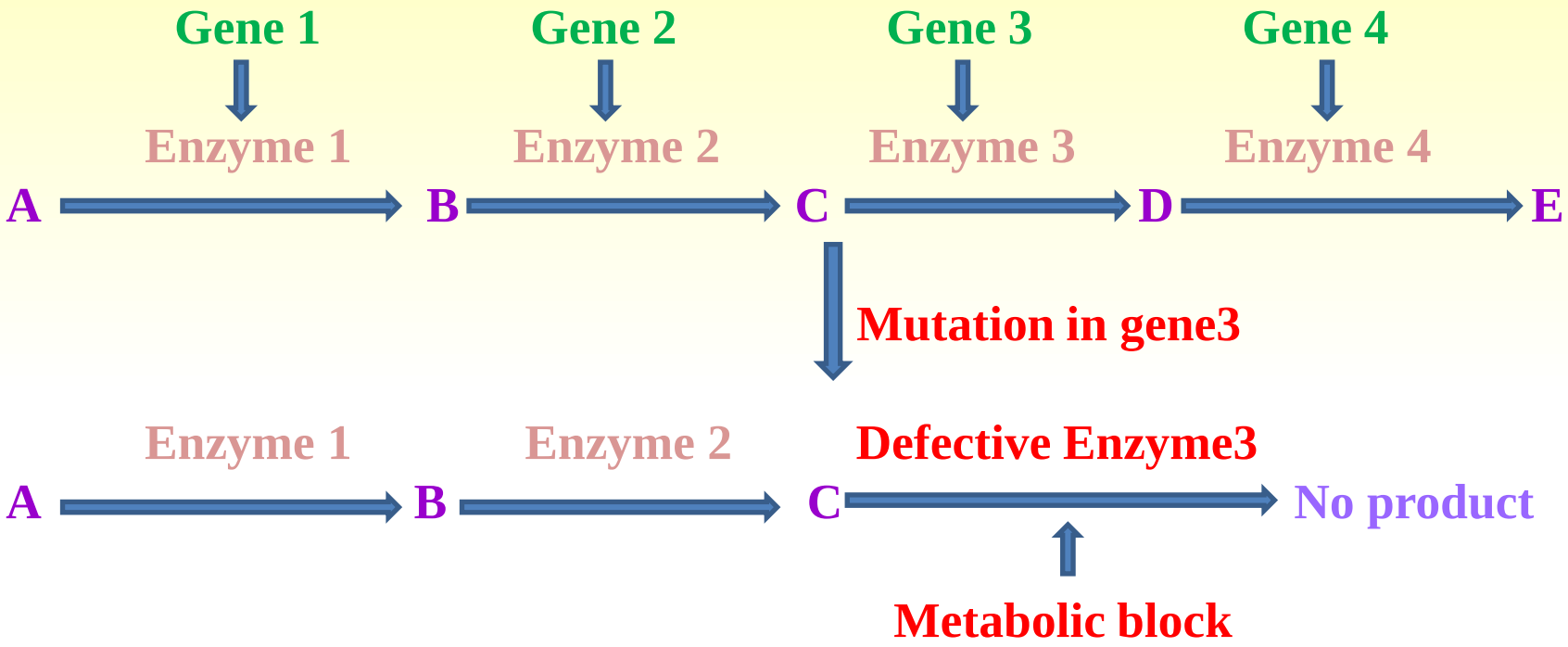
Structural Organization of Gene

ONE GENE-ONE POLYPEPTIDE CONCEPT

A gene - determines or affects a single character or phenotype

A gene - determines or codes for one enzyme

Garrod's study:



Accumulation of C

Thus concept might be best stated as one mutant gene-one metabolic block

Many proteinaceous Enzymes have multiple polypeptide chain of identical or non-identical nature

Identical chains - encoded by the same gene

Non-identical chains - encoded by different genes

Examples-

1. Haemoglobins - consist of two or more different polypeptide chains - each polypeptide was found to be the product of separate gene.

2. Tryptophan synthetase of *E.coli*, for example, contains an α-polypeptide, the product of the trpA gene, and a β-polypeptide, the product of the trpB gene.

