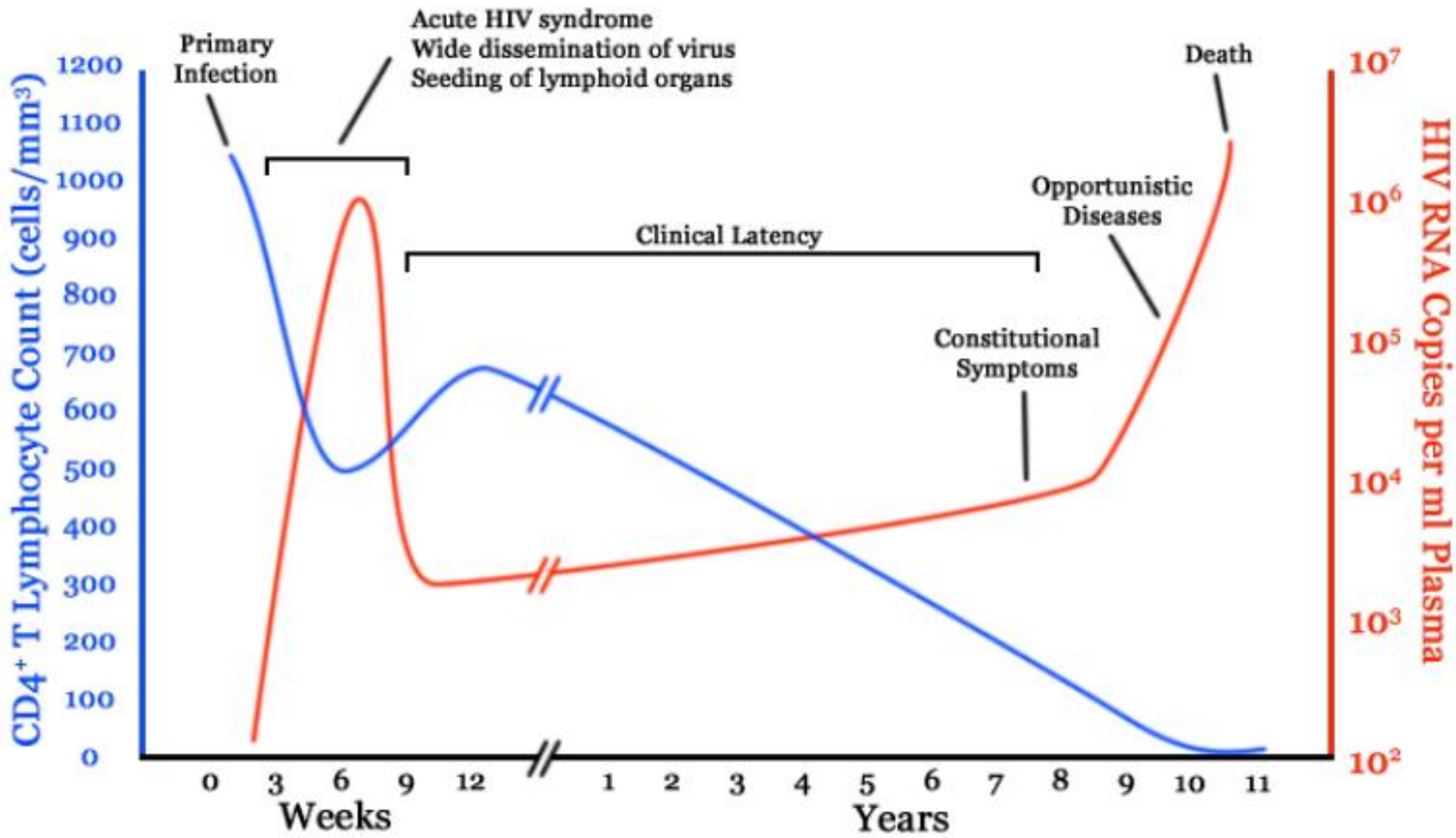
Gehad Elkady
Assistant Lecturer of Virology

There are 10^{16} HIV genomes on the planet today

With this number of genomes, it is highly probable that HIV genomes exist that are resistant to every one of the antiviral drugs that we have now, or EVER WILL HAVE!

