

Genetic Basis Of Life Cheat Sheet

Hankertrix

December 4, 2023

Contents

1	Definitions	4
1.1	Genome	4
1.2	The central dogma	5
1.3	Transcription	6
1.4	Translation	7
1.5	The genetic code	7
1.6	Codon	8
1.7	Start codon	8
1.8	Stop codon	8
1.9	Anti-codon	8
1.10	Gene expression	8
1.11	Coding region	9
1.12	Non-coding region	9
1.13	Gene	9
1.14	Template strand	9
1.15	Coding strand	9
1.16	Promoter	9
1.17	Transcription-translation coupling	10
1.18	Post-transcriptional processing	10
1.19	Pre-mRNA	11
1.20	Introns	11
1.21	Exons	11
1.22	Post-translation modification	11
1.23	Reverse transcription	12
1.24	Regulation of gene expression	13
1.25	Control of gene expression	13
1.26	Control of protein activity	14

1.27	Default (basal) state	14
1.28	Response state	14
1.29	LacZ gene	14
1.30	Constitutive expression	14
1.31	Activated response (expression)	14
1.32	Inhibited response (expression)	15
1.33	Regulatory proteins	15
1.34	Regulatory DNA sequences (regions)	15
1.35	Operon structure (poly-cistronic operon structure)	15
1.36	Heterochromatin	15
1.37	Euchromatin	15
1.38	Mutation	16
1.39	Wild type gene or wild type sequence	16
1.40	Parental sequence or parental gene	16
1.41	Gene variant	16
1.42	Mutant	16
1.43	Mutant protein	16
1.44	Single-base mutation or point mutation	16
1.45	Mutagenesis	16
1.46	Mutagen	17
1.47	DNA replication	17
1.48	Mutagenic activity	17
1.49	Transposons	17
1.50	Carcinogens	17
2	Features of the genetic code	18
2.1	Unambiguous	18
2.2	Redundant	18
2.3	Start/stop enabled	18
3	Transcription process	19
4	Translation process	20
4.1	Post translation	21
5	Site of transcription and translation	21
5.1	Prokaryotic cells	21
5.2	Eukaryotic cells	21

6	Exceptions to the central dogma	22
6.1	RNA Virus	22
6.2	Uncommon amino acids	22
7	Control of chromatin dynamics	23
8	Types of mutations	24
8.1	Classification according to the type of base changes	24
8.2	Classification based on the effect on the coding region	25
8.3	Classification based on phenotypic outcome	26
8.4	Classification based on the cause of the mutation	27
9	Mutations are not always all bad	29

1 Definitions

1.1 Genome

The genome refers to a collection of DNA that holds complete information about an organism, such as:

- What it will be like
- How it will grow
- Where it should grow
- What will its functions be
- How these parts function together

To make a genome functional, it must be **read and understood** by specific molecules in a specific environment. These molecules then **carry out** the series of actions that produces proteins that can bring about the results that are intended for the organism, such as:

- Running a biochemical reaction
- Procuring nutrients
- Coordinating metabolic activities
- Cell repair

Genetic information must also be **read with precision** by the cell machinery to carry out the tasks correctly and in a highly controlled and regulated manner, to give a distinct identity to the activities.

Two genomes from different individuals can result in **conflicting instructions**, which means that one genome has a set of instructions that specify a **particular outcome in a certain way**, while the other genome has the set of instructions that specify the **same outcome but in an alternative way**.

In other words, it is possible to have genes defining the **same type of tasks, but varying in terms of the details**. For example, a gene with the same function of defining the shape of seeds may have 2 variations, giving either smooth or rough surface seeds.

In short, the **genome** is the blueprint with information about the **collection of tasks** that can bring about the structural features, cellular activities, functions, and behaviour of an organism.

1.2 The central dogma

The central dogma describes the flow of information and states that the flow of information starts from genes, in the form of DNA, to a transient copy of the gene, in the form of messenger RNA, which then directs the sequence of amino acids in a polypeptide or protein.

This flow of information is absolute and unchanging, and cannot be reversed.

