs. All known transcription factors may recognize either **DNA sequences, another factor or RNA polymerase.**

Formation of transcriptosome with RNA pol II.

A promoter sequence which is responsible for constitutive expression of common genes (also **called house keeping genes**) in all cells, is called **generic promoter.**

The generic promoter cannot bring about regulated expression (i.e., tissue or stimulus specific expression of genes, called **luxary genes**).



13.17 Transcription is initiated at RNA polymerase II promoters. Transcription is initiated when the TFIID transcription factor binds to the TATA box, followed by the binding of a preassembled holoenzyme containing general transcription factors, RNA polymerase II, and the mediator. TBP stands for TATA-binding protein.

Initiation of transcription on the generic promoter by RNA polymerase II requires the action of diverse transcription factors (TFs) in the following order :

(i) **TF II D** binds at TATA box;

(ii) the step (i) permits the association of **TF IIA** and **TF IIB**;

(iii) TF II B forms the so-called DB complex and RNA polymerase II associates to promoter site;

(iv) RNA pol II is accompanied to the promoter by **TF II F** to form a transcription complex ;

(v) orderly addition of **TF II E, TF II H** and **TF II J** helps the initiation process.

Elongation of RNA Chain in Eukaryotes

There are certain accessory proteins of transcription, called **elongation factors**, which enhance the overall activity of RNA polymerase II and lead to increase in the elongation rate. At least two such proteins are known:

(1) The **TF II F** accelerates RNA chain growth relatively uniformly in concord with RNA polymerase II.

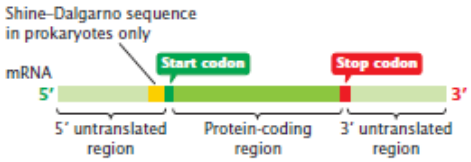
(2) The **TF II S** (also called **S II**) helps in elongation of RNA chain by unburdening the obstruction in the path of such elongation. TF II S is known to act by first causing hydrolytic cleavage at 3´ end of RNA chain, thereby, helping in the forward movement of RNA polymerase through any block to elongation.

Termination of Eukaryotic Transcription

(1) In eukaryotes, the actual termination of RNA polymerase II activity during transcription may take place through termination sites similar to those found in prokaryotes.

(2) However, the nature of individual sites is not known. Such termination sites are believed to be present away (sometimes up to one kilobase away from the site of the 3´end of mRNA).

POST TRANSCRIPTIONAL MODIFICATIONS



14.5 Three primary regions of mature mRNA are the 5′ untranslated region, the protein-coding region, and the 3′ untranslated region.

Pre-mRNA Processing

In bacterial cells, transcription and translation take place simultaneously; while the 3′ end of an mRNA is undergoing transcription, ribosomes attach to the Shine–Dalgarno sequence near the 5′

end and begin translation. Because transcription and translation are coupled, bacterial mRNA has little opportunity to be modified before protein synthesis. In contrast, transcription and translation are separated in both time and space in eukaryotic cells. Transcription takes place in the nucleus, whereas translation takes place in the cytoplasm; this separation provides an opportunity for eukaryotic RNA to be modified before it is translated. Indeed, eukaryotic mRNA is extensively altered after transcription. Changes are made to the 5′ end, the 3′ end, and the protein coding section of the RNA molecule.

Table 14.2 Posttranscriptional modifications to eukaryotic pre-mRNA

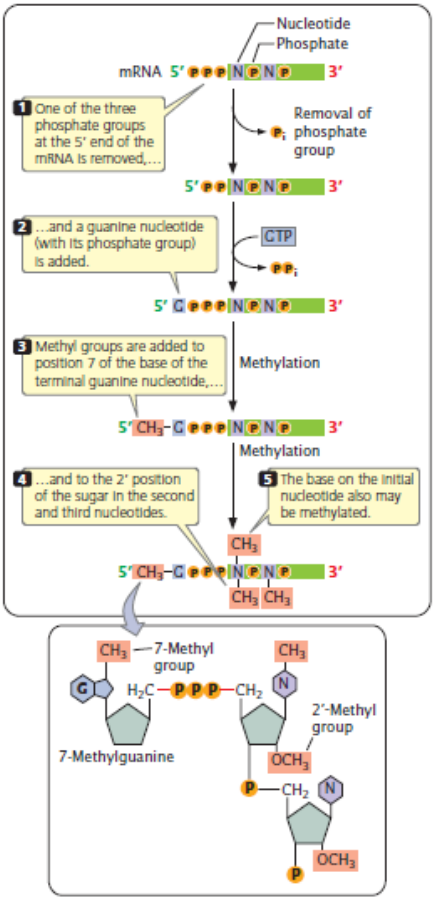
Modification	Function
Addition of 5′ cap	Facilitates binding of ribosome to 5′ end of mRNA, increases mRNA stability, enhances RNA splicing
3′ cleavage and addition of poly(A) tail	Increases stability of mRNA, facilitates binding of ribosome to mRNA
RNA splicing	Removes noncoding introns from pre-mRNA, facilitates export of mRNA to cytoplasm, allows for multiple proteins to be produced through alternative splicing
RNA editing	Alters nucleotide sequence of mRNA

The Addition of the 5′ Cap

One type of modification of eukaryotic pre-mRNA is the addition at its 5′ end of a structure called a **5′ cap**. The cap consists of an extra nucleotide at the 5′ end of the mRNA and methyl groups (CH3) on the base in the newly added nucleotide and on the 2′-OH group of the sugar of one or more nucleotides at the 5′ end (**Figure 14.6**). The addition of the cap takes place rapidly after the initiation of transcription and, as will be discussed in more depth in Chapter 15, the 5′ cap functions in the initiation of translation. Cap-binding proteins recognize the cap and attach to it; a ribosome then binds to these proteins and moves downstream along the mRNA until the start codon is reached and translation begins. The presence of a 5′ cap also increases the stability of mRNA and influences the removal of introns.

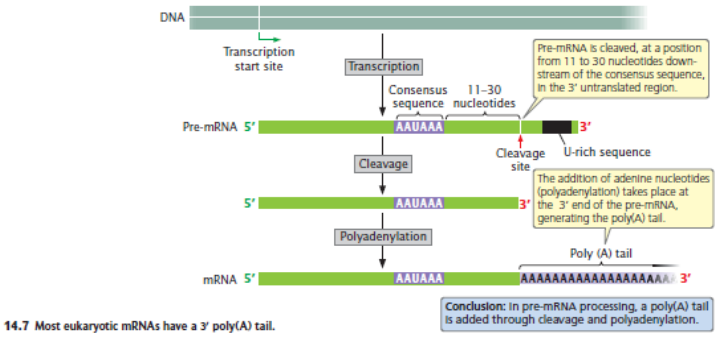
Three phosphate groups are present at the 5′ end of all RNA molecules because phosphate groups are not cleaved from the first ribonucleoside triphosphate in the transcription reaction. The 5′ end of pre-mRNA can be represented as 5′–pppNpNpN . . . , in which the letter “N” represents a ribonucleotide and “p” represents a phosphate. Shortly after the initiation of transcription, one of

these phosphate groups is removed and a guanine nucleotide is added. This guanine nucleotide is attached to the premRNA by a unique 5′–5′ bond, which is quite different from the usual 5′–3′ phosphodiester bond that joins all the other nucleotides in RNA. One or more methyl groups are then added to the 5′ end; the first of these methyl groups is added
to position 7 of the base of the terminal guanine nucleotide, making the base 7-methylguanine. Next, a methyl group may be added to the 2′ position of the sugar in the second and third nucleotides, as shown in Figure 14.6. Rarely, additional methyl groups may be attached to the bases of the second and third nucleotides of the pre-mRNA. Several different enzymes take part in the addition of the 5′ cap. The initial step is carried out by an enzyme that associates with RNA polymerase II. Because neither RNA polymerase I nor RNA polymerase III have this associated enzyme, RNA molecules transcribed by these polymerases (rRNAs, tRNAs, and some snRNAs) are not capped.



14.6 Most eukaryotic mRNAs have a 5′ cap. The cap consists of a nucleotide with 7-methylguanine attached to the pre-mRNA by a unique 5′–5′ bond (shown in detail in the bottom box).

The Addition of the Poly(A) Tail



14.7 Most eukaryotic mRNAs have a 3′ poly(A) tail.

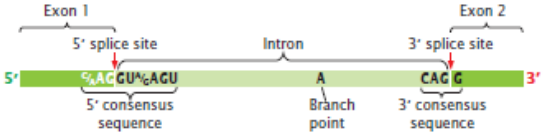
A second type of modification to eukaryotic mRNA is the addition of 50 to 250 or more adenine nucleotides at the 3′ end, forming a **poly(A) tail**. These nucleotides are not encoded in the DNA but are added after transcription in a process termed polyadenylation. Many eukaryotic genes transcribed by RNA polymerase II are transcribed well beyond the end of the coding sequence; most of the extra material at the 3′ end is then

cleaved and the poly(A) tail is added. For some pre-mRNA molecules, more than 1000 nucleotides may be removed from the 3′ end before polyadenylation. Processing of the 3′ end of pre-mRNA requires sequences both upstream and downstream of the cleavage site. The consensus sequence AAUAAA is usually from 11 to 30 nucleotides upstream of the cleavage site and determines the point at which cleavage will take place. A sequence rich in uracil nucleotides (or in guanine and uracil nucleotides) is typically downstream of the cleavage site. A large number of proteins take part in finding the cleavage site and removing the 3′ end. After cleavage has been completed, adenine nucleotides are added to the new 3′ end, creating the poly(A) tail.

The poly(A) tail confers stability on many mRNAs, increasing the time during which the mRNA remains intact

and available for translation before it is degraded by cellular enzymes. The stability conferred by the poly(A) tail depends on the proteins that attach to the tail and on its length. The poly(A) tail also facilitates attachment of the ribosome to the mRNA. Poly(U) tails are added to the 3′ ends of some mRNAs, microRNAs, and small nuclear RNAs. Although the function of poly(U) tails is still under investigation, evidence suggests that poly(U) tails on some mRNAs may facilitate their

degradation.



14.8 Splicing of pre-mRNA requires consensus sequences. Critical consensus sequences are present at the 5′ splice site and the 3′ splice site. A weak consensus sequence (not shown) exists at the branch point.

RNA Splicing

The other major type of modification of eukaryotic premRNA is the removal of introns by **RNA splicing**. This

modification takes place in the nucleus, before the RNA moves to the cytoplasm.

Consensus sequences and the spliceosome

Splicing requires the presence of three sequences in the intron. One end of the intron is referred to as the **5′ splice site**, and the other end is the **3′ splice site** (**Figure 14.8**); these splice sites possess short consensus sequences. Most introns in premRNAs begin with GU and end with AG, indicating that these sequences play a crucial role in splicing. Indeed, changing a single nucleotide at either of these sites prevents splicing.

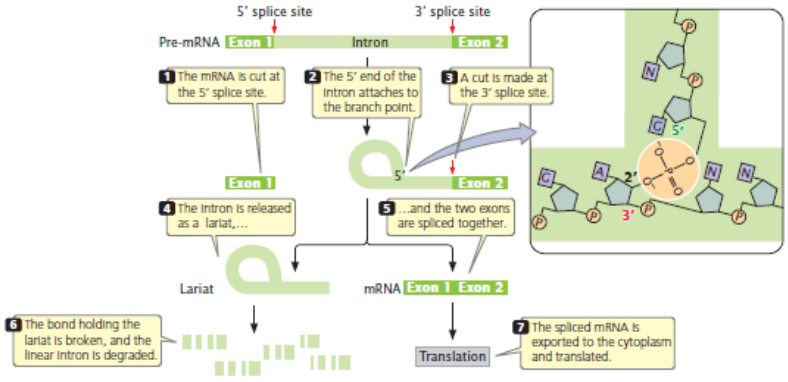
The third sequence important for splicing is at the **branch point**, which is an adenine nucleotide that lies

from 18 to 40 nucleotides upstream of the 3′ splice site The sequence surrounding the branch point does not have a strong consensus. The deletion or mutation of the adenine nucleotide at the branch point prevents splicing. Splicing takes place within a large structure called the **spliceosome**, which is one of the largest and most complex of all molecular complexes. The spliceosome consists of five RNA molecules and almost 300 proteins.

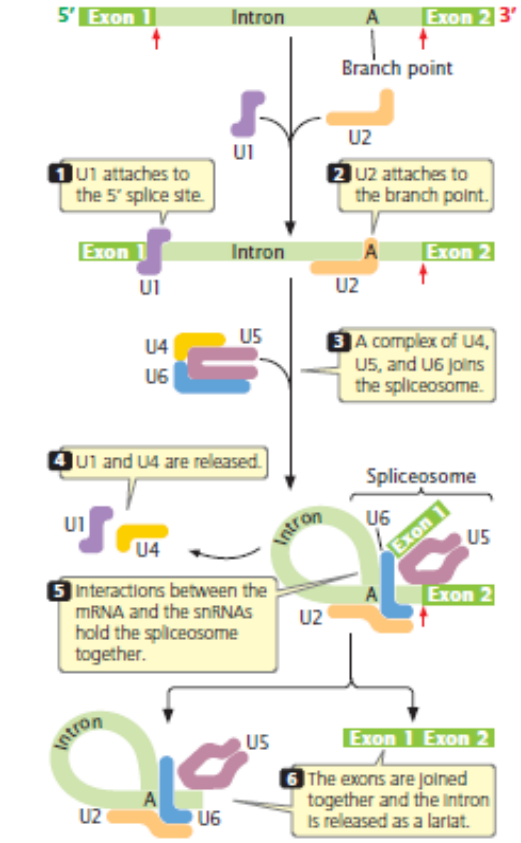
The RNA components are small nuclear RNAs ranging in length from 107 to 210 nucleotides; these snRNAs associate with proteins to form small ribonucleoprotein particles (snRNPs).

Each snRNP contains a single snRNA molecule and multiple proteins.

The spliceosome is composed of five snRNPs (U1, U2, U4, U5, and U6), and some proteins not associated with an snRNA.



14.9 The splicing of nuclear introns requires a two-step process.



14.10 RNA splicing takes place within the spliceosome. The spliceosome assembles sequentially.

The process of splicing Before splicing takes place, an intron lies between an upstream exon (exon 1) and a

downstream exon (exon 2), PremRNA is spliced in two distinct steps. In the first step of splicing, the pre-mRNA is cut at the 5′ splice site. This cut frees exon 1 from the intron, and the 5′ end of the intron attaches to the branch point; that is, the intron folds back on itself, forming a structure called a **lariat**. In this reaction, the guanine nucleotide in the consensus sequence at the 5′splice site bonds with the adenine nucleotide at the branch point through a transesterification reaction. The result is that the 5′ phosphate group of the guanine nucleotide is now attached to the 2′-OH group of the adenine nucleotide at the branch point.

In the second step of RNA splicing, a cut is made at the 3′ splice site and, simultaneously, the 3′ end of exon 1 becomes covalently attached (spliced) to the 5′ end of exon 2. The intron is released as a lariat. The intron becomes linear when the bond breaks at the branch point and is then rapidly degraded by nuclear enzymes. The mature mRNA consisting of the exons spliced together is exported to the cytoplasm, where it is translated. These splicing reactions take place within the spliceosome, which assembles on the pre-mRNA in a step-bystep fashion and carries out the splicing reactions. A key feature of the process is a series of interactions between the mRNA and

the snRNAs and between different snRNAs. These interactions depend on complementary base pairing between the different RNA molecules and bring the essential components of the pre-mRNA transcript and the spliceosome close together, which make splicing possible.

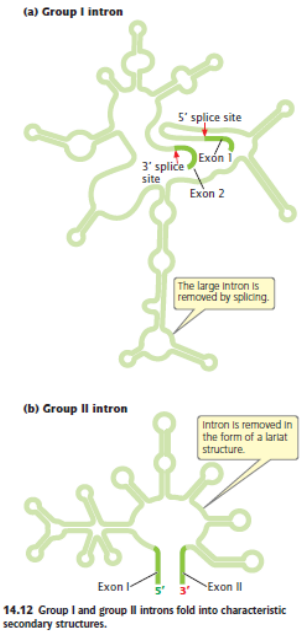
Self-splicing introns

Some introns are self-splicing, meaning that they possess the ability to remove themselves from an RNA molecule. These self-splicing introns fall into two major categories.

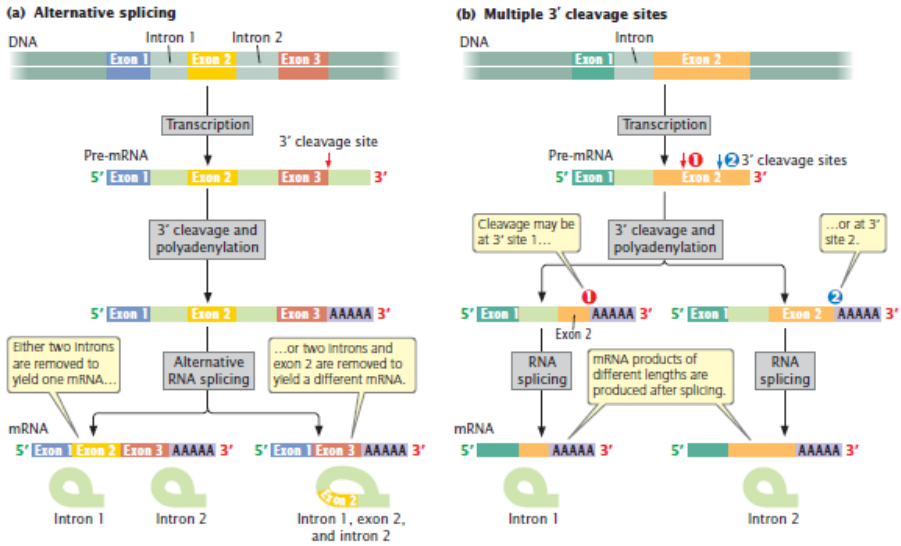
Group I introns are found in a variety of genes, including some rRNA genes in protists, some mitochondrial genes in fungi, and even some bacteriophage genes. Although the lengths of group I introns vary, all of them fold into a common secondary structure with nine looped stems, which are necessary for splicing.

Group II introns, present in some mitochondrial genes, also have the ability to self-splice. All group II introns also fold into secondary structures. The splicing of group II introns is accomplished by a mechanism that has some similarities to the spliceosomal-mediated splicing of nuclear genes, and splicing generates a lariat structure.

Because of these similarities, group II introns and nuclear pre-mRNA introns have been suggested to be evolutionarily related; perhaps the nuclear introns evolved from self-splicing group II introns and later adopted the proteins and snRNAs of the spliceosome to carry out the splicing reaction.



14.12 Group I and group II introns fold into characteristic secondary structures.



14.13 Eukaryotic cells have alternative pathways for processing pre-mRNA. (a) With alternative splicing, pre-mRNA can be spliced in different ways to produce different mRNAs. (b) With multiple 3′ cleavage sites, there are two or more potential sites for cleavage and polyadenylation; use of the different sites produces mRNAs of different lengths.

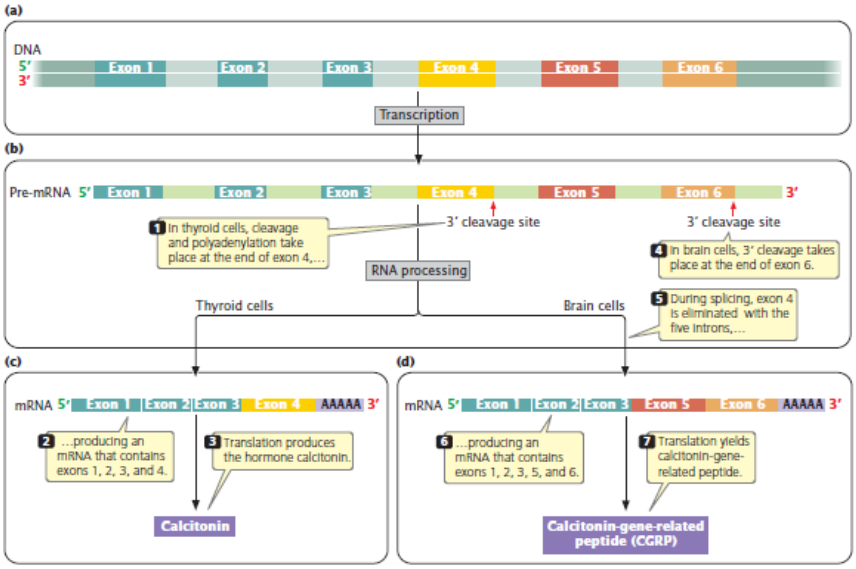
Alternative Processing Pathways

A finding that complicates the view of a gene as a sequence of nucleotides that specifies the amino acid sequence of a protein (see section on The Concept of the Gene Revisited) is the existence of **alternative processing pathways**. In these pathways, a single pre-mRNA is processed in different ways to produce alternative types of mRNA, resulting in the production of different proteins from the same DNA sequence.

One type of alternative processing is **alternative splicing,** in which the same pre-mRNA can be spliced in more than one way to yield multiple mRNAs that are translated into different amino acid sequences and thus different proteins. Another type of alternative processing requires the use of **multiple 3′ cleavage sites**, where two or more potential sites for cleavage and polyadenylation are present in the pre-mRNA. In the example, cleavage at the first site produces a relatively short mRNA compared with the mRNA produced through cleavage at the second site. The use of an alternative cleavage site may or may not produce a different protein, depending on whether the position of the site is before or after the termination codon.