Time course of the HIV infection



- When a gene has been identified, insight into its function in principle can be gained by generating a mutant organism that entirely lacks the gene.

Viral genome mutations fall into_
1- Evident…
2- Hidden…

1- cis-acting…
2- trans-acting….

1- null mutation: a mutation that completely eliminate gene function, usually because the gene has been deleted. So if a gene is essential, a null mutation is lethal.

2- leaky mutation: a mutation fails to affect the phenotype, i.e: enough active product is made to fulfill its function, even though the activity is quantitatively reduced or qualitatively different from the wild type.

- But if a null mutant fails to affect a phenotype, we may safely conclude that the gene function is not necessary.

3- Loss-of- function mutations: it is a recessive mutation
It is a null mutations, or other mutations that impede gene function (but do not necessarily abolish it entirely).

4- A gain- of- function mutation: it is a dominant mutation,
Sometimes a mutation has the opposite effect and causes a protein to acquire a new function.

5- Silent mutations: Mutations without apparent effect. Not all mutations in DNA lead to a detectable change in the phenotype due to (neutral substitutions)
- Neutral substitutions fall into two:

a- Some involve base changes in DNA that do not cause any change in the amino acid present in the corresponding protein.

B- Others change the amino acid, but the replacement in the protein does not affect its activity.

6- Forward mutations: mutations that inactivate a gene. Their effects are reversed by back mutations.

back mutations fall into two types:

a-True reversion: an exact reversal of the original mutation.
So if an A.T pair has been replaced by a G.C pair, another mutation to restore the A.T pair will exactly regenerate the wild –type sequence.

B- Second- site reversion: another mutation may occur elsewhere in the gene, and its effects compensate for the first mutation.
For example, one amino acid change in a protein may abolish gene function, but a second alteration may compensate for the first and restore protein activity.

NB:
A forward mutation results from any change that inactivates a gene, whereas a back mutation must restore function to a protein damaged by a particular forward mutation.

-So the demands for back mutation are much more specific than those for forward mutation.

- The rate of back mutation is correspondingly lower than that of forward mutation, typically of a factor ~ 10

- The genetic code is triplet

7- The point mutation: due to substitution of a single base pair will affect only one amino acid because only the second codon has been changed.

8- The frameshift mutation: due to incorporated or omitted single base pair, so this will change the reading frame for the entire subsequent sequence.

- Because the new sequence of triplets is completely different from the old one, the entire amino acid sequence of the protein is altered beyond the site of mutation.

- So the function of the protein is likely to be lost completely.

- However, frameshift mutations are induced by the acridines, compounds that bind to DNA and distort the structure of the double helix, causing additional bases to be incorporated or omitted during replication. Each mutagenic event sponsored by an acridine results in the addition or removal off a single base pair.

- If an acridine mutant is produced by, say, addition of a nucleotide, it should revert to wild type by deletion of the nucleotide. But reversion can also be caused by deletion of a different base, at a site close to the first. Combinations of such mutations provided revealing evidence about the nature of the genetic code.

- However the combination of an insertion and a deletion causes the code to be read in the incorrect frame only between the two sites of mutation; correct reading resumed after the second site.

- The original analysis was performed by genetic analysis of mutations in rII region of the phage T6.
-All acridine mutations could be classified into one of two sets, described as (+) and (-). Either type of mutation by itself causes a frameshift.
- The (+) type by virtue of a base addition.
- The (-) type by virtue of a base deletion.

- Double mutant combinations of the types:
(++) and (--) = continue to show mutant behavior.

- But combinations of the types:
(+-) or (-+)= suppress one another, called suppressor.

- When triple mutants are constructed, only (+++) and (---) combinations show the wild phenotype, while other combinations remain mutant.
If we take three additions or three deletions to correspond respectively to the addition or omission overall of a single amino acid, this implies that the code is read in triplets. An incorrect amino acid sequence is found between the two outside sites of mutation, and the sequence on either side remains wild type.

1- cis-acting sites (control sites): DNA close to the coding region which bind to it a controlled protein.

(Control sites in DNA provide binding sites for proteins; coding regions are expressed via the synthesis of RNA).

