

Gene Expression

We now know that genes encode **proteins** and proteins control the functions of a cell. Are all the genes in a cell expressed at the same time? Also, are all genes expressed all the time? No! This will not only lead to wastage of cellular energy but also affect the balance within a **cell**. This is why gene expression is regulated. How exactly are genes regulated? Let's find out.

Regulation Of Gene Expression

Protein synthesis begins at transcription, ends at translation and involves multiple steps. Therefore, regulation of gene **expression** can happen at any of these steps. In eukaryotes, gene regulation occurs at any of the following steps:

- Transcriptional level i.e. during the formation of the primary transcript.
- Processing level i.e. at the stage of splicing.
- During transport of mRNA from the nucleus to the cytoplasm.
- Translational level.

A great example of coordinated gene regulation is the development and differentiation of embryo into adult organisms. Metabolic, physiological and environmental conditions govern the regulation of gene expression. For example, *E. coli* uses lactose as a source of energy.

To do so, it synthesizes an enzyme called beta-galactosidase which hydrolyzes lactose into galactose and glucose. However, if there is no lactose around to be used as an energy source, the *E. coli* does not need to synthesize beta-galactosidase.

Prokaryotic Gene Regulation

In prokaryotes, the main site for regulation of gene expression is transcription initiation. Within a transcription unit, the activity of

RNA polymerase at the promoter is regulated by ‘accessory proteins’. These proteins affect the ability of RNA polymerase to recognize start sites. These proteins can act both positively (activators) or negatively (repressors).

In prokaryotic DNA, the accessibility of the promoter depends on the interaction of proteins with sequences called operators. In most operons, the operator is adjacent to the promoter elements. Moreover, in most cases, the operator has a repressor protein bound to it. Therefore, each operon has its own, specific operator and repressor. Let's understand this better using lac operon as an example.

The *Lac* Operon

Here, ‘lac’ refers to lactose. Francois Jacob and Jacques Monod were the first to elucidate the *lac* operon – a transcriptionally regulated system. *Lac* operon consists of a polycistronic structural gene regulated by a common promoter and regulatory genes. Such arrangements are common in bacteria and are called operons. Other examples include *trp* operon, *val* operon, *his* operon etc.

The *lac* operon has the following parts:

- One regulatory gene – The *i* gene where ‘i’ is derived from ‘inhibitor’. This gene codes for the repressor of the *lac* operon.
- Three structural genes –
 1. The *z* gene that codes for the enzyme beta-galactosidase that hydrolyzes lactose to glucose and galactose.
 2. The *y* gene codes for the enzyme permease that increases the permeability of the cell to beta-galactosides.
 3. The *a* gene codes for transacetylase.

Lactose metabolism requires gene products of all three genes mentioned above. Lactose, the substrate for the enzyme beta-galactosidase, regulates the switching on and off of the operon. Therefore, lactose is the inducer. Let's understand how lactose switches the operon on or off.

In the absence of lactose, the *i* gene synthesizes the repressor which then binds to the operator region of the operon. This prevents RNA polymerase from transcribing the genes (*z*, *y*, *a*) on the operon. Therefore, if there is no lactose, the operon does not synthesize genes

for its utilization. The action of the repressor on the *lac* operon is negative regulation.

In the presence of lactose, the repressor interacts with lactose and gets inactivated. Thus, RNA polymerase is free and can transcribe the genes in the operon. Therefore, if lactose is present, the operon synthesizes the genes for its utilization. Therefore, essentially, the presence of the substrate i.e. lactose regulates the synthesis of enzymes for its utilization.

Solved Example For You

Question: Match the columns –

Gene of *lac* operon

Produces

(i) *a* gene

(a) beta-galactosidase

(ii) *i* gene

(b) permease

(iii) *z* gene

(c) transacetylase

(iv) y gene

(d) repressor

Solution: The answers are: i \rightarrow c, ii \rightarrow d, iii \rightarrow a, iv \rightarrow b.

The DNA

You may have had a relative say to you, “You look exactly like your father” or “You have the same eyes as your mother!” Have you ever wondered what is the reason for this? It is because of your DNA! You get 50% of your DNA from your father and the other 50% from your mother. Want to learn more about the components and structure of this interesting molecule? Let’s dive in.

Structure of DNA

The genetic material in most organisms is DNA or Deoxyribonucleic acid; whereas in some viruses, it is RNA or Ribonucleic acid. A DNA molecule consists of two polynucleotide chains i.e. chains with multiple nucleotides. Let’s understand the structure of this chain in detail.

Structure Of Polynucleotide Chain

A nucleotide is made of the following components:

- Pentose sugar – A pentose sugar is a 5-carbon sugar. In case of DNA, this sugar is deoxyribose whereas, in RNA, it is ribose.
- Phosphate group
- Nitrogenous base – These can be of two types – Purines and Pyrimidines. Purines include Adenine and Guanine whereas pyrimidines include Cytosine and Thymine. In RNA, thymine is replaced by Uracil.

Nitrogenous base + pentose sugar (via N-glycosidic linkage) =
Nucleoside.

Nucleoside + phosphate group (via phosphoester linkage) =
Nucleotide.

Nucleotide + Nucleotide (via 3'-5' phosphodiester linkage) =
Dinucleotide.

Many nucleotides linked together = Polynucleotide.

A polynucleotide has a free phosphate group at the 5' end of the sugar and this is called the 5' end. Similarly, the sugar also has a free 3'-OH group at the other end of the polynucleotide which is called the 3' end. The backbone of a polynucleotide chain consists of pentose sugars and phosphate groups; whereas the nitrogenous bases project out of this backbone.

Polynucleotide chains of DNA and its components [Source: Wikimedia Commons]

Check out our other article on [DNA](#) here.

Double Helix Structure

DNA is a long polymer and therefore, difficult to isolate from cells in an intact form. This is why it is difficult to study its structure.