The 2 stages in this flow of information are referred to as **transcription** and **translation**.

1.3 Transcription

Transcription refers to the process whereby a region of DNA is copied into the form of an RNA. The product of the transcription, which is an RNA, is called a **transcript**.

In eukaryotes, there is an extra step of processing the RNA transcript to mature mRNA, and hence, the RNA transcript in this case is referred to as **pre-mRNA**.

You can think of transcription as simply the copying of a part of the larger nucleic acid document, originally set in the DNA dialect, into a transient mobile form, using an RNA dialect as DNA and RNA possess similar complementary pairing relationship of $A : T/U$ and $C : G$.

1.3.1 Key starting materials

1. Deoxyribonucleic acid (DNA)
2. RNA polymerase, an enzyme
3. Ribonucleotides (also referred to as NTPs)

1.3.2 Resulting product

RNA transcript.

1.3.3 Methods of control

1. Altering the binding efficiency of the RNA polymerase to the promoter
2. Altering the efficiency of RNA polymerase to unwind the DNA duplex
3. Affecting the movement of the RNA polymerase

1.4 Translation

Translation refers to the process whereby the information on an mRNA is used to assemble a sequence of amino acids into a chain. The product of translation is a **polypeptide**, which may be a complete unit of a protein or part of a protein.

You can think of translation in the language sense, as the 4-base language of DNA/RNA needs to be used to "describe" proteins, which are made up of 20 types of amino acids. As such, proteins can be said to use a "different language" from DNA and RNA, so there must be a system to "translate" the information encoded in the 4-base language to the 20-amino acid language.

The "dictionary" of this translation system is known as **the genetic code**.

1.4.1 Key starting materials

1. Messenger ribonucleic acid (mRNA)
2. Ribosome
3. Transfer ribonucleic acids (tRNAs)
4. Amino acids

1.4.2 Resulting product

Polypeptide.

1.4.3 Methods of control

- Targetting the ribosomal binding site

For example, having a specific mRNA sequence that stabilises, bends, and folds mRNA to block the ribosomal binding site, or using regulatory proteins that interfere or assist with ribosome binding.

1.5 The genetic code

The genetic code translates the information of RNA presented as **a set of 3 bases** (referred to as a codon) into the information of **a specific amino acid**.

1.6 Codon

Codon is a set of 3 bases. Codons run in a **continuous, unbroken sequence**. There are $4^3 = 64$ codons in total, with **3 stop codons** and **61 codons** that specify the 20 different **amino acids**.

1.7 Start codon

The start codon is the first codon in the transcribed mRNA that undergoes translation. It is AUG, and it codes for the amino acid methionine (Met) in eukaryotes and formyl methionine (fMet) in prokaryotes.

Mnemonic for the start codon:

- Are u gay? (AUG)

1.8 Stop codon

The stop codon is a codon that signals the termination of the translation process of the current protein. There are 3 stop codons, UAG, UAA and UGA.

Mnemonic for the stop codons:

- U are gay. (UAG)
- U are annoying. (UAA)
- U go away. (UGA)

1.9 Anti-codon

An anti-codon is a tRNA sequence of 3 nucleotides that is complimentary to a corresponding codon in mRNA. The anti-codon binds to its complimentary codon pair in protein synthesis.

1.10 Gene expression

Gene expression refer to the outcome when a gene is transcribed, translated and the polypeptides are finally compiled into a protein product and made biologically active.

1.11 Coding region

The coding region is the part of a gene that defines the amino acid sequence for the protein product.

The complete coding region begins with a **start codon** and ends with **any one of the stop codons**. There are various other ways to refer to the coding region, such as Open reading frame (ORF). In eukaryotes, the term **exons** are used to refer to coding regions when they are associated with the non-coding regions (**introns**).

1.12 Non-coding region

The non-coding region is the part of the DNA that are important for the control of gene expression. **DNA consists of both coding and non-coding regions.**

1.13 Gene

The region of the DNA that defines a polypeptide or protein is called a **gene**. This region includes the complete coding region for that polypeptide or protein as well as some associated non-coding regions.

