Post-Graduate Degree Programme (CBCS)

in

ZOOLOGY

SEMESTER-IV

ELECTIVE THEORY PAPER

CYTOGEETICS AND MOLECULAR BIOLOGY

ZDSE(MJ)T-403

SELF LEARNING MATERIAL

DIRECTORATE OF OPEN AND DISTANCE LEARNING

UNIVERSITY OF KALYANI

KALYANI, NADIA

W.B. INDIA

Content Writer:

Dr. Subhabrata Ghosh, Assistant Professor, Department of Zoology (UG & PG), Asutosh College, Kolkata

---

May 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.

Director's Message

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 Distance Learning (ODL) systems. The onus lies on writers, editors, production professionals and other personnel involved in the process to overcome the challenges inherent to curriculum 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 Distance 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 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 the SLMs, making them useful to the learners, 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 criticism 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 everymember of PGBOS (DODL), University of Kalyani. Heartfelt thanks are 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 coordinated efforts have resulted in the compilation of comprehensive, learner-friendly, flexible texts that meet the 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 and Distance Learning
University of Kalyani

ELECTIVE THEORY PAPER (ZDSE(MJ)T -403)

CYTOGENETICS AND MOLECULAR BIOLOGY

PART I -Epigenetics

Module	Unit	Content	Credit	Class	Time (h)	Page No.
ZDSE(MJ)T - 403 (CYTOGENETICS AND MOLECULAR BIOLOGY)	I	**A brief history of epigenetics - overview and concepts; chromatin modifications and their mechanism of action, concept of 'histone-code' hypothesis.**				
	II	**Epigenetics in saccharomyces cerevisiae, position effect variegation, heterochromatin formation, and gene silencing in Drosophila**				
	III	**Fungal models for epigenetic research: *Schizosaccharomyces pombe* and *Neurospora crassa*; RNAi and heterochromatin assembly, role of noncoding RNAs;**				
	IV	**Chromatin structure and epigenetics marks - transcriptional silencing by polycomb group proteins , transcriptional regulation by trithorax group proteins, histone variants and epigenetics, epigenetic**				

		regulation of chromosome inheritance.				
	V	**Epigenetic regulation of the X chromosomes in *C.elegans*, dosage compensation in *Drosophila*, dosage compensation in mammals.**				
	VI	**Types and mechanism of chromatin remodeling.**				
	VII	**Epigenetics and genome imprinting - DNA methylation in mammals, genomic imprinting in mammals.**				
	VIII	**Nuclear transplantation and the reprogramming of the genome. epigenetics and human disease, epigenetic determinants of cancer.**				
PART- II **Transposons and Extra-nuclear inheritance**						
	IX	**Mobile genetic elements: Characteristics of transposable elements in prokaryotes and eukaryotes; AC/DS system in maize**				
	X	**P element in Drosophila; Salmonella phase variation; retrosposons**				

	XI	**Extra-nuclear inheritance: Streptomycin resistance in *Chlamydomonus***				
	XII	**Kappa particles; criteria for extra-chromosomal inheritance, infectious heredity.**				
	XIII	**Recombination: Homologous recombination, Mechanism of recombination in bacteria and mammals, Gene conversion, Classes of recombinases and types of site-specific recombination,**				
	XIV	**Transpositional recombination, Mitotic and meiotic recombination, Recombination and genomic instability, Application in genetic engineering.**				

UNIT-I

A brief history of epigenetics - overview and concepts; chromatin modifications and their mechanism of action, concept of 'histone-code' hypothesis

Objective: In this unit we will discuss brief history of epigenetics. We will give emphasis to chromatin modifications and their mechanism of action.

Epigenetics: Epigenetics is the study of heritable changes in gene expression (active versus inactive genes) that do not involve changes to the underlying DNA sequence — a change in phenotype without a change in genotype — which in turn affects how cells read the genes. Epigenetic change is a regular and natural occurrence but can also be influenced by several factors including age, the environment/lifestyle, and disease state. Epigenetic modifications can manifest as commonly as the manner in which cells terminally differentiate to end up as skin cells, liver cells, brain cells, etc. Or, epigenetic change can have more damaging effects that can result in diseases like cancer. At least three systems including DNA methylation, histone modification and non-coding RNA (ncRNA)-associated gene silencing are currently considered to initiate and sustain epigenetic change. New and ongoing research is continuously uncovering the role of epigenetics in a variety of human disorders and fatal diseases.

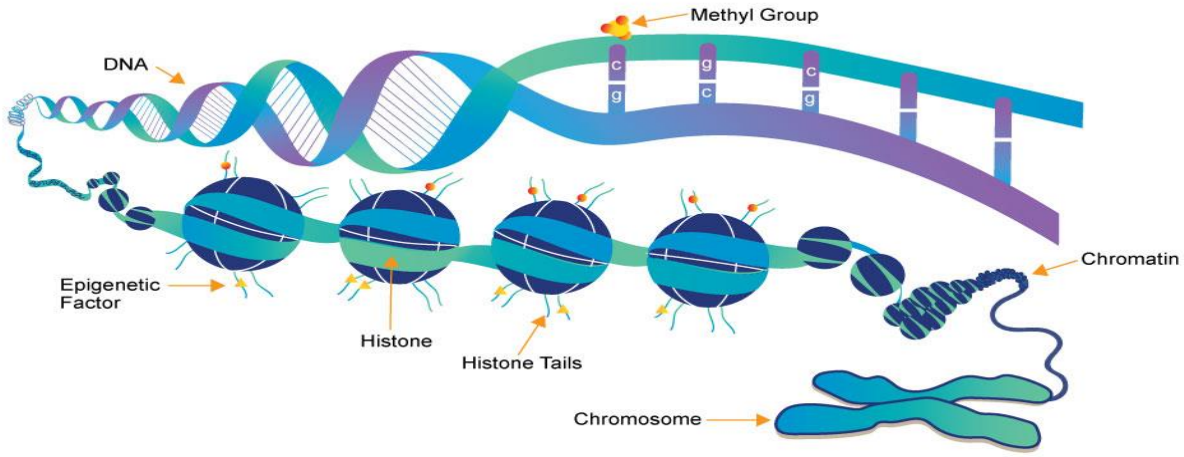


Fig: Representation of the chromatin structure, including histones and DNA, which become available to epigenetic marks.

The Evolving Landscape of Epigenetic Research: A Brief History

What began as broad research focused on combining genetics and developmental biology by well-respected scientists including **Conrad H. Waddington** and Ernst Hadorn during the mid-twentieth century has evolved into the field we currently refer to as epigenetics. The term epigenetics, which was coined by Waddington in 1942, was derived from the Greek word "epigenesis" which originally described the influence of genetic processes on development. During the 1990s there became a renewed interest in genetic assimilation. This led to elucidation of the molecular basis of Conrad Waddington's observations in which environmental stress caused genetic assimilation of certain phenotypic characteristics in *Drosophila* fruit flies. Since then, research efforts have been focused on unraveling the epigenetic mechanisms related to these types of changes.

Currently, DNA methylation is one of the most broadly studied and well-characterized epigenetic modifications dating back to studies done by Griffith and Mahler in 1969 which suggested that DNA methylation may be important in long term memory function.4 Other major modifications include chromatin remodeling, histone modifications, and non-coding RNA mechanisms. The renewed interest in epigenetics has led to new findings about the relationship between epigenetic changes and a host of disorders including various cancers, mental retardation associated disorders, immune disorders, neuropsychiatric disorders and paediatric disorders.

The human genome is composed of billions of sequence arrangements containing a bioinformatics code that controls how genes are expressed. This code is further dependent upon heritable non-static epigenetic arrangement of histone scaffolding that surrounds the DNA and comprises the "epigenome." The historical transitional evolution of the human genome is believed to have occurred through a number of processes, one being the altered sequence and re-arrangement of transposable elements located at segments of non-coding DNA. It is believed that the greater the complexity of an organism, the greater amount of non-coding DNA. In humans, protein-coding regions of DNA account for <1.6% of the genome. Transposable elements, also referred to as "jumping genes," have accumulated throughout millions of years as evolutionary ancient DNA in the form of transposons and retrotransposons, which are reverse transcribed long-terminal repeat (LTR) retroviruses. Today, active non-LTR retrotransposons (i.e., Alu and LINEs) perpetuate transgenerational genetic diversities through genomic DNA variation among humans. While the evolution of DNA occurs at a slow pace, expedient heritable changes to the epigenome allow dynamic and flexible modification to suit rapid environmental adaptation. While the epigenome has more influence on the temporal phenotype, the collective effects of change to the genome and the epigenome contribute to observable physical or biochemical characteristics of an organism.

Throughout the life cycle, dynamic epigenetic control over the phenotype is influenced by a time component responsible for maturation and senescence from conception to adulthood. Environmental epigenetic factors affecting long-term phenotypic change are largely initiated during in utero/perinatal periods, when introduction to the external world is being established. It is believed that since epigenetic patterns are inherited through mitosis, the earlier the stage of development, the more critical the environmental impact on the resulting phenotype. During fetal development, environmental cues can induce the modification of a pliable epigenome, which can result in long-term changes in gene expression that occur in a self-sustaining manner in the absence of the original stimulus. Adverse gestational conditions that arise from inadequate healthcare, poor nutrition, socioeconomic disadvantage and racial disparities are often associated with long-lasting phenotypic consequences in adults, yielding greater risk of diabetes and heart disease, as well as low birth weight and congenital defects in progeny. It is now becoming evident that these effects are inextricably linked to altered epigenetic patterns. Offspring exposed to gestational malnutrition due to extended famine in certain populations also show higher prevalence of adult onset obesity and schizophrenia, tantamount to altered DNA methylation patterns for specific genes such as *insulin like growth factor 1/2* and the obesity factor gene *leptin*. Altered epigenetic patterns acquired during early development involve changes in DNA methylation patterns, genomic imprinting, histone modifications and the establishment of specific expression profiles of non-coding miRNAs.

Given the enormous impact of early epigenetic programming and the serious nature of related developmental conditions, such as Prader-Willi, Angelman and Beckwith-Wiedemann syndromes, neural tube defects, adult onset psychiatric disorders, obesity, cancer and schizophrenia, considerable attention is given to the "nurture of the epigenome" prior to birth. A number of community outreach projects such as the CDC's National Centre on Birth Defects and Developmental Disabilities promote awareness about reducing the risk of epigenetic related defects such as spina bifida by suggesting an adequate intake of epigenetic-related nutrients (e.g., choline, vitamin B_{12}, B_6 and folate) during pregnancy. The epigenome appears to remain pliable after birth and during the first years of life. This is evidenced by correlations described in infants exposed to stress or lack of emotional nurturing who show overactive hypothalamic-pituitary-adrenal stress response, glucocorticoid feedback or decreased hypothalamic corticotropin-releasing factor. Once established in the offspring, epigenetic marks can become transgenerational, continuing transmittance to future descendants—including the very trait of maternal nurturing in females. The longevity of transgenerational epigenomic inheritance pattern is further influenced by the severity and repetition of a similar environmental stimulus among individuals of the same lineage. If the stimuli are discontinued, phenotypic traits could dissipate after the first or second generation. In other instances, longer lasting epigenetic changes adversely affect the phenotype of the third or fourth generation, often initiated by environmental factors adverse to human health that perpetuate aberrant patterns of transgenerational transmission of phenotype.

The purpose of this review is to simplify the enormous complexity of epigenetic biochemistry that links nuclear DNA to the environment. On one hand, the concept of epigenetics is relatively simple in that it describes a means by which genes are either turned on or off by a heritable epigenome. On the other hand, the environmental and biological controls that mediate these events are extraordinary in number, compounded by instances of similar methylation events that have opposite effects when occurring at different histone amino acids (e.g., H3K36me3 and H3K9me3) and by variation in the interpretation of studies performed in diverse organisms such as flies, plants, worms, yeast, ciliated protozoans, tumour cells and mammals.

Various mechanisms of epigenetic regulation: An overview

1. Genomic Imprinting
2. X – Chromosome inactivation
3. Bookmarking
4. Gene Silencing or Activation
5. Paramutation

6. Histone modification and heterochromatin regulation
7. Position effect and variegation
8. Teratogenic effects etc.

Mechanisms of Histone Modification Function

There are two characterized mechanisms for the function of modifications. The first is the disruption of contacts between nucleosomes in order to ‘‘unravel’’ chromatin and the second is the recruitment of nonhistone proteins. The second function is the most characterized to date. Thus, depending on the composition of modifications on a given histone, a set of proteins are encouraged to bind or are occluded from chromatin. These proteins carry with them enzymatic activities (e.g., remodelling ATPases) that further modify chromatin. The need to recruit an ordered series of enzymatic activities comes from the fact that the processes regulated by modifications (transcription, replication, repair) have several steps. Each one of these steps may require a distinct type of chromatin-remodelling activity and a different set of modifications to recruit them. Below is a more detailed description of the different mechanisms by which modifications work. Modifications may affect higher-order chromatin structure by affecting the contact between different histones in adjacent nucleosomes or the interaction of histones with DNA. Of all the known modifications, acetylation has the most potential to unfold chromatin since it neutralizes the basic charge of the lysine. This function is not easy to observe in vivo, but biophysical analysis indicates that intern-nucleosomal contacts are important for stabilization of higher-order chromatin structure. Thus, any alteration in histone charge will undoubtedly have structural consequences for the chromatin architecture. Furthermore, the recent development of strategies to make recombinant nucleosomes modified at specific sites has allowed this question to be addressed in vitro. By chemically ligating modified tail peptides onto recombinant histone core preparations, it has been possible to show that acetylation of H4K16 has a negative effect on the formation of a 30-nanometer fibre and the generation of higher-order structures. Phosphorylation is another modification that may well have important consequences for chromatin compaction via charge changes. The role of this modification has not been demonstrated rigorously in vitro but demonstrations of its role in mitosis, apoptosis, and gametogenesis are suggestive of such a role. Proteins are recruited to modifications and bind via specific domains. Methylation is recognized by chromo-like domains of the Royal family (chromo,tudor, MBT) and nonrelated PHD domains, acetylation is recognized by bromodomains, and phosphorylation is recognized by a domain within 14-3-3 proteins. A number of proteins have been identified that are recruited to specific modifications. The recent isolation of several proteins that recognize H3K4me has highlighted the fact that their purpose is to tether enzymatic activities onto chromatin. BPTF, a component of the

NURF chromatin-remodelling complex, recognizes H3K4me3 via a PHD domain. This recruitment tethers the SNF2L ATPase to activate H0XC8 gene expression. The PHD-finger proteinING2 tethers the repressive mSin3a-HDAC1 histone deacetylases complex to highly active, proliferation-specific genes after the exposure of cells to DNA-damaging agents. This finding represents a new mechanism of active shut-off of highly transcribed,H3K4-methylated genes. Two other H3K4mebindingproteins JMJD2A and CHD1 also tether enzymatic activities to chromatin, but in these instances the enzymatic activity resides within the methyl-binding protein:JMJD2A is a histone lysine demethylase that binds via a tudor domain and CHD1 is an ATPase that binds via a chromodomain. One other protein, WDR5, has been demonstrated to bind H3K4me1 and H3K4me2. however, structural analysis of this interaction does not support a purely methyl-recognition based interaction but suggests that this protein binds most avidly to the residues preceding H3K4 and in particular to H3R2. Perhaps this protein provides an adaptor function, augmenting the recognition of H3K4me.Proteins that bind other modified residues also deliver enzymes: H3K27me recruits the chromodomain containing polycomb protein PC2, which is associated with ubiquitin ligase activity specific for H2A; the chromocontainingHP1 protein binds H3K9me and is associated with deacetylase activity and methyltransferase activity. Equally important may be the effectiveness of histone modifications in preventing the docking of nonhistone proteins onto chromatin. The study of such pathways is less detailed, but examples include H3K4me disrupting the binding of the NuRD complex and H3T3ph preventing the binding of the INHAT complex. Both complexes have a repressive capability for transcription, so their occlusion by positively acting modifications makes sense. The abundance of modifications on the histone tail make ‘‘crosstalk’’ between modification is very likely. Mechanistically such communication between modifications may occur at several different levels. Firstly, many different types of modification occur on lysine residues. This will undoubtedly result in some form of antagonism since distinct types of modifications on lysines are mutually exclusive. Secondly, the binding of a protein could be disrupted by an adjacent modification.

The best example of this is phosphorylation ofH3S10 that affects the binding of HP1 to methylated H3K9. Thirdly, the catalytic activity of an enzyme could be compromised by modification of its substrate recognition site; for example, isomerization ofH3P38 affects methylation of H3K36 by Set2. Fourthly, an enzyme could recognize its substrate more effectively in the context of a second modification; the example here is the GCN5 acetyltransferase, which may recognize H3 more effectively when it is phosphorylated at H3S10. Communication between modifications can also occur when the modifications are on different histone tails. The best studied example is the case of ubiquitinoylation of H2B being required for methylation of H3K4me3.

Histone code hypothesis:

The histone code is a hypothesis that the transcription of genetic information encoded in DNA is in part regulated by chemical modifications to histone proteins, primarily on their unstructured ends that is in histone tails, particularly H3 and H4 by Acetylation, methylation, phosphorylation, ubiquitination etc. Together with similar modifications such as DNA methylation it is part of the epigenetic code. Histones associate with DNA to form nucleosomes, which themselves bundle to form chromatin fibres, which in turn make up the more familiar chromosome. Histones are globular proteins with a flexible N-terminus that protrudes from the nucleosome. Many of the histone tail modifications correlate very well to chromatin structure and both histone modification state and chromatin structure correlate well to gene expression levels. The critical concept of the histone code hypothesis is that the histone modifications serve to recruit other proteins by specific recognition of the modified histone via protein domains specialized for such purposes, rather than through simply stabilizing or destabilizing the interaction between histone and the underlying DNA. These recruited proteins then act to alter chromatin structure actively or to promote transcription.

Probable Questions:

1. Define epigenetics. How it evolve?
2. Explain histone code hypothesis.
3. Explain Position-Effect Variegation.

Suggested Readings:

1. Molecular Cell Biology by Lodish, Fourth Edition.
2. The Cell – A Molecular Approach by Cooper and Hausman, Fourth Edition
3. Principles of Genetics by Snustad and Simmons, Sixth Edition.
4. Molecular Biology of the Cell – by Bruce Alberts
5. Cell and Molecular Biology by Gerald Karp, 7th Edition.
6. Gene cloning and DNA Analysis by T. A. Brown, Sixth Edition.
7. Genetics . Verma and Agarwal.
8. Brown TA. (2010) Gene Cloning and DNA Analysis. 6th edition. Blackwell Publishing, Oxford, U.K.

UNIT-II

Epigenetics in *Saccharomyces cerevisiae*, position effect variegation, heterochromatin formation and gene silencing in Drosophila

Objectives: Epigenetics in *Saccharomyces cerevisiae*, gene silencing in *Drosophila* will also be discussed in this unit.

Epigenetics in *Saccharomyces cerevisiae*

Saccharomyces cerevisiae provides a well-studied model system for heritable silent chromatin, in which a nonhistone protein complex—the SIR complex—represses genes by spreading in a sequence-independent manner, much like heterochromatin in higher eukaryotes. The ability to study mutations in histones and to screen genome-wide for mutations that impair silencing has yielded an unparalleled depth of detail about this system. Recent advances in the biochemistry and structural biology of the SIR-chromatin complex bring us much closer to a molecular understanding of how Sir3 selectively recognizes the deacetylated histone H4 tail and demethylated histone H3 core. The existence of appropriate mutants has also shown how components of the silencing machinery affect physiological processes beyond transcriptional repression.

The fraction of chromatin in a eukaryotic nucleus that bears active genes is termed euchromatin. This chromatin condenses in mitosis to allow chromosomal segregation and decondenses in interphase of the cell cycle to allow transcription to occur. However, some chromosomal domains were observed by cytological criteria to remain condensed in interphase, and this constitutively compacted chromatin was called heterochromatin. With the development of new techniques, molecular rather than cytological features have been used to define this portion of the genome, and heterochromatin, which is often found at centromeres and telomeres, was shown to contain many thousands of simple repeat sequences, particularly in higher eukaryotic organisms. The repeat-rich genomic DNA tends to replicate late in S phase of the cell cycle, is found clustered at the nuclear periphery or near the nucleolus, and is resistant to nuclease attack. Importantly, the characteristic chromatin structure that is formed on repeat DNA tends to spread and repress nearby genes. In the case of the fruit fly locus *white*, a gene that determines red eye colour, epigenetic repression yields a red and white sectored eye through a phenomenon called position effect variegation (PEV). Mechanistically, PEV in flies reflects the

recognition of methylated histone H3K9 by heterochromatin protein 1 (HP1), which can spread along the chromosomal arm. In *Saccharomyces cerevisiae*, also known as budding yeast, a distinct mechanism of heterochromatin formation has evolved, yet it achieves a very similar result.

S. cerevisiae is a microorganism commonly used for making beer and baking bread. However, unlike bacteria, it is a eukaryote. The chromosomes of budding yeast, like those of more complex eukaryotes, are bound by histones, enclosed in a nucleus and replicated from multiple origins during S phase. Still, the yeast genome is tiny with only 14 megabase pairs of genomic DNA divided among its 16 chromosomes, some not much larger than a bacteriophage genome. There are approximately 6000 genes in the yeast genome, closely packed along chromosomal arms, generally with less than 2 kb spacing between them. The vast majority of yeast genes are in an open chromatin state, meaning that they are either actively transcribed or can be rapidly induced. This, coupled with a very limited amount of simple repeat DNA, makes the detection of heterochromatin by cytological techniques very difficult in yeast.

Nonetheless, budding yeast has distinct heterochromatin-like regions adjacent to all 32 telomeres and at two silent mating loci on chromosome III, shown using molecular tools. Transcriptional repression at telomeres and the silent mating loci can spread into adjacent DNA and repression of the silent mating loci is essential for maintaining a mating-competent haploid state. Both the subtelomeric regions and the silent mating type loci repress integrated reporter genes in a position-dependent, epigenetic manner; they replicate late in S phase and are present at the nuclear periphery. Thus, these loci bear most of the functional characteristics of heterochromatin, without having cytologically visible condensation in interphase. By exploiting the advantages afforded by the small genome of yeast and its powerful genetic and biochemical tools, many basic principles of chromatin-mediated repression that are relevant to heterochromatin in more complex organisms have been discovered. Nonetheless, silent chromatin in budding yeast is dependent on a unique set of nonhistone proteins that do not deposit nor recognize histone H3 lysine 9 methylation.

Position-Effect Variegation, Heterochromatin Formation, and Gene Silencing in *Drosophila*

Position-effect variegation (PEV) results when a gene normally in euchromatin is juxtaposed with heterochromatin by rearrangement or transposition. When heterochromatin packaging spreads across the heterochromatin/euchromatin border, it causes transcriptional silencing in a stochastic pattern. PEV is intensely studied in *Drosophila* using the *white* gene. Screens for dominant mutations that suppress or

enhance *white* variegation have identified many conserved epigenetic factors, including the histone H3 lysine methyltransferase SU(VAR). Heterochromatin protein HP1a binds H3K9me2/3 and interacts with SU(VAR), creating a core memory system. Genetic, molecular, and biochemical analysis of PEV in *Drosophila* has contributed many key findings concerning establishment and maintenance of heterochromatin with concomitant gene silencing.

Genes that are abnormally juxtaposed with heterochromatin, either by rearrangement or transposition, show a variegating phenotype. This is a result of the gene being silenced in some of the cells in which it is normally active. Because the change is caused by a change in the position of the gene in the genome, rather than a change in the gene itself this phenomenon is termed "position-effect variegation" (PEV). The silencing that occurs in PEV can be attributed to the packaging of the reporter gene in a heterochromatic form, indicating that endogenous heterochromatin formation, once initiated, can spread to encompass nearby genes. Genetic, cytological, and biochemical analyses are all possible in *Drosophila melanogaster*. In this article we will show how these different approaches have converged to identify many contributors to this system, leading to characterization of both structural proteins and modifying enzymes that play key roles in establishing and maintaining heterochromatin.

Heterochromatin formation depends critically on methylation of histone H3 at lysine 9 (H3K9me2/3), with concomitant association of heterochromatin protein 1 (HP1a) and other interacting proteins, including H3K9 methyltransferases (HKMTs); the multiple interactions of these proteins are required for the spreading and maintenance of heterochromatin. Targeting of heterochromatin formation to particular regions of the genome appears to involve multiple mechanisms, from satellite DNA-specific binding proteins to utilization of the RNA interference (RNAi) machinery. Although heterochromatic regions (pericentric regions, telomeres, the Y chromosome, and the small fourth chromosome) share a common biochemistry, each is distinct, and each is complex in different ways. Heterochromatin in *Drosophila* is gene poor, but it is not devoid of genes, and counterintuitively, those genes that reside in heterochromatin are often dependent on this environment for full expression. A complete understanding of heterochromatin formation and maintenance (including targeting and spreading) will need to include an explanation for the varying responses of different genes to this chromatin environment.

Probable Questions:

1. How epigenetics affect *Saccharomyces cerevisiae*?
2. How epigenetics causes gene silencing in *Drosophla* ?

Suggested Readings:

1. Molecular Cell Biology by Lodish, Fourth Edition.
2. The Cell – A Molecular Approach by Cooper and Hausman, Fourth Edition
3. Principles of Genetics by Snustad and Simmons, Sixth Edition.
4. Molecular Biology of the Cell – by Bruce Alberts
5. Cell and Molecular Biology by Gerald Karp, 7th Edition.
6. Gene cloning and DNA Analysis by T. A. Brown, Sixth Edition.
7. Genetics . Verma and Agarwal.
8. Brown TA. (2010) Gene Cloning and DNA Analysis. 6th edition. Blackwell Publishing, Oxford, U.K.

UNIT-III

Fungal models for epigenetic research: *Schizosaccharomyces pombe* and *Neurospora crassa*; RNAi and heterochromatin assembly, role of noncoding RNAs.

Objective: In this unit we will discuss epigenetics of fungal models such as *Schizosaccharomyces pombe* and *Neurospora crassa*. We will also discuss epigenetic regulation in plants. RNAi and heterochromatin assembly as well as role of non coding RNAs will also e discussed.

A. Epigenetic study in *Schizosaccharomyces pombe*:

Assembly of heterochromatin, a dense chromatin structure that represses the expression of embedded genes, is vital for the establishment and maintenance of cell identity. A hallmark of heterochromatin is methylation of histone H3 at Lys-9 (H3K9me), a modification that is conserved from the fission yeast *Schizosaccharomyces pombe* to humans. Studies using *S. pombe* as a model organism have established the concept that the RNA interference (RNAi) pathway contributes to the assembly of heterochromatin. In fission yeast, the RNAi pathway is required predominantly at the pericentromeric regions, while the pathway is dispensable for the maintenance of the heterochromatin assembled at the subtelomeric regions and the mating-type locus. Notably, defects in the RNAi pathway lead to great loss of H3K9me and derepression of silencing at the pericentromeric regions but not at the subtelomeric regions or the mating-type locus. These regional differences in dependence on the RNAi pathway have provided researchers with clues to ascertain whether the factors of interest act specifically in the RNAi pathway or act more generally in the assembly of heterochromatin.

In the *S. pombe* RNAi pathway, formation of the small interfering RNA (siRNA)-containing effector complex is coupled to heterochromatin assembly. siRNA is generated, by the Dicer family endoribonuclease Dcr1, from double-stranded non-coding RNA that is complementary to heterochromatin. The siRNA duplex is loaded onto a non-chromatin-associated complex called Argonaute small interfering RNA chaperone (ARC), which contains the Argonaute family endoribonuclease Ago1. The loading of the siRNA duplex onto the Ago1 subunit requires the two ARC-specific subunits Arb1 and Arb2, which also inhibit the release of the passenger strand. This complex then changes its subunit composition to form a chromatin-associated effector complex called RNA-induced transcriptional silencing (RITS). The RITS complex is composed of Ago1, now binding

single-stranded siRNA as a guide for target recognition, and the two RITS-specific subunits Chp1 and Tas3. Chp1 uses a chromodomain to recognize H3K9me , whereas Tas3 bridges Ago1 and Chp1.

With the ability to interact with both H3K9me and target RNA, RITS plays a central role in the self-enforcing cycle of RNAi-dependent heterochromatin assembly. RITS' function depends on two major interactions. On the one hand, RITS interacts with the RNA-dependent RNA polymerase complex, which synthesizes double-stranded RNA for secondary siRNA generation. On the other hand, RITS interacts (via bridging by the linker protein Stc1) with the Clr4 histone methyltransferase-containing complex that methylates the H3 histone to create the H3K9me epigenetic marker. Thus, the formation of RITS is crucial for the generation of siRNA and for the assembly of RNAi-dependent heterochromatin. The formation of small RNA-containing effector complexes is generally assisted by heat-shock molecular chaperones. However, the heat-shock molecular chaperones responsible for the RNAi-dependent heterochromatin assembly remain unidentified. The candidates may belong to one or more of the distinct families of heat-shock proteins 40, 70, and 90 (Hsp40, Hsp70, and Hsp90, respectively).

Among the three Hsp families, the proteins belonging to the Hsp90 family promote the in vitro formation of small RNA-containing complexes in all species that have been tested. Notably, however, Hsp90-family proteins appear to act in species-specific manners. For example, the steps that require ATP hydrolysis by Hsp90-family proteins appear to differ among various species. For instance, Hsp90-mediated ATP hydrolysis is required for siRNA duplex loading in animal cells, but is instead required for passenger strand removal in plant cells. Similarly, the formation of small RNA-containing complexes does not necessarily require Hsp70-family proteins. An Hsp70 protein is essential for complex formation in the fruit fly *Drosophila melanogaster*, but not in the ciliated protozoan *Tetrahymena thermophila*. Therefore, the differences between species should be acknowledged in examining how such chaperones act in RNAi-dependent heterochromatin assembly.

The *S. pombe* genome encodes six Hsp70 proteins. These Hsps show high sequence similarity to their counterparts in the budding yeast *Saccharomyces cerevisiae*, where the cellular roles of Hsp70 s have been thoroughly examined. Among the six S. pombe Hsp70 proteins, Ssa1 and Ssa2, which show high sequence similarity to each other (identity: 94%), are recognized as nucleocytoplasmic Hsp70 protein. Ssa1 and Ssa2 also exhibit the strongest sequence similarity to the *D. melanogaster* Hsp70 protein Hsc70-4 (identity: 75% each), which is essential for the formation of a small RNA-containing complex in that organism.

The *S. pombe* genome encodes 26 Hsp40 family proteins, all of which harbour a characteristic DnaJ domain. These Hsp40 proteins can be divided into three classes: types I, II, and III. Type-I proteins are also found in *S. cerevisiae* and have the same names in the two yeast species. Mdj1 and Scj1 localize in mitochondria and the lumen of the ER,

respectively. In contrast, Mas5 (also known as Ydj1 in *S. cerevisiae*) and Xdj1 localize in the cytosol and nucleus and are categorized as nucleocytoplasmic type-I Hsp40 proteins. Among the 26 Hsp40 proteins in S. pombe, Mas5 shows the greatest sequence similarity to the *D. melanogaster* protein Droj2 (identity: 41%), a protein that promotes the formation of a small RNA-containing complex in vitro.

RNAi and the RNA Pol II Machinery in Heterochromatin Assembly in *S. pombe* :

RNAi is an important mechanism contributing to heterochromatin formation in *S. pombe*. The phenomenon of RNAi was first discovered in Caenorhabditis elegans in which the expression of double-stranded RNA (dsRNA) abolished the expression of a homologous gene. It soon became apparent that this form of RNAi is related to the process of transcriptional gene silencing (TGS) described in plants and quelling in *N. crassa*. These are processes of silencing that occur when a region is transcriptionally active, and transcripts that generate regions of dsRNA (e.g., through the self-annealing of inverted repeats) can be processed into small RNA fragments (termed small RNA biogenesis). These small RNAs are taken up by effector complexes and can trigger silent chromatin via targeting activities, which cause DNA methylation and histone modification. This process of silencing appears to be in operation from *S. pombe* to plants and metazoans, including mammals. Studies of the components of the RNAi machinery in *S. pombe* have led to significant advances in our understanding of RNAi-mediated chromatin modification and silencing. Mutants in the RNAi machinery in *S. pombe* result in reduced H3K9me2 and loss of silencing over the outer repeats of centromeres. Surprisingly at the time, these RNAi mutants revealed overlapping noncoding RNA (ncRNA) transcripts of a discrete size, originating from centromeric outer repeats. These ncRNAs were homologous to naturally occurring small dsRNAs called small interfering RNAs (siRNAs; 21 nt) that had been isolated and sequenced from S. pombe. These long noncoding double-stranded centromere repeat transcripts are cleaved by the Dicer (Dcr1) enzyme to generate siRNAs. These siRNAs then act to guide the RNAi machinery to homologous transcripts. Mutation of either subunit of RNA Pol II (Rpb2 and Rpb7) results in defective centromere silencing although these mutations display very different phenotypes. The rpb7-1 mutant shows reduced levels of centromere repeat transcription, resulting in less ncRNA and, consequently, less siRNA production and a loss of silent chromatin. This implies that RNA Pol II is required for the transcription of centromere repeats, which then provides the primary substrate for RNAi. In contrast, centromeric transcripts in the rpb2- m203 mutant are produced but not processed into siRNA, and H3K9 methylation at centromeres is reduced. These studies indicate that RNAi not only requires an RNA Pol II transcript but that, like other RNA-processing events, the production of centromeric siRNA may be coupled to transcription by interactions between

the RNAi machinery, chromatin, histone-modifying enzymes, and RNA Pol II. The RNAi machinery in S. pombe is complex and not yet fully understood. In addition to transcription of noncoding centromeric outer repeats by RNA Pol II and the processing of transcripts into siRNAs by Dcr1, the key activities of the RNAi machinery involve two complexes: RITS and RDRC (RNA-directed RNA polymerase complex). The RITS complex incorporates siRNAs to then direct it to centromere outer repeats via sequence recognition as well as H3K9me2/3 recognition through the Chp1 chromodomain. RDRC is recruited to amplify the process of TGS by generating more long double stranded ncRNAs through the action of Rdp1 (RNA-directed RNA polymerase 1). Rdp1 transcribes from transcripts primed with siRNAs presented by the RITS complex. The chromatin-modifying machineries that execute chromatin changes include the CLRC (for Clr4-Rik1-Cul4 complex) and the SHREC complex. CLRC is recruited by the RITS complex via the Stc1 protein. Once recruited, Clr4 methylates H3K9 over the outer repeats and this allows it to bind directly to H3K9me2/3 via its chromodomain. The HP1 homologs Swi6 and Chp2 play further roles in establishing and maintaining heterochromatin. Initiation of transcription, transcriptional elongation, and transcript processing are as important for heterochromatin assembly as they are for euchromatic gene expression. Several associated factors and activities are important for these different steps of the RNA Pol II transcription cycle, in addition to RNA Pol II itself. FACT, an RNA Pol II-associated chromatin assembly factor, and Spt6, another RNA Pol II-associated protein both colocalize to pericentric repeats. Spt6 is specifically required for facilitating trimethylation of H3K9, Swi6 binding, siRNA production, and recruitment of the HDAC enzyme Clr3. Mutations in the FACT component, Pob3, have a similar phenotype to spt6 implicating FACT in the same processes. Interestingly, mutations in chromatin-modifying activities or RNA-processing factors have been shown to suppress the need for Dcr1. For example, loss of Mst2 activity, a H3K14-specific acetyltransferase, completely suppresses dcr1 mutants; that is, it eliminates the need for the RNAi machinery in heterochromatin maintenance, but not in the establishment of new heterochromatin. This suggests that an important role of the RNAi directed heterochromatin process is to prevent Mst2 activity, which might interfere with CLRC recruitment. Another example of bypassing the RNAi pathway is by knocking out the gene encoding Mlo3, involved in mRNP biogenesis and RNA quality control. This also suppresses the need for Dcr1. It was suggested that in mlo3 ago1 knockout cells there is an aberrant accumulation of centromeric transcripts, which results in the recruitment of the CLRC complex via Rik1. In wild-type cells, CLRC acts downstream from the RNAi pathway and is required for heterochromatin silencing by recruiting and promoting the activity of the Clr4 enzyme. Thus, this RNAi-independent recruitment mechanism of CLRC allows the induction of heterochromatin assembly at repeat sequences in mlo3 ago1 knockout cells. Another RNA Pol II linked event is splicing of pre mRNA by the spliceosome. There is not a general requirement for the spliceosome or the splicing process in heterochromatin formation, although some specific splicing factors are required for siRNA production. It remains to be determined how exactly these splicing

factors contribute to RNAi, but their physical interaction with the RDRC suggests a direct role in RNAi.

B. Epigenetic study in *Neurospora crassa*:

The filamentous fungus *Neurospora crassa* has provided a rich source of knowledge on epigenetic phenomena that would have been difficult or impossible to gain from other systems. *Neurospora* sports features found in higher eukaryotes but absent in both budding and fission yeast, including DNA methylation and H3K27 methylation, and also has distinct RNA interference (RNAi)-based silencing mechanisms operating in mitotic and meiotic cells. This has provided an unexpected wealth of information on gene silencing systems. One silencing mechanism, named repeat-induced point mutation (RIP), has both epigenetic and genetic aspects and provided the first example of a homology-based genome defence system. A second silencing mechanism, named quelling, is an RNAi-based mechanism that results in silencing of transgenes and their native homologs. A third, named meiotic silencing, is also RNAi-based but is distinct from quelling in its time of action, targets, and apparent purpose.

Fungi provide excellent models for understanding the structure and function of chromatin both in actively transcribed regions (euchromatin) and in transcriptionally silent regions (heterochromatin). The budding yeast, *Saccharomyces cerevisiae*, has been an invaluable eukaryotic model for studying chromatin structure associated with transcription at euchromatic regions and providing a paradigm for silent chromatin. The fission yeast, *Schizosaccharomyces pombe*, has some epigenetic machinery that is absent from *S. cerevisiae* but common in higher organisms—most notably for RNA interference (RNAi) and for methylation of lysine 9 of histone H3 (H3K9me). Research using *S. pombe* has provided invaluable information on the structure and function of heterochromatin, principally found in regions of the centromeres, telomeres, and silent mating-type genes. This article focuses on a third model system, namely the filamentous fungus *Neurospora crassa*. Although not as commonly studied as the yeasts, *Neurospora* has proved to be a remarkably rich source of knowledge that would have been difficult or impossible to gain from other systems. *Neurospora* sports features found in higher eukaryotes, including DNA methylation and the H3K27 methylation ("Polycomb") system that both budding and fission yeasts lack, as well as RNAi and other epigenetic processes found in the yeasts. This has provided an unexpected wealth of information on gene silencing systems, some of which operate at distinct stages of its life cycle. The first such mechanism, named repeat-induced point mutation (RIP), has both epigenetic and genetic aspects and provided the first example of a homology-based genome defence system. The second, named quelling, is an RNAi-based mechanism that results in silencing of transgenes and their native homologs. The third, named meiotic silencing (or meiotic silencing by unpaired DNA), is

also RNAi-based but is distinct from quelling in its time of action, targets, and apparent purpose. Although we are still in the early days of epigenetic studies in all organisms, it is already clear that yeasts and filamentous fungi such as *N. crassa* will continue to serve as rich sources of information on epigenetic mechanisms operative in a broad range of eukaryotes.

RNAi and Heterochromatin Assembly :

The intersection between RNA interference (RNAi) and heterochromatin formation has brought together two areas of gene regulation that had previously been thought to operate by different, perhaps even unrelated, mechanisms. Heterochromatin was originally defined nearly 80 years ago using cytological staining methods as those chromosome regions that retain a condensed appearance throughout the cell cycle. Early investigators studying the relationship between chromosome structure and gene expression noticed that certain chromosome rearrangements resulted in the spreading of heterochromatin into adjacent genes, which then became silent. But, the seemingly stochastic patterns of spreading gave rise to genetically identical populations of cells that had different phenotypes. This phenomenon, initially described in *Drosophila* as position-effect variegation, provides a striking example of epigenetic regulation. The term RNAi was first used to describe gene silencing that resulted from the introduction of homologous antisense or double-stranded RNA (dsRNA) into the nematode *Caenorhabditis elegans*. But, it was soon recognized that a related mechanism involving RNA accounted for posttranscriptional transgene silencing (PTGS) described earlier in petunia and tobacco. In contrast, heterochromatin was widely believed to operate directly at the chromatin level to cause transcriptional repression by a mechanism referred to as transcriptional gene silencing (TGS). This article focuses on the relationship between the RNAi pathway and the formation of epigenetically heritable heterochromatin at specific chromosome regions. It draws on recent examples that show this relationship in the fission yeast *Schizosaccharomyces pombe* and the mustard plant *Arabidopsis thaliana*.

The fission yeast nuclear genome is composed of three chromosomes that range in size from 3.5 to 5.7 Mb. Each chromosome contains large blocks of repetitive DNA, particularly at centromeres, which are packaged into heterochromatin. The mating-type loci (which control cell type) and subtelomeric DNA regions also contain repetitive sequences that are packaged into heterochromatin. We now know that the assembly of DNA into heterochromatin plays both regulatory and structural roles. In the case of the mating-type loci in yeast, regulation of gene transcription by heterochromatin is important for cell-type identity. In the case of telomeres and centromeres, heterochromatin plays a structural role that is important for proper chromosome segregation during cell division. Moreover, repetitive DNA sequences and transposable elements account for a large fraction, in some cases more than half, of the genomes of many eukaryotic cells. Heterochromatin and

associated mechanisms play a critical role in regulating the activity of repeated sequences, thus maintaining genome stability.

Recent studies have uncovered a surprising requirement for components of the RNAi pathway in the process of heterochromatin formation in fission yeast and have provided insight into how these two pathways can work together at the chromatin level. Briefly, small interfering RNA (siRNA) molecules and their Argonaute-binding proteins assemble into the RNA-induced transcriptional silencing (RITS) complex and direct epigenetic chromatin modifications and heterochromatin formation at complementary chromosome regions. RITS uses siRNA-dependent base pairing to guide association with nascent RNA sequences at the target locus destined to be silenced, an association that is stabilized by direct binding to methylated histone H3 at lysine (K)9 (H3K9me). The presence of these two activities in RITS (i.e., siRNA base-pairing and association with chromatin via methylated H3K9) triggers heterochromatin formation in concert with well-known heterochromatin-associated factors, and RNA polymerase II (Pol II) directly linking RNA silencing to heterochromatin modification and silencing.

In *A. thaliana* and many other eukaryotes, repeat sequences such as retroelements and other transposons are targeted for inactivation at the chromatin level by mechanisms that couple small RNA-mediated targeting with histone H3K9, but also DNA methylation. Although the existence of a RITS complex is not always clear, components of the RNAi and related pathways are required for the initiation and maintenance of these repressive methylation events, along with Pol II and related polymerases. In this article, we will discuss how heterochromatic siRNAs are produced, and how they mediate DNA and/or chromatin modifications in fission yeast and *A. thaliana*.

Role of Non Coding RNA

Non-coding RNA (ncRNA) is a functional RNA molecule that is transcribed from DNA but not translated into proteins.