However, in 1953, James Watson and Francis revealed the ‘double helix’ model of the structure of DNA, based on X-ray diffraction data from Maurice Wilkins and Rosalind Franklin.

This model also reveals a unique property of polynucleotide chains – Base pairing. It refers to the hydrogen bonds that connect the nitrogen bases on opposite DNA strands. This pairing gives rise to complementary strands i.e. if you know the sequence of bases on one strand, you can predict the bases on the other strand. Additionally, if each DNA strand acts as a template for synthesis (parent) of a new strand, then the new double-stranded DNA (daughters) produced are identical to the parental DNA strand.

Salient Features of DNA Double-Helix

- It consists of two polynucleotide chains where the sugar and phosphate group form the backbone and the nitrogenous bases project inside the helix.

- The two polynucleotide chains have anti-parallel polarity i.e. if one strand has $5' \rightarrow 3'$ polarity, the other strand has $3' \rightarrow 5'$ polarity.
- The bases on the opposite strands are connected through hydrogen bonds forming base pairs (bp). Adenine always forms two hydrogen bonds with thymine from the opposite strand and vice-versa. Guanine forms three hydrogen bonds with cytosine from the opposite strand and vice-versa. Therefore, a purine always pairs with a pyrimidine on the other strand, giving rise to a uniform distance between the two strands of the helix.
- The two strands coil in a right-handed fashion. Each turn of the helix is 3.4nm (or 34 Angstrom units) consisting of 10 nucleotides. These nucleotides are at a distance of 0.34nm (or 3.4 Angstrom units).
- The helix is stable because of the base pairs that stack over one another and hydrogen bonds that hold the bases together.

DNA double helix [Source: Wikimedia Commons]

Packaging of DNA Helix

If you calculate the length of DNA in a typical mammalian cell, it is approximately 2.2 meters. The dimension of a typical nucleus is only about 10^{-6} meters! Then, how does such a long polymer fit in the nucleus of a cell?

Prokaryotes like *E. coli*, do not have a defined nucleus. Here, the negatively-charged DNA is held together in large loops by positively-charged proteins in a structure called ‘nucleoid’. In

Eukaryotes, however, the organization of DNA in the nucleus is much more complex and is as follows:

- The negatively-charged DNA is wrapped around a positively-charged histone octamer i.e. a unit of 8 histone molecules. This forms a ‘Nucleosome’. Histones are positively-charged proteins that are rich in basic amino acids – arginines and lysines. A typical nucleosome has 200bp of DNA helix.
- Many nucleosomes join together to form a thread-like structure – Chromatin in the nucleus. The nucleosomes in chromatin appear as ‘beads-on-string’ under the electron microscope.
- The chromatin is packaged to form chromatin fibres which are further coiled and condensed to form chromosomes. The higher level packaging of chromatin requires another set of proteins – Non-histone Chromosomal (NHC) proteins.

Nucleosome Structure [Source: Wikimedia Commons]

Note: Euchromatin is the region of chromatin that is loosely packed and therefore stains lightly; whereas Heterochromatin is the densely packed region and therefore stains dark.

Solved Example For You

Q1: In DNA, adenine pairs with thymine. What does adenine pair with, in RNA?

- a. Cytosine
- b. Uracil
- c. Thymine
- d. Guanine

Solution: The answer is ‘b’. Adenine pairs with Uracil in RNA.

The Genetic Material

Today, we all know that DNA is the genetic material that carries information from generation to generation. But, have you ever wondered how and when DNA was discovered? What experiments and observations led to the discovery of DNA? Let's explore this fun journey of the discovery of the genetic material here.

Discovery of DNA

Scientists had narrowed down that the genetic material was on chromosomes in the nucleus of a cell. However, the exact molecule was discovered only much later. Let's take a look at the series of experiments that scientists undertook that brought us closer to the discovery of DNA.

Frederick Griffith

While working with *Streptococcus pneumoniae* (the bacterium that causes pneumonia) in 1928, Frederick Griffith observed a miraculous transformation in this bacterium. When you grow this bacterium on a

culture plate, some produce shiny colonies (denoted as 'S') and some produce rough colonies (denoted as 'R').

The S strain bacteria have a polysaccharide coat which gives rise to smooth, shiny colonies. The R strain lacks this coat and hence, it gives rough colonies. Also, the S strain is virulent and causes pneumonia; while the R strain is non-virulent. He performed the following experiment with these strains and saw different observations.

1. S strain → Inject into mice → Mice develop pneumonia and die.
2. R strain → Inject into mice → Mice live.
3. Heat-killed S strain → Inject into mice → Mice live. (Griffith found that heating kills the bacteria).
4. Heat-killed S strain + R strain → Inject into mice → Mice die.

Frederick Griffith's experiment [Source: Wikimedia Commons]

- Observations – Not only did the mice injected with the heat-killed S strain + R strain die, but Griffith also recovered live S strain bacteria from these dead mice.
- Conclusions – He concluded that this was because the R strain had somehow been ‘transformed’ by the heat-killed S strain. This he argued was due to the transfer of a ‘transforming principle’ from the S strain to the R strain, which made the R strain virulent. Although significant, his observations did not identify the biochemical nature of the transforming principle.

Oswald Avery, Colin MacLeod & Maclym McCarty

Avery, MacLeod, and McCarty, together set out to determine the biochemical nature of the ‘transforming principle’ identified by Griffith. These people purified DNA, RNA, and **proteins** from the heat-killed S strain and determined which macromolecule converted the R strain into the S strain.

- Experiment – They first treated the heat-killed S strain with proteases to break down proteins. Subsequently, they treated it with RNAses and then DNAses to break down RNA and DNA, respectively.
- Observations – Both protease and RNase treatments did not affect the transformation of the R strain into the virulent one. Finally, treatment with DNAses inhibited the transformation of the R strain.
- Conclusions – They concluded that the genetic material is not protein or RNA, but it is DNA. However, this discovery was not accepted by all biologists.