Both alternative splicing and multiple 3′ cleavage sites can exist in the same pre-mRNA transcript. An example is seen in the mammalian gene that encodes calcitonin; this gene contains six exons and five introns. The entire gene is transcribed into pre-mRNA.

There are two possible 3′ cleavage sites. In cells of the thyroid gland, 3′ cleavage and polyadenylation take place after the fourth exon to produce a mature mRNA consisting of exons 1, 2, 3, and 4. This mRNA is translated into the hormone calcitonin.



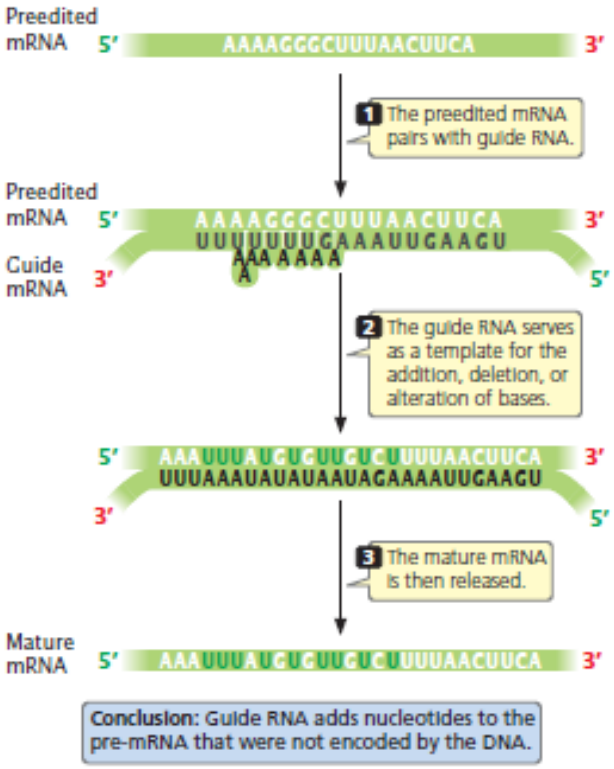
14.14 Pre-mRNA encoded by the gene for calcitonin undergoes alternative processing.

RNA Editing

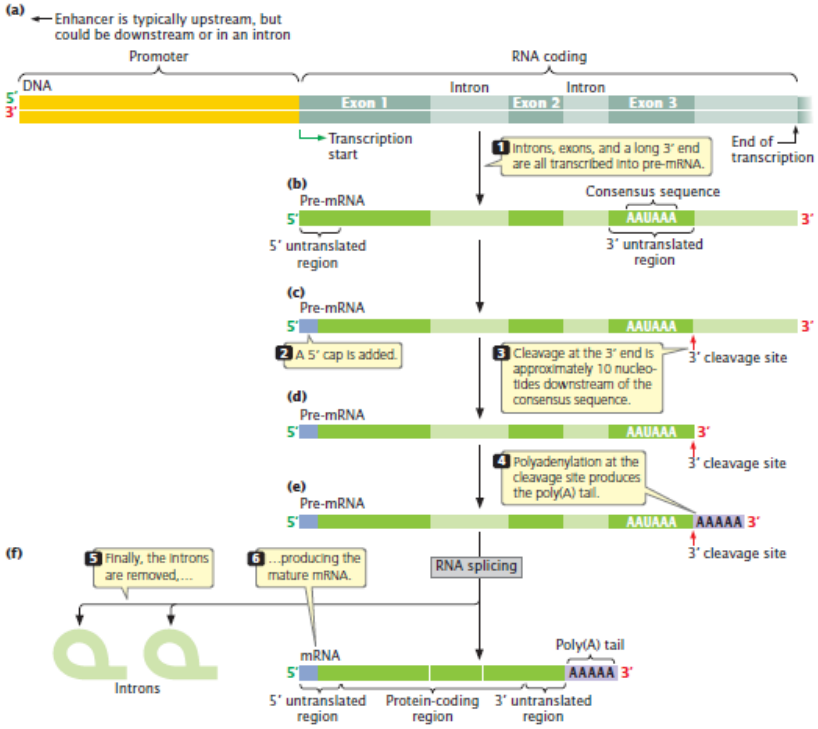
In humans, for example, a gene is transcribed into mRNA that encodes a lipid-transporting polypeptide called apolipoprotein-B100, which has 4563 amino acids and is synthesized in liver cells.

A truncated form of the protein called apolipoprotein-B48—with only 2153 amino acids—is synthesized in intestinal cells through editing of the apolipoprotein-B100 RNA.

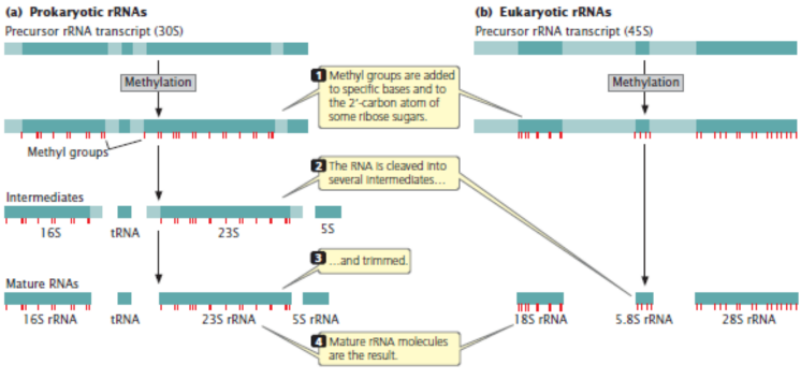
In this editing, an enzyme deaminates a cytosine base, converting it into uracil. This conversion changes a codon that specifies the amino acid glutamine into a stop codon that prematurely terminates translation, resulting in the shortened protein.



14.16 RNA editing is carried out by guide RNAs. The guide mRNA has sequences that are partly complementary to those of the preedited mRNA and pairs with it. After pairing, the mRNA undergoes cleavage and new nucleotides are added, with sequences in the gRNA serving as a template. The ends of the mRNA are then joined together.



14.17 Mature eukaryotic mRNA is produced when pre-mRNA is transcribed and undergoes several types of processing.



14.22 Ribosomal RNA is processed after transcription. Prokaryotic rRNA (a) and eukaryotic rRNA (b) are produced from precursor RNA transcripts that are methylated, cleaved, and processed to produce mature rRNAs. Eukaryotic 5S rRNA is transcribed separately from a different gene.

RNA interference (RNAi) is a powerful and precise mechanism used by eukaryotic cells to limit the invasion of foreign genes (from viruses and transposons) and to censor the expression of their own genes.

RNA interference is triggered by double-stranded RNA molecules, which may arise in several ways:

by the transcription of inverted repeats into an RNA molecule that then base pairs with itself to form double-stranded RNA;

by the simultaneous transcription of two different RNA molecules that are complementary to one another and that pair, forming double-stranded RNA;

or by infection by viruses that make double-stranded RNA. T

these double-stranded RNA molecules are chopped up by an enzyme appropriately called Dicer, resulting in tiny RNA molecules that are unwound to produce siRNAs and miRNAs (see Figure 14.23).

Some geneticists speculate that RNA interference evolved as a defense mechanism against RNA viruses and transposable elements that move through RNA intermediates (see Chapter 11); indeed, some have called RNAi the immune system of the genome.

However, RNA interference is also responsible for regulating a number of key genetic and developmental processes, including changes in chromatin structure, translation, cell fate and proliferation, and cell death.

Geneticists also use the RNAi machinery as an effective tool for blocking the expression of specific genes

Types of Small RNAs

Two abundant classes of RNA molecules that function in RNA interference in eukaryotes are small interfering RNAs and microRNAs.

They have a number of features in common and their functions overlap considerably. Both are about 22 nucleotides long.

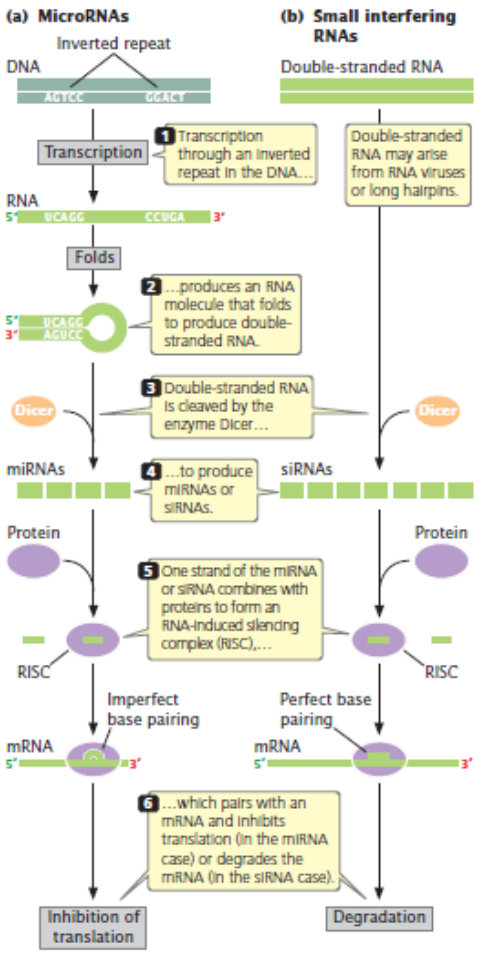
Small interfering RNAs arise from the cleavage of mRNAs,RNA transposons, and RNA viruses. Some miRNAs are cleaved from RNA molecules transcribed from sequences that encode miRNA only, but others are encoded in the introns and exons of mRNAs.

Each miRNA is cleaved from a single-stranded RNA precursor that forms small hairpins, whereas multiple siRNAs are produced from the cleavage of an RNA duplex consisting of two different RNA molecules.

Usually, siRNAs have exact complementarity with their target mRNA or DNA sequences and suppress gene expression by degrading mRNA or inhibiting transcription, whereas miRNAs often have limited complementarity with their target mRNAs and often suppress gene expression by inhibiting translation.

Finally, miRNAs usually silence genes that are distinct from those from which the miRNAs were transcribed, whereas siRNAs typically silence the genes from which the siRNAs were transcribed.

Note, however, that these differences between siRNAs and miRNAs are not hard and fast, and scientists are increasingly finding small RNAs that exhibit characteristics of both. For example, some miRNAs have exact complementarity with mRNA sequences and cleave these sequences, characteristics that are usually associated with siRNAs.



14.23 Small interfering RNAs and microRNAs are produced from double-stranded RNAs.

Table 14.5 Differences between siRNAs, miRNAs, and piRNAs

Feature	siRNA	miRNA	piRNA
Origin	mRNA, transposon, or virus	RNA transcribed from distinct gene	Transposons
Cleavage of	RNA duplex or single-stranded RNA that forms long hairpins	Single-stranded RNA that forms short hairpins of double-stranded RNA	Single-stranded RNA from transposons
Size	21–25 nucleotides	21–25 nucleotides	24–31 nucleotides
Action	Degradation of mRNA, inhibition of transcription, chromatin modification	Degradation of mRNA, inhibition of translation, chromatin modification	Degradation of RNA, chromatin modification
Target	Genes from which they were transcribed	Genes other than those from which they were transcribed	Transposons

Processing and Function of MicroRNAs

MicroRNAs have been found in all eukaryotic organisms examined to date, as well as viruses; they control the expression of genes taking part in many biologicalprocesses, including growth, development, and metabolism.

Humans have more than 450 distinct miRNAs; scientists estimate that more than one-third of all human genes are regulated by miRNAs.

Most miRNA genes are found in regions of noncoding DNA or within the introns of other genes.

The genes that encode miRNAs are transcribed into longer precursors, called primary miRNA (pri-miRNA), that range from several hundred to several thousand nucleotides in length.

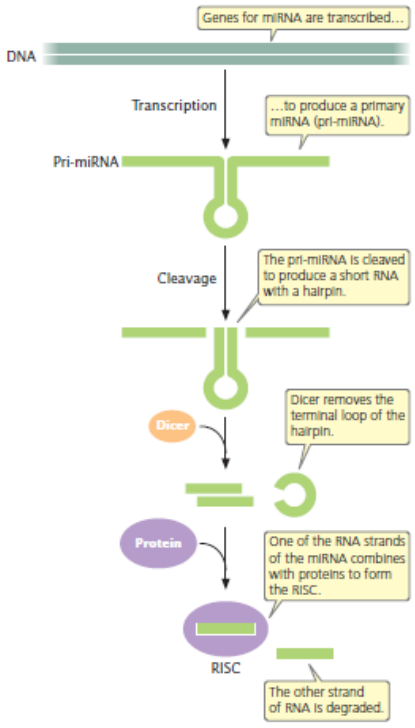
The pri-miRNA is then cleaved into one or more smaller RNA molecules with a hairpin.

Dicer (endoribonuclease Dicer or helicase with RNase motif) binds to this hairpin structure and removes the terminal loop. One of the miRNA strands is incorporated into the RISC (RNA induced Silencing Complex); the other strand is released and degraded.

The RISC attaches to a complementary sequence on the mRNA, usually in the 3′ untranslated region of the mRNA.

The region of close complementarity, called the seed region, is quite short, usually only about seven nucleotides long.

Because the seed sequence is so short, each miRNA can potentially pair with sequences on hundreds of different mRNAs. Furthermore, a single mRNA molecule may possess multiple miRNA-binding sites. The inhibition of translation may require binding



14.24 MicroRNAs are cleaved from larger precursors (pri-miRNAs).

INHIBITORS OF TRANSCRIPTION

Rifampicin- binds with Beta subunit of prokaryotic RNA polymerase,

• It is an inhibitor of prokaryotic transcription initiation.

• It binds only to bacterial RNA polymerase but not to eukaryotic RNA polymerases.

• Therefore, Rifampicin is a powerful drug for treatment of bacterial infections.

• Used for the treatment of tuberculosis and leprosy

Actinomycin D

• Actinomycin D- Intercalates with DNA strands

• Actinomycins inhibit both DNA synthesis and RNA synthesis by blocking chain elongation.

• They interact with G·C base pairs as they require the 2-amino group of guanine for binding.

• Actinomycins are used as anticancer Drugs

Mitomycin

• Mitomycin- Intercalates with DNA strands

• blocks transcription,

• used as anticancer drug

Alpha amanitin

• Alpha amanitin is a molecule made from the “death cap” mushroom and is a known potent inhibitor RNA polymerase.

• One single mushroom could very easily lead to a fast death in 10 days.

• The mechanism of action is that alpha amanitin inhibits RNA polymerase –II at both the initiation and elongation states of transcription.

REFERENCES

GENETICS: A conceptual approach, 4th Edition, Benjamin A. Pierce, W. H. Freeman and company England; 2006

Cell Biology, Genetics, Molecular Biology, Evolution and Ecology, P.S. Verma, V.K. Agarwal, S. Chand & Company Ltd, 2005

Principles of Molecular Biology, Veer Bala Rastogi, Medtech, 2016

Molecular Biology of the Cell. 4th edition**.** Alberts B, Johnson A, Lewis J, et al. New York: Garland Science; 2002.

www.namrata.com

http://highered.mcgrawhill. com/olcweb/cgi/pluginpop.cgi?it=swf:: 535::535::/sites/dl/free/ 0072437316/120077/bio30.swf::How%20Spliceosomes%20Process%20RNA

SATHYABAMA
INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)
Accredited "A" Grade by NAAC | 12B Status by UGC | Approved by AICTE
www.sathyabama.ac.in

SCHOOL OF BIO AND CHEMICAL ENGINEERING

DEPARTMENT OF BIOTECHNOLOGY

SBTA1301 – MOLECULAR BIOLOGY AND GENETICS

SBTA1301	MOLECULAR BIOLOGY AND GENETICS	L	T	P	Credit	Total Marks
		3	*	0	3	100

COURSE OBJECTIVES

- The course aims to give an understanding on the fundamentals of conventional genetics and the intricate molecular mechanisms of heredity and variations.
- To understand storage of genetic information and its translation at molecular level in prokaryotic and eukaryotic systems.

UNIT 1 CLASSICAL GENETICS **9 Hrs.**

Fundamental principles of genetics- Mendel's principles and experiments, gene interaction, multiple alleles, complementation, linkage, sex linked, sex limited and sex influenced inheritance; Chromosomes basis of heredity- extra-chromosomal inheritance; Linkage and crossing over; Hardy-Weinberg equilibrium, Extensions of Hardy- Weinberg equilibrium.

UNIT 2 STRUCTURE AND PROPERTIES OF NUCLEIC ACIDS **9 Hrs.**

Introduction to nucleic acids: Evidence for DNA&RNA as a genetic material; Structure and physicochemical properties of elements in DNA and RNA, Biological significance of differences in DNA and RNA. Primary structure of DNA: Chemical and structural qualities of 3',5'-Phosphodiester bond. Secondary Structure of DNA: Watson & Crick model, Chargaff's rule; DNA replication- Overview of differences in prokaryotic and eukaryotic DNA replication, D-loop and rolling circle mode of replication, Telomere replication in eukaryotes; Okazaki fragments, Fidelity of DNA replication, Inhibitors of DNA replication, DNA repair- Mutagens, DNA mutations and various types of repair mechanisms.

UNIT 3 TRANSCRIPTION **9 Hrs.**

Central Dogma in molecular biology -Structure and function of mRNA, rRNA tRNA and micro RNAs. Characteristics of promoter and enhancer sequences. RNA synthesis: Initiation, elongation and termination of RNA synthesis, Proteins of RNA synthesis, Fidelity of RNA synthesis, Inhibitors of transcription, Differences in prokaryotic and eukaryotic transcription. Basic concepts in RNA world: RNA processing: 5'-Capping, Splicing-Alternative splicing, Poly 'A' tail addition and base modification.

UNIT 4 TRANSLATION **9 Hrs.**

Introduction to Genetic code: Elucidation of genetic code, Codon degeneracy, Wobble hypothesis and its importance, Prokaryotic and eukaryotic ribosomes. Steps in translation: Initiation, Elongation and termination of protein synthesis. Inhibitors of protein synthesis. Post-translational modifications and its importance.

UNIT 5 REGULATION OF GENE EXPRESSION **9 Hrs.**

Organization of genes in prokaryotic and eukaryotic chromosomes- operon concept; Gene expression and regulation- Hierarchical levels of gene regulation, Prokaryotic gene regulation –lac and trp operon, Eukaryotic gene regulation- gene silencing.

Max.45 Hrs.

COURSE OUTCOMES

On completion of the course, student will be able to

- CO1 - Explain the foundations of Mendelian genetics and chromosomal theory and apply these, with appropriate terminology, to contemporary concepts in genetics.
- CO2 - Emphasize the molecular mechanism of DNA replication and repair in various organisms
- CO3 - Explain the properties of genetic materials and storage and processing of genetic information.
- CO4 - Analyze the processes of transcription and translation in both prokaryotes and eukaryotes at molecular level.
- CO5 - Understand the redundant and universal qualities of the genetic code and how it is used to determine the amino acid sequence of a polypeptide.
- CO6 - Compare the mechanisms of gene regulation in prokaryotes and eukaryotes.

TEXT / REFERENCE BOOKS

1. Lewin B., Genes XI, International Edition, Jocelyn Krebs, Stephen Kilpatrick and Elliott Goldstein, Jones & Bartlett Learning, 2017, ISBN 978-1-4496-5985-1
2. Tropp, Burton E., Molecular Biology: Genes to Proteins, 3rd Edition, Jones and Bartlett, 2008.
3. Glick B.R. and Pasternak J.J., Molecular Biotechnology: Principles and Applications of Recombinant DNA, 4th Edition. ASM, 2010

UNIT – IV – TRANSLATION

GENETIC CODE

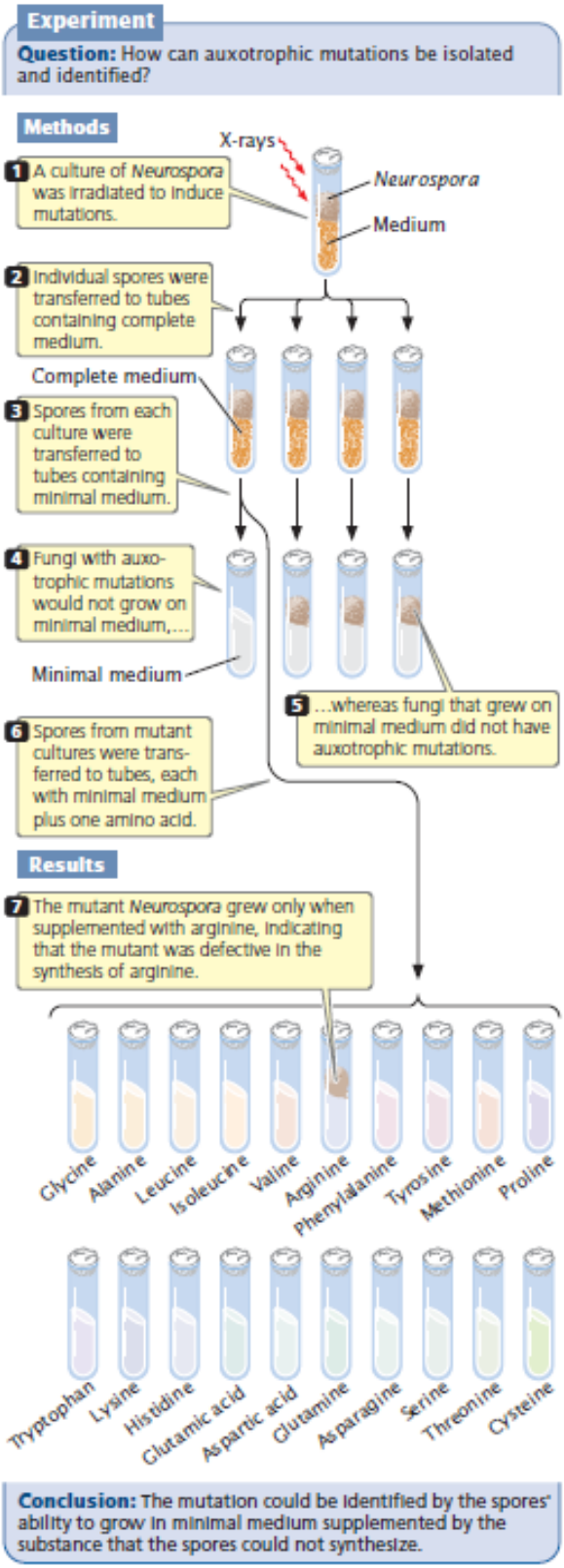
The first person to suggest the existence of a relation between genotype and proteins was English physician Archibald Garrod. In 1908, Garrod correctly proposed that genes encode enzymes, but, unfortunately, his theory made little impression on his contemporaries. Not until the 1940s, when George Beadle and Edward Tatum examined the genetic basis of biochemical pathways in *Neurospora*, did the relation between genes and proteins become widely accepted.