-So figure 1.31: mRNA can be synthesized only when the protein is bound to the DNA. Now suppose that a mutation occur in the DNA sequence to which this protein binds (that is called control sites), so that the protein can no longer recognize the DNA. As a result, the DNA can no longer be expressed.

-So a gene can be inactivated either by a mutation in a control site or by a mutation in a coding region.
-The mutations can not be distinguished genetically, because both have the property of acting only on the DNA sequence of the single allele in which they occur.

-They will therefore have identical properties in the cis/trans test, and a mutation in a control region is therefore defined as comprising part of the gene in the same way as a mutation in the coding region.

- figure 1.32 shows that a deficiency in the control site affects only the coding region to which it is connected; it does not affect the ability of the other allele to be expressed.

-So cis acting mutation: a mutation that acts solely by affecting the properties of the contagious sequence of DNA.

or. Cis-acting mutation: it must function by affecting directly the properties of the contagious DNA, which means that it is not expressed in the form of RNA or pt.

Figure 1.33 shows trans-acting mutation: that the absence of regulator protein would prevent both alleles from being expressed.

-OR. Trans- acting mutation: its effects must be exerted through some diffusible products (typically pt) that acts on multiple targets within a cell.

RF fall into_

1- ORF: An open reading frame _A reading frame that consists exclusively of triplets that represent amino acid
- A sequence that is translated into protein has a reading frame that starts with a special initiation codon (AUG) and that extends through a series of triplets representing amino acids until it ends at one of three types of termination codon (UAA, UAG, UGA).

2- BRF: Blocked reading frame_ A reading frame that cannot be read into protein because termination codons occur frequently.
- If a sequence is blocked in all three reading frames, it cannot have the function of coding for protein.
- When the sequence of a DNA region of unknown function is obtained, each possible reading frame is analyzed to determine whether it is open or blocked.

3- URF: Unidentified reading frame_ An open reading frame for which no protein product has been identified.

Mutations are concentrated at hotspots

eg: lacI gene of E.coli

- The statistical probability that more than one mutation occurs at a particular site is given by random-hit kinetics.

So some sites will gain one, two, or three mutations, while others will not gain any.

Hotspots: some sites gain far more than the number of mutations expected from a random distribution; they may have 10X or even 100X more mutations than predicted by random hits.

The Repair system

Types_

1- Post-replication repair.
2- Excision repair.
3- Recombinant- like mechanism.

- The deamination of 5-methylcytosine leaves thymine. Because this base is a respectable constituent of DNA in its own right, the system does not recognize the change, and the mutation results.
The conversion creates a mispaired G.T partnership, whose separation at the subsequent replication produces one wild-type G.C pair and one mutant A.T pair.

The operation of this system casts a interesting light on the use of T in DNA compared with U in RNA.

1- Mutations change the sequence of DNA.
2- Mutations are concentrated at hotspots.
3- A cistron is a single stretch of DNA.
4- The nature of multiple alleles.
5- Recombination occurs by physical exchange of DNA.
6- The genetic code is triplet.

3- A cistron is a single stretch of DNA.

could also represent mutations in two *different* genes whose products are involved in the same function. The **complementation test** is used to determine whether two mutations lie in the same gene or in different genes. The test consists of making a heterozygote for the two mutations (by mating parents homozygous for each mutation).

If the mutations lie in the same gene, the parental genotypes can be represented as:

$$\frac{m_1}{m_1} \quad \text{and} \quad \frac{m_2}{m_2}$$

The first parent provides an m_1 mutant allele and the second parent provides an m_2 allele, so that the heterozygote has the constitution:

$$\frac{m_1}{m_2}$$

No wild-type gene is present, so the heterozygote has mutant phenotype.

If the mutations lie in different genes, the parental genotypes can be represented as:

$$\frac{m_1 \; +}{m_1 \; +} \quad \text{and} \quad \frac{+ \; m_2}{+ \; m_2}$$

Each chromosome has a wild-type copy of one gene (represented by the plus sign) and a mutant copy of the other. Then the heterozygote has the constitution:

$$\frac{m_1 \; +}{+ \; m_2}$$

in which the two parents between them have provided a wild-type copy of each gene. The heterozygote has wild phenotype; the two genes are said to **complement**.

