TWO-YEAR
POST GRADUATE DEGREE PROGRAMME (CBCS)
IN
BOTANY

SEMESTER - IV

Course: BOTDSE T402.2

Molecular Genetics, Advanced Cell Biology, Molecular Breeding & Plant Tissue Culture
(COURSE – I)

Self-Learning Material

DIRECTORATE OF OPEN AND DISTANCE LEARNING
UNIVERSITY OF KALYANI
KALYANI – 741235, WEST BENGAL

Course Preparation Team

Dr. Ankita Pramanik
Assistant Professor
Department of Botany, DODL
Kalyani University

Dr. Ikbal Azahar
Assistant Professor
Department of Botany
Sree Chaitanya College
West Bengal State University

Dr. Bapi Ghosh
Assistant Professor
Department of Botany
Acharya Brojendra Nath Seal College
Cooch Behar Panchanan Barma University

April 2024

Directorate of Open and Distance Learning, University of Kalyani

Published by the Directorate of Open and Distance Learning, University of Kalyani, Kalyani 741235, West Bengal and Printed by New School Book Press, Kolkata - 14.

Director's Messeage

Satisfying the varied needs of distance learners, overcoming the obstacle of distance and reaching the unreached students are the threefold functions catered by Open and Distane Learning (ODL) systems. The onus lies on writers, editors, production professionals and other personal involved in the process to overcome the challenges inherent to curriculam design and production of relevant Self Learning Materials (SLMs). At the University of Kalyani a dedicated team under the able guidance of the Hon'ble Vice-Chancellor has invested its best efforts, professionally and in keeping with the demands of Post Graduate CBCS Programmes in Distance Mode to devise a self-sufficient curriculum for each course offered by the Directorate of Open and Distane Learning (DODL) University of Kalyani.

Development of printed SLMs for students admitted to the DODL within a limited time to cater to the academic requirements of the Course as per standards set by Distance Education Bureau of the University Grants Commission, New Delhi, India under Open and Distance Mode UGC-DEB Regulations, 2020 had been our endeavour. We are happy to have achieved our goal.

Utmost care and precision have been ensured in the development of SLMs, making them useful to the lesrners, besides avoiding errors as far as practicable. Further, suggestions from the stakeholders in this would be welcome.

During the production-process of the SLMs, the team continuously received positive stimulations and feedback from **Professor (Dr.) Amalendu Bhunia**, Hon'ble Vice-Chancellor, University of Kalyani, who kindly accorded directions, encouragements and suggestions, offered constructive criticisms to develop it within proper requirements. We gracefully, acknowledge his inspiration and guidance.

Sincere gratitude is due to the respective chairpersons as well as each and every Members of PGBOS (DODL), University of Kalyani, Heartfelt thanks is also due to the Course Writers- faculty members at the DODL, subject-experts serving at University Post Graduate departments and also to the authors and academicians whose academic contributions have enriched the SLMs. We humbly acknowledge their valuable academic contributions. I would especially like to convey gratitude to all other University dignitaries and personnel involved either at the conceptual or operational level of the DODL of University of Kalyani.

Their persistent and co-ordinated efforts have resulted in the compilation of comprehensive, learners friendly, flexible text that meets curriculum requirements of the Post Graduate Programme through distance mode.

All the Self Learning Materials are self-writing and collected from e-Book, journals and websites.

Director
Directorate of Open & Distance Learning
University of Kalyani

SYLLABUS

COURSE – BOTDSE T402.2

Molecular Genetics, Advanced Cell Biology, Molecular Breeding & Plant Tissue Culture (Course − I)

(Full Marks – 80)

Course	Group	Details Contents Structure		Study hour
BOTDSE T402.2	**Molecular Genetics, Advanced Cell Biology, Molecular Breeding & Plant Tissue**	**Unit 1. Cell Cycle Regulation and Cancer – I**	1. Cell Cycle Regulation and Cancer: Role of proteins in controlling cell cycle; apoptosis; oncogenes and protooncogenes.	1
		Unit 2. Cell Cycle Regulation and Cancer – II	2. Cell Cycle Regulation and Cancer: tumour suppressor genes; role of E2F and p53 in controlling cell cycle.	1
		Unit 3. Genes Directing Flower Development in *Arabidopsis* – I	3. Genes Directing Flower Development in *Arabidopsis*: ABC model, mutations.	1
		Unit 4. Genes Directing Flower Development in *Arabidopsis* – II	4. Genes Directing Flower Development in *Arabidopsis*: floral quartet model of floral organ specification.	1
		Unit 5. Replication of Chromosome Termini – I	5. Replication of Chromosome Termini: End-replication problem and aging in human.	1
		Unit 6. Replication of Chromosome Termini – II	6. Replication of Chromosome Termini: Telomerase.	1
		Unit 7. Epigenetics – I	7. Epigenetics: Introduction, methylation.	1
		Unit 8. Epigenetics – II	8. Epigenetics: histone modifications, epialleles.	1
		Unit 9. RNA Biology – I	9. RNA Biology: Gene silencing through antisense RNA technology and Ribozymes; RNA interference (RNAi) by small regulatory RNAs, different types of small non-coding RNAs.	1

Course	Group	Details Contents Structure		Study hour
BOTDSE T402.2	**Molecular Genetics, Advanced Cell Biology, Molecular Breeding & Plant Tissue**	**Unit 10. RNA Biology – II**	10. RNA Biology: their biogenesis and functions in posttranscriptional gene silencing; applications of RNAi in crop quality improvement.	1
		Unit 11. Genomes and Genomics – I	11. Genomes and Genomics: Concept of genome; Genome sequencing strategies.	1
		Unit 12. Genomes and Genomics – II	12. Genomes and Genomics: Genomes of Yeast, *Arabidopsis* and rice, Genome annotation, Genome duplication, Approaches to analyse differential gene expression- ESTs, Microarrays and their applications.	1
		Unit 13. Genomes and Genomics – III	13. Genomes and Genomics: Reverse genetics-Gene tagging, Gene trapping, Gene silencing and Gene knockout.	1
		Unit 14. Genomes and Genomics – IV	14. Genomes and Genomics: Reverse genetics- Gene silencing and Gene knockout; Metagenomics.	1
		Unit 15. Genome Editing Technologies – I	15. Genome Editing Technologies: CRISPR, and their applications in crop improvements.	1
		Unit 16. Genome Editing Technologies – II	16. Genome Editing Technologies: TALEN, and their applications in crop improvements.	1
		Unit 17. Genome Editing Technologies – III	17. Genome Editing Technologies: LEAPER and their applications in crop improvements.	1
		Unit 18: Proteomics – I	18. Proteomics: Concept of proteome; Functional, structural and differential proteomics; Principle of 2D gel electrophoresis (2-DE); advantages and limitations of 2-DE.	1

Course	Group	Details Contents Structure		Study hour
BOTDSE T402.2	Molecular Genetics, Advanced Cell Biology, Molecular Breeding & Plant Tissue	**Unit 19. Proteomics – II**	19. Proteomics: Protein Fingerprinting; Gel free proteomics (iTRAQ); Mass spectrometry (MALDI-TOF MS).	1
		Unit 20. Proteomics – III	20. Proteomics: Posttranslational modifications of proteins; Applications of proteomics in agriculture.	1
		Unit 21. Membrane Transport – I	21. Membrane Transport: Lipid bilayer, Membrane transport proteins.	1
		Unit 22. Membrane Transport – II	22. Membrane Transport: Active and passive membrane transport, Ion channels.	1
		Unit 23. Intracellular Compartments and Protein Sorting – I	23. Intracellular Compartments and Protein Sorting: Compartmentalization of Higher Cells, Signal peptides and signal patches; Transport of proteins into nucleus.	1
		Unit 24. Intracellular Compartments and Protein Sorting – II	24. Intracellular Compartments and Protein Sorting: Transport of proteins into mitochondria and chloroplasts.	1
		Unit 25. Intracellular Compartments and Protein Sorting – III	25. Intracellular Compartments and Protein Sorting: Transport of proteins from E.R. through the golgi apparatus.	1
		Unit 26. Intracellular Compartments and Protein Sorting – IV	26. Intracellular Compartments and Protein Sorting: Role of M6P (Mannose 6-Phosphate) receptor in lysosomal enzyme sorting.	1
		Unit 27: Intracellular Compartments and Protein Sorting – V	27. Intracellular Compartments and Protein Sorting: Transport from the Plasma membrane via Endosomes- Endocytosis.	1

Course	Group	Details Contents Structure		Study hour
BOTDSE T402.2	**Molecular Genetics, Advanced Cell Biology, Molecular Breeding & Plant Tissue**	**Unit 28. Cell Signaling – I**	28. Cell Signaling: Cell surface and intracellular receptors; Ion channel linked, Signaling via Gprotein linked cell surface receptors.	1
		Unit 29. Cell Signaling – II	29. Cell Signaling: Ion channel linked cell surface receptors.	1
		Unit 30. Cell Signaling – III	30. Cell Signaling: Signaling via Gprotein linked cell surface receptors.	1

Content

COURSE – BOTDSE T402.2

Molecular Genetics, Advanced Cell Biology, Molecular Breeding & Plant Tissue Culture (Course – I)

Hard Core Theory Special Paper **Credits = 8**

Content Structure

1. Introduction
2. Course Objectives
3. Cell Cycle Regulation and Cancer: Role of proteins in controlling cell cycle; apoptosis; oncogenes and protooncogenes; tumour suppressor genes; role of E2F and p^{53} in controlling cell cycle.
4. Genes Directing Flower Development in Arabidopsis: ABC model, mutations; floral quartet model of floral organ specification.
5. Replication of Chromosome Termini: End-replication problem and aging in human; Telomerase.
6. Epigenetics: Introduction, methylation, histone modifications, epialleles.
7. RNA Biology: Gene silencing through antisense RNA technology and Ribozymes; RNA interference (RNAi) by small regulatory RNAs: different types of small non-coding RNAs, their biogenesis and functions in posttranscriptional gene silencing; applications of RNAi in crop quality improvement.
8. Genomes and Genomics: Concept of genome; Genome sequencing strategies, Genomes of Yeast, Arabidopsis and rice, Genome annotation, Genome duplication, Approaches to analyze differential gene expression- ESTs, Microarrays and their applications, Reverse genetics Gene tagging, Gene trapping, Gene silencing and Gene knockout; Metagenomics.
9. Genome Editing Technologies: CRISPR, TALEN, LEAPER and their applications in crop improvements.

10. Proteomics: Concept of proteome; Functional, structural and differential proteomics; Principle of 2D gel electrophoresis (2-DE); advantages and limitations of 2-DE; Protein Fingerprinting; Gel free proteomics (iTRAQ); Mass spectrometry (MALDI-TOF MS);Post translational modifications of proteins; Applications of proteomics in agriculture.
11. Membrane Transport: Lipid bilayer, Membrane transport proteins, Active and passive membrane transport, Ion channels.
12. Intracellular Compartments and Protein Sorting: Compartmentalization of Higher Cells, Signal peptides and signal patches; Transport of proteins into nucleus, mitochondria and chloroplasts; Transport of proteins from E.R. through the golgi apparatus; Role of M6P (Mannose 6-Phosphate) receptor in lysosomal enzyme sorting; Transport from the Plasma membrane via Endosomes- Endocytosis.
13. Cell Signaling: Cell surface and intracellular receptors; Ion channel linked, Signaling via G protein linked cell surface receptors.
14. Let's sum up
15. Suggested Readings
16. Assignments

1. Introduction

Genetics has always been concerned with the fact of how the hereditary information in DNA controls, what an organism looks like and how it works. Classically this involved the use of genetic variants (mutants) to upset the biological function of the cells or organisms and from the effect of these mutations, to make deductions about the way cells and organisms worked. At the molecular end of the subject, the availability of sequence information and genomic analysis, together with sophisticated techniques for gene replacement, and analysis of gene expression patterns gives us much more powerful tools for looking at the way genes work to make us what we are. At the other extreme of the subject, knowledge of genetics is fundamental to an understanding of how organisms, populations and species evolve. Molecular Biology is the branch of biology that studies the structure and function of macro molecules that encode and regulate the flow of genetic information used by living organisms. This course will focus on the structure and content of the three genomes found in plant cells, gene structure, expression, and regulation. Cell biology is a branch of biology that studies the structure, function, and behavior of cells. All living organisms are made of cells. A cell is the basic unit of life that is responsible for the living and functioning of organisms. Cell biology encompasses both prokaryotic and eukaryotic cells and has many subtopics which may include the study of cell metabolism, cell communication, cell cycle, biochemistry, and cell composition. The study of cells is performed using several microscopy techniques, cell culture, and cell fractionation. These have allowed for and are currently being used for discoveries and research pertaining to how cells function, ultimately giving insight into understanding larger organisms. Knowing the components of cells and how cells work is fundamental to all biological sciences while also being essential for research in biomedical fields such as cancer, and other diseases. Research in advanced cell biology is interconnected to other fields such as genetics, molecular genetics, molecular biology.

2. Course Objectives

At the end of the course the learners will be able to:

- To know cell signalling, cell cycle, cancer biology
- Gather knowledge about restriction enzyme, cloning vector , gel electrophoresis and other technique
- To know the methods and techniques of genetics are applicable throughout the spectrum of biological activity
- To learn about epigenetics, gene knockout, genomics, concept of proteome, protein separation and identification techniques, and post-translational modifications of proteins
- Perceive the role of compartmentalization and signalling in cellular biology
- Provide students with a solid understanding of the relationship between structure and function of protein, protein sorting
- Interpret and explain key experiments in the history of cell biology
- To learn and gather knowledge about the basic methods and approaches used in molecular biology, antisense RNA technology, genome editing techniques
- To know membrane transport mehods, membrane transport proteins
- Evaluate and apply knowledge of modern techniques in cellular biology

3. Cell Cycle Regulation and Cancer: Role of proteins in controlling cell cycle; apoptosis; oncogenes and protooncogenes; tumour suppressor genes; role of E2F and p^{53} in controlling cell cycle.

Cell cycle is the sequence of events that occur between two successive cell divisions. DNA replication and the division of a cell are the major events in cell cycle. Eukaryotic cell cycle is divided into four non overlapping phases – G1 (gap1), S (synthesis), G2 & M phase. The DNA component of the cell gets duplicated in the S phase whereas other cellular organelles are continuously synthesized in sufficient quantity to be distributed equally between the daughter cells when the parental cell is big enough to divide in the M phase. The events of chromosome separation occur during the M phase. When a cell withdraws from the process of cell division, it enters a quiescent state termed Go phase. In eukaryotes, DNA replication and mitosis never occur simultaneously. Conversely, in bacteria the analogous process, replication and partition, are coordinated so that partially replicated chromosomes can segregate during rapid growth. All discrete events during the progression of cell are controlled at checkpoints by the regulatory proteins to insure that DNA content in the daughter cells remains constant.

Cell cycle checkpoints

A checkpoint is a stage in the eukaryotic cell cycle at which the cell examines internal and external cues and "decides" whether or not to move forward with division.

There are a number of checkpoints, but the three most important ones are:

The G_1checkpoint, at the G_1/S transition.

The G_2checkpoint, at the G_2/M transition.

The spindle checkpoint, at the transition from metaphase to anaphase.

1. G1 checkpoint (G1/S)

The replication of genetic material is one of the most crucial stage of the cell cycle and hence the major checkpoint in eukaryotic cells lies late in G1 (G1/S) and it controls the

entry to S phase. This regulatory point is called Restriction (R) Point in animal cells and START in yeast. The cell integrates both external and internal signals before taking a decision to divide. Some of these factors include growth factors, nutrient availability, cell size and above all the intactness of the genome. The cells will remain in G1 until it has built the requisite level of enzymes / proteins necessary for DNA replication. Once a cell crosses this checkpoint it will replicate and complete a round of division.

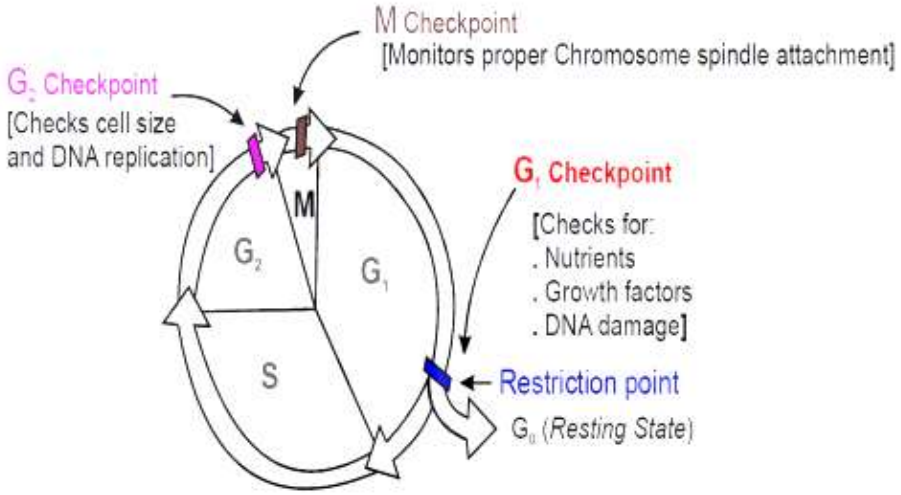


Fig.: Cell cycle checkpoints and restriction points.

2. G2 Checkpoint

- The G2 checkpoint is the second checkpoint in the cell cycle where is present at the transition between G2 and S phase.
- The checkpoint prevents the entry of cells into the S phase of the cycle by preventing the activation of regulators like cyclins and CDKs.
- This checkpoint, like the G1 checkpoint, looks for DNA damage and breaks to prevent the proliferation of mutated or damaged cells.
- As the checkpoint helps maintain genomic stability, studies on the checkpoint help to understand the molecular mechanism of cancer.
- The target of the G2 checkpoint arrest is the CDK2 that usually drives the transition from G2 to the S phase.
- In the checkpoint, DNA damage triggers the activation of the ATM pathway, which causes phosphorylation of ATM and inactivation of checkpoint kinases.

- The checkpoint also involves the p53 genes which inactivate enzymes by the expression of p21 proteins.
- Additional pathways in the G2 checkpoint ensure the stability of the arrest by the expression of proteins like Rb and downregulation of several genes that code for proteins required for the S phase.

3. Metaphase Checkpoint (Spindle checkpoint)

- The metaphase checkpoint or M phase checkpoint or Spindle checkpoint is the checkpoint during mitosis which checks if all the sister chromatids are correctly attached to the spindle fibers.
- The checkpoint ensures that all the chromosomes of cells entering the anaphase are firmly attached to at least two spindle fibers from opposite poles of the cell.
- The separation of chromosomes in anaphase is an irreversible process, which is why this checkpoint is crucial in mitosis.
- The proteins in the checkpoint look for straggler chromosomes that can be detected in the cytoplasm.
- The checkpoint acts by negative regulation of CDC20 which prevents the activation of ubiquitin tag by the anaphase-promoting complex.
- There are different mechanisms to deactivate the checkpoint once all chromosomes are correctly attached.
- One of the important mechanisms is by transporting the motor complex proteins away from the kinetochores. The proteins are then redistributed to the spindle poles.

Controlling Proteins of the Cell Cycle

The main cell cycle regulatory proteins are Cyclin-dependent kinases and Cyclins. Human cells contain multiple loci encoding CDKs and cyclins (13 and 25 loci, respectively). However, only a certain subset of CDK–cyclin complexes is directly involved in driving the cell cycle. They include three interphase CDKs (CDK2, CDK4 and CDK6), a mitotic CDK [CDK1, also known as cell division control protein 2 (CDC2)] and ten cyclins that belong to four different classes (the A-, B-, D- and E-type cyclins).

Cyclin-dependent kinases (Cdks)

Cyclin-dependent kinases (CDKs) are a group of enzymes that work to regulate different processes in the cell cycle after activation by the binding of a cyclin molecule. CDKs are a part of the CMGC group of enzymes consisting of serine or threonine units that are characterized by their dependency on protein subunits. The activity of these enzymes is only observed after the binding of a cyclin molecule followed by the phosphorylation of the threonine unit.

- a G1 Cdk (Cdk4, CDK6)
- an S-phase Cdk ((Cdk2)
- an M-phase Cdk (Cdk1)

Their levels of Cdk in the cell remain fairly stable, but each must bind the appropriate cyclin (whose levels fluctuate) in order to be activated. They add phosphate groups to a variety of protein substrates that control processes in the cell cycle.

Cyclins

Cyclins are a group of proteins that together work to regulate different phases of the cell cycle as core regulators. These proteins regulate the various phases of the cell cycle by either activating the cyclin-dependent kinases or by activating some other enzymes or complexes. Cyclins are specific to different phases as work to regulate different phases of the cycle. The concentration of these cyclins usually remains low for the most part but peaks dramatically if they are needed during the cycle. The activation of the cyclin proteins is stimulated by the binding of the growth factors to the receptors on the cell, which activate the transcription of the cyclin genes. Most of the cyclin proteins act by binding themselves to the cyclindependent kinases, which form a complex. The complex is then responsible for the regulation of the cell cycle.

- a G1 cyclin (cyclin D)
- S-phase cyclins (cyclins E and A)
- mitotic cyclins (cyclins B and A)

The levels of cyclins in the cell rise and fall with the stages of the cell cycle.

Regulation of Cdks-Cyclin activity:

The substrate specificity of CDK is conferred by cyclins. CDK activity not only requires the binding of cyclin molecules for activation but also requires phosphorylation of some target amino acid residues in the CDK molecule itself. For example, CDC28 and Cdc2 kinase in yeast require phosphorylation at Tyr-15 and Thr-161. Phosphorylation at Thr161 induces the kinase activity whereas phosphorylation at Tyr-15 inhibits the kinase activity and dominant over Thr161 phosphorylation.

In S. pombe, Wee1 is the tyrosine kinase which phophorylates Tyr-15. The phosphatase Cdc25 antagonizes the activity of Wee1 and removes the phosphate group from Tyr-15. Both Wee1 and Cdc25 are themselves regulated by intrinsic and extrinsic signals. The decision to proceed to Mitosis or arrest at G2 depends on the levels of the intrinsic and extrinsic signals and regulatory networks which feed into this checkpoint. Apart from cyclin and phosphorylation mediated regulation, the activity of CDKs is regulated by inhibitory proteins called CKIs (CDK-cyclin Inhibitors). Rum1 is a protein which is synthesized to inhibit CDK-cyclin complex and is synthesized throughout the G1 and S phase and prevents the cycle skipping DNA replication and entering mitosis prematurely.

Maturation-promoting factor (MPF)

Maturation-promoting factor or M-phase promoting factor (MPF) is a large-sized diffusible protein that regulates the M-phase of a cell cycle. The protein consists of two subunits; an inert subunit and a kinase subunit. The kinase subunit is capable of activating the inert subunit as well as other molecules. MPF is the regulator of the G2/M transition where it activates activities like nuclear envelope breakdown and chromosome condensation.

Mechanism

During the interphase, the inert subunit of MPF is inactive due to the presence of an enzyme, Wee1. The activation of the MPF unit is brought about by CDC25, which results in the binding of the cyclin molecule to the kinase subunit. After the binding of cyclin to cyclin-dependent kinase, and the activation of CDK, transition into the M phase begins.

The MPF molecules then act by adding phosphate molecules to the nuclear envelope, which causes the breakdown of the membrane. Besides, it also triggers the formation of spindle fibers as a result of microtubule instability. The MPF kinase also phosphorylases several substances like histone H1, which then promotes chromosome condensation. The activity of MPF is further regulated by other components like p34. The phosphorylation of p34 regulates the activity of MPF.

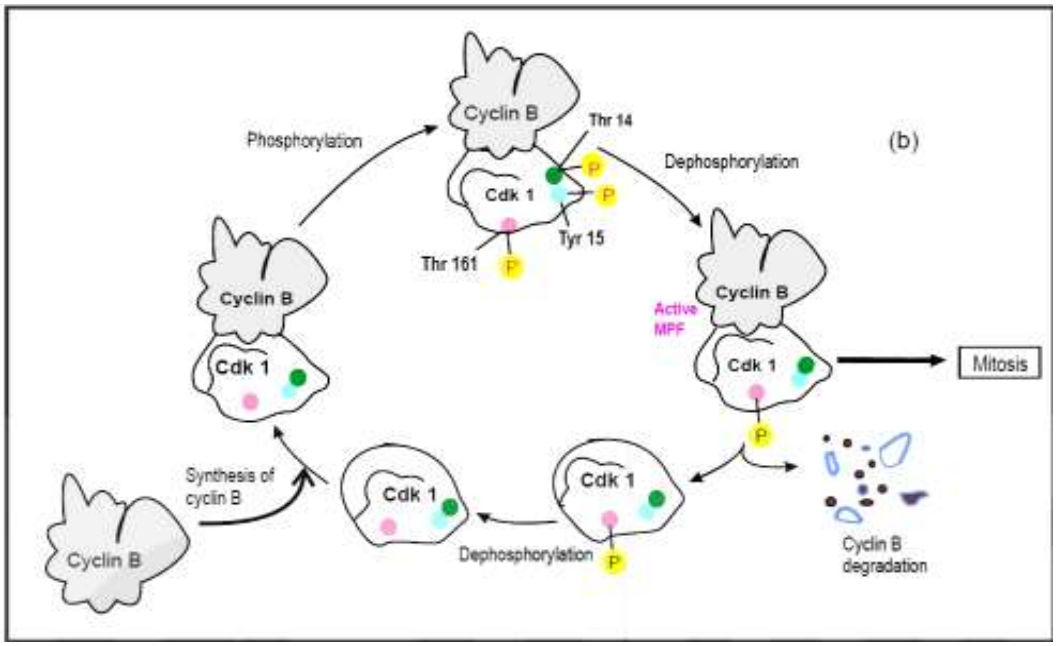


Fig.: Molecular mechanism of MPF regulation.

The anaphase-promoting complex (APC).

The APC is also called the cyclosome, and the complex is often designated as the APC/C.) The APC/C

- triggers the events leading to destruction of the cohesins thus allowing the sister chromatids to separate;
- degrades the mitotic cyclin B.

MPF (cyclin/CDK complex) induces the degradation of cyclin by activating the anaphasepromoting complex/ cyclosome (APC/C) late in metaphase. It is an E3 ubiquitin ligase that attaches a polyubiquitin tag on specific protein regulators including

mitotic cyclins and securin that are then directed for degradation to peptides by a large multiprotein assembly called proteasome (26S). APC also plays a similar role in meiosis II but it requirement in meiosis I depends on the organism.

APC controls anaphase entry, progression and exit from mitosis. It is a large complex that has at least 11 core subunits in vertebrates. Its activity is regulated during mitosis by two activators, Cdc20 and Cdh1. APC-cdc20 is responsible for degradation of cyclin B and securin. Let us see how the key event of anaphase – sister chromatid separation, is initiated by an active APC. Prior to anaphase the activity of separase is inhibited by a protein, securing that associates to form an inactive complex. At anaphase entry, APC degrades securin to relieve separase inhibition resulting in the cleavage of the cohesin subunits

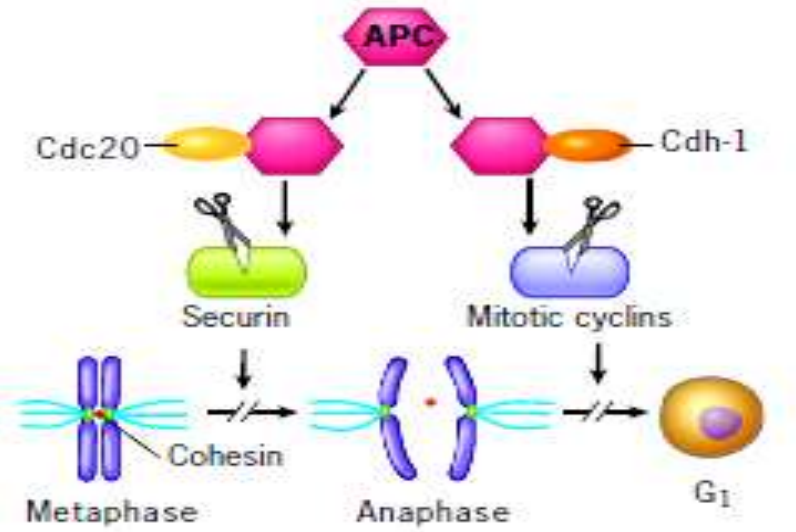


Fig.: APC activities during the cell cycle.

Tumour suppressor genes:

RB

Recent research has revealed that the Retinoblastoma (RB) tumor suppressor protein plays a key role in regulation of the cell cycle. Although the RB gene was discovered

through its association with retinoblastoma, mutations in this gene are also associated with other types of cancer, including small-cell lung carcinomas, osteosarcomas, and bladder, cervical, and prostate carcinomas. Furthermore, mice that are homozygous for an RB knockout mutation die during embryonic development. Thus, the RB gene product is essential for life. The RB gene product, symbolized pRB, is a 105-kilodalton nuclear protein that is involved in cell-cycle regulation. Two genes homologous to RB have been found in mammalian genomes, and their protein products, p107 and p130 (each named for its mass in kilodaltons), may also play key roles in cell-cycle regulation. No human tumors are known to have inactivating mutations in either of these two genes, and mice homozygous for a knockout mutation in either of them do not show abnormal phenotypes. However, mice that are homozygous for knockout mutations in both of these genes die shortly after birth. Thus, together the p107 and p130 members of the RB family of proteins are involved in important cellular processes.

Early in the G1 phase of the cell cycle, pRB binds to the E2F proteins, a family of transcription factors that control the expression of several genes whose products move the cell through its cycle. When E2F transcription factors are bound to pRB, they cannot bind to specific enhancer sequences in their target genes. Consequently, the cell-cycle factors encoded by these genes are not produced, and the machinery for DNA synthesis and cell division remains quiescent. Later in G1, pRB is phosphorylated through the action of cyclin-dependent kinases. In this changed state, pRB releases the E2F transcription factors that have bound to it. These released transcription factors are then free to activate their target genes, which encode proteins that induce the cell to progress through S phase and into mitosis. After mitosis, pRB is dephosphorylated, and each of the daughter cells enters the quiescent phase of a new cell cycle.

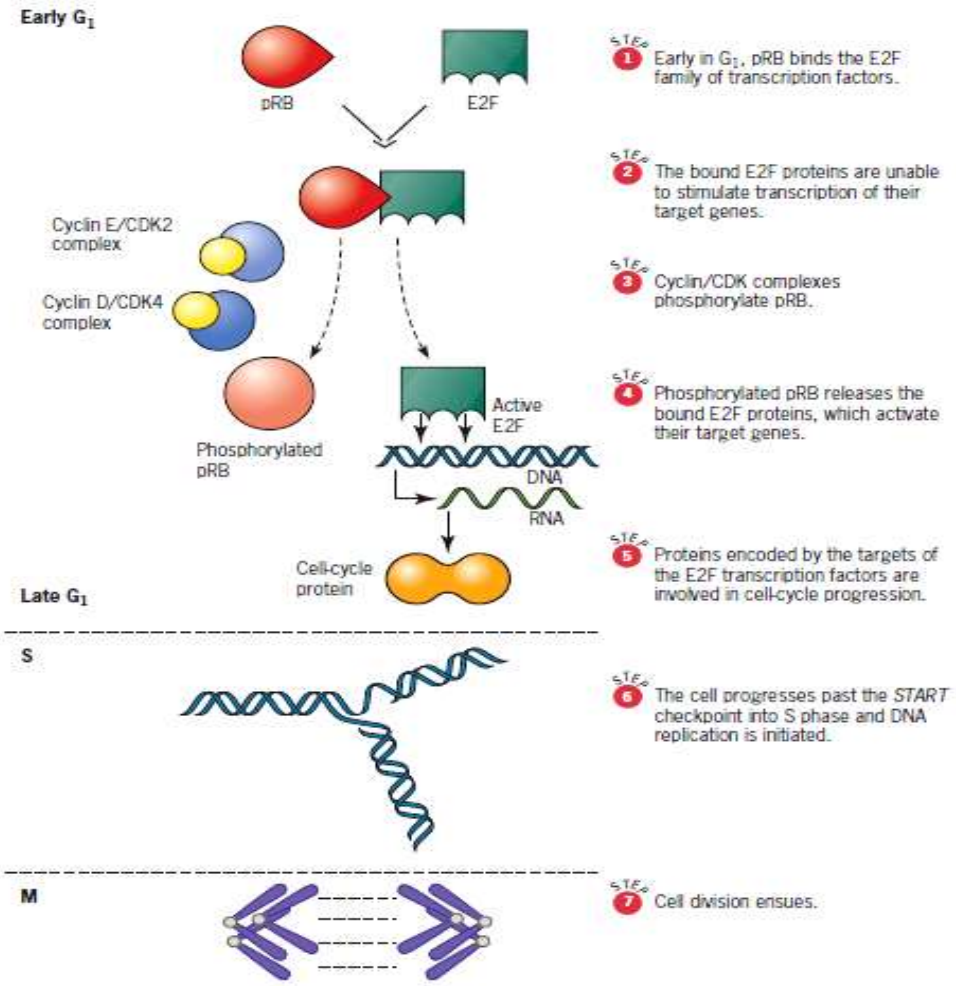


Fig.: Mode of action of RB gene.

p53

The 53-kilodalton tumor suppressor protein p53 was discovered through its role in the induction of cancers by certain DNA viruses. This protein is encoded by a tumor suppressor gene called TP53. Inherited mutations in TP53 are associated with the Li-Fraumeni syndrome, a rare dominant condition in which any of several different types of cancer may develop. Somatic mutations that inactivate both copies of the TP53 gene are also associated with a variety of cancers. In fact, such mutations are found in a

majority of all human tumors. Loss of p53 function is therefore a key step in carcinogenesis.

In response to DNA damage, p53 is phosphorylated, converting it into a stable and active form. Once activated, p53 either stimulates the transcription of genes whose products arrest the cell cycle, thereby allowing the damaged DNA to be repaired, or it activates another set of genes whose products ultimately cause the damaged cell to die.

One prominent factor in the response that arrests the cell cycle is p21, a protein encoded by a gene that is activated by the p53 transcription factor. The p21 protein is an inhibitor of cyclin/CDK protein complexes. When p21 is synthesized in response to cell stress, the cyclin/CDK complexes are inactivated and the cell cycle is arrested. During this timeout, the cell's damaged DNA can be repaired. Thus, p53 is responsible for activating a brake on the cell cycle, and this brake allows the cell to maintain its genetic integrity.

The p53 protein can also mediate another response to cell stress. Instead of orchestrating efforts to repair damage within a cell, p53 may trigger a suicidal response in which the damaged cell is programmed for destruction. The way in which p53 programs cell death is not well understood. One mechanism seems to involve the protein product of the BAX gene. The BAX protein is an antagonist of another protein called BCL-2, which normally suppresses the apoptotic, or cell-death, pathway. When the BAX gene is activated by p53, its protein product releases the BCL-2 protein from its suppressing mode. This release then opens the apoptotic pathway, and the cell proceeds to its own destruction.

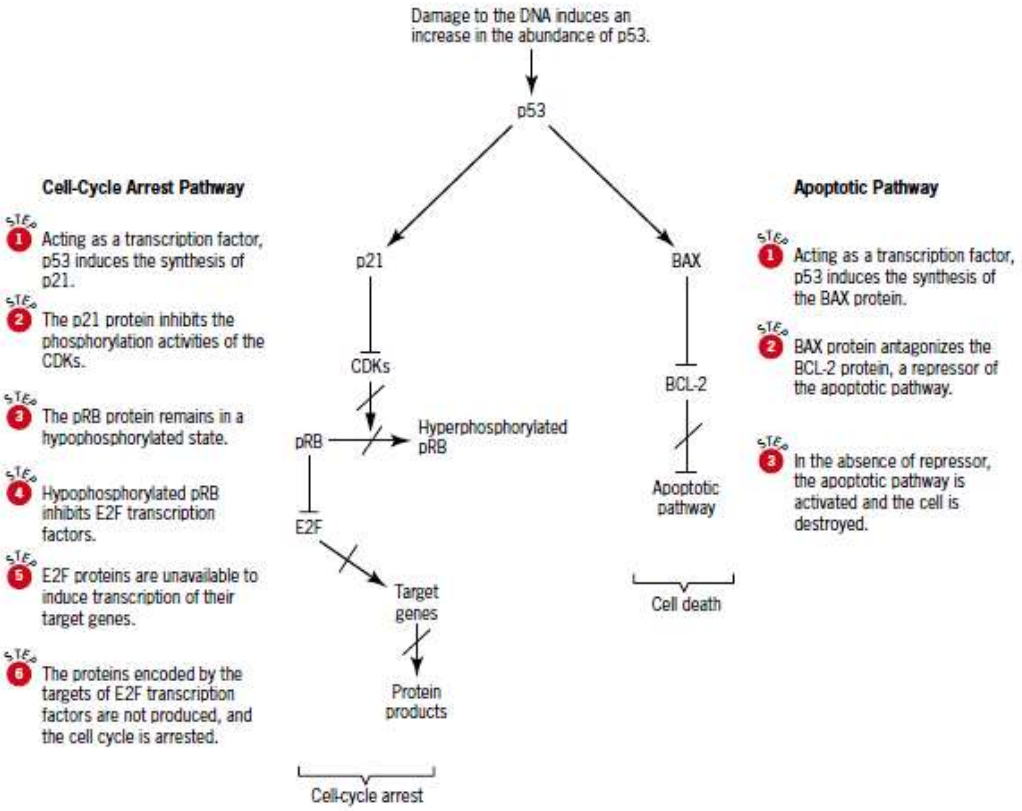


Fig.: Mode of action of p53.

4. Genes Directing Flower Development in *Arabidopsis*: ABC model, mutations; floral quartet model of floral organ specification.

Transition from vegetative to reproductive development in *Arabidopsis*:

Floral meristems can usually be distinguished from vegetative meristems by their larger size. In the vegetative meristem, the cells of the central zone complete their division cycles slowly. The transition from vegetative to reproductive development is marked by an increase in the frequency of cell divisions within the central zone of the shoot apical meristem. The increase in the size of the meristem is largely a result of the increased division rate of these central cells.

During the vegetative phase of growth, the *Arabidopsis* apical meristem produces leaves with very short internodes, resulting in a basal rosette of leaves. When reproductive development is initiated, the vegetative meristem is transformed into the primary inflorescence meristem. The **primary inflorescence meristem** produces an elongated inflorescence axis bearing two types of lateral organs: stem-borne (or inflorescence) leaves and flowers.

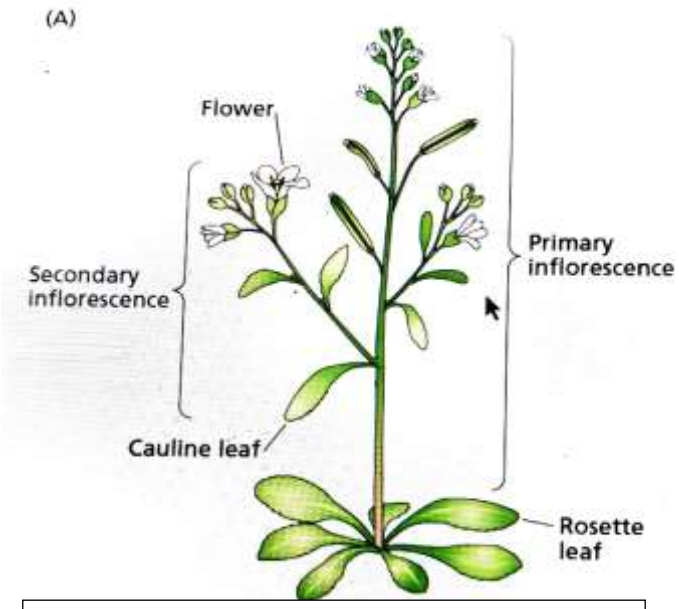


Fig: The shoot apical meristem in *Arabidopsis* generates different organs at different stages of development.

The axillary buds of the stem-borne leave develop into **secondary inflorescence meristems**, and their activity repeats the pattern of development of the primary inflorescence meristem. The Arabidopsis inflorescence meristem has the potential to grow indefinitely and thus exhibits indeterminate growth. Flowers arise from **floral**

meristems that form on the flanks of the inflorescence meristem. In contrast to the inflorescence meristem, the floral meristem is determinate.

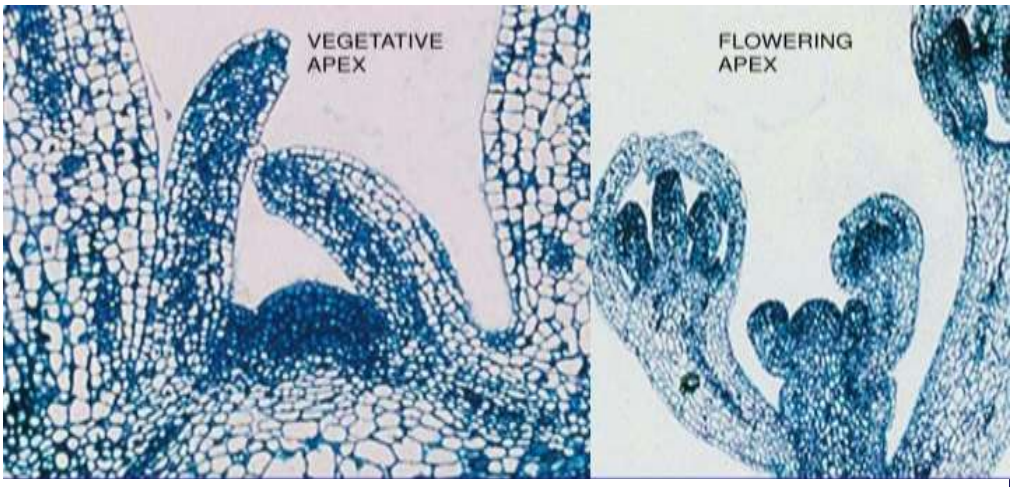


Fig: Vegetative vs. Flowering Shoot Apex in *Arabidopsis*.

Floral organs in *Arabidopsis*:

Floral meristems initiate four different types of floral organs: sepals, petals, stamens, and carpels. These sets of organs are initiated in concentric rings, called **whorls**, around the flanks of the meristem. The initiation of the innermost organs, the carpels, consumes all of the meristematic cells in the apical dome, and only the floral organ primordia are present as the floral bud develops. In ***Arabidopsis***, the whorls are arranged as follows:

- The first (outermost) whorl consists of four sepals, which are green at maturity.
- The second whorl is composed of four petals, which are white at maturity.
- The third whorl contains six stamens, two of which are shorter than the other four.
- The fourth (innermost) whorl is a single complex organ, the gynoecium or pistil, which is composed of an ovary with two fused carpels, each containing numerous ovules, and a short style capped with a stigma

Genes regulate floral development

Studies of mutations have enabled identification of two key categories of genes that regulate floral development: floral meristem identity genes and floral organ identity genes.

1. **Floral meristem identity genes** encode transcription factors that are necessary for the initial induction of floral organ identity genes. They are the positive regulators of floral organ identity in the developing floral meristem.
2. **Floral organ identity genes** directly control floral organ identity. The proteins encoded by these genes are transcription factors that interact with other protein cofactors to control the expression of downstream genes whose products are involved in the formation or function of floral organs.

Floral meristem identity genes:

Floral meristem identity genes must be active for the immature primordia formed at the flanks of the shoot apical meristem or inflorescence meristem to become floral meristems. For example, mutants of snapdragon (*Antirrhinum*) that have a defect in the floral meristem identity gene FLORICAULA (FLO) develop an inflorescence that does not produce flowers. Instead of developing floral meristems in the axils of the bracts, flo mutants develop additional inflorescence meristems in the bract axils. Thus, the wild-type FLO gene controls the determination step that establishes floral meristem identity.

In *Arabidopsis*, LEAFY (LFY), FLOWERING D (FD), SUPPRESSOR OF OVEREXPRESSION OF CONSTANS1 (SOC1), and APETALA1 (AP1) are among the critical genes in the genetic pathway that must be activated to establish floral meristem identity. LFY is the Arabidopsis version of the snapdragon FLO gene. As we saw earlier in the chapter, LFY, FD, and SOC1 play central roles in floral evocation by integrating signals from several different pathways involving both environmental and internal cues. lfy and fd double mutants fail to form flowers, highlighting the roles of LFY and FD as floral meristem identity genes that serve as master regulators for the initiation of floral development.

Table: Genes that regulate flowering in *Arabidopsis*

Gene	**Transcription Factor Family**	**Functions**
CONSTANS (CO)	Zinc finger	Activates flowering in response to long photoperiods
FLOWERING D (FD)	bZIP	Receptor for florigen, activates flowering via AP1
SUPPRESSOR OF OVEREXPRESSION OF CONSTANS1 (SOC1)	MADS	Activates flowering downstream of florigen
FLOWERING LOCUS C	MADS	Floral repressor
LEAFY (LFY)	LFY	Floral meristem identity gene
APETALA1 (AP1)	MADS	Class A homeotic gene, meristem identity
APETALA2 (AP2)	AP2/EREBP	Class A homeotic gene, floral meristem identity
PISTILLATA (PI)	MADS	Class B homeotic gene
AGAMOUS (AG)	MADS	Class C homeotic gene
SEPALLATA (SEP) 1, 2, 3, 4	MADS	Class E homeotic genes
CAULIFLOWER (CAL)	MADS	Meristem identity
FRUITFULL (FUL)	MADS	Floral meristem identity

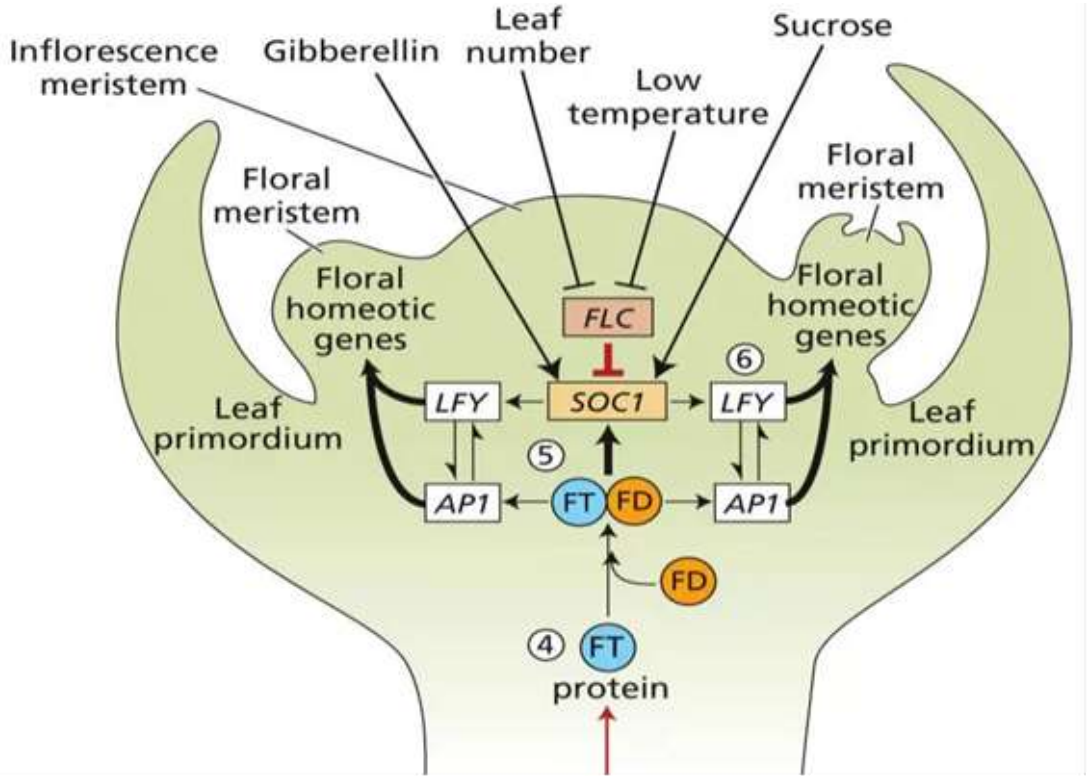


Fig.: Multiple factors regulate flowering in *Arabidopsis*.

The genes that determine floral organ identity were discovered as floral homeotic mutants. The floral organ identity genes were first identified as single-gene homeotic mutations that altered floral organ identity, causing some of the floral organs to appear in the wrong places. Five key genes initially were identified in Arabidopsis that specify floral organ identity: APETALA1 (AP1), APETALA2 (AP2), APETALA3 (AP3), PISTILLATA (PI), and AGAMOUS (AG). Mutations in these genes dramatically altered the structure, and thus the identity, of the floral organs produced in two adjacent whorls. For example, plants with the ap2 mutation lacked sepals and petals. Plants bearing ap3 or pi mutations produced sepals instead of petals in the second whorl, and carpels instead of stamens in the third whorl. Plants homozygous for the ag mutation lacked both stamens and carpels. Because mutations in these genes change floral organ identity without affecting the initiation of flowers, they are, by definition, homeotic genes.

The role of organ identity genes in floral development is dramatically illustrated by experiments in which two or three activities are eliminated by loss-of-function mutations. In quadruple-mutant Arabidopsis plants (ap1, ap2, ap3/pi, and ag) floral meristems no longer produce floral organs but rather produce green leaflike structures; these leaflike organs are produced with a whorled phyllotaxy typical of normal flowers. This experimental result shows that leaves are the "ground state" of organs produced by shoot meristems, and that the activity of additional genes such as AP1 and AP2 are required to convert the leaflike, "ground-state" organs into petals, sepals, stamens, and pistils.

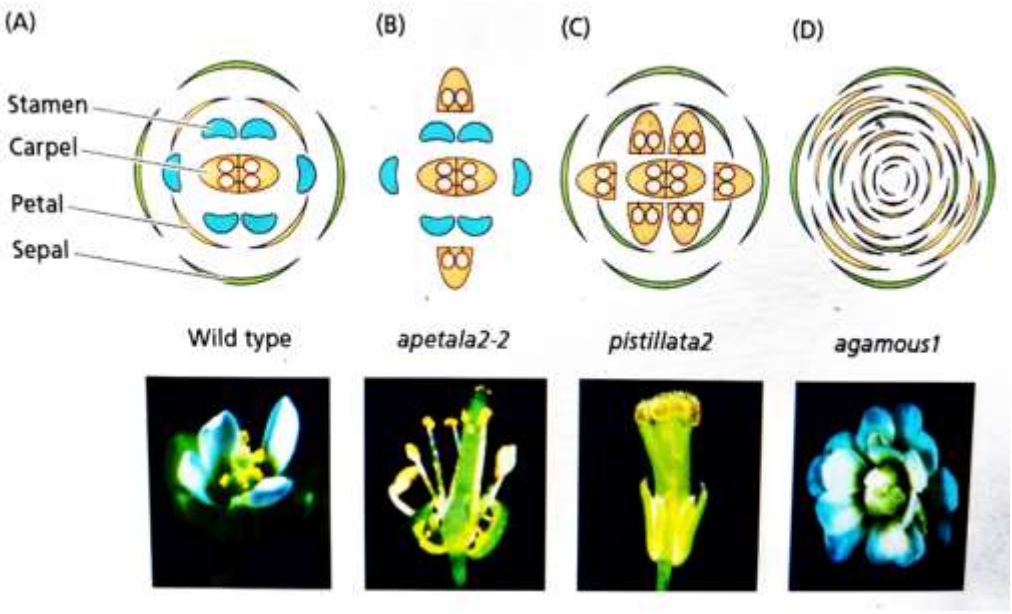
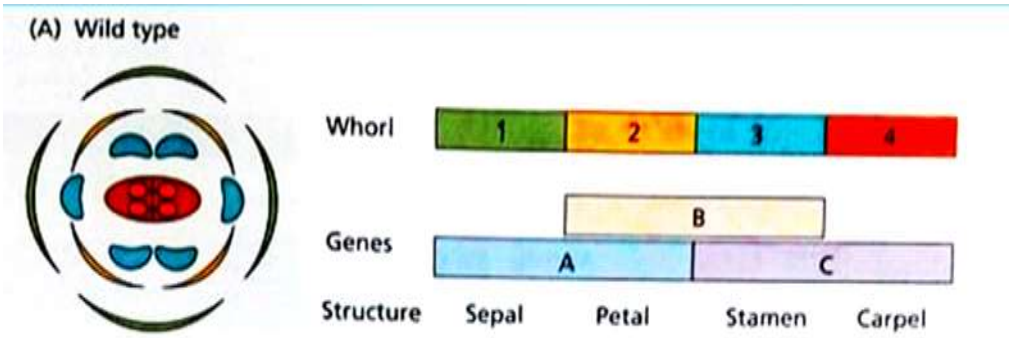


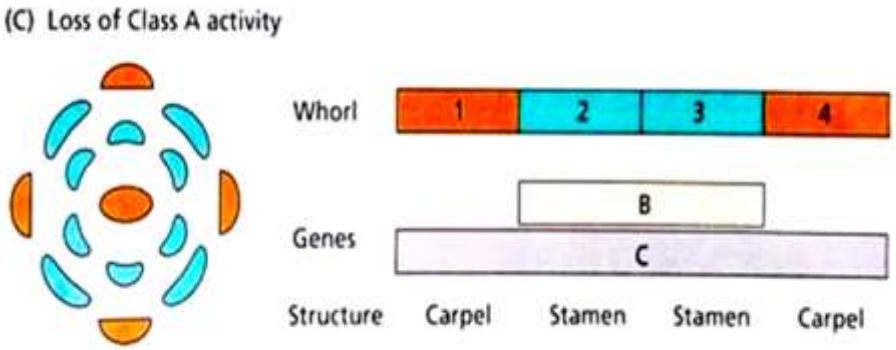
Fig.: Mutations in the floral organ identity genes dramatically alter the structure of the flower.

The ABC model and its Mutations:

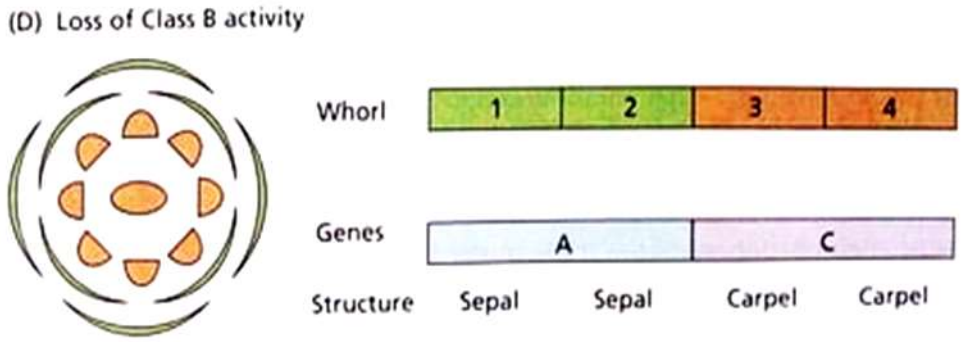
The ABC model partially explains the determination of floral organ identity. The ABC model was discovered by Elliot Meyerowitz and Enrico Coen in 1991. The five floral organ identity genes described above fall into three classes—A, B, and C—defining three different kinds of activities encoded by three distinct types of genes.



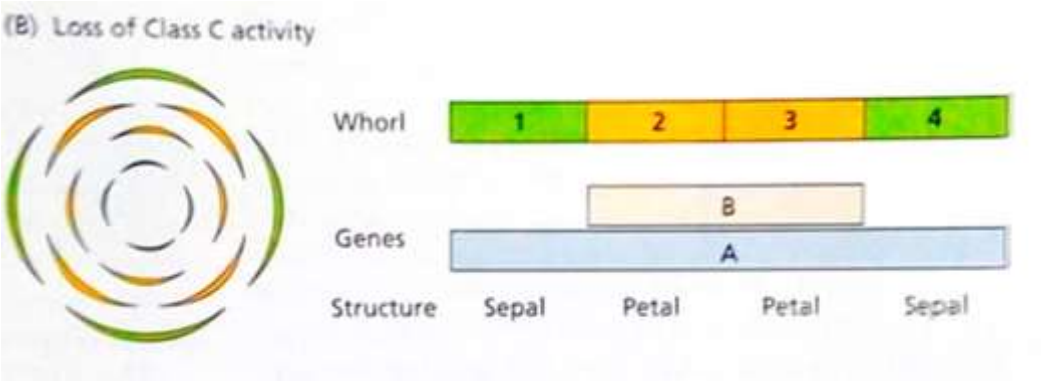
- **Class A activity**, encoded by AP_1 and AP_2, controls organ identity in the first and second whorls. Loss of Class A activity results in the formation of carpels instead of sepals in the first whorl, and of stamens instead of petals in the second whorl.



- **Class B activity**, encoded by AP_3 and PI, controls organ determination in the second and third whorls. Loss of Class B activity results in the formation of sepals instead of petals in the second whorl, and of carpels instead of stamens in the third whorl.



- **Class C activity**, encoded by AG, controls events in the third and fourth whorls. Loss of Class C activity results in the formation of petals instead of stamens in the third whorl. Moreover, in the absence of Class C activity, the fourth whorl (normally a carpel) is replaced by a new flower. As a result, the fourth whorl of an ag mutant flower is occupied by sepals. The floral meristem is no longer determinate. Flowers continue to form within flowers, and the pattern of organs (from outside to inside) is: sepal, petal, petal; sepal, petal, petal; and so on.



The **ABC model** accounts for many observations in two distantly related eudicot species (snapdragon and Arabidopsis), and provides a way of understanding how relatively few key regulators can combinatorically provide a complex outcome. The ABC model postulates that organ identity in each whorl is determined by a unique combination of the three organ identity gene activities.

- **Class A activity** alone specifies sepals.
- **Class A and B activities** are required for the formation of petals.
- **Class B and C activities** form stamens.
- **Class C activity** alone specifies carpels.

The model further proposes that Class A and C activities mutually repress each other; that is, both A- and C-class genes exclude each other from their expression domains, in addition to their function in determining organ identity.

Although the patterns of organ formation in wildtype flowers and most of the mutants are predicted and explained by this model, not all observations can be accounted for by the ABC genes alone. For example, expression of the ABC genes throughout the plant

does not transform vegetative leaves into floral organs. Thus, the ABC genes, while necessary, are not sufficient to impose floral organ identity onto a leaf developmental program. The transcription factors encoded by floral meristem identity genes are also required for the formation of petals, stamens, and carpels.

ABCE model:

Since the A, B, and C genes were identified, another class of floral homeotic genes, the Class E genes, has been discovered. Mutations in three of the other genes identified in mutant screens for floral homeotic mutants, AGAMOUS-LIKE1-3 (AGL1-3), produced only subtle phenotypes when mutated individually. However, the flowers of the agl1/agl2/agl3 triple mutants consisted of sepal-like structures only, suggesting that the subtle phenotypes previously observed in the three individually mutated AGL genes were due to functional redundancy. Because of the sepal-rich phenotype of the triple mutant, the three AGL genes were renamed SEPALLATA1-3 (SEP1-3) and were added to the ABC model as Class E genes.

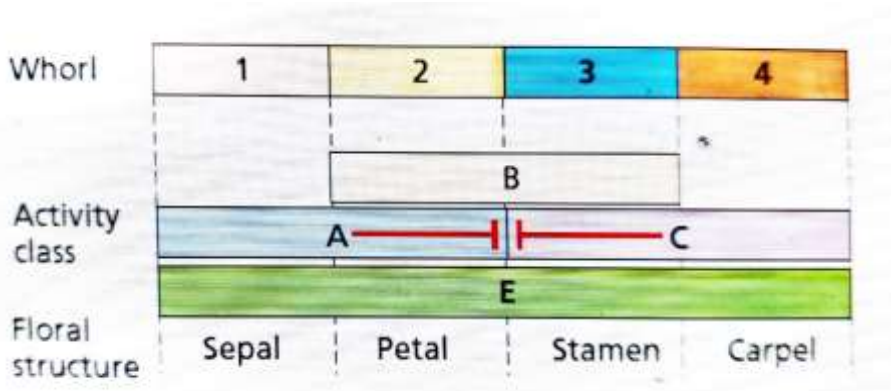


Fig.: ABCE model for floral organ determination in which SEPs act as Class E genes required for floral organ identity.

Another SEPALLATA gene, SEP4, is required redundantly with the other three SEP genes to confer sepal identity, and contributes to the development of the other three organ types. sep quadruple mutants show a conversion of all four floral organ types into leaflike structures, similar to the ap1, ap2, ap3/pi, and ag quadruple mutant. Remarkably, by expressing Class E genes in combination with Class A and B genes, it is possible to convert both cotyledons and vegetative leaves into petals.

The ABCE model was formulated based on genetic experiments in *Arabidopsis* and *Antirrhinum*. Flowers from different species have evolved diverse structures by modifying the regulatory networks described by the ABCE model.

Quartet Model:

All homeotic genes that have been identified so far, in both plants and animals, encode transcription factors. However, unlike animal homeotic genes, which contain homeobox sequences, most plant homeotic genes belong to a class of related sequences known as **MADS box genes**. The acronym MADS is based on the four founding members (MCM1, AGAMOUS, DEFICIENS, and SRF) of a large gene family.

Many of the genes that determine floral organ identity are MADS box genes, including the DEFICIENS gene of snapdragon and the AGAMOUS (AG), PISTILLATA (PI), and APETALA3 (AP3) genes of Arabidopsis. The MADS box genes share a characteristic, conserved nucleotide sequence known as a MADS box, which encodes a protein structure known as the MADS domain. Adjacent to the MADS domain is an intervening region followed by the K domain, which is a coiled-coil region primarily involved in protein–protein interactions. The MADS box gene transcription factors form tetramers that bind to $CC(A/T)_6GG$ sequences, the so-called CArG-box, in the regulatory regions of their target genes. When the tetramers bind two different CArGboxes on the same target gene, the boxes are brought

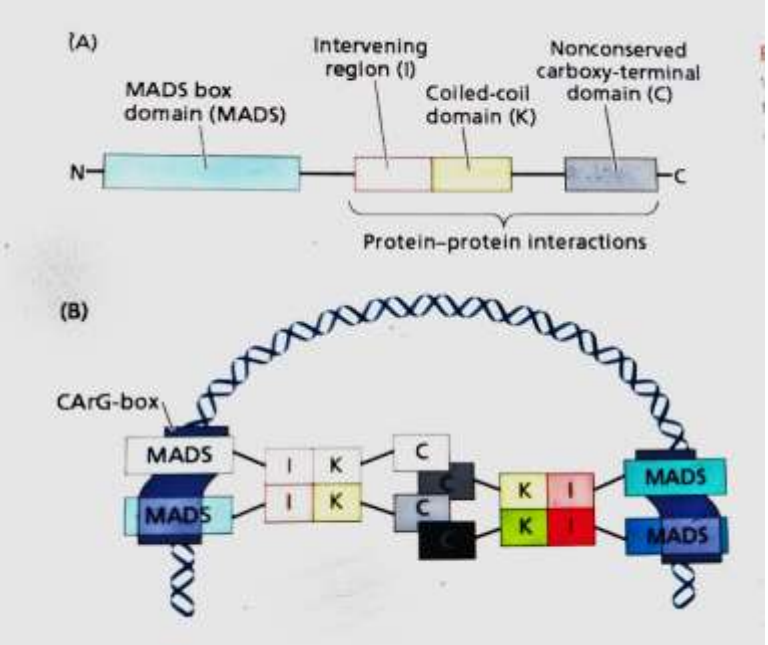


Fig: Model of MADS box domain interaction with target genes. (A) Domain structure of MADS box transcription factors. (B) Tetramers of MADS box transcription factors bind to a pair of CArG-box motifs in the regulatory regions of their target genes, which causes DNA bending.

into close proximity, causing DNA bending. Not all homeotic genes are MADS box genes, and not all genes containing the MADS box domain are homeotic genes. For example, the homeotic gene AP2 is a member of the AP2/ERF (ethylene-responsive element-binding factor) family of transcription factors, and the floral meristem identity gene SOC1 is a MADS box gene.