The One Gene, One Enzyme Hypothesis

Beadle and Tatum used the bread mold *Neurospora* to study the biochemical results of mutations. *Neurospora* is easy to cultivate in the laboratory. The main vegetative part of the fungus is haploid, which allows the effects of recessive mutations to be easily observed.

Wild-type *Neurospora* grows on minimal medium, which contains only inorganic salts, nitrogen, a carbon source such as sucrose, and the vitamin biotin. The fungus can synthesize all the biological molecules that it needs from these basic compounds. However, mutations may arise that disrupt fungal growth by destroying the fungus's ability to synthesize one or more essential biological molecules. These nutritionally deficient mutants, termed auxotrophs, will not grow on minimal medium, but they can grow on medium that contains the substance that they are no longer able to synthesize.

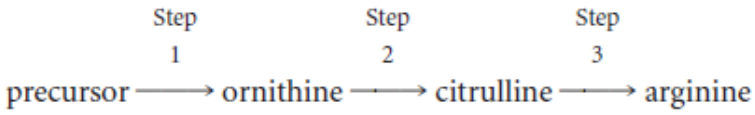
Beadle and Tatum first irradiated spores of *Neurospora* to induce mutations. After irradiation, they placed individual spores into different culture tubes containing complete medium (medium having all the biological substances needed for growth). Next, they transferred spores from each culture to tubes containing minimal medium. Fungi containing auxotrophic mutations grew on complete medium but would not grow on minimal medium, which allowed Beadle and Tatum to identify cultures that contained mutations.



After they had determined that a particular culture had an auxotrophic mutation, Beadle and Tatum set out to determine the specific *effect* of the mutation. They transferred spores of each mutant strain from complete medium to a series of tubes, each of which possessed minimal medium plus one of a

variety of essential biological molecules, such as an amino acid. If the spores in a tube grew, Beadle and Tatum were able to identify the added substance as the biological molecule whose synthesis had been affected by the mutation. For example, an auxotrophic mutant that would grow only on minimal medium to which arginine had been added must have possessed a mutation that disrupts the synthesis of arginine.

Adrian Srb and Norman H. Horowitz patiently applied this procedure to genetically dissect the multistep biochemical pathway of arginine synthesis. They first isolated a series of auxotrophic mutants whose growth required arginine. They then tested these mutants for their ability to grow on minimal medium supplemented with three compounds: ornithine, citrulline, and arginine. From the results, they were able to place the mutants into three groups on the basis of which of the substances allowed growth. Group I mutants grew on minimal medium supplemented with ornithine, citrulline, or arginine. Group II mutants grew on minimal medium supplemented with either arginine or citrulline but did not grow on medium supplemented only with ornithine. Finally, group III mutants grew only on medium supplemented with arginine. Biochemical pathway leading to the amino acid arginine has at least three steps:



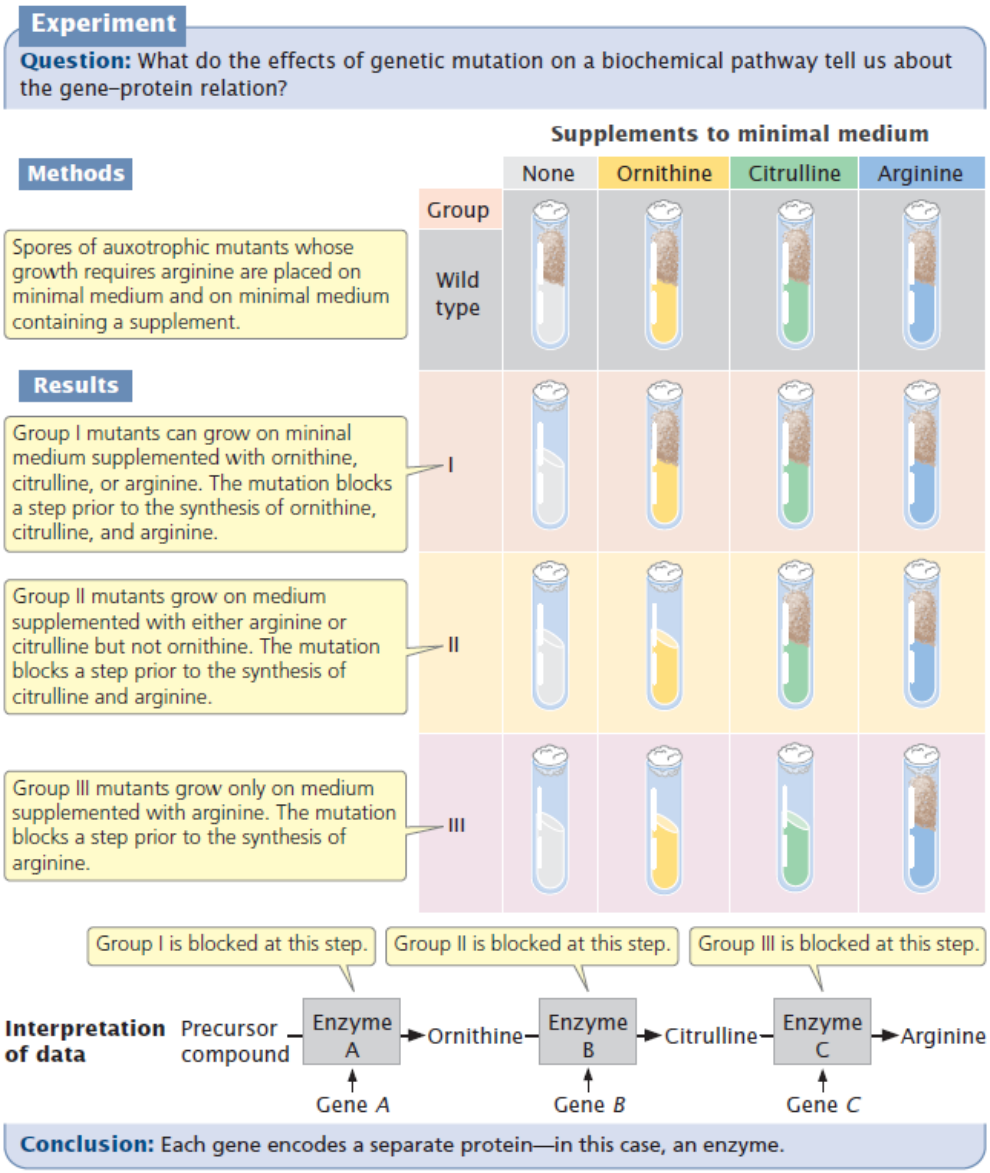
They concluded that the mutations in group I affect step 1 of this pathway, mutations in group II affect step 2, and

mutations in group III affect step 3. But how did they know that the order of the compounds in the biochemical pathway was correct?

Notice that, if step 1 is blocked by a mutation, then the addition of either ornithine or citrulline allows growth, because these compounds can still be converted into arginine (see Figure 15.3). Similarly, if step 2 is blocked, the addition of citrulline allows growth, but the addition of ornithine has no effect. If step 3 is blocked, the spores will grow only if arginine is added to the medium. The underlying principle is that an auxotrophic mutant cannot synthesize any compound that comes after the step blocked by a mutation.

Notice that, if step 1 is blocked by a mutation, then the addition of either ornithine or citrulline allows growth, because these compounds can still be converted into arginine. Similarly, if step 2 is blocked, the addition of citrulline allows growth, but the addition of ornithine has no effect. If step 3 is blocked,

the spores will grow only if arginine is added to the medium. The underlying principle is that an auxotrophic mutant cannot synthesize any compound that comes after the step blocked by a mutation.

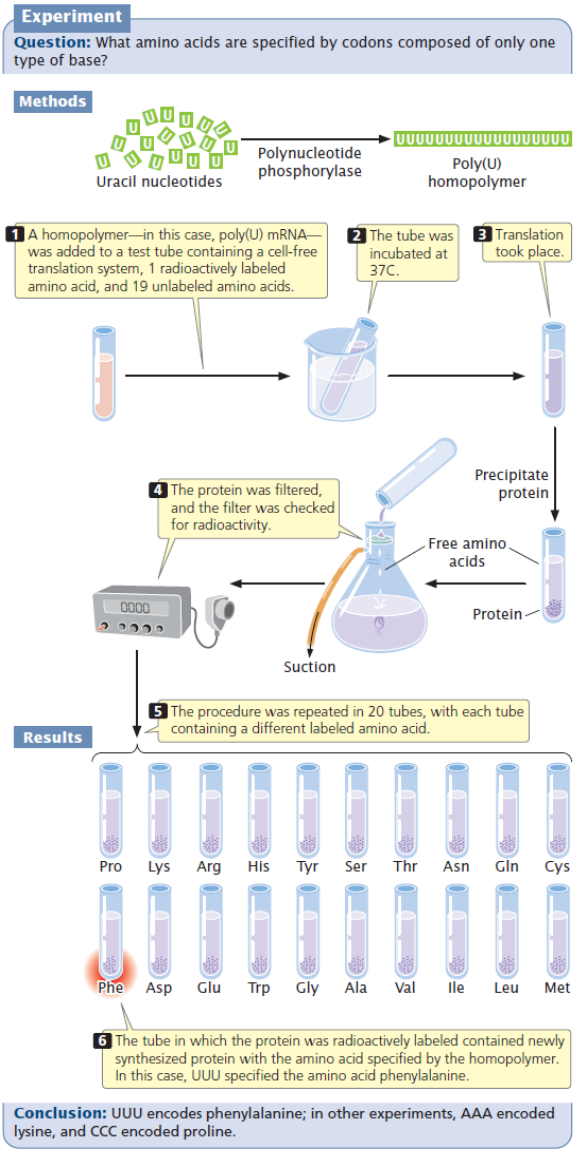


In 1953, James Watson and Francis Crick solved the structure of DNA and identified the base sequence as the carrier of genetic information. However, the way in which the base sequence of DNA specifies the amino acid sequences of proteins (the genetic code) was not immediately obvious and remained elusive for another 10 years.

One of the first questions about the genetic code to be addressed was: How many nucleotides are necessary to specify a single amino acid? This basic unit of the genetic code—the set of bases that encode a single amino acid—is a *codon* (see p. 380 in Chapter 14). Many early investigators recognized that codons must contain a minimum of three nucleotides. Each nucleotide position in

mRNA can be occupied by one of four bases: A, G, C, or U. If a codon consisted of a single nucleotide, only four different codons (A, G, C, and U) would be possible, which is not enough to encode the 20 different amino acids commonly found in proteins. If codons were made up of two nucleotides each (i.e., GU, AC, etc.), there would be 4 × 4 = 16 possible codons—still not enough to encode all 20 amino acids. With three nucleotides per codon, there are 4 × 4 × 4 = 64 possible codons, which is more than enough to specify 20 different amino acids.

Therefore, a *triplet code* requiring three nucleotides per codon is the most efficient way to encode all 20 amino acids. Using mutations in bacteriophage, Francis Crick and his colleagues confirmed in 1961 that the genetic code is indeed a triplet code.



15.8 Nirenberg and Matthaei developed a method for identifying the amino acid specified by a homopolymer.

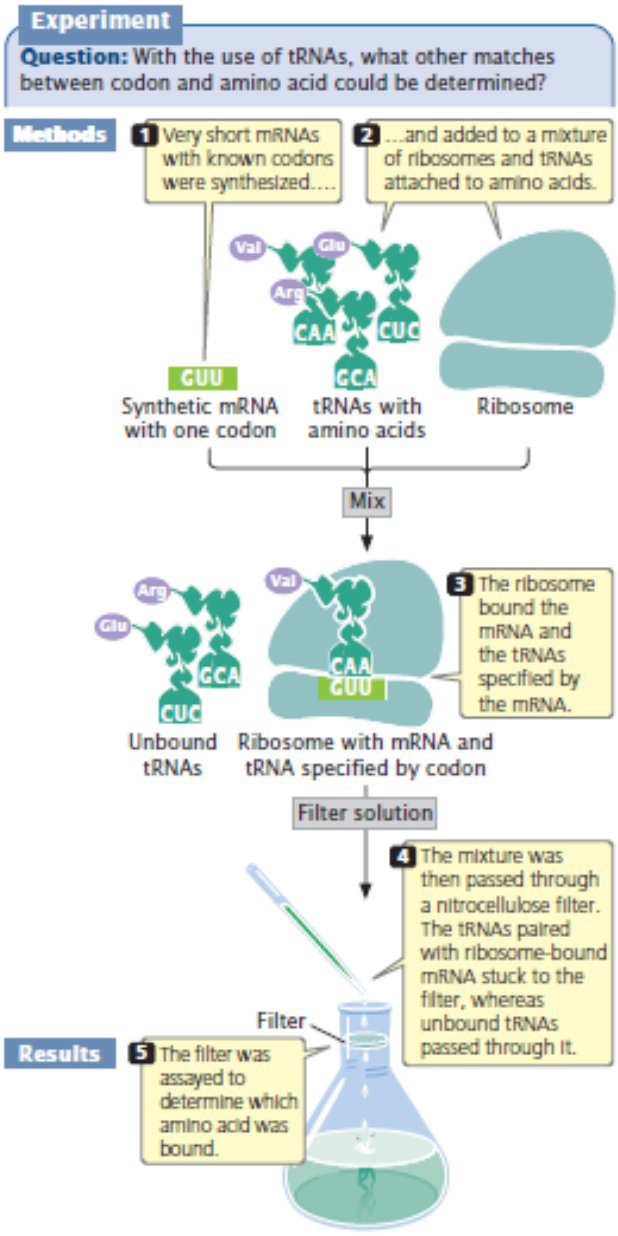
The use of homopolymers The first clues to the genetic code came in 1961, from the work of Marshall Nirenberg and Johann Heinrich Matthaei. These investigators created synthetic RNAs by using an enzyme called polynucleotide phosphorylase. Unlike RNA polymerase, polynucleotide phosphorylase does not require a template; it randomly links together any RNA nucleotides that happen to be available. The first synthetic mRNAs used by Nirenberg and Matthaei were homopolymers, RNA molecules consisting of a single type of nucleotide. For example, by adding polynucleotide phosphorylase to a solution of uracil nucleotides, they generated RNA molecules that consisted entirely of uracil nucleotides and thus contained only UUU codons. These poly(U) RNAs were then added to 20 tubes, each containing the components necessary for translation and all 20 amino acids. A different amino acid was radioactively labeled in each of the 20 tubes. Radioactive protein appeared in only one of the tubes—the one containing labeled phenylalanine. This result showed that the codon UUU specifies the amino acid phenylalanine. The results of similar experiments using poly(C) and poly(A) RNA demonstrated that CCC

The use of homopolymers The first clues to the genetic code came in 1961, from the work of Marshall Nirenberg and Johann Heinrich Matthaei. These investigators created synthetic RNAs by using an enzyme called polynucleotide phosphorylase. Unlike RNA polymerase, polynucleotide phosphorylase does not require a template; it randomly links together any RNA nucleotides that happen to be available. The first synthetic mRNAs used by Nirenberg and Matthaei were homopolymers, RNA molecules consisting of a single type of nucleotide. For example, by adding polynucleotide phosphorylase to a solution of uracil nucleotides, they generated RNA molecules that consisted entirely of uracil nucleotides and thus contained only UUU codons. These poly(U) RNAs were then added to 20 tubes, each containing the components necessary for translation and all 20 amino acids. A different amino acid was radioactively labeled in each of the 20 tubes. Radioactive protein appeared in only one of the tubes—the one containing labeled phenylalanine (see Figure 15.8). This result showed that the codon UUU specifies the amino acid phenylalanine. The results of similar experiments using poly(C) and poly(A) RNA demonstrated that CCC

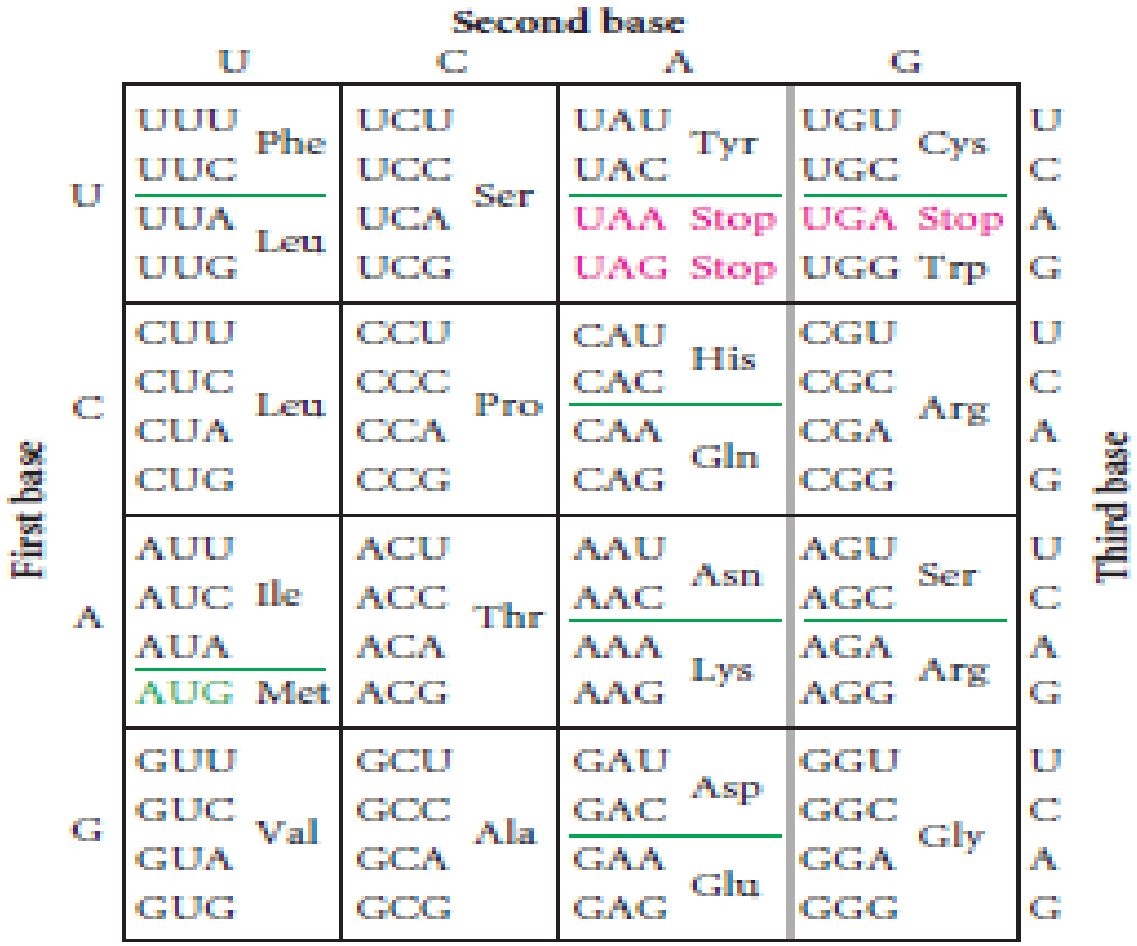
The use of random copolymers To gain information about additional codons, Nirenberg and his colleagues created synthetic RNAs containing two or three different bases.

Because polynucleotide phosphorylase incorporates nucleotides randomly, these RNAs contained random mixtures of the bases and are thus called random copolymers. For example, when adenine and cytosine nucleotides are mixed with polynucleotide phosphorylase, the RNA molecules produced have eight different codons: AAA, AAC, ACC, ACA, CAA, CCA, CAC, and CCC. These poly(AC)

RNAs produced proteins containing six different amino acids: asparagine, glutamine, histidine, lysine, proline, and threonine.