The complementation test is shown in more detail in **Figure 1.20.** The basic test consists of the comparison shown in the top part of the figure. If two mutations lie in the same gene, we will see a difference in the phenotypes of the *trans* configuration and the *cis* configuration. The *trans* configuration is mutant, because each

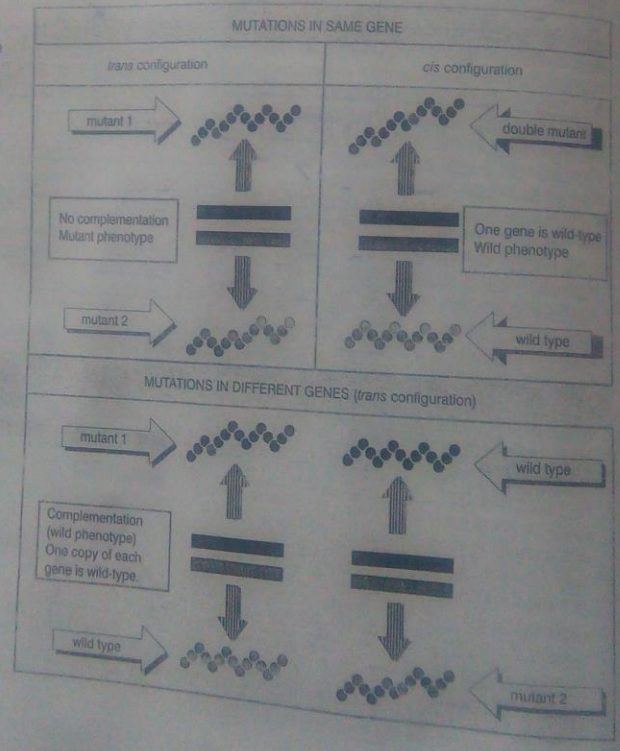
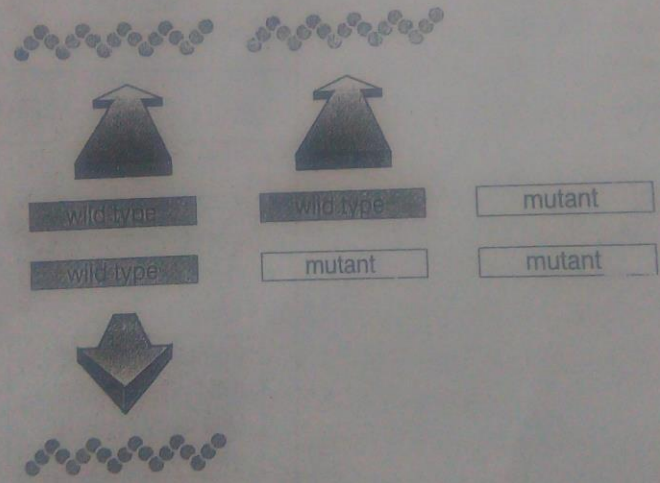


Figure 1.20 The cistron is defined by the complementation test. Genes are represented by bars; red stars identify sites of mutation.

Figure 1.19 Genes code for proteins; dominance is explained by the properties of mutant proteins. A recessive allele does not contribute to the phenotype because it produces no protein (or protein that is nonfunctional).

Wild-type homozygote	Wild-type/mutant heterozygote	Mutant homozygote
Both alleles produce protein	Only dominant allele produces protein	Neither allele produces protein



Wild phenotype	Wild phenotype	Mutant phenotype

4- The nature of multiple alleles.

5- Recombination occurs by physical exchange of DNA.

Figure 1.24 Recombination involves pairing between complementary strands of the two parental duplex DNAs.