Thus in new concept the gene-protein relationship is more accurately described by the phrase "one gene-one polypeptide"

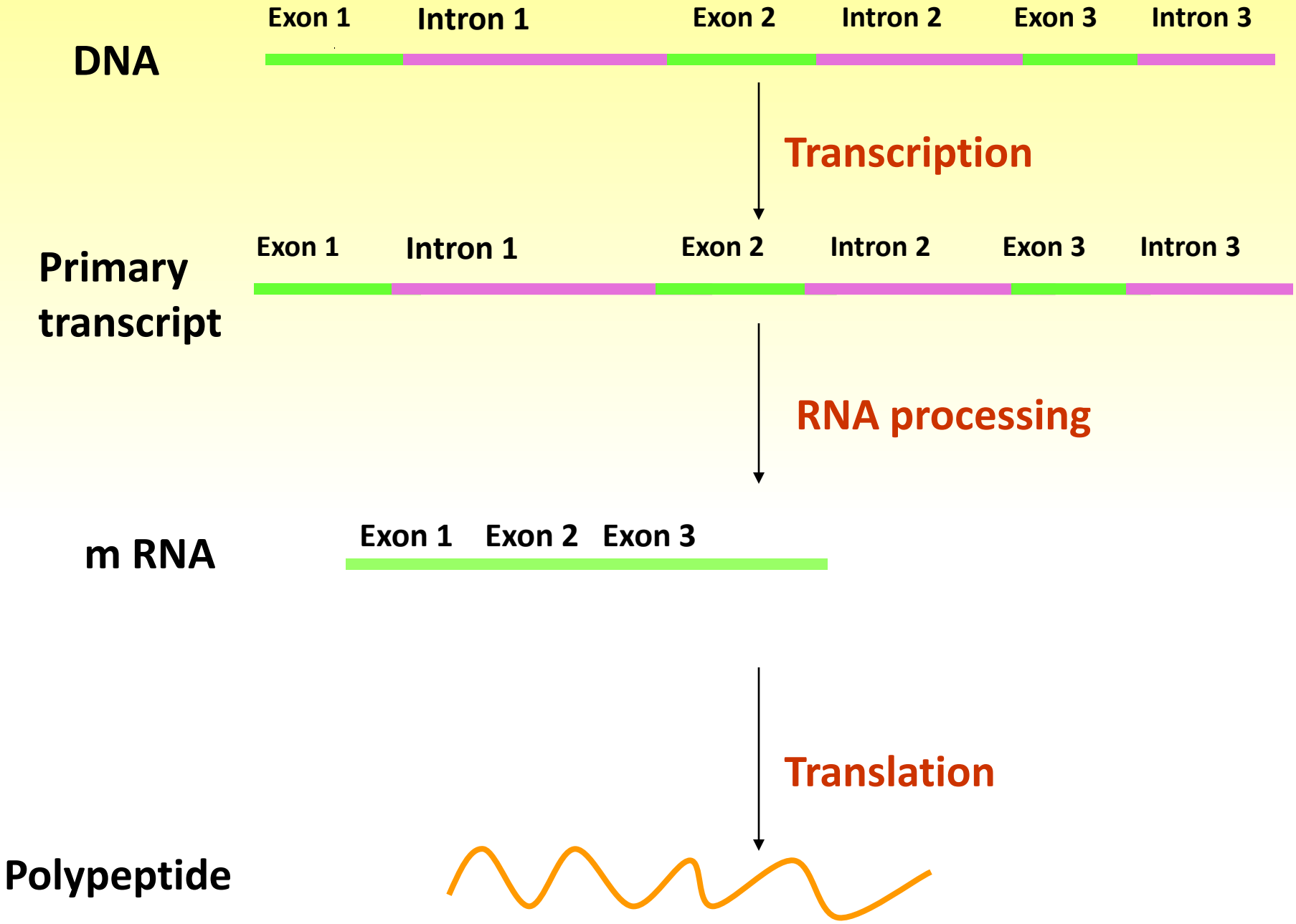
Present biochemical definition of a gene - one gene-one polypeptide