Epigenetic related ncRNAs include miRNA, siRNA, piRNA and lncRNA. In general, ncRNAs function to regulate gene expression at the transcriptional and post-transcriptional level. Those ncRNAs that appear to be involved in epigenetic processes can be divided into two main groups; the short ncRNAs (<30 nts) and the long ncRNAs (>200 nts). The three major classes of short non-coding RNAs are microRNAs (miRNAs), short interfering RNAs (siRNAs), and piwi-interacting RNAs (piRNAs). Both major groups are shown to play a role in heterochromatin formation, **histone modification**, **DNA methylation** targeting, and gene silencing.

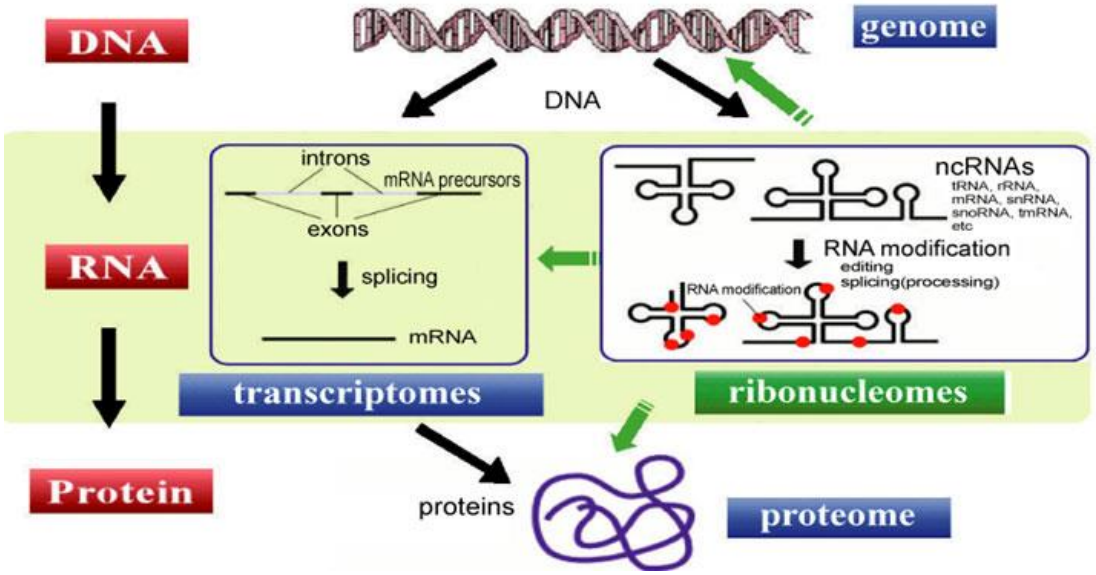


Fig: In vivo function of NcRNA in living organisms

Short ncRNAs

MicroRNAs (miRNA) generally bind to a specific target messenger RNA with a complementary sequence to induce cleavage, or degradation or block translation. This may be done in the context of a feedback mechanism that involves chromosome methylation. For example, miRNA genes mir-127 and mir-136 were found to be involved in regulating the genetic imprinting of Rtl1, a key gene involved in placenta formation in mice. Methylation of a specific region in the paternal chromosome results in expression of Rtl1. If the chromosome is not methylated, as on the maternal chromosome, mir-127 and mir-136 are produced and bind to the Rtl1 transcript and induce degradation. Lack of Rtl1 protein expression due to improper epigenetic modifications can result in foetal death in mice.

Short interfering RNAs (siRNA) function in a similar way as miRNAs to mediate post-transcriptional gene silencing (PTGS) as a result of mRNA degradation. In addition to this function, siRNAs have also been shown to induce heterochromatin formation via an RNA-induced transcriptional silencing (RITS) complex which when bound to siRNA promotes H3K9 methylation and chromatin condensation.

Piwi-interacting RNAs (piRNA) are so named due to their interaction with the piwi family of proteins. The primary function of these RNA molecules involves chromatin regulation and suppression of transposon activity in germline and somatic cells. PiRNAs that are antisense to expressed transposons target and cleave the transposon in complexes with PIWI-proteins. This cleavage generates additional piRNAs which target and cleave

additional transposons. This cycle continues to produce an abundance of piRNAs and augment transposon silencing.

Long ncRNAs

Many lncRNAs can complex with chromatin-modifying proteins and recruit their catalytic activity to specific sites in the genome, thereby modifying chromatin states and influencing gene expression. The majority of non-coding RNA transcripts belong to the group lncRNAs. Long ncRNAs function in chromatin remodelling, transcriptional regulation, post-transcriptional regulation, and as precursors for siRNAs.**7** One particular subgroup of lncRNAs, the large intergenic non-coding RNAs (lincRNAs), has been associated with chromatin modifying complexes which can target specific genomic loci to promote specific epigenetic states. One widely known example of this is the role of X-inactive specific transcript gene (Xist), in X-chromosome inactivation (XCI). This process involves two lncRNAs; Xist and its antisense transcript Tsix, a negative regulator of Xist. Prior to differentiation, Xist and Tsix are actively transcribed due to H3K4 dimethylation of the Xist gene. In this state XCI is a random event. Upon differentiation, Xist expression is elevated resulting in Xist RNA coating the future inactive X chromosome which triggers extensive histone methylation and chromosome inactivation.

Probable Questions:

1. How epigenetics affect *Schizosaccharomyces pombe* ?
2. What is the role of RNAi and RNA Pol II Machinery in Heterochromatin Assembly in *S. pombe ?*
3. Describe different types of short non coding RNAs .
4. Describe different types of long non coding RNAs .
6. How epigenetics affect *Neurospora crassa* ?

Suggested Readings:

1. Molecular Cell Biology by Lodish, Fourth Edition.
2. The Cell – A Molecular Approach by Cooper and Hausman, Fourth Edition
3. Principles of Genetics by Snustad and Simmons, Sixth Edition.
4. Molecular Biology of the Cell – by Bruce Alberts
5. Cell and Molecular Biology by Gerald Karp, 7th Edition.
6. Gene cloning and DNA Analysis by T. A. Brown, Sixth Edition.
7. Brown TA. (2010) Gene Cloning and DNA Analysis. 6th edition. Blackwell Publishing, Oxford, U.K.

UNIT-IV

Chromatin structure and epigenetics marks - transcriptional silencing by polycomb group proteins , transcriptional regulation by trithorax group proteins, histone variants and epigenetics, epigenetic regulation of chromosome inheritance

Objective: In this unit we will discuss Chromatin structure and epigenetics marks - transcriptional silencing by polycomb group proteins, transcriptional regulation by trithorax group proteins, histone variants and epigenetics, epigenetic regulation of chromosome inheritance

Underlying mechanism of Chromosomal epigenetic inheritance:

Based on of the nucleosome modifications that occur at the silent mating-type region in yeast, Thon and colleagues developed a mathematical model of histone PTM regulation in the context of a two state process. The authors assumed that the model is applicable for bi-stable conditions meaning that transition from one state to another occurs but at a relatively low frequency. The bi-stability is efficiently maintained through replication assuming that histone PTMs are randomly segregated during replication and that newly deposited nucleosomes are naïve (i.e. devoid of the histone modifications regulating the transition from one state to another and are therefore potential targets for these modifications). We mentioned in the previous section that newly incorporated histones do not carry the PTMs potentially involved in transmitting epigenetic information. The bi-stability requires cooperativity between the histone modifications and positive feedback for the spreading of a defined mark. Furthermore this positive feedback should propagate the mark not only linearly, i.e. to adjacent nucleosomes, but also reach targets several nucleosomes away, thus highlighting the importance of higher order chromatin structure. In the next sections, we will consider known mechanisms that could be critical for the propagation of epigenetic information in light of these theoretical predictions, which we believe to be of great importance in understanding transmission of histone marks and the establishment of chromatin domains that are inherited.

Importantly, while we have some idea about the propagation of DNA methylation and models are emerging for the duplication of histone PTMs, very little is known about inheritance of nucleosome occupancy or histone variants. It was reported that the silencing of the *MLH1* promoter in cancer is associated with specific changes in nucleosome occupancy. Considering that this promoter is also specifically DNA methylated in cancer cells and that changes in nucleosome occupancy are reversed by inhibition of DNA methylation, nucleosome occupancy in this case might be regulated by DNA methylation.

Regarding histone variants, a better characterization of the histone deposition machinery will be required to understand how they might be stably maintained at a defined locus.

Polycomb Group (PcG) of Proteins
What are PcG proteins?

Polycomb-group (PcG) genes encode chromatin proteins involved in stable and heritable transcriptional silencing. PcG proteins participate in distinct multimeric complexes that deposit, or bind to, specific histone modifications (e.g., H3K27me3 and H2AK119ub1) to prevent gene activation and maintain repressed chromatin domains. PcG proteins are evolutionary conserved and play a role in processes ranging from vernalization and seed development in plants, over X-chromosome inactivation in mammals, to the maintenance of stem cell identity. PcG silencing is medically relevant as it is often observed in human disorders, including cancer, and tissue regeneration, which involve the reprogramming of PcG-controlled target genes.

Brief idea about transcriptional memory

Organs of humans, animals, and plants are constructed from a large pool of distinct cell types, each performing a specialized physiological or structural function. With very few exceptions, all cell types contain the same genetic information encoded in their DNA. Thus, the distinctiveness of a given cell type is achieved through specific gene expression programs. As a consequence, cell lineages need to have these programs maintained during growth and cell division. This implies the existence of a memory system that ensures the faithful transmission of information (i.e., which gene is active or repressed) from mother to daughter cells. The existence of such a system is illustrated by the fact that cultured tissues of plants and animals usually maintain their differentiated characters even if grown in a foreign environment. By way of example, ivy plants regenerated after tissue culture produce the type of leaf corresponding to the phase of development from which the original tissue was taken (i.e., juvenile or adult leaf).

Brief idea on the roles and effects of PcG proteins

The major question to be addressed here concerns the molecular identity of factors contributing to the mechanism(s) of “cellular” or “transcriptional memory,” which maintains a determined state over many cell divisions. Genetic analyses in *Drosophila melanogaster* have identified regulators crucial in maintaining the morphology of individual body segments that are determined by the action of the HOX genes. In *Drosophila*

males, the first thoracic segment has legs with sex combs. Legs on the second and third thoracic segment lack these structures. In the 1940s, *Drosophila* mutants were identified (Polycomb and extra sex combs) wherein males had sex combs on all legs. These morphological alterations reflect homeotic transformations of the second and third leg identities into the first leg identity. Subsequent molecular studies showed that these mutations did not affect the products of the HOX genes themselves, but rather the way HOX gene activity was spatially controlled. Throughout the years, a large number of similar regulatory genes were identified, and were classified into two antagonistic groups: the Polycomb (PcG) and Trithorax (TrxG) group. Whereas the PcG proteins are required to maintain the silenced state of developmental regulators such as the HOX genes, the TrxG proteins are generally involved in maintaining the active state of gene expression. Thus PcG and TrxG proteins embody the molecular components of cellular memory.

Proteins of both groups form large multimeric protein complexes that act on their target genes by modulating chromatin structure. In *Drosophila*, it was shown that transcription factors recruit PcG complexes to a DNA sequence called a PcG response element (PRE). Once recruited, they establish a silent chromatin state that can be inherited over many cell divisions. Members of PRC2 are highly conserved between plants and animals, whereas PRC1 proteins are less well conserved. This implies conservation, but also diversity, in the basic building blocks of the cellular memory system. In addition to the function of PcG complexes in the maintenance of cell types, they may also play important roles in stem cell plasticity and regeneration. Also, their deregulation can lead to neoplastic transformation and cancer. Thus, PcG proteins play a crucial role in many fundamental processes of normal development and disease in multicellular eukaryotes.

Transcriptional Regulation by Trithrorax group (trxG) of proteins

What are trxG proteins?

The trithorax group of genes (trxG) was identified in mutational screens that examined developmental phenotypes and suppression of *Polycomb* mutant phenotypes. The protein products of these genes are primarily involved in gene activation, although some can also have repressive effects. There is no central function for these proteins. Some move nucleosomes about on the genome in an ATP-dependent manner, some covalently modify histones such as methylating lysine 4 of histone H3, and some directly interact with the transcription machinery or are a part of that machinery. It is interesting to consider why these specific members of large families of functionally related proteins have strong developmental phenotypes.

Brief idea about the roles and effects of trxG in Transcriptional Memory:

All cells in an organism must be able to "remember" what type of cell they are meant to be. This process, referred to as "cellular memory" or "transcriptional memory," requires two basic classes of mechanisms. The first class, functions to maintain an "off" state for genes that, if turned on, would specify an inappropriate cell type. The Polycomb-group (PcG) proteins have as their primary function a repressive role in cellular memory. The second class of mechanisms is composed of those that are required to maintain key genes in an "on" state. Any cell type requires the expression of master regulatory proteins that direct the specific functions required for that cell type. The genes that encode these master regulatory proteins must be maintained in an "on" state throughout the lifetime of an organism to maintain the proper cell types within that organism.

The proteins that are involved in maintaining the "on" state are called trithorax-group (trxG) proteins in honor of the trithorax gene, the founding member of this group of regulatory proteins. A large group of proteins with diverse functions make up the trxG. The roles these proteins play in the epigenetic mechanisms that maintain the "on" state appear more complex at this juncture than the roles for PcG proteins in repression. The first complexity is that a very large number of proteins and mechanisms are needed to actively transcribe RNA from any gene. Thus, in contrast to repression, which might be accomplished by comparatively simple mechanisms that block access of all proteins, activation of a gene requires numerous steps, any of which might play a role in maintaining an "on" state. Thus, there are numerous possible stages in which a trxG protein might work.

A second complexity in thinking about trxG proteins is that proteins that function in activation can also, in different contexts, function in repression. This might appear counterintuitive, but depending on the precise architecture of a gene, the same protein performing its function might in one case help a gene become activated, and in another case help a different gene become repressed. At this time it does not appear that trxG proteins are dedicated solely to the maintenance of gene expression, but that these proteins can also play multiple roles in the cell. These complexities make for several interesting unanswered questions. Why are only some of the proteins needed to activate transcription also critical for maintenance of transcription? Do these proteins have functions that are uniquely suited to maintaining the active state? Or are some of these proteins needed for maintenance solely because an evolutionary accident that made them key regulators of gene(s) particularly important to development?

As we will see when we discuss mechanisms of action, some of the trxG proteins are involved in regulating chromatin structure in opposition to the mechanisms used by the PcG proteins. trxG proteins can place covalent posttranslational modifications (PTMs) on chromatin or can alter chromatin by changing the structure and position of the nucleosomes that are the building blocks of chromatin. Other trxG proteins function as part of the transcription machinery. Thus, these proteins are found in a wider variety of

complexes than the PcG proteins and are likely to play more complicated roles in epigenetic mechanisms.

Numerous developmental decisions—including the determination of cell fates—are made in response to transient positional information in the early embryo. These decisions are dependent on changes in gene expression. This allows cells with identical genetic blueprints to acquire unique identities and follow distinct pathways of differentiation. The changes in gene expression underlying the determination of cell fates are heritable; a cell's fate rarely changes once it is determined, even after numerous cell divisions and lengthy periods of developmental time. Understanding the molecular mechanisms underlying the maintenance of the determined state has long been a goal of developmental and molecular biologists.

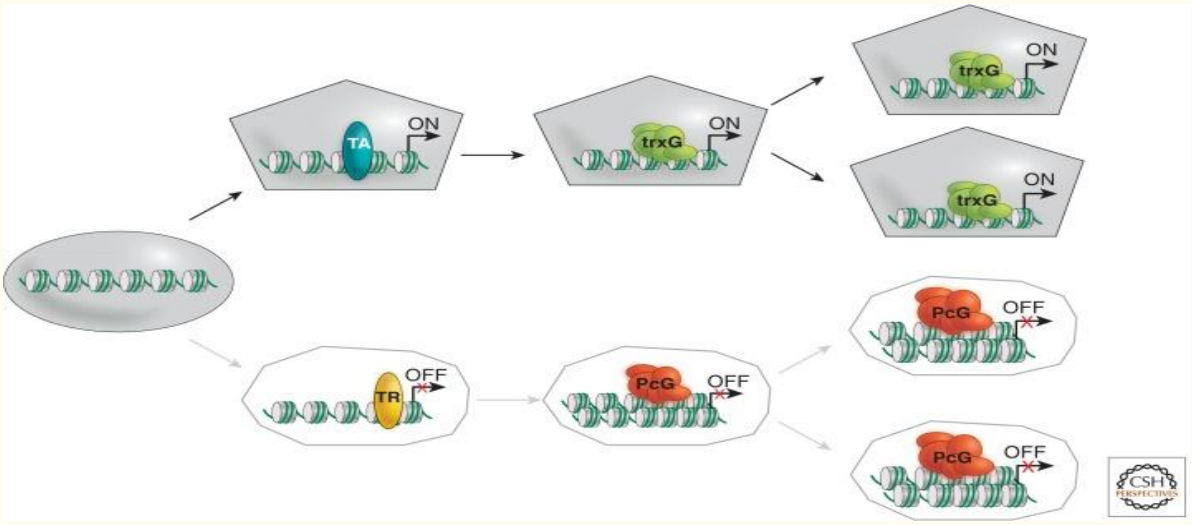


Figure 1.

The concept of cellular memory. Schematic illustration highlighting the role of trxG complexes in maintaining heritable states of active gene expression in contrast to heritable silencing by PcG complexes, as defined originally for the *Drosophila* homeotic (Hox) gene cluster.

Many of the regulatory proteins involved in the maintenance of heritable states of gene expression were identified in studies of *Drosophila* homeotic (Hox) genes. Hox genes encode homeodomain transcription factors that regulate the transcription of batteries of downstream target genes, which in turn specify the identities of body segments. In *Drosophila*, Hox genes are found in two gene complexes: the Antennapedia complex (ANT-C), which contains the Hox genes *labial* (*lab*), *Deformed*, *Sex combs reduced* (*Scr*),

and *Antennapedia* (*Antp*); and the bithorax complex, which contains the Hox genes *Ultrabithorax* (*Ubx*), *abdominalA* (*abdA*), and *AbdominalB* (*AbdB*;). Each Hox gene specifies the identity of a particular segment, or group of segments, along the anterior–posterior axis of the developing fly. For example, *Antp* specifies the identity of the second thoracic segment, including the second pair of legs, whereas *Ubx* specifies the identity of the third thoracic segment, including the balancer organs located behind the wings. Thus, the transcription factors encoded by Hox genes function as master regulatory switches that direct the choice between alternative pathways of development.

The transcription of Hox genes must be regulated precisely because dramatic alterations in cell fates can result from their inappropriate expression. For example, the derepression of *Antp* in head segments transforms antennae into legs, whereas the inactivation of *Ubx* in thoracic segments transforms balancer organs into wings. In *Drosophila*, the initial patterns of Hox transcription are established early in embryogenesis by transcription factors encoded by segmentation genes. The proteins encoded by segmentation genes—including the gap, pair-rule, and segment polarity genes—subdivide the early embryo into 14 identical segments. These proteins also establish the initial patterns of Hox transcription, the first step toward the development of segments with distinct identities and morphology. The majority of segmentation genes, however, are transiently expressed during early development. Once established, the segmentally restricted patterns of Hox transcription must be maintained throughout subsequent embryonic, larval, and pupal stages to maintain the identities of the individual body segments. This function is performed by two other groups of regulatory proteins: the Polycomb group of repressors (PcG) and the trithorax group of transcriptional regulators (trxG). The regulation of Hox transcription therefore consists of at least two distinct phases: establishment (by segmentation genes) and maintenance (by PcG and trxG genes; Fig. 2).

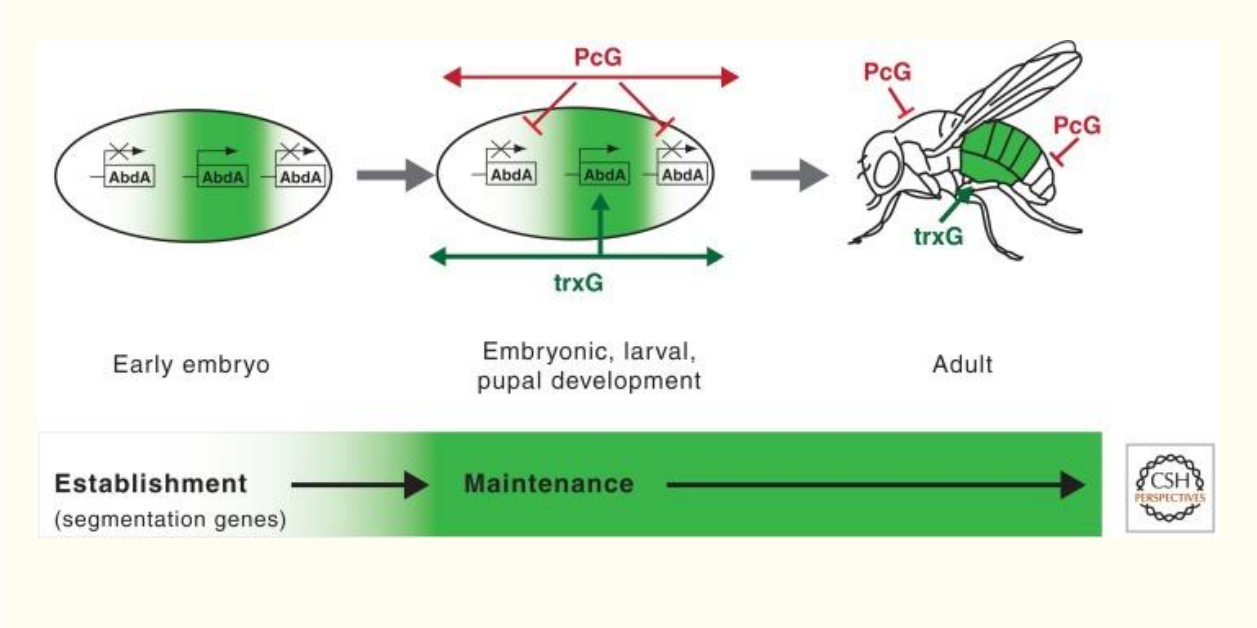


Figure 2.

Regulation of Hox transcription. The boundaries of *abd-A* transcription and other Hox genes are established by segmentation proteins. These include the products of gap and pair-rule genes, which subdivide the embryo into 14 identical segments. During subsequent development, the "off" or "on" states of Hox transcription are maintained by the ubiquitously expressed members of the trxG of activators and the PcG of repressors via mechanisms that remain poorly understood.

Histone Variants and Epigenetics

Histones package DNA by assembling into nucleosome core particles, whereas the double helix wraps around. Over evolutionary time, histone-fold domain proteins have diversified from archaeal ancestors into the four distinct subunits that comprise the familiar octamer of the eukaryotic nucleosome. Further diversification of histones into variants results in differentiation of chromatin that can have epigenetic consequences. Investigations into the evolution, structure, and metabolism of histone variants provide a foundation for understanding the participation of chromatin in important cellular processes and in epigenetic memory. Most histones are synthesized at S phase for rapid deposition behind replication forks to fill in gaps resulting from the distribution of pre-existing histones. In addition, the replacement of canonical S-phase histones by variants, independent of replication, can potentially differentiate chromatin. The replacement of a canonical histone by a noncanonical variant is a dynamic process that changes the composition of chromatin.

The differentiation of chromatin by a histone variant is especially conspicuous at centromeres, in which the H3 variant, CENP-A, is assembled into specialized nucleosomes that form the foundation for kinetochore assembly. A centromeric H3 (cenH3) counterpart of CENP-A is found in all eukaryotes. In plants and animals, the faithful assembly of cenH3-containing nucleosomes at centromeres does not appear to require centromeric DNA sequences, a spectacular example of epigenetic inheritance. Some cenH3s have evolved adaptively in regions that contact DNA, which suggests that centromeres compete with each other, and cenH3s and other centromere-specific DNA-binding proteins have adapted in response. This process could account for the large size and complexity of centromeres in plants and animals.

Chromatin can also be differentiated outside of centromeres by incorporation of a constitutively expressed form of H3, called H3.3, which is the substrate for replication-independent nucleosome assembly. Replacement with H3.3 occurs at active genes, a dynamic process with potential epigenetic consequences. Differences between H3 and H3.3 in their complement of covalent modifications might underlie changes in the properties of chromatin at actively transcribed loci.

Several H2A variants can also differentiate or regulate chromatin. H2A.X is defined as a variant by a four-amino-acid carboxy-terminal motif whose serine residue is the site for phosphorylation at sites of DNA double-stranded breaks. Phosphorylation of H2A.X is an early event in double-strand break repair, in which it is thought to concentrate components of the repair machinery. H2A.X phosphorylation also marks the inactive XY bivalent during mammalian spermatogenesis and is required for condensation, pairing, and fertility.

H2A.Z is a structurally diverged variant that has long presented an enigma. Studies in yeast have implicated H2A.Z in establishing transcriptional competence and in counteracting heterochromatic silencing. The biochemical complex that replaces H2A with H2A.Z in nucleosomes is an ATP-dependent nucleosome remodeler, providing the first example of a specific function for a member of this diverse class of chromatin-associated machines.

Two vertebrate-specific variants, macroH2A and H2A.B (also called H2A.Bbd), display contrasting features when packaged into nucleosomes in vitro, with macroH2A impeding and H2A.B facilitating transcription. These features are consistent with their localization patterns on the epigenetically inactivated mammalian X chromosome: macroH2A showing enrichment and H2A.B showing depletion. The emerging view from these studies is that histone variants and the processes that deposit them into nucleosomes provide a primary differentiation of chromatin that might serve as the basis for epigenetic processes.

Probable questions:

1. What are PcG proteins?
2. Discuss Briefly about transcriptional memory.
3. Discuss roles and effects of PcG proteins.
4. What are trxG proteins?
5. Discuss roles and effects of trxG in Transcriptional Memory.
6. Discuss Histone Variants and Epigenetics in brief.

Suggested Readings:

1. Molecular Cell Biology by Lodish, Fourth Edition.
2. The Cell – A Molecular Approach by Cooper and Hausman, Fourth Edition
3. Principles of Genetics by Snustad and Simmons, Sixth Edition.
4. Molecular Biology of the Cell – by Bruce Alberts
5. Cell and Molecular Biology by Gerald Karp, 7th Edition.
6. Gene cloning and DNA Analysis by T. A. Brown, Sixth Edition.
7. Genetics . Verma and Agarwal.
8. Brown TA. (2010) Gene Cloning and DNA Analysis. 6th edition. Blackwell Publishing, Oxford, U.K.

UNIT-V

Epigenetic regulation of the X chromosomes in *C.elegans*, dosage compensation in *Drosophila*, dosage compensation in mammals

UNIT-VI

Types and mechanism of chromatin remodeling

Objective: In this unit we will discuss about Epigenetic regulation of X chromosome, dosage compensation in human and Drosophila. We will also discuss types and mechanism of chromatin remodelling.

THE PHENOMENON OF DOSAGE COMPENSATION WAS DISCOVERED IN DROSOPHILA:

The karyotypes (i.e., ensemble of chromosomes) of many organisms include a pair of sex chromosomes. In Drosophila, females have two sex chromosomes called the X chromosomes that are identical in shape and genetic content; both X chromosomes are active in all somatic cells. Males have one X and a Y chromosome that differs from the X in morphology and genetic information that it contains. On the sex chromosomes there are genes that are responsible for sex determination and sexual differentiation. The Y chromosome is male specific, but the X chromosome carries many genes involved in basic cellular housekeeping functions or developmental pathways. Females with two X chromosomes have twice the number of these genes; males with a single X have only one dose. Yet, the level of the products of most of these genes is the same in the two sexes.

In the early 1930s, this paradox was first noticed in Drosophila by H.J. Muller while he was studying the eye pigment level of individuals carrying partial loss-of-function X-linked mutations (Muller 1932). Muller reasoned that there must be a regulatory mechanism that helps flies to compensate for the difference in dosage of X-linked genes in males and females by equalizing the level of X-linked gene products between the two sexes. He called this hypothetical regulatory mechanism “dosage compensation”(Fig. 1).

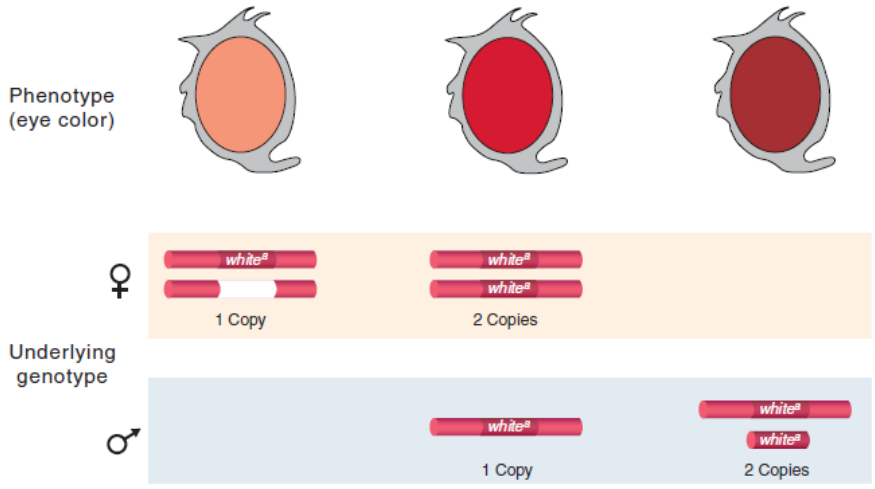


Figure 1. Diagrammatic representation of the results that led H.J. Muller to formulate the hypothesis of dosage compensation. The mutant allele of the X-linked *white* gene (w^a) is a hypomorph and allows partial eye-pigment synthesis; its presence on the X chromosomes is indicated. The level of pigmentation is directly proportional to the dosage of the w^a allele within each sex; yet, males with one dose and females with two doses have comparable amounts of pigment because of dosage compensation.

Following its discovery in Drosophila, the phenomenon of dosage compensation was observed in additional species. In organisms belonging to distantly related groups—from round worms to mammals—transcriptional regulation leading to equal products of X-linked genes in males and females has been achieved in different ways: by decreasing the level of transcription of the two doses of X-linked genes in hermaphrodites relative to males (Caenorhabditis elegans) or by hypertranscribing the X chromosome in both males and females and then shutting down one of the two X chromosomes throughout most of its length in the somatic cells of females (mammals). The mechanisms underlying dosage compensation in these forms are described in Strome et al. (2014) and Brockdorff and Turner (2014).

The first evidence that dosage compensation in Drosophila is achieved by regulating the transcription of X-linked genes was obtained more than 30 years after Muller's seminal observations, by A.S. Mukherjee and W. Beermann (Mukherjee and Beermann 1965). Using transcription autoradiography of the giant polytenic chromosomes of larval salivary glands, a molecular technique that represented the state of the art at that time, these investigators observed that the level of [3H]uridine incorporation by the single X in males and both Xs in females was equivalent. It appeared, therefore, that the rate of RNA synthesis by the single X chromosome in males was approximately twice the rate of each of the two Xs in females. The

next experimental breakthrough consisted of the genetic identification by J. Belote and J. Lucchesi of four genes: msl1, msl2, msl3, and mle, with loss-of-function mutations that appeared inconsequential in females but lethal in males; notably, the mutant males showed approximately half of the normal level of [3H]uridine incorporation by their X chromosome

(Belote and Lucchesi 1980a,b). Furthermore, the X chromosome had lost its normal paler and
somewhat puffed appearance that had been interpreted as an indication of an enhanced level of transcriptional activity in relation to each of the two X chromosomes in females. These results suggested that the equalization of X-linked gene products was achieved by doubling, on average, the transcriptional activity of the X chromosome in males rather than by halving the transcriptional activity of each X in females.

An alternate hypothesis was proposed based on an "inverse dosage effect," in which the activity of all chromosomes is set by general transcriptional regulators (reviewed in Birchler et al. 2011). In males, because of the absence of one X chromosome, a greater concentration of these regulators would be available than in females, driving the activity of all chromosomes to higher levels. For appropriate compensation to occur, the products of the msl loci would sequester some of these regulators away from the autosomes in males, thus leaving only the X chromosome with increased expression. In this model, msl gene mutations result in elevation of the expression of autosomal genes rather than a reduction of X-linked gene expression.

However, a number of experimental results are incompatible with the inverse hypothesis (Arkhipova et al. 1997; Hamada et al. 2005; Straub et al. 2005; Deng et al. 2011). Particularly compelling is the recent observation that ectopic MSL complex on autosomes leads to a localized increase in transcription and suppression of phenotypes caused by haplo-insufficient mutants in the same region (Park et al. 2010).

Among the four genes introduced above, two were newly discovered (male-specific lethal 1, msl1; and male-specific lethal 2, msl2), whereas the other two (maleless, mle; and male-specific lethal 3, msl3) had been previously identified by other investigators in natural populations (specific references to this early phase of the study of dosage compensation can be found in Lucchesi and Manning 1987). For ease of reference, all of the gene products identified to date, on the basis of the male-specific lethal phenotype of their loss-of-function mutations, are called the MSLs. The next phase in the study of dosage compensation was initiated with the cloning of mle and the three msl genes, and the discovery and cloning of the mof histone acetyltransferase gene. By cytoimmunofluorescence, the five gene products were found to associate in an identical pattern at numerous sites along the polytene X chromosome in males
(reviewed in Gelbart and Kuroda 2009). This observation and the interdependence of the different gene products for X-chromosome binding suggested that they form a complex. It is crucial for viability that the complex is present in males (XY) and absent in females (XX); therefore, the first step in dosage compensation is to establish this sex specificity.

REGULATORS OF DOSAGE COMPENSATION:

Regulation of Dosage Compensation Starts with Counting the Number of X Chromosomes:

Each embryo needs to count its X chromosomes to make the critical decision whether or not to implement dosage compensation. An incorrect decision, such as failure to upregulate the single male X or aberrant up-regulation of both female XXs, results in lethality. In Drosophila, the X-counting process is coordinated with the sex determination decision. Phenotypic sex is determined by the number of X chromosomes per nucleus, such that XX embryos are females and XY embryos are male. The Y chromosome is required for male fertility, but unlike in mammals, it plays no role in phenotypic sex. Formally it is the X:autosome ratio that controls both sex and dosage compensation, as the X counting mechanism is sensitive to the number of sets of autosomes. This becomes apparent in 2X:3A triploids, which have an intermediate X:A ratio between XY:2A males and XX:2A females. 2X:3A triploids differentiate as intersexes with a mixture of both male and female cells.

The X:A ratio controls both sex determination and dosage compensation by regulating a critical binary switch gene, Sex lethal (Sxl). Sxl encodes a female-specific RNA binding protein that regulates splicing and translation of key messenger RNAs (mRNAs) in the sex determination

and dosage compensation pathways respectively (Fig. 2). The Sex lethal gene resides on the X chromosome and is positively regulated by transcription factors encoded by the X, such that embryos with two X chromosomes are able to initiate Sxl expression from an early, regulated promoter, Pe, whereas embryos with a single X per nucleus fail to express Sxl from Pe. This initial transient difference in activation of Sxl in early embryos is stabilized by an autoregulatory loop in which SXL protein positively regulates splicing of its own mRNA from a maintenance promoter that is expressed constitutively. SXL initiates differentiation n the female mode by regulating the splicing of the transformer (tra) gene in a sex-specific manner. In turn, this gene product (together with the product of another gene, transformer2 (tra2), present in both sexes) directs the splicing of the doublesex (dsx) primary transcript to yield a regulatory protein that acts to repress genes required for male development, thus achieving female sexual differentiation. In male embryos, an alternate mode of splicing of the dsx transcripts occurs by default and leads to a product that represses genes required for female development, resulting in male sexual differentiation.

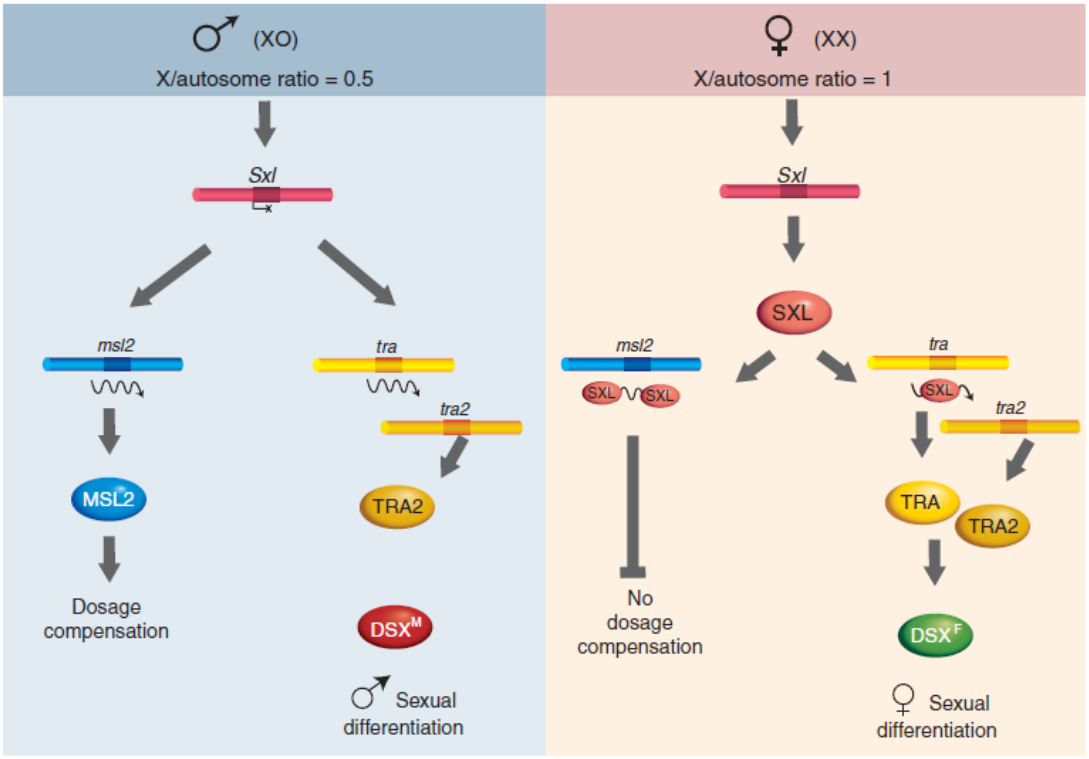


Figure 2. Diagram of the control of sex determination and dosage compensation. If the X/A ratio is equal to 1, a regulatory cascade leads to female sexual development. In females, the presence of the *Sxl* gene product prevents the translation of the *msl2* message and the assembly of the MSL complex. If the X/A ratio is only 0.5, absence of the cascade leads by default to male sexual development and to the formation of the MSL complex.

The SXL Protein Prevents Formation of the MSL Complex in Females:

The key target of SXL in the dosage compensation pathway is msl2 mRNA (Bashaw and Baker 1997; Kelley et al. 1997). SXL binding sites are located in both the 5′ and 3′ untranslated regions (UTRs) of msl2 mRNA. SXL is normally present only in females, in which it represses translation of the msl2 mRNA through association with its UTRs (see Fig. 2). If SXL is absent in females, dosage compensation is aberrantly turned on and these females die. Conversely, if SXL is ectopically expressed in males, dosage compensation is turned off and males die. Ectopic expression of MSL2 in females is sufficient to assemble MSL complexes on both female X chromosomes, indicating that all other MSL components are either turned on or stabilized by expression of MSL2.

In summary, dosage compensation must respond to the number of X chromosomes in the nucleus, and these are counted early in embryonic development. Females repress MSL2 translation, preventing inappropriate dosage compensation when two X chromosomes are present. In the absence of SXL-mediated repression, males express MSL2 protein and this leads to the assembly of a functional MSL complex.

CHROMATIN MODIFICATIONS ASSOCIATED WITH DOSAGE COMPENSATION:

A key modification that is correlated with the association of the MSL complex with the X chromosome in males is the presence of a high level of histone H4 acetylated at lysine 16 (Turner et al. 1992; Bone et al. 1994). This chromatin mark occurs throughout active transcriptional units with a bias toward the middle and the 3′ end (Fig. 7) (Kind et al. 2008; Gelbart et al. 2009).

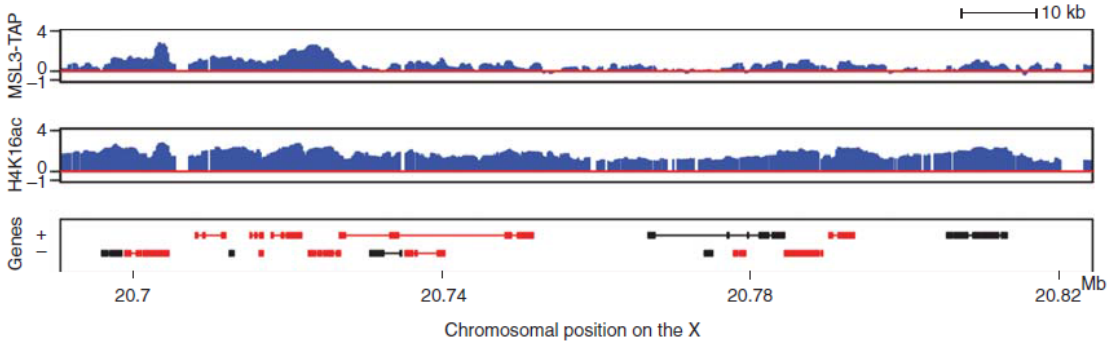


Figure 7. Correlation of H4K16 acetylation and MSL complex binding on the male X chromosomes. The distribution of H4K16ac on the male X chromosome is broader than MSL complex; active genes that lack stable MSL binding are nonetheless associated with H4K16ac. See Figure 5 for explanation of gene representation. (Adapted from Gelbart et al. 2009.)

In yeast, this particular covalent modification of histone H4 plays a key role in maintaining the boundary between silent and active chromatin; loss of function of Sas2, the HAT responsible for H4K16ac, allows the spreading of telomeric heterochromatin into adjacent subtelomeric chromatin (Suka et al. 2002). Structural studies have indicated that a key internucleosomal interaction may occur between an acidic patch of the histone H2A-H2B dimer on one nucleosome and a positively charged segment of the histone H4 tail (residues 16–26) extending from a neighbouring nucleosome (Luger et al. 1997; Schalch et al. 2005). When lysine 16 is acetylated, its positive charge becomes neutral suggesting that weakening a repressive internucleosomal structure could play a key role in dosage compensation. This contention was supported by the demonstration that reconstituted nucleosomal arrays acetylated at lysine 16 of histone H4 cannot achieve the level of salt-induced condensation of nonacetylated arrays (Shogren-Knaak et al. 2006; Robinson et al. 2008) and that this acetylation also weakens the self-association of reconstituted single nucleosome particles, reflecting the specific role of H4K16 in nucleosome–nucleosome stacking (Liu et al.2011). Using molecular force spectroscopy, the acetylation of H4K16 was observed to weaken nucleosome packing in reconstituted chromatin fibers and to result in a more disordered architecture (Dunlap et al. 2012). Whether, in vivo, it is the formation of the 30-nm fiber (intramolecular compaction) or the higher-order 100- to 400-nm fibers (intermolecular compaction) that are affected by the presence of H4K16ac is not known

(Shogren-Knaak et al. 2006). In either case, the presence of H4K16ac renders the chromatin of dosage compensated genes more accessible to factors or complexes. This is evidenced by the significantly greater accessibility of the compensated male X chromosome to an extrinsic DNA-binding protein (e.g., the bacterial DNA methyltransferase). The elevated accessibility
of this protein follows the distribution of H4K16ac along the X (Bell et al. 2010). Given the long-standing correlation between active chromatin and early DNA replication (Hiratani and Gilbert 2009; Schubeler et al. 2002), it is not surprising that another feature of the compensated X chromosome is that it initiates replication earlier in S phase than the rest of the genome (Lakhotia and Mukherjee 1970; Bell et al. 2010). Another finding is that the X chromosome is more susceptible to mechanical shearing than the autosomes in both males and females, indicating that it has a more open chromatin structure. In fact, the histone marks associated with active gene transcription (H3K4me2 and H3S10ph), as well as those specifically enriched on the dosage compensated X (H4K16ac) in males, are also slightly enriched on the X chromosomes of females. These findings suggest that the evolution of the distinct chromatin structure responsible for dosage compensation in males has affected the female X (Zhang and Oliver 2010).