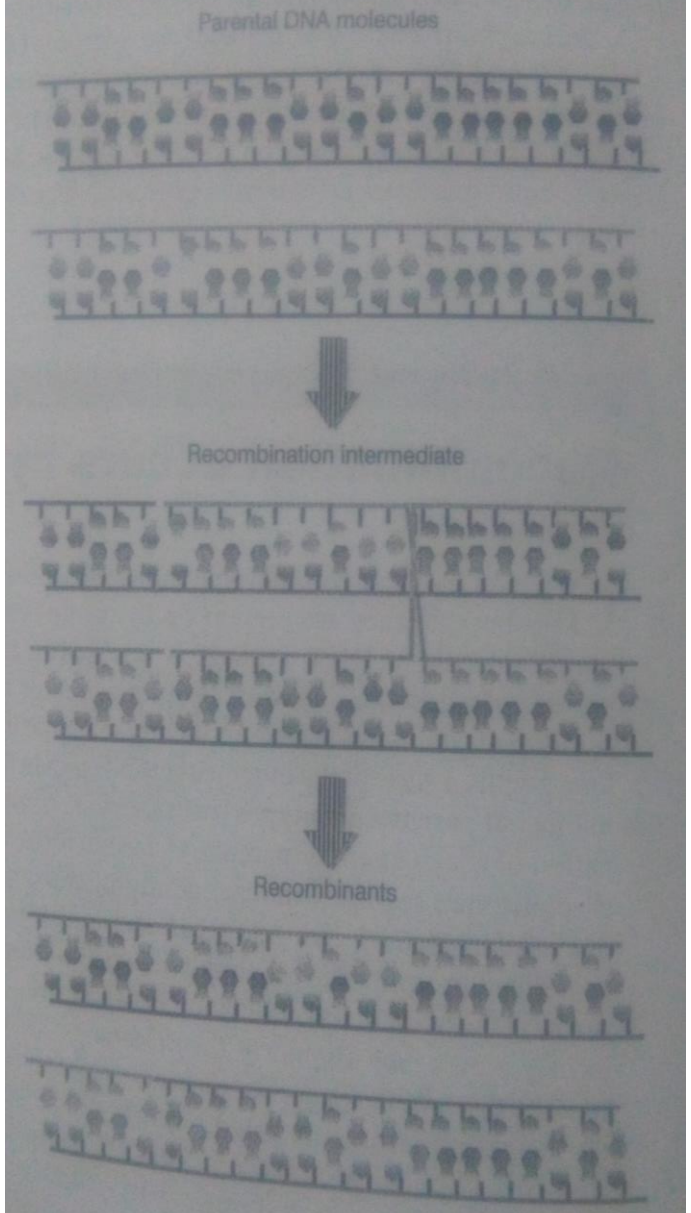
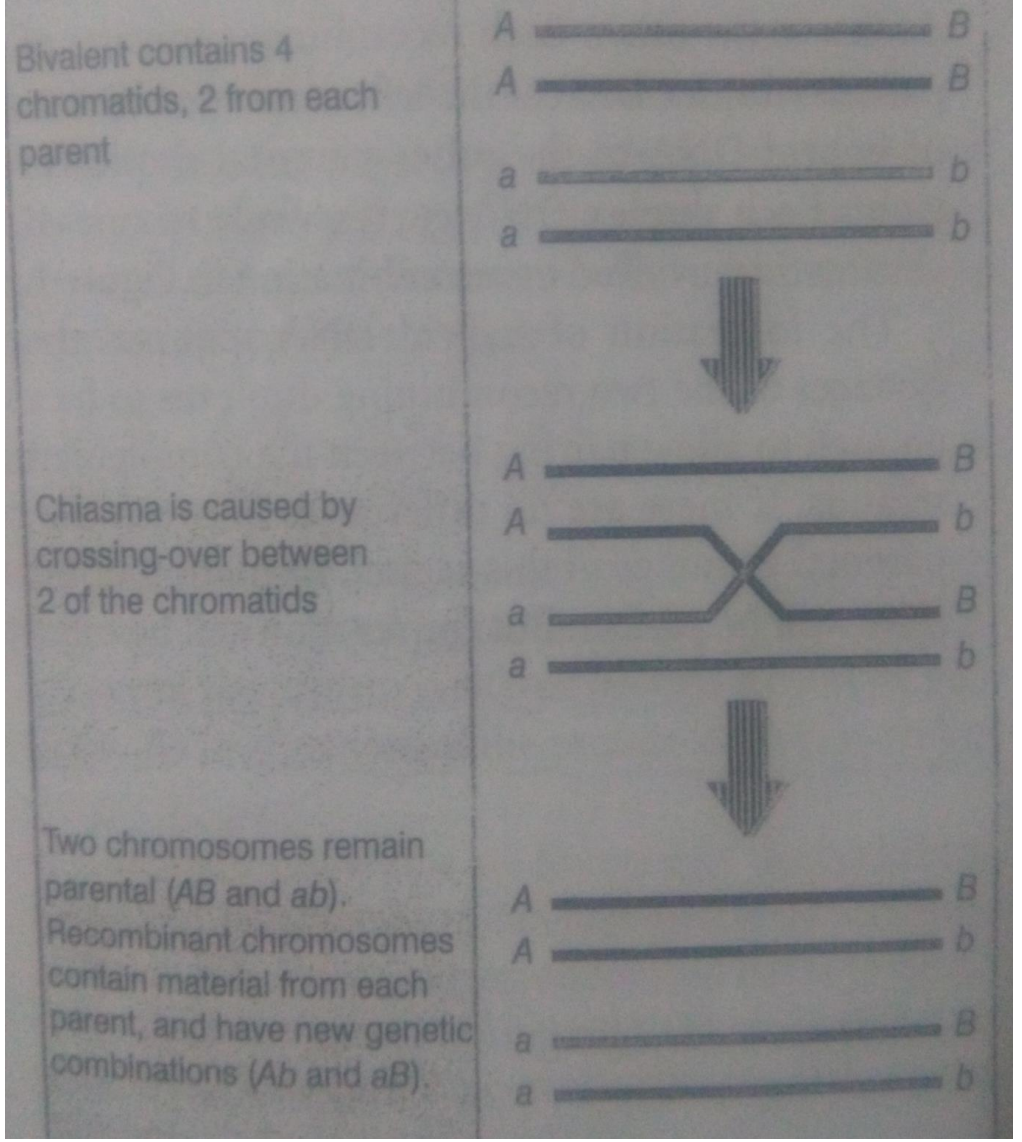