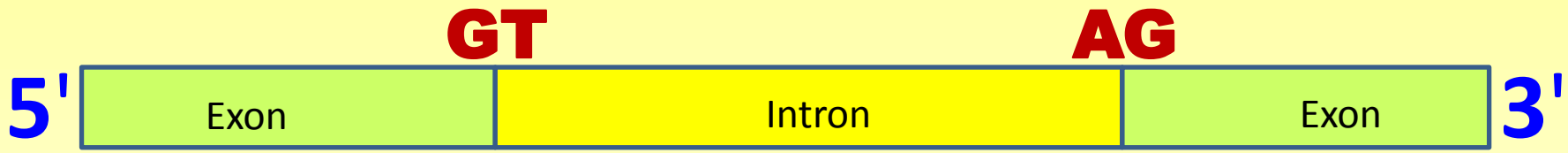
SPLIT GENE

Coding regions (exons) are interrupted by non-coding sequences (introns). Such genes are called split genes since their sequences are split into several parts due to the intervening introns

All the nuclear gene and also mitochondrial and chloroplast gene have split organisation.
Exception - histone genes of Sea Urchin and four Heat Shock genes of *Drosophila*.

First described by Chambon et.al. and that gene was ovalbumin gene of chicken.


DNA
Exon 1
Intron 1
Exon 2
Intron 2
Exon 3
Intron 3
Transcription
Primary transcript
Exon 1
Intron 1
Exon 2
Intron 2
Exon 3
Intron 3
RNA processing
m RNA
Exon 1
Exon 2
Exon 3
Translation
Polypeptide



Introns begin with the sequence GT (5' end) and end with bases AG (3' end) - termed as consensus sequences - believed to be important in the recognition of the limits of introns during splicing

The significance of split organisation of eukaryotic genes is not clear -

In some cases, different exons of a gene code for different active regions of the protein molecules, e.g. in the case of antibodies.

Introns brought together different ancestral genes to form new larger genes.

Introns may also provide for increased recombination rates between exons of a gene and thus be of some significance in the creation of genetic variation.

OVERLAPPING GENE

Each codon is triplet and non-overlapping

AUGUGGAGGCGGACUCAU……AGUUUUUCCCGUACG

Met-Trp-Arg-Arg-Thr-His……...Ser-Phe-Ser-Arg-Thr
Leu-Phe-Pro-Met

Thus 900 bp long DNA can give maximum 300 amino acid long polypeptide

F. Sanger observed that it is not always true
DNA of bacteriophage ϕX174, contains 5386 nucleotide residues, not long enough to code for the nine different proteins
Sometimes, the number of amino acid is greater than the number of respective codons available in the genome.
One gene may be packed partly or fully within another gene.
Same segment could be utilized by two different genes (cistrons) coding for different proteins – but the two cistrons have to function at different times and their nucleotide sequence are translated in two different reading frames.

Such genes sharing their nucleotide sequence either partially or fully are called as overlapping gene.

In plasmids, gene coding antisense RNA tend to overlap with the gene they regulate.
Overlapping gene discovered also from viruses like- MS_2, SV40, λ phage, tryptophan mRNA of *E.coli*.

Sometimes overlapping genes may reflect a mechanism of regulation.

The phenomenon of overlapping gene is an economic device to make better use of genetic material, through packing of more genetic information in less DNA. Thus DNA can potentially code for many more amino acid in different reading frames than the number of amino acid expected to be coded, if no overlapping was present.

REPETITIVE DNA – TANDEM AND INTERSPERSED

Present in multiple copies are called as repetitive DNA
Highly repetitive DNA are also called as satellite DNA

Several times repetitions form a second or 'satellite' band when genomic DNA is separated on a density gradient.

Most satellite DNA is localized to the telomeric or the centromeric region
Play important role only in chromosome organisation but not in convey of functional genetic information.

Satellite DNA are of two types –

Minisatellite - GC- rich, repeating sequence is 10-100 bp long.

Microsatellite – More smaller than minisatellites, repeating sequence is 1-6 bp long.

Repetitive DNA types – tandem and interspersed

Tandem repeats - **a pattern of two or more nucleotides is repeated and the repetitions are directly adjacent to each other.**
Example: ATTCGATTCGATTCG – in which the sequence ATTCG is repeated three times.
Tandem repeats can be very useful in determining parentage.