1.14 Template strand

The template strand is the strand of the DNA duplex that the RNA polymerase **interacts** with during transcription.

1.15 Coding strand

The coding strand is the strand of the DNA duplex that the RNA polymerase **does not** interact with during transcription.

1.16 Promoter

The promoter refers to a region on the template strand of the DNA where the RNA polymerase binds during transcription.

1.17 Transcription-translation coupling

Transcription-translation coupling refers to ribosomes binding to an mRNA transcript and starting the process of translation before the transcription is even completed. This can only happen in prokaryotic cells as they have no nucleus and hence the mRNA transcript is always exposed to ribosomes present in the cytoplasm. This allows the ribosome to bind to the mRNA polymerase as long as the ribosome binding site on the mRNA is exposed.

1.18 Post-transcriptional processing

Post-transcriptional processing refers to the process mRNA transcripts in eukaryotic cells undergo to modify them prior to translation.

Within the nucleus, the pre-mRNA is processed in the following ways to form a mature mRNA transcript:

1. Modification of the 5' end by adding a 5' cap of a modified G
2. Modification of the 3' end by adding a poly-*A* tail
3. Removal of the introns to join the exons together, which is also known as **splicing**

After this process, the mature mRNA transcript then migrates out to the cytoplasm where it gets bound by a ribosome to start the process of translation.

This process doesn't occur in prokaryotes as the genes are not organised in an exon-intron manner.

1.18.1 Control of post-transcriptional processing

The splicing process makes use of DNA sequences at the intron and exon junctions and several proteins. Hence, the control of gene expression can be achieved by targetting any of these components.

For example, a particular junction sequence may be such that it can delay splicing, which means that the completion of the mature mRNA is slowed down, and the gene expression level decreases. To ramp up, a regulatory protein can be used to speed up the delayed splicing process. This then increases the level of gene expression.

1.19 Pre-mRNA

Pre-mRNA refers to the RNA transcript first produced within the nucleus of a eukaryotic cell. It is called this because it carries regions which are not meant to be translated.

1.20 Introns

Introns are regions on pre-mRNA that are **not** meant to be translated.

1.21 Exons

Exons are regions on pre-mRNA that are **meant** to be translated.

1.22 Post-translation modification

Post-translation modification refers to the further modifications to polypeptides created from the translation process to become complete and functional proteins.

Some of the post-translational modifications include:

- Cleavage of the polypeptide chain
- Formation of disulphide bonds
- Binding of more than one polypeptide chain together
- Addition of chemical groups like sugars, methyl groups and acetyl groups to the amino acid residues of the polypeptide

1.22.1 Post-translation modification of insulin

- Insulin is first formed as a single polypeptide called preproinsulin.
- Disulphide bonds are formed to secure the *A* and *B* segments, and the signal peptide is cleaved off, giving rise to proinsulin.
- Cleavage of the *C* segment follows, giving rise to a fully functional insulin molecule.

1.22.2 Control of post-translational modification

Post-translational modification often requires the activity of other enzymes to form the covalent bonds necessary in the modification. So, mechanisms of regulation may target these enzymes, such as:

1. Keeping the enzyme from accessing the pro-protein, such as confining the pro-protein within membrane vesicles until the time when it needs to be activated.
2. Controlling the availability of substrates needed for the modification, such as specific sugars for glycosylation.
3. Controlling the conditions necessary for enzyme catalysis.

1.23 Reverse transcription

Reverse transcription is a process in RNA viruses where the viral RNA creates the DNA transcript for further replication of the virus.

1.24 Regulation of gene expression

The regulation of gene expression refers to the control of when and how much of a particular protein to produce, which controls how much the gene is expressed.

1.24.1 Why is regulation needed?

- Expressing a gene requires cellular energy, so expressing it all the time will lead to a **waste of energy and resources**, such as the transcription and translation machinery.
- This waste of cellular energy will lead to a **shortage of energy or cellular machinery** for the expression of other genes.
- Expressing a gene continuously **may conflict with other gene activities**.