To gain a more mechanistic understanding of the ABCE model, a biochemical interaction model, called the **Quartet Model**, has been proposed (G. Theissen and H. Saedler, 2001). In the Quartet Model, tetramers of combinations of the ABCE genes directly bind DNA and specify floral organs. The model is based on the observation that MADS box genes dimerize, and two dimers can come together, forming a tetramer. These tetramers are hypothesized to bind CArG-boxes on target genes and modify their expression. Although all MADS box proteins can form higher-order complexes, not all of these are able to bind DNA. For example, Class B factors (AP3 and PI) bind DNA only as heterodimers, whereas both homodimers and heterodimers of Classes A, C, and E can bind DNA. According to the model, tetramers composed of different homodimers and heterodimers of MADS domain proteins can exert combinatorial control over floral organ identity. For example, the AP3–PI heterodimer interacts directly with AP1 and SEP3 to promote petal formation, and indirectly with AG with the help of SEP3 acting as a scaffold. In general, the SEP proteins seem to act as cofactors that provide flower-specific activity to the ABC genes by making complexes of their products.

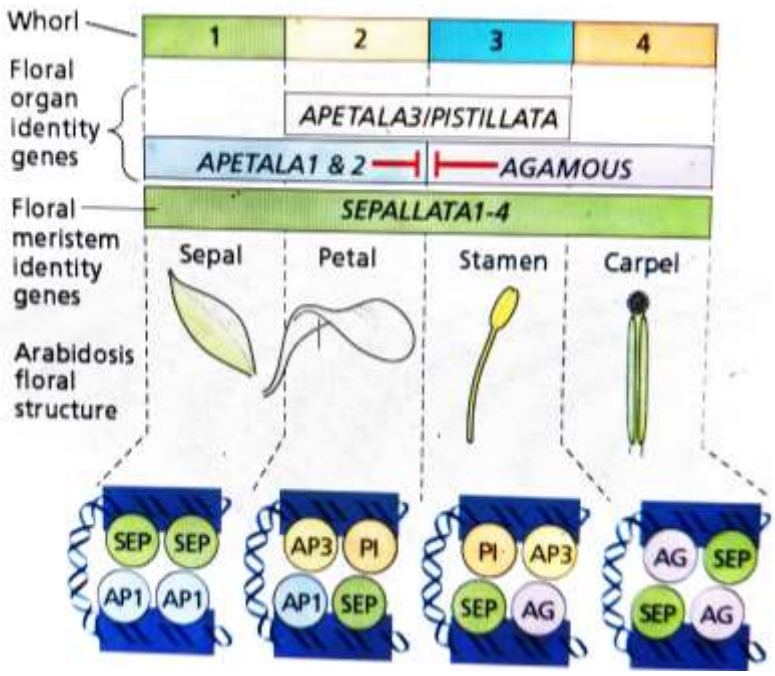


Fig.: Quartet model of floral organ specification in *Arabidopsis*.

- In whorl 1, expression of Class A (AP1 andAP2) and E (SEP) genes results in the formation of sepals.
- In whorl 2, expression of Class A (AP1, AP2), B (AP3, PI), and E (SEP) genes results in the formation of petals.
- In whorl 3, the expression of Class B (AP3, PI), C (AG), and E (SEP) genes causes the formation of stamens.
- In whorl 4, Class C (AG) and E (SEP) genes specify carpels.

In addition, Class A activity (AP1 and AP2) represses Class C activity (AG) in whorls 1 and 2, while Class C activity represses Class A activity in whorls 3 and 4. According to the Quartet Model, the identity of each of the floral organs is determined by four combinations of the floral homeotic proteins known as the MADS box proteins. Two dimers of each tetramer recognize two different DNA sites (termed CArG-boxes, shown here in yellow) on the same strand of DNA, which are brought into close proximity by DNA bending. Note that SEPALLATA proteins are present in all four complexes, serving to recruit the other proteins to the complex.

Class D genes are required for ovule formation

According to the ABCE model, carpel formation requires the activities of the Class C and E genes. However, it appears that a third group of MADS box genes closely related to the Class C genes is required for ovule formation. These ovule-specific genes have been called Class D genes. Since the ovule is a structure within the carpel, Class D genes are not, strictly speaking, "organ identity genes," although they function in much the same way in specifying ovules. Class D activities were first discovered in *Petunia*. Silencing two MADS box genes known to be involved in floral development in petunia, FLORALBINDING PROTEIN7/11 (FBP7/11), resulted in the growth of styles and stigmas in the locations normally occupied by ovules. When the FBP11 was overexpressed in *petunia*, ovule primoridia formed on the sepals and petals.

In Arabidopsis, the ectopic expression of either SHATTERPROOF1 or SHATTERPROOF2 (SHP1, SHP2) or SEEDSTICK (STK) is sufficient to induce the transformation of sepals into carpeloid organs bearing ovules. Moreover, stk/shp1/shp2 triple mutants lack normal ovules. Thus, in addition to the Class C and E genes, Class D genes are required for normal ovule development.

5. Replication of Chromosome Termini: End-replication problem and aging in human; Telomerase.

End replication problem:

The requirement for an RNA primer to initiate all new DNA synthesis creates a dilemma for the replication of the ends of linear chromosomes, called the **end replication problem**. This difficulty is not observed during the duplication of the leading-strand template. In that case, a single internal RNA primer can direct the initiation of a DNA strand that can be extended to the extreme 5' terminus of its template. In contrast, the

requirement for multiple primers to complete lagging-strand synthesis means that a complete copy of its template cannot be made. Even if the end of the last RNA primer for Okazaki fragment synthesis anneals to the final base of the lagging-strand template, once this RNA molecule is removed, there will remain a short region of unreplicated ssDNA at the end of the chromosome.

Although this shortening would only occur on one of the two strands of the daughter molecule, after the next round of replication occurs both strands of the daughter molecule would be shorter. This means that each round of DNA replication would result in the shortening of one of the two daughter DNA molecules. Obviously, this scenario would disrupt the complete propagation of the genetic material from generation to generation. Slowly, but surely, genes at the end of the chromosomes would be lost.

Organisms solve the end replication problem in a variety of ways. One solution is to use a protein instead of RNA as the primer for the last Okazaki fragment at each end of the chromosome. In this situation, the "priming protein" binds to the lagging-strand template and uses an amino acid to provide an OH (typically a tyrosine) that replaces the 3'-OH normally provided by an RNA primer. By priming the last lagging strand, the priming protein becomes covalently linked to the 5' end of the chromosome. Terminally attached replication proteins of this kind are found at the end of the linear chromosomes of certain species of bacteria (most bacteria have circular chromosomes) and at the ends of the linear chromosomes of certain bacterial and animal viruses.

But most eukaryotic cells use an entirely different solution to replicate their chromosome ends. The ends of eukaryotic chromosomes are called telomeres, and they are generally composed of head-to-tail repeats of a TG-rich DNA sequence. For example, human telomeres consist of many head-to-tail repeats of the sequence 5'-TTAGGG-3'. Although many of these repeats are double-stranded, the 3' end of each chromosome extends beyond the 5' end as ssDNA. This unique structure acts as a novel origin of replication that compensates for the end replication problem. This origin does not interact with the same proteins as other eukaryotic origins, but it instead recruits a specialized DNA polymerase called telomerase.

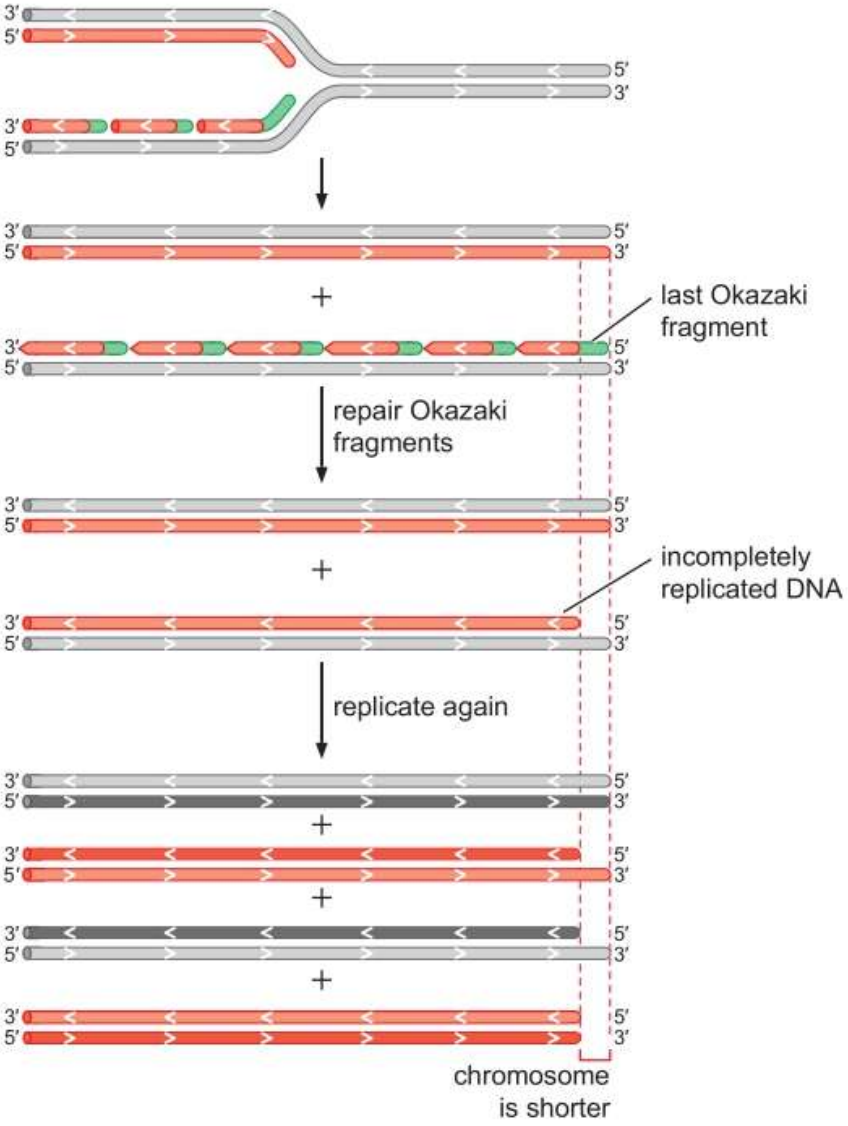


Fig.: The End replication problem.

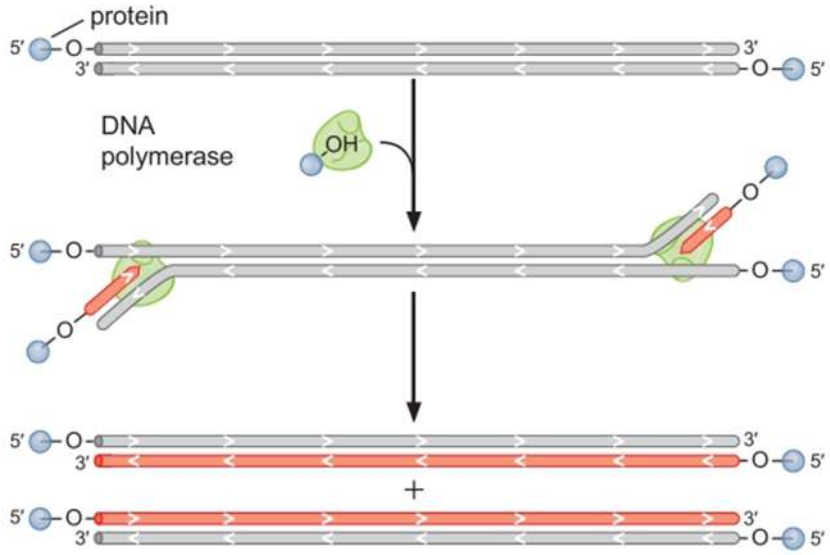


Fig.: Protein priming as a solution to the end replication problem.

Telomerase:

Telomerase is a remarkable enzyme that includes multiple protein subunits and an RNA component (an example of a ribonucleoprotein). This unique enzyme was discovered in 1985 by Elizabeth Blackburn and Carol Greider. They shared the 2009 Nobel Prize in Physiology or Medicine with Jack Szostak, who, along with Blackburn, determined how the unique structures of telomeres protected them from degradation Like all other DNA polymerases, telomerase acts to extend the 3' end of its DNA substrate. But unlike most DNA polymerases, telomerase does not need an exogenous DNA template to direct the addition of new dNTPs. Instead, the RNA component of telomerase serves as the template for adding the telomeric sequence to the 3' terminus at the end of the chromosome. Telomerase specifically elongates the 3'-OH of telomeric ssDNA sequences using its own RNA as a template. As a result of this unusual mechanism, the newly synthesized DNA is single-stranded.

The key to telomerase's unusual functions is revealed by the RNA component of the enzyme, called **"telomerase RNA" (TER)**. Depending on the organism, TER varies in size from 150 to 1300 bases. In all organisms, the sequence of the RNA includes a short

region that encodes about 1.5 copies of the complement of the telomere sequence (for humans, this sequence is 5'-AAUCCCAAUC-3'). This region of the RNA can anneal to the ssDNA at the 3' end of the telomere. Annealing occurs in such a way that a part of the RNA template remains single-stranded, creating a primer:template junction that can be acted on by telomerase. Interestingly, one of the protein subunits of telomerase is a member of a class of DNA polymerases that use RNA templates called "reverse transcriptases" (this subunit is called "telomerase reverse transcriptase," or TERT). These enzymes "reverse-transcribe" RNA into DNA instead of the more conventional transcription of DNA into RNA. Using the associated RNA template, TERT synthesizes DNA to the end of the TER template region but cannot continue to copy the RNA beyond that point. At this point, the RNA template disengages from the DNA product, reanneals to the last four nucleotides of the telomere, and then repeats this process.

The characteristics of telomerase are in some ways distinct and in other ways similar to those of other DNA polymerases.

- The inclusion of an RNA component, the lack of a requirement for an exogenous template, and the ability to use an entirely ssDNA substrate to produce an ssDNA product sets telomerase apart from other DNA polymerases.
- In addition, telomerase must have the ability to displace its RNA template from the DNA product to allow repeated rounds of template-directed synthesis. Formally, this means that telomerase includes an RNA.DNA helicase activity.
- On the other hand, like all other DNA polymerases, telomerase requires a template to direct nucleotide addition, can only extend a 3'-OH end of DNA, uses the same nucleotide precursors, and acts in a processive manner, adding many sequences repeats each time it binds to a DNA substrate.

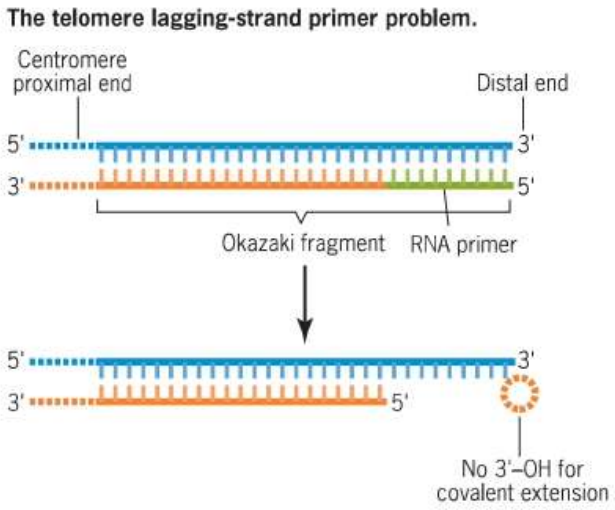
Replication of chromosome termini:

An early reason for thinking that telomeres must have special structures was that DNA polymerases cannot replicate the terminal DNA segment of the lagging strand of a linear chromosome. At the end of the DNA molecule being replicated discontinuously, there would be no DNA strand to provide a free 3'-OH (primer) for polymerization of deoxyribonucleotides after the RNA primer of the terminal Okazaki fragment has been

excised. Either (1) the telomere must have a unique structure that facilitates its replication or (2) there must be a special enzyme that resolves this enigma of replicating the terminus of the lagging strand. Indeed, evidence has shown that both are correct. The special structure of telomeres provides a neat mechanism for the addition of telomeres by an RNA-containing enzyme called **telomerase**.

The telomeres of humans, which contain the tandemly repeated sequence TTAGGG, will be used to illustrate how telomerase adds ends to chromosomes. Telomerase recognizes the G-rich telomere sequence on the 3' overhang and extends it 5' → 3' one repeat unit at a time. Telomerase does not fill in the gap opposite the 3' end of the template strand; it simply extends the 3' end of the template strand. The unique feature of telomerase is that it contains a built-in RNA template. After several telomere repeat units are added by telomerase, DNA polymerase catalyzes the synthesis of the complementary strand. Without telomerase activity, linear chromosomes would become progressively shorter. If the resulting terminal deletions extended into an essential gene or genes, this chromosome shortening would be lethal.

One change observed in many cancer cells is that the genes encoding telomerase are expressed, whereas they are not expressed in most somatic cells. Thus, one approach to cancer treatments has been to try to develop telomerase inhibitors, so that the chromosomes in cancer cells will lose their telomeres and the cells will die. However, other cancer cells do not contain active telomerase, making this approach problematic.



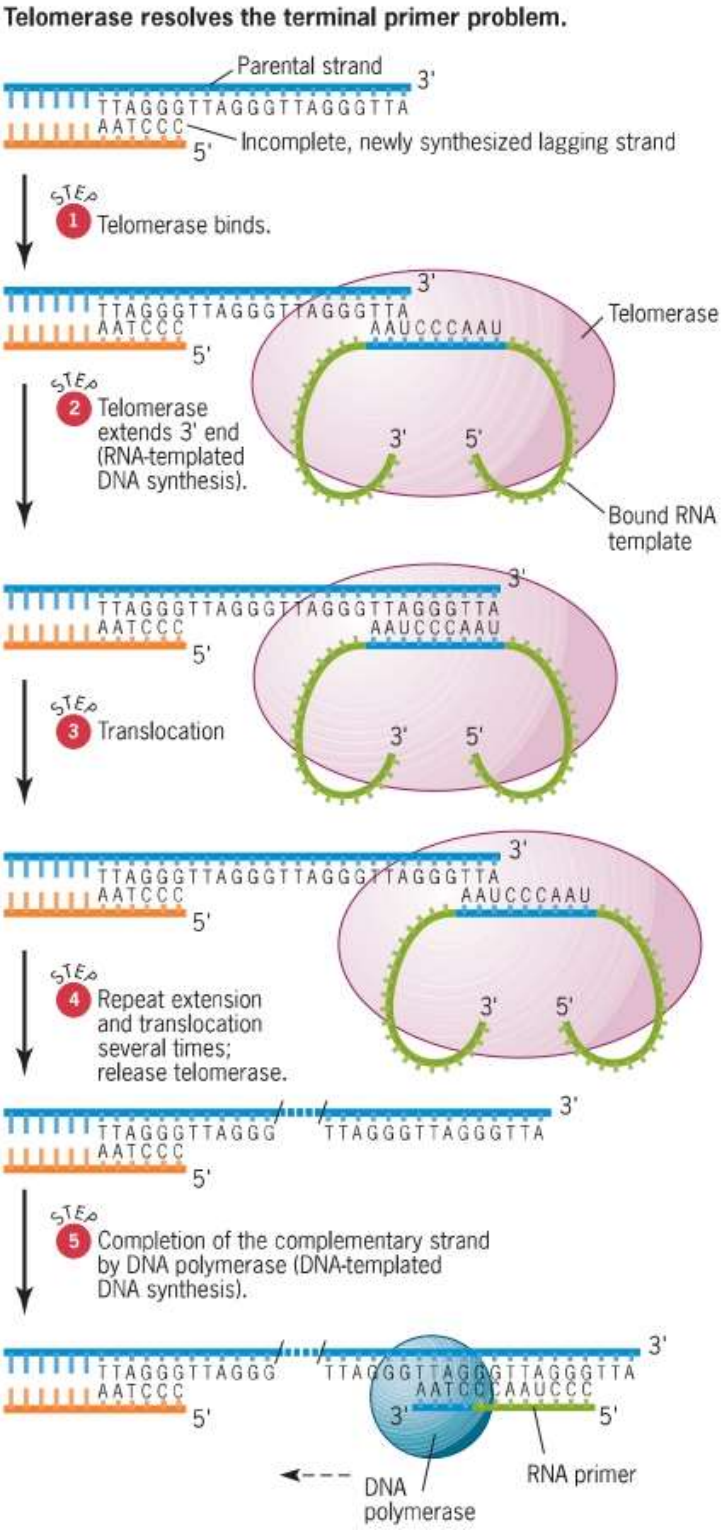


Fig.: Replication of chromosome telomeres by telomerase.

Telomere-Binding Proteins Regulate Telomerase Activity and Telomere Length:

Although extension of telomeres by telomerase could theoretically go on indefinitely, proteins bound to the double-strand regions of the telomere regulate telomere length. In *S. cerevisiae* cells, proteins bound to the telomere act as weak inhibitors of telomerase activity. When there are relatively few copies of the telomere sequence repeat, few of these proteins are bound to the telomere, and telomerase can extend the 3'-OH end of the telomere. As the telomere becomes longer, more of the telomere-binding proteins accumulate and inhibit telomerase extension of the 3'-OH end of the telomere. This simple negative-feedback loop mechanism (longer telomeres inhibit telomerase) is a robust method to maintain a similar telomere length at the ends of all chromosomes.

Proteins that recognize the single-stranded form of the telomere can also modulate telomerase activity. In *S. cerevisiae* cells, the Cdc13 protein binds to single-stranded regions of the telomere. Studies of this protein indicate that it recruits telomerase to the telomeres. Thus, Cdc13 is a positive activator of telomerase. In contrast, the human protein that binds to single stranded telomeric DNA, POT1, acts in the opposite manner—that is, as an inhibitor of telomerase activity. In vitro studies show that POT1 binding to single-stranded telomere DNA inhibits telomere activity. Cells that lack this protein show dramatically increased telomere DNA length. Interestingly this protein interacts indirectly with the double-strand telomere binding proteins in human cells. It has been proposed that as telomeres increase in length, more POT1 is recruited, thereby increasing the likelihood that it binds to the ssDNA ends of the telomere and inhibits telomerase.

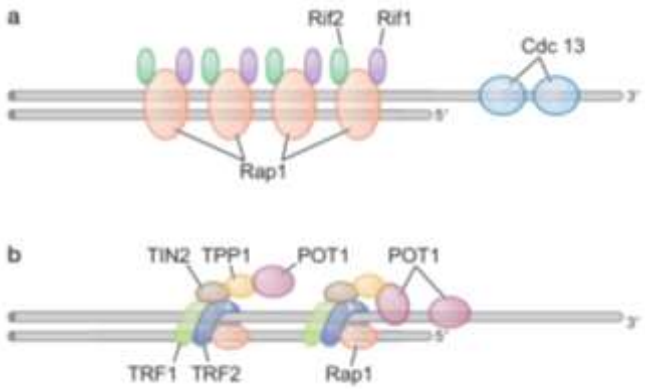


Fig: Telomere-binding proteins.

Telomere-Binding Proteins Protect Chromosome Ends:

In addition to their role in regulating telomerase function, telomere-binding proteins also play a crucial role in protecting the ends of chromosomes. Ordinarily in a cell, the presence of a DNA end is considered the sign of a double-stranded break in the DNA, which is targeted by the DNA repair machinery. The most common outcome of this repair is to initiate recombination with other DNA in the genome. (In a diploid cell this recombination is targeted to the intact copy of the broken chromosome.) Whereas this response is appropriate for random DNA breaks, it would be disastrous for the telomeres to participate in the same events. Attempts to repair telomeres in the same manner as double-stranded DNA breaks would lead to chromosome fusion events, which eventually result in random chromosome breaks.

The proteins bound at the telomere distinguish telomeres from other DNA ends in the cell. Elimination of these proteins leads to the recognition of the telomeres as normal DNA breaks. It is possible that protection is conferred simply by coating the telomere with binding proteins. Studies of the structure of the human telomere have led to an alternative possibility. Telomeres isolated from human cells were observed by electron microscopy and found to form a loop rather than a linear structure. Subsequent analysis indicated that this structure, called a **t-loop**, was formed by the 3'-ssDNA end of the telomere invading the dsDNA region of the telomere. It has been proposed that by forming a t-loop, the end of the telomere is masked and cannot be recognized as a normal DNA end. Interestingly, purified TRF2 is capable of directing t-loop formation with purified telomere DNA.

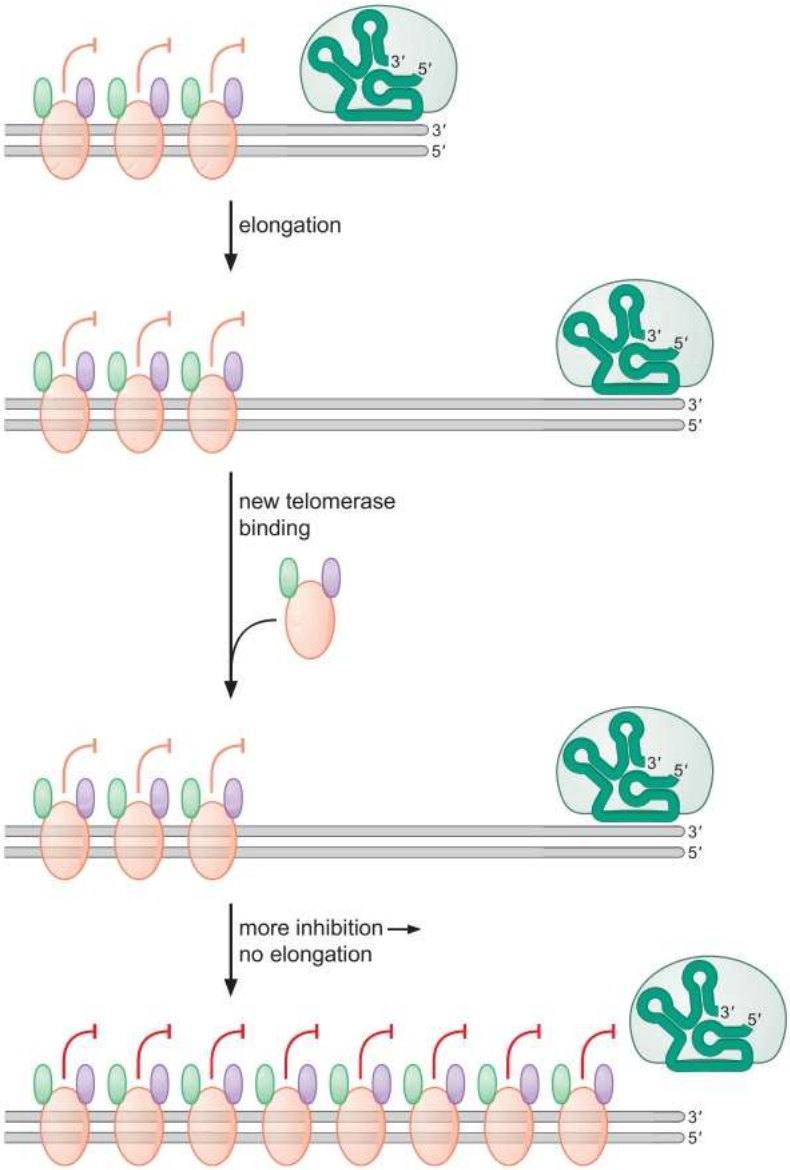


Fig.: Telomere length regulation by telomere-binding proteins.

The t-loop structure may also be relevant to telomere length control. Just as the loop structure may protect the telomere from DNA repair enzymes, it is also likely that telomerase cannot recognize this form of the telomere, because it lacks an obvious single-strand 3' end. It has been proposed that as telomeres shorten, they would have

an increasingly difficult time forming the t-loop, thereby allowing increased access to the 3' end of the telomere.

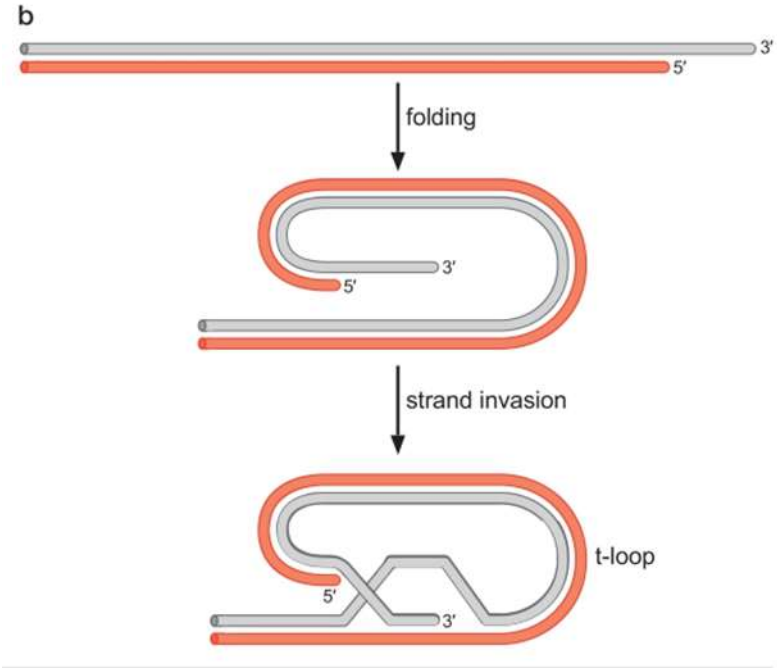


Fig.: Telomeres form a looped structure in the cell.

Telomere length and aging in humans:

Unlike germ-line cells, most human somatic cells lack, or have very low levels of, telomerase activity. When human somatic cells are grown in culture, they divide only a limited number of times (usually only 20 to 70 cell generations) before senescence and death occur. When telomere lengths are measured in various somatic cell cultures, a correlation is observed between telomere length and the number of cell divisions preceding senescence and death. Cells with longer telomeres survive longer—go through more cell divisions— than cells with shorter telomeres. As would be expected in the absence of telomerase activity, telomere length decreases as the age of the cell culture increases. Occasionally, somatic cells are observed to acquire the ability to proliferate in culture indefinitely, and these immortal cells have been shown to contain telomerase activity, unlike their progenitors. Since the one common feature of all

cancers is uncontrolled cell division or immortality, scientists have proposed that one way to combat human cancers would be to inhibit the telomerase activity in cancer cells.

Further evidence of a relationship between telomere length and aging in humans has come from studies of individuals with disorders called progerias, inherited diseases characterized by premature aging. In the most severe form of progeria, Hutchinson–Gilford syndrome, senescence—wrinkles, baldness, and other symptoms of aging—begins immediately after birth, and death usually occurs in the teens. This syndrome is caused by a dominant mutation in the gene encoding lamin A, a protein involved in the control of the shape of nuclei in cells. Why this mutation leads to premature aging is unknown. In a less severe form of progeria, Werner syndrome, senescence begins in the teenage years, with death usually occurring in the 40s. Werner syndrome is caused by a recessive mutation in the WRN gene, which encodes a protein involved in DNA repair processes. Again, we still do not understand how the loss of this protein leads to premature aging. However, the somatic cells of individuals with both forms of progeria have short telomeres and exhibit decreased proliferative capacity when grown in culture, which is consistent with the hypothesis that decreasing telomere length contributes to the aging process. At present, the relationship between telomere length and cell senescence is entirely correlative. There is no direct evidence indicating that telomere shortening causes aging. Nevertheless, the correlation is striking, and the hypothesis that telomere shortening contributes to the aging process in humans warrants further study.

6. Epigenetics: Introduction, methylation, histone modifications, epialleles.

Epigenetics refers to heritable changes in gene expression that arise from changes in chromosomes without alteration of DNA sequence. These changes occur throughout all stages of development or in response to environmental factors such as exposure to toxins or chronic stress and are implicated in diseases such as cancer. Epigenetic mechanisms of gene regulation, which collectively make up the epigenome, include modifications to DNA and histone components of nucleosomes as well as expression of noncoding RNAs (ncRNAs). These modifications can affect gene accessibility to DNA-binding and regulatory proteins such as methyl-CpG-binding proteins, transcription factors, RNA polymerase II and other components of the transcriptional machinery, ultimately altering transcription patterns, often in tissue- and cell-specific ways. A schematic diagram showing the most well characterized epigenetic modifications are shown in Figure 1.

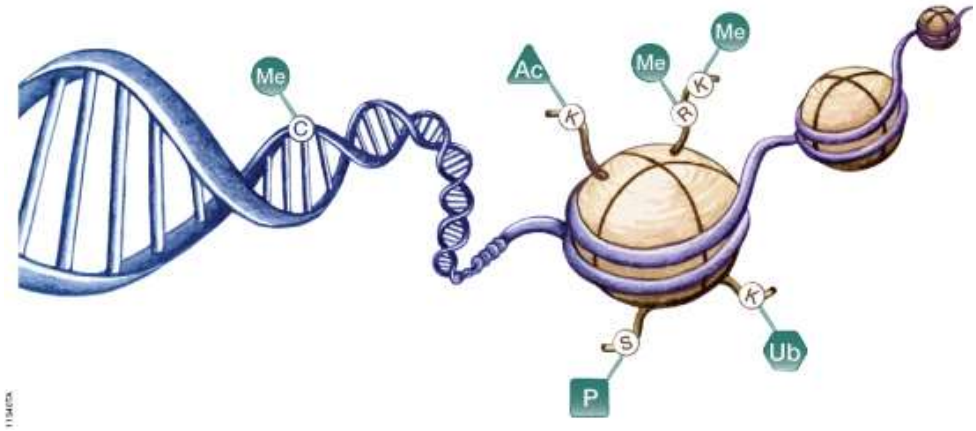


Fig.: Epigenetic mechanisms involved in regulation of gene expression. Cytosine residues within DNA can be methylated, and lysine and arginine residues of histone proteins can be modified. Me = methylation, Ac = acetylation, P = phosphorylation, Ub = ubiquitination.

DNA Methylation

In vertebrates, DNA methylation occurs on the 5C position of cytosine residues to yield 5-methylcytidine. This occurs almost exclusively within CpG dinucleotides, although nonCpG methylation does occur in plants (primarily CpNpG and CpHpH methylation, where H = A, T, C) and to a lesser extent, mammals. Other forms of cytosine exist, including 5-hydroxymethylcytosine, 5-formylcytosine and 5-carboxylcytosine, which may be intermediates in a pathway for DNA demethylation.

In mammalian genomes, approximately 70–80% of CpG dinucleotides are methylated. However, stretches of CpG-rich sequences with low levels of DNA methylation, known as CpG islands, exist. DNA methylation is typically associated with epigenetic gene repression, and many targets of de novo DNA methylation during differentiation are promoters of stem cell- and germline-specific genes. DNA methylation also recruits methyl-CpG-binding proteins, which recruit additional proteins that add silencing modifications to neighboring histones. This coordination between DNA methylation and silencing histone marks leads to compaction of chromatin and gene repression.

CpG Islands

CpG islands (CGIs) make up only 0.7% of the human genome but contain 7% of the CpG dinucleotides. CpG islands often are highly enriched at gene promoters, and approximately 60% of all mammalian gene promoters are CpG-rich. CpG islands are typically unmethylated, open regions of DNA with low nucleosome occupancy. As such, CpG islands promote relaxed chromatin structure that favors active transcription, known as euchromatin, and increases accessibility of RNA polymerase II and other components of the basal transcription machinery to the transcription start site. Most CGI promoters have heterogeneous transcription start sites and lack TATA boxes, so transcription factors with CpG in their recognition sites, such as SP1, can help recruit TATA-binding protein to promoters without TATA boxes. Without additional regulatory signals, transcription from CGI promoters results in nonproductive, bidirectional cycles of initiation and premature termination. The regulatory signals

required for the transition from this nonproductive state to productive, directional synthesis of full-length transcripts are not yet well characterized.

The mechanisms that keep CpG islands free of methylation appear to involve binding of transcription factors and other transcriptional machinery or the act of transcription itself. However, CpG islands can become hypermethylated to silence specific genes during cellular differentiation, genomic imprinting and X chromosome inactivation.

DNA Methylases

DNA methylation is catalyzed by DNA methyltransferases (DNMTs). DNMT3A and DNMT3B are involved in de novo methylation and are targeted to particular genomic regions by specific histone modifications. During DNA replication, the protein Np95 recognizes hemimethylated DNA and directs DNMT1 to the replication fork to maintain patterns of DNA methylation.

Techniques to Assess DNA Methylation

Methylation-Sensitive Restriction Enzymes

The methylation status of a DNA sequence can be determined using a variety of techniques such as the use of restriction enzymes (REs), which recognize short DNA sequences and cleave double-stranded DNA at specific sites within or adjacent to these sequences. Some REs are sensitive to methylation and will not cleave DNA if a cytosine in their recognition sites is methylated, while other REs are insensitive to methylation. The methylation-sensitive RE HpaII was used in early epigenetics studies to determine that 55–70% of all HpaII sites (5'-CCGG-3') are methylated in the mammalian genome and to identify CpG-rich, hypomethylated DNA regions [known as HpaII tiny fragments (HTFs)].

Pairs of isoschizomers where one RE is insensitive to methylation and the other is sensitive (Table 1) are often used to query methylation status. DNA fragments generated by a methylation-sensitive isoschizomer will differ in size from fragments generated by a methylation-insensitive isoschizomer. The extent of cytosine methylation can be estimated by calculating the ratio of the different DNA fragments.

Table 1. Methylation Sensitivity of Isoschizomer and Neoschizomer Pairs.

Methylated Sequence	Cleaved by	Not Cleaved by
^{m4}CCGG	MspI (C/CGG)	HpaII (C/CGG)
C^{m5}CGG	MspI (C/CGG)	HpaII (C/CGG)
C^{m4}CGG	MspI (C/CGG)	HpaII (C/CGG)
CC^{m5}CGGG	XmaI (C/CCGGG)	SmaI (CCC/GGG)
G^{m6}ATC	Sau3AI (/GATC)	MboI, NdeII (/GATC)
GAT^{m5}C	MboI, NdeII (/GATC)	Sau3AI (/GATC)
GAT^{m4}C	MboI (/GATC)	Sau3AI (/GATC)
GGTAC^{m5}C	KpnI (GGTAC/C)	Acc65I (G/GTACC)

Bisulfite Sequencing

Bisulfite sequencing refers to techniques that assess DNA methylation through bisulfite conversion, which converts unmethylated cytosine residues to uracil residues. Methylated cytosine residues remain unmodified. The target DNA is purified, alkaline- or heat-denatured, treated with sodium bisulfite, cleaned up, treated with alkaline, then cleaned up again to remove salts and other components that can inhibit downstream applications. DNA purification kits, such as the Wizard® SV Gel and PCR Clean-Up System (Cat.# A9281), are commonly used for this purpose. After bisulfite conversion and DNA cleanup, the DNA is amplified by whole genome PCR, and the amplified products are analyzed using a technique that distinguishes products derived from unmethyled DNA, which contain thymine residues, from products derived from methylated DNA, which contain cytosine residues. These techniques include pyrosequencing, methylation-specific PCR, methylation-sensitive single-strand conformation analysis, high-resolution melting analysis, methyl cytosine immunoprecipitation, bisulfite methylation profiling and MALDI-TOF mass spectrometry. For high-throughput analysis, bisulfite-treated DNA can be analyzed

using microarrays with two sets of oligonucleotide probes, one of which is complementary to cytosine-containing DNA and the other complementary to thymine-containing DNA.

Typical bisulfite conversion protocols involve long incubation times under harsh conditions, resulting in highly fragmented DNA. Promega offers the MethylEdge™ Bisulfite Conversion System (Cat.# N1301), which results in efficient DNA conversion and recovery with reduced template fragmentation using a protocol that can be completed in less than two hours, including desulphonation and cleanup. The MethylEdge™ Bisulfite Conversion System does not require an additional cleanup kit.

Luciferase-Based Sensors of DNA Methylation

The firefly luciferase reporter protein (Fluc) can be used to assess DNA methylation at the genome level or at specific DNA sequences. Researchers have developed split-luciferase biosensors composed of two fusion proteins: a DNA-binding domain fused to the N-terminal portion of Fluc, and a second DNA-binding domain fused to the Fluc C-terminus. To assess levels of global DNA methylation, both fusion proteins are constructed using the DNA-binding domain of a methyl-CpG-binding domain protein such as MBD2, which has a 100-fold preference for methylated CpG sites over unmethylated CpG sites. The fusion proteins are expressed in a cell-free expression system, and then incubated with the target DNA to allow DNA binding. If multiple methylated CpG sites exist in proximity, the N-terminal and C-terminal portions of Fluc will interact (Fig). The level of restored Fluc activity is measured using a firefly luciferase assay, such as the Steady-Glo® or Dual-Glo® Luciferase Assay System, and luminescence levels are indicative of DNA methylation levels throughout the genome. To measure site-specific DNA methylation levels, the N-terminus of Fluc is coupled to the MBD DNA-binding domain, but the C-terminus is coupled to a sequence-specific DNA-binding domain.

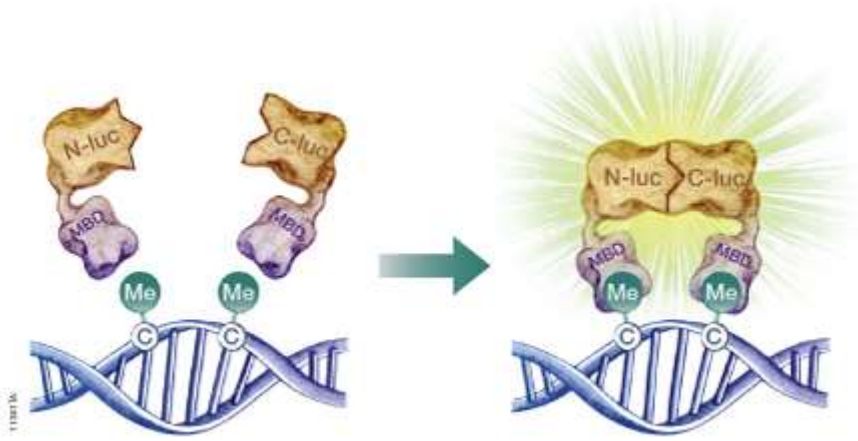


Fig.: A schematic diagram showing the use of a split-luciferase biosensor to assess DNA methylation at the genome level. To assess DNA methylation at a specific DNA sequence, one of the methyl-CpG-binding domains (MBD) is replaced with a sequence-specific DNA-binding domain. N-luc is the N-terminal portion of firefly luciferase; C-luc is the C-terminal portion of firefly luciferase.

Histone Modification and Histone Variants

Epigenetic gene regulation also is controlled by changes in histones that make up the nucleosome and histone modification. Canonical nucleosomes are octamers that consist of H2A, H2B, H3 and H4 proteins. However, there are several histone variants that can vary by a small number of amino acids or include large insertions. Often these histone variants are found at specific locations within the chromatin or are used to demarcate boundaries between heterochromatin and euchromatin regions.

The majority of histone-mediated regulation stems from histone modification, most often modification of the exposed amino termini of histones protruding from the nucleosome core. The predominant histone modifications include acetylation, methylation, phosphorylation, ubiquitination and sumoylation, with thousands of potential combinations of modifications within a single nucleosome. Of these, histone acetylation and methylation are the best understood, and some general trends have been observed. Trimethylation of histone H3, specifically the lysine at position 4

(H3K4me3), is a mark associated with transcriptionally active chromatin, whereas H3K27me3 leads to compact chromatin, which represses gene expression. The term "histone code" is used to describe how different combinations of histone modifications affect transcription levels.

Identification of proteins that read, write or erase these marks is critical to help unravel the complexities of epigenetic regulation. Chromatin immunoprecipitation (ChIP) is a powerful assay to identify proteins that bind to chromatin and map protein binding throughout the genome using techniques such as microarray analysis or high-throughput sequencing.

In ChIP analysis, protein:protein and protein:DNA complexes are crosslinked, immunoprecipitated using an antibody against the protein of interest and purified. The DNA sequence of interest then is amplified from the immunoprecipitated material using PCR. Alternatively, the immunoprecipitated DNA can be sequenced (ChIP-seq) or analyzed using microarrays (ChIP-chip) to identify target sequences.

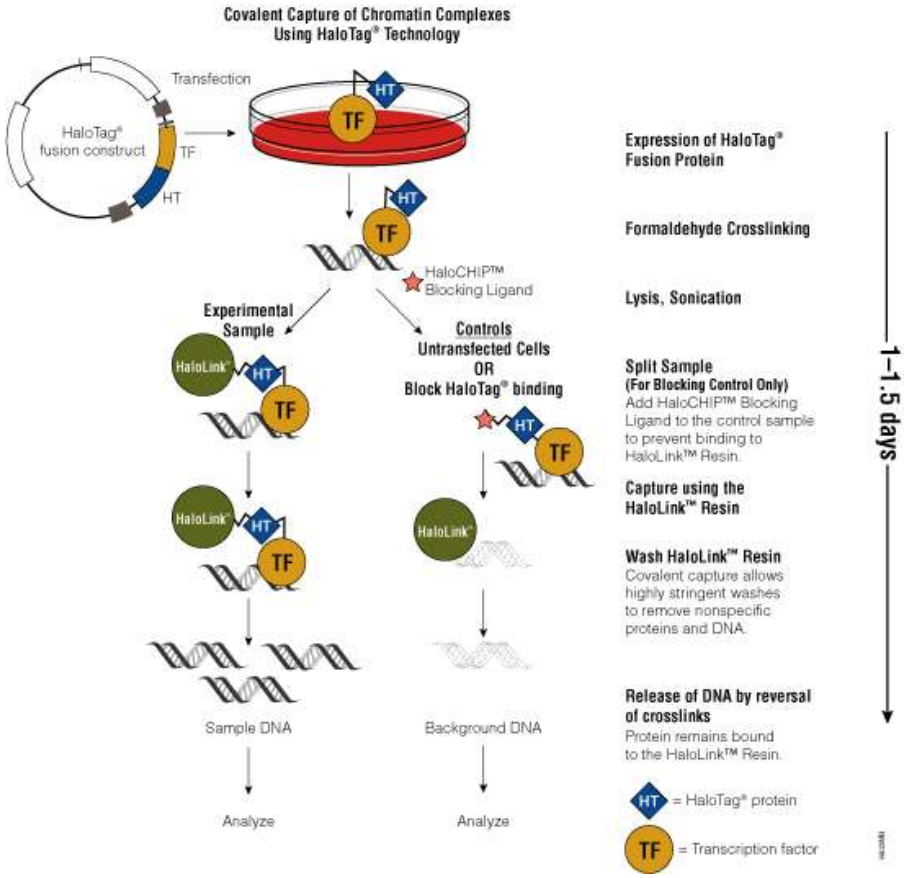


Fig.: Schematic diagram of the HaloCHIP™ System.

One challenge of the traditional ChIP method is the availability of specific antibodies that recognize crosslinked epitopes. To overcome the need for suitable antibodies, Promega scientists developed the HaloCHIP™ System (Figure 3). This system takes advantage of the HaloTag® protein, which is a mutated hydrolase (that catalyzes a covalent attachment to a variety of ligands, including a resin-based ligand for immobilization. This tag can be fused to any protein; for ChIP, the DNA-binding protein of interest is fused to the HaloTag® protein by cloning the protein-coding region into a HaloTag® vector. The recombinant construct is transfected into cells for stable or transient expression, then cells are treated with formaldehyde to induce covalent protein:DNA and protein:protein crosslinks, lysed and sonicated to shear the DNA into smaller fragments. The crosslinked complexes are captured directly from the lysate through covalent binding of the HaloTag® moiety to the HaloLink™ Resin. Covalent binding allows more extensive and stringent washing than is possible with noncovalent interactions, resulting in reduced background and increased signal-to-noise ratio. Subsequent heating of the purified complexes reverses the crosslinks and releases captured DNA fragments, which can be purified and analyzed using PCR, sequencing or microarray analysis. For more information, see the *HaloCHIP™ System Technical Manual* #TM075.

Histone Acetylation and Deacetylation

Acetylation of a lysine residue neutralizes a positive charge on a histone protein, reducing the electrostatic interaction with negatively charged DNA. This reduction in affinity leads to increased accessibility of the DNA to protein complexes, which can lead to increased gene expression. In addition, lysine acetylation can recruit nucleosome-remodeling complexes, such as Swi2/Snf2, via their bromodomains to promote and maintain euchromatin structure. However, the factors controlling gene expression are complex, and histone acetylation also can lead to reduced gene expression through indirect mechanisms.

Lysine acetylation occurs on the N-terminal tails of core histones and is controlled primarily by two enzyme families: histone acetyl transferases (HATs) and histone deactylases (HDACs). HATs use acetyl CoA as a coenzyme to transfer an acetyl group to

the epsilon amino group of the lysine side chain. These enzymes are grouped into three families: GNAT, p300/CBP and MYST. HDACs reverse histone acetylation and promote gene silencing. HDACs are often components of large protein complexes and are recruited to sites of DNA methylation by methyl DNA-binding proteins. HDACs fall into four categories: Class I, which includes HDAC1, 2, 3 and 8; Class, II, which includes HDAC4, 5, 6, 7, 9 and 10; Class III, which includes the NAD+-dependent sirtuins (SIRTs); and Class IV, which includes HDAC11.

Misregulation of HATs and HDACs often is associated with development and progression of cancer and other diseases such as neurodegenerative disorders and cardiovascular diseases, making these enzymes attractive therapeutic drug targets. Many HDAC inhibitors promote cell cycle arrest at the G1/S phase, and studies have shown that tumor cells generally are more sensitive to HDAC inhibitors than normal cells. Also, HDAC inhibitors can restore the ability of animals to recall memory that had been lost in Alzheimer's and Parkinson's disease models, possibly by changing chromatin structure in neurons.

To facilitate screening of potential HDAC inhibitors, Promega offers the HDAC-Glo™ I/II Assays and Screening Systems and SIRT-Glo™ Assays Systems. The HDAC-Glo™ I/II and SIRT-Glo™ Assays are single-reagent-addition, homogeneous, luminescent assays that measure relative activities of HDAC class I and II enzymes and sirtuins, respectively. The HDAC-Glo™ I/II Assays use an acetylated, live-cell-permeant, luminogenic peptide substrate that is deacetylated by HDAC activities from cells, extracts or purified enzyme sources. The SIRT-Glo™ Assay uses a similar substrate to detect SIRT activities from purified enzyme sources. Deacetylation of the peptide substrate is measured using a coupled enzymatic system in which a protease in the Developer Reagent cleaves the deacetylated peptide from aminoluciferin, which is quantified in a luciferase-based reaction. The HDAC-mediated luminescent signal is proportional to enzyme activity and persistent, allowing batch processing of multiwell plates in high-throughput screening.

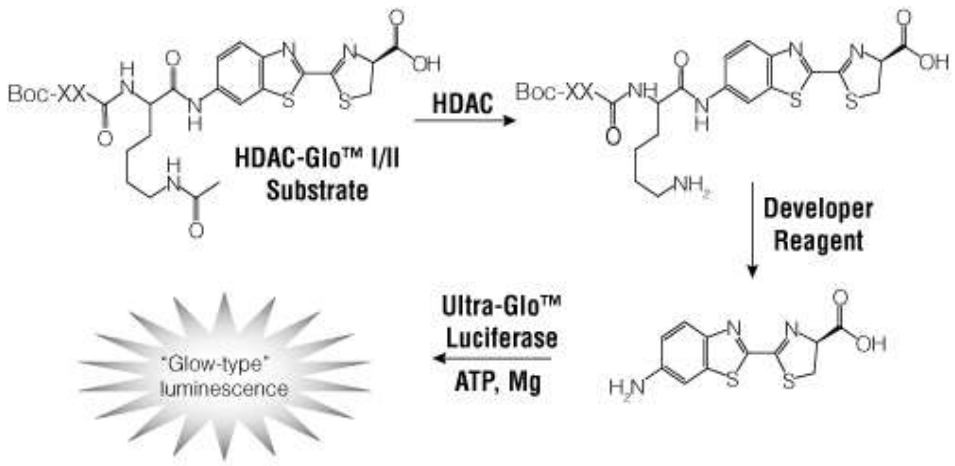


Fig.: The single-reagent-addition HDAC-Glo™ I/II Assay. HDAC activity deacetylates the luminogenic HDAC-Glo™ I/II Substrate, making the peptide sensitive to specific proteolytic cleavage that is mediated by the HDAC-Glo™ I/II Developer Reagent and liberates aminoluciferin. Free aminoluciferin then can be measured using Ultra-Glo™ firefly luciferase to produce stable, persistent light emission. Boc represents an amino-terminal blocking group that protects the substrate from nonspecific cleavage. XXLysine is an HDAC I/II-optimized consensus sequence derived from histone 4.

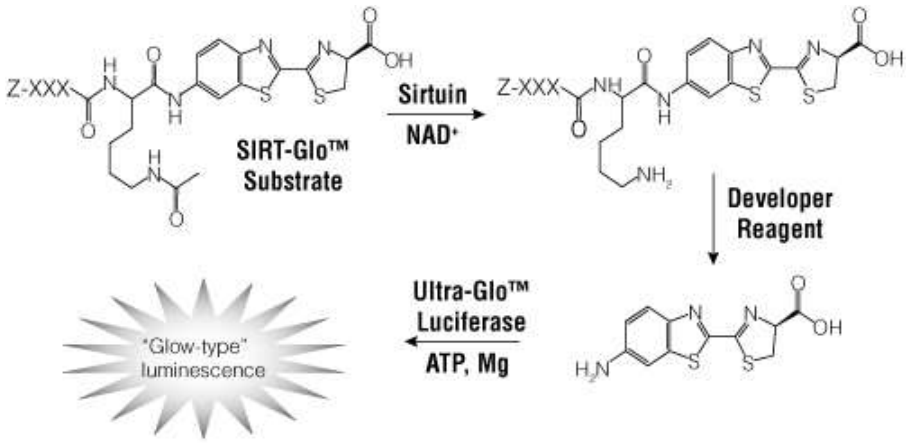


Fig.: The single-reagent-addition SIRT-Glo™ Assay. SIRT activity deacetylates the luminogenic SIRT-Glo™ Substrate, making the peptide sensitive to specific proteolytic cleavage that is mediated by the SIRT-Glo™ Developer Reagent and liberates aminoluciferin. Free aminoluciferin then can be measured using Ultra-Glo™ firefly luciferase to produce stable, persistent light emission. Z represents an amino-terminal blocking group that protects the substrate from nonspecific cleavage. XXXLysine is a SIRT-optimized amino acid sequence based on a consensus sequence derived from p53.

Histone Methylation

Histone methylation occurs at lysine residues, which can be mono-, di- or trimethylated, and arginine residues, which can be mono- or dimethylated. Histone methylation is catalyzed by protein lysine methyltransferases (PKMTs) and protein arginine methyltransferases (PRMTs) but can be reversed by protein demethylases. To date, researchers have identified >30 demethylating enzymes, >50 protein lysine methyltransferases and >10 protein arginine methyltransferases, suggesting that protein methylation is a dynamic and complex process. Histone methylation has different effects on transcriptional activity, depending on the number of methyl groups and position of the amino acid being modified. In general, the H3K9me1 mark is activating, whereas H3K9me2 and H3K9me3 are repressive; H3K4me3 and H3K36me3 are associated with active chromatin, whereas H3K9me3, H3K27me3, H3K36me2 and H4K20me1 often are found in transcriptionally repressed heterochromatin.

The downstream effects of histone methylation are largely determined by proteins that bind to the modified histones. For example, H3K9me3 acts as a binding site for heterochromatin protein 1 (HP1), which then can recruit histone methyltransferases, histone deacetylases and other proteins that affect chromatin structure. H3K4me3 recruits proteins that promote euchromatin, whereas H3K9me1, H3K9me2 and H3K27me3 interact with proteins that promote heterochromatin. Two such groups of proteins are the polycomb group (PcG) proteins and their antagonists, the trithorax (trxG) group proteins, which were first identified as regulators of *hox* gene expression in *Drosophila*. More recent studies have shown that related proteins exist in mammals and plants. PcG proteins repress transcription; trxG proteins activate transcription. Some PcG and trxG proteins possess histone methyltransferase activity and can modify histones directly, while others bind to and interpret histone modifications.

In embryonic stem (ES) cells, CpG islands that are regulated by PcG proteins often are "bivalent" in that they retain the permissive H3K36me2-depleted and H3K4me3-enriched environment but also exhibit H3K27me3. Genes with bivalent promoters often are actively silenced in ES cells but lose the repressive H3K27me3 mark while retaining the activating H3K4me3 mark later during differentiation.

Histone Phosphorylation

Histones can be phosphorylated on serine, threonine and tyrosine residues. Many of the serine and threonine phosphorylation events play a role in DNA repair or DNA condensation, segregation and decondensation during mitosis, but some are involved in epigenetic regulation of transcription, including H3T3ph, H3T6ph, H3T11ph, H2.AS1ph, H3S10ph and H4S41ph. H3S10ph is one of the best characterized of these histone modifications. In addition to its DNA-restructuring responsibilities during mitosis, H3S10ph seems important for chromatin decondensation associated with transcriptional activation of target genes. H3S10ph recruits chromatin-modifying enzymes and chromatin-remodeling complexes and prevents binding of HP1 to neighboring H3K9me3 marks at the onset of mitosis. H3S10ph, as well as H3T3ph and H3T11ph, can block binding of DNMT3a to H3, reducing methylation of nearby chromatin.

Several kinases are involved in phosphorylation of H3S10, including IkB kinase a (IKKa), proviral integration site for Moloney murine leukemia virus 1 (PIM1) (Zippo *et al*. 2007) and ribosomal S6 kinase 2 (RSK2). Addition of the H3S10ph mark to H3K9me3 is catalyzed by Aurora B kinase, which also modulates chromosome structure during mitosis and mediates chromosome alignment and attachment to microtubules of the mitotic spindle.

Histones contain many highly conserved tyrosine residues, many of which can be phosphorylated. Phosphorylation of H3Y99 is critical for polyubiquitination and subsequent proteolysis of excess histones, which can increase a cell's sensitivity to DNA-damaging agents, cause genomic instability and induce apoptosis. Another tyrosine residue, H3Y41, is important in chromatin structure and oncogenesis. In human hematopoietic cell lines, phosphorylation of H3Y41 by Janus kinase 2 (JAK2) destabilizes binding of HP1a to histone H3, leading to a more open chromatin structure around certain gene promoters such as leukemia oncogene *LMO2*, which can trigger oncogenesis in hematopoietic cells. Overexpression or aberrant activation of JAK2 activity leads to higher levels of H3Y41, loss of HP1a binding and higher expression of *LMO2*.

Promega offers a number of kinase enzyme systems to monitor the activity or identify inhibitors of different kinases involved in histone phosphorylation, including IKKa, PIM1, RSK2 and several cyclin-dependent kinases (CDKs) such as CDK1, CDK2 and CDK5. These luminescent assays convert ADP produced by these kinases to ATP, which is then converted to light by Ultra-Glo™ Luciferase. The resulting luminescent signal positively correlates with ADP amount and kinase activity.

Histone Ubiquitination

Conjugation of ubiquitin, a 76-amino acid protein, to lysine residues of histone proteins can affect transcription activity as well as nucleosome stability and, as a result, gene accessibility. The consequences of histone ubiquitination depend on the histone substrate and degree of ubiquitination (reviewed by Weake and Workman, 2008). Mono-ubiquitination of histone H2A (H2Aub1) is often considered a repressive mark, while H2B mono-ubiquitination can play a role in both transcriptional activation and silencing. In addition, there is evidence of cross-talk between histone ubiquitination and other forms of histone modification. For example, ubiquitinated H2B has been identified as a docking site for the COMPASS protein complex, which includes the histone methyltransferase responsible for H3K4 methylation. Also, H2Aub, but not H2A, specifically represses di- and trimethylation of H3K4, and ubiquitin-specific protease 21 (USP21) relieves this repression.

Ubiquitination of histones can be reversed by cleaving the peptide bond between ubiquitin and the ubiquitinated protein. Several deubiquitinases (DUBs) have been reported to deubiquitylate histones 2A, 2A.Z and 2B, including USP3, USP10, USP21, USP22 and Bap1. Histone deubiquitination has been associated with both transcription activation and repression.

Sumoylation as a Mechanism of Epigenetic Regulation

Another post-translational modification that plays an important role in epigenetic regulation is sumoylation, the addition of the small ubiquitin-related modifier SUMO. This modification can stabilize proteins, alter subcellular localization, affect enzyme activity and mediate interactions with other proteins. Many transcription factors and

cofactors can be sumoylated, which is generally indicative of transcription repression. In *Drosophila*, the sumoylated form of Sp3 recruits the polycomb protein Sfmbt and HP1α, β and γ to repress transcription.

Many histone-modifying enzymes, nucleosome-remodeling complexes and their associated enzyme cofactors contain one or more SUMO interaction motifs (SIMs). This motif allows these proteins to interact with sumoylated transcription factors and cofactors, which can direct these enzymes to specific promoters. Two such groups of histone-modifying enzymes recruited by SUMO are histone deacetylases, which decrease histone acetylation at the target promoter, and histone demethylases such as lysine-specific demethylase 1, which catalyzes the removal of methyl groups from H3K4.

Epigenetic Inheritance

Maintenance and inheritance of epigenetic marks during cell division is critical to maintain a committed cell lineage and cellular phenotype in progeny cells, and set a memory of transcriptional status. The transmission of epigenetic information through multiple cell divisions involves many of the mechanisms discussed in this chapter: DNA methylation, histone modification, histone variants and expression of noncoding RNAs. These same mechanisms govern the inheritance of epimutations, which can lead to changes in chromatin structure and transcription levels of genes important to diseases such as cancer and imprinting disorders.