Interspersed repeats - **scattered throughout the genome, major contributor to genome size.**
Interspersed repetitive sequences are mainly of two types –

SINES (Short Interspersed Elements) and LINES (Long Interspersed Elements) - Both are retrotransposons and their transposition is mediated by reverse transcription.

SINEs: 100-300 bp long, represent reverse transcribed tRNA, rRNA and snRNA molecules, can transcribe but do not encode proteins, function unknown.
LINEs: >5kb long, represent reverse transcribed mRNA molecule. LINEs also can transcribed and some of them encode proteins, function unknown.

In human being SINEs and LINEs accounts 13% and 21% respectively of the total genome.

TRANSPOSONS (AC-DS SYSTEM)

Discovered by B. McClintock.

Transposons (transposable element, mobile gene or jumping gene) - segments of DNA that move or "hop" from one place to another on the same or a different chromosome.

The movement is called transposition.

Examples - IS elements, Tn3 family, Yeast TY elements, Maize Ac-Ds elements, Maize spm and dspm elements, P elements in *Drosophila* etc.

No homology is required for transposition to occur.

In the process of transposition they can cause mutation.

All transposons have inverted repeats at their ends

When transposition occurs, a short sequence at the target site (5-10 bp) is duplicated to form an additional short repeat flanking each end of the inserted transposon.

Transposition requires enzyme - transposase encoded by gene found in the middle of a transposon, can use the gene of another transposon.

Bacteria - four kinds of transposons:

Insertion sequences, transposons, composite transposons, and bacteriophage elements

In eukaryotes - additional type namely retroposons or retrotransposons

Transposition two ways:

1. Conservative transposition - entire double-stranded DNA segment moves to a different site in the genome.
2. Replicative transposition - transposon first replicated, one copy stays, other relocates.

Ac-Ds system – Discovered in Maize by B. McClintock

Breaks →Dissociation

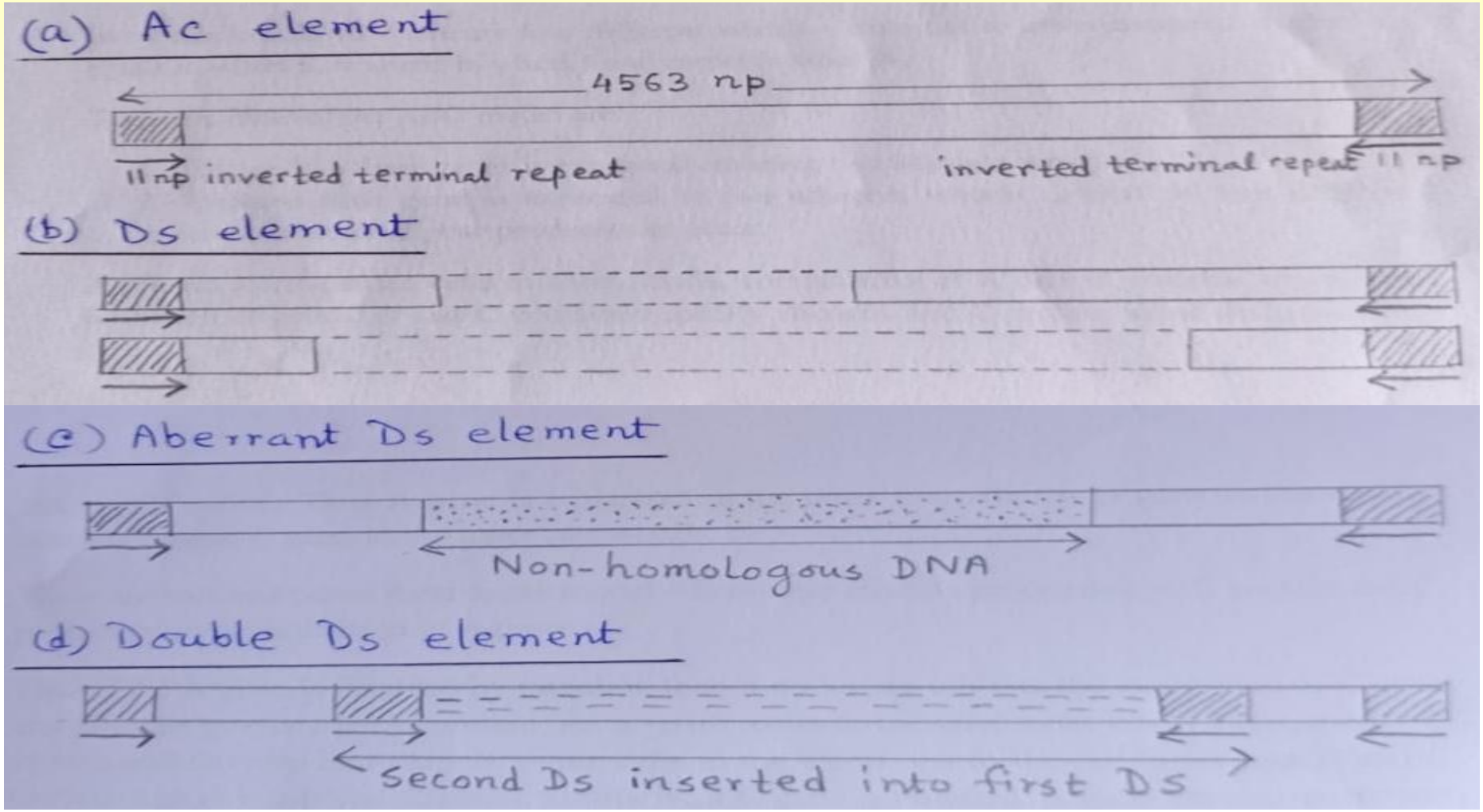
Breaks done by 'Ds' (non-autonomous), activated by 'Ac' (autonomous).