Alfred Hershey & Martha Chase

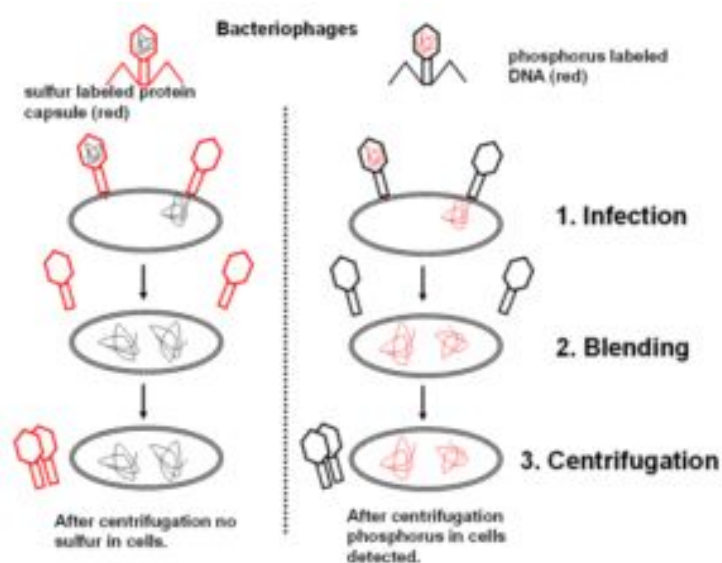
Much earlier, scientists believed that the genetic material was protein. In 1952, Hershey & Chase were the ones to conclusively prove that DNA is the genetic material. They worked with bacteriophages – **viruses** that infect bacteria. A bacteriophage attaches and delivers its genetic material into a bacterial cell, where it generates more virus particles. Hersey & Chase used bacteriophages to experiment as follows:

- Labelling – Some viruses were grown on a medium containing radioactive **phosphorus** and some on a medium with radioactive sulfur.
- Viruses – grown on radioactive phosphorus have radioactive DNA but not protein since DNA contains phosphorus but protein does not. Contrarily, viruses grown on radioactive **sulfur** have radioactive protein but not DNA since DNA does not contain sulfur.
- Infection – The radioactive phages were then allowed to infect the bacteria – *E. coli*.
- Blending and Centrifugation – As the infection progressed, the viral coats were removed from the bacteria by blending. Then,

centrifugation was used to separate the viral particles from the bacteria.

Observations – Bacteria infected with viruses that have radioactive DNA, were radioactive, while bacteria infected with viruses that have radioactive protein, were not radioactive.

Conclusions – This experiment conclusively showed that DNA is the genetic material transferred from virus to bacteria, and not protein.



Hershey and Chase experiment [Source: Wikimedia Commons]

Properties Of Genetic Material

For a **molecule** to act as the genetic material, it should have the following characteristics:

- Be capable of replication i.e. create its own replica.
- It should be stable, structurally and chemically.
- It must have the scope for slow changes (**mutations**) to evolve.
- Be expressed in the form of ‘Mendelian Characters’.

Although DNA is the genetic material in most organisms, in some viruses, RNA is the genetic material. In fact, according to studies, RNA was the first genetic material. But, since RNA is unstable, DNA evolved from RNA with chemical modifications, making it more stable and more fit to carry genetic information.

Solved Example For You

Q1: Hershey and Chase labelled viral DNA with which radioactive **element** in their experiment?

- a. Potassium
- b. Sulfur
- c. Phosphorus

d. [Calcium](#)

Sol: The answer is 'c'. They used radioactive phosphorus to label DNA since DNA contains phosphorus.

Human Genome Project and DNA Fingerprinting

Did you know that scientists have determined the complete [DNA](#) sequence of humans! Yes, it's true, through an ambitious project called the Human Genome Project (HGP). Also, did you know that just like your fingerprint, you also have a DNA fingerprint that is unique to you! Want to know more about these concepts? Let's find out.

Human Genome Project

The Human Genome Project (HGP) was a mega project launched in the year 1990. The advances in genetic engineering techniques have made this project possible. The aims of this project reveal the magnitude and the requirements of this project.

The human genome (i.e. the complete set of genes) has approximately 3×10^9 base pairs. If the cost of sequencing is US \$ 3 per base pair, then the cost of the entire project would be approximately US \$ 9 billion! Moreover, let's say the sequencing data were to be stored in books. Then if each page has 1000 letters and each book has 1000 pages, we will need 3300 such books to store the **genetic** information from a single cell!

This large amount of data would also need computational devices with high speed to store, retrieve and analyze the data. Therefore, HGP aided the rapid development of another field in **biology** – Bioinformatics.

Goals Of HGP

- Identify all the genes in the human genome (approximately 20,000-25,000 genes).
- Provide a complete and accurate sequence of the 3 billion base pairs that make up the human genome.
- Store all the sequencing data in databases.
- Develop new tools to obtain and analyze data and make the information widely available.

- Necessitate technology transfer to other sectors like industries.
- Address the ethical, legal and social implications of the project on society.

HGP was a 13-year project, coordinated by the National Institute of [Health](#) (NIH) and the U.S. Department of [Energy](#). It involved contributions from other countries too such as Japan, Germany, China, France etc. The benefits of this project are that it can lead to revolutionary new ways to diagnose, treat and prevent human diseases.

Besides the human genome, information about the genomes of non-human organisms can also be very helpful. We can understand their natural capabilities and apply them towards solving problems in human healthcare, [agriculture](#), energy production etc. Therefore, scientists have also sequenced many non-human organisms such as bacteria, yeast, fruit fly, plants etc.