1.24.2 Levels of control

1. Transcription
2. Post-transcriptional processing (eukaryotes only)
3. Translation
4. Post-translational modification
5. Chromatin dynamics (eukaryotes only)

1.25 Control of gene expression

Gene expression being controlled refers to the gene being transcribed and translated to produce the protein at the right time, place, and at the right amount.

1.26 Control of protein activity

A protein can be present constantly but placed in a stand-by mode, which means it is not functionally normally but only becomes active under certain circumstances.

1.26.1 Example

A low concentration of antibodies always exists in our blood. This means that a low level of antibody gene expression has been occurring to have antibody molecules on stand-by. When an antibody binds to a foreign particle, such as bacterium, it is triggered to start a series of defence mechanisms to protect us from the foreign particle. This situation describes a **control of the activity** of the antibodies.

In the process, signals are released which result in more expression of genes encoding the antibody, so that more antibodies are produced. This situation describes a **control of the gene expression** of antibodies.

1.27 Default (basal) state

The state a gene is in **before** any kind of stimulus.

1.28 Response state

The state the gene is in **after** a stimulus.

1.29 LacZ gene

The lacZ gene encodes β -galactosidase, which is an enzyme that breaks down lactose into glucose and galactose. The lacZ gene is only expressed when **glucose is not present**, but **lactose is present**.

1.30 Constitutive expression

A constitutive expression just means that a gene is being expressed at a **moderate to high** level even in the **absence** of any kind of stimulus.

1.31 Activated response (expression)

An activated response just means that gene expression after stimulation is **higher** compared to the default state.

1.32 Inhibited response (expression)

An inhibited response just means that gene expression after stimulation is **lower** compared to the default state.

1.33 Regulatory proteins

Regulatory proteins are proteins that **regulate the transcription process**. They are also referred to as regulators, and more specifically, activators, repressors, or transcriptional factors.

Regulatory proteins often need to bind to specific sequences of DNA to carry out its function. However, this is not always the case and some may act by binding to RNA polymerase or other regulatory proteins.

1.34 Regulatory DNA sequences (regions)

Regulatory DNA sequences are DNA sequences that **regulate the transcription process**. These sequences are known as **operators** in **prokaryotes** and **enhancers** in **eukaryotes**. Regulatory DNA sequences are often for regulatory proteins to bind to, and some may serve their function in other ways, such as by affecting the way DNA bends.

1.35 Operon structure (poly-cistronic operon structure)

The operon structure refers to a structure that has genes of related functions lined up sequentially and are then transcribed from one promoter. This is usually the case in **prokaryotic chromosomes**. This structure allows all genes in the same operon to be regulated together at the transcriptional level. Furthermore, this structure means that only a single mRNA is transcribed which is then translated into the respective proteins.

1.36 Heterochromatin

Heterochromatin is a form of chromatin that is **more densely compacted**.

1.37 Euchromatin

Euchromatin is a form of chromatin that is **less densely compacted**.

1.38 Mutation

A mutation is a change in the base DNA sequence. A mutation is only meaningful when comparing to an "original" version. A mutation can occur in both the coding and non-coding regions of a gene. **Mutations** in the **coding region** may lead to a change in the amino acid sequence, resulting in a **mutant protein**, while **mutations** in the **non-coding region** may lead to a change in **gene expression**, which would cause a mutant cell to exhibit a change in its gene expression pattern.

1.39 Wild type gene or wild type sequence

A wild type gene is the reference gene that most other genes are compared to.

1.40 Parental sequence or parental gene

A parental sequence or parental gene is a reference gene that has gone through several rounds of mutations, and is no longer the same as the wild type gene.

1.41 Gene variant

A gene variant is a gene that contains mutations.

1.42 Mutant

A mutant is a cell or organism carrying the mutated version of a gene.

1.43 Mutant protein

A mutant protein is a protein encoded from the mutated version of a gene.

1.44 Single-base mutation or point mutation

A single-base mutation is a change of one base in the DNA sequence.

1.45 Mutagenesis

Mutagenesis refers to the process of generating mutations.