Activating function of Ac elements associated with a protein - transposase. It is diffusible.

Ds have no transposase gene, making it non-functional, can move by using transposase enzyme coded for by Ac. Thus, Ac has to be present in the cell if Ds has to transpose.

Two elements transpose after the chromosome has replicated and the transposition is through a non-replicative mechanism. Thus, transposon disappears from the donor site.

Ac - 4563 nucleotide pairs bounded by 8-nucleotide pair direct repeat.
Other repeat sequences found within the element - 11bp long inverted terminal repeat, crucial for transposition.
All Ac elements structurally similar, but Ds have heterogeneity



HOMOEOTIC GENE (ABC MODEL)

Homoeobox - DNA sequence within some genes involved in regulation of development (morphogenesis).
Genes that have a homeobox are called as homoeotic gene.
Feature - larger size(50 to 100 kb), contain large introns.

Homeobox - 180 bp long, situated at the 3' end of gene, encodes protein domain (the homeodomain) - a transcription factor - can bind DNA – mainly promoter region in a specific manner which typically switch on cascades of other genes.

MADS box - a homeobox -168 bp or 180 bp.
Name derived from -
MCM 1 gene from budding yeast (*Saccharomyces cerevisiae*), AGAMOUS gene from *Arabidopsis thaliana*,
DEFICIENS gene from the Snapdragon (*Antirrhinum majus*) and SRF gene from *Homo sapiens*.

ABCE Quartet model of flowering- Development of floral organs

Floral organ identity genes divided into 5 classes - A, B, C, D and E

Basic idea:

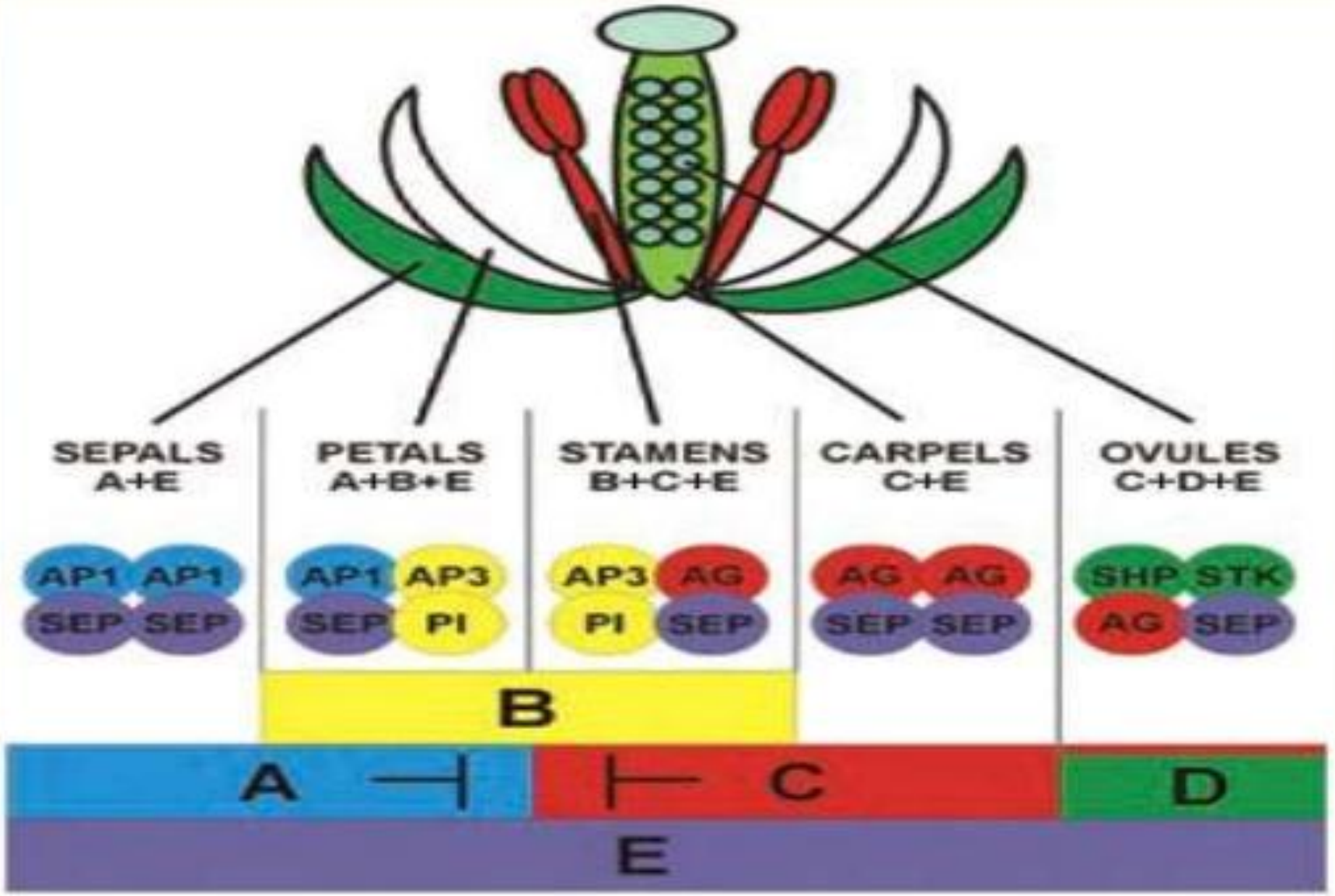
Each of the A, B & C genes is expressed covering two adjacent whorls.