Methodologies Of HGP

HGP involved two major approaches:

- Expressed Sequence Tags (ESTs) – This approach focussed on identifying all genes expressed as RNA.
- Sequence Annotation – This blind approach involved sequencing the whole genome (coding and non-coding) and later assigning functions to the different regions.

DNA sequencing involves the following steps:

- First, total DNA is isolated from a cell and converted into random, small-size fragments since it is difficult to sequence long pieces of DNA. These fragments are then cloned into a suitable host (bacteria or yeast) using special vectors such as Bacterial Artificial Chromosomes (BAC) or Yeast Artificial Chromosomes (YAC). This amplifies each DNA fragment so that it can be sequenced easily.
- Next, the fragments are sequenced using automated DNA sequencers. These sequences work on the principle of Frederick Sanger's method.
- Special computer-based programs are used to arrange and align the DNA sequences based on overlapping regions present in them.

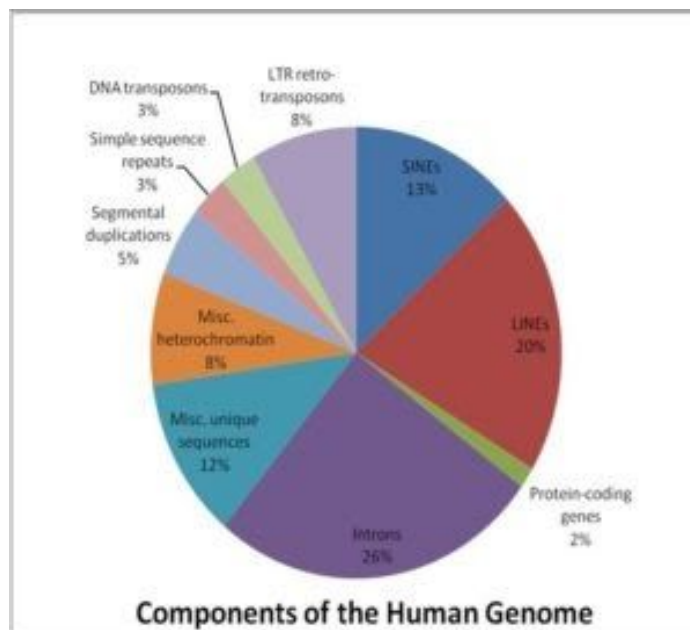
- Subsequently, the sequences are annotated and assigned to each chromosome.

This is a time-consuming process. Therefore, the sequence of chromosome 1 (the last chromosome to be sequenced) was completed only in May 2006.

Salient Features of Human Genome

- There are 3164.7 million nucleotide bases in the human genome.
- An average gene has 3000 bases. However, sizes vary greatly, with the largest human gene being ‘dystrophin’ that has 2.4 million bases.
- The original estimate of the number of genes was 80,000 to 1,40,000 genes. However, HGP gave an estimate of about 30,000 genes. About 99.9% nucleotide bases are the same in all people.
- For over 50% of the discovered genes, the functions are unknown.
- Less than 2% of the genome codes for proteins.

- Repeated sequences form a large part of the human genome.
- Stretches of DNA sequences that are repeated many times (sometimes 100 to 1000 times) are repetitive sequences. Although they don't code for proteins, they shed light on chromosome structure, evolution, and dynamics.
- Chromosome 1 has the most number of genes (2968), and chromosome Y has the least (231).
- HGP has identified 1.4 million locations with single base DNA differences in humans. This information will revolutionize the identification of disease-associated sequences and tracking of human history.



Components of the Human Genome [Source: Wikimedia Commons]

Applications of HGP and Future Challenges

The need to derive meaningful knowledge from genomic sequences and better understand biological systems will drive future research. This enormous task will require the coordinated effort of scientists from various fields.

A major impact of HGP is providing a radically new approach in biological research. Earlier, researchers studied one or a few genes at a time. Now, with new technologies and whole genome sequences, they can study all the genes in a genome i.e. all the transcripts in a [tissue](#) or

organ. They can also study how thousands of genes work together in networks to make a system function.

DNA Fingerprinting

As we know, 99.9% of nucleotide bases are the same in all humans. However, there are some differences in DNA sequences among people, which make them unique. This is their DNA fingerprint. How do we determine these differences? If we compare the whole DNA sequences of two individuals, it'll take way too long. DNA fingerprinting is a quicker way to compare the sequences of two individuals.

This technique involves identifying differences in the repetitive DNA regions. The peaks on a density gradient centrifugation help to separate the repetitive part from the bulk DNA. Here, the bulk DNA forms a major peak, while the small peaks are called satellite DNA.

Satellite DNA is classified into micro-satellites and mini-satellites based on multiple factors such as – base composition (A:T rich or G:C rich), number of repetitive units, length of segment etc. These sequences, don't code for any protein but are abundant in the human genome. They also show a high degree of polymorphism i.e.

differences in DNA sequence and therefore, form the basis of DNA fingerprinting.

DNA from every tissue such as hair follicle, saliva, skin, bone etc show the same degree of polymorphism. Thus, these are very important as an identification tool in forensic applications. Moreover, since polymorphisms are passed on from parents to children, this fingerprinting technique is also the basis of paternity testing.

Let's understand exactly what polymorphisms are.

Polymorphism

Polymorphisms are variations at the genetic level that arise due to mutations. In an individual, new mutations can arise either in somatic cells or germ cells i.e. cells that generate sperm and ovum. If the germ cell mutation doesn't affect the individual's ability to reproduce, then it is passed on to the next generation and thus, spreads in the population.

DNA polymorphism is an inheritable mutation observed at a high frequency in a population. The probability of these variations is higher on non-coding DNA since mutations in them will not impact an

individual's reproductive ability. This then passes from generation to generation and is one of the basis of variation in human evolution. Polymorphisms can be changes in a single nucleotide or large scale changes.