1.46 Mutagen

Mutagen is the agent that causes mutagenesis, or causes mutations to happen.

1.47 DNA replication

DNA replication is the process by which the parental chromosomal DNA is replicated to be distributed to the daughter cells.

1.48 Mutagenic activity

This refers to the property of certain chemical compounds to alter the structure of DNA.

1.49 Transposons

Transposons are DNA sequences that can jump onto other regions in the DNA. They're usually called "jumping genes" due to this characteristic. They can range from a few hundred bases long to several thousand bases long, and are characterised by having inverted repeats at both ends.

1.50 Carcinogens

Carcinogens are cancer-causing substances. Cancer is triggered through the accumulation of certain types of mutations in our cells, thus chemicals that can alter DNA structures and cause mutations are carcinogens as well.

2 Features of the genetic code

2.1 Unambiguous

Unambiguous means that each codon only has one meaning. For example, the codon *UCU* is meant to represent the amino acid Serine and no other amino acids.

2.2 Redundant

The redundancy of the genetic code (also referred to as the "degeneracy" of codons), refers to the fact that more than one codon may specify a particular amino acid. For example, Leucine is represented by 6 codons and Glycine by 4 codons.

2.3 Start/stop enabled

This means that there are specific codons that signal the start and the stop of a polypeptide formation.

3 Transcription process

Transcription involves DNA and requires the action of RNA polymerases. During transcription, the RNA polymerase interacts with one stretch of sequence on one strand of the DNA duplex.

The region where the RNA polymerase binds is called the promoter and transcription starts some distance downstream from it. The RNA polymerase will unwind the DNA duplex so that the DNA will stay single-stranded for that short region. It then synthesises RNA by adding on sequential ribonucleotides which are pairing with the bases on the DNA. As it does that, it moves along the template strand. At the end of the process, an mRNA chain is synthesised.

The mRNA chain is synthesised in the 5' to 3' direction. The sequence of the RNA transcript produced will be **identical** to that of the coding strand of the DNA.

As the RNA polymerase moves along the template DNA strand and the RNA transcript grows longer, the part of DNA behind the RNA polymerase will rewind (the DNA becomes double-stranded again) while the part in front of it will unwind.

4 Translation process

The mRNA transcript produced through transcription will need to be bound by a ribosome to start the translation process. The ribosome is the tool that has the ability to translate the RNA transcript. A ribosome consists of a large subunit and a small subunit.

The two subunits come together as an mRNA is bound to the small subunit through complementary binding with a sequence of rRNA on the small subunit. Within the ribosome, there are parts which we refer to as *A* site, *P* site and *E* site. The rRNAs that make up the sites are responsible for catalysing the formation of the polypeptide.

Once the ribosome are bound with the mRNA transcript, translation begins.

When the small subunit of ribosome binds to the mRNA, translation is initiated. The translation starts with bringing the start codon (*AUG*) amino acid, Methionine, to the *P* site by tRNA. Thereafter, the amino acids carried on by their respective tRNA (amino-acyl tRNA, also referred to as "charged tRNA") are brought into the *A* site, through the complementary binding of the tRNA's anti-codon to the codon on the mRNA transcript.

The second amino acid, defined by the codon next to the start codon, brought into the *A* site, allowing a peptide bond to be formed between the two amino acids through the enzymatic action of the rRNAs. Concurrently, the first amino acid residue's bond to its tRNA is broken.

The ribosome then moves 3 bases to the right (i.e. downstream) on the mRNA, bringing with it the tRNA-peptide bound to the codon. This effectively moves the first tRNA to the *E* site and the second tRNA to the *P* site, along with its attached peptide.

The first tRNA is then released from the *E* site, and the next tRNA enters the *A* site. This process continues until the ribosome moves into the position of a stop codon.

4.1 Post translation

After translation, all polypeptides undergo folding. Some polypeptides become "complete" and functional proteins immediately after folding, while some need to undergo post-translation modification to become complete and functional proteins.

5 Site of transcription and translation

5.1 Prokaryotic cells

Since the chromosome in prokaryotes are not separated from the cytoplasm by a nuclear envelope, transcription and translation takes place in one common cellular space.