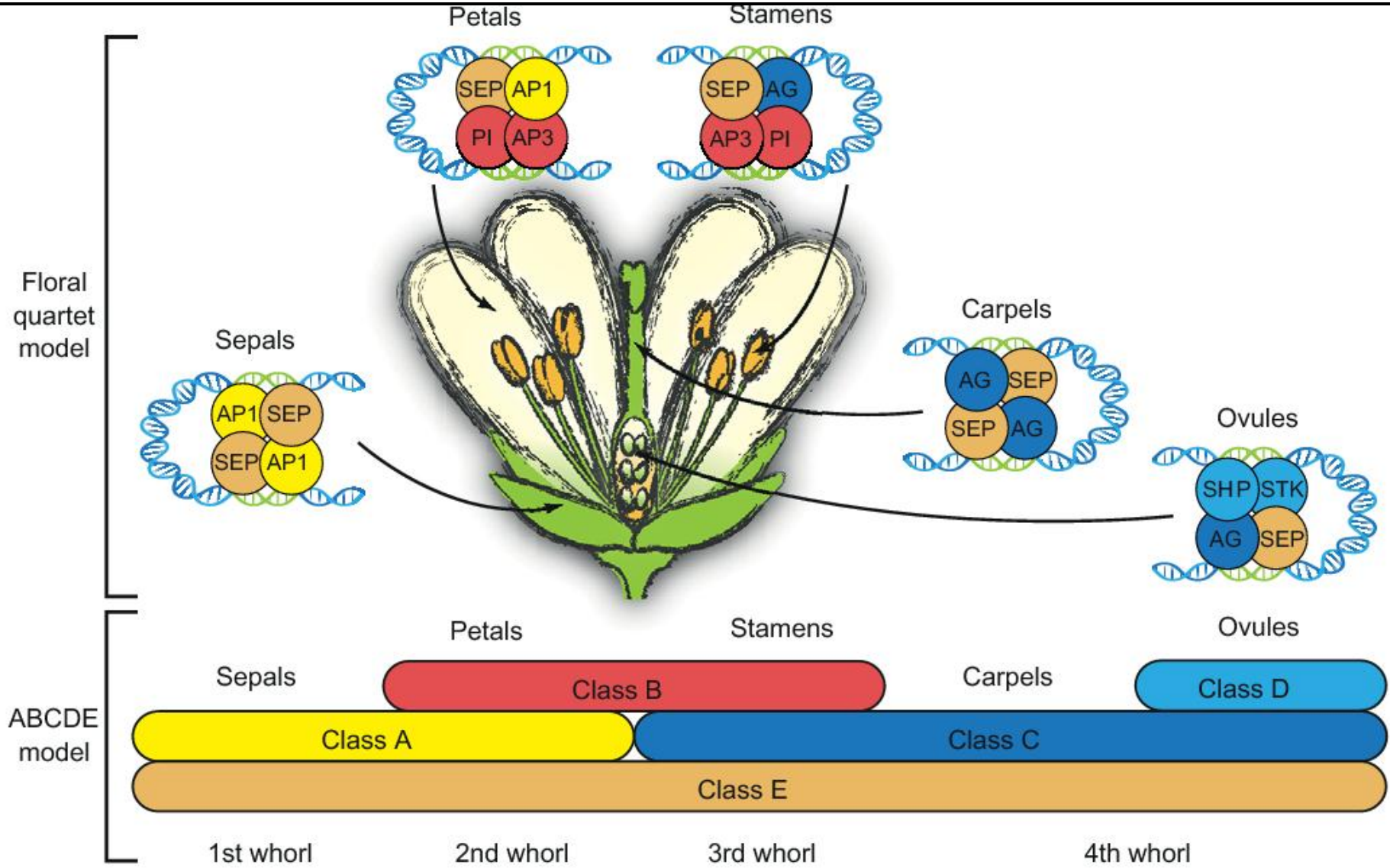
Because each gene is expressed in two adjacent whorls, a total of four different combinations of gene products can occur.

A protein – sepals
A and B proteins – petals
B and C protein – stamens
C protein - carpel
D protein – specify ovule identity
E protein – expressed in all floral whorl

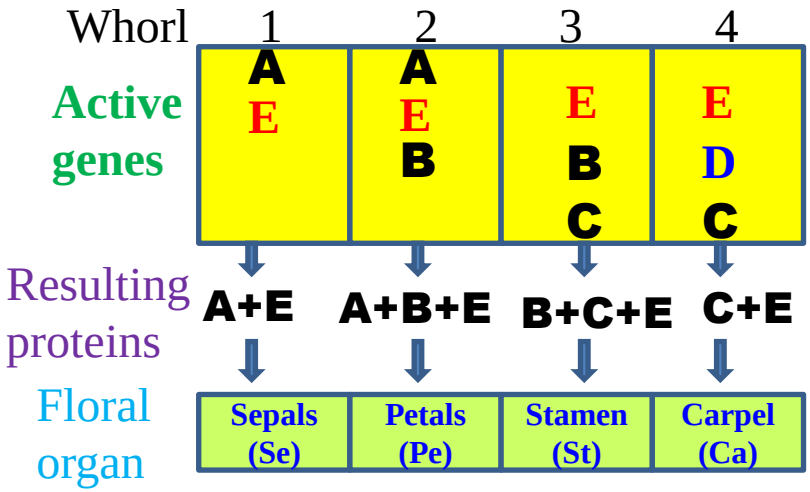
- **A protein inhibits production of C protein and C protein inhibits production of A protein.**