Technique

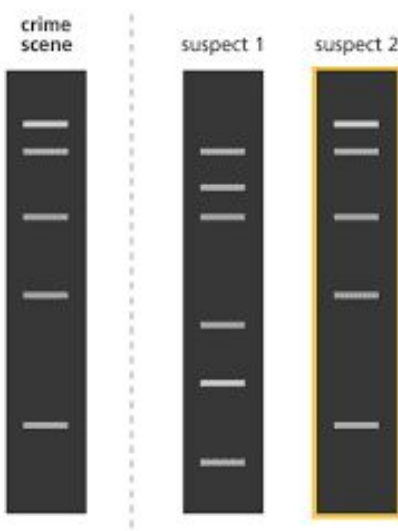
Alec Jeffreys initially developed the technique of DNA fingerprinting using a satellite DNA that shows a very high degree of polymorphism, as a probe. It is called Variable Number of Tandem Repeats (VNTR). VNTR belongs to the class of mini-satellites. Here, a small DNA sequence is arranged in many copies. The copy number varies between individuals and the number of repeats shows a high degree of polymorphism.

The technique of DNA fingerprinting involves Southern blot hybridization using radiolabelled VNTR as a probe. The steps are:

- Sample collection
- DNA isolation.
- DNA **digestion** using restriction endonucleases.
- Separation of DNA fragments using electrophoresis.

- Blotting (transferring) of separated DNA fragments on to synthetic membranes like nylon or nitrocellulose.
- Hybridization with the labelled VNTR probe.
- Detection of the hybridized DNA fragments using autoradiography.

The size of VNTR ranges from 0.1 to 20 kilobases. Therefore, the autoradiogram results show bands of multiple sizes. These bands give a characteristic pattern which differs between individuals except for monozygotic twins. Further, polymerase chain reaction (PCR) increases the sensitivity of fingerprinting i.e. DNA from a single cell is enough to perform fingerprinting.



DNA fingerprint [Source: flickr]

Apart from forensic science and paternity testing, this technique is also useful in determining population and genetic diversities. Therefore, many different probes are used currently to generate DNA fingerprints.

Solved Example For You

Question: Which of the following statements is false?

- Less than 2% of the human genome codes for proteins.
- Chromosome Y has the most genes.
- An average gene in the human genome has 3000 bases.
- Repeated sequences make up a large part of the human genome.

Solution: Statement 'b' is false because chromosome Y has the fewest genes (231).

Replication

We know that existing cells transfer genetic information to the new cells during cell division. What do you think will happen if each new cell receives only half the set of genetic information? Yes, the cell will not be able to function properly. This is why the DNA in an existing cells needs to be doubled (or copied) so that each new cell gets the full set of instructions. This process is DNA replication. Let's learn more about it.

Meselson And Stahl Experiment

Along with the double-helix structure of DNA, Watson and Crick also proposed a scheme for DNA replication. They suggested that the two DNA strands would separate and become a template for synthesis of complementary DNA strands. Therefore, each new DNA molecule would have one parental and one new DNA strand.

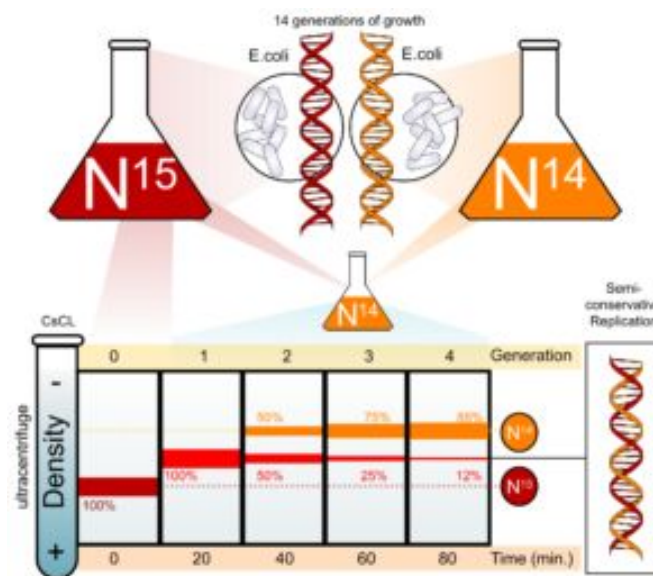
This scheme was referred to as semiconservative DNA replication. In 1958, Matthew Meselson and Franklin Stahl performed the following experiment to prove this mode of replication:

Experiment

- Meselson and Stahl grew *E.coli* on a medium that contains $^{15}\text{NH}_4\text{Cl}$ as the only nitrogen source for many generations [Note: ^{15}N is the heavy isotope of nitrogen]. As a result, all newly synthesized DNA had ^{15}N which can be differentiated from normal DNA by centrifugation in a cesium chloride (CsCl) density gradient.
- They then transferred the cells to a medium containing normal $^{14}\text{NH}_4\text{Cl}$.
- As the cells multiplied, they collected samples at time intervals (20mins, 40mins, 60mins etc). This is because *E. coli* cells divide every 20 minutes. They then extracted the double-stranded DNA from the samples and separated them on CsCl gradients to measure the DNA densities.

Results

- DNA extracted after 20 minutes (one generation) in the $^{14}\text{NH}_4\text{Cl}$ medium had an intermediate density. This is because it contained one parental DNA strand with the heavy ^{15}N and one new DNA strand with the light ^{14}N to give $^{15}\text{N}^{14}\text{N}$.
- DNA extracted after 40 minutes (two generations) in the $^{14}\text{NH}_4\text{Cl}$ medium showed equal amounts of intermediate density and light density. This is because it contained equal amounts of the hybrid $^{15}\text{N}^{14}\text{N}$ DNA (intermediate) and $^{14}\text{N}^{14}\text{N}$ DNA (light).