In addition to the effect of H4K16 acetylation, there is mounting evidence that dosage compensation involves changes in the torsional stress of X-linked genes. Reducing the level of supercoiling factor, a protein known to associate at the 5′ end of active genes (Ogasawara et al. 2007), preferentially affects male viability because of a sex-specific decrease in the transcription of X-linked genes (Furuhashi et al. 2006). Compensated chromatin is topologically different from noncompensated chromatin. The difference requires the function of topoisomerase II, which associates with the MSL complex and is recruited to compensated
genes in excess of the amount present on autosomal genes with similar transcription levels (Cugusi et al. 2013). The rate of histone variant H3.3 incorporation into the X chromosome in male cells is enhanced in relation to the autosomes (Mito et al. 2005). This is to be expected, as the replication-independent nucleosome deposition of H3.3 occurs in transcriptionally active regions of chromatin and involves the replacement of histone H3with the variant H3.3 (see Henikoff and Smith 2014). However, contrary to the expectation that this increased level of H3.3 on the X may contribute to the mechanism of dosage compensation is the observation that the absence of the two genes that encode H3.3, although causing sterility in both sexes, has no effect on viability of mutant flies (Hodl and Basler 2009).

DOSAGE COMPENSATION AND NUCLEAR ORGANIZATION:

During interphase, chromosomes are seen to occupy individual territories rather than an intermingling of unravelled chromatin strands (Cremer and Cremer 2010; see Dekker and Misteli 2014). This organization is particularly evident in cells of Drosophila males in which the X chromosome can be identified by the presence of the MSL complex (Strukov et al. 2011). Within this compartment, the chromatin modifications that underlie the mechanism of dosage compensation appear to induce a particular higher order topography to the X chromosome. Throughout development, X-linked sites that are separated by approximately a dozen megabases are located much closer in male than in female nuclei (Fig. 10) (Grimaud and Becker 2009). This difference is dependent on the presence of the MSL1-MSL2 chromatin-recognition component of the complex and is not affected by the absence of the other three MSL proteins. Because, in the absence of any one of the latter, the partial complex that includes MSL1-MSL2 is found only at HASs, the proximity of X-linked loci in male cells must be mediated by their clustering (Grimaud and Becker 2009). It is interesting to note that the MSL proteins copurify with the nuclear pore complex proteins Nup153 and Megator (Mtor). Regions of the genome at the nuclear periphery that are proximal to the nuclear pore complex contain groups of active genes, suggesting this compartment may have a regulatory effect on transcription (Vaquerizas et al. 2010). Depletion of the Nup153 and Mtor nucleoporins leads to the loss of dosage compensation (Mendjan et al. 2006), although this analysis could be complicated by general viability issues (Grimaud and Becker 2009).

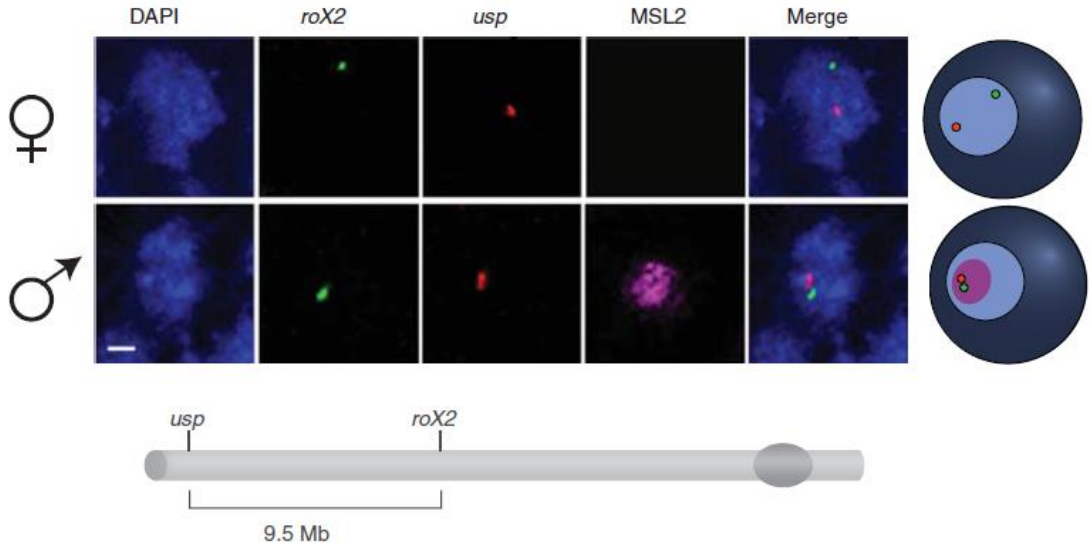


Figure 10. Male-specific conformation of the dosage-compensated X chromosome. A pair of high-affinity chromosomal sites (*roX2* and *usp*) were visualized by two-color FISH (fluorescence in situ hybridization) in female or male embryos. DNA was stained with DAPI (blue) and the X-chromosome territory (magenta) was painted with an antibody against MSL2 in male nuclei (there is no MSL2 in female nuclei). A merge of the channels reveals the proximity of the HASs and their residence relative to the MSL2 territory in male nuclei, clearly summarized in the cartoon on the right. The schematic diagram showing part of the X chromosome below indicates the distances separating the different HASs. (Modified, with permission, from Grimaud and Becker 2009, © Cold Spring Harbor Laboratory Press.)

Dosage Compensation in Human:

Dosage Compensation and Sex-Chromatin Bodies:

In man it has been found that Y-chromosomes are genetically inert in comparison to the X-chromosomes and other chromosomes and only a few genes are present in the human Y-chromosome. From the above discussion on the chromosome numbers of male and female human, it appears that females contain a higher dose of functional gene containing chromosome than males (Female chromosome numbers = 44 + XX and Male chromosome number = 44 + XY).

For many years, geneticists have observed that in some case, female homozygous for the genes in the X-chromosomes do not express a trait more markedly than do hemizygous males. So, it must be a mechanism of "dosage compensation", through which the effective dosage of genes of the two sexes is made equal or nearly so.

This mechanism of compensating the differential doses of functional sex chromosomes in male and female human is effected by the inactivation of one X-chromosome in the normal female. The genetically inactive X- chromosome or condensed X-chromosome is called

hetero-pychnotic X-chromosome or heterochromatin or sex-chromatin body or Barr body (according to the name of the geneticist M. L. Barr who first observed it) or Drum-stick (according to the shape of the inactive X-chromosome). Of the two X-chromosomes in females, which X-chromosome becomes inactive is a matter of chance, but it should be remembered that once an X- chromosome has become inactivated, all cells arising from that cell will keep the same inactive X-chromosome. In humans, inactive form of X-chromosome as a Barr-body have been observed by the sixteenth day of gestation. X-chromosome inactivation occurs in human when two or more X-chromosomes are present.

Details about Dosage Compensation or Lyon's Hypothesis:

The inactive X hypothesis or the Lyon's hypothesis or the Dosage Compensation is widely known from 1961 which states that only one of the two X chromosomes in the homogametic sex is functional while the other condenses and is inactivated. The X inactivated in some cells would be that from the father, in other cells it would be that from the mother.

Hence any tissue in the body of a woman would be a mosaic of cells which would show dominance of all genes having diffusible products but would remain a fine-grained mosaic for other intracellular differences.

Such a mosaic of cells might be difficult to demonstrate, particularly among rigid tissues, although cells which can be separated and cloned might show antigenic differences. This hypothesis has stimulated many new investigations, some of which are currently being completed.

Objectives behind the Proposition of Lyon's Hypothesis:

Lyon was impressed by three observations relating to X chromosome:

1. In normal mammalian females, one of the two X's is genetically inactive in the somatic cells (single active X-hypothesis).
2. Inactivation is random i.e., irrespective of paternal and maternal origin (random inactivation).
3. (a) The inactivation occurs during early ontogeny (early ontogenic differentiation) and (b) The particular X which has thus become inactivated, remains inactive in all the succeeding cell generation (fixed differentiation).

Evidences in Support of Lyon's Hypothesis:

A. For Single Active X Hypothesis:

1. Bengham (1958) and Russell (1961) noticed that an XO mouse is normal and fertile female indicating that the activity of the single X is sufficient for the normal development of this species.

2. McNeil (1956) recorded the case of "calico-cat" or tortoise-shell cat, usually a heterogygote female with black and yellow patches. Here the dominant X linked coat-colour gene producing yellow-colour becomes inactivated in some cells, whereas in others this mutant gene produces yellow-colour, thus causing a mosaic appearance. Exceptional male "calico" is XXY.:

3. X-linked ocular albinism in female heterogygotes causes the mosaic pigmentation of retina showing one active and another inactive X.

4. The late replicating nature of sex chromatin by H_3-TdR and very little or no RNA synthesized by Barr body in human body indicate the metabolic activity of one X chromosome.

5. The DNA replication pattern in mammalian females, for example:
Taylor (1960) — in Chinese haunter,
German (1962) — in human being
Mukherjee & Sinha (1966) — in cow etc. shows a late-replicating X (or inactivated) chromosome.

6. In individuals having XXXY or XXX polysomic conditions:
i) There are 2 late Xs.
ii) 2 sex chromatins and
iii) an apparent inactivity of G6PD genes in all but one X chromosome.

From above discussions it is clear that:

$$\begin{bmatrix}\text{No. of}\\ \text{Barr Body}\end{bmatrix} = \begin{bmatrix}\text{No. of}\\ \text{late X's}\end{bmatrix} = \begin{bmatrix}\text{No. of}\\ \text{Inactive X}\end{bmatrix} = \begin{bmatrix}\text{Total No. of}\\ \text{'X's—1}\\ \text{(in diploid)}\end{bmatrix}$$

For polyploid cells, Harnden gave a formula as :

1. $\text{No. of sex chromatin} = \text{No. of X} \frac{\text{Polyploidy}}{2}$
2. $(\# X - \frac{P}{2})$

B. For Random Inactivation:

1. Ohno and Catlanah (1962) examined that prophase skin cells of "variegated" mouse where dominant autosomal (eighth chromosome) coat colour gene had been translocated to an X chromosome as used by Russell. In light coloured area the sex chromatin size (obviously the translocated X) was larger than that of sex chromatin in darker patches.

2. Ohno (1963) worked on chinchilla mouse in which the dominant autosomal gene for chinchilla had been translocated onto an X chromosome. In the heterozygote female, the same size variation was observed.

3. Very distinct and ingenious first autoradiographic evidence for random inactivation was provided by Mukherjee and Sinha (1964) who had labelled in vitro the mule (hybrid of female horse + male donkey) leucocytes with H_3 thymidine at the terminal part of the "S"-period. In about 50% cells, the metacentric horse X was late replicating while submetacentric donkey — Xs were so in the remaining cells. Similar observations had been made in several hybrids such as "Zepony" (male zebra + female pony). "Gazel", "Nikosia" etc.

It has conclusively been noticed that any structurally abnormal X, e.g., ring-X, iso-X, deleted-X etc. is consistently late replicating and, as such, is heavily labelled in autoradiographic experiments. Random inactivation of either the maternal X or the paternal X chromosome seems to be the rule and the continued typical functioning of the single X-chromosome seems to take place early in development. Selection during embryonic growth may favour those cells which retains the normal X on active status. However, the early decision apparently is final for the inactivation of the one of the two X chromosomes.

C. For Early Embryonic Differentiation:

1. Graham (1954) while surveying the sex chromatin, a various species of mammalian females showed the absence of this structure in early embryos.

2. Despite its absence during early ontogeny, Barr bodies are noticed in late embryos of female cat.
3. Sex chromatins are deleted not earlier than five days in rabbit blastocyst and sixteen and nineteen days in human and macaque embryos, respectively.

4. Hills and Yunis (1966, 68) studied the golden hamster embryos from two-celled stage to about five to six days with H_3-TdR and found that late replicating X in female embryos is absent up to eight-celled stage and did identify the late X along the time of implantation at about the fourth day of gestation.

D. For Fixed Differentiation:

1. G6PD, which is much common in Negros and Mediterranean people but is virtually absent in North American whites in deficient activity, is quantitated by its ability for reducing methaemoglobin with the consequent destruction of glutathione.

2. The enzymatic activity is now very conveniently assessed by the action on the substrate and also the detection is made by starch-gel electrophoresis.

3. By "cloning" (aggregation of cells presumably originating from a single progenitor cell), the cells from females, heterozygotes for G6PD activity and employing mostly electrophoretic technique, two cell populations- one enzymatically active and other deficient— were detected.

Recruitment of the Dosage Compensation Complex to the X chromosome in *C. elegans:*

In *C. elegans*, the Dosage Compensation Complex (DCC) binds to and represses both X chromosomes in XX hermaphrodites by an average of two-fold. At the core of the DCC is a specialized condensin complex. Condensins are evolutionarily conserved five-subunit protein complexes that are essential for proper chromosome condensation and segregation. In metazoans, two types of condensin complexes (named I and II) share two Structural Maintenance of Chromosomes (SMC) protein subunits, and a set of three different non-SMC subunits. The condensin core of the DCC shares four out of five subunits with condensin I, but includes an SMC variant called DPY-27.

DPY-27 interacts with at least five other non-condensin proteins including SDC-1, SDC-2, SDC-3, DPY-30 and DPY-21. Sex-specificity of the DCC is provided by SDC-2 protein, which is expressed only in XX hermaphrodites during early embryogenesis. SDC-2, SDC-3 and DPY-30 are required for the recruitment of the condensin portion of the DCC to the X chromosome.

C. elegans DCC first binds to a number of recruitment sites on the X (*rex*), and then spreads onto the X chromosome. There are approximately 100 predicted *rex* sites along the length of the 17.5 Mb X-chromosome. Initially, *rex* sites were identified by assaying the ability of DNA fragments, in the form of multi-copy extrachromosomal arrays to recruit the DCC. ChIP-chip analysis of DCC identified additional recruitment sites, and defined a 10 bp DNA sequence motif that is enriched at the *rex* sites. This motif was later extended to 12-bp and named the motif enriched on the X (MEX). Although a 35 bp DNA fragment containing the motif as shown to recruit the DCC on extrachromosomal arrays, it is still unknown if the same fragment could recruit as a single copy insertion on an autosome. Nevertheless, the extrachromosomal recruitment assays show that MEX is important, because mutation of the motif resulted in loss of DCC recruitment. It is not known if any of the DCC subunits bind directly to MEX. Therefore, it remains unclear which proteins specify X-recruitment of the DCC via interaction with the MEX motif.

Mechanism of chromatin remodelling:

Nucleosomes inhibit transcription, DNA repair, and other chromosome transactions. SWI/SNF and RSC chromatin-remodelling complexes relieve this inhibition by sliding or disassembling nucleosomes. In the case of sliding, the structure of the nucleosome is unaltered at the end of the reaction. Studies to date have revealed an important principle of the remodelling process, DNA translocation driven by ATP hydrolysis, but the underlying mechanism remains obscure. Translocation slides the DNA around the histone octamer. Sliding exposes DNA at one end of an isolated nucleosome particle and in the linker regions between nucleosomes in an array. Exposure of the DNA may be detected by an increase in susceptibility to attack by nucleases.

Translocation is effected by the Sth1 subunit of RSC, a member of the DEAD/H-box family of helicase/translocases. These enzymes contact DNA through two domains and step along one strand by a scissors-like motion between them. A gap in one strand stops this stepping and blocks translocation. It could be inferred from the effect of gaps in nucleosomal DNA that Sth1 contacts one strand about two turns from the dyad of the nucleosome. A favored notion for remodelling is that the translocase draws DNA into the nucleosome from one side, creating a bulge, which is expelled on the other side. The alternative of DNA twisting

strain rather than a bulge traversing the nucleosome has been excluded on the basis of studies with nicked or gapped nucleosomal DNA.
An outstanding question for remodelling by DNA translocation is how the remodeler effects DNA sliding in the face of histone–DNA interaction. Translocation by Sth1 entails DNA sliding from the end of the nucleosome to a point near the dyad, requiring the disruption of all histone–DNA contacts along the way. Neither the mechanism nor the energetics of this process has been described.

Probable questions:

1. Discuss how dosage compensation was discovered in *Drosophila*?
2. How number of X chromosome regulate the dosage compensation in *Drosophila*?
3. How chromatin modification is associated with dosage compensation?
4. how dosage compensation is related with nuclear organization.
5. Discuss Lyon's hypothesis.
6. Discuss some evidences in support of Lyon's hypothesis.
7. How dosage compensation occurs in *C. elegans*?
8. Discuss the mechanism of chromatin remodelling.

Suggested Readings:

1. Molecular Cell Biology by Lodish, Fourth Edition.
2. The Cell – A Molecular Approach by Cooper and Hausman, Fourth Edition
3. Principles of Genetics by Snustad and Simmons, Sixth Edition.
4. Molecular Biology of the Cell – by Bruce Alberts
5. Cell and Molecular Biology by Gerald Karp, 7th Edition.
6. Gene cloning and DNA Analysis by T. A. Brown, Sixth Edition.
7. Genetics . Verma and Agarwal.
8. Brown TA. (2010) Gene Cloning and DNA Analysis. 6th edition. Blackwell Publishing, Oxford, U.K.

UNIT-VII

Epigenetics and genome imprinting - DNA methylation in mammals, genomic imprinting in mammals

Objective: In this unit we will discuss about genomic imprinting and DNA methylation in mammals.

Introduction:

The presence of genomic imprinting in mammals has considerable medical, societal, and intellectual implications in terms of (1) the clinical management of genetic traits and diseases, (2) the capacity to control human and animal breeding by assisted reproductive technologies, and (3) the progress of biotechnology and postgenomic medical research. Any modern day discussion of genetic problems, whether in research or medicine, must consider if a gene shows a biparental (i.e., diploid) mode of expression, or, is subject to genomic imprinting and shows parental-specific (i.e., haploid) expression. Despite the importance of genomic imprinting to human health and well-being, it is surprising that widespread acceptance of its existence and significance did not happen until the early nineties after three genes were unequivocally shown to display parental specific expression in mice.

Parental-specific behaviour of whole chromosomes had been observed in cytogenetic studies of chromosomes in Arthropods as early as the 1930s (Chandra and Nanjundiah 1990). Interestingly, the term “chromosome imprinting” was first coined to describe paternal-specific chromosome elimination that plays a role in sex determination in some Arthropod species (Crouse et al. 1971). Chromosomal imprinting of the mammalian X chromosome was also noted, which leads to paternal-specific inactivation of one of the two X chromosomes in all cells of female marsupials and the extraembryonic tissues of the mouse (Cooper et al. 1971). During the same period, classical geneticists were generating mouse mutants carrying chromosomal translocations that laid the foundation for the observation of imprinted gene expression. Some of these “translocation” mice, initially used to map the position of genes on chromosomes, showed a parental-specific phenotype when certain chromosomal regions were inherited as duplications of one parental chromosome in the absence of the other parental chromosome (known as uniparental disomy or UPD; Fig. 1). These results indicated the possibility “that haploid expression of particular maternal or paternal genes is important for normal mouse development” (Searle and Beechey 1978). At the same time, other geneticists used an unusual mouse mutant known as the “hairpin-tail” mouse that carried a large deletion of chromosome 17 to unequivocally set aside a basic tenet of genetics “that organisms heterozygous at a given locus are phenotypically identical irrespective of which gamete contributes which allele to the genotype” (Johnson 1974). Instead, offspring who received the Hairpin-tail deletion from a

maternal parent were increased in size and died midway through embryonic development, whereas paternal transmission of the genetically identical chromosome produced viable and fertile mice (Fig. 1). It is notable with hindsight that in spite of the previously published description of imprinted X-chromosome inactivation in mammals, the favoured interpretation of these genetic translocation and deletion experiments was not that the regions contained imprinted genes, but that genes on these autosomes primarily acted in the haploid egg or sperm to modify proteins used later in embryonic development. Despite this, the concept of differential functioning of the maternal and paternal genome was gaining ground and a suggestion made that "the maternal genome might be normally active at the Hairpin-tail chromosomal region while its paternal counterpart is preferentially inactivated" (McLaren 1979).

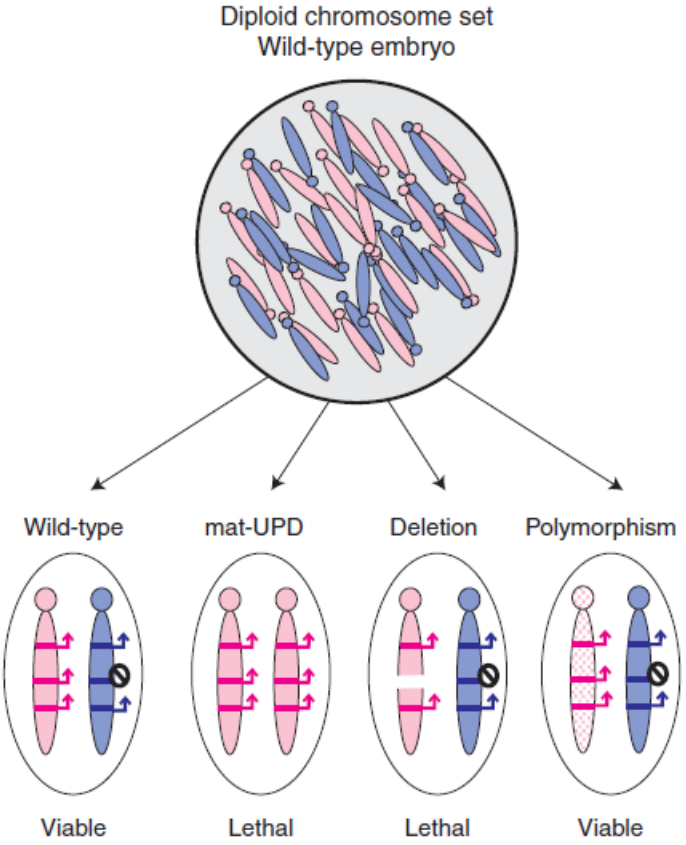


Figure 1. Mouse models to study genomic imprinting that allow the maternal and paternal chromosome to be distinguished. Mammals are diploid and inherit a complete chromosome set from the maternal and paternal parent. However, mice can be generated that (1) inherit two copies of a chromosome pair from one parent and no copy from the other parent (known as UPD), (2) inherit a partial chromosomal deletion from one parent and a wild-type chromosome from the other parent, and (3) inherit chromosomes carrying single-nucleotide polymorphisms (known as SNPs) from one parent and a wild-type chromosome from the other parent. Offspring with UPDs or deletions are likely to display lethal phenotypes, whereas SNPs will allow the production of viable offspring.

A major step forward in establishing the existence of genomic imprinting in mammals came several years later with the development of an improved nuclear transfer technology being used to test the possibility of generating diploid uniparental embryos solely from mouse egg nuclei.
The nuclear transfer technique took a donor male or female pronucleus from a newly fertilized egg and used a fine micropipette to place it inside a host fertilized egg from which either the maternal or paternal pronucleus had been removed. This regenerated diploid embryos, but with two maternal or two paternal genomes (known, respectively, as gynogenetic and androgenetic embryos; Fig. 2). The technique was first used to show that nuclei from fertilized Hairpin-tail mutant embryos could not be rescued when transferred into a wild-type host egg. This provided proof that the embryonic genome, and not the oocyte cytoplasm, carried the Hairpin-tail defect. It also confirmed the suggestion that genes on the maternal and paternal copy of chromosome 17 functioned differently during embryonic development (McGrath and Solter 1984b). Subsequently, nuclear transfer was used to show that embryos, reconstructed from two maternal pronuclei (known as gynogenetic embryos) or two paternal pronuclei (androgenetic embryos), failed to survive; whereas only embryos reconstructed from one maternal and one paternal pronucleus produced viable and fertile offspring (McGrath and Solter 1984a; Surani et al. 1984). This work overturned a previous claim that uniparental mice could develop to adulthood (Hoppe and Illmensee 1982). Gynogenetic embryos at the time of death were defective in extraembryonic tissues that contribute to the placenta, whereas androgenetic embryos were defective in embryonic tissue. These outcomes led to the hypothesis that embryonic development required imprinted genes expressed from the maternal genome, whereas the paternal genome expressed imprinted genes required for extraembryonic development (Barton et al. 1984). Subsequent identification of imprinted genes in the mouse did not confirm a bias in the function of imprinted genes, but indicated that the observed differences between gynogenetic and androgenetic embryos may be explained by a dominant effect of one or a few imprinted genes. The nuclear transfer experiments, combined with supporting data from mouse genetics, provided convincing evidence that both parental genomes were required for embryogenesis in mice, laying a strong foundation for the existence of genomic imprinting in mammals (Fig. 2). An extensive survey of parental chromosome contribution to embryonic development, using “translocation” mice to create UPD chromosomes (Fig. 1), identified two regions on mouse chromosomes 2 and 11 that showed opposite phenotypes when present either as two maternal or two paternal copies.

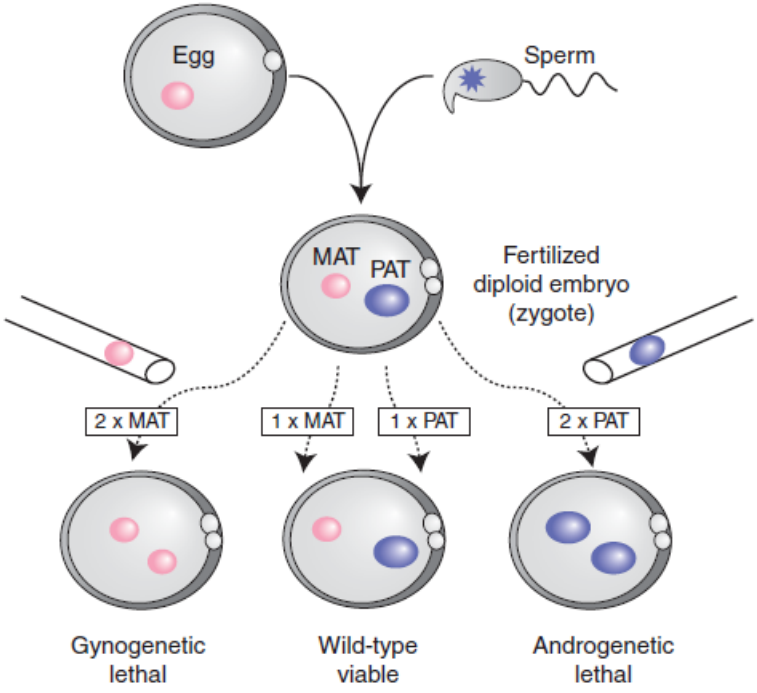


Figure 2. A maternal and paternal genome are needed for mammalian reproduction. The nuclear transfer technique used micropipettes and high-powered microscopes to remove the male or female nuclei from a newly fertilized egg and place them in various combinations into a second "host" fertilized egg that had already been enucleated, thereby generating anew diploid embryos with two maternal (gynogenetic) or two paternal (androgenetic) genomes or a biparental genome (wild-type). Gynogenetic and androgenetic embryos were lethal at early embryonic stages. Only reconstituted embryos that received both a maternal and paternal nucleus (wild-type) survived to produce living young. These experiments show the necessity for both the maternal and paternal genome in mammalian reproduction, and indicate the two parental genomes express different sets of genes needed for complete embryonic development.

This further strengthened the argument for parental- specific gene expression in mammals (Cattanach and Kirk 1985). In addition, human data strongly indicated that some genetic conditions, most notably the Prader– Willi syndrome, which appears to arise exclusively by paternal transmission, could best be explained by parental-specific gene expression (Reik 1989). Further clues came from experiments applying the newly developed technology for making transgenic mice by microinjecting gene sequences into a fertilized mouse egg. This was often beset by the problem of DNA methylation unexpectedly inducing silencing of the transgene in somatic tissues. Some transgenes even showed parental-specific differences in their ability to acquire DNA methylation, adding weight to the argument that parental chromosomes behave differently. This normally followed the pattern that maternally transmitted transgenes were

methylated whereas paternally transmitted transgenes were not. However, only in a few cases did DNA methylation differences correlate with parental-specific expression. Although many similarities were later found between “transgene” methylation imprinting and genomic imprinting of endogenous mouse genes, several features distinguish them (Reik et al. 1990). This includes a high susceptibility to strain-specific background effects, an inability to maintain imprinted expression at different chromosomal integration sites, and a requirement for foreign DNA sequences to produce the imprinted effect (Chaillet
et al. 1995).

Despite the wealth of supportive data, final proof of the existence of genomic imprinting in mammals depended on the identification of genes showing imprinted parental specific expression. This occurred in 1991 when three imprinted mouse genes were described. The first of these, Igf2r (insulin-like growth factor type 2 receptor that is a “scavenger” receptor for the growth hormone insulin-like growth factor type 2 [Igf2]) was identified as a maternally expressed imprinted gene. This gene was later shown to explain the overgrowth phenotype of the Hairpin-tail mutant mouse (Barlow et al. 1991). A few months later, the Igf2 gene was identified as a paternally expressed imprinted gene (DeChiara et al. 1991; Ferguson-Smith et al. 1991). Finally, the H19 gene (cDNA clone number 19 isolated from a foetal hepatic library), an unusual long noncoding RNA (lncRNA) was subsequently shown to be a maternally expressed imprinted gene (Bartolomei et al. 1991). Diverse strategies were used to identify these three imprinted genes, each of which depended on emerging technologies in mouse genetics. For Igf2r, positional cloning was used to identify genes that mapped to the Hairpin-tail deletion on chromosome 17. Mice then inheriting the deletion from one parent were used to identify those genes showing maternal-specific expression (Fig. 1). For Igf2, the physiological role of this growth factor in embryonic development was being tested by gene knockout technology.

Surprisingly, mice carrying the mutant non-functional allele showed a phenotype following paternal transmission, but no phenotype on maternal transmission. The H19 lncRNA was identified as an imprinted gene after this gene was mapped close to the Igf2 locus on chromosome 7, proving the hypothesis that imprinted genes could be clustered together. Although these strategies were to prove useful in subsequent attempts to identify imprinted genes, the demonstration that imprinted genes were closely clustered has proven to be a pivotal discovery in understanding the mechanism controlling genomic imprinting in mammals.

Genomic Imprinting—An Epigenetic Gene Regulatory System:

The defining characteristic of genomic imprinting is that it is cis acting. Thus, the imprinting mechanism acts only on one chromosome. The two parental chromosomes will normally contain many single base pair differences (known as single-nucleotide polymorphisms [SNPs]) if the population is outbred, but they can be genetically identical if inbred mouse strains are used. Because genomic imprinting occurs in inbred mice that have genetically identical parental chromosomes, it was concluded that the process must use an epigenetic mechanism to modify

the information carried by the DNA sequence, yet create an expression difference between the two parental gene copies. These observations also indicate that a cis-acting silencing mechanism, which is restricted to one chromosome, is operating so that the silencing factors cannot freely diffuse through the nucleus to reach the active gene copy. Although imprinted genes are repressed on one parental chromosome relative to the other, genomic imprinting is not necessarily a silencing mechanism and has the potential to operate at any level of gene regulation (i.e., at the promoter, enhancers, splicing junctions, or polyadenylation sites) to induce parental-specific differences in expression. Genomic imprinting must therefore depend on an epigenetic system that modifies or "imprints" one of the two parental chromosomes (Fig. 3). This imprint is subsequently used to attract or repel transcriptional factors or mRNA processing factors, thereby changing expression of the imprinted gene on one parental chromosome. Because inbred mice with genetically identical chromosomes also show genomic imprinting, parental imprints are not likely to be acquired after the embryo becomes diploid because there would be no way for the cells' epigenetic machinery to distinguish between identical parental gene copies. Thus, parental imprints must be acquired when the two parental chromosome sets are separate and this only occurs during gamete formation and for 12h postfertilization (Fig-3).

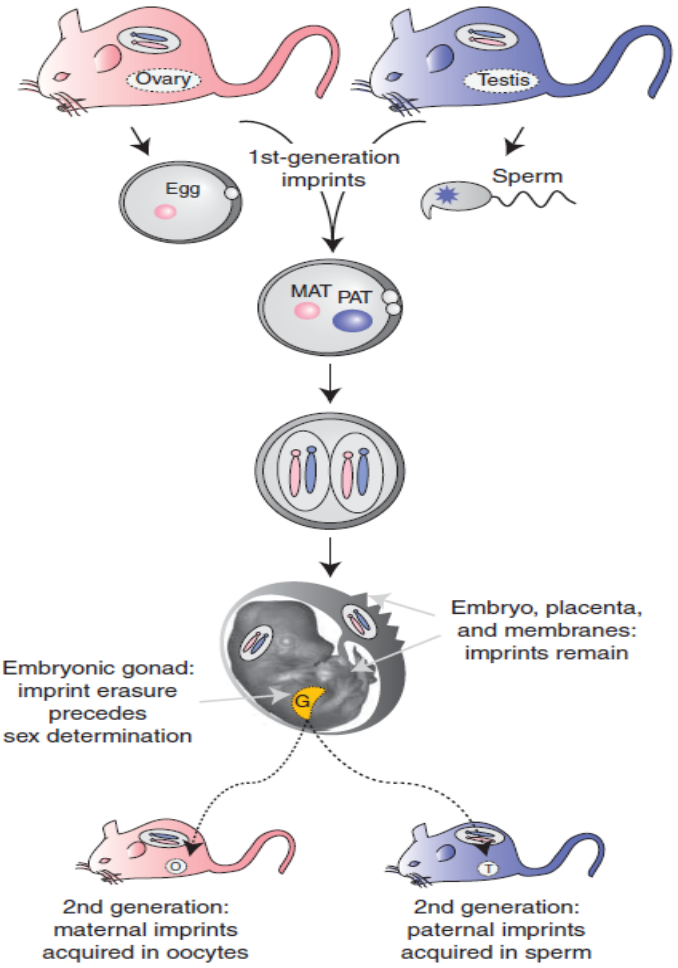


Figure 3. Imprint acquisition and erasure in mammalian development. Imprints are acquired by the gametes; thus, oocytes and sperm already carry imprinted chromosomes (first-generation imprints). After fertilization when the embryo is diploid, the imprint is maintained on the same parental chromosome after each cell division in cells of the embryo, yolk sac, placenta, and also in the adult. The germ cells are formed in the embryonic gonad and the imprints are erased only in these cells before sex determination. As the embryo develops into a male, the gonads differentiate to testes that produce haploid sperm that acquire a paternal imprint on their chromosomes. Similarly, in developing females, chromosomes in the ovaries acquire maternal imprints (second-generation imprints).

The most likely scenario is that gametic imprints are placed on paternally imprinted genes during sperm production and on maternally imprinted genes during egg formation.

A key feature of the "imprinted" DNA sequence is that it would only be modified in one of the two parental gametes; thus, two types of recognition system are required, one sperm-specific and one oocyte-specific, each directed toward a different DNA sequence. Several other features are required of the imprint. First, once established, it must remain on the same parental chromosome after fertilization when the embryo is diploid. Second, the imprint must be stably inherited through mitosis of the embryo and adult animal. Last, it must be erasable. The latter is necessary

because the embryo will follow either a male or female developmental path midway through development and its gonads will need to produce only one type of imprinted haploid parental gamete. Thus, germ cells that have arisen from embryonic diploid cells (Fig. 3) must first lose their inherited maternal and paternal imprints before they gain that of the gamete.

An imprint can be defined as the epigenetic modification that distinguishes the two parental copies of a given gene. Once formed, the imprint must also allow the transcription machinery to treat the maternal and paternal gene copy differently within the same nucleus. A gametic imprint is predicted to be continuously present at all developmental stages (Fig. 3), thus imprints can be found by comparing epigenetic modifications on maternal and paternal chromosomes in embryonic or adult tissues (using strategies outlined in Fig. 1) and tracing them back in development to one of the two gametes. Candidates for gametic imprints could be modifications of DNA or histone proteins that package DNA into chromosomes (Allis et al. 2014). There are now two types of epigenetic DNA modification known in mammals; 5- methylcytosine and 5-hydroxymethylcytosine (Li and Zhang 2014). Histones can bear multiple types of modification including methylation, acetylation, phosphorylation, sumoylation, and ubiquitylation (Allis et al. 2014). They can also be replaced by variant histones with specific functions (Henikoff and Smith 2014). Any of these epigenetic modifications could qualify as an imprint. One would predict that enzymes responsible for these epigenetic modifications or an essential cofactor would be exclusively expressed in one of the two gametes, and specifically associate with one parental chromosome to copy the modification when the cell divides. However, as will be described in Section 3 on "key discoveries," only 5-methylcytosine has been clearly shown to function as the gametic imprint for imprinted genes in mammals and, to date, is the only known heritable modification.

How does a gametic imprint control imprinted expression?

To understand how the imprint operates, three pieces of information are required: which parental chromosome carries the imprint, which parental chromosome carries the expressed allele of the imprinted gene, and the position of the imprinted sequence relative to the expressed or silenced allele of the imprinted gene. Using this type of approach it has been shown that gametic imprints can act on whole clusters of genes at once. These imprinted clusters contain 3–12 imprinted genes and span from 100–3700 kb of genomic DNA. The majority of genes in any one cluster are imprinted protein-coding mRNA genes; however, at least one is always an imprinted lncRNA.

Because of the arrangement of imprinted genes in clusters, with some genes expressed from one parental chromosome and some from the other, it is not trivial to determine how the imprint operates. It is possible to study the effect of the imprint on single genes in the cluster, but it may prove more informative to study the effects of the imprint on the entire cluster. One thing, however, is clear. Nature has not chosen the simplest mechanism whereby the imprint is directed toward a promoter to pre-emptively silence an imprinted gene in one gamete. Instead, imprints

appear, in general, to be directed toward long-range cis-acting regulators that influence the expression of multiple genes, and are located a long distance away on the same chromosome.

Imprinted Genes Control Embryonic and Neonatal Growth:

What is the function of genomic imprinting in mammals?

One way to answer this question is to determine the function of known imprinted genes in vivo. This can be performed by mutating the gene sequence to impair its function using the "homologous recombination" technique. The function of many of the known imprinted genes has been determined in this fashion. The most significantly represented function among imprinted genes includes genes that affect growth of the embryo, placenta, and neonate. In this category are paternally expressed imprinted genes that function as growth promoters (i.e.,
Igf2, Peg1, Peg3, Rasgrf1, Dlk1) and show growth retardation in embryos deficient for the gene. There are also maternally expressed imprinted genes that function as growth repressors (i.e., Igf2r, Gnas, Cdkn1c, H19, Grb10), as shown by a growth enhancement in embryos deficient for the gene. Another significant category includes genes with behavioural or neurological defects (e.g., Nesp, Ube3a, Kcnq1). These results are, at one level, disappointing because they do not identify one function for all imprinted genes. Nevertheless, these results show that the majority of imprinted genes function as embryonic or neonatal growth regulators.
More interestingly, the ability to regulate growth appears to be neatly divided with maternally expressed growth regulating genes acting to repress growth of the offspring, whereas paternally expressed genes in this category act to increase growth. Moreover, numerous tested imprinted genes are active in neurological processes, some of which affect neonatal growth rate by altering maternal behaviour.

The Function of Genomic Imprinting in Mammals:

Assessment of genomic imprinting in different types of mammals has been informative. Placental mammals such as mice and humans, and marsupials such as opossum and wallaby, have genomic imprinting. Egg-laying mammals, such as platypus and echidna, appear to lack imprinted genes, although extensive studies have not yet been performed (Renfree et al. 2009). Placental mammals and marsupials are distinguished from egg-laying mammals by a reproductive strategy that allows the embryo to directly influence the amount of maternal resources used for its own growth. In contrast, embryos that develop within eggs are unable to directly influence maternal resources. Most invertebrates and vertebrates use an egg-laying reproductive strategy. Notably, they can also undergo parthenogenesis—a form of reproduction in which the female gamete develops into a new diploid individual without fertilization by a male gamete (note that parthenogenetic embryos arise from the duplication of the same maternal genome, whereas the gynogenetic embryos described in Fig. 2 arise from two different maternal genomes). The ability of organisms to undergo parthenogenesis most likely indicates a complete

absence of genomic imprinting as it shows the paternal genome is dispensable. In mammals, however, a direct consequence of imprinted gene expression controlling foetal growth is that parthenogenesis is not possible. Both parents are necessary to produce viable offspring making mammals completely reliant on sexual reproduction to reproduce (Fig. 4). Parthenogenesis has thus not yet been observed in mammals despite claims to the contrary, although manipulating expression of the Igf2 and Dlk1 imprinted clusters has generated some rare mice with a diploid maternal genome (Kawahara et al. 2007).

Why should genomic imprinting have evolved only in some mammals, but not in vertebrates in general? Three features of genomic imprinting—the growth regulatory function of many imprinted genes, the restriction of imprinted genes to placental and marsupial mammals, and last, the necessity of the paternal genome for foetal development, provide evidence that can fit two equally attractive hypotheses. The first hypothesis proposes that genomic imprinting evolved in response to a "parental conflict" situation. This arises from the opposing interests of the maternal and paternal genome: Embryonic growth is dependent on one parent, but influenced by an embryo whose genome comes from two parents. Paternally expressed imprinted genes are proposed to increase embryonic growth, thereby maximizing the fitness of an individual offspring bearing a particular paternal genome. Maternally expressed imprinted genes are proposed to suppress foetal growth. This would allow a more equal distribution of maternal resources to all offspring and increase transmission of the maternal genome to multiple offspring, which may have different paternal genomes. The second hypothesis is named "trophoblast defence" (Varmuza and Mann 1994). This proposes that the maternal genome is at risk from the consequences of being anatomically equipped for internal reproduction should spontaneous oocyte activation lead to full embryonic development. Because males lack the necessary anatomical equipment for internal reproduction, they do not share the same risks should spontaneous activation of spermatozoa occur. Imprinting is thus proposed to either silence genes on the maternal chromosome that promote placental development or to activate genes that limit this process. The genes necessary for placental invasion of the maternal uterine vasculature would consequently only be expressed from a paternal genome after fertilization has occurred.