Epigenetics and Disease

Aberrant regulation of epigenetic mechanisms can result in genomic imprinting disorders, such as Angelman syndrome and Prader-Willi syndrome, and may contribute to the heritability of many forms of cancer, asthma, Alzheimer's disease and autoimmune diseases such as systemic lupus erythematosus, rheumatoid arthritis and multiple sclerosis. Epimutations can interfere with epigenetic regulation at many levels, including DNA methylation, histone modification and noncoding RNAs. Some epimutations are inherited, but many accumulate due to environmental factors or with age. For example, even though monozygotic twins are epigenetically indistinguishable

at birth, their patterns of DNA methylation and histone acetylation can differ dramatically as they age.

Epialleles:

Metastable epialleles are alleles that are variably expressed in genetically identical individuals due to epigenetic modifications established during early development and are thought to be particularly vulnerable to environmental influences.

These naturally occurring epialleles are often stably inherited and independent of genetic variation and may be valuable material for altering agronomic traits for crop improvement.

7. RNA Biology: Gene silencing through antisense RNA technology and Ribozymes; RNA interference (RNAi) by small regulatory RNAs: different types of small non-coding RNAs, their biogenesis and functions in posttranscriptional gene silencing; applications of RNAi in crop quality improvement.

Antisense RNA Technology:

Antisense RNA (asRNA), also referred to as antisense transcript, natural antisense transcript (NAT) or antisense oligonucleotide, is a single stranded RNA that is complementary to a protein coding messenger RNA (mRNA) with which it hybridizes, and thereby blocks its translation into protein. asRNAs (which occur naturally) have been found in both prokaryotes and eukaryotes, and belong to a subtype of long noncoding RNA (lncRNA) that is larger than 200 nucleotides. The primary function of asRNA is regulating gene expression. asRNAs may also be produced synthetically and

have found wide spread use as research tools for gene knockdown. They may also have therapeutic applications.

In cases where a gene has been identified and assigned a particular phenotype, additional approaches are often required to exactly probe the function of gene.

Such hurdles in gene identification and manipulation can be overcome by antisense RNA technology.

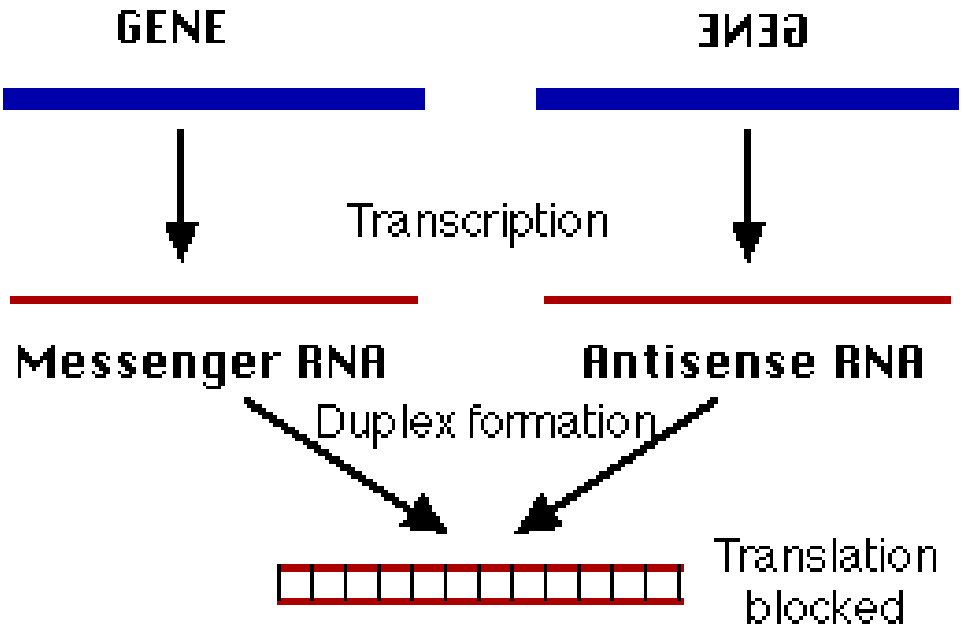
Many genomic loci contain transcription units on both strands, therefore two oppositely oriented transcripts can overlap. While one strand codes for a protein, the transcript from the other strand is non-encoding. Such natural antisense transcripts (NATs) can negatively regulate the conjugated sense transcript.

NATs are highly prevalent in a wide range of species—for example, around 15% of human protein-encoding genes have an associated NAT. NATs can be divided into cis-NATs, which are transcribed from opposing DNA strands at the same genomic locus. trans-NATs are transcribed from separate loci. cis-NAT pairs display perfect sequence complementarity, whereas trans- NAT pairs display imperfect complementarity and can target many sense targets to form complex regulation network.

One ingenious and promising approach exploits the specificity of hybridization reactions between two complementary nucleic acid chains. Normally, only one of the two DNA strands in a given portion of double helix is transcribed into RNA and it is always the same strand for a given gene.

If a cloned gene is engineered so that the opposite DNA strand is transcribed instead, it will produce antisense RNA molecules that have a sequence complementary to the normal RNA transcripts. Antisense RNA, when synthesized in large enough amounts, will often hybridize with the “sense” RNA made by the normal genes and thereby inhibit the synthesis of the corresponding protein .

A related method is to synthesize short antisense nucleic acid molecules by chemical or enzymatic means and then inject (or otherwise deliver) them into cells, again blocking (though only temporarily) production of the corresponding protein.



Formation of RNA/RNA strand

The first effort that definitely demonstrated the blockage in translation due to the use of antisense RNA in cell-free extracts (CFEs) was carried out by Singer et al. (1963) They showed that synthesis of polyphenylalanine in CFE with polyuridylic acid as template was completely inhibited when polyadenylic acid was added to the translation mixture. This write-up briefly describes the basic and applied aspects of antisense RNA technology, particularly in plant systems.

Natural Antisense RNA Regulation of Gene Expression:

Naturally occurring antisense RNA was involved in gene regulation and this was demonstrated during the study of replication of *E. coli* ColE1 plasmid. The replication of E. coli plasmid ColE1 involves formation of a RNA primer which is processed by RNase-H while bound to the DNA template. Antisense RNA binds the primer inhibiting the processing of RNA primer and replication of the plasmid, hence the plasmid copy number may be regulated (Tomizawa et al. 1981). Likewise, Staphylococcus aureus plasmid (pT 181) replication and copy number appear to be controlled by antisense RNA (Kumar and Novick, 1985). Translation of *E. coli Tn10* transposase mRNA is inhibited by antisense mRNA.

Ribozymes:

Ribozymes are catalytic RNA enzymes that act to alter covalent structure in other classes of RNAs and certain molecules. They occur in ribosomes, nucleus and chloroplasts of eukaryotic organisms. Some viruses including several bacteriophages also have ribozymes. An optimum concentration of metal ions such as Mg^{+2} and K^{+2} is associated with their effective functioning. Ribozymes generally act as molecular scissors cutting precursor RNA molecules at specific sites. Surprisingly, they also serve as molecular staplers, which ligate or join two RNA molecules together. Ribozymes are involved in the transformation of large precursor molecules of tRNA, rRNA and mRNA into smaller final products. In their active form, ribozymes are complexed with protein molecules, e.g., the enzyme ribonuclease-P (RNAse-P) is found in all living cells.

Biogenesis and functions of small RNAs in posttranscriptional gene silencing

RNA interference

RNA interference (RNAi) is a mechanism where the presence of certain fragments of dsRNA interferes with the expression of a particular gene which shares a homologous sequence with this dsRNA.

Before RNAi was well characterized, it was called by other names, including:

- Post Transcriptional Gene Silencing (PTGS)
- Co-supression
- Virus Induced Gene Silencing (VIGS) In Virus
- Quelling In Fungi
- Transgene Silencing

RNAi has become a powerful and more reliable technique to inhibit the expression of targeted genes in a precise manner and also determine gene loss-of-function phenotype, when no mutant alleles are unavailable.

- ❖ RNAi pathway is well conserved across eukaryotes.

- The use of RNAi to reduce expression in plants has been a common procedure for many years. Single-stranded antisense RNA was introduced into plant cells that hybridized to the cognate, single-stranded, sense messenger RNA. While scientists first believed that the resulting dsRNA helix could not be translated into a protein, it is now clear that the dsRNA triggered the RNAi response. The use of dsRNA became more widespread after the discovery of the RNAi machinery, first in petunia and later in round worms (*Caenorhabditis elegans*).

Discovery of small RNAs

The first small RNA: **lin-4**

- In 1993 Rosalind Lee was studying a non-coding gene in *Caenorhabditis elegans*, **lin-4**, that was involved in silencing of another gene, **lin-14**, at the appropriate time in the development of the worm *C. elegans*.
- Two small transcripts of **lin-4** (22 nt and 61 nt) were found to be complementary to a sequence in the 3' UTR of lin-14.
- Because lin-4 encoded no protein, she deduced that it must be these transcripts that are causing the silencing by RNA-RNA interactions.
- The second small RNA wasn't discovered until 2000!

What are small ncRNAs?

- Two flavors of small non-coding RNA:
    1. micro RNA (miRNA)
    2. short interfering RNA (siRNA)
- Properties of small non-coding RNA:
    1. Involved in silencing other mRNA transcripts.
    2. Called “small” because they are usually only about 21-24 nucleotides long.
    3. Synthesized by first cutting up longer precursor sequences (like the 61 nt one that Lee discovered).
    4. Silence an mRNA by base pairing with some sequence on the mRNA.

Salient features of RNA interference machinery

- The RNA interference machinery processed long dsRNA into small dsRNAs (19–24 nt) using an endonuclease (Dicer), separates the two strands, and then proceeds to destroy other single-stranded RNA molecules that are complementary to one of those sequences.
- The dsRNAs direct the creation of small interfering RNAs (siRNAs) which target RNA-degrading enzymes (RNAses) to destroy transcripts complementary to the siRNAs.
- The life cycle and replication of many RNA viruses involves a double-stranded RNA stage, so it is likely that part of the RNA interference machinery evolved as a defense against these viruses.
- The machinery is however also used by the cell itself to regulate gene activity: certain parts of the genome are transcribed into microRNA, short RNA molecules that fold back on themselves in a hairpin shape to create a double strand. When the RNA interference machinery detects these double strands, it will also destroy all mRNAs that match the microRNA, thus preventing their translation and lowering the activity of many other genes.

Small Regulatory RNAs and their role in gene silencing:

Small Regulatory RNAs in eukaryotes fall into two main groups-

1. miRNAs (microRNAs)
2. siRNAs (short interfering RNAs)

Both of these RNAs are non-coding, i.e., they are untranslated and therefore do not specify a polypeptide product.

1. miRNAs:

- MicroRNAs are ssRNA regulatory molecules about 21-23 nucleotide (nt) long that derivefrom RNA transcripts.
- These miRNAs are coded by genes in the genome of all multicellular eukaryotes, as well as some unicellular ones (budding yeast- *Saccharomyces cerevisiae*).

- In humans, several hundred miRNA genes are scattered throughout all the chromosomesexcept of the Y chromosome.
- ~30% of mammalian miRNA genes are located in intergenic regions i.e. between protein-coding genes of the genome.
- They are transcribed by RNA polymerase II resulting in capped, polyadenylated transcripts.
- Some miRNA genes are located in transposons.
- Rests are situated within other genes- many are in introns of protein-coding genes, while some are in introns and exons of non-protein coding genes.
- In all cases, the miRNA sequence is transcribed by an RNA polymerase as part of thetranscript of the host gene.
- In few cases, an intron-located miRNA gene is transcribed independently by RNA polymerase II.

2. MicroRNAs (miRNAs) mediated gene silencing:

- The transcript containing a miRNA is called the primary miRNA transcript or pri-miRNA.
- The pri-miRNA molecule contains a hairpin structure about 70 nt long, within which is theeventual miRNA.
- The hairpin is cut out of the pri-miRNA in the nucleus by the dsRNA-specific endonuclease Drosha complexed to an accessory protein (*Pasha* in *Drosophila*).
- Drosha makes staggered cuts resulting in a ~2 nt 3' single-stranded overhang.
- The excised hairpin- pre-miRNA is exported rapidly to the cytoplasm.
- In the cytoplasm, another dsRNA-specific endonuclease, Dicer, complexed to an accessory protein (*Loq* in *Drosophila*), makes staggered cuts in the pre-miRNA, releasing a short miRNA:miRNA* dsRNA consisting of some the former paired sides of the hairpin.
- The two RNA strands are imperfectly paired: "miRNA" is the mature miRNA

strand that subsequently functions in the cell for RNA silencing, while miRNA* is its partial complement and does not function in RNA silencing.

- Because the miRNA directs RNA silencing, it is termed the guide strand, while the miRNA* is termed the passenger strand.
- Next the dsRNA, Dicer and accessory protein bind to Ago1, a member of the Argonaute family of protein and other proteins to form the pre-microRNA-induced silencing complex, orpre-miRISC.
- Ago1 is another RNA endonuclease; more generally called Slicer. It makes a single cut withinthe miRNA* passenger strand.
- A helicase that is part of the pre-miRISC then unwinds to two pieces from the miRNA guidestrand, and they dissociate from the complex.
- The result is the mature miRISC, the ribonucleoprotein complex that can silence geneexpression.

How does a miRISC function in Post-transcriptional gene silencing?

(i) The miRNA in the miRISC is a trans-acting RNA regulatory molecule, meaning that it targetsmRNAs that are not the same as the RNA molecules from which the miRNA is derived. This is one distinguishing feature of miRNAs compared with siRNAs.

(ii) A miRISC binds to a target mRNA through complementary base pairing involving the miRNA.

(iii) Usually, the sequence to which the miRNA binds are short sequence in the 3' UTR of themRNA.

(iv) An mRNA molecule may have one or more sequences in its 3' UTR to which the same miRNA can bind and/or it may have several sequences in its 3' UTR to which several different miRNAs can bind. The latter raises the possibility of regulating the expression of the same gene (through its mRNA) by various combinations of miRNA regulator molecules.

(v) Here, one miRISC is shown binding to a 3' UTR sequence for simplicity.

(vi) Binding of most of the miRISC to their target mRNAs involves imperfect pairing

between themiRNA and the 3’ UTR region of the mRNA.

(vii) Such pairing triggers translational repression – translation of that mRNA becomes inhibited.

(viii) The translationally repressed mRNA with its associated miRISC(s) is then sequestered from the translation machinery by becoming or moving into a P body.

(ix) P body is a cytoplasmically located aggregate of translationally repressed mRNAs complexed with proteins, and proteins for mRNA decapping and mRNA degradation.

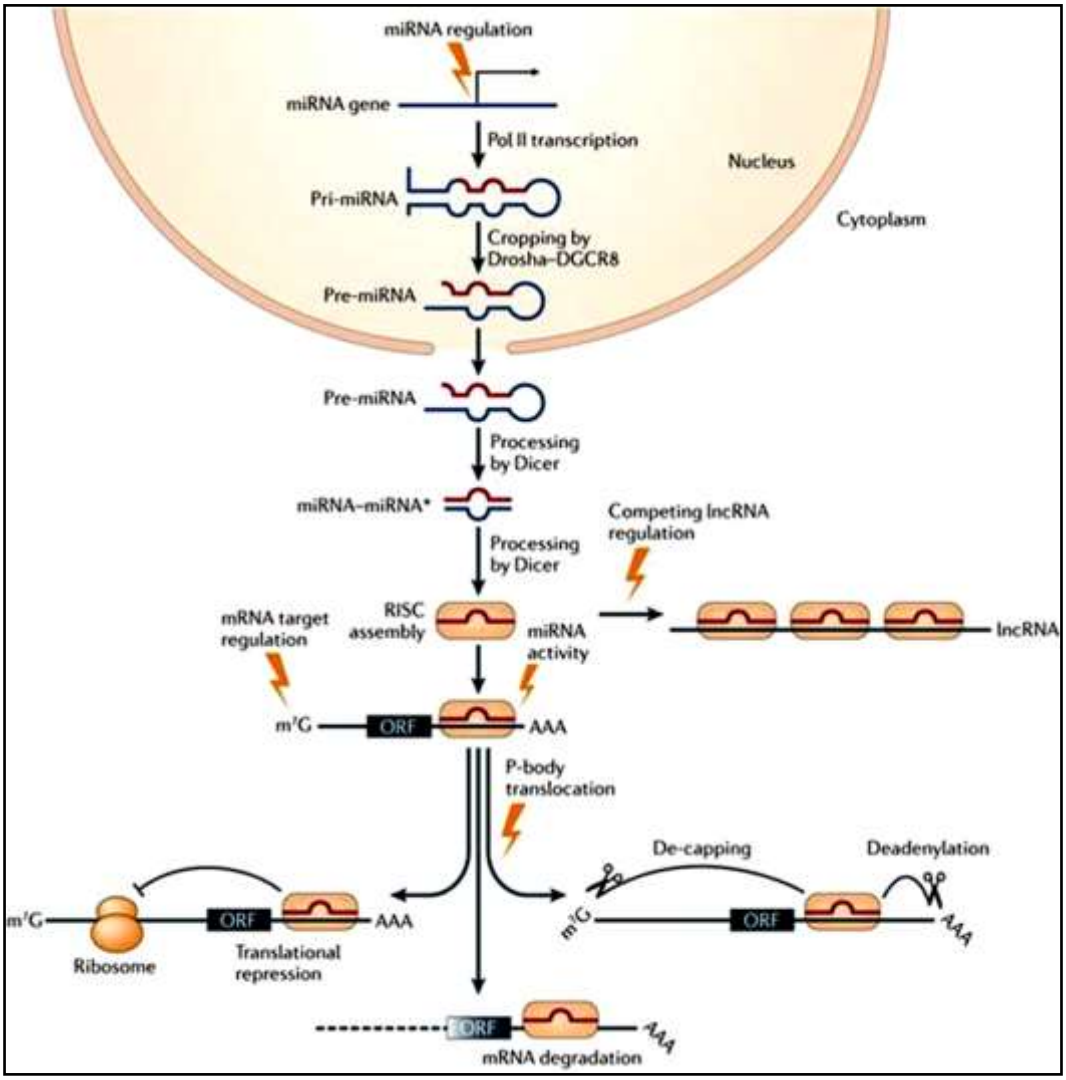
(x) The mRNAs is P bodies may be degraded using the contained mRNA degradation machinery or stored in ribonucleoprotein complexes.

(xi) Stored mRNAs can be returned to translation at a later time. Whether degraded or stored the effect of miRNA action is to reduce the expression of the gene encoding the targeted mRNAat the translational level.

(xii) In plants, binding of most miRISCs to their target mRNAs involves perfect or near-perfect pairing between much of the miRNA and the 3’ UTR region of the mRNA.

(xiii) Perfect pairing triggers mRNA degradation rather than translational repression.

(xiv) Here, the Ago1 Slicer protein cuts the target mRNA into two and the mRNA-miRISC complex forms, or it is moved to a P body where degradation of the mRNA is completed.



Biogenesis and mode of action of miRNAs in gene silencing

siRNAs (Short interfering RNAs/ Small interfering RNA)

- Small interfering RNA (siRNA), sometimes known as short interfering RNA, are a class of **20-25 nucleotide-long RNA molecules** that interfere with the expression of genes. They are naturally produced as part of the RNA interference (RNAi) pathway by the enzyme **Dicer**. They can also be exogenously (artificially) introduced by investigators to bring about the **knockdown of a particular gene**.

- siRNA's have a well defined structure. Briefly, this is a short (usually 21-nt) double-strand of RNA (dsRNA) with 2-nt overhangs on either end, including a 5' phosphate group and a 3' hydroxy (-OH) group.
- siRNAs are produced from the cytoplasmically located dsRNA molecules that are hundreds to thousands of base pair long.
- Source of these long dsRNA molecules include intermediates in the replication of viruses with RNA genome, naturally generated molecules from complementary or partially complementary sense and antisense transcripts from regions of the genome, and transcripts that fold into long, extended hairpins.
- The long dsRNA is processed using a pathway highly similar to that for processing pre-miRNA in the cytoplasm.
- First, the molecule is processed by a Dicer-protein complex into many ~22 nt duplexes, each with 2 nt 3' overhangs.
- One strand of each duplex is the siRNA guide strand that will carry out RNA silencing, while the complementary strand is the passenger strand that will be discarded.
- Then the dsRNA-Dicer-protein complex binds to Ago2 (RNA endonuclease, another member of the Argonaute family), and other proteins to form the pre-si-RNA-induced silencing complex (pre-siRISC).
- Ago2 cleaves one of the two RNA strands, a helicase unwinds the two pieces, and they dissociate from the complex.

Applications of RNAi in crop quality improvement

- Improvement in the crop quality has been done through conventional breeding, but this approach is time consuming and labor intensive.
- RNAi, being a novel approach has great potential to modify the gene expression in plants for better quality traits and nutritional improvement in different crops.

- RNAi enables repression of gibberellic acid and auxin signal pathways after a reduction in the level of *SlARF7 transcript responsible* for pollination and fertilization in tomato plants. These results by-pass the auxin signaling-fertilization pathway that leads to the development of parthenocarpic fruits having great commercial value.

- Carotenoid's production such as β-carotene and lutein were reported higher in potato through gene silencing of β-carotene hydroxylase.

- The post-harvest life can enhance by knowing-out genes responsible for ethylene production in tomato. This was achieved through introducing dsRNA and blocking the gene expression of ACC-oxidase which significantly reduced the ethylene formation and enhanced shelf-life in tomato.

- RNAi suppression of α-mannosidase and β-acetylhexosaminidase associated with fruit softening also increased the shelf-life in tomato fruits .

- Increase in amylose contents in wheat by suppressing two genes (*SBEIIa and SBEIIb) meant for* starch-branching enzyme was well demonstrated by authors.

- RNAi could be exploited as a metabolic engineering tool for the production and synthesis of commercially valuable plant products such as alkaloid production (codeine, quinine, vincristine, scopolamine), biosynthesis of essential oil and flavoring agents (vanillin).

8. Genomes and Genomics: Concept of genome; Genome sequencing strategies, Genomes of Yeast, *Arabidopsis* and rice, Genome annotation, Genome duplication, Approaches to analyse differential gene expression- ESTs, Microarrays and their applications, Reverse genetics-Gene tagging, Gene trapping, Gene silencing and Gene knockout; Metagenomics.

Concept of genome:

Genome refers to the basic set of chromosomes. In a genome, each type of chromosome is represented only once. Now **genomics** is being developed as a sub discipline of genetics which is devoted to the mapping, sequencing and functional analysis of genomes. The term genomics was first used by **Thomas Roderick in 1986.** It refers to the study of structure and function of entire genome of a living organism.

i. It is a computer aided study of structure and function of entire genome of an organism.

ii. It deals with mapping and sequencing of genes on the chromosomes.

iii. It is a rapid and accurate method of gene mapping. It is more accurate than recombination mapping and deletion mapping techniques.

iv. The genomic techniques are highly powerful, efficient and effective in solving complex genetic problems.

v. Now use of genomic techniques has become indispensible in plant breeding and genetics.

Types of Genomics:

i. Structural Genomics:

It deals with the study of the structure of entire genome of a living organism. In other words, it deals with the study of the genetic structure of each chromosome of the genome. It determines the size of the genome of a species in mega-bases [Mb] and also the genes present in the entire genome of a species.

ii. Functional Genomics:

The study of function of all genes present in the entire genome is known as functional genomics. It deals with transcriptome and proteome. The transcriptome refers to complete set of RNAs transcribed from a genome and proteome refers to complete set of proteins encoded by a genome.

In addition to structural and functional genomics, there are other subfields of genomics, including:

Epigenomics involves the study of epigenetic modifications or epigenome, which refers to the collection of chemical compounds that attach to DNA and influence its activity. The epigenome plays an important role in determining the differences between various cell types in the body. Epigenomic modifications include DNA methylation and histone modification.

Metagenomics involves the study of genetic material from entire biological communities rather than just individual organisms. It is generally applied to microorganisms. Metagenomics can be used to study the diversity and function of microbial communities in diverse environments, such as the human gut, soil, or ocean.

Pharmacogenomics is the subfield of genomics that uses an individual's genetic information to study and customize the choice and dosage of drugs in medical treatment. It can be used to predict drug response or toxicity and to develop a more personalized approach to prescribing drugs.

Comparative genomics involves the comparison of genomes from different species that can provide insights into evolutionary relationships, functional elements, and genetic variations among species. It uses various tools that help to identify and understand the similarities and differences in the genomes of various species.

Role of Genomics in Crop Plants:

The discipline of genomics is of recent origin. The genome mapping was first completed in a free-living bacteria *Haemophilus influenza* in 1995. Later on, genome sequencing work was intensified both in prokaryotes and eukaryotes. In plants, genome sequencing

was first completed in *Arabidopsis thaliana* (a weedy relative of mustard followed by rice (*Oriza sativa*). Now genome sequencing work has been completed in more than 40 crop plants.

Genome sequencing strategies:

Genome sequencing is the technique that allows researchers to read the genetic information found in the DNA of anything from bacteria to plants to animals. Sequencing involves determining the order of bases, the nucleotide subunits- adenine(A), guanine(G), cytosine(C) and thymine(T), found in DNA. Genome sequencing is figuring out the order of DNA nucleotides.

Step of genome sequencing:

- Break genome into smaller fragments
- Sequence those smaller pieces
- Piece the sequences of the short fragments together

Genome sequencing approaches:

Two different methods used

1. **Hierarchical shotgun sequencing**:

The method preferred by the Human Genome Project is the hierarchical shotgun sequencing method. It is also known as – The Clone-by-Clone Strategy. This method was the first to be developed. It begins with the construction of genomic libraries of restriction fragments covering all the genomic DNA (or genomic clones) of an organism. By using genetic markers, overlapping clones are assembled into genetic and physical maps encompassing the entire genome. The nucleotide sequence is determined on a clone-by- clone basis until the entire genome is sequenced.

Step:

1) Markers for regions of the genomes are identified.

2) The genome is split into larger fragments (50-200kb) using restriction/cutting enzymes that contain a known marker.

3) These fragments are cloned in bacteria (*E. coli*) using BACs (Bacterial Artificial Chromosomes), where they are replicated and stored.

4) The BAC inserts are isolated and the whole genome is mapped by finding markers regularly spaced along each chromosome to determine the order of each cloned.

5) The fragments contained in these clones have different ends, and with enough coverage finding a scaffold of BAC contigs. This scaffold is called a tiling path. BAC contig that covers the entire genomic area of interest makes up the tiling path.

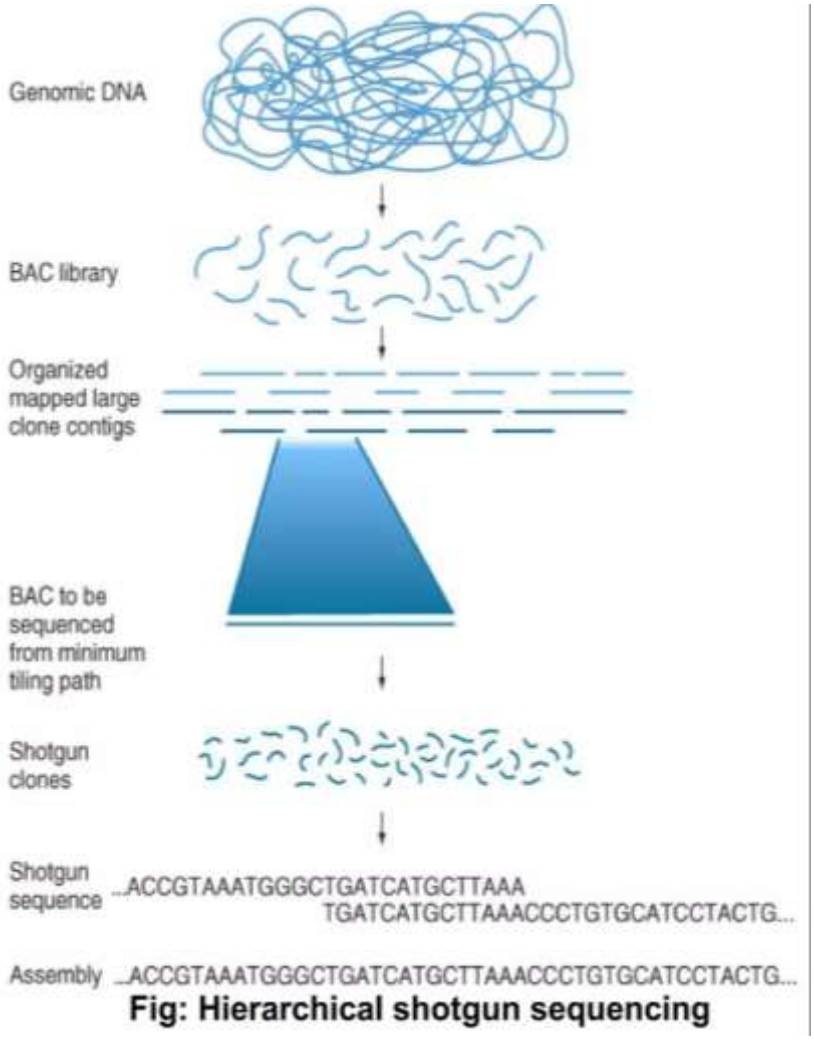


Fig: Hierarchical shotgun sequencing

6) Each BAC fragment in the Golden Path is fragmented randomly into smaller pieces and these fragments are individually sequenced using automated Sanger sequencing and sequenced on both strands.

7) These sequences are aligned so that identical sequences are overlapping. Assembly of the genome is done on the basis of prior knowledge of the markers used to localize sequenced fragments to their genomic location. A computer stitches the sequences up using the markers as a reference guide.

2. **Whole genome Shotgun Sequencing**:

In this method, genomic libraries are prepared and randomly selected clones are sequenced until all clones in the library are analysed. Assembler software organises the nucleotide sequence information into a genome sequence.

This method, developed **by Craig Venter** and his colleagues at the Institute for Genome Research (TIGR), was used to sequence the genome of the bacterium *Haemophilus influenzae* in 1995, the first organism to have its genome completely sequenced.

After refining the method and using it to sequence the genomes of other prokaryotes, the shotgun method was used to sequence eukaryotic genomes, including Drosophila and humans. Using the shotgun method, venter and his colleagues started a privately funded human genome project. The project begun in September 1999 and sequencing was finished in June 2000.

Step:

Library construction:

The large bacterial chromosomes were randomly broken into fairly small fragments, about the size of a gene or less, using ultrasonic waves; the fragments were then purified. These fragments were attached to plasmid vectors, and plasmids with a single insert were isolated. Special *E. coli* strains lacking

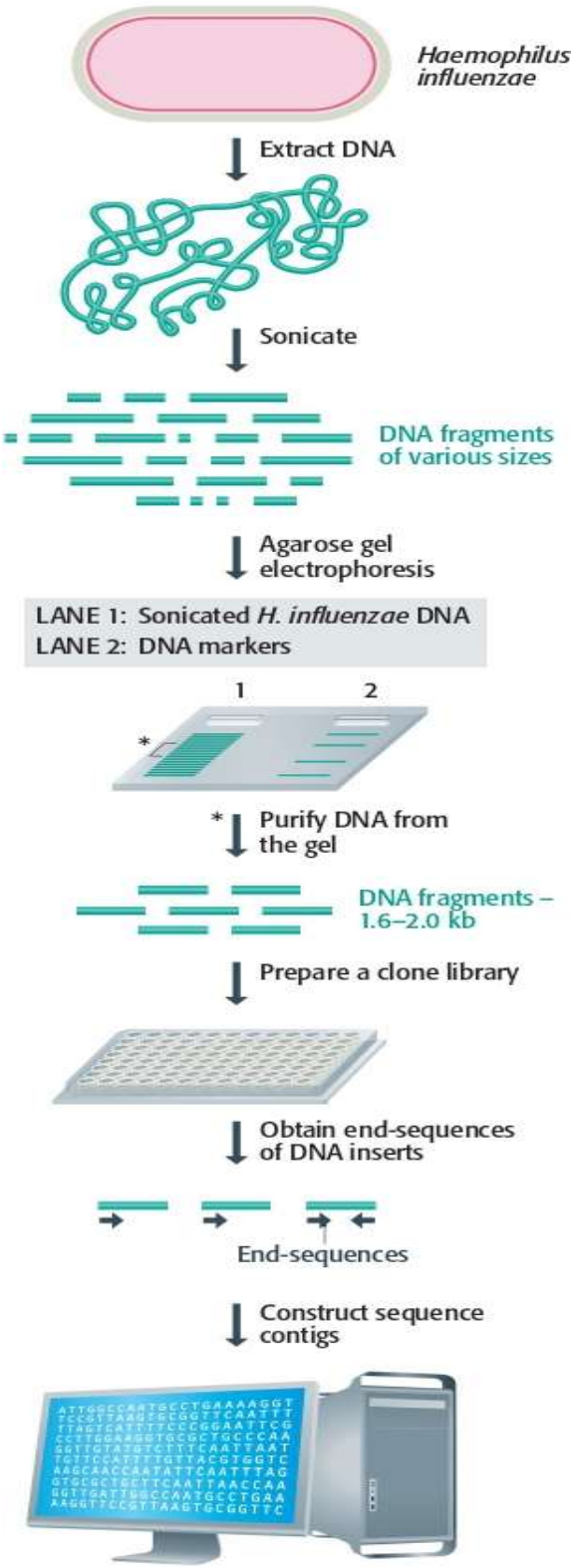


Fig.: Flow chart of whole genome Shotgun Sequencing.

restriction enzymes were transformed with the plasmids to produce a library of the plasmid clones.

Random sequencing:

After the clones were prepared and the DNA purified, thousands of bacterial DNA fragments were sequenced with automated sequencers, employing special dye-labeled primers. Thousands of templates were used, normally with universal primers that recognized the plasmid DNA sequences just next to the bacterial DNA insert. The nature of the process is such that almost all stretches of genome are sequenced many times, and this increases the accuracy of the final results.

Fragment alignment and gap closure:

By the Using specialized computer programs, the sequenced DNA fragments were clustered and assembled into longer stretches of sequence by comparing nucleotide sequence overlie in between the fragments. Two fragments were joined together to form a larger stretch of DNA if the ends of the sequences overlapped and matched (i.e., were the same). This overlap comparison process resulted in a set of larger contiguous nucleotide sequences or contigs.

Finally, the contigs were aligned in the proper order to form the completed genome sequence. If gaps existed between two contigs, sometimes fragment samples with their ends in the two adjacent contigs were available. These fragments could be analyzed and the gaps filled in with their sequences. When this approach was not possible, a variety of other techniques were used to align contigs and fill in gaps.

For example, Phage libraries containing large bacterial DNA fragments were constructed. The large fragments in these libraries overlapped the previously sequenced contigs. These fragments were then combined with oligonucleotide probes that matched the ends of the contigs to be aligned. If the probes bound to a library fragment, it could be used to prepare a stretch of DNA that represented the gap region. Overlapping of the sequence new fragment with two contigs would be placed side-by-side and fill in the gap between them.

Editing:

Proof reading of the sequence is done carefully to resolve any ambiguities in the sequence. Also, the sequence was checked for unwanted frameshift mutations and corrected if necessary.

Genomes of Yeast, *Arabidopsis* and rice

Genomes of Yeast:

S. cerevisiae contains a haploid set of **16 well-characterized chromosomes**, ranging in size from **200 to 2,200 kb**. The total sequence of chromosomal DNA, constituting **12,052 kb,** was released in April, 1996. A total of **6,183 open-readingframes (ORF)** of over **100 amino acids** long were reported, and approximately **5,800** of them were predicated to correspond to actual **protein-coding genes**. A larger number of ORFs were predicted by considering shorter proteins. In contrast to the genomes of multicellular organisms, the yeast genome is highly compact, with genes representing 72% of the total sequence (<2% in the human genome!). The **average size** of yeast genes is **1.45 kb, or 483 codons**, with a range from 40 to 4,910 codons. A total of 3.8% of the ORF contain introns. Approximately 30% of the genes already have been characterized experimentally. Of the remaining 70% with unknown function, approximately one half either contain a motif of a characterized class of proteins or correspond to genes encoding proteins that are structurally related to functionally characterized gene products from yeast or from other organisms.

Ribosomal RNA is coded by approximately 120 copies of a single tandem array on chromosome XII. The DNA sequence revealed that yeast contains 262 tRNA genes, of which 80 have introns. In addition, chromosomes contain movable DNA elements, retrotransposons (Ty elements), that vary in number and position in different strains of *S. cerevisiae*, with most laboratory strains having approximately 30.

Other nucleic acid entities, presented in Figure 3.1, also can be considered part of the yeast genome. Mitochondrial DNA encodes components of the mitochondrial translational machinery and approximately 15% of the mitochondrial proteins.

Genome of *Arabidopsis:*

Arabidopsis is a small plant from the mustard family (Brassicaceae), native to Eurasia and Africa, commonly found along the shoulders of roads and in disturbed land, it is generally considered a weed. It is a popular model organism in plant biology and genetics.

Due to the small size of its genome, and because it is diploid, Arabidopsis thaliana is useful for genetic mapping and sequencing — with about 157 megabase pairs and five chromosomes, *A. thaliana* has one of the smallest genomes among plants. It was the first plant genome to be sequenced, completed in 2000 by the Arabidopsis Genome Initiative. In December of 2000, the *Arabidopsis* research community announced a major accomplishment: the completion of the sequence of a flowering plant.

The genome of ***Arabidopsis***:

- Contains about **125 megabases** of sequence
- Encodes approximately **25,500 genes**
- Contains a similar number of gene functional classifications as other sequenced eukaryotic genomes (*Drosophila melanogaster* and *Ceanorhabditis elegans*)
- Has **35% unique genes**
- Has **37.5% genes that exist as members of large gene families** (families of 5 or more members)
- Shows evidence of **ancient polyploidy**: an estimated 58-60% of the *Arabidopsis* genome exists as **large segmental duplications**.

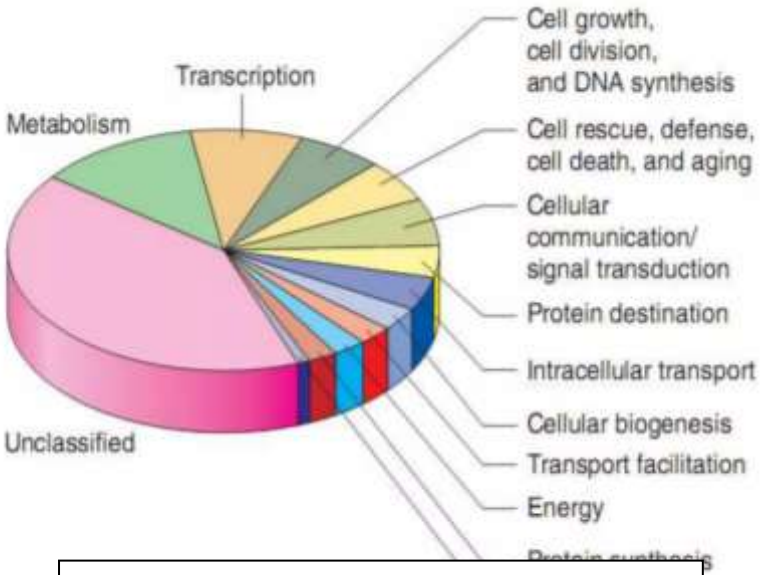


Fig.: Functional analysis of *Arabidopsis* gene.

Table General features of genes encoded by the three genomes in *Arabidopsis*

	Nucleus/cytoplasm	Plastid	Mitochondria
Genome size	125 Mb	154 kb	367 kb
Genome equivalent/cell	2	560	26
Duplication	60%	17%	10%
Number of protein genes	25,498	79	58
Gene order	Variable, but syntenic	Conserved	Variable
Density (kb per protein gene)	4.5	1.2	6.25
Average coding length	1,900 nt	900 nt	860 nt
Genes with introns	79%	18.4%	12%
Genes/pseudogenes	1/0.03	1/0	1/0.2–0.5
Transposons (% of total genome size)	14%	0%	4%

Analysis of the sequence of the *Arabidopsis* genome tells us that the genome of a higher plant is similar in several important ways to the genome of other sequenced multicellular organisms. It also points out several important differences, which may not be too surprising, considering that plants differ in many important ways from the animals whose genomes have been analyzed.

- *Arabidopsis* centromeric regions, although largely **heterochromatic, overall contain at least 47 expressed genes**
- *Arabidopsis* contains several classes of proteins that are used in animal systems for processes not present in the plant, underscoring the idea that evolution makes use of the tools it is given to accomplish different tasks in different organisms
- Plants have evolved a host of signal transduction apparati, perhaps to enable them to deal with their sessile nature
- Arabidopsis genome contains genes encoding RNA polymerase subunits not seen in other eukaryotic organisms

- *Arabidopsis* has genes unique to plants – approximately 150 unique protein families were found, including 16 unique families of transcription factors
- *Arabidopsis* has many gene families common to plants and animals which have been greatly expanded in plants – for instance, Arabidopsis contains 10fold as many aquaporin (water channel) proteins than any other sequenced organism

TAIR (The Arabidopsis Information Resource):

The Arabidopsis Information Resource (TAIR) maintains a database of genetic and molecular biology data for the model higher plant *Arabidopsis thaliana*. TAIR is located at Phoenix Bioinformatics.

Genomes of rice:

The rice genome is about 389 mb, the map-based sequence covered 95% of this genome. Sequenced segment represents 99% of euchromatin. A total of **37,544 genes** with an average gene density of one gene per 9.9 kb and **average gene length of 2,699 bp**. Chromosomes **1 and 3 have the highest gene density**. Chromosomes 11 and 12 have the lowest gene density. 29% in clustered gene families and 2,859 genes seem to be unique to rice and the other cereals, some of which might differentiate monocot and dicot lineages. Between 0.38 and 0.43% of the nuclear genome contains organellar DNA fragments, representing repeated and ongoing transfer of organellar DNA to the nuclear genome.

Transposable elements were maximum for chromosome 8 (38%) and 12 (38.3%) and least for chromosome 1 (31%), 2 (29.8%) and 3 (29%). Contains at least 35% repeat elements. japonica genome sequence showed almost 60% of the genome is duplicated. 421 chloroplast and 909 mitochondrial DNA insertions contributing to ~0.2% each of the nuclear genome. GC content of 43.6% with 54.2% and 38.3% of exons and introns respectively.

Application and Impact of Rice Genome:

- Understanding – plant evolution and the differences between monocots and dicots.
- Improve efficiency of rice breeding.

- Improve nutritional value of rice.
- Enhance crop yield by improving seed quality, resistance to pests and diseases and plant hardiness.
- Useful in identifying plant specific genes that can be potential herbicide targets.
- Development of gene-specific markers for marker-assisted breeding of new and improved rice varieties.

Genome Annotation:

Genome annotation is the process of identifying the location and function of genes and other features in a genome sequence. It involves predicting the locations of genes, coding regions, regulatory sequences, and other functional elements. Annotation is crucial for understanding the genetic makeup of an organism and its potential biological functions. Genome annotation process involves two main steps: gene prediction and functional assignment.

Methods: -

- **Computational Annotation:** Utilizes software algorithms to predict gene locations, splice sites, promoter regions, and other features based on statistical models and sequence similarity.
- **Experimental Annotation:** Employs laboratory techniques such as gene expression analysis, protein sequencing, and functional assays to validate and refine computational predictions.

Tools and Databases for Genome Annotation: -

Gene Mark: Widely used software for ab initio gene prediction in prokaryotic and eukaryotic genomes.

NCBI RefSeq: A curated database of reference sequences with annotations for genes, transcripts, and proteins from a wide range of organisms.

Applications:

Biomedical Research: Annotation facilitates the study of genetic variation, gene regulation, and disease mechanisms, leading to insights into human health and disease.

Biotechnology: Annotated genomes aid in genetic engineering, synthetic biology, and bioprospecting for novel enzymes and metabolic pathways.

Evolutionary Studies: Comparative genomics using annotated genomes reveals evolutionary relationships, gene family expansions, and adaptive traits across different species.

Genome duplication:

Gene duplication is a major mechanism through which new genetic material is generated during molecular evolution. It can be defined as any duplication of a region of DNA that contains a gene. Gene duplications can arise as products of several types of errors in DNA replication and repair machinery as well as through fortuitous capture by selfish genetic elements. Common sources of gene duplications include **ectopic recombination, retrotransposition event, aneuploidy, polyploidy, and replication slippage.**

Whole-Genome Duplication:

Whole-genome duplication is the process by which an organism's entire genetic information is copied, once or multiple times, which is known as polyploidy. This may provide an evolutionary benefit to the organism by supplying it with multiple copies of a gene, thus, creating a greater possibility of functional and selectively favoured genes.

Evidence of Whole-Genome Duplication

In 1997, Wolfe & Shields gave evidence for an ancient duplication of the *Saccharomyces cerevisiae* (Yeast) genome. It was initially noted that this yeast genome contained many individual gene duplications. Wolfe & Shields hypothesized that this was actually the result of an entire genome duplication in the yeast's distant evolutionary history. They found 32 pairs of homologous chromosomal regions, accounting for over half of the yeast's genome.

Evolutionary importance

Paleopolyploidization events lead to massive cellular changes, including doubling of the genetic material, changes in gene expression and increased cell size. Gene loss during

diploidization is not completely random, but heavily selected. Genes from large gene families are duplicated. On the other hand, individual genes are not duplicated. Overall, paleopolyploidy can have both short-term and long-term evolutionary effects on an organism's fitness in the natural environment.

Genome diversity

Genome doubling provides organisms with redundant alleles that can evolve freely with little selection pressure. The duplicated genes can undergo neofunctionalization or subfunctionalization which could help the organism adapt to the new environment or survive different stress conditions.

Speciation

Sympatric speciation can begin with a chromosomal error during meiosis or the formation of a hybrid individual with too many chromosomes, such as polyploidy which can occur during whole-genome duplication. Scientists have identified types of polyploidy that can lead to reproductive isolation of an individual in the polyploid state. In some cases, a polyploid individual will have two or more complete sets of chromosomes from its own species in a condition called autopolyploidy. The other form of polyploidy occurs when individuals of two different species reproduce to form a viable offspring called an allopolyploid. The prefix "allo" means "other" (recall from allopatric); therefore, an allopolyploid occurs when gametes from two different species combine.

Differential gene expression- ESTs:

Expressed Sequence Tags (EST):

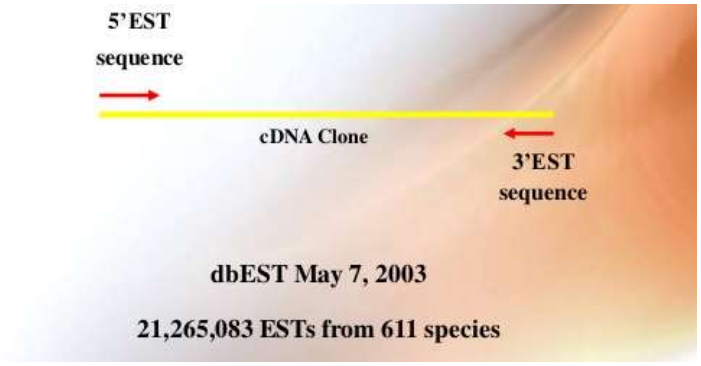
Expressed Sequence Tags (ESTs) are small pieces of DNA and their location and sequence on the chromosome are known. The variations which are found at a single nucleotide position are known. The term Expressed Sequence Tags (ESTs) was first used by Venter and his colleagues in 1991. Main features of EST markers are given below –

1. ESTs are short DNA sequences (200-500 nucleotide long).

2. They are a type of sequence tagged sites (STS).

3. ESTs consist of exons only.

Single-pass sequencing reads from randomly selected cDNA clone.



Steps for ESTs

- •cDNA libraries (containing many of the expressed genes of an organism)
- •pick cDNA clones randomly
- •rapidly determine some of the sequence of nucleotides from the end of each clone.
- •These ESTs could then be compared to all known sequences using a program called ***BLAST***.

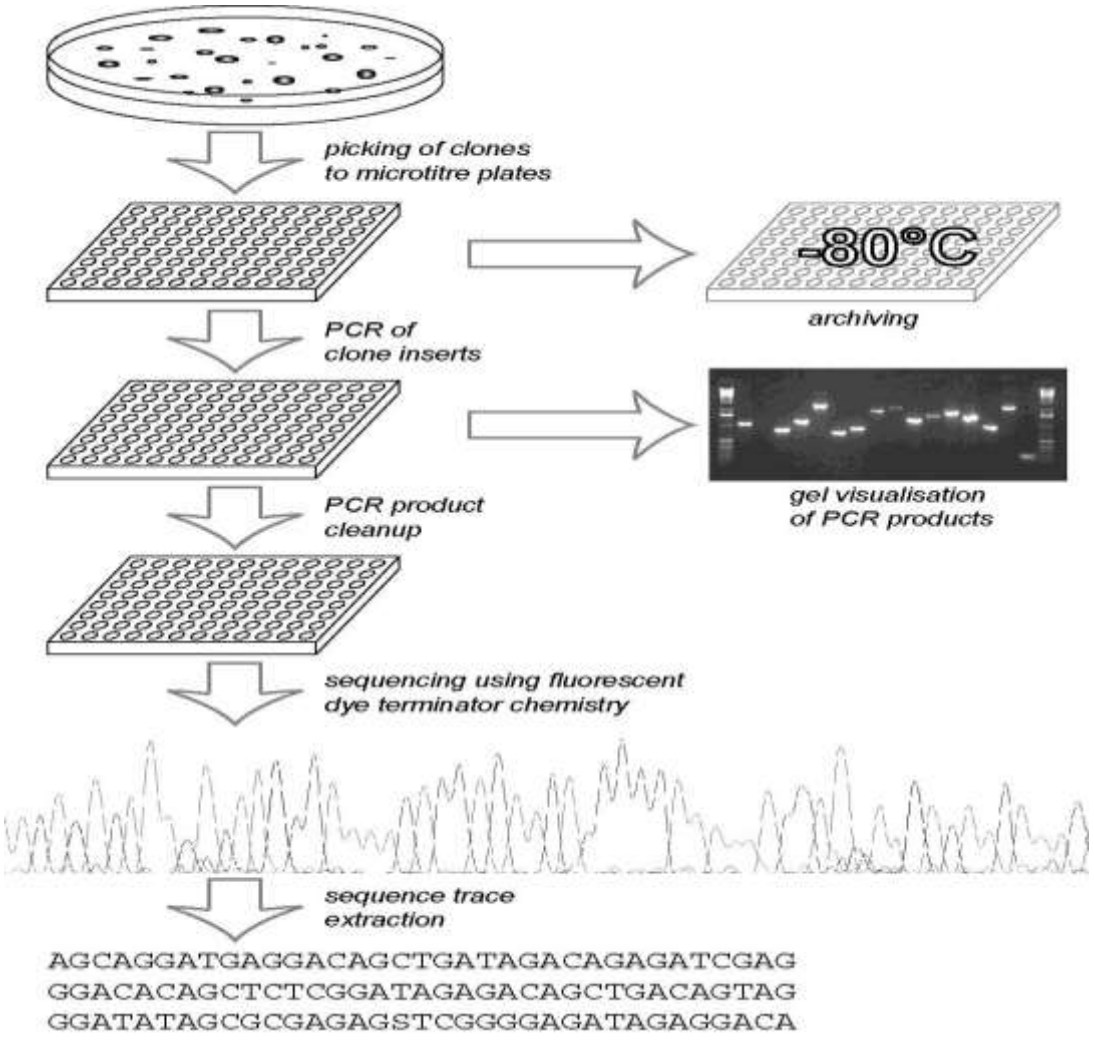
An exact match to a sequenced gene means that the gene encoding that EST is already known.

If the match was close but not exact one could conclude that the EST is derived from a gene with a function similar to that of the known gene.

The EST sequences with their putative identification are then deposited in the GenBank and the clones from which they were derived are kept in a freezer for later use.

Overview of the EST sequencing process

Clones are picked from petriedishes into microtitre plates, and archived for later use. All subsequent manipulations (PCR, clean up and sequencing) are carried out in microtitre plates to yield medium-throughput.



Advantages:

It is a rapid and inexpensive technique of locating a gene. ESTs are useful in discovering new genes related to genetic diseases. They can be used for tissue specific gene expression.

Disadvantages:

ESTs have lack of prime specificity. It is a time consuming and labour-oriented technique. The precision is lesser than other techniques. It is difficult to obtain large (> 6kb) transcripts. Multiplexing is not possible for all loci.

Uses:

ESTs are commonly used to map genes of known function. They are also used for phylogenetic studies and generating DNA arrays.

Gene tagging:

Gene tagging refers to the identification of existing DNA or the introduction of new DNA that can function as a tag or label for the gene of interest. Gene tagging is a most common method used today for selection against different biotic and abiotic stress resistances studies in crop plants. There are four different strategies used for gene tagging.

Types of gene tagging:

1. Marker based gene tagging
2. Transposon tagging
3. T-DNA tagging
4. Epitope tagging

1. Marker based gene tagging:

A Molecular marker is a DNA sequence which is readily detected and whose inheritance can be easily monitored. The effectiveness of molecular markers depends on their ability to identify variation in the DNA of a population, known as marker polymorphism. Molecular markers are detected as differences in DNA fragment size, which arise from differences in DNA sequence. Molecular markers are widely used in marker-assisted breeding for tagging of an important trait or traits in a breeding program.

Molecular tagging of genes for biotic stress resistance

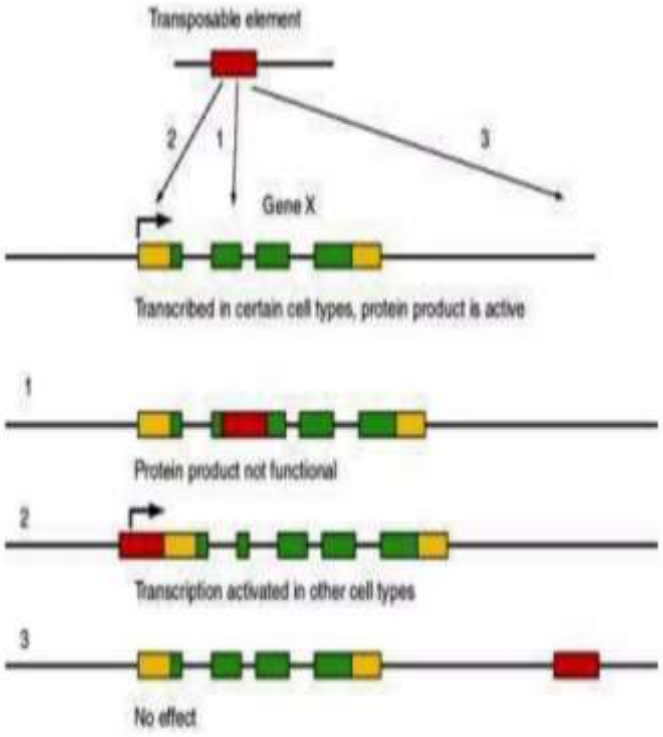
Biotic Stress is stress that occurs as a result of damage done to plants by other living organisms, such as bacteria, viruses, fungi, parasites, beneficial and harmful insects, weeds, and cultivated or native plants. To combat resistance to biotic stresses different

molecular techniques are used these days. Restriction fragment length polymorphism (RFLP) have been used for the molecular tagging of various agronomic traits in different crop species.

Bacterial blight (BB) elicited by Xanthomonas oryzae pv. oryzae (Xoo) is one of the most devastating diseases across the tropics and semi-tropics. Molecular markers are becoming essential components in breeding programs involving gene pyramiding.

Marker-assisted selection for BB resistant genes: Nine F_1 plants from each of the crosses between Mahsuri/IRBB60 , KMR3/IRBB60, PRR78/IRBB60, IR58025B/IRBB60, Pusa 6B/IRBB60 were tested for their heterozygosity for the R gene linked markers and were backcrossed using the female parent.

2. Transposon Tagging: Transposon tagging describes isolation of novel genes using transposable elements as tags. The transposon sequence is used to identify the flanking sequences after insertional mutagenesis. The strategy was initially developed to clone the drosophila white locus. In plants, this strategy was first used to identify and clone bronze1 (bz1) locus in maize.



Schematic of a transposon-tagged mutation.

Types:

Directed-Gene Tagging: Directed tagging identifies the transposon-induced alleles by crossing transposon active plants with a reference homozygous mutant. The mutable alleles are separated from the reference allele by crossing to a standard line(hybrid, inbred or tester).

Non-directed Gene Tagging: In this technique one has to go for M_2 population as compared to F_1 in targeted gene tagging. The main advantage of this technique is that it can also be used to study lethal or infertile mutants, ultimately identifying the gene responsible for it. But targeted tagging is only restricted to mutants non-essential for a plant to complete its life cycle.

3. T-DNA tagging: T-DNA of *Agrobacterium*, generally used as a vector in genetic transformation, could also be considered as a insertional mutagen and with its sequence characterized T-DNA could also be used as a tag.

Insertion of foreign DNA can alter the expression of neighbouring gene, resulting in gain or loss of function which produces a screenable phenotype the sequence of the tag provided a landmark allowing its isolation along with the mutated gene

Many integrate within transcriptional units so that promoters, enhancers, exons and introns can be tagged.

The integrated T-DNA is genetically stable and however from time to time, regions within the T-DNA can become methylated and the resultant reduction in gene expression can result in phenotypic instability. DNA transferred to the plant genome is precisely defined by the border sequences. The genes encoded by the T-DNA can be replaced without interfering with the transfer process.

Types: (a) Promoter tagging (b) Activation tagging

4. Epitope tagging:

An epitope (also called an antigenic determinant) is any structure or sequence that is recognized by an antibody. A single large molecule such as a protein may have many epitopes.

Epitope tagging offers several advantages over other methods of analyzing and purifying proteins: Epitope tagging is much faster than the traditional method of producing a new antibody to every protein studied. The same tag-specific antibody will recognize the epitope tag in many different proteins. Epitope tagging is much less costly and labor intensive than setting up and maintaining antibody-producing facilities. Adding a small (3–14 amino acid) epitope tag generally does not affect the function of the tagged protein, allowing the study of the tagged protein's role in the cell. Epitope tagging makes it possible to gather information about proteins that would otherwise be too difficult to purify or too similar to other proteins to be distinguished in vivo.

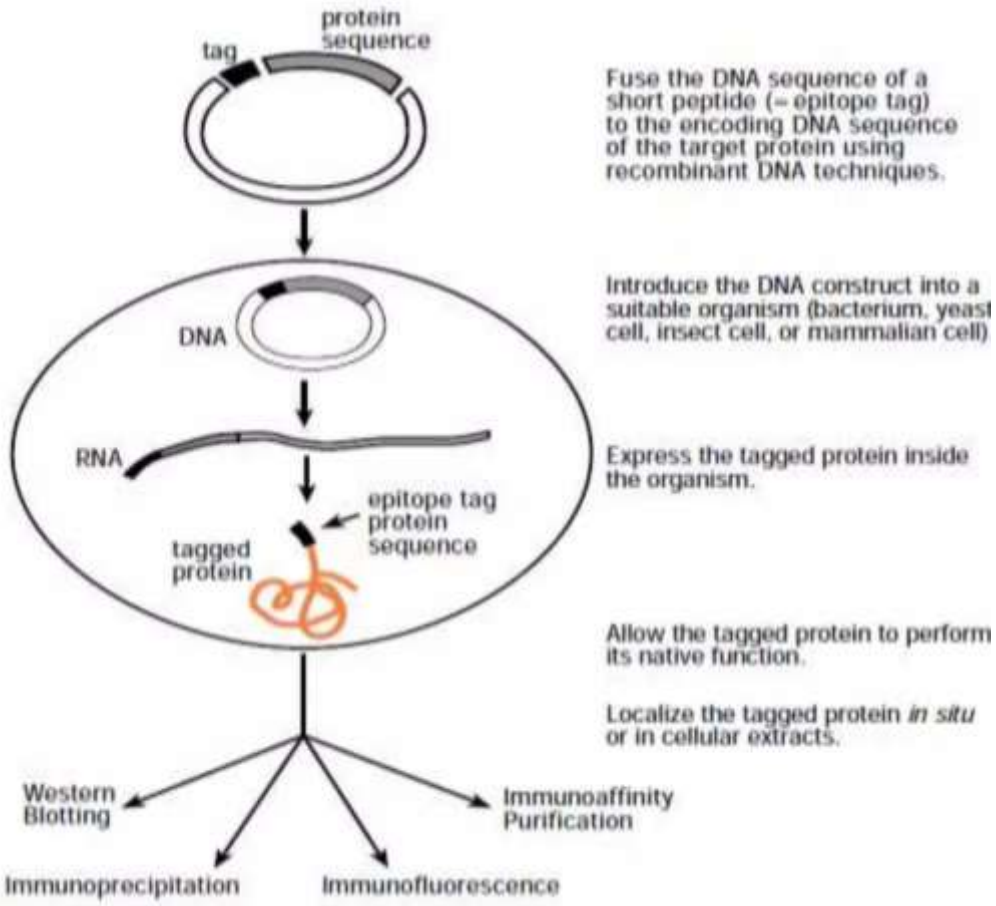
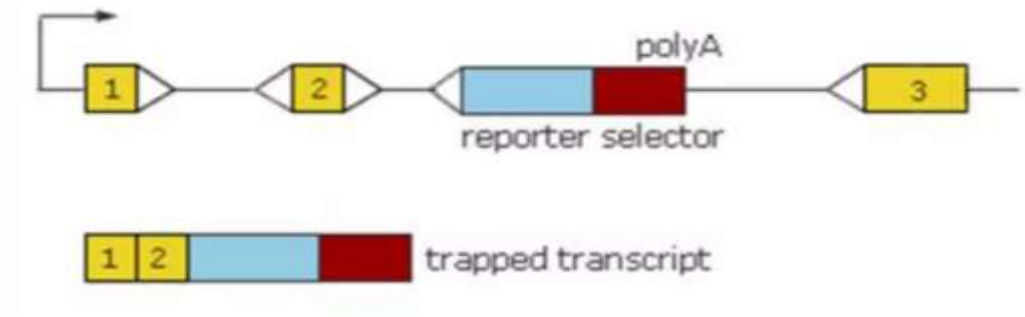


Fig.: Schematic of Epitope tagging technique.

Gene trapping:

Gene trapping is a unique method that helps to identify novel gene by producing random gene disruption by inserting a DNA element, contains a reporter gene and a selectable marker, throughout the genome.