Figure 1.23 Chiasma formation is responsible for generating recombinants.



7- Cis-acting sites and trans-acting molecules

Figure 1.31 Control sites in DNA provide binding sites for proteins; coding regions are expressed via the synthesis of RNA.

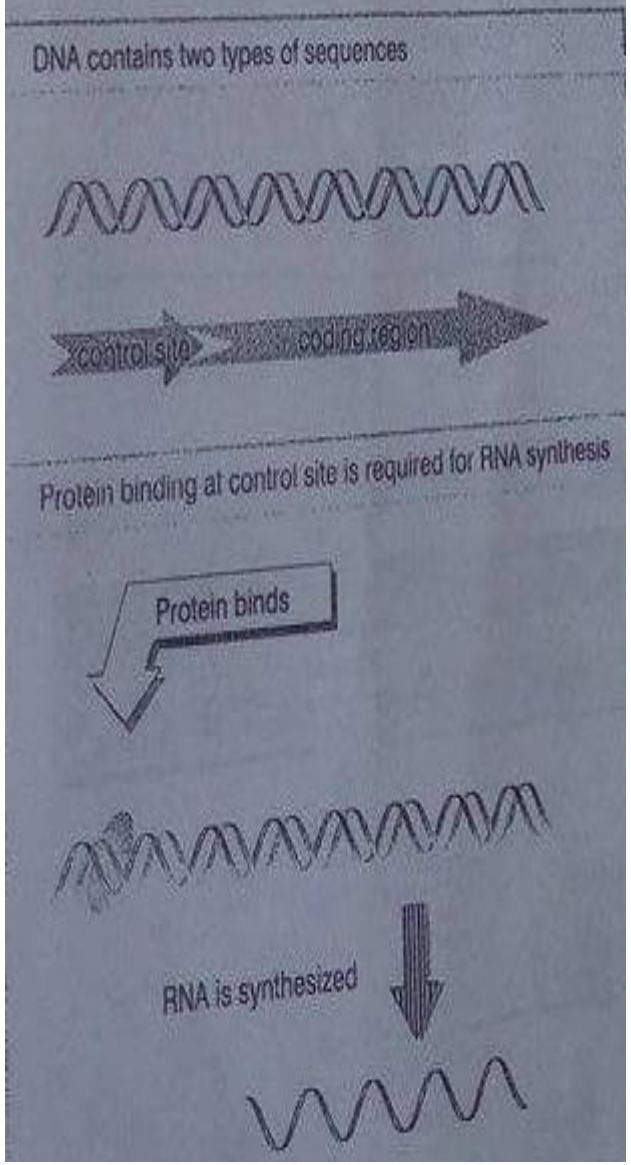


Figure 1.32 A *cis*-acting site controls the adjacent DNA but does not influence the other allele.