15.9 Nirenberg and Leder used ribosome-bound tRNAs to provide additional information about the genetic code.



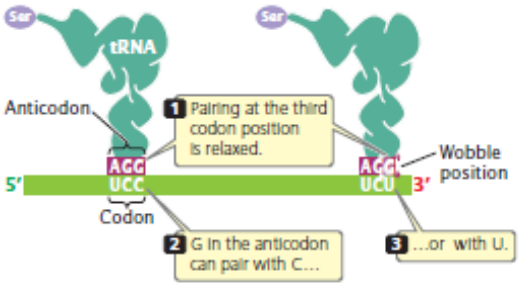
15.10 The genetic code consists of 64 codons. The amino acids specified by each codon are given in their three-letter abbreviation. The codons are written 5′⟶3′, as they appear in the mRNA. AUG is an initiation codon; UAA, UAG, and UGA are termination (stop) codons.

The Degeneracy of the Code

One amino acid is encoded by three consecutive nucleotides in mRNA, and each nucleotide can have one of four possible bases (A, G, C, and U) at each nucleotide position, thus permitting 43 = 64 possible codons. Three of these codons are stop codons, specifying the end of translation. Thus, 61 codons, called **sense codons**, encode amino acids. Because there are 61 sense codons and only 20 different amino acids commonly found in proteins, the code contains more information than is needed to specify the amino acids and is said to be a **degenerate code**.

The degeneracy of the genetic code means that amino acids may be specified by more than one codon. Only tryptophan and methionine are encoded by a single codon. Other amino acids are specified by two codons, and some, such as leucine, are specified by six different codons. Codons that specify the

same amino acid are said to be **synonymous**, just as synonymous words are different words that have the same meaning.



15.11 Wobble may exist in the pairing of a codon and anticodon. The mRNA and tRNA pair in an antiparallel fashion. Pairing at the first and second codon positions is in accord with the Watson-and-Crick pairing rules (A with U, G with C); however, pairing rules are relaxed at the third position of the codon, and G on the anticodon can pair with either U or C on the codon in this example.

The cells of most organisms possess from about 30 to 50 different tRNAs, and yet there are only 20 different amino acids in proteins. Thus, some amino acids are carried by more than one tRNA. Different tRNAs that accept the same amino acid but have different anticodons are called **isoaccepting tRNAs**. Some synonymous codons specify different isoacceptors.

Many synonymous codons differ only in the third position. For example, alanine is encoded by the codons GCU, GCC, GCA, and GCG, all of which begin with GC. When the codon on the mRNA and the anticodon of the tRNA join, the first (5′) base of the codon pairs with the third (3′) base of the anticodon, strictly according to Watson-and-Crick rules: A with U; C with G. Next, the middle bases of codon and anticodon pair, also strictly following the Watson-and-Crick rules. After these pairs have hydrogen bonded, the third bases pair weakly: there may be flexibility, or **wobble**, in their pairing.

the code is generally **nonoverlapping**. An overlapping code is one in which a single nucleotide may be included in more than one codon, as follows:

Nucleotide sequence A U A C G A G U C

Nonoverlapping code A U A | C G A | G U C
Ile Arg Val

Overlapping code A U A C G A G U
Ile
U A C
Tyr
A C G
Thr

Termination Codons

Three codons—UAA, UAG, and UGA—do not encode amino acids. These codons signal the end of the protein in both bacterial and eukaryotic cells and are called **stop codons, termination codons,** or **nonsense codons**. No tRNA molecules have anticodons that pair with termination codons.

The Universality of the Code

For many years the genetic code was assumed to be **universal**, meaning that each codon specifies the same amino acid in all organisms. We now know that the genetic code is almost, but not completely, universal; a few exceptions have been found. Most of these exceptions are termination codons, but there are a few cases in which one sense codon substitutes for another. Most exceptions are found in mitochondrial genes; a few nonuniversal codons have also been detected in the nuclear genes of protozoans and in bacterial DNA

Table 15.3 Some exceptions to the universal genetic code

Genome	Codon	Universal Code	Altered Code
Bacterial DNA			
Mycoplasma capricolum	UGA	Stop	Trp
Mitochondrial DNA			
Human	UGA	Stop	Trp
Human	AUA	Ile	Met
Human	AGA, AGG	Arg	Stop
Yeast	UGA	Stop	Trp
Trypanosomes	UGA	Stop	Trp
Plants	CGG	Arg	Trp
Nuclear DNA			
Tetrahymena	UAA	Stop	Gln
Paramecium	UAG	Stop	Gln

Characteristics of Genetic code

1. The genetic code consists of a sequence of nucleotides in DNA or RNA. There are four letters in the code, corresponding to the four bases—A, G, C, and U (T in DNA).

2. The genetic code is a triplet code. Each amino acid is encoded by a sequence of three consecutive nucleotides, called a codon.

3. The genetic code is degenerate; that is, of 64 codons, 61 codons encode only 20 amino acids in proteins (3 codons are termination codons). Some codons are synonymous, specifying the same amino acid.

4. Isoaccepting tRNAs are tRNAs with different anticodons that accept the same amino acid; wobble allows the anticodon on one type of tRNA to pair with more than one type of codon on mRNA.

5. The code is generally nonoverlapping; each nucleotide in an mRNA sequence belongs to a single reading frame.

6. The reading frame is set by an initiation codon, which is usually AUG.

7. When a reading frame has been set, codons are read as successive groups of three nucleotides.

8. Any one of three termination codons (UAA, UAG, and UGA) can signal the end of a protein; no amino acids are encoded by the termination codons.

9. The code is almost universal.

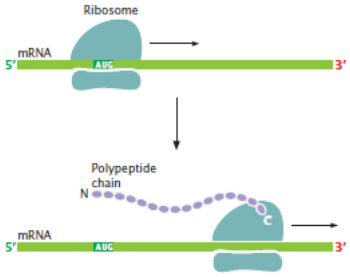
TRANSLATION

Amino Acids Are Assembled into a Protein Through the Mechanism of Translation

Now that we are familiar with the genetic code, we can begin to study the mechanism by which amino acids are assembled into proteins. Because more is known about translation in bacteria, we will focus primarily on bacterial translation. In most respects, eukaryotic translation is similar, although some significant differences will be noted as we proceed through the stages of translation. Translation takes place on ribosomes; indeed, ribosomes can be thought of as moving protein-synthesizing machines. Through a variety of techniques, a detailed view of the structure of the ribosome has been produced in recent years, which has greatly improved our understanding of the translational process. A ribosome attaches near the 5′ end of an mRNA strand and moves toward the 3′ end, translating the codons as it goes. Synthesis begins at the amino end of the protein, and the protein is elongated by the addition of new amino acids to the carboxyl end. Protein synthesis includes a series of RNA–RNA interactions: interactions between the mRNA and the rRNA that hold the mRNA in the ribosome, interactions between the codon on the mRNA and the anticodon on the tRNA, and interactions

between the tRNA and the rRNAs of the ribosome Protein synthesis can be conveniently divided into four stages: (1) tRNA charging, in which tRNAs bind to amino acids; (2) initiation, in which the

components necessary for translation are assembled at the ribosome; (3) elongation, in which amino acids are joined, one at a time, to the growing polypeptide chain; and (4) termination, in which protein synthesis halts at the termination codon and the translation components are released from the ribosome.

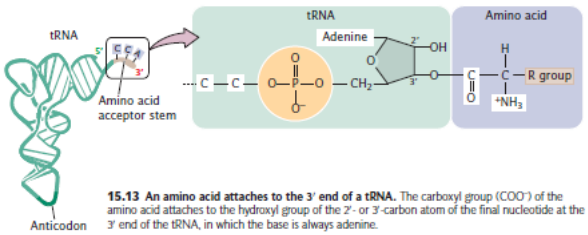


15.12 The translation of an mRNA molecule takes place on a ribosome. The letter N represents the amino end of the protein; C represents the carboxyl end.

The Binding of Amino Acids to Transfer RNAs

The first stage of translation is the binding of tRNA molecules to their appropriate amino acids, called tRNA charging. Each tRNA is specific for a particular amino acid. All tRNAs have the sequence CCA at the 3′ end, and the carboxyl group (COO–) of the amino acid is attached to the adenine nucleotide at the 3′ end of the tRNA. If each tRNA is specific for a particular amino acid but all amino acids are attached to the same nucleotide (A) at the 3′ end of a tRNA, how does a tRNA link up with its appropriate amino acid?

tRNA is a set of enzymes called **aminoacyl-tRNA synthe tases**. A cell has 20 different aminoacyl-tRNA synthetases, one for each of the 20 amino acids. Each synthetase recognizes a particular amino acid, as well as all the tRNAs that accept that amino acid. Recognition of the appropriate amino acid by a synthetase is based on the different sizes, charges, and R groups of the amino acids. The recognition of tRNAs by a synthetase depends on the differing nucleotide sequences of the tRNAs. Researchers have identified which nucleotides are important in recognition by altering different nucleotides in a particular tRNA and determining whether the altered tRNA is still recognized by its synthetase.



15.13 An amino acid attaches to the 3′ end of a tRNA. The carboxyl group (COO⁻) of the amino acid attaches to the hydroxyl group of the 2′- or 3′-carbon atom of the final nucleotide at the 3′ end of the tRNA, in which the base is always adenine.

The attachment of a tRNA to its appropriate amino acid, termed **tRNA charging**, requires energy, which is supplied by adenosine triphosphate (ATP):

amino acid + tRNA + ATP → aminoacyl-tRNA + AMP + PPi

This reaction takes place in two steps. To identify the resulting aminoacylated tRNA, we write the three-letter abbreviation for the amino acid in front of the tRNA; for example, the amino acid alanine (Ala) attaches to its tRNA (tRNAAla), giving rise to its aminoacyl-tRNA (Ala-tRNAAla). Errors in tRNA charging are rare; they occur in only about 1 in 10,000 to 1 in 100,000 reactions. This fidelity is due in part to the presence of editing (proofreading) activity
in many of the synthetases. Editing activity detects and removes incorrectly paired amino acids from the tRNAs. Some antifungal chemical agents work by trapping tRNAs in the editing site of the enzyme, preventing their release and thus inhibiting the process of translation in the fungi.

The Initiation of Translation

The second stage in the process of protein synthesis is initiation. At this stage, all the components necessary for protein synthesis assemble: (1) mRNA; (2) the small and large subunits of the ribosome; (3) a set of three proteins called initiation factors; (4) initiator tRNA with *N*-formylmethionine attached (fMet-tRNAfMet); and (5) guanosine triphosphate (GTP). Initiation comprises three major steps. First, mRNA binds to the small subunit of the ribosome. Second, initiator tRNA binds to the mRNA through base pairing between the codon and the
anticodon. Third, the large ribosome joins the initiation complex. Let's look at each of these steps more closely.

Initiation in bacteria The functional ribosome of bacteria exists as two subunits, the small 30S subunit and the large 50S subunit. An mRNA molecule can bind to the small ribosome subunit only when the subunits are separate. **Initiation factor 3** (IF-3) binds to the small subunit of the ribosome and prevents the large subunit from binding during initiation . Another factor, **initiation factor 1** (IF-1), enhances the disassociation of the
large and small ribosomal subunits.

Key sequences on the mRNA required for ribosome binding have been identified in experiments designed to allow the ribosome to bind to mRNA but not proceed with protein synthesis; the ribosome is thereby stalled at the initiation site. After the ribosome is allowed to attach to the mRNA, ribonuclease is added, which degrades all the mRNA except the region covered by the ribosome. The intact mRNA can be separated from the ribosome and studied. The sequence covered by the ribosome during initiation is from 30 to 40 nucleotides long and includes the AUG initiation codon. Within the ribosome-binding site is the Shine–Dalgarno consensus sequence, which is complementary to a sequence of nucleotides at the 3′ end of 16S rRNA (part of the small subunit of the ribosome). During initiation, the nucleotide in the Shine–Dalgarno sequence pair with their complementary nucleotides in the 16S rRNA, allowing the small subunit of the ribosome to attach to the mRNA and positioning the ribosome directly over the initiation codon.

Next, the initiator tRNA, fMet-tRNAfMet, attaches to the initiation codon. This step requires **initiation factor 2** (IF-2), which forms a complex with GTP. At this point, the initiation complex consists of (1) the small subunit of the ribosome; (2) the mRNA; (3) the initiator tRNA with its amino acid (fMet-tRNAfMet); (4) one molecule of GTP; and (5) several initiation factors. These components are collectively known as the **30S initiation complex**. In the final step of initiation, IF-3 dissociates from the small subunit, allowing the large subunit of the ribosome to join the initiation complex. The molecule of GTP (provided by IF-2) is hydrolyzed to guanosine diphosphate (GDP), and the initiation factors dissociate. When the large subunit has joined the initiation complex, the complex is called the **70S initiation complex**.

Initiation in eukaryotes

Similar events take place in the initiation of translation in eukaryotic cells, but there are some important differences. In bacterial cells, sequences in 16S rRNA of the small subunit of the ribosome bind to the Shine–Dalgarno sequence in mRNA. No analogous consensus sequence exists in eukaryotic mRNA. Instead, the cap at the 5′ end of eukaryotic mRNA plays a critical role in the initiation of translation. In a series of steps, the small subunit of the eukaryotic ribosome, initiation factors, and the initiator tRNA with its amino acid (Met-tRNAi Met) form an initiation complex that recognizes the cap and binds there.

The initiation complex then moves along (scans) the mRNA until it locates the first AUG codon. The identification of the start codon is facilitated by the presence of a consensus sequence (called the Kozak sequence) that surrounds the start codon:

5 –ACCAUGG–3

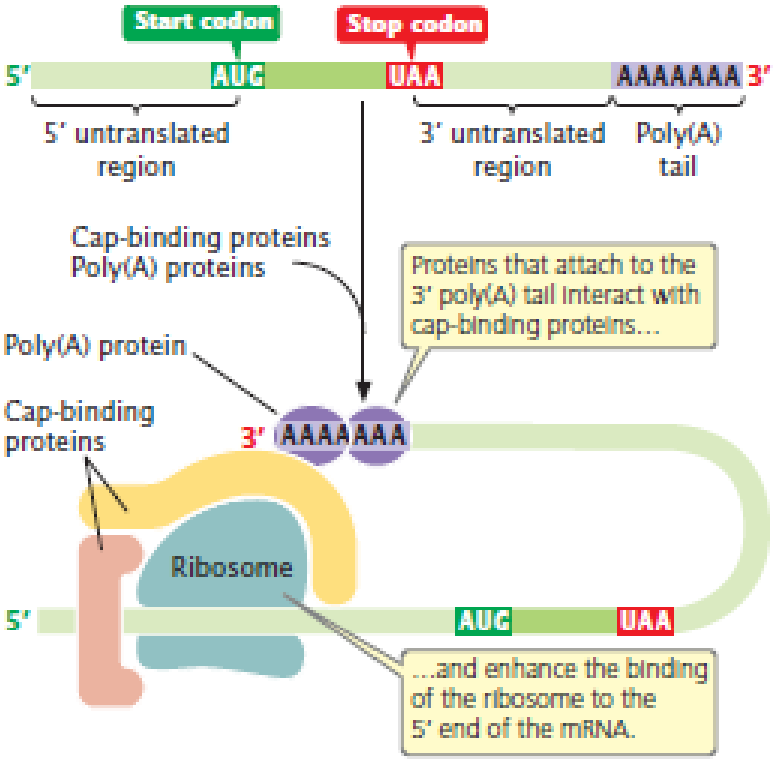
Kozak sequence

Start codon

Another important difference is that eukaryotic initiation requires at least seven initiation factors. Some factors keep the ribosomal subunits separated, just as IF-3 does in bacterial cells. Others recognize the 5′ cap on mRNA and allow the small subunit of the ribosome to bind there. Still others possess RNA helicase activity, which is used to unwind secondary structures that may exist in the 5′ untranslated region of mRNA, allowing the small subunit to move down the mRNA until the initiation codon is reached.

Other initiation factors help bring Met-tRNAi Met to the initiation complex. The poly(A) tail at the 3′ end of eukaryotic mRNA also plays a role in the initiation of translation. During initiation, proteins that attach to the poly(A) tail interact with proteins that bind to the 5′ cap, enhancing the binding of the small subunit of the ribosome to the 5′ end of the mRNA. This interaction indicates that the 3′ end of mRNA bends over and associates with the 5′ cap during the initiation of translation, forming

a circular structure known as the closed loop. A few eukaryotic mRNAs contain internal ribosome entry sites, where ribosomes can bind directly without first attaching to the 5′ cap.



15.18 The poly(A) tail of eukaryotic mRNA plays a role in the initiation of translation.

Elongation

The next stage in protein synthesis is elongation, in which amino acids are joined to create a polypeptide chain. Elongation requires (1) the 70S complex just described; (2) tRNAs charged with their amino acids; (3) several elongation factors; and (4) GTP.

A ribosome has three sites that can be occupied by tRNAs; the **aminoacyl**, or **A**, **site**, the **peptidyl**, or **P**, **site**, and the **exit**, or **E**, **site**. The initiator tRNA immediately occupies the P site (the only site to which the fMet-tRNAfMet is capable of binding), but all other tRNAs first enter the A site. After initiation, the ribosome is attached to the mRNA, and fMet-tRNAfMet is positioned over the AUG start codon in the P site; the adjacent A site is unoccupied (see Figure 15.19a).

Elongation takes place in three steps. In the first step, a charged tRNA binds to the A site. This binding takes place when **elongation factor Tu** (EF-Tu) joins with GTP and then with a charged tRNA to form a three-part complex. This complex enters the A site of the ribosome, where the anticodon on the tRNA pairs with the codon on the mRNA. After the charged tRNA is in the A site, GTP is cleaved to GDP, and the EF-Tu–GDP complex is released. **Elongation factor Ts** (EF-Ts)

regenerates EF-Tu–GDP to EF-Tu–GTP. In eukaryotic cells, a similar set of reactions delivers the charged tRNA to the A site.

The second step of elongation is the formation of a peptide bond between the amino acids that are attached to tRNAs in the P and A sites. The formation of this peptide bond releases the amino acid in the P site from its tRNA. Evidence indicates that the catalytic activity is a property of the ribosomal RNA in the large subunit of the ribosome; this rRNA acts as a ribozyme.

The third step in elongation is **translocation**, the movement of the ribosome down the mRNA in the 5′→3′ direction. This step positions the ribosome over the hydrolysis of GTP to GDP. Because the tRNAs in the P and A sites are still attached to the mRNA through codon– anticodon pairing, they do not move with the ribosome as it translocates. Consequently, the ribosome shifts so that the tRNA that previously occupied the P site now occupies the E site, from which it moves into the cytoplasm where it can be recharged with another amino acid. Translocation also causes the tRNA that occupied the A site (which is attached to the growing polypeptide chain) to be in the P site, leaving the A site open. Thus, the progress of each tRNA through the ribosome in the course of elongation can be summarized as follows:

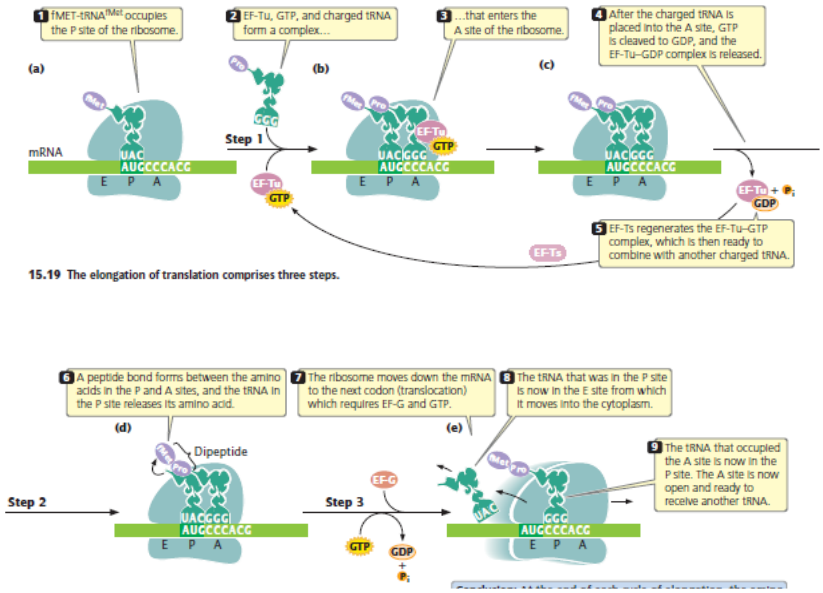
cytoplasm → A site → P site → E site → cytoplasm.

As discussed earlier, the initiator tRNA is an exception: it attaches directly to the P site and never occupies the A site. After translocation, the A site of the ribosome is empty and ready to receive the tRNA specified by the next codon. The elongation cycle (see Figure 15.19b through e) repeats itself: a charged tRNA and its amino acid occupy the A site, a peptide bond is formed between the amino acids in the A and P sites, and the ribosome translocates to the next codon. Throughout the cycle, the polypeptide chain remains attached to the tRNA in the P site. Messenger RNAs, although single stranded, often contain secondary structures formed by pairing of complementary bases on different parts of the mRNA. As the ribosome moves along the mRNA, these secondary structures are unwound by helicase activity located in the small subunit of the ribosome.