Basic Strategy in Gene Trapping:

1. Choosing proper vector and delivery system
2. Selecting the clones with markers
3. Identification of location of the insert in the clone
4. Studying biological questions: Production of chimeras

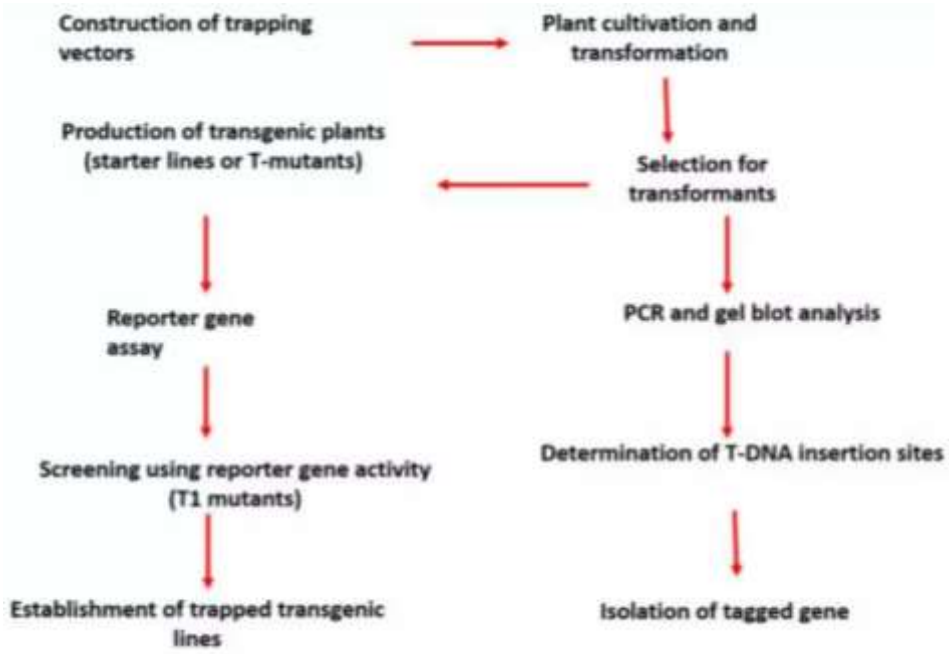


Fig: Gene trapping in plants.

Trap vectors: It should provide a reporter as a tag to easily detect endogenous gene expression and should allow rapid identification of the trapped gene by serving as a molecular tag.

Types:

1. Enhancer trap vectors,
2. Promoter trap vectors,
3. Secretory trap vectors,
4. Poly A trap vectors,
5. Gene trap vectors.

Choice of reporter gene: GUS – Bacterial gene uidA encodes a soluble β- glucuronidase enzyme which breaks down glucuronide substrate to give a coloured reaction

GFP – This gene has been obtained from jellyfish *Aequorea Victoria* It emits green fluorescent light in the blue to ultraviolet range

Luc- Luciferase

Choice of insertion vehicle: T DNA and transposable elements

Creation of gene trapped lines: 1. Identification of insertion events 2. Genetic screen

Table: List of plants where Gene trapping used

S.No.	Promoter	Specificity of expression	Source species	Reference
1.	A1EM	Embryo	*Arabidopsis*	Topping et al. 1994
2.	Cryptic	Seed coat specific	*Nicotiana tabacum*	Fobert et al.1994
3.	HVT1	Tapetum and vascular tissue	*Arabidopsis*	Wel et al.1997
4.	Pyk20	Nematode fonding structure	*Arabidopsis*	Puzio et al.1998
5.	tcup	Constitutive	*Nicotiana tabacum*	Foster et al.1999
6.	Cryptic	Guard cell	*Arabidopsis*	Pleach et al.2000
7.	Cryptic	Roots	*Arabidopsis*	Moher et al.2000
8.	Lj Gbp 1	Roots	*Lotus japonicus*	Webb et al.2000
9.	Eif4A 1	Growing tissues, young leaves	*Arabidopsis*	De Grave et
10.	EXORDIUM(EXO)	Meristematic cells	*Arabidopsis*	Farrar et al.2003

DNA Microarray:

DNA microarrays are solid supports, usually of glass or silicon, upon which DNA is attached in an organized, pre-determined grid fashion. Each spot of DNA, called a probe, represents a single gene. DNA microarrays can analyze the expression of tens of thousands of genes simultaneously. There are several synonyms for DNA microarrays, such as **DNA chips, gene chips, DNA arrays, gene arrays**, and **biochips**.

Principle of DNA Microarray Technique:

- The principle of DNA microarrays lies on the hybridization between the nucleic acid strands.
- The property of complementary nucleic acid sequences is to specifically pair with each other by forming hydrogen bonds between complementary nucleotide base pairs.
- For this, samples are labelled using fluorescent dyes.
- At least two samples are hybridized to chip.
- Complementary nucleic acid sequences between the sample and the probe attached on the chip get paired via hydrogen bonds.
- The non-specific bonding sequences while remain unattached and washed out during the washing step of the process.
- Fluorescently labelled target sequences that bind to a probe sequence generate a signal.
- The signal depends on the hybridization conditions (ex: temperature), washing after hybridization etc while the total strength of the signal, depends upon the amount of target sample present.
- Using this technology the presence of one genomic or cDNA sequence in 1,00,000 or more sequences can be screened in a single hybridization.

Types:

There are 2 types of DNA Chips/Microarrays:

1. **cDNA based microarray**
2. **Oligonucleotide based microarray**

1. Spotted DNA arrays ("cDNA arrays")

Chips are prepared by using cDNA, called cDNA chips or cDNA microarray or probe DNA. The cDNAs are amplified by using PCR. Then these immobilized on a solid support made up of nylon filter of glass slide (1 x 3 inches). The probe DNA are loaded into a spotting spin by capillary action. Small volume of this DNA preparation is spotted on solid surface making physical contact between these two. DNA is delivered mechanically or in a robotic manner.

2. Oligonucleotide arrays (Gene Chips)

In oligonucleotide microarrays, short DNA oligonucleotides are spotted onto the array. Small number of 20-25mers/gene. The main feature of oligonucleotide microarray is that each gene is normally represented by more than one probe and enabled by photolithography from the computer industry.

Requirements of DNA Microarray Technique:

There are certain requirements for designing a DNA microarray system, viz:

- DNA Chip
- Target sample (Fluorescently labelled)
- Fluorescent dyes
- Probes
- Scanner

Steps Involved in cDNA-based Microarray:

The reaction procedure of DNA microarray takes places in several steps:

- **Collection of samples:**

The sample may be a cell/tissue of the organism that we wish to conduct the study on. Two types of samples are collected: healthy cells and infected cells, for comparison and to obtain the results.

- **Isolation of mRNA:**

RNA is extracted from the sample using a column or solvent like phenol-chloroform. From the extracted RNA, mRNA is separated leaving behind rRNA and tRNA.

As mRNA has a poly-A tail, column beads with poly-T-tails are used to bind mRNA.
After the extraction, the column is rinsed with buffer to isolate mRNA from the beads.

- **Creation of labeled cDNA:**

To create cDNA (complementary DNA strand), reverse transcription of the mRNA is done.

Both the samples are then incorporated with different fluorescent dyes for producing fluorescent cDNA strands. This helps in distinguishing the sample category of the cDNAs.

- **Hybridization:**

The labeled cDNAs from both the samples are placed in the DNA microarray so that each cDNA gets hybridized to its complementary strand; they are also thoroughly washed to remove unbounded sequences.

- **Collection and analysis**

The collection of data is done by using a microarray scanner. This scanner consists of a laser, a computer, and a camera. The laser excites fluorescence of the cDNA, generating signals. When the laser scans the array, the camera records the images produced. Then the computer stores the data and provides the results immediately. The data thus produced are then analyzed. The difference in the intensity of the colors for each spot determines the character of the gene in that particular spot.

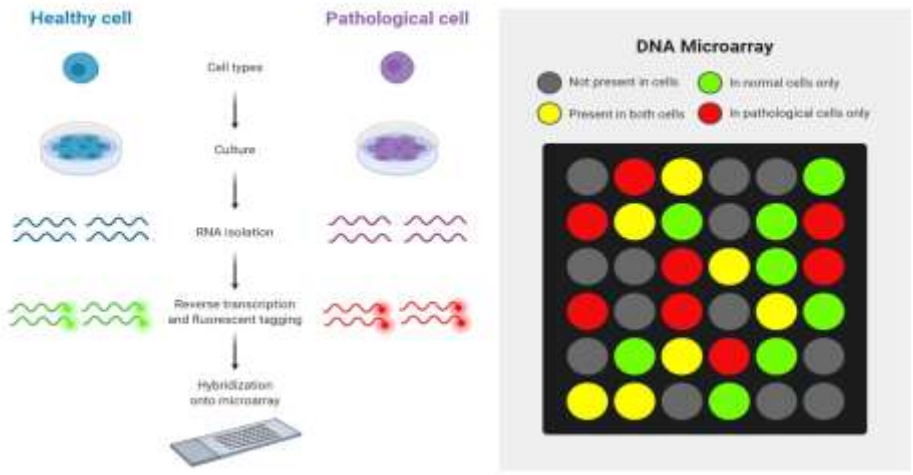


Fig: Steps of DNA microarray.

Applications of DNA Microarray:

1. In humans, they can be used to determine how particular diseases affect the pattern of gene expression (the expression profile) in various tissues, or the identity (from the expression profile) of the infecting organism. Thus, in clinical medicine alone, DNA microarrays have huge potential for diagnosis.
2. Besides, it has applications in many fields such as:
3. Discovery of drugs
4. Diagnostics and genetic engineering
5. Alternative splicing detection
6. Proteomics
7. Functional genomics
8. DNA sequencing
9. Gene expression profiling
10. Toxicological research (Toxicogenomics)

Gene knockout:

A **gene knockout** is a genetic technique in which one of an organism's genes is made inoperative ("knocked out" of the organism). However, KO can also refer to the gene that is knocked out or the organism that carries the gene knockout. The technology of gene knockout is based on gene targeting a useful technique that utilizes homologous recombination to modify the genome of a living organism. **Knockout organisms** or simply **knockouts** are used to study gene function, usually by investigating the effect of gene loss. Researchers draw inferences from the difference between the knockout organism and normal individuals.

The Nobel Prize in Physiology or Medicine 2007 was awarded jointly to **Mario R Capecchi, Sir Martin J Evans** and **Oliver Smithies** "for their discoveries of principles for introducing specific gene modifications in mice by the use of embryonic stem cells"

Molecular

Methods:

Knockouts are accomplished through a variety of techniques. Originally, naturally occurring mutations were identified and then gene loss or inactivation had to be established by DNA sequencing or other methods.

1. Homologous recombination
2. Site-Specific Nucleases

Knockout mutations in the mouse:

Normally, the transgenes are inserted into the genome at random sites. However, if the injected or transfected DNA contains a sequence homologous to a sequence in the mouse genome, it will sometimes be inserted into that sequence by homologous recombination. The insertion of this foreign DNA into a gene will disrupt or "**knock out**" the function of the gene just like the insertion of a transposable genetic element. Indeed, this approach has been used to generate knockout mutations in hundreds of mouse genes.

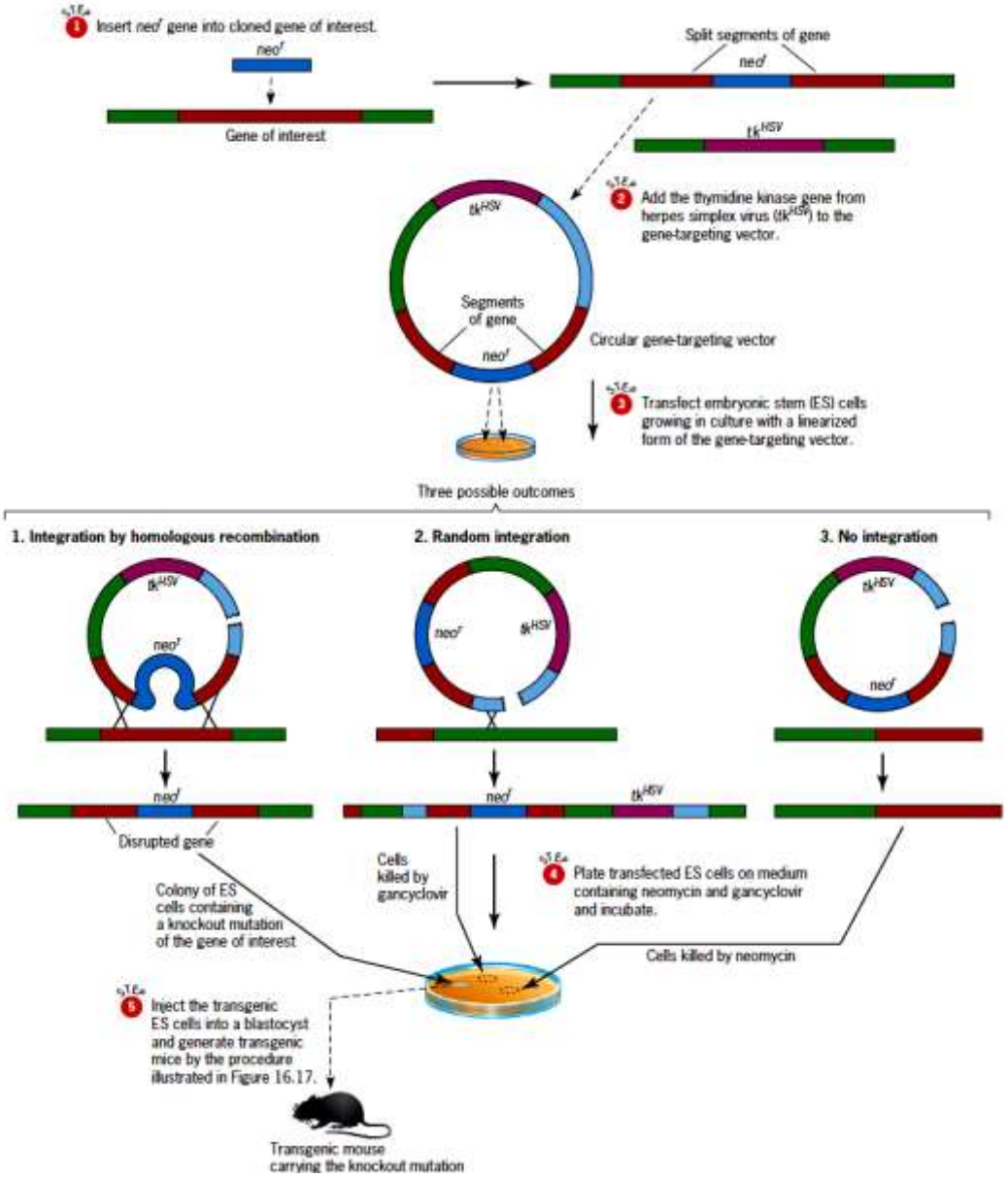
The first step in the production of mice carrying a knockout mutation in a gene of interest is to construct a gene-targeting vector, a vector with the potential to undergo homologous recombination with one of the chromosomal copies of the gene and, in so doing, insert foreign DNA into the gene and disrupt its function. A gene (neor) that confers resistance to the antibiotic neomycin is inserted into a cloned copy of the genemof interest, splitting it into two parts and making it nonfunctional (step 1). The presence of the neor gene in the vector will allow neomycin to be used to eliminate cells not carrying an integrated copy of the gene-targeting vector or the neor gene. The segments of the gene retained on either side of the inserted neor gene provide sites of homology for recombination with chromosomal copies of the gene. The thymidine kinase gene (tkHSV) from herpes simplex virus is inserted into the cloning vector for subsequent use in eliminating transgenic mouse cells resulting from the random integration of the vector. The thymidine kinase from herpes simplex virus (HSV) phosphorylates the drug gancyclovir, and when this phosphorylated nucleotide-analog is incorporated into DNA, it kills the host cell. In the absence of the HSV thymidine kinase, gancyclovir is harmless to the host cell. The next step is to transfect embryonic

stem (ES) cells (from dark-colored mice) growing in culture with linear copies of the gene-targeting vector (step 3) and subsequently plate them on medium containing neomycin and gancyclovir (step 4). Three different events will occur in the transfected ES cells.

(1) Homologous recombination may occur between the split sequences of the gene in the vector and a chromosomal copy of the gene inserting the neor gene into the chromosomal gene and disrupting its function. When this event occurs, the tkHSV gene will not be inserted into the chromosome. As a result, these cells will be resistant to neomycin, but not sensitive to gancyclovir.

(2) The gene-targeting vector may integrate at random into the host chromosome. When this occurs, both the neor gene and the tkHSV gene will be present in the chromosome. These cells will be resistant to neomycin, but killed by gancyclovir.

(3) There may be no recombination between the gene-targeting vector and the chromosome and, thus, no integration of any kind. In this case, the cells will be killed by neomycin. Thus, only the ES cells with the knockout mutation produced by the insertion of the neor gene into the gene of interest on the chromosome will be able to grow on medium containing both neomycin and gancyclovir.



The generation of knockout mutations in the mouse by homologous recombination

The selected ES cells containing the knockout mutation are injected into blastocysts from light-colored parents, and the blastocysts are implanted into light-colored females. Some of the offspring will be chimeric with patches of light and dark fur. The chimeric offspring are mated with light-colored mice, and any dark-colored progeny produced by

this mating are examined for the presence of the knockout mutation. In the last step, male and female offspring that carry the knockout mutation are crossed to produce progeny that are homozygous for the mutation. Depending on the function of the gene, the homozygous progeny may have normal or abnormal phenotypes. Indeed, if the product of the gene is essential early during development, homozygosity for the knockout mutation will be lethal during embryonic development.

Gene knockout Vs gene knockin:

Gene knockout technology aims to either delete part of the DNA sequence or insert irrelevant DNA sequence information to disrupt the expression of a specific genetic locus. Gene knock-in technology, on the other hand, alters the genetic locus of interest via a one-for-one substitution of DNA sequence information or by the addition of sequence information that is not found on said genetic locus. A gene knock-in therefore can be seen as a gain-of-function mutation and a gene knockout a loss-of-function mutation, but a gene knock-in may also involve the substitution of a functional gene locus for a mutant phenotype that results in some loss of function.

Gene knockout Vs gene knockdown:

Making a gene nonfunctional is called gene knockout while reducing the expression of a gene is known as gene knockdown

In a simple language, we can say in the gene knockout method we are making a gene inactive totally, while in the gene knockdown, the gene is active but the expression of a gene is reduced to know its activity in a particular cell type One of the best methods for inactivating a gene is by introducing a mutation gene knockout.

Conditional Knockout:

The conditional knockout method is used to inactivate the gene in a specific tissue at a specific time for a specific function In addition to this, it is done in the adult animal instead of during the embryonic stage through the process of homologous recombination.

Cancer like lethal conditions can be studied by the conditional gene knockout method using mammalian model organism.

Recombination with Cre-loxP-cre or cyclization recombination enzyme is able to recognize and cut DNA sequence specifically which is further followed by recombination with second enzyme loxP Within DNA, Cre recombinase is able to bind with two loxP site resulting a recombination between Depending on the arrangement of loxP site, the recombination produces deletion or inversions of the target genes Due to the inducible nature of cre loxP the knockout can be induced by chemicals such as tetracyclin and tamoxifane.

Applications of gene knockout:

- One of the important application of gene knockout is to study the function of a particular gene.
- It also enables scientists to monitor and control the effect of a gene.
- Gene knockout method is used for constructing genetically modified organism such as GM plants, GM bacteria and GM animals.
- It is also used to study the effect and contribution of a particular gene and its role in the development of a disease.
- It is likewise employed in drug discovery: using gene knockout like genetic engineering tools, drug screening can be done.

Gene silencing:

Gene silencing is a technique that aims to reduce or eliminate the production of a protein from its corresponding gene. It generally describes **the "switching off"** of a gene by a mechanism other than genetic modification. It occurs when RNA is unable to make a protein during translation. Gene silencing is same as gene **knock down** but is totally different from **gene knock out**.

Types of Gene silencing:

Genes are regulated at either the transcriptional level or post-transcriptional level, therefore silencing can be induced either at transcriptional level or post- transcriptional level. There are mainly two types of gene silencing

1. Transcriptional gene silencing
2. Post transcriptional gene silencing

1. Transcriptional gene silencing:

It is the result of histone modifications, creating an environment of heterochromatin around a gene that makes it inaccessible to transcriptional machinery (RNA polymerase, transcription factors, etc.).

Genomic imprinting:

Genomic imprinting is a genetic phenomenon by which certain genes are expressed in a parent-of-origin-specific manner. It is an inheritance process independent of the classical Mendelian inheritance. genomic imprinting has been demonstrated in insects, mammals and flowering plants.

Step 1: Alleles of the Igf2 gene are imprinted in the parental germ lines—methylated in the female germ line and not methylated in the male germ line.

Step 2: Imprinted alleles of the Igf2 gene from each parent are combined in the zygote at fertilization.

Step 3: During development of the somatic tissues, the maternally contributed allele remains methylated while the paternally contributed allele remains unmethylated. In somatic cells, only the unmethylated, paternally contributed allele is expressed. The methylated, maternally contributed allele is silent.

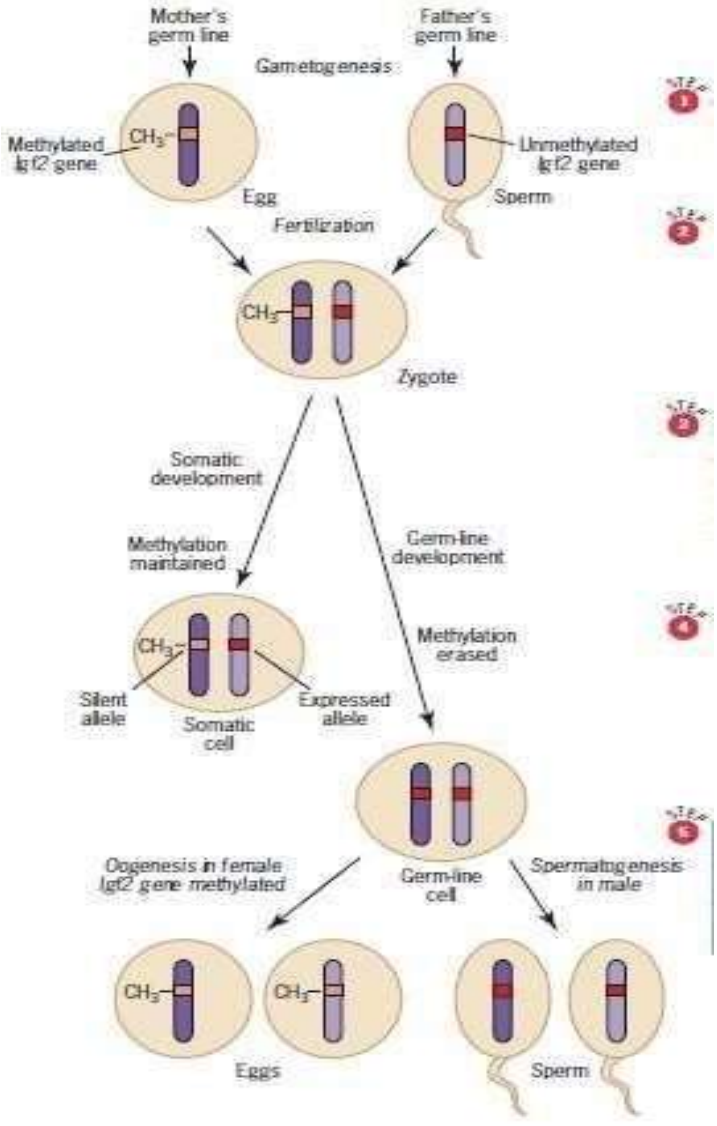


Fig: Methylation and imprinting of the *Igf*2 gene in mice.

Step 4: During development of the germ line, the methylation imprint is erased.

Step 5: Methylation is reestablished during oogenesis, but not during spermatogenesis. Thus, if the mouse is female, all Igf2 genes will be methylated, even if they are copies of the unmethylated Igf2 allele inherited from the father. If the mouse is male, none of the Igf2 genes will be methylated even if they are copies of the methylated Igf2 allele inherited from the mother.

Paramutation:

Paramutation is an interaction between two alleles of a single locus, resulting in a heritable change of one allele that is induced by the other allele. Paramutation was first observed by the effect it had on the colour of corn kernels in maize plants. This change may be in the pattern of DNA methylation. For example – Anthocyanin pigment in corn plant B allele – Anthocyanin pigment coded.

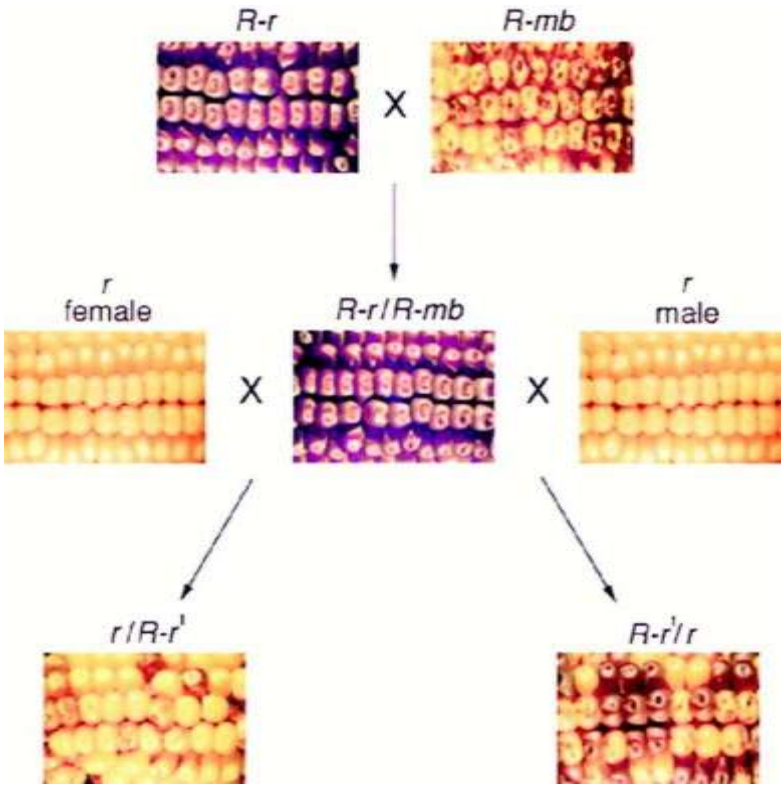


Fig: Paramutation in the maize plant.

- Paramutagenic allele at this locus(B') cause reduced pigment production
- B allele is silenced by the B' allele in the first generation
- In next generation, the newly silence B allele is paramutagenic and silence.

Position effect:

Position effect is the effect on the expression of a gene when its location in a chromosome is changed, often by translocation. This has been well described in Drosophila with respect to eye colour and is known as position effect variegation (PEV).

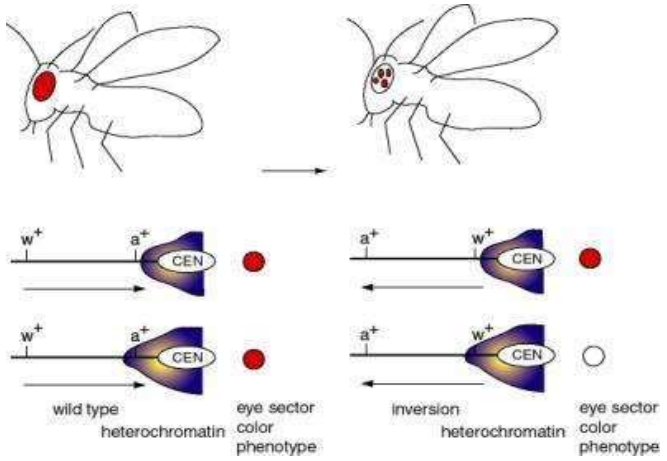


Fig: Position effect in *Drosophila*.

Post transcriptional gene silencing:

The ability of exogenous or sometimes endogenous RNA to supress the expression of the gene which corresponds to the m-RNA sequence.

Anti sense RNA technology:

Antisense RNA has the opposite sense to m-RNA. The presence of complimentary sense and antisense RNA in the in the same cell can lead to the formation of a stable duplex which interferes with gene expression at the level of RNA processing or possible translation. This technology widely used in plants Antisense RNA inhibits protein synthesis by blocking Translation.

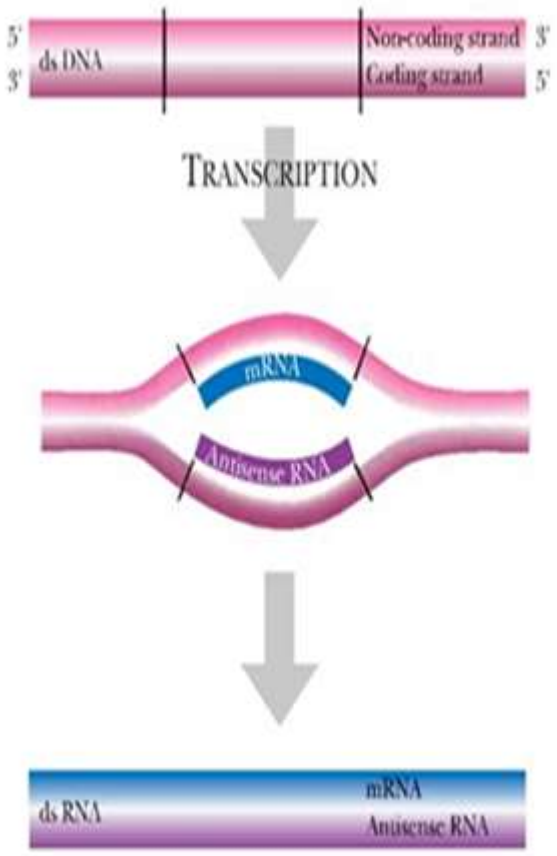


Fig: Anti sense RNA technology

Application of Gene silencing:

- Specific gene silencing using RNAi in cell culture.
- Cancer treatments
- RNA interference has been used for applications in biotechnology, particularly in the engineering of food plants that produce lower levels of natural plant toxins.
- Such techniques take advantage of the stable and heritable RNAi phenotype in plant stocks. For example, cotton seeds.
- Modulation of HIV-I replication by RNAi.
- Small RNA and its application in andrology and urology.
- Developing technologies for epigenomic analysis and clinical application of molecular diagnosis.

- Currently there are at least six oligonucleotide drugs inducing Specific gene silencing using RNA i in cell culture.
- Cancer treatments
- RNA interference has been used for applications in biotechnology, particularly in the engineering of food plants that produce lower levels of natural plant toxins.
- Such techniques take advantage of the stable and heritable RNA i phenotype in plant stocks. For example, cotton seeds.
- Modulation of HIV-I replication by RNA i.
- Small RNA and its application in andrology and urology.
- Developing technologies for epigenomic analysis and clinical application of molecular diagnosis.
- Currently there are at least six oligonucleotide drugs inducing RNAi for illness including cancer.
- RNAi for illness including cancer.

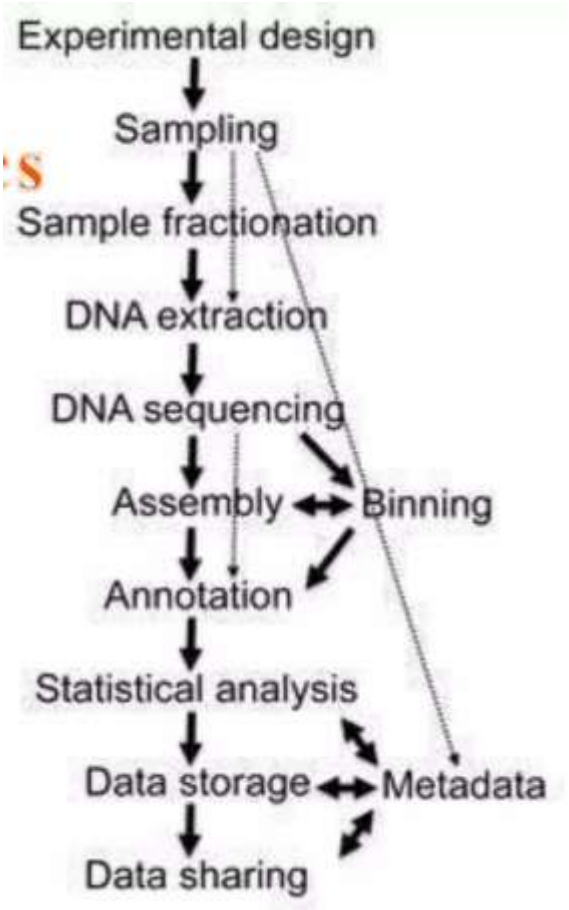
Metagenomics

Metagenomics is the study of metagenome, genetic material, recovered directly from environmental sample such as soil, water or faeces. Metagenomics is based on the genomics analysis of microbial DNA directly from the communities present in samples Metagenomics technology – genomics on a large scale will probably lead to great advances in medicine, agriculture, energy production and bioremediation. Metagenomics can unlock the massive uncultured microbial diversity present in the environment for new molecule for therapeutic and biotechnological application. These studies have identified many novel microbial genes coding for metabolic pathways such as energy acquisition, carbon and nitrogen metabolism in natural environments that were previously considered to lack such metabolism.

The term **metagenomics** first used by **Jo Handelsman, Jon Clarly, Robert M. Goodma**n and first appeared in publication in 1998. Metagenomics defined as **"the genomics analysis of microorganism by direct extraction and cloning DNA from an assemblage of microorganism."** o In **Greek, meta** means **"transcendent"**

(combination of separate analysis) Genomics refers to the study of the genome Jo Handelsman.

The science of metagenomics, only a few years old, will make it possible to investigate microbes in their natural environments, the complex communities in which they normally live. It will bring about a transformation in biology, medicine, ecology, and biotechnology that may be as profound as that initiated by the invention of the microscope. All plants and animals have closely associated microbial communities that make necessary nutrients (carbon, nitrogen, oxygen, and sulfur) metals, and vitamins available to their hosts. We depend on microbes to remediate toxins in the environment—both the ones that are produced naturally and the ones that are the byproducts of human activities, such as oil and chemical spills.



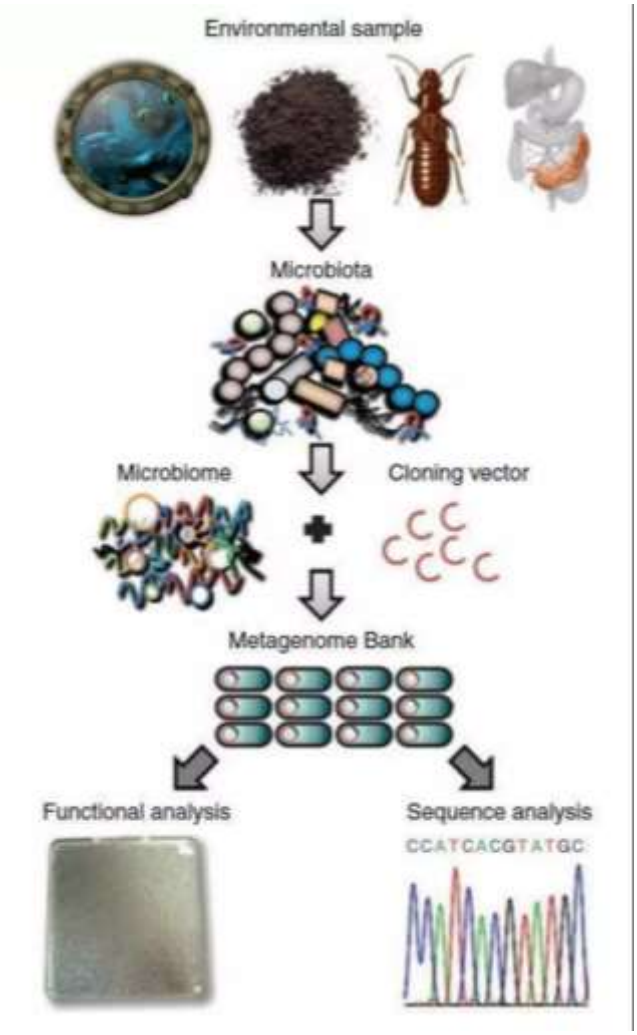


Fig: Overview of Metagenomics.

Type: There are two basic types of Metagenomics studies **I. Sequence-based Metagenomics**- involve sequencing and analysis of DNA from environmental samples **II. Function-based Metagenomics** involves screening for a particular function or activity.

I. Sequence-based metagenomics studies can be used to assemble genomes, identify genes, find complete metabolic pathways, and compare organisms of different communities. Genome assembly requires lots of computer power but it can lead to a better understanding of how certain genes help organisms survive in a particular environment. Sequence-based metagenomics can also be used to establish the degree of diversity and the number of different bacterial species existing in a particular sample.

II. Functional metagenomics is a powerful experimental approach for studying gene function, starting from the extracted DNA of mixed microbial populations. A functional approach relies on the construction and screening of metagenomic libraries—physical libraries that contain DNA cloned from environmental metagenomes. Functional metagenomics begins with the construction of a metagenomic library, Cosmid- or fosmid-based libraries are often preferred due to their large and consistent insert size and high cloning efficiency. The information obtained from functional metagenomics can help in future annotation of gene function and serve as a complement to sequence – based metagenomics. Using this function-based approach allows for discovery of novel enzymes whose functions would not be predicted based on DNA sequence alone.

Application of metagenomics:

Metagenomics has the potential to advance knowledge in a wide variety of field, like Medicine, Engineering Agriculture, Ecology, and Biotechnology.

- Metagenomics can improve strategies for monitoring the impact of pollutants on ecosystems and for cleaning up contaminated environments. Increased understanding of bioaugmentation or biostimulation trials to succeed.
- Recent progress in mining the rich genetic resource of nonculturable microbes has led to the discovery of new gene, enzymes and natural products. The impact of metagenomics is witnessed in the development of commodity and fine chemicals, agrochemicals and pharmaceuticals where the benefit of enzyme catalyzed chiral synthesis is increasingly recognized.
- Metagenomics libraries are, indeed, an essential tool for the discovery of new enzymatic activities, facilitating genetic tracking for all biotechnological applications of interest for the future.
- Metagenomics sequencing is being used to characterize the microbial communities. This is part of the human micro-biome initiative with primary goals to determine if there is a core human micro-biome, to understand the changes in the human micro-biome that can be correlated with human health, and to develop new technological and bioinformatics tools to support these goals.

- It is well known that the vast majority of microbes have not been cultivated. Functional metagenomics strategies are being used to explore the interactions between plants and microbes through cultivation-independent study of the microbial communities.

9. Genome Editing Technologies: CRISPR, TALEN, LEAPER and their applications in crop improvements.

Genome editing (gene editing):

It is a type of genetic engineering in which DNA is inserted, deleted, modified or replaced in the genome of a living organism. Unlike early genetic engineering techniques that randomly inserts genetic material into a host genome, genome editing targets the insertions to **site-specific locations**.

The basic mechanism involved in genetic manipulations through programmable **nucleases** is the recognition of target genomic loci and binding of effector DNA-binding domain (DBD), double-strand breaks (DSBs) in target DNA by the restriction endonucleases (FokI and Cas), and the repair of DSBs through **homology-directed recombination** (HDR) **or non-homologous end joining** (NHEJ).

History:

- Genome editing was pioneered in the 1990s, before the advent of the common current nuclease-based gene editing platforms but its use was limited by low efficiencies of editing.
- All three major classes of these enzymes—**zinc finger nucleases (ZFNs)**, **transcription activator-like effector nucleases (TALENs)** and engineered **meganucleases**—were selected by **Nature Methods as the 2011**

- ❖ The CRISPR-Cas system was selected by **Science as 2015** Breakthrough of the Year.
- ❖ As of 2015 four families of engineered nucleases were used: meganucleases, ZFNs, TALEN and CRISPR/Cas9 system.
- ❖ Nine genome editors were available as of 2017.

Why genome editing?

- ➢ To understand the function of a gene or a protein, one interferes with it in sequence-specific way and monitors its effects on the organism.
- ➢ In some organisms, it is difficult or impossible to perform site-specific mutagenesis, and therefore more indirect methods must be used such as silencing the gene of interest by short RNA interference (siRNA).
- ➢ Nucleases such as CRISPR can cut any targeted position in the genome and introduce a modification of the endogenous sequences for genes that are impossible to specifically target using conventional RNA interference (RNAi).
- ➢ To increase the nutritional content and pest resistance in the crop with knock-out therapies.

Tools for Genome Editing:

The core technologies now most commonly used to facilitate genome editing are

(1) clustered regularly interspaced short palindromic repeats (CRISPR)-CRISPR-associated protein 9 (Cas9)

(2) transcription activator-like effector nucleases (TALENs)

(3) zinc-finger nucleases (ZFNs)

(4) homing endonucleases or meganucleases.

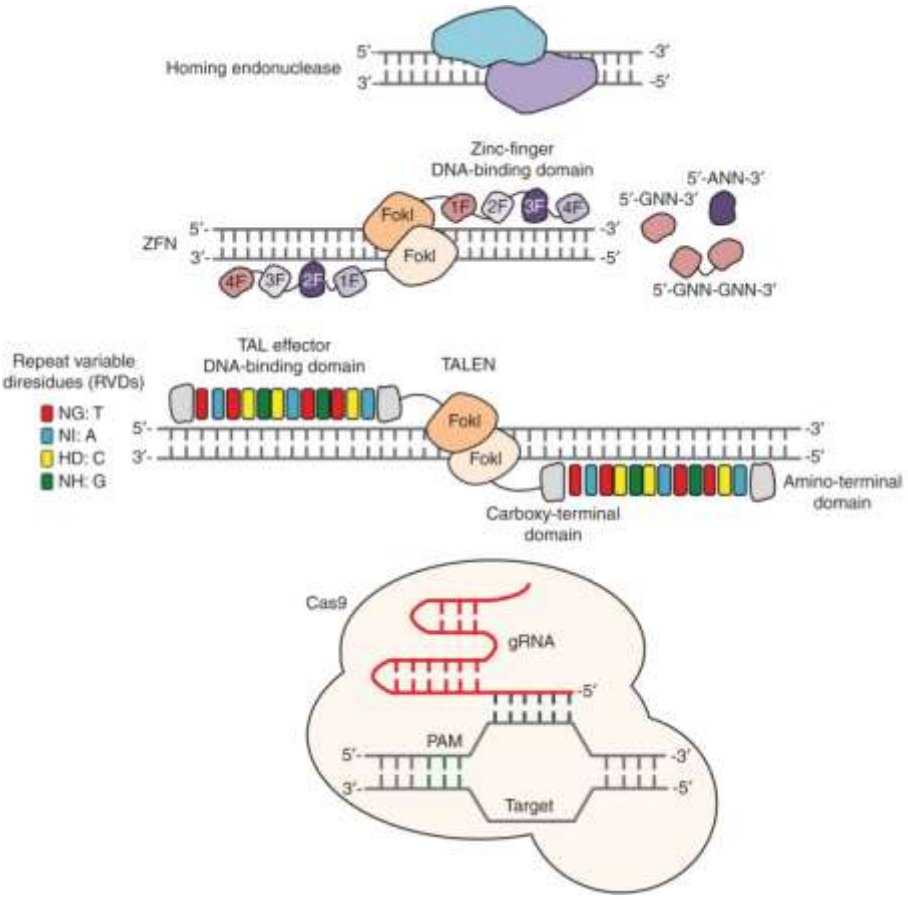


Fig.: Genome-editing technologies – Meganucleases, ZFNs, TALENs, CRISPR-Cas9 (top to bottom).

CRISPR:

CRISPR (an acronym for **clustered regularly interspaced short palindromic repeats**) is a family of DNA sequences found in the genomes of prokaryotic organisms such as bacteria and archaea. These sequences are derived from DNA fragments of bacteriophages that had previously infected the prokaryote. They are used to detect and destroy DNA from similar bacteriophages during subsequent infections. Hence these sequences play a key role in the antiviral defense system of prokaryotes and provide a form of acquired immunity. CRISPR is found in approximately 50% of sequenced bacterial genomes and nearly 90% of sequenced archaea.

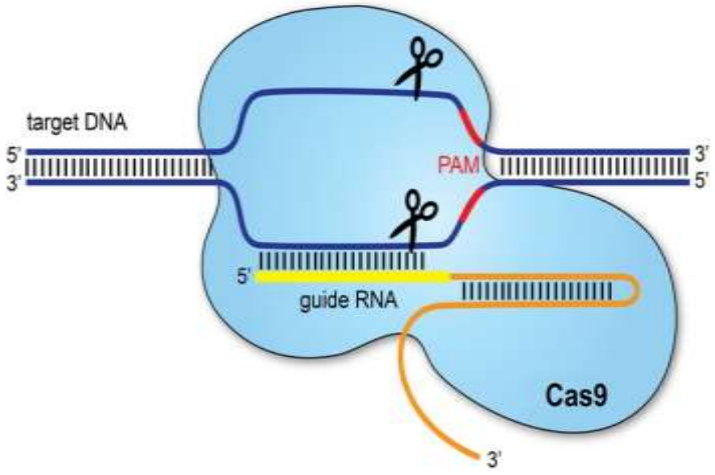
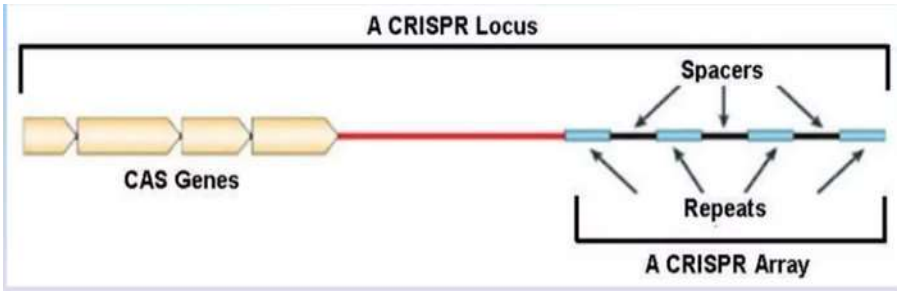


Fig.: The CRISPR/Cas9 system.

Cas9 (or "CRISPR-associated protein 9") is an enzyme that uses CRISPR sequences as a guide to recognize and open up specific strands of DNA that are complementary to the CRISPR sequence. Cas9 enzymes together with CRISPR sequences form the basis of a technology known as CRISPR-Cas9 that can be used to edit genes within the organisms.



A CRISPR array is composed of a series of repeats interspaced by spacer sequences acquired from invading genomes.

History:

- 1987- CRISPR sequences were first discovered in Escherichia coli. (Ishino et al., 1987)
- 2002- Identification of Cas genes that are associated with DNA repeats in prokaryotes. (Jansen et al., 2002).

- 2007- CRISPR provides acquired resistance against viruses in prokaryotes. (Barrangou et al., 2007)
- 2012- Idea of using CRISPR- Cas9 as a genome engineering tool was published by **Jennifer Doudna** and **Emmanuelle Charpentier** and got **Nobel Prize** in Chemistry in 2020.

Components of CRISPR:

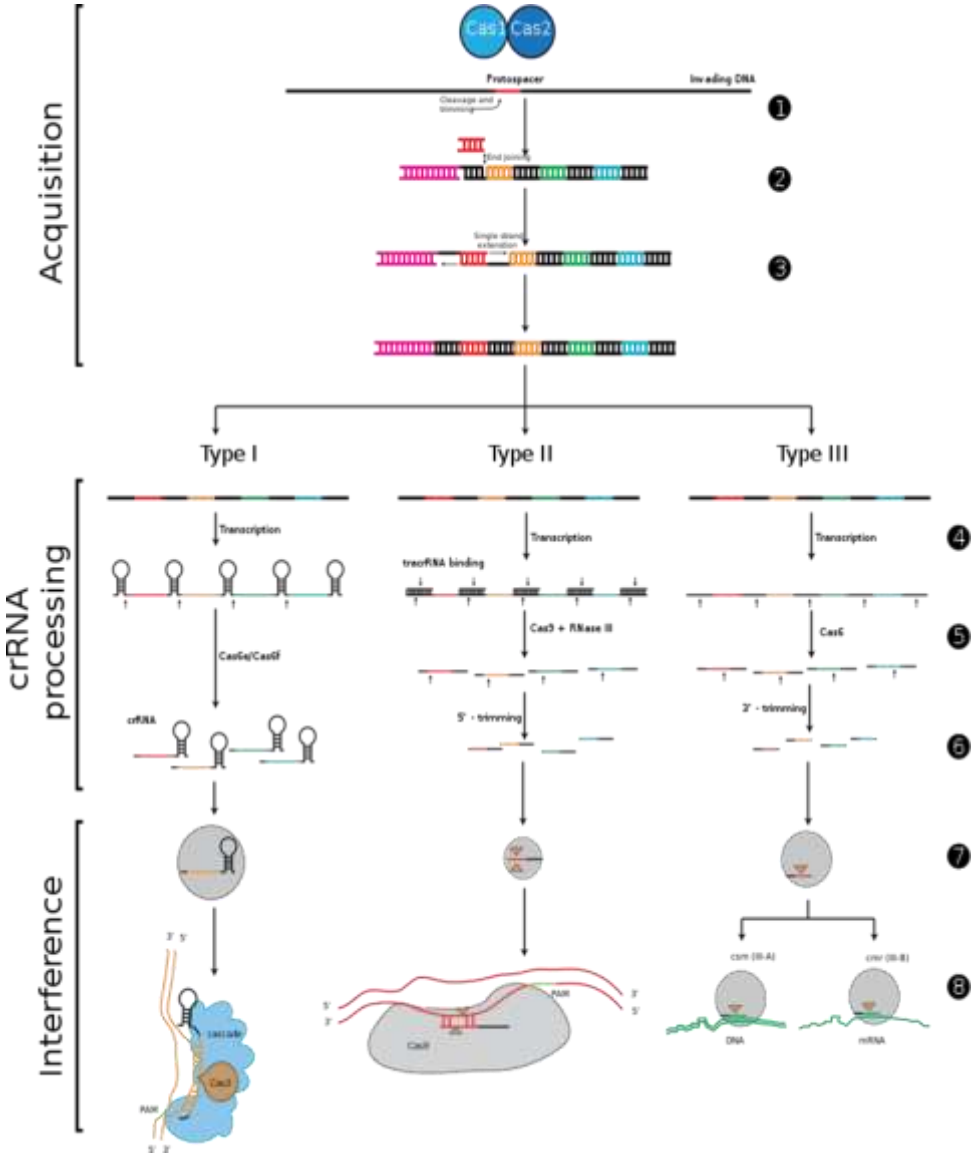
1. **crRNA/ CRISPR RNA:** The crRNA is an 18–20 base pair in length that specifies the target DNA by pairing with the target sequence. It is a Guide RNA present in Host DNA that binds with tracrRNA and form hairpin complex.
2. **tracrRNA:** Trans active RNA that bind with crRNA form active complex.
3. **sgRNA:** Single guide RNA (crRNA + tracrRNA)
4. **Cas9:** The first Cas protein used in genome editing was extracted from **Streptococcus pyogenes** (SpCas-9). It is a large (1368 amino acids) multi-domain DNA endonuclease responsible for cleaving the target DNA to form a double- stranded break and is called a **genetic scissor**. Cas-9 consists of two regions, called the **recognition (REC) lobe** and the **nuclease (NUC) lobe**. The REC lobe consists of REC1 and REC2 domains responsible for binding guide RNA, whereas the NUC lobe is composed of RuvC, HNH, and **Protospacer Adjacent Motif (PAM)** interacting domains.
5. **Protospacer adjacent motif (PAM) —** required for Cas9 function, this sequence motif is immediately downstream of the target sequence. Cas9 recognizes the PAM sequence 5'-NGG, where N can be any nucleotide (A, T, C, or G). When Cas9 binds the PAM, it separates the DNA strands of the adjacent sequence to allow binding of the sgRNA. If the sgRNA is complementary to that sequence, Cas9 cuts the DNA

Different CRISPR-Cas system in Bacterial Adaptive Immunity:

- **Class 1- type I** (CRISPR-Cas3) and **type III** (CRISPR- Cas10) uses several Cas proteins and the crRNA

- **Class 2- type II** (CRISPR-Cas9) and **type V** (CRISPR- Cpf1): employ a large single-component Cas-9 protein in conjunction with crRNA and tracerRNA. Zetsche et al., (2015) functioning of type II CRISPR system

  CRISPR-Cas Defense Mechanism in Bacteria: The CRISPR-Cas mediated Défense process can be divided into three stages:



1. The first stage, **adaptation**, leads to insertion of new spacers in the CRISPR locus.

2. In the second stage, **expression**, the system gets ready for action by expressing the Cas genes and transcribing the CRISPR into a long precursor CRISPR RNA (pre-crRNA). The pre-crRNA is subsequently processed into mature crRNA by Cas proteins and accessory factors.
3. In the third and last stage, **interference**, target nucleic acid is recognized and destroyed by the combined action of crRNA and Cas proteins complex.

Mechanism of CRISPR-Cas9 mediated genome editing:

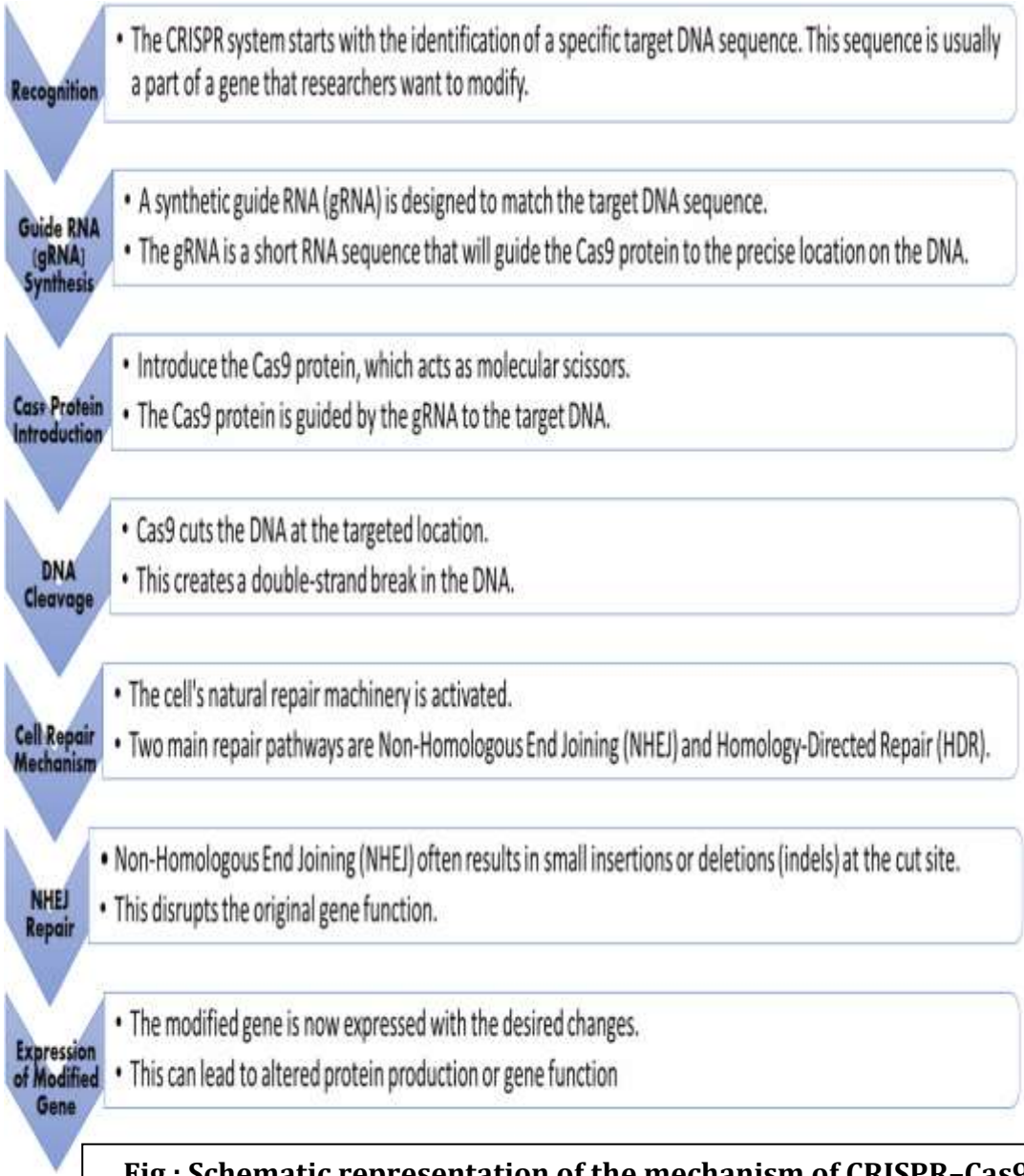
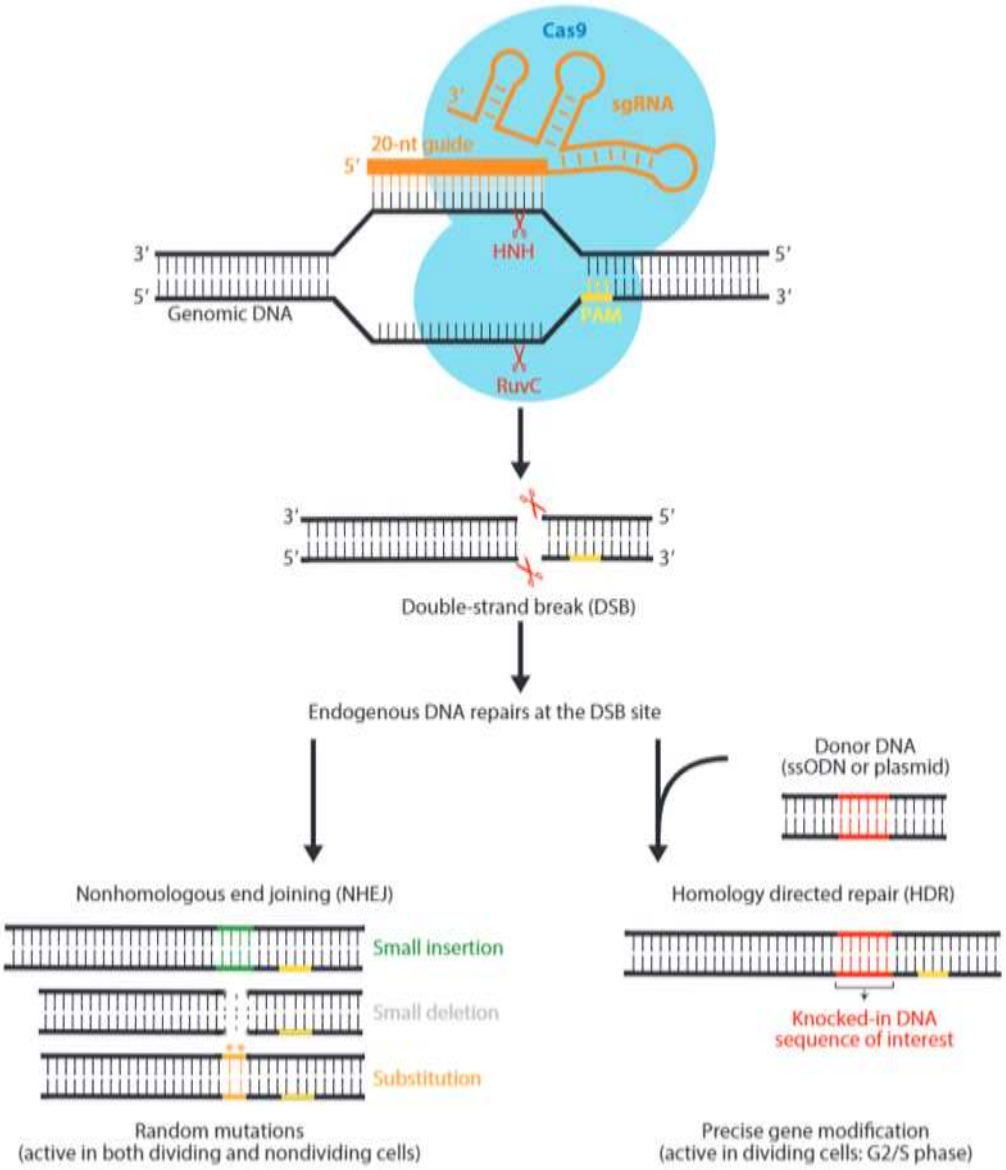


Fig.: Schematic representation of the mechanism of CRISPR–Cas9–mediated genome editing.



Application of CRISPR-Cas9 in crop improvements:

Researchers have found that CRISPR-Cas9 can be applied to nearly every organism. Early studies using CRISPR - Cas9 for gene editing have focused on crops important for agriculture.

1. **Improvement of Crop Yield:**

Increasing yield is a major objective of studies aimed at improving crops. Specific grain yield-related traits include grain number and size per panicle, tiller number per panicle,

grain weight, and grain size. Miao et al. generated a **pyl1/4/6 triple knockout rice mutant** using the CRISPR/Cas9 system. Compared with the wild-type control, the mutant had a higher yield, longer panicles, more primary and secondary branches in the panicles, and fewer tillers per plant. Zhang et al. (2016) targeted **TaGASR7 by transiently expressing CRISPR/Cas9** DNA or RNA in the calli of **hexaploid wheat and tetraploid durum wheat** and observed that the **1000-grain weight increased** in the T_0 mutant. Deleting the **TaGW2 gene** encoding a **RING E3 ligase in wheat** increases the grain length and width, thereby increasing the grain yield.

2. **Improvement of Crop Quality:**

With the general improvement in global living standards, the demand for high-quality crops is increasing. The market value of crops is greatly influenced by crop quality, which is determined by external and internal traits. The CRISPR/Cas9 system can be used to decrease the amylose content and improve the nutritional value and flavor of rice grains. Ma et al. successfully **reduced the amylose content of rice grains from 14.6% to 2.6%** by knocking out the **WAXY (Wx) gene,** thereby obtaining waxy rice mutant plants. Similarly, the **maize Wx1 gene** encodes a starch synthase affecting the grain composition. Knocking out **Wx1 via the CRISPR/Cas9 system** can increase the amylopectin content of maize grains to almost 100%.

A **low-gluten non-transgenic wheat line** was developed by knocking out the most conserved domains of the α-gliadin family members, which decreased the genetically predisposed intestinal immune response. Using an HDR-based method, Dahan et al. edited the **CRTISO gene in tomato**, which significantly **increased the carotene content.**

Knock out **GmFATB1 gene** (encodes a fatty acyl carrier protein thioesterase) by CRISPR/Cas9 can significantly decrease the abundance of two saturated fatty acids in soybean mutants.

Increase Disease Resistance:

Table: Examples of genes from various crops that have been modified by the CRISPR/Cas9 system to increase disease resistance.