The Machinery Of DNA Replication

DNA Polymerase

It is the main enzyme needed for DNA replication. It uses DNA as a template to catalyze the polymerization of deoxynucleotides. DNA Polymerases should have the following properties:

- Highly efficient to carry out polymerization of a large number of nucleotides in a very short time. The process of replication in *E. coli*, that has only 4.6×10^6 bp (base pairs) is completed within 18 minutes. This indicates that the average rate of polymerization is approximately 2000 bp per second.
- High degree of accuracy since any mistake will lead to mutations.

Deoxyribonucleoside triphosphates

They act as substrates for polymerization. Replication is also a very energy-expensive process. Therefore, the triphosphates also provide energy for polymerization.

Origin of replication

Replication cannot start randomly at any place in DNA. It starts at a specific place in *E.coli* DNA called the ‘origin of replication’.

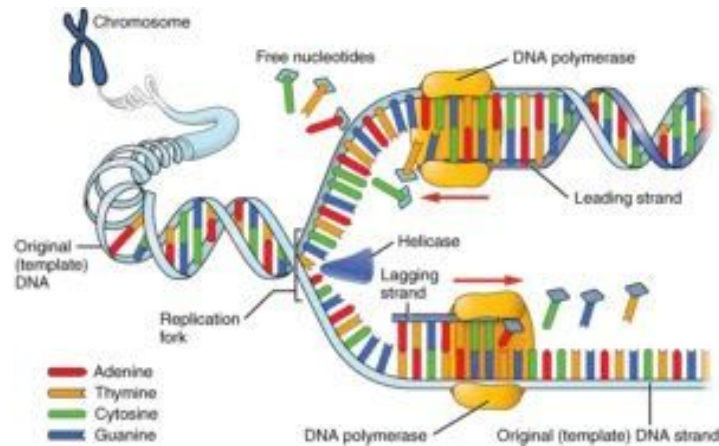
Replication fork

Since it will require a lot of energy to separate the two DNA strands for their entire length, replication usually starts within a small opening in the DNA helix. This is the ‘replication fork’. DNA polymerase can function only in one direction $5' \rightarrow 3'$. Therefore, replication of the DNA strand with polarity $3' \rightarrow 5'$ is continuous.

This is also called the ‘leading strand’. While on the DNA strand with polarity $5' \rightarrow 3'$, it is discontinuous and results in the formation of fragments of new DNA. This strand is the ‘lagging strand’.

DNA Ligase

This is the enzyme that joins the discontinuously synthesized fragments of DNA. In eukaryotes, DNA replication occurs in the S phase of the cell cycle. Moreover, the replication of DNA and cell division should be highly coordinated. If a cell fails to divide after DNA replication, it results in polyploidy – a condition in which a cell has more than the normal number of chromosomes.



DNA Replication [Wikimedia Commons]

Solved Example For You

Q1: In the Meselson and Stahl experiment, if DNA was extracted after 60 minutes in the $^{14}\text{NH}_4\text{Cl}$ medium, what would be the DNA densities?

- a. 75% light and 25% intermediate
- b. 25% light and 75% intermediate
- c. 50% light and 50% intermediate
- d. 15% light and 85% intermediate

Sol: The answer is option 'a'. After 60 minutes, i.e. three generations in the $^{14}\text{NH}_4\text{Cl}$ medium, the extracted DNA will show 75% light density and 25% intermediate density.

Transcription

How does DNA provide the information required by a cell to function properly? DNA or genes give instructions for the synthesis of proteins, which constitute most of the substances in our body, such as enzymes, hormones, antibodies etc. Transcription is the first step in the synthesis of proteins from DNA. Let's learn about this process in detail.

Transcription Unit

The central dogma of molecular biology is as follows:



Transcription is the copying of genetic information from one DNA strand into RNA by RNA polymerase. Like replication, it is also governed by the principle of complementarity. However, unlike replication, only one DNA strand is copied to RNA in transcription. Why is this?

Firstly, if both DNA strands act as templates, it will result in two different RNA sequences and in turn give rise to two different

proteins. This complicates the transfer of genetic information.

Secondly, the two RNA sequences produced will be complementary to each other and produce a double-stranded RNA. This will prevent RNA from being translated into protein. Now, let's look at the different regions of a transcription unit in DNA.

1. Structural Gene

The two DNA strands within the structural gene have different names. Since RNA polymerase catalyzes polymerization in only one direction $5' \rightarrow 3'$, the strand with $3' \rightarrow 5'$ polarity becomes the template strand. The other strand with $5' \rightarrow 3'$ polarity is displaced during transcription and is called the coding strand even though it does not code for anything. In a transcription unit, the promoter and terminator regions lie on either side of the structural gene.

2. Promoter

It is a DNA sequence located towards the $5'$ end (upstream) of the coding strand. It is the binding site for RNA polymerase and is the site that tells the polymerase to start transcription. Additionally, the presence of the promoter defines the template and coding strand in a transcription unit.

3. Terminator

It is a DNA sequence located towards the 3' end (downstream) of the coding strand. It provides the stop signal and defines the end of transcription. Additional regulatory sequences may be present upstream or downstream of the promoter.

Transcription unit [Source: Wikimedia Commons]

The Gene

The functional unit of inheritance is a gene. Although genes are located on DNA, it is difficult to define a gene in terms of DNA sequence. A DNA sequence that codes for tRNA (transfer RNA) or rRNA (ribosomal RNA) is also a gene.

A cistron is a segment of DNA that codes for a polypeptide (a polymer of amino acids). A cistron can be polycistronic (mostly in prokaryotes and bacteria), i.e. it can code for several proteins. It can also be monocistronic (mostly in eukaryotes) i.e. it codes for a single protein. The monocistronic genes in eukaryotes consist of coding sequences

called exons and intervening sequences called introns. Exons appear in mature or processed RNA whereas introns do not.