Recently, researchers have developed methods for following a single ribosome as it translates individual codons of an mRNA molecule. These studies revealed that translation does not take place in a smooth continuous fashion. Each translocation step typically requires less than a tenth of a second, but sometimes there are distinct pauses, often lasting a few seconds, between each translocation event when the ribosome moves from one codon to another. Thus, translation takes place in a series of quick translocations interrupted by brief pauses. In addition to the short pauses between translocation events, translation may be interrupted by longer pauses—lasting from 1 to 2 minutes—that may play a role in regulating the process of translation. Elongation in eukaryotic cells

takes place in a similar manner. Eukaryotes possess at least three elongation factors, one of which also acts in initiation and termination. Another

of the elongation factors used in eukaryotes, called elongation factor 2 (EF-2) is the target of a toxin produced by bacteria that causes diphtheria, a disease that, until recently, was a leading killer of children. The diphtheria toxin inhibits EF-2, preventing translocation of the ribosome along the mRNA, and protein synthesis ceases.

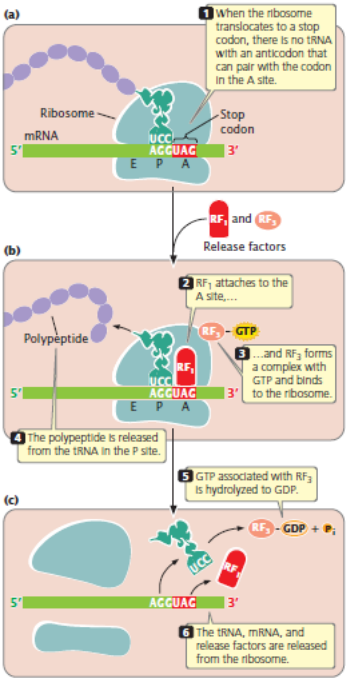


15.19 The elongation of translation comprises three steps.

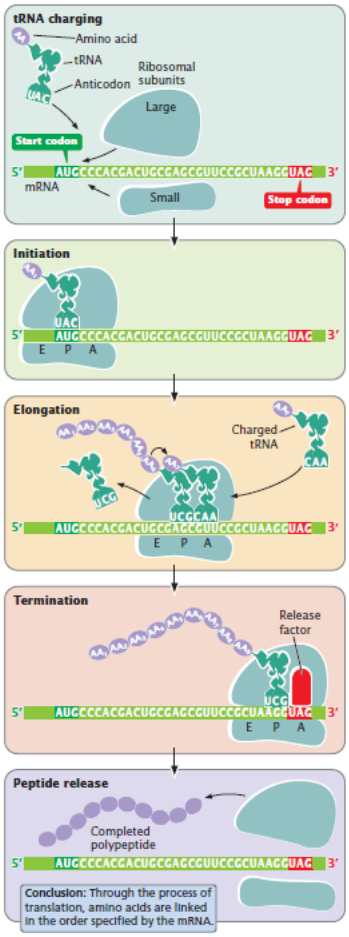
Termination

Protein synthesis terminates when the ribosome translocates to a termination codon. Because there are no tRNAs with anticodons complementary to the termination codons, no tRNA enters the A site of the ribosome when a termination codon is encountered. Instead, proteins called **release factors** bind to the ribosome. *E. coli* has three release factors—RF1, RF2, and RF3. Release factor 1 binds to the termination codons UAA and UAG, and RF2 binds to UGA and UAA. The binding of release factor RF1 or RF2 to the A site of the ribosome promotes the cleavage of the tRNA in the P site from the polypeptide chain and the release of the polypeptide. Release factor 3 binds to the ribosome and forms a complex with GTP. This binding brings about a conformational change in the ribosome, releasing RF1 or RF2 from the A site and causing the tRNA in the P site to move to the E site; in the process, GTP is hydrolyzed to GDP. Additional factors help bring about the release of the tRNA from the P site, the release of the mRNA from the ribosome, and the dissociation of the ribosome.

Translation in eukaryotic cells terminates in a similar way, except that there are two release factors: eRF1, which recognizes all three termination codons, and eRF2, which binds GTP and stimulates the release of the polypeptide from the ribosome.



15.20 Translation ends when a stop codon is encountered. Because UAG is the termination codon in this illustration, the release factor is RF_1.



15.21 Translation consists of tRNA charging, initiation, elongation, and termination. In this process, amino acids are linked together in the order specified by mRNA to create a polypeptide chain. A number of initiation, elongation, and release factors take part in the process, and energy is supplied by ATP and GTP.

www.whfreeman.com/pierce4e Explore the process of bacterial translation by examining the consequences of various mutations in the coding region of a gene in Animation 15.1.

POST TRANSLATIONAL MODIFICATIONS

After synthesis is completed, proteins can be modified by various methods such as phosphorylation, glycosylation, ADP ribosylation, hydroxylation, and addition of other groups.

Proteolysis

As the newly synthesized protein is released in the lumen of the ER, signal peptidases cleave peptide sequence. Apart from signal peptide, some polypeptide sequence of the protein is also cleaved resulting in the final sequence.

Example:

Insulin is synthesized in the cells in its inactive form which cannot perform its function. Post translational modifications ensure proper function which involves the removal of the part of protein to convert it into a three dimensional and fully active form.

Phosphorylation

Phosphoryalation is the addition of one or more phosphate groups to the protein. Post Translational Phosphorylation is one of the most common protein modifications that occur in animal cells. Majority of phosphorylation occurs as a mechanism to regulate the biological activity of a protein. In animal cells Serine, tyrosine and thereonine are the amino acids that subjected to the phosphorylation.

Glycosylation

Glycosylation is the addition of carbohydrate molecules to the polypeptide chain and modifying it into glycoproteins. Many of the proteins that are destined to become a part of plasma membrane or to be secreted from the cell, have carbohydrate chains attached to the amide nitrogen of asparagine(N linked) or the hydroxyl groups of serine, threonine(O linked). N glycosylation occurs in ER and O glycosylation occurs in the golgi complex.

Sulfation

Sulfate modidication takes place by the addition of sulphate molecules and these modifications of proteins occurs at tyrosine residues. Tyrosine sulfation accomplished via the activity of tyrosylproteinsulfotransferases (TPST) which are membrane associated enzymes of trans-Golgi network. There are two known TPSTs. TPST-1 TPST-2 The universal phosphate donor is 3'-phosphoadenosyl- 5'-phosphosulphate (PSPA).

Methylation

The transfer of one-carbon methyl groups to nitrogen or oxygen 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.

Hydroxylation

The biological process of addition of a hydroxy group to a protein amino acid is called Hydroxylation. Protein hydroxylation is one type of PTM that involves the conversion of –CH group into –COH group and these hydroxylated amino acids are involved in the regulation of some important factors called transcription factors. Among the 20 amino acids, the two amino acids regulated by this method are proline and lysine.

Others

a) SUMOylation

SUMO (small ubiquitin related modifier) proteins are 100 amino acid residue proteins which bind to the target protein in the same way as ubiquitin. They also confer the transcription regulatory activity of the protein and help in the transport of the target protein from cytosol to the nucleus.

b) Disulfide bond formation

Stabilizes protein structure and involved in redox processes.

c) Lipidylation, Acetylation, Prenylation etc.

SIGNIFICANCE

Proteins are synthesized by ribosomes translating mRNA into polypeptide chains, which may then undergo modifications to form the mature protein product.

Post-translational modifications of proteins, which are not gene- template based, can regulate the protein functions, by causing changes in protein activity, their cellular locations and dynamic interactions with other proteins.

PTMs have significant biological functions which include:

Aids in proper protein folding – few lectin molecules called calnexin binds to glycosylated proteins and assist in its folding.

Confers stability to the protein- glycosylation can modify the stability of the protein by increasing protein half life.

It protects the protein against cleavage by proteolytic enzyme by blocking the cleavage sites.

Protein sorting or translocation- If phosphorylated mannose residues are present in the protein it always goes to lysosome.

It regulates protein activity and function- phosphorylation of protein is a reversible PTM which activates the protein.

Acetylation regulates many diverse functions, including DNA recognition, protein-protein interaction and protein stability.

Redox-dependent PTM of proteins is emerging as a key signaling system conserved through evolution, influences many aspects of cellular homeostasis.

PTMs are important components in cell signaling, as for example when prohormones are converted to hormones.

It significantly increases the diversity and complexity in the proteome.

INHIBITING PROTEIN SYNTHESIS

- Protein synthesis inhibitors usually act at the ribosome level, taking advantage of the major differences between prokaryotic and eukaryotic ribosome structures.
- 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.
- By targeting different stages of the mRNA translation, antimicrobial drugs can be changed if resistance develops.

A protein synthesis inhibitor 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. It usually refers to substances, such as antimicrobial drugs, that act at the ribosome level. The substances take advantage of the major differences between prokaryotic and eukaryotic ribosome structures which differ in their size, sequence, structure, and the ratio of protein to RNA. The differences in structure allow some antibiotics to kill bacteria by inhibiting their ribosomes, while leaving human ribosomes unaffected.

Translation in prokaryotes involves the assembly of the components of the translation system which are: the two ribosomal subunits (the large 50S & small 30S subunits), the mRNA to be translated, the first aminoacyl tRNA, GTP (as a source of energy), and three initiation factors that help the assembly of the initiation complex. The ribosome has three sites: the A site, the P site, and the E site (not shown in). The A site is the point of entry for the aminoacyl tRNA. The P site is where the peptidyl tRNA is formed in the ribosome. The E site which is the exit site of the now uncharged tRNA after it gives its amino acid to the growing peptide chain.

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. The following is a list of common antibacterial drugs and the stages which they target.

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

- Tetracyclines and Tigecycline (a glycylcycline related to tetracyclines) block the A site on the ribosome, preventing the binding of aminoacyl tRNAs.
- Aminoglycosides, among other potential mechanisms of action, interfere with the proofreading process, causing an increased rate of error in synthesis with premature termination.
- Chloramphenicol blocks the peptidyl transfer step of elongation on the 50S ribosomal subunit in both bacteria and mitochondria.
- Macrolides, clindamycin, and aminoglycosides have evidence of inhibition of ribosomal translocation.
- Streptogramins also cause premature release of the peptide chain.

By targeting different stages of the mRNA translation, antimicrobial drugs can be changed if resistance develops to one or many of the drugs.

REFERENCES

GENETICS: A conceptual approach, 4th Edition, Benjamin A. Pierce, W. H. Freeman and company England; 2006

Cell Biology, Genetics, Molecular Biology, Evolution and Ecology, P.S. Verma, V.K. Agarwal, S. Chand & Company Ltd, 2005

Principles of Molecular Biology, Veer Bala Rastogi, Medtech, 2016

https://microbenotes.com/post-translational-modification/

https://bio.libretexts.org/Bookshelves/Microbiology/Book%3A_Microbiology_(Boundless)/13%3A_Antimicrobial_Drugs/13.2%3A_Functions_of_Antimicrobial_Drugs/13.2D%3A_Inhibiting_Protein_Synthesis

SATHYABAMA
INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)
Accredited "A" Grade by NAAC | 12B Status by UGC | Approved by AICTE
www.sathyabama.ac.in

SCHOOL OF BIO AND CHEMICAL ENGINEERING

DEPARTMENT OF BIOTECHNOLOGY

SBTA1301 – MOLECULAR BIOLOGY AND GENETICS

SBTA1301	MOLECULAR BIOLOGY AND GENETICS	L	T	P	Credit	Total Marks
		3	*	0	3	100

COURSE OBJECTIVES

- The course aims to give an understanding on the fundamentals of conventional genetics and the intricate molecular mechanisms of heredity and variations.
- To understand storage of genetic information and its translation at molecular level in prokaryotic and eukaryotic systems.

UNIT 1 CLASSICAL GENETICS **9 Hrs.**

Fundamental principles of genetics- Mendel's principles and experiments, gene interaction, multiple alleles, complementation, linkage, sex linked, sex limited and sex influenced inheritance; Chromosomes basis of heredity- extra-chromosomal inheritance; Linkage and crossing over; Hardy-Weinberg equilibrium, Extensions of Hardy- Weinberg equilibrium.

UNIT 2 STRUCTURE AND PROPERTIES OF NUCLEIC ACIDS **9 Hrs.**

Introduction to nucleic acids: Evidence for DNA&RNA as a genetic material; Structure and physicochemical properties of elements in DNA and RNA, Biological significance of differences in DNA and RNA. Primary structure of DNA: Chemical and structural qualities of 3',5'-Phosphodiester bond. Secondary Structure of DNA: Watson & Crick model, Chargaff's rule; DNA replication- Overview of differences in prokaryotic and eukaryotic DNA replication, D-loop and rolling circle mode of replication, Telomere replication in eukaryotes; Okazaki fragments, Fidelity of DNA replication, Inhibitors of DNA replication, DNA repair- Mutagens, DNA mutations and various types of repair mechanisms.

UNIT 3 TRANSCRIPTION **9 Hrs.**

Central Dogma in molecular biology -Structure and function of mRNA, rRNA tRNA and micro RNAs. Characteristics of promoter and enhancer sequences. RNA synthesis: Initiation, elongation and termination of RNA synthesis, Proteins of RNA synthesis, Fidelity of RNA synthesis, Inhibitors of transcription, Differences in prokaryotic and eukaryotic transcription. Basic concepts in RNA world: RNA processing: 5'-Capping, Splicing-Alternative splicing, Poly 'A' tail addition and base modification.

UNIT 4 TRANSLATION **9 Hrs.**

Introduction to Genetic code: Elucidation of genetic code, Codon degeneracy, Wobble hypothesis and its importance, Prokaryotic and eukaryotic ribosomes. Steps in translation: Initiation, Elongation and termination of protein synthesis. Inhibitors of protein synthesis. Post-translational modifications and its importance.

UNIT 5 REGULATION OF GENE EXPRESSION **9 Hrs.**

Organization of genes in prokaryotic and eukaryotic chromosomes- operon concept; Gene expression and regulation- Hierarchical levels of gene regulation, Prokaryotic gene regulation –lac and trp operon, Eukaryotic gene regulation- gene silencing.

Max.45 Hrs.

COURSE OUTCOMES

On completion of the course, student will be able to

- CO1 - Explain the foundations of Mendelian genetics and chromosomal theory and apply these, with appropriate terminology, to contemporary concepts in genetics.
- CO2 - Emphasize the molecular mechanism of DNA replication and repair in various organisms
- CO3 - Explain the properties of genetic materials and storage and processing of genetic information.
- CO4 - Analyze the processes of transcription and translation in both prokaryotes and eukaryotes at molecular level.
- CO5 - Understand the redundant and universal qualities of the genetic code and how it is used to determine the amino acid sequence of a polypeptide.
- CO6 - Compare the mechanisms of gene regulation in prokaryotes and eukaryotes.

TEXT / REFERENCE BOOKS

1. Lewin B., Genes XI, International Edition, Jocelyn Krebs, Stephen Kilpatrick and Elliott Goldstein, Jones & Bartlett Learning, 2017, ISBN 978-1-4496-5985-1
2. Tropp, Burton E., Molecular Biology: Genes to Proteins, 3rd Edition, Jones and Bartlett, 2008.
3. Glick B.R. and Pasternak J.J., Molecular Biotechnology: Principles and Applications of Recombinant DNA, 4th Edition. ASM, 2010

UNIT – V –REGULATION OF GENE EXPRESSION

Chromosome

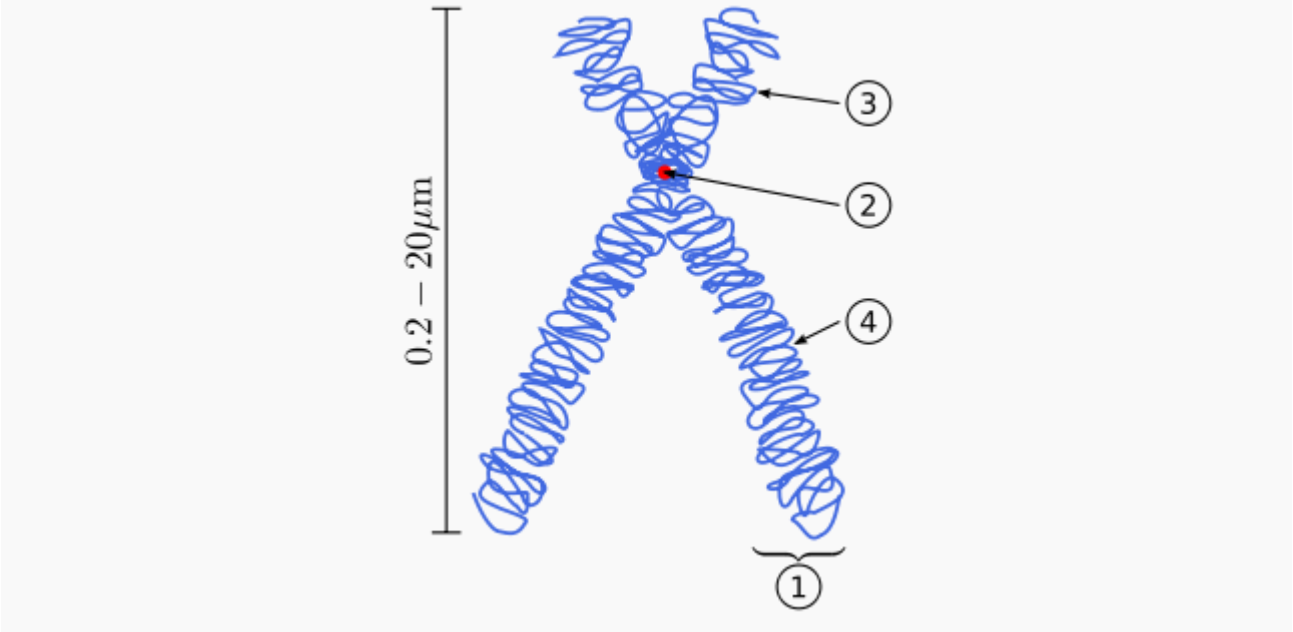


Diagram of a replicated and condensedmetaphase eukaryotic chromosome. (1)Chromatid – one of the two identical parts of the chromosome after S phase. (2)Centromere – the point where the two chromatids touch. (3) Short arm. (4) Long arm.

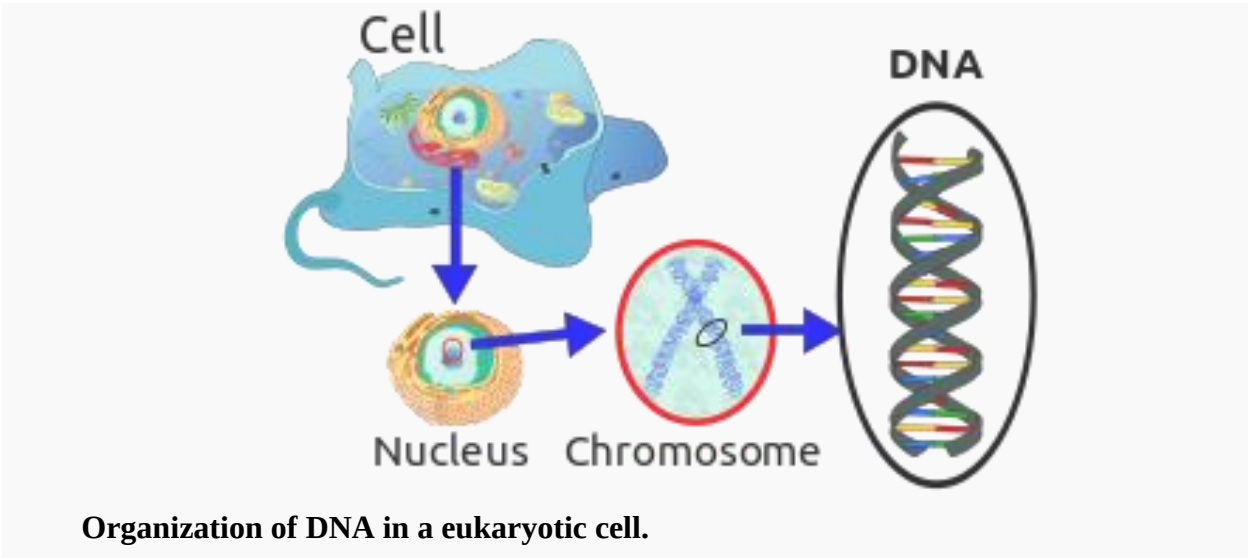
A **chromosome** (*chromo-* + *-some*) is a packaged and organized structure containing most of the DNA of a living organism. It is not usually found on its own, but rather is structured by being wrapped around protein complexes called nucleosomes, which consist of proteins called histones. The DNA in chromosomes is also associated with transcription (copying of genetic sequences) factors and several other macromolecules. During most of the duration of the Cell cycle, a chromosome consists of one long double-stranded DNA molecule (with associated proteins). During S phase, the chromosome gets replicated, resulting in an 'X'-shaped structure called a metaphase chromosome. Both the original and the newly copied DNA are now called chromatids. The two "sister" chromatids join together at a protein junction called a centromere. Chromosomes are normally visible under a light microscope only when the cell is undergoing mitosis (cell division). Even then, the full chromosome containing both joined sister chromatids becomes visible only during a sequence of mitosis known as metaphase (when chromosomes align together, attached to the mitotic spindle and prepare to divide). This DNA and its associated proteins and macromolecules is collectively known as chromatin, which is further packaged along with its associated molecules into a discrete structure called a nucleosome. Chromatin is present in most cells, with a few exceptions - erythrocytes for example. Occurring only in the nucleus of eukaryotic cells, chromatin

composes the vast majority of all DNA, except for a small amount inherited maternally which is found in mitochondria. In prokaryotic cells, chromatin occurs free-floating in cytoplasm, as these cells lack organelles and a defined nucleus. Bacteria also lack histones. The main information-carrying macromolecule is a single piece of coiled double-stranded DNA, containing many genes, regulatory elements and other noncoding DNA. The DNA-bound macromolecules are proteins, which serve to package the DNA and control its functions. Chromosomes vary widely between different organisms. Some species such as certain bacteria also contain plasmids or other extrachromosomal DNA. These are circular structures in the cytoplasm which contain cellular DNA and play a role in horizontal gene transfer.