Which, if any, of these hypotheses explains the evolution of genomic imprinting in mammals? Both hypotheses indicate a role for imprinted genes in regulating the development and function of the placenta, however, neither the parental conflict nor the trophoblast defence models can provide a full explanation for all the data (Wilkins and Haig 2003). It is interesting to note that imprinted genes have also been identified in the plant endosperm, a tissue that has been compared to the placenta by virtue that it transfers nutrient resources from the parent plant to the embryo (Grossniklaus and Paro 2014). This finding strengthens arguments that genomic imprinting evolved as a means to regulate nutrient transfer between the parent and offspring, but it does not tell us why. Fuller or alternative explanations of the function of genomic imprinting in mammals could come from two sources. The first would be to examine the function of "imprinting" across a complete gene cluster in contrast to examining the phenotype of mice lacking a single imprinted gene product. This would require an ability to reverse an

imprint and generate biparental gene expression across the whole imprinted cluster. The second approach is to learn exactly how genes are imprinted. It is possible that not all genes in a cluster are deliberate targets of the imprinting mechanism and that some may just be "innocent bystanders" of the process, and their function would not be informative about the role of genomic imprinting. The existence of innocent bystander genes affected by the imprinting mechanism may satisfactorily explain the curious abundance of imprinted genes with no obvious biological function in development.

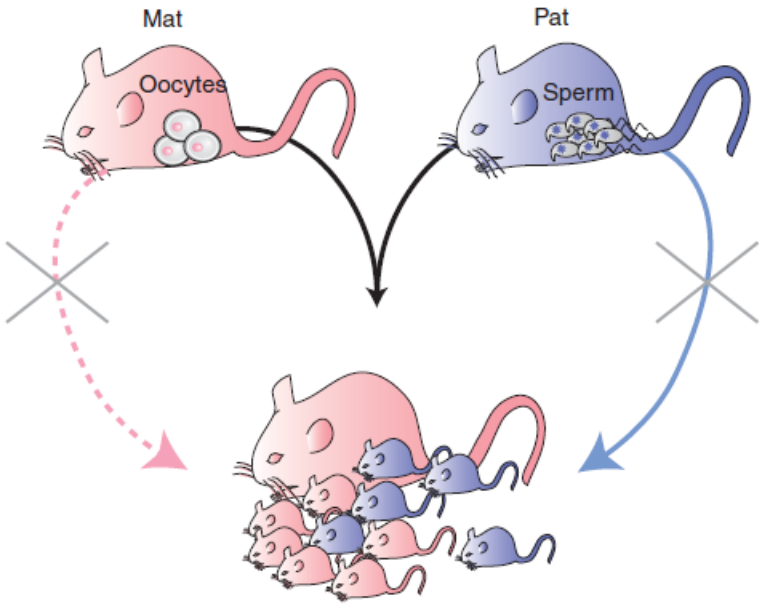


Figure 4. Imprinted genes play a role in mammalian reproduction. Mammals are diploid and reproduction requires fertilization of a haploid female egg by a haploid male sperm to recreate a diploid embryo. Only females are anatomically equipped for reproduction, but they cannot use parthenogenesis to reproduce (the possibility of which is represented by a pink dashed line) because essential imprinted genes needed for fetal growth are imprinted and silenced on maternal chromosomes. These genes are expressed only from paternal chromosomes; thus, both parental genomes are needed for reproduction in mammals. Parthenogenesis is the production of diploid offspring from two copies of the same maternal genome.

Imprinted Genes Are Clustered and Controlled by Imprint Control Elements (ICEs):

To date, about 150 imprinted genes have been mapped to 17 mouse chromosomes including the X chromosome. More than 80% of the identified imprinted genes are clustered into 16 genomic regions that contain two or more genes (Wan and Bartolomei 2008). The discovery of clusters of imprinted genes was a strong indication that a common DNA element may regulate imprinted expression of multiple genes in cis. To date, seven of the 16 imprinted clusters have been well

characterized, and these are listed in Table 1 by the name of the principle imprinted mRNA gene in the cluster or after a disease association (e.g., the Pws cluster for Prader–Willi syndrome; Beaudet and Zoghbi 2014). These seven clusters contain three to 12 (or more) imprinted genes and are spread over 80–3700 kb of DNA. A common feature of these seven clusters is the presence of a DNA sequence carrying a gametic methylation imprint that is known as a gametic DMR (differentially DNA-methylated region). A gametic DNA methylation imprint is defined as a methylation imprint established in one gamete and maintained only on one parental chromosome in diploid cells of the embryo. In five clusters (Igf2r, Kcnq1, Gnas, Grb10, and Pws), the gametic DMR has a maternal methylation imprint acquired in oogenesis, whereas in two clusters (Igf2 and Dlk1), it has a paternal methylation imprint acquired during spermatogenesis. In these examples, the gametic DMR controls imprinted expression of the whole or part of the cluster and is therefore designated as the imprint control element, or ICE, for the cluster (Barlow 2011).

Table 1. Features of imprinted gene clusters in the mouse genome

Cluster name	Chromosome mouse/human	ICE (gametic methylation imprint)	Cluster size (kb)	Gene number in cluster	Parental expression M/P	lncRNA and expression (M or P)
Igf2r	17/6	Region 2 (M)	490	4	3 M (pc) 1 P (nc)	*Airn* (P)
Kcnq1	7/11	*KvDMR1* (M)	780	12	11 M (pc) 1 P (nc)	*Kcnq1ot1* (P)
Pws	7/15	*Snrpn-CGI* (M)	3700	>8	2 M (pc)/ >7 P (nc and pc)	*Ube3aas* (P)[a] *Ipw* (P)[a] *Zfp127as* (P)[a] *PEC2* (P)[a] *PEC3* (P)[a] *Pwcr1* (P)[a]
Gnas	2/20	*Nespas* DMR (M)	80	7	2 M (pc) 5 P (4 nc and 1 pc)	*Nespas* (P)[b] *Exon1A* (P) *miR-296* (P)[b] *miR-298* (P)[b]
Grb10	11/7	Meg1/Grb10 DMR (M)	780	4	2 M (pc)/ 2 P (pc)	NI
Igf2	7/11	H19-DMD (P)	80	3	1 M (nc)/ 2 P (pc)	*H19* (M)
Dlk1	9/14	IG-DMR (P)	830	>5	>1 M (nc)/ 4 P (pc)	*Gtl2* (M)[c] *Rian* (M)[c] *Rtl1as* (M)[c] *Mirg* (M)[c] miRNAs (M)[c] snoRNAs (M)[c]

Note that cluster size and number of genes in the cluster are provisional and await a genome-wide analysis of imprinted expression. Pws and Dlk1 clusters contain overlapping transcripts in which the number of distinct genes is not yet clear. Details are given in the text.

M, maternal; P, paternal; DMR, differentially methylated region; pc, protein coding; nc, noncoding RNA; NI, none identified; miRNA, micro RNA; snoRNA, small nucleolar RNAs.

[a] May be one long lncRNA.

[b] Part of *Nespas* transcript.

[c] May be part of one or multiple lncRNAs.

Table 1 shows that each imprinted gene cluster contains multiple mRNAs and, with the exception of Grb10, at least one lncRNA. Two trends emerge. First, the imprinted protein-coding genes in each cluster are expressed, for the most part, from the same parental

chromosome, whereas the lncRNA is expressed from the opposite parental chromosome (as illustrated in Fig. 5 for a maternal gametic DMR).

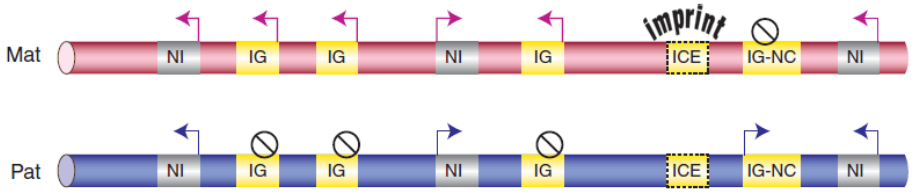


Figure 5. Imprinted genes are expressed from one parental allele and often clustered. Most imprinted genes (yellow) are found in clusters that include multiple protein-coding mRNAs (IG) and at least one noncoding RNA (IG-NC). Nonimprinted genes can also be present (NI in gray). The imprinting mechanism is *cis* acting and imprinted expression is controlled by an imprint control element (ICE) that carries an epigenetic imprint inherited from one parental gamete. One pair of diploid chromosomes is shown: the pink is of maternal origin and the blue of paternal origin. Arrow, expressed gene; slashed circle, repressed gene.

Second, the ICE deletion causes loss of imprinted expression only when deleted from the parental allele expressing the lncRNA. Table 1 shows that in three clusters (Igf2r, Kcnq1, and Gnas) the lncRNA promoter sits in an intron of one of the imprinted mRNAs, whereas in the remaining clusters the lncRNA promoter is separated, but lies close to the imprinted mRNA genes. This close intermingling of active and silent genes in an imprinted cluster indicates that the silencing and activating mechanisms affecting imprinted genes do not spread and may be restricted to the affected gene. In particular, the fact that the promoter of a silent lncRNA can reside in the intron of an actively transcribed gene indicates that silencing mechanisms may not even spread throughout the length of a gene, but may just be restricted to regulatory elements.

What is the role of the gametic DMR?

Despite the fact the gametic DMRs can be maternally or paternally methylated, experiments that deleted these elements have produced broadly similar results albeit with a few interesting exceptions (Fig. 6). For three clusters (Igf2r, Kcnq1, Dlk1), experimental deletion of the methylated gametic DMR produced no effect. In contrast, deletion of the unmethylated gametic DMR eliminated parental-specific expression causing a loss of lncRNA expression in cis and biallelic mRNA expression (Lin et al. 1995; Zwart et al. 2001; Fitzpatrick et al. 2002). Two clusters (Gnas and Pws) appear to contain more than one gametic DMR and show a more complex behaviour, yet they still share some similarities with the pattern presented in Figure 6 (Williamson et al. 2006). The Igf2 cluster, however, behaves differently: deletion of both the methylated and unmethylated gametic DMR causes changes in mRNA and lncRNA expression in cis (Thorvaldsen et al. 1998).

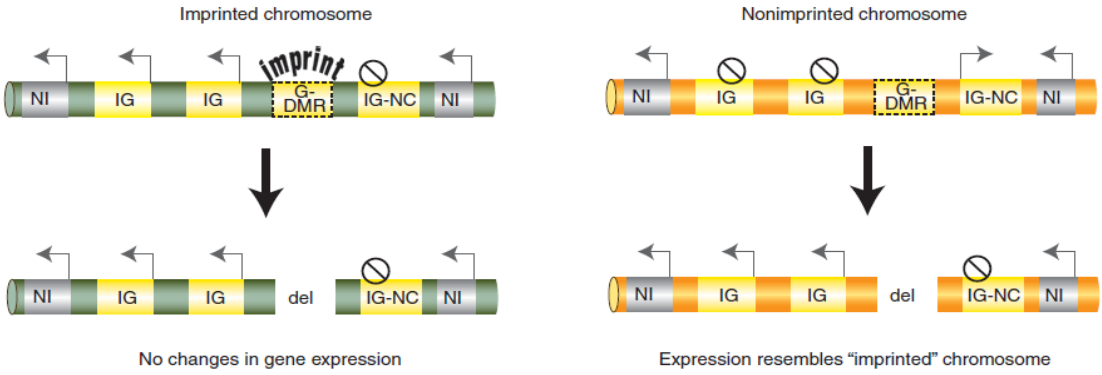


Figure 6. Imprinted expression is regulated by gametic DMRs (G-DMR). (*Left*) The effect of deleting the gametic DMR from the imprinted chromosome (green). (*Right*) The effect of deleting the G-DMR from the nonimprinted chromosome (yellow). In many imprinted clusters (e.g., *Igf2r*, *Kcnq1*, and *Dlk1*), experimental deletion of the G-DMR only affects the chromosome carrying the nonimprinted G-DMR. This results in a loss of repression of the imprinted protein-coding mRNA genes (IG) and a gain of repression of the imprinted lncRNA gene (IG-NC). Note that in some imprinted clusters (*Igf2* and *Pws*) that are not illustrated here, the methylated G-DMR appears also to be required for expression of some of the imprinted mRNAs in *cis*. del, deleted DNA; G-DMR, gametic differentially DNA-methylated region; NG, nonimprinted gene; arrow, expressed allele; slashed circle, repressed allele; imprint, epigenetic modification leading to a change in gene expression in *cis*.

The results from the above gametic DMR deletion experiments do not at first glance indicate a common function for gametic DMRs. However, an understanding of their exact function depends on knowing the position of the DMR with respect to the imprinted genes in each cluster. In the three clusters with the simplest pattern (Igf2r, Kcnq1, and Dlk1), the gametic DMR either contains or controls expression of the lncRNA, thus deletion of this element will clearly lead to loss of lncRNA expression. The gametic DMR in the Igf2 cluster, however, does not directly promote H19 transcription, but changes the interaction between Igf2 and H19 and their shared enhancers, and in this way regulates their expression. Despite these differences, in general, the unmethylated gametic DMR is implicated in all six clusters as a positive regulator of lncRNA expression, and the presence of the DNA methylation imprint is associated with repression of the lncRNA. The conclusion from the data obtained from deletion of gametic DMRs clearly identifies these regions as an ICE, whose activity is regulated by DNA methylation.

The Role of DNA Methylation in Genomic Imprinting:

The identification of the first three endogenous imprinted genes in 1991 enabled investigators to study how the cell's epigenetic machinery marked an imprinted gene with its parental identity. The first and most easily testable candidate was DNA methylation, a modification in mammals that covalently adds a methyl group to the cytosine residue in CpG dinucleotides. DNA methylation is acquired through the action of de novo methyltransferases and maintained in situ each time the cell divides by the action of maintenance methyltransferases (see Li and Zhang 2014).Hence, this modification fulfils the criteria outlined in Fig. 3 for a parental identity mark or "imprint" because (1) it can be established in either the sperm or oocyte by de novo methyltransferases that act only in one gamete, (2) it can be stably propagated at each embryonic cell division by a maintenance methyltransferase, and (3) it can be erased in the germline to reset the imprint in the next generation, either by passive demethylation (DNA replication followed by

the failure to undergo maintenance methylation) or through the action of a demethylating activity (possibly through conversion of 5-methylcytosine to 5-hydroxymethylcytosine by the ten-eleven translocation family of enzymes or through excision of 5-methylcytosine by the DNA repair machinery. DNA methylation could potentially perform two different functions in genomic imprinting. It could act as the imprinting mark by being acquired de novo only by the chromosomes in one gamete. It could also serve to silence one of the parental alleles because DNA methylation is associated with gene repression (Li and Zhang 2014). To determine which function it has, it is first necessary to show that DNA methylation is present only on one parental chromosome (i.e., that it is a DMR). Second, it is necessary to identify which imprinted gene in the cluster and which regulatory sequences are marked by DNA methylation.

The location of methylation marks on a promoter, or on distant positive or negative regulatory elements will have different consequences for gene expression. Finally, it is necessary to identify when the DMR forms during development. If it forms during gametogenesis and is continuously maintained in place in somatic cells (known as a gametic DMR), it may serve as the imprinting mark. If, however, it is placed on the gene after the embryo has become diploid when both parental chromosomes are in the same cell (known as a somatic DMR), it is unlikely to serve as the identity mark, but may serve to maintain parental specific silencing.

Parental allele-specific DNA methylation has been found at most imprinted clusters that have been examined. For example, the Igf2 cluster has a gametic DMR located 2 kb upstream of the H19 lncRNA promoter that is methylated only in the paternal gamete and is maintained thereafter in all somatic tissues (Bartolomei et al. 1993).A similar gametic DMR was identified covering the promoter of the Airn lncRNA, present only on the silent maternal gene copy, and acquired in the female gamete (Stoger et al. 1993). Surprisingly, gametic DMRs were not identified at the promoters of the principal imprinted protein-coding genes in these clusters (respectively, Igf2 and Igf2r). Instead, the silenced Igf2 promoter is free of DNA methylation, whereas the silenced Igf2r promoter lies within a somatic DMR that is placed after fertilization (Sasaki et al. 1992; Stoger et al. 1993). Similar findings of gametic DMRs methylated on the chromosome carrying the silent copy of the imprinted lncRNA (Fig. 6) have been made for other well-studied imprinted gene clusters, including Pws, Kcnq1, Gnas, Dlk1, and Grb10. Somatic DMRs are relatively rare but have been reported for some imprinted clusters, which suggests that this type of epigenetic modification plays a limited role in maintaining imprinted gene expression. Deletions of gametic DMRs in mice result in complete loss of imprinting for multiple genes, thereby proving that this class of DMRs also serves as a major ICE for the whole cluster (Fig. 6). In contrast, deletion of the somatic DMRs affects expression of the adjacent imprinted gene, but imprinted expression is maintained by other genes in the cluster (Constancia et al. 2000; Sleutels et al. 2003).

Genome-wide deficiency in DNA methylation caused by mutations in the Dnmt gene family underscores the essential role of DNA methylation in regulating imprinted gene expression. Mutations in the de novo DNA methyltransferase Dnmt3a, the DNA methyltransferase stimulatory factor Dmnt3L, or the Dnmt1 maintenance DNA methyltransferase generate DNA

methylation deficient embryos that show alterations in imprinted gene expression. The type of perturbations shown for four imprinted clusters (Igf2, Igf2r, Kcnq1, and Dlk1) indicates that DNA methylation is generally acting to suppress the action of the gametic DMR. Thus, in the absence of DNA methylation, the gametic DMR cannot function appropriately (i.e., cannot silence the lncRNA). As a consequence, the lncRNA is ab-errantly expressed and several imprinted protein-coding genes, including Igf2, Igf2r, Kcnq1, and Dlk1, become repressed on both parental chromosomes. This indicates that these mRNA genes require epigenetic modification of a cis-regulatory element to be expressed. Notably, the H19 lncRNA that is normally only expressed on the chromosome carrying the unmethylated gametic DMR becomes expressed on both parental chromosomes. Some exceptions to this general pattern have been reported for genes that show imprinted expression only in the placenta (Lewis et al. 2004).

Are other types of epigenetic modification used as gametic imprints? Given the sheer abundance of epigenetic mechanisms acting to modify genetic information in the mammalian genome, DNA methylation is unlikely to be the only imprinting mechanism. Histone modifications that affect chromatin activity states are also likely candidates for parental imprints because they could fulfil many of the prerequisites shown in Fig. 3. In one example, the Polycomb group protein known as EED (part of the PRC2 complex that catalyses methylation of H3K27, i.e., histone H3 at lysine 27) has been shown to affect a few paternally repressed genes in the placenta. The effects of Eed mutation on genomic imprinting, however, are relatively minor compared to that of DNA methylation (Mager et al. 2003). In another example, the EHMT2 histone methyltransferase acting specifically on H3K9 is required to repress a few imprinted genes, but also only in the placenta (Nagano et al. 2008). Thus, evidence to date suggests that histone modifications and modifying enzymes play a minor role in genomic imprinting.

Although much is known about the identity and epigenetic modifications of gametic DMRs, much less is known about how these sequences are chosen for methylation in the gametes. To date, many more maternally than paternally methylated gametic DMRs have been identified (Bartolomei and Ferguson-Smith 2011). The maternally methylated DMRs are methylated during oocyte growth and the paternally methylated DMRs are methylated prenatally in pro-spermatogonia (Fig. 7) (Lucifero et al. 2002). For maternal gametic DMRs, a sequence comparison of known gametic DMRs reveals no striking sequence conservation although some contain a series of direct repeats that may adopt a secondary structure that attracts DNA methylation (Neumann et al. 1995). The tandem direct repeats in the Igf2r cluster gametic DMR have, for instance, been shown to be essential for oocyte-specific DNA methylation (Koerner et al. 2012). Those in the Kcnq1 cluster gametic DMR, however, are not essential (Mancini-DiNardo et al. 2006). Another feature of maternal DMRs is that they are markedly CpG rich compared to the remainder of the genome. One idea for how these regions are recognized comes from the structural analysis of the complexed carboxy-terminal domains of DNMT3A and DNMT3L, which was obtained by X-ray crystallography (Jia et al. 2007).A tetrameric complex consisting of these two enzymes preferentially methylate a pair of CpGs that are 8–10 base pairs

apart (Cheng 2014). Such spacing is found in maternally methylated, but not in paternally methylated imprinted loci. This CpG spacing, however, is widespread in the genome, questioning the specificity of such a mechanism or indicating that additional features are required (Ferguson-Smith and Greally 2007). Additional specificity has been suggested by the demonstration that DNMT3L interacts with the amino terminus of histone H3 if the H3K4 residue is unmethylated, and promotes local DNA methylation. Another factor contributing to the specificity of de novo DNA methylation at DMRs in the oocyte is transcription across differentially methylated regions (Chotalia et al. 2009). Importantly, only protein-coding transcripts traversing the germline ICEs are thought to be involved in DNA methylation establishment. Although it is, as yet, unclear how this transcription may be attracting the DNA methylation machinery, it has been suggested that transcription across ICEs is required to establish or maintain open chromatin domains that are permissive for the establishment of DNA methylation. To investigate and define the mechanism further, it will be necessary to describe the temporal relationship between transcription and de novo DNA methylation in greater detail. Nevertheless, CpG spacing, posttranslational histone modifications, and transcription in oocytes could provide a starting point for the acquisition of maternal-specific DNA methylation imprints.

There is far less information regarding how paternal specific DNA methylation imprints are established in the male germline. Nevertheless, early experiments suggest there could be some similarities with the female germline. It has recently been shown that high transcriptional readthrough, predominantly from one strand, is detected at two paternal gametic DMRs in primordial germ cells, H19- DMD and IG-DMR, at the time of imprint establishment (Henckel et al. 2011). It also appears that maternal gametic DMRs, which are protected from DNA methylation, are enriched for H3 lysine 4 trimethylation (H3K4me3) in male primordial germ cells. One of the most mysterious questions in genomic imprinting is how the DNA methylation marks at imprinted genes escape the genome-wide reprogramming that occurs after fertilization, including the DNA demethylation that occurs in the preimplantation embryo and the subsequent wave of de novo DNA methylation (Fig. 7). It is likely that a combination of cis-acting sequences and trans-acting factors are mediating the protection. One maternal factor, PGC7/ STELLA, appears to have a general role in maintaining DNA methylation in the early mouse embryo through interactions with H3K9me2 (Nakamura et al. 2012). However, a factor that may be more specific for imprinted genes is ZFP57. Studies have shown that ZFP57mutations identified in transient neonatal diabetes patients are associated with defects in DNA methylation at multiple imprinted loci (Mackay et al. 2008). Additionally, Zfp57 null mice show embryonic lethality and loss of imprinting at many (but not all) loci (Li et al. 2008). More recently, it has been shown that ZFP57 binds to cofactor KAP1, which can then recruit
other epigenetic regulators (Quenneville et al. 2011). Thus, sequence- and DNA methylation-dependent binding of ZFP57 could act as an anchor to specify allelic binding of KAP1, which would subsequently recruit other major repressive epigenetic regulators such as SETDB1, HP1,

DNMT1, DNMT3A, and DNMT3B to the heterochromatic, silenced allele at imprinted loci. It is possible that other yet-to-be-identified proteins also maintain DNA methylation at imprinted loci in the early embryo.

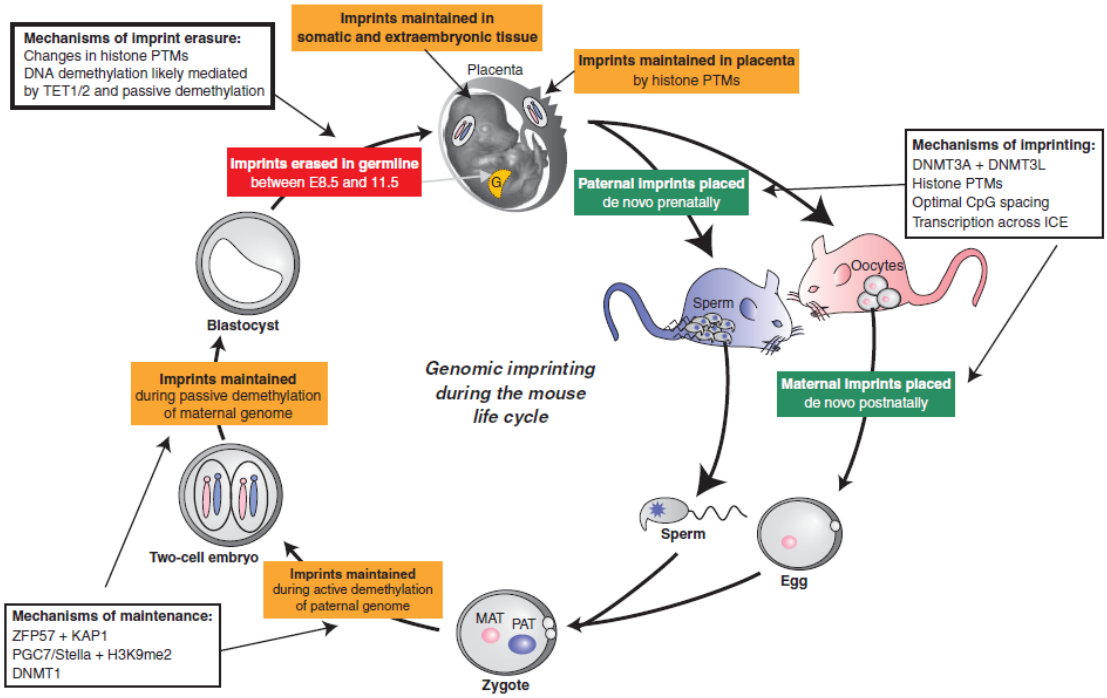


Figure 7. Establishment, maintenance, and erasure of genomic imprints in mouse development. In the germline, primordial germ cells (PGCs) undergo multiple changes in chromatin structure and DNA demethylation during migration into the genital ridge (gonad). Imprints are then acquired in a sex-specific manner in the germline (green shading). DNA methylation is targeted specifically to paternally and maternally DNA-methylated ICEs—prenatally in prospermatogonia and postnatally during oocyte maturation. These imprints are maintained despite global changes in DNA methylation after fertilization (orange shading): active demethylation of the paternal genome in the zygote and passive maternal demethylation in the preimplantation embryo. Candidates for protection of methylation regions include ZFP57 and PGC7/STELLA. De novo DNA methylation of the genome begins at the morula stage, during which time unmethylated alleles of imprinted genes must be protected. These imprints are maintained in somatic cells throughout the lifetime of the organism, whereas imprinting in extraembryonic tissues is thought to be less dependent on maintenance of DNA methylation. In the germline, imprints are erased and reset for the next generation (red shading). PTM, post-translational modification; MAT, maternal genome; PAT, paternal genome.

Two Types of cis-Acting Silencing Identified in Imprinted Gene Clusters:

Currently, two major classes of cis-acting silencing mechanisms are hypothesized to govern imprinting at various clusters: the insulator model applicable to the Igf2 cluster and the lncRNA-mediated silencing model applicable to the Igf2r and Kcnq1 clusters. Although not yet completely defined, most of the clusters in Table 1 incorporate aspects of one of the two models. The breakthrough that led to the definition of the insulator model at the Igf2 locus was the deletion of the gametic DMR (H19-DMD) that is located 2 kb upstream of the start of H19 transcription and 80 kb downstream of Igf2 (Fig. 8) (Thorvaldsen et al. 1998).

When deleted, H19 and Igf2 showed a loss of imprinting regardless of whether the deletion was inherited maternally or paternally, identifying this DMR as an ICE. It was subsequently shown that this ICE bound CTCF, a protein shown to mediate insulator activity at the beta-globin locus, and that the ICE itself functioned as an insulator (Bell and Felsenfeld 2000; Hark et al.

2000). In this context, an insulator s defined as an element that blocks enhancer and promoter interactions when placed between them. Thus, the model for imprinted gene expression at this locus is as follows: on the maternal allele, CTCF binds to the ICE and blocks the access of Igf2 and Ins2 to enhancers shared with the H19 lncRNA that are located downstream of the three genes. This thereby allows H19 exclusive access to the enhancers (Fig. 8). On the paternal allele, the ICE acquires DNA methylation in the male germline, preventing CTCF from binding to it. Thus, on the paternal chromosome, Igf2 and Ins2 interact with the enhancers and are expressed from this chromosome. The presence of DNA methylation on the paternal ICE leads to secondary methylation of the H19 promoter by an unknown mechanism and it becomes silenced on the paternal chromosome. Although the insulator model is widely accepted, it is unclear how the insulator acts at this locus. One of the most widely held views is that CTCF interacts with DNA molecules in cis to insulate genes through the formation of chromatin loops. Moreover, it has been shown that cohesin interacts with CTCF to form these
loops (Nativio et al. 2009). The involvement of CTCF in the insulator model has led to the identification of CTCF binding sites at other imprinted genes such as Rasgrf1, Grb10, and Kcnq1ot1, indicating that the insulator model may operate in other imprinted clusters. The lncRNA class of imprinting model may, however, be more common. The breakthrough that led to the identification of functional ncRNAs in imprinted clusters was an experiment that truncated the 108-kb Airn lncRNA to 3 kb (Sleutels et al. 2002). This shortened lncRNA retained imprinted expression and the Airn promoter retained imprinted DNA methylation—yet silencing of all three mRNA genes in the Igf2r cluster was lost (Fig. 8). lncRNA-mediated
silencing has also now been shown to operate at the Kcnq1 cluster(Mancini-DiNardo et al. 2006), although in a tissue specific manner, suggesting that another mechanism such as one that uses insulators may also be involved at this cluster (Shin et al. 2008), and in the Gnas imprinted cluster (Williamson et al. 2011). At this time it is not precisely known how lncRNAs silence genes but many models are possible. Two possibilities arise from the sense–antisense
overlap between an mRNA and the lncRNA that occurs in each cluster. The first possibility is that double-stranded RNA can form between them RNA and lncRNA and induce RNA-interference (RNAi) (described in Martienssen and Moazed 2014). Absence of the RNAi machinery, however, does not affect imprinted expression in the Kcnq1 cluster (Redrup et al. 2009). Thus, a second possibility is that this sense–antisense overlap causes a form of transcriptional interference of a promoter or an enhancer, which affects transcription from the mRNA promoter (Pauler et al. 2012). In this case, the first event could be silencing of the overlapped promoter or enhancer followed by accumulation of repressive chromatin that can spread and induce transcriptional gene silencing throughout the cluster. Evidence for this model comes from a series of recombinant endogenous chromosomes generated at the Igf2r/Airn locus in ES (embryonic stem) cells (Latos et al. 2012). The onset of allele-specific expression at this locus in the embryo can be recapitulated by ES cell differentiation, in which Igf2r is initially biallelically expressed, but the initiation of Airn expression results in Igf2r imprinting (Latos et al. 2009). To test whether Airn transcription or the lncRNA itself was required for Igf2r

silencing, Airn was shortened to different lengths, with the result that silencing only required Airn transcription overlap of the Igf2r promoter, which interferes with RNA polymerase II recruitment (Latos et al. 2012). This model suggests that Airn acts predominantly through its transcription rather than as an lncRNA. It is, however, also possible that imprinted lncRNAs act by coating the local chromosomal region and directly recruit repressive chromatin proteins to the imprinted cluster, in a manner similar to that described for the action of the Xist lncRNA in X-chromosome inactivation (Brockdorff and Turner 2014). Evidence for a function of the lncRNA in recruitment of histone posttranslational modification machinery comes from experiments in placental tissues. RNA fluorescence in situ hybridization experiments showed that Airn and Kcnq1ot1 form RNA clouds at their site of transcription (Nagano et al. 2008; Pandey et al. 2008; Terranova et al. 2008; Redrup et al. 2009). Terranova and colleagues show that these long ncRNAs are associated with a repressive histone compartment and Polycomb group proteins (Terranova et al. 2008). This nuclear compartment is also devoid of RNA polymerase II and exists in a three-dimensionally contracted state. Other studies on the Airn lncRNA go further in suggesting that the lncRNAs actively recruit repressive histone modifications (Nagano et al. 2008), but only in the placenta. In this latter case, Airn was shown to actively recruit the EHMT2 H3K9 methyltransferase. This resulted in the paternal-specific silencing of the Slc22a3 gene but not the Igf2r gene. These experiments indicate that lncRNA mediated silencing of imprinted genes may depend on different downstream mechanisms.

Importantly, other mechanisms of imprinted gene regulation are likely. For example, Wood and colleagues described a new imprinted locus (H13) in which alternative polyadenylation sites are used in an allele-specific manner (Wood et al. 2008). The H13 gene contains a maternally methylated internal CpG island that acquires DNA methylation in oocytes (it has not been tested for ICE activity yet). Hypermethylation of this CpG island ensures synthesis of the full length and functional H13 gene transcript from the maternal chromosome. Experiments showed that the unmethylated CpG island on the paternal allele allowed transcription from the promoter for the Mcts2 retrogene. Mcts2 expression, in turn, correlates with the premature polyadenylation of H13 and, hence, expression of truncated H13 transcripts. This locus raises the possibility that other less widely used mechanisms of genomic imprinting will be identified once the full catalogue of imprinted genes is elucidated.

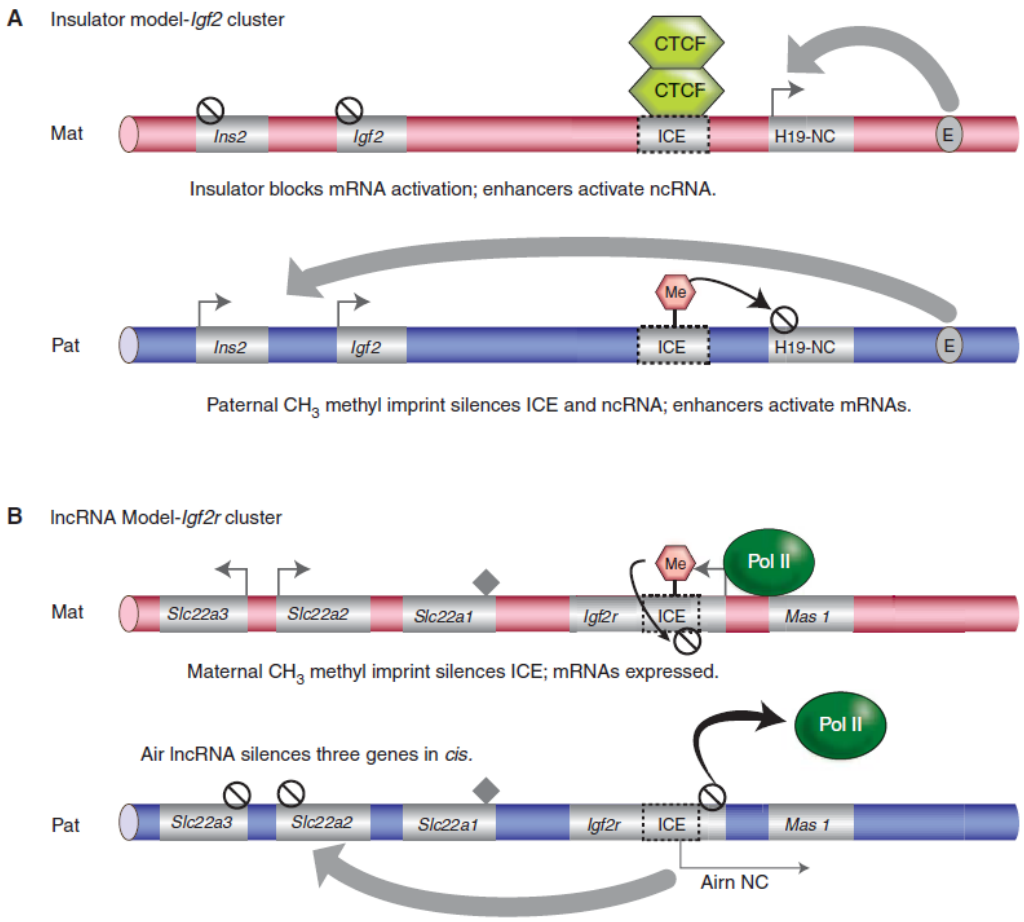


Figure 8. Two *cis*-acting silencing mechanisms at imprinted gene clusters. (*A*) Insulator model for the *Igf2* cluster. The expression pattern for endoderm is shown. On the maternal chromosome, the unmethylated ICE binds the CTCF protein and forms an insulator that prevents the common endoderm enhancers (E) from activating *Igf2* and *Ins2*. Instead the enhancers activate the nearby *H19* lncRNA promoter. On the paternal chromosome, the methylated ICE cannot bind CTCF and an insulator does not form; hence the *Igf2* and *Ins2* mRNA genes are expressed only on this chromosome. The *H19* lncRNA is methylated, most likely because of spreading from the 2-kb distant methylated ICE, and silenced. (*B*) lncRNA model for the *Igf2r* cluster. The expression pattern for placenta is shown. On the maternal chromosome, the methylated ICE contains the *Airn* lncRNA promoter that is directly silenced by the DNA methylation imprint. The *Igf2r*, *Slc22a2*, and *Slc22a3* mRNA genes are expressed only on this chromosome. *Mas1* and *Slc22a1* are not expressed in placenta (filled diamond). On the paternal chromosome, the Airn lncRNA promoter lying in the unmethylated ICE is expressed and silences *Igf2r* (in part by kicking off RNA polymerase II), *Slc22a2*, and *Slc22a3* in *cis*. Note that in both models, the DNA methylation imprint silences the lncRNA and permits mRNA expression. ICE, imprint control element; gray arrow, expressed allele of an imprinted gene; slashed circle, repressed allele of an imprinted gene; thick gray arrows, long distance effect in *cis*.

Genomic Imprinting - A model for mammalian epigenetic regulation:

Studying genomic imprinting has an advantage over other mammalian epigenetic gene regulation models because both the active and inactive parental allele reside in the same nucleus and are exposed to the same transcriptional environment (Bartolomei 2009; Barlow 2011). As a result, any epigenetic difference between the two parental alleles is more likely to correlate to their transcriptional state in contrast to “before and after” epigenetic systems, in which epigenetic changes may also reflect the altered differentiation state of the cell. The presence of both the

active and silent parental allele in the same nucleus makes genomic imprinting an ideal system to study epigenetic gene regulation. At the same time, it imposes a difficulty because it is necessary to first distinguish between the parental alleles so that specific features associated with gene activity and silencing can be attributed to the right parental allele. This difficulty has been largely overcome in the mouse by the development of model systems that allow the maternal and paternal chromosomes to be distinguished (Fig. 1). Despite the fact that epigenetic gene regulatory mechanisms are highly conserved in evolution, there are likely
to be differences that relate to the type of genome organization for each organism. The mammalian genome shows an unusual organization that intersperses genes with high copy number repeats (also known as transposable elements). This greatly increases the length of most genes as well as the distance between adjacent genes. This contrasts with other model organisms such as yeast, nematodes, plants, and Drosophila, whose genomes show a tendency
toward remaining repeat-free or, at least to separate repeats from genes. How can genomic imprinting contribute to an understanding of mammalian epigenetics? Although the characterization of imprinted gene clusters is far from complete, they clearly have the potential to provide information about how genes are controlled in local regions or domains. To date, imprinted gene clusters have already provided examples of cis-acting DNA sequences that are regulated by DNA methylation, genes that are silenced by default in the mammalian genome
and require epigenetic activation to be expressed, long range regulatory elements that can act as insulators, and unusual lncRNAs that silence large domains of genes in cis. Time will tell whether these types of epigenetic regulatory mechanisms are unique to imprinted clusters or whether they can also be found regulating expression of nonimprinted genes in the mammalian genome.

Conclusion:

Genomic imprinting has been the focus of intense interest since the discovery of the first imprinted genes in mammals in 1991. Whereas early experiments relied on molecular and genetic strategies to identify imprinted genes, high throughput technology on polymorphic individuals is allowing the complete determination of imprinted genes (Deveale et al. 2012) and regions containing parental-specific DNA methylation (Xie et al. 2012). These experiments
are indicating that most genes showing ubiquitous imprinted expression have already been identified.

However, it is possible that some genes showing tissue specific imprinted expression remain to be identified (Prickett and Oakey 2012). Some questions still await conclusive answers, particularly those concerning why mammals alone among vertebrates use imprinted genes to
regulate embryonic and neonatal growth. This lack of knowledge contrasts with the extensive progress during the intervening 20 years on elucidating the epigenetic mechanisms controlling imprinted expression in mammals. From this information, we think we understand the general principles of how the imprinting mechanism operates at imprinted gene clusters, although all the details are still not clear. At this stage, it is clear that genomic imprinting uses the cell’s normal

epigenetic machinery to regulate parental-specific expression, and that everything is set in motion by restricting this machinery in the gamete to just one parental allele. Although there are general similarities in the mechanism controlling imprinted expression at different gene clusters, it is not yet understood how many variants of this mechanism exist in the mammalian genome. In the future, it will also be of interest to determine to what degree nonimprinted genes are controlled by the epigenetic mechanisms described for imprinted gene clusters. Ultimately, transferring this knowledge for therapeutic use in humans, for example, by inducing re-expression of the silent parental alleles in patients with the Prader-Willi and Angelman syndromes to ameliorate their symptoms would be of great benefit. An understanding of the way the cell controls epigenetic information is of increasing importance, with the realization that epigenetic regulation can also be disturbed in cancers (Jones and Baylin 2014), in assisted reproductive technologies and also in the aging process (Rando and Chang 2012; Berger and Sassone-Corsi 2014). An improved understand ing of genomic imprinting will undoubtedly continue to provide an important model to discover how the mammalian genome uses epigenetic mechanisms to regulate gene expression.

Probable questions:

1. What is Genomic imprinting? How it affect gene expression?
2. How does a gametic imprint control imprinted expression?
3. What is the function of genomic imprinting in mammals?
4. What is the role of the gametic DMR?
5. Discuss the Role of DNA Methylation in Genomic Imprinting.
6. Briefly discuss two types of cis-Acting Silencing identified in imprinted Gene Clusters.

Suggested Readings:

1. Molecular Cell Biology by Lodish, Fourth Edition.
2. The Cell – A Molecular Approach by Cooper and Hausman, Fourth Edition
3. Principles of Genetics by Snustad and Simmons, Sixth Edition.
4. Molecular Biology of the Cell – by Bruce Alberts
5. Cell and Molecular Biology by Gerald Karp, 7th Edition.
6. Gene cloning and DNA Analysis by T. A. Brown, Sixth Edition.
7. Genetics . Verma and Agarwal.
8. Brown TA. (2010) Gene Cloning and DNA Analysis. 6th edition. Blackwell Publishing, Oxford, U.K.

Unit-VIII

Nuclear transplantation and the reprogramming of the genome. epigenetics and human disease, epigenetic determinants of cancer

Objective: In this unit we will discuss on nuclear transplantation and the reprogramming of the genome. We will also discuss epigenetics and human disease and also epigenetic determinants of cancer.

Nuclear Transplantation:

Nuclear transplantation is a method in which the nucleus of a donor cell is relocated to a target cell that has had its nucleus removed (enucleated). Nuclear transplantation has allowed experimental embryologists to manipulate the development of an organism and to study the potential of the nucleus to direct development. Nuclear transplantation, as it was first called, was later referred to as somatic nuclear transfer or cloning.