Crop Name	Gene Name	Gene Function	Editing Methods	Mutant Features
barley	*MLO*	reduced resistance to powdery mildew	knockout	improved resistance to powdery mildew
wheat	*TaMLO-A1*	reduced resistance to powdery mildew	knockout	improved resistance to powdery mildew
tomato	*SlMLO1*	reduced resistance to powdery mildew	knockout	improved resistance to powdery mildew
tomato	*MLO1*	reduced resistance to powdery mildew	knockout	improved resistance to powdery mildew
tomato	*EDR1*	encoded MAPKKK protein kinase	knockout	improved resistance to powdery mildew
rice	*OsERF922*	involved in the modulation of multiple stress tolerance	knockout	enhancing blast resistance
rice	*OsSEC3A*	interacted with rice SNAP25-type SNARE protein OsSNAP32 and phosphatidylinositol-3-phosphate	knockout	enhanced resistance to the fungal pathogen Magnaporthe oryzae
grape	*VvWRKY52*	play roles in biotic stress responses	knockout	increased the resistance to Botrytis cinerea
rice	*SWEET11, SWEET1113, SWEET1114*	transporter genes required for disease susce	knockout	increased broad spectrum resistance to different physiological races of Xoo
grape	*CsLOB1*	a critical citrus disease susceptibility gene	editing the PthA4 effector binding element	increased canker-resistance
grape	*CsLOB1*	a critical citrus disease susceptibility gene	remove the effector binding region of CsLOB1	enhanced resistance to Xcc
apple	*DIPM-1, DIPM-2, DIPM-4*	disease susceptibility genes	knockout	increased resistance to fire blight disease
potatoes	*StDND1, StCHL1, StDMR6-1*	disease susceptibility genes	knockout	increased resistance against late blight
chili pepper	*CaERF28*	disease susceptibility genes	knockout	increased anthracnose resistance

Improvement of Crop Herbicide Tolerance:

The use of CRISPR/Cas technology can accelerate the creation of crops tolerant to multiple herbicides. Several herbicide-tolerant materials have been created by using CRISPR/Cas to edit the **ALS gene** in model plants, rice, wheat, and other crops. Studies of naturally occurring point mutations in the A. thaliana ALS gene suggest that substituting specific bases in ALS (Acetolactate synthase) may lead to herbicide tolerance. The **wheat TaALS gene** can be edited by CBE to produce wheat mutant lines able to grow after being treated with herbicides

Improvement of Crop Abiotic Stress Tolerance:

Abiotic stresses, such as drought, salinity, high temperatures, and soil pollution, severely affect crop growth and greatly hinder efforts to increase crop yield and quality. A mutation to the **maize ZmSRL5 gene**, which is associated with the formation of the maize cuticle wax structure, can enhance **maize drought tolerance**. Editing the promoter region of ZmARGOS8, which encodes a negative regulator of the maize response to ethylene, can also positively affect drought tolerance. Zhou et al. used CRISPR/Cas9 technology to decrease the sensitivity of the elite rice restorer line **'Hua Zhan'** to abscisic acid and minimize the leaf water loss rate, leading to increased tolerance to drought, high temperatures, and osmotic stress. Kumar et al. mutated the drought tolerance-related gene in indica rice using CRISPR/Cas9; the mutants were moderately tolerant to osmotic stress and highly tolerant to salinity stress at the seedling stage, implying their method was appropriate for **improving the drought and salinity tolerance of indica rice varieties**.

Improvement of Other Crop Traits:

In addition to stress resistance/tolerance as well as traits related to yield and quality, CRISPR/Cas9 technology has also been used to improve other crop traits (e.g., fertility), leading to the development of novel plant types and haploid materials.

The creation of male-sterile materials is extremely important for hybrid seed production. Researchers have conducted a series of investigations regarding the editing of pollen fertility genes using CRISPR/Cas9 technology. Li et al. used the T. aestivum TaU3 RNA polymerase III U3 promoter to drive the optimized CRISPR/Cas9 vector.

Three homologous alleles encoding the wheat redox enzyme **NO POLLEN 1 (NP1)** were edited to produce fully **male-sterile wheat mutants**. Chen et al. constructed a CRISPR/Cas9 vector to delete **Male sterility gene 8 (MS8)** in maize. Li et al. edited the carbon starvation gene CSA in the pollen grains of **rice variety 'Kongyu 131'** and reported that the csa mutant exhibited a male-sterile phenotype under short-day conditions and a male-fertile phenotype under long-day conditions.

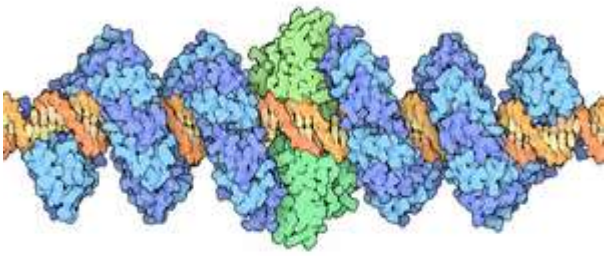
Hu et al. used the CRISPR/Cas9 system to edit the **semi-dwarf gene SD1**, which decreased the plant height of the sd1 mutant by about 25%. Li et al. used CRISPR/Cas9 technology to edit the upright **panicle-type gene DEP1** and the ideal plant-type gene **IPA1 in the 'Zhonghua 11' rice variety**. The dep1 mutant was characterized by an upright plant and compact panicle phenotype. The number of tillers either increased or decreased in the ipa1 mutant, reflecting the two extreme phenotypes induced by the mutation.

TALEN:

Transcription activator-like effector nucleases (TALEN) are restriction enzymes that can be engineered to cut specific sequences of DNA.

They are made by fusing a

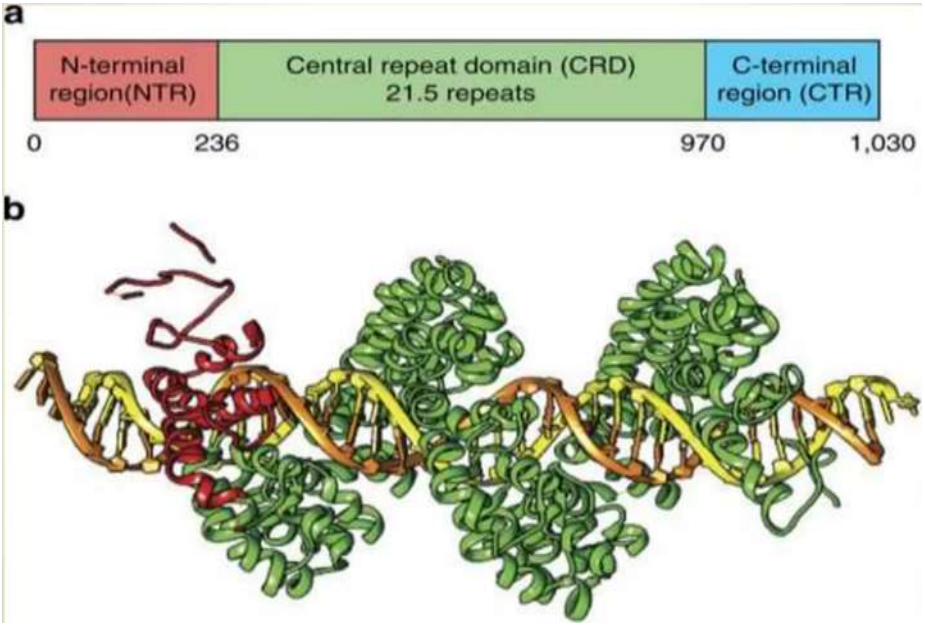
1. DNA-binding domain (TAL effector)
2. a DNA cleavage domain (a nuclease which cuts DNA strands)

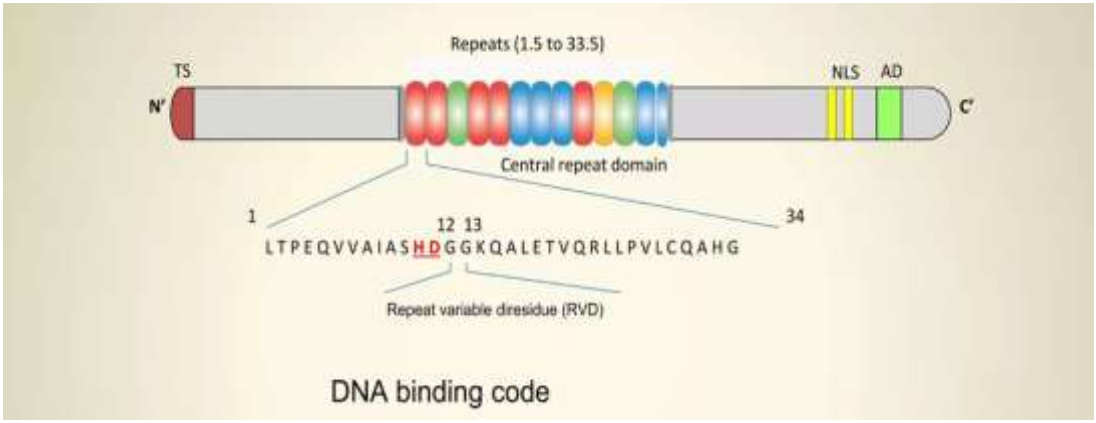


Transcription activator-like effectors (TALEs) can be engineered to bind to practically any desired DNA sequence, so when combined with a nuclease, DNA can be cut at specific locations. The restriction enzymes can be introduced into cells, for use in gene editing or for genome editing in situ, a technique known as genome editing with engineered nucleases. Alongside zinc finger nucleases and CRISPR/Cas9, TALEN is a prominent tool in the field of genome editing.

TALE DNA-binding domain

TAL effectors are proteins that are secreted by Xanthomonas bacteria via their type III secretion system when they infect plants. The DNA binding domain contains a repeated highly conserved 33–34 amino acid sequence with divergent 12th and 13th amino acids. These two positions, referred to as the **Repeat Variable Diresidue (RVD)**, are highly variable and show a strong correlation with specific nucleotide recognition. This straightforward relationship between amino acid sequence and DNA recognition has allowed for the engineering of specific DNA-binding domains by selecting a combination of repeat segments containing the appropriate RVDs.





DNA cleavage domain

The non-specific DNA cleavage domain from the end of the **FokI endonuclease** can be used to construct hybrid nucleases that are active in a yeast assay. These reagents are also active in plant cells and in animal cells. Initial TALEN studies used the wild-type FokI cleavage domain, but some subsequent TALEN studies also used FokI cleavage domain variants with mutations designed to improve cleavage specificity and cleavage activity. The FokI domain functions as a dimer, requiring two constructs with unique DNA binding domains for sites in the target genome with proper orientation and spacing. Both the number of amino acid residues between the TALE DNA binding domain and the FokI cleavage domain and the number of bases between the two individual TALEN binding sites appear to be important parameters for achieving high levels of activity.

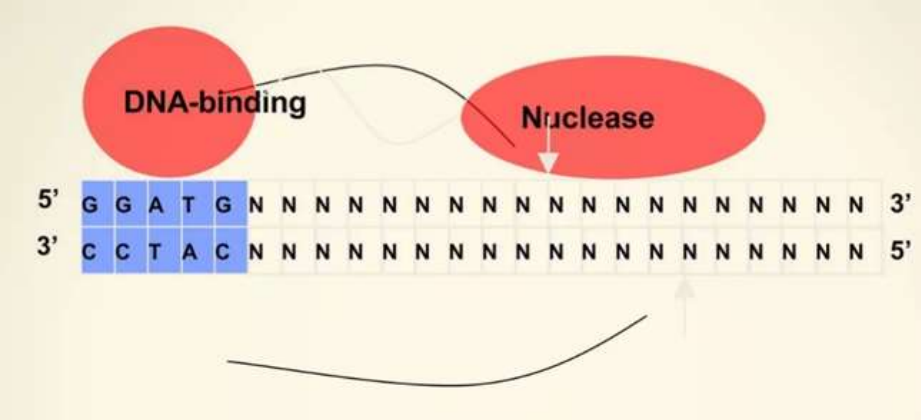


Fig.: FOK I Domain.

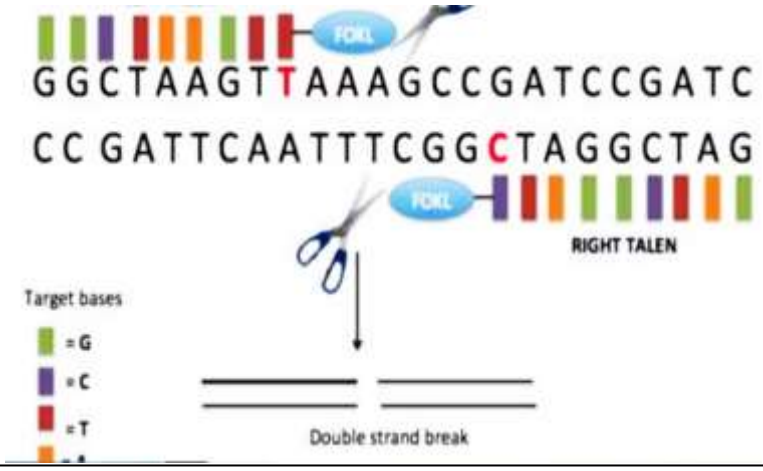


Fig.: Cleavage of DNA by FOK I.

TALEN constructs

The simple relationship between amino acid sequence and DNA recognition of the TALE binding domain allows for the efficient engineering of proteins. In this case, artificial gene synthesis is problematic because of improper annealing of the repetitive sequence found in the TALE binding domain.

One solution to this is to use a publicly available software program to calculate oligonucleotides suitable for assembly in a two-step PCR oligonucleotide assembly followed by whole gene amplification. A number of modular assembly schemes for generating engineered TALE constructs have also been reported. Both methods offer a systematic approach to engineering DNA binding domains that is conceptually similar to the modular assembly method for generating zinc finger DNA recognition domains.

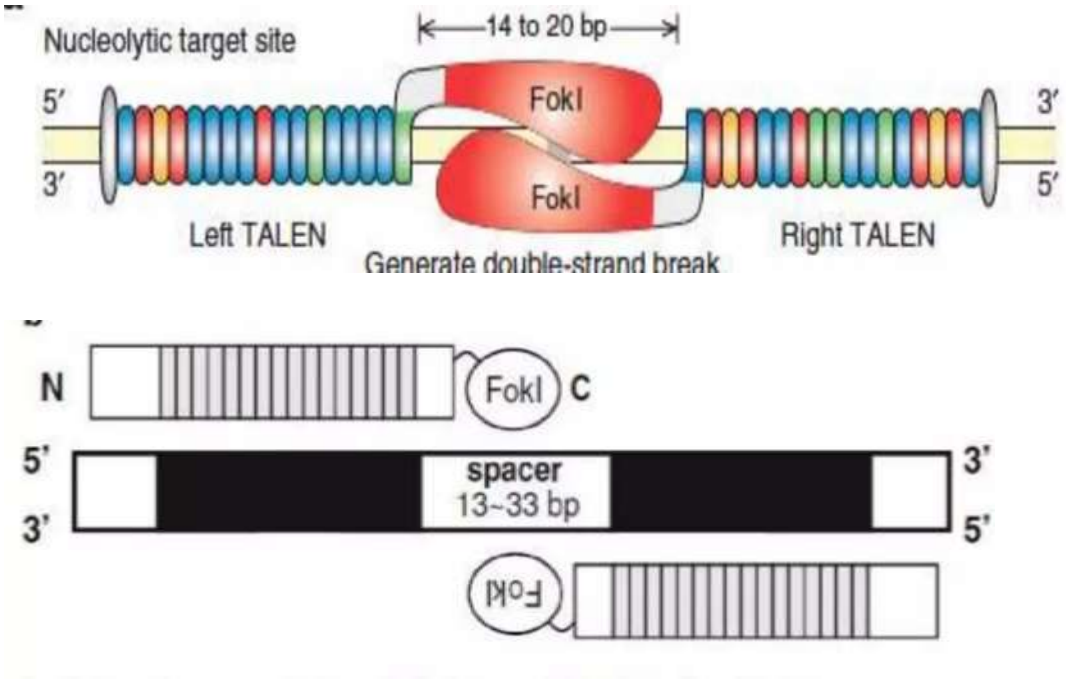


Fig.: Structure of TALENs.

Transfection:

Once the TALEN constructs have been assembled, they are inserted into plasmids; the target cells are then transfected with the plasmids, and the gene products are expressed and enter the nucleus to access the genome. Alternatively, TALEN constructs can be delivered to the cells as mRNAs, which removes the possibility of genomic integration of

the TALEN-expressing protein. Using an mRNA vector can also dramatically increase the level of homology directed repair (HDR) and the success of introgression during gene editing.

Genome editing Mechanism:

By inducing DNA double-strand break (DSB) following binding of TALEN to the target site, two DNA repair pathways are induces: 1) HDR and 2) non-homologous end joining (NHEJ)

Non-homologous end joining (NHEJ):

Non-homologous end joining (NHEJ) directly ligates DNA from either side of a double-strand break where there is very little or no sequence overlap for annealing. This repair mechanism induces errors in the genome via indels (insertion or deletion), or chromosomal rearrangement; any such errors may render the gene products coded at that location non-functional. Because this activity can vary depending on the species, cell type, target gene, and nuclease used, it should be monitored when designing new systems. A simple heteroduplex cleavage assay can be run which detects any difference between two alleles amplified by PCR. Cleavage products can be visualized on simple agarose gels or slab gel systems.

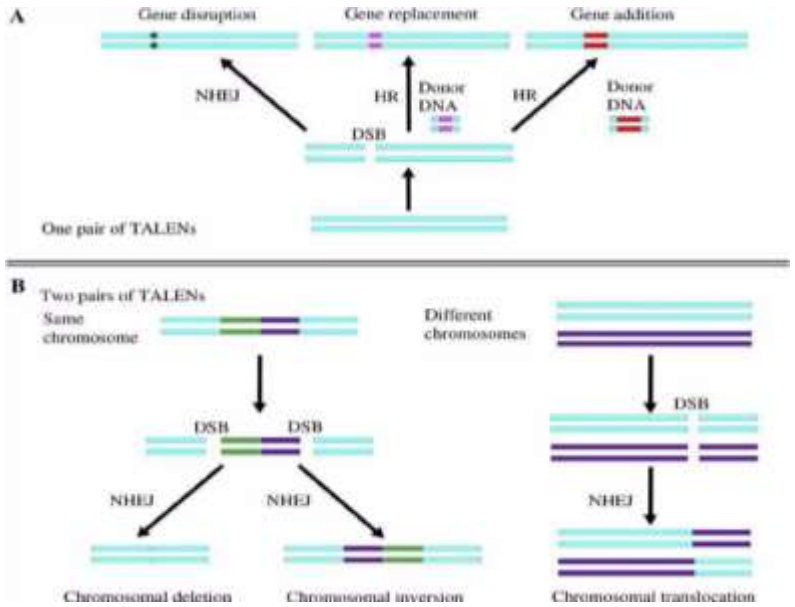


Fig.: Genome editing Mechanism.

Alternatively, DNA can be introduced into a genome through NHEJ in the presence of exogenous double-stranded DNA fragments.

Homology directed repair:

Homology directed repair can also introduce foreign DNA at the DSB as the transfected double-stranded sequences are used as templates for the repair enzymes.

Applications of TALENs in Crop Improvement:

TALENs can be applied to help protect plants from the effects of climate change. Genes that play a role in resistance to pests, diseases, or protection from harsh environmental conditions, such as drought and salinity, can be edited to enhance resilience. TALENS have also been used to improve the quality of products derived from crops.

Soybean:

Soybean lines with low levels of polyunsaturated fats were developed using TALENs by introducing stacked mutations in two fatty acid desaturase 2 genes (**FAD2-1A and FAD2-1B**) which confer changes in fatty acid desaturase 3A (**FAD3A**). This led to oleic acid levels of over 80% and linoleic acid levels under 4%. Now gene-edited soybean plants produce premium quality high-oleic soybean oil sold as Calyno by Calyxt. It is the first commercialized product from a gene-edited plant.

Rice:

TALENs have been used in rice to engineer resistance to the destructive disease, bacterial blight, caused by Xanthomonas oryzae. The rice gene **OsSWEET14** encodes an essential function for Xanthomonas to infect and cause bacterial blight. By editing the regulatory region of the gene, researchers generated heritable disease resistance.

Additionally, non-aromatic rice varieties were transformed into aromatic rice varieties using TALENs. TALENs were engineered to target and disrupt the OsBADH2 gene, resulting in the production of the major fragrance compound 2-acetyl-1-pyrrolin.

Potato:

The commercial potato variety, Ranger Russet, suffers from browning, bitter taste, and high levels of potentially carcinogenic acrylamides. TALENs were used to knockout the

vacuolar invertase gene (Vlnv) producing improved properties. The amounts of reducing sugars were undetectable, and chips produced from these potatoes had a reduced level of acrylamide and a more desirable light colour.

To reduce cholesterols and related toxic steroidal glycoalkaloids, TALENs targeting the **SSR2 gene** were delivered into potato using a transient Agrobacterium transformation method. In the resulting potato lines, the SSR2 gene was inactivated and no TALEN transgene remained in the plant

Wheat:

Scientists attempted to engineer resistance to powdery mildew, a widespread disease, by testing both TALENS and CRISPR-Cas9 on the **MILDEW-RESISTANCE LOCUS (mlo) gene**. They succeeded in making TALEN-induced mutations in all six sets of mlo genes, resulting in plants with essentially complete resistance to the wheat powdery mildew pathogen.

Maize:

To develop TALEN tools for corn (maize), scientists made stable, heritable mutations at a visible marker, the **glossy2 (gl2)** locus. Transgenic lines containing mutations were obtained and three of the novel alleles were able to confer a glossy phenotype. The use of TALENs in maize is a powerful tool for genome mutagenesis, discovery of gene function, and trait improvement.

Plant species	Gene name	Modification	TALEN construction	Mutation	Reference type platform
Arabidopsis	ADH1	NHEJ	Golden Gate	Gene knockout	Cermak et al., 2011
Tobacco	EBE of Hax3	NHEJ	Not available	Gene knockout	Mahfouz et al., 2011
Tobacco	SurA, SurB	NHEJ, HR	Golden Gate	Gene knockout, insertion and replacement	Zhang et al., 2013
Rice	EBE of AvrXa7 and PthXo3	NHEJ	Golden Gate	Gene knockout	Li et al., 2012b
Rice	OsDEP1	NHEJ	Golden Gate	Gene knockout, large deletion, inversion	Shan et al., 2013
Rice	OsBADH2	NHEJ	Golden Gate	Gene knockout	Shan et al., 2013
Rice	OsCKX2	NHEJ	Golden Gate	Gene knockout	Shan et al., 2013
Rice	OsSD1	NHEJ	Golden Gate	Gene knockout	Shan et al., 2013
Brachypodium	BdABA1	NHEJ	Golden Gate	Gene knockout	Shan et al., 2013
Brachypodium	BdHTA1	NHEJ	Golden Gate	Gene knockout	Shan et al., 2013

LEAPER (Leveraging endogenous ADAR for programmable editing of RNA):

It is a genetic engineering technique in molecular biology by which RNA can be edited. The technique relies on engineered strands of RNA to recruit native ADAR enzymes to swap out different compounds in RNA. Developed by researchers at Peking University in 2019, the technique, some have claimed, is more efficient than the CRISPR gene editing technique. Initial studies have claimed that editing efficiencies of up to 80%.

LEAPER targets editing messenger RNA (mRNA) for the same gene which is transcribed into a protein. Post-transcriptional RNA modification typically involves the strategy of converting adenosine-to-inosine (A-to-I) since inosine (I) demonstrably mimics guanosine (G) during translation into a protein. A-to-I editing is catalyzed by adenosine deaminase acting on RNA (ADAR) enzymes, whose substrates are double-stranded RNAs. Three human ADAR genes have been identified with ADAR1 (official symbol ADAR) and ADAR2 (ADARB1) proteins developed activity profiles. LEAPER achieves this targeted RNA editing through the use of short engineered ADAR-recruiting RNAs (arRNAs). arRNAs consist of endogenous ADAR1 proteins with several RNA binding domains (RBDs) fused with a peptide, CRISPR-Cas13b protein, and a guide RNA (gRNA) between 100 and 150 nt in length for high editing efficiency designed to recruit the chimeric ADAR protein to a target site.

This results in a change in which protein is synthesized during translation.

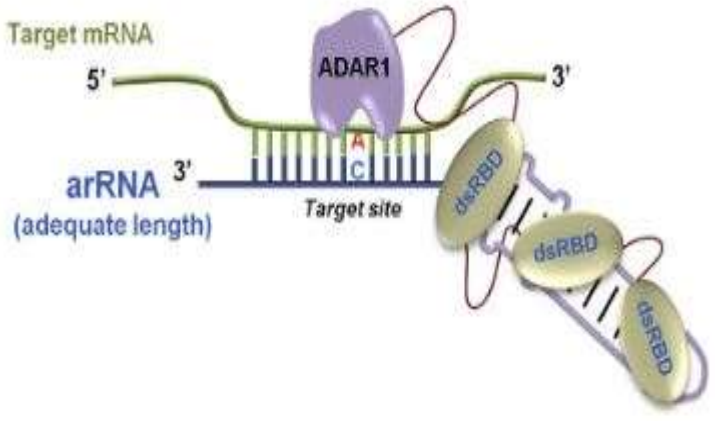


Fig.: LEAPER-mediated RNA Editing.

Applications:

Chinese researchers have utilized LEAPER to restore functional enzyme activity in cells from patients with Hurler syndrome. They have claimed that LEAPER could have the potential to treat almost half of all known hereditary disorders.

Highly specific editing efficiencies of up to 80% can be achieved when LEAPER editing using arRNA151 is delivered via a plasmid or viral vector or as a synthetic oligonucleotide, though this efficiency varied significantly across cell types. Based on these preliminary results, LEAPER may have the most therapeutic promise with no production of functional protein but if a partial restoration of protein expression would provide therapeutic benefit. For example, in human cells with defective α-L-iduronidase (IDUA) expression in cells from patients with IDUA-defective Hurler syndrome, LEAPER resulted in a W53X truncation mutant of p53 being edited using arRNA151 to achieve a "normal" p53 translation and functional p53-mediated transcriptional responses.

---

10. Proteomics: Concept of proteome; Functional, structural and differential proteomics; Principle of 2D gel electrophoresis (2-DE); advantages and limitations of 2-DE; Protein Fingerprinting; Gel free proteomics (iTRAQ); Mass spectrometry (MALDI-TOF MS); Posttranslational modifications of proteins; Applications of proteomics in agriculture.

Basic idea of Proteomics:

Proteomics is the large-scale study of proteins. Proteins are vital parts of living organisms, with many functions. The term proteomics was coined in 1997 in analogy with genomics, the study of the genome. The word proteome is a portmanteau of protein and genome, and was coined by Marc Wilkins in 1994 while he was a PhD student at Macquarie University. Macquarie University also founded the first dedicated

proteomics laboratory in 1995.

The proteome is the entire set of proteins that are produced or modified by an organism or system. This varies with time and distinct requirements, or stresses, that a cell or organism undergoes. Proteomics is an interdisciplinary domain that has benefitted greatly from the genetic information of various genome projects, including the Human Genome Project. It covers the exploration of proteomes from the overall level of protein composition, structure, and activity. It is an important component of functional genomics.

Proteomics generally refers to the large-scale experimental analysis of proteins and proteomes but is often specifically used to refer to protein purification and mass spectrometry.The entire protein component of a given organism is called 'proteome', the term coined by Wasinger in 1995. A proteome is a quantitatively expressed protein of a genome that provides information on the gene products that are translated, amount of products and any post translational modifications.

Proteomics is an emerging area of research in the post-genomic era, which involves identifying the structures and functions of all proteins of a proteome. It is sometimes also treated as structural based functional genomics.

Complexity of the problem

After genomics and transcriptomics, proteomics is the next step in the study of biological systems. It is more complicated than genomics because an organism's genome is more or less constant, whereas the proteome differs from cell to cell and from time to time. Distinct genes are expressed in different cell types, which means that even the basic set of proteins that are produced in a cell needs to be identified.

In the past this phenomenon was done by RNA analysis, but it was found not to correlate with protein content. It is now known that mRNA is not always translated into protein, and the amount of protein produced for a given amount of mRNA depends on the gene it is transcribed from and on the current physiological state of the cell. Proteomics confirms the presence of the protein and provides a direct measure of the quantity present.

Post-translational modifications

Not only does the translation from mRNA cause differences, but many proteins are also subjected to a wide variety of chemical modifications after translation. Many of these post-translational modifications are critical to the protein's function.

- **Phosphorylation**

One such modification is phosphorylation, which happens to many enzymes and structural proteins in the process of cell signaling. The addition of a phosphate to particular amino acids—most commonly serine and threonine[12] mediated by serine/threonine kinases, or more rarely tyrosine mediated by tyrosine kinases—causes a protein to become a target for binding or interacting with a distinct set of other proteins that recognize the phosphorylated domain.

Because protein phosphorylation is one of the most-studied protein modifications, many "proteomic" efforts are geared to determining the set of phosphorylated proteins in a particular cell or tissue-type under particular circumstances. This alerts the scientist to the signaling pathways that may be active in that instance.

- **Ubiquitination**

Ubiquitin is a small protein that can be affixed to certain protein substrates by enzymes called E3 ubiquitin ligases. Determining which proteins are poly-ubiquitinated helps understand how protein pathways are regulated. This is, therefore, an additional legitimate "proteomic" study. Similarly, once a researcher determines which substrates are ubiquitinated by each ligase, determining the set of ligases expressed in a particular cell type is helpful.

- **Additional modifications**

In addition to phosphorylation and ubiquitination, proteins can be subjected to (among others) methylation, acetylation, glycosylation, oxidation and nitrosylation. Some proteins undergo all these modifications, often in time-dependent combinations. This illustrates the potential complexity of studying protein structure and function.

Types of Proteomics:

(i) Structural Proteomics:

One of the main targets of proteomics investigation is to map the structure of protein complexes or the proteins present in a specific cellular organelle known as cell map or structural proteins. Structural proteomics attempt to identify all the proteins within a protein complex and characterization all protein-protein interactions. Isolation of specific protein complex by purification can simplify the proteomic analysis.

(ii) Functional Proteomics:

It mainly includes isolation of protein complexes or the use of protein ligands to isolate specific types of proteins. It allows selected groups of proteins to be studied its chracteristics which can provide important information about protein signalling and disease mechanism etc.

(iii) Expression proteomics:

Expression proteomics includes the analysis of protein expression at larger scale. It helps identify main proteins in a particular sample, and those proteins differentially expressed in related samples— such as diseased vs. healthy tissue. If a protein is found only in a diseased sample then it can be a useful drug target or diagnostic marker. Proteins with same or similar expression profiles may also be functionally related. There are technologies such as 2D-PAGE and mass spectrometry that are used in expression proteomics.

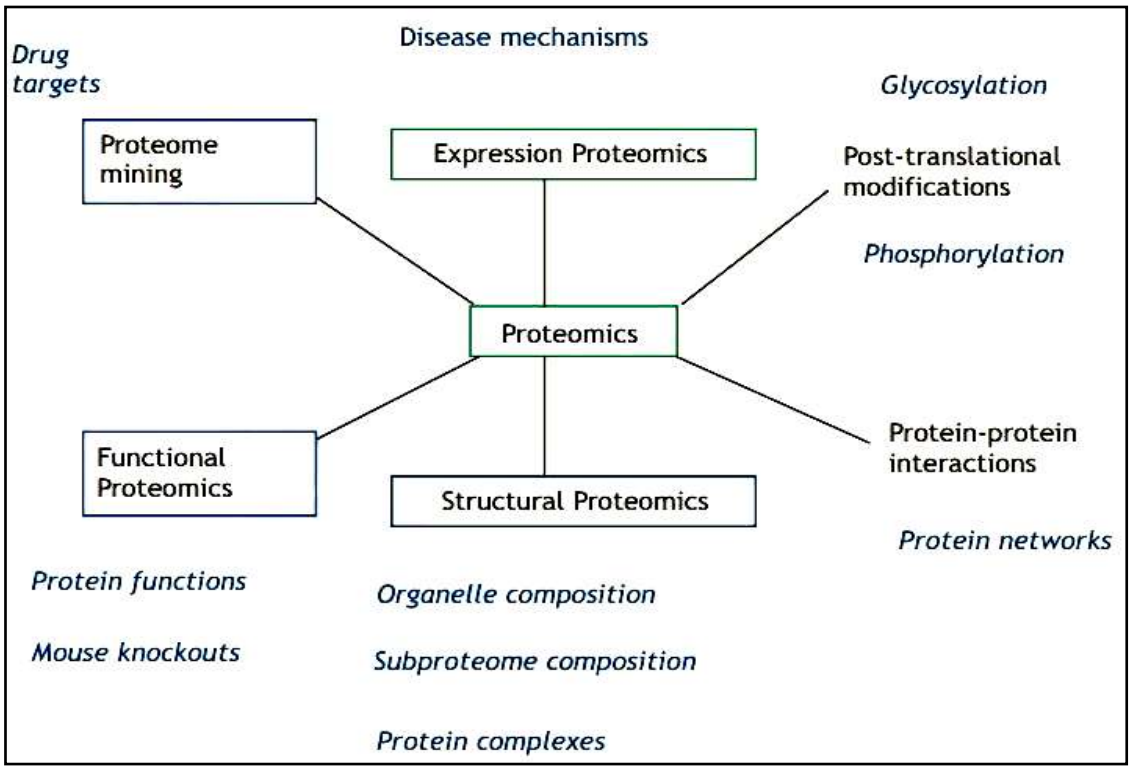


Fig.: Types of Proteomics and the scientific knowledge that can be gained from them.

Significance of Proteomics:

(i) Protein profiling

Bioinformatics has been widely employed in protein-profiling, where question of protein structural information for the purpose of protein identification, characterization and database is carried out. The spectrum of protein expressed in a cell type provides the cell with its unique identity. It explores how the protein complement changes in a cell type during development in response to environmental stress.

(ii) Protein arrays

Protein microarrays facilitate the detection of protein protein interaction and protein expression profiling. Several protein microarray examples indicate that protein arrays hold great promise for the global analysis of protein-protein and protein-ligand interaction.

(iii) Proteomics to a phosphorylation

In post-translational modification of protein, mass spectrometer (MS) can be used to identify novel phosphorylation. Measure changes in phosphorylation state of protein takes place in response to an effective and determining phosphorylation sites in proteins.

Identification of phosphorylation sites can provide information about the mechanism of enzyme regulation and protein kinase and phosphotases involved. A proteomics approach for this process has an advantage that one can study all the phosphorylating proteins in a cell at the same time.

(iv) Proteome mining

Proteome mining is a functional proteomic approach used to extract information from the analysis of specific sub-proteomics. In principle, it is based on the assumption. In principle, it is based oil the assumption that all drug like molecule selectively compete with a natural cellular ligand for a binding site on a protein target.

Two-dimensional gel electrophoresis

Two-dimensional gel electrophoresis, abbreviated as 2-DE or 2-D electrophoresis, is a form of gel electrophoresis commonly used to analyze proteins. Mixtures of proteins are separated by two properties in two dimensions on 2D gels. 2-DE was first independently introduced by O'Farrell and Klose in 1975.

Basis for separation

2-D electrophoresis begins with electrophoresis in the first dimension and then separates the molecules perpendicularly from the first to create an electropherogram in the second dimension. In electrophoresis in the first dimension, molecules are separated linearly according to their isoelectric point. In the second dimension, the molecules are then separated at 90 degrees from the first electropherogram according to molecular mass. Since it is unlikely that two molecules will be similar in two distinct properties, molecules are more effectively separated in 2-D electrophoresis than in 1-D electrophoresis.

The two dimensions that proteins are separated into using this technique can be isoelectric point, protein complex mass in the native state, or protein mass.

Separation of the proteins by isoelectric point is called isoelectric focusing (IEF). Thereby, a gradient of pH is applied to a gel and an electric potential is applied across the gel, making one end more positive than the other. At all pH values other than their isoelectric point, proteins will be charged. If they are positively charged, they will be pulled towards the more negative end of the gel and if they are negatively charged they will be pulled to the more positive end of the gel. The proteins applied in the first dimension will move along the gel and will accumulate at their isoelectric point; that is, the point at which the overall charge on the protein is 0 (a neutral charge).

For the analysis of the functioning of proteins in a cell, the knowledge of their cooperation is essential. Most often proteins act together in complexes to be fully functional. The analysis of this sub organelle organisation of the cell requires techniques conserving the native state of the protein complexes. In native polyacrylamide gel electrophoresis (native PAGE), proteins remain in their native state and are separated in the electric field following their mass and the mass of their complexes respectively. To obtain a separation by size and not by net charge, as in IEF, an additional charge is transferred to the proteins by the use of Coomassie Brilliant Blue or lithium dodecyl sulfate. After completion of the first dimension the complexes are destroyed by applying the denaturing SDS-PAGE in the second dimension, where the proteins of which the complexes are composed of are separated by their mass.

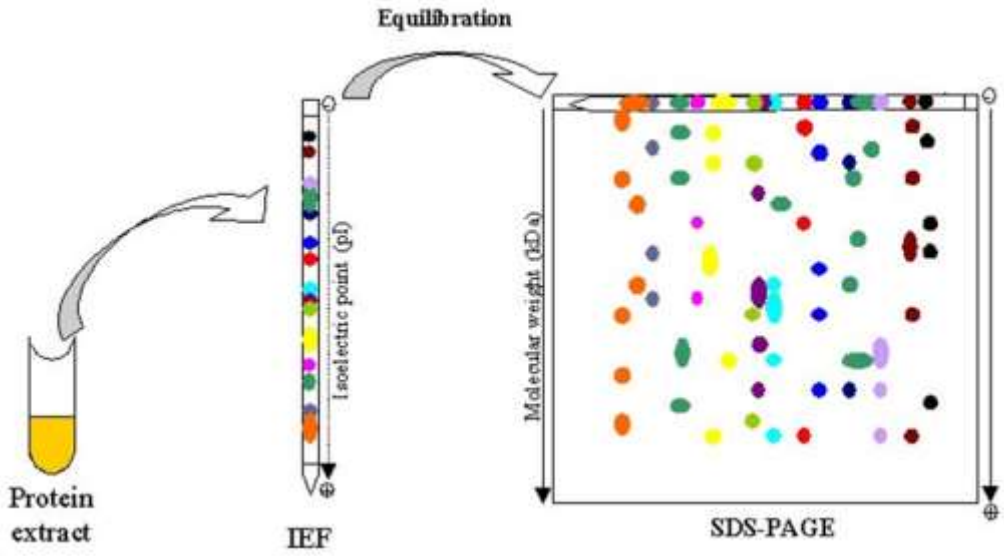


Fig.: Two-dimensional gel electrophoresis.

Before separating the proteins by mass, they are treated with sodium dodecyl sulfate (SDS) along with other reagents (SDS-PAGE in 1-D). This denatures the proteins (that is, it unfolds them into long, straight molecules) and binds a number of SDS molecules roughly proportional to the protein's length. Because a protein's length (when unfolded) is roughly proportional to its mass, this is equivalent to saying that it attaches a number of SDS molecules roughly proportional to the protein's mass. Since the SDS molecules are negatively charged, the result of this is that all of the proteins will have approximately the same mass-to-charge ratio as each other. In addition, proteins will not migrate when they have no charge (a result of the isoelectric focusing step) therefore the coating of the protein in SDS (negatively charged) allows migration of the proteins in the second dimension (SDS-PAGE, it is not compatible for use in the first dimension as it is charged and a nonionic or zwitterionic detergent needs to be used). In the second dimension, an electric potential is again applied, but at a 90 degree angle from the first field. The proteins will be attracted to the more positive side of the gel (because SDS is negatively charged) proportionally to their mass-to-charge ratio. As previously explained, this ratio will be nearly the same for all proteins. The proteins' progress will be slowed by frictional forces. The gel therefore acts like a molecular sieve when the current is applied, separating the proteins on the basis of their molecular weight with larger proteins being retained higher in the gel and smaller proteins being able to pass through the sieve and reach lower regions of the gel.

Advantages of 2-DE Gel:

- ✓ It is used to separate DNA fragments, RNA fragments, and proteins based on their charge and sizes.
- ✓ Gel electrophoresis is used to determine the size of the fragments and to visualize it.
- ✓ This technique is often used to analyze DNA samples that have gone through PCR techniques to ensure that the target DNA sequence was appropriately amplified.
- ✓ It starts with preparing the DNA, RNA, or protein samples to be used in the process. The gel solution would then be prepared by mixing agarose gel to a buffer solution.

- ✓ It is essential to understand that the percentage of the agarose gel present in the buffer will affect the separation ability of the electrophoresis.
- ✓ After the gel hardens, it will be placed in a gel chamber with a buffer solution to maintain an environment where the sample can be separated. The sample will be loaded to the wells in the gel, and the gel chamber will be supplied with an electrical force to enable the separation. The last step in the gel electrophoresis would be staining the gel with a dye to visualize and analyze the samples.
- ✓ 2D DIGE provides higher sensitivity, accuracy, and reproducibility. It offers a broader spectrum of protein detection and requires fewer gels, making it a cost-efficient choice.

Limitations of 2-DE Gel:

Some of the inherent caveats/disadvantages associated with 2DGE are as follows:

(i) many protein spots are likely comprised of multiple proteins;

(ii) individual proteins may migrate as multiple spots based on differential digestion;

(iii) labor-intensive image analysis requires gel matching

Detecting proteins

The result of this is a gel with proteins spread out on its surface. These proteins can then be detected by a variety of means, but the most commonly used stains are silver and Coomassie Brilliant Blue staining. In the former case, a silver colloid is applied to the gel. The silver binds to cysteine groups within the protein. The silver is darkened by exposure to ultra-violet light. The amount of silver can be related to the darkness, and therefore the amount of protein at a given location on the gel. This measurement can only give approximate amounts, but is adequate for most purposes. Silver staining is 100x more sensitive than Coomassie Brilliant Blue with a 40-fold range of linearity.

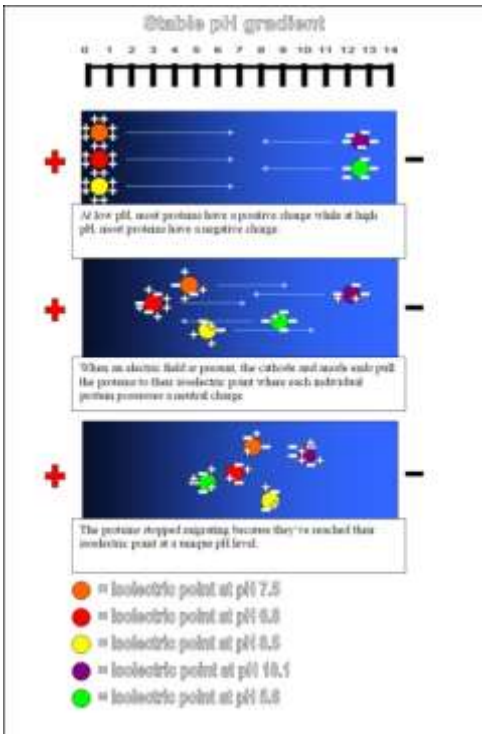
Molecules other than proteins can be separated by 2D electrophoresis. In supercoiling assays, coiled DNA is separated in the first dimension and denatured by a DNA intercalator (such as ethidium bromide or the less carcinogenic chloroquine) in the

second. This is comparable to the combination of native PAGE /SDS-PAGE in protein separation.

Common techniques

IPG-DALT

A common technique is to use an Immobilized pH gradient (IPG) in the first dimension. This technique is referred to as IPG-DALT. The sample is first separated onto IPG gel (which is commercially available) then the gel is cut into slices for each sample which is then equilibrated in SDS-mercaptoethanol and applied to an SDS-PAGE gel for resolution in the second dimension. Typically IPG-DALT is not used for quantification of proteins due to the loss of low molecular weight components during the transfer to the SDS-PAGE gel.



IEF SDS-PAGE

Isoelectric focusing (IEF), also known as electrofocusing, is a technique for separating different molecules by differences in their isoelectric point (pI). It is a type of zone

electrophoresis, usually performed on proteins in a gel, that takes advantage of the fact that overall charge on the molecule of interest is a function of the pH of its surroundings.

2D gel analysis software

In quantitative proteomics, these tools primarily analyze bio-markers by quantifying individual proteins, and showing the separation between one or more protein "spots" on a scanned image of a 2-DE gel. Additionally, these tools match spots between gels of similar samples to show, for example, proteomic differences between early and advanced stages of an illness. Software packages include Delta2D, ImageMaster, Melanie, PDQuest, Progenesis and REDFIN – among others.[citation needed] While this technology is widely utilized, the intelligence has not been perfected. For example, while PDQuest and Progenesis tend to agree on the quantification and analysis of well-defined well-separated protein spots, they deliver different results and analysis tendencies with less-defined less-separated spots.

Challenges for automatic software-based analysis include incompletely separated (overlapping) spots (less-defined and/or separated), weak spots / noise (e.g., "ghost spots"), running differences between gels (e.g., protein migrates to different positions on different gels), unmatched/undetected spots, leading to missing values, mismatched spots , errors in quantification (several distinct spots may be erroneously detected as a single spot by the software and/or parts of a spot may be excluded from quantification), and differences in software algorithms and therefore analysis tendencies.

Proteomics aims at the simultaneous analysis of all proteins expressed by a cell, tissue or organism in a specific physiological condition. Because proteins are the effector molecules in all organisms, it is evident that changes in the physiological condition of an organism will be reflected by changes in protein expression and/or processing. Since the formulation of the concept of proteomics in the mid 90's proteomics has relied heavily on 2 dimensional gel electrophoresis (2DGE) for the separation and visualization of proteins. 2DGE, however, has a number of inherent drawbacks. 2DGE is costly, fairly insensitive to low copy proteins and cannot be used for the entire proteome. Therefore, over the years, several gel-free proteomics techniques have been

developed to either fill the gaps left by 2DGE or to entirely abolish the gel based techniques.

Methods of Protein Identification by Mass Spectrometry

Two methods are widely used for protein identification by mass spectrometry- MALDI-TOF based protein fingerprinting and LC-MS/MS based peptide sequencing.

In MALDI-TOF based protein fingerprinting method, a sample is digested with certain proteolytic enzyme (usually trypsin) and one MS spectrum is acquired which generates the masses of all peptides, and these masses are used as a fingerprint to search proteins in a database.

In contrast, protein identification by LC-MS/MS peptide sequencing method follows different principle. Typically, the peptide mixture from a proteolytic digestion (usually trypsin digestion) is separated by a HPLC. A Tandem mass spectrometer is coupled on-line with HPLC, so the peptides eluted from HPLC column is fragmented by a process called collision-induced dissociation (CID) and a MS/MS spectra is acquired for each fragmented peptide (there are often several thousands of MS/MS spectra from each sample). Each MS/MS spectrum (corresponding to a specific peptide sequence) is used to search protein database for matched peptides.

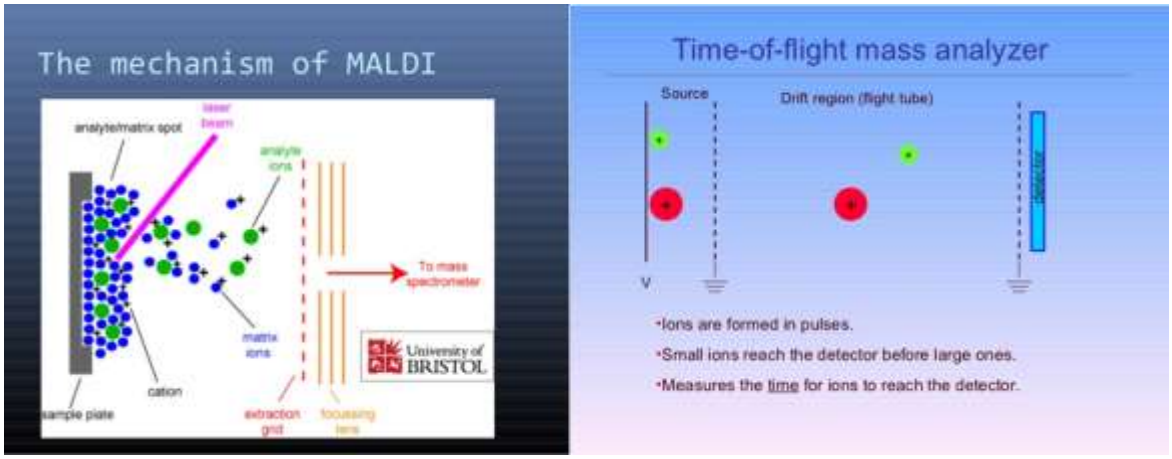
Protein identification based on MALDI-TOF MS fingerprinting has some advantages over the Nano-LC-MS/MS based method. The first advantage is its speed. Acquiring a MALDI-TOF MS data often takes less than one minute. Subsequent database search often takes less than three minutes, and analysis of search result is often rapid and straightforward. In contrast, Nano LC-MS/MS data acquisition typically takes 1-2 hours with several thousands of MS/MS spectra generated, and subsequent database search may take several hours, and analyzing search result could take even longer time. Secondly, operating a MALDI-TOF mass spectrometer is a much easier task than operating a Nano LC-MS/MS system.

However, there are two major factors that make the Nano LC-MS/MS a superior platform for protein identification: sensitivity and reliability. The high sensitivity of Nano LC-MS/MS is achieved mainly by 3 factors: first, the use of Nano LC with 75 um ID reverse phase C18 column can effectively concentrate peptides (50-200 fold) before MS

detection. Since mass spectrometry detection is concentration-dependent, such concentration increase is effectively translated into the increase of MS signal. Consequently, it significantly increases the sensitivity in identifying proteins. Second, protein identification by LC-MS/MS is based on independent sequencing of peptides. It is often possible to confidently identify a protein based on MS/MS sequencing of only one peptide, in contrast to 10-20 peptides required for MALDI-TOF fingerprinting. Since number of peptides observed from a protease digestion is often related to amount of sample for a giving protein, the ability to make a positive protein identification based on much less peptides is translated into the real sensitivity gain. Third, during MALDI based MS analysis, only very small percentage (typically 1/1000 to 1/10000) of sample loaded on a MALDI plate are utilized, while almost all the sample are utilized during electrospray process in LC-MS/MS. Another important advantage of Nano LC-MS in protein identification is its reliability. First, the outcome from a MALDI-TOF fingerprinting is a list of candidates each with a ranking score. In many cases scores of many protein candidates are so close that hit is picked up based on a range of constrains such as MW, pI, and species. Although computer database search with some MALDI-TOF MS data can always generate a list of "hits", it is often very difficult evaluate the fidelity of these hits. In contrast, identification of a protein based on one peptide sequenced by MS/MS with 70%-80% sequence coverage (a typical case) is at about 90% confident level. The confident level is about 99% if the protein identification is based on MS/MS sequencing of two independent peptides. It is usually 100% certain is three peptides are sequenced. Majority of the protein identification carried out at ProtTech is based on large than 10 peptides sequenced. Second, if a gel band contains two or more proteins, although database search will still generate a hit list, the reliability of protein identification based on MALDI-TOF fingerprinting will become a very serious issue. Unfortunately, roughly 80% of all the samples we have analyzed contain two or more proteins. In contrast, nano LC-MS/MS is able to analyze very complex mixture since each peptide is independently sequenced. Theoretically a LC-MS/MS can identify a thousand protein mixtures with the same reliability as that in identifying one protein, but such capacity is limited by one-dimensional HPLC, which is often difficult to separate a very large number of peptides.

Principles of MALDI-TOF Mass Spectrometry:

MALDI is the abbreviation for "Matrix Assisted Laser Desorption/Ionization." The sample for MALDI is uniformly mixed in a large quantity of matrix. The matrix absorbs the ultraviolet light (nitrogen laser light, wavelength 337 nm) and converts it to heat energy. A small part of the matrix heats rapidly (in nano seconds) and is vaporized, together with the sample.



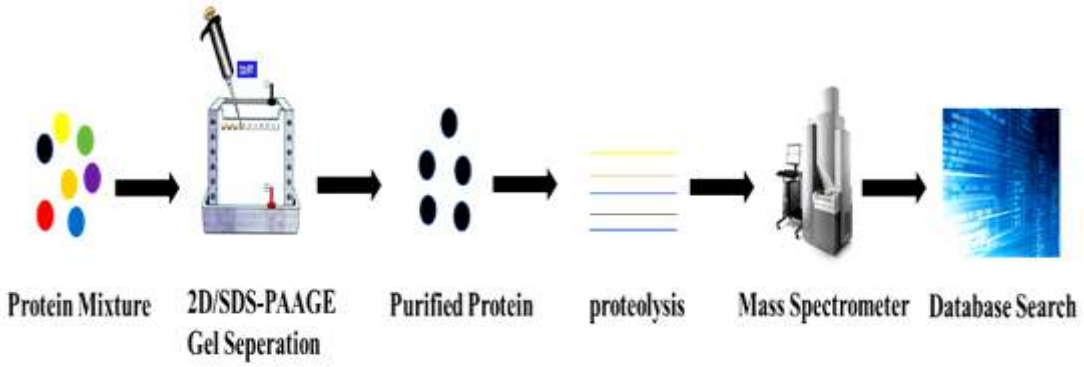
TOF MS is the abbreviation for Time of Flight Mass Spectrometry. Charged ions of various sizes are generated on the sample slide. A potential difference V_0 between the sample slide and ground attracts the ions in the direction shown in the diagram. The velocity of the attracted ions v is determined by the law of conservation of energy. As the potential difference V_0 is constant with respect to all ions, ions with smaller mass-to-charge ratio (m/z) value (lighter ions) and more highly charged ions move faster through the drift space until they reach the detector. Consequently, the time of ion flight differs according to the mass-to-charge ratio (m/z) value of the ion. The method of mass spectrometry that exploits this phenomenon is called Time of Flight Mass Spectrometry.

Protein Fingerprinting:

Peptide mass fingerprinting (PMF), also known as protein fingerprinting, is a high-throughput analytical method that developed in 1933 to identify proteins. Endoproteases first cleaves the unknown target protein into smaller peptides. The absolute mass of these peptides can be accurately measured using a mass spectrometer (such as MALDI-TOF), and a list of peptide peaks of the unknown protein is also obtained. This list of peaks is compared to a list of theoretical peptide peaks from a protein database using a computer program. This program translates all organism's known genomes into proteins, theoretically digests the proteins into small peptides, and calculates the absolute mass of the peptide from each protein. At the same time, the mass of the peptide from the lysed unknown protein is also compared with the theoretical peptide mass of each protein encoded in the genome. The results are statistically analyzed to find the best match.

Process of Protein Fingerprinting or Peptide Mass Fingerprinting:

After the proteins in biological tissues or cells are purified, they are separated by 2D or SDS-PAGE gel and compared using imaging techniques. Proteins with significant differences are selected for further analysis. A series of peptide compounds that reflect the characteristics of the protein are obtained by degumming and enzymatic hydrolysis of the selected protein. The fingerprint of the peptide mixture is accurately measured by a mass spectrometer to obtain the molecular weight of the protein. Searching through a known database provides quasi-determinism in the selected protein.



Applications of Protein Fingerprinting:

With the development of analytical chemistry technology, PMF has been widely used in various fields such as drug authenticity identification, food quality control and human disease diagnosis due to its high sensitivity, high stability and low sample loss. The World Health Organization and the State Food and Drug Administration of China use PMF as a material identification and process control technology to provide objective indicators for quality evaluation.

Advantages of Protein Fingerprinting

- ✓ Does not rely on protein sequencing for protein identification, but the peptide quality.
- ✓ Simple and fast, a classic method for the identification of proteins by first-order mass spectrometry

Disadvantages of Protein Fingerprinting:

- The protein sequence of interest must exist in the target database.
- Good for single protein analysis. Presence of mixed proteins greatly increases the complexity of the analysis.
- Peptides of similar quality will greatly increase the difficulty of matching.

MALDI-TOF Protein Fingerprinting:

Rapid and accurate protein identification is essential in proteomics research. PMF can be measured by matrix-assisted laser desorption ionization ion source (MALDI) and time-of-flight (TOF) mass analyzer. It is the most commonly used method in proteomics research.

MALDI uses a laser to irradiate a co-crystal thin film formed by a sample and a substrate. The substrate absorbs energy from the laser and transfers it to biomolecules. In the ionization process, the protons are transferred to or obtained from the biomolecules, and the biomolecules are ionized. Therefore, it is a soft ionization technique suitable for the determination of mixtures and biological macromolecules. The principle of TOF is that ions are accelerated to fly through the flight duct under the

action of an electric field, and the ions are detected based on the time of flight to the detector. MALDI-TOF-MS has the characteristics of high sensitivity, high accuracy, and high resolution. It provides an increasingly important role for sample analysis in life sciences and many other fields.

Post-translational modifications of proteins:

In molecular biology, **post-translational modification** (**PTM**) is the covalent process of changing proteins following protein biosynthesis. PTMs may involve enzymes or occur spontaneously. Proteins are created by ribosomes, which translate mRNA into polypeptide chains, which may then change to form the mature protein product. PTMs are important components in cell signalling, as for example when prohormones are converted to hormones.

Post-translational modifications can occur on the amino acid side chains or at the protein's C-orN-termini. They can expand the chemical set of the 22 amino acids by changing an existing functional group or adding a new one such as phosphate. Phosphorylation is highly effective for controlling the enzyme activity and is the most common change after translation. Many eukaryotic and prokaryotic proteins also have carbohydrate molecules attached to them in a process called glycosylation, which can promote protein folding and improve stability as well as serving regulatory functions. Attachment of lipid molecules, known as lipidation, often targets a protein or part of a protein attached to the cell membrane.

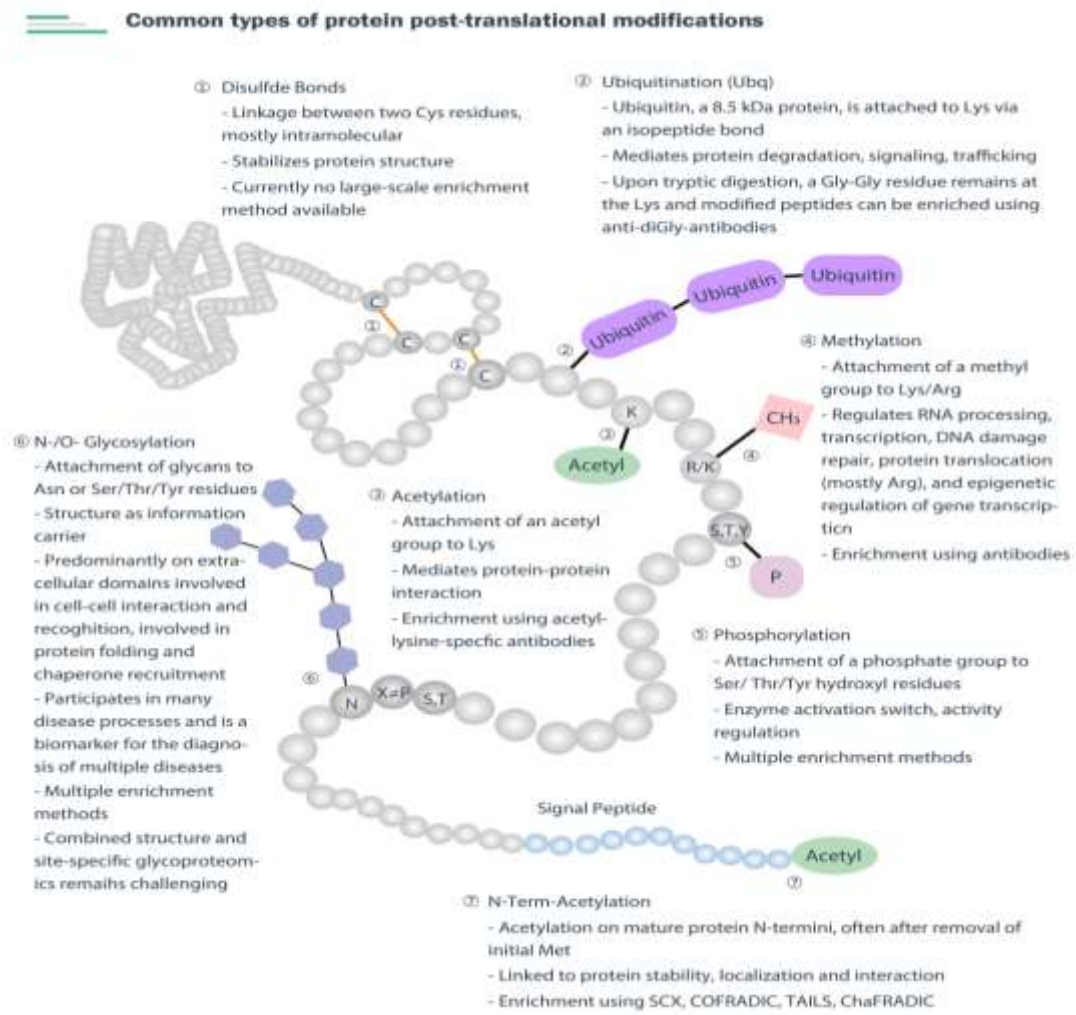
Other forms of post-translational modification consist of cleaving peptide bonds, as in processing a propeptide to a mature form or removing the initiator methionine residue. The formation of disulfide bonds from cysteine residues may also be referred to as a post-translational modification. For instance, the peptide hormone insulin is cut twice after disulfide bonds are formed, and a propeptide is removed from the middle of the chain; the resulting protein consists of two polypeptide chains connected by disulfide bonds.