ABCDE Model of Flower Development




Floral quartet model
Petals
SEP
AP1
PI
AP3
Stamens
SEP
AG
AP3
PI
Sepals
AP1
SEP
SEP
AP1
Carpels
AG
SEP
SEP
AG
Ovules
SHP
STK
AG
SEP
ABCDE model
Petals
Stamens
Ovules
Sepals
Class B
Carpels
Class D
Class A
Class C
Class E
1st whorl
2nd whorl
3rd whorl
4th whorl

ABCE model



Products of 3 genes expressed in two adjacent whorls

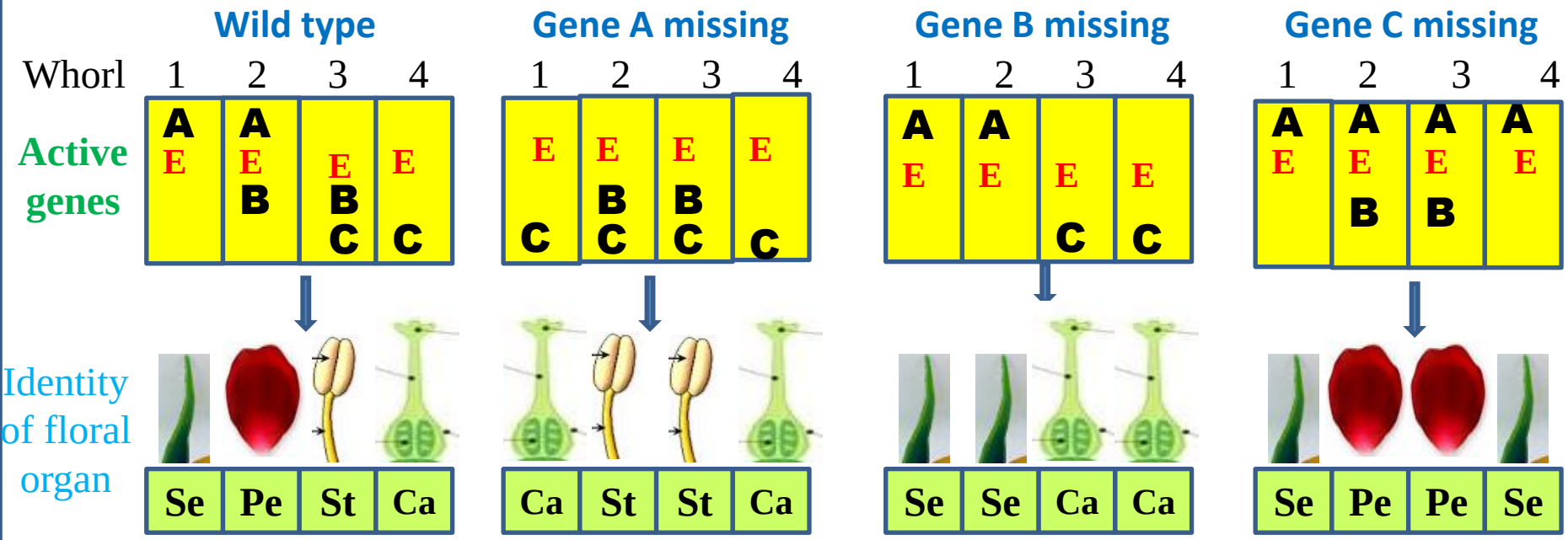
D for ovule identity

Four different combination of protein

Each combination triggers a different floral organ

Assume:
Protein A inhibits C
Protein C inhibits A
(thus no A+C) ⟶

Predictions of ABCE model in case of mutations



If D gene is mutated : fruit/seed shattering is affected

If E gene is mutated: unique type of flower is developed