Yves Delage first wrote about nuclear transplantation in 1895, speculating that if one were to replace an egg nucleus with another egg's nucleus, full development would occur. Later in 1938, Hans Spemann suggested an experiment whereby, using technologies not yet available to him, one could remove the nucleus of an egg and replace it with a different nucleus extracted from a developed cell. Thomas King and Robert Briggs were the first to perform experimental nuclear transplantation. The technique was soon after used by John Gurdon and eventually led to the first clone of a mammal, "Dolly" the sheep, by Ian Wilmut in 1996.

Nearly fifteen years after Spemann wrote about the possibility of nuclear transplantation, Briggs and King, using northern leopard frogs (*Rana pipiens*), performed the first nuclear transplantation experiment. They transplanted the nucleus from an early stage embryo to an unfertilized egg that had been enucleated. The egg cell was pricked with a clean glass needle in order to induce a fertilization-like response. The faux activation of fertilization allowed for extraction of the nuclear material inside while also activating the host egg cell. Meanwhile, the nucleus of a donor cell was extracted and then inserted into the newly enucleated and activated egg cell. That process induced development of the host egg according to the instructions of the newly inserted nucleus, resulting in the formation of an organism with the same genetic material as the donor cell, or a clone. Briggs and King continued to examine the potential of differentiated cells throughout the 1950s. They found that if the donor nucleus was extracted later in development, the potential of directing full development in the activated egg cell was greatly reduced. After the Briggs and King experiments it was generally accepted that the nuclear material in developing cells slowly loses its potential for full development.

That view was challenged in 1958 when Gurdon's experiments with African claws frogs (*Xenopus laevis*) produced fully developed frogs from the transferred nucleus of cells much later in development. Gurdon allowed the cloned frogs to develop to sexual maturity and was then able to mate two sexually mature clones, suggesting that the donor nuclei were able to fully redirect development. Gurdon's experiments were widely accepted by the scientific community but questions remained for several decades. Scientists were concerned about whether the nucleus of the host egg cell was truly enucleated. The question of whether remnants of the host egg cell or the inserted nucleus directed development remained unanswered from 1958 to 2002, despite many attempts by Gurdon to prove it was the inserted nucleus. In 2002, however, Konrad Hochedlinger and Rudolf Jaenisch published an experiment using nuclear transplantation of mature white blood cells to generate mouse clones. Hochedlinger and Jaenisch were able to show that the inserted nucleus induced development in the host egg cell.

Although experimental embryologists continued to use nuclear transplantation to create clones of several species, Ian Wilmut's cloning experiment in 1996 was a controversial and widely publicized cloning experiment. Dolly was cloned using the nucleus of a mammary gland cell from an adult sheep and transplanting it into an enucleated egg cell from another sheep. The activated egg cell was then transferred into a third surrogate sheep that carried Dolly to term. Dolly died at the age of to six due to lung disease and severe arthritis, and although her death was not attributed to the fact that she was a clone, many believe that the relationship between telomeres and aging was the reason for her demise.

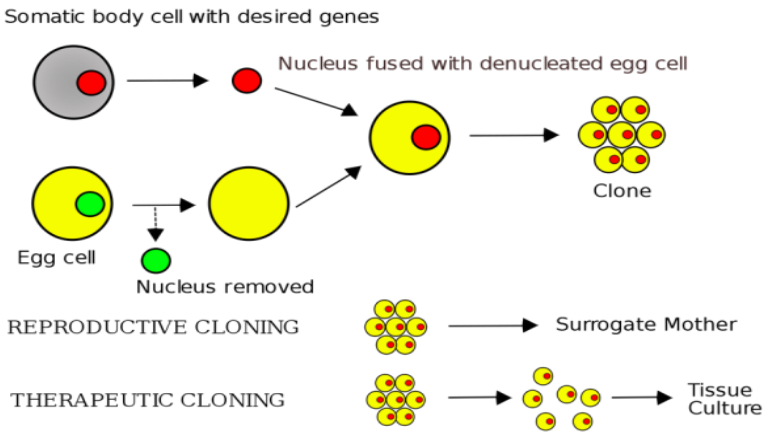
Nuclear transplantation may have begun as a subtle idea in the late 19th and early 20th centuries, but it evolved into a feasible and widely used process by experimental embryologists in the late 1990s. The cloning of Dolly the sheep worried many about the possibility of human cloning and the moral boundaries of modern advances in science. In the context of the embryonic stem cell discourse of the late 1990s and early twenty-first century, somatic nuclear transfer has been contrived into moral arguments about rights of the human embryo. Furthermore, nuclear transplantation has spurred ethical discussion on the value of a human life during all stages of development. Many scientists have abandoned the methods involved in nuclear transplantation and have adopted methods set forth by Shinya Yamanaka in his experiments involving induced pluripotent stem cells.

Brief Conventional Process:

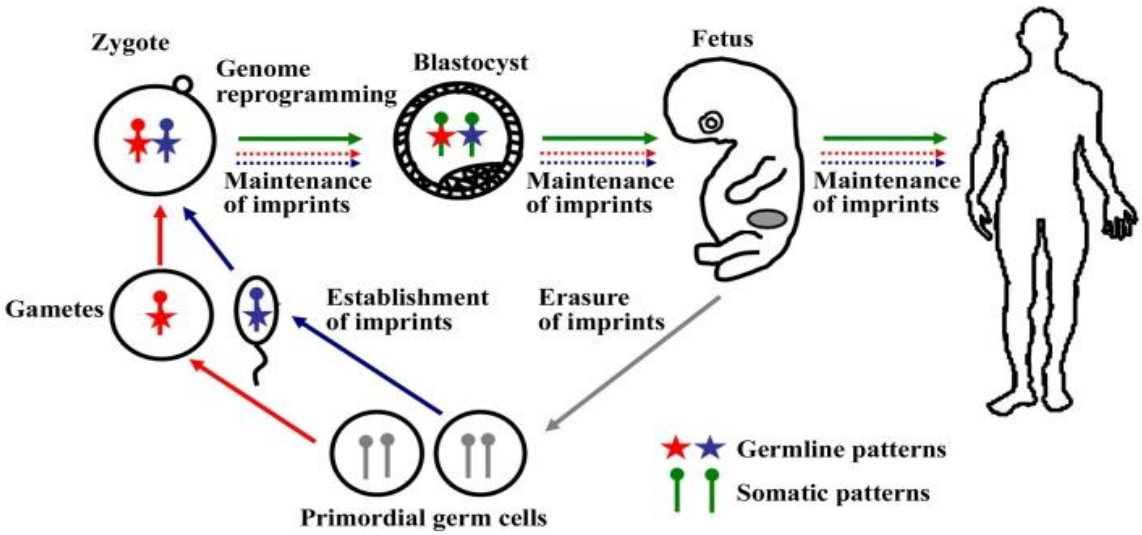
Nuclear Transfer Cloning:

Also known as "somatic cell nuclear transfer" (SCNU), the process of nuclear transfer cloning requires two cells. The somatic cell, the cell collected from the animal that is to be cloned (the genetic donor), is any cell other than an egg or sperm cell but contains the complete DNA (Deoxyribonucleic Acid) or genetic blueprint. This can be easily obtained from a skin biopsy performed by a veterinarian. The second cell required is the egg cell, which is collected from any

female of the same species (the egg donor). In the figure below, the somatic cell is in grey and the egg cell, in yellow. Once both cells are obtained, the cloning process can begin. As shown in the diagram, the nucleus (green circle) of the egg cell is extracted and discarded of. By the doing this, it removes the egg cell of the genetic information of the original host, as the genes are contained within the nucleus. As the genetic information from the somatic cell is needed, the same is done; the nucleus of the somatic cell (red circle) is extracted, and the resulting egg cell and nucleus are then "fused" together through electricity, resulting in the egg cell containing the DNA of the genetic donor. The cell is then stimulated, which then causes it to divide, just as any zygote (fertilised egg) would during reproduction, as shown in the part of figure 1 labelled "Clone". The egg is then placed in a culture medium, a liquid or gel designed to support growth micro-organisms or cells. The cellular division continues over several days until the blastocyst (early stage embryo) is formed. Within a week, an embryo transfer specialist inserts the blastocyst into a surrogate mother. Once it has fully developed, the mother gives birth to a clone of the genetic donor.



The concept of nuclear transfer cloning was created in 1928 by a German embryologist by the name of Hans Spemann. In his initial experiments, Spemann had transferred salamander embryonic cell nucleic into egg cells. It was only decades later that the embryologist conceived the concept of generating clones by transferring nuclei from differentiated cells into enucleated egg cells, nuclear transfer cloning. At the time, Spemann thought this process impossible, due to the fact that the microsurgical tools necessary to perform such delicate procedure without causing damage to the genetic material or egg cells did not exist. Eventually in the 1990's, a team of scientists from Roslin Institute Scotland used the proposed process to produce sheep clones, and thus in 1996, Dolly was born (figure below). This process has helped make major strides in the area of stem cell research and nuclear reprogramming.



Basic principles of epigenetics:

DNA methylation and histone modifications The human genome contains 23,000 genes that must be expressed in specific cells at precise times. Cells manage gene expression by wrapping DNA around clusters (octamers) of globular histone proteins to form nucleosomes (Fig. 1A). These nucleosomes of DNA and histones are organized into chromatin. Changes to the structure of chromatin influence gene expression: genes are inactivated (switched off) when the chromatin is condensed (silent), and they are expressed (switched on) when chromatin is open (active) (Fig. 1B). These dynamic chromatin states are controlled by reversible epigenetic patterns of DNA methylation and histone modifications. Enzymes involved in this process include DNA methyltransferases (DNMTs), histone deacetylases (HDACs), histone acetylases, histone methyltransferases and the methyl-binding domain protein MECP2. Alterations in these normal epigenetic patterns can deregulate patterns of gene expression, which results in profound and diverse clinical outcomes.

The loss of normal DNA methylation patterns is the best understood epigenetic cause of disease, based on the initial studies during the 1980s that focused on X chromosome inactivation, genomic imprinting and cancer. DNA methylation involves the addition of a methyl group to cytosines within CpG (cytosine/guanine) pairs (Fig. 1A). Typically, unmethylated clusters of CpG pairs are located in tissue specific genes and in essential “housekeeping” genes, which are involved in routine maintenance roles and are expressed in most tissues. These clusters, or CpG “islands,” are targets for proteins that bind to unmethylated CpGs and initiate gene transcription. In contrast, methylated CpGs are generally associated with silent DNA, can block methylation sensitive proteins and can be easily mutated. DNA methylation patterns are established and maintained by DNMTs, enzymes that are essential for proper gene expression patterns. In animal experiments, the removal of genes that encode DNMTs is lethal; in humans, overexpression of these enzymes has been linked to a variety of cancers. In addition to DNA methylation, changes to histone proteins orchestrate DNA organization and gene expression. Histone-modifying

enzymes are recruited to ensure that a receptive DNA region is either accessible for transcription or that DNA is targeted for silencing (Fig. 1B). Active regions of chromatin have unmethylated DNA and have high levels of acetylated histones, whereas inactive regions of chromatin contain methylated DNA and deacetylated histones. Thus, an epigenetic "tag" is placed on targeted DNA, marking it with a special status that specifically activates or silences genes. These reversible modifications ensure that specific genes can be expressed or silenced depending on specific developmental or biochemical cues, such as changes in hormone levels, dietary components or drug exposures.

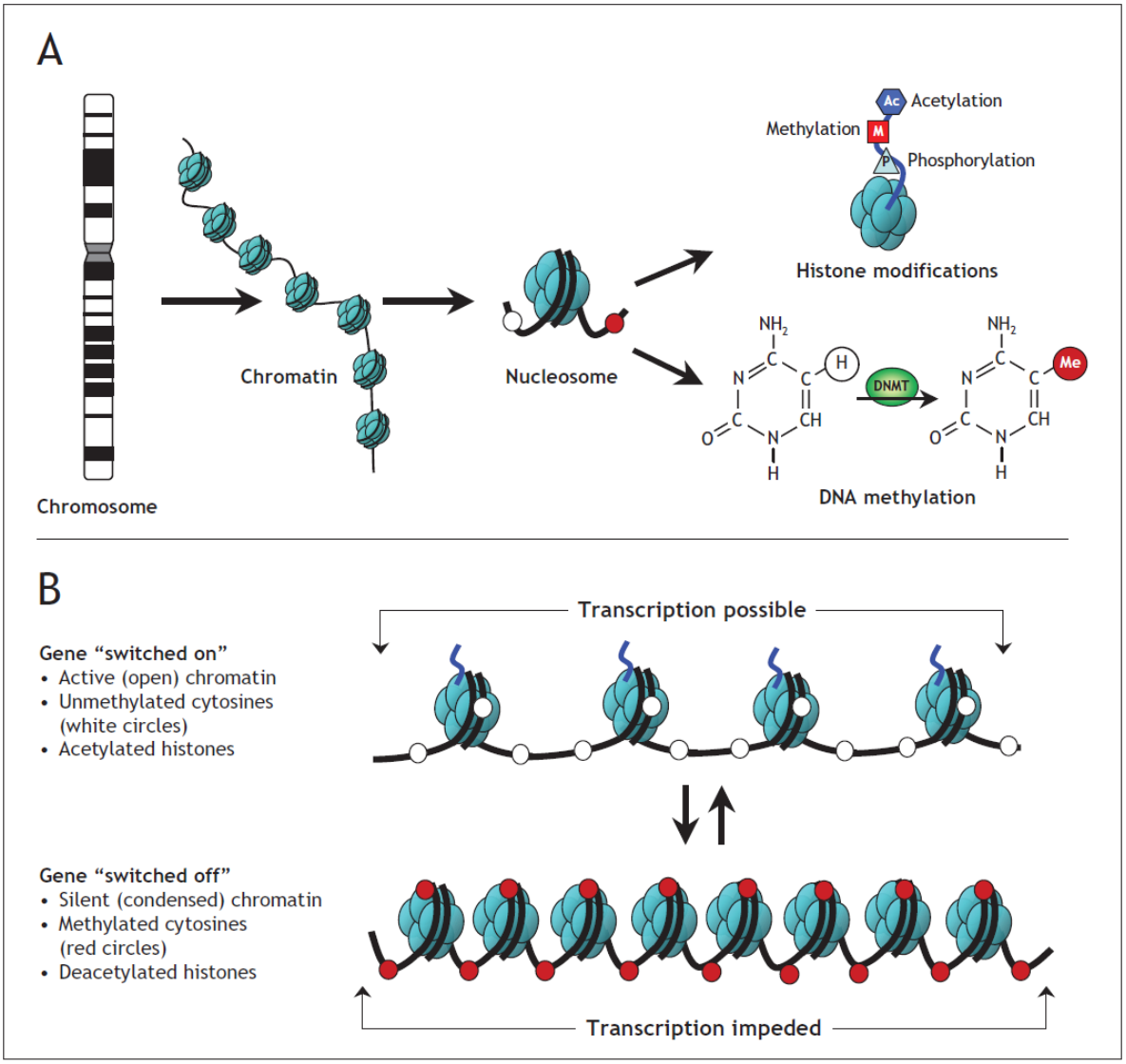


Fig 1: (A) Schematic of epigenetic modifications. Strands of DNA are wrapped around histone octamers, forming nucleosomes. These nucleosomes are organized into chromatin, the building block of a chromosome. Reversible and site-specific histone modifications occur at multiple sites through acetylation, methylation and phosphorylation. DNA methylation occurs at 5-position of cytosine residues in a reaction catalyzed by DNA methyltransferases (DNMTs). Together, these modifications provide a unique epigenetic signature that regulates chromatin organization and gene expression. (B) Schematic of the reversible changes in chromatin organization that influence gene expression: genes are expressed (switched on) when the chromatin is open (active), and they are inactivated (switched off) when the chromatin is condensed (silent).[3] White circles = unmethylated cytosines; red circles = methylated cytosines.

Genomic imprinting and imprinting disorders:

Genomic imprinting allows genes to “remember” whether they were inherited from the mother or the father so that only the maternally or paternally inherited allele is expressed. Imprinting is regulated by DNA methylation and histone modifications and is important in the context of a variety of developmental and paediatric disorders. Prader–Willi, Angelman and Beckwith–Weidemann syndromes best characterize congenital imprinting disorders. Prader–Willi and Angelman syndromes are caused by genetic and epigenetic errors to the same part of chromosome; errors inherited from the father result in Prader–Willi syndrome, and those inherited from the mother, Angelman syndrome. Beckwith–Wiedemann syndrome is caused by genetic or epigenetic mutations resulting in loss of imprinting on chromosome 11. Besides gene-specific imprinting effects, global imprinting changes can occur in embryos that completely lack one parental genome. For example, spontaneous activation of oocytes in situ leads to ovarian teratomas that lack a paternal genome. In contrast, complete hydatidiform moles have been found that lack a maternal genome and an embryo and exhibit hyperproliferation of trophectoderm tissues, with the potential of forming choriocarcinoma. The recessive disorder “familial biparental complete hydatidiform mole” also leads to recurrent development of moles when maternal specific imprints fail to be established during oogenesis.

Interestingly, imprinting effects that appear to target trophoblast cells have been recently implicated as a cause of preeclampsia.

Aging and epigenetics:

Both increases and decreases in DNA methylation are associated with the aging process, and evidence is accumulating that age-dependent methylation changes are involved in the development of neurologic disorders, autoimmunity and cancer in elderly people.88 Methylation changes that occur in an age-related manner may include the inactivation of cancer related genes. In some tissues, levels of methylated cytosines decrease in aging cells, and this demethylation may promote chromosomal instability and rearrangements, which increases the risk of neoplasia. In other tissues, such as the intestinal crypts, increased global hypermethylation may be the predisposing event that accounts for the increased risk of colon cancer with advancing age.

Table 1: Associations between epigenetic modifications and human diseases and conditions

Disease/condition	Gene	Biological process	Disease/condition	Gene	Biological process
Cancer			Neurologic		
Bladder	Multiple genes	Hypermethylation[20]	Schizophrenia	*RELN*	Hypermethylation[46,47]
Brain (glioma)	*RASSF1A*	Hypermethylation[28,29]	Bipolar disorder	*11p?*	Unknown[48]
Brain (glioblast)	*MGMT*	Hypermethylation[30]	Memory formation	Multiple genes	Hypo-, hypermethylation[49]
Breast	*BRCA1*	Hypermethylation[31]	Lupus	Retroviral DNA	Hypomethylation[50]
Breast	Multiple genes	Hypermethylation[32,33]	Cardiovascular		
Cervix	*p16*	Hypermethylation[34]	Atherosclerosis	Multiple genes	Hypo-, hypermethylation[19,51]
Colon	Multiple genes	Hypermethylation[20]	Homocysteinemia	Multiple genes	Hypomethylation[52]
Colorectal	L1 repeats	Hypomethylation[35]	Vascular endothelium	*eNOS*	Hypomethylation[53]
Esophagus	*CDH1*	Hypermethylation[20]	Imprinting and pediatric syndromes		
Head/neck	*p16, MGMT*	Hypermethylation[20]	PWS or AS	15q11-q13	Imprinting[54]
Kidney	*TIMP-3*	Hypermethylation[20]	BWS	11p15	Imprinting[55]
Leukemia	p15	Hypermethylation[20]	SRS	Chromosome 7	Imprinting[56]
Liver	Multiple genes	Hypermethylation[36]	UPD14	14q23-q32	Imprinting[57]
Lung	*p16, p73*	Hypermethylation[20]	PHP, AHO, MAS	20q13.2	Imprinting[58]
Lymphoma	*DAPK*	Hypermethylation[20]	Rett syndrome	*MECP2*	Mutation[59]
Myeloma	*DAPK*	Hypermethylation[37]	ICF syndrome	*DNMT3B*	Mutation[60]
Ovary	*BRCA1*	Hypermethylation[38]	ATRX	*ATRX*	Chromatin structure[61]
Ovary	*Sat2*	Hypomethylation[39]	FraX	Triplet repeat	Silencing[62]
Pancreas	*APC*	Hypermethylation[20]	FSHD	3.3 kb repeat	Chromatin structure[63]
Pancreas	Multiple genes	Hypomethylation[40]	Reproductive		
Prostate	*BRCA2*	Hypermethylation[20,41]	Ovarian teratoma	No paternal genome	Imprinting[64]
Rhabdomyosarcoma	*PAX3*	Hypermethylation[42]	CHM	No maternal genome	Imprinting[65]
Stomach	*Cyclin D2*	Hypomethylation[43]	BiCHM	Maternal genome	Imprinting[65]
Thymus	*POMC*	Hypomethylation[44]	Aging	Chromatin	Hypo-, hypermethylation[66]
Urothelial	Satellite DNA	Hypomethylation[45]			
Uterus	*hMLH1*	Hypermethylation[20]			

Note: PWS = Prader-Willi syndrome; AS = Angelman syndrome; BWS = Beckwith-Weidemann syndrome; SRS = Silver-Russell syndrome; UPD14 = uniparental disomy 14; PHP = pseudohypoparathyroidism; AHO = Albright hereditary osteodystrophy; MAS = McCune-Albright syndrome; ICF = immunodeficiency, centromeric instability and facial anomalies; ATRX = α-thalassemia/mental retardation syndrome, X-linked; FraX = Fragile X syndrome; FSHD = facioscapulohumeral muscular dystrophy, CHM = complete hydatidiform mole, BiCHM = familial biparental CHM.

Immunity and related disorders:

The activation of the immune response involves stepwise epigenetic changes, which allow individual cells to mount a specific immune response that can be maintained over multiple cell generations.90,91 For example, shifts in both acetylation and methylation are required to coordinate DNA accessibility and permit recombination, thereby allowing cells to mount an immune response against a specific antigen. Recent reports suggest that loss of epigenetic control over this complex process contributes to autoimmune disease. Abnormal DNA methylation has been observed in patients with lupus whose T cells exhibit decreased extracellular signal-regulated kinase pathway signalling, decreased methyltransferase activity and hypomethylated DNA. Disregulation of this pathway apparently leads to overexpression of methylation-sensitive genes such as the leukocyte function-associated factor (LFA1), which causes lupus-like autoimmunity. Interestingly, LFA1 expression is also required for the development of arthritis,

which raises the possibility that altered DNA methylation patterns may contribute to other diseases displaying idiopathic autoimmunity.

Neuropsychiatric disorders:

Recent reports have begun to address the role of epigenetic errors in the causation of complex adult psychiatric, autistic and neurodegenerative disorders (Table 1). Several reports have associated schizophrenia and mood disorders with DNA rearrangements that include the DNMT genes. DNMT1 is selectively overexpressed in gamma-aminobutyric acid (GABA)-ergic interneurons of schizophrenic brains, whereas hypermethylation has been shown to repress expression of Reelin (a protein required for normal neurotransmission, memory formation and synaptic plasticity) in brain tissue from patients with schizophrenia and patients with bipolar illness and psychosis. In addition, the HDAC inhibitor valproic acid has been shown to prevent Reelin promoter hypermethylation in a mouse model of schizophrenia. A role for aberrant methylation mediated by folate levels has been suggested as a factor in Alzheimer's disease; however, there is contradictory evidence regarding hypomethylation and overexpression of the presenilin-1 gene that is involved in synaptic plasticity, long-term memory and neuronal survival. As well, some preliminary evidence supports a model that incorporates both genetic and epigenetic contributions in the causation of autism. Autism has been linked to the region on chromosome 15 that is responsible for Prader–Willi syndrome and Angelman syndrome. Findings at autopsy of brain tissue from patients with autism have revealed deficiency in MECP2 expression that appears to account for reduced expression of several relevant genes. These results suggest that MECP2 deficiency plays a role in chromosome organization in the developing brain in autism, Rett syndrome and several other neurodevelopmental disorders. There may be a role for epigenetics in the diagnosis and treatment of complex neuropsychiatric disorders in the future.

Table 4. Summary of epigenetic aberrations reported in mental diseases

Disease	Epigenetic change (tissues)
Fragile X syndrome	Hyper-methylation at the FMR-1 gene with an expanded (CCG)n repeat
Huntington	Histone modification in HDACs and histone KDM5D/Kdm5d
Rett syndrome	Mutation in the gene encoding *MeCP2*
Autistic patients and their parents	Abnormal trans-methylation, trans-sulfuration metabolism, genome-wide DNA hypo-methylation and elevated blood homocysteine level (blood)
Down syndrome	miR-99a, let-7c, miR-125b-2, miR-155, and miR-802 up-regulation
SCZ	DNA hyper-methylation of the RELN promoter and SOX10 promoter (brain)
SCZ and BD	DNA hypo-methylation of the MB-COMT promoter (brain)
SCZ	Histone 3 lysine 4 hypo-methylation at the GAD1 promoter due to mixed-lineage leukemia 1 gene dysfunction (brain)
SCZ (male)	DNA hyper –methylation of the WDR18 gene (brain)
SCZ (male)	Global DNA hypo-methylation (blood)
SCZ & Psychotic BD	DNM T1 hyperexpression and increase in SAM content (corticalinter-neurons)
Bipolar ll	DNA hypo-methylation of *PPIEL* gene (blood)
BD (female)	Hypo-methylation of *RPL39* (brain)
Dementia	Hyper-methylation of circadian genes, *PER1* and *CRY1* (blood)
Alcoholism	DNA hyper-methylation of alpha synuclein promoter, HERP gene promoter and dopamine transporter gene (blood)

SCZ, schizophrenia; BD, bipolar disorders

Paediatric syndromes and epigenetics:

In addition to epigenetic alterations, specific mutations affecting components of the epigenetic pathway have been identified that are responsible for several syndromes: DNMT3B in the ICF (immunodeficiency, centromeric instability and facial anomalies) syndrome, MECP2 in Rett syndrome, 105 ATRX in ATR-X syndrome (a-thalassemia/mental retardation syndrome, X linked) and DNA repeats in facioscapulohumeral muscular dystrophy (Table 1). In Rett syndrome, for example, MECP2 encodes a protein that binds to methylated DNA; mutations in this protein cause abnormal gene expression patterns within the first year of life. Girls with Rett syndrome display reduced brain growth, loss of developmental milestones and profound mental disabilities. Similarly, the ATR-X syndrome also includes severe developmental deficiencies due to loss of ATRX, a protein involved in maintaining the condensed, inactive state of DNA. Together, this constellation of clinical paediatric syndromes is associated with alterations in genes and chromosomal regions necessary for proper neurologic and physical development.

Cancer and epigenetic therapies:

Cancer is a multistep process in which genetic and epigenetic errors accumulate and transform a normal cell into an invasive or metastatic tumour cell. Altered DNA methylation patterns change the expression of cancer-associated genes (Table 1). DNA hypomethylation activates oncogenes and initiates chromosome instability, whereas DNA hypermethylation initiates silencing of tumour suppressor genes. The incidence of hypermethylation, particularly in sporadic cancers, varies with respect to the gene involved and the tumour type in which the event occurs. For example, p16ink4A promoter hypermethylation occurs in varying degrees (9%–49%) in as many as 15 cancer types; in contrast, BRCA1 hypermethylation is primarily associated with 10%–20% of sporadic breast and ovarian cancers.20 These epigenetic changes can be used in the molecular diagnosis of a variety of cancers. To date, epigenetic therapies are few in number, but several are currently being studied in clinical trials or have been approved for specific cancer types. Nucleoside analogues such as azacitidine are incorporated into replicating DNA, inhibit methylation and reactivate previously silenced genes. Azacitidine has been effective in phase I clinical trials in treating myelodysplastic syndrome and leukaemia characterized by gene hypermethylation. For example, 54% of patients with chronic myelogenous leukaemia resistant to imatinib exhibited a complete or partial hematologic response, and 46% had a major or minor cytogenetic response to 5-aza-2'-deoxycytidine. The antisense oligonucleotide MG98 that downregulates DNMT1 is showing promising results in phase I clinical trials and in targeting solid tumours and renal cell cancer. Molecular analysis of biopsies of head and neck cancer following MG98 treatment revealed demethylation of 2 methylated tumour suppressor genes and methylation of an oncogene. Similarly, small molecules such as valproic acid that downregulate HDACs are being used to induce growth arrest and tumour cell death. Combination epigenetic therapies (demethylating agents plus HDAC inhibitors) or epigenetic therapy followed by conventional chemotherapy (or immunotherapy) may be more effective since they reactivate silenced genes, including tumour suppressor genes, resensitize drug-resistant cells to standard therapies and act synergistically to kill cancer cells. The key challenge for the future will be to limit toxic effects in normal cells and ensure that these novel drug effects reach critical target genes in tumour cells.

Epigenetic modifications have a considerable effect on cancer. Hypermethylation of promoter regions in tumour suppressor genes can inactivate many tumour suppressor functions. Methylation levels also play an important role in cell divisions, DNA repair, differentiation, apoptosis, angiogenesis, metastasis, growth factor response, detoxification, and drug resistance. Such features have promoted huge advances in the early detection of cancer using methylation levels. For example, hypermethylation of promoter regions in APC and RASSF1A genes are considered as common epigenetic markers for early detection of cancer. Also, hypermethylation of TP53 promoter region has been reported as a common marker for evaluation of cancer development. There are also some other types of epigenetic changes in cancer. In recent years, dysregulation of miRNAs has been confirmed in breast cancer, which has a potential to be used

as diagnostic biomarkers. Also, hyper- and hypo-methylation of several genes in breast cancer have been confirmed.

Microsatellite instability, chromosomal instability, and CpG island methylator phenotype have been identified as three major mechanisms affecting gene function in colorectal cancer (CRC). Microsatellite instability occurs in 15% of CRCs, which can result in instability phenotype by mutated or methylated mismatch repair genes[. In a comprehensive analysis of CRC tumours in Iranian patients, Brim et al. demonstrated a high microsatellite instability rate (18%). From 15 known methylation target genes, APC2, PTPRD, EVL, GPNMB, MMP2, and SYNE1 were found to be methylated in most samples, which can be potentially used as specific clinical and pathological markers of CRC in this population.

The pathogenesis of CRC has been reported to be controlled by miRNAs, which can act as regulators of oncogenic and tumour suppressor pathways, responsible for the development of cancer. It has been confirmed that different miRNAs can be useful as biomarkers and are potentially applicable in prognosis evaluation and the detection of CRC stages. It has been also observed that in the absence of O6-methylguanine-DNMTs activity as a DNA repair protein, the specific genes, such as K-ras and p53, might be accumulated by G-to-A transition. Furthermore, hypermethylation near the methylguanine-DNMT start codon in the specific locus is critical for cancer progression, which may have a prognostic value in CRC patients.

It has been indicated that miRNAs play an important role in many types of cancer: acute myeloid leukaemia, acute lymphocytic leukaemia, chronic myeloid leukaemia, chronic lymphocytic leukaemia, endometrial carcinoma, gastrointestinal cancer, lung cancer, bladder cancer, thyroid tumours, and oesophageal adenocarcinomas. Hence, the potential applications of miRNAs in diagnosis and prognosis of these cancers would be highlighted in the near future. Isocitrate dehydrogenase 1 (IDH1) and IDH2 genes are frequently mutated in low-grade gliomas, de-novo acute myeloid leukaemia in adult and in the subsets of chondrosarcomas and lymphomas. Interestingly, high correlation between histone and DNA methylation phenotype in IDH mutant gliomas has been reported.

Table 1. Promoter methylation in different types of cancer

Cancer type	Gene	Promoter methylation	Reference
Breast	RARB2, MSH2, ESR1B, AKR1B1, COL6A2, GPX7, HIST1H3C, HOXB4, RASGRF2,TM6SF1, ARHGEF7, TMEFF2, RASSF1, BRCA1, STRATIFIN, RASSF1A	Hypermethylation	[102]
Gastric	RUNX3	Hypermethylation	[102]
Liver	CDKN2A	Hypermethylation	[102]
Esophageal	APC	Hypermethylation	
Colorectal	SEPT9, hMLH1, CDKN2A/p16, HTLF, ALX4, TMEFF2/HPP1, NGFR, SFRP2, NEUROG1, RUNX3,UBE2Q1	Hypermethylation	[103,104]
Lung	RARB2, RASSF1A, CHFR, STRATI-FIN, SHOX2, RASSF1A APC1	Hypermethylation	[102]

Table 2. Histone modifications in different types of cancer

Cancer type	Type of histone modification
Lung adenocarcinoma	Up-regulation of α-2 glycoprotein 1 in consequence of global histone acetylation[105]
Non-small cell lung	Global H3 deacetylation[106]
Gastric	Global H3K9 trimethylation[107] Silencing of RUNX3 in the consequence of increased H3K9 dimethylation and decreased H3 acetylation[108]
Prostate	Global H3K9, H3K18, and H4K12 acetylation and H4K3 and H3K4 dimethylation[109] Activation of PTEN, CYLD, p53, and FOX03a by modulating histone H3K9 methylation and deacetylation[110]
Colorectal	Global H3K9 deacetylation[111]
Pancreatic	Acetylation of histone H3 promoter region of C/EPBα[112]

Table 3. miRNA changes in different types of cancer[3,65,113]

Cancer type	Types of miRNA[(+)=up-regulation/(-)=down-regulation]
Oesophageal squamous cell carcinomas	miR-21(+)
Lung	miR-17-92 (+) miR-34c, miR-145, and miR-142-5p, let-7(-)
Primary head and neck squamous cell carcinoma	miR-1, miR-133a, miR-205, and led-7d(-) bsa-miR-21(+)
Gastric	miR-106a(+) miR-433 and miR-9(-)
Prostate	miR-135b and miR-194(+) miR-23b, miR-100, miR-145, miR-221, miR-222(-)
Melanoma	miR-182(+)
Hepatocellular	miR-18a(-)
Colorectal	miR-let 7g, miR-21, miR-20a, miR-17-19 family, miR31, miR 135, miR-181b, and miR 200c (+) miR-34, miR-let7, miR-143, miR-145, miR-133b, and miR-126(-)
Bladder	miR-2 23, miR-26b, miR-221, miR-103-1, miR-185, miR-23 b, miR- 203, miR 17-5p, miR-23, miR-205(+) miR-29c, miR-26a, miR-30c, miR-30e-5p, miR-45, miR-30a-3p, miR-133a, miR-133b, miR-195, miR-125b,and miR-199a (-)
Breast	miR-21, miR-155, miR-23, and miR-191(+) miR-205, miR-145, miR-10b, and miR-125b (-)

Epigenetic Determinants of Cancer

Epigenetic changes are present in all human cancers and are now known to cooperate with genetic alterations to drive the cancer phenotype. These changes involve DNA methylation, histone modifiers and readers, chromatin remodelers, microRNAs, and other components of

chromatin. Cancer genetics and epigenetics are inextricably linked in generating the malignant phenotype; epigenetic changes can cause mutations in genes, and, conversely, mutations are frequently observed in genes that modify the epigenome. Epigenetic therapies, in which the goal is to reverse these changes, are now one standard of care for a preleukemic disorder and form of lymphoma. The application of epigenetic therapies in the treatment of solid tumors is also emerging as a viable therapeutic route.

Cancer is caused by the somatically heritable deregulation of genes that control the processes governing when cells divide, die, and move from one part of the body to another. During carcinogenesis, genes can become activated in such a way that enhances division or prevents cell death (oncogene). Alternatively, genes can become inactivated so that they are no longer available to apply the brakes to these processes (tumour suppressor gene). It is the interplay between these two classes of genes that results in the formation of cancer. Tumour-suppressor genes (TSGs) can become inactivated by at least three pathways: (1) through mutations, in which their functions become disabled; (2) a gene can be completely lost and thus not be available to work appropriately (loss of heterozygosity); and (3) a gene can be switched off in a somatically heritable fashion by epigenetic changes, rather than by mutation of the DNA sequence. Epigenetic silencing can occur by deregulation of the epigenetic machinery at several different levels; it may involve inappropriate methylation of cytosine (C) residues in CpG sequence motifs that reside within control regions governing gene expression. Also, changes to histone posttranslational modifications (PTMs) or aberrations in the way histone-modifying enzymes function may occur. A change in a protein's ability to read histone marks, and hence bind to chromatin, or alterations in the way nucleosome-remodelling or histone exchange complexes function can result. Finally, changes in regulatory microRNA (miRNA) expression patterns have been noted. The basic molecular mechanisms responsible for maintaining the silenced state are quite well understood, as outlined in this collection. Consequently, we also know that epigenetic silencing has profound implications for cancer prevention, detection, and therapies. We now have drugs approved by the U.S. Food and Drug Administration (FDA) that are used to reverse epigenetic changes and restore gene activity to cancer cells. Also, because changes in DNA methylation can be detected with a high degree of sensitivity, many strategies are able to detect cancer early by finding changes in DNA methylation. The translational opportunities for epigenetics in human cancer research, detection, prevention, and treatment are, therefore, quite extraordinary.

Probable Questions:

1. What is nuclear transplantation. Briefly explain.
2. Describe the procedure of nuclear transplantation.
3. How epigenetics affects aging?
4. How epigenetics affects immunity in human?
5. How epigenetics affects Neuropsychiatric disorders
6. Discuss pediatric syndromes and epigenetics.
7. Discus role of epigenetics in cancer formation.
8. What do you know about epigenetic determinants of cancer ?

Suggested Readings:

1. Molecular Cell Biology by Lodish, Fourth Edition.
2. The Cell – A Molecular Approach by Cooper and Hausman, Fourth Edition
3. Principles of Genetics by Snustad and Simmons, Sixth Edition.
4. Molecular Biology of the Cell – by Bruce Alberts
5. Cell and Molecular Biology by Gerald Karp, 7^{th} Edition.
6. Gene cloning and DNA Analysis by T. A. Brown, Sixth Edition.
7. Genetics . Verma and Agarwal.
8. Brown TA. (2010) Gene Cloning and DNA Analysis. 6th edition. Blackwell Publishing, Oxford, U.K.

UNIT-IX

Mobile genetic elements: Characteristics of transposable elements in prokaryotes and eukaryotes; AC/DS system in maize

Objective: In this unit we will discuss about different types of transposable genetic elements both in prokaryotes and eukaryotes.

Definition of Transposons:

Presence of transposable elements was first predicted by Barbara McClintock in maize (corn) in late 1940s. After several careful studies, she found that certain genetic elements were moving from one site to an entirely different site in the chromosome. She called this phenomenon of changing sites of genetic elements as transposition and those genetic elements were called by her as controlling elements. These controlling elements were later on called as transposable elements by Alexander Brink. In late 1960s this phenomenon was also discovered in bacteria.

Consequently, the molecular biologists called them as Transposons. A transposon may be defined as: **"a DNA sequence that is able to move or insert itself at a new location in the genome."** The phenomenon of movement of a transposon to a new site in the genome is referred to as transposition.

Transposons are found to encode a special protein named as transposase which catalyses the process of transposition. Transposons are particular to different groups of organisms. They constitute a fairly accountable fraction of genome of organisms like fungi, bacteria, plants, animals and humans. Transposons have had a major impact on changing or altering the genetic composition of organisms. Transposons or transposable genetic elements are often referred to as 'mobile genetic elements' also. They can be categorized on different bases like their mode of transposition or on the basis of the organisms in which they are present.

Types of Transposons:

Different transposons may change their sites by following different transposition mechanisms.

On the basis of their transposition mechanism, transposons may be categorized into following types:

(i) Cut-and-Paste Transposons:

They transpose by excision (cutting) of the transposable sequence from one position in the genome and its insertion (pasting) to another position within the genome (Fig. 1).

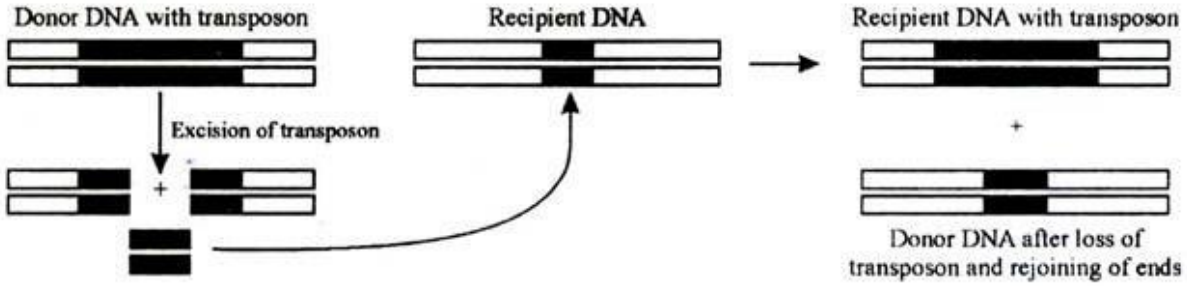


Fig. 1. Cut and Paste Transposons.

The cut-and-paste transposition involves two transposase subunits. Each transposase submit binds to the specific sequences at the two ends of transposon. These subunits of transposase protein then come together and lead to the excision of transposon.

This excised 'transposon-Transposase Complex' then gets integrated to the target recipient site. In this manner, the transposon is cut from one site and then pasted on other site by a mechanism mediated by transposase protein (Fig. 2).

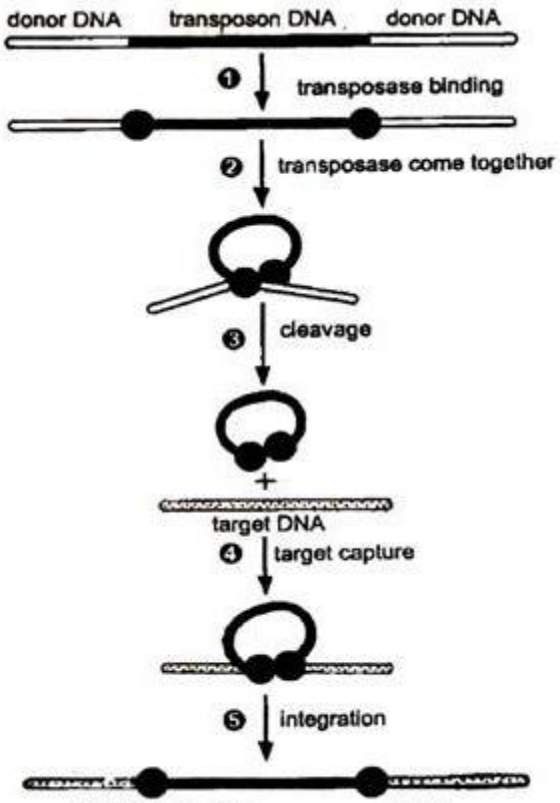


Fig. 2. Role of Transposase protein in cut-and-paste transposition.

Examples of cut-and-paste type of transposons are IS-elements, P-elements in maize, hobo-elements in Drosophila etc.

(ii) Replicative Transposons:

They transpose by a mechanism which involves replication of transposable sequence and this copy of DNA, so formed, is inserted into the target site while the donor site remains unchanged (Fig. 3). Thus, in this type of transposition, there is a gain of one copy of transposon and both-the donor and the recipient DNA molecule are having one-one transposable sequence each, after transposition.