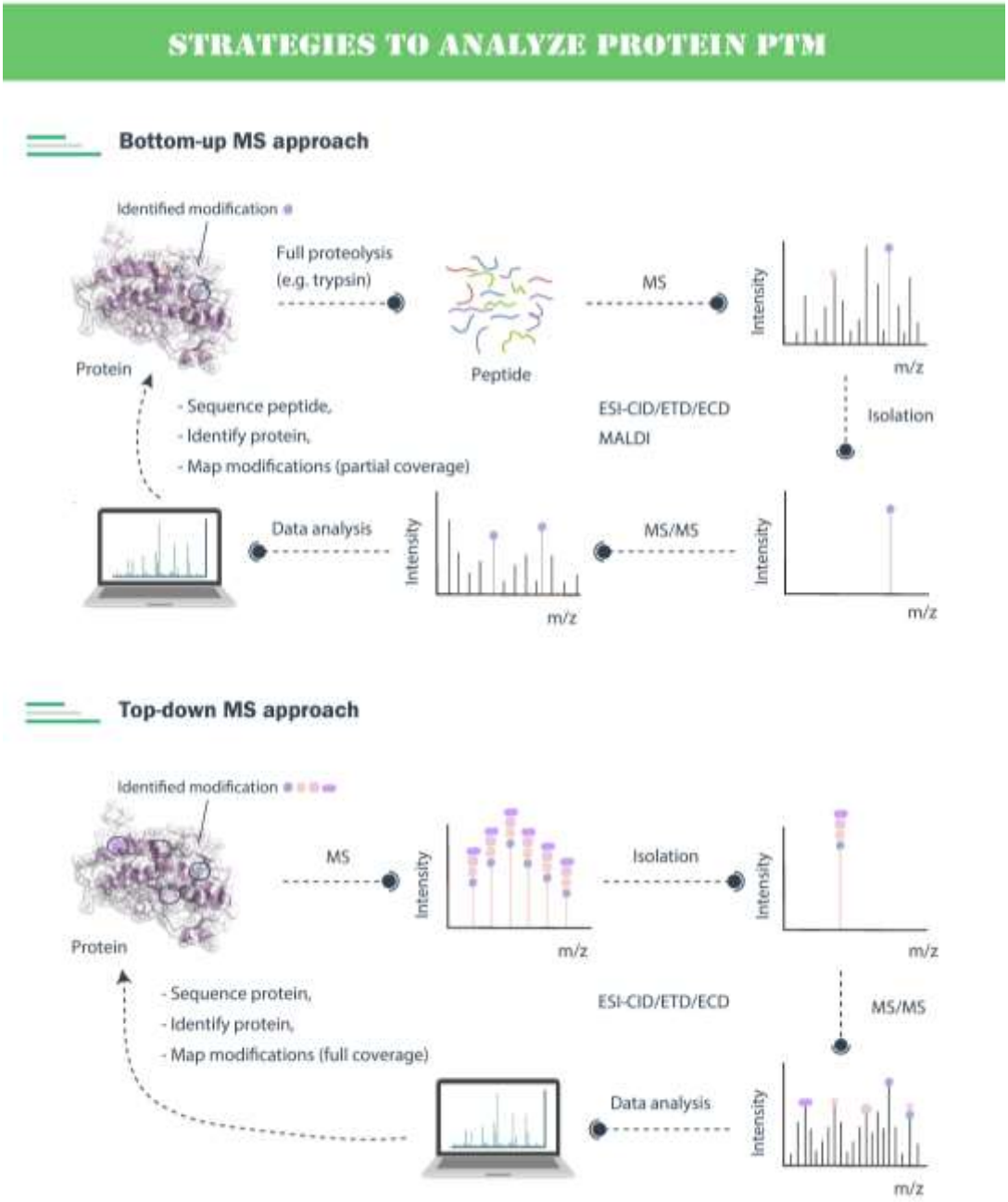
Some types of post-translational modification are consequences of oxidative stress. Carbonylation is one example that targets the modified protein for degradation and can

result in the formation of protein aggregates. Specific amino acid modifications can be used as biomarkers indicating oxidative damage.

Sites that often undergo post-translational modification are those that have a functional group that can serve as a nucleophile in the reaction: the hydroxyl groups of serine, threonine, and tyrosine; the amine forms of lysine, arginine, and histidine; the thiolate anion of cysteine; the carboxylates of aspartate and glutamate; and the N- and C-termini. In addition, although the amide of asparagine is a weak nucleophile, it can serve as an attachment point for glycans. Rarer modifications can occur at oxidized methionines and at some methylene groups in side chains.

Post-translational modification of proteins can be experimentally detected by a variety of techniques, including mass spectrometry, Eastern blotting, and Western blotting.





Applications of proteomics in agriculture:

Proteomics can be harnessed readily to improve agricultural productivity, increase food nutrition, and reduce allergen content, contributing to better food security in a rapidly changing world. Proteomics has enabled it to identify a wide range of proteins that play key roles in maintaining crop health, resistance to stressors, and improving product quality. The characterization of specific genes provides researchers the ability to

monitor crop response under adverse conditions or even predict the nutritional value of crops. Specifically, targeted genes may be more or less expressed under certain conditions, and these changes can now be identified.

- Rice under drought stress was shown to have proteins involved in photosynthesis, cell elongation, antioxidant metabolism and lignifications. Drought-responsive proteins, APX and Cu-Zn SOD, have been identified and can be used as markers to research plant responses to drought stress.
- OsCYP2, a salt-induced rice cyclophilin was isolated and identified and having Peptidyl-Prolyl cis-trans Isomerase activity. Cyclophilin 2 (OsCYP2) transgenic rice seedlings were found to be more resistant to salt stress than wild-type seedlings.
- During Magnaporthe infestation in rice, an apoplast localized cyclophilin group protein was discovered, suggesting that its activity is critical for dealing with most cellular stressors (Shenton et al., 2012).
- The investigation of the process of programmed cell death during salt stress begins with the mitochondrial proteome.
- Now, the proteomic technique attempts to characterize the proteins responsible for resistance caused by PAMPs, which will be critical in the plant defence era.
- Characterization of the process with the final gene product responsible for the bio control features, such as lytic enzymes and plant resistance elicitors, can be done directly.
- Identification of bacteria might be done accurately and has been utilized for taxonomic studies by acquiring the profile of those proteins and analyzing quantitatively.
- Ekramoddoullah and Hunt in 1993 used 2D PAGE to differentiate between susceptible and resistant seedlings of *Pinus lambertiana* (sugar-pine) to white pine blister rust *Cronartium ribicola*.
- The Pierre de Wit research group's pioneering work on the *Cladosporium fulvum* and tomato interaction from 1985 exemplifies the importance of proteomics in

plant fungal diseases research. The first avirulence gene product (Avr9) was characterized using preparative polyacrylamide gel electrophoresis in 1986 from extraction of tomato apoplastic fluids.

- Meiotic proteomics can provide you a lot of information on how meiocytes work. It reveals the recombination mechanism, highlighting the effects of male sterility genes and elucidating cell cycle regulation.

11. Membrane Transport: Lipid bilayer, Membrane transport proteins, Active and passive membrane transport, Ion channels.

The cytoplasmic membrane is a hydrophobic barrier that limits the entry and exit of molecules from the cell. Nonetheless, living cells require the exchange of many types of molecules between the cytoplasm and the outside environment. Cells need nutrients and inorganic ions, and they must remove waste products and expel other ions. This chapter focuses on membrane impermeability and on membrane proteins that facilitate and control the transport of molecules across the lipid bilayer with a special emphasis on ions (e.g., K+ and Na+).

Fick's First Law

The tendency for solutes to move from a region of higher concentration to one of lower concentration was first defined in 1855 by the physiologist Adolf Fick .His work is summarized in what is now the very well-known Fick's Laws of Diffusion. The laws apply to both free solution and diffusion across membranes. Fick developed his laws by measuring concentrations and fluxes of salt diffusing between two reservoirs through connecting tubes of water.

Fick's First Law describes diffusion as:

Diffusion rate = - DA (dc/dx)

Where D = diffusion coefficient (bigger molecules have lower Ds); A =cross-sectional area over which diffusion occurs; dc/dx is the solute concentration gradient (diffusion occurs from a region of higher concentration to one of lower concentration).

1. OSMOSIS

Osmosis is a special type of diffusion, namely the diffusion of water across a semipermeable membrane. Water readily crosses a membrane down its potential gradient from high to low. Osmotic pressure is the force required to prevent water movement across the semipermeable membrane. Net water movement continues until its potential reaches zero. An early application of the basic principles of osmosis came from the pioneering work on hemolysis of red blood cells by William Hewson in the 1770s. It has also been discussed that MLVs (multilamellar vesicles, liposomes) behave as almost perfect osmometers, swelling in hypotonic solutions and shrinking in hypertonic solutions. Liposome swelling and shrinking can be easily followed by changes in absorbance due to light scattering using a simple spectrophotometer. Therefore, osmosis has been investigated for many years using common and inexpensive methodologies and a lot is known about the process.

Membranes are rarely, if ever, perfectly semipermeable. Deviation from ideality is defined by a reflection coefficient (s). For an ideal semipermeable membrane where a solute is totally impermeable, s = 1. If a solute is totally permeable (its permeability is equal to water), s = 0. Biological membranes are excellent semipermeable barriers with s = 0.75 to 1.0.

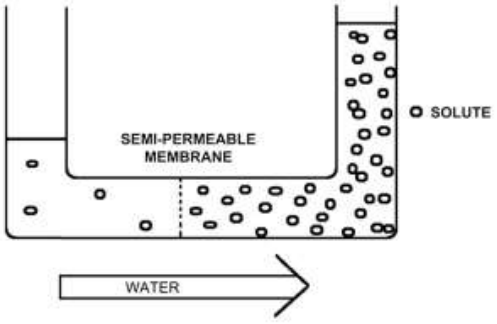


Fig.: Osmosis and osmotic pressure.

2. SIMPLE PASSIVE DIFFUSION

Movement of solutes across membranes can be divided into two basic types: passive diffusion and active transport .Passive diffusion requires no additional energy source other than what is found in the solute's electrochemical (concentration) gradient and results in the solute reaching equilibrium across the membrane. Passive diffusion can be either simple passive diffusion where the solute crosses the membrane anywhere by simply dissolving into and diffusing through the lipid bilayer, or facilitated passive diffusion where the solute crosses the membrane at specific locations where diffusion is assisted by solute-specific facilitators or carriers. Active transport requires additional energy, often in the form of ATP, and results in a non-equilibrium, net accumulation (uptake) of the solute on one side of the membrane. The basic types of membrane transport, simple passive diffusion, facilitated diffusion (by channels and carriers) and active transport are summarized in Fig.below. There are countless different examples of each type of membrane transport process .Only a few representative examples will be discussed here. Even simple passive diffusion requires energy to cross a bilayer membrane. In order to cross a membrane, the solute must first lose its waters of hydration, diffuse across the membrane, and then regain its waters on the opposite side.

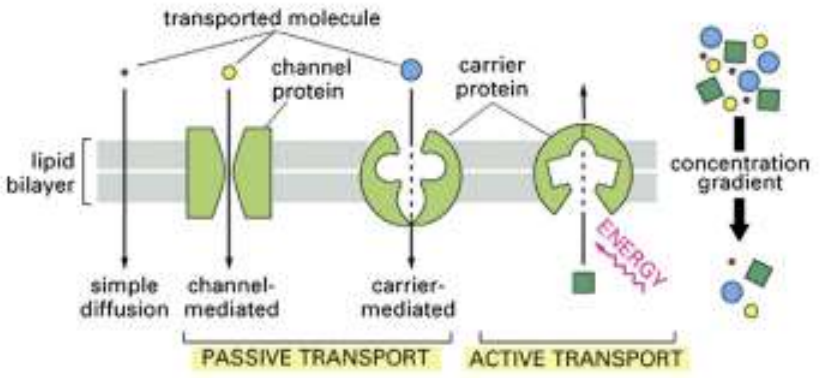


Fig.: Basic types of membrane transport, simple passive diffusion, facilitated diffusion (by channels and carriers), and active transport.

3. FACILITATED DIFFUSION

Facilitated diffusion (also known as carrier-mediated diffusion) is, like simple passive diffusion, dependent on the inherent energy in a solute gradient. No additional energy is required to transport the solute and the final solute distribution reaches equilibrium across the membrane. Facilitated diffusion, unlike simple passive diffusion, requires a highly specific transmembrane integral protein or carrier to assist in the solute's membrane passage. Facilitators come in two basic types: carriers and gated channels. Facilitated diffusion exhibits Michaelis-Menton saturation kinetics (Figure Part A, right) indicating the carrier has an enzyme-like active site. Like enzymes, facilitated diffusion carriers exhibit saturation kinetics and recognize their solute with exquisite precision, easily distinguishing chemically similar isomers like D-glucose from L-glucose. Figure (Part A) compares simple passive diffusion to facilitated diffusion. The figure is not to scale, however, as facilitated diffusion is orders of magnitude faster than simple passive diffusion.

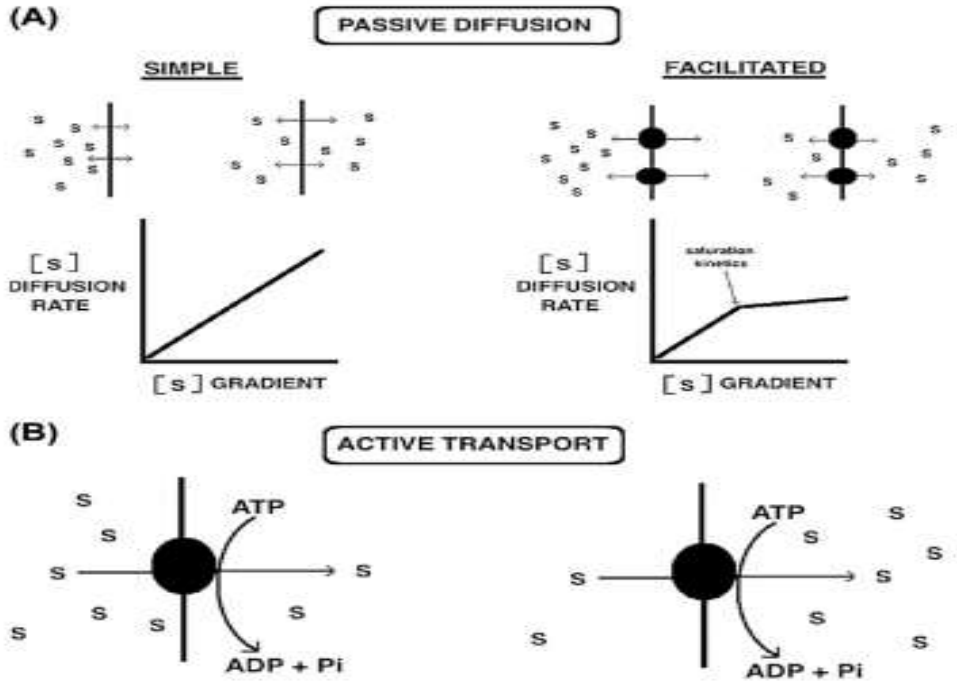


Fig.: (A) Simple passive diffusion (top, left) and facilitated passive diffusion (top, right).

❖ Glucose Transporter

A well-studied example of a facilitated diffusion carrier is the glucose transporter, or GLUT. GLUTs occur in nearly all cells and are particularly abundant in cells lining the small intestine. GLUTs are integral membrane proteins whose membrane-spanning region is composed of 12 a-helices. GLUTs function through a typical membrane transport mechanism. Glucose binds to the membrane outer surface site causing a conformational change associated with transport across the membrane. At the inner side of the membrane, glucose is released into the internal aqueous solution.

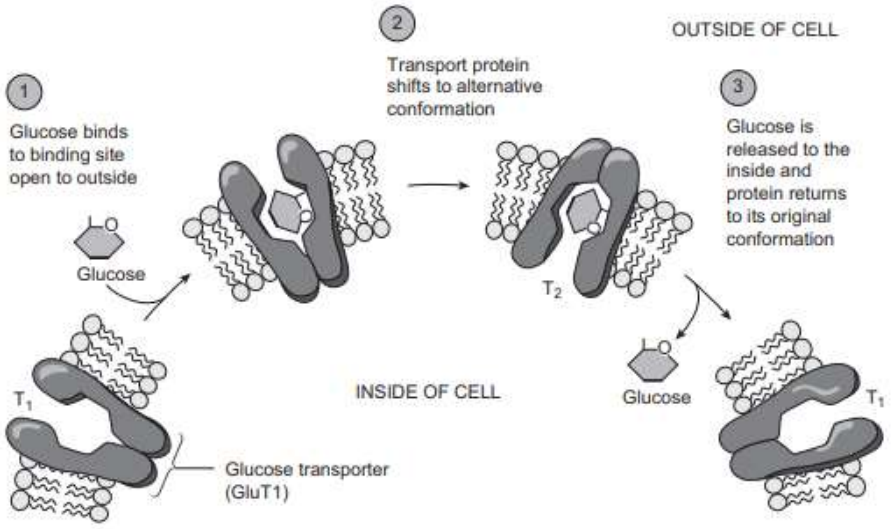


Fig.: Glucose-facilitated diffusion transporter GLUT-1.

❖ Potassium Channels

In virtually all organisms there exists a wide variety of ion channels, the most widely distributed being potassium channels. There are four basic classes of potassium channels, all of which provide essential membrane-associated functions including setting and shaping action potentials and hormone secretion:

1. Calcium-activated potassium channel

2. Inwardly rectifying potassium channel

3. Tandem pore domain potassium channel

4. Voltage-gated potassium channel

Potassium channels are composed of four protein subunits that can be the same (homotetramer) or closely related (heterotetramer). All potassium channel subunits have a distinctive pore-loop structure that sits at the top of the channel and is responsible for potassium selectivity. This is often referred to as a selectivity or filter loop. The selectivity filter strips the waters of hydration from the potassium ion, allowing it into the channel.

- **Sodium Channel**

In some ways, Na^+ channels parallel the action of K^+ channels. They are both facilitated diffusion carriers that conduct the cation down the ion's electrochemical gradient. In excitable cells such as neurons, myocytes, and some glia, Na^+ channels are responsible for the rising phase of action potentials .Therefore agents that block Na^+ channels also block nerve conduction and so are deadly neurotoxins. There are two basic types of Na^+ channels: voltage-gated and ligand-gated. The opening of a Na^+ channel has a selectivity filter that attracts Na^+. Of particular interest are two extremely potent biological toxins, tetrodotoxin (TTX) and saxitoxin (STX) that, in seafood, have killed and injured many humans. Both toxins shut down Na^+ channels by binding from the extracellular side.

- **Aquaporins**

Aquaporins are also known as water channels and are considered to be "the plumbing system for cells". Aquaporins are usually specific for water permeability and exclude the passage of other solutes. A type of aquaporin known as aqua-glyceroporins can also conduct some very small uncharged solutes such as glycerol, CO_2, ammonia, and urea across the membrane. However, all aquaporins are impermeable to charged solutes. Water molecules traverse the aquaporin channel in single file.

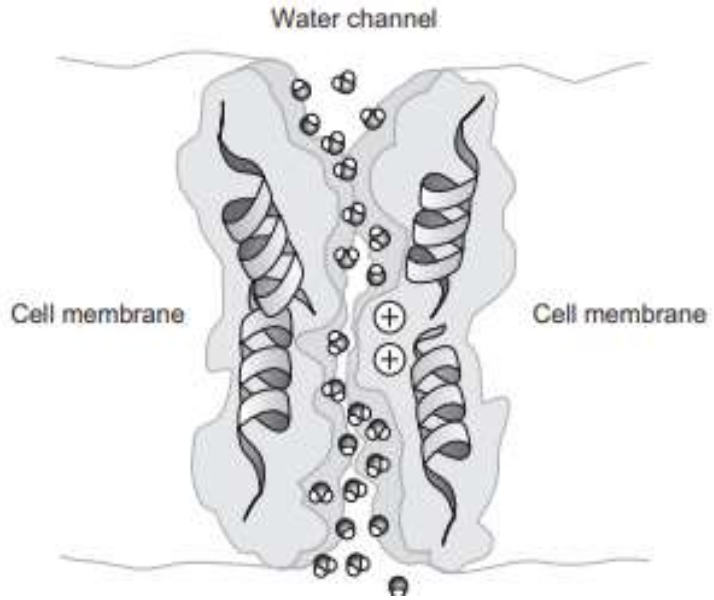


Fig.: Aquaporin. Water molecules pass through the aquaporin channel in single file.

4. ACTIVE TRANSPORT

Active transport requires a form of energy (often ATP) to drive the movement of solutes against their electrochemical gradient, resulting in a nonequilibrium distribution of the solute across the membrane. A number of nonexclusive and overlapping terms are commonly used to describe the different types of active transport. Some of these are depicted in Figure below.

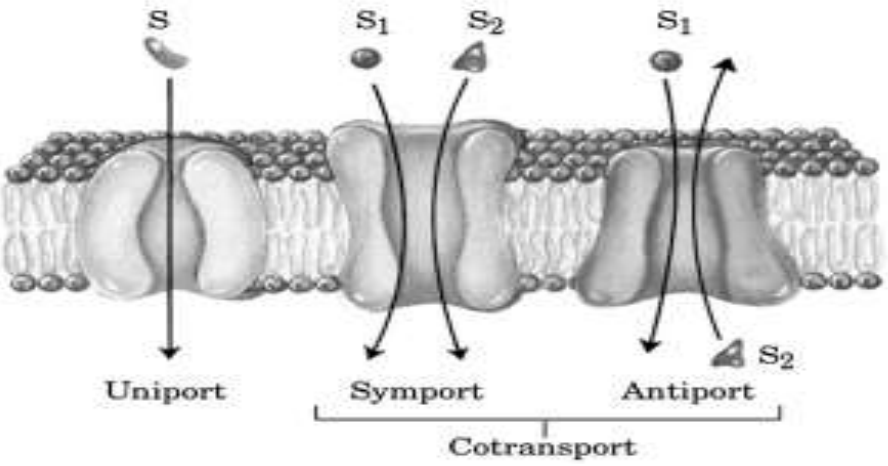


Fig.: Basic types of active transport.

❖ Primary Active Transport

Primary active transport is also called direct active transport or uniport. It involves using energy (usually ATP) to directly pump a solute across a membrane against its electrochemical gradient. The most studied example of primary active transport is the plasma membrane Naþ,K⁺ATPase. Other familiar examples of primary active transport are the redox H^+-gradient generating system of mitochondria.

❖ Na^+,K^+-ATPase

Arguably the most important active transport protein is the plasma membrane-bound Na^+,K^+-ATPase. This single enzyme accounts for one-third of human energy expenditure and is often referred to as the "pacemaker for metabolism."

ATP-using primary active transport systems	Example
P-type	Na^+,K^+-ATPase Ca^{2+} pump H^+ acid pump
F-type	Mitochondrial ATP synthase Chloroplast ATP synthase
V-type	Vacuolar ATPase
ABC (ATP binding cassette transporter)	Many

❖ Secondary Active Transport

Secondary active transport (also known as cotransport) systems are composed of two separate functions. The energy-dependent movement of an ion (eg, H+, Na+, or K+) generates an electrochemical gradient of the ion across the membrane. This ion gradient is coupled to the movement of a solute in either the same direction (symport) or in the opposite direction (antiport). . The purpose of both types of co-transport is to use the energy in an electrochemical gradient to drive the movement of another solute against its gradient. An example of symport is the SGLT1 (sodium-glucose transport protein-1) in the intestinal epithelium.

- **Bacterial Lactose Transport**

Lactose uptake is driven through a channel by a H+ gradient generated by the bacterial electron transport system.

- **Vectorial Metabolism, Group Translocation**

Vectorial metabolism has been used to describe the mechanism for several membrane transport systems. For example, it has been reported in some cases the uptake of glucose into a cell may be faster if the external source of glucose is sucrose rather than free glucose. Through a vectorial transmembrane reaction, membrane-bound sucrase may convert external sucrose into internal glucose plus fructose more rapidly than the direct transport of free glucose through its transport system. Mitchell defined one type of vectorial transport as group translocation, the best example being the PTS (phosphotransferase system) discovered by Saul Roseman in 1964. PTS is a multicomponent active transport system that uses the energy of intracellular phosphoenol pyruvate (PEP) to take up extracellular sugars in bacteria. Transported sugars include glucose, mannose, fructose, and cellobiose.

5. IONOPHORES

The term ionophore means "ion bearer." Ionophores are small, lipid-soluble molecules, usually of microbial origin, whose function is to conduct ions across membranes. They are facilitated diffusion carriers that transport ions down their electrochemical gradient. Ionophores can be divided into two basic classes: channel formers and mobile carriers. Channel formers are long lasting, stationary structures that allow many ions at a time to rapidly flow across a membrane. Mobile carriers bind to an ion on one side of a membrane, dissolve in and cross the membrane bilayer and release the ion on the other side. They can only carry one ion at a time. Four representative ionophores will be discussed: the K+ ionophore valinomycin, the proton ionophore 2,4-dinitrophenol, synthetic crown ethers, and the channel-forming ionophore nystatin.

- **Valinomycin**

Superficially valinomycin resembles a cyclic peptide. Valinomycin has an oily surface that readily dissolves in a membrane lipid bilayer, carrying K+across the membrane down its electrochemical gradient.

- **2, 4-Dinitrophenol**

2, 4-Dinitrophenol (DNP) is considered to be the classic uncoupler of oxidative phosphorylation). It is a synthetic lipid-soluble proton ionophore that dissipates proton gradients across bioenergetic membranes (mitochondrial inner, thylakoid, bacterial plasma). An uncoupler is therefore an H+-facilitated diffusion carrier.

- **Nystatin**

Nystatin is a channel-forming ionophore that creates a hydrophobic pore across a membrane. Channel-forming ionophores allow for the rapid facilitated diffusion of various ions that depend on the dimensions of the pore. Nystatin, like other channel-forming ionophores (eg, amphotericin B and natamycin), is a commonly used antifungal agent.

6. GAP JUNCTIONS

Gap junctions are a common structural feature of many animal plasma membranes. In plants similar structures are known as plasmodesmata. Gap junctions represent a primitive type of intercellular communication that allows transmembrane passage of small solutes like ions, sugars, amino acids, and nucleotides while preventing migration of organelles and large polymers like proteins and nucleic acids. Gap junctions connect the cytoplasms of two adjacent cells through nonselective channels. Each channel in a gap junction is made up of 12 proteins called connexins. Six hexagonally arranged connexins are associated with each of the adjacent cell plasma membranes that the gap junction spans. Each set of six connexins is called a connexon and forms half of the gap junction channel. Therefore, one gap junction channel is composed of 2 aligned connexons and 12 connexins. Each connexin has a diameter of about 7 nm and the hollow center formed between the 6 connexins (the channel) is about 3 nm in diameter. Gap junctions allow adjacent cells to be in constant electrical and chemical

communication with one another. Of particular importance is the rapid transmission of small second messengers, such as inositol triphosphate (IP3) and Ca2+.

12. Intracellular Compartments and Protein Sorting: Compartmentalization of Higher Cells, Signal peptides and signal patches; Transport of proteins into nucleus, mitochondria and chloroplasts; Transport of proteins from E.R. through the golgi apparatus; Role of M6P (Mannose 6-Phosphate) receptor in lysosomal enzyme sorting; Transport from the Plasma membrane via Endosomes- Endocytosis.

Compartmentalization of Higher Cells:

Many vital biochemical processes take place in or on membrane surfaces. Lipid metabolism, for example, is catalyzed mostly by membrane-bound enzymes, and oxidative phosphorylation and photosynthesis both require a membrane to couple the transport of H+ to the synthesis of ATP. Intracellular membrane systems, however, do more for the cell than just provide increased membrane area: they create enclosed compartments that are separate from the cytosol, thus providing the cell with functionally specialized aqueous spaces.

The major intracellular compartments common to eucaryotic cells are

1. The **nucleus** contains the main genome and is the principal site of DNA and RNA synthesis.
2. The **cytoplasm** consists of the cytosol and the cytoplasmic organelles suspended in it. The cytosol, constituting a little more than half the total

volume of the cell, is the site of protein synthesis and degradation. It also performs most of the cell's intermediary.

3. About half the total area of membrane in a eucaryotic cell encloses the labyrinthine spaces of the **endoplasmic reticulum (ER)**. The ER has many ribosomes bound to its cytosolic surface; these are engaged in the synthesis of both soluble and integral membrane proteins, most of which are destined either for secretion to the cell exterior or for other organelles. The ER also produces most of the lipid for the rest of the cell and functions as a store for Ca^{2+} ions. The ER sends many of its proteins and lipids to the Golgi apparatus.
4. The **Golgi apparatus** consists of organized stacks of disclike compartments called Golgi cisternae; it receives lipids and proteins from the ER and dispatches them to a variety of destinations, usually covalently modifying them en route.
5. **Mitochondria and chloroplasts (in plants}** generate most of the ATP used by cells to drive reactions that require an input of free energy; chloroplasts are a specialized version of plastids, which can also have other functions in plant cells, such as the storage of food or pigment molecules.
6. **Lysosomes** contain digestive enzymes that degrade defunct intracellular organelles, as well as macromolecules and particles taken in from outside the cell by endocytosis. On their way to lysosomes, endocytosed material must first pass through a series of organelles called endosomes.
7. **Peroxisomes** are small vesicular compartments that contain enzymes utilized in a variety of oxidative reactions.

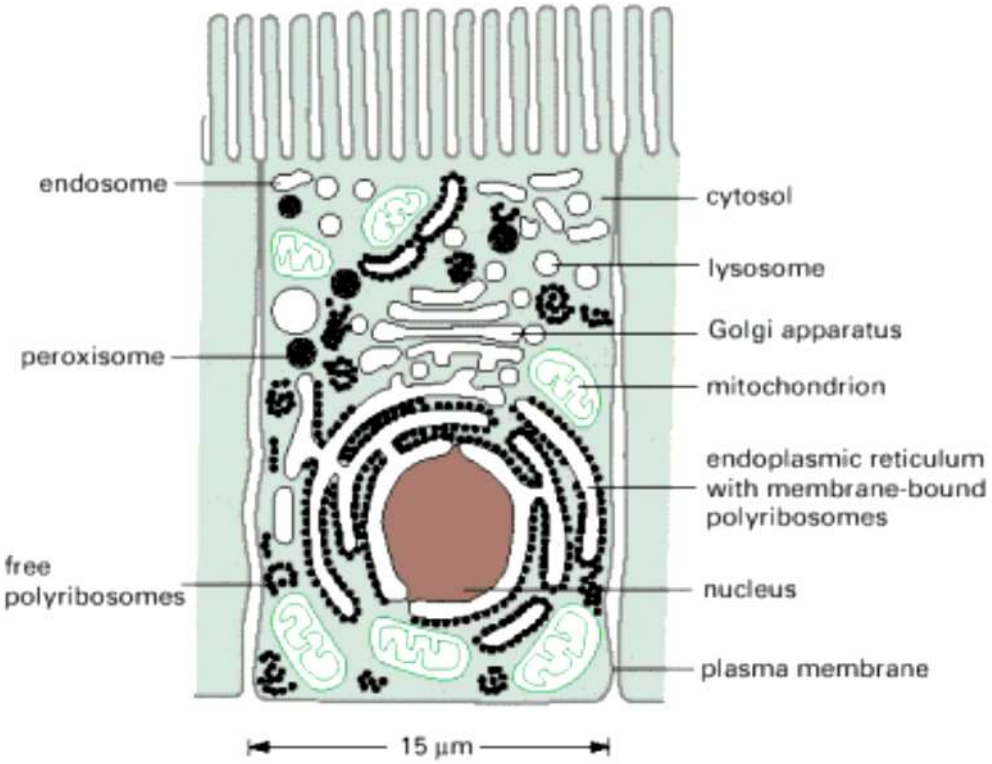


Fig.: The major intracellular compartments of a cell.

In general, each membrane-enclosed organelle performs the same set of basic functions in all cell types. But to serve the specialized functions of cells, these organelles will vary in abundance and can have additional properties that differ from cell type to cell type.

Membrane-enclosed organelles often have characteristic positions in the cytosol. In most cells, for example, the Golgi apparatus is located close to the nucleus, whereas the network of ER tubules extends from the nucleus throughout the entire cytosol.

Protein sorting:

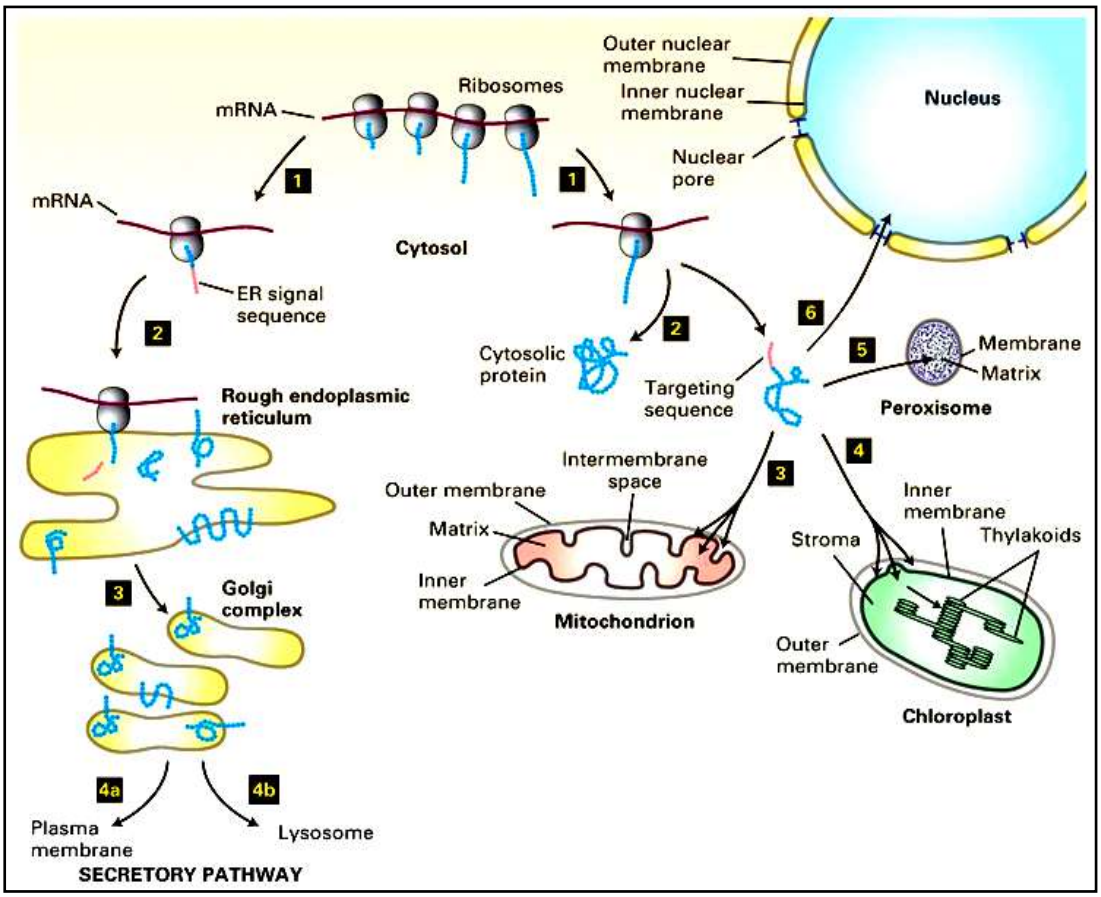
Protein targeting or protein sorting is the biological mechanism by which proteins are transported to their appropriate destinations in the cell or outside it. Proteins can be targeted to the inner space of an organelle, different intracellular membranes, plasma membrane, or to exterior of the cell via secretion.

Both in prokaryotes and eukaryotes, newly synthesized proteins must be delivered to a specific sub cellular location or exported from the cell for correct activity. Protein

targeting is necessary for proteins that are destined to work outside the cytoplasm. This delivery process is carried out based on information contained in the protein itself. Correct sorting is crucial for the cell; errors can lead to diseases.

The endomembrane system and secretory pathway:

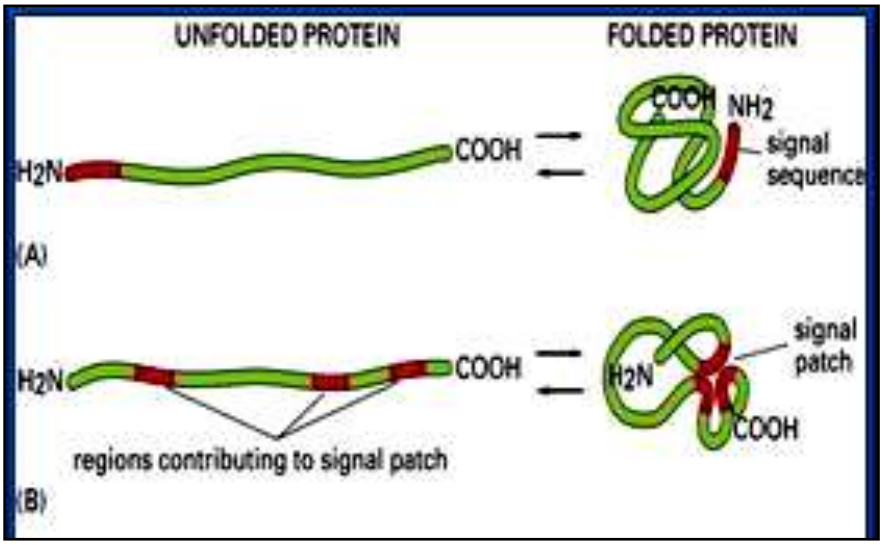
Proteins destined for any part of the endomembrane system (or the outside of the cell) are brought to the ER during translation and fed in as they're made.



Types of Sorting Signals in Proteins

1. **Signal Sequence**
   - *continuous sequence of 15-60 aminoacids (aa's)*
   - *sometimes removed from finishedprotein*
   - *sometimes a part of finished protein*

2. **Signal Patch**
   - *specific 3D arrangement of atoms onprotein surface; aa's distant*
   - *persist in finished protein*



Signal peptide:

The signal peptide that sends a protein into the endoplasmic reticulum during translation is a series of hydrophobic ("water-fearing") amino acids, usually found near the beginning (N- terminus) of the protein. When this sequence sticks out of the ribosome, it's recognized by a protein complex called the signal-recognition particle (SRP), which takes the ribosome, to the ER. There, the ribosome feeds its amino acid chain into the ER lumen (interior) as it's made.

(i) Signal recognition particle (SRP) binds to the signal peptide as it emerges from the ribosome.

(ii) SRP brings the ribosome to the ER by binding to a receptor on the ER surface. The receptor isassociated with other proteins that make a pore.

(iii) The ribosome resumes translating, feeding the polypeptide through the pore and into the ERlumen (interior).

(iv) An enzyme associated with the pore snips off the signal peptide.

(v) Translation continues, and the growing amino acid chain slides into the ER lumen.

(vi) The completed polypeptide is released into the ER lumen, where it floats freely.

(vii) In some cases, the signal peptide is snipped off during translation and the finished protein is released into the interior of the ER (as shown above). In other cases, the signal peptide or another stretch of hydrophobic amino acids gets embedded in the ER membrane. This creates a transmembrane (membrane-crossing) segment that anchors the protein to the membrane.

Table: Some Typical Signal Sequences

FUNCTION OF SIGNAL SEQUENCE	EXAMPLE OF SIGNAL SEQUENCE
Import into nucleus	-Pro-Pro-Lys-Lys-Lys-Arg-Lys-Val-
Export from nucleus	-Leu-Ala-Leu-Lys-Leu-Ala-Gly-Leu-Asp-Ile-
Import into mitochondria	$^{+}H_3N$-Met-Leu-Ser-Leu-Arg-Gln-Ser-Ile-Arg-Phe-Phe-Lys-Pro-Ala-Thr-Arg-Thr-Leu-Cys-Ser-Ser-Arg-Tyr-Leu-Leu-
Import into plastid	$^{+}H_3N$-Met-Val-Ala-Met-Ala-Met-Ala-Ser-Leu-Gln-Ser-Ser-Met-Ser-Ser-Leu-Ser-Leu-Ser-Ser-Asn-Ser-Phe-Leu-Gly-Gln-Pro-Leu-Ser-Pro-Ile-Thr-Leu-Ser-Pro-Phe-Leu-Gln-Gly-
Import into peroxisomes	-Ser-Lys-Leu-COO^{-}
Import into ER	$^{+}H_3N$-Met-Met-Ser-Phe-Val-Ser-Leu-Leu-Leu-Val-Gly-Ile-Leu-Phe-Trp-Ala-Thr-Glu-Ala-Glu-Gln-Leu-Thr-Lys-Cys-Glu-Val-Phe-Gln-
Return to ER	-Lys-Asp-Glu-Leu-COO^{-}

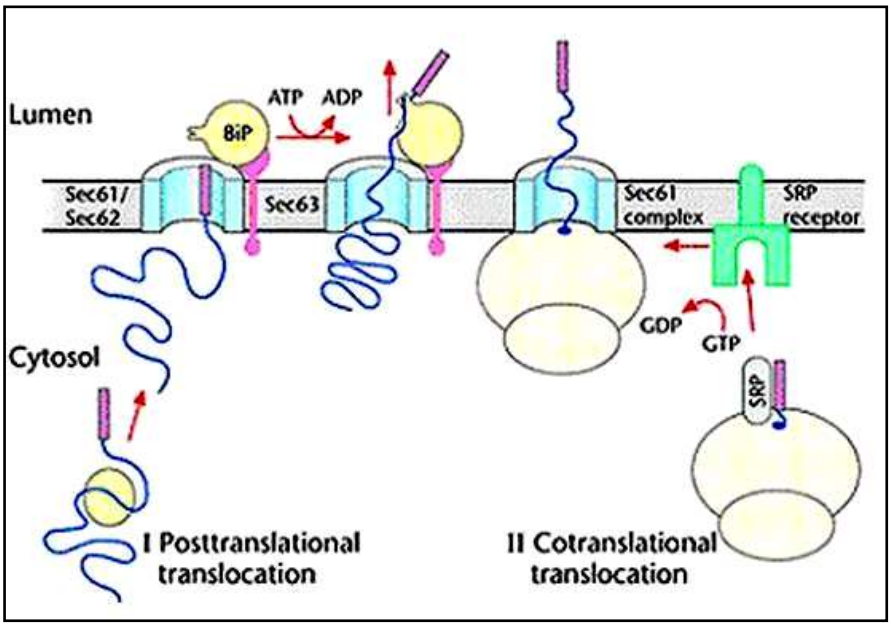
Some characteristic features of the different classes of signal sequences are highlighted in color. Where they are known to be important for the function of the signal sequence, positively charged amino acids are shown in *red* and negatively charged amino acids are shown in green. Similarly, important hydrophobic amino acids are shown in *yellow* and hydroxylated amino acids are shown in *blue*. $^{+}H_3N$ indicates the N-terminus of a protein; COO^{-} indicates the C-terminus.

Protein translocation:

In 1970, **Günter Blobel** conducted experiments on the translocation of proteins across membranes. He was awarded the Nobel Prize (1999) for his findings. He discovered that many proteins have a signal sequence, that is, a short amino acid sequence at one end that functions like a postal code for the target organelle.

Sorting or translocation of protein can occur in two ways:

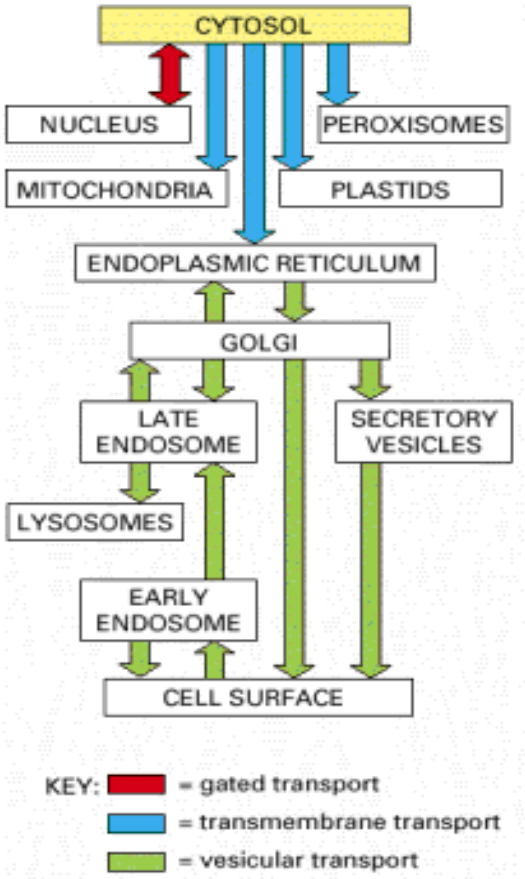
1. **Co-translational:** Synthesised protein is transferred to an SRP receptor on the endoplasmic reticulum (ER), a membrane-enclosed organelle. There, the nascent protein is inserted into the translocation complex
2. **Post-translational:** Even though most proteins are co translationally translocated, some are translated in the cytosol and later transported to their destination. This occurs for proteins that go to a mitochondrion, a chloroplast, or a peroxisome.



Proteins can move between compartments in different ways –

1. Gated transport(Nucleus)
2. Transmembrane transport(Mitochondria, Peroxisomes)
3. Vesicular transport (E.R)

1. **Gated transport:** The protein traffic between the cytosol and nucleus occurs between topologically equivalent spaces, which are in continuity through the nuclear pore complexes. The nuclear pore complexes function as selective gates that actively transport specific macromolecules and macromolecular assemblies.
2. **Transmembrane transport:** Membrane-bound protein translocators directly transport specific proteins across a membrane from the cytosol into a space that is topologically distinct. The transported protein molecule usually must unfold to snake through the translocator. The initial transport of selected proteins from the cytosol into the ER lumen or from the cytosol into mitochondria.
3. **Vesicular transport**: Proteins from the ER to the Golgi apparatus and proteins to E.R, for example, occurs in this way, transport intermediates— which may be small, spherical transport vesicles or larger, irregularly shaped organelle fragments— ferry proteins from one compartment to another. The transfer of soluble recognized by a complementary receptor in the appropriate membrane.



Transport into the nucleus:

All proteins found in the nucleus are synthesized in the cytoplasm. The flow of biomolecules between the nucleus and cytosol is bidirectional and energy dependent. The nucleocytoplasmic transport is through a large number of nuclear pore complexes (NPC). A variety of proteins, enzymes and ribonucleoprotein (RNP) complexes are continuously moving between the two compartments. These include DNA and RNA polymerases; histones, topoisomerases, different species of RNA and snRNPs. Inherent in the bidirectional movement of substances across nucleus membrane is the need to have a way to provide directionality to the transport process.

NPC is a macromolecular complex that forms an aqueous channel of approximately 9nm in diameter facilitating the diffusion of metabolites up to 50kD. The movement of larger macromolecules (up to 26-28nm in diameter) is an active process and is characterised by signal dependence and saturability.

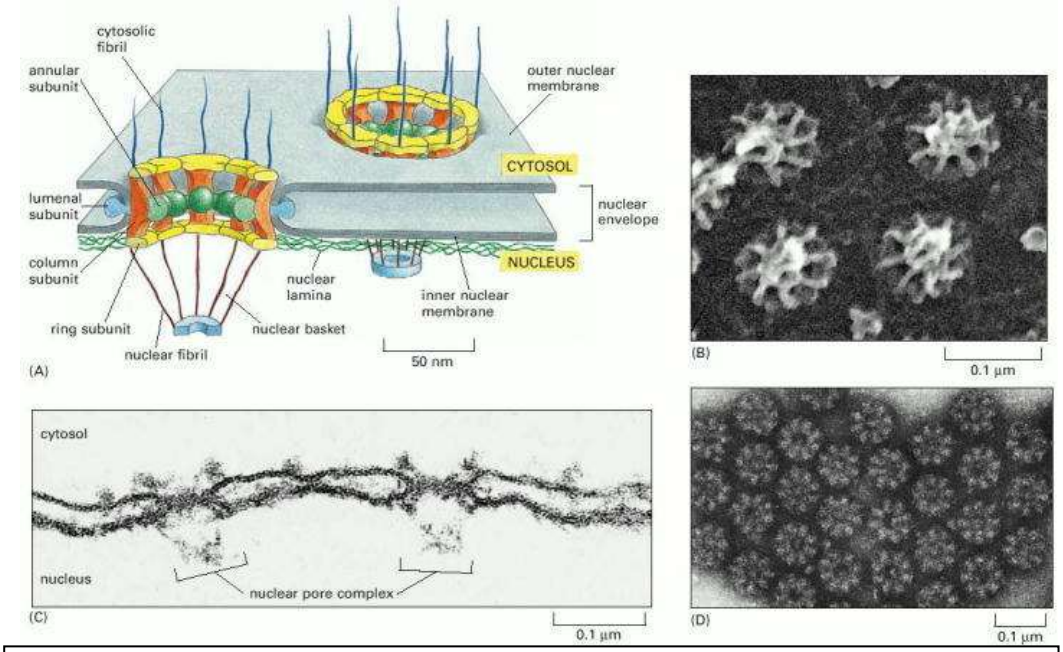


Fig.: The arrangement of nuclear pore complexes in the nuclear envelope for nuclear transport.

NPC is thought to be composed of more than 50 different proteins, called nucleoporins, that are arranged with a striking octagonal symmetry.

Nuclear Localization Signals:

When proteins are experimentally extracted from the nucleus and reintroduced into the cytosol (e.g., through experimentally induced perforations in the plasma membrane), even the very large ones reaccumulate efficiently in the nucleus. The selectivity of this nuclear import process resides in **nuclear localization signals (NLSs)**, which are present only in nuclear proteins. In many nuclear proteins they consist of one or two short sequences that are rich in the **positively charged amino acids lysine and arginine**, the precise sequence varying for different nuclear proteins.

Nuclear Import and export Receptors:

To initiate nuclear import, most nuclear localization signals must be recognized by nuclear import receptors, which are encoded by a family of related genes. The import receptors are soluble cytosolic proteins that bind both to the nuclear localization signal on the protein to be transported and to **nucleoporins**, some of which form the tentaclelike fibrils that extend into the cytosol from the rim of the nuclear pore complexes. The fibrils and many other **nucleoporins contain a large number of short amino-acid repeats that contain phenylalanine and glycine** and are therefore called **FG-repeats**. FG-repeats serve as binding sites for the import receptors. They are thought to line the path through the nuclear pore complexes taken by the import receptors and their bound cargo proteins.

Nuclear Export Works like Nuclear Import, But in Reverse. The nuclear export of large molecules, such as new ribosomal subunits and RNA molecules, also occurs through nuclear pore complexes and depends on a selective transport system. The transport system relies on nuclear export signals on the macromolecules to be exported, as well as on complementary nuclear export receptors. These receptors bind both the export signal and nucleoporins to guide their cargo through the pore complex to the cytosol.

Nuclear export receptors are structurally related to nuclear import receptors, and they are encoded by the same gene family of nuclear transport receptors, or karyopherins.

The import receptors bind their cargo molecules in the cytosol, release them in the nucleus, and are then exported to the cytosol for reuse, while the export receptors function in reverse.

Nuclear import and export cycle:

The import of nuclear proteins through the pore complex concentrates specific proteins in the nucleus, thereby increasing order in the cell, which must consume energy. The energy is thought to be provided by the hydrolysis of GTP by the monomeric GTPase Ran. Ran is found in both the cytosol and the nucleus, and it is required for both the nuclear import and export systems.

Like other GTPases, Ran is a molecular switch that can exist in two conformational states, depending on whether GDP or GTP is bound. Conversion between the two states is triggered by two Ran-specific regulatory proteins: a cytosolic GTPase-activating protein (GAP) that triggers GTP hydrolysis and thus converts Ran-GTP to Ran-GDP, and a nuclear guanine exchange factor (GEF) that promotes the exchange of GDP for GTP and thus converts Ran-GDP to Ran-GTP. Because Ran-GAP is located in the cytosol and Ran-GEF is located in the nucleus, the cytosol primarily contains Ran-GDP, and the nucleus primarily contains Ran-GTP.

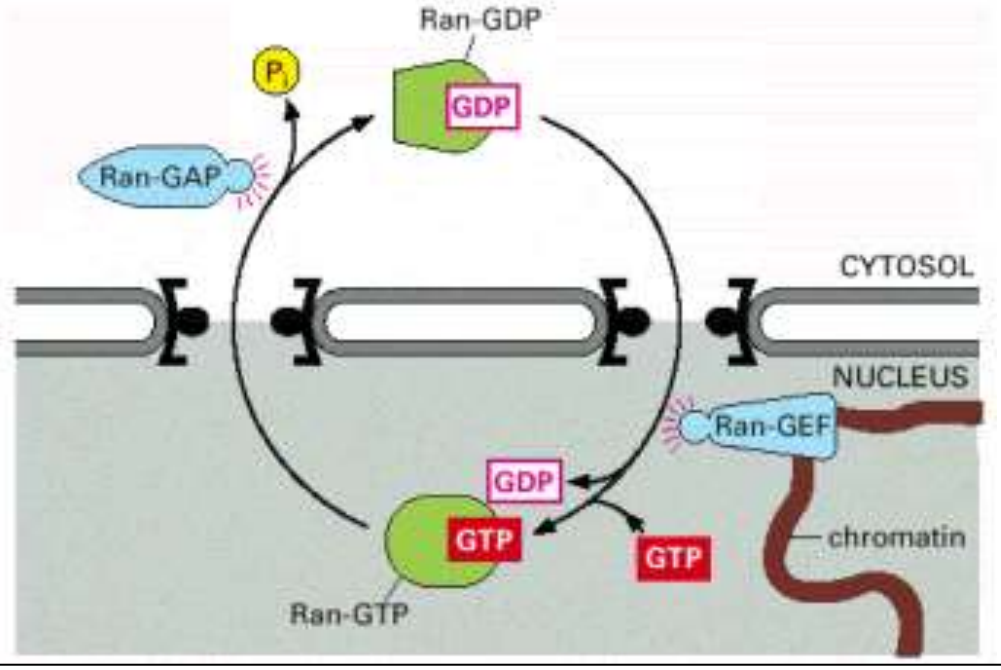


Fig.: The compartmentalization of Ran-GDP and Ran-GTP for nuclear transport.

This gradient of the two conformational forms of Ran drives nuclear transport in the appropriate direction. Docking of nuclear import receptors to FG-repeats on the cytosolic side of the nuclear pore complex, for example, occurs only when these receptors are loaded with an appropriate cargo. The import receptors with their bound cargo then move along tracks lined by FG-repeat sequences until they reach the nuclear side of the pore complex, where Ran-GTP binding causes the import receptors to release their cargo. By favouring cargo-dependent loading of import receptors onto the FG-repeat track in the cytosol and Ran-GTP-dependent cargo release in the nucleus, the nuclear localization of Ran-GTP imposes directionality.

Having discharged its cargo in the nucleus, the empty import receptor with Ran-GTP bound is transported back through the pore complex to the cytosol. There, two cytosolic proteins, Ran Binding Protein and Ran-GAP collaborate to convert Ran-GTP to Ran-GDP. The Ran Binding Protein first displaces Ran-GTP from the import receptor, which allows Ran-GAP to trigger Ran to hydrolyze its bound GTP. The Ran-GDP then dissociates from the Ran Binding Protein and is reimported into the nucleus, thereby completing the cycle.

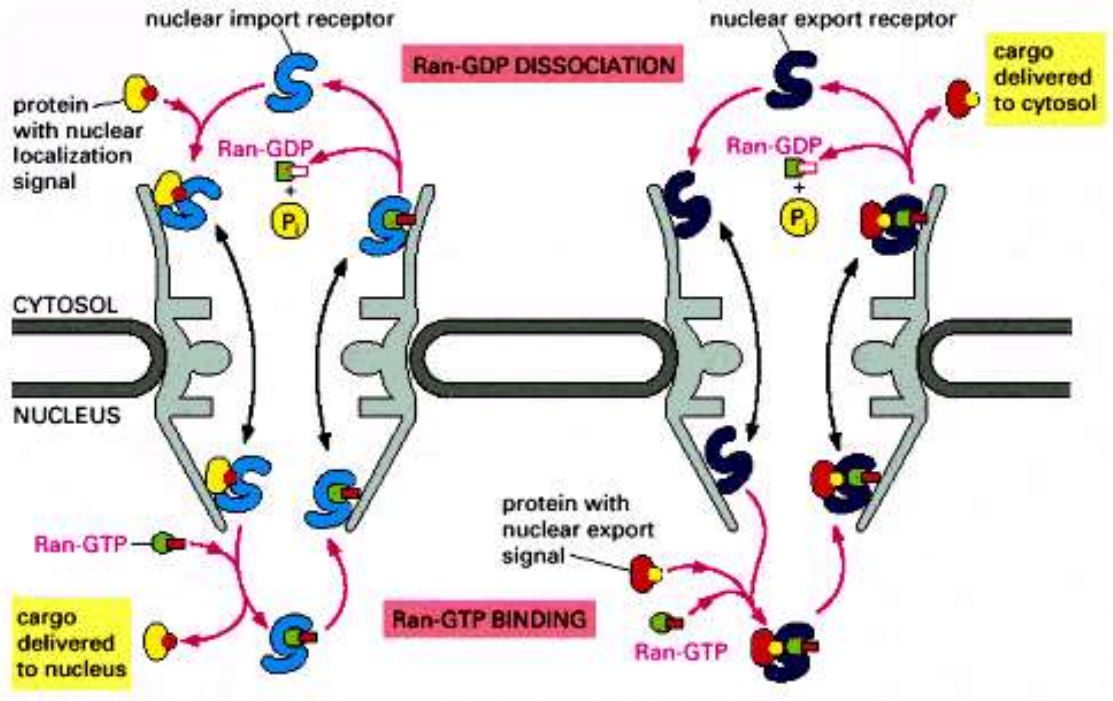


Fig.: A model for how GTP hydrolysis by Ran provides directionality for nuclear transport.

Nuclear export occurs by a similar mechanism, except that Ran-GTP in the nucleus promotes cargo binding to the export receptor and the binding of the loaded receptor to the nuclear side of the pore complex. Once in the cytosol, Ran encounters Ran-GAP and Ran Binding Protein and hydrolyses its bound GTP. The export receptor then releases both its cargo and Ran-GDP in the cytosol and dissociates from the pore complex, and free export receptors are returned to the nucleus to complete the cycle.

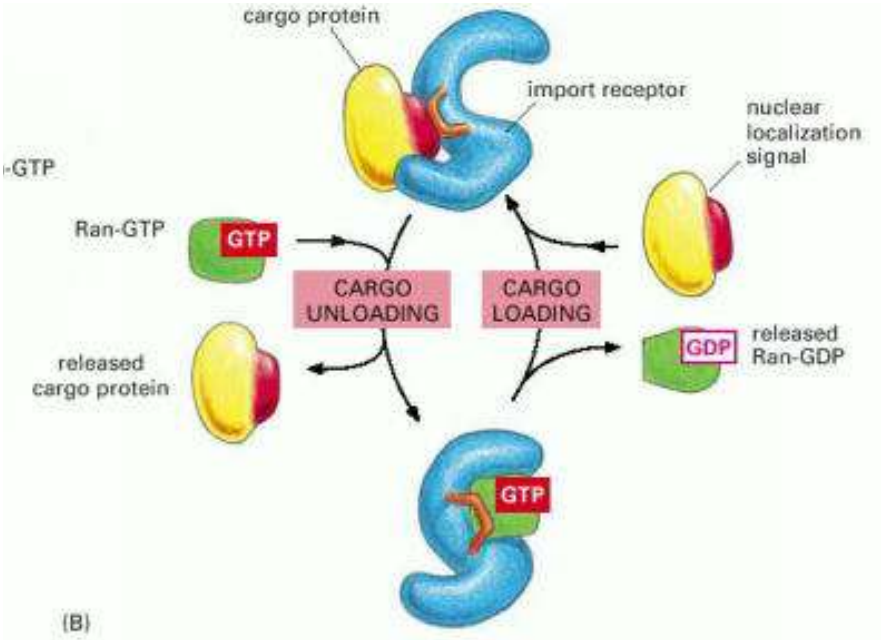


Fig.: A model for how the binding of Ran-GTP might cause nuclear import receptors to release.

Transport of proteins into mitochondria:

Mitochondria are double-membrane-enclosed organelles and are two subcompartments: the internal matrix space and the intermembrane space. These compartments are formed by the two concentric mitochondrial membranes: the inner membrane, which forms extensive invaginations, the cristae, and encloses the matrix space, and the outer membrane, which is in contact with the cytosol. Although the organelle contains their own DNA, ribosomes, and other components required for protein synthesis, most of their proteins are encoded in the cell nucleus and imported from the cytosol.

Mitochondrial signal Sequence and Protein Translocators:

Mitochondrial proteins are first fully synthesized as precursor proteins in the cytosol and then translocated into mitochondria by a posttranslational mechanism. Most of the mitochondrial precursor proteins have a **signal sequence at their N terminus** that is rapidly removed after import by a protease (the signal peptidase) in the mitochondrial matrix. These signal sequences are actually **amphipathic α helix**, in which positively charged residues are clustered on one side of the helix, while uncharged hydrophobic residues are clustered on the opposite side. This configuration—rather than a precise amino acid sequence—is recognized by specific receptor proteins that initiates protein translocation.

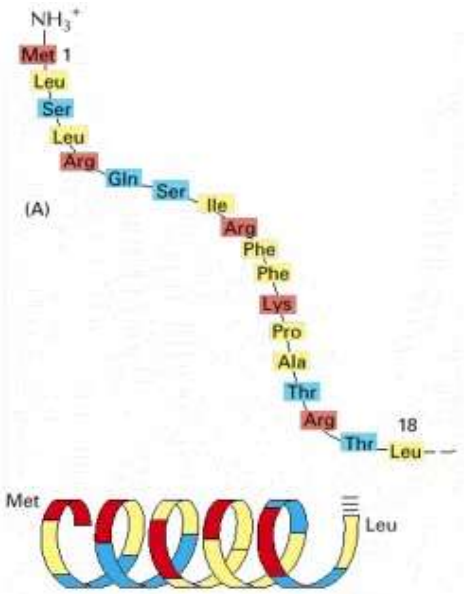


Fig.: A signal sequence for mitochondrial protein import.

The transport of proteins is mediated by membrane spanning protein channel (commonly known as translocon) that are driven by electrochemical gradient across the mitochondrial inner membrane. There are three major multiprotein membrane translocons in the mitochondria: TOM, TIM and OXA.

TOM complex is present in the outer mitochondrial membrane (OMM). This complex includes Tom40 (protein lined pore) and receptor proteins (Tom20, Tom22 and Tom5) which recognize and bind presequences on mitochondrial proteins.

TIM complex is an inner mitochondrial membrane (IMM) protein complex that contains two major protein channels, Tim23 and Tim22 complex. Tim23 protein channel has a Tim44 receptor protein which is used to translocate the mitochondrial protein into the matrix while Tim22 protein complex is used to import mitochondrial membrane protein to the IMM.

OXA complex is also located in the inner mitochondrial membrane that mediates the insertion of IMM and intermembrane space (IMS) proteins. These proteins are synthesised either in the mitochondrial matrix or cytosol. Those that are synthesised by cytosolic ribosomes generally have both a presequence and a second sorting signal while others have only one signal. It is primarily involved in the translocation of transmembrane proteins to the inner membrane.