Compaction of the duplicated chromosomes during cell division (mitosis or meiosis) results either in a four-arm structure (pictured to the right) if the centromere is located in the middle of the chromosome or a two-arm structure if the centromere is located near one of the ends. Chromosomal recombination during meiosis and subsequent sexual reproduction plays a vital role in genetic diversity. If these structures are manipulated incorrectly, through processes known as chromosomal instability and translocation, the cell may undergo mitotic catastrophe and die, or it may unexpectedly evade apoptosis leading to the progression of cancer.

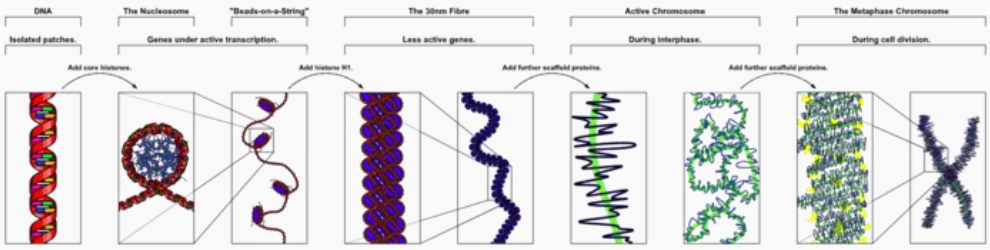
In prokaryotes (see nucleoids) and viruses, the DNA is often densely packed and organized: in the case of archaea, by homologs to eukaryotic histones, and in the case of bacteria, by histone-like proteins. Small circular genomes called plasmids are often found in bacteria and also in mitochondria and chloroplasts, reflecting their bacterial origins.

Eukaryotes



Organization of DNA in a eukaryotic cell.

In eukaryotes, nuclear chromosomes are packaged by proteins into a condensed structure called chromatin. This allows the very long DNA molecules to fit into the cell nucleus. The structure of chromosomes and chromatin varies through the cell cycle. Chromosomes are even more condensed than chromatin and are an essential unit for cellular division. Chromosomes must be replicated, divided, and passed successfully to their daughter cells so as to ensure the genetic diversity and survival of theirprogeny. Chromosomes may exist as either duplicated or unduplicated. Unduplicated chromosomes are single linear strands, whereas duplicated chromosomes contain two identical copies (called chromatids or sister chromatids) joined by a centromere.



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

Eukaryotes (cells with nuclei such as those found in plants, yeast, and animals) possess multiple large linear chromosomes contained in the cell's nucleus. Each chromosome has one centromere, with one or two arms projecting from the centromere, although, under most circumstances, these arms are not visible as such. In addition, most eukaryotes have a small circular mitochondrial genome, and some eukaryotes may have additional small circular or linear cytoplasmic chromosomes.

In the nuclear chromosomes of eukaryotes, the uncondensed DNA exists in a semi-ordered structure, where it is wrapped around histones (structural proteins), forming a composite material called chromatin.

Chromatin

Chromatin is the complex of DNA and protein found in the eukaryotic nucleus, which packages chromosomes. The structure of chromatin varies significantly between different stages of the cell cycle, according to the requirements of the DNA.

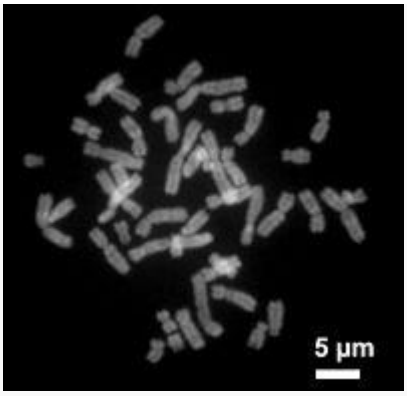
Interphase chromatin

During interphase (the period of the cell cycle where the cell is not dividing), two types of chromatin can be distinguished:

- Euchromatin, which consists of DNA that is active, e.g., being expressed as protein.

- Heterochromatin, which consists of mostly inactive DNA. It seems to serve structural purposes during the chromosomal stages. Heterochromatin can be further distinguished into two types:
  - *Constitutive heterochromatin*, which is never expressed. It is located around the centromere and usually contains repetitive sequences.
  - *Facultative heterochromatin*, which is sometimes expressed.

Metaphase chromatin and division



Human chromosomes duringmetaphase

In the early stages of mitosis or meiosis (cell division), the chromatin strands become more and more condensed. They cease to function as accessible genetic material (transcription stops) and become a compact transportable form. This compact form makes the individual chromosomes visible, and they form the classic four arm structure, a pair of sister chromatids attached to each other at the centromere. The shorter arms are called *p arms* (from the French *petit*, small) and the longer arms are called *q arms* (*q* follows *p* in the Latin alphabet; q-g "grande"; alternatively it is sometimes said q is short for *queue* meaning tail in French[19]). This is the only natural context in which individual chromosomes are visible with an optical microscope.

During mitosis, microtubules grow from centrosomes located at opposite ends of the cell and also attach to the centromere at specialized structures called kinetochores, one of which is present on each sister chromatid. A special DNA base sequence in the region of the kinetochores provides, along with special proteins, longer-lasting attachment in this region. The microtubules then pull the chromatids apart toward the centrosomes, so that each daughter cell inherits one set of chromatids. Once the cells have divided, the chromatids are uncoiled and DNA can again be transcribed. In spite of their appearance, chromosomes are structurally highly condensed, which enables these giant DNA structures to be contained within a cell nucleus (Fig. 2).

Human chromosomes

Chromosomes in humans can be divided into two types: autosomes and sex chromosomes. Certain genetic traits are linked to a person's sex and are passed on through the sex chromosomes. The autosomes contain the rest of the genetic hereditary information. All act in the same way during cell division. Human cells have 23 pairs of chromosomes (22 pairs of autosomes and one pair of sex chromosomes), giving a total of 46 per cell. In addition to these, human cells have many hundreds of copies of the mitochondrial genome.Sequencing of the human genome has provided a great deal of information about each of the chromosomes. Below is a table compiling statistics for the chromosomes, based on the Sanger Institute's human genome information in the Vertebrate Genome Annotation (VEGA) database. Number of genes is an estimate as it is in part based on gene predictions. Total chromosome length is an estimate as well, based on the estimated size of unsequenced heterochromatin regions.

The Operon

Within its tiny cell, the bacterium *E. coli* contains all the genetic information it needs to metabolize, grow, and reproduce. It can synthesize every organic molecule it needs from glucose and a number of inorganic ions.

Many of the genes in *E. coli* are expressed constitutively; that is, they are always turned "on". Others, however, are active only when their products are needed by the cell, so their expression must be regulated.

Two examples:

- If the amino acid tryptophan (**Trp**) is added to the culture, the bacteria soon stop producing the five enzymes previously needed to synthesize Trp from intermediates produced during the respiration of glucose. In this case, the presence of the products of enzyme action **represses** enzyme synthesis.
- Conversely, adding a new substrate to the culture medium may **induce** the formation of new enzymes capable of metabolizing that substrate. If we take a culture of *E. coli* that is feeding on glucose and transfer some of the cells to a medium contain lactose instead, a revealing sequence of events takes place.
    - At first the cells are quiescent: they do not metabolize the lactose, their other metabolic activities decline, and cell division ceases.
    - Soon, however, the culture begins growing rapidly again with the lactose being rapidly consumed. What has happened? During the quiescent interval, the cells began to produce **three enzymes**.

The three enzymes are

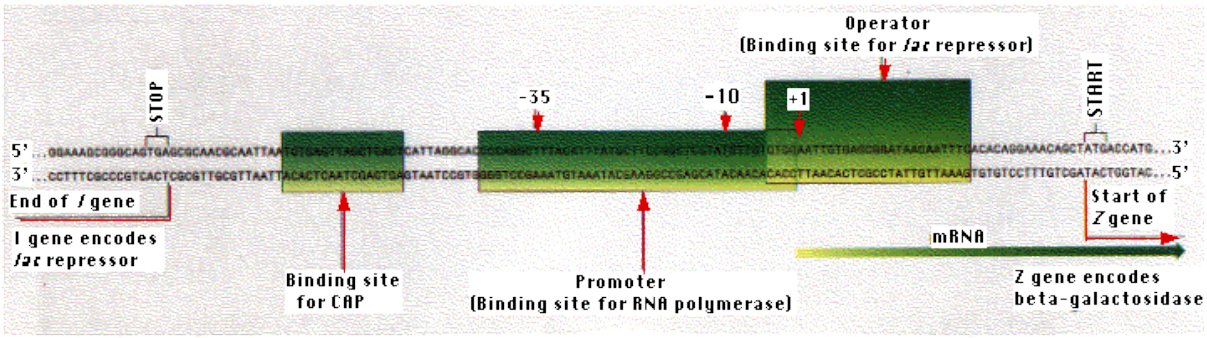
- a **permease** that transports lactose across the plasma membrane from the culture medium into the interior of the cell
- **beta-galactosidase** which converts lactose into the intermediate allolactose and then hydrolyzes this into glucose and galactose. Once in the presence of lactose, the quantity of beta-galactosidase in the cells rises from a tiny amount to almost 2% of the weight of the cell.
- a **transacetylase** whose function is still uncertain.

The *lac* operon

The capacity to respond to the presence of lactose was always there. The genes for the three induced enzymes are part of the genome of the cell. But until lactose was added to the culture medium, these genes were not expressed (β-galactosidase was expressed weakly — just enough to convert lactose into allolactose).

The most direct way to control the expression of a gene is to **regulate its rate of transcription**; that is, the rate at which RNA polymerase transcribes the gene into molecules of messenger RNA (mRNA).

Gene transcription begins at a particular nucleotide shown in the figure as "**+1**". RNA polymerase actually binds to a site "upstream" (i.e., on the 5' side) of this site and opens the double helix so that transcription of one strand can begin.



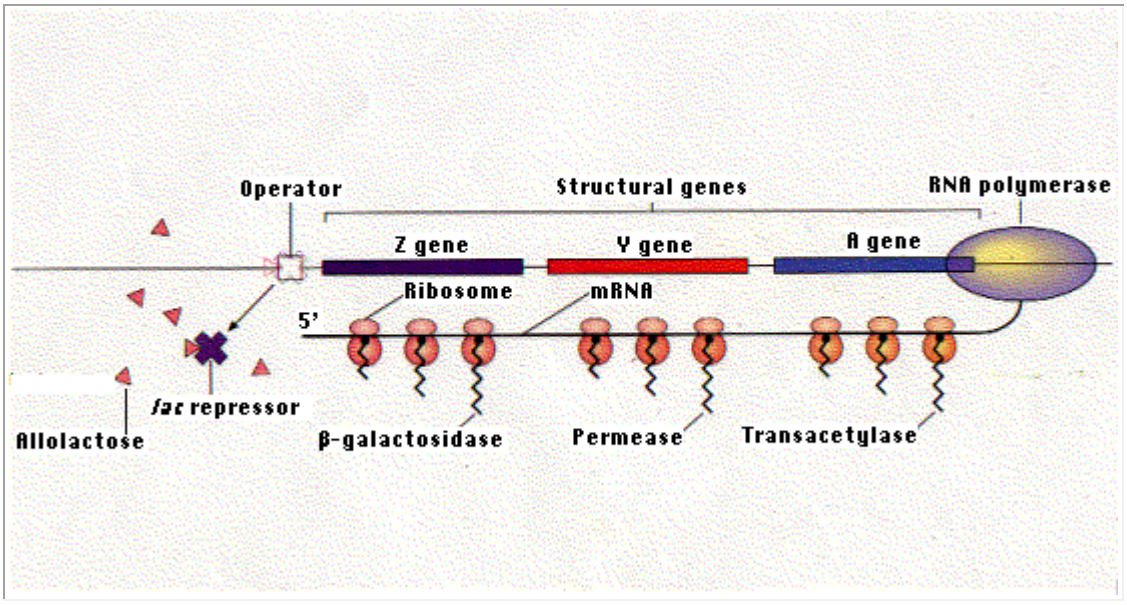
The binding site for RNA polymerase is called the **promoter**. In bacteria, two features of the promoter appear to be important:

- a sequence of TATAAT (or something similar) centered 10 nucleotides upstream of the **+1** site and
- another sequence (TTGACA or something quite close to it) centered 35 nucleotides upstream.

The exact DNA sequence between the two regions does not seem to be important.

Each of the three enzymes synthesized in response to lactose is encoded by a separate gene. The three genes are arranged in tandem on the bacterial chromosome.

The *lac* operon. In the absence of lactose, the repressor protein encoded by the I gene binds to the *lac* operator and prevents transcription. Binding of allolactose to the repressor causes it to leave the operator. This enables RNA polymerase to transcribe the three genes of the operon. The single mRNA molecule that results is then translated into the three proteins.



The ***lac* repressor** binds to a specific sequence of two dozen nucleotides called the **operator**. Most of the operator is downstream of the promoter. When the repressor is bound to the operator, RNA polymerase is unable to proceed downstream with its task of gene transcription.

The operon is the combination of the

- operator and
- the three protein-encoding genes associated with it.

The gene encoding the *lac* repressor is called the *I* gene. It happens to be located just upstream of the *lac* promoter. However, its precise location is probably not important because it achieves its effect by means of its protein product, which is free to diffuse throughout the cell. And, in fact, the genes for some repressors are not located close to the operators they control.

Although repressors are free to diffuse through the cell, how does — for example — the lac repressor find the single stretch of 24 base pairs of the operator out of the 4.6 million base pairs of

DNA in the *E. coli* genome? It turns out the repressor is free to bind **anywhere** on the DNA using both

- hydrogen bonds and
- ionic (electrostatic) interactions between its positively-charged amino acids (Lys, Arg) and the negative charges on the deoxyribose-phosphate backbone of the DNA.

Once astride the DNA, the repressor can move along it until it encounters the operator sequence. Now an allosteric change in the tertiary structure of the protein allows the same amino acids to establish bonds — mostly hydrogen bonds and hydrophobic interactions — with particular bases in the operator sequence.

The *lac* repressor is made up of four identical polypeptides (thus a "homotetramer"). Part of the molecule has a site (or sites) that enable it to recognize and bind to the 24 base pairs of the lac operator. Another part of the repressor contains sites that bind to allolactose. When allolactose unites with the repressor, it causes a change in the shape of the molecule, so that it can no longer remain attached to the DNA sequence of the operator. Thus, when lactose is added to the culture medium,

- it causes the repressor to be released from the operator
- RNA polymerase can now begin transcribing the 3 genes of the operon into a **single** molecule of **messenger RNA**.

Hardly does transcription begin, before ribosomes attach to the growing mRNA molecule and move down it to **translate** the message into the three proteins. You can see why punctuation codons — UAA, UAG, or UGA — are needed to terminate translation between the portions of the mRNA coding for each of the three enzymes.

This mechanism is characteristic of bacteria, but differs in several respects from that found in eukaryotes:

- Genes in eukaryotes are not linked in operons (except for nematodes like *C. elegans* and tunicates like *Ciona intestinalis*).
- Primary transcripts in eukaryotes contain the transcript of only a single gene (with the above exceptions).

- Transcription and translation are not physically linked in eukaryotes as they are in bacteria; transcription occurs in the nucleus while translation occurs in the cytosol (with a few exceptions).

Corepressors

As mentioned above, the synthesis of tryptophan from precursors available in the cell requires 5 enzymes. The genes encoding these are clustered together in a single operon with its own promoter and operator. In this case, however, the **presence** of tryptophan in the cell **shuts down** the operon. When Trp is present, it binds to a site on the Trp repressor and **enables** the Trp repressor to bind to the operator. When Trp is not present, the repressor leaves its operator, and transcription of the 5 enzyme-encoding genes begins.

The usefulness to the cell of this control mechanism is clear. The presence in the cell of an essential metabolite, in this case tryptophan, turns off its own manufacture and thus stops unneeded protein synthesis.

As its name suggests, repressors are **negative control** mechanisms, shutting down operons

- in the absence of a substrate (lactose in our example) or
- the presence of an essential metabolite (tryptophan is our example).

However, some gene transcription in *E. coli* is under positive control.

Positive Control of Transcription: CAP

Absence of the lac repressor is essential but not sufficient for effective transcription of the lac operon. The activity of RNA polymerase also depends on the presence of another DNA-binding protein called **catabolite activator protein** or **CAP**. Like the lac repressor, CAP has two types of binding sites:

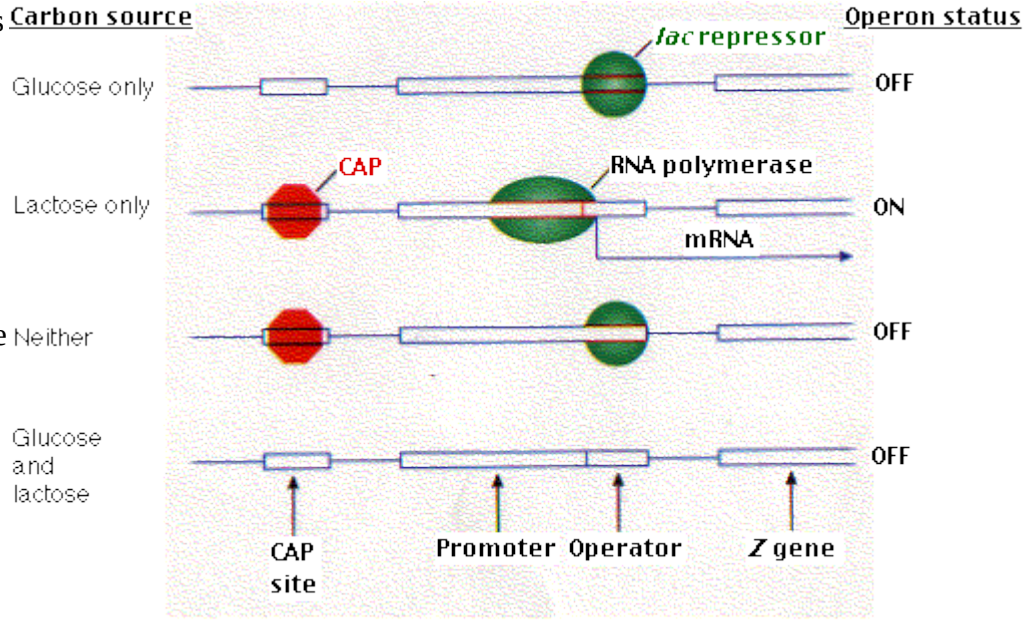
- One binds the nucleotide cyclic AMP; the other
- binds a sequence of 16 base pairs upstream of the promoter

However, CAP can bind to DNA only when cAMP is bound to CAP. so when cAMP levels in the cell are low, CAP fails to bind DNA and thus RNA polymerase cannot begin its work, even in the absence of the repressor.

So the *lac* operon is under both **negative** (the **repressor**) and **positive** (**CAP**) control.

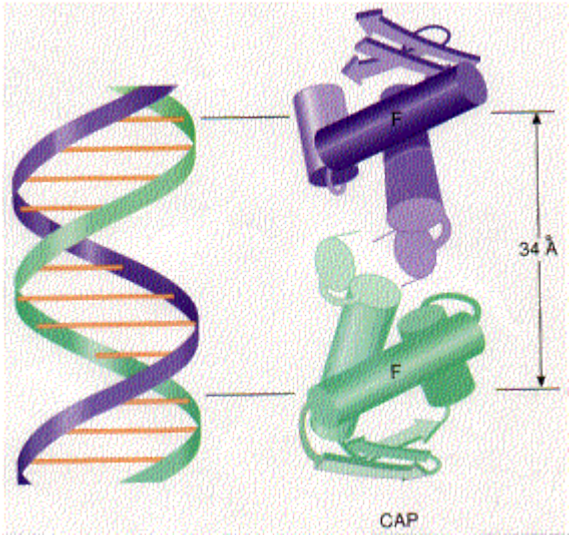
Why?

It turns out that it is not simply a matter of belt and suspenders. This dual system enables the cell to **make choices**. What, for example, should the cell do when fed both glucose and lactose? Presented with such a choice, *E. coli* (for reasons about which we can only speculate) chooses glucose. It makes its choice by using the interplay between these two control devices.



- Although the presence of lactose removes the repressor,
- the presence of glucose lowers the level of cAMP in the cell and thus removes CAP.