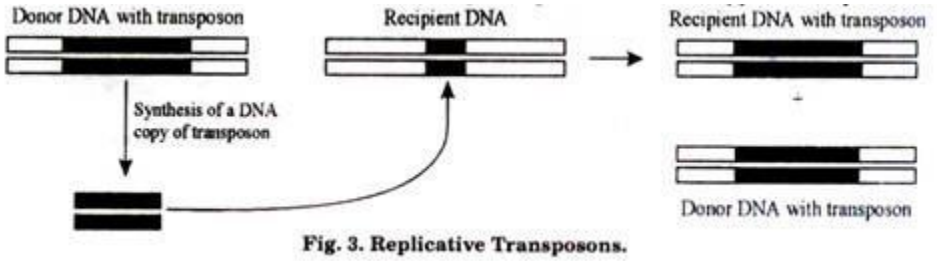


Fig. 3. Replicative Transposons.

Figure: Tn3-elements found in bacteria are good examples of such type of transposons.

(iii) Retro Elements:

Their transposition is accomplished through a process which involves the synthesis of DNA by reverse transcription (i.e. RNA DNA) by using elements RNA as the template (Fig. 4). This type of transposition involves an RNA intermediate, the transposable DNA is transcribed to produce an RNA molecule.

This RNA is then used as a template for producing a complementary DNA by the activity of enzyme reverse transcriptase. This single stranded DNA copy so formed, is then made double stranded and then inserted into the target DNA site. The transposable elements which require reverse transcriptase tor their movement are called retro transposons.

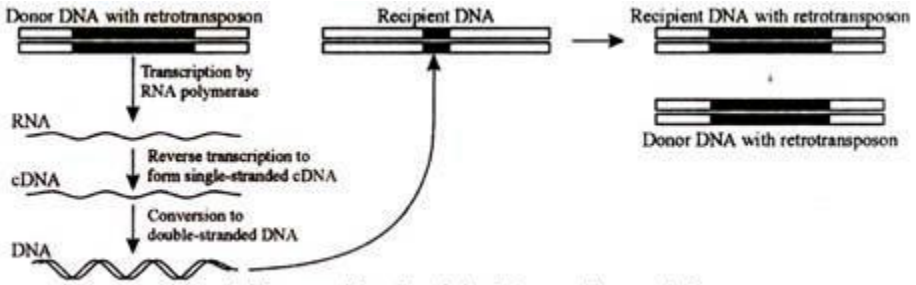


Fig. 4. Transposition Involving Reverse Transcription.

The Retro elements may be viral or non-viral. Out of these two, the non-viral retro elements are important and may further be classified as:

(A) Retrovirus like elements:

They carry long terminal repeats (LTR). Examples are copia, gypsy elements in Drosophila.

Retroposons:

LTR are absent. Examples are LINEs and SINEs in humans.

Structure of a Transposon:

Transposons are stretches of DNA that have repeated DNA segments at either end. A transposon consists of a central sequence that has transposes gene and additional genes. This is flanked on both sides by short repeated DNA segments. The repeated segments may be direct repeats or inverted repeats. These terminal repeats help in identifying transposons.

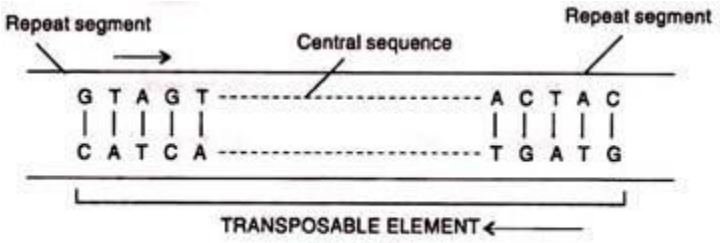


Fig. 18.2.

The number of repeated nucleotides is uneven 5 or 7 or 9 nucleotides are due to its method of insertion at the target site.

Target Sequence:

The site where a transposon is inserted is called target site or recipient site. Before the transposon is moved into the target site, the target sequence is duplicated. The two copies formed move apart. The transposon is inserted in between the two copies of the target sequences.

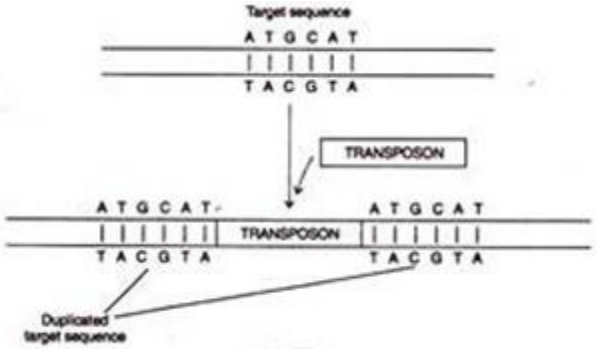


Fig. 18.3.

Mechanism of Transposition:

The enzyme transposase present in the transposon itself makes nicks or cuts in each strand of the target sequence. The target sequence is duplicated and two copies move away to make way for the transposon in the centre. The transposon then fixes itself into the two free ends generated. The nicks are sealed by ligase and two strands become continuous.

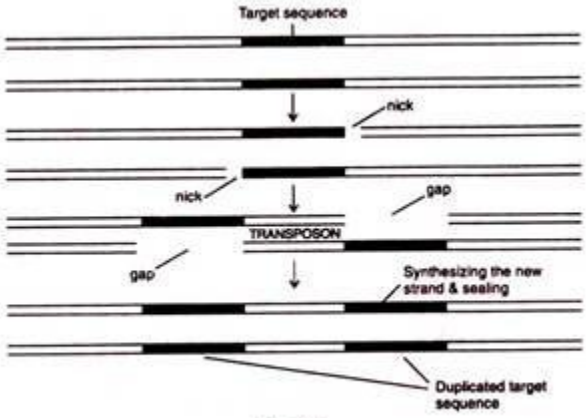


Fig. 18.4.

Transposable genetic elements in Prokaryotes:

The four transposable genetic elements in prokaryotes are: (1) Bacterial Insertion Sequences (2) Prokaryotic Transposons (3) Insertion-Sequence Elements and Transposons in Plasmids and (4) Phage mu.

(1) Bacterial Insertion Sequences:

Insertion Sequences or Insertion-Sequence (IS) Elements:

Insertion sequences, or insertion-sequence (IS) elements, are now known to be segments of bacterial DNA that can move from one position on a chromosome to a different position on the same chromosome or on a different chromosome. An IS element contains only genes required for mobilizing the element and inserting the element into a chromosome at a new location. Is elements are normal constituents of bacterial chromosome and plasmids. When IS elements appear in the middle of genes, they interrupt the coding sequence and inactivate the expression of that gene. Owing to their size and in some cases the presence of transcription and translation termination signals, IS elements can also block the expression of other genes in the same operon if those genes are downstream from the promoter of the operon. IS elements were first found in E. coli as a result of their affects on the expression of a set of three genes whose products are needed to metabolize the sugar galactose as a carbon source. Careful investigations showed that the mutant phenotypes resulted from the insertion of an approximately 800 base pairs (bp) DNA segment into a gene. This particular DNA segment is now called insertion sequencel (IS1).

Properties of IS Elements:

Is1 is the genetic element capable of moving around the genome. It integrates into the chromosome at locations with which it has no homology, thereby distinguishing it from recombination. This event is an example of transposition event. There are number of IS elements that have been identified in E. coli, including IS1, IS2, and IS 10, each present in 0 to 30 copies per genome, and each with a characteristic length and unique nucleotide sequence.

IS 1 is 768 bp long, and is present in 4 to 19 copies on the E. coli chromosomes. IS2 is present in 0 to 12 copies on the E. coli chromosome and in one copy on the F plasmid, and IS 10 is found in a class of plasmids called R plasmid that can replicate in E. coli (Fig. 12.1).
Among prokaryotes, the IS elements are normal cell constituents, that is, they are found in most cells. Altogether, IS elements constitute approx. 0.3% of the cell's genome. All IS elements that have been sequenced, end with perfect or nearly perfect inverted terminal repeats (IRs) of between 9 and 41 bp. This means that essentially the same sequence is found at each end of an IS but in opposite orientations.

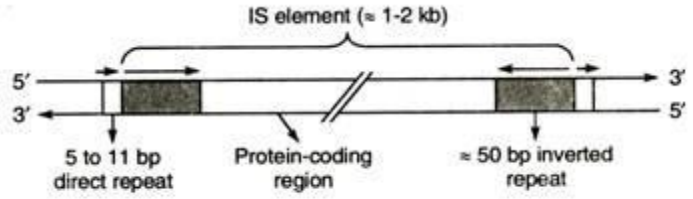


Fig. 12.1. Structure of bacterial IS elements.

IS Transposition:

When transposition of an IS element takes place, a copy of the IS element inserts into a new chromosome location while the original IS elements remains in place. That is, transposition requires the precise replication of the original IS element, using the replication enzymes of the host cell. The actual transposition also requires an enzyme encoded by the Is element called transposase.
The IR sequences are essential for the transposition process, that is, those sequences are recognized by transposase to initiate transposition. Is elements insert into the chromosomes at sites with which they have no sequence homology?
Genetic recombination between non-homologous sequences is called illegitimate recombination. The sites into which IS elements insert are called target sites. The process of IS insertion into a chromosome is shown in Figure 12.2. Firstly, a staggered cut is made in the target site and the IS element is then inserted, becoming joined to the jingle-stranded ends.
The gaps are filled in by DNA polymerase and DNA ligase, producing an integrated IS element with two direct repeats of the target site sequence flanking the IS element. 'Direct'

in this case means that the two sequences are repeated in the same orientation. The direct repeats are called target site duplications. The sizes of target site duplication vary with the IS elements, but tend to be small. Integration of some IS elements show preference for certain regions, while others integrate only at particular sequences.

All copies of a given IS element have the same sequence, including that of the inverted terminal repeats. Mutations that affect the inverted terminal repeat sequence of IS elements affect transposition, indicating that the inverted terminal repeat sequences are the key sequences recognized by transposase during a transposition event.

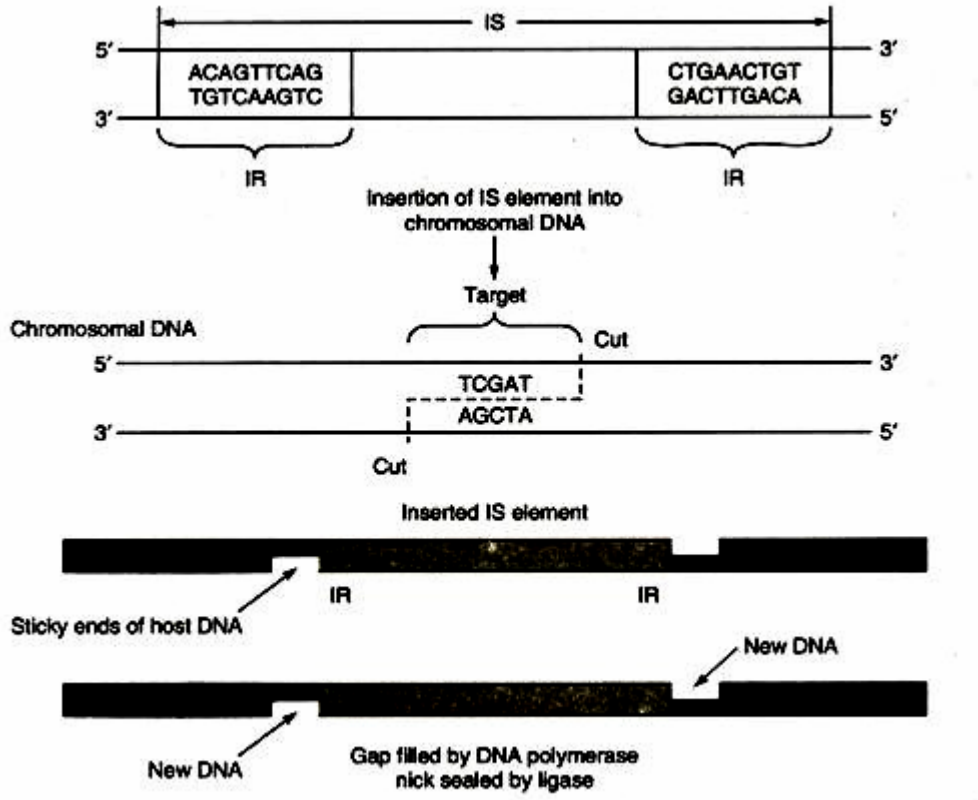


Fig. 12.2. Schematic presentation of the integration of an IS elements into chromosomal DNA.

(2) Prokaryotic Transposons:

A transposon (Tn) is more complex than an IS elements. A transposon is a mobile DNA segment that, like an IS element, contains genes for the insertion of the DNA segment into the chromosome and for the mobilization of the element to other locations on the chromosome. There are two types of prokaryotic transposons: composite transposons and non-composite transposons.

(a) Composite Transposons:

They are complex transposons with a central region containing genes, e.g., drug resistance genes, flanked on both sides by IS elements (also called IS modules). Composite transposons may be thousands of base pairs long. The IS elements are both of the same types and are called IS-L (for "left") and IS-R (for "right"). Depending upon the transposon, IS-L and IS-R may be in the same or inverted orientation relative to each other. Because the ISs themselves have terminal inverted repeats, the composite transposons also have terminal inverted repeats.

Figure 12.3 shows the structure of the composite transposon Tn 10 to illustrate the general features of such transposons. The Tn 10 transposon is 9,300 bp long and consists of 6,500 bp of central, nonrepeating DNA containing the tetracycline resistance gene flanked at each end with a 1,400-bp IS element. These IS elements are designated IS10L and IS10R and are arranged in an inverted orientation. Cells containing Tn 10 are resistant to tetracycline resistance gene contained within the central DNA sequence. Transposition of composite transposon occurs because of the function of the IS elements they contain. One or both IS element supplies the transposase. The inverted repeats of the IS elements at the two ends of the transposon are recognized by transposase to initiate transposition (as with transposition of IS elements).

Transposition of Tn 10 is rare, occurring once in 10 cell generations. This is the case because less than one transposase molecule per cell generation is made by Tn 10. Like IS elements, composite transposons produce target site duplications after transposition.

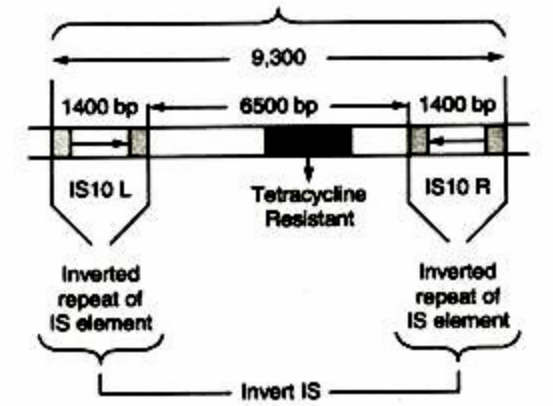


Fig. 12.3. Detailed structure of Tn10 transposon.

(b) Non-composite Transposons:

They like composite transposons, contain genes such as those for drug resistance. Unlike composite transposons, they do not terminate with IS elements. However, they do have the repeated sequences at their ends that are required for transposition. Tn3 is a non-composite transposon.

Tn3 has 38 bp inverted terminal repeats and contains three genes in its central region. One of those genes, bla, encodes β-lactamase which breaks down ampicillin and therefore makes cells containing Tn3 resistant to ampicillin. The other two genes, tnpA and tnpB, encode the enzymes transposase and resolvase that are needed for transposition of Tn3 (Fig. 12.4). Transposase catalyzes insertion of the Tn into new sites, and resolvase is an enzyme involved in the particular re-combinational events associated with transposition.

Resolvase is not found in all transposons. The genes for transposition are in the central region for non-composite transposons, while they are in the terminal IS elements for composite transposons. Non composite transposons also cause target site duplications when they move.

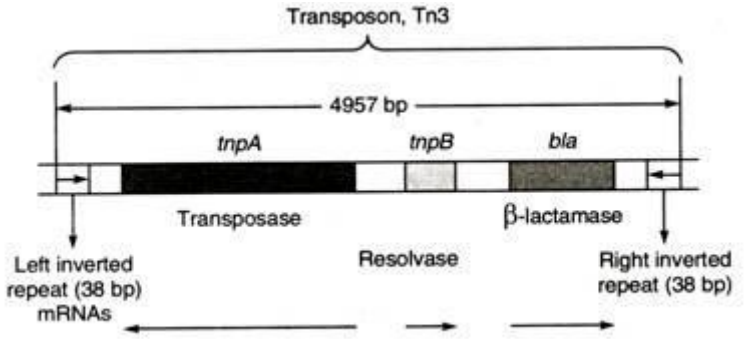


Fig. 12.4. Detailed structure of Tn3 transposon.

(c) Mechanism of Transposition in Prokaryotes:

Several different mechanisms of transposition are employed by prokaryotic transposable elements. And, as we shall see later, eukaryotic elements exhibit still additional mechanisms of transposition. In E. coli, we can identify replicative and conservative (non-replicative) modes of transposition. In the replicative pathway, a new copy of the transposable element is generated in the transposition event. The results of the transposition are that one copy appears at the new site and one copy remains at the old site. In the conservative pathway, there is no replication. Instead, the element is excised from the chromosome or plasmid and is integrated into the new site (Fig. 12.5).

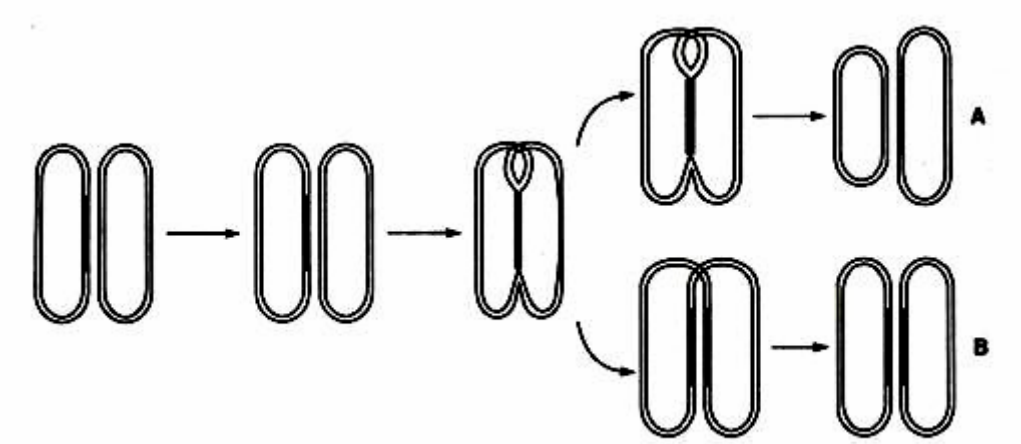


Fig.12.5. A. Conservative mode of transposition B. Replicative mode of transposition.

Replicative Transposition:

The transposition of Tn3 occurs in two stages. Firstly, the transposase mediates the fusion of two molecules, forming a structure called a cointegrate. During this process, the transposon is replicated, and one copy is inserted at each junction in the cointegrate. The two Tn3 are oriented in the same direction. In the second stage of transposition, the tnpR-encoded resolvase mediates a site-specific recombination event between the two Tn3 elements. This event occurs at a sequence in Tn3 called res, the resolution site, and generates two molecules, each with a copy of the transposon.

The tnpR gene-product also has another function, namely, to repress the synthesis of both the transposase and resolvase proteins. This repression occurs because the res site is located in between the tnpA and tnpR genes. By binding to this site, the tnpR protein interferes with the synthesis of both gene-products, leaving them in chronic short supply. Consequently, the Tn3 element tends to remain immobile (Fig. 12.6).

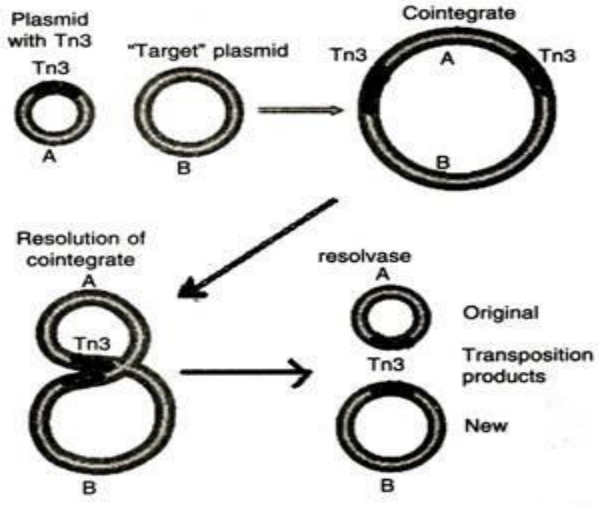


Fig. 12.6. Process of Tn3 transposition through cointegrate intermediate.

Conservative transposition:

Some transposons, such as Tn10, excise from the chromosome and integrate into the target DNA. In these cases, DNA replication of the element does not occur, and the element is lost from the site of the original chromosome. This mechanism is called conservative (non replicative) transposition or simple insertion. Tn 10, e.g., transposes by conservative transposition.

Insertion of a transposon into the reading frame of a gene will disrupt it, causing a loss of function of that gene. Insertion into gene's controlling region can cause changes in the level of expression of the gene. Deletion and insertion events also occur as a result of activities of the transposons, and from crossing-over between duplicated transposons in the genome.

(3) IS Elements and Transposons in Plasmids:

The transfer of genetic material between conjugating E. coli is the result of the function of the fertility factor F. The F factor, a circular double stranded DNA molecule, is one of the example of bacterial plasmid. Plasmids such as F that are also capable of integrating into the bacterial chromosomes are called episomes. F factor consists of 94,500 bp of DNA that code for a variety of proteins.

The important elements are:

(i) Transfer gene (tra) required for the conjugation transfer of the DNA.

(ii) Genes that encode proteins required for the plasmid's replication,

(iii) Four IS elements, two copies of IS3, one of IS2, and one of an insertion sequence element called gamma- delta.

It is because the E. coli chromosome has copies of these four insertion sequence at various positions that the F factor can integrate into the E. coli chromosome at different sites and in different orientations with homologous sequence of the insertion elements.

Another class of plasmids that has medical significance is the R plasmid group, which was discovered in Japan in the 1950s, during the cure for dysentery. The disease is the result of infection by the pathogenic bacterium Shigella. Shigella was found to be resistant to most of the commonly used antibiotics.

Subsequently, they found that the genes responsible for the drug resistances were carried on R plasmids, which can promote the transfer of genes between bacteria by conjugation, just as the F factor. One segment of an R plasmid that is homologous to a segment in the F factor is the part needed for the conjugal transfer of genes. That segment and the plasmid-specific genes for DNA replication constitute what is called the RTF (resistance transfer factor) region (Fig. 12.7). The rest of the R plasmid differs from type to type and includes the antibiotic-resistance genes or other types of genes of medical significance, such as resistance to heavy metal ions.

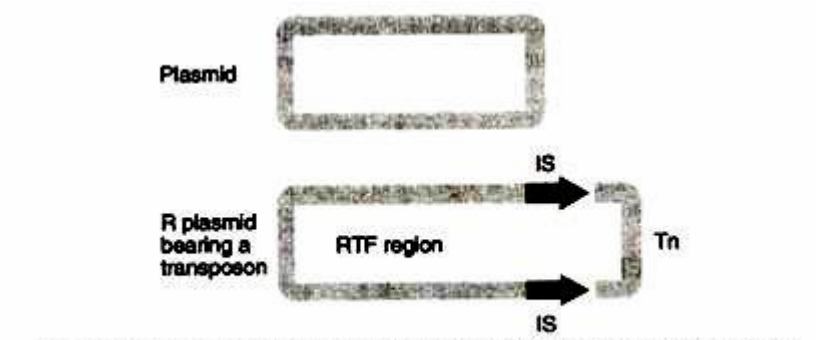


Fig. 12.7. The insertion of a transposon (Tn) into a plasmid. RTF represents the resistance-transfer functional genes of the plasmid.

The resistance genes in R plasmid are, in fact, transposons, that is each resistance gene is located between flanking, directly repeated segments such as one of the IS modules (Fig. 12.8). Thus, each transposon with its resistance gene in the R plasmid can be inserted into new location on other plasmids or on the bacterial chromosome, while at the same time leaving behind a copy of itself in the original position.

Fig. 12.8. Two different transposons having different inverted repeat (IR) regions and carrying different drug-resistance genes. (a) Tn9 has a short IR region, because the two IS1 elements are in the same orientation and each element has a short inverted repeat. (b) Tn10 has a large IR region because the two IS10 components have opposite orientations, and the entire IS10 sequence constitutes the inverted repeat.

(4) Phage mu:

Phage mu is a normal-appearing phage. We consider it here because, although it is a true virus, it has many features in common with IS elements. The DNA double helix of this phage is 36,000 nucleotides long-much larger than an IS element. However, it does appear to be able to insert itself anywhere in a bacterial or plasmid genome in either orientation. Once inserted, it causes mutation at the locus of insertion-again like an IS element. (The phage was named for this ability: mu stands for “mutator.”)

Normally, these mutations cannot be reverted, but reversion can be produced by certain kinds of genetic manipulation. When this reversion is produced, the phages that can be

recovered showing no deletion, proving that excision is exact and that the insertion of the phage therefore does not involve any loss of phage material either. Each mature phage particle has on each end a piece of flanking DNA from its previous host (Fig. 12.9). However, this DNA is not inserted anew into the next host. Its function is unclear. Phage mu also has an IR sequence, but neither of the repeated elements is at a terminus.

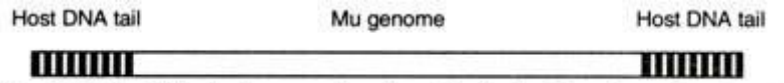


Fig. 12.9. The DNA of a free mu phage has tails derived from its previous host.

Mu can also act like a genetic snap fastener, mobilizing any kind of DNA and transposing it anywhere in a genome. For example, it can mobilize another phage (such as λ) or the F factor. In such situations, the inserted DNA is flanked by two mu genomes (Fig. 12.10).

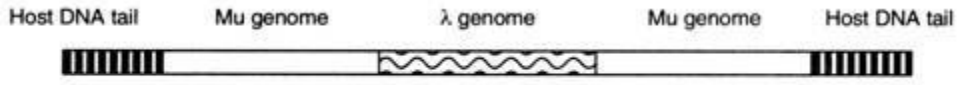


Fig. 12.10. Phage mu can mediate the insertion of phage λ into a bacterial chromosome, resulting in a structure like the one shown here.

Probable Questions:

1. What is Transposable genetic element?
2. Describe types of transposons.
3. What is retrotransposons?
4. Describe the properties of IS elements.
5. What is composite transposons?
6. What is non composite transposons?
7. Describe the mechanism of transpositions?
8. Describe phage mu transposons.

Suggested Readings:

1. Molecular Cell Biology by Lodish, Fourth Edition.
2. The Cell – A Molecular Approach by Cooper and Hausman, Fourth Edition
3. Principles of Genetics by Snustad and Simmons, Sixth Edition.
4. Molecular Biology of the Cell – by Bruce Alberts
5. Cell and Molecular Biology by Gerald Karp, 7th Editi

UNIT-X

Mobile genetic elements: Characteristics of transposable elements in eukaryotes; AC/DS system in maize, P element in Drosophila; Salmonella phage variation; retrosposons

Objective: In this unit we will discuss about transposable elements in eukaryotes; AC/DS system in maize, P element in Drosophila; Salmonella phage variation and also about retrosposons.

Transposable elements in the eukaryotic organisms:

Transposons have been discovered in eukaryotic organisms also, e.g., controlling elements in maize, Tam1 elements in *Antirrhinum majus*, Ty elements in Yeast and FB elements in Drosophila. These elements can be divided into two main classes. 1. This class includes the transposable elements that are similar to those found in bacteria. These elements contain inverted repeats at their ends and generate short direct repeats of the target DNA at the sites of their insertion. These elements are always located in the host genome and cannot survive outside the genome. Controlling elements in maize and P elements in Drosophila belong to this class of transposable elements. 2. Retroposons (Retro transposons). Retroposons are DNA elements formed by the reverse transcription of retroviruses. This class of retroviruses and other sequences are transposed via RNA. Transposition of retroposons occurs through RNA intermediates.

Controlling Elements in Maize:

In 1940's Barbara McClintock discovered changes in maize genome during somatic cell division. The changes were genetically controlled aberrations, such as, deficiencies, duplications, inversions, translocations and ring chromosomes. These changes were found to be caused by a genetic system named Dissociation-Activator (Ds-Ac) system.

McClintock termed these genetic elements Ds and Ac as controlling elements in 1956. Since then, several systems of controlling elements have been discovered in maize. These elements are classified into two groups: autonomous and non-autonomous.

1. Autonomous elements:

The controlling elements which have the ability of their own excision and transposition are called autonomous elements, e.g., Activator (Ac), Suppressor mutator (Spm) and Enhancer (En).

2. Non-autonomous elements:

These elements do not have the ability of transposition. Non-autonomous elements have originated from autonomous elements through the loss of transacting functions which are required for transposition. A single type of autonomous element and different non-autonomous elements derived from it form a family.

Deletions of different lengths and different regions from an autonomous element give rise to different types of non-autonomous elements. Such elements change their position in response to an autonomous element of the same family present in the genome.

Non-autonomous element is activated in trans by its related autonomous element. Examples of non-autonomous elements are Dissociation (Ds), defective suppressor mutator (dSpm), and Inhibitor (I).

Dissociation-Activator (DS-Ac) System:

The main features of the Ds-Ac system in maize are as follows:

(i) An Ac element can exist in a number of states similar to other genes, and it controls the activity and time of action of the Ds element.

(ii) Ac and Ds, both exhibit inter-chromosomal as well as intra-chromosomal movements (transposition). The movement occurs through excision of these elements from one site and their insertion at a new site.

(iii) Ds element is unstable in the presence of the Ac element in the same nucleus. When both the elements are present, loss (deletion) of a part of the chromosome 9 occurs if the chromosome 9 carries the Ds element. The deletion is caused by breakage of the chromosome at the site of Ds.

(iv) The genes lying adjacent to the Ds become inactivated.

(v) The number of Ac elements present in a genome has a negative relationship with the time of Ds action during the development. Therefore, the presence of Ac in a greater number delays the transposition of Ds during the development. This can be well explained in the maize endosperm which is triploid.

In maize endosperm, the number of Ac element may vary from 0 to 6. The dominant allel I at the C locus on chromosome 9 inhibits colour formation in the aleurones of kernels so that the kernels having the I allele are colourless. In the presence of recessive allele i, colour develops normally in the aleurone. Therefore, an ii secondary nucleus fertilized by a pollen carrying i allele will produce coloured aleurone iii. But an ii secondary nucleus fertilized by a pollen carrying the dominant allele I will produce colourless endosperm Iii.

If both, Ac and Ds are present in the above I pollen, and Ds occupies a place within or near the C locus, coloured spots would be observed in many kernels. The coloured spots develop due to the transposition of the Ds element from the allele I during the stages of seed development which permits the c locus to produce aleurone colour. An increase in the

number of Ac elements delays the dissociation of Ds. Thus variegation pattern in the kernel will differ according to the number of Ac. (Table 5.3).

TABLE 5.3. Effect of number of *AC* elements on the time of *DS* dissociation leading to chromosome breakage and kernel colour in maize

Genotype of endosperm	Time of dissociation of DS	Degree of colour spotting in kernel
Iii - DS	No dissociation	Colourless kernel
Iii-DS AC	Early dissociation	Large colour spots
Iii- DS AC AC	Dissociation at the stage	Small colour spots
Iii – DS AC AC AC	Dissociation at very late stage	Very few tiny colour spots

Organization of Ac and Ds elements:

Activator (Ac):

Activator (Ac) element is 4563 bp long and is autonomous in action. It has 11 bp inverted repeats at its both ends (Fig. 5.11). The target site for Ac insertion is 8 bp long; this target sequence is duplicated during the insertion as direct repeats. The Ac element has 5 exons (Fig. 5.11); transcription produces and mRNA of 35 00 bases which has a coding sequence for 807 codons. This element has two open reading frames.

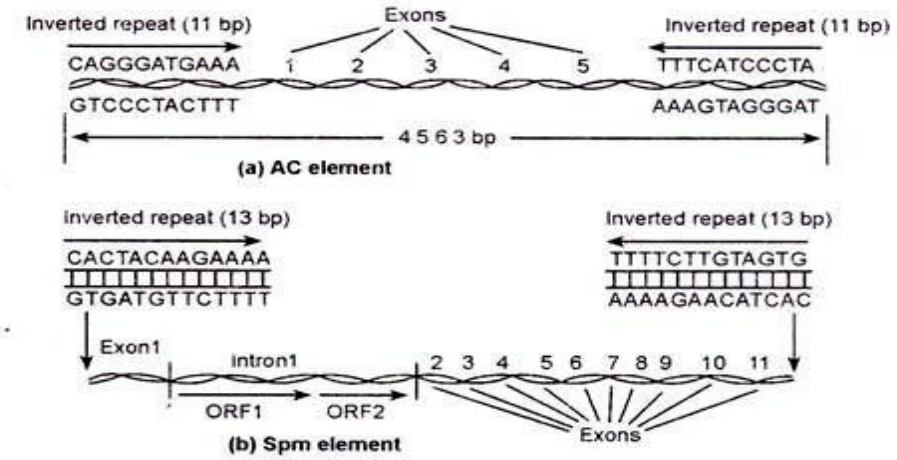


Fig. 5.11. Diagrammatic representation of *AC* and *Spm* elements in maize. Different *DS* elements are produced due to interstitial deletion in *AC*, while different *dSPm* elements are produced due to interstitial deletion in *Spm*.

Dissociation (Ds) element:

These elements are non-autonomous and are produced through interstitial deletions in the Ac element (Fig. 5.11). Based on the length and the region of deletion, Ds elements are grouped into several types as, Ds1, Ds2, Ds6, Ds9 Ds 2dl, and Ds2d2 etc. All the Ds elements contain the 11 bp inverted repeats at their ends. The Ds1 element represents an extreme

case in that it has a large interstitial deletion so that only the terminal 11bp inverted repeats are retained.

Ds6 element possesses 1000 bp from each end of the Ac, the rest portion being deleted. Ds9 on the other hand, represents a very short deletion of about 194 bp. Further changes may also occur in the non-autonomous elements leaving them incapable of transposition, i.e., they become permanently stabilized.

Autonomous elements may also be subject to changes. During the different developmental periods of an individual these elements may undergo cycles of active and inactive phases; the phase changes are brought about by methylation of their DNA. A methylation in the target sequence $\frac{CAG}{GTC}$ of an element leads to a reversible inactivation of the element.

Effects of transposition of Ds:

Transposition of Ds causes breakage in the chromosome at the site from which the Ds element moves out. The mechanism of transposition is non-replicative. Following breakage, the acentric chromosome fragment is lost.

If the chromosome carrying Ds has dominant alleles, e.g. A, B, C and its homologue carries the recessive alleles a, b, c, the transposition of Ds will lead to breakage and loss of the fragment carrying the dominant alleles (Fig. 5 12.). In the progeny cells as a result only the recessive alleles a, b, c will be expressed.

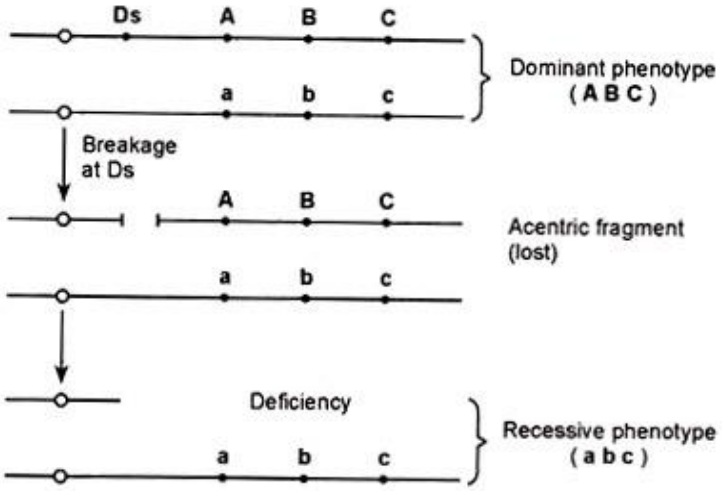


Fig. 5.12. Breakage at *DS* site produces an acentric fragment which is lost. As a consequence, the cell becomes deficient for genes *A*, *B* and *C* carried in the acentric fragment. Therefore, the expression of their recessive alleles present in the homologous chromosome is observable.

Alternatively, the broken ends of the two sister chromatids may join together as they are produced through chromosome replication. The acentric fragment, as a result, will form a

U- shaped structure which would be lost. The centric fragment, on the other hand, will form a dicentric chromatid bridge during anaphase. As a result of the tension created due to the centromere movement, the chromatid bridge will break at some point between them producing two dissimilar chromatids. In the next cycle, the broken chromatid ends will again fuse during chromosome replication.

This will, as earlier lead to the formation of dicentric chromatid bridges in both the daughter cells. Thus a chromatid-fusion-bridge-breakage cycle is generated (Fig. 5.13). In such a condition cells contain duplication for one or more genes along with deficiency for some other genes.

Duplication and deletion:

When the transposons are located on both the homologues but at different positions, pairing and recombination between them leads to the production of one deficient chromosome and one chromosome with duplication (Fig. 5.14).

Translocations caused by transposons:

Transposons located on non-homologous chromosomes can pair, and crossing over between them produces reciprocal translocations (Fig. 5.15).

Suppressor-Mutator (Spm) Elements:

Spm is nearly similar to the En (enhancer). It is a larger transposon than Ac and contains inverted repeats of 13bp at its ends. Its promoter is situated at the left and is responsible for the transcription of 8300bp of DNA. Spm is composed of two genes, tnpA and tnpB. The tnpA is a split gene containing 11 exons, the first intron being very long (Fig. 5.11). After splicing, a 2500 base mRNA is obtained which is translated into a protein of 621 amino acids. The first intron of the tnpA gene possesses two additional reading frames ORF l and ORF 2; both the reading frames are jointly called the tnpB gene. The proteins coded by tnpA and tnpB are required for several functions related to transposition. A deletion in the ORF 1 and ORF 2 regions produce defective Spm elements (dSpm).

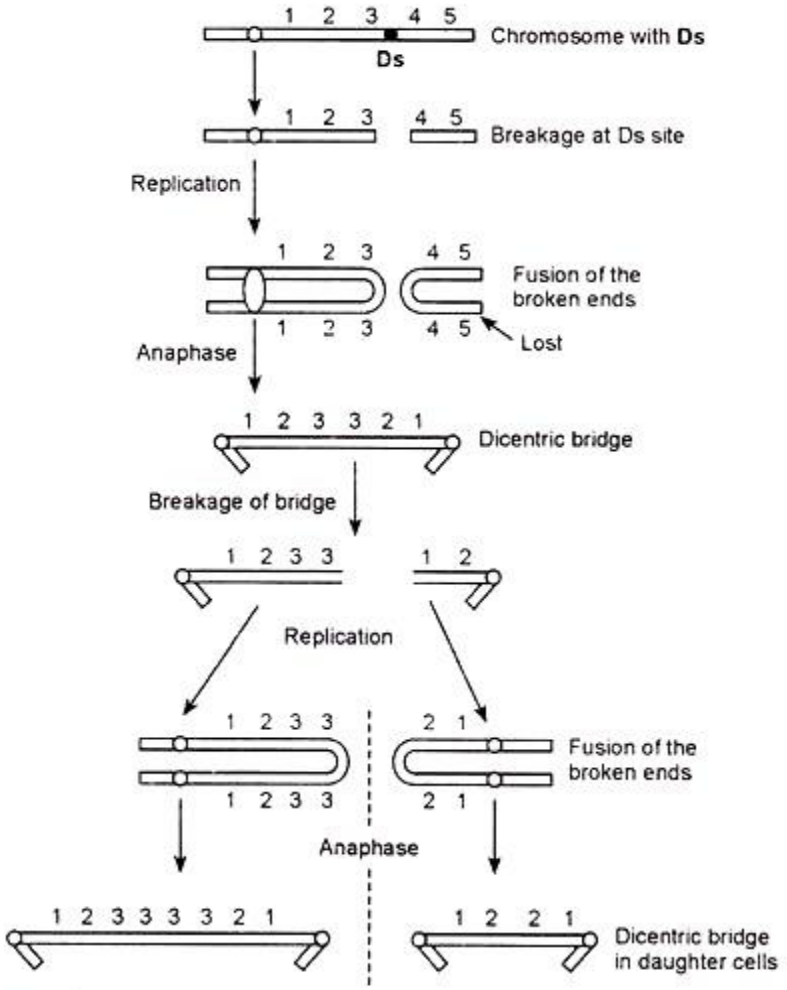


Fig. 5.13. Breakage-fusion-bridge cycle generated by chromosome breakage caused by the transposition of a *DS* element.

Transposition of the Spm into a gene completely inhibits the expression of the gene. But when dSpm insertion occurs within a gene the expression of the gene is reduced; such a gene is called dSpm- suppressive gene. Insertion of dSpm in the vicinity of a gene does not inhibit or reduce its expression but, on the other hand, enhances its expression. Therefore, such a gene is called "dSpm- dependent gene".

P Elements in Drosophila Melanogaster:

In Drosophila, certain strains when mated together produce hybrid dysgenesis (mutations, chromosome aberrations, distorted segregation at meiosis and sterility). In such crosses, the F, flies have normal somatic tissues but, their gonads do not develop.

A transposable genetic element known as “P element” or “P factor” has been found to be responsible for this condition. On the basis of the presence or the absence of the P factors, the flies are divided into two types.

I. P-type (Paternal Contributing):

Strains containing P elements in their chromosomes are designated as P-type. The number of P elements varies from 30 to 40 in the genome.

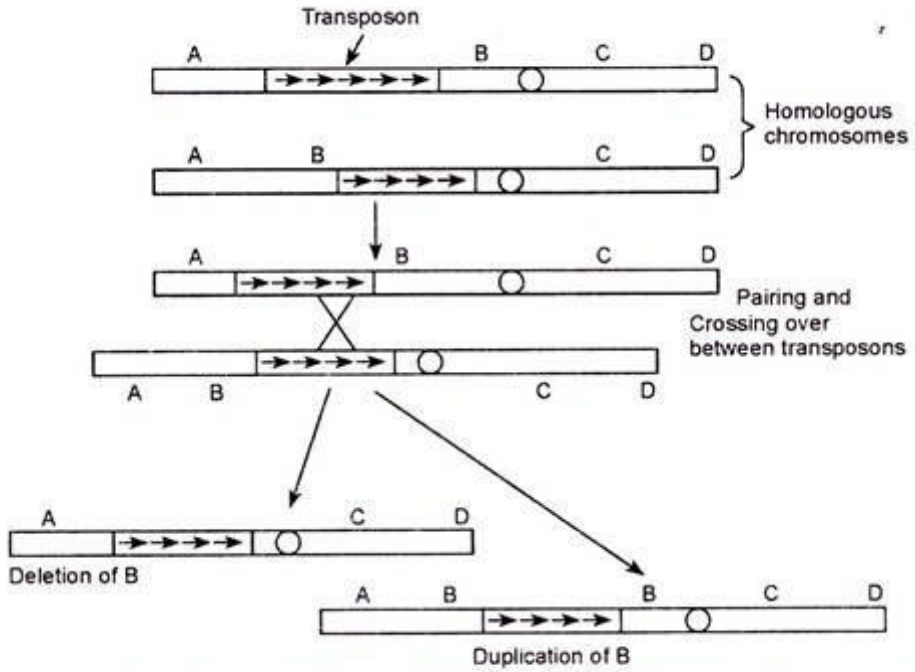


Fig. 5.14. Crossing over in the transposons located on homologous chromosomes produces deletion and duplication.