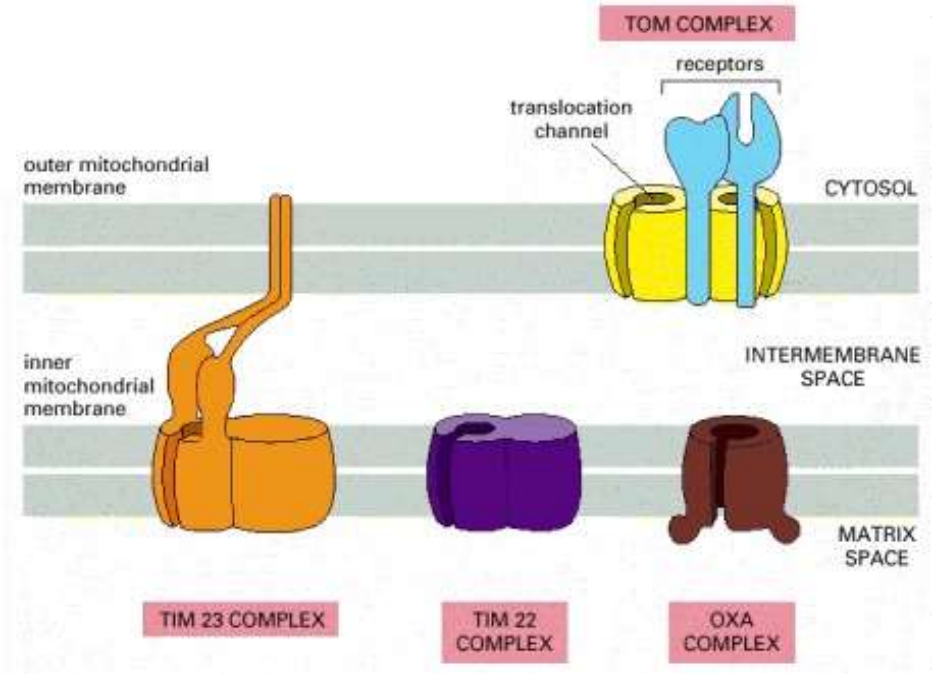


Fig.: Three protein translocators in the mitochondrial membranes.

Protein import into mitochondria:

In principle, a protein could reach the mitochondrial matrix by crossing the two membranes one at a time, or it could pass through both at once. ss through both mitochondrial membranes at once to enter the matrix. It is thought that the TOM complex first transports the mitochondrial targeting signal across the outer membrane. Once it reaches in the intermembrane space, the targeting signal binds to a TIM complex, opening the channel in the complex through which the polypeptide chain either enters the matrix or inserts into the inner membrane. Although the functions of the TOM and TIM complexes are usually coupled to allow protein transport across both membranes at the same time, both protein types of translocator can work

independently. Despite the independent functional roles of the TOM and TIM translocators, the two mitochondrial membranes at contact sites may be permanently held together by the TIM23 complex, which spans both membranes.

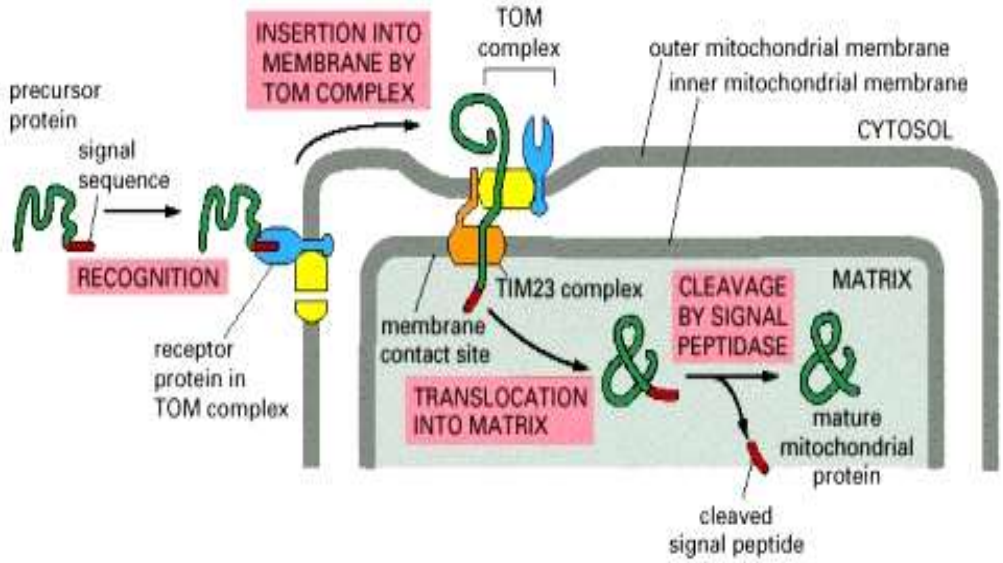


Fig.: Protein import by mitochondria.

ATP Hydrolysis and a H^+ Gradient are Used to Drive Protein Import into Mitochondria:

Mitochondrial protein import is fueled by ATP hydrolysis at two discrete sites, one outside the mitochondria and one in the matrix. In addition, another energy source is required: an electrochemical H^+ gradient across the inner mitochondrial membrane. the requirement for hsp70 and ATP in the cytosol can be bypassed if the precursor protein is artificially unfolded prior to adding it to purified mitochondria. Once the signal sequence has passed through the TOM complex and has become bound to either TIM complex, further translocation through the TIM requires an electrochemical H+ gradient across the inner membrane. The electrochemical gradient is maintained by the pumping of H+ from the matrix to the intermembrane space, driven by electron transport processes in the inner membrane.

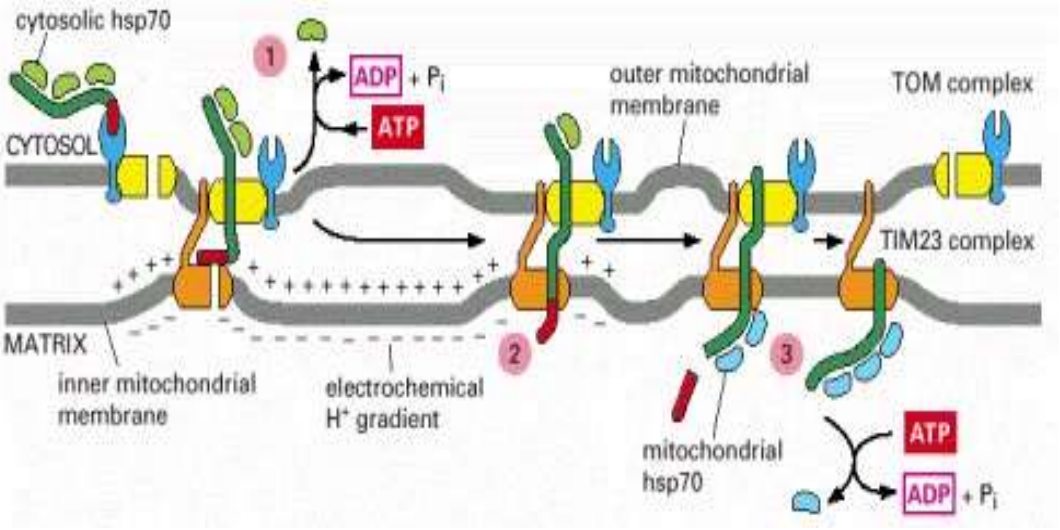
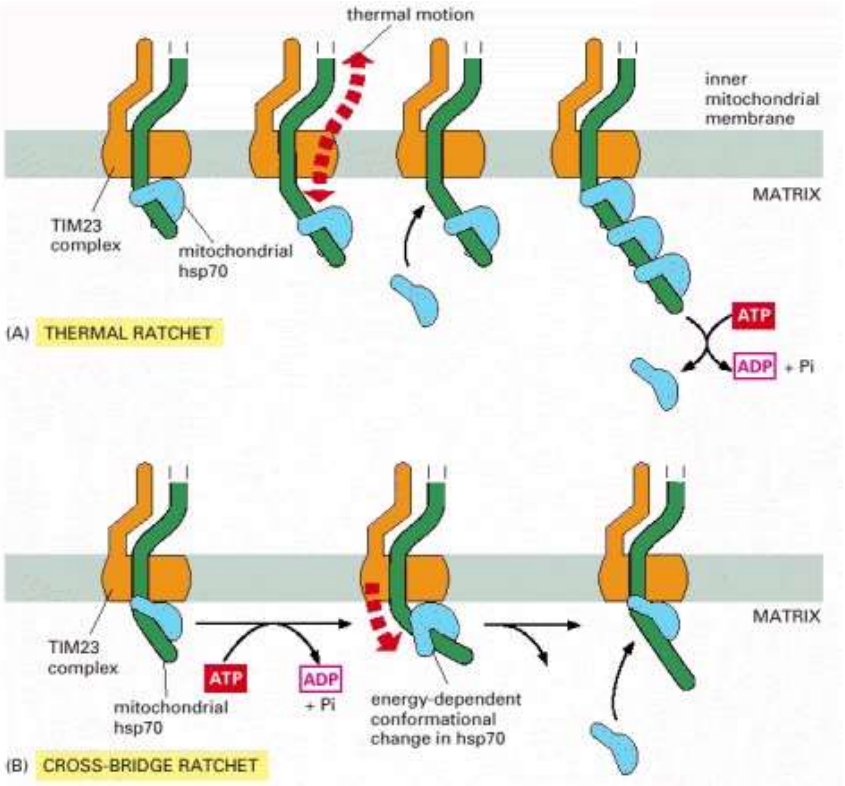


Fig.: The role of energy in protein import into the mitochondrial matrix.

Repeated Cycles of ATP Hydrolysis by Mitochondrial Hsp70 Complete the Import Process:

Thermal ratchet model: The emerging chain slides back and forth in the TIM23 translocation channel by thermal motion. Each time a sufficiently long portion of the chain is exposed in the matrix, an hsp70 molecule binds to it, preventing further backsliding and thereby making the movement directional. Thus, a hand-over-hand binding of multiple



hsp70 proteins translocates the polypeptide chain into the matrix.

Cross-bridge ratchet model: The hsp70 proteins that bind to the emerging polypeptide chain undergo a conformational change, driven by ATP hydrolysis, that actively pulls a segment of the polypeptide chain into the matrix.

Protein Transport into the Inner Mitochondrial Membrane and the Intermembrane Space Requires Two Signal Sequences

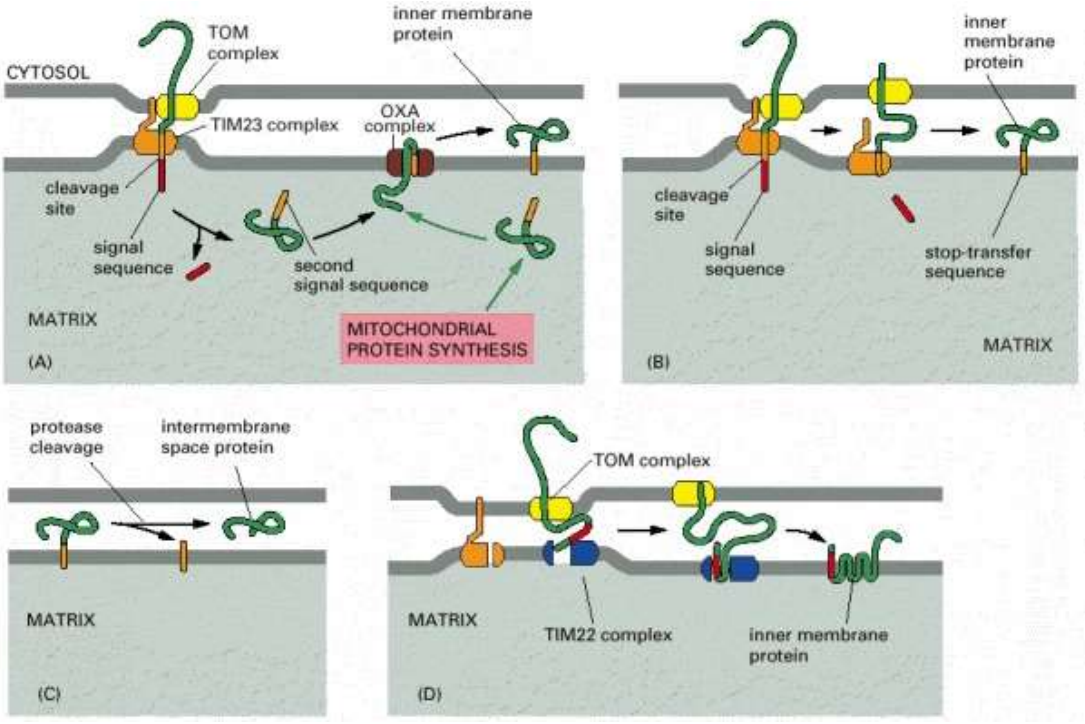


Fig.: Protein import from the cytosol into the inner mitochondrial membrane or intermembrane space.

(A) A pathway that requires two signal sequences and two translocation events is thought to be used to move some proteins from the cytosol to the inner membrane. The precursor protein is first imported into the matrix space. Cleavage of the signal sequence used for the initial translocation, however, unmasks an adjacent hydrophobic signal sequence at the new N terminus. This signal then directs the protein into the

inner membrane, presumably by the same OXA-dependent pathway that is used to insert proteins encoded by the mitochondrial genome.

(B) In some cases, the hydrophobic sequence that follows the matrix-targeting signal binds to the TIM23 translocator in the inner membrane and stops translocation. The remainder of the protein is then pulled into the intermembrane space through the TOM translocator in the outer membrane, and the hydrophobic sequence is released into the inner membrane.

(C) Some soluble proteins of the intermembrane space may also use the pathways shown in (A) and (B) before they are released into the intermembrane space by a second signal peptidase, which has its active site in the intermembrane space and removes the hydrophobic signal sequence.

(D) The import pathway used to insert metabolite carrier proteins into the inner mitochondrial membrane utilizes the TIM22 complex, which is specialized for the translocation of multipass membrane proteins.

Transport of proteins into Chloroplast:

About 90% proteins in chloroplast are encoded by nuclear genes. A chloroplast has at least six subcompartments that include the outer and inner membrane, intermembrane space, stroma, thylakoid membrane and lumen into which proteins are specifically targeted post translationally. Due to the complex organisation of the chloroplast, some proteins have to be translocated through three distinct membrane systems.

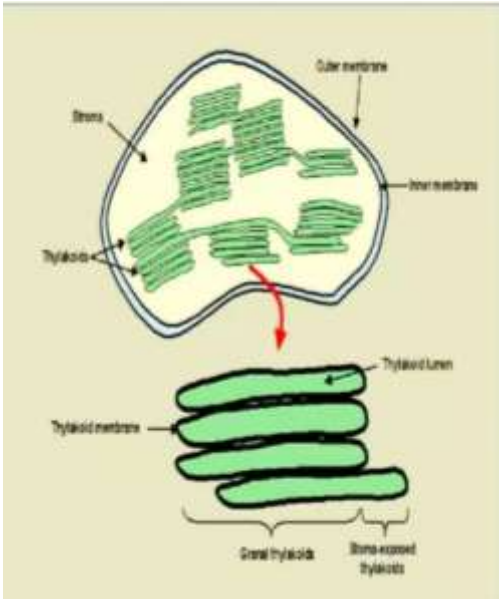


Fig.: Structure of Chloroplast.

Chloroplast N-terminal sequence and translocation factors:

Most proteins destined to the chloroplast have an **N-terminal cleavable sequence** that functions as a chloroplast targeting peptide (CTP) sequence. A number of CTPs have been identified and they are highly divergent in length, composition and organisation. Inspite of the apparent variation, CTPs do share some common features. They have a high proportion of hydroxylated amino acids, lack acidic amino acids and can fold to form αhelical structure.

Import of protein into chloroplast:

The preprotein for chloroplasts may contain a stromal import sequence or a stromal and thylakoid targeting sequence. The majority of preproteins are translocated through the Toc and Tic complexes located within the chloroplast envelope. In the stroma the stromal import sequence is cleaved off and folding as well as intra-chloroplast sorting to thylakoids continues. Proteins targeted to the envelope of chloroplasts usually lack cleavable sorting sequence.

- The vast majority of chloroplast proteins are synthesized as precursor proteins (preproteins) in the cytosol and are imported post- translationally into the organelle.
- Most proteins that are destined for the thylakoid membrane,
- Preproteins that contain a cleavable transit peptide are recognized in a GTP-regulated manner12 by receptors of the outer-envelope translocon, which is called the **TOC complex.**
- The preproteins cross the outer envelope through an aqueous pore and are then transferred to the translocon in the inner envelope, which is called the **TIC complex.**
- The TOC and TIC translocons function together during the translocation process Completion of import requires energy, which probably comes from the ATP-dependent functioning of molecular chaperones in the stroma.
- The stromal processing peptidase then cleaves the transit sequence to produce the mature form of the protein, which can fold into its native form.

Protein transport into the thylakoid Membrane:

The thylakoid signal sequence initiates transport across the thylakoid membrane. There are at least four routes for proteins to cross or become integrated into the thylakoid membrane. They are:

(i) **Sec-dependent:** This pathway uses proteins that are homologous to the bacterial Sec proteins for translocating proteins across the bacterial plasma membrane. It imports proteins into the thylakoid space (lumen) in an ATP dependent manner.

(ii) **SRP (Signal Recognition Particle) like pathway:** The chloroplast homologue of SRP binds to the protein signal sequence to import proteins into thylakoid membrane. This process requires ATP.

(iii) **ΔpH pathway**: It is dependent on a proton gradient across the thylakoid membrane to transport protein into thylakoid space.

(iv) **Spontaneous insertion:** Some proteins with a signal sequence specific for thylakoid membrane can insert spontaneously. It does not depend on a protein translocator for membrane integration.

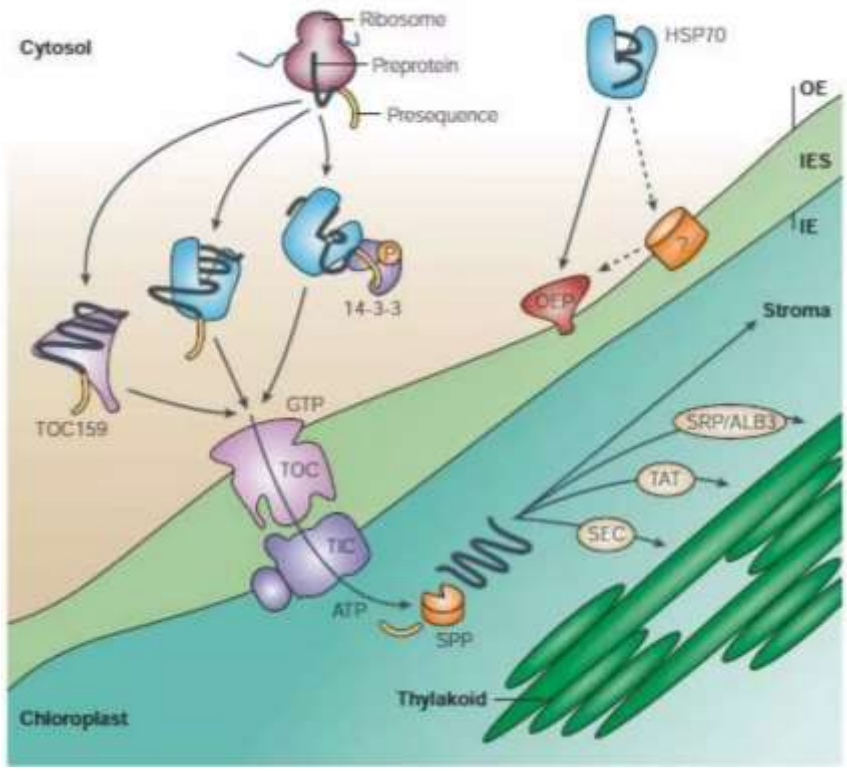


Fig.: Transport of proteins into Chloroplast.

Transport of proteins from E.R. through the golgi apparatus:

The signal sequences cover about twenty amino acids, including a stretch of hydrophobic residues, usually at the amino terminus of the polypeptide chain. As soon as the signal sequences of the growing polypeptide chain emerge from the ribosome, they are recognised and bound by a signal recognition particle (SRP) consisting of six polypeptides and a small cytoplasmic RNA (srp RNA). Then the complex containing the growing polypeptide chain, ribosome, and SRP is specifically targeted to the endoplasmic reticulum membrane by an interaction with a membrane-bounrd receptor, the SRP receptor or docking protein. In the next step, the SRP is released from both the ribosome and the signal sequence, where GTP (guanosine triphosphate) plays a key role. The ribosome then binds to a protein translocation complex in the membrane of the endoplasmic reticulum, and the signal sequence is inserted into a membrane channel or translocon. The translocons are complexes of three transmembrane proteins, known as Sec61 proteins. Transfer of the ribosome from the SRP to the translocon allows translation to resume, and the growing polypeptide chain is transferred directly into the translocon channel and across the membrane of the endoplasmic reticulum as translation proceeds. As translocation proceeds, the signal sequence is cleaved by the signal peptidase and the polypeptide is released into the lumen of the endoplasmic reticulum. Finally, GTP hydrolysis leads to the dissociation of the SRP from its receptor, and a new targeting cycle can begin. The actual transfer of the polypeptide through the membrane does not require the SRP or its receptor and commences only after their disengagement.

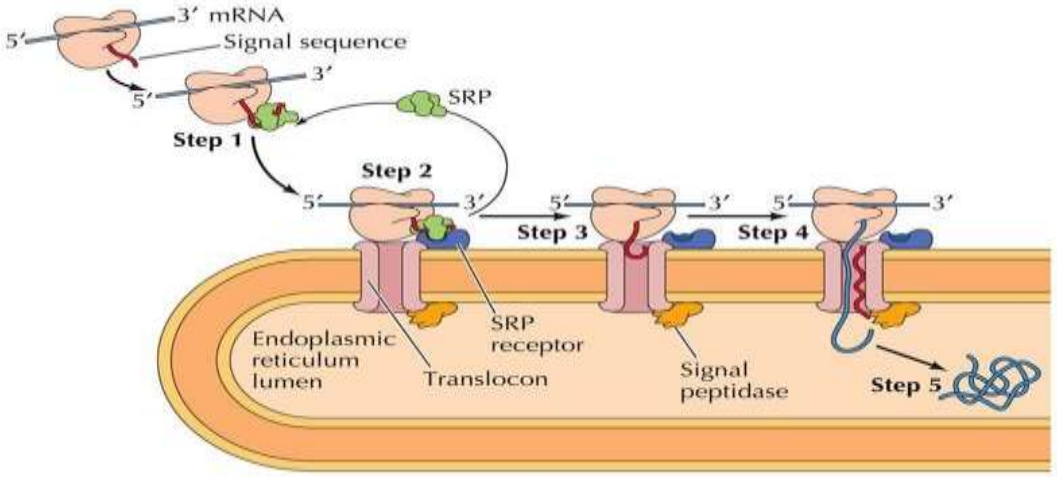
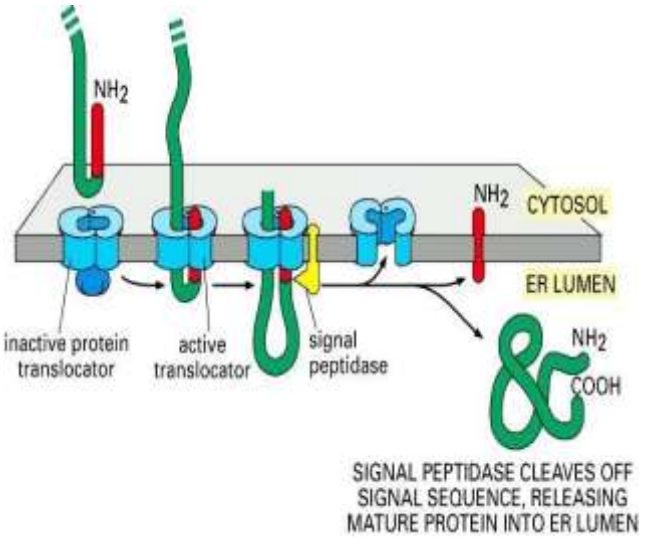


Fig.: The co-translational pathway of transport of secretory proteins to the endoplasmic reticulum.

Two basic functions are done by the SRP, where first it targets the polypeptide chain to the Endoplasmic reticulum membrane by interacting both with the signal sequence and with the translocation apparatus and secondly it keeps the bound signal sequence segregated from the rest of the polypeptide chain and thereby prevents aberrant, premature folding.

Some proteins in mammals and many proteins in yeast are transported through post-translational pathway. These proteins are synthesized on free cytosolic ribosomes and these proteins do not require a signal recognition particle (SRP) for their transport. Their signal sequences are recognised by distinct receptor proteins associated with the translocon in the endoplasmic reticulum membrane. The polypeptide chains are remained in an unfolded conformation by the cytosolic Hsp70 chaperones.



ER and Protein Trafficking

Signal Sequence is removed from Soluble Proteins

Two signalling functions:

1) Directs protein to ER membrane
2) Serves as “start transfer signal” to open pore

Signal peptidase removes terminal ER signal sequence upon release of protein into the lumen

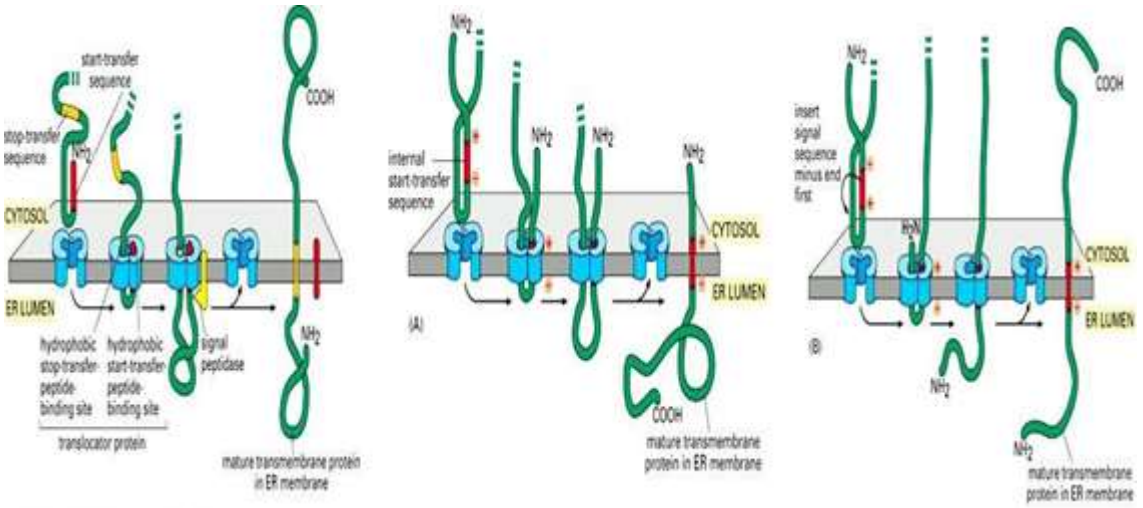
Single Pass Transmembrane Proteins

1. N-terminal signal sequence initiates trans-location and additional hydrophobic “stop sequence anchors protein in membrane

2. Signal sequence is internal and remains in lipid bilayer after release from translocator

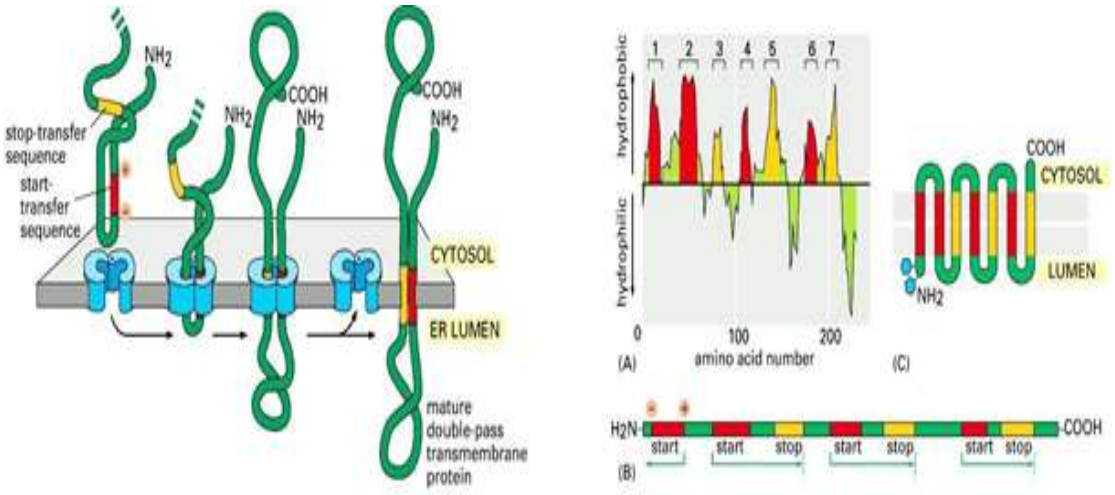
3. Internal signal sequence in opposite orientation

4. Orientation of start-transfer sequence governed by distribution of nearby charged aa



Multipass Transmembrane Proteins

- Combinations of start- and stop-transfer signals determine topology
- Whether hydrophobic signal sequence is a start- or stop-transfer sequence depends upon its location in polypeptide chain
- All copies of same polypeptide have same orientation



Protein sorting and export from Golgi apparatus:

Proteins, as well as lipids and polysaccharides, are transported from the Golgi apparatus to their final destinations through the secretory pathway. This involves the sorting of proteins into different kinds of transport vesicles, which bud from the trans Golgi network and deliver their contents to the appropriate cellular locations.

Some proteins are carried from the Golgi to the plasma membrane by a constitutive secretory pathway, which accounts for the incorporation of new proteins and lipids into the plasma membrane, as well as for the continuous secretion of proteins from the cell.

Other proteins are transported to the cell surface by a distinct pathway of regulated secretion or are specifically targeted to other intracellular destinations, such as lysosomes in animal cells or vacuoles in yeast.

The anterograde movement of proteins intended to be targeted through the secretory pathway is:

Rough ER → ER-to-Golgi transport vesicles → Golgi body (cisternae) → trans Golgi network (secretory or transport vesicles) → cell surface → exocytosis

Transport from the Golgi apparatus takes place by two pathway -

a. Constitutive secretory pathway:

Proteins are secreted from a cell continuously, regardless of external signals or factors. Proteins are stored in vesicles in the Golgi and move directly to the cell surface and fuse with the PM and release the soluble proteins.

b. Regulated secretory pathway: A distinct regulated secretory pathway in which specific proteins are secreted in response to environmental signals. Proteins are sorted in trans Golgi network and packed into secretory vesicles. These secretory vesicles are usually larger than other transport vesicles. and stored until specific signals are received, then fuse with plasma membrane to release the proteins.

Transport from the Trans Golgi Network to the Cell Exterior: Exocytosis

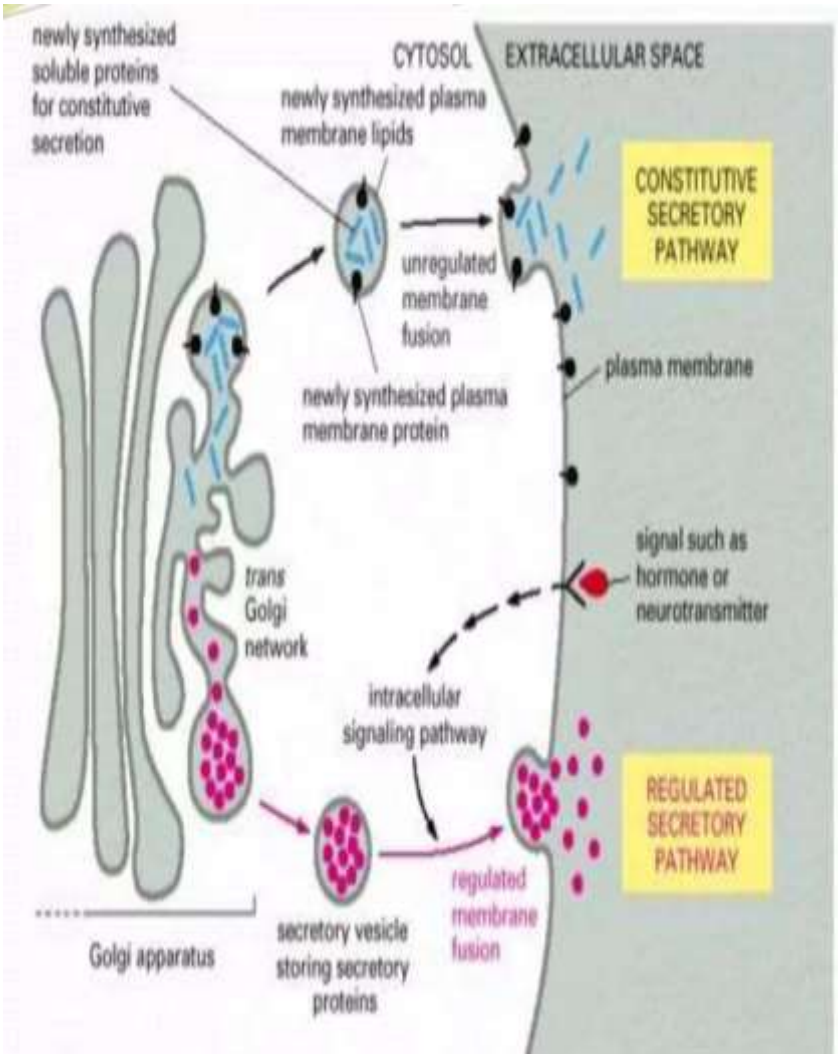


Fig.: The constitutive and regulated secretory pathways.

Vesicular Tubular Clusters (VTC): VTC are formed when ER-derived vesicles fuse with one another to form large compartment These compartments differ from ER and lacks many functional proteins present in ER They are generated continually and function as transport packages that bring material from the ER to the Golgi apparatus. The clusters are relatively short-lived because they quickly move along microtubules to the Golgi apparatus, where they fuse and deliver their contents.

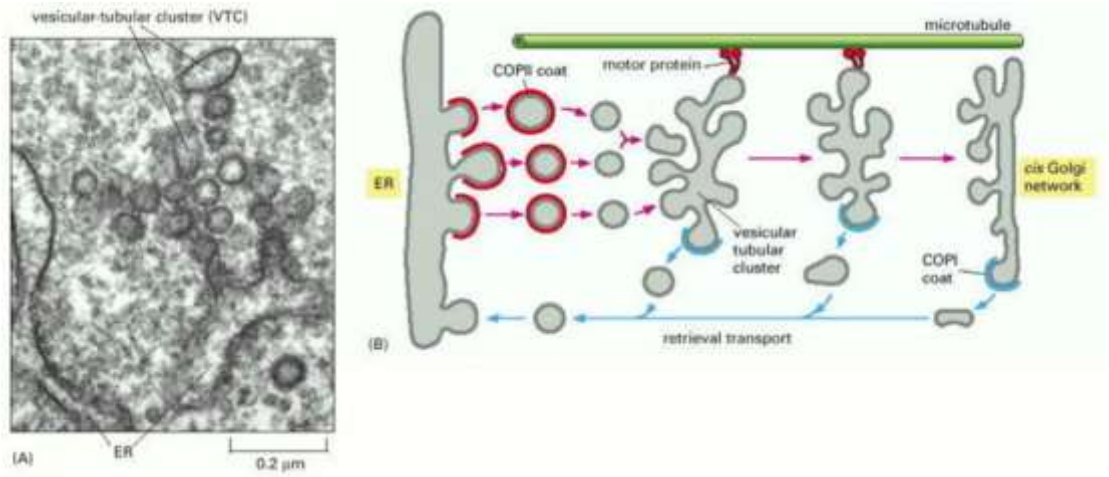


Fig.: (A)-An electron micrograph section of vesicular tubular clusters forming from the ER membrane
(B)- Vesicular tubular clusters (VTC) move along microtubules to carry proteins from ER to GA. COPI coats mediate the budding of vesicles that return to ER from these clusters. Coats disassemble after vesicles have formed.

Vesicular transport:

It is a process in which membrane enclosed transport vesicles transport proteins from one membrane enclosed compartment to another proteins do not move across the lipid bilayer of any membranes Move from Lumen of ER to lumen of Golgi to exterior of the cell. **Principles of vesicular transport**

1. A protein-coated membrane-enclosed transport vesicle buds off from the membrane of donor compartment carrying a variety of specifically selected cargo molecules.

2. Transport vesicle binds to the target compartment and fuse with the membrane of the target compartment.

3. Cargo molecules transfer into lumen of the target compartment and inserting the vesicular membrane components into the target compartment membrane.

Steps a. Budding, b.Uncoating, c.Transport, d.Docking, e.Fusion

Transport vesicles can recognize the correct target membrane with which to fuse.

The recognition step is thought to be controlled mainly by two classes of proteins

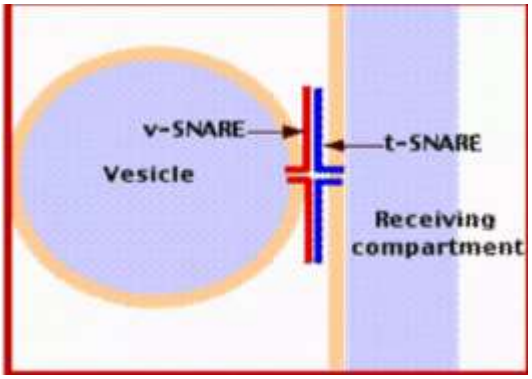
i) SNAREs (Soluble NSF Attachment Protein Receptor) play a central role both in providing specificity and in catalyzing the fusion of vesicles with the target membrane. This involves pairs of complementary integral membrane SNAREs proteins.

NSF (N-ethylmaleimide sensitive fusion)

SNARE proteins are fusion proteins that regulate docking of granules and vesicles to target membranes including the plasma membrane

v-SNAREs = "vesicle SNAREs" — on the vesicle surface.

t-SNAREs = "target SNAREs" — on the surface of the target membrane.



v-SNAREs and t-SNAREs bind specifically to each other, Binding is followed by fusion of the two membranes.

ii) Rabs (GTPase) - work together with other proteins to regulate the initial docking and tethering of the vesicle to the target membrane.

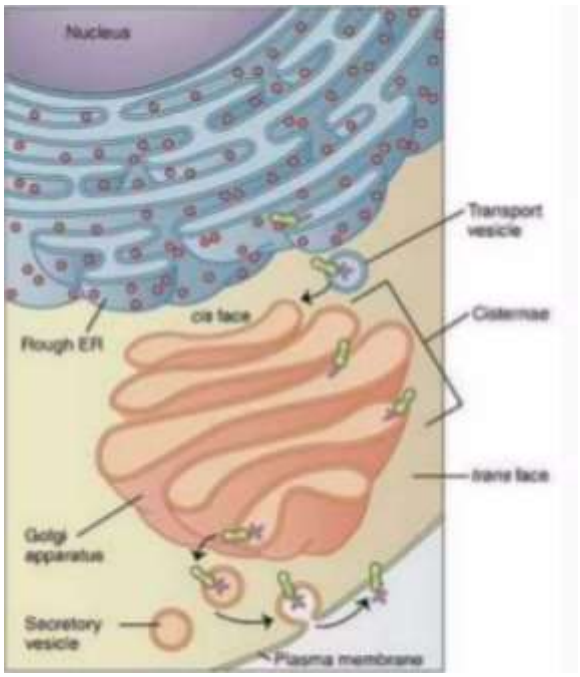
Small GTPase proteins play essential roles in the regulation of vesicular trafficking systems in eukaryotic cells. Two types of small GTPases, secretion-associated Ras-related protein (Sar) and ADP-ribosylation factor (Arf), act in the biogenesis of transport vesicles. Sar/Arf GTPases function as molecular switches by cycling between active, GTP-bound and inactive, GDP-bound forms, catalyzed by guanine nucleotide exchange factors and GTPase-activating proteins, respectively. Activated Sar/Arf GTPases undergo a conformational change, exposing the N-terminal amphipathic α-helix for insertion into membranes. The process triggers the recruitment and assembly

of coat proteins to the membranes, followed by coated vesicle formation Sar1 protein strictly controls anterograde transport from the endoplasmic reticulum (ER) through the recruitment of plant COPII coat components onto membranes. COPII vesicle transport is responsible for the organization of highly conserved polygonal ER networks.

There are two types of vesicles

- **Inbound path**: Using a variety of signals, the Golgi separates the products from the processing enzymes that made them and returns the enzymes back to the endoplasmic reticulum. This transport is also done by pinching off vesicles, but the inbound vesicles are coated with COPI (coat protein I) .
- **Out bound Path:** Transition vesicles pinch off from the surface of the endoplasmic reticulum carrying integral membrane proteins soluble proteins awaiting processing enzymes

Pinching off requires that the vesicle be coated with COPII (Coat Protein II). The transition vesicles move toward the cis Golgi on microtubules. As they do so, their COPII coat is removed and they may fuse together forming larger vesicles. These fuse with the cis Golgi Sugars is added to proteins in small packets so many glycoproteins have to undergo a large number of sequential steps of glycosylation, each requiring its own enzymes. These steps take place as shuttle vesicles carry the proteins from cis to medial to the trans Golgi compartments. At the outer face of the trans Golgi, vesicles pinch off and carry their completed products to their various destinations.



Vesicular transport is active movement of substances across the cell membrane through vesicles. Vesicles form from the

Endoplasmic Reticulum,

Golgi Apparatus, &

Plasma Membrane

Vesicles leaving RER transported to the cis face of GA, fuse with the membrane and empty the contents into the lumen. Molecules inside the lumen are modified and sorted for transport to the next destination. Proteins destined for places other than ER and GA, moves to trans face. Gets placed on either of the 3 vesicles, i**.e. Exocytotic , Secretory and Lysosomal vesicles.**

Transport from the ER to the Golgi apparatus is mediated by vesicular tubular clusters:

Transport vesicles after budding from an ER exit site, shed their coat and began to fuse with one another. Fusion of membranes from same compartment is called homotypic fusion.

It differs from heterotypic fusion in which a membrane from one compartment fuses with the membrane of different compartment.

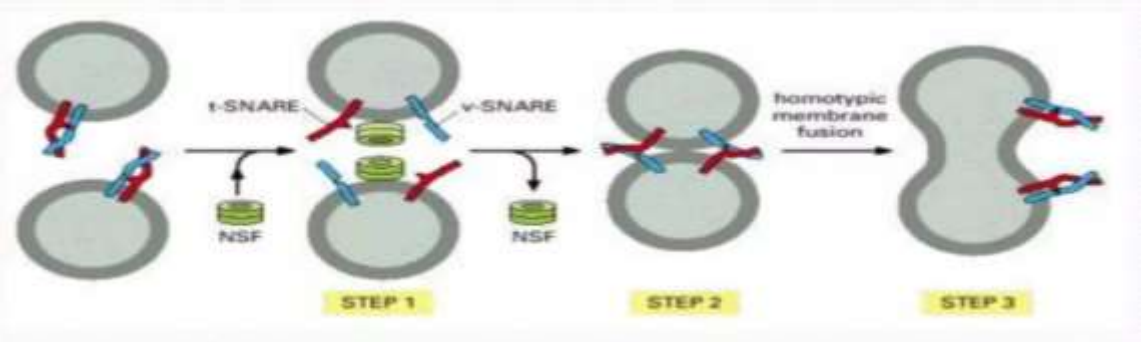


Fig.: Homotypic membrane fusion.

In step 1, identical pairs of v-SNAREs and t-SNAREs in both membranes are pried apart by NSF. (N- ethylmaleimide sensitive fusion)

In steps 2 and 3, the separated matching SNAREs on adjacent identical membranes interact, which leads to membrane fusion and the formation of one continuous compartment.

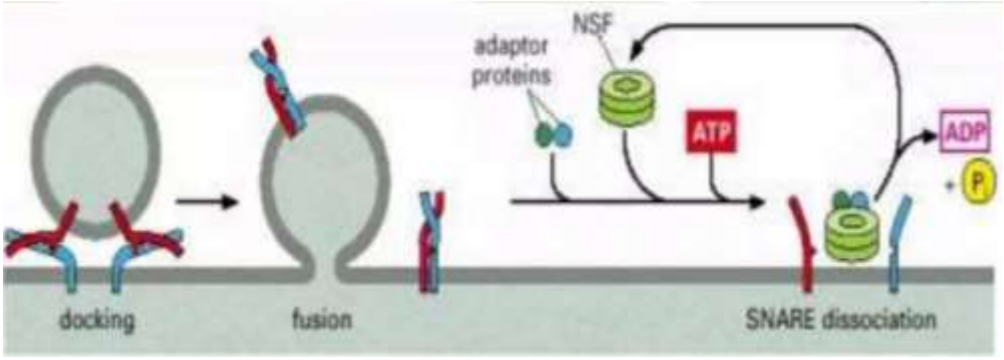


Fig.: Dissociation of SNARE pairs by NSF after a membrane fusion cycle is completed.

After the v-SNAREs and t-SNAREs have mediated the fusion of a vesicle on a target membrane, the NSF binds to the SNARE complex via adaptor proteins and hydrolyzes ATP to pry the SNAREs apart.

Only proteins that are properly folded and assembled can leave the ER:

Proteins must be properly folded and if they are subunits of multimeric protein complexes, must be completely assembled to exit from Endoplasmic Reticulum. Misfolded or incompletely assembled proteins are retained in ER, where they are bound to chaperone proteins such as Bip (Binding immunoglobulin protein) or calexin. Such proteins are transported back to the cytosol where they are degraded by proteasomes Such misfolded and incompletely assembled proteins may interfere with normal proteins when they are transported forward from ER to Golgi body. The cells make a large excess of many protein molecules from which those which fold and assemble properly are selected.

Retrieval pathway to ER uses sorting signals:

Resident ER membrane proteins (misfolded or not completely processed proteins) move back from Golgi apparatus to ER, by retrieval pathway. These proteins contain retrieval signals (two lysines followed by any two other amino acids at the extreme C-

terminal end of the ER membrane protein) that bind directly to COPI coats and are packed into COPI-coated vesicles . It is called a KKXX sequence, based on the single letter amino acid code. The retrieval (retrograde) continues as vesicular tubular clusters. Selected proteins are returned to ER, after vesicular tubular clusters have delivered their cargo.

Soluble ER resident proteins, such as BiP, also contain a short retrieval signal at their C-terminal end, but it is different: it consists of a **Lys-Asp-Glu-Leu** or similar sequence. If this signal (**called the KDEL sequence**) is removed from BiP by genetic engineering, the protein is slowly secreted from the cell. If the signal is transferred to a protein that is normally secreted, the protein is now efficiently returned to the ER, where it accumulates.

Unlike the retrieval signals on ER membrane proteins that can interact directly with the COPI coat, soluble ER resident proteins must bind to specialized receptor proteins such as the KDEL receptor—a multipass transmembrane protein that binds to the KDEL sequence and packages any protein displaying it into COPI-coated retrograde transport vesicles. To accomplish this task, the KDEL receptor itself must cycle between the ER and the Golgi apparatus.

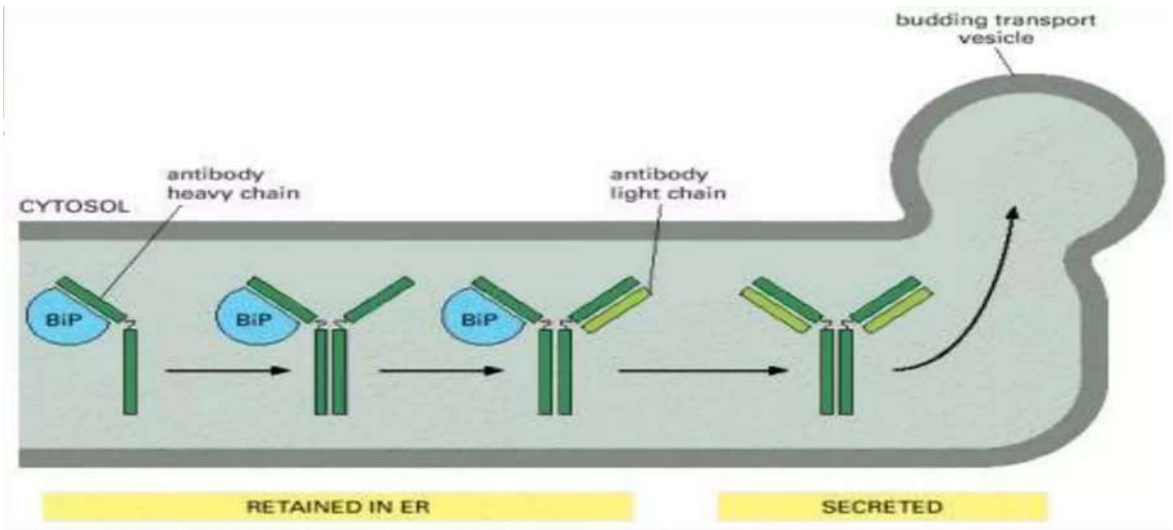


Fig.: Retention of incompletely assembled antibody molecules in the ER.

Antibodies are made up of two heavy and two light chains which assemble in the ER. The chaperone BiP is thought to bind to all incompletely assembled antibody molecules and to cover up an exit signal. Only completely assembled antibodies leave the ER and are secreted.

Newly synthesized proteins enter the biosynthetic- secretory pathway in the ER by crossing the ER membrane from the cytosol. During their subsequent transport, from the ER to the Golgi apparatus and from the Golgi apparatus to the cell surface and elsewhere, these proteins pass through a series of compartments, where they are successively modified. Transfer from one compartment to the next involves a delicate balance between forward and backward (retrieval) transport pathways. Some transport vesicles select cargo molecules and move them to the next compartment in the pathway, while others retrieve escaped proteins and return them to a previous compartment where they normally function. The pathway from the ER to the cell surface involves many sorting steps, which continually select membrane and soluble lumenal proteins for packaging and transport—in vesicles or organelle fragments that bud from the ER and Golgi apparatus.

Small membrane vesicles are created by the action of coat proteins that deform membranes into the shape of vesicles and simultaneously select cargo proteins for inclusion into these vesicles

There are three types of protein coated vesicles

1. Clathrin coated vesicles mediate transport from TGN (trans Golgi network) to endosome. These vesicles bud from the trans-Golgi network and the plasma membrane and then fuse with late endosomes.

2. COP I coated vesicles mediate transport retrograde movement from Golgi cisterna to ER (8 protein complex including ARF GTPase). COPI vesicles mediate retrograde transport from trans Golgi to medial to cis-golgi, as well as from the cis-Golgi/cis-Golgi network to the rough ER. It may also mediate forward transfer of vesicles from the rough ER to the cis-Golgi network

3. COPII coated vesicles mediate transport from ER to cis-Golgi complex /cis-Golgi network (5 protein complex including SARI GTPase)

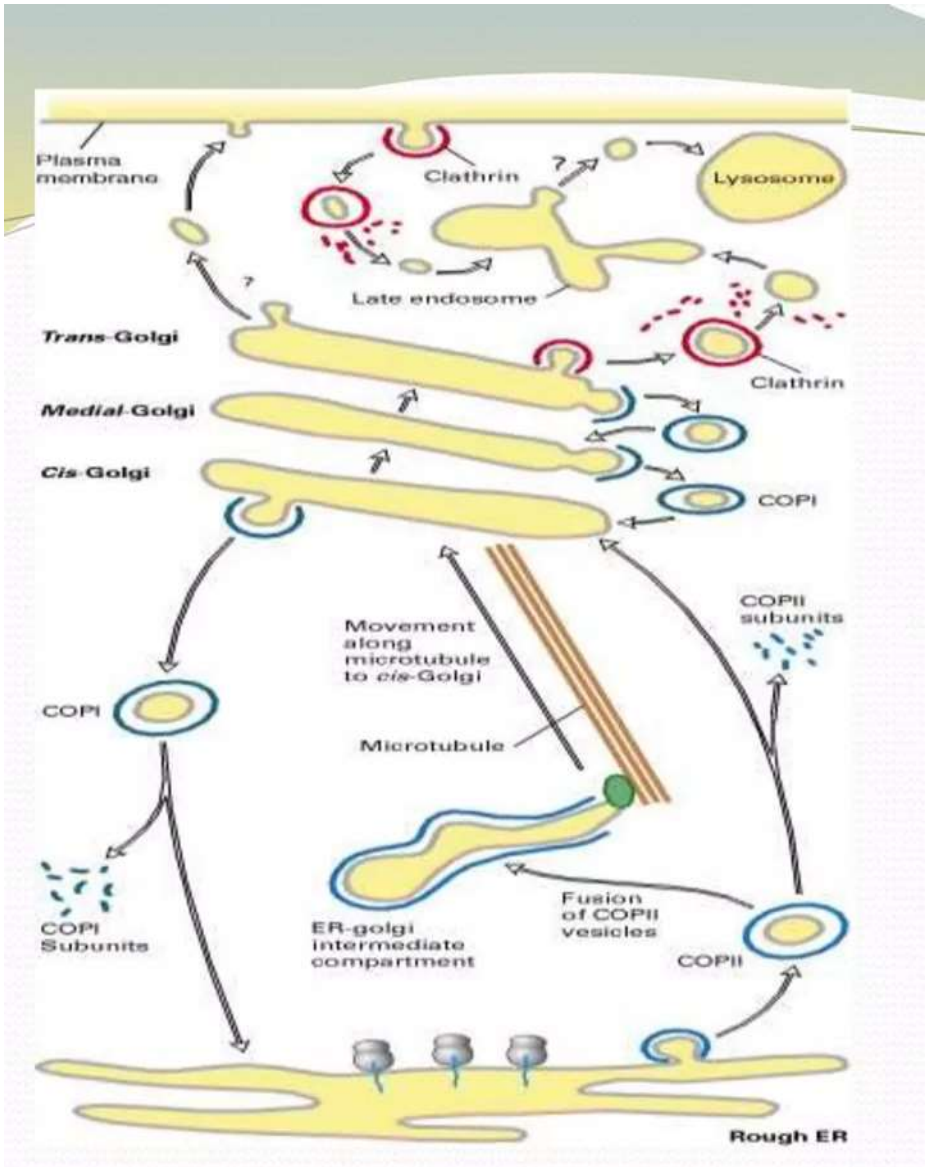


Fig.: 50 Involvement of the three known types of coat proteins.

Role of three types of coat proteins (**COP I, COP II and clathrin** in vesicular traffic in secretory and endocytic pathways):

COPII protein mediates transfer of vesicles from rough ER to cis-golgi/ cis-Golgi network COPI protein mediates retrograde transport from trans golgi to medial to cis-

golgi, as well as from the cis-Golgi/cis-Golgi network to the rough ER. It may also mediate forward transfer of vesicles from the rough ER to the cis-Golgi network

Clathrin mediates transfer of vesicles that bud from the trans-Golgi network and the plasma membrane and that then fuse with late endosomes.

Proteins leave the er in COPII-coated transport vesicles

Proteins that enter the ER and are destined for the Golgi apparatus or beyond are first packaged into small COPII-coated transport vesicles. These transport vesicles bud from specialized regions of the ER called ER exit sites, whose membrane lacks bound ribosomes Coat protein complex II (COPII) is a set of highly conserved proteins that is responsible for creating small membrane vesicles that originate from the endoplasmic reticulum COPII is a coatomer, a type of vesicle coat protein that transports proteins from the rough endoplasmic reticulum to the Golgi apparatus. The name "COPII" refers to the specific coat protein complex that initiates the budding process. COPII coat consists of five cytosolic proteins: Sar1, Sec23, Sec24, Sec13, and Sec31.

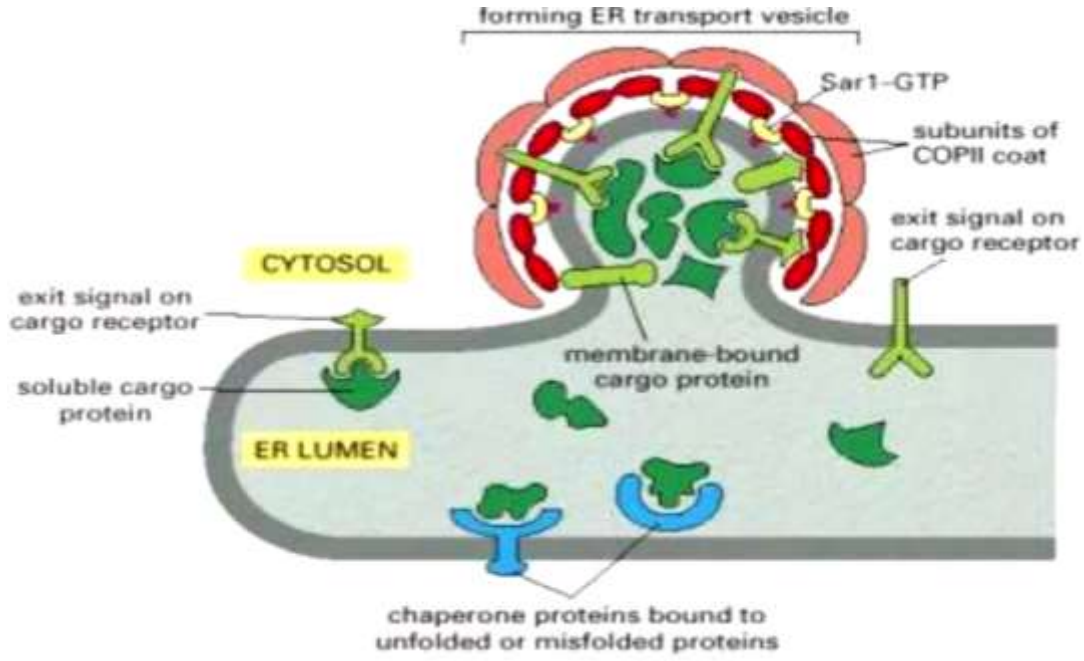


Fig.: The recruitment of cargo molecules into ER transport vesicles.

By binding to the COPII coat, membrane and cargo proteins become concentrated in the transport vesicles as they leave the ER. Membrane proteins are packaged into budding transport vesicles through the interactions of exit signals on their cytosolic tails with the COPII coat. Some of the membrane proteins trapped by the coat in turn function as cargo receptors, binding soluble proteins in the lumen and helping to package them into vesicles. A typical 50-nm transport vesicle contains about 200 membrane proteins, which can be of many different types. Unfolded or incompletely assembled proteins are bound to chaperones and are thereby retained in the ER compartment.

Role of M6P (Mannose 6-Phosphate) receptor in lysosomal enzyme sorting:

The lysosomal enzymes and lysosomal membrane proteins are synthesized in rough ER and transported to Golgi cisternae and ultimately to lysosomes. The sorting signal that directs the lysosomal enzymes from the trans- Golgi network (TGN) to lysosomes is **mannose 6-phosphate (M6P)**. The attachment of M6P to lysosomal enzymes prevents their further modification.

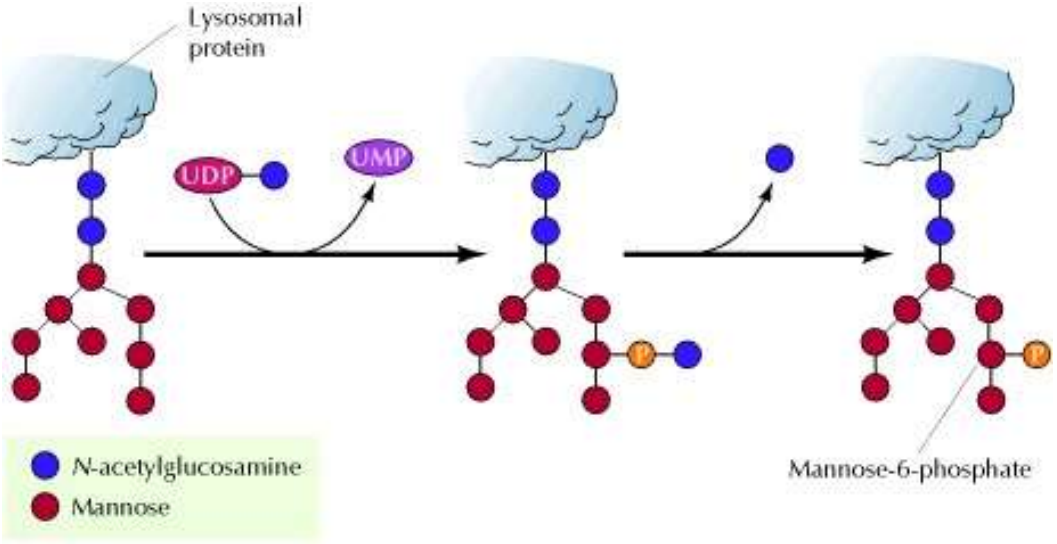


Fig.: Targeting of lysosomal proteins by phosphorylation of mannose residues.

The processing of N-linked oligosaccharides of lysosomal proteins differs from that of plasma membrane and secreted proteins. The lysosomal proteins are modified by **mannose phosphorylation.** First, there is addition of N-acetylglucosamine phosphates to specific mannose residues, and this happens probably while the protein is still in the cis Golgi network. After this N-acetylglucosamine group is removed, leaving mannose-6-phosphate residues on the N-linked oligosaccharide. The phosphorylated mannose residues are specifically recognised by a mannose-6- phosphate receptor in the trans Golgi network, which directs the transport of these proteins to lysosomes.

Separation of M6P bearing lysosomal enzymes from other proteins takes place in TGN. The wall of TGN contains M6P receptors. These M6P receptors bind to lysosomal proteins. The vesicles containing these receptor bearing proteins bud off from TGN. These vesicles are called lysosomes. Later these vesicles fuse with vesicles which have arisen by pinocytosis and phagocytic to form secondary lysosomes. Low pH of Lysosomes triggers the dissociation of enzymes from the receptors.

The M6P receptors are recycled back to trans-golgi network in vesicles. Lysosomes contain hydrolyzing proteolytic enzyme, which digests proteins meant for degradation. A protein named ubiquitin marks the proteins meant for destruction. Ubiquitin is present in all eukaryotic cells. This mechanism degrades only those proteins which are meant for destruction and not the proteins which are to be left alone.

Transport from the Plasma membrane to Endosomes:

Endocytosis:

Endocytic mechanisms control the lipid and protein composition of the plasma membrane, thereby regulating how cells interact with their environments. Endocytosis describes the de novo production of internal membranes from the plasma membrane lipid bilayer. In so doing, plasma membrane lipids and integral proteins and extracellular fluid become fully internalized into the cell. Endocytosis a key process in regulating processes such as mitosis, antigen presentation, and cell migration. There are number of endocytic mechanisms each of which involves unique proteins.

The endocytic pathway comprises two distinct kinds of **endosome, early endosomes** and **late endosomes**. Material taken up by endocytosis passes from the early

endosomes to the late endosomes and from there may intersect with trafficking pathways from the Golgi apparatus, or may be directed to lysosomes or to the Golgi. The exact pathway depends on the cell and the material that has been internalised. Endocytic mechanisms are differentiated into three subcategories: namely, macropinocytosis, phagocytosis, receptor-mediated endocytosis.