Types Of RNA

There are three major types of RNAs in bacteria:

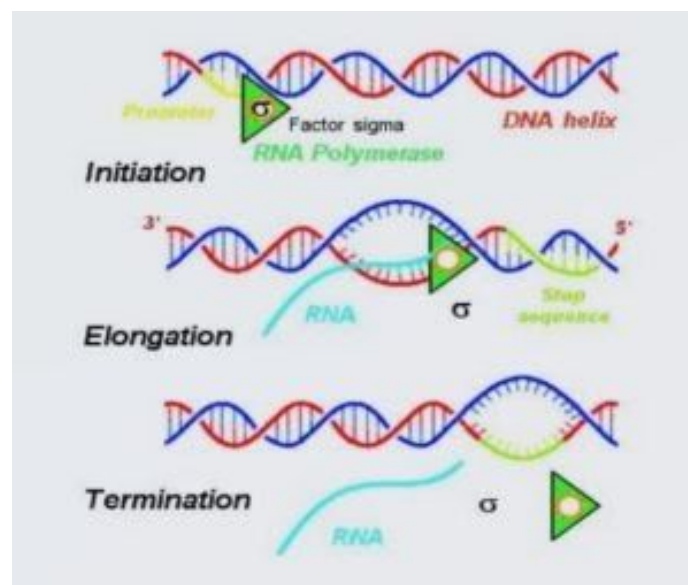
- Messenger RNA (mRNA) – It provides the template to make protein.
- Transfer RNA (tRNA) – It reads the genetic code and transfers amino acids for protein synthesis.
- Ribosomal RNA (rRNA) – It has a structural and catalytic role in protein synthesis.

The Process Of Transcription

Transcription has the following steps:

- Initiation: Here, RNA polymerase binds to the promoter region and transiently binds to the ‘initiation factor’ to initiate transcription.

- **Elongation:** This is the step where the RNA strand starts getting longer. RNA polymerase “walks” along one strand of DNA. For every nucleotide recognized on the DNA template, it adds a complementary RNA nucleotide to the growing RNA transcript.
- **Termination:** Transcription stops once the RNA polymerase reaches the terminator region. At this region, the RNA transcript and the RNA polymerase, both fall off. RNA polymerase transiently associates with the ‘termination factor’ to stop transcription.



Complexity In Eukaryotic Transcription

In bacteria, since the mRNA does not need to be processed and since transcription and translation occur in the same cell compartment, the two processes can occur simultaneously. Also, the RNA Polymerase catalyzes transcription of all kinds of RNA. Eukaryotes, however, differ and show two main complexities. There are 3 types of RNA polymerases –

- RNA Polymerase I that transcribes rRNA.
- Type II that transcribes a precursor of mRNA – heterogenous nuclear RNA (hnRNA).
- RNA Polymerase III that transcribes tRNA and small nuclear RNAs (snRNA).

The primary transcript in eukaryotes is non-functional since it contains exons and introns. It undergoes splicing, a process that removes introns and joins the exons together in a specific order. The precursor hnRNA undergoes additional processing called capping and tailing.

An unusual nucleotide is added to the 5' end of hnRNA during capping. In tailing, 200-300 adenylate residues are added to the 3' end

of hnRNA. This fully processed hnRNA, called mRNA is now transported out of the nucleus for translation.

Post-transcriptional modifications in Eukaryotes [Source: Wikimedia Commons]

Solved Example For You

Q1: If the sequence of the coding strand is 5'-TACGTACGTACGTA-3', what is the sequence of the transcribed RNA?

Sol: The answer is 5'-UACGUACGUACGUA-3'. RNA has the same sequence as the coding strand, except in place of thymine it has uracil.

Translation

We have learned that the process of transcription converts the information in DNA to RNA. How is the information in the RNA then used to synthesize proteins? Transcription is followed by a process called translation which involves decoding the instructions in the RNA to make proteins. How exactly does this work? Let's find out below.

Genetic Code

In replication and transcription, a nucleic acid is copied to make another nucleic acid. Both these processes work on the basis of complementarity. Translation, however, converts genetic information from a polymer of nucleotides to a polymer of amino acids. This is not based on complementarity. Then, how does this process take place?

You may have seen barcodes on items in a grocery store. These barcodes are used to identify the item. Scanning the barcode helps the store to track that item. In the same way, there also exists a genetic code. It is a set of rules on the mRNA that tells the cell which amino acid should be added to the polymer.

The process of deciphering this genetic code was extremely challenging and involved multiple scientists from different areas. Let's learn who they were and how they contributed.

- George Gamow – a physicist who proposed that if 4 bases (adenine, guanine, cytosine, uracil) have to code for 20 amino acids, the genetic code must be a combination of three bases (a triplet). According to this, a permutation combination of 4^3 ($4 \times 4 \times 4$) will give 64 codons, which is more than the number of amino acids.
- Har Gobind Khorana – a biochemist who developed the chemical method to synthesize RNA molecules with defined combination of bases. These molecules were instrumental in deciphering the genetic code.
- Marshall Nirenberg – a biochemist and geneticist who developed a cell-free system for protein synthesis that helped to finally decipher the genetic code.
- Severo Ochoa – a biochemist who discovered an enzyme that allowed him to synthesize RNA with defined sequences, without the use of a DNA template.

The work of all these scientists finally gave a checkerboard for the genetic code as shown below:

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

The codons for different amino acids [Source: Wikimedia Commons]

Salient Features of the Genetic Code

- Each codon is a triplet of bases. There are 64 codons in total, of which 61 code for amino acids while 3 act as stop codons during translation.
- One codon codes for only one amino acid. Therefore, it is specific and unambiguous.
- Some amino acids are coded for by more than one codon. For example, GUU, GUC, GUA, and GUG – all code for valine (Val). Therefore, the code is degenerate.