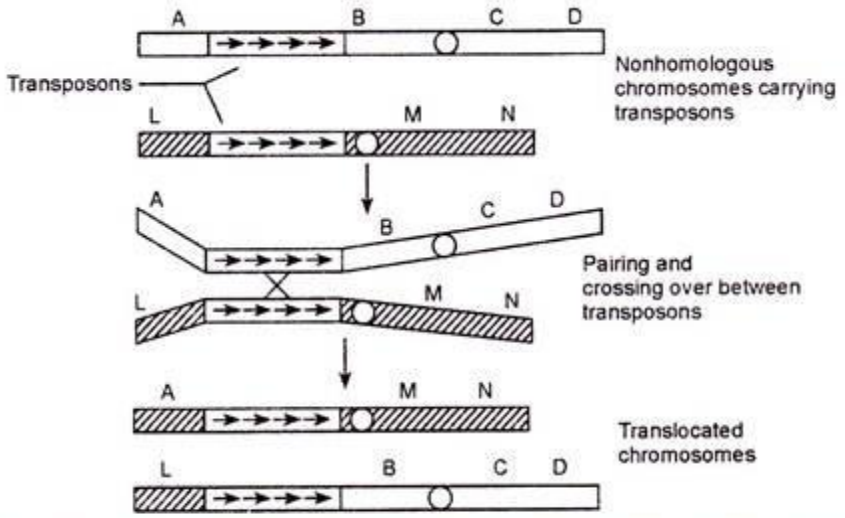


Fig. 5.15. Translocation caused by crossing over between the transposons located on non-homologous chromosomes.

P elements occupy several different positions on the chromosome. These factors are present on chromosomes as inactive component in the P-strain.

2. M-Type (Maternal Contributing):

M-strains do not contain P-elements in their chromosomes. Crosses involving P-Type male flies and M-type female flies produce F, flies showing hybrids dysgenesis. But when the female flies are P-type, the hybrids are normal irrespective of the male being P-type or M-type. (Fig. 5.16.).

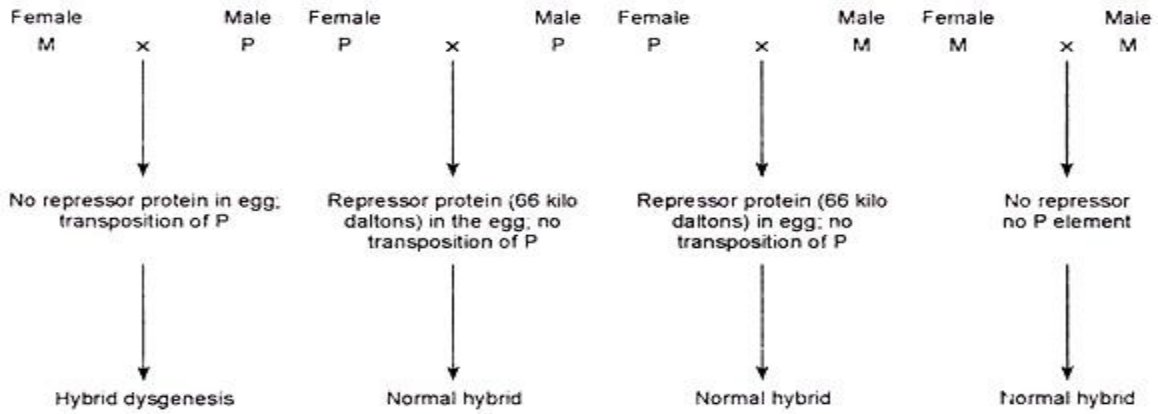


Fig. 5.16. Results of different types of P-M crosses in *Drosophila*. Hybrid dysgenesis occurs only when a P male is mated to an M-female. A 66 kilo dalton repressor protein is present in the egg cytoplasm of P-females : This repressor inhibits the transposition of P elements resulting into normal F_1 flies.

Organisation of P elements:

There are different types of the P element. The full size elements are about 3000 bp long and contain inverted repeats of 31 bp at their ends. Transposition of the P elements generates direct repeats of 8 bp on both sides of the site where transposition has taken place.

Interstitial deletions in P element produce different P elements of smaller sizes. Some of the small P element possess the gene coding for transposase while others do not have the complete gene. The latter type of P elements (which have an incomplete transposase gene) are activated by the transposase enzyme produced by another normal P element. The full size P element has 4 open reading frames designated as ORFO, ORF1, ORF2 and ORF3. The processing of the primary RNA transcript occurs by separate mechanisms in somatic and germinal tissues.

1. Processing in somatic tissues:

In somatic tissues, splicing occurs producing the mRNA which contains only three open reading frames, namely, ORFO, ORF1 and ORF2; the intron 3 is not spliced. Somatic cells contain a protein that binds to the intron 3, thus inhibiting the removal of this intron. This mRNA produces a 66,000 Dalton protein which functions as a repressor for transposition so that somatic tissues are not affected by P element.

2. Processing in the germ line:

In the germ line, the intron 3 binding protein is absent and therefore, all the introns are removed during processing of the primary transcript. The four reading frames ORFO, ORF1, ORF2 and ORF3 are joined together to produce a large mRNA. This mRNA produces the 87,000 Dalton protein (the enzyme transposase) which leads to transposition. Transposition occurs by the no-replicative mechanisms similar to that of Tn10 transposon. The enzyme transposase binds to 10 bp sequence adjacent to the 31 bp inverted repeats at ends of P. Transposition leaves a break at the original site of P and it produces gene mutation at the new (insertion) site of P. Both the events, therefore, generate adverse effects on the individual.

The P line contains P-cytotype, while the M line contains M-cytotype. When chromosomes bearing P factors come into the M-cytotype, transposition occurs leading to hybrid dysgenesis. However, when chromosomes carrying the P elements come into the P cytoplasm, there is no transposition. This can be explained as follows (Fig. 5.16).

A repressor protein called 66,000 Dalton protein is present in the egg cytoplasm of flies containing P elements. However, the repressor protein is absent from the egg cytoplasm of M- females. When a P-male is crossed to an M-female, in the F_1 their P elements present in the paternal chromosome undergo transposition due to the absence of the repressor; this produces hybrid dysgenesis. But in crosses involving P-females, the transposition of P

elements is prevented by the 66,000 Dalton repressor protein present in the egg cytoplasm; this yields normal fertile hybrids.

Salmonella Phage variation:

There are over 1500 recognized Salmonella enterica ssp. enterica serovars, and these are responsible for over 98% of human clinical Salmonella infections. Infection mostly occurs as a result of contaminated water and food and is facilitated by animal reservoirs that for many serovars include cattle and poultry. Combating this problem will require a more complete understanding of the basis of serovar diversity and of virulence and persistence in both humans and animal reservoirs. In addition, typing of isolates is an essential tool in Salmonella epidemiology and management strategies. New molecular approaches are being sought for diagnostic and epidemiological analyses, including ones based on specific genome sequences, and how to relate these to the classical serotyping schemes.

Salmonella serotyping is based on the White-Kaufmann-Le Minor scheme, which is a modification of the original scheme from the 1930s (Salmonella Subcommittee of the Nomenclature Committee of the International Society of Microbiology, 1934; Grimont and Weill, 2007; Guibourdenche et al., 2010). Serotyping is mainly based on agglutination with specific sera to identify antigenic variants of the flagellar antigen (H factor) and the O-antigen of the lipopolysaccharide (LPS), which defines the O factor. The lipid A tail and the core polysaccharide of Salmonella LPS have little structural or compositional variation compared with the high degree of variability in the O-antigen. The correlations between serotype and the chemistry of the corresponding O-antigen is based on a large body of work on biochemical analysis of purified LPS (Luderitz et al., 1966; Hellerqvist et al., 1969; Knirel and Kochetkov, 1994; Raetz and Whitfield, 2002; Wang et al., 2002; Guibourdenche et al., 2010). O-antigens can differ both in the composition of the polysaccharides of the repeating units and the linkage between the individual sugar moieties. Further variation of Salmonella O-antigen composition can occur by modification of these repeating units, specifically by linkage of an acetyl group (Hellerqvist et al., 1969; Slauch et al., 1996) or glucose moiety (Reeves, 1994; Guibourdenche et al., 2010). The recipient moiety and the chemical linkage for these modifications can also vary. Some of these variable O-antigen modifications are recognized in the serotyping scheme. Together, these variables contribute to the large number of Salmonella serovars.

A generic model for the biochemical pathway of O-antigen glucosylation is based on studies from the 1970s identifying biochemical intermediates in Salmonella (Nikaido et al., 1971; Wright, 1971), supplemented with genetic studies on Shigella flexneri (Lehane et al., 2005; Korres and Verma, 2006). A glucosyltransferase (gtr) gene cluster, consisting of three genes, is required for O-antigen glucosylation [reviewed in Allison and Verma (2000)]. The

gtrA and gtrB genes are predicted to encode membrane proteins, for the bactoprenol-linked glucosyl translocase or 'flippase' and the bactoprenol glucosyl transferase respectively. The third, variable, gene in the cluster is referred to generically as gtrC and encodes the glucosyltransferase that mediates the attachment of the glucose group to the O-antigen. This is specific for each gtr operon and it is this variable gene product that determines the attachment residue in the O-antigen and the nature of the linkage, and thus defines the serotype-specific modification associated with each gtr operon (Lehane et al., 2005; Korres and Verma, 2006).

Genetic determinants for a few of the Salmonella serotypes that depend on O-antigen modification have been identified. The O1 serotype arises as a result of lysogenization by the temperate phage P22 of S. enterica subspecies enterica serovar Typhimurium (Salmonella Typhimurium). Seroconversion from O:4,5,12 to O:4,5,12,1 (originally designated 'antigen 1') is mediated by a P22-genome encoded gtrABC gene cluster (previously designated 'con' or 'a1') (Fukazawa and Hartman, 1964; Van der Byl and Kropinski, 2000) and is the result of the addition of a glucose group to the galactose moiety of the O-antigen (Makela, 1973; Grimont and Weill, 2007). A modification that is associated with the O12 subtype 2 (O122) serotype was recently attributed to the gtr cluster STM0557-0559 on the S. Typhimurium genome (Bogomolnaya et al., 2008). Other gtr gene clusters are found in a significant number of P22-like Salmonella phage and on Salmonella genomes (Allison and Verma, 2000; Vernikos and Parkhill, 2006; Villafane et al., 2008) (M. Davies, unpublished). For the majority of these, the modification reactions have not been elucidated.

The modification of the O-antigen may contribute to immune evasion, as indicated by seroconversion, but it has also been implicated in other bacteria–host interactions. Specifically, the O122 modification within a S. Typhimurium mouse model for infection may facilitate gut persistence (Bogomolnaya et al., 2008). These roles of O-antigen modification and its role in serotyping, along with the prevalence of the gtr gene clusters, indicate a necessity to understand the modification process and the regulation of gtr expression in Salmonella. Studies from the 1940s acknowledged that expression of Salmonella serotype-specific O-antigens may not be uniform among colonies of a given isolate, stating that this 'undermines the theoretical basis of serological standardization of Salmonella O-suspensions … ' (Kauffmann, 1941; Hayes, 1947) and quantitative variation was recognized in the biochemical studies on O-antigen composition from the 1970s (Nikaido et al., 1971). The prevalence and molecular basis of this variation for different O-factors is not known, but we hypothesized that this may reflect what is currently termed 'phase variation'. This is a reversible yet heritable form of gene regulation that results in heterogeneous clonal populations and can be mediated by a variety of molecular mechanisms (van der Woude and Baumler, 2004).

Retroposons (Retro-transposons):

Retroposons are transposable genetic elements which are mobilized through an RNA form. The DNA element is transcribed into RNA and then the RNA is copied by the enzyme reverse transcriptase into DNA which is inserted at a new site into the host genome. Retroposons include processed pseudo genes, small RNA pseudo genes (SnRNA) and Alu family in primates and rodents.

Some of the eukaryotic transposons are related to retroviral proviruses and mobilize through RNA intermediates. Retroposons differ from retroviruses in the sense that they do not pass through an independent infectious form. However, they do use the reverse transcription process to produce DNA. In order to understand retroposons, it is desirable to study the life cycle of a retrovirus.

Retroviruses:

They are infectious viruses containing single-stranded RNA (+ strand) which infect eukaryotic cells. Through reverse transcription, the viral RNA produces a complementary DNA (-) strand. The enzyme reverse transcriptase has exonuclease activity (RNAase H) by which it degrades the RNA strand from the RNA-DNA hybrid so produced. The same enzyme also synthesizes, by its polymerase activity, the complementary DNA strand (+ strand) to (-) strand.

This double-stranded DNA moves to the nucleus of the cell where one or more copies of it become integrated into the host genome; the enzyme involved in the insertion is integrase. This viral DNA genome integrated into eukaryotic host chromosome is called a provirus or retroposon. It remains as a endogenous provirus in the germ line. In other cells, the pro-viral DNA is transcribed to produce RNAs which function as (i) viral genome and (ii) mRNA to produce proteins that are structural components of the retrovirus.

In every viral particle, 2 copies of RNAs are packaged, making it a diploid virion. When two different retroviruses infect a single cell, the new viral particle may contain one chromosome from each of the two viruses; thus some virions may be heterozygous.

Organization of retroviral RNA and pro-viral DNA (Retroposon):

The retroviral RNA has direct repeats (R) varying from 10 to 80 nucleotides at its both ends (Fig. 5.17). An 80-100 base long unique region (U5) lies next to the R segment at the 5′-end. Similarly, at the 3′-end, left to the R segment, there occurs a unique region (U3) containing 170- 1260 nucleotides.

The coding region of the virus contains the genes gag (2000 bases), pol (2900 bases) and env (1800 bases). The reverse-transcribed DNA has a long terminal repeat (LTR) that s composed of the sequences “U3-R-U5” at both the ends (Fig. 5.17).

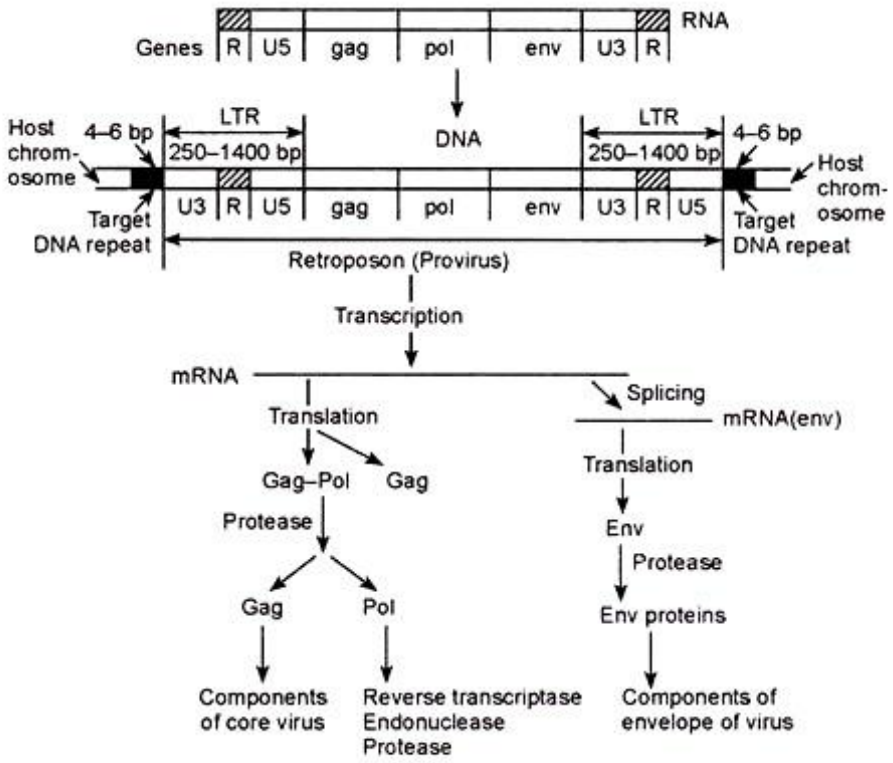


Fig. 5.17. Retroviral RNA, proviral DNA and the synthesis of polyproteins. Retroviral RNA has ends of short direct repeats. Genes are expressed as polyproteins that are processed into individual proteins. Number of bases in different regions of the retroviral RNA are 10-80 in R 170-1260 in U3, 80-100 in U5, ~2000 in *gag*, ~2900 in *pol* and ~1800 in ***env***.

Integration:

Integration of DNA into the host chromosome occurs through the linear form of DNA. Integration events are similar to those of transposable elements. The enzyme integrase makes staggered cuts at the target site which may be 4-6 bp in length. Direct repeats of the target DNA is produced during the integration. During this process, the U3 sequence loses 2 bp from left end and the U5 sequence loses 2 bp from the right end. A single cell receives up to 10 copies of a provirus.

Expression of viral genes:

The coding region of retroviruses consists of 3-4 genes, such as, gag pol, env sequences. Transcription of provirus produces the genomic RNA from which env mRNA is obtained by splicing (Fig. 5.17). Translation of the genomic RNA yields Gag, Gag-Pol and Env poly-proteins. Specific proteases cleave the poly-proteins into individual proteins through processing. After about 8 hour of infection, the poly-proteins together with viral genomic RNA begin to assemble under the plasma membrane. They attract the envelope proteins already present in the membrane. Nucleocapsid is formed by protein cleavages. A segment of the host cell membrane is pinched off (like budding) and viral particle is released. During

the process of infection, the viral particle becomes fused with the plasma membrane as a result of which, the RNA is released into the cell.

Transducing viral particles:

A retrovirus may carry genes from its host cell. This occurs when a deletion in the provirus (retroposon) occurs, thus fusing the viral and host genes. As shown in the Figure 5.18, the deletion of a part of the pro-viral DNA causes the transcription of DNA containing both, pro-viral DNA and the host DNA to produce a "fused mRNA."

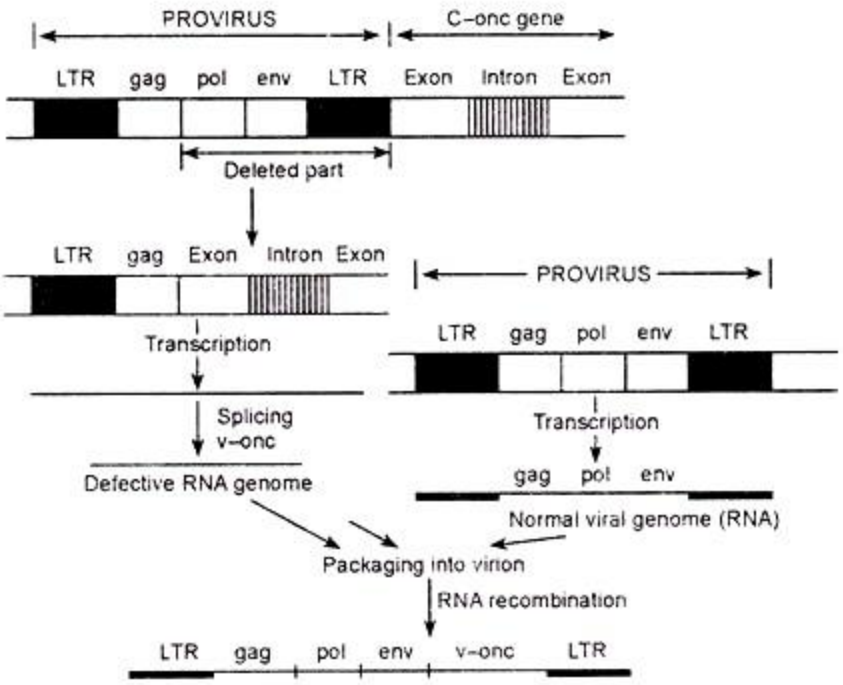


Fig. 5.18. Deletion in proviral DNA may produce fused mRNA containing some viral and cellular genes; this is defective viral RNA. A helper viral RNA produces the viral capsid and both the RNA genomes may become included in a single virion. Nonhomologous recombination will produce an RNA genome which contains both viral and host genes flanked by LTR sequences.

After splicing of the host mRNA the fused mRNA becomes shorter. In some cases, the c-onc gene of the host may be transcribed and fused with the viral mRNA. But this RNA is defective and cannot produce new virus. If the cell contains some normal provirus, it acts as a helper.

Some of the viral particles produced in such a way will contain one fused (defective) and one normal viral genome. Recombination between the two RNAs will produce an RNA genome that contains LTR along with the viral genes and host genes.

The c-onc gene is called v-onc gene when present in the viral genome. The properties of the host cell are drastically changed when it is infected by such a viral particle; it becomes a cancerous cell.

Retroposon-like Elements in Eukaryotes:

Retroposons or retro-transposon like elements are found in different eukaryotic organisms, such as, yeast, Drosophila, and mammals including human. These elements are classified into the following two groups.

I. Viral super family:

The retroposons that code for reverse transcriptase and integrase, and possess the ability of transposition belong to this family. They have long terminal repeats. Many retroposons also contain introns. They generate direct repeats of 4-6 bases in the target DNA. Examples of such elements are: Ty elements in yeast, copia in Drosophila and LINES LI in mammals.

II. Non-viral super family:

The retroposons belonging to this family do not code for proteins that have role in transposition. They are believed to have originated from RNA sequences through the process of reverse transcription; they do not contain either terminal repeats or introns. They generate direct repeats of target DNA containing 7-21 bp. Examples are SINES B1/Alu family in mammals, processed pseudo genes, transcript of RNA polymerase II.

Ty Elements in Yeast (Saccharomyces Cerevisiae):

Ty (Transposon yeast) elements are of divergent types and made a family of dispersed repetitive sequences on yeast genome. These elements are 63000 bases long and are grouped into two main classes, Ty1 and Ty917. A typical yeast genome contains about 30 copies of Ty] type and about 6 copies of Ty917 type elements. They have direct repeats of 350 bp at each end; these repeats are called delta (8).

A Ty element has two open reading frames TyA and TyB. The TyA protein represents the TyA reading frame, while the TyB protein represents the joint TyA and TyB regions. The TyA region codes for DNA-binding proteins, while TyB codes for reverse transcriptase, protease and integrase (like retrovirus). Ty elements are mobilized through an RNA intermediate and transposition is controlled by its genes. The element behaves like a retrovirus which has lost the coding region for the viral envelope.

Copia Elements in Drosophila:

The term copia denotes a large number of closely related sequences in Drosophila. The number of these retroposons per genome varies from 20 to 60. The copia element is 5146 bp long with terminal direct repeats of 276 bp and terminal inverted repeats of 13 bp.

It generates direct repeats of 5 bp in the target DNA at the site of insertion. Copia elements are dispersed and take different locations in different strains of Drosophila.

Sometimes copia elements are found as circular molecules of 5000 bp and 4700 bp in length. They contain a single reading frame of 4227 bp which shows homologous relationship with gag and pol sequences of retrovirus but the env sequence is absent. Therefore, copia cannot produce a virus particle.

Retroposon-Like Elements in Mammals:

In mammals, a large part of the repetitive DNA consists of retroposons. These are two main groups of these elements, called LINES and SINES. The LINES sequences are also called LI. They are long interspersed sequences dispersed in the genome. The average length of LINES is 6.5 kb. At their end, they contain sequences rich in adenine (A) nucleotides. LINES sequences are derived from the transcripts of RNA polymerase II. The number of copies of LINES ranges from 20,000 to 50,000 per mammalian genome.

Short interspersed sequences in mammalian genomes are called SINES. They are derived from the transcripts of RNA polymerase III. These elements do not posses a coding region, and are about 300 bp in length. Probably they originated from a transposition event like retroviruses and the RNA was copied into DNA by reverse transcriptase. The SINES family includes the "Alu- family". Alu family is a set of dispersed, related sequences (about 300 bp long) formed in human genome. Individual sequences have Alu cleavage sites at each end. There are about 300,000 (3×10^5) Alu sequences dispersed in the haploid human genome. The Alu sequences are flanked ' by short direct repeats indicating their resemblance to transposons.

Transposons in Humans:

Transposons in humans are in the form of repetitive DNA which consists of sequences that are interspersed within the entire human genome. These sequences are transposable and can move to different locations within the genome.

These are of following two types:

(1) SINEs (Short Interspersed Elements):

They are ~ 300 bp long and may be present about 5 lakh times in human genome. Alu sequences are the best characterized SINEs in humans.

These are termed as **'Alu'** elements because they contain specific nucleotide sequences which are cleaved by the restriction enzyme named Alul. Alu elements contain Direct Terminal Repeats (DTR) of 7-20 bp length. These DTRs help them in the insertion process during transposition.

(2) LINEs (Long Interspersed Elements):

They are ~ 6400 bp long and are present about 1 lakh times in the human genome. Most prominent example is LI sequence. These transposable elements are some of the most abundant and common families of moderately repeated sequences in human DNA.

Significance of Transposable Elements:

1. Transposons may change the structural and functional characteristics of genome by changing their position in the genome.
2. Transposable elements cause mutation by insertion, deletion, etc.
3. Transposons make positive contribution in evolution as they have tremendous impact on the alteration of genetic organisation of organisms.
4. They are useful as cloning vectors also, in gene cloning. For example, P-elements are frequently used as vector for introducing transgenes into Drosophila.
5. Transposons may also be used as genetic markers while mapping the genomes.
6. Transposon-mediated gene tagging is done for searching and isolation of a particular gene.

Probable Questions:

1. What is P element in Drosophila.
2. Describe copia elements of Drosophila.
3. Describe Ty Elements in Yeast.
4. Discuss transposons present in human.
5. What are the significance of Transposable genetic elements?
6. Briefly discuss about Salmonella phage variation.
7. What is Ac/Ds element? Describe its mechanism of transposition

Suggested Readings:

1. Molecular Cell Biology by Lodish, Fourth Edition.
2. The Cell – A Molecular Approach by Cooper and Hausman, Fourth Edition
3. Principles of Genetics by Snustad and Simmons, Sixth Edition.
4. Molecular Biology of the Cell – by Bruce Alberts
5. Cell and Molecular Biology by Gerald Karp, 7th Edition.

Unit-XI

Extra-nuclear inheritance: Streptomycin resistance in *Chlamydomonus*

Objective: In this unit we will discuss about extra nuclear inheritance such as Streptomycin resistance in *Chlamydomonus*

Cytoplasmic Inheritance:

The existence of genes as segments of nucleic acid molecules, located in chromosome of nucleus, has been demonstrated by several experiments. The nuclear genes control the phenotypes of the organisms and are concerned with the transmission of hereditary character from one generation to next generation is known and predictable Mendelian fashion.

The inheritance of genes of nuclear chromosomes is characterised by the fact that the genes from male and female parents contribute equally to the genetic constitution of the offspring. Therefore, in it the reciprocal crosses between parents of different homozygous genotype will produce offspring's of identical phenotypes except for sex-linked genes. However, in certain cases, although male and female parents contribute equally their nuclear genes to the offspring's, the results show a non-Mendelian inheritance pattern and the result of reciprocal crosses varies.

These variations suggest that the genes for the inheritance of certain characters do not occur within the nucleus, but they are present ill cytoplasm and play an important role in transmission of certain specific traits, which are not controlled by nuclear genes. Therefore, it builds up the concept of cytoplasmic inheritance. The genes for cytoplasmic inheritance are independent, self-replicating nucleic acids.

Evidence for cytoplasmic inheritance was first reported by Correns in *Mirabilis jalapa* and by Bar in Pelargonium zonule in 1908. Rhoades described cytoplasmic male sterility in maize in 1933. In 1943, Sonneborn discovered kappa particles in *Paramoecium* and described its cytoplasmic inheritance. Presence of DNA in chloroplasts was first demonstrated by Ris in plant cell. In 1963, Nass and his co-workers proved the existence of DNA in mitochondria. Subsequently, from time to time, observations by several scientists have been reported the important role of cytoplasm in genetics. Thus, on the basis of observations made on cytoplasmic inheritance of some specific traits, it has been suggested that cytoplasm is also genetically active.

Terms and Definitions of Cytoplasmic Inheritance:

Extra-chromosomal inheritance, extra-nuclear inheritance, somal inheritance and maternal inheritance are all synonyms. All these terms can be defined as the inheritance of characteristics of only one of the two parents, usually the female parent to the progeny. The reciprocal crosses show consistent differences as well as there is a lack of segregation in F_2 and subsequent generations. The genes controlling cytoplasmic inheritance are present outside the nucleus and,

in the cytoplasm, they are known as plasma genes, cytoplasmic genes, cytogeneses, extra nuclear genes or extra chromosomal genes. The sum total of the genes present in cytoplasm of a cell is known as Plasmon. All the genes present in a plastid are known as plastoms. Similarly, all the genes present in a mitochondrion are known as chondrioms. The genes present in plastid and in mitochondrion are located in their own DNAs and are known as cp DNA and mtDNA, respectively. These DNAs are collectively termed organelle DNA.

Characteristics and Detection of Cytoplasmic Inheritance:

Cytoplasmic inheritances do not show Mendelian inheritance.

They show the following characteristic features:

i. Hereditary traits which are transmitted by cytoplasm do not show Mendelian segregation in crosses and in reciprocal crosses with respect to a particular set of characteristics controlled by a set of cytoplasmic genes produce dissimilar hybrids.

ii. Most of the recorded cytoplasmically inherited characteristics would follow the maternal line, i.e., uniparental mode of transmission. In higher plants and animals, ovum or egg cell is comparatively large and contains large amount of cytoplasm. But male gametes or sperms have very little amount of cytoplasm. So, under this situation, most of cytoplasmic factors are transmitted to the progeny through the ovum of mother.

It is known as maternal inheritance or trans-ovarian transmission. In this mode of transmission, all the offspring's of the parents have maternal condition and only female progeny can transmit the cytoplasmic characteristics to the succeeding generations. Hence the reciprocal crosses yield different or non-Mendelian results.

Characteristics of Mendelian Inheritance:

The inheritance pattern of characters of an organism as proposed by Mendel on the basis of monohybrid and di-hybrid crosses is referred to as Mendelian inheritance.

It shows the following characteristic features:

i. Contribution of both male and female is equal, hence results from reciprocal crosses are similar.

ii. Segregation produces the phenotypes ratio 3 : 1 and genotype ratio 1 : 2 : 1 in the F_2 generation of a monohybrid cross and a typical phenotype ratio 9 : 3 : 3 : 1 in di-hybrid crosses.

Mendelian inheritance pattern is regarded as a sufficient evidence for a gene to be located in chromosomes; such genes are called nuclear genes or simply as genes.

Maternal Inheritance*:*

Maternal inheritance means the inheritance controlled by extra-chromosomal, i.e., cytoplasmic, factors that are transmitted to the succeeding generation through the egg of female organism.

They show the following features:

i. reciprocal differences in F_1;
ii. which in most cases disappears in F_2;
iii. a smaller variation in F_2 as compared to that in F_3.

Maternal inheritance may be, broadly speaking, of two kinds:

i. If some treatments (chemical poison, heat shock etc.) are applied to the female parent, it may affect the egg's cytoplasm. As a result subsequent offspring's are modified in some way. Effects of this kind are called Dauer-modifications or persisting modifications.
It is observed that when protozoa are treated experimentally with chemical poisons or heat shocks, the treatments induce several morphological abnormalities in them. Such abnormalities go on decreasing generation after generation and, eventually, disappear completely through cell division if the treatments are removed. Further evidences also come from fruit flies subjected to heat treatment and from bacteria treated with chemicals.

ii. Other kinds of maternal inheritance are also known which do not depend upon the repeated application of an external stimulus to the cytoplasm. In this case, maternal inheritance is truly controlled by independent cytoplasmic genes. Maternal effects reflect the influence of the mother's gene on developing tissues. Many important characteristics of both animal and plants show maternal effects of which some examples axe described next.

Difference between Nuclear Traits and Extra-Nuclear Traits:

1. Reciprocal Differences:

Differences in the results of reciprocal crosses would suggest a deviation from the pattern of Mendelian autosomal gene transmission. According to Mendelian inheritance, the chromosome complement in male and female gametes obtained from the same species would be similar; reciprocal crosses should give same results (♀A x♂ B =♀B x ♀'A). The only exception to this expectation is sex- linked inheritance which can be explained on the basis of transmission of sex chromosomes.
If sex linkage is ruled out, differences in the result of reciprocal crosses would indicate that one parent (maternal) is exerting a greater influence than the other on a particular trait. This is because cytoplasm does not divide in a precise manner like the chromosomes during the process of cell division during gametogenesis. Female gametes usually contribute more cytoplasm to the zygote.

Consequently for characters having cytoplasmic control, differences in reciprocal crosses are observed. As shown in the Fig. 10.1, if two strains A and B respectively having genotypes AA and BB and cytoplasm's a and b are crossed reciprocally, we will get two hybrids AB (a) and AB (b) [cytoplasm is indicated in parentheses]. In case of maternal inheritance, AB (a) and AB (b), despite having same nuclear genotype, will differ. AB (a) will resemble strain A or AA (a) and AB (b) will resemble strain B or BB (b). Since such effects are solely produced by cytoplasm of the egg, they are described as maternal inheritance (uniparental inheritance).

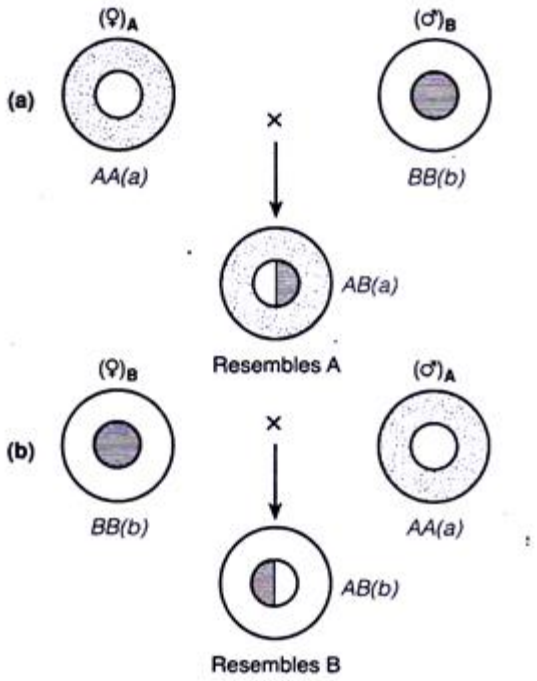


Fig. 10.1: Maternal inheritance showing difference in reciprocal crosses, (a) ♀ A × ♂ B, (b) ♀ B × ♂ A

2. Lack of Segregation:

Irregular Segregation; Somatic Segregation; Lack of Mendelian segregation and characteristic Mendelian ratios that depend on chromosomal transmission in meiosis would suggest extra-chromosomal transmission. Cytoplasmic genes inherited from both the parents sometimes give rise to irregular segregation ratios. They generally show somatic segregation during mitosis, a feature uncommon to nuclear genes.

3. Lack of Chromosomal Location:

The chromosomal genes occupy particular loci and specifically linked to other genes. The failure to find linkage to known nuclear genes may rule out chromosomal inheritance and suggests extra-nuclear inheritance.

4. Association with Organellar DNA:

The cytoplasmic inheritance or extra-chromosomal inheritance is defined as non-Mendelian inheritance, usually involving DNA in replicating cytoplasmic organelles, such as mitochondria and chloroplastids. The presence of DNA in cell- organelles, found outside the nucleus, is a strong evidence to suggest that genetic information does exist in cytoplasm also.

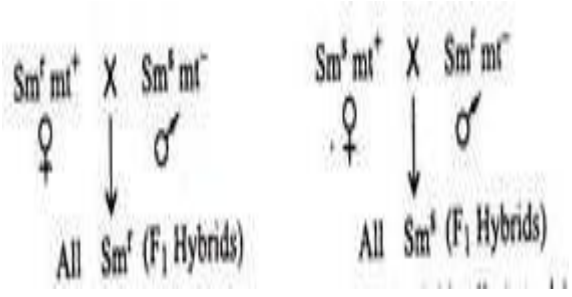
5. Transfer of Nuclear Genome through Backcrosses:

The transfer of nucleus of a variety to the cytoplasm' of another variety through repeated backcrossing, results in lines having nucleus and cytoplasm from two different varieties. A comparison of these lines with original lines having nucleus and cytoplasm of same variety demonstrates cytoplasmic effects on these traits.

Streptomycin Resistance in Chlamydomonas:

Ruthsagar (1965) has reported some cases of extrachromosomal inheritance in green alga *Chlamydomonas reinhardi*. The alga reproduces by asexual as well as sexual means. It does not have different sexes but has positive and negative strains or mating types (mt^+ and mt). The sexual reproduction involves fusion between two morphologically similar but physiologically dissimilar gametes of two different mating types (+ strain or mt^+ and – strain or mt^-) and the gametic fusion results in zygote. The sex is determined by a single chromosomal gene. When meiosis occurs in zygote, four haploid daughter protoplasts are formed which give rise to new plants. Out of 4 new plants resulted from a zygote two are of + strain and the other two are of negative (-) strain. Although both the sexes contribute equally to the zygote, there is maternal transmission of certain cytoplasmic traits. Ruthsagar (1965) isolated two strains of *Chlamydomonas*: one strain was resistant (Sm^r) to 500 jig of streptomycin per ml. of culture solution and the other was sensitive (Sm^s).

When the reciprocal crosses were made between the streptomycin resistant (Sm^r) and streptomycin sensitive (Sm^s) strains, the following results were obtained:



The diploid cells undergo meiosis and give rise to four haploid cells (tetrads) as shown in Fig. 18.3.

From these crosses the following two inferences with respect to Sm resistance can be drawn:

1. The F_1, reciprocal crosses differ from each other.
2. The phenotype of F_1, is governed by mt^+ strain i.e., it is maternal inheritance.

Using the analogy of higher organisms, mt^+ is referred to as female and mt^- as male. The mating type genes mt^+ and mt^- segregate in 1: 1 ratio as expected for the Mendelian inheritance. In the higher organisms formation of zygote involves fusion between an egg and a sperm and the contribution of cytoplasm to the zygote by sperm is negligible. Under such condition it is easy to comprehend the mechanism of maternal inheritance.

But in Chlamydomonas, male (mt^-) and female (mt^+) gametes being identical in size contribute equal amount of cytoplasm to the zygote, even then the cytoplasmic features of only mt^+ strain is expressed in F_1, i.e., it is uniparental inheritance. Now the question arises, what happens to the cytoplasmic determinants of mt^- gametes. This problem was solved by ruthsagar who discovered that the chloroplast DNA of mt^- strain becomes degraded in zygote and the mt^+ gene or a gene closely associated to it specifies a restriction-modification system. Here restriction implies degradation and the modification means protection.

The system encoding the DNA modifying enzyme, modifies its own DNA which cannot be degraded by the restriction system. The mt^+ chloroplast DNA which is not modified or protected is degraded by restriction system of mt^+ gamete. It is suggested that mt^+ linked gene encodes an endonuclease enzyme (enzyme which degrades DNA) which differentiates plastid DNA of its own cell from that of mt^- cell or digested due to modification.

Since plastid DNA of mt^- strain is degraded after sexual union, no expression of mt^- chloroplast DNA is possible and hence uniparental pattern of inheritance is observed. In rare cases (one in a thousand), however, the plastid DNA of mt^- cell escapes degradation by restriction system of mt^+ cell and the zygote contains the plastid DNAs from both the cells. Such a zygote is referred to as cytohet or cytoplasmic heterozygote. Cytohets are important from the view point of studying the recombination of cytoplasmic genes. Ruthsagar has constructed the genetic map of *Chlamydomonas* chloroplast DNA using the genie analysis of cytohets. This suggests that the inheritance of streptomycin resistance is uniparental and the factor for streptomycin resistance resides in the cytoplasm of + strain or mt^+ (Fig 18.3).

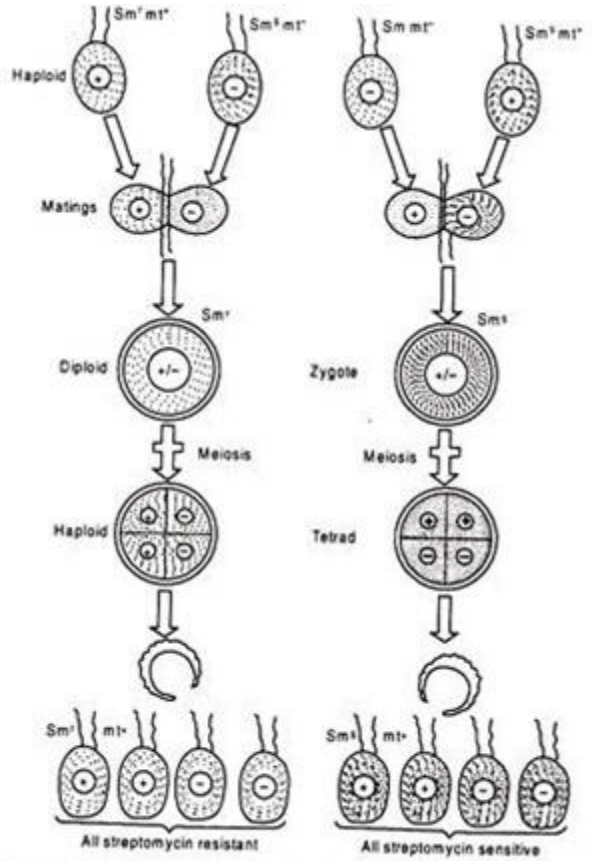


Fig. 18.3 Inheritance of resistance to streptomycin. The plus and minus signs refer to mating types (mt) which is inherited as a single gene difference. The progeny is with rare exceptions always like the plus parent in its reaction to streptomycin, but in these crosses the mating type difference segregates in every tetrad.

Probable Questions:

1. What is cytoplasmic inheritance? How it differs from nuclear inheritance?
2. What are the characteristics of cytoplasmic inheritance?
3. What are the differences between Nuclear Traits and Extra-Nuclear Traits?
4. How streptomycin resistance occur in *Chlamydomonas*?

Suggested Readings:

1. Molecular Cell Biology by Lodish, Fourth Edition.
2. The Cell – A Molecular Approach by Cooper and Hausman, Fourth Edition
3. Principles of Genetics by Snustad and Simmons, Sixth Edition.
4. Molecular Biology of the Cell – by Bruce Alberts
5. Cell and Molecular Biology by Gerald Karp, 7th Edition

UNIT-XII

Extra Nuclear inheritance: Kappa particles; criteria for extra-chromosomal inheritance.

Objective: In this unit we will discuss about Kappa and other extra nuclear inheritance particles of *Paramoecium* and will also discuss about criteria for extra chromosomal inheritance.

Kappa Particles in *Paramoecium*:

One of the most striking and spectacular example of cytoplasmic inheritance due to symbiont bacteria is noted in the most common ciliate protozoa *Paramoecium aurelia*. In 1943, T. M. Sonneborn reported that some strains of P. aurelia contain kappa particles and are known as killer strain.