In **macropinocytosis** small soluble molecules get the uptake in the form of vesicles (0.5–5 μm in diameter), which usually occurs from highly ruffled regions of the plasma membrane. These vesicles are filled with a large volume of extracellular fluid and molecules within it. No specificity is required in filling of the vesicle. These vesicle travels into the cytosol and get fused with endosomes and lysosomes in later stages.

Phagocytosis is an endocytic process in which cells bind and internalize larger insoluble particles (0.75 μm in diameter) like cell debris, micro-organisms and apoptotic cells.

Receptor-mediated endocytosis is an endocytic uptake of specific molecule from the external medium using a receptor which generally present on the membrane surface and have specificity for the internalized molecule. In general, receptor-mediated endocytosis may occur by two different types that are **'clathrin-mediated'** and **'caveolae' endocytosis**.

Clathrin-mediated endocytosis (CME) is mediated by the production of small (approx. 100 nm in diameter) vesicles that contain a coat made up of the cytosolic protein 'clathrin'. In this process cargo recruited into developing clathrin coated pits (CCPs) and result in the formation of clathrin-coated vesicles (CCVs). Coated pits can concentrate large extracellular molecules that have different receptors responsible for the receptor-mediated endocytosis of ligands, e.g. low density lipoprotein, transferrin, growth factors, antibodies and many others. Clathrin consists of a heavy chain and a light chain and successively assemble into a polyhedral, cage-like coat on the surface of the coated pit. The clathrin coat is made of sub-assemblies, each consisting of a threepronged protein complex, a triskelion, each leg of which is made of one heavy and one light chain.

The triskelion forms a lattice-like network of hexagons and pentagons, which attaches to the membrane via an adaptor protein (AP) complex. Adaptor proteins bind both to clathrin and to integral membrane proteins which are destined to be internalized and stimulate its assembly. Much more importantly, by binding to the molecules in the membrane of the vesicle, adaptor proteins appear to be responsible for recognizing the appropriate cargo molecules.

Caveolae endocytosis is another common non-clathrin-coated endocytotic process in smooth muscle, type I pneumocytes, fibroblasts, adipocytes, and endothelial cells. In this process, plasma membrane buds are present which consist of the cholesterol-binding protein caveolin (Vip21) with a bilayer enriched in cholesterol and glycolipids. Caveolae are small (approx. 50 nm in diameter) flask-shape pits in the membrane that resemble the shape of a cave (hence the name caveolae). Extracellular molecules are taken-up by specific receptors in caveolae. Potocytosis is a kind of receptor-mediated endocytosis in which caveolae vesicles bring molecules of various sizes into the cell. However, material endocytosed via potocytosis is released into the cytosol not in lysosomes or any other organelle as done in other endocytosis processes.

Examples of Endocytosis.

Low density lipoprotein (LDL) acquisition by endocytic pathway:

Low density lipoprotein (LDL) consists of a spherical outer phospholipid shell having a single molecule of large apoB100 protein and the core in which cholesterol is packed in the form of cholesteryl esters. Cells membrane contains surface receptors that specifically bind to apoB-100 and internalize LDL particles by receptor-mediated endocytosis. Cytosolic tail of the LDL receptor contains NPXY sorting signal for the binding with AP2 complex that help in the incorporation of receptor-ligand complex into developing endocytic vesicles. Dynamin-mediated mechanism is responsible for clathrin-coated pits (or buds) containing receptor-LDL complexes to pinched off. After internalization, the LDL particles are transported to endosome and the acidic pH in late endosome causes a conformational change in the LDL receptor that is responsible for the release of the bound LDL particle. The late endosome fuses with the lysosome, and

the proteins and lipids of the free LDL particle then broken down to their constituent parts by enzymes in the lysosome. LDL receptors, which dissociate from their ligands in the late endosome, recycle to the cell surface where at the neutral pH of the exterior medium the receptor undergoes a conformational change so that it can bind another LDL particle. The LDL receptor makes one round trip into and out of the cell every 10–20 minutes, for a total of several hundred trips in its 20-hour life span.

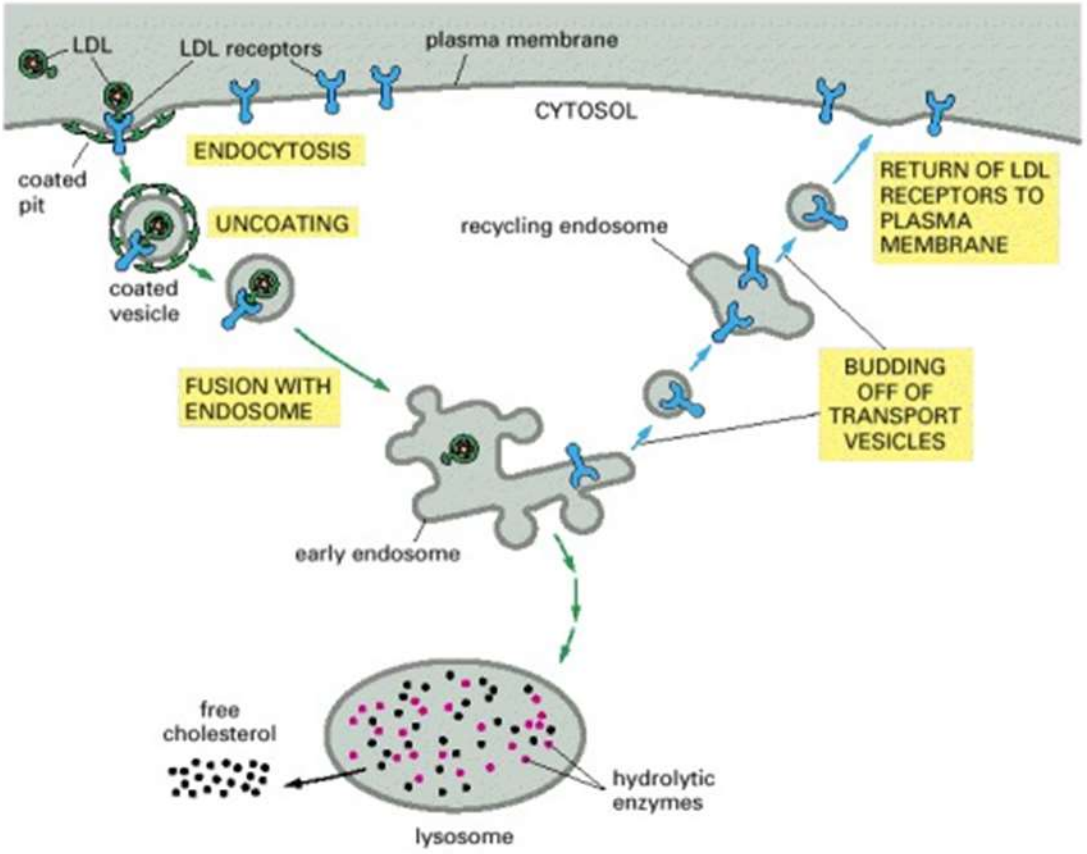


Fig.: The receptor-mediated endocytosis of LDL.

13. Cell Signaling: Cell surface and intracellular receptors; Ion channel linked, Signaling via Gprotein linked cell surface receptors.

Intracellular receptors are located in the cytoplasm of the cell and are activated by hydrophobic ligand molecules that can pass through the plasma membrane. Cell-surface receptors bind to an external ligand molecule and convert an extracellular signal into an intracellular signal.

Types of Receptors

Receptors are protein molecules in the target cell or on its surface that bind ligands. There are two types of receptors: internal receptors and cell-surface receptors.

Internal receptors:

Internal receptors, also known as intracellular or cytoplasmic receptors, are found in the cytoplasm of the cell and respond to hydrophobic ligand molecules that are able to travel across the plasma membrane. Once inside the cell, many of these molecules bind to proteins that act as regulators of mRNA synthesis to mediate gene expression. Gene expression is the cellular process of transforming the information in a cell's DNA into a sequence of amino acids that ultimately forms a protein. When the ligand binds to the internal receptor, a conformational change exposes a DNA-binding site on the protein. The ligand-receptor complex moves into the nucleus, binds to specific regulatory regions of the chromosomal DNA, and promotes the initiation of transcription. Internal receptors can directly influence gene expression without having to pass the signal on to other receptors or messengers.

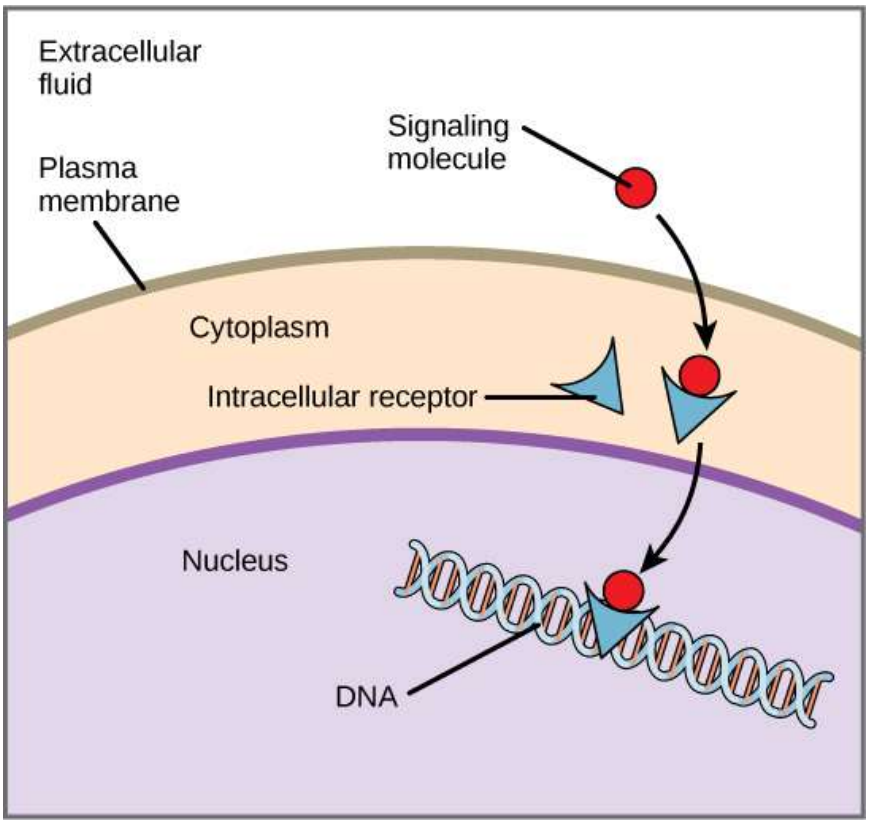


Fig.: Intracellular Receptors: Hydrophobic signaling molecules typically diffuse across the plasma membrane and interact with intracellular receptors in the cytoplasm. Many intracellular receptors are transcription factors that interact with DNA in the nucleus and regulate gene expression.

Cell-Surface Receptors:

Cell-surface receptors, also known as transmembrane receptors, are cell surface, membrane-anchored, or integral proteins that bind to external ligand molecules. This type of receptor spans the plasma membrane and performs signal transduction, converting an extracellular signal into an intracellular signal. Ligands that interact with cell-surface receptors do not have to enter the cell that they affect. Cell-surface receptors are also called cell-specific proteins or markers because they are specific to individual cell types.

Each cell-surface receptor has three main components: an external ligand-binding domain (extracellular domain), a hydrophobic membrane-spanning region, and an intracellular domain inside the cell. The size and extent of each of these domains vary widely, depending on the type of receptor.

Cell-surface receptors are involved in most of the signaling in multicellular organisms. There are three general categories of cell-surface receptors: ion channel-linked receptors, G-protein-linked receptors, and enzyme-linked receptors.

Ion Channel-Linked Receptors

Ion channel-linked receptors bind a ligand and open a channel through the membrane that allows specific ions to pass through. To form a channel, this type of cell-surface receptor has an extensive membrane-spanning region. In order to interact with the phospholipid fatty acid tails that form the center of the plasma membrane, many of the amino acids in the membrane-spanning region are hydrophobic in nature. Conversely, the amino acids that line the inside of the channel are hydrophilic to allow for the passage of water or ions. When a ligand binds to the extracellular region of the channel, there is a conformational change in the protein's structure that allows ions such as sodium, calcium, magnesium, and hydrogen to pass through.

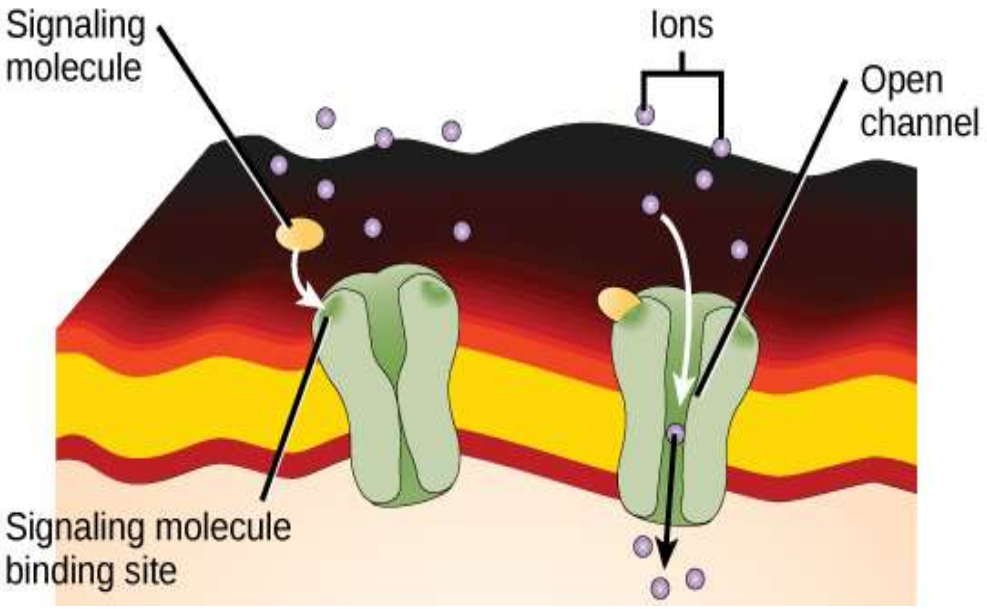


Fig.: Gated-Ion Channels: Gated ion channels form a pore through the plasma membrane that opens when the signaling molecule binds. The open pore then allows ions to flow into or out of the cell.

G-Protein Linked Receptors

G-protein-linked receptors bind a ligand and activate a membrane protein called a G-protein. The activated G-protein then interacts with either an ion channel or an enzyme in the membrane. All G-protein-linked receptors have seven transmembrane domains, but each receptor has its own specific extracellular domain and G-protein-binding site.

Cell signaling using G-protein-linked receptors occurs as a cyclic series of events. Before the ligand binds, the inactive G-protein can bind to a newly-revealed site on the receptor specific for its binding. Once the G-protein binds to the receptor, the resultant shape change activates the G-protein, which releases GDP and picks up GTP. The subunits of the G-protein then split into the α subunit and the β subunit. One or both of these G-protein fragments may be able to activate other proteins as a result. Later, the GTP on the active α subunit of the G-protein is hydrolyzed to GDP and the β subunit is deactivated. The subunits reassociate to form the inactive G-protein, and the cycle starts over.

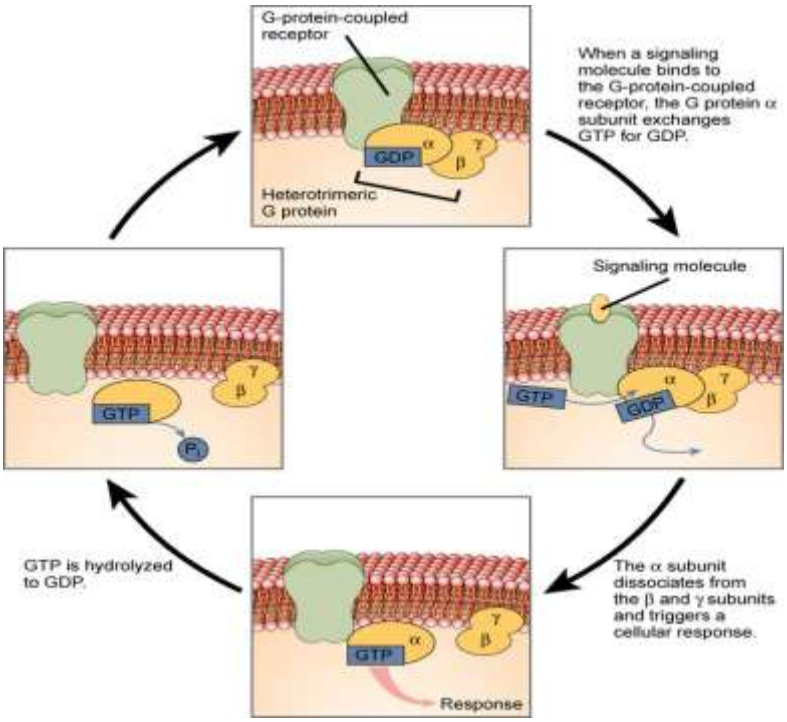


Fig.: G-proteins: Heterotrimeric G proteins have three subunits: α, β, and γ. When a signaling molecule binds to a G-protein-coupled receptor in the plasma membrane, a GDP molecule associated with the α subunit is exchanged for GTP. The β and γ subunits dissociate from the α subunit, and a cellular response is triggered either by the α subunit or the dissociated β pair. Hydrolysis of GTP to GDP terminates the signal.

Enzyme-Linked Receptors

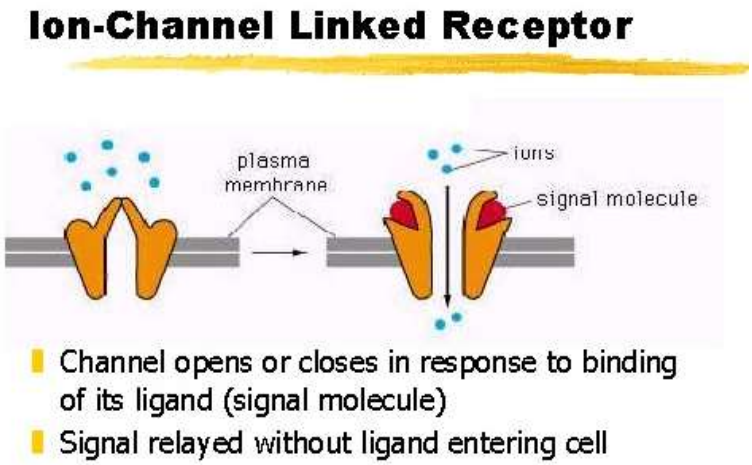
Enzyme-linked receptors are cell-surface receptors with intracellular domains that are associated with an enzyme. In some cases, the intracellular domain of the receptor itself is an enzyme or the enzyme-linked receptor has an intracellular domain that interacts directly with an enzyme. The enzyme-linked receptors normally have large extracellular and intracellular domains, but the membrane-spanning region consists of a single alpha-helical region of the peptide strand. When a ligand binds to the extracellular domain, a signal is transferred through the membrane and activates the enzyme, which sets off a chain of events within the cell that eventually leads to a response. An example

of this type of enzyme-linked receptor is the tyrosine kinase receptor. The tyrosine kinase receptor transfers phosphate groups to tyrosine molecules. Signaling molecules bind to the extracellular domain of two nearby tyrosine kinase receptors, which then dimerize. Phosphates are then added to tyrosine residues on the intracellular domain of the receptors and can then transmit the signal to the next messenger within the cytoplasm.

Ion -channel-linked receptors bind a ligand and open a channel through the membrane that allows specific ions to pass through.

Ion channel-linked receptors and cell signalling:

Also called transmittergated ion channels or ionotropic receptors. Found in nerve and muscle cells (electrically excitable). This type of signaling is mediated through neurotransmitters.



When a ligand binds to the extracellular region of the ionchannel-linked receptors, there is a conformational change in the receptor protein's structure that allows ions such as sodium, calcium, magnesium, and hydrogen to pass through.

Signaling via Gprotein linked cell surface receptors.

G-protein coupled receptors are cell surface receptors that pass on the signals that they receive with the help of guanine nucleotide binding proteins (G-proteins). Before thinking any further about the signaling pathways downstream of GPCRs, it is necessary

to know a few important facts about these receptors and the G-proteins that assist them. Though there are hundreds of different G-protein coupled receptors, they all have the same basic structure: they all consist of a single polypeptide chain that threads back and forth seven times through the lipid bilayer of the plasma membrane. For this reason, they are sometimes called seven- pass transmembrane (7TM) receptors.

One end of the polypeptide forms the extracellular domain that binds the signal while the other end is in the cytosol of the cell.

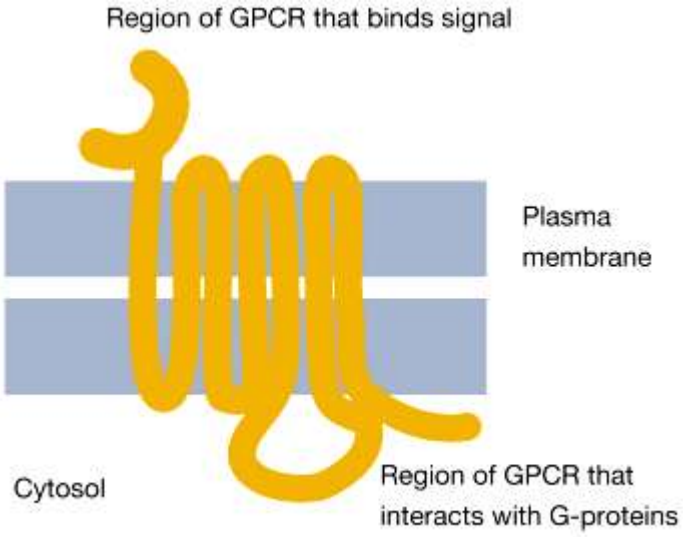


Fig.: *G*-protein coupled receptor.

A G-protein is a guanine nucleotide-binding protein that can interact with a G-protein linked receptor. G-proteins are associated with the cytosolic side of the plasma membrane, where they are ideally situated to interact with the cytosolic tail of the GPCR, when a signal binds to the GPCR.

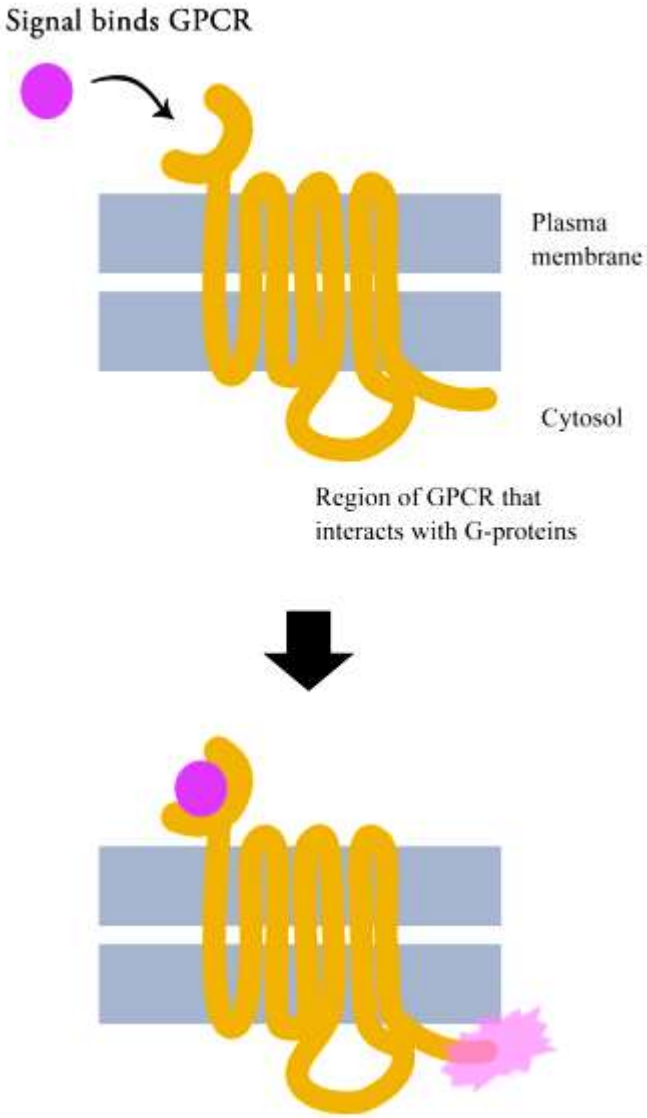


Fig.: *G*-protein coupled receptor signalling.

There are many different G-proteins, all of which share a characteristic structure- they are composed of three subunits called alpha, beta and gamma (aß.). Because of this, they are sometimes called heterotrimeric G proteins (hetero=different, trimeric= having three parts). The alpha subunit of such proteins can bind GDP or GTP and is capable of hydrolyzing a GTP molecule bound to it into GDP. In the unstimulated state of the cell, that is, in the absence of a signal bound to the GPCR, the G-proteins are found in the trimeric form (aβ bound together) and the alpha subunit has a GDP molecule bound to it.

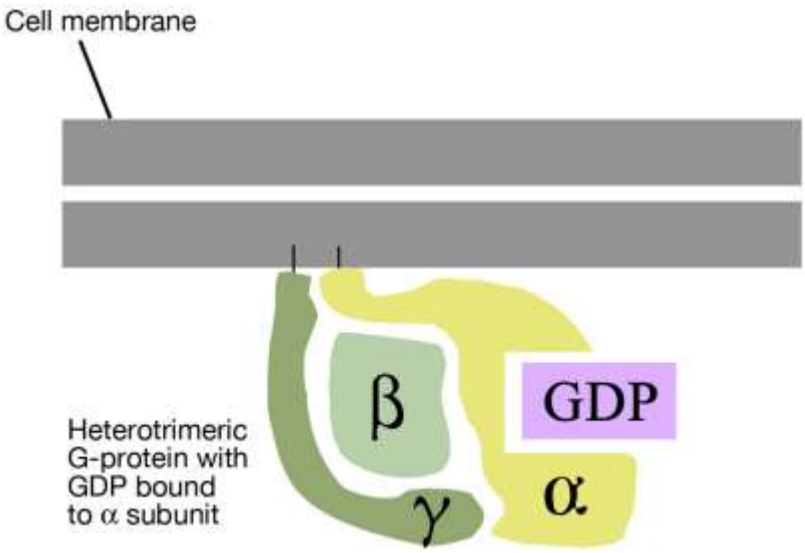


Fig.: *G*-protein with GDP bound.

With this background on the structure and general properties of the GPCRs and the G-proteins, it is necessary to know when a signal arrives at the cell surface and binds to a GPCR. The binding of a signal molecule by the extracellular part of the G-protein linked receptor causes the cytosolic tail of the receptor to interact with, and alter the conformation of, a G-protein. This has two consequences:

- First, the alpha subunit of the G- protein loses its GDP and binds a GTP instead.
- Second, the G-protein breaks up into the GTP-bound a part and the β part.

These two parts can diffuse freely along the cytosolic face of the plasma membrane and act upon their targets.

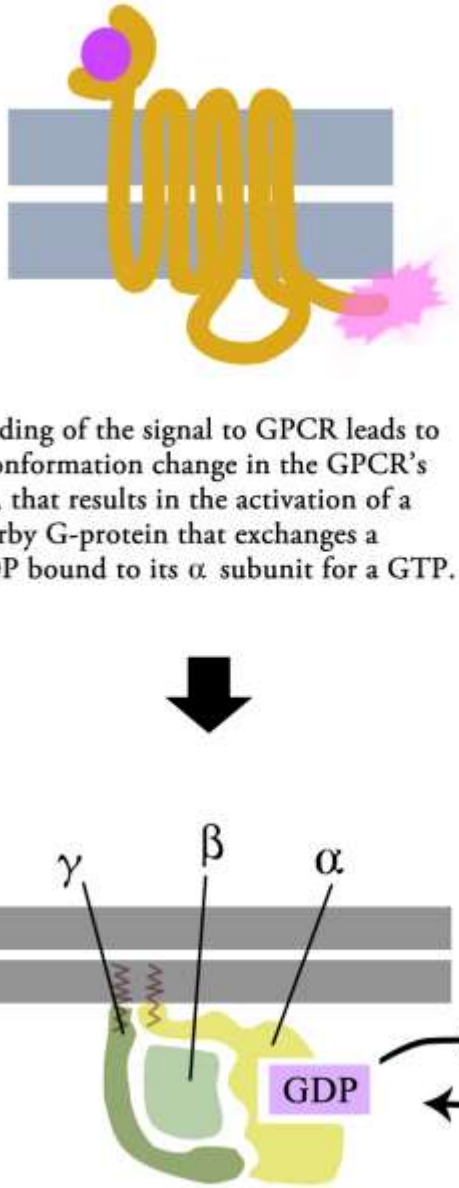


Fig.: *G*-protein activation.

The interaction of G-proteins with their target enzymes can regulate the activity of the enzyme, either increasing or decreasing its activity. Often the target enzyme will pass the signal on in another form to another part of the cell. As you might imagine, this kind of response takes a little longer than the kind where an ion channel is opened instantaneously. Two well-studied examples of enzymes whose activity is regulated by a

G-protein are adenylate cyclase and phospholipase C. When adenylate cyclase is activated, the molecule cAMP is produced in large amounts.

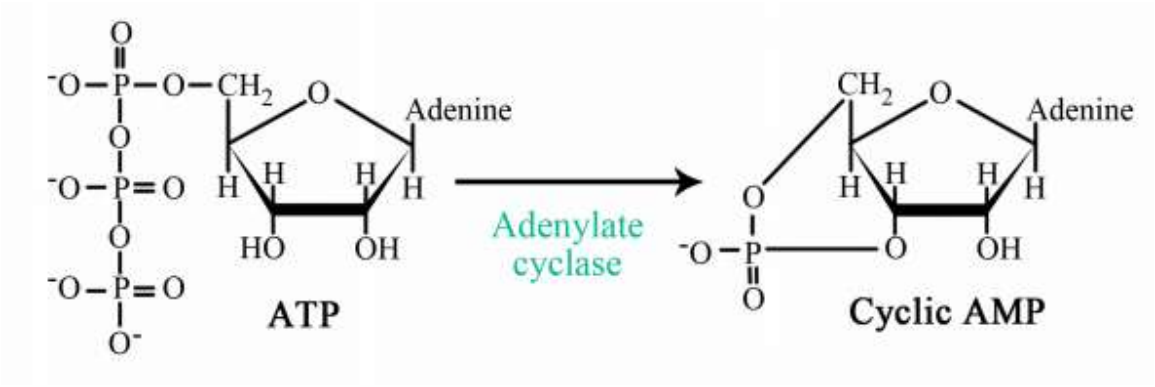


Fig.: Synthesis of cAMP.

When phospholipase C is activated, the molecules inositol trisphosphate (IP3) and diacylglycerol (DAG) are made. cAMP, IP3 and DAG are second messengers, small, diffusible molecules that can "spread the message" brought by the original signal, to other parts of the cell.

In these cases, the binding of a signal to the GPCR activated a G- protein, which in turn, activated an enzyme that makes a second messenger that can amplify the message in the cell. We will first trace the effects of activating adenylate cyclase and the resulting increase in cAMP.

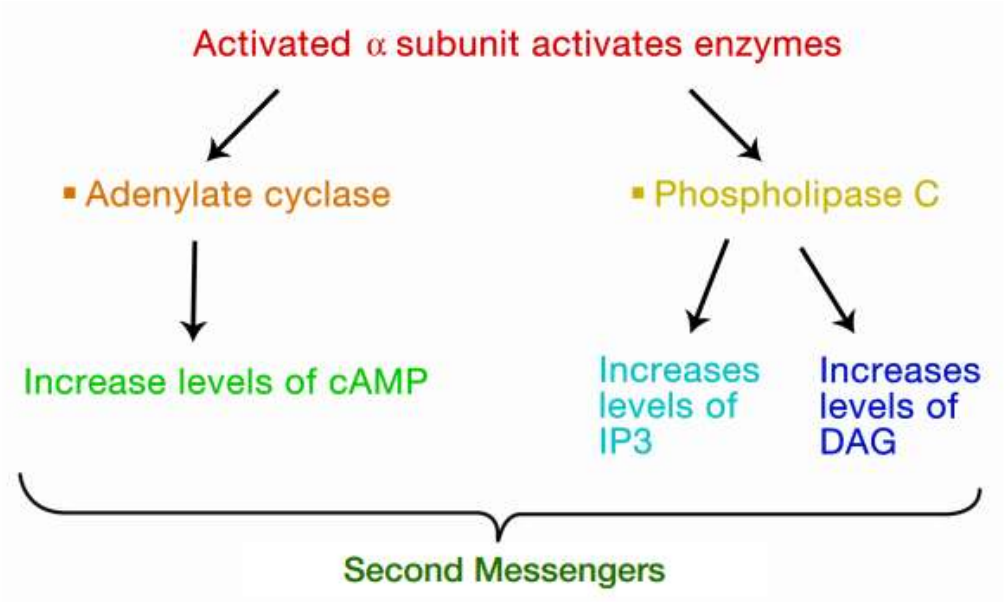


Fig.: Second messengers.

What is the effect of elevated cAMP levels?

cAMP molecules bind to, and activate an enzyme, protein kinase A (PKA). PKA is composed of two catalytic and two regulatory subunits that are bound tightly together. Upon binding of cAMP the catalytic subunits are released from the regulatory subunits, allowing the enzyme to carry out its function, namely phosphorylating other proteins.

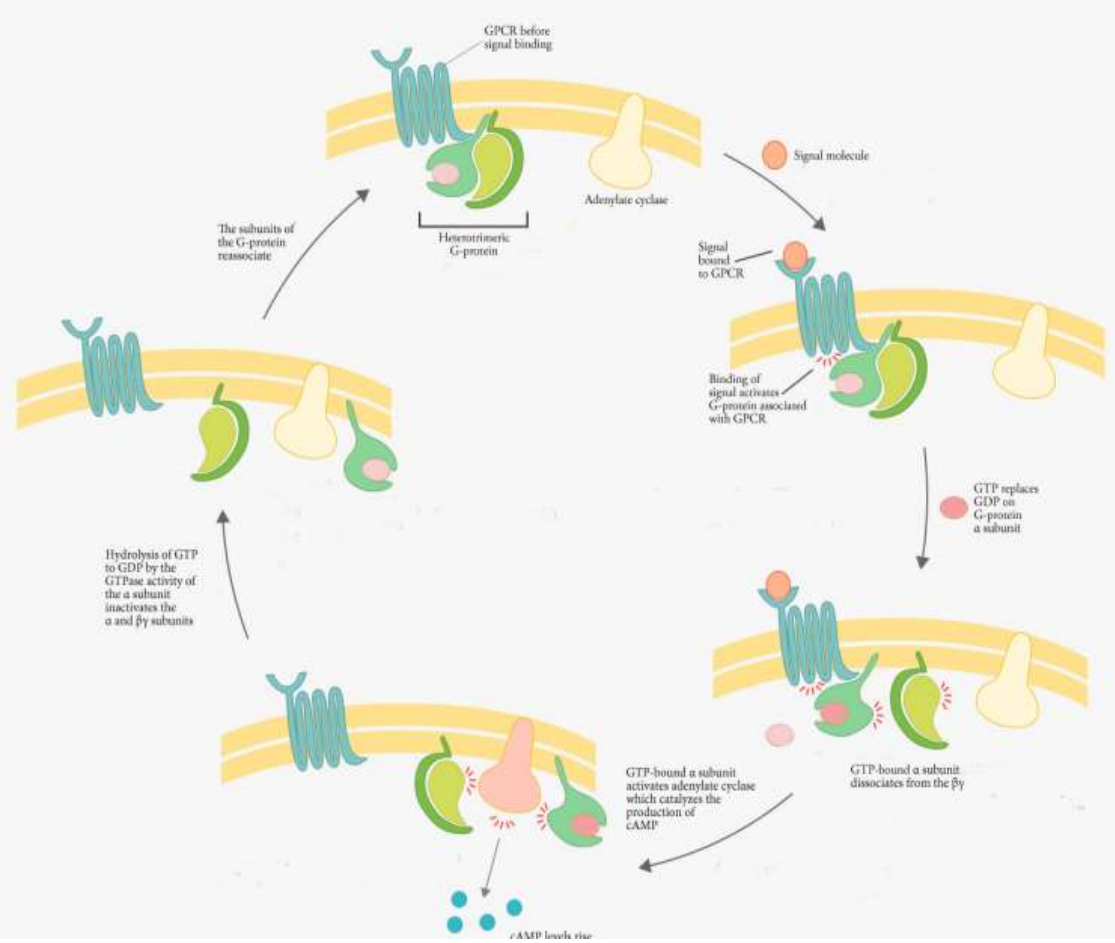


Fig.: *G*-protein signalling cycle.

Thus, cAMP can regulate the activity of PKA, which in turn, by phosphorylating other proteins can change their activity. The targets of PKA may be enzymes that are activated by phosphorylation, or they may be proteins that regulate transcription. The phosphorylation of a transcriptional activator, for example, may cause the activator to bind to a regulatory sequence on DNA and to increase the transcription of the gene it controls. The activation of previously inactive enzymes alters the state of the cell by changing the reactions that are occurring within the cell.

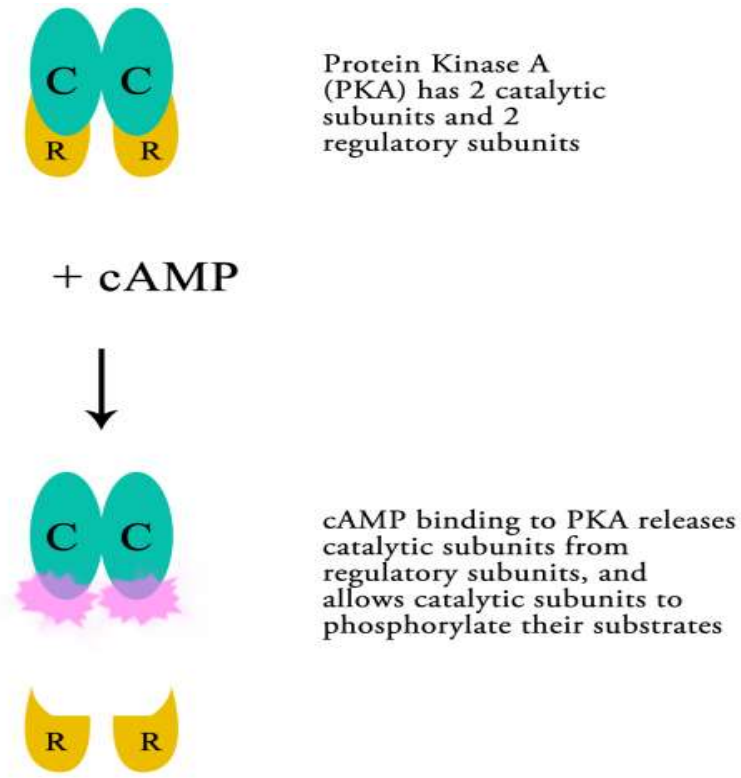


Fig.: **Protein Kinase A Activation.**

For example, the binding of epinephrine to its receptor on the cell surface, activates, through the action of G-proteins, and subsequent activation of PKA, the phosphorylation of glycogen phosphorylase. The resulting activation of glycogen phosphorylase leads to the breakdown of glycogen, releasing glucose (in the form of glucose-1-phosphate) for use by the cell. Changes in gene expression, likewise, lead to changes in the cell by altering the production of particular proteins in response to the signal.

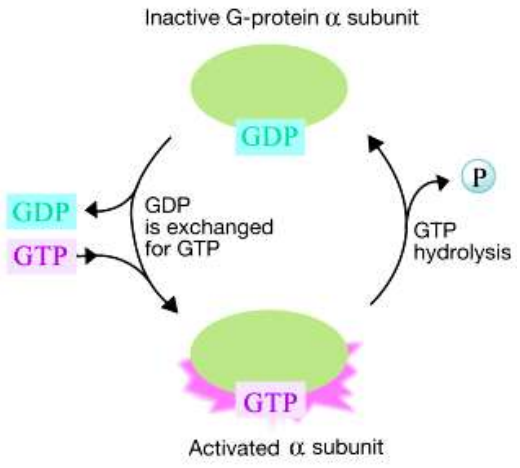


Fig.: **G-protein nucleotide swapping.**

- Binding of signal to receptor
- Several steps where the signal is passed on through intermediate molecules (G-proteins, adenylate cyclase, cAMP, and finally, PKA)
- Phosphorylation of target proteins by the kinase, leading to changes in the cell.

Finally, if the signal binding to the receptor serves as a switch that sets these events in motion, there must be mechanisms to turn the pathway off. The first is at the level of the G-protein. Recall that the alpha subunit of the G-protein is in its free and activated state when it has GTP bound and that it associates with the beta- gamma subunits and has a GDP bound when it is inactive. We also know that the alpha subunit has an activity that enables it to hydrolyze GTP to GDP, as shown in the figure above left. This GTP-hydrolyzing activity makes it possible for the alpha subunit, once it has completed its task, to return to its GDP bound state, re-associate with the beta-gamma part and become inactive again.

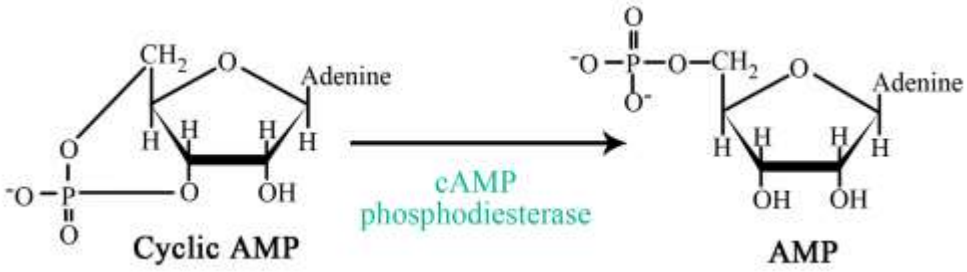


Fig.: **cAMP Breakdown.**

The second "off switch" is further down the signaling pathway, and controls the level of cAMP. We just noted that cAMP levels increase when adenylate cyclase is activated. When its job is done, cAMP is broken down by an enzyme called phosphodiesterase. When cAMP levels drop, PKA returns to its inactive state, putting a halt to the changes brought about by the activation of adenylate cyclase by an activated G-protein.

Let us now examine the events that follow the activation of Phospholipase C (PLC) by a G-protein. As we noted earlier, the activation of PLC results in the production of the second messengers IP3 and DAG. What do these molecules do?

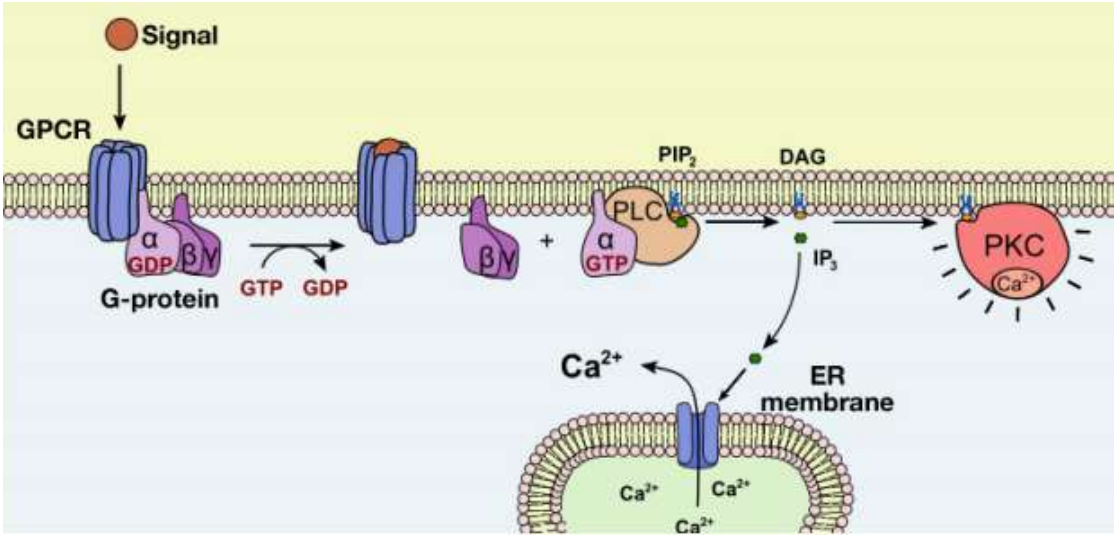


Fig.: **Phospholipase C Signaling.**

The IP3 and DAG produced by activated phospholipase C work together to activate a protein kinase. First, IP3 diffuses to the endoplasmic reticulum membrane where it binds to gated calcium ion channels. This causes calcium channels in the ER membrane to open and release large amounts of calcium into the cytoplasm from the ER lumen, as shown in the figure below.

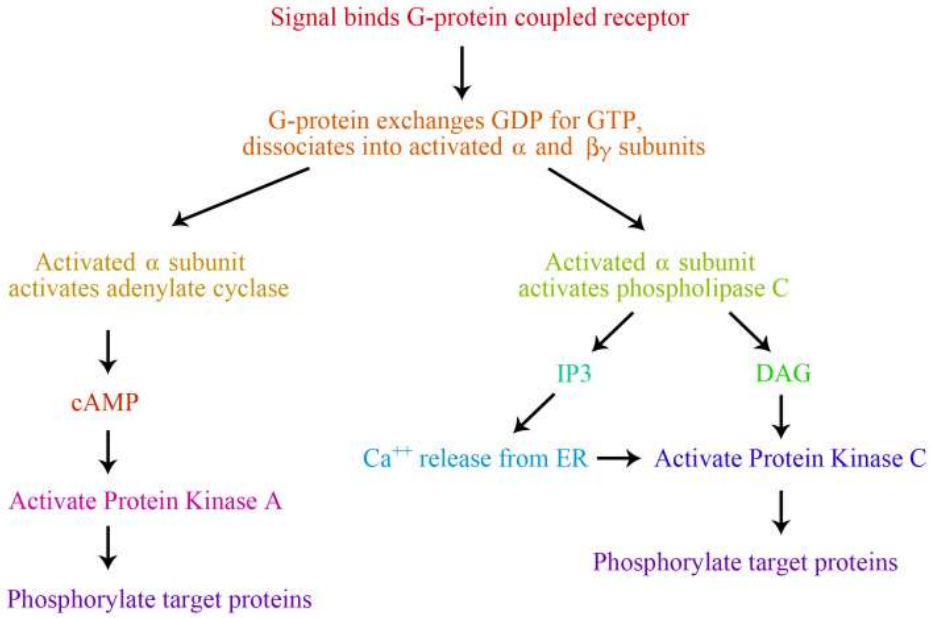


Fig.: **Signaling Outcomes.**

The increase in cytosolic calcium ion concentration has various effects, one of which is to activate a protein kinase called protein kinase C (C for calcium), together with the DAG made in the earlier step. Like PKA, Protein kinase C phosphorylates a variety of proteins in the cell, altering their activity and thus changing the state of the cell.

14. Let's sum up

- Intracellular receptors are located in the cytoplasm of the cell and are activated by hydrophobic ligand molecules that can pass through the plasma membrane.
- Cell-surface receptors bind to an external ligand molecule and convert an extracellular signal into an intracellular signal.
- Three general categories of cell-surface receptors include: ion -channel, G- protein, and enzyme -linked protein receptors.
- Ion channel -linked receptors bind a ligand and open a channel through the membrane that allows specific ions to pass through.
- G-protein-linked receptors bind a ligand and activate a membrane protein called a G-protein, which then interacts with either an ion channel or an enzyme in the membrane.
- Enzyme-linked receptors are cell-surface receptors with intracellular domains that are associated with an enzyme.
  - Cell cycle is the sequence of events that occur between two successive cell divisions. DNA replication and the division of a cell are the major events in cell cycle. Eukaryotic cell cycle is divided into four non overlapping phases – G1 (gap1), S (synthesis), G2 & M phase.

- Cyclin-dependent kinases (CDKs) are a group of enzymes that work to regulate different processes in the cell cycle after activation by the binding of a cyclin molecule.
- CDKs are a part of the CMGC group of enzymes consisting of serine or threonine units that are characterized by their dependency on protein subunits. The activity of these enzymes is only observed after the binding of a cyclin molecule followed by the phosphorylation of the threonine unit.
- The 53-kilodalton tumor suppressor protein p53 was discovered through its role in the induction of cancers by certain DNA viruses.
- Floral meristems can usually be distinguished from vegetative meristems by their larger size.
- During the vegetative phase of growth, the *Arabidopsis* apical meristem produces leaves with very short internodes, resulting in a basal rosette of leaves.
- In *Arabidopsis*, LEAFY (LFY), FLOWERING D (FD), SUPPRESSOR OF OVEREXPRESSION OF CONSTANS1 (SOC1), and APETALA1 (AP1) are among the critical genes in the genetic pathway that must be activated to establish floral meristem identity.
- All homeotic genes that have been identified so far, in both plants and animals, encode transcription factors. However, unlike animal homeotic genes, which contain homeobox sequences, most plant homeotic genes belong to a class of related sequences known as **MADS box genes**.
- Telomerase is a remarkable enzyme that includes multiple protein subunits and an RNA component (an example of a ribonucleoprotein). The key to telomerase's unusual functions is revealed by the RNA component of the enzyme, called **"telomerase RNA" (TER)**. Depending on the organism, TER varies in size from 150 to 1300 bases. In all organisms, the sequence of the RNA includes a short region that encodes about 1.5 copies of the complement of the telomere sequence (for humans, this sequence is 5'-AAUCCCAAUC-3').
- Epigenetics refers to heritable changes in gene expression that arise from changes in chromosomes without alteration of DNA sequence. These changes occur throughout all stages of development or in response to environmental

factors such as exposure to toxins or chronic stress and are implicated in diseases such as cancer.

- CpG islands (CGIs) make up only 0.7% of the human genome but contain 7% of the CpG dinucleotides. CpG islands often are highly enriched at gene promoters, and approximately 60% of all mammalian gene promoters are CpG-rich. CpG islands are typically unmethylated, open regions of DNA with low nucleosome occupancy.
- DNA methylation is catalyzed by DNA methyltransferases (DNMTs). DNMT3A and DNMT3B are involved in de novo methylation and are targeted to particular genomic regions by specific histone modifications. During DNA replication, the protein Np95 recognizes hemimethylated DNA and directs DNMT1 to the replication fork to maintain patterns of DNA methylation.
- Acetylation of a lysine residue neutralizes a positive charge on a histone protein, reducing the electrostatic interaction with negatively charged DNA. This reduction in affinity leads to increased accessibility of the DNA to protein complexes, which can lead to increased gene expression.
- RNA interference (RNAi) is a mechanism where the presence of certain fragments of dsRNA interferes with the expression of a particular gene which shares a homologous sequence with this dsRNA.
- RNAi enables repression of gibberellic acid and auxin signal pathways after a reduction in the level of *SlARF7 transcript responsible* for pollination and fertilization in tomato plants. These results by-pass the auxin signaling-fertilization pathway that leads to the development of parthenocarpic fruits having great commercial value.
- The basic mechanism involved in genetic manipulations through programmable **nucleases** is the recognition of target genomic loci and binding of effector DNA-binding domain (DBD), double-strand breaks (DSBs) in target DNA by the restriction endonucleases (FokI and Cas), and the repair of DSBs through **homology-directed recombination** (HDR) **or non-homologous end joining** (NHEJ).

- **CRISPR** (an acronym for **clustered regularly interspaced short palindromic repeats**) is a family of DNA sequences found in the genomes of prokaryotic organisms such as bacteria and archaea. These sequences are derived from DNA fragments of bacteriophages that had previously infected the prokaryote. They are used to detect and destroy DNA from similar bacteriophages during subsequent infections.
- Transcription activator-like effector nucleases (TALEN) are restriction enzymes that can be engineered to cut specific sequences of DNA.
- TAL effectors are proteins that are secreted by *Xanthomonas* bacteria via their type III secretion system when they infect plants.
- The proteome is the entire set of proteins that are produced or modified by an organism or system. This varies with time and distinct requirements, or stresses, that a cell or organism undergoes. Proteomics is an interdisciplinary domain that has benefitted greatly from the genetic information of various genome projects, including the Human Genome Project. It covers the exploration of proteomes from the overall level of protein composition, structure, and activity. It is an important component of functional genomics.
- 2-D electrophoresis begins with electrophoresis in the first dimension and then separates the molecules perpendicularly from the first to create an electropherogram in the second dimension. In electrophoresis in the first dimension, molecules are separated linearly according to their isoelectric point. In the second dimension, the molecules are then separated at 90 degrees from the first electropherogram according to molecular mass. Since it is unlikely that two molecules will be similar in two distinct properties, molecules are more effectively separated in 2-D electrophoresis than in 1-D electrophoresis.
- A common technique is to use an Immobilized pH gradient (IPG) in the first dimension. This technique is referred to as IPG-DALT. The sample is first separated onto IPG gel (which is commercially available) then the gel is cut into slices for each sample which is then equilibrated in SDS-mercaptoethanol and applied to an SDS-PAGE gel for resolution in the second dimension.

- MALDI is the abbreviation for "Matrix Assisted Laser Desorption/Ionization." The sample for MALDI is uniformly mixed in a large quantity of matrix. The matrix absorbs the ultraviolet light (nitrogen laser light, wavelength 337 nm) and converts it to heat energy. A small part of the matrix heats rapidly (in nano seconds) and is vaporized, together with the sample.
- Peptide mass fingerprinting (PMF), also known as protein fingerprinting, is a high-throughput analytical method that developed in 1933 to identify proteins. Endoproteases first cleaves the unknown target protein into smaller peptides. The absolute mass of these peptides can be accurately measured using a mass spectrometer (such as MALDI-TOF), and a list of peptide peaks of the unknown protein is also obtained. This list of peaks is compared to a list of theoretical peptide peaks from a protein database using a computer program.

- The cytoplasmic membrane is a hydrophobic barrier that limits the entry and exit of molecules from the cell. Nonetheless, living cells require the exchange of many types of molecules between the cytoplasm and the outside environment. Cells need nutrients and inorganic ions, and they must remove waste products and expel other ions. This chapter focuses on membrane impermeability and on membrane proteins that facilitate and control the transport of molecules across the lipid bilayer with a special emphasis on ions (e.g., K+ and Na+).
- A well-studied example of a facilitated diffusion carrier is the glucose transporter, or GLUT. GLUTs occur in nearly all cells and are particularly abundant in cells lining the small intestine. GLUTs are integral membrane proteins whose membrane-spanning region is composed of 12 a-helices.
- The term ionophore means “ion bearer.” Ionophores are small, lipid-soluble molecules, usually of microbial origin, whose function is to conduct ions across membranes. They are facilitated diffusion carriers that transport ions down their electrochemical gradient.
- **Protein targeting or protein sorting** is the biological mechanism by which proteins are transported to their appropriate destinations in the cell or outside it. Proteins can be targeted to the inner space of an organelle, different

intracellular membranes, plasma membrane, or to exterior of the cell via secretion.

- **TOM complex** is present in the outer mitochondrial membrane (OMM). This complex includes Tom40 (protein lined pore) and receptor proteins (Tom20, Tom22 and Tom5) which recognize and bind presequences on mitochondrial proteins.

- **TIM complex** is an inner mitochondrial membrane (IMM) protein complex that contains two major protein channels, Tim23 and Tim22 complex. Tim23 protein channel has a Tim44 receptor protein which is used to translocate the mitochondrial protein into the matrix while Tim22 protein complex is used to import mitochondrial membrane protein to the IMM.
- **Vesicular Tubular Clusters (VTC)**: VTC are formed when ER-derived vesicles fuse with one another to form large compartment These compartments differ from ER and lacks many functional proteins present in ER They are generated continually and function as transport packages that bring material from the ER to the Golgi apparatus. The clusters are relatively short-lived because they quickly move along microtubules to the Golgi apparatus, where they fuse and deliver their contents.
- Cell-surface receptors, also known as transmembrane receptors, are cell surface, membrane-anchored, or integral proteins that bind to external ligand molecules. This type of receptor spans the plasma membrane and performs signal transduction, converting an extracellular signal into an intracellular signal. Ligands that interact with cell-surface receptors do not have to enter the cell that they affect. Cell-surface receptors are also called cell-specific proteins or markers because they are specific to individual cell types.
- Internal receptors, also known as intracellular or cytoplasmic receptors, are found in the cytoplasm of the cell and respond to hydrophobic ligand molecules that are able to travel across the plasma membrane. Once inside the cell, many of

these molecules bind to proteins that act as regulators of mRNA synthesis to mediate gene expression.

- G-protein-linked receptors bind a ligand and activate a membrane protein called a G-protein. The activated G-protein then interacts with either an ion channel or an enzyme in the membrane. All G-protein-linked receptors have seven transmembrane domains, but each receptor has its own specific extracellular domain and G-protein-binding site.
- G-protein coupled receptors are cell surface receptors that pass on the signals that they receive with the help of guanine nucleotide binding proteins (G-proteins). Before thinking any further about the signaling pathways downstream of GPCRs, it is necessary to know a few important facts about these receptors and the G-proteins that assist them.
- Though there are hundreds of different G-protein coupled receptors, they all have the same basic structure: they all consist of a single polypeptide chain that threads back and forth seven times through the lipid bilayer of the plasma membrane. For this reason, they are sometimes called seven- pass transmembrane (7TM) receptors.

15. Suggested Readings

1. Snustad, D.P. & Simmons, M.J. Principles of Genetics (6th ed.), 2012, John Wiley & Sons.

2. Watson, J.D., Baker, T.A., Bell, S.P., Gann, A., Levine, M. & Losick, R. Molecular Biology of the Gene (5th ed.) 2004. Pearson Education Inc .

3. Taiz L., Zeiger, E, Moller, I. M. & Murphy, M. (2015) Plant Physiology and Development, 6th edition, Sinauer Associates, USA

4. Cottrell J S. Protein identification by peptide mass fingerprinting. *Peptide research*, 1994, 7(3): 115-124.

5. Hjernø K. Protein identification by peptide mass fingerprinting//Mass Spectrometry Data Analysis in Proteomics. *Humana Press*, 2007: 61-75.

6. Suckau D, Resemann A, Schuerenberg M, *et al*. A novel MALDI LIFT-TOF/TOF mass spectrometer for proteomics. *Analytical and bioanalytical chemistry*, 2003, 376(7): 952-965.

7. Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K. & Walter, P. Molecular Biology of the Cell, 2002 (4th ed.), Garland Sciences.

8. Molecular Cell Biology- Lodish H, Berk A, Kaiser C.A., Krieger M., Scott M.P., Bretscher A., Ploegh H. &Matsudaira P. (W.H. Freeman & Co.)

9. Cooper, G.M. and Hausman, R.E. (2009) The Cell: A Molecular Approach. 5th edition. ASM Press & Sunderland, Washington, D.C.; Sinauer Associates, MA.

10. Karp, G. (2010). Cell Biology, John Wiley & Sons, U.S.A. 6th edition.

11. https://www.slideshare.net/purakichha/structure-and-functon-of-golgi-apparatus

12. https://egyankosh.ac.in/bitstream/123456789/68491/1/Block-3.pdf

13. https://en.wikipedia.org/wiki/CRISPR_gene_editing

14. https://www.slideshare.net/MUHAMMADAQUIB3/final-ppt-90944683

15. https://www.isaaa.org/resources/publications/pocketk/foldable/Pocket%20K59%20(English).pdf

16. https://www.slideshare.net/MuhammadMujahid58/talens-transcription-activatorlike-effector-nucleases-220134170

17. Alaa A.A. Aljabali, Mohamed El-Tanani, Murtaza M. Tambuwala (2024) Principles of CRISPR-Cas9 technology: Advancements in genome editing and emerging trends in drug delivery, Journal of Drug Delivery Science and Technology, Volume 92, 105338, ISSN 1773-2247, https://doi.org/10.1016/j.jddst.2024.105338

18. Jiang F, Doudna JA (2017) CRISPR-Cas9 structures and mechanisms. Annu Rev Biophys 46:505–529

19. Pierce, Benjamin A. Genetics (2nd ed.), 2005, W.H. Freeman & Company.

20. Griffiths, A.I.F., Miller, J.H., Suzuki, D.T., Lewentin, C.R. & Gilbert, M.W. An Intrduction to Genetic Analysis, 2005 (8th ed.), W.H. Freeman & Company.

21. Bruce Alberts, Alexander Johnson, Julian Lewis, David Morgan, Martin Raff, Keith Roberts, Peter Walter. Molecular Biology of the Cell. 2015. 6th Ed: Garland Science.

22. Gerald Karp. Cell Biology. 2013. 7th Ed. International Student Version. Wiley.

23. Lewin, B. Genes VIII, 2004, Pearson Educational International.

24. Kar, D.K. and Halder, S. Cell Biology, Genetics and Molecular Biology 2008, New Central Book Agency.

25. The Multinational Coordinated Arabidopsis *thaliana* Functional Genome Project: Annual Report 2010, Multinational *Arabidopsis* Steering Committee, 2010·

26. http://www.biologydiscussion.com/

27. https://microbenotes.com/genomics/

28. http://www.deskuenvis.nic.in/pdf/yeast1.pdf

29. https://microbiologynotes.org/whole-genome-shotgun-sequencing-overview-steps-and-achivements/#google_vignette.

30. https://www.slideshare.net/shitalpal3/genome-sequencing-41108752

31. https://www.collegesidekick.com/study-docs/5690092

32. https://bio.libretexts.org/Bookshelves/Introductory_and_General_Biology/Book%3A_General_Biology_(Boundless)/18%3A_Evolution_and_the_Origin_of_Species/18.04%3A_Evolution_of_Genomes/18.4C%3A_Whole-Genome_Duplication

33. https://www.slideshare.net/kirtimehta16/molecular-tagging

34. https://www.slideshare.net/kcyaadav/gene-traps-for-plant-development-and-genomics

35. http://www.nsf.gov/news/news_sum.jsp?cntnid=10307

36. https://nowgonggirlscollege.co.in/attendence/classnotes/files/1628944240.pdf

37. https://www.slideshare.net/berciyalgolda1/metagenomics-251652133

16. Assignments

1. Write short notes on any four of the following:
   i. Proteomics. ii. Application of RNAi in crop quality improvement iii. RNAi iv. p53.
2. Describe the function of small RNAs in post-transcriptional gene silencing.
3. Write short notes on any four of the following:
   i. Functional genomics ii. G-protein
4. Describe the basic principle of 2-DE with limitation.
5. What are G-proteins?
6. Write short note on cell signalling pathway via G proteins.
7. What are ionophores?
8. Write the significance of si-RNA and mi-RNA?
9. What is epiallels?
10. What is epigenetics?
11. Define functional and structural proteomics.
12. Write a short note on genomics.
13. Differentiate between genomics and proteomics
14. What is meant by expression proteomics?
15. Enumerate the steps of recombinant DNA construction.

All the materials are self writing and collected from e-book, journals and websites.