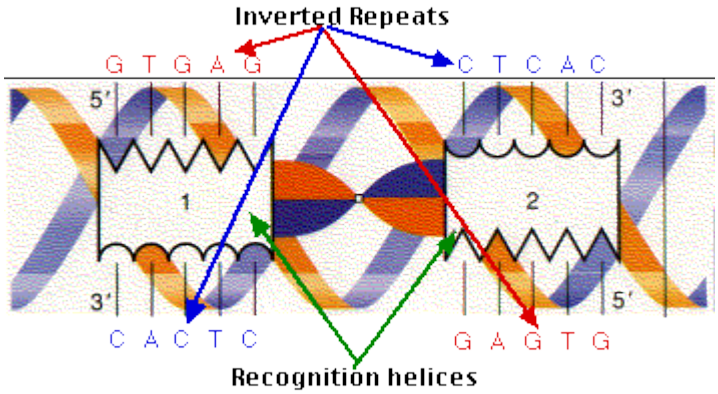
Without CAP, binding of RNA polymerase is inhibited even though there is no repressor to interfere with it if it could bind. The molecular basis for its choices is shown in the graphic.



CAP consists of two identical polypeptides (hence it is a homodimer). Toward the C-terminal, each has two regions of alpha helix with a sharp bend between them. The longer of these is called the **recognition helix** because it is responsible for recognizing and binding to a particular sequence of bases in DNA.

The graphic shows a model of CAP. The two monomers are identical. Each monomer recognizes a sequence of nucleotides in DNA by means of the region of alpha helix labeled **F**. Note that the two

recognition helices are spaced 34Å apart, which is the distance that it takes the DNA molecule (on the left) to make precisely one complete turn.



The recognition helices of each polypeptide of CAP are, of course, identical. But their orientation in the dimer is such that the sequence of bases they recognize must run in the opposite direction for each recognition helix to bind properly. This arrangement of two identical sequences of base pairs running in opposite directions is called an **inverted repeat**.

The strategy illustrated by CAP and its binding site has turned out to be used widely. As more and more DNA-regulating proteins have been discovered, many turn out to share the traits we find in CAP:

- They usually contain two subunits. Therefore, they are **dimers**.
- They recognize and bind to DNA sequences with **inverted repeats**.
- In bacteria, recognition and binding to a particular sequence of DNA is accomplished by a segment of alpha helix. Hence these proteins are often described as **helix-turn-helix** proteins. The Trp repressor shown above is a member of this group.

Riboswitches

Protein repressors and corepressors are not the only way in which bacteria control gene transcription. It turns out that the regulation of the level of certain metabolites can also be controlled by riboswitches. A riboswitch is section of the 5'-untranslated region (5'-UTR) in a molecule of messenger RNA (mRNA) which has a specific binding site for the metabolite (or a close relative).

Some of the metabolites that bind to riboswitches:

- the purines adenine and guanine
- the amino acids glycine and lysine
- flavin mononucleotide (the prosthetic group of NADH dehydrogenase)
- S-adenosyl methionine (that donates methyl groups to many molecules, including
    - DNA
    - the cap at the 5' end of messenger RNA [Link]
- tRNAs. When these are bound to their amino acid (aminoacyl-tRNA), they bind to the riboswitch in the mRNA that encodes the enzyme (an aminoacyl-tRNA synthetase) responsible for loading the amino acid onto the tRNA. This causes transcription of the mRNA to terminate prematurely. tRNAs with no amino acid attached also bind to the riboswitch but in such a way that transcription of the mRNA continues. Its translation (in bacteria, translation begins while transcription is still going on [Link]) produces the aminoacyl-tRNA synthetase used to load the amino acid onto the tRNA. Thus these riboswitches regulate the level of aminoacyl-tRNAs producing more when needed, less when not (a kind of feedback inhibition.)

In each case, the riboswitch regulates **transcription** of genes involved in the metabolism of that molecule. The metabolite binds to the growing mRNA and induces an allosteric change that

- for some genes causes further synthesis of the mRNA to terminate before forming a functional product and
- for other genes, enhances completion of synthesis of the mRNA.
- In both cases, one result is to control the level of that metabolite.

Some riboswitches control mRNA **translation** rather than its transcription.

It has been suggested that these regulatory mechanisms, which **do not involve any protein**, are a relict from an "RNA world".

REGULATION OF GENE EXPRESSION IN EUKARYOTES

Many features of gene regulation are common to both bacterial and eukaryotic cells. For example, in both types of cells, DNA-binding proteins influence the ability of RNA polymerase to initiate transcription. However, there are also some differences.

First, most eukaryotic genes are not organized into operons and are rarely transcribed together into a single mRNA molecule, although some operon-like gene clusters have been discovered in eukaryotes. In eukaryotic cells, each structural gene typically has its own promoter and is transcribed separately.

Second, chromatin structure affects gene expression in eukaryotic cells; DNA must unwind from the histone proteins before transcription can take place.

Third, the presence of the nuclear membrane in eukaryotic cells separates transcription and translation in time and space.

Therefore, the regulation of gene expression in eukaryotic cells is characterized by a greater diversity of mechanisms that act at different points in the transfer of information from DNA to protein.

Eukaryotic gene regulation is less well understood than bacterial regulation, partly owing to the larger genomes in eukaryotes, their greater sequence complexity, and the difficulty of isolating and manipulating mutations that can be used in the study of gene regulation. Nevertheless, great advances in our understanding of the regulation of eukaryotic genes have been made in recent years, and eukaryotic regulation continues to be a cutting-edge area of research in genetics.

Changes in Chromatin Structure Affect the Expression of Genes

One type of gene control in eukaryotic cells is accomplished through the modification of chromatin structure. In the nucleus, histone proteins associate to form octamers, around which helical DNA tightly coils to create chromatin. In a general sense, this chromatin structure represses gene expression. For a gene to be transcribed, transcription factors, activators, and RNA polymerase must bind to the DNA. How can these events take place with DNA wrapped tightly around histone

proteins? The answer is that, before transcription, chromatin structure changes and the DNA becomes more accessible to the transcriptional machinery.

DNase I Hypersensitivity

Several types of changes are observed in chromatin structure when genes become transcriptionally *active*. As genes become transcriptionally active, regions around the genes become highly sensitive to the action of DNase I. These regions, called **DNase I hypersensitive sites**, frequently develop about 1000 nucleotides upstream of the start site of transcription, suggesting that the chromatin in these regions adopts a more open configuration during transcription. This relaxation of the chromatin structure allows regulatory proteins access to binding sites on the DNA. Indeed, many DNase I hypersensitive sites correspond to known binding sites for regulatory proteins. At least three different processes affect gene regulation by altering chromatin structure: (1) the modification of histone proteins; (2) chromatin remodeling; and (3) DNA methylation. Each of these mechanisms will be discussed in the sections that follow.

Histone Modification

Histones in the octamer core of the nucleosome have two domains: (1) a globular domain that associates with other histones and the DNA and (2) a positively charged tail domain that probably interacts with the negatively charged phosphate groups on the backbone of DNA. The tails of histone proteins are often modified by the addition or removal of phosphate groups, methyl groups, or acetyl groups. These modifications have sometimes been called the **histone code**, because they encode information that affects how genes are expressed.

Methylation of histones

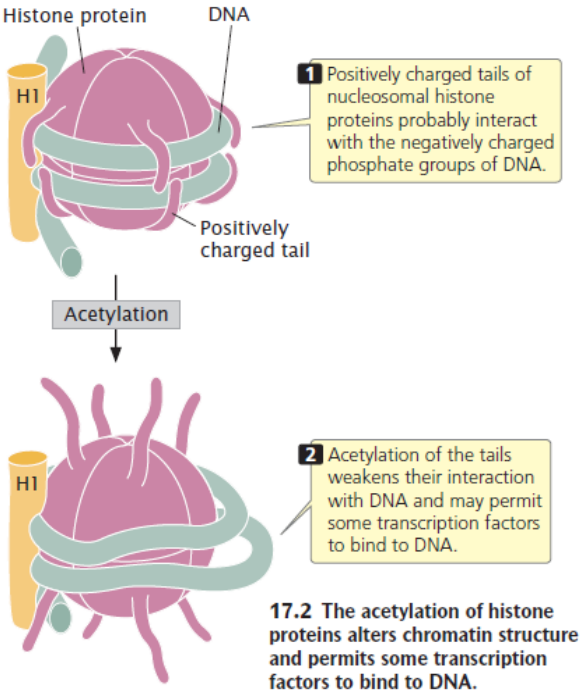
One type of histone modification is the addition of methyl groups to the tails of histone proteins. These modifications can bring about either the activation or the repression of transcription, depending on which particular amino acids in the histone tail are methylated. A common modification is the addition of three methyl groups to lysine 4 in the tail of the H3 histone protein, abbreviated H3K4me3 (K is the abbreviation for lysine). The H3K4me3 modification is frequently found in promoters of transcriptionally active genes in eukaryotes.

Recent studies have identified proteins that recognize and bind to H3K4me3, including a protein called nucleosome remodeling factor (NURF). NURF and other proteins that recognize H3K4me3 have a common protein-binding domain that binds to the H3 histone tail and then alters chromatin packing, allowing transcription to take place.

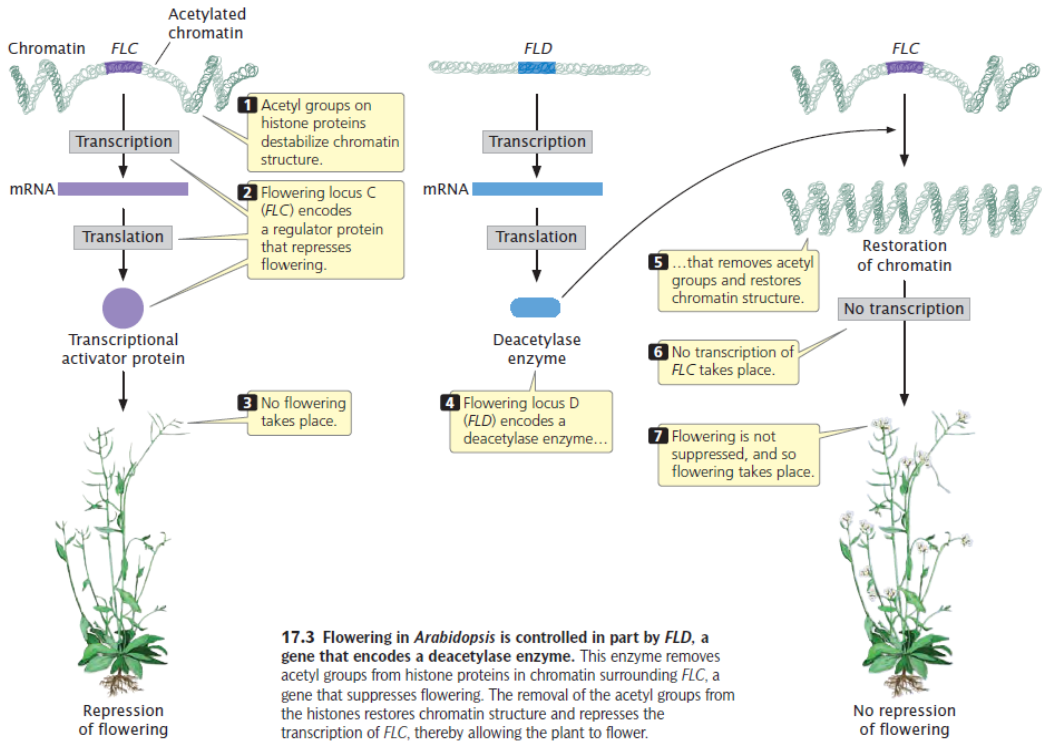
Research has also demonstrated that some transcription factors, which are necessary for the initiation of transcription, directly bind to H3K4me3.

Acetylation of histones

Another type of histone modification that affects chromatin structure is acetylation, the addition of acetyl groups (CH3CO) to histone proteins (**Figure 17.2**). The acetylation of histones usually stimulates transcription. For example, the addition of a single acetyl group to lysine 16 in the tail of the H4 histone prevents the formation of the 30-nm chromatin fiber (see Figure 11.4), causing the chromatin to be in an open configuration and available for transcription. In general, acetyl groups destabilize chromatin structure, allowing transcription to take place. Acetyl groups are added to histone proteins by acetyltransferase enzymes; other enzymes called deacetylases strip acetyl groups from histones and restore chromatin structure, which represses transcription. Certain transcription factors (see Chapter 13) and other proteins that regulate transcription either have acteyltransferase activity or attract acteyltransferases to DNA.



17.2 The acetylation of histone proteins alters chromatin structure and permits some transcription factors to bind to DNA.



17.3 Flowering in *Arabidopsis* is controlled in part by *FLD*, a gene that encodes a deacetylase enzyme. This enzyme removes acetyl groups from histone proteins in chromatin surrounding *FLC*, a gene that suppresses flowering. The removal of the acetyl groups from the histones restores chromatin structure and represses the transcription of *FLC*, thereby allowing the plant to flower.

Chromatin Remodeling

The changes to chromatin structure discussed so far have been through alteration of the structure of the histone proteins. Some transcription factors and other regulatory proteins alter chromatin structure without altering the chemical structure of the histones directly. These proteins are called **chromatin-remodeling complexes**. They bind directly to particular sites on DNA and reposition the nucleosomes, allowing transcription factors two mechanisms by which remodeling complexes reposition nucleosomes. First, some remodeling complexes cause the nucleosome to slide along the DNA, allowing DNA that was wrapped around the nucleosome to occupy a position in between nucleosomes, where it is more accessible to proteins affecting gene expression. Second, some complexes cause conformational changes in the DNA, in nucleosomes, or in both so that DNA that is bound to the nucleosome assumes a more exposed configuration.

Chromatin-remodeling complexes are targeted to specific DNA sequences by transcriptional activators or repressors that attach to a remodeling complex and then bind to the promoters of specific genes. There is also evidence that chromatin-remodeling complexes work together with enzymes that alter histones, such as acetyltransferase enzymes, to change chromatin structure and expose DNA for transcription.

DNA Methylation

Another change in chromatin structure associated with transcription is the methylation of cytosine bases, which yields 5-methylcytosine (see Figure 10.18). The methylation of cytosine in DNA is distinct from the methylation of histone proteins mentioned earlier. Heavily methylated DNA is associated with the repression of transcription in vertebrates and plants, whereas transcriptionally active DNA is usually unmethylated in these organisms. Abnormal patterns of methylation are also associated with some types of cancer.

DNA methylation is most common on cytosine bases adjacent to guanine nucleotides (CpG, where p represents the phosphate group in the DNA backbone); so two methylated cytosines sit diagonally across from each other on opposing strands:

$$5'-\overset{m}{C}G-3'$$
$$3'-G\underset{m}{C}-5'$$

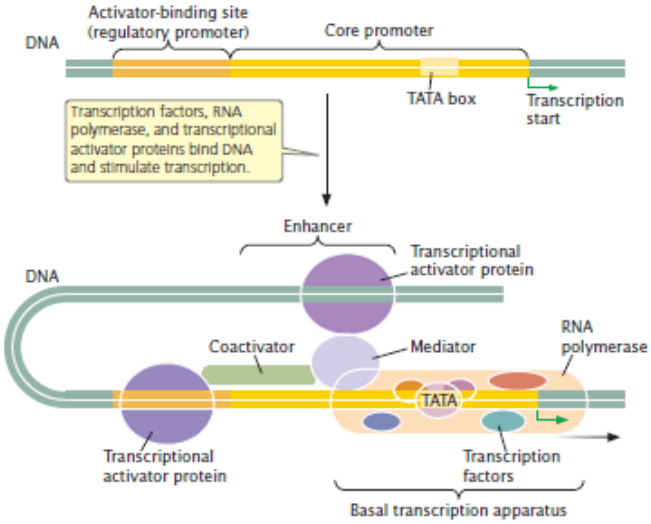
DNA regions with many CpG sequences are called **CpG islands** and are commonly found near transcription start sites. While genes are not being transcribed, these CpG islands are often methylated, but the methyl groups are removed before the initiation of transcription. CpG methylation is also associated with long-term gene repression, such as on the inactivated X chromosome of female mammals.

Evidence indicates that an association exists between DNA methylation and the deacetylation of histones, both of which repress transcription. Certain proteins that bind tightly to methylated CpG sequences form complexes with other proteins that act as histone deacetylases. In other words, methylation appears to attract deacetylases, which remove acetyl groups from the histone tails, stabilizing the nucleosome structure and repressing transcription. The demethylation of DNA allows acetyltransferases to add acetyl groups, disrupting nucleosome structure and permitting transcription.

The Initiation of Transcription Is Regulated by Transcription Factors and Transcriptional Regulator Proteins

another important level of control—that is, control through the binding of proteins to DNA sequences that affect transcription. Transcription is an important level of control in eukaryotic cells, and this control requires a number of different types of proteins and regulatory elements. The initiation of eukaryotic transcription was discussed in detail in Chapter 13. Recall that general transcription factors and RNA polymerase assemble into a *basal transcription apparatus*, which binds to a *core promoter* located immediately upstream of a gene. The basal transcription apparatus is capable of minimal

levels of transcription; *transcriptional regulator proteins* are required to bring about normal levels of transcription. These proteins bind to a regulatory promoter, which is located upstream of the core promoter (**Figure 17.5**), and to *enhancers*, which may be located some distance from the gene. Some transcriptional regulator proteins are activators, stimulating transcription. Others are represssors, inhibiting transcription.



The basal transcription apparatus is capable of minimal levels of transcription; *transcriptional regulator proteins* are required to bring about normal levels of transcription. These proteins bind to a regulatory promoter, which is located upstream of the core promoter (**Figure 17.5**), and to *enhancers*, which may be located some distance from the gene. Some transcriptional regulator proteins are activators, stimulating transcription. Others are represssors, inhibiting transcription.

Transcriptional Activators and Coactivators

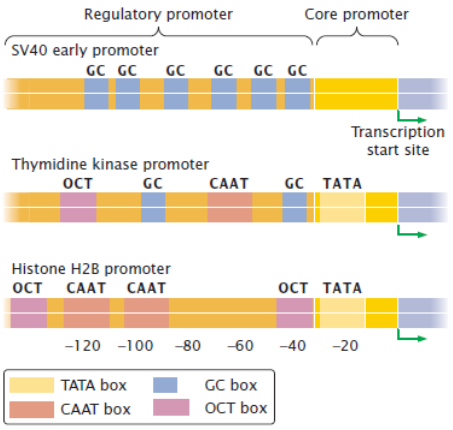
Transcriptional activator proteins stimulate and stabilize the basal transcription apparatus at the core promoter. The activators may interact directly with the basal transcription apparatus or indirectly through protein coactivators.

Some activators and coactivators, as well as the general transcription factors, also have acteyltransferase activity and so further stimulate transcription by altering chromatin structure. Transcriptional activator proteins have two distinct functions (see Figure 17.5). First, they are capable of binding DNA at a specific base sequence, usually a consensus sequence in a regulatory promoter or enhancer; for this function, most transcriptional activator proteins contain one or more DNA-binding motifs, such as the helix-turn-helix, zinc finger, and leucine zipper. A second function is the

ability to interact with other components of the transcriptional apparatus and influence the rate of transcription.

Within the regulatory promoter are typically several different consensus sequences to which different transcriptional activators can bind. Among different promoters, activator-binding sites are mixed and matched in different combinations (**Figure 17.6**), and so each promoter is regulated by a unique combination of transcriptional activator proteins.

Transcriptional activator proteins bind to the consensus sequences in the regulatory promoter and affect the assembly or stability of the basal transcription apparatus at the core promoter. One of the components of the basal transcription apparatus is a complex of proteins called the **mediator** (see Figure 17.5). Transcriptional activator proteins binding to sequences in the regulatory promoter (or enhancer, see next section) make contract with the mediator and affect the rateat which transcription is initiated. Some regulatory promoters also contain sequences that are bound by proteins that lower the rate of transcription through inhibitory interactions with the mediator.



17.6 The consensus sequences in the promoters of three eukaryotic genes illustrate the principle that different sequences can be mixed and matched in different combinations. A different transcriptional activator protein binds to each consensus sequence, and so each promoter responds to a unique combination of activator proteins.

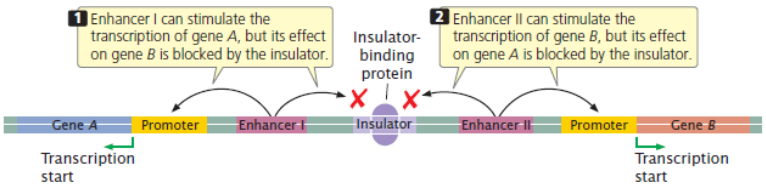
Transcriptional Repressors

Some regulatory proteins in eukaryotic cells act as repressors, inhibiting transcription. These repressors bind to sequences in the regulatory promoter or to distant sequences called *silencers*, which, like enhancers, are position and orientation independent. Unlike repressors in bacteria, most eukaryotic repressors do not directly block RNA polymerase. These repressors may compete with activators for DNA binding sites: when a site is occupied by an activator, transcription is activated, but, if a repressor occupies that site, there is no activation. Alternatively, a repressor may bind to sites

near an activator site and prevent the activator from contacting the basal transcription apparatus. A third possible mechanism of repressor action is direct interference with the assembly of the basal transcription apparatus, thereby blocking the initiation of transcription.