Kappa particles are the symbiont bacteria called *Caedobacter taeniospiralis*. The diameter of kappa particles are about 0.2µ. They are bounded by a membrane and contain a little bit of cyto-plasm with DNA. The strain of *Parameocium* in which the kappa particles are absent are called sensitive strain. The sensitive strains are killed by the killer strain. The destruction of sensitive strain occurs through secretion of a toxic substance called paramecin. This toxic substance is believed to breakdown the food vacuole membrane of the sensitive strain. Paramecin is diffusible in the liquid medium (Fig. 22.8). When killers are allowed to remain in a medium for a time, they are not killed. It means that paramecin has no effect on killers. Paramecin is associated with a particular kind of kappa that occurs in about 20 percent of a kappa population.

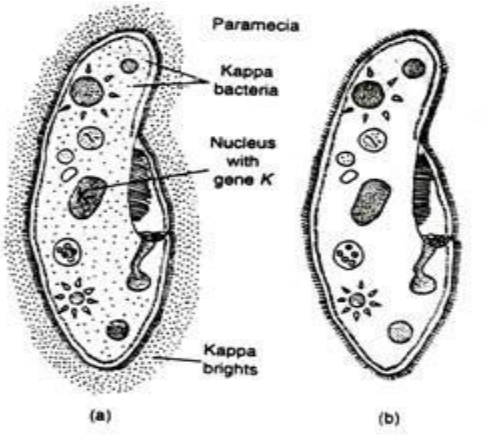


Fig. 22.8: (a) Killer strain Paramoecium with Kappa particles and nucleus with gene K; (b) Sensitive Paramoecium with no Kappa particles and nucleus with gene k.

These kappa bacteria possess a refractile protein containing 'R' body and are called brights because they are infected with a virus that controls the synthesis of a viral protein as well as R protein body in kappa bacterium. The virus may act as the toxin in the killing response and R body facilitates the penetration of the toxin. The non-bright kappa bacteria may also contain virus but the virus may be in provirus state in them.

The killer character of *Paramoecium* has a nuclear as well as cytoplasmic basis. The existence of kappa particles is determined by presence of a nuclear dominant gene K. Kappa particles, like other bacteria, multiply through fission. But their multiplication in the cytoplasm of *Paramoecium* depends on the presence of a dominant nuclear gene K which helps to make an environment necessary for the bacteria to reproduce. When killer strain of *Paramoecium* conjugates with sensitive strain under appropriate condition for brief period and no cytoplasm exchange occurs, two kinds of clones result- one from the original killer cell which contains allele K (Kk) and kappa particles and the other from the original sensitive cell which carries the allele K (Kk) and lacks kappa particles. It indicates that homozygous (either KK or kk) strains become heterozygous following an exchange of K and k genes without cytoplasmic exchange.
Following autogamy (a process of self- fertilisation within one undivided cell resulting in homozygosity), half the progeny (50%) are sensitive *Paramecia*. But all progenies of sensitives following autogamy will be sensitive.

In this conjugation, following autogamy of killers, 50% progeny will receive Kk genotype with cytoplasmic kappa particles other 50% progeny will receive kk genotype with cytoplasmic kappa particles. But it will be sensitive, because kappa cannot reproduce in the cells unless a K allele is present in the nucleus and, as a consequence the kappa are eliminated.
On the other hand, in this conjugation the product of autogamy of sensitive strain obtained after conjugation are all sensitive. All through, 50% progeny of autogamy have KK genotype without cytoplasmic kappa particles because no cytoplasm has been transferred in this conjugation. Remaining 50% progeny of autogamy of sensitive's have kk genotype and no cytoplasmic kappa particles. Under some conditions of conjugation persists much longer; a long connection is established between conjugants (killer and sensitive). In this conjugation, cytoplasm as well as nuclear genes are exchanged (Fig. 22.9). As a consequence both ex-conjugants will receive the genotype Kk and the cytoplasm with kappa particles.

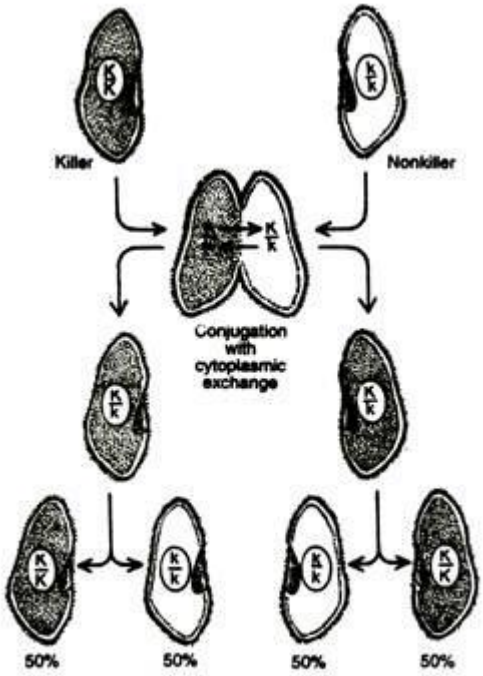


Fig. 22.9: Effect of conjugation for longer period with cytoplasmic exchange followed by autogamy.

Therefore, conjugation for longer period with cytoplasmic exchange will produce all killer strains. Autogamy of both ex-conjugants produces homozygotes KK (killer) and kk (sensitive) cell in the 1 : 1 ratios, respectively, as expected from Mendelian segregation. Therefore, conjugation for shorter period without cytoplasmic exchange does not follow the Mendelian pattern of inheritance. Hence it confirms the cytoplasmic basis of inheritance of killer trait.

Mu Particles in *Paramoecium*:

There is another type of killer trait found in certain strain of *Paramoecium* due to presence of 'mu' particles in the cytoplasm. A *Paramoecium* with a 'mu' particle is called mate killer. On the other hand, a *Paramoecium* having no 'mu' particles is called mate sensitive.

It is so named because when a *Paramoecium* with 'mu' particle conjugates with a partner *Paramoecium* without 'mu' particle then the former kills the latter. The 'mu' particles exist only in those cells whose micronucleus contains at least one dominant of either of the two pairs of unlinked chromosomal genes such as M_1 and M_2. The 'mu' particles are symbionts which are made of DNA, RNA and other substances. The maintenance of the 'mu' symbiont in a *Paramoecium* is dependent upon the genotype of the *Paramoecium*. In fact, the mate-killers of few genotypes maintain their normal number of particles for about seven generations. From the eighth generation, the particles suddenly and completely disappear from the a small fraction of the cell.

Gibson and Beale (1962) suggested that the maintenance of 'mu' particle in *Paramoecium* was due to the presence of another cytoplasmic particle called metagon. It is possibly a long-lived messenger RNA or informosome and may be a product of M_1 and M_2 gene. One metagon may be necessary for the maintenance of hundred 'mu' particles.

Probable Questions:

1. How kappa particle is maternally inherited in *Paramecium*?
2. Describe the role of Mu Particles in *Paramoecium ?*

Suggested Readings:

1. Molecular Cell Biology by Lodish, Fourth Edition.
2. The Cell – A Molecular Approach by Cooper and Hausman, Fourth Edition
3. Principles of Genetics by Snustad and Simmons, Sixth Edition.
4. Molecular Biology of the Cell – by Bruce Alberts
5. Cell and Molecular Biology by Gerald Karp, 7^{th} Edition

UNIT-XIII

Recombination: Homologous recombination, Mechanism of recombination in bacteria and mammals, Gene conversion, Classes of recombinases and types of site-specific recombination

UNIT-XIV

Transpositional recombination, Mitotic and meiotic recombination, Recombination and genomic instability, Application in genetic engineering

Objective: In unit XIII we will discuss about recombination process in bacteria and mammals. We will also discuss about Classes of recombinases and types of site-specific recombination. In XIV unit, Transpositional recombination, Mitotic and meiotic recombination, Recombination and genomic instability, Application in genetic engineering will be discussed.

Definition of Recombination:

The most important features of organisms are to adapt in the environment and to maintain their DNA sequence in the cells generation to generations with very little alterations. In long term survival of organisms depends on genetic variations, a key feature through which the organism can adapt to an environment which changes with time.

This variability among the organisms occurs through the ability of DNA to undergo genetic rearrangements resulting in a little change in gene combination. Rearrangement of DNA occurs through genetic recombination. Thus, recombination is the process of formation of new recombinant chromosome by combining the genetic material from two organisms. The new recombinants show changes in phenotypic characters.

Most of the eukaryotes show a complete sexual life cycle including meiosis, an important event that generates new allelic combinations by recombination. It is made possible through chromosomal exchange resulting from crossing over between the two homologous

chromosomes containing identical gene sequences. Much work was done on eukaryotic genetics until 1945 that laid the foundation of classical genetics. The work on bacterial genetics was done between 1945 and 1965 that advanced the understanding of microbial genetics at molecular level.

Bacterial Recombination:

Bacteria are haploid, therefore do not undergo meiosis. They possess only one double stranded DNA molecule or chromosome. There are several types of genetic recombination in microorganisms. The most common recombination is the reciprocal exchange between homologous DNA sequences.

During genetic recombination usually only a part of the genetic material of a donor cell is transferred to a recipient cell. The DNA of the recipient cell and the donor pair with each other and reciprocally exchange DNA strands by crossing over. This gives rise to a new genetic constitution of the recipient cell with new characters. Subsequent daughter cells that contain only recombined chromosome. There are following main methods by which recombination of genetic material takes place in bacteria.

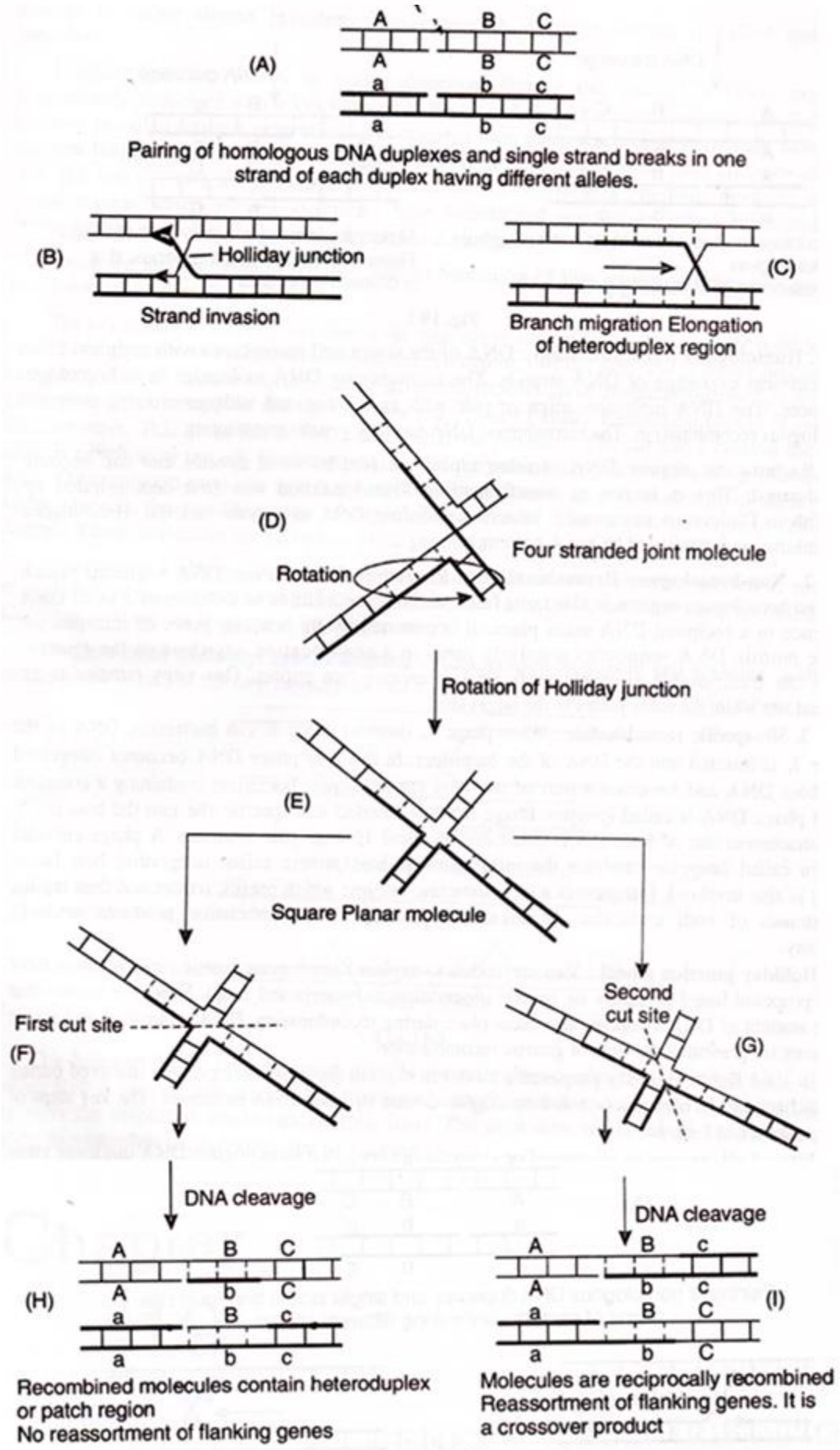

(A)
A B C
A B C
a b c
a b c
Pairing of homologous DNA duplexes and single strand breaks in one strand of each duplex having different alleles.
(B)
Holliday junction
Strand invasion
(C)
Branch migration Elongation of heteroduplex region
(D)
Four stranded joint molecule
Rotation
Rotation of Holliday junction
(E)
Square Planar molecule
First cut site
(F)
Second cut site
(G)
DNA cleavage
DNA cleavage
(H)
A B C
A b C
a B c
a b c
(I)
A B c
A b c
a B C
a b C
Recombined molecules contain heteroduplex or patch region
No reassortment of flanking genes
Molecules are reciprocally recombined
Reassortment of flanking genes. It is a crossover product

Fig. 19.1

Mechanism of Recombination:

Basically, there are three theories viz., breakage and reunion, breakage and copying and complete copy choice that explain the mechanism of recombination (Fig.8.23).

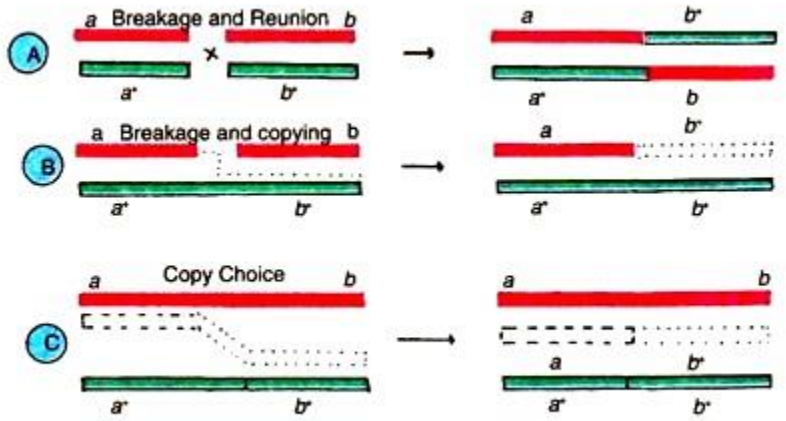


Fig. 8.23 : Three possible mechanisms of recombination.

(i) Breakage and Reunion:

Two homologous duplex of chromosome laying in paired form breaks between the gene loci a and b, and a^+ and b^+ (Fig. 8.23A). The broken segments rejoin crosswise and yield recombinants containing a and b^+ segment, and a^+ and b segment. This type of recombination does not require the synthesis of new DNA. This concept has been used to explain genetic recombination.

(ii) Breakage and Copying:

One helix of paired homologous chromosome (ab and a^+ b^+) breaks between a and b (Fig. 8.23B). Segment b is replaced by a newly synthesized segment copied from b^+ and attached to a section. Thus the recombinants contain and ab^+ and a^+ b^+.

(iii) Complete Copy Choice:

In, 1931, Belling proposed this theory for recombination of chromosome in higher animals. However, it has been questioned by several workers. Therefore, it has only historical importance. According to this theory a portion of one parental strand of homologous chromosome acts as template for the synthesis of a copy of its DNA molecule. The process of copying shifts to the other parental strand. Thus, the recombinants contain some genetic information of one parental strand and some of the other strand (Fig. 8.23 C).

Types of Recombination:

Many kinds of recombination occur in microorganisms.

These are classified basically into the following three groups:

(i) General recombination,

(ii) Non-reciprocail recombination, and

(iii) Site specific recombination.

(i) General Recombination:

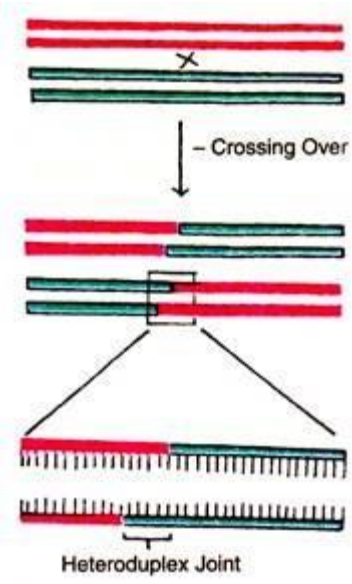


Fig. 8.24 :Formation of hetero-duplex joint consisting of base pairs of two different DNA helices.

General recombination occurs only between the complementary strands of two homologous DNA molecules. Smith (1989) reviewed the homologous recombination in prokaryotes. General recombination in E. coli is guided by base pairing interactions between the complementary strands of two homologous DNA molecules.

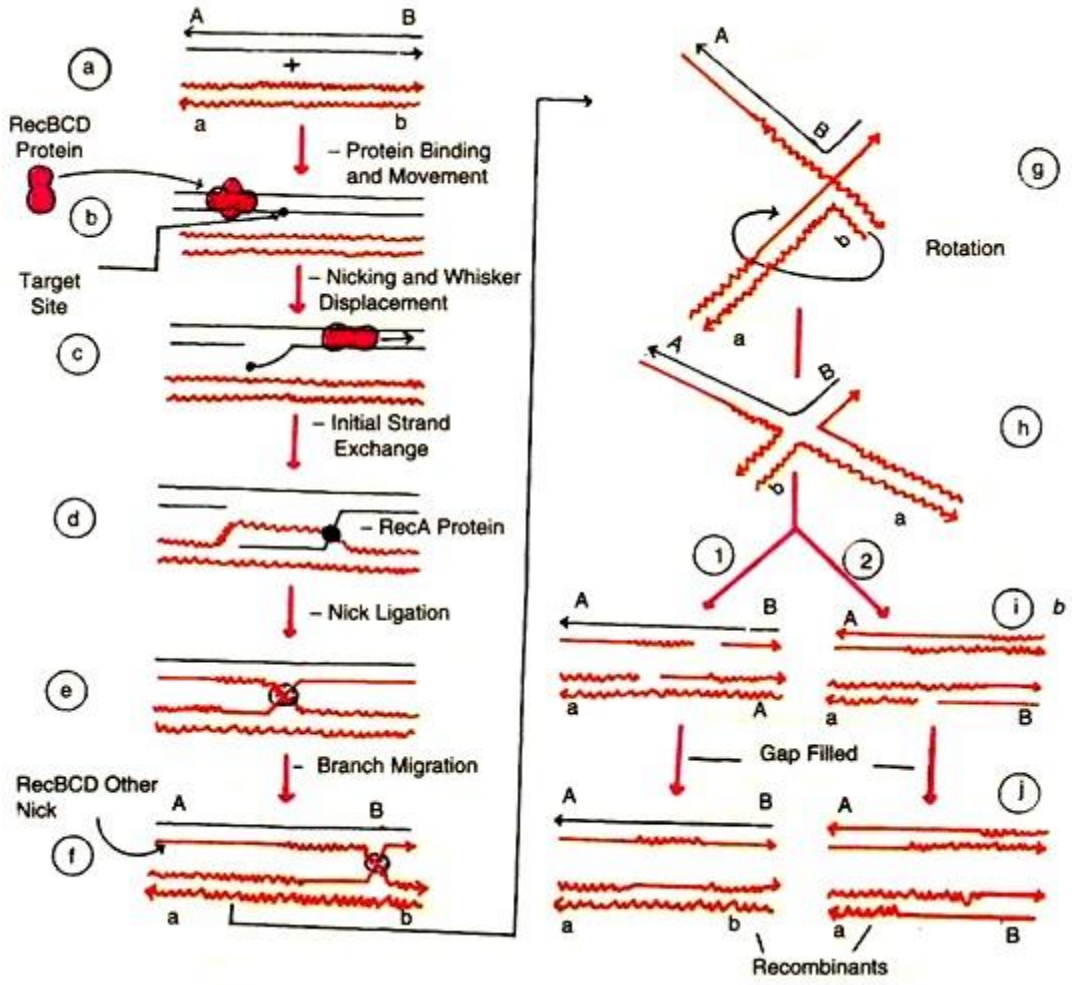
Double helix of two DNA molecules breaks and the two broken ends join to their opposite partners to reunite to form double helix. The site of exchange can occur anywhere in the homologous nucleotide sequence where a strand of one DNA molecule becomes base paired to the second strand to yield heteroduplex just between two double helices (Fig.

8.24). In the heteroduplex no nucleotide sequences are changed at the site of exchange due to cleavage and rejoining events. However, heteroduplex joints can have a small number of mismatched base pairs.

General recombination is also known as homologous recombination as it requires homologous chromosomes. In bacteria and viruses general recombination is carried out by the products of rec genes such as RecA protein. The RecA protein is very important for DNA repair; therefore, it is recA dependent recombination.

Holliday Model for General Recombination:

Holliday (1974) presented a model to show the general recombination (Fig. 8.25). According to this model recombination occurs in five steps such as strand breakage, strand pairing, strand invasion/assimilation, chiasma (crossing over) formation, breakage and reunion and mismatch repair.



(a) Strand breakage:

General recombination occurs through crossing over by pairing between the complementary single strands of DNA duplex (a). Two homologous regions of DNA double helix undergo an exchange reaction.

The homologous region contains a long sequence of complementary base pairing between a strand from one or two original double helices and a complementary strand from the other. However, it is unknown how the homologous region of DNA recognises each other.

A list of recombination genes and their function have been given in Table 8.2. The RecBCD proteins of recBCD or recJ genes are required for recombination in E. coli. This protein enters the DNA from one end of double helix, and travels along the DNA at double helix, at the rate of about 300 nucleotides per second. It creates a loop of ssDNA along travelling DNA (b). It uses energy derived from hydrolysis of ATP molecules. A special recognition site (a) sequence of eight nucleotides scattered throughout E. coli chromosome (b) is nicked in the travelling loop of DNA formed by RecBCD protein.

Gene	*Map location*	*Gene Function*
*rec*A	58	Complete recombination deficiency and many other phenotype defects including suppression of *tif* DNA dependent ATPase.
*rec*B	61	Structural gene of exonuclease V couples ATP hydrolysis to DNA unwinding.
*rec*C	61	Structural gene of exonuclease V
*rec*D	61	α subunit of exo V.
*rec*E	30	Exonuclease VIII, 5'——>3' dsDNA
*rec*F	83	Recombination deficiency of recB⁻ rec C⁻ sbc B⁻ strains blocks UV induction of λ prophage.
*rec*J	64.6	Recombination deficiency of *rec*B⁻ *rec*C⁻ *sbc*B⁻ strains.
*rec*G	82.6	ATPase, distrupts Holiday structure.
*rec*R	11	Help recA utilize SSB-ssDNA complex as substrate.
*rec*O	56	Promotes renaturation of complementary ssDNA.
*ruv*A	41.6	Complexes with Holiday junction.
*ruv*B	41.6	ATPase, dissociates Holiday junction.
*ruv*C	—	Endonuclease, Holiday junction, resolvase.
*ruv*Q	86.5	DNA helicase.
*ruv*L	83	Recombination deficiency of *rec*B⁻ recC⁻ *sbc*B⁻ strain.
*sbc*A	30	Suppress or of *rec*B⁻ and *rec*C⁻ mutations, controlling gene of *rec*E.
*sbc*B	44	Structural gene for exonuclease I

Table 8.2 : Recombination (rec) genes and their function.

(b) Strand pairing:

The RecBCD proteins act as DNA helicase because these hydrolyse ATP and travel along DNA helix. Thus, the RecBCD proteins result in formation of single stranded whisker at the recognition site which is displaced from the helix (c). This initiates a base pairing interaction between the two complementary sequences of DNA double helix.

(c) Strand invasion/assimilation:

A single strand (whisker) generated from one DNA double helix invades the another double helix (d). In E. coli recA gene produces RecA protein which is important for recombination between the chromosomes like single strand binding (SSB) proteins, The RecA protein binds firmly to single stranded DNA to form a nucleoprotein filament.

Roca and Cox (1990) have reviewed the structure and function of RecA protein. RecA protein promotes rapid renaturation of complementary ssDNA hydrolyzing ATP in the process. RecA protein has several binding sites; therefore, it can bind a ssDNA and subsequently a dsDNA. RecA protein binds first to ssDNA, then search for homology between the donor strand and the recipient molecule.

Due to the presence of these sites RecA protein catalyses a multistep reaction (called synapsis) between the homologous region of ssDNA and a DNA double helix. E. coli SSB protein helps the Rec protein to carry out these reactions. When a region of homology is identified by an initial base pairing between the complementary sequences, the crucial step in synapsis occurs. In vivo experiments have shown that several types of complexes are formed between a ssDNA covered with RecA protein and a dsDNA helix. First a non-base paired complex is formed which is converted into a three stranded structure (ssDNA, dsDNA and RecA protein) when a homologous region is found. This complex is unstable and spins out a DNA heteroduplex plus a displaced ssDNA from the original helix. Once the homologous regions are encountered and the ssDNA and dsDNA are complexed, a stable D-loop is formed (d).

(d) Branch migration:

The next step is the assimilation of strand and nick ligation (e). The donor strand gradually displaces the recipient strand which is called branch migration. After formation of synapsis, the heteroduplex region is enlarged through protein-directed branch migration catalysed by RecA protein.

RecA protein directed branch migration proceeds at a uniform rate in one direction due to addition of more RecA protein to one end of RecA protein filament on the ssDNA. Branch migration can take place at any point where two single strands with the sequence make attempts to pair with the same complementary strand.

An unpaired region of the other single strand resulting in movement of branch point without changing the total number of DNA base pairs. Special DNA helicases that catalyse protein directed branch migration are involved in recombination. In contrast, the spontaneous branch migration proceeds in both the directions almost at the same rate. Therefore, it makes a little progress over a long distance.

(e) Chiasma or crossing over formation:

Exchange of a single strand between two double helices is a different step in a general recombination event. After the initial cross strand exchange, further strand exchanges between the two closely opposed helices is thought to proceed rapidly. A nuclease cleaves and partly degrades the D-loop at some points.

At this stage possibly different organisms follow different pathways. However, in most of the cases an important structure called cross-strand exchange (also called Holliday Juncture or chi form or chiasmas, is formed by the two participating DNA helices (g). A chi form of single stranded connections in the cross over region has also been observed under the electron microscope by Dressier and Potter (1982). The chi form of two homologous helices that initially paired and held together by mutual exchange of two of the four strands where one strand originates from each of the helices (g).

The chi form has two important properties, (i) the point of exchange can migrate rapidly back and forth along the helices by a double branch migration, and (ii) it contains two pairs of strands, one pair of crossing strands and the other pair of non-crossing strands.

(f) Breakage and reunion:

The chi structure can isomerize several rotations (h). This results in alteration of two original non-crossing strands into the crossing strands, and the crossing strands into the non-crossing strands. In order to regenerate two separate DNA helices, breakage and reunion in two crossing strands are required.

If breakage and reunion occur before isomerization the two crossing strands would not occur. Therefore, isomerization is required for the breakage and reunion of two homologous DNA double helices resulting from general genetic recombination. Breakage and reunion occur either in the vertical or horizontal plane. If breakage occurs horizontally the recombinants would contain genotype ABlab with a little change in base sequences at the inner region (i).

However, if breakage occurs vertically the recombinants would contain Ab/aB (J). The RurC protein and RecG protein expressed from ruvC and recG genes respectively are thought to be alternative endonucleases specific for Holliday structure.

Enzymes of Homologous Recombination:

There are various proteins that catalyse various steps in the process of homologous recombination in E. coli.

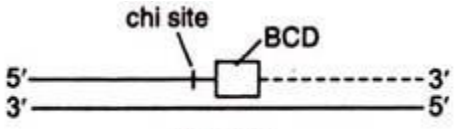


Fig. 19.4.

Enzymes Rec BCD load onto one end of DNA of double stranded break and move along DNA. (Rec-recombination). In the process it unwinds DNA (helical activity) and degrades one or both DNA strands (nuclease activity). Rec BCD is an endonuclease enzyme.

It is encoded by three genes, Rec B, Rec C and Rec D. Rec BCD continues its degrading activity until it reaches a chi site (%). At this point activities of Rec BCD are stopped. The chi site has eight nucleotides. 5′ GCTGGTGG 3′. The chi sites promote recombination.

The single strand DNA tails generated by BCD enzymes are coated by Rec A enzyme. Rec A stimulates pairing or synopsis between two homologous DNA molecules Rec A also promotes strand invasion, displacing one strand of unbroken DNA molecule and forming D-loop. The displaced strand invades the broken DNA molecule. The missing portions of DNA strands are synthesized using homologous strand as template and gaps are sealed by ligase enzyme. Ruv AB enzymes recognize and bind to Holliday junction and performs branch migration. Ruv C enzyme cuts DNA strands at Holliday junction and causes separation and resolution of Holliday junction.

Different biological processes like replication, recombination and repair occur in a co-ordinated manner. In this way new DNA can be synthesized, damaged DNA repaired and genetic recombination takes place. Nucleotide sequences can be replaced through heteroduplexes and gene conversion.

Role of Rec A protein in Homologous Genetic Recombination:

In the various homologous genetic recombination models, the central features are similar in all recombination models.

These include breaks or nicks in DNA molecules. Alignment or pairing or synapsis of homologous sequences of two different DNA molecules. Formation of a crossover structure or Holliday junction in which DNA strand from each molecule creates short regions of heteroduplux DNA. Extension of heteroduplex DNA, which is called branched migration Lastly, resolution of crossover junction to yield end products. This is an extremely complex process involving the action of several different enzymes. The first event of creating breaks

or micks in DNA strands and the last event of resolution are undertaken by various enzymes like helicase, nuclease, and ligases.

But the event starting from pairing of DNA molecules, formation of Holliday junction branch migration are the central features in recombination process. These events are undertaken by a special protein called Rec A protein. Rec A protein is involved in pairing, exchange of strands and branch migration. It is also known as strand exchange protein. Rec A protein plays a major role in homologous recombination. It is a special protein a completely distinct class of enzymes.

Rec A protein binds quickly to single stranded DNA along the phosphate backbone of DNA helix. DNA is completely covered by Rec A protein. Alongside rec A, a second protein called single strand binding protein (SSB protein) is also involved. Each Rec A molecule has 352 amino acids. There is one rec A monomer every 3-4 nucleotides, of DNA. Then, the ssDNA in duplex is aligned with homologous sequence of the other DNA molecule. Several steps occur in this process. Two types of homologous interactions occur. The first is the formation of paranemic joints in aligned homologous strands. The end second interaction involves formation of plactonemic joints.

Rec A protein is a DNA dependent ATPase. ATP hydrolysis is required for branch migration, in which strands are replaced and strand exchange occurs. It exhibits polarity as branch migration proceeds in 5′ -> 3′ direction only.

Exchange of DNA Strands:

Role of strand exchange in post-replication repair is very prominent in E. Coli. Strand exchange plays a very prominent, role in repair of DNA damage. As the advancing replication fork comes across a lesion or damaged site such as thymine dimers, it is bypassed during replication process. The damaged protein may be cleaved which may prove to be lethal.

Repair of this lesion requires conversion of this DNA into double stranded DNA and this is achieved by rec A protein. Rec A protein plays its role in retrieving a portion of the complementary strand from other side of the replication fork to fill the gap. This involves branch migration by Rec A protein. This proves that branch migration is essential activity of the cell.

(g) Mismatch Repair (Mismatch Proof Reading System):

It is such a repair system which corrects mismatched base pairs of unpaired regions after recombination. This system recognises mismatched function of DNA polymerase. The mechanism involves the excision of one of the other mismatched bases along with about

3,000 nucleotides. This RecFJO is involved in the repair of short mismatch either in the initial stage or at the end of recombination. The two proteins MutS and MutL are present in bacteria and eukaryotes. The MutS protein binds to mismatched base pair, whereas MutL scan the DNA for a nick (Fig. 8.26).

When a nick is formed MutL triggers the degradation of the nicked strand all the way back through the mismatch, because the nicks are largely confined to the newly replicated strands in eukaryotes, replication errors are selectively removed. In bacteria the mechanism is the same except that an additional protein MutH nicks the un-methylated GATC sequences and begins the process.

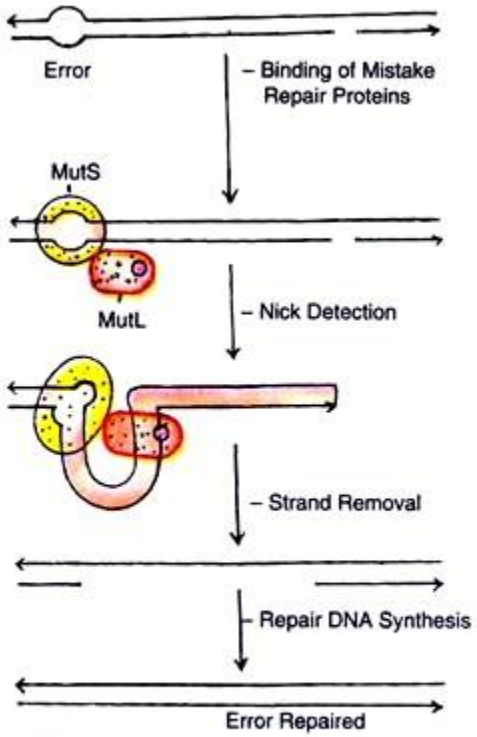


Fig. 8.26 : Mechanism of removal of error in newly made strand by mismatch repair system.

It has been demonstrated in yeast and bacteria that the same mismatch repair system which removes replication errors as in Fig. 8.26 also interrupts the genetic recombination events between imperfectly matched DNA sequences. It is known that homologous genes in two closely related bacteria (E. coli and S.typhimurium) generally will not recombine, even after having 80% identical nucleotide sequences. However, when mismatch repair system is inactivated by mutation, the frequency of such interspecies recombination increases by 100-fold. This mechanism protects the bacterial genome from sequence changes that would be caused by recombination with foreign DNA molecules entering in the cell.

(ii) Non-reciprocal Recombination (Gene Conversion):

The fundamental law of genetics is that the two partners contribute the equal amount of genes to the offsprings. It means that the offsprings inherit the half complete set of genes from the male and half from the female. One diploid cell undergoes meiosis producing four haploid cells; therefore, the number of genes contributed by male gets halved and so the genes of female.

In higher animals like man it is not possible to analyse these genes taking a single cell. However, in certain organisms such as fungi it is possible to recover and analyse all the four daughter cells produced from a single cell through meiosis. Occasionally, three copies of maternal allele and only one copy of paternal allele is formed by meiosis. This indicates that one of two copies of parental alleles has been altered to the maternal allele. This gene alteration is of non-reciprocal type and is called gene conversion. Gene conversion is thought to be an important event in the evolution of certain genes and occurs as a result of the mechanism of general recombination and DNA repair.

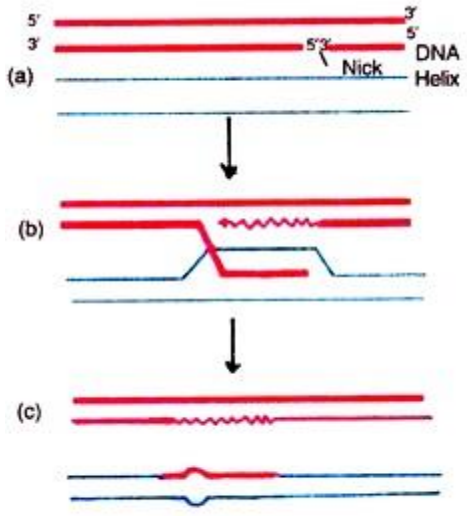


Fig. 8.27 : A model for non-reciprocal general recombination.

Non-reciprocal general recombination is given in Fig. 8.27. Kobayashi (1992) has discussed the mechanism for gene conversion and homologous recombination.

This process starts when a nick is made in one of the strands (a). From this point DNA polymerase synthesizes an extra copy of a strand and displaces the original copy as a single strand (b). This single strand starts pairing with the homologous region as in lower duplex of DNA molecule (b). The short unpaired strand produced in step (b) is degraded when the

transfer of nucleotide sequence is completed. The results are observed (in the next cycle) when DNA replication has separated the two non-matching strands (c).

(iii) Site-Specific Recombination:

Site specific recombination alters the relative position of nucleotide sequences in chromosome. The base pairing reaction depends on protein mediated recognition of the two DNA sequences that will combine. Very long homologous sequence is not required.

Unlike general recombination, site specific recombination is guided by a recombination enzyme that recognises specific nucleotide sequences present on one of both recombining DNA molecules. Base pairing is not involved, however, if occurs the heteroduplex joint is only a few base pair long. It was first discovered in phage λ by which its genome moves into and out of the E. coli chromosome. After penetration phage encoded an enzyme, lambda integrase which catalyses the recombination process (Fig. 8.28). Lambda integrase binds to a specific attachment site of DNA sequence on each chromosome.

It makes cuts and breaks a short homologous DNA sequences. The integrase switches the partner strands and rejoins them to form a heteroduplex joint of 7 bp long. The integrase resembles a DNA topoisomerase in rejoining the strands which have previously been broken.

Site specific recombination is of the following two types:

(a) Conservative site-specific recombination:

Production of a very short heteroduplex by requiring some DNA sequence that is the same on the two DNA molecules is known as conservative site-specific recombination. The detail procedure is described in Fig. 8.28

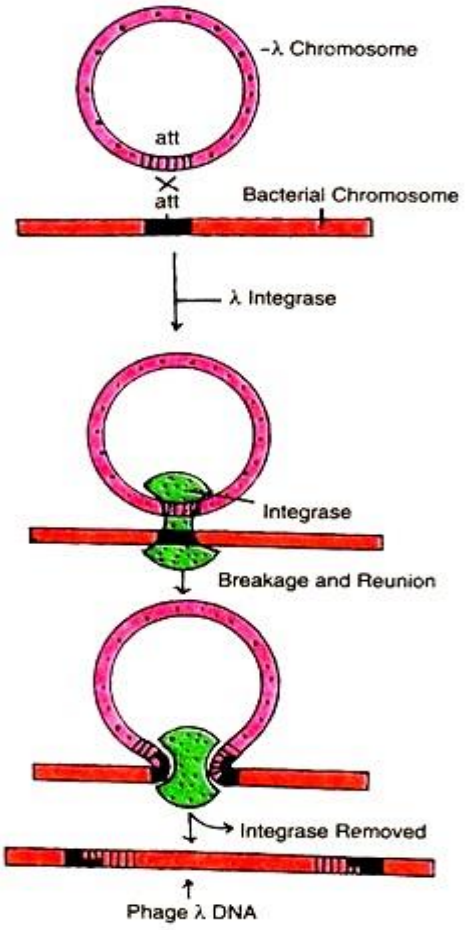


Fig. 8.28 : Diagrammatic representation of site specific genetic recombination.

(b) Trans-positional site-specific recombination:

There is another type of recombination system known as trans-positional site-specific (TSS) recombination. The TSS recombination does not produce heteroduplex and requires no specific sequences on the largest DNA.

There are several mobile DNA sequences including many viruses and transposable elements that encode integrates. The enzyme integrates by involving a mechanism different from phage λ insert its DNA into a chromosome. Each enzyme of integrates recognises a specific DNA sequence like phage λ. K. Mizuuchi (1992a) reviewed the mechanism of trans-positional recombination based on the studies of bacteriophage Mu and the other elements. The enzyme integrase was first purified from Mu. Similar to integrase of phage λ, the Mu integrase also carries out of its cutting and rejoining reactions without requirement of ATP. Also they do not require a specific DNA sequence in the target chromosome and do not form a joint of heteroduplex.

Different steps of TSS recombinational events are shown in Fig. 8.29. The integrase makes a cut in one strand at each end of the viral DNA sequences, and exposes the 3′-OH group that protrudes out. Therefore, each of these 3′-OH ends directly invades a phosphodiester bond on opposite strands of a randomly selected site on a target chromosome. This facilitates to insert the viral DNA sequence into the target chromosome, leaving two short single stranded gaps on each side of recombinational DNA molecule.

These gaps are filled in later on by DNA repair process (i.e. DNA polymerase) to complete the recombination process. This mechanism results in formation of short duplication (short repeats of about 3 to 12 nucleotide long) of the adjacent target DNA sequence. Formation of short repeats is the hall-marks of a TSS recombination.

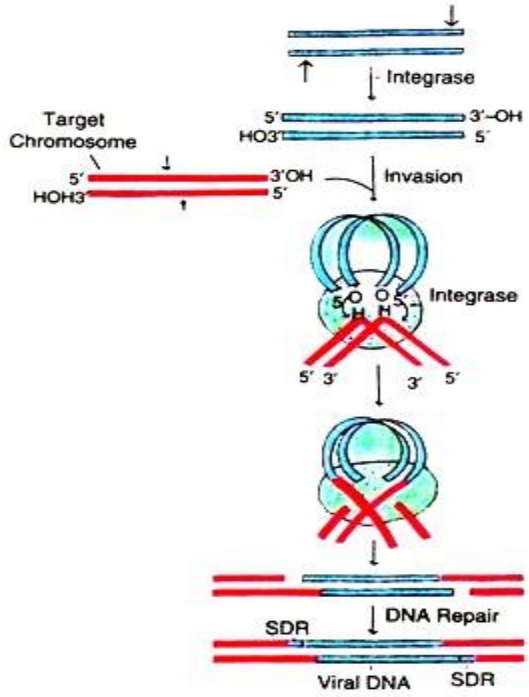


Fig. 8.29 : Mechanism of trans-positional site-specific recombination; SDR, short direct repeats of target DNA sequence.

Probable Questions:

1. How recombination occurs in bacteria?
2. Discuss mechanism of recombination.
3. Discuss Holiday junction model for general recombination.
4. Discuss Role of Rec A protein in Homologous Genetic Recombination
5. What is non reciprocal recombination or gene conversion?
6. Discuss types of site specific recombination.

Suggested Readings:

1. Molecular Cell Biology by Lodish, Fourth Edition.
2. The Cell – A Molecular Approach by Cooper and Hausman, Fourth Edition
3. Principles of Genetics by Snustad and Simmons, Sixth Edition.
4. Molecular Biology of the Cell – by Bruce Alberts
5. Cell and Molecular Biology by Gerald Karp, 7^{th} Edition

DISCLAIMER: This Self Learning Material (SLM) has been compiled from various authentic books, Journals articles, e-journals and other web sources.