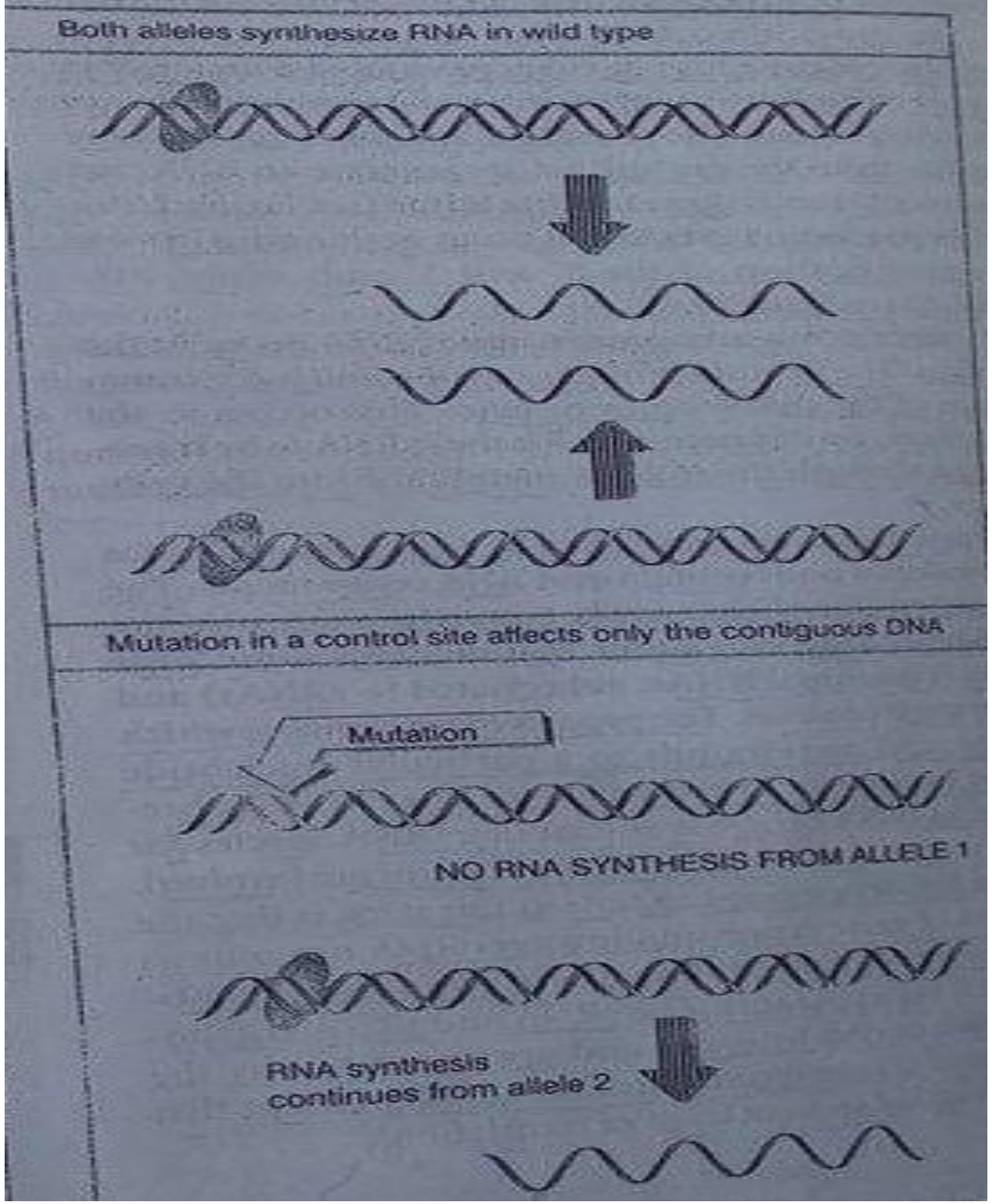


Figure 1.33 A *trans*-acting mutation in a protein affects both alleles of a gene that it controls.