Enhancers and Insulators

Enhancers are capable of affecting transcription at distant promoters. For example, an enhancer that regulates the gene encoding the alpha chain of the T-cell receptor is located 69,000 bp downstream of the gene's promoter. Furthermore, the exact position and orientation of an enhancer relative to the promoter can vary. How can an enhancer affect the initiation of transcription taking place at a promoter that is tens of thousands of base pairs away? In many cases, regulator proteins bind to the enhancer and cause the DNA between the enhancer and the promoter to loop out, bringing the promoter and enhancer close to each other, and so the transcriptional regulator proteins are able to directly interact with the basal transcription apparatus at the core promoter (see Figure 17.5). Some enhancers may be attracted to promoters by proteins that bind to sequences in the regulatory promoter and "tether" the enhancer close to the core promoter. A typical enhancer is some 500 bp in length and contains 10 binding sites for proteins that regulate transcription. Most enhancers are capable of stimulating any promoter in their vicinities. Their effects are limited, however, by **insulators** (also called boundary elements), which are DNA sequences that block or insulate the effect of enhancers in a position-dependent manner. If the insulator lies between the enhancer and the promoter, it blocks the action of the enhancer; but, if the insulator lies outside the region between the two, it has no effect (**Figure 17.8**). Specific proteins bind to insulators and play a role in their blocking activity. Some insulators also limit the spread of changes in chromatin structure that affect transcription.



17.8 An insulator blocks the action of an enhancer on a promoter when the insulator lies between the enhancer and the promoter.

Regulation of Transcriptional Stalling and Elongation

Transcription in eukaryotes is often regulated through factors that affect the initiation of transcription, including changes in chromatin structure, transcription factors, and transcriptional regulatory proteins. Research indicates that transcription may also be controlled through factors that affect stalling and elongation of RNA polymerase after transcription has been initiated.

The basal transcription apparatus—consisting of RNA polymerase, transcription factors, and other proteins— assembles at the core promoter. When the initiation of transcription has taken place, RNA polymerase moves downstream, transcribing the structural gene and producing an RNA product. At some genes, RNA polymerase initiates transcription and transcribes from 24 to 50 nucleotides of RNA but then pauses or stalls. For example, stalling is observed at genes that encode **heat-shock proteins** in *Drosophila*—proteins that help to prevent damage from stressing agents such as extreme heat.

Heat-shock proteins are produced by a large number of different genes. During times of environmental stress, the transcription of all the heat-shock genes is greatly elevated. RNA polymerase initiates transcription at heat-shock genes in *Drosophila* but, in the absence of stress, stalls downstream of the transcription initiation site. Stalled polymerases are released when stress is encountered, allowing rapid transcription of the genes and the production of heat-shock proteins that facilitate adaptation to the stressful nvironment.

Stalling was formerly thought to take place at only a small number of genes, but recent research indicates that stalling is widespread and common throughout eukaryotic genomes. For example, stalled RNA polymerases were found at hundreds of genes in *Drosophila*. Factors that promote stalling have been identified; a protein called negative elongation factor (NELF) causes RNA polymerase to stall after initiation. Reducing the amount of NELF in cultured cells stimulates the elongation of RNA molecules at many genes with stalled RNA polymerases.

Table 17.1 A few response elements found in eukaryotic cells

Response Element	Responds to	Consensus Sequence
Heat-shock element	Heat and other stress	CNNGAANNTCCNNG
Glucocorticoid response element	Glucocorticoids	TGGTACAAATGTTCT
Phorbol ester response element	Phorbal esters	TGACTCA
Serum response element	Serum	CCATATTAGG

Coordinated Gene Regulation

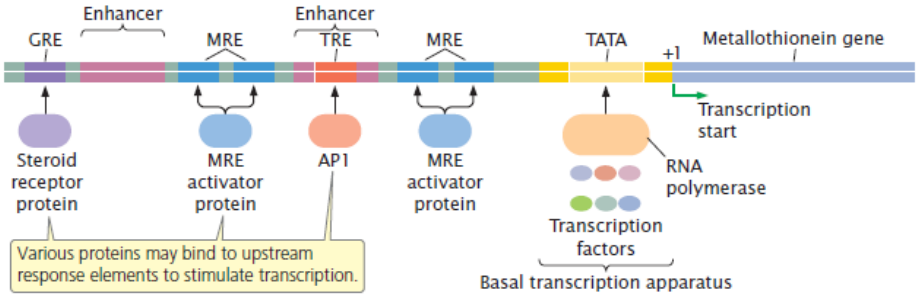
Although most eukaryotic cells do not possess operons, several eukaryotic genes may be activated by the same stimulus. Groups of bacterial genes are often coordinately expressed (turned on and off together) because they are physically clustered as an operon and have the same promoter, but coordinately expressed genes in eukaryotic cells are not clustered. How, then, is the transcription of eukaryotic genes coordinately controlled if they are not organized into an operon?

Genes that are coordinately expressed in eukaryotic cells are able to respond to the same stimulus because they have short regulatory sequences in common in their promoters or enhancers. For example, different eukaryotic heat-shock genes possess a common regulatory element upstream of their start sites. Such DNA regulatory sequences are called**response elements**; they are short sequences that typically contain consensus sequences (**Table 17.1**) at varying distances from the gene being regulated. The response elements are binding sites for transcriptional activators. A transcriptional activator protein binds to the response element and elevates transcription. The same response element may be present in different genes, allowing multiple genes to be activated by the same stimulus.

A single eukaryotic gene may be regulated by several different response elements. For example, the metallothionein gene protects cells from the toxicity of heavy metals by encoding a protein that binds to heavy metals and removes them from cells. The basal transcription apparatus assembles around the TATA box, just upstream of the transcription start site for the metallothionein gene, but the apparatus alone is capable of only low rates of transcription.

Other response elements found upstream of the metallothionein gene contribute to increasing its rate of transcription. For example, several copies of a metal response element (MRE) are upstream of the metallothionein gene (**Figure 17.9**). Heavy metals stimulate the binding of activator proteins to MREs, which elevates the rate of transcription of the metallothionein gene. Because there are multiple copies of the MRE, high rates of transcription are induced by metals. Two enhancers also are located in the upstream region of the metallothionein gene; one enhancer contains a response element known as TRE, which stimulates transcription in the presence of an activated protein called AP1. A third response element called GRE is located approximately 250 nucleotides upstream of the metallothionein gene and stimulates transcription in response to certain hormones.

This example illustrates a common feature of eukaryotic transcriptional control: a single gene may be activated by several different response elements found in both promoters and enhancers. Multiple response elements allow the same gene to be activated by different stimuli. At the same time, the presence of the same response element in different genes allows a single stimulus to activate multiple genes. In this way, response elements allow complex biochemical responses in eukaryotic cells.



17.9 Multiple response elements (MREs) are found in the upstream region of the metallothionein gene. The basal transcription apparatus binds near the TATA box. In response to heavy metals, activator proteins bind to several MREs and stimulate transcription. The TRE response element is the binding site for transcription factor AP1, which is stimulated by phorbol esters. In response to glucocorticoid hormones, steroid-receptor proteins bind to the GRE response element located approximately 250 nucleotides upstream of the metallothionein gene and stimulate transcription.

Some Genes Are Regulated by RNA Processing and Degradation

In bacteria, transcription and translation take place simultaneously. In eukaryotes, transcription takes place in the nucleus and the pre-mRNAs are then processed before moving to the cytoplasm for translation, allowing opportunities for gene control after transcription. Consequently, posttranscriptional gene regulation assumes an important role in eukaryotic cells. A common level of gene regulation in eukaryotes is RNA processing and degradation.

Gene Regulation Through RNA Splicing

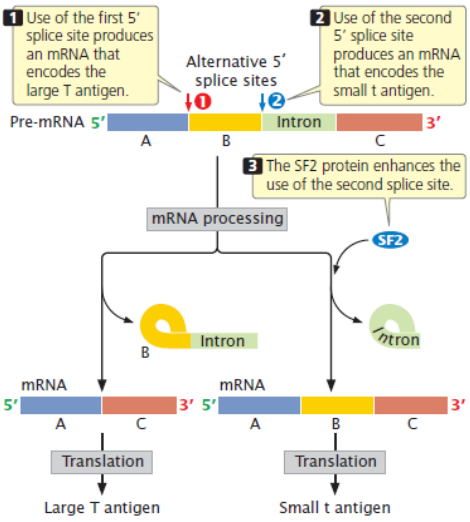
Alternative splicing allows a pre-mRNA to be spliced in multiple ways, generating different proteins in different tissues or at different times in development. Many eukaryotic genes undergo alternative splicing, and the regulation of splicing is probably an important means of controlling gene expression in eukaryotic cells.

Alternative splicing in the T-antigen gene

The T-antigen gene of the mammalian virus SV40 is a well studied example of alternative splicing. This gene is capable of encoding two different proteins, the large T and small t antigens. Which of the two proteins is produced depends on which of two alternative 5′ splice sites is used in RNA splicing (**Figure 17.10**). The use of one 5′ splice site produces mRNA that encodes the large T antigen, whereas the use of the other 5′ splice site (which is farther downstream) produces an mRNA encoding the small t antigen.

A protein called splicing factor 2 (SF2) enhances the production of mRNA encoding the small t antigen. Splicing factor 2 has two binding domains: one domain is an RNA-binding region and the other has alternating serine and arginine amino acids. These two domains are typical of **SR** (serine- and arginine-rich) **proteins**, which often play a role in regulating splicing. Splicing factor 2 stimulates

the binding of small nuclear ribonucleoproteins (snRNPs) to the 5′ splice site, one of the earliest steps in RNA splicing (see Chapter 14). The precise mechanism by which SR proteins influence the choice of splice sites is poorly understood. One model suggests that SR proteins bind to specific splice sites on mRNA and stimulate the attachment of snRNPs, which then commit the site to splicing.



17.10 Alternative splicing leads to the production of the small t antigen and the large T antigen in the mammalian virus SV40.

The Degradation of RNA

The amount of a protein that is synthesized depends on the amount of corresponding mRNA available for translation. The amount of available mRNA, in turn, depends on both the rate of mRNA synthesis and the rate of mRNA degradation.

Eukaryotic mRNAs are generally more stable than bacterial mRNAs, which typically last only a few minutes before being degraded. Nonetheless, there is great variability in the stability of eukaryotic mRNA: some mRNAs persist for only a few minutes; others last for hours, days, or even months. These variations can result in large differences in the amount of protein that is synthesized.

Cellular RNA is degraded by ribonucleases, enzymes that specifically break down RNA. Most eukaryotic cells contain 10 or more types of ribonucleases, and there are several different pathways of mRNA degradation. In one pathway, the 5′ cap is first removed, followed by 5′→3′ removal of nucleotides. A second pathway begins at the 3′ end of the mRNA and removes nucleotides in the 3′→5′ direction. In a third pathway, the mRNA is cleaved at internal sites.

Messenger RNA degradation from the 5′ end is most common and begins with the removal of the 5′ cap. This pathway is usually preceded by the shortening of the poly(A) tail. Poly(A)-binding proteins (PABPs) normally bind to the poly(A) tail and contribute to its stability-enhancing effect. The

presence of these proteins at the 3′ end of the mRNA protects the 5′ cap. When the poly(A) tail has been shortened below a critical limit, the 5′ cap is removed, and nucleases then degrade the mRNA by removing nucleotides from the 5′ end.

These observations suggest that the 5′ cap and the 3′ poly(A) tail of eukaryotic mRNA physically interact with each other, most likely by the poly(A) tail bending around so that the PABPs make contact with the 5′ cap (see Figure 15.18).

Much of RNA degradation takes place in specialized complexes called P bodies. However, P bodies appear to be more than simply destruction sites for RNA. Recent evidence suggests that P bodies can temporarily store mRNA molecules, which may later be released and translated. Thus, P bodies help control the expression of genes by regulating which RNA molecules are degraded and which are sequestered for later release. RNA degradation facilitated by small interfering RNAs (siRNAs) also may take place within P bodies (see next section).

Other parts of eukaryotic mRNA, including sequences in the 5′ untranslated region (5′ UTR), the coding region, and the 3′ UTR, also affect mRNA stability. Some short-lived eukaryotic mRNAs have one or more copies of a consensus sequence consisting of 5′-AUUUAUAA-3′, referred to as the AU-rich element, in the 3′ UTR. Messenger RNAs containing AU-rich elements are degraded by a mechanism in which microRNAs take part.

RNA Interference Is an Important Mechanism of Gene Regulation

The expression of a number of eukaryotic genes is controlled through RNA interference, also known as RNA silencing and posttranscriptional gene silencing (see Chapter 14). Recent research suggests that as much as 30% of human genes are regulated by RNA interference. Although many of the details

of this mechanism are still being investigated, RNA interference is widespread in eukaryotes, existing in fungi, plants, and animals. This mechanism is also widely used as a powerful technique for artificially regulating gene expression in genetically engineered organisms

Small Interfering RNAs and MicroRNAs

RNA interference is triggered by microRNAs (miRNAs) and small interfering RNAs (siRNAs), depending on their origin and mode of action (see Chapter 14). An enzyme called Dicer cleaves and processes double-stranded RNA to produce siRNAs or miRNAs that are from 21 to 25 nucleotides in length (**Figure 17.13**) and pair with proteins to form an RNA-induced silencing complex (RISC). The RNA component of RISC then pairs with complementary base sequences of specific mRNA

molecules, most often with sequences in the 3′ UTR of the mRNA. Small interfering RNAs tend to base pair perfectly with the mRNAs, whereas miRNAs often form less-than-perfect pairings.

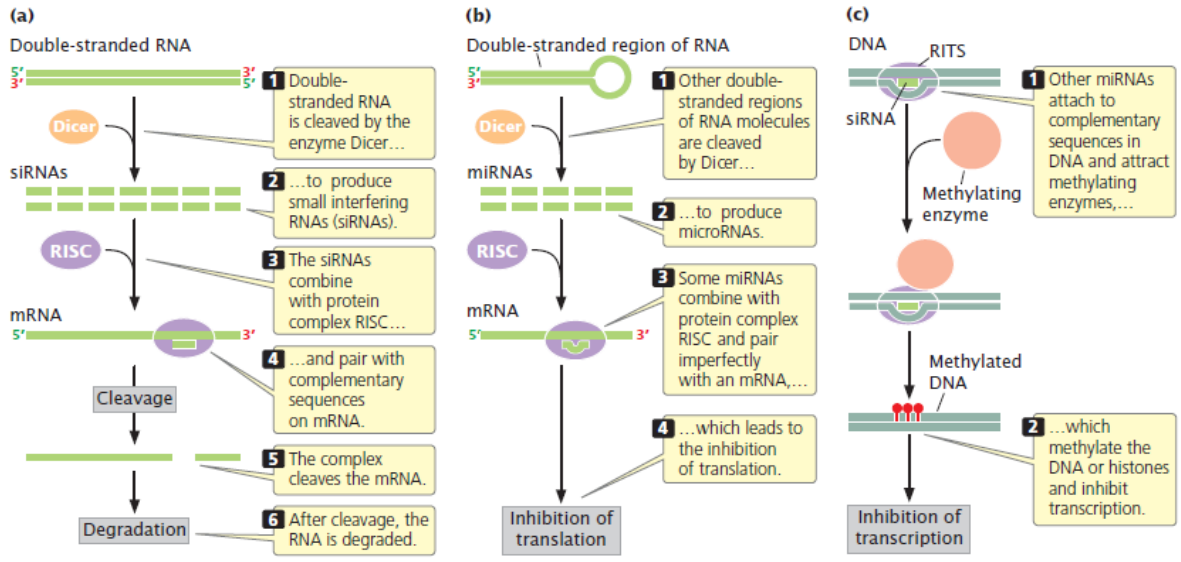
Mechanisms of Gene Regulation by RNA Interference

Small interfering RNAs and microRNAs regulate gene expression through at least four distinct mechanisms: (1)cleavage of mRNA, (2) inhibition of translation, (3) transcriptional silencing, or (4) degradation of mRNA.

RNA cleavage RISCs that contain an siRNA (and some that contain an miRNA) pair with mRNA molecules and cleave the mRNA near the middle of the bound siRNA (see Figure 17.13a). This cleavage is sometimes referred to as "Slicer activity." After cleavage, the mRNA is further degraded. Thus, the presence of siRNAs and miRNAs increase the rate at which mRNAs are broken down and decrease the amount of protein produced.

Inhibition of translation Some miRNAs regulate genes by inhibiting the translation of their complementary mRNAs (see Figure 17.13b). For example, an important gene in flower development in *Arabidopsis thaliana* is *APETALA2.* The expression of this gene is regulated by an miRNA that base pairs with nucleotides in the coding region of *APETALA2* mRNA and inhibits its translation.

The exact mechanism by which miRNAs repress translation is not known, but some research suggests that it can inhibit both the initiation step of translation and steps after translation initiation such as those causing premature termination. Many mRNAs have multiple miRNA-binding sites, and translation is most efficiently inhibited when multiple miRNAs are bound to the mRNA.



17.13 RNA silencing leads to the degradation of mRNA or to the inhibition of translation or transcription. (a) Small interfering RNAs (siRNAs) degrade mRNA by cleavage. (b) MicroRNAs (miRNAs) lead to the inhibition of translation. (c) Some small interfering RNAs (siRNAs) methylate histone proteins or DNA, inhibiting transcription.

Transcriptional silencing Other siRNAs silence transcription by altering chromatin structure. These siRNAs combine with proteins to form a complex called RITS (for RNA transcriptional silencing; see Figure 17.13c), which is analogous to RISC. The siRNA component of RITS then binds to its complementary sequence in DNA or an RNA molecule in the process of being transcribed and represses transcription by attracting enzymes that methylate the tails of histone proteins. The addition of methyl groups to the histones causes them to bind DNA more tightly, restricting the access of proteins and enzymes necessary to carry out transcription (see earlier section on histone modification). Some miRNAs bind to complementary sequences in DNA and attract enzymes that methylate the DNA directly, which also leads to the suppression of transcription (see earlier section on DNA methylation).

Slicer-independent degradation of mRNA A final mechanism by which miRNAs regulate gene expression is by triggering the decay of mRNA in a process that

does not require Slicer activity. For example, a shortlived mRNA with an AU-rich element in its 3′ UTR is degraded by an RNA-silencing mechanism. Researchers have identified an miRNA with a sequence that is complementary to the consensus sequence in the AU-rich element. This miRNA binds to the AU-rich element and, in a way that is not yet fully understood, brings about the degradation of the mRNA in a process that requires Dicer and RISC.

A Comparison of Bacterial and Eukaryotic Gene Control

Now that we have considered the major types of gene regulation in bacteria (see Chapter 16) and eukaryotes (this chapter), let's pause to consider some of the similarities and differences in bacterial and eukaryotic gene control.

1. Much of gene regulation in bacterial cells is at the level of transcription (although it does exist at other levels). Gene regulation in eukaryotic cells takes place at multiple levels, including chromatin structure, transcription, mRNA processing, RNA stability, RNA interference, and posttranslational control.

2. Complex biochemical and developmental events in bacterial and eukaryotic cells may require a cascade of gene regulation, in which the activation of one set of genes stimulates the activation of another set.

3. Much of gene regulation in both bacterial and eukaryotic cells is accomplished through proteins that bind to specific sequences in DNA. Regulatory proteins come in a variety of types, but most can be characterized according to a small set of DNA-binding motifs.

4. Chromatin structure plays a role in eukaryotic (but not bacterial) gene regulation. In general, condensed chromatin represses gene expression; chromatin structure must be altered before transcription can take place. Chromatin structure is altered by the modification of histone proteins, chromatinremodeling proteins, and DNA methylation.

5. Modifications to chromatin structure in eukaryotes may lead to epigenetic changes, which are changes that affect gene expression and are passed on to other cells or future generations.

6. In bacterial cells, genes are often clustered in operons and are coordinately expressed by transcription into a single mRNA molecule. In contrast, most eukaryotic genes typically have their own promoters and are transcribed independently.

Coordinate regulation in eukaryotic cells takes place through common response elements, present in the promoters and enhancers of the genes. Different genes that have the same response element in common are influenced by the same regulatory protein.

7. Regulatory proteins that affect transcription exhibit two basic types of control: *repressors* inhibit transcription (negative control); *activators* stimulate transcription (positive control). Both negative control and positive control are found in bacterial and eukaryotic cells.

8. The initiation of transcription is a relatively simple process in bacterial cells, and regulatory proteins function by blocking or stimulating the binding of RNA polymerase to DNA. In contrast, eukaryotic transcription requires complex machinery that includes RNA polymerase, general transcription factors, and transcriptional activators, which allows transcription to be influenced by multiple factors.

9. Some eukaryotic transcriptional regulatory proteins function at a distance from the gene by binding to enhancers, causing the formation of a loop in the DNA, which brings the promoter and enhancer into close proximity. Some distant-acting sequences analogous to enhancers have been described in bacterial cells, but they appear to be less common.

10. The greater time lag between transcription and translation in eukaryotic cells compared with that in bacterial cells allows mRNA stability and mRNA processing to play larger roles in eukaryotic gene regulation.

11. RNA molecules (antisense RNA) may act as regulators of gene expression in bacteria. Regulation by siRNAs and miRNAs, which is extensive in eukaryotes, is absent from bacterial cells.

REFERENCES

GENETICS: A conceptual approach, 4th Edition, Benjamin A. Pierce, W. H. Freeman and company England; 2006

Cell Biology, Genetics, Molecular Biology, Evolution and Ecology, P.S. Verma, V.K. Agarwal, S. Chand & Company Ltd, 2005

Principles of Molecular Biology, Veer Bala Rastogi, Medtech, 2016

Molecular Biology of the Cell. 4th edition. Alberts B, Johnson A, Lewis J, et al. New York: Garland Science; 2002.