- The codons on the mRNA are read in a continuous manner, without any punctuations.
- The genetic code is universal i.e. from bacteria to humans, the code UUU refers to phenylalanine (Phe). However, there are some exceptions to this rule, such as mitochondrial codons.
- The codon AUG has dual functions. It codes for the amino acid methionine (Met) and is also the start/initiator codon.

The Genetic Code And Mutations

You may have already learned about mutations that involve large deletions or rearrangement of segments of DNA. These result in loss or gain of a gene and therefore, a function. What about mutations or changes in a single base pair of a codon? These are called point mutations. A classic example is a point mutation in a gene for beta globin chain.

A change in a single base pair in this gene changes the amino acid glutamine to valine. This results in a disease called sickle cell anaemia. There are other types of mutations too. Let's use the

following statement with three-letter words like the genetic code, as an example:

TOM HAS BIG TOE

- Now, insert one letter in the middle – TOM HAS **O**BI GTO E
- Now, insert two letters – TOM HAS **ONB** IGT OE

In both the above cases, the frame of reading is shifted changing the meaning of the sentence. If similar insertions of one or two bases happen in the genetic code, the frame will shift and change the sequence of amino acids added. This is called a frameshift insertion.

- Now, delete the letter B from the above sentence – TOM HAS IGT OE
- Now, delete I from the above sentence – TOM HAS GTO E

Again, the frame has shifted changing the meaning of the sentence.

But, this time letters were deleted, so this is a frameshift deletion.

Inserting or deleting three letters (or multiples of three) adds or deletes

one codon and therefore, one amino acid. This does not affect the reading frame. Let's use the above sentence as an example.

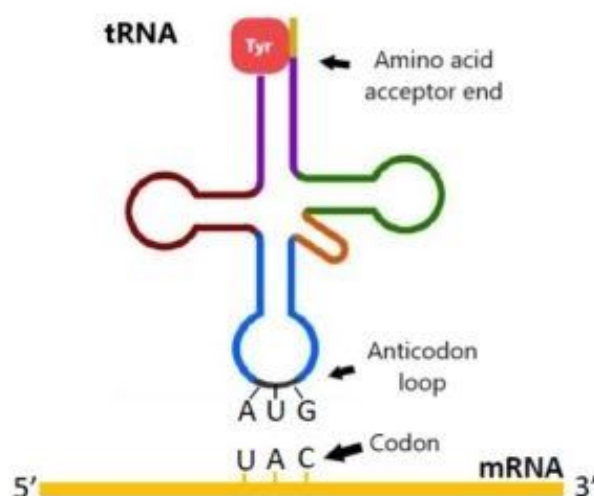
- Inserting three letters – TOM HAS **ONE** BIG TOE
- Deleting three letters (B, I, G) – TOM HAS TOE

The Adaptor Molecule – tRNA

Amino acids have no special ability to read the codons. Then how are the codons read? Francis Crick proposed that there is an 'adaptor molecule' that reads the codons and also binds to specific amino acids. This adaptor molecule was found to be tRNA or transfer RNA. The tRNA has the following parts:

- Anticodon loop – that has bases that are complementary to the codon on the mRNA.
- Amino acid acceptor end – It uses this end to bind to amino acids.

There are specific tRNAs for each amino acid. A special tRNA called initiator tRNA initiates the process of translation. There are no tRNAs for stop codons.



Cloverleaf structure of tRNA [Source: Wikimedia Commons]

Translation

It is the process of polymerization of amino acids to give rise to a polypeptide. The amino acids in the polypeptide are joined by peptide bonds. The sequence of amino acids is governed by the sequence of bases on the mRNA. In mRNA, a translational unit is a region that begins with the start codon (AUG), ends with the stop codon and codes for a polypeptide.

A mRNA also has some additional, untranslated regions or UTRs. They exist before the start codon at the 5' end and after the stop codon at the 3' end and are essential for efficient translation. Like transcription, translation has the following three steps.

Initiation

Translation occurs in the ribosome. In the inactive state, the ribosome consists of a large and small subunit. Translation begins when the small subunit encounters a mRNA. The ribosome binds to the start codon (AUG) on the mRNA to initiate translation. The initiator tRNA brings the first amino acid encoded by AUG i.e. methionine.

This phase also involves activation of amino acids in the presence of ATP and linking to their specific tRNA. This is called charging of tRNA or aminoacylation of tRNA. When two charged tRNAs are close enough, it favours formation of a peptide bond between them. The large subunit of the ribosome has two sites for subsequent amino acids to bind and to be close enough to form peptide bonds. The ribosome also acts as a catalyst for peptide bond formation.

Elongation

Elongation involves the stepwise addition of amino acids to the growing polypeptide chain. In this phase, the new amino acid-tRNA complex binds to the complementary codon on the mRNA via the anticodon on the tRNA. Next, the bond between the tRNA and amino acid breaks and a new peptide bond forms between the new and previous amino acids on the growing chain.

After adding each new amino acid to the polypeptide, the ribosome moves down to a new codon on the mRNA. This releases the previous tRNA, which is now free to bring another amino acid. These steps repeat till the ribosome reaches the stop codon on the mRNA.

Termination

Elongation continues till the ribosome reaches the stop codons (UAA or UAG or UGA). At this point, a release factor binds to the stop codon and terminates translation by releasing the polypeptide and mRNA from the ribosome.

Translation [Source: Wikimedia Commons]

Solved Example For You

Q1: Use the genetic code checkerboard to list the amino acid sequence resulting from the following mRNA sequence: 5'-A U A G C A G G A C U U-3'.

Sol: The answer is 5'- isoleucine-alanine-glycine-leucine -3'. If you use the checkerboard, you can see that AUA codes for isoleucine, GCA codes for alanine, GGA codes for glycine and CUU codes for leucine.