5.2 Eukaryotic cells

Since the RNA transcripts are produced within the nucleus of a eukaryotic cell, which is separated by a nuclear envelope, the RNA transcript needs to migrate out from the nucleus to the cytoplasm to bind to ribosomes to start the translation process.

6 Exceptions to the central dogma

The central dogma should universally apply to all biological beings on Earth. However, this is not absolute. It is important to know where these exceptional cases exist, and to keep in mind that there may be more exceptions which our current advances in science has not yet surfaced.

6.1 RNA Virus

A class of virus carries RNA as its genome, which means that the full information necessary to define the biological activities of this virus is passed from one generation to the next in the form of RNA, instead of DNA.

The virus doesn't start from the halfway point in the central dogma, instead, the virus relies on its host cell to replicate and express its genes, so it first converts its RNA genetic materials into DNA, and the host cell then follows the central dogma sequence of DNA being transcribed into RNA, which is then translated into proteins.

Since viral RNA creates the DNA, the "transcription" process is **reversed** compared to the usual transcription process in the central dogma, we refer to this process as **reverse transcription**. The DNA transcript is referred to as complementary DNA (cDNA) to distinguish it from DNA that originated from other sources and processes.

6.2 Uncommon amino acids

The central dogma is universal across all living things. Since the genetic code itself is exactly the same among all organisms, if someone obtains a DNA sequence from new or unknown sources, they can deduce the possible sequence of the protein encoded. However, the uncommon amino acids **Selenocysteine** and **Pyrrolysine** are not simply incorporated through the universal codon-anti-codon matching of amino-acyl-tRNA based on the genetic code. Both of them are incorporated in place of certain **stop codons**, but require a specific set of conditions and biochemical reactions to complete the incorporation.

7 Control of chromatin dynamics

This only occurs in **eukaryotes**. This level of control can be considered to be part of "control at the transcriptional level" because what chromatin structure ultimately affects is the accessibility of the RNA polymerase to the promoter site. However, the complex eukaryotic chromosome structure referred to as chromatin, involves more than simply winding or unwinding DNA supercoils. Thus, it makes sense to consider this as a completely different level of control for gene expression.

Eukaryotic chromosome is organised into nucleosomes which consists of a portion of DNA wrapped around histone molecules. Whether the tails of histones are methylated ($-CH_3$) or acetylated ($-COCH_3$) determines whether the chromatin exists as heterochromatin (more densely compacted) or euchromatin (less densely compacted) respectively.

Genes in euchromatin are more accessible for transcription but the nucleosomes act as a neutral barrier to RNA polymerase access. This barrier can be removed through the action of a chromatin-remodelling complex which pushes the histones apart. Hence, control of gene expression can be achieved by:

- Methylation or acetylation of the histones
- Controlling the chromatin-remodelling complex.

8 Types of mutations

8.1 Classification according to the type of base changes

8.1.1 Insertion mutation type

The addition of an extra base, or a sequence of extra bases, is called insertion.

8.1.2 Deletion mutation type

The removal of an existing base, or a sequence of bases, is called deletion.

8.1.3 Substitution mutation type

The replacement of an existing base, or a sequence of bases, with another base or sequence of bases is called a substitution.

8.1.4 Inversion mutation type

If a stretch of bases is inverted, meaning that the upper strand of the DNA has been rotated to take the position of the lower strand or vice versa, it is called an inversion.

8.1.5 Reciprocal translocation mutation type

A reciprocal translocation mutation is a mutation where two segments carrying the same number of bases, but are located some distance apart, are exchanged.

8.2 Classification based on the effect on the coding region

This classification of mutations is based on the outcome generated by the mutation with respect to the amino acid sequence in the coding region.

8.2.1 Nonsense mutation

A nonsense mutation is generated when a stop codon takes the place of a codon for an amino acid residue. As a result, a truncated polypeptide is produced after translation.

8.2.2 Missense mutation

A missense mutation is generated when a particular codon is changed to become another codon of a different amino acid residue.

8.2.3 Silent mutation

Silent mutation refers to a type of mutation where there is a change in the base sequence, but the change does not result in any amino acid changes.

8.2.4 Frameshift mutation

Frameshift mutation refers to a type of mutation where a deletion or an insertion in a DNA sequence shifts the way the sequence is read (also known as the "frame of reading").

8.3 Classification based on phenotypic outcome

This classification of mutations describes the characteristics of the mutant protein compared to the wild type protein, or that of the mutant organism compared to the wild type organism. In this classification, how the mutation is generated is not relevant.

8.3.1 Loss of function mutation

A loss of function mutation is a mutation that causes a protein that has reduced or no function to be produced.

8.3.2 Gain of function mutation

A gain of function mutation is a mutation that causes a protein that has new or enhanced activity to be produced.

8.3.3 Null mutation

A null mutation is a mutation that stops the production of a protein.

8.3.4 Constitutive mutation

A constitutive mutation is a mutation that causes gene expression to become unregulated and constitutive.

8.3.5 Lethal mutation

A lethal mutation is a mutation that causes the death of the cell or organism.

8.4 Classification based on the cause of the mutation

This classification of mutations is based on the way that the mutation has been generated.

8.4.1 Spontaneous mutation

A spontaneous mutation comes about during the DNA replication process, specifically during the process where new bases are incorporated into the growing DNA strand. Sometimes, a mismatched base is added to DNA strand and the error is not corrected, resulting in a spontaneous mutation.

8.4.2 Chemical mutagenesis

Chemical mutagenesis refers to the exposure of DNA to compounds with mutagenic activity, which is the ability to alter the structure of DNA. Some examples of such compounds are ethidium bromide, which is commonly used to stain DNA in the laboratory, and nitrosamine, which is found in tobacco smoke.

Chemical compounds may modify the bases directly through:

1. Depurination, deamination or oxidation.
2. Binding of DNA to alter its backbone structure.

In both cases, the changes enhances error-prone incorporation of bases during DNA replication, leading to mutation in the next generation of cells.

8.4.3 Mutagenesis by radiation

High energy radiation is able to:

1. Cause damage to the DNA backbone. For example, X-rays can break the sugar-phosphate backbone
2. Alter the structure of bases. For example, UV light can cause thymine dimers to form.

A break in the DNA backbone usually results in the removal of short stretches of bases, leading to deletion, if the cell is able to repair the break. Sometimes the number of breaks within a genome can be so numerous that the cell is unable to repair them adequately. In this case, the cell will simply die instead of ending up with mutations.

Altering the structure of DNA through thymine dimer formation will have similar effects as chemical mutagenesis, resulting in error-prone DNA replication and incorporation of bases which are different from the original sequence.

8.4.4 Transposon mutagenesis (transposition)

In a biological system, a large change in sequence may occur due to the activities of transposons. These are DNA sequences which can range from a few hundred bases long to several thousand bases long, and are characterised by having inverted repeats at both ends.

It is able to copy or cut itself out from its original position to somewhere else in another region of DNA due to the activities of the transposase enzyme, which is encoded in the transposon's DNA sequence.

9 Mutations are not always all bad

A mutation is **not always bad** and is **not completely undesirable**. A mutation is simply a variation, and the variation may not always have negative consequences. An example of this would be a gene that carries a silent mutation, as there is essentially no difference in the resulting protein. Some mutations may simply make the gene different but not worse, like producing pigments that give rise to a red flower instead of an orange flower. Sometimes, the mutation may even make the gene perform better, usually in the case of gain of function mutations. We are also artificially introducing mutations to fulfil our needs. Thus, mutations are not always all bad.

Furthermore, mutations and mutants serve a very important role in biological research as the function of a particular gene cannot be studied by just observing what a "normal" organism does, as it only reflects the collective actions of many genes that make up the organism. A mutation in the gene of interest helps to isolate its function.

Essentially, a mutation is the only observable difference in an experiment, with the rest of the wild type genes being the control group. With this, it is possible to trace the